On the Role of Valence in Public Health Surveillance and Response

Srinivasan Parthasarathy

Department of Computer Science and Engineering

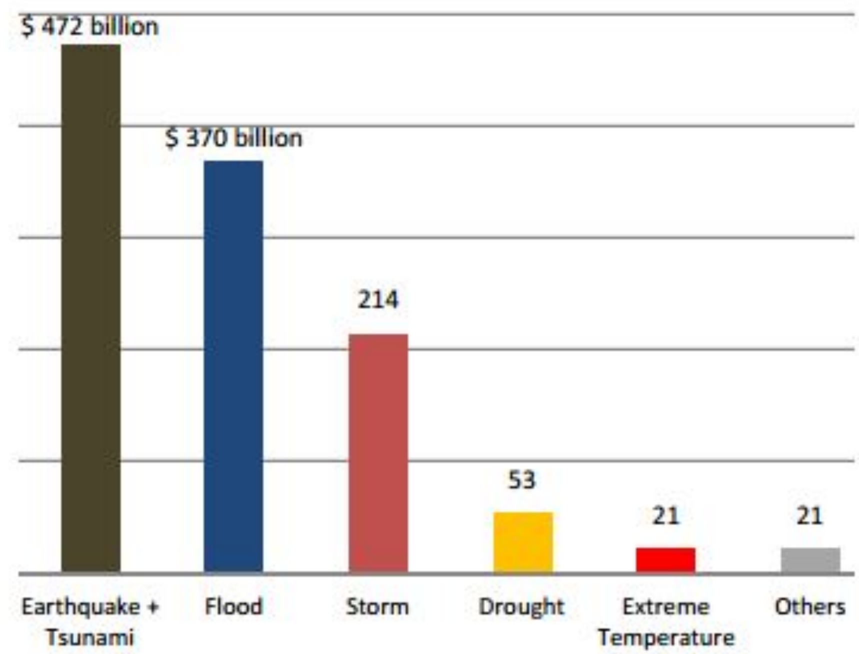
The Ohio State University

Disasters and Epidemics: A Quick Primer

- In 2015 ten weather and climate related disasters that exceeded 1 billion USD in costs – NCEI scorekeeper
  - Hurricanes and Floods are a dominant concern

- Epidemics are also costly
  - Average cost of a Dengue case in Brazil -- $889; Prevalence 270/100,000

Figure 3.3 Economic Losses from Natural Disasters in Asia and the Pacific by Type (1970 – 2013)



Emergency Response and Health Surveillance

- **Emergency Response Informatics** is the study of the use of information and technology in the different phases of disasters or crisis
- **Public health surveillance** is the continuous, systematic collection, and analysis of **health**- data needed for the planning, implementation, and evaluation of **public health** practice

- Key Question: How to combine different modalities and extract useful nuggets for on-the-ground personnel for planning and prioritization purposes in near real-time?

Challenges

A. Need to model trustworthiness of sensing device and modeling framework
   - Need to fuse Physical Sensors, Social Sensors, Models

B. Need to extract, filter and integrate in real time
   - Not all data is equally relevant!

C. Need to model functional intent of sensed or model data
   - More of an issue for social sensing

Focus of this talk: Social Sensing

- Part I: Depression Surveillance: Emotional and Linguistic Cues of Depression in Social Media

- Part II: Predicting Trust Relations Within a Social Network: A Case Study on Emergency Response

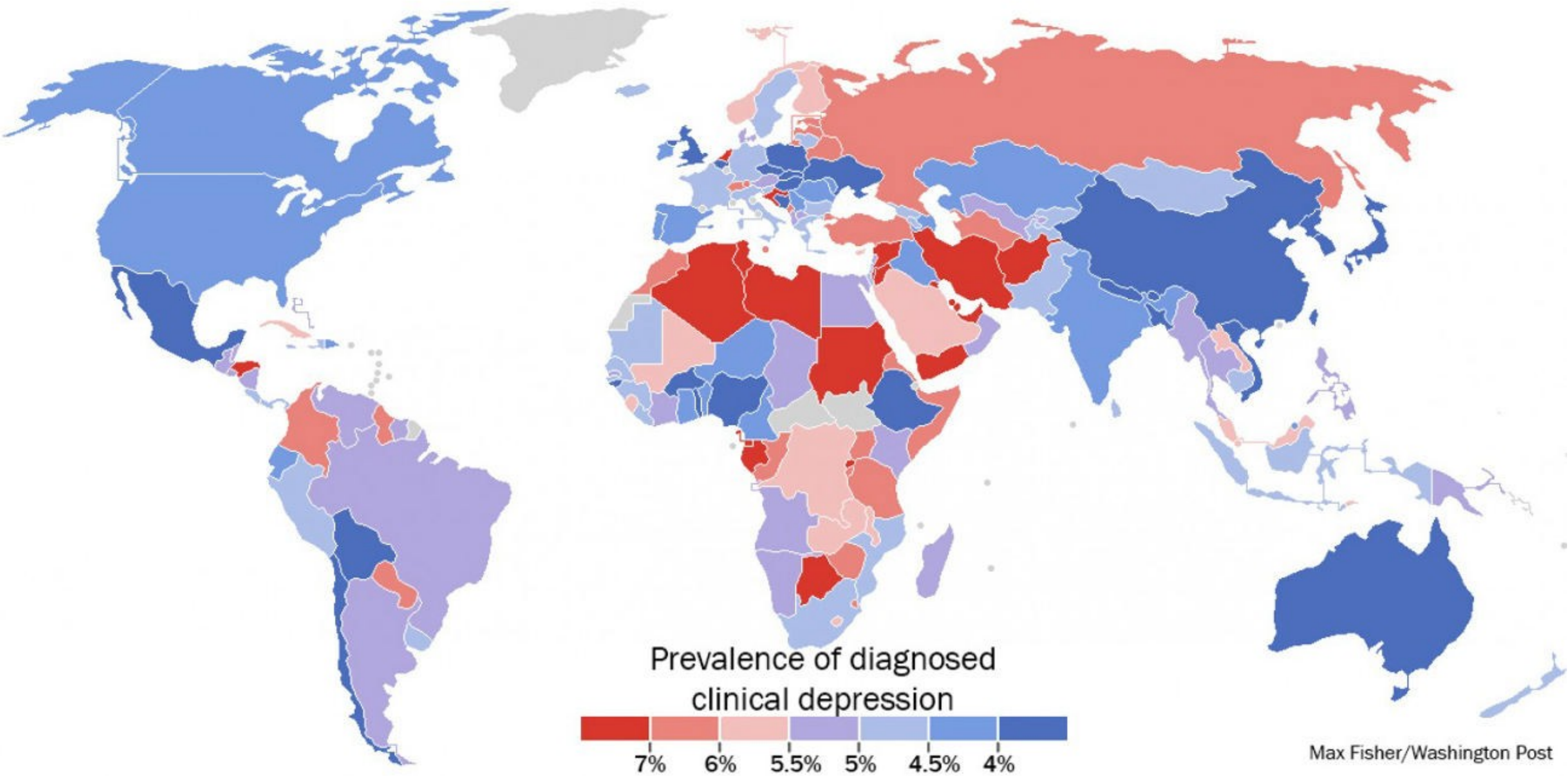
Central role of Sentiment/Valence

I: Emotional and Linguistic Cues of Depression in Social Media

Nikhita Vedula and Srinivasan Parthasarathy

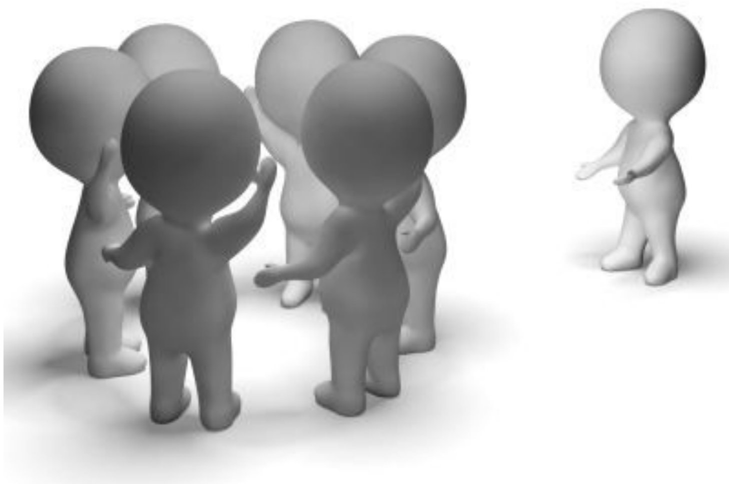
ACM Digital Health 2017

Motivation



Prevalence of diagnosed clinical depression
7%
6%
5.5%
5%
4.5%
4%
Max Fisher/Washington Post

Motivation



A large fraction of users use social media so:

- Can we use social media to detect clinical depression?
- Are online social media communication patterns similar to offline interactions in depressed users?

Gold Standard Data Collection

- Target Users (and Egonet)
  - ‘Depressed’ user profiles (50)
    - Minimize ego-net overlap/interference
    - Each user explicit on clinical anti-depressant use
  - Control group ‘Normal’ users (100)
    - Minimize network interference with depressed users
- Content Data Crawl
  - July 2016 to January 2017
  - Content includes: Target users, 1-hop and 2-hop
  - <u>Purely observational study</u>

Network Activity and Participation

Type	Measure	Depressed	Normal
Activity	# posts daily (period)	5.84 (2041)	7.95 (3145)
	Retweet rate (daily)	4.61	7.28
	Retweet rate (entire period)	1366.54	2742.32
	Mention rate (daily)	1.68	4.25
	Mention rate (entire period)	359.78	1048.45
	Median time of posting	11:51 p.m	5:36 p.m.
	Regional entropy of ego-net [Compton et al., 2014]	3.761	4.483
Specific ego-net properties	Size (1-hop)	1196	3215
	Size (2-hop)	210850	987098
	Density (1-hop)	8.59×10^{-5}	2.67×10^{-5}
	Density (2-hop)	3.44×10^{-7}	1.41×10^{-7}
	User clustering coeff (1-hop)	0.208	0.073
	Eccentricity of user (1-hop)	4.4	2.6

Network Responsiveness to User

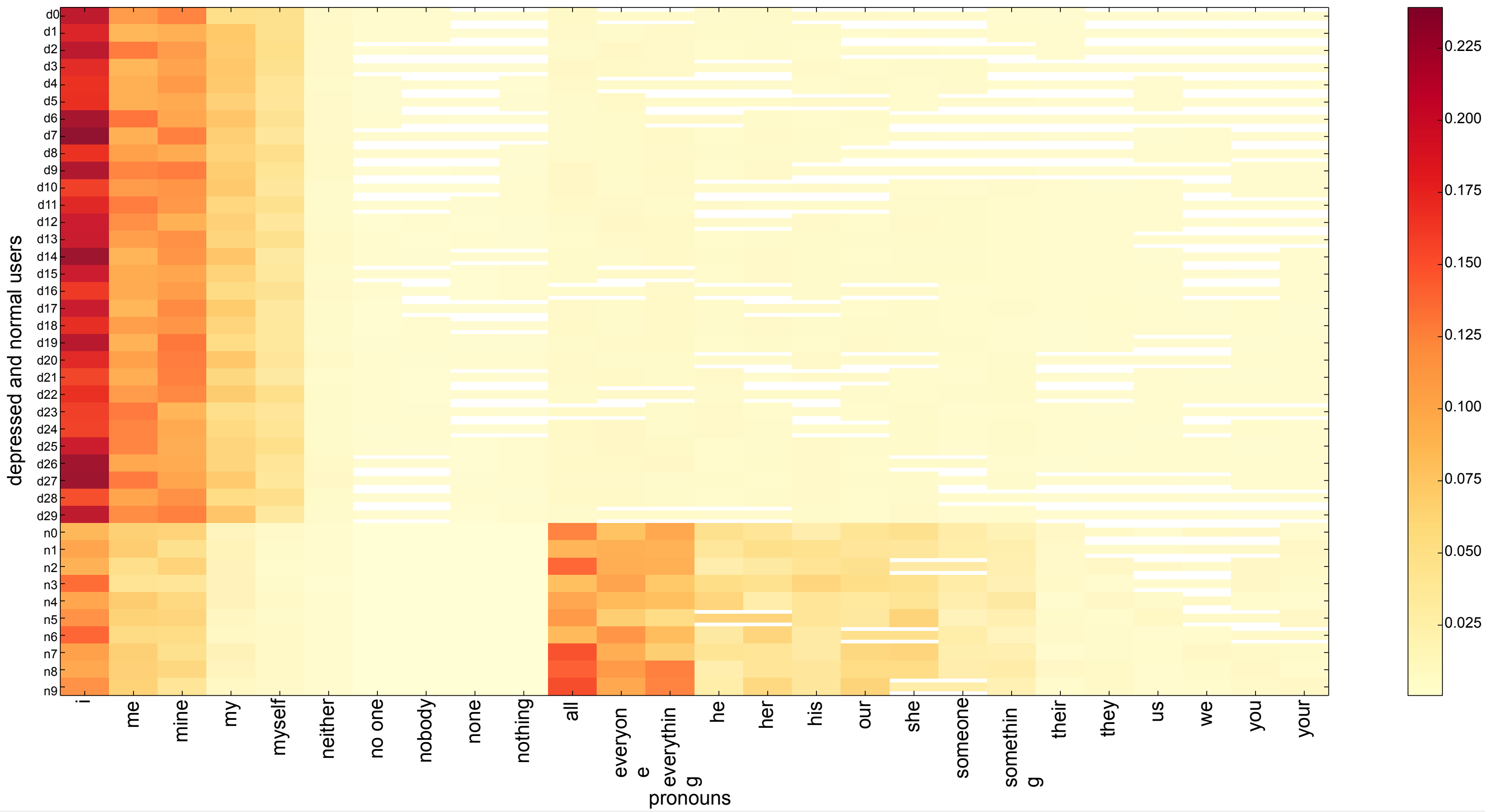
User	% of egonet reacting to user
d0	0.221%
d1	7.69%
d5	13.5%
d15	1.45%
d17	0.56%
d19	4.34%
d21	1.66%
d22	0.989%
d23	0

Theory

- Homophilic [Hogue and Steinberg, 1995, Rosenblatt and Greenberg, 1991]
- Individuals suffering from depression may be socially isolated Joiner et al., 1992]
- Some online users are socially isolated, some are adequately engaged
  - Additional social capital for some!

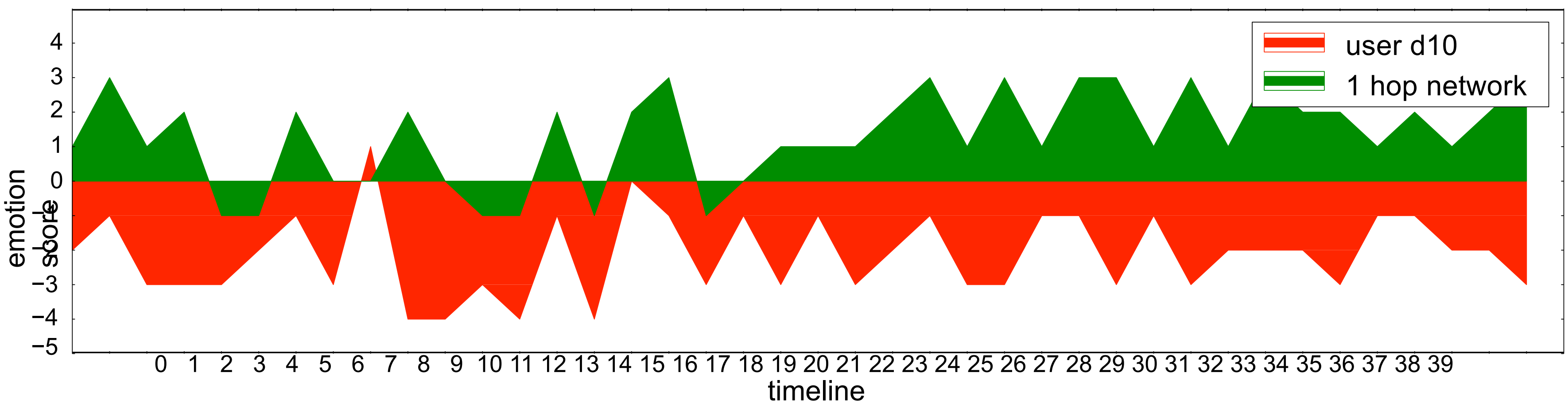
Linguistic Content (Pronoun) Analysis

- Higher self-focus, lesser third-person or collective pronouns [Rude et al., 2004, Ingram et al., 1988, Bucci et al., 1981]



Content Based Emotion Analysis

- Negative emotion dominant for depressed users [Nunn et al., 1997, Ingram et al., 1988].
- Overall emotion of their ego-net is positive.
- Use SentiStrength [Thelwall et al., 2012].



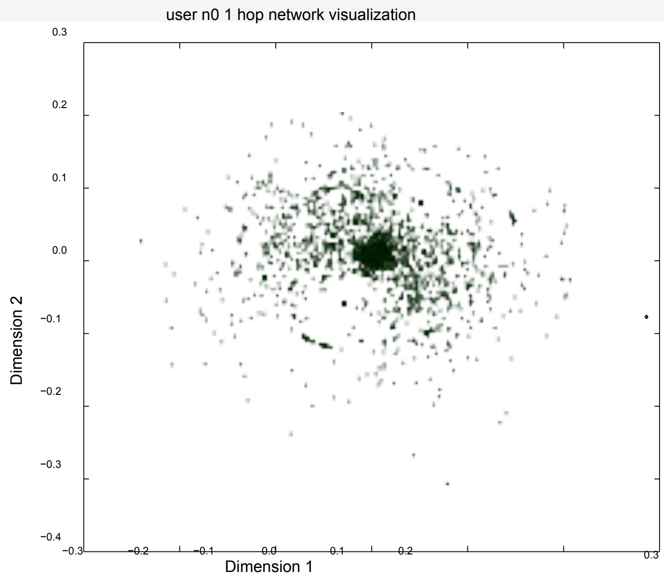
Content Based Emotion Analysis

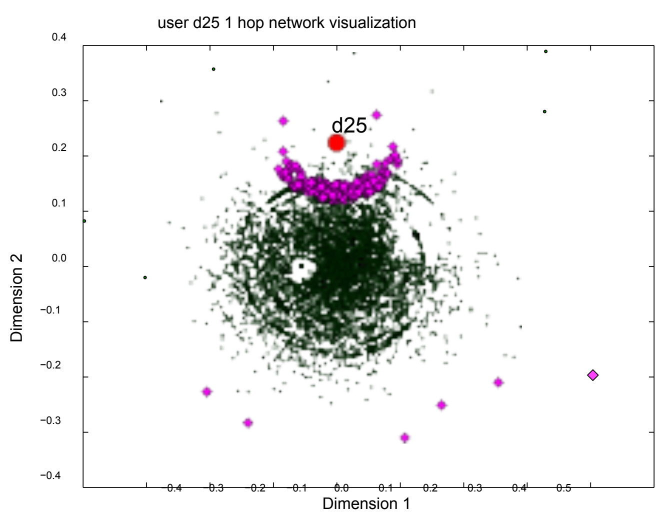
Table 2: Cross correlation values over time w.r.t the daily aggregated avg emotion scores between users and their ego-net.

User [neg, pos]	1-hop [neg, pos]	Corr 1-hop (Lag/Lead)	Time 1-hop (Lag/Lead)	Corr 2-hop (Lag/Lead)	Time 2-hop (Lag/Lead)
d5 [-3, 1]	[-1, 3]	0.209	-1	0.061	-1
d10 [-4, 1]	[-1.5, 3]	0.282	1	0.132	-1
d25 [-3.5, 1]	[-2, 3]	0.277	1	0.218	-1
d29 [-4, 1]	[-1.5, 3]	0.301	1	0.237	-1
n0 [-2, 3]	[-1, 5]	0.516	-1	0.467	-1
n8 [-1, 3]	[-1, 3.5]	0.517	0	0.511	-1

- Low correlation values of depressed users' aggregated avg emotion with their ego-net ($\leq$ 0.275, normal > 0.5)
- No temporal contagion effect for negative emotions for depressed users – unaffected by network [Fowler et al., 2008].
- Normal users tend to follow (lag) their network.

Predictive Model for Depression





- Users represented by word2vec [Mikolov et al., 2013] vectors of their aggregated tweets and visualized using MDS
- Normal users tend to be more central w.r.t ego-net
- Depressed users sit at the fringes.
  - Depressed users also tend to congregate (see pink points) → higher clustering coefficient.

Feature Selection and Predictive Performance

Two sample t-test of significance comparing both user classes

Feature	**p-value**
Average user emotion	**0.000142**
Clustering coefficient	**0.05**
Pronoun usage	**0.004**
User Activity	0.112
No. of mentions	**0.00243**
No. of retweets	**0.00121**
Location entropy of egonet	0.078
% of reaction obtained from egonet	**0.043**
Correlation of user emotion with 1-hop network (Table 2)	**0.00713**
Correlation of user emotion with 2-hop network (Table 2)	**0.00904**

Gradient Boosted Decision Tree (5-fold) Accuracy: 0.90; F1-measure: 0.90

Related Work I

- Social media has been used to study dissemination of health information, diseases and their symptoms [Hawn, 2009, Scanfeld et al., 2010, Paul and Dredze, 2011, Coppersmith et al., 2015].
- Textual content analysis helps identify signs of mental disorders [Bucci and Freedman, 1981, Neuman et al., 2012, Oxman et al., 1982, Pennebaker et al., 2003, Rude et al., 2004, Weintraub, 1981, Ingram et al., 1988].
- Changes in mood and emotional state of individuals is reflected on their social media profiles [Golder et al. 2011, Bollen et al. 2011].

Related Work II

- Various social, psychological, linguistic studies on clinical depression.
- ‘Facebook Depression’ [Jelenchick et al. 2013, Moreno et al. 2011].
- Social media as a tool to study postpartum depression in women [De Choudhury et al., 2013a], and to estimate an individual’s risk of having Major Depressive Disorder [De Choudhury et al., 2013b].

Summary

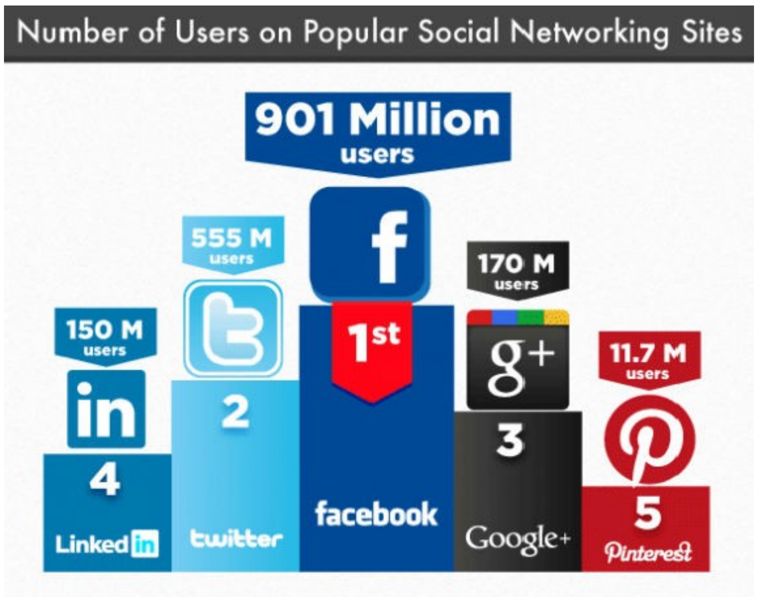
- Wide range of social media signals levered:
  - linguistic style, emotion/valence signals, user engagement, geo-location and network topology
- Significant deviations in depressed user behavior from control group – largely in agreement with offline studies with a few exceptions.
  - social capital from online interactions levered by some depressed users.
- Predictive model achieves an F1-score of 90%.

II. Predicting Trust Relations Within a Social Network: A Case Study on Emergency Response

Nikhita Vedula, Srinivasan Parthasarathy and Valerie Shalin

ACM Web Science 2017

Motivation



Users of popular social media websites as of 2014
(Credit: [Helmick 2014])

- Over last decade:
  - Growth in social media usage
  - Rise of citizen sensors
- Can we use this data for hazard response?
- Need to extract trustworthy signal
  - Trust in facts (not this talk)
  - **Trust in users**
  - **Quantifying trust**

Problem Formulation

- **Input:** Bipartite graph $G = (U \cup T, E)$
  - U = set of users
  - T = set of topics/genres/themes
  - E = connections between the users and topics.
- **Output:** Predict
  - Pairwise trust value between every pair of users.
  - Aggregated rank ordering of users based on trustworthiness.

Overview of Approach

Factors we study pertaining to Trust among users:

- Influence
- Social Cohesion
- Content and Valence
  - **Valence:** used to characterize emotions, ranging from negative to positive.

Influence

- As a proxy measure for trust relationships [Berg et al. 1995, Meyer et al. 1995].
- Measured using the IP algorithm [Romero et al. 2011]

Influence

$$Infl_i = \sum_{j:(i,j)\in E} u_{ij} Passiv_j$$

$$Passiv_i = \sum_{j:(j,i)\in E} v_{ji} Infl_j$$

$$u_{ij} = \frac{w_{ij}}{\sum (k,j)\in E\, w_{kj}} \text{ and } v_{ji} = \frac{1-w_{ji}}{\sum (j,k)\in E(1-w_{jk})}$$

$$w_{ij} = \frac{S_{ij}}{Q_i}$$

Q_i = no. of posts by i, S_{ij} = no. of posts by i and re-posted by j.

Related Ideas

Influence computation

- Twitter-Rank [Weng et al. 2010] and Trust-Rank [Wu et al. 2006]: pagerank style, using global network structure
- Twitter-specific measures of influence [Weng et al. 2010, Kwak et al. 2010, Cha et al. 2010]: no. of followers, retweets and mentions, user pagerank
- Influence-Passivity (IP) [Romero et al. 2011]: uses frequency with which users' posts are re-posted, user passivity and prior content history of posts.
- Influence Maximization [Kempe et al. 2003, Barbieri et al. 2013, Aslay et al. 2014]: identifies seed users such that the no. of network users they influence is maximized.

Social Cohesion Based User Behavior

- Members of a ‘group’ share emotional and behavioral characteristics [Lott and Lott, 1965]
- Strong correlation between user similarity and trust [Ziegler and Golbeck, 2007]
- Jaccard similarity: approximates local triangle density (community cohesiveness) – fast estimation via LSH [Broder’98]

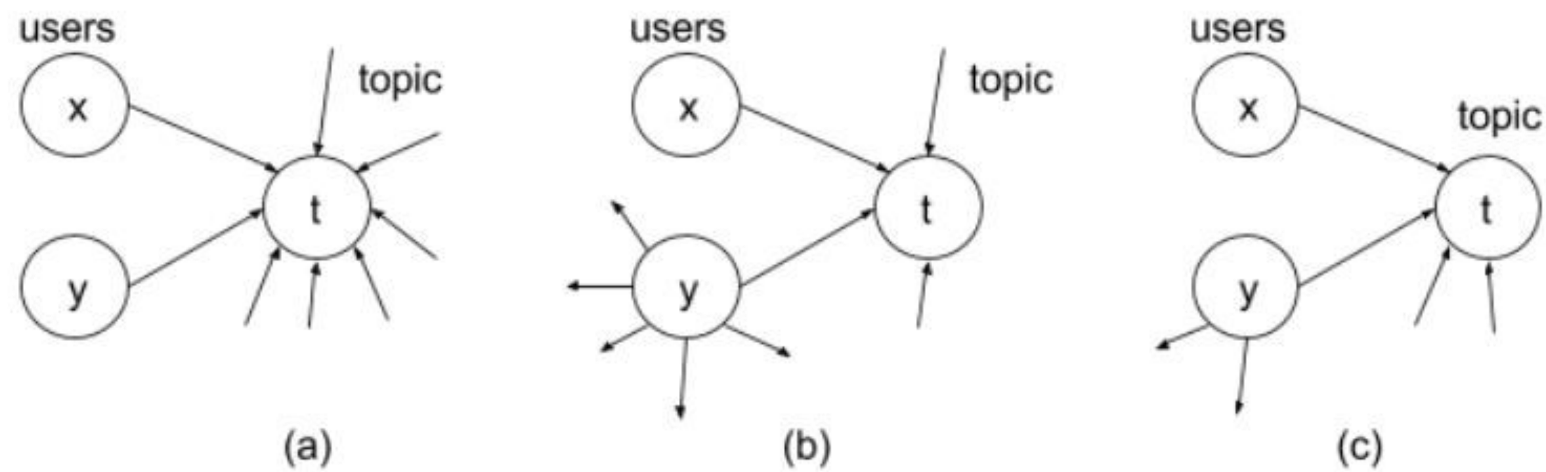
$$Jacc(x,y) = \frac{V_x \cap V_y}{V_x \cup V_y}$$

V_x = vector of topics to which user x has a directed edge.

Content and Valence

- Positive interactions and interpersonal agreement are critical to trust [Lott and Lott, 1965]
- *Shared topic*: Both users post on it
  - With and without valence
- Content valence measured using SentiStrength [Thelwall et al. 2012]
- Accounts for the relative *popularity* of the topics

Content and Valence: Degree Discounting



Intuition: Similarity between x and y:

(a) Popular topic (less important)

(b) Involves non-discriminative user (y) (less important)

(c) Discriminative users on focused topic (important)

Degree Discounting based user similarity

$$sim_d(x,y) = \frac{1}{\sqrt{D_o(x,x)}\sqrt{D_o(y,y)}} \sum_t \frac{A(x,t)A(y,t)}{\sqrt{D_i(t,t)}}$$

Pairwise/Global Trust Computation

Trust Equation

$$Trust(x,y) = \alpha Infl(y,x) + \beta Jacc(x,y) + \gamma sim_d(x,y)$$

such that $\alpha + \beta + \gamma = 1$.
α, β and γ are regularization parameters, learned using a grid search.

[Above]: Pairwise (directional) trust relationship

Develop a global user ranking via:

- Normalized aggregation of pairwise trust of x from all other users.

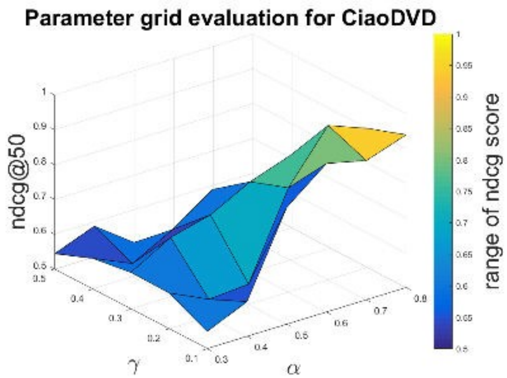
Evaluation: Datasets

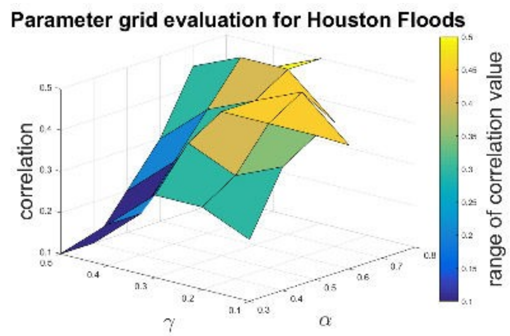
Dataset	#Users	#Topics	#Edges	#Tweets
India Anti-Corruption	2104	15	7180	100K
Mumbai Blast	466	10	932	10K
Phone and Tablet	7506	15	16265	100K
Houston Floods	301	15	875	100K
Hurricane Irene	6960	20	17593	200K
Nice Attack	57470	20	166943	800K

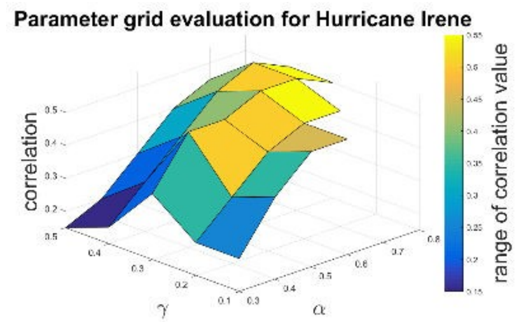
Dataset	#Users	#Topics	#Edges	#Trust relations
CiaoDVD	7375	17	111781	40133
Epinions	114467	27	442787	717667

Evaluation: Parametric Grid Search: Valence

W/O

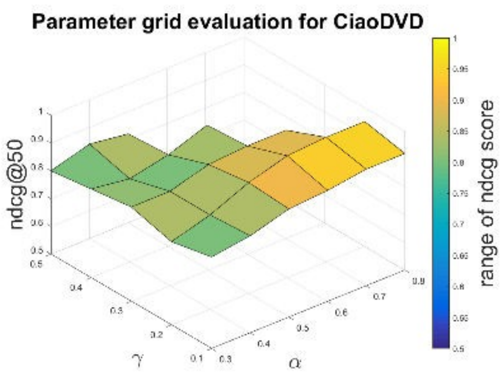


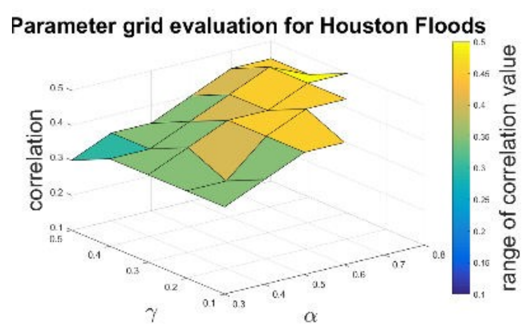


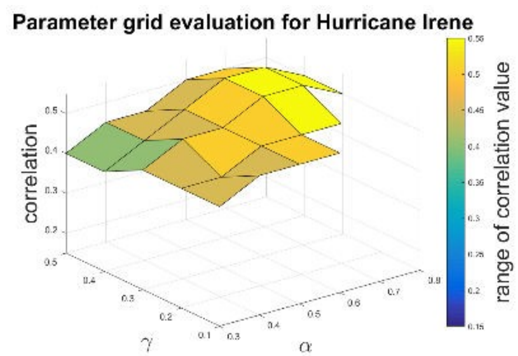


Correlation of Length of conversation between pairs of users as an approximate measure of mutual trust [Adali et al. 2010, 2012].

With







Valence matters for quality and algorithmic stability!

Evaluation: Factor Impact and Analysis

Dataset	α	β	γ
India Anti-Corruption	0.6	0.05	0.35
Mumbai Blast	0.65	0.05	0.3
Phone and Tablet	0.7	0.2	0.1
Houston Floods	0.65	0.1	0.25
Hurricane Irene	0.65	0.05	0.3
Nice Attacks	0.7	0.05	0.25
CiaoDVD	0.7	0.1	0.2
Epinions	0.65	0.1	0.25

Table 1: Tuned parameter values whose trust ranking best matches with ground truth

- Strongest factor contributing to Trust = Influence (α)
- Next strongest = content/valence based user similarity (γ)

Baseline Algorithms for Trust Computation

1. EigenTrust [Kamvar et al. 2003] and MoleTrust [Massa 2005]: Peer-to-peer global trustworthiness.

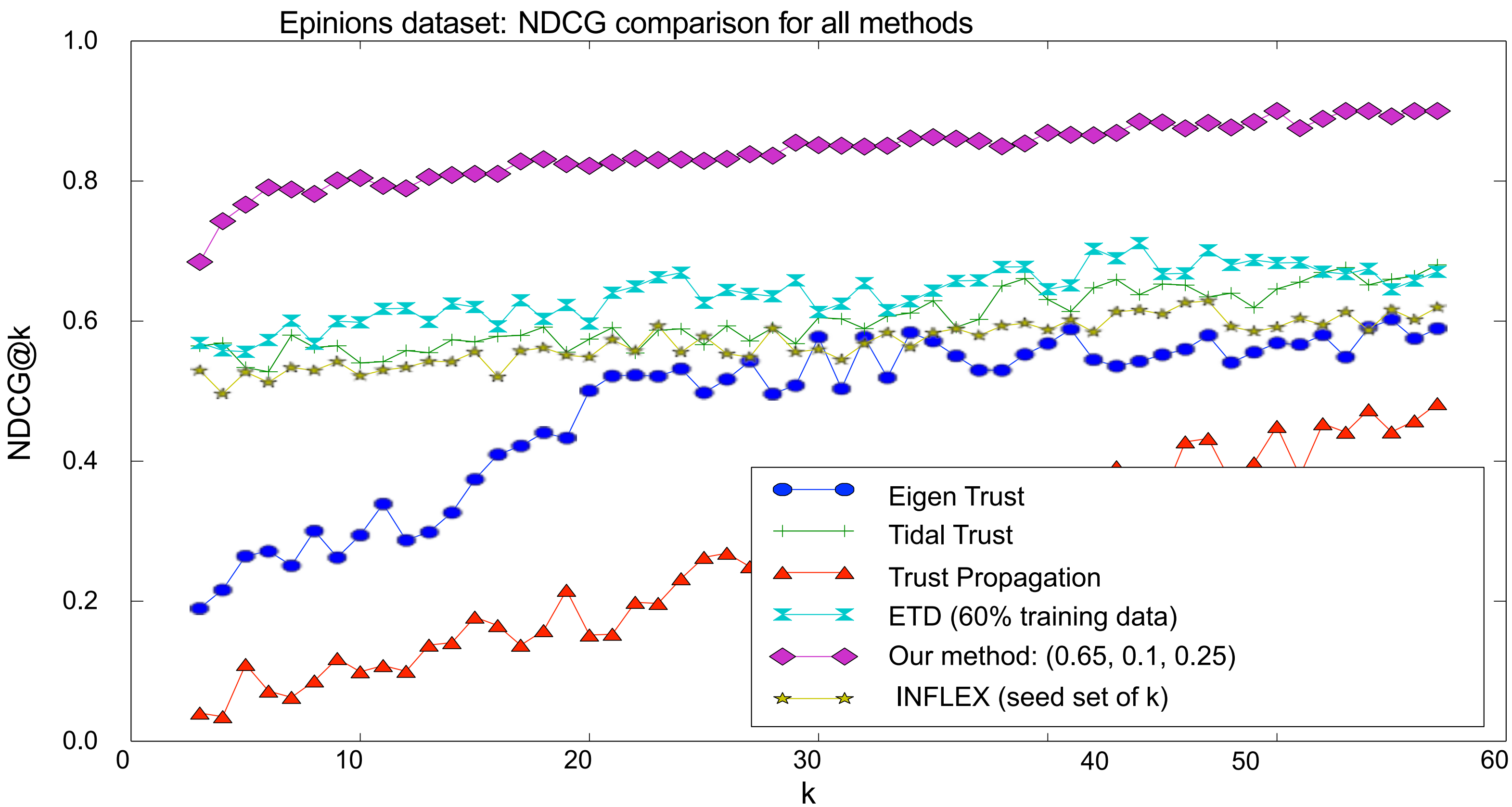
2. Rule-based network trust propagation [Guha et al. 2004]

3. TidalTrust [Golbeck et al. 2006]: BFS based trust propagation.

4. INFLEX – Topic aware trust (influence) modeling [Aslay'14]

5. ETD [Beigi et al. 2016]: supervised, uses similarity in emotion expressed by users as a signal of trust between them.

Evaluation: Comparison with baselines: Epinions



Our method significantly outperforms baselines

Evaluation: Comparison with Baselines: Social Media

Algorithm	**Avg NDCG@50 for Twitter datasets**	**Avg F-score for Twitter datasets**
EigenTrust	0.444	N/A
TidalTrust	0.549	0.511
TrustPropagation	0.321	0.303
INFLEX	0.647	N/A
Our method	**0.876**	**0.803**

Our method significantly outperforms baselines

Case Study

Trust Dynamics and Crisis Response: Hurricane Irene

Before	During	After
mashable	cnnbrk	cnnbrk
lfmccullough	BreakingNews	BreakingNews
richmintz	USArmy	BarackObama
FrommersTravel	severestudios	HumaneSociety
nydailynews	shibanijoshi	RedCross
6abc	JimCantore	JimCantore
travelingmoms	xanpearson	*atmanes*
severestudios	CraigatFEMA	CraigatFEMA
BreakingNews	MikeBloomberg	*Fanua*
Daily_Press	*atmanes*	RedCrossPhilly
cnnbrk	*Fanua*	SamaritansPurse
funnyordie	HumaneSociety	RedCrossSAZ
afreedma	BarackObama	mashable
BarackObama	RedCross	USGS

Case Study

Trust Dynamics and Crisis Response: Houston Floods

Before	During	After
GlitchxCity	Breaking911	Breaking911
Sportsnaut	NWSHouston	NWSHouston
TriCityHerald	AlertHouston	BarackObama
Nick_Anderson_	WSOCWeather	AlertHouston
HOUBizJournal	StormViewLIVE	*TexasTsunami*
StarfishGawdess	BarackObama	RedCross
Breaking911	ArchCollegeTAMU	ArchCollegeTAMU
BarackObama	JohnCornyn	WSOCWeather
JohnCornyn	*TexasTsunami*	JohnCornyn
NWSHouston	GlitchxCity	GlitchxCity

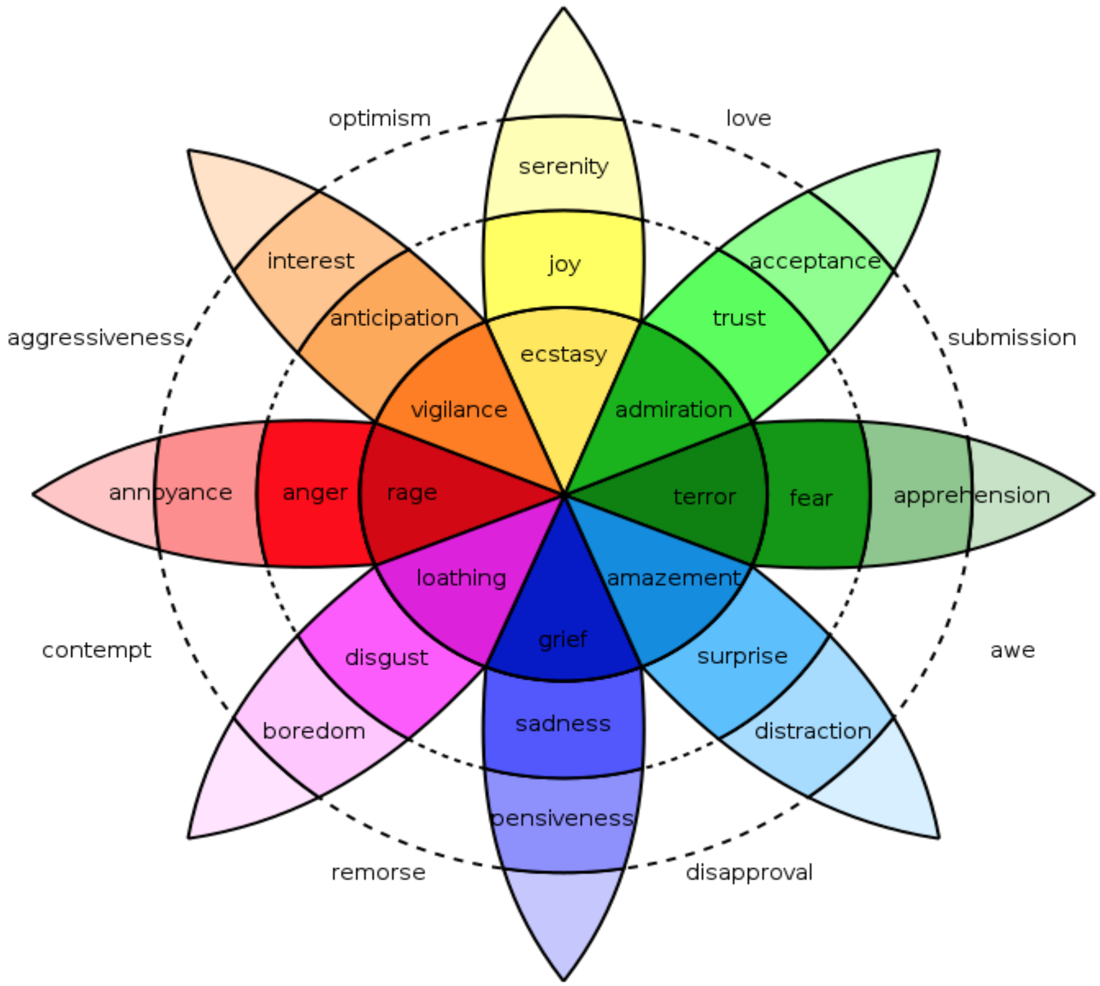
Summary

Main Contributions

- Influence, followed by content and valence-based user similarity are important in trust estimation.
- Valence agreement among users increases the robustness of the method under various parametric settings.
- Our method outperforms trust computation baselines on real-world datasets.
- It captures trustworthy users who are well known/popular.
- Effectively identifies emergent, newly trustworthy users.

III. Ongoing and Future Directions

How can one detect and lever personal state



Robert Plutchik's Wheel of Emotions (1980)

Dengue Surveillance in Brazil

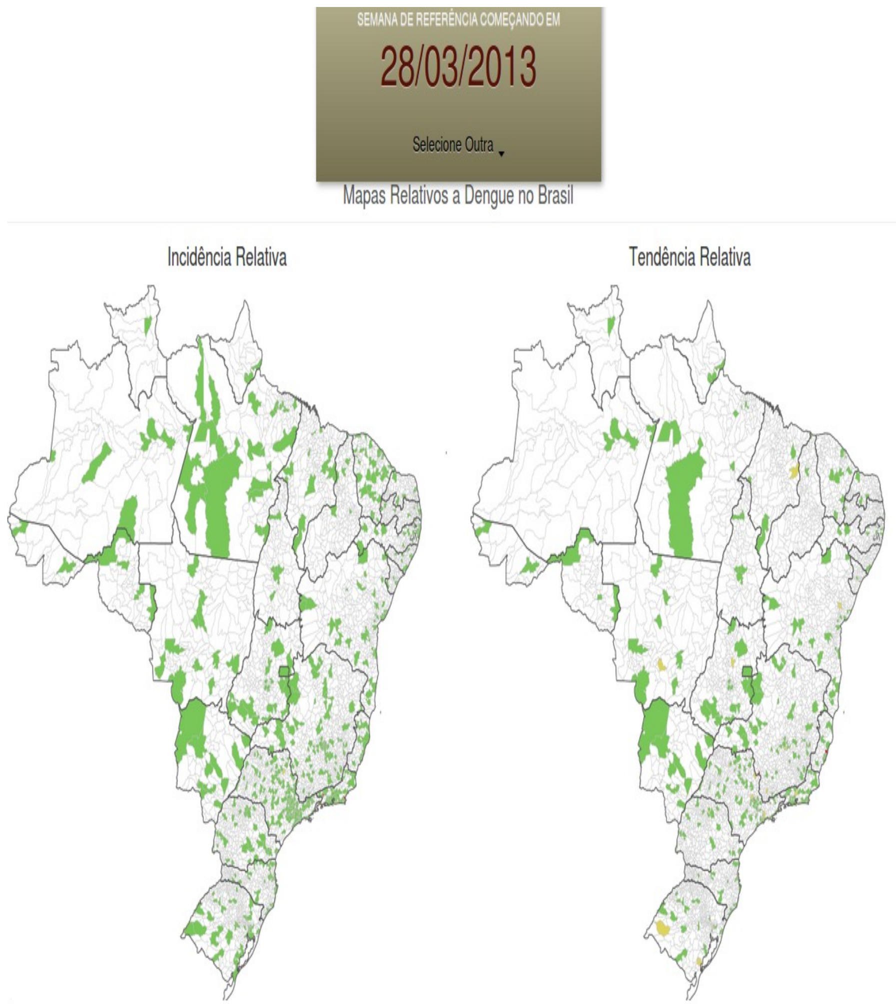
Mosquito-borne infection that causes a severe flu-like illness, and sometimes a potentially lethal complications

- 390 million affected per year worldwide

Early prediction → Better Resource Management

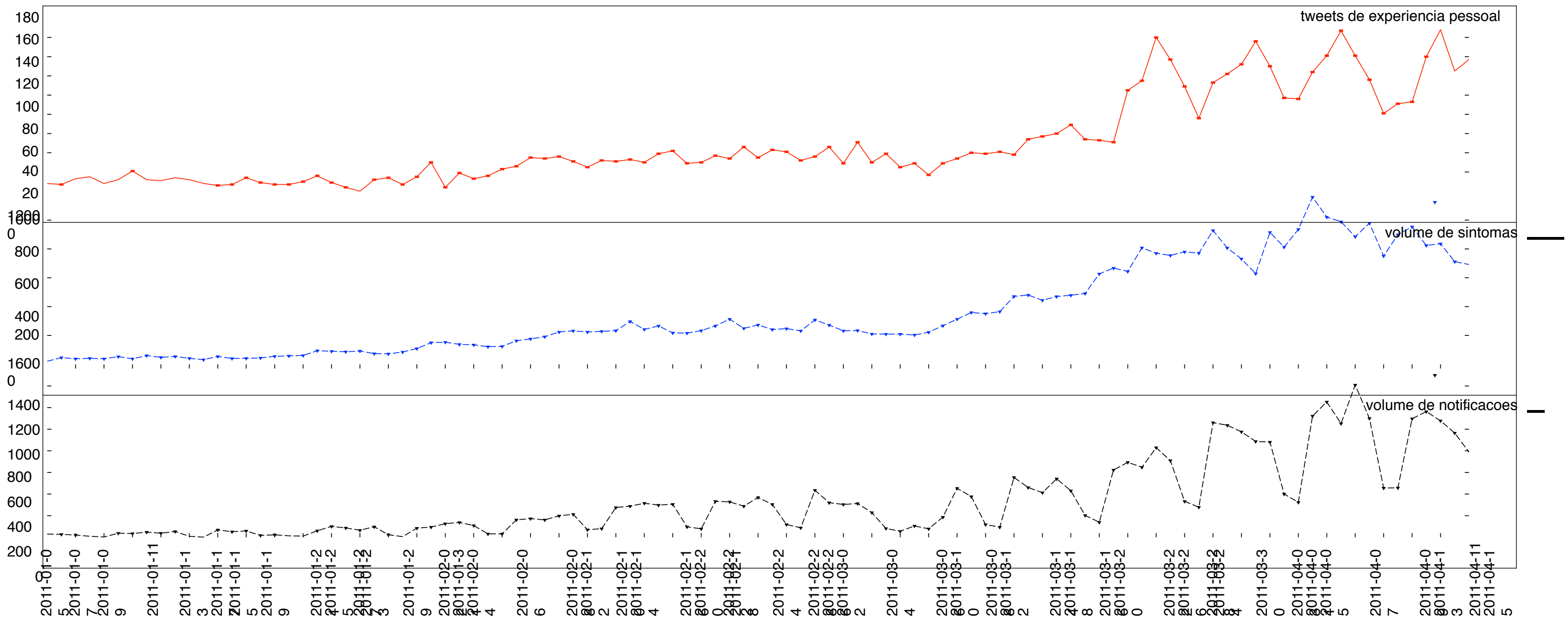
- – Current practice based on insect surveillance – slow

Can social sensing help? Yes!



Drill Down: Rio de Janeiro Dec’10-May’11 [Meira’12]

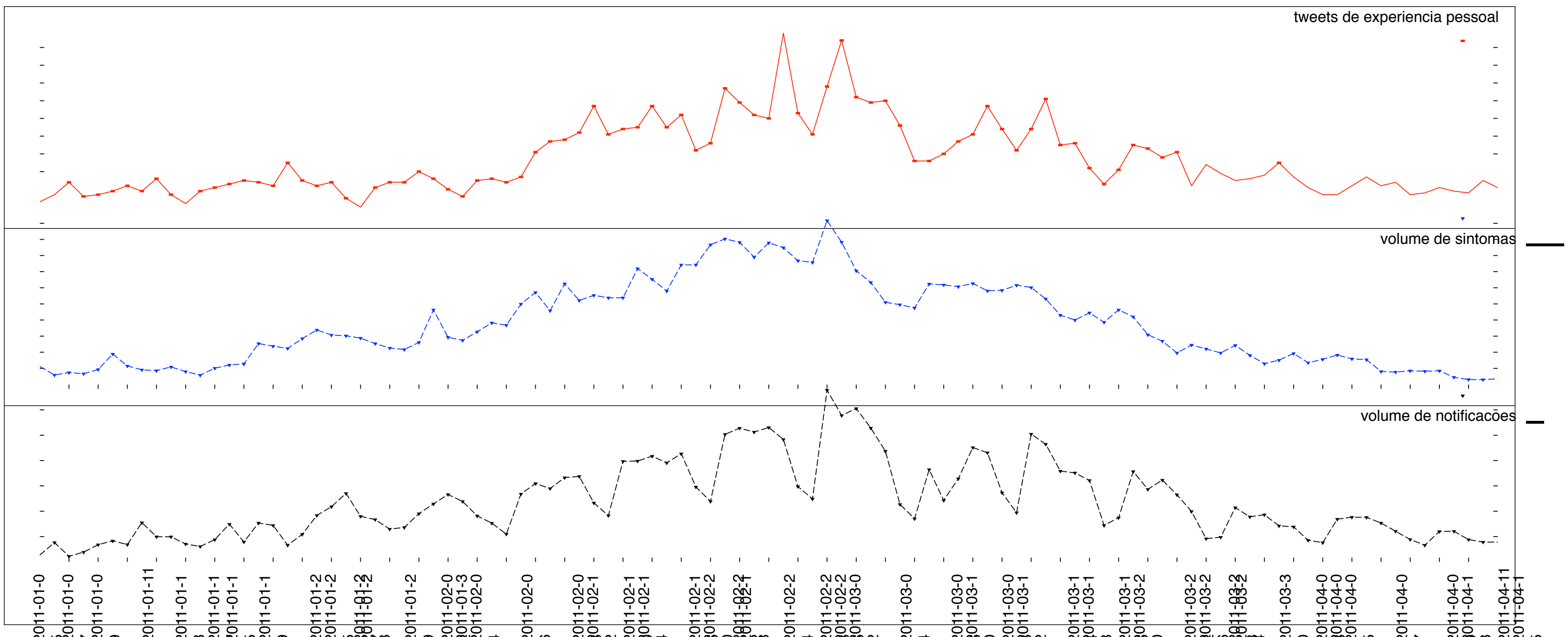
Personal experience and valence, notifications and **symptom perception**



Highly Correlated (>0.88); Cross-Correlation maximized ~2 days prior to notification
Signal observable 6-8 days prior to notificaton.

Drill Down: Manaus Dec'10-May'11

Personal experience (from social media), notifications and **symptom perception**



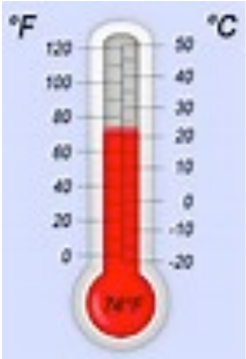
Dengue prevalent earlier in the season but correlation still strong (>0.9)

Towards a Theory of Latent Pragmatics

- *Can we go from social data to represent environmental change according to known functions and standardized measurement units?*

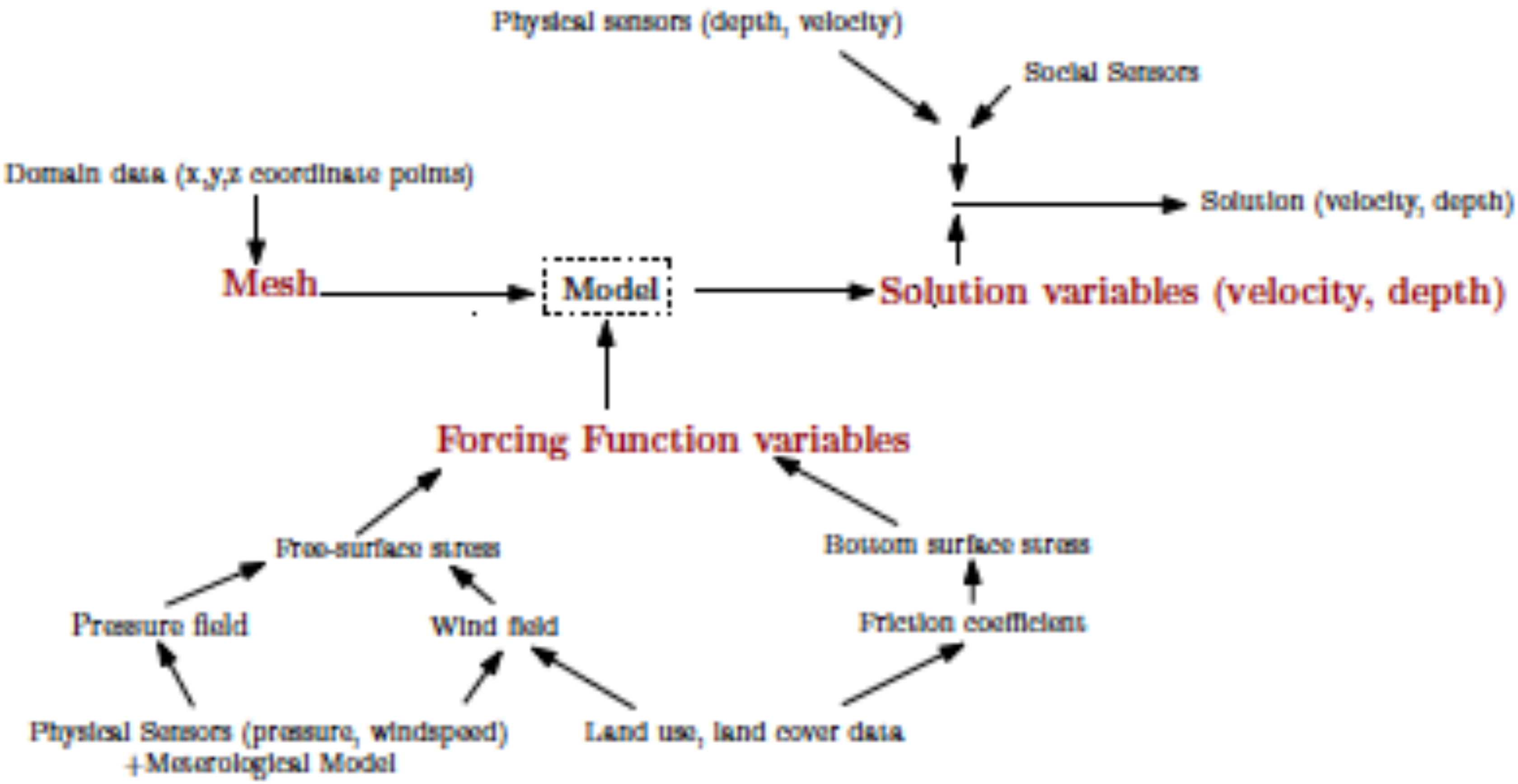
- *How is measured change in the world reflected in human language during an emergency?*

- *In other words what is the functional intent?*

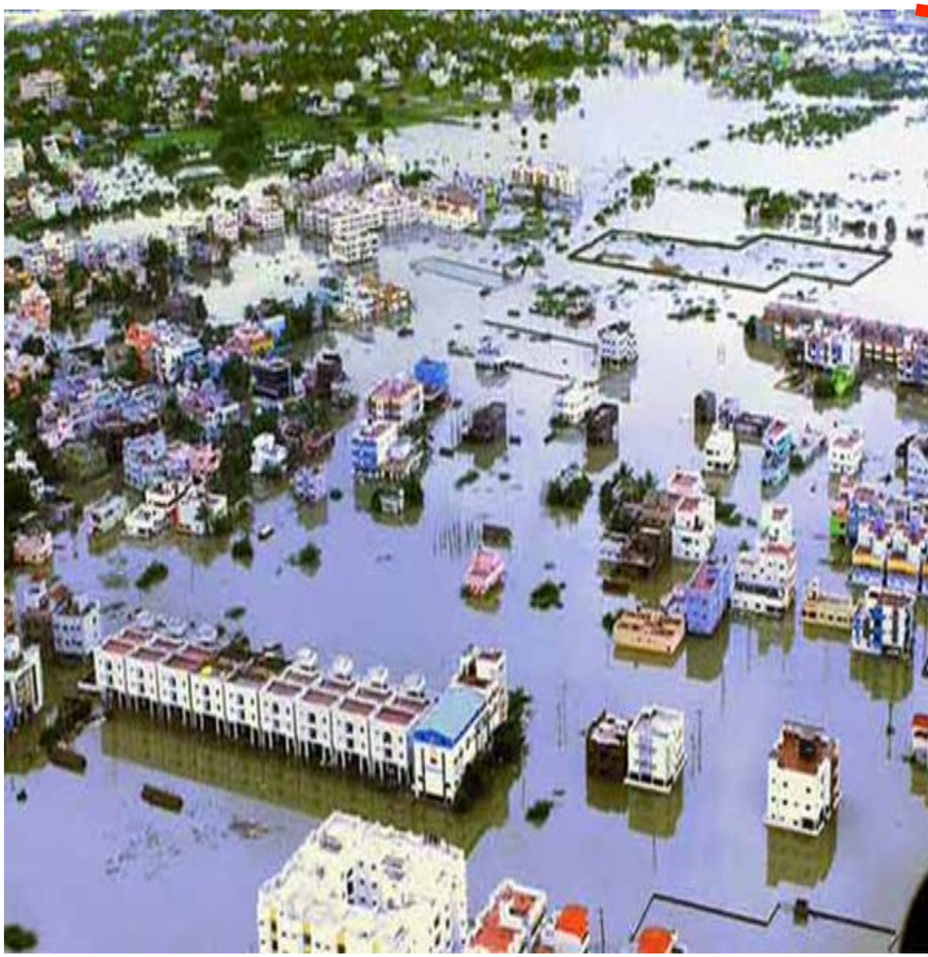





Adaptive Models for Storm Surge [From Citizen and Physical Sensing]



Chennai Floods (2015): Social Sensing Enhanced Flood Mapping



After 3rd Depression

Take Home Message

- Social sensing a critical part of the ER/PHS equation
  - Along with physical sensing, modeling and analysis
- Must model valence and emotional state
  - Accounting for linguistic, social network theory essential
  - Latent Pragmatics – a key direction for the future
- Demonstrable benefits for end applications
  - Dengue surveillance (deployed)
  - Emergency response (field tested)
  - Flood mapping (ongoing)

THANK YOU

Organizers: Erik, Iti and other organizers of this workshop

Audience: For your attention

Students: N. Vedula, J Liang, P. Jacobs,

Colleagues: V. Shalin, E. Kubatko, A. Sheth, T. Prasad, W. Meira, D. Liu, C. Browning

Acknowledgments: This work is supported by the grants NSF-EAR-1520870 and NIH-1R01 HD088545-01A1