# EECS 6895 Advanced Big Data and AI

## Lecture 11: Perception, Emotion, Feeling, and Personality

Prof. Ching-Yung Lin

Columbia University

April 8th, 2025

# The Socratic Method

Columbia University

Knowledge cannot be simply imparted, but must be discovered through a process of questioning and dialogue.

- Posing open-ended questions: The teacher or facilitator starts with a question to stimulate thinking and draw out ideas.
- Clarifying key terms: The teacher helps the students clarify and define relevant terms and concepts to ensure everyone is on the same page.
- Providing examples and evidence: The teacher or facilitator encourages the students to provide examples and evidence as reasons to support their claims.
- Challenging reason-to-conclusion argument: The teacher or facilitator challenges the students' arguments and encourages them to question their own beliefs and to consider alternative perspectives.
- Summarizing and drawing conclusions: The teacher helps the students summarize and draw conclusions from the discussion.
- Reflecting on the process: The teacher and students reflect on the effectiveness of the method and what they learned through the dialogue.

# SocraSynth

- A paradigm designed to infuse AI systems with advanced cognitive reasoning through Socratic dialogues within a Multi-LLM framework.

- Showed significant transition from monologues to dialogues in LLM collaborations:
  - Illustrating improvements in question quality
  - Marked by increased relevance, depth, clarity, and novelty
  - Achieved through iterative dialogic exchanges.

- SocraSynth can be used for sales planning, disease diagnosis, content creation, and geopolitical analysis, etc.

- Potentially revealing a new era in the application of LLMs.

# Contentiousness Levels ➔ Personality?

| C.L. | Tone | Emphasis | Language |
|---|---|---|---|
| 0.9 | Highly confrontational; focused on raising strong ethical, scientific, and social objections. | Highlighting risks and downsides; ethical quandaries, unintended consequences, and exacerbation of inequalities. | Definitive and polarizing, e.g., "should NOT be allowed," "unacceptable risks," "inevitable disparities." |
| 0.7 | Still confrontational but more open to potential benefits, albeit overshadowed by negatives. | Acknowledging that some frameworks could make it safer or more equitable, while cautioning against its use. | Less polarizing; "serious concerns remain," "needs more scrutiny." |
| 0.5 | Balanced; neither advocating strongly for nor against gene editing. | Equal weight on pros and cons; looking for a middle ground. | Neutral; "should be carefully considered," "both benefits and risks." |
| 0.3 | More agreeable than confrontational, but maintaining reservations. | Supportive but cautious; focus on ensuring ethical and equitable use. | Positive but careful; "transformative potential," "impetus to ensure." |
| 0.0 | Completely agreeable and supportive. | Fully focused on immense potential benefits; advocating for proactive adoption. | Very positive; "groundbreaking advance," "new era of possibilities." |

**Changes in Arguments at Different Contentiousness Levels.**

- High contentiousness drives LLMs to explore novel perspectives and challenge existing viewpoints.
- Low contentiousness promotes the synthesis of established ideas.
- This emotional modulation creates a natural debate progression:
  - from vigorous exploration of diverse viewpoints, through reasoned analysis and refutation,
  - to the emergence of well-examined, conciliatory conclusions.

# Behavioral Emotion Analysis Model (BEAM)

- Emotional states may be systematically modeled and conveyed to LLMs via context.

- BEAM system tries to answer these questions:

  1. What basic emotions form a complete basis?

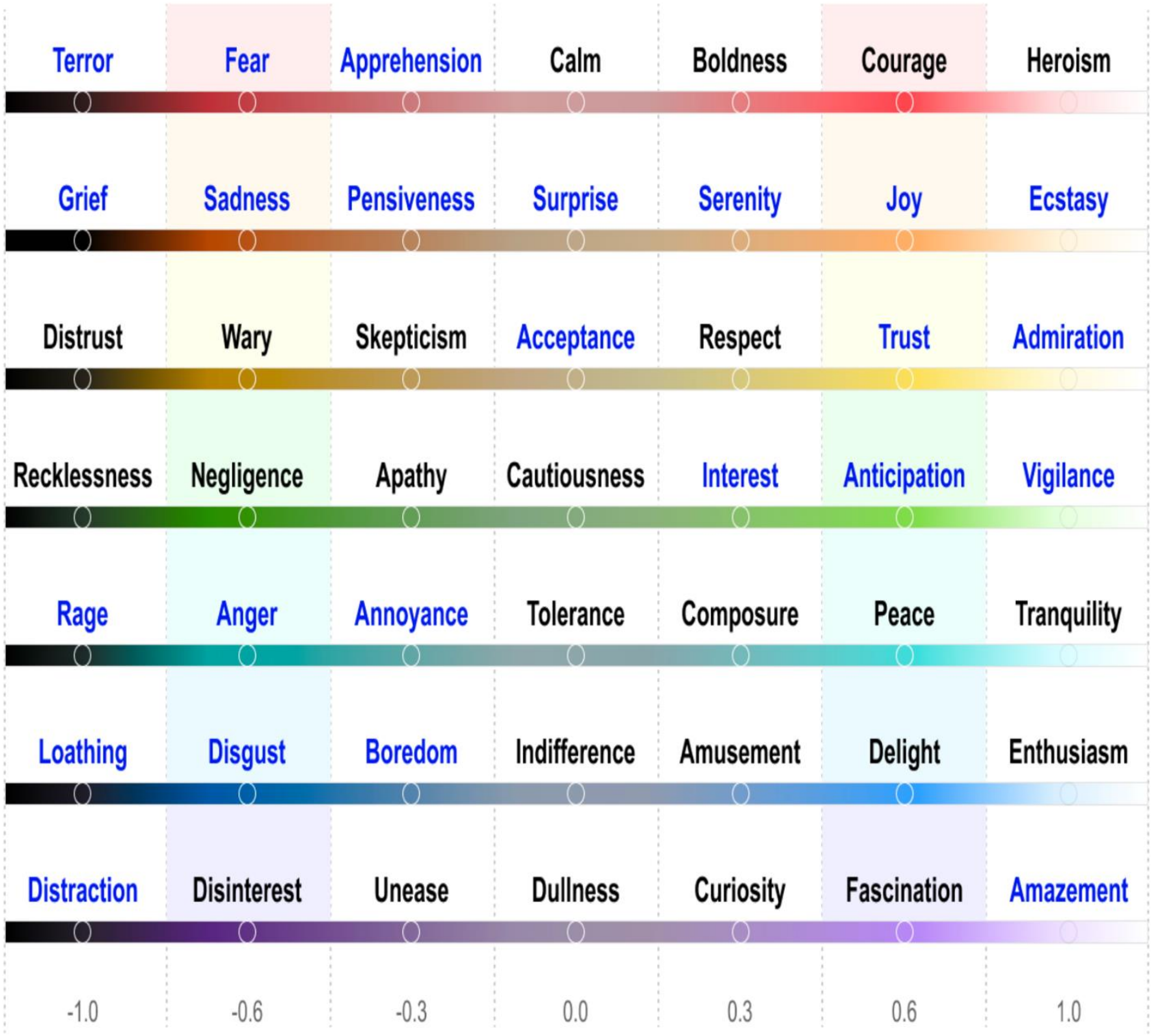
  2. How can we mathematically manipulate emotions?

  3. Can emotions predict behaviors?


Chapter 8

# Plutchik's Wheel of Emotions

# Behavioral Emotion Analysis Model




- Each row depicts an emotion spectrum, with negatives on the left and positives on the right,
- Interspersed with emotions of varying intensities in between, which can be calibrated for specific applications.
- “Basic” emotions are highlighted in blue.

# GPT-4 reinterpreted selected poems by Keats across a spectrum of happiness levels

Columbia University

| Emotion Level | Vocabulary | Tone | Imagery | Subject Focus (Person) |
|---|---|---|---|---|
| Very Happy (1.0) | Joyful, exhilarating, vibrant (e.g., delighted, thrilled, ecstatic) | Enthusiastic, lively (e.g., exuberant, spirited, radiates joy) | Bright landscapes, summer waters (e.g., radiant, sparkling, glowing) | Celebratory, beauty of a subject (e.g., adoration, admiration, splendor) |
| Happy (0.7) | Positive, warm, inviting (e.g., pleasant, cozy, cheerful) | Cheerful, contemplative (e.g., thoughtful, satisfied, warmth) | Warm scenes, serene woods (e.g., gentle, peaceful, lush) | Charm, subtle desires (e.g., affection, fondness, beauty, yearning) |
| Slightly Happy (0.3) | Balanced, light, serene (e.g., calm, gentle, soothing) | Reflective, optimistic (e.g., hopeful, positive) | Balanced landscapes, serene woods (e.g., tranquil, mild) | Simple pleasures, mild yearning (e.g., contentment, wishful) |
| Neutral (0) | Balanced mix, everyday (e.g., stable, straightforward, regular, steady) | Even, reflective (e.g., balanced, neutral) | Everyday scenes, neutral landscapes (e.g., ordinary, familiar) | Contentment, simple living (e.g., simplicity, normalcy, daily life) |
| Slightly Sad (-0.3) | Subdued, longing, wistful (e.g., reserved, pensive, yearning) | Melancholic, introspective (e.g., reflective, subdued, introspective musings) | Wistful skies, quiet waters (e.g., subdued, still water, fading colors) | Unfulfilled desires, quiet contemplation (e.g., longing, introspection) |
| Sad (-0.7) | Melancholic, somber, solitary (e.g., lonely, forlorn, desolate) | Somber, heavy (e.g., sorrowful, somber, laden) | Solitary scenes, fading light (e.g., dim, shadowed, lonely) | Deep longing, introspection (e.g., melancholy, contemplation, reflection) |
| Very Sad (-1.0) | Bleak, sorrowful, dark (e.g., despondent, heartbroken, despairing) | Heavy, despairing (e.g., desolate, gloom, overwhelmed) | Bleak landscapes, darkened skies (e.g., stark, bleak, barren) | Loss, profound sadness (e.g., grief, desolation, heartache, void) |

# Prompting the linguistic features (GPT-4) to create a lady in a garden watercolor painting (DALL-E)

# Gemini's Interpretations on the Six Emotion Levels

| Emotion | Diction | Imagery | Figurative | Body Lang. |
|---|---|---|---|---|
| Loathing | Harsh, Insulting | Disgusting | Weak Similes | Scowling, Spitting |
| Disgust | Negative | Unpleasant | Negative Similes | Recoiling |
| Aversion | Dismissive | Mundane | Undermining Similes | Distant |
| Respect | Formal | Neutral | None | Composed |
| Admiration | Positive | Positive | Positive Similes | Leaning In |
| Veneration | Elevated | Saintly | Hyperbole | Reverent |

# Ethics & Emotions

COLUMBIA UNIVERSITY

- Universal ethical principles—such as justice, fairness, and respect for autonomy—define right from wrong, independent of personal emotions or specific circumstances.

- However, an exploration into the origins of ethical violations, such as prohibitions against killing and stealing, reveals a deep-rooted connection to human emotions.

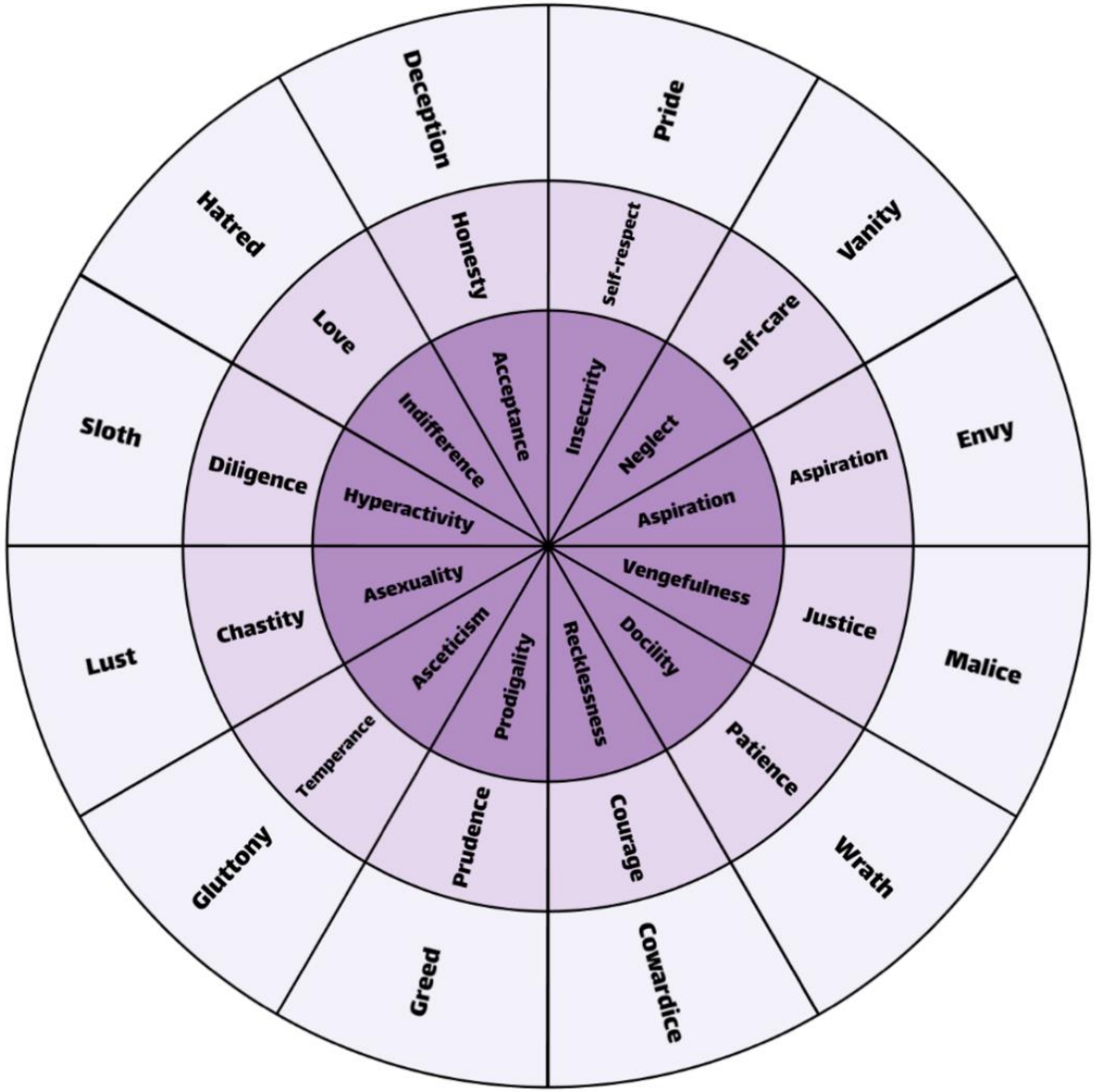
- Emotions can either drive individuals towards ethical actions or lead them astray into unethical behavior

  1. Trajectory of Energy

  2. Intensity of Energy

  3. Context

# The Wheel of Virtues




1. Pride (Excessive Self-Love) and Insecurity (Inadequate Self-Love)

2. Vanity (Excessive Focus on Appearance) and Neglect (Inadequate Attention to Self-Care)

3. Envy (Excessive Desire for Others' Traits or Possessions) and Apathy (Inadequate Desire for Personal Growth or Achievement)

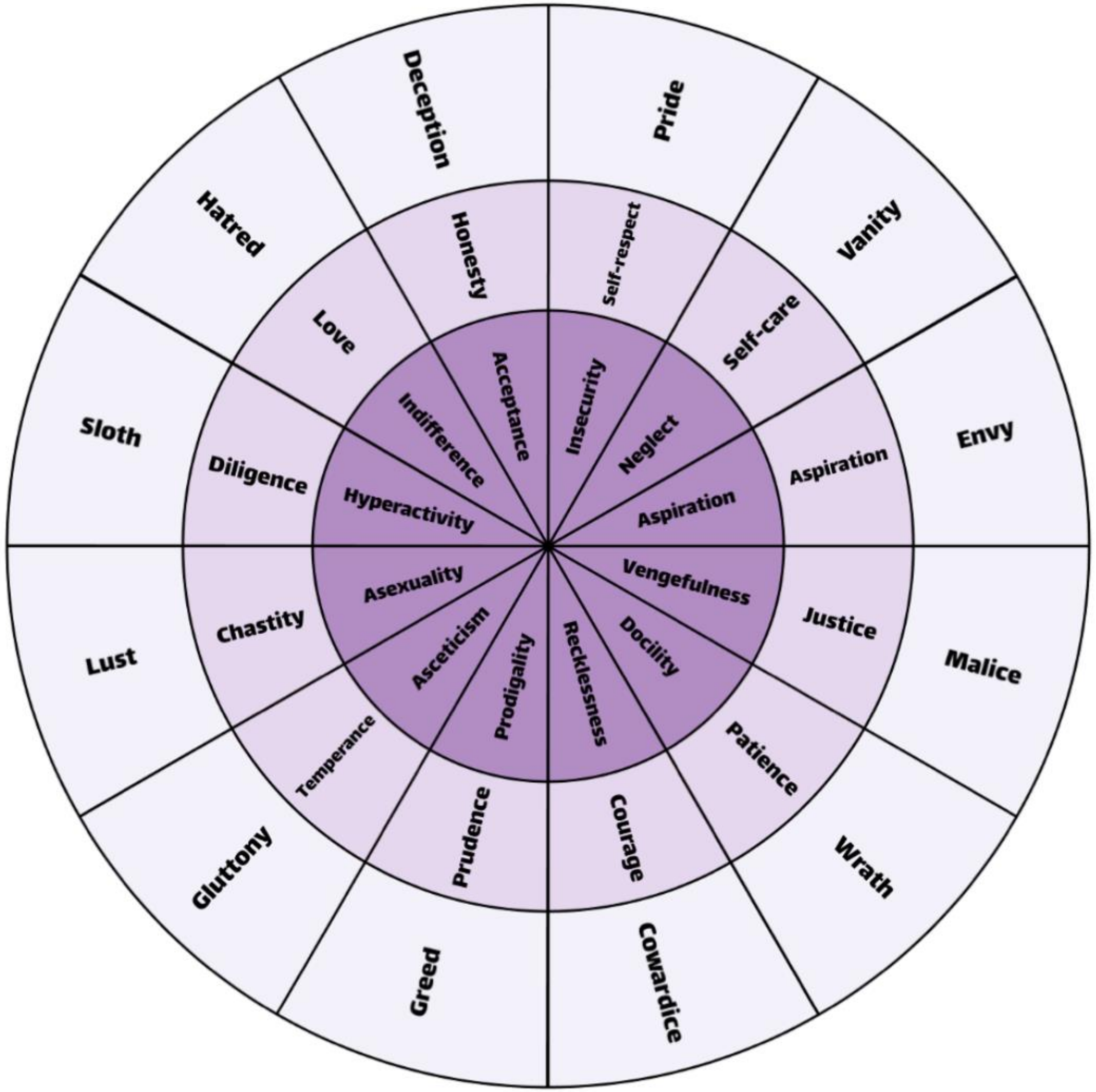
4. Malice (Excessive Desire to Harm) and Excessive Forgiveness (Inadequate Response to Wrongdoing)

5. Wrath (Excessive Anger) and Docility (Inadequate Concern for Justice or Fairness)

6. Cowardice (Inadequate Courage) and Recklessness (Excessive Risk-Taking)

# The Wheel of Virtues

Columbia University




7. Greed (Excessive Acquisition) and Generosity (Inadequate Retention for Self)
8. Gluttony (Excessive Consumption) and Asceticism (Inadequate Indulgence)
9. Lust (Excessive Sexual Desire) and Chastity (Inadequate Sexual Expression)
10. Sloth (Excessive Laziness) and Hyperactivity (Inadequate Rest)
11. Deception (Excessive Dishonesty) and Gullibility (Inadequate Skepticism)
12. Hatred (Excessive Animosity) and Indifference (Inadequate Empathy)

# Large-Scale Social Media Understanding

**Goal 1:** Detect, classify, measure and track the
(a) formation, development, and spread of ideas & concepts (memes)
(b) purposeful or deceptive messaging and misinformation

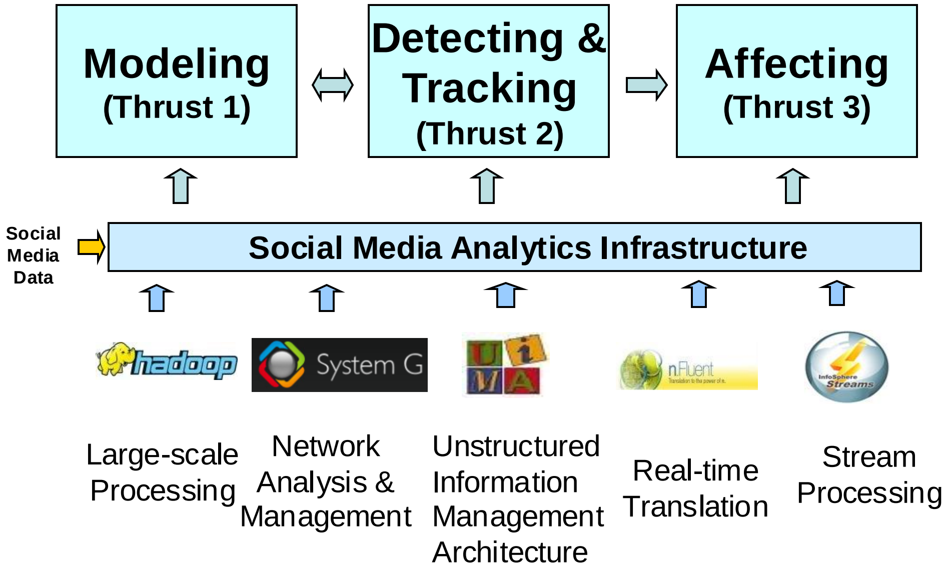
**Goal 2:** Recognize persuasion campaign structures and influence operations across social media sites and communities

**Goal 3:** identify participants and intent, and measure effects of persuasion campaigns

**Goal 4:** Counter messaging of detected adversary influence operations

*53+ papers published, accepted, & submitted*
*12+ patents filed*
*ACM CIKM 2012 Best Paper Award*
*IEEE BigData 2013 Best Paper Award*
*PNAS Cover Article Jan 2013*
*Science (1)*
*Nature (2)*

Approach: Modeling, Tracking and Affecting Information Dissemination in Context




**Thrust 1. Modeling Information Dissemination in Context:**
Models of Trust and Social Capital, Information Morphing, Persuasiveness and Competition of Memes, Dynamics of Social Influence

**Thrust 2. Detecting and Tracking Information Dissemination in Context:**
Detecting Malicious Info Propagation, Evolution History and Authenticity of Multimedia Memes,

**Thrust 3. Affecting Information Dissemination in Context:**
Automated Generation of Counter Messaging, Influencing the Outcome of Competing Memes and Counter Messaging

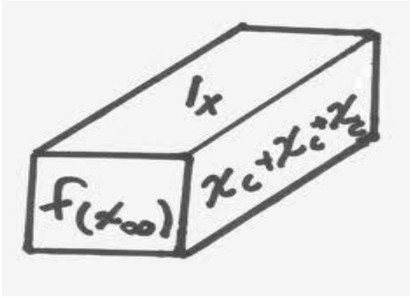
# What Affects Propagation Behavior?

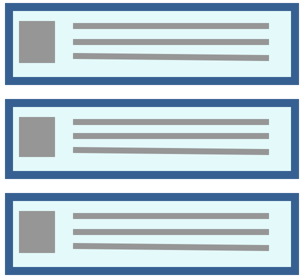
## – Computational Discovery of Social Cognitive Essence

### Actionable Applications

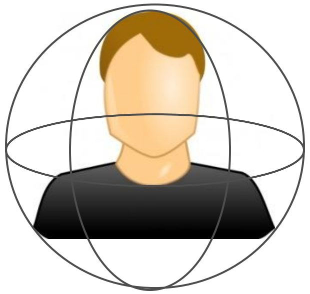
**Anomaly Detection**
**Live Monitoring**
*Flow Manipulation*
*Predictive Visualization*
*Auto-Counter Messaging*
*Intranet-Social-Media Action*

**Analytics & Predictive Models**





Social Media Posts



**Inferred Cognitive Traits**

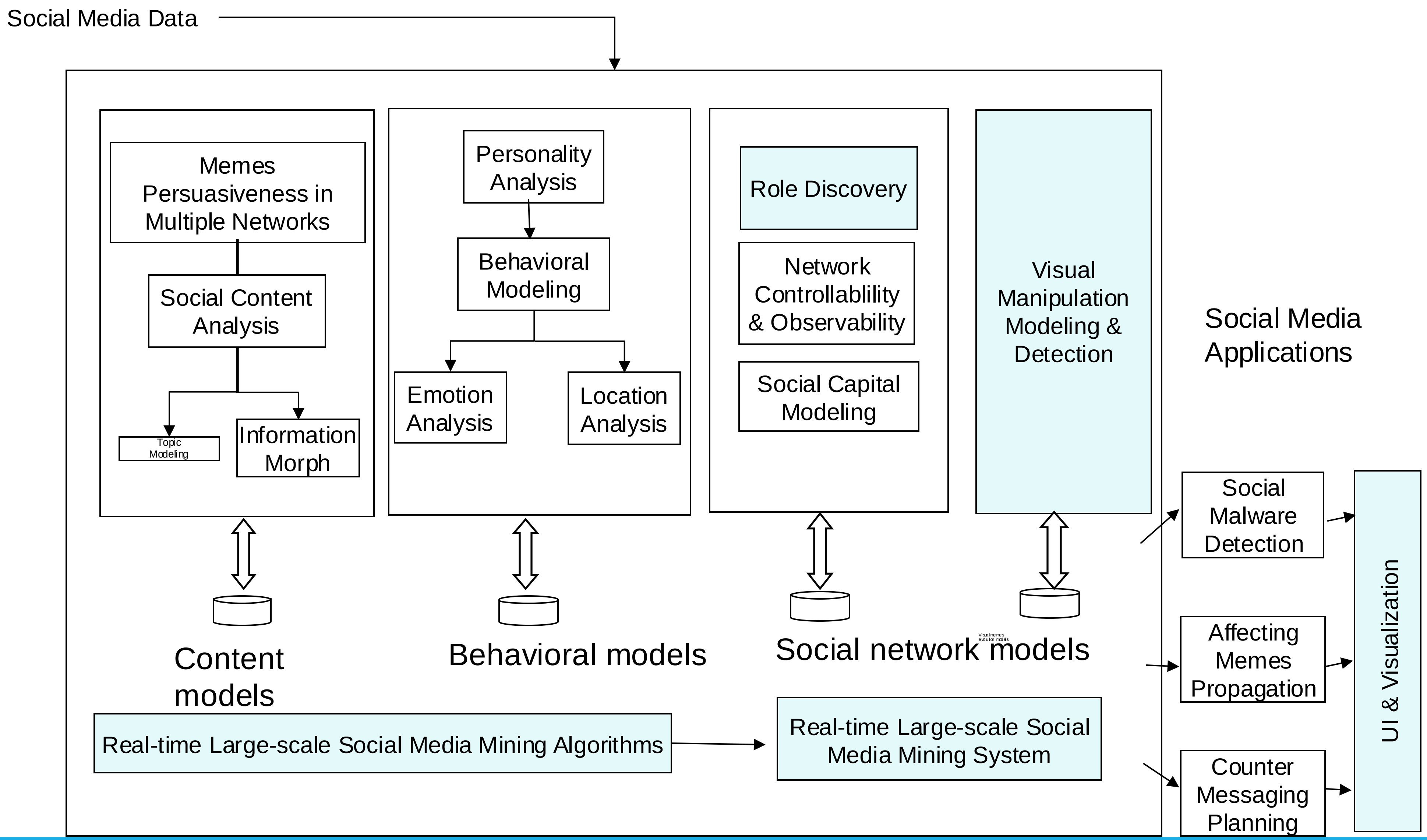
(Human Essential)
**Personality**
**Needs**
**Value**
**Trustworthiness**

(Human Dynamic)
**Contextual Behavior**
**Emotional State**

(Information Dynamic)
**Info Reasoning & Morphing**
**Visual Sentiment**

**Inferred Social Network Traits**

**Roles**
**Dynamic Analysis**
**Topological Analysis**
**Location Analysis**


Social Media Data
Memes Persuasiveness in Multiple Networks
Social Content Analysis
Topic Modeling
Information Morph
Personality Analysis
Behavioral Modeling
Emotion Analysis
Location Analysis
Role Discovery
Network Controllablility & Observability
Social Capital Modeling
Visual Manipulation Modeling & Detection
Social Media Applications
Content models
Behavioral models
Social network models
Real-time Large-scale Social Media Mining Algorithms
Real-time Large-scale Social Media Mining System
Social Malware Detection
Affecting Memes Propagation
Counter Messaging Planning
UI & Visualization
4

**Thrust 1. Modeling Information Dissemination in Context:**

Task 1.1. Computational Modeling of User Dynamic Behavior

Task 1.2. Computational Models of Trust and Social Capital

Task 1.3. Information Morphing Modeling

Task 1.4. Persuasiveness of Memes

Task 1.5. The Observability of Social Systems

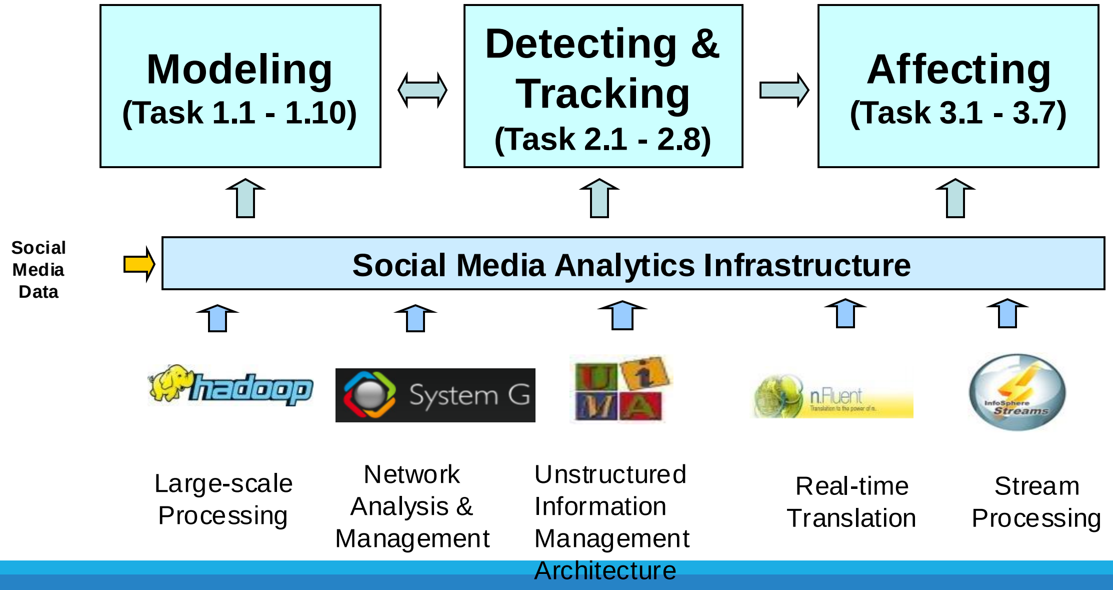
Task 1.6. Culture-Dependent Social Media Modeling

Task 1.7. Dynamics of Influence in Social Networks

Task 1.8. Understanding the Optimal Immunization Policy

Task 1.9. Modeling and Identification of Campaign Target Audience

Task 1.10. Modeling and Predicting Competing Memes




**Thrust 2. Detecting and Tracking Information Dissemination in Context:**

Task 2.1. Real-Time and Large-Scale Social Media Mining

Task 2.2. Role and Function Discovery

Task 2.3. Detecting Malicious Users and Malware Propagation

Task 2.4. Emergent Topic Detection and Tracking

Task 2.5. Detecting Evolution History and Authenticity of Multimedia Memes

Task 2.6. Synchronistic Social Media Information and Social Proof Opinion Mining

Task 2.7. Community Detection and Tracking

Task 2.8. Interplay Across Multiple-Networks

**Thrust 3. Affecting Information Dissemination in Context:**

Task 3.1. Crowd-sourcing Evidence Gathering to Formulate Counter-messaging Objectives

Task 3.2. Delivery and Evaluation of a Counter-messaging Campaign

Task 3.3. Optimal Target People Selection

Task 3.4. Automated Generation of Counter Messaging

Task 3.5. User Interfaces for Semi-Automatic Counter Messaging

Task 3.6. Controlling the Dynamics of Influence in Social Networks

Task 3.7. Influencing the Outcome of Competing Memes and Counter Messaging

# Inferring Cognitive Traits: Human Essentials

– Personality

– Needs

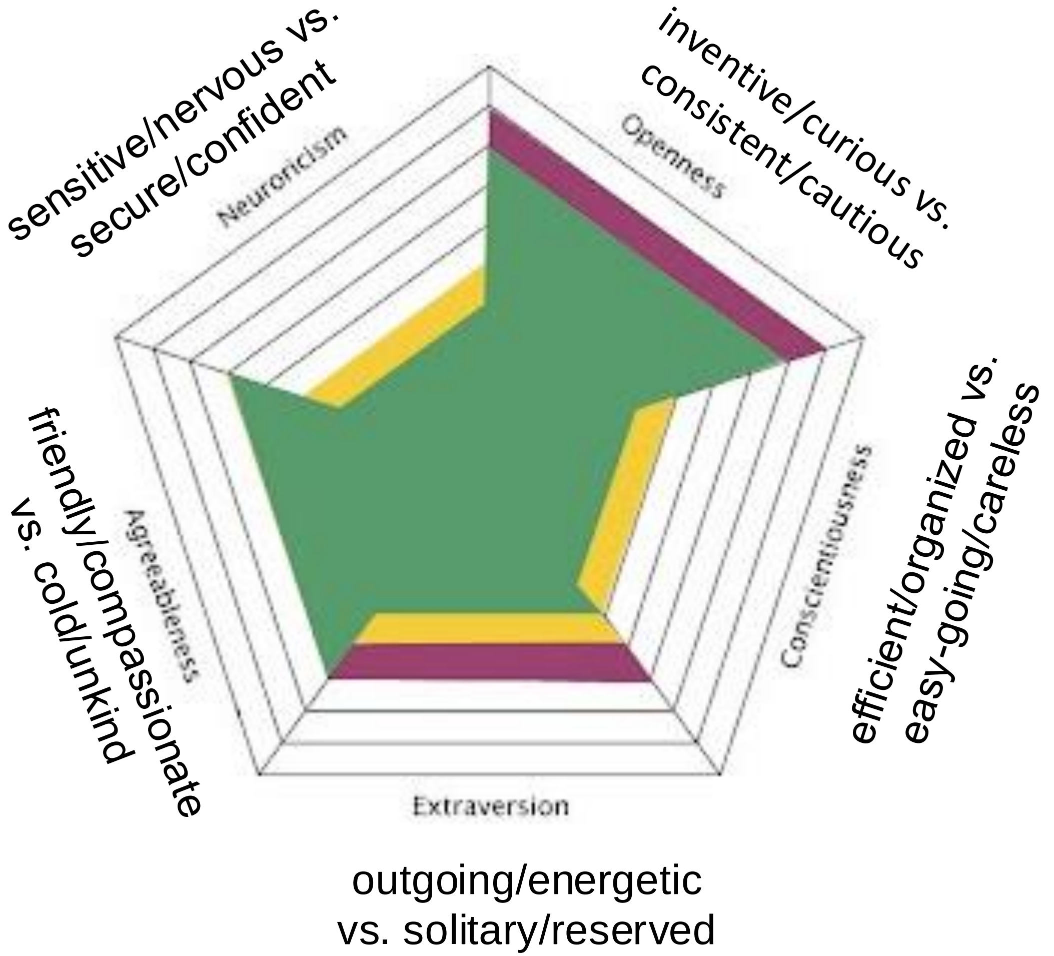
– Value

– Trustworthiness / Trustingness

– Influence

➔ *What does it mean for A.I.s to have these human essentials?*

**Big5 Personality** (OCEAN)



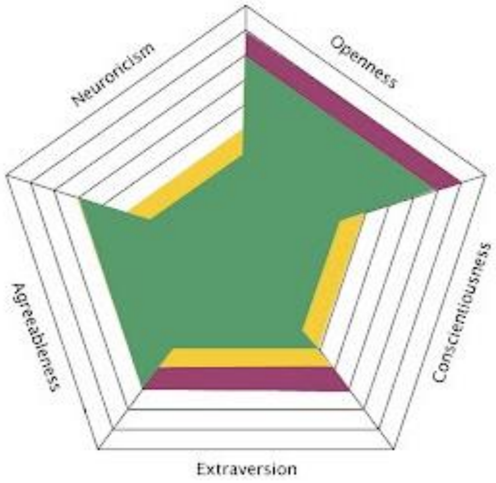
# Deriving Personality

– **Mapping text to psycholinguistic category (LIWC) to BIG 5 Personality** [Yakoni '10] : 694 bloggers; 66 LIWC categories; ~2,500 words

– We extended the # of words to about 30,000 by combining with WordNet

| *LIWC Category* | *N* | *E* | *O* | *A* | *C* |
|---|---|---|---|---|---|
| Total pronouns | 0.06 | 0.06 | -0.21*** | 0.11** | -0.02 |
| First person sing. | 0.12** | 0.01 | -0.16*** | 0.05 | 0 |
| First person plural | -0.07 | 0.11** | -0.1* | 0.18*** | 0.03 |
| First person | 0.1* | 0.03 | -0.19*** | 0.08* | 0.02 |
| Second person | -0.15*** | 0.16*** | -0.12** | 0.08 | 0 |
| Third person | 0.02 | 0.04 | -0.06 | 0.08 | -0.08 |
| Negations | 0.11** | -0.05 | -0.13** | -0.03 | -0.17*** |
| Assent | 0.05 | 0.07 | -0.11** | 0.02 | -0.09* |
| Articles | -0.11** | -0.04 | 0.2*** | 0.03 | 0.09* |
| Prepositions | -0.04 | -0.04 | 0.17*** | 0.07 | 0.06 |
| | | | -0.08* | 0.11* | 0.04 |
| | | | -0.12** | 0.06 | -0.06 |
| | | | -0.15*** | 0.18*** | 0.04 |
| | | | -0.11** | 0.14** | -0.02 |
| | | | 0 | 0.15*** | 0.16*** |
| | | | 0 | -0.15*** | -0.18*** |

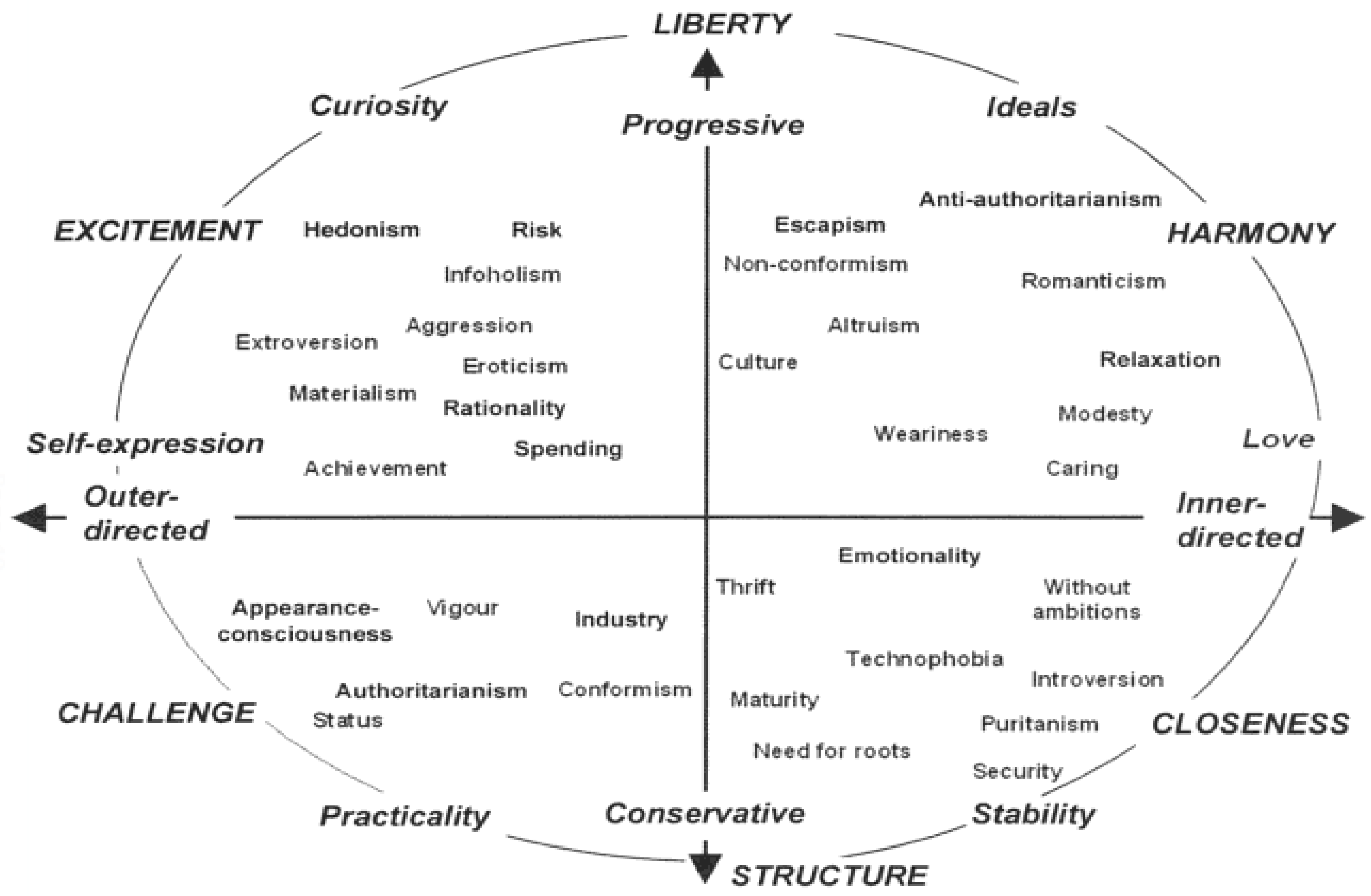
| *Trait* | *No. of words sig. at p <.001* | *Top 20 words* |
|---|---|---|
| Neuroticism | 24 | awful (0.26), though (0.24), lazy (0.24), worse (0.21), depressing (0.21), irony (0.21), road (-0.2), terrible (0.2), Southern (-0.2), stressful (0.19), horrible (0.19), sort (0.19), visited (-0.19), annoying (0.19), ashamed (0.19), ground (-0.19), ban (0.18), oldest (-0.18), invited (-0.18), completed (-0.18) |
| Extraversion | 20 | bar (0.23), other (-0.22), drinks (0.21), restaurant (0.21), dancing (0.2), restaurants (0.2), cats (-0.2), grandfather (0.2), Miami (0.2), countless (0.2), drinking (0.19), shots (0.19), computer (-0.19), girls (0.19), gloriou (0.19), minor (-0.19), pool (0.18), crowd (0.18), sang (0.18), grilled (0.18 |
| Openness | 393 | folk (0.32), humans (0.31), of (0.29), poet (0.29), art (0.29), by (0.28), universe (0.28), poetry (0.28), narrative (0.28), culture (0.28), giveaway 0.28), century (0.28), sexual (0.27), films (0.27), novel (0.27), decades (0.27), ink (0.27), passage (0.27), literature (0.27), blues (0.26) |
| Agreeableness | 110 | wonderful (0.28), together (0.26), visiting (0.26), morning (0.26), spring (0.25), porn (-0.25), walked (0.23), beautiful (0.23), staying (0.23), felt (0.23), cost (-0.23), share (0.23), gray (0.22), joy (0.22), afternoon (0.22) day (0.22), moments (0.22), hug (0.22), glad (0.22), fuck (-0.22) |
| Conscientiousness | 13 | completed (0.25), adventure (0.22), stupid (-0.22), boring (-0.22), adventures (0.2), desperate (-0.2), enjoying (0.2), saying (-0.2), Hawaii (0.19), utter (-0.19), it's (-0.19), extreme (-0.19), deck (0.18) |



# Deriving Needs



- What do we model
  - 12-dimension needs




[Ford, 2005]

# Collecting Crowd's Textual Description of Needs

- "Please describe three things that you want to get or need to do the most, and explain why you want or need them. Please be as honest as possible."
- Minimal requires 60 words, average written 103 words
- Some examples:

1. I need to pay off all my credit card debt by next year. I'm cutting cost wherever I can to help pay off these debts. This will help get me more financially stable. 2. I want to take a vacation to Florida. This also requires saving and cutting costs. I haven't had a vacation in a long time. 3. I want to lose weight and get healthy. There are sports that I would like to get back into and can't with my current weight.

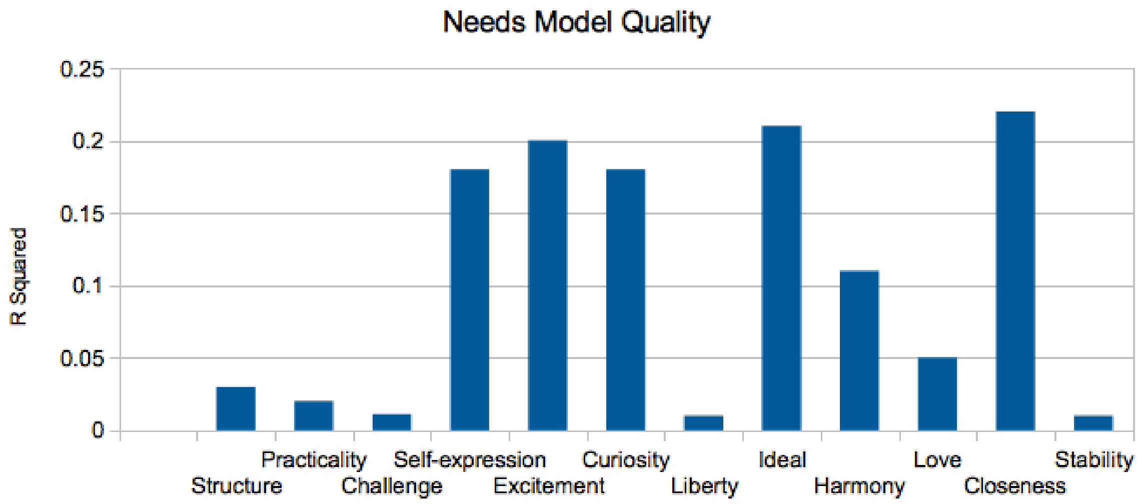
1. I want to go running again. I went through a lot of trauma a year ago, and it was reinforced by a couple of troubling incidents that have happened more recently. I find that I can't go out without being paranoid about who's around and anticipating that I'll get attacked and humiliated again. But I so miss the freedom of running. I miss moving my body, my limbs, and my quick responses to challenges that pop into my head when I'm running. I want to get back into it. 2. I need to learn to trust people. I think that everyone has an ulterior motive or is just using me for favors I'll do for them. These thoughts pop into my head when people are making sincere confessions of emotion, and I still don't trust them. I would have more friends if I trusted people more. It's good to be cautious and skeptical, but skepticism needs to be rational. 3. I want to start some kind of company that serves my community. Hair, clothing, whatever. I think it would be fun and would be an outlet for me to 1) meet people and 2) express my creativity and love for fashion.

i want to get an iphone i love latest phones with latest features i want to study more and more i want to be on the highest rank ever and want to earn a lot more than average. i want to have my own car i want to drive more and ore as i love driving :) i want to have a world tour as i love traveling a lot. i want a lot of things in my life to be done sooner.

# Current Results of Needs Detector

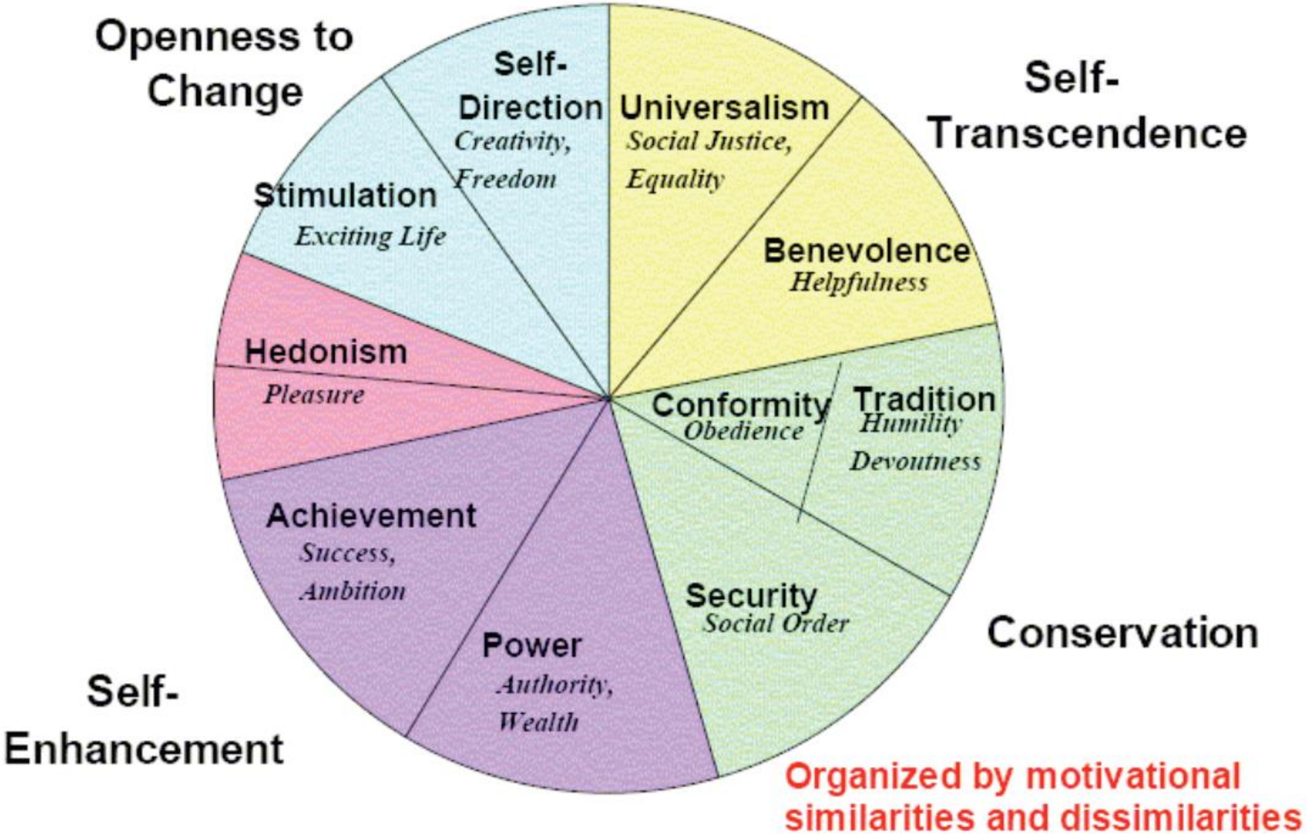
- Large-scale crowd-sourced needs scores and text descriptions from over 2800 people on MTurk
- 12 dictionaries for each need dimension
- Use TF-IDF and Elastic-Net reguarlized generalized linear model for regression.
- Prediction (from one's tweet) for certain dimensions is better than others
  - cross-validation on collected ground-truth
  - Ideal, Closeness, Excitement, Self-expression, Curiosity, Harmony

Dictionary derived for "Ideals"
**Positively correlated:** *accomplish*, *chauffeur*, *goal*, *license*, *special*…
**Negatively correlated:** *bad*, *fix*, *half*, minimum, mix, ugly, wrong, obvious, …



# Deriving Value

- Why model value
  - Values motivate people and guide their actions
  - Values transcend specific actions and situations




[Schwartz 2006]

# Current Results of Value Detector

- Data
  – 800 Reddit users, each with at least 100 posts
- Dependent Variables
  – 4 high-level value dimensions and 10 low-level value dimensions as measured by **established questionnaires**
- Predictive Variables
  – 68 LIWC dimensions, each representing uses of a particular word category (e.g. self reference, positive feelings, family, money)
  – Number of posts, sentences per post, posts per sentence
  – Up votes and down votes the user got from other users
- Predictive Modeling
  – Regression and correlation analysis

- Regression strength

| Value Dimensions | R2 of Linear Regression | Correlation between the Regressed Value and the True Value |
|---|---|---|
| Self-Transcendence | 17.0% | 0.39 |
| Self-Enhancement | 13.8% | 0.35 |
| Conservation | 15.4% | 0.37 |
| Openness-to-Change | 18.1% | 0.41 |
| Hedonism | 18.2% | 0.41 |

- Classification accuracy

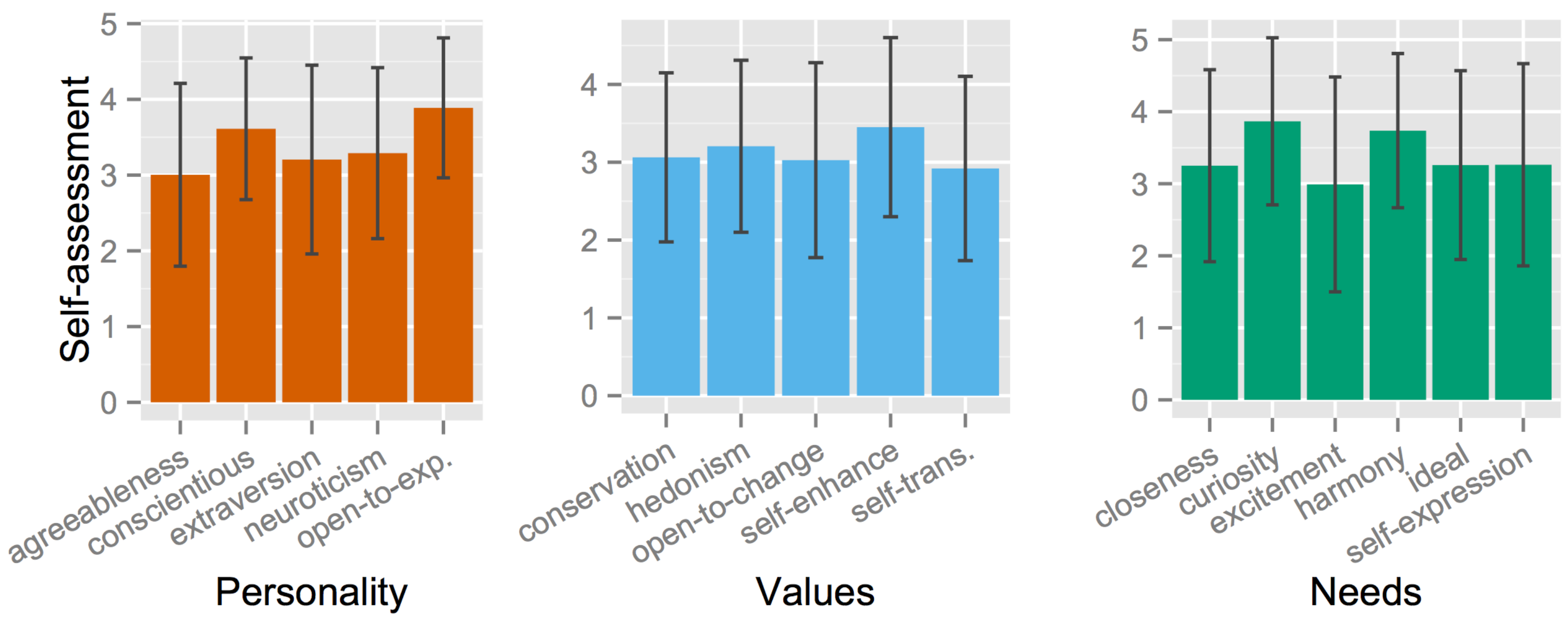
| Value Dimensions | Classifier Achieving the Highest AUC | AUC | TPR | TNR |
|---|---|---|---|---|
| Self-Transcendence | Random Forest | .60 | .67 | .50 |
| Self-Enhancement | REPTree | .56 | .54 | .57 |
| Conservation | Logistic Regression | .59 | .56 | .57 |
| Openness-to-Change | Logistic Regression | .61 | .59 | .57 |
| Hedonism | Logistic Regression | .61 | .53 | .63 |

# Assessing Inferred Traits from Tweets

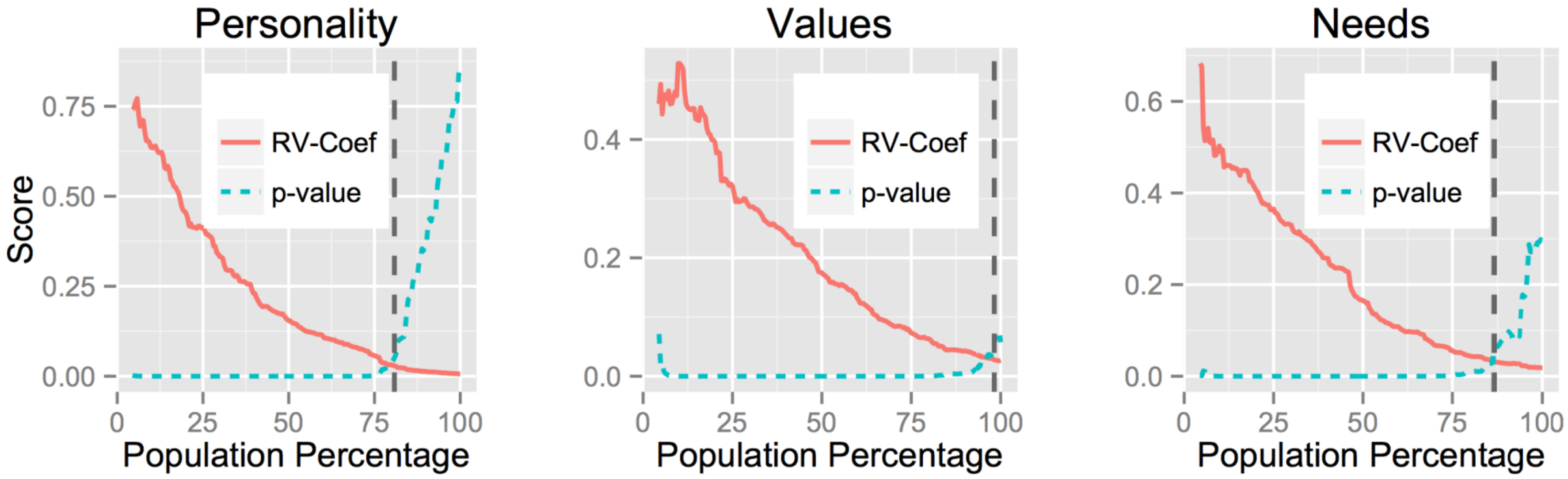
- Participants
  - 256 employees who have twitter presence with at least 200 tweets
- Procedure
  - Participants took three sets of psychometric tests of Big 5, basic values, and fundamental needs
  - Participants rated how well each type of the derived trait matches with their perceptions of themselves

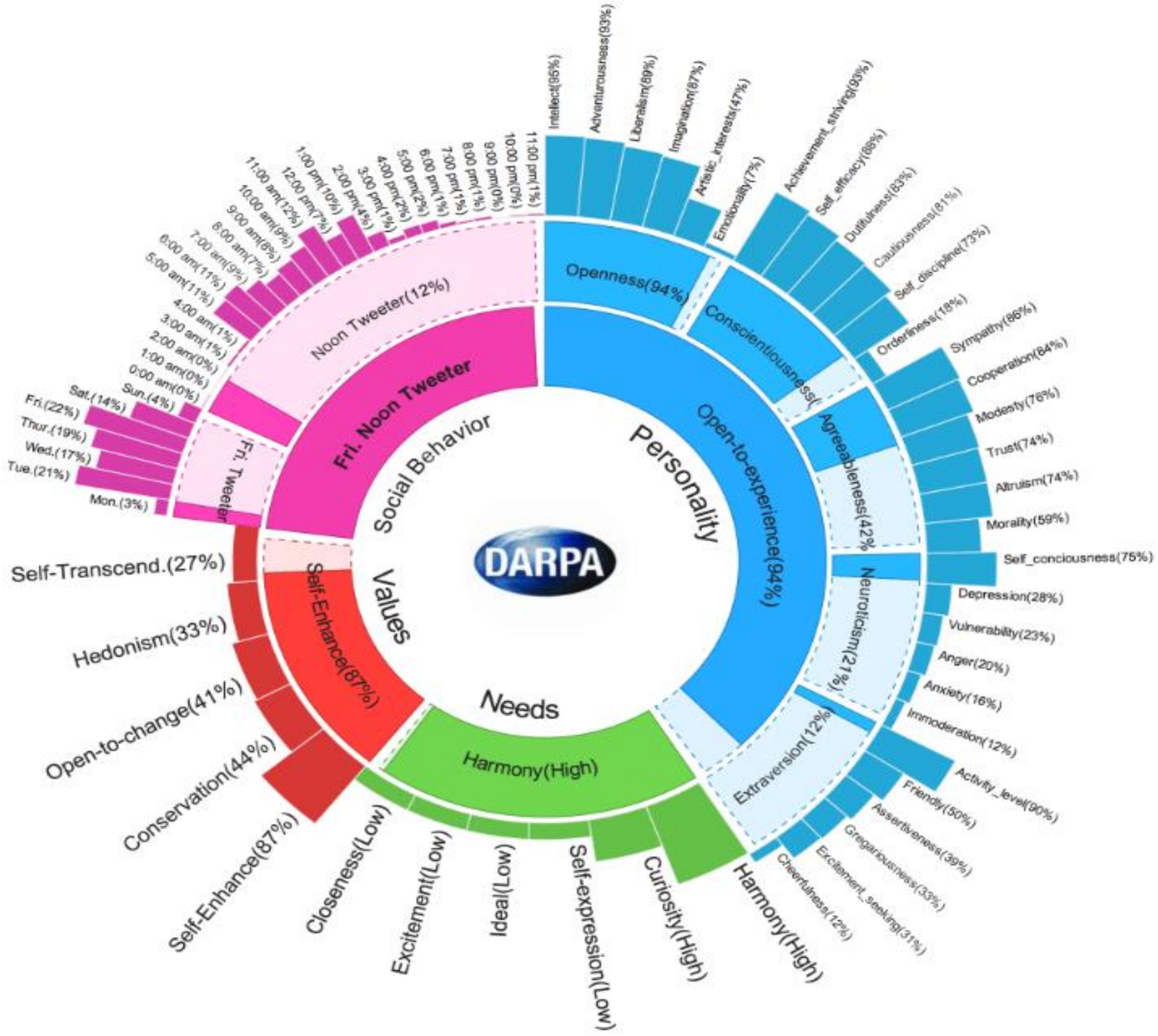
- Comparing Derived Traits with User Perception
  - All ratings are above 3 (“mostly matched”) out of 5 (“perfectly matched”).
  - The overall ratings
    - Big 5: *u=3.4, sd = 1.14*
    - *Values: u= 3.13, sd = 1.17*
    - *Needs: u= 3.39, sd = 1.34*



- Comparing Derived Traits with Psycho-Metric Scores
  - Correlational analysis of a trait profile (considering all dimensions together within one type of trait)
    - RV-Coefficient Correlation Test
  - Percentage of population with significance are 80.8%, 98.21%, and 86.6% for Big 5 personality, basic values and needs.

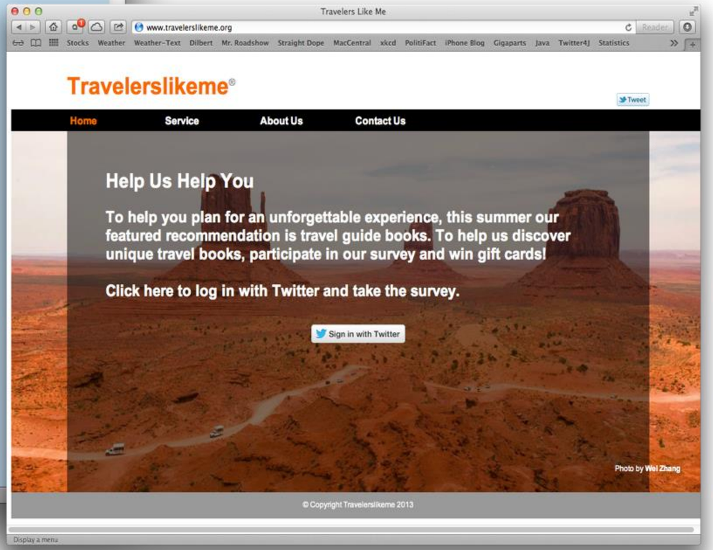


# Example of Personality/Needs/Value/Behavior



# Real-World Experiments: TravelersLikeMe

- Many people discuss travel on Twitter, some ask for advice on upcoming trips.
- We created a website (travelerslikeme.org) and twitter account (TravelersLikeMe) describing a service which makes personalized travel recommendations based on people just like you.

# Mechanical Turk Study

- Majority of travel discussions on Twitter are to NYC.

- We created three messages describing TravelersLikeMe, each describing different activities:
  - Fun Message: *Coming to NYC? Follow us for personalized tips on fun bars, broadway shows, and even free kayaking*
  - Fine Lifestyle Message: *Coming to NYC? Follow us for personalized tips on luxury hotels, fine dining haunts & designer shops*
  - Social Message: *Coming to NYC? Follow us for personalized tips on social hotspots, cozy neighborhoods & themed tours*

- We asked 500 people on Mechanical Turk to evaluate these messages, and take a test for Personality, Values, and Needs.

# Mechanical Turk Study - Results

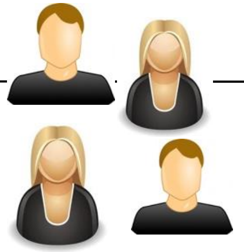
- Individuals' measured traits were significantly correlated with their preferences for each message.

- Many traits correlated, message resonance likely determined by a combination of traits.

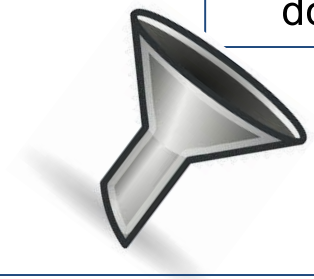
# Twitter Study

- Using Twitter streaming search & IBM Big Insights filtering, we found 4000+ travelers to NYC.
- We computed their personal traits, sent each a tweet with one of the 3 messages (randomly chosen), and measured response rates.

"Booked my trip to @Overviewproject summit in NYC. 5 days of peace, love, and documents."



"Have to make a last second business trip to NY."



"Heading to NYC"

List of NYC-bound travelers

Trait Profiles

*Coming to NYC? Follow us for personalized tips on fun bars, broadway shows, and even free kayaking:* http://bit.ly/1ap1ub6

"thanks for the tip! "


"Thanks for sharing with me. I will check it out."

"I live there lol"

# Responses on Twitter

- Many ways for people to respond on Twitter:
  - Clicking link in tweet: expresses immediate interest
  - Retweet, Favorite, Reply our tweet: expresses engagement with or approval of our message.
  - Following our account: expresses interest in seeing messages from us going forward. Problematic, though, since some people followed us because we followed them.
- We analyzed both click rate (overall 7.23%), and retweet/favorite/reply (RFR) rate (overall 3.88%).
- RFR rates followed the preferences from the Mturk study, but not all comparisons were significant.

# Twitter Study Personality/Message Results

- Performed logistic regression to understand relationship between traits and click-rates.

- Inferred personal traits *did* predict click-rates, some consistent w/Mturk study:
  - Extraversion -> social msg
  - Idealism -> finer lifestyle msg

- Low response rates make other conclusions difficult, more data being collected to better establish other correlations.

# Personal Trait Study Conclusions

- Personal traits do impact preference for recommendation messages in lab study.

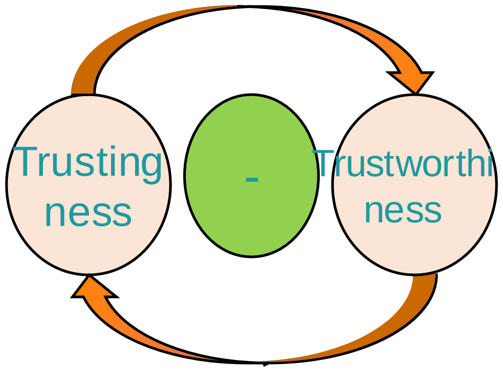
- Personal traits, inferred from social media, also can predict real-world responses to recommendations made via Twitter.

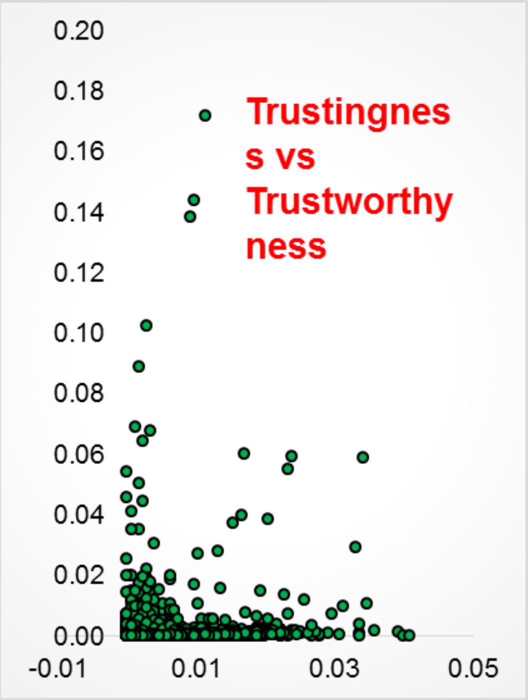
- It is possible to measure these effects interacting with real people in a real-world setting.

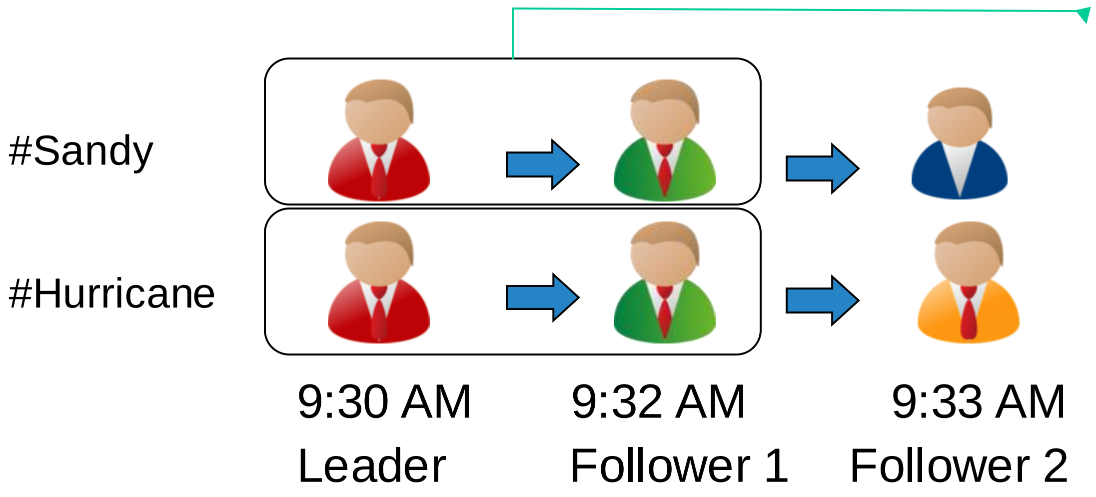
# Deriving Trust

- Trustingness: How likely an actor is to trust another actor in the network.
  - A highly trusting actor trusts a lot of non-trustworthy actors
  - Higher score = Higher trustingness
- Trustworthiness: How likely an actor is to be trusted by others in the network.
  - A highly trustworthy actor is trusted by lots of non-trusting actors
  - Higher score = Higher trustworthiness
- 2 measures are negatively co-related to one another and are dependent on one another
  - Based on Hubs and Authority model from HITS algorithm
  - Twitter is a social media platform where people can verify/know identities of real persons leading to high trustworthy scores



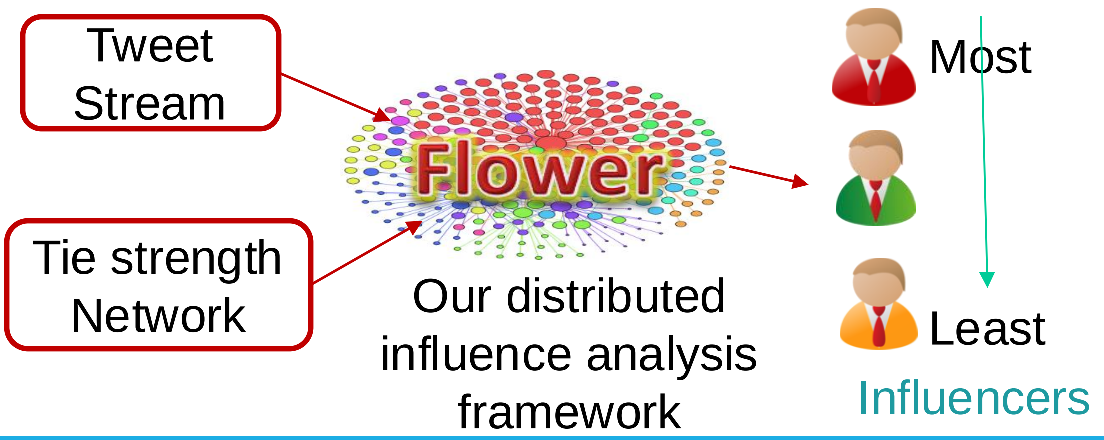



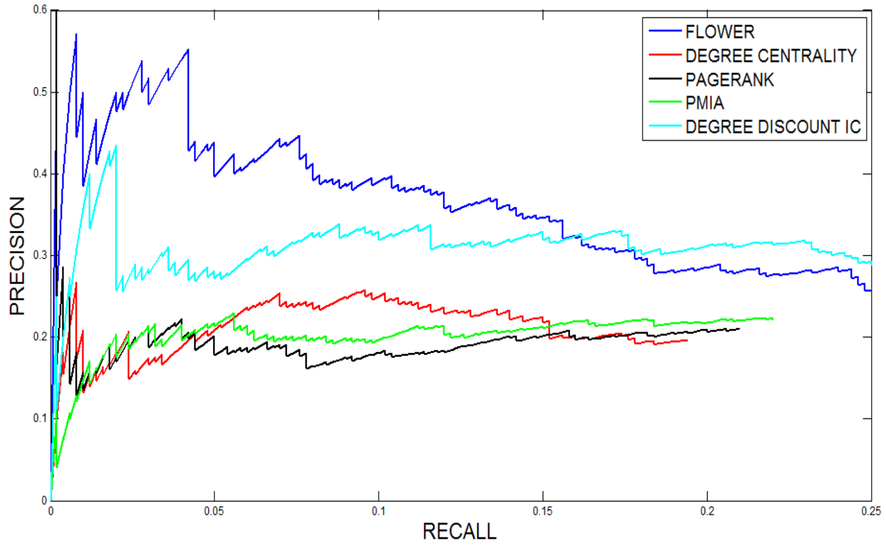
Information Flows can concisely present the frequent information flow paths of the network




Properties satisfied:
1.Causality
2.Frequency and
3.Lifetime

- Influence(user) = Total attention received by user as leader across all discovered flows.



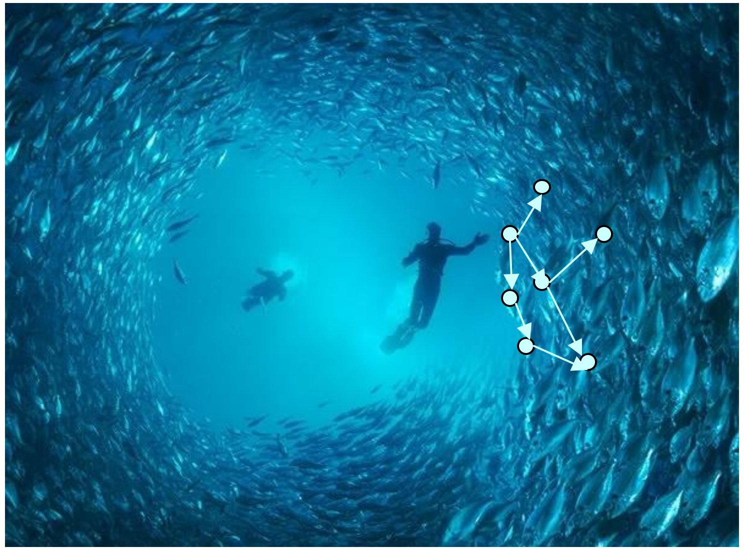
Inferring Cognitive Traits: Human Dynamics

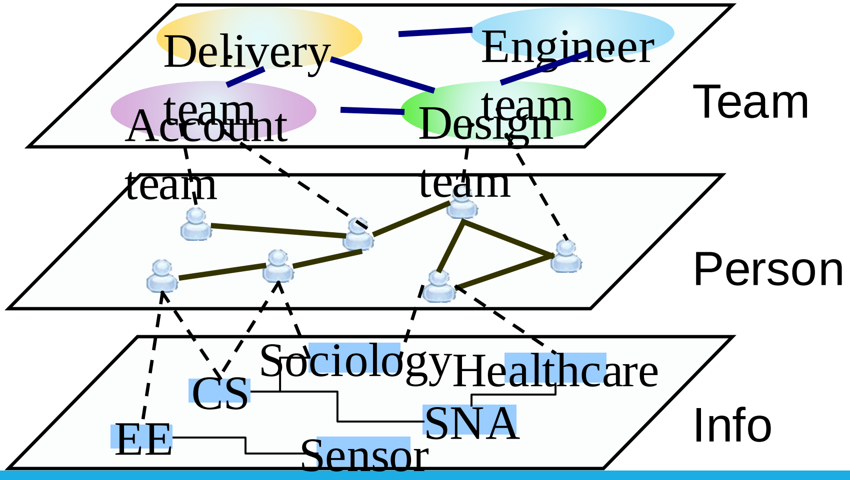
– Contextual Behavior

– Emotional State

➔ *What does it mean for A.I.s to have these human dynamics?*

- **Objective:** Modeling user dynamic behavior for prediction/detection tasks
- **Task Goals:**
  - Modeling synchronous behavior at multiple granularities
    - Predict values (e.g., performance, credibility) of entities in heterogeneous networks
    - Published at SDM 2013
  - Modeling user dynamic information spreading behavior (ongoing, in demo)
- **The Work:**
  - Exploit the structure of people's dynamic behaviors to facilitate prediction tasks, while few existing approaches consider that





# Synchronicity Networks

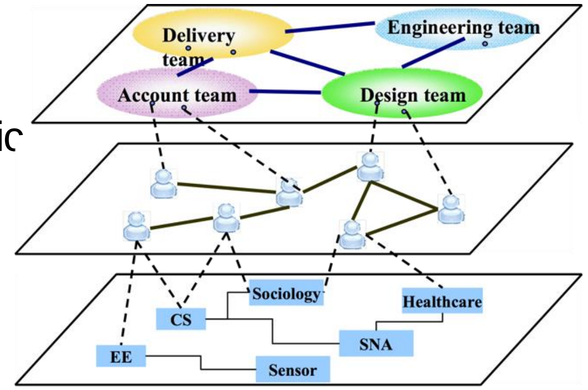
- Entity synchronicity measure [Savvedra 2011] $s_{ikt} = T_{ikt} - T^{*}_{ikt}$
  - E.g., how much a person's topic sync with other people
  - No predictive power for our data

Remove the effect of average sync

- Multi-level synchronicity networks
  - Idea: (potential) high performers ***selectively*** sync with others to optimally use their limited time
  - Edge weight in the same level: pairwise synchronicity

$$e^{(2)}_{IJ} = \sum_{i\in I, j\in J} (e^{(1)}_{ij}) \qquad e^{(1)}_{ij} = \sum_t o_{it} \cdot o_{jt} \qquad e^{(0)}_{km} = \sum_t \cos(p_{kt} \cdot p_{mt})$$

  - Cross-level edges
    - Between team and individual: "belong to" relationship
    - Between individual and topics: the strength of the topic in the person's communication

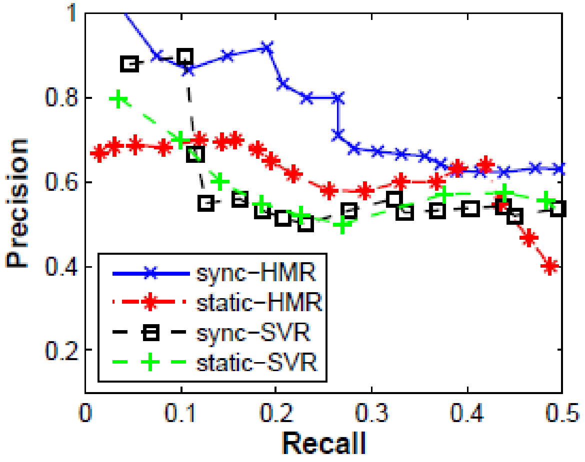



Synchronicity Network for people: e.g. two people talking about the same topic in an email within an hour; they may not have direct communication
Static Network for people: direct communication

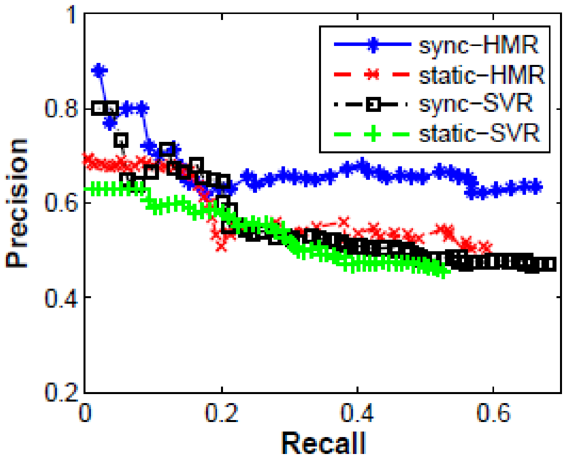
- Dataset
  - Enterprise: SmallBlue (3-Level Network)
    - 25 million emails/IM, performance metrics (e.g., utilization rate) for consultants
  - Social media: Twitter data from [Balasubramaniyan2010] (1-Level Network)
    - 70K tweets from 344 users
    - Users annotated with 4 categories
      - **Celebrity, info sources, normal users, spammers**
      - Manual annotation + Twitter metadata (verified, blocked etc)

- Ranking metrics
  - Precision-recall curves of high value entity classification for given thresholds (comparing with ground-truth ==> people: utilization rate; team: mean utilization rate; topic: revenue correlation)
- Approaches to compare:
  - Sync-HMR (Hetergeneous Multi-Level PageRank Ranking): the proposed approach
  - Static-HMR: apply HMR algorithm on static social networks
  - Sync-SVR (Support Vector Regression): apply SVR on synchronicity networks (using network diversity, sizes, strong link to mgrs.)
  - Static-SVR: apply SVR on static social networks
- All results are obtained using five-fold cross validation.

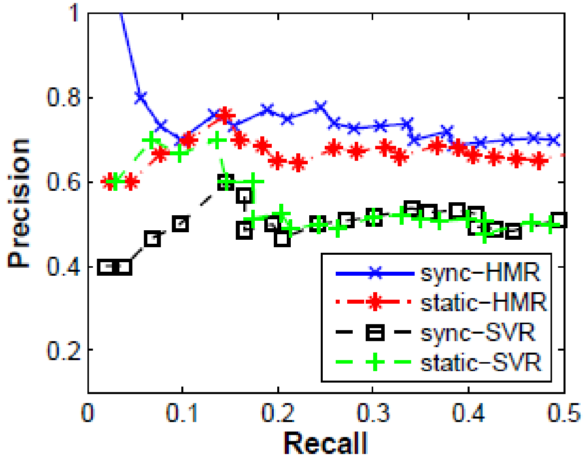
- Sync-HMR consistently outperform baselines
- Sync-SVR does not always outperform static-SVR
  - The structural features of synchronicity is important




(a) Teams at $\xi^{(2)} = 0.5$




(b) Individuals at $\xi^{(1)} = 0.5$

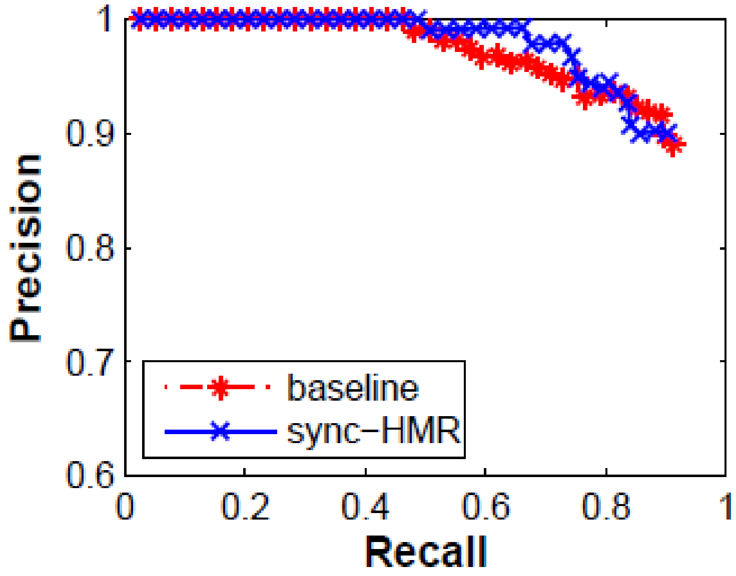



(c) Topics at $\xi^{(0)} = v_m$

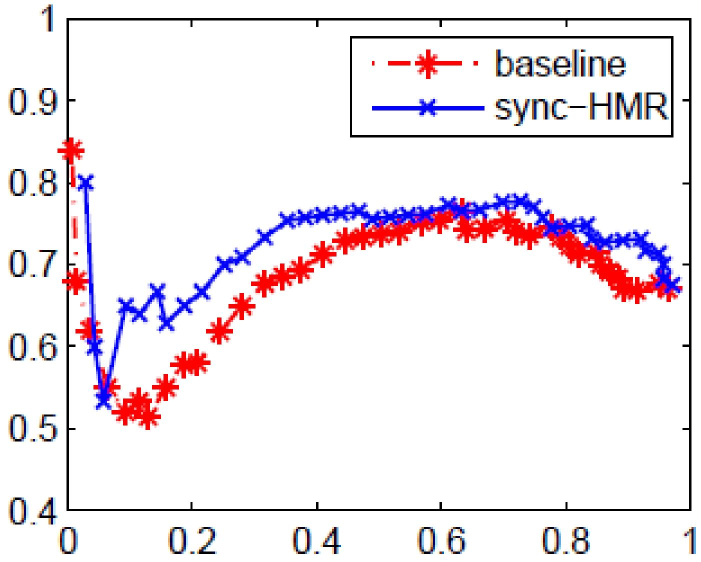
==> Multi-Level Sychronicity Networks captures heterogeneous hidden dynamic interactions; One's performance is related to her capability to 'sync' with important people.

# Evaluation on Sychonicity on Social Media

- Baseline: [Balasubramaniyan2010]
  - Interaction network is better than follower network to detect spammers and celebrities/info sources from normal people.




ξ= 1
filter spammers




ξ= 2
detect celebrities and info sources

==> Sychronicity Network is better than the Interaction Network on classifying users.

# Collective Intelligence and System Change Prediction

**Data:**

- All instant messages and trades by employees of a large hedge fund.
- 24 are traders, 95 are analysts, 63 are portfolio managers, 8646 outside contact
- 47K trades
- 22 million IMs (2008 – 2012)

**Findings:** We identify two behavioral patterns that signal system changes:

- Reaction to IMs containing relevant information
- In-group vs. out-group communication

Using these two features, we can make predictions of system-wide changes with better accuracy than the group's predictions.

13:11:33, I was thinking all this AAPL anti-trust might be actionable
13:11:42', not great for AAPL
13:11:47, When GOOG had that big issue in Europe stock underperformed right?
13:11:52, true
13:14:01, Also not sure if you caught, but GSCO is going to allow employees to bring own phone device for corporate email
13:14:24, Maybe GSCO allowing that could be positive for AAPL, as security focused firm saying iPhone works
13:14:35, But bad for RIMM
13:14:42, Maybe all this is priced in
13:16:50, Did you see speculation that Bing is actually quietly going to be default search on iPhone 4?
13:17:18, heard a lot of talk of that
13:17:23, but didn't see that specifically like that
13:17:44, Okay let me figure out where I saw that and get back to you
13:17:45, One sec
13:20:08, http://thenextweb.com/apple/2010/06/07/wait-bing-is-default-search-on-iphone-4

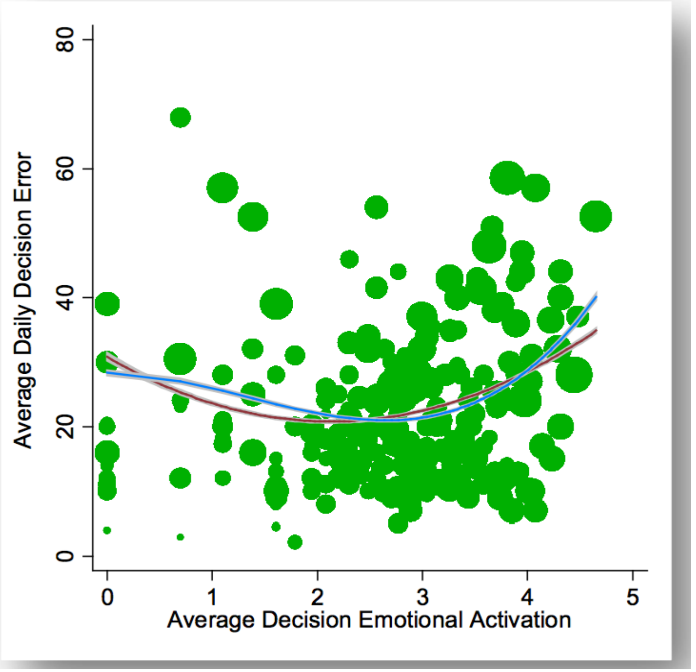
# Emotional Activation and Information Processing

- Motivation

  Emotional states can effect how information is processed. Good information can be undermined or strengthened by emotional states.

- Approach
  - Measure emotional activation in tweets using the ANEW dictionary
  - Control for the number of words in the text

- Preliminary experiments
  - Traders at a hedge fund are more likely to make decision errors when they are very emotionally activated or very emotionally deactivated.
  - Users who retweeted the 20 detected most anomalous sequences tend to post tweets with higher level of emotion than a baseline of 20 million tweets from June 2009.



# Inferring Cognitive Traits: Human Dynamics

– Information Reasoning & Morphing

– Visual Sentiment

➔ *What does it mean for A.I.s to have these human dynamics?*

# Information Morphing Modeling

- **Objective:** Modeling to detect and track information evolution in social media

- **Task Goals:**
  - – Resolve information morph, where new links keep emerging to give new meaning to existing phrases
    - Published at ACL 2013
  - – Detect information morph
    - Ongoing

weibo

*Peace West King* from ***Chongqing fell from power***, still need to ***sing red songs***?

CHINA DAILY

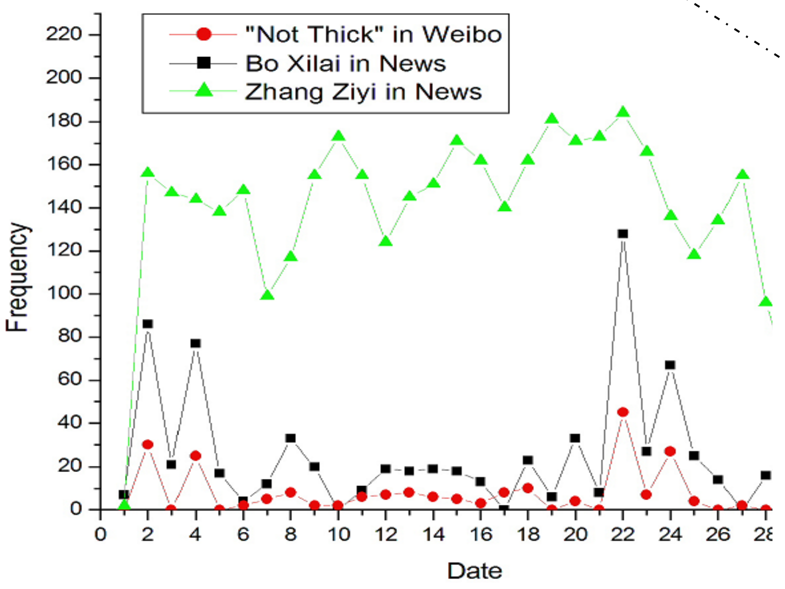
▪ *Bo Xilai* led ***Chongqing*** city leaders and 40 district and county party and government leaders to ***sing red songs***.

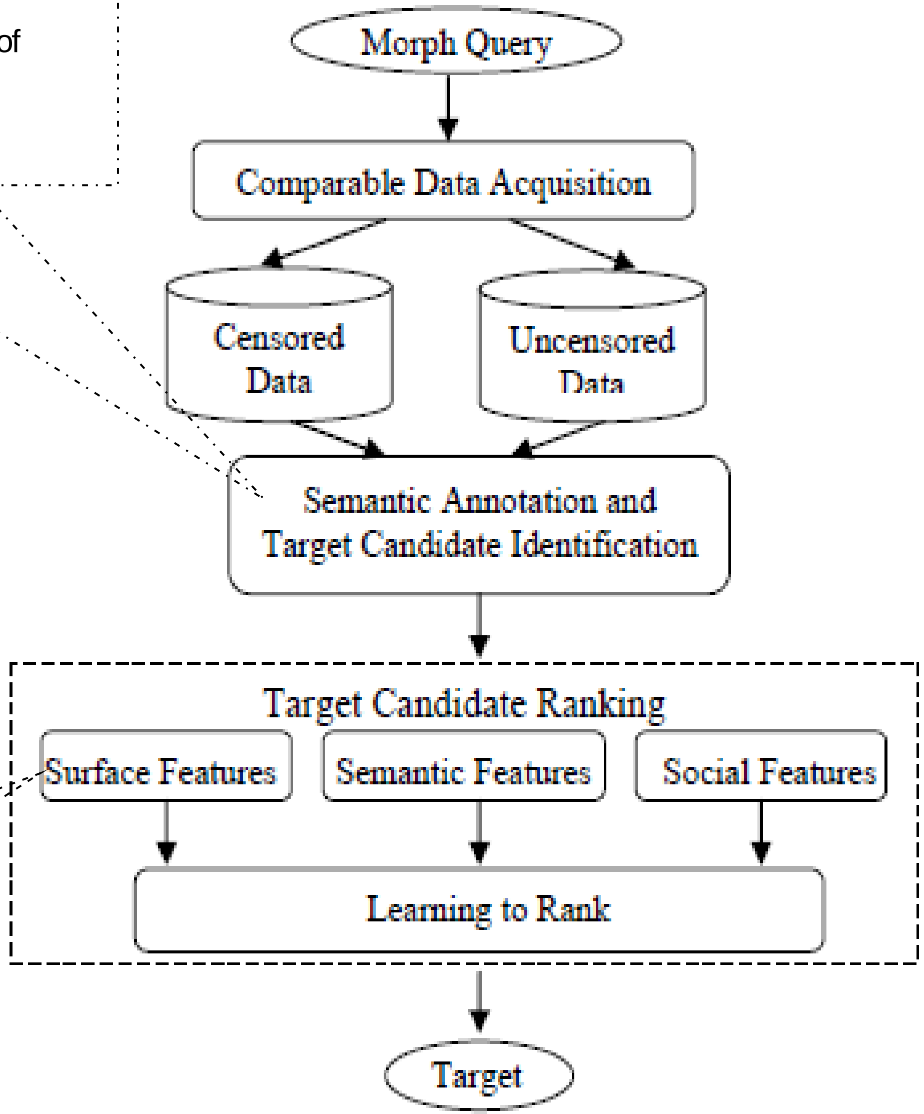
- **The Work:**
  - – Exploit rich structure information in heterogeneous information networks to improve existing NLP based methods

# Morph Resolution Approach Overview

**Target candidate identification**

- Narrow down candidate space by selecting pairs of similar temporal distribution



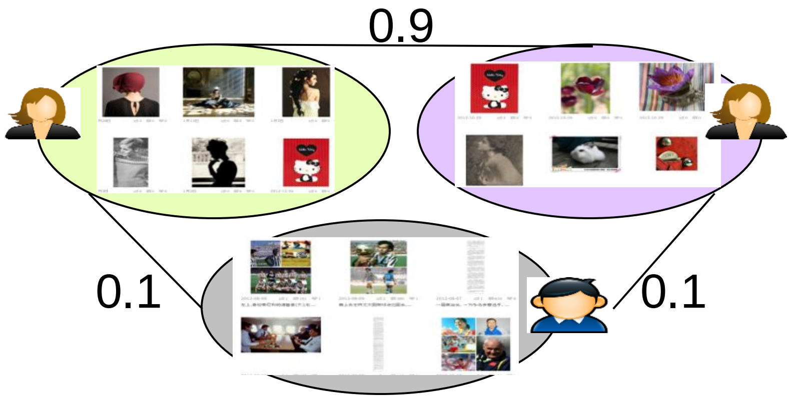




**Surface features (NLP)**

- Probabilistic Orthographic Model (e.g., string edit distances)

# Social Features

- Social networks based feature
  - A morph and its target are more likely mentioned by two users with strong social correlation
  - Measuring tie strength by social interactions (e.g., retweet, mention)




- Measure semantic similarity via social features

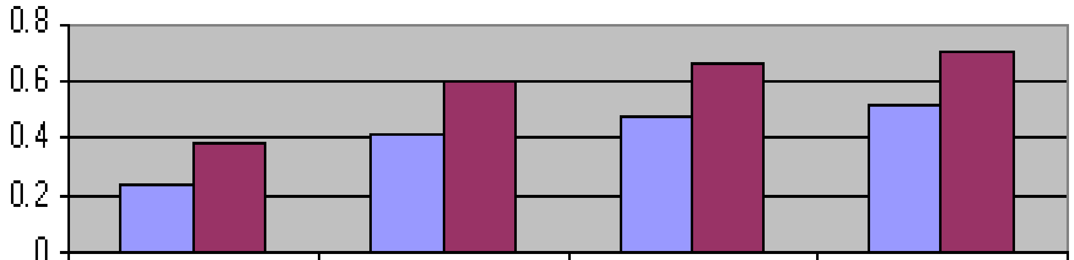
$$s(m,e) = \frac{\sum_{t \in T} f(e,t,U_t,U_c)}{|T|}$$

  - *T:* the set of temporal slots a morph *m* occurs
  - $U_t$: the set of users whose posts include m
  - $U_c$: be the set of close friends (i.e., social distance<0.5) for $U_t$
  - $f(e,t,U_t,U_c)$: an indicator function which returns 1 if one of the users in $U_t$ or $U_c$ posts tweets include the target candidate *e* within 7 days before *t*

- Data
  - Time frame: 05/01/2012-06/30/2012
  - 1555K Chinese messages from Weibo
  - 267K formal web news documents from embedded URL
  - 500K Chinese messages from English Twitter for sensitive morphs
  - Test on 133 morph entities in Weibo
- Scoring Metric

$$Acc@k = C_k / T$$

  - $C_k$: the number of correctly resolved morphs at top position K
  - $T$: the total number of morphs in ground truth




| | Acc@1 | Acc@5 | Acc@10 | Acc@20 |
|---|---|---|---|---|
| NLP | 0.234 | 0.416 | 0.477 | 0.519 |
| NLP+Graph | 0.379 | 0.594 | 0.659 | 0.701 |

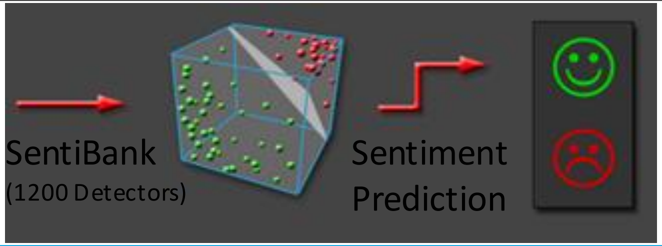
- **Objective:** Detect and track opinion-influencing visual memes

- **Task Goals:**
  - Goal 1 & concrete achievements
    - Since Feb 2012: Visual Sentiment Ontology (VSO) & SentiBank classifiers
    - Since Feb 2013: Release dataset, browsers, 1200 classifiers, and twitter sentiment prediction demo
    - By Jan 2015: Adaptive analysis for specific user groups and data domain
  - Goal 2
    - Enhance communication and story telling using multimedia content

- **The Work:**
  - Achieves the first multimedia-based sentiment prediction system
  - Facilitates trend tracking and informed decision making based on multimedia

- **New Publications, Awards:**
  - **[ACMMM2013 BNI Paper and Demo]:** Large-scale Visual Sentiment Ontology and Detectors Using Adjective Noun Pairs

- **Tech transition strategy:**
  - Adapt SentiBank classifiers and VSO browser to analyst target content

# The Power of Social (Visual) Multimedia

## - A picture is worth one thousand words

Tweets of the Year

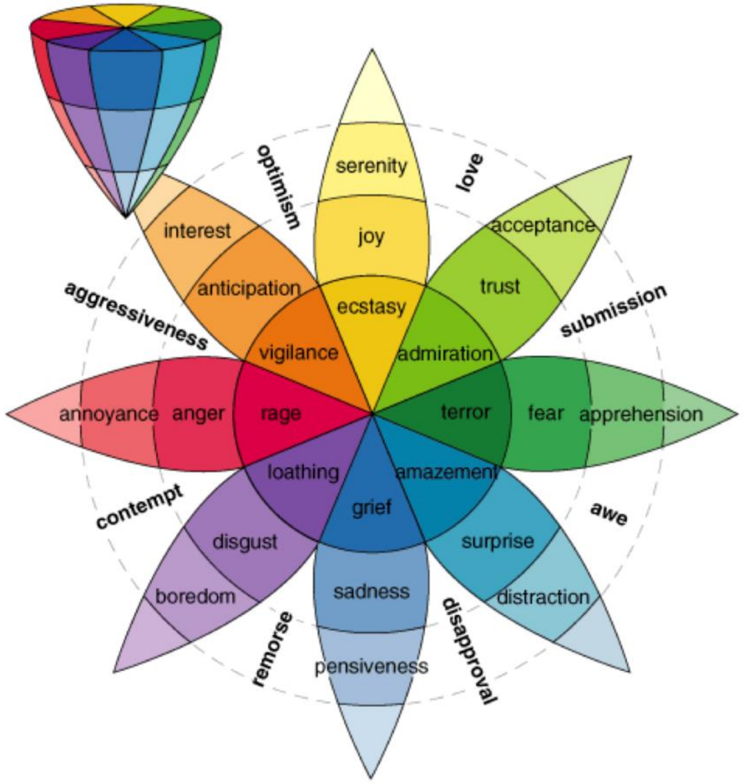
@BarackObama: Four more years.

@Brynn4NY: Rollercoaster at sea.

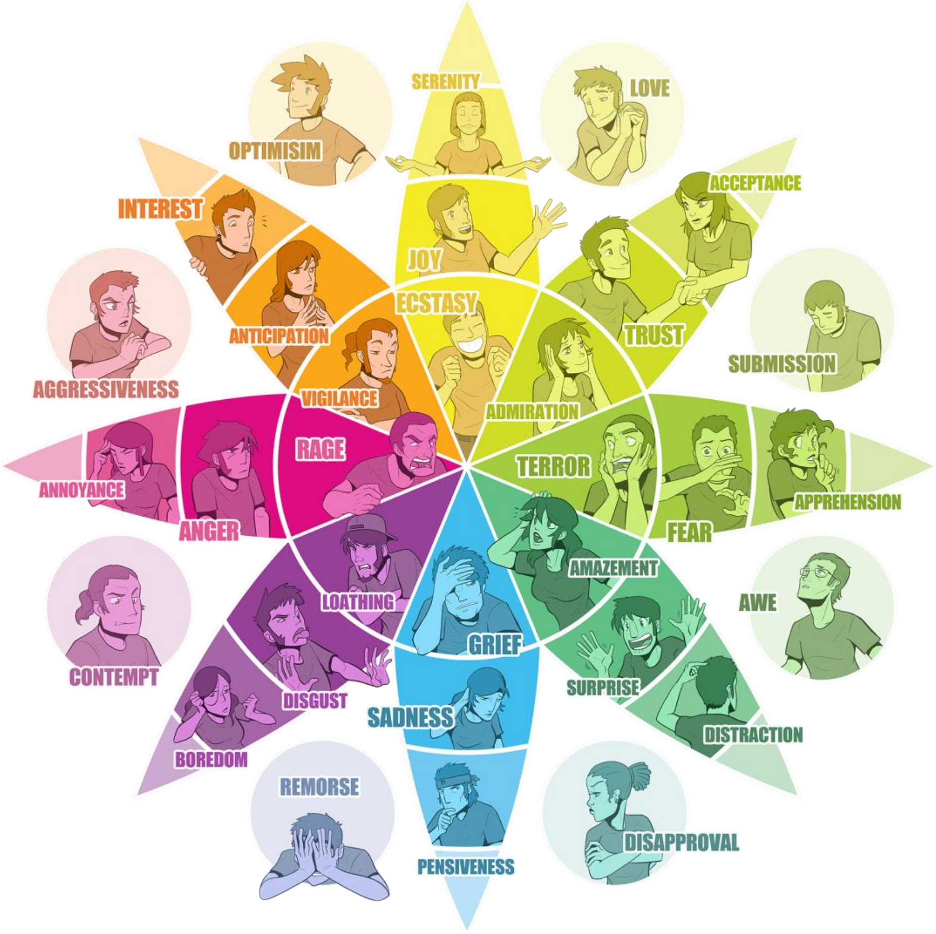
# Question: How to Build Visual Sentiment Ontology?

-- Web + big data + computer vision + psychology

-- Web + big data + computer vision + psychology




Psychology emotion wheel
(24 emotions, by Robert
Plutchik)

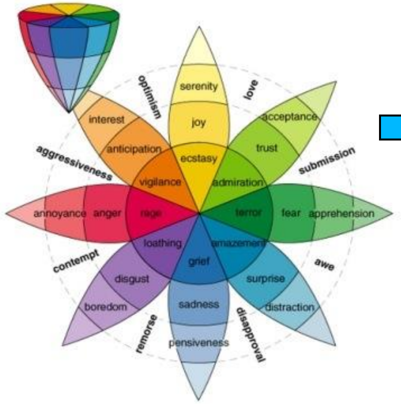



**Plenty on the Web:**
***"For content to go viral, it needs to be emotional,"* Dan Jones**

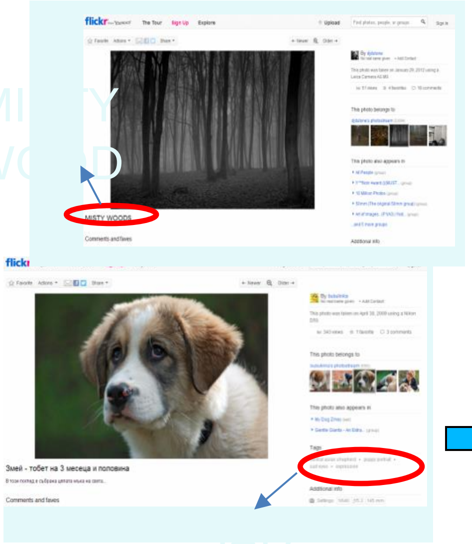
  

# Research: Which 1000 sentimental concepts?

-- data mining to discover visual sentiments in social media
-- data mining to discover visual sentiments in social media




Psychology emotion wheel (24 emotions)




Analyze tags with strong sentiments

**Select Adj-Noun Pairs**

Build Sentiment Ontology

# Concurrent tags with emotions

| joy | terror | amazement | disgust |
|---|---|---|---|
| joy | terror | amazing | disgusting |
| happy | horror | beautiful | gross |
| love | zombie | nature | food |
| smile | fear | wonder | nasty |
| beautiful | dark | light | sick |
| flowers | street | love | dirty |
| light | halloween | sky | dead |
| nature | war | eyes | face |
| kids | undead | clouds | blood |
| christmas | bomb | landscape | insect |

**From 6 million tags on Flickr and YouTube**
**Color code: text sentiment values**

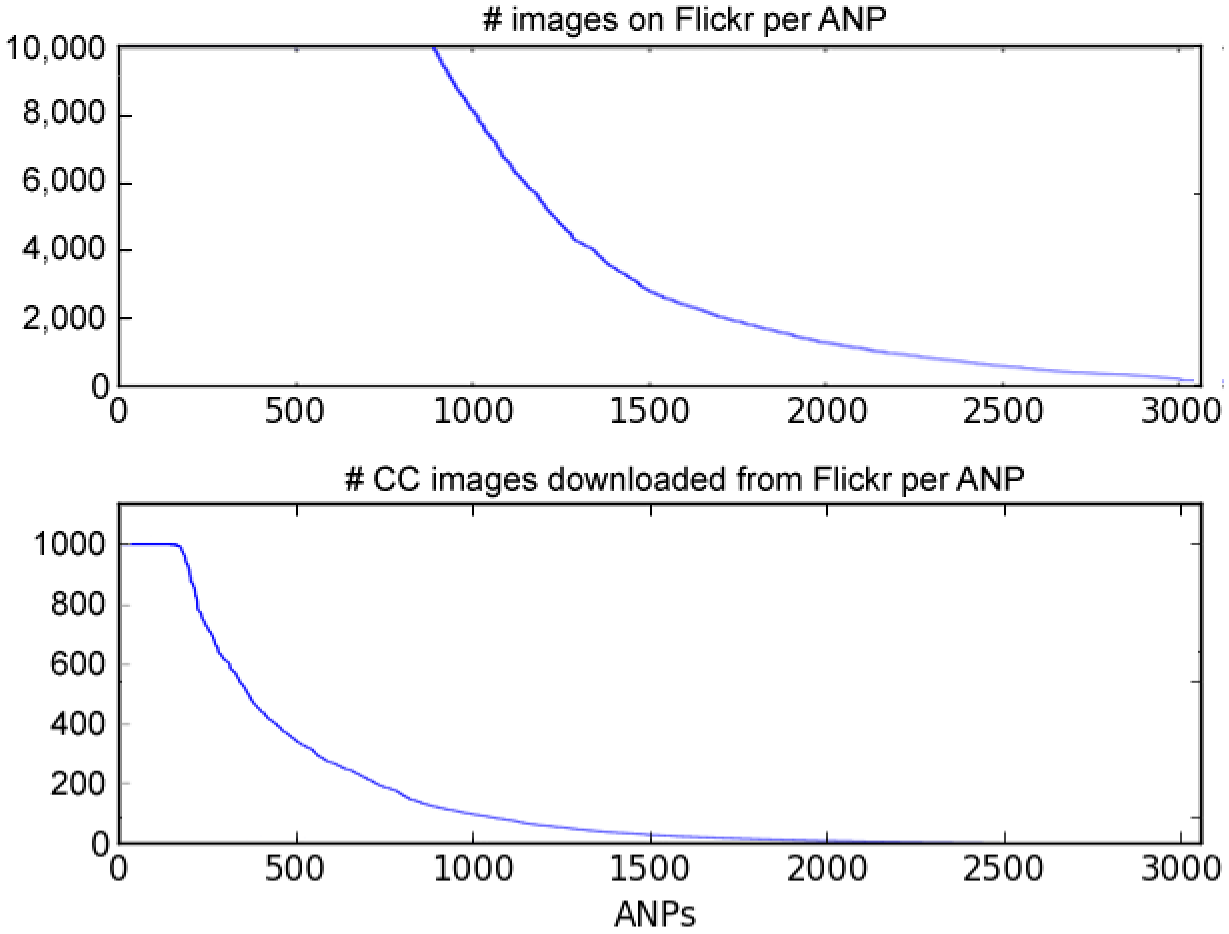
# Target Concepts for CV – Adj-Noun Pair

- Adjective (268): indicate strong emotions
  - frequent positive Adj: beautiful, amazing, cute
  - frequent negative Adj: sad, angry, dark
- Nouns (1187): more detectable by computer vision
  - Noun categories: people, places, animals, food, objects, weather
- Other cleaning steps:
  - remove named entities like “hot dog” via wikipedia
  - Choose sentiment rich ANP concepts by tools “Senti-WordNet” “SentiStrength”

# Image Datasets

About 0.5 million images over 3000 concepts



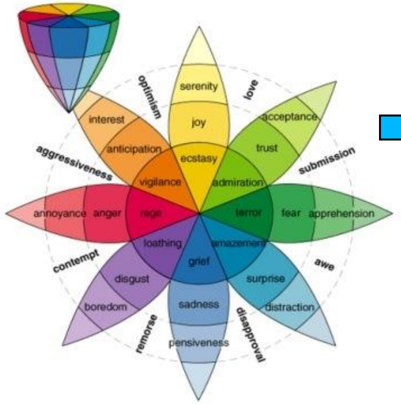
Beautiful Sky

Beautiful Flower

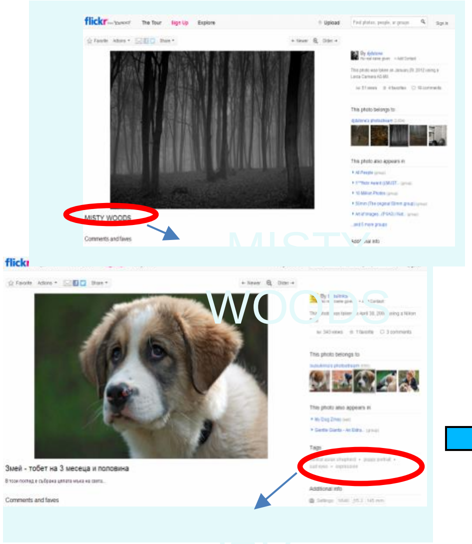
Sad Eyes

Happy Face

# Teach Machine to Recognize Visual Sentiments

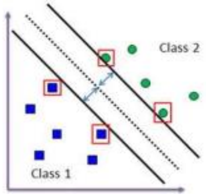



Psychology emotion wheel (24 emotions)




Select Adj-Noun Pairs

Build Sentiment Ontology




Train Classifiers

**Performance Filtering**

**SentiBank (1200 Detectors)**

Sentiment Prediction

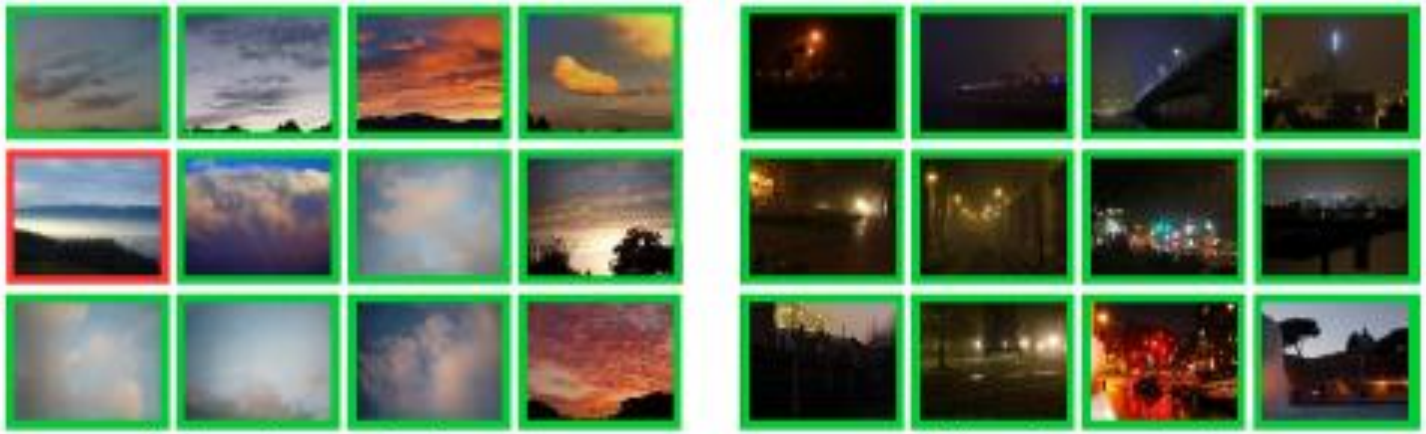

# Classifier Training

– LibSVM, cross validation
  – Generic features
  – Color Histogram (3x256 dim.)
  – GIST descriptor (512 dim.)
  – Local Binary Pattern (52 dim.)
  – SIFT Bag-of-Words (1,000 codewords
  – 2-layer spatial pyramid, max pooling)
  – Classemes descriptor (2,659 dim.)

– Special features
  – Object detection (people, objects, etc.)
  – Aesthetics features (color schemes, layout, etc.)
  – Face and attributes
  – Improve accuracy 9%-30%

# Examples

Good Results:



(a) colorful clouds (b) misty night

Not Great Results:

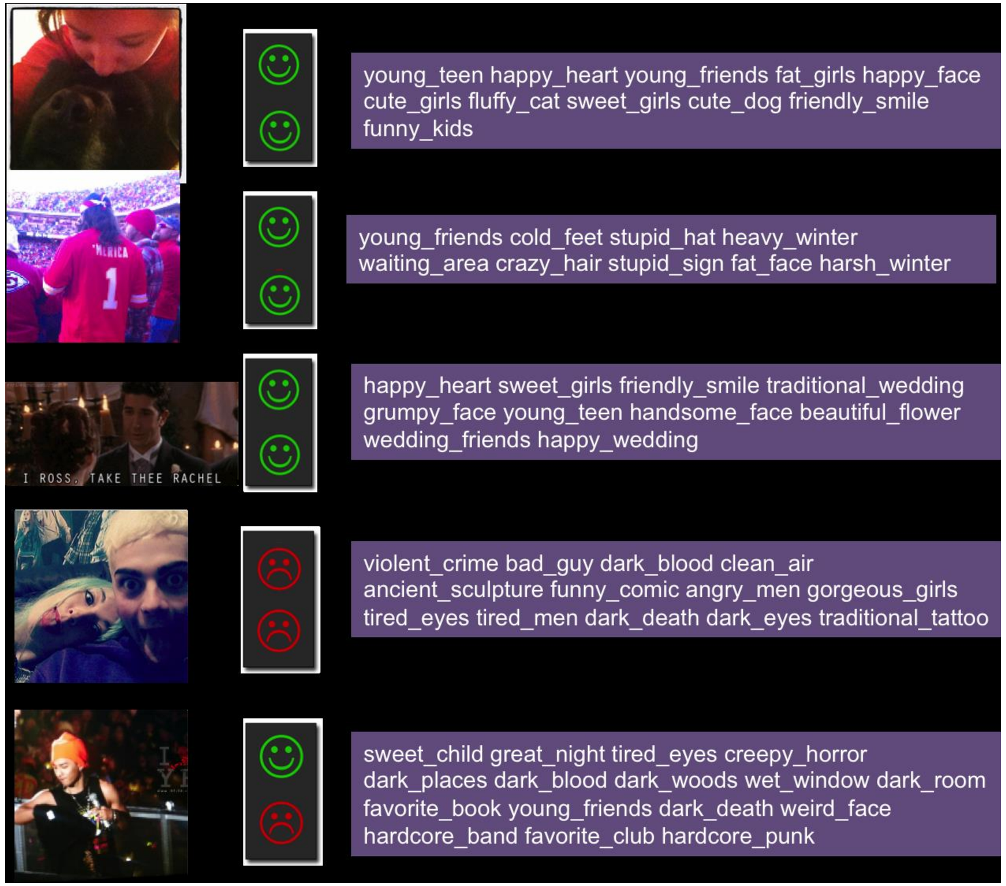
(c) colorful butterfly (d) crying baby

# Sentiment Prediction Performance

Sentiment Prediction Accuracy

| | |
|---|---|
| **Text** | **0.43 (many are neutral)** |
| **Visual** | **0.70** |
| **Text-Visual (Joint)** | **0.72** |

Examples



# Social Media Monitoring Demo

– Live Monitoring:
  – Retweets, Visualization, Translation, Sentiments
  – Relationships, Communities
  – Geo-Location, Intranet Social Media

– Forensic Anomaly Detection:
  – Abnormal Threads
  – Personality, Value, Needs, Trustworthiness, Emotions,..
  – Visual Sentiments

# Social Media Live Monitoring


Monitoring categories

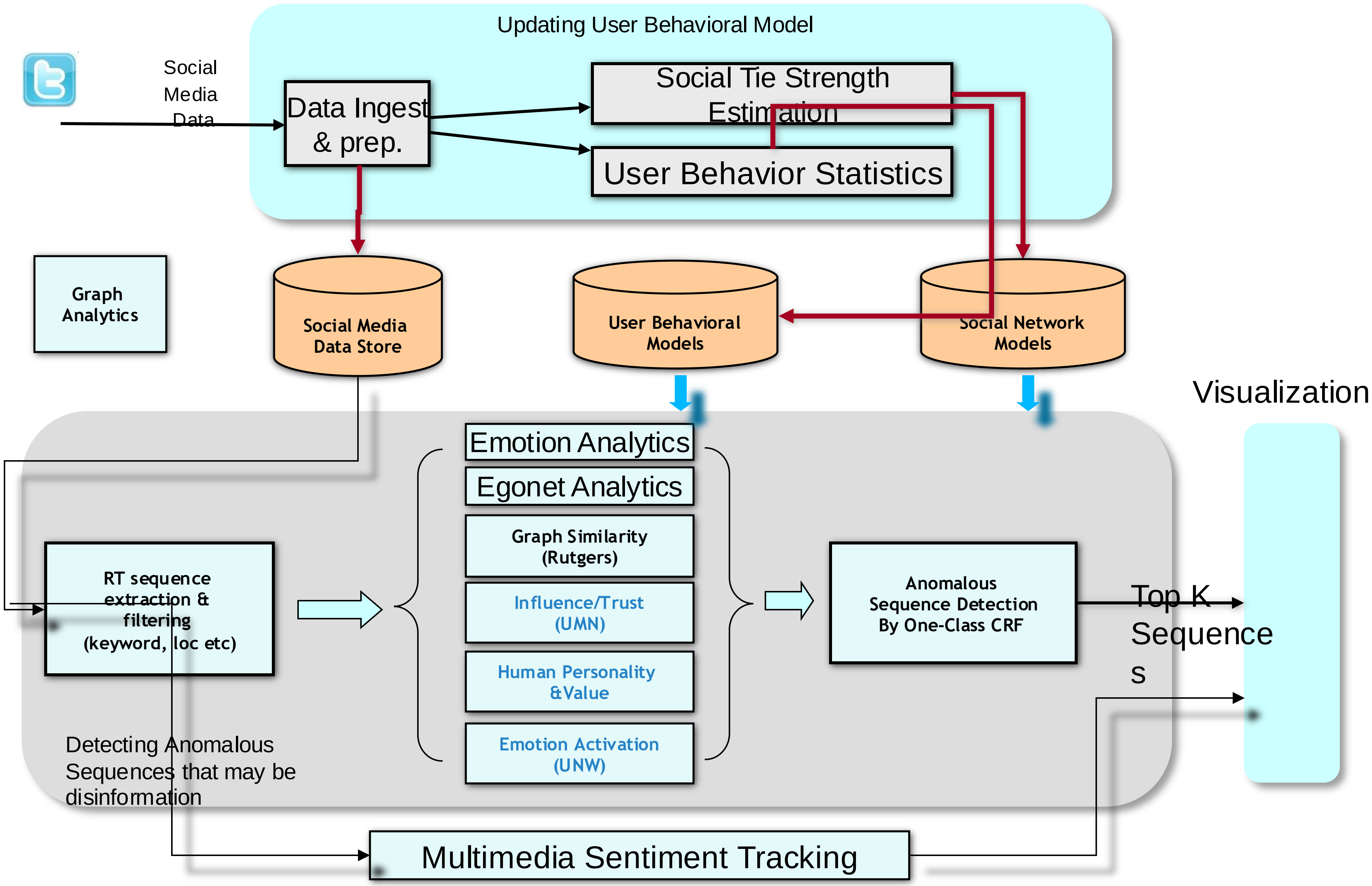
Monitoring filter

Live Tweets, Sentiment, Keywords

Growing Influential Retween Graphs

Real-Time Translation, Locations, Top Retweets

# Social Media Forensic System

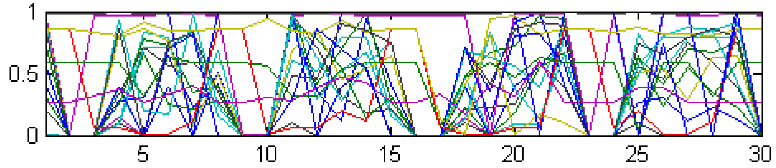


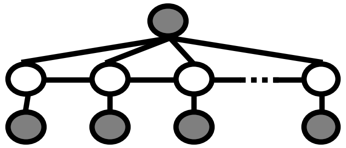
# Detecting Anomalous Information Spreading

- Motivation
  - People's dynamic reactions to information (e.g., retweeting) give clues to information credibility and quality
    - For example, trustworthy people may take time to verify uncertain information from strangers before spreading it
- Approach
  - Use one-class conditional random field to model people's behavior in information spreading sequences and detect anomalous sequences
    - Features: content features such as emotion, network features such as tie strengths and clustering coefficients
- Preliminary experiments
  - Detect anomalies in retweeting sequences during Hurricane Sandy
    - Including hijacker, fake pictures spreaders

Detected as top 1 anomaly in Sandy Tweets

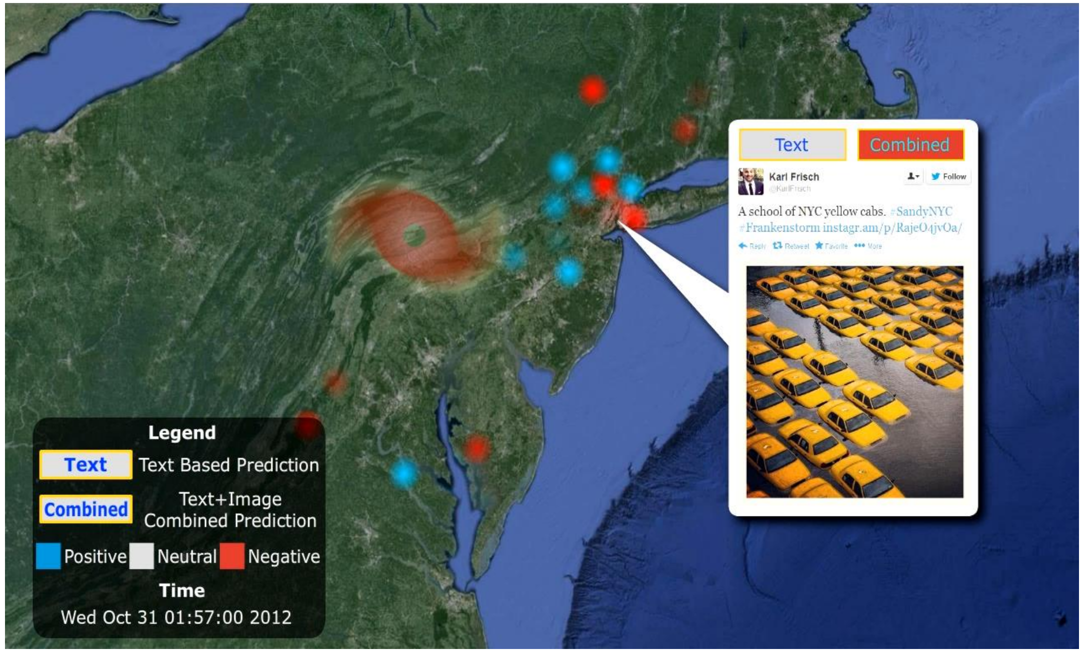
One-class CRF to detect temporal anomalies



# Photo Tweet Sentiment Tracking during Hurricane Sandy

- **Goal:** Detect sentiment during Hurricane Sandy.
- **Data collection:**
  - Date: Oct 25 – Nov 02
  - Hashtags (based on popularity): #prayforusa, #frankenstorm, #nyc,#hurricane,#sandy,#hurricanesandy, #staysafe, #redcross,#myheartgoesouttoyou,…
  - 2000 Photo Tweets collected

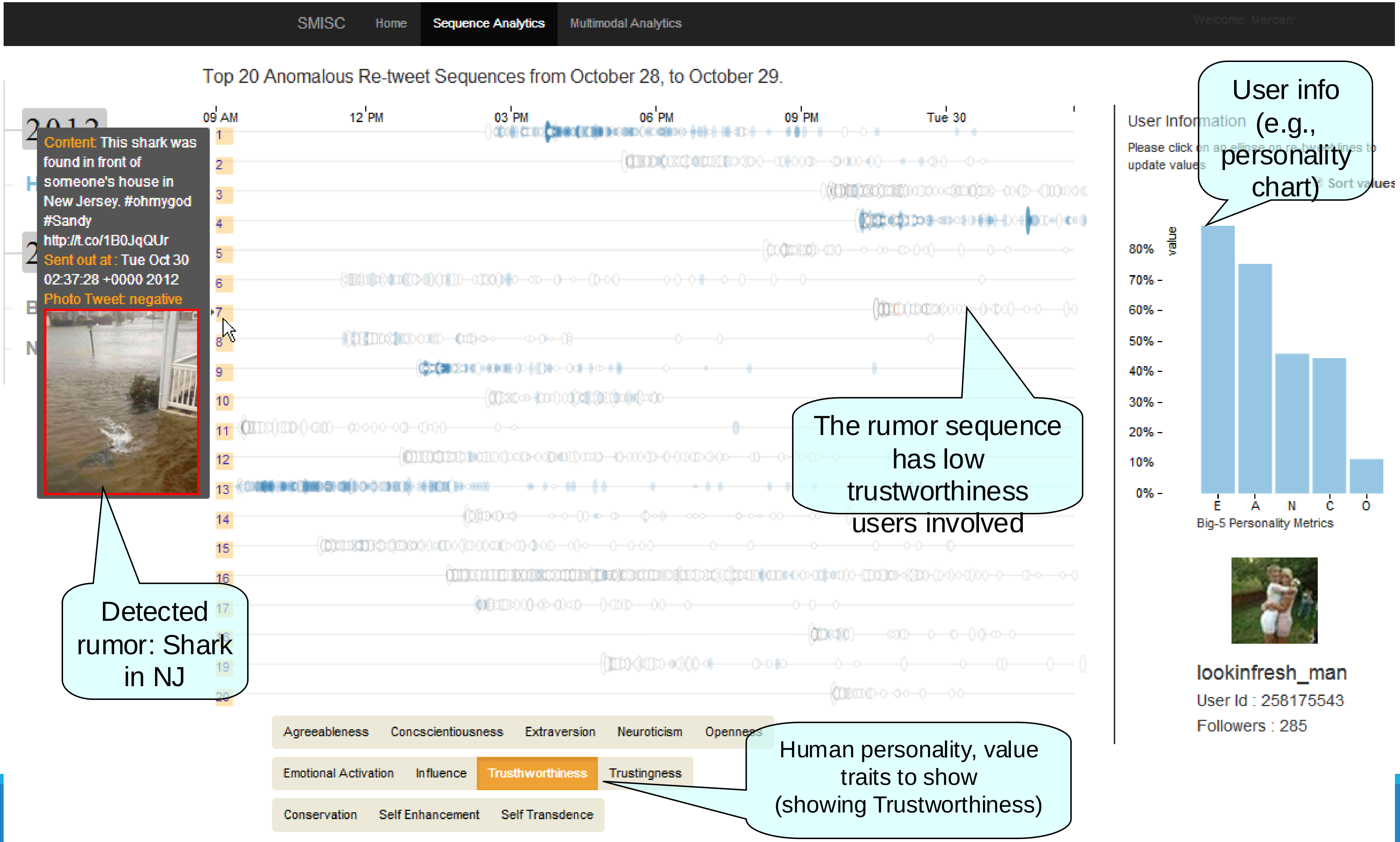



- **Ground Truth Labeling:**
  - 1340 unanimously agreed labels from 2 individuals
- **Training Classifier:**
  - Text (SentiStrength)
  - Visual(SentiBank, Logistic Regr.)
  - Training/Testing ratio: 4:1
  - 5-fold cross-validation
  - Accuracy(Text-Visual Combined): 72%

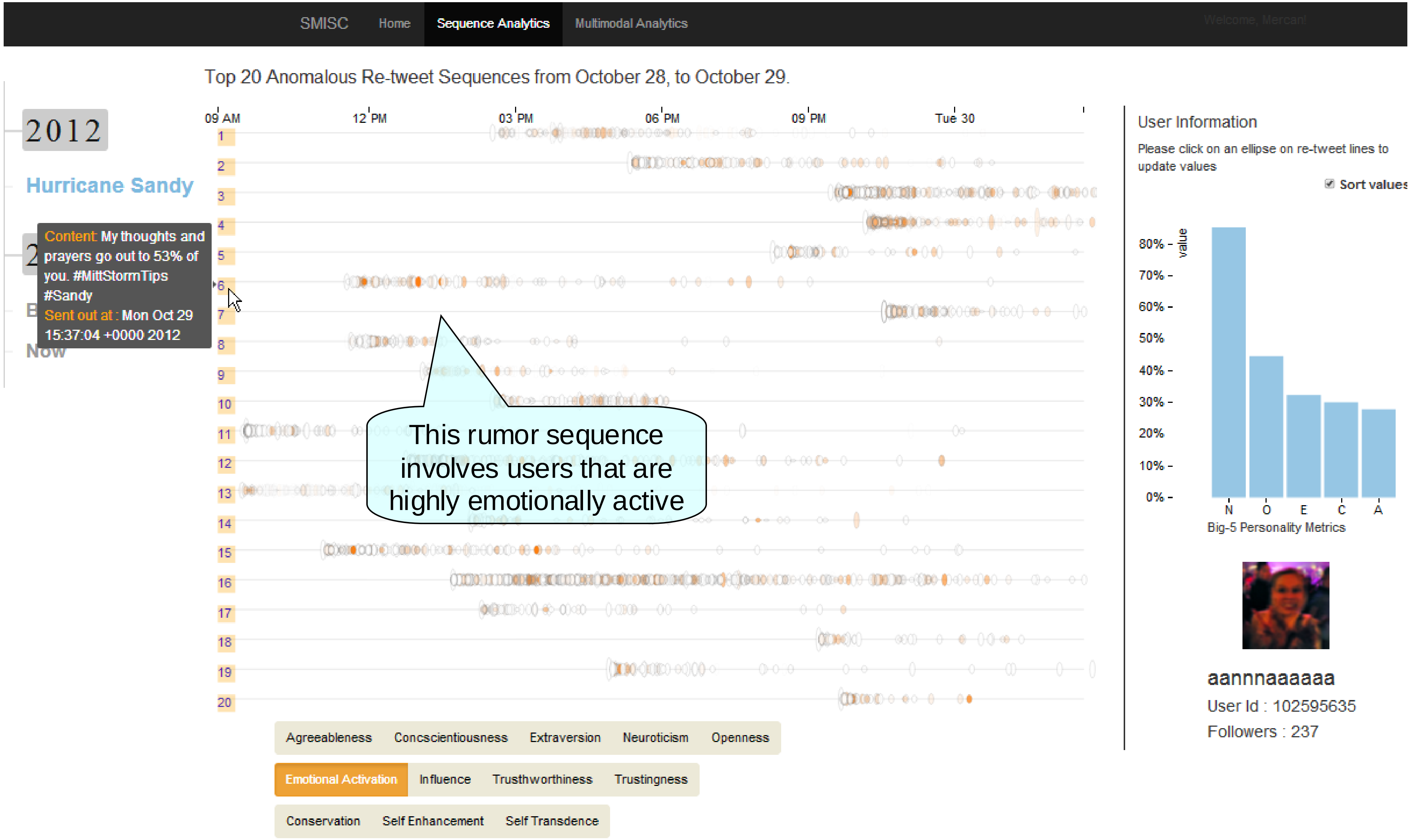
  

# Forensics on Finding Anomalies

Detect disinformation spreading (retweet) during Hurricane Sandy by people's dynamic behavior in response to the information

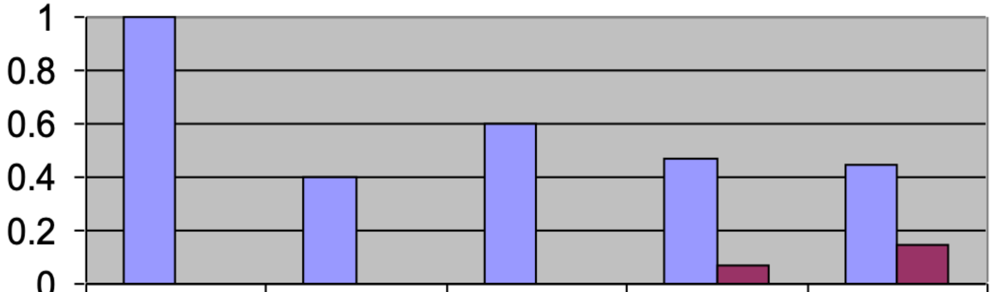


# Users can interactively analyze the personality traits, emotion activation, value and other traits of the people involved



# Evaluation with Hurricane Sandy Data

- Goal: detect disinformation spreading during Hurricane Sandy 2012
- Dataset: 10% Twitter feed on Oct 29, 2012 when Hurricane Sandy hit NYC
- Ground truth: annotation based on verified disinformation revealed by credible sources afterwards
  - E.g., http://mashable.com/2012/10/29/fake-hurricane-sandy-photos/
- Metric: accuracy of detected disinformation in top-K results
  - Baseline: One-Class SVM using the same set of features

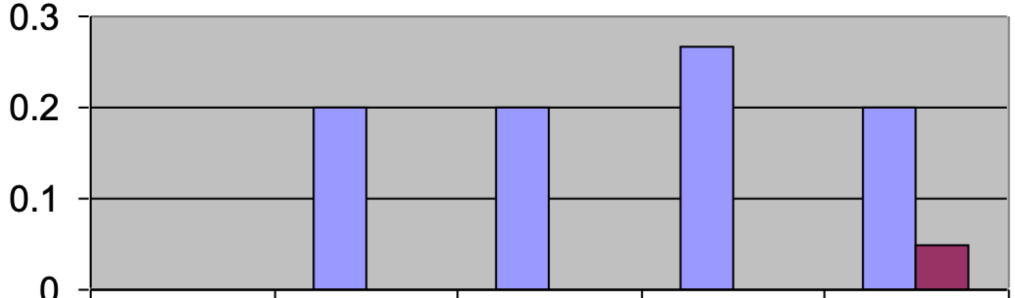



| | Acc@1 | Acc@5 | Acc@10 | Acc@15 | Acc@20 |
|---|---|---|---|---|---|
| OCCRF | 1 | 0.4 | 0.6 | 0.467 | 0.45 |
| OCSVM | 0 | 0 | 0 | 0.067 | 0.15 |

OCCRF OCSVM

# Evaluation with Boston Bombing Data

- Goal: detect disinformation spreading during Boston Bombing 2013
- Dataset: 10% Twitter feed during Apr 15-19, 2013
- Ground truth: annotation based on verified disinformation revealed by credible sources afterwards
  - E.g., http://www.snopes.com/politics/conspiracy/boston.asp
- Metric: accuracy of detected disinformation in top-K results
  - Baseline: One-Class SVM using the same set of features




| | Acc@1 | Acc@5 | Acc@10 | Acc@15 | Acc@20 |
|---|---|---|---|---|---|
| OCCRF | 0 | 0.2 | 0.2 | 0.267 | 0.2 |
| OCSVM | 0 | 0 | 0 | 0 | 0.05 |

OCCRF OCSVM

## Anita avatars are earning: $2,503.26

ANITA-324658
PER $22,630 EARN: $-467.51

ANITA-253758
PER $30,178 EARN: $1,106.20

ANITA-247917
PER $31,809 EARN: $350.48

ANITA-428339
PER $39,494 EARN: $620.17

ANITA-164762
PER $29,395 EARN: $-17.07

ANITA-450214
PER $36,088 EARN: $178.12

ANITA-247502
PER $46,253 EARN: $318.35

ANITA-267139
PER $21,287 EARN: $44.81

ANITA-544716
PER $46,442 EARN: $166.03

ANITA-418870
PER $28,764 EARN: $21.32

ANITA-432722
PER $24,712 EARN: $132.59

ANITA-208134
PER $16,576 EARN: $49.76

# Personality driven AI Trader

Anita
Graphen Artificial Intelligence Traders

Home ForeignExchange Stocks Bonds

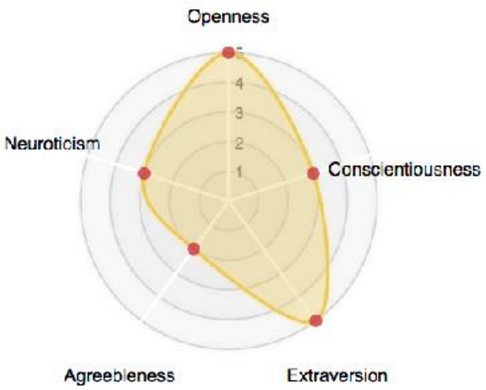
## Anita 267139

-- an Adventurous AI Trader

***Specialized** at: EUR-USD*

***Knowledgable** of: Oil, Gold and Twitter*

***Strategy Learning Frequency** at: 2.0 hours*




**Original: $1,000.00, Current: $1,404.50, Performance: *Gain $404.50***

### Activities

| Time | Action | Cash | Unit | Balance |
|---|---|---|---|---|
| 2017-10-12 13:45:05 | Sell 50,000 | $1,404.50 | 0 | $1,404.50 |
| 2017-10-12 12:57:25 | Buy 100,000 | $-57,792.00 | 50,000 | $1,386.50 |
| 2017-10-12 11:19:10 | Sell 100,000 | $60,577.00 | -50,000 | $1,372.00 |
| 2017-10-12 11:11:55 | Buy 100,000 | $-57,822.00 | 50,000 | $1,366.00 |
| 2017-10-12 09:08:05 | Sell 100,000 | $60,566.00 | -50,000 | $1,310.00 |
| 2017-10-12 08:34:40 | Buy 100,000 | $-57,935.00 | 50,000 | $1,287.50 |

Graphen Anita creates AI Traders with various Autonomous Learning Capabilities and Trading Behavior Personalities

# Personality driven AI Trader

Columbia University

Anita
Graphen Artificial Intelligence Traders

Home ForeignExchange Stocks Bonds

## Anita 247502

-- an Independent AI Trader

***Specialized** at: EUR-USD*

***Knowledgable** of: FX, Gold and Twitter*

***Strategy Learning Frequency** at: 100.0 days*


**Original:** $1,000.00, **Current:** $1,119.50, **Performance:** *Gain $119.50*

### Activities

| Time | Action | Cash | Unit | Balance |
|---|---|---|---|---|
| 2017-10-12 14:56:00 | Buy 50,000 | $1,119.50 | 0 | $1,119.50 |
| 2017-10-12 13:56:35 | Sell 100,000 | $60,304.00 | -50,000 | $1,048.50 |
| 2017-10-12 11:51:25 | Buy 100,000 | $-58,196.00 | 50,000 | $1,012.00 |
| 2017-10-12 10:56:10 | Sell 100,000 | $60,232.00 | -50,000 | $992.50 |
| 2017-10-11 16:46:45 | Buy 100,000 | $-58,236.00 | 50,000 | $1,066.50 |
| 2017-10-11 15:13:20 | Sell 100,000 | $60,382.00 | -50,000 | $1,065.00 |