Credibility in Context

An Analysis of Feature Distributions in Twitter

John O'Donovan, Byungkyu Kang, Greg Meyer, Tobias Höllerer
University of California
Santa Barbara, USA

Sibel Adalı
Rensselaer Polytechnic Institute
Troy, New York, USA

September 5th 2012, Amsterdam, The Netherlands. (SocialCom 2012)

Outline

- **Background**
  - Motivation
  - Research Questions
  - Contribution
  - Credibility
  - Related Work
- Features and Contexts
- Experimental Framework
- Results
- Conclusion

Motivation

- Growth of Social Media
  - User-Generated Content (UGC)
  - Information Overload
- “Credibility” models can help to identify useful information. They can leverage **historical** and **current** information available through social web APIs
- But… Indicators of credibility vary across contexts. There is a need for more adaptive models.

What is Credibility?

- Broad use, many different definitions:

  Social (Golbeck, Ziegler), Cognitive (Gray, Todorov), Computational (Marsh, Josang), Psychological (Dellorcas, Erikson)

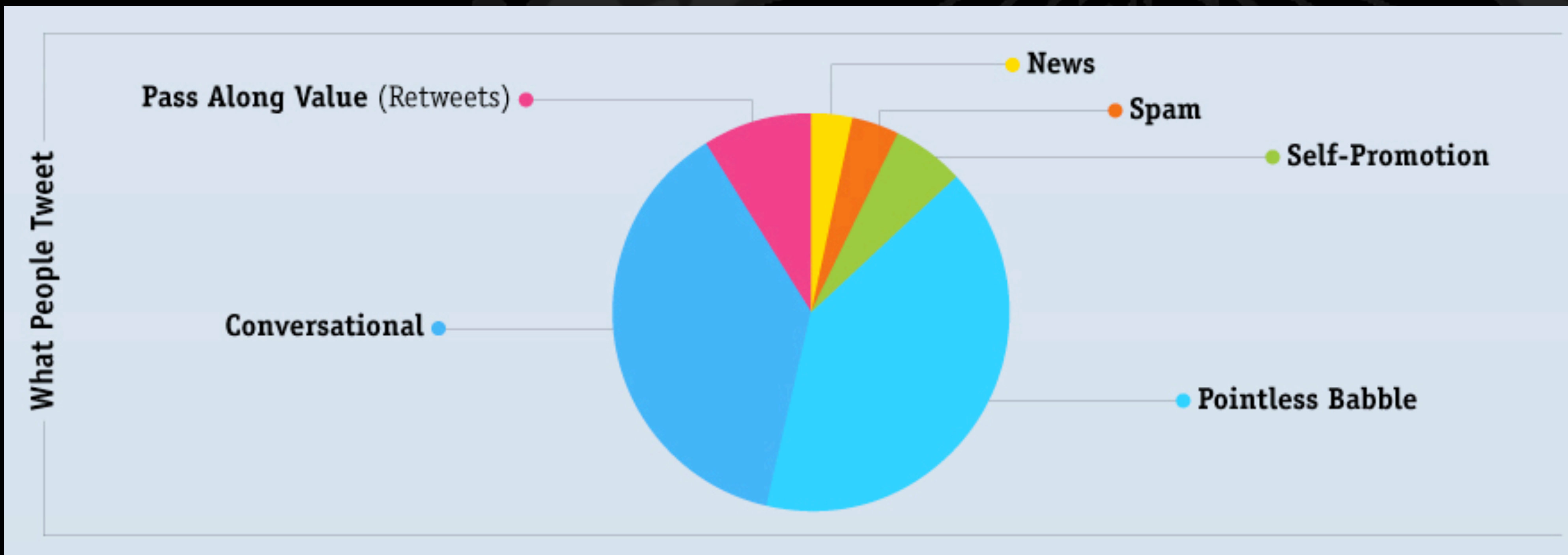
Message-level Credibility

A degree of believability that can be assigned to a tweet about a target topic, i.e.: an indication that the tweet contains believable information.

Social Credibility

The expected believability imparted on a user as a result of their standing in the social network, based on any and all available metadata.

Lots of useless information?



(Excerpt from mashable.com infographic)

Examples

- Useless / Nonsensical Tweets
  - *"yo yo yo, looky here!!"*
- Spam Tweets
  - *'Have you heard millions of people are making $5k+/Mo from home? heres how...t.co/blah'*
- Credible / Newsworthy Tweets
  - *Great keynote by Todorov at #SocialCom2012 in #Amsterdam*
  - *#LADodgers commentator #vinscully back for another season!*
- Personal / Conversational
  - *@anTusail: thanks for the info!*

Related Work - *Credibility Evaluation*

Classification-based

Supervised	Semi-supervised	Clustering
Kang et al. 2012 Castillo et al., 2011	Bian et al., 2009 Yin & Tan, 2011	Gupta et al., 2011 Canini et al., 2011

Graph Models

Agichtein et al., 2008

Similarity-based Approaches

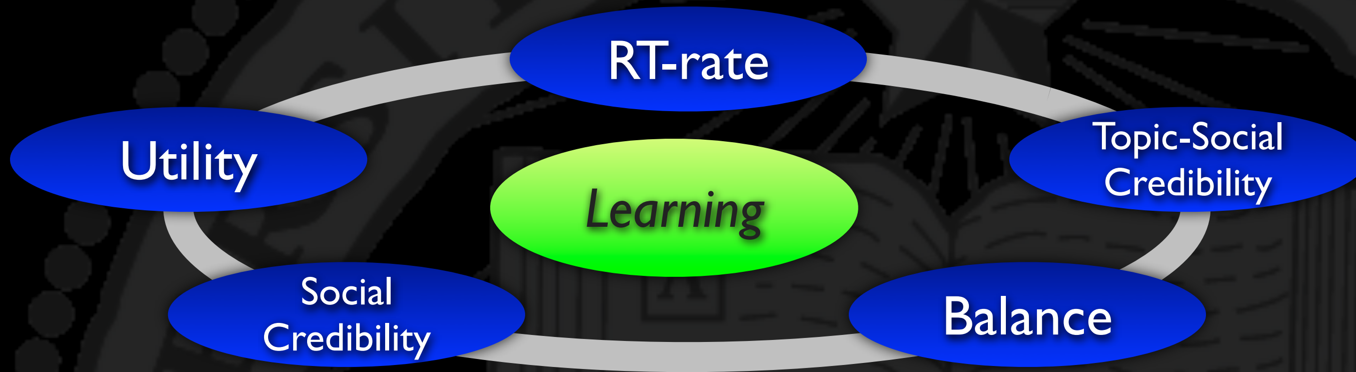
Juffinger et al. 2009, O'Donovan 2005

Game Theory Models

Ghosh & McAfee, 2011

Credibility Models

Social Model



Content Model

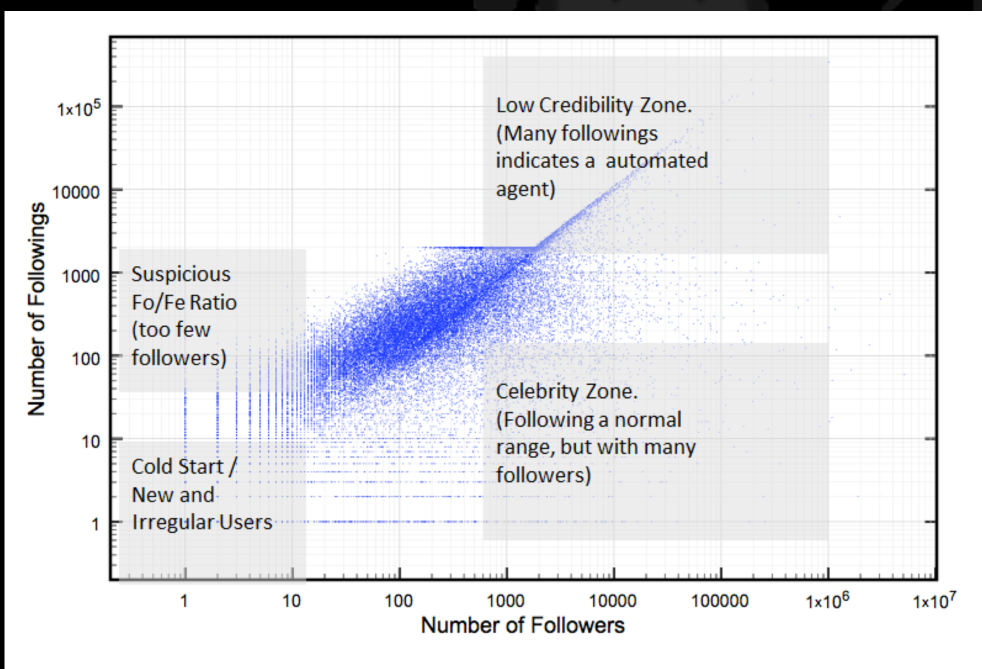
Numeric Indicators	Binary Indicators
Positive Sentiment Factor	Is Only Urls
Negative Sentiment Factor	**Is a Retweet**
Sentiment Polarity	**Has a Question Mark**
Number of intensifiers	**Has an Exclamation Mark**
Age of Profile	**Has multiple Questions/**
Number of popular topic-specific terms	Has a positive emoticon
Number of Uppercase Chars	Has a negative emoticon

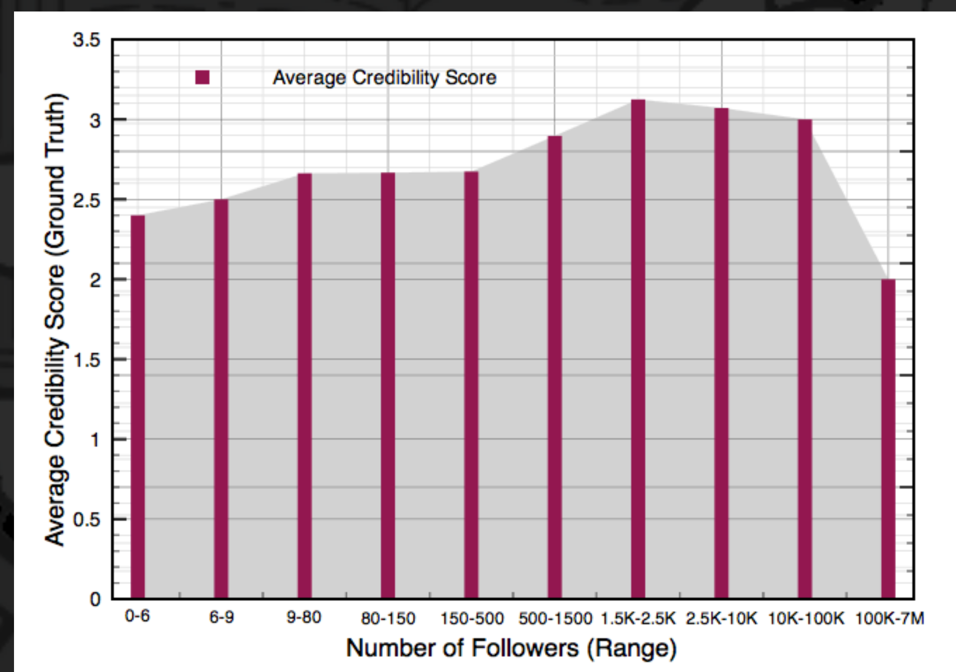
Hybrid Model

Social Features A, B C… | Content Features X,Y,Z…

Initial Experiments (Kang '12)

- Social model outperforms content-based and hybrid model
- Approximately 88.5% accuracy predicting manually labeled tweets using J48 Learner using our Social Model
- However, results varied greatly across different topics.

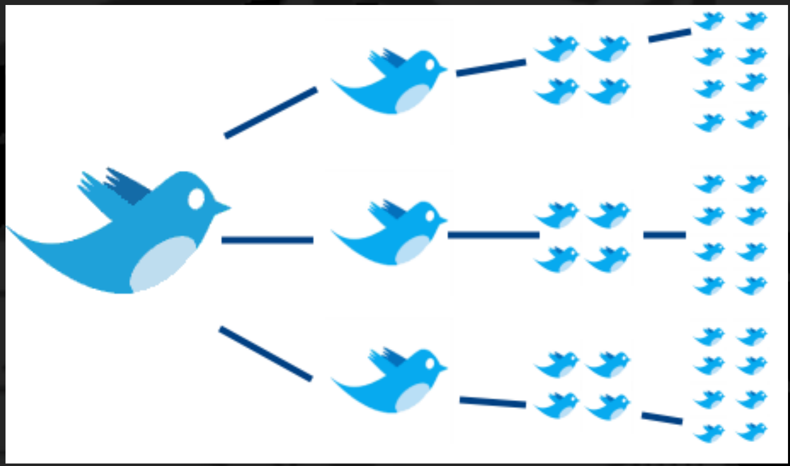




Research Questions

- How are the features that indicate credibility distributed in Twitter?
- How and why do they vary across different contexts?
- How do we use knowledge of feature distribution to create more adaptive, better performing credibility-based information filters?

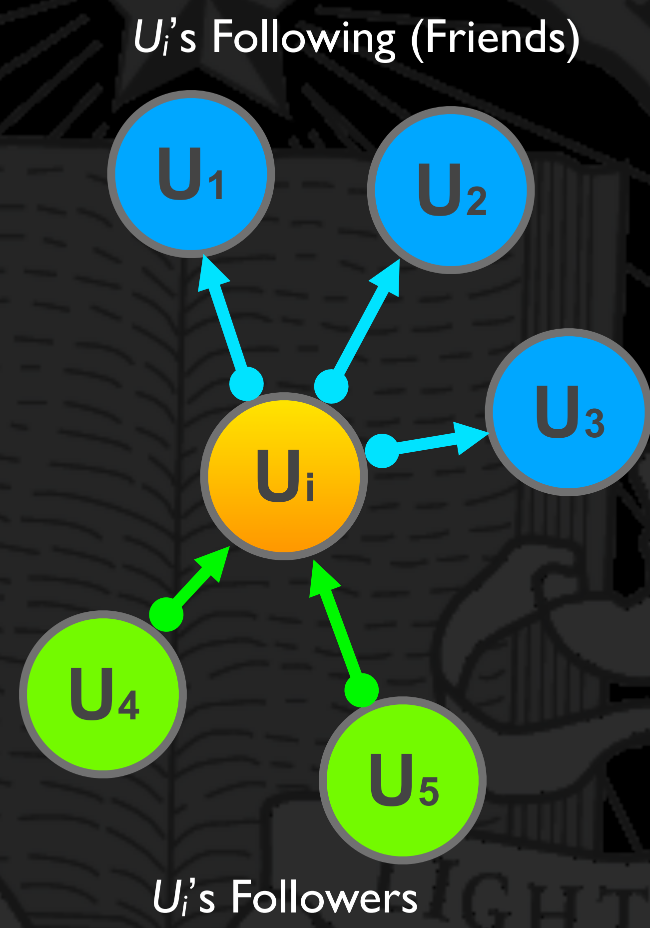
Outline



- Background
- **Features and Contexts**
  - Terminology
  - The Twitter Graph
  - Details of Social, Content, and Hybrid Model
- Experimental Framework
- Results
- Conclusion

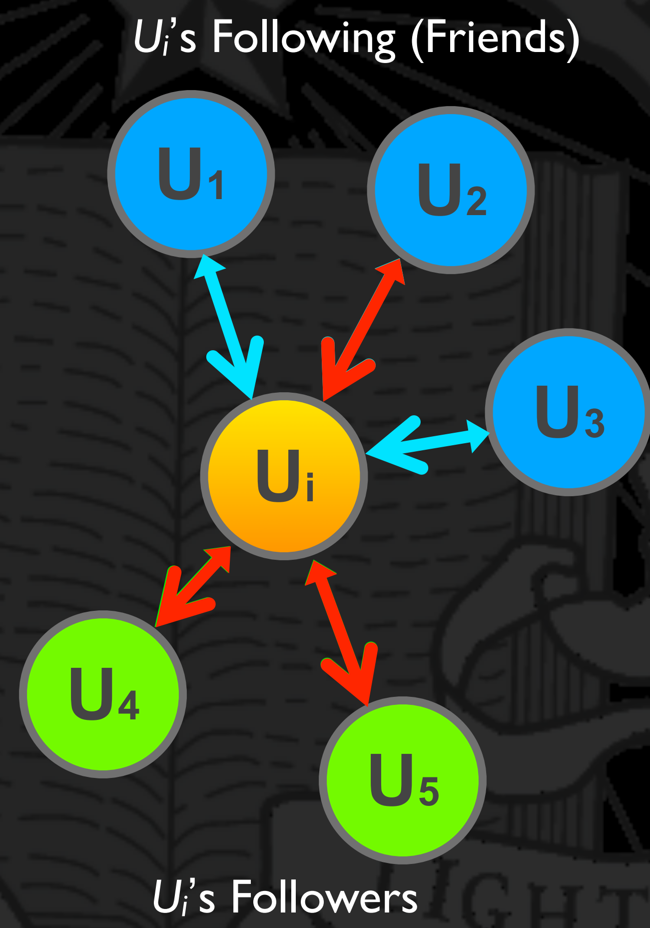
The Twitter Graph

- Follower Group
  - *the people who receive my Twitter updates*
- Following Group
  - *the people I follow (their Twitter updates appear in my personal timeline)*

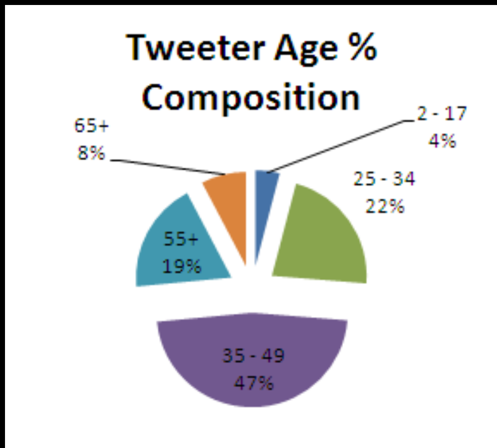


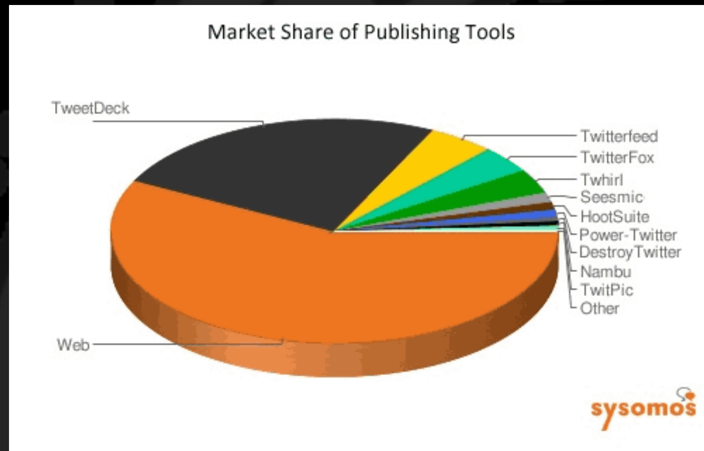
The Twitter Graph

- Follower Group
  - *the people who receive my Twitter updates*
- Following Group
  - *the people I follow (their Twitter updates appear in my personal timeline)*



Slicing the Twitter Graph:





Class	*Description*	*# of Contexts*
Diverse Topics	Diverse topics in Twiter; eg: #Romney #Facebook	8 different topics (see Table II)
Credibility	Manually provided assessments of tweets	Credible or non credible
Chain length	Mined retweet chains and classified based on length	Long or short
Dyadic pairs	Mined interpersonal interaction and classified	Dyadic or not dyadic

Feature Sets

Three classes of features were used: Social, Content-based and Behavioral/Dynamic.

Social

Name	*% Present*	*Average score*	*Class*
Age	100.00	610.64	Social
listed_count	100.00	11.82	Social
status_count	100.00	554.49	Social
status_rt_count	100.00	10.17	Social
favourites_count	100.00	57.96	Social
followers	100.00	295.15	Social
followings	100.00	315.03	Social
fofe_ratio	100.00	5.81	Social

Feature Sets

Content-based

Name	*% Present*	*Average score*	*Class*
char	100.00	120.55	Content
word	100.00	18.69	Content
question	7.95	0.10	Content
excl	10.10	0.15	Content
uppercase	10.23	11.27	Content
pronoun	92.84	4.22	Content
smile	42.24	0.02	Content
frown	1.81	0.43	Content
url	14.17	0.42	Content
retweet	8.71	0.74	Content
sentiment_pos	71.51	1.53	Content
sentiment_neg	59.07	1.23	Content
sentiment	74.20	0.29	Content
num_hashtag	42.09	0.83	Content
num_mention	19.25	0.25	Content
tweet_type	100.00	1.10	Content
ellipsis	2.11	0.29	Content
news	5.13	2.03	Content

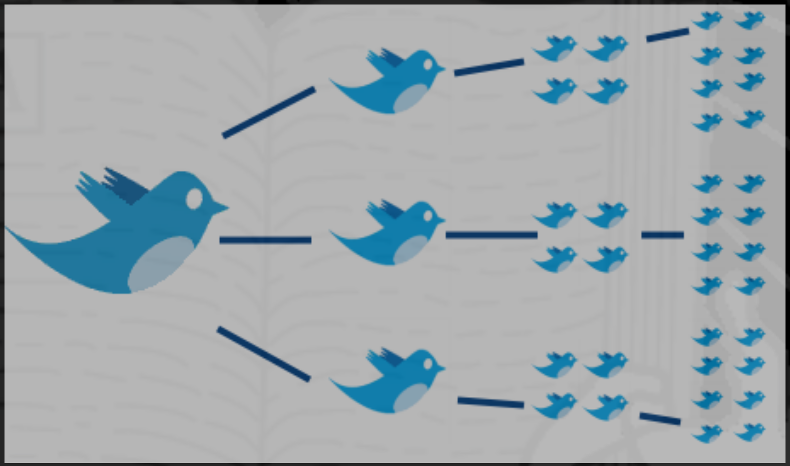
Feature Sets

Behavioral / Dynamic

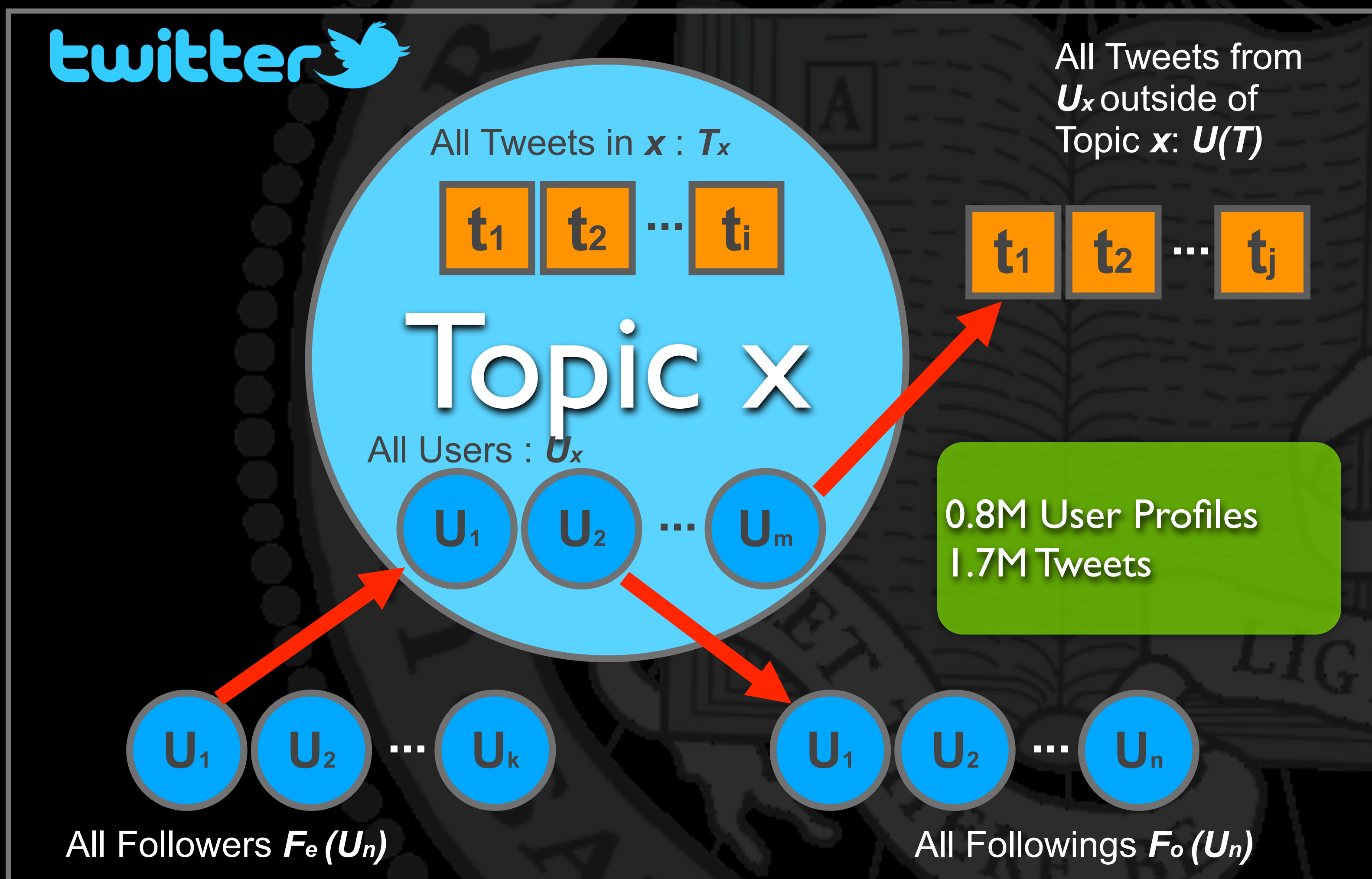
Name	*% Present*	*Average score*	*Class*
average balance of conversation	100.00	0.32	Behavioral
average number of friends in time-line	100.00	2086.28	Behavioral
average spacing between statuses in seconds in timeline	100.00	21959.07	Behavioral
average text length in timeline	100.00	104.52	Behavioral
average general response time	100.00	3.27	Behavioral
average number of messages per conversation	100.00	4.34	Behavioral
average trust value in conversation	100.00	0.10	Behavioral
fraction of statuses in timeline that are retweets	100.00	0.55	Behavioral

Outline

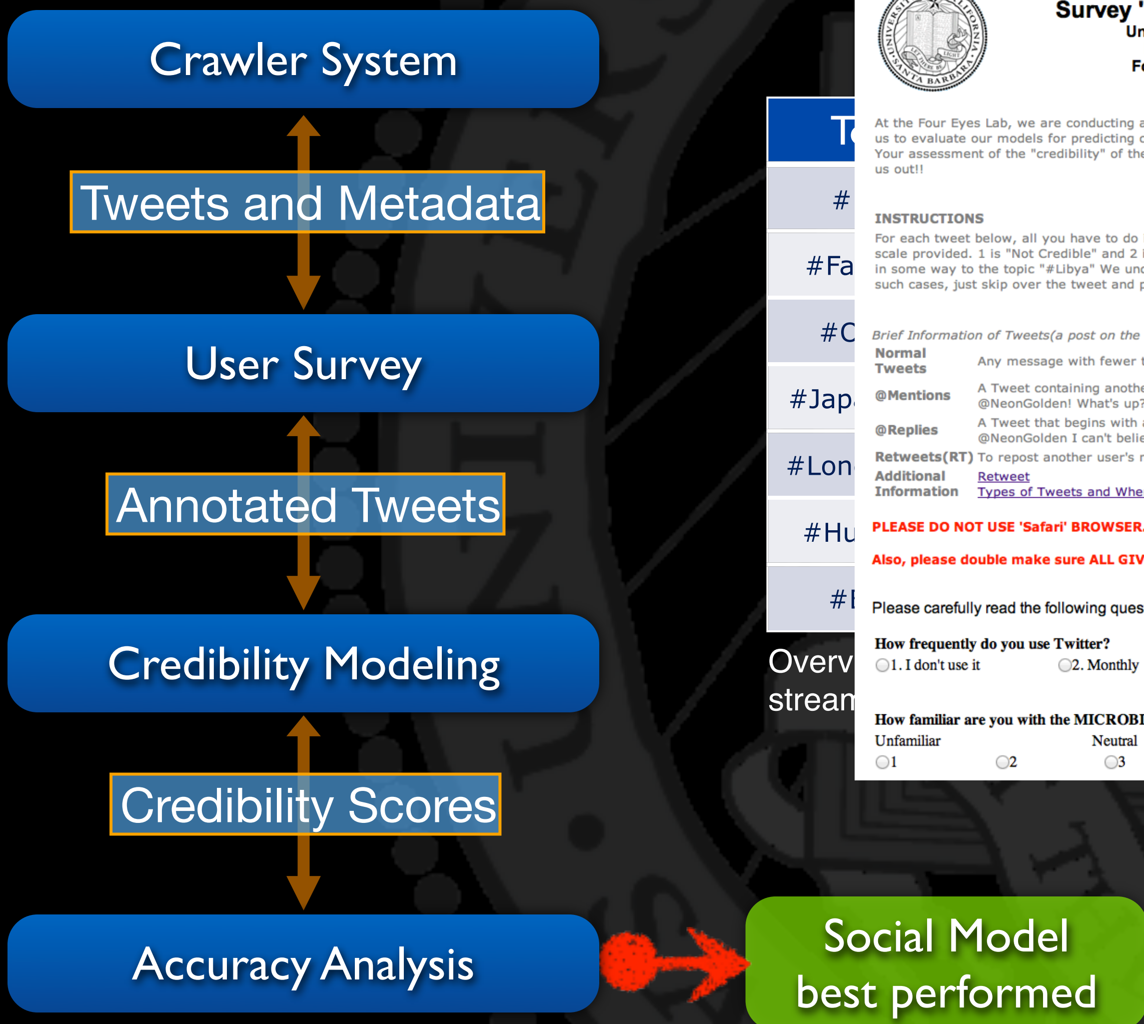
- Background
- Features and Contexts
- **Experimental Framework**
  - Crawler System
  - Data
  - Credibility Assessments
- Results
- Conclusion

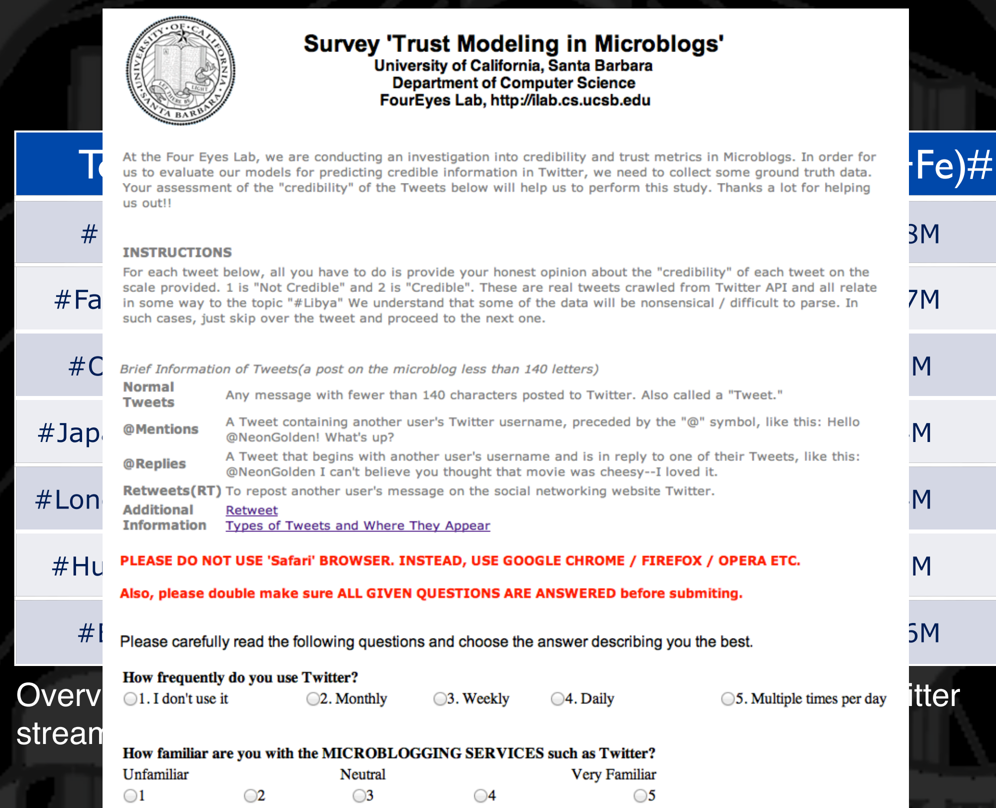


Crawling Strategy



Segmenting based on Credibility





Survey 'Trust Modeling in Microblogs'
University of California, Santa Barbara
Department of Computer Science
FourEyes Lab, http://ilab.cs.ucsb.edu

At the Four Eyes Lab, we are conducting an investigation into credibility and trust metrics in Microblogs. In order for us to evaluate our models for predicting credible information in Twitter, we need to collect some ground truth data. Your assessment of the "credibility" of the Tweets below will help us to perform this study. Thanks a lot for helping us out!!

INSTRUCTIONS

For each tweet below, all you have to do is provide your honest opinion about the "credibility" of each tweet on the scale provided. 1 is "Not Credible" and 2 is "Credible". These are real tweets crawled from Twitter API and all relate in some way to the topic "#Libya" We understand that some of the data will be nonsensical / difficult to parse. In such cases, just skip over the tweet and proceed to the next one.

Brief Information of Tweets(a post on the microblog less than 140 letters)

Normal Tweets	Any message with fewer than 140 characters posted to Twitter. Also called a "Tweet."
@Mentions	A Tweet containing another user's Twitter username, preceded by the "@" symbol, like this: Hello @NeonGolden! What's up?
@Replies	A Tweet that begins with another user's username and is in reply to one of their Tweets, like this: @NeonGolden I can't believe you thought that movie was cheesy--I loved it.
Retweets(RT)	To repost another user's message on the social networking website Twitter.
Additional Information	Retweet Types of Tweets and Where They Appear

PLEASE DO NOT USE 'Safari' BROWSER. INSTEAD, USE GOOGLE CHROME / FIREFOX / OPERA ETC.

Also, please double make sure ALL GIVEN QUESTIONS ARE ANSWERED before submiting.

Please carefully read the following questions and choose the answer describing you the best.

How frequently do you use Twitter?
○1. I don't use it ○2. Monthly ○3. Weekly ○4. Daily ○5. Multiple times per day

How familiar are you with the MICROBLOGGING SERVICES such as Twitter?
Unfamiliar — Neutral — Very Familiar
○1 ○2 ○3 ○4 ○5

Method

- Algorithm used
  - Use Weka3 toolkit
  - Train a J48(C4.5) Decision Tree Algorithm
  - 70:30 train-test ratio (both kept separate)
  - 10 Fold Cross Validation

Segmenting based on Topics

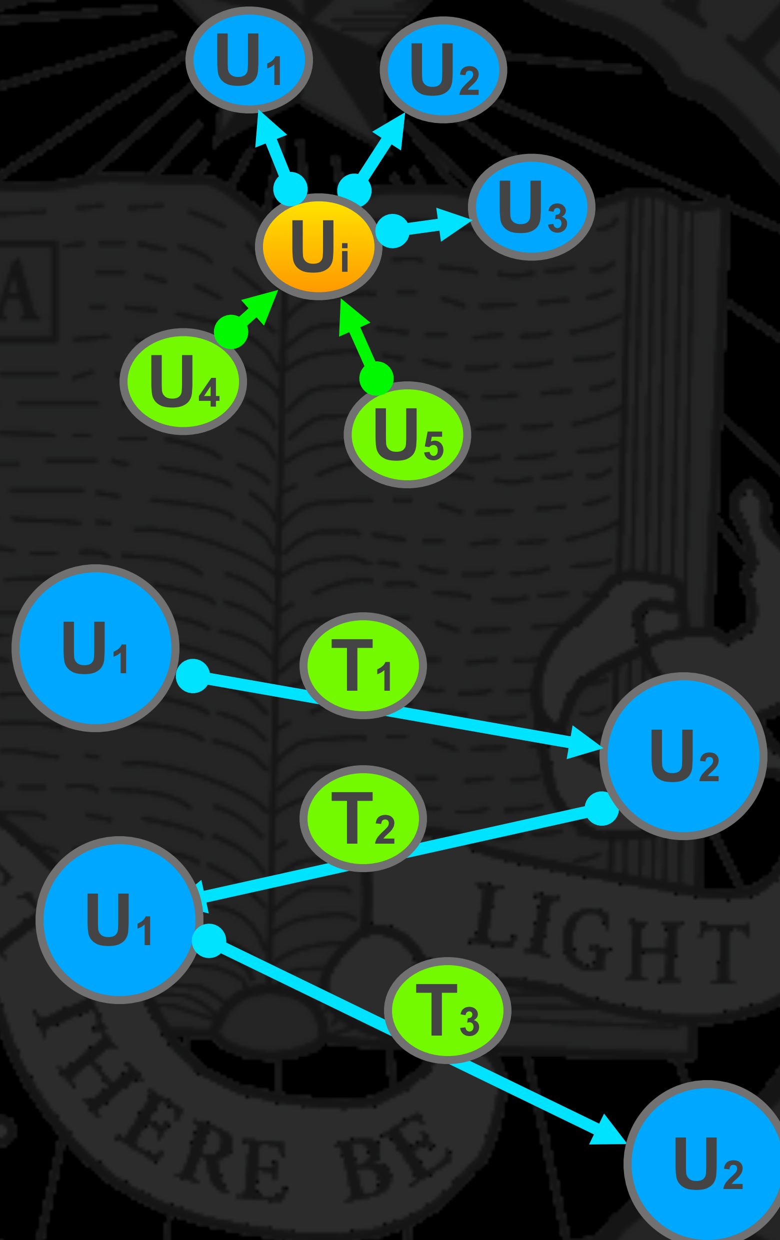
Set Name	*Core Tweeters*	*Core Tweets*	F_o *and* F_e (overlapped)	F_o *and* F_e (distinct)
Libya	37K	126K	94M	28M
Superbowl	191K	227K	N/A	N/A
Romney	226K	705K	N/A	N/A
Facebook	433K	217K	62M	37M
EnoughIsEnough	85K	129K	13M	4M
Egypt	49K	217K	73M	36M
Earthquake	67K	131K	15M	5M

TABLE II

OVERVIEW OF 7 TOPIC-SPECIFIC DATA COLLECTIONS MINED FROM THE TWITTER STREAMING API.

Segmenting based on Behavior:

- For our experiments, a "dyadic pair" is a conversation between two twitter users that contains at least three messages. Tweets from such conversations make up the "dyadic pair" data set.

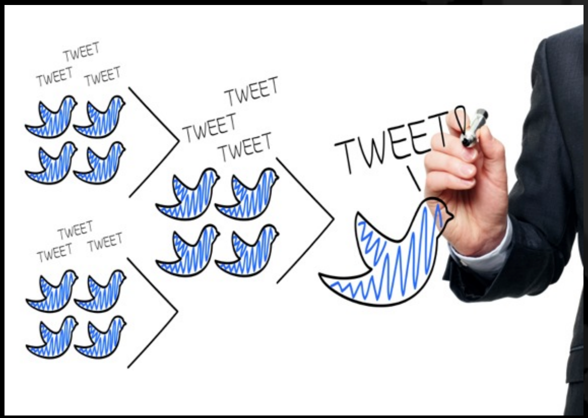


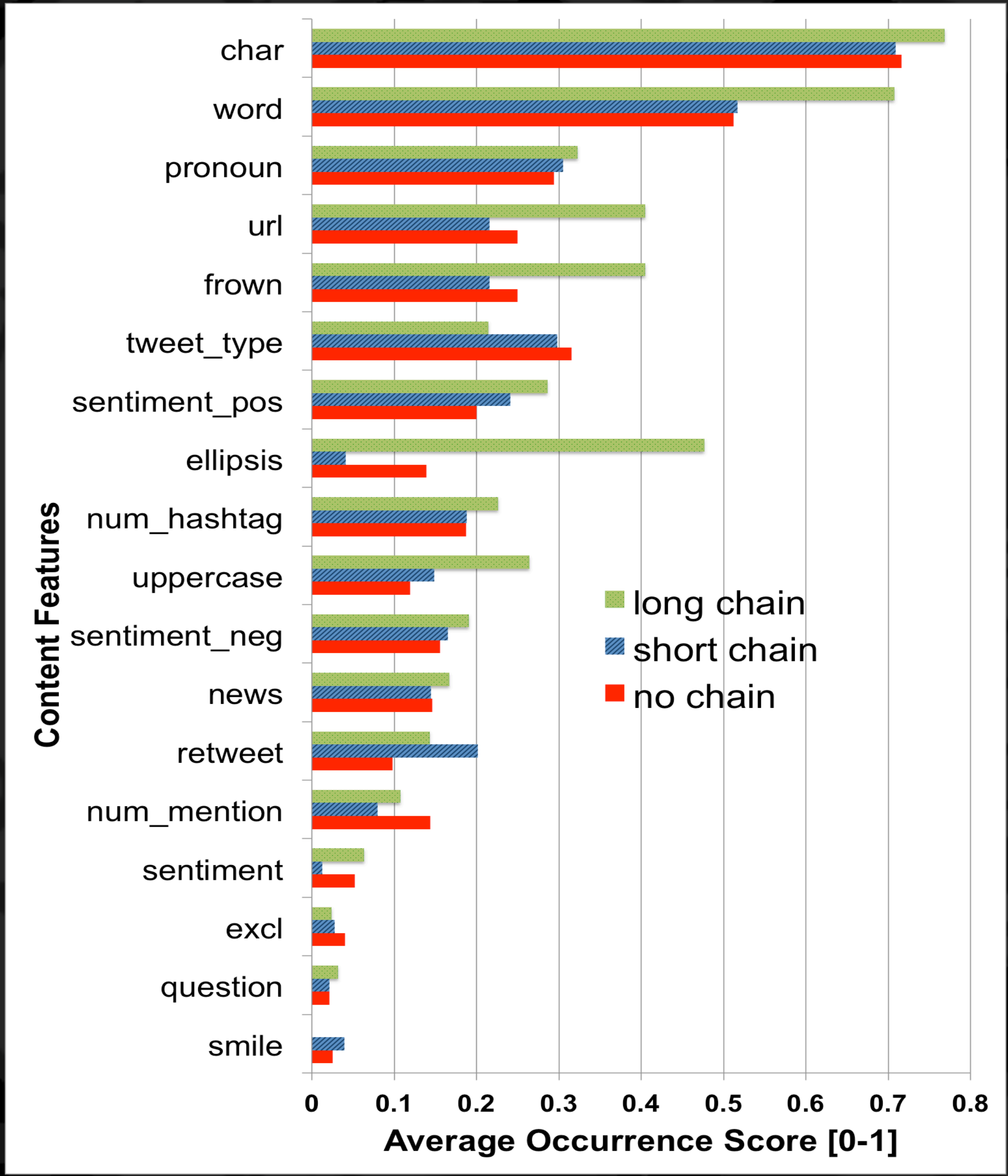
Outline

- Background
- Experimental Framework
- Credibility Models
- **Results**
  - Results
  - Credibility Predictions
  - Location and Devices
- Conclusion

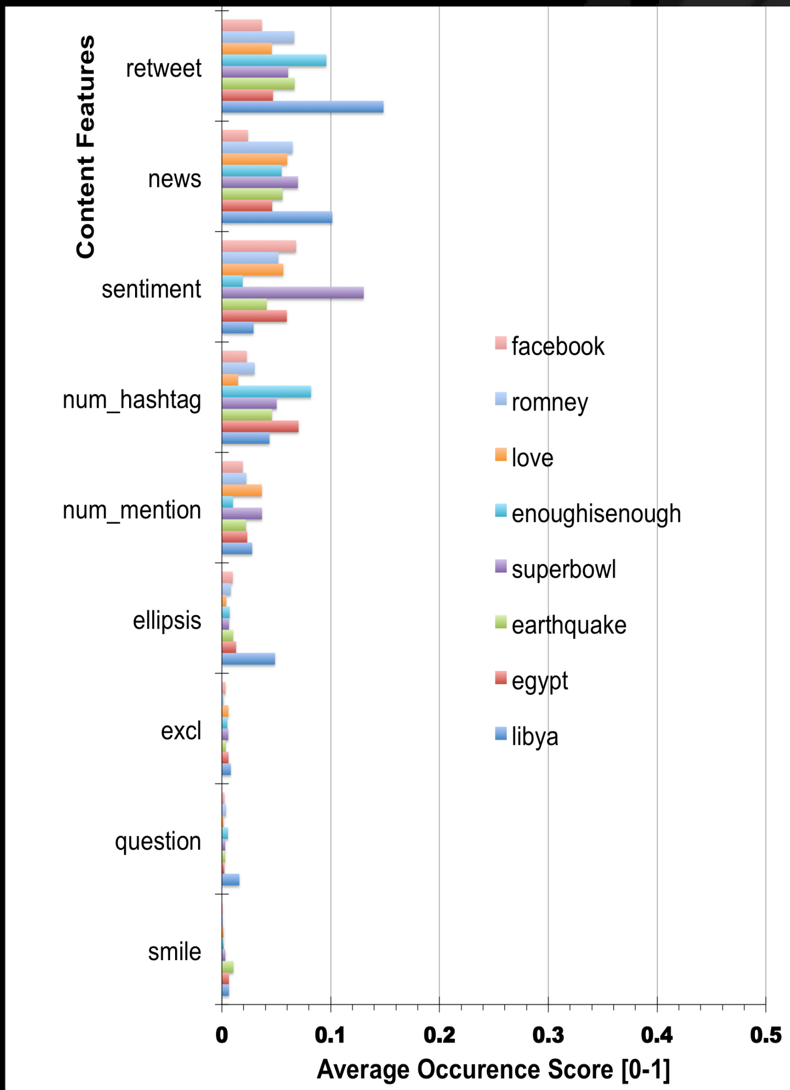
Results: Retweet Chains

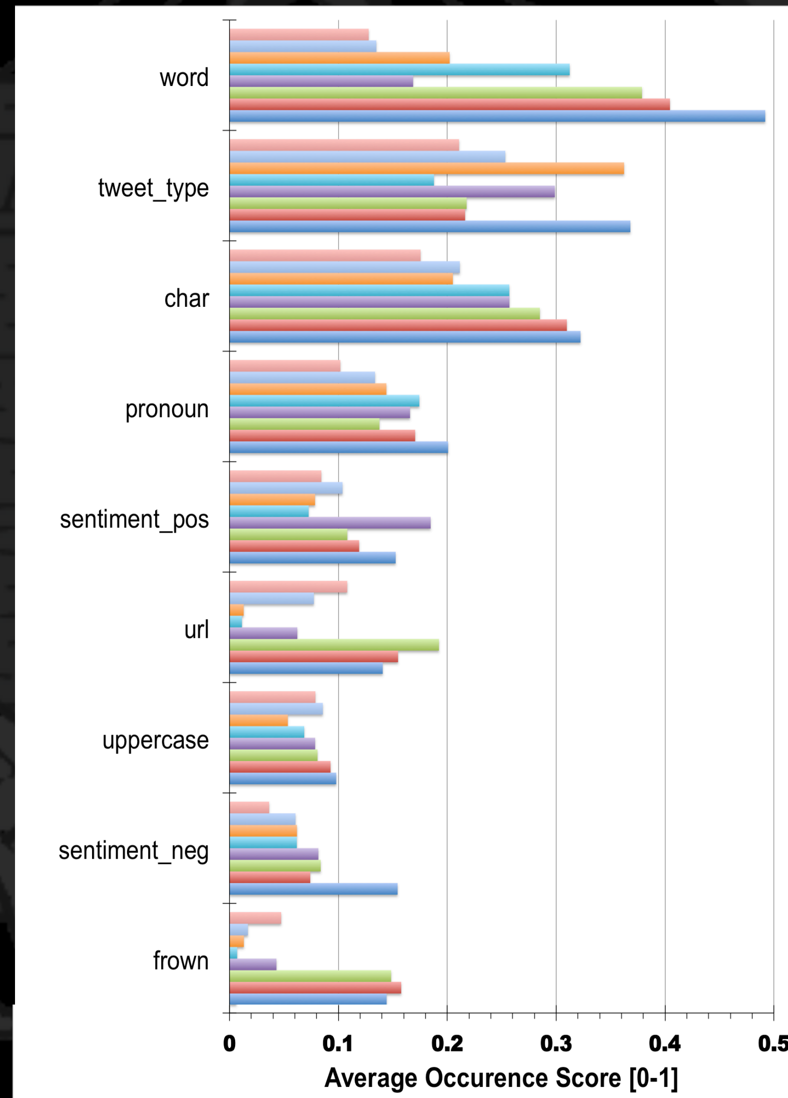
- Longer Tweets and tweets with URLs tend to be retweeted more frequently





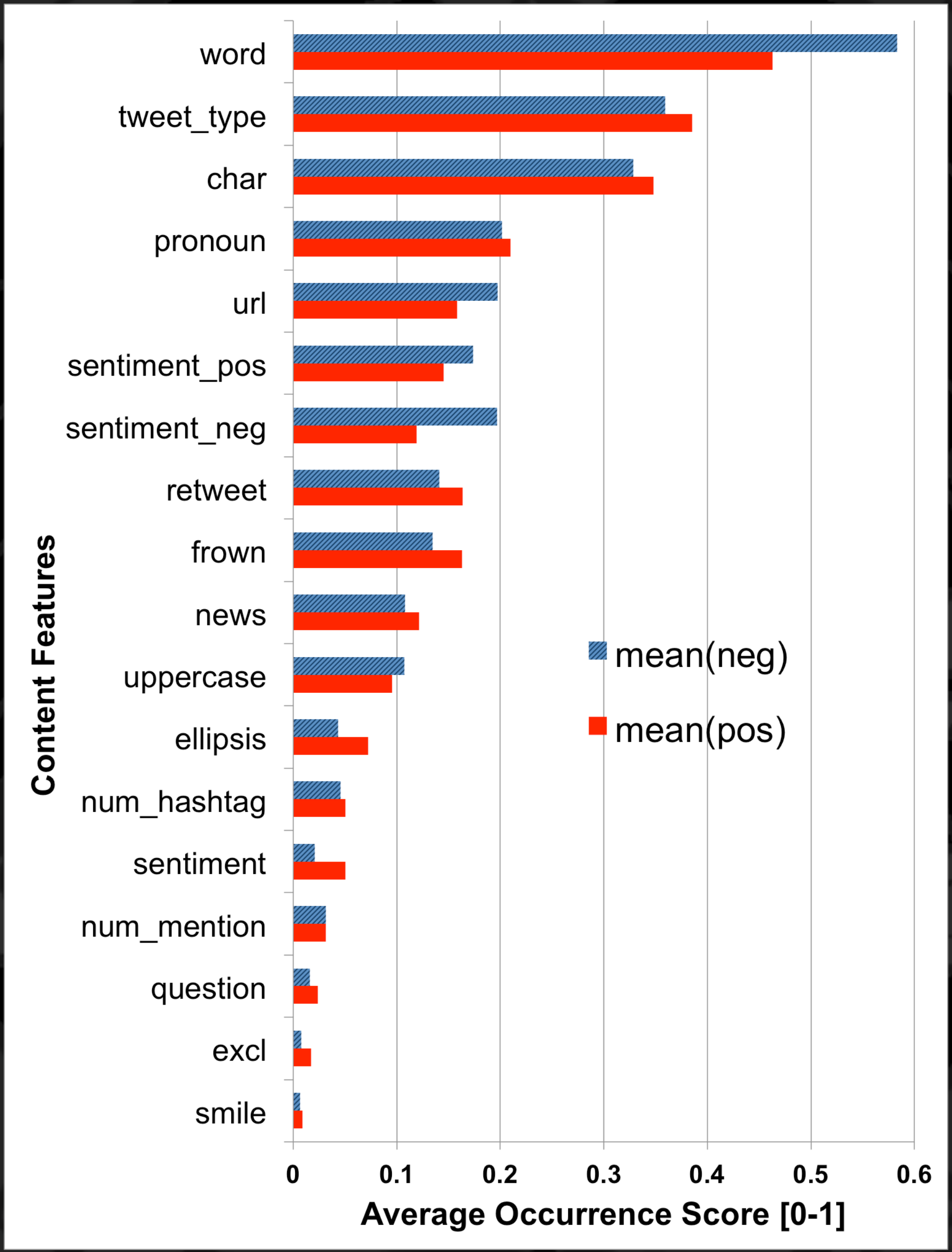
Results: Features Across Topics





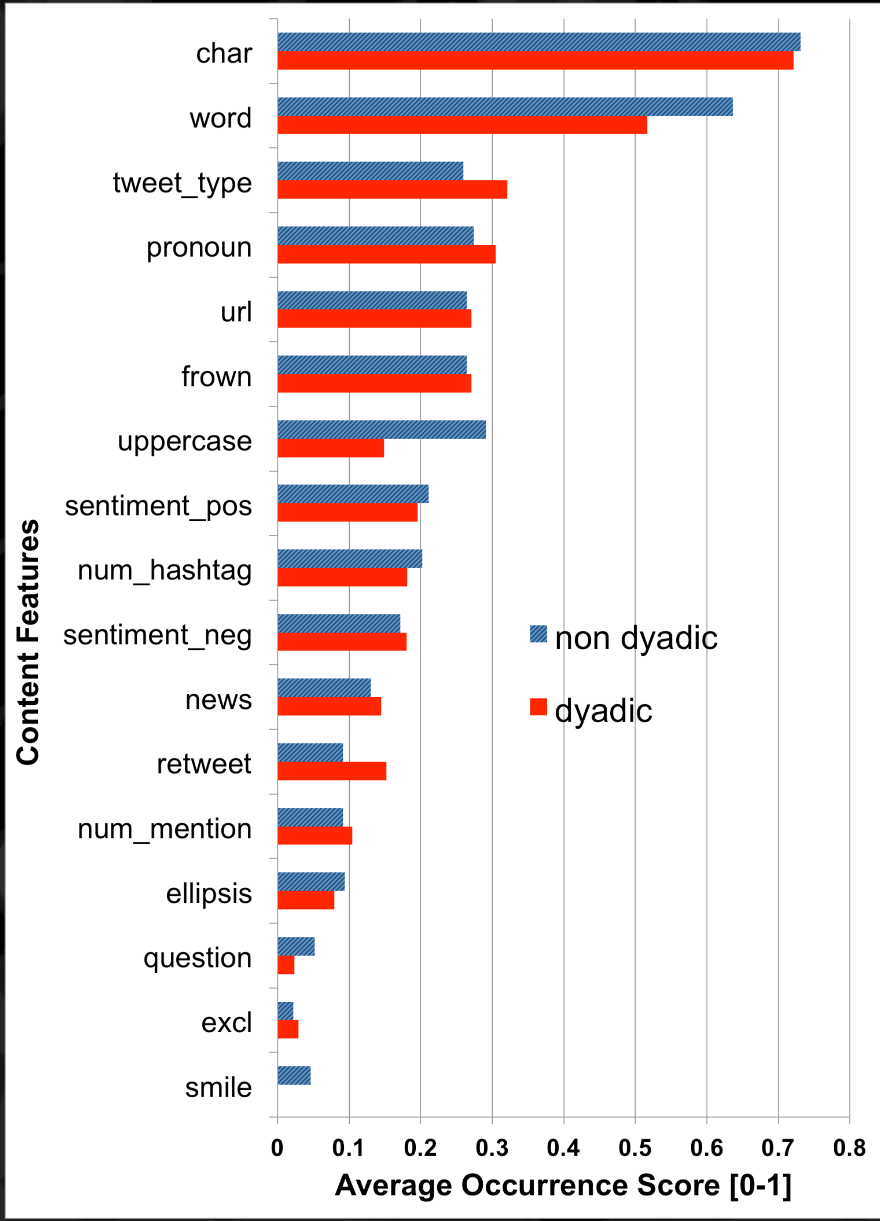
Results: Credibility Distribution

- Analyzed feature distribution across credible and non-credible sets of tweets.
- E.g. Long tweets are usually more credible
- E.g. Negative sentiment occurred more in tweets that were tagged as “not credible”.



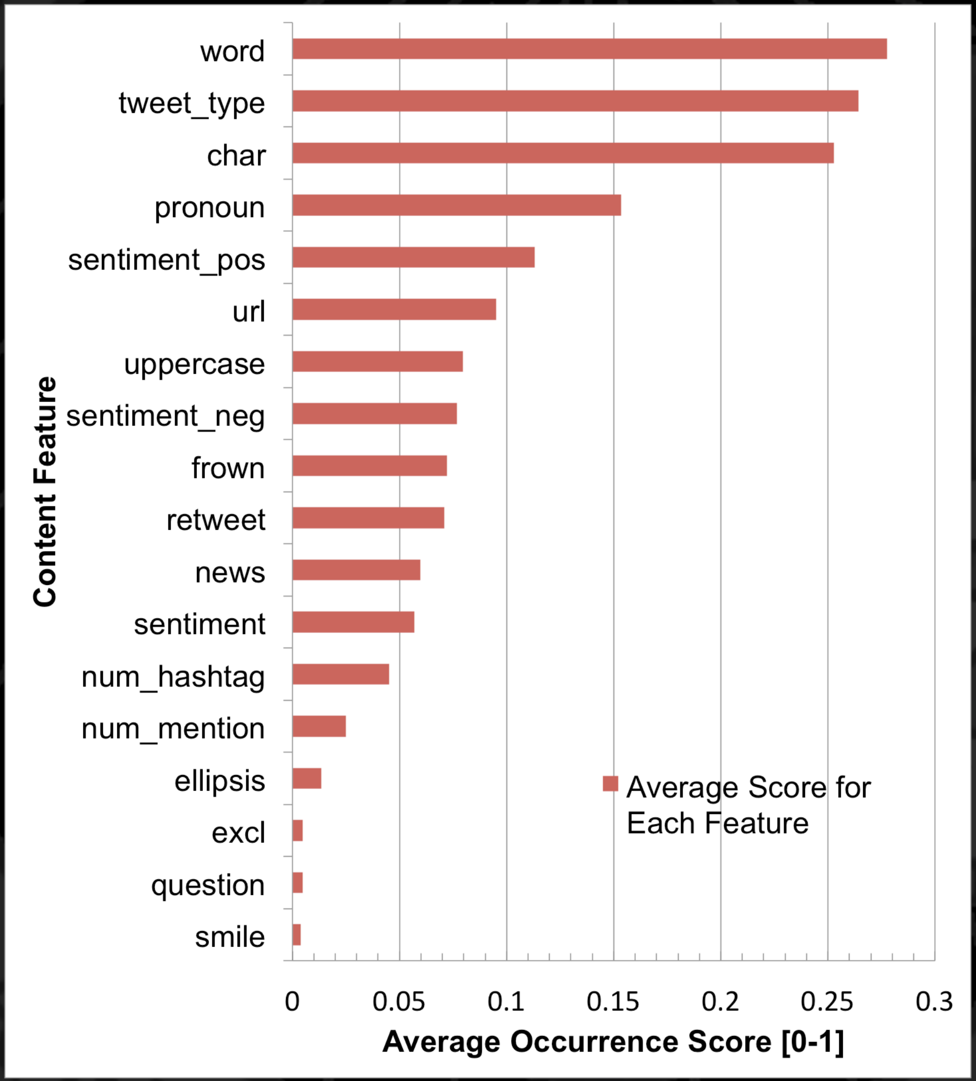
Results: Dyadic Pairs

- Analyzed sets of tweets that were part of pairwise conversations with at least three messages
- Conversational tweets tended to be shorter
- More use of uppercase terms in non-conversational tweets
- More retweet tags in conversational tweets



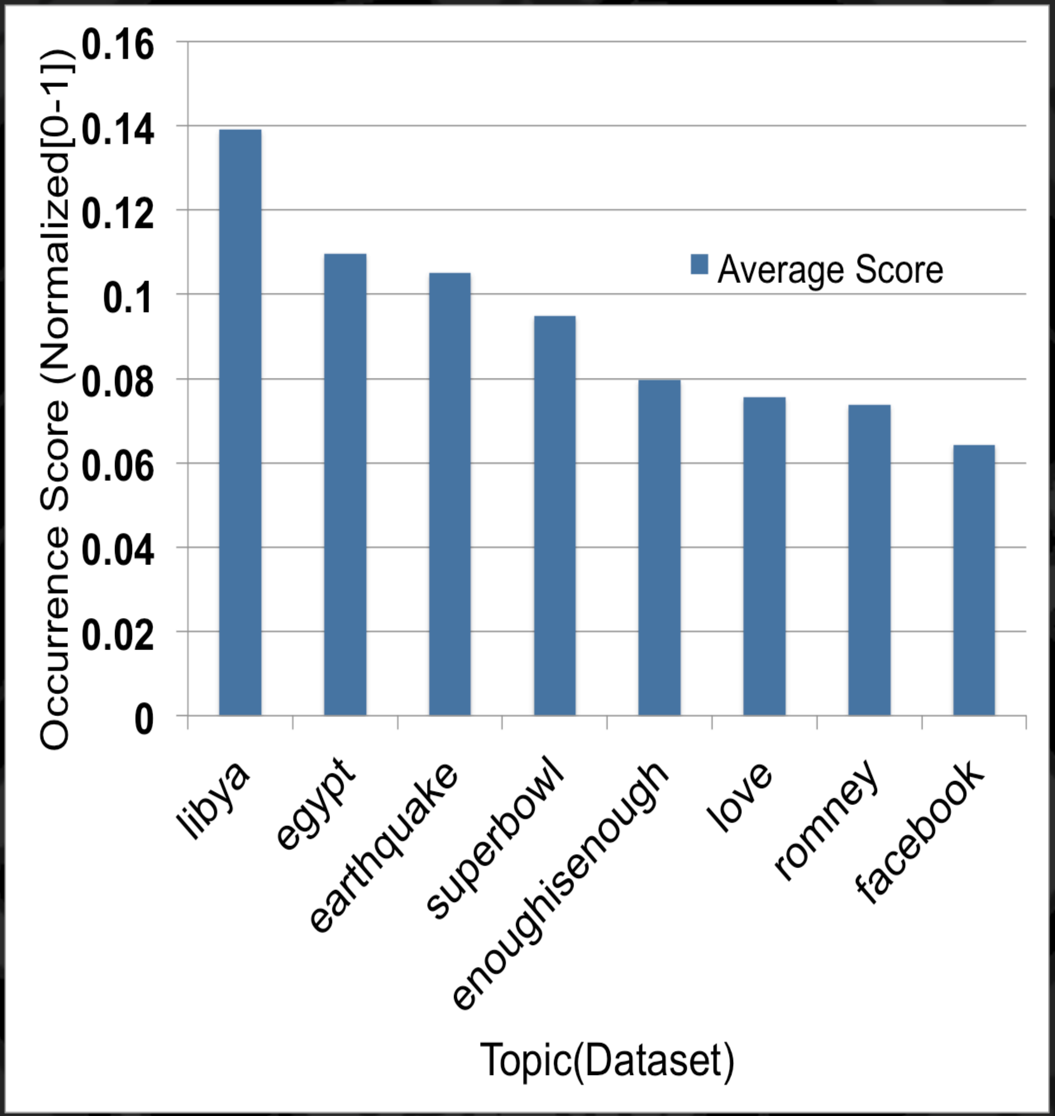
Results: Feature Utility Scores

- Computed the utility of each feature based on occurrence across all contexts in our experiments.
- Most useful features include tweet length, sentiment, url, use of uppercase.



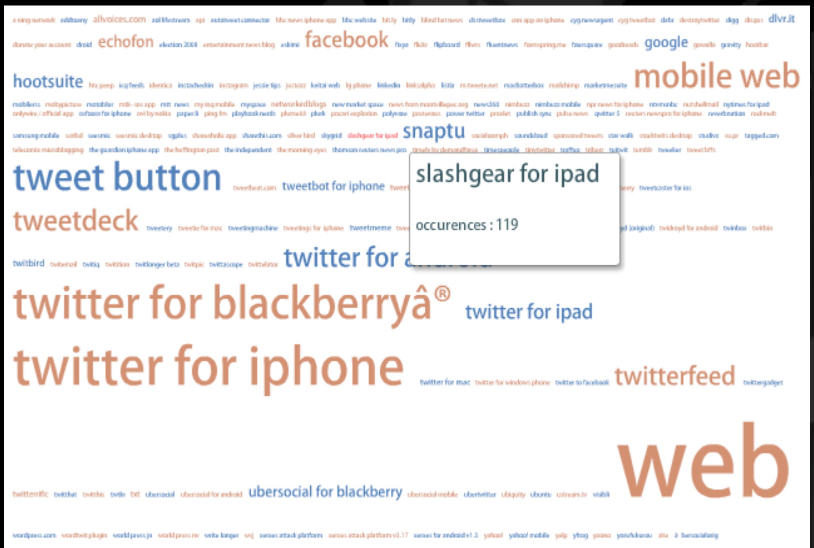
Results: Per-Topic Features

- Analyzed how often our credibility indicators occurred in each of our topic-based slices.
- Credibility indicating features tended to be used more in emergency and unrest situations.
- Interestingly, less credibility-indicating features in the political data set "#Romney".

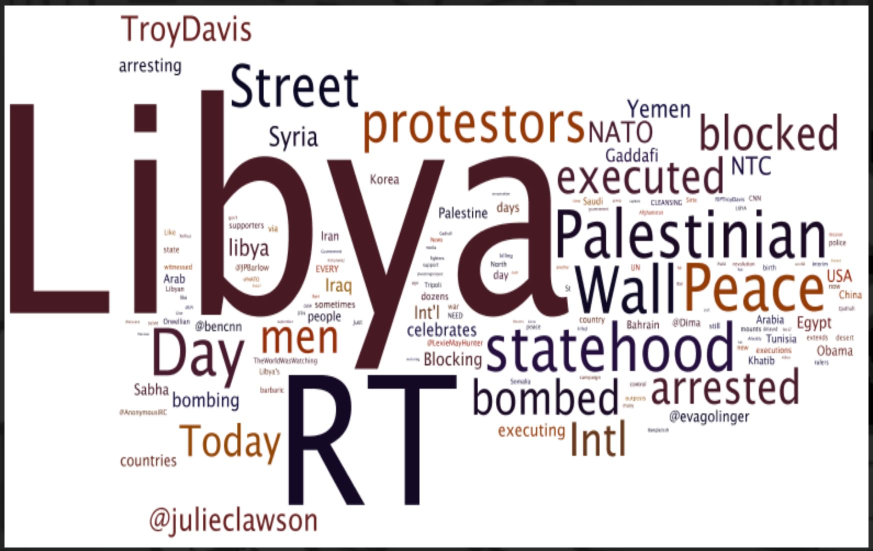


Location and Devices

- Analysis on the Crawled Data Set shows the Distribution of **Frequent** *Information Sources* and *Topics*.



Word cloud showing origin of tweets in the Libya data set



Word cloud showing distribution of popular terms in the Libya data set.

Outline

- Background
- Features and Contexts
- Experimental Framework
- Results
- **Conclusion**
  - Research Question (revisited)
  - Conclusion
  - Future Work

Future Work

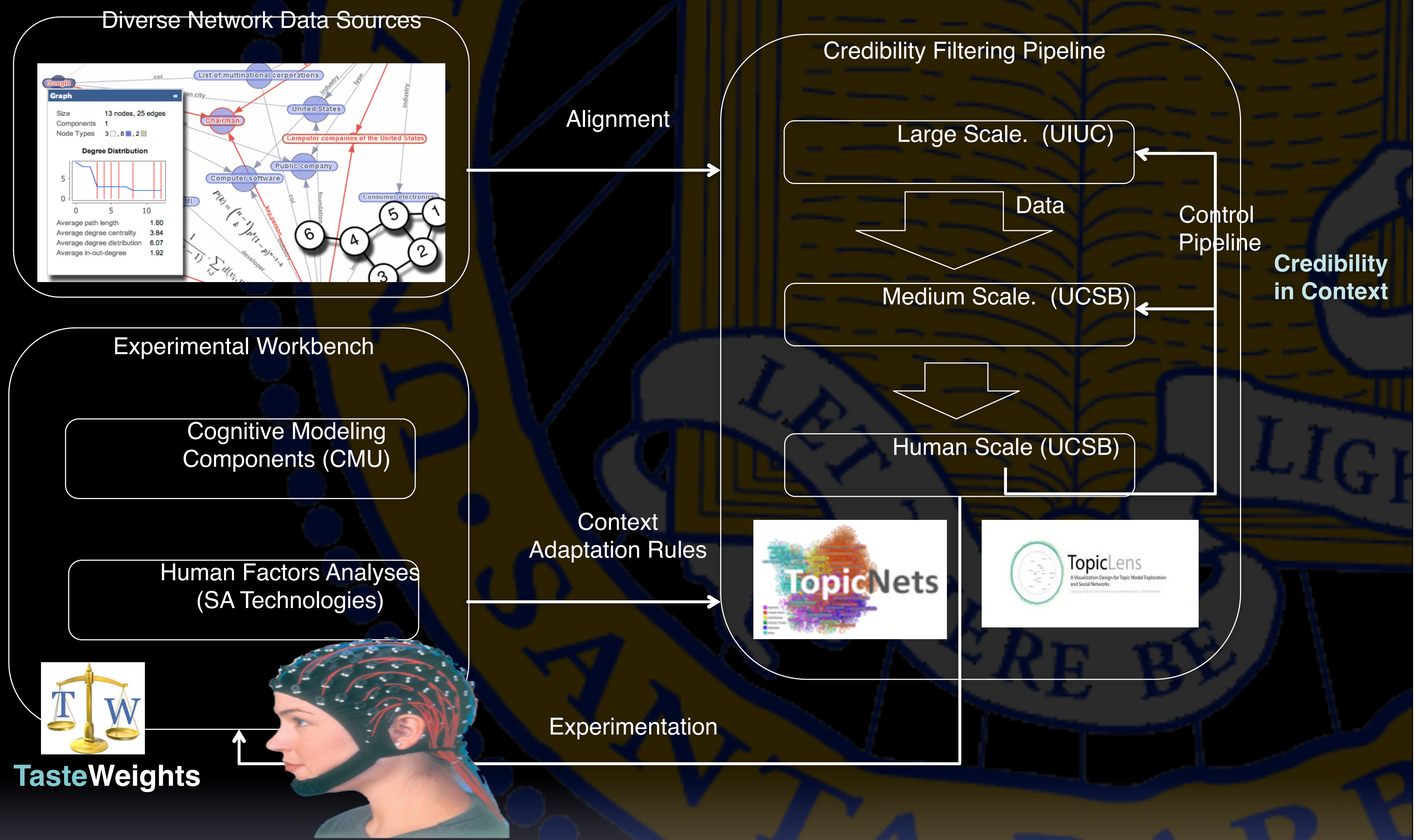
- Integration of distribution knowledge into credibility-based filtering algorithms.
- Analysis of behavioral patterns for groups of features (a correlation-based analysis).
- Cognitive modeling of users while interacting with data from different filters.

Conclusion

- How are the features that indicate credibility distributed in Twitter?
- Feature distribution changes substantially across different slices of the network. (Dyadic, Topic-based, Chain-based segmentations)
- How/Why do they vary across different contexts?
- Many influencing factors. For example, strong indicators tend to occur more frequently in conversational tweets, and in topics about emergency or social unrest situations

Thank you!

Overview of Experimental Framework



Social Impact

- As of February 2010