

TargetVue

Analysis of Online Anomalous User

Hung-li Chen (Henry)

Target Vue: *Visual* Analysis of *Anomalous* *User Behaviours* in *Online Communication* Systems (TVCG, 2016)

Nan Cao, Conglei Shi, Sabrina Lin,
Jie Lu, Yu-Ru Lin, Ching-Yung Lin

First five authors are from IBM T.J. Watson Research Centre
The last author is from University of Pittsburgh not Pissburg...

Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue
- Evaluation and Comments

Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue
- Evaluation and Comments

Context

- Anomaly Detection is important.
- Challenging to find completely automated solutions

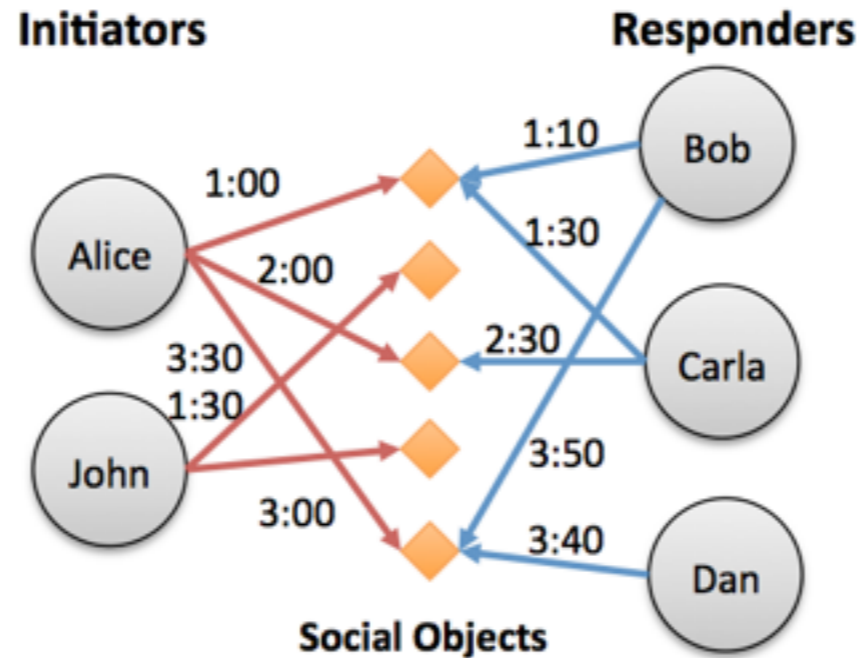
Contribution

- TargetVue: a system that **detects** and supports **interactive exploration** of **anomalous users**
- New **glyph design** and the grid layout
- Evaluation through a bot detection challenge and case a case study

Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue
- Evaluation and Comments

Data Model



- Initiator - Social Objects - Responders
- High-level features: Behaviour, Content, Interaction, Temporal, Network, User Profile
- Data: **time series** of **feature vectors** (for each user)

Requirements

1. Feature Selection
2. Anomaly Detection in Context
3. Ranking Threats
4. Learn from User Feedback

Tasks

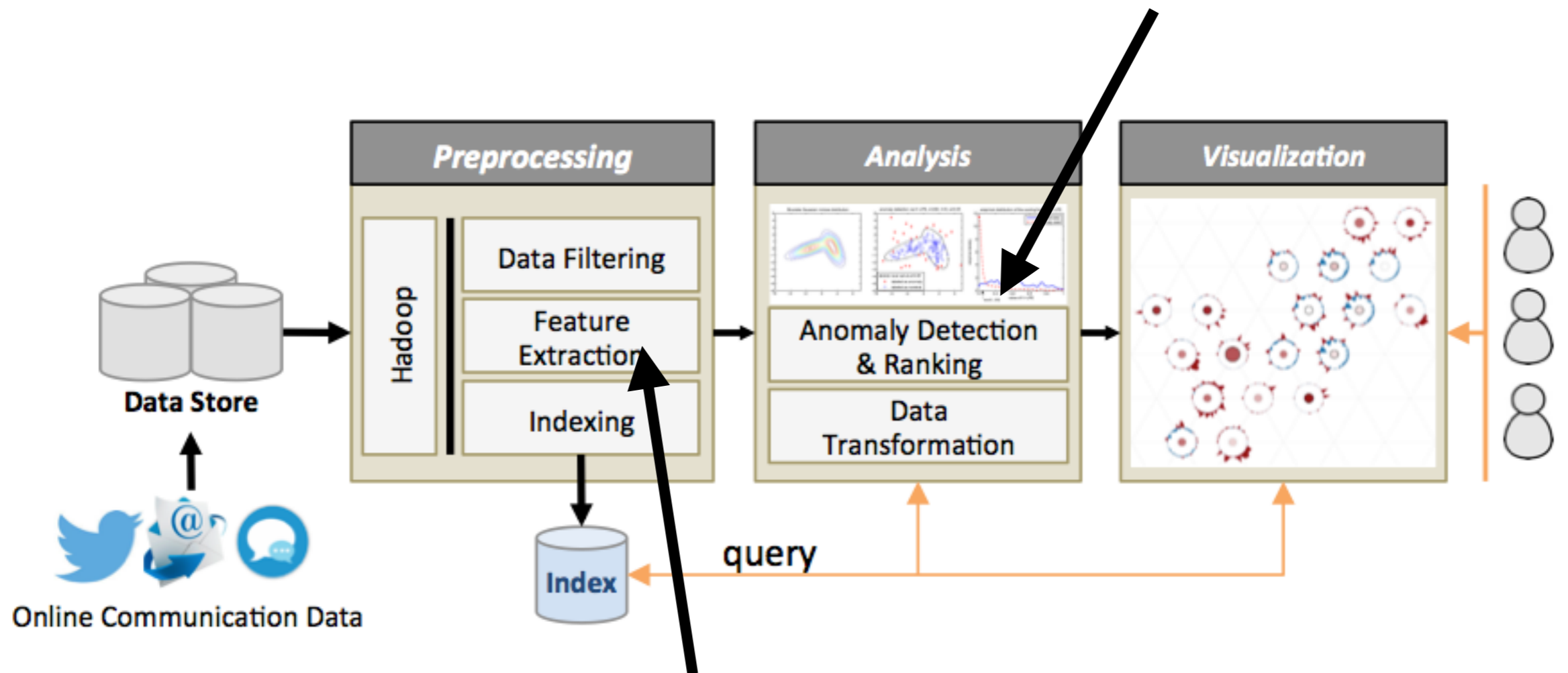
- Showing the data **overview** and detection results
- Interpreting user behaviours from **different perspectives**
- Facilitate visual data **comparisons**
- Revealing users' impacts in social communication
- Easy **browsing** of raw data
- Flexible data labeling

Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue
- Evaluation and Comments

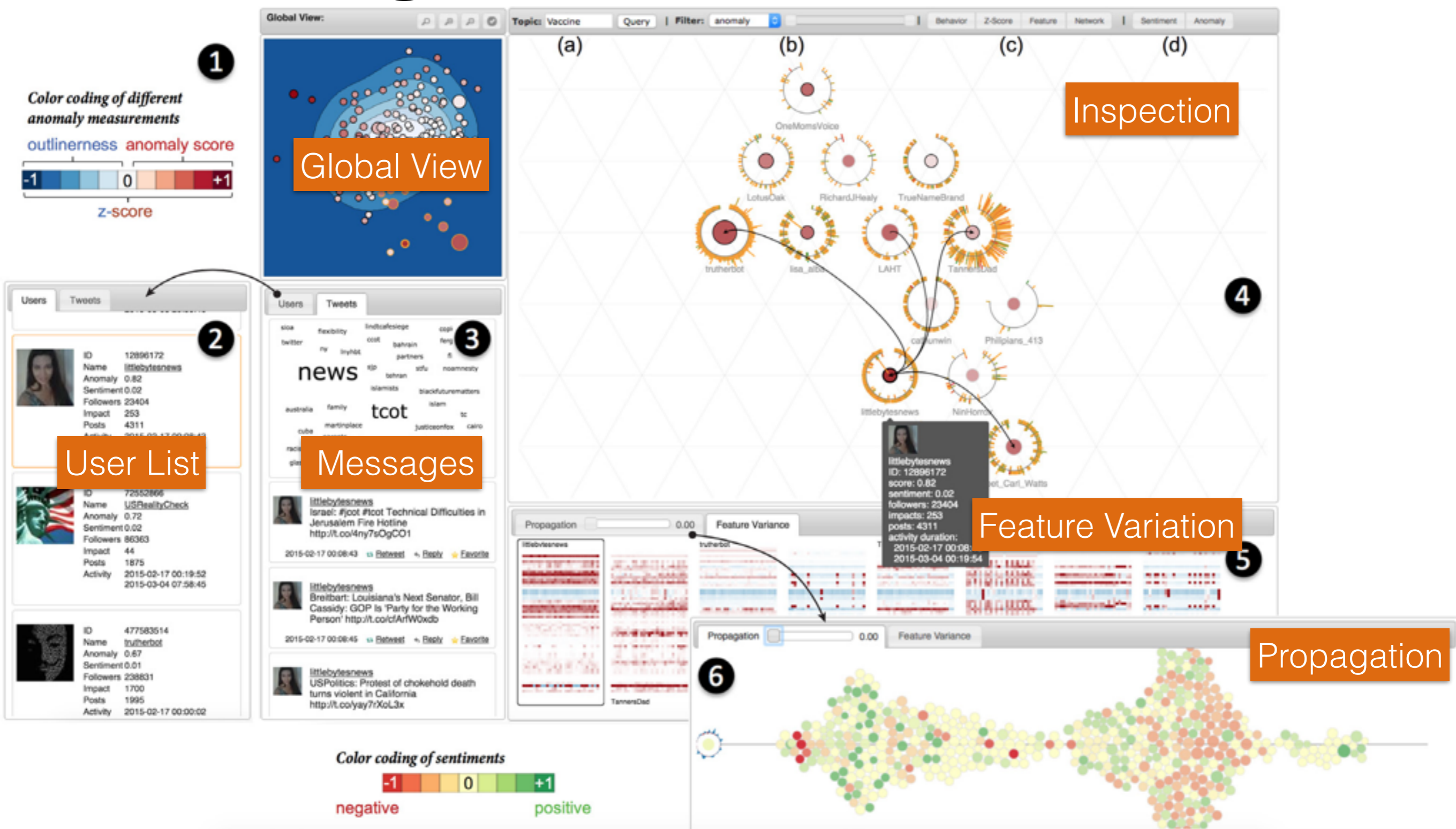
TargetVue: System Design

Time Adaptive Local Outlier Factor



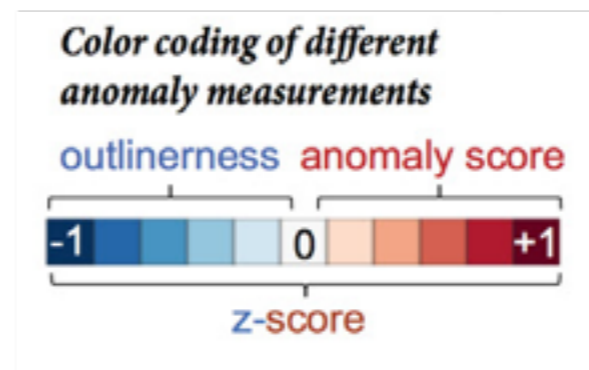
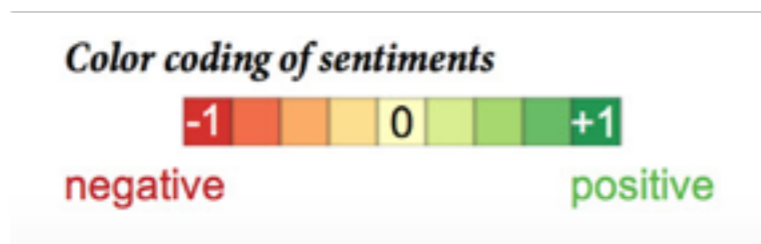
Choosing Features

TargetVue: Interface



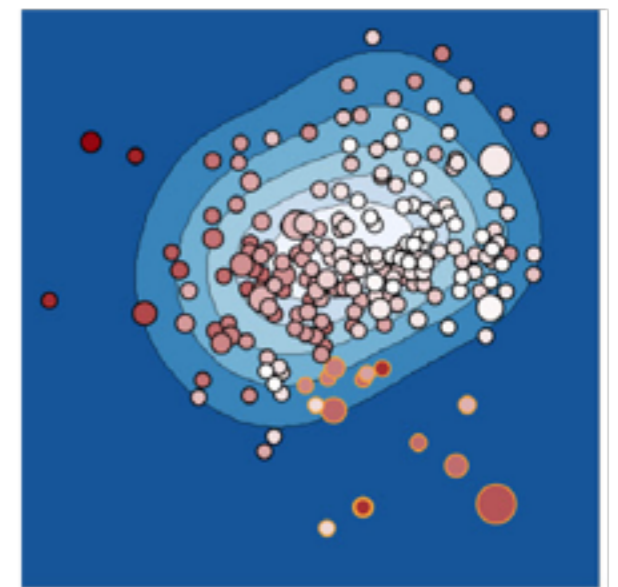
Global Encodings

- **Users** as circular nodes
- **Importance** as the size of the nodes
- **Sentiments** or **anomaly scores** as color



Global View

- Data: dimensionality reduced mean feature vector, kernel density estimation
- Encoding: location, contour map (white to blue)
- Task: overview
- Outliers are in the low density areas



User list and Messages

- Data: user profile information, raw messages
- Encoding: high frequency tag cloud, list of messages and user profiles
- Task: browsing raw data and overview

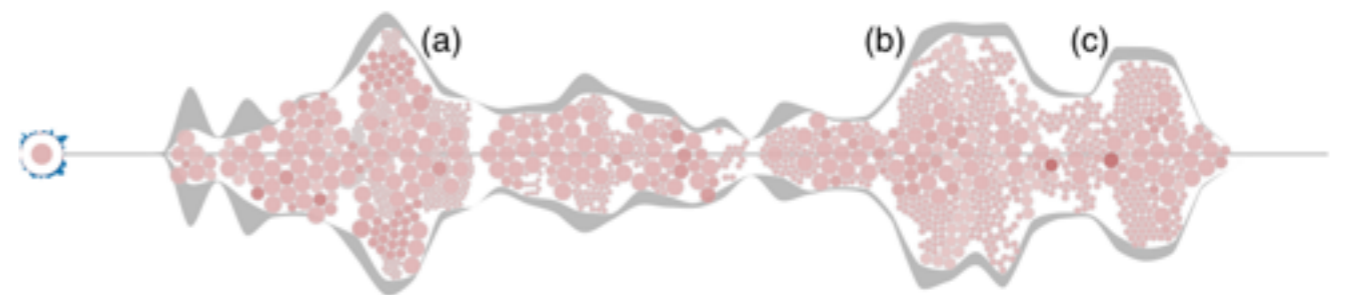
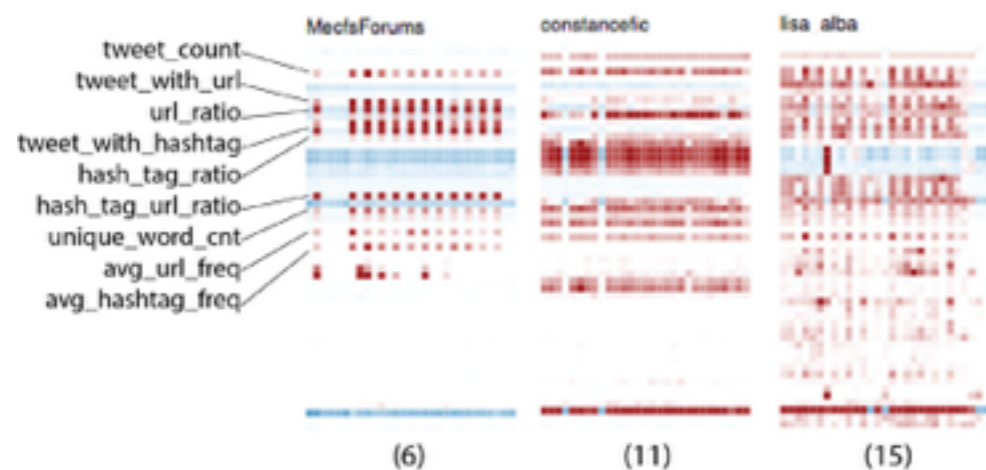
The image shows a screenshot of a Twitter interface. On the left, there are three user profiles listed under the 'Users' tab. The first profile is highlighted with a red box and a '2' in a black circle. The second profile has a picture of the Statue of Liberty. The third profile has a picture of a person's face. On the right, there is a 'Tweets' tab with a tag cloud and three tweets. The tag cloud includes words like 'news', 'tcot', 'israel', 'jerusalem', 'fire', 'hotline', 'breitbart', 'louisiana', 'senator', 'bill cassidy', 'gop', 'party', 'working person', 'protest', 'chokehold', 'death', 'turns violent', 'california'. A '3' in a black circle is next to the tag cloud. The tweets are from the user 'littlebytesnews' and include links to external content.

User Profile	ID	Name	Anomaly	Sentiment	Followers	Impact	Posts	Activity
2	12896172	littlebytesnews	0.82	0.02	23404	253	4311	2015-02-17 00:08:43 2015-03-04 00:19:54
	72552866	USRealityCheck	0.72	0.02	86363	44	1875	2015-02-17 00:19:52 2015-03-04 07:58:45
	477583514	truthbot	0.67	0.01	238831	1700	1995	2015-02-17 00:00:02

Tweet	User	Text	Activity
3	littlebytesnews	Israel: #jcot #tcot Technical Difficulties in Jerusalem Fire Hotline http://t.co/4ny7sOgCO1	2015-02-17 00:08:43 Retweet Reply Favorite
	littlebytesnews	Breitbart: Louisiana's Next Senator, Bill Cassidy: GOP Is 'Party for the Working Person' http://t.co/cfArfW0xdb	2015-02-17 00:08:45 Retweet Reply Favorite
	littlebytesnews	USPolitics: Protest of chokehold death turns violent in California http://t.co/yay7rXoL3x	

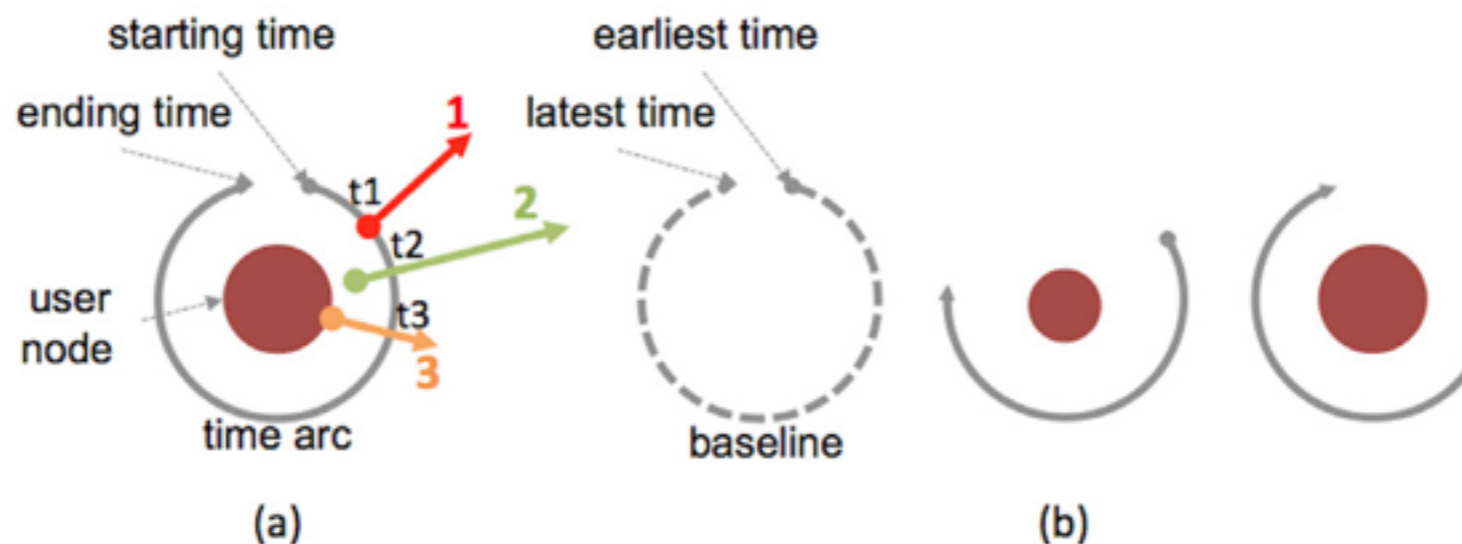
Feature Variation and Propagation View

- Data: derived feature z-score (difference from baseline), users in communication threads
- Encoding: temporal heatmap, propagation view (introduced in FluxFlow)
- High impact users have many other users in the thread



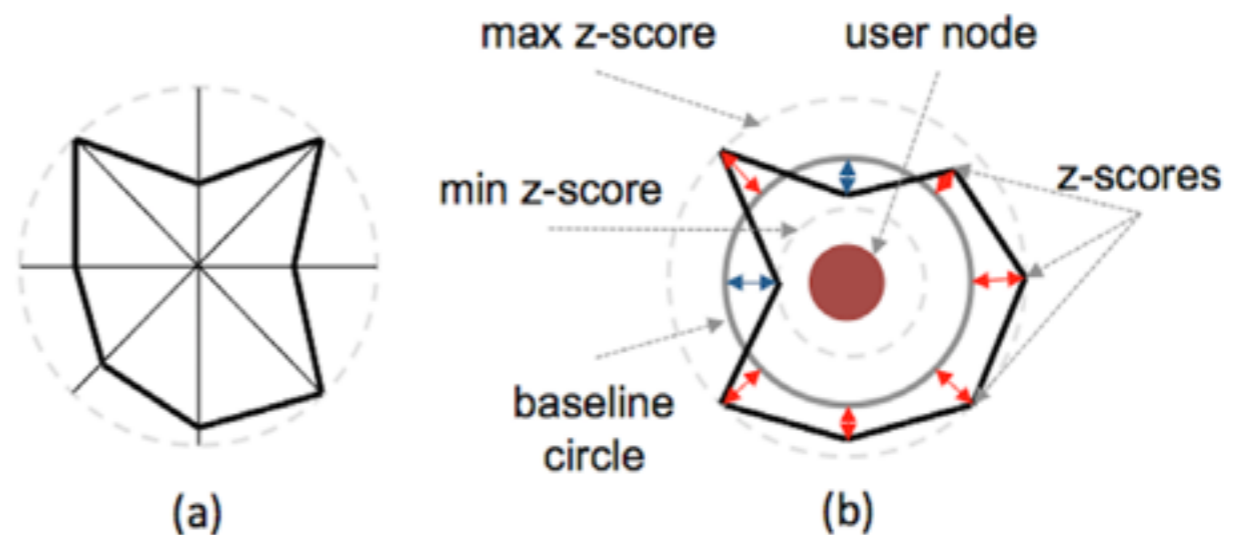
Inspection: Behaviour Glyph

- Data: posting and responding activity timeline, duration, number of users involved, sentiment of the threads
- Encoding: circular timeline, line mark (see below)
- line mark: thickness (number of users), length (duration), color (sentiment), intersection (time when the user join).



Inspection: Z-glyph

- Data: derived z-score of different features, (based on mean and standard deviation of features)
- Encoding: baseline circle, color coded area mark



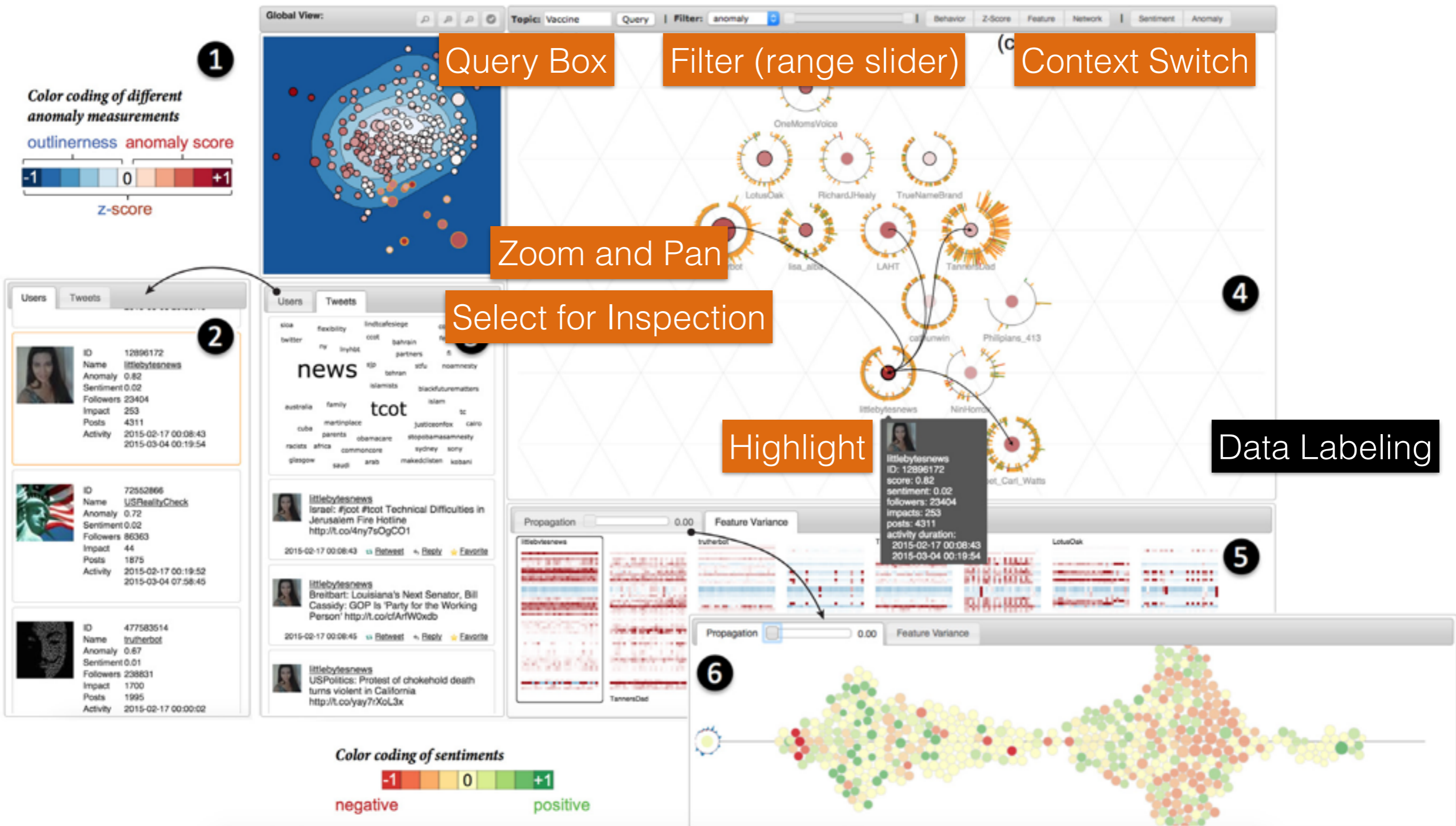
Inspection: Relation glyph

- Data: interaction relationships between users
- Encoding: directed links

Inspection: Layout

- Triangle mesh for placement
- Fast linear layout
- Preserve topology
- **Maximize average similarity** (clustering) between neighbouring glyphs

Interaction



Agenda

- Context and Contribution
- Requirements, Data and Tasks
- Design of TargetVue
- Evaluation and Comments

Evaluation

- The investigators used the system in social bot detection challenge.
 - Use global view to pick out outliers and anomalies
 - Inspect the users, and study their behaviour
 - Inspect specific features of users
 - Tune the model
- Example usage on Email data.
- Domain expert interview (2 experts): “Comprehensive”, “very powerful”

Comments

- Delivers what are promised (Explicit reference to the requirements and tasks).
- Glyph design is information dense, effective for identifying anomalies, encoding may not be the most visually effective.
- Scaling limit is unclear (mentioned that the pipeline is built on hadoop, used the system for twitter data of 8000 users and 4M tweets)
- Evaluation in the future would be helpful.

Questions?