Department of Computer Science
University of British Columbia

MultiConVis: A Visual Text Analytics System for Exploring a Collection of Online Conversations

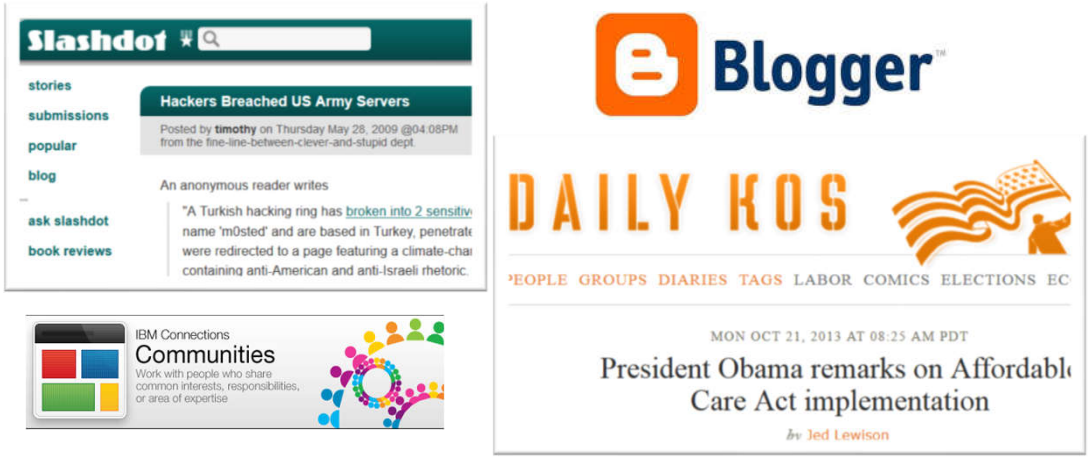
Enamul Hoque, Giuseppe Carenini

{enamul, carenini}@cs.ubc.ca

NLP group @ UBC

Rise of Text Conversations

- People engage in asynchronous conversations frequently
- e.g., blogs, forums.



Blogs:

- More than 100 millions of blogs
- The audience is rising exponentially
- Many different categories: politics, technology, business, sports,…

Problem Scenario

- Lot of articles and comments were posted on Macrumors.

- John is interested about buying iPhone6.
- He decides to explore blogs about this issue to **verify** whether the bending issue is serious.

Problem Scenario

Existing Interfaces

•Lack of high-level abstraction

- Only show conversations/comments as paginated lists ordered by recency

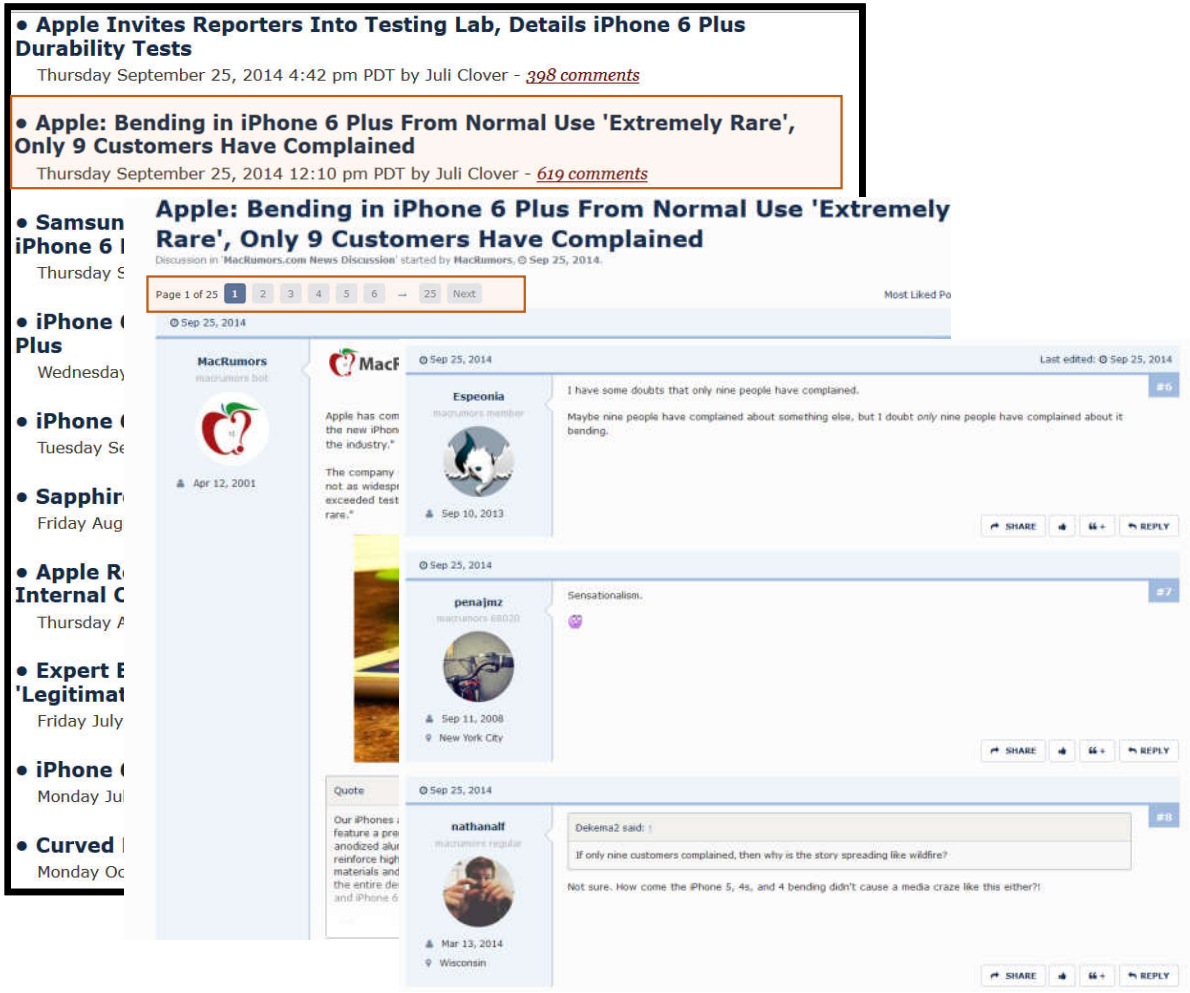
•Too many conversations

•Too many comments

=> **Information Overload**

Users

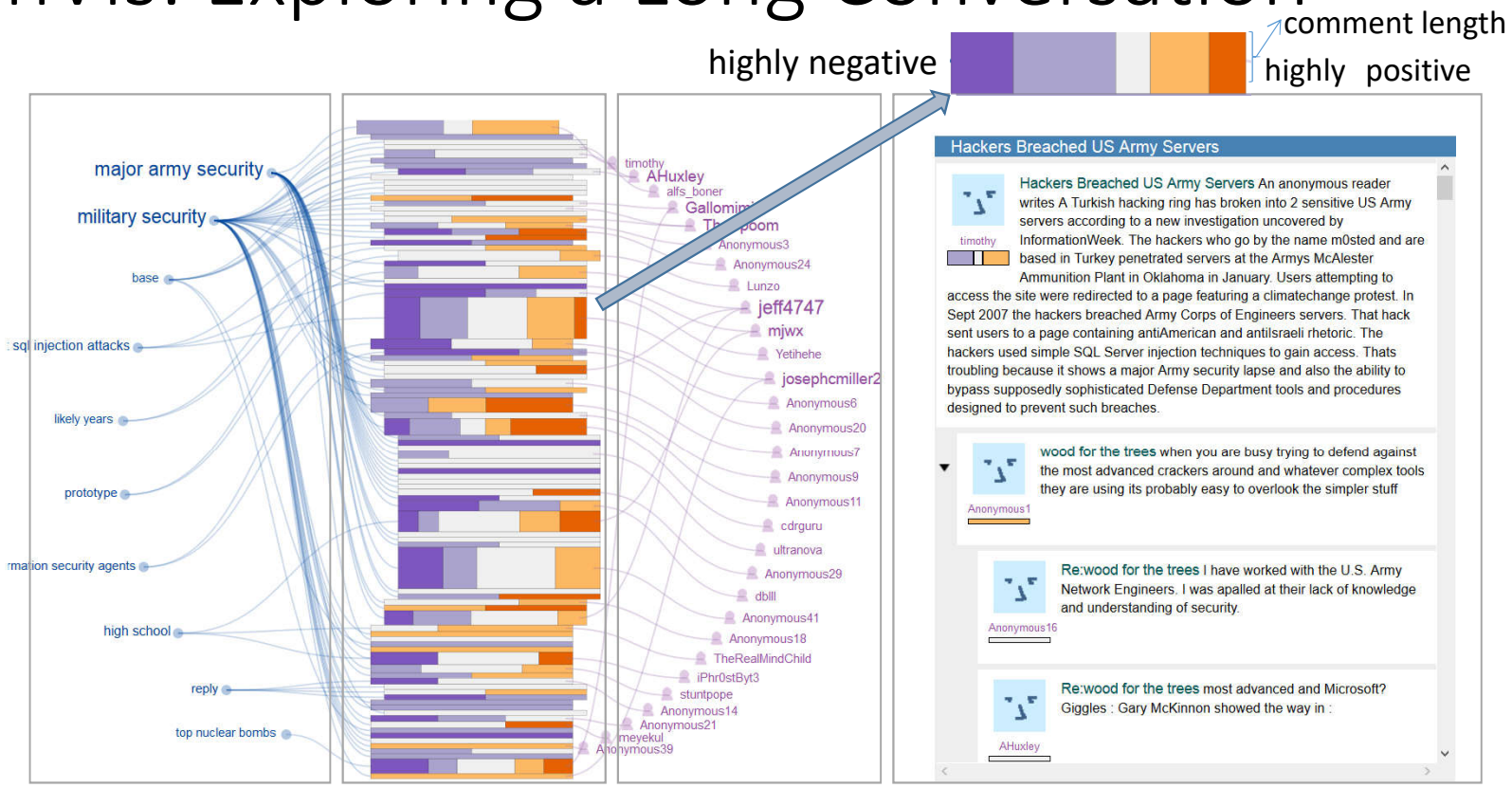
- Focus on most recent conversations/comments
- Generate short responses
- Leave conversations prematurely



Our Solution

*tightly integrate **text analysis** and **interactive visualization** to support users in exploring **collection of online conversations**.*

ConVis: Exploring a Long Conversation

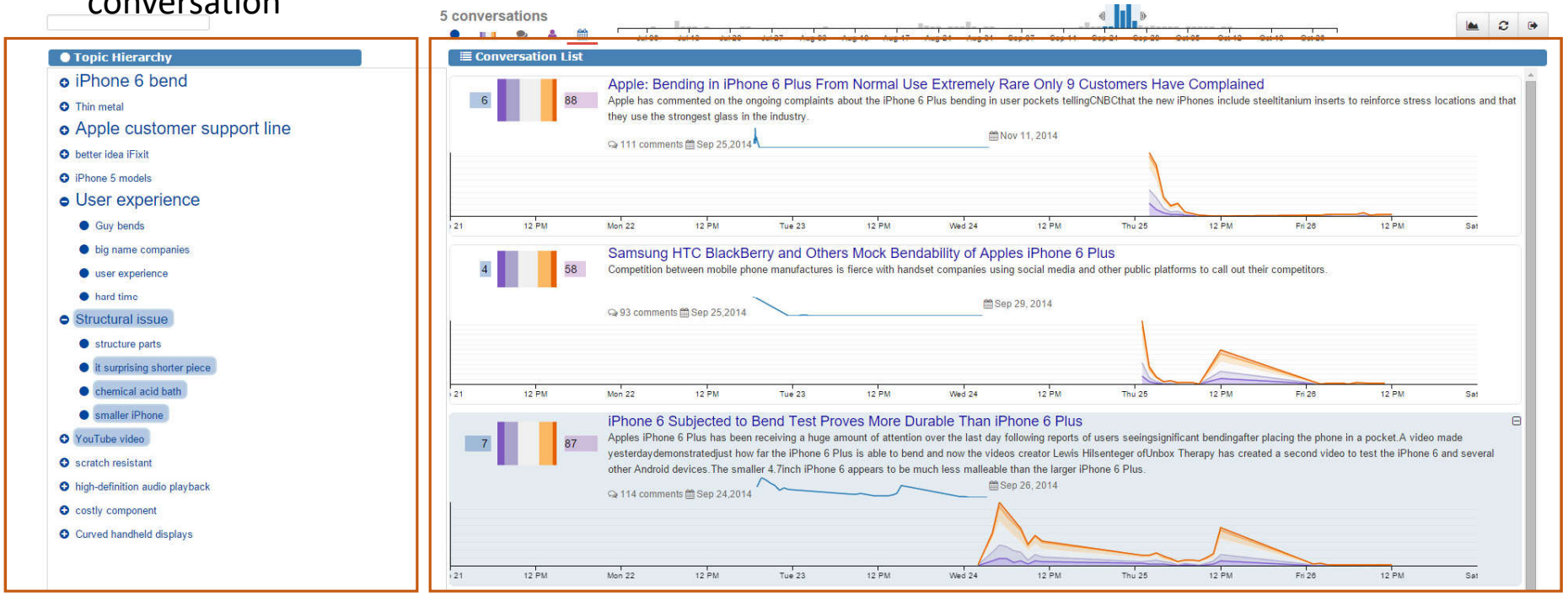


Topics | *Conversation Overview* | *Authors* | *Conversation view*

Enamul Hoque and Giuseppe Carenini *(EuroVis 2014, IUI 2015)*.

MultiConVis: Exploring a Collection of Conversations

- Large number of topics-> organize topics into **hierarchy**
- Designed on top of ConVis: **switch** from exploring a **collection** of conversations to a **single** conversation



Contributions

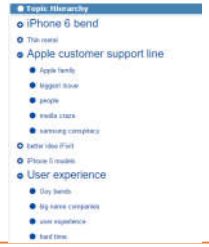
- Hierarchical topic modeling method
  - organizes large set of topics from multiple conversations

- User-centered design of MultiConVis.
  - multi-scale exploration of a collection of conversations

- Evaluation of MultiConVis :
  - user performance and subjective opinions compared to a traditional interface

Contributions

- Hierarchical topic modeling method
  - organizes large set of topics from multiple conversations



- User-centered design of MultiConVis.
  - multi-scale exploration of a collection of conversations

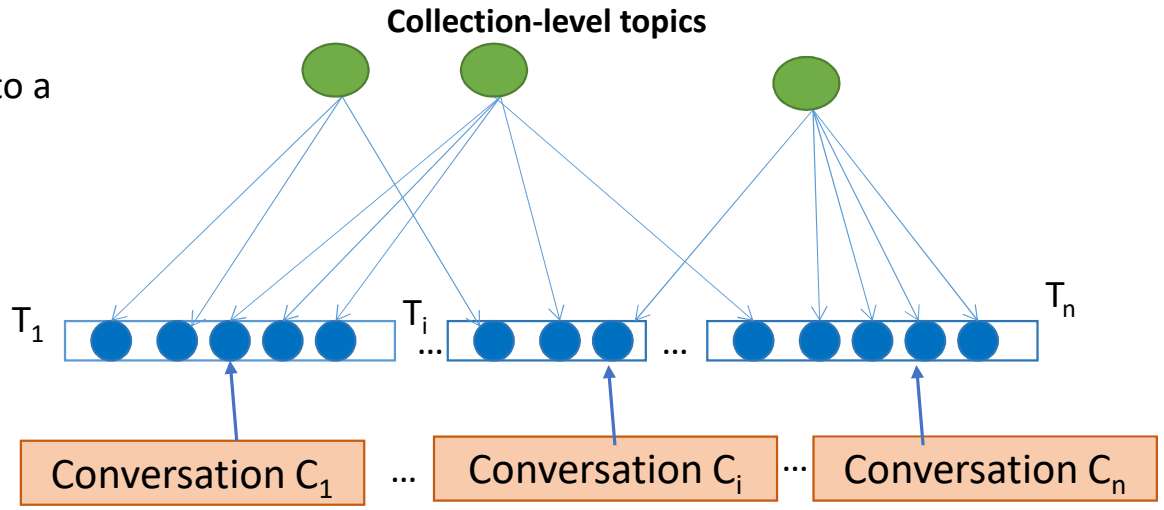
- Evaluation of MultiConVis :
  - user performance and subjective opinions compared to a traditional interface

Topic Hierarchy Generation for Multiple Conversations

Bottom-up approach:

2 The sets of topics $\{T_1, T_i, T_n\}$ are clustered into a hierarchical topic structure

1 Generate topics for each conversation
Taking conversational features into account
(Joty et al., 2013)



Topic Hierarchy Generation for Multiple Conversations

1) Create a weighted undirected graph: $G\ (V, E)$

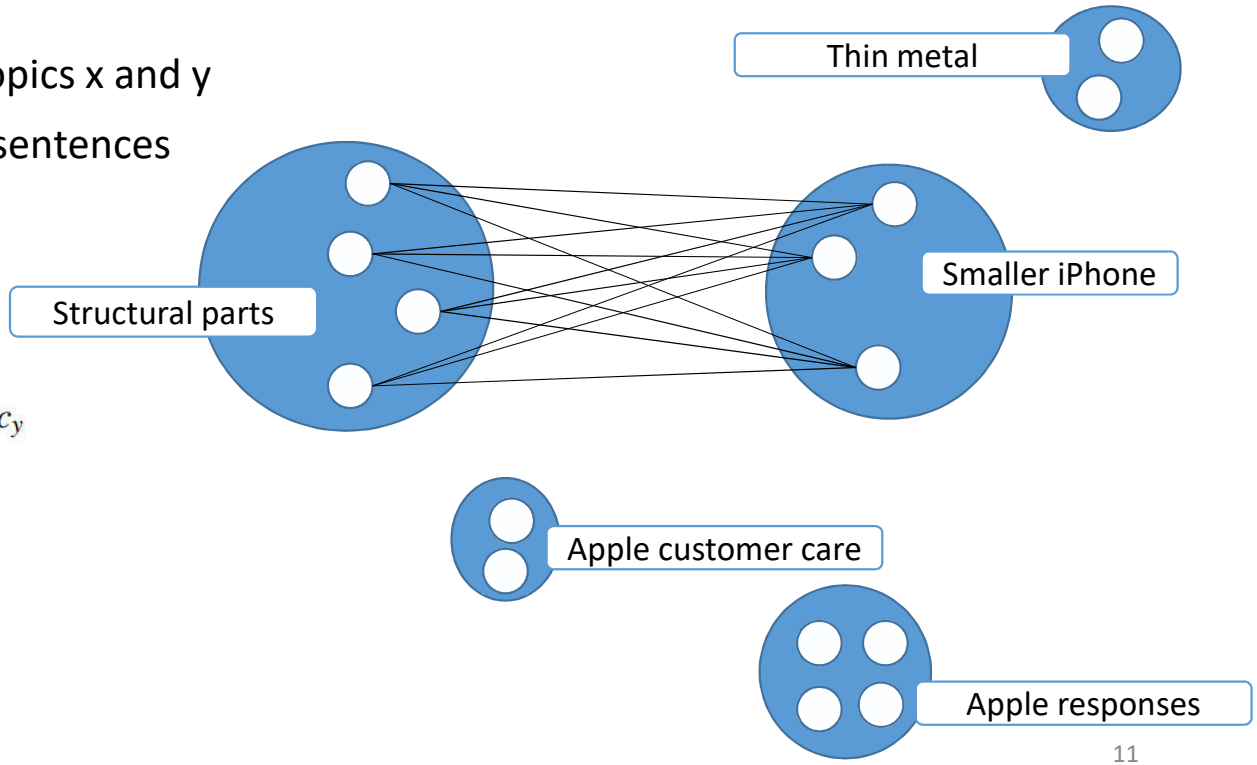
Nodes: **Topics** from conversations

Edge weight w(x,y): Similarity between two topics x and y

Sum of the pairwise similarity between their sentences

$$w(x;y) = \frac{1}{l \times m} \sum_{s_i \in S_x, s_j \in S_y} sim(s_i, s_j)$$

$$sim(s_i, s_j) = \begin{cases} \begin{cases} CosineSim(s_i, s_j) & \text{if } c_x \neq c_y \\ k & \text{if } k >= 1 \\ CosineSim(s_i, s_j) & \text{if } k = 0 \end{cases} & \text{else} \end{cases}$$



Topic Hierarchy Generation for Multiple Conversations

1) Create a weighted undirected graph: $G\ (V, E)$

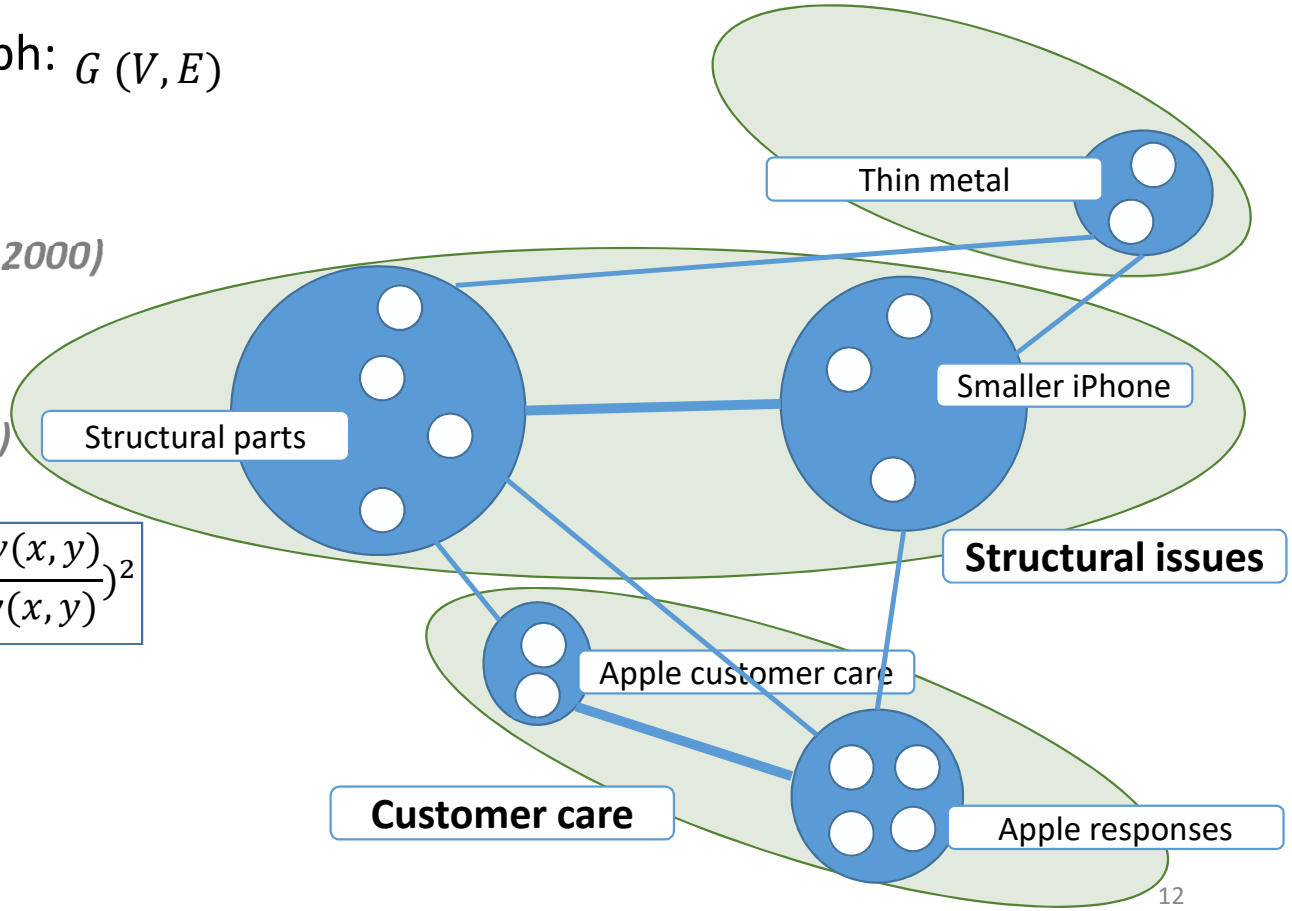
2) Apply Graph based clustering
   - Normalized cut criteria ***(Shi & Malik, 2000)***

Num. of topics: ***(Newman and Girvan, 2004)***

Maximize:

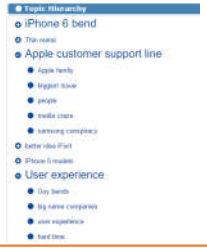
$$Q_n(G) = \sum_{c=1}^{n} \frac{\sum_{x \in V_c, y \in V_c} w(x,y)}{\sum_{x \in V, y \in V} w(x,y)} - \left(\frac{\sum_{x \in V_c, y \in V} w(x,y)}{\sum_{x \in V, y \in V} w(x,y)}\right)^2$$

3) Label each cluster



Contributions

- Hierarchical topic modeling method
  - organizes large set of topics from multiple conversations



- User-centered design of MultiConVis.
  - multi-scale exploration of a collection of conversations

- Evaluation of MultiConVis :
  - user performance and subjective opinions compared to a traditional interface

User Requirements Analysis

Why
and
how
people explore a collection of conversations?

- Topics
- Sentiment
- Time
- Authors

- **Information** seeking
- **Fact** checking
- **Guidance** seeking
- Keep track of arguments and evidences

- **When** aspect: *Find out what are people thinking or feeling about X over time"* **(Hearst 08)**
- Have **fun** and **enjoyment**

User Requirements Analysis

Why
and
how
people explore a collection of conversations?

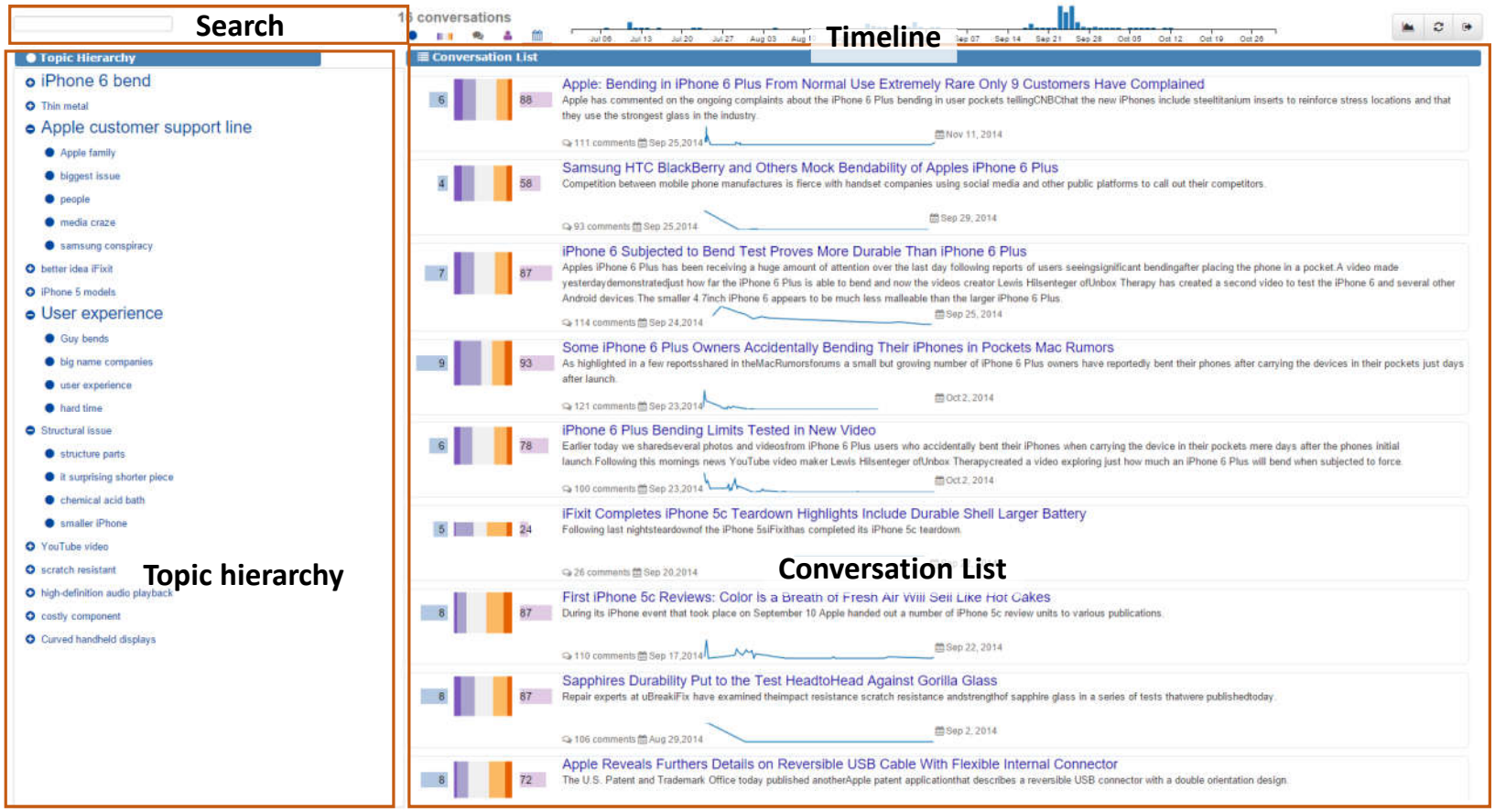
- **Variety seeking** behaviour:
  - Read various sub-topics of a topic
- **Skimming** behaviour: Explore vs. focused reading
- **Switching** between **multiple-levels** of granularity:

Various levels
All Conversations
Subset of relevant Conversations
One Conversation
-> Comments

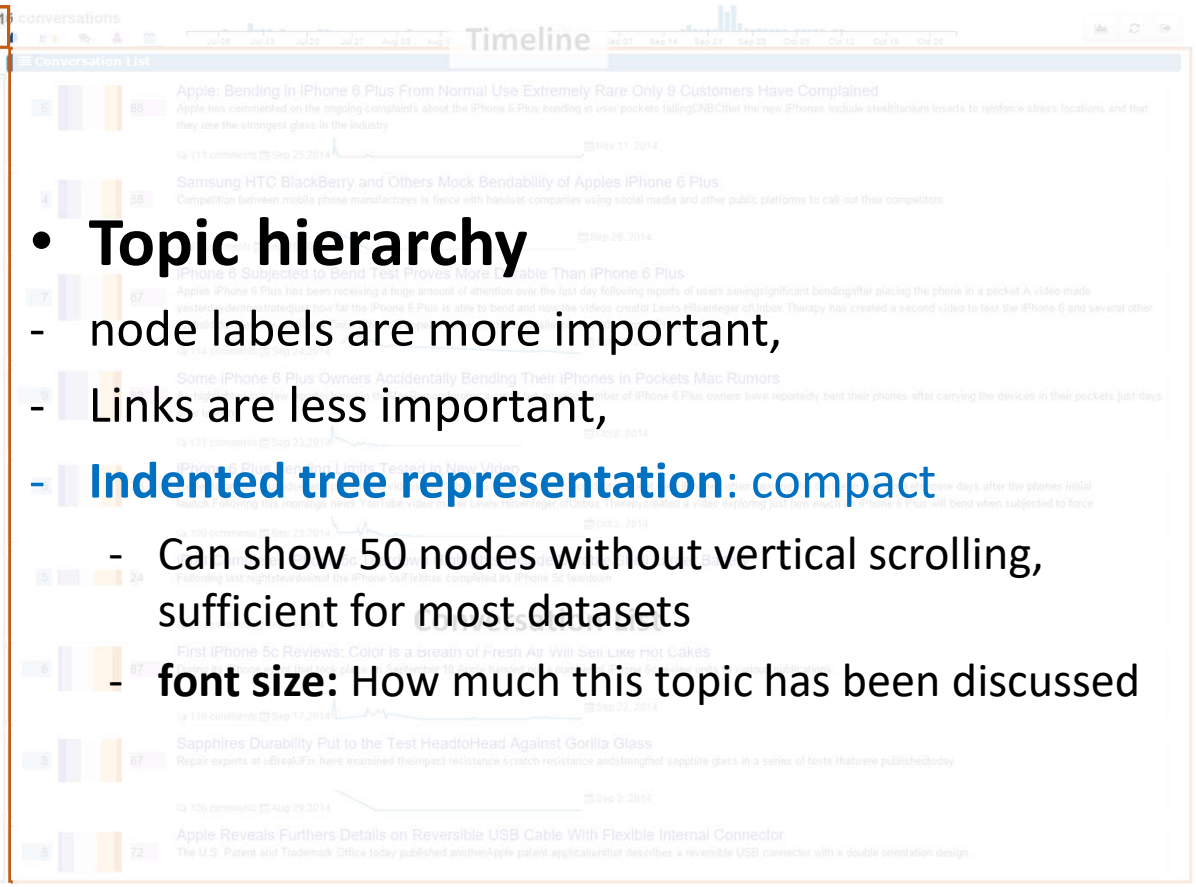
Data Abstractions

Levels / Facets	Collection of Conversations	One Conversation
Topics		
Time		
Sentiment		
Authors		

Visual Encoding: Set of Conversations

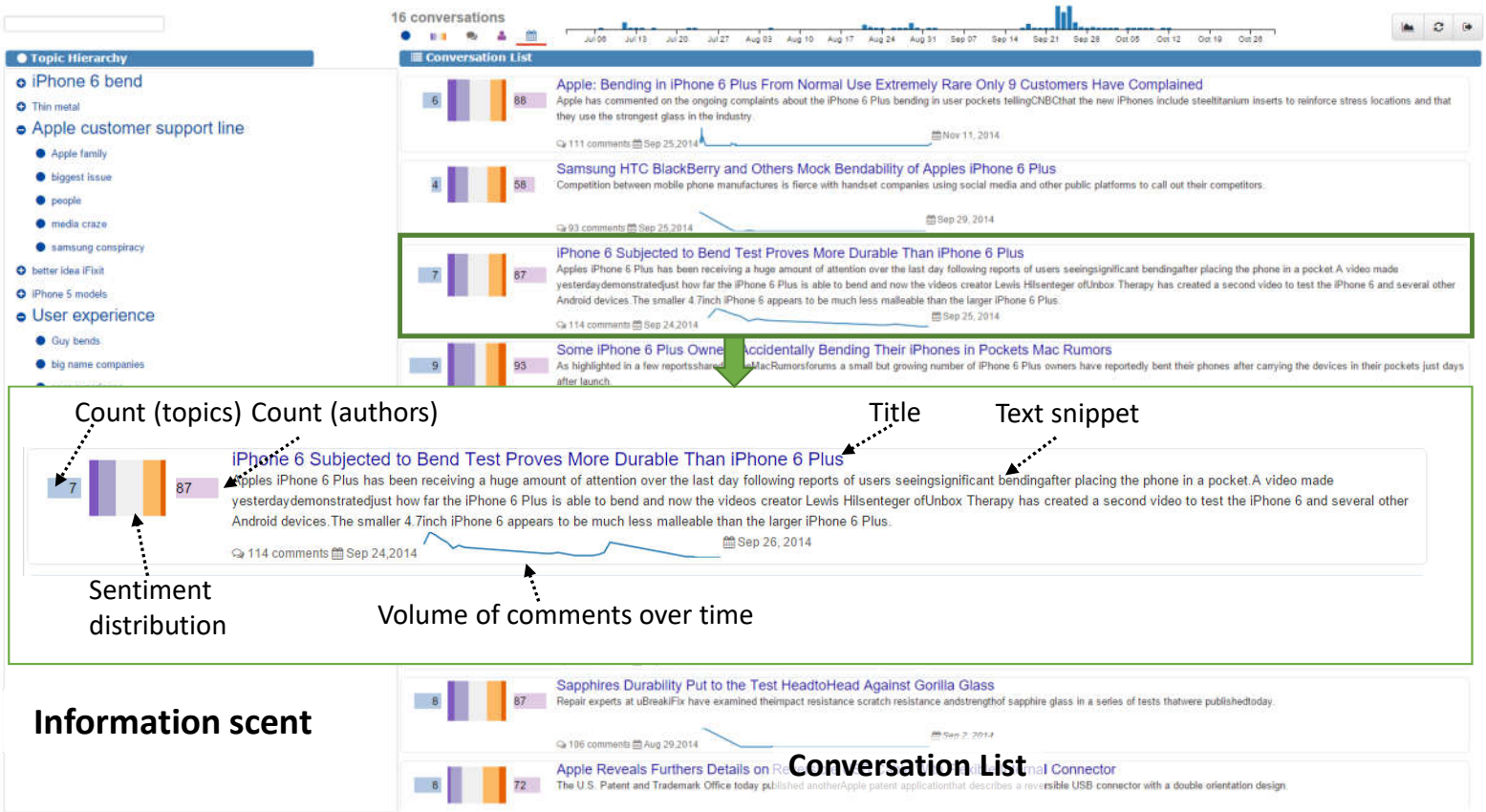


Visual Encoding: Set of Conversations



- **Topic hierarchy**
- node labels are more important,
- Links are less important,
- **Indented tree representation**: compact
  - Can show 50 nodes without vertical scrolling, sufficient for most datasets
  - **font size:** How much this topic has been discussed

Visual Encoding: Set of Conversations



Video Demo

Contributions

- Hierarchical topic modeling method
  - organizes large set of topics from multiple conversations

- User-centered **design** of MultiConVis.
  - **multi-scale exploration** of a collection of conversations

Evaluation of MultiConVis :

- user performance and subjective opinions compared to a traditional interface

User Evaluation

Case studies:

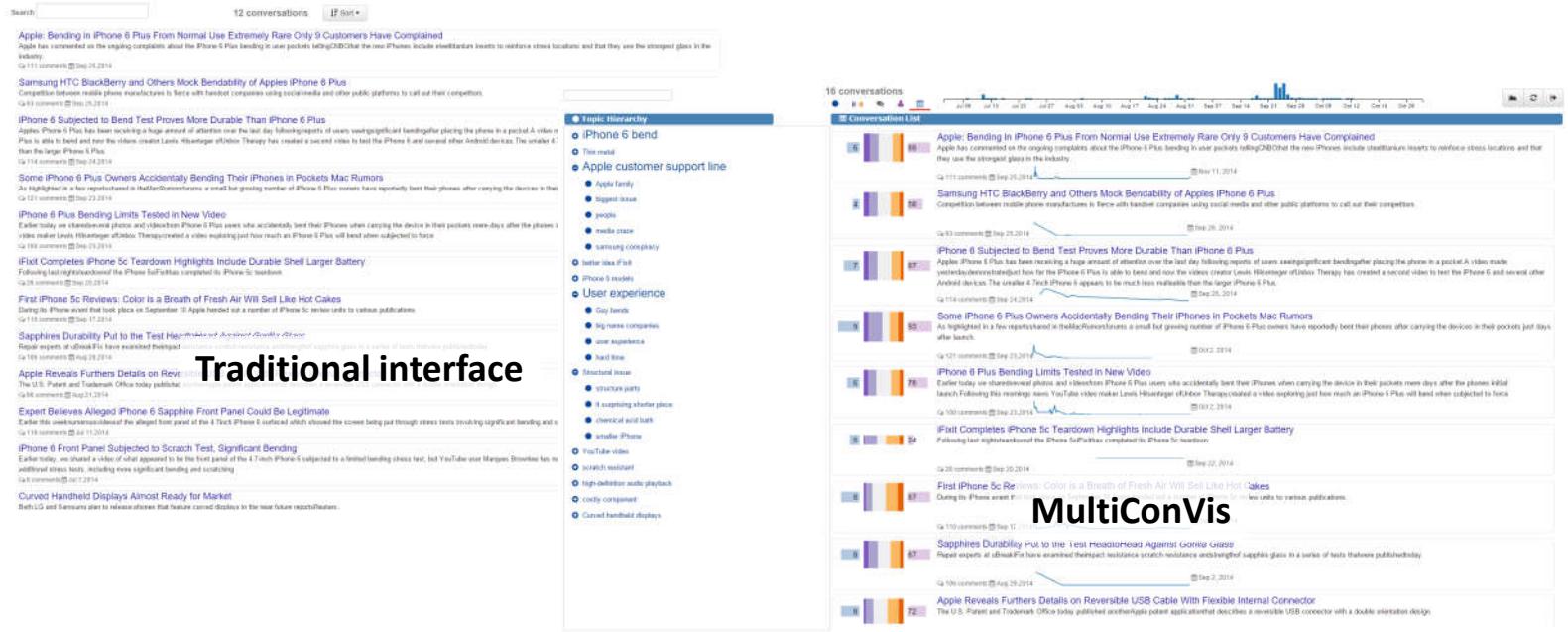
- Participants explored the datasets according to their information needs
  - Regular blog reader: iPhone bending
  - Journalist: ObamaCare health reform
  - Business analyst: iWatch release
- In follow-up interviews: topic hierarchy was extremely useful

Laboratory study:

- Compare with a traditional interface
- **Task:** Explore the given set of conversations, write a summary of major keypoints

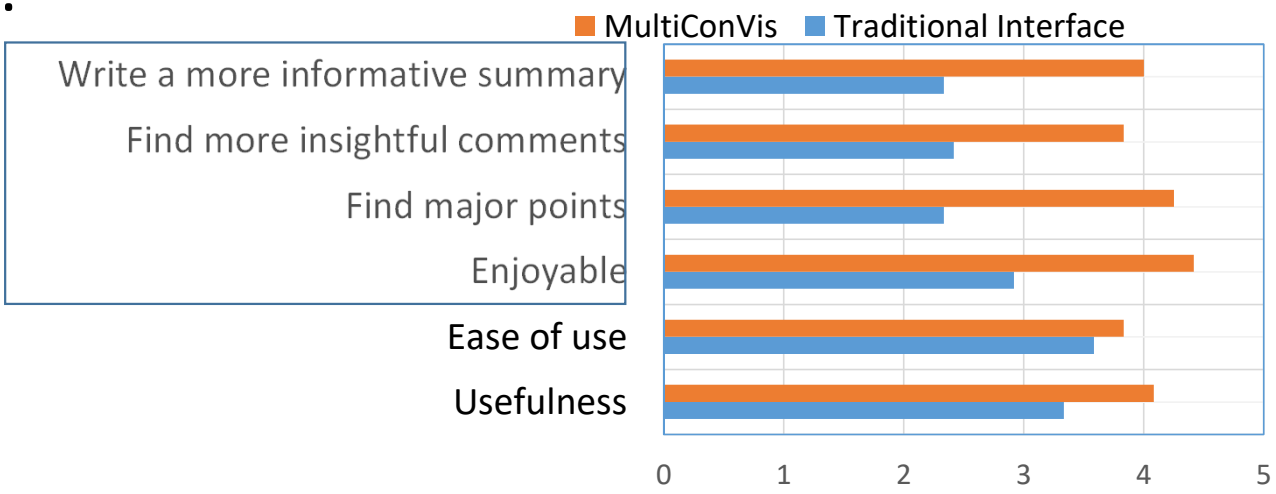
Evaluation: Lab Study

- 16 subjects (aged 18-37, 6 females)
- Within subjects



User Study: Selected Results

- **Time-to-task completion:** No significant difference
- **Subjective ratings:**



- **Preference:**
  - MultiConVis (75%): topic organization, visual overview of conversations
  - Traditional interface (25%): simplicity and familiarity

Conclusions

1) **Hierarchical topic modeling** for a collection of online conversations
   - consider **unique features of conversations.**

2) **Design** of MultiConVis.
   - **Multi-scales exploration** of a collection of conversation
     - Consistency of encoding among various scales

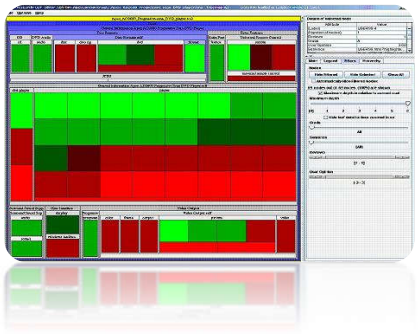
3) **Evaluation**
   - MultiConVis was preferred by majority of participants
   - Assessment of different interface features

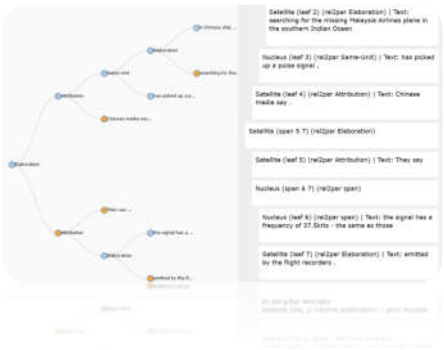
Future Work

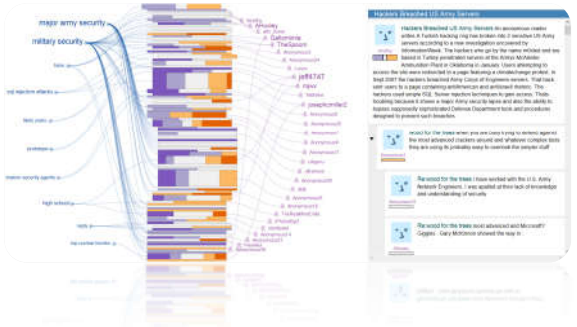
- Interactive topic hierarchy revisions
  - Allow user to modify topic hierarchy
- Apply and tailor to specific conversational genres
  - Community question answering forums
  - MOOC forums
  - ….
- Online longitudinal study
  - For ecologically validity

For More Information…

www.cs.ubc.ca/cs-research/lci/research-groups/natural-language-processing/







Thanks:

Tamara Munzner

Raymond T. Ng

NSERC CRSNG