a place of mind
THE UNIVERSITY OF BRITISH COLUMBIA

Visual Text Analytics for Online Conversations

Enamul Hoque
PhD Candidate, Computer Science, UBC
enamul@cs.ubc.ca

Problem Scenario

- Lot of articles and comments were posted on Macumers.

- 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

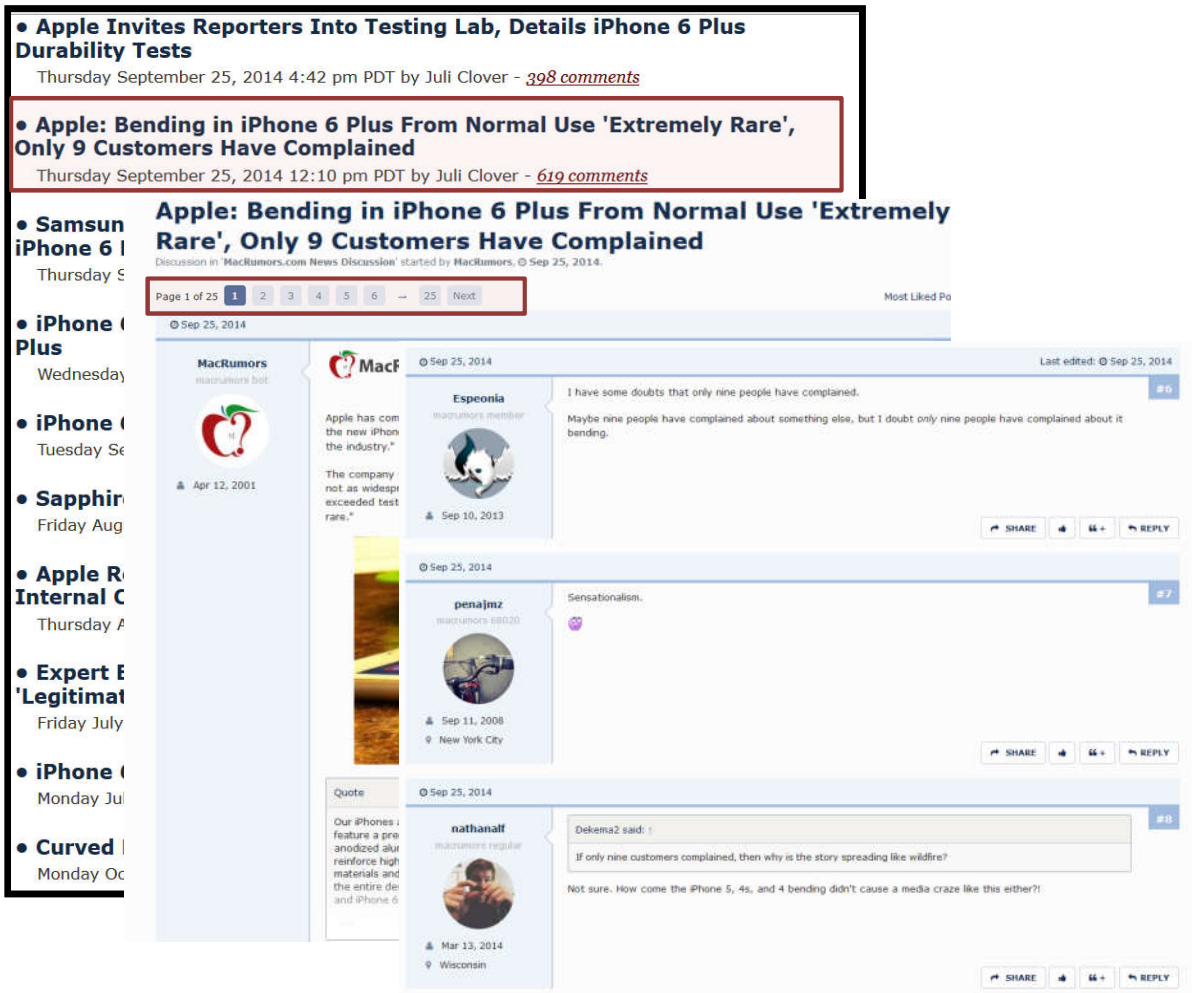
- Lack of high-level abstraction
  - Only show conversations/comments as paginated lists ordered by recency

- Too many conversations
- Too many comments

=> **Information Overload**

Users

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

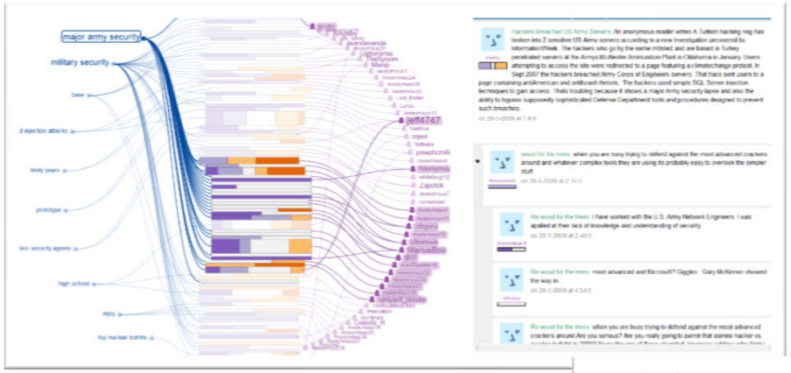


Our Goal

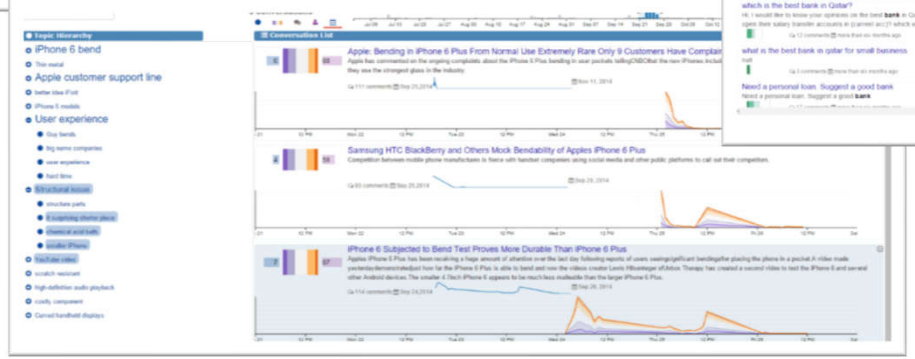
tightly integrate ***text analysis*** *and* ***interactive visualization*** *to support users in exploring* ***collection of online conversations****.*

Interactive visualization

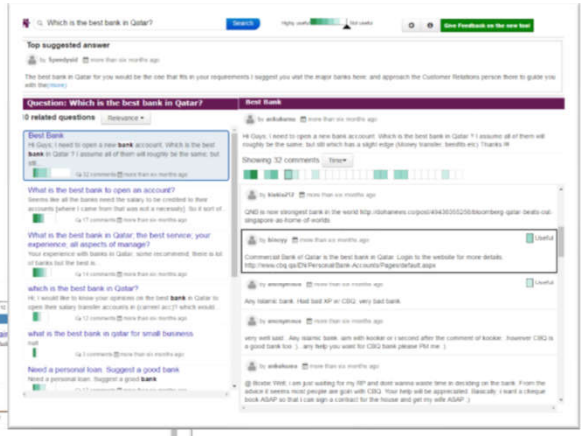
Tools for Exploring Online Conversations



ConVis, Eurovis 2014

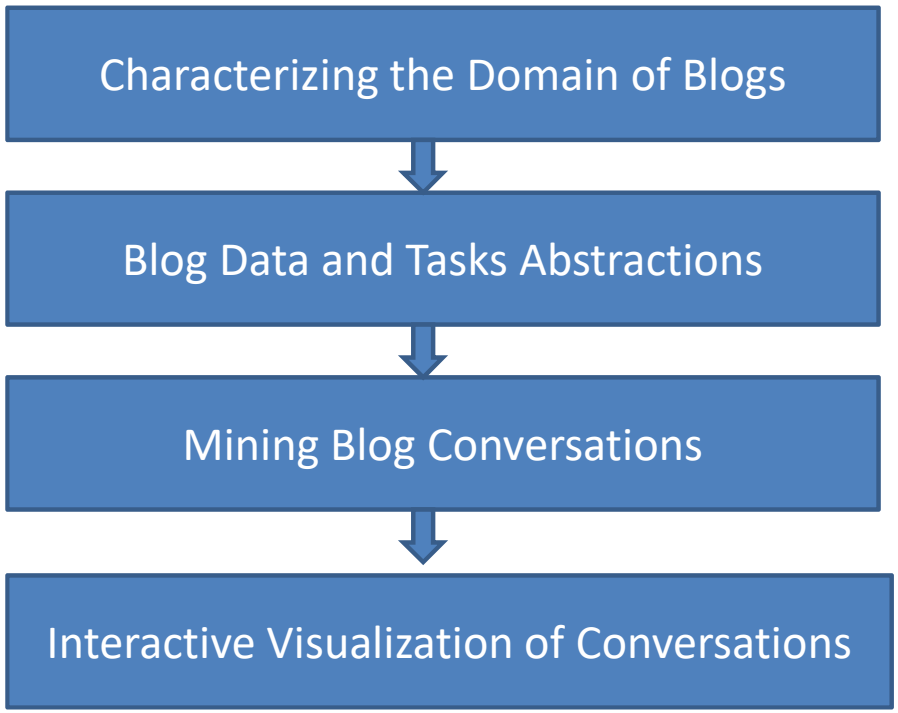


MultiConVis, IUI 2016



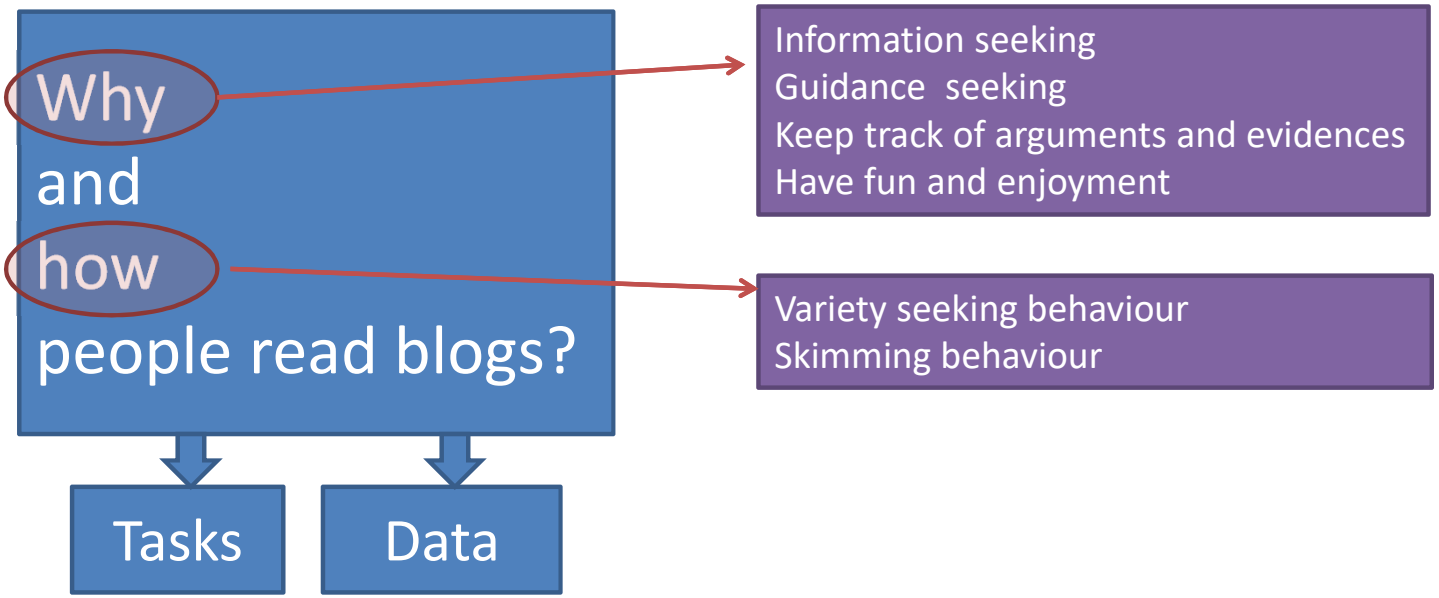
CQAVis, IUI 2017

Overall Approach



Characterizing the Domain of Blogs

- **Computer mediated communications**
- **Social media**
- **Human computer interactions (HCI)**
- **Information retrieval**



Blog Data and Tasks Abstractions

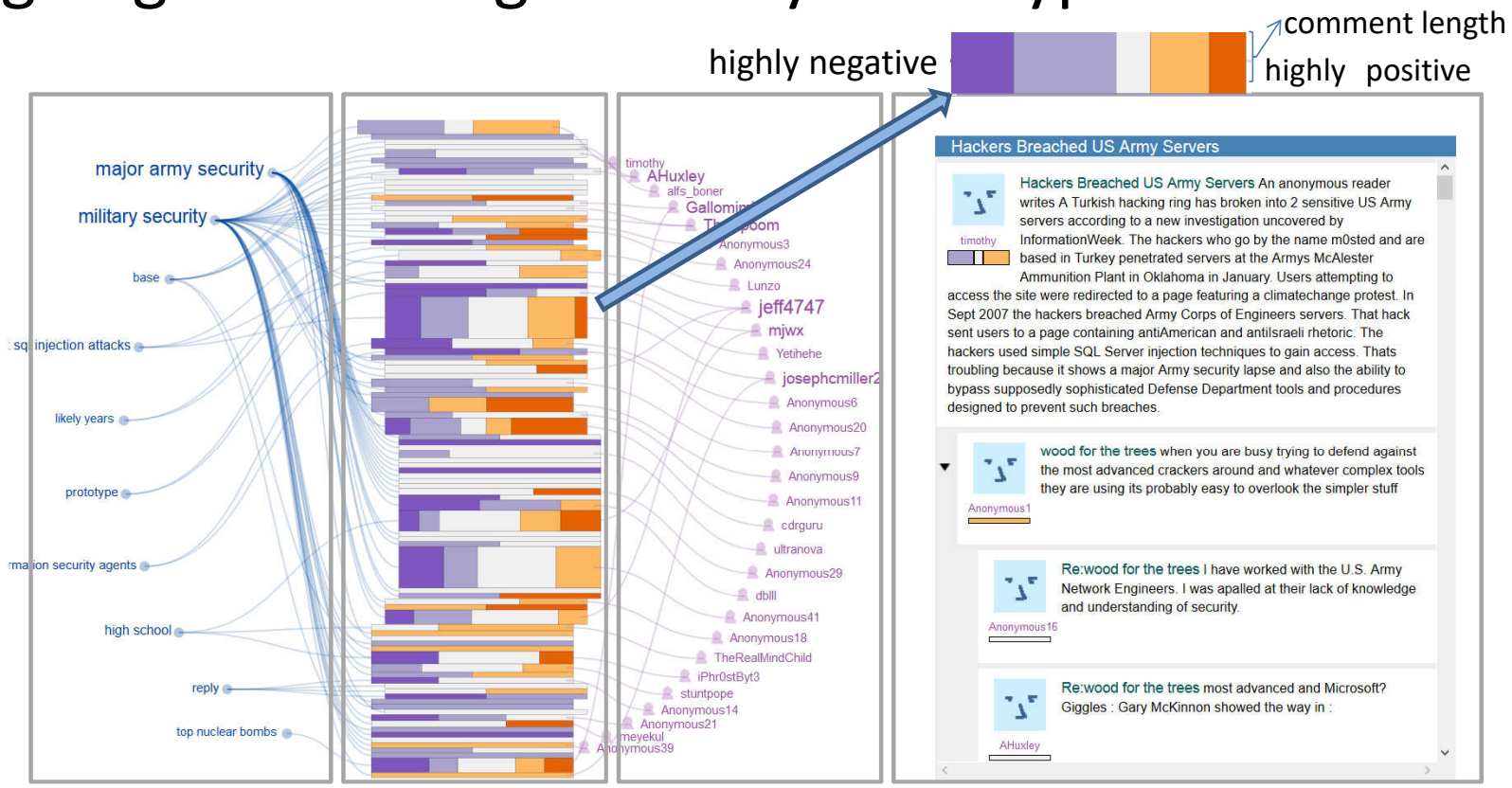
	Data Variables				
TASKS	**Topic**	**Author**	**Opinion**	**Thread**	**Comment**
What this conversation is about?	x				x
Which topics are generating more discussions?	x				
What do people say about topic X?			x	x	x
How controversial was the conversation? Were there substantial differences in opinion?	x	x	x	x	x
Why are people supporting/ opposing an opinion?			x	x	
…	…	…	…	…	…

Text Analysis for Conversations

- **Topic modeling** *(Joty et al., 2013)*
  - Take advantage of the conversational structure
  - Graph based clustering (normalized n-cut)
  - Generate keyphrases for each cluster
    - Co-ranking

- **Sentiment analysis** *(Taboada et al., JCL 2011)*
  - So-CAL: Lexicon-based approach
  - Compute polarity distribution for each comment

Designing ConVis: High-Fidelity Prototype

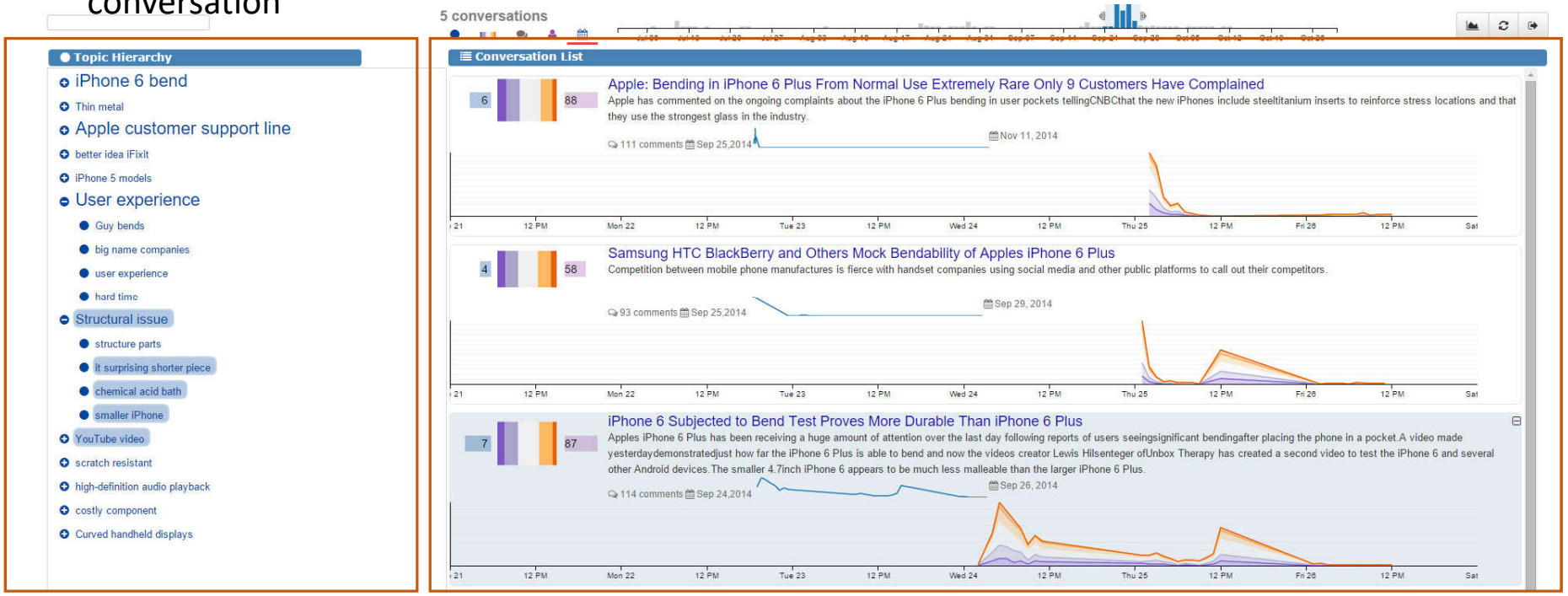


Topics *Conversation Overview* *Authors* *Conversation view*

For particular tasks such as document comprehension, **overview + details** has been found more effective. *(Cockburn et al. 2008)*

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

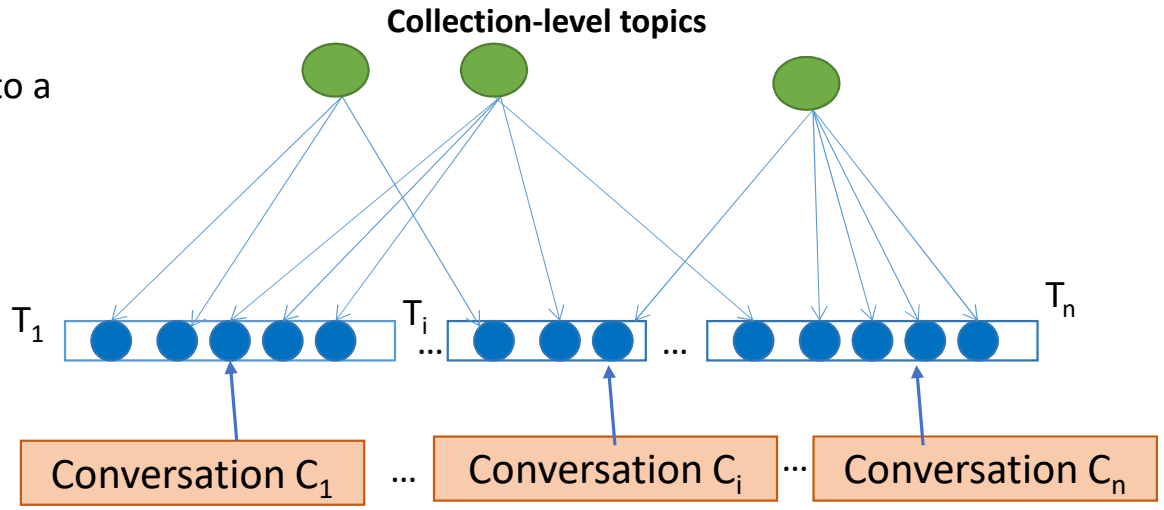


Topic Hierarchy Generation for Multiple Conversations

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

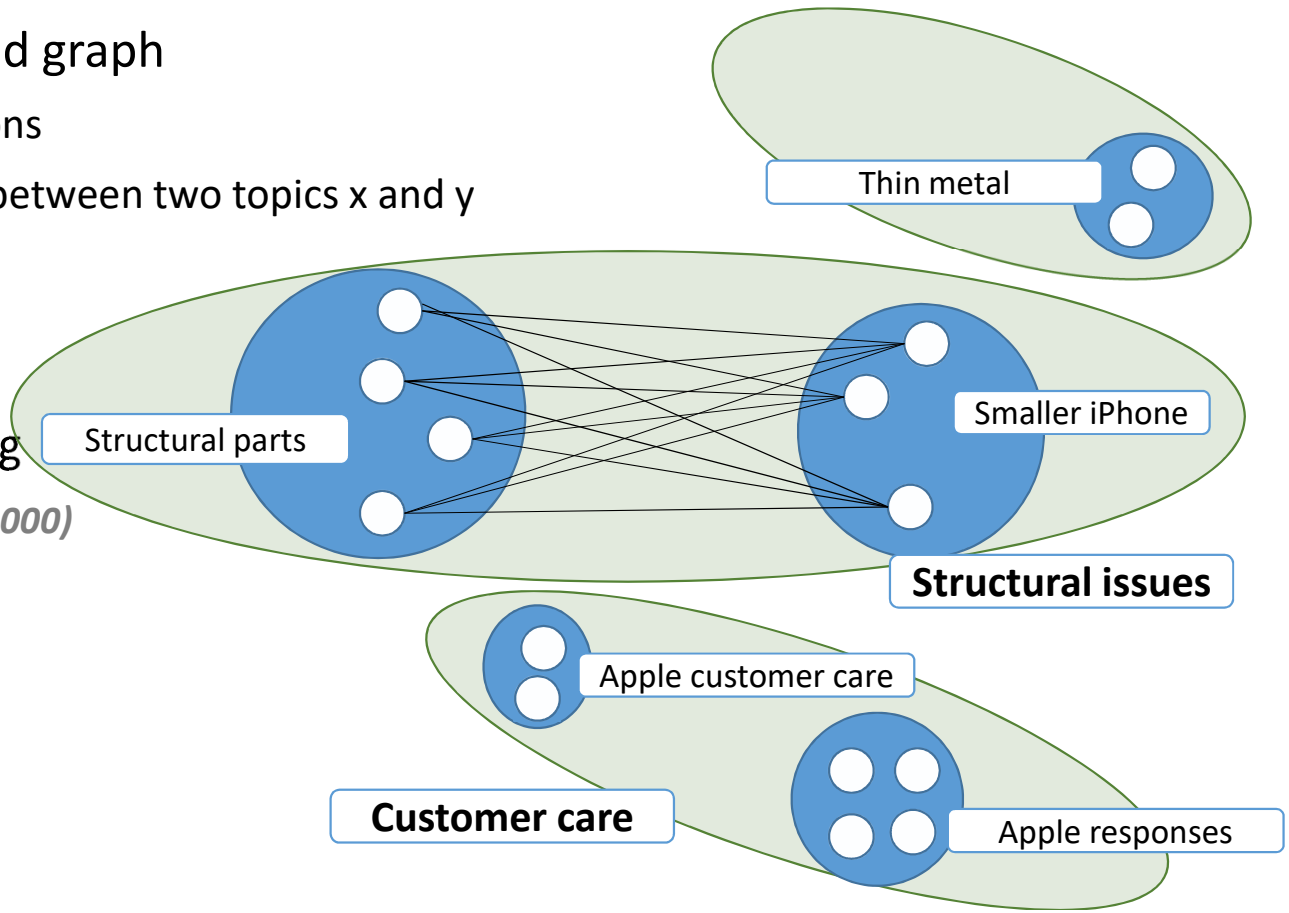
1) Create a weighted undirected graph

- Nodes: **Topics** from conversations
- Edge weight w(x,y): Similarity between two topics x and y

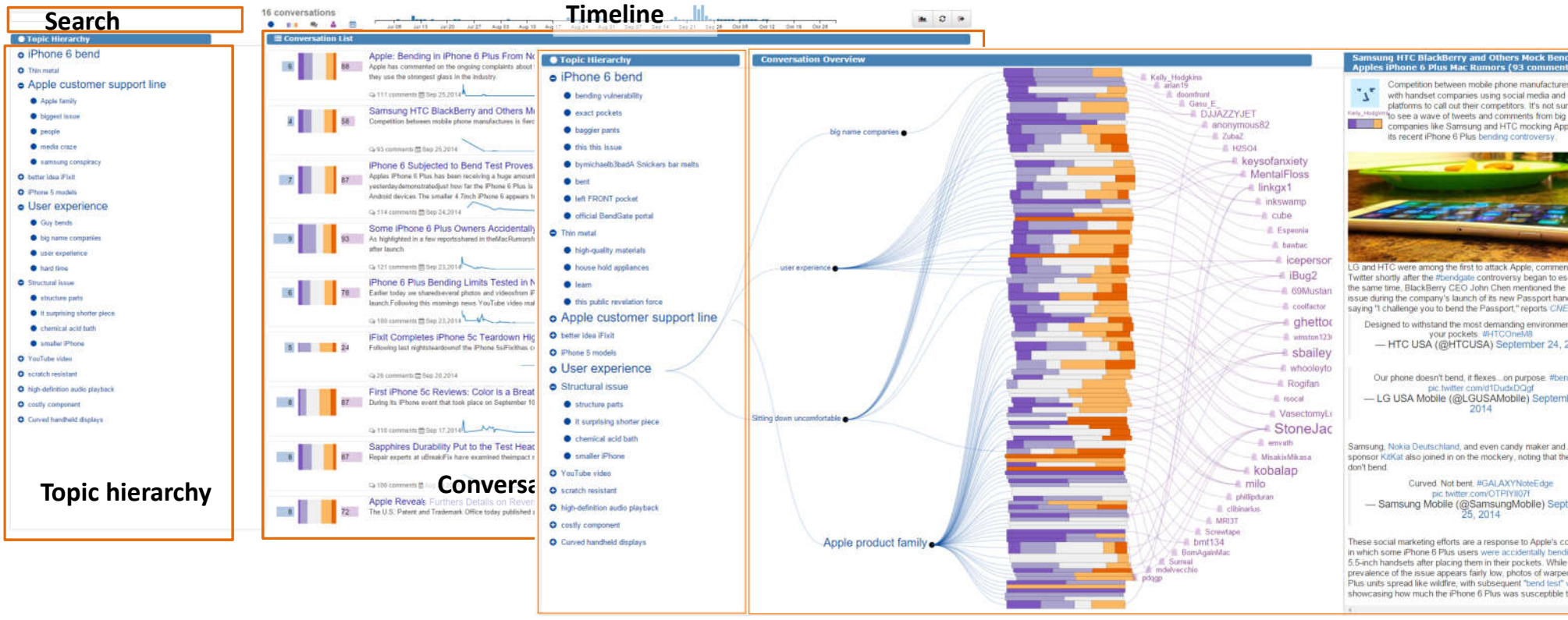
2) Apply Graph based clustering

- N-cut criteria *(Shi & Malik, 2000)*

3) Label each cluster



User-centered Design of MultiConVis



Further Information

UBC @ NLP