



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

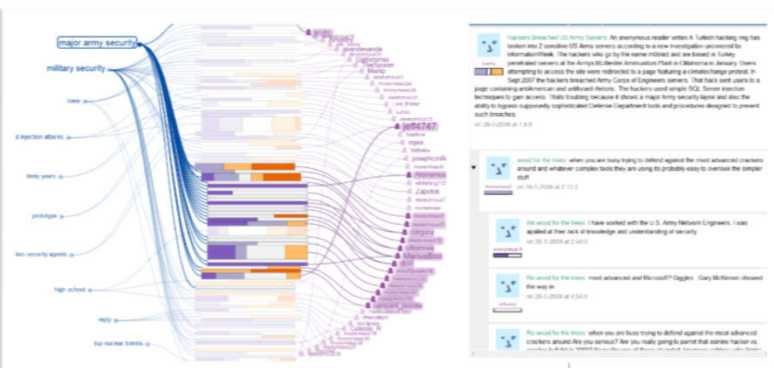
The screenshot shows a forum thread on MacRumors.com. The thread title is "Apple: Bending in iPhone 6 Plus From Normal Use 'Extremely Rare', Only 9 Customers Have Complained". It was started by Juli Clover on Thursday, September 25, 2014, at 4:42 pm PDT, with 398 comments. A second post by Juli Clover at 12:10 pm PDT has 619 comments. The thread is on page 1 of 25. The forum post by Juli Clover reads: "Apple has com the new iPhone the industry." "The company not as widespr exceeded test rare." Below this, there are several replies. One reply from Espeonia says: "I have some doubts that only nine people have complained. Maybe nine people have complained about something else, but I doubt only nine people have complained about it bending." Another reply from pena1mz says: "Sensationalism." A third reply from nathanalt says: "Dekema2 said: 'If only nine customers complained, then why is the story spreading like wildfire?' Not sure. How come the iPhone 5, 4s, and 4 bending didn't cause a media craze like this either?!" The forum interface includes a sidebar with navigation links, a main content area with the thread title and posts, and a right sidebar with user avatars and profile information.

Our Goal

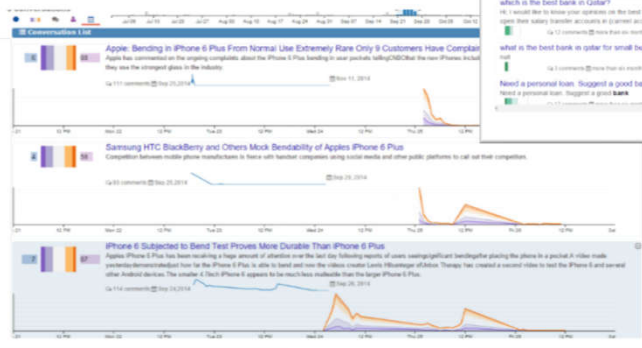
*tightly integrate **text analysis** and **interactive visualization** to support users in exploring **collection of online conversations**.*



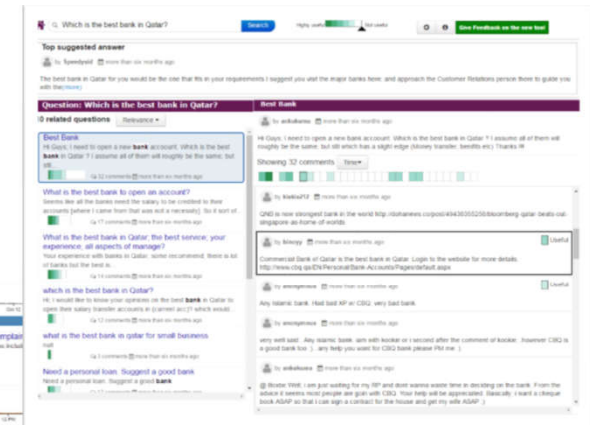
Tools for Exploring Online Conversations



ConVis, Eurovis 2014

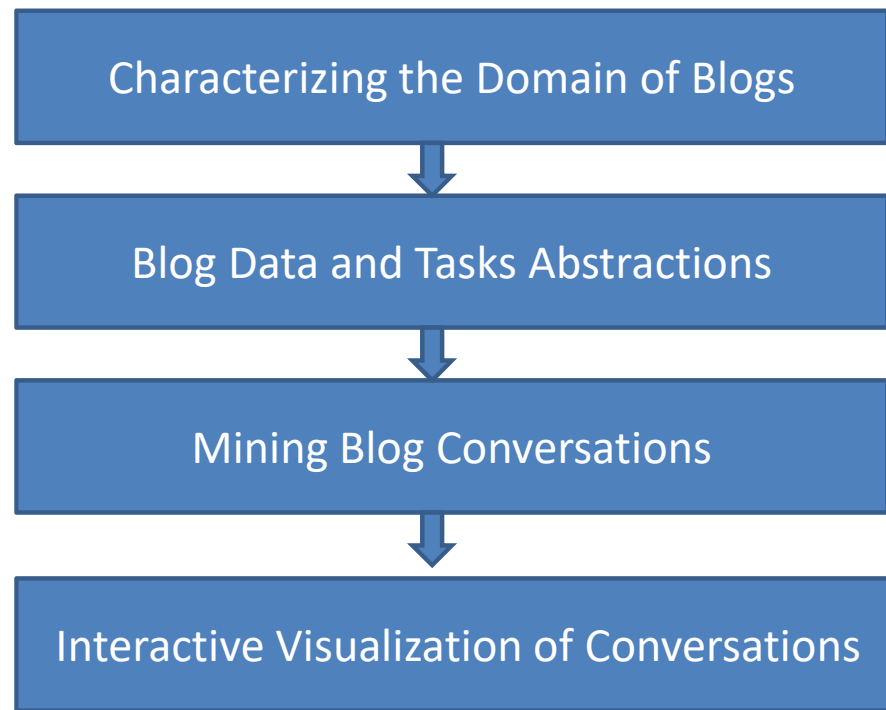


MultiConVis, IUI 2016



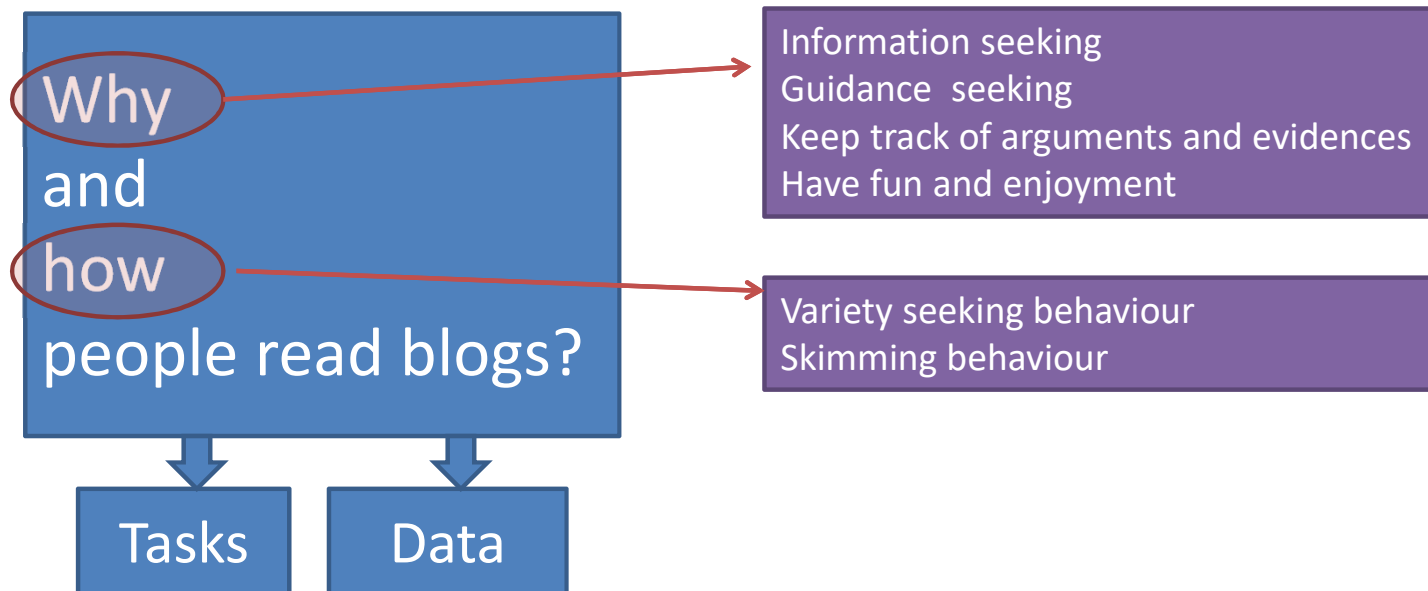
CQAVis, IUI 2017

Overall Approach



Characterizing the Domain of Blogs

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



Blog Data and Tasks Abstractions

TASKS	Data Variables				
	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)

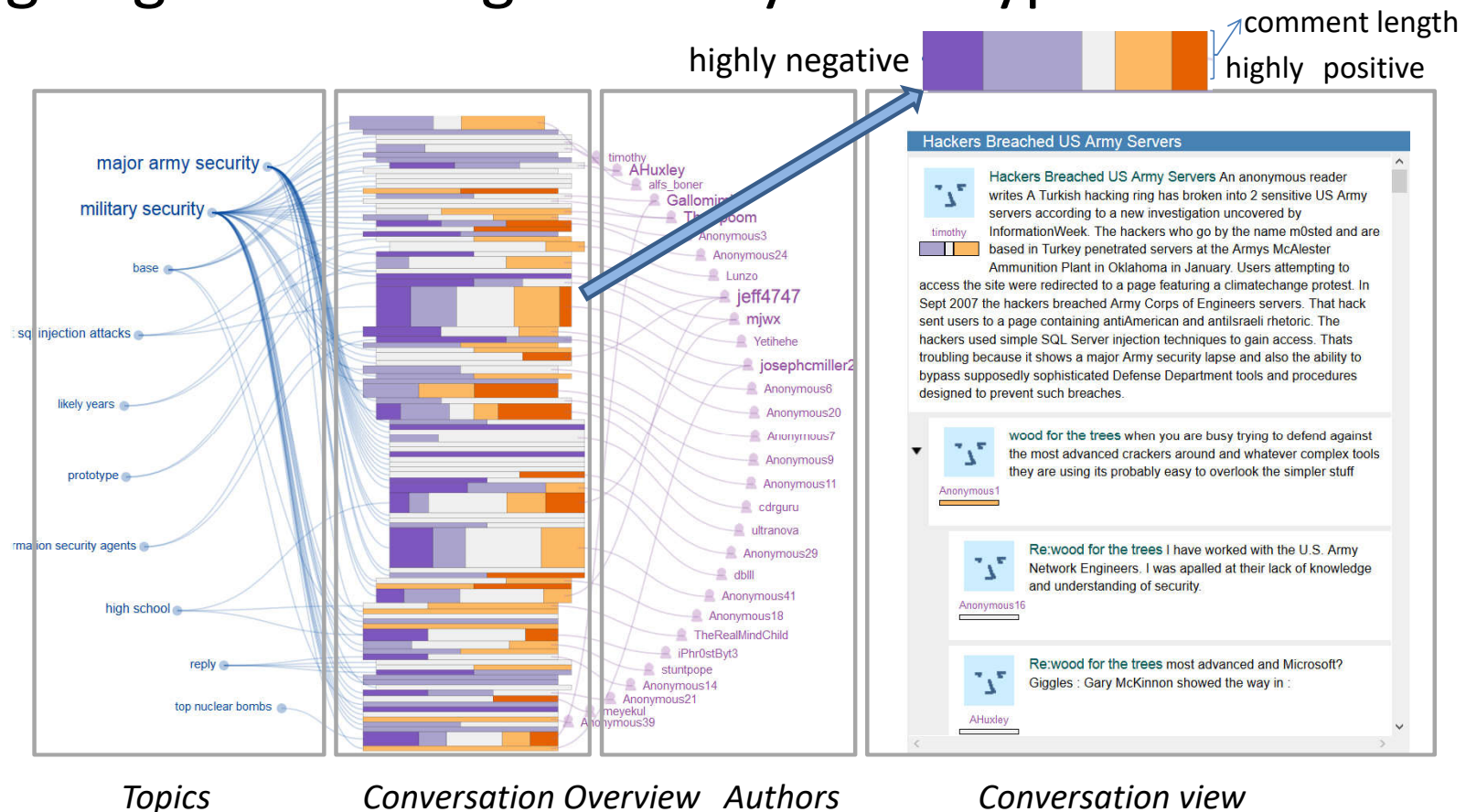
- Take advantage of the conversational structure
- Graph based clustering (normalized n-cut)
- Generate keyphrases for each cluster
 - Co-ranking

- **Sentiment analysis**

(Taboada et al., JCL 2011)

- So-CAL: Lexicon-based approach
- Compute polarity distribution for each comment

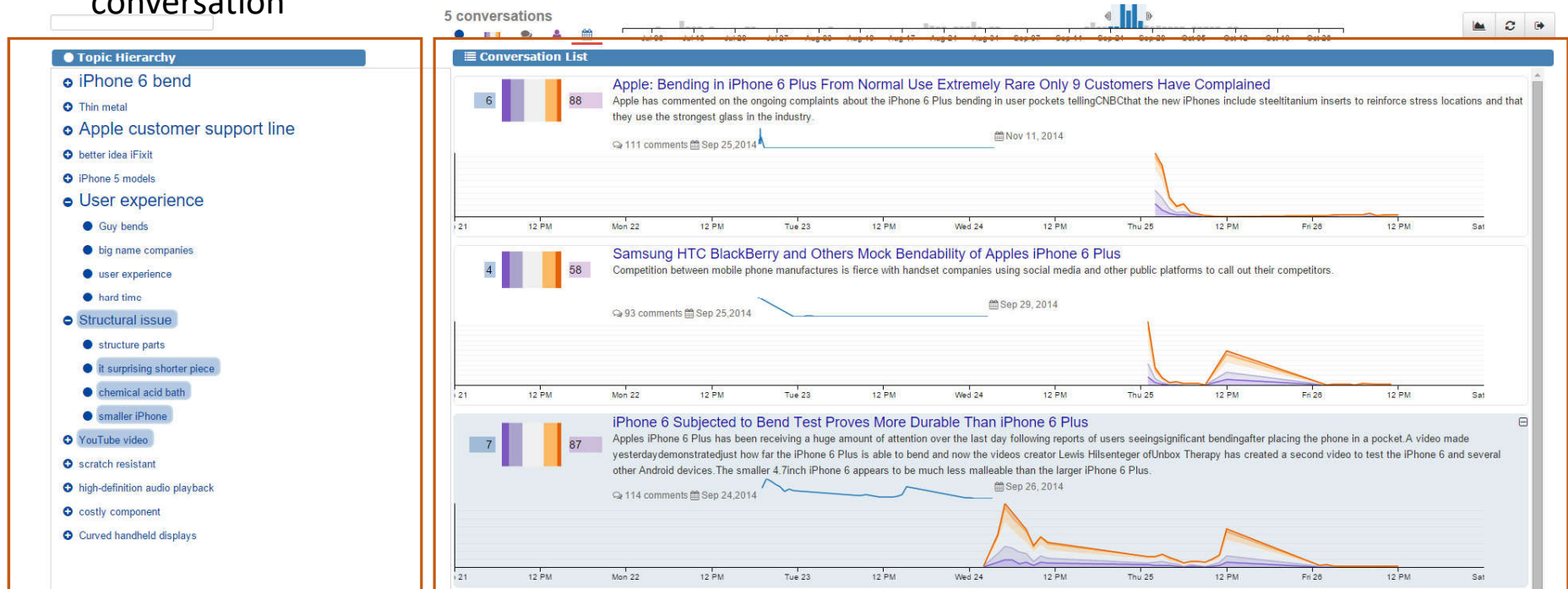
Designing ConVis: High-Fidelity Prototype



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

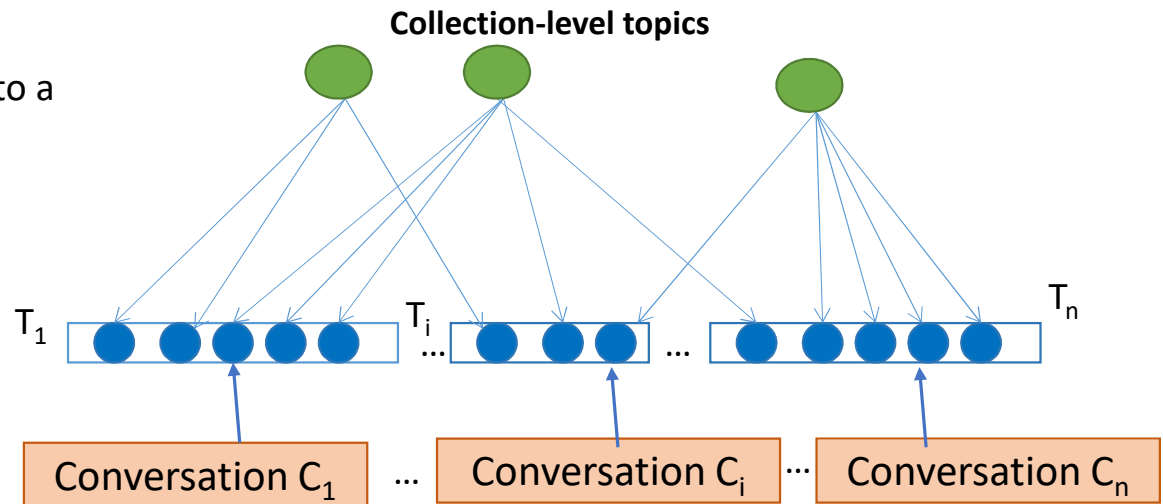
Bottom-up approach:

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

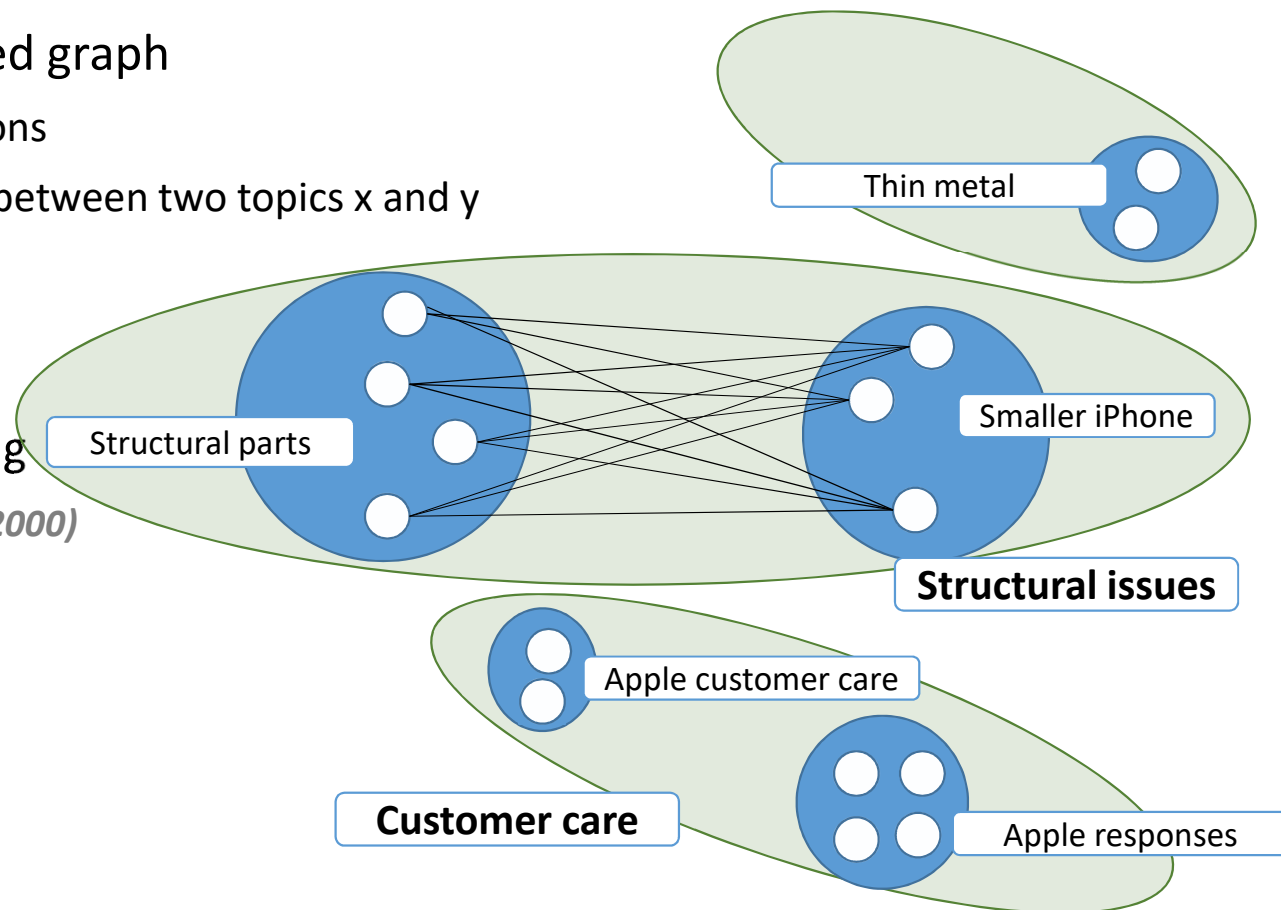
1) Create a weighted undirected graph

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

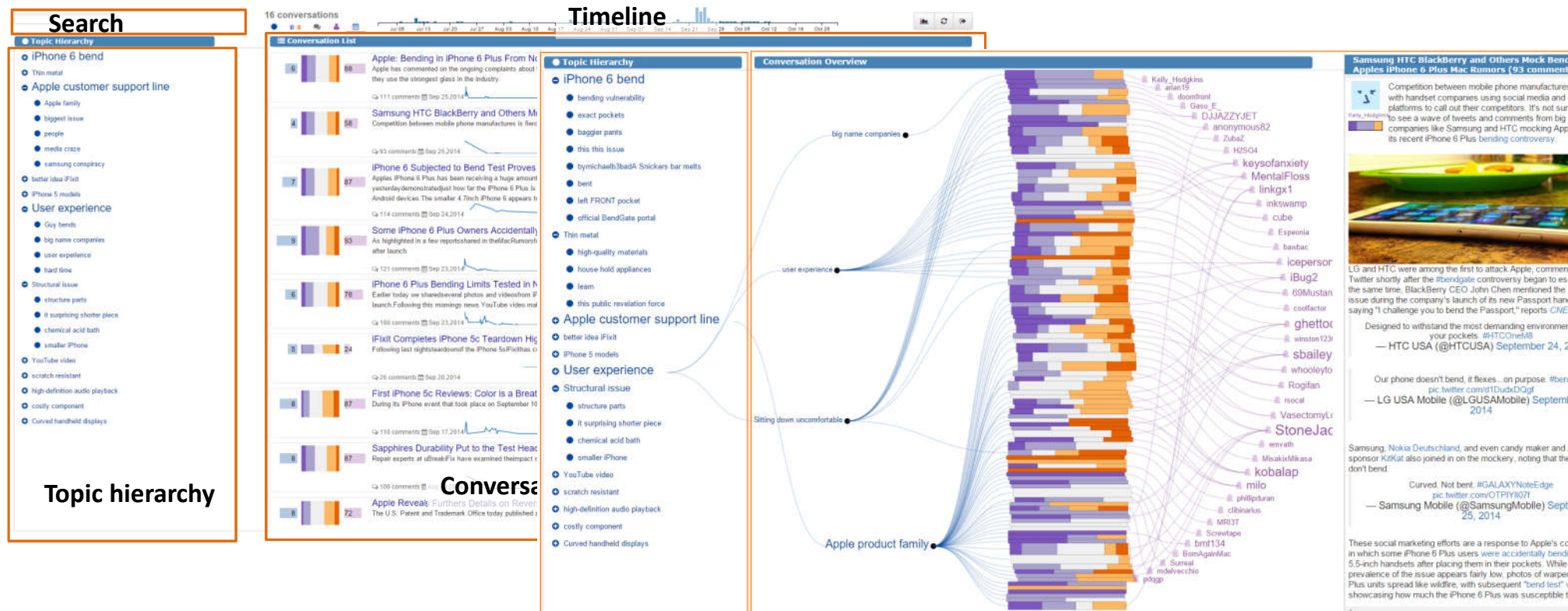
2) Apply Graph based clustering

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

3) Label each cluster



User-centered Design of MultiConVis



Further Information

UBC @ NLP

