THE UNIVERSITY
OF BRITISH COLUMBIA

Visual analytics for exploring online conversations

Soheil Alavi
salavis@cs.ubc.ca

Felipe González-Pizarro
felipegp@cs.ubc.ca

Topic modeling and InfoVis

- It is needed to **summarize** and **understand** textual data

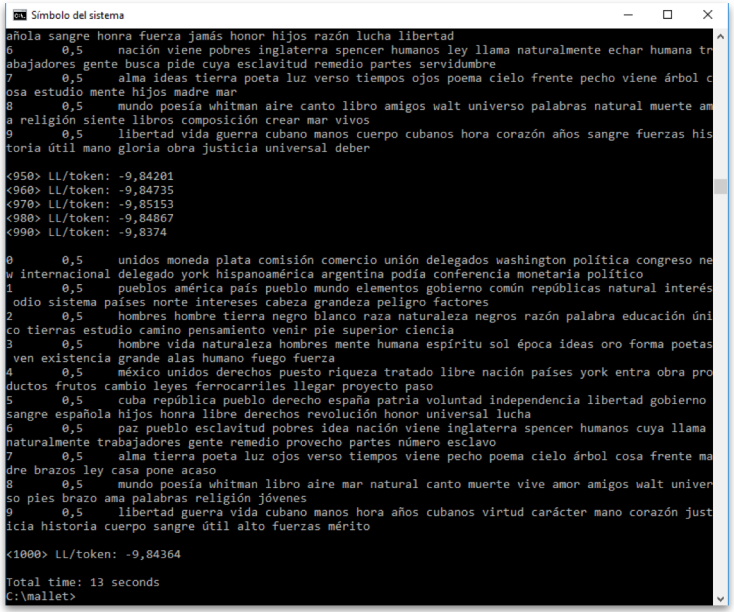
- Promising solution: Topic modeling
  - Statistical approach for extracting **topics** from large text corpora.
  - Topic models do not provide meanings and interpretation directly
    - **humans must be involved [1]**

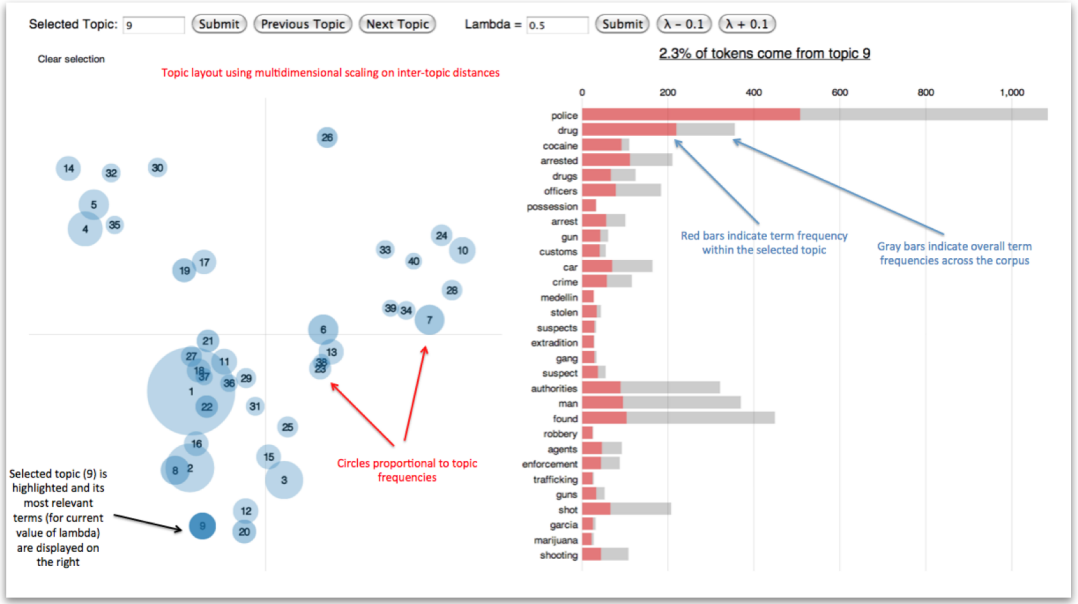
- Humans who directly interact with and interpret the output of topic modeling **may rely on visualization tools to better interpret the results [2]**

[1] Dou, W., Wang, X., Chang, R., & Ribarsky, W. (2011, October). Paralleltopics: A probabilistic approach to exploring document collections. In *2011 IEEE conference on visual analytics science and technology (VAST)* (pp. 231-240). IEEE.

[2] Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems* (pp. 288-296).

Topic modeling interpretation

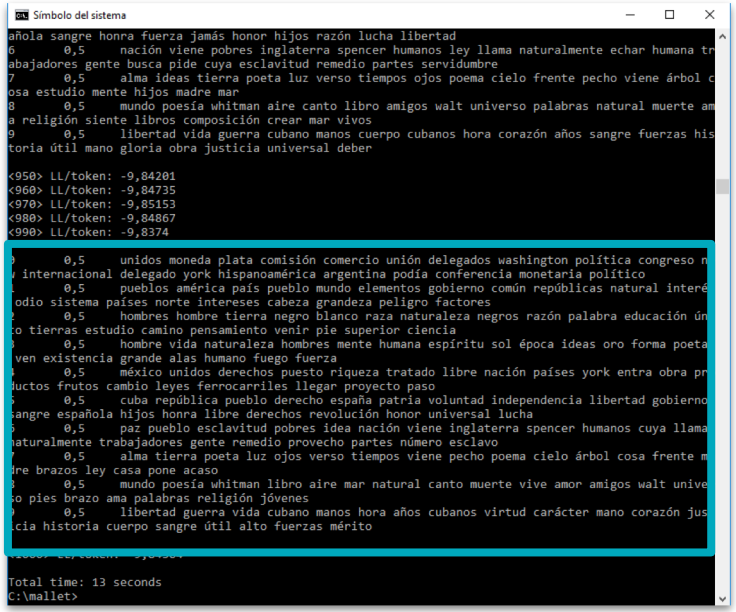


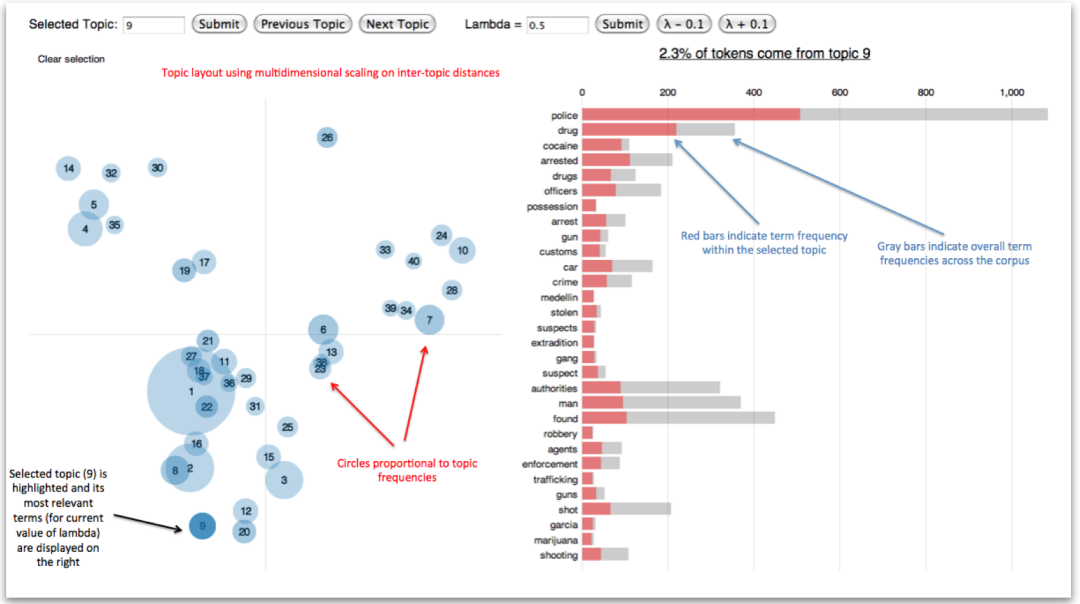


[1] Graham, S., Weingart, S., & Milligan, I. (2012). Getting started with topic modeling and MALLET. The Editorial Board of the Programming Historian.

[2] Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

Topic modeling interpretation

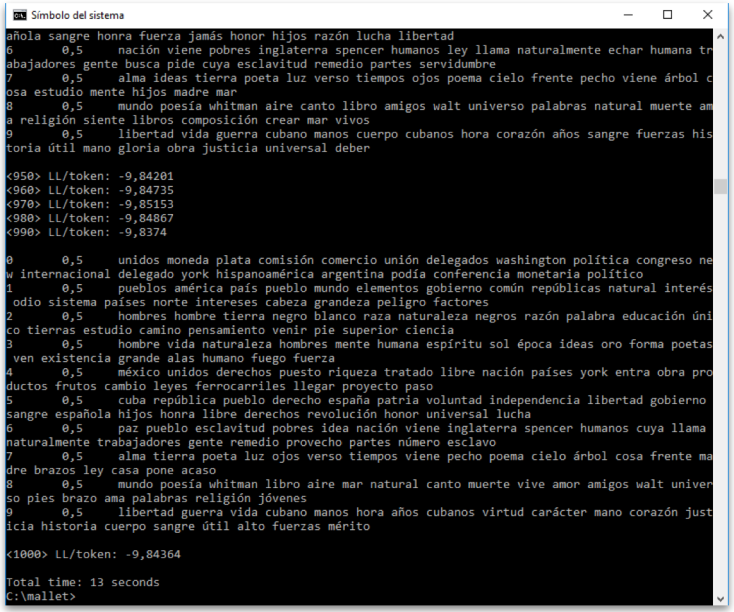


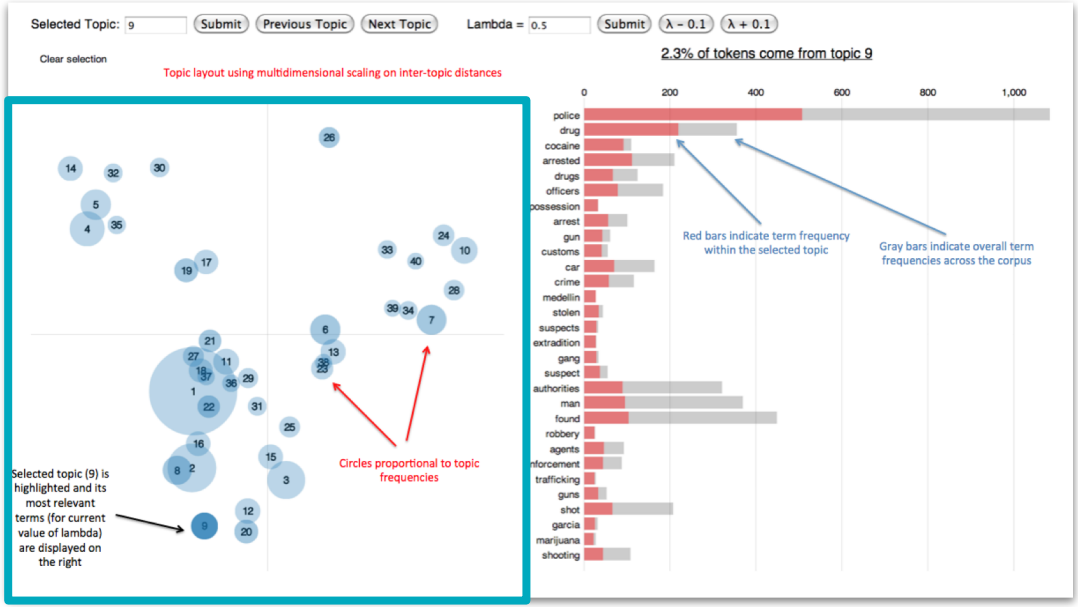


[1] Graham, S., Weingart, S., & Milligan, I. (2012). Getting started with topic modeling and MALLET. The Editorial Board of the Programming Historian.

[2] Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

Topic modeling interpretation

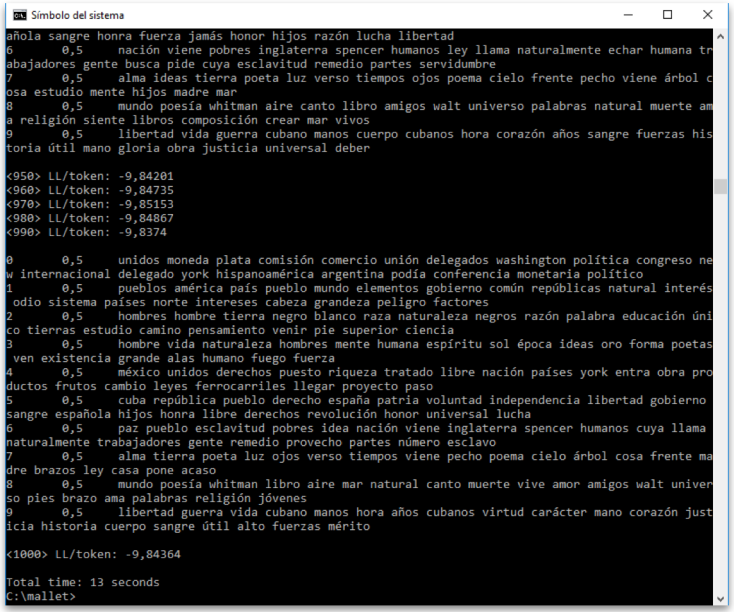


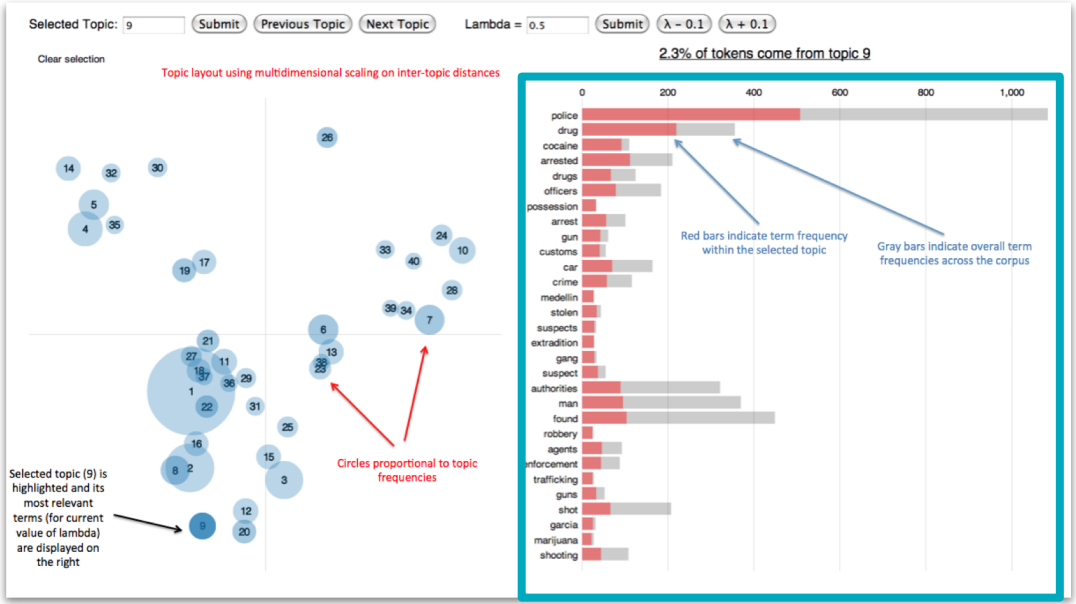


[1] Graham, S., Weingart, S., & Milligan, I. (2012). Getting started with topic modeling and MALLET. The Editorial Board of the Programming Historian.

[2] Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

Topic Modeling Interpretation





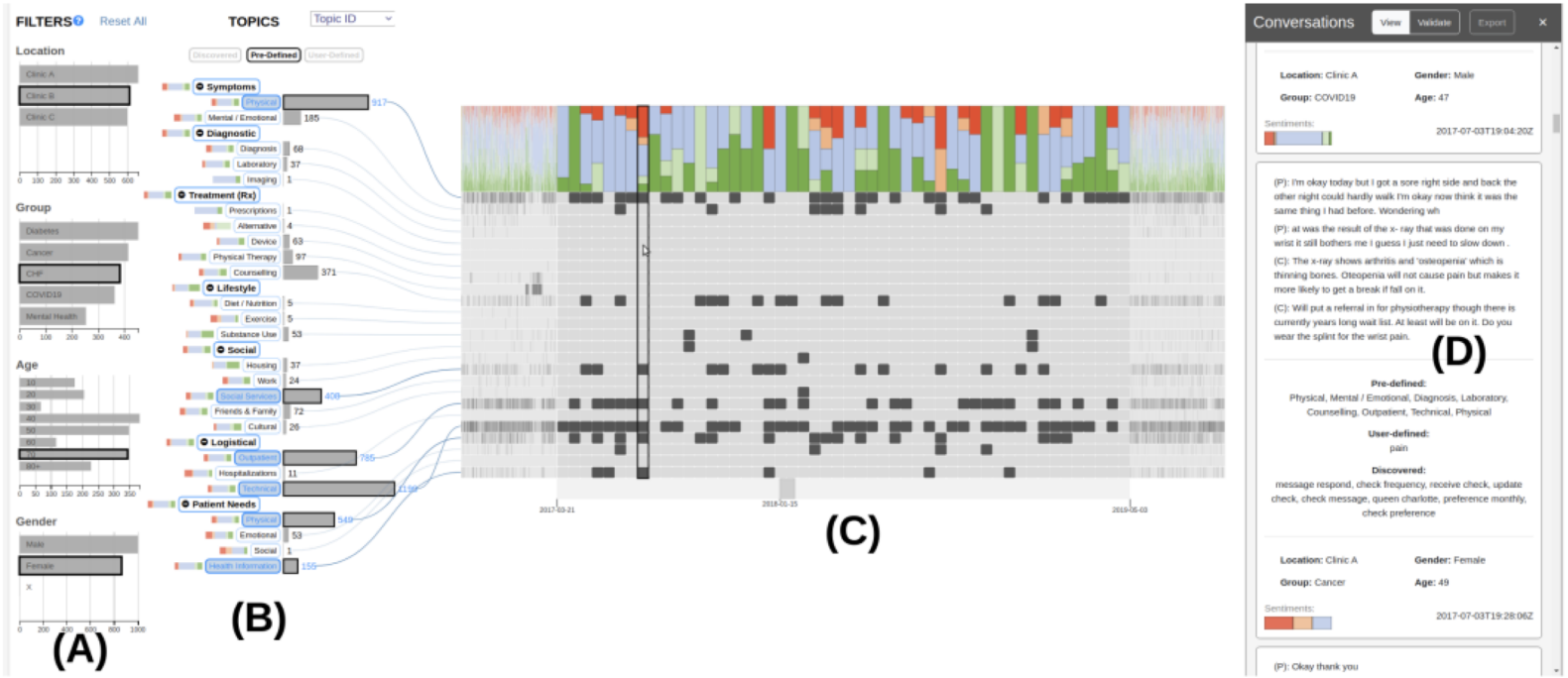
[1] Graham, S., Weingart, S., & Milligan, I. (2012). Getting started with topic modeling and MALLET. The Editorial Board of the Programming Historian.

[2] Sievert, C., & Shirley, K. (2014, June). LDAvis: A method for visualizing and interpreting topics. In Proceedings of the workshop on interactive language learning, visualization, and interfaces (pp. 63-70).

Finding mechanisms to improve these visual representations is still an open challenge [1]

[1] Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey. Multimedia Tools and Applications, 78(11), 15169-15211.

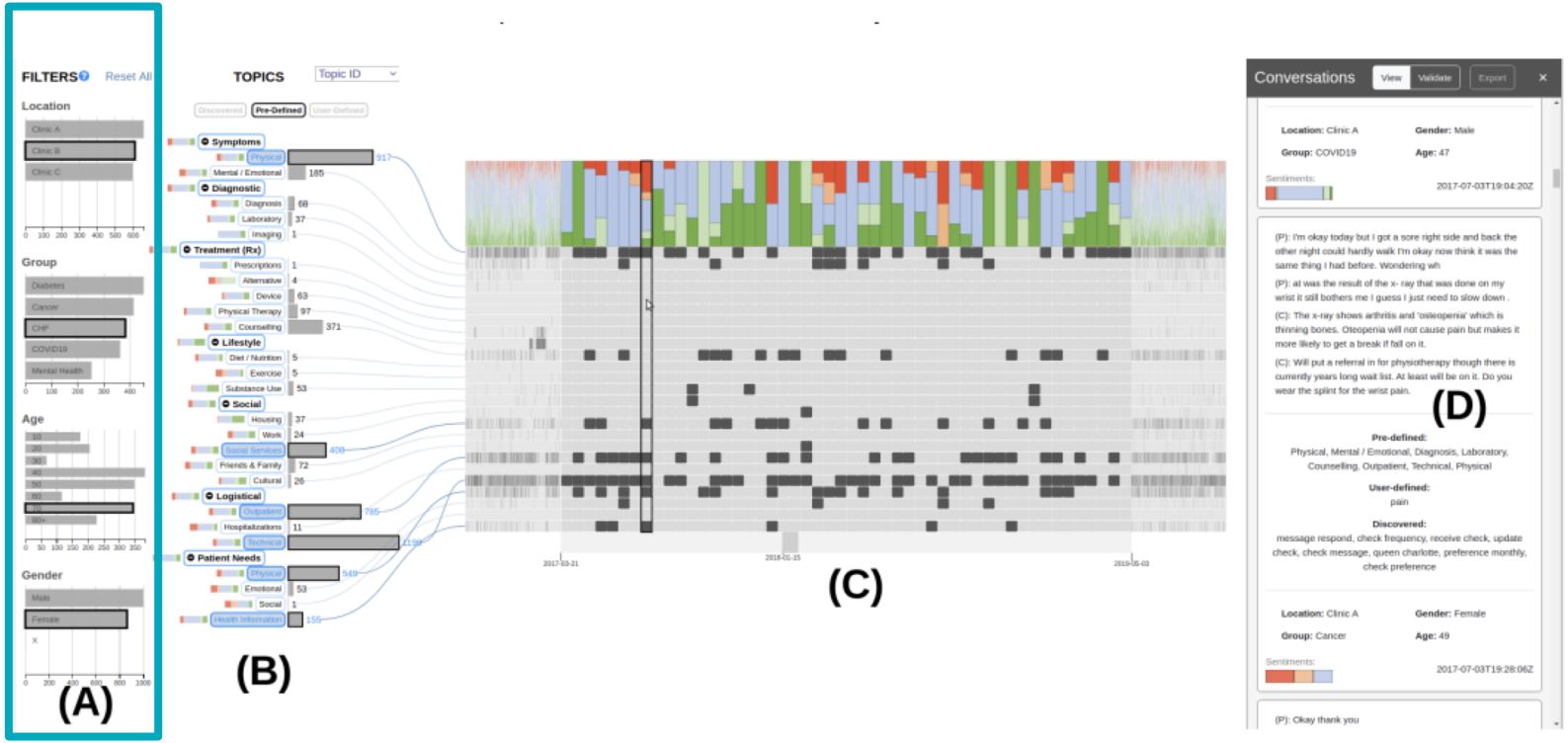
Recent research efforts



ConVIScope Views (see §4): Metadata (A), Topics (B), Analysis (C) and Conversations (D).

Li, R., Hoque, E., Carenini, G., Lester, R., & Chau, R. (2021). ConVIScope: Visual Analytics for Exploring Patient Conversations. *arXiv preprint arXiv:2108.13514*.

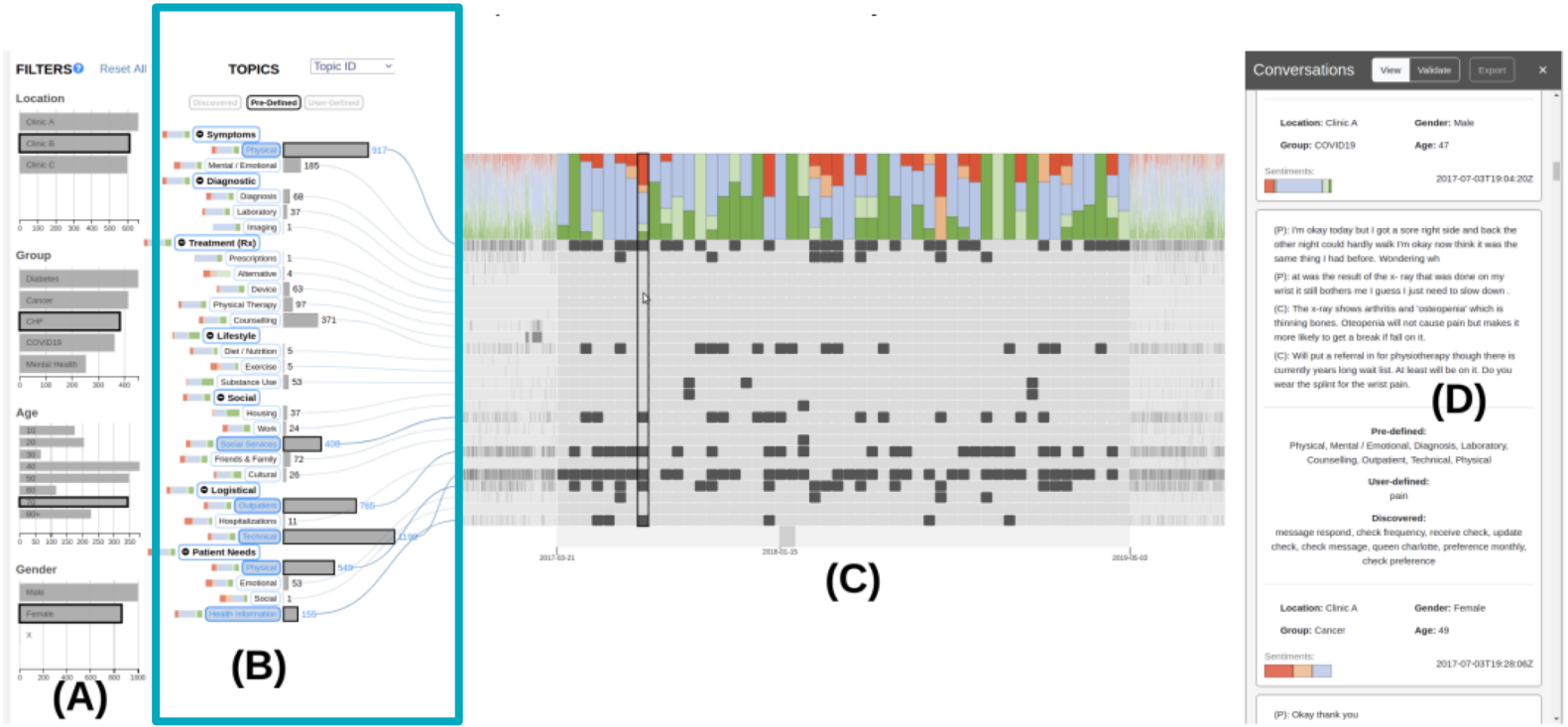
Recent research efforts



ConVIScope Views (see §4): Metadata (A), Topics (B), Analysis (C) and Conversations (D).

Li, R., Hoque, E., Carenini, G., Lester, R., & Chau, R. (2021). ConVIScope: Visual Analytics for Exploring Patient Conversations. *arXiv preprint arXiv:2108.13514*.

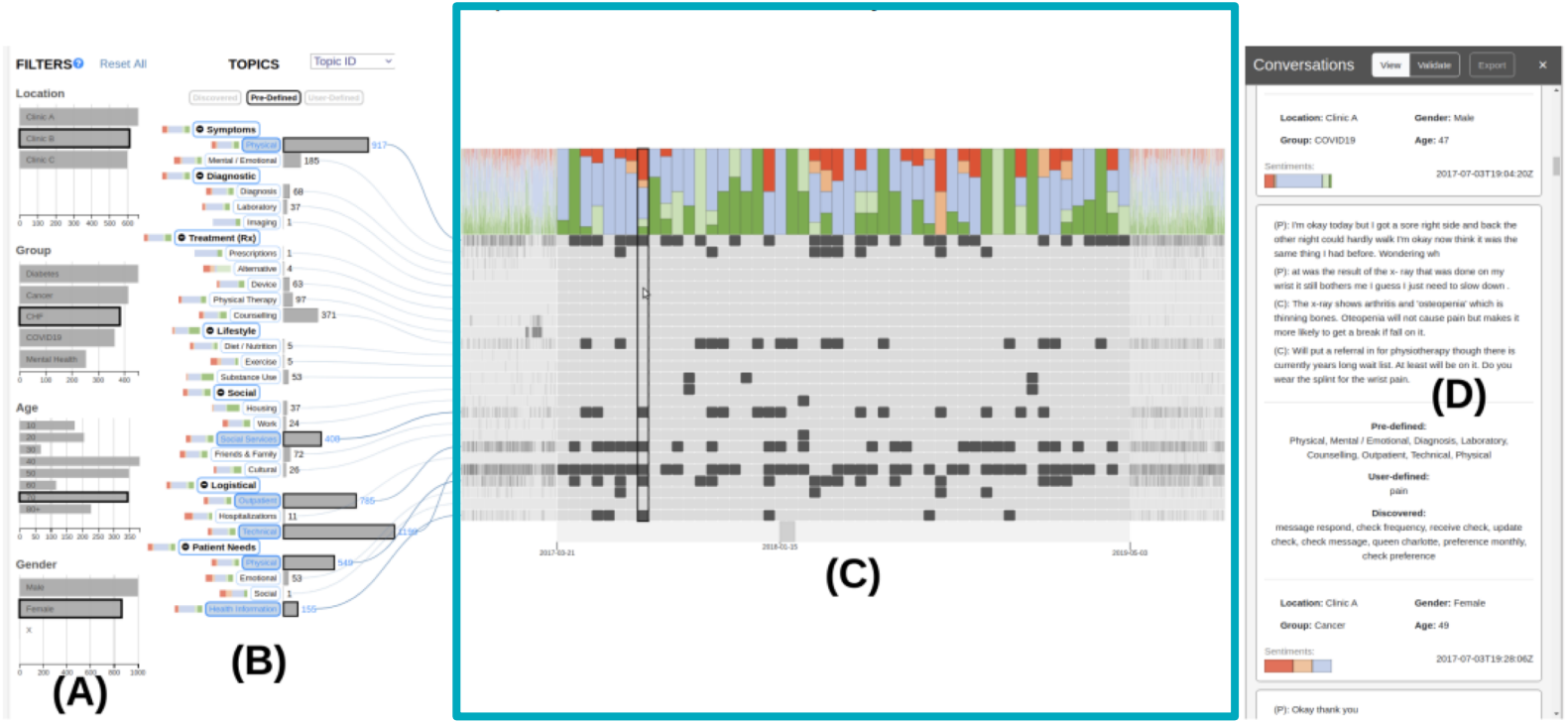
Recent research efforts



ConVIScope Views (see §4): Metadata (A), Topics (B), Analysis (C) and Conversations (D).

Li, R., Hoque, E., Carenini, G., Lester, R., & Chau, R. (2021). ConVIScope: Visual Analytics for Exploring Patient Conversations. *arXiv preprint arXiv:2108.13514*.

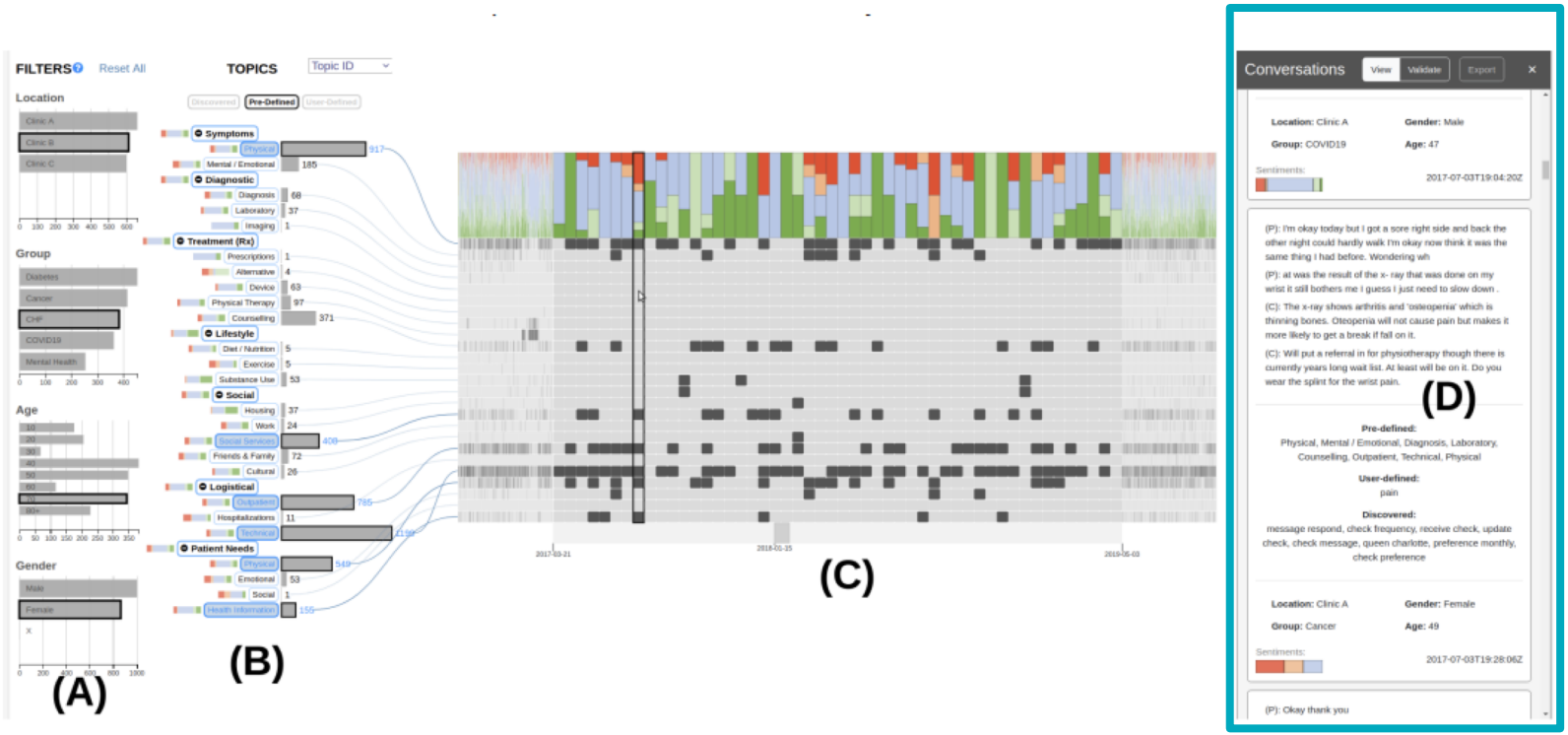
Recent research efforts



ConVIScope Views (see §4): Metadata (A), Topics (B), Analysis (C) and Conversations (D).

Li, R., Hoque, E., Carenini, G., Lester, R., & Chau, R. (2021). ConVIScope: Visual Analytics for Exploring Patient Conversations. *arXiv preprint arXiv:2108.13514*.

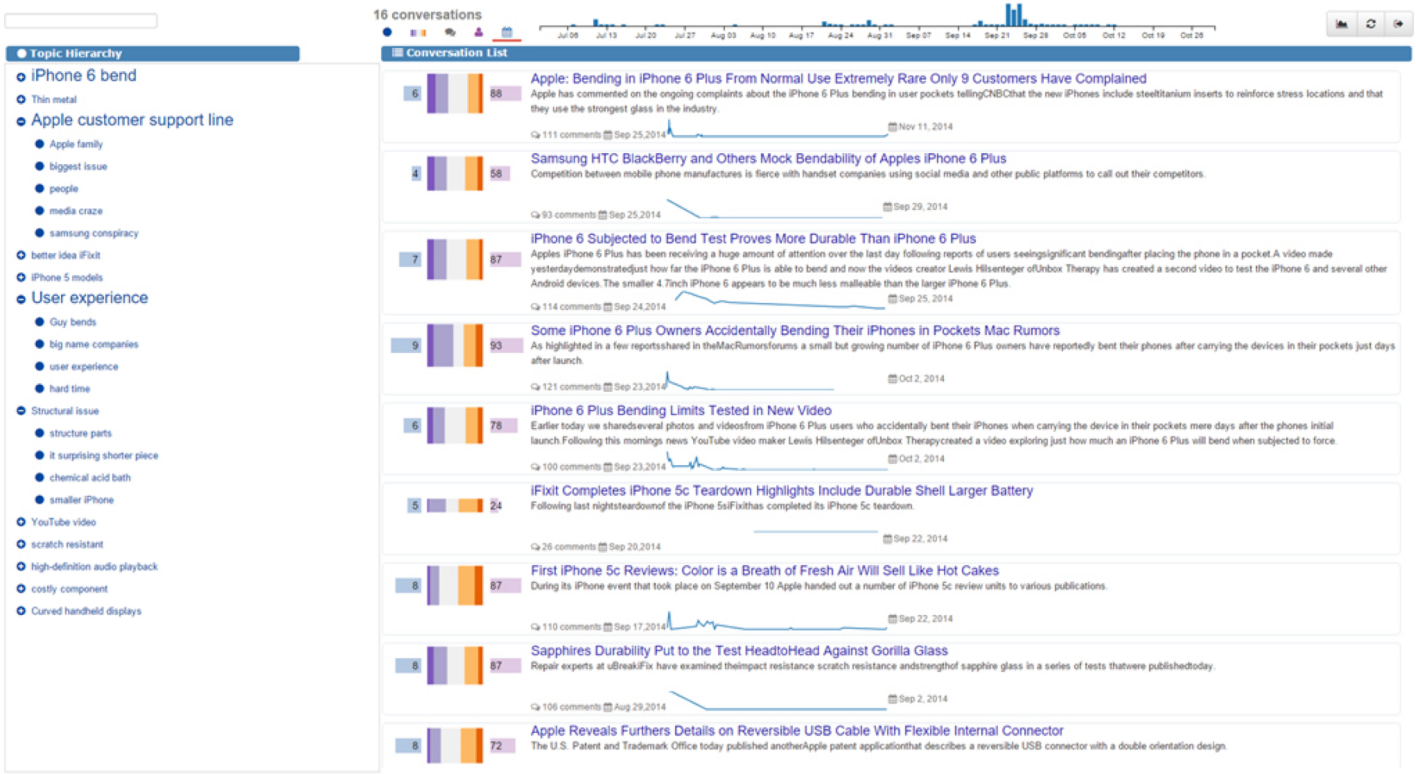
Recent research efforts



ConVIScope Views (see §4): Metadata (A), Topics (B), Analysis (C) and Conversations (D).

Li, R., Hoque, E., Carenini, G., Lester, R., & Chau, R. (2021). ConVIScope: Visual Analytics for Exploring Patient Conversations. *arXiv preprint arXiv:2108.13514*.

Recent research efforts



Hoque, E., & Carenini, G. (2019). Interactive topic hierarchy revision for exploring a collection of online conversations. Information Visualization, 18(3), 318–338. https://doi.org/10.1177/1473871618757228

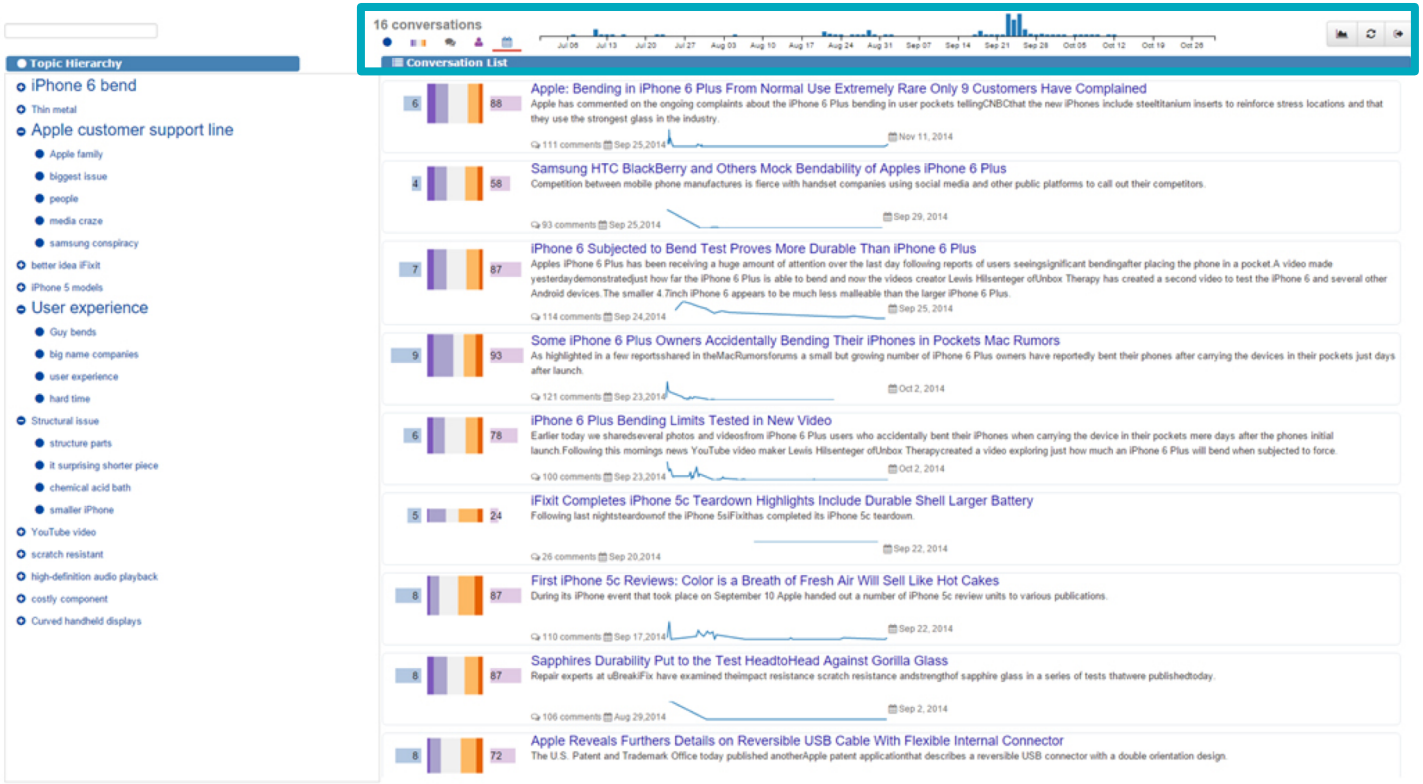
Recent research efforts

Hoque, E., & Carenini, G. (2019). Interactive topic hierarchy revision for exploring a collection of online conversations. Information Visualization, 18(3), 318–338. https://doi.org/10.1177/1473871618757228

Recent research efforts

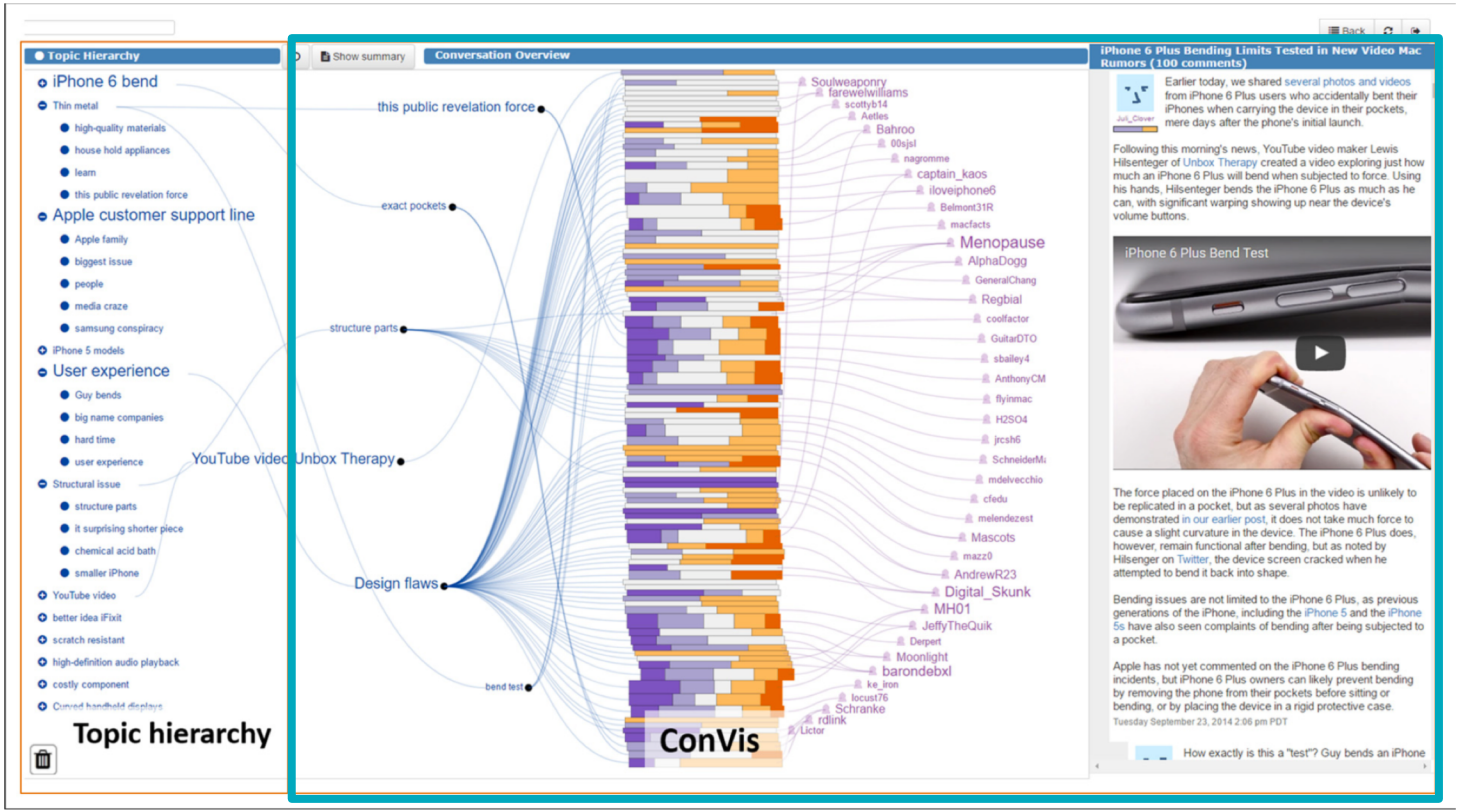
Hoque, E., & Carenini, G. (2019). Interactive topic hierarchy revision for exploring a collection of online conversations. Information Visualization, 18(3), 318–338. https://doi.org/10.1177/1473871618757228

Recent research efforts

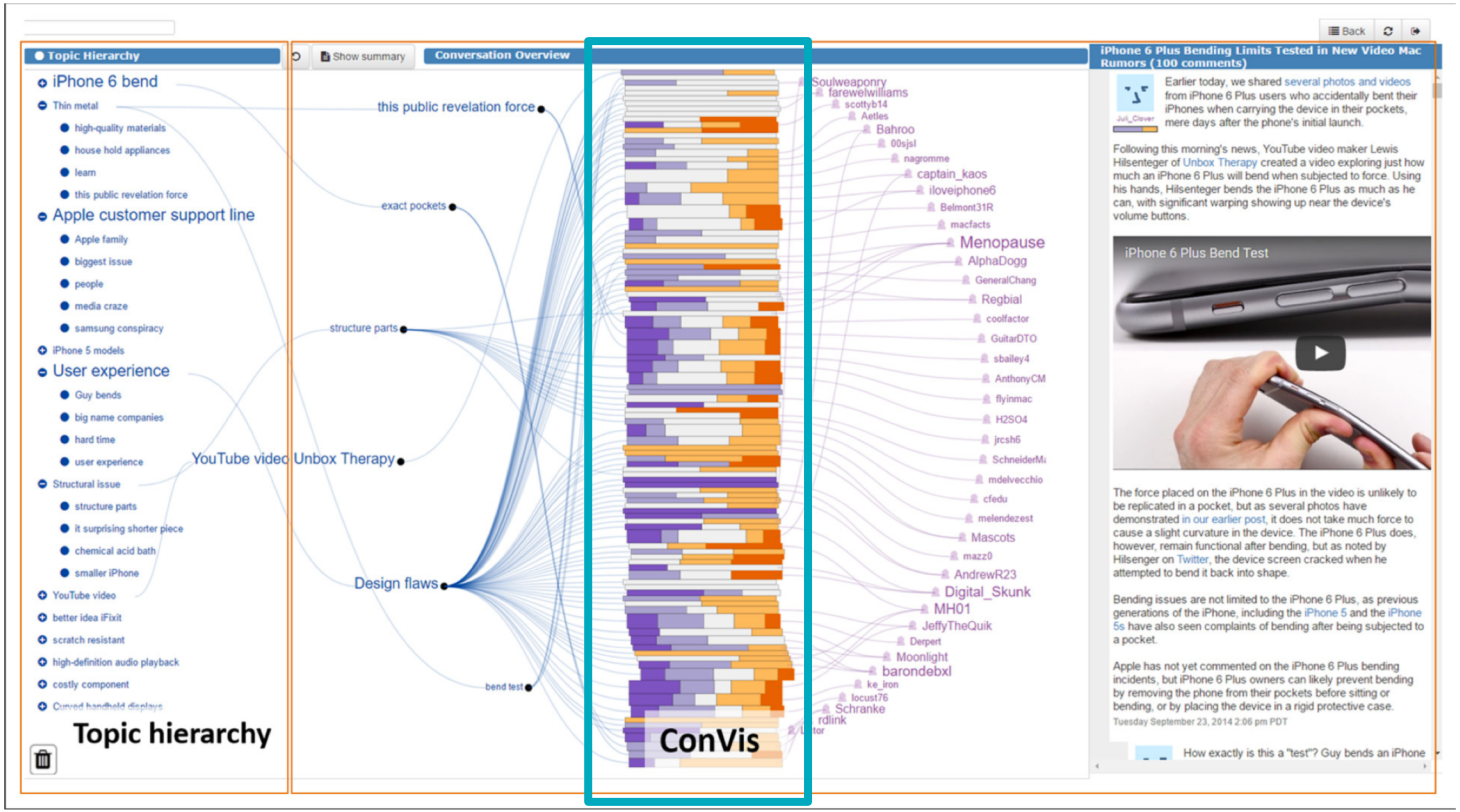


Hoque, E., & Carenini, G. (2019). Interactive topic hierarchy revision for exploring a collection of online conversations. Information Visualization, 18(3), 318–338. https://doi.org/10.1177/1473871618757228

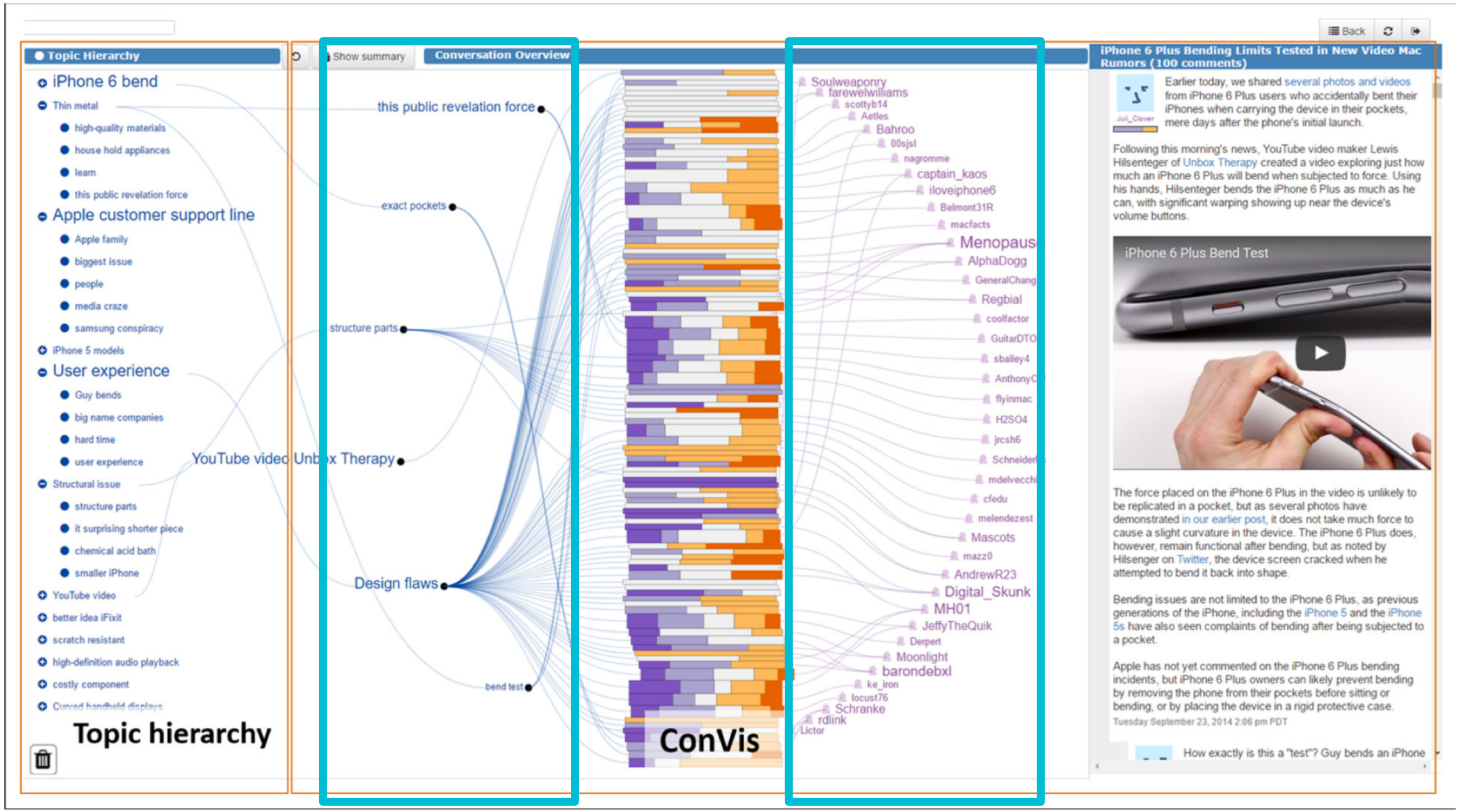
Recent research efforts



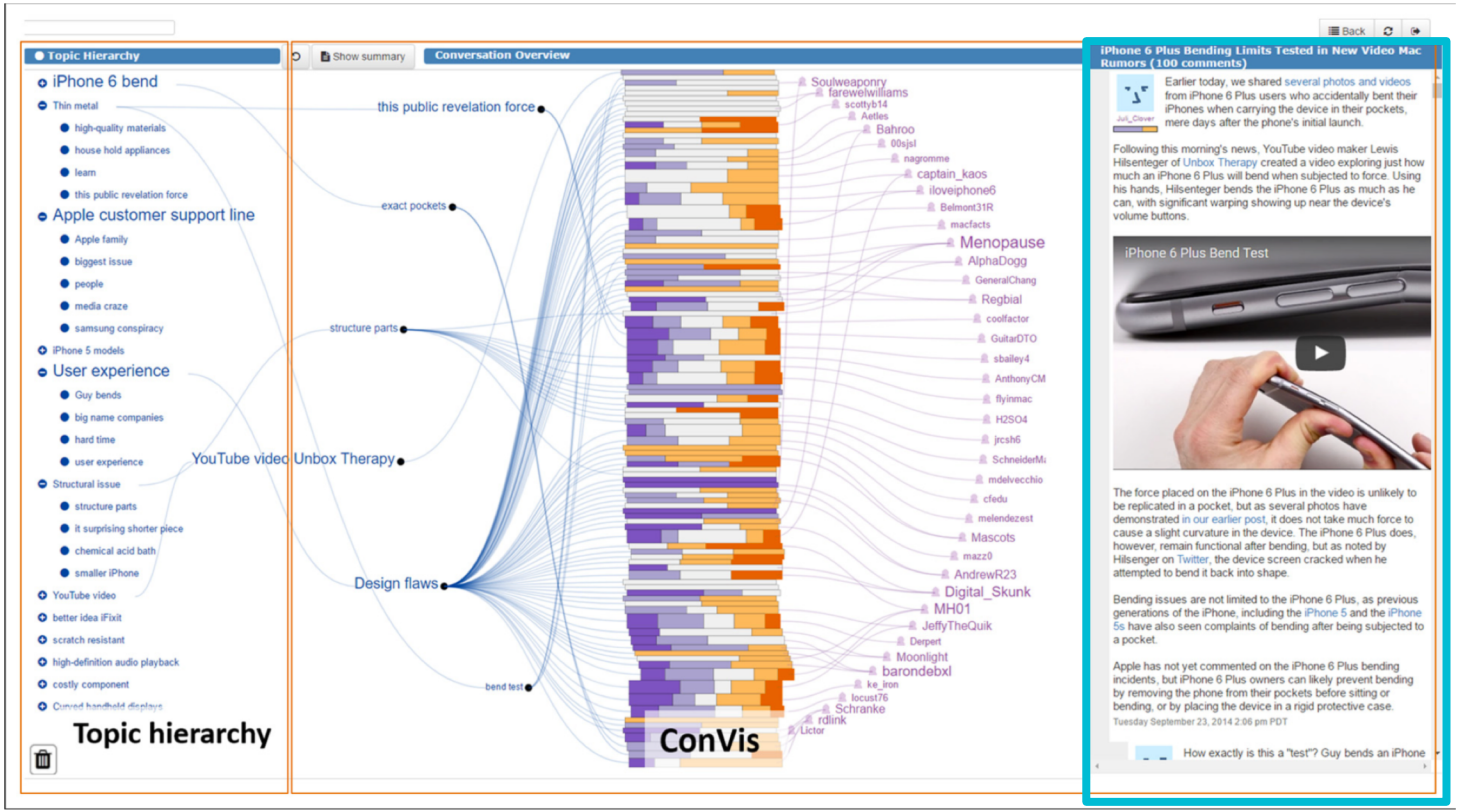
Recent research efforts



Recent research efforts



Recent research efforts



Datasets

Text:

- Ubuntu Dialogue Dataset
- Amazon QA dataset: https://jmcauley.ucsd.edu/data/amazon/qa/
- Alexa Topical Dataset (Alexa prize competition)
- Movie Dialogue Corpus
- More?
  - ParlAI Framework has incorporated a variety of text datasets for downstream NLP Tasks: https://github.com/facebookresearch/ParlAI

Multimodal (Image + Text) :

- 4chan https://zenodo.org/record/3606810#.YU-wSLhKiUk
- MMHS150K Dataset. Hate speech in Twitter. https://gombru.github.io/2019/10/09/MMHS/

Potential future work

- Provide support for Multimodal (Image - text) datasets
  - Adapt previous tools to support image content

- Seek options to:
  - Visualize emotions in addition to the sentiment analysis
  - Scalability: Visualize larger number of topics (e.g, 100 topics)

Questions?

Thank You! :)