

Topic Modeling and Network Visualization to Explore Patient Experiences

Annie T. Chen¹, Laura Sheble¹, Gabriel Eichler²

1. University of North Carolina, Chapel Hill; 2. PatientsLikeMe, Cambridge, MA

Background

Online support groups and health-related social networking sites can be powerful ways for patients to connect with each other and seek ways to improve their health. This content can also help clinicians and scientists understand patient experience and unmet needs. However, it can be difficult to navigate and make sense of the large volume of content.

Project rationale

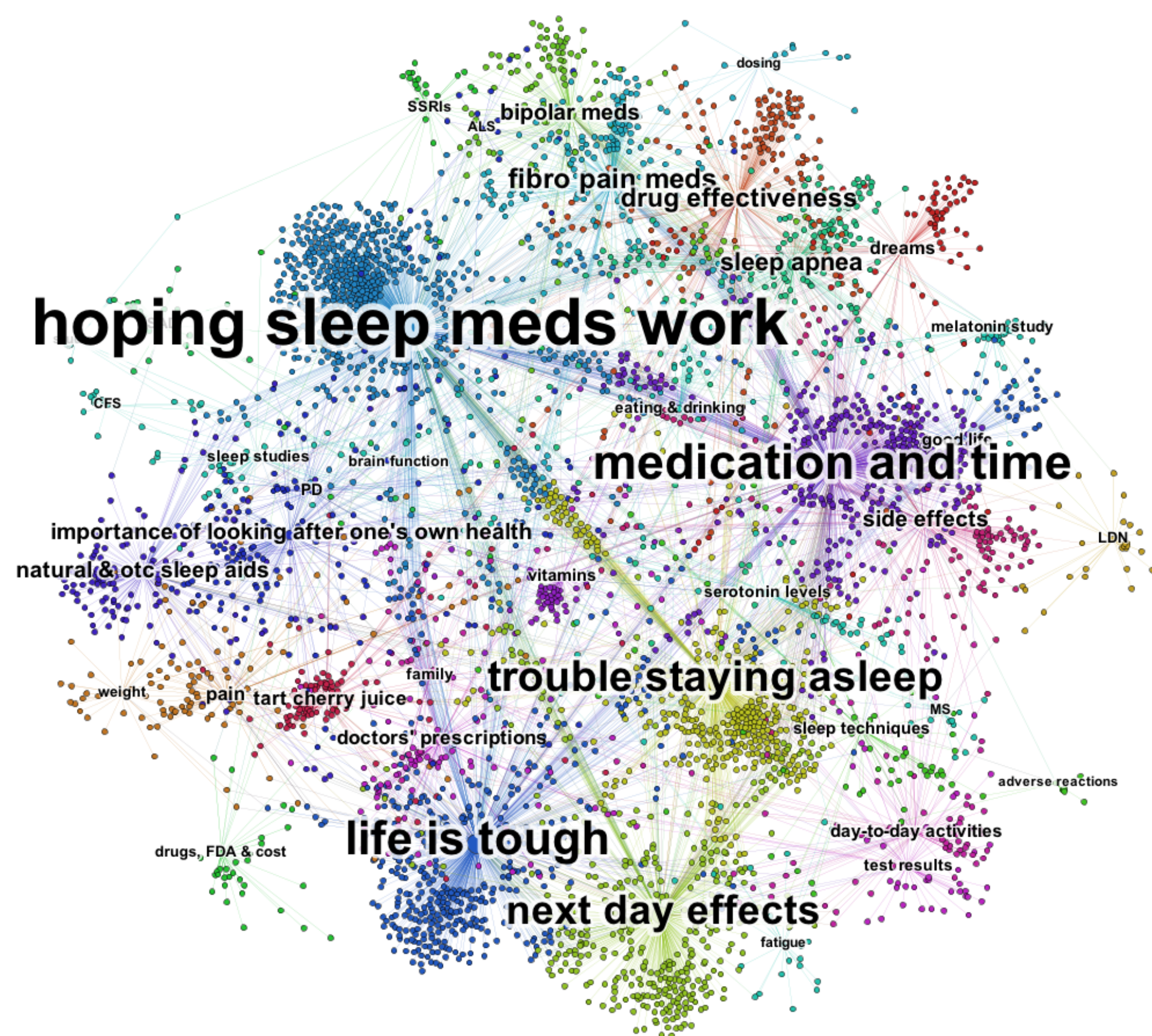
Topic modeling in combination with network visualization can be used to provide overviews and details of online discussion forum content through a series of visualizations. Decisions made when processing text and rendering visualizations affect content representation and subsequent interpretations. We present a case study of a lightweight approach to model health discussion content from PatientsLikeMe using open source tools.



THE UNIVERSITY
of NORTH CAROLINA
at CHAPEL HILL

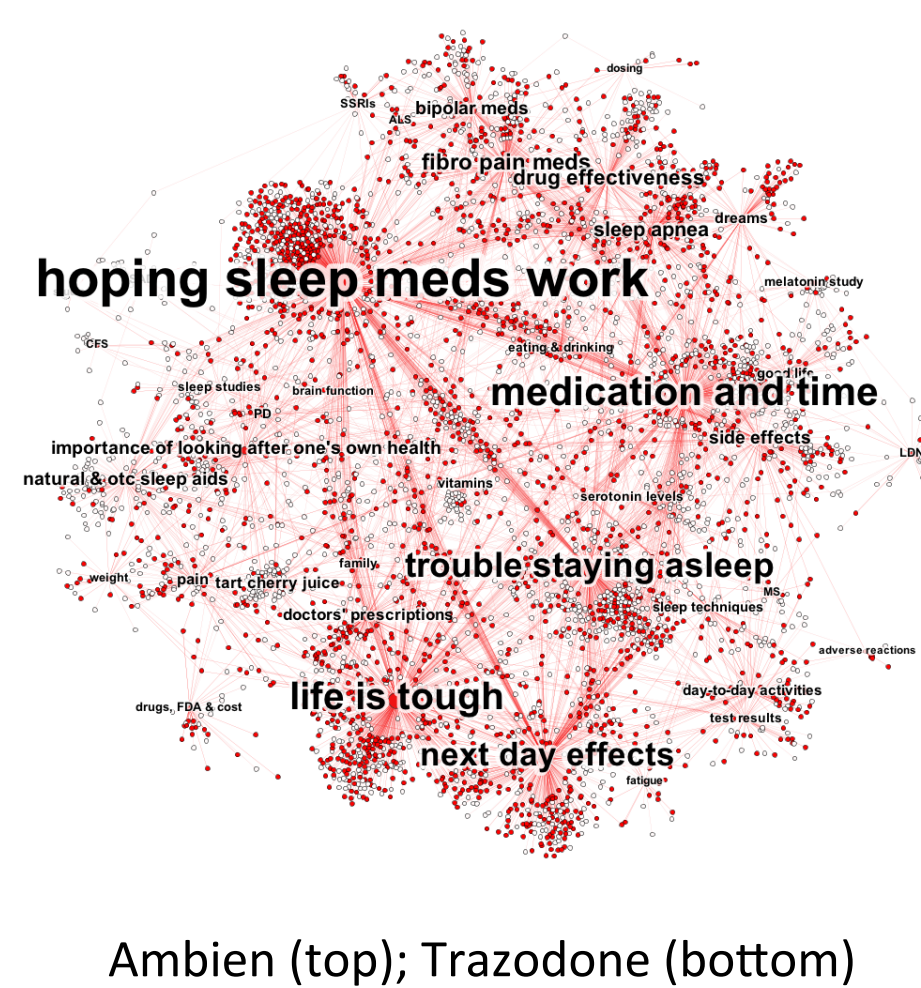
patientslikeme®

Visualizing patient experiences

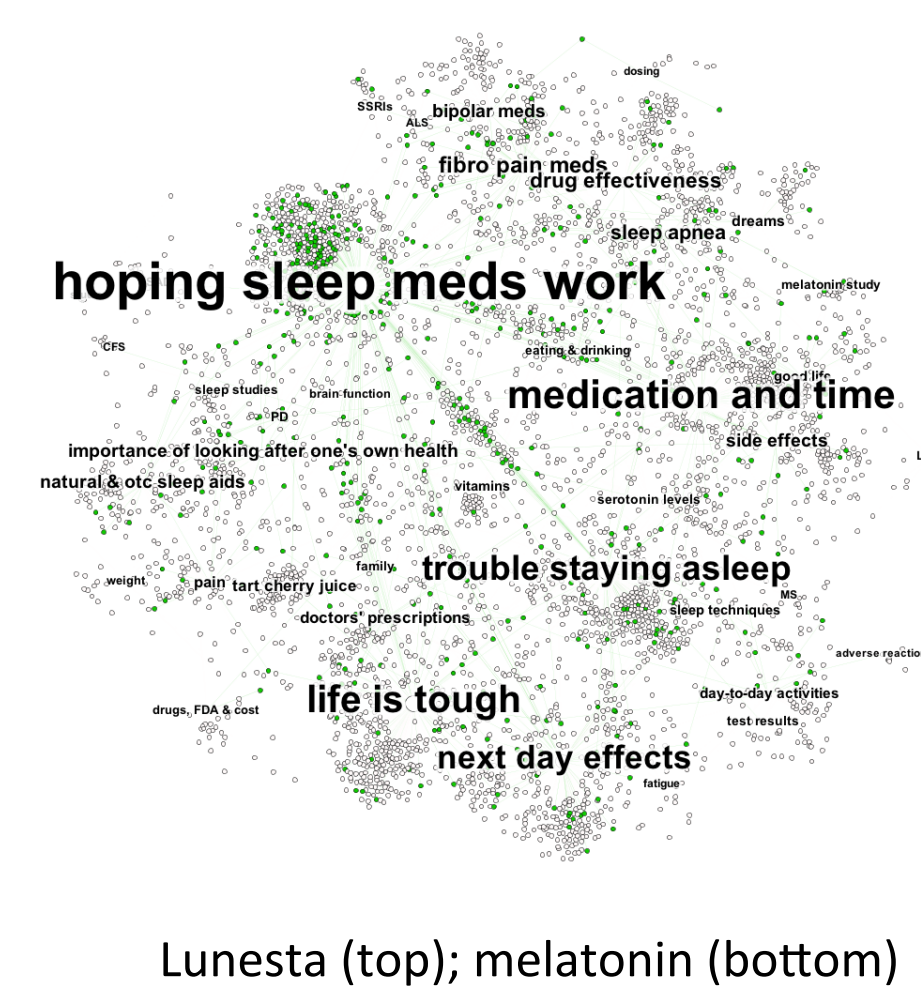


Overview: Composite network

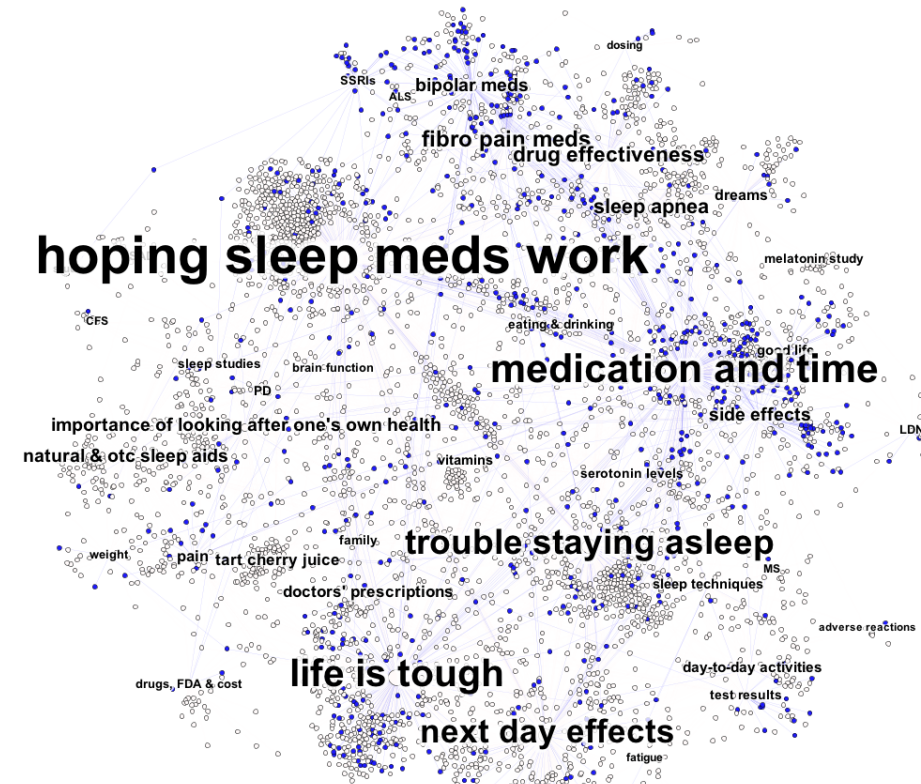
Corpus: forum posts mentioning four sleep aids: Ambien, Lunesta, Trazodone and melatonin.
Visualization: Bi-modal network of topics and forum posts
Text is proportionate to topic prevalence.



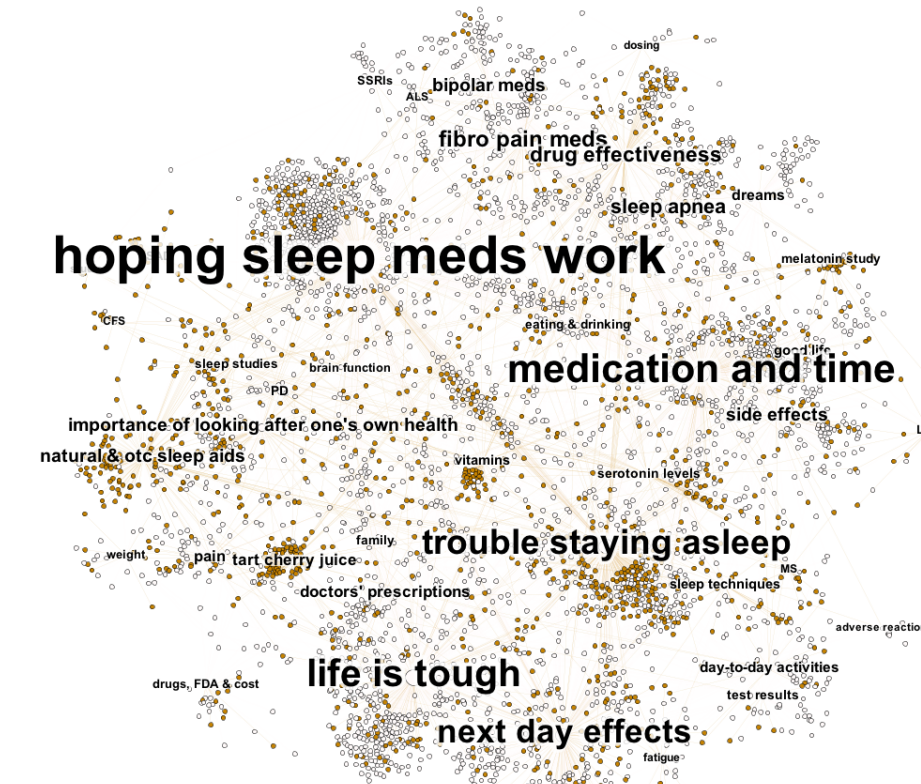
Ambien (top); Trazodone (bottom)



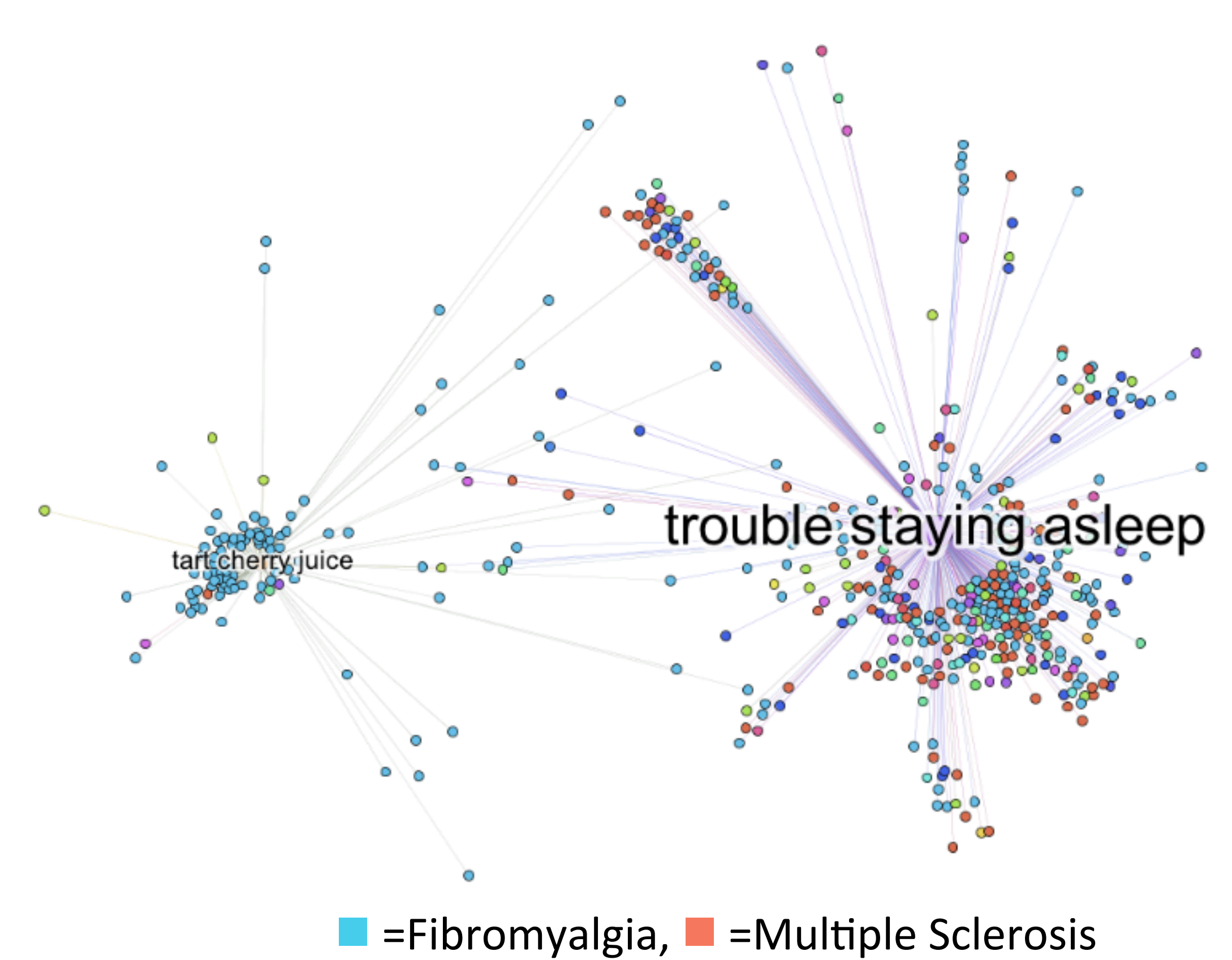
Lunesta (top); melatonin (bottom)



hoping sleep meds work



hoping sleep meds work



■ =Fibromyalgia, ■ =Multiple Sclerosis

Contrasting subsets

Graphs with highlighted subsets expose differences between the highlighted and overall discussion. When juxtaposed with other subsets, can illuminate differences between conversations.

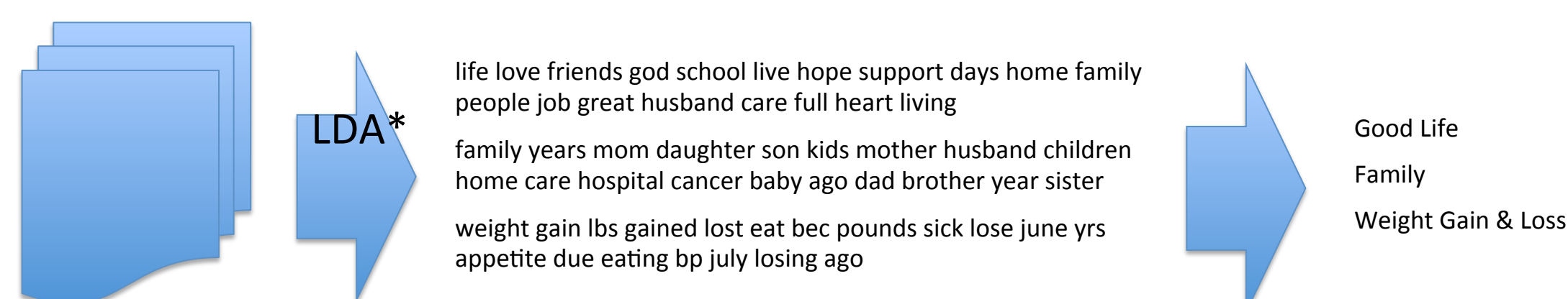
- Ambien: similar to those of the composite network
- Lunesta: concerned but hopeful
- Trazodone: timing of medication use, but also appears in discussions about mood disorders and fibromyalgia
- Melatonin is discussed in terms of its effectiveness

In more detail: Topic relationships

- The proportion of patients with Multiple Sclerosis (MS) who post about “Sleep Apnea” (32%) is nearly twice the proportion of MS patients who authored posts in the network overall (18%)
- Research indicates a higher prevalence of sleep apnea among those with MS [1]

Process

- Topic modeling can be used to quickly scan large volumes of text
- A “topic” is a cluster of words that frequently occurs together



- The algorithm provides estimates of the topic composition of each document
- This is then used to classify the posts by topic
- Gephi is used to visualize a bi-modal network of topic and post nodes.
- Document-associated metadata and network properties are used to generate visualizations that provide different perspectives

Decision points that may influence representation and subsequent interpretation include:

- The number of topics
- The threshold set for topic proportion that determines whether a post is classified as pertaining to a particular topic

Use scenarios

| Use segments | Example scenarios |
|-----------------------------|--|
| Patient education | Overview of and relationships between topics for navigation and/or education. Simplified and interactive visualizations could be presented online or integrated into PHRs |
| Informing clinicians | Alert services could provide information on patient concerns (e.g., side effects); and inform physicians about information patients consult online, but are unlikely to share with physicians |
| Clinical discovery | Discovery of novel associations (e.g., [2]) Illustrative patient experiences could be used to communicate research to broader audiences Less obtrusive data collection may enable greater participation, i.e., Citizen science [3] |
| Surveillance and monitoring | Overviews enable distant readings of topics associated with drugs and treatments and provide insight into patients perceptions, perhaps suggesting marketing and communication needs Contribute to diffusion analyses of health information across personal and institutional sources |

References

- [1] Braley TJ, Segal BM, Chervin RD. Sleep-disordered breathing in multiple sclerosis. *Neurology* 2012;79(9):929-36.
- [2] MacLean D, Seltzer M. Mining the web for medical hypotheses. *Proc HEALTHINF* 2011.
- [3] Swan M. Crowdsourced health research studies: An important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res* 2012;14:e46.