

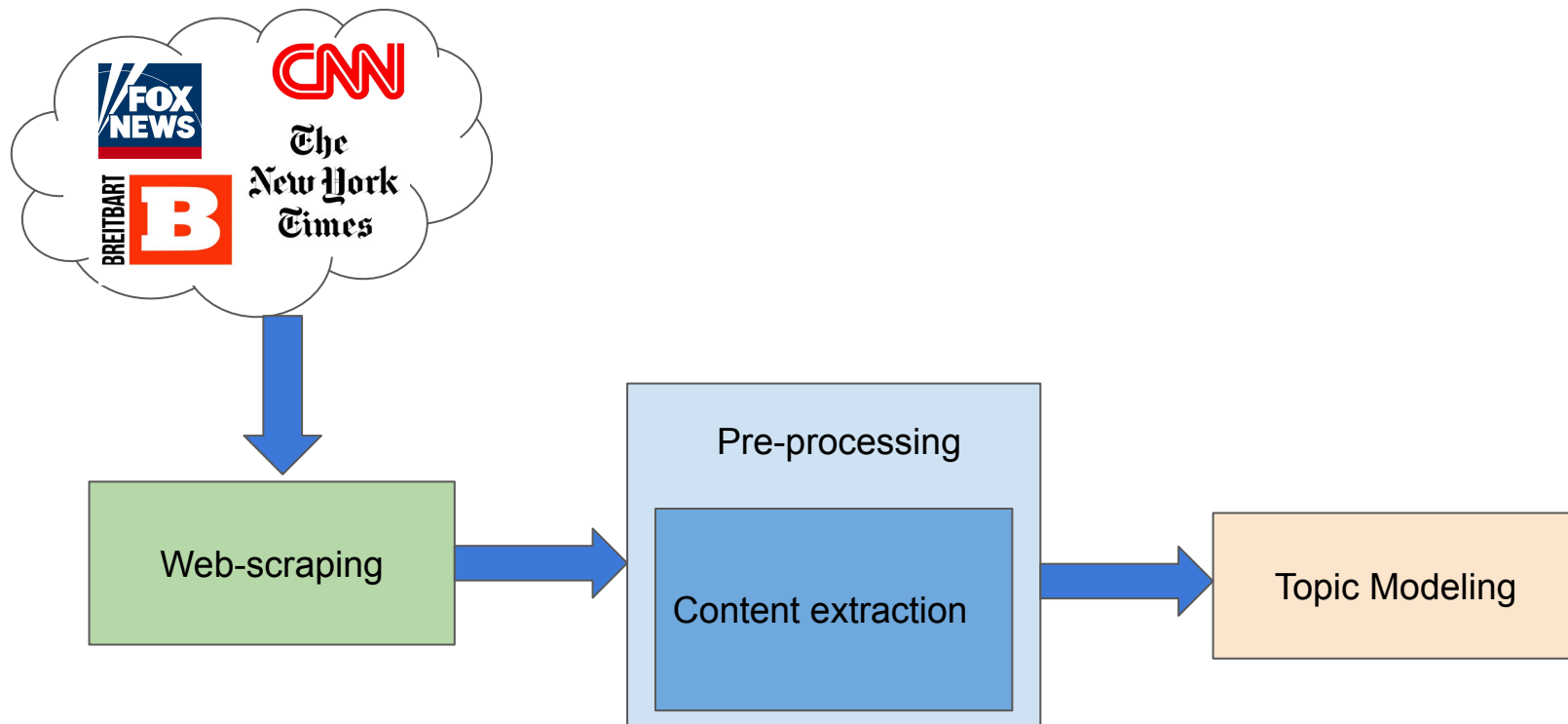
NLP on news

Duc Vu - Metis Chicago, Summer 2019

What kind of stories are covered in news?

- Can we find general topics ?
- Given a new article, can we identify the topic for that one?

Method

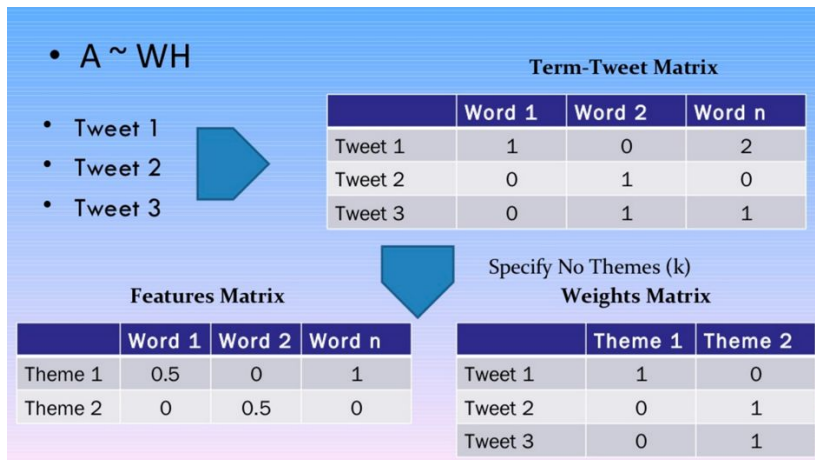


Dataset

- Collection from 4 main mainstream media
- Total 3020 articles
- From July 19, 2019 to August 19, 2019

Topic Modeling

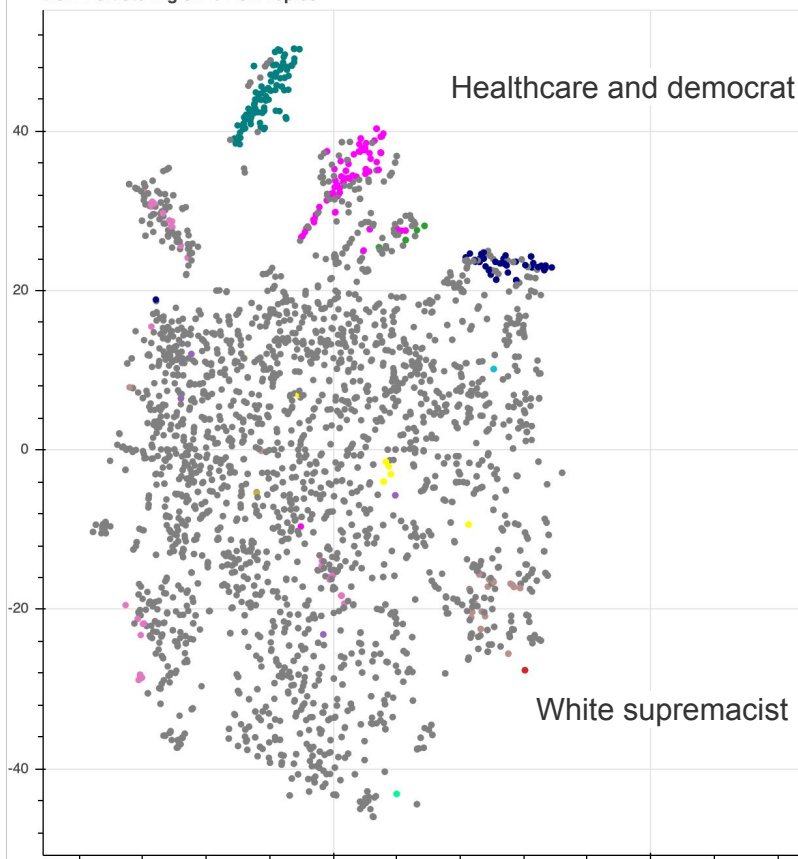
- Latent Semantic Analysis (LSA) is to take a matrix of documents and terms and decompose it into a separate document-topic matrix and a topic-term matrix.
- Non-negative Matrix Factorization



- Latent Dirichlet Allocation (LDA) uses dirichlet priors for the document-topic and word-topic distributions

Topic modeling with LSA and visualization in t-SNE

t-SNE Clustering of 20 LSA Topics



Healthcare and democrat

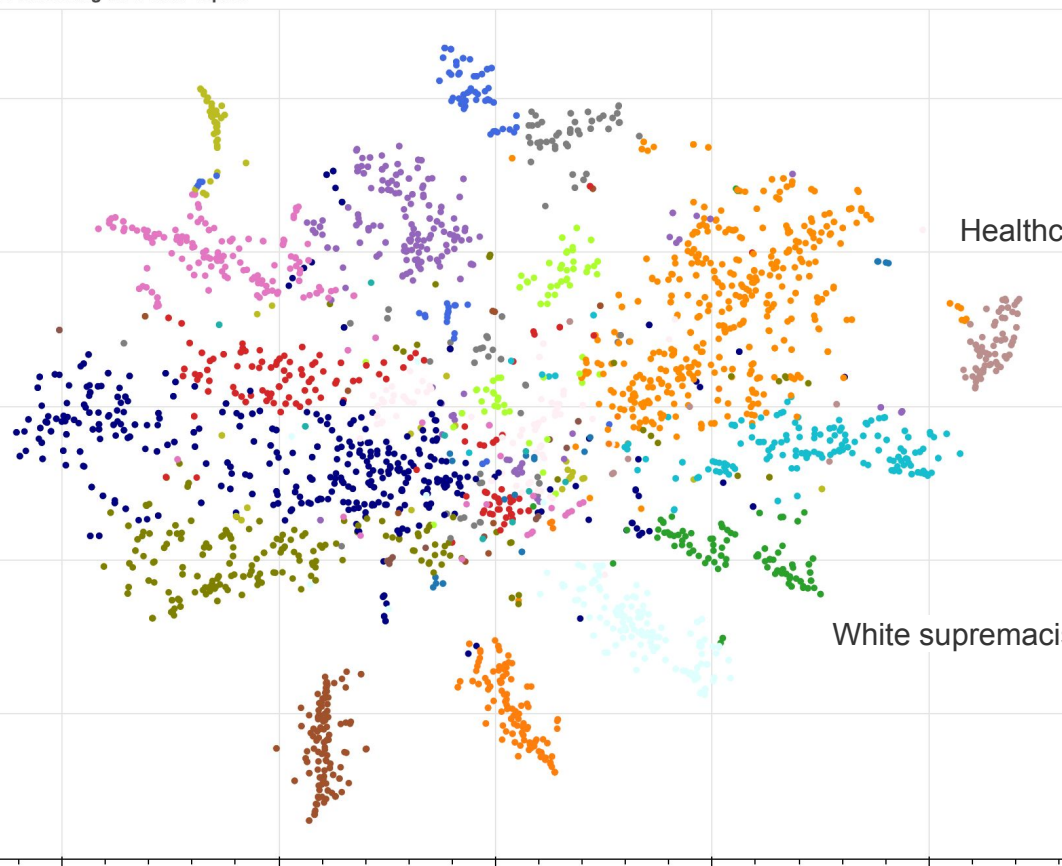
sanders vermont biden joseph warren senator democratic massachusetts bernie

White supremacist

guy dornic matt orchestrate

Topic modeling with NMF and visualization in t-SNE

t-SNE Clustering of 20 NMF Topics



Healthcare and democrat

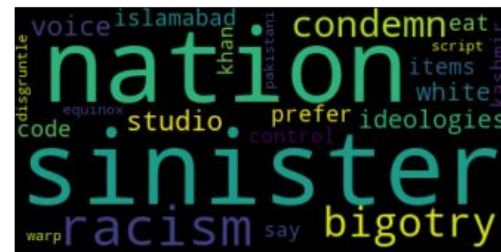
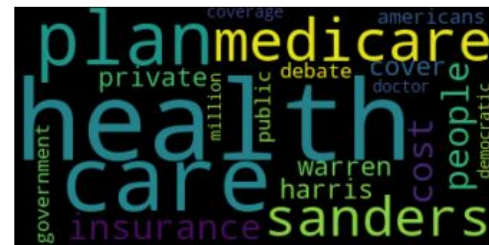
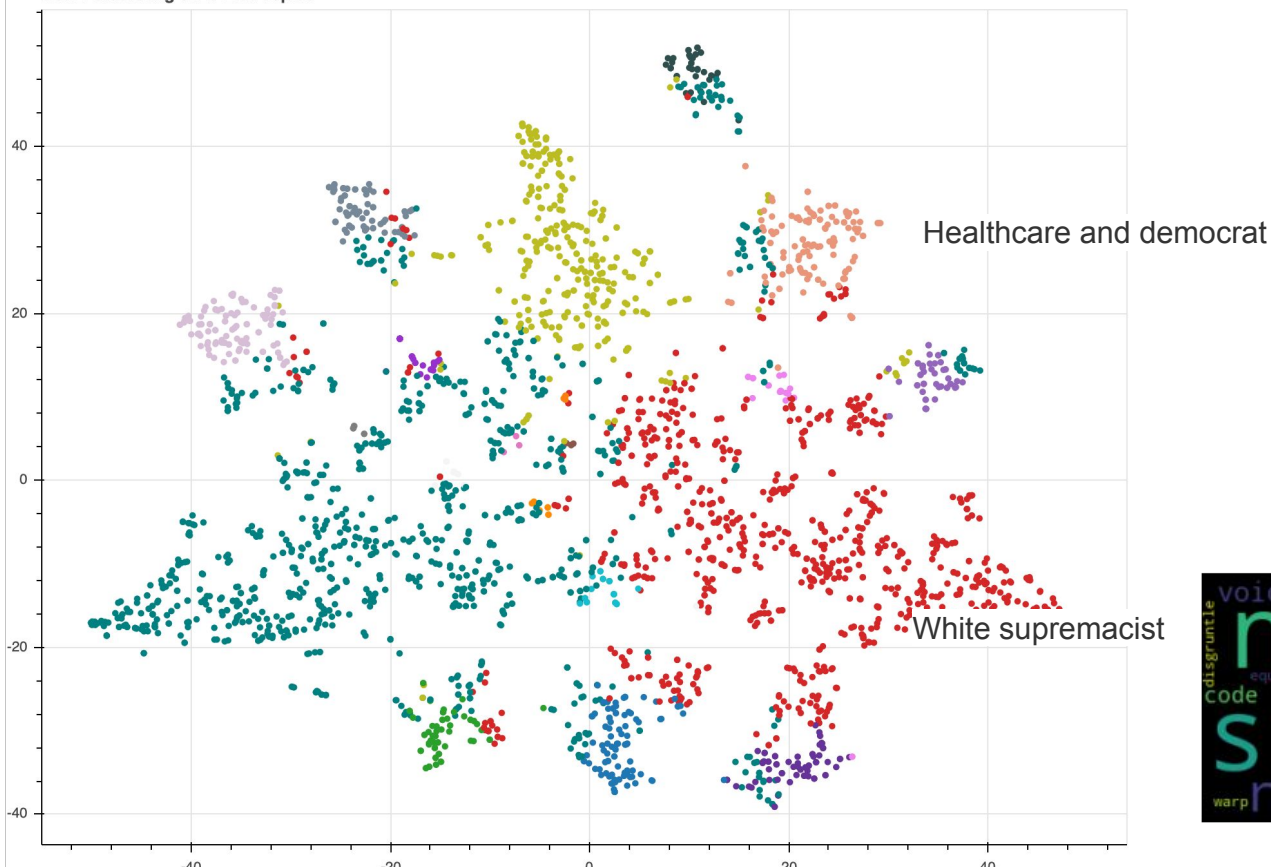


White supremacist



Topic modeling with LDA and visualization in t-SNE

t-SNE Clustering of 20 LDA Topics



Identify topic from new article

CNN politics 45 Congress Supreme Court Facts First 2020 Election

Scaramucci says he's recruiting ex-cabinet officials to speak out against Trump 02:17

Does that seem like someone who is totally unbothered by Scaramucci's public rejection of him in a variety of interviews and tweets this month?

Remember, always, that Trump's Twitter feed is a direct line to his mental and emotional state. Trump's allies -- in the White House and outside of it -- can say whatever they want, but what really matters is what Trump tweets. If he tweets about it, he cares about it. Bigly.

Which begs the question: Why does Trump care so much about Scaramucci's rebellion?

The answer is, as always with Trump, tied directly to how he views the world: There are only two kinds of people -- the pro-Trump ones and the anti-Trump ones. People who have been pro-Trump for a long time, or at least before he was elected president, he keeps closer and values more. Those are his OG people, the ones who have been with him from the jump.

It's not a big group -- and it's mostly family. But there are a few non-family members in there -- and Scaramucci was one. When those pro-Trumpers turn on him, Trump goes bananas.

Score: 0.5793884992599487

Topic: 0.019*"say" + 0.017*"trump" + 0.008*"people" + 0.008*"president" + 0.008*"state" + 0.005*"white" + 0.004*"issue" + 0.004*"support" + 0.004*"know" + 0.003*"unite"

Score: 0.2874836325645447

Topic: 0.018*"epstein" + 0.016*"say" + 0.009*"trump" + 0.006*"time" + 0.006*"president" + 0.005*"charge" + 0.005*"federal" + 0.005*"state" + 0.005*"know" + 0.004*"people"

Score: 0.09673985838890076

Topic: 0.032*"mueller" + 0.020*"say" + 0.011*"trump" + 0.008*"report" + 0.007*"president" + 0.007*"question" + 0.006*"hear" + 0.006*"july" + 0.005*"house" + 0.005*"robert"

Score: 0.023268956691026688

Topic: 0.013*"say" + 0.009*"space" + 0.007*"time" + 0.006*"moon" + 0.005*"book" + 0.004*"people" + 0.004*"know" + 0.004*"black" + 0.004*"think" + 0.004*"right"

Takeaway

- **LSA, NMF and LDA are not able to automatically determine the number of topics**
- **Recommend stories based on collaborative reader behavior**



- **Classify the bias news**

Appendix