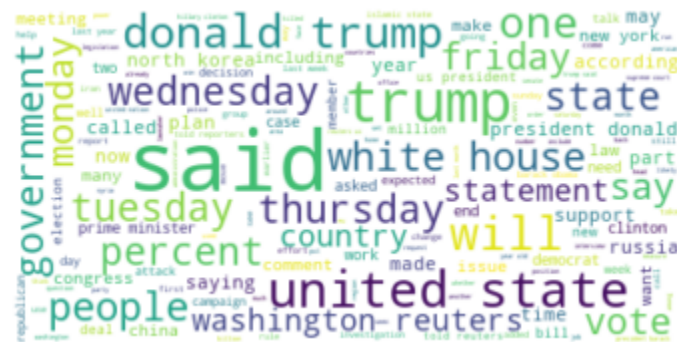# Word Cloud

- **Fake News Dataset**
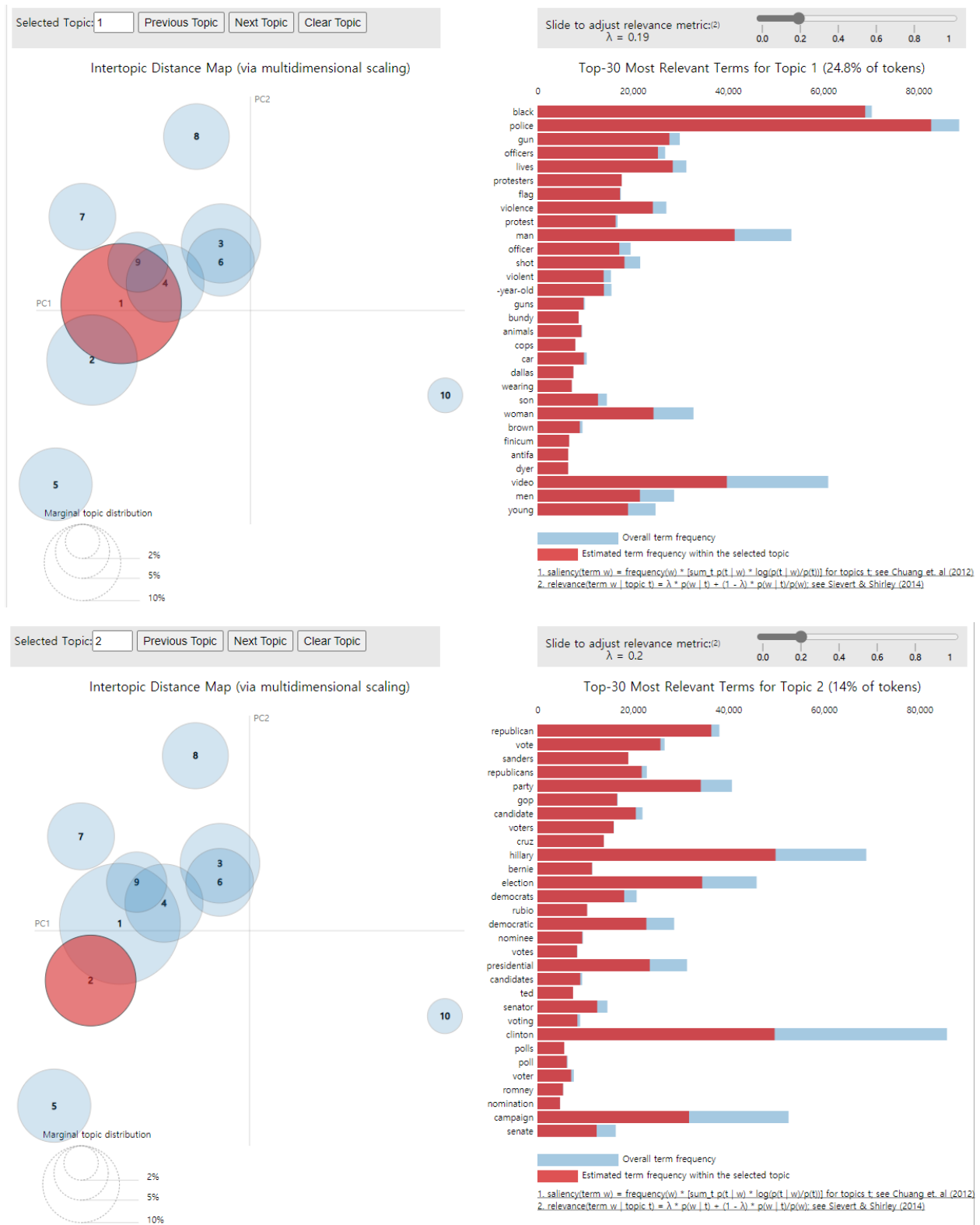

- **True News Dataset**

LDA model

- Fake News Dataset

Intertopic Distance Map (via multidimensional scaling)

PC2

PC1

Marginal topic distribution

2%

5%

10%

Slide to adjust relevance metric:(2)
λ = 0.19

0.0 0.2 0.4 0.6 0.8 1

Top-30 Most Relevant Terms for Topic 1 (24.8% of tokens)

0   20,000   40,000   60,000   80,000

- black
- police
- gun
- officers
- lives
- protesters
- flag
- violence
- protest
- man
- officer
- shot
- violent
- -year-old
- guns
- bundy
- animals
- cops
- car
- dallas
- wearing
- son
- woman
- brown
- finicum
- antifa
- dyer
- video
- men
- young

■ Overall term frequency
■ Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Intertopic Distance Map (via multidimensional scaling)

PC2

PC1

Marginal topic distribution

2%

5%

10%

Slide to adjust relevance metric:(2)
λ = 0.2

0.0 0.2 0.4 0.6 0.8 1

Top-30 Most Relevant Terms for Topic 2 (14% of tokens)

0   20,000   40,000   60,000   80,000

- republican
- vote
- sanders
- republicans
- party
- gop
- candidate
- voters
- cruz
- hillary
- bernie
- election
- democrats
- rubio
- democratic
- nominee
- votes
- presidential
- candidates
- ted
- senator
- voting
- clinton
- polls
- poll
- voter
- romney
- nomination
- campaign
- senate

■ Overall term frequency
■ Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

- True News Dataset

Topic 1

Selected Topic: 1 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric: (2)
λ = 0.28
0.0 0.2 0.4 0.6 0.8 1

Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Relevant Terms for Topic 1 (15.6% of tokens)

korea
north
nuclear
islamic
iran
china
military
syria
forces
syrian
un
iraq
korean
missile
kurdish
militants
iraqi
weapons
beijing
troops
war
sanctions
pyongyang
south
ministry
chinese
fighters
japan
air
civilians

Marginal topic distribution
2%
5%
10%

■ Overall term frequency
■ Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

Topic 4

Selected Topic: 4 | Previous Topic | Next Topic | Clear Topic

Slide to adjust relevance metric: (2)
λ = 0.28
0.0 0.2 0.4 0.6 0.8 1

Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Relevant Terms for Topic 4 (11.7% of tokens)

billion
hurricane
companies
irma
puerto
million
storm
rico
market
industry
brazil
investment
company
prices
financial
hun
investors
temer
sen
banks
fund
workers
firms
products
caribbean
labor
bank
per
infrastructure
growth

Marginal topic distribution
2%
5%
10%

■ Overall term frequency
■ Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)