Does Gender Matter in the News? Detecting and Examining Gender Bias in News Articles

Jamell Dacon[1] and Haochen Liu[1]

1: Michigan State University, East Lansing, MI, USA

WWW 2021

Presenter: Jamell Dacon
daconjam@msu.edu

Online News Recommendation

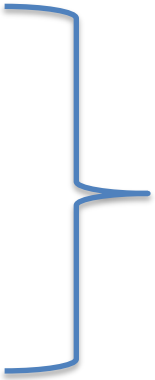
Benefits:

- Constantly updating
- No **costs**
- Ability to discuss the news with peers quickly
- A great number of news choice

Online News Recommendation

Cons:

- The existence of several viewpoints
- Ideological bias
- Coverage bias
- Selection bias
- Presentation bias



Media Bias

Speculative Sentences/Clauses

Title example(s):

- *"Women who want to succeed at work should shut up - while men who want the same should keep talking, research says"*
- *"Men have been promoted 3 times more than women during the pandemic, study finds"*

- *"Women in the workplace."*

Goal

- To detect and examine the phenomenon of implicit and explicit gender bias in the abstracts of news articles

- To gain a sense of understanding of the gender representation in the news by examining the relationships between social hierarchies and news content.

Motivation

- To identify how several forms of media bias (i.e., **coverage bias**, **selection bias**, and **presentation bias**) contribute to the problem of gender bias.

Contributions

- We construct two large benchmark datasets: (1) possessive (gender-specific and gender-neutral) nouns dataset and (2) attribute (career-related and family-related);

- We conduct three large scale analyses to detect and examine gender biases in distribution, content, and labeling and word choice;

- We introduce a series of measurements to better understand gender representation in news articles quantitatively and qualitatively.

Datasets

- **MIND**: The MIND dataset was collected from the Microsoft News website, randomly sampled news for 6 weeks from October 12th to November 22th, 2019. We obtained 96,112 abstracts.

- **NCD**: The NCD dataset was collected from Huffpost. The news articles were sampled from news headlines from the year 2012 to 2018 totaling in 202,372 news articles. We obtained 200,853 abstracts.

Bias in Gender Distribution

Table 1: Gender distribution test on the news datasets.

Dataset	Abstracts	Category	M	F
MIND	96,112	18	22,760	6,817
NCD	200,853	41	21,250	15,856

Table 2: Illustration of four intersecting career words (prefixes) across the two datasets for *females* compared to their respective *male* counter parts. The results are reported in terms of no. of gender-specific career words mentioned in each dataset per gender with their corresponding *Woman/Man* suffixes.

	MIND		NCD	
Career Words	# Man	# Woman	# Man	# Woman
Spokes	192	121	112	42
Congress	191	49	94	25
Chair	225	20	102	5
Business	66	3	31	4

Jamell Dacon (daconjam@msu.edu),
https://www.cse.msu.edu/~daconjam

Bias in Content

Table 3: The average number of the attribute words observed in each news abstract.

	Dataset			
	MIND		NCD	
	M	F	M	F
Diversity (%)	23.68	7.09	10.22	7.88
Avg. Career Words per Abstract	**0.1258**	0.0907	**0.0657**	0.0554
Avg. Family Words per Abstract	0.6406	**0.6954**	0.4431	**0.4723**

Bias in Wording

- **Sentiment Analysis**

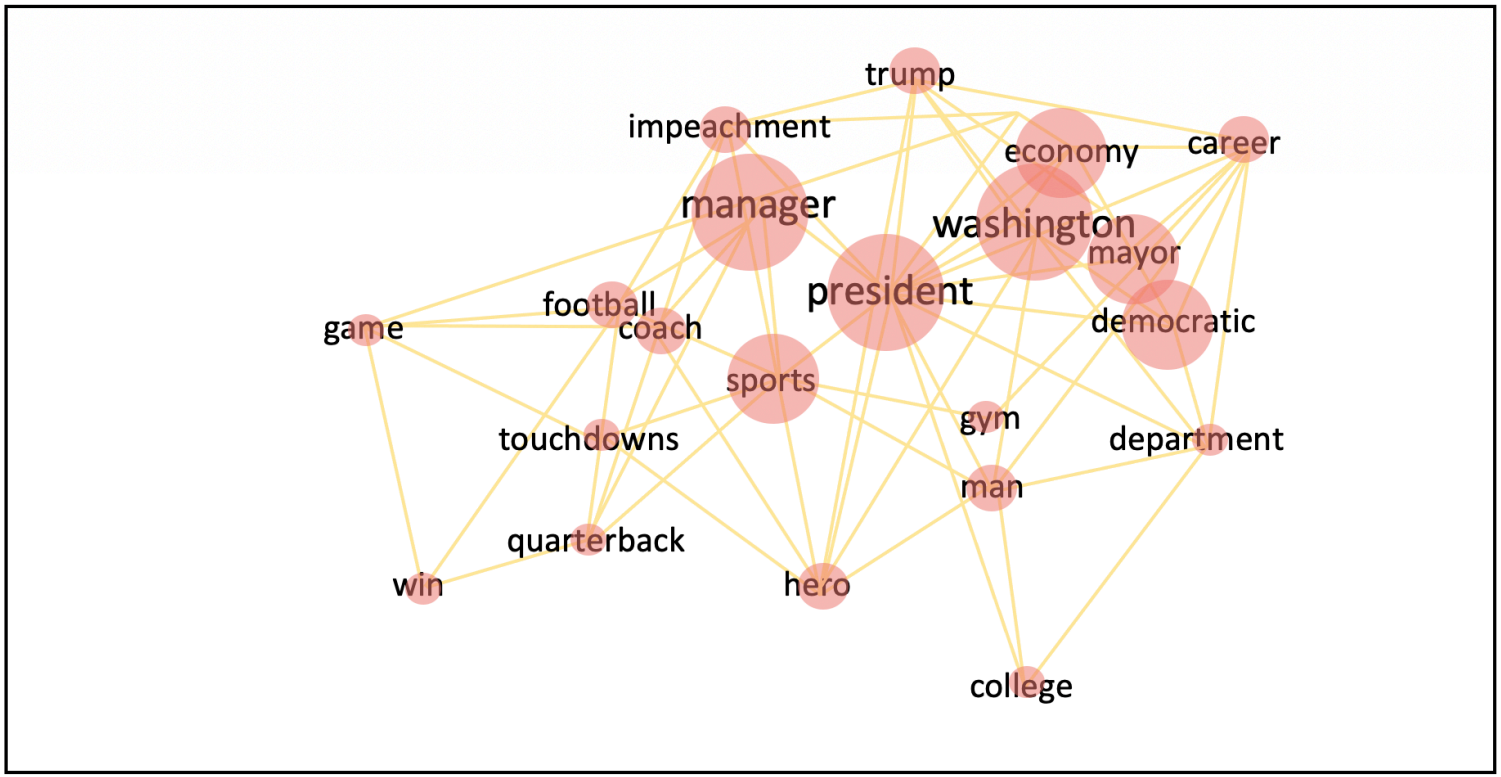
A popular, well known sentiment analysis tool, VADER to measure the sentiment of each news abstracts. VADER computes a normalized, weighted *compound* score of each word in a sentence by summing their valence scores between -1 (being extremely negative) and +1 (being extremely positive).

- **Centering Resonance Analysis**

A network word-based method that constructs a network representation of correlated words. This method exploits rich textual data and expresses the intention and meandering behaviors of authors (or columnists). To determine textual "centers" without the use of dictionaries i.e., ***identifying the most central nouns that mostly contribute to the meaning of a document or corpora***.

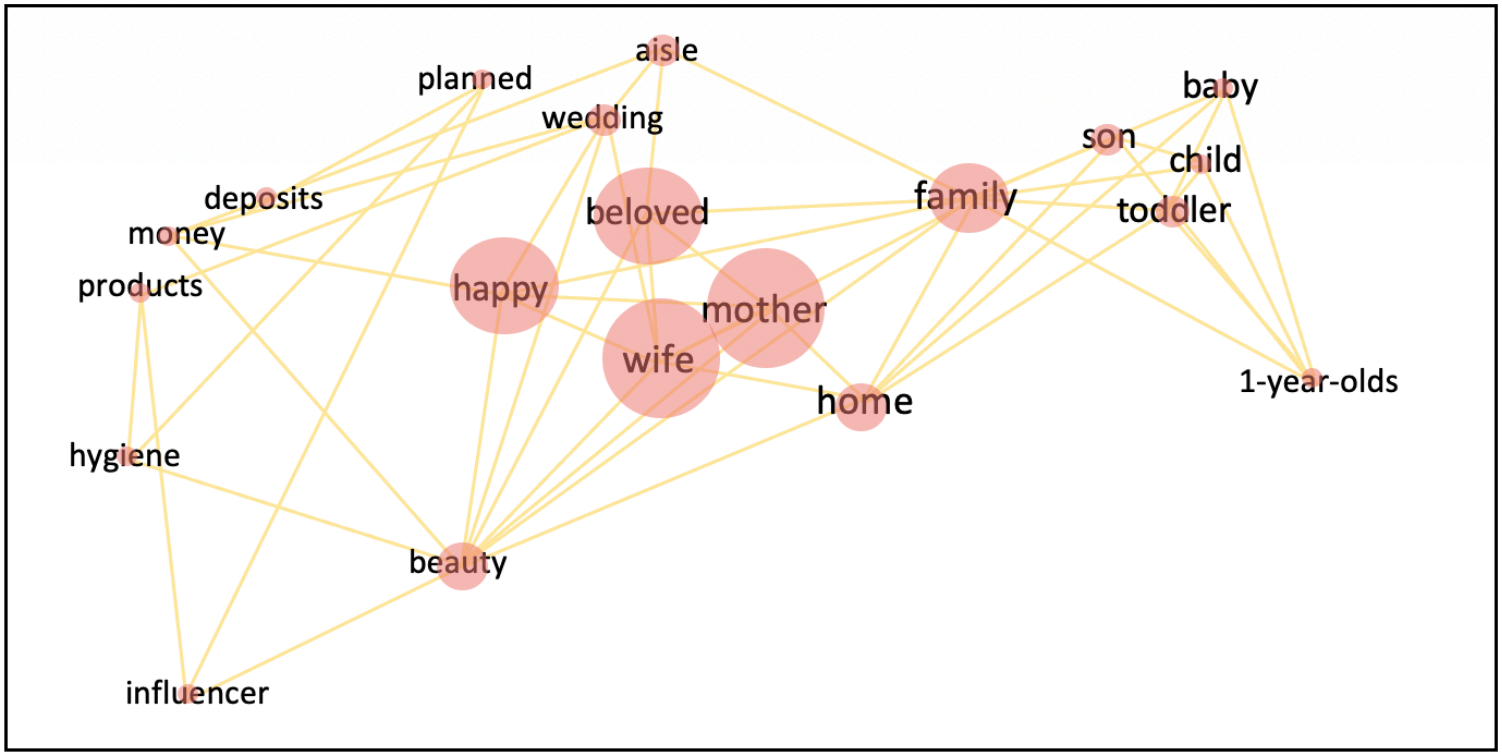
Bias in Wording

Figure 1: (a) The resulting CRA network for the top 20 nouns in the M tagged abstracts.



Bias in Wording

Figure 1: (b) The resulting CRA network for the top 20 nouns in the F tagged abstracts.



Conclusion

We have investigated that gender bias in media appears in different forms.

- Bias in gender distribution across all news categories and exploring the top four intersecting career words (prefixes) for ***females*** compared to their respective ***male*** counterparts
- Bias in content in terms of attribute words
- Bias in wording by constructing CRA networks for the top 20 most central nouns for both gender-tagged abstracts. Each graph illustrated the compound nouns that contributed the most tagged abstracts.

Thank You!

Code and Data: https://github.com/daconjam/Detecting_Gender_Bias

DSE Lab@MSU: http://dse.cse.msu.edu/