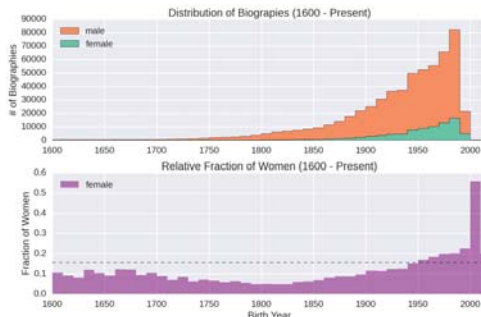




Gender Inequality in Wikipedia

Data

- Three external reference lists of notable people: Freebase [1], Human Accomplishment [2] and Pantheon [3]
- The DBpedia 2014 dataset.
- Inferred gender for Wikipedia biographies by [4]

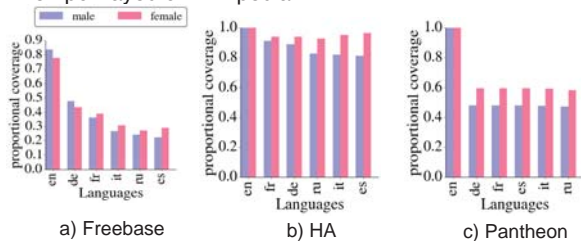


Methods

- Coverage and Visibility Bias:** Who is covered by Wikipedia? Which articles make it to the front page?
- Topical and Linguistic Bias:** Differences in the word usage in articles about men and women? Linguistic Intergroup Bias theory [5] suggests that we generalize positive aspects of people in our ingroup and negative aspects of people in our outgroup.
- Structural and Metadata Bias:** Differences in the metadata, link structure and network position of articles about men and women.

Coverage Bias

Differences between the number of notable women and men portrayed on Wikipedia.

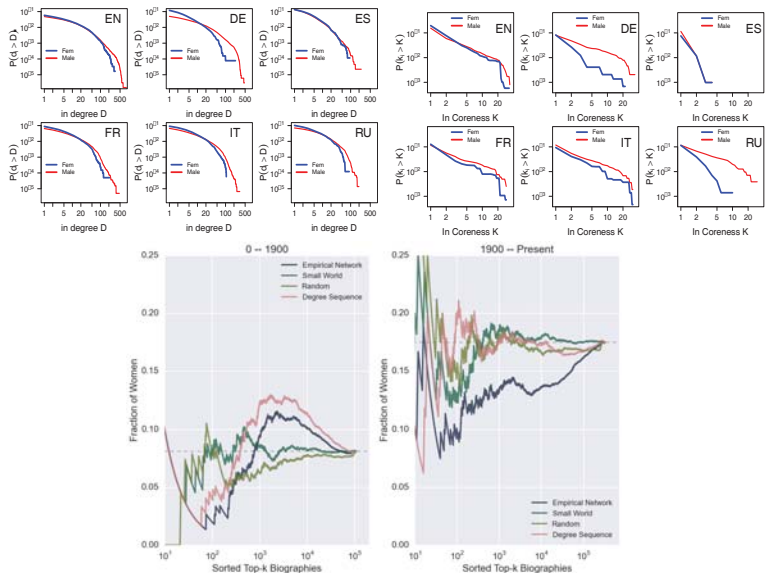


Are women who are covered on Wikipedia more notable than men?

- Neg. Binomial Regression: Number of language editions (DV)
- Being female has a sign. positive effect (IRR=1.13) also if we control profession and birth decade. Women are on average covered by 13% more language editions.
- Neg. Binomial Regression: number of regions and number of months with search volume above the Google trend threshold

	Num. regions			Num. months		
	β	Std. err.	p	β	Std. err.	p
Intercept	0.4417	0.048	***	3.4120	0.021	***
C(gender)(Tfemale)	0.2792	0.117	*	0.1220	0.052	**
AIC	20,939.81			53,351.84		
Num. obs.	5,998			5,998		

Structural Bias



Topical and Linguistic Bias

Inequalities in the word usage of articles about notable men and women.

Before 1900



After 1900



	Family	Gender	Relationship	Other
0-1900				
Men	0.5	1.5	0	98
Women	5.0	7	3	85
1900-present				
Men	0.5	2.5	0	97
Women	3	4.5	2	90.5

	% in men	% in women	χ^2	w	% change
Abstract positive	27.96	25.53	933.7***	0.04	8.69
Abstract negative	13.47	13.69	6.26**	0.005	-1.62

References

- [1] Scheiner-Fisher, C., and Russell, W. B. 2012. Using Historical Films to Promote Gender Equity in the History Curriculum. The Social Studies 103(6):221-225.
- [2] Murray, C. 2003. Human Accomplishment. The pursuit of excellence in the Arts and Sciences.
- [3] Amy Yu, Kevin Hu, S. R. D. G., and Hidalgo, C. A. 2013. The Pantheon Multilingual Wikipedia Expression Dataset. MIT project.
- [4] Bamman D, Smith NA (2014) Unsupervised discovery of biographical structure from text. Trans Assoc Comput Linguist 2:363-376
- [5] Maass A, Salvi D, Arcuri L, Semin GR (1989) Language use in intergroup contexts: the linguistic intergroup bias. J Pers Soc Psychol 57(6):981-993

Claudia Wagner, Eduardo Graells-Garrido, David Garcia and Filippo Menczer, **Women through the glass ceiling: gender asymmetries in Wikipedia**, EPJ Data Science Journal, 2016

Claudia Wagner, David Garcia, Mohsen Jadidi and Markus Strohmaier, **It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia**, The International AAAI Conference on Web and Social Media (ICWSM2015)