Replication of Wu (2018): Gendered Language on EJMR

Alejandro De La Torre

February 11, 2026

Paper overview

- Question: How are women and men discussed on EJMR, and do the words differ systematically?
- Approach: Lasso logistic models on word counts to identify gendered language.
- Replication targets: Table 1, Table 2, Figure 1 from Wu (2018).

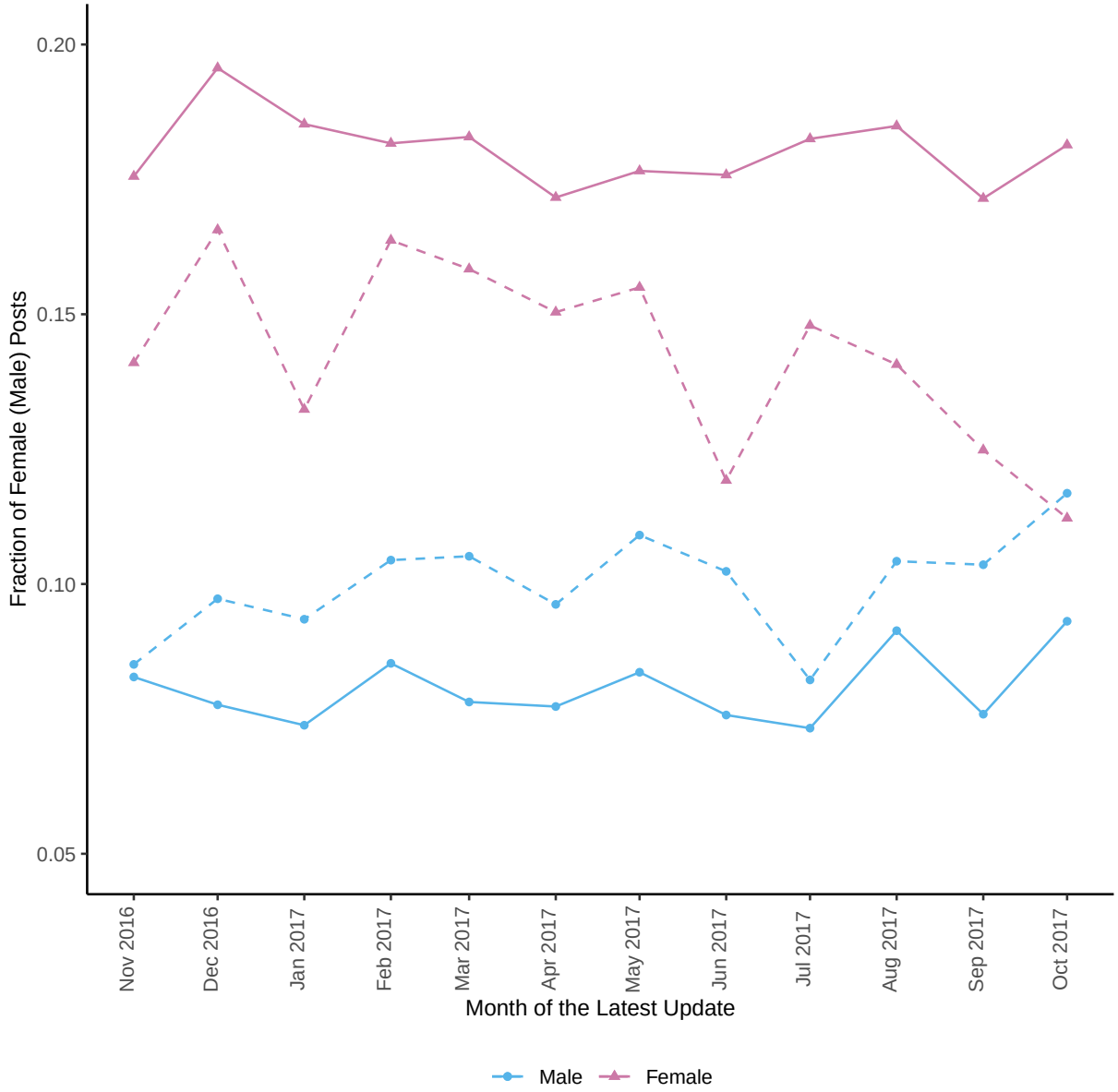
Replication pipeline

- Data: OpenICPSR package in `data/raw/openicpsr_wu2018_replication-pkg/`.
- Command: `python src/run_all.py`.
- Outputs: `output/figures/figure1.pdf`, `output/tables/table1.csv`, `output/tables/table2.csv`, and Lasso intermediates.

Obstacles and fixes

- Pandas deprecation: replaced `as_matrix()` with `to_numpy()`.
- NumPy security change: added `allow_pickle=True` to `np.load(...)`.
- These changes restore compatibility only; analysis logic unchanged.

Results snapshot



Modest extension

- Compared overlap of top 50 female- and male-associated words between full and pronoun samples.
- Overlap counts: 29 (female-associated) and 18 (male-associated).
- Interpretation: core signals are stable across labeling schemes, but not identical.

Conclusion and next steps

- Replication targets were reproduced with a single-command pipeline.
- Next step: implement an alternative NLP model for Table 1 (assignment extension requirement).