

# Does Victim Gender Matter for Justice Delivery? Police and Judicial Responses to Women's Cases in India

Chuchu Wen & Priscila Stisman

March 28, 2025

- 1 Introduction
- 2 Methods
- 3 Results
- 4 Extensions
- 5 Suggested Improvements
- 6 Autopsy

- **Research Question:** Are women disadvantaged when accessing justice in India compared to men, and if so how?
- **Hypothesis:** Women in India face 'multi-stage' discrimination when accessing justice, encountering systemic barriers at sequential stages of the justice process, including police registration, investigation, trial, and verdict.
- **Data:** 418,190 police reports in Haryana from January 2015 to November 2018.

- OLS Modeling
- **Topic Modeling**
- Text Matching

# Structural Topic Models STMs

## What is?

- Can predict whether cases devoted to a topic (e.g. rape) are functions of covariates
- Disaggregates what citizens told the police happened to them using statistical associations between words. e.g.: distinguish between domestic violence versus domestic violence that also included murder attempt.

## Goals

- (1) give voice to victims by utilizing their own words,
- (2) highlight the severity of claims, especially VAW, and
- (3) coarsen high-dimensional data to allow for text matching techniques.

Are there topics in the victims' testimonies—including inside women's complaints—that yield low convictions for suspects?

- **Step 1:** Data Preprocessing. Cleans and tokenizes the text and retains metadata, which contains additional information about each document (e.g., gender of the complainant, whether the person was convicted, etc.).
- **Step 2:** Preparing Documents for STM: refine data, filtering out rare words that appear in fewer than 50 documents.
- **Step 3:** Running the STM Model ( $K = 32$  topics).
- **Step 4:** Assign Topic Labels.

Run STMs in different corpuses and analyze topics

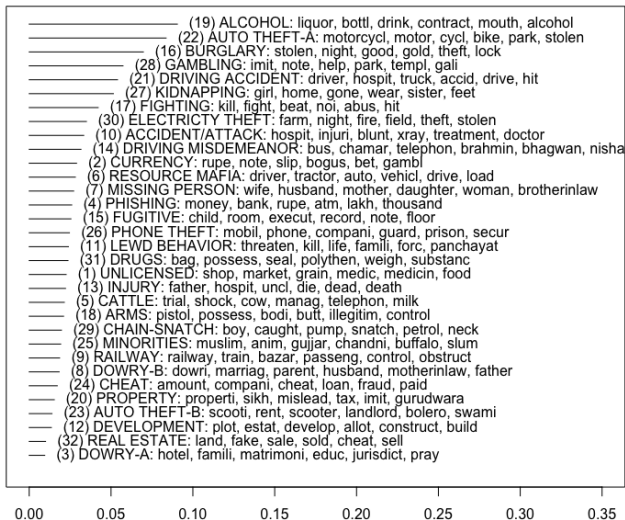
- All crime
- Female Compainant
- VAW reports

FREX (short for Frequency and Exclusivity) is a measure used in STMs to identify words that are both frequent within a topic and exclusive to that topic (highlight words that make a topic more distinguishable from others)

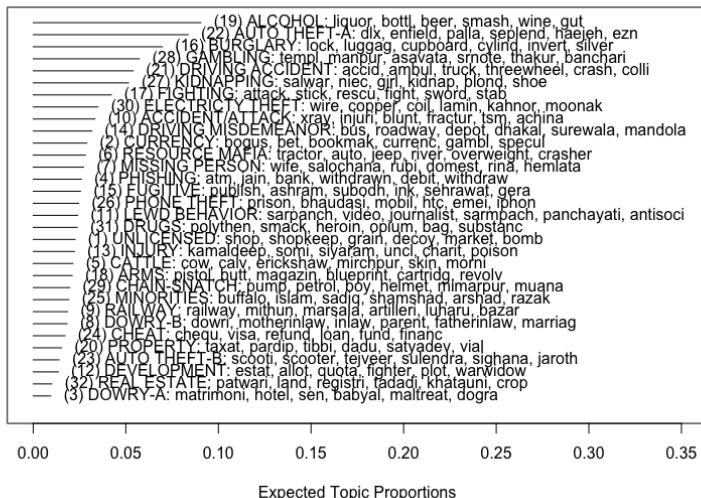
- High Probability Words: These words appear frequently in a topic but might also appear in other topics.
- Exclusive Words: These words are more unique to a single topic but might not be the most frequent ones.
- FREX Score: FREX balances frequency and exclusivity to identify words that are both common and distinctive for a topic.



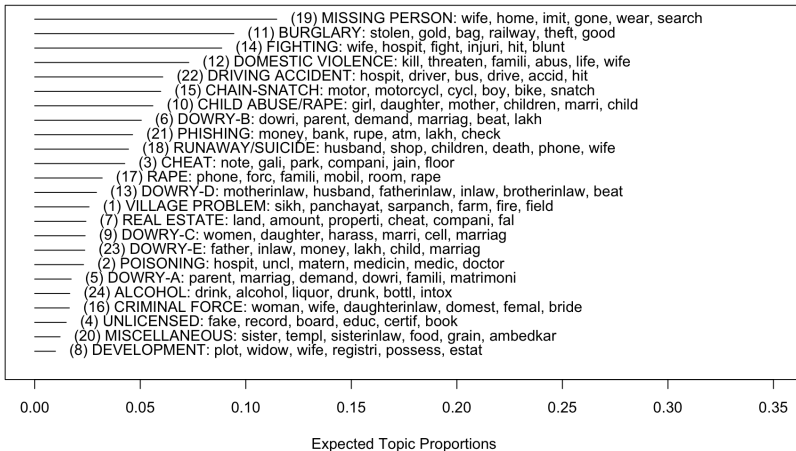
## All Crime Top Topics



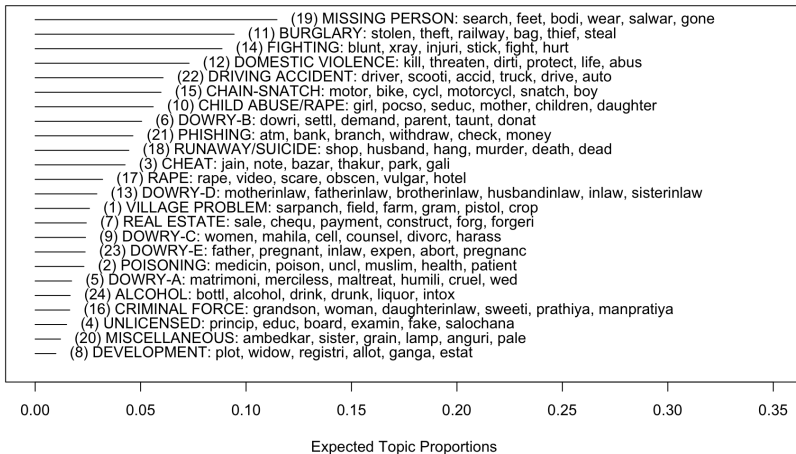
## All Crime Top FREX



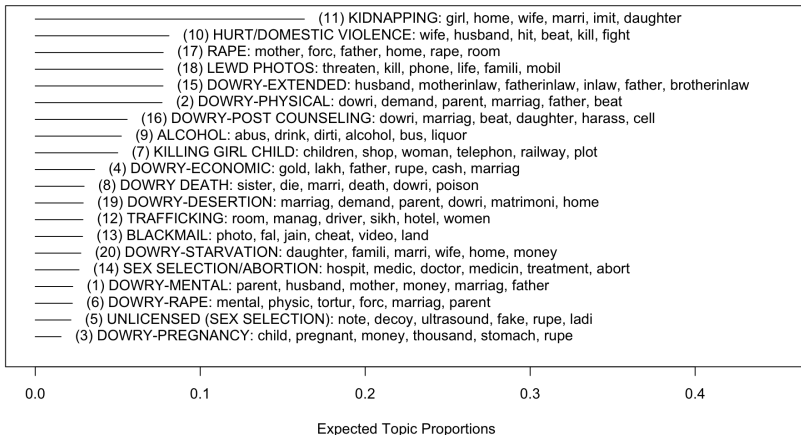
## Female Complainant Top Topics



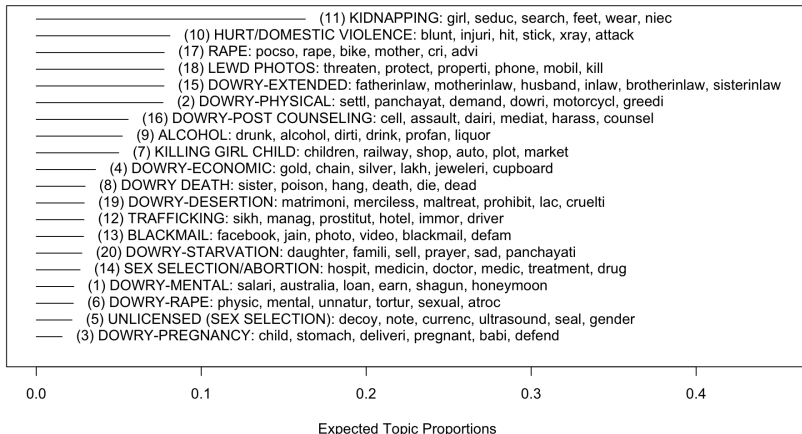
## Female Complainant Top FREX



## VAW Top Topics



## VAW Top FREX



- We analyze the most frequently used words by men and women within the same topic in police reports.
- Rationale: Do men and women use certain words differently or with varying frequency when reporting a criminal case?
- For example, in alcohol-related incidents, what words are predominantly used by men, and which ones are more common among women?
- We perform this exercise in three topics in non-gendered police reports

# Extensions Results

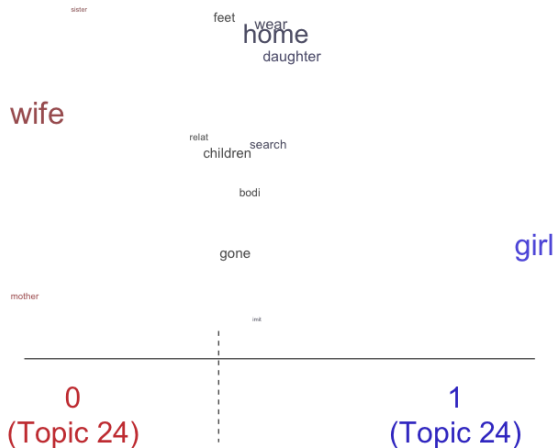
Word Used for theft related crime, by gender of the complaints





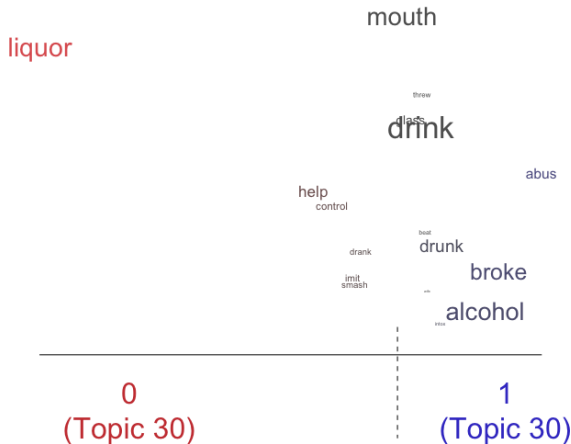
# Extensions Results

Word Used for abduction related crime, by gender of the complaints

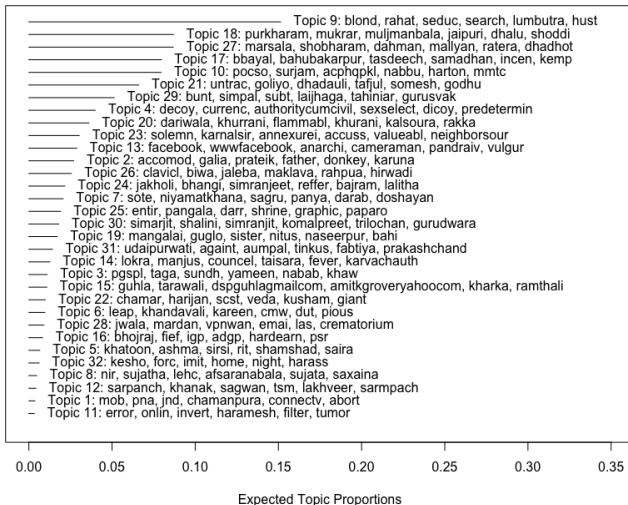


# Extensions Results

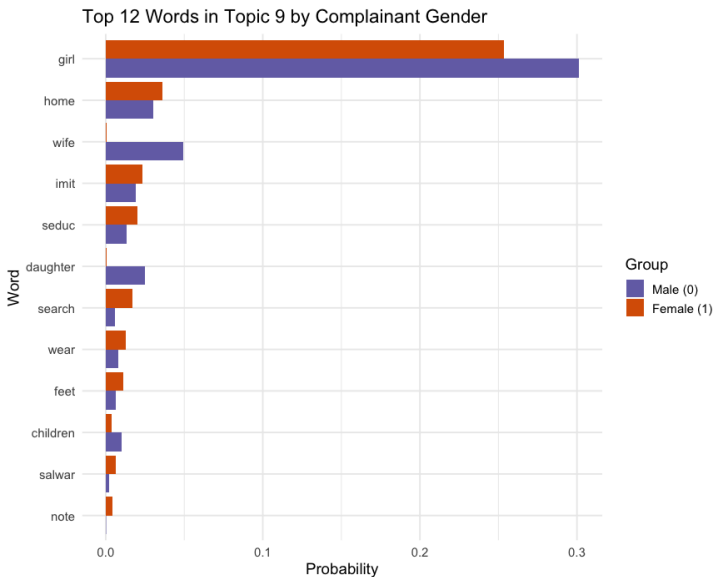
Word Used for alcohol related crime, by gender of the complaints



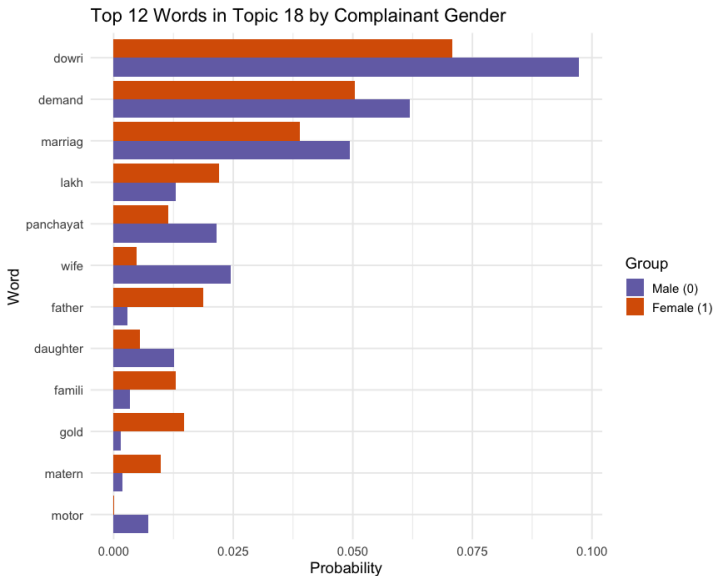
## Gendered Crime Content Covariate



# Extensions Results



# Extensions Results



# Suggested Improvements

- It is unclear how and where the translation from Hindi to English occurs in the author's code
- The gendered crime corpus was not translated, which makes challenging to easily interpret results

- Overall, finding a paper with well-documented replication materials was challenging
- We are highly satisfied with our STM results. It was also valuable to deepen our understanding of STMs and FREX
- We included the STM for all types of crimes, an analysis presented in the author's code but not in the paper
- All STM results were successfully replicated, producing identical outcomes since we used the same seed
- Our extensions were straightforward to implement and provided valuable insights