Gender Homophily in Open Source Software Networks
Tielin Katy Yu, Patrick Park, Huilian Sophie Qiu, Zihe H Zhao, and Bogdan Vasilescu

Carnegie Mellon University

Introduction

While gender diversity in open-source software (OSS) has been widely studied, the seemingly opposite concept, gender homophily, where individuals tend to collaborate with those of the same gender, is less explored.

Research Questions: Does the OSS community exhibit gender homophily in its collaborative networks? If so, what are the reasons behind it? Does it benefit gender diversity in OSS or work against it?

Methodology

Social Network Analysis
- Nodes: Individual contributors.
- Edges: Collaborations based on joint project contributions.
- Network Filtering: Employing the backbone method to filter out less significant connections based on edge weight significance scores.
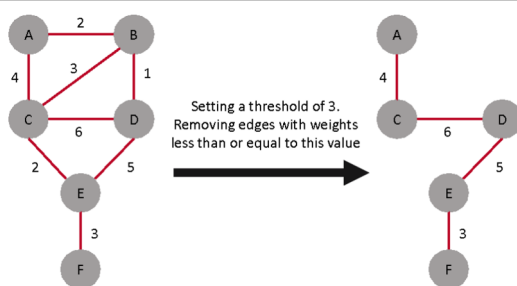

Fig 1. Toy Example of the Backbone Method

Analytical Approaches
- Shortest Path Distance Analysis: Estimating the mean shortest path distance for gender dyads as a proxy for social proximity.
- Exponential Random Graph Models (ERGMs): Utilized to estimate probabilities of same-gender ties, adjusting for network density and other potential confounders.

Data

- Data Source: Libraries.io (~8M packages) & World of Code (~18B Git objects)
- Gender Data: Binary gender inferred from contributor names using namsor API by World of Code
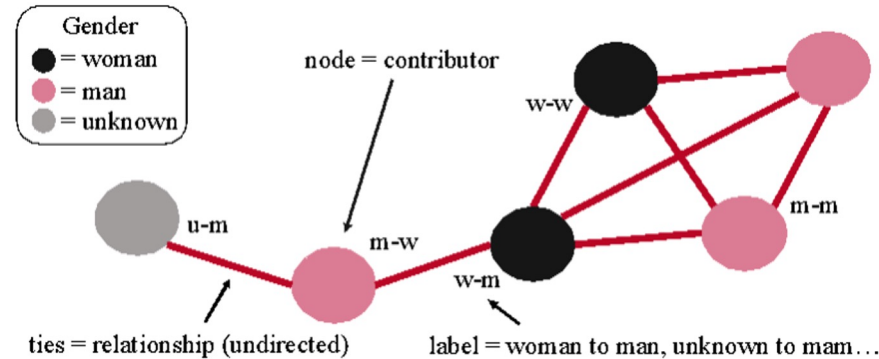

Fig 2. OSS Network Construction Method

Results

Network Composition: Significant gender skew (66.38% men, 2.70% women, 30.92% unknown).

Homophily Patterns:
- Path Distance Metrics: Women-to-women dyads exhibited significantly shorter path lengths than other gender pairings, suggesting closer collaboration networks among female participants.
- ERGM Findings: Women's probability of forming ties with other women was 28.3% higher compared to men, indicating strong gender homophily among female contributors.
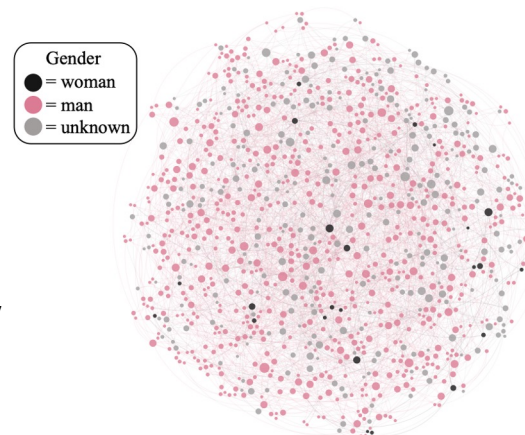

Fig 3. Visualization of Gender-Based Collaboration Patterns in OSS Networks Layout Using Force Atlas 2 Algorithm

Discussion

Implications:
- These patterns of gender homophily might influence the inclusivity and dynamics within OSS projects, potentially limiting diversity.
- Strong gender homophily among women could be a response to the underrepresentation and potential marginalization in predominantly male OSS environments.
- Recognizing gender homophily can help OSS communities devise strategies to foster a more inclusive environment. Initiatives may include mentorship programs, anonymous review processes, and policies that promote gender diversity in project teams.

Limitations and Considerations:
- The significant gender skew in OSS (high men dominance) complicates direct estimations of homophily and may bias results towards more visible gender-specific behaviors.

Future Work

- **Dataset Expansion:** The study will include a broader dataset across different time frames to validate and generalize findings.
- **Causation Analysis:** Investigating the causative factors behind gender homophily, focusing on elements of social (e.g., community culture, mentorship roles) and technical (e.g., project types, leadership roles).
- **Develop Practical Solutions: A key focus will be translating our insights into action**. We aim to design and propose targeted interventions and policy recommendations that will actively foster greater diversity and mitigate gender bias within OSS communities.

Contact:
tieliny@andrew.cmu.edu