

# Gender, Representation and Online Participation: A Quantitative Study of StackOverflow

David Arroyo Menéndez

March 8, 2020

- Published in: 2012 International Conference on Social Informatics
- Citas: 64 en Google Scholar (visto 2019/05/09)
- Autores:

Bogdan Vasilescu Eindhoven University of Technology, Eindhoven,  
Netherlands Daryl Posnett University of California, Davis, Davis, CA, USA  
Baishakhi Ray University of California, Davis, Davis, CA, USA Mark G.J.  
van den Brand Eindhoven University of Technology, Eindhoven,  
Netherlands Alexander Serebrenik University of Eindhoven, Eindhoven,  
Netherlands Premkumar Devanbu University of California, Davis, Davis,  
CA, USA Vladimir Filkov University of California, Davis, Davis, CA, USA

- URL: <https://dl.acm.org/citation.cfm?id=2702549>

- Importancia de las comunidades online (aspecto de calidad en software libre, descripción de sitios online relevantes)
- La cuestión social de la baja representación de mujeres en la industria del software (citaciones)
- Organización del paper

- Goal Question Metric
- Quizás se podría haber hecho un estudio cualitativo previo de que esas preguntas eran las adecuadas para el objetivo planteado. Así, las preguntas no están justificadas para el objetivo de estudio.

# Empirical approach (I)

- Hay pasos automáticos (nombre, localización) y manuales (person's avatar picture, based on additional data sources)

## Obtaining the Data

- Descripción de funcionalidades/componentes Stackoverflow
- Data dump

# Empirical approach (II)

## Automatic gender resolution

- Preprocessing

We first convert the names in Leet to Latin. For example, w35l3y is converted to Wesley. Identify real names of the participants that choose not to disclose them, using nicknames

- Gender resolution. The tool takes a (name, country) tuple as input, and returns one of “female”, “male”, or “x”

La crítica que le hago es que si escribieron una herramienta, escriban el artículo. Una herramienta de detección de género es una compañía actualmente ¿Por qué no eligieron gender guesser?

## Manual gender resolution

- Tras inspección manual concluyen que los avatares github, o los nicknames no son fáciles de identificar como male, female, debido a que un avatar puede ser una forma geométrica y un nickname algo como CoffeCode.
- Nombres, nicknames y avatares pueden ser usados en otras comunidades on-line conocidas.

Mejor entrevistas y grupos de discusión antes que la encuesta piloto.  
Críticas filosóficas/cualitativas que podrían haber tenido:

- justificar en la elección de las preguntas es el tema de las identidades trans/homosexuales/etc.
- el concepto de verdad en el género en Internet, en el censo, la validez de mi enfoque.
- una explicación más fuerte de Stackoverflow como muestra de ciencia y tecnología.



En caso de no hacer esas técnicas cualitativas, podrían haber hecho una justificación de conocimiento del campo de estudio. Por ejemplo, si tuviera que hacer algo así personal de Libresoft, pueden alegar libros escritos, u otras publicaciones.

Otra justificación para no usar cualitativismo es que en muchos estudios de mercado no se hace por motivos económicos. El dinero es limitado y el cliente quiere números.

Una encuesta piloto es una muy buena idea en cualquier caso.

## Qualitative analysis

- 2297 males
- 291 females
- 1556 not known

---

7% of women in Stackoverflow

## Hypothesis Testing

`/home/davidam/git/articles/lectures/mann-whitney.png`

## Discussion and Implications

- women do not engage with the software development online community;
- women are turned off by the blatant sexism of participants and leave these communities;
- women use neuter names or “male profiles” to cope and be accepted by the mostly male audiences.

Various reasons have been given to the under-representation of women in STEM subjects: a general **lack of interest** in STEM subjects, **stereotyped thinking** by family and teachers, lack of role **models**, and most often a combination of **various causes** together.

# Threats to Validity

- **Internal validity** (whether confounding factors can influence the findings). Example: names associated with one gender in the birth country may be associated with a different gender in the residence country.
- **External validity** (whether results can be generalized). Example: online communities act with similar gender barriers (e.g., gaming communities), while others are more gender- and minorities- friendly.
- **Construct validity** (relationship between theory and observation). Example: we note potential human error when inferring gender from an avatar picture, or when deciding whether a certain profile in another data source (e.g., Twitter) belongs to the same SO participant

# Conclusions and Future Work

## Conclusions

The main objective of the study was to add facts to current anecdotal evidence, that suggests that StackOverflow actively discourages the participation of women

A large proportion of SO users are not identifiable

- It was found that the percentage of women engaged in SO is greatly imbalanced, and men represent the vast majority of contributors
- men participate more, earn more reputation, and engage in the “game” more than women do

## Future work

Future work should expand on the current notion of gender as a binary phenomenon (male/female), an approach that has been already criticised by some of the gender-technology students