

Gendered creative careers in software development

Extended Abstract

Balazs Vedres
Center for Network Science
Central European University
Budapest, Hungary

Orsolya Vasarhelyi
Center for Network Science
Central European University
Budapest, Hungary
Vasarhelyi_Orsolya@phd.ceu.edu

Michael Szell
Center for Network Science
Central European University
Budapest, Hungary

ABSTRACT

While women has been shown to be vital in contributing to collective intelligence of a team, they are also marginalized in many respects. The field of software development is an area with especially strong male dominance. In this article, we analyze the paths to success for men and women based on activity traces on GitHub. We collected data on more than ten million users, active in more than ten million repositories. We analyze the careers of the most active ten thousand women, men, and users of unidentified gender. We found that women can succeed to the extent if they adopt a collaboration and activity pattern characteristic of men.

Keywords

Gender bias, career, segregation, success, software development, GitHub

The empirical basis of this study is a data set acquired via

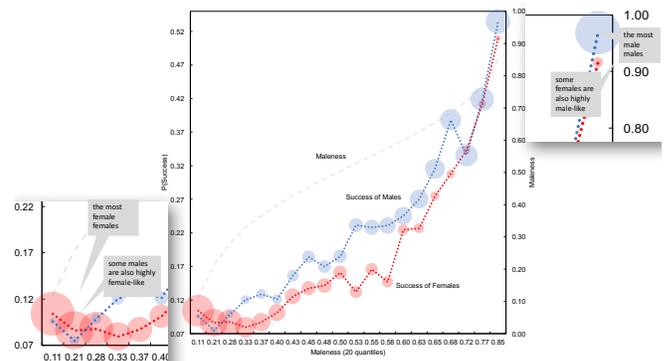
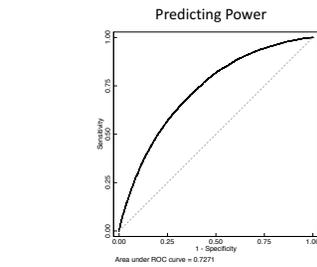


Figure 2: How Maleness Correlates with Success

GitHubarchive.org about the activities on the social coding platform GitHub, encompassing the following types of events that took place between 2009-02-19 and 2016-10-21: creation of repository, addition of new member to a repository, push to a repository, opening, closing and merging a pull request, user watching a repository, user following another user. To collect information about users' name, e-mail address, number of followers, number of people they follow, number of public repositories and the date they joined GitHub, we sent calls to the official GitHub users API. Since users do not list their gender directly, we infer each person's gender based on their first name. We take an approach pioneered by Wachs and coauthors [Wachs et al. 2017], to operationalize the extent to which someone's behavior follows a gendered pattern, as opposed to operationalize gender as a binary hard-encoded category. For gender prediction we ran a logistic regression model. Success is defined as the number of stars on users' own repositories.

	Odds ratio	SE	p
Collaboration			
N of collaborators (log)			
Female	.081*	.019	.000
Male	1.025	.233	.850
Unknown	.615*	.078	.000
Female * Female	.322*	.120	.002
Male * Male	.856	.136	.326
Unknown * Unknown	.861	.143	.366
Female * Male	3.098*	.919	.000
Female * Unknown	1.521	.484	.188
Male * Unknown	1.875*	.380	.002
Activity Controls			
N of repositories (log)	1.174	.478	.093
Proportion of own repositories	.763	.064	.201
Pushes (log)	1.587	.384	.016
Pushes / N of repositories (log)	.487	.211	.269
N of followers (log)	.813*	.038	.000
Collaborators / N of repositories	.983	.082	.430
Versatility	1.028*	.008	.000
Tenure	1.200	.128	.000
Tenure * Tenure	1.003	.007	.853
N of repositories / Tenure (log)	.548	.274	.229
Positive significant			
TypeScript	1.093*	.032	.003
Groovy	1.074*	.036	.034
Shell	1.059*	.036	.000
C#	1.029*	.008	.000
PHP	1.014*	.005	.002
JavaScript	1.013*	.002	.000
Negative significant			
Ruby	.992*	.003	.001
Objective C	.988*	.004	.002
Other	.987*	.005	.000
HTML	.944*	.008	.000
CSS	.934*	.008	.000
XSLT	.699*	.088	.004
Unknown	.611*	.114	.031
Unknown	.608*	.146	.039



Collaboration network homophily and specialization of the activity profile contribute to a soft form of gender segregation, which is rather a female characteristic. Women have more followers, but men are more versatile.

Figure 1: Maleness Prediction with Logit Model

womENCourage 2017 Barcelona, Spain

ACM ISBN 978-1-4503-2138-9.

DOI: 10.1145/1235

1. REFERENCES

[Wachs et al. 2017] Johannes Wachs, Anikó Hannák, András Vörös, and Bálint Daróczy. 2017. Why Do Men Get More Attention? Exploring Factors Behind Success in an Online Design Community. *arXiv preprint arXiv:1705.02972* (2017).