Gendered Success in Developer Careers

Orsolya Vásárhelyi*, Balázs Vedres, Micheal Széll

Vasarhelyi_Orsolya@phd.ceu.edu

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

Introduction

The field of software development is an area with especially strong male dominance. We analyzed the paths to success of the most active ten thousand women, men, and users of unidentified gender based on activity traces on GitHub. We found that women can succeed to the extent if they adopt a collaboration and activity pattern characteristic of men.

Data

Data about collaboration and personal GitHub profiles was acquired via GitHubarchive.org. Using publicly available personal information, such as real names, e-mail addresses and logins we inferred gender for 1.6 million users.

Collaboration activity was used to build up individual careers. After filtering for users with at least 10 actions we created stratified samples of each gender group, 10 000 each three categories, to run our analysis.

Based on users' activity we predicted with a logistic regression model users' *genderedness*. We used the gender prediction results *(maleness)* to understand the correlation between success and gendered behavior. Success is defined as the number of stars on users' own repositories.

Years	2010-2016
Users	7.8 Million
Inferred Gender	
Male	1.4 million
Female	180 000
Filtered Data*	
Male	305 000
Female	26 000
Unknown	300 000

We filtered users with at least 10 actions (creating a 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)

Research Questions

How does the genderedness of behavior relate to success?

Do women need to follow male patterns of behavior to succeed and survive?

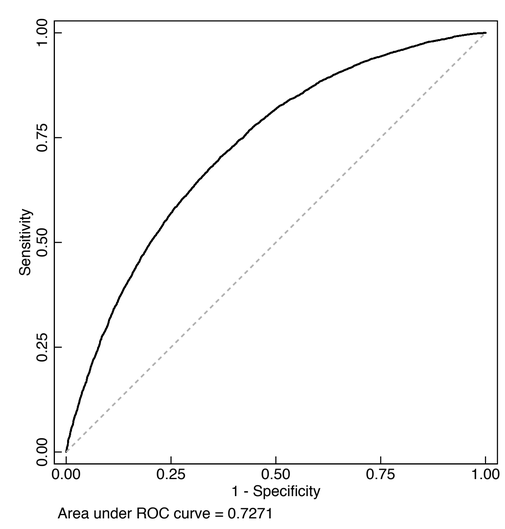
Even if they do, can they achieve the same success as men?

Results

Gender Prediction

	Odds ratio	SE	p	
N of collaborators (log)				*Collaboration*
Female	.087*	.019	.000	
Male	1.025	.133	.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	
N of repositories (log)	1.174	.478	.693	*Activity Controls*
Proportion of own repositories	.915	.064	.201	
Pushes (log)	1.587	.384	.056	
Pushes / N of repositories (log)	.667	.215	.209	
N of followers (log)	.813*	.028	.000	
Collaborators / N of repositories	.933	.082	.430	
Versatility	1.028*	.008	.000	
Tenure	1.220	.128	.060	
Tenure * Tenure	1.001	.007	.853	
N of repositories / Tenure (log)	.548	.274	.229	
Positive significant				*Languages*
Typescript	1.093*	.032	.003	
Groovy	1.074*	.036	.034	
Shell	1.059*	.016	.000	
C#	1.029*	.008	.000	
PHP	1.015*	.005	.002	
Javascript	1.013*	.002	.000	
Negative significant				
Ruby	.992*	.003	.001	
Objective C	.988*	.004	.002	
Other	.967*	.005	.000	
HTML	.944*	.008	.000	
CSS	.935*	.008	.000	
XSLT	.699*	.088	.004	
Livescript	.671*	.124	.031	
Unknown	.608*	.146	.039	

Predicting Power



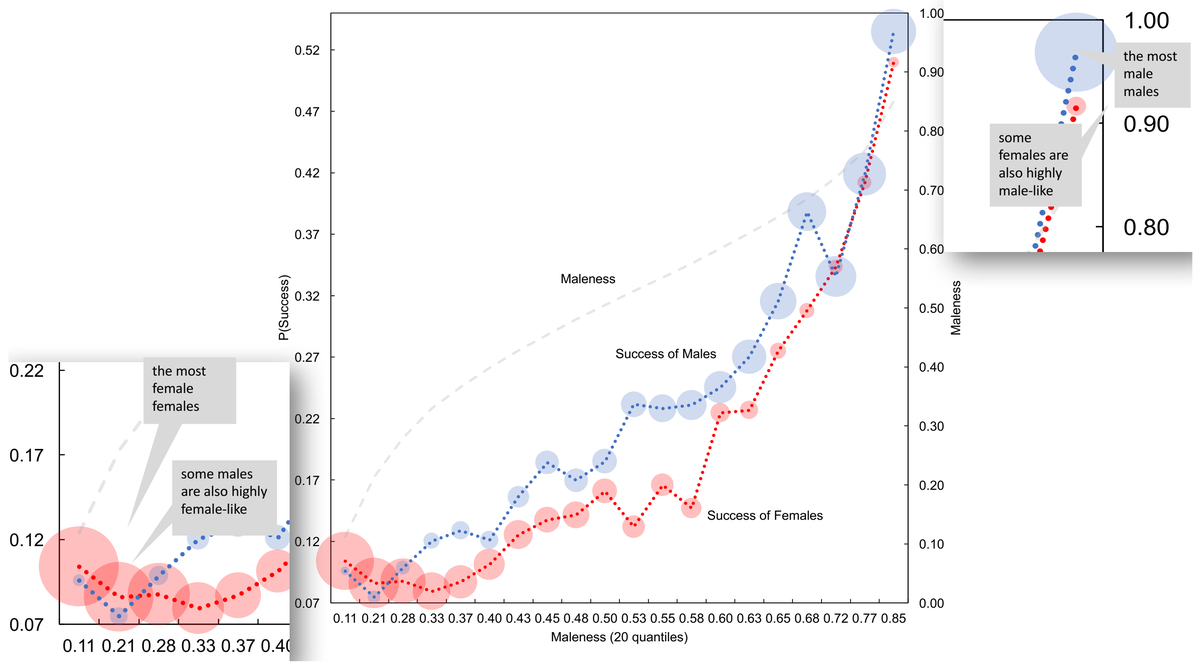
Gender homophily in collaboration is more likely among women.

Women have more followers, men use more different languages.

Maleness is positively related to success (# repository stars) but female developers systematically less successful.

Female developers' quit rate is double compared to men.

Maleness by Success



Background Image Source: http://cameronmcefee.com/img/work/the-octocat/femalecodercat.jpg