

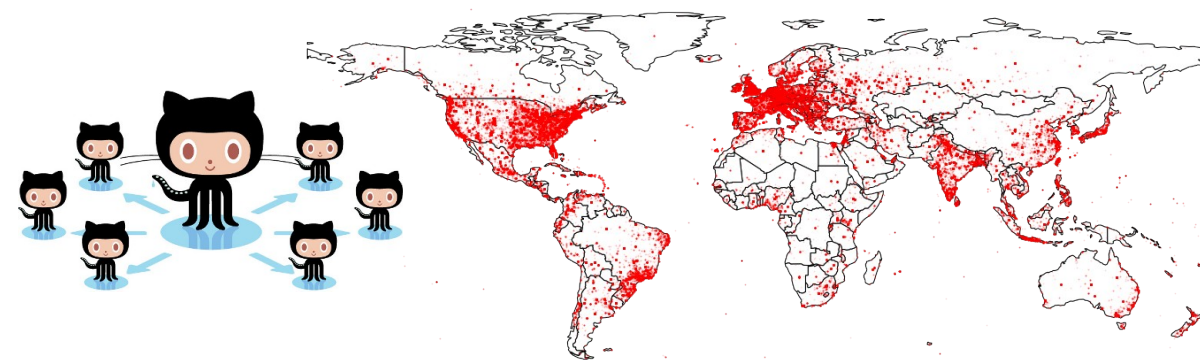
Modeling popularity dynamics on GitHub

Nikolas Zöller

Jacobs University Bremen, University of Applied Sciences Potsdam

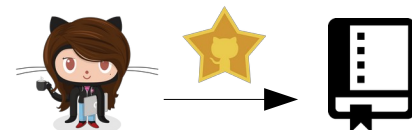
What is GitHub?

- Most important platform for collaborative Free and Open Source software development.
- Serves as a hosting service for software repositories and as a social network for developers.
- More than 50 million reported user accounts and over 200 million code repositories hosted.



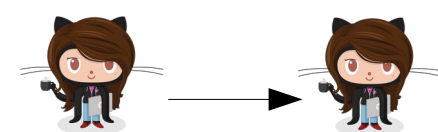
Popularity on GitHub

Starring:



- Developers have the possibility to star software repositories.
- Use cases: 1. Show appreciation, 2. Bookmark projects, 3. Current or past usage [1].

Following



- Users can follow other users in order to be notified about their activities on the platform.

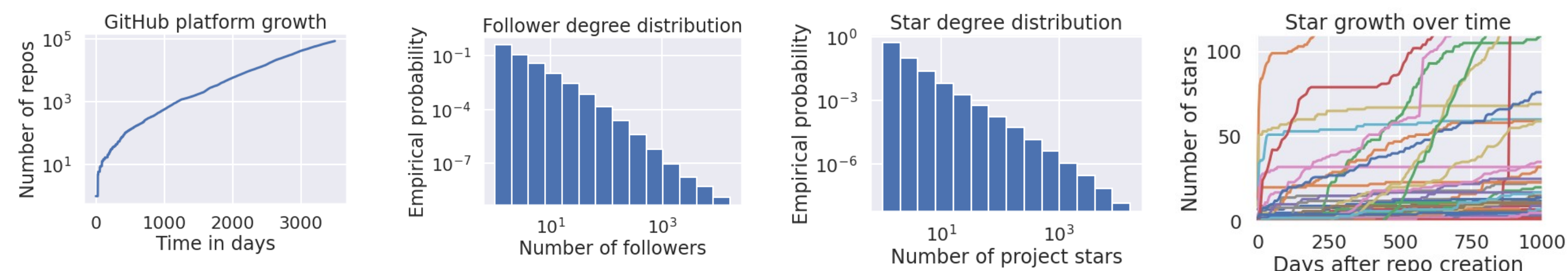
Why is popularity important?

- Maintaining or contributing to a popular project provides prestige and status among peers [2].
- Users take a high star count as a signal for quality and a marker of a project's success
- Popular users influence their followers. If they engage with a new project a significant portion of their followers is attracted to that project as well [3].

Research question

- What are the mechanisms that drive popularity dynamics on GitHub?

Empirical data analysis



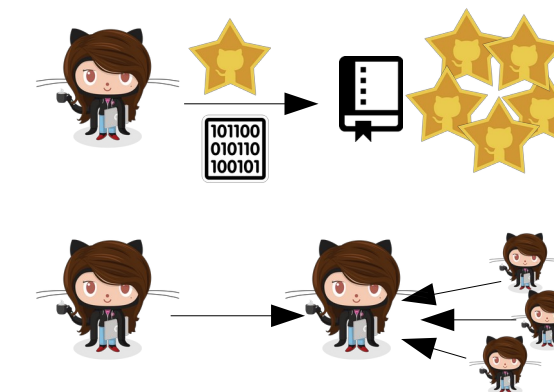
- The number of repositories on the platform grows exponentially (after an initial phase).
- The degree distributions for the number of followers per user and the number of stars per repository are **heavy-tailed**.
- Looking at the star growth over time of randomly sampled projects we find projects whose star count grows steadily and almost linearly. However, we also find projects that exhibit sudden bursts in their star growth.

Hypotheses

In growing social networks, heavy-tailed degree distributions can often be traced back to preferential attachment mechanisms [4].

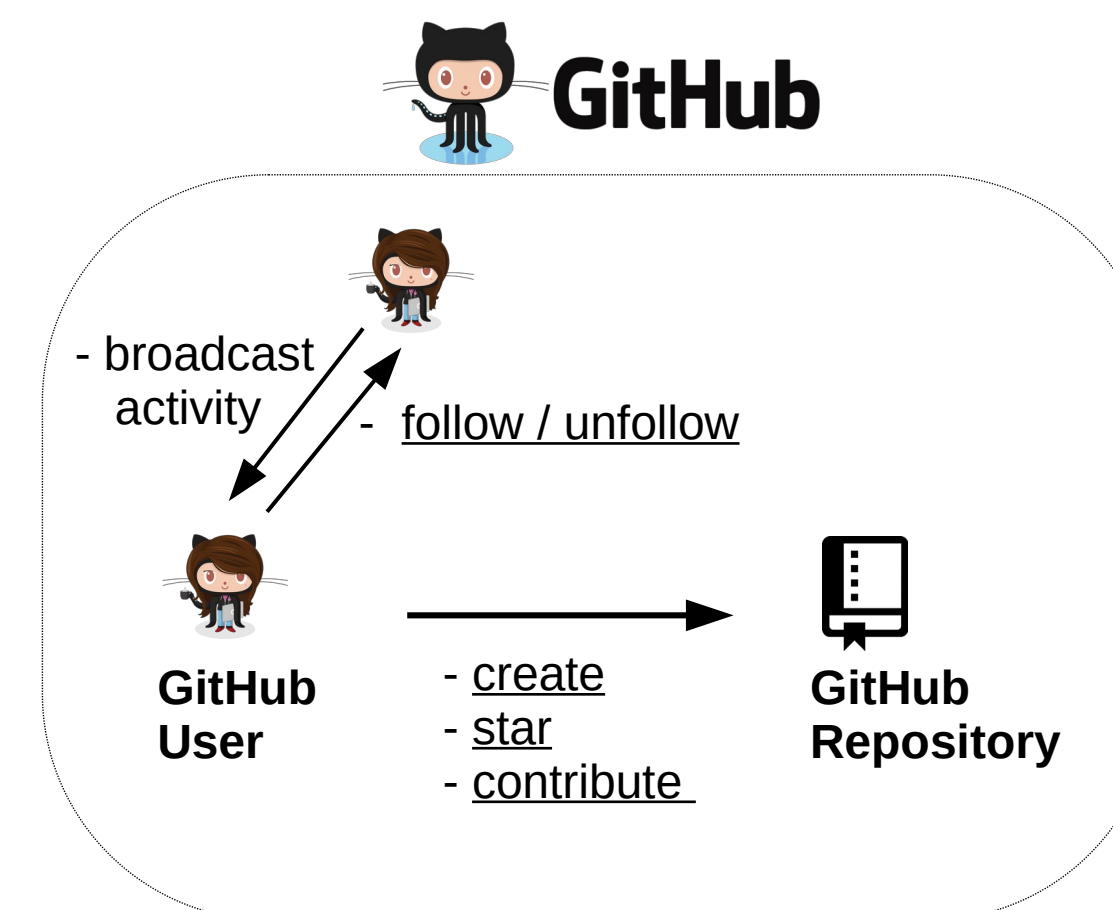
We claim that there are three relevant **preferential attachment** mechanisms that influence popularity:

- 1) Users will more likely star projects that already have many stars.
- 2) Users will more likely contribute to projects that already have many stars.
- 3) Users will more likely follow users who already have many followers

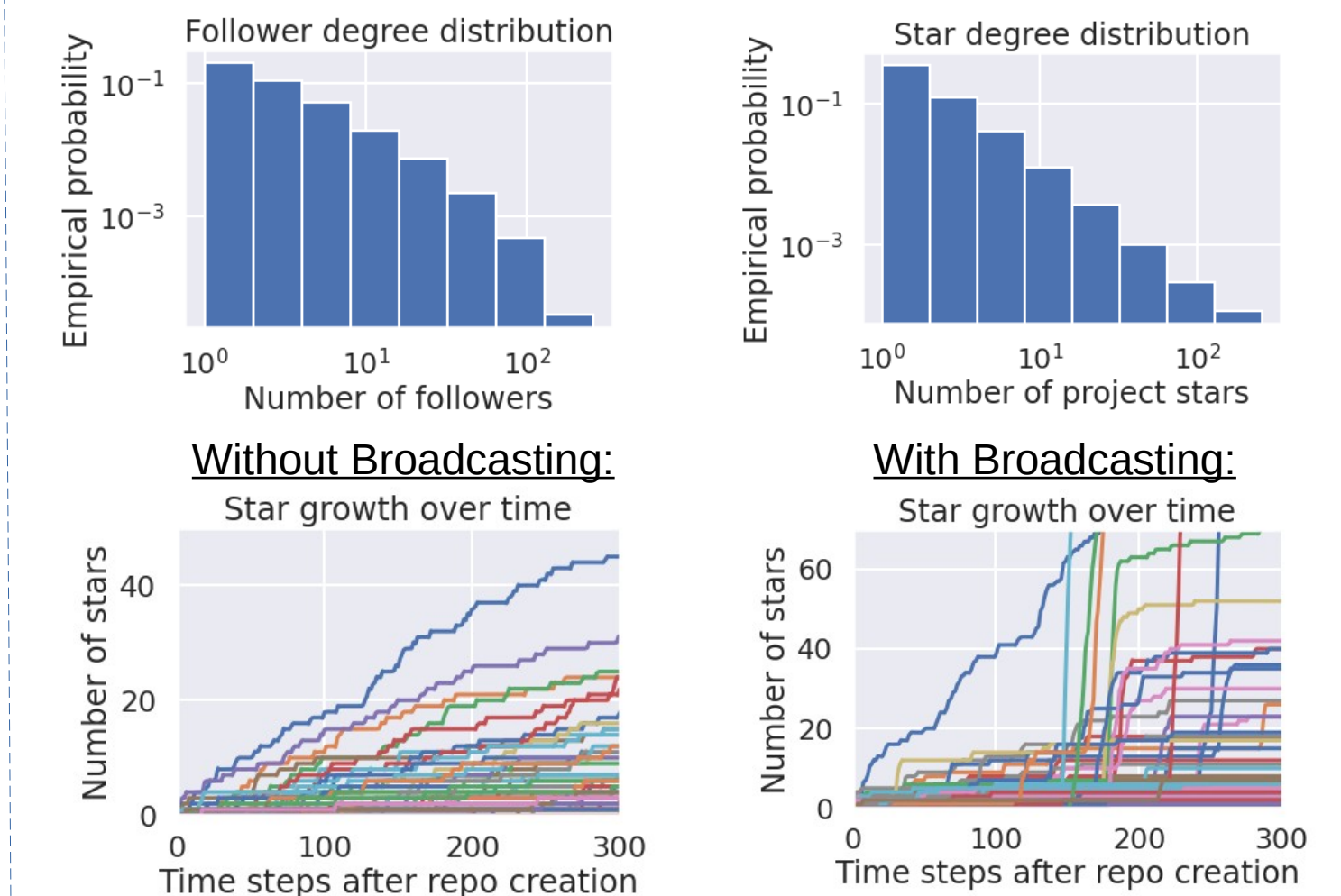


Agent Based Model

- **Model entities:** GitHub Users, Repositories and the GitHub platform which provides the global environment.
- **At each time step:**
 - Users probabilistically perform several actions: they create new repositories, contribute to and star existing repositories and start to follow other users.
 - New users can join the platform.
- **Preferential attachment** probabilistically determines whom a user decides to follow and which repository a user decides to star or to contribute to according to the hypotheses.
- **Broadcasting:** If a user engages with a new repository their followers are informed and in the next time step there is a higher probability for them to also engage with that repository.



Simulation results



- Preferential attachment mechanisms indeed lead to heavy tailed follower and star degree distributions.
- A broadcasting mechanism cause occasional sudden bursts in individual projects' star growth also observed in the empirical data.

Conclusions

We use an agent based model to simulate popularity dynamics on GitHub and find that preferential attachment mechanisms in GitHub users' activities (starring, contributing and following) can explain how empirically observed heavy-tailed degree distributions arise. Furthermore, we simulate a broadcasting mechanism that informs followers about popular user's activity and show that it can lead to sudden bursts in popularity also observed empirically on GitHub. A good strategy for maintainers to boost their projects might therefore be to reach out to popular users in order to draw their attention (and consequentially the attention of their followers) to a specific project.

References

- [1] Borges, H. and Tulio Valente, M.: . What's in a GitHub Star? Understanding Repository Starring Practices in a Social Coding Platform. Journal of Systems and Software, 146:11–129. (2018)
- [2] Eckhardt, E., Kaats, E., Jansen, S., and Alves, C.: The merits of a meritocracy in open source software ecosystems. ACM International Conference Proceeding Series (2014)
- [3] Blincoe K., Sheoran J., Goggins S., Petakovic E., Damian D.: Understanding the popular users: Following, affiliation influence and leadership on GitHub. Information and Software Technology, 70:30–39. (2016)
- [4] Newman, M. E. J.: Clustering and preferential attachment in growing networks. Phys. Rev. E, 64:025102. (2001)