

# Tracing Community Genealogy: How New Communities Emerge from the Old

Chenhao Tan  
University of Colorado Boulder  
[chenhao@chenhaot.com](mailto:chenhao@chenhaot.com)  
@ChenhaoTan

# A world of diverse communities



politics



changemyview



Babybumps



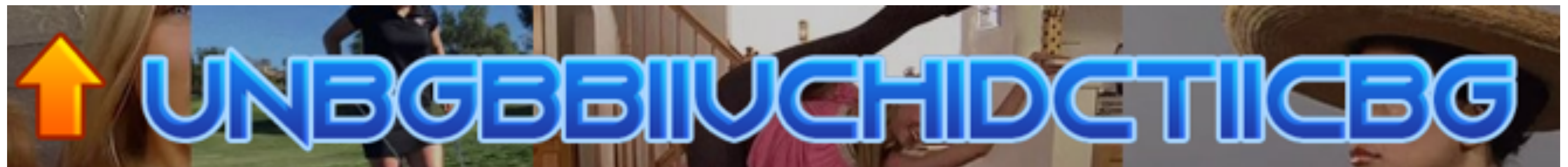
The\_Donald



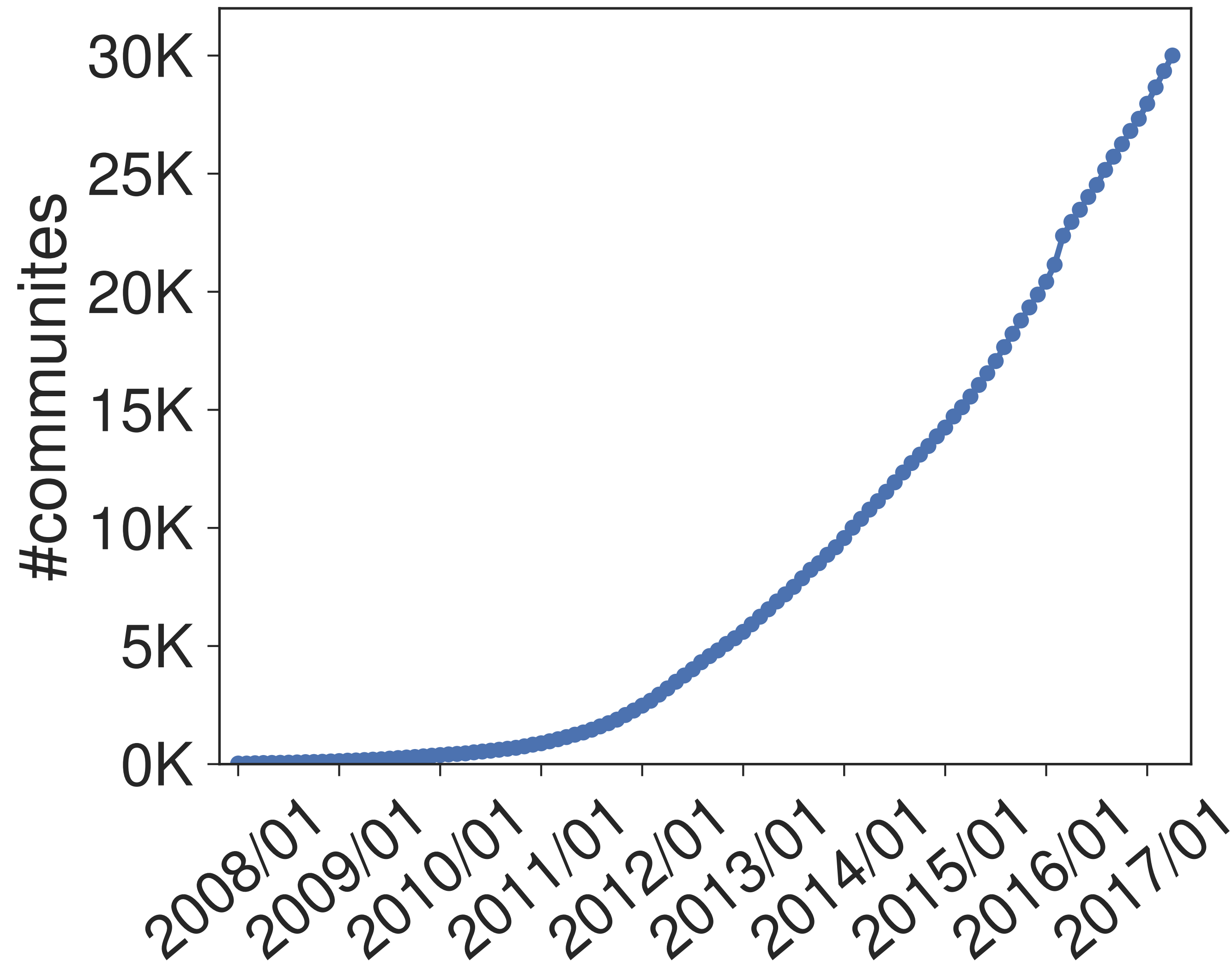
SandersForPresident



hillaryclinton



# Continually emerging communities



**Unfold the emerging process of  
new communities**



# A community does not emerge from a vacuum



politics



The\_Donald



AskThe\_Donald



AskReddit

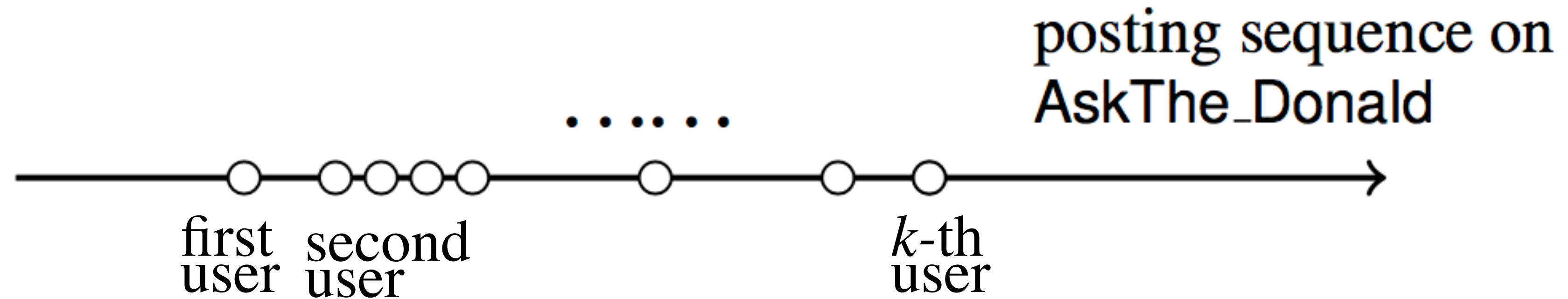
How can we trace the community genealogy?

*The key is to define the (parent-child) edges.*

[Padgett and Powell 2012;  
Fleming et al. (2007)]

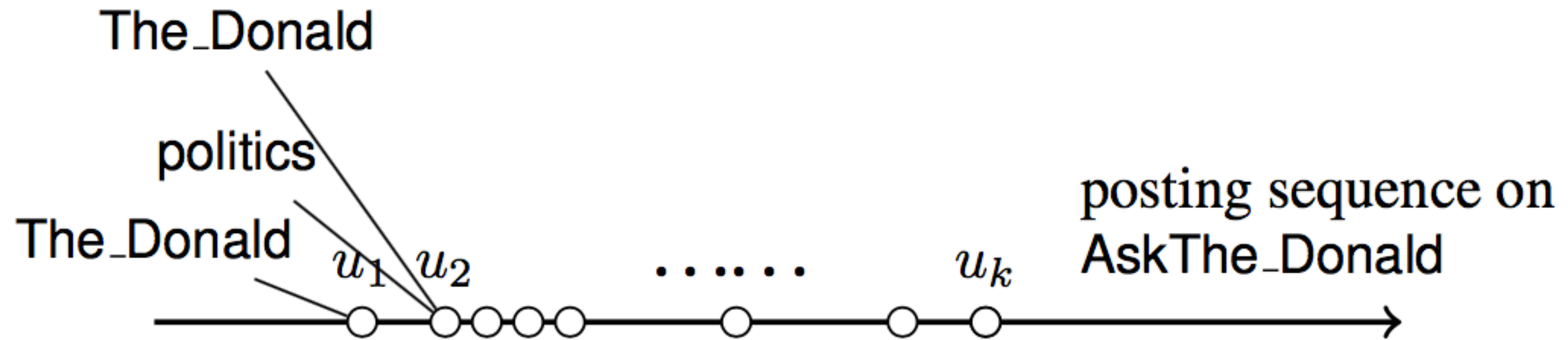
# Define the parents of a community

Focus on early users in a new community



# Define the parents of a community

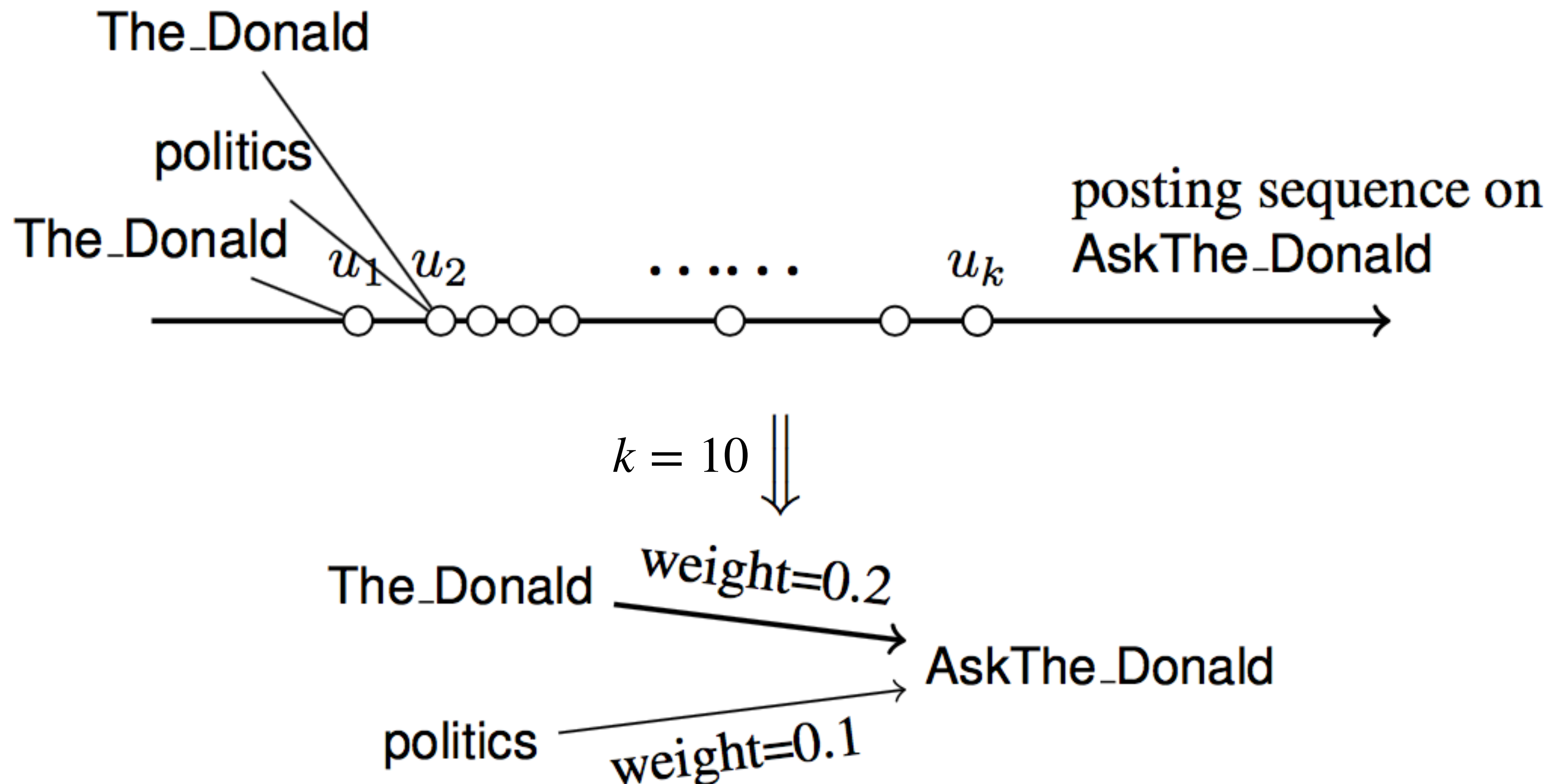
Focus on early users in a new community



Subreddits that users  
were recently active in

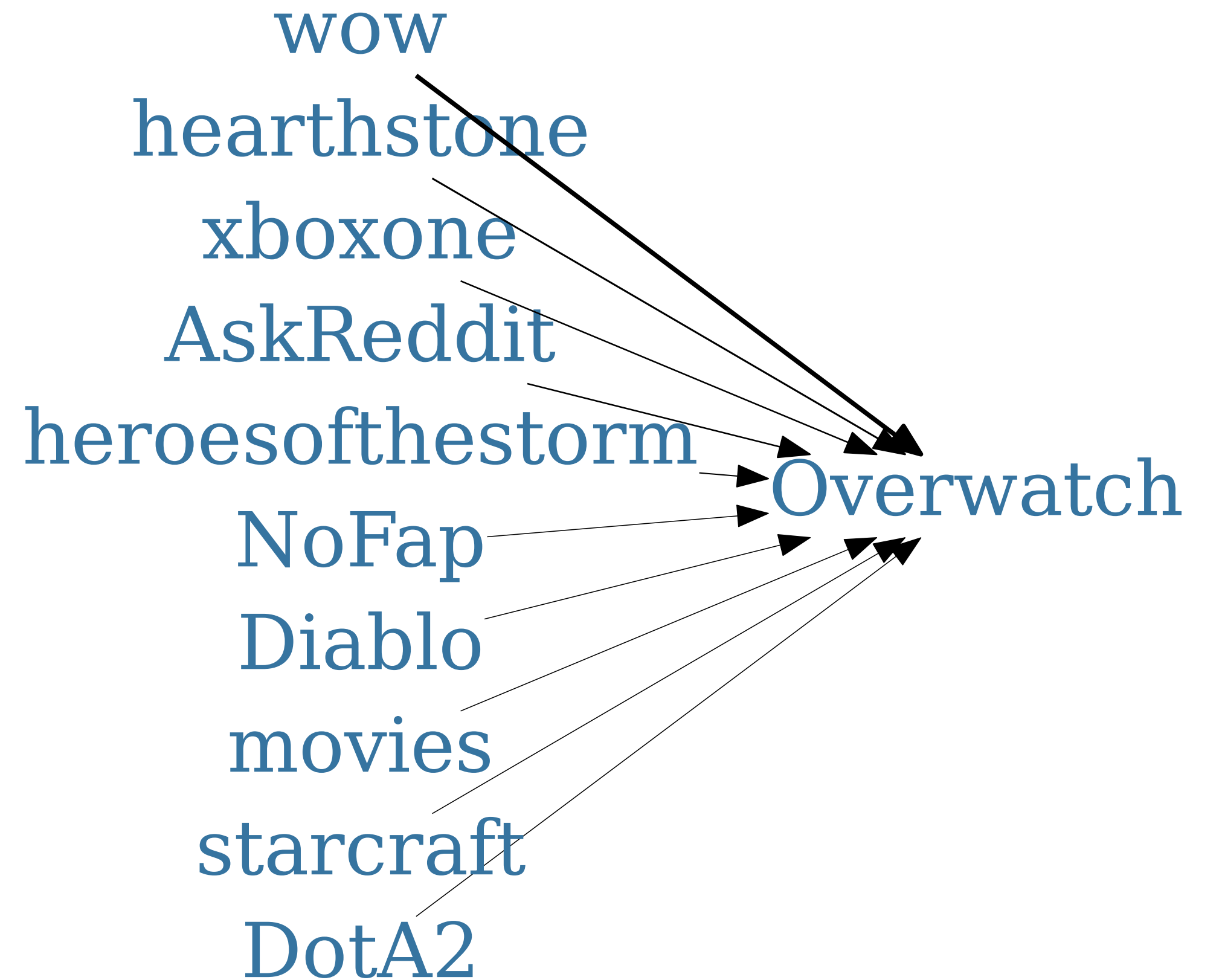
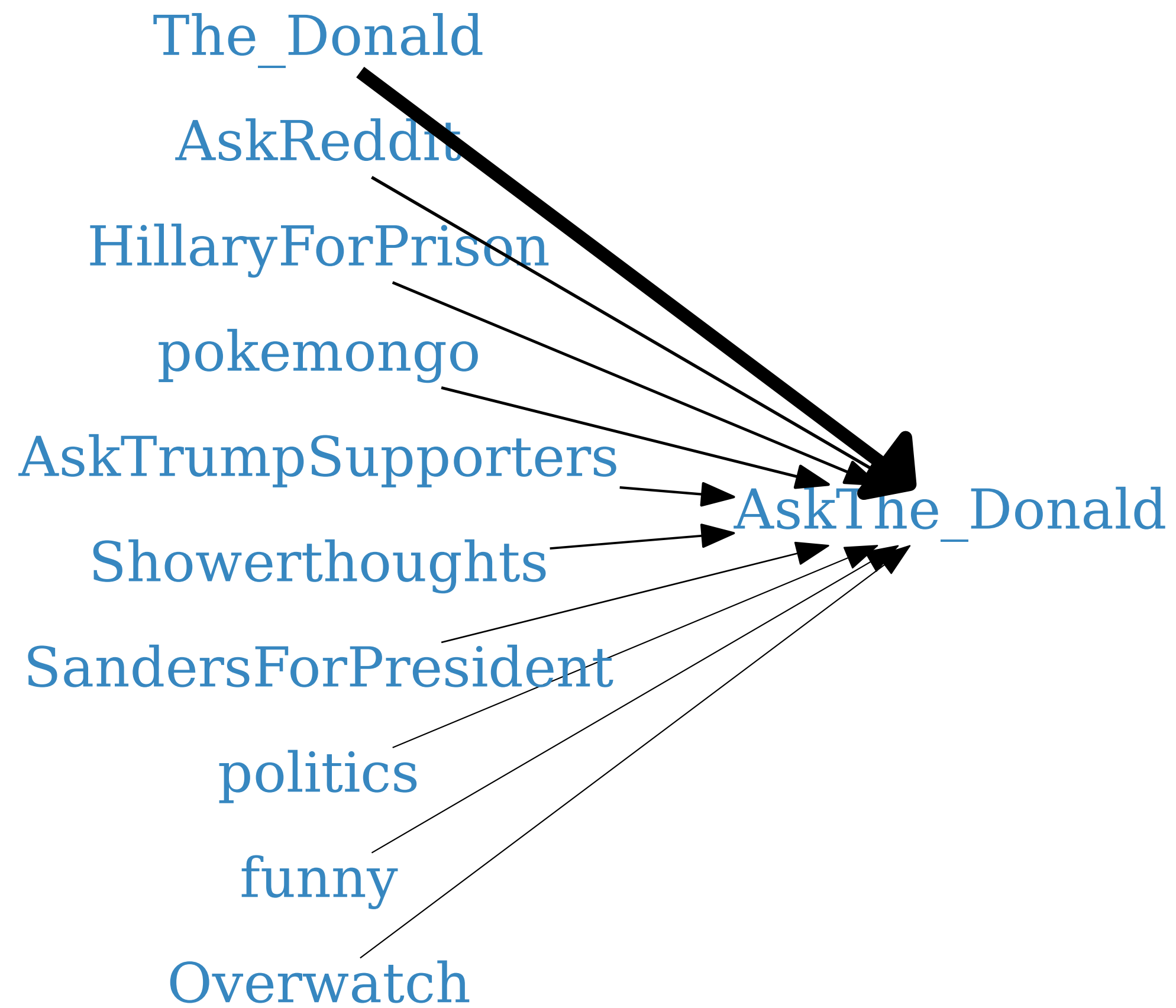
# Define the parents of a community

Focus on early users in a new community

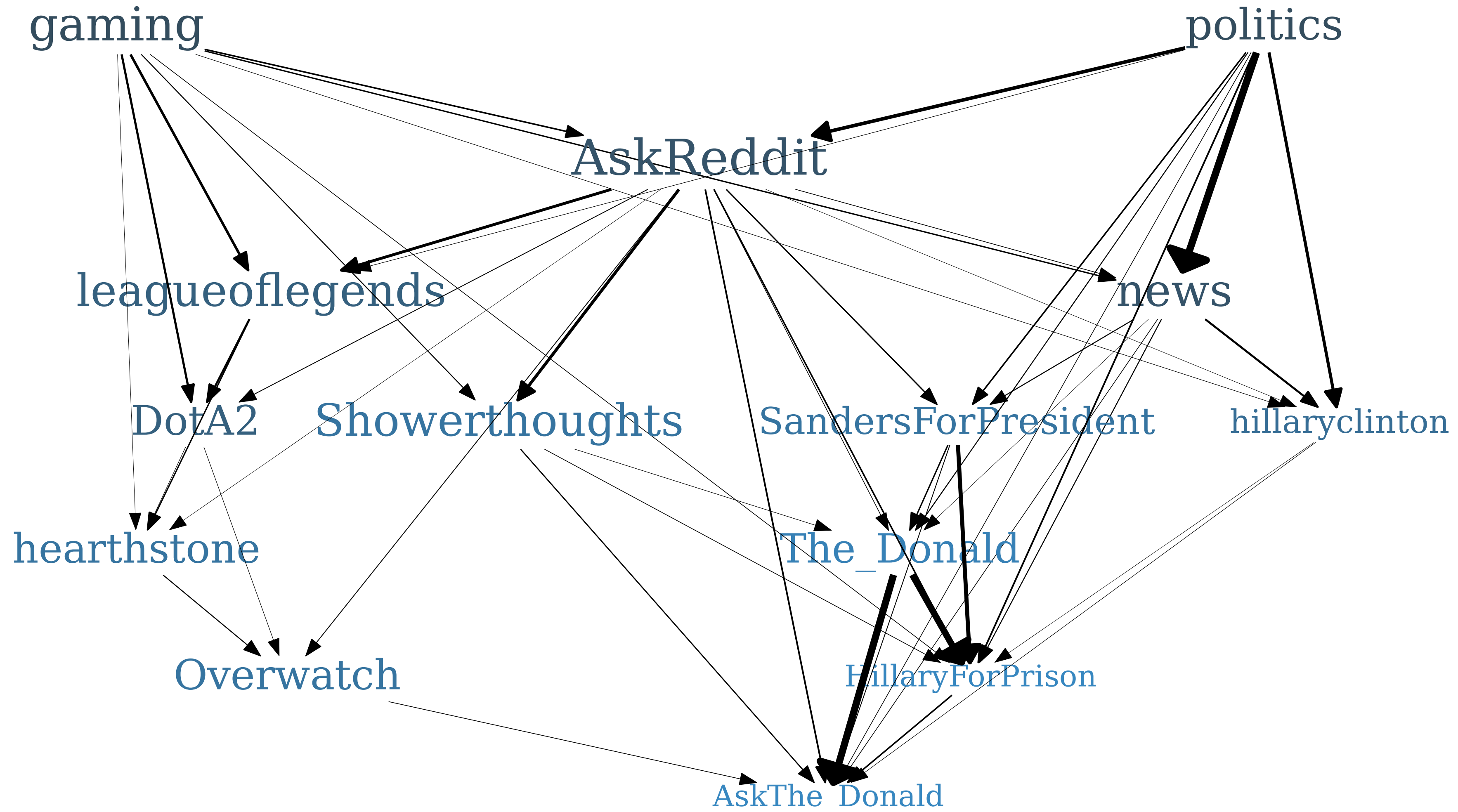




# Top parents



# Genealogy of Communities

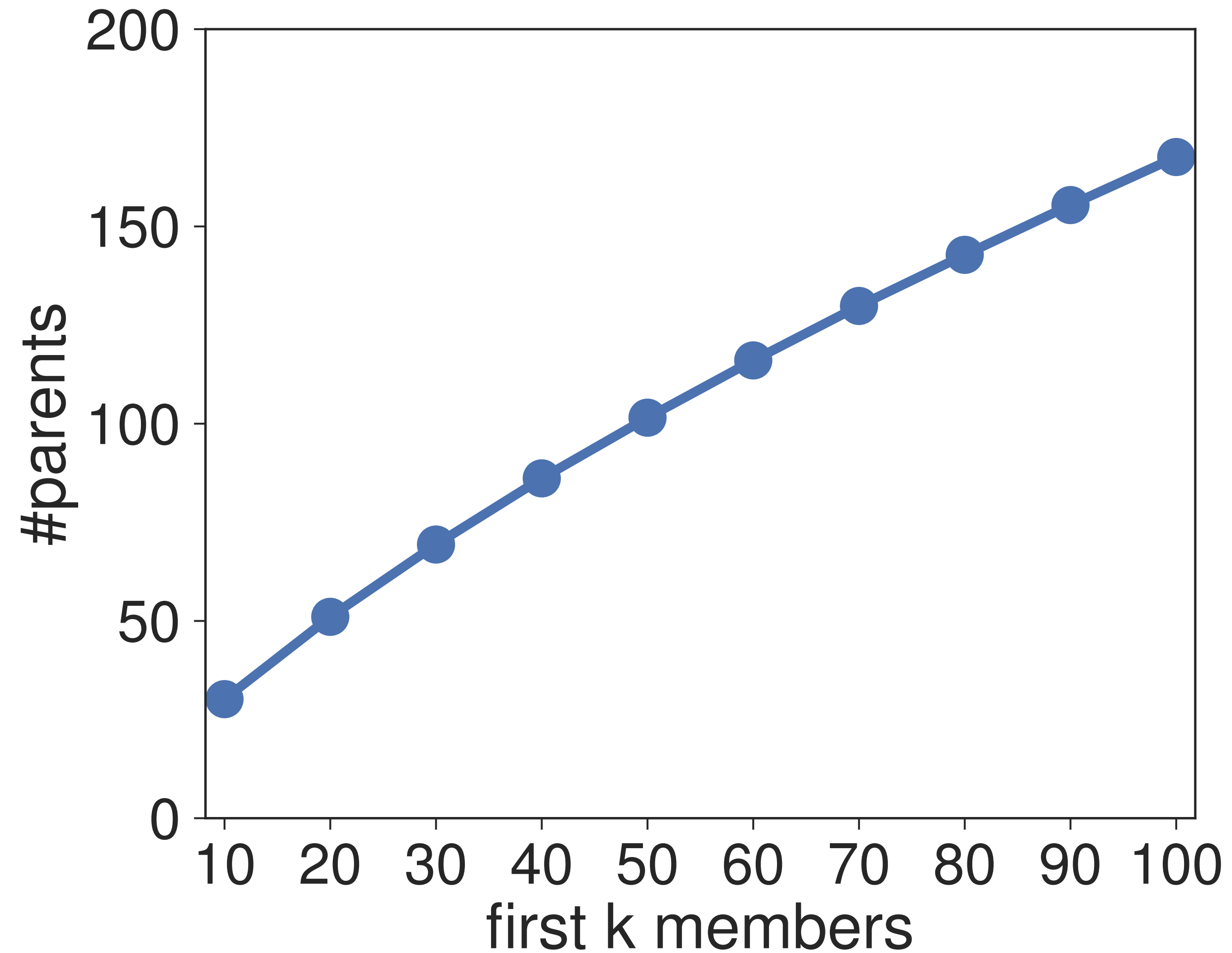


*Demo: <https://redditvisualization.herokuapp.com/>  
(thanks to Kumar Srinivasan)*

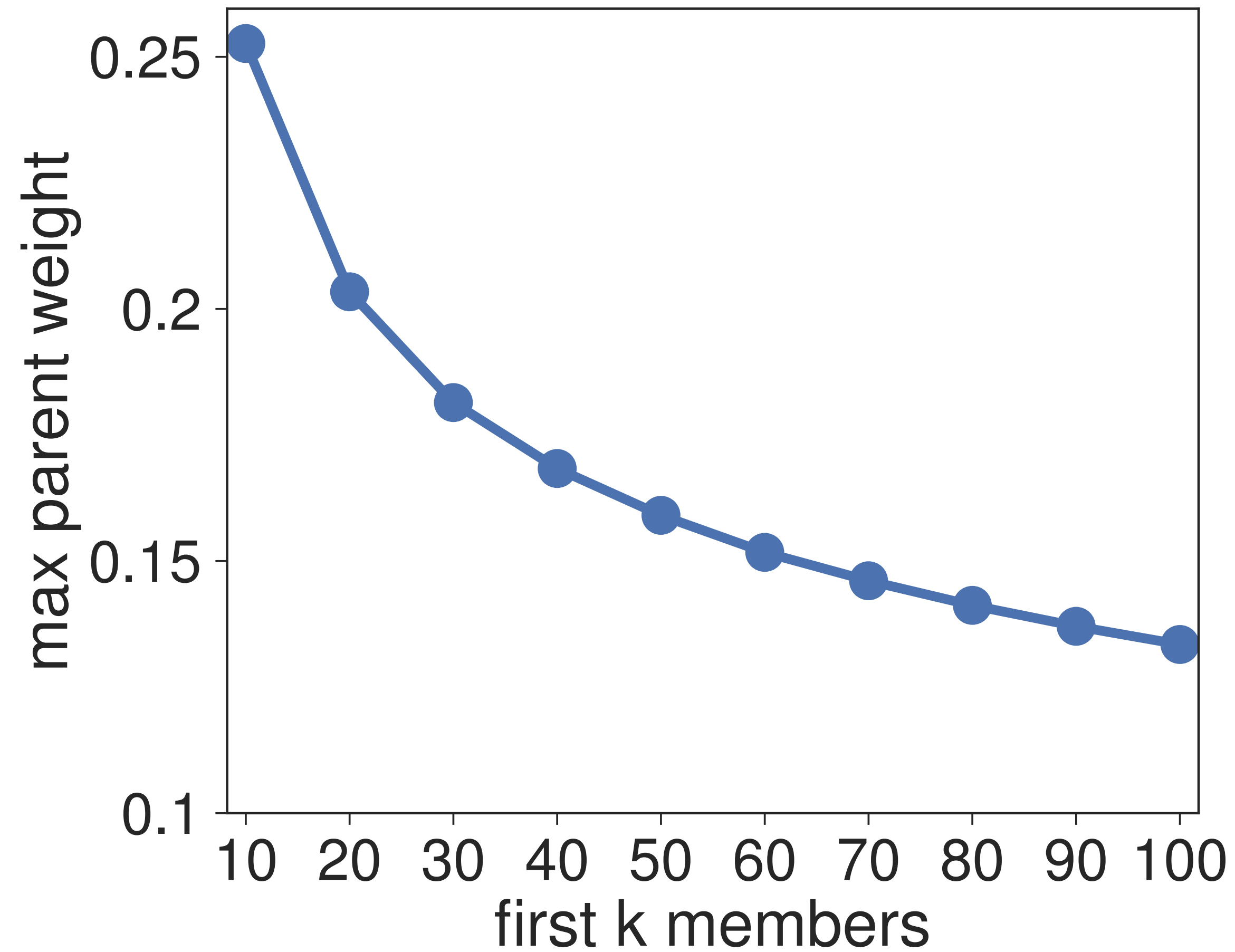
# Characterizing the genealogy graphs

- As a community gathers more users and matures ( $k$  increases), how does that change parent-child relationships?
- How do graph properties evolve as more and more communities emerge ( $t$  increases)?
- Number of parents
- Maximum parent weight

# Number of parents increases as a community matures



# Maximum parent weight decreases as a community matures



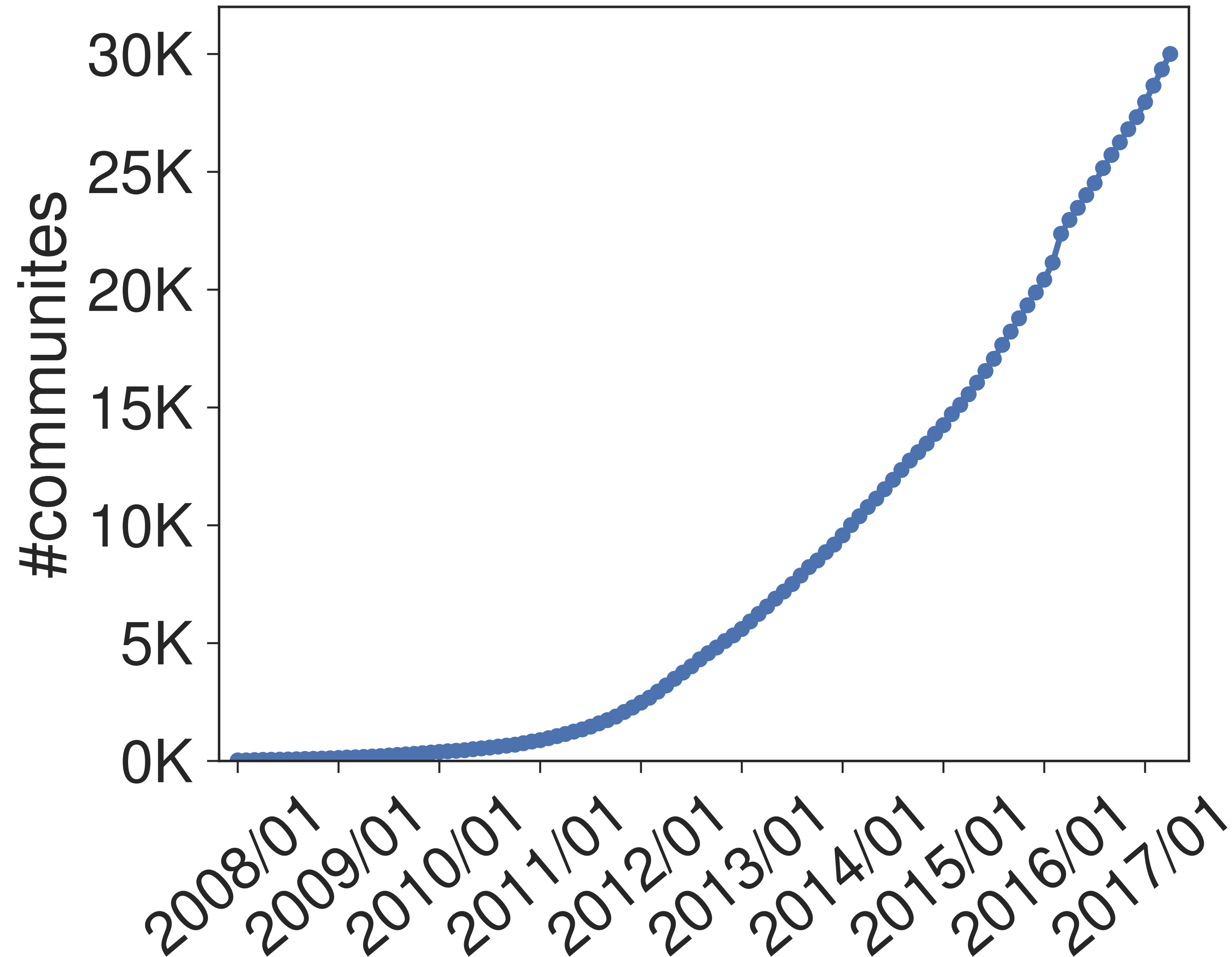


# Ego-networks in genealogy graphs

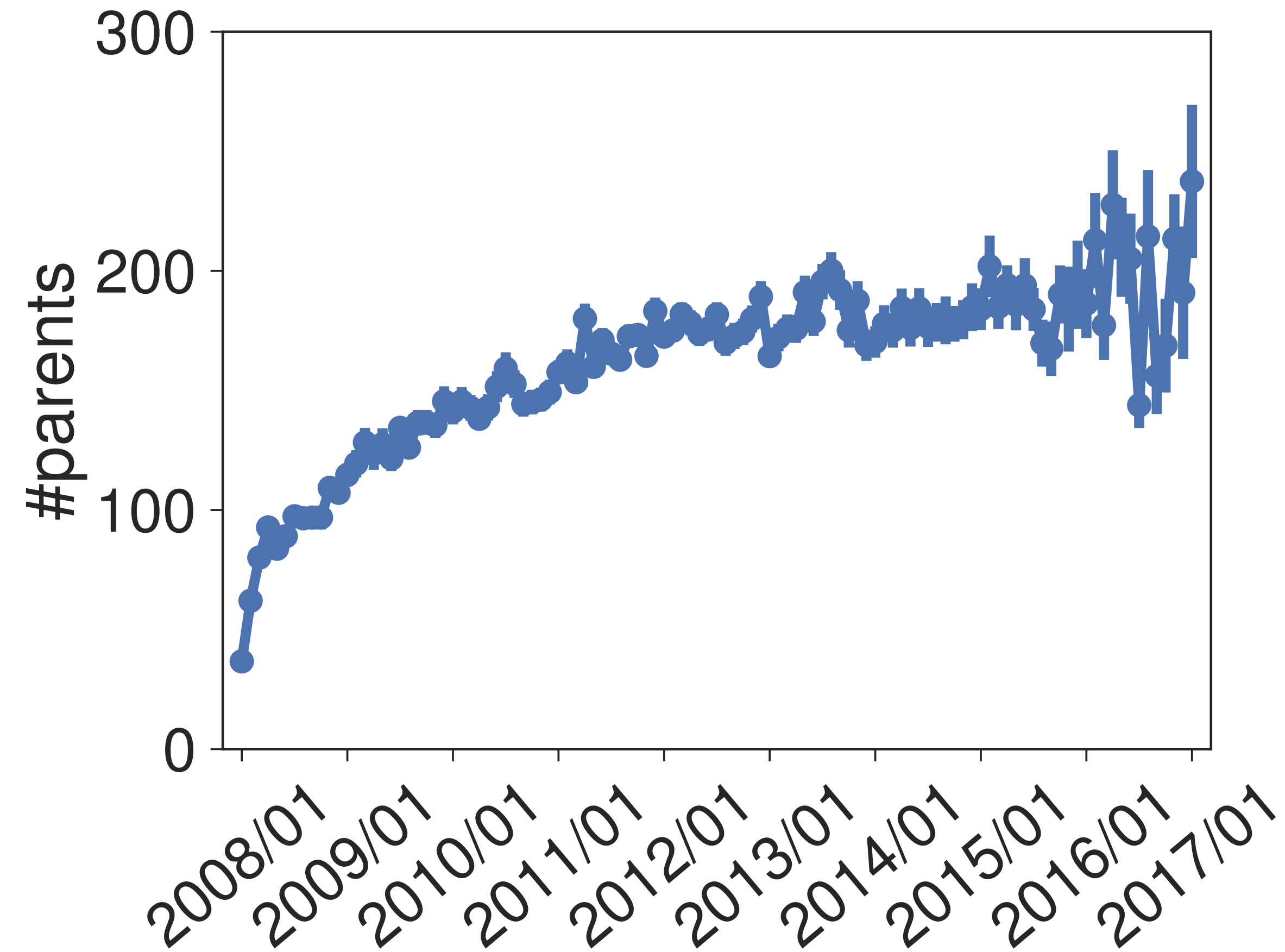
As a community matures

- The number of parents increases
- The influence of any parent decreases
- The fraction of new users increases

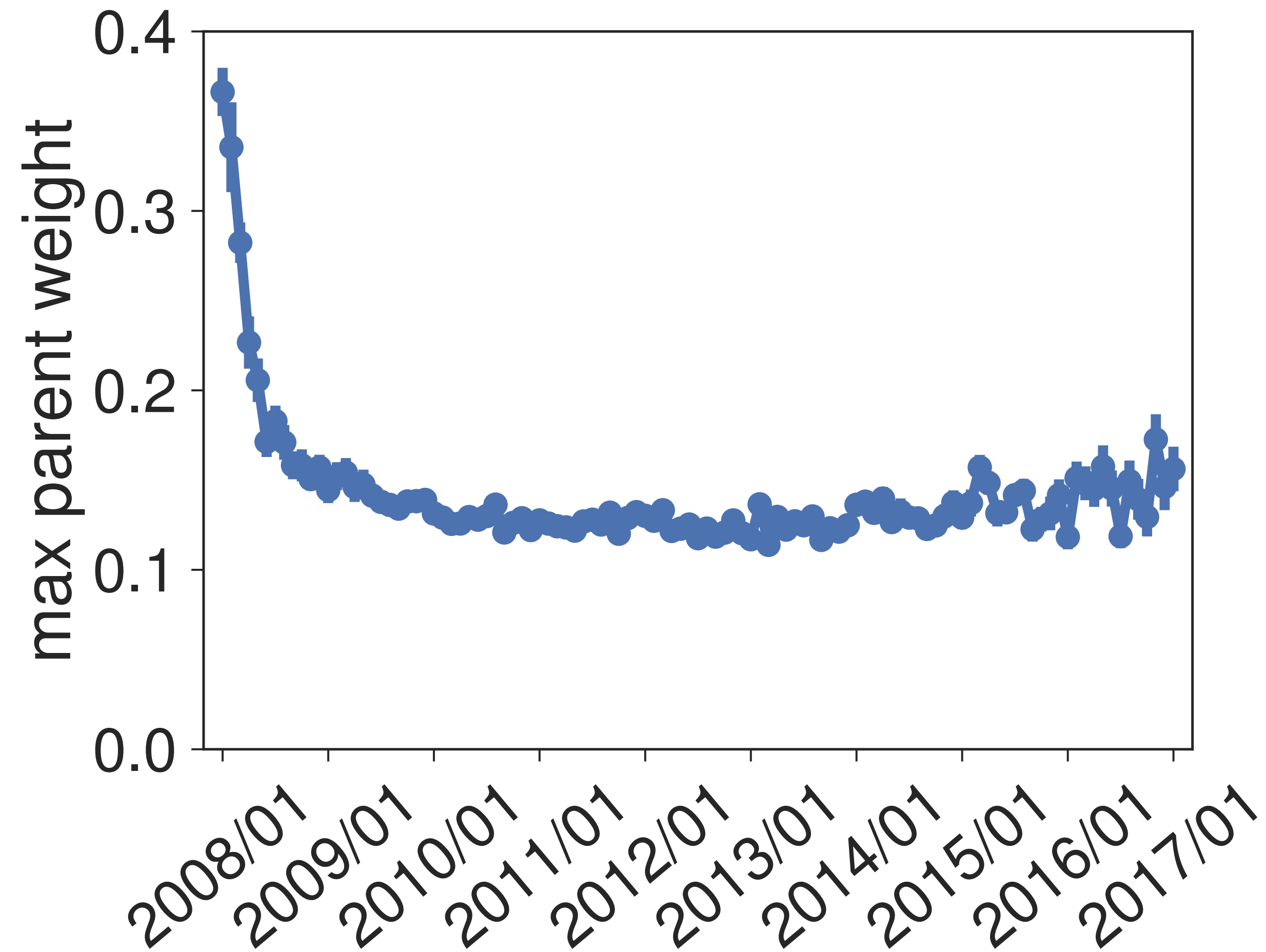
# The number of communities grows rapidly over time



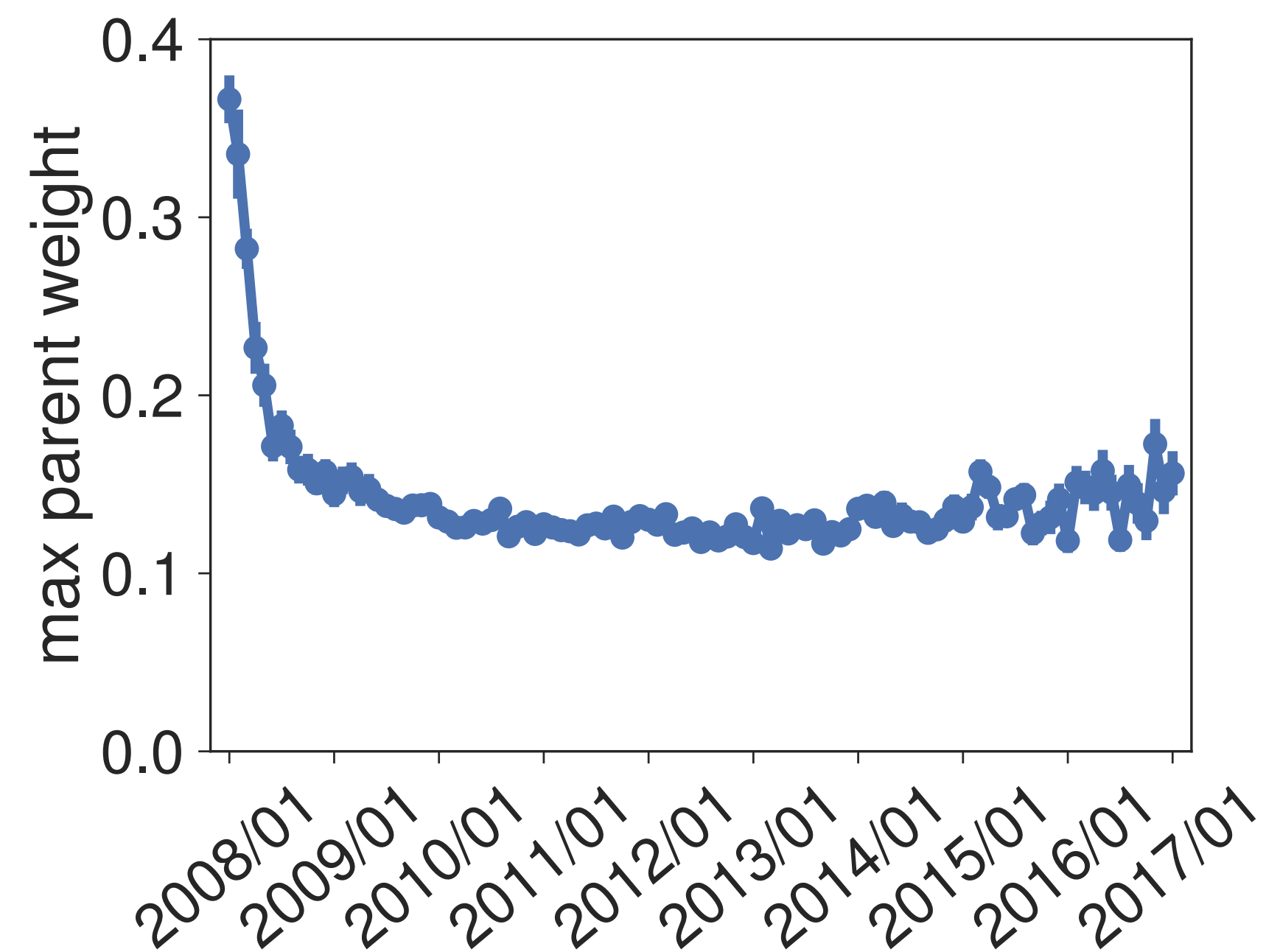
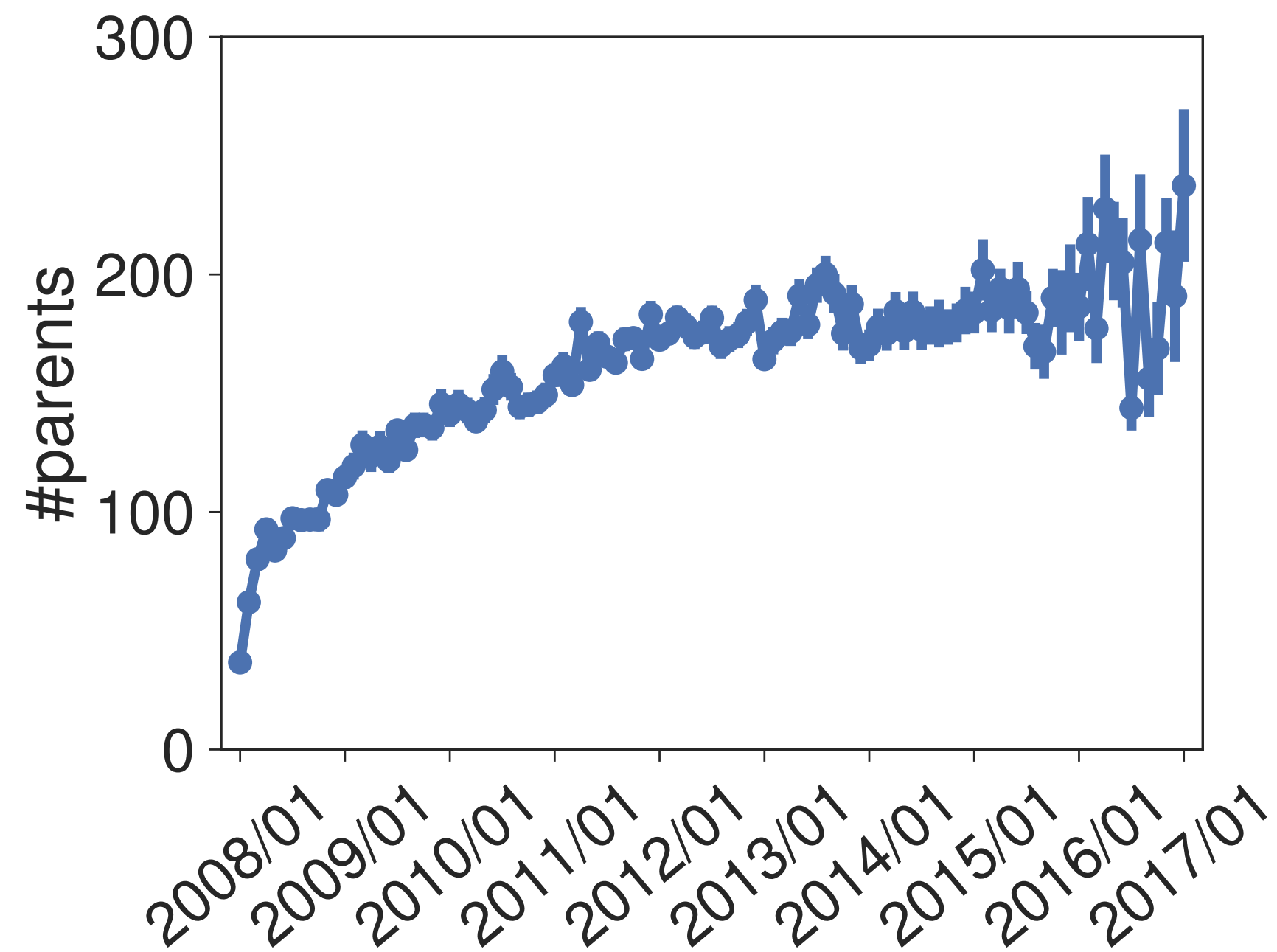
# Number of parents grows and converges to ~180



# Max parent weight declines and converges to 0.1



# Properties of genealogy graphs converge quickly despite the rapid growth in the number of communities





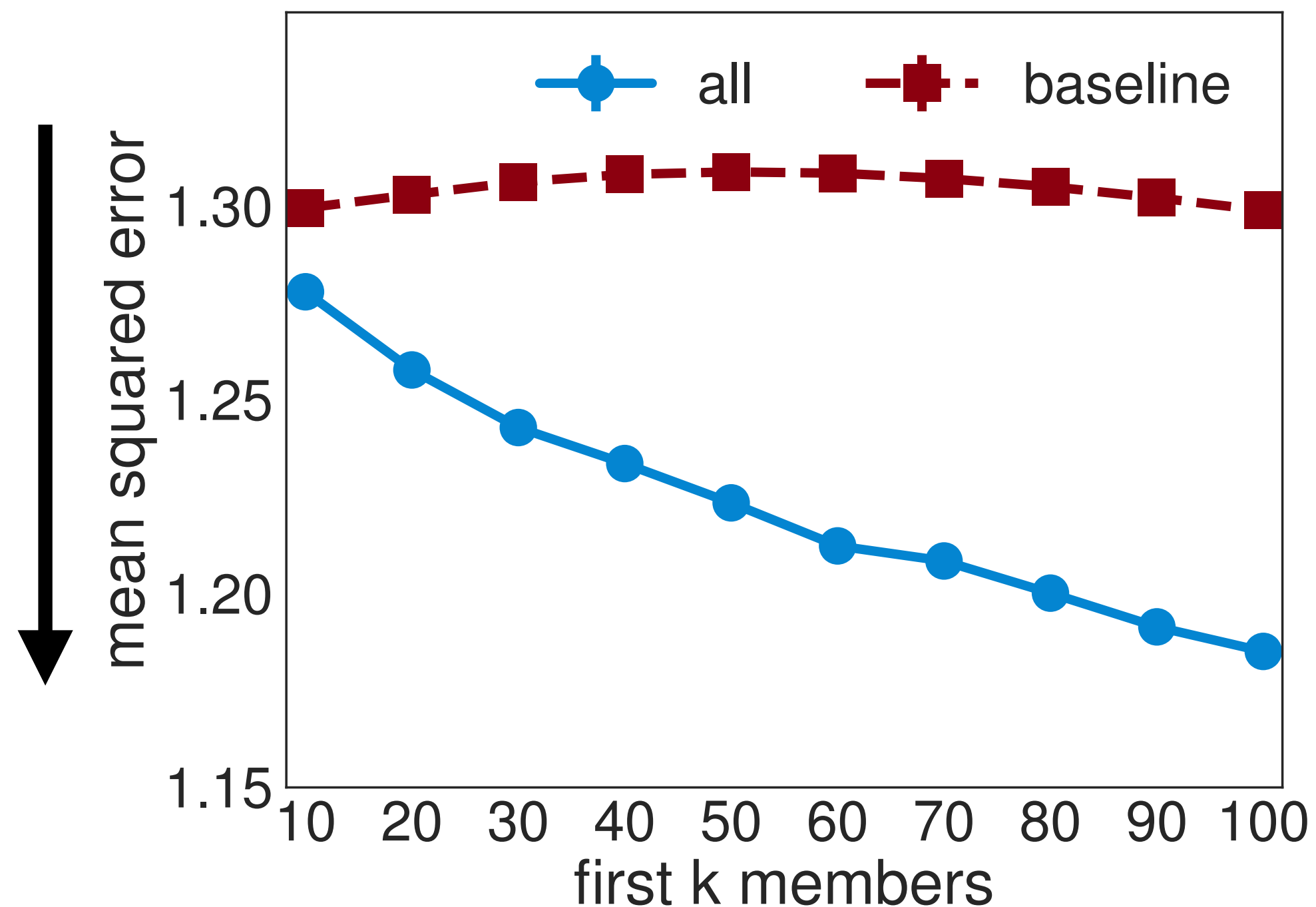
# Predicting future community growth

Measuring community growth [Cheng et al. 2014]

- Does the community size exceed the median?
- How long does it take?

- Temporal features (baseline) [Cheng et al. 2014; Kairam, Wang, and Leskovec 2012]
- Genealogy related features
  - Basic parent properties
  - Meta information of parents
  - Fraction of new users

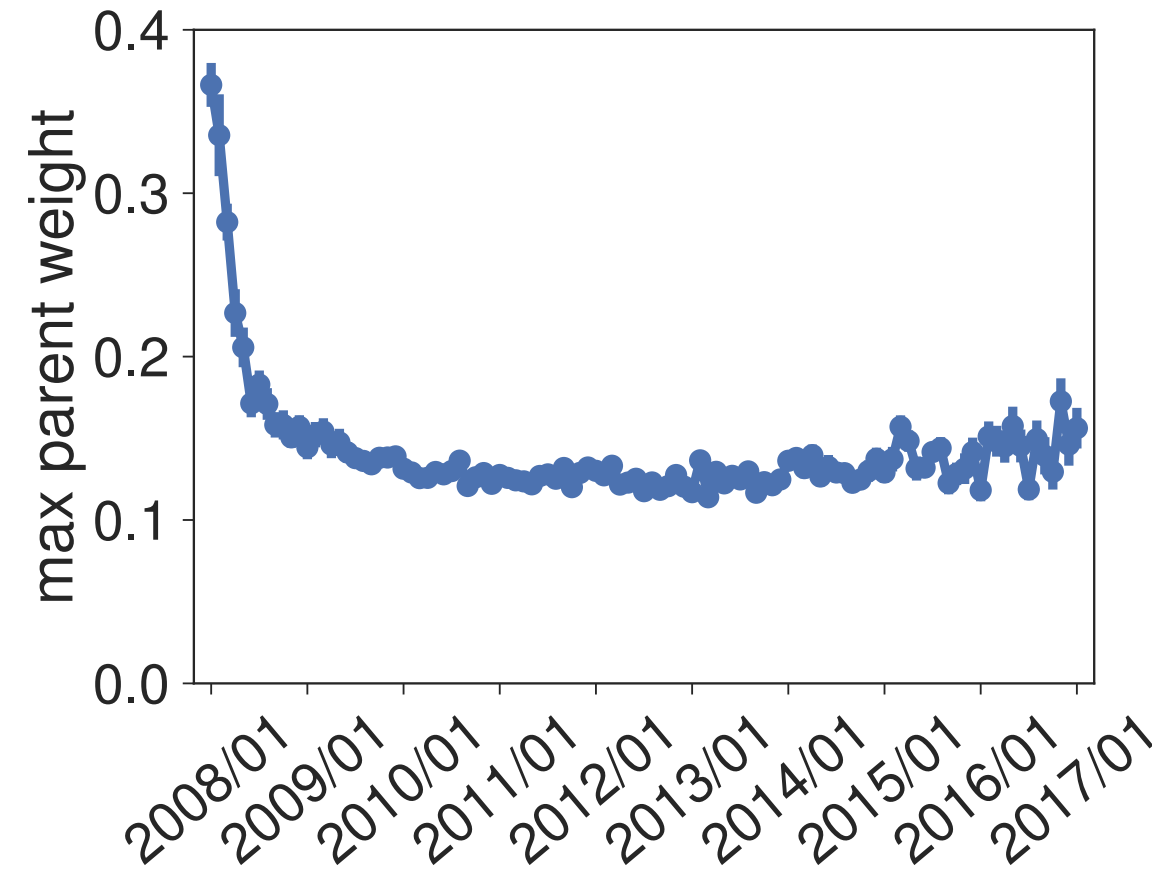
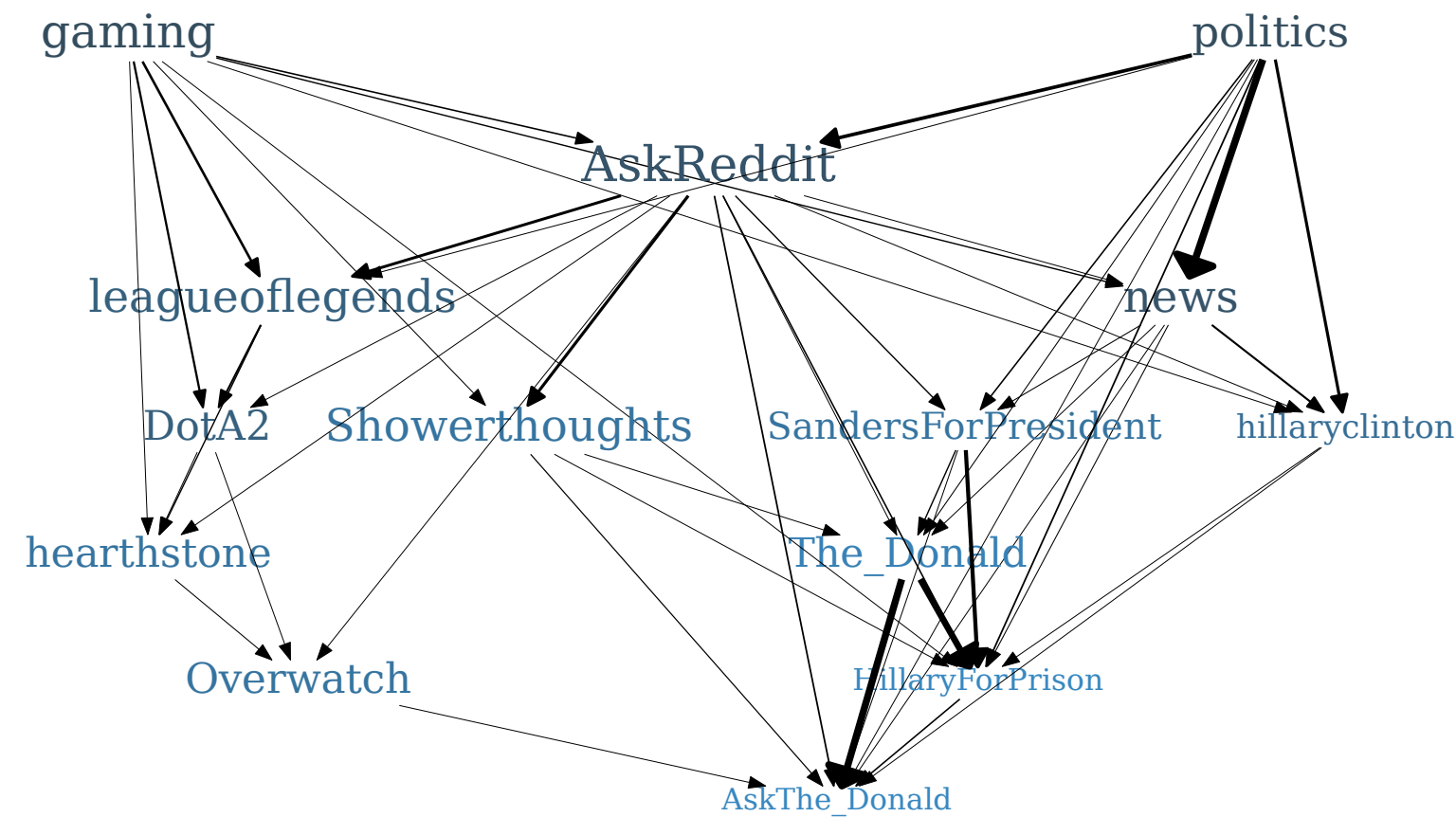
# Prediction performance



**Growth rate regression**

- Strong parent connections are important for future community growth
- The emerging process of a community echoes the idea of complex contagion [Centola and Macy 2007; Fink et al. 2015; Romero, Meeder, and Kleinberg 2011]

# Takeaway



Predict future  
community growth

Predict early adopters

First attempt to trace  
community genealogy

Intriguing convergence of  
graph properties over time

*Demo: <https://redditvisualization.herokuapp.com/>  
(thanks to Kumar Srinivasan)*

*Data & paper: <https://chenhaot.com/papers/community-genealogy.html>*

*@ChenhaoTan, [chenhao@chenhaot.com](mailto:chenhao@chenhaot.com)*