The Size Conundrum

Why Online Knowledge Markets Can Fail at Scale

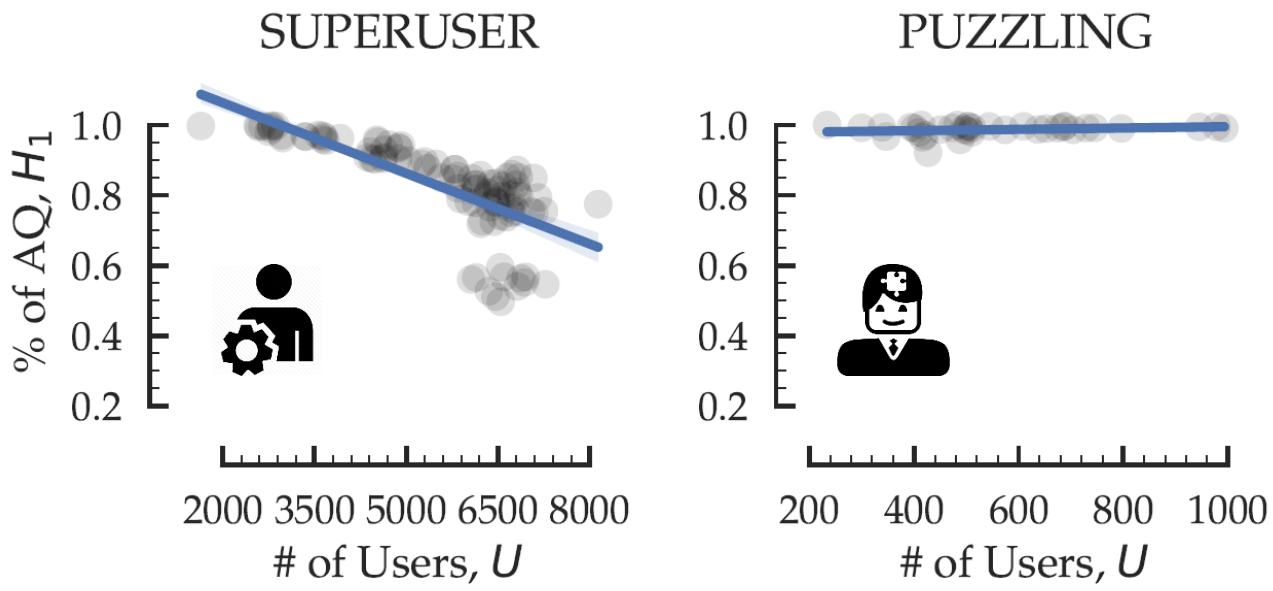
Himel Dev, Chase Geigle, Qingtao Hu, Jiahui Zheng, Hari Sundaram

University of Illinois at Urbana–Champaign

ILLINOIS

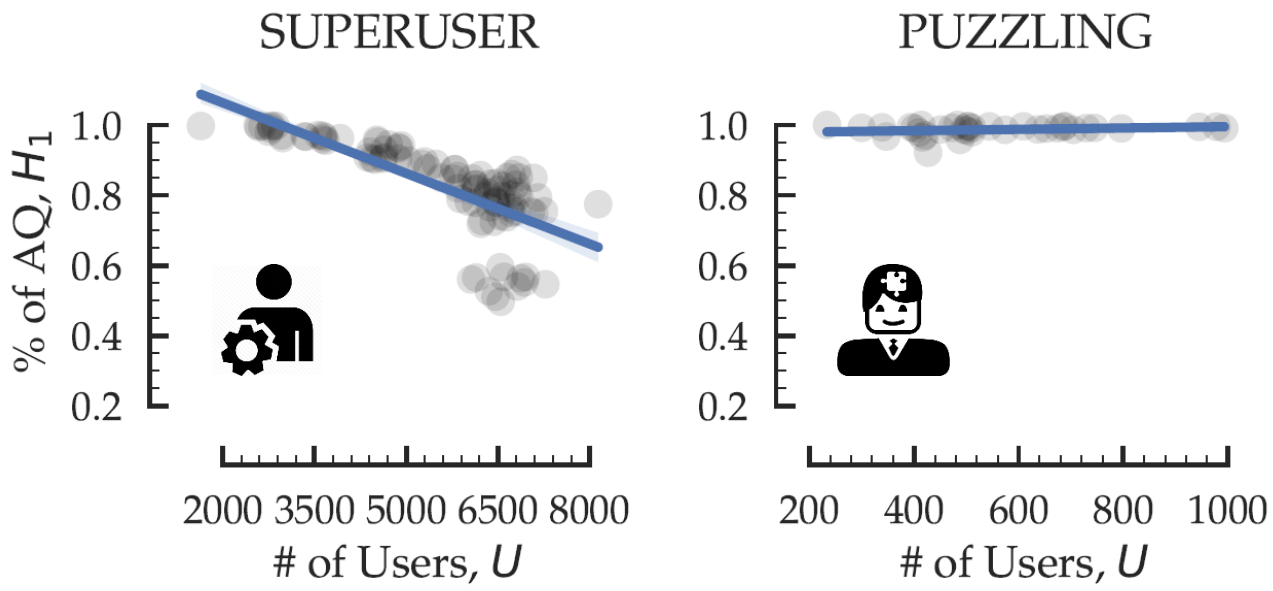
Let's examine user growth at two StackExchanges: **superuser** & **puzzling**.

Why is superuser **failing at scale**, while puzzling continues to thrive?



What should you do as a social media operator?

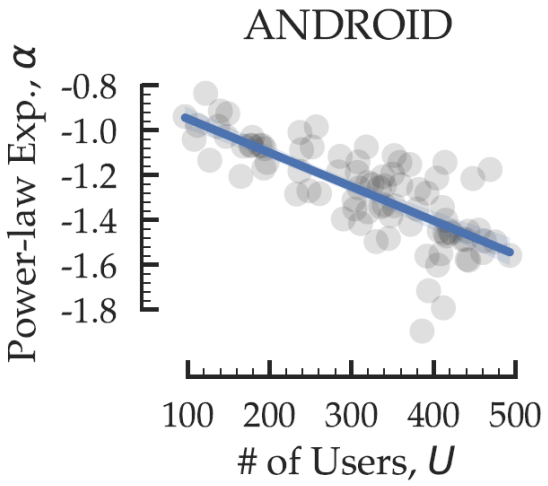
Should you **expand** your platform, **or not**?



Our key idea: **social media** platforms are **markets** where participants exchange information

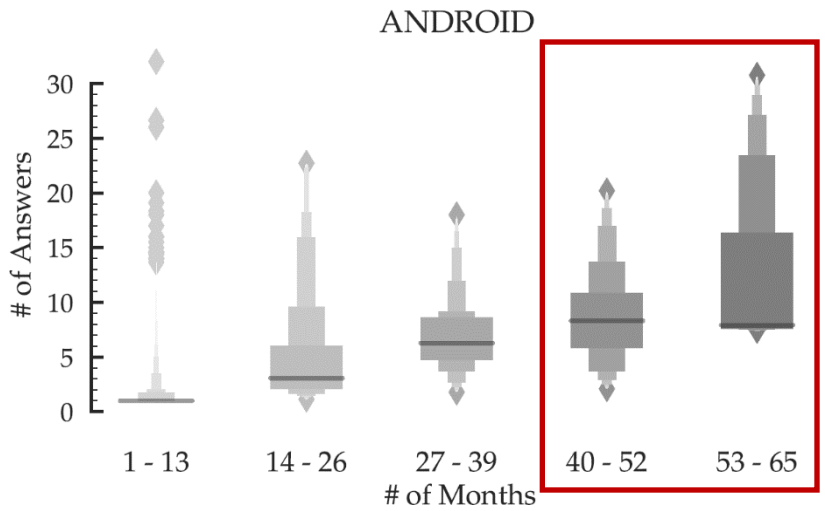
Findings

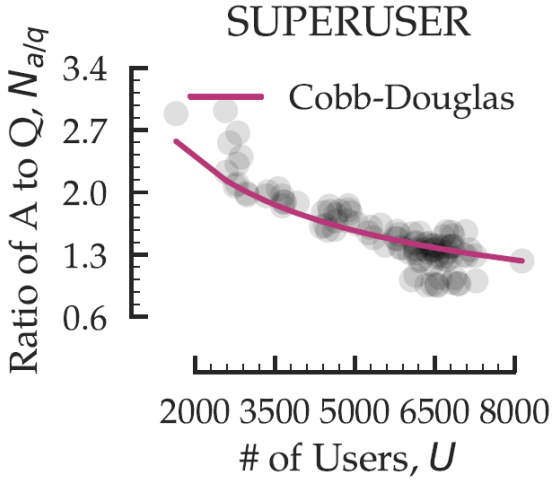
Our analysis provides three critical insights: **stable core**, **size dependent distribution**, and **diseconomies of scale**



Size Dependent Distribution

Stable Core

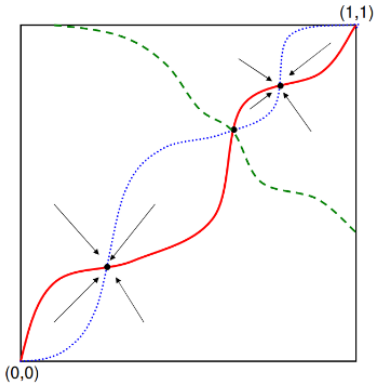




Diseconomies of Scale

Related Work

1. Evolution of two-sided markets. Ravi Kumar, Yury Lifshits, and Andrew Tomkins. WSDM (2010).



2. Knowledge Market Design: A Field Experiment at Google Answers. Yan Chen, Teck-Hua Ho, and Yong-Mi Kim. Journal of Public Economic Theory (2009).

Table 4: Determinants of Answer Length (Effort)

	Dependent Variable: Word Count		
	(1) IPL	(2) GA	(3) All
Price	7.472	13.097	12.575
	(23.035)	(2.545)***	(2.128)***
Tip	16.862	25.519	27.115
	(21.164)	(19.357)	(13.796)*
Reputation	1,368.709	1,073.285	1,143.625
	(434.286)***	(613.795)*	(413.810)***
Experience	-0.244	-0.136	-0.168
	(0.128)*	(0.126)	(0.095)*
Constant	-5,083.371	-4,259.175	-4,448.396
	(2,011.381)**	(2,687.905)	(1,801.455)**
Observations	75	125	200

Motivation and Problem Statement

Modeling Knowledge Markets

Knowledge Market Characteristics

Knowledge Market Failures

Implications and Summary

Market

Information Market

Social media enable users to exchange information

Problem Statement

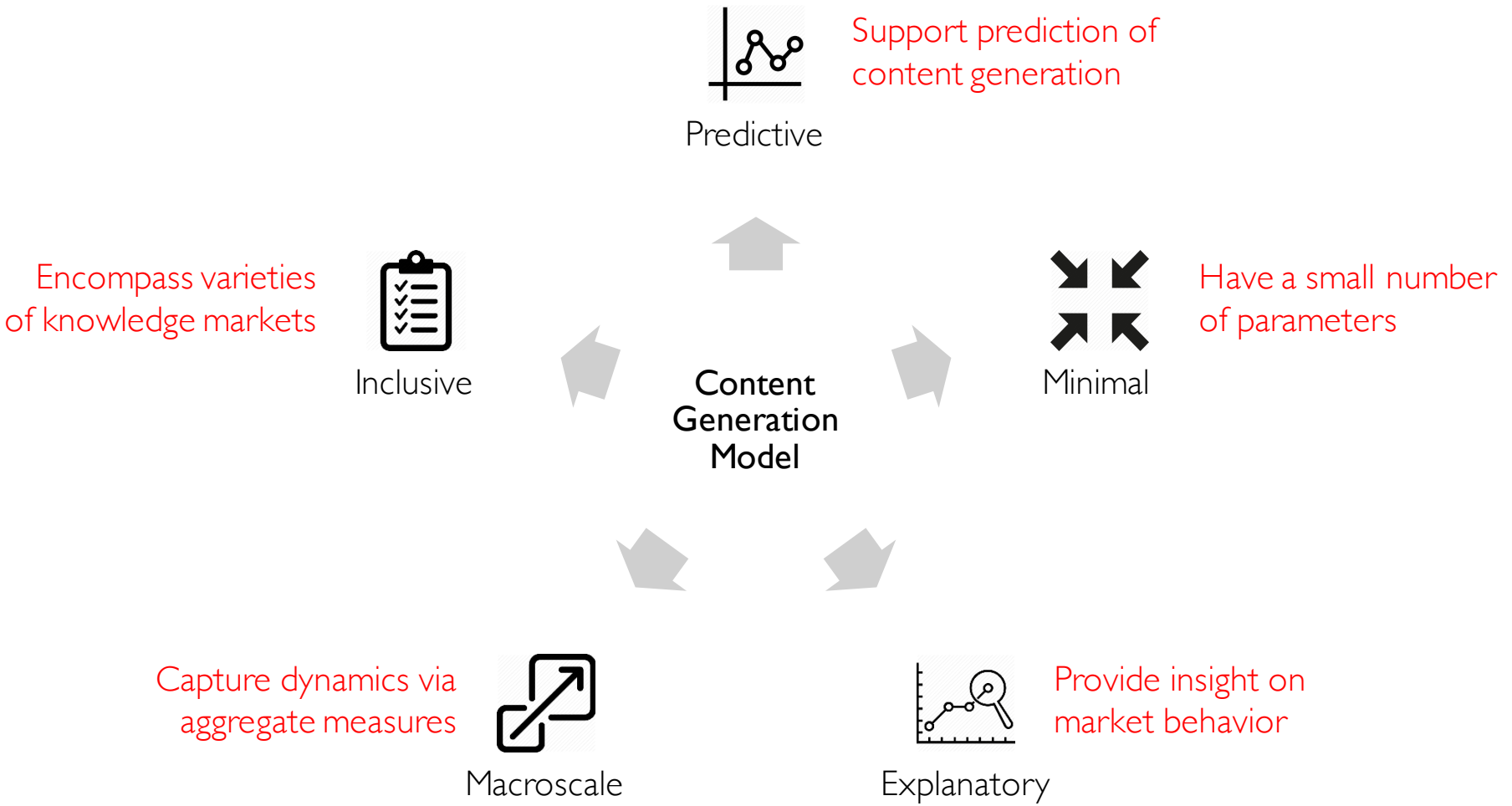
We concentrate on a particular type of social media platform—community question answering (CQA) websites that we interpret as **knowledge markets**.

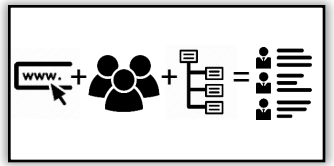
Our **goal** is to design a model for **content generation** in knowledge markets, and reason about market **sustainability**.

What factors drive the production and consumption processes?

How can we predict future production and consumption?

Modeling Desiderata





Social Media Platform + User Participation + Content Dependency = Social Media Content

Content Generation in StackExchange

Social Media Content

Content Dependency

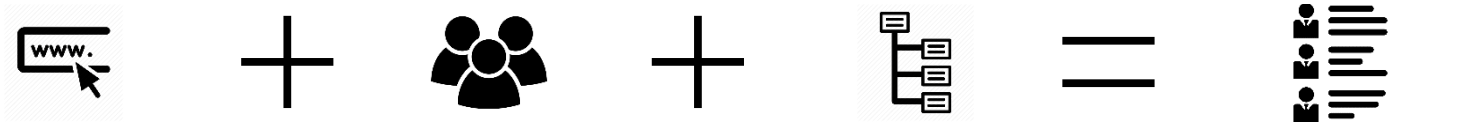
Social Media Platform

User Participation

Model Overview

We model content generation using Macroeconomic production functions which concentrate on key **factors**.

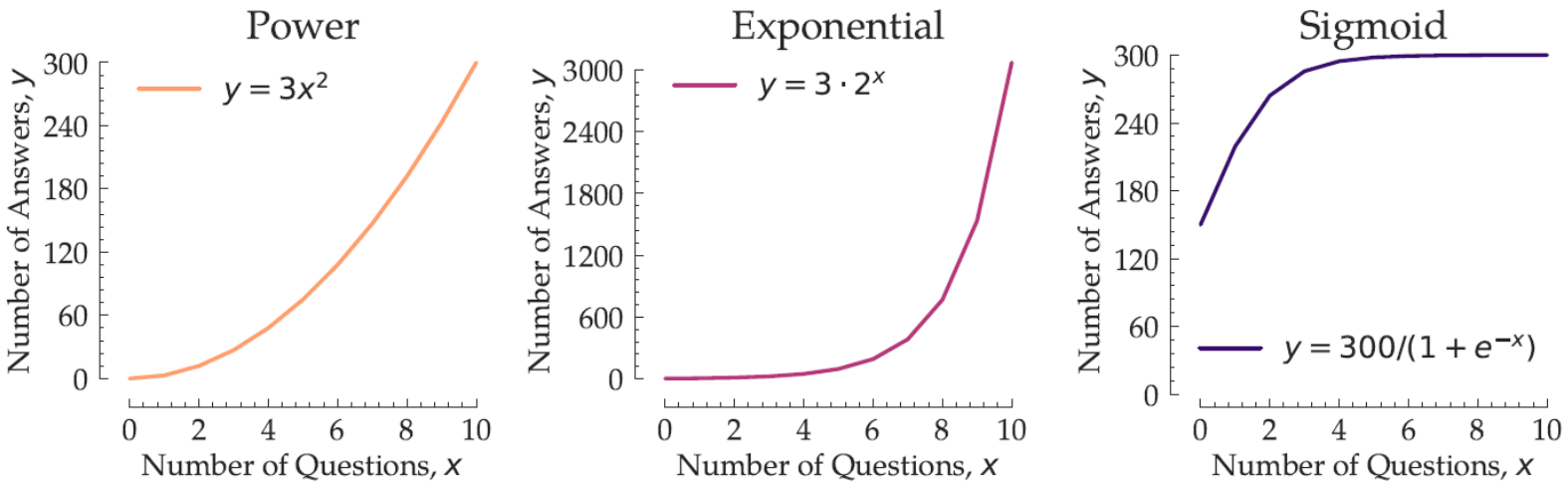
We combine a **basis function** and an **interaction type** to capture these factors.



Social Media Platform + User Participation + Content Dependency = Social Media Content

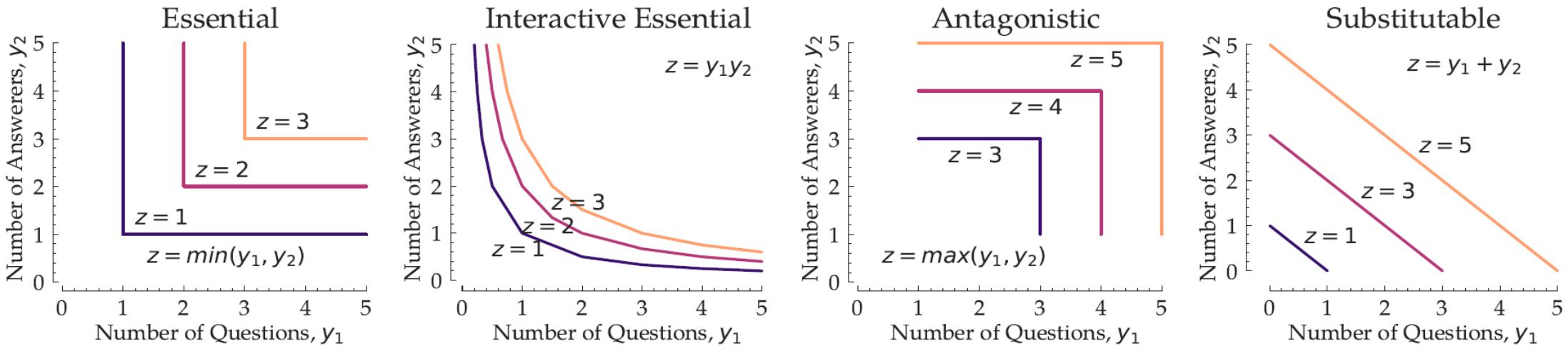
Basis Function

Basis function captures the relationship between an **input** and **output**

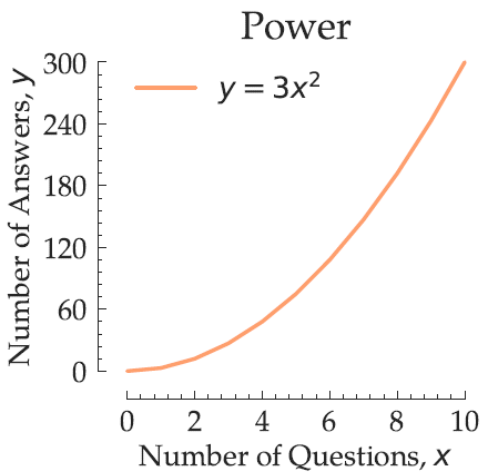


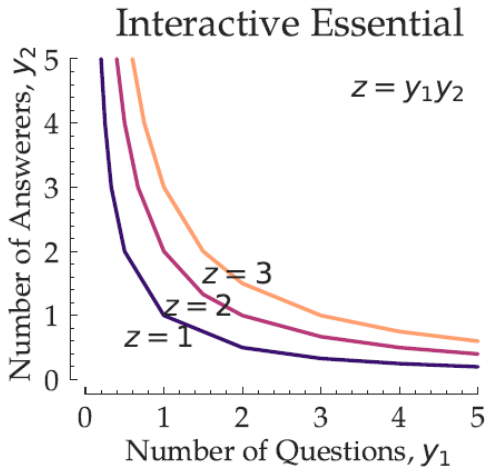
Interaction Type

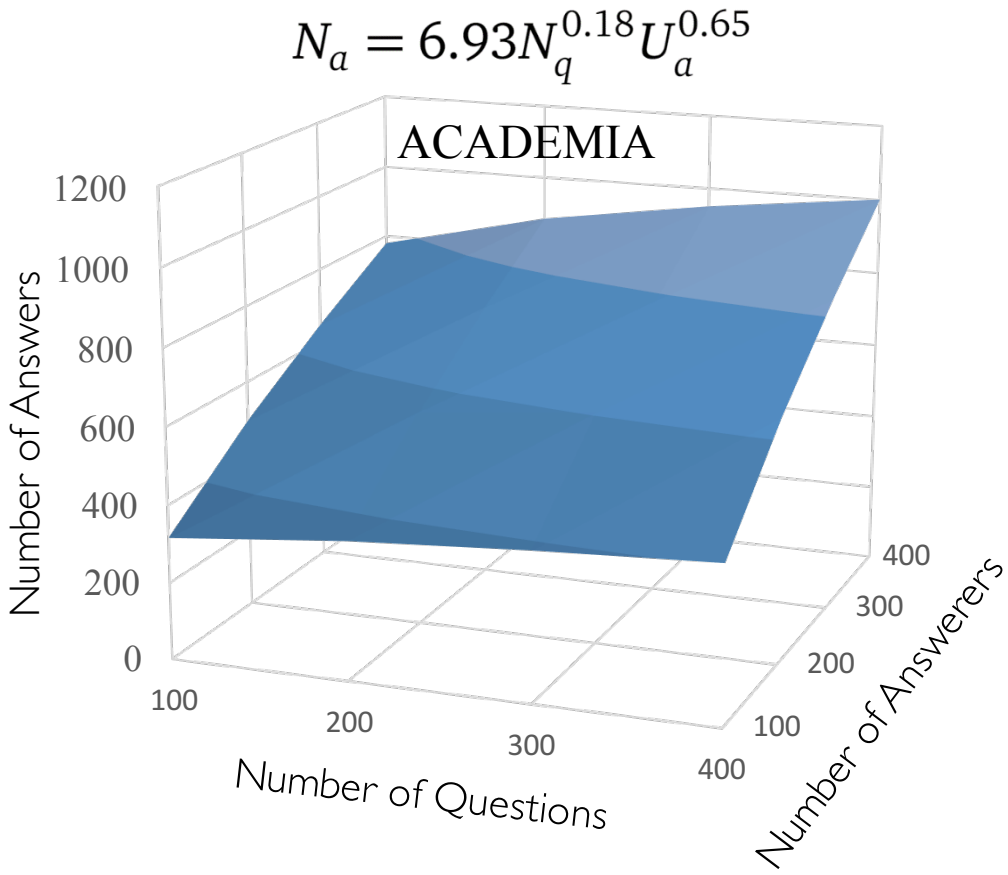
Interaction type captures how different inputs interact to generate output



Cobb-Douglas Model

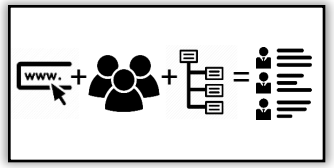


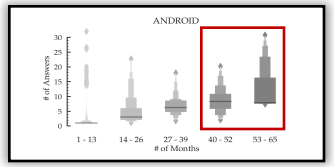




For an output z with two inputs x_1 and x_2, the Cobb-Douglas function is

$$z = a x_1^{\lambda_1} x_2^{\lambda_2}$$





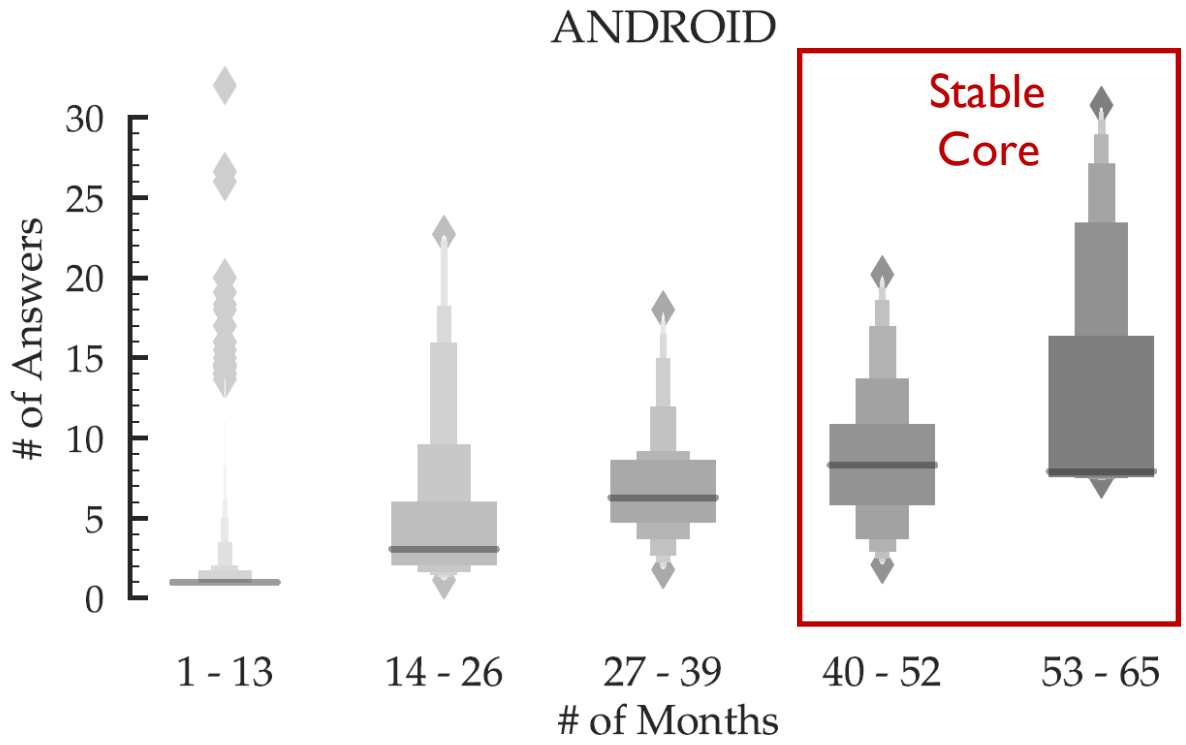
Motivation and Problem Statement

Modeling Knowledge Markets

Knowledge Market Characteristics

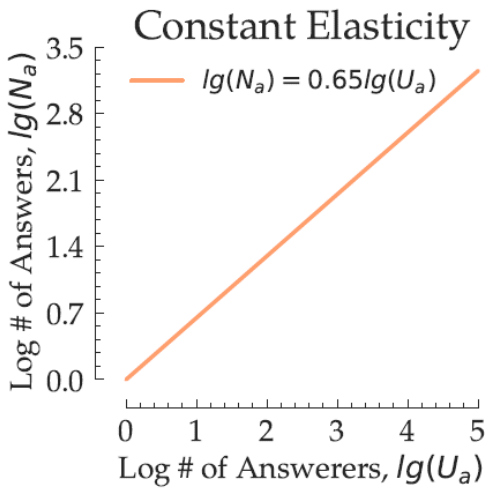
Knowledge Market Failures

Conclusion and Future Work

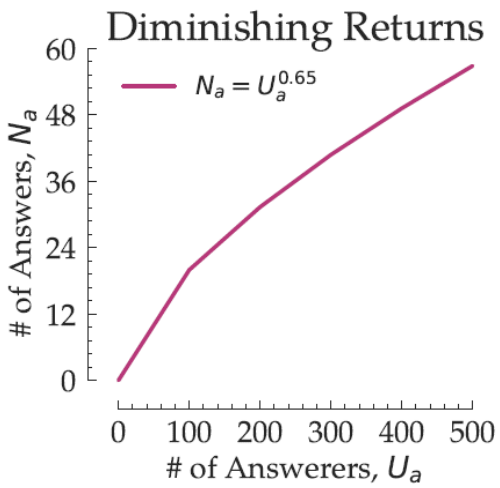


Model Interpretation

A percentage increase in **any** of the inputs leads to a constant percentage increase in the output

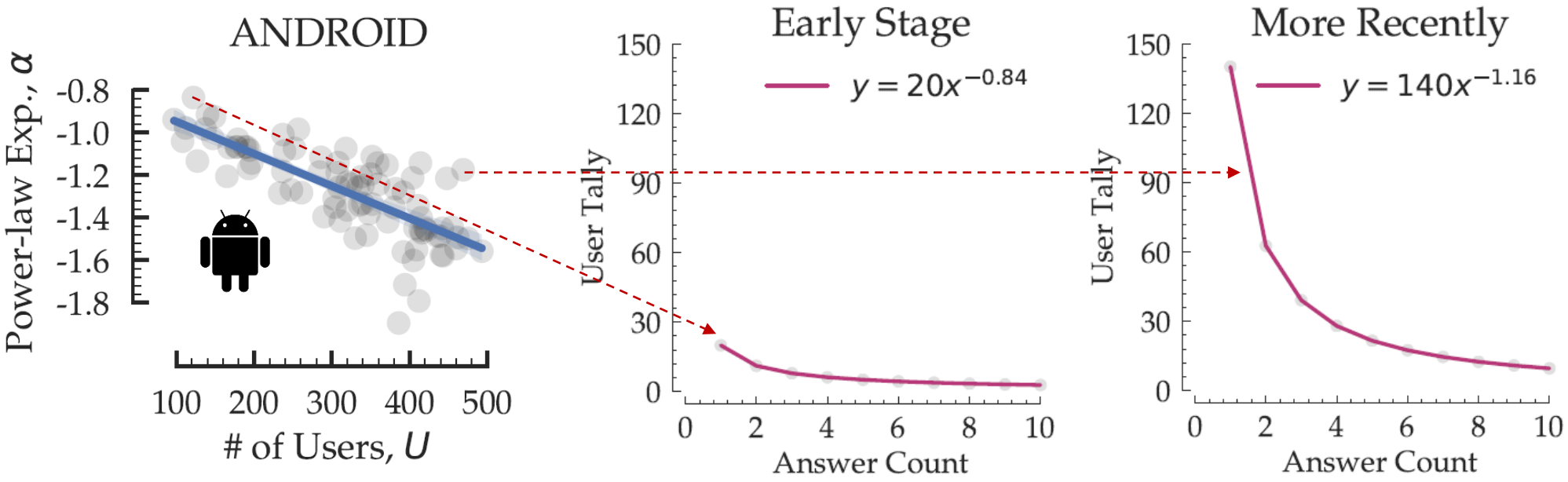


The **marginal** output decreases as an input is incrementally increased



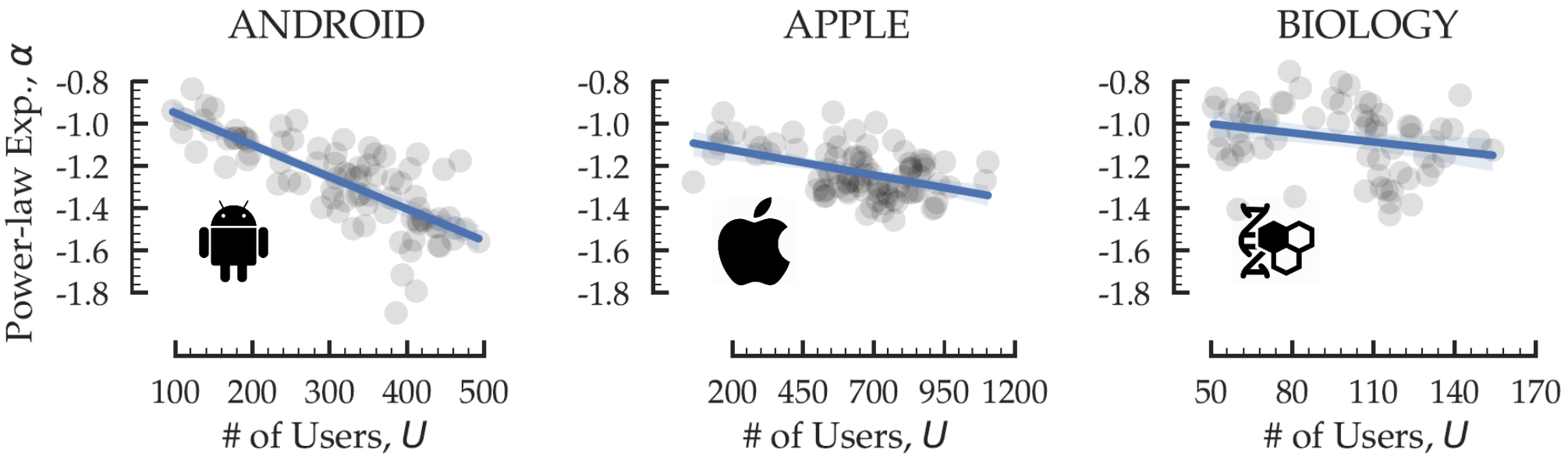
Size Dependent Distribution

The distribution of participant activities **evolves** with the number of participants.



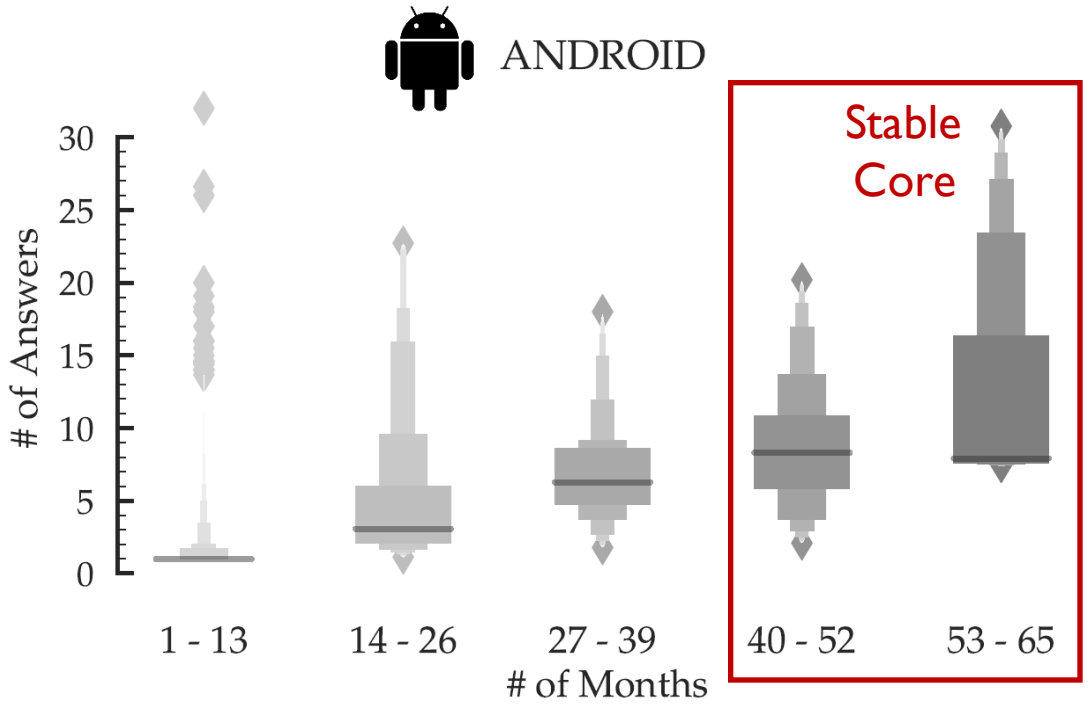
Size Dependent Distribution

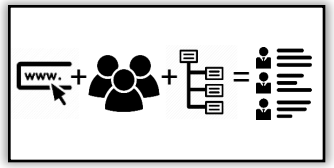
The distribution of participant activities **evolves** with the number of participants.

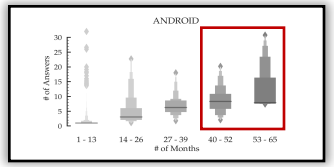


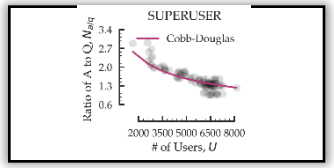
Stable Core

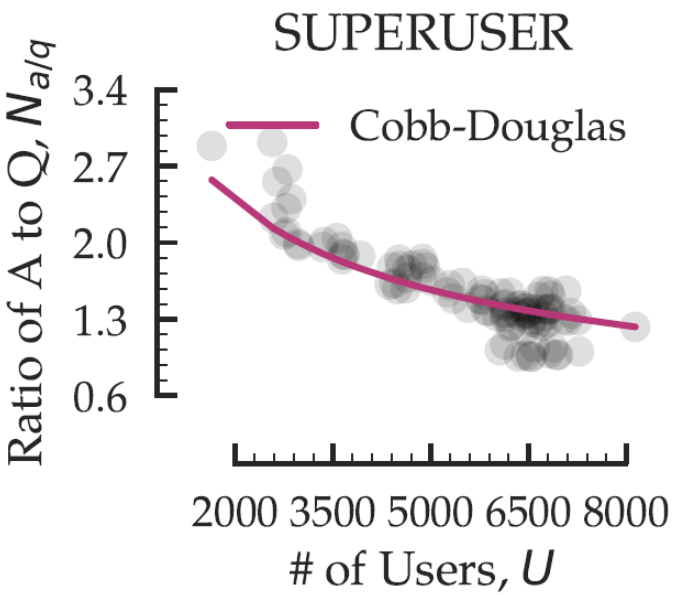
Stable core: users who contribute **many** answers for a **long** period





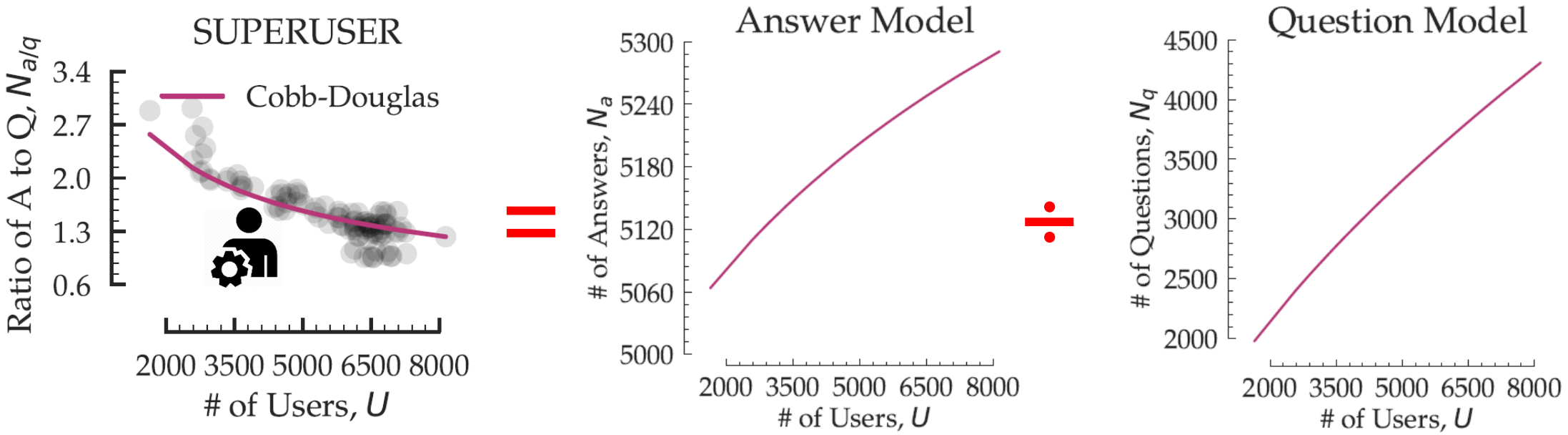






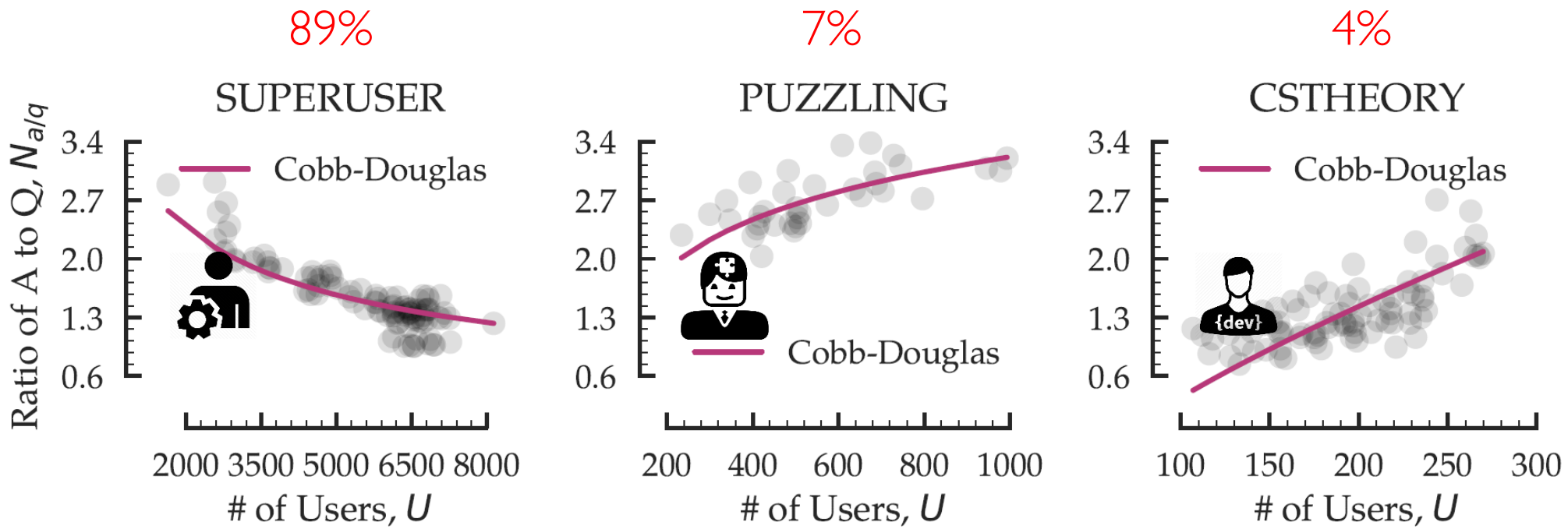
Dis(economies) of Scale

The ratio of answers to questions ($N_{a/q}$) **decreasing** (increasing) with the **increase** in number of users (U).



Dis(economies) of Scale

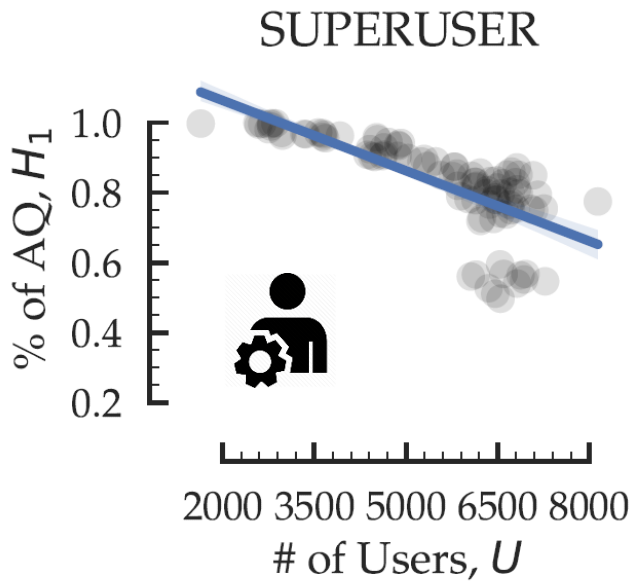
The ratio of answers to questions ($N_{a/q}$) **decreasing** (increasing) with the **increase** in number of users (U).



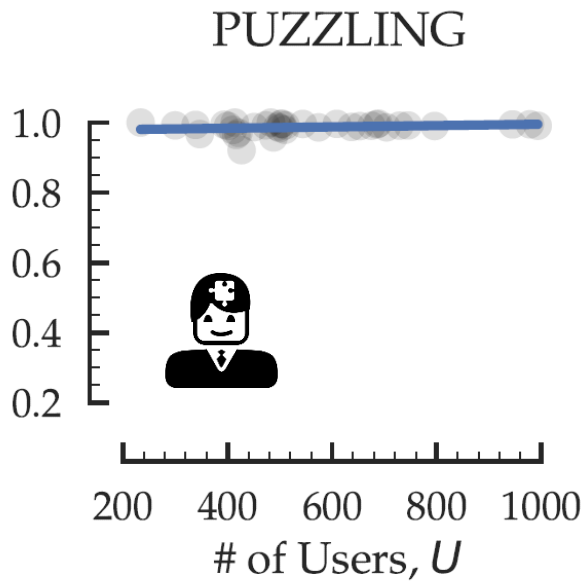
Health at Scale - I

In most StackExchanges, the fraction of questions with at least one answer (H_1) **decreases** with the **increase** in number of users (U).

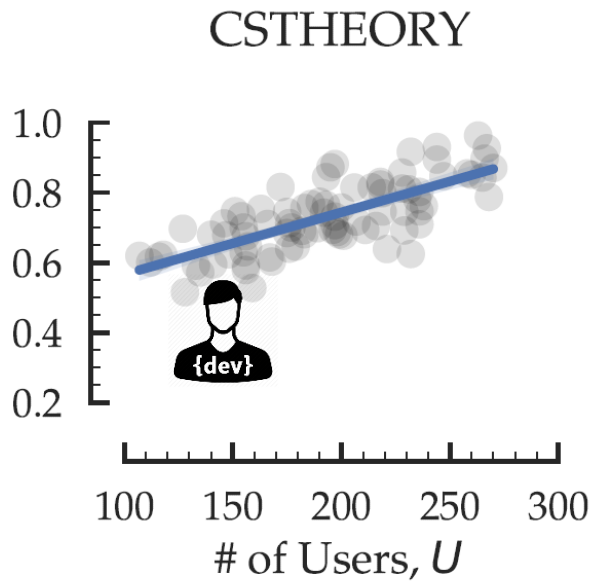
89%



7%



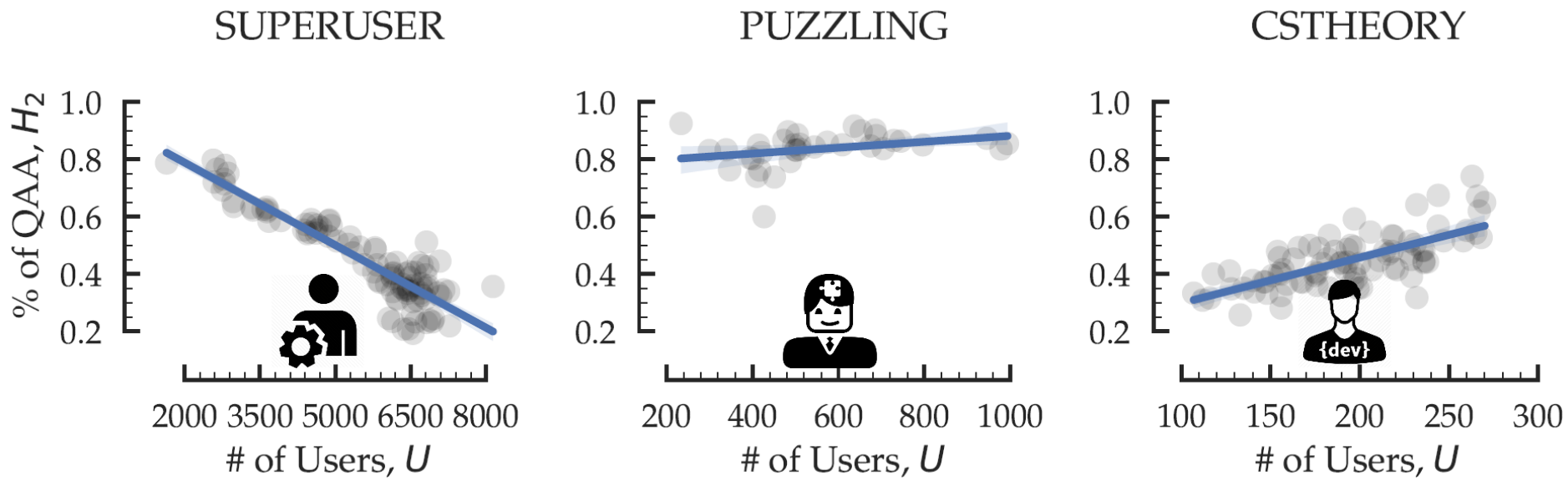
4%

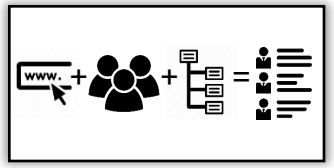


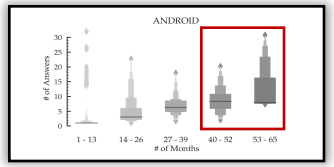
Health at Scale - II

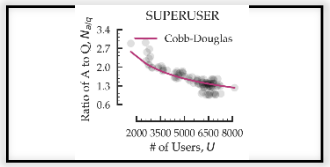
In most StackExchanges, the fraction of questions with an accepted answer (H_2) **decreases** with the **increase** in number of users (U).

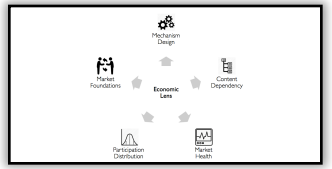
89% 7% 4%











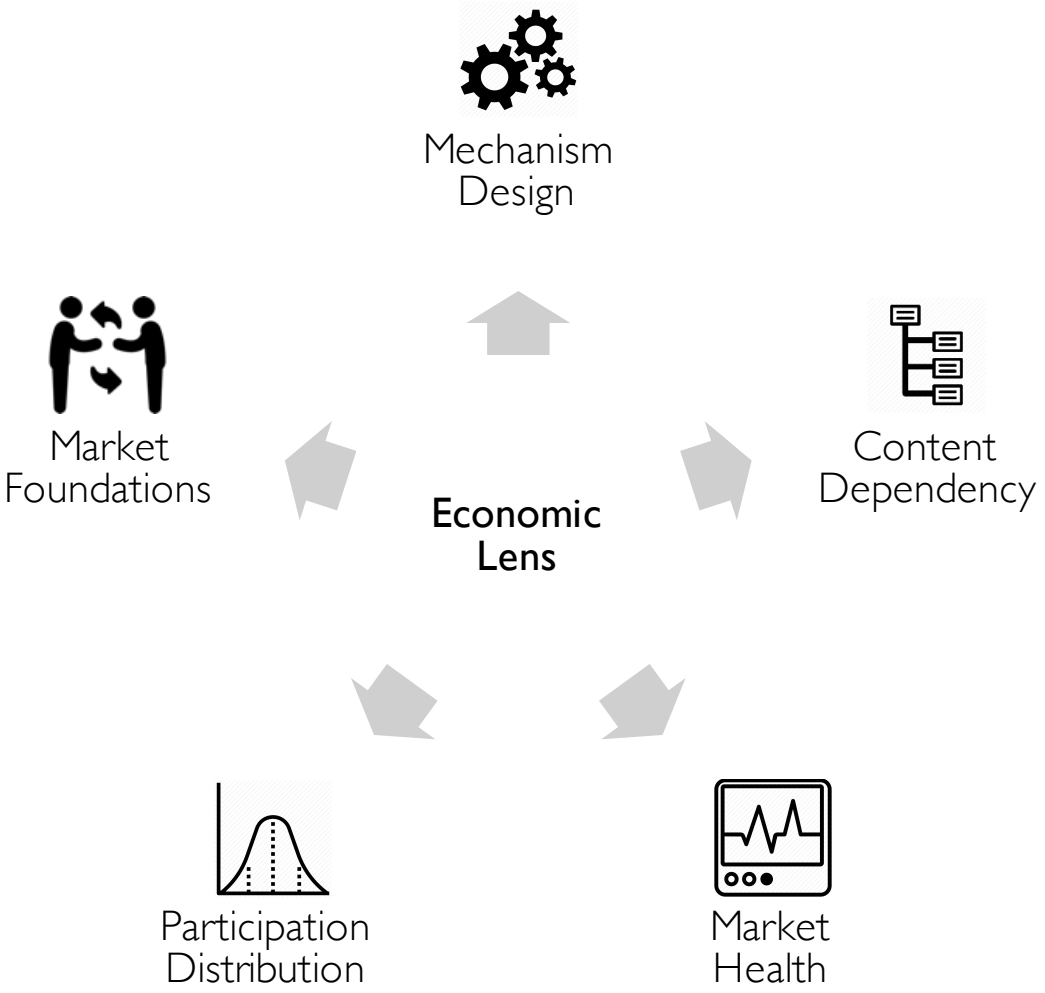
Motivation and Problem Statement

Modeling Knowledge Markets

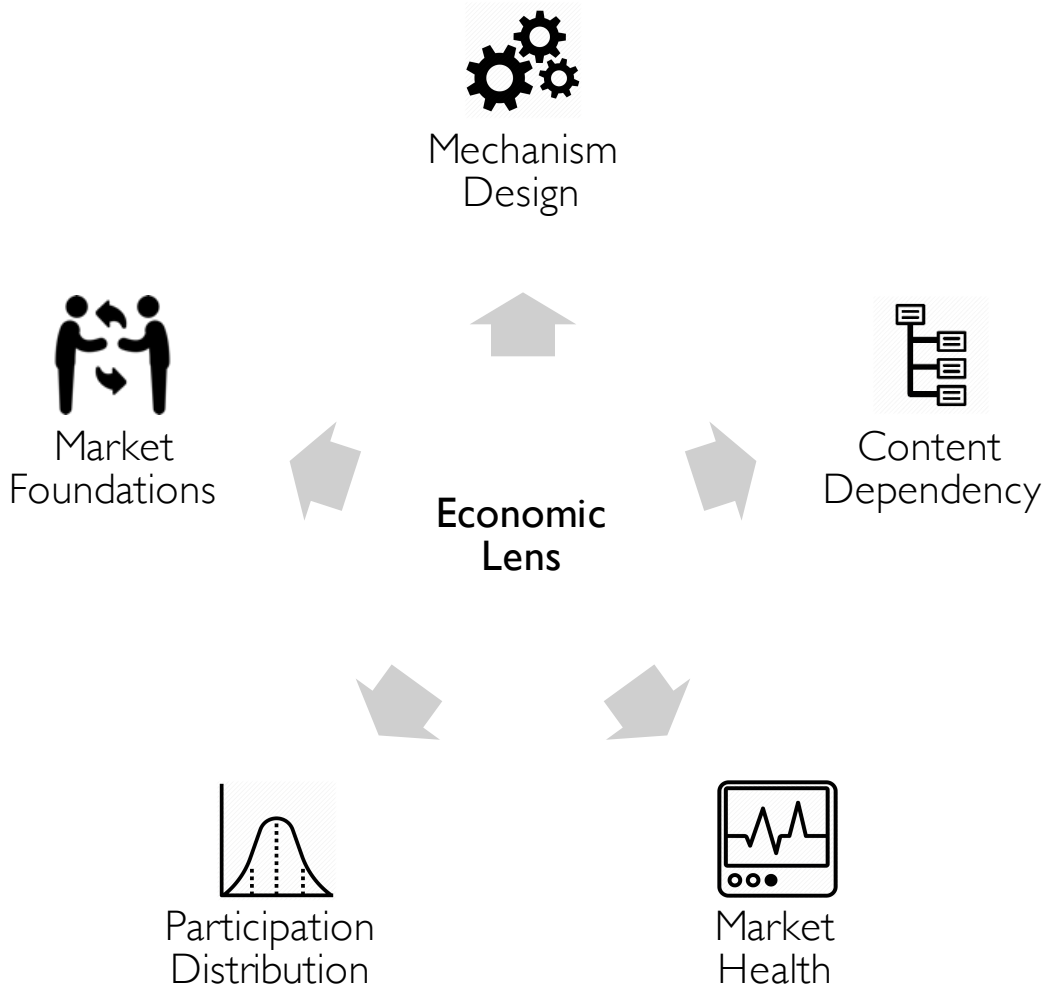
Knowledge Market Characteristics

Knowledge Market Failures

Implications and Summary



Implications

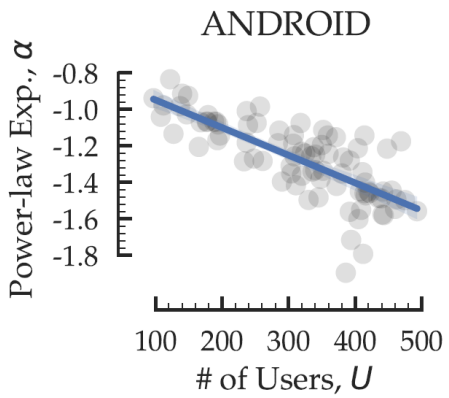


Summary

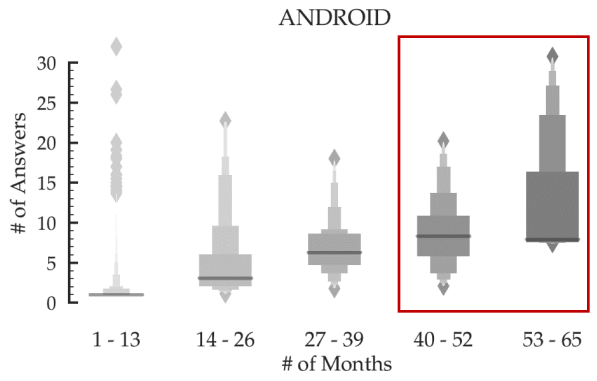
Key Idea:

=

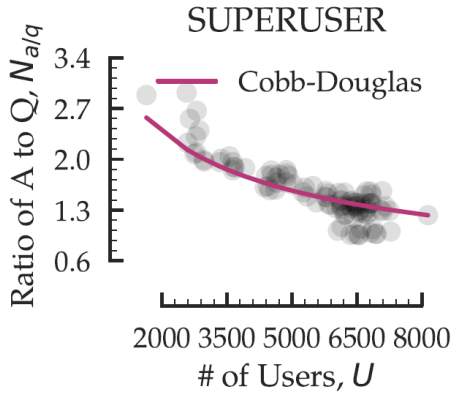
Findings:



Size Dependent Distribution



Stable Core



Diseconomies of Scale

GitHub : https://github.com/information-market/knowledge-market-sustainability