

Is User's Centrality Related to Hardness of Location Estimation?

Toyohashi University of Technology

Shiori Hironaka
s143369@edu.tut.ac.jp

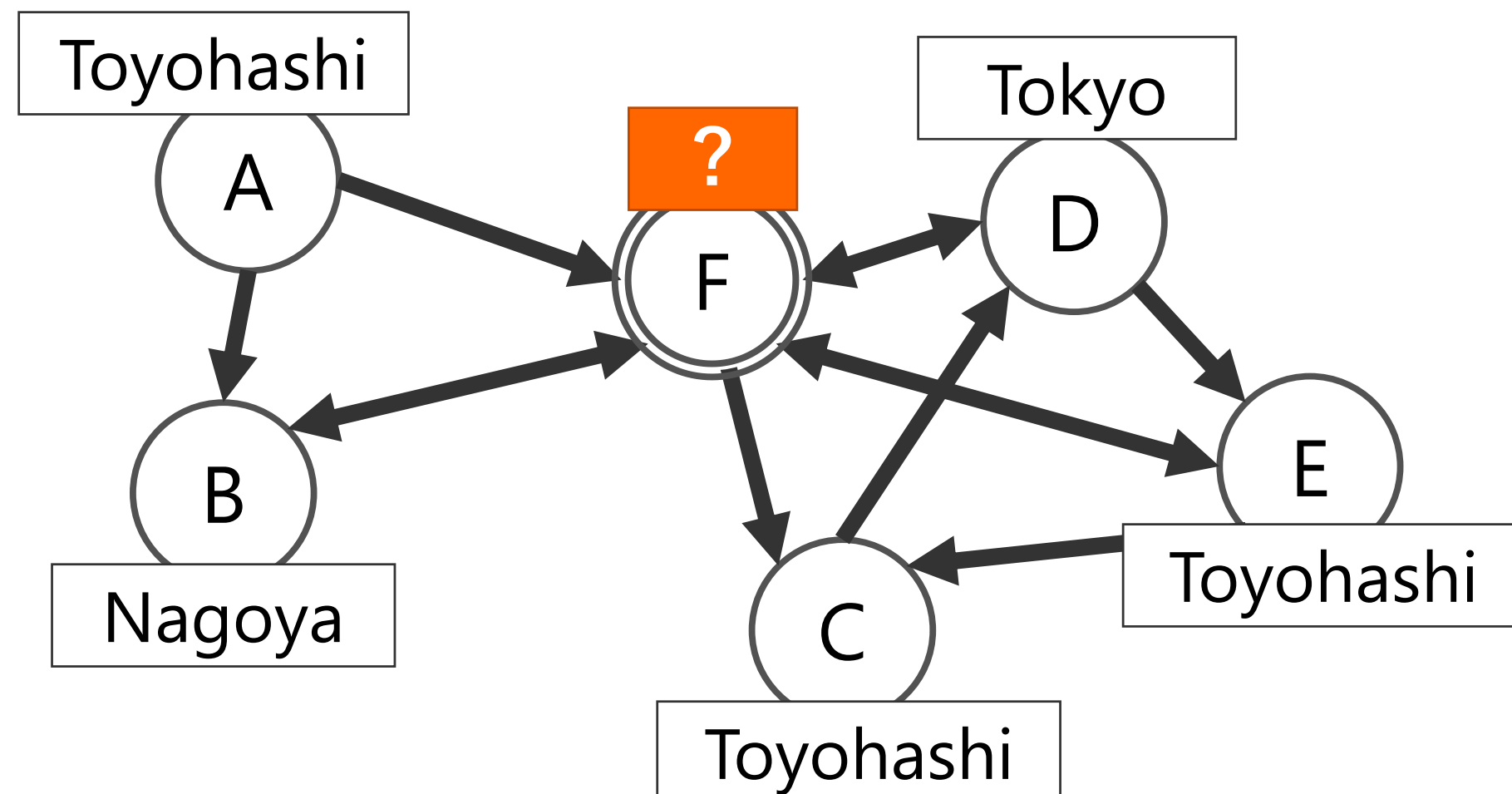
Mitsuo Yoshida
yoshida@cs.tut.ac.jp

Kyoji Umemura
umemura@tut.jp



Background

User's home location can be estimated by their relationships

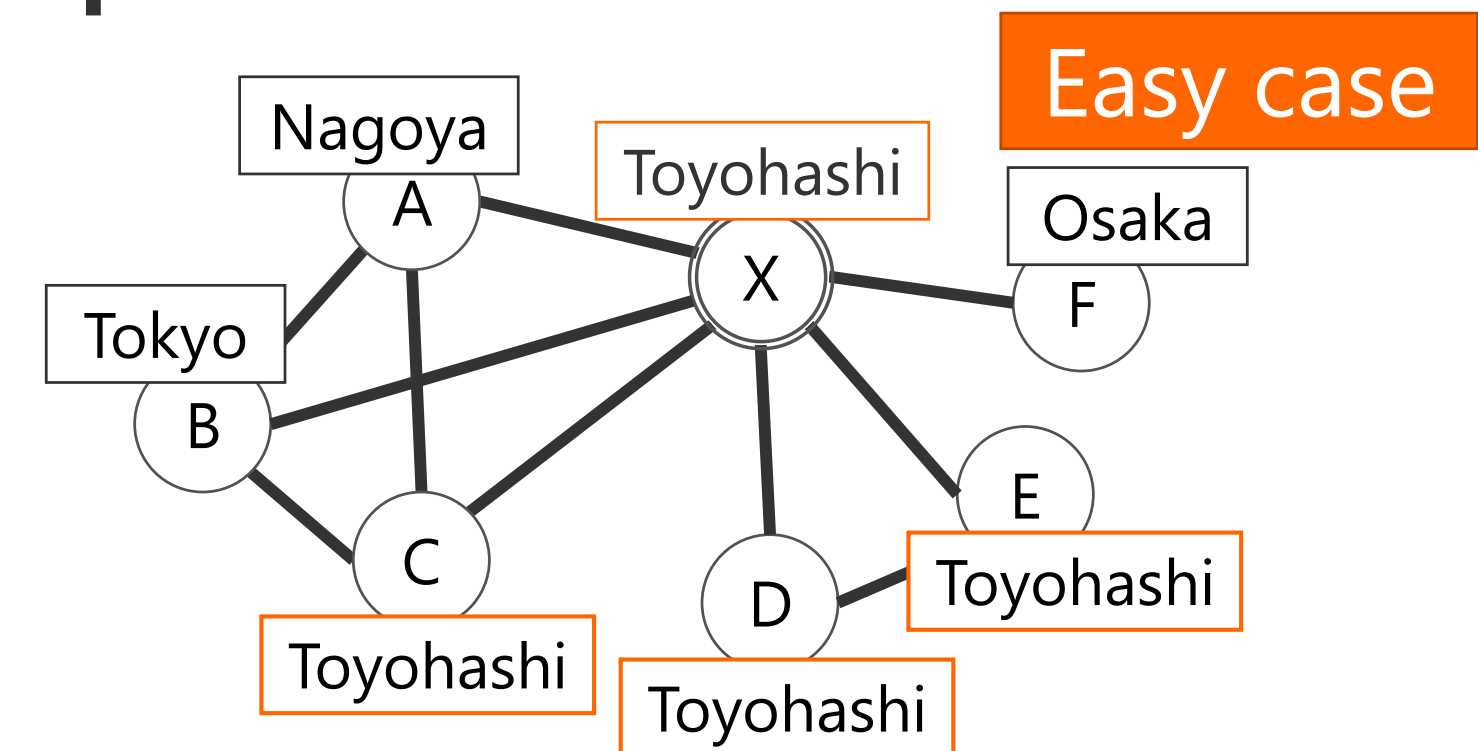


Assumption: Connected users on the graph are located geographically close

Example social graphs

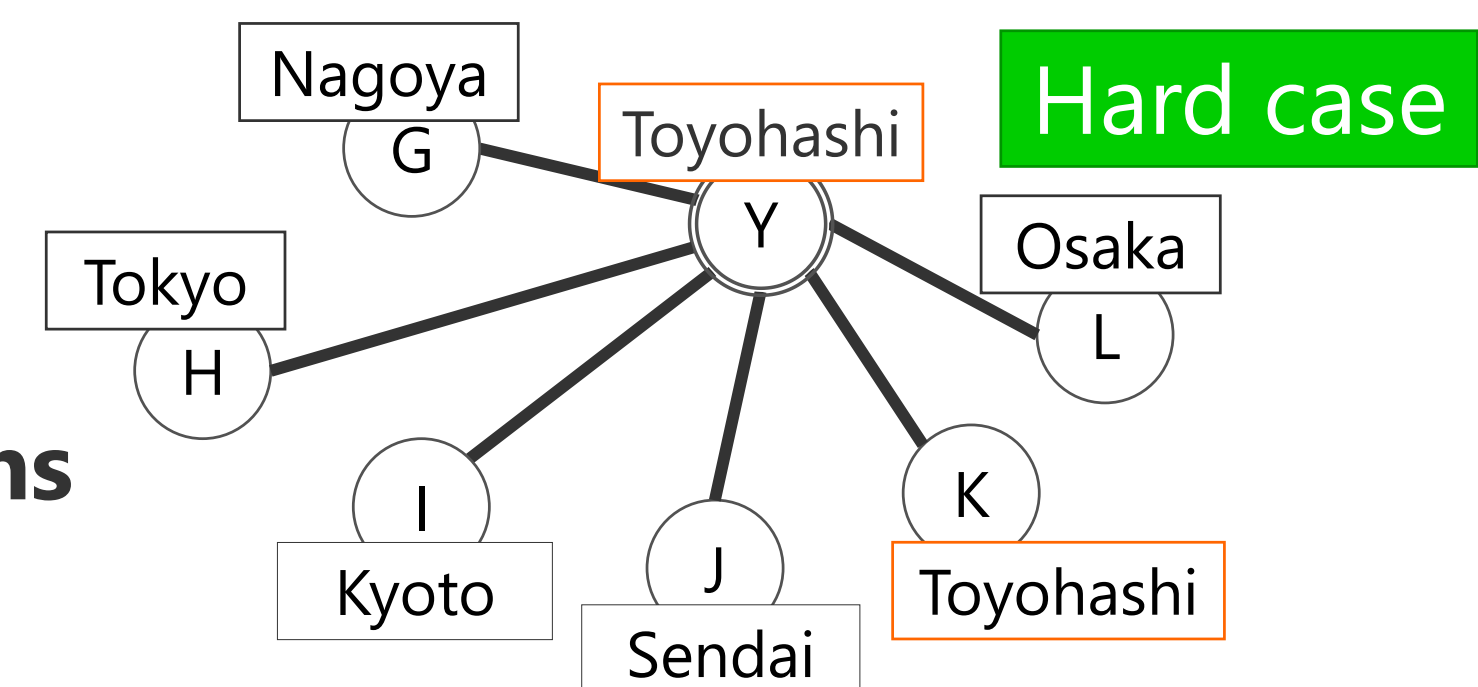
Common user:

User X has the **same location** as many friends



Celebrity:

The friends of user Y have **different locations**



Problem

Some users do not satisfy the assumption.
e.g., celebrities, influencers, ...

Purpose

We analyze the user characteristics by centralities whose home locations are hard to estimate

Method

Location data: assigned using geo-tagged tweets
Follow data: Collected following data on Twitter

Categorize users into three groups:

Easy

Same location with the majority of friends

Hard

Different location with the majority of friends

Unknown

Friends' locations are not available

Calculate five centrality scores: in-/out-degree centrality, PageRank, HITS Authority and HITS Hub

RQ. What is the centrality score that finds users who satisfy the assumption?

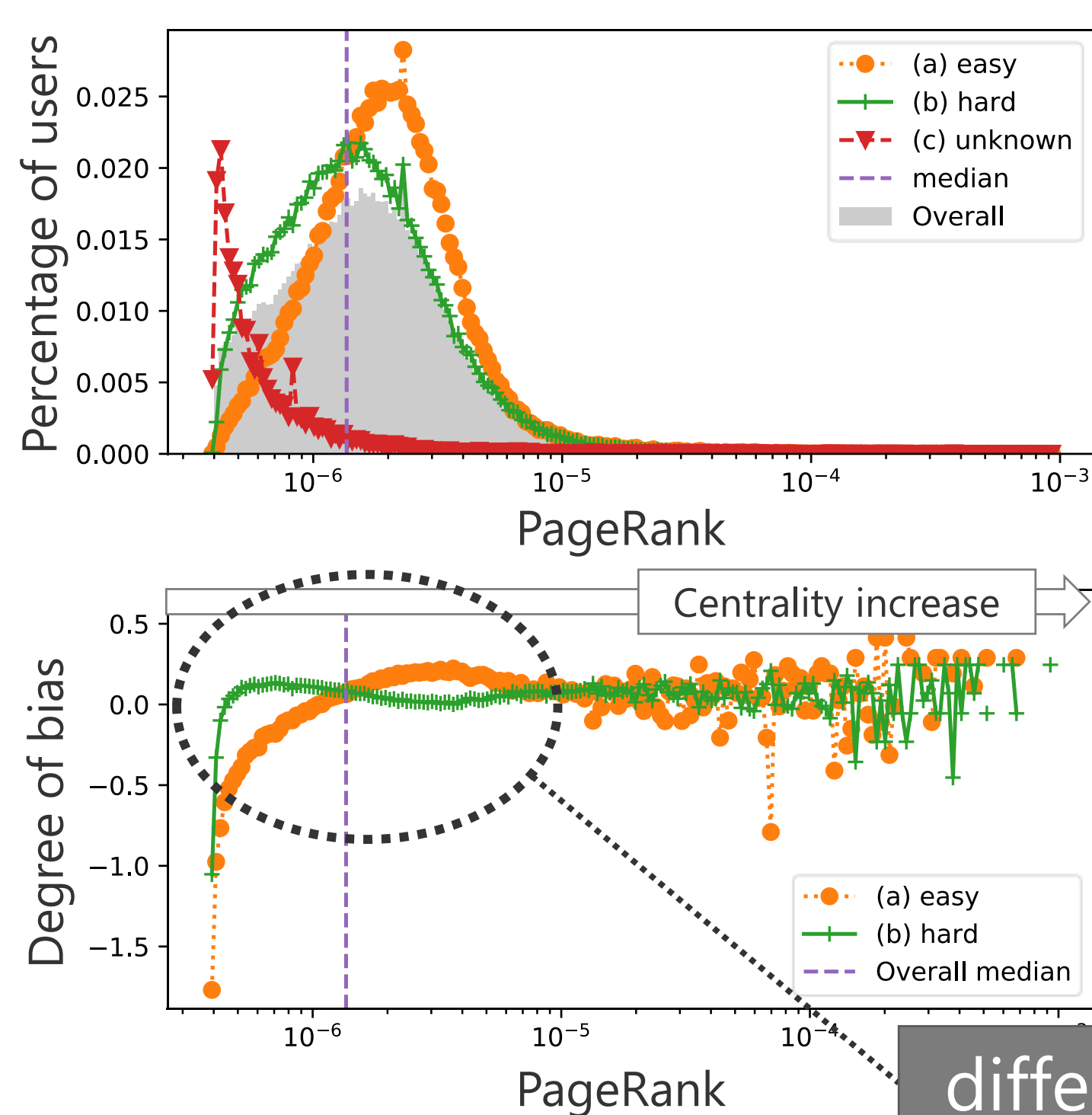
Calculate score distribution of each group : $f(x; \text{group})$

$$f(x; U)$$

$$\text{Degree_of_bias}(x) = \log \left(\frac{\% \text{users}}{\% \text{total users}} \right) f(x; V)$$

Result

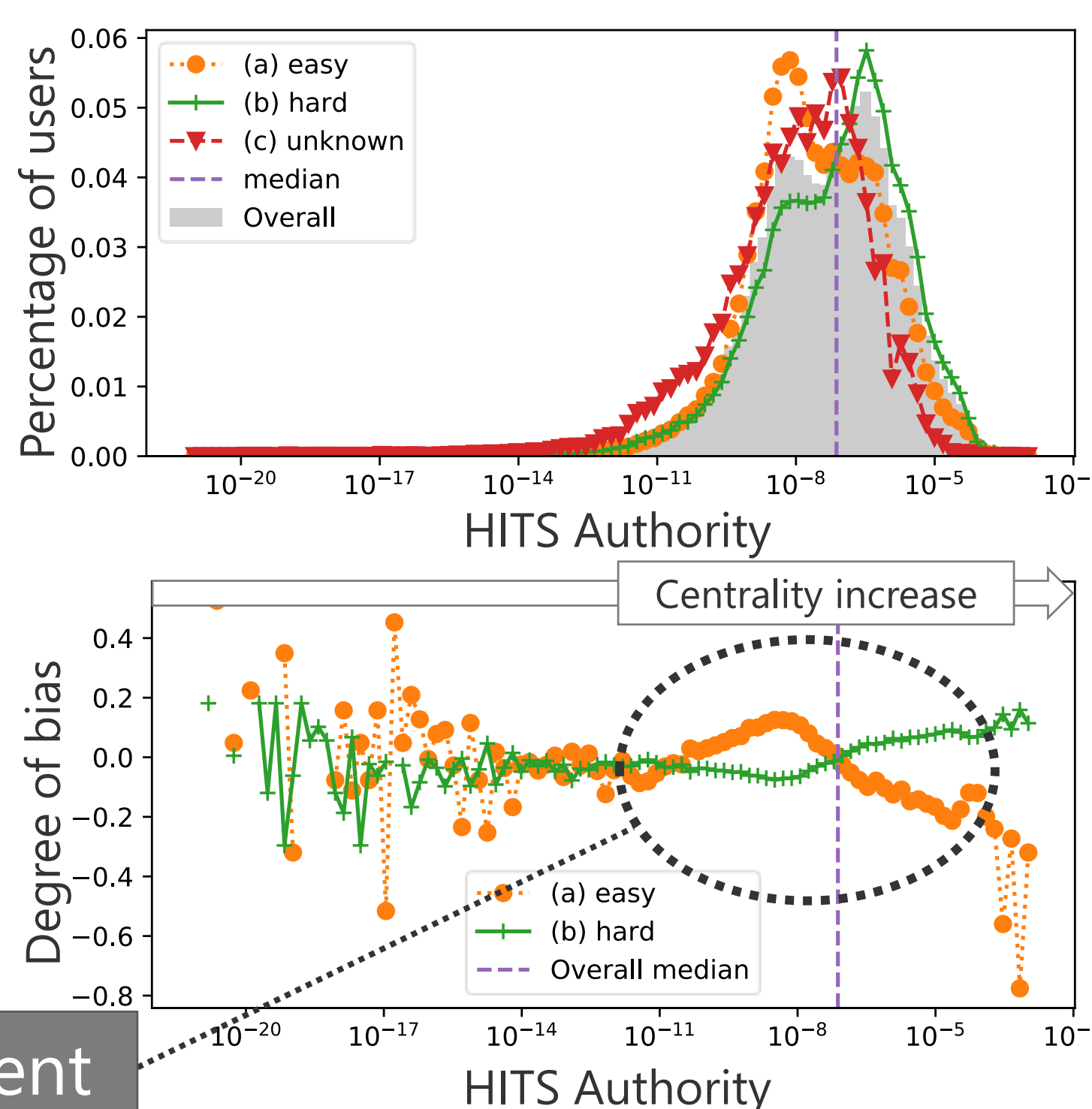
PageRank



Hard users tend to have lower PageRank scores than easy users.

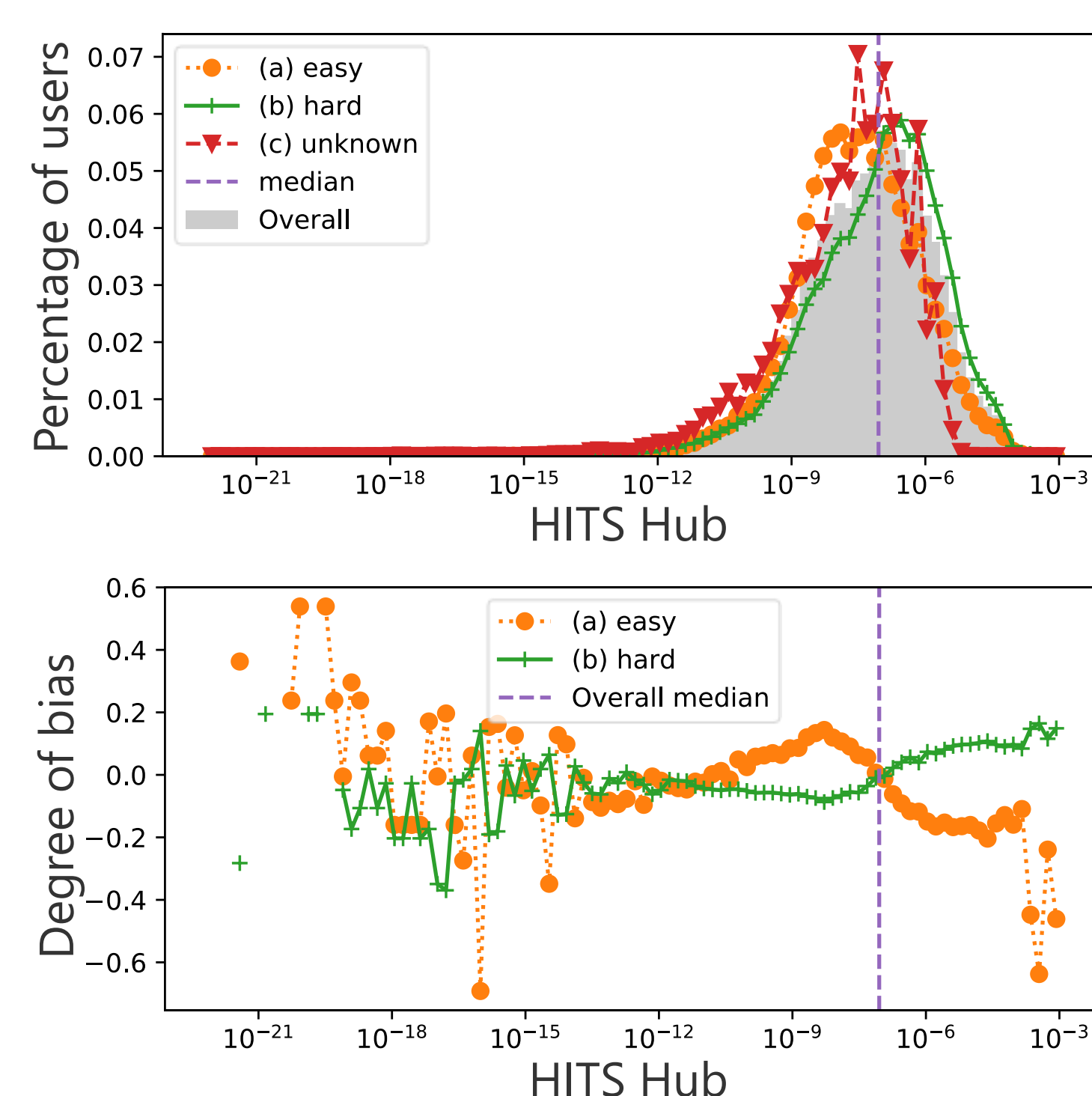
But the ratio of hard users do not increase in high PageRank, therefore PageRank cannot measure the hardness of estimation.

HITS Authority



Hard users tend to have higher authority scores.

HITS Hub



Hard users tend to have higher hub scores.

There are two types of users having hardness: high authority users and high hub users.