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

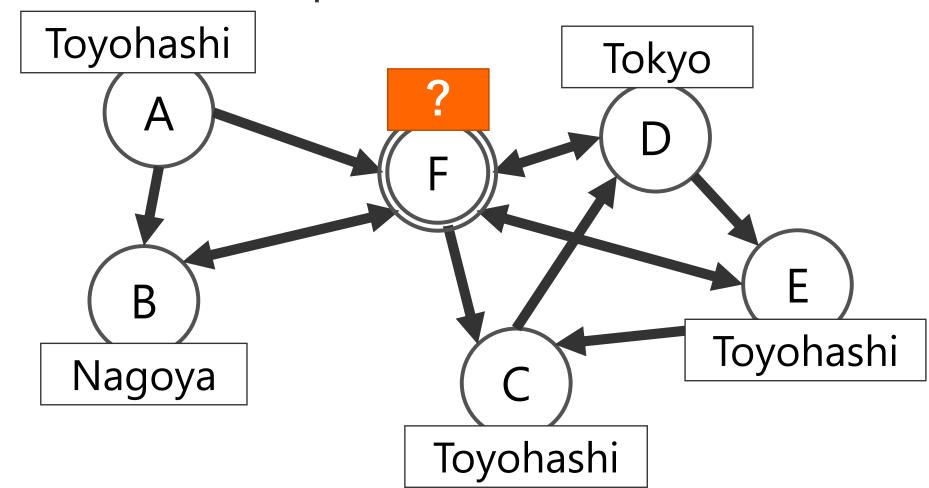
Kyoji Umemura umemura@tut.jp



Easy case

Background

User's home location can be estimated by their relationships



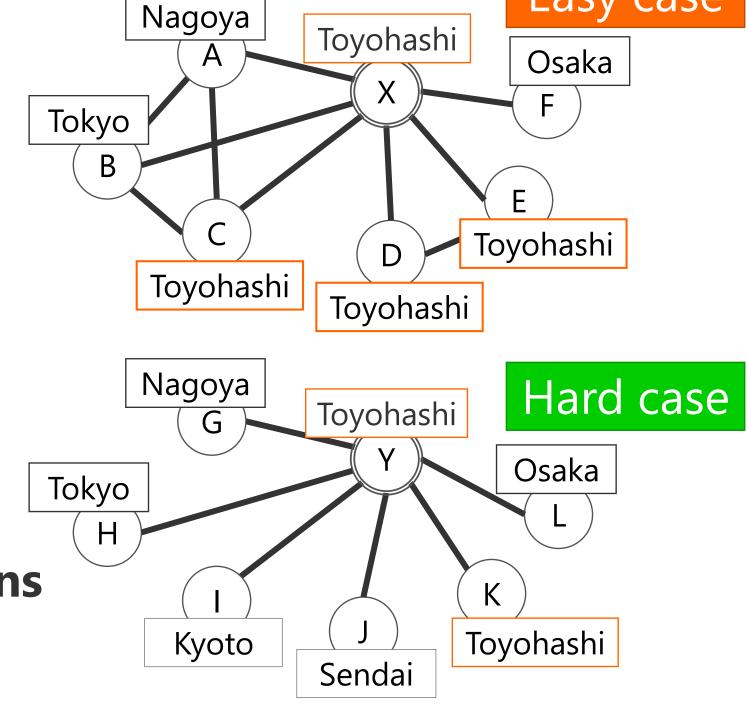
Assumption: Connected users on the graph are located geographically close

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, ...

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

Purpose

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

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

Calculate score distribution of each group : f(x; group)f(x; U)

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

HITS Hub

Result

PageRank HITS Authority O.025 · 0.020 ••• (a) easy ── (b) hard -▼- (c) unknown --- median Percentage 0.000 Overall Overall ercentage 0.02 0.01 10⁻³ 10⁻⁵ 10-20 10^{-11} 10^{-8} PageRank **HITS Authority** Centrality increase Centrality increase 0.4 of bias of bias 0.2 0.0 Degree (Degree -0.2-0.4(a) easy (b) hard -1.5b) hard -0.6Overall median Overall median 10^{-6} 10^{-5} 10^{-11} 10-5 different PageRank

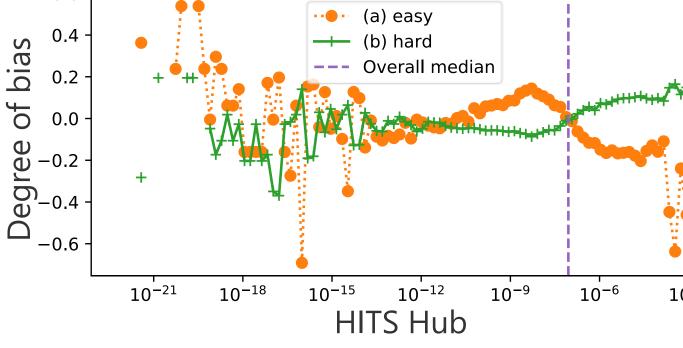
Hard users tend to have lower PageRank scores than easy users.

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



Hard users tend to have higher authority scores.

- (c) unknown Overall 10-18 10^{-12} HITS Hub (a) easy (b) hard bias Overall median 0.2 of 0.0



Hard users tend to have higher hub scores.

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

[ABCSS 19] S. Hironaka, M. Yoshida, K. Umemura. User's Centrality Analysis for Home Location Estimation. Companion Proceedings of the 2019 IEEE/WIC/ACM International Conference on Web Intelligence, pp. 59–63, 2019.