

Social Network Model based on Keyword Categorization



Prantik Bhattacharyya, Ankush Garg, S. Felix Wu
Computer Science Department
University of California, Davis

Introduction



- ▶ Bob
 - ▶ a newcomer in city
 - ▶ soccer enthusiast

Introduction



- ▶ Bob
 - ▶ a newcomer in city
 - ▶ soccer enthusiast
- ▶ Individuals become friends
 - ▶ common interests or passions
- ▶ Online Social Networks
 - ▶ profile characteristics

Previous Approaches

- ▶ Lattice Model [Kleinberg et. al. 2000]
 - ▶ geodesic distance between individuals
- ▶ Hierarchical Model [Kleinberg et. al. 2001]
 - ▶ least common ancestor of individuals
- ▶ Interest Model [Sandberg et. al. 2007]
 - ▶ interest between individuals

Contributions

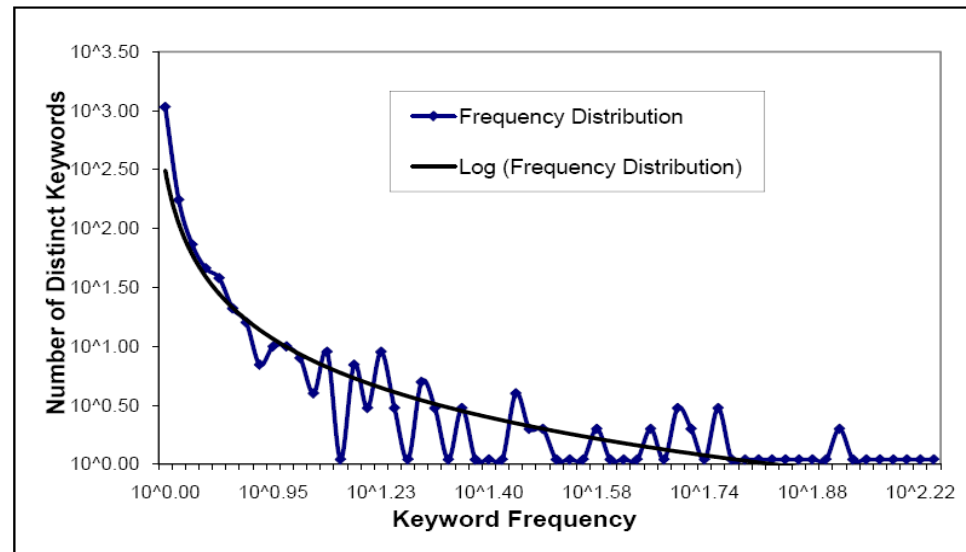
- ▶ Proposed
 - ▶ a 'forest' model to categorize keywords
 - ▶ social network model based on keyword categorization

- ▶ Validated
 - ▶ Relationship between semantic similarity of user keywords and social network topology
 - ▶ With Facebook data

Our Approach

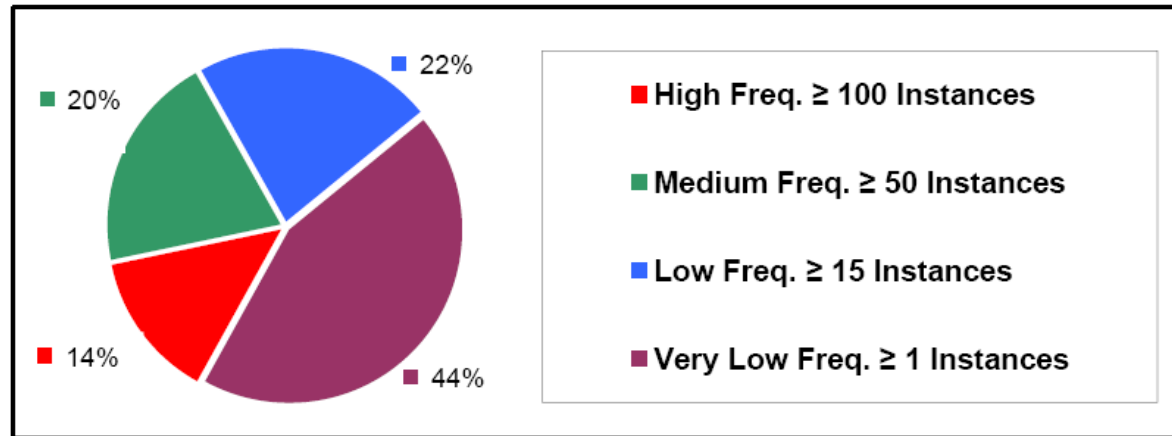
- ▶ analyze Facebook user profiles
 - ▶ \approx 1265 unique Facebook profiles
 - ▶ translates to \approx 1million user pairs
 - ▶ \approx 1500 unique keywords
 - ▶ 'Interests' field of a Facebook profile
 - ▶ single word keywords only
- ▶ \approx 6000 total keywords

Number of Distinct Keywords vs Frequency



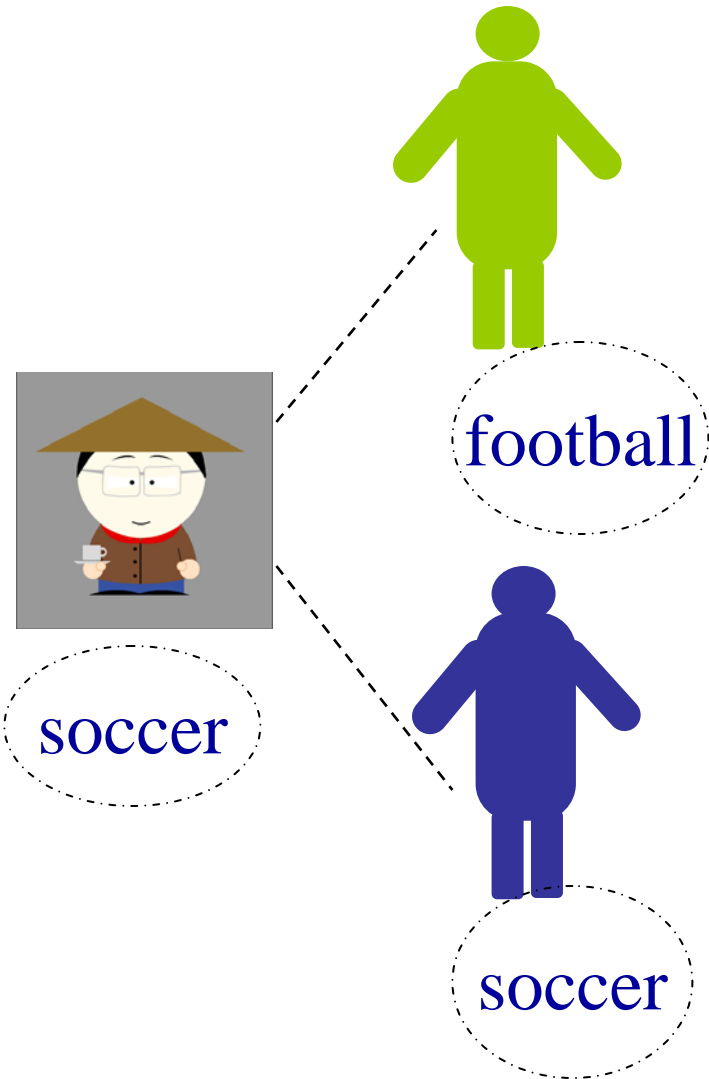
- ▶ drop rate as power-log distribution
- ▶ consistency with results on tag distribution

Keyword Frequency Distribution

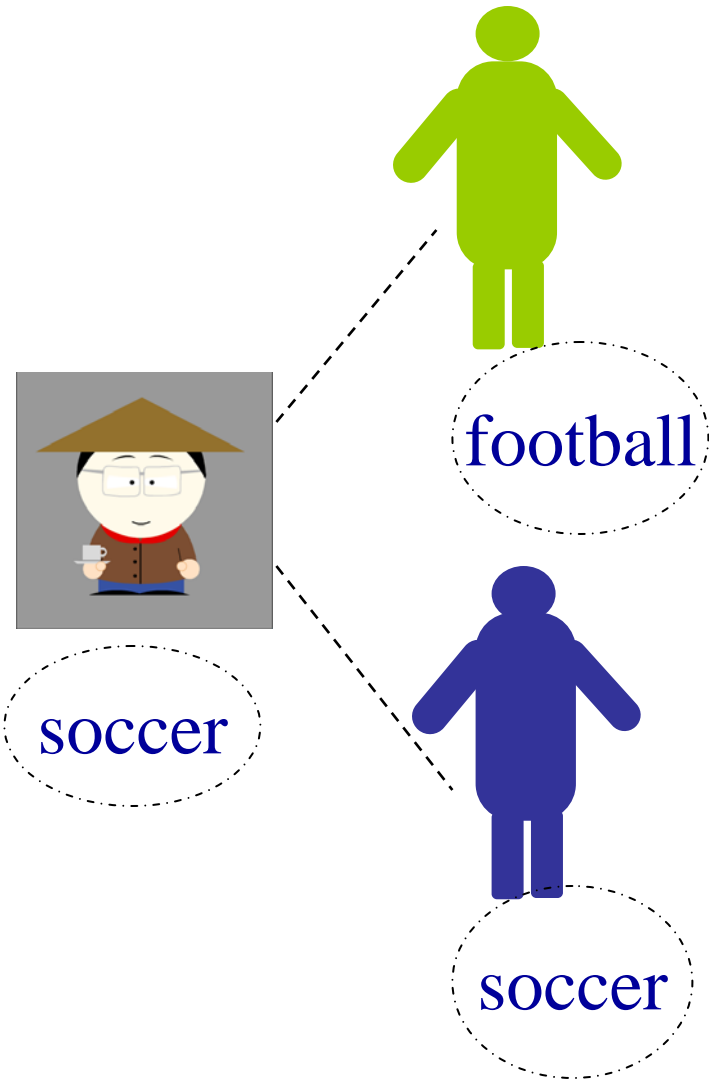


- ▶ only a fraction used repeatedly
- ▶ large percentage occurs scarcely

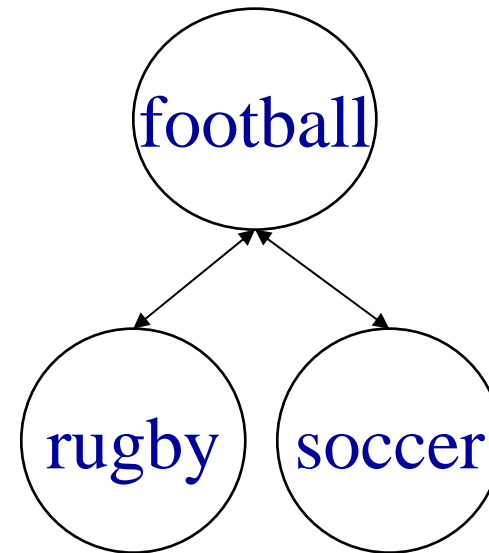
Keyword Categorization



Keyword Categorization



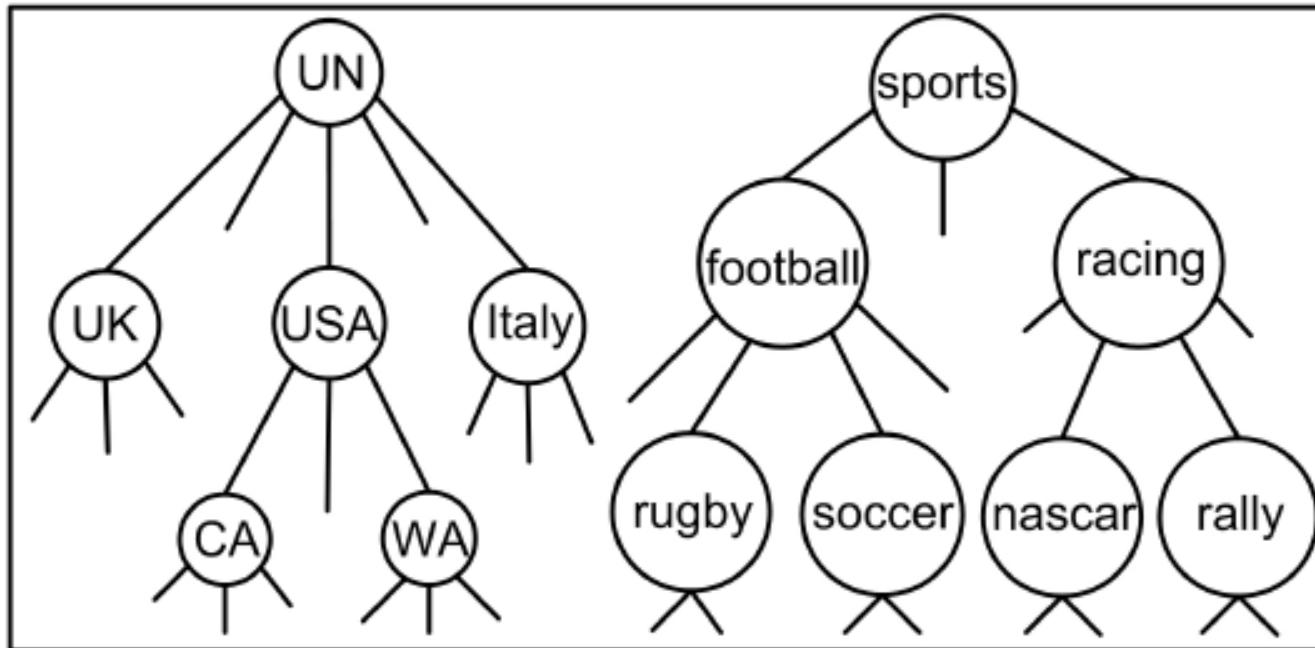
Hierarchy Model



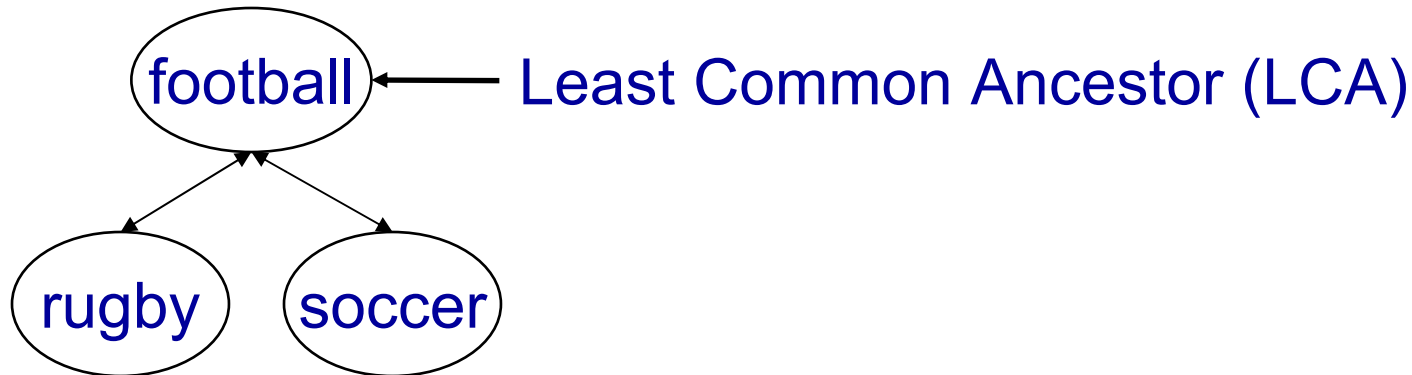
How to capture varied usage of keywords in hierarchical fashion?

Keyword Categorization: Forest Structure

Multiple Trees



Formulation: Keyword Distances



► Distance between Keywords

$$D(K_1, K_2) = \begin{cases} \text{max of distance from LCA} & \text{if } K_1, K_2 \in T_i \\ \infty & \text{if no such } T_i \text{ exists} \end{cases}$$

Formulation: User Keywords

- ▶ user u has N_u keywords
- ▶ keyword pairs between u and v : $k(u, v) = N_u \times N_v$
- ▶ keyword pairs in same tree: $n(u, v)$

Similarity functions

- ▶ Weak Similarity:

- ▶ Strong Similarity:

Similarity functions

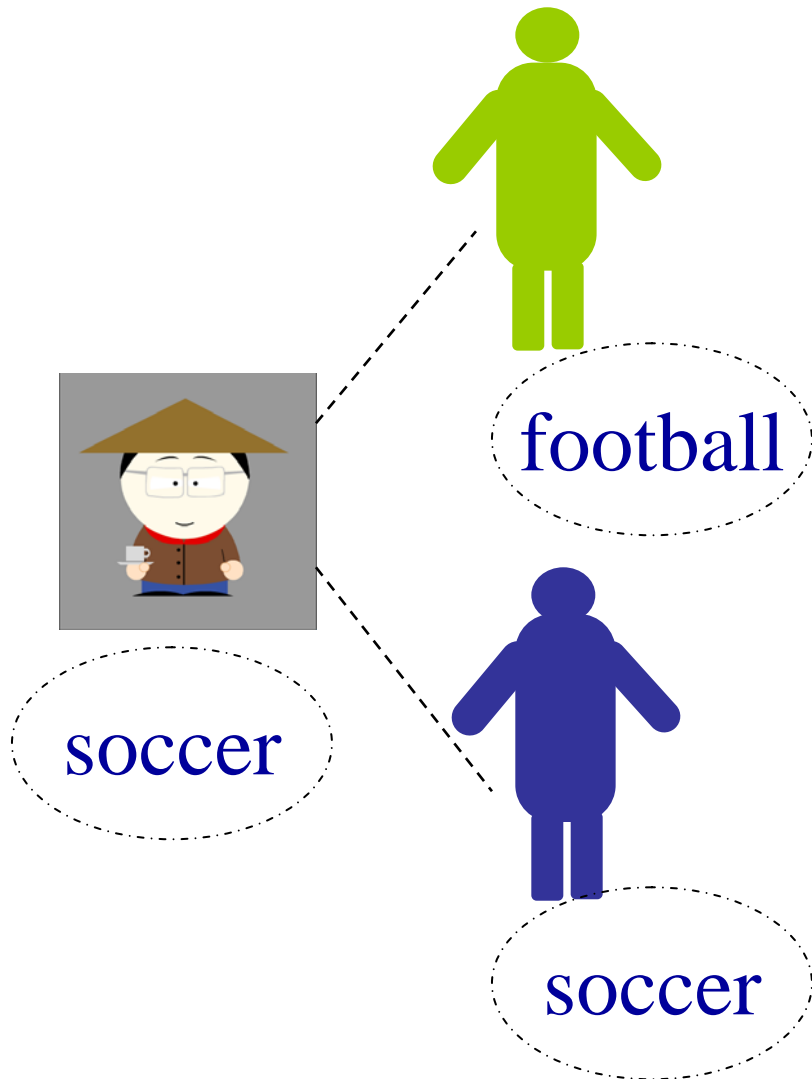
- ▶ Weak Similarity:

$$s(u, v) = \frac{n(u, v)}{k(u, v)}$$

- ▶ Strong Similarity:

$$S(u, v) = \frac{\sum_{1 \leq i \leq Nu, 1 \leq j \leq Nv} e^{-D(K^u_i, K^v_j)}}{k(u, v)}$$

Similarity functions



- ▶ Weak Similarity = 1
- ▶ Strong Similarity = e^{-1}

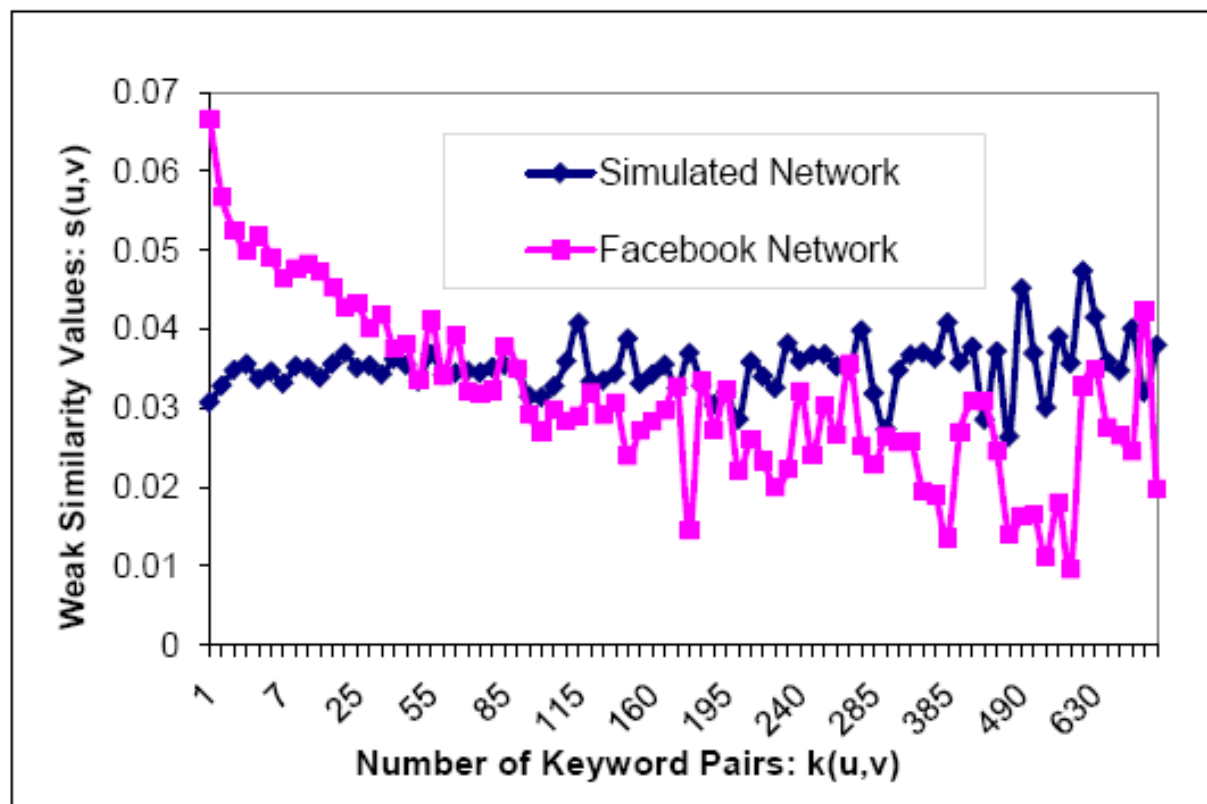
- ▶ Weak Similarity = 1
- ▶ Strong Similarity = e^{-0}

Social Network Graph

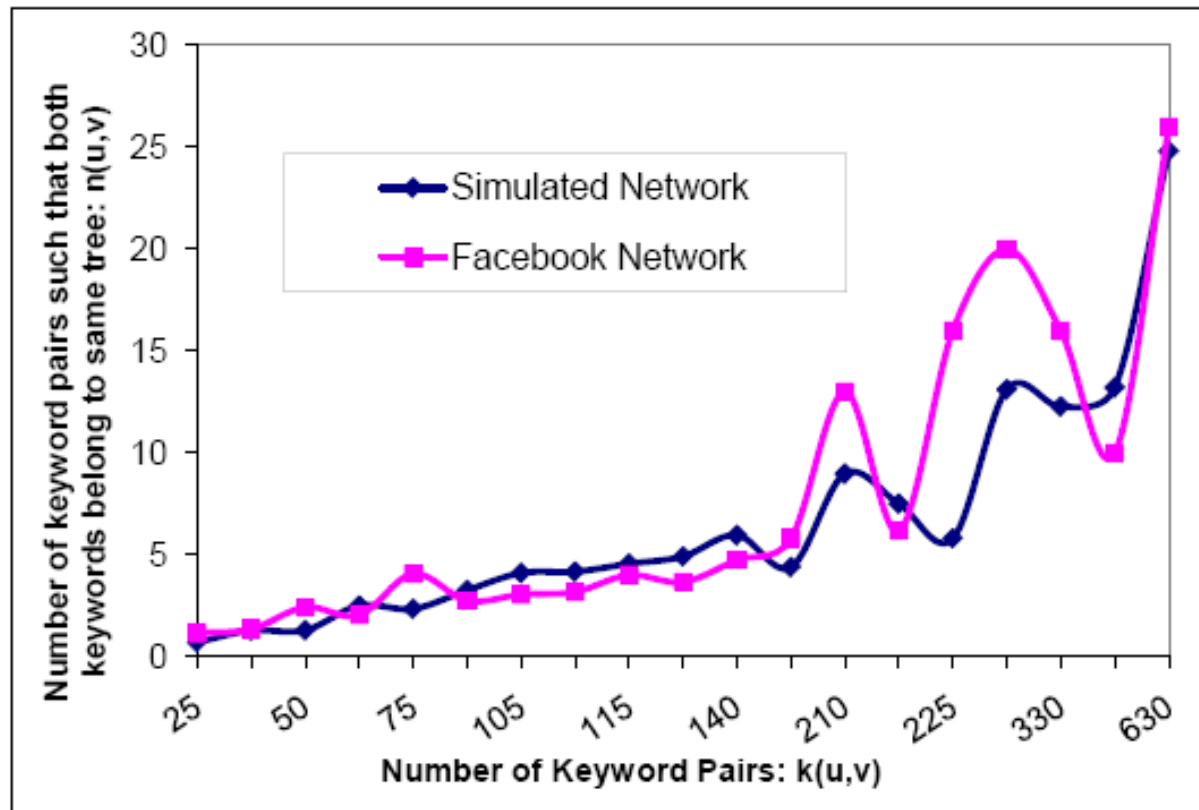
$$p(u, v) = \frac{S(u, v)}{\sum_v S(u, v)}$$

- ▶ $p(u, v) \neq p(v, u)$
- ▶ *Social Network Graph* \equiv *Users, Friendships*

Weak similarity: All User Pairs



Weak Similarity: Direct Friends



Concluding Remarks

- ▶ analyzed profile characteristics
 - ▶ model usage scope of keywords
- ▶ correlated with social network topology
- ▶ analysis on larger and varied dataset
 - ▶ Orkut, Youtube, Twitter..

Thank you!

Questions?

{pbhattacharyya,garg,sfwu}@ucdavis.edu