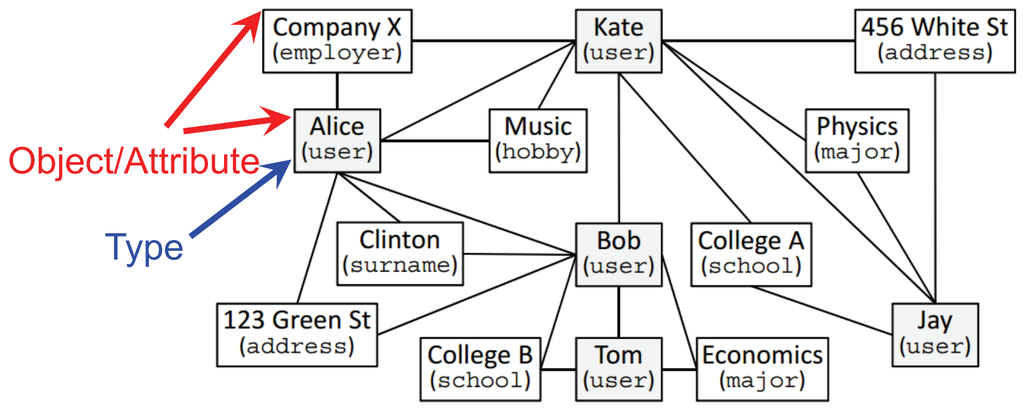
Semantic Proximity Search on Graphs with Metagraph-based Learning

Yuan Fang[1] Wenqing Lin[1] Vincent Zheng[2] Min Wu[1] Kevin Chang[23] Xiao-Li Li[1]

[1] Institute for Infocomm Research, Singapore
[2] Advanced Digital Sciences Center, Singapore
[3] University of Illinois at Urbana-Champaign, USA

Problem: *Semantic Proximity Search on Heterogeneous Graph*



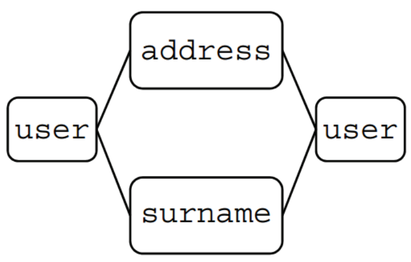
On a **"typed" object graph** that captures users and their attributes on a social network:

Which users are ~~close to / related~~ to Bob?

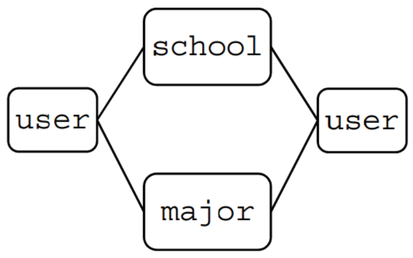
Family? (Alice)
Classmates? (Tom)

Insights: *Metagraphs to "Explain" Different Semantic Classes*

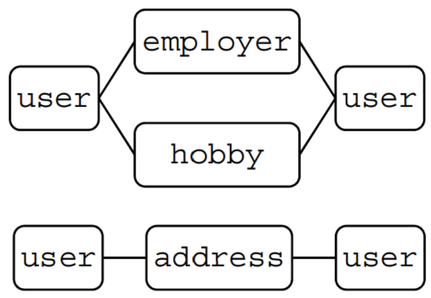
Family
[Bob & Alice]



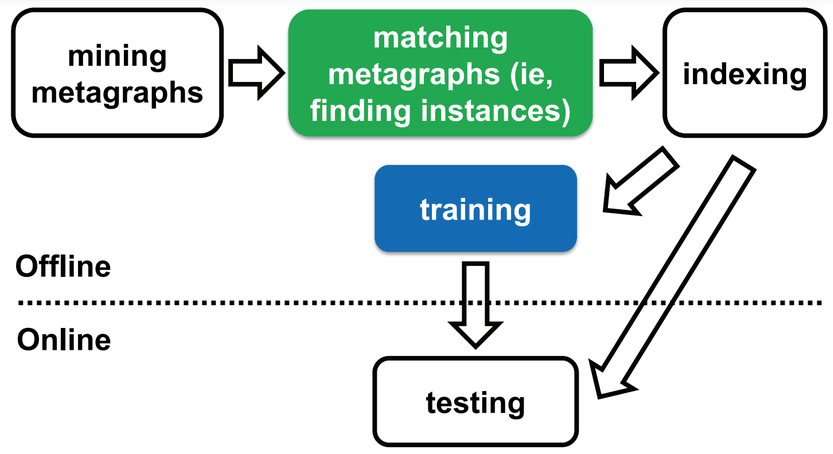
Classmates
[Kate & Jay, Bob & Tom]



Close friends
[Kate & Alice] [Kate & Jay]



Overall Framework



Training

Definition of Proximity

Proximity of two nodes x, y on graph

$$\pi(x, y; \mathbf{w}) \triangleq \frac{2\ \mathbf{m}_{xy} \cdot \mathbf{w}}{\mathbf{m}_x \cdot \mathbf{w} + \mathbf{m}_y \cdot \mathbf{w}}$$

$\mathbf{m}_{xy}[\mathrm{i}]$: # times x, y co-occur in instances of metagraph i

$\mathbf{m}_x[\mathrm{i}]$: # times x occurs in instances of metagraph i

$\mathbf{w}[i]$: weight for metagraph i

Basic Learning Model

Pairwise learning to rank

$$P(q, x, y; \mathbf{w}) \triangleq \frac{1}{1 + e^{-\mu(\pi(q,x;\mathbf{w}) - \pi(q,y;\mathbf{w}))}}$$

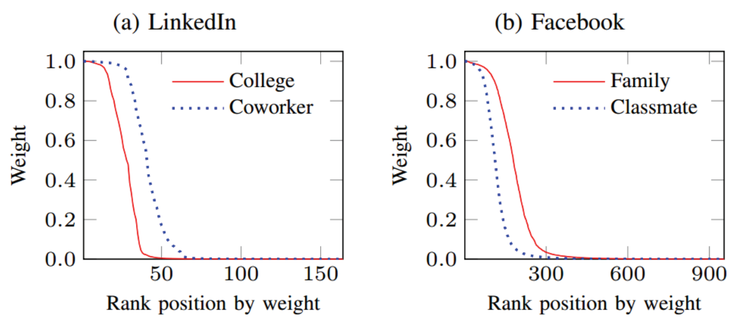
Each example is a triplet:
for query q, x is ranked before y.

Objective function

$$L(\mathbf{w}; \Omega) = \sum_{(q,x,y)\in\Omega} \log P(q, x, y; \mathbf{w})$$

Dual-Stage Training

- Expensive to process/match all metagraphs
- Yet not all metagraphs are useful



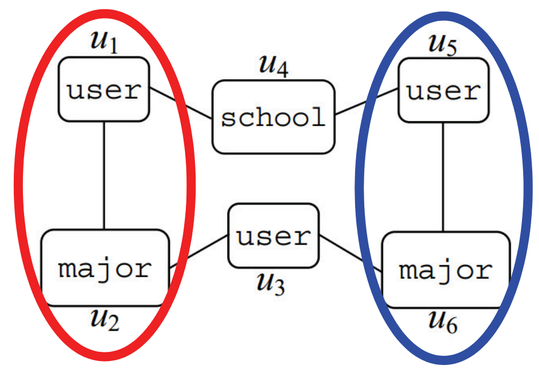
identify seed metagraphs → **learn with seed metagraphs**

based on weights of seed metagraphs and their structural relationship with other metagraphs

select more metagraphs → **re-learn with seed + selected metagraphs**

Matching Metagraphs

- Existing method
  - Backtracking DFS search
  - Node by node until an entire matched instance is found
  - Fail to leverage symmetric components
- Symmetry-based matching
  - Many metagraphs are symmetric
  - Avoid redundant computation



Main Results

Datasets:
- College & Coworkers *(labelled on LinkedIn)*
- Family & Classmate *(rule on Facebook)*

Baselines:
- **MGP: metagraph-based proximity (ours)**
- MPP: metapath-based proximity
- MGP-U: all metagraphs have uniform weights
- MGP-B: only use the best metagraph
- SRW: supervised random walk