Replacing the Irreplaceable: Fast Algorithms for Team Member Recommendation

Presenter: **Liangyue Li (ASU)**
Joint work with

Hanghang Tong
ASU

Nan Cao
IBM

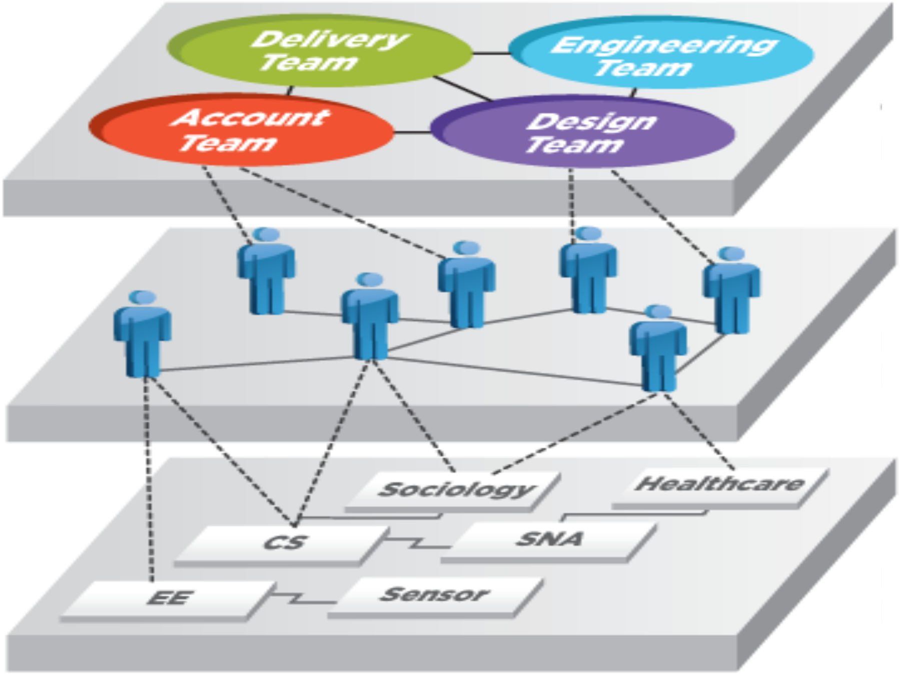
Kate Ehrlich
IBM

Yu-Ru Lin
U Pitts

Norbou Buchler
ARL

Collaboration Teams in Network

People collaborate as a team to collectively perform some complex tasks



Team Level Network

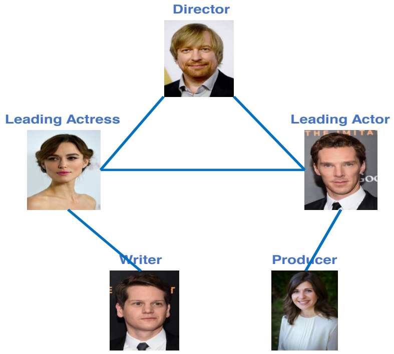
Person Level Network

Information Topic Level Network

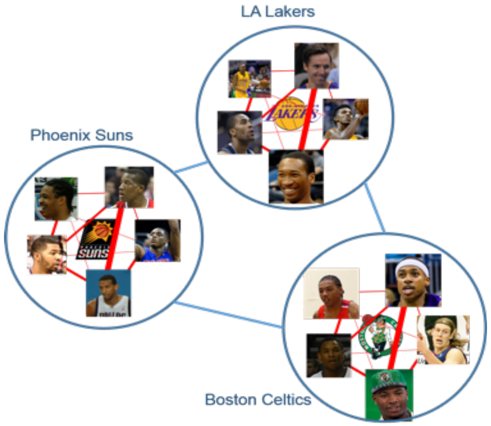
Wuchty, Stefan, Ben Jones, and Brian Uzzi. "The Increasing Dominance of Teams in the Production of Knowledge," Science, May 2007, 316:1036-1039.

Teams Are Everywhere

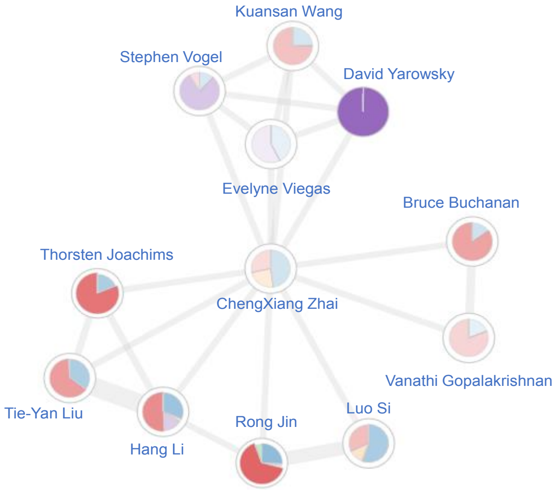
1. Film Crew



2. Sports Team



3. Research Team



4. Military Team

Research Questions

- **What do high-performing teams share in common?** [Uzzi+Science13]
- **What drives long-term scientific impact?**[Wang+Science13]
- **What's the optimal design for a team in the context of network?**[Lappas+KDD09, Anagnostopoulos+WWW10]

Churn of A Team Member

Case 1: Employee resigns in a sales team

Case 2: Task force down in a SWAT team

Case 3: Rotation tactic between benches in NBA team

Q: How to find the best alternative when a team member leaves? [This paper!]

Team Member Replacement

Problem Definition:

Adj. Matrix

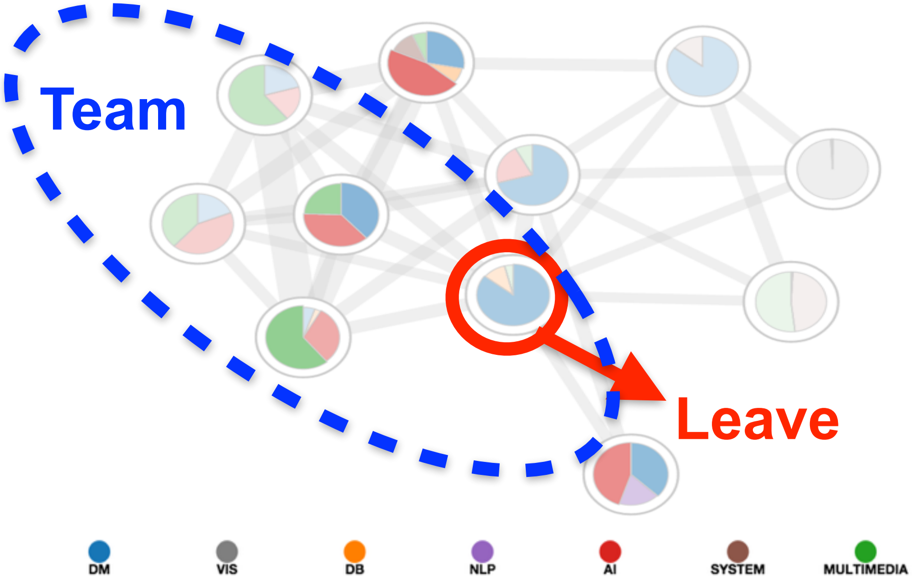
Given: (1) A labelled social network $G := \{A, L\}$

(2) A team $G(\mathcal{T})$

(3) A team member $p \in \mathcal{T}$

Skill Indicator

Recommend: A "best" alternative $q \notin \mathcal{T}$ to replace the person p's role in the team $G(\mathcal{T})$



Q: who is a good candidate to replace the person to leave

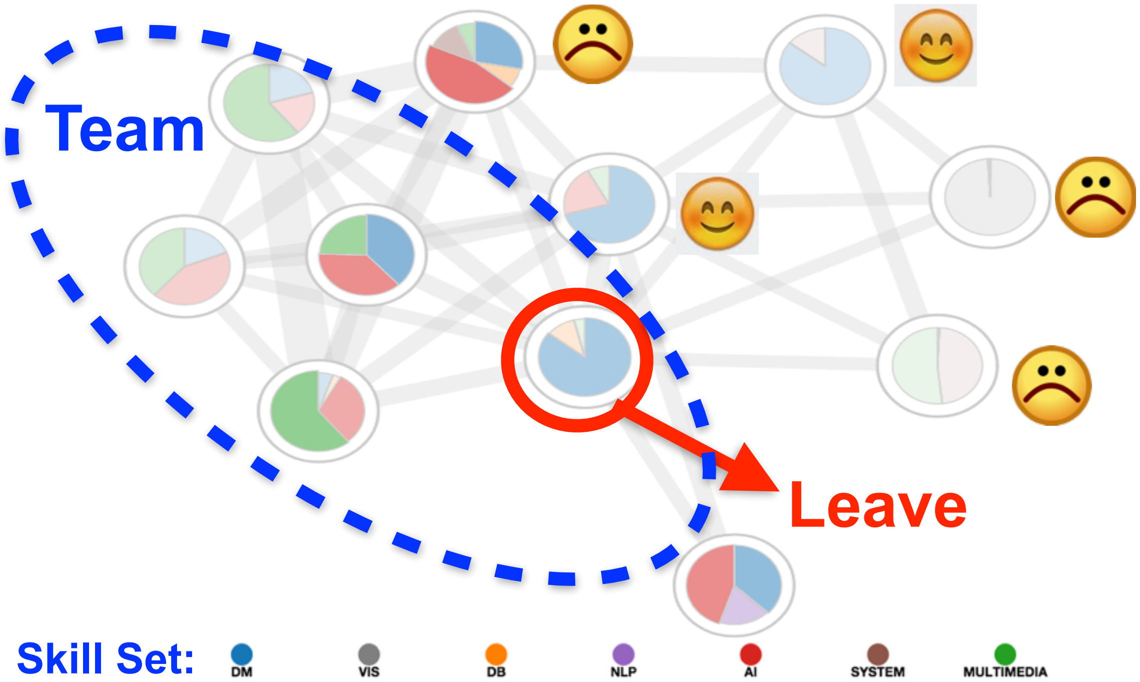
Social Studies

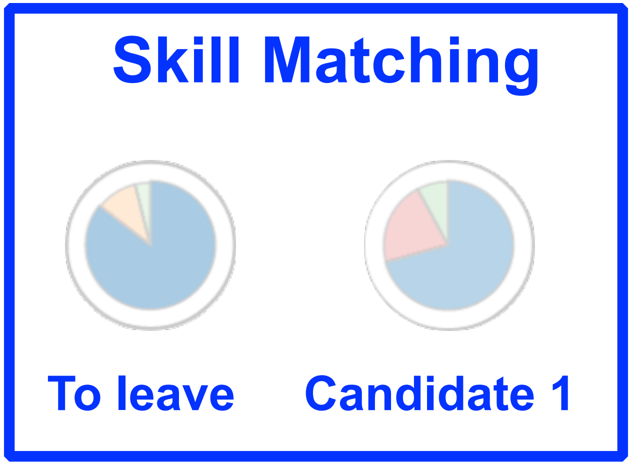
- Team members prefer to work with people they have worked before [Hinds+OBHDP00]
- Distributed teams perform better when members know each other [Cummings+CSCW08]
- Specific communication patterns amongst team members are critical for performance [Cataldo +CHI12]

Conjecture: The similarity should be measured in the **context of the team itself**

Design Objectives

Objective 1: A good candidate should have a similar skill set

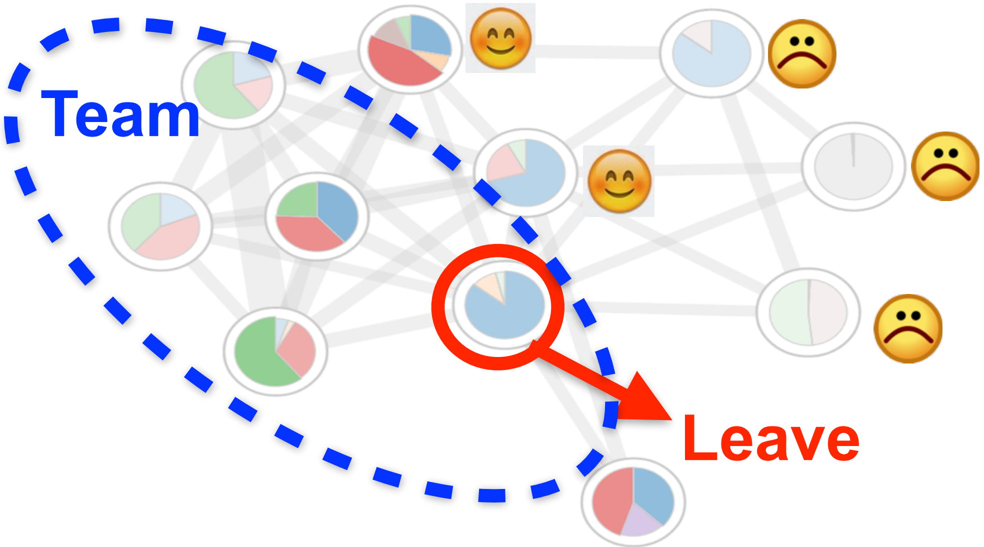


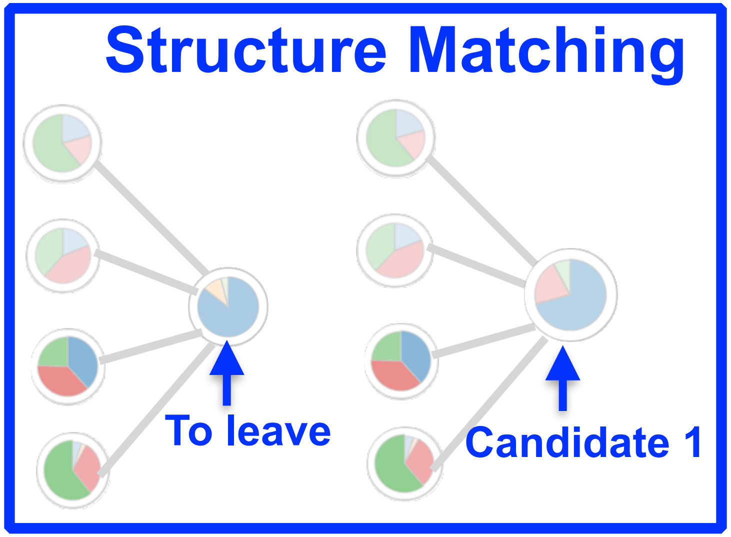


New team will have similar skill set as the old team to complete the task

Design Objectives

Objective 2: A good candidate should have a similar network structure



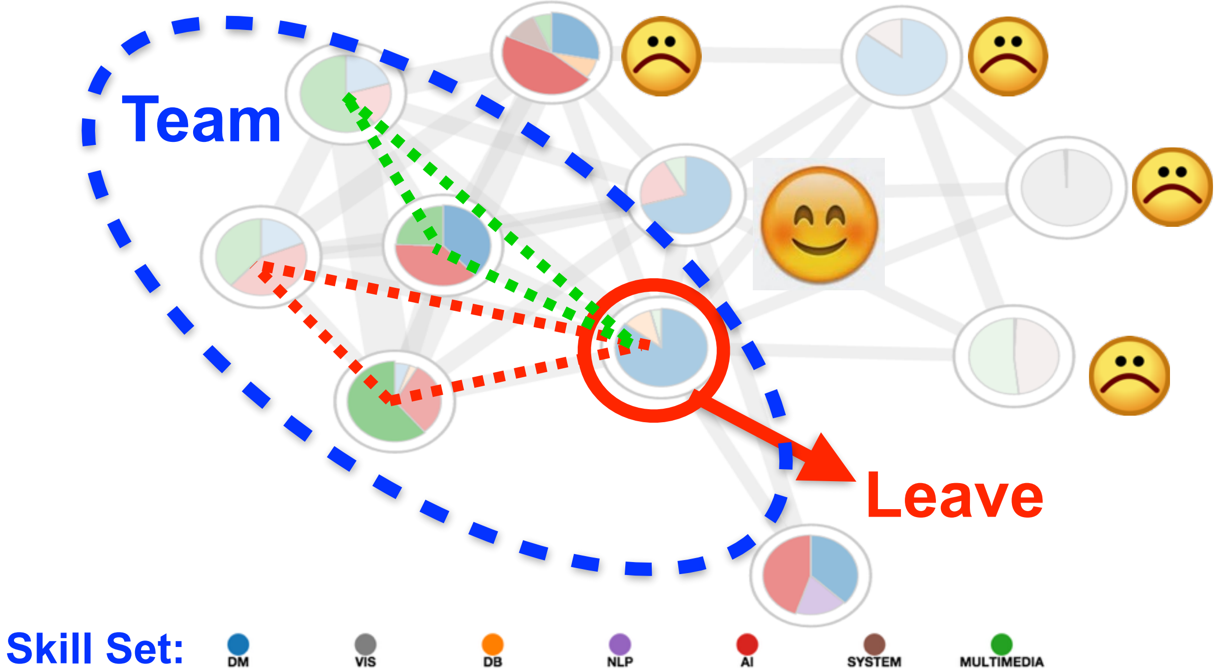


Skill Set: ● DM ● VIS ● DB ● NLP ● AI ● SYSTEM ● MULTIMEDIA

New team will have similar network structure as the old team to collaborate effectively

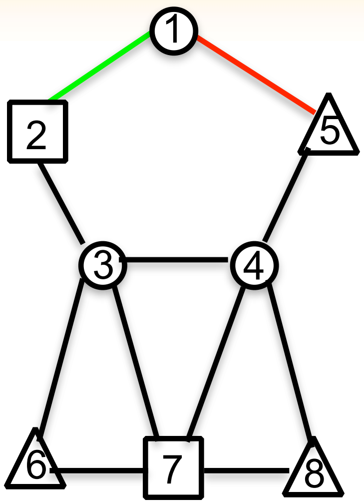
Design Objectives

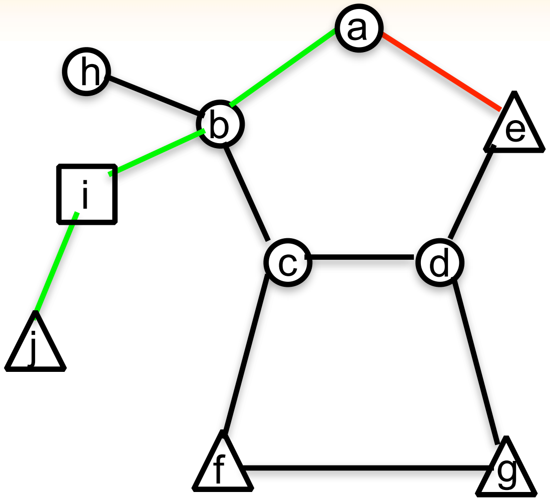
The two objectives should be fulfilled simultaneously!



New team will have similar skill and communication configuration for each sub-task

Random Walk based Graph Kernel





Graph 1

Graph 2

Details:

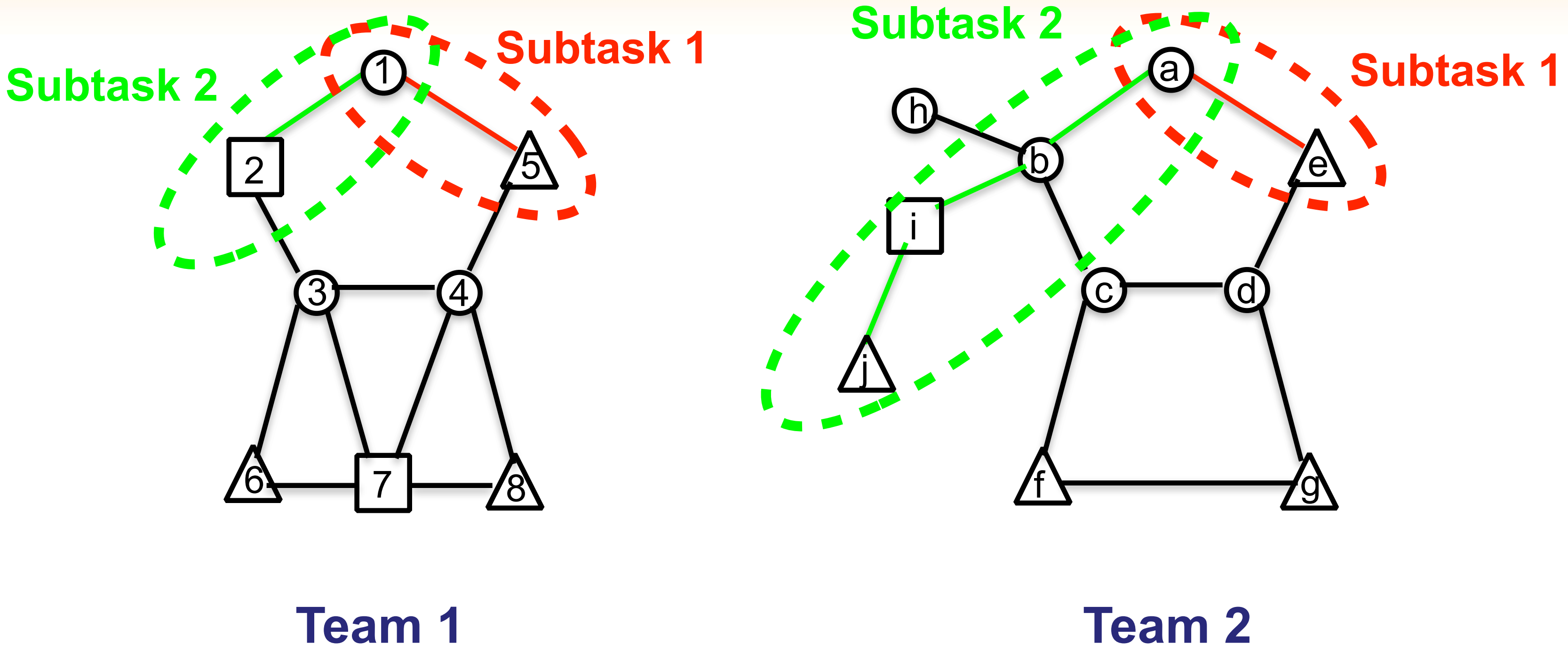
1. Compare similarity of every pair of nodes from each graph

— Eg: (1,2) vs (a, j) → less similar

(1,5) vs (a,e) → more similar

2. Node pair similarity is measured by random walks
3. Two graphs are similar if they share many similar node pairs

Random Walk based Graph Kernel

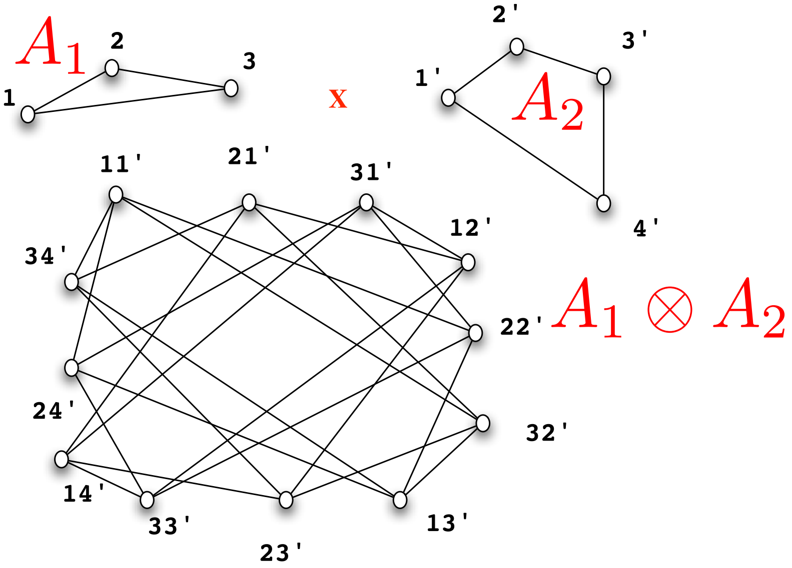


Remarks:

- Incorporates both **attributes** and **structures** similarity
- Ideal fit for our two design objectives **simultaneously**

Kronecker Product Graph w/o Attribute

Graph Illustration



Matrix Description

$$A_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \qquad A_2 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

$$A_1 \otimes A_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Kronecker product

One Random Walk on A_1

$+$

One Random Walk on A_2

$=$ **One Random Walk on** $A_1 \otimes A_2 = A_\times$

S. V. N. Vishwanathan, Nicol N. Schraudolph, Imre Risi Kondor, and Karsten M. Borgwardt. Graph Kernels. Journal of Machine Learning Research, 11:1201–1242, April 2010.

RWR Graph Kernel — Formulation

Taking expectations instead of summing

$$\begin{aligned}\mathrm{Ker}(G_1, G_2) &= \textstyle\sum_k c^k q'_\times (L_\times A_\times)^k L_\times p_\times \\ &= q'_\times (I - cL_\times A_\times)^{-1} L_\times p_\times\end{aligned}$$

Attribute Indicator

Computational challenge:

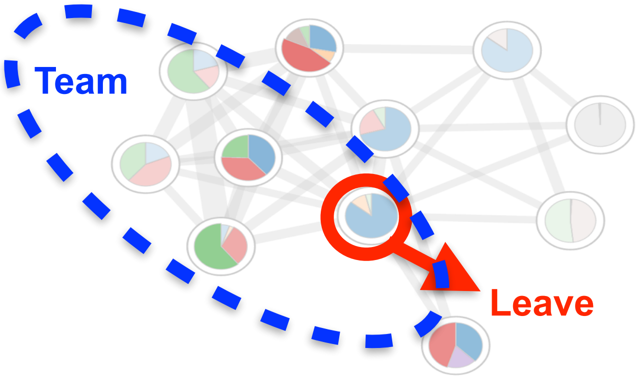
- $A_\times$ is of size $n^2 \times n^2$
- **Computational cost**
  - *Exact methods:* $O(n^6)$ (Direct computation) or $O(n^3)$ (Sylvester equation)
  - *Approx methods:* $O(n^2r^4 + mr + r^6)$ (Kang+SDM12)

- U. Kang, Hanghang Tong, Jimeng Sun. Fast Random Walk Graph Kernel. SDM 2012
- S. V. N. Vishwanathan, Nicol N. Schraudolph, Imre Risi Kondor, and Karsten M. Borgwardt. Graph Kernels. Journal of Machine Learning Research, 11:1201–1242, April 2010.

TEAMREP-BASIC

Find a new member *q* not in the current team that satisfies:

$$q = \arg\max_{j, j \notin \mathcal{T}} \operatorname{Ker}(G(\mathcal{T}), G(\mathcal{T}_{p \to j}))$$



One graph kernel computation for every possible candidate

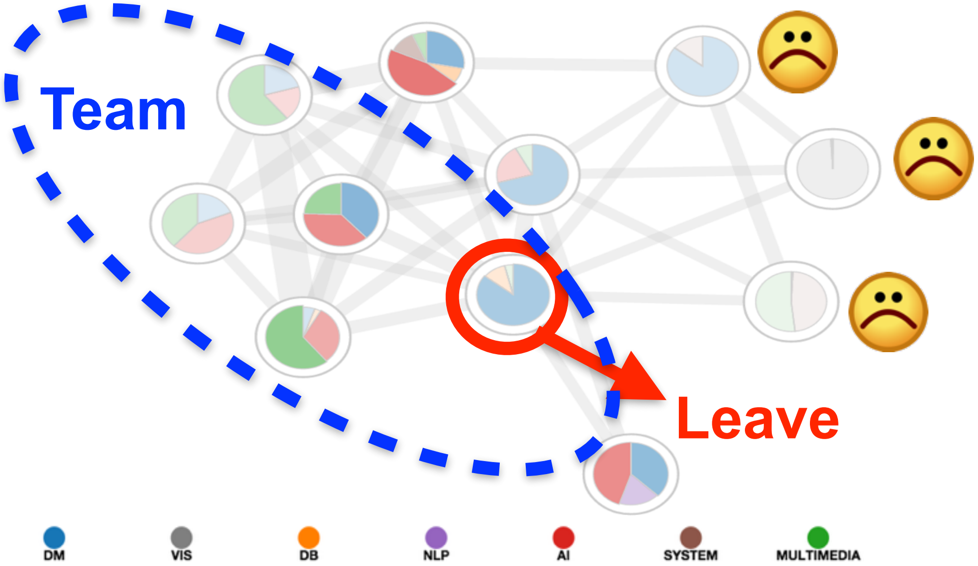
- **Challenge:** need to compute many graph kernel
  overall complexity: $O(nt^3)$
- **Questions:**
  - **Q1:** how to reduce the number of graph kernels
  - **Q2:** how to speed up the computation for each graph kernel

Roadmap

- Motivations
- Proposed Solutions
- Experimental Results
- Conclusion

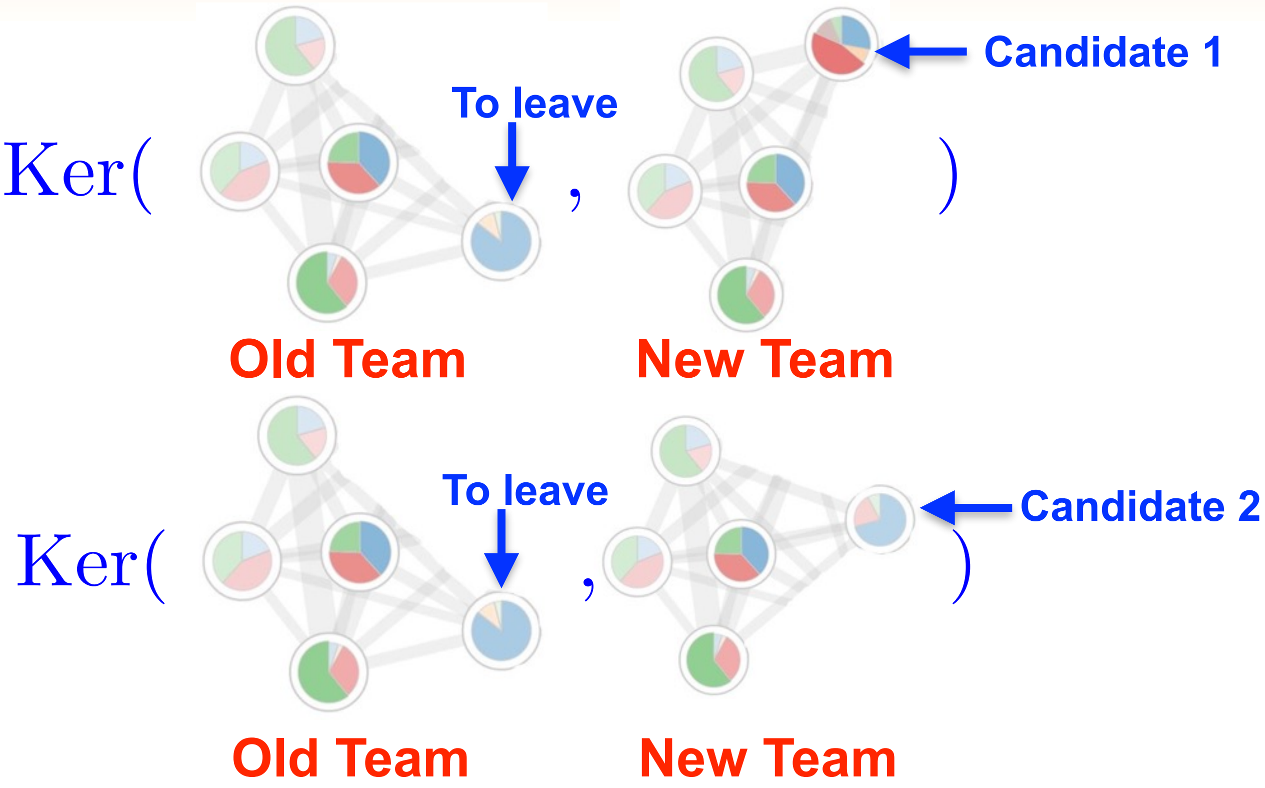
Scale-up: Candidate Filtering

***Pruning Strategy*:** Filter out all the candidates who do not have any connections to any of the rest team members.



- **Theorem:** The pruning is safe: wont' miss any potentially good replacement
- **Benefit:** The number of graph kernel computations is reduced to $O(\sum_{i \in \mathcal{T}/p} d_i)$

Speedup — Observation


$\mathrm{Ker}($ To leave $,$ Candidate 1 $)$

Old Team **New Team**

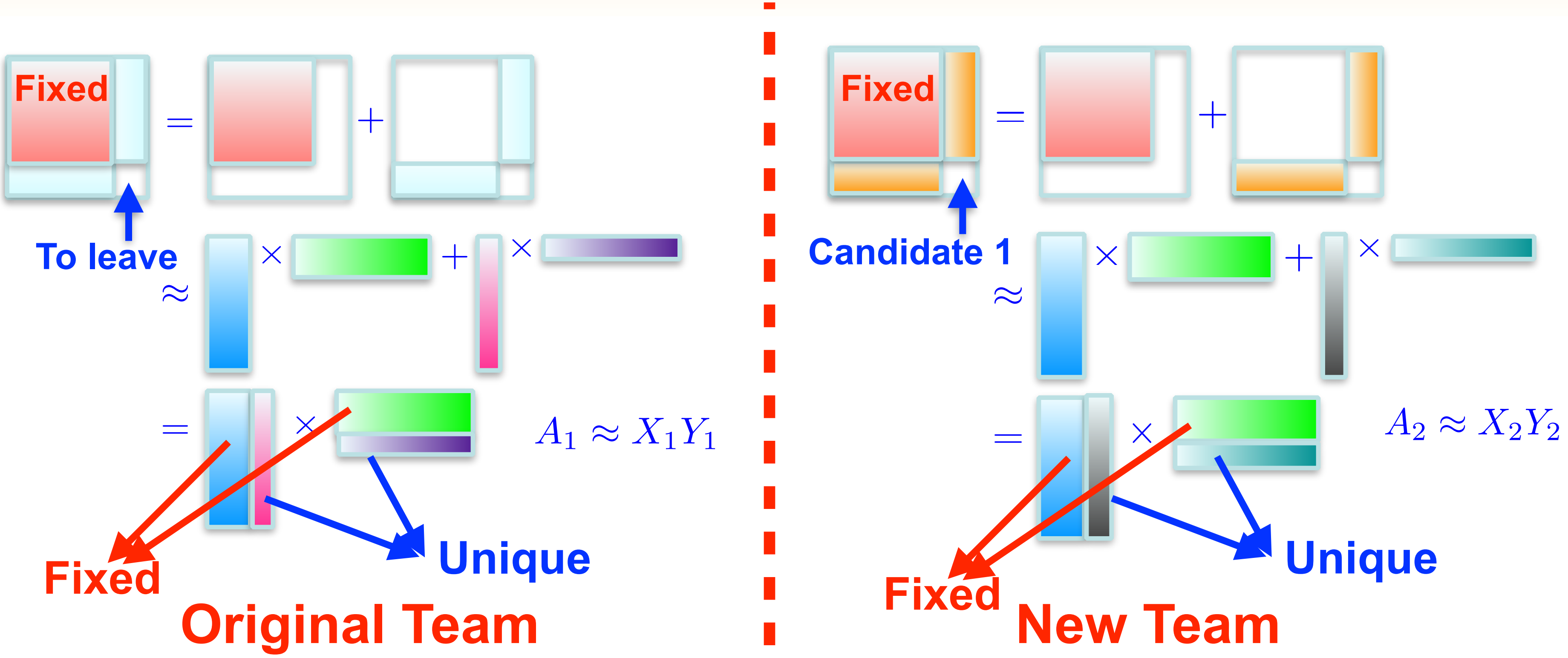
$\mathrm{Ker}($ To leave $,$ Candidate 2 $)$

Old Team **New Team**

Observation:

Many redundancies — *the nodes and edges within the rest team members remain the same*

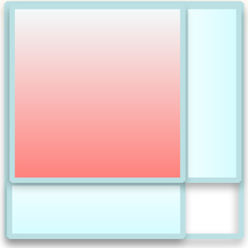
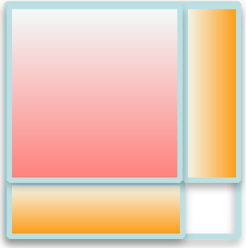
Speedup — Approx Approach



The common part is the adjacency matrix of the rest team members

Speedup — Approx Approach

Details

$\mathrm{Ker}($  ,  $)$

$$\approx y'(1 - cL_{\times}(X_1Y_1)\otimes(X_2Y_2))^{-1}L_{\times}x$$

$$= y'L_{\times}x + cy'L_{\times}(X_1\otimes X_2)\boxed{M}(Y_1\otimes Y_2))L_{\times}x$$

$$M = (I - c(\sum_{j=1}^{l} Y_1L_1^{(j)}X_1 \otimes Y_2L_2^{(j)}X_2))^{-1}$$

M is of size $(r+2)^2 \times (r+2)^2$

Time Complexity: $O((\sum_{i\in\mathcal{T}/p} d_i)(lt^2r + r^6))$

Original Complexity: $O(nt^3)$

$[\sum_{i\in\mathcal{T}/p} d_i \ll n, r \ll t]$

Roadmap

- Motivations
- Proposed Solutions
- Experimental Results
- Conclusion

Datasets

- *DBLP*: For a given paper, treat co-authors as a team, use conferences as skills (e.g., WWW, KDD, etc)
- *Movie*: For a given movie, treat actors/actresses as a team, use movie genres as skills (e.g.,action, comedy, etc)
- *NBA*: team of a season, use position as skill (guard, forward, center)

Data	n	m	# of teams
DBLP	916,978	3,063,244	1,572,278
Movie	95,321	3,661,679	10,197
NBA	3,924	126,994	1,398

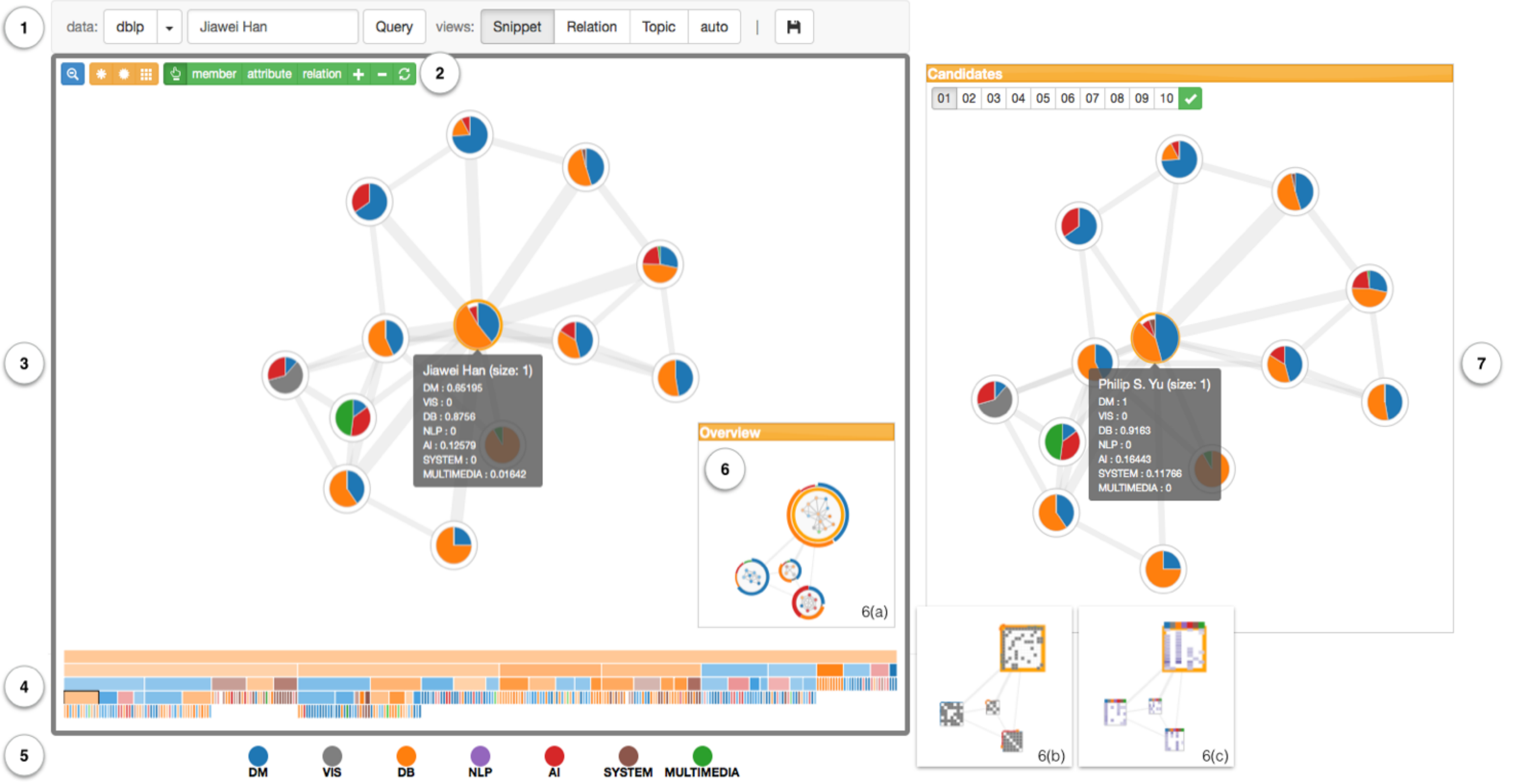
Questions

- **Q1**: How effective is skill + structure?
- **Q2**: How fast is pruning?
- **Q3:** How fast is proposed solution?
- **Q4**: How is the scalability?

A Case Study on DBLP

Questions

- **Q1: How effective is skill + structure?**
- Q2: How fast is pruning?
- Q3: How fast is proposed solution?
- Q4: How is the scalability?



data: dblp

Jiawei Han

Jiawei Han (size: 1)
DM : 0.65195
VIS : 0
DB : 0.8756
NLP : 0
AI : 0.12579
SYSTEM : 0
MULTIMEDIA : 0.01642

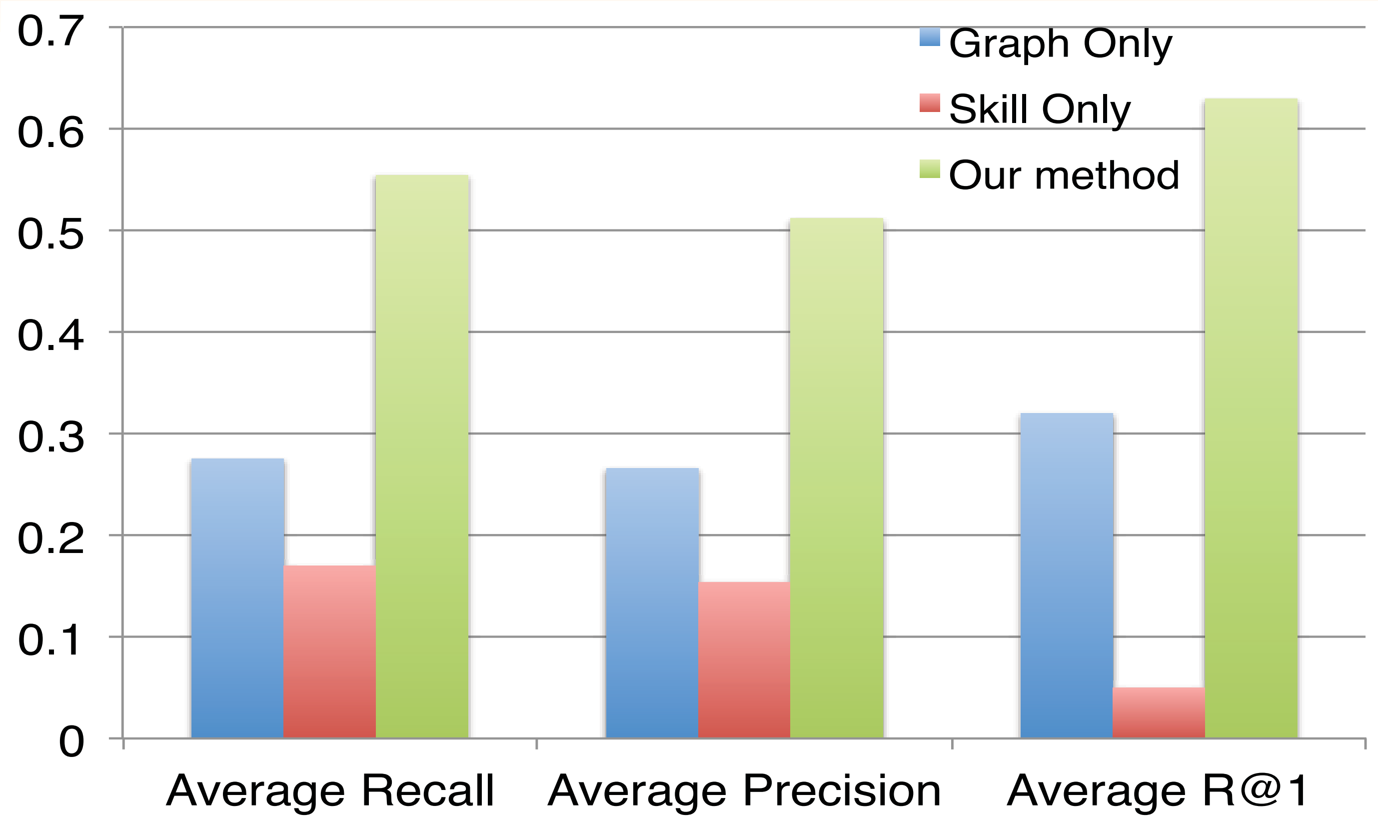
Philip S. Yu (size: 1)
DM : 1
VIS : 0
DB : 0.9163
NLP : 0
AI : 0.16443
SYSTEM : 0.11766
MULTIMEDIA : 0

prototype: http://team-net-work.org

User Studies

- Perform a user study with 20 people aged from 22-35
- Choose 10 papers from various fields, replace one author of each paper, run comparison methods and each recommends top five candidates
- Mix the outputs and ask users to (a) mark one best replacement (b) mark all good replacements

User Studies

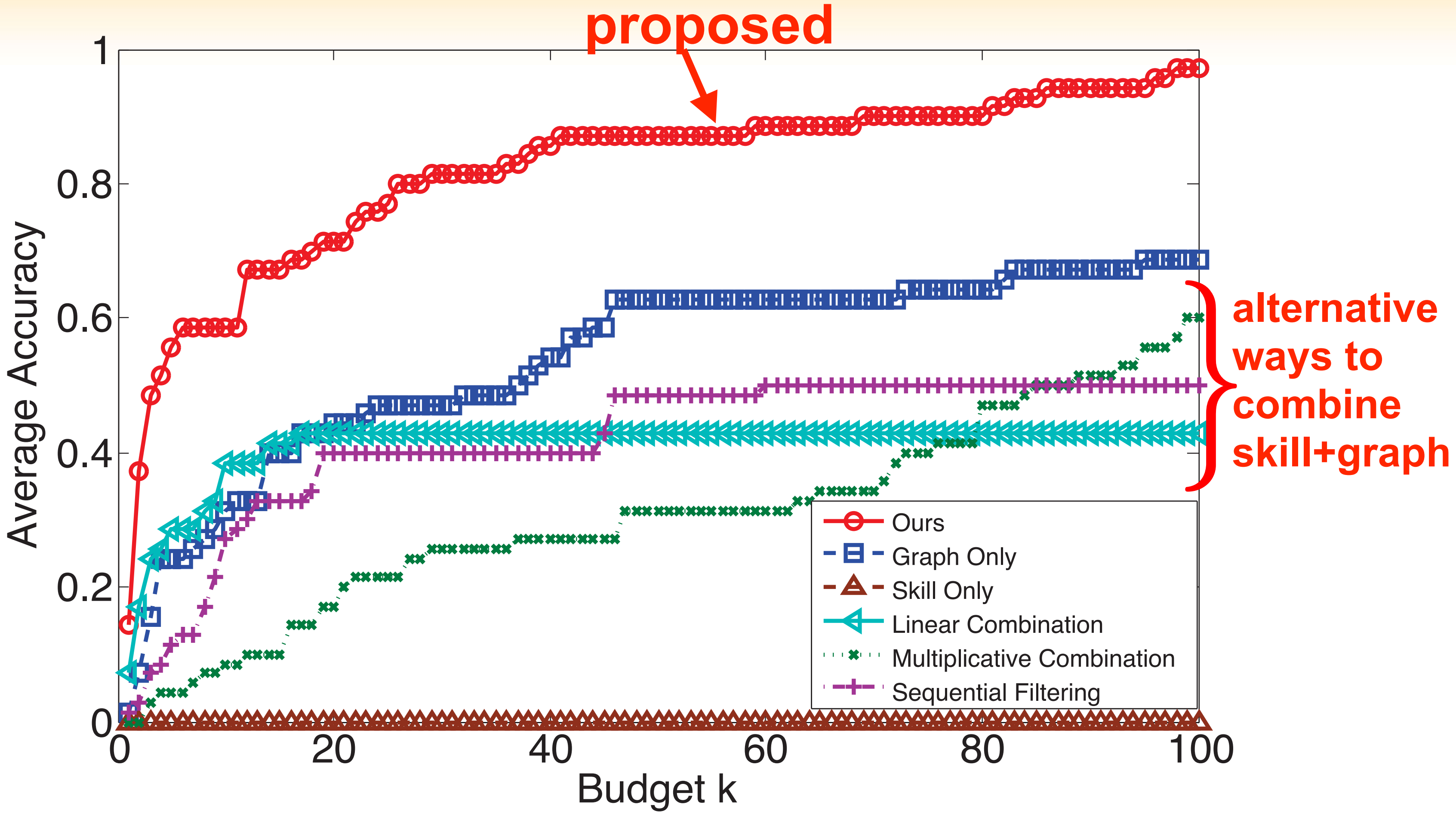


Our method achieves the best average recall, precision and R@1

Author Alias Prediction

- Author Alias, e.g., *Alexander J. Smola* vs. *Alex J. Smola*
- For such an author, run the team replacement algorithms on papers s/he was involved
- If the other alias appears in the top-k list, treat it as ***hit***

Author Alias Prediction

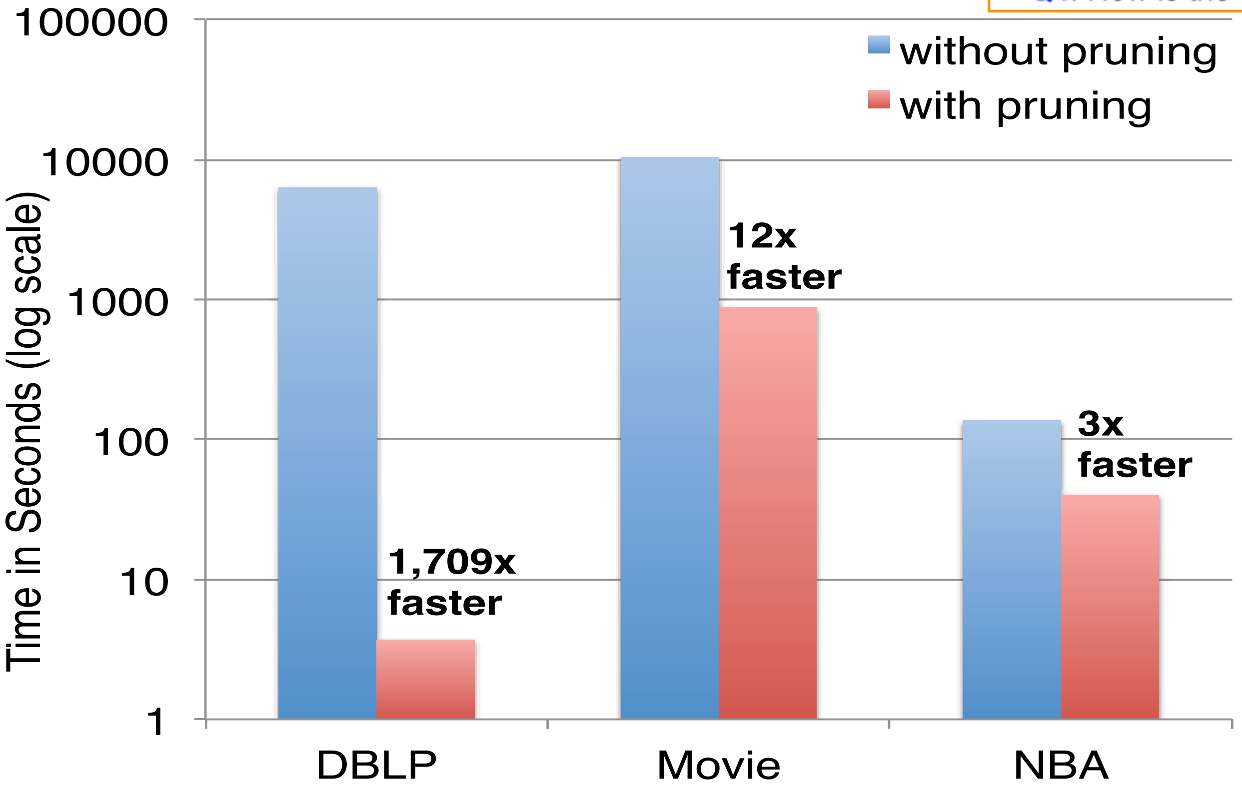


Our method achieves the highest accuracy

Speed-up by Pruning

Questions

- **Q1**: How effective is skill + structure?
- **Q2: How fast is pruning?**
- **Q3:** How fast is proposed solution?
- **Q4**: How is the scalability?

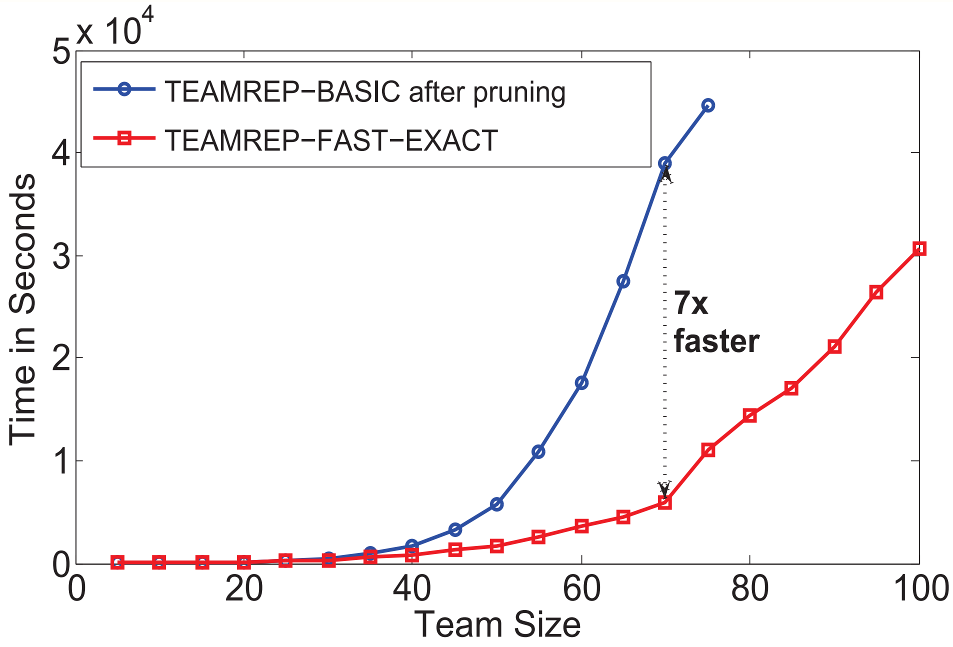


Pruning has dramatic speed improvement

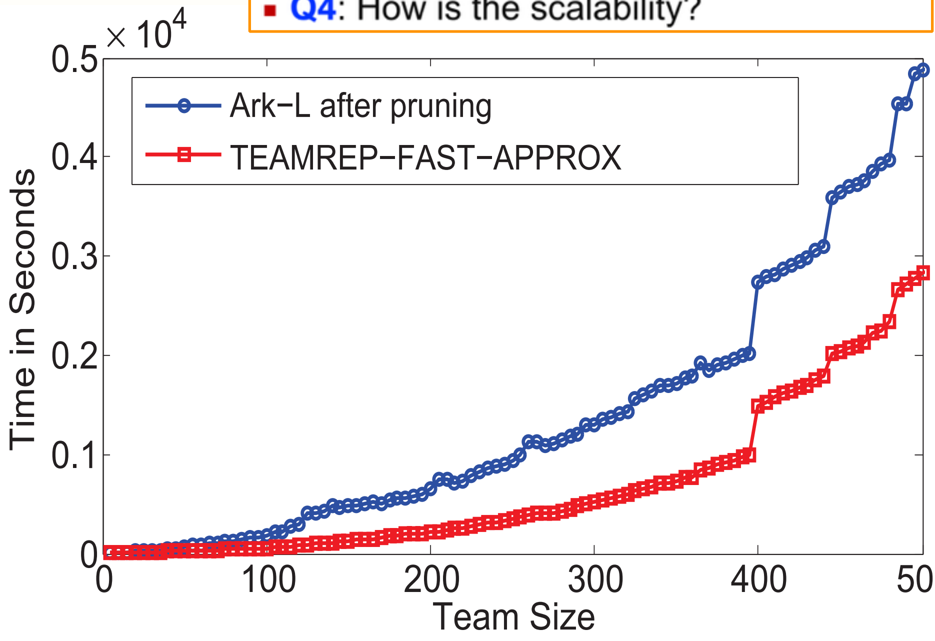
Further Speed-up

Questions

- **Q1**: How effective is skill + structure?
- **Q2**: How fast is pruning?
- **Q3: How fast is proposed solution?**
- **Q4**: How is the scalability?



Exact methods



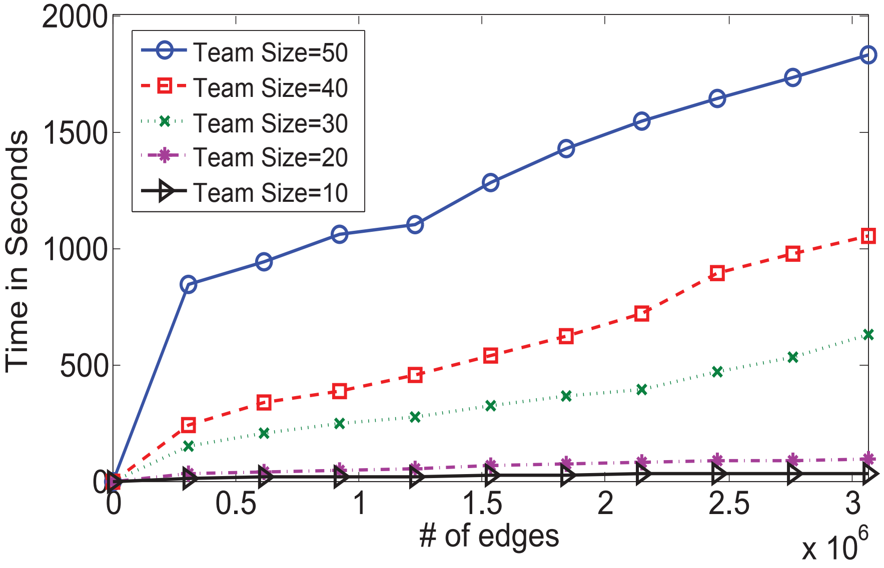
Approximate methods

Our fast solutions achieve significant speed improvement

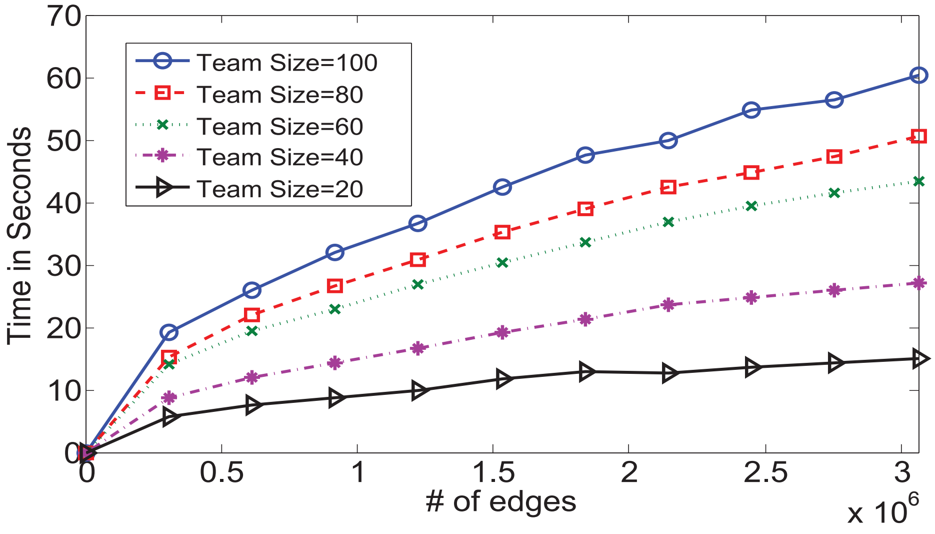
Scalability

Questions

- **Q1**: How effective is skill + structure?
- **Q2**: How fast is pruning?
- **Q3:** How fast is proposed solution?
- **Q4**: **How is the scalability?**



TEAMREP-FAST-EXACT



TEAMREP-FAST-APPROX

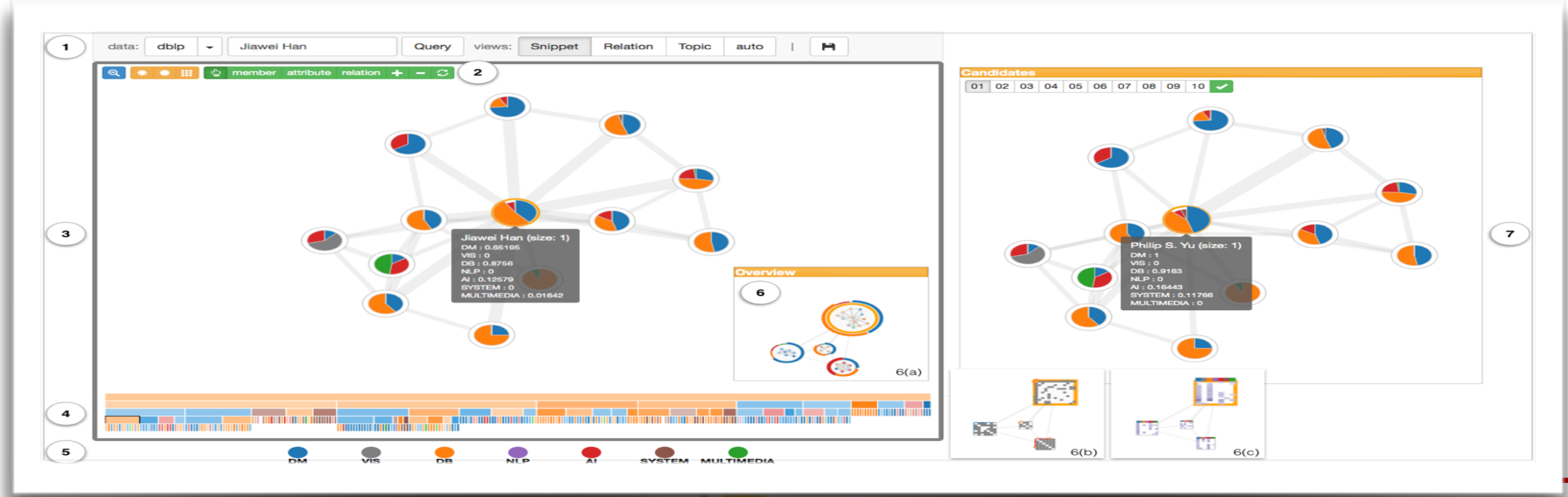
Our fast solutions scale sub-linearly

Roadmap

- Motivations
- Proposed Solutions
- Experimental Results
- Conclusions

Conclusions

- **Problem Def**: Team Member Replacement
- **Design Objectives**: skill matching & structural matching
- **Solutions**: graph kernel and fast algorithms
- **Systems:** http://team-net-work.org/



Jiawei Han (size: 1)
DM : 0.65195
VIS : 0
DB : 0.8756
NLP : 0
AI : 0.12579
SYSTEM : 0
MULTIMEDIA : 0.01642

Philip S. Yu (size: 1)
DM : 1
VIS : 0
DB : 0.9163
NLP : 0
AI : 0.16443
SYSTEM : 0.11766
MULTIMEDIA : 0