KDD2018

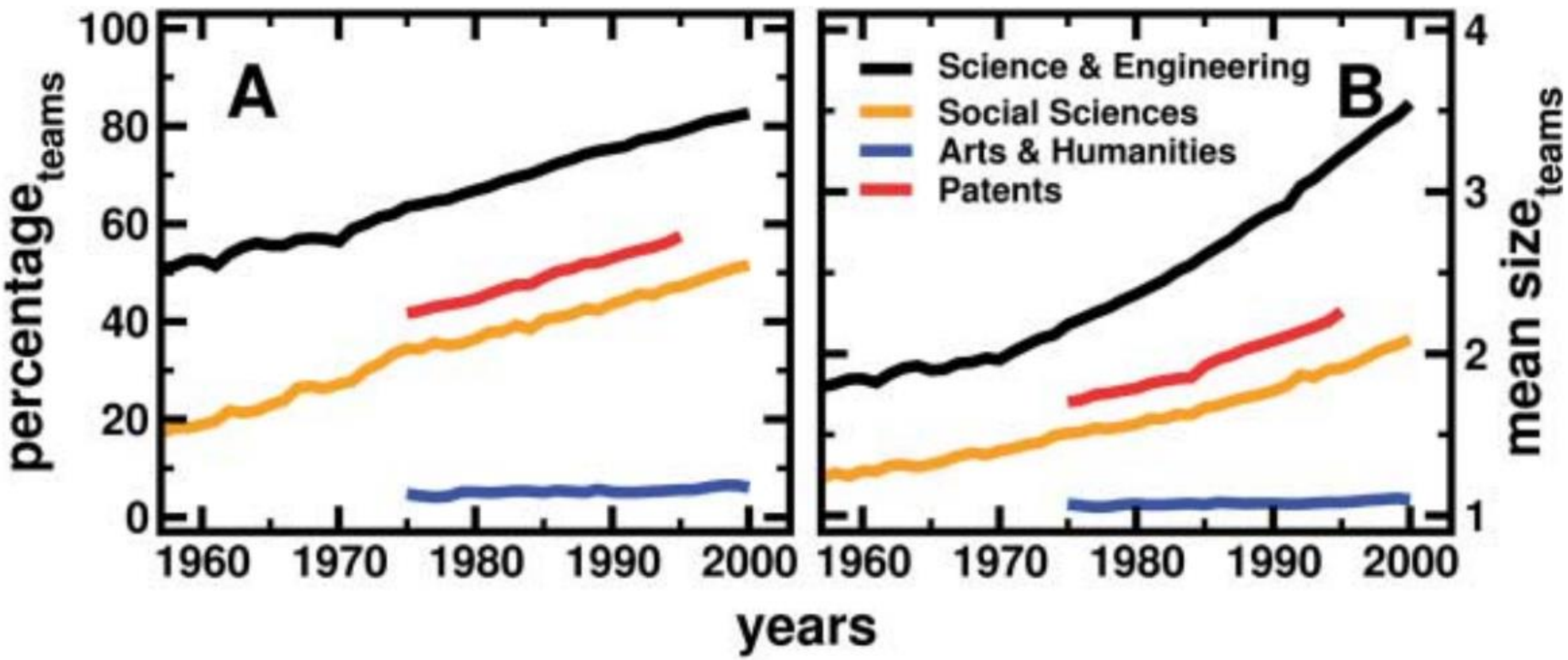
Network Science of Teams: Characterization, Prediction, and Optimization

Liangyue Li, Hanghang Tong

Arizona State University

Slides can be downloaded from:
http://www.public.asu.edu/~liangyue/team-tutorial.html

Shift from Individuals to Teams



Teams increasingly dominate solo authors in the production of knowledge

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. Sales Team

4. Research Team

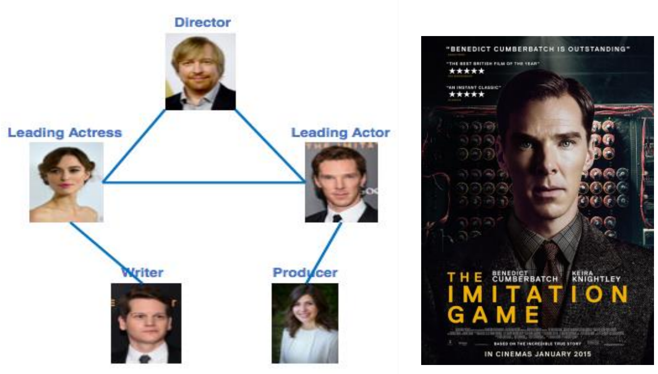
5. Military Team

6. Development Team

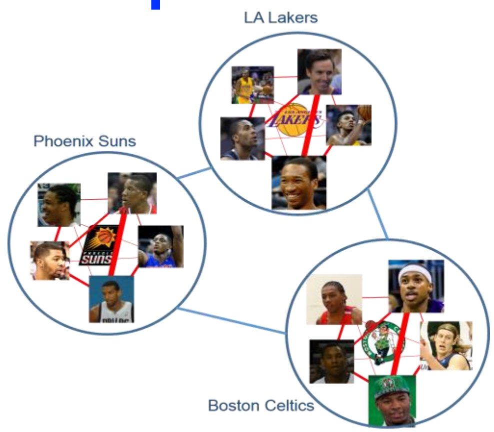
Wuchty, Stefan, Ben Jones, and Brian Uzzi. "The Increasing Dominance of Teams in the Production of Knowledge," Science, May 2007, 316:1036-1039.

Networks Are Everywhere in Teams

1. Film Crew



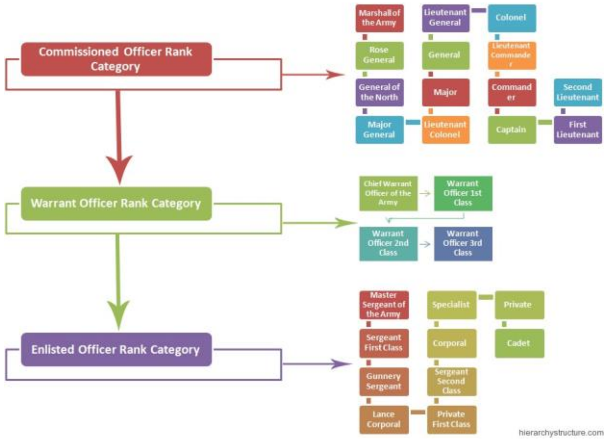
2. Sports Team



3. Sales Team

4. Research Team

5. Military Team

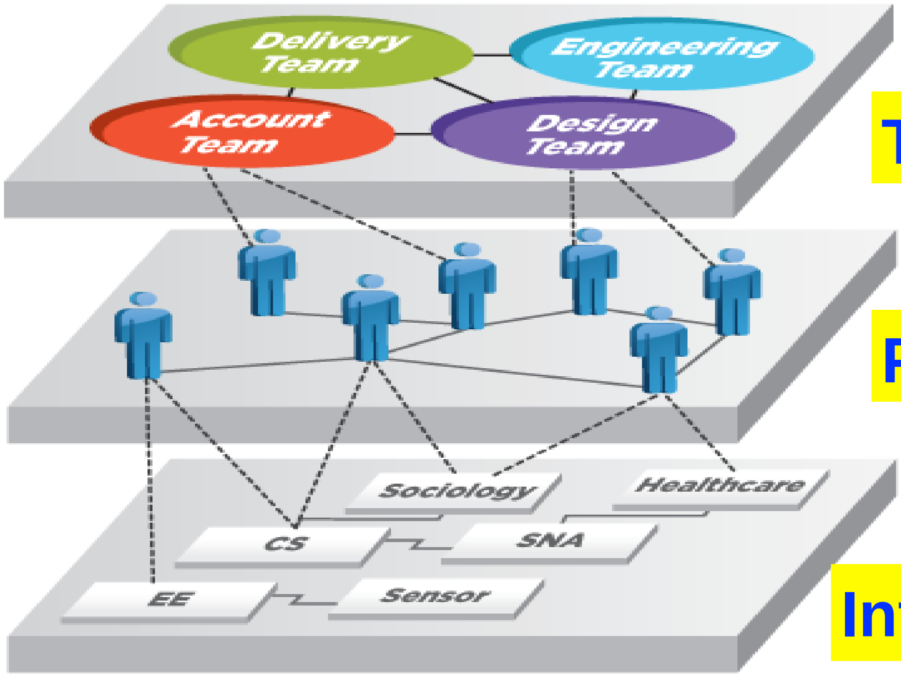


6. Development Team

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

Network Science of Teams

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.

Research Questions

- Q1: What do high-performing teams share in common? [Uzzi+Science13]
- Q2: How to foresee the success at an early stage? [Wang+Science13]
- Q3: What's the optimal design for a team in the context of networks?[Lappas+KDD09, Rangapuram+WWW13]

- S. Wuchty, B. Jones, and B. Uzzi. The Increasing Dominance of Teams in the Production of Knowledge, Science, 2007
- D. Wang, C. Song, and A.-L. Barabasi. Quantifying long-term scientific impact. Science, 342(6154): 127-132, 2013.
- T. Lappas, K. Liu, and E. Terzi. Finding a team of experts in social networks. In KDD, pages 467–476, 2009.
- S. S. Rangapuram, T. Buhler, and M. Hein. Towards realistic team formation in social networks based on densest subgraphs. WWW

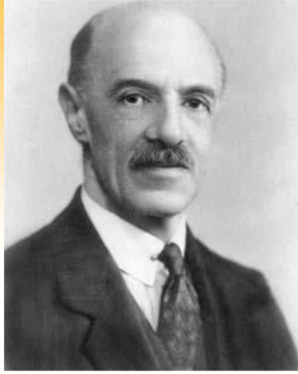
Roadmap

- Motivations and Background
- Part I: Team Performance Characterization
- Part II: Team Performance Prediction
- Part III: Team Performance Optimization
- Part IV: Open Challenges
- Demo

Part I: Team Performance Characterization

- Collective Intelligence
- Virtual Teams in online games
- Network in Sports Teams
- Network in Github Teams

Individual Intelligence



- Spearman's *g*
  - Individuals take a diverse set of cognitive tasks
  - The first factor extracted in a factor analysis of these scores accounts for 30% to 50% of the variance

Woolley, Anita Williams, et al. "Evidence for a collective intelligence factor in the performance of human groups." science 330.6004 (2010): 686-688

Collective Intelligence

- Definition: general ability of the group to perform a wide variety of tasks
- Question: Is there a single factor for groups?

Woolley, Anita Williams, et al. "Evidence for a collective intelligence factor in the performance of human groups." science 330.6004 (2010): 686-688

Study 1

- 40 groups spend five hours together in the laboratory
- Work together on a diverse set of tasks, plus a more complex criterion task
- Also measured individual intelligence

Woolley, Anita Williams, et al. "Evidence for a collective intelligence factor in the performance of human groups." science 330.6004 (2010): 686-688

Example Tasks

Task	Description	Scoring
Generate	Brainstorming. Come up with as many uses for a brick as possible.	Scored on quantity and quality of ideas.
Choose	Intellective. Members answer a set of Raven's Matrices questions as a group.	Scored on correctness.
Negotiate	Devise a shopping trip using a shared car so that all members can get as many of their items at the best places possible.	Cumulative score of all group members.
Execute	Typing task. Members must collectively type difficult text into a shared online document.	Scored on number of words typed minus errors and skipped words in limited time period.

Woolley, Anita Williams, et al. "Evidence for a collective intelligence factor in the performance of human groups." science 330.6004 (2010): 686-688

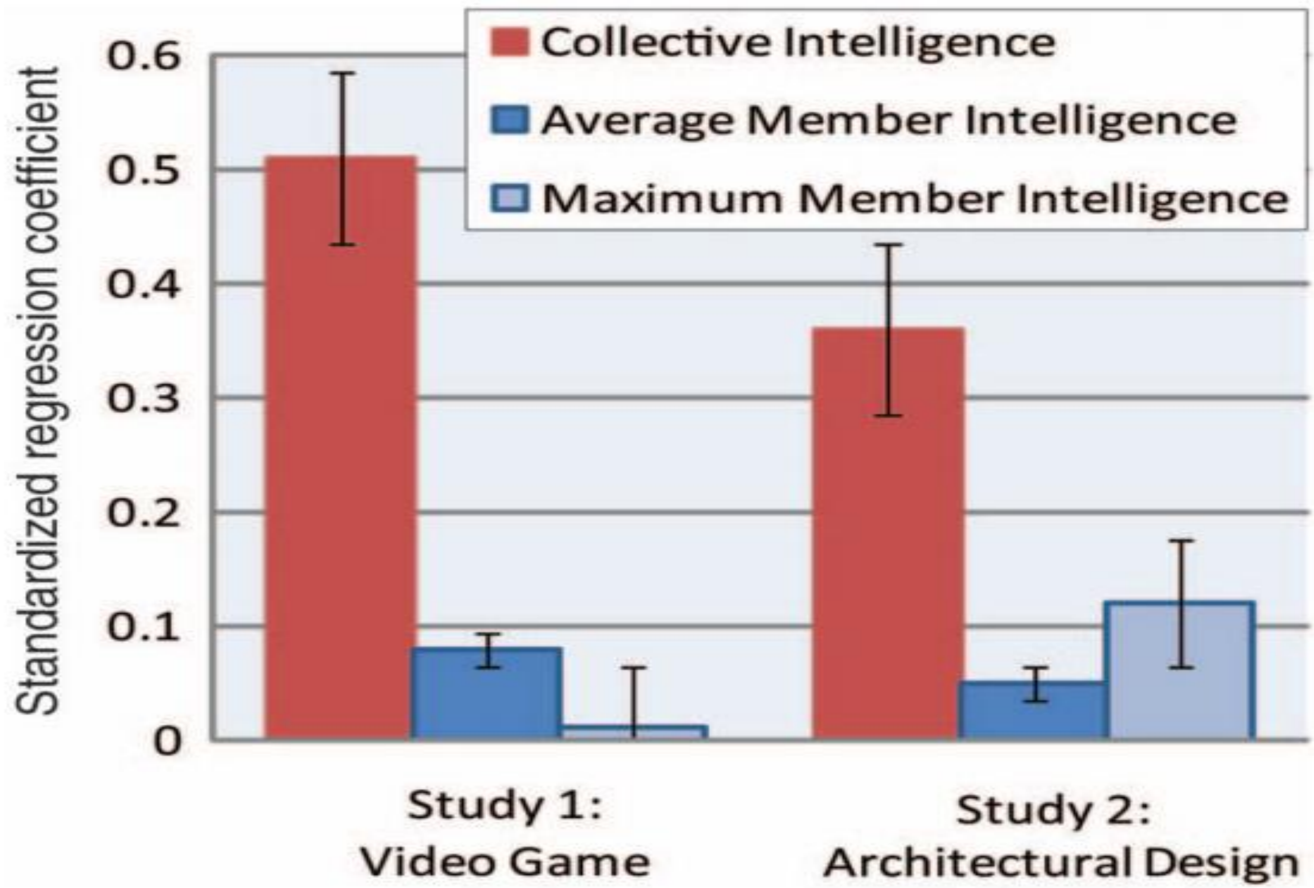
Study 1

- Average inter-item correlation = .28
- First principal component accounts for 43% of variance
- Factor loadings on the first factor are used to calculate *c* score – strongly predicts the performance on the criterion task
- Avg and max individual intelligence not predictive of criterion task performance

Woolley, Anita Williams, et al. "Evidence for a collective intelligence factor in the performance of human groups." science 330.6004 (2010): 686-688

Study 2

- 152 groups ranging from 2-5 members
- Replicate findings using broader tasks



But what can predict *c*

- Average social perceptiveness

"Reading the Mind in the Eyes" Baron-Cohen et al., 2001

Woolley, Anita Williams, et al. "Evidence for a collective intelligence factor in the performance of human groups." science 330.6004 (2010): 686-688

But what can predict *c*

- The proportion of females positively correlate with *c*
  - Might be mediated by social perceptiveness
- The variance in the number of speaking turns negatively correlate with *c*

Woolley, Anita Williams, et al. "Evidence for a collective intelligence factor in the performance of human groups." science 330.6004 (2010): 686-688

Virtual Teams

- Does collective intelligence exist in virtual teams where face-to-face interaction is not available?
  - Multiplayer Online Battle Arena (MOBA) teams

Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

League of Legends

- A match is between two five-person teams
- Matchmaking algorithms vs. self-organize
- A team’s goal is to destroy the opponent team’s base

Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

Study Hypotheses

- **H1**: CI will predict team performance in League of Legends
- **H2**: social perceptiveness and proportion of woman will be positively associated with CI in League of Legends teams
- **H3**: CI will not be associated with equality of contribution to conversation or decision making in LOL teams.

Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

Method

- Data for CI, game performance, team characteristics
  - All team members individually completed a questionnaire (demographic, psychological variables, cognition,affect)
  - Test of Collective Intelligence
  - In game data (performance metrics, play history, statistics)

Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

Sample

- Research advertisement on official community board
- 248 teams completed all components
- 97% male, avg age is 22

Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

Results

- CI factor analysis
  - Factor analysis of scores on all tasks in TCI yielded one factor accounting for 28.28% of the variance

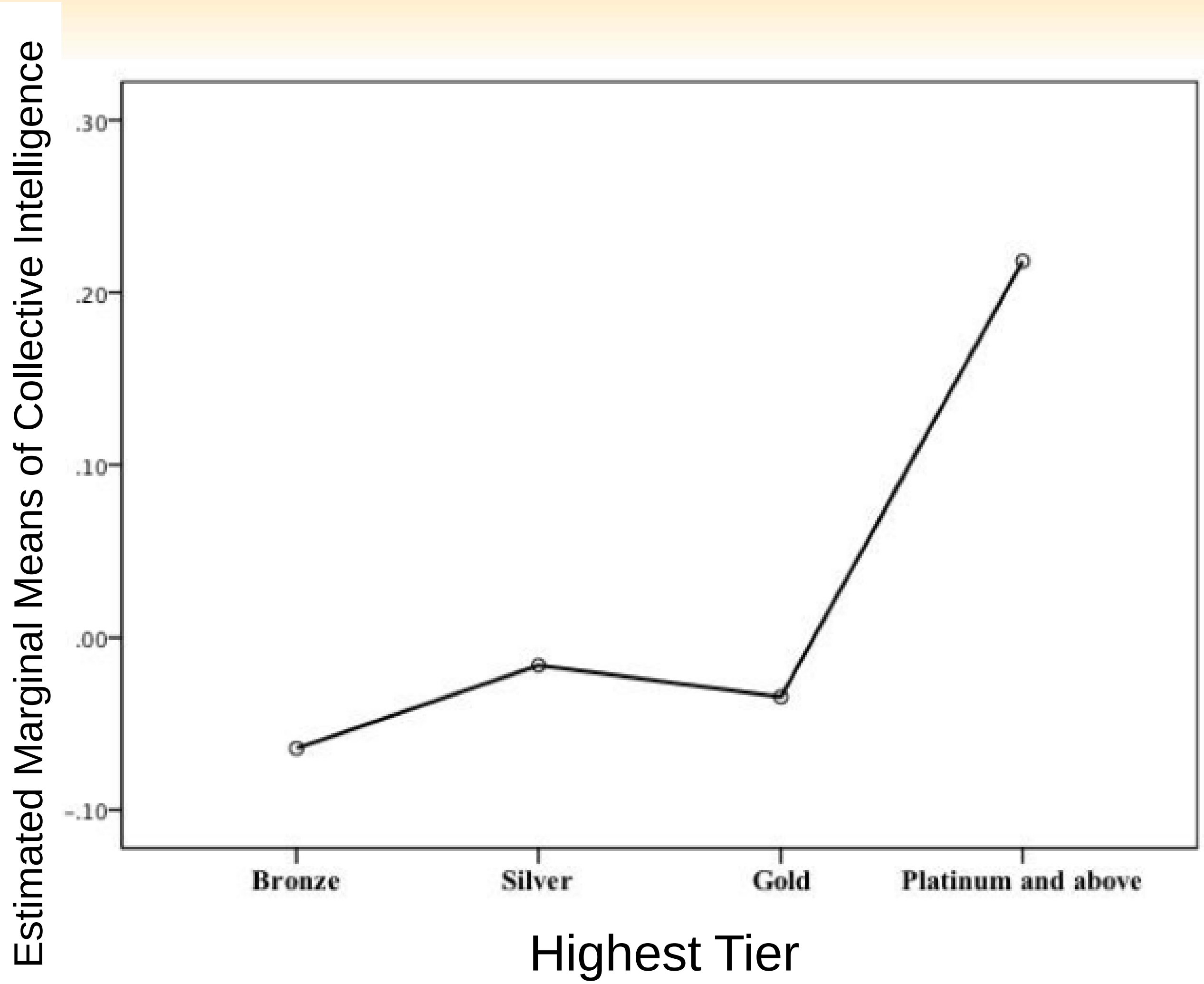
Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

H1:CI and Game Performance

	MMR at Time of Study		MMR after 6 Months	
	Step 1	Step 2	Step 1	Step 2
Individual Play Time	.30***	.32***	.27***	.28***
Team Play Time	-.22***	-.22***	-.21**	-.22***
Collective Intelligence		.14*		.15*
R^2	.14	.16	.11	.14
R^2 Change		.02*		.02*

Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

H1:CI and Game Performance



Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

H2: Women, Social Perceptiveness and CI

- CI is positively correlated with the number of woman in the team ($r=0.18$, $p=0.005$)
- CI is positively correlated with social perceptiveness ($r=0.14$, $p=0.03$)

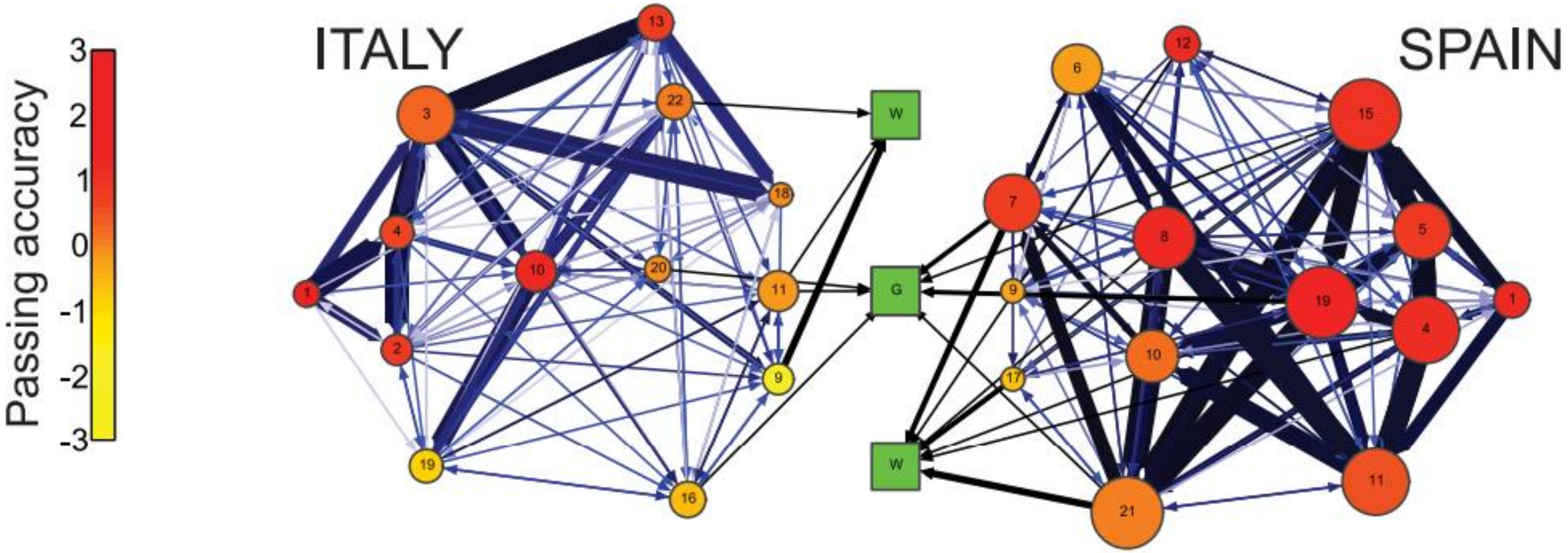
Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

H3: Communication Processes and CI

- Standard deviation of chat lines and chat word count, is not significantly correlated with CI
- CI negatively correlates with
  - perceived equality in decision making,
  - frequency of game-specific communication
  - strategy-related process
  - team learning behavior

Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

Network in Sports Teams



Flow Network:

Node: players

Arc weights: passing success rate btw two players

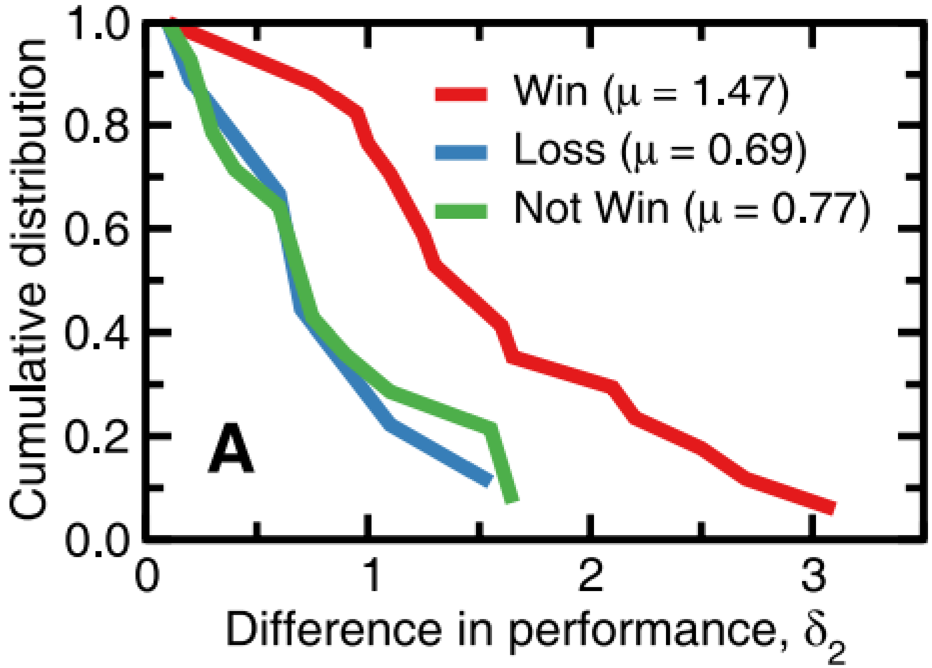
Duch, Jordi, Joshua S. Waitzman, and Luís A. Nunes Amaral. "Quantifying the performance of individual players in a team activity." PloS one 5.6 (2010): e10937.

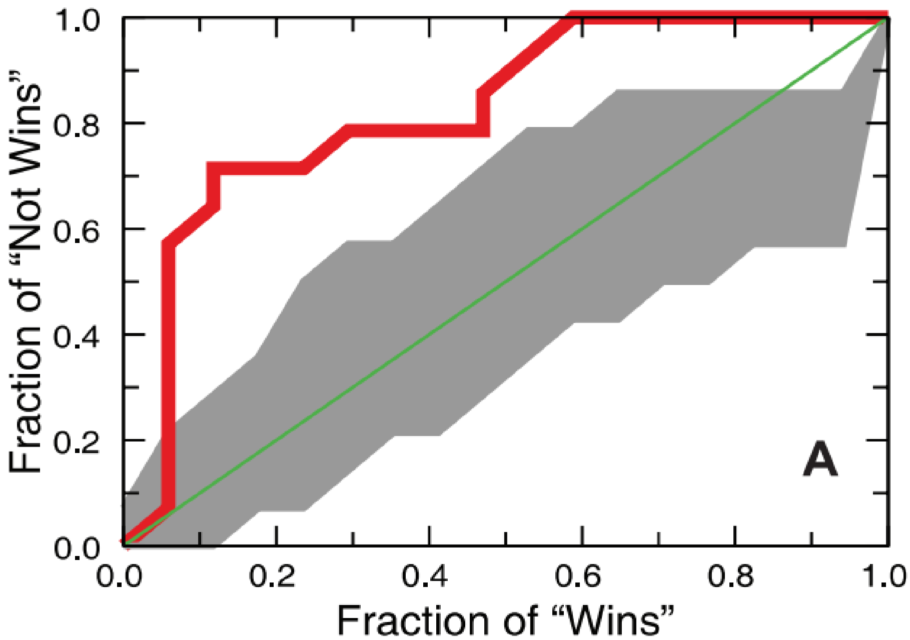
Team Performance

- Match performance of player: normalized value of log of the player’s betweenness centrality
- Team performance: avg performance of the top k players
- Difference between two teams indicate winning probability

Duch, Jordi, Joshua S. Waitzman, and Luís A. Nunes Amaral. "Quantifying the performance of individual players in a team activity." PloS one 5.6 (2010): e10937.

Results





Duch, Jordi, Joshua S. Waitzman, and Luís A. Nunes Amaral. "Quantifying the performance of individual players in a team activity." PloS one 5.6 (2010): e10937.

Network Structure in Github

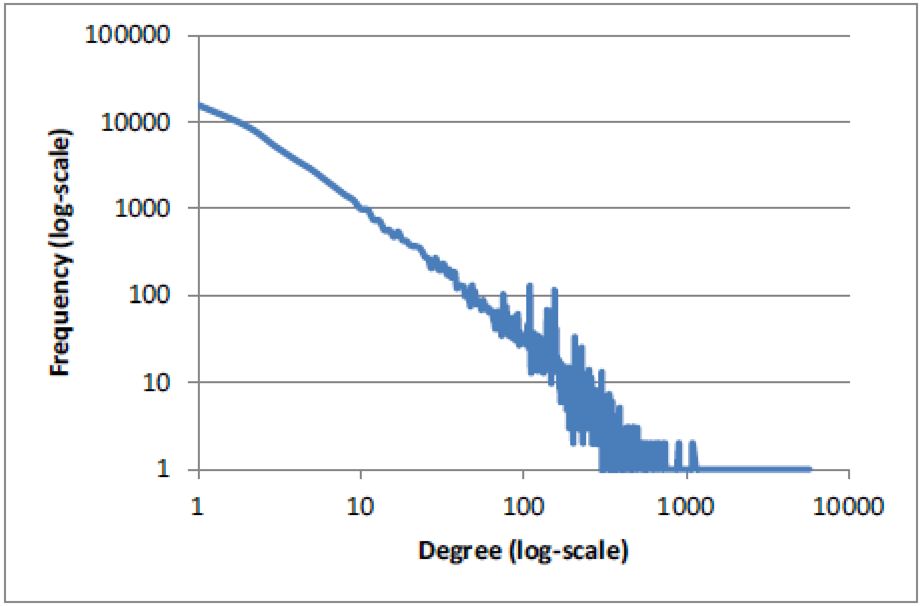
- Network Construction
  - Project-project network: two projects are connected if they share at least one developer
  - Developer-developer network: two developers are connected if they work together in at least one project

Github Data

- 100,000 projects retrieved from GitHub API
- 1,161,522 edges in the project-project network
- 23,678,455 edges in the developer-developer network

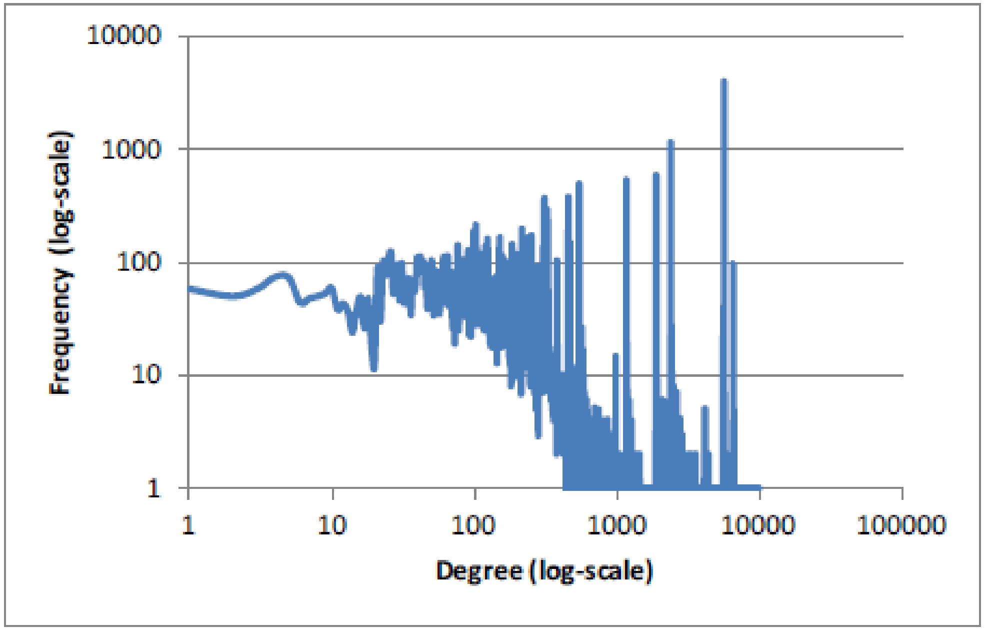
F. Thung, T. F. Bissyande, D. Lo, and L. Jiang, Network Structure of Social Coding in GitHub. CSMR 2013

Project-project network



The diameter of the largest connected component: 9
Avg shortest path: 3.7

Developer-developer network



The diameter of the largest connected component: 5
Avg shortest path: 2.47
-> compare with avg shortest path of Facebook: 4.7
Social coding enables substantially more collaborations among developers

F. Thung, T. F. Bissyande, D. Lo, and L. Jiang, Network Structure of Social Coding in GitHub. CSMR 2013

Influential Projects

Project url	PageRank
https://github.com/mxcl/homebrew	0.0009862
https://github.com/rails/rails	0.0006378
https://github.com/lifo/docrails	0.0006370
https://github.com/joyent/node	0.0002161
https://github.com/rubinius/rubinius	0.0001678

Table I
TOP 5 MOST INFLUENTIAL PROJECTS

F. Thung, T. F. Bissyande, D. Lo, and L. Jiang, Network Structure of Social Coding in GitHub. CSMR 2013

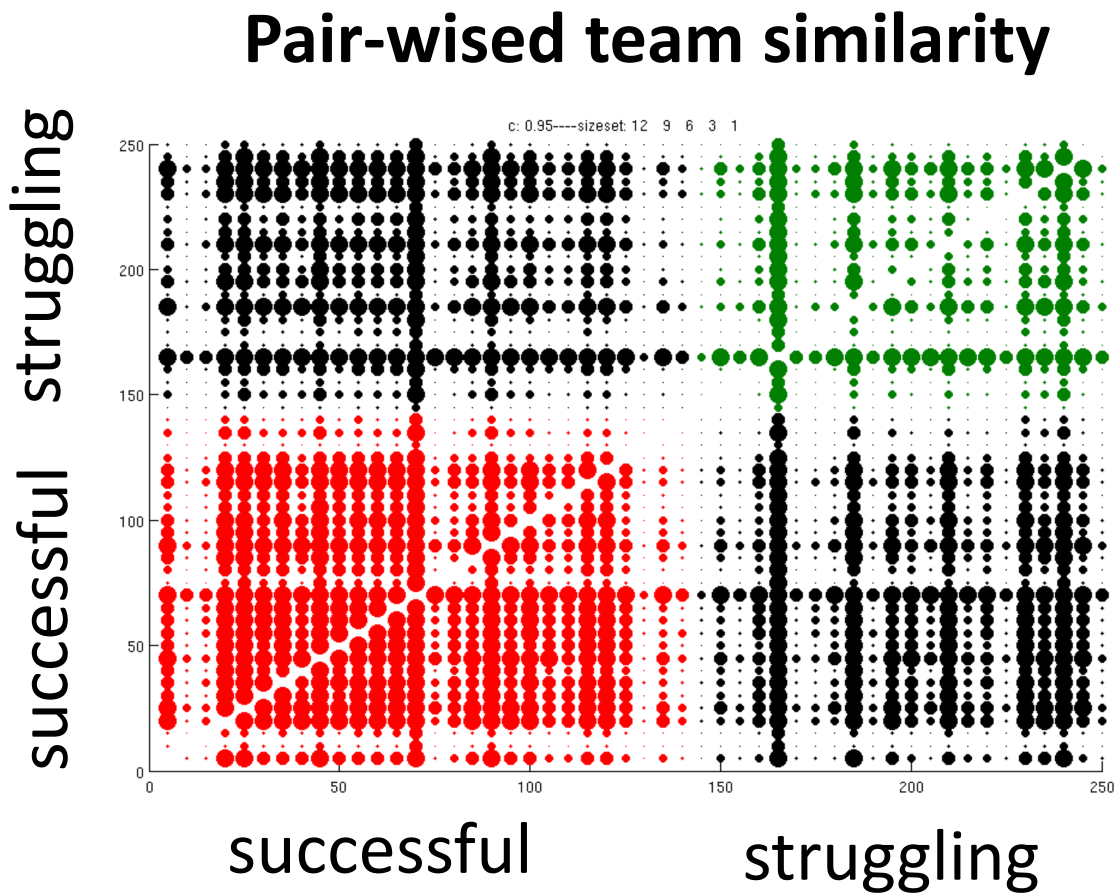
Influential Developers

Developer		PageRank
Joshua Peek	josh[AT]joshpeek.com	0.00009536
Aman Gupta	aman[AT]tmm1.net	0.00008860
Steve Richert	steve.richert[AT]gmail.com	0.00008850
Michael Klishin	michaelklishin[AT]me.com	0.00008170
Josh Kalderimis	josh.kalderimis[AT]gmail.com	0.00008163

Table II
TOP 5 MOST INFLUENTIAL DEVELOPERS

F. Thung, T. F. Bissyande, D. Lo, and L. Jiang, Network Structure of Social Coding in GitHub. CSMR 2013

The Effect of Team Network Connectivity



"Happy families resemble each other; each unhappy family is unhappy in its own way."
- Leo Tolstoy, Russian writer

Roadmap

- Motivations and Background
- Part I: Team Performance Characterization
- Part II: Team Performance Prediction
- Part III: Team Performance Optimization
- Open Challenges
- Demo

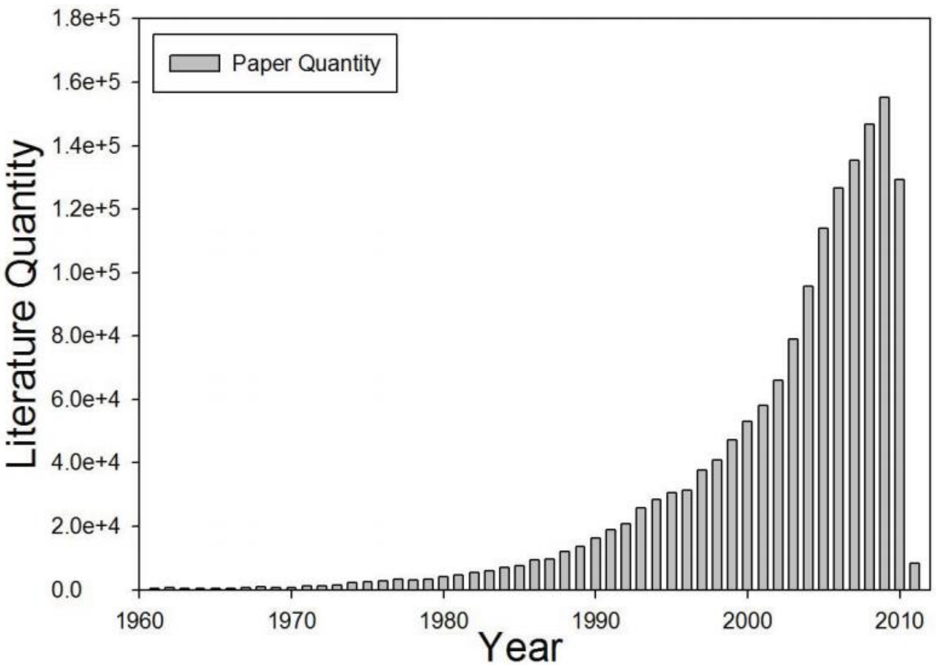
Part II: Team Performance Prediction

- Citation Count Prediction
- Mechanistic Model for Scientific Impact
- Long-term Performance
- Performance Trajectory
- Joint Modeling of Parts and Whole

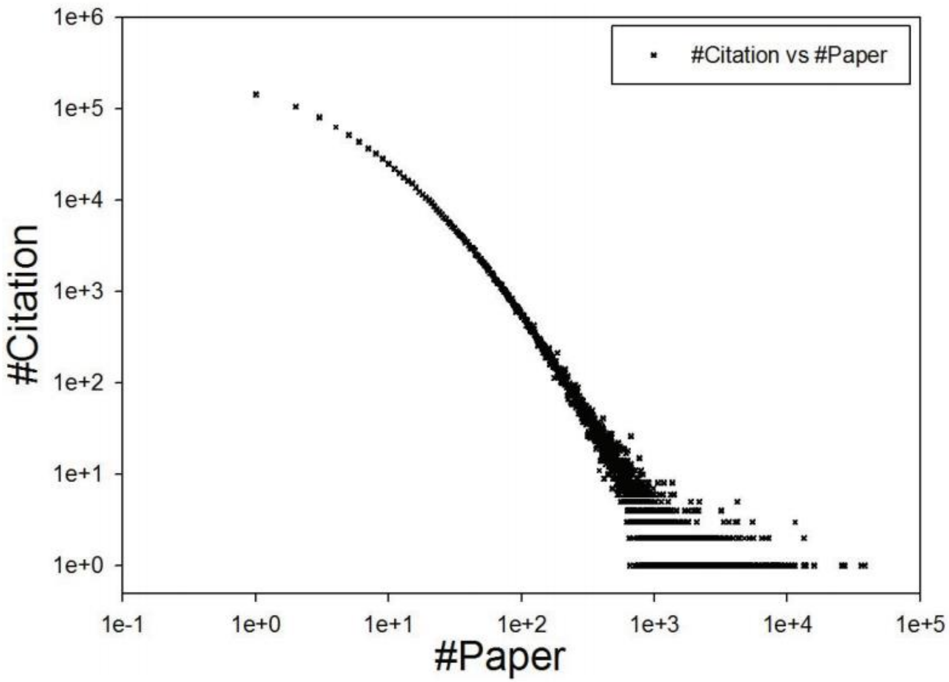
Scientific Teams

- Science of science
  - Prediction of future impact of scientific works
- Implications
  - Research grants evaluation
  - Scholarly awards dispensing

Scientific Impact



(a). The growing volume of literatures.



(b). Distribution of literature citation.

Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Factors driving scientific impact

- Content
- Author
  - Collaboration social network
- Venue
- Temporal

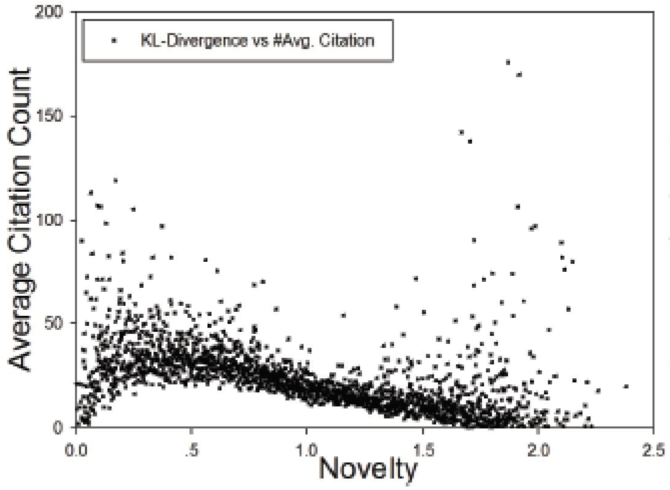
Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.
Yuxiao Dong, Reid A. Johnson, and Nitesh V. Chawla. 2015. Will This Paper Increase Your h-index?: Scientific Impact Prediction. WSDM, 2015.

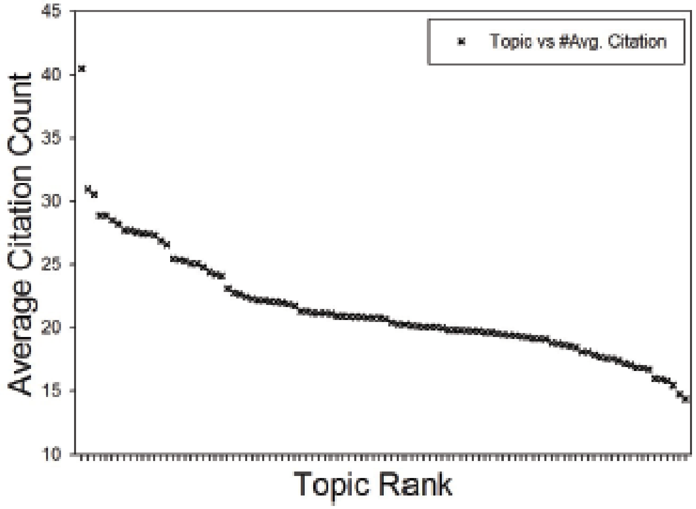
Content Features

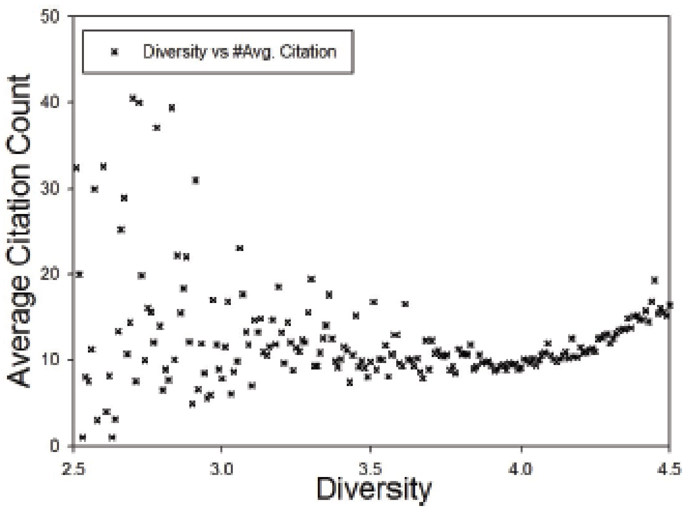
- Novelty: difference between a particular paper and the other publications
- Topic Rank: popular topics accumulate more citation counts than unpopular ones
- Diversity: the breadth of an article from its topic distributions

Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Content Features







Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Author Features

- Author Rank: “fame” of an author ensures the amount of citations
- H-index
- Past influence of authors
  - Maximum past influence
  - Total past influence
- Productivity: the number of published papers

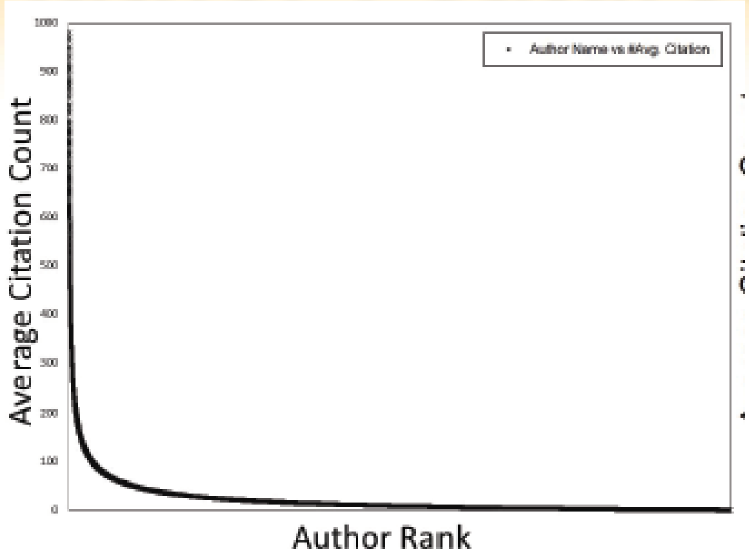
Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

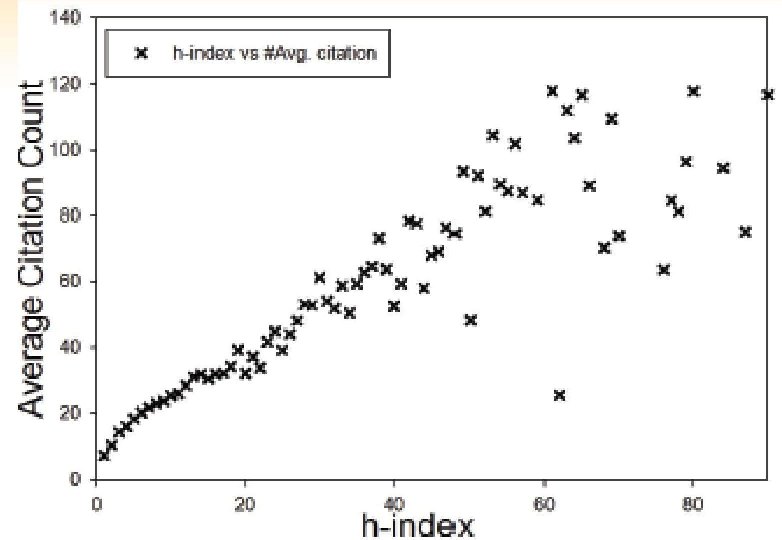
Author Features – con’t

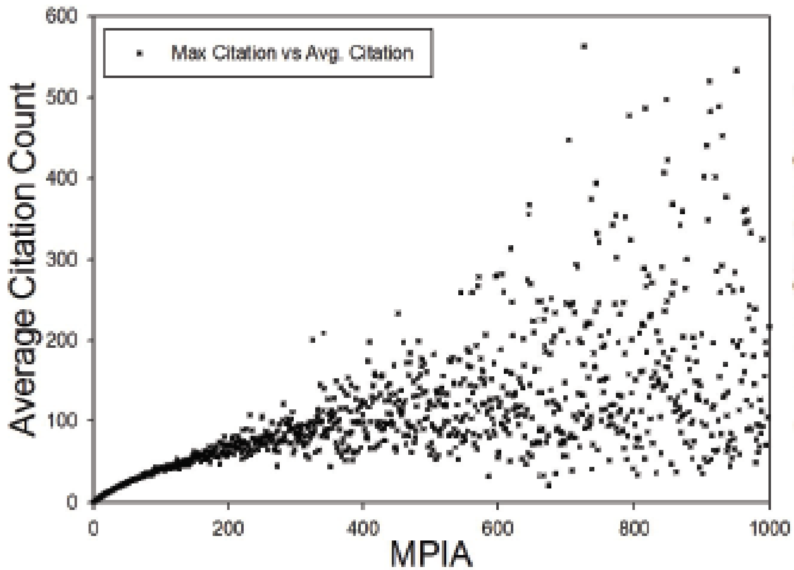
- Sociality: PageRank-like measure in co-author network
- Authority: PageRank-like measure in paper citation network and transmit paper authority to all its authors
- Versatility: topic breadth of an author’s research

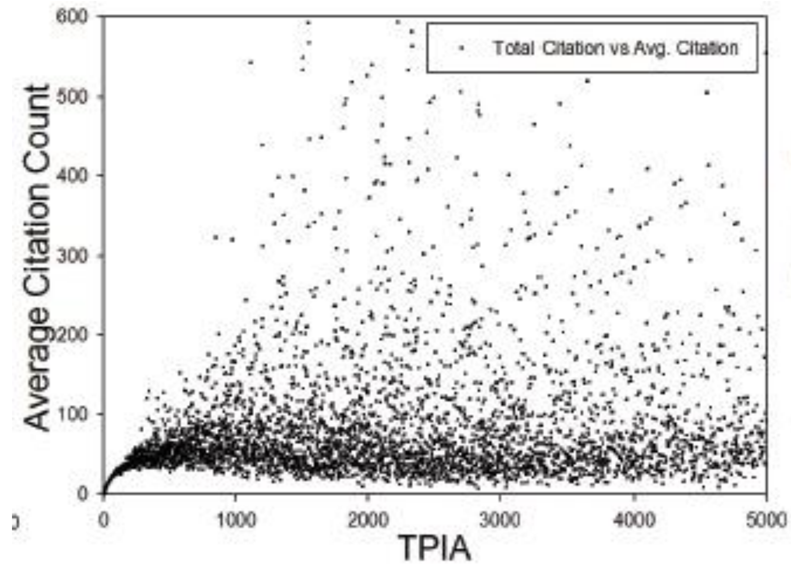
Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Author Features



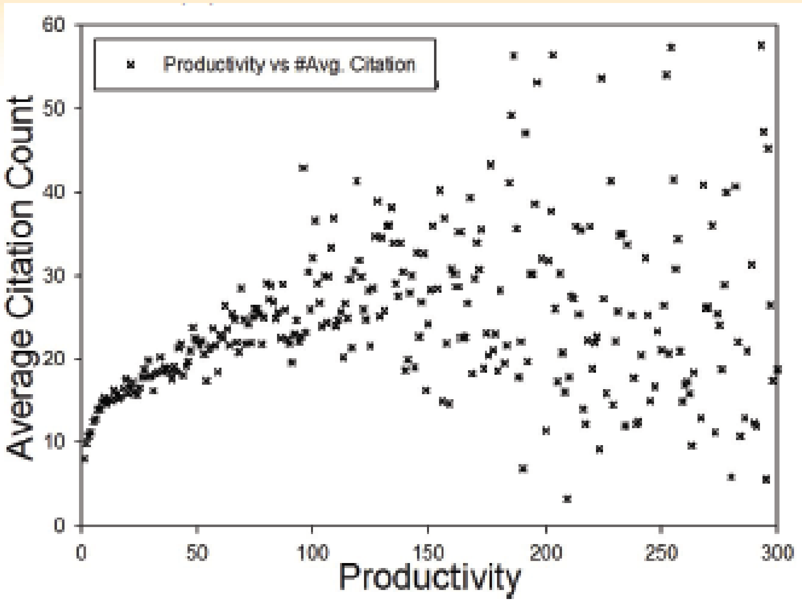


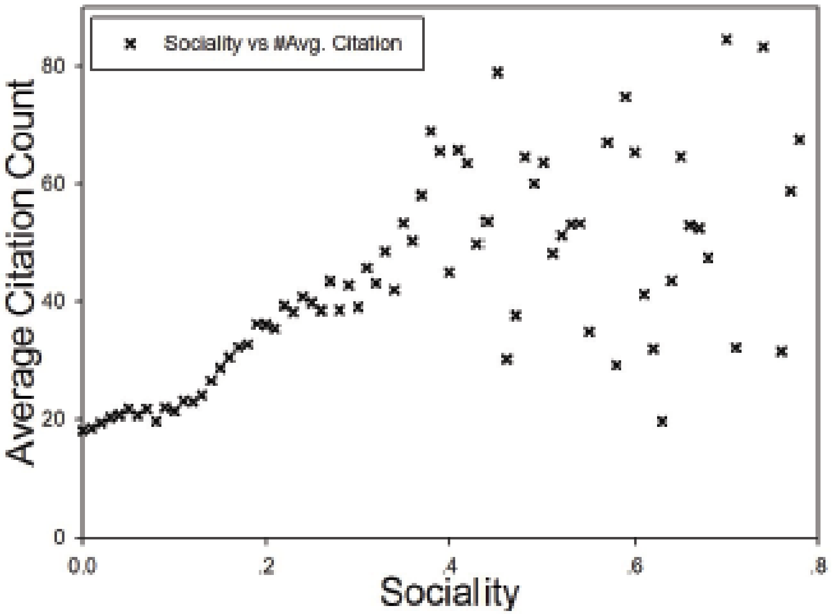


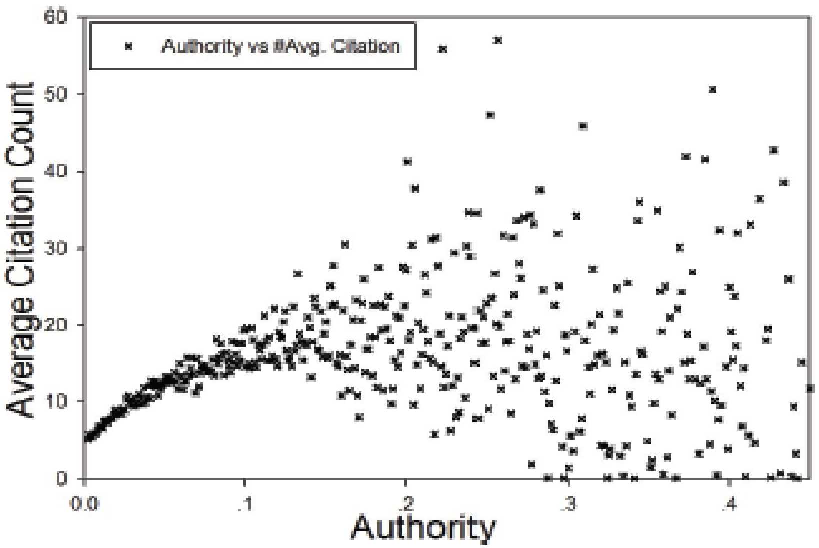


Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Author Features







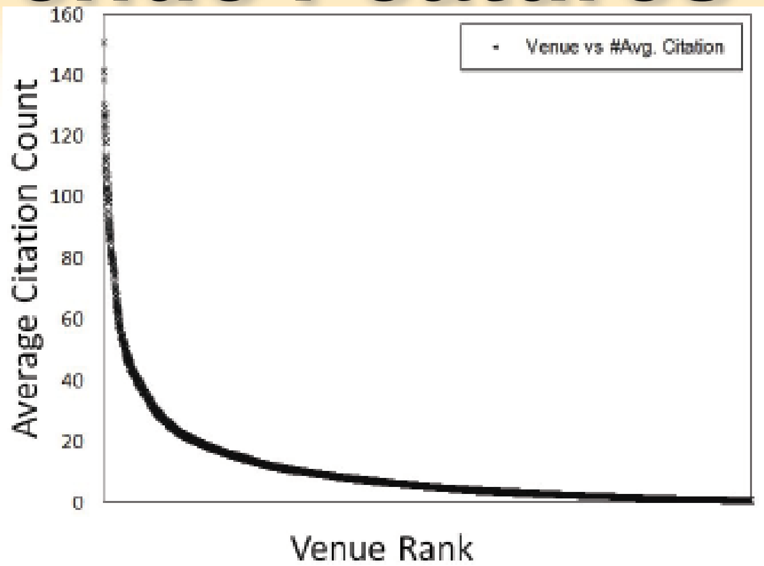
Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

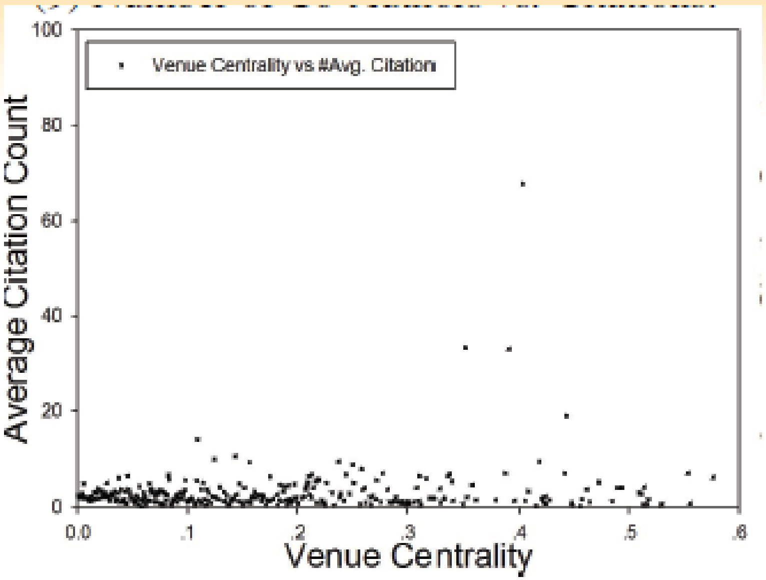
Venue Features

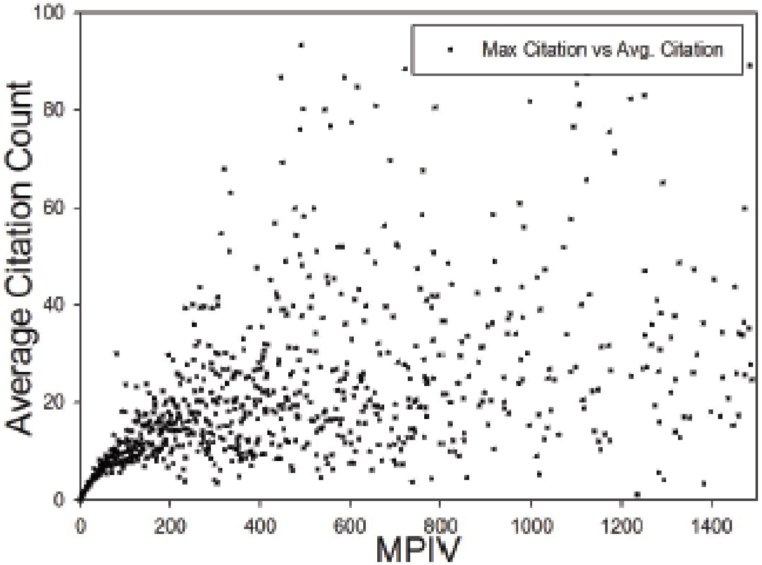
- Venue Rank: prestigious venues attract more focus
- Venue Centrality: PageRank-like measure in the venue citation network
- Past Influence of venues:
  - Maximum past influence
  - Total past influence

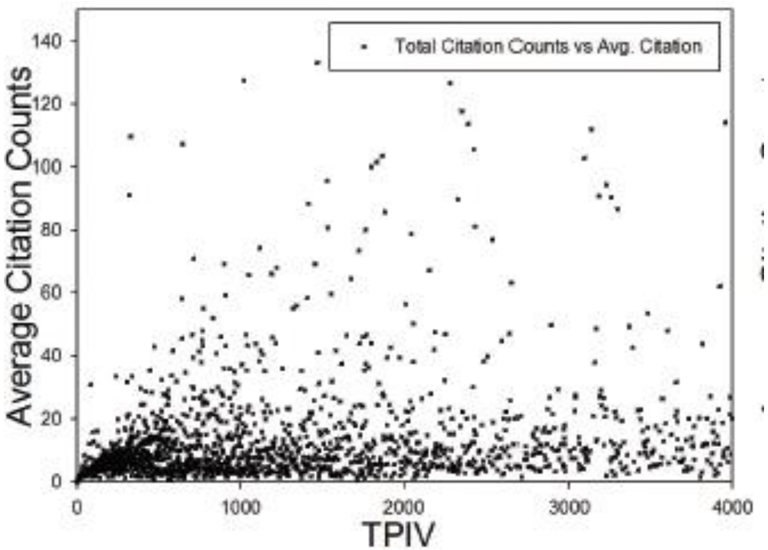
Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Venue Features





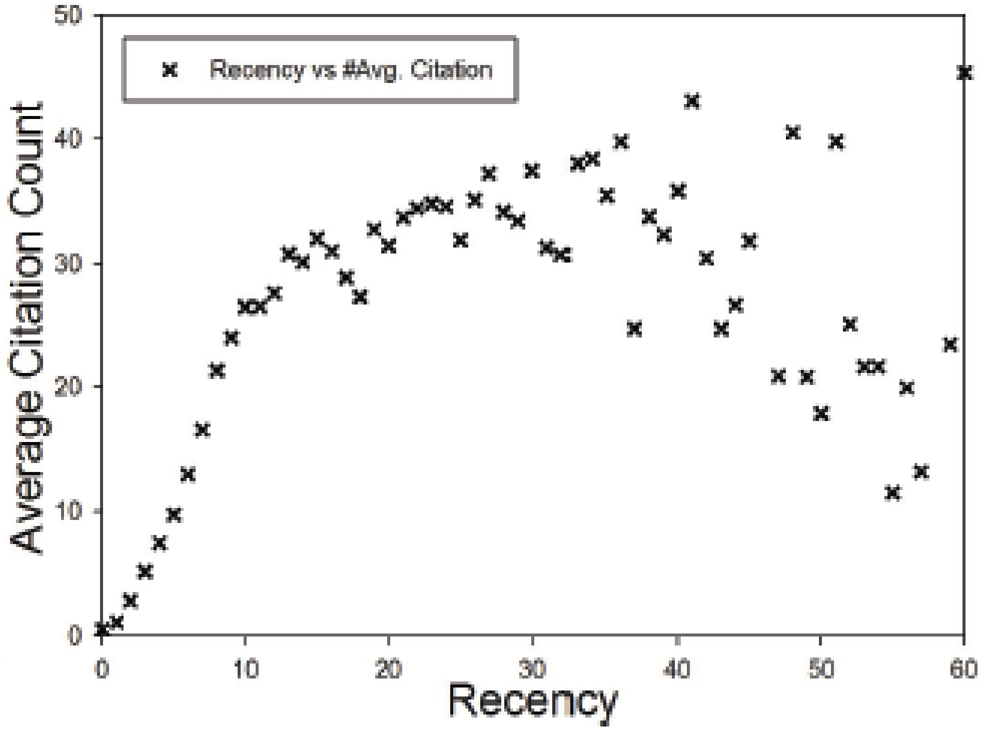




Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Temporal Feature

- Recency: the number of years since the article was published



Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Data Description

- AMiner (https://aminer.org/citation)
  - 1,558,499 papers in CS
  - 916,946 researchers (from 1960-2010)
  - Co-author network (3,063,257 edges)
  - Citation network (20,083,947 edges)

Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Set-up

- Test set: 10,000 papers from year 2009
- For training, only use features available up to year 2008
- Evaluation Metric
  - Coefficient of determination R^2
  - $R^2 = \frac{\sum(\hat{y}-\bar{y})^2}{\sum(y-\bar{y})^2}$

Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Predictive Models

- kNN
- Linear Regression
- Support Vector Regression
- Classification and Regression Tree (CART)
- Gaussian Process Regression (GPR)

Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Performance Comparisons

	1-Year FIP (Δt=1)			5-Year FIP (Δt=5)			10-Year FIP (Δt=10)		
Methods	FData	RData	Combined	FData	RData	Combined	FData	RData	Combined
kNN	0.515	0.311	0.593	0.681	0.268	0.734	0.649	0.161	0.767
LR	0.625	0.479	0.692	0.798	0.134	0.811	0.885	0.123	0.912
SVR	0.590	0.268	0.644	0.723	0.162	0.771	0.813	0.111	0.861
CART	0.679	0.441	**0.713**	0.797	0.203	0.834	0.852	0.128	0.905
GPR	0.601	0.349	0.668	0.823	0.153	**0.869**	0.894	0.130	**0.927**

Accuracy increase as ∇t increases
Non-linear regression achieves better performance
GPR performs the best

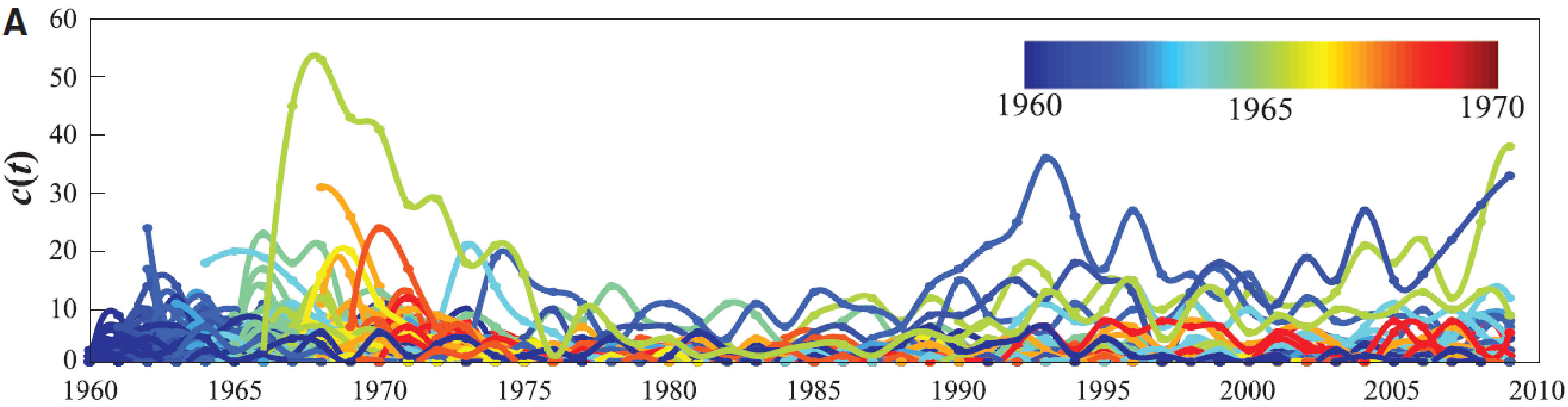
Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Feature Analysis

	FData		RData	
Feature	+Add	−Drop	+Add	−Drop
Novelty	0.059	0.754	0.066	0.751
T.Rank	0.079	0.783	0.135	0.678
Diversity	0.157	0.661		
A.Rank	**0.593**	**0.406**	0.227	0.626
H-Index	0.244	0.611	0.186	0.663
Productivity	0.198	0.652	0.187	0.684
MPIA	0.585	0.419	**0.363**	**0.596**
TPIA	0.048	0.805	0.037	0.811
NOCA	0.056	0.794	0.158	0.643
Sociality	0.249	0.597	0.181	0.632
Authority	0.155	0.668	0.178	0.615
Versatility	0.160	0.649	0.139	0.665
Recency	0.101	0.738		
V.Rank	0.337	0.603	0.225	0.648
V.Centrality	0.049	0.793	0.067	0.776
MPIV	0.329	0.616	0.196	0.667
TPIV	0.023	0.815	0.021	0.823

Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.

Lack of predictability in citation patterns

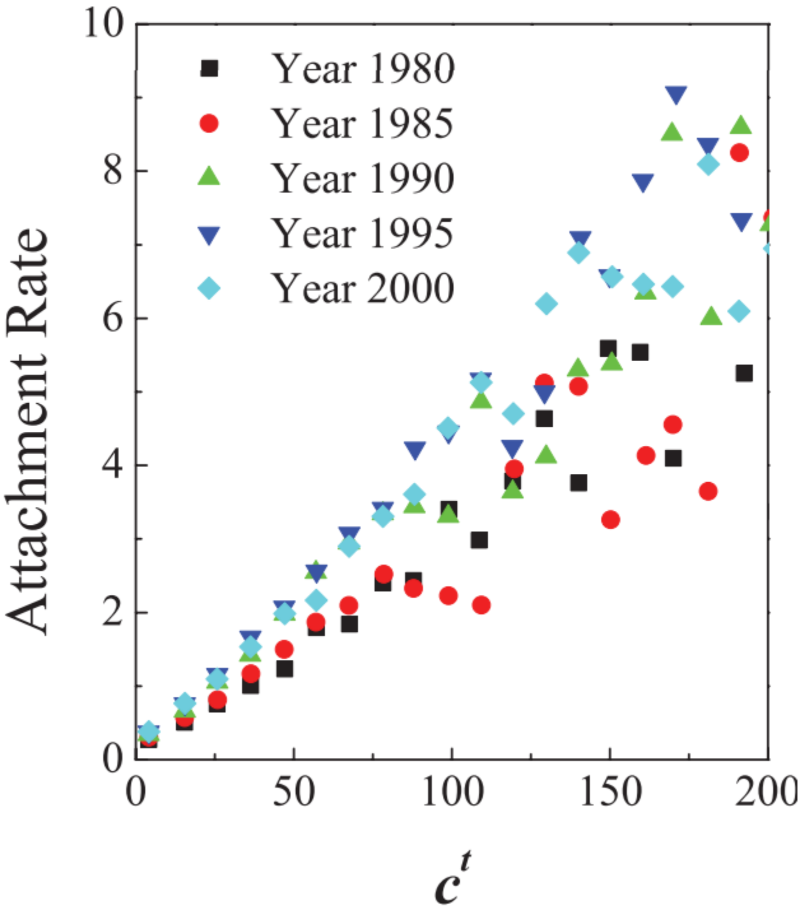


Citation history of 463,348 papers extracted from the *Physical Review* corpus

Wang, Dashun, Chaoming Song, and Albert-László Barabási. "Quantifying long-term scientific impact." Science 342.6154 (2013): 127-132.

Preferential attachment

- Highly cited papers are more likely to be cited again

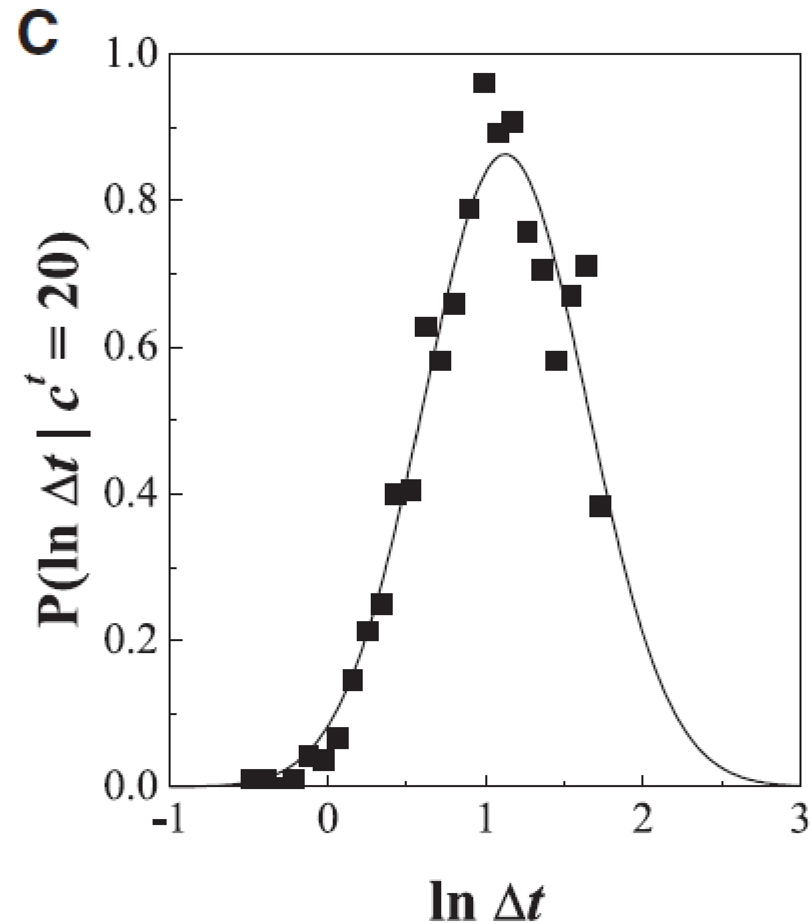


Wang, Dashun, Chaoming Song, and Albert-László Barabási. "Quantifying long-term scientific impact." Science 342.6154 (2013): 127-132.

Temporal Citation Trend

- Long-term decay follows a log-normal survival probability

$$P_i(t) = \frac{1}{\sqrt{2\pi}\sigma_i t} \exp\left[-\frac{(\ln t - \mu_i)^2}{2\sigma_i^2}\right]$$



Wang, Dashun, Chaoming Song, and Albert-László Barabási. "Quantifying long-term scientific impact." Science 342.6154 (2013): 127-132.

Fitness η of a paper

- The paper's importance relative to its peers

Wang, Dashun, Chaoming Song, and Albert-László Barabási. "Quantifying long-term scientific impact." Science 342.6154 (2013): 127-132.

Mechanistic Model

- The probability that paper i is cited at time t after publication is

$$\Pi_i(t) \sim \eta_i c_i^t P_i(t)$$

- Solving for the cumulative number of citations acquired by paper i at time t

$$c_i^t = m\left[e^{\frac{\beta\eta_i}{A}\Phi\left(\frac{\ln t-\mu_i}{\sigma_i}\right)}-1\right] \equiv m\left[e^{\lambda_i\Phi\left(\frac{\ln t-\mu_i}{\sigma_i}\right)}-1\right] \qquad (3)$$

$$\tilde{c} = \Phi(\tilde{t})$$

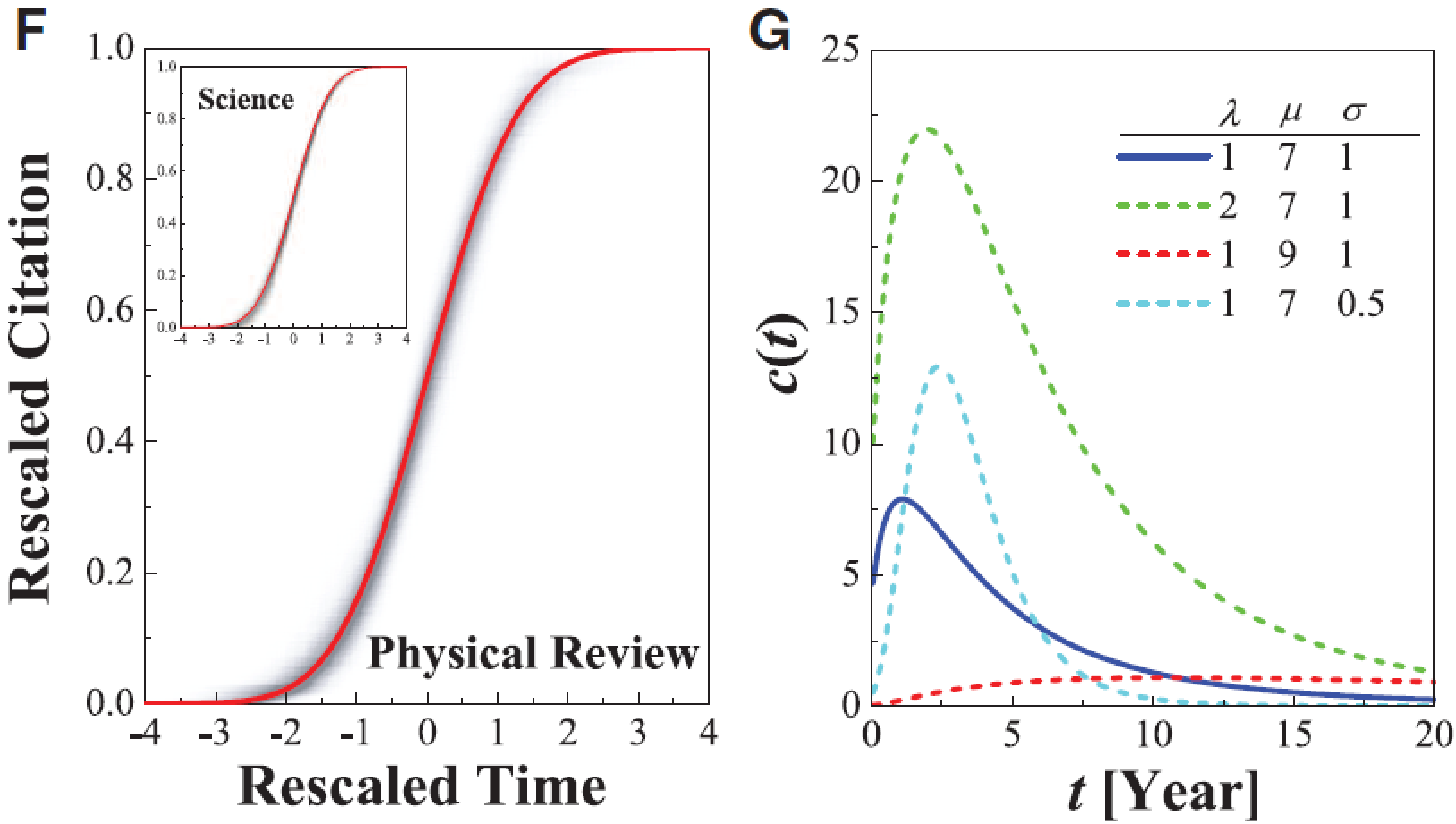
$$\tilde{c} = \frac{\ln\left(1+\frac{c_i^t}{m}\right)}{\lambda_i}$$

$$\tilde{t} = (\ln t - \mu_i)/\sigma\_i$$

where

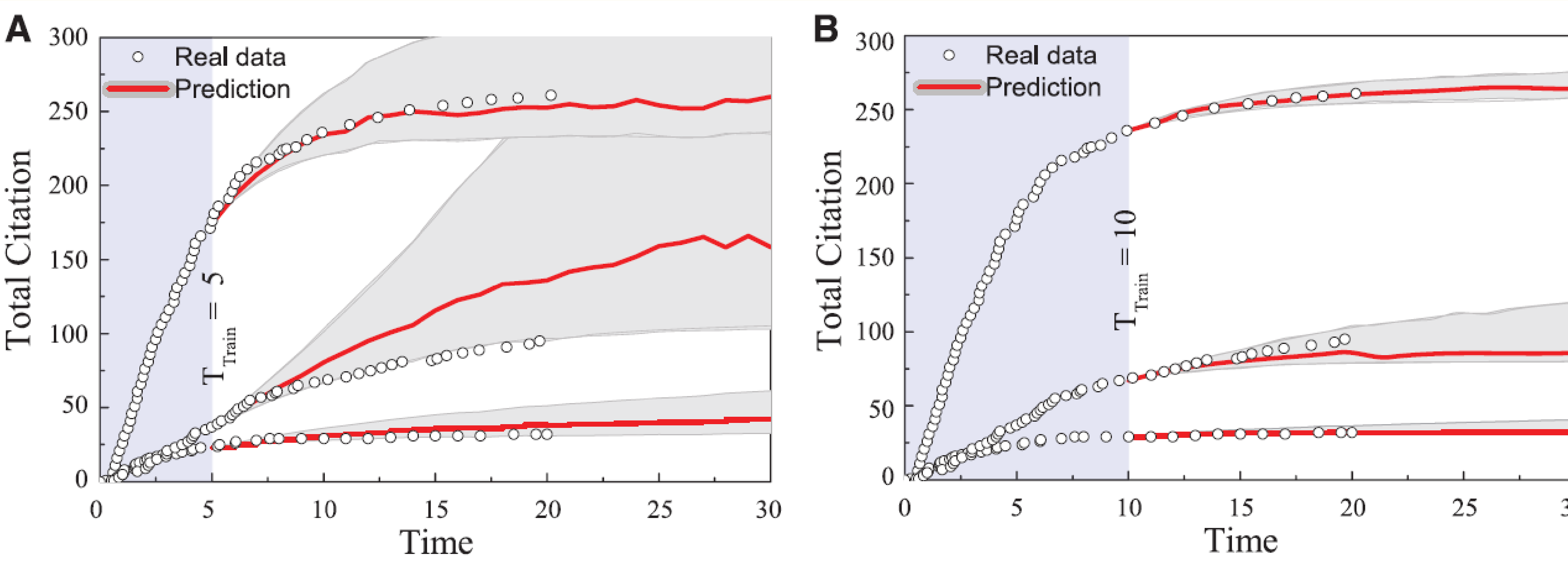
$$\Phi(x) \equiv (2\pi)^{-1/2}\int_{-\infty}^{x} e^{-y^2/2}\,dy \qquad (4)$$

Model's validity



Wang, Dashun, Chaoming Song, and Albert-László Barabási. "Quantifying long-term scientific impact." Science 342.6154 (2013): 127-132.

Predicting future impact



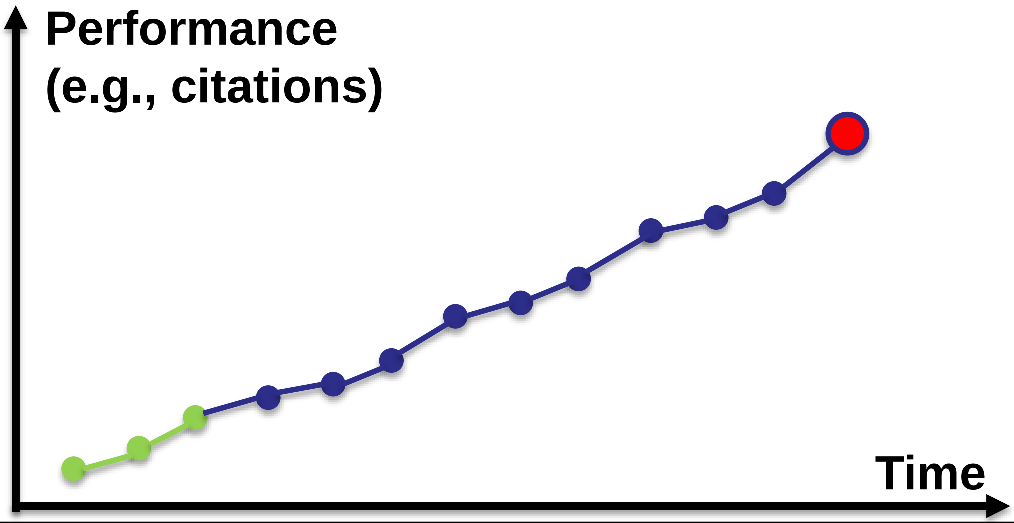
Measure the fraction of papers that fall within the envelope for all PR papers published in 1960

With T-train=5, 6.5% left the envelope 30 years later

Wang, Dashun, Chaoming Song, and Albert-László Barabási. "Quantifying long-term scientific impact." Science 342.6154 (2013): 127-132.

Performance Prediction: Setup

- **Given:** Initial Performance of a team
- **Predict:**
  - (1) Long-Term Performance [KDD15]
  - (2) Performance Trajectory [SDM16]



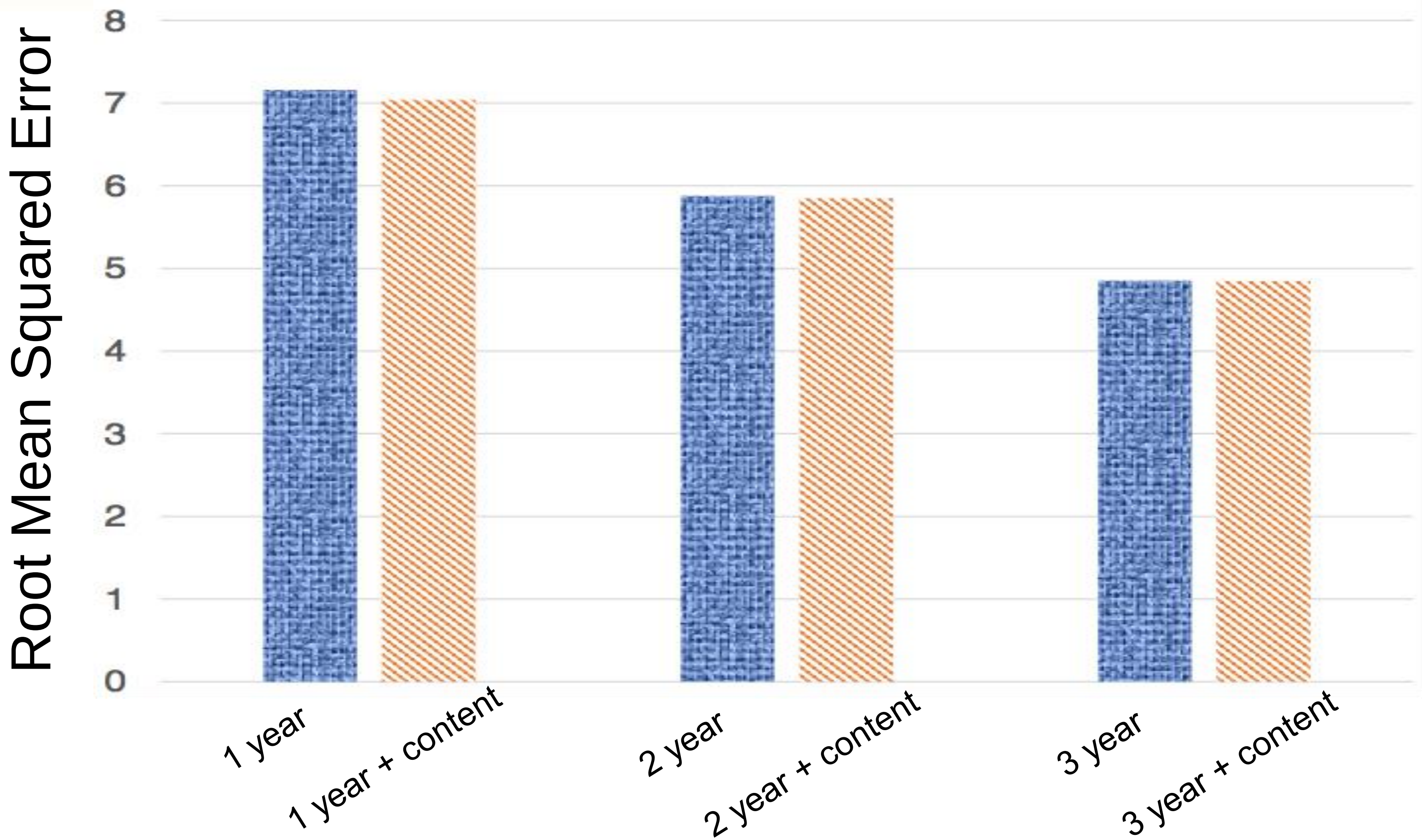
- L. Li, and H. Tong: The Child is Father of the Man: Foresee the Success at the Early Stage. KDD 2015: 655-664
- L. Li, *H. Tong*, J. Tang and W. Fan: “iPath: Forecasting the Pathway to Impact”. SDM 2016

Performance Prediction: Challenges

- C1: Scholarly feature design
- C2: Non-linearity
- C3: Domain heterogeneity
- C4: Dynamics

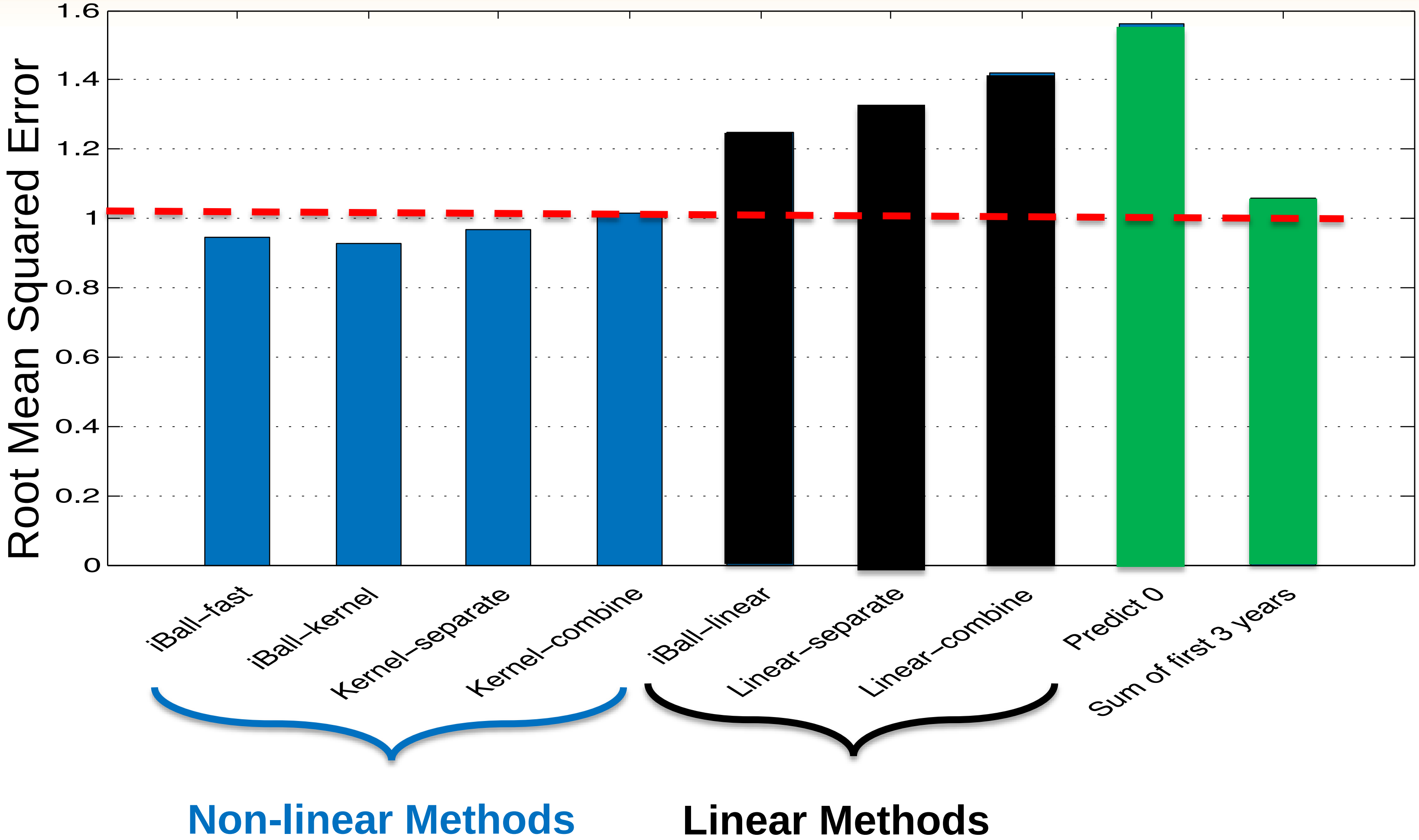
- L. Li, and H. Tong: The Child is Father of the Man: Foresee the Success at the Early Stage. KDD 2015: 655-664

C1: Scholarly Feature Design



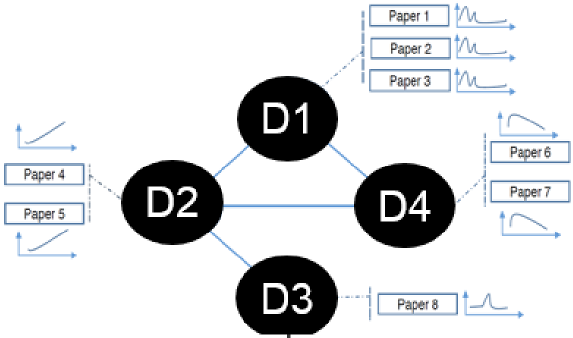
Obs.: Adding content features brings little improvement

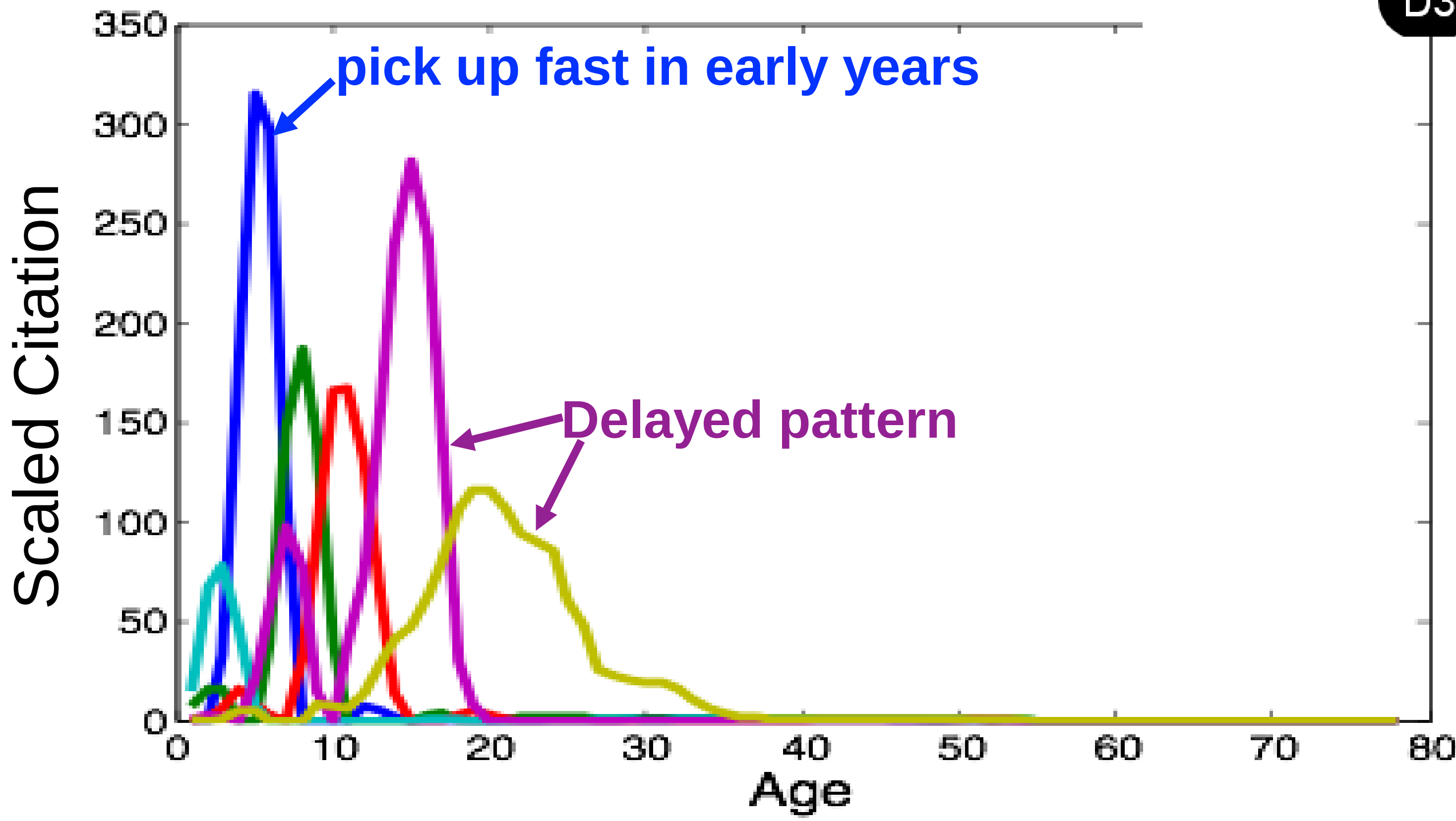
C2: Non-linearity



Obs.: Non-linear methods > linear ones

C3: Domain heterogeneity

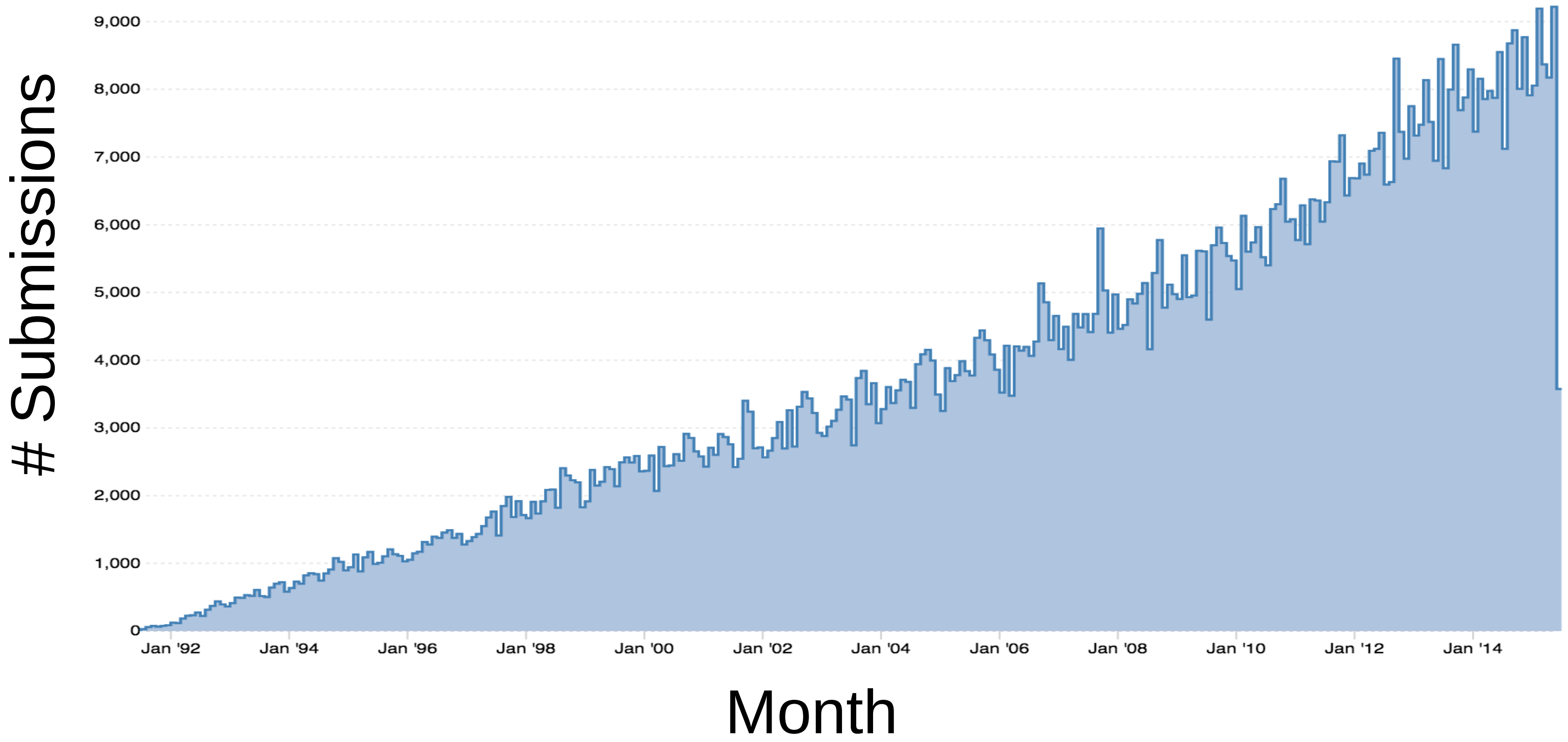




Obs.: Impact of scientific work from different domains behaves differently

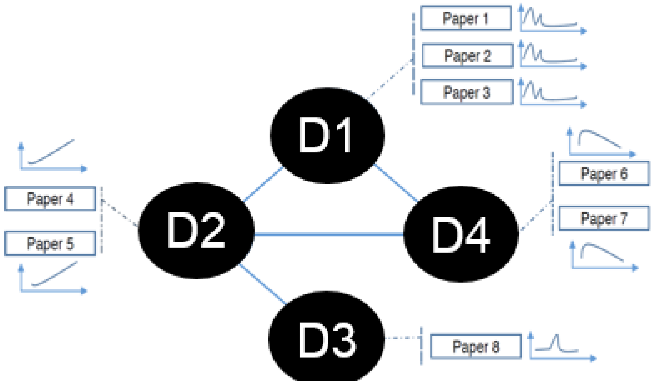
C4: Dynamics

arXiv monthly submission rates



Q: How to quickly update the predictive model?

iBall — Formulations



- Optimization Formulation

$$\min_{\mathbf{w}^{(\mathbf{i})}, i=1,\ldots,n_d} \overbrace{\sum_{i=1}^{n_d} \mathcal{L}[f(\mathbf{X}^{(\mathbf{i})}, \mathbf{w}^{(\mathbf{i})}), \mathbf{Y}^{(\mathbf{i})}] + \lambda \sum_{i=1}^{n_d} \Omega(\mathbf{w}^{(\mathbf{i})})}^{\text{Within-Domain Model}}$$

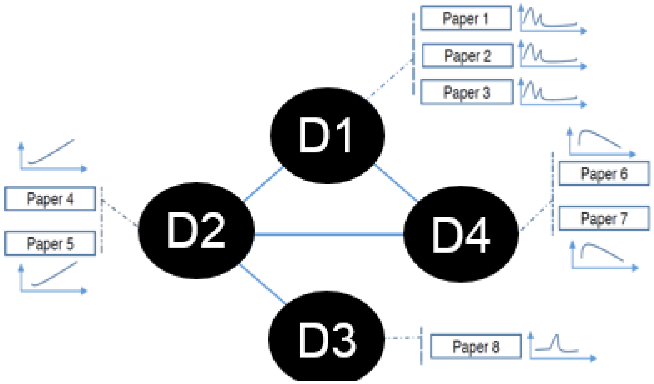
$$+\theta \sum_{i=1}^{n_d} \sum_{j=1}^{n_d} \underbrace{\mathbf{A}_{\mathbf{ij}} g(\mathbf{w}^{(\mathbf{i})}, \mathbf{w}^{(\mathbf{j})})}_{\text{Cross-Domain Consistency}}$$

- Remarks
  - **Within-Domain Model**: regression/classification, linear/non-linear
  - **Cross-Domain Consistency**: similar domains have similar models

Question: how to instantiate such consistency?

- L. Li, and H. Tong: The Child is Father of the Man: Foresee the Success at the Early Stage. KDD 2015: 655-664

iBall — linear formulation



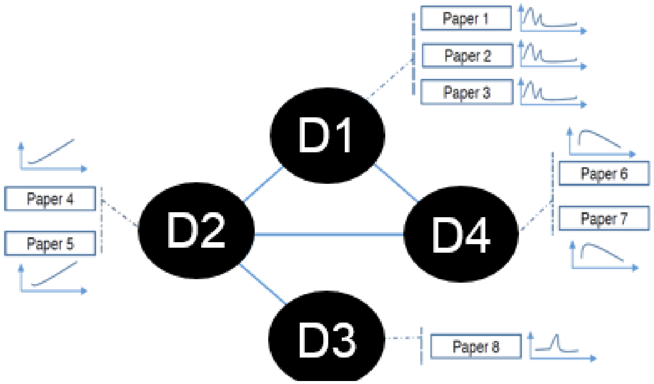
Details:

$$\min_{\mathbf{w}^{(i)}, i=1,\ldots,n_d} \sum_{i=1}^{n_d} \|\mathbf{X}^{(i)}\mathbf{w}^{(i)} - \mathbf{Y}^{(i)}\|_2^2 + \lambda \sum_{i=1}^{n_d} \|\mathbf{w}^{(i)}\|_2^2$$
$$+\theta \sum_{i=1}^{n_d} \sum_{j=1}^{n_d} \underbrace{\mathbf{A}_{\mathbf{ij}} \|\mathbf{w}^{(i)} - \mathbf{w}^{(j)}\|_2^2}$$

Intuitions: similar domains (large $\mathbf{A}_{ij}$)

→ same feature has similar effect (small $\|\mathbf{w}^{(i)} - \mathbf{w}^{(j)}\|_2^2$)

iBall — non-linear formulation



Details:

$$\min_{\mathbf{w}^{(\mathbf{i})}, i=1,\ldots,n_d} \sum_{i=1}^{n_d} \|\mathbf{K}^{(\mathbf{i})}\mathbf{w}^{(\mathbf{i})} - \mathbf{Y}^{(\mathbf{i})}\|_2^2 + \lambda \sum_{i=1}^{n_d} \mathbf{w}^{(\mathbf{i})\prime}\mathbf{K}^{(\mathbf{i})}\mathbf{w}^{(\mathbf{i})}$$
$$+\theta \sum_{i=1}^{n_d}\sum_{j=1}^{n_d} \mathbf{A}_{\mathbf{ij}} \|\underbrace{\mathbf{K}^{(\mathbf{i})}\mathbf{w}^{(\mathbf{i})}} - \underbrace{\mathbf{K}^{(\mathbf{ij})}\mathbf{w}^{(\mathbf{j})}}\|_2^2$$

Predicted output (domain *i* → domain *i*)

Predicted output (domain *j* → domain *i*)

Intuitions: similar domains (large $\mathbf{A}_{ij}$)

→ similar predicted outputs (small $\|\mathbf{K}^{(\mathbf{i})}\mathbf{w}^{(\mathbf{i})} - \mathbf{K}^{(\mathbf{ij})}\mathbf{w}^{(\mathbf{j})}\|_2^2$)

iBall — Closed-form Solutions

- Closed-form Solution

$$\mathbf{w} = \mathbf{S}^{-1}\mathbf{Y}$$

  - iBall — linear:

$$\mathbf{w} = [\mathbf{w}^{(1)}; \ldots ; \mathbf{w}^{(k)}] \quad \mathbf{Y} = [\mathbf{X}^{(1)\prime}\mathbf{Y}^{(1)}; \ldots ; \mathbf{X}^{(k)\prime}\mathbf{Y}^{(k)}]$$

$$\begin{array}{cccc} & & \text{i-th block column} & \text{j-th block column} & \\ \mathbf{S} = & \left[\begin{array}{c} \cdots \\ \cdots \\ \cdots \end{array}\right. & \begin{array}{c} \cdots \\ \mathbf{X}^{(\mathbf{i})\prime}\mathbf{X}^{(\mathbf{i})} + (\theta \sum_{j=1}^{k} \mathbf{A}_{\mathbf{ij}} + \lambda)\mathbf{I} \\ \cdots \end{array} & \left.\begin{array}{c} \cdots \\ -\theta \mathbf{A}_{\mathbf{ij}}\mathbf{I} \\ \cdots \end{array}\right] & \begin{array}{l} \text{i-th block} \\ \text{row} \end{array} \end{array}$$

Time Complexity: $O((dk)^3)$

d: # of features; k: # of domains
(*dk:* in the order of 10 or 100)

iBall — Closed-form Solutions

- Closed-form Solution

$$\mathbf{w} = \mathbf{S}^{-1}\mathbf{Y}$$

  - iBall — non-linear:

$$\mathbf{w} = [\mathbf{w}^{(1)}; \ldots; \mathbf{w}^{(k)}] \qquad \mathbf{Y} = [\mathbf{Y}^{(1)}; \ldots; \mathbf{Y}^{(k)}]$$

$$\begin{array}{cccc} & \text{i-th block column} & \text{j-th block column} & \\ \mathbf{S} = \left[\begin{array}{c} \cdots \\ \cdots \\ \cdots \end{array}\right. & \begin{array}{c} \cdots \\ (1+\theta\sum_{j=1}^{k}\mathbf{A_{ij}})\mathbf{K^{(i)}} + \lambda\mathbf{I} \\ \cdots \end{array} & \left.\begin{array}{c} \cdots \\ -\theta\mathbf{A_{ij}}\mathbf{K^{(ij)}} \\ \cdots \end{array}\right] & \text{i-th block row} \end{array}$$

Time Complexity: $O(n^3)$

n: total # of training examples (in the order of millions)

iBall — Scale-up with Dynamic Update

- **Key idea #1:** Approx **S** by low-rank approx
- **Details:**

$$\mathbf{S}_{t+1} \approx \mathbf{U}_{t+1}\mathbf{\Lambda}_{t+1}\mathbf{U}'_{t+1} \rightarrow \mathbf{w}_{t+1} \begin{aligned} &= \mathbf{S}^{-1}_{t+1}\mathbf{Y}_{t+1} \\ &= \mathbf{U}_{t+1}\mathbf{\Lambda}^{-1}_{t+1}\mathbf{U}'_{t+1}\mathbf{Y}_{t+1} \end{aligned}$$

(**Overall:** $O(n^2r)$)　　(**Overall:** $O(nr)$)

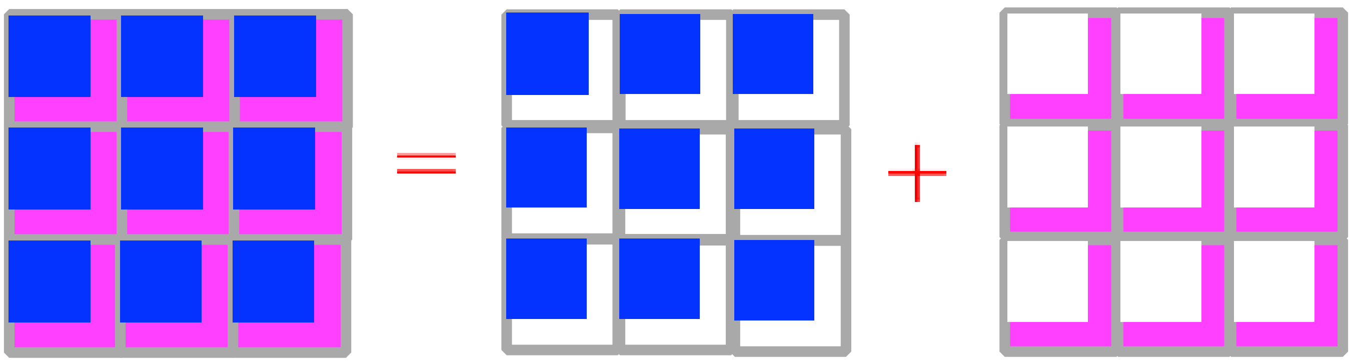
- **Complexity:** $O(n^3) \rightarrow O(n^2r + nr)$
- **Benefit:** avoid matrix inverse

Question: how to avoid re-computing low-rank approx at each time step?

iBall — Scale-up with Dynamic Update

- **Key idea #2:** Incrementally update the low rank structure of **S**
- **Details:**

white: zeros
blue: old at *t*
pink: new at *t+1*



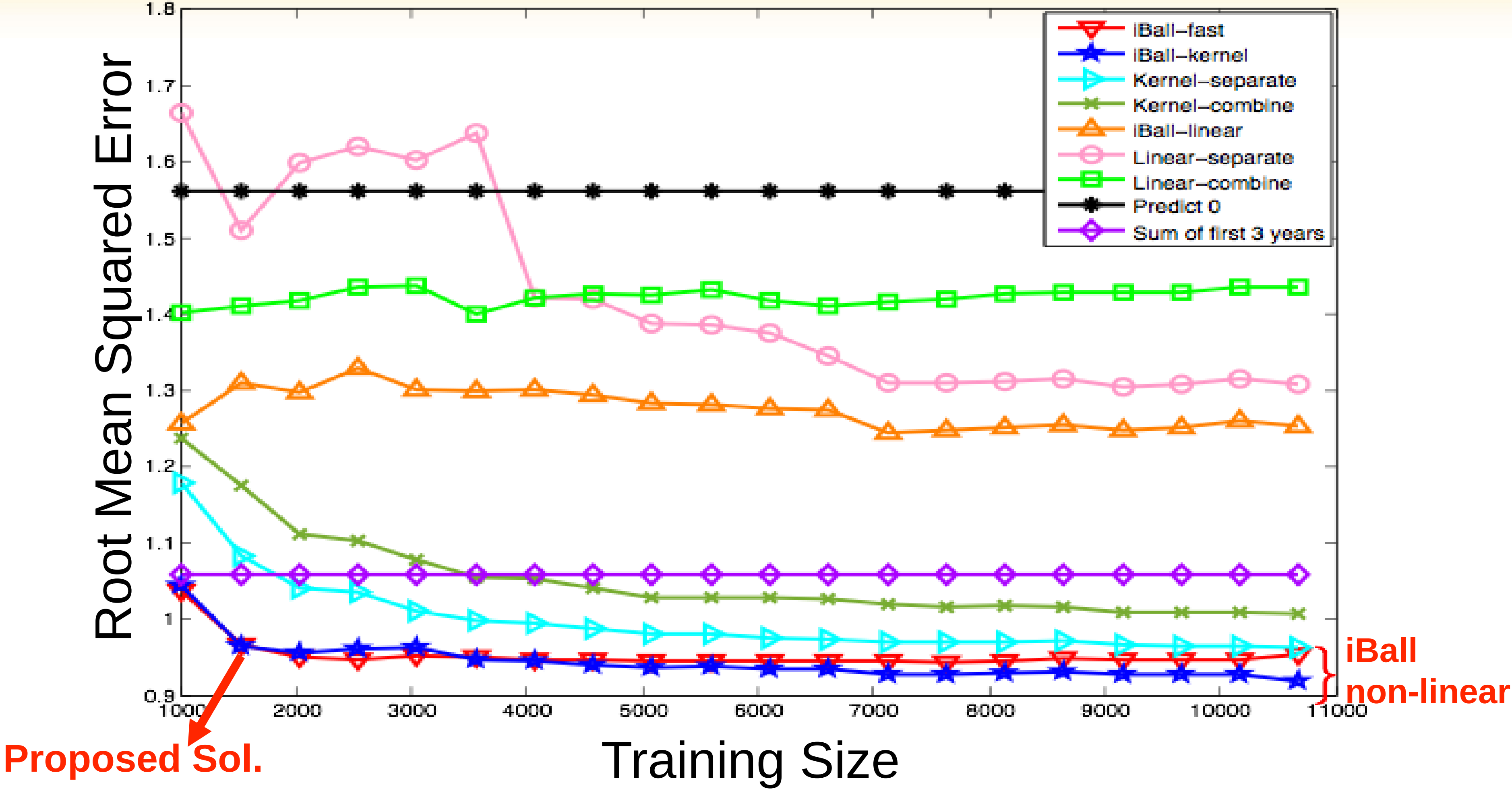
$\mathbf{S}_{t+1} = \tilde{\mathbf{S}}_t + \Delta\mathbf{S}$

(low rank, sparse)

- **Complexity:** $O(n^2 r) \rightarrow O((n+m)(r^2 + r'^2)), r \ll n$
- **Benefit:** avoid re-computing low-rank approx

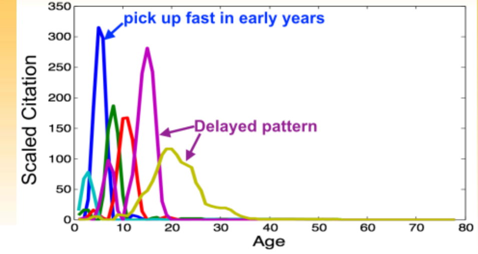
- L. Li, H. Tong, Y. Xiao, W. Fan. Cheetah: Fast Graph Kernel Tracking on Dynamic Graphs. SDM 2015.

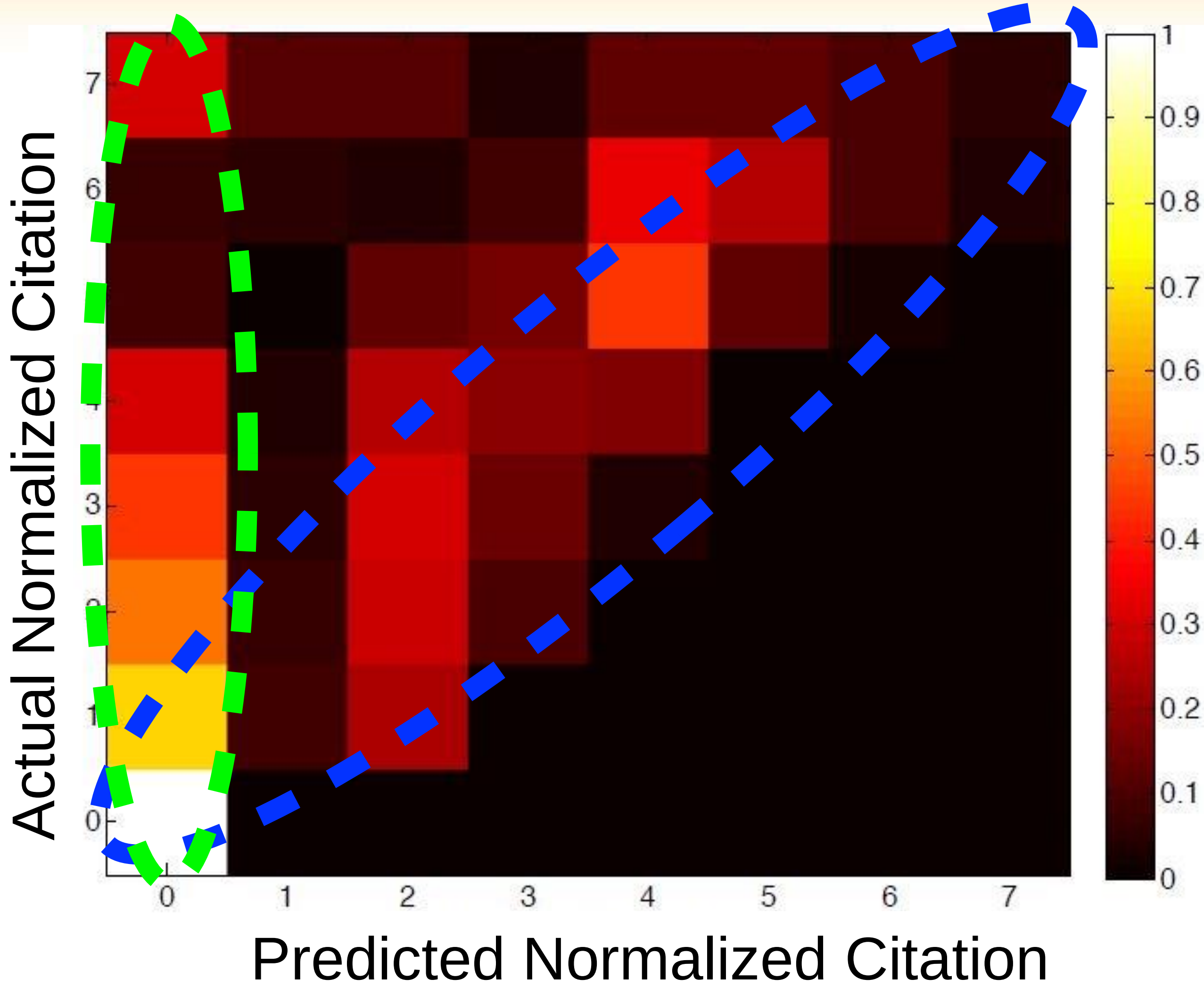
Paper Citation Prediction Performance



Datasets: AMiner (2,243,976 papers, 1,274,360 authors, 8,882 venues)

Error Analysis

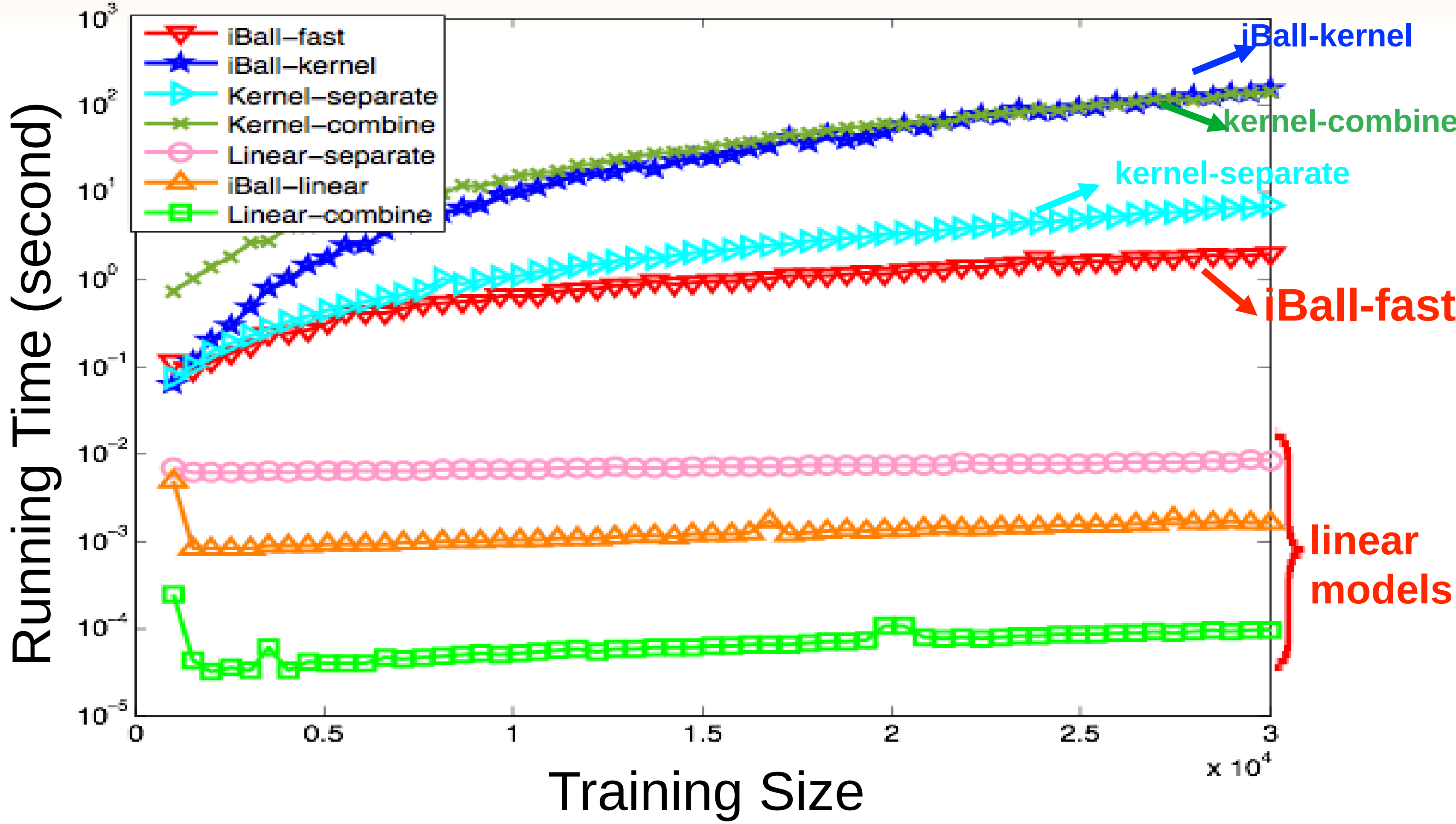




Obs.: bright region at x = y

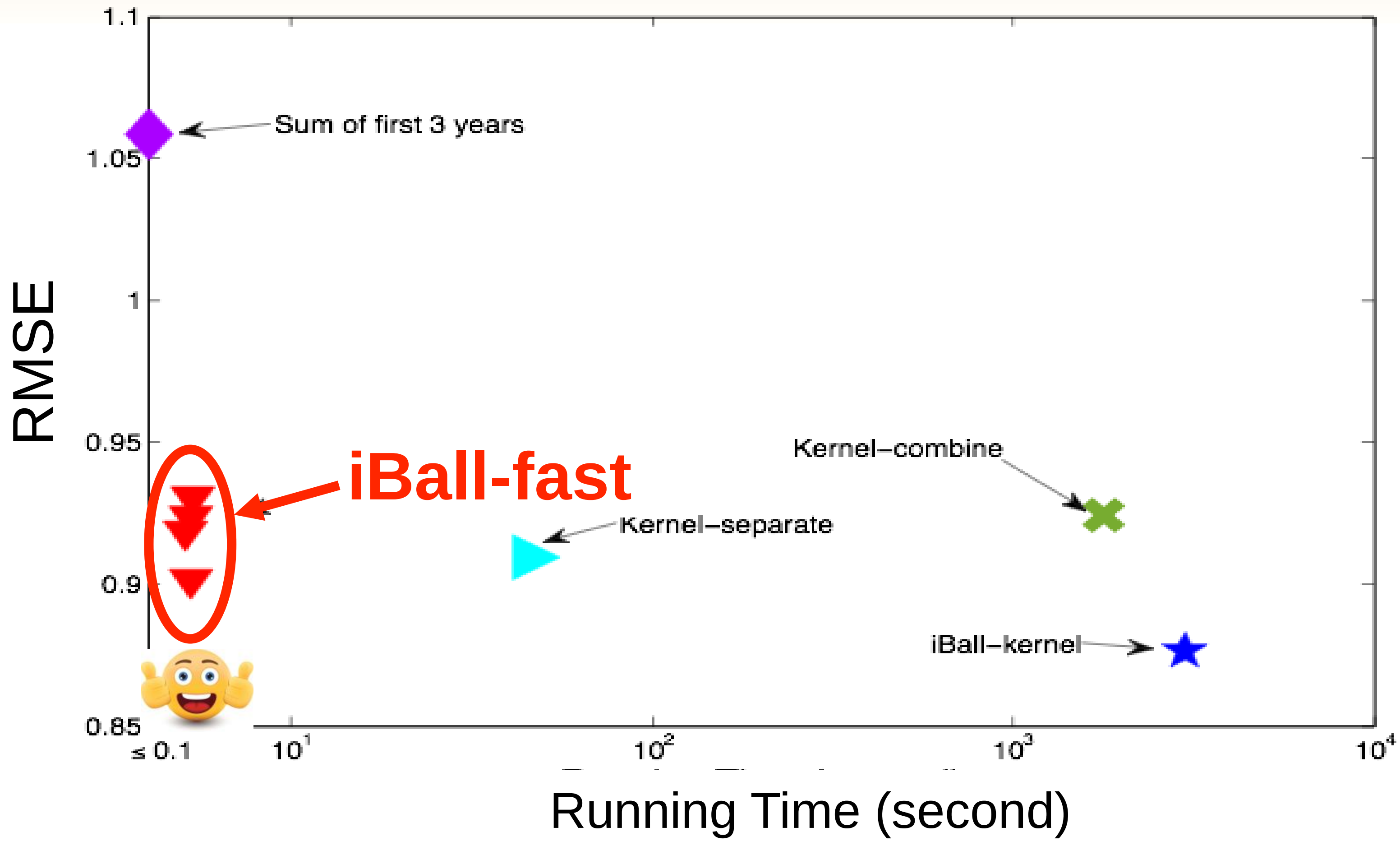
- L. Li, and H. Tong: The Child is Father of the Man: Foresee the Success at the Early Stage. KDD 2015

Running Time Comparison



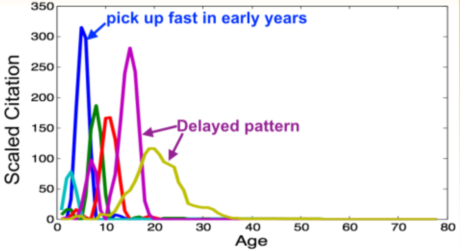
Obs.: iBall-fast outperforms other non-linear methods

Quality vs. Speed



Obs.: iBall-fast: good trade-off between quality and speed

iBall: Summary



- **Goal**: predict long-term impact of scholarly entities
- **Solutions**: joint predictive model (**iBall**)

Challenges	C1 *feature design*	C2 *non-linearity*	C3 *domain-heterogeneity*	C4 *dynamics*
Tactics	*first 3 years' citation*	*kernel trick*	*domain consistency*	*low-rank approximation*

- **Results:**
  - iBall joint models better than separate versions
  - iBall-fast updates efficiently and accurately

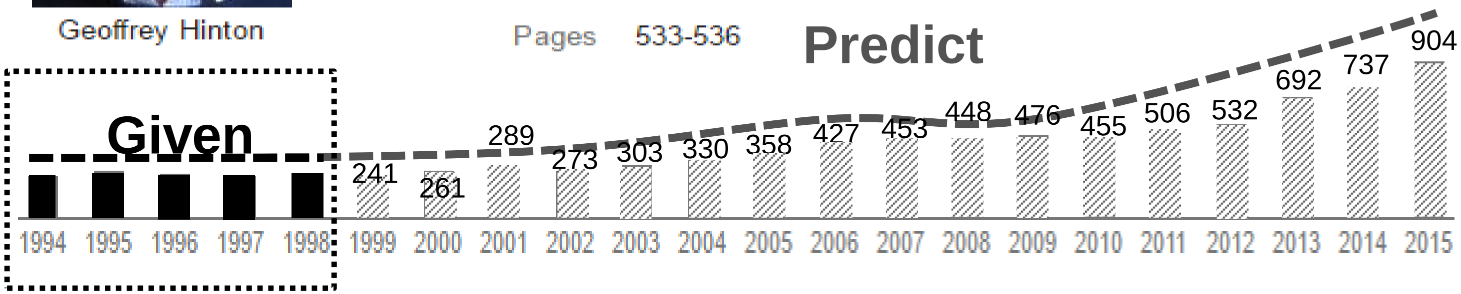
• L. Li, and H. Tong: The Child is Father of the Man: Foresee the Success at the Early Stage. KDD 2015: 655-664

Foresee the Pathway to Impact

Geoffrey Hinton

Learning representations by back-propagating errors

Authors	David E Rumelhart, Geoffrey E Hinton, Ronald J Williams
Publication date	1986
Journal	Nature
Volume	323
Pages	533-536



Implications of forecasting the pathway to impact

- Tracking research frontier
- Invoking early intervention

Question: how to foresee the impact pathway at the early stage?

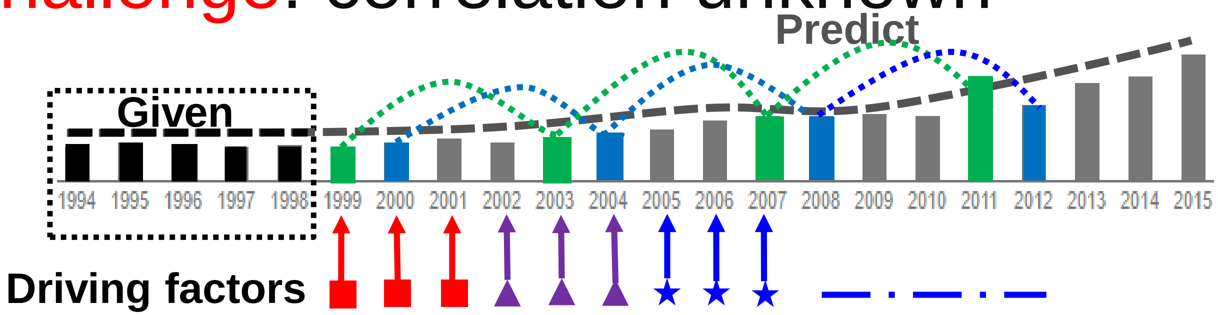
Modeling Scientific Impact

- **Effective scholarly feature design** [Yan+CIKM11]
- **Mechanistic model for the citation dynamics of individual papers** [Wang+Science13]
- ***iBall*- Joint Predictive Model for long-term impact prediction** [Li+KDD15]

All for Point Prediction

Challenges

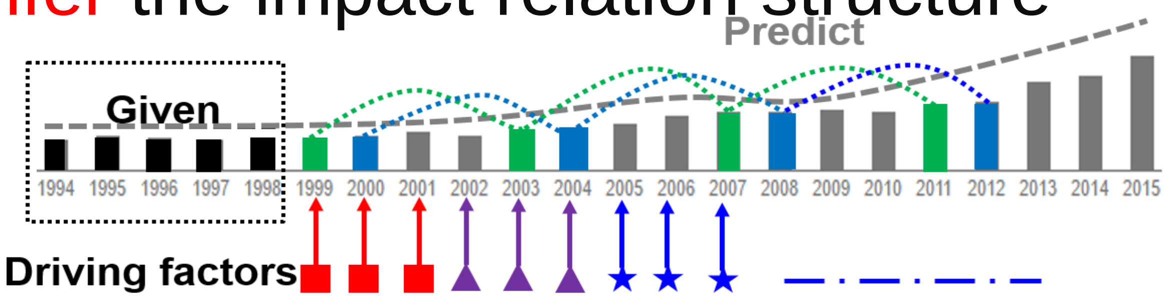
- C1: Output Space -- Correlation
  - Possible solution: multi-label/task learning
  - Challenge: correlation unknown



- C2: Parameter Space -- Smoothness
  - Possible solution: linear dynamic system
  - Challenge: transition process unknown

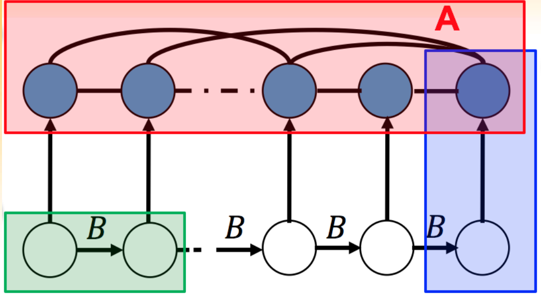
Design Objectives

- D1: Prediction Consistency (for C1)
  - Exploit the correlation in output space
  - Infer the impact relation structure



- D2: Parameter Smoothness (for C2)
  - Apply linear transition to adjacent parameters
  - Learn the linear transition process

iPath -- Formulations



- Optimization Formulations

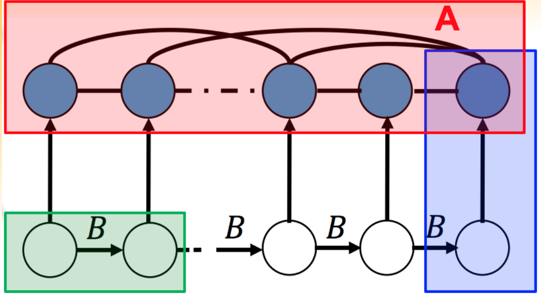
$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \overbrace{\left\{\mathcal{L}[f(\mathbf{X},\mathbf{W}),\mathbf{Y}]\right\}}^{\text{Empirical loss}} + \overbrace{\left\{\alpha\sum_{i=1}^{l}\sum_{j=1}^{l}\mathbf{A}_{ij}\, g(\mathbf{w}_i,\mathbf{w}_j)\right\}}^{\text{Prediction Consistency}}$$

$$+\left\{\beta\sum_{t=2}^{l}\|\mathbf{w}_t-\mathbf{B}\mathbf{w}_{t-1}\|_2^2\right\} \text{ Parameter Smoothness}$$

$$+\,\gamma\underbrace{\|\mathbf{B}-\mathbf{I}\|_F^2+\delta\sum_{i=1}^{l}\Omega(\mathbf{w}_i)+\epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2}_{\text{Regularizations}}$$

- Remarks
  - **Prediction Consistency**: similar impacts have similar models
  - **Parameter Smoothness**: model parameters at adjacent time steps have linear transformation

iPath – linear formulation



- **Details**:

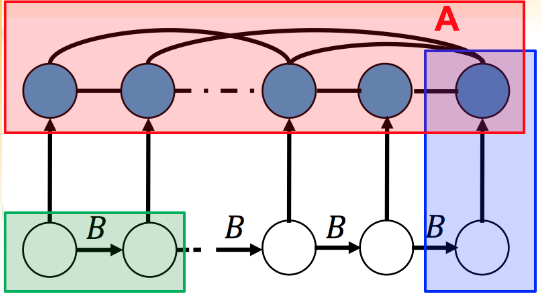
$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \|\mathbf{XW}-\mathbf{Y}\|_F^2 + \alpha\sum_{i=1}^{l}\sum_{j=1}^{l}\boxed{\mathbf{A}_{ij}\|\mathbf{X}\mathbf{w}_i-\mathbf{X}\mathbf{w}_j\|_2^2}$$
$$+\beta\sum_{t=2}^{l}\|\mathbf{w}_t-\mathbf{B}\mathbf{w}_{t-1}\|_2^2+\gamma\|\mathbf{B}-\mathbf{I}\|_F^2$$
$$+\delta\sum_{i=1}^{l}\|\mathbf{w}_i\|_2^2+\epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2$$

- **Intuition**:

Similar impacts (large $\mathbf{A}_{ij}$)

→ Similar Predictions (small $\|\mathbf{X}\mathbf{w}_i-\mathbf{X}\mathbf{w}_j\|_2^2$)

iPath – non-linear formulation



- **Details**:

$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \quad \|\mathbf{KW}-\mathbf{Y}\|_F^2 + \alpha \sum_{i=1}^{l}\sum_{j=1}^{l} \boxed{\mathbf{A}_{ij}\|\mathbf{Kw}_i - \mathbf{Kw}_j\|_2^2}$$

$$+\beta \sum_{t=2}^{l} \|\mathbf{w}_t - \mathbf{Bw}_{t-1}\|_2^2 + \gamma\|\mathbf{B}-\mathbf{I}\|_F^2$$

$$+\delta \sum_{i=1}^{l} \mathbf{w}_i'\mathbf{Kw}_i + \epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2$$

- **Intuition**:

Similar Impacts (large $\mathbf{A}_{ij}$)

➡ Similar Predictions (small $\|\mathbf{Kw}_i - \mathbf{Kw}_j\|_2^2$)

iPath – Optimization Solutions

- Alternating Optimization Strategy

Optimize for W

$$\min_{\mathbf{W}} \; \|\mathbf{X}\mathbf{W} - \mathbf{Y}\|_F^2 + \alpha \sum_{i=1}^{l} \sum_{j=1}^{l} \mathbf{A}_{ij} \|\mathbf{X}\mathbf{w}_i - \mathbf{X}\mathbf{w}_j\|_2^2 + \beta \sum_{t=2}^{l} \|\mathbf{w}_t - \mathbf{B}\mathbf{w}_{t-1}\|_2^2 + \delta \sum_{i=1}^{l} \|\mathbf{w}_i\|_2^2$$

Optimize for B

$$\min_{\mathbf{B}} \; \beta \sum_{t=2}^{l} \|\mathbf{w}_t - \mathbf{B}\mathbf{w}_{t-1}\|_2^2 + \gamma \|\mathbf{B} - \mathbf{I}\|_F^2$$

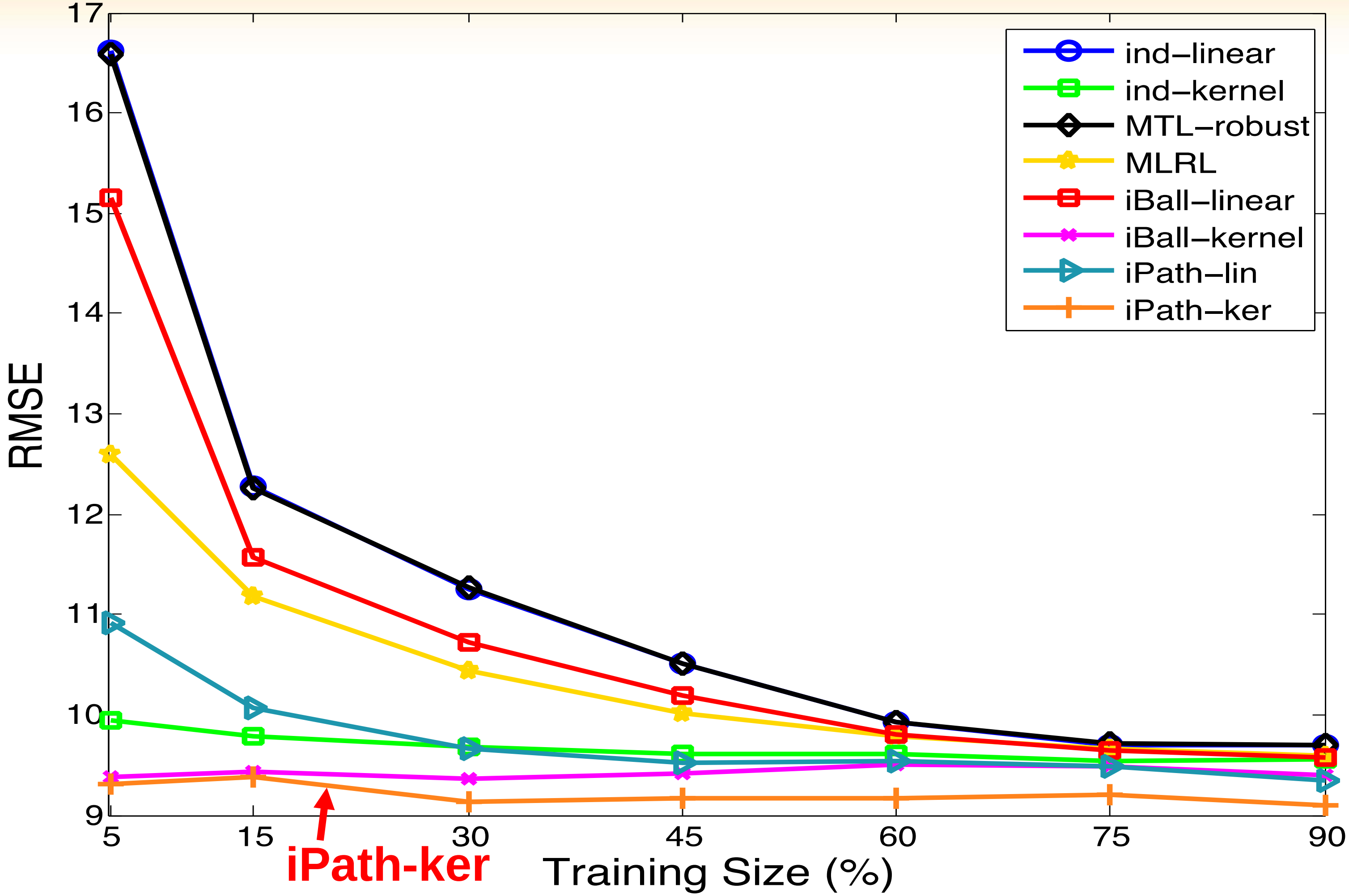
Optimize for A

$$\min_{\mathbf{A}} \; \alpha \sum_{i=1}^{l} \sum_{j=1}^{l} \mathbf{A}_{ij} \|\mathbf{X}\mathbf{w}_i - \mathbf{X}\mathbf{w}_j\|_2^2 + \epsilon \|\mathbf{A} - \mathbf{A}_0\|_F^2$$

Experiment Setup

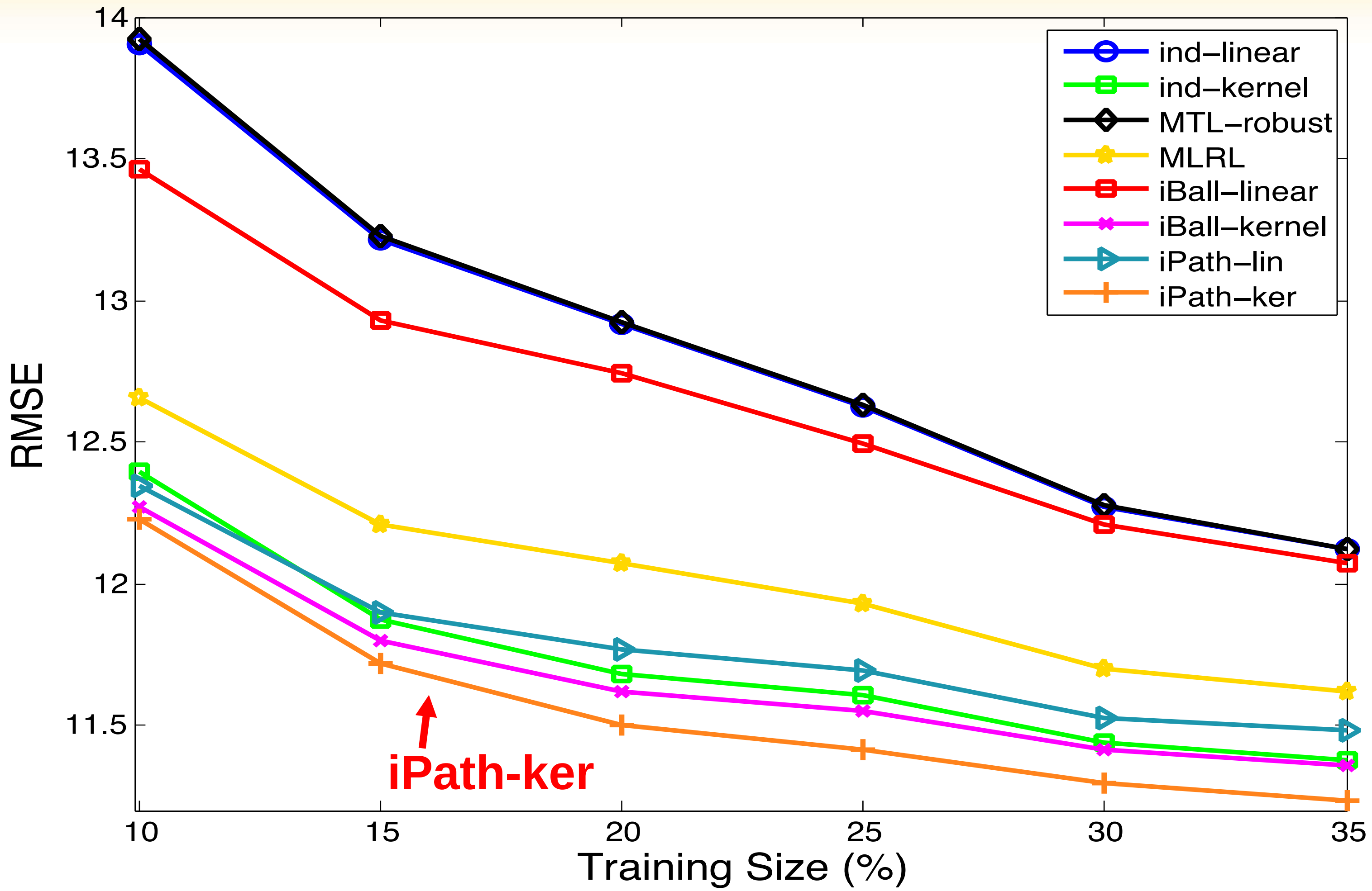
- **Datasets**: *AMiner* (2,243,976 papers, 1,274,360 authors, 8,882 venues)
- **Task**: Observing the first 5 years' citations, predict yearly citations from year 6 – 15
- **Evaluation Metric**: Root Mean Squared Error (RMSE)

Paper Impact Pathway Prediction



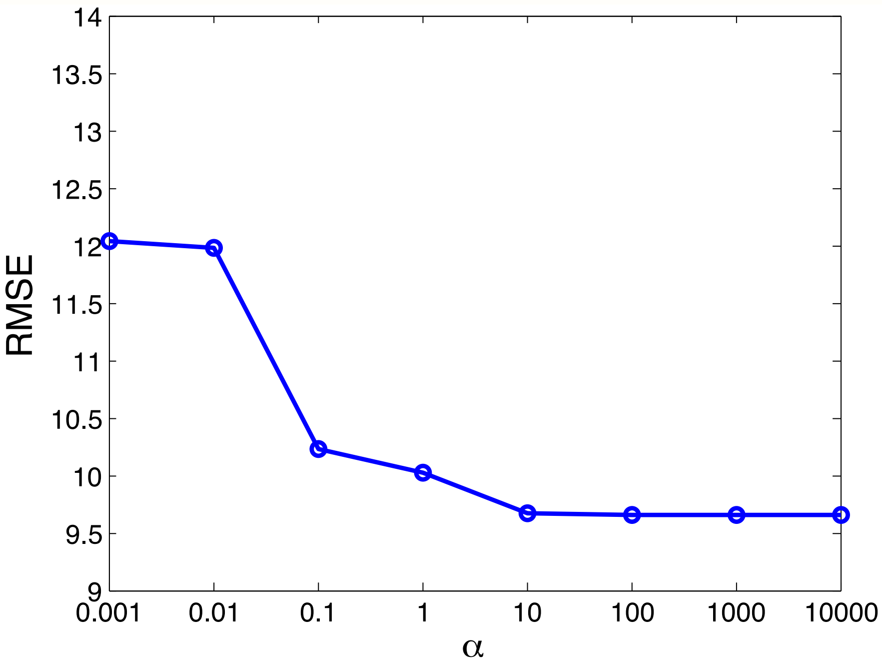
Obs: iPath-ker performs the best among all the competitors

Author Impact Pathway Prediction

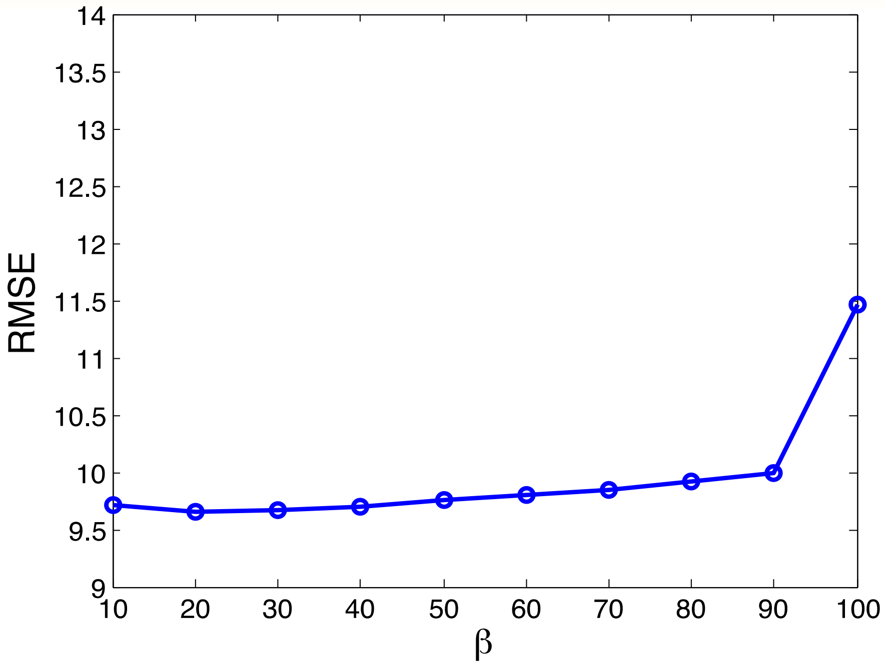


Obs: iPath-ker performs the best among all the competitors

Sensitivity Analysis



RMSE vs. α



RMSE vs. β

Obs: iPath is stable in a large range of parameter spaces

Performance Gain Analysis

RMSE	Paper Impact	Author Impact
1	9.602	11.608
1 + 2	9.507	11.548
1 + 2 + 3	9.335	11.489
1 + 2 + 3 + 4	9.171	11.391

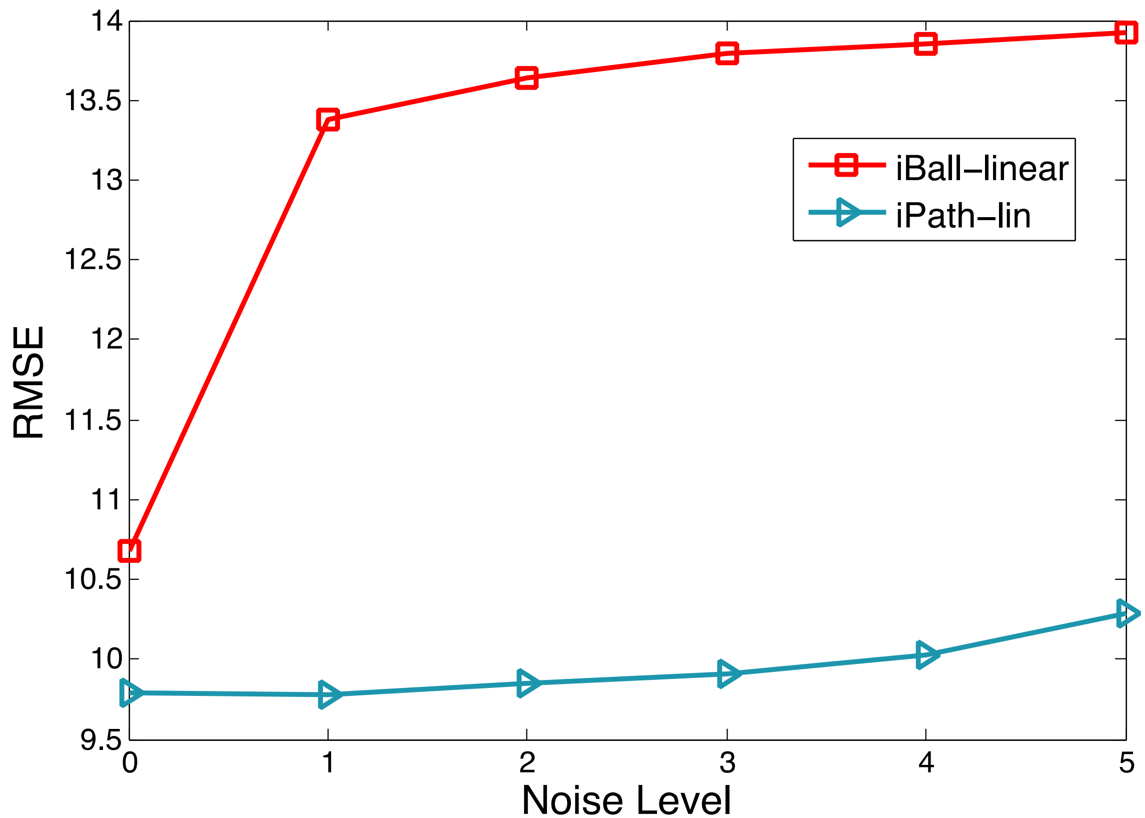
1 Basic form, 2 relation, 3 transition, 4 inferring

$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \left(\|\mathbf{KW} - \mathbf{Y}\|_F^2 \right) + \left(\alpha \sum_{i=1}^{l} \sum_{j=1}^{l} \mathbf{A}_{ij} \|\mathbf{Kw}_i - \mathbf{Kw}_j\|_2^2 \right) + \left\{ \beta \sum_{t=2}^{l} \|\mathbf{w}_t - \mathbf{Bw}_{t-1}\|_2^2 \right\} + \gamma \|\mathbf{B} - \mathbf{I}\|_F^2 + \delta \sum_{i=1}^{l} \mathbf{w}_i' \mathbf{Kw}_i + \left(\epsilon \|\mathbf{A} - \mathbf{A}_0\|_F^2 \right)$$

Obs: relation, transition and inferring are all beneficial in improving the prediction

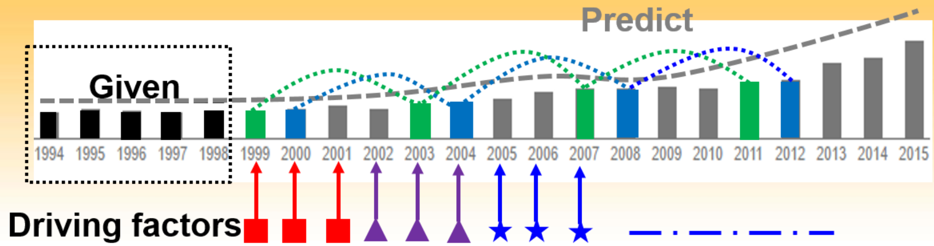
Robustness to Noise in A_0

$$\min_{\mathbf{W},\mathbf{B},\mathbf{A}} \|\mathbf{X}\mathbf{W}-\mathbf{Y}\|_F^2 + \alpha \sum_{i=1}^{l}\sum_{j=1}^{l} \mathbf{A}_{ij}\|\mathbf{X}\mathbf{w}_i - \mathbf{X}\mathbf{w}_j\|_2^2$$
$$+\beta\sum_{t=2}^{l}\|\mathbf{w}_t - \mathbf{B}\mathbf{w}_{t-1}\|_2^2 + \gamma\|\mathbf{B}-\mathbf{I}\|_F^2$$
$$+\delta\sum_{i=1}^{l}\|\mathbf{w}_i\|_2^2 + \epsilon\|\mathbf{A}-\mathbf{A}_0\|_F^2$$

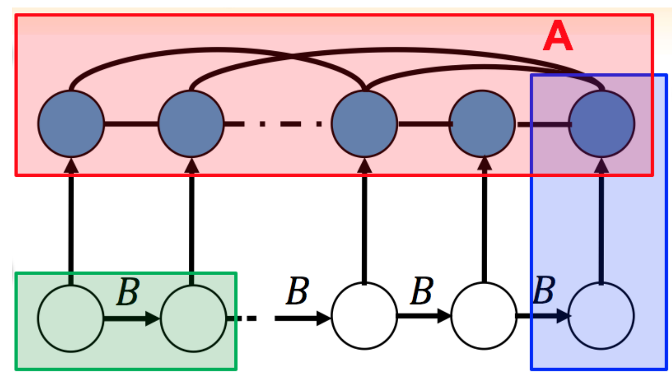


Obs: iPath degenerates gradually with the noise level

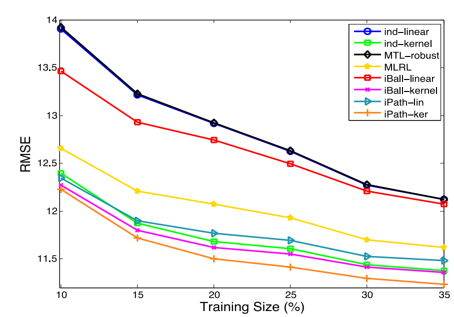
iPath: Summary



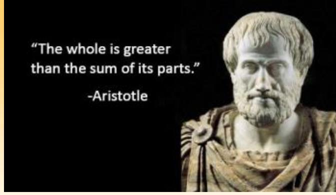
- **Goals**: predict the pathway to impact
- **Solutions**: *i*Path prediction model
  - Design objectives:
    - Prediction Consistency
    - Parameter Smoothness



  - Results:
    - Lower error than competitors
    - Robust to noise in impact relations



From the Ancient Philosophy



The whole is greater than the sum of its parts. -- **Aristotle**

- **Whole**: a collection of parts
- **Parts**: individual elements
- **Aristotle's hypothesis**:
  - whole > sum of parts

Part-Whole in Team Science

Research Team

Sports Team

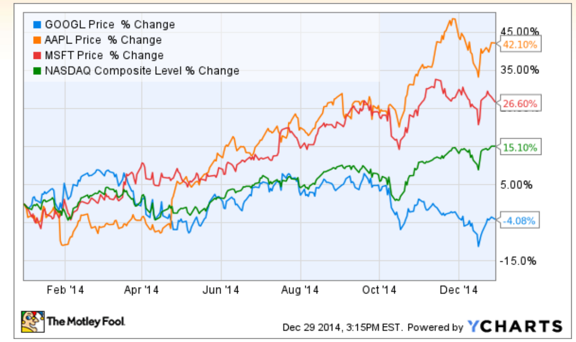
Film Crew

Sales Team

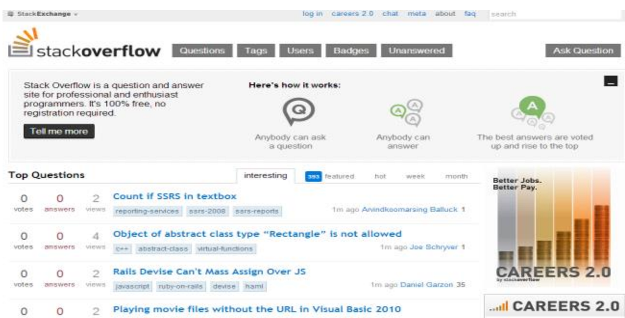
Whole – Team
Parts – Team members

Part-Whole Beyond Teams

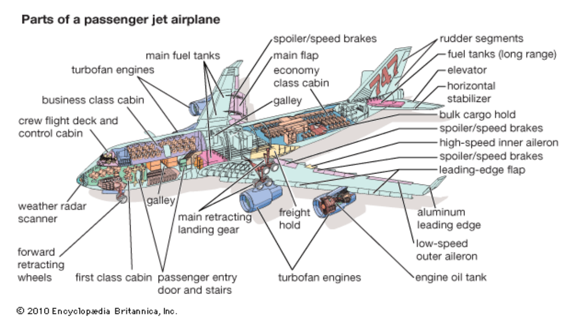
Autonomous System
Whole: system
Parts: individual drones



Stock Market
Whole: DJIA
Parts: individual stock



Community Question Answering
Whole: question
Parts: individual answers



System Reliability
Whole: system
Parts: individual component

Outcome of Part-Whole

Whole: Team
Part: Members

Whole outcome: Team's performance
Part outcome: each member's performance

Whole: Researcher
Part: Publications

Whole outcome: h-index
Part outcome: #citations of publications

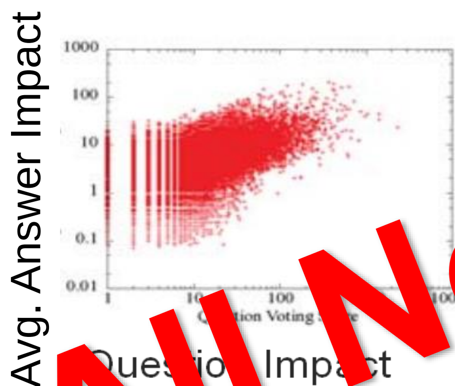
Question: how can we predict the outcome of whole/parts?

Predict the Part-Whole Outcomes

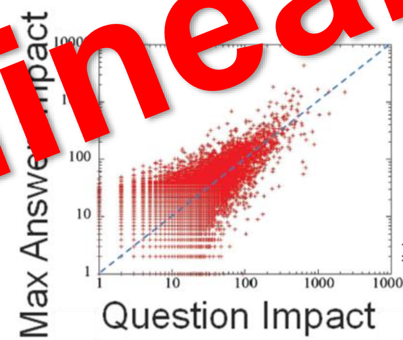
- Existing Algorithmic Work
  - Separate models for parts and whole
  - Joint linear models
- Aristotle’s hypothesis: whole>sum(parts)
- Question: how to jointly predict part/whole
  - by leveraging the part-whole relationship *beyond* the linear models?

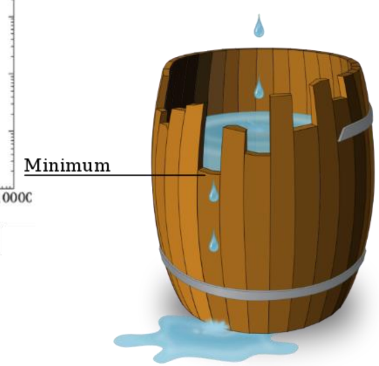
Challenges -- Modeling

- Nonlinear Part-whole Relationship
  - **Max**: impact of a question is strongly correlated with that of the *best* answer



All Non-linear!





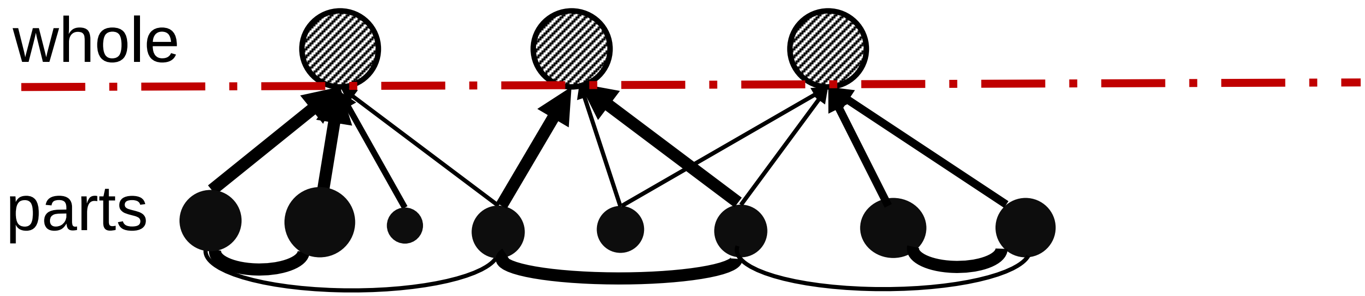
  - **Min**: classic Wooden Bucket Theory
  - **Sparsity**: team performance often dominated by a few top-performing team members

Challenges – Modeling (con't)

- Part-part Interdependency
  - Parts are connected via underlying network
  - Impact of such interdependency on outcomes

Hypothesis-1: similar parts -> similar contribution to whole

Hypothesis-2: similar parts -> similar parts outcome



Question: how can we utilize

1. nonlinear part-whole relationship
2. part-part interdependency

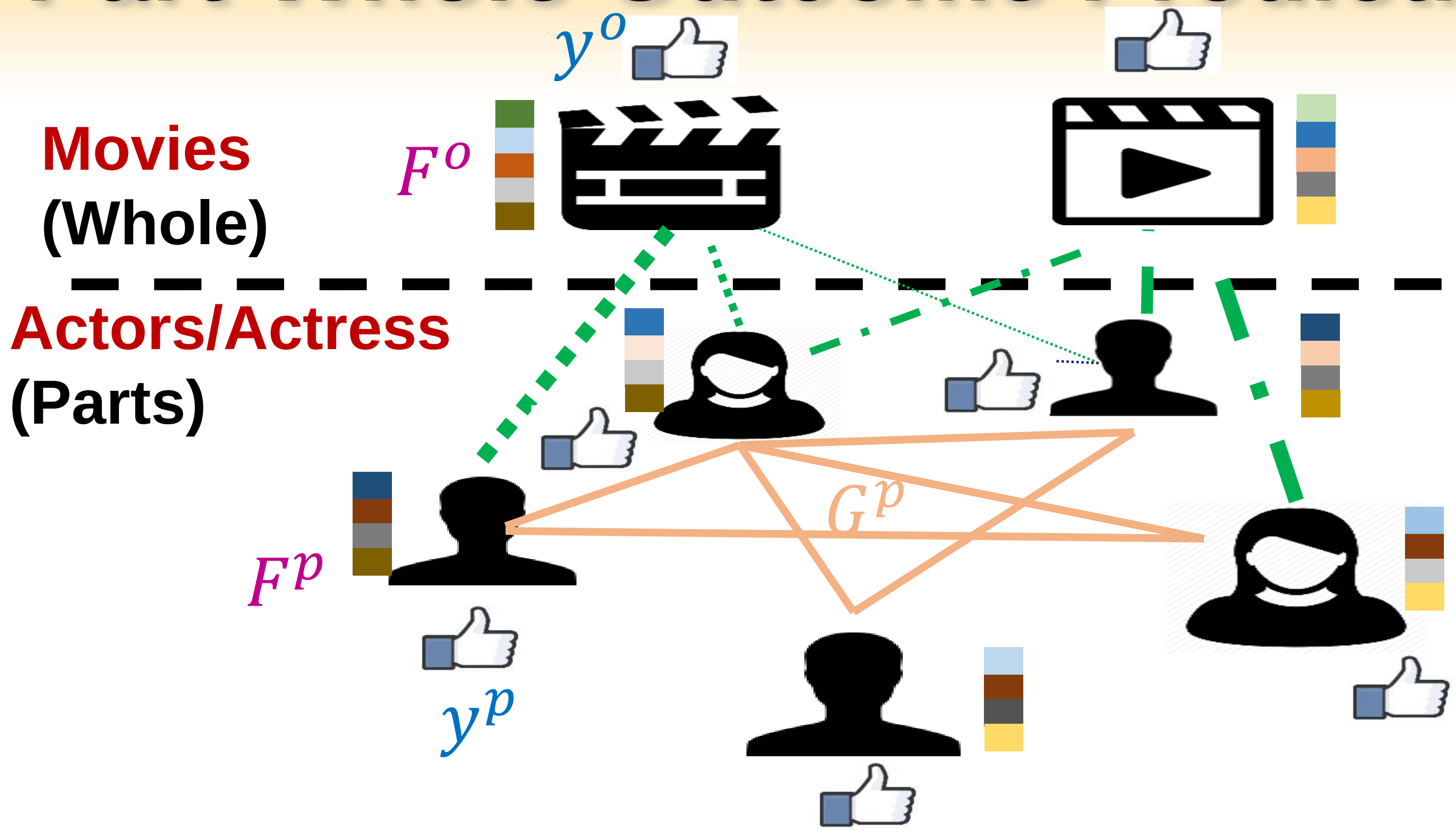
Challenges -- Algorithm

Non-linearity
+
Interdependency

high complexity

Question: how to scale up the computation?

Part-Whole Outcome Prediction



Given: 1. feature matrix for whole/part F^o/F^p

2. outcome vector for whole/part y^o/y^p

3. whole to part mapping ϕ

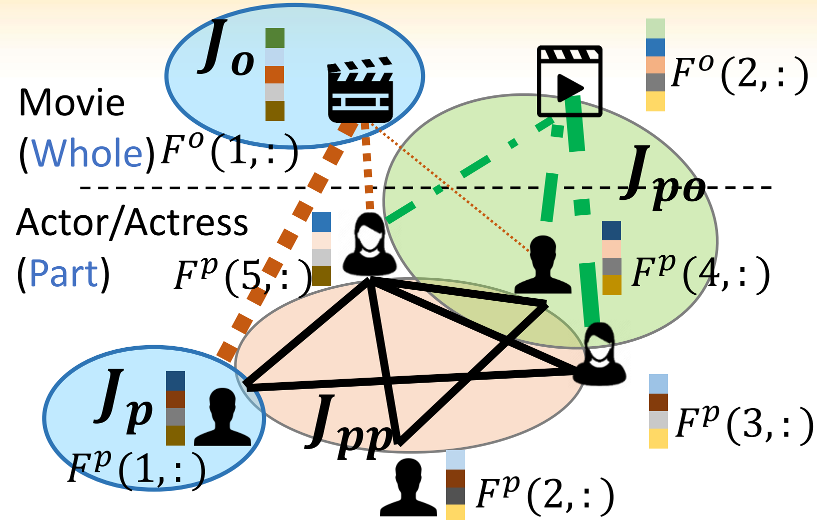
4. parts' network G^p (optional)

Predict: outcome of new whole/parts

A Generic Joint Prediction Framework -- PAROLE

- Formulation

$$\min J = J_o + J_p + J_{po} + J_{pp} + J_r$$



$$= \frac{1}{n_o}\sum_{i=1}^{n_o} L[f(F^o(i,:),w^o),y^o(i))]$$

$$+ \frac{1}{n_p}\sum_{i=1}^{n_p} L[f(F^p(i,:),w^p),y^p(i))]$$

$$+ \frac{\alpha}{n_o}\sum_{i=1}^{n_o} h(f(F^o(i,:),w^o),Agg(\phi(o_i)))$$

$$+ \frac{\beta}{n_p}\sum_{i=1}^{n_p}\sum_{j=1}^{n_p} G_{ij}^p g(f(F^p(i,:),w^p),f(F^p(j,:),w^p))$$

$$+ \gamma(\Omega(w^o) + \Omega(w^p))$$

J_o: Predictive Model for Whole

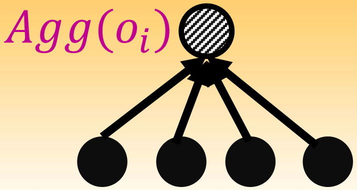
J_p: Predictive Model for Part

J_{po}: Part-whole Relationship

J_{pp}: Part-part Interdependency

J_r: parameter regularizer

Modeling Part-Whole Relationship

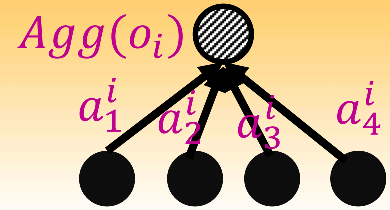


- **Overview**: for each whole entity $\mathrm{o_i}$, define

$$\underline{e_i} = \mathbf{F^o}(\mathrm{i},:)\mathbf{w^o} - \mathrm{Agg}(\mathrm{o_i})$$

  - e_i: Measure the difference between
    - predicted whole outcome using whole feature
    - predicted whole outcome using aggregated parts outcome
- **Key idea**: model part-whole relations by
  - Different loss functions on e_i
  - Different aggregation functions $Agg(\cdot)$

Overview



- **Intuition:** whole ← (weighted) sum of parts
- **Details**:

$$e_i = F^o(i,:)w^o - Agg(o_i)$$

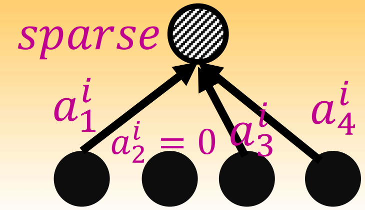
$$Agg(o_i) = \sum_{j \in \phi(o_i)} a_j^i F^p(j,:)w^p$$

  - a_j^i: weight of part j's contribution to the whole o_i's outcome
- **Remark**:
  - Characterize part-whole relationships
    - Use different loss functions on e_i
    - Use different norms on a_i

Linear Part-Whole Relation

- **Intuition**: Whole ← linear combination of parts
  - some parts play more important roles than the others in contributing to the whole outcome
- **Details**: $J_{po} = \frac{\alpha}{2n_o}\sum_{i=1}^{n_o} e_i^2$
- **Remark**:
  - $a_j^i = 1$: the whole is the sum of its parts
  - $a_j^i = \frac{1}{|o_i|}$: average coupling

Sparse Part-Whole Relation



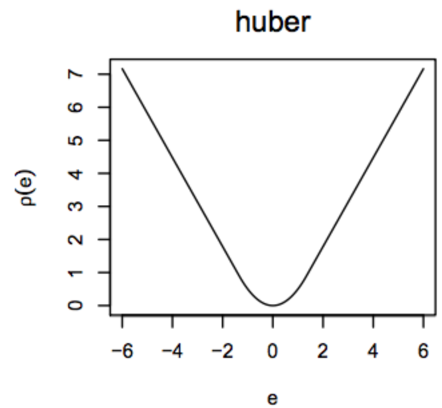
- **Intuition**: Whole ← a few parts
  - some parts have little or no effect on the whole outcome
- **Details**: $J_{po} = \frac{\alpha}{n_o} \sum_{i=1}^{n_o} (\frac{1}{2} e_i^2 + \lambda |\mathbf{a}_i|_1)$
- **Remark**:
  - The l_1 norm can shrink some part contributions a_j^i to exactly zero
  - **Nonlinear** part-whole relation

Ordered Sparse Part-Whole Relation

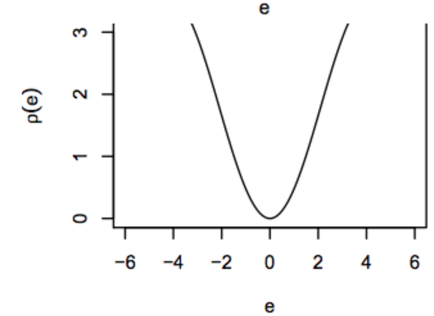
- **Intuition**: Whole ← a few top parts
  - team performance is determined by not only a few key members, but also the structural hierarchy between them
- **Details:** $J_{po} = \frac{\alpha}{n_o} \sum_{i=1}^{n_o} (\frac{1}{2} e_i^2 + \lambda \Omega_w(\mathbf{a}_i))$
  - $\Omega_w(x) = \sum_{i=1}^{n} |x|_{[i]} w_i = \boldsymbol{w}^T |\boldsymbol{x}|_{\downarrow}$: ordered weighted l_1 norm
  - $w \in \mathcal{K}_{m+}$: vector of non-increasing non-negative weights

Robust Part-Whole Relation

- **Intuition**: Whole ← parts that are not outliers
  - squared loss is sensitive to outliers
- **Solution**: robust regression model
- **Details**: $J_{po} = \frac{\alpha}{n_o} \sum_{i=1}^{n_o} \rho(e_i)$
  - $\rho(\cdot)$ is robust estimator



Method \ Case	$\lvert e\rvert \le t$	$\lvert e\rvert > t$
Huber $\rho_H(e)$	$\frac{1}{2}e^2$	$t\lvert e\rvert - \frac{1}{2}t^2$
Bisquare $\rho_B(e)$	$\frac{t^2}{6}\left\{1-\left[1-\left(\frac{e}{t}\right)^2\right]^3\right\}$	$\frac{t^2}{6}$



Maximum Part-Whole Relation

- **Intuition**: Whole ← max(parts)
  - team performance dominated by the best team member/leader
- **Details**:
  - $Agg(o_i) = \max(parts'outcome)$ [not differentiable]
  - Soft max function: $\max(x_1, x_2, \ldots, x_n) \approx \ln(\exp(x_1) + \exp(x_2) + \cdots + \exp(x_n))$
  - Aggregation: $Agg(o_i) = \ln(\sum_{j \in \phi(o_i)} \exp(F^p(j,:)w^p))$

$$J_{po} = \frac{\alpha}{2n_o} \sum_{i=1}^{n_o} e_i^2$$

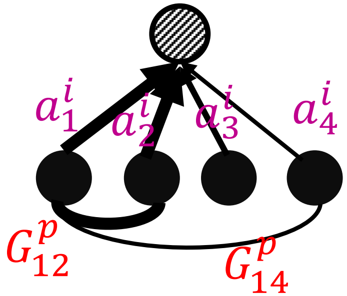
Summarize Part-Whole Relations

Name	$Agg(o_i)$ Aggregation of parts	J_{po} Sub-objective	Remark
Maximum	$\ln(\sum \exp(F^p(j,:)w^p))$	$\frac{\alpha}{2n_o}\sum e_i^2$	Nonlinear Whole ← max(parts)
Linear	$\sum a_j^i F^p(j,:)w^p$	$\frac{\alpha}{2n_o}\sum e_i^2$	Linear Whole ← linear combination of parts
Sparse	$\sum a_j^i F^p(j,:)w^p$	$\frac{\alpha}{n_o}\sum(\frac{1}{2}e_i^2 + \lambda\|a_i\|_1)$	Nonlinear Whole ← a few parts
Ordered Sparse	$\sum a_j^i F^p(j,:)w^p$	$\frac{\alpha}{n_o}\sum(\frac{1}{2}e_i^2 + \lambda\Omega_w(a_i))$	Nonlinear Whole ← a few top parts
Robust	$\sum a_j^i F^p(j,:)w^p$	$\frac{\alpha}{n_o}\sum \rho(e_i)$	Nonlinear Whole ← parts that are not outliers

Modeling Part-Part Interdependency

- **Effect on the whole outcome**
  - **Intuition**: closely connected parts might have similar contribution to the whole outcome
  - **Details**:

$$\mathcal{J}_{po} = \frac{\alpha}{n_o} \sum_{i=1}^{n_o} \left[\frac{1}{2} e_i^2 + \lambda |\mathbf{a}_i|_1 + \boxed{\frac{1}{2} \sum_{k,l \in \phi(o_i)} G^p_{kl} (a^i_k - a^i_l)^2} \right]$$



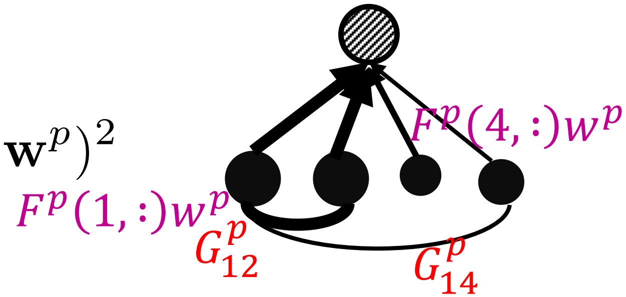
- Similar parts (large G^p_{kl})

→ similar contributions ($a^i_k \approx a^i_l$)

Modeling Part-Part Interdependency

- **Effect on the parts outcome**
  - **Intuition**: closely connected parts might share similar outcomes themselves
  - **Details**:

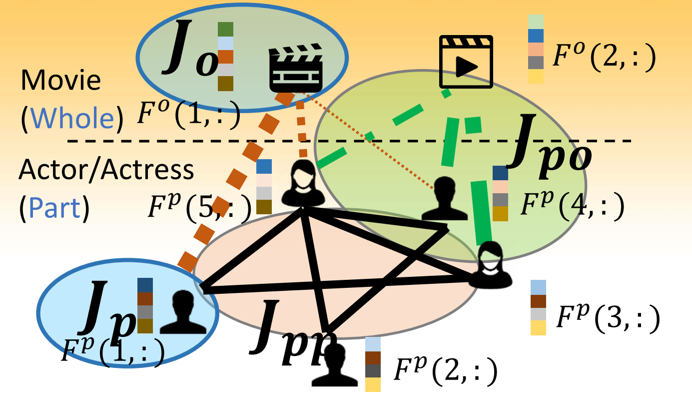
$$\mathcal{J}_{pp} = \frac{\beta}{2n_p} \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} G_{ij}^{p} (\mathbf{F}^p(i,:)\mathbf{w}^p - \mathbf{F}^p(j,:)\mathbf{w}^p)^2$$



  - Similar parts (large G_{ij}^p)

→ similar predicted outcomes ($F^p(i,:)w^p \approx F^p(j,:)w^p$)

Optimization Solution



- **Formulation**:
  - $J = J_o(w^o) + J_p(w^p) + J_{po}(w^o, w^p, a_j^i) + J_{pp}(w^p) + J_r(w^o, w^p)$
- **Observation**:
  - not jointly convex w.r.t. w^o, w^p, a_i^j
  - Convex w.r.t. to one block while fixing others
- **Solution**: block coordinate descent

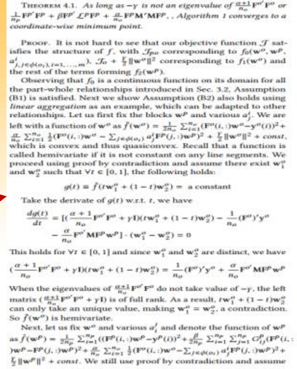
Block Coordinate Descent

- Three coordinate blocks: w^o, w^p, a_j^i
- Update one block while fixing others
- Update each block
  - (proximal) gradient descent

	$\frac{\partial J_{po}}{\partial w^o}$	$\frac{\partial J_{po}}{\partial w^p}$	$\frac{\partial J_{po}}{\partial a_i}$ **or proximal gradient update**
Maximum Agg	$\frac{\alpha}{n_o}\sum_{i=1}^{n_o} e_i (F^o(i,:))'$	$\frac{\alpha}{n_o}\sum_{i=1}^{n_o} e_i \frac{\sum_{j\in\phi(o_i)}(F^p(j,:))'\tilde{y}_i^p}{\sum_{j\in\phi(o_i)}\tilde{y}_i^p}$	**N/A**
Linear Agg	$\frac{\alpha}{n_o}(\mathrm{F^o})'(\mathrm{F^o w^o} - \mathrm{MF^p w^p})$	$-\frac{\alpha}{n_o}(\mathrm{F^p})'\mathrm{M}'(\mathrm{F^o w^o} - \mathrm{MF^p w^p})$	$e_i(-F^p(\phi(o_i),:)w^p) + L_i^p a_i$
Sparse Agg	$\frac{\alpha}{n_o}(\mathrm{F^o})'(\mathrm{F^o w^o} - \mathrm{MF^p w^p})$	$-\frac{\alpha}{n_o}(\mathrm{F^p})'\mathrm{M}'(\mathrm{F^o w^o} - \mathrm{MF^p w^p})$	$z = a_i - \tau\left[e_i(-F^p(\phi(o_i),:)w^p) + L_i^p a_i\right]$ $a_i \leftarrow prox_{\lambda\tau l_1}(z)$
Order Sparse Agg	$\frac{\alpha}{n_o}(\mathrm{F^o})'(\mathrm{F^o w^o} - \mathrm{MF^p w^p})$	$-\frac{\alpha}{n_o}(\mathrm{F^p})'\mathrm{M}'(\mathrm{F^o w^o} - \mathrm{MF^p w^p})$	$z = a_i - \tau\left[e_i(-F^p(\phi(o_i),:)w^p) + L_i^p a_i\right]$ $a_i \leftarrow prox_{\lambda\tau\Omega_w}(z)$
Robust Agg	$\frac{\alpha}{n_o}\sum_{i=1}^{n_o}\frac{\partial\rho(e_i)}{\partial e_i}F^o(i,:)'$	$\frac{\alpha}{n_o}\sum_{i=1}^{n_o}\frac{\partial\rho(e_i)}{\partial e_i}(-\sum_{j\in\phi(o_i)} a_j F^p(j,:)')$	$\frac{\alpha}{n_o}\left[\frac{\partial\rho(e_i)}{\partial e_i}(-F^p(\phi(o_i),:)w^p) + L_i^p a_i\right]$

Optimization Properties

details



Theorem 4.1. *As long as $-\gamma$ is not an eigenvalue of $\frac{\alpha+1}{n_o}\mathbf{F}^{o'}\mathbf{F}^o$ or $\frac{1}{n_p}\mathbf{F}^{p'}\mathbf{F}^p + \beta\mathbf{F}^{p'}\mathcal{L}^p\mathbf{F}^p + \frac{\alpha}{n_o}\mathbf{F}^p\mathbf{M}'\mathbf{M}\mathbf{F}^p$, , Algorithm 1 converges to a coordinate-wise minimum point.*

Proof. It is not hard to see that our objective function $\mathcal{J}$ satisfies the structure of f, with $\mathcal{J}_{po}$ corresponding to $f_0(\mathbf{w}^o, \mathbf{w}^p, a^i_{j, j\in\phi(o_i), i=1,\dots,m})$, $\mathcal{J}_o + \frac{\gamma}{2}\|\mathbf{w}^o\|^2$ corresponding to $f_1(\mathbf{w}^o)$ and the rest of the terms forming $f_2(\mathbf{w}^p)$.

Observing that f_0 is a continuous function on its domain for all the part-whole relationships introduced in Sec. 3.2, Assumption (B1) is satisfied. Next we show Assumption (B2) also holds using *linear aggregation* as an example, which can be adapted to other relationships. Let us first fix the blocks $\mathbf{w}^p$ and various a^i_j. We are left with a function of $\mathbf{w}^o$ as $\tilde{f}(\mathbf{w}^o) = \frac{1}{2n_o}\sum_{i=1}^{n_o}(\mathbf{F}^o(i,:)\mathbf{w}^o - \mathbf{y}^o(i))^2 + \frac{\alpha}{n_o}\sum_{i=1}^{n_o}\frac{1}{2}(\mathbf{F}^o(i,:)\mathbf{w}^o - \sum_{j\in\phi(o_i)} a^i_j\mathbf{F}^p(j,:)\mathbf{w}^p)^2 + \frac{\gamma}{2}\|\mathbf{w}^o\|^2 + const$, which is convex and thus quasiconvex. Recall that a function is called hemivariate if it is not constant on any line segments. We proceed using proof by contradiction and assume there exist $\mathbf{w}^o_1$ and $\mathbf{w}^o_2$ such that $\forall t \in [0,1]$, the following holds:

$$g(t) \equiv \tilde{f}(t\mathbf{w}^o_1 + (1-t)\mathbf{w}^o_2) = \text{ a constant}$$

Take the derivate of $g(t)$ w.r.t. t, we have

$$\frac{dg(t)}{dt} = [(\frac{\alpha+1}{n_o}\mathbf{F}^{o'}\mathbf{F}^o + \gamma\mathbf{I})(t\mathbf{w}^o_1 + (1-t)\mathbf{w}^o_2) - \frac{1}{n_o}(\mathbf{F}^o)'\mathbf{y}^o - \frac{\alpha}{n_o}\mathbf{F}^{o'}\mathbf{M}\mathbf{F}^p\mathbf{w}^p]\cdot(\mathbf{w}^o_1 - \mathbf{w}^o_2) = 0$$

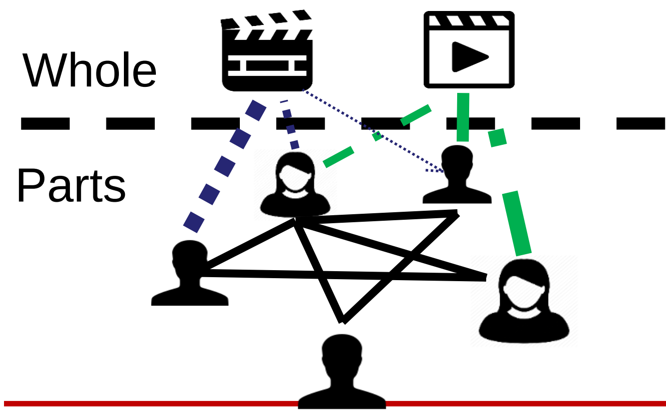
This holds for $\forall t \in [0,1]$ and since $\mathbf{w}^o_1$ and $\mathbf{w}^o_2$ are distinct, we have

$$(\frac{\alpha+1}{n_o}\mathbf{F}^{o'}\mathbf{F}^o + \gamma\mathbf{I})(t\mathbf{w}^o_1 + (1-t)\mathbf{w}^o_2) = \frac{1}{n_o}(\mathbf{F}^o)'\mathbf{y}^o + \frac{\alpha}{n_o}\mathbf{F}^{o'}\mathbf{M}\mathbf{F}^p\mathbf{w}^p$$

When the eigenvalues of $\frac{\alpha+1}{n_o}\mathbf{F}^{o'}\mathbf{F}^o$ do not take value of $-\gamma$, the left matrix $(\frac{\alpha+1}{n_o}\mathbf{F}^{o'}\mathbf{F}^o + \gamma\mathbf{I})$ is of full rank. As a result, $t\mathbf{w}^o_1 + (1-t)\mathbf{w}^o_2$ can only take an unique value, making $\mathbf{w}^o_1 = \mathbf{w}^o_2$, a contradiction. So $\tilde{f}(\mathbf{w}^o)$ is hemivariate.

Next, let us fix $\mathbf{w}^o$ and various a^i_j and denote the function of $\mathbf{w}^p$ as $\hat{f}(\mathbf{w}^p) = \frac{1}{2n_p}\sum_{i=1}^{n_p}((\mathbf{F}^p(i,:)\mathbf{w}^p - \mathbf{y}^p(i))^2 + \frac{\beta}{2n_p}\sum_{i=1}^{n_p}\sum_{j=1}^{n_p} G^p_{ij}(\mathbf{F}^p(i,:)\mathbf{w}^p - \mathbf{F}^p(j,:)\mathbf{w}^p)^2 + \frac{\alpha}{n_o}\sum_{i=1}^{n_o}\frac{1}{2}(\mathbf{F}^o(i,:)\mathbf{w}^o - \sum_{j\in\phi(o_i)} a^i_j\mathbf{F}^p(j,:)\mathbf{w}^p)^2 + \frac{\gamma}{2}\|\mathbf{w}^p\|^2 + const$. We still use proof by contradiction and assume

- **Convergence and Optimality**
  - Under mild conditions, the optimization alg converges to a coordinate-wise minimum point
- **Complexity**
  - The alg scales linearly w.r.t. the size of part-whole graph in both time and space



Complexity: $O(n_o d_o + n_p d_p + m_{po} + m_{pp})$

n_o: #whole entities

n_p: #part entities

m_{po}: #links from whole to parts

m_{pp}: #links in part-part network

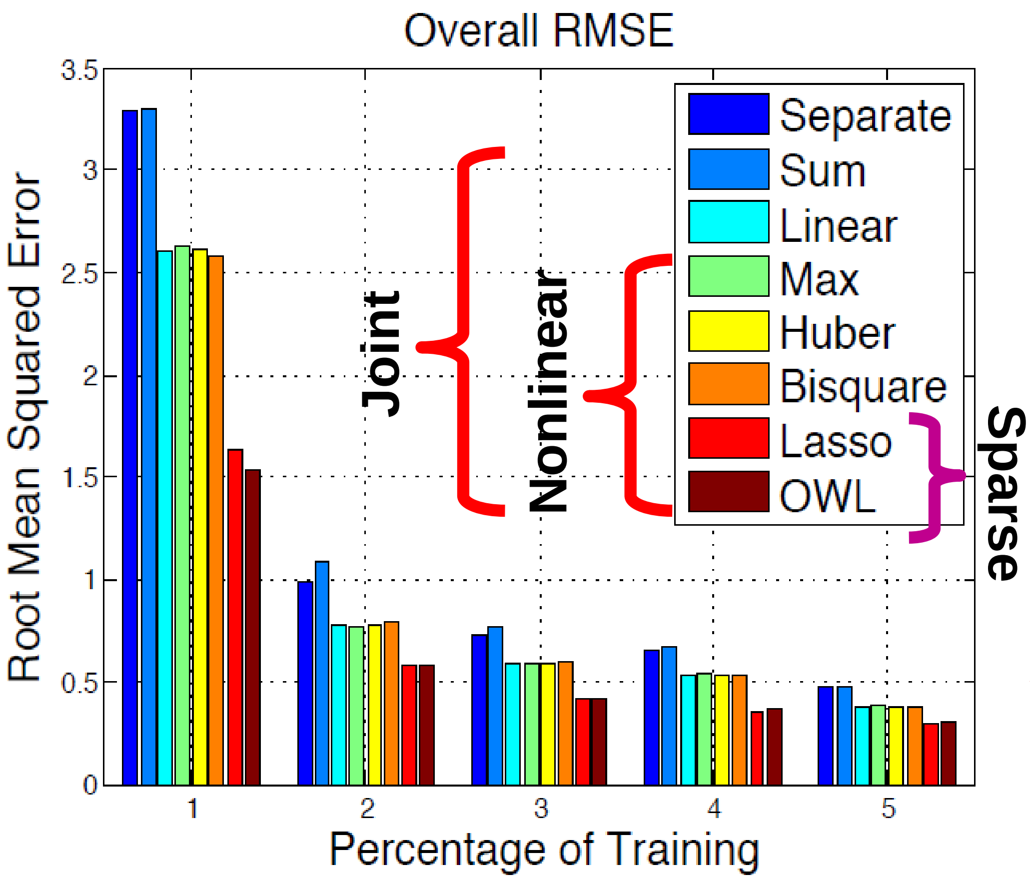
d_o, d_p: feature dimension of whole, parts

Datasets

Data	Whole	Part	#Whole	#Part
Math	Question (#votes)	Answer (#votes)	16,638	32,876
SO	Question (#votes)	Answer (#votes)	1,966,272	4,282,570
DBLP	Author (h-index)	Paper (#citation)	234,681	129,756
Movie	Movie (# )	Actors/Actress (# )	5,043	37,365

- **Setup**: sort whole in chronological order, gather first x percent and corresponding parts as training, test on last 10%
- **Metric**: root mean squared error (RMSE)

Outcome Prediction Performance



Math

Observations

1. Joint prediction models > separate models
2. Non-linear part-whole relationships (max, Huber, Bisquare, Lasso, OWL) > linear relationships (Sum, Linear)
3. Lasso and OWL are the best methods in most cases

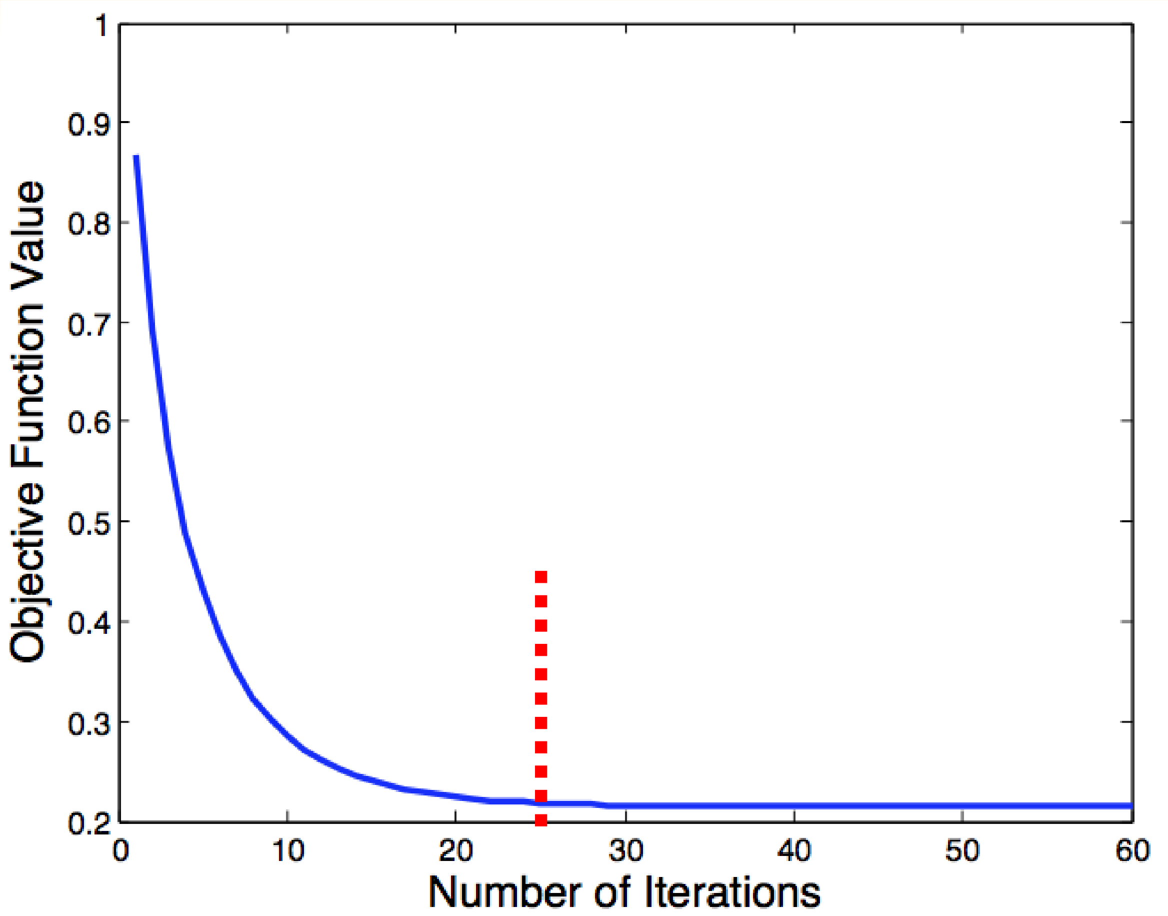
Effect of part-part interdependency

Movie

- PAROLE-Basic – no network information
- Part-part interdependency on whole outcome and parts outcome both boost the performance

Convergence Analysis

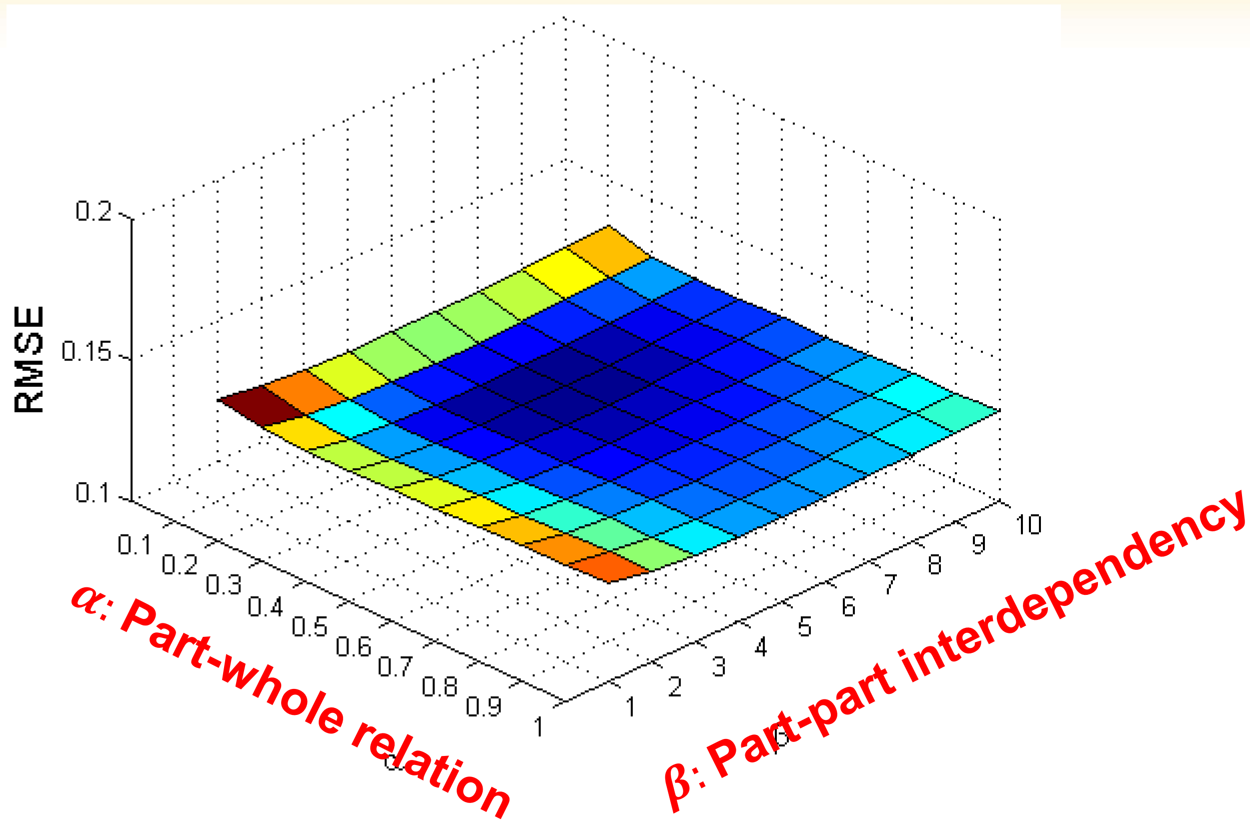
SO



- PAROLE converges fast (25-30 iterations)

Parameter Sensitivity

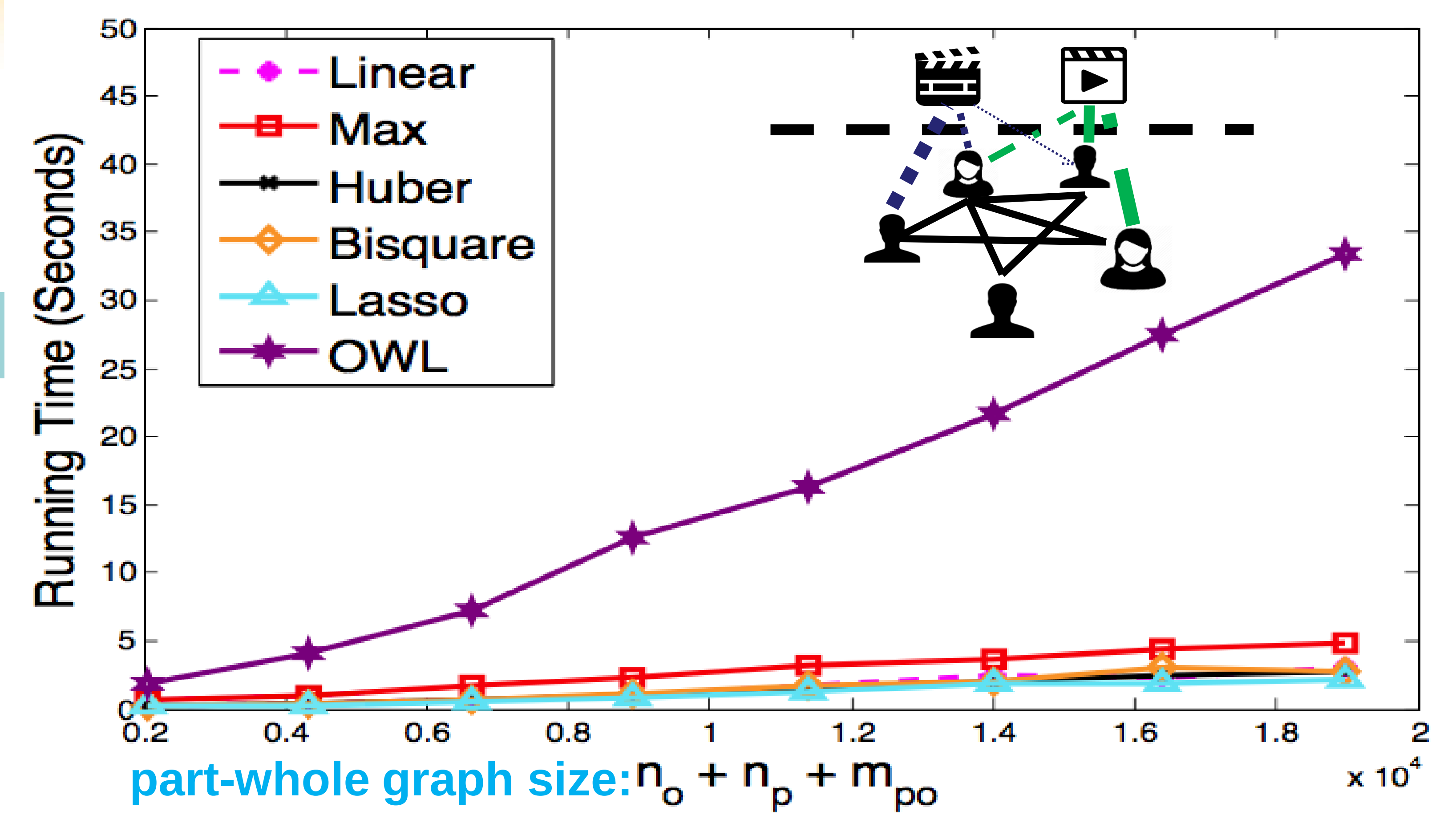
Movie



- α controls importance of part-whole relation
- β controls importance of part-part interdependency
- Stable in a relatively large parameter space

Scalability of PAROLE

SO

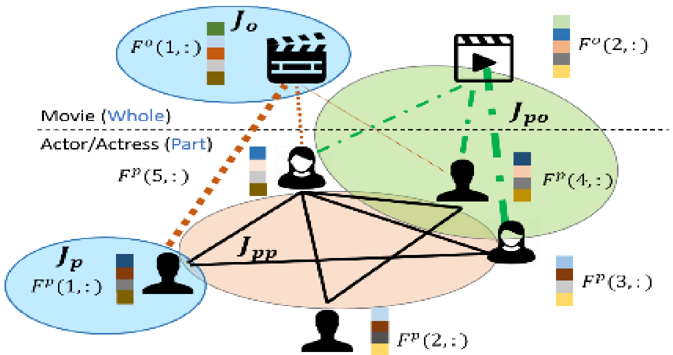


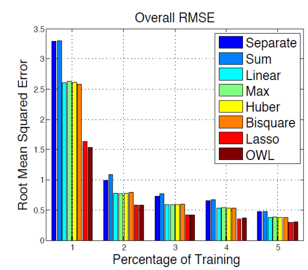
part-whole graph size: $n_o + n_p + m_{po}$

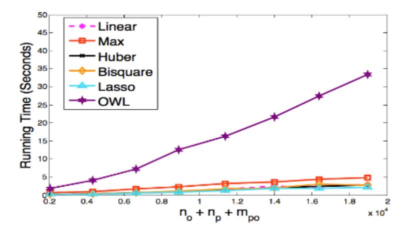
- PAROLE scales linearly w.r.t. part-whole graph size

Conclusions -- PAROLE

- **Goals**: leverage potentially non-linear part-whole relationships for outcome prediction
- **Solutions**: PAROLE
  - **Modeling**
    - Part-whole relationship
    - Part-part interdependency
  - **Optimization**
    - Block coordinate descent
    - Converges to a coordinate-wise minimum point
    - Scales linearly w.r.t. the part-whole graph size







Roadmap

- Motivations and Background
- Part I: Team Performance Characterization
- Part II: Team Performance Prediction
- Part III: Team Performance Optimization
- Open Challenges
- Demo

Part III: Team Performance Optimization

- Team Formation and its variants
- Team Member Replacement
- Team Enhancement

Simple Team formation Problem

- Input:
  - A task T, consisting of a set of skills
  - A set of candidate experts each having a subset of skills

T = {algorithms, java, graphics, python}

Alice	Bob	Cynthia	David	Eleanor
{algorithms}	{python}	{graphics, java}	{graphics}	{graphics,java,python}

- Problem: Given a task and a set of experts, find the smallest subset (team) of experts that together have all the required skills for the task

Set Cover

- The Set Cover problem:
  - We have a universe of elements $U = \{x_1, \ldots, x_N\}$
  - We have a collection of subsets of U, $\boldsymbol{S} = \{S_1, \ldots, S_n\}$, such that $\bigcup_i S_i = U$
  - We want to find the smallest sub-collection $\boldsymbol{C} \subseteq \boldsymbol{S}$ of $\boldsymbol{S}$, such that $\bigcup_{S_i \in \boldsymbol{C}} S_i = U$
    - The sets in $\boldsymbol{C}$ cover the elements of U

Coverage

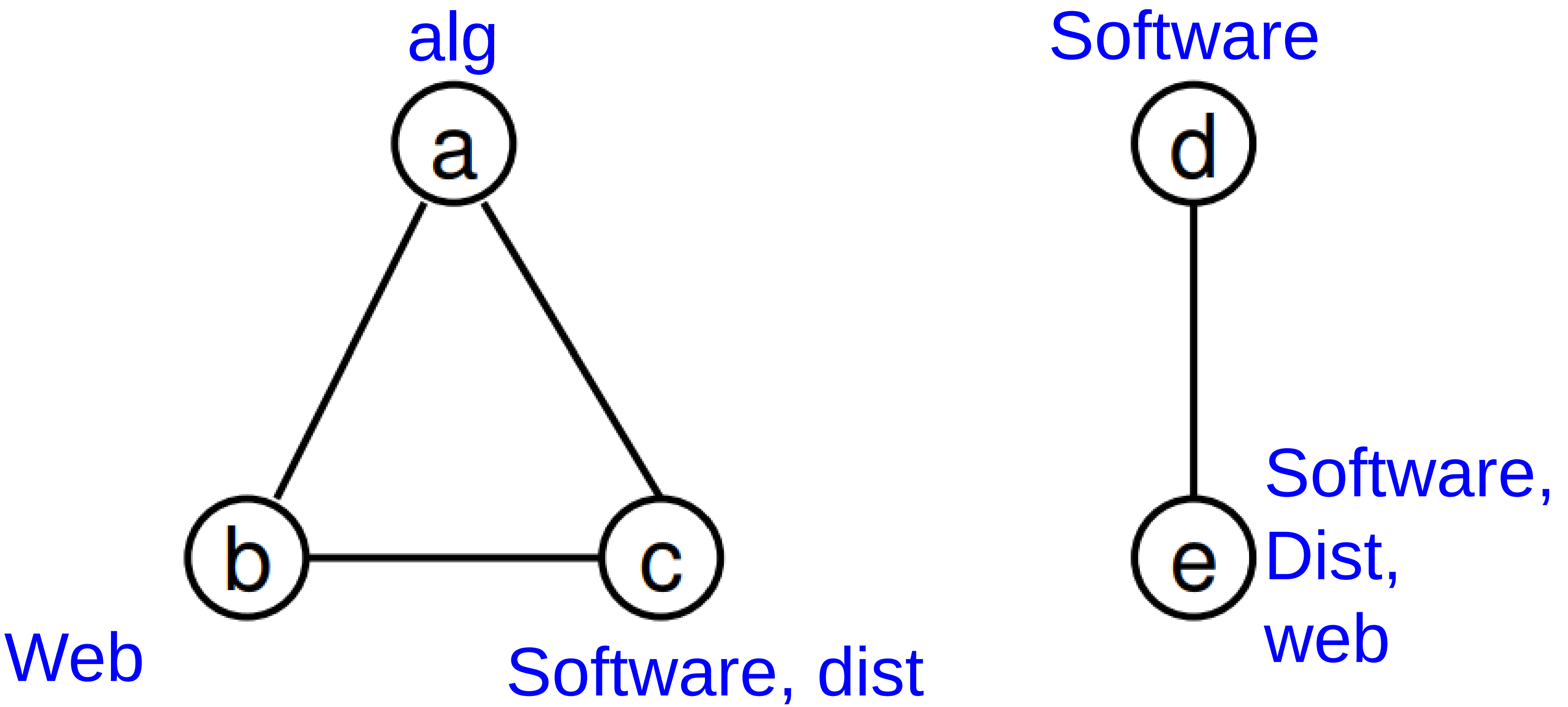
- The Simple Team Formation Problem is a just an instance of the Set Cover problem
  - Universe U of elements = Set of all skills
  - Collection $\boldsymbol{S}$ of subsets = The set of experts and the subset of skills they possess.

T = {algorithms, java, graphics, python}

Alice	Bob	Cynthia	David	Eleanor
{algorithms}	{python}	{graphics, java}	{graphics}	{graphics,java,python}

Team Formation with Networks

- T = {algorithms, software engineering, distributed systems, web programming}



Theodoros Lappas, Kun Liu, and Evimaria Terzi. 2009. Finding a team of experts in social networks. KDD, 2009.

Problem Definition

- Given:
  - Task requiring a set of skills
  - Set of individuals
  - Skills possessed by each individual
  - Graph of communication cost between individuals
- Find
  - A subset of individuals containing all required skills with minimized communication cost

Theodoros Lappas, Kun Liu, and Evimaria Terzi. 2009. Finding a team of experts in social networks. KDD, 2009.

Communication cost

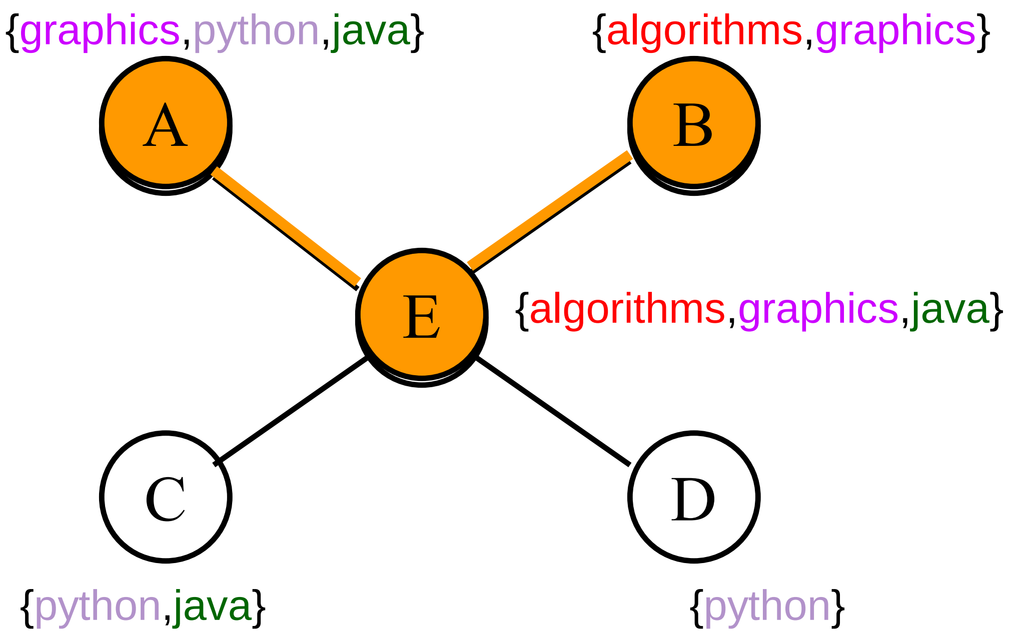
- Diameter (*CC-R*)
  - Diameter of the subgraph of the selected individuals
  - NP-complete (reduce to Multiple-Choice Cover)
- Minimum Spanning Tree (*CC-MST*)
  - Cost of the MST on the subgraph of the selected individuals
  - NP-complete (reduce to Group Seiner Tree)

Algorithm for Diameter-TF

- For every skill a required by the task T, compute $S(a)$: the individuals with a
- Pick the skill a_{rarest} with lowest cardinality
- Among all candidates from the set $S(a_{rarest})$, pick the one that leads to the smallest diameter

The RarestFirst algorithm

T={algorithms,java,graphics,python}



Skills:

- algorithms
- graphics
- java
- python

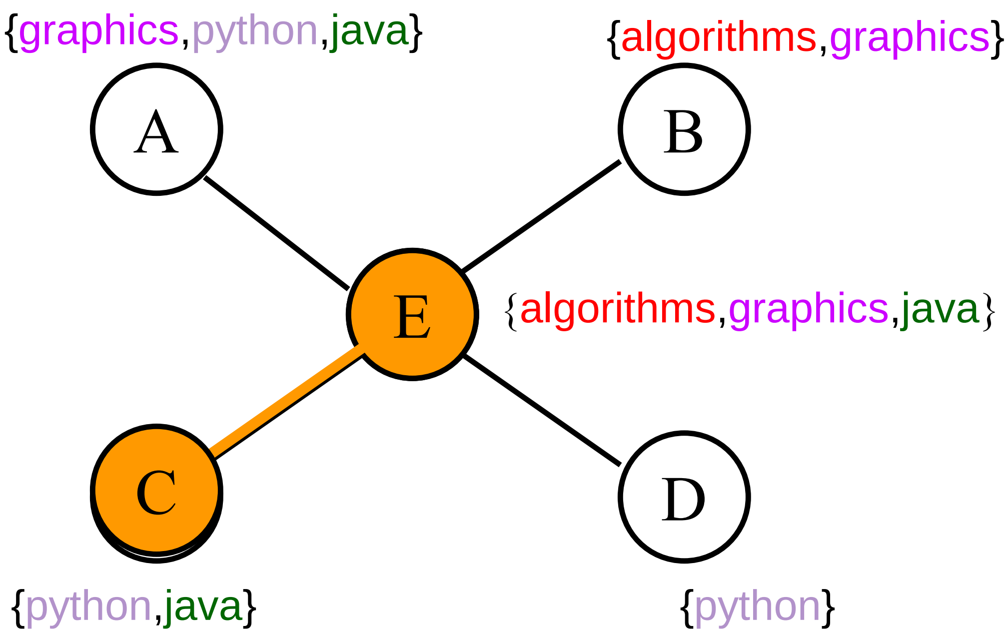
α_{rare} = algorithms

S_{rare} ={Bob, Eleanor}

Diameter = 2

The RarestFirst algorithm

T={algorithms,java,graphics,python}



Skills:

- algorithms
- graphics
- java
- python

α_{rare} = algorithms

S_{rare} ={Bob, Eleanor}

Diameter = 1

Algorithm for MST-TF

- CoverSteiner
  - $X_0 \leftarrow$ GreedyCover
    - Add individuals with most uncovered skills
  - $X' \leftarrow$ SteinerTree(G, X_0)

$\mathcal{X}' \leftarrow v$, where v is a random node from $\mathcal{X}_0$.
while $(\mathcal{X}_0 \setminus \mathcal{X}') \neq \emptyset$ **do**
    $v* \leftarrow \arg\min_{u \in \mathcal{X}_0 \setminus \mathcal{X}'} d(u, \mathcal{X}')$
    if $Path(v^*, \mathcal{X}') \neq \emptyset$ **then**
        $\mathcal{X}' \leftarrow \mathcal{X}' \cup \{Path(v^*, \mathcal{X}')\}$
    else
        Return Failure

Another algorithm for MST-TF

1: $H \leftarrow \texttt{EnhanceGraph}(G, T)$
2: $\mathcal{X}_H \leftarrow \texttt{SteinerTree}(H, \{Y_1, \ldots, Y_k\})$
3: $\mathcal{X}' \leftarrow \mathcal{X}_H \setminus \{Y_1, \ldots, Y_k\}$

EnhanceGraph:

For every skill a_j in T

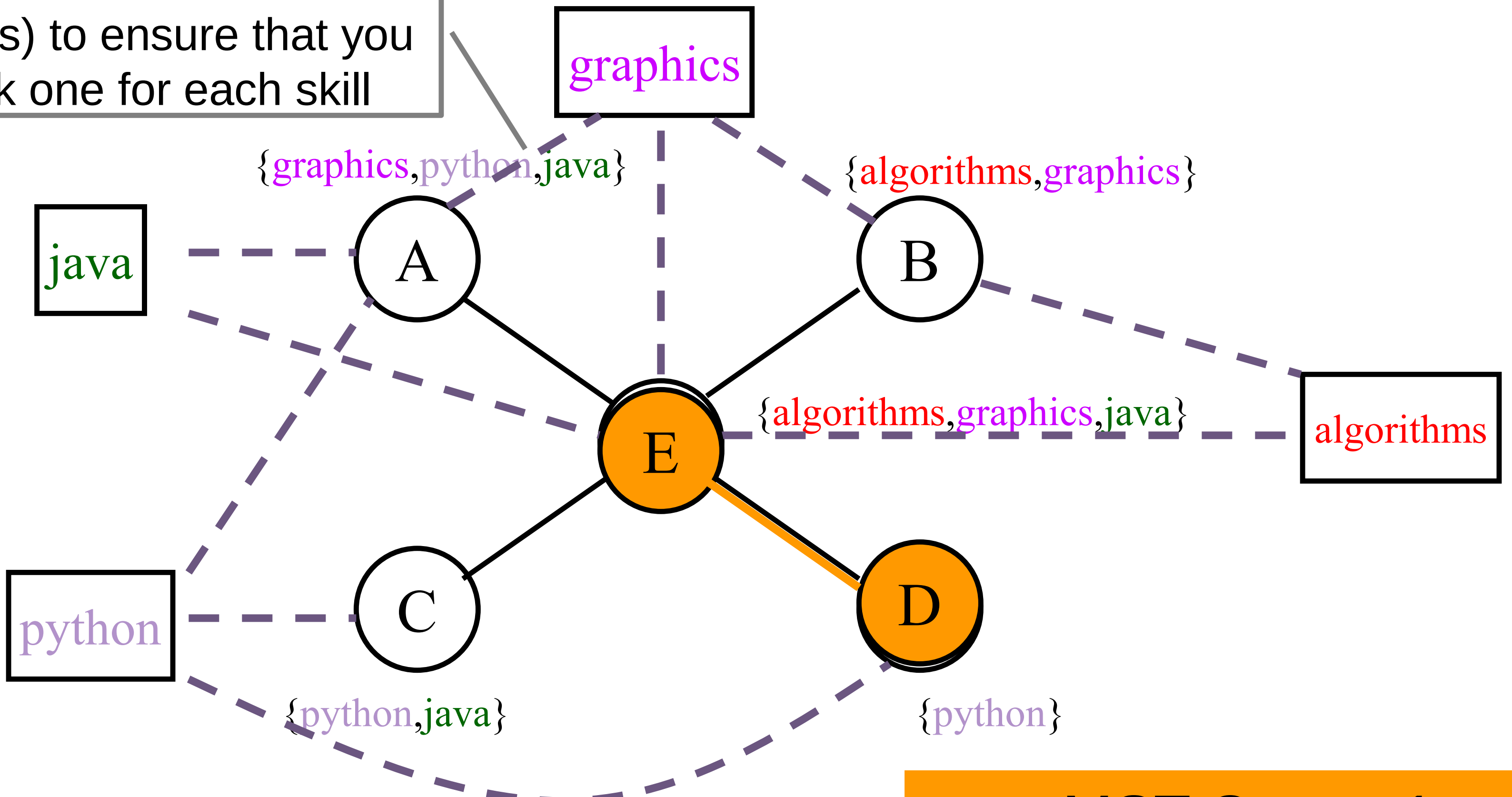
1. Create an additional node Y_j

2. Connect Y_j to all individuals with a_j with large weight

The EnhancedSteiner algorithm

Put a large weight on the new edges (more than the sum of all edges) to ensure that you only pick one for each skill

T={algorithms,java,graphics,python}

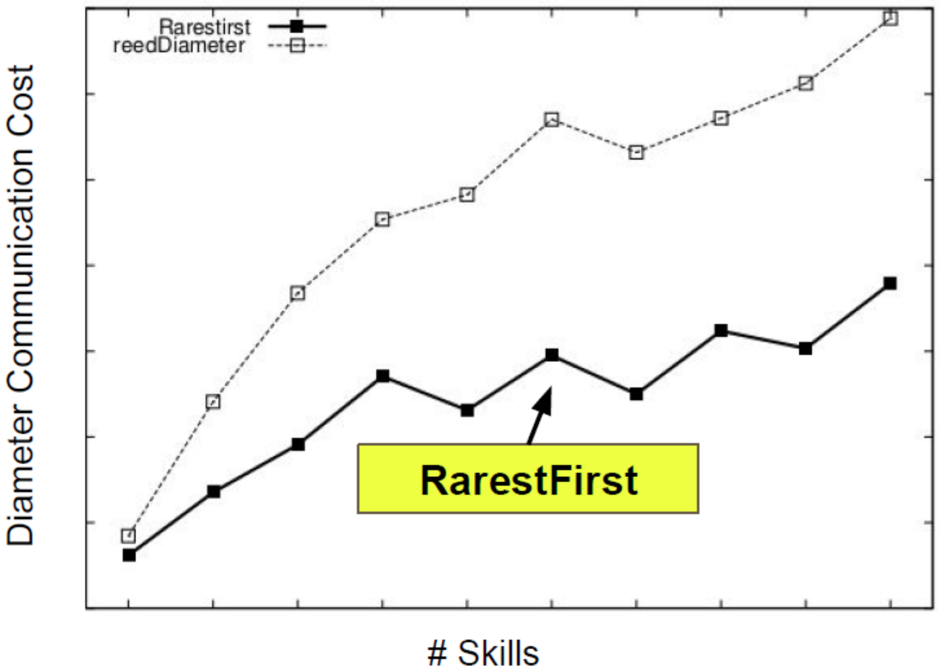


MST Cost = 1

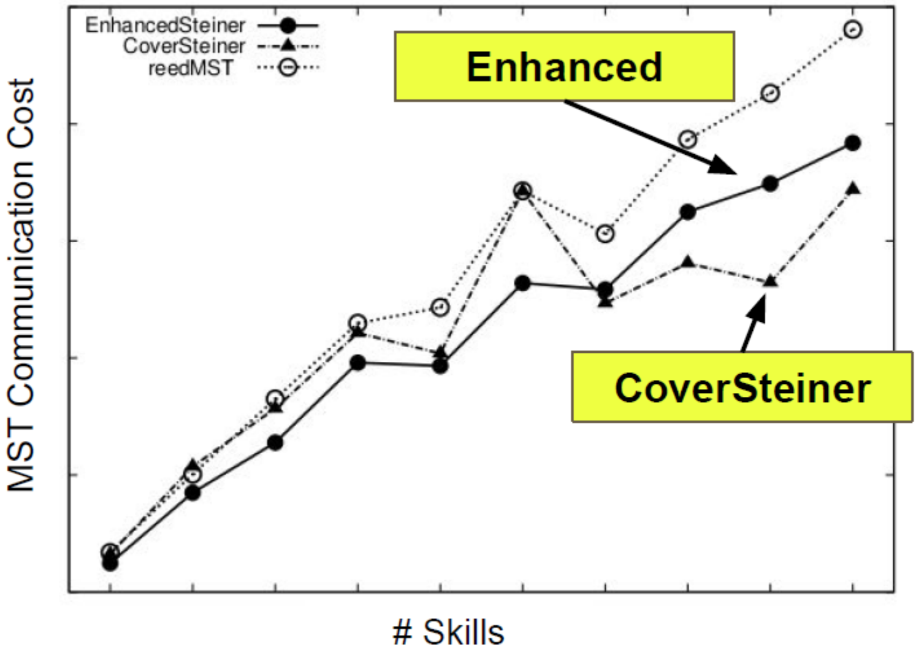
Experimental Evaluation

- DBLP: papers in database, data mining, AI, theory
- Skills derived from common terms in paper titles
- Communication weights determined by co-authorship
- 5509 individuals, 1792 skills
- Tasks generated with 2 to 20 skills
- Average over 100 combinations

Communication Cost



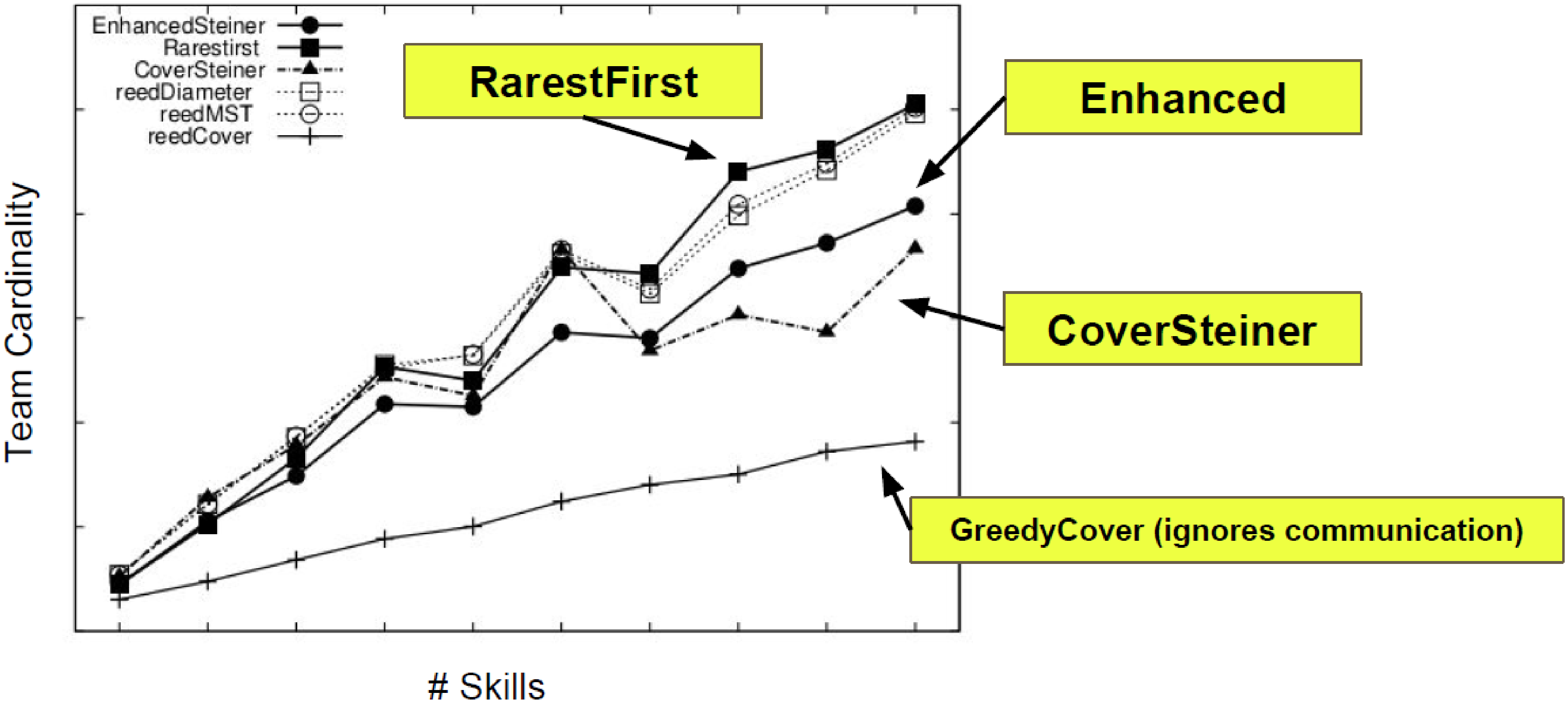
(a) Cc-R cost



(b) Cc-Mst cost

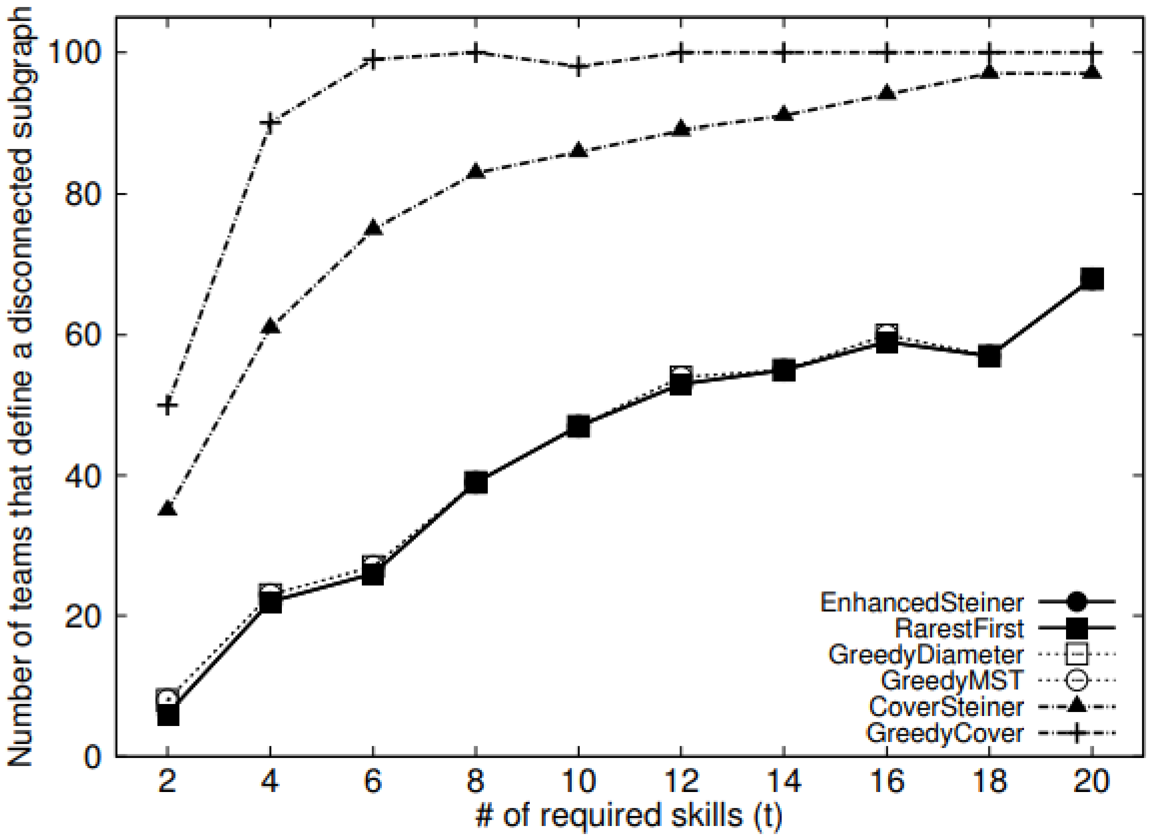
Theodoros Lappas, Kun Liu, and Evimaria Terzi. 2009. Finding a team of experts in social networks. KDD, 2009.

Cardinality of the team



(a) Cardinality of the team.

Connectivity of the team



(b) Number of disconnected teams.

Theodoros Lappas, Kun Liu, and Evimaria Terzi. 2009. Finding a team of experts in social networks. KDD, 2009.

Case study on 10 papers

Rank	Paper title
1	The anatomy of a large-scale hypertextual Web search engine
2	Fast algorithms for mining association rules
3	Mining association rules between sets of items in large databases
4	Text categorization with support vector machines: Learning with many relevant features
5	Conditional random fields: Probabilistic models for segmenting and labeling sequence data
6	Mining frequent patterns without candidate generation
7	A survey of approaches to automatic schema matching
8	Automatic subspace clustering of high dimensional data for data mining applications
9	Models and issues in data stream systems
10	NiagaraCQ: A Scalable Continuous Query System for Internet Databases

Case Study Results

Rank	Actual authors	RarestFirst result	EnhancedSteiner result
1	S. Brin, L. Page	**Paolo Ferragina, Patrick Valduriez, H. V. Jagadish, Alon Y. Levy, Daniela Florescu** Divesh Srivastava, S. Muthukrishnan	**P. Ferragina ,J. Han, H. V. Jagadish, Kevin Chen-Chuan Chang, A. Gulli**, S. Muthukrishnan, Laks V. S. Lakshmanan
2	R. Agrawal, R. Srikant	**R. Agrawal**	**Philip S. Yu**
3	R. Agrawal, T. Imielinski, A. N. Swami	**Philip S. Yu**	**Wei Wang, Philip S. Yu**
4	T. Joachims	**Wei-Ying Ma, Gui-Rong Xue, H. Liu, J. Han, H. Lu, Z. Chen,** Q.Yang, H. Cheng	**J. Han, H. Lu, Wei-Ying Ma, Z. Chen, H. Liu, Gui-Rong Xue**, Q. Yang
5	J. Lafferty, F. Pereira, A. McCallum	**A. McCallum**	**A. McCallum**
6	J. Han, J. Pei, Y. Yin	**F. Bonchi**	**A. Gionis, H. Mannila, R. Motwani**
7	E. Rahm, P. A. Bernstein	**C. Bettini, R. Agrawal, Kevin Chen-Chuan Chang**, T. Imielinski, H. Garcia-Molina, D. Barbara, S. Jajodia	**C. Bettini, P. A. Bernstein,** H. Garcia-Molina, S. Jajodia, D. Maier, D. Barbara
8	R. Agrawal, J. Gehrke, D. Gunopulos, P. Raghavan	**D. Gunopulos, R. Agrawal**	**R. Agrawal, D. Gunopulos**
9	B. Babcock, S. Babu, M. Datar, R. Motwani, J. Widom	**M. T. Ozsu**	**H. V. Jagadish, D. Srivastava**
10	J. Chen, D. J. DeWitt, F. Tian, Y. Wang	**Donald Kossmann, David J. DeWitt, Michael J. Franklin, Michael J. Carey**	**M. J. Carey, M. J. Franklin, D. Kossmann, D. J. DeWitt**

Theodoros Lappas, Kun Liu, and Evimaria Terzi. 2009. Finding a team of experts in social networks. KDD, 2009.

Steaming Tasks

- Steam of tasks arriving online
- Create teams on-the-fly for each task
  - Teams should be fit for the tasks
  - Allocation should be fair to people

Aris Anagnostopoulos, Luca Becchetti, Carlos Castillo, Aristides Gionis, and Stefano Leonardi. 2010. Power in unity: forming teams in large-scale community systems. CIKM, 2010.

Balanced Task Covering

$$\min L \tag{1}$$

$$\sum_{j=1}^{n} \mathbf{p}_\ell^j X_{ji} \geq \mathbf{J}_\ell^i, \qquad \forall i = 1, \ldots, k,\ \ell = 1, \ldots, m \tag{2}$$

All tasks are executed

$$\sum_{i=1}^{k} X_{ji} \leq L, \qquad \forall j = 1, \ldots, n \tag{3}$$

Load balancing

$$L \geq 0, X_{ji} \in \{0, 1\} \tag{4}$$

Aris Anagnostopoulos, Luca Becchetti, Carlos Castillo, Aristides Gionis, and Stefano Leonardi. 2010. Power in unity: forming teams in large-scale community systems. CIKM, 2010.

Online TF in Social Networks

- Forming teams that can accomplish the specified tasks while optimizing:
  - **Load**: number of tasks one expert participates
  - **Coordination Cost**:
    - Steiner tree
    - Diameter

Aris Anagnostopoulos, Luca Becchetti, Carlos Castillo, Aristides Gionis, and Stefano Leonardi. Online Team Formation in Social Networks. WWW, 2012.

Balanced Social Task

$$\min \max_{i \in \mathcal{P}} L(\mathbf{p}^i)$$

Load balancing

$$\mathbf{cov}(\mathbf{J}^j, \mathbf{q}^j) = 1 \qquad \forall j \in \mathcal{J}$$

All tasks are executed

$$c(Q^j) \leq B \qquad \forall j \in \mathcal{J}$$

Bound on the communication cost

Aris Anagnostopoulos, Luca Becchetti, Carlos Castillo, Aristides Gionis, and Stefano Leonardi. Online Team Formation in Social Networks. WWW, 2012.

Realistic Team Formation

- Realistic Requirements
  - Inclusion of a designated team leader and/or a group of experts
  - Skill requirement
  - Team size, or team cost
  - Locality of the team, e.g., in a geographical sense

Measure of collaborative compatibility

- Generalized form of subgraph density
  - $density(C) = \frac{assoc(C)}{vol_g(C)} = \frac{\sum_{i,j \in C} w_{ij}}{\sum_{i \in C} g_i}$
  - Strict monotonicity
  - Robustness

Syama Sundar Rangapuram , Thomas Bühler , Matthias Hein, Towards realistic team formation in social networks based on densest subgraphs. WWW, 2013.

Problem Formulation

$$\max_{C \subseteq V} \frac{\mathrm{assoc}(C)}{\mathrm{vol}_g(C)}$$

$$\text{subject to} : S \subseteq C$$ Required inclusion

Skill Requirement $$\kappa_j \leq \mathrm{vol}_{M_j}(C) \leq \iota_j , \quad \forall j \in \{1, \ldots, p\}$$

$$|C| \leq b$$ Team size

$$\mathrm{vol}_c(C) \leq B$$ Budget constraint

$$\mathrm{dist}(u, v) \leq d_0 , \quad \forall u, v \in C,$$ Team locality

Syama Sundar Rangapuram , Thomas Bühler , Matthias Hein, Towards realistic team formation in social networks based on densest subgraphs. WWW, 2013.

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?

- L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Replacing the Irreplaceable: Fast Algorithms for Team Member Recommendation, WWW 2015
- N. Cao, Y.-R. Lin, L. Li, H. Tong: g-Miner: Interactive Visual Group Mining on Multivariate Graphs, ACM CHI 2015
- System prototype & video demo: http://team-net-work.org

Team Member Replacement

Problem Definition:

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

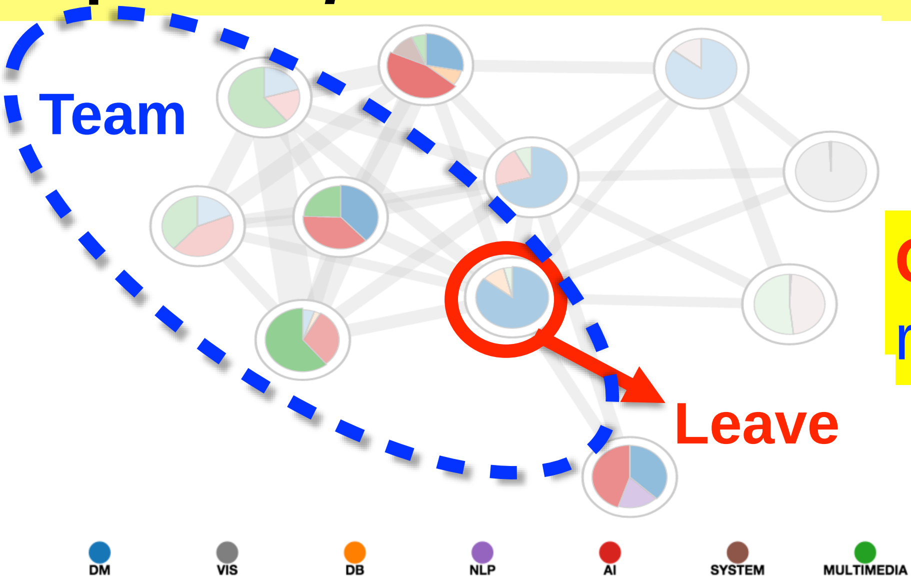
Adj. Matrix

Skill Indicator

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

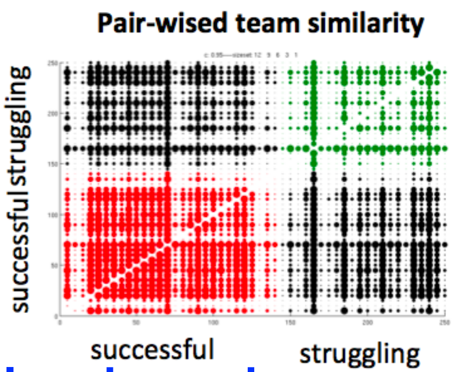
(3) A team member $p \in \mathcal{T}$

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 Science Literature

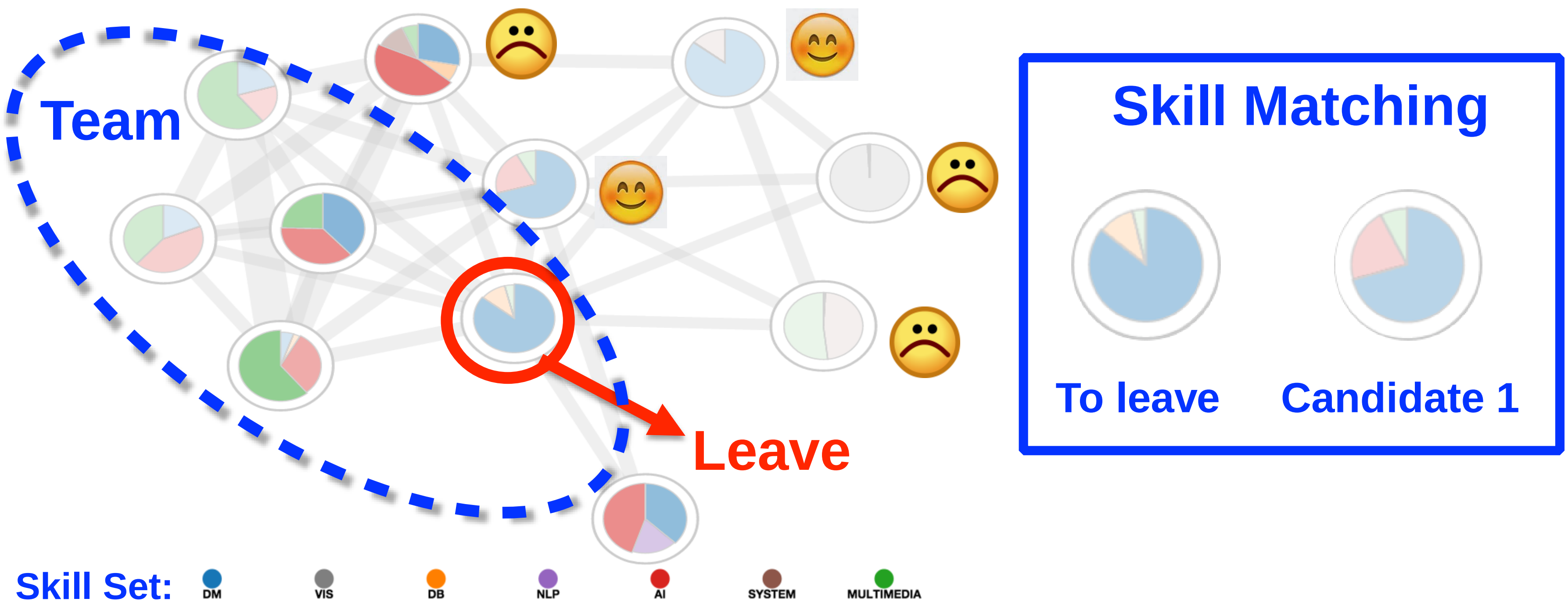


- Team members prefer to work with people they have worked before [Hinds+OBHDP00]
- Distributed teams perform better when members know each other [Cummings+CSCW08]
- Specific communication patterns amongst team members are critical for performance [Cataldo+CHI12]

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

Design Objectives

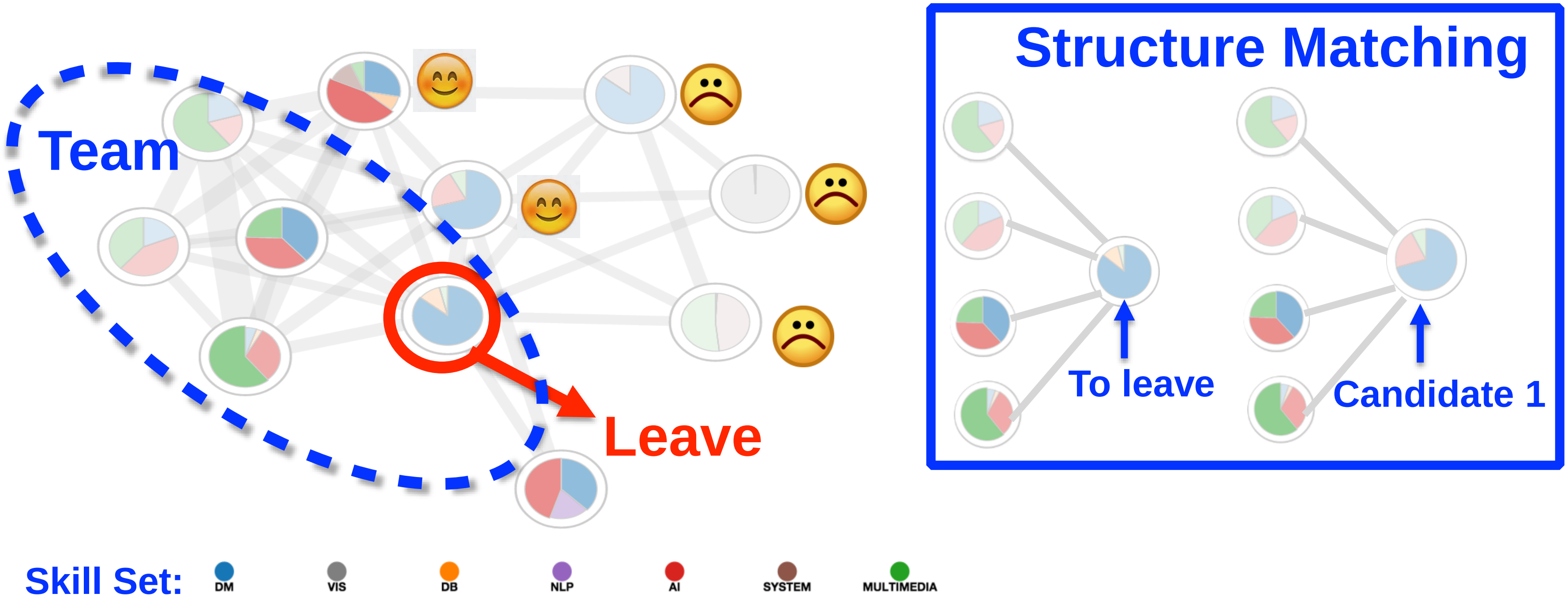
Objective 1: A good candidate should have a similar skill set



New team would have a similar skill set as the old team to continue to complete the task

Design Objectives

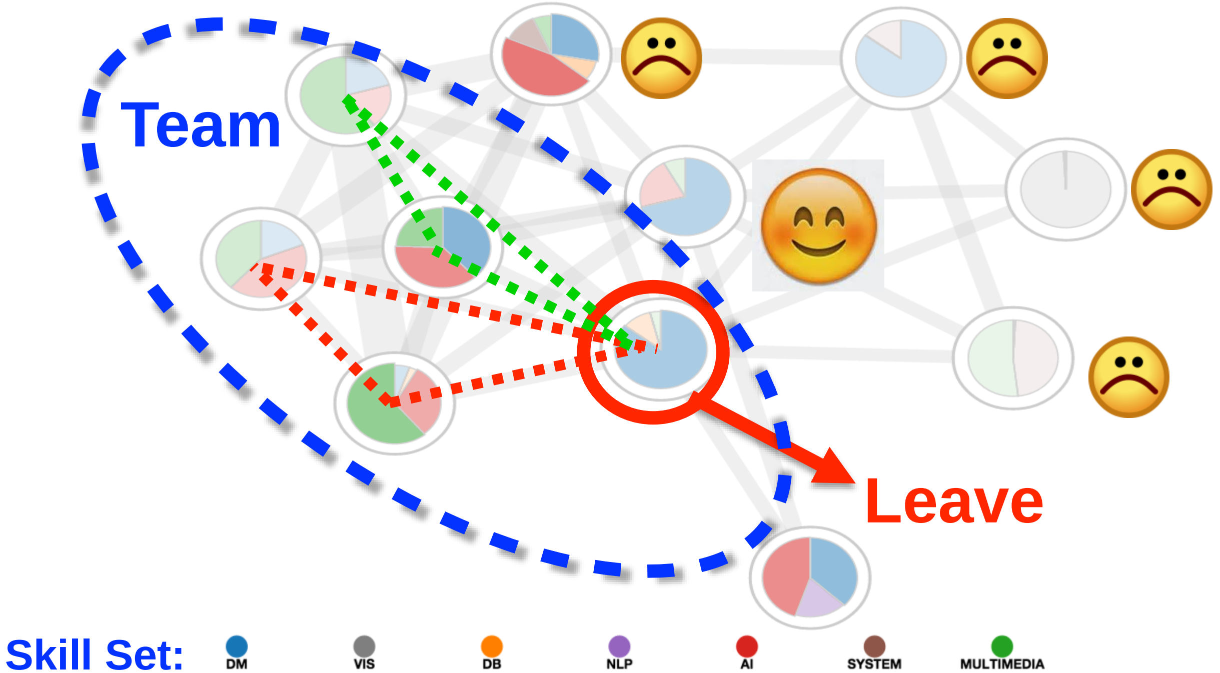
Objective 2: A good candidate should have a similar network structure



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

Design Objectives

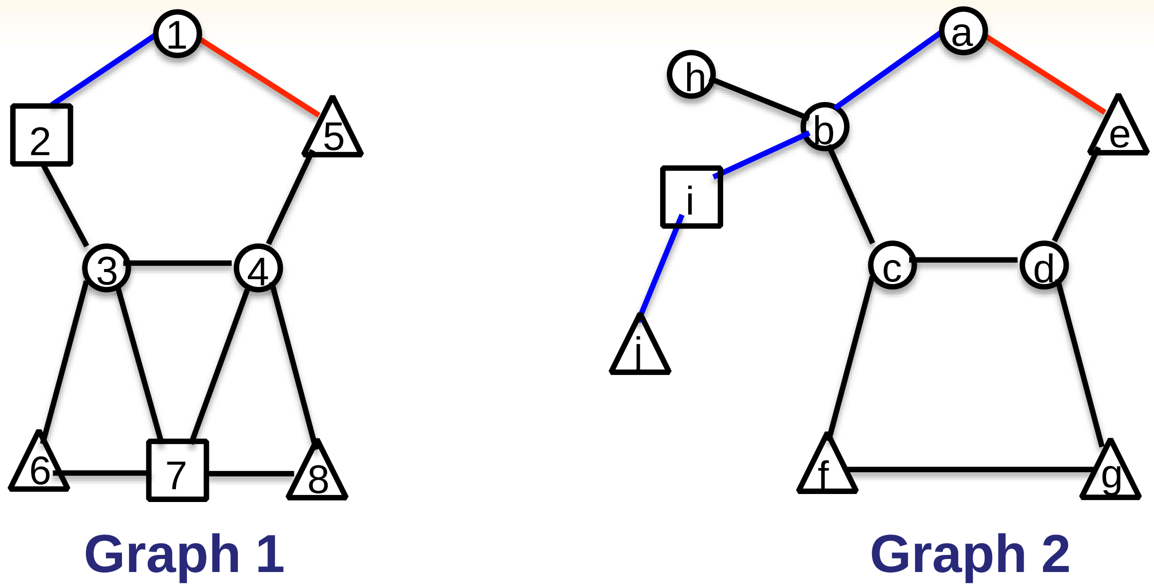
The skill and structure match should be fulfilled simultaneously!



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

- L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Replacing the Irreplaceable: Fast Algorithms for Team Member Recommendation, WWW 2015

Random Walk based Graph Kernel



Details:

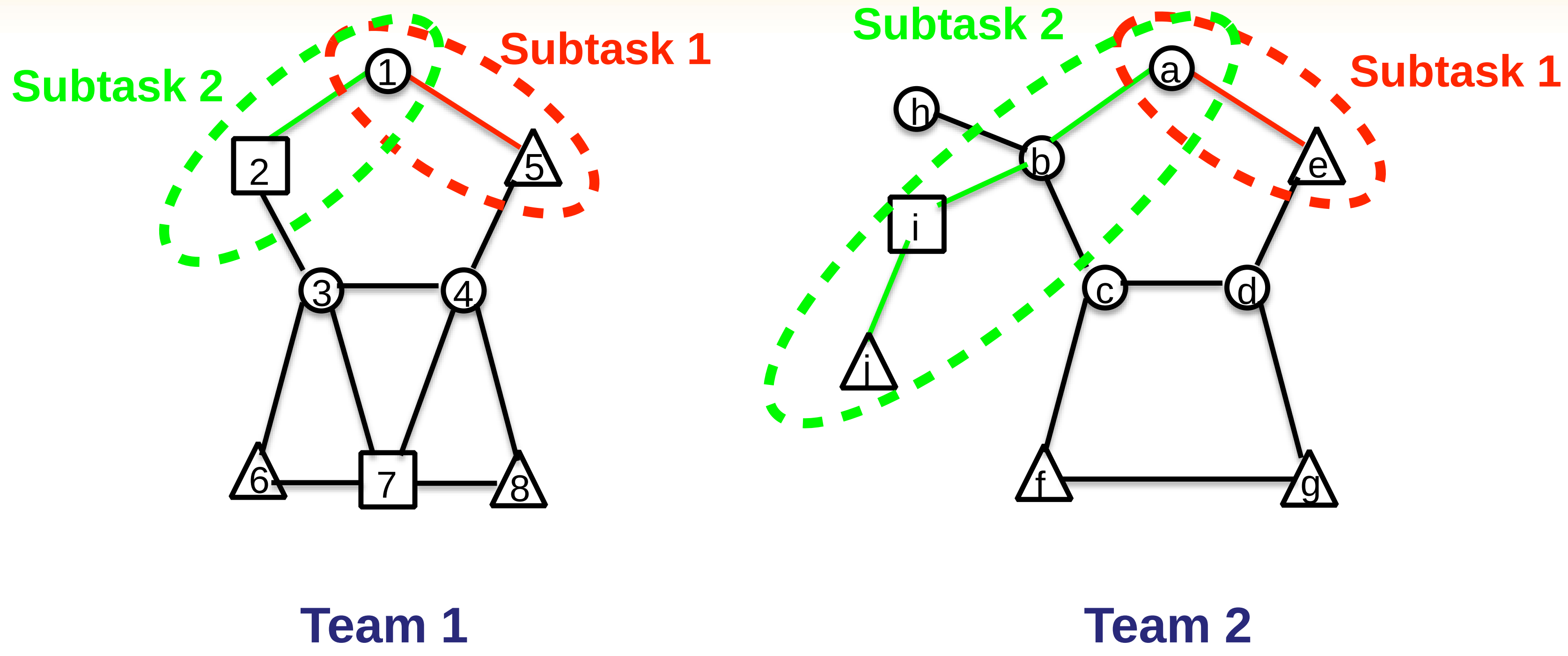
1. Compare similarity of every pair of nodes from each graph

— Eg: (1,2) vs (a, j) $\longrightarrow$ less similar

(1,5) vs (a,e) $\longrightarrow$ 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

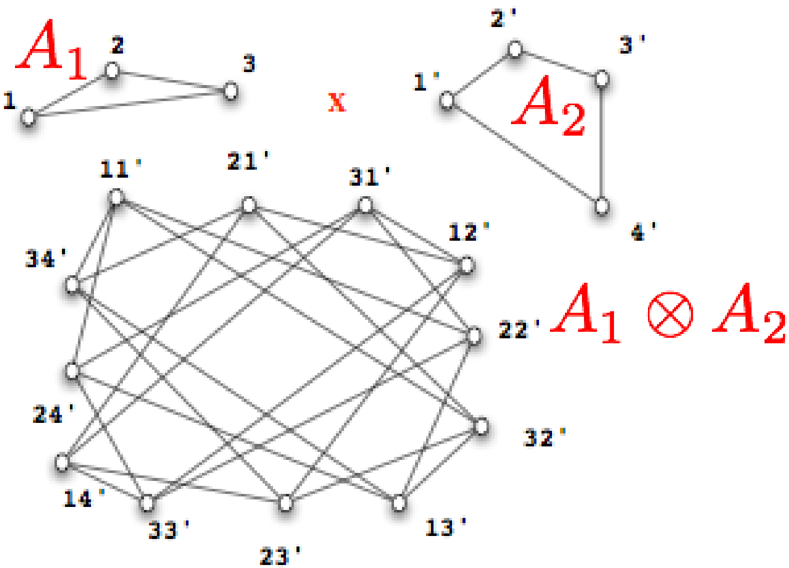


Remarks:

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

Kronecker Product Graph w/o Attribute

Graph Illustration



Matrix Description

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

$$\underbrace{A_1 \otimes A_2}_{\text{Kronecker product}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

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.

RW Graph Kernel — Formulation

Taking expectations instead of summing

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

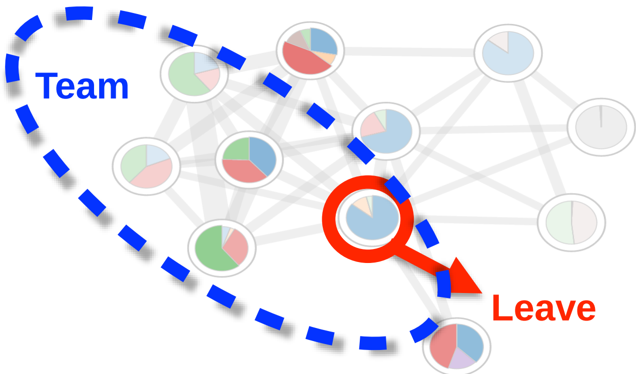
- **Computational cost (*A_x: t^2 x t^2*)**
  - *Exact methods:* [Vishwanathan+JMLR2010]
    - *$O(t^6)$* **-** Direct computation
    - *$O(t^3)$* - Sylvester equation
  - *Approx methods:* *$O(t^2 r^4 + mr + r^6)$* [Kang+SDM12]

- U. Kang, Hanghang Tong, Jimeng Sun. Fast Random Walk Graph Kernel. SDM 2012
- S. V. N. Vishwanathan, N. N. Schraudolph, I. Kondor, and K. M. Borgwardt. Graph Kernels. JMLR 2010.

TEAMREP-BASIC

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

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



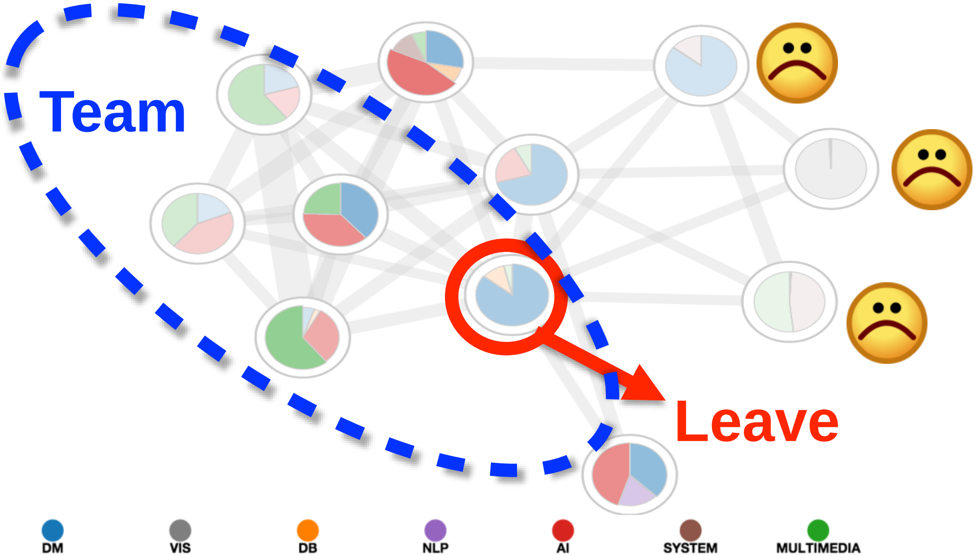
One graph kernel computation for every possible candidate

- **Challenge:** need to compute many graph kernel
  overall complexity: ***O(nt³)***
- **Questions:**
  - **Q1:** how to reduce the number of graph kernels
  - **Q2:** how to speed up the computation for each graph kernel

- L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Replacing the Irreplaceable: Fast Algorithms for Team Member Recommendation, WWW 2015

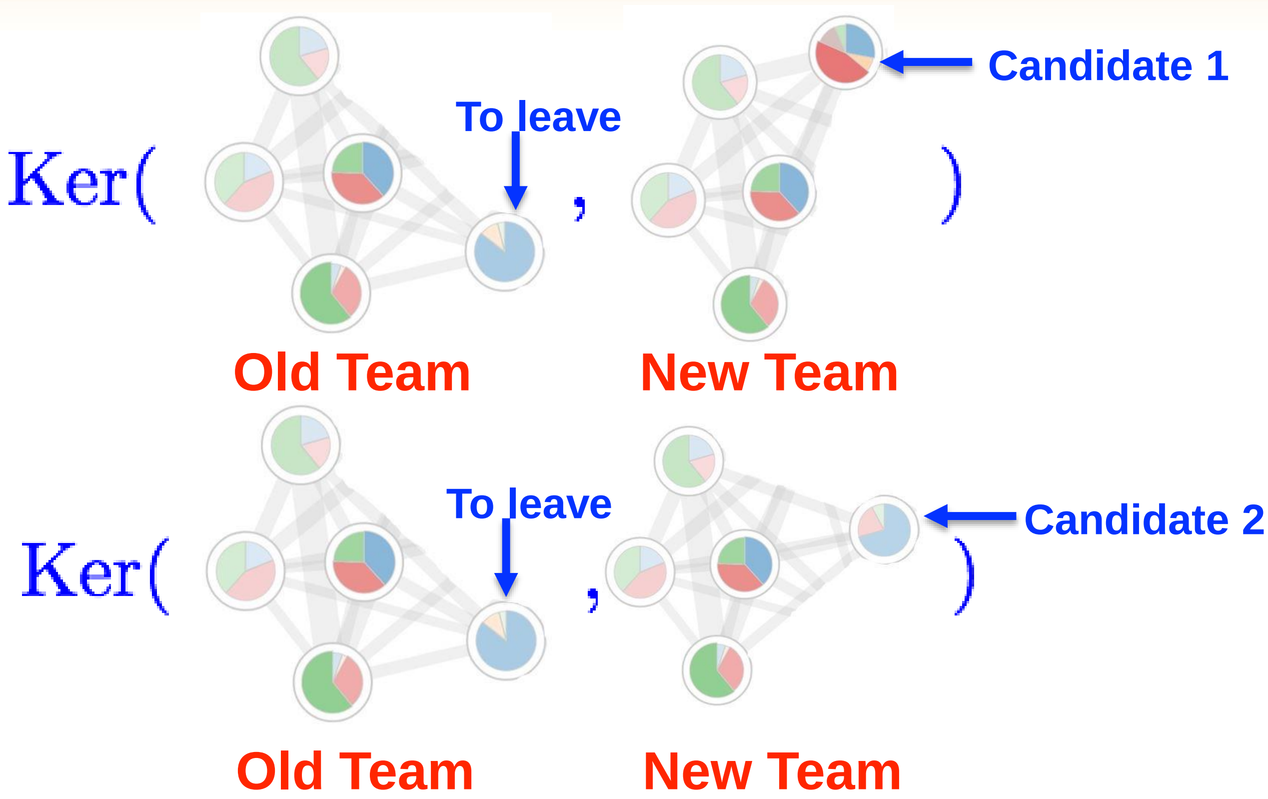
Scale-up: Candidate Filtering

***Pruning Strategy*:** Filter out all the candidates w/o 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(size of the neighborhood of T) $O(\sum_{i \in \mathcal{T}/p} d_i)$

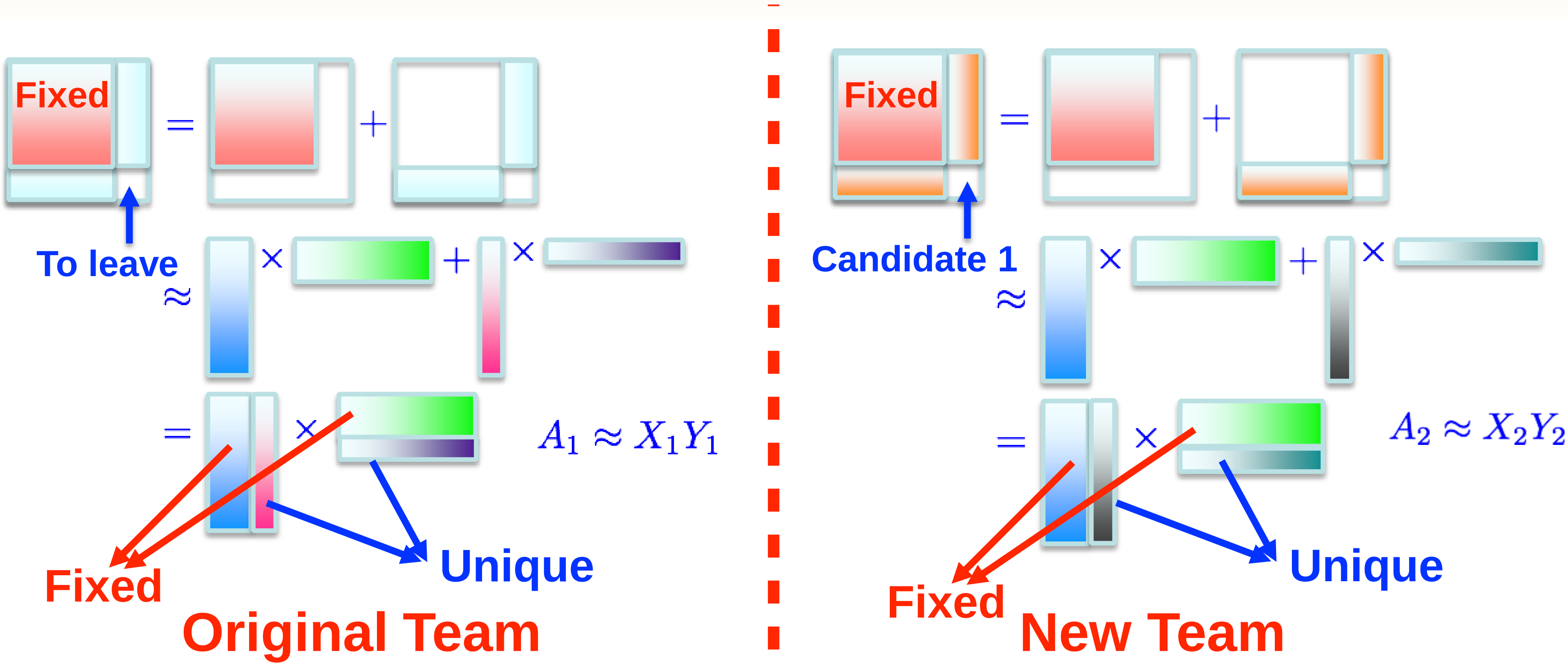
Speedup — Observation



Observation:

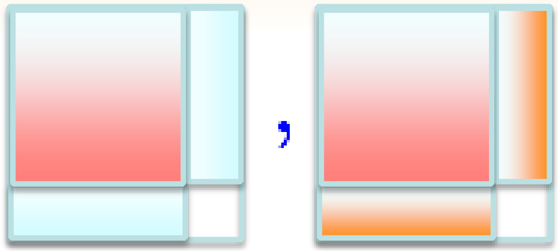
Many redundancies — *the nodes and edges within the rest team members remain the same*

Speedup — Approx Approach



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

Speedup — Approx Approach

$\mathrm{Ker}($  $)$

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

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

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

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

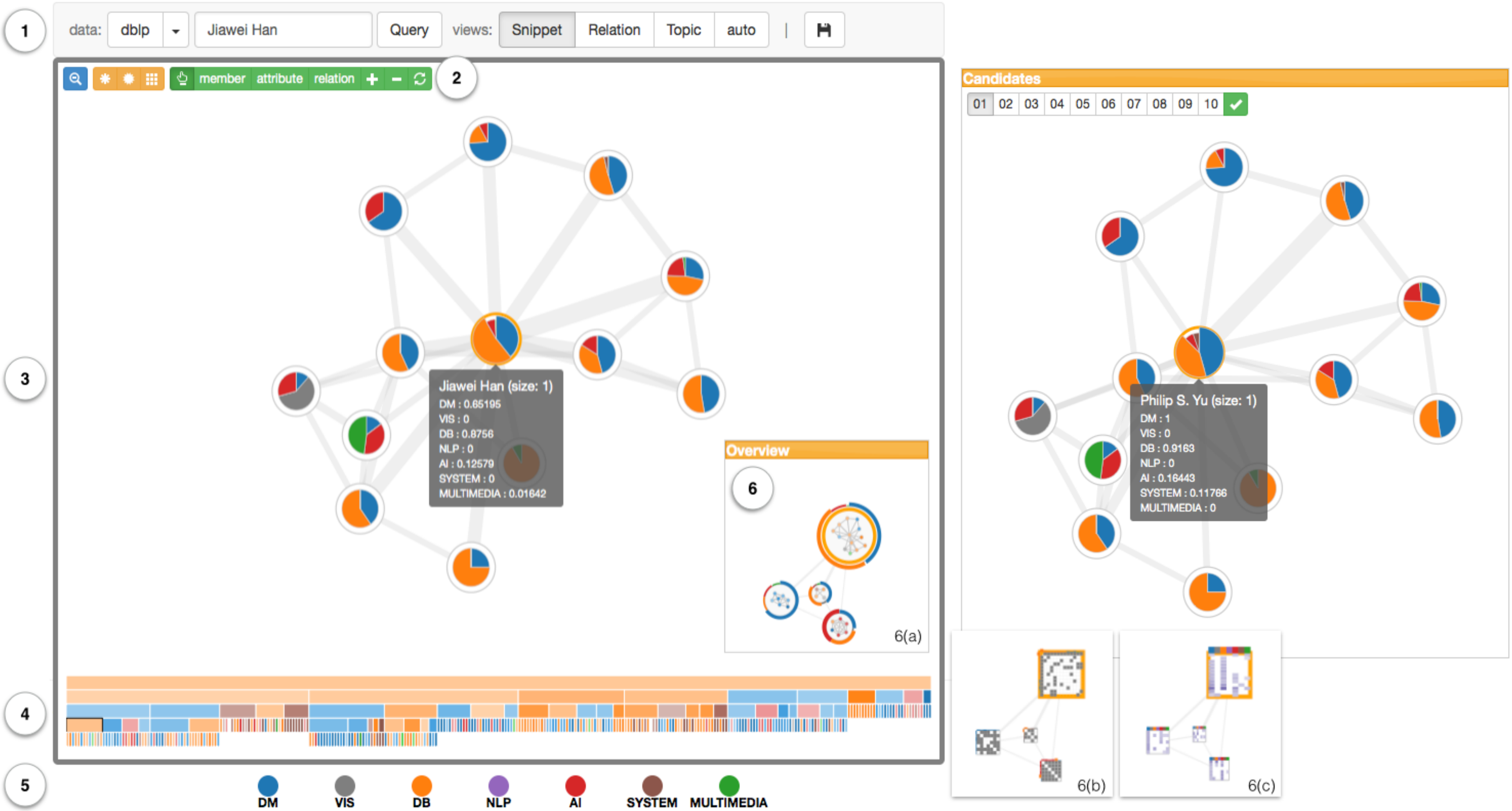
Time Complexity: $O((\sum_{i\in\mathcal{T}/p} d_i)(lt^2r + r^6))$

Original Complexity: $O(nt^3)$

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

Prototype Systems

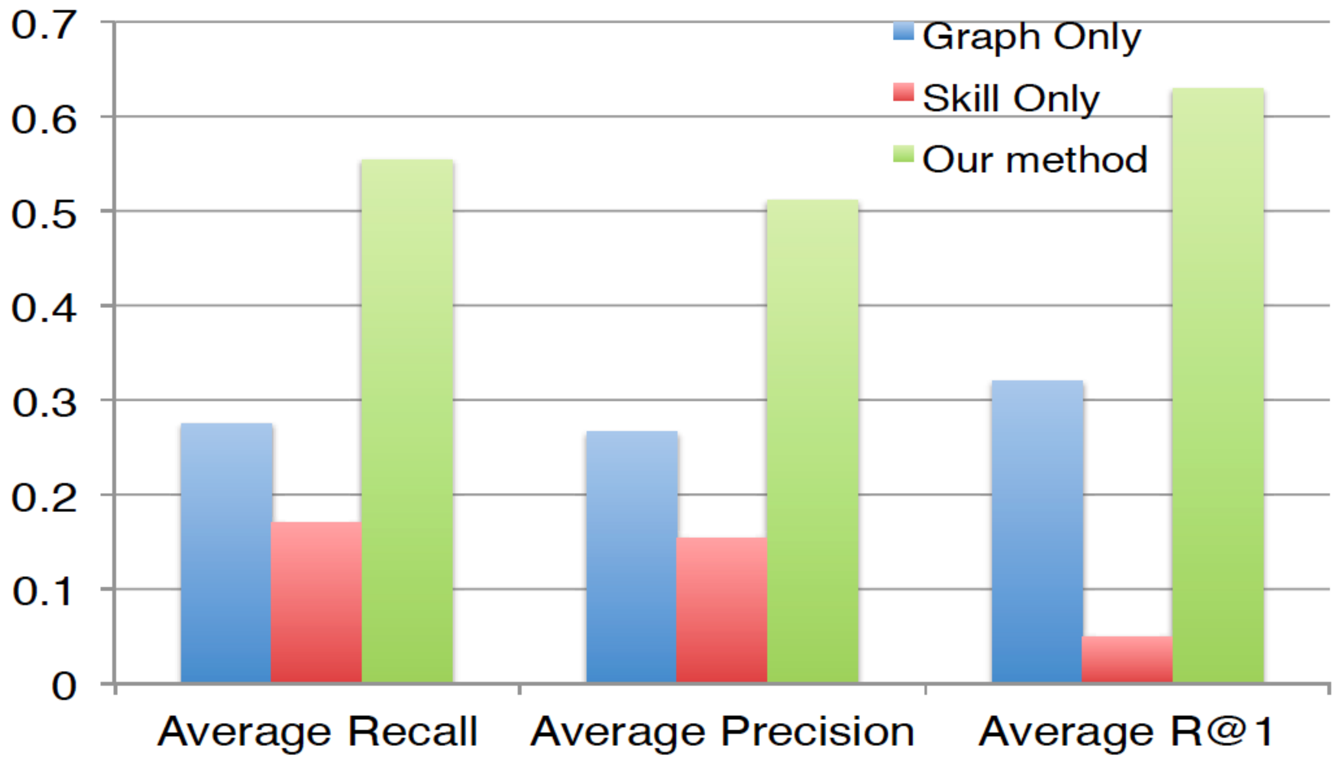
Questions
- **Q1: How effective is skill + structure?**
- **Q2**: How fast is pruning?
- **Q3:** How fast is proposed solution?
- **Q4**: How is the scalability?



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

- Nan Cao, Yu-Ru Lin, Liangyue Li, Hanghang Tong."g-Miner: Interactive Visual Group Mining on Multivariate Graphs", ACM **CHI** 2015.

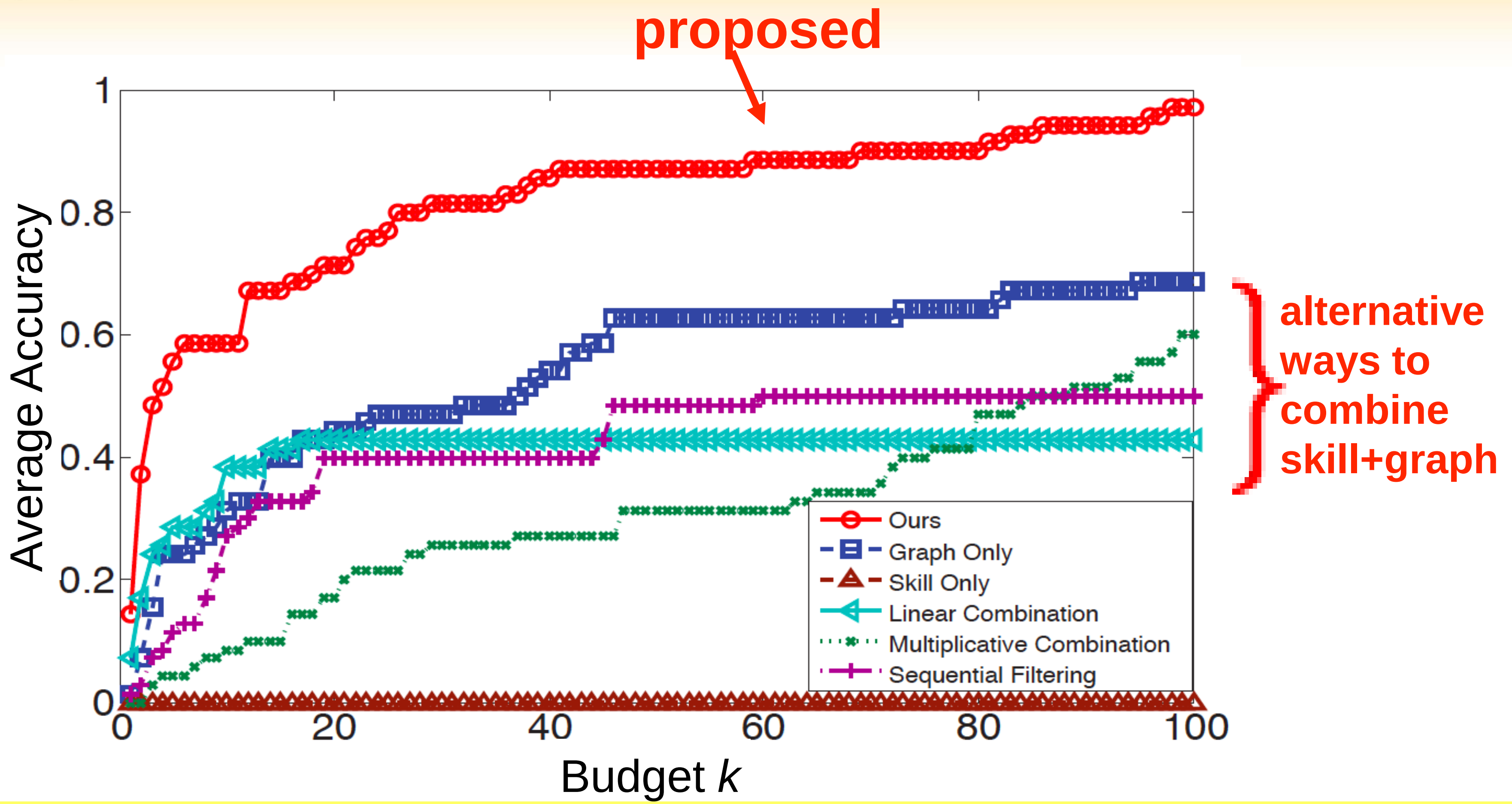
User Studies



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

- L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Replacing the Irreplaceable: Fast Algorithms for Team Member Recommendation, WWW 2015

Application in Author Alias Prediction

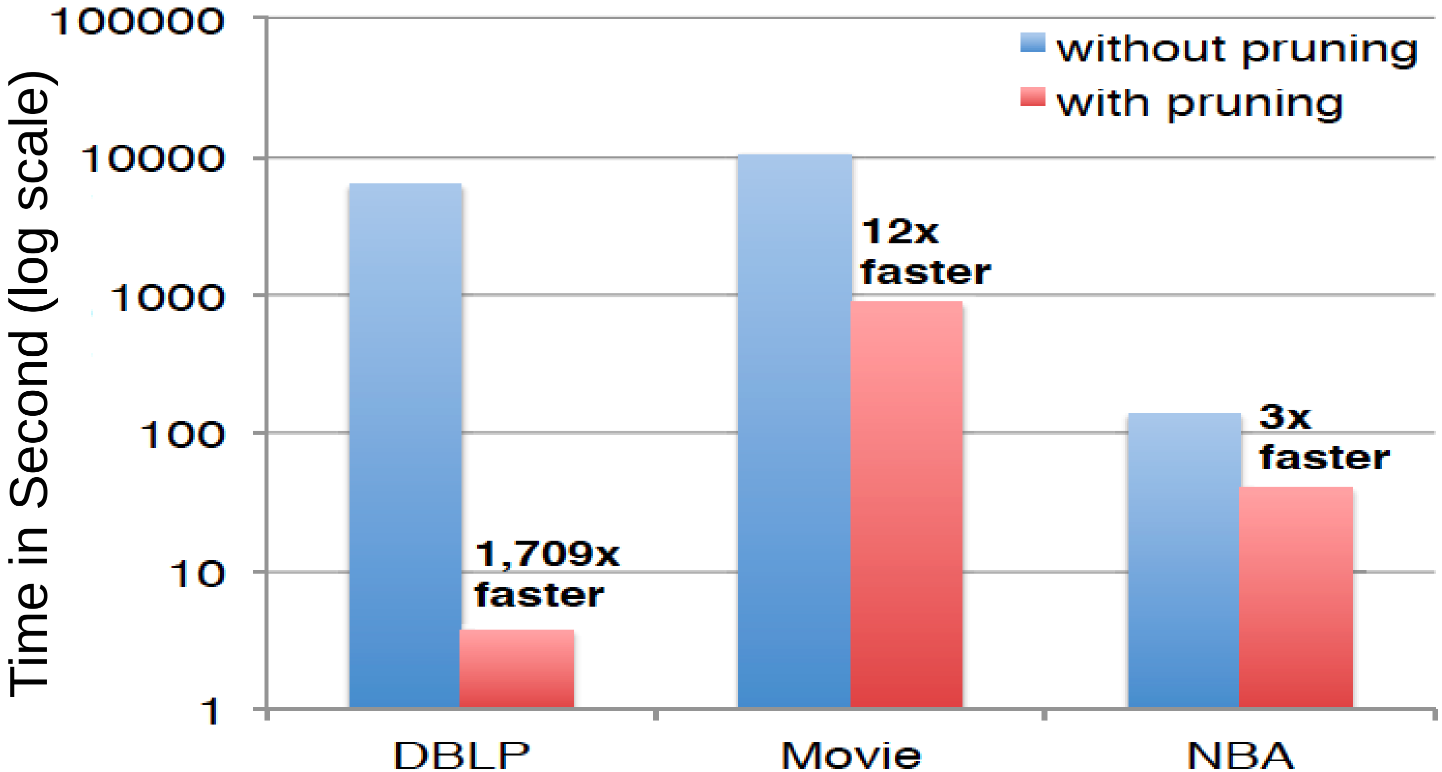


Our method achieves the highest accuracy

Author Alias: *Alexander J. Smola* vs. *Alex J. Smola*

Speed-up by Pruning

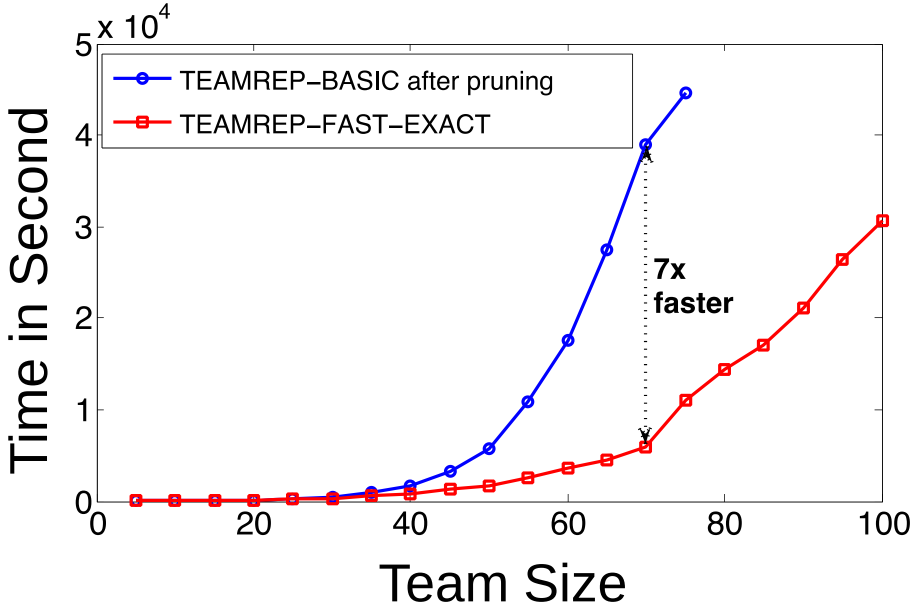
Questions
- **Q1**: How effective is skill + structure?
- **Q2: How fast is pruning?**
- **Q3:** How fast is proposed solution?
- **Q4**: How is the scalability?



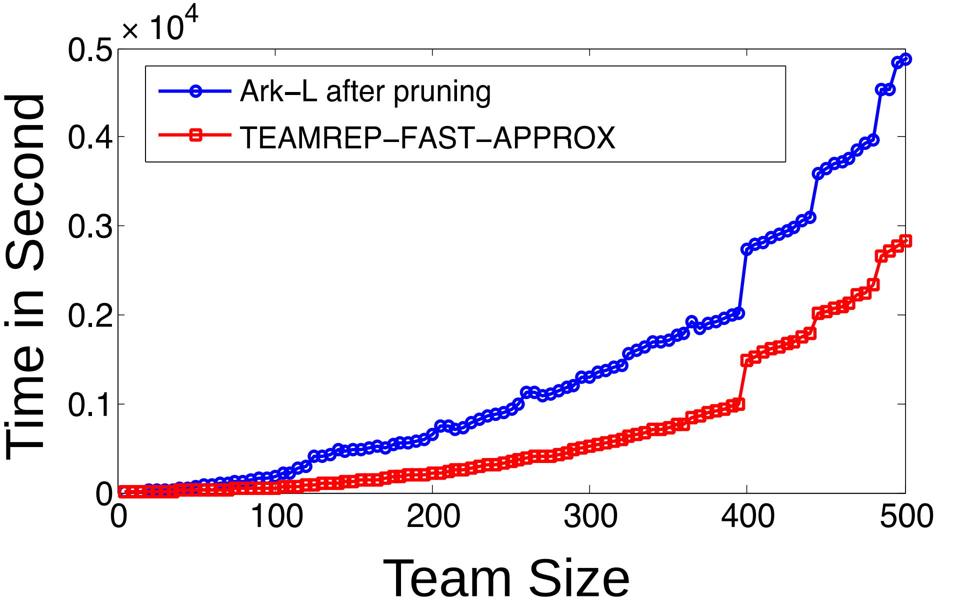
Pruning has dramatic speed improvement

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



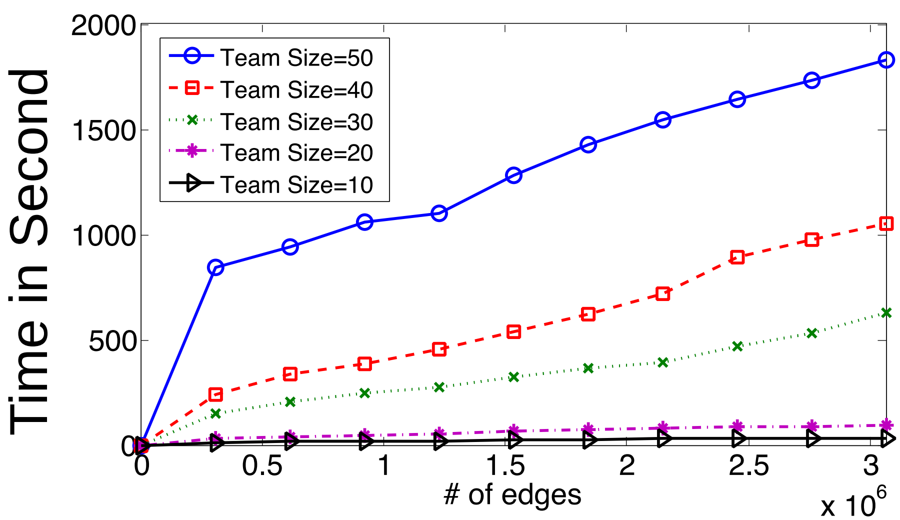
Approximate methods

Exploiting redundancy leads to additional speed-up!

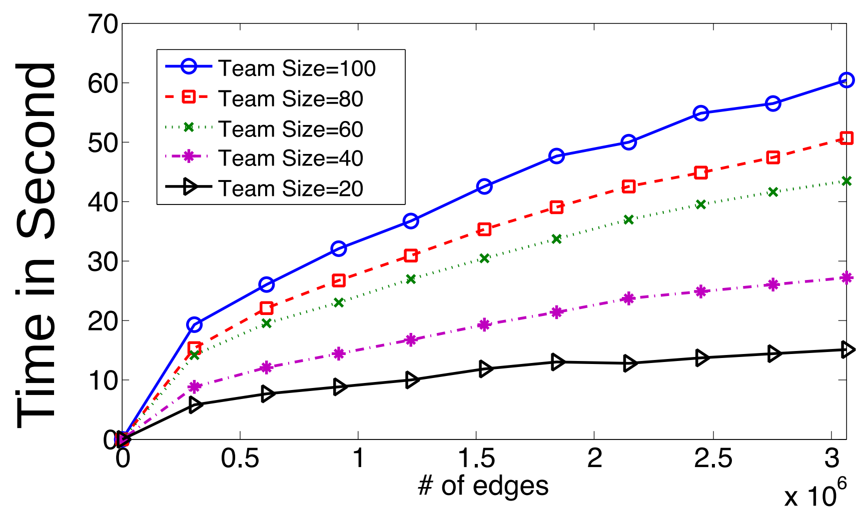
Scalability

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

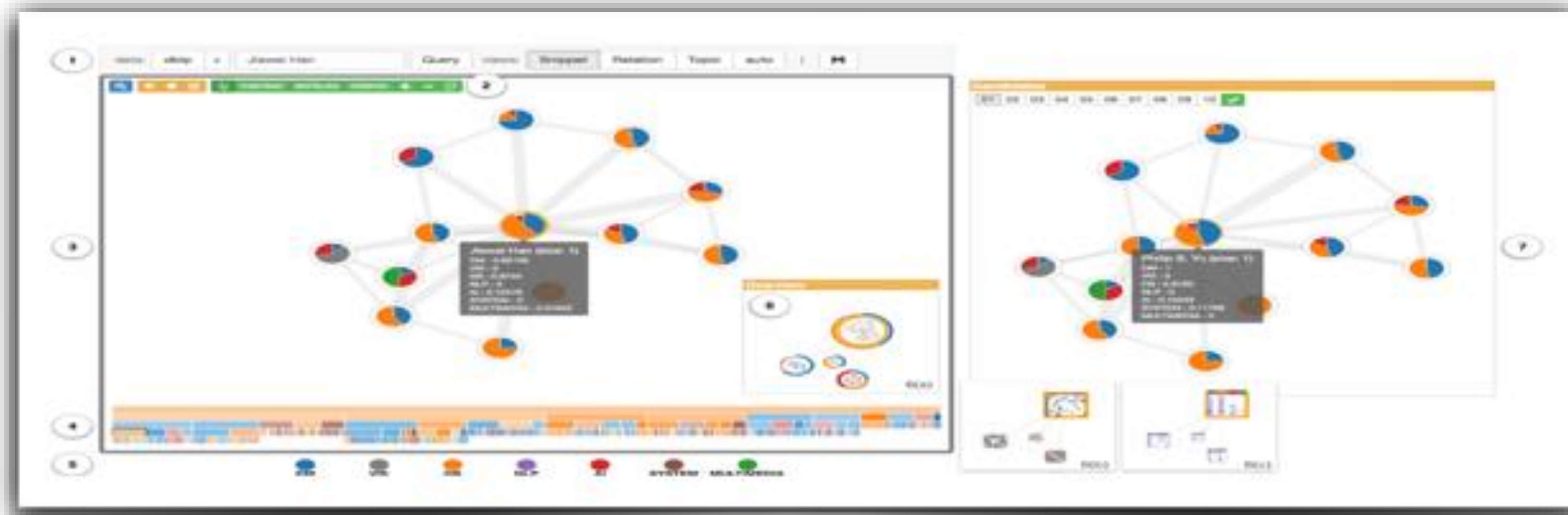


TEAMREP-FAST-APPROX

Our fast solutions scale sub-linearly

Team Member Replacement - Summary

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



- L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Replacing the Irreplaceable: Fast Algorithms for Team Member Recommendation, WWW 2015
- N. Cao, Y.-R. Lin, L. Li, H. Tong: g-Miner: Interactive Visual Group Mining on Multivariate Graphs, ACM CHI 2015

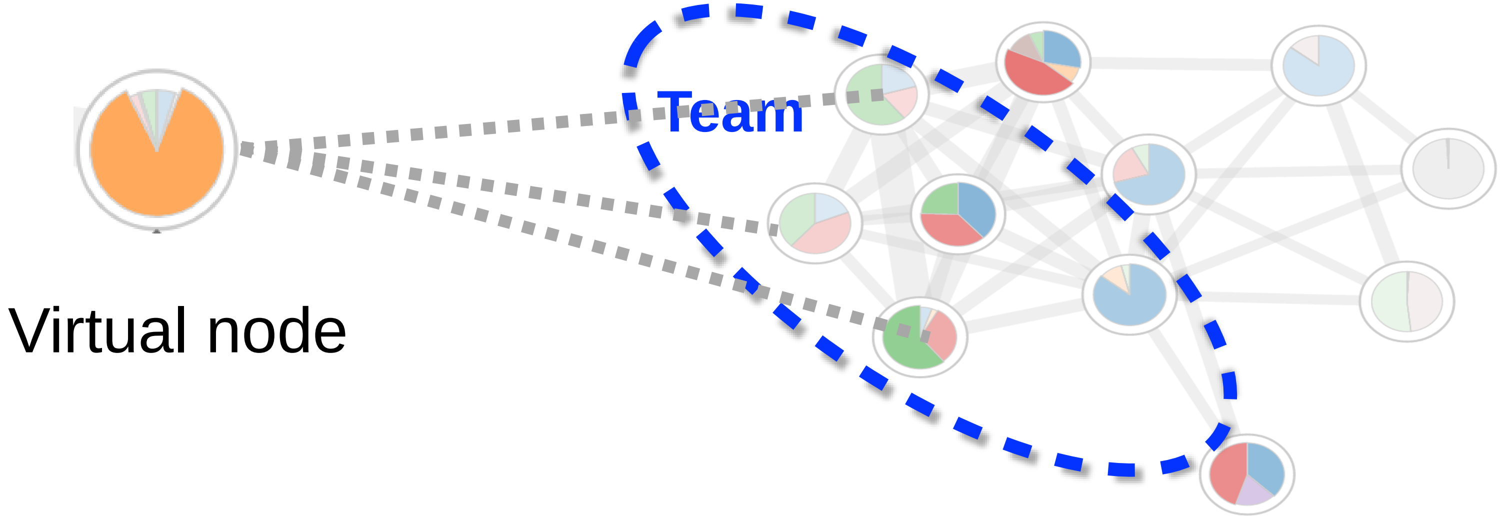
Beyond Team Member Replacement

- **Team Shrinkage**
  - If we need to reduce the size of an existing team (e.g., for the purpose of cost reduction), who shall leave the team?
- **Team Expansion**
  - If the team leader perceives the need to enhance certain expertise of the entire team, who shall we bring into the team?
- **Team Conflict Resolution**
  - If the team leader sees a conflict between certain team members, how shall we resolve it?

Key Idea: Solve all these team enhancement scenarios by team member replacement !

L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Enhancing Team Composition in Professional Networks: Problem Definitions and Fast Solutions, TKDE, 2016

Team Expansion



L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Enhancing Team Composition in Professional Networks: Problem Definitions and Fast Solutions, TKDE, 2016

Team Expansion – Case Study

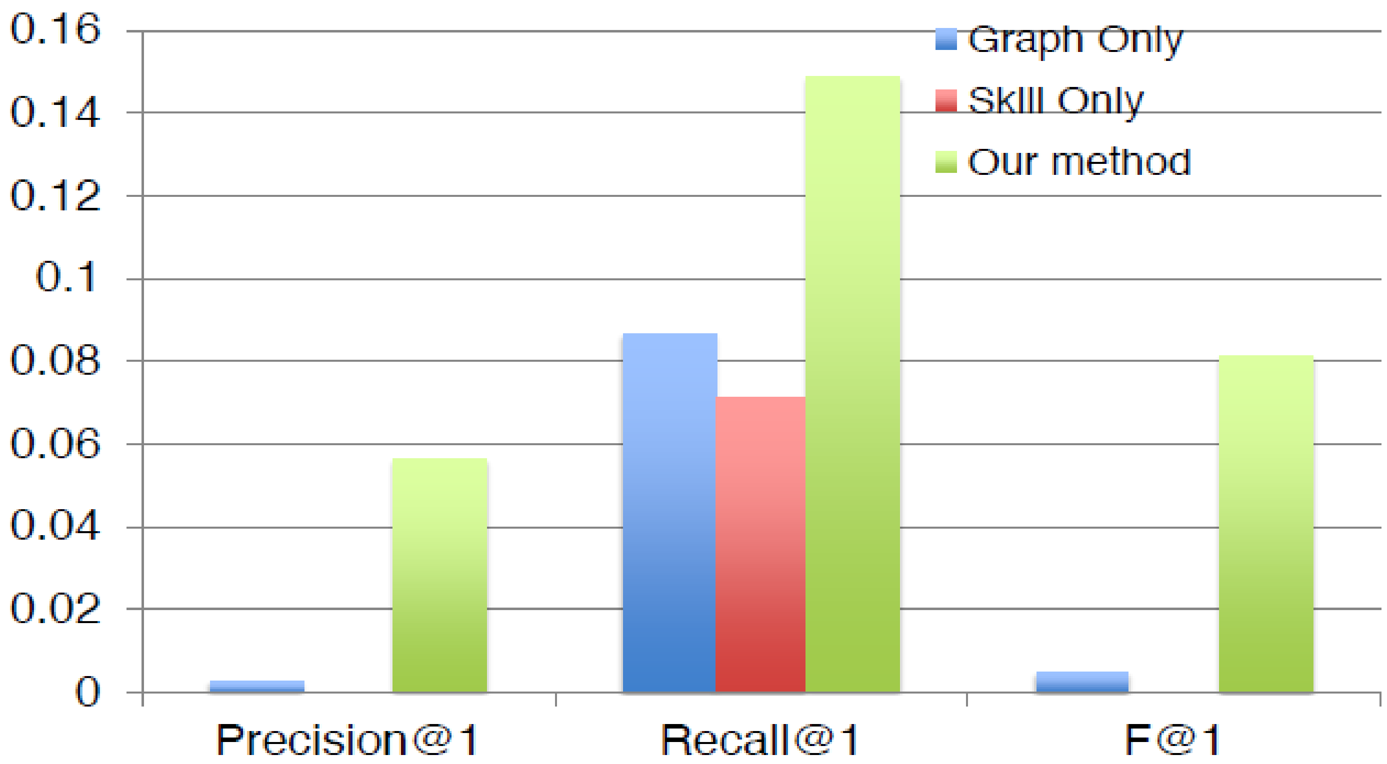
- Expand the organizing committee of KDD 2013 by hiring some
  - strong expertise in AI
  - collaborated with as many existing committee members as possible
- Top five candidate:
  - *Qiang Yang, Zoubin Ghahramani, Eric Horvits, Thomas Dietteirich, Raymond J. Mooney*

L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Enhancing Team Composition in Professional Networks: Problem Definitions and Fast Solutions, TKDE, 2016

Team Shrinkage

- Select teams with over 10 members and manually inject a “noisy” individual
  - Connect to all team members w/ random weights
  - Random skill vector
  - “best” candidate to leave the team

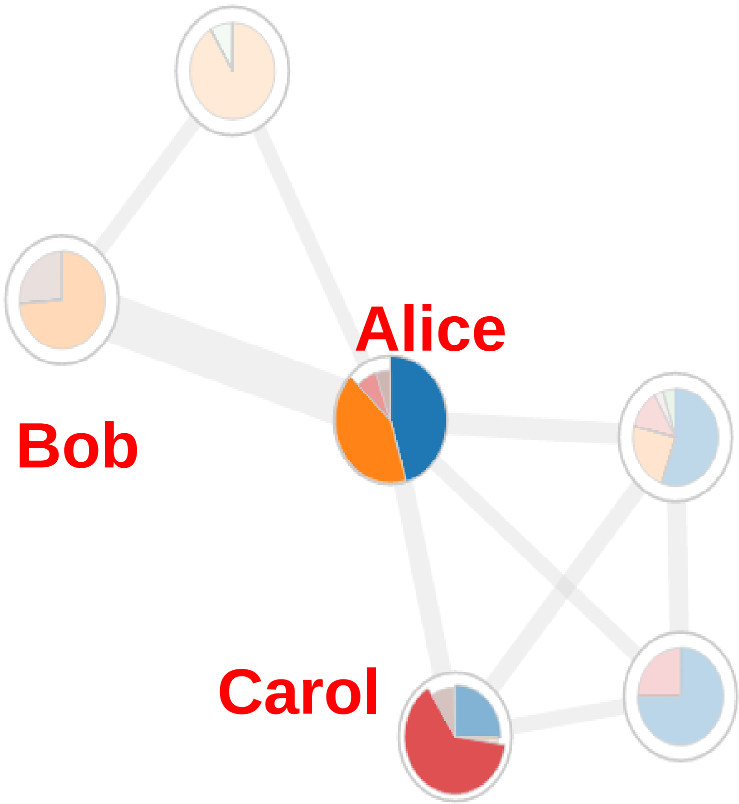
Team Shrinkage -- Results



L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Enhancing Team Composition in Professional Networks: Problem Definitions and Fast Solutions, TKDE, 2016

Team Conflict Resolution

- E.g., Bob has a conflict with Alice
  - Replace either
  - Remove either



Roadmap

- Motivations and Background
- Part I: Team Performance Characterization
- Part II: Team Performance Prediction
- Part III: Team Performance Optimization
- Part IV: Open Challenges
- Demo

Open Challenges

- Prediction Explanation
- Optimization Explanation
- Multiple Teams Optimization

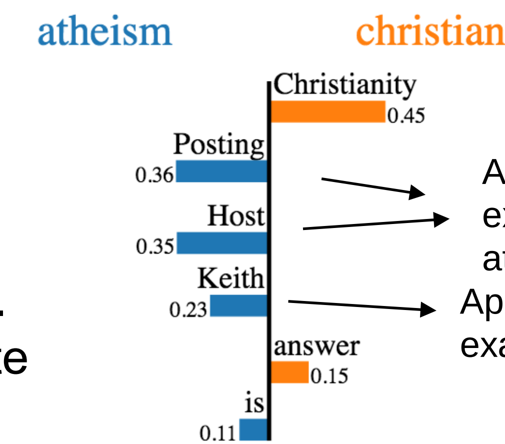
Prediction Explanation

- **Observations**:
  - Predictive models are mostly black-box or too complicated to understand reasons behind
  - Interpretable models cannot achieve satisfactory prediction accuracy
- **Goal**:
  - Provide explanations to performance prediction
  - Assess trust of models

What an explanation looks like

From: Keith Richards
Subject: Christianity is the answer
NTTP-Posting-Host: x.x.com

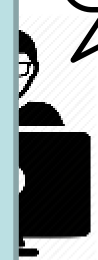
I think Christianity is the one true religion.
If you'd like to know more, send me a note



Appear in 21% of training examples, almost always in atheism

Appears in 11% of training examples, **always** in atheism

Why did this happen? How do I fix it?



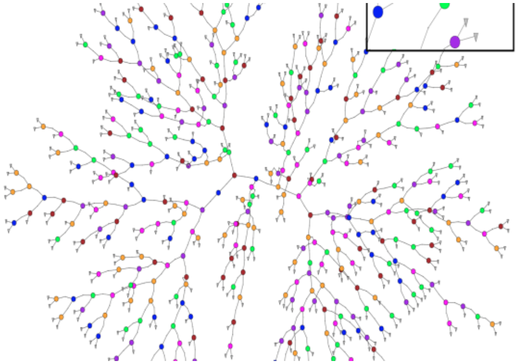
➔ Will not generalize
➔ Don't trust this model!

Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

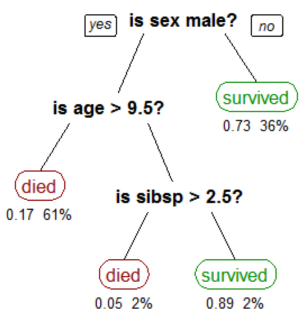
Three must-haves for a good explanation

Interpretable

- Humans can easily interpret reasoning



Definitely not interpretable



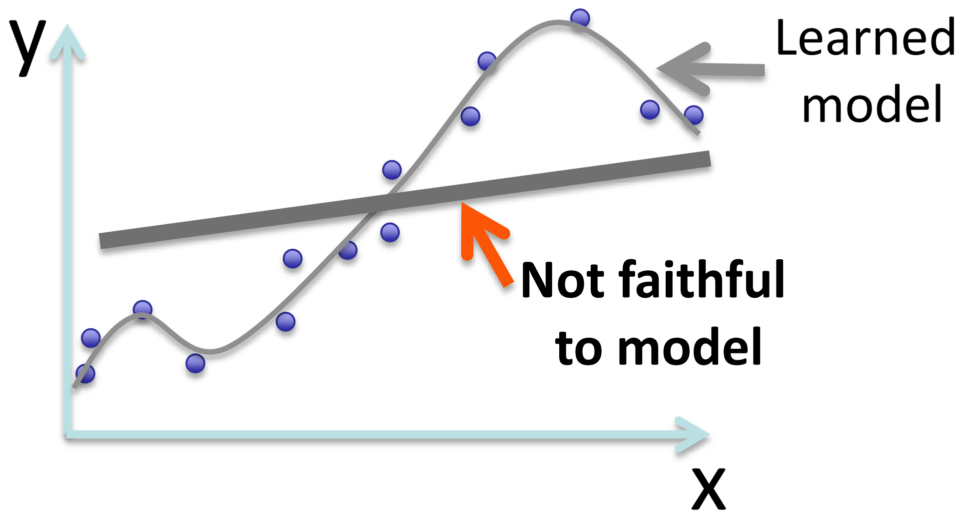
Potentially interpretable

Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.

Slides credit: Marco Tulio Ribeiro.

Three must-haves for a good explanation

Interpretable	• Humans can easily interpret reasoning
Faithful	• Describes how this model actually behaves



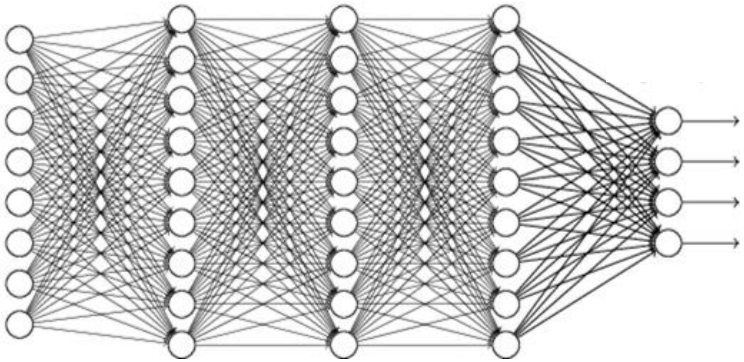
Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.

Slides credit: Marco Tulio Ribeiro.

Three must-haves for a good explanation

Interpretable	• Humans can easily interpret reasoning
Faithful	• Describes how this model actually behaves
Model agnostic	• Can be used for *any* ML model

Can explain this mess ☺



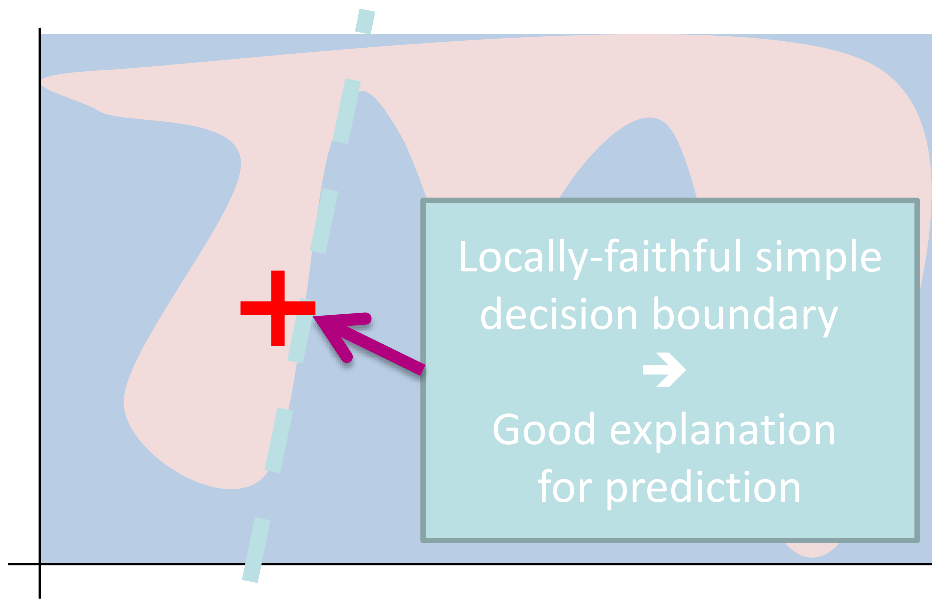
Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

LIME – Key Ideas

1. Pick a model class interpretable by humans
   - Not globally faithful… ☹

Line,
shallow decision tree,
sparse features, …

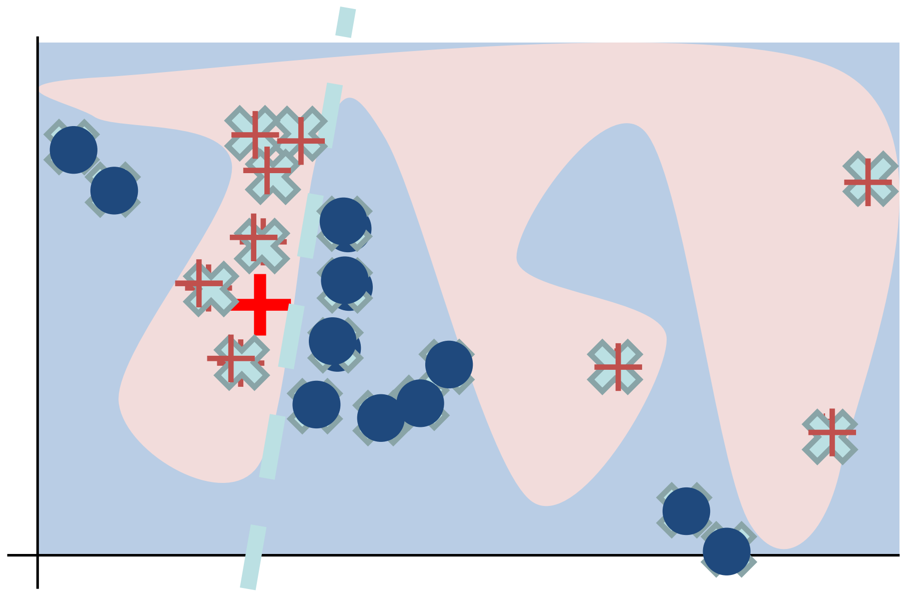
2. Locally approximate global (blackbox) model
   - Simple model globally bad, but locally good



Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

Using LIME to explain a complex model's prediction for input x_i

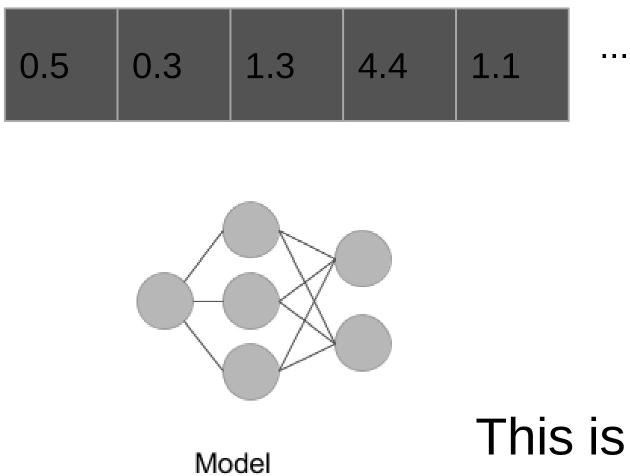
1. Sample points around x_i
2. Use complex model to predict labels for each sample
3. Weigh samples according to distance to x_i
4. Learn new simple model on weighted samples
5. Use simple model to explain



Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

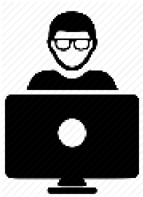
Interpretable representations

x (embeddings)



x' (words)

This is a horrible movie.

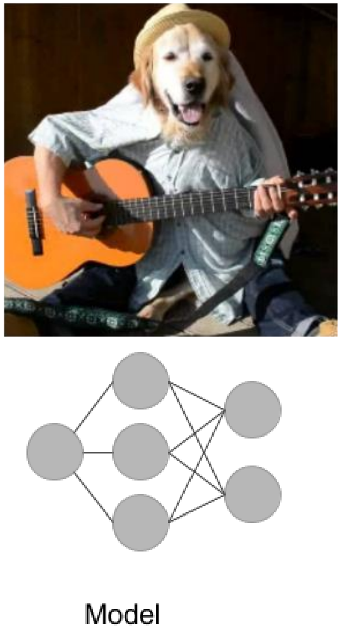


This is what we perturb, and this is what we use in the interpretable approximation

Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

Interpretable representation: images

x (3 color channels / pixel)



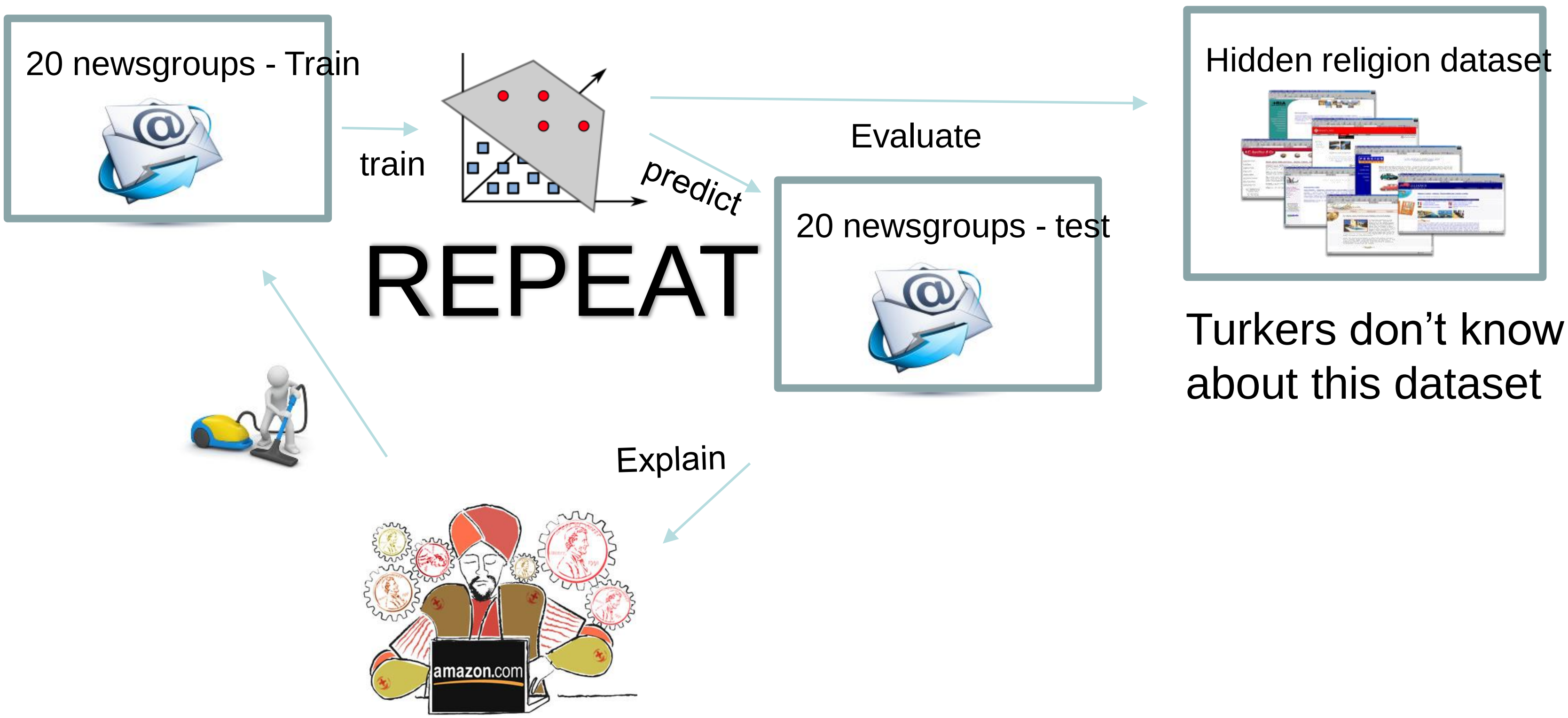
x' (contiguous superpixels)

Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

Explaining Google's Inception NN

Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

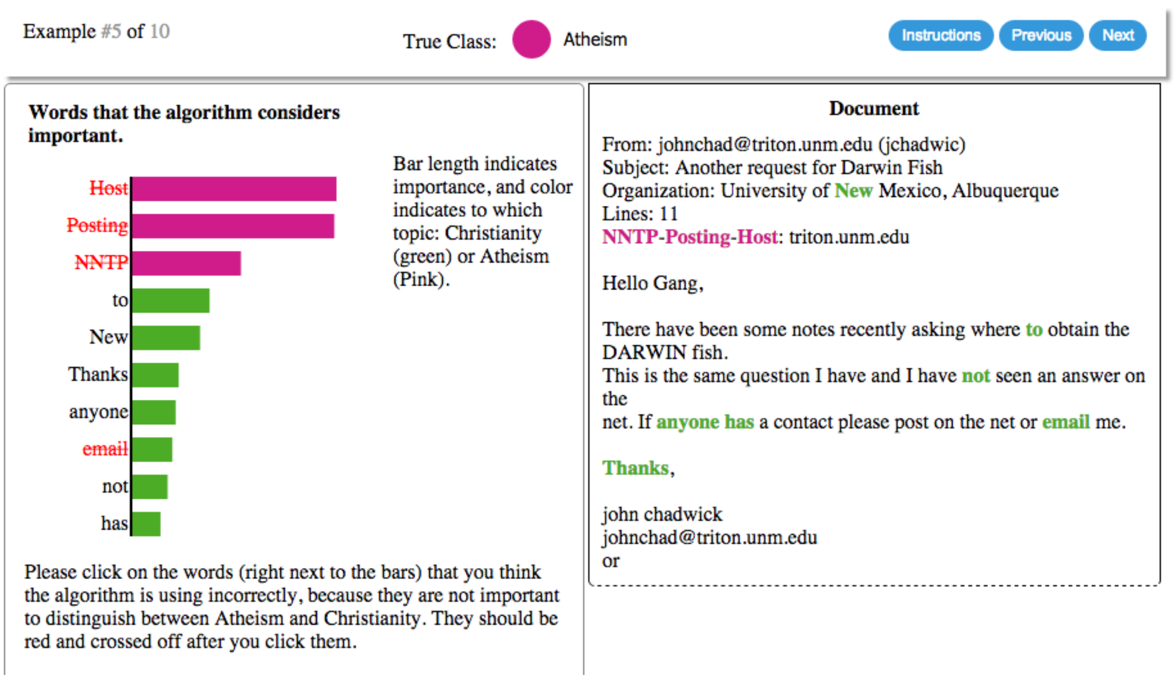
Fixing bad classifiers



Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

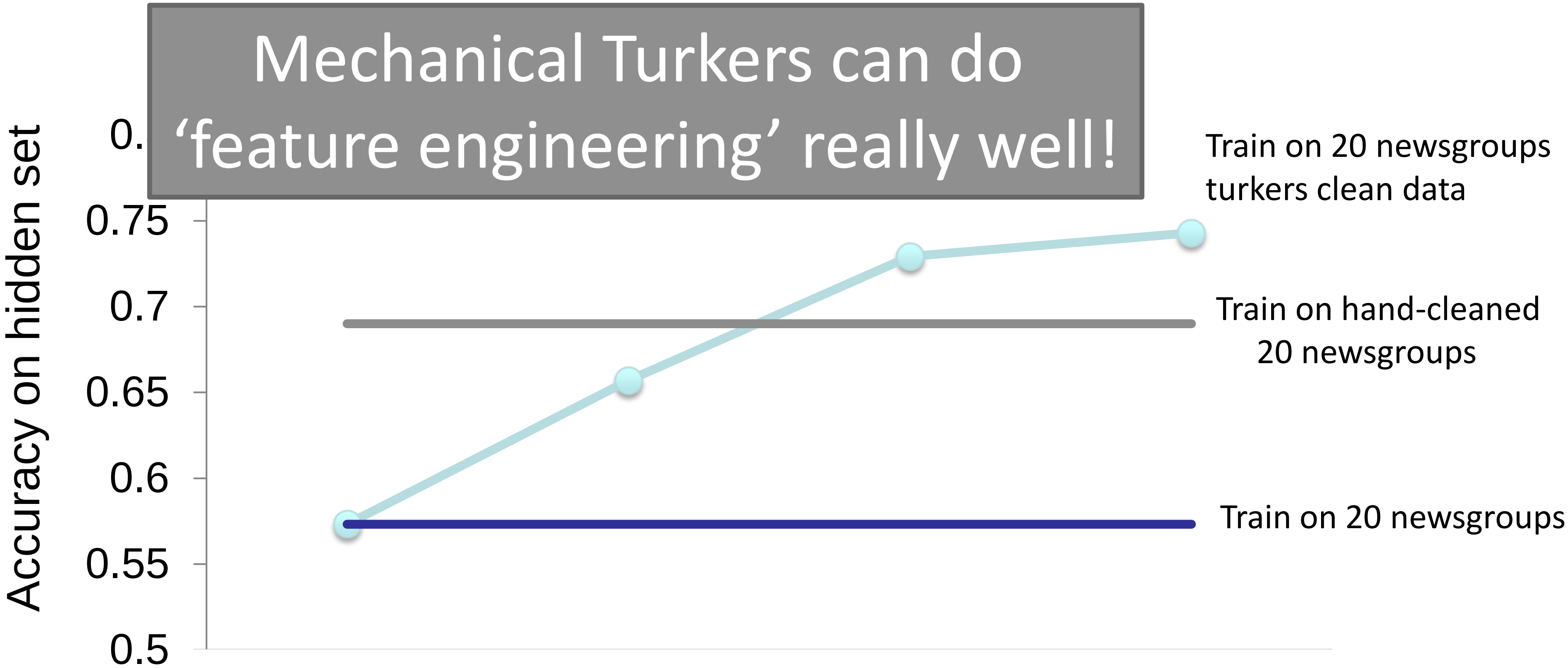
Fixing bad classifiers

- Turkers click on 'useless' words for the task in each round



Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

Fixing bad classifiers



Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
Slides credit: Marco Tulio Ribeiro.

Explain through examples

- Consider the following learning task

Training points: $z_1, \ldots, z_n$

Loss: $L(z_i, \theta)$

Params: $\hat{\theta} \triangleq \arg\min_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} L(z_i, \theta)$

- Upweighting a training example

$$\mathcal{I}_{\text{up,params}}(z) \triangleq \left.\frac{d\hat{\theta}_{\epsilon,z}}{d\epsilon}\right|_{\epsilon=0} = -H_{\hat{\theta}}^{-1} \nabla_\theta L(z, \hat{\theta})$$

Koh PW and Liang Percy Liang. Understanding black-box predictions via influence functions. ICML, 2017.

Explain through examples

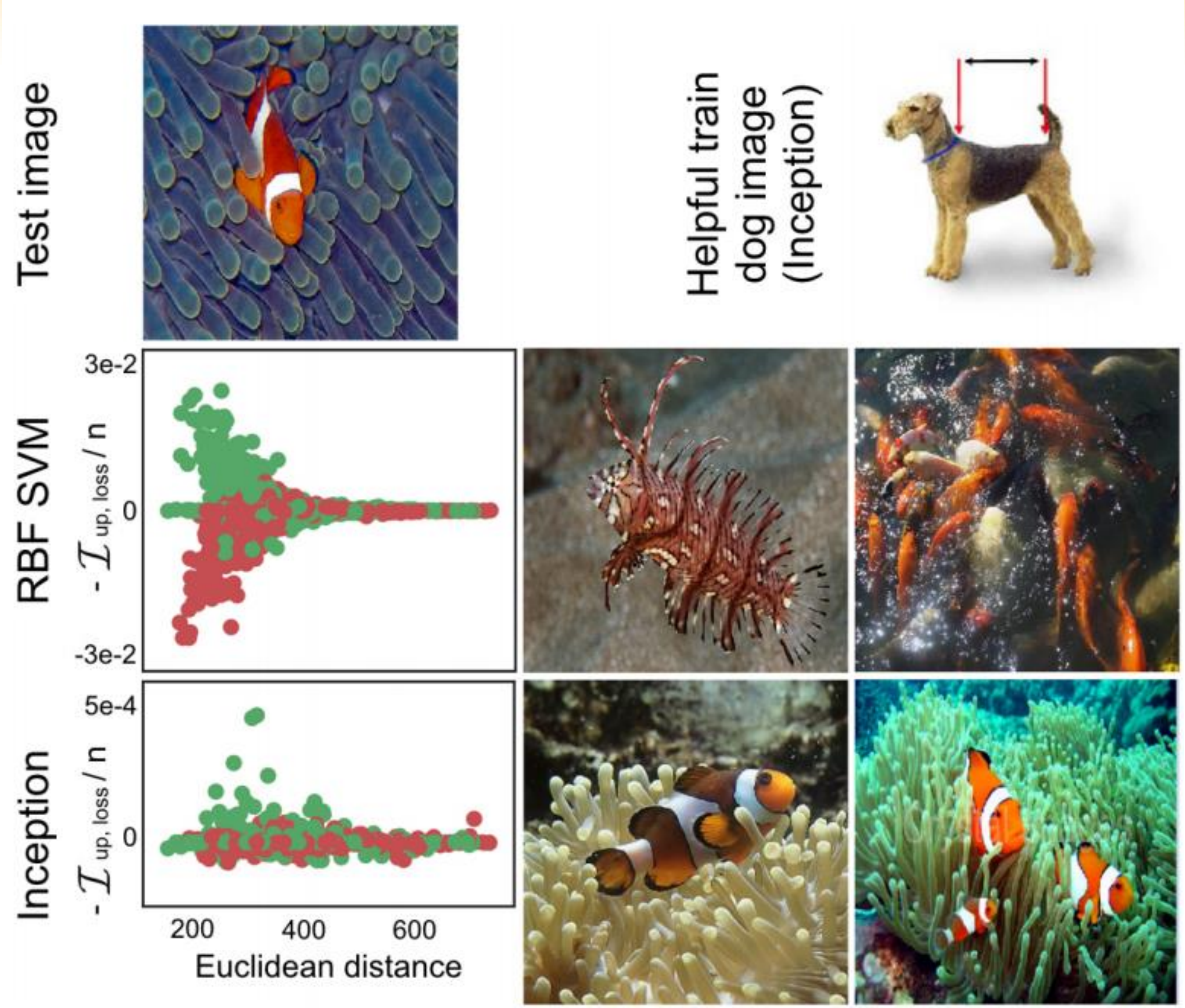
- Effect on the loss of a test example

$$\mathcal{I}_{\text{up,loss}}(z, z_{\text{test}}) \triangleq \left.\frac{dL(z_{\text{test}}, \hat{\theta}_{\epsilon,z})}{d\epsilon}\right|_{\epsilon=0}$$

$$= -\nabla_\theta L(z_{\text{test}}, \hat{\theta})^\top H_{\hat{\theta}}^{-1} \nabla_\theta L(z, \hat{\theta})$$

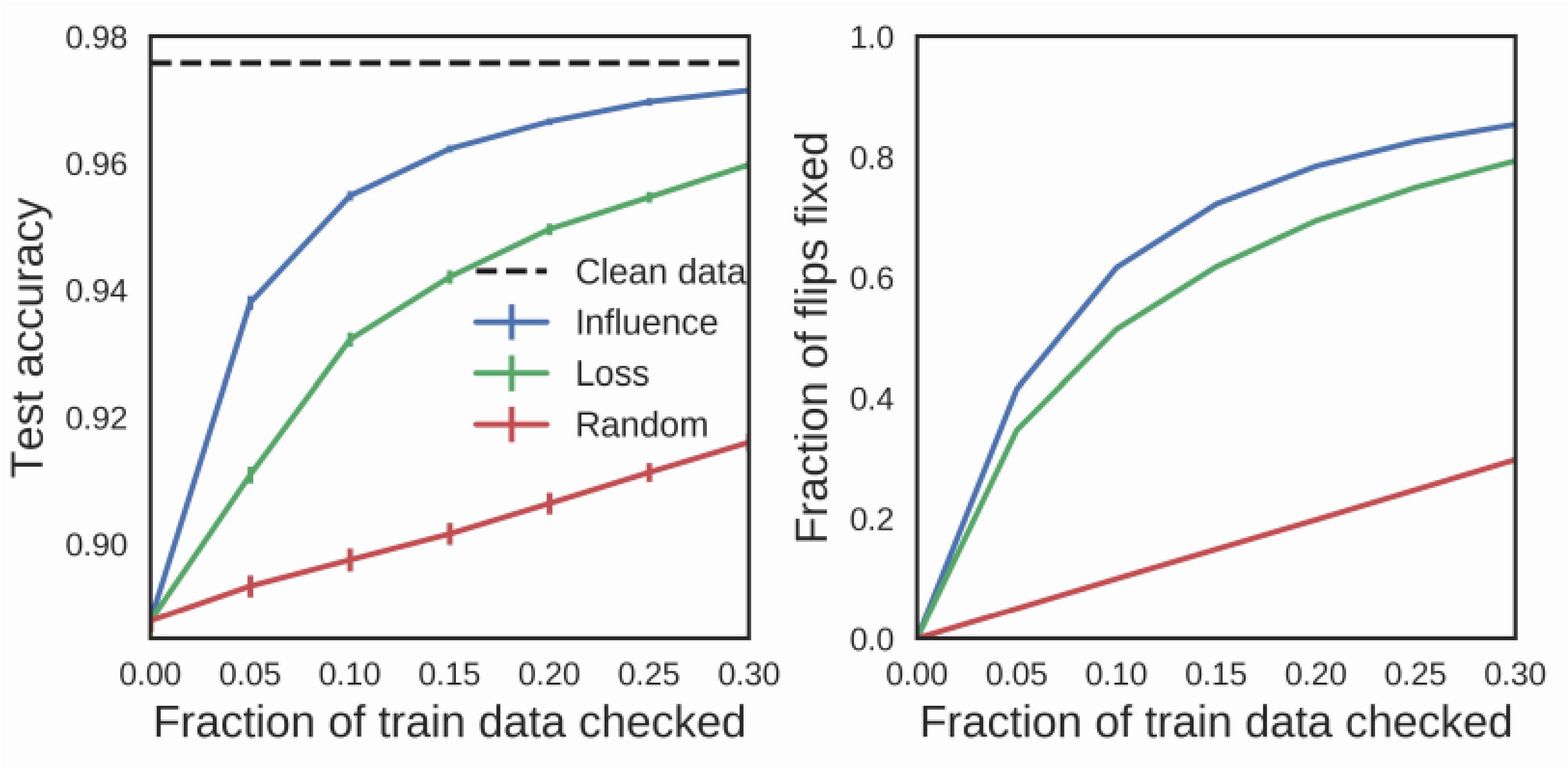
Koh PW and Liang Percy Liang. Understanding black-box predictions via influence functions. ICML, 2017.

Understanding Model Behavior



Koh PW and Liang Percy Liang. Understanding black-box predictions via influence functions. ICML, 2017.

Fixing Mislabeled Examples



Koh PW and Liang Percy Liang. Understanding black-box predictions via influence functions. ICML, 2017.

Optimization Explanation

- **Observation:** existing work focus on recommending candidates for optimization
- **Goal**: provide explanations for team optimization algorithms
  - Convince the manager to make appropriate decisions
  - Example explanations for replacement
    - The candidate also participates in the key subtasks that the person leaving is involved in

Multiple Teams Optimization

- How to optimally shrink one team while expanding another team?
- How to recruit a new player from several other teams?
- Enhance all teams within an organization and/or form new teams by collectively imposing a series of team enhancement operations

Data

- *AMiner*: https://aminer.org/data
- *Semantic Scholar*: http://labs.semanticscholar.org/corpus/
- *MovieLens*: https://grouplens.org/datasets/movielens/
- *NBA*: https://www.basketball-reference.com
- *Github*: https://www.githubarchive.org/

Resources

- Project Website: http://team-net-work.org/ for papers, code, slides
- Prototype System: http://team-net-work.org/system.html

References

- Wuchty, Stefan, Ben Jones, and Brian Uzzi. "The Increasing Dominance of Teams in the Production of Knowledge," Science, May 2007, 316:1036-1039.
- Woolley, Anita Williams, et al. "Evidence for a collective intelligence factor in the performance of human groups." science 330.6004 (2010): 686-688
- Rui Yan, Jie Tang, Xiaobing Liu, Dongdong Shan, and Xiaoming Li. 2011. Citation count prediction: learning to estimate future citations for literature. CIKM, 2011.
- D. Wang, C. Song, and A.-L. Barabasi. Quantifying long-term scientific impact. Science, 342(6154): 127-132, 2013.
- T. Lappas, K. Liu, and E. Terzi. Finding a team of experts in social networks. In KDD, pages 467–476, 2009.
- S. S. Rangapuram, T. Buhler, and M. Hein. Towards realistic team formation in social networks based on densest subgraphs. WWW 2013.
- Kim, Young Ji, et al. "What Makes a Strong Team?: Using Collective Intelligence to Predict Team Performance in League of Legends." CSCW. 2017.

References

- Duch, Jordi, Joshua S. Waitzman, and Luís A. Nunes Amaral. "Quantifying the performance of individual players in a team activity." PloS one 5.6 (2010): e10937.
- F. Thung, T. F. Bissyande, D. Lo, and L. Jiang, Network Structure of Social Coding in GitHub. CSMR 2013
- Yuxiao Dong, Reid A. Johnson, and Nitesh V. Chawla. 2015. Will This Paper Increase Your h-index?: Scientific Impact Prediction. WSDM, 2015.
- L. Li, and H. Tong: The Child is Father of the Man: Foresee the Success at the Early Stage. KDD 2015: 655-664
- L. Li, H. Tong, J. Tang and W. Fan: “iPath: Forecasting the Pathway to Impact”. SDM 2016
- Liangyue Li, Hanghang Tong, Yong Wang, Conglei Shi, Nan Cao and Norbou Buchler. Is the Whole Greater Than the Sum of Its Parts? KDD, 2017.
- Aris Anagnostopoulos, Luca Becchetti, Carlos Castillo, Aristides Gionis, and Stefano Leonardi. 2010. Power in unity: forming teams in large-scale community systems. CIKM, 2010.

References

- Duch, Jordi, Joshua S. Waitzman, and Luís A. Nunes Amaral. "Quantifying the performance of individual players in a team activity." PloS one 5.6 (2010): e10937.
- Aris Anagnostopoulos, Luca Becchetti, Carlos Castillo, Aristides Gionis, and Stefano Leonardi. Online Team Formation in Social Networks. WWW, 2012.
- Syama Sundar Rangapuram , Thomas Bühler , Matthias Hein, Towards realistic team formation in social networks based on densest subgraphs. WWW, 2013.
- N. Cao, Y.-R. Lin, L. Li, H. Tong: g-Miner: Interactive Visual Group Mining on Multivariate Graphs, ACM CHI 2015
- L. Li, H. Tong, N. Cao, K. Ehrlich, Y.-R. Lin and N. Buchler: Enhancing Team Composition in Professional Networks: Problem Definitions and Fast Solutions, TKDE, 2016
- Marco Tulio Ribeiro, Sameer Singh, Carlos Guestrin. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. KDD, 2016.
- Koh PW and Liang Percy Liang. Understanding black-box predictions via influence functions. ICML, 2017.