Recent Advances in **Applied** ***Matrix*** Technologies

Fei Wang

IBM Research

Hanghang Tong

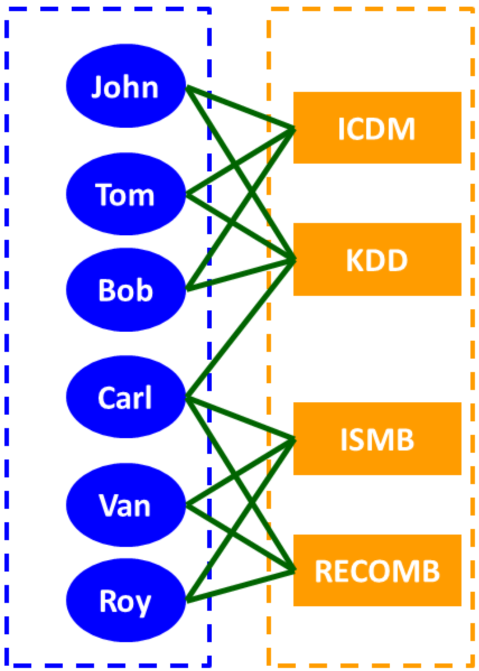
City College, CUNY

Outline

- Introduction
- Overview of the Technologies
- Applications in Health Informatics
- Applications in Social Informatics
- Conclusions and Future Works

What are Matrices?

Matrix: A Natural Representation for Networks/Graphs/Relational Data



Author **Conference**

Conference

Author

1	1	0	0
1	1	0	0
1	1	0	0
0	1	1	1
0	0	1	1
0	0	1	1

Adjacency matrix: *A*

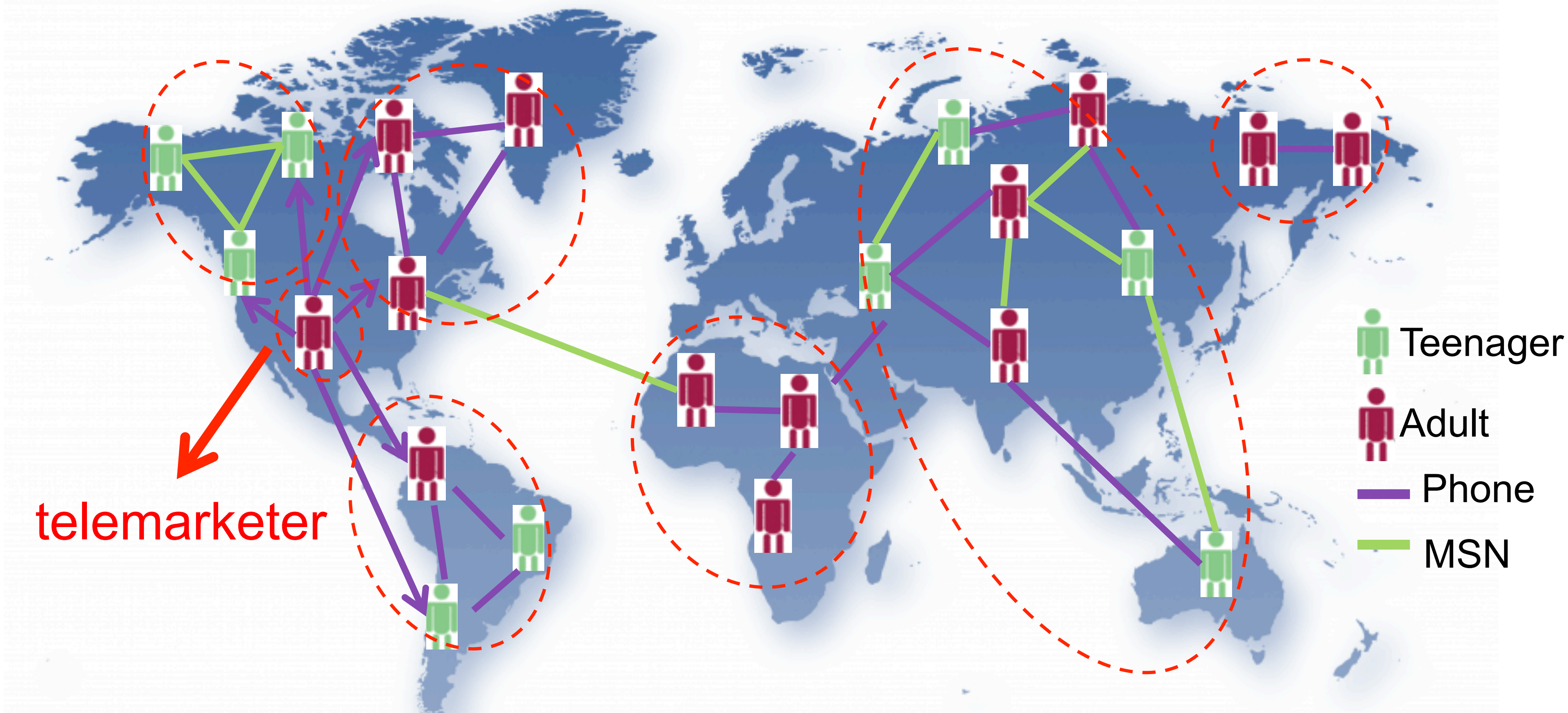
Matrices in Social Networks

Research Qs: How to find common friends?

Matrices: rows/columns: users; entries: friendship

Matrix Tools: graph proximity

Matrices in Social Networks

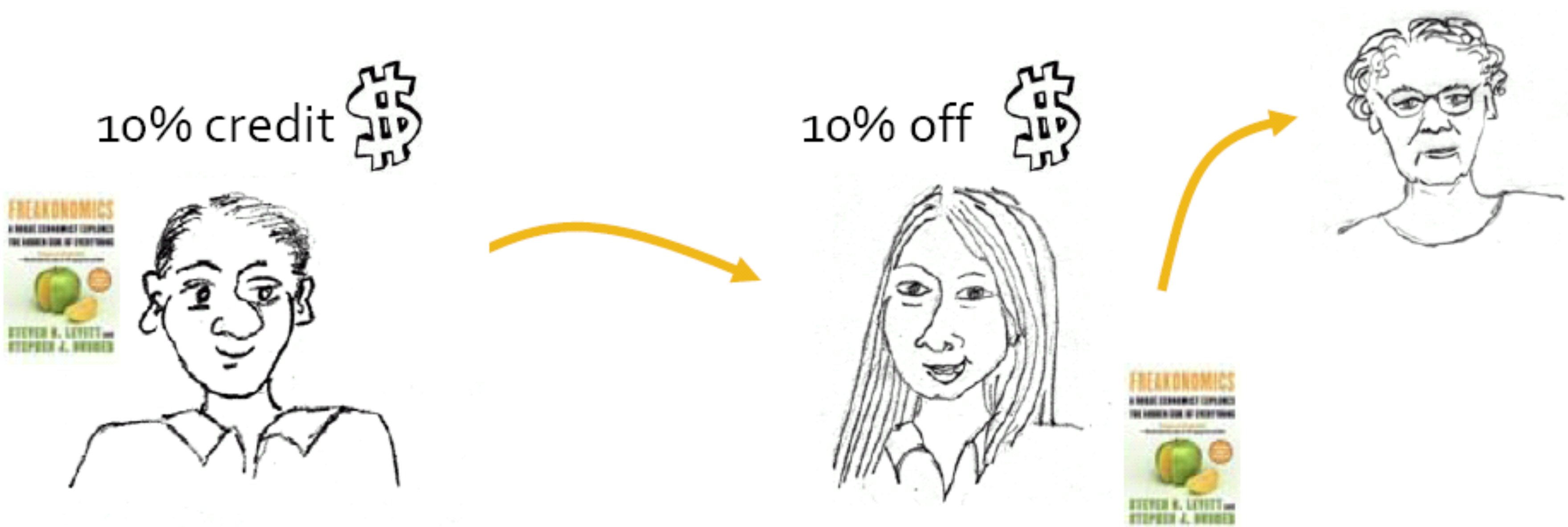


Research Qs: How to spot abnormal calling activities?

Matrices: rows/columns: users; entries: phone calls

Matrix Tools: graph proximity; low-rank approximation

Matrices in Social Networks [Leskovec+ 2007]



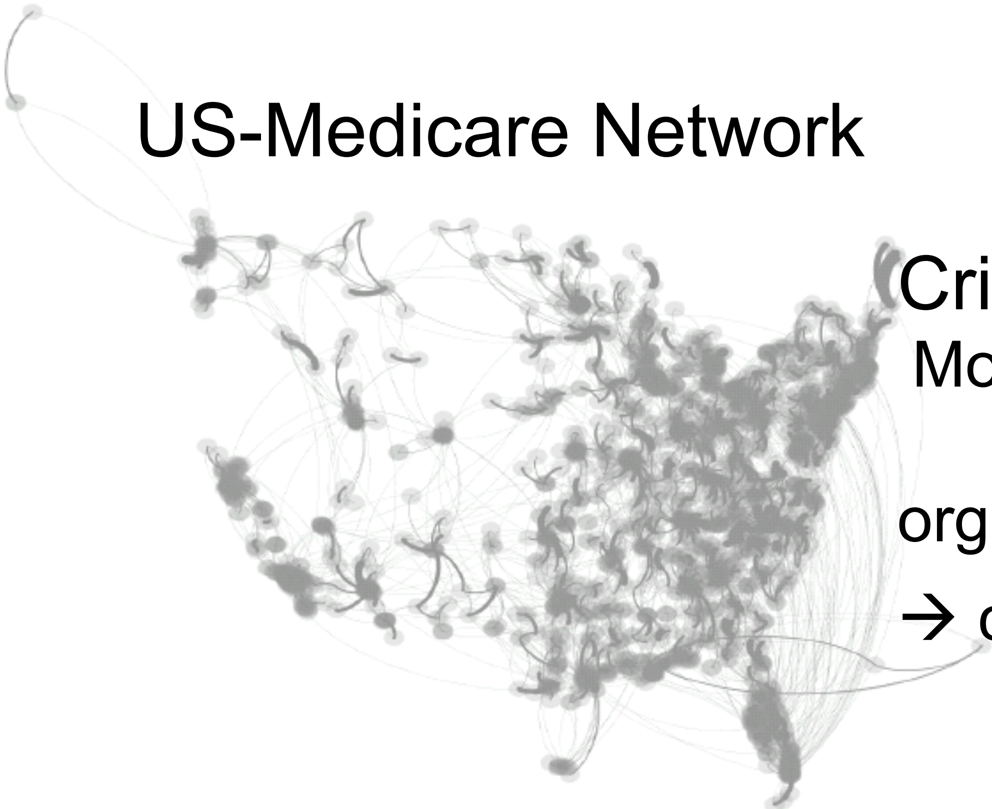
Research Qs: Can we boost the purchase?

Matrices: rows/columns: people; entries: recommendation

Matrix Tools: eigenvalue optimization

Matrices in Healthcare [Prakash+ 2013]

US-Medicare Network



Critical Patient transferring

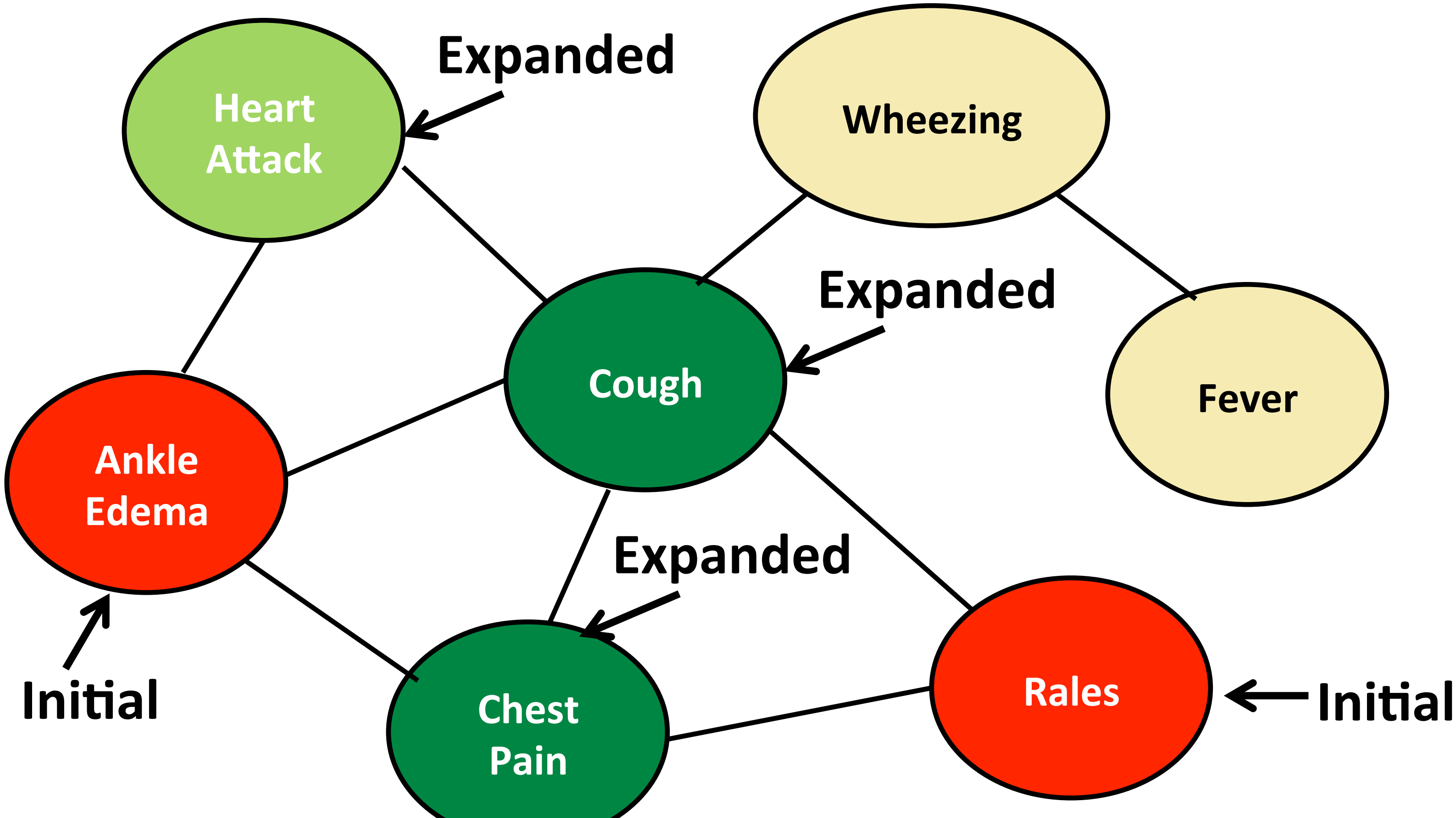
Move patients → specialized care → highly resistant micro-organism → Infection controlling → costly & limited

Research Qs: How to optimally allocate resources?

Matrices: rows/columns: hospitals; entries: patient transfer

Matrix Tools: eigenvalue optimization

Matrices in Healthcare [Parikshit+ 2012]

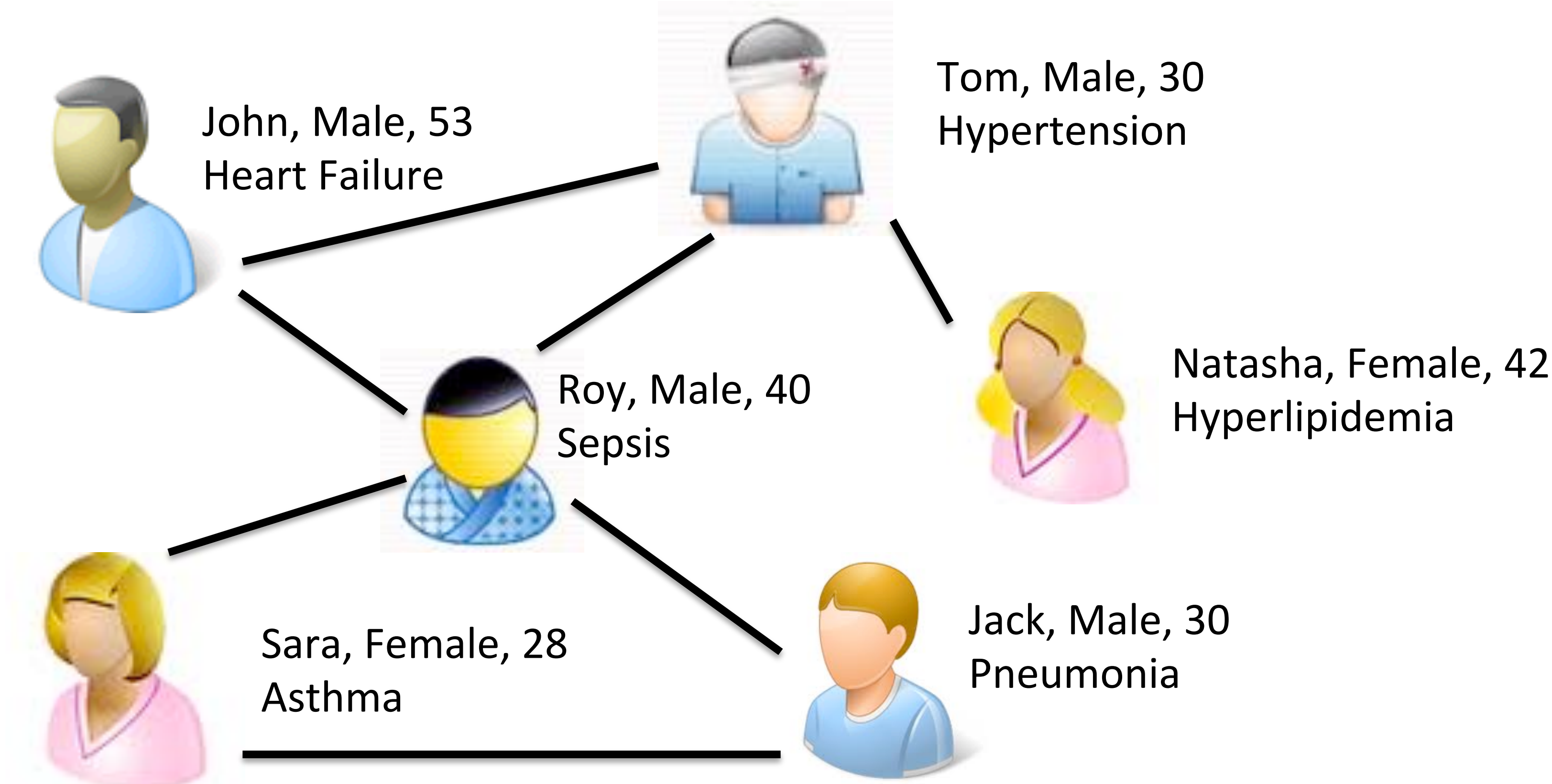


Research Qs: How to find more, related symptoms?

Matrices: rows/columns: symptoms; entries: co-occurrence

Matrix Tools: graph proximity

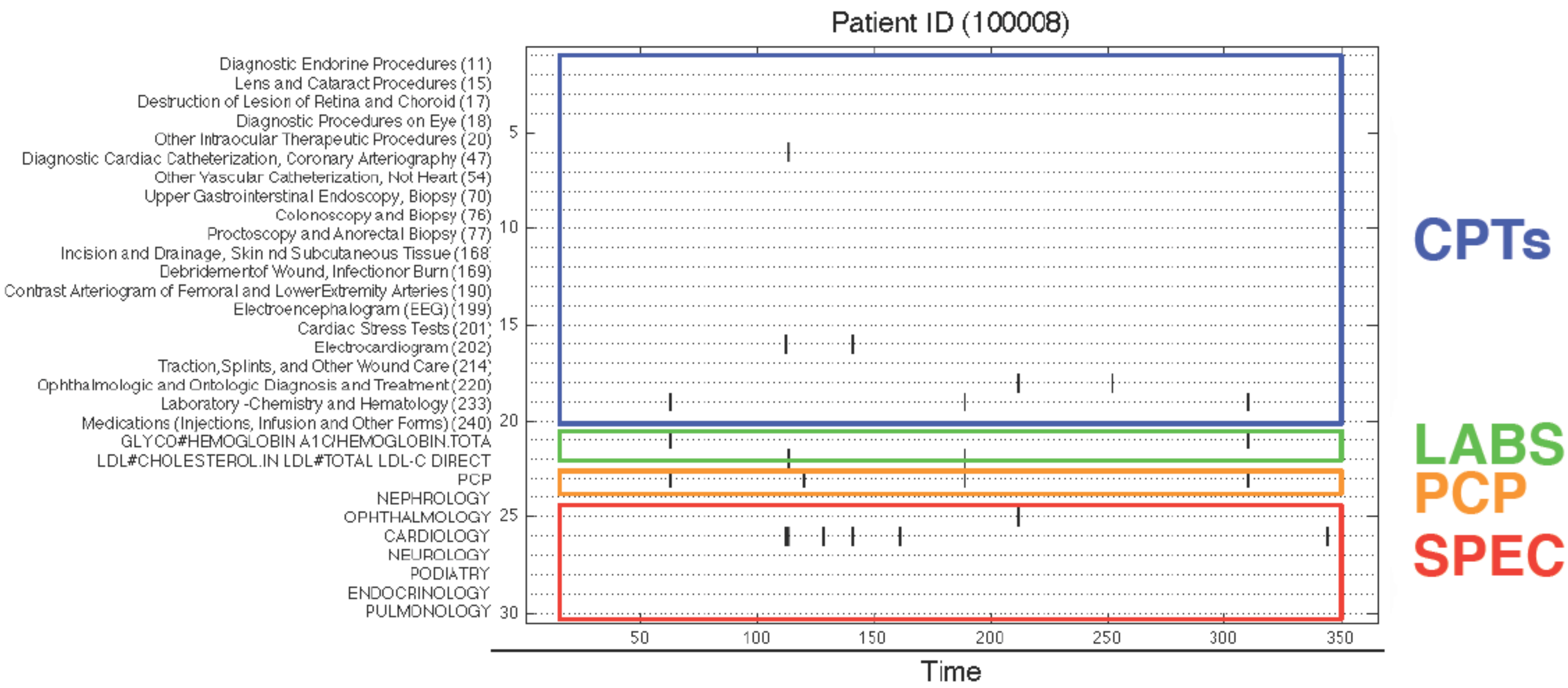
Matrices in Healthcare [Fei+ 2011]



Research Qs: How to find clinically similar patients?

Matrices: rows: patients; cols: clinical features; entries: values

Matrices in Healthcare [Fei+ 2012]



Research Qs: How to find frequent event subsequences?

Matrices: rows: events; cols: time; entries: indicator

Matrix Tools: Low rank approximation

Outline

- Introduction
- Overview of the Technologies
- Applications in Health Informatics
- Applications in Social Informatics
- Conclusions and Future Works

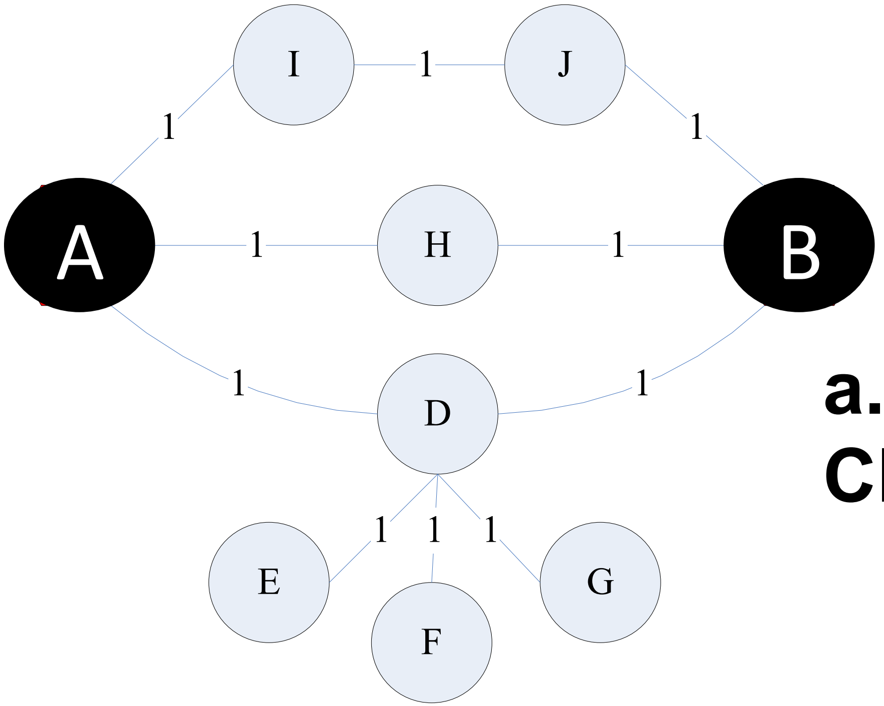
Overview of the Technologies

- T1: Graph Proximity
- T2: Low-Rank Approximation
- T3: Sparse Learning
- T4: Large-Scale Learning
- T5: Eigenvalue Opt. (in Section 4)

T1: Graph Proximity

- Basic Techniques: RWR
- Recent Advance #1: Supervision
- Recent Advance #2: Graph Kernel

Basic Tech.: Node Proximity Measurement

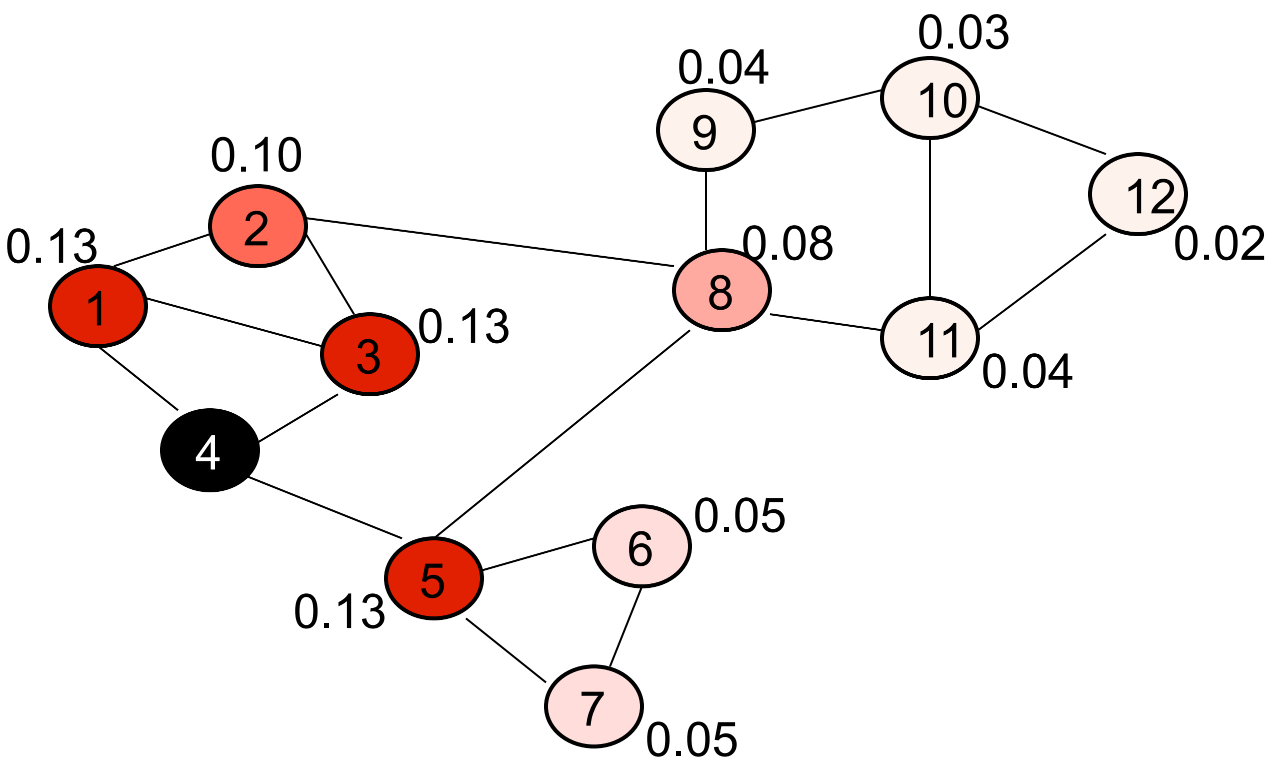


a.k.a Relevance, Closeness, 'Similarity'…

Q: How close is A to B?

Basic Tech. : Random Walk with Restart

[Tong+ ICDM 2006]



	Node 4
Node 1	0.13
Node 2	0.10
Node 3	0.13
Node 4	0.22
Node 5	0.13
Node 6	0.05
Node 7	0.05
Node 8	0.08
Node 9	0.04
Node 10	0.03
Node 11	0.04
Node 12	0.02

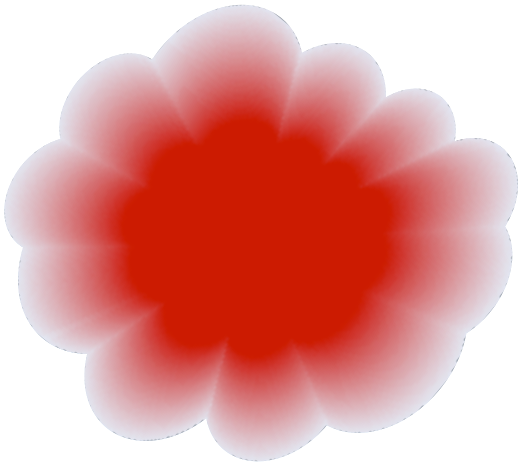
Nearby nodes, higher scores

More red, more relevant

Ranking vector

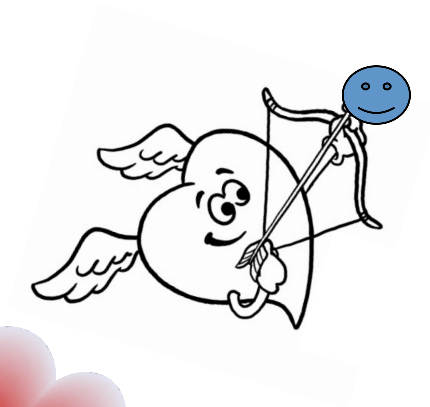
$\vec{r}_4$

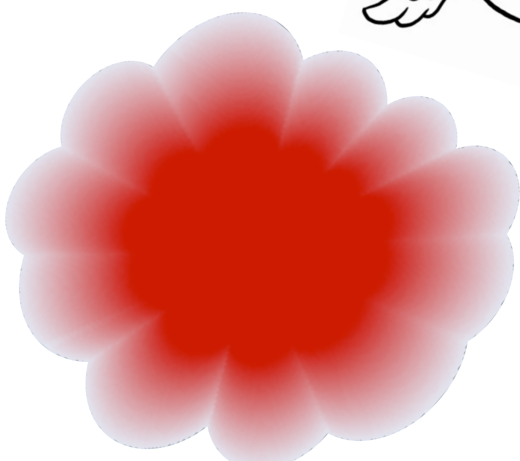
RWR: Think of it as Wine Spill

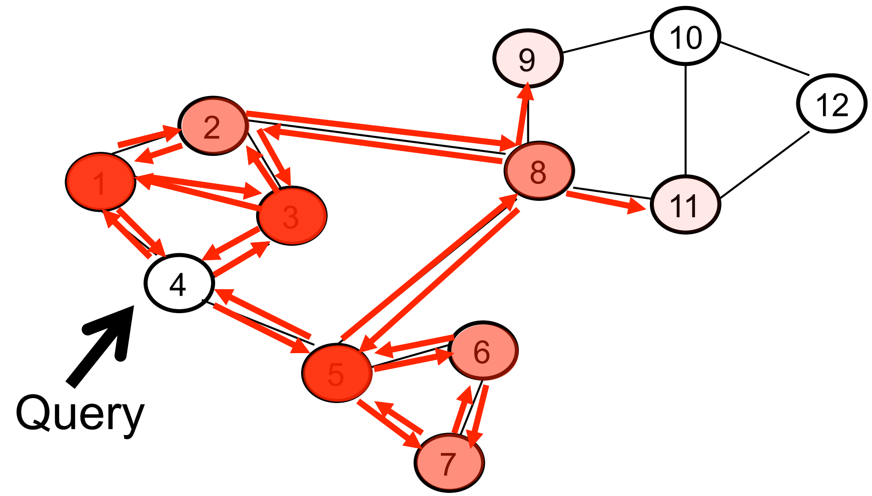


1. Spill a drop of wine on cloth
2. Spread/diffuse to the neighborhood

RWR: Wine Spill on a Graph





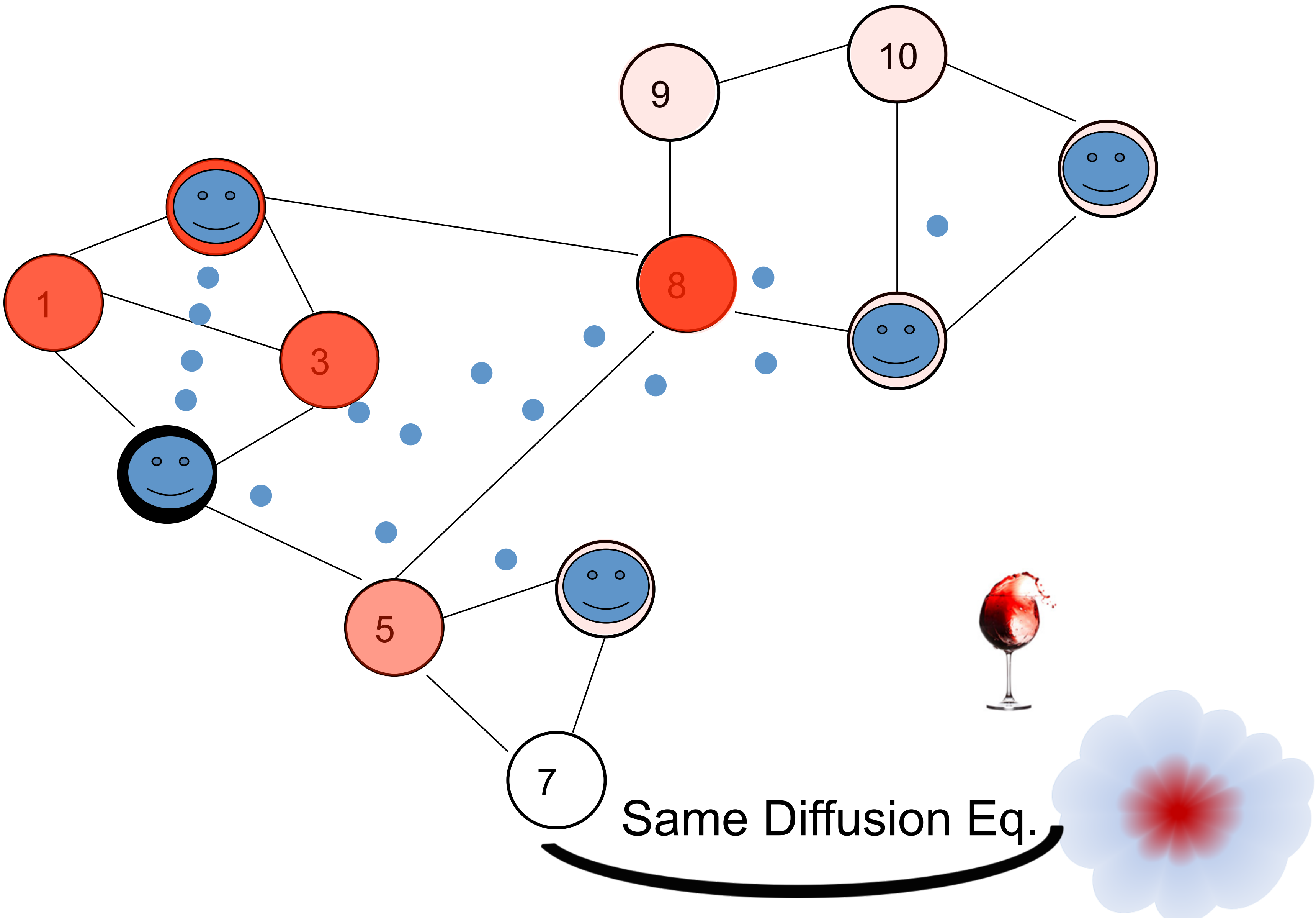


wine spill on cloth

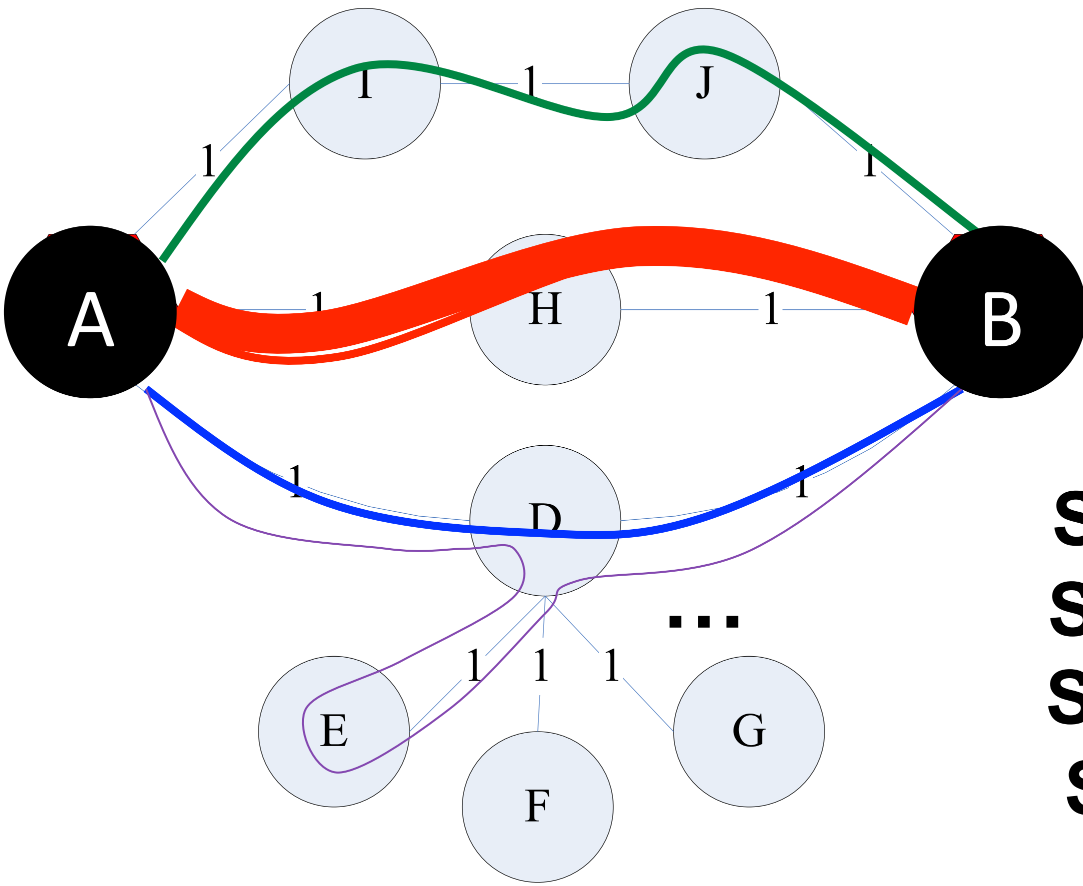
RWR on a graph

Same Diffusion Eq.

Random Walk with Restart



Intuition: Why *RWR* is A Good Score?



Prox (A, B) =

Score (Red Path) +

Score (Green Path) +

Score (Blue Path) +

Score (Purple Path) +

...

High proximity ⟺ many, short, heavy-weighted paths

Computing RWR

$$\vec{r}_i = cW\vec{r}_i + (1-c)\vec{e}_i$$

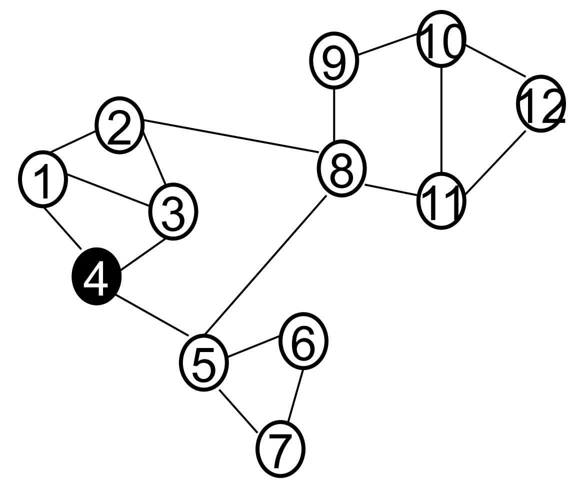
Ranking vector

(Normalized) Adjacency matrix

Restart p

Starting vector

$$\begin{pmatrix}0.13\\0.10\\0.13\\0.22\\0.13\\0.05\\0.05\\0.08\\0.04\\0.03\\0.04\\0.02\end{pmatrix} = 0.9 \times \begin{pmatrix} & 1/3 & 1/3 & 1/3 & & & & & & & & \\ 1/3 & & 1/3 & & & & & 1/4 & & & & \\ 1/3 & 1/3 & & 1/3 & & & & & & & & \\ 1/3 & & 1/3 & & 1/4 & & & & & & & \\ & & & 1/3 & & 1/2 & 1/2 & 1/4 & & & & \\ & & & & 1/4 & & 1/2 & & & & & \\ & & & & 1/4 & 1/2 & & & & & & \\ & 1/3 & & & 1/4 & & & & 1/2 & & 1/3 & \\ & & & & & & & 1/4 & & 1/3 & & \\ & & & & & & & & 1/2 & & 1/3 & 1/2 \\ & & & & & & & 1/4 & & 1/3 & & 1/2 \\ & & & & & & & & & 1/3 & 1/3 & \end{pmatrix} \begin{pmatrix}0.13\\0.10\\0.13\\0.22\\0.13\\0.05\\0.05\\0.08\\0.04\\0.03\\0.04\\0.02\end{pmatrix} + 0.1 \times \begin{pmatrix}0\\0\\0\\1\\0\\0\\0\\0\\0\\0\\0\\0\end{pmatrix}$$



nx1 *nxn* *nx1*

Footnote: "Maxwell Equation" for Web [Soumen Chakrabarti]

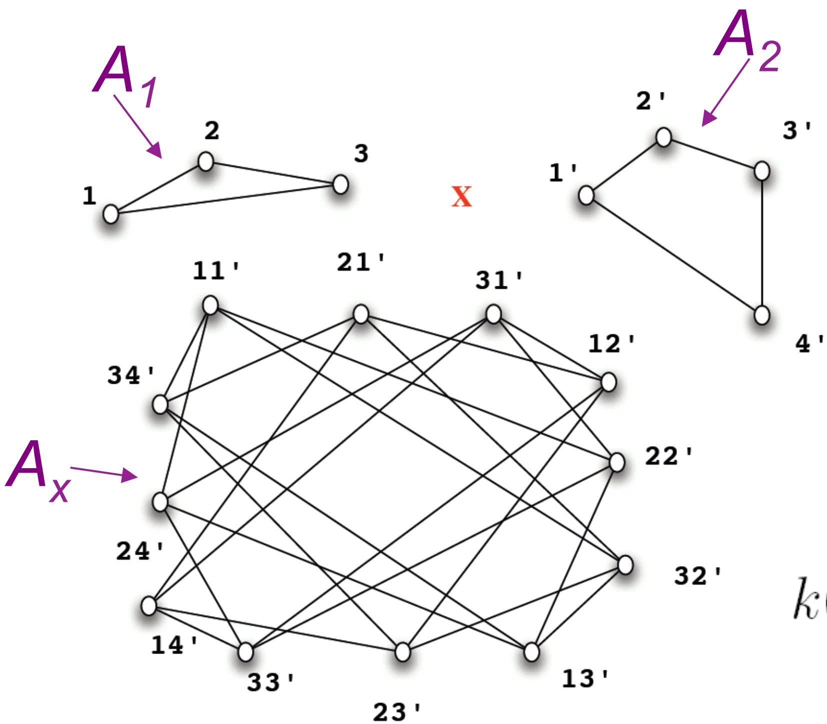
Recent Advance #1: Supervision

$$\vec{r}_i = cW\vec{r}_i + (1-c)\vec{e}_i$$

- Q: What is the optimal *W*?
- A: Learning optimal weights from supervision
- Key Idea: if we know some preference, we use such supervision to guild random walks to minimize
  - Penalty of preference violation + model complexity

L. Backstrom, J. Leskovec.: Supervised Random Walks: Predicting and Recommending Links in Social Network. WSDM 2011
A. Agarwal, S. Chakrabarti: Learning random walks to rank nodes in graphs. ICML 2007

Recent Advance #2: Node Proximity → Graph Similarity/Kernel



- Q: Sim(A_1, A_2) ?
- A: Do ***two*** random walks (A_1, A_2)!
- … = one random walk on A_x

$$k(G, G') = \sum_k \lambda^k q_\times^\top A_\times^k p_\times = q_\times^\top (\mathbf{I} - \lambda A_\times)^{-1} p_\times$$

SVN Vishwanathan. Fast computation of random walk graph kernels. NIPS 2006
K. Borgwardt and X. Yan: Graph Mining and Graph Kernels. KDD 2008 Tutorial
U. Kang, H. Tong and J. Sun: Fast Random Walk Graph Kernel. SDM 2012
~20+ other graph similarity measures

Overview of the Technologies

T1: Graph Proximity

T2: Low-Rank Approximation

T3: Sparse Learning

T4: Large-Scale Learning

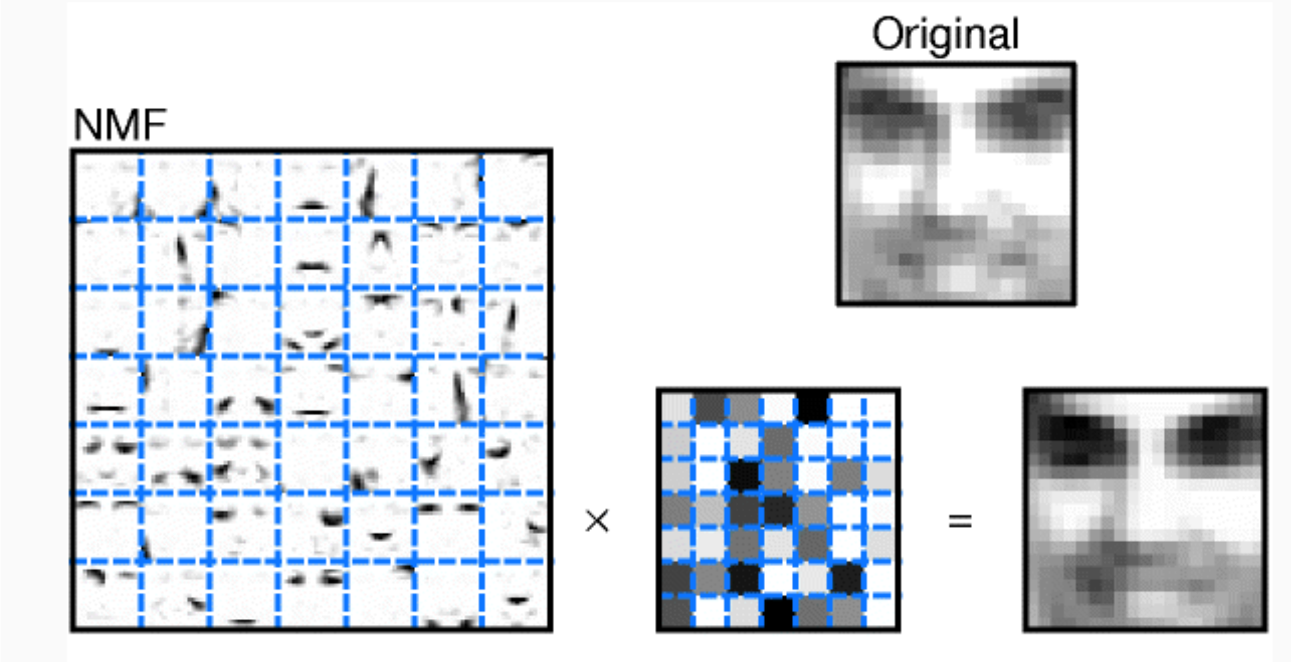
T5: Eigenvalue Opt. (in Section 4)

Why low rank approximation

- Collaborative Filtering
  - it is commonly believed that only a few factors contribute to anyone's taste or preference.
- Health Informatics
  - Usually the progression of disease is highly associated with a certain set of risk factors
- …

Low Rank Approximation

- Nonnegative Matrix Factorization (NMF)
- Nuclear norm related technologies

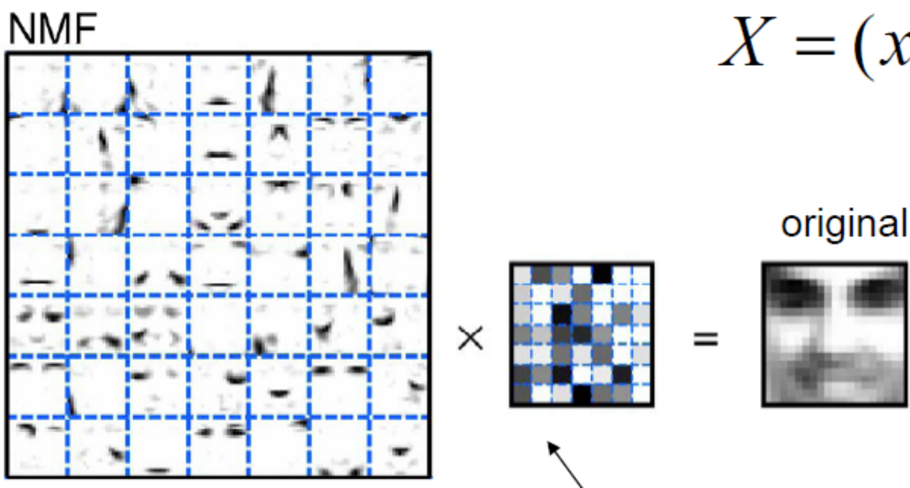


Daniel D. Lee and H. Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788-791 (21 October 1999)

Nonnegative Matrix Factorization (NMF)

- Factorizing a nonnegative matrix to the product of two low-rank matrices

$$X = (x_1, x_2, \cdots, x_n)$$



$$\min_{\mathbf{F}\in\mathbb{R}_+^{d\times r},\mathbf{G}\in\mathbb{R}_+^{n\times r}} \mathcal{L}(\mathbf{F},\mathbf{G})$$

$$\mathcal{L}(\mathbf{F},\mathbf{G}) = \left\|\mathbf{X} - \mathbf{F}\mathbf{G}^\top\right\|_F^2$$

$$FG^T \approx X$$

$$F = (f_1, f_2, \cdots, f_k) \quad G = (g_1, g_2, \cdots, g_k)$$

NMF Solutions: Multiplicative Updates

- Multiplicative update method

$$\mathbf{F}_{ij} \longleftarrow \mathbf{F}_{ij} \frac{(\mathbf{X}\mathbf{G})_{ij}}{(\mathbf{F}\mathbf{G}^T\mathbf{G})_{ij}}$$

$$\mathbf{G}_{ij} \longleftarrow \mathbf{G}_{ij} \frac{(\mathbf{X}^T\mathbf{F})_{ij}}{(\mathbf{G}\mathbf{F}^T\mathbf{F})_{ij}}$$

Daniel D. Lee and H. Sebastian Seung (2001). Algorithms for Non-negative Matrix Factorization. NIPS 2001.
H Zhou, K Lange, and M Suchard. (2010) Graphical processing units and high-dimensional optimization, *Statistical Science*, 25:311-324

NMF Solutions: Alternating Nonnegative Least Squares

- Initialize F and G with nonnegative values
- Iterate the following procedure:
  - Fixing $\mathbf{G}^{(t)}$, Solve $\min_{\mathbf{F}} \quad J(\mathbf{F}, \mathbf{G}^{(t)}) = \left\| \mathbf{X} - \mathbf{F}(\mathbf{G}^{(t)})^T \right\|_F^2$
  - Fixing $\mathbf{F}^{(t)}$, Solve $\min_{\mathbf{G}} \quad J(\mathbf{F}^{(t)}, \mathbf{G}) = \left\| \mathbf{X} - \mathbf{F}^{(t)}\mathbf{G}^T \right\|_F^2$

(1) Projected Gradient: http://www.csie.ntu.edu.tw/~cjlin/nmf/
(2) Newtown Type of Method:
http://www.cs.utexas.edu/users/dmkim/Source/software/nnma/index.html
(3) Block Principal Pivoting: https://sites.google.com/site/jingukim/nmf_bpas.zip?attredirects=0

P. Paatero and U. Tapper. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics, 5(1):111–126, 1994

C.-J. Lin. Projected gradient methods for non-negative matrix factorization. *Neural Computation*,19(2007), 2756-2779.

D. Kim, S. Sra, I. S. Dhillon, Fast Newton-type Methods for the Least Squares Nonnegative Matrix Approximation Problem. SDM 2007.

J. Kim and H. Park. Toward Faster Nonnegative Matrix Factorization: A New Algorithm and Comparisons. ICDM 2008.

NMF: Extensions

- General loss
  - Bregman Divergence
- Different constraints
  - Semi-NMF, Convex NMF, Symmetric NMF
- Incorporating supervisions
  - Pairwise constraints, label
- Multiple factorized matrices
  - Tri-factorization

I. S. Dhillon and S. Sra. Generalized Nonnegative Matrix Approximations with Bregman Divergences. NIPS 2005.
Chris H. Q. Ding, Tao Li, Michael I. Jordan: Convex and Semi-Nonnegative Matrix Factorizations. IEEE Trans. Pattern Anal. Mach. Intell. 32(1): 45-55 (2010)
Chris H. Q. Ding, Tao Li, Wei Peng, Haesun Park: Orthogonal nonnegative matrix t-factorizations for clustering. KDD 2006.
Fei Wang, Tao Li, Changshui Zhang: Semi-Supervised Clustering via Matrix Factorization. SDM 2008: 1-12
Yuheng Hu, Fei Wang, Subbarao Kambhampati. Listen to the Crowd: Automated Analysis of Live Events via Aggregated Twitter Sentiment. IJCAI 2013.

Low Rank Approximation

- Nonnegative Matrix Factorization
- Nuclear norm related technologies

Rank Minimization and Nuclear Norm

- Matrix completion with rank minimization

$$\min_{\mathbf{X}} \operatorname{rank}(\mathbf{X}) \quad s.t.\ X_{ij} = M_{ij}\ \forall (i,j) \in \Omega$$

NP hard

- Convex relaxation

$$\min_{\mathbf{X}} \|\mathbf{X}\|_* \quad s.t.\ X_{ij} = M_{ij}\ \forall (i,j) \in \Omega$$

$$\boxed{\|\mathbf{X}\|_* = \sum_i \sigma_i(X)}$$

M. Fazel, H. Hindi, S. Boyd. A Rank Minimization Heuristic with Application to Minimum Order System Approximation. Proceedings American Control Conference, 6:4734-4739, June 2001.

Nuclear Norm Minimization

- Singular Value Thresholding
  - http://svt.stanford.edu/
- Accelerated gradient
  - http://www.public.asu.edu/~jye02/Software/SLEP/index.htm
- Interior point methods
  - http://abel.ee.ucla.edu/cvxopt/applications/nucnrm/

J-F. Cai, E.J. Candès and Z. Shen. A Singular Value Thresholding Algorithm for Matrix Completion. SIAM Journal on Optimization. Volume 20 Issue 4, January 2010 Pages 1956-1982.

Shuiwang Ji and Jieping Ye. An Accelerated Gradient Method for Trace Norm Minimization. The Twenty-Sixth International Conference on Machine Learning (ICML 2009)

Z. Liu, Lieven Vandenberghe. Interior-point method for nuclear norm approximation with application to system identification. SIAM Journal on Matrix Analysis and Applications (2009)

Overview of the Technologies

T1: Graph Proximity

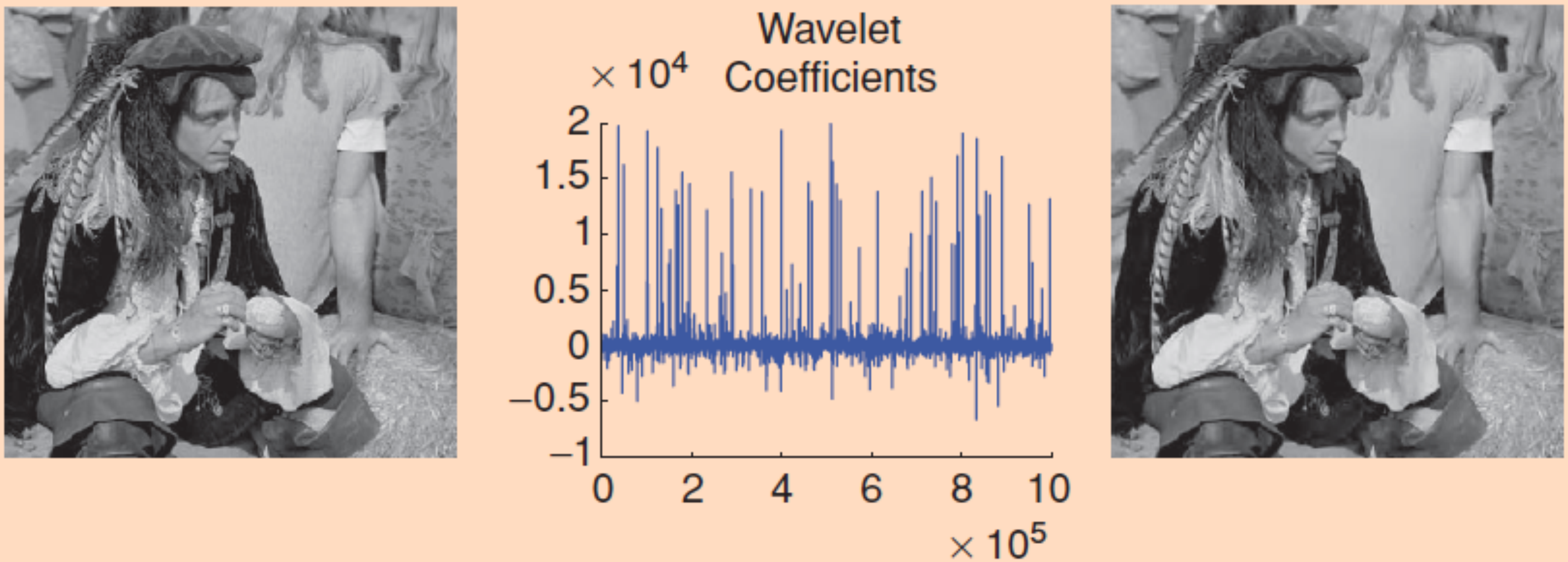
T2: Low-Rank Approximation

T3: Sparse Learning

T4: Large-Scale Learning

T5: Eigenvalue Opt. (in Section 4)

Why Sparse Learning



Candes, E.J., Wakin, M.B. An Introduction To Compressive Sampling. Signal Processing Magazine, IEEE. Volume: 25, Issue: 2. Page(s): 21 – 30.

Sparsity: L0 Norm & L1 Norm

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{w}\|_0 \\ s.t. \quad & \mathbf{w} \in \mathcal{C} \end{aligned}$$

Lower Bound

$$\begin{aligned} \min_{\mathbf{w},\mathbf{z}} \quad & \mathbf{1}^\top \mathbf{z} \\ s.t. \quad & |w_i| \leqslant R z_i \ \ \forall i = 1, 2, \cdots, d \\ & \mathbf{w} \in \mathcal{C}, \ \ z_i \in \{0, 1\} \ \ \forall i = 1, 2, \cdots, d \end{aligned}$$

$$\begin{aligned} \min_{\mathbf{w},\mathbf{z}} \quad & \mathbf{1}^\top \mathbf{z} \\ s.t. \quad & |w_i| \leqslant R z_i \ \ \forall i = 1, 2, \cdots, d \\ & \mathbf{w} \in \mathcal{C} \\ & z_i \in [0, 1] \ \ \forall i = 1, 2, \cdots, d \end{aligned}$$

$$\begin{aligned} \min_{\mathbf{w}} \quad & 1/R\|\mathbf{w}\|_1 \\ s.t. \quad & \mathbf{w} \in \mathcal{C} \end{aligned}$$

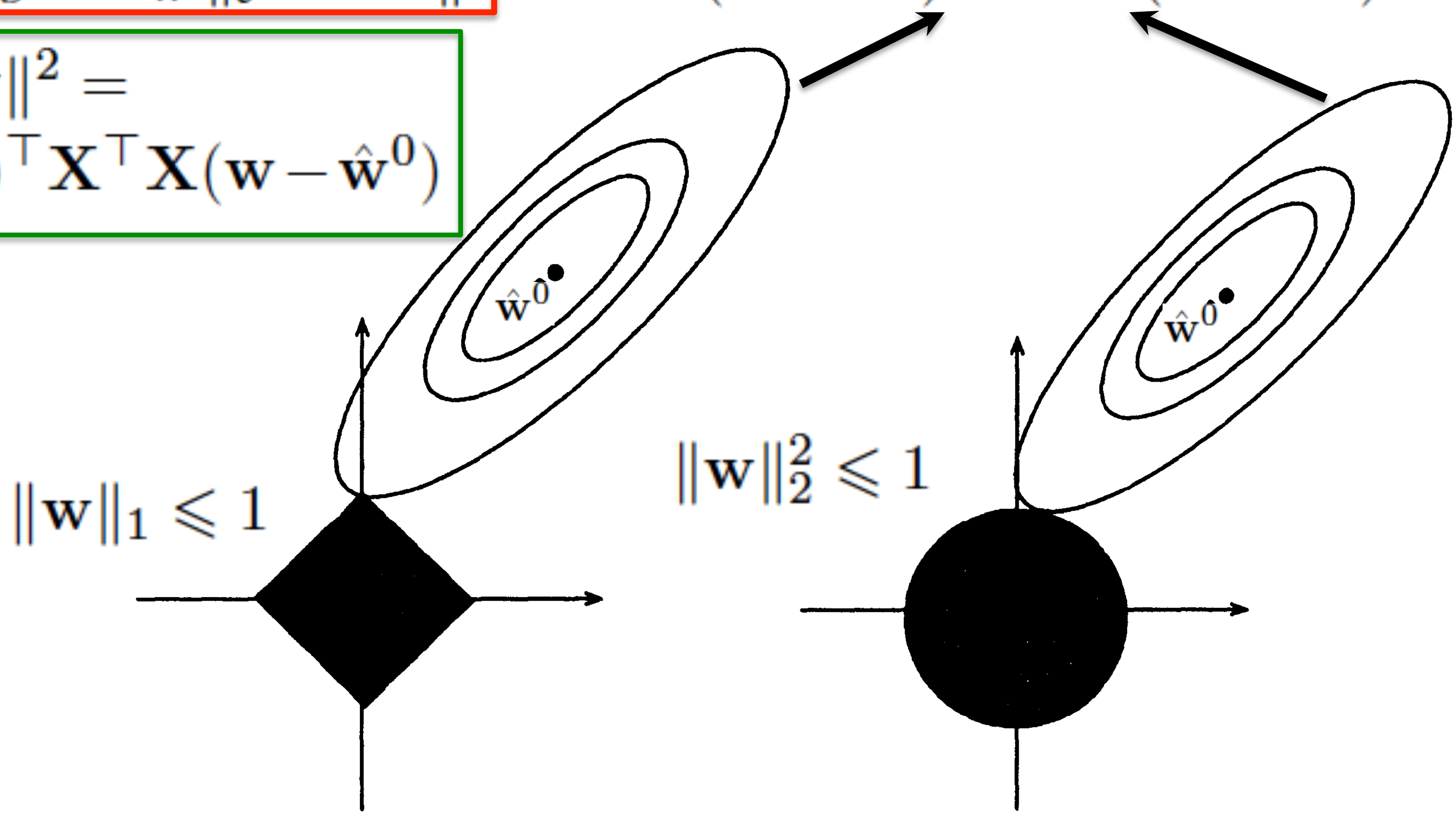
http://www.stanford.edu/class/ee364b/lectures/l1_slides.pdf

Why L1 Norm Can Achieve Sparsity

$$\boxed{\hat{\mathbf{w}}^0 = arg\,\min_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2}$$

$$\boxed{\|\mathbf{y} - \mathbf{X}\mathbf{w}\|^2 = (\mathbf{w} - \hat{\mathbf{w}}^0)^\top \mathbf{X}^\top \mathbf{X} (\mathbf{w} - \hat{\mathbf{w}}^0)}$$

$$(\mathbf{w} - \hat{\mathbf{w}}^0)^\top \mathbf{X}^\top \mathbf{X} (\mathbf{w} - \hat{\mathbf{w}}^0)$$



Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society, Series B. 1994.

Other Sparsity Penalties

- Group Lasso: L1/2 norm
- Exclusive Lasso: L2/1 norm
- Elastic Net Regularization
- Fused Lasso
- Tree Structured Group Lasso

SLEP: A Sparse Learning Package
http://www.public.asu.edu/~jye02/Software/SLEP/

Lukas Meier, Sara Van De Geer, Peter Bühlmann. The group lasso for logistic regression. Journal of the Royal Statistical Society: Series B, 70(1), 53–71, 2008.

Y. Zhou, *R. Jin*, and S. C. H. Hoi. Exclusive Lasso for Multi-task Feature Selection. AISTATS 2010.

Zou, Hui; Hastie, Trevor. Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society, Series B: 301–320. 2005.

R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B. 67(1), 91–108. 2005.

J. Liu, J. Ye. Moreau-Yosida Regularization for Grouped Tree Structure Learning. NIPS 2010.

Overview of the Technologies

T1: Graph Proximity

T2: Low-Rank Approximation

T3: Sparse Learning

T4: Large-Scale Learning

T5: Eigenvalue Opt. (in Section 4)

Distributed Learning

- Parallel Matrix Factorization
  - H. F. Yu, C. J. Hsieh, S. Si, and I. S. Dhillon. Scalable Coordinate Descent Approaches to Parallel Matrix Factorization for Recommender Systems. ICDM 2012.
  - Lester Mackey, Ameet Talwalkar, and Michael I. Jordan. Divide-and-Conquer Matrix Factorization. NIPS 2011.
- Parallel Spectral Clustering
  - W. Chen, Y. Song, H. Bai, C. Lin, E. Y. Chang. Parallel Spectral Clustering in Distributed Systems. IEEE TPAMI 33(3), 568-586. 2011.
- Parallel SVD
  - M. W. Berry, D. Mezher, B. Philippe, and A. Sameh. Parallel Algorithms for the Singular Value Decomposition. In Erricos John: Handbook on Parallel Computing and Statistics. https://www.irisa.fr/sage/bernard/publis/SVD-Chapter06.pdf
- Parallel Optimization
  - Y. Censor, S. A. Zenios. Parallel Optimization: Theory, Algorithms and Applications. Oxford University Press. 1998

Online Learning

- Online Matrix Factorization
  - J. Mairal, F. Bach, J. Ponce and G. Sapiro. Online Learning for Matrix Factorization and Sparse Coding. Journal of Machine Learning Research, volume 11. pages 19-60. 2010.
  - Fei Wang, Chenhao Tan, Christian Konig, Ping Li. Online Nonnegative Matrix Factorization for Document Clustering. SDM 2011.
  - A. Lefèvre, F. Bach, and C. Févotte. Online algorithms for Nonnegative Matrix Factorization with the Itakura-Saito divergence. WASPAA 2011.
- General Online Learning
  - Shai Shalev-Shwartz. Online Learning: Theory, Algorithms, and Applications. The Hebrew University of Jerusalem. PH.d. thesis. July 2007.
- Parallel Online Learning
  - Daniel Hsu, Nikos Karampatziakis, John Langford, Alex Smola. Parallel Online Learning. http://arxiv.org/abs/1103.4204v1.

Matrix Tools vs Applications

Tools	C' Patterns	Anomalies	Influ' Prop	Immunization	Symptom Exp'	Similar Patient	Clinical Patterns	Risk Factor
Prox.	✓				✓			
LRA		✓				✓	✓	
Sparse L'								✓
Large L'						✓		
Eigen. Opt.			✓	✓				

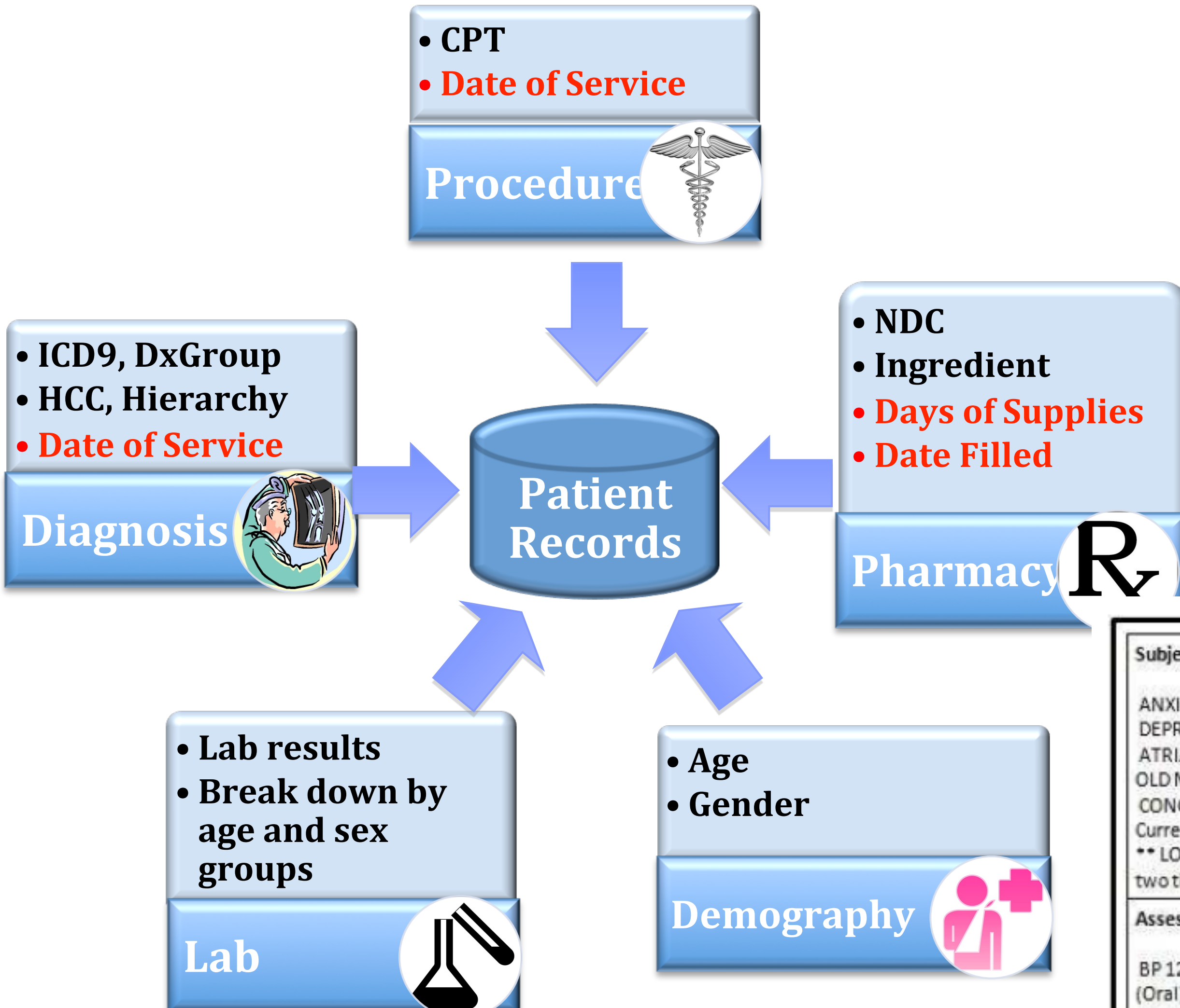
Social Networks

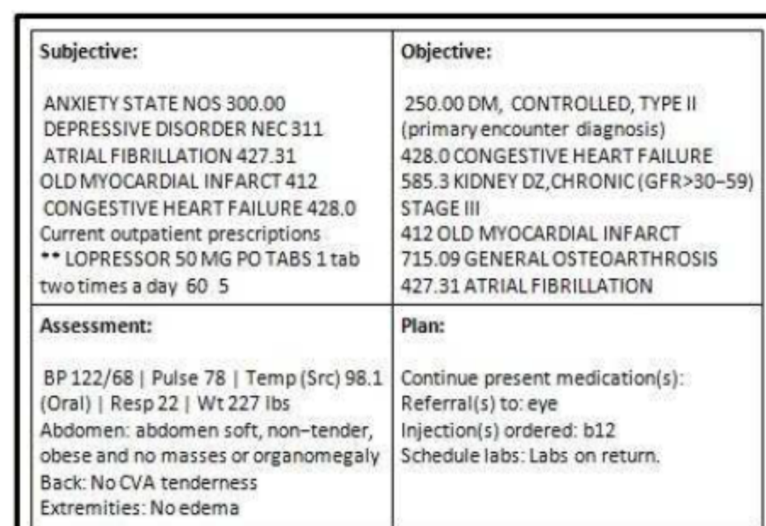
Healthcare

Outline

- Introduction
- Overview of the Technologies
- Applications in Health Informatics
- Applications in Social Informatics
- Conclusions and Future Works

Longitudinal Medical Records





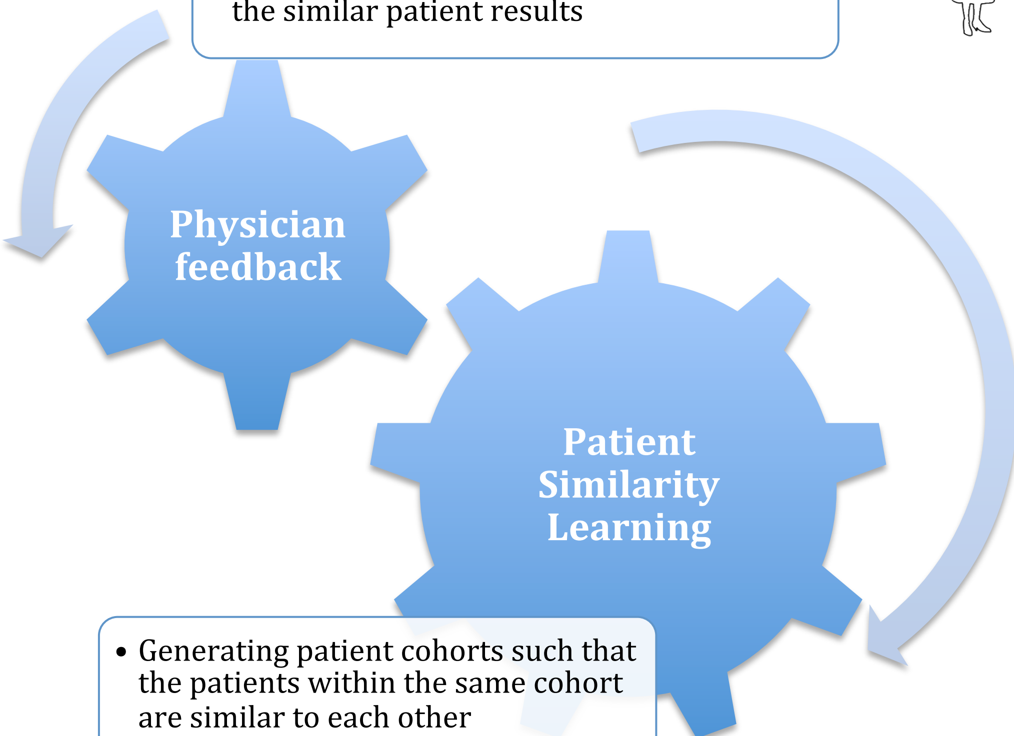
Subjective:	Objective:
ANXIETY STATE NOS 300.00 DEPRESSIVE DISORDER NEC 311 ATRIAL FIBRILLATION 427.31 OLD MYOCARDIAL INFARCT 412 CONGESTIVE HEART FAILURE 428.0 Current outpatient prescriptions ** LOPRESSOR 50 MG PO TABS 1 tab two times a day 60 5	250.00 DM, CONTROLLED, TYPE II (primary encounter diagnosis) 428.0 CONGESTIVE HEART FAILURE 585.3 KIDNEY DZ,CHRONIC (GFR>30-59) STAGE III 412 OLD MYOCARDIAL INFARCT 715.09 GENERAL OSTEOARTHROSIS 427.31 ATRIAL FIBRILLATION
Assessment:	**Plan:**
BP 122/68 | Pulse 78 | Temp (Src) 98.1 (Oral) | Resp 22 | Wt 227 lbs Abdomen: abdomen soft, non-tender, obese and no masses or organomegaly Back: No CVA tenderness Extremities: No edema	Continue present medication(s): Referral(s) to: eye Injection(s) ordered: b12 Schedule labs: Labs on return.

Applications in Health Informatics

- Patient Similarity Learning
- Risk Factor Identification
- Clinical Pattern Detection

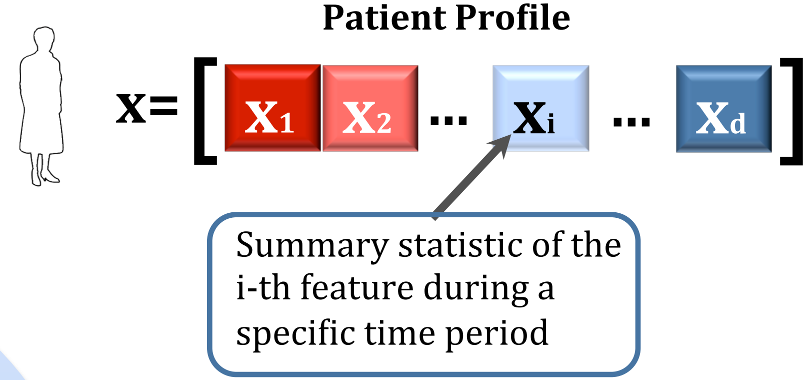
Patient Similarity Assessment

- Leverage historical data about the similar patients to diagnose the query patient
- Provide positive/negative feedback about the similar patient results



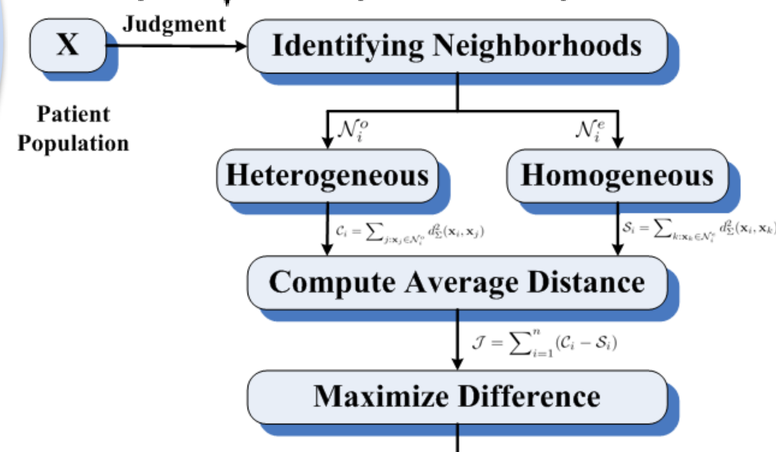
- Generating patient cohorts such that the patients within the same cohort are similar to each other
- Explain why they are similar

Vector Space Model



Locally Supervised Metric Learning (LSML)

$$d_{\Sigma}(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{(\mathbf{x}_i - \mathbf{x}_j)^{\top} \Sigma (\mathbf{x}_i - \mathbf{x}_j)} \qquad \Sigma = \mathbf{W}\mathbf{W}^{\top}$$

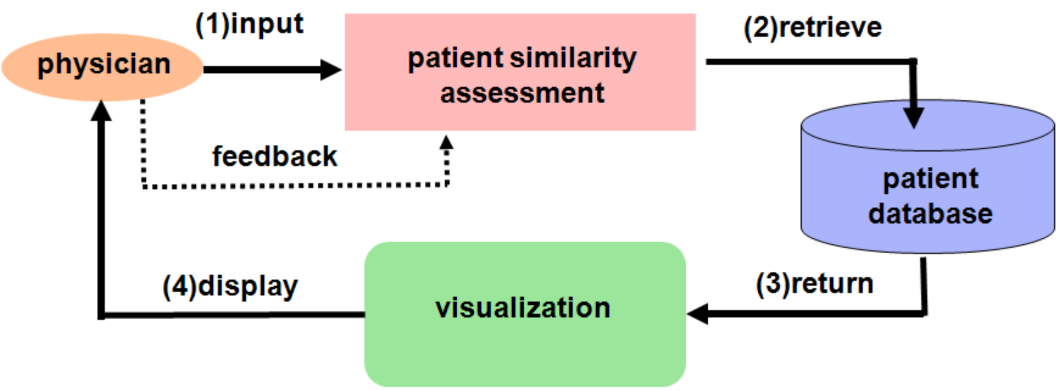


$$\min_{\mathbf{W}^{\top}\mathbf{W}=\mathbf{I}} tr\left(\mathbf{W}^{\top} \boxed{\mathbf{X}(\mathbf{L}^o - \mathbf{L}^e)\mathbf{X}^{\top}} \mathbf{W}\right)$$

Fei Wang. Semi-Supervised Metric Learning by Maximizing Constraint Margin. IEEE TSMC-B. 2011.
Fei Wang, Changshui Zhang. Feature Extraction by Maximizing the Average Neighborhood Margin. CVPR 2007.

Online Adjustment of Patient Similarity

Physician Decision Support System



How to adjust the learned patient similarity by incorporating physician's feedback in real time?

Learning the distance metric is equivalent to learn the projection matrix W, which can be solved by doing eigenvalue decomposition on $\mathbf{X}(\mathbf{L}^o - \mathbf{L}^e)\mathbf{X}^\top$

Any physicians' feedback is an increment on the matrix $\mathbf{L} = \mathbf{L}^o - \mathbf{L}^e$

Efficient eigensystem update of a matrix: matrix perturbation theory

iMet: interactive Metric Learning

matrix		eigensystem		perturbation
$\mathbf{X}\mathbf{L}\mathbf{X}^\top$ $\mathbf{X}(\mathbf{L}+\Delta\mathbf{L})\mathbf{X}^\top$	$\Rightarrow$	$(\lambda_i, \mathbf{w}_i)$ $(\widetilde{\lambda}_i, \widetilde{\mathbf{w}}_i)$	$\Rightarrow$	$\widetilde{\lambda}_i = \lambda_i + \Delta\lambda_i$ $\widetilde{\mathbf{w}}_i = \mathbf{w}_i + \Delta\mathbf{w}_i$

$$\mathbf{X}(\mathbf{L}+\Delta\mathbf{L})\mathbf{X}^\top(\mathbf{w}_i + \Delta\mathbf{w}_i) = (\lambda_i + \Delta\lambda_i)(\mathbf{w}_i + \Delta\mathbf{w}_i)$$

solution

$$\Delta\lambda_i = \mathbf{w}_i^\top \mathbf{X}\Delta\mathbf{L}\mathbf{X}^\top \mathbf{w}_i$$

$$\Delta\mathbf{w}_i = -\frac{1}{2}\mathbf{w}_i + \sum_{j\neq i} \frac{\mathbf{w}_j^\top \mathbf{X}\Delta\mathbf{L}\mathbf{X}^\top \mathbf{w}_i}{\lambda_i - \lambda_j}\mathbf{w}_j$$

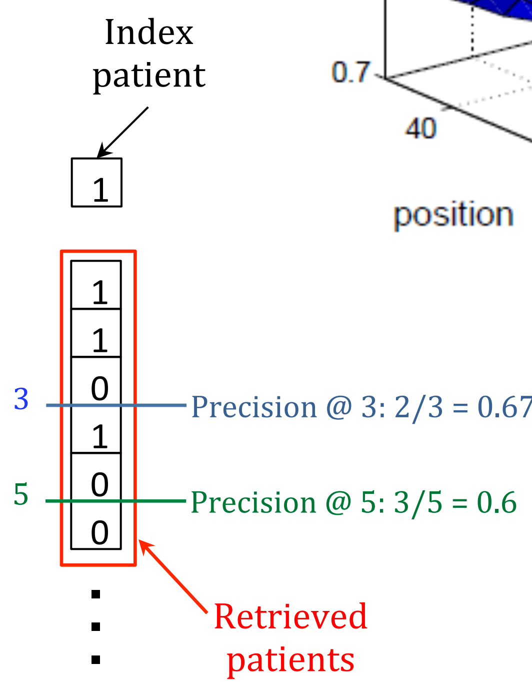
Fei Wang, J. Sun, J. Hu, S. Ebadollahi. iMet: Interactive Metric Learning in Healthcare Applications. SDM 2011.

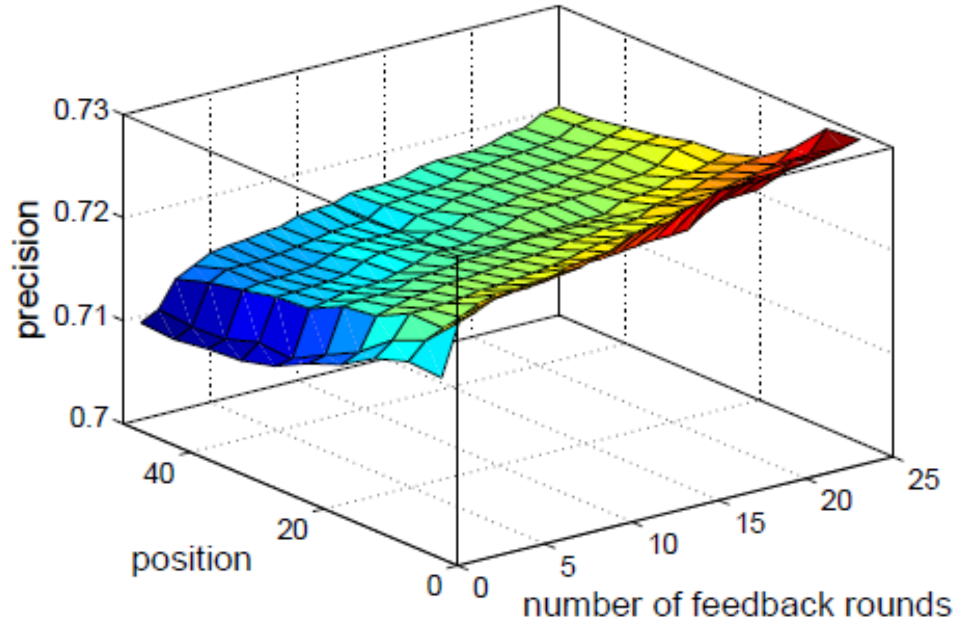
Performance Evaluation

Initial Metric: The patient population was clustered into 10 clusters using Kmeans with the features as counts of the HCC codes over one year. An initial distance metric was then learned using LSML

Feedback: For each round of simulated feedback, an index patient was randomly selected and 20 similar patients were retrieved based on current distance metric. The feedback is based on whether these retrieved patients have the same label as the index patient

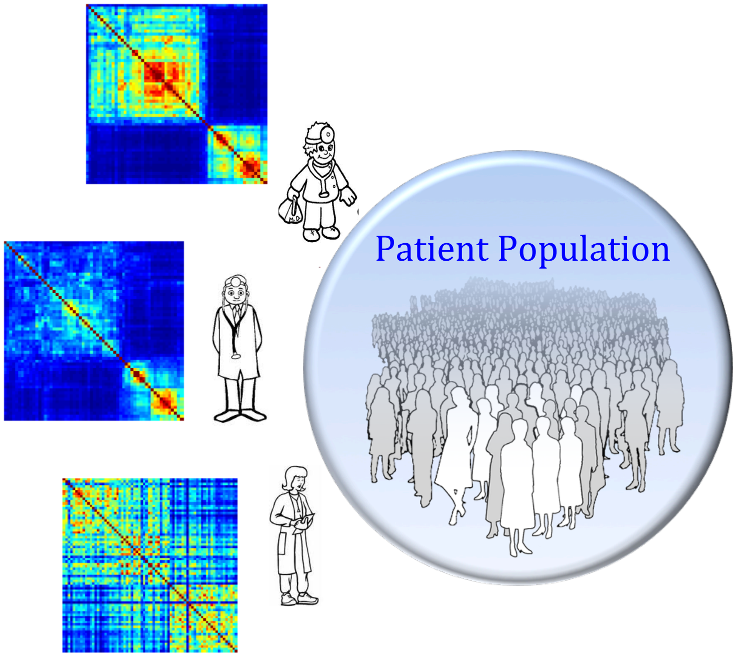
Performance metric: precision@position measure





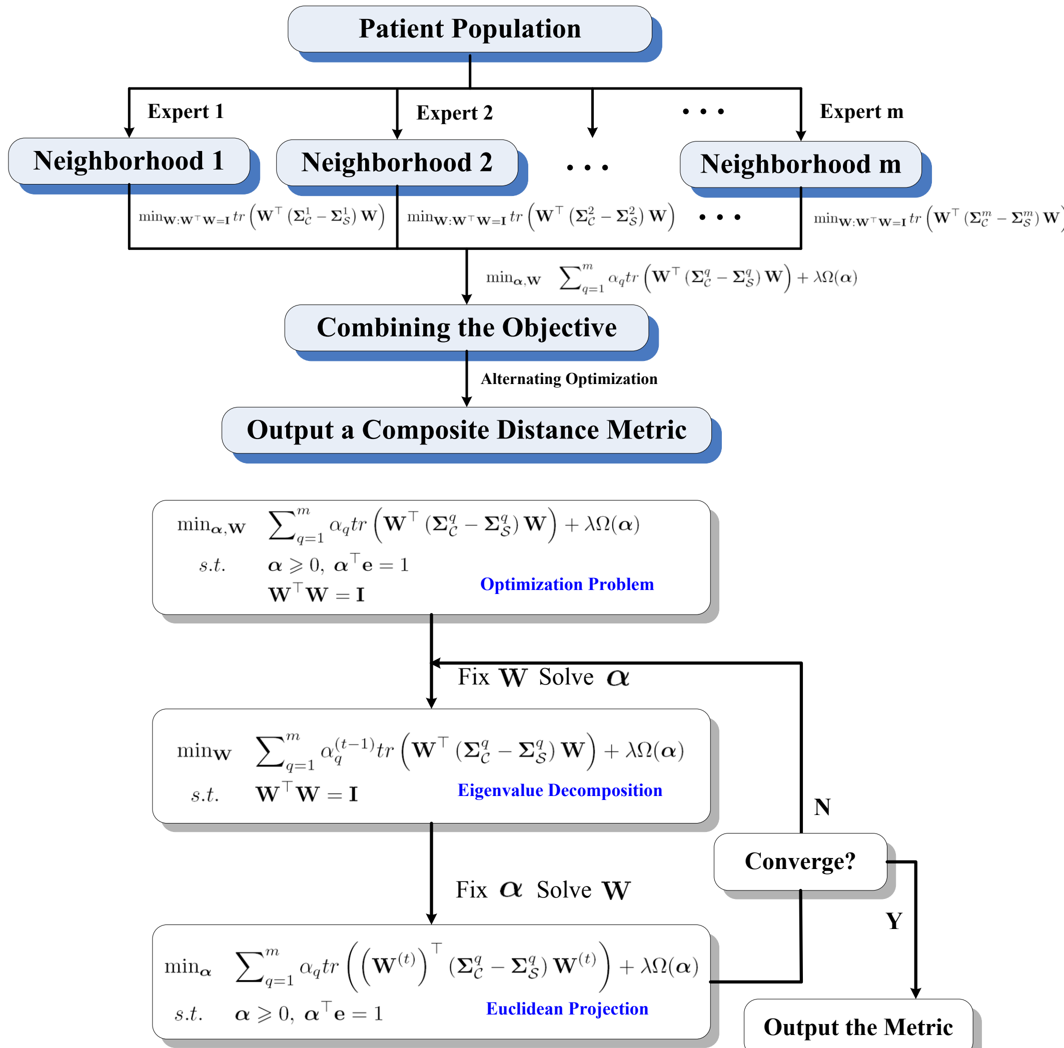
Integrating Multiple Physicians' Inputs

Different physicians have their own opinions on patient similarities



How to integrate these judgments from multiple physicians to a consistent similarity measure?

Comdi: Composite Distance integration



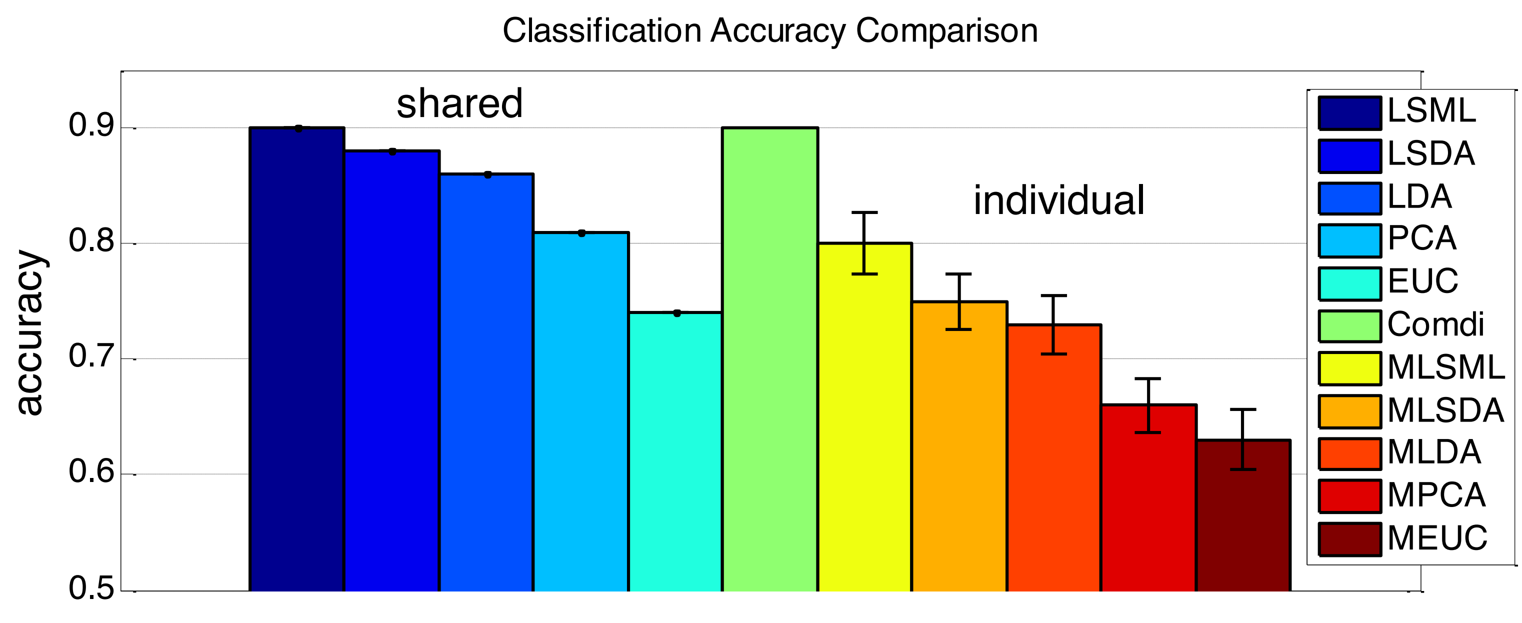
Fei Wang, Jimeng Sun, Shahram Ebadollahi. Integrating Distance Metrics Learned from Multiple Experts and its Application in Inter-Patient Similarity Assessment. SDM 2011.

Comdi: Experimental Evaluation

Data:
- Scale: 135 k
- Aggregation period: 1 year
- Cohorts: 247, select 30
- Feature: HCC codes

Experimental Setting:
- Share versions: learning on all 30 cohorts
- Individual versions: learning on 1 cohort

Observations:
- Shared version perform better than individual versions
- Comdi is comparable to LSML, which is the best among sharing versions



Applications in Health Informatics

- Patient Similarity Learning
- Risk Factor Identification
- Clinical Pattern Detection

Scalable Orthogonal Regression

- Scalable Orthogonal Regression is to minimize

$$J(\boldsymbol{\alpha}) = \underbrace{\frac{1}{2}\|\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}\|^2}_{\text{Regression Error}} + \underbrace{\lambda\|\boldsymbol{\alpha}\|_1}_{\text{Sparse Penalty}} + \boxed{\underbrace{\frac{\beta}{4}\sum_{ij}\left(\alpha_i \mathbf{x}_i^T \mathbf{x}_j \alpha_j\right)^2}_{\text{Redundancy Penalty}}}$$

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_p]^T \in \mathbb{R}^p \qquad \mathbf{y} \in \mathbb{R}^n$$

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p] \in \mathbb{R}^{n \times p}$$

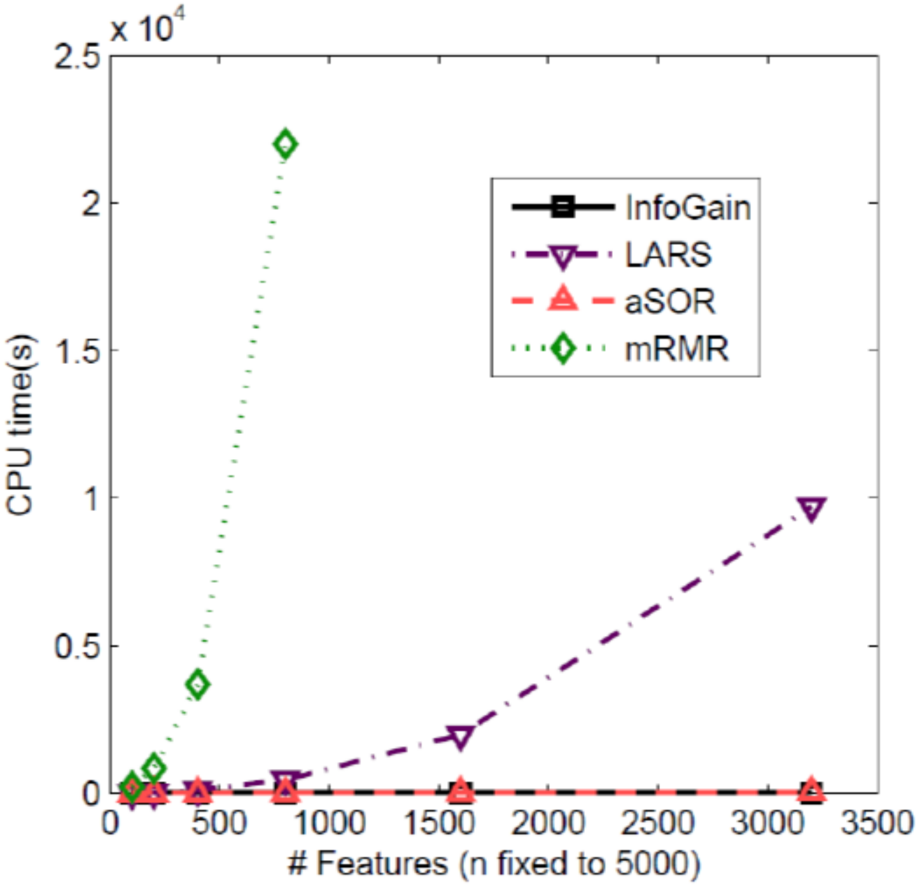
λ and β are model parameters and assume x_j is normalized, $j = 1,2 \ldots p$.

- Notice that the sparse penalty is non-smooth --Generating sparse solution of α.
- Feature selection
  - If $\alpha_j \neq 0$, feature j is selected, $j = 1,2, \ldots, p$, where p is the number of feature
  - For those $\alpha_j \neq 0$, rank the features according to $|\alpha_j|$.

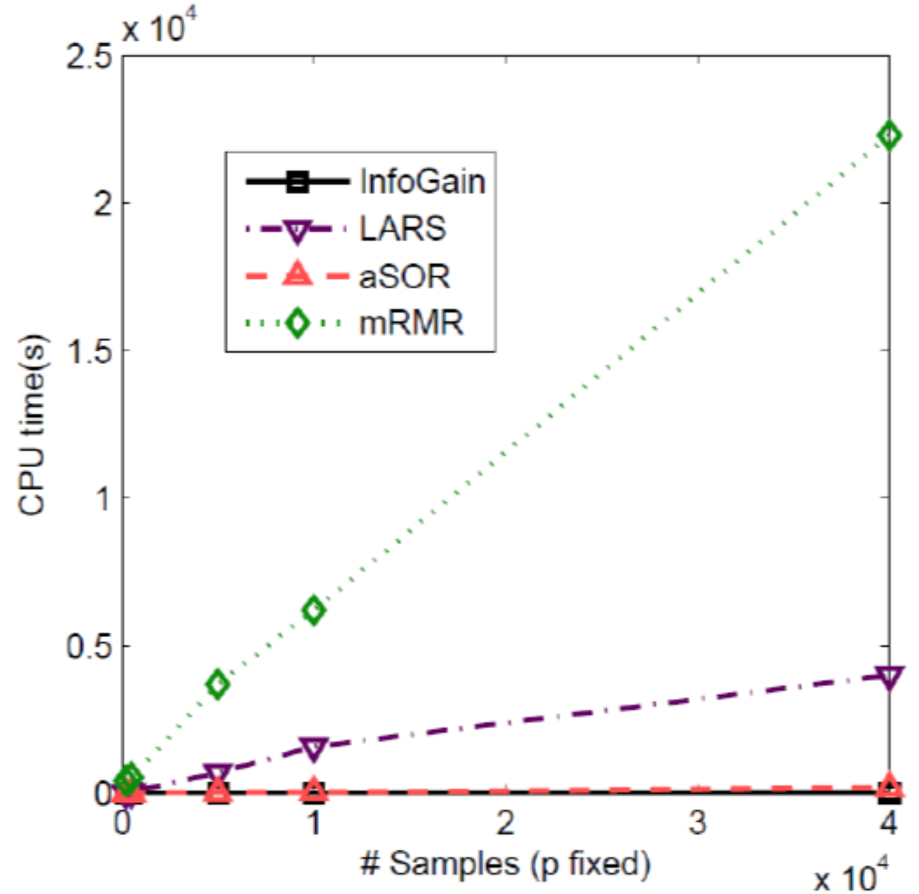
Dijun Luo, Fei Wang, Jimeng Sun, Marianthi Markatou, Jianying Hu, Shahram Ebadollahi. SOR: Scalable Orthogonal Regression for Low-Redundancy Feature Selection and its Healthcare Applications. SDM 2012.

Scalability Comparison

- Metric: Computational time

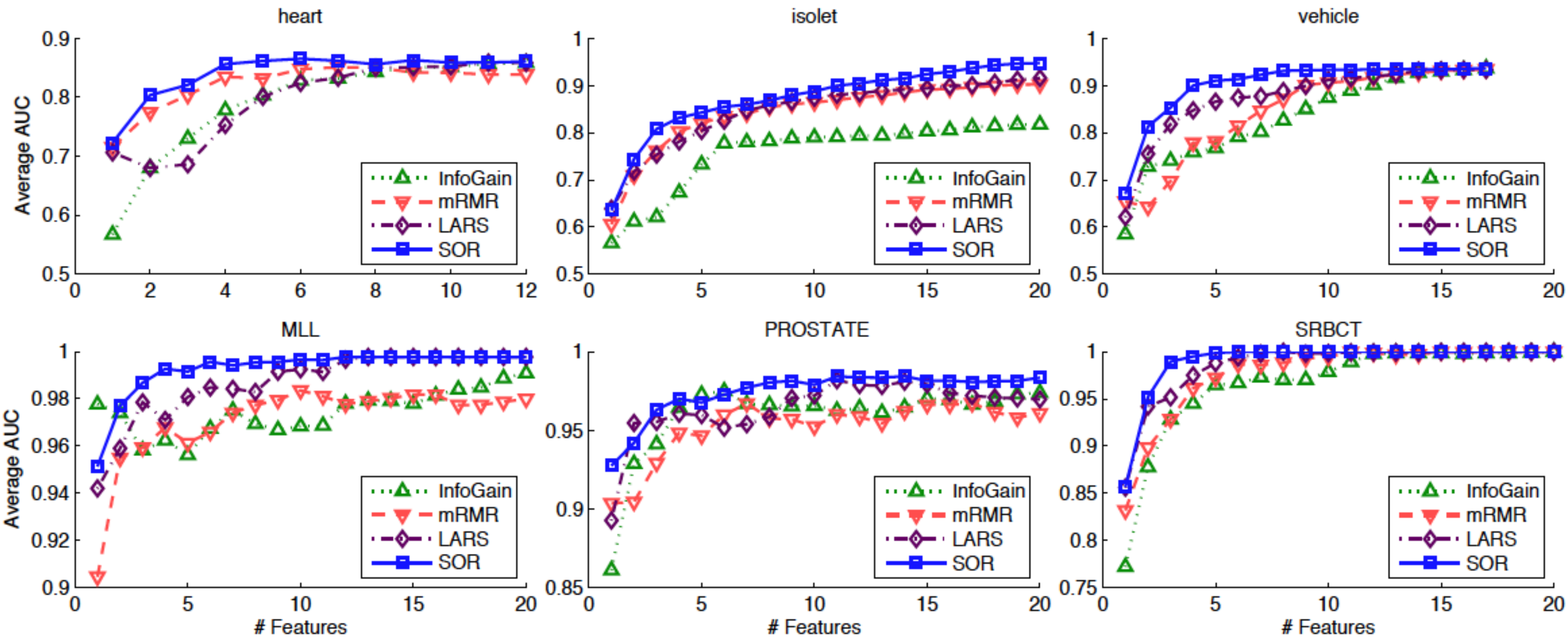


Fixing n



Fixing p

AUC Comparison



Augmented SOR

- SOR with pre-selected feature set

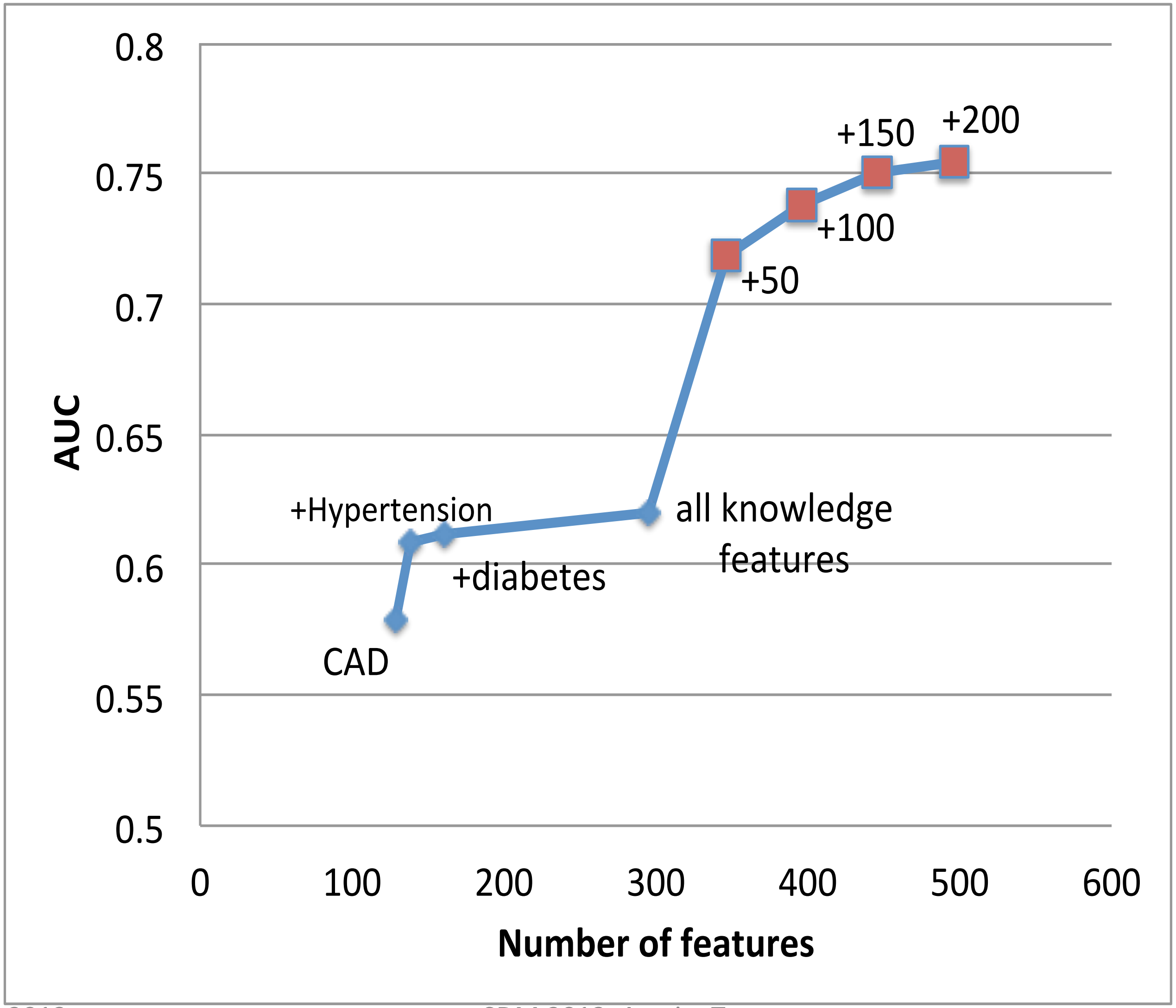
$$f_p(\boldsymbol{\alpha}_Q) = \frac{1}{2}\|\mathbf{y} - \mathbf{X}_Q\boldsymbol{\alpha}_Q\|^2 + \lambda \|\boldsymbol{\alpha}\|_1$$

$$+\frac{\beta}{4}\left[\underbrace{\sum_{ij\in\mathcal{Q}} (\alpha_i \mathbf{x}_i^T \mathbf{x}_j \alpha_j)^2}_{\text{Redundancy among features to be selected}} + \underbrace{\sum_{i\in\mathcal{Q}, j\in\mathcal{P}} (\alpha_i \mathbf{x}_i^T \mathbf{x}_j \alpha_j)^2}_{\text{Redundancy between preselected and features to be selected}}\right]$$

$$\boldsymbol{\alpha}_{\mathcal{P}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{X}_{\mathcal{P}}\boldsymbol{\alpha}\|^2 = (\mathbf{X}_{\mathcal{P}}^T\mathbf{X}_{\mathcal{P}})^{-1}\mathbf{X}_{\mathcal{P}}^T\mathbf{y}$$

- P: pre-selected feature set
- Q: Feature set to be selected from
- Algorithms of SOR and aSOR still apply
  - With different computation of the gradient

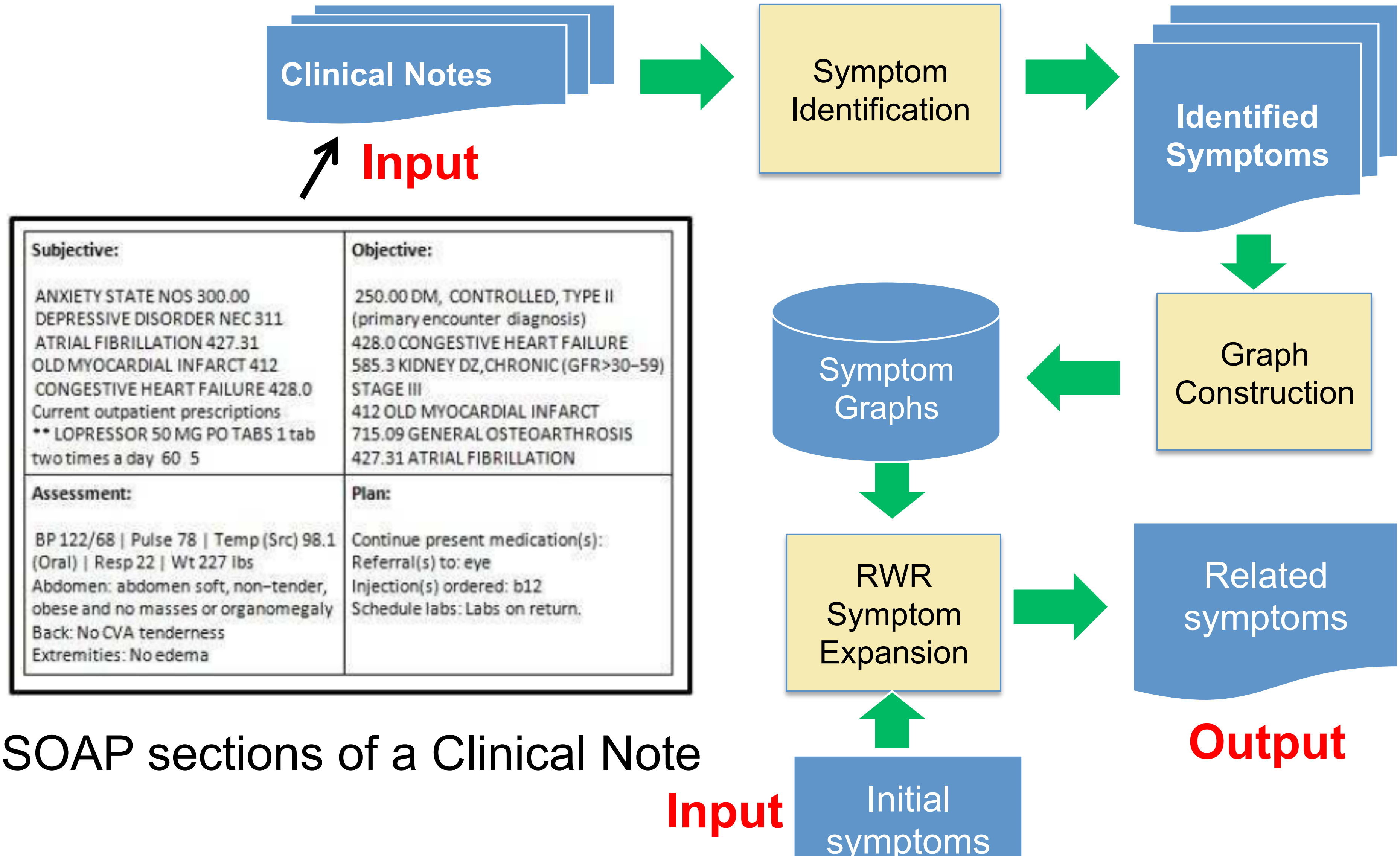
Performance of aSOR



the City College of New York

IBM

Finding Relevance: Mining Clinical Notes



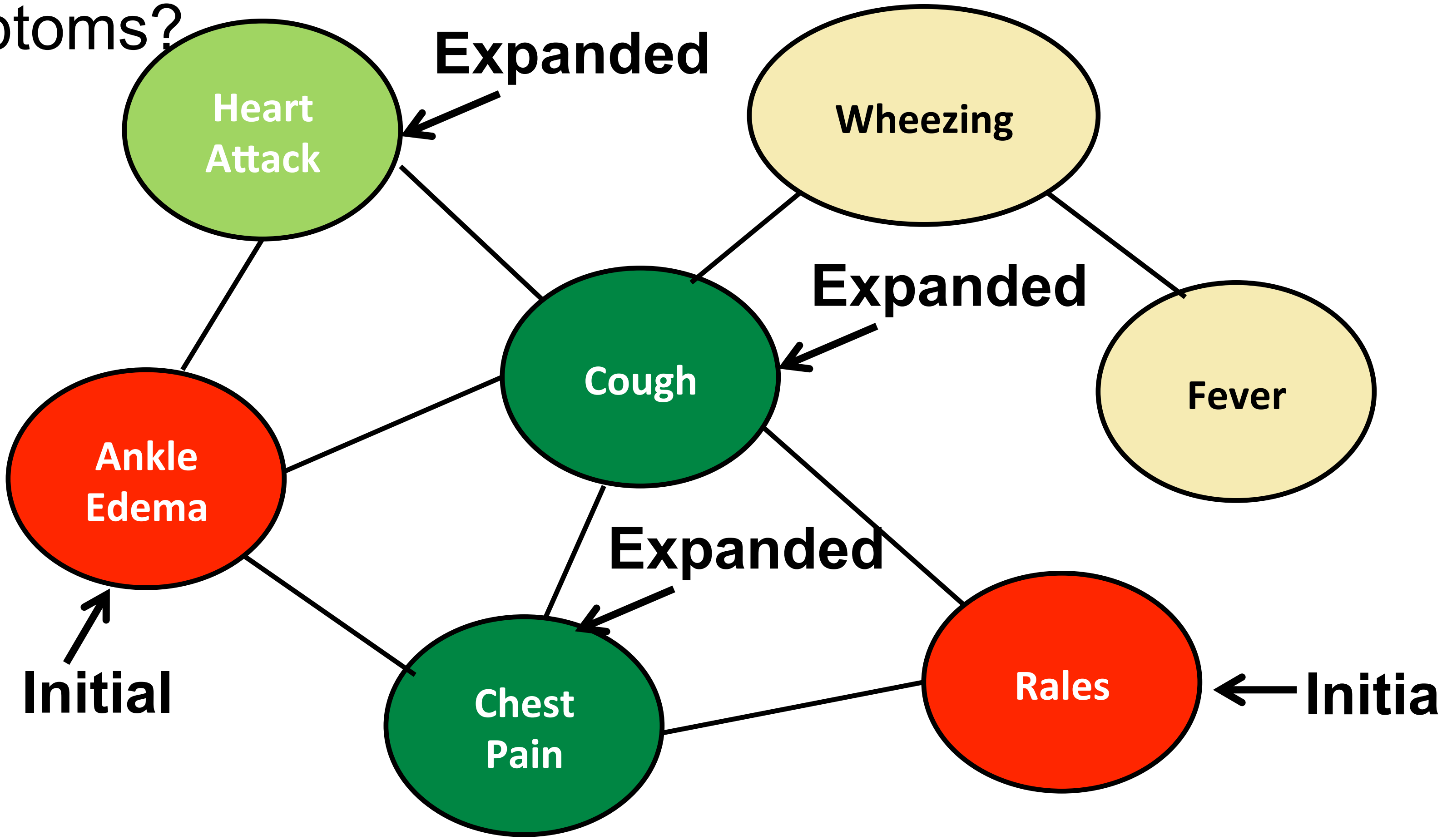
Subjective:	Objective:
ANXIETY STATE NOS 300.00 DEPRESSIVE DISORDER NEC 311 ATRIAL FIBRILLATION 427.31 OLD MYOCARDIAL INFARCT 412 CONGESTIVE HEART FAILURE 428.0 Current outpatient prescriptions ** LOPRESSOR 50 MG PO TABS 1 tab two times a day 60 5	250.00 DM, CONTROLLED, TYPE II (primary encounter diagnosis) 428.0 CONGESTIVE HEART FAILURE 585.3 KIDNEY DZ,CHRONIC (GFR>30-59) STAGE III 412 OLD MYOCARDIAL INFARCT 715.09 GENERAL OSTEOARTHROSIS 427.31 ATRIAL FIBRILLATION
Assessment:	**Plan:**
BP 122/68 | Pulse 78 | Temp (Src) 98.1 (Oral) | Resp 22 | Wt 227 lbs Abdomen: abdomen soft, non-tender, obese and no masses or organomegaly Back: No CVA tenderness Extremities: No edema	Continue present medication(s): Referral(s) to: eye Injection(s) ordered: b12 Schedule labs: Labs on return.

SOAP sections of a Clinical Note

P. Sondhi, J. Sun, H. Tong, C. Zhai: SympGraph: a framework for mining clinical notes through symptom relation graphs. KDD 2012

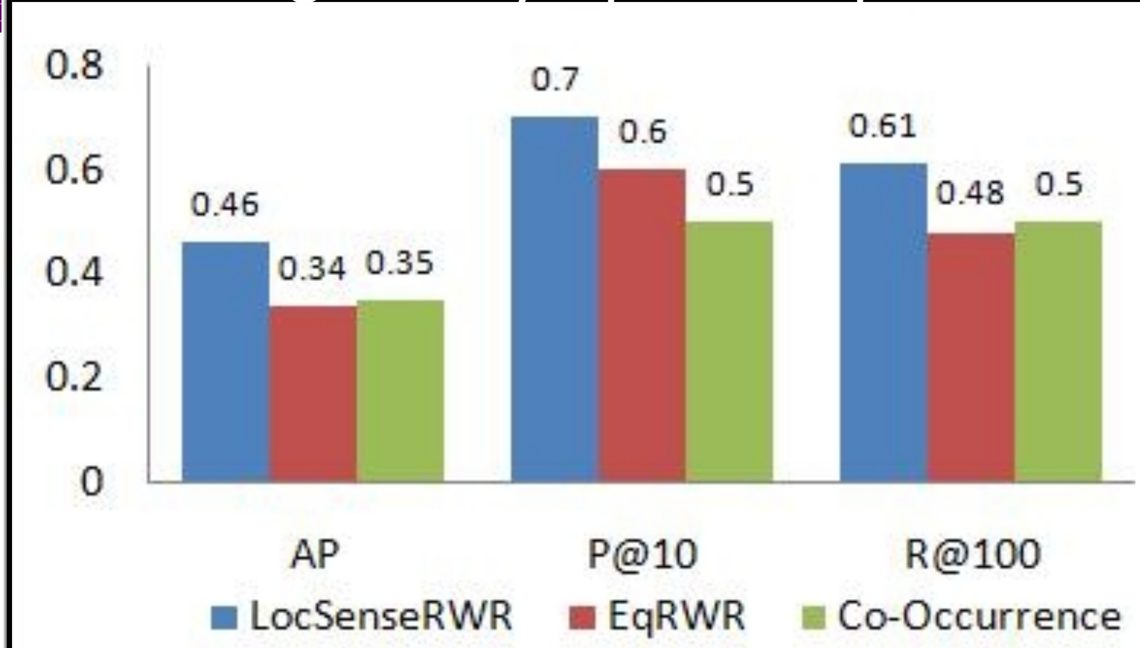
Mining Clinical Notes: Symptom Expansion

Key Idea: Symptom Expansion → graph node proximity.
i.e., which symptoms are most relevant to initial symptoms?



Evaluations

Framingham Symptom Expansion



[Parikshit+ KDD 2012]

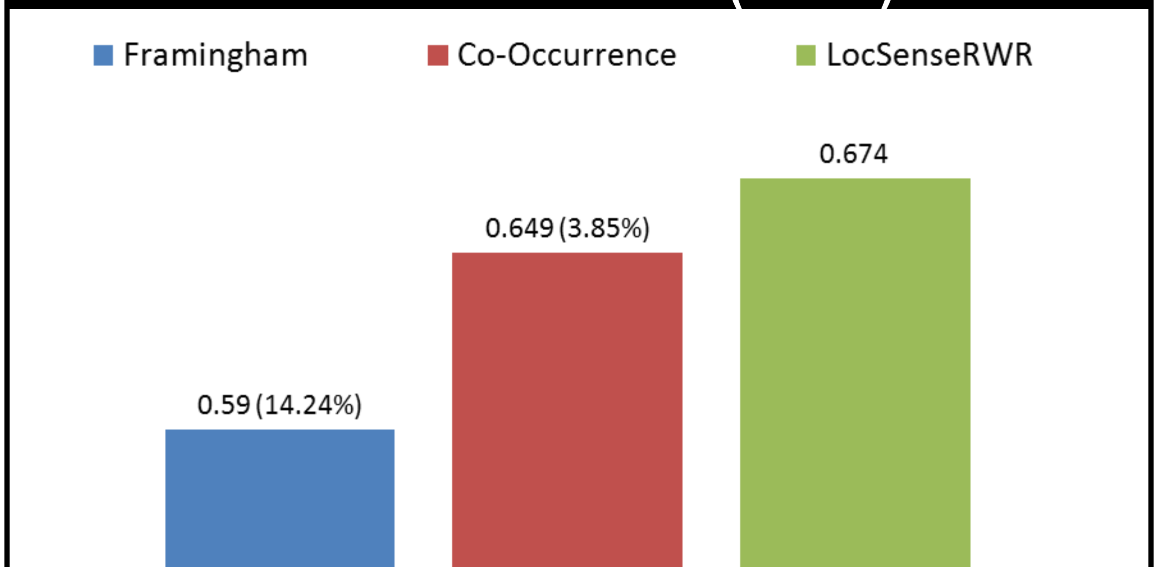
Evaluation Details:

Experts: 2; 175 symptoms judged

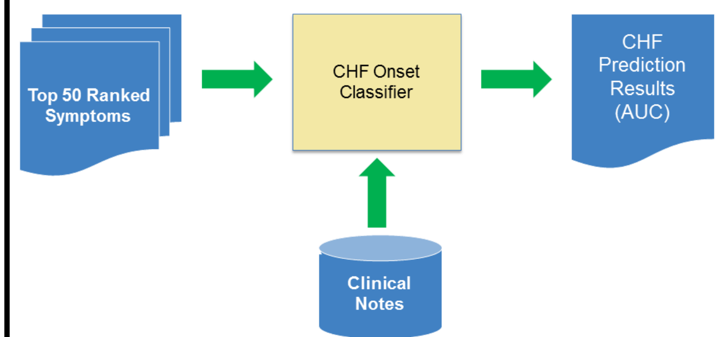
Relevant: 72 , Irrelevant: 103
Inter-annotator agreement : 81.8%

Symptoms labeled as related by both experts were considered as relevant.

CHF Prediction (AUC)



Evaluation Details:

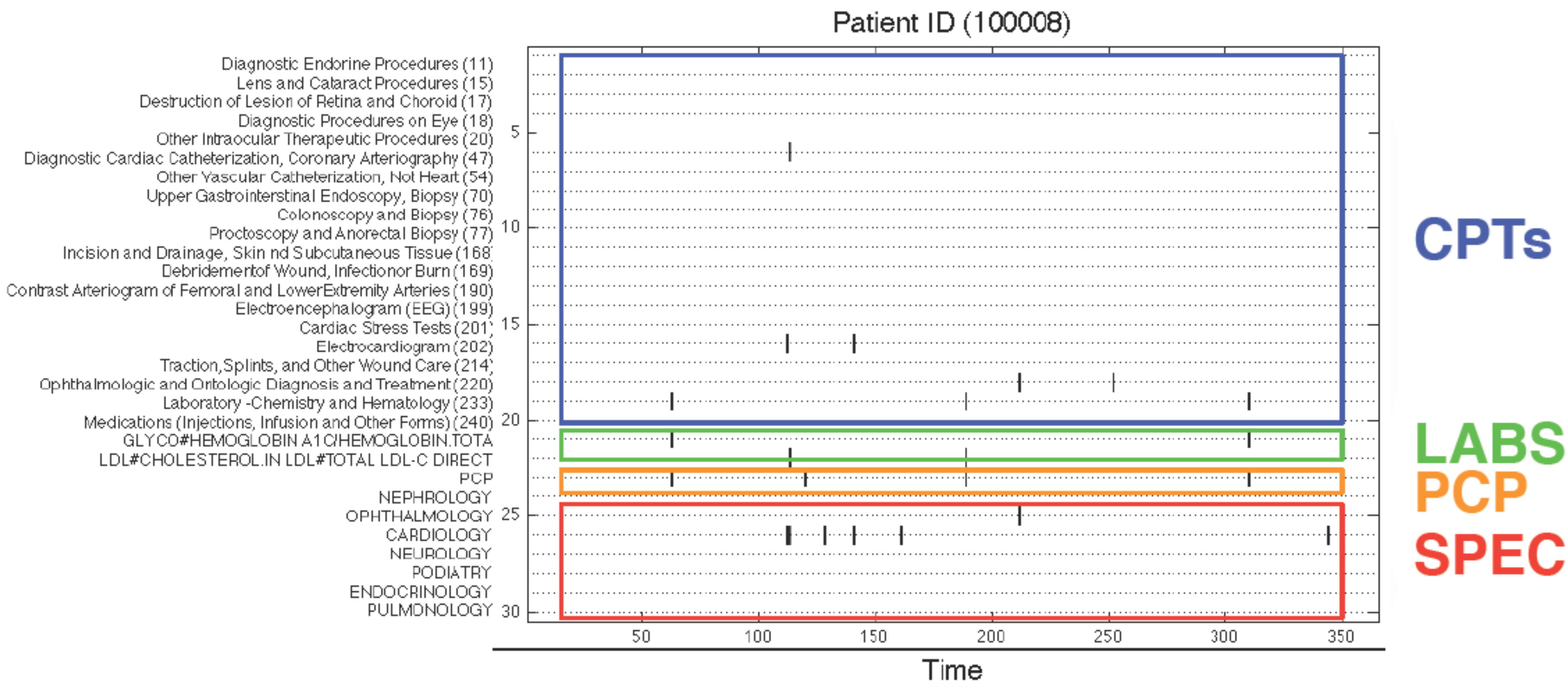


CHF: affecting 1 out of 5 adults in US; most costly in CMS Framingham, 1971 → 50s, 60s

Applications in Health Informatics

- Patient Similarity Learning
- Risk Factor Identification
- Clinical Pattern Detection

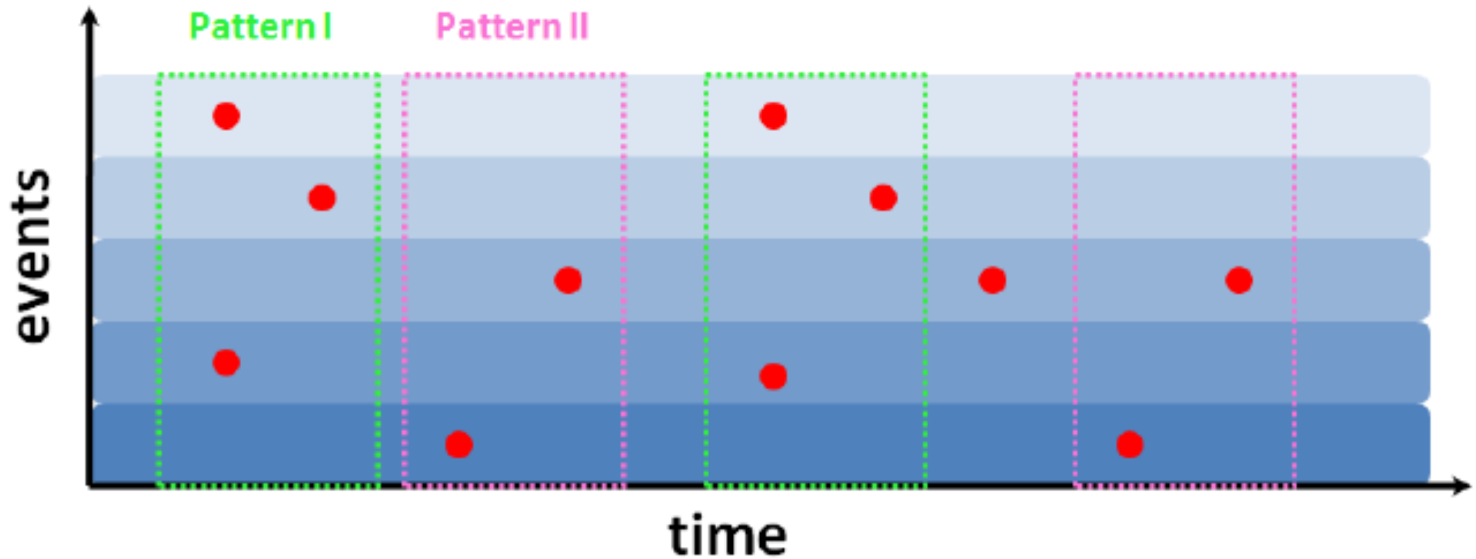
Matrix Representation of a Patient



Fei Wang, Noah Lee, Jianying Hu, Jimeng Sun, Shahram Ebadollahi. Towards Heterogeneous Temporal Clinical Event Pattern Discovery:A Convolutional Approach. KDD 2012.

Fei Wang, Noah Lee, Jianying Hu, Jimeng Sun, Shahram Ebadollahi. A Framework for Mining Signatures from Event Sequences and Its Applications in Healthcare Data. TPAMI 2012.

Temporal Patterns in Longitudinal Patient Records

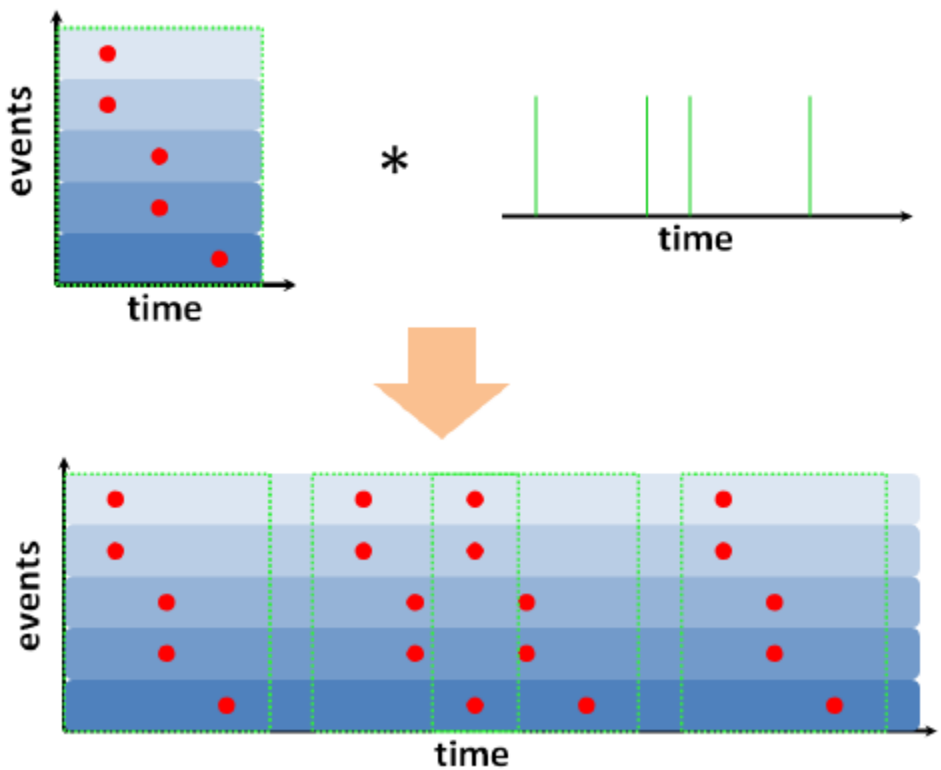


One-Side Convolution

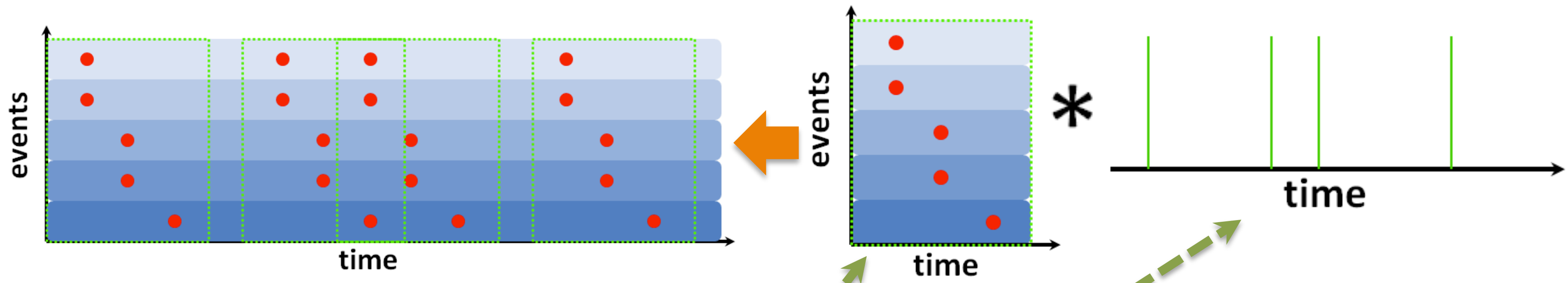
Definition (One-Side Convolution). *The one-side convolution of* $\mathbf{F} \in \mathbb{R}^{n\times m}$ *and* $\mathbf{g} \in \mathbb{R}^{t\times 1}$ *is an* $n \times t$ *matrix with*

$$(\mathbf{F} * \mathbf{g})_{ij} = \sum\nolimits_{k=1}^{t} g_{j-k+1} F_{ik}$$

Note that $g_j = 0$ *if* $j \leqslant 0$ *or* $j > t$*, and* $F_{ik} = 0$ *if* $k > m$.



One-Side Convolutional NMF



$$\min_{\mathcal{F},\mathcal{G}} \quad \mathcal{J}$$

$$s.t. \quad \forall r = 1, \cdots, R; c = 1, \cdots, C$$

$$\mathbf{F}^{(r)} \geqslant 0, \mathbf{g}_c^{(r)} \geqslant 0$$

$$\mathcal{J} = \sum_{c=1}^{C} d_\beta \left(\mathbf{A}_c \odot \mathbf{X}_c, \mathbf{A}_c \odot \left(\sum_{r=1}^{R} \mathbf{F}^{(r)} * \mathbf{g}_c^{(r)} \right) \right) + \lambda_1 \sum_{r=1}^{R} \|\mathbf{F}^{(r)}\|_1 + \lambda_2 \sum_{c=1}^{C} \sum_{r=1}^{R} \|\mathbf{g}_c^{(r)}\|_1$$

Definition (β-divergence) *The β-divergence between two matrices $\mathbf{A}$ and $\mathbf{B}$ with the same size is*

$$d_\beta(\mathbf{A}, \mathbf{B}) = \frac{1}{\beta(\beta - 1)} \sum_{ij} \left(A_{ij}^\beta + (\beta - 1) B_{ij}^\beta - \beta A_{ij} B_{ij}^{\beta - 1} \right)$$

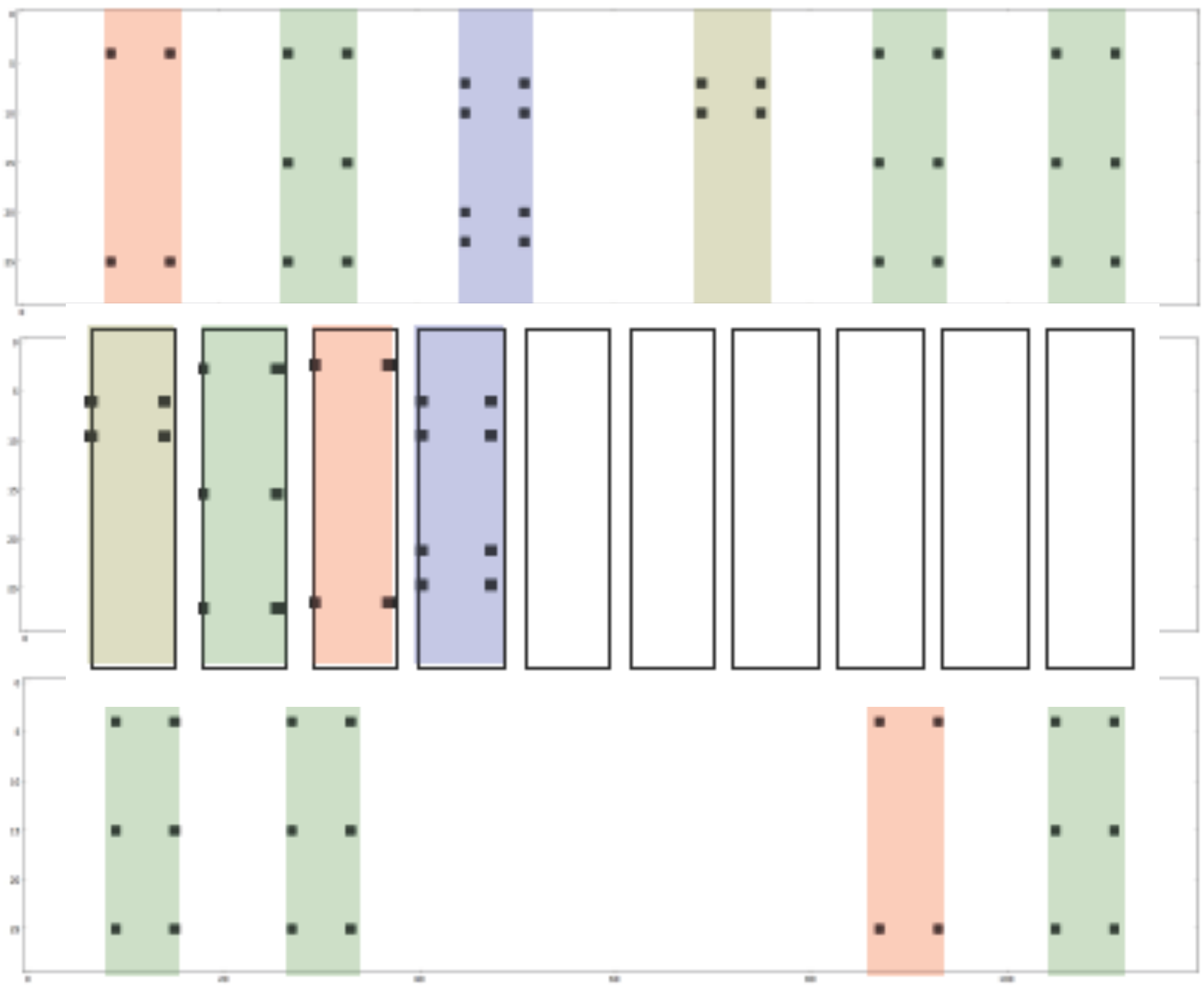
Multiplicative Updates

$$F_{ik}^{(r)} \leftarrow F_{ik}^{(r)} \left(\frac{\sum_{c=1}^{C} \sum_{j=1}^{t} A_{c_{ij}}^{\beta-1} X_{c_{ij}} Y_{c_{ij}}^{\beta-2} g_{c_{j-k+1}}^{(r)}}{\sum_{c=1}^{C} \sum_{j=1}^{t} A_{c_{ij}} Y_{c_{ij}}^{\beta-1} g_{c_{j-k+1}}^{(r)} + \lambda_1} \right)^{\eta(\beta)}$$

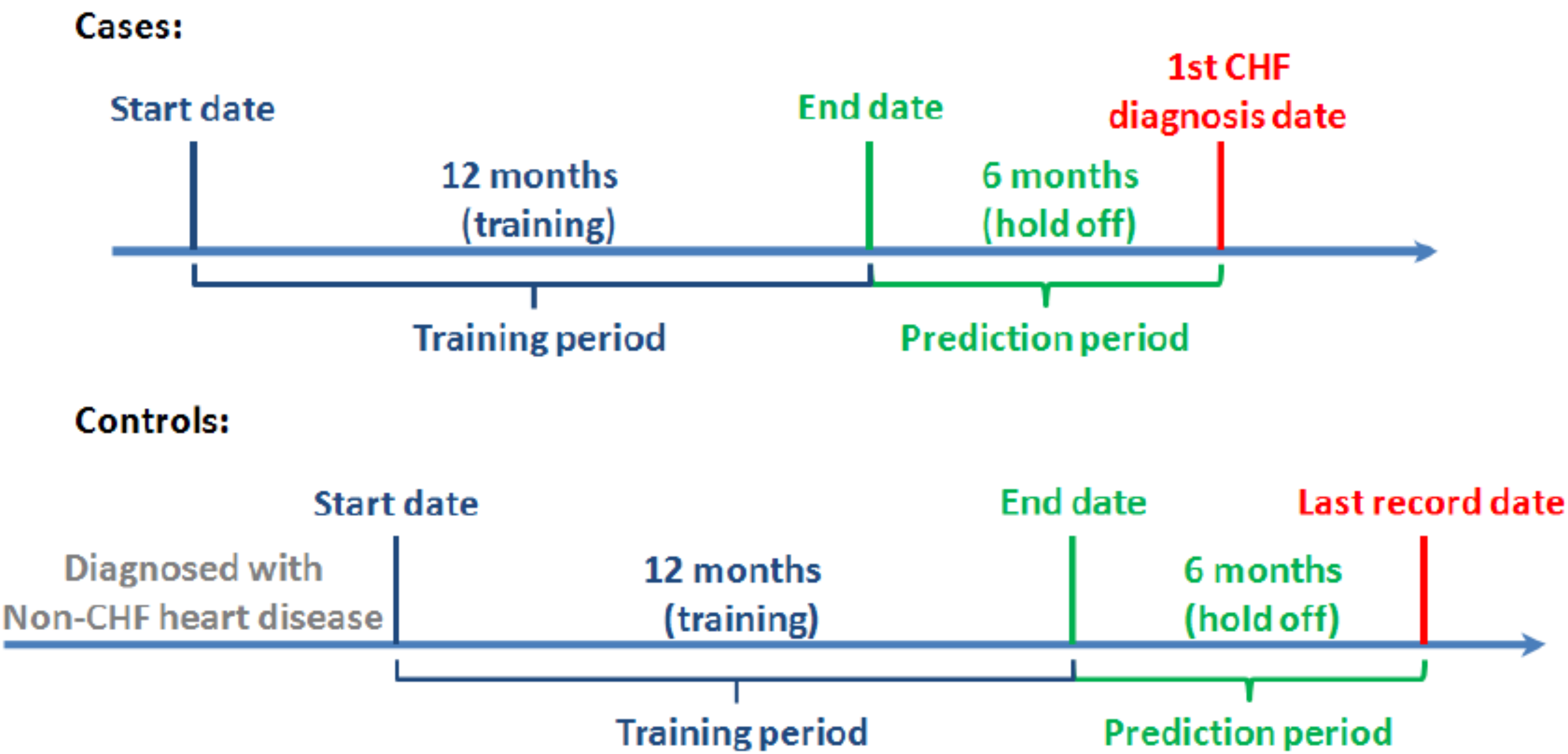
$$g_{c_k}^{(r)} \leftarrow g_c^{(r)} \left(\frac{\sum_{i=1}^{n} \sum_{j=1}^{t} A_{c_{ij}}^{\beta-1} X_{c_{ij}} Y_{c_{ij}}^{\beta-2} F_{i,j-k+1}^{(r)}}{\sum_{i=1}^{n} \sum_{j=1}^{t} A_{c_{ij}} Y_{c_{ij}}^{\beta-1} F_{i,j-k+1}^{(r)} + \lambda_2} \right)^{\eta(\beta)}$$

$$\eta(\beta) = \begin{cases} \frac{1}{2-\beta}, & \beta < 1 \\ 1, & 1 \leqslant \beta \leqslant 2 \\ \frac{1}{\beta-1}, & \beta > 2 \end{cases}$$

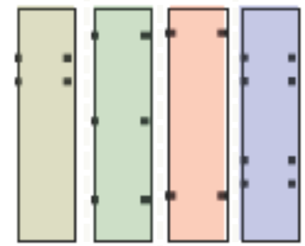
A Synthetic Example

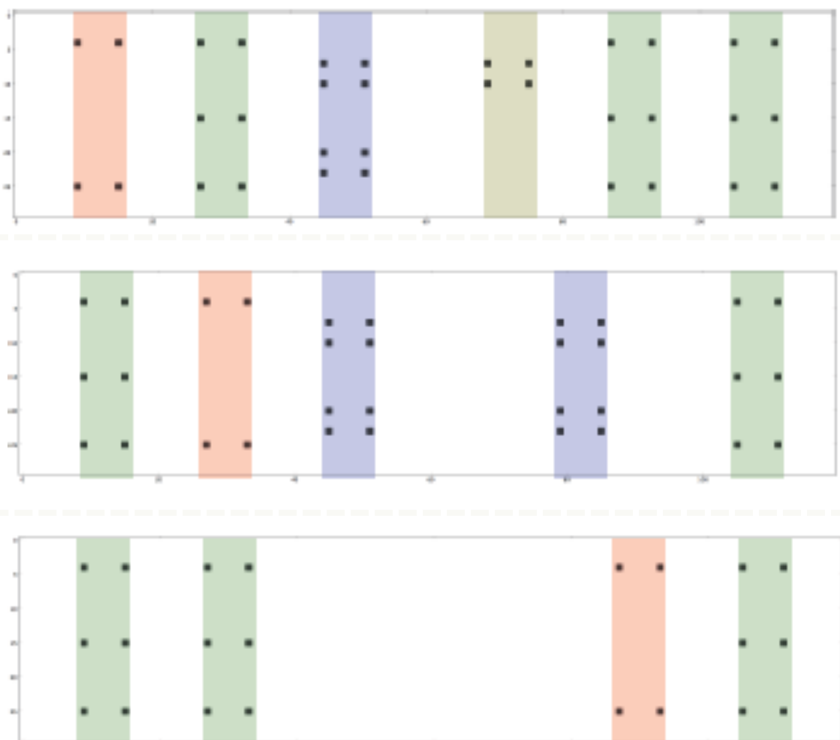


An Evaluation Case



Bag-of-Pattern Representation



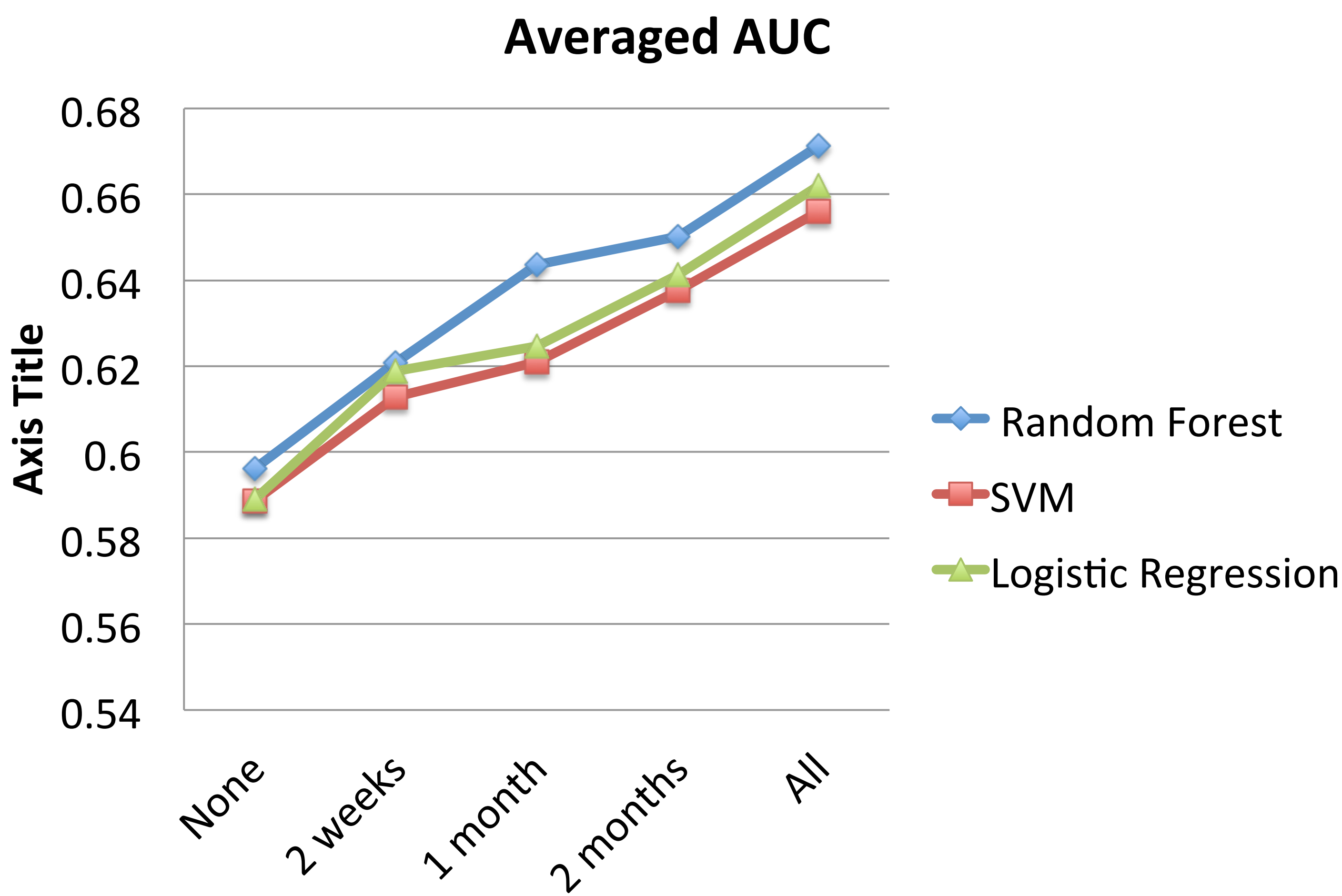


[1 3 1 1]

[0 2 1 2]

[0 3 1 0]

Results


Averaged AUC
0.68
0.66
0.64
0.62
0.6
0.58
0.56
0.54
Axis Title
None
2 weeks
1 month
2 months
All
Random Forest
SVM
Logistic Regression

Outline

- Introduction
- Overview of the Technologies
- Applications in Health Informatics
- Applications in Social Informatics
- Conclusions and Future Works

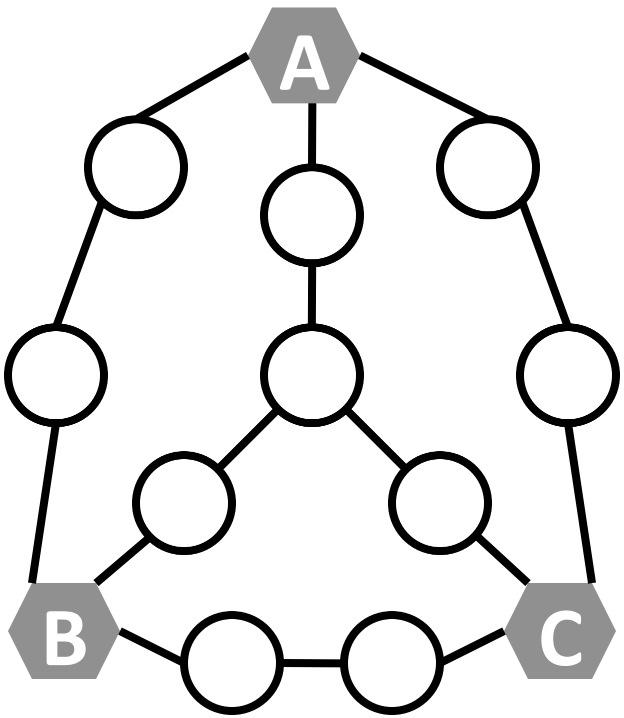
Applications in Social Informatics

- Finding Complex User Patterns
- (Matrix-based) Anomaly Detection
- Influence and Virus Propagation

Finding Commonality: Center-Piece Subgraph Discovery

[Tong+ KDD06, VLDB06, TKDE13]

- **Given**: a graph *W*, and a query set
- **Find**: the most central node (wrt the query set)

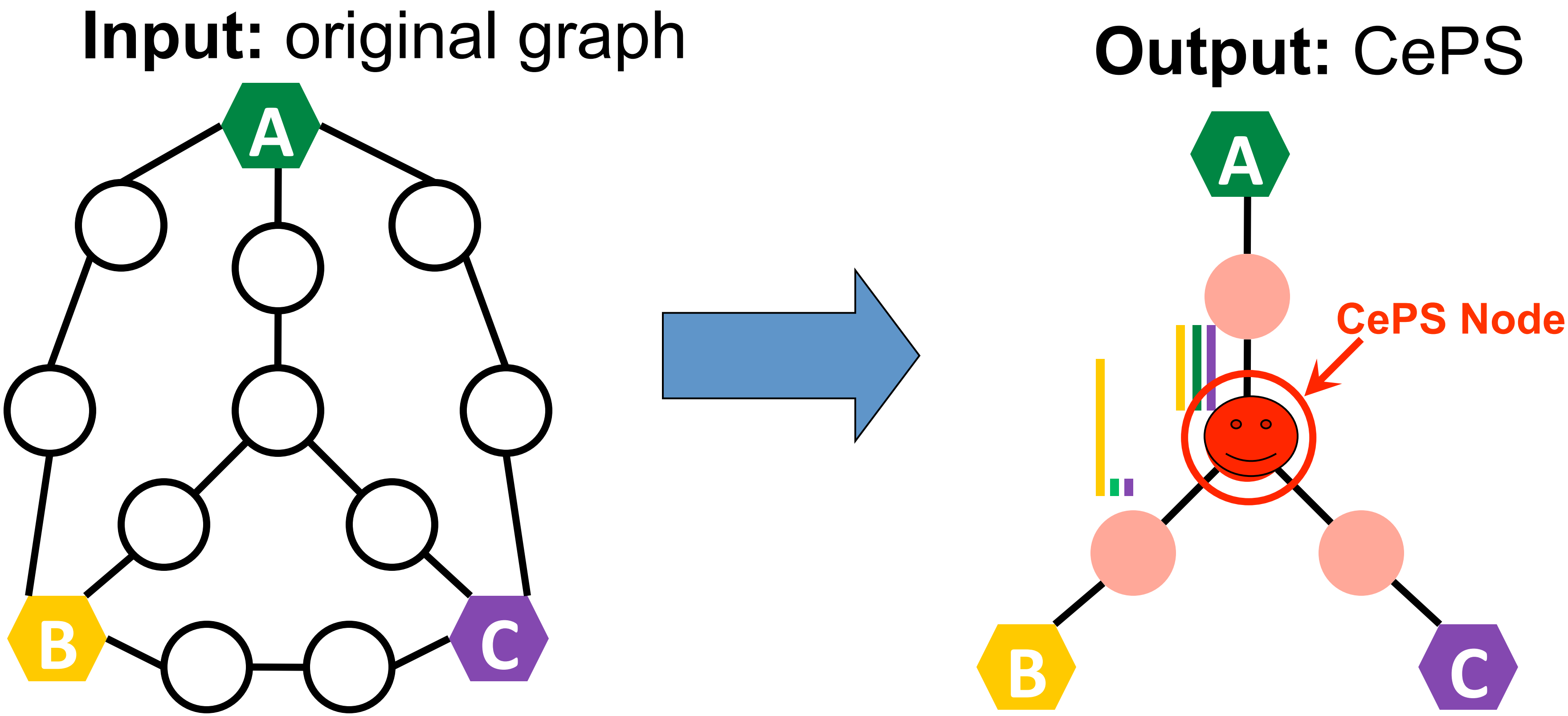


Q: Who is the most central node wrt the black nodes?

(e.g., master-mind criminal, common advisor/collaborator, etc)

Center-Piece Subgraph Discovery

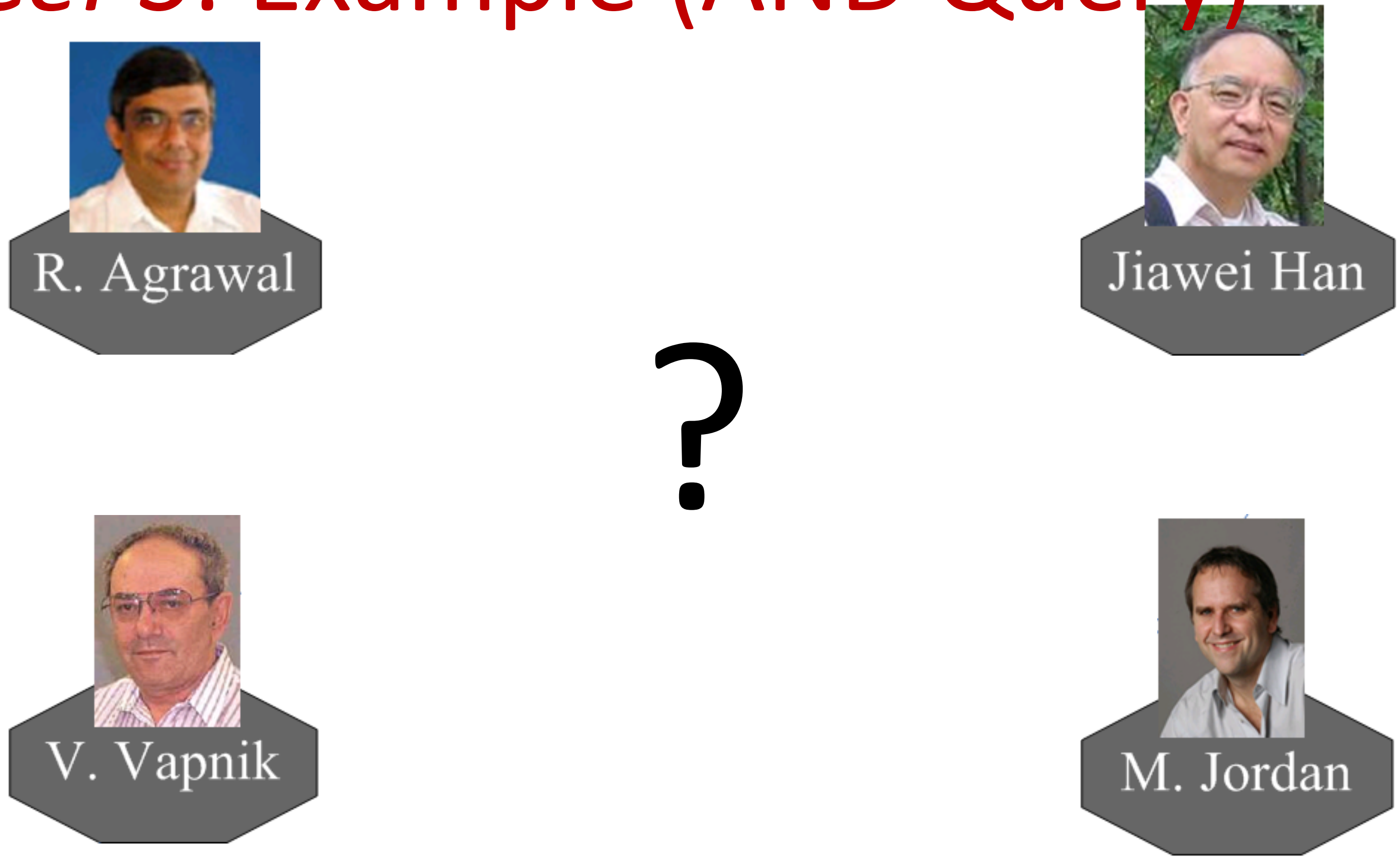
[Tong+ KDD06, VLDB06, TKDE13]



Q: How to find hub for nodes A, B, C?

Our Solution: **Max** (Prox(A, Red) x Prox(B, Red) x Prox(C, Red))

CePS: Example (AND Query)

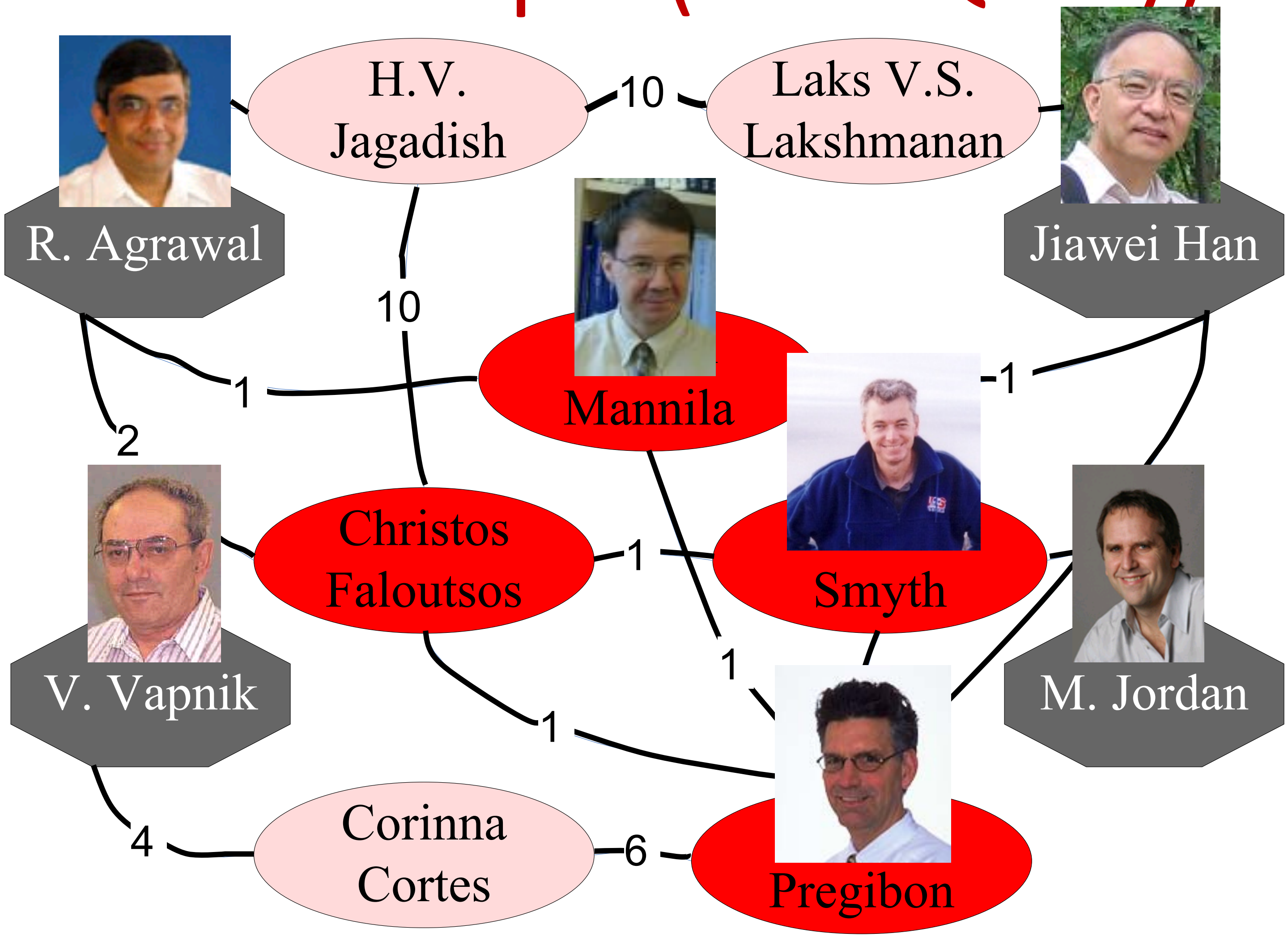


DBLP co-authorship network:

-400,000 authors, 2,000,000 edges

Code at: http://www.cs.cmu.edu/~htong/soft.htm

CePS: Example (AND Query)



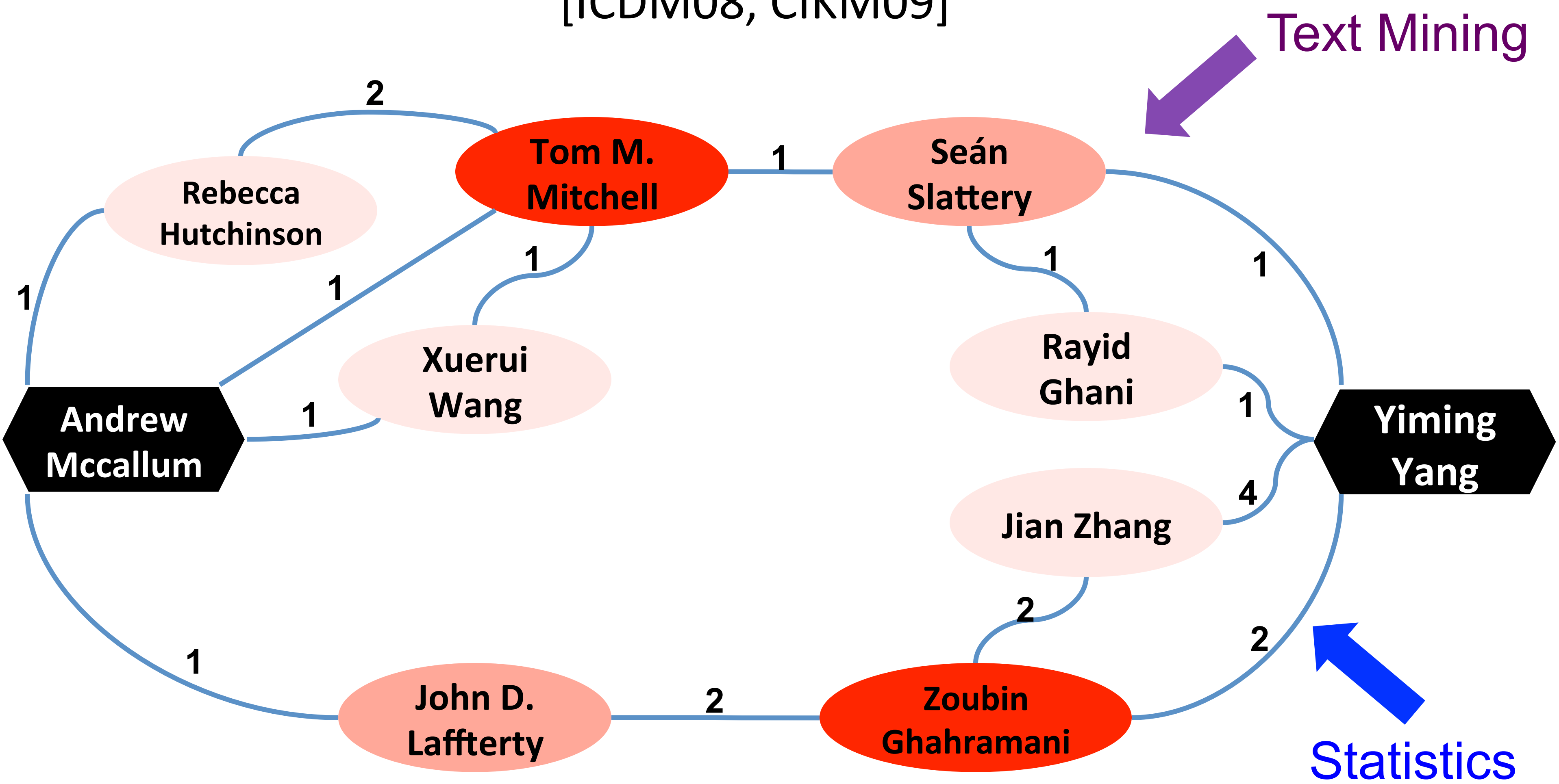
DBLP co-authorship network:

-400,000 authors, 2,000,000 edges

Code at: http://www.cs.cmu.edu/~htong/soft.htm

Negation: CePS - Initial Result

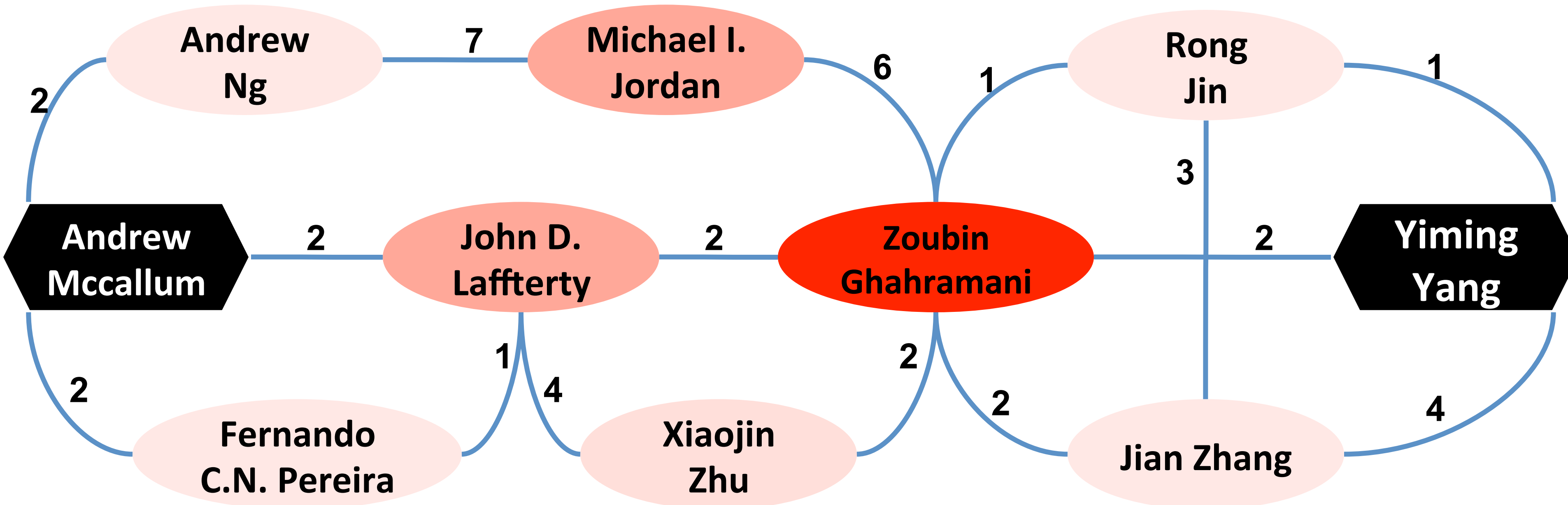
[ICDM08, CIKM09]



CePS between “Andrew Mccallum” and “Yiming Yang”

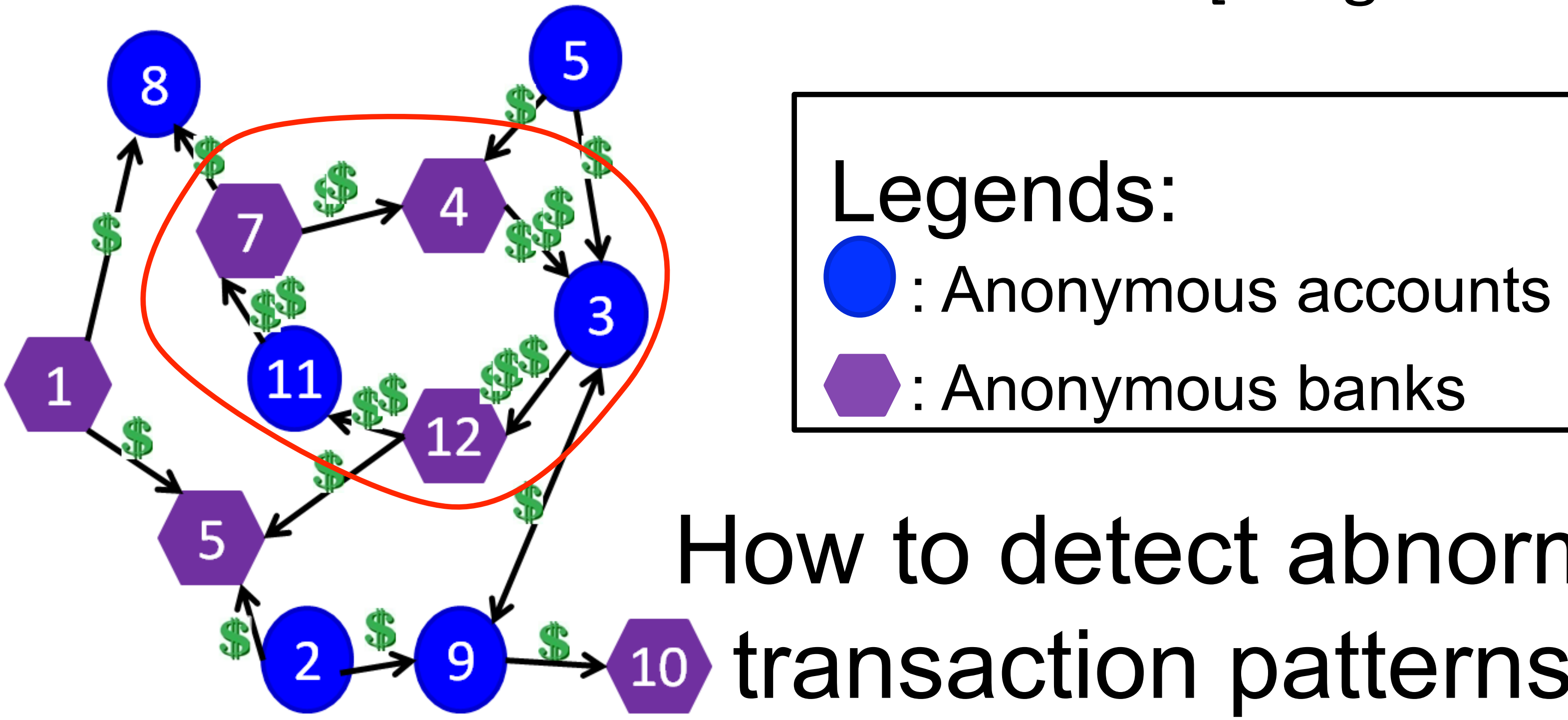
Negation: CePS – After Feedback

[ICDM08, CIKM09]



CePS between "Mccallum" and "Yang", avoiding "Mitchell"
entire 'Text' connection gone, and more connections on 'Statistics'

Best-Effort Pattern Match [Tong+ KDD 2007]



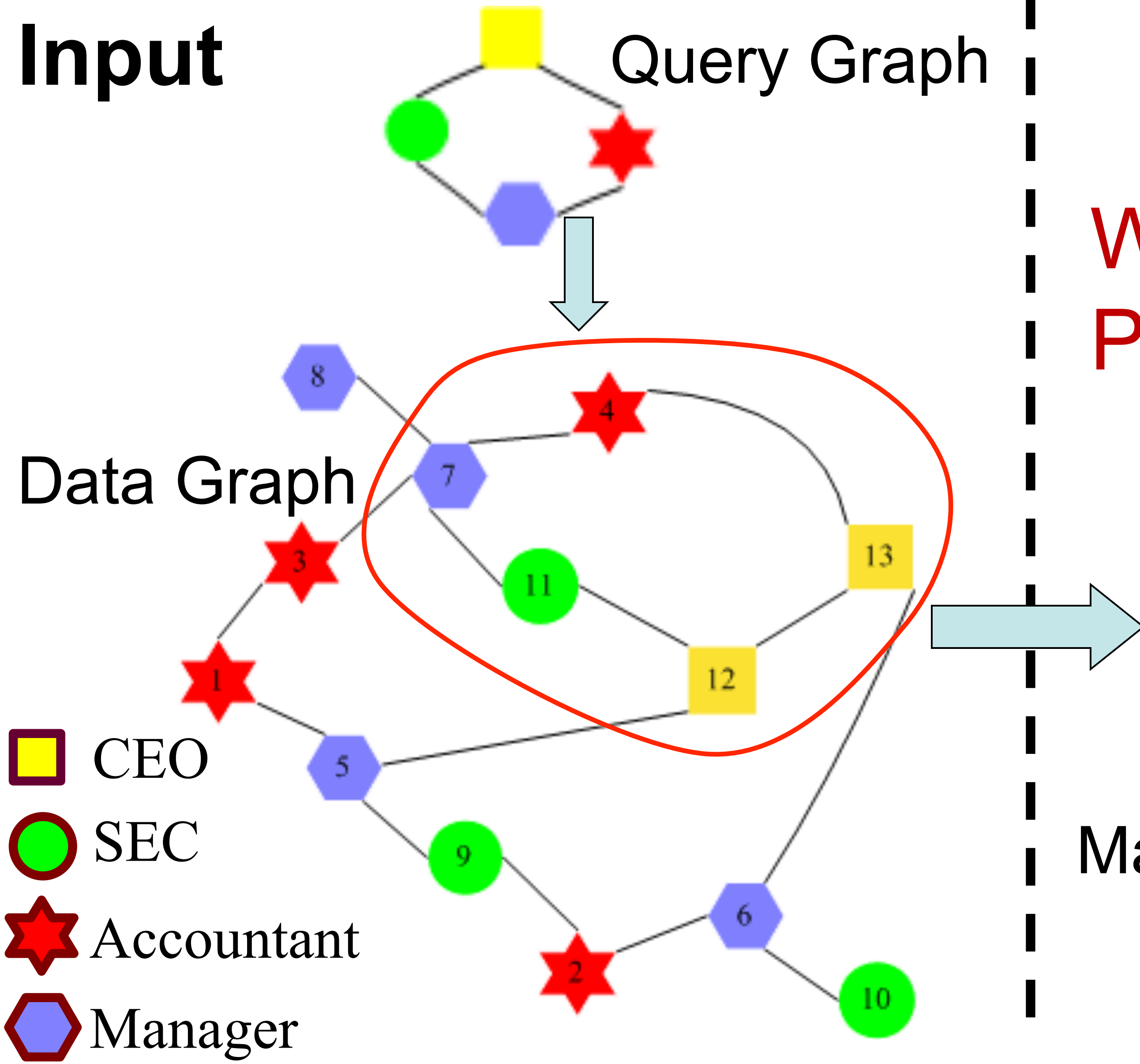
How to detect abnormal transaction patterns?
(e.g., money-laundering ring)

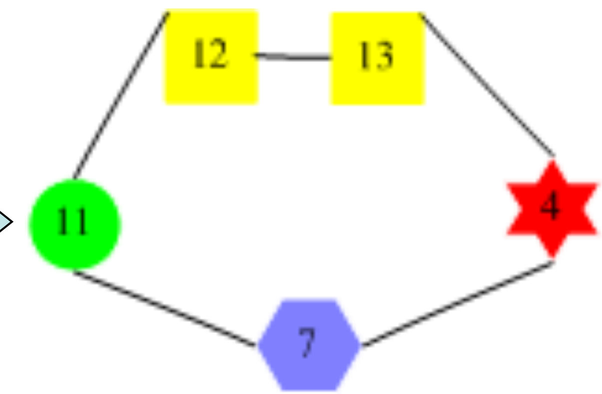
- 7.5% of U.S. adults lost money for financial fraud
- 50%+ US corporations lost >= $500,000
  - e.g., Enron ($70bn) [Albrecht+ 2001]
- Total cost of financial fraud: $1trillion [Ansari 2006]

Input

Output

W4. Best-Effort Pattern Match



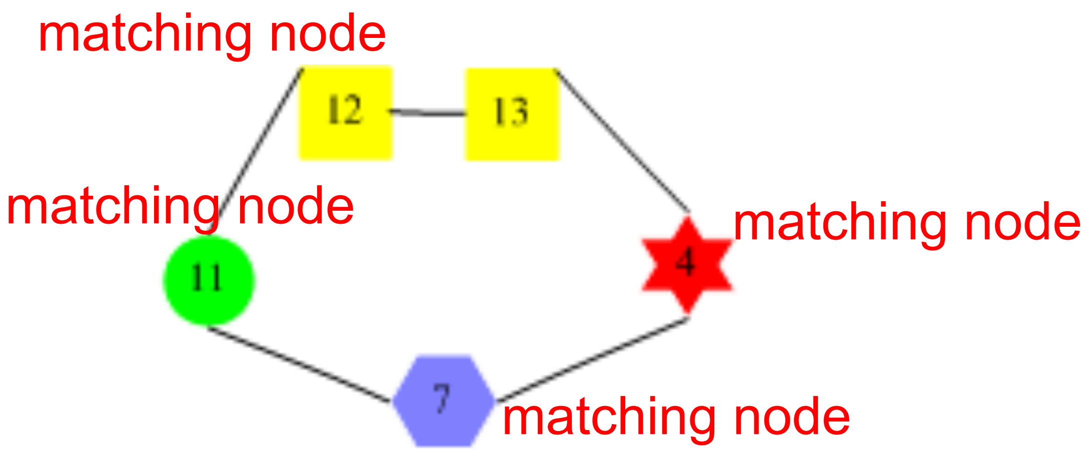


Matching Subgraph

Q: How to find matching subgraph?

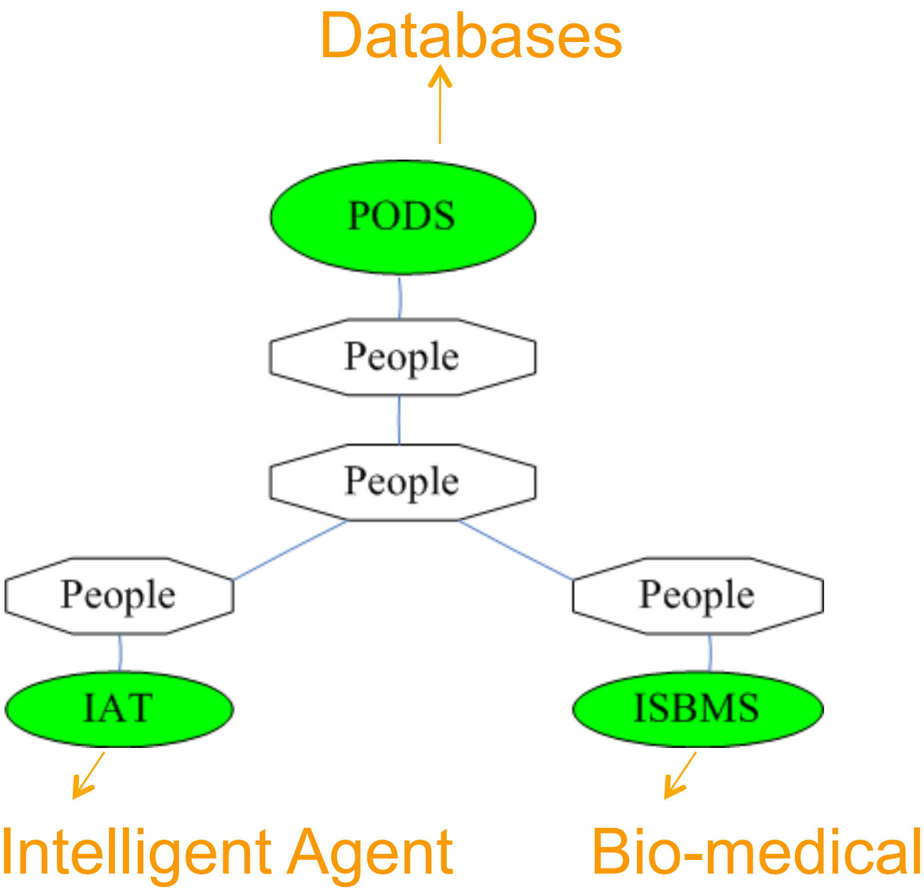
A: Proximity! [Tong+ KDD 2007 b]

G-Ray: How to?

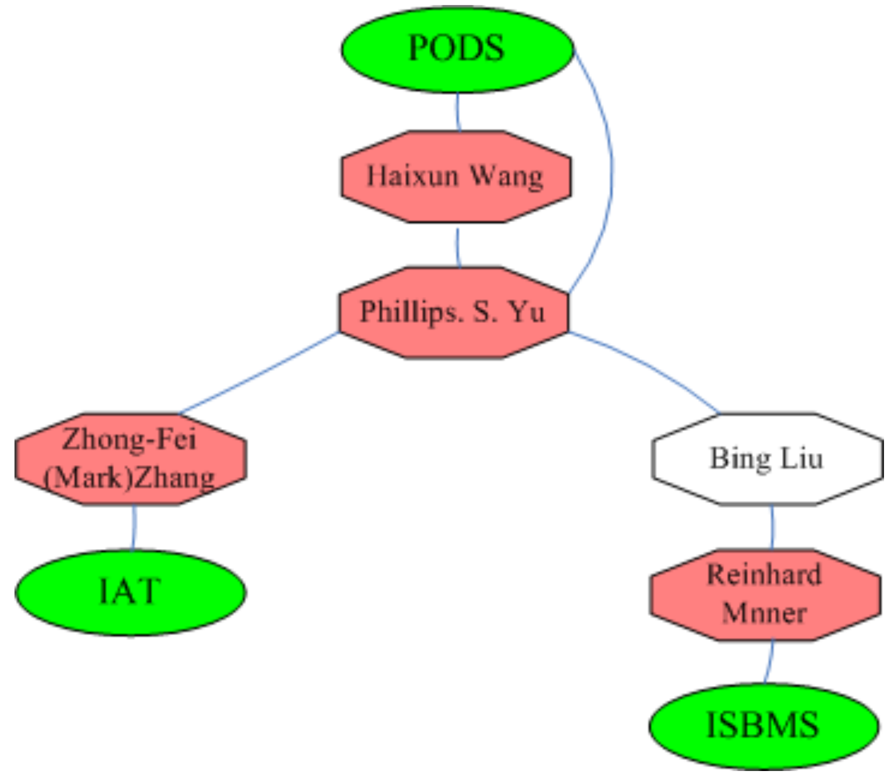


Goodness = Prox (12, 4) x Prox (4, 12) x
Prox (7, 4) x Prox (4, 7) x
Prox (11, 7) x Prox (7, 11) x
Prox (12, 11) x Prox (11, 12)

Effectiveness: star-query



Query

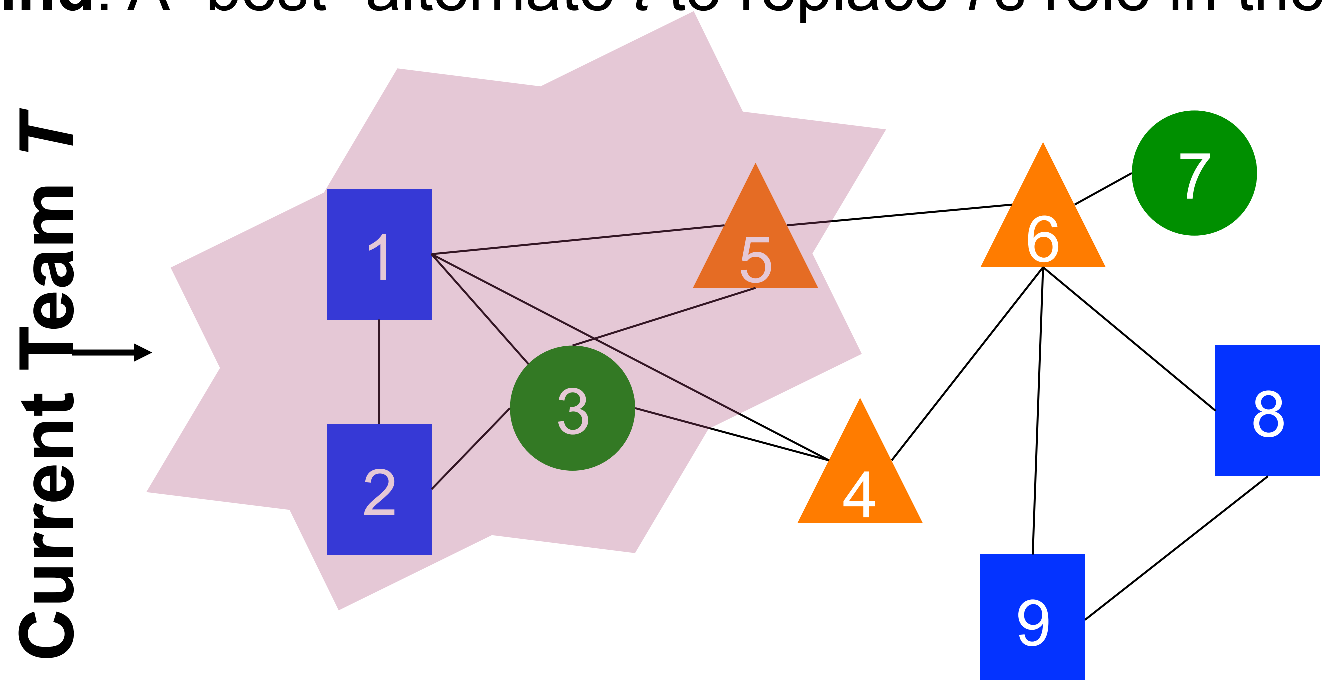


Result

Team Replacement [Tong+ SDM12b, WIDS12]

Problem Definitions

- **Given**: (1) A social network *A*; (2) The skill indicator for each person *S*; (3) a Team *T*; and (4) A team member;
- **Find**: A “best” alternate *t* to replace *i*’s role in the team *T*.



Q: Who is the best alternative of `5’ in *T* ?

A: Team-Aware Similarity!

Team Replacement

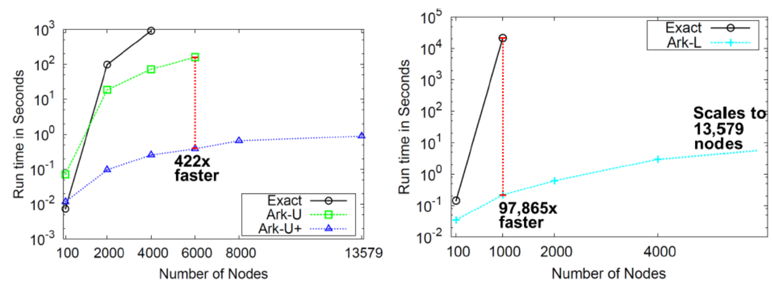
Key Observation: Graph Kernel → Team-Aware Similarity

$$t = \mathrm{argmax}_{j, j \notin \mathcal{T}} \ \mathrm{Ker}(\mathbf{A}(\mathcal{T}, \mathcal{T}), \mathbf{A}(\mathcal{T}_{i_0 \to j}, \mathcal{T}_{i_0 \to j}))$$

Our Contributions: A Family of Fast Algorithms for Random Walk based Graph Kernel.

Input Graphs	Time Complexity (Our methods)	Time Complexity (Existing methods)
Normalized, unlabelled	$O(n^2r^4+r^6+mr)$	$O(n^3)$
Unnormalized, unlabelled	$O(nr+r^2+mr)$	$O(n^3)$
Normalized, labelled	$O(d_n n^2 r^4+r^6+mr)$	$O(m^2 i_F)$
Unnormalized, labelled	$O(d_n n^2 r^4+r^6+mr)$	$O(m^2 i_F)$

Complexity Comparison



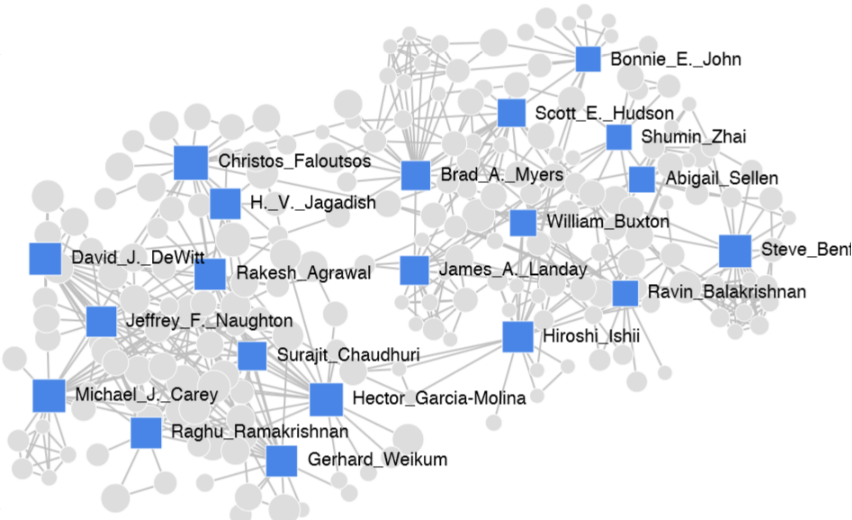
efficiency

HEP-TH			Oregon		
Ark-U	Ark-U+	Ark-L	Ark-U	Ark-U+	Ark-L
0.999	0.999	0.999	0.998	0.999	0.999
0.977	0.999	0.995	0.959	0.999	0.980
0.962	0.999	*	0.939	0.999	*
0.952	0.999	*	0.934	0.999	*
0.946	0.998	*	0.928	0.999	*

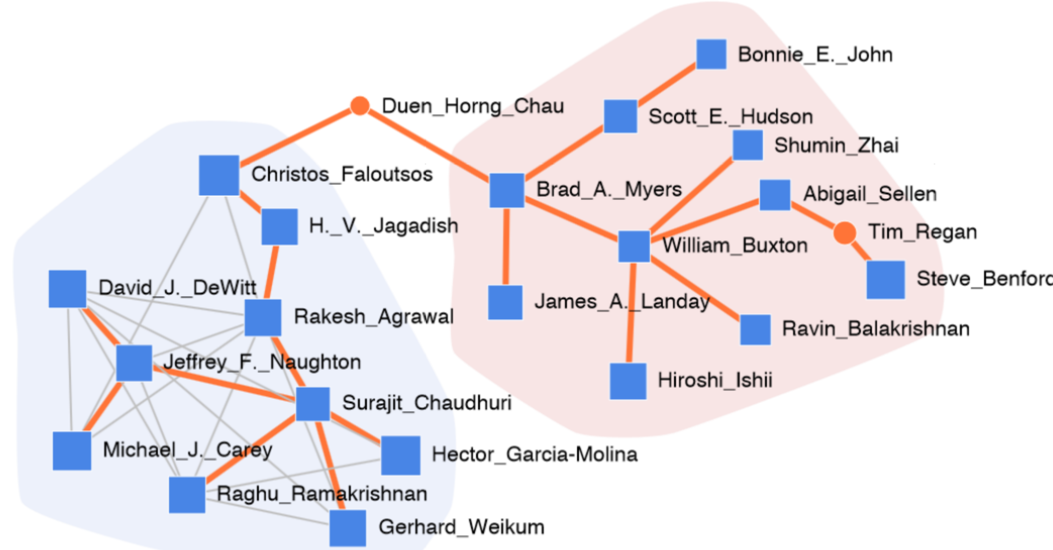
accuracy

Empirical Evaluations

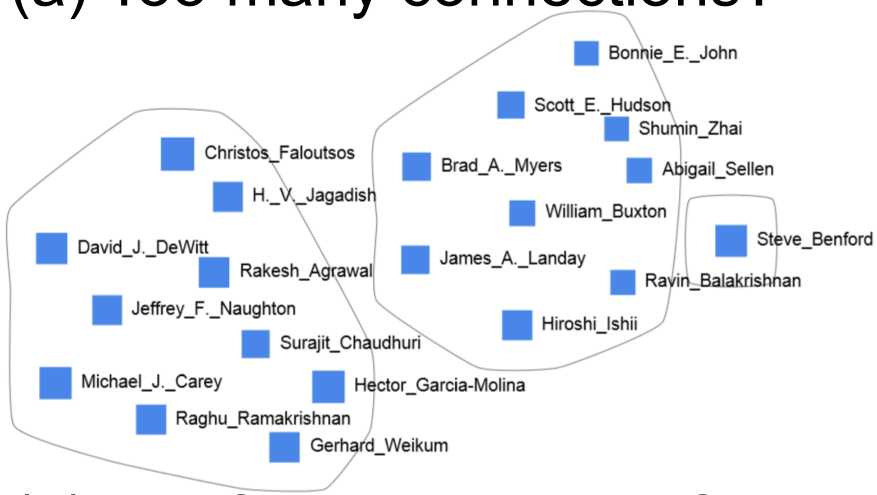
Sense-Making of Marked Nodes [Akoglu+ SDM 2013]



(a) Too many connections?



(c) Our sol.: 'right' connections → better sense-making



(b) Too few connections?

+ 'right' connections = most succinct way to describe marked nodes

+ MDL-based formulation, NP-Hard

+ Effective Approximate Algorithms

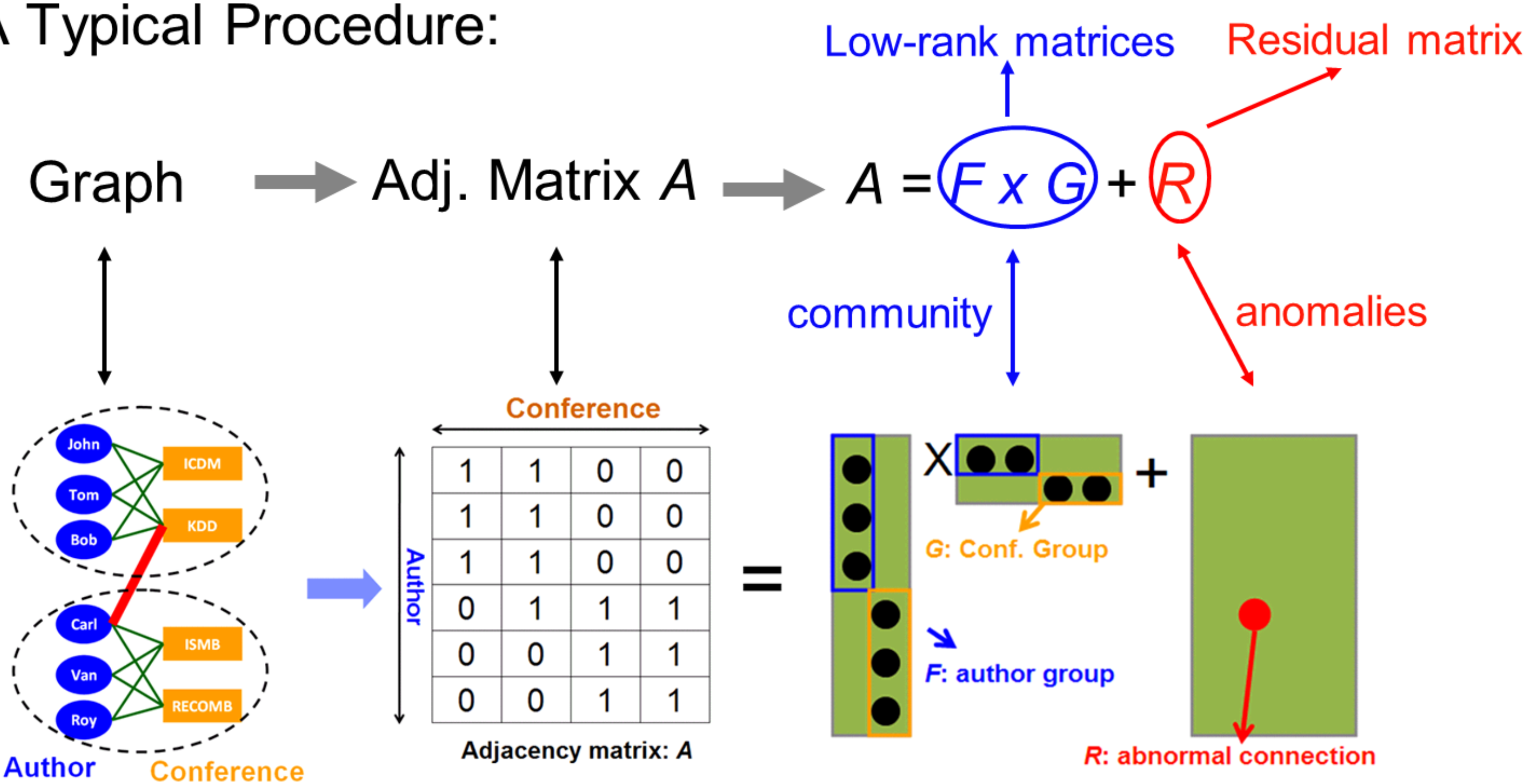
L. Akoglu, J. Vreeken, H. Tong, D. Chau, N. Tatti, and C. Faloutsos: Mining Connection Pathways for Marked Nodes in Large Graphs. SDM 2013

Applications in Social Informatics

- Finding Complex User Patterns
- (Matrix-based) Anomaly Detection
- Influence and Virus Propagation

Graph Anomalies by Low-Rank Approximation

- A Typical Procedure:

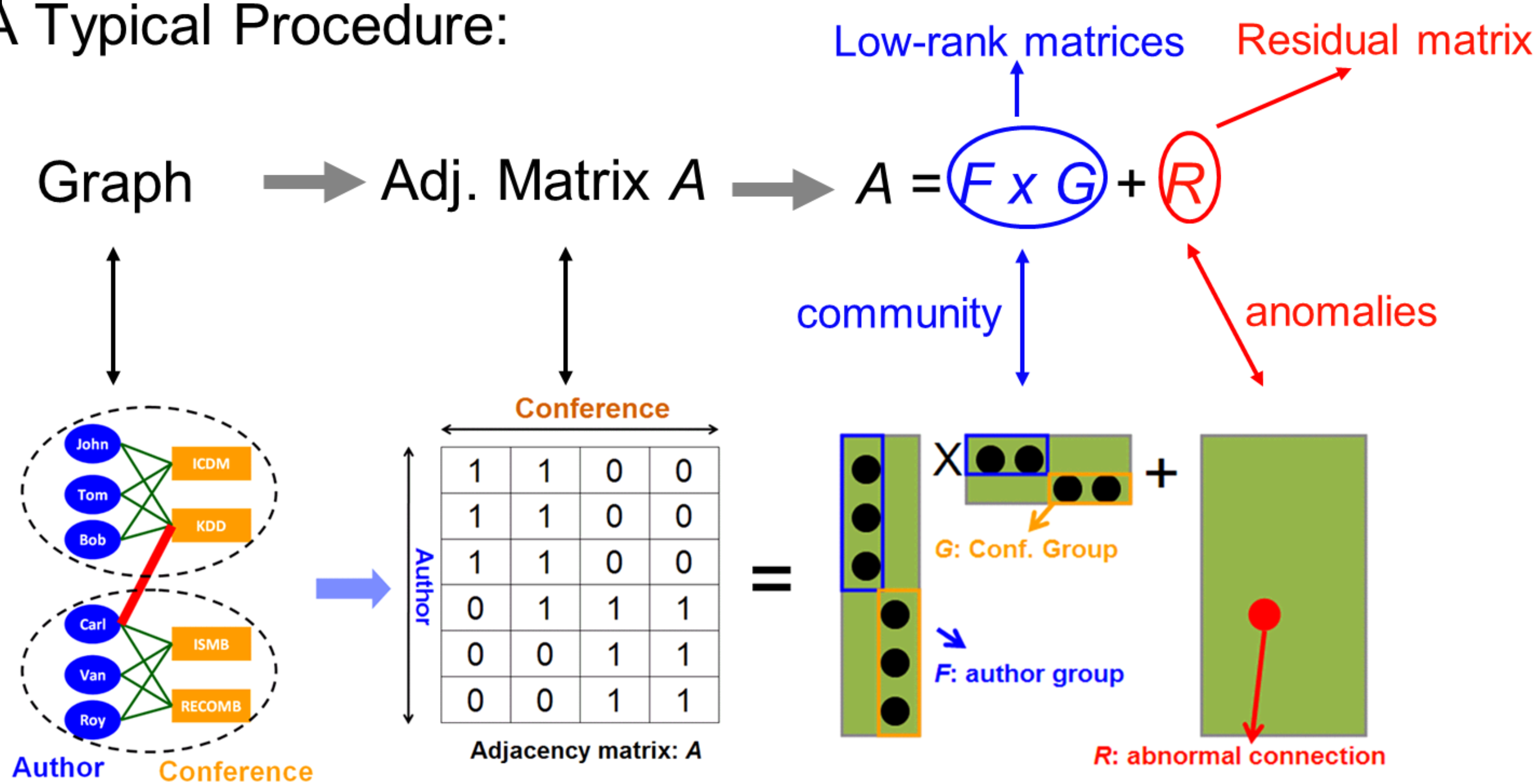


An Illustrative Example

Q: How to get the low-rank matrix approximations?

Graph Anomalies by Low-Rank Approximation

- A Typical Procedure:



An Illustrative Example

Q: How to get the low-rank matrix approximations?

A1: Example-based LRA

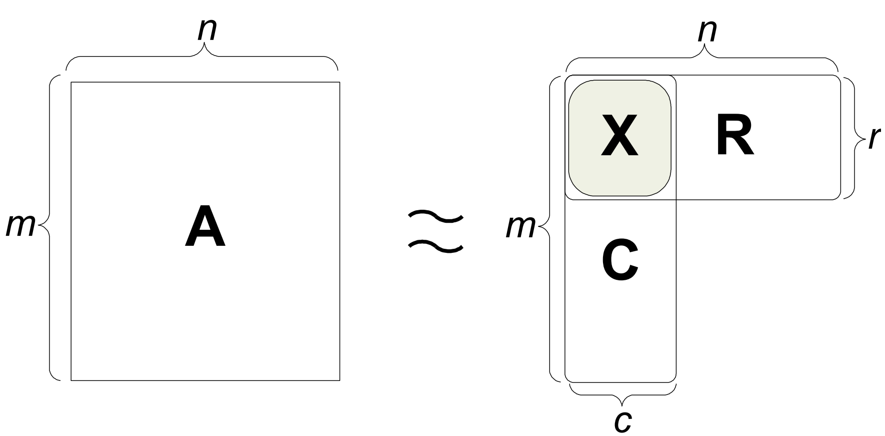
A2: Non-negative Residual Matrix Factorization

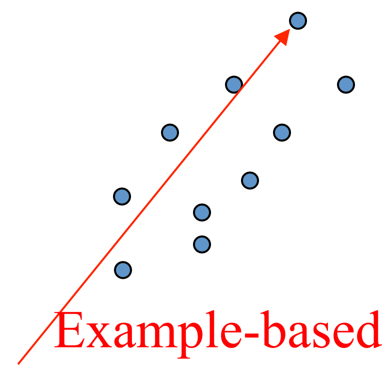
A1: Example-Based LRA

- Why Not SVD, PCA? both transform data into some abstract space (specified by a set basis)
  – Interpretability problem
  – Loss of sparsity (space cost)
  – Efficiency (time cost)

A1: Example-Based LRA -- CUR/CX

- Example-based projection: use actual rows and columns to specify the subspace
- Given a matrix $A \in R^{m \times n}$, find three matrices $C \in R^{m \times c}$, $U \in R^{c \times r}$, $R \in R^{r \times n}$, such that $||A-CUR||$ is small





- Two recent variants:
  - CMD: removing duplicates
  - Colibri: removing linear correlations (and tracking)

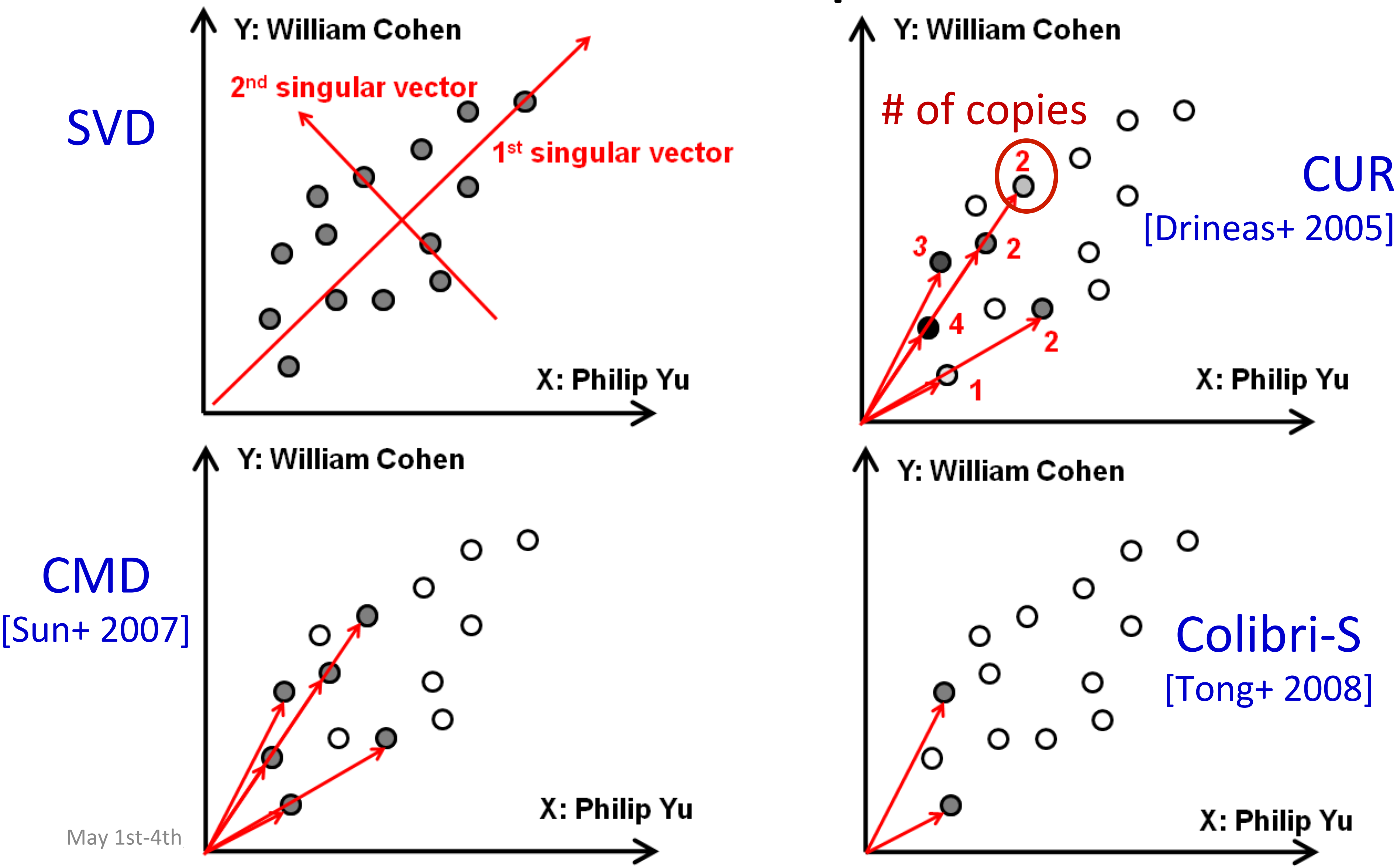
U is the pseudo-inverse of X: $U = X^{\dagger} = (U^T U)^{-1} U^T$

H. Tong, S. Papadimitriou, J. Sun, P.S. Yu, C. Faloutsos: Colibri: fast mining of large static and dynamic graphs. KDD 2008

J. Sun, Y. Xie, H. Zhang, C. Faloutsos: Less is More: Compact Matrix Decomposition for Large Sparse Graphs. SDM 2007

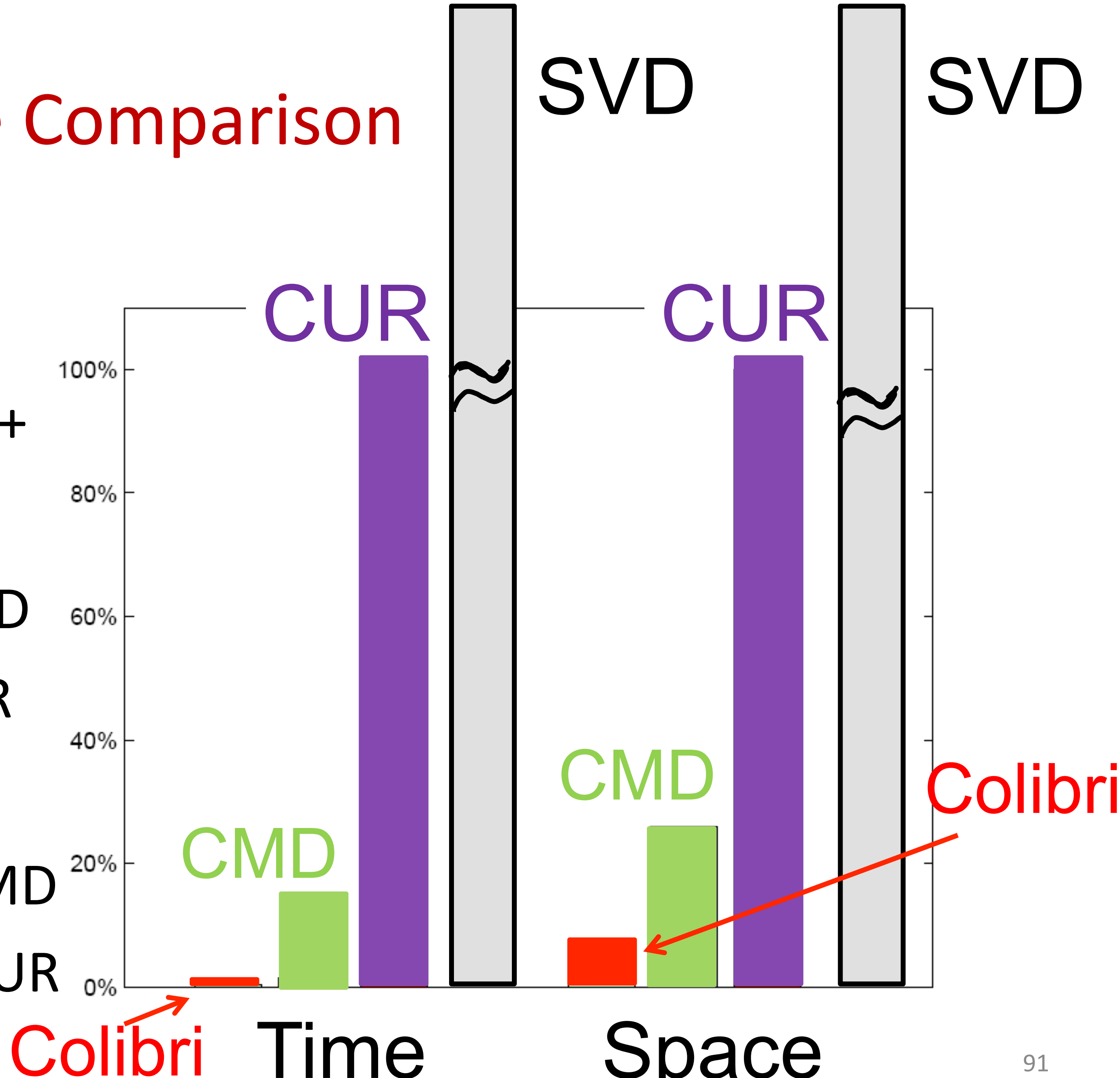
P. Drineas, R. Kannan, M.W. Mahoney: Fast Monte Carlo Algorithms for Matrices II: Computing a Low-Rank Approximation to a Matrix. SIAM J. Comput. (SIAMCOMP) 2006

A Pictorial Comparison



Performance Comparison

- Accuracy
  - <u>Same</u> 91%+
- Time
  - 12x of CMD
  - 28x of CUR
- Space
  - ~1/3 of CMD
  - ~10% of CUR



A2: Non-negative Residual MF

- Observations: anomalies ←→ actual activities
- Examples: popularity contest, port scaner, etc
- NrMF formulation

$$\begin{aligned} \text{argmin}_{\mathbf{F},\mathbf{G}} &= \|\mathbf{R}_{n\times l} \otimes \mathbf{W}_{n\times l}\|_F^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{l}(\mathbf{A}(i,j) - \mathbf{F}(i,:)\mathbf{G}(:,j))^2\mathbf{W}(i,j)^2 \\ \text{s.t.} \quad & \text{for all } \mathbf{A}(i,j) > 0 : \\ & \mathbf{F}(i,:)\mathbf{G}(:,j) \leq \mathbf{A}(i,j) \end{aligned}$$

Weighted Frobenius Form

Common in Any MF

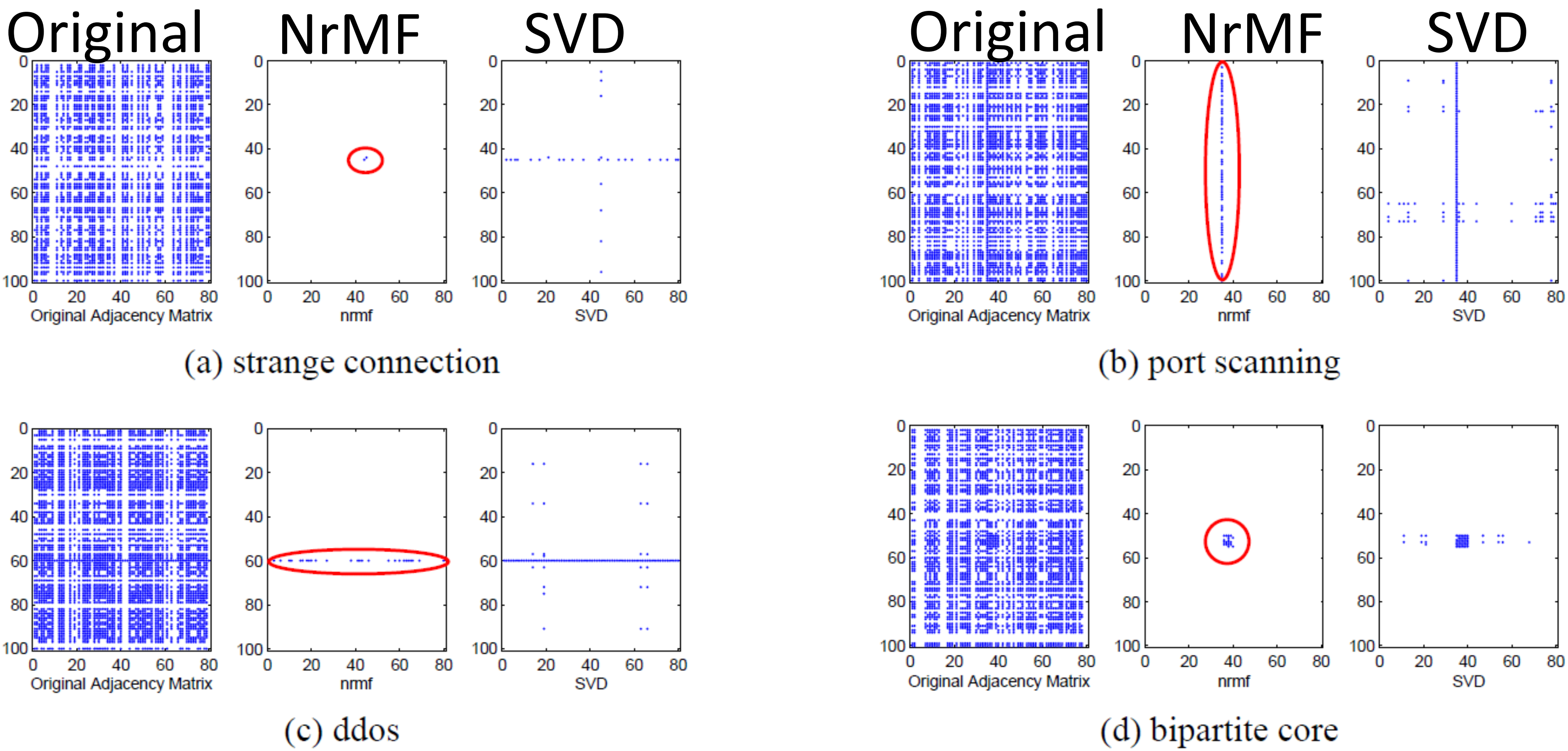
Weight

Unique in NrMF

Non-negative residual

H. Tong, C. Lin: Non-Negative Residual Matrix Factorization with Application to Graph Anomaly Detection. SDM 2011

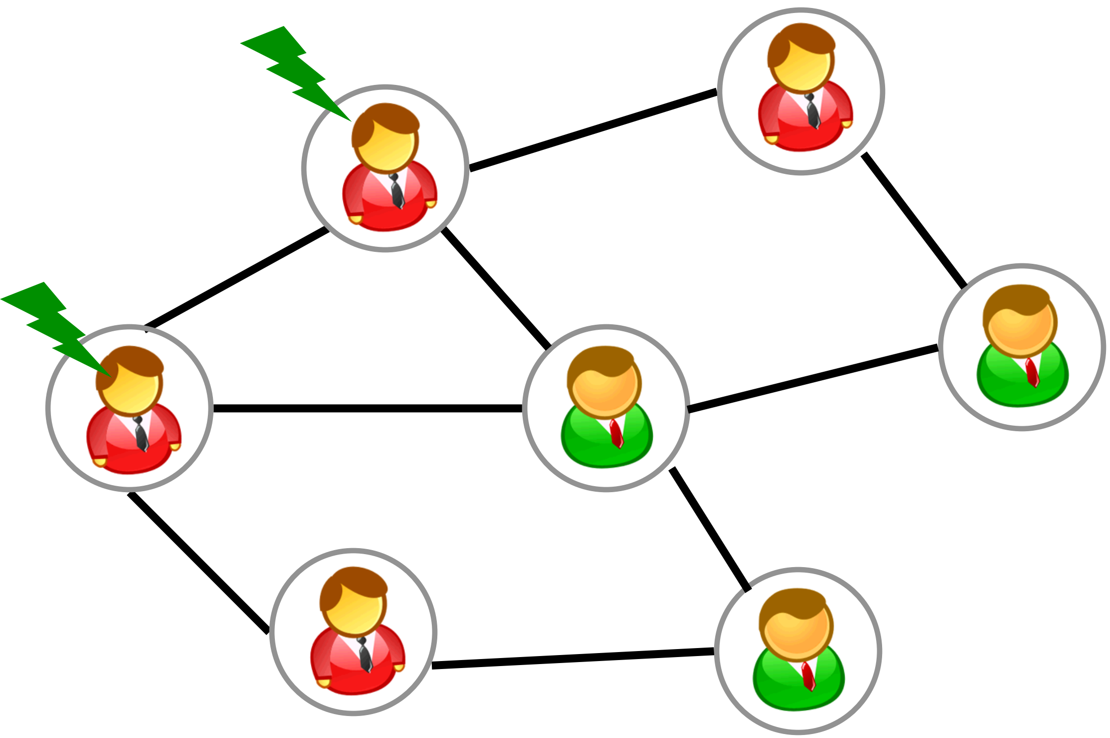
Visual Comparisons



Applications in Social Informatics

- Finding Complex User Patterns
- (Matrix-based) Anomaly Detection
- Influence and Virus Propagation

An Example: Flu/Virus Propagation



1: Sneeze to neighbors

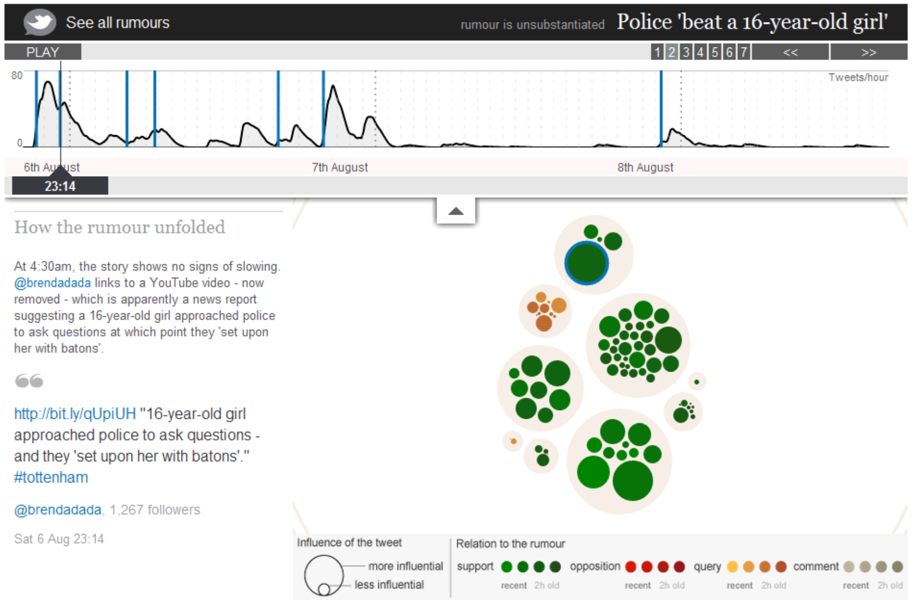
2: Some neighbors → Sick

3: Try to recover

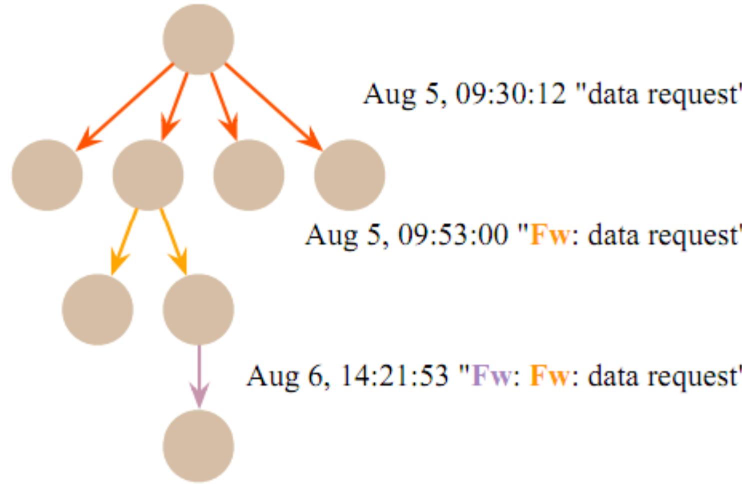
Q: How to minimize infected population?

- Q1: Understand tipping point
- Q2: Affecting algorithms

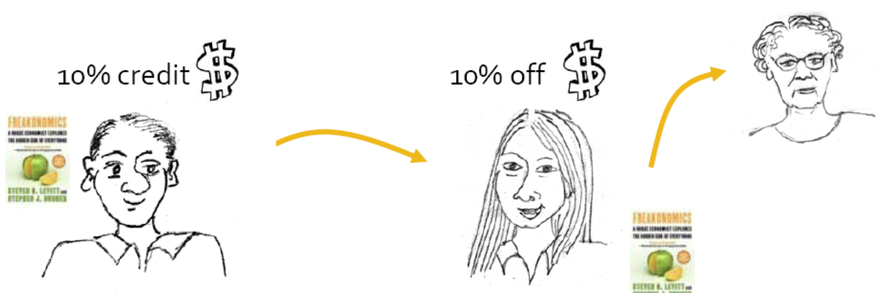
Why Do We Care?



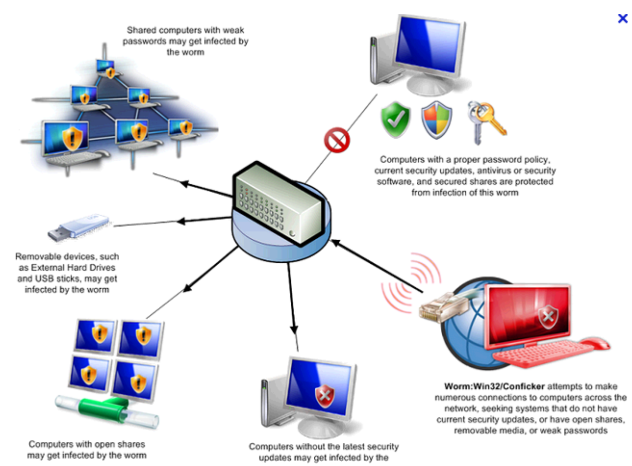
Rumor Prop on Twitter in UK riots



Email Fwd in Organization



Viral Marketing

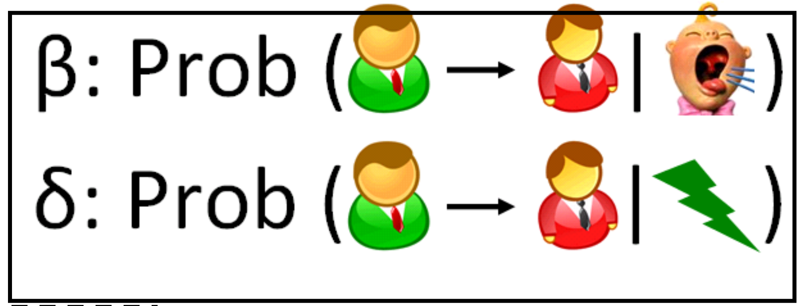


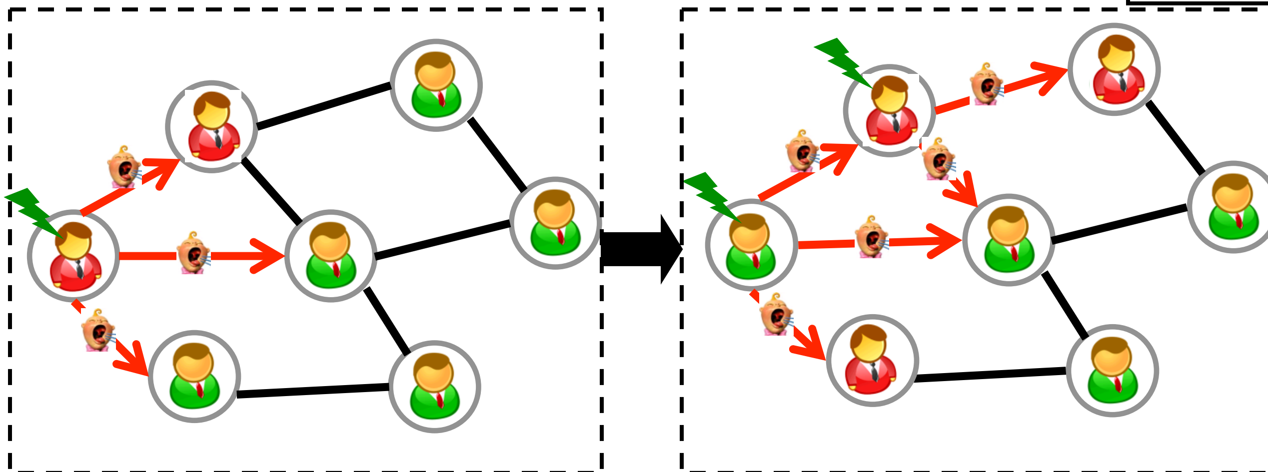
Malware Infection

- Q1: Understand tipping point

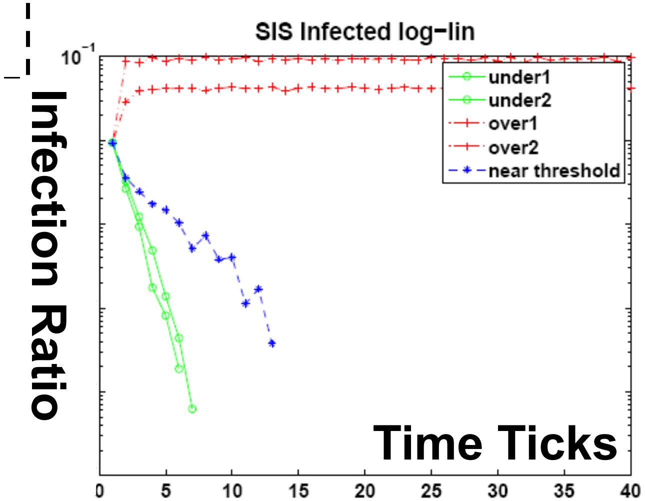
- Q2: Affecting algorithms

SIS Model (e.g., Flu)





$$p_{t+1} = H(p_t)$$



Theorem [Chakrabarti+ 2003, 2007]:

If $\lambda \times (\beta/\delta) \leq 1$; no epidemic for any initial conditions

λ: largest eigenvalue of the graph (~ connectivity of the graph)

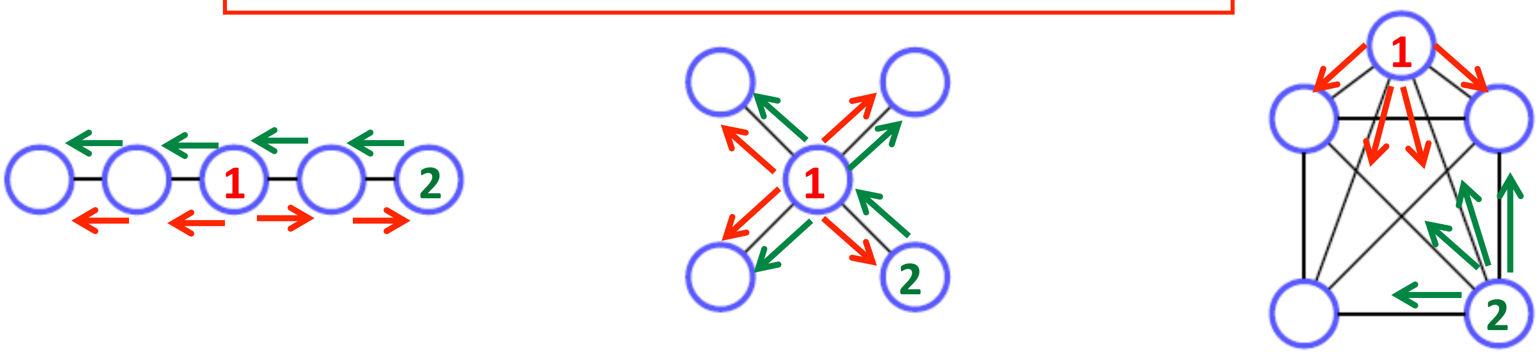
β, δ : virus parameters (~strength of the virus)

Generalize to ~ 25 other models; to partial immunity; to dynamic networks

Why is λ So Important?

- λ → Capacity of a Graph:

$$\boxed{\left(\vec{1}^{*} A^{k} \vec{1}\right)^{1/k} \xrightarrow[k \to \infty]{} \lambda}$$



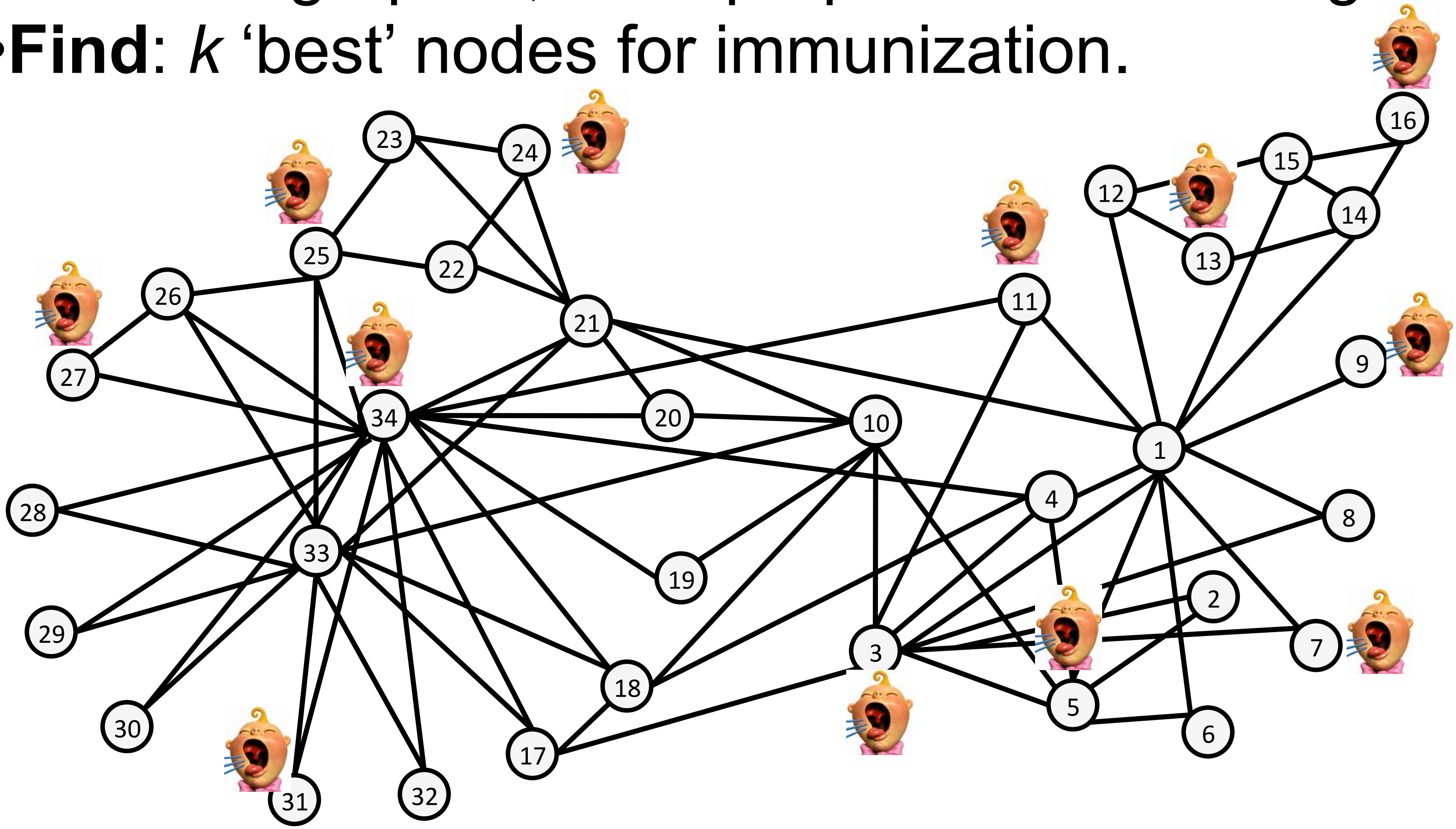
(a)Chain($\lambda_1 = 1.73$) **(b)Star**($\lambda_1 = 2$) **(c)Clique**($\lambda_1 = 4$)

Larger λ → better connected

- Q1: Understand tipping point
- Q2: Affecting algorithms

Minimizing Propagation: Immunization

- **Given**: a graph *A*, virus prop model and budget *k*;
- **Find**: *k* 'best' nodes for immunization.

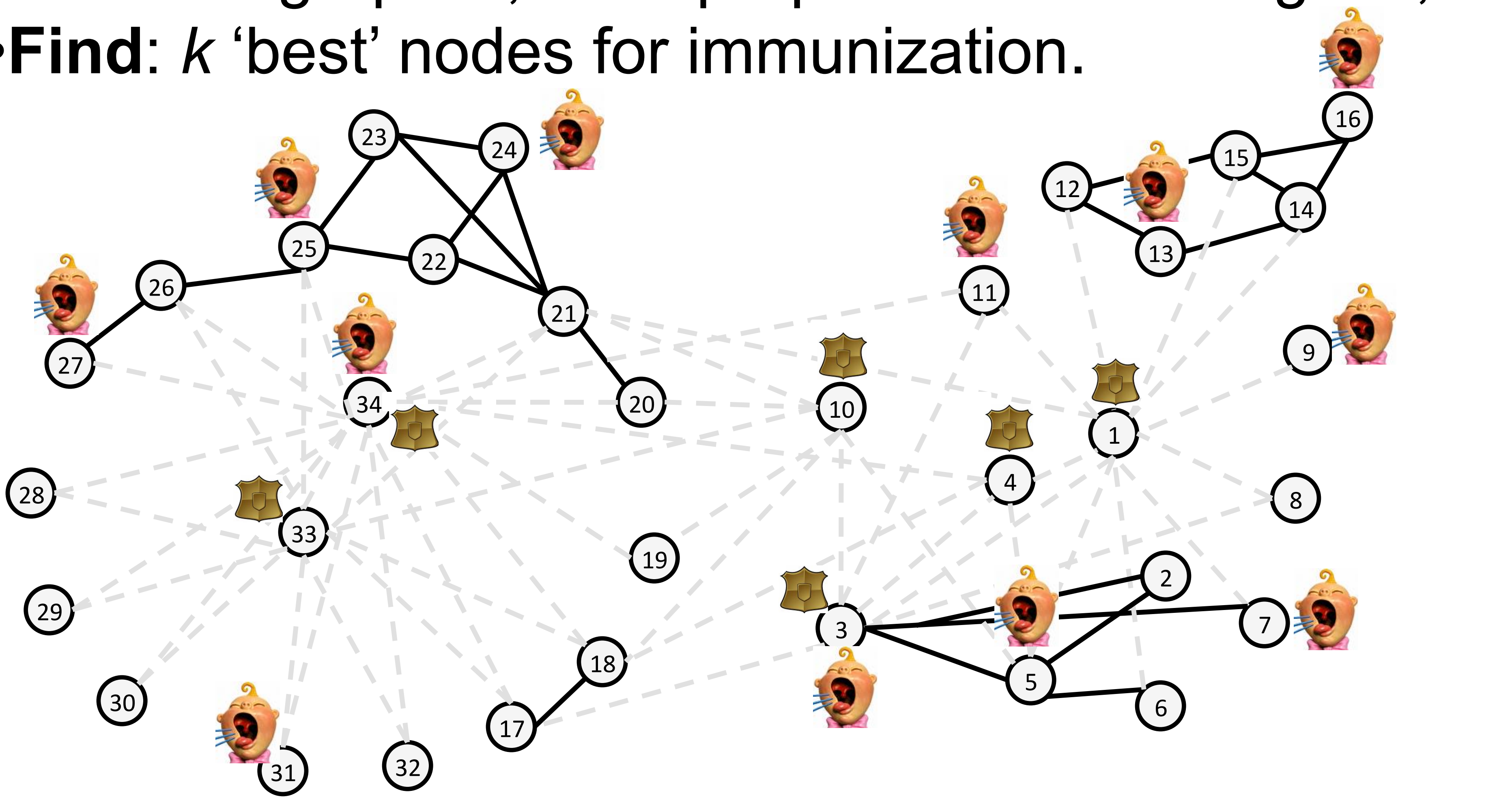


SARS costs 700+ lives; $40+ Bn; H1N1 costs Mexico $2.3bn

- Q1: Understand tipping point
- Q2: Affecting algorithms

Minimizing Propagation: Immunization

- **Given**: a graph *A*, virus prop model and budget *k*;
- **Find**: *k* 'best' nodes for immunization.

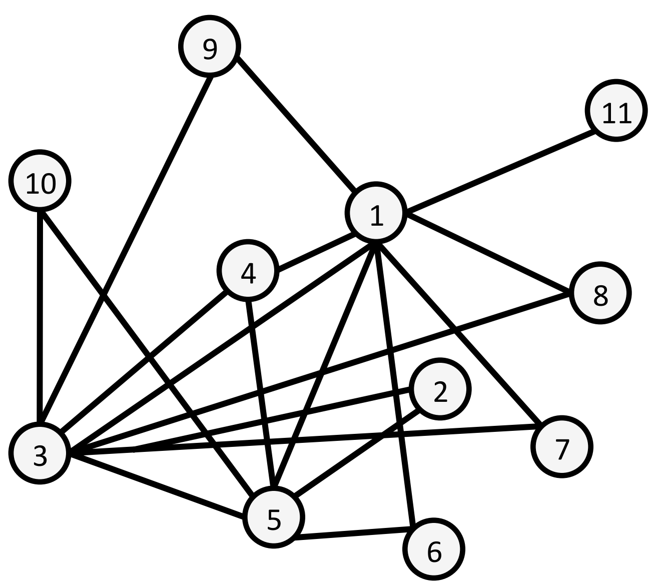


SARS costs 700+ lives; $40+ Bn; H1N1 costs Mexico $2.3bn

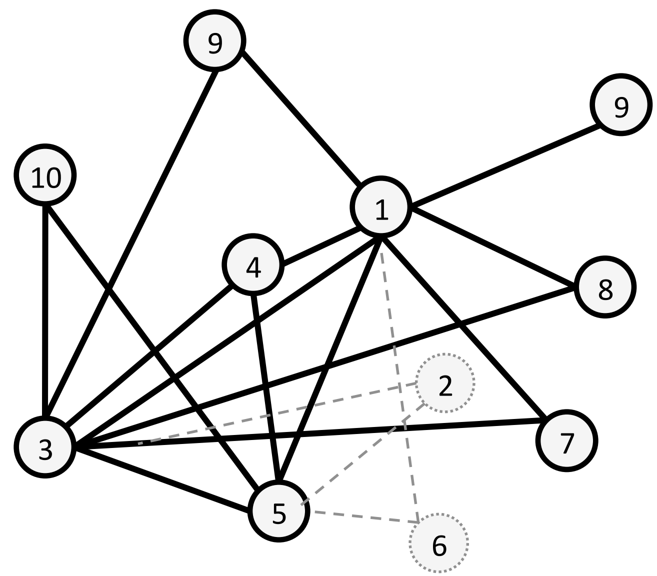
Optimal Method

- Select *k* nodes, whose absence creates the largest drop in λ

$$S = \arg\max_{|S|=k} \lambda\text{-}\lambda_S$$



Original Graph: λ



Without {2, 6}: λ_s

Optimal Method

- Select *k* nodes, whose absence creates the largest drop in λ

$$S = \arg\max_{|S|=k} \lambda - \lambda_S$$

Largest eigenvalue w/o subset of nodes *S*

- But, we need $O(\binom{n}{k} \cdot m)$ in time
  - Example: 1,000 nodes, with 10,000 edges
    - It takes 0.01 seconds to compute λ
    - It takes **2,615 years** to find best-*5* nodes !

Theorem: (Tong+ CIKM 2012)
Find Optimal k-node Immunization is NP-Hard

Netshield to the Rescue

Theorem: (Tong+ 2010)

(1) $\lambda - \lambda_s \approx Sv(S) = \sum_{i \in S} 2\lambda u(i)^2 - \sum_{i,j \in S} A(i,j)u(i)u(j)$

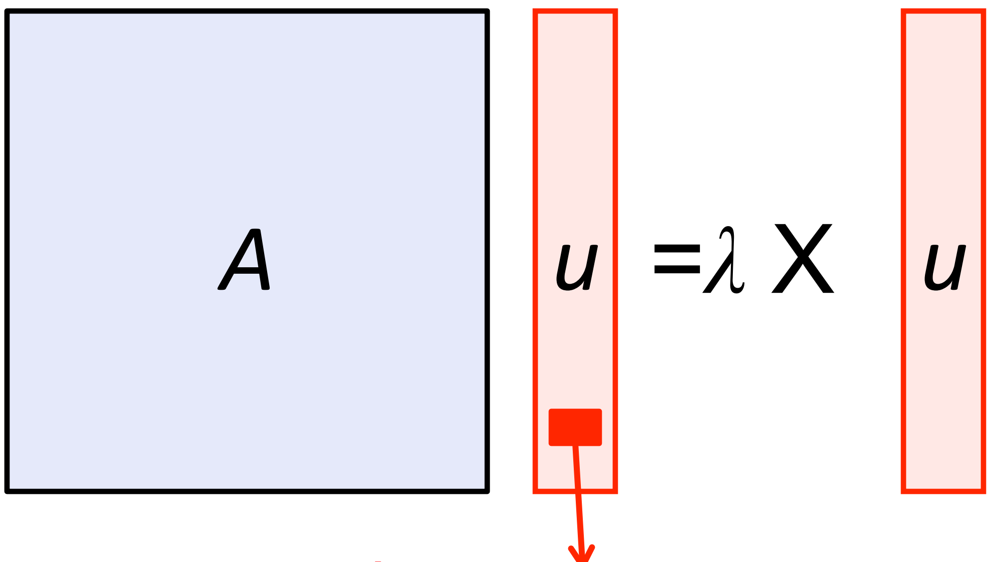
$A_S = A - E$

$= A - (F + F' - G)$

$\lambda_s = \lambda - u' Eu/(u'u) + O(|E|^2)$

$= \lambda - 2u' Fu + 2u' Eu + O(|E|^2)$

$= \lambda - (\sum_{i \in S} 2\lambda u(i)^2 - \sum_{i,j \in S} A(i,j)u(i)u(j)) + O(|E|^2)$



u(i): eigen-score

Footnote: *u(i)* ~ PageRank*(i)* ~ in-degree*(i)*

- Q1: Understand tipping point
- Q2: Affecting algorithms

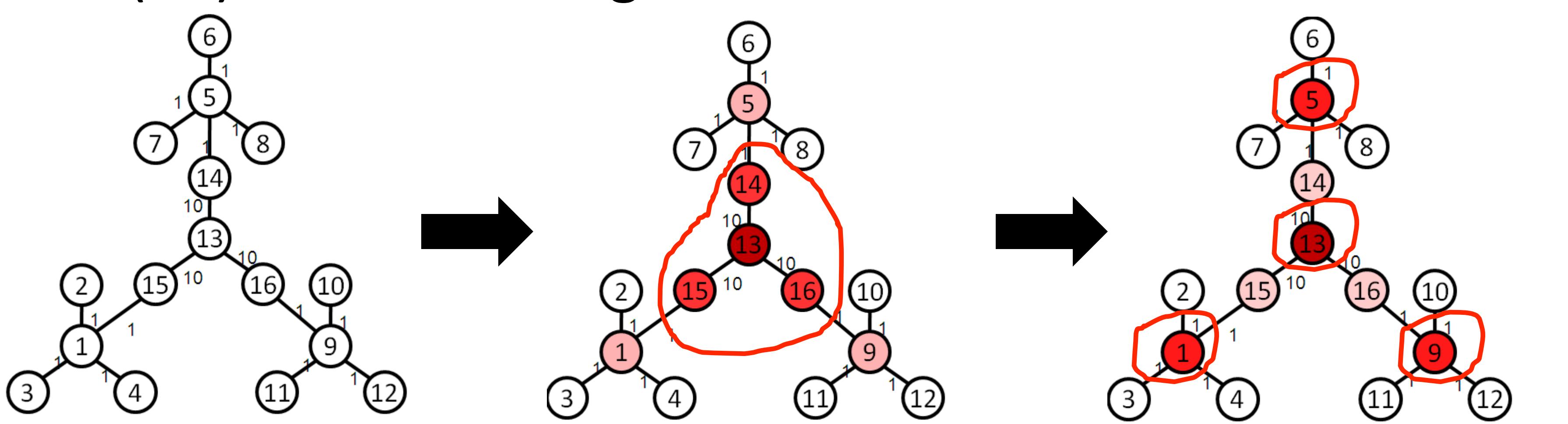
Intuition

Netshield to the Rescue

Theorem: (Tong+ 2010)

(1) $\lambda - \lambda_s \approx Sv(S) = \sum_{i \in S} 2\lambda u(i)^2 - \sum_{i,j \in S} A(i,j)u(i)u(j)$

- find a set of nodes *S (e.g. k=4),* which
  - (C1) each has high eigen-scores
  - (C2) diverse among themselves



Original Graph　　Select by C1　　Select by C1+C2

Netshield to the Rescue

Theorem: (Tong+ ICDM 2010)

(1) $\lambda - \lambda_s \approx Sv(S) = \sum_{i \in S} 2\lambda u(i)^2 - \sum_{i,j \in S} A(i,j) u(i) u(j)$

(2) Sv(S) is sub-modular (+monotonically non-decreasing)

Corollary: (Tong+ ICDM 2010)

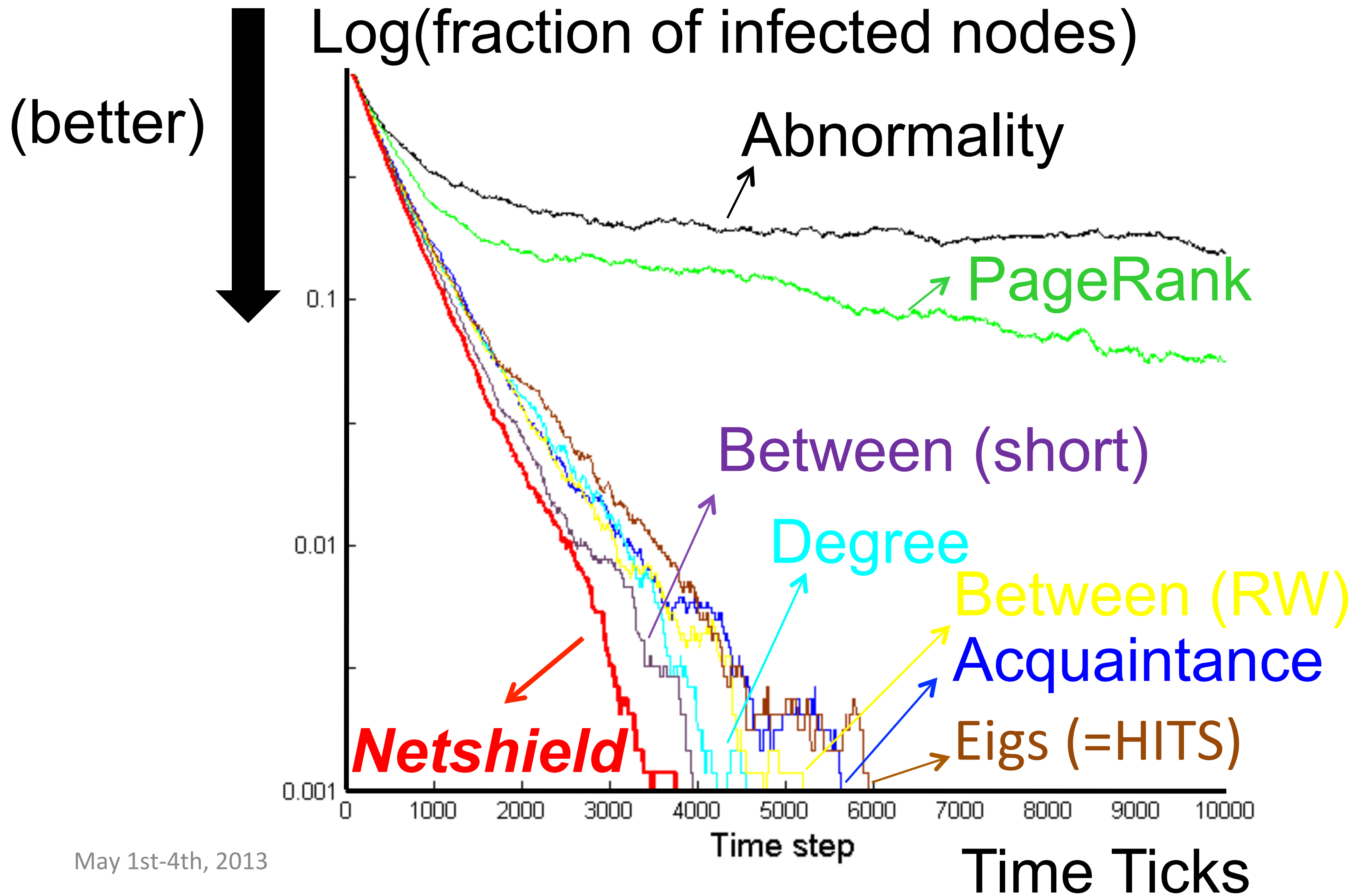
(3) *Netshield* is near-optimal (wrt max Sv(*S*))

(4) *Netshield* is $O(nk^2+m)$

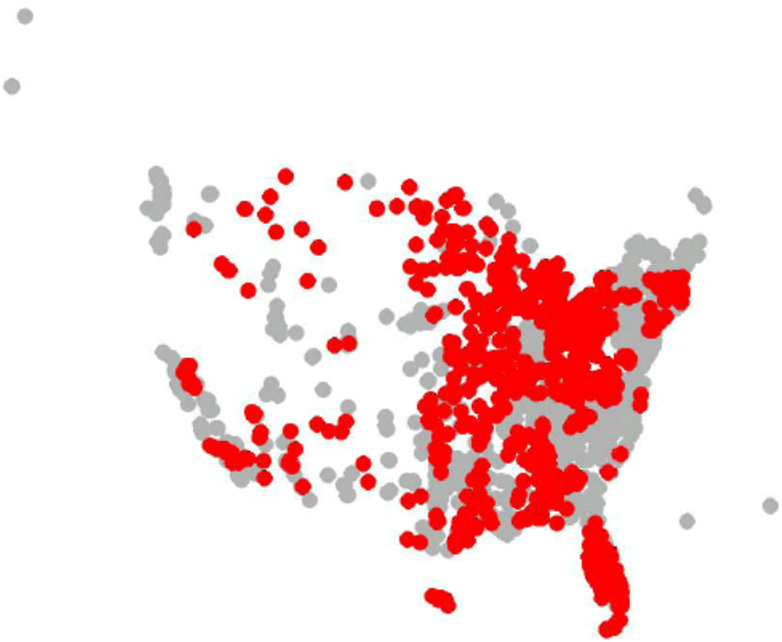
- Example: 1,000 nodes, with 10,000 edges
  - *Netshield* takes **< 0.1 seconds** to find best-*5* nodes !
  - … as opposed to **2,615 years**

Footnote: near-optimal means $Sv(S^{Netshield}) >= (1-1/e)\, Sv(S^{Opt})$

Comparison of Immunization



Hospital Infection Controlling (US-Medicare Network)



Current Method

Out Method

Red: Infected Hospitals after 365 days

Using a variant of Netshield (adapting to partial immunity)
B.A. Prakash, L. Adamic, T. Iwashynaz, H. Tong and C. Faloutsos: Fractional Immunization in Networks. SDM 2013.

Maximizing Propagation: Edge Addition

[Tong+ CIKM 2012]

•**Given**: a graph *A*, virus prop model and budget *k*;

•**Find**: add *k* 'best' new edges into *A*.

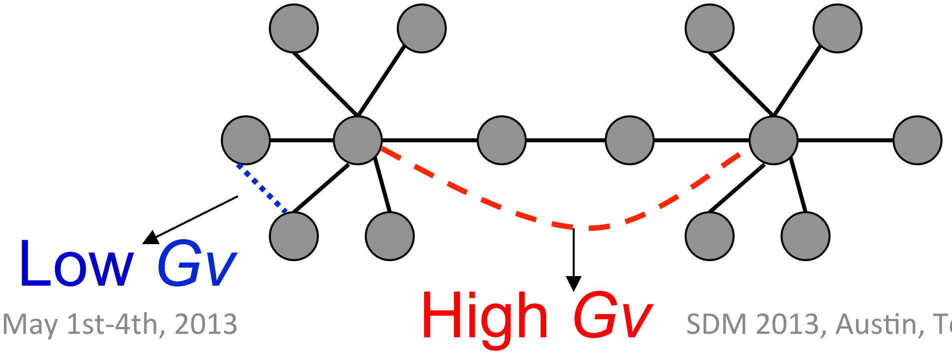
- By 1st order perturbation, we have

$$\lambda_s - \lambda \approx G\mathrm{v}(S) = c \sum_{e \in S} u(i_e) v(j_e)$$

Left eigen-score of source

Right eigen-score of target

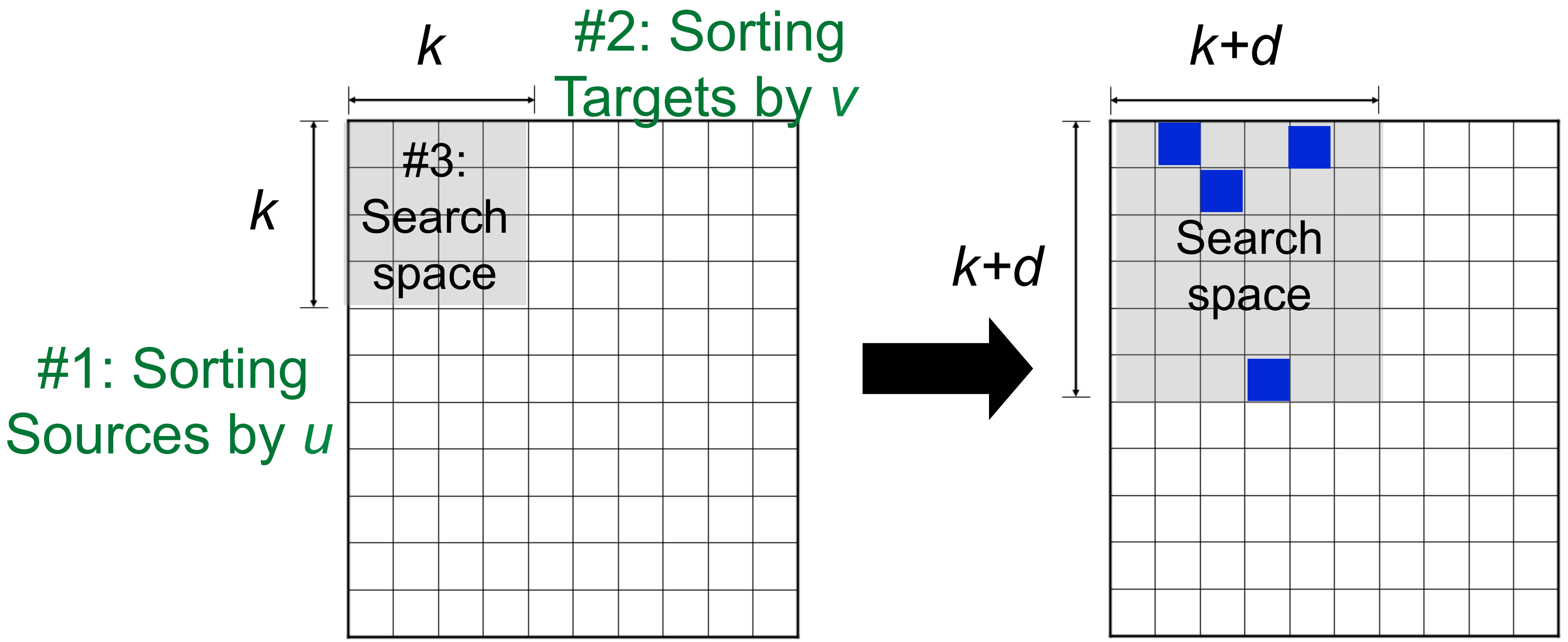
- So, we are done → need O(n^2-m) complexity



Maximizing Propagation: Edge Addition

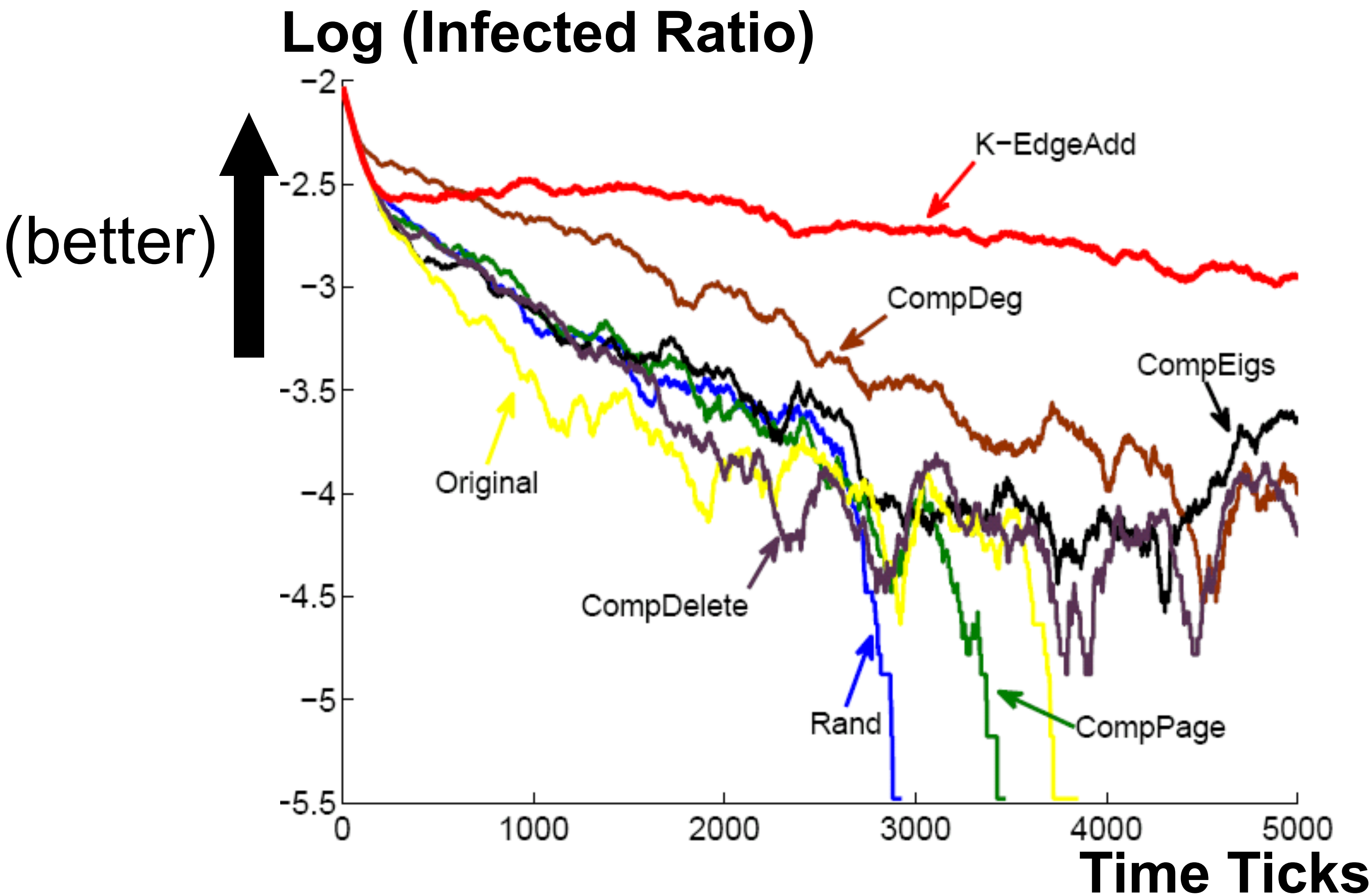
$$\lambda_s - \lambda \approx Gv(S) = c \sum_{e \epsilon S} u(i_e) v(j_e)$$

- Q: How to Find k new edges w/ highest $Gv(S)$?
- A: Modified Fagin's algorithm



Time Complexity: $O(m+nt+kt^2)$, $t = \max(k,d)$:existing edge

Maximizing Propagation: Evaluation



Influence & Virus Propagation: Summary

- **Goal**: Guild Dissemination by Opt. Link Structure
- **Theory**: Opt. Dissemination = Opt. λ
- **Algorithms**:
  - Netshield to Minimize Dissemination
  - NetGel to Maximize Dissemination
- **More on This Topic**
  - Beyond Link Structure (content, attribute) [WWW11]
  - Beyond Full Immunity [SDM13b]
  - Higher Order Variants [CIKM12a]
  - Equivalence (node deletion vs. edge deletion) [CIKM12a]
  - Immunization on Dynamic Graphs [PKDD10]

Conclusion & Remarks

Tools	C' Patterns	Anomalies	Influ' Prop	Immunization	Symptom Exp'	Similar Patient	Clinical Patterns	Risk Factor
Prox.	✓				✓			
LRA		✓				✓	✓	
Sparse L'								✓
Large L'						✓		
Eigen. Opt.			✓	✓				

Social Networks

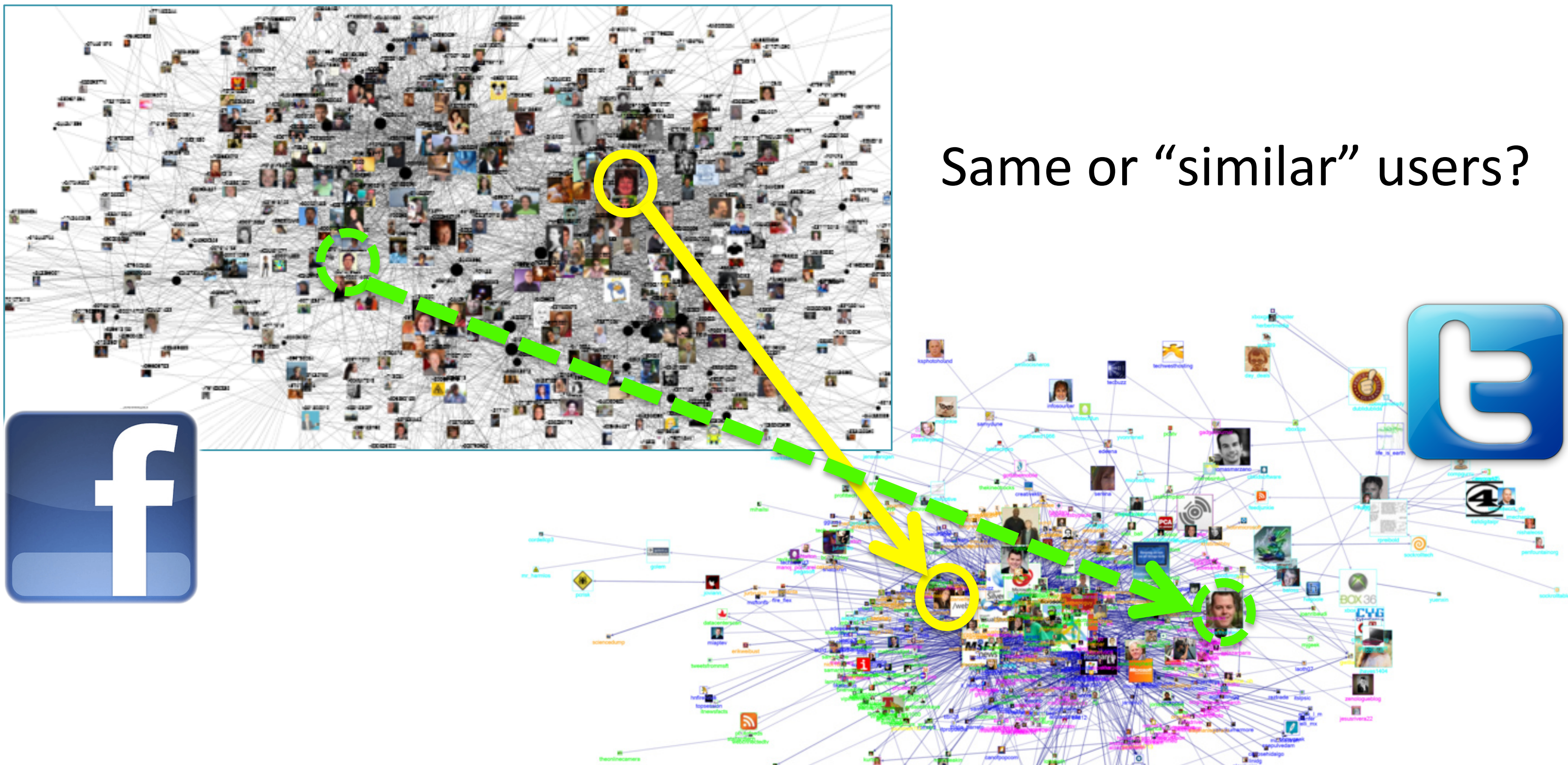
Healthcare

Conclusion

- Recent Advances in Applied Matrix Technologies
  - Low rank approximation
  - Sparse learning
  - Large scale learning
- Applications in healthcare informatics
- Applications in social informatics

backup

Matrices in Social Networks [Koutra+ 2013]

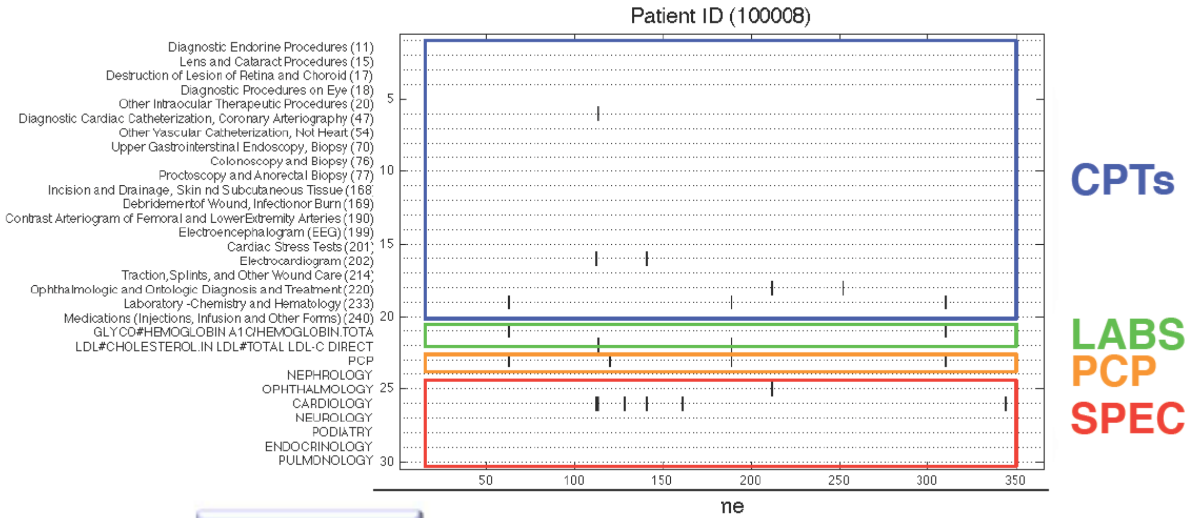


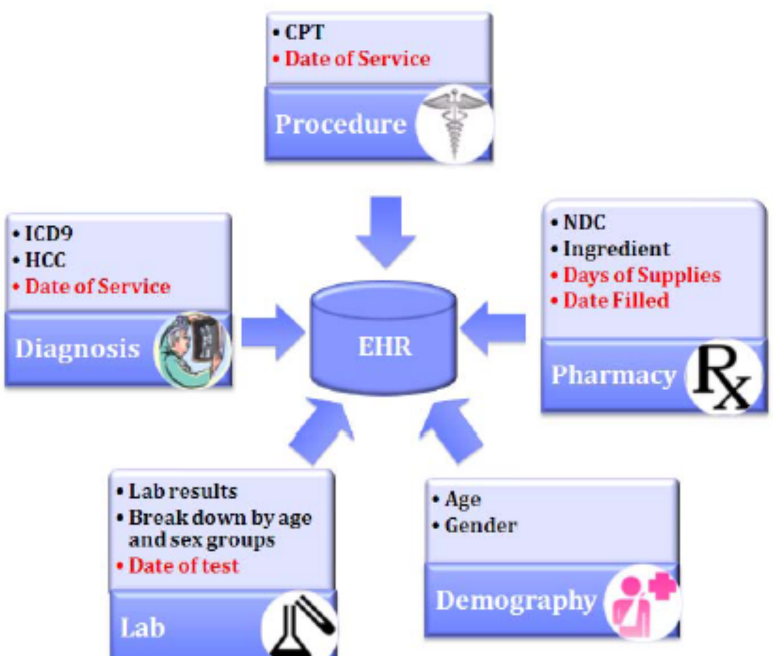
Research Qs: Can we identify users across social networks?

Matrices: rows/columns: people; entries: friendship

Matrix Tools: graph alignment

Matrices in Healthcare



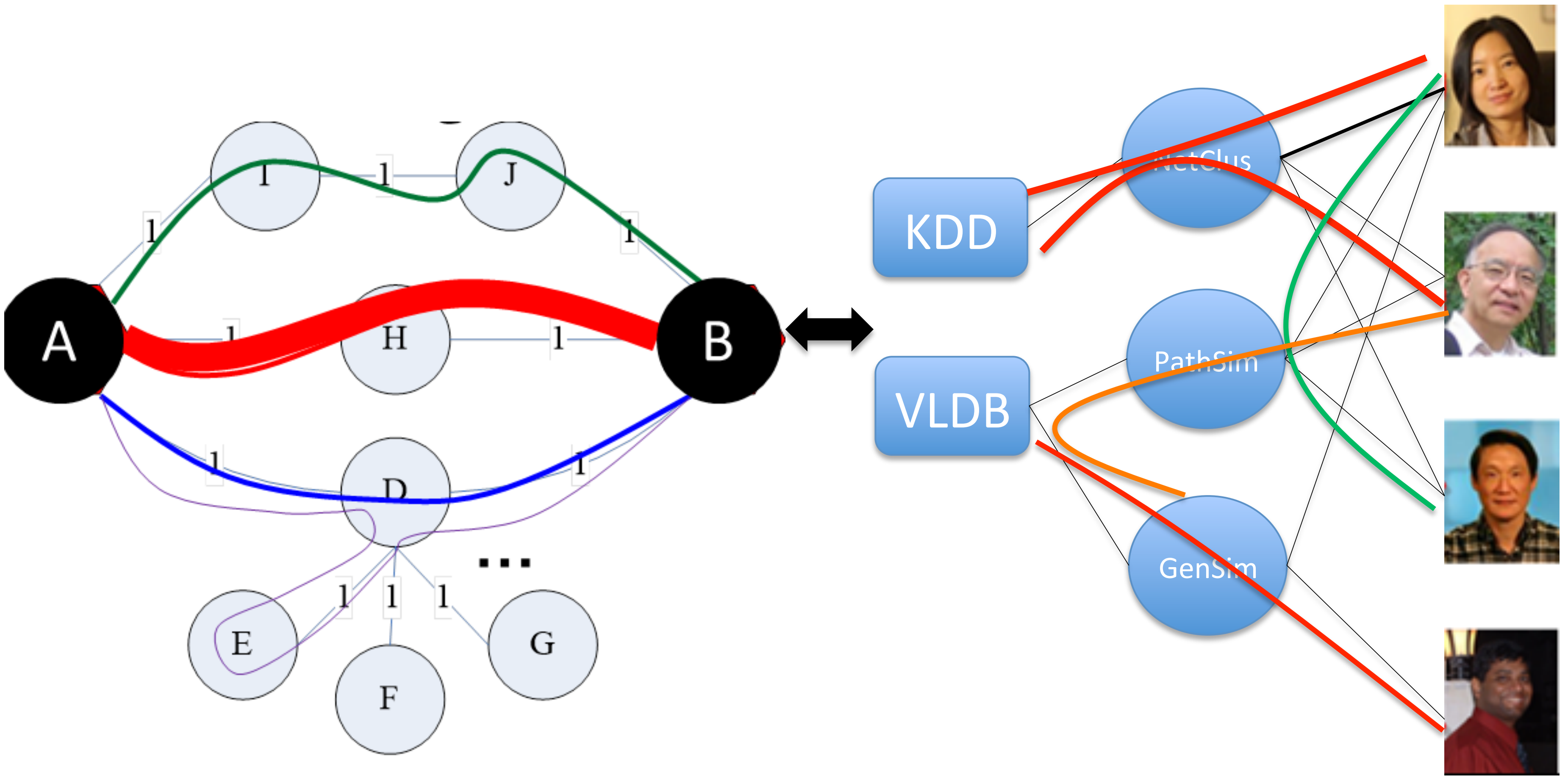


How to identify clinically similar patients?

How to utilize EMR data to perform predictive modeling?

How to characterize the progression course of a specific disease?

Recent Advance #3: Prox. on H. Networks



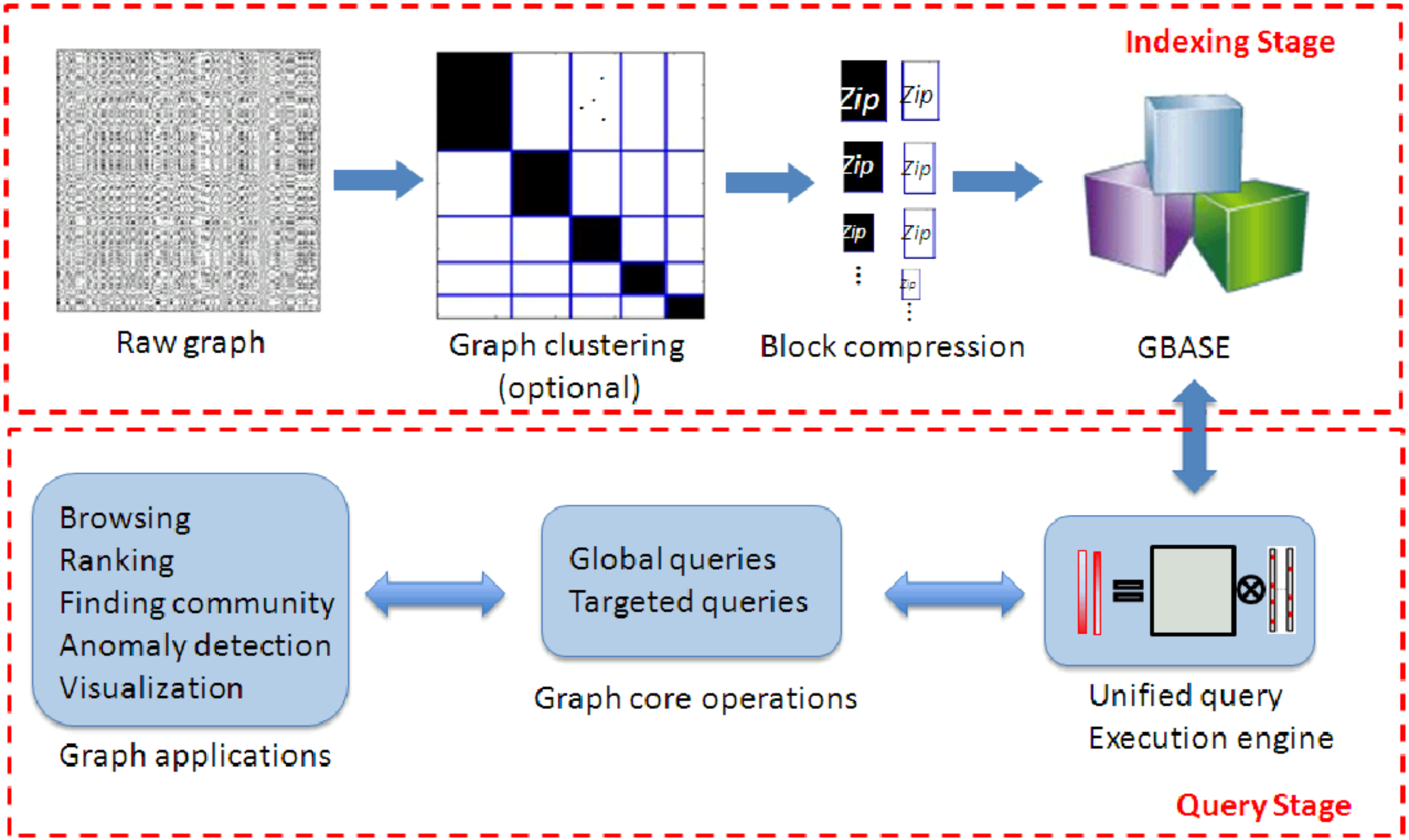
RWR: Path w/ Different Length ➡ Meta Path: Path w/ Different Types

Y. Sun, B. Norick, J. Han, X.Yan, P.S. Yu, and X. Yu: Integrating Meta-Path Selection with User-Guided Object Clustering in Heterogeneous Information Networks.KDD'12

Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu: PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. PVLDB 2011

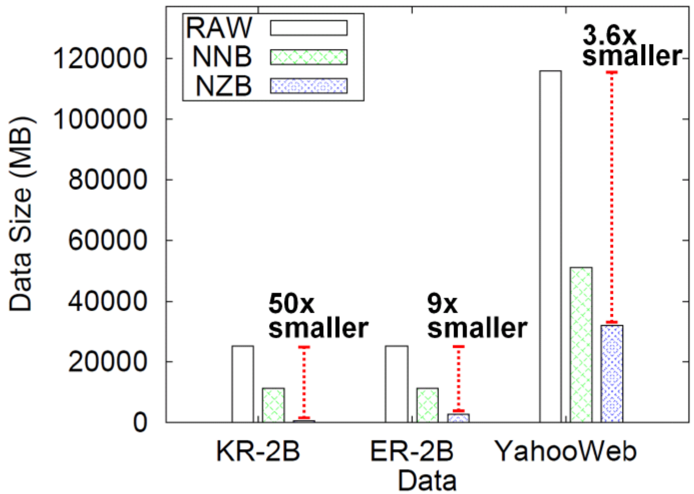
K. Chiang, N. Natarajan, A. Tewari, I. S. Dhillon: Exploiting longer cycles for link prediction in signed networks. CIKM 2011

Recent Advance #4: Scale-Up

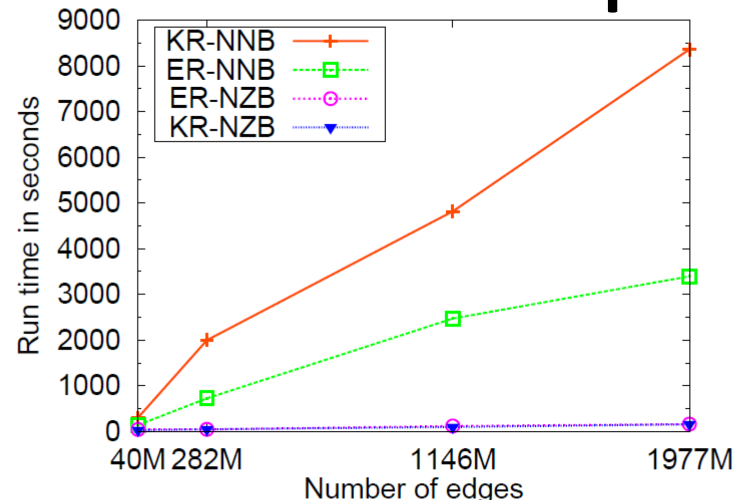


U. Kang, H. Tong, J. Sun, C. Lin, C. Faloutsos: GBASE: a scalable and general graph management system. KDD 2011
U. Kang, C. Tsourakakis, C. Faloutsos: PEGASUS: mining peta-scale graphs. KAIS 2011

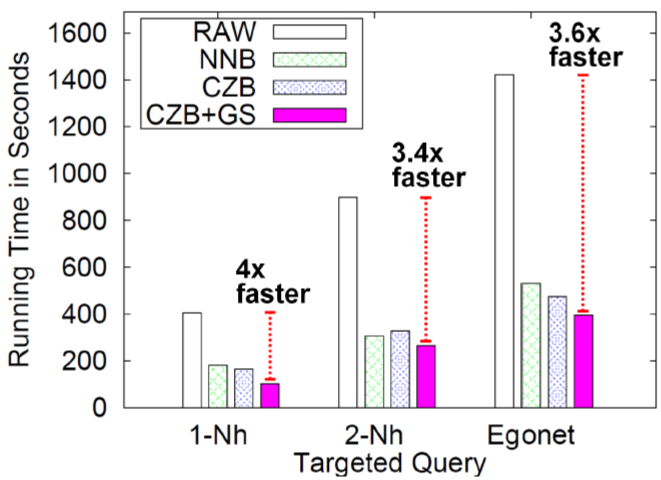
Recent Advance #4: Scale-Up



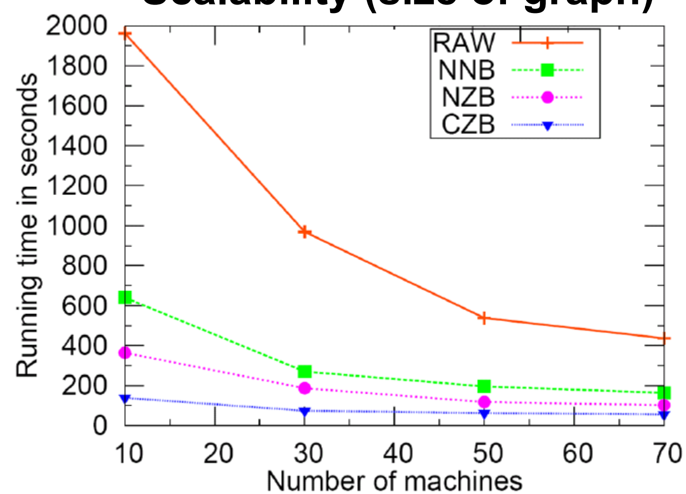
Storage Savings



Scalability (size of graph)



Running Time Savings



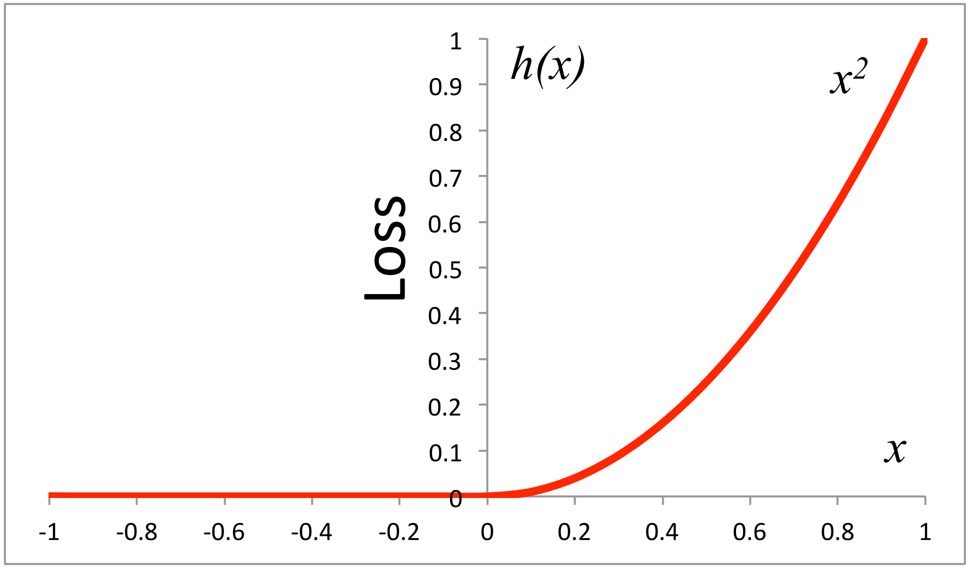
Scalability (# of Machine)

U. Kang, H. Tong, J. Sun, C. Lin, C. Faloutsos: GBASE: a scalable and general graph management system. KDD 2011

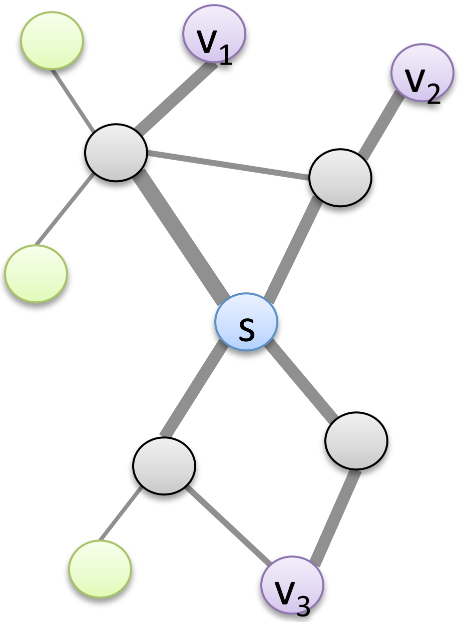
U. Kang, C. Tsourakakis, C. Faloutsos: PEGASUS: mining peta-scale graphs. KAIS 2011

Learning W: $\min_w F(w) = ||w||^2 + \lambda \sum_{ld} h(p_l - p_d)$

- w: the parameter to learn
- *h: loss function to penalize the violation*



$p_l < p_d$ $p_l = p_d$ $p_l > p_d$



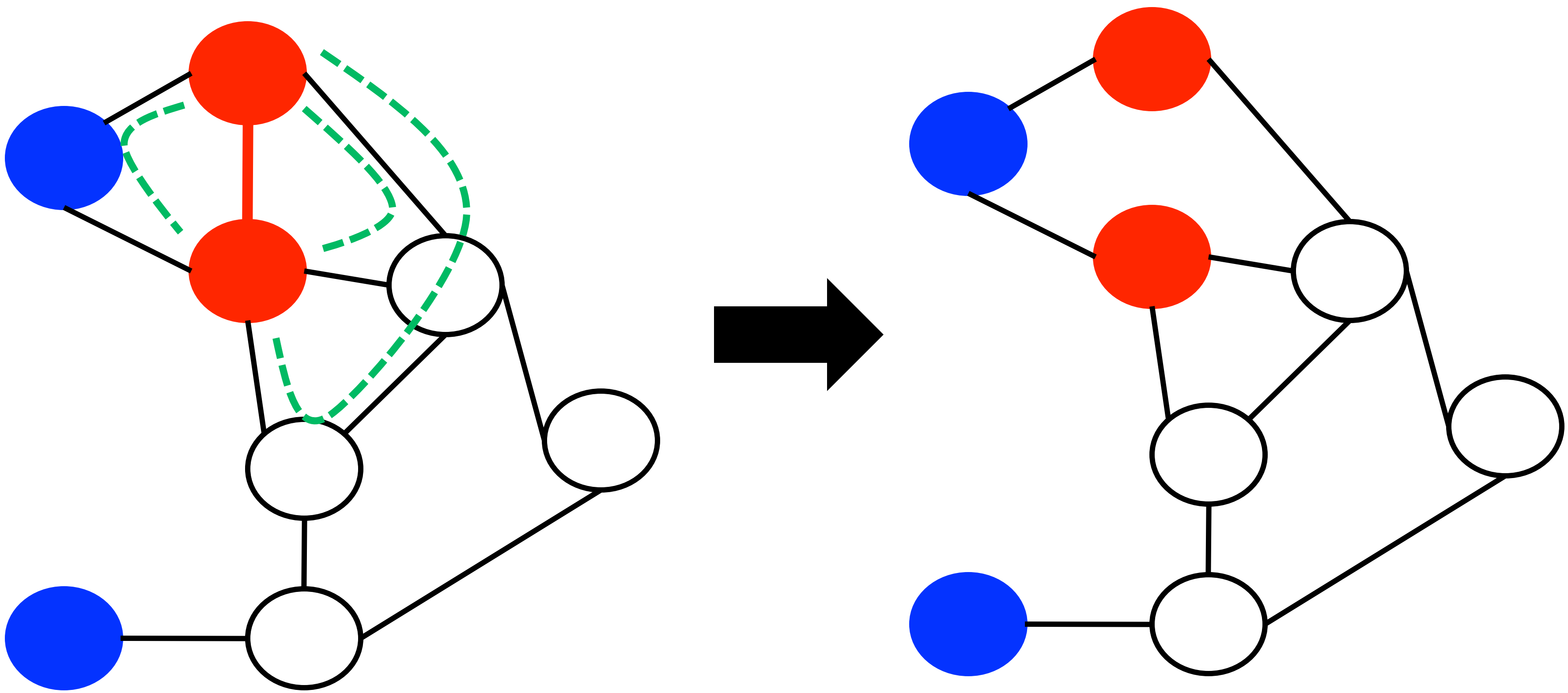
s: be the center node; *l*: destination; *d*: no-link

edge strength $a_{uv} = f_w(u,v) = exp(-w^T \Psi_{uv})$

feature vector Ψ_{uv} (Features of node *u;* Features of node *v;* Features of edge *(u,v)*

Thanks to Jure Leskovec: Social Media Analytics (KDD '11 tutorial)

Link Prediction



Prox (deleted) >> Prox (absent) !

Footnote:

- Red pair: ``deleted”;
- Blue pair: ``absent”

Experimental setting

- Node and Edge features for learning:
  - Node:
    - Age; Gender; Degree
  - Edge:
    - Age of an edge; Communication; Profile visits; Co-tagged photos
- Baselines:
  - Decision trees and logistic regression:
    - Above features + 10 network features (PageRank, common friends)
- Evaluation:
  - AUC and precision at Top20

Results: Facebook Iceland

- Facebook: predict future friends
  - Adamic-Adar already works great
  - Logistic regression also strong
  - SRW gives slight improvement

Learning Method	AUC	Prec@20
Random Walk with Restart	0.81725	6.80
Adamic-Adar	0.81586	7.35
Common Friends	0.80054	7.35
Degree	0.58535	3.25
DT: Node features	0.59248	2.38
DT: Network features	0.76979	5.38
DT: Node+Network	0.76217	5.86
DT: Path features	0.62836	2.46
DT: All features	0.72986	5.34
LR: Node features	0.54134	1.38
LR: Network features	0.80560	7.56
LR: Node+Network	0.80280	7.56
LR: Path features	0.51418	0.74
LR: All features	0.81681	7.52
SRW: one edge type	**0.82502**	6.87
SRW: multiple edge types	**0.82799**	**7.57**

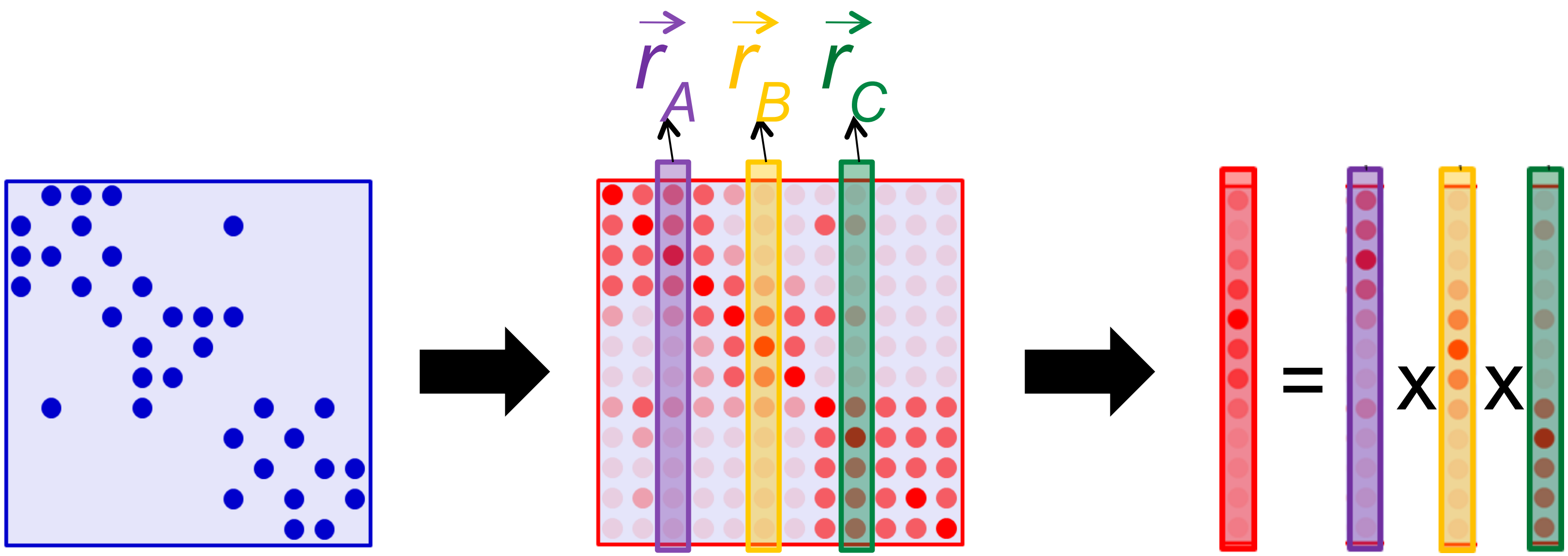
Results: Co-authorship

- Arxiv Hep-Ph collaboration network:
  - Poor performance of unsupervised methods
  - Logistic regression and decision trees don't work to well
  - SRW gives 10% boos in Prec@20

Learning Method	AUC	Prec@20
Random Walk with Restart	0.63831	3.41
Adamic-Adar	0.60570	3.13
Common Friends	0.59370	3.11
Degree	0.56522	3.05
DT: Node features	0.60961	3.54
DT: Network features	0.59302	3.69
DT: Node+Network	0.63711	3.95
DT: Path features	0.56213	1.72
DT: All features	0.61820	3.77
LR: Node features	0.64754	3.19
LR: Network features	0.58732	3.27
LR: Node+Network	0.64644	3.81
LR: Path features	0.67237	2.78
LR: All features	0.67426	3.82
SRW: one edge type	0.69996	4.24
SRW: multiple edge types	**0.71238**	**4.25**

details

Computing CePS Score (AND)



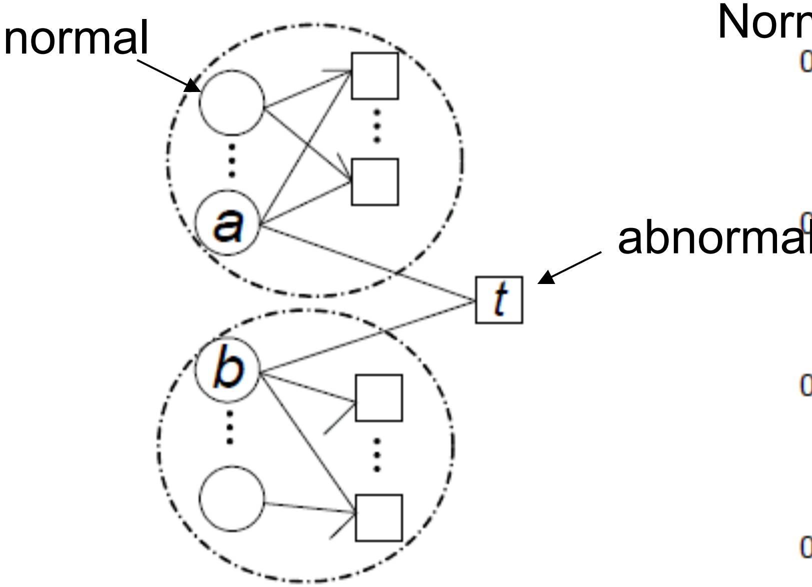
Normalized adj. matrix : W

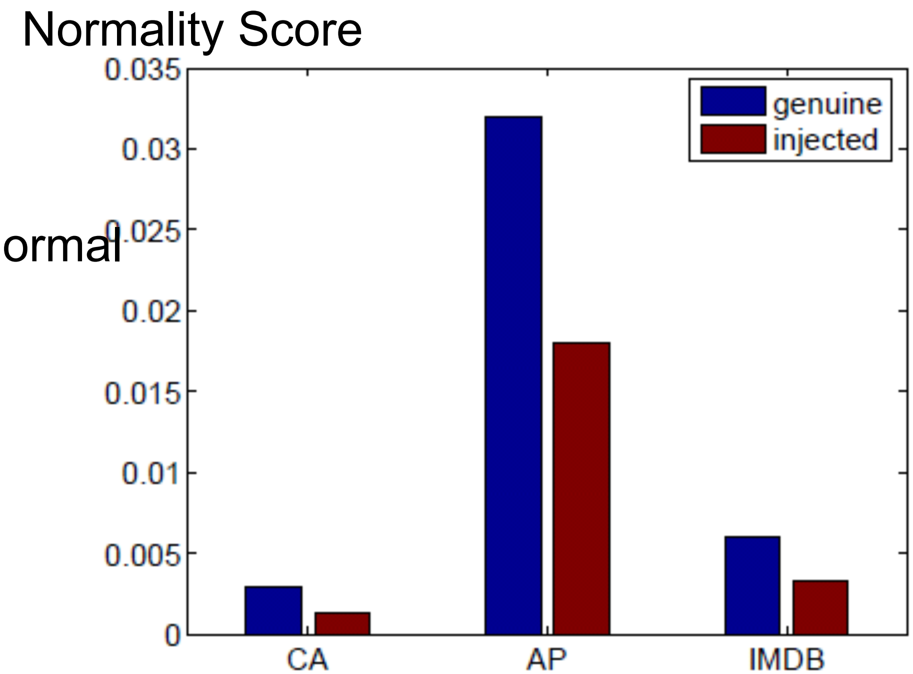
Proximity matrix: $Q=(I-cW)^{-1}$

CePS Score: $\vec{r}_{ceps} = \vec{r}_A \times \vec{r}_B \times \vec{r}_C$

Graph Anomalies by Proximity

[Sun+ ICDM 2005]



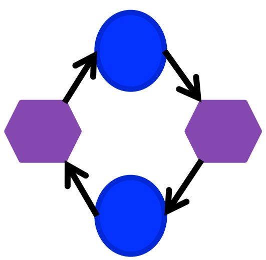


Normality Score = Average proximity among neighbors

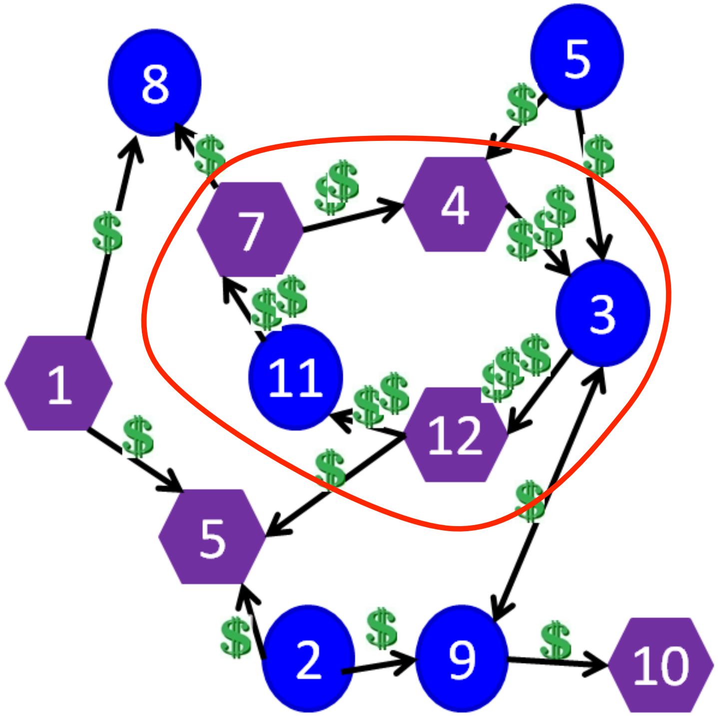
→ Essentially tries to find bridging nodes/edges

J. Sun, H. Qu, D. Chakrabarti, C. Faloutsos: Neighborhood Formation and Anomaly Detection in Bipartite Graphs. ICDM 2005

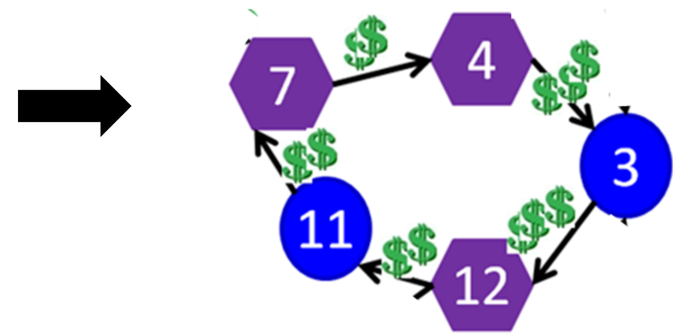
Graph Anomalies by Proximity



Query Graph



Transaction Network



Resulting Matching subgraph

Given: a query graph (pattern)

Find: best matching subgraph

Blue Node: anonymous account; Purple Node: Anonymous banks; Edge: Transaction

H. Tong, C. Faloutsos, B. Gallagher, T. Eliassi-Rad: Fast best-effort pattern matching in large attributed graphs. KDD 2007

Graph Anomalies by Belief Propagation

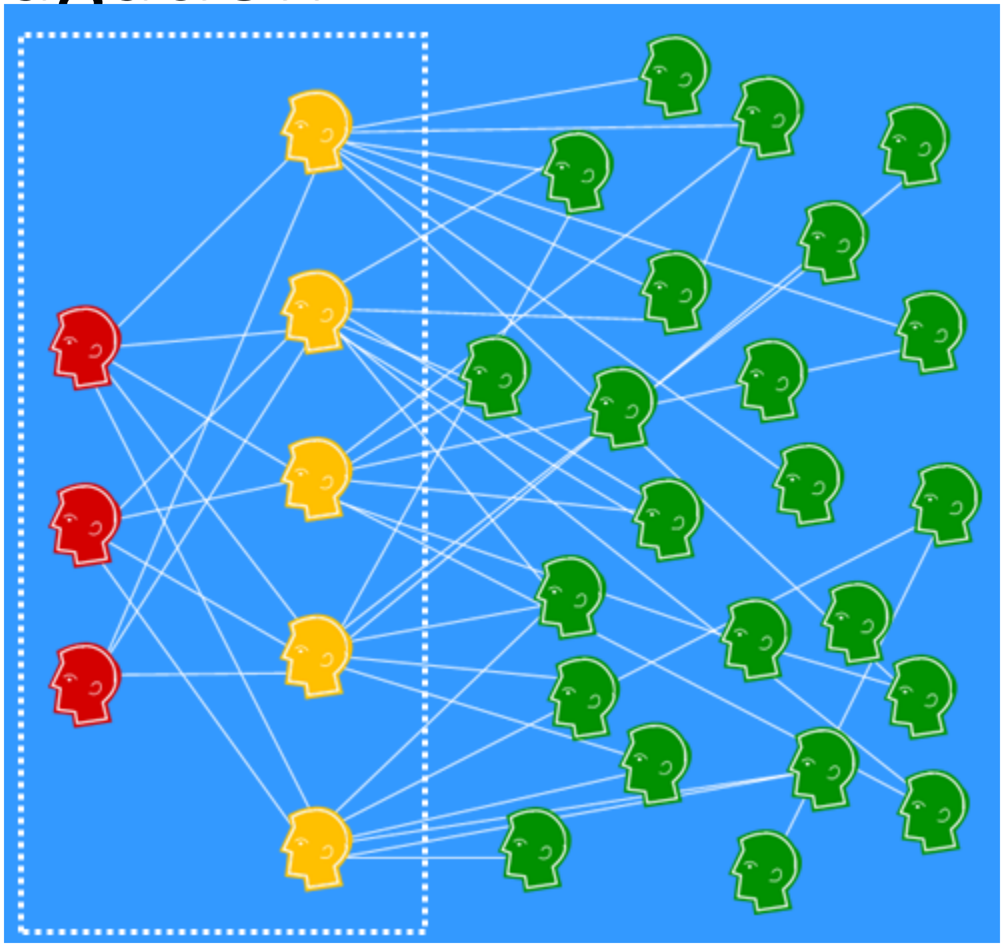
The bad guys (humans) create 2 types of users

Accomplice

- Trade mostly with honest users
- Looks legitimate

Fraudster

- Trade mostly with accomplices
- Don't trade with other fraudsters



Node: account; Edge: Positive Rating. (Graph constructed from e-bay on-line auction data)

S. Pandit, D. H. Chau, S. Wang, C. Faloutsos: Netprobe: a fast and scalable system for fraud detection in online auction networks. WWW 2007: 201-210