Recent Advances in \textbf{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

Conference

Author

Adjacency matrix: A

SDM 2013, Austin, Texas

May 1st-4th, 2013
Matrices in Social Networks

Research Qs: How to find common friends?
Matrices: rows/columns: users; entries: friendship
Matrix Tools: graph proximity
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 microorganism → 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
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

Q: How close is A to B?

a.k.a Relevance, Closeness, ‘Similarity’…
Basic Tech. : Random Walk with Restart

[Tong+ ICDM 2006]

Nearby nodes, higher scores
More red, more relevant

Ranking vector

\[ \mathbf{r}_4 \]
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

Same Diffusion Eq.
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

\[ r_i = c W^T p_i + (1 - c) e_i \]

- 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} \times \begin{pmatrix}
1/3 & 1/3 & 1/3 \\
1/3 & 1/3 & 1/4 \\
1/3 & 1/3 & 1/3 \\
1/3 & 1/2 & 1/2 & 1/4 \\
1/4 & 1/2 \\
1/4 & 1/2 \\
1/3 & 1/4 \\
1/2 & 1/3 \\
1/3 & 1/3 & 1/2 \\
1/3 & 1/3
\end{pmatrix} \times \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} = \begin{pmatrix} 0.9 \times \\
0.1 \times \end{pmatrix} + 0.1 \times \begin{pmatrix} 1 \end{pmatrix}

Footnote: “Maxwell Equation” for Web [Soumen Chakrabarti]
Recent Advance #1: Supervision

\[ r_i = c W r_i + (1 - c) e_i \]

- 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 $\rightarrow$ Graph Similarity/Kernel

- **Q:** $\text{Sim}(A_1, A_2)$?
- **A:** Do *two* random walks $(A_1, A_2)$!
- $\ldots = \text{one random walk on } A_x$

$$k(G, G') = \sum_k \lambda^k q_x A_x^k p_x = q_x^T(I - \lambda A_x)^{-1}p_x$$

SVN Vishwanathan. Fast computation of random walk graph kernels. NIPS 2006
K. Borgwardt and X. Yan: Graph Mining and Graph Kernels. KDD 2008 Tutorial
~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

• ...

May 1st-4th, 2013

SDM 2013, Austin, Texas
Low Rank Approximation

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

Nonnegative Matrix Factorization (NMF)

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

\[ X = (x_1, x_2, \cdots, x_n) \]

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

\[ FG^T \approx X \]
NMF Solutions: Multiplicative Updates

• Multiplicative update method

\[
F_{ij} \leftarrow F_{ij} \frac{(XG)_{ij}}{(FG^T G)_{ij}}
\]

\[
G_{ij} \leftarrow G_{ij} \frac{(X^TF)_{ij}}{(GFT^T F)_{ij}}
\]

NMF Solutions: Alternating Nonnegative Least Squares

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

(3) Block Principal Pivoting: [https://sites.google.com/site/jingukim/nmf_bpas.zip?attredirects=0](https://sites.google.com/site/jingukim/nmf_bpas.zip?attredirects=0)

NMF: Extensions

• General loss
  – Bregman Divergence

• Different constraints
  – Semi-NMF, Convex NMF, Symmetric NMF

• Incorporating supervisions
  – Pairwise constraints, label

• Multiple factorized matrices
  – Tri-factorization

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
Low Rank Approximation

• Nonnegative Matrix Factorization
• Nuclear norm related technologies
Rank Minimization and Nuclear Norm

- Matrix completion with rank minimization
  \[
  \min_{X} \text{rank}(X) \quad \text{s.t. } X_{ij} = M_{ij} \quad \forall (i, j) \in \Omega
  \]
  \[
  \text{NP hard}
  \]

- Convex relaxation
  \[
  \min_{X} \|X\|_* \quad \text{s.t. } X_{ij} = M_{ij} \quad \forall (i, j) \in \Omega
  \]
  \[
  \|X\|_* = \sum_i \sigma_i(X)
  \]

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/

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

L0-norm minimization

choose in this scenario?

variables with nonzero model parameters to be orthogonal to each other. Then what should we

we can construct different types of regularizers. In this part we will first introduce the various con-

As most of the data are sparse in nature, sparse learning has been a hot research direction in recent

nuclear norm, which is the summation of the singular values of a matrix, serves as a good surrogate to

fit different types of data, such as Tri-NMF, Convex NMF, Semi-NMF and Binary NMF. We will

low-rank nonnegative matrices. A lot of extensions of NMF have been proposed in recent years to

One popular research area along this line is Nonnegative Matrix Factorization (NMF), which was

Rank deficiency is a very common assumption for data-driven analytics based on matrix based rep-

4.2.1 Low-Rank Approximation

Relaxed Mixed boolean Convex Problem

L0-norm minimization

choose when we do sparse learning. For example, we may have two complete strategies to enforce

responding to similar variables to nonzero, another is minimum redundancy type which want the

duce the complexity of those data would be too high.

Another hot research area for low-rank approximation of data matrices we want to talk about is

4.2.2 Nuclear Norm Approximation

Matrix completion problem

nuclear-norm approximation. Instead of approximating the data matrix as the product of some small

introduce each of them briefly and talk about their application scenarios.

One popular research area along this line is Nonnegative Matrix Factorization (NMF), which was

originally proposed with the purpose of approximating a nonnegative matrix with the product of two

presentations. This assumption is very reasonable as it has been validated on many different types

http://www.stanford.edu/class/ee364b/lectures/l1_slides.pdf
Why L1 Norm Can Achieve Sparsity

\[ \hat{w}^0 = \arg \min_w \| y - Xw \|^2 \]

\[ \| y - Xw \|^2 = (w - \hat{w}^0)^\top X^\top X (w - \hat{w}^0) \]

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/


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

• Parallel Spectral Clustering

• Parallel SVD

• Parallel Optimization
Online Learning

• Online Matrix Factorization

• General Online Learning

• Parallel Online Learning
# Matrix Tools vs Applications

<table>
<thead>
<tr>
<th></th>
<th>C’ Patterns</th>
<th>Anomalies</th>
<th>Infl’ Prop</th>
<th>Immunization</th>
<th>Symptom Exp’</th>
<th>Similar Patient</th>
<th>Clinical Patterns</th>
<th>Risk Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prox.</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRA</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sparse L’</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large L’</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Social Networks**
- **Healthcare**

May 1st-4th, 2013
Outline

• Introduction
• Overview of the Technologies
• Applications in Health Informatics
• Applications in Social Informatics
• Conclusions and Future Works
Longitudinal Medical Records

- CPT
  - Date of Service

- ICD9, DxGroup
- HCC, Hierarchy
- Date of Service

Diagnosis
- Lab results
- Break down by age and sex groups

Lab
- Age
- Gender

Demography
- NDC
  - Ingredient
  - Days of Supplies
  - Date Filled

Pharmacy
- Date of Service
  - Procedure

Patient Records

Subjective:
- 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

Objective:
- 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 OSTEOARTHRITIS
- 427.31 ATRIAL FIBRILLATION

Assessment:
- BP 122/68 | Pulse 78 | Temp (Soc) 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

Plan:
- 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

Vector Space Model

\[ x = \begin{bmatrix} x_1 & x_2 & \cdots & x_i & \cdots & x_d \end{bmatrix} \]

Summary statistic of the i-th feature during a specific time period

Locally Supervised Metric Learning (LSML)

\[ d_{\Sigma}(x_i, x_j) = \sqrt{(x_i - x_j)^T \Sigma^{-1} (x_i - x_j)} \]

\[ \Sigma = WW^T \]

Identifying Neighborhoods

Patient Population

Heterogeneous

\[ c_i = \sum_{j \in \mathbb{N}_i} w_{ij} d(x_i, x_j) \]

\[ \mathbb{N}_i^H \]

Homogeneous

\[ s_i = \sum_{j \in \mathbb{N}_i^H} d(x_i, x_j) \]

\[ \mathbb{N}_i^L \]

Compute Average Distance

\[ \mathcal{S} = \sum_{i=1}^{n} (c_i - s_i) \]

Maximize Difference

\[ \min_{W} \text{tr}(W^T X (L^o - L^c) X^T W) \]

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


Online Adjustment of Patient Similarity

Physician Decision Support System

(1) input

physician

(2) retrieve

patient similarity assessment

(3) return

patient database

(4) display

visualization

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 $X(L^o - L^e)X^T$

Any physicians’ feedback is an increment on the matrix $L = L^o - L^e$

Efficient eigensystem update of a matrix: matrix perturbation theory

iMet: interactive Metric Learning

matrix

$$X(L + \Delta L)X^T$$

$X(L + \Delta L)X^T(w_i + \Delta w_i) = (\lambda_i + \Delta \lambda_i)(w_i + \Delta w_i)$

eigensystem

$$\begin{pmatrix} \lambda_i & w_i \\ \lambda_i & \tilde{w}_i \end{pmatrix}$$

perturbation

$$\begin{cases} \tilde{\lambda}_i &= \lambda_i + \Delta \lambda_i \\ \tilde{w}_i &= w_i + \Delta w_i \end{cases}$$

solution

$$\begin{align*}
\Delta \lambda_i &= w_i^T X \Delta LX^T w_i \\
\Delta w_i &= -\frac{1}{2} w_i + \sum_{j \neq i} \frac{w_j^T X \Delta LX^T w_i}{\lambda_i - \lambda_j} w_j
\end{align*}$$

**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

- Precision @ 3: \( \frac{2}{3} = 0.67 \)
- Precision @ 5: \( \frac{3}{5} = 0.6 \)
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

Patient Population

Expert 1

Expert 2

... 

Expert m

Neighborhood 1

Neighborhood 2

... 

Neighborhood m

\[ \min_{W, \alpha} \sum_{q=1}^{m} \alpha_q tr \left( W^T (\Sigma_q^0 - \Sigma_q^0) W \right) + \lambda \Omega(\alpha) \]

s.t. \[ \alpha \geq 0, \alpha^T e = 1 \]

Optimization Problem

Fix \( W \) Solve \( \alpha \)

Fix \( \alpha \) Solve \( W \)

Converge?

Output the Metric

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

Classification Accuracy Comparison
Applications in Health Informatics

• Patient Similarity Learning
• Risk Factor Identification
• Clinical Pattern Detection
Scalable Orthogonal Regression

Scalable Orthogonal Regression is to minimize

\[ J(\alpha) = \frac{1}{2} \| y - X\alpha \|^2 + \lambda \| \alpha \|_1 + \frac{\beta}{4} \sum_{ij} (\alpha_i x_i^T x_j \alpha_j)^2 \]

- **Regression Error**
- **Sparse Penalty**
- **Redundancy Penalty**

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

\[ X = [x_1, x_2, \cdots, x_p] \in \mathbb{R}^{n \times p} \]

\( \lambda \) and \( \beta \) 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 \( \alpha \).

- **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 \( |\alpha_j| \).
Scalability Comparison

- Metric: Computational time

**Fixing n**

- **# Features (n fixed to 5000)**
  - InfoGain
  - LARS
  - aSOR
  - mRMR

**Fixing p**

- **# Samples (p fixed)**
  - InfoGain
  - LARS
  - aSOR
  - mRMR
AUC Comparison
Augmented SOR

- SOR with pre-selected feature set

\[ f_p(\alpha_Q) = \frac{1}{2} \| y - X_Q \alpha_Q \|_2^2 + \lambda \| \alpha \|_1 \]
\[ + \frac{\beta}{4} \left[ \sum_{ij \in Q} (\alpha_i x_i^T x_j \alpha_j)^2 + \sum_{i \in Q, j \in P} (\alpha_i x_i^T x_j \alpha_j)^2 \right] \]

Redundancy among features to be selected

Redundancy between preselected and features to be selected

\[ \alpha_P = \arg \min_{\alpha} \| y - X_P \alpha \|_2^2 = (X_P^T X_P)^{-1} X_P^T y \]

- \( P \): 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

AUC vs Number of features

- CAD
- Diabetes
- Hypertension
- all knowledge features

May 1st-4th, 2013
SDM 2013, Austin, Texas
Finding Relevance: Mining Clinical Notes

SOAP sections of a Clinical Note

Key Idea: Symptom Expansion $\rightarrow$ graph node proximity.

i.e., which symptoms are most relevant to initial symptoms?

- Heart Attack
- Ankle Edema
- Chest Pain
- Rales
- Cough
- Wheezing
- Fever

Initial

Expanded

May 1st-4th, 2013

SDM 2013, Austin, Texas
**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)**

- Framingham: 0.59 (14.24%)
- Co-Occurrence: 0.649 (3.85%)
- LocSenseRWR: 0.674

**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


Temporal Patterns in Longitudinal Patient Records
One-Side Convolution

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

$$(F \ast g)_{ij} = \sum_{k=1}^{t} g_{j-k+1} F_{ik}$$

Note that $g_{j} = 0$ if $j \leq 0$ or $j > t$, and $F_{ik} = 0$ if $k > m$. 
One-Side Convolutional NMF

\[ \min_{F, G} J \]

s.t. \( \forall r = 1, \ldots, R; c = 1, \ldots, C \)
\( F^{(r)} \geq 0, g_c^{(r)} \geq 0 \)

\[ J = \sum_{c=1}^{C} d_\beta \left( A_c \odot X_c, A_c \odot \left( \sum_{r=1}^{R} F^{(r)} \star g_c^{(r)} \right) \right) + \lambda_1 \sum_{r=1}^{R} \| F^{(r)} \|_1 + \lambda_2 \sum_{c=1}^{C} \sum_{r=1}^{R} \| g_c^{(r)} \|_1 \]

**Definition** (\( \beta \)-divergence) The \( \beta \)-divergence between two matrices \( A \) and \( B \) with the same size is

\[ d_\beta(A, B) = \frac{1}{\beta(\beta - 1)} \sum_{ij} \left( A_{ij}^\beta + (\beta - 1)B_{ij}^\beta - \beta A_{ij}B_{ij}^{\beta - 1} \right) \]
Multiplicative Updates

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

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

\[
\eta(\beta) = \begin{cases} 
\frac{1}{2-\beta}, & \beta < 1 \\
1, & 1 \leq \beta \leq 2 \\
\frac{1}{\beta-1}, & \beta > 2 
\end{cases}
\]
A Synthetic Example
An Evaluation Case

Cases:

- Start date
- 12 months (training)
- End date
- 6 months (hold off)
- 1st CHF diagnosis date

Training period
Prediction period

Controls:

- Start date
- Diagnosed with Non-CHF heart disease
- 12 months (training)
- End date
- 6 months (hold off)
- Last record date

Training period
Prediction period
Bag-of-Pattern Representation

\[
\begin{bmatrix}
1 & 3 & 1 & 1 \\
0 & 2 & 1 & 2 \\
0 & 3 & 1 & 0 \\
\end{bmatrix}
\]
<table>
<thead>
<tr>
<th>Axis Title</th>
<th>Random Forest</th>
<th>SVM</th>
<th>Logistic Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 weeks</td>
<td>0.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 month</td>
<td>0.58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 months</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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]

Input: original graph

Output: CePS

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

Our Solution: \( \text{Max} (\text{Prox}(A, \text{Red}) \times \text{Prox}(B, \text{Red}) \times \text{Prox}(C, \text{Red})) \)
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”
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]
Q: How to find matching subgraph?
A: Proximity! [Tong+ KDD 2007 b]
G-Ray: How to?

Goodness = \text{Prox}(12, 4) \times \text{Prox}(4, 12) \times \text{Prox}(7, 4) \times \text{Prox}(4, 7) \times \text{Prox}(11, 7) \times \text{Prox}(7, 11) \times \text{Prox}(12, 11) \times \text{Prox}(11, 12)
Effectiveness: star-query

Query

May 1st-4th, 2013

Result

SDM 2013, Austin, Texas
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!
Key Observation: Graph Kernel $\Rightarrow$ Team-Aware Similarity

$$t = \arg\max_{j, j \notin \mathcal{T}} \text{Ker}(A(\mathcal{T}, \mathcal{T}), A(\mathcal{T}_{i_0 \rightarrow j}, \mathcal{T}_{i_0 \rightarrow j}))$$

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

<table>
<thead>
<tr>
<th>Input Graphs</th>
<th>Time Complexity (Our methods)</th>
<th>Time Complexity (Existing methods)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized, unlabelled</td>
<td>$O(n^2 r^4 + r^6 + mr)$</td>
<td>$O(n^3)$</td>
</tr>
<tr>
<td>Unnormalized, unlabelled</td>
<td>$O(nr + r^2 + mr)$</td>
<td>$O(n^3)$</td>
</tr>
<tr>
<td>Normalized, labelled</td>
<td>$O(d_n n^2 r^4 + r^6 + mr)$</td>
<td>$O(m^2 i_F)$</td>
</tr>
<tr>
<td>Unnormalized, labelled</td>
<td>$O(d_n n^2 r^4 + r^6 + mr)$</td>
<td>$O(m^2 i_F)$</td>
</tr>
</tbody>
</table>

Complexity Comparison

Empirical Evaluations

<table>
<thead>
<tr>
<th></th>
<th>HEP-TH</th>
<th>Oregon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ark-U</td>
<td>Ark-U+</td>
</tr>
<tr>
<td></td>
<td>0.999</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>0.977</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>0.962</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>0.952</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>0.946</td>
<td>0.998</td>
</tr>
</tbody>
</table>
Sense-Making of Marked Nodes [Akoglu+ SDM 2013]

(a) Too many connections?

(b) Too few connections?

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

+ ‘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:

1. Graph \(\rightarrow\) Adj. Matrix \(A\)
2. \(A = F \times G + R\)
3. Low-rank matrices
4. Residual matrix
5. Community
6. Anomalies

An Illustrative Example

Q: How to get the low-rank matrix approximations?
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 $A \in \mathbb{R}^{m \times n}$, find three matrices $C \in \mathbb{R}^{m \times c}$, $U \in \mathbb{R}^{c \times r}$, $R \in \mathbb{R}^{r \times n}$, such that $||A - CUR||$ is small

![Diagram showing matrix decomposition](image)

- 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

A Pictorial Comparison

SVD

CMD [Sun+ 2007]

May 1st-4th, 2013

2\textsuperscript{nd} singular vector

1\textsuperscript{st} singular vector

CUR [Drineas+ 2005]

Colibri-S [Tong+ 2008]
Performance Comparison

- **Accuracy**
  - Same 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 scanner, etc
- NrMF formulation

\[
\text{argmin}_{F,G} = \|R_{n \times l} \otimes W_{n \times l}\|_F^2 \quad \text{Weighted Frobenius Form}
\]

Common in Any MF

\[
= \sum_{i=1}^{n} \sum_{j=1}^{l} (A(i,j) - F(i,:)G(:,j))^2 W(i,j)^2 \quad \text{Weight}
\]

s.t. for all \( A(i,j) > 0 \):

\[
F(i,:)G(:,j) \leq A(i,j) \quad \text{Non-negative residual}
\]

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

(a) strange connection

(b) port scanning

(c) ddos

(d) bipartite core
Applications in Social Informatics

- Finding Complex User Patterns
- (Matrix-based) Anomaly Detection

Influence and Virus Propagation
An Example: Flu/Virus Propagation

Q: How to minimize infected population?
- Q1: Understand tipping point
- Q2: Affecting algorithms
Why Do We Care?

Email Fwd in Organization

Aug 5, 09:30:12 "data request"

Aug 5, 09:53:00 "Fw: data request"

Aug 6, 14:21:53 "Fw: Fw: data request"

Rumor Prop on Twitter in UK riots

10% credit $10% off $10% credit $10% off $10% credit $10% off $10% credit $10% off

Viral Marketing

SDM 2013, Austin, Texas

Malware Infection
Theorem [Chakrabarti+ 2003, 2007]:
If $\lambda \times (\beta / \delta) \leq 1$; no epidemic for any initial conditions

$\beta$: Prob (\text{infection} \rightarrow \text{infected})

$\delta$: Prob (\text{infected} \rightarrow \text{recovery})

$\rho_{t+1} = H(p_t)$

$\lambda$: largest eigenvalue of the graph (~connectivity of the graph)
$\beta, \delta$: virus parameters (~strength of the virus)

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

- Q1: Understand tipping point
- Q2: Affecting algorithms

SIS Model (e.g., Flu)
Why is $\lambda$ So Important?

- $\lambda \rightarrow$ Capacity of a Graph:

$$\left( \bar{1}^* A^k 1 \right)^{1/k} \xrightarrow{k \to \infty} \lambda$$

- Q1: Understand tipping point
- Q2: Affecting algorithms

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

Larger $\lambda \rightarrow$ better connected
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.

- Q1: Understand tipping point
- Q2: Affecting algorithms

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

- Select $k$ nodes, whose absence creates the largest drop in $\lambda$

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

**Original Graph: $\lambda$**

**Without $\{2, 6\}$: $\lambda_S$**

- Q1: Understand tipping point
- Q2: Affecting algorithms
Optimal Method

• Select \( k \) nodes, whose absence creates the largest drop in \( \lambda \)

\[
S = \arg \max_{|S|=k} \left( \lambda - \lambda_S \right)
\]

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

Theorem: (Tong+ CIKM 2012)
Find Optimal \( k \)-node Immunization is NP-Hard
Theorem: (Tong+ 2010)
(1) \( \lambda - \lambda_s \approx \Sigma_v(S) = \Sigma_{i \in S} 2\lambda u(i)^2 - \Sigma_{i,j \in S} A(i,j)u(i)u(j) \)

Footnote: \( u(i) \sim \text{PageRank}(i) \sim \text{in-degree}(i) \)
Netshield to the Rescue

Theorem: (Tong+ 2010)
(1) \( \lambda - \lambda_s \approx Sv(S) = \sum_{i \in S} 2\lambda u(i)^2 \cdot A(i,j)u(i)u(j) \)

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

Original Graph \[ \rightarrow \]
Select by C1 \[ \rightarrow \]
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 \( Sv(S^{Netshield}) \geq (1-1/e) \) \( Sv(S^{Opt}) \)
Comparison of Immunization

Log(fraction of infected nodes)

Netshield

PageRank

Between (short)

Degree

Between (RW)

Acquaintance

Eigs (=HITS)

Time Ticks

Time step
Hospital Infection Controlling (US-Medicare Network)

Using a variant of Netshield (adapting to partial immunity)
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 1\textsuperscript{st} order perturbation, we have

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

• So, we are done $\Rightarrow$ need $O(n^2-m)$ complexity
Maximizing Propagation: Edge Addition

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

• Q: How to Find k new edges w/ highest $G_v(S)$?

• A: Modified Fagin’s algorithm

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

- Q1: Understand tipping point
- Q2: Affecting algorithms
Maximizing Propagation: Evaluation

- Q1: Understand tipping point
- Q2: Affecting algorithms

Log (Infected Ratio)

Time Ticks

May 1st-4th, 2013
SDM 2013, Austin, Texas
Influence & Virus Propagation: Summary

• **Goal**: Guild Dissemination by Opt. Link Structure

• **Theory**: Opt. Dissemination = Opt. \( \lambda \)

• **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

<table>
<thead>
<tr>
<th>C' Patterns</th>
<th>Anomalies</th>
<th>Influ' Prop</th>
<th>Immunization</th>
<th>Symptom Exp’</th>
<th>Similar Patient</th>
<th>Clinical Patterns</th>
<th>Risk Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tools</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prox.</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRA</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Sparse L’</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Large L’</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

### Social Networks

- C' Patterns
- Anomalies
- Influ' Prop
- Immunization
- Symptom Exp’
- Similar Patient
- Clinical Patterns

### Healthcare

- C' Patterns
- Anomalies
- Influ' Prop
- Immunization
- Symptom Exp’
- Similar Patient
- Clinical Patterns

---

May 1st-4th, 2013
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
Recent Advance #4: Scale-Up

GBASE: a scalable and general graph management system. KDD 2011

PEGASUS: mining peta-scale graphs. KAIS 2011
Recent Advance #4: Scale-Up

Storage Savings
- RAW
- NNB
- NZB

Scalability (size of graph)
- KR-NNB
- ER-NNB
- ER-NZB
- KR-NZB

Running Time Savings
- 1-Nh
- 2-Nh
- Egonet

Scalability (# of Machine)
- RAW
- NNB
- NZB
- CZB

U. Kang, C. Tsourakakis, C. Faloutsos: PEGASUS: mining peta-scale graphs. KAIS 2011
Learning $\mathcal{W}$: 

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

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

$s$: be the center node; $l$: destination; $d$: no-link
edge strength $a_{uv} = f_w(u,v) = \exp(-w^T \Psi_{uv})$
feature vector $\Psi_{uv}$ (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

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

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

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

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

Thanks to Jure Leskovec: Social Media Analytics (KDD '11 tutorial)
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% boost in Prec@20

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

Thanks to Jure Leskovec: Social Media Analytics (KDD '11 tutorial)
Computing CePS Score (AND)

Normalized adj. matrix: $W$

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

CePS Score:
$r_{ceps} = r_A \times r_B \times r_C$
Graph Anomalies by Proximity

[Sun+ ICDM 2005]

Normality Score

\[ \text{Normality Score} = \text{Average proximity among neighbors} \]

\[ \rightarrow \text{Essentially tries to find bridging nodes/edges} \]

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

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 eBay online auction data)