

# MultiRank and HAR for Ranking Multi-relational Data, Transition Probability Tensors, and Multi-Stochastic Tensors

Michael K. Ng

Centre for Mathematical Imaging and Vision and  
Department of Mathematics  
Hong Kong Baptist University  
Email: [mng@math.hkbu.edu.hk](mailto:mng@math.hkbu.edu.hk)

July 23, 2016

- Background
- Algorithms: PageRank and HITs
- Algorithms: MultiRank and HAR
- Numerical Examples: Information Retrieval, Community Network and Image Retrieval
- Theoretical Results: Transition Probability Tensors
- Theoretical Results: Multi-Stochastic Tensors
- Concluding Remarks

- Matrix can be used to describe the relationship between objects, and objects with different attributes:

	$O_1$	$O_2$	$\dots$	$O_n$		$A_1$	$A_2$	$\dots$	$A_m$
$O_1$	$a_{1,1}$	$a_{1,2}$	$\dots$	$a_{1,n}$	$O_1$	$a_{1,1}$	$a_{1,2}$	$\dots$	$a_{1,m}$
$O_2$	$a_{2,1}$	$a_{2,2}$	$\dots$	$a_{2,n}$	$O_2$	$a_{2,1}$	$a_{2,2}$	$\dots$	$a_{2,m}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
$O_n$	$a_{n,1}$	$a_{n,2}$	$\dots$	$a_{n,n}$	$O_n$	$a_{n,1}$	$a_{n,2}$	$\dots$	$a_{n,m}$

Examples: (left) a similarity matrix, an image, a Google matrix; (right) a gene expression data, multivariate data, terms and documents.

- large data ( $n$  is large); high-dimensional data ( $m$  is large)

# Tensor

- Tensor can be used to describe the multiple relationships between objects. A tensor is a multidimensional array. Here a three-way array is used:

	$O_1$	$O_2$	$\dots$	$O_n$
$O_1$	$a_{1,1,1}$	$a_{1,2,1}$	$\dots$	$a_{1,n,1}$
$O_2$	$a_{2,1,1}$	$a_{2,2,1}$	$\dots$	$a_{2,n,1}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
$O_n$	$a_{n,1,1}$	$a_{n,2,1}$	$\dots$	$a_{n,n,1}$

	$O_1$	$O_2$	$\dots$	$O_n$
$O_1$	$a_{1,1,2}$	$a_{1,2,2}$	$\dots$	$a_{1,n,2}$
$O_2$	$a_{2,1,2}$	$a_{2,2,2}$	$\dots$	$a_{2,n,2}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
$O_n$	$a_{n,1,2}$	$a_{n,2,2}$	$\dots$	$a_{n,n,2}$

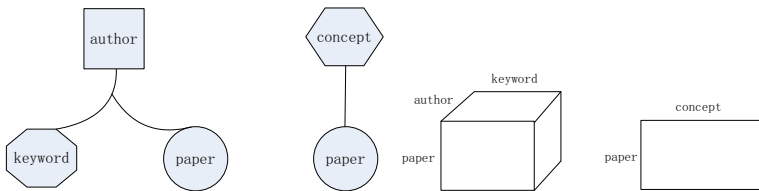
	$O_1$	$O_2$	$\dots$	$O_n$
$O_1$	$a_{1,1,p}$	$a_{1,2,p}$	$\dots$	$a_{1,n,p}$
$O_2$	$a_{2,1,p}$	$a_{2,2,p}$	$\dots$	$a_{2,n,p}$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
$O_n$	$a_{n,1,p}$	$a_{n,2,p}$	$\dots$	$a_{n,n,p}$

- $p$  relationships among  $n$  objects

- Web information retrieval is significantly more challenging than that based on web hyperlink structure
- One main difference is the multiple links based on the other features (text, images, etc)
- Example: 100,000 webpages from .GOV Web collection in 2002 TREC and 50 topic distillation topics in TREC 2003 Web track as queries
- Multiple links among webpages via different anchor texts
- 39,255 anchor terms (multiple relations), and 479,122 links with these anchor terms among the 100,000 webpages

# Tensor

- In a social network where objects are connected via multiple relations, via sharing, comments, stories, photos, tags, keywords, topics, etc
- In a publication network where the interactions among items in three entities: author, keyword and paper



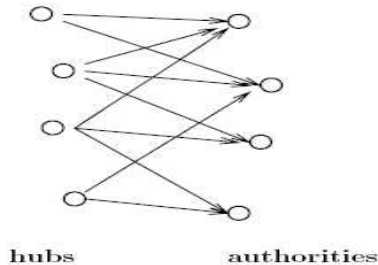
- A tensor: the interactions among items in three dimensions/entities: author, keyword and paper; A matrix: the interactions between items in two dimensions/entities: concept and paper

- The hyperlink structure is exploited by three of the most frequently cited Web IR methods: HITS (Hypertext Induced Topic Search), PageRank and SALSA.
- **PageRank**: L. Page, S. Brin, R. Motwani and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. 1998.
- **HITS**: J. Kleinberg. Authoritative Sources in a Hyperlinked Environment. Journal of the ACM, 46: 604-632, 1999.
- **SALSA**: R. Lempel and S. Moran. The Stochastic Approach for Link-structure Analysis (SALSA) and the TKC effect. The Ninth International WWW Conference, 2000.

[The survey given by A. Langville and C. Meyer, A Survey of Eigenvector Methods for Web Information Retrieval, SIAM Review, 2005.]

# HITS

Each page/document on the Web is represented as a node in a very large graph. The directed arcs connecting these nodes represent the hyperlinks between the documents.





- The HITS IR method defines authorities and hubs. An authority is a document with several inlinks, while a hub has several outlinks.
- The HITS thesis is that good hubs point to good authorities and good authorities are pointed to by good hubs. HITS assigns both a hub score and authority score to.
- Page  $i$  has both a hub score  $x_i$  and an authority score  $y_i$ . Let  $E$  be the set of all directed edges in the Web graph and let  $e_{i,j}$  represent the directed edge from node  $i$  to node  $j$ .

Given that each page has been assigned an initial hub score  $x_i^{(0)}$  and authority score  $y_i^{(0)}$ , HITS successively refines these scores by computing

$$y_j^{(k)} = \sum_{i:e_{i,j}} x_i^{(k-1)} \quad \text{and} \quad x_i^{(k)} = \sum_{j:e_{i,j}} y_j^{(k-1)}$$

With the help of the adjacency matrix  $L$  of the directed Web graph:  $L_{i,j} = 1$ , if there exists an edge from node  $i$  to node  $j$ ,  $L_{i,j} = 0$ , otherwise.

$$\mathbf{y}^{(k)} = L^T \mathbf{x}^{(k-1)} \quad \text{and} \quad \mathbf{x}^{(k)} = L \mathbf{y}^{(k)}$$

- These two equations define the iterative power method for computing the dominant eigenvector for the matrices  $L^T L$  and  $LL^T$ . Since the matrix  $L^T L$  determines the authority scores, it is called the authority matrix and  $LL^T$  is known as the hub matrix.
- $L^T L$  and  $LL^T$  are symmetric positive semi-definite matrices. Computing the hub vector  $x$  and the authority vector  $y$  can be viewed as finding dominant right-hand eigenvectors of  $LL^T$  and  $L^T L$ , respectively.
- The structure of  $L$  allows to be a repeated root of the characteristic polynomial, in which case the associated eigenspace would be multi-dimensional. This means that the different limiting authority (and hub) vectors can be produced by different choices of the initial vector.

- PageRank importance is determined by “votes” in the form of links from other pages on the web. The idea is that votes (links) from important sites should carry more weight than votes (links) from less important sites, and the significance of a vote (link) from any source should be tempered (or scaled) by the number of sites the source is voting for (linking to).
- The rank  $r(s)$  of a given page  $s$  to be

$$r(s) = \sum_{t \in B_s} \frac{r(t)}{\#(t)}$$

where  $B_s = \{ \text{all pages pointing to } s \}$ , and  $\#(t)$  is the number of out links from  $t$ .

- Compute

$$\pi = P\pi \quad (\text{column vector})$$

where  $P$  is the matrix with  $p_{i,j} = 1/\#(s_j)$  if  $s_j$  links to  $s_i$ , otherwise  $p_{i,j} = 0$ .

# PageRank

- The raw Google matrix  $P$  is nonnegative with column sums equal to one or zero.
- Zero column sums correspond to pages that have no outgoing links such pages are sometimes referred to as dangling nodes.
- Dangling nodes can be accounted for by artificially adding appropriate links to make all column sums equal one, then  $P$  is a column stochastic matrix, which in turn means that the PageRank iteration represents the evolution of a Markov chain.
- This Markov chain is a random walk on the graph defined by the link structure of the web pages in the Google database.

An irreducible Markov chain is one in which every state is eventually reachable from every other state. That is, there exists a path from node  $j$  to node  $i$  for all  $i, j$ .

Irreducibility is a desirable property because it is precisely the feature that guarantees that a Markov chain possesses a unique (and positive) stationary distribution vector  $\pi = P\pi$ , the Perron-Frobenius theorem at work:

$$\tilde{P} = \alpha P + (1 - \alpha)E$$

where  $\alpha$  is a scalar between 0 and 1, and  $E$  is a matrix of all ones by the number of pages. The matrix  $E$  can guarantee irreducible matrix  $\tilde{P}$ .

- Stochastic Approach for Link Structure Analysis (SALSA). SALSA, developed by Lempel and Moran in 2000, was spawned by the combination of ideas from both HITS and PageRank.
- Like HITS, both hub and authority scores for webpages are created, and like PageRank, they are created through the use of Markov chains:

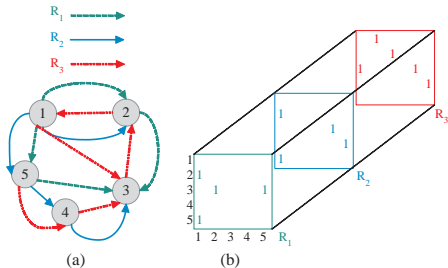
$$\mathbf{y}^{(k)} = \text{Normalize}(L^T)\mathbf{x}^{(k-1)} \quad \text{and} \quad \mathbf{x}^{(k)} = \text{Normalize}(L)\mathbf{y}^{(k)}$$

Both  $\text{Normalize}(L^T)$  and  $\text{Normalize}(L)$  are transition probability matrices (column sum is equal to 1).

# Multi-relation Data: The Representation

Example: five objects and three relations (**R1: green**, **R2: blue**, **R3: red**) among them.

A tensor is a multidimensional array. Example: a three-way array is used, where each two dimensional slice represents an adjacency matrix for a single relation. The data can be represented as a tensor of size  $5 \times 5 \times 3$  where  $(i, j, k)$  entry is nonzero if the  $i$ th object is connected to the  $j$ th object by using the  $k$ th relation.





# Notation

Let  $\mathbb{R}$  be the real field. We call  $\mathcal{T} = (t_{i_1, i_2, j_1})$  where  $t_{i_1, i_2, j_1} \in \mathbb{R}$ , for  $i_k = 1, \dots, m$ ,  $k = 1, 2$  and  $j_1 = 1, \dots, n$ , a real  $(2, 1)$ th order  $(m \times n)$ -dimensional rectangular tensor.

$(i_1, i_2)$  to be the indices for **objects** and  $j_1$  to be the indices for **relations**.

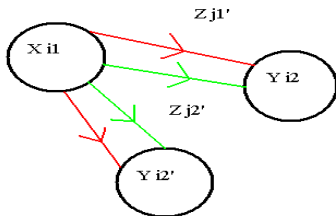
For instance, five objects ( $m = 5$ ) and three relations ( $n = 3$ ) are used in the example. When there is a link from the  $i_1$ th object to the  $i_2$ th object when the  $j_1$ th relation is used, we set  $t_{i_1, i_2, j_1} = 1$ , otherwise  $t_{i_1, i_2, j_1} = 0$ .

$\mathcal{T}$  is called non-negative if  $t_{i_1, i_2, j_1} \geq 0$ .

# The Contribution

- Propose a framework to study the hub and authority scores of objects in multi-relational data for query search.
- Besides hub score  $x$  and authority score  $y$  for each object, we assign a relevance score  $z$  for each relation to indicate its importance in multi-relational data.
- These three scores have the following mutually-reinforcing relationships: (i) An object that points to many objects with high authority scores through relations of high relevance scores, receives a high hub score. (ii) An object that is pointed by many objects with high hub scores through relations of high relevance scores, receives a high authority score. (iii) A relation that is connected in between objects with high hub scores and high authority scores, receives a high relevance score.

# The Idea



$$\sum_{i_2=1}^m \sum_{j_1=1}^n h_{i_1, i_2, j_1} y_{i_2} z_{j_1} = x_{i_1}, \quad 1 \leq i_1 \leq m$$

$$\sum_{i_1=1}^m \sum_{j_1=1}^n a_{i_1, i_2, j_1} x_{i_1} z_{j_1} = y_{i_2}, \quad 1 \leq i_2 \leq m$$

$$\sum_{i_1=1}^m \sum_{i_2=1}^m r_{i_1, i_2, j_1} x_{i_1} y_{i_2} = z_{j_1}, \quad 1 \leq j_1 \leq n$$

# The HAR Model

Transition Probability Tensors:

$\mathcal{H} = (h_{i_1, i_2, j_1})$ ,  $\mathcal{A} = (a_{i_1, i_2, j_1})$  and  $\mathcal{R} = (r_{i_1, i_2, j_1})$  with respect to hubs, authorities and relations by normalizing the entry of  $\mathcal{T}$  as follows:

$$h_{i_1, i_2, j_1} = \frac{t_{i_1, i_2, j_1}}{m}, \quad i_1 = 1, 2, \dots, m,$$
$$\sum_{i_1=1} t_{i_1, i_2, j_1}$$

$$a_{i_1, i_2, j_1} = \frac{t_{i_1, i_2, j_1}}{m}, \quad i_2 = 1, 2, \dots, m,$$
$$\sum_{i_2=1} t_{i_1, i_2, j_1}$$

$$r_{i_1, i_2, j_1} = \frac{t_{i_1, i_2, j_1}}{n}, \quad j_1 = 1, 2, \dots, n.$$
$$\sum_{j_1=1} t_{i_1, i_2, j_1}$$

# The Interpretation

- To compute hub and authority scores of objects and relevance scores of relations by considering a random walk in a multi-relational data/tensor, and studying the limiting probabilities having objects as hubs or as authorities, and using relations respectively.
- $h_{i_1, i_2, j_1}$  (or  $a_{i_1, i_2, j_1}$ ) can be interpreted as the probability of having the  $i_1$ th (or  $i_2$ th) object as an hub (or as an authority) by given that the  $i_2$ th (or  $i_1$ th) object as an authority (or as a hub) is currently considered and the  $j_1$ th relation is used;
- $r_{i_1, i_2, j_1}$  can be interpreted as the probability of using the  $j_1$ th relation given that the  $i_2$ th object as an authority is visited from the  $i_1$ th object as a hub.

# The Interpretation

$$h_{i_1, i_2, j_1} = \text{Prob}[X_t = i_1 | Y_t = i_2, Z_t = j_1]$$

$$a_{i_1, i_2, j_1} = \text{Prob}[Y_t = i_2 | X_t = i_1, Z_t = j_1]$$

$$r_{i_1, i_2, j_1} = \text{Prob}[Z_t = j_1 | Y_t = i_2, X_t = i_1]$$

$X_t$ ,  $Y_t$  and  $Z_t$  are random variables referring to consider at any particular object as a hub and as an authority, and to use at any particular relation respectively at the time  $t$  respectively. Here the time  $t$  refers to the time step in a Markov chain (generalization).

We remark that the construction of  $a_{i_1, i_2, j_1}$  is related to the transpose of  $h_{i_1, i_2, j_1}$ . This is similar to the construction of SALSA algorithm to incorporate the link structure among objects for the role of hub and authority in the single relation data.

# HAR - Tensor Equations

hub score:  $\bar{x}$

authority score:  $\bar{y}$

relevance score:  $\bar{z}$

$$\mathcal{H}\bar{y}\bar{z} = \bar{x}, \quad \mathcal{A}\bar{x}\bar{z} = \bar{y}, \quad \mathcal{R}\bar{x}\bar{y} = \bar{z},$$

with

$$\sum_{i_1=1}^m \bar{x}_{i_1} = 1, \quad \sum_{i_2=1}^m \bar{y}_{i_2} = 1, \quad \sum_{j_1=1}^n \bar{z}_{j_1} = 1.$$

hub and authority scores for objects and relevance scores for relations by solving tensor equations based on mutually-reinforcing relationship among hubs, authorities and relations.

# Generalization

When we consider a single relation type, we can set  $\bar{\mathbf{z}}$  to be a vector  $\mathbf{1}/n$  of all ones, and thus we obtain two matrix equations

$$\mathcal{H}\bar{\mathbf{y}}\mathbf{1}/n = \bar{\mathbf{x}} \quad \mathcal{A}\bar{\mathbf{x}}\mathbf{1}/n = \bar{\mathbf{y}}.$$

We remark that  $\mathcal{A}$  can be viewed as the transpose of  $\mathcal{H}$ . This is exactly the same as that we solve for the singular vectors to get the hub and authority scoring vectors in SALSA. As a summary, the proposed framework HAR is a generalization of SALSA to deal with multi-relational data.



# The MultiRank Model

Transition Probability Tensors:

$\mathcal{A} = (a_{i_1, i_2, j_1})$  and  $\mathcal{R} = (r_{i_1, i_2, j_1})$  with respect to nodes and relations by normalizing the entry of  $\mathcal{T}$  as follows:

$$a_{i_1, i_2, j_1} = \frac{t_{i_1, i_2, j_1}}{m}, \quad i_2 = 1, 2, \dots, m,$$
$$\sum_{i_2=1} t_{i_1, i_2, j_1}$$

$$r_{i_1, i_2, j_1} = \frac{t_{i_1, i_2, j_1}}{n}, \quad j_1 = 1, 2, \dots, n.$$
$$\sum_{j_1=1} t_{i_1, i_2, j_1}$$

$$\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} = \bar{\mathbf{x}}, \quad \mathcal{R}\bar{\mathbf{x}}^2 = \bar{\mathbf{z}},$$

with

$$\sum_i^m \bar{x}_i = 1, \quad \sum_j^n \bar{z}_j = 1.$$

# HAR - Query Search

To deal with query processing, we need to compute hub and authority scores of objects and relevance scores of relations with respect to a query input (like topic-sensitive PageRank):

$$(1 - \alpha)\mathcal{H}\bar{\mathbf{y}}\bar{\mathbf{z}} + \alpha\mathbf{o} = \bar{\mathbf{x}},$$

$$(1 - \beta)\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} + \beta\mathbf{o} = \bar{\mathbf{y}},$$

$$(1 - \gamma)\mathcal{R}\bar{\mathbf{x}}\bar{\mathbf{y}} + \gamma\mathbf{r} = \bar{\mathbf{z}},$$

where  $\mathbf{o}$  and  $\mathbf{r}$  are two assigned probability distributions that are constructed from a query input, and  $0 \leq \alpha, \beta, \gamma < 1$ , are three parameters to control the input.

$$F(\mathbf{x}, \mathbf{y}, \mathbf{z}) = \begin{pmatrix} (1 - \alpha)\mathcal{H}\bar{\mathbf{y}}\bar{\mathbf{z}} + \alpha\mathbf{o} \\ (1 - \beta)\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} + \beta\mathbf{o} \\ (1 - \gamma)\mathcal{R}\bar{\mathbf{x}}\bar{\mathbf{y}} + \gamma\mathbf{r} \end{pmatrix}$$

$$\Omega_m = \{\mathbf{u} = (u_1, u_2, \dots, u_m) \in R^m \mid u_i \geq 0, 1 \leq i \leq m, \sum_{i=1}^m u_i = 1\}$$

and

$$\Omega_n = \{\mathbf{w} = (w_1, w_2, \dots, w_n) \in R^n \mid w_j \geq 0, 1 \leq j \leq n, \sum_{j=1}^n w_j = 1\}$$

## Theorem

Suppose  $\mathcal{H}$ ,  $\mathcal{A}$  and  $\mathcal{R}$  are constructed from  $\mathcal{T}$ ,  $0 \leq \alpha, \beta, \gamma < 1$ , and  $\mathbf{o} \in \Omega_m$  and  $\mathbf{r} \in \Omega_n$  are given. If  $\mathcal{T}$  is irreducible, then there exist  $\bar{\mathbf{x}} > 0$ ,  $\bar{\mathbf{y}} > 0$  and  $\bar{\mathbf{z}} > 0$  such that  $(1 - \alpha)\mathcal{H}\bar{\mathbf{y}}\bar{\mathbf{z}} + \alpha\mathbf{o} = \bar{\mathbf{x}}$ ,  $(1 - \beta)\mathcal{A}\bar{\mathbf{x}}\bar{\mathbf{z}} + \beta\mathbf{o} = \bar{\mathbf{y}}$ , and  $(1 - \gamma)\mathcal{R}\bar{\mathbf{x}}\bar{\mathbf{y}} + \gamma\mathbf{r} = \bar{\mathbf{z}}$ , with  $\bar{\mathbf{x}}, \bar{\mathbf{y}} \in \Omega_m$  and  $\bar{\mathbf{z}} \in \Omega_n$ .

Using Brouwer's Fixed Point Theorem.

## Theorem

Suppose  $\mathcal{T}$  is irreducible, and  $\mathcal{H}$ ,  $\mathcal{A}$  and  $\mathcal{R}$  are constructed from  $\mathcal{T}$ ,  $0 \leq \alpha, \beta, \gamma < 1$  and  $\mathbf{o} \in \Omega_m$  and  $\mathbf{r} \in \Omega_n$  are given. If 1 is not the eigenvalue of the Jacobian matrix of the mapping  $F$ , then the solution vectors  $\bar{\mathbf{x}}$ ,  $\bar{\mathbf{y}}$  and  $\bar{\mathbf{z}}$  are unique.

Using Kellogg's Unique Fixed Point Results.

# HAR - Theory ( $\alpha, \beta, \gamma > 1/2$ )

## Theorem

Suppose  $\mathcal{H}$ ,  $\mathcal{A}$  and  $\mathcal{R}$  are constructed from  $\mathcal{T}$ , and  $\mathbf{o} \in \Omega_m$  and  $\mathbf{r} \in \Omega_n$  are given. If  $1/2 < \alpha, \beta, \gamma < 1$  then the solution vectors  $\bar{\mathbf{x}}$ ,  $\bar{\mathbf{y}}$  and  $\bar{\mathbf{z}}$  are unique.

Check the Jacobian matrix

$$J = \begin{pmatrix} 0 & (1 - \alpha)J_{12} & (1 - \alpha)J_{13} \\ (1 - \beta)J_{21} & 0 & (1 - \beta)J_{23} \\ (1 - \gamma)J_{31} & (1 - \gamma)J_{32} & 0 \end{pmatrix}$$

$J_{st}$  are transition probability matrices (their column sum are equal to 1). The largest magnitude of the eigenvalue of  $J$  is less than 1.

# The HAR Algorithm

**Input:** Three tensors  $\mathcal{H}$ ,  $\mathcal{A}$  and  $\mathcal{R}$ , two initial probability distributions  $\mathbf{y}_0$  and  $\mathbf{z}_0$  with  $(\sum_{i=1}^m [\mathbf{y}_0]_i = 1$  and  $\sum_{j=1}^n [\mathbf{z}_0]_j = 1)$ , the assigned probability distributions of objects and/or relations  $\mathbf{o}$  and  $\mathbf{r}$  ( $\sum_{i=1}^m [\mathbf{o}]_i = 1$  and  $\sum_{j=1}^n [\mathbf{r}]_j = 1$ ), three weighting parameters  $0 \leq \alpha, \beta, \gamma < 1$ , and the tolerance  $\epsilon$

**Output:** Three limiting probability distributions  $\bar{\mathbf{x}}$  (authority scores),  $\bar{\mathbf{y}}$  (hub scores) and  $\bar{\mathbf{z}}$  (relevance values)

**Procedure:**

- 1: Set  $t = 1$ ;
- 2: Compute  $\mathbf{x}_t = (1 - \alpha)\mathcal{H}\mathbf{y}_{t-1}\mathbf{z}_{t-1} + \alpha\mathbf{o}$ ;
- 3: Compute  $\mathbf{y}_t = (1 - \beta)\mathcal{A}\mathbf{x}_{t-1}\mathbf{z}_{t-1} + \beta\mathbf{o}$ ;
- 4: Compute  $\mathbf{z}_t = (1 - \gamma)\mathcal{R}\mathbf{x}_{t-1}\mathbf{y}_{t-1} + \gamma\mathbf{r}$ ;
- 5: If  $\|\mathbf{x}_t - \mathbf{x}_{t-1}\| + \|\mathbf{y}_t - \mathbf{y}_{t-1}\| + \|\mathbf{z}_t - \mathbf{z}_{t-1}\| < \epsilon$ , then stop, otherwise set  $t = t + 1$  and goto Step 2.

## Convergence ( $\alpha, \beta, \gamma > 1/2$ )

### Theorem

*Suppose  $\mathcal{H}$ ,  $\mathcal{A}$  and  $\mathcal{R}$  constructed, and  $\mathbf{o} \in \Omega_m$  and  $\mathbf{r} \in \Omega_n$  are given. If  $1/2 < \alpha, \beta, \gamma < 1$  then the HAR algorithm converges to the unique vectors  $\bar{\mathbf{x}}$ ,  $\bar{\mathbf{y}}$  and  $\bar{\mathbf{z}}$ .*

Gauss Seidel iterations:

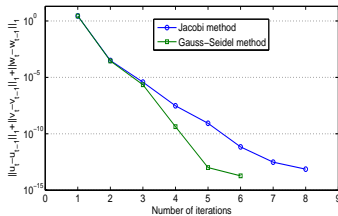
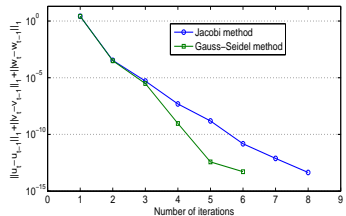
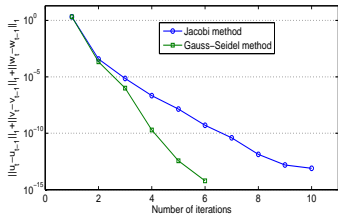
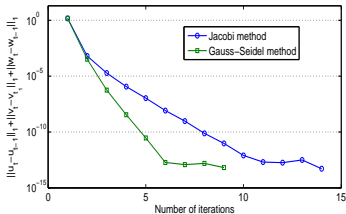
Compute  $\mathbf{x}_t = (1 - \alpha)\mathcal{H}\mathbf{y}_{t-1}\mathbf{z}_{t-1} + \alpha\mathbf{o}$ ;

Compute  $\mathbf{y}_t = (1 - \beta)\mathcal{A}\mathbf{x}_t\mathbf{z}_{t-1} + \beta\mathbf{o}$ ;

Compute  $\mathbf{z}_t = (1 - \gamma)\mathcal{R}\mathbf{x}_t\mathbf{y}_t + \gamma\mathbf{r}$ ;

# Gauss Seidel is faster than Jacobi

Example in Information Retrieval:





# Experiment 1

- 100,000 webpages from .GOV Web collection in 2002 TREC and 50 topic distillation topics in TREC 2003 Web track as queries
- Links among webpages via different anchor texts
- 39,255 anchor terms (multiple relations), and 479,122 links with these anchor terms among the 100,000 webpages
- If the  $i_2$ th webpage links to the  $i_1$ th webpage via the  $j_1$ th anchor term, we set the entry  $t_{i_1, i_2, j_1}$  of  $\mathcal{T}$  to be one. The size of  $\mathcal{T}$  is  $100,000 \times 100,000 \times 39,255$
- Sparse tensors: the percentage of nonzero entries is  $1.22 \times 10^{-7}\%$

	P@10	P@20	NDCG@10	NDCG@20	MAP	R-prec
HITS	0.0000	0.0000	0.0000	0.0000	0.0041	0.0000
SALSA	0.0160	0.0140	0.0157	0.0203	0.0114	0.0084
TOPHITS (500-rank)	0.0020	0.0010	0.0044	0.0028	0.0008	0.0002
TOPHITS (1000-rank)	0.0040	0.0020	0.0088	0.0057	0.0016	0.0010
TOPHITS (1500-rank)	0.0040	0.0030	0.0063	0.0049	0.0011	0.0018
BM25+ DeplnOut	0.0280	0.0180	0.0419	0.0479	0.0370	0.0370
<b>HAR</b> (rel. query)	0.0560	0.0410	0.0659	0.0747	0.0330	0.0552
<b>HAR</b> (rel. and obj. query)	<b>0.1100</b>	<b>0.0800</b>	<b>0.1545</b>	<b>0.1765</b>	<b>0.1035</b>	<b>0.1051</b>

The results of all comparison algorithms on TREC data set.

ranks	HAR
1	<a href="http://www.tempe.gov/library/youth/teach.htm">www.tempe.gov/library/youth/teach.htm</a>
2	<a href="http://www.get.wa.gov/reading.htm">www.get.wa.gov/reading.htm</a>
3	<a href="http://www.tempe.gov/library/content.htm">www.tempe.gov/library/content.htm</a>
4	<a href="http://www.ed.gov/pubs/Paraprofessionals/norfolk.html">www.ed.gov/pubs/Paraprofessionals/norfolk.html</a>
5	<a href="http://www.loc.gov/rr/international/int-gateway.html">www.loc.gov/rr/international/int-gateway.html</a>
6	<a href="http://graham.sannet.gov/public-library/searching-the-net/childrensites.shtml">graham.sannet.gov/public-library/searching-the-net/childrensites.shtml</a>
7	<a href="http://www.cde.ca.gov/ci/literature/">www.cde.ca.gov/ci/literature/</a>
8	<a href="http://www.cde.ca.gov/news/releases2001/rel36.asp">www.cde.ca.gov/news/releases2001/rel36.asp</a>
9	<a href="http://www.hud.gov/lea/lestand.html/">www.hud.gov/lea/lestand.html/</a>
10	<a href="http://libwww.library.phila.gov/databases/keywordsearch.taf">libwww.library.phila.gov/databases/keywordsearch.taf</a>
ranks	TOPHITS
1	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139239&amp;cftoken=44790442&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1139239&amp;cftoken=44790442&amp;Topic=0101</a>
2	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139177&amp;cftoken=3776877&amp;Topic=0404">www.students.gov/link_search/listlinks.cfm?cfid=1139177&amp;cftoken=3776877&amp;Topic=0404</a>
3	<a href="http://www.crh.noaa.gov/dlh/firewx.htm">www.crh.noaa.gov/dlh/firewx.htm</a>
4	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139251&amp;cftoken=50170076&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1139251&amp;cftoken=50170076&amp;Topic=0101</a>
5	<a href="http://www.bls.gov/oes/2000/oestec2000.htm">www.bls.gov/oes/2000/oestec2000.htm</a>
6	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139205&amp;cftoken=97602529&amp;Topic=0203">www.students.gov/link_search/listlinks.cfm?cfid=1139205&amp;cftoken=97602529&amp;Topic=0203</a>
7	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1130081&amp;cftoken=49786299&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1130081&amp;cftoken=49786299&amp;Topic=0101</a>
8	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139251&amp;cftoken=50170076&amp;Topic=0103">www.students.gov/link_search/listlinks.cfm?cfid=1139251&amp;cftoken=50170076&amp;Topic=0103</a>
9	<a href="http://www.crh.noaa.gov/sgf/hydro/reports/hydrorpt.shtml">www.crh.noaa.gov/sgf/hydro/reports/hydrorpt.shtml</a>
10	<a href="http://graham.sannet.gov/directories/privacy.shtml">graham.sannet.gov/directories/privacy.shtml</a>

ranks	HAR
1	<a href="http://www.npwrc.usgs.gov/">www.npwrc.usgs.gov/</a>
2	<a href="http://pastel.npsc.nbs.gov/resource/2001/impplan/appenE.htm">pastel.npsc.nbs.gov/resource/2001/impplan/appenE.htm</a>
3	<a href="http://water.usgs.gov/eap/env_guide/fish_wildlife.html">water.usgs.gov/eap/env_guide/fish_wildlife.html</a>
4	<a href="http://fedlaw.gsa.gov/legal2a.htm">fedlaw.gsa.gov/legal2a.htm</a>
5	<a href="http://envirotext.eh.doe.gov/data/uscode/16/">envirotext.eh.doe.gov/data/uscode/16/</a>
6	<a href="http://envirotext.eh.doe.gov/data/uscode/16/ch5A.html">envirotext.eh.doe.gov/data/uscode/16/ch5A.html</a>
7	<a href="http://envirotext.eh.doe.gov/data/uscode/16/ch6.html">envirotext.eh.doe.gov/data/uscode/16/ch6.html</a>
8	<a href="http://www.nwr.noaa.gov/1salmon/salmesa/pubs.htm">www.nwr.noaa.gov/1salmon/salmesa/pubs.htm</a>
9	<a href="http://www.dfg.ca.gov/hcpb/conplan/mitbank/banking_report.shtml">www.dfg.ca.gov/hcpb/conplan/mitbank/banking_report.shtml</a>
10	<a href="http://www.hawaii.gov/dlnr/ldxConservation.htm">www.hawaii.gov/dlnr/ldxConservation.htm</a>
ranks	TOPHITS
1	<a href="http://www.npwrc.usgs.gov/">www.npwrc.usgs.gov/</a>
2	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1133639&amp;cftoken=90444830&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1133639&amp;cftoken=90444830&amp;Topic=0101</a>
3	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1133456&amp;cftoken=20839758&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1133456&amp;cftoken=20839758&amp;Topic=0101</a>
4	<a href="http://www.fedstats.gov/qf/states/29000.html">www.fedstats.gov/qf/states/29000.html</a>
5	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1133457&amp;cftoken=22178553&amp;Topic=0404">www.students.gov/link_search/listlinks.cfm?cfid=1133457&amp;cftoken=22178553&amp;Topic=0404</a>
6	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139247&amp;cftoken=47516940&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1139247&amp;cftoken=47516940&amp;Topic=0101</a>
7	<a href="http://www.crh.noaa.gov/sgf/products/mtruno.shtml">www.crh.noaa.gov/sgf/products/mtruno.shtml</a>
8	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139201&amp;cftoken=34746433&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1139201&amp;cftoken=34746433&amp;Topic=0101</a>
9	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1130149&amp;cftoken=13070299&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1130149&amp;cftoken=13070299&amp;Topic=0101</a>
10	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139266&amp;cftoken=37252845&amp;Topic=0203">www.students.gov/link_search/listlinks.cfm?cfid=1139266&amp;cftoken=37252845&amp;Topic=0203</a>

ranks	HAR
1	<a href="http://usembassy.state.gov/seoul/wwwwh42x4.html">usembassy.state.gov/seoul/wwwwh42x4.html</a>
2	<a href="http://usembassy.state.gov/seoul/wwwhe404.html">usembassy.state.gov/seoul/wwwhe404.html</a>
3	<a href="http://www.state.gov/r/pa/bgn/2792pf.htm">www.state.gov/r/pa/bgn/2792pf.htm</a>
4	<a href="http://www.state.gov/r/pa/bgn/2792.htm">www.state.gov/r/pa/bgn/2792.htm</a>
5	<a href="http://www.voa.gov/korean/">www.voa.gov/korean/</a>
6	<a href="http://www.st.nmfs.gov/st3/multilateral.html">www.st.nmfs.gov/st3/multilateral.html</a>
7	<a href="http://www.usinfo.state.gov/regional/ea/easec/nkoreapg.htm">www.usinfo.state.gov/regional/ea/easec/nkoreapg.htm</a>
8	<a href="http://usembassy.state.gov/seoul/wwwwh42yt.html">usembassy.state.gov/seoul/wwwwh42yt.html</a>
9	<a href="http://usembassy.state.gov/seoul/wwwwh0108.html">usembassy.state.gov/seoul/wwwwh0108.html</a>
10	<a href="http://citrus.sbaonline.sba.gov/nd/ndoutreach.html">citrus.sbaonline.sba.gov/nd/ndoutreach.html</a>
ranks	TOPHITS
1	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139239&amp;cftoken=44790442&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1139239&amp;cftoken=44790442&amp;Topic=0101</a>
2	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139177&amp;cftoken=3776877&amp;Topic=0404">www.students.gov/link_search/listlinks.cfm?cfid=1139177&amp;cftoken=3776877&amp;Topic=0404</a>
3	<a href="http://www.crh.noaa.gov/dlh/firewx.htm">www.crh.noaa.gov/dlh/firewx.htm</a>
4	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139251&amp;cftoken=50170076&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1139251&amp;cftoken=50170076&amp;Topic=0101</a>
5	<a href="http://www.bls.gov/oes/2000/oestec2000.htm">www.bls.gov/oes/2000/oestec2000.htm</a>
6	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139205&amp;cftoken=97602529&amp;Topic=0203">www.students.gov/link_search/listlinks.cfm?cfid=1139205&amp;cftoken=97602529&amp;Topic=0203</a>
7	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1130081&amp;cftoken=49786299&amp;Topic=0101">www.students.gov/link_search/listlinks.cfm?cfid=1130081&amp;cftoken=49786299&amp;Topic=0101</a>
8	<a href="http://www.students.gov/link_search/listlinks.cfm?cfid=1139251&amp;cftoken=50170076&amp;Topic=0103">www.students.gov/link_search/listlinks.cfm?cfid=1139251&amp;cftoken=50170076&amp;Topic=0103</a>
9	<a href="http://www.crh.noaa.gov/sgf/hydro/reports/hydrorpt.shtm">www.crh.noaa.gov/sgf/hydro/reports/hydrorpt.shtm</a>
10	<a href="http://graham.sannet.gov/directories/privacy.shtml">graham.sannet.gov/directories/privacy.shtml</a>

## Experiment 2

- ① five conferences (SIGKDD, WWW, SIGIR, SIGMOD, CIKM)
- ② Publication information includes title, authors, reference list, and classification categories associated with publication
- ③ 6848 publications and 617 different categories
- ④ 100 category concepts as query inputs to retrieve the relevant publications
- ⑤ Tensor:  $6848 \times 6848 \times 617$ , If the  $i_2$ th publication cites the  $i_1$ th publication and the  $i_1$ th publication has the  $j_1$ th category concept, then we set the entry  $t_{i_1, i_2, j_1}$  of  $\mathcal{T}$  to be one, otherwise we set the entry  $t_{i_1, i_2, j_1}$  to be zero.
- ⑥ 24901 nonzero entries, the percentage of the nonzero entries is  $8.61 \times 10^{-5}\%$

	P@10	P@20	NDCG@10	NDCG@20	MAP	R-prec
HITS	0.2260	0.1815	0.3789	0.3792	0.2522	0.2751
SALSA	0.4100	0.3105	0.5606	0.5352	0.3462	0.3929
TOPHITS (50-rank)	0.1360	0.1145	0.1684	0.1557	0.0566	0.0617
TOPHITS (100-rank)	0.1640	0.1340	0.2012	0.1857	0.0646	0.0732
TOPHITS (150-rank)	0.1920	0.1410	0.2315	0.1998	0.0732	0.0765
BM25+ DeplnOut	0.0170	0.0145	0.0147	0.0138	0.0162	0.0109
<b>HAR</b> <i>(rel. query)</i>	<b>0.5880</b>	<b>0.4155</b>	<b>0.7472</b>	<b>0.6760</b>	<b>0.4731</b>	<b>0.4683</b>

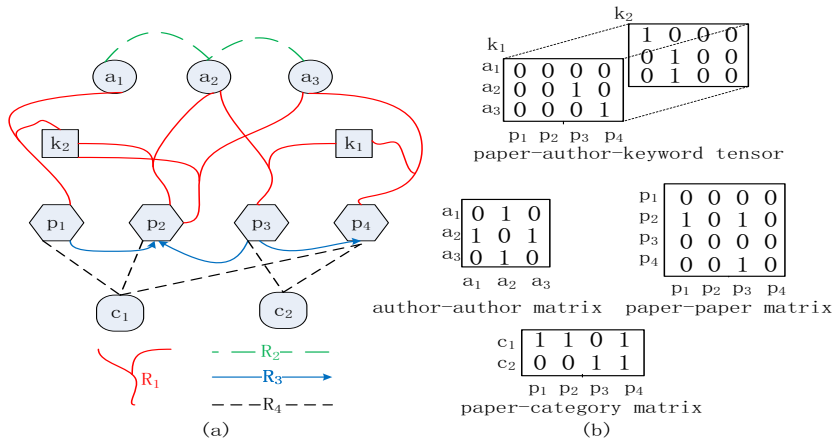
The results of all comparison algorithms on DBLP data set.

	P@10	P@20	NDCG@10	NDCG@20	MAP	R-prec
HAR (rel. query)	0.2000	0.1312	0.1488	0.1647	0.1312	0.2075
HAR (rel. and obj. query)	<b>0.6000</b>	<b>0.7500</b>	<b>0.6995</b>	<b>0.7697</b>	<b>0.5422</b>	<b>0.6226</b>

The results of HAR with two settings when we query “clustering” concept and “document” related papers. In this case, we judge a paper to be relevant if it has “clustering” concept and a “document” related title for evaluation.



# Community Discovery



(a) An example of an academic publication network where  $R_1$  indicates paper-author-keyword interaction,  $R_2$  indicates author-author collaboration,  $R_3$  indicates paper-paper citation,  $R_4$  indicates paper-category interaction.

(b) One tensor and three matrices are used to represent the interactions among authors, papers, keywords and categories.

For SIAM journal data, we consider the papers published in SIMAX, SISC and SINUM in the recent 15 years

Multi-dimensional networks:

the paper-author-keyword ( $x_1-x_2-x_3$ ) tensor  $\mathcal{P}^{(1)}$

[3736  $\times$  1807  $\times$  7630;  $nz = 31257$  6.07  $\times 10^{-5}\%$ ];

the paper-paper ( $x_1-x_1$ ) citation tensor  $\mathcal{P}^{(2)}$  [3736  $\times$  3736;

$nz = 4523$  0.032%];

the author-author ( $x_2-x_2$ ) collaboration tensor  $\mathcal{P}^{(3)}$  [1807  $\times$  1807;

$nz = 4410$  0.14%];

the paper-category ( $x_1-x_4$ ) concept tensor  $\mathcal{P}^{(4)}$  [3736  $\times$  1202;

$nz = 11897$  0.26%]

$\mathcal{P}^{(1)} \rightarrow \mathcal{P}^{(1,1)}, \mathcal{P}^{(1,2)}, \mathcal{P}^{(1,3)}$

$\mathcal{P}^{(2)} \rightarrow \mathcal{P}^{(2,1)}$

$\mathcal{P}^{(3)} \rightarrow \mathcal{P}^{(3,1)}$

$\mathcal{P}^{(4)} \rightarrow \mathcal{P}^{(4,1)}, \mathcal{P}^{(4,2)}$

Multi-dimensional networks tensor equations:

$$\mathbf{x}_1 = (1 - \alpha) \left[ \frac{1}{3} \mathcal{P}^{(1,1)} \mathbf{x}_2 \mathbf{x}_3 + \frac{1}{3} \mathcal{P}^{(2,1)} \mathbf{x}_1 + \frac{1}{3} \mathcal{P}^{(4,1)} \mathbf{x}_4 \right] + \alpha \mathbf{z}_1$$

$$\mathbf{x}_2 = (1 - \alpha) \left[ \frac{1}{2} \mathcal{P}^{(1,2)} \mathbf{x}_1 \mathbf{x}_3 + \frac{1}{2} \mathcal{P}^{(3,1)} \mathbf{x}_2 \right] + \alpha \mathbf{z}_2$$

$$\mathbf{x}_3 = (1 - \alpha) \mathcal{P}^{(1,3)} \mathbf{x}_1 \mathbf{x}_2 + \alpha \mathbf{z}_3$$

$$\mathbf{x}_4 = (1 - \alpha) \mathcal{P}^{(4,2)} \mathbf{x}_1 + \alpha \mathbf{z}_4$$

## Numerical Example

Community	MetaFac	Clauset	LWP	MultiComm
SIMAX	0.3544	0.0612	0.0022	<b>0.6396</b>
SINUM	0.4591	0.4321	0.0001	<b>0.5113</b>
SISC	0.3868	0.3047	0.0056	<b>0.4321</b>
average F-measure	0.4001	0.2660	0.0026	<b>0.5277</b>
KDD	0.5741	0.7662	0.0020	<b>0.7722</b>
SIGIR	0.6497	0.6664	0.0015	<b>0.7350</b>
average F-measure	0.6119	0.7163	0.0018	<b>0.7536</b>

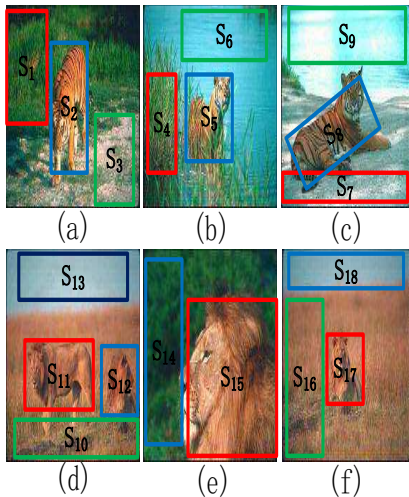
Community(AMS code)	MetaFac	Clauset	LWP	MultiComm
65F10	0.2123	0.1143	0.0090	<b>0.4602</b>
65F15	0.1641	0.2197	0.0222	<b>0.4282</b>
65N30	0.2133	0.0701	0.0059	<b>0.3672</b>
65N15	0.1341	0.0802	0.0152	<b>0.3244</b>
65N55	0.1743	0.2555	0.0001	<b>0.3321</b>
average F-measure	0.1796	0.1480	0.0105	<b>0.3824</b>

# Image Retrieval

- An image can be represented by several visual concepts, and a tensor is built based on visual concepts as its entry contain the affinity between two images in the corresponding visual concept.
- A ranking scheme can be used to compute the association scores of images and the relevance scores of visual concepts by employing input query vectors to handle image retrieval: MultiRank and HAR algorithms.

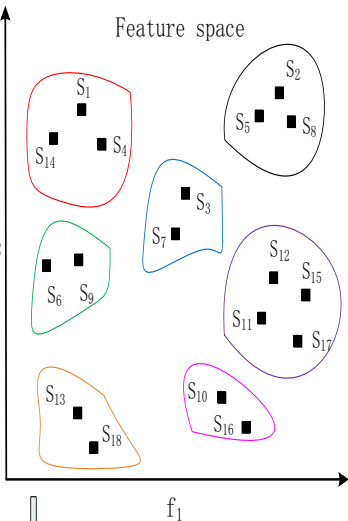
# Image Retrieval

Image collection



(1)  $\rightarrow f_2$

Feature space



(2)

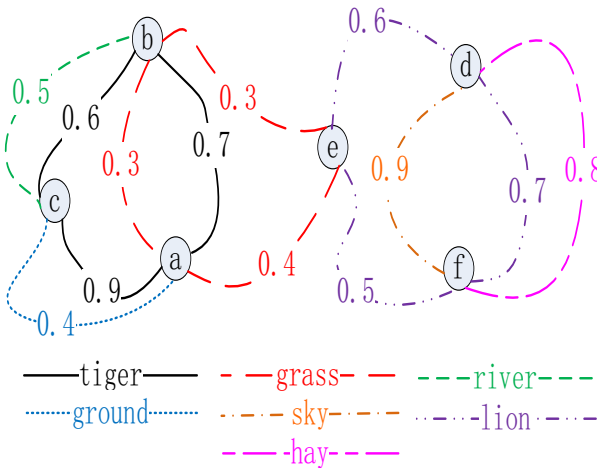
# Image Retrieval

Query image



(3)

Hypergraph



(4)

Ranking result

# Image Retrieval

ground

sky  
hay

lion

(4)

Ranking result



(a)

score=0.3908



(b)

score=0.2238



(c)

score=0.1588



(e)

score=0.1362



(d)

score=0.0452



(f)

score=0.0452



# Image Retrieval

5000 images from Corel data are used. The images in this collection belong to 50 categories and each category has 100 images.

Methods	P@5	P@10	P@20	NDCG@5	NDCG@10	NDCG@20	MAP	R-prec
MultiVCRankI	0.4888	0.4494	0.3935	0.5042	0.4715	0.4245	0.1938	0.2252
MultiVCRankII	<b>0.5004</b>	<b>0.4578</b>	<b>0.4003</b>	<b>0.5142</b>	<b>0.4798</b>	<b>0.4318</b>	<b>0.2058</b>	<b>0.2317</b>
TOPHITS	0.3408	0.3266	0.2909	0.3463	0.3346	0.3069	0.1183	0.1552
HypergraphRank	0.4188	0.3668	0.3122	0.4432	0.3986	0.3491	0.1597	0.1813
ManifoldRank	0.4244	0.3690	0.3140	0.4485	0.4016	0.3515	0.1604	0.1820
SimilarityRank	0.4312	0.3776	0.3200	0.4561	0.4101	0.3580	0.1620	0.1837

# Image Retrieval

