Latent factor model: methodology, theory and applications

Junhui Wang

City University of Hong Kong

Collaborators: Annie Qu(UCI/UIUC), Xiaotong Shen(UMN)
Ben Dai(CUHK), Xin He(SUFE), Shirong Xu(Purdue), Jingnan Zhang, Yaoming Zhen(CityU)

Supported in part by HK GRF Grants 11303918, 11300919, 11304520, 11206821

Latent factor model

- LFM is a major statistical tool for multivariate data, and goes back to Spearman (1904) on latent intelligence factor
- LFM is particularly suitable for modeling dyadic and multi-adic relational data
- Entities in the relational data can be embedded with low-dimensional latent factors
  - customers' ratings on products/services
  - test-takers' responses to test items
  - users' linking behavior on social networks
  - and many others ...

Application I: recommender system

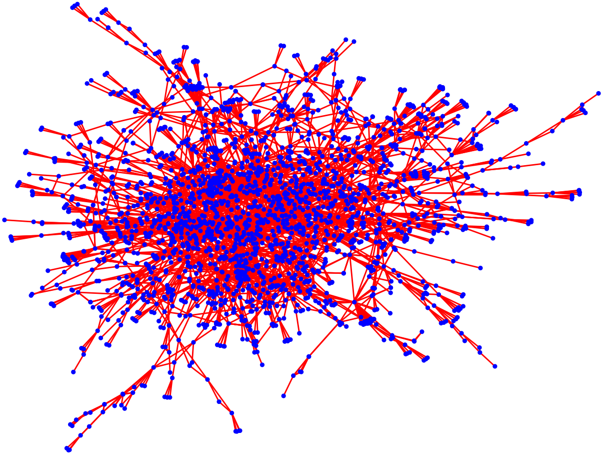
A system that recommends items to users, by tracking users' preferences and making personalized predictions

					
	2	5	?	4	?
	5	4	?	?	?
	?	1	5	?	?
	?	?	?	3	?
	?	?	?	?	?
	?	4	?	5	?

Netflix competition

- Netflix initiated a million-dollar competition in 2006
- Data: movie ratings (from 1 to 5)
  - Training: $\sim$100 million ratings, 480,000 users, 18,000 movies
  - Test: $\sim$1.5 million ratings but withheld
- Observations are scarse; about 99% missing
- “Cold-start" issue

Application II: co-authorship network



Remark: dataset from Ji & Jin (2016), which consists of 3248 papers and 3607 authors.

Network structures

- Network data captures pairwise interactions among nodes of interest
- Interaction can be undirected or directed
- Community structures
- Sparse networks; small averaged node degree

A general LFM framework

- Let $\mathbf{R} = (r_{ui})_{n \times m}$ be the available dyadic relational data
  - In RS, $r_{ui} \in \mathcal{R}$ is user u's rating on item i
  - In directed network, $r_{ui} \in \{0, 1\}$ denotes if user u sends an edge to user i
  - In undirected network, $r_{ui} = r_{iu}$
- Low-rank assumption: $\text{rank}(\mathbf{R}) \leq K$
- The LFM model assumes that

$$E(r_{ui}) = \theta_{ui} = \mathbf{p}_u^T \mathbf{q}_i,$$

where $\mathbf{p}_u$ and $\mathbf{q}_i$ are K-dim latent factors

Smooth RS (Dai, W., Shen and Qu, 2019)

- If **R** is completely observed, SVD of **R** suffices
- In practice, only a very small part of **R** is observed,

$$\min_{\boldsymbol{P},\boldsymbol{Q}} \sum_{u=1}^{n} \sum_{i=1}^{m} \Big(\sum_{(u',i')\in\Omega} \omega_{ui,u'i'} r_{u'i'} - \boldsymbol{p}_u^T \boldsymbol{q}_i \Big)^2 + \lambda J(\mathbf{P}, \mathbf{Q})$$

  where Ω is the index set with observed ratings
  - $\omega_{ui,u'i'}$ is defined based on social network and covariates
  - $\sum_{(u',i')\in\Omega} \omega_{ui,u'i'} r_{u'i'}$ echoes the idea of data imputation
  - $J(\mathbf{P}, \mathbf{Q}) = \|\mathbf{P}\|_F^2 + \|\mathbf{Q}\|_F^2$; other penalties are possible
- Smooth RS enlarges the effective sample size, and provides an effective treatment for the “cold start" issue

Directed community detection (Zhang, He and W., 2021)

- Given a directed network with $r_{ij} \sim \text{Bern}(p_{ij})$,

$$logit(p_{ij}) = \theta_{ij} = \boldsymbol{\alpha}_i^T \boldsymbol{\beta}_j$$

where $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_j$ are r-dimensional latent factors

- Likelihood of the network is

$$P(\mathcal{G}) = \prod_{i,j=1}^{n} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}} = \prod_{i,j=1}^{n} \frac{\exp(\boldsymbol{\alpha}_i^T \boldsymbol{\beta}_j a_{ij})}{1 + \exp(\boldsymbol{\alpha}_i^T \boldsymbol{\beta}_j)}$$

- The regularized formulation is

$$L_\lambda(\boldsymbol{\alpha}, \boldsymbol{\beta}) = -\frac{1}{n^2} \log\left(P(\mathcal{G})\right) + \lambda_0 J_0(\boldsymbol{\alpha}, \boldsymbol{\beta}) + \lambda_1 J_1(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

Regularization terms

- $J_0(\alpha, \beta) = \|\alpha\|_F^2 + \|\beta\|_F^2$ helps overcome non-identifiability
- $J_1(\alpha, \beta)$ encourages community structure,

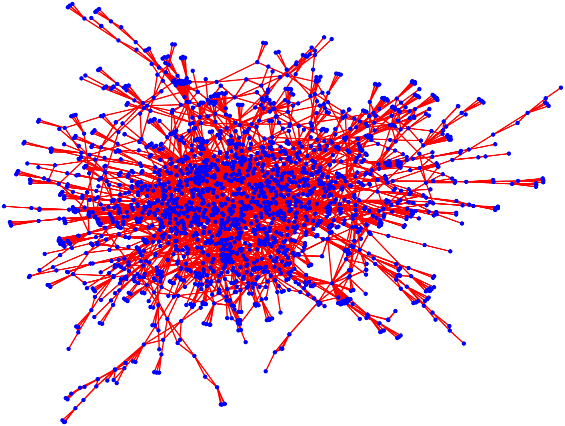
$$\min_{\{\mathbf{c}_l^{out}\}_{l=1}^{k_1}} \sum_{i=1}^{n} \min_{1 \le l \le k_1} \|\alpha_i - \mathbf{c}_l^{out}\|_\gamma^\gamma + \min_{\{\mathbf{c}_m^{in}\}_{m=1}^{k_2}} \sum_{j=1}^{n} \min_{1 \le m \le k_2} \|\beta_j - \mathbf{c}_m^{in}\|_\gamma^\gamma,$$

  which echoes the idea of K-means clustering
- Achieved consistencies in both network estimation and community detection, even when

$$\max_{1 \le i,j \le n} p_{ij} \le c_u e^{-T_n}$$

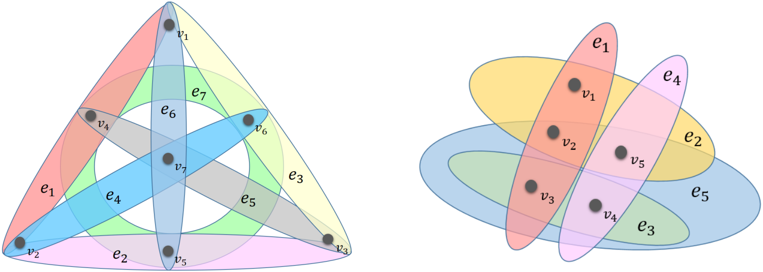
  with some constant c_u and $T_n \to \infty$

Another look at the coauthorship network



Remark: more than 2 co-authors in a paper, leading to hypergraph

Some toy hypergraphs

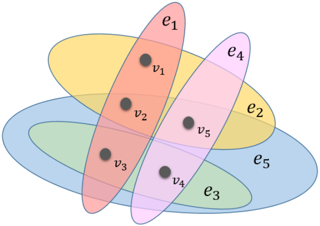


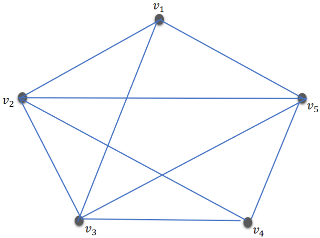
Left: Fano plane, a 3-uniform hypergraph; Right: a non-uniform hypergraph

General hypergraph

- In a hypergraph $\mathcal{H}(V, E)$, where $V = [n]$ is a vertex set and E consists of all hyperedges
- A hyperedge is a non-empty subset of $[n]$, and may contain multiple vertices in V
- A hypergraph is called *m*-uniform if the cardinality of every hyperedge equals m, denoted by an order-m tensor $\mathcal{A} = (a_{i_1 \ldots i_m}) \in \{0, 1\}^{n \times \ldots \times n}$
- A non-uniform hypergraph contains hyperedges whose cardinalities can vary from one to another

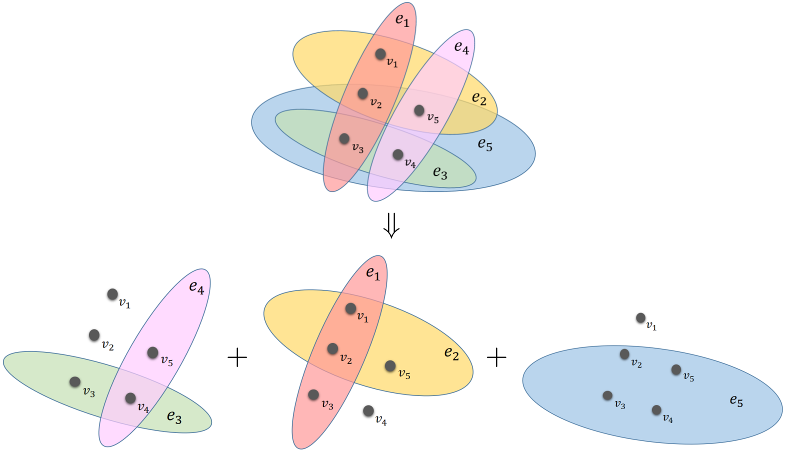
Projection approach (Ghoshdastidar and Dukkipati, 2017)





- Binary graph: two nodes are connected if they are both contained in at least one hyperedge
- Weighted graph: the weight is proportional to the number of hyperedges both nodes are contained in

Decomposition approach (Ke et al., 2021; Yuan et al., 2021)



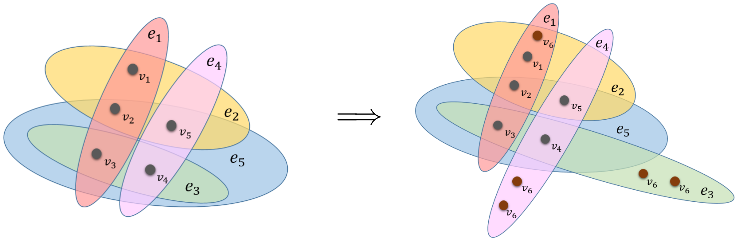
A non-uniform hypergraph is decomposed into three uniform hypergraphs

General hypergraph embedding (Zhen and W., 2021)

Given $\mathcal{H}(V, E)$ with $V = [n]$ and range $m \geq 2$, where range is the largest hyperedge cardinality

- Hypergraph augmentation
  - Introduce a null vertex v_{n+1}
  - Any hyperedge with cardinality less than m is augmented to a multi-set with m elements by adding one or more v_{n+1}
- Hypergraph embedding
  - Embeds all vertices in $\mathcal{H}$ into a low-dim Euclidean space
  - Vertices belonged to the same community tend to have shorter distance

Hypergraph augmentation



- Each augmented hyperedge has 4 vertices, some with multiple null vertices v_6
- It becomes a uniform multi-hypergraph, but only v_6 can appear more than once

Hypergraph augmentation

- With the introduced null vertex, $\mathcal{H}$ is transformed into an m-uniform multi-hypergraph
- Its adjacency tensor $\mathcal{A} = (a_{i_1 \ldots i_m})$ with entries

$$a_{i_1 \ldots i_m} = \begin{cases} 1, & \text{if } \{\{i_1, \ldots, i_m\}\} \setminus \{n+1\} \in E; \\ 0, & \text{otherwise} \end{cases}$$

- Let $\mathcal{P} = (p_{i_1 \ldots i_m})$ with $p_{i_1 \ldots i_m} = P(a_{i_1 \ldots i_m} = 1)$, and

$$\theta_{i_1 \ldots i_m} = \log\left(\frac{p_{i_1 \ldots i_m}}{s_n - p_{i_1 \ldots i_m}}\right),$$

where s_n captures network sparsity, and $0 \leq p_{i_1 \ldots i_m} \leq s_n$

Hypergraph embedding

- We consider a hypergraph embedding model (HEM),

$$\theta_{i_1\ldots i_m} = \log\left(\frac{p_{i_1\ldots i_m}}{s_n - p_{i_1\ldots i_m}}\right) = \mathcal{I} \times_1 \boldsymbol{\alpha}_{i_1}^T \times_2 \ldots \times_m \boldsymbol{\alpha}_{i_m}^T$$

  - $\mathcal{I}$ is the m-th order identity tensor of dimension r
  - $\boldsymbol{\alpha}_i$ is an r-dimensional embedding vector of vertex i
  - $\boldsymbol{\alpha}_{n+1} = \mathbf{1}_r$ is the embedding of the null vertex v_{n+1}
- It has an equivalent CP decomposition,

$$\Theta = \mathcal{I} \times_1 \boldsymbol{\alpha} \times_2 \ldots \times_m \boldsymbol{\alpha} = \sum_{j=1}^{r} \boldsymbol{\alpha}_{.j} \circ \ldots \circ \boldsymbol{\alpha}_{.j},$$

  where $\boldsymbol{\alpha}_{.j}$ is the j-th column of $\boldsymbol{\alpha}$, and r can be understood as the symmetric rank of Θ.

A regularization formulation

- With HEM, the negative log-likelihood function of $\mathcal{H}$ becomes

$$\mathcal{L}(\Theta; \mathcal{A}) = \frac{1}{\varphi(n,m)} \sum_{\delta_{i_1 \dots i_m}^{n+1,ord}=0} L(\theta_{i_1 \dots i_m}; a_{i_1 \dots i_m}),$$

where $\varphi(n,m) = \sum_{k=1}^{m} \binom{n}{k}$ is # of potential hyperedges, and

$$L(\theta; a) = \log\left(1 + \frac{s_n}{1 - s_n + e^{-\theta}}\right) - a \log\left(\frac{s_n}{1 - s_n + e^{-\theta}}\right)$$

- The proposed regularized formulation is

$$\mathcal{L}_\lambda(\boldsymbol{\alpha}; \mathcal{A}) = \mathcal{L}(\Theta; \mathcal{A}) + \lambda_n \min_{Z,C} \frac{1}{n} ||\boldsymbol{\alpha} - ZC||_F^2,$$

Community detection consistency

Theorem 1

Suppose all the assumptions in Theorem 1 and Assumptions A and B are satisfied. If $\lim_{n\to+\infty} \lambda_n \epsilon_n s_n^{-2} (\log s_n^{-1})^{-1} > 0$ *and* $K = o(\gamma_n (s_n \epsilon_n^{-1})^{1/2})$*, then*

$$err(\psi^*, \hat{\psi}) = O_p(K \epsilon_n s_n^{-1}).$$

Remark: Theorem 2 holds true as long as $s_n \gg \epsilon_n \gg n^{1-m} \log n$, matching up the best sparsity result in literature.

MeSH hypergraph networks

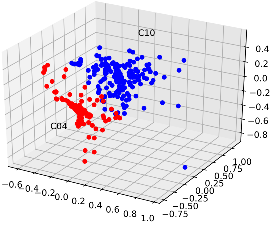
- Medical Subject Headings (MeSH) hypergraph network (Ke et al., 2021) consists of 318 MeSH terms of two diseases: Neoplasms (C04) and Never System Diseases (C10)
- Each vertex represents a Mesh term and a hyperedge is a research paper where one or more of the above MeSH terms are annotated
- After pre-processing, we obtain a hypergraph network with 281 vertices and 1057 hyperedges

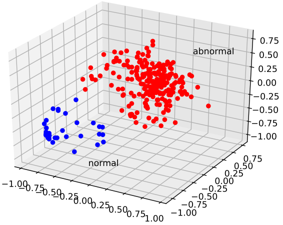
SPECT hypergraph networks

- The cardiac Single Proton Emission Computed Tomography (SPECT) hypergraph network (Dua and Graff, 2017) consists of SPECT images of 267 patients, where each image has been processed to 44 categorical features to discriminate abnormal patients from the normal ones
- Each vertex represents a patient and a hyperedge contain all the patients sharing a particular set of feature values
- After pre-processing, we obtain a hypergraph network with 264 vertices and 2950 hyperedges

Real examples

	HEM	Tensor-SCORE	SHP	WPTG
MeSH	**0.0427**	0.0819	0.0498	0.3630
SPECT	**0.1931**	0.3409	0.3181	0.2652





Take-home messages

- LFM provides a general framework for modeling dyadic and multi-adic relational data
- Identification of latent factors improves prediction accuracy and model interpretability

- Other structures beyond low-rank?
- Some ongoing work on multi-layer biological networks, dynamic social networks ...

Take-home messages

- LFM provides a general framework for modeling dyadic and multi-adic relational data
- Identification of latent factors improves prediction accuracy and model interpretability

- Other structures beyond low-rank?
- Some ongoing work on multi-layer biological networks, dynamic social networks ...

Thank you!