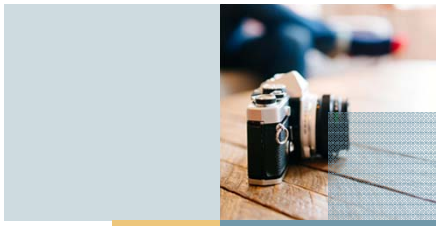
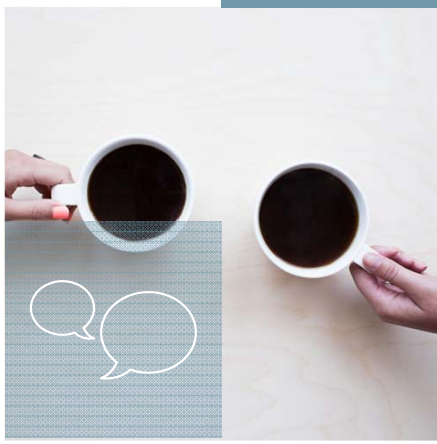




Semantic Data Mining Using RDF Hypergraphs

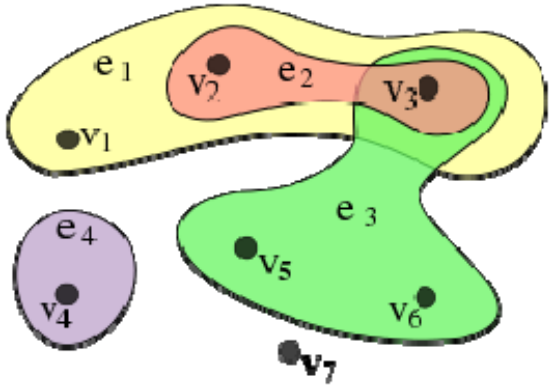
Wei Yu
Han Xiao Liu

5/27/2016

The definition of hypergraph

(1)	(2)
• In a graph, an edge, only two vertices • In a hyper-graph, an edge, more than two vertices.	• A hypergraph HG=(V,E) • V : vertex set. • V : vertex set. • e ∈ E is a subset of V.

An example of hypergraph



An example of a hypergraph

$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$

$E = \{e_1, e_2, e_3, e_4\}$

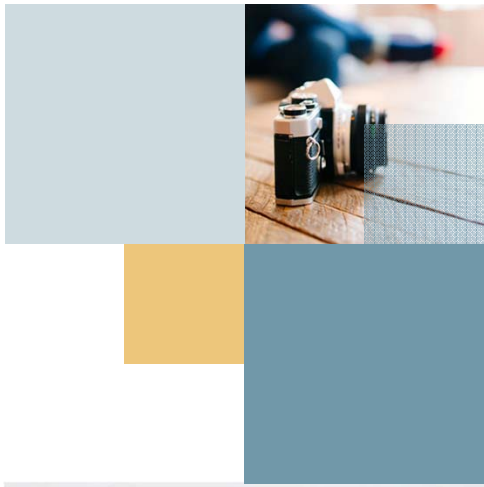
$= \{\{v_1, v_2, v_3\}, \{v_2, v_3\}, \{v_3, v_5, v_6\}, \{v_4\}\}$.

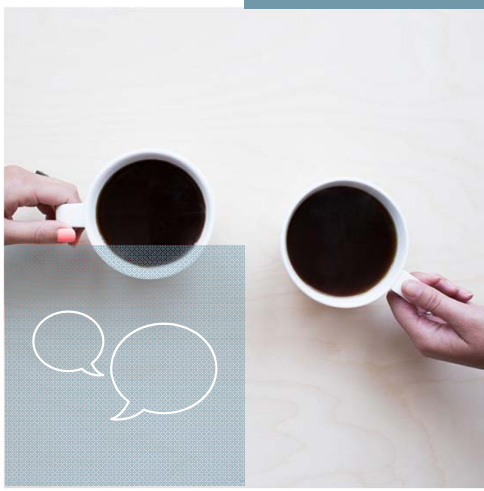
The application of hypergraph in data mining

Discovering semantic associations between items

Clustering

Spatial data mining

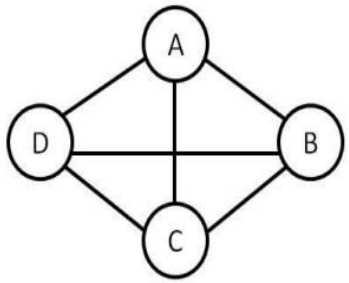




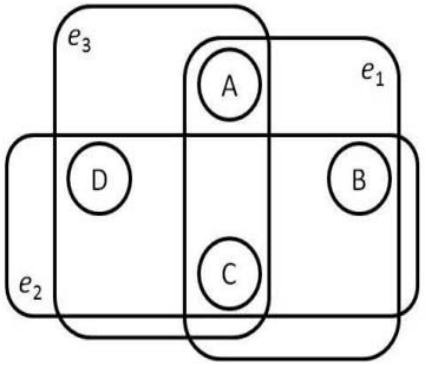
Why we use hypergraph

	A	B	C	D
e_1:	1	1	1	0
e_2:	0	1	1	1
e_3:	1	0	1	1

(a)



(b)

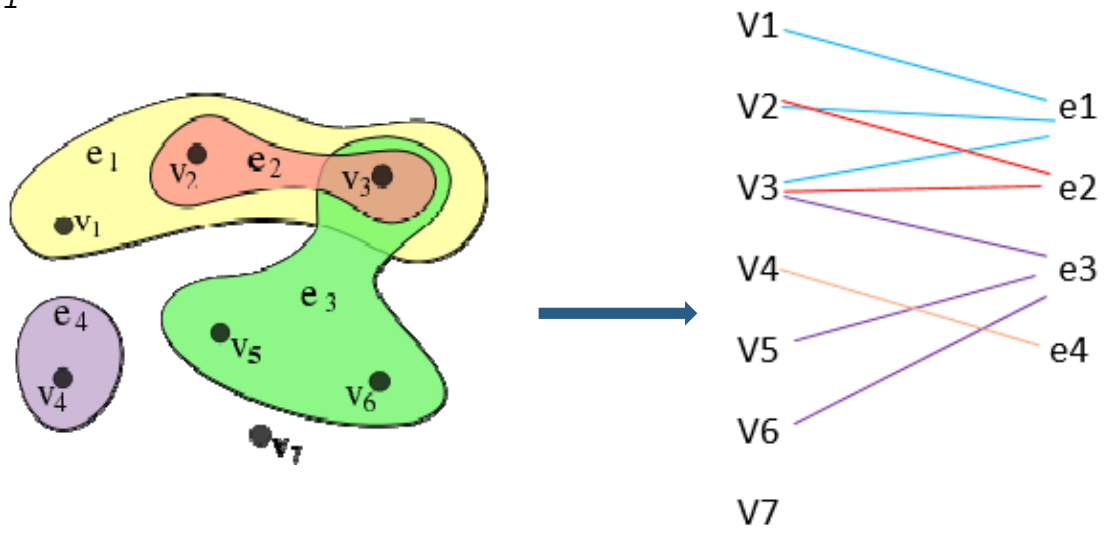


(c)

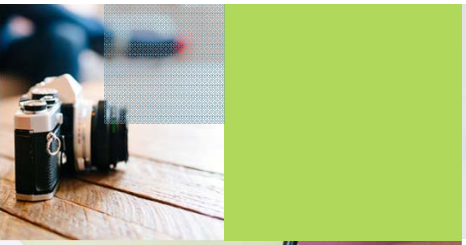
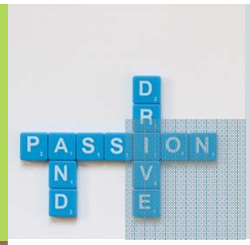
Figure 1. (a) an example transaction table; (b) the Gaifman graph representation of the table; (c) The hypergraph representation of the table

How to transform hypergraph to bipartite graph(BG)

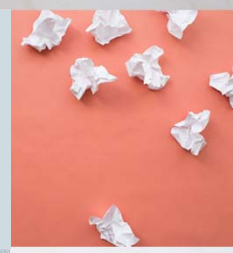
- Let V and E be the two parts of BG
- (v_1, e_1) are connected with an edge if and only if vertex v_1 is contained in edge e_1 in H



Hypergraph BG







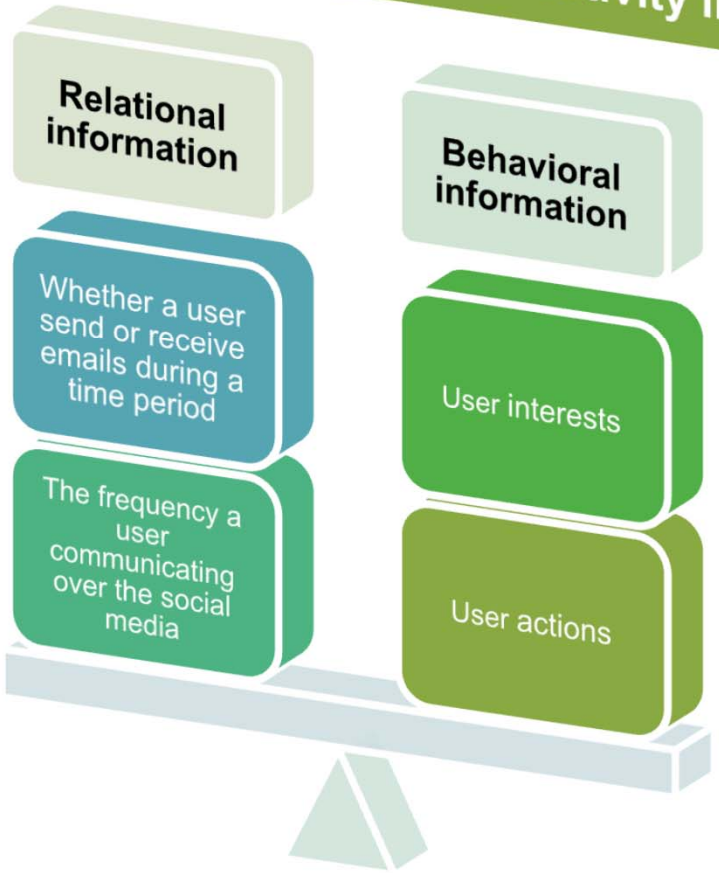


The Application In Dynamic Social Networks of Multi-Layer Graph

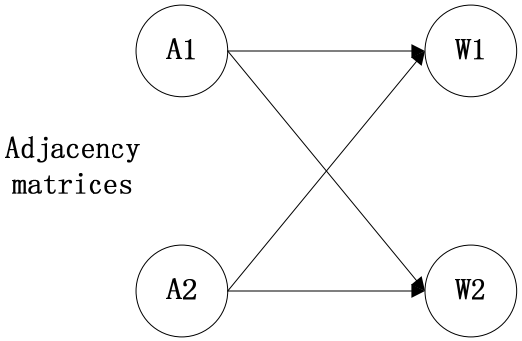
We are married

JD.COM 京东

Two type of sources in the connectivity information

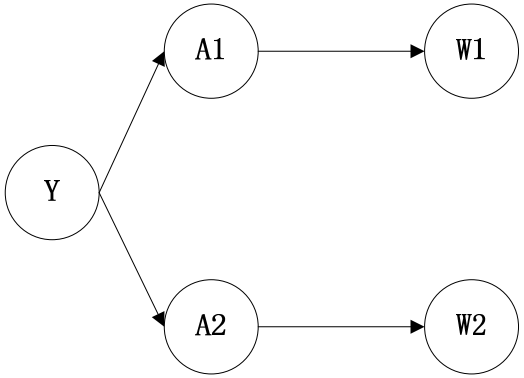


The evolution of solutions of the multi-layer graph



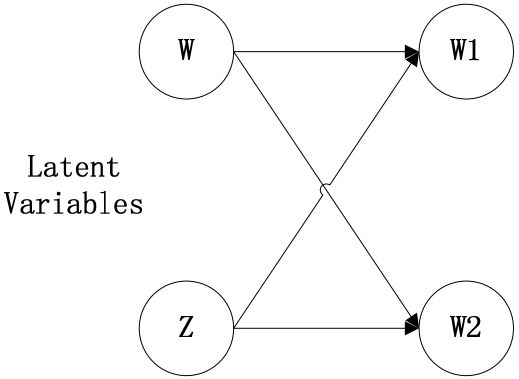
Adjacency matrices

Observed matrices



Latent Variables

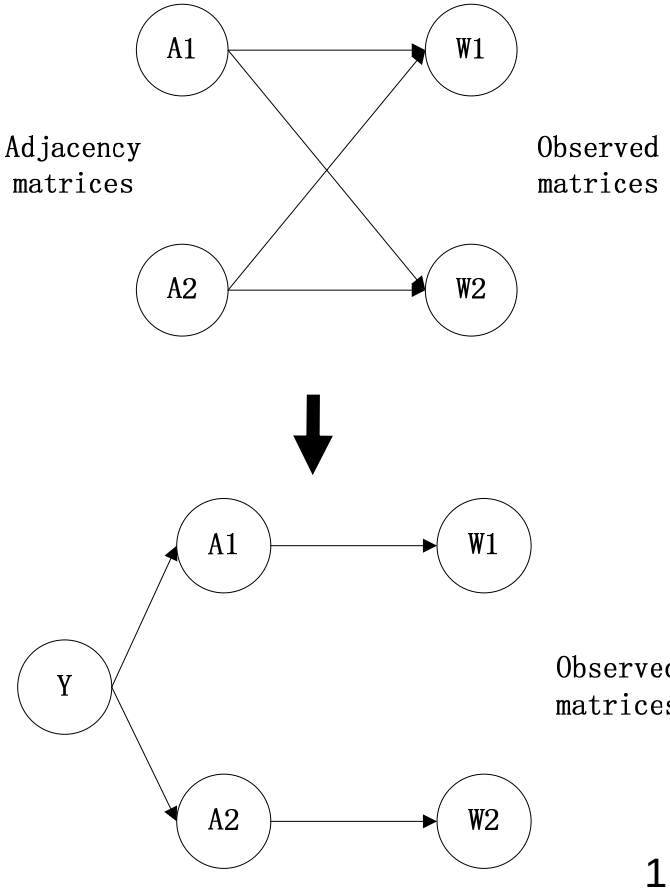
Observed matrices



Latent Variables

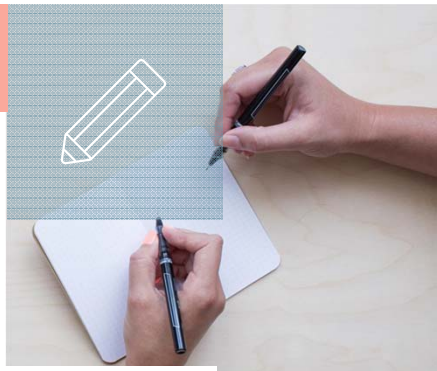
Observed matrices

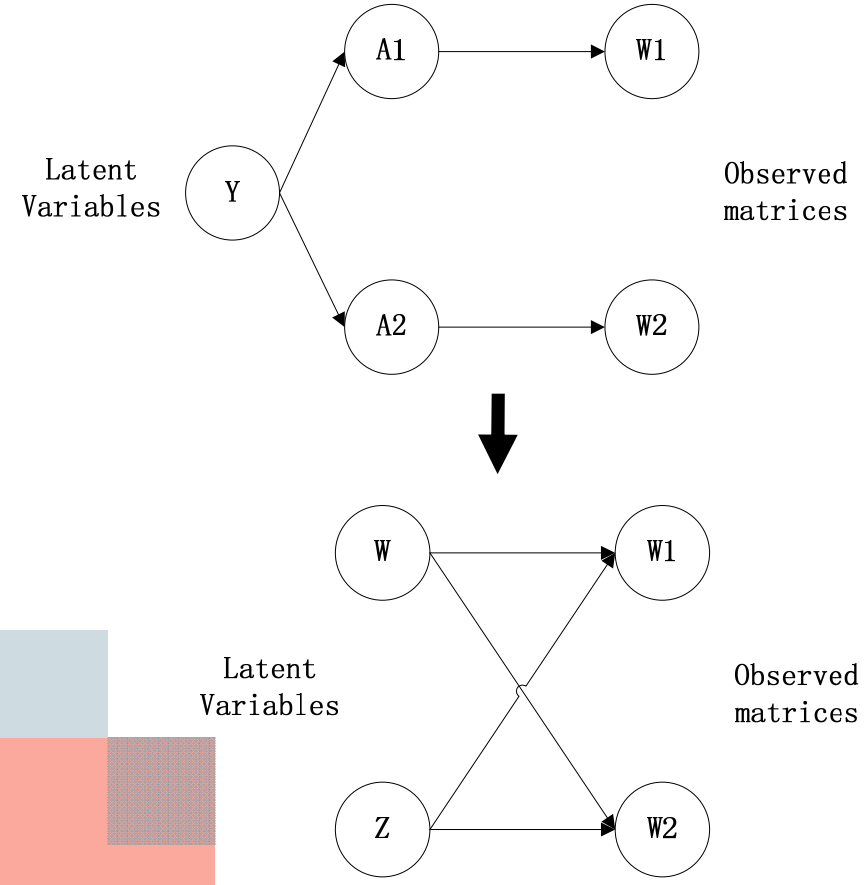
Hierarchical Latent-variable Model



$$P(Y|W_1, W_2) = \sum_{A_1, A_2} P(Y|A_1, A_2)P(A_1, A_2|W_1, W_2).$$

Posterior Mixture Model





$$P(Y|W_1, W_2) = \sum_{A_1, A_2} P(Y|A_1, A_2) P(A_1, A_2|W_1, W_2).$$

$$P(W|W_1, W_2) \tag{12}$$

$$= \xi \frac{P(W)P_1(W_1|W)}{P_1(W_1)} + (1-\xi)\frac{P(W)P_2(W_2|W)}{P_2(W_2)} \tag{13}$$

$$= P(W)\left[\gamma_1 P_1(W_1|W) + \gamma_2 P_2(W_2|W)\right], \tag{14}$$

$$\arg\max_W \left[\gamma_1 P_1(W_1|W) + \gamma_2 P_2(W_2|W)\right]. \tag{15}$$

References

[1] Morales, A., & Serodio, M. (2007, June). A directed hypergraph model for RDF. In Proc. of Knowledge Web PhD Symposium.

[2] Hayes, J., & Gutierrez, C. (2004). Bipartite graphs as intermediate model for RDF. In The Semantic Web–ISWC 2004 (pp. 47-61). Springer Berlin Heidelberg.

[3] Ausiello, G., & Laura, L. (2016). Directed hypergraphs: Introduction and fundamental algorithms—A survey. Theoretical Computer Science.

[4]Nettleton, D. F. (2013). Data mining of social networks represented as graphs. Computer Science Review, 7, 1-34.

[5] L.P. Cordella, P. Foggia, C. Sansone, M. Vento, An Improved Algorithm for Matching Large Graphs, in: Proc. 3rd IAPRTC-15 International Workshop on Graph based Representations, Cuen, Italy, 2001, pp. 149–159.

[6] V.D. Blondel, J.L. Guillaume, R. Lambiotte, E. Lefebure, Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment (10) (2008) 1000.

References

[7] Oselio, B., Kulesza, A., & Hero, A. O. (2014). Multi-layer graph analysis for dynamic social networks. Selected Topics in Signal Processing, IEEE Journal of, 8(4), 514-523. [8] A. Raftery, "Bayesian model selection in social research," Sociol. Methodol., vol. 25, pp. 111–164, 1995.

[9] K. S. Xu and A. O. H. III, "Dynamic stochastic blockmodels: Statistical models for time-evolving networks," CoRR, vol. abs/1304.5974, 2013.

[10] Liu, H. (2012). A graph-based approach for semantic data mining.

[11] Yang, S., Yan, X., Zong, B., & Khan, A. (2012, May). Towards effective partition management for large graphs. In Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data (pp. 517-528). ACM.

[12] Liu, H., Dou, D., Jin, R., LePendu, P., & Shah, N. (2013, December). Mining biomedical ontologies and data using rdf hypergraphs. In Machine Learning and Applications (ICMLA), 2013 12th International Conference on (Vol. 1, pp. 141-146). IEEE.

Thank You