Learning with Hypergraphs

B. Ravindran

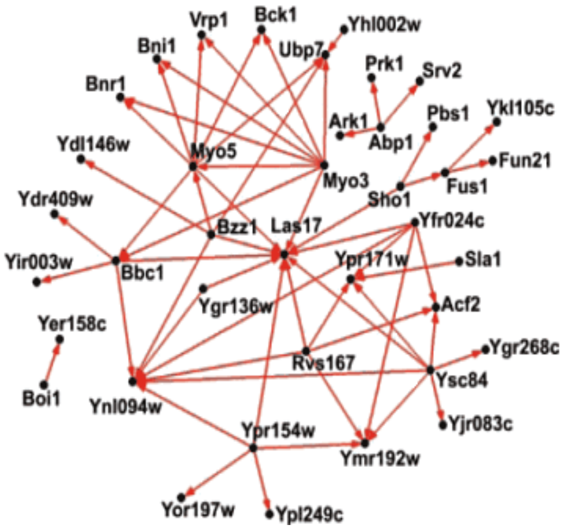
Joint Work with Sai Nageshwar S.

Reconfigurable and Intelligent Systems Engineering (RISE) Group

Department of Computer Science and Engineering

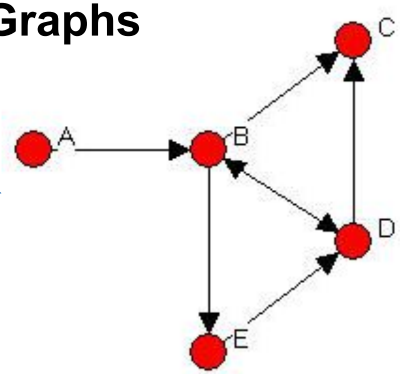
Indian Institute of Technology Madras

We live in a connected World!

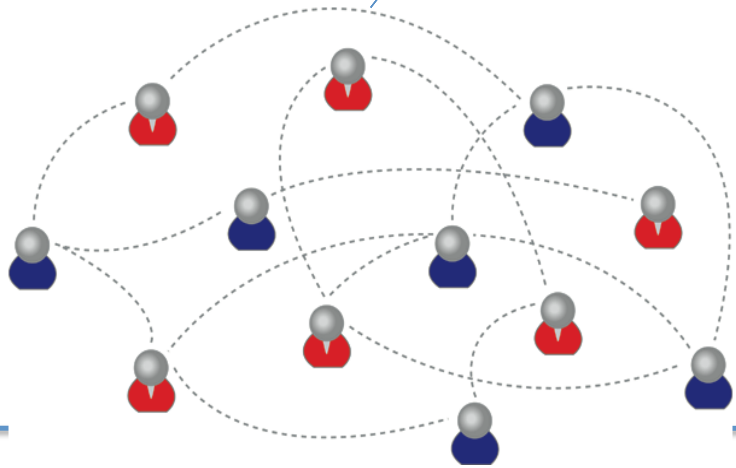


Protein-Protein Interaction Network

Represented As Graphs



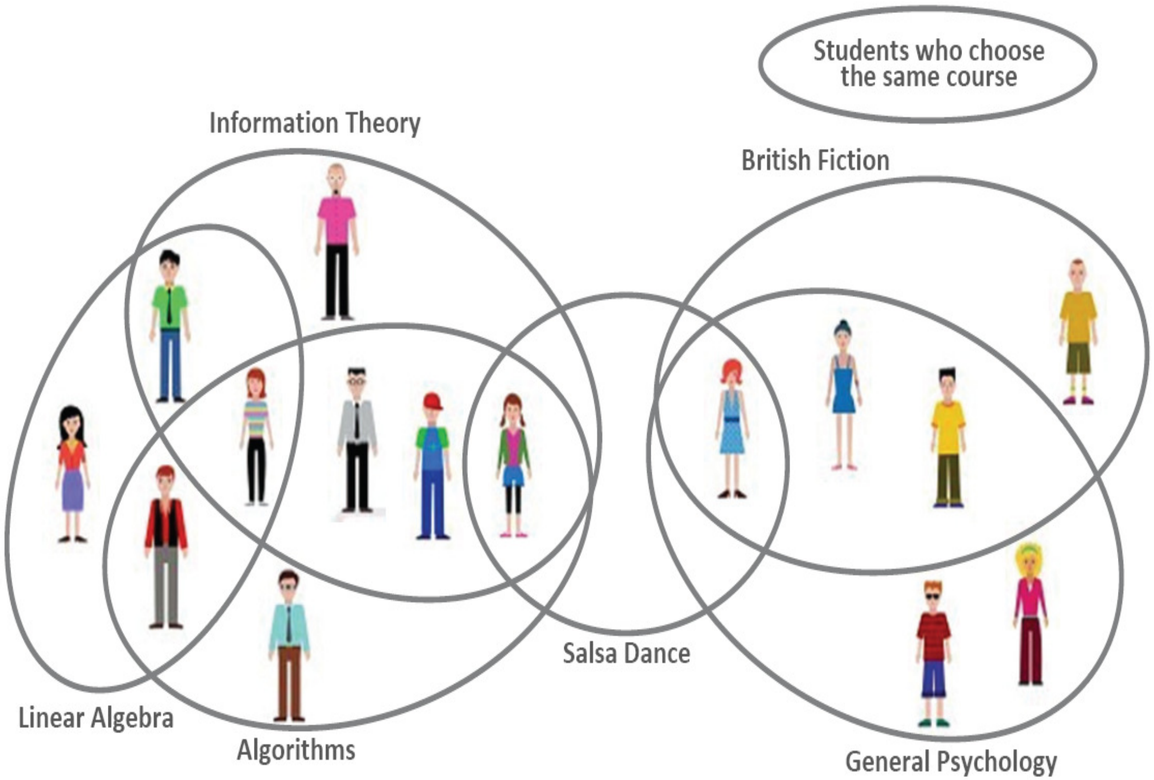
World Wide Web



Social Networks

Super-Dyadic Relations

Relationships might involve multiple entities.

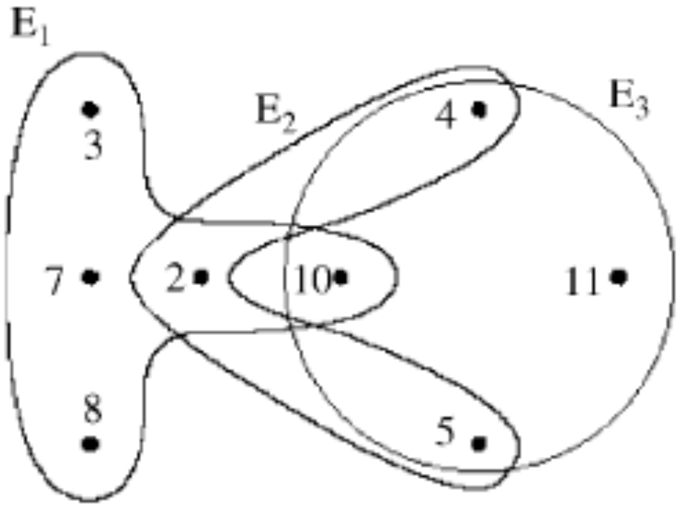


Other examples

- Publication data
  - Co-authorship
  - Co-citation
- Collaborations
  - Committee memberships
  - Movies
- Chemical processes
- Social interactions

Hypergraphs

Hypergraphs are generalization of graphs in which an edge can connect any number of vertices



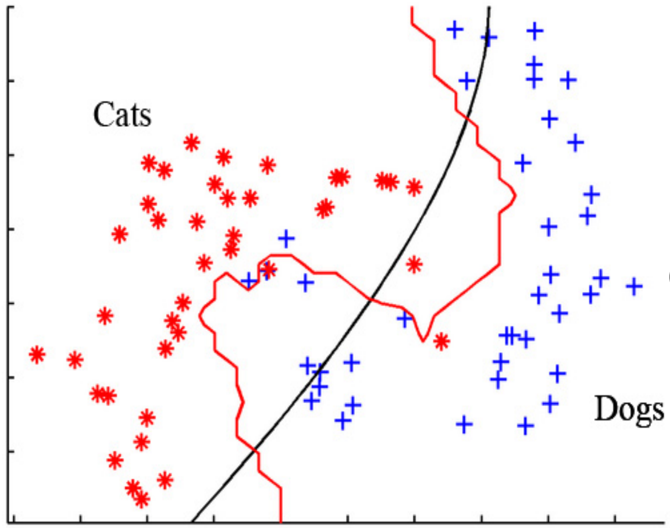
H = (V, E)

V = {2, 3, 4, 5, 7, 8, 10, 11}

E = {E_1, E_2, E_3}
= { {3, 7, 8, 2, 10}, {2, 4, 5}, {4, 5, 10, 11} }

Learning With Hypergraphs

Segmentation



Classification

Segmentation

- Based on Hypergraph Distance
  - Create normal graph by connecting nodes with weighted edges
  - Weight is the number of hyperedges traversed in the shortest path
- Find clusters in the normal graph
  - Maximize modularity Newman '06

Modularity

$$Q = \frac{1}{4m} \sum_{\substack{i,j \\ \textit{in same} \\ \textit{module}}} \left(A_{ij} - \frac{k_i k_j}{2m} \right)$$

modularity

normalization

adjacency matrix

probability a random edge would go between i and j

m = # edges in graph
k_i = degree(i)

Lectures on Bio-informatics
Carl Kingsford, CMU

Indian Railways Network

- Each Station is a node in the graph
- Each train is considered as an hyperedge!
  - Spans all the stations that the train visits
- Remove edges that are subsets of other edges
  - Eliminates express trains like Rajdhani, Shathabdi, etc.
  - Alternative explored: Assign weights based on "speed"

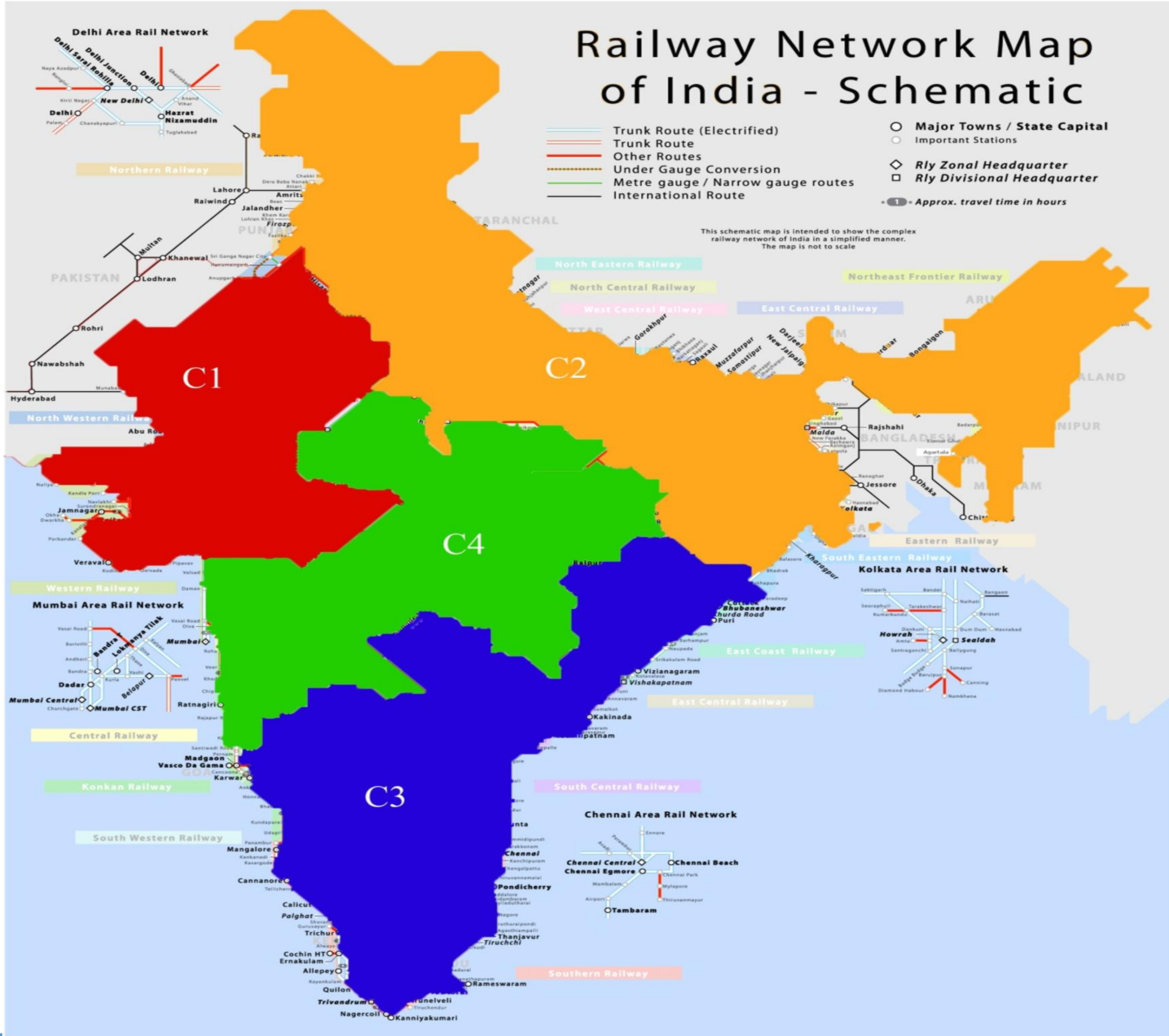
Transport Network Model

- Link two stations if there is a track between them.
- Most stations are of degree two
  - Except major junctions
- Inferred from the time tables

Hypergraph Model vs Normal Graph Model

Property	Base Graph Model	Hypergraph Model
Connected Components	2	2
Diameter	247	5
Communities	**75** latent Communities with 0.956 modularity	**5** latent communities mapping with Indian Adminstrative zones

Communities and Zones



Future Work

- Weighted Analysis of Indian Railways Network
- Network properties for hypergraph relations
  - Centrality
  - Conductance, etc.
- …relevance to network analysis

Collective Classification with Hypergraphs

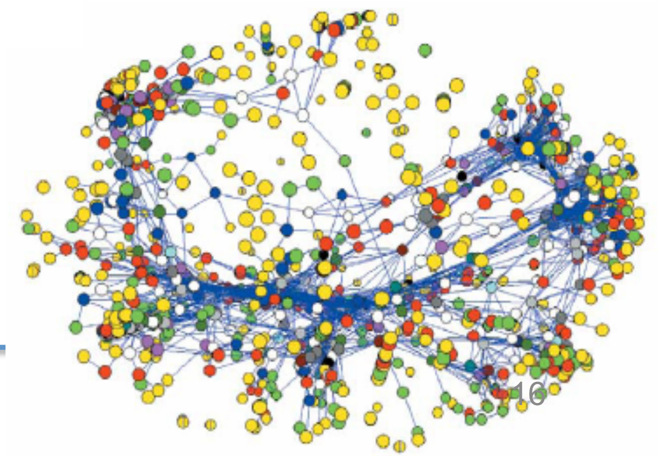
Machine Learning

- Traditional Machine Learning
  - Use node attributes (content) alone

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Collective Approach

- Traditional Machine Learning
  - Use node attributes (content) alone
- Collective approaches
  - Use content and link information
- Label of a node depends on labels of neighbors on the link structure (graph)
  - *How many friends bought computers?*
  - Web page text and hyperlinks
  - Content of papers and citations
  - Tweet text and followers



Collective Classification

- Typically work with aggregate label statistics
- Build classifiers on networked data using label distribution in neighborhood as well
  - *Collective learning*
- Predict labels on partially labeled test data
  - *Collective inference*
  - Each network is partially labeled
- Eg.: Iterative Classification Algorithm; Graph Cuts; etc.

A Naive Approach

- Train a classifier on known instances and then infer labels of unknowns
- Use relation for smoothing
  - Repeated local voting
  - A form of relaxation labeling

Graph Cut Approach

- For each class, add a node in the graph
- Connect all instances to the class nodes with weights equal to probability of the class
  - Given by classifier trained on known instances
  - For known instances, weight will be 1.
- Partition the graph for *smoothing*
  - Multi-way Cuts

Multi Way Relations In Collective Learning

- Various problem exhibit super dyadic relationships
- Examples :
  - In classifying research papers, besides words in the papers co-citation relationship can also be used for classification. A paper cites many papers and a co-citation graph involves a multi-way relationship.
  - In Social Networks, for clustering/classifying users besides posts, group memberships can be used for learning. These groups are multi-way relationships

Transductive Inference

- Transductive inference is a semi-supervised learning approach
  - Both known and unknown instances are used for inferring class labels
  - Answer questions about specific unknown instances
  - Induction: infer general rules
- Our hypergraph classification approach is a form of transductive inference
  - Repeat the inference process assigning labels to unknown instances until convergence

Classification With Hypergraphs

Steps :

- Using a random walk operator on a hypergraph, obtain node to node transition probability, P.
- Find stationary distribution, Π, of the random walk operator on the hypergraph
- Let $y(u)$ be label of u then for an unknown node v, label can be inferred as

$$y(v) = \sum_{u \in V} \pi(u) p(u, v) y(u)$$

Weights in the Hypergraph

- Weight of an edge is the *purity* of the edge
  - Fraction of labeled nodes in the edge that belong to the majority class
- Class imbalance: the fraction of data points belonging to the critical class is very small
  - Medical data
  - Dissatisfied customers
- Problem: Critical class edges will not have significant weights

Handling class imbalance

- Hyperedges are weighted to represent relative richness of a class
- Hellinger distance for a 2 class problem (positive and negative) is defined as :

$$W(e) = \left(\sqrt{\frac{|y_e^+|}{|y^+|}} - \sqrt{\frac{|y_e^-|}{|y^-|}} \right)^2$$

where, y_e^+ is set of positive instances in hyperedge e

y^+ is set of positive instances in training set

y_e^- is set of negative instances in hyperedge e

y^- is set of negative instances in training set

Random Walk on Hypergraph

- Prob. of transition between two nodes u and v is given as product of:
  - Prob. of taking edges e' containing u and v:

$$p1 = \frac{w(e')}{\sum_{v \in e} w(e)}$$

  - Prob. of choosing v, once e' is chosen:

$$p2 = \frac{1}{deg(e') - 1}$$

$$p(u,v) = \sum_{u,v \subseteq e'} p1 * p2$$

Results – Single Relation

		Macro F1		
Dataset	Notes	hmetis	Hypergraph Expansion	Hypergraph MRMVCC
Cora (link only)	Citation links of 2708 research papers	0.7061	0.7435	0.7720
20 Newsgroup	Categorization of 1642 documents	0.4260	0.639	0.744
WebKB (link only)	Classification of 195 webpages from Cornell University using only hyperlinks	0.3290	0.474	0.544

Multi Relation Multi View Collective Classification

- A *view* is a description of the data
  - Words appearing in a document
  - Pixels in an image
- Some data naturally have multiple views
  - Image pixels and tags
  - Webpage text and incoming hyperlink text
- Convert each view into a hypergraph relation as follows :
  - For each attribute: For each unique value of attribute connect all instances which have same value for the attribute by an hyperedge .
- Both relations and views can be modeled using hypergraphs now!

Attribute hypergraph construction

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

- 14 nodes
- 10 hyperedges
  - 3 each for age and income
  - 2 each for student and credit_rating
- Example:
  - income-high={1,2,3,13};
  - income-medium ={4,8,10,11,12,14};
  - income-low={5,6,7,9}

Random Walk with Multiple Hypergraphs

- Calculate probability transition matrix, Π_i for each hypergraph
  - One corresponding to each relation and view
- Define probability of transition to each graph as α_i
- Stationary distribution of resultant multi graph random walk will be given by

$$\Pi = \sum_i \alpha_i \Pi_i$$

- Resultant probability transition matrix is given by,

$$P = \sum_i \Pi^{-1} \left(\alpha_i \Pi_i \right) * P_i$$

- Perform transductive inference using resultant stationary distribution and probability transition matrix
  - The α_i are determined empirically using cross validation

Results – Single View

WebKB: A dataset of 877 Webpages gathered from four different universities contains keywords and has five classes namely course, faculty, student, project and staff.

	Macro - F1 Score	
Num of Unknowns	**Decision Tree**	**Hypergraph CC**
352(40%)	0.5490	0.6364
527(60%)	0.528	0.6202
616(70%)	0.510	0.566

Results – Single view, single relation

Cora: A dataset of 2708 research papers classified into 7 classes. The dataset consists of keywords and citation links among them.

	Macro F1		
Num Of Unknowns	**Iterative Classification Algorithm**	**Hypergraph MRMVCC**	**Content Only**
677 (25%)	0.8178	0.8277	0.6239
1354 (50%)	0.8164	0.8146	0.6086
2031 (75%)	0.7848	0.7871	0.5609

Results – Single View, Single relation

Citeseer: A dataset of 3312 research papers classified into 6 classes. The dataset consists of keywords and citation links.

	Macro F1		
Num Of Unknowns	**Iterative Classification Algorithm**	**Hypergraph MRMVCC**	**Content Only**
663 (20%)	0.7265	0.7327	0.7060
994 (30%)	0.7264	0.7243	0.7014
1656 (50%)	0.7054	0.7137	0.6837
2319 (70%)	0.6731	0.6874	0.6482
2816 (85%)	0.624	0.614	0.597

Results – Multiple Views, Multiple relations

Twitter-Olympics: A dataset of 464 athletes and organizations that were involved in the London 2012 Summer Olympics. We have considered 8 views: tweets, lists, follows, followed-by, mentions, mentioned-by, retweeted-by, retweets. The first two views, i.e., tweets and lists are constructed as hypergraphs and rest are constructed as directed graphs.

	Macro F1	
Num Of Unknowns	**Collective Ensemble**	**Hypergraph MRMVCC**
186 (40%)	0.7366	0.9132
232 (50%)	0.6388	0.8425
279 (60%)	0.558	0.7718
325 (70%)	0.4525	0.625

Conclusion

- Hypergraphs offer a convenient mechanism to model super dyadic relations
- They can also represent multi-view multi-relational data effectively
- Many interesting questions:
  - What is the interpretation of the combined random walk operator on the set of hypergraphs?
  - Any insight from the spectra of the graphs?
  - A more robust optimization procedure for determining the α's
  - Beyond transductive inference: Learn general classifiers

Questions?

http://www.cse.iitm.ac.in/~ravi