A Tensor Spectral Approach to Learning Mixed Membership Community Models

Anima Anandkumar

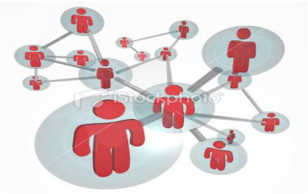
U.C. Irvine

Joint work with Rong Ge, Daniel Hsu and Sham Kakade.

Community Models in Social Networks

Social Network Modeling

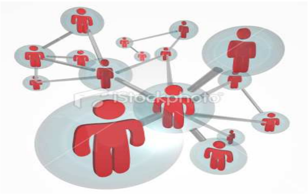
- Community: group of individuals
- Community formation models: how people form communities and networks
- Community detection: Discovering hidden communities from observed network



Community Models in Social Networks

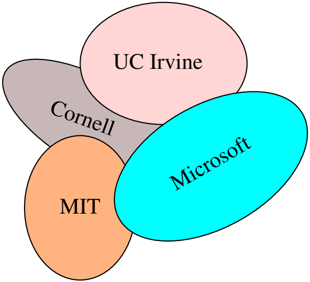
Social Network Modeling

- Community: group of individuals
- Community formation models: how people form communities and networks
- Community detection: Discovering hidden communities from observed network



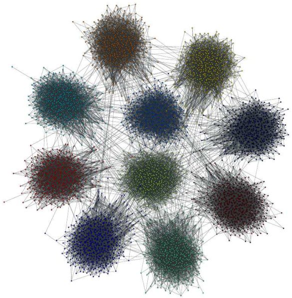
Modeling Overlapping Communities

- People belong to multiple communities
- Challenging to model and learn such overlapping communities



Stochastic Block Model: Classical Approach

Generative Model



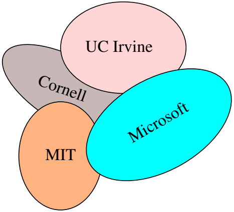
- k communities and network size n
- Each node belongs to one community: $\boxed{\pi_u = e_i}$ if node u is in community i.
- e_i is the basis vector in i^{th} coordinate.
- Probability of an edge from u to v is $\boxed{\pi_u^\top P \pi_v}$.
- Notice that $\pi_u^\top P \pi_v = P_{i,j}$ if $\pi_u = e_i$ and $\pi_v = e_j$.
- Independent Bernoulli draws for edges.

- Pros: Guaranteed algorithms for learning block models, e.g. spectral clustering, d_2 distance based thresholding
- Cons: Too simplistic. Cannot handle individuals in multiple communities

Mixed Membership Block Model

Generative Model

- k communities and network size n
- Nodes in multiple communities: for node u, π_u is community membership vector
- Probability of an edge from u to v is $\boxed{\pi_u^\top P \pi_v}$, where P is block connectivity matrix
- Independent Bernoulli draws for edges



Dirichlet Priors

- Each π_u drawn independently from $\mathrm{Dir}(\alpha)$: $\boxed{\mathbb{P}[\pi_u] \propto \prod_{j=1}^{k} \pi_u(j)^{\alpha_j - 1}}$
- Stochastic block model: special case when $\alpha_j \to 0$.
- Sparse regime: $\alpha_j < 1$ for $j \in [k]$.

Airoldi, Blei, Fienberg, and Xing. Mixed membership stochastic blockmodels. J. of Machine Learning Research, June 2008.

Learning Mixed Membership Models

Advantages

- Mixed membership models incorporate overlapping communities
- Stochastic block model is a special case
- Model sparse community membership

Challenges in Learning Mixed Membership Models

- Not clear if guaranteed learning can be provided.
- Potentially large sample and computational complexities
- Identifiability: when can parameters be estimated?

Learning Mixed Membership Models

Advantages

- Mixed membership models incorporate overlapping communities
- Stochastic block model is a special case
- Model sparse community membership

Challenges in Learning Mixed Membership Models

- Not clear if guaranteed learning can be provided.
- Potentially large sample and computational complexities
- Identifiability: when can parameters be estimated?

Solution: Method of Moments Approach

Method of Moments

- Inverse moment method: solve equations relating parameters to observed moments
- Spectral approach: reduce equation solving to computing the "spectrum" of the observed moments
- Non-convex but computationally tractable approaches

Method of Moments

- Inverse moment method: solve equations relating parameters to observed moments
- Spectral approach: reduce equation solving to computing the "spectrum" of the observed moments
- Non-convex but computationally tractable approaches

Spectral Approach to Learning Mixed Membership Models

- Edge and Subgraph Counts: Moments in a network
- Tensor Spectral Approach: Low rank tensor form and efficient decomposition methods

Summary of Results and Technical Approach

Contributions

- First guaranteed learning algorithm for overlapping community models
- Correctness under exact moments.
- Explicit sample complexity bounds.
- Results are tight for Stochastic Block Models

Summary of Results and Technical Approach

Contributions

- First guaranteed learning algorithm for overlapping community models
- Correctness under exact moments.
- Explicit sample complexity bounds.
- Results are tight for Stochastic Block Models

Approach

- Method of moments: edge counts and 3-star count tensors
- Tensor decomposition: Obtain spectral decomposition of the tensor
- Tensor spectral clustering: Project nodes on the obtained eigenvectors and cluster.

Related Work

Stochastic Block Models

- Classical approach to modeling communities (White et. al ‘76, Fienberg et. al 85)
- Spectral clustering algorithm (McSherry ‘01, Dasgupta ‘04)
- d_2-distance based clustering (Frieze and Kannan ‘98)

  weak regularity lemma: any dense convergent graph can be fitted to a block model

Random graph models based on subgraph counts

- Exponential random graph models
- NP-hard in general to learn and infer these models

Overlapping community models

Many empirical works but no guaranteed learning

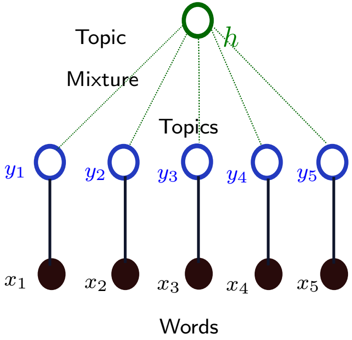
Outline

Outline

Connection to LDA Topic Models

Exchangeable Topic Models

- l words in a document $x_1, \ldots, x_l$.
- Document: topic mixture (draw of h).
- Word x_i generated from topic y_i.
- Exchangeability: $\boxed{x_1 \perp\!\!\!\perp x_2 \perp\!\!\!\perp \ldots | h}$
- LDA: $\boxed{h \sim \mathrm{Dir}(\alpha)}$.
- Learning from bigrams and trigrams

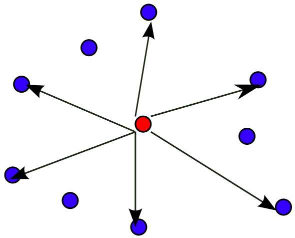


A. Anandkumar, R. Ge, D. Hsu, S.M. Kakade and M. Telgarsky "Tensor Decompositions for Learning Latent Variable Models," Preprint, October 2012.

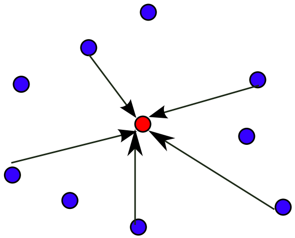
Viewing Community Models as Topic Models

- Analogy for community model: each person can function both as a document and a word.
- Outgoing links from a node u: node u is a document.
- Incoming links to a node v: node v is a word.

Node as a document



Node as a word



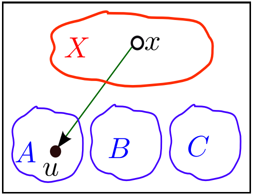
Outline

Moments for Spectral Method

- Subgraph counts as moments of a random graph distribution

Edge Count Matrix

- Consider partition X, A, B, C.
- Adjacency Submatrices $G_{X,A}$, $G_{X,B}$, $G_{X,C}$

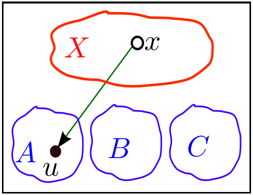


Moments for Spectral Method

- Subgraph counts as moments of a random graph distribution

Edge Count Matrix

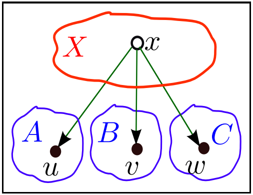
- Consider partition X, A, B, C.
- Adjacency Submatrices $G_{X,A}$, $G_{X,B}$, $G_{X,C}$



3-Star Count Tensor

- $\#$ of 3-star subgraphs from X to A, B, C.

$$\boxed{M_3(u, v, w) := \frac{1}{|X|}\#\text{ of 3-stars with leaves u,v,w}}$$



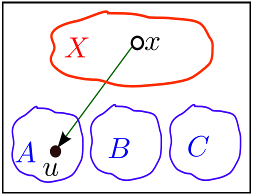
- Nodes in A, B, C: words and X: documents.

Moments for Spectral Method

- Subgraph counts as moments of a random graph distribution

Edge Count Matrix

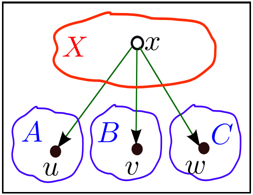
- Consider partition X, A, B, C.
- Adjacency Submatrices $G_{X,A}$, $G_{X,B}$, $G_{X,C}$



3-Star Count Tensor

- $\#$ of 3-star subgraphs from X to A, B, C.

$$\boxed{M_3(u,v,w) := \frac{1}{|X|} \# \text{ of 3-stars with leaves u,v,w}}$$



- Nodes in A, B, C: words and X: documents.

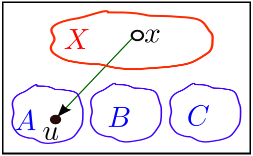
Learning via Edge and 3-Star Counts

Recall Stochastic Block Model..

- k communities and network size n
- Each node belongs to one community: for node u, $\boxed{\pi_u = e_i}$ if u is in community i. e_i is the basis vector in i^{th} coordinate.
- Probability of an edge from u to v is $\boxed{\pi_u^\top P \pi_v}$, where P is block connectivity matrix
- Independent Bernoulli draws for edges
- Probability of edges from X to A is $\boxed{\Pi_X^\top P \Pi_A}$, where Π_A has π_a, $a \in A$ as column vectors.
- Denote $\boxed{F_A := \Pi_A^\top P^\top}$ and $\boxed{\lambda_i = \mathbb{P}[\pi = e_i]}$.

Moments for Stochastic Block Model

- Denote $\boxed{F_A := \Pi_A^\top P^\top}$ and $\boxed{\lambda_i = \mathbb{P}[\pi = e_i]}$.



Edge Count Matrix

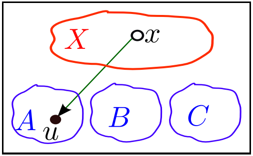
- Adjacency Submatrices $G_{X,A}$, $G_{X,B}$, $G_{X,C}$

$$\boxed{\mathbb{E}[G_{X,A}^\top | \Pi_{A,X}] = \Pi_X^\top P \Pi_A = \Pi_A^\top P^\top \Pi_X = F_A \Pi_X}$$

Moments for Stochastic Block Model

- Denote $\boxed{F_A := \Pi_A^\top P^\top}$ and $\boxed{\lambda_i = \mathbb{P}[\pi = e_i]}$.

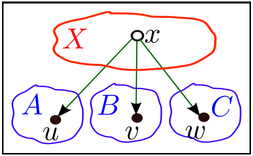
Edge Count Matrix



- Adjacency Submatrices $G_{X,A}$, $G_{X,B}$, $G_{X,C}$

$$\boxed{\mathbb{E}[G_{X,A}^\top|\Pi_{A,X}] = \Pi_X^\top P \Pi_A = \Pi_A^\top P^\top \Pi_X = F_A \Pi_X}$$

3-Star Count Tensor



- $\#$ of 3-star subgraphs from X to A, B, C.

$$\boxed{M_3 := \frac{1}{|X|} \sum_{i \in X} [G_{i,A}^\top \otimes G_{i,B}^\top \otimes G_{i,C}^\top]}$$

$$\boxed{\mathbb{E}[M_3|\Pi_{A,B,C}] = \sum_i \lambda_i [(F_A)_i \otimes (F_B)_i \otimes (F_C)_i]}$$

Goal: Recover $F_A, F_B, F_C, \vec{\lambda}$

Outline

Outline

1. Introduction
2. Tensor Form of Subgraph Counts
   - Connection to Topic Models
   - Tensor Forms for Network Models
3. Tensor Spectral Method for Learning
   - Tensor Preliminaries
   - Spectral Decomposition: Tensor Power Method
4. Conclusion

Tensor Basics: Multilinear Transformations

- For a tensor T, define (for matrices V_i of appropriate dimensions)

$$\boxed{[T(W_1, W_2, W_3)]_{i_1,i_2,i_3} := \sum_{j_1,j_2,j_3} (T)_{j_1,j_2,j_3} \prod_{m\in[3]} W_m(j_m, i_m)}$$

- For a matrix M, $\boxed{M(W_1, W_2) := W_1^\top M W_2}$.
- For a symmetric tensor T of the form

$$\boxed{T = \sum_{r=1}^{k} \lambda_r \phi_r^{\otimes 3}}$$

$$\boxed{\begin{aligned} T(W, W, W) &= \sum_{r\in[k]} \lambda_r (W^\top \phi_r)^{\otimes 3} \\ T(I, v, v) &= \sum_{r\in[k]} \lambda_r \langle v, \phi_r\rangle^2 \phi_r. \\ T(I, I, v) &= \sum_{r\in[k]} \lambda_r \langle v, \phi_r\rangle \phi_r \phi_r^\top. \end{aligned}}$$

Whiten: Convert to Orthogonal Symmetric Tensor

- Assume exact moments are known.

$$\boxed{\mathbb{E}[G_{X,A}^\top|\Pi_{A,X}] = F_A \Pi_X}$$

$$\boxed{\mathbb{E}[M_3|\Pi_{A,B,C}] = \sum_i \lambda_i [(F_A)_i \otimes (F_B)_i \otimes (F_C)_i]}$$

Whiten: Convert to Orthogonal Symmetric Tensor

- Assume exact moments are known.

$$\boxed{\mathbb{E}[G_{X,A}^\top|\Pi_{A,X}] = F_A \Pi_X}$$

$$\boxed{\mathbb{E}[M_3|\Pi_{A,B,C}] = \sum_i \lambda_i [(F_A)_i \otimes (F_B)_i \otimes (F_C)_i]}$$

- Use SVD of $G_{X,A}, G_{X,B}, G_{X,C}$ to obtain whitening matrices W_A, W_B, W_C
- Apply multi-linear transformation on M_3 using W_A, W_B, W_C.

$$\boxed{T := \mathbb{E}[M_3(W_A, W_B, W_C)|\Pi_{A,B,C}] = \sum_i w_i \mu_i^{\otimes 3}}$$

- T is symmetric orthogonal tensor: $\{\mu_i\}$ are orthonormal.

Spectral Tensor Decomposition of T

Outline

1. Introduction
2. Tensor Form of Subgraph Counts
   - Connection to Topic Models
   - Tensor Forms for Network Models
3. Tensor Spectral Method for Learning
   - Tensor Preliminaries
   - Spectral Decomposition: Tensor Power Method
4. Conclusion

Orthogonal Tensor Eigen Analysis

- Consider orthogonal symmetric tensor $T = \sum_i w_i \mu_i^{\otimes 3}$

$$\boxed{T = \sum_{i=1}^{k} w_i \mu_i^{\otimes 3}. \quad T(I, \mu_i, \mu_i) = w_i \mu_i}$$

Orthogonal Tensor Eigen Analysis

- Consider orthogonal symmetric tensor $T = \sum_i w_i \mu_i^{\otimes 3}$

$$\boxed{T = \sum_{i=1}^{k} w_i \mu_i^{\otimes 3}. \quad T(I, \mu_i, \mu_i) = w_i \mu_i}$$

Obtaining eigenvectors through power iterations

$$\boxed{u \mapsto \frac{T(I, u, u)}{\|T(I, u, u)\|}}$$

Orthogonal Tensor Eigen Analysis

- Consider orthogonal symmetric tensor $T = \sum_i w_i \mu_i^{\otimes 3}$

$$\boxed{T = \sum_{i=1}^{k} w_i \mu_i^{\otimes 3}. \quad T(I, \mu_i, \mu_i) = w_i \mu_i}$$

Obtaining eigenvectors through power iterations

$$\boxed{u \mapsto \frac{T(I, u, u)}{\|T(I, u, u)\|}}$$

Challenges and Solution

- Challenge: Other eigenvectors present

  Solution: Only stable vectors are basis vectors $\{\mu_i\}$

Orthogonal Tensor Eigen Analysis

- Consider orthogonal symmetric tensor $T = \sum_i w_i \mu_i^{\otimes 3}$

$$\boxed{T = \sum_{i=1}^{k} w_i \mu_i^{\otimes 3}. \quad T(I, \mu_i, \mu_i) = w_i \mu_i}$$

Obtaining eigenvectors through power iterations

$$\boxed{u \mapsto \frac{T(I, u, u)}{\|T(I, u, u)\|}}$$

Challenges and Solution

- Challenge: Other eigenvectors present

  Solution: Only stable vectors are basis vectors $\{\mu_i\}$
- Challenge: empirical moments

  Solution: robust tensor decomposition methods

Optimization Viewpoint for Tensor Eigen Analysis

Consider Norm Optimization Problem for Tensor T

- $\boxed{\max_u \; T(u,u,u) \qquad s.t. \; u^\top u = I}$
- Constrained stationary fixed points $T(I,u,u) = \lambda u$ and $u^\top u = I$.
- u is a local isolated maximizer if $w^\top (T(I,I,u) - \lambda I) w < 0$ for all w such that $w^\top w = I$ and w is orthogonal to u.

Review for Symmetric Matrices $M = \sum_i w_i \mu_i^{\otimes 2}$

- Constrained stationary points are the eigenvectors
- Only top eigenvector is a maximizer and stable under power iterations

Orthogonal Symmetric Tensors $T = \sum_i w_i \mu_i^{\otimes 3}$

- Stationary points are the eigenvectors (up to scaling)
- All basis vectors $\{\mu_i\}$ are local maximizers and stable under power iterations

Tensor Decomposition: Perturbation Analysis

- Observed tensor $\widetilde{T} = T + E$, where $T = \sum_{i \in k} w_i \mu_i^{\otimes 3}$ is orthogonal tensor and perturbation E, and $\|E\| \leq \epsilon$.
- Recall power iterations $\boxed{u \mapsto \dfrac{\widetilde{T}(I, u, u)}{\|\widetilde{T}(I, u, u)\|}}$

Tensor Decomposition: Perturbation Analysis

- Observed tensor $\widetilde{T} = T + E$, where $T = \sum_{i \in k} w_i \mu_i^{\otimes 3}$ is orthogonal tensor and perturbation E, and $\|E\| \leq \epsilon$.
- Recall power iterations $\boxed{u \mapsto \dfrac{\widetilde{T}(I,u,u)}{\|\widetilde{T}(I,u,u)\|}}$
- "Good" initialization vector $\boxed{\langle u^{(0)}, \mu_i \rangle^2 = \Omega\left(\dfrac{\epsilon}{w_{\min}}\right)}$

Tensor Decomposition: Perturbation Analysis

- Observed tensor $\widetilde{T} = T + E$, where $T = \sum_{i \in k} w_i \mu_i^{\otimes 3}$ is orthogonal tensor and perturbation E, and $\|E\| \leq \epsilon$.
- Recall power iterations $\boxed{u \mapsto \dfrac{\widetilde{T}(I, u, u)}{\|\widetilde{T}(I, u, u)\|}}$
- "Good" initialization vector $\boxed{\langle u^{(0)}, \mu_i \rangle^2 = \Omega\left(\dfrac{\epsilon}{w_{\min}}\right)}$

Perturbation Analysis

After N iterations, eigen pair (w_i, μ_i) is estimated up to $O(\epsilon)$ error, where

$$\boxed{N = O\left(\log k + \log\log \frac{w_{\max}}{\epsilon}\right).}$$

A. Anandkumar, R. Ge, D. Hsu, S.M. Kakade and M. Telgarsky "Tensor Decompositions for Learning Latent Variable Models," Preprint, October 2012.

Robust Tensor Power Method

$$\boxed{\widetilde{T} = \sum_i w_i \mu_i^{\otimes 3} + E}$$

Basic Algorithm

- Pick random initialization vectors
- Run power iterations $\boxed{u \mapsto \dfrac{\widetilde{T}(I,u,u)}{\|\widetilde{T}(I,u,u)\|}}$
- Go with the winner, deflate and repeat

Robust Tensor Power Method

$$\boxed{\widetilde{T} = \sum_i w_i \mu_i^{\otimes 3} + E}$$

Basic Algorithm

- Pick random initialization vectors
- Run power iterations $\boxed{u \mapsto \dfrac{\widetilde{T}(I,u,u)}{\|\widetilde{T}(I,u,u)\|}}$
- Go with the winner, deflate and repeat

Further Improvements

- Initialization: Use neighborhood vectors for initialization
- Stabilization: $\boxed{u^{(t)} \mapsto \alpha \dfrac{\widetilde{T}(I,u^{(t-1)},u^{(t-1)})}{\|\widetilde{T}(I,u^{(t-1)},u^{(t-1)})\|} + (1-\alpha)u^{(t-1)}}$

Efficient Learning Through Tensor Power Iterations

Outline

Conclusion

Mixed Membership Models

- Can model overlapping communities
- Efficient to learn from low order moments: edge counts and 3-star counts.

Tensor Spectral Method

- Whitened 3-star count tensor is an orthogonal symmetric tensor
- Efficient decomposition through power method
- Perturbation analysis: tight for stochastic block model