A Latent Space Model for Hypergraphs with Diversity and Heterogeneous Popularity

Ji Zhu

Department of Statistics, University of Michigan

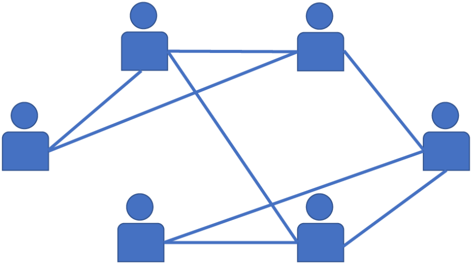
IASM-BIRS Workshop
December 11-15, 2023, Hangzhou, China

Joint work with Xianshi Yu (University of Wisconsin)

Network Data

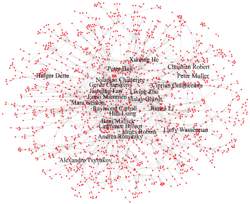
Network encodes relationships among individuals:

- A node represents an individual.
- A link between two nodes represents some connectivity.

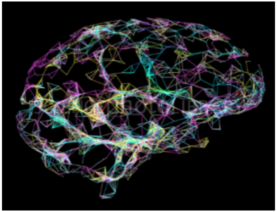


Network data are ubiquitous

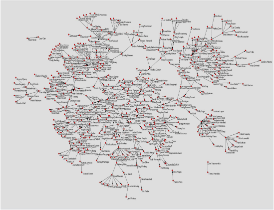
- Social network: Facebook, Twitter, collaboration network
- Biological network: brain network, gene network
- Information network: World Wide Web, Email network



Coauthor Network



Brain Network



Email Network

Dyadic vs Polyadic Relations

- Traditional network data analysis mainly focuses on dyadic relations. However, polyadic relations that involve more than two individuals are even more common in real-world interactions, e.g. co-author relationships, protein-protein interactions, product purchased together in supermarket transactions.
- In current practice, polyadic relations are often projected into a dyadic network before any analysis, which causes substantial loss of information.

Dyadic vs Polyadic Relations

- Traditional network data analysis mainly focuses on dyadic relations. However, polyadic relations that involve more than two individuals are even more common in real-world interactions, e.g. co-author relationships, protein-protein interactions, product purchased together in supermarket transactions.

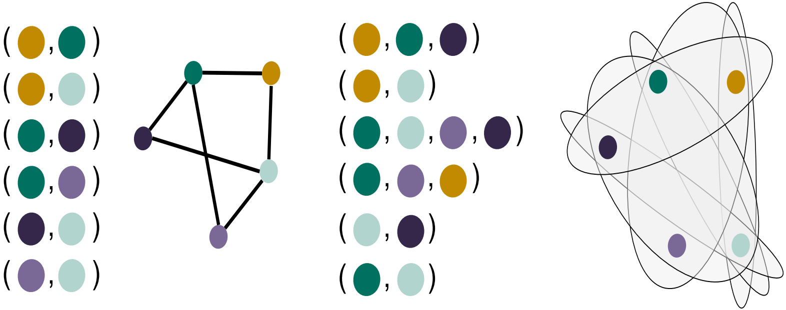
- In current practice, polyadic relations are often projected into a dyadic network before any analysis, which causes substantial loss of information.

Hypergraphs

- Relations that involve more than two individuals can be naturally represented using a hypergraph, which generalizes the traditional network.
- A hypergraph is defined by a set of nodes $\mathcal{V}$ and a (multi)set of hyperedges. Each hyperedge is a subset of $\mathcal{V}$, indicating existence of a relation among the nodes in the hyperedge.
- A subset of $\mathcal{V}$ of any size can be a hyperedge, in contrast to a traditional network which only allows edges to represent relations between exactly two nodes.

Hypergraphs

- Relations that involve more than two individuals can be naturally represented using a hypergraph, which generalizes the traditional network.
- A hypergraph is defined by a set of nodes $\mathcal{V}$ and a (multi)set of hyperedges. Each hyperedge is a subset of $\mathcal{V}$, indicating existence of a relation among the nodes in the hyperedge.
- A subset of $\mathcal{V}$ of any size can be a hyperedge, in contrast to a traditional network which only allows edges to represent relations between exactly two nodes.



Existing Methods for Hypergraphs

- Node clustering using generalization of graph cuts or modularity (Li and Milenkovic, 2017; Benson et al., 2020, 2021); hypergraph embedding (Zhou et al., 2006; Tu et al., 2018; Maleki et al., 2021). These algorithms are heuristic and lack statistical models or principles.
- Much of hypergraph modeling effort has been on a special type of hypergraph, the k-uniform hypergraph, where all hyperedges have the same number of nodes k (Ghoshdastidar and Dukkipati, 2017b; Chien et al., 2018; Kim et al., 2018; Lyu et al., 2021; Yuan and Qu, 2021).
- For non-uniform hypergraphs
  - Node clustering (Ghoshdastidar and Dukkipati, 2017a; Ke et al., 2019; Chodrow et al., 2021; Ng and Murphy, 2021)
  - Hereditary hypergraphs (Zhang and McCullagh, 2015; Lunagomez et al., 2017), where all subsets of a hyperedge are required to appear.
  - Notable works not limited to k-uniform hypergraphs, hereditary hypergraphs or clustering include Turnbull et al. (2019) and Zhen and Wang (2021), both of which assign latent positions to nodes. However, they do not allow for multiplicity of hyperedges.

Existing Methods for Hypergraphs

- Node clustering using generalization of graph cuts or modularity (Li and Milenkovic, 2017; Benson et al., 2020, 2021); hypergraph embedding (Zhou et al., 2006; Tu et al., 2018; Maleki et al., 2021). These algorithms are heuristic and lack statistical models or principles.
- Much of hypergraph modeling effort has been on a special type of hypergraph, the k-uniform hypergraph, where all hyperedges have the same number of nodes k (Ghoshdastidar and Dukkipati, 2017b; Chien et al., 2018; Kim et al., 2018; Lyu et al., 2021; Yuan and Qu, 2021).
- For non-uniform hypergraphs
  - Node clustering (Ghoshdastidar and Dukkipati, 2017a; Ke et al., 2019; Chodrow et al., 2021; Ng and Murphy, 2021)
  - Hereditary hypergraphs (Zhang and McCullagh, 2015; Lunagomez et al., 2017), where all subsets of a hyperedge are required to appear.
  - Notable works not limited to k-uniform hypergraphs, hereditary hypergraphs or clustering include Turnbull et al. (2019) and Zhen and Wang (2021), both of which assign latent positions to nodes. However, they do not allow for multiplicity of hyperedges.

Existing Methods for Hypergraphs

- Node clustering using generalization of graph cuts or modularity (Li and Milenkovic, 2017; Benson et al., 2020, 2021); hypergraph embedding (Zhou et al., 2006; Tu et al., 2018; Maleki et al., 2021). These algorithms are heuristic and lack statistical models or principles.
- Much of hypergraph modeling effort has been on a special type of hypergraph, the k-uniform hypergraph, where all hyperedges have the same number of nodes k (Ghoshdastidar and Dukkipati, 2017b; Chien et al., 2018; Kim et al., 2018; Lyu et al., 2021; Yuan and Qu, 2021).
- For non-uniform hypergraphs
  - Node clustering (Ghoshdastidar and Dukkipati, 2017a; Ke et al., 2019; Chodrow et al., 2021; Ng and Murphy, 2021)
  - Hereditary hypergraphs (Zhang and McCullagh, 2015; Lunagomez et al., 2017), where all subsets of a hyperedge are required to appear.
  - Notable works not limited to k-uniform hypergraphs, hereditary hypergraphs or clustering include Turnbull et al. (2019) and Zhen and Wang (2021), both of which assign latent positions to nodes. However, they do not allow for multiplicity of hyperedges.

Our goal:

- A latent space model for non-uniform, non-hereditary hypergraphs which allows multiple hyperedges.
- Motivated by observations in real-world scenarios, the proposed model aims to promote hyperedges among sets of nodes with high diversity in their latent positions, e.g. selection of products to purchase and selection of tags to assign to an online item.
- Account for variation in node popularity, as some nodes appear in hyperedges much more frequently than others.

Our goal:

- A latent space model for non-uniform, non-hereditary hypergraphs which allows multiple hyperedges.
- Motivated by observations in real-world scenarios, the proposed model aims to promote hyperedges among sets of nodes with high diversity in their latent positions, e.g. selection of products to purchase and selection of tags to assign to an online item.
- Account for variation in node popularity, as some nodes appear in hyperedges much more frequently than others.

Our goal:

- A latent space model for non-uniform, non-hereditary hypergraphs which allows multiple hyperedges.
- Motivated by observations in real-world scenarios, the proposed model aims to promote hyperedges among sets of nodes with high diversity in their latent positions, e.g. selection of products to purchase and selection of tags to assign to an online item.
- Account for variation in node popularity, as some nodes appear in hyperedges much more frequently than others.

Notation

- We use $\mathcal{V}$ to denote the set of all nodes and let $n_v := |\mathcal{V}|$ be the number of nodes in $\mathcal{V}$.
- The collection of observed hyperedges is denoted by $\mathcal{E} = \{e_1, e_2, \ldots, e_{n_e}\}$, where $e_\ell \subset \mathcal{V}$ for $\ell = 1, \ldots, n_e$ and n_e denotes the number of observed hyperedges.
- Note that elements in $\mathcal{E}$ can be duplicated, as the same hyperedge can be observed repeatedly.

Notation

- We use $\mathcal{V}$ to denote the set of all nodes and let $n_v := |\mathcal{V}|$ be the number of nodes in $\mathcal{V}$.
- The collection of observed hyperedges is denoted by $\mathcal{E} = \{e_1, e_2, \ldots, e_{n_e}\}$, where $e_\ell \subset \mathcal{V}$ for $\ell = 1, \ldots, n_e$ and n_e denotes the number of observed hyperedges.
- Note that elements in $\mathcal{E}$ can be duplicated, as the same hyperedge can be observed repeatedly.

A Latent Space Model

A generative latent space model for hypergraphs, which is driven by diversity within hyperedges and heterogeneous popularity among nodes.

- Each node is associated with
  - a latent position $v_i \in R^d$
  - a popularity parameter $\alpha_i \in R^+$
- Combining v_i and α_i
  - Define $\widetilde{v}_i^\top = (v_i^\top, \sqrt{\alpha_i} w_i^\top) \in R^{d+n_v}$, where w_i is the ith standard basis vector of R^{n_v}, i.e., the ith element is equal to 1 and all other elements are zero.
  - Example: $n_v = 3$

$$\begin{pmatrix} \tilde{v}_1^\top \\ \tilde{v}_2^\top \\ \tilde{v}_3^\top \end{pmatrix} = \begin{pmatrix} v_1^\top & \sqrt{\alpha_1} & 0 & 0 \\ v_2^\top & 0 & \sqrt{\alpha_2} & 0 \\ v_3^\top & 0 & 0 & \sqrt{\alpha_3} \end{pmatrix}$$

A Latent Space Model

A generative latent space model for hypergraphs, which is driven by diversity within hyperedges and heterogeneous popularity among nodes.

- Each node is associated with
  - a latent position $v_i \in R^d$
  - a popularity parameter $\alpha_i \in R^+$
- Combining v_i and α_i
  - Define $\widetilde{v}_i^\top = (v_i^\top, \sqrt{\alpha_i} w_i^\top) \in R^{d+n_v}$, where w_i is the ith standard basis vector of R^{n_v}, i.e., the ith element is equal to 1 and all other elements are zero.
  - Example: $n_v = 3$

$$\begin{pmatrix} \tilde{v}_1^\top \\ \tilde{v}_2^\top \\ \tilde{v}_3^\top \end{pmatrix} = \begin{pmatrix} v_1^\top & \sqrt{\alpha_1} & 0 & 0 \\ v_2^\top & 0 & \sqrt{\alpha_2} & 0 \\ v_3^\top & 0 & 0 & \sqrt{\alpha_3} \end{pmatrix}$$

A Latent Space Model

A generative latent space model for hypergraphs, which is driven by diversity within hyperedges and heterogeneous popularity among nodes.

- Each node is associated with
  - a latent position $v_i \in R^d$
  - a popularity parameter $\alpha_i \in R^+$
- Combining v_i and α_i
  - Define $\widetilde{v}_i^\top = (v_i^\top, \sqrt{\alpha_i} w_i^\top) \in R^{d+n_v}$, where w_i is the ith standard basis vector of R^{n_v}, i.e., the ith element is equal to 1 and all other elements are zero.
  - Example: $n_v = 3$

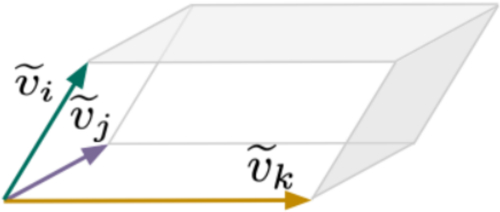
$$\begin{pmatrix} \tilde{v}_1^\top \\ \tilde{v}_2^\top \\ \tilde{v}_3^\top \end{pmatrix} = \begin{pmatrix} v_1^\top & \sqrt{\alpha_1} & 0 & 0 \\ v_2^\top & 0 & \sqrt{\alpha_2} & 0 \\ v_3^\top & 0 & 0 & \sqrt{\alpha_3} \end{pmatrix}$$

A Latent Space Model

We use E to denote a generic random hyperedge and e to denote its realization. Given any hyperedge $e \subset \mathcal{V}$, we consider the parallelotope formed with vectors $\widetilde{v}_i$'s, $i \in e$ and define the distribution $\mathcal{P}$ of a random hyperedge E by setting

$$P(E = e) \propto \text{vol}^2\left(\{\widetilde{v}_i, i \in e\}\right),$$

the square of the volume of the parallelotope formed by $\widetilde{v}_i$'s, $i \in e$.



Diversity and Popularity Hypergraph (DiPH) Model

- Let $L_{n_v \times n_v} := (\widetilde{v}_i^\top \widetilde{v}_j)_{i,j=1}^{n_v}$. It can be shown that

$$\mathrm{vol}^2\left(\{\widetilde{v}_i, i \in e\}\right) = \det(L_e),$$

where L_e denotes the submatrix of L indexed by e, and $\det(\cdot)$ denotes the determinant of a matrix.

- Moreover, one can also show that

$$\sum_{e \subset \mathcal{V}} \det(L_e) = \det(L + I).$$

Therefore, we have

$$P(E = e) = \frac{\det(L_e)}{\det(L + I)}, \text{ for any } e \subset \mathcal{V}.$$

- In addition, it is not difficult to see

$$L_{n_v \times n_v} = (v_i^\top v_j)_{i,j=1}^{n_v} + \mathrm{diag}(\alpha), \text{ where } \alpha := (\alpha_1, \ldots, \alpha_{n_v}).$$

Diversity and Popularity Hypergraph (DiPH) Model

- Let $L_{n_v \times n_v} := (\widetilde{v}_i^\top \widetilde{v}_j)_{i,j=1}^{n_v}$. It can be shown that

$$\mathrm{vol}^2\left(\{\widetilde{v}_i, i \in e\}\right) = \det(L_e),$$

  where L_e denotes the submatrix of L indexed by e, and $\det(\cdot)$ denotes the determinant of a matrix.

- Moreover, one can also show that

$$\sum_{e \subset \mathcal{V}} \det(L_e) = \det(L + I).$$

  Therefore, we have

$$P(E = e) = \frac{\det(L_e)}{\det(L + I)}, \text{ for any } e \subset \mathcal{V}.$$

- In addition, it is not difficult to see

$$L_{n_v \times n_v} = (v_i^\top v_j)_{i,j=1}^{n_v} + \mathrm{diag}(\alpha), \text{ where } \alpha := (\alpha_1, \ldots, \alpha_{n_v}).$$

Diversity and Popularity Hypergraph (DiPH) Model

- Let $L_{n_v \times n_v} := (\widetilde{v}_i^\top \widetilde{v}_j)_{i,j=1}^{n_v}$. It can be shown that

$$\mathrm{vol}^2\left(\{\widetilde{v}_i, i \in e\}\right) = \det(L_e),$$

where L_e denotes the submatrix of L indexed by e, and $\det(\cdot)$ denotes the determinant of a matrix.

- Moreover, one can also show that

$$\sum_{e \subset \mathcal{V}} \det(L_e) = \det(L + I).$$

Therefore, we have

$$P(E = e) = \frac{\det(L_e)}{\det(L + I)}, \text{ for any } e \subset \mathcal{V}.$$

- In addition, it is not difficult to see

$$L_{n_v \times n_v} = (v_i^\top v_j)_{i,j=1}^{n_v} + \mathrm{diag}(\alpha), \text{ where } \alpha := (\alpha_1, \ldots, \alpha_{n_v}).$$

Model Identifiability

To ensure identifiability, we constrain $\|v_i\|_2$, $i = 1, \ldots, n_v$ to be an (unknown) constant, i.e.

$$\|v_i\|_2 = \|v_{i'}\|_2 > 0 \text{ for any } i, i' \in 1, \ldots, n_v.$$

Under this constraint, it can be shown that α is identifiable and v_i's are identifiable up to a universal orthogonal transformation and individual sign changes.

Remarks

The parallelotope has a comparatively large volume when $\|\widetilde{v}_i\|_2$'s ($i \in e$) are large and the directions of vectors $\{\widetilde{v}_i, i \in e\}$ are well separated. Recall $\widetilde{v}_i^\top = (v_i^\top, \sqrt{\alpha_i} w_i^\top) \in R^{d+n_v}$, and $\|\widetilde{v}_i\|_2^2 = \|v_i\|_2^2 + \alpha_i$. Hence

- $\{\widetilde{v}_i, i \in e\}$ are well separated when $\{v_i, i \in e\}$ are well separated.
- Since $\|v_i\|_2$ is a constant as i varies, $\|\widetilde{v}_i\|_2$ depends solely on α_i. Nodes with large α_i values are more likely to form hyperedges.

Example: let $\beta := \|v_i\|_2^2$ which is a constant across i. We have

$$P(E = \{i, i'\}) = \frac{(\alpha_i + \beta)(\alpha_{i'} + \beta) - \cos^2(v_i, v_{i'})\beta^2}{\det(L + I)}.$$

We refer our model as the Diversity and Popularity Hypergraph (DiPH) model.

Remarks

The parallelotope has a comparatively large volume when $\|\widetilde{v}_i\|_2$'s ($i \in e$) are large and the directions of vectors $\{\widetilde{v}_i, i \in e\}$ are well separated. Recall $\widetilde{v}_i^\top = (v_i^\top, \sqrt{\alpha_i} w_i^\top) \in R^{d+n_v}$, and $\|\widetilde{v}_i\|_2^2 = \|v_i\|_2^2 + \alpha_i$. Hence

- $\{\widetilde{v}_i, i \in e\}$ are well separated when $\{v_i, i \in e\}$ are well separated.
- Since $\|v_i\|_2$ is a constant as i varies, $\|\widetilde{v}_i\|_2$ depends solely on α_i. Nodes with large α_i values are more likely to form hyperedges.

Example: let $\beta := \|v_i\|_2^2$ which is a constant across i. We have

$$P(E = \{i, i'\}) = \frac{(\alpha_i + \beta)(\alpha_{i'} + \beta) - \cos^2(v_i, v_{i'})\beta^2}{\det(L + I)}.$$

We refer our model as the Diversity and Popularity Hypergraph (DiPH) model.

Remarks

The parallelotope has a comparatively large volume when $\|\widetilde{v}_i\|_2$'s ($i \in e$) are large and the directions of vectors $\{\widetilde{v}_i, i \in e\}$ are well separated. Recall $\widetilde{v}_i^\top = (v_i^\top, \sqrt{\alpha_i} w_i^\top) \in R^{d+n_v}$, and $\|\widetilde{v}_i\|_2^2 = \|v_i\|_2^2 + \alpha_i$. Hence

- $\{\widetilde{v}_i, i \in e\}$ are well separated when $\{v_i, i \in e\}$ are well separated.
- Since $\|v_i\|_2$ is a constant as i varies, $\|\widetilde{v}_i\|_2$ depends solely on α_i. Nodes with large α_i values are more likely to form hyperedges.

Example: let $\beta := \|v_i\|_2^2$ which is a constant across i. We have

$$P(E = \{i, i'\}) = \frac{(\alpha_i + \beta)(\alpha_{i'} + \beta) - \cos^2(v_i, v_{i'})\beta^2}{\det(L + I)}.$$

We refer our model as the Diversity and Popularity Hypergraph (DiPH) model.

Remarks

The parallelotope has a comparatively large volume when $\|\widetilde{v}_i\|_2$'s ($i \in e$) are large and the directions of vectors $\{\widetilde{v}_i, i \in e\}$ are well separated. Recall $\widetilde{v}_i^\top = (v_i^\top, \sqrt{\alpha_i} w_i^\top) \in R^{d+n_v}$, and $\|\widetilde{v}_i\|_2^2 = \|v_i\|_2^2 + \alpha_i$. Hence

- $\{\widetilde{v}_i, i \in e\}$ are well separated when $\{v_i, i \in e\}$ are well separated.
- Since $\|v_i\|_2$ is a constant as i varies, $\|\widetilde{v}_i\|_2$ depends solely on α_i. Nodes with large α_i values are more likely to form hyperedges.

Example: let $\beta := \|v_i\|_2^2$ which is a constant across i. We have

$$P(E = \{i, i'\}) = \frac{(\alpha_i + \beta)(\alpha_{i'} + \beta) - \cos^2(v_i, v_{i'})\beta^2}{\det(L + I)}.$$

We refer our model as the Diversity and Popularity Hypergraph (DiPH) model.

Remarks

The parallelotope has a comparatively large volume when $\|\widetilde{v}_i\|_2$'s $(i \in e)$ are large and the directions of vectors $\{\widetilde{v}_i, i \in e\}$ are well separated. Recall $\widetilde{v}_i^\top = (v_i^\top, \sqrt{\alpha_i} w_i^\top) \in R^{d+n_v}$, and $\|\widetilde{v}_i\|_2^2 = \|v_i\|_2^2 + \alpha_i$. Hence

- $\{\widetilde{v}_i, i \in e\}$ are well separated when $\{v_i, i \in e\}$ are well separated.
- Since $\|v_i\|_2$ is a constant as i varies, $\|\widetilde{v}_i\|_2$ depends solely on α_i. Nodes with large α_i values are more likely to form hyperedges.

Example: let $\beta := \|v_i\|_2^2$ which is a constant across i. We have

$$P(E = \{i, i'\}) = \frac{(\alpha_i + \beta)(\alpha_{i'} + \beta) - \cos^2(v_i, v_{i'})\beta^2}{\det(L + I)}.$$

We refer our model as the Diversity and Popularity Hypergraph (DiPH) model.

Remarks

- Many existing hypergraph models assume independent Bernoulli distributions for whether there exists a hyperedge with configuration e for all $e \subset \mathcal{V}$ (Stasi et al., 2014; Ghoshdastidar and Dukkipati, 2017a; Zhen and Wang, 2021; Ke et al., 2019; Lyu et al., 2021), while the DiPH model considers hyperedges as i.i.d. realizations of one hyperedge distribution $\mathcal{P}$.
- The hyperedge distribution $\mathcal{P}$ in the DiPH model is a specially structured (discrete) determinantal point process (DPP) (Kulesza and Taskar, 2012). A DPP is a type of distribution over the power set of a point set (e.g. $\mathcal{V}$), and the term L for a generic DPP can be any positive semi-definite matrix. In the DiPH model, we require L to take the special form of $L_{n_v \times n_v} = (v_i^\top v_j)_{i,j=1}^{n_v} + \mathrm{diag}(\alpha)$, which reduces the the number of parameters from $n_v(n_v+1)/2$ to $n_v d + 1$.

Remarks

- Many existing hypergraph models assume independent Bernoulli distributions for whether there exists a hyperedge with configuration e for all $e \subset \mathcal{V}$ (Stasi et al., 2014; Ghoshdastidar and Dukkipati, 2017a; Zhen and Wang, 2021; Ke et al., 2019; Lyu et al., 2021), while the DiPH model considers hyperedges as i.i.d. realizations of one hyperedge distribution $\mathcal{P}$.

- The hyperedge distribution $\mathcal{P}$ in the DiPH model is a specially structured (discrete) determinantal point process (DPP) (Kulesza and Taskar, 2012). A DPP is a type of distribution over the power set of a point set (e.g. $\mathcal{V}$), and the term L for a generic DPP can be any positive semi-definite matrix. In the DiPH model, we require L to take the special form of $L_{n_v \times n_v} = (v_i^\top v_j)_{i,j=1}^{n_v} + \mathrm{diag}(\alpha)$, which reduces the the number of parameters from $n_v(n_v+1)/2$ to $n_v d + 1$.

Model Parameter Estimation

To fit the DiPH model, we apply the maximum likelihood estimation (MLE) to an observed hypergraph with $\mathcal{E} = \{e_1, e_2, \ldots, e_{n_e}\}$. We assume that $e_\ell \sim_{\text{i.i.d.}} \mathcal{P}$ for $\ell = 1, \ldots, n_e$. Under the reparameterization using $V_{n_v \times d}$ and β by setting

$$V_{i\cdot} := \frac{v_i}{\|v_i\|_2} \text{ for all } i, \text{ and } \beta := \|v_i\|_2^2,$$

we have

$$\underset{\substack{V_{n_v \times d}, \|V_{i\cdot}\|_2 = 1, \\ \beta > 0, \alpha_i > 0}}{\arg\max} \quad -\log\det\left(\beta VV^\top + \mathrm{diag}(\alpha) + I\right) + \frac{1}{n_e}\sum_{\ell=1}^{n_e} \log\det\left(\beta (VV^\top)_{e_\ell} + \mathrm{diag}(\alpha)_{e_\ell}\right),$$

which allows a standard application of the projected gradient descent (ascent) algorithm.

Theoretical Results

- Consistency: under certain regularity conditions, if $n_v > 2d$ and $\{v_1^*, \cdots, v_{n_v}^*\}$ span $\mathbb{R}^d$, then as $n_e \to \infty$, we have

$$\min_{O \in \mathbb{O}_d} \sum_{i=1}^{n_v} \min_{s=\pm 1} \|\hat{v}_i - sOv_i^*\|_2 \xrightarrow{p} 0,$$

$$\|\hat{\alpha} - \alpha^*\|_2 \xrightarrow{p} 0,$$

$$\min_{S \in \mathbb{D}_{n_v}} \|\hat{L} - SL^*S\|_F \xrightarrow{p} 0,$$

  where $\mathbb{O}_d$ is the set of all d-by-d orthogonal matrices and $\mathbb{D}_{n_v}$ is the set of all n_v-by-n_v diagonal matrices with diagonal entries in $\{-1, 1\}$.

- Asymptotic normality: let $\tilde{L} := \arg\min_{M \in \{S\hat{L}S | S \in \mathbb{D}_{n_v}\}} \|M - L^*\|_F$. Under certain regularity conditions, we have

$$\sqrt{n_e} \cdot \mathrm{vec}(\tilde{L} - L^*) \xrightarrow{dist.} N(0, \Sigma),$$

  where Σ is a function of L^* that can be derived.

Theoretical Results

- Consistency: under certain regularity conditions, if $n_v > 2d$ and $\{v_1^*, \cdots, v_{n_v}^*\}$ span $\mathbb{R}^d$, then as $n_e \to \infty$, we have

$$\min_{O \in \mathbb{O}_d} \sum_{i=1}^{n_v} \min_{s=\pm 1} \|\hat{v}_i - sOv_i^*\|_2 \xrightarrow{p} 0,$$

$$\|\hat{\alpha} - \alpha^*\|_2 \xrightarrow{p} 0,$$

$$\min_{S \in \mathbb{D}_{n_v}} \|\hat{L} - SL^*S\|_F \xrightarrow{p} 0,$$

  where $\mathbb{O}_d$ is the set of all d-by-d orthogonal matrices and $\mathbb{D}_{n_v}$ is the set of all n_v-by-n_v diagonal matrices with diagonal entries in $\{-1, 1\}$.

- Asymptotic normality: let $\tilde{L} := \underset{M \in \{S\hat{L}S | S \in \mathbb{D}_{n_v}\}}{\arg\min} \|M - L^*\|_F$. Under certain regularity conditions, we have

$$\sqrt{n_e} \cdot \mathrm{vec}(\tilde{L} - L^*) \xrightarrow{dist.} N(0, \Sigma),$$

  where Σ is a function of L^* that can be derived.

Simulation Studies

We evaluate the following relative errors

$$\frac{\ell(\hat{V}, V^*)}{\|V^*\|_F}, \frac{\ell(\hat{\beta}, \beta^*)}{|\beta^*|}, \frac{\ell(\hat{\alpha}, \alpha^*)}{\|\alpha^*\|_2} \text{ and } \frac{\ell(\hat{L}, L^*)}{\|L^*\|_F},$$

where

$$\ell(\hat{V}, V^*) := \min_{O \in \mathbb{O}_d, S \in \mathbb{D}_{n_v}} \|\hat{V} - SV^*O\|_F, \qquad \ell(\hat{\beta}, \beta^*) := |\hat{\beta} - \beta^*|,$$
$$\ell(\hat{L}, L^*) := \min_{S \in \mathbb{D}_{n_v}} \|\hat{L} - SL^*S\|_F, \qquad \ell(\hat{\alpha}, \alpha^*) := \|\hat{\alpha} - \alpha^*\|_2,$$

and

$$\hat{V}_{i\cdot} = \frac{\hat{v}_i}{\|\hat{v}_i\|_2}, \ \hat{\beta} = \|\hat{v}_i\|_2^2 \ \text{ for all } i.$$

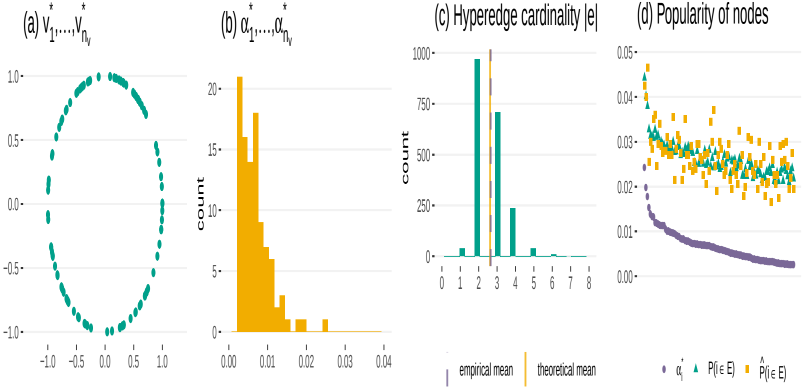
Note that we evaluate the direction and the length of $\hat{v}_i$ separately.

Simulation Setting I: Uniformly Distributed v_i^*'s

- $v_1^*, \cdots, v_{n_v}^*$ are uniformly distributed on S^{d-1}.
- $n_v = 100$ and $d = 2$, 3 and 4
- Generate $\gamma_i \sim_{i.i.d.}$ Beta$(1, 4)$ and then set $\sqrt{\alpha_i^*} = 0.15\gamma_i + 0.05$ for $i = 1, \ldots, n_v$. As a result, most nodes have a small popularity parameter value (with α_i^* close to 0.05^2), while a few nodes have a relatively large popularity parameter value (with α_i^* approaching 0.2^2).

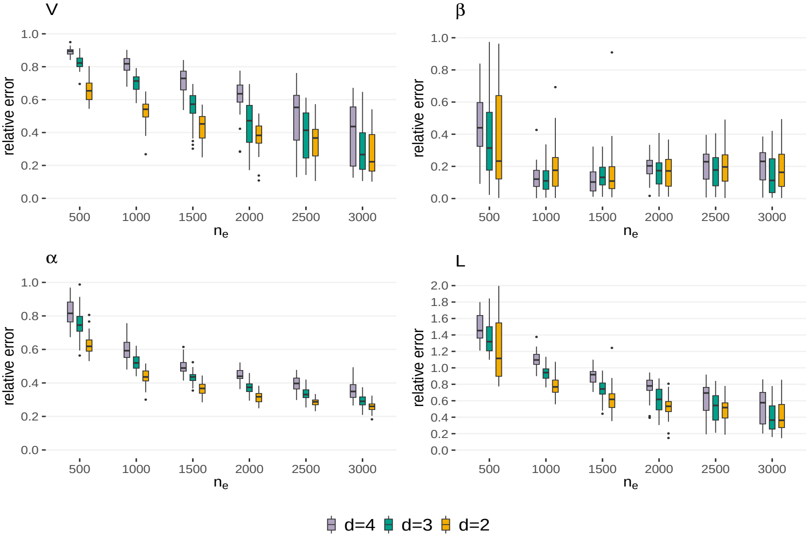
Setting I: Results

$d = 2,\ n_v = 100$ and $n_e = 2000$



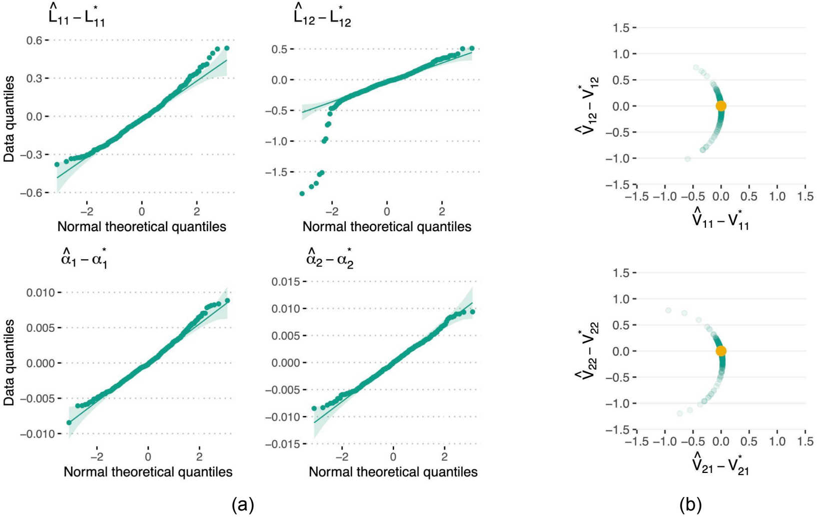
Setting I: Results

Relative errors of $\hat{V}$, $\hat{\beta}$, $\hat{\alpha}$ and $\hat{L}$.



Setting I: Results

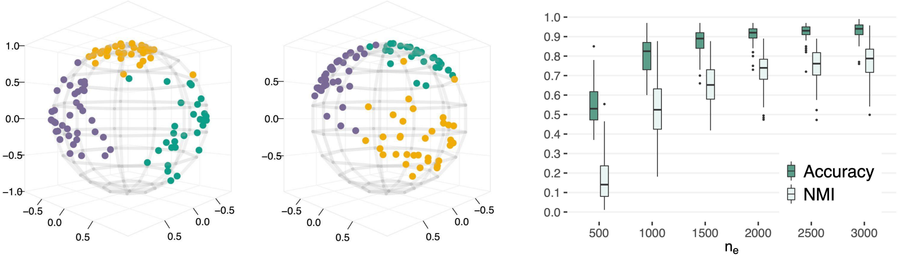
$d = 2$, $n_v = 100$ and $n_e = 3000$



Simulation Setting II: Clustered v_i^*'s

- $v_1^*, \cdots, v_{n_v}^*$ have a clustered structure ($d = 3$).
- $n_v = 100$ nodes; each is randomly assigned to one of three clusters with equal probability.
- The latent position of node v_i^* is then generated from a von Mise-Fisher distribution parameterized by the concentration parameter $\kappa = 10$ and one of the three mean directions $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$.

Setting II: Results



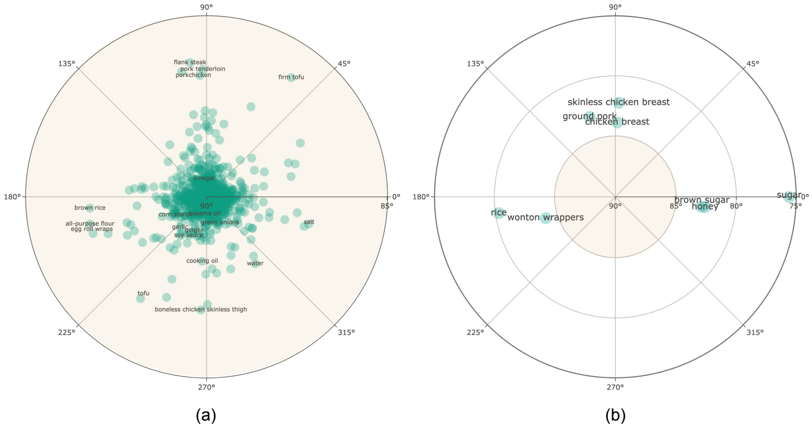
(a) $v_1^*, \ldots, v_{100}^*$ (b) $\hat{v}_1, \ldots, \hat{v}_{100}$ (c) Clustering results

What's Cooking Data

- Recipes on yummly.com; 39,774 recipes in total.
- We focus on recipes in the Chinese cuisine, which consist of $n_e = 2,673$ recipes with $n_v = 906$ ingredients (after preprocessing).
  - East Asian cuisines tend to avoid compound sharing ingredients. (Ahn et al., 2011)

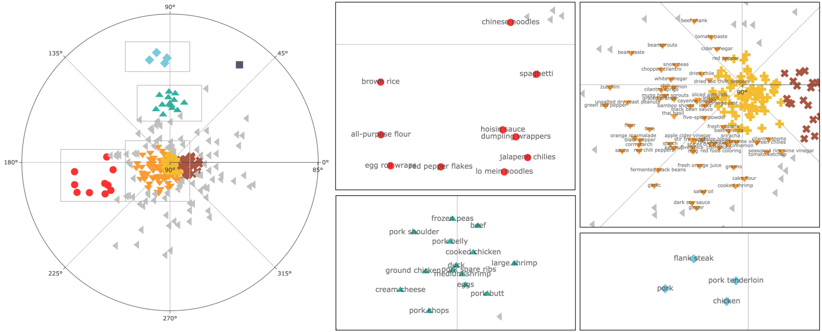
```
recipe                       Large list (2673 elements, 2.4 Mb)
  : chr [1:15] "low sodium soy sauce" "fresh ginger" "dry mustard" "green beans" ...
  : chr [1:12] "sesame seeds" "red pepper" "yellow peppers" "water" ...
  : chr [1:15] "sugar" "lo mein noodles" "salt" "chicken broth" ...
  : chr [1:14] "green bell pepper" "egg roll wrappers" "sweet and sour sauce" "corn starch" ...
  : chr [1:8] "eggs" "mandarin oranges" "water" "orange liqueur" ...
  : chr [1:16] "chinese rice wine" "shiitake" "sesame oil" "oil" ...
  : chr [1:13] "water" "salt" "garlic cloves" "cold water" ...
  : chr [1:14] "sugar" "napa cabbage leaves" "garlic" "chicken leg quarters" ...
  : chr [1:11] "water" "daikon" "sugar" "pork ribs" ...
  : chr [1:10] "fresh ginger" "garlic" "low sodium soy sauce" "mirin" ...
  : chr [1:13] "scallion greens" "Sriracha" "carrots" "soy sauce" ...
  : chr [1:13] "coconut oil" "ground red pepper" "scallions" "curry powder" ...
  : chr [1:14] "chicken stock" "minced garlic" "seasoned rice wine vinegar" "snow peas" ...
  : chr [1:16] "soy sauce" "egg whites" "red wine vinegar" "peanut oil" ...
  : chr [1:6] "Philadelphia Cream Cheese" "powdered sugar" "oil" "crab meat" ...
```

What's Cooking



Clustering of Ingredients

Focus on 298 ingredients that have appeared 10 or more times.



Conclusion

- We proposed a new hypergraph latent space model, which allows hyperedges with varying cardinality.
- It is driven by diversity (rather than similarity) of nodes within hyperedges.
- The proposed model admits heterogeneity in the popularity of nodes.
- We have established the consistency and asymptotic normality for the MLE estimates of the model parameters.

Thank you!