High-order Joint Embedding for Multi-Level Link Prediction

Annie Qu

University of California, Irvine
Department of Statistics

2025 SNAB Workshop

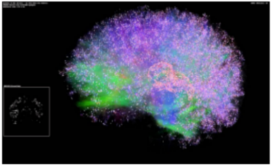
*This is a joint work with Yubai Yuan (Penn State)

SNAB 2025, Tokyo, JP

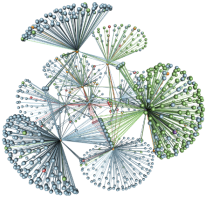
Network Data

social networks

brain networks



HIV protein interaction networks



- Traditional network data: collection of two-way relation
  - interactions between a pair of nodes
  - two-way relations are **independent** to each other
  - Real world complex network: multi-way (subgroup) interaction

Motivation: Ego Network in Social Media

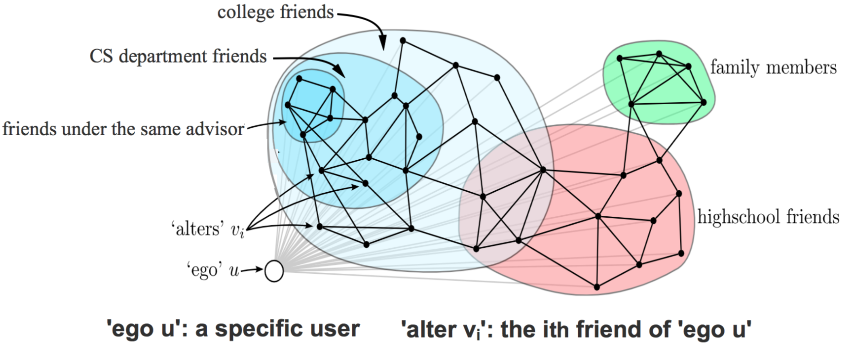


Fig. 1: Facebook Ego-network, adapted from McAuley and Leskovec (2012)

- Two-way relations: friendships among people
- **Social circles**: multi-way relations among people

Network Beyond Two-way Relation

- **Multi-way relations:** protein complex, social circle, authorship, ...
  - relations among a group of nodes
  - capture **higher-order** interactions among nodes
  - subgroup information in network
- Two-way and multi-way relations coexist among the same set of nodes

Model Multi-Way Relations as Hyperlink

- **Hyperlink: links to connect nodes in a subgroup**
- *M*-**order hyperlink**: links connecting M nodes

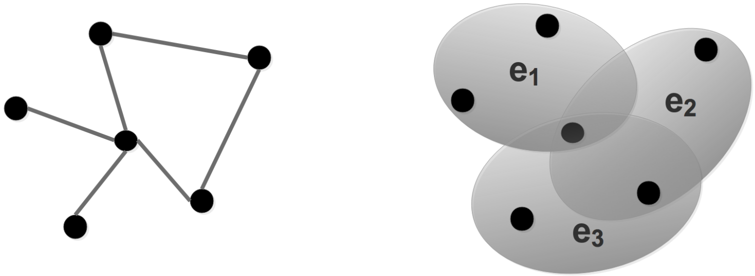


Fig. 2: Left: pairwise links; Right: three 3-order hyperlinks

- A pairwise link is a special case: 2-order hyperlink

Hyperlink Encodes Subgroup Similarity

- Capture nodes' similarity at different levels
- Pairwise link:
  - common features shared by two nodes only
- Hyperlink:
  - common features shared by all nodes in a subgroup

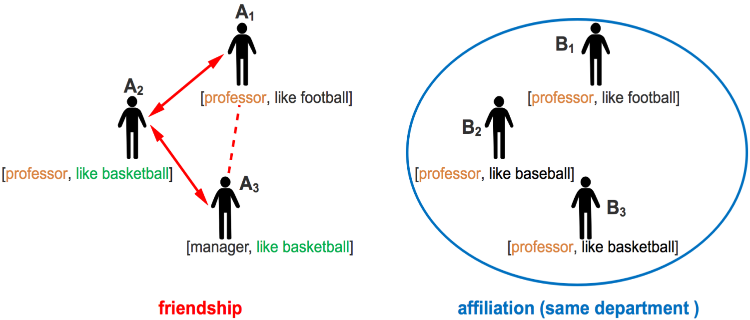
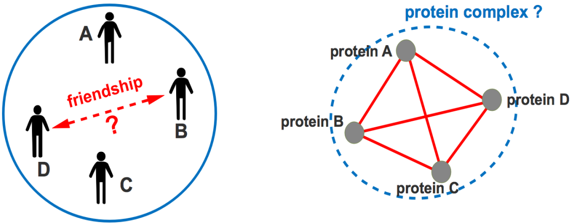


Fig. 3: Differences between pairwise similarity and subgroup similarity

Dependency among Pairwise Links and Hyperlinks

- Dependency between pairwise links and hyperlinks
  - sharing the same set of nodes
  - high-order relations arise from specific connection patterns
- Incorporate mutual information



observe {A, B, C, D} in the same department

observe pairwise interaction clique among protein {A, B, C, D}

Fig. 4: B and D are more likely to be friends or enemies given that they are **in the same department**; protein A, B, C, and D are more likely to form a protein complex given that they are **pairwisely interacted with each other**

Goal: Link Prediction

- Informal scoring methods (Adamic and Adar, 2003; Katz, 1953; Kossinet, 2006)
- Exponential-family random graph models (Holland and Leinhardt, 1981; Hunter et al., 2012)
- Latent variable models (Hoff et al., 2002; Handcock et al., 2007; Kim et al., 2018)
- Embedding-based methods:
  - matrix factorization (Ahmed et al., 2013; Cao et al., 2015)
  - random walk (Grover and Leskovec, 2016; Perozzi et al., 2014)
  - graph neural networks (Scarselli et al., 2009)
- **Our goals**
  - **Predict pairwise links and hyperlinks jointly**
  - **Borrow information between pairwise links and hyperlinks**

Network Embedding

- **Embedding**: map nodes into latent factors $\boldsymbol{Z}_i \in R^r, \boldsymbol{Z} = \{\boldsymbol{Z}_i\}_{i=1}^N$
- Z_{ir} represents **hidden features** for node i

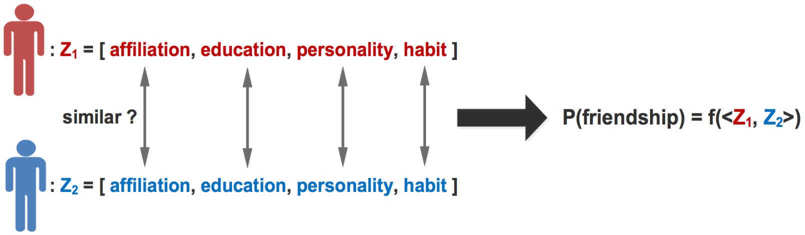


Fig. 5: The probability of potential link depends on the concordance via inner product between latent factors Z_1, Z_2, and $f(\cdot)$ is a link function.

- Measure concordance among nodes: transform a binary link to a continuous probability

Proposed Embedding Framework

- Observed data:
  - $\mathcal{V}$: a set of nodes $\{v_i\}_{i=1}^{N}$
  - $\Omega_{\mathbf{A}}$: a set of pairwise links and non-links on $\mathcal{V}$
  - $\Omega_{\mathcal{A}}$: a set of hyperlinks and non-hyperlinks on $\mathcal{V}$
  - $|\Omega_{\mathbf{A}}|$, $|\Omega_{\mathcal{A}}|$: cardinality of $\Omega_{\mathbf{A}}$ and $\Omega_{\mathcal{A}}$
- Obtain latent factors $\boldsymbol{Z} = \text{argmin}_{\boldsymbol{Z}}\, Loss(\boldsymbol{Z}; \Omega_{\boldsymbol{A}}, \Omega_{\mathcal{A}})$

$$Loss(\boldsymbol{Z}; \Omega_{\boldsymbol{A}}, \Omega_{\mathcal{A}}) = Loss_{pair}(\boldsymbol{Z}; \Omega_{\boldsymbol{A}}) + Loss_{hyper}(\boldsymbol{Z}; \Omega_{\mathcal{A}}) + \lambda Penalty(\boldsymbol{Z})$$

- $Loss_{pair}$: mismatch between observed and predicted pairwise links
- $Loss_{hyper}$: mismatch between observed and predicted hyperlinks
- $Penalty$: regularizations to filter out spurious links
- Integrate different-order moment information of $\boldsymbol{Z}$

Encode Pairwise Links

- Pairwise link network $\Rightarrow$ **adjacent matrix** $A = \{-1, 0, 1\}^{N^2}$

$$A_{ij} = A_{ji} = \begin{cases} 1, & i \text{ and } j \text{ are connected} \\ 0, & i \text{ and } j \text{ are not connected} \\ -1, & \text{not observed} \end{cases}$$

- Minimize $Loss_{pair}(\boldsymbol{Z})$ by encouraging concordance between embeddings of connected nodes:

$$Loss_{pair}(\boldsymbol{Z}) = \frac{1}{|\Omega_{\boldsymbol{A}}|} \sum_{A_{ij} \in \Omega_{\boldsymbol{A}}} \Big(\underbrace{A_{ij}}_{\text{observed}} - \underbrace{f\left[\boldsymbol{Z}_i^T \boldsymbol{Z}_j\right]}_{P(A_{ij}=1)} \Big)^2$$

where $f(\cdot)$ is a link function, such as logit

- If v_i and v_j are connected $\Longleftrightarrow$ a larger $\boldsymbol{Z}_i^T \boldsymbol{Z}_j$
- Preserve observed **pairwise similarity** in latent space

$({}^*\Omega_{\boldsymbol{A}} = \{A_{ij} | A_{ij} \in \{0, 1\}\})$

Tensor Modeling for Hyperlink Network

- A m-order tensor representation $\mathcal{A} = \{-1, 0, 1\}^{N^m}$:

$$\mathcal{A}_{i_1 \cdots i_m} = \begin{cases} 1, \text{hyperlink among } \{i_1, \cdots, i_m\} \\ 0, \text{non-hyperlink among } \{i_1, \cdots, i_m\} \\ -1, \text{not observed} \end{cases}$$

- No order among nodes in hyperlink: $\mathcal{A}$ is **supersymmetric**

$$\mathcal{A}_{i_1 \cdots i_m} = \{\mathcal{A}_{\sigma(i_1) \cdots \sigma(i_m)} | \sigma(\cdot) \text{ is any index permutation}\}$$

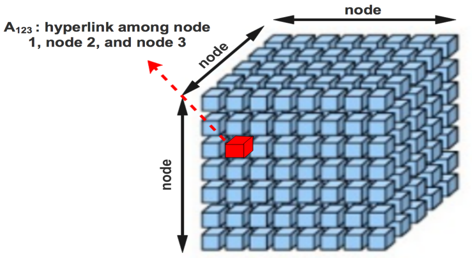


Fig. 6: A 3-uniform hypergraph formulates as a third-order tensor.

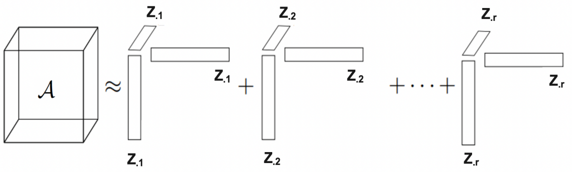
Tensor Decomposition

- CANDECOMP/PARAFAC decomposition (Hitchcock, 1927):

$$\mathcal{A} = \sum_{k=1}^{r} \underbrace{\boldsymbol{Z}_{\cdot k} \circ \boldsymbol{Z}_{\cdot k} \cdots \circ \boldsymbol{Z}_{\cdot k}}_{m},$$

$$\Longrightarrow \mathcal{A}_{i_1 i_2 \cdots i_m} = \sum_{k=1}^{r} Z_{i_1 k} \times Z_{i_2 k} \times \cdots \times Z_{i_m k},$$

where $\boldsymbol{Z}_{\cdot k} (k = 1, \cdots r)$ are N-dimensional vectors, $\circ$ is the outer product, and r is the rank of tensor

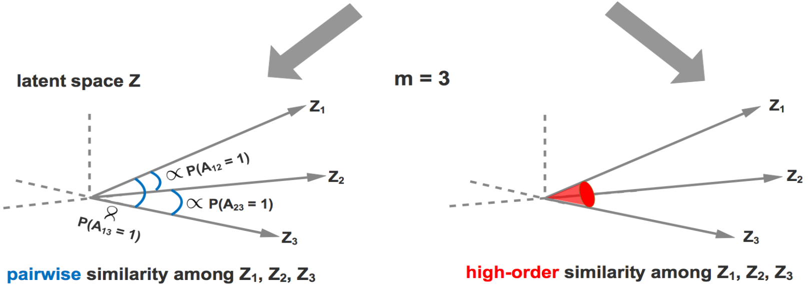


- Need to incorporate dependency between pairwise links and hyperlinks

Proposed Hyperlink Tensor Modeling

- Model $\mathcal{A}_{i_1 i_2 \cdots i_m} \in \{0, 1\}$ via $\mathbf{P}_{\boldsymbol{Z}}(\mathcal{A}_{i_1 i_2 \cdots i_m} = 1) \in [0, 1]$

$$\mathbf{P}_{\boldsymbol{Z}}(\mathcal{A}_{i_1 i_2 \cdots i_m} = 1) = \mathbf{f}\left(\textbf{pairwise concordance on } \{Z_{i_1}, \cdots, Z_{i_m}\} + \textbf{high-order concordance on } \{Z_{i_1}, \cdots, Z_{i_m}\}\right)$$



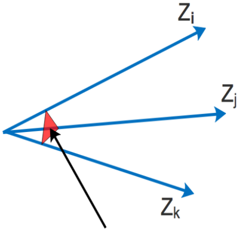
- Incorporate connectivity information at different resolution levels
- $f(\cdot)$ can be a non-linear link function, e.g., logit function and probit function

High-Order Concordance Modeling

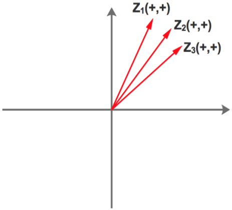
$$\mathbf{P}_Z(\mathcal{A}_{i_1 i_2 \cdots i_m} = 1) = f\left(\text{pairwise concordance on } \{Z_{i_1}, \cdots, Z_{i_m}\} + \text{high-order concordance on } \{Z_{i_1}, \cdots, Z_{i_m}\}\right)$$

$$\text{m-order concordance} = \sum_{k=1}^{r} \psi_k \left| Z_{i_1 k} \times Z_{i_2 k} \times \cdots \times Z_{i_m k} \right|,$$

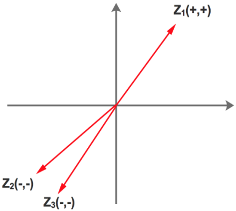
$$\psi_k = \begin{cases} 1, & \text{signs of } Z_{i_1 k}, Z_{i_2 k}, \cdots, Z_{i_m k} \text{ are the same} \\ -1, & \text{otherwise} \end{cases}$$



high-order concordance among Z_i , Z_j, and Z_k



3-order concordance > 0



3-order concordance < 0

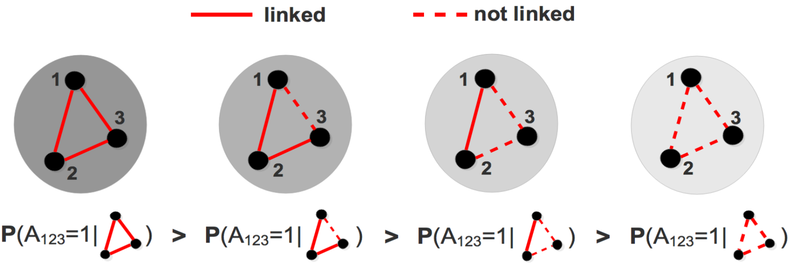
- ψ_k captures the common latent features

(*$|\cdot|$ is absolute value)

Dependency between Pairwise Link and Hyperlink

$$\mathbf{P}_{\mathbf{Z}}(\mathcal{A}_{i_1 i_2 \cdots i_m} = 1) = \mathbf{f}\Big(\underbrace{\sum_{(i,j)\in\{i_1 i_2 \cdots i_m\}} Z_i^T Z_j}_{\text{pairwise concordance on } \{Z_{i_1}, \cdots, Z_{i_m}\}} + \text{high-order concordance on } \{Z_{i_1}, \cdots, Z_{i_m}\}\Big)$$

- Density of **pairwise connections** within $\{Z_{i_1}, \cdots, Z_{i_m}\}$
- Model dependency between pairwise link and hyperlink



P(A_{123}=1|) > P(A_{123}=1|) > P(A_{123}=1|) > P(A_{123}=1|)

Embed Hyperlinks

- Represent observed hyperlink statuses on latent space by minimizing

$$Loss_{hyper}(\boldsymbol{Z};\Omega_{\mathcal{A}}) = \frac{1}{|\Omega_{\mathcal{A}}|}\sum_{(i_1 i_2\cdots i_m)\in\Omega_{\mathcal{A}}}\Big\{\underbrace{\mathcal{A}_{i_1 i_2\cdots i_m}}_{\text{observed}} - \underbrace{\mathbf{P}_{\boldsymbol{Z}}(\mathcal{A}_{i_1 i_2\cdots i_m}=1)}_{\text{predicted prob.}}\Big\}^2$$

- If $\mathcal{A}_{i_1 i_2\cdots i_m}=1$, encourage large concordance among $\{\boldsymbol{Z}_{i_1}, \boldsymbol{Z}_{i_2}\cdots \boldsymbol{Z}_{i_m}\}$
- Size of subgroup **m**: features shared by all **m** nodes
- Decompose hyperlink tensor $\{\mathbf{P}(\mathcal{A}_{i_1 i_2\cdots i_m}=1)\}$ on $\boldsymbol{Z}$

Predict Hyperlinks and Pairwise Links

- Obtain embedding representation of nodes via $\hat{\boldsymbol{Z}}$:

$$\hat{\boldsymbol{Z}} = \underset{\boldsymbol{Z}}{\operatorname{argmin}}\, Loss_{pair}(\boldsymbol{Z}; \Omega_{\boldsymbol{A}}) + Loss_{hyper}(\boldsymbol{Z}; \Omega_{\mathcal{A}}) + \lambda \|\boldsymbol{Z}\|_F^2$$

- Predict probability of a pairwise link:

$$\hat{P}(A_{ij} = 1) = f\left(\hat{\boldsymbol{Z}}_i^T \hat{\boldsymbol{Z}}_j\right)$$

- Predict probability of a m-order hyperlink:

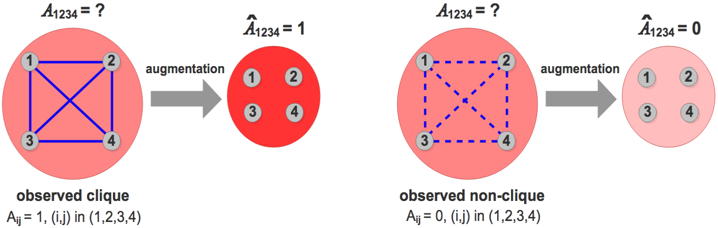
$$\hat{P}(\mathcal{A}_{i_1 i_2 \cdots i_m} = 1) = f\Big(\sum_{(i,j)\in\{i_1 i_2 \cdots i_m\}} \hat{Z}_i^T \hat{Z}_j + \sum_{k=1}^{r} \psi_k |\hat{Z}_{i_1 k} \hat{Z}_{i_2 k} \cdots \hat{Z}_{i_m k}|\Big)$$

- Joint link embedding (JLE) estimator:

$$\hat{\boldsymbol{\Theta}} = \{\hat{P}(A_{ij} = 1), \hat{P}(\mathcal{A}_{i_1 i_2 \cdots i_m} = 1)\}$$

Hyperlink Augmentation from Observed Networks

- Hyperlinks might be sparse: infer unobserved hyperlink statuses
- Observed network + **clique dependency** = augmented training data

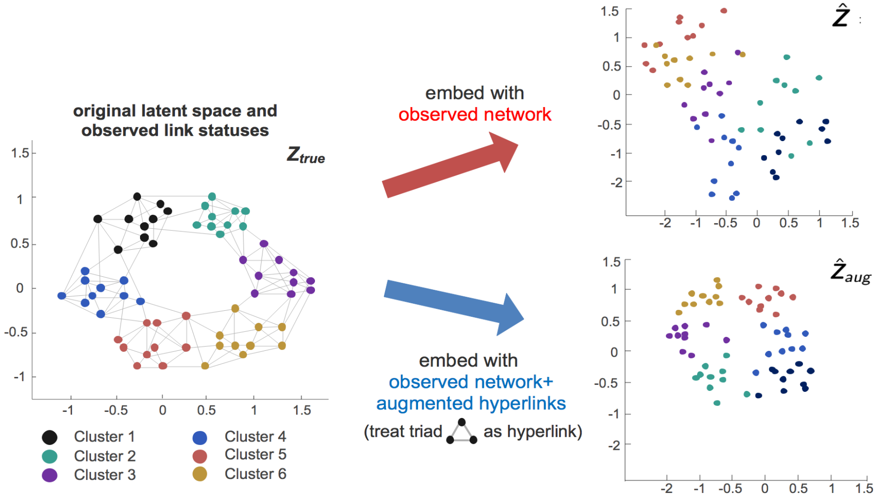


- **clique dependency**
  $P(\mathcal{A}_{i_1,\cdots,i_m} = 1 | \text{clique}, Z) - P(\mathcal{A}_{i_1,\cdots,i_m} = 1 | \text{non-clique}, Z) > 0$
- Augmented Embedding: embedding with a set of augmented hyperlink statuses $\Omega_{\hat{\mathcal{A}}}$

$$\hat{\boldsymbol{Z}}_{aug} = \underset{\boldsymbol{Z}}{\operatorname{argmin}}\, Loss(\boldsymbol{Z}; \Omega_{\mathbf{A}}, \Omega_{\mathcal{A}}, \Omega_{\hat{\mathcal{A}}})$$

Augmentation as Regularization on Embedding

- Hyperlink augmentation as **implicit regularization** on latent space
- Shrink distances among within-subgroup nodes: increase dependence between hyperlinks and cliques



Asymptotic Property of Augmented JLE

- Goal: recover $\boldsymbol{\Theta}_0 = \{P(A_{ij} = 1), P(\mathcal{A}_{i_1 i_2 \cdots i_m} = 1)\}$ via observed links
- $\rho := P(\mathcal{A}_{i_1 \cdots i_m} = 1 | \text{clique}, \boldsymbol{Z}) - P(\mathcal{A}_{i_1 \cdots i_m} = 1 | \text{no clique}, \boldsymbol{Z})$
- $|\Omega_{\hat{\mathcal{A}}}|$: number of inferred hyperlinks from hyperlink augmentation procedure

Theorem 1

Under regularity conditions, we establish the convergence rate for $\hat{\boldsymbol{\Theta}}_{aug}$:

$$P\left(\frac{\|\hat{\boldsymbol{\Theta}}_{aug} - \boldsymbol{\Theta}_0\|_F}{\sqrt{n_{N,m}}} \geq \eta\right) \leq 11 \exp\left\{-c \frac{|\Omega_{aug}|}{(1 + C_0\rho)^2} \left(1 + c_1 + c_1 \frac{\delta^2}{\epsilon^2}\right)^{-1} \eta^2\right\}$$

where $|\Omega_{aug}| = |\Omega_{\mathbf{A}}| + |\Omega_{\mathcal{A}}| + |\Omega_{\hat{\mathcal{A}}}|$, $n_{N,m} = \binom{N}{2} + \binom{N}{m}$, ϵ *is prediction MSE from Aug JLE, and* $\eta = \max\left(\epsilon, \lambda^{1/2}\right)$.

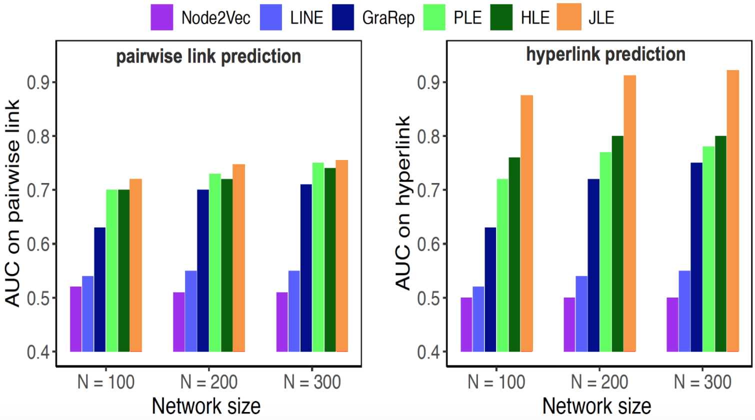
- Hyperlink augmentation accelerates convergence rate of prediction MSE ϵ:
  - pairwise links + hyperlinks + augmentation: $\epsilon \sim \frac{1+C_0\rho}{\left(|\Omega_{\mathbf{A}}|+|\Omega_{\mathcal{A}}|+|\Omega_{\hat{\mathcal{A}}}|\right)^{1/2}}$
  - pairwise links + hyperlinks: $\epsilon \sim \frac{1+C_0\rho}{(|\Omega_{\mathbf{A}}|+|\Omega_{\mathcal{A}}|)^{1/2}}$
  - pairwise links only: $\epsilon \sim \frac{1}{|\Omega_{\mathbf{A}}|^{1/2}}$

Simulation Study

- Comparing methods
  - **JLE**: the **proposed** joint pairwise and hyperlink embedding
  - **Aug JLE**: the **proposed** JLE incorporating augmented hyperlinks
  - **PLE**: embedding only through pairwise links $Loss_{pair}(\boldsymbol{Z})$
  - **HLE**: embedding only through hyperlinks $Loss_{hyper}(\boldsymbol{Z})$
  - **GraRep**: graph representations with global structural information (Cao et al., 2015)
  - **LINE**: large-scale information network embedding (Tang et al., 2015)
  - **Node2Vec** (Grover et al., 2016)
- Performance criterion
  - AUC: Area under the ROC Curve

Comparison under Conditional Independent Model

- $\{A_{ij}\}$ and $\{\mathcal{A}_{ijk}\}$ are **independent** conditioning on $\boldsymbol{Z}$
- Training on observed networks (without augmented hyperlink)
- **AUC** of link prediction on pairwise link and hyperlink testing sets

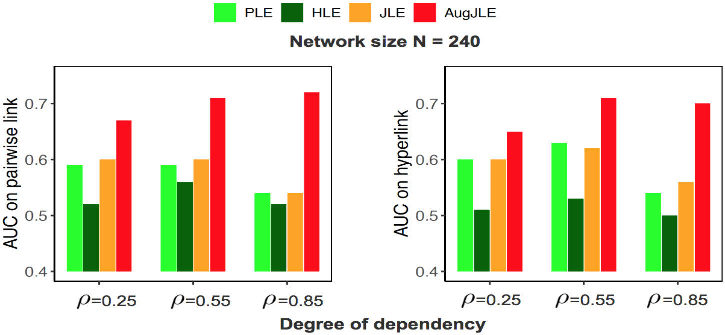


Comparison under Conditional Dependent Model (Cont.)

- Incorporate **clique structure dependency**

$$\rho = P(\mathcal{A}_{ijk} = 1 | \text{clique}, Z) - P(\mathcal{A}_{ijk} = 1 | \text{no clique}, Z) > 0$$

- $\{A_{ij}\}$ and $\{\mathcal{A}_{ijk}\}$ are **dependent** conditioning on $\boldsymbol{Z}$
- Utilize hyperlink augmentation (Aug JLE)



- Performances using observed networks are poor with large ρ
- Improvement from Aug JLE increases as ρ increases

Real Data: Facebook Ego-Network

- Ego-Network: social network among user's friends
- Social circles
  - categorization of users sharing similar features
  - **multi-way relations among users**
- Applications on detecting underlying social circles
  - online advertising
  - recommendation/content filtering
  - personalized social network organization
- Current methods: organize manually or by pre-specified attributes
  - time consuming
  - not update automatically
  - missing user profiles

Ego-Network Summary

- 224 people (N = 224)
- 6384 friendships (**two-way relations**)
- 14 social circles (**multi-way relations**)
  - defined by the ego user
  - describe different social relations: family, classmates, colleges

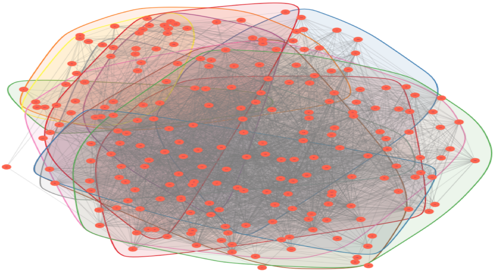


Fig. 7: The ego network in the Facebook dataset, where the social circles are marked as polygons with different colors.

Prediction Results

- Prediction
  - pairwise friendship (pairwise link)
  - whether users $\{i_1, i_2, \cdots, i_M\}(M = 6, 10)$ belong to the same circle (M-order hyperlink)

Table 1: **AUC** of link prediction for ego-network

	Link Prediction		
	Pairwise Link	**6-order Hyperlink**	**10-order Hyperlink**
Aug JLE	0.80	0.95	0.97
JLE	0.79	0.91	0.89
HLE	0.57	0.89	0.82
PLE	0.80	0.62	0.63
GraRep	0.79	0.77	0.51
LINE	0.62	0.51	0.75
Node2Vec	0.49	0.49	0.48

Concluding Remarks

- Hierarchical modeling for **multi-way** relations
  - detect subgroup structures and high-order interactions
  - capture nodes' similarities at different network levels
- Joint embedding of two-way and multi-way links
  - borrow **mutual** information for predictions
  - increase prediction performance
- Hyperlink augmentation
  - model structural dependency between pairwise and hyperlinks in latent space
  - infer potential unobserved hyperlinks
  - achieve fast convergence rate of estimation

References

- Ahmed, A., Shervashidze, N., Narayanamurthy, S., Josifovski, V., & Smola, A. J. (2013, May). Distributed large-scale natural graph factorization. In Proceedings of the 22nd international conference on World Wide Web (pp. 37-48). ACM.
- Cao, S., Lu, W., & Xu, Q. (2015, October). Grarep: Learning graph representations with global structural information. In Proceedings of the 24th ACM international on conference on information and knowledge management (pp. 891-900). ACM.
- Chang, S., Han, W., Tang, J., Qi, G. J., Aggarwal, C. C., & Huang, T. S. (2015, August). Heterogeneous network embedding via deep architectures. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 119-128). ACM.
- Dong, Y., Chawla, N. V., & Swami, A. (2017, August). metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 135-144). ACM.
- Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864). ACM.
- Handcock, M. S., Raftery, A. E., & Tantrum, J. M. (2007). Model-based clustering for social networks. Journal of the Royal Statistical Society: Series A (Statistics in Society), 170(2), 301-354.
- Hunter, D. R., Krivitsky, P. N., & Schweinberger, M. (2012). Computational statistical methods for social network models. Journal of Computational and Graphical Statistics, 21(4), 856-882.
- Ji, P., & Jin, J. (2016). Coauthorship and citation networks for statisticians. The Annals of Applied Statistics, 10(4), 1779-1812.
- Kim, B., Lee, K. H., Xue, L., & Niu, X. (2018). A review of dynamic network models with latent variables. Statistics surveys, 12, 105.
- Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 701-710). ACM.
- Qiu, J., Dong, Y., Ma, H., Li, J., Wang, K., & Tang, J. (2018, February). Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining (pp. 459-467). ACM.
- Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015, May). Line: Large-scale information network embedding. In Proceedings of the 24th international conference on world wide web (pp. 1067-1077). International World Wide Web Conferences Steering Committee.

Appendix

- Proportion of hyperlinks with high certainty: $f_{\text{clique}}(\varphi), f_{\text{non-clique}}(\varphi)$:$[0,1]\to[0,1]$

$$f_{\text{clique}}(\varphi) = \frac{|\{Y_{i_1 i_2 \cdots i_m} \in \Omega_{\text{clique}} | P(Y_{i_1 i_2 \cdots i_m} = 1 | \boldsymbol{Z}) \in [1-\varphi, 1)\}|}{|\Omega_{\text{clique}}|}$$

$$f_{\text{non-clique}}(\varphi) = \frac{|\{Y_{i_1 i_2 \cdots i_m} \in \Omega_{\text{non-clique}} | P(Y_{i_1 i_2 \cdots i_m} = 1 | \boldsymbol{Z}) \in (0, \varphi]\}|}{|\Omega_{\text{non-clique}}|}$$

- Size from augmented hyperlinks $\left|\Omega_{\hat{\mathcal{A}}}(\epsilon_{aug}, \rho)\right| = |\Omega_{\text{link}}| + |\Omega_{\text{non-link}}|$

$$|\Omega_{\text{link}}| = f_{\text{clique}}\left(\min\left\{\frac{\epsilon}{1-\rho}, 1\right\}\right)|\Omega_{\text{clique}}|$$

$$|\Omega_{\text{non-link}}| = f_{\text{non-clique}}\left(\min\left\{\frac{\epsilon}{1-\rho}, 1\right\}\right)|\Omega_{\text{non-clique}}|$$

Appendix: Hyperlink Augmentation Procedure

- Hyperlink augmentation procedure on three-order hyperlinks
- Infer $\{\mathcal{A}_{ijk}\} \in \Omega_{\mathcal{A}}^{c}$ based on $\{\{\mathcal{A}_{ijk}\} \in \Omega_{\mathcal{A}}, \{A_{ij}\} \in \Omega_{\boldsymbol{A}}\}$
- **Step 1**: construct candidate for hyperlink status: Ω_{clique} and $\Omega_{\text{non-clique}}$

$$\text{(hyperlink) } \Omega_{\text{clique}} : \{(i,j,k) | A_{ij} = A_{ik} = A_{jk} = 1\} \cap \Omega_{\mathcal{A}}^{c}$$
$$\text{(non-hyperlink) } \Omega_{\text{non-clique}} : \{(i,j,k) | A_{ij} = A_{ik} = A_{jk} = 0\} \cap \Omega_{\mathcal{A}}^{c}$$

- $|\Omega_{\text{clique}}|$ and $|\Omega_{\text{non-clique}}|$: estimable from observed network
- Use hierarchical dependency prior, not involve network model

Appendix: Hyperlink Augmentation Procedure (Continue)

- **Step 2**: select candidate hyperlink status with **high certainty**
  1. obtain $\{\hat{P}(\mathcal{A}_{i_1 i_2 i_3})\}$ via JLE based on **observed networks**
  2. construct the set of augmented hyperlink statuses as

$$\hat{\mathcal{A}}_{ijk} = \begin{cases} 1, & (i,j,k) \in \Omega_{\text{clique}} \text{ and } \hat{P}(\mathcal{A}_{ijk}) \geq 1-\delta, \\ 0, & (i,j,k) \in \Omega_{\text{non-clique}} \text{ and } \hat{P}(\mathcal{A}_{ijk}) \leq \delta. \end{cases}$$

- Augmented hyperlink $\Omega_{\hat{\mathcal{A}}} := \{\hat{\mathcal{A}}_{ijk}\}$: $|\Omega_{\hat{\mathcal{A}}}| = f(\delta, \rho, |\Omega_{\text{cliq}}|, |\Omega_{\text{non-cliq}}|)$[1]
  - larger δ (require higher inference certainty) $\Longrightarrow$ smaller $|\Omega_{\hat{\mathcal{A}}}(\delta,\rho)|$
  - larger ρ (stronger hierarchical dependency) $\Longrightarrow$ larger $|\Omega_{\hat{\mathcal{A}}}(\delta,\rho)|$ (larger $\rho \Longrightarrow$ smaller $|P(\mathcal{A}_{ijk}=1|\boldsymbol{Z}) - P(\triangle_{ijk}=1|\boldsymbol{Z})|$)
- **Step 3**: embedding with augmented hyperlinks

$$\hat{\boldsymbol{Z}}_{aug} = \underset{\boldsymbol{Z}}{\operatorname{argmin}}\, Loss(\boldsymbol{Z}; \Omega_{\mathbf{A}}, \Omega_{\mathcal{A}}, \Omega_{\hat{\mathcal{A}}})$$

[1] detailed $|\Omega_{\hat{\mathcal{A}}}|$ is provided in paper

Appendix: Model Setup for Theoretical Analysis

- Hierarchical link generation $(\mathcal{A}, \mathbf{A}) \sim \mathbf{P}_{\boldsymbol{Z}}(\mathcal{A}, \mathbf{A}) = \mathbf{P}_{\boldsymbol{Z}}(\mathcal{A}|\mathbf{A})\mathbf{P}_{\boldsymbol{Z}}(\mathbf{A})$
- Pairwise link $\mathbf{P}_{\boldsymbol{Z}}(\mathbf{A})$: $A_{ij} \sim \text{Bern}\big(P(A_{ij} = 1|\boldsymbol{Z})\big)$
- Hyperlink $\mathbf{P}_{\boldsymbol{Z}}(\mathcal{A}|\mathbf{A})$: $\mathcal{A}_{i_1 \cdots i_m} \sim \text{Bern}(P(\mathcal{A}_{i_1 \cdots i_m} = 1|\triangle, \boldsymbol{Z}))$
  - $P(\mathcal{A}_{i_1 \cdots i_m} = 1|\triangle, \boldsymbol{Z})$: dependent on $\mathbf{A}$ via clique indicator $\triangle_{i_1 \cdots i_m}$

$$\triangle_{i_1 \cdots i_m} = \begin{cases} 1, & \text{if } A_{ij} = 1,\ (i,j) \in \{i_1, \cdots, i_m\}, \\ 0, & \text{otherwise.} \end{cases}$$

  - functional form of $P(\mathcal{A}_{i_1 \cdots i_m} = 1|\triangle, \boldsymbol{Z})$ is not specified
  - degree of structural dependency

$$\rho_{i_1 \cdots i_m} := P(\mathcal{A}_{i_1 \cdots i_m} = 1|\triangle_{i_1 \cdots i_m} = 1, \boldsymbol{Z}) - P(\mathcal{A}_{i_1 \cdots i_m} = 1|\triangle_{i_1 \cdots i_m} = 0, \boldsymbol{Z})$$

  - $\rho_{i_1 \cdots i_m} = 0$: conditional independent, $\rho_{i_1 \cdots i_m} > 0$: conditional correlated

Appendix: Simulation Settings

- Hierarchical network generation
- Latent position $\boldsymbol{Z} = \{Z_i = (Z_{ir})_{r=1}^5\}_{i=1}^N$, N: number of nodes

$$Z_{ir} \sim \mu \times \text{Unif}(-1, -0.6) + (1-\mu) \times \text{Unif}(0.6, 1), \mu \sim Bern(1, 0.5)$$

- Pairwise links $\mathbf{A}$: $\{Z_i^{(\alpha)} = Z_i \odot \alpha\}_{i=1}^N$, $\alpha = (1, 1, 1, 0.2, 0.2)$

$$A_{ij} \sim \text{Bern}(\sigma(Z_i^{(\alpha)T} Z_j^{(\alpha)})),\ \sigma(\cdot) \text{ is logistic function}$$

- Hyperlinks $\mathcal{A}$: $\{Z_i^{(\beta)} = Z_i \odot \beta\}_{i=1}^N$, $\beta = (0.2, 0.2, 0.2, 1, 1)$

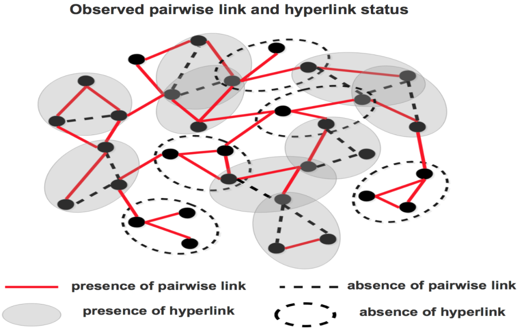
$$\mathcal{A}_{ijk} \sim \text{Bern}(\theta_{ijk}), \theta_{ijk} = \sigma\Big(\sum_{(s,t)\in\{i,j,k\}} Z_s^{(\beta)T} Z_t^{(\beta)} + \sum_{r=1}^{5} \psi_r |Z_{ir}^{(\beta)} Z_{jr}^{(\beta)} Z_{kr}^{(\beta)}|\Big)$$

- $\{A_{ij}\}$ and $\{\mathcal{A}_{ijk}\}$ are **independent** conditioning on $\boldsymbol{Z}$

($^*\odot$ denotes element-wise multiplication)

Appendix: Simulated Multi-Level Network

- Randomly split $\mathbf{A}$
  - training set ($\Omega_{\mathbf{A}}$): 60%
  - validation and testing set: 20% and 20%
- Sampling for hyperlinks: $\Omega_{\text{pool}} = \{\mathcal{A}_{ijk} | A_{ij} \in \Omega_{\mathbf{A}}, A_{ik} \in \Omega_{\mathbf{A}}, A_{jk} \in \Omega_{\mathbf{A}}\}$
  - sample 5% from Ω_{pool} as a training set $\Omega_{\mathcal{A}}$
  - split $\Omega_{\text{pool}} \cap \Omega_{\mathcal{A}}^c$ into validation (50%) and testing (50%) set



Appendix: Comparison of Conditional Dependent Model

- Incorporate **structural dependency**

$$\rho = P(\mathcal{A}_{ijk} = 1 | \text{clique}, Z) - P(\mathcal{A}_{ijk} = 1 | \text{no clique}, Z) > 0$$

- Hyperlink generating model

$$P(\mathcal{A}_{ijk} = 1 | \text{clique}) = \theta_{ijk} + \rho\{1 - \prod_{(p,q)\in\{ijk\}} \theta_{pq}\}$$

$$P(\mathcal{A}_{ijk} = 1 | \text{no clique}) = \theta_{ij} - \rho\{\prod_{(p,q)\in\{ijk\}} \theta_{pq}\}$$

- $\{A_{ij}\}$ and $\{\mathcal{A}_{ijk}\}$ are **dependent** conditioning on $\boldsymbol{Z}$

Appendix: Asymptotic Results of the Proposed Estimator

Theorem 2

Denote $n_{N,m} := \binom{N}{2} + \binom{N}{m}$ *as the number of possible pairwise and hyperlinks, under regularity conditions, we establish the convergence rate for a* **JLE** *estimator* $\hat{\mathbf{\Theta}}$ *using* **observed network**. *That is,*

$$P\left(\frac{\|\hat{\mathbf{\Theta}} - \mathbf{\Theta}_0\|_F}{\sqrt{n_{N,m}}} \geq \eta\right) \leq 11 \exp\left(-c\frac{|\Omega_{\boldsymbol{A}}| + |\Omega_{\mathcal{A}}|}{(1 + C_0\rho)^2}\eta^2\right),$$

where $\|\cdot\|_F$ *indicates the Frobenius norm,* $\eta = \max\left(\varepsilon, \lambda^{1/2}\right)$, *and* λ *is the penalty parameter,* $c \geq 0$ *is a constant,* C_0 *is the degree of link overlap, and the best possible rate is* $\varepsilon \sim \left(\frac{1+C_0\rho}{(|\Omega_{\boldsymbol{A}}|+|\Omega_{\mathcal{A}}|)^{1/2}}\right)$ *when* $\lambda \asymp \varepsilon^2$.