High-Order Joint Embedding for Multi-Level Link Prediction

Yubai Yuan and Annie Qu

Department of Statistics, University of California, Irvine, Irvine, CA

1. Introduction

Hyperlinks or hyperedges generalize traditional pairwise links through capturing interactions among groups of nodes. For example, hyperlinks occur frequently in online social networks and recommender systems such as Delicious, Last.fm and Flickr, where the system involves not only pairwise relations between users and items, but also three-way user-tag-item relations that cannot be captured by pairwise relations using the traditional network formulation.

Another example of a hyperlink is in a gene interaction network, where it is essential to identify subgroups of genes which are functionally associated with each other to potentially formulate a protein complex (Das and Yu 2012; Razick, Magklaras, and Donaldson 2008; Croft et al. 2013; Fabregat et al. 2013; Shojaaie and Michailidis 2009; Rinaldo et al. 2005). In contrast, pairwise relations between genes are inadequate to capture the collaborative high-order interactions of one gene with a subgroup of other genes. A hyperlink is called $m$-order if it is a set containing $m$ nodes. Figure 1 illustrates the differences between pairwise links and hyperlinks. Motivated by previous examples, it is critical to identify potential multi-way or high-order relations among multiple nodes represented by hyperlinks. Meanwhile, the coexistence of hyperlinks and pairwise links within a complex network system also motivates us to model and utilize their relations simultaneously for inference. In contrast, existing link prediction methods are designed for inferring either two-way or multi-way relations separately, which ignore the joint information between pairwise links and hyperlinks within the same network. For example, most existing methods utilize the information only from observed two-way relations (Hristova et al. 2016; Jallili 2017; Song et al. 2009; Al Hasan and Zaki 2011; Menon and Elkan 2011; Liben-Nowell and Kleinberg 2007; Lü and Zhou 2011; Al Hasan et al. 2006; Zhao et al. 2017; Zhang, Levina, and Zhu 2017) or decompose multi-way relations into pairwise relations (Cao, Lu, and Xu 2015; Tang et al. 2015; Grover and Leskovec 2016). In general, these methods are unable to capture high-order relations among nodes and suffer from information loss in the process of decomposing multi-way relations to two-way relations.

Alternatively, hyperlink prediction methods are proposed through directly modeling observed multi-way relations (Klamt, Haus, and Theis 2009; Li et al. 2013; Zlatić, Ghoshal, and Caldarelli 2009; Agarwal, Branson, and Belongie 2006; Zhou, Huang, and Schölkopf 2007; Zhu et al. 2016; Ghoshal 2009). One unique challenge for these methods is that hyperlinks are only partially observed or highly sparse in practice, which is quite distinct from the pairwise relations which are likely to have more observed complete information. For example, identification of a hyperlink for a protein complex connecting multiple genes requires additional biological experiments and validation, and in social networks, inferences about local social circles may require additional information involving high-order interactions among users. Therefore, inferring high-order relations with a limited training size could be unreliable without borrowing information from observed pairwise links.

Existing works on hyperlink modeling are also considered in node classification and community detection. For instance,
(Chitra and Raphael 2019; Pu and Faltings 2012; Agarwal, Branson, and Belongie 2006) suggest methods based on hyperlink expansions or random walks to reconstruct hyperlinks from pairwise links. These methods use a principle of generating hyperlinks based on prespecified relations among hyperlinks and pairwise links, while treating hyperlinks as a subgraph with a certain configuration such as a fully connected clique or star-shaped subgroup of nodes. However, this heuristic principle is not adaptive to different structures of hyperlinks, which tend to lead misspecified hyperlinks, especially in the absence of the accurate knowledge of the pairwise links and hyperlinks. In addition, they mainly focus on node classifications and detection of global community structures rather than identifying subgroup structures.

To overcome the foregoing difficulties of modeling relations with heterogeneous orders, in this article we develop a novel network embedding procedure to jointly model pairwise links and high-order hyperlinks simultaneously to capture complex interactions among nodes. In particular, we proceed in a hierarchical fashion in that pairwise and multiway relations are modeled at different resolutions. For example, pairwise relations are structured by low-level node-wise network features, while hyperlinks capture multiway relations via high-level subgroup-wise features. Jointly they can identify the subgroup configuration and capture the hyperlink-generating features effectively. This enables the incorporation of more complete and rich information from observed networks.

One advantage of this hierarchical modeling is the mutual information borrowing between hyperlink prediction and pairwise link prediction. More specifically, in the presence of a hyperlink, nodes from the same hyperlink are more likely to form a pairwise relation as compared with nodes in the absence of a hyperlink. On the other hand, nodes that are highly connected by pairwise links may suggest the presence of a potential hyperlink among them. In addition, the proposed joint embedding framework follows a principle of data augmentation for hyperlinks based on pairwise links, which improves inference accuracy as the effective sample size of links increases, and consequently captures subgroup structures associated with potential hyperlinks more efficiently. Furthermore, this principle of network connectivity also reflects the phenomenon in practice. In summary, the proposed hyperlink prediction framework provides cohesive statistical modeling for both pairwise links and hyperlinks, which enhance the mutual inferences between pairwise links and hyperlinks.

In theory, we provide the consistency of the proposed estimation and show that our method achieves a faster convergence rate compared to embedding procedures using pairwise link information only, since we are able to incorporate at least one of the observed or inferred hyperlinks via the joint embedding procedure. The results are valid in cases when either pairwise link and hyperlink are independent or dependent conditioning on the latent space. In addition, our theoretical analyses show that hyperlink augmentation can further boost the convergence rate of the joint embedding method, especially when the conditional dependency among pairwise and hyperlinks is strong. The theoretical development using large deviation theory is nontrivial here, since pairwise links and hyperlinks are correlated intrinsically in that the independent model assumption cannot be used. In contrast, most existing probability concentration properties are established under the independent model assumption.

This article is organized as follows: Section 2 introduces the background and notation of the proposed method. Section 3 introduces the proposed joint embedding method and the inference procedure from observed pairwise network to hyperlinks. Section 4 establishes the theoretical properties of the proposed embedding method under the scenarios where the hyperlinks are either observed or inferred from pairwise links. Section 5 demonstrates simulation studies, and Section 6 illustrates an application of the Facebook ego-network. The last section provides conclusions and some final discussion.

2. Background and Notations

We define an observed network $G = (V, E)$, where $V = \{v_i\}_{i=1}^N$ denotes a set of $N$ nodes and $E \subseteq \binom{V}{2}$ is a set of node pairs with observed pairwise link status, which represents the presence or absence of a pairwise link. For an undirected and unweighted network, $G$ can be represented through a symmetric binary adjacent matrix $Y = \{Y_{ij}\}_{1 \leq i \neq j \leq N}$, in that

$$Y_{ij} = \begin{cases} 1, & \{i, j\} \in E \text{ and } (i, j) \text{ is presence}, \\ 0, & \{i, j\} \in E \text{ and } (i, j) \text{ is absence}, \\ -1, & \{i, j\} \notin E. \end{cases}$$

In addition, we define an $m$-order uniform hypergraph on a set of nodes $V$ as $G_H = (V, H \subseteq \binom{V}{m})$, where $e \in H$ is an index set of $m$-tuple indices $e = \{i_1, i_2, \ldots, i_m\}$ with observed hyperlink status, indicating the presence or absence of a hyperlink. To represent the $m$-order hyperlink status, we introduce an $m$-order tensor $Y \in \{0, 1, -1\}^{Rm}$ such that

$$Y_{i_1 i_2 \ldots i_m} = \begin{cases} 1, & \{i_1, i_2, \ldots, i_m\} \in H \text{ and } \{i_1, i_2, \ldots, i_m\} \text{ is presence}, \\ 0, & \{i_1, i_2, \ldots, i_m\} \in H \text{ and } \{i_1, i_2, \ldots, i_m\} \text{ is absence}, \\ -1, & \{i_1, i_2, \ldots, i_m\} \notin H. \end{cases}$$

We introduce observation indicator $A = \{A_{ij} \in \{0, 1\}, A_{i_1 \ldots i_m} \in \{0, 1\}\}$, where $A_{ij} = 1$ if $\{i, j\} \in E$ and $A_{ij} = 0$ if $\{i, j\} \notin E$, and $A_{i_1 \ldots i_m}$ are similarly defined. Given that $A$ is independent from link statuses, the network modeling $P(Y, Y')$ can be decomposed into two submodels, where one models the underlying link statuses: $P(Y \in \{0, 1\}, Y' \in \{0, 1\})$, and the other models whether a link status is observed or not: $P(A \in \{0, 1\})$. In the following, we focus on modeling the link statuses based on observed link statuses where, $\{Y_{ij}\} \in \{0, 1\}$ and $\{Y_{i_1 \ldots i_m}\} \in \{0, 1\}$. For the observation modeling $P(A \in \{0, 1\})$, we proceed in a hierarchical fashion in that pairwise and multiway relations are modeled at different resolutions.
we follow the convention in literature (Huisman 2009) and assume the observed link statuses are uniformly and randomly sampled from all possible links on networks, that is, \{A_{ij}, A_{i1...im}\} are independent binary random variables. However, we also consider the scenario of observed link statuses with missing not at random in that dependency exists among \{A_{ij}\}, and provide the theoretical and numerical analysis in later sections.

Denote the sets of observed pairwise link statuses and hyperlink statuses as \(\Omega_1\) and \(\Omega_2\). Accordingly, the number of observed pairwise link statuses and hyperlink statuses are \(|\Omega_1|\) and \(|\Omega_2|\). The proposed link prediction framework is based on network latent space modeling. Specifically, we introduce an \(r\)-dimensional latent vector representation \(Z_i = (Z_{i1}, \ldots, Z_{ir})\) for each node \(v_i\), where each element \(Z_{ik}, k = 1, \ldots, r\) represents a latent feature of node \(v_i\). The latent factor modeling enables us to capture the relation between nodes \(v_i\) and \(v_j\) on a network through their concordance on the latent space spanned by \(\{Z_i\}_{i=1}^N\). In addition, the latent factor model allows us to capture the low-rank structure feature by assuming that the dimension of the latent space \(r\) is much smaller than \(N\).

Through a dimension reduction we are able to increase the estimation efficiency of latent factors while reducing the variation of prediction. Based on the node-wise latent factors \(\{Z_i\}_{i=1}^N\), we propose a probabilistic link modeling. That is, we assume that both pairwise link \(Y_{ij}\) and hyperlink \(Y_{i1...im}\) follow a Bernoulli distribution with probability \(P(Y_{ij}|Z_i, Z_j)\) and \(P(Y_{i1...im}|Z_{i1}, \ldots, Z_{im})\), respectively. Therefore, once the positions of nodes in the latent space are estimated, we can predict the probabilities for both pairwise links and hyperlinks among these nodes.

3. Methodology

In this section, we introduce the proposed joint network embedding framework to perform prediction for both pairwise links and hyperlinks. The key idea is to project the nodes in the network into points in a latent space, and preserve both the observed pairwise and high-order relations in the latent space. In the following, we present the methods for embedding pairwise links and hyperlinks.

3.1. Reconstruction of Pairwise Relations in Latent Space

To incorporate observed pairwise relations into a latent space, we propose to minimize a loss function to estimate node-wise latent factors \(Z = \{Z_i \in \mathbb{R}^r\}_{i=1}^N\):

\[
L(Z) = \frac{1}{|\Omega_1|} \sum_{Y_{ij} \in \Omega_1} (Y_{ij} - \sigma(Z_i^T Z_j))^2,
\]

where \(|\Omega_1|\) is the number of observed pairwise links, and \(Z_i^T Z_j\) is the second-order concordance, measuring the degree of similarity between two nodes' latent features. The \(\sigma(\cdot)\) is a logistic link function that transforms the second-order concordance into the probability of a pairwise link. Intuitively, two latent factors \(Z_i\) and \(Z_j\) are expected to be close in the latent space given that a pairwise link exists between nodes \(v_i\) and \(v_j\), that is, \(Y_{ij} = 1\). In contrast, two latent factors \(Z_i\) and \(Z_j\) are far away from each other given \(Y_{ij} = 0\). In this way, we transform the binary link statuses of the network into continuous measures of distance in the latent space.

The probabilistic prediction loss in (1) has the same form as the Brier score (Brier 1950), for example, \((Y - \hat{P})^2\) where \(Y\) is a binary outcome and \(\hat{P}\) is the corresponding predicted probability. Compared with the negative log-likelihood loss, for example, \(-\log(\hat{P} + (1 - Y)\log(1 - \hat{P}))\), both of them have a similar U-shape landscape and reach the same minimizer of \(\hat{P} = P(Y = 1)\) at the population level. In addition, an empirical study (Menon and Elkan 2011) demonstrates that the choice of loss function has minor influence on link prediction performance using latent factor models.

3.2. Reconstruction of High-Order Relations in Latent Space

Hyperlink prediction is more challenging since the configurations of high-order relation involve a high degree of uncertainty and require high-dimensional modeling tools. One key element of the proposed method is to model hyperlinks via the structure of latent factors expressed in high-order tensors. We formulate the \(m\)-order hyperlinks by an \(m\)-order tensor \(Y = (Y_{i1...im}) \in \{0, 1\}^{N^m}\). Specifically, if there is a hyperlink connecting \(m\) nodes \((v_{i1}, \ldots, v_{im})\), then \(Y_{i1...im} = 1\). To map the high-order relations of the network into the latent space, we model the \(m\)-order hyperlinks as a low-rank concordance structure on the latent feature \(Z\) based on the CANDECOMP/PARAFAC (CP) tensor decomposition. Specifically, we introduce a high-order concordance measure function among \(m\) nodes as

\[
f(Z_{i1}, Z_{i2}, \ldots, Z_{im}) = \sum_{k=1}^{r} \psi_{k,i1,i2,...,im}|Z_{i1}Z_{i2} \cdots Z_{im}|^k,\]

where

\[
\psi_{k,i1,i2,...,im} = \begin{cases} 1 & Z_{ik} \geq 0, l = 1, \ldots, m \text{ or } Z_{ik} < 0, l = 1, \ldots, m, \\ -1 & \text{otherwise}. \end{cases}
\]

The high-order concordance measures the degree of similarity among a group of nodes. Here we constrain the scope of concordance among multiple latent features only when their signs are all the same, thereby increasing the power for discriminating between high-order concordance and high-order discordance. Based on the high-order concordance among the latent factors of nodes \(v_{i1}, v_{i2}, \ldots, v_{im}\), we model the probability of hyperlink \(Y_{i1...im}\) as:

\[
P(Y_{i1...im} = 1) = \sigma(\sum_{(i,j)\in\{i1...im\}} Z_{ij}^T Z_j + \beta f(Z_{i1}, Z_{i2}, \ldots, Z_{im})),\]

where \(\beta > 1\) is a weighting parameter. Analogous to measuring the pairwise concordance via an inner product among latent factors, we apply the generalized inner product \(\sum_{k=1}^{\gamma} \psi_k|Z_{i1}Z_{i2} \cdots Z_{im}|^k\) to measure the joint concordance among a group of \(m\) latent factors. Notice that the proposed measure \(f(Z_{i1}, Z_{i2}, \ldots, Z_{im})\) is equivalent to the CP tensor decomposition given that \(\psi_k = 1, k = 1, \ldots, r\). The purpose of introducing sign consistency function \(\psi_k\) for each latent factor dimension \(k = 1, \ldots, r\) is to adjust the sign of multiplication among latent factors so that the sign is consistent with the interpretation of \(f(Z_{i1}, Z_{i2}, \ldots, Z_{im})\). For example, consider
the kth latent feature for three nodes $Z_{i1k} > 0, Z_{i2k} < 0, Z_{i3k} < 0$. Since the signs of the latent features are not consistent, the $P(Y_{i1j2} = 1)$ decreases given other latent features. However, applying CP decomposition would lead to $Z_{i1k}Z_{i2k}Z_{i3k} > 0$, which increases the $P(Y_{i1j2} = 1)$. Therefore, the introduction of $\psi_k, k = 1, \ldots, r$ encourages a large value of $f(Z_{i1}, Z_{i2}, \ldots, Z_{im})$ only when all $Z_{i1}, Z_{i2}, \ldots, Z_{im}$ are close to each other in the latent space, which is consistent with the interpretation of the high-order concordance. Note that the joint concordance cannot be directly inferred by the pairwise concordance in the sense that even $Z_{i1}Z_{i2}$ is large for any pair $(i,j) \in \{i_1, i_2, \ldots, i_m\}$, though it is still possible that $f(Z_{i1}, Z_{i2}, \ldots, Z_{im})$ is small. This also implies that the joint concordance capturing high-order relations cannot be substituted for by pairwise concordance.

The latent-factor-based hyperlink modeling in (3) introduces a hierarchical dependency between pairwise links and hyperlinks since (3) also incorporates the pairwise concordance among nodes in (1). In addition, the high-order concordance is correlated with $\sum_{(i,j) \in \{i_1, i_2, \ldots, i_m\}} Z_{i1}^T Z_{i2}$ via the shared latent factors. Through utilizing the dependency between hyperlink and pairwise link, we borrow information from each for more accurate prediction. In addition, the proposed hierarchical dependency facilitates better interpretations in real applications as compared to the methods of using pairwise links or hyperlinks separately. Subsequently, we incorporate the hyperlink information into the latent space of nodes via obtaining the latent factors $Z$ via minimizing the following hyperlink loss function:

$$L_2(Z) = \frac{1}{|\Omega_Y|} \sum_{Y_{i12} = \ldots = Y_{im} \in \Omega_Y} w_{i12} \ldots \sum_{Y_{im}} \left( Y_{i12} \ldots \sum_{Y_{im}} w_{i12} \ldots \sum_{Y_{im}} \right)$$

$$- \sum_{(i,j) \in \{i_1, i_2, \ldots, i_m\}} \left( Z_{i1}^T Z_{i2} + \beta f(Z_{i1}, Z_{i2}, \ldots, Z_{im}) \right)^2,$$

(4)

where $w_{i12} \ldots \sum_{Y_{im}}$ is the weight for hyperlink $Y_{i12} = \ldots = Y_{im}$, $\Omega_Y$ is the set of incorporated $m$-order hyperlinks, and $|\Omega_Y|$ is the total number of hyperlinks. If there exists a potential hyperlink connecting nodes $\{i_1, i_2, \ldots, i_m\}$ in a hypergraph, then decreasing the loss function in (4) leads to a strong joint concordance among latent factors $\{Z_{i1}, Z_{i2}, \ldots, Z_{im}\}$. Consequently, we preserve high-order relations among nodes associated with their embedding in latent space. In terms of latent factors estimation, (1) and (4) serve as the second-order moments and the $m$-order moment estimations of $Z$, respectively. Intuitively, incorporating additional moment information generated from latent factors $Z$ reduces estimation bias while increasing efficiency, which leads to a more accurate estimation of latent factor $Z$.

### 3.3. Joint Network Embedding for Pairwise Link and Hyperlink Prediction

Based on the link embedding (1) and (4), we estimate the latent features of nodes by jointly incorporating pairwise links and hyperlinks. Due to randomness or noisy sources of information, some pairwise links and hyperlinks can be incorrectly observed or inferred, which requires regularization during the embedding process. In addition, in a network system with many potential hyperlinks, we are interested in detecting more significant hyperlinks in the sense that they capture local subgroup structures with high certainty in the network. To achieve these goals, we propose the following loss function:

$$L(Z) = \frac{1}{|\Omega_Y|} \sum_{Y_{i1} = \ldots = Y_{im} \in \Omega_Y} w_{i1} \ldots \sum_{Y_{im}} \left( Y_{i1} \ldots \sum_{Y_{im}} w_{i1} \ldots \sum_{Y_{im}} \right)$$

$$- \sum_{(i,j) \in \{i_1, i_2, \ldots, i_m\}} \left( Z_{i1}^T Z_{i2} + \beta f(Z_{i1}, Z_{i2}, \ldots, Z_{im}) \right)^2$$

$$+ \frac{1}{|\Omega_Y|} \sum_{Y_{ij} \in \Omega_Y} (Y_{ij} - \sigma(z_{ij}^T z_j)^2 + \lambda ||Z||^2).$$

Specifically, we impose the weight function $w_{i12} \ldots \sum_{Y_{im}}$ to down-weight spurious links through penalization. We also adopt the penalization $||Z||^2$ to control the magnitude of the latent factors of nodes in order to alleviate the overfitting of latent factors due to spurious pairwise links and hyperlinks. Furthermore, the penalization $||Z||^2$ imposes a low-rank structure of latent factors which can mitigate the singularity problem when the degree of nodes is smaller than the rank of latent factors. To tune the penalty parameter $\lambda$ in (5), we first split both the observed pairwise link status and hyperlink status into a training and validation dataset. Then we fit the loss function (5) on the training dataset and perform a grid search for $\lambda$ on a predefined region, for example, $[0, 1]$. We can select $\lambda$ to maximize the AUC on the validation dataset.

After mapping each node into the latent space spanned by column vectors of $Z$ consisting of the pairwise links and hyperlinks, we predict potential pairwise links and hyperlinks through an estimated degree of concordance among the latent factors of nodes. Specifically, we predict a pairwise link between nodes $v_i$ and $v_j$ through

$$P(Y_{ij} = 1 | (v_i, v_j)) = \sigma(Z_{i1}^T Z_{j1}).$$

(6)

Similarly, to predict an $m$-order hyperlink among a group of nodes $v_1, v_2, \ldots, v_m$, we have

$$P(Y_{i12} = \ldots = Y_{im} = 1 | (v_1, v_2, \ldots, v_m)) = \sigma \left( \sum_{(i,j) \in \{i_1, i_2, \ldots, i_m\}} Z_{i1}^T Z_{j1} + \beta f(Z_{i1}, Z_{i2}, \ldots, Z_{im}) \right).$$

(7)

Although the main focus in this article is link prediction, identifying the latent factor space of nodes is also fundamentally important as it permits exploration of other types of network structures. For example, in community detection, detection of the community structures of a network leads to discoveries in biology, marketing, and social science. In other situations, we may develop a clustering algorithm to identify homogeneous subgroups of nodes in a latent space based on their estimated similar embeddings of learned nodes.

**Remark.** The proposed joint embedding procedure can be generalized to the case when orders of hyperlink statuses are different. Instead of directly encoding hyperlink statuses via a series of tensors with different orders, an alternative solution is to decompose hyperlink statuses of different orders into hyperlink statuses with the same and low order, which could be then
formulated into a single hyperlink tensor. In this way, an original high-order hyperlink statuses is approximated by a set of low-order hyperlink statuses, and then the objective function (5) can be directly applied.

3.4. Inferring Potential Hyperlinks through Observed Links

An innovation of the proposed method is that the framework enables a data augmentation procedure on hyperlinks based on the observed network. Although in many applications, only a small number of high-order relations represented by hyperlinks can be directly observed, it is feasible to infer potential hyperlinks from observed links through link dependency information.

The rationale is that pairwise links and hyperlinks characterize the same types of relations within a network but at different group levels. Therefore, observing whether two nodes are bilaterally connected or not is informative for inferring whether they belong to the same subgroup or hyperlink. This hierarchical dependency among links is implicitly introduced into the proposed framework as the node-wise latent factors and their concordances are shared by both pairwise link (1) and hyperlink modeling (3), which enables us to infer the unobserved hyperlinks along with estimating latent factors. The inferred hyperlinks can be integrated into the embedding procedure to increase the effective sample size and enrich the subgroup structures in the latent space.

In the following, we illustrate an inference procedure example for a three-order hyperlink, which can be similarly generalized to the high-order hyperlinks following a similar procedure. Intuitively, we tend to infer hyperlink structures whose underlying probability is either close to 1 or close to 0 due to their reduced inference uncertainty. Based on the hyperlink modeling in (3), the inference of a hyperlink status follows the three steps:

Step 1: Embedding with observed network: We first obtain node-wise latent factors based on the observed pairwise links \( \Omega_Y \) and hyperlinks \( \Omega_Y \) through

\[
Z^{obs} = \arg \min_L L(Z, \Omega_Y, \Omega_Y), \tag{8}
\]

where \( L(\cdot) \) is the joint embedding loss defined in (5). The \( Z^{obs} \) and encodes the position of each node and its geometric relation in the latent space, which is used by the downstream hyperlink inference. The penalty term \( ||Z||^2 \) is not considered, that is, \( \lambda = 0 \) in step 1 for fully utilizing the observed network.

Step 2: Construct candidate for augmented hyperlink statuses: Given the prior knowledge such that the probability of establishing a hyperlink among a set of nodes is higher when there are more observed pairwise links among these nodes compared to those with fewer observed pairwise links, we construct a candidate pool for augmented hyperlink statuses based on the hierarchical dependence structure and observed pairwise link information. Specifically, we introduce \( \Omega^{pair}_1 = \{(i,j,k) | Y_{ik} = Y_{jk} = 1\} \cap \Omega_Y \) and \( \Omega^{pair}_2 = \{(i,j,k) | Y_{ig} = Y_{ik} = Y_{jk} = 0\} \cap \Omega_Y \) to indicate the candidate pool for augmented hyperlinks and nonhyperlinks, respectively, where \( \Omega_Y \) is the observed hyperlink statuses. We denote the total candidate pool as \( \Omega_{pool} := \Omega^{pair}_1 \cup \Omega^{pair}_2 \).

Step 3: Select informative augmented hyperlink statuses: We select augmented hyperlink statuses with low inference uncertainty from \( \Omega_{pool} \). Based on \( Z^{obs} \) we estimate the hyperlink probability \( \hat{P}_{ijk} = P(Y_{ijk}) \) via (3) for \( Y_{ijk} \in \Omega_{pool} \). Note that \( \hat{P}_{ijk} \) only utilizes the information from observed network, which is an independent information source from the hierarchical dependency prior in Step 2. We construct the set of augmented hyperlink statuses \( \hat{\Omega}_Y := \{\hat{Y}_{ijk}\} \) as follows:

\[
\hat{Y}_{ijk} = \begin{cases}
1, & (i,j,k) \in \Omega^{pair}_1 \text{ and } \hat{P}_{ijk} \geq 1 - \delta, \\
0, & (i,j,k) \in \Omega^{pair}_2 \text{ and } \hat{P}_{ijk} \leq \delta,
\end{cases}
\]

where \( \delta \in (0,0.5) \) is a cutoff that can be tuned on a validation set. Finally, we refine the node-wise latent factors \( Z^{obs} \) through incorporating the set of augmented hyperlink statuses \( \hat{\Omega}_Y \) into the joint embedding loss function in (5) as \( Z^{aug} = \arg \min_L L(Z, \Omega_Y, \Omega_Y, \Omega_Y) \).

The effectiveness of the augmentation procedure relies on the existence of dependency between pairwise links and hyperlinks on the same set of nodes, and the improvement from link augmentation becomes more significant as the degree of dependency increases. Given the dependency exists and it is moderate, we can capture it via modeling the pairwise links in (1) and hyperlinks in (4) on shared latent factors. When the dependency is stronger in that two types of links are still correlated conditioning on the latent positions, then the augmentation procedure can incorporate the conditional dependency via imposing subgroup structure on the latent positions; and introducing structured regularization on the parameter space of latent factors can enhance the estimation efficiency.

3.5. Embeddings Estimation

We embed each node into a latent feature \( Z \) estimated by minimizing the joint loss function in (5). In contrast to existing pairwise link or hyperlink embedding approaches involving matrix factorization (Menon and Elkan 2011; Qiu et al. 2018; Ahmed et al. 2013; Belkin and Niyogi 2002; Cao, Lu, and Xu 2015; Ou et al. 2016; Zhou, Huang, and Schölkopf 2007; Zhu et al. 2016) or tensor decomposition, which could entail high computational cost especially when the order of tensor \( m \) is high. Notice that the hyperlink tensor \( Y \) is super-symmetric in that \( Y_{ijkl} = Y_{jikl} = ... = Y_{ijlk} \), where \( \varphi \) is the order permutation mapping. On this ground, we develop a scalable algorithm to minimize (5) while taking advantage of the super-symmetry of a hyperlink tensor to reduce the computational cost.

In general, we estimate the embedding of nodes \( Z \) through the coordinate gradient descent algorithm where both the gradient and Hessian matrix have explicit forms. Alternatively, we can replace the gradient update by the first-order gradient descent with an adaptive learning rate strategy such as ADAM (Kingma and Ba 2014) to accelerate the convergence and avoid local minima. The detailed estimation procedure is summarized as Gradient Descent Algorithm with Parallel Computing in the supplemental material. One advantage of the proposed algorithm is that it uses the super-symmetry property of a
hyperlink tensor to update the latent vectors corresponding to
different tensor modes simultaneously instead of updating each
node modally. In addition, updating gradients of the node-
wise latent vectors can be performed independently of each
other, which makes it feasible for parallelization to accelerate
computation.

4. Theoretical Results

In this section, we investigate the theoretical properties for
the proposed joint embedding method. Specifically, we focus
on establishing the asymptotic properties of link probability
estimation as it directly associates with prediction accuracy. We
consider a class of link-generating processes where the pairwise
link statuses $Y = \{Y_{ij}\}$ and $m$-order hyperlink statuses $Y' = \{Y_{ij_1j_2 \ldots j_m}\}$ can be hierarchically dependent conditioning on the
latent position $Z$.

Given the node-wise latent factors $Z = \{Z_i\}_{i=1}^N$, we decom-
pose the joint distribution as $P(Y, Y|Z) = P(Y|Y, Z)P(Y|Z)$, and assume that the pairwise links $(P(Y_{ij}|Z))$ independently follow

$$Y_{ij} \sim \text{Bern}(\theta_{ij}), \text{ where } \theta_{ij} = \frac{\exp(Z_iZ_j^T)}{1 + \exp(Z_iZ_j^T)}; \quad 1 \leq i < j \leq N.$$ (9)

For the generation of an $m$-order hyperlink status $Y_{ij_1j_2 \ldots j_m}$, we introduce a clique indicator variable $\Delta_{ij_1j_2 \ldots j_m}$ as

$$\Delta_{ij_1j_2 \ldots j_m} = \begin{cases} 1, \text{ if } Y_{ij} = 1, (i, j) \in \{i_1, i_2, \ldots, i_m\}, \\ 0, \text{ otherwise}. \end{cases}$$ (10)

In other words, $\Delta_{ij_1j_2 \ldots j_m}$ captures whether all nodes $i_1, i_2, \ldots, i_m$ are pairwisely connected. We assume that $Y_{i_1i_2 \ldots i_m}$ generates from a Bernoulli distribution as

$$Y_{i_1i_2 \ldots i_m} \sim P(Y_{i_1i_2 \ldots i_m}|Y, Z) = \text{Bern}(P(Y_{i_1i_2 \ldots i_m}|1 \Delta_{i_1i_2 \ldots i_m}|Z)),$$

$$1 \leq i_1 < \cdots < i_m \leq N,$$ (11)

where the probability relies on the clique indicator $\Delta_{i_1i_2 \ldots i_m}$, therefore, introduces an additional hierarchical dependency
between the pairwise links and hyperlinks in addition to sharing the same set of latent position $Z$. We provide an explicit formulation of $P(Y_{i_1i_2 \ldots i_m} = 1|\Delta_{i_1i_2 \ldots i_m}|Z)$ in the
supplementary. We model the marginal distribution $P(Y|Z) = \{P(Y_{i_1i_2 \ldots i_m}|Z)\}$ by Bernoulli distribution with probability

$$\theta_{ij_1j_2 \ldots j_m} = P(Y_{ij_1j_2 \ldots j_m} = 1|Z) = \sigma[\sum_{(i,j) \in \{i_1, i_2, \ldots, i_m\}} Z_iZ_j^T + \beta f(Z_{i_1}, Z_{i_2} \ldots Z_{i_m})],$$ (12)

which has the same formulation as (3) in Section 3.2. Therefore, given pairwise links and hyperlinks follows the joint distribution $P(Y', Y|Z)$ as (9) and (11), the joint link embedding
method (JLE) objective function combining (9) and (12) serves as the composite likelihood for $P(Y'|Z)P(Y|Z)$ as a surrogate of
$P(Y', Y|Z)$ (Lindsay 1988; Varin, Reid, and Firth 2011).

It is natural to consider a nonnegative dependency in (11) in
that the probability of $Y_{ij_1i_2 \ldots i_m} = 1$ is higher given that a clique
exists among nodes $\{i_1, \ldots, i_m\}$, and the degree of dependency can be quantified as

$$\rho_{ij_1i_2 \ldots i_m} := P(Y_{ij_1i_2 \ldots i_m} = 1|\Delta_{i_1i_2 \ldots i_m} = 1, Z) - P(Y_{ij_1i_2 \ldots i_m} = 1|\Delta_{i_1i_2 \ldots i_m} = 0, Z),$$ (13)

where $0 \leq \rho_{ij_1i_2 \ldots i_m} \leq 1$. In practice, a positive dependency
between a hyperlink $Y_{ij_1i_2 \ldots i_m}$ and the corresponding clique $\Delta_{i_1i_2 \ldots i_m}$ in (11) is extensively observed within complex
networks. For example, the existence of a protein complex is
highly associated with a densely connected subgraph within
protein–protein interaction networks (Xu et al. 2013; Ma et al.
2017). In addition, a high-order social interaction is generally
built upon pairwise interactions in social networks (Cencetti
et al. 2021; Alvarez-Rodriguez et al. 2021).

In addition, instead of assuming the link statuses are ran-
domly observed within the pairwise link network, we allow that
the observation dependency exists and is formulated as

$$\rho_{\text{obs}} := \text{corr}\left\{I(Y_{ij}, Y_{ik} \text{ are observed}) (Y_{ij}, Y_{ik}), I(Y_{jk} \text{ is observed}) (Y_{jk})\right\}.$$ Specifically, we consider the case when $\rho_{\text{obs}} \geq 0$, and $\rho_{\text{obs}}$
measures an increase in the probability of $Y_{jk}$ being observed
compared to being randomly observed given $Y_{ij}$ and $Y_{ik}$ are
observed. In many applications, the observation of networks
is missing at random in that the probability of a link's
missingness is related to its neighborhood links connecting to
the same nodes (Huisinga 2009). Therefore, $\rho_{\text{obs}}$ quantifies the
missing pattern information from the pairwise links.

Given the parameter set $\Theta = \{\theta_{ij}, \theta_{ij_1i_2 \ldots i_m}; \ 1 \leq i < j \leq N, 1 \leq i_1 < \cdots < i_m \leq N\} \subseteq \mathbb{R}^{N^2 \times N} \cup \mathbb{R}^{N_m}$,
the link prediction accuracy of the proposed method can be
established through investigating the convergence property of
the $\Theta$ estimator. In the following, we establish the consistency
and convergence rate of the $\Theta$ estimator that minimizes the
proposed joint embedding loss function:

$$l_{\text{joint}}(\Theta(Z); Y, Y') = l_{\text{pair}}(\Theta(Z); Y) + l_{\text{hyper}}(\Theta(Z); Y) + \lambda ||Z||^2,$$ (14)

where $l_{\text{pair}}(\Theta; Y)$ and $l_{\text{hyper}}(\Theta; Y')$ are the loss functions in (1) and (4) representing the pairwise link and hyperlink embed-
ding, respectively. Because of the nonconvex nature of the loss
function (14) with respect to $\Theta$, obtaining the global minimizer
of (14) is generally infeasible or computationally expensive in
practice. Instead, we establish the convergence property for a
class of alternative estimators $\hat{\Theta}$ satisfying

$$l_{\text{joint}}(\hat{\Theta}(Y, Y')) \leq \inf_{\Theta \in \mathcal{S}} l_{\text{joint}}(\Theta; Y, Y') + \tau,$$ (15)

where $\tau$ goes to zero as the number of observed links increases.
We show that the alternative estimator $\hat{\Theta}$ in (15) is still able to
achieve the convergence rate of the global minimizer of (14) as
the sample size increases. We introduce the following regularity
assumption:

(C1): The node-wise latent factors are uniformly bounded such
that $||Z||_\infty \leq C$ for some positive constant $C$, where $|| \cdot ||_\infty$ indicates the infinity norm. The weighting parameter $\beta$
is bounded such that $1 \leq \beta \leq C'$ for some positive constant $C'$.
Remark 4.1. Assumption (C1) requires that the underlying search space for the latent factors is bounded, hence, the parameter space $\mathcal{S}$ is also uniformly bounded. This is a standard assumption for latent factor modeling.

In the following, we establish the consistency and convergence rates for the estimator $\hat{\Theta}$ from the proposed JLE when the pairwise links and hyperlinks are hierarchically dependent on each other given their latent information. We first introduce the following quantity:

$$
C_0 := \max_{\{i,j\} \in \Omega_Y} \left\{ \sum_{\{i_1,i_2,\ldots,i_m\} \in \Omega_m} \mathbf{1}_{\{i_1,i_2,\ldots,i_m\}}((i,j)) \right\},
$$

where $\mathbf{1}_{A}(e)$ is $1$ if $e \subseteq A$ and $\mathbf{1}_{A}(e) = 0$ otherwise. Intuitively, $C_0$ measures the largest number of observed hyperlinks such that these hyperlinks overlap in the same pairwise link, and therefore, measures the degree of overlapping between the observed pairwise network and the hyperlink network. Notice that $0 \leq C_0 \leq \min(\binom{m-1}{2}, |\Omega_Y|)$ as $C_0 \leq \frac{|\Omega_Y|}{\binom{m-1}{2}}$, given that the observed hyperlink statuses are sampled uniformly from all the $C_N^m$ possible tuples of nodes, where $\simeq$ denotes the same order. Without loss of generality, we assume that $\{\rho_{i_1,i_2,\ldots,i_m}\}$ are all equal and denoted as $\rho$. Denote $n_{N,m} := C_N^m + C_N^m$. In the following, we establish the consistency and convergence rate for the JLE using the observed network.

Theorem 4.1. Denote $\Theta_0$ as the underlying true link probabilities, and $n_{N,m} := C_N^m + C_N^m$. Under the assumption (C1) and $\tau = O(\epsilon^2)$, we establish the convergence rate for a JLE estimator $\hat{\Theta}$. That is,

$$
P\left( \frac{||\hat{\Theta} - \Theta_0||_F}{\sqrt{n_{N,m}}} \geq \eta \right) \leq \exp\left(-n_{N,m} \left( \frac{1}{\rho} \right)^2 \right),
$$

where $|| \cdot ||_F$ indicates the Frobenius norm, $c \geq 0$ is a constant, $\eta = \max(\epsilon, \lambda^{1/2})$, $C_0$ is the degree of overlap links in (16), and the best possible rate $\epsilon \sim \left( \frac{1+C_0\rho}{(1+C_0\rho)^{1+1/2}} \right)$ is achieved when $\lambda \simeq \epsilon^2$.

Theorem 4.1 provides the convergence property for the JLE under a broad class of link-generating models which include both scenarios when the hyperlink statuses and pairwise link statuses are independent or hierarchically dependent given the latent position $Z$. Specifically, Theorem 4.1 states that given the magnitude of penalty term $||Z||$ in (14) shrinking to zero with an appropriate rate, the proposed estimator for $\Theta$ can achieve the convergence rate at $\frac{1+C_0\rho}{(1+C_0\rho)^{1+1/2}}$. In addition, Theorem 4.1 shows that the JLE estimator achieves a faster convergence rate compared to methods using either pairwise link statuses $\Omega_Y$ only pairwise link embedding (PLE), or hyperlink statuses $\Omega_Y$ only hyperlink embedding (HLE) with the corresponding rate of $\frac{1}{(1+\frac{1+C_0\rho}{\rho})^{1+1/2}}$ or $\frac{1}{(1+\frac{1}{\rho})^{1+1/2}}$, when the observed pairwise and hyperlinks are conditionally independent, that is, $\rho = 0$ or $C_0 = 0$. When these two types of links are hierarchically dependent, that is, $\rho > 0$, the convergence rates of PLE and HLE are $\frac{1}{(1+\frac{1+C_0\rho}{\rho})^{1+1/2}}$ and $\frac{1+C_0\rho}{(1+C_0\rho)^{1+1/2}}$, respectively. Therefore, the JLE estimator still converges faster than HLE and PLE when $\frac{\Omega_Y}{\Omega_Y} \geq 2C_0\rho + C_0^2\rho^2$. We provide discussion on the comparison between existing convergence rate of low-rank matrix completion and Theorem 4.1 in Appendix.

In the following Theorem 4.2, we further develop the convergence property of the proposed joint embedding estimator incorporating augmented hyperlink statuses. We first define the candidate pool of augmented hyperlink statuses $\Omega_{\text{clique}} \cup \Omega_{\text{non-clique}}$ as

$$
\Omega_{\text{clique}} := \{Y_{i_1i_2,\ldots,i_m} | Y_{ij} = 1, \{i,j\} \subset \{i_1,i_2,\ldots,i_m\} \} \cap \Omega_{\text{pool}},
$$

$$
\Omega_{\text{non-clique}} := \{Y_{i_1i_2,\ldots,i_m} | Y_{ij} = 0, \{i,j\} \subset \{i_1,i_2,\ldots,i_m\} \} \cap \Omega_{\text{pool}},
$$

where $\Omega_{\text{pool}}$ representing the set $\{Y_{i_1i_2,\ldots,i_m} | Y_{ij} \text{ is observed}, \{i,j\} \subset \{i_1,i_2,\ldots,i_m\} \}$. In other words, we infer the potential hyperlink statuses on the node tuple $\{i_1,i_2,\ldots,i_m\}$, where the pairwise link statuses among $\{i_1,i_2,\ldots,i_m\}$ are observed, and are either all linked together or not linked at all. Given the set of observed pairwise link statuses $\Omega_Y$, the size of $\Omega_{\text{pool}}$ is determined by the observation dependency $\rho_{\text{obs}}$ among pairwise link statuses as

$$
|\Omega_{\text{pool}}(\rho_{\text{obs}})| = \Omega_Y^{m-1} (1-\rho_{\text{obs}})^{(m-1)\epsilon}.
$$

where $\epsilon = \frac{2}{|\Omega_Y|} \in (0,1)$ is the observation density in the pairwise link network. To ensure $|\Omega_{\text{pool}}(\rho_{\text{obs}})| > 1$, we consider the case where the order of hyperlink is bounded in that $m = O(\log N/\log \kappa^{-1})$. Therefore, the size of the candidate pool of hyperlink statuses for augmentation $|\Omega_{\text{clique}} \cup \Omega_{\text{non-clique}}|$ increases as $\rho_{\text{obs}}$ becomes larger due to more cliques $Y_{i_1i_2,\ldots,i_m} = 1$, $\{i,j\} \subset \{i_1,i_2,\ldots,i_m\}$ or noncliques $Y_{i_1i_2,\ldots,i_m} = 0$, $\{i,j\} \subset \{i_1,i_2,\ldots,i_m\}$ being observed with a larger $\rho_{\text{obs}}$.

To quantify the number of augmented hyperlink statuses within the candidate pool, we introduce two monotone increasing distribution functions $f_{\text{clique}}(\varphi), f_{\text{non-clique}}(\varphi) : [0,1] \rightarrow [0,1]$ to indicate the proportions:

$$
f_{\text{clique}}(\varphi) = \frac{|\{Y_{i_1i_2,\ldots,i_m} \in \Omega_{\text{clique}} | P(Y_{i_1i_2,\ldots,i_m} = 1 | Z \in [1-\varphi,1])\}|}{|\Omega_{\text{clique}}|},
$$

$$
f_{\text{non-clique}}(\varphi) = \frac{|\{Y_{i_1i_2,\ldots,i_m} \in \Omega_{\text{non-clique}} | P(Y_{i_1i_2,\ldots,i_m} = 1 | Z \in [0,\varphi])\}|}{|\Omega_{\text{non-clique}}|}.
$$

Intuitively, $f_{\text{clique}}(\cdot)$ and $f_{\text{non-clique}}(\cdot)$ quantify the proportion of hyperlinks with a large generating probability within $\Omega_{\text{clique}}$ and hyperlinks with small generating probability within $\Omega_{\text{non-clique}}$. Based on (17), (18), and (19), we can show that the size of augmented hyperlink statuses with inference bias smaller than $\epsilon$ is formulated as

$$
|\hat{\Omega}_Y(\epsilon, \rho_{\text{obs}}) \setminus \Omega_Y| = f_{\text{clique}}(\frac{\epsilon}{1-\rho_{\text{obs}}})|\Omega_{\text{clique}}(\rho_{\text{obs}})|
$$

$$
+ f_{\text{non-clique}}(\frac{\epsilon}{1-\rho_{\text{obs}}})|\Omega_{\text{non-clique}}(\rho_{\text{obs}})|.
$$
Let $\epsilon_{\text{ini}}$ denote the mean square error of hyperlink probability estimation from the JLE using observed network, the convergence property for the estimator of joint link embedding incorporating augmented hyperlink statuses is established in the following Theorem 4.2.

**Theorem 4.2.** Assume that the hyperlink statuses are inferred through the proposed augmentation procedure in Section 3.4. Denote $\hat{\Theta}$ as the minimizer of the JLE objective function with augmented hyperlink statuses, and $N_{N,m} := C_N^2 + C_m^2$. Under the assumption (C1) and $\tau = O(\epsilon_{\text{aug}})$, we establish the convergence rate for $\hat{\Theta}$:

$$P \left( \frac{\|\hat{\Theta} - \Theta\|_{\text{JLE}}}{\sqrt{N_{N,m}}} \geq \eta \right) \leq 1 - \exp \left( -c (1 + \epsilon_{\text{aug}}^2)^2 \right),$$

where $c$ is the constant as in Theorem 4.1, $\eta = \max(\epsilon_{\text{aug}}(\lambda^{1/2}), 2)$, $\epsilon_{\text{ini}}$ and $\epsilon_{\text{obs}}$ denote the mean square error of hyperlink probability estimation from the JLE using observed network, the convergence rate of $\hat{\Theta}$ can be achieved when $\lambda \geq \epsilon_{\text{aug}}^2$ with $\epsilon_{\text{aug}} = \frac{1 + \epsilon_{\text{ini}}^2}{(\Omega_1 + |\Omega_2|)^{1/2}}$.

Theorem 4.2 shows that the proposed JLE can achieve a faster convergence rate via incorporating the augmented hyperlink statuses. Compared with the convergence rate of JLE at $1 + \frac{\epsilon_{\text{ini}}^2}{(\Omega_1 + |\Omega_2|)^{1/2}}$ from Theorem 4.1, the augmented JLE is $\sqrt{1 + \frac{\epsilon_{\text{ini}}^2}{(\Omega_1 + |\Omega_2|)^{1/2}}}$ times faster, and the improvement increases as the link dependency $\rho$ and observation dependency $\rho_{\text{obs}}$ increase.

In addition, when the estimation bias $\epsilon_{\text{aug}}$ is larger than $(1 - \rho)\epsilon_{\text{ini}}$, the term $(1 + c_1 + c_1^2 \epsilon_{\text{aug}}^{-1})^{-1}$ degenerates to 1. For an $\epsilon_{\text{aug}}$ smaller than $(1 - \rho)\epsilon_{\text{ini}}$, the term $(1 + c_1 + c_1^2 \epsilon_{\text{aug}}^{-1})^{-1}$ decreases as the $\epsilon_{\text{aug}}$ becomes smaller, hence, reducing the probability of achieving the convergence rate in (20). Intuitively, the quantity $\epsilon_{\text{ini}}/\epsilon_{\text{aug}}$ reflects the trade-off between bias and variance in that the hyperlink augmentation introduces both additional signals and noise in selecting an informative augmented hyperlink from $\Omega_{\text{pool}}$. However, Theorem 4.2 implies that the augmented JLE can achieve the best convergence rate with a higher probability than JLE as long as $\left\{ f_{\text{ clique}}(\epsilon_{\text{ini}}) - f_{\text{ non-clique}}(\epsilon_{\text{ini}}) \right\} \leq \epsilon_{\text{ini}} \left( \frac{1}{1 - \rho} \right)$, which can be satisfied in general when $\rho$ or the size of observed links $|\Omega_2| + |\Omega_2|$ is larger. Asymptotically, the JLE estimator and the augmented JLE estimator achieve the same convergence rate toward the true link probability. However, the augmented JLE is more efficient than the regular JLE.

In addition, Theorem 4.2 implies that the augmented JLE performs better when the pairwise link status and hyperlink status are conditionally dependent compared to the case when they are conditionally independent. Specifically, the convergence rate of augmented JLE is $\epsilon_{\text{aug}}(0) = \left( \frac{\Omega_1 + |\Omega_2| + |\Omega_2|^{1/2}(\epsilon_{\text{ini}}, \rho_{\text{obs}}))}{\sqrt{N_{N,m}}} \right)$ under the independent model and is $\epsilon_{\text{aug}}(\rho) = \frac{1 + c_0 \rho}{(\Omega_1 + |\Omega_2| + |\Omega_2|^{1/2}(\epsilon_{\text{ini}}, \rho_{\text{obs}}))^{1/2}}$ under dependent model. We can show that $\epsilon_{\text{aug}}(\rho) < \epsilon_{\text{aug}}(0)$ given that

$$\lim_{\rho \rightarrow 0} \frac{f_{\text{ clique}}(\rho)}{\sqrt{2} \Omega_1 (L + 1)} = 0$$

and

$$\lim_{\rho \rightarrow 0} \frac{f_{\text{ non-clique}}(\rho)}{\sqrt{2} \Omega_1 (L + 1)} = 0,$$

that is, both the $f_{\text{ clique}}(\rho)$ and $f_{\text{ non-clique}}(\rho)$ do not have a heavy left-tail, where $L := \frac{1}{2} |\Omega_2|^{1/2}(\epsilon_{\text{ini}}, \rho_{\text{obs}})$ and $L$ decreases as observation dependency $\rho_{\text{obs}}$ increases. Furthermore, the selection bias term $1 + c_1 + c_1^2 \epsilon_{\text{aug}}$ in (20) becomes smaller when $\rho > 0$. Therefore, the augmented JLE achieves a faster convergence rate with a higher probability when the pairwise links and hyperlinks are conditional dependent compared with independent model. We provide discussion on the generalization of Theorem 4.2 in supplementary material.

5. Numerical Study

In this section, we conduct simulation studies to illustrate the performance of the proposed method on pairwise link and hyperlink predictions on a network. In particular, we investigate three scenarios where the incorporated hyperlinks are directly observed, and are augmented through the observed network. In addition, we consider the scenario when hyperlink statuses are dependent on pairwise link statuses conditioning on the latent factor.

5.1. Study 1: Link Predictions with Observed Hyperlinks

In the first simulation study, we consider the network generated from an underlying latent space model to compare the performance of various methods based on different criteria of prediction performance.

Suppose there are $N$ nodes in a network and the rank of nodewise latent factors $Z = \{Z_i\}_{i=1}^N$ is 5, where $Z_i = (Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}, Z_{i5})$, and each latent feature $Z_{ir}$ is generated from a mixture uniform distribution $Z_{ir} \sim \mu \times \text{unif}(-1,0.6) + (1 - \mu) \times \text{unif}(0.6,1)$, $i = 1, \ldots, N$, $r = 1, \ldots, 5$, with $\mu \sim \text{Bern}(1,0.5)$. We introduce a feature weighting vector $\alpha = (1,1,1,0.2,0.2)$ and the weighted latent feature $Z_{a} = \{Z_{a,i}\}_{i=1}^N$ where $Z_{a,i} = (\alpha_1 Z_{i1}, \alpha_2 Z_{i2}, \alpha_3 Z_{i3}, \alpha_4 Z_{i4}, \alpha_5 Z_{i5})$. Given the latent factors $Z_{a}$, pairwise link statuses are generated from the latent factor model in (9). Once pairwise link network $Y$ is generated, we randomly split $Y$ into training, validation and testing sets with corresponding proportions of 60%, 20%, and 20%, respectively.

In the following simulations, we consider a network consisting of 3-order hyperlinks where each hyperlink status $Y_{ijk}$ is independently generated to mimic the case where hyperlinks are directly observed, that is,

$$Y_{ijk} \sim \text{Bern}(P_{ijk}),$$

$$P_{ijk} = \sigma \left\{ c(Z_{a,i}^T Z_{a,j} + Z_{a,j}^T Z_{a,k} + Z_{a,k}^T Z_{a,i}) + \beta f(Z_{a,i}, Z_{a,j}, Z_{a,k}) \right\},$$

where $c = 0.2, 0.2, 0.2, 1, 1$, $c = 1$, $\beta = 3$. For generating the training and testing sets of hyperlinks, we use a
hierarchical sampling procedure and the size of the training 3-order hyperlinks is about 0.2% of all possible hyperlinks. This is consistent with real applications in that the multi-way relations are more difficult to observe and cost more to verify in contrast to the pairwise links. The details about the hierarchical sampling procedure is provided in supplemental material.

We then randomly divide the complement set \{Y_{ijk} | Y_i \in \Omega_Y, Y_j \in \Omega_Y, Y_k \in \Omega_Y\} into the validation set \Omega_Y^{valid} and the testing set \Omega_Y^{test} with a proportion of 50% and 50%. Similar to the pairwise link, we further differentiate the links in the testing set \Omega_Y^{test} through defining

\[
A_1^{\text{hyper}} = \{ Y_{ijk} \in \Omega_Y^{test} | 0.2 \leq P(Y_{ijk}) \leq 0.8 \},
\]

\[
A_2^{\text{hyper}} = \{ Y_{ijk} \in \Omega_Y^{test} | P(Y_{ijk}) < 0.2 \text{ or } P(Y_{ijk}) > 0.8 \}.
\]

The rationale for such a testing set formulation is to investigate the prediction performance of different methods on links with intrinsic uncertainty at different levels. We also illustrate performance comparisons on the complete testing sets of pairwise link statuses and hyperlink statuses, which are denoted as \(A_{\text{test}}^{\text{pair}}\), and \(A_{\text{test}}^{\text{hyper}}\), respectively.

To investigate the performance of incorporating hyperlinks on improving link prediction, we compare six different methods. The first three methods are based on the proposed framework. Specifically, the first one obtains the estimation of the latent factor \(Z\) through the loss function in (1), which is equivalent to only utilizing observed pairwise links for embedding, and is denoted as pairwise link embedding (PLE).

The second method estimates the latent factor \(Z\) through the loss function defined in (4) where the latent concordance within the link function is replaced by symmetric CP decomposition \[ \sum_{r=1}^{5} Z_{ir}Z_{jr}Z_{kr}, \] and is denoted as hyperlink embedding (HLE). The proposed method is denoted as joint link embedding (JLE) which incorporates both pairwise links and hyperlinks to jointly estimate \(Z\) through the proposed loss function (5), where we set all the hyperlink weights \(w_{li(j)} = 1\).

In addition, we also compare the proposed method with three other popular and state-of-the-art network embedding methods including learning graph representations with global structural information (GraRep) (Cao, Lu, and Xu 2015), large-scale information network embedding (LINE) (Tang et al. 2015) and scalable feature learning for networks (Node2Vec) (Grover and Leskovec 2016). In general, they encode observed pairwise or high-order relations into node-wise latent factors through decomposing a series of graphical laplacian matrices with different orders, or encouraging \(Z\) to conform with the similarity among nodes calculated based on biased random walks on the observed network. We obtain node-wise embedding estimation through the different methods mentioned above and then predict the probability of testing links following (6) and (7). The performance of prediction is measured by the AUC (area under the ROC curve).

The prediction results are provided in Table 1 and Figure 2, showing that the proposed joint embedding method consistently outperforms the other methods in terms of achieving higher AUC. For the pairwise link prediction, the proposed joint embedding (JLE) slightly outperforms PLE. For hyperlink prediction, the joint embedding also achieves more than 10% improvement on testing set \(A_1^{\text{hyper}}\), and more than 16% improvement on testing set \(A_2^{\text{hyper}}\) compared with using hyperlinks (HLE) alone. This indicates that the joint embedding strategy takes advantage of dependency to encode both two-way relations and multi-way relations into latent factor \(Z\) simultaneously, hence, achieving better prediction performance, especially for low-uncertainty links.

Compared with the three other existing embedding methods, the proposed method achieves comparable performance to the best competing method GraRep on \(A_1^{\text{test}}\), but more than 12% improvement when \(N = 100\). In addition, the improvement on hyperlink prediction is more than 16% on \(A_1^{\text{hyper}}\) and \(A_2^{\text{hyper}}\), and more than 22% on \(A_3^{\text{hyper}}\). The improvement with the proposed method demonstrates that the high-order approximation captured by hyperlinks is crucial for inferring potential pairwise links. In terms of hyperlink prediction, the collection of two-way relations alone is not sufficient to determine the underlying multi-way relations. In addition, we provide the more detailed simulation results, and the ROC for comparisons between PLE, HLE, and JLE in the supplementary material.

### 5.2. Study 2: Link Prediction with the Hyperlinks Inferred

In this subsection, we mimic the situation where few hyperlinks are observed, but most of them are inferred from the observed pairwise links. The inferred hyperlinks might be misspecified due to the error propagation from misspecified pairwise links, random sampling, and gaps between two-way relations and multi-way relations. Nevertheless, the intrinsic hierarchical dependency between hyperlinks and pairwise links can still benefit the recovery of partial high-order information via incorporating information inferred beyond the observed two-way relations. In this study we investigate the case where the pairwise link statuses and hyperlink statuses are independent conditioning on the latent factors \(Z\).

The training dataset and testing dataset are generated via the same procedure in Study 1. Specifically, the hyperlink training set \(\Omega_Y\) consists of 1% random samples from the set \(\Omega^{valid} = \{ Y_{ijk} | Y_i \in \Omega_Y, Y_j \in \Omega_Y, Y_k \in \Omega_Y \}\), which leads to a sparser hyperlink scenario compared with Study 1. We denote the joint embedding method with augmented hyperlinks proposed in Section 3.4 as Aug JLE. Table 2 provides the link prediction performance.

| Network size | Node2Vec | LINE | GraRep | PLE | HLE | JLE |
|--------------|----------|------|--------|-----|-----|-----|
| N = 100      |          |      |        |     |     | 0.77|
| N = 200      |          |      |        |     |     | 0.78|
| N = 300      |          |      |        |     |     | 0.92|

Table 1. The link prediction performance of different methods on different hyperlink testing sets when hyperlinks are observed.
JLE outperforms PLE and JLE on both $A_{\text{pair}}^{\text{test}}$ and $A_{\text{hyper}}^{\text{test}}$, and achieves more than 7% improvement on $A_{\text{pair}}^{\text{test}}$, and 10% improvement on $A_{\text{hyper}}^{\text{test}}$ compared with the other competing methods. The simulation results show the advantage of introducing dependency between pairwise link and hyperlink, which is the key step for hyperlink inference. However, the dependency are relatively weak when pairwise links and hyperlinks are independent conditioning on latent factors, which leads to a limited improvement from the hyperlink augmentation procedure. Therefore, we investigate the performance of the proposed methods when the dependency between pairwise links and hyperlinks are stronger. The detailed simulation results, and the ROC for comparisons between PLE, HLE, JLE, and Aug JLE are included in supplementary material.

We also investigate the improvement of augmented JLE over JLE in Table 3 under different missing rates of pairwise link status and a fixed sample size of hyperlink statuses. The missing rates for the observed pairwise link statuses are set as 20%, 30%, 40%, and 50%, and the sample size of hyperlink statuses are 1% of $\Omega_{\text{train}}^{\text{hyper}}$. Table 3 shows that the augmented JLE outperforms the JLE for both link prediction and link probability recovery under various sample sizes, and the improvement from the augmented JLE decreases as the size of observed link status increases.

### 5.3. Study 3: Link Prediction under the Conditional Dependent Model

In the previous simulation studies, the link-generating process follows a conditional independent model in that the pairwise...
link statuses and hyperlink statuses are independent conditioning on the nodewise latent factors Z; that is, we have $p_{i1...im} = 0$ in (13) and $P(Y_{i1...im} = 1 | \Delta_{i1...im}, \mathbf{Z}) = P(Y_{i1...im} = 1 | \mathbf{Z})$ for $1 \leq i_1 \leq \cdots \leq i_m \leq N$. In this subsection, we consider a setting where pairwise link statuses and hyperlink statuses are conditional dependent given latent factors Z, that is, $p_{i1...im} > 0$ and the hyperlink statuses are generated based on both Z and the pairwise link clique $\Delta_{i1...im}$ in (10). We investigate the performance comparisons between the augmented joint link-embedding method (Aug JLE) and other link-embedding methods using observed networks only.

In the following, we assume that the $N$ latent positions $\mathbf{Z} \in \mathbb{R}^{N \times S} = \{Z_{i1}, \ldots, Z_{iS}\}$ evenly distribute across $K = 6$ clusters. Then the pairwise link network is generated from (9), and we adjust the observation dependency within a pairwise link network so that $p_{\text{obs}}$ can be equal to different values. To incorporate the conditional dependency between pairwise link statuses and hyperlink statuses given latent factors, we adopt a random effect model for generating hyperlink statuses. Specifically, we adjust the magnitude of random effect such that empirical estimation of condition dependency (13) is equal to a specific value. We provide the more details regarding the link generating process and settings of dependency in the supplemental material.

We randomly split pairwise link statuses into training, validation and testing sets with corresponding proportions of 40%, 20%, and 40%, respectively. We consider a sparse hyperlink network scenario in that the hyperlink training set consists of only 30 hyperlink statuses. The hyperlink testing set $Y_{\text{test}}^{\text{hyper}} = \{Y_{i1j} \mid Y_{i1j} \neq 0\}$ are collected in a way that $\{Y_{i1j} \mid Y_{i1j} = 1\}$ and the hyperlink statuses $|\{Y_{i1j} = 1\}|$ and $|\{Y_{i1j} = 0\}|$ are balanced.

We increase the link dependency $\rho$ from 0.25 to 0.85, and increase the observation dependency $p_{\text{obs}}$ from 0.15 to 0.35. We compare the augmented JLE with other methods on both pairwise link prediction and hyperlink predictions under various degrees of dependency. Table 4, Table 5 and Figure 4 show that the augmented JLE consistently outperforms the JLE and other methods for both pairwise link prediction and hyperlink prediction under different sizes of networks. In addition, the improvement from the augmented JLE is more significant when $\rho$ and $p_{\text{obs}}$ are large. Specifically, the improvement is up to 33% on pairwise link prediction and 25% on hyperlink prediction over JLE when $p_{\text{obs}} = 0.35$ and $\rho = 0.85$. Furthermore, Table 6 shows that the link dependency $\rho$ can also affect the performance of the augmented JLE. That is, given $p_{\text{obs}}$ is fixed, the improvement from the augmented JLE is more significant when $\rho$ is large. Intuitively, as the conditional dependency becomes stronger, the augmentation procedure imposes stronger structural constraints on the latent factor $\mathbf{Z}$ to capture the underlying subgroup structure on the latent space.

To better illustrate the strength of the hyperlink augmentation, we also investigate the performance of the proposed methods given the hyperlink generating model is misspecified. Similarly, we consider a pairwise link network with $K = 12$ clusters among $N$ nodes such that nodewise latent factors Z within the same cluster are closer. The pairwise links $\{Y_{ij}\}$ are independently generated based on the latent positions as in (9). Therefore, the probability of pairwise link within the same cluster is about 0.90 and is about 0.25 for between-cluster links.

### Table 4. The link predictions performance of different methods.

| $p_{\text{obs}} = 0.15, \rho = 0.25$ | PLE | HLE | JLE | Aug JLE |
|----------------------------------|-----|-----|-----|---------|
| AUCPair | 0.56(0.02) | 0.52(0.02) | 0.57(0.01) | 0.61(0.03) |
| AUCCyper | 0.57(0.02) | 0.53(0.01) | 0.57(0.02) | 0.62(0.02) |
| AUCCyper | 0.55(0.01) | 0.53(0.02) | 0.59(0.02) | 0.67(0.02) |
| AUCCyper | 0.59(0.01) | 0.55(0.02) | 0.60(0.02) | 0.68(0.01) |
| AUCCyper | 0.54(0.02) | 0.51(0.02) | 0.57(0.02) | 0.68(0.02) |
| AUCCyper | 0.54(0.01) | 0.50(0.03) | 0.57(0.01) | 0.68(0.01) |

### Table 5. The link predictions performance of different methods.

| $p_{\text{obs}} = 0.15, \rho = 0.25$ | PLE | HLE | JLE | Aug JLE |
|----------------------------------|-----|-----|-----|---------|
| AUCPair | 0.59(0.01) | 0.52(0.01) | 0.60(0.01) | 0.67(0.01) |
| AUCCyper | 0.60(0.02) | 0.51(0.01) | 0.60(0.01) | 0.65(0.01) |
| AUCCyper | 0.59(0.01) | 0.56(0.01) | 0.60(0.01) | 0.71(0.01) |
| AUCCyper | 0.63(0.02) | 0.53(0.01) | 0.62(0.01) | 0.71(0.01) |
| AUCCyper | 0.54(0.01) | 0.52(0.01) | 0.54(0.01) | 0.72(0.01) |
| AUCCyper | 0.54(0.02) | 0.50(0.01) | 0.56(0.01) | 0.70(0.01) |

Based on 3-order $\Delta_{i1j1}$, we assume that a hyperlink status $Y_{i1j1}$ is independently generated from the Bernoulli distribution with probability $P(Y_{i1j1} = 1)$ as:

\[
P(Y_{i1j1} = 1) = \begin{cases} 0.90, & \text{if } i_1, i_2, i_3 \text{ are in same cluster and } \\
0.10, & \text{if } i_1, i_2, i_3 \text{ are not in same cluster and } \\
0, & \text{otherwise,} 
\end{cases}
\]

and the link dependency can be quantified as $\rho := \text{corr} \{1_{\{Y_{i1j}=1\}}, 1_{\{Y_{i1j1}=1\}}, 1_{\{\Delta_{i1j1}=1\}}\}$. Intuitively, $\rho$ measures the marginal correlation between a hyperlink among $i_1, i_2, i_3$ and the corresponding clique among $i_1, i_2, i_3$. Under this generating model, the dependency between pairwise links and hyperlinks is $\rho = 0.8$. Notice that the generation of hyperlinks is not based on the latent factors which are used to formulate pairwise links. Instead, hyperlinks are generated upon clique $\Delta$ to preserve the dependency with pairwise links. Figure 5 shows that the augmented JLE has a significant improvement on both pairwise link prediction and hyperlink prediction compared with other methods given that hyperlink model is misspecified while the link dependency is strong. Specifically, the improvement from the augmented JLE over JLE is 25% on pairwise link prediction and 32% in hyperlink prediction when the network size is $N = 360$.

### 6. Real Data Application

In this section, we apply the proposed joint embedding method to the Facebook social circles network dataset (Leskovec and Mcauley 2012), which contains the ego-network (Kim and Rinaldo 2017). Each ego denotes a specific user in Facebook, and his or her associated ego-network is the social network corresponding to this user's friends in Facebook. One of the
important attributes of ego-network is that it contains the social circles defined by the user, which leads to subgroups among people in the ego-network. Currently, the social circle is a common functionality for many popular social media such as Facebook, Google and Twitter. The purpose of introducing the social circle is to allow users to organize their own social network to mitigate the “information overload” through filtering contents or status updates posted by friends in specific groups. In addition, it allows users to protect their privacy by restricting or sharing personal information for specific groups of friends.

The major distinction between social circles and traditional social network communities is that social circles are in general highly overlapped and can be hierarchically nested, and therefore, different people in an ego-network might belong to multiple circles. In addition, unlike the social network communities identified with dense internal connections, social circles are formulated through specific attributes of friends selected by the user. For example, a user might cluster his or her social networks according to categories such as college friends, high school friends, department friends or colleagues.

Currently, social media adopts two methods to formulate social circles in an ego-network. The first one requires the user to manually group people, which is time-consuming and cannot be updated automatically when the user adds more friends. The second approach categorizes social circles through identifying people sharing common predefined attributes. However, it fails to incorporate the user’s individual preferences and suffers from missing profile information. In this section, we apply the proposed approach and investigate the performance of learning the user-specified clustering through network embedding.

The network data we consider here includes 224 nodes as users and 6384 undirected pairwise links as friend relationships. In addition, there are 14 overlapped circles within the network, which are formulated according to the similarity of social relation features among people defined by the users, such as common university affiliations, sports teams and relatives. These circles usually contain a large number of nodes so that the average size of each circle is 40 nodes and the largest circle contains 201 nodes. The ego-network is illustrated in Figure 6.
We first randomly split the pairwise links from the ego-network into training, validation and test sets with the proportion of 40%, 20%, and 40%. Incorporating multi-way information requires more elaborate preprocessing. We discard the most noninformative social circles, that is, the largest social circle with a size of 201, since the large social circle is noninformative in differentiating subgroup-specific attributes, and denote the with a size of 201, since the large social circle is noninformative in differentiating subgroup-specific attributes, and denote the 13 remaining social circles as \( \{C_l\}_{l=1,\ldots,13} \). We then decompose each social circle into a set of three-way hyperlinks \( \Omega^\text{pool}_Y = \{Y_{ijk}\} \) according to the following rule:

\[
Y_{ijk} = \begin{cases} 
1, & |\{l | i \in C_l, j \in C_l, k \in C_l, l = 1, \ldots, 13\}| \geq 1 \\
0, & |\{l | i \in C_l, j \in C_l, k \in C_l, l = 1, \ldots, 13\}| = 0 
\end{cases},
\]

where \( | \cdot | \) denotes cardinality of a set. Through this preprocessing, the local multi-way relations and the global subgroups in the original social circles can be represented by the three-way relations constructed in (21). Finally, we generate training hyperlink statuses \( \hat{\Omega}_Y \) through randomly sampling 1% observed hyperlink statuses from \( \Omega^\text{pool}_Y \). The size of \( \hat{\Omega}_Y \) is about 600 after the preprocessing, and the proportions for present links \( \{Y_{ijk} \in \hat{\Omega}_Y | Y_{ijk} = 1\} \) and absent links \( \{Y_{ijk} \in \hat{\Omega}_Y | Y_{ijk} = 0\} \) are relatively balanced. For applying the augmented JLE, we also generate augmented hyperlinks statuses via utilizing the pairwise friendships \( \hat{Y} = \{Y_{ij} \in [0,1]_{1 \leq i < j \leq 224}\} \). Specifically, we collect the candidate pool for augmented hyperlink statuses \( \Omega^\text{pool}_Y = \{\hat{Y}_{ijk}\} \) as follows:

\[
\hat{Y}_{ijk} = \begin{cases} 
1, & Y_{ij} = Y_{ik} = Y_{jk} = 1, \text{ and } (i,j,k) \text{ in the same circle,} \\
0, & Y_{ij} = Y_{ik} = Y_{jk} = 0, \text{ and } (i,j,k) \text{ not in the same circle.}
\end{cases}
\]

Then we construct the augmented hyperlink statuses set \( \check{\Omega}_Y \) via randomly sampling 2000 augmented hyperlink statuses from \( \Omega^\text{pool}_Y \) with a balanced proportion between presence and absence of links.

We investigate the performance of the proposed method and other competing network embedding methods on predicting both two-way relations on the ego-network. In addition, we also investigate the performance of hyperlink prediction to evaluate whether the social circle information has been encoded into nodes’ latent factors. Instead of directly predicting the original social circles, we predict the joint memberships of specific \( m \) nodes, that is, whether they belong to the same social circle or not, and compare the result with the original circles. We choose the order of testing hyperlinks as \( m = 6,10, \) and the prediction is based on estimated \( Z \) through \( P(Y_{(i_1\ldots i_m)} = 1) = \frac{\exp(\sum_{i_(j\in[i_1\ldots i_m]}Z_{ik}^TZ_{lj}^T)}{1 + \exp(\sum_{i_(j\in[i_1\ldots i_m]}Z_{ik}^TZ_{lj}^T)} \). For the proposed joint embedding method JLE, and the counter-method of only using PLE, the rank of \( Z \) is chosen at \( r = 5 \). For the tuning parameters of competing methods, we adopt the strategies recommended by the original papers or the packages.

The comparison of performance is illustrated in Table 7. For the two-way relation prediction, the proposed JLE and augmented JLE method achieve top performances compared with competing methods. It achieves about 30% improvement over the LINE and 60% improvement over the Node2Vec. The real data analyses show the importance of borrowing high-order relation information for the two-way relation predictions. The incorporated third-order hyperlinks recover the underlying subgroup structure, and the estimated latent features characterize the subgroup attributes. If partial subgroup information is recovered, the two-way relations from the subgroup level provide additional information regarding potential friendships. The inferior performance of LINE and Node2Vec might be due to the relatively sparse ego-network as the pairwise links only account for 13% of the total possible friendships.

In terms of the multi-way relation prediction, the proposed augmented JLE method achieves the best performance for both 6-order and 10-order hyperlink prediction. Specifically, the Aug JLE achieves 23% improvement over the best existing method GraRep on the 6-order hyperlink prediction and 29% improvement over the best existing method LINE on the 10-order hyperlink prediction. Although the partial embedding method PLE and the GraRep perform well for the two-way relation prediction, they do not have consistent performance for hyperlink predictions, indicating that the social circles indeed encode significant high-order relations information which cannot be represented through two-way relations. This demonstrates the importance of incorporating high-order information for predicting multi-way relations.

7. Discussion

In this article, we propose a new network link prediction method. The major innovation of the proposed method is to incorporate the multi-way relation into the network embedding process and therefore, jointly embed both the hyperlinks and pairwise links. It allows the node-wise latent factors to encode the pairwise similarity to their neighborhood, and induce cohesive high-order subgroups. In addition, the proposed method formulates hierarchical modeling for the link-generating

![Figure 6. The ego-network in the Facebook dataset, where the social circles are marked as polygons with different colors.](image)
process to introduce the dependency between pairwise links and hyperlinks. In terms of estimating latent factors, the link dependency allows borrowing the mutual information between pairwise links and hyperlinks such that prediction for both pairwise links and hyperlinks can be improved. In term of model interpretability, the link dependency reflects the principle that high-order interaction among nodes in a network can in general be built on the low-order interactions. In theory, we establish the consistency of the link probability estimator based on the proposed joint embedding loss function. In addition, we show that the convergence rate can be improved through incorporating the observed hyperlinks and hyperlink augmentation.

In this article, we only consider a latent-factor based hyperlink augmentation approach. However, it is worth further exploration to develop an inference procedure to learn the hierarchical dependency of hyperlinks on pairwise links. In addition, generalization of the proposed method to varying hyperlink orders is also one of the important directions for our future study.

Acknowledgments

The authors thank the Associate Editor and two anonymous reviewers for their suggestions and helpful feedback which improved the paper significantly.

Funding

This work is supported by NSF grants DMS 2019461 and DMS 1952406.

Supplementary Materials

The supplementary materials provide proofs of Theorem 4.1, Theorem 4.2, optimization algorithm in Section 3.5, explicit formulation for hyperlink modeling, detailed discussion on the theoretical results, detailed settings and results of simulations, and ROC graphs.

References

Agarwal, S., Branson, K., and Belongie, S. (2006). “Higher Order Learning with Graphs,” in Proceedings of the 23rd International Conference on Machine learning, pp. 17–24. ACM. [1692,1693]
Ahmed, A., Shervashidze, N., Narayanamurthy, S., Josifovski, V., and Smola, A. J. (2013). “Distributed Large-Scale Natural Graph Factorization,” in Proceedings of the 22nd International Conference on World Wide Web, pp. 37–48. ACM. [1696]
Al Hasan, M., Chaoji, V., Salem, S., and Zaki, M. (2006). “Link Prediction Using Supervised Learning,” in SDM06: Workshop on Link Analysis, Counter-terrorism and Security. [1692]
Al Hasan, M., and Zaki, M. I. (2011). “A Survey of Link Prediction in Social Networks,” in Social Network Data Analytics, ed. C. C. Aggarwal, pp. 243–275. London: Springer. [1692]
Alvarez-Rodriguez, U., Battistoni, F., de Arruda, G. F., Moreno, Y., Perc, M., and Latora, V. (2021). “Evolutionary Dynamics of Higher-Order Interactions in Social Networks,” Nature Human Behaviour, 5, 586–595. [1697]
Belkin, M., and Niyogi, P. (2002). “Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering,” in Advances in Neural Information Processing Systems, pp. 585–591. [1696]
Brier, G. W. (1950). “Verification of Forecasts Expressed in Terms of Probability,” Monthly Weather Review, 78, 1–3. [1694]
Qiu, J., Dong, Y., Ma, H., Li, J., Wang, K., and Tang, J. (2018). “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. [1696]
Razick, S., Magklaras, G., and Donaldson, I. M. (2008). “Irefindex: A Consolidated Protein Interaction Database with Provenance,” BMC Bioinformatics, 9, 1–9. [1692]
Rinaldo, A., Bacanu, S.-A., Devlin, B., Sonpar, V., Wasserman, L., and Roeder, K. (2005). “Characterization of Multilocus Linkage Disequilibrium,” Genetic Epidemiology, 28, 193–206. [1692]
Shojaie, A., and Michailidis, G. (2009). “Analysis of Gene Sets based on the Underlying Regulatory Network,” Journal of Computational Biology, 16, 407–426. [1692]
Song, H. H., Cho, T. W., Dave, V., Zhang, Y., and Qiu, L. (2009). “Scalable Proximity Estimation and Link Prediction in Online Social Networks,” in Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, pp. 322–335. ACM. [1692]
Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., and Mei, Q. (2015). “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. [1692,1700]
Varin, C., Reid, N., and Firth, D. (2011). “An Overview of Composite Likelihood Methods,” Statistica Sinica, 21, 5–42. [1697]
Xu, B., Lin, H., Chen, Y., Yang, Z., and Liu, H. (2013). “Protein Complex Identification by Integrating Protein-Protein Interaction Evidence from Multiple Sources,” PLoS One, 8, e3841. [1697]
Zhang, Y., Levina, E., and Zhu, J. (2017). “Estimating Network Edge Probabilities by Neighbourhood Smoothing.” Biometrika, 104, 771–783. [1692]
Zhao, Y., Wu, Y.-J., Levina, E., and Zhu, J. (2017). “Link Prediction for Partially Observed Networks,” Journal of Computational and Graphical Statistics, 26, 725–733. [1692]
Zhou, D., Huang, J., and Schölkopf, B. (2007). “Learning with Hypergraphs: Clustering, Classification, and Embedding,” in Advances in Neural Information Processing Systems, pp. 1601–1608. [1692,1696]
Zhu, Y., Guan, Z., Tan, S., Liu, H., Cai, D., and He, X. (2016). “Heterogeneous Hypergraph Embedding for Document Recommendation,” Neurocomputing, 216, 150–162. [1692,1696]
Zlatić, V., Ghoshal, G., and Caldarelli, G. (2009). “Hypergraph Topological Quantities for Tagged Social Networks,” Physical Review E, 80, 036118. [1692]