ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10448101

A BAYESIAN APPROACH TO HIGH-ORDER LINK PREDICTION

Georgios V. Karanikolas,[†] *Alba Pagès-Zamora,*[‡] *and Georgios B. Giannakis*[†]

[†]Dept. of ECE, University of Minnesota, Minneapolis, MN
[‡] SPCOM Group, Universitat Politècnica de Catalunya, Barcelona, Spain

ABSTRACT

Using a subset of observed network links, high-order link prediction (HOLP) infers missing hyperedges, that is links connecting three or more nodes. HOLP emerges in several applications, but existing approaches have not dealt with the associated predictor's performance. To overcome this limitation, the present contribution develops a Bayesian approach and the relevant predictive distributions that quantify model uncertainty. Gaussian processes model the dependence of each node to the remaining nodes. These nonparametric models yield predictive distributions, which are fused across nodes by means of a pseudo-likelihood based criterion. Performance is quantified by proper measures of dispersion, which are associated with the predictive distributions. Tests on benchmark datasets demonstrate the benefits of the novel approach.

***Index Terms*—** Link prediction, hypergraphs, Gaussian processes

1. INTRODUCTION

Link prediction aims at predicting edges (pairwise links) that are missing from a graph [11, 20, 10]. Missing edges may either be present, albeit not observed, or they may appear in the future, in the case of evolving networks [6].

Going beyond pairwise links, hyperedges connect three or more nodes of a hypergraph [3, 1, 17]. They model higher-than second-order nodal dependencies, and HOLP aims at predicting such dependencies [23, 24]. HOLP arises naturally, in e.g., drug substance networks, where nodes represent substances, and a (hyper-)edge models the pertinent drug substances. As drugs can contain more than two substances, going beyond predictions of pairs of nodes, is well justified.

Prior works. HOLP methods can be broadly categorized into informal scoring, supervised, and unsupervised learning based methods. Examples of supervised methods include structural support vector machine [23] and neural network based approaches [21], whereas matrix factorizations [24] and distribution similarities [5] are leveraged in unsupervised approaches. A common limitation shared by existing approaches, regardless of category, is the inability to provide estimates of the uncertainty associated with the predictions made. Such estimates can be used in subsequent decision making, or, in order to guide data collection for further model improvement. An additional limitation faced by supervised methods is the requirement for artificially generated "absent hyperedges," that is sets of nodes that are assumed not to be connected by a hyperedge. To see why this is the case, note that supervised methods tackle HOLP as a classification problem of "present" versus "absent" hyperedges. Training a classifier requires examples from both classes. Real world networks however, typically comprise only "present" hyperedges, thereby necessitating artificially examples for the "absent" class, what of course biases the learned predictor.

To overcome the limitations of existing methods, our novel approach utilizes Gaussian processes to obtain expressive probabilistic models of the dependencies amongst nodes. Combining the resulting node-wise models through a pseudo-likelihood yields an effective predictor, whose learning does not require artificially generated "absent hyperedges," unlike supervised approaches. Finally, we leverage the predictive distributions, learned through our Bayesian formulation, to extract measures of uncertainty on the participation of a node to a hyperedge, as well as on hyperedge presence prediction. All in all, the novel method does not entail unnecessary biases, while also additionally offering uncertainty estimates.

Contributions. The main novelty here is a Gaussian process based approach to HOLP, and its performance analysis. Nonparametric learning of nodal dependencies, and the adoption of pseudolikelihood in the predictive model are also novel in the context of HOLP.

Notation. Scalar $\mathbf{A}_{ij}$ denotes the (i,j)-th entry of the matrix $\mathbf{A}$; $[\mathbf{a}]_i$ the i-th entry of vector $\mathbf{a}$; superscript $^\top$ transposition; and, $\mathbf{I}$ stands for the identity matrix. Moreover, $|\cdot|$ denotes set cardinality, and $\mathbb{1}\{\cdot\}$ is the indicator function. The expectation and variance of a random variable Z are denoted by $\mathbb{E}[Z]$ and $\mathbb{V}[Z]$, respectively. Finally, $\mathcal{N}(x;\mu,\sigma^2)$ represents the probability density function of a Gaussian random variable (with mean μ and variance σ^2) taking the value x.

This work was supported in part by NSF grants 1901134, 2126052, 2212318, 2220292. A. Pagès-Zamora was supported by grants PID2022-137099NB-C41 funded by MCIN/AEI/ 10.13039/501100011033; and by 2021 SGR 01033 funded by Dept. de Recerca i Univ. de la Generalitat de Catalunya 10.13039/501100002809.

2. PRELIMINARIES

Let $\mathcal{G}$ be an undirected hypergraph $\mathcal{G} := (\mathcal{V}, \mathcal{E}_o)$, where $\mathcal{V}$ is the set of nodes (vertices), and $\mathcal{E}_o$ is the set of both observed edges (pairwise links) and hyperedges (links among three or more nodes). Conceptually, the goal is to retrieve the set $\mathcal{E}_u$ of *unobserved* hyperedges, where each hyperedge is allowed to connect an arbitrary number of nodes. In practice, a set of candidate hyperedges $\mathcal{E}_c$ is considered, and the HOLP approach assigns a score $S(e^*)$ to each candidate hyperedge $e^* \in \mathcal{E}_c$, where higher scores shall be given to candidate hyperedges that are deemed more likely to exist.

As is customary, the structure of hypergraph $\mathcal{G}$ will be represented by the incidence matrix $\mathbf{H} \in \{0,1\}^{N \times |\mathcal{E}_o|}$, where $N = |\mathcal{V}|$ denotes the number of nodes, and $\mathbf{H}_{im} = \mathbb{1}\{i \in e_m\}$, that is $\mathbf{H}_{im} = 1$, if node i participates in (hyper) edge $e_m \in \mathcal{E}_o$, and $\mathbf{H}_{im} = 0$ otherwise [3]. Let also $\mathbf{H} =: [\mathbf{h}^{(1)}, \ldots, \mathbf{h}^{(M)}]$, where we shall refer to $\mathbf{h}^{(m)}$ as the incidence vector associated with (hyper) edge e_m, and $M = |\mathcal{E}_o|$.

3. GP-BASED NODAL MODELS

As a starting point, a probabilistic model is sought for each node i that relates the participation of node i in a hyperedge to the participation of the remaining nodes. Fusing the learned models across nodes, and providing hyperedge predictions, is discussed in the ensuing section.

Consider an arbitrary (hyper) edge $e_\circ$, the associated incidence vector $\mathbf{h}^{(\circ)}$ and let also $\mathbf{h}^{(\circ)}_{\setminus i} := [h^{(\circ)}_1, \ldots, h^{(\circ)}_{i-1}, h^{(\circ)}_{i+1}, \ldots, h^{(\circ)}_N]^\top$ where nonzero entries correspond to nodes, other than i, that participate in (hyper) edge $e_\circ$. The likelihood function of our model is given by

$$p(h^{(\circ)}_i = 1 | \mathbf{h}^{(\circ)}_{\setminus i}) = \Phi(f_i(\mathbf{h}^{(\circ)}_{\setminus i})) \tag{1}$$

where $f_i(\cdot)$ is a nonlinear real-valued function, and the Gaussian cumulative density function (CDF) $\Phi(\cdot)$ is used to map f_i to the unit interval, thereby yielding a valid probability[1] [22]. In words, the participation of node i in hyperedge $e_\circ$ is modeled by a Bernoulli distribution, whose probability of success is a function of the participation of the remaining nodes (cf. (1)).

In order to obtain a highly expressive Bayesian model, a Gaussian process (GP) prior is placed on f_i, with covariance structure captured by a kernel $\kappa_{\boldsymbol{\theta}_i}$, where $\boldsymbol{\theta}_i$ collects the kernel hyperparameters [22]. Letting $\mathbf{f}_i := [f_i(\mathbf{h}^{(1)}_{\setminus i}), \ldots, f_i(\mathbf{h}^{(M)}_{\setminus i})]^\top \in \mathbb{R}^M$, the GP prior over functions translates to a multivariate Gaussian prior over function values, that is $p(\mathbf{f}_i | \mathbf{H}_{\setminus i}, \boldsymbol{\theta}_i) = \mathcal{N}(\mathbf{f}_i; \mathbf{0}, \mathbf{K}_i)$, where $\mathbf{H}_{\setminus i} := [\mathbf{h}^{(1)}_{\setminus i}, \ldots, \mathbf{h}^{(M)}_{\setminus i}]$ and the kernel matrix $\mathbf{K}_i \in \mathbb{R}^{M \times M}$ is formed as $[\mathbf{K}_i]_{l,l'} = \kappa_{\boldsymbol{\theta}_i}(\mathbf{h}^{(l)}_{\setminus i}, \mathbf{h}^{(l')}_{\setminus i})$. With $\mathbf{h}_i^\top := [\mathbf{H}_{i1}, \ldots, \mathbf{H}_{iM}]$ collecting the participation status of node i across observed (hyper) edges, the posterior over latent function values $\mathbf{f}_i$ is obtained through Bayes' rule as

$$p(\mathbf{f}_i | \mathbf{H}_{\setminus i}, \mathbf{h}_i^\top, \boldsymbol{\theta}_i) = \mathcal{N}(\mathbf{f}_i; \mathbf{0}, \mathbf{K}_i) \times \tag{2}$$
$$\times \frac{\prod_{m=1}^M [\Phi(f_i(\mathbf{h}^{(m)}_{\setminus i}))]^{\mathbf{H}_{im}} [1 - \Phi(f_i(\mathbf{h}^{(m)}_{\setminus i}))]^{1-\mathbf{H}_{im}}}{p(\mathbf{h}_i^\top | \mathbf{H}_{\setminus i}, \boldsymbol{\theta}_i)}$$

where the factorization $p(\mathbf{h}_i^\top | \mathbf{f}_i) = \prod_{m=1}^M p(\mathbf{H}_{im} | f_i(\mathbf{h}^{(m)}_{\setminus i}))$ was used, as $\{\mathbf{H}_{im}\}_{m=1}^M$ are considered conditionally independent given the latent function f_i [22]. The non-Gaussianity of (2), stemming from the non-Gaussian likelihood (1), however, makes inference challenging. As is customary in the context of GPs, a Gaussian approximation to the posterior, let $q(\mathbf{f}_i | \mathbf{H}_{\setminus i}, \mathbf{h}_i^\top, \boldsymbol{\theta}_i) = \mathcal{N}(\mathbf{f}_i; \mathbf{m}_i, \mathbf{A}_i)$, is leveraged [14]. Here we will rely on the popular expectation propagation (EP) approach [13, 9]; see also the Appendix.

Our end goal is, of course, to make predictions about unobserved hyperedges $e^* \notin \mathcal{E}_o$. With regards to node i, this translates to potentially unseen $\mathbf{h}^*_{\setminus i}$. To that end, first, the predictive distribution for the latent function value $f_i^* := f_i(\mathbf{h}^*_{\setminus i})$ is obtained as

$$p(f_i^* | \mathbf{h}^*_{\setminus i}, \mathbf{H}_{\setminus i}, \mathbf{h}_i^\top, \boldsymbol{\theta}_i) = \int p(f_i^* | \mathbf{f}_i, \mathbf{h}^*_{\setminus i}, \mathbf{H}_{\setminus i}, \mathbf{h}_i^\top, \boldsymbol{\theta}_i)$$
$$q(\mathbf{f}_i | \mathbf{H}_{\setminus i}, \mathbf{h}_i^\top, \boldsymbol{\theta}_i) d\mathbf{f}_i \tag{3}$$

with $p(f_i^* | \mathbf{f}_i, \mathbf{h}^*_{\setminus i}, \mathbf{H}_{\setminus i}, \mathbf{h}_i^\top, \boldsymbol{\theta}_i) = \mathcal{N}(f_i^*; \mathbf{k}_*^\top \mathbf{K}_i^{-1} \mathbf{m}_i, k_{**} - \mathbf{k}_*^\top (\mathbf{K}_i^{-1} - \mathbf{K}_i^{-1} \mathbf{A}_i \mathbf{K}_i^{-1}) \mathbf{k}_*)$, where $\mathbf{k}_* := [\kappa_{\boldsymbol{\theta}_i}(\mathbf{h}^*_{\setminus i}, \mathbf{h}^{(1)}_{\setminus i}), \ldots, \kappa_{\boldsymbol{\theta}_i}(\mathbf{h}^*_{\setminus i}, \mathbf{h}^{(M)}_{\setminus i})]^\top$ and $k_{**} := \kappa_{\boldsymbol{\theta}_i}(\mathbf{h}^*_{\setminus i}, \mathbf{h}^*_{\setminus i})$ [9]. The predictive distribution for the participation of node i in the potential hyperedge e^* is then obtained as

$$p(h_i^* = 1 | \mathbf{h}^*_{\setminus i}, \mathbf{H}, \boldsymbol{\theta}_i) = \int \Phi(f_i^*) p(f_i^* | \mathbf{h}^*_{\setminus i}, \mathbf{H}, \boldsymbol{\theta}_i) df_i^*. \tag{4}$$

where we used the fact that $\{\mathbf{H}_{\setminus i}, \mathbf{h}_i^\top\}$ is equivalent to $\mathbf{H}$ to simplify notation.

The GP-based formulation adopted provides a principled way for hyperparameter selection [22]. This is accomplished by maximizing over $\boldsymbol{\theta}_i$ the marginal likelihood, as per

$$\hat{\boldsymbol{\theta}}_i := \arg\max_{\boldsymbol{\theta}_i} \int p(\mathbf{h}_i^\top | \mathbf{f}_i) q(\mathbf{f}_i | \mathbf{H}_{\setminus i}, \mathbf{h}_i^\top, \boldsymbol{\theta}_i) d\mathbf{f}_i. \tag{5}$$

To illustrate the importance of hyperparameters in our context, let us turn our attention to the squared exponential automatic relevance determination (SE-ARD) kernel [22]

$$\kappa_{\boldsymbol{\theta}_i}(\mathbf{h}^{(l)}_{\setminus i}, \mathbf{h}^{(l')}_{\setminus i}) \propto \exp\left(-\frac{1}{2} \sum_{v=1}^{N-1} \left([\mathbf{h}^{(l)}_{\setminus i}]_v - [\mathbf{h}^{(l')}_{\setminus i}]_v \right)^2 / \sigma_{iv}^2 \right)$$

that is leveraged here, where $\boldsymbol{\theta}_i := [\sigma_{i1}^2, \ldots, \sigma_{i(N-1)}^2]$. The inverse of each scale parameter σ_{iv} indicates how relevant

[1] This is also referred to as the probit likelihood.

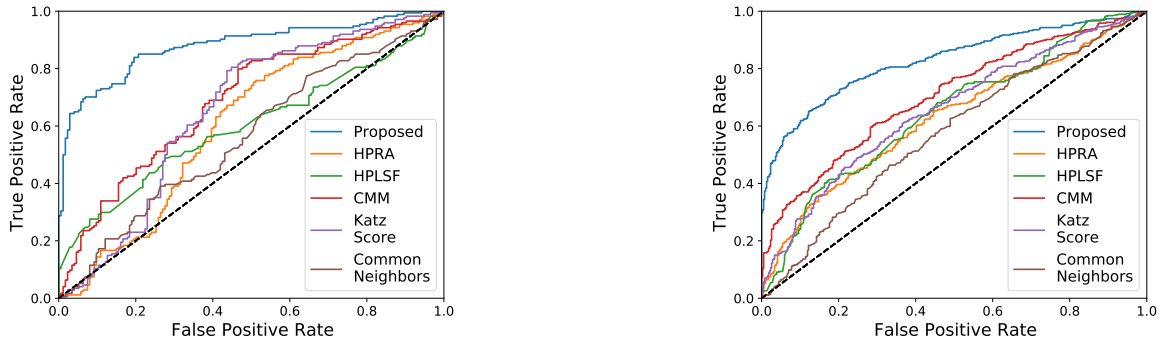

Fig. 1: ROC curves for hyperedge prediction on the (a) `tags-ask-ubuntu` and (b) `tags-math-sx` datasets. Comparison of the proposed HOLP approach against HPRA [8], CMM [24], HPLSF [23], Katz index and number of common neighbors.

node in position v is to predicting the participation of node i. Therefore, having a principled way to optimize over $\boldsymbol{\theta}_i$ allows for a nonparametric model of nodal dependencies, that is not only adaptable, but also explainable. Note finally that we will drop $\boldsymbol{\theta}_i$ from our notation, as $\boldsymbol{\theta}_i \to \hat{\boldsymbol{\theta}}_i$ hereafter.

Remark 1. As the cubic complexity associated with computing $\mathbf{K}_i^{-1}$ hinders scalability, low-rank approximants of the kernel matrix are leveraged, based on the deterministic training conditional approximation [18, 16].

4. HOLP AND UNCERTAINTY ESTIMATES

The goal in HOLP is to assign to each candidate hyperedge $e^* \in \mathcal{E}_c$ a score $S(e^*)$, that is predictive of hyperedge presence. Our scoring function amounts to fusing the node-wise predictors (cf. (4)) by relying on the notion of pseudolikelihood [4]. Formally, we have

$$S(e^*) = \prod_{i=1}^{N} p(h_i^*|\mathbf{h}_{\backslash i}^*, \mathbf{H}) \approx p(\mathbf{h}^*|\mathbf{H}) \tag{6}$$

where $\mathbf{h}^*$ is the incidence vector associated with e^*. Note that this scoring rule is rather intuitive, as it amounts to the approximate likelihood of the candidate hyperedge e^*, given the observed edges and hyperedges collected in $\mathbf{H}$.

Estimates of various notions of uncertainty, both at the node and at the hyperedge level, can be obtained within the proposed framework. As a starting point, consider the predictive likelihood for the participation of node i in candidate hyperedge e^*, as per (4). The key observation here is that $\Phi(f_i^*)$ is a random variable, due to the random function f_i, and thus random variable f_i^*. Intuitively, a lower variance $\mathbb{V}[\Phi(f_i^*)]$ indicates that the Bernoulli parameter concentrates more tightly around its mean $\mathbb{E}[\Phi(f_i^*)] \equiv p(h_i^* = 1|\mathbf{h}_{\backslash i}^*, \mathbf{H})$, thereby indicating lower model uncertainty with regards to the participation of node i in candidate hyperedge e^*.

To render the uncertainty estimates obtained across different nodes and hyperedges comparable, it is prudent to bring them to a common scale by using a normalized measure of dispersion, such as the coefficient of variation [19], defined here as $\mathrm{C}_i(e^*) = \sqrt{\mathbb{V}[\Phi(f_i^*)]}/\mathbb{E}[\Phi(f_i^*)]$. See also [7] for related concepts.

At the hyperedge level, we can proceed analogously. First, in order to compactly write (4) as a function of the participation of node i in e^*, let us introduce $\bar{h}_i^* := 2h_i^* - 1$. It follows that

$$p(h_i^*|\mathbf{h}_{\backslash i}^*, \mathbf{H}) = \int \Phi(\bar{h}_i^* f_i^*) p(f_i^*|\mathbf{h}_{\backslash i}^*, \mathbf{H}) df_i^* \tag{7}$$

where we used the fact that $p(h_i^* = 0|\cdot) = 1 - p(h_i^* = 1|\cdot)$ and the Gaussian CDF property $\Phi(-t) = 1 - \Phi(t)$. Substituting for the per-node predictive likelihoods as per (7) in (6) and applying Fubini's theorem we obtain

$$S(e^*) = \int \cdots \int \Phi(\bar{h}_1^* f_1^*) \ldots \Phi(\bar{h}_N^* f_N^*) \; p(f_1^*|\mathbf{h}_{\backslash 1}^*, \mathbf{H}) \ldots p(f_N^*|\mathbf{h}_{\backslash N}^*, \mathbf{H}) \, df_1^* \ldots df_N^*. \tag{8}$$

As $S(e^*|f_1^*, \ldots, f_N^*) := \Phi(\bar{h}_1^* f_1^*) \ldots \Phi(\bar{h}_N^* f_N^*)$ is a random variable, the scoring function in (8) is now viewed as an expectation $S(e^*) = \mathbb{E}_{\tilde{p}}[S(e^*|f_1^*, \ldots, f_N^*)]$ with respect to the measure $\tilde{p} := p(f_1^*|\mathbf{h}_{\backslash 1}^*, \mathbf{H}) \ldots p(f_N^*|\mathbf{h}_{\backslash N}^*, \mathbf{H})$. A notion of uncertainty at the hyperedge level can be obtained as $U_{\mathcal{V}}(e^*) = \sqrt{\mathbb{V}_{\tilde{p}}[S(e^*|f_1^*, \ldots, f_N^*)]}/\mathbb{E}_{\tilde{p}}[S(e^*|f_1^*, \ldots, f_N^*)]$. However, $U_{\mathcal{V}}(e^*)$ may inflate uncertainty as it tries to capture the uncertainty of both nodes $v \in e^*$ being present in e^* and nodes $v \notin e^*$ being absent from e^*. A perhaps more reasonable metric is obtained by conditioning on $v \notin e^*$ being absent, and focusing on the participating nodes, that is letting $U(e^*) = \sqrt{\mathbb{V}_{\tilde{p}}[\Phi(\bar{h}_{v_1}^* f_{v_1}^*) \ldots \Phi(\bar{h}_{v_{|e^*|}}^* f_{v_{|e^*|}}^*)]}/\mathbb{E}_{\tilde{p}}[\Phi(\bar{h}_{v_1}^* f_{v_1}^*) \ldots \Phi(\bar{h}_{v_{|e^*|}}^* f_{v_{|e^*|}}^*)]$ where $\{v_1, \ldots, v_{|e^*|}\} := \{v : v \in e^*\}$.

5. NUMERICAL TESTS

To assess the performance of our novel approach, tests were performed on the first $N = 100$ vertices and $M = 100,000$ (hyper) edges of the a) `tags-ask-ubuntu`, and

(a) (b)

Fig. 2: Uncertainty estimates in the `tags-ask-ubuntu` dataset.

b) `tags-math-sx` datasets. Nodes correspond to tags and each hyperedge connects the tags associated with a single question on the "Ask Ubuntu" and "Mathematics Stack Exchange" websites; see [2] for a detailed description.

In our tests, several HOLP alternatives were considered, including the hyperedge prediction using resource allocation (HPRA) method [8]; coordinated matrix minimization (CMM) [24]; the supervised hyperlink prediction using latent social features (HPLSF) approach [23]; as well as higher-order counterparts of standard link prediction scores, such as the Katz index and the number of common neighbors; see [24] for a detailed description.

With regards to the proposed approach, the SE-ARD kernel was used. Regarding CMM, all hyperparameters were set as per the original work [24], and we report the best results achieved across the embedding dimensionalities considered therein, namely $\{10, 20, 30\}$.

The set of unobserved hyperedges $\mathcal{E}_u$ was obtained by holding out 20% of hyperedges that were the last to appear in each dataset. As the goal is to assess hyperedge prediction performance, our candidate set $\mathcal{E}_c$ in the testing phase should also include "absent hyperedges." It is important to note that $\mathcal{E}_c$ is used here only for evaluation purposes. In particular, we relied on the clique negative sampling (CNS) scheme of [15], which is a higher-order counterpart of the widely used approach of [12], to generate the "absent hyperedge" set $\mathcal{E}_a$ with $|\mathcal{E}_a| = |\mathcal{E}_u|$. Finally, our test set is obtained as $\mathcal{E}_c = \mathcal{E}_u \cup \mathcal{E}_a$.

In the first test, the hyperedge prediction performance of the proposed approach was compared to that of competing alternatives. The results are depicted in the form of receiver operating characteristics (ROC) curves in Fig. 1. The proposed approach offers the best performance in both datasets. This can be attributed to the highly expressive model considered here, that also does not introduce unnecessary biases in the form of artificial "absent hyperedges."

As no ground truth is available regarding uncertainty, the purpose of the second test was to perform a sanity check on the uncertainty estimates provided by the proposed approach. Let $\bar{C}(e^*) := \frac{1}{|e^*|}\sum_{l\in e^*} \mathrm{C}_l(e^*)$ denote the coefficient of variation, averaged across nodes participating in each hyperedge $e_* \in \mathcal{E}_c$. In Fig. 2a, $\bar{C}(e^*)$ is plotted versus the (average) number of (hyper) edges in $\mathbf{H}$ in which pairs of nodes contained in e_*, that is $(i, j) \subset e_*$, coappeared. The results are provided in the form of box plots. As expected, the uncertainty is lower for hyperedges whose comprising nodes featured a greater number of pairwise interactions in the training data. Similar observations can be made when considering the hyperedge level uncertainty $U(e^*)$; see Fig. 2b.

6. CONCLUSIONS

The present work introduced a novel approach to high-order link prediction. Gaussian process based learning was leveraged in order to build probabilistic models for the dependence structure of each node. Candidate hyperedge scores were obtained by fusing the predictions of node-wise models through a pseudolikelihood function. Finally, estimates of the uncertainty, both at the node and hyperedge level, were introduced. Tests on benchmark datasets demonstrated the superior hyperedge prediction performance of the novel approach, relative to alternatives, as well as its uncertainty quantification capabilities.

7. APPENDIX

At a high level, EP first considers an approximation of the likelihood as per $q(\mathbf{H}_{im}|[\mathbf{f}_i]_m) \approx \tilde{Z}_m^{(i)}\mathcal{N}([\mathbf{f}_i]_m; \tilde{\mu}_m^{(i)}, \tilde{v}_m^{(i)})$. EP then optimizes over parameters $\boldsymbol{\tau}_i := \{\tilde{\mu}_m^{(i)}, \tilde{v}_m^{(i)}, \tilde{Z}_m^{(i)}\}_{m=1}^M$ so that, for each m, the marginals $q([\mathbf{f}_i]_m|\mathbf{h}_i^\top, \mathbf{H}_{\setminus i}) \propto \int \mathcal{N}(\mathbf{f}_i; \mathbf{0}, \mathbf{K}_i) \prod_{m'=1}^M \tilde{Z}_{m'}^{(i)}\mathcal{N}([\mathbf{f}_i]_{m'}; \tilde{\mu}_{m'}^{(i)}, \tilde{v}_{m'}^{(i)}) d[\mathbf{f}_i]_{\setminus m}$ agree with marginals $\int \mathcal{N}(\mathbf{f}_i; \mathbf{0}, \mathbf{K}_i) p(\mathbf{H}_{im}|[\mathbf{f}_i]_m) \prod_{m'\neq m} \tilde{Z}_{m'}^{(i)} \mathcal{N}([\mathbf{f}_i]_{m'}; \tilde{\mu}_{m'}^{(i)}, \tilde{v}_{m'}^{(i)}) d[\mathbf{f}_i]_{\setminus m}$ whereby the exact likelihood $p(\mathbf{H}_{im}|[\mathbf{f}_i]_m)$ is used. This is accomplished by moment matching. The approximants $\{\mathbf{m}_i, \mathbf{A}_i\}$ are finally obtained based on $\boldsymbol{\tau}_i$; see [9] for a more elaborate description.

8. REFERENCES

[1] S. Barbarossa and M. Tsitsvero, "An introduction to hypergraph signal processing," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 6425–6429.

[2] A. R. Benson, R. Abebe, M. T. Schaub, A. Jadbabaie, and J. Kleinberg, "Simplicial closure and higher-order link prediction," *Proc. Natl. Acad. Sci.*, vol. 115, no. 48, pp. E11 221–E11 230, 2018.

[3] C. Berge, *Hypergraphs: Combinatorics of Finite Sets*. North-Holland, 1989, vol. 45.

[4] J. Besag, "Statistical analysis of non-lattice data," *Journal of the Royal Statistical Society Series D: The Statistician*, vol. 24, no. 3, pp. 179–195, 1975.

[5] G. V. Karanikolas, A. Pagès-Zamora, and G. B. Giannakis, "Higher-order link prediction via learnable maximum mean discrepancy," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023.

[6] E. D. Kolaczyk, *Statistical Analysis of Network Data: Methods and Models*. Springer, 2009.

[7] A. Kucukelbir, Y. Wang, and D. M. Blei, "Evaluating Bayesian models with posterior dispersion indices," in *Proc. of the 34th International Conference on Machine Learning*, 2017, pp. 1925–1934.

[8] T. Kumar, K. Darwin, S. Parthasarathy, and B. Ravindran, "HPRA: Hyperedge prediction using resource allocation," in *Proc. of 12th ACM conference on Web Science*, 2020, pp. 135–143.

[9] M. Kuss, C. E. Rasmussen, and R. Herbrich, "Assessing approximate inference for binary Gaussian process classification," *Journal of Machine Learning Research*, vol. 6, no. 57, pp. 1679–1704, 2005.

[10] H. Kwak and H. B. K. Jung, "Subgraph representation learning with hard negative samples for inductive link prediction," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 4768–4772.

[11] D. Liben-Nowell and J. Kleinberg, "The link prediction problem for social networks," in *Proc. of the 12th ACM International Conference on Information and Knowledge Management*, 2003, pp. 556–559.

[12] R. N. Lichtenwalter, J. T. Lussier, and N. V. Chawla, "New perspectives and methods in link prediction," in *Proc. of ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*, 2010, pp. 243–252.

[13] T. P. Minka, "A family of algorithms for approximate bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[14] H. Nickisch and C. E. Rasmussen, "Approximations for binary Gaussian process classification," *Journal of Machine Learning Research*, vol. 9, no. 67, pp. 2035–2078, 2008.

[15] P. Patil, G. Sharma, and M. N. Murty, "Negative sampling for hyperlink prediction in networks," in *Proc. of the 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2020, pp. 607–619.

[16] J. Quiñonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *Journal of Machine Learning Research*, vol. 6, no. 65, pp. 1939–1959, 2005.

[17] M. T. Schaub *et al.*, "Signal processing on higher-order networks: Livin' on the edge... and beyond," *Signal Processing*, vol. 187, p. 108149, 2021.

[18] M. W. Seeger, C. K. I. Williams, and N. D. Lawrence, "Fast forward selection to speed up sparse gaussian process regression," in *Proc. of the 9th International Workshop on Artificial Intelligence and Statistics*, 2003, pp. 254–261.

[19] R. R. Sokal and F. J. Rohlf, *Biometry*. W. H. Freeman and Company, 1995.

[20] Y. Tao, Y. Li, and Z. Wu, "Temporal link prediction via reinforcement learning," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 3470–3474.

[21] K. Tu, P. Cui, X. Wang, F. Wang, and W. Zhu, "Structural deep embedding for hyper-networks," in *Proc. of AAAI Conference on Artificial Intelligence*, 2018.

[22] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*. MIT press, 2006.

[23] Y. Xu, D. Rockmore, and A. M. Kleinbaum, "Hyperlink prediction in hypernetworks using latent social features," in *Proc. of Discovery Science*, 2013, pp. 324–339.

[24] M. Zhang, Z. Cui, S. Jiang, and Y. Chen, "Beyond link prediction: Predicting hyperlinks in adjacency space," in *Proc. of AAAI Conf. on Artificial Intelligence*, 2018.