Advances in Knowledge-Based and Intelligent Information and Engineering Systems
M. Graña et al. (Eds.)
IOS Press, 2012

doi: 10.3233/978-1-61499-105-2-1233

Latent Informative Links Detection

Liang Hu[1], Jian Cao[1], Guandong Xu[2], Zhiping Gu[3]

[1]Dept. of Computer Science & Engineering, Shanghai Jiaotong University, China
[2]School of Engineering and Science Victoria University, Australia
[3]Dept. of Electrical Engineering, Shanghai Technical Institute of Electronics & Information

[1]{lianghu,cao-jian}@sjtu.edu.cn
[2]guandong.xu@vu.edu.au
[3]raingzp@sohu.com

Abstract. Sometimes, explicit relationships between entities do not provide sufficient information or can be unavailable in the real world. Unseen latent relationships may be more informative than explicit relationships. Thereby, we provide a method for constructing latent informative links between entities, using their common features, where entities are regarded as vertices on a graph. First, we employ a hierarchical nonparametric model to infer shared latent features for entities. Then, we define a filter function based on information theory to extract significant features and control the density of links. Finally, a couple of stochastic interaction processes are introduced to simulate dynamics on the networks so that link strength can be retrieved from statistics in a natural way. In experiments, we evaluate the usage of filter function. The results of two examples based on mixture networks show how our method is capable of providing latent informative relationships in comparison to explicit relationships.

1 Introduction

Networks, such as World Wide Web, social networking sites (SNS), and bio-networks, have become integral parts of our everyday life. In order to understand such complex networks in-depth, it is necessary to reveal their subtle structures. Numerous research methods such as machine learning, graph theory, information theory, and statistics have been used to study complex networks [1]. Most current research on complex networks are based on the topology analysis of the network, e.g. the small world phenomenon [2]. All these methods are sensitive to link distribution because the complexity of a network is mainly determined by links other than vertices. In practice, links are often defined according to some intuitive relationships such as hyperlinks and citations. Unfortunately, the explicit links may not provide sufficient information and can even be unavailable in the real world. Some reasons are: link absence is ubiquitous, e.g. a web page can only accommodate a limited amount of hyperlinks and there is no guarantee two relevant pages are linked; the presence and strength of a link is volatile with given conditions, e.g. the partnerships, for researchers, may vary with different research topics. Sometimes, unseen or unformed links also play significant roles, e.g. two researchers may both focus on a common research topic but have never realized it;

learning of the latent relationship may help them to find each other for collaboration. Therefore, discovering latent informative links is an essential and meaningful task.

Bayesian network (BN) is applied to model the relationships among features, whose structure can be constructed by some learning methods [3, 4]. BN assumes that an underlying structure encodes the joint distribution of variables by conditional independence assertions. The data are generated over the joint distribution so that it may recover the structure from observed data. A BN can be viewed as an ecosystem where a set of dependent features resides, whereas in our model the observed data between entities are independent so that every entity can be regarded as an ecosystem. An event observed on a BN only has impact on intra-ecosystem features and will never propagate to other ecosystems, so structure learning methods for BN cannot help infer relationships over inter-ecosystems like our model.

Most current link prediction methods [5] predict unseen links by analyzing observed links, where the observed links often come from structural properties of networks [6], relational features inside relational databases or linked data on the web [7, 8]. As analyzed above, given a set of entities, neither may well-collected observed links be available nor does the relational structure across entities exist, since the observed data for each entity are independent. Under these conditions, none of these methods can work. Some recent works which predict links through infinite statistical models [9, 10] are most similar to our method. However, these prediction methods completely depend on observed links over latent features. In this paper, we provide a generative method to enable inferring linkages from observed data lack of well-collected observed links.

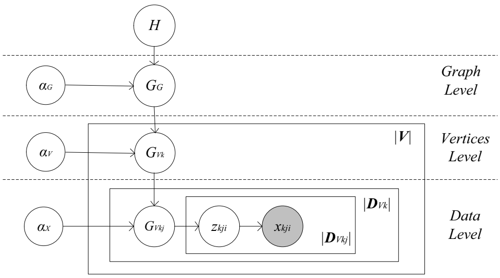
2 Latent Features Inference

Statistical models are often used to infer latent features for a set of entities from abundant observed data, e.g. Latent Dirichlet allocation (LDA) [11] is a representative and prevalent finite parametric model for document modeling and it has also been applied in many other areas, e.g. image processing [12]. However, choosing the number of mixture components is difficult for finite models and also for LDA [13]. In comparison, the nonparametric mixture model, typically, Dirichlet process (DP) model [14, 15] provides a solution to infer the number of latent components automatically, alleviating the difficulties of model selection. Our method would like to find common features shared across entities, but the utilization of independent DP for each entry are not qualified [16]. We hence resort to hierarchical models [16, 17], where the parent model in the hierarchies serves as a discrete base measure so that its child models may draw common atoms with a positive probability.

2.1 Model Construction

Since links in the real world are usually generated over common features that entities contain or work on, this assumption leads us to infer common features shared across entities by a hierarchical model. Recently, many hierarchical statistical models have been developed, typically, the hierarchical Dirichlet process (HDP) [16] which has been applied in many areas such as information retrieval, topic modeling, statistical genetics, and speech recognition [17].

We construct a three-level HDP to infer latent features as shown in Fig. 1. The top level is graph level, which provides a global view of all latent features on graph. The second level is vertices level, which focuses on features for each vertex. The bottom level is data level where the features for each observed data group can be retrieved.



$$G_G|\alpha_G, H \sim DP_G(\alpha_G, H)$$
$$G_{Vk}|\alpha_V, G_G \sim DP_{Vk}(\alpha_V, G_G) \qquad k = 1, \cdots, |\boldsymbol{V}|$$
$$G_{Vkj}|\alpha_X, G_{Vk} \sim DP_{Vkj}(\alpha_X, G_{Vk}) \qquad j = 1, \cdots, |\boldsymbol{D}_{Vk}|$$
$$z_{kji}|G_{Vkj} \sim G_{Vkj} \qquad i = 1, \cdots, |\boldsymbol{D}_{Vkj}|$$
$$x_{kji}|z_{kji} \sim F(z_{kji}) \qquad x_{kji} \in \boldsymbol{D}_{Vkj}$$

Fig. 1. The graphical model of 3-level HDP for inferring shared latent features

In Fig. 1, H is the prior distribution over all infinite latent features. G_G is distributed as a DP with concentration parameter α_G and base measure H. G_G is a global discrete measure over all latent features for the whole graph, which acts to tie together common features across the vertices. $|\boldsymbol{V}|$ is the number of vertices. Each G_{Vk} is distributed as a DP with concentration parameter α_V and G_G. G_{Vk} is a discrete measure over the latent features associated with vertex V_k. $|\boldsymbol{D}_{Vk}|$ is the number of training data groups for vertex V_k. Each G_{Vkj} is distributed as a DP with concentration parameter α_X and G_{Vk}. G_{Vkj} is a discrete measure over the latent features for a training data group $\boldsymbol{D}_{Vkj}$ at data level. Latent features z_{kji} are sampled from G_{Vkj}. $F(z_{kji})$ is the distribution of the observations x_{kji} given z_{kji}.

2.2 Model Inference

We present the inference based on the metaphor of Chinese restaurant franchise [16, 17]. First, we denote latent features $\boldsymbol{Z}_G$ for the graph, $\boldsymbol{Z}_{Vk}$ for each vertex V_k and $\boldsymbol{Z}_{Vkj}$ for every data group j associated with V_k. Then, let $\boldsymbol{\Theta}$ denote unique dishes (mixture components) served in the entire franchise, v^θ denote the number of tables serving dish θ ($\theta \in \boldsymbol{\Theta}$) in all vertices-level DPs, d_{Vk}^θ denote the number of tables serving dish θ in all DP_{Vk}'s child DPs at data level and n_{Vkj}^θ denote the number of customers (observations) eating dish θ in DP_{Vkj}. Analogous to inferring the posterior by HDP samplers [16, 17], a brief Gibbs sampling scheme is presented as follows:

— *Sampling weights for* $\boldsymbol{Z}_G, \boldsymbol{Z}_{Vk}, \boldsymbol{Z}_{Vkj}$:

$$\beta_G^{\theta 1}, \dots, \beta_G^{\theta N}, \beta_G^{\theta new}|\alpha_G, \boldsymbol{Z}_G \sim Dirichlet(v^{\theta 1}, \dots, v^{\theta N}, \alpha_G) \qquad N = |\boldsymbol{\Theta}|$$
$$\beta_{Vk}^{\theta 1}, \dots, \beta_{Vk}^{\theta N}, \beta_{Vk}^{\theta new}|\alpha_V, \boldsymbol{Z}_{Vk} \sim Dirichlet(\alpha_V \beta_G^{\theta 1} + d_{Vk}^{\theta 1}, \dots, \alpha_V \beta_G^{\theta N} + d_{Vk}^{\theta N}, \alpha_V \beta_G^{\theta new})$$
$$\beta_{Vkj}^{\theta 1}, \dots, \beta_{Vkj}^{\theta N}, \beta_{Vkj}^{\theta new}|\alpha_X, \boldsymbol{Z}_{Vkj} \sim Dirichlet(\alpha_X \beta_{Vk}^{\theta 1} + n_{Vkj}^{\theta 1}, \dots, \alpha_X \beta_{Vk}^{\theta N} + n_{Vkj}^{\theta N}, \alpha_X \beta_{Vk}^{\theta new})$$

– *Sampling* $z_{kji}, d_{Vk}^{\theta}, v^{\theta}$ *and update* $n_{Vkj}^{\theta}, \boldsymbol{\Theta}$

$z_{kji}|\alpha_X, x_{kji}, \{\beta_{Vkj}^{\theta}, n_{Vkj}^{\theta}\} \sim$

$$Discrete\left\{\left[\alpha_X\beta_{Vkj}^{\theta} + n_{Vkj}^{\theta} - \delta^{\theta}(x_{kji})\right]f_{-x_{kji}}^{\theta}(x_{kji})\right\}_{\theta=\theta_1,\dots,\theta_N,\theta_{new}}$$

where $n_{Vkj}^{\theta_{new}} = 0$, $f_{-x_{kji}}^{\theta}(x_{kji})$ is conditional density of x_{kji} under component θ given all observations except x_{kji}; $\delta^{\theta}(x_{kji}) = 1$ if x_{kji} belongs to θ and 0 else; After drawing dish for every x_{kji}, n_{Vkj}^{θ} needs to be updated; $\boldsymbol{\Theta}$ also should be updated if a new dish is added or some dish is deleted from the franchise.

$d_{Vk}^{\theta} = \sum_{j=1}^{|\boldsymbol{D}_{vk}|} d_{Vkj}^{\theta}$, where $d_{Vkj}^{\theta}|\alpha_X, \beta_{Vk}^{\theta}, n_{Vkj}^{\theta} \sim \left|CRP(\alpha_X\beta_{Vk}^{\theta}, n_{Vkj}^{\theta})\right|$, $CRP(\pi, n)$ is Chinese restaurant process with the concentration parameter π and n customers entering, d_{Vkj}^{θ} is the number of tables sampled from $CRP(\alpha_X\beta_{Vk}^{\theta}, n_{Vkj}^{\theta})$

$v^{\theta} = \sum_{k=1}^{|\boldsymbol{V}|} v_k^{\theta}$, where $v_k^{\theta}|\alpha_V, \beta_G^{\theta}, d_{Vk}^{\theta} \sim \left|CRP(\alpha_V\beta_G^{\theta}, d_{Vk}^{\theta})\right|$

After a sufficient burn-in period, a group of samples $\boldsymbol{S}$ are drawn. Then, we select a sample with N features, where N is the most frequent number of features occurring in $\boldsymbol{S}$, and then the weights for $\boldsymbol{Z}_{Vkj}, \boldsymbol{Z}_{Vk}, \boldsymbol{Z}_G$ are estimated as follows:

$$\widehat{w}_{Vkj}^{\theta} = n_{Vkj}^{\theta}/\sum_{i=1}^{N} n_{Vkj}^{\theta i},\ \widehat{w}_{Vk}^{\theta} = \hat{d}_{Vk}^{\theta}/\sum_{i=1}^{N}\hat{d}_{Vk}^{\theta i},\ \widehat{w}_G^{\theta} = \hat{v}^{\theta}/\sum_{i=1}^{N}\hat{v}^{\theta i} \quad (1)$$

where $\hat{d}_{Vk}^{\theta} = \sum_{j=1}^{|\boldsymbol{D}_{vk}|}\hat{d}_{Vkj}^{\theta}$, $\hat{d}_{Vkj}^{\theta} = \mathbb{E}\left[CRP(\alpha_X\beta_{Vk}^{\theta}, n_{Vkj}^{\theta})\right]$, $\hat{v}^{\theta} = \sum_{k=1}^{|\boldsymbol{V}|} v_k^{\theta}$, $v_k^{\theta} = \mathbb{E}\left[CRP(\alpha_V\beta_G^{\theta}, d_{Vk}^{\theta})\right]$, $\mathbb{E}[CRP(\pi, n)] = \sum_{k=1}^{n}\pi/(\pi + k - 1)$ for $n \geq 1$ and 0 else. Finally, the latent features for graph, vertices and data groups are denoted:

$$F_G = \sum_{\theta\epsilon\boldsymbol{\Theta}}\widehat{w}_G^{\theta}\delta(z_G^{\theta}), F_{Vk} = \sum_{\theta\epsilon\boldsymbol{\Theta}}\widehat{w}_{Vk}^{\theta}\delta(z_{Vk}^{\theta}), F_{Vkj} = \sum_{\theta\epsilon\boldsymbol{\Theta}}\widehat{w}_{Vkj}^{\theta}\delta(z_{Vkj}^{\theta}) \quad (2)$$

where $\widehat{w}^{\theta}\delta(z)$ represents weight $\widehat{w}^{\theta}$ focusing on feature z.

3 Latent Links Reconstruction

The most straightforward method to create links between two vertices is to check if existing common features. A single link cannot indicate which common features making two vertices connected so we need to construct multilink over each shared feature and label their strengths.

3.1 Link Inclination

We name the link constructed on a common feature θ as θ-oriented links. Intuitively, if an entity focuses attention on θ, it tends to build θ-oriented links with other entities which also highlight θ, even more so if they have parallel θ-relevant backgrounds (their θ-relevant features are similar). For example, if an economist needs to find partners who can help him with mathematical modeling, he is more willing to collaborate with a professor of mathematics with an economics background than a professor who is skillful in mathematical model construction for biology. Another factor affecting interaction is preference. That is, an entity may prefer to interact with some entities and decline to interact with others, e.g. due to a confidentiality agreement. Hereby, we measure the inclination of building a link from V_i to V_k by two parts: 1. the attraction for interaction; 2. the preference for interaction. With such assumptions, we formally define a quantified function to score the link inclination:

$$ls_\theta(V_i \to V_k) = \left[sig_\theta(V_i, V_k)\left(\alpha + \beta \cdot sim_\theta\left(rel_\theta(V_i), rel_\theta(V_k)\right)\right)\right] \cdot I_\theta(V_i, V_k) \quad (3)$$

where $sig_\theta(\cdot,\cdot)$ is a function to measure how significant the feature θ is to the given nodes; $rel_\theta(\cdot)$ is a function to retrieve θ-relevant features; $sim_\theta(\cdot,\cdot)$ is to measure similarity between two sets; α gives a base weight for $sig_\theta(\cdot,\cdot)$ and β is to control the proportion of $sim_\theta(\cdot,\cdot)$ to enhance $sig_\theta(\cdot,\cdot)$ function; $I_\theta(\cdot,\cdot)$ is to measure the interaction preference. For different applications, the link inclination function may have different definition. The implementation of this paper is given as follows:

$$sig_\theta(V_i, V_k) = f(w_i)f(w_k), \text{ where } w_i = w_{Vi}^{\theta}/max_{\theta i \in \boldsymbol{z}_{Vi}}\left(w_{Vi}^{\theta i}\right),\ f(w) = \frac{e^{-w}}{1+e^{-w}}$$

$$rel_\theta(V_k) = \boldsymbol{\eta}\left(z_{Vk}^{\theta}\right),\ sim_\theta\left(rel_\theta(a), rel_\theta(b)\right) = \frac{\left|\boldsymbol{\eta}\left(z_{Vi}^{\theta}\right) \cap \boldsymbol{\eta}\left(z_{Vk}^{\theta}\right)\right|}{\left|\boldsymbol{\eta}\left(z_{Vi}^{\theta}\right)\right|}$$

$$I_\theta(V_i, V_k) = \epsilon + log[\#(V_i \to V_k)]/log\left[max_{Vn \in \boldsymbol{V}}\left(\#(V_i \to V_n)\right)\right]$$

where $f(w_i)$ is the logistic function given the relative weight w_i of feature z_{Vi}^{θ} as input; $\boldsymbol{\eta}\left(z_{Vi}^{\theta}\right)$ is the direct neighbors of z_{Vi}^{θ} on a Chow-Liu tree [4] induced by mutual information between each pair of features of V_i, i.e. $MI\langle z_{Vi}^{\theta m}, z_{Vi}^{\theta n}\rangle_{\theta m \neq \theta n}$; ϵ of $I_\theta(V_i, V_k)$ is a constant to give a base measure and $\#(V_i \to V_k)$ is the times of interaction in history and $log\left[max_{Vn \in \boldsymbol{V}}\left(\#(V_i \to V_n)\right)\right]$ is used for normalization.

Given a latent feature θ, a network can be constructed by connecting each pair of vertices (a, b) with two links $a \to b$ and $b \to a$. Note that if selflinking is allowed, there is an additional selflink on each vertex. However , weak links are often negligible in the real world instead of such a complete graph. The inclination of building a link between vertices is measured by Eq. 3, so we can prune the links with low inclination scores. For example, we first computer the $ls_\theta(a \to b)$ of each pair of vertices, and then we order these scores so as to remove the lowest scored links with some given percentage. The pruned network is denoted as ψ_θ, named θ-oriented base network. Hence, we can obtain total N base networks, where $N = |\boldsymbol{Z}|$.

3.2 Dynamics on Mixture Network

The union of the base networks ψ_θ forms a multigraph, formally $\Psi = \bigcup_\theta \psi_\theta$, which is named Mixture Network. Now we need to determine the strength of each link on Ψ and the bundle of links from one node to another are the mixture of θ-oriented links union of all base networks, so the interaction process on Ψ need to involve the mixture of features. Normally, the formation of links in the real world is under the action of some long-run stochastic process. If we can simulate the dynamics on Ψ, then the statistics over interactions will show link strength. In this paper, we borrow the flow network to interpret the interactions on Ψ because each interaction can be regarded as a unit of flow and calculating the strengths of interactions is equivalent to retrieving the distribution of flow. Following we designed a stochastic interaction process (SIP) based on random walk.

Briefly, this SIP can be described a recursive 2-step interaction, *1st step*: a node first chooses a feature θ for interaction; *2nd step:* it choose an interaction target for θ (the choices of θ and targets are under some conditional distributions). If all nodes on Ψ follow such 2-step interaction constantly, the statistics of the flow distribution can be retrieved. To model such SIP and retrieve the long-run steady distribution $\bar{\boldsymbol{\varepsilon}}_{\Psi}$, we simulate the dynamics as a process of flow regulation system (FRS), where each node is regard as a switch and following tactics are always taken by FRS:

T1. Switch a outputs flow by two selective steps: 1. choosing a base network $\psi_\theta \sim f_{FRS}(\boldsymbol{\psi_L}|a)$ for outputting; 2. choosing the outflow from a to b on ψ_θ with a probability $P_\theta(b|a)$;

T2. If Switch a selects ψ_θ as the first step in ***T1***, but a has no outflow on ψ_θ, then a will transport flow to a uniformly random selected flow on ψ_θ;

T3. Each flow may be pumped into a uniformly random selected flow on some ψ_θ with a small probability λ, where $\psi_\theta \sim f_{FRS}(\boldsymbol{\psi_G})$.

In above tactics, ***T2*** can be viewed as a special case of ***T1***. In ***T1,T2***, the selection of base network ψ_θ for outputting flow is determined by each switch under conditional densities $f_{FRS}(\boldsymbol{\psi_L}|\cdot)$, so ***T1,T2*** are local regulation tactics taken by each switch. ***T3*** is a global regulation tactic; the system regulates all flows on Ψ with probability λ, and $f_{FRS}(\boldsymbol{\psi_G})$ is a global distribution for system to select a network ψ_θ for output. The definitions for $f_{FRS}(\boldsymbol{\psi_L}|\cdot)$ and $f_{FRS}(\boldsymbol{\psi_G})$ can be various, as shown later.

FRS implies the Markov chain model $\boldsymbol{\varepsilon}_\Psi = \boldsymbol{P}_\Psi \boldsymbol{\varepsilon}_\Psi$, where $\boldsymbol{\varepsilon}_\Psi$ is the flow distribution over all links of Ψ. For clarity, we can rewrite above Markov chain model as the form using block matrices:

$$\begin{bmatrix} \boldsymbol{\varepsilon}_{\psi_{\theta 1}} \\ \vdots \\ \boldsymbol{\varepsilon}_{\psi_{\theta N}} \end{bmatrix} = \left((1-\lambda) \left(\begin{bmatrix} \bar{\boldsymbol{P}}_{\psi_{\theta 1}} \\ \vdots \\ \bar{\boldsymbol{P}}_{\psi_{\theta N}} \end{bmatrix} + \begin{bmatrix} \mathbf{1}_{\psi_{\theta 1}} \boldsymbol{h}^T_{\psi_{\theta 1}} / \#E_{\psi_{\theta 1}} \\ \vdots \\ \mathbf{1}_{\psi_{\theta N}} \boldsymbol{h}^T_{\psi_{\theta N}} / \#E_{\psi_{\theta N}} \end{bmatrix} \right) + \lambda \begin{bmatrix} \mathbf{1}_{\psi_{\theta 1}} \boldsymbol{t}^T_{\psi_{\theta 1}} / \#E_{\psi_{\theta 1}} \\ \vdots \\ \mathbf{1}_{\psi_{\theta N}} \boldsymbol{t}^T_{\psi_{\theta N}} / \#E_{\psi_{\theta N}} \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{\varepsilon}_{\psi_{\theta 1}} \\ \vdots \\ \boldsymbol{\varepsilon}_{\psi_{\theta N}} \end{bmatrix}$$

where N is the number of base networks composing Ψ, $\#E_{\psi_{\theta i}}$ is the number of arcs on $\psi_{\theta i}$, $\#E_\Psi = \sum \#E_{\psi_{\theta i}}$ is the number of arcs on Ψ; $\boldsymbol{\varepsilon}_{\psi_{\theta i}}$ is a $\#E_{\psi_{\theta i}}$-dimensional vector to describe the flow volume of each link on $\psi_{\theta i}$; $\bar{\boldsymbol{P}}_{\psi_{\theta i}}$ is a $\#E_{\psi_{\theta i}} \times \#E_\Psi$ matrix, $\bar{\boldsymbol{P}}_{\psi_{\theta i}}(k \to l) = f_{FRS}(\psi_{\theta i}|v_{to}(k)) \cdot P_{\theta i}(v_{to}(l)|v_{to}(k))$ here $v_{to}(l)$ is a function to retrieve the to-node of the link l and $P_{\theta i}(v_{to}(l)|v_{to}(k)) = ls_\theta(v_{to}(l) \to v_{to}(k)) / \sum_{v \in \boldsymbol{V}_\theta} ls_\theta(v_{to}(l) \to v)$; $\mathbf{1}_{\psi_{\theta i}}$ is a $\#E_{\psi_{\theta i}}$-dimensional vector of ones; $\boldsymbol{h}^T_{\psi_{\theta i}}$ is a $\#E_\Psi$-dimensional row vector, $h^T_{\psi_{\theta i}}(k) = f_{FRS}(\psi_{\theta i}|v_{to}(k))$ if $v_{to}(k)$ has no outflow on $\psi_{\theta i}$ and $h^T_{\psi_{\theta i}}(k) = 0$ otherwise (see T2); $\boldsymbol{t}^T_{\psi_{\theta i}}(k) = f_{FRS}(\psi_{\theta i})$ (see T3). The transition probabilities for the Markov chain are aperiodic and irreducible under such construction [18], so a unique solution of the stationary distribution $\bar{\boldsymbol{\varepsilon}}_\Psi$ is guaranteed. There are various methods to solve such ergodic Markov chain problems [19, 20], e.g. the Power Method.

Trivial Mixture Network (TMN). If Ψ is the union of all base networks ψ_θ and $f_{FRS}(\psi_{\theta i}|V_k) = \widehat{w}^{\theta i}_{Vk} / \sum_{\theta \in \boldsymbol{Z}_{Vk}} \widehat{w}^{\theta}_{Vk}$ (local regulation tactics, biased with the features weights of each V_k), $f_{FRS}(\psi_{\theta i}) = \widehat{w}_G^{\theta i} / \sum_{\theta \in \mathbf{Z}_G} \widehat{w}_G^{\theta}$ (global regulation tactics, biased with the features weights of graph level), where $\mathbf{Z}_{Vk}$, $\mathbf{Z}_G$, $\widehat{w}_{Vk}^{\theta}$, $\widehat{w}_G^{\theta}$ are given in Eq. 2, then the mixture network derived under such setting is named Trivial Mixture Network, denoted as Ψ_{TMN}. Thus naming is based on the fact that the construction of Ψ_{TMN} only depends on the observed data of vertices. TMN depicts the intrinsic relations among all vertices.

Conditional Mixture Network (CMN). Given a mixture of features $\boldsymbol{F_Z} = \sum_{\theta \in \mathbf{Z}} w_c^{\theta} \delta(z^{\theta})$, we set $f_{FRS}(\psi_{\theta i} | V_k) = f_{FRS}(\psi_{\theta i}) = w_c^{\theta i} / \sum_{\theta \in \mathbf{Z}} w_c^{\theta}$ (the selection of output network is fully biased with the given condition $\boldsymbol{F_Z}$ for both local regulation tactics and global regulation tactics), where w_c^{θ} is the conditional weight of feature θ given in $\boldsymbol{F_Z}$. Thus constructed mixture network is named Conditional Mixture Network, denoted $\Psi | \boldsymbol{F_Z}$. CMN depicts the conditional interactions in the context of $\boldsymbol{F_Z}$.

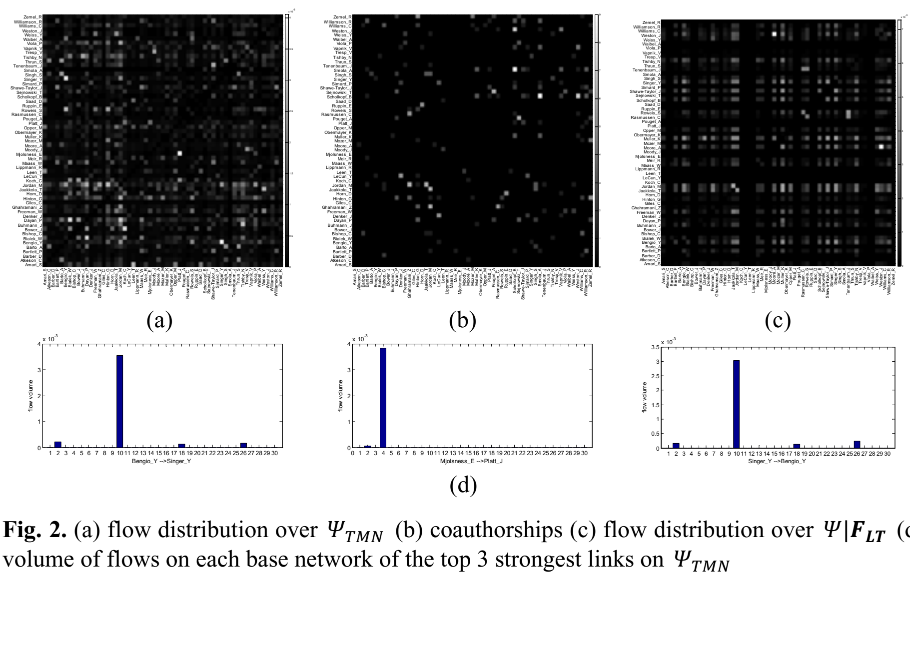
4 Experiments

We describe two experiments based on TMN and CMN respectively in this section. The NIPS dataset [21] is used for experiments, where we took the researchers who have published more than 10 papers on NIPS from 1987 to 2003..

Example for TMN over Researchers. In this example, we try to find latent links among researchers. The latent links between researchers may be coauthorships, citations, learning others' work, potential collaborations and so on. We searched the NIPS dataset, totally retrieving 57 authors who published more than 10 papers. These 57 authors are set as vertices; the observed data groups for each vertex are the papers the author published. The HDP used to infer research topics (latent features) has been depicted in Fig. 1, where H is symmetric Dirichlet distribution with hyperparameter 0.5 over research topics. For graph level DP, the concentration parameter α_G is sampled from $\Gamma(5,0.1)$ following the setting in [16]. 57 DPs in vertices-level are used to infer research topics of authors accordingly, where $\alpha_V \sim \Gamma(5,0.1)$. The data-level DPs infer the features for each paper, $\alpha_X \sim \Gamma(0.1,0.1)$. As to the link inclination function (Eq. 3), here we set $\alpha = 0.8$, $\beta = 0.2$ and selflinking is not allowed for the function. Each base network ψ_θ is to model the interactions over topic θ, which is constructed according to section 3.1. The construction of transition matrix for TMN follows section 3.2 and λ is set to 0.1. Finally, we took Power Method to compute out the solution of $\bar{\boldsymbol{\varepsilon}}_{\Psi_{TMN}}$.

Fig. 2(a) shows the latent interaction distribution over Ψ_{TMN}. The ith row depicts the states of latent interactions with all authors initialized by the ith author. The lighter an entry, the more latent interactions the author has, with the other author in the corresponding column. Fig. 2(b) shows the true coauthorships among 57 authors in the dataset; the lighter an entry, the more papers were coauthored. In comparison to the work for predicting coauthorships [10], Ψ_{TMN} enables to discover more underlying latent relationships other than explicit coauthorships. For instance, it can tell what research topics most probably cause the latent interactions between two authors and how much amount of interaction is over each topic; it also may help to find the most potential coauthors, where no coauthorship exists between two researchers but the latent interactions are strong; vice versa, it can detect anomalous coauthorships, where latent interactions are weak but true collaborations exist. As the interactions between two nodes are model by a mixture of links on Ψ_{TMN}, Fig. 2(d) illustrates the top 3 strongest interactions among authors, where the histograms display the volume of flows over each base network. Each research topic of the base network is indexed by a number as shown in Fig. 2 (d). Additionally, the predicted topological map of Ψ_{TMN} is demonstrated in Fig. 3, where the thicker the line, the more possible collaboration between two authors (a) and Fig.4 (a) illustrates the latent top 10 most popular authors (receiving the largest inflows volume) for collaboration in all given areas. **Example for CMN over Researchers.** In this example, we try to find latent interactions under the research area "Learning Theory (LT)". The training data of this area are all papers in the section of Learning Theory, NIPS 2001-2003. When constructing HDP, besides 57 DPs for authors, we add an additional DP into vertices level for inferring the latent features for LT and its child DPs at data level are used to model all the papers of LT accordingly. Other settings are the same as TMN. The inferred 57 mixtures of features for authors are still used to construct base networks $\boldsymbol{\psi}_{Z_{base}}$ while the latent features for LT ($\boldsymbol{F}_{LT}$) is set as the condition to construct the CMN $\Psi|\boldsymbol{F}_{LT}$.

Fig. 2(c) shows the result of $\Psi|\boldsymbol{F}_{LT}$, which demonstrates the latent interaction distribution under the research area LT. We find that it is apparently different from Ψ_{TMN} (Fig. 2(a)). TMN depicts the interactions biased with the research interests of authors themselves, and it covers all base networks without additional constraints. In contrast, $\Psi|\boldsymbol{F}_{LT}$ depicts the interactions constrained under the area LT where authors tend to have latent interactions with others who focus on the area LT (light columns). If some author has no research interest under LT, then he neither initializes interaction nor accepts interaction on $\Psi|\boldsymbol{F}_{LT}$ (both the row and column are dark). Akin to Ψ_{TMN}, the predicted topological map of Ψ_{CMN} is demonstrated in Fig. 3 (b) and Fig.4 (b) illustrates the latent top 10 most popular authors for collaboration in the area LT.

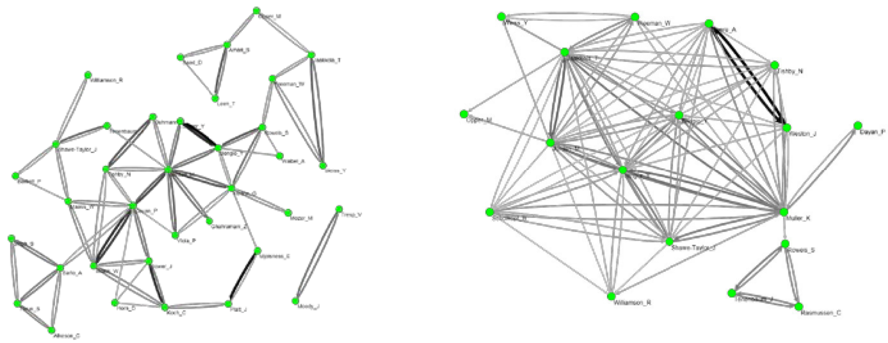


(d)

Fig. 2. (a) flow distribution over Ψ_{TMN} (b) coauthorships (c) flow distribution over $\Psi|\boldsymbol{F}_{LT}$ (d) volume of flows on each base network of the top 3 strongest links on Ψ_{TMN}

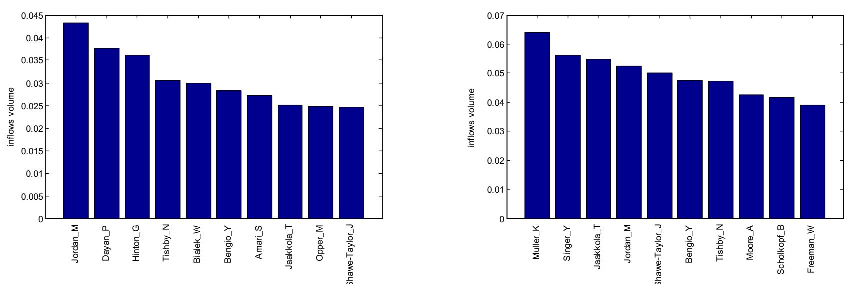
5 Discussion and Conclusion

We have presented the latent interaction network detection method which combines HDP and SIP. In this paper, we fully infer latent links from observed data, but the observed links may also help to improve performance, e.g. we can model preference distribution $I_\theta(\cdot,\cdot)$ in Eq. 3 over observed links. Moreover, the latent features are homogenous in this paper, e.g. topics and interests. However, the relationships in a complex system are always determined by various factors, so we must describe it using a group of heterogeneous features. HDP combined with India buffet process [22] may model such complex system as a hyper-mixture model over a set of heterogeneous mixtures. Apart from the examples given in experiments, we have been applying the reformed models in SNS, web services and bio-networks.



(a) (b)

Fig. 3. (a) Topological map of Ψ_{TMN}, only showing the top 100 strongest links; the thicker the line, the greater the volume of interactions; actually, each link on this map is composed of a mixture of links coming from each base network. This map does not show all of the 57 authors, where it only shows the authors connected by these top 100 strongest links. (b) Topological map of $\Psi|\boldsymbol{F_{LT}}$, only showing the top 100 strongest links; the thicker the line, the greater the volume of interactions under area LT. This map only shows the authors connected by these top 100 strongest links.



(a) (b)

Fig. 4. (a) Top 10 authors who receive the largest inflows volume on TMN (b) Top 10 authors who receive largest inflows volume on CMN in the area LT

6 References

1. Newman, M.: Networks: an introduction. Oxford Univ Pr (2010)
2. Newman, M.E.J., Watts, D.J.: The structure and dynamics of networks. Princeton Univ Pr (2006)
3. Heckerman, D.: A tutorial on learning with Bayesian networks. Innovations in Bayesian Networks 33-82 (2008)
4. Friedman, N., Geiger, D., Goldszmidt, M.: Bayesian network classifiers. Machine learning 29, 131-163 (1997)
5. Getoor, L., Diehl, C.P.: Link mining: a survey. ACM SIGKDD Explorations Newsletter 7, 3-12 (2005)
6. Liben-Nowell, D., Kleinberg, J.: The link-prediction problem for social networks. Journal of the American Society for Information Science and Technology 58, 1019-1031 (2007)
7. Popescul, A., Ungar, L.H.: Statistical relational learning for link prediction. In: IJCAI Workshop on Learning Statistical Models from Relational Data, pp. 172. (2003)
8. Taskar, B., Wong, M.F., Abbeel, P., Koller, D.: Link prediction in relational data. In: Advances in Neural Information Processing Systems (NIPS). (2003)
9. Kemp, C., Tenenbaum, J.B., Griffiths, T.L., Yamada, T., Ueda, N.: Learning systems of concepts with an infinite relational model. In: Proceedings of the 21st national conference on Artificial intelligence - Volume 1, pp. 381-388. AAAI Press, 1597600 (2006)
10. Miller, K.T., Griffiths, T.L., Jordan, M.I.: Nonparametric latent feature models for link prediction. In: Advances in Neural Information Processing Systems (NIPS), pp. 1276-1284. (2009)
11. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. The Journal of Machine Learning Research 3, 993-1022 (2003)
12. Fei-Fei, L., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: Computer Vision and Pattern Recognition, pp. 524-531. IEEE, (2005)
13. Griffiths, T.L., Steyvers, M.: Finding scientific topics. Proceedings of the National Academy of Sciences of the United States of America 101, 5228 (2004)
14. Ferguson, T.S.: A Bayesian analysis of some nonparametric problems. The annals of statistics 1, 209-230 (1973)
15. Escobar, M.D., West, M.: Bayesian Density Estimation and Inference Using Mixtures. Journal of the american statistical association 90, (1995)
16. Teh, Y.W., Jordan, M.I., Beal, M.J., Blei, D.M.: Hierarchical dirichlet processes. Journal of the american statistical association 101, 1566-1581 (2006)
17. Teh, Y.W., Jordan, M.I.: Hierarchical Bayesian nonparametric models with applications. Bayesian Nonparametrics 158 (2009)
18. Motwani, R., Raghavan, P.: Randomized algorithms. Cambridge Univ Pr (1995)
19. Langville, A.N., Meyer, C.D.: Deeper inside pagerank. Internet Mathematics 1, 335-380 (2004)
20. Stewart, W.J.: Introduction to the Numerical Solution of Markov Chains. Princeton University Press, Princeton (1994)
21. Globerson, A., Chechik, G., Pereira, F., Tishby, N.: Euclidean embedding of co-occurrence data. Journal of Machine Learning Research 8, 2265-2295 (2007)
22. Griffiths, T., Ghahramani, Z.: Infinite latent feature models and the Indian buffet process. In: Advances in Neural Information Processing Systems (NIPS). (2006)