Are latent feature/class models adapted for link prediction in social networks?

***Abstract*—**

I. Introduction

We provide formal definitions of fundamental properties of social networks which are design to be consistent with the probabilistic framework. We study those properties on two general models based on Bayesian nonparametric prior namely the Hierarchical Dirichlet Process (HDP) and the Indian Buffet Process (IBP). We show the relation between those properties and the models. Thus it provides a better comprehension of the models and their limitation in order to capture those properties in a learning problem. Additionally, we propose an adaptation of priors which gives a better interpretation of models in terms of assumptions on social networks and lead to better prediction performance.

Recently, several complex Bayesian models based on latent variables to explain the structure of social networks have been introduced [mmsb, ilfrm, etc]. This work was mainly evaluated on prediction tasks, such as link prediction or communities detection. However, few works have been done concerning the study of the intrinsic capacity of the models to model basic properties that arise in social networks, such as the dynamics of degree distribution, known to exhibit the preferential attachment effect [barabasi, web..] or the homophily effect[ref].

(++ Indeed the most heavily studied properties in social networks was the degree distribution and the mixing pattern (homophily/assortativity) tableaux !)

(++ not clear consensus of the formalism of properties and their evaluation, and whatsoever for the homophily property, the feature the definition are usually for single attribute... We consider a general vector . (with a measure working for both latent and real features)

(++ Probabilistic models we are interested in provide two ways of representing the data or network. One fall in the paradigm of mixture models and the other in the latent feature modeling. A motivation of those two modeling paradigm is that they are consistent with two key nonparametric prior for discrete data, namely the Dirichlet process (DP) and the the Indian Buffet Process (IBP). Many baysian model can be view as equivalent to truncated models with nonparametric priors. This provide a motivation to study those models. Furthermore, they are used as priors to generate latent features, either as proposition vector (class/DP) or binary vector (feature/IBP). It is admitted that those priors gives bursty features [accounting for burstiness in topic model]. We seek to clarify why this is true and how the burstiness can propagate at the degree level.

In the next section we will, first, explain the mathematical background in a machine learning context. Secondly, we will review the models of interest for dyadic data. Then, we will introduce the formal definition of properties of interest in social networks within the Bayesian frameworks, and how this is translated in terms of assumptions within Bayesian priors. Finally, we will show empirical results (on synthetic and real datasets) to support our claims.

II. Background

Without loss of generality, we focus on social networks with binary relationships. Our object of interest is the topology of the network representing the presence or absence of links between nodes in the graph. The network can be either directed or not. For a network with N nodes, we represent the topology by an adjacency matrix $Y \in \{0,1\}^{N\times N}$ associated to a graph $G = (V, E)$, where V is a set of nodes representing entities, $E \in V \times V$ is a set of edges who represents relationships between pairs of entities. From a probabilistic point of view, the network topology is modeled using a kernel with a Bernoulli density. The parameters of the Bernoulli is the probability to observe a link between two nodes.

We define a matrix of weight interactions $\Phi \in W^{K\times K}$ with W the space of weights, where K is the number of classes or features. Let $\Theta \in \mathcal{F}^{N\times K}$, be a matrix where each row i represents the latent feature vector associated to the node i, and $\mathcal{F}$ the latent feature space. Hence for the MMSB and ILFM, the latent feature vectors are respectively proportion vectors (who sum to one) and binary vectors. In this framework the network is generated with the following density:

$$Y \sim \mathrm{Bern}(\sigma(\Theta\Phi\Theta^T)) \tag{1}$$

where σ is a function that map values to a probability space. When σ is the identity function, the expectation of the observation reduces to a matrix factorization (bilinear) expression, and is related to Discrete Component Analysis (DCA) [**?**]:

$$E_{y\sim p(y|\Theta,\Phi)}[Y] = \Theta\Phi\Theta^T \tag{2}$$

This matrix factorization approach of the Bayesian model is in due to the likelihood of the model when applying the sum rule over the latent variables. Indeed the probability to have a link for the interaction (i, j) is:

$$\mathbf{p}(y_{ij} = 1 \mid \Theta, \Phi) = \sum_{k,k'} \mathbf{p}(y_{ij} = 1 \mid \phi_{k,k'})\mathbf{p}(k \mid \theta_i)\mathbf{p}(k' \mid \theta_j) \tag{3}$$

The questions that arise are:

- What kind of properties the model can capture or learn on networks ?
- Which constraint on the models can come with an consistent interpretation of latent variables along with the concepts of communities structure and homophily in social networks ?

In the next session we review the models of interest.

III. Models

As mentioned before, we focus in this study on two major representatives of the latent models used for link prediction in social networks, namely the latent feature model [**?**] and the mixed-membership stochastic block model [**?**]. To be as general as possible, we consider non-parametric extensions of these models, respectively based on the Indian Buffet Process (IBP) and the Hierarchical Dirichlet Process (HDP). Similar extensions have already been considered in the past, *e.g.* through the Infinite Latent Feature model [**?**] and through conditional random fields [**?**] or a dynamic version of the Hierarchical Dirichlet Process [**?**].

We now briefly describe the two models retained.

A. Infinite Latent Feature Model (ILFM)

In the latent feature model, each node is represented by a vector of binary features. The probability of linking two nodes is then based on a weighted similarity between their feature vectors, the weight matrix being generated according to a normal distribution. In its non-parametric version, the feature vectors are now generated according to an IBP, leading feature vectors of infinite dimensions (even though only a finite number of dimensions are actually active). The following steps summarizes this process:

1) Generate a feature matrix $\mathbf{F}_{N\times\infty}$ representing the feature vector of each node: $\mathbf{F} \sim \text{IBP}(\alpha)$
2) Generate a weight matrix for each latent feature: $\phi_{mn} \sim N(0, \sigma_w)$, $m, n \in \mathbb{N}^{+*}$
3) Generate or not a link between any node i and any node j according to:

$$y_{ij} \sim \text{Bern}(\sigma(\mathbf{f}_i \mathbf{\Phi} \mathbf{f}_j^\top)) \tag{4}$$

where $\sigma()$ is the sigmoid function, mapping $[-\infty, +\infty]$ values to [0,1], and where y_{ij} is a binary variable indicating that a link has been generated ($y_{ij} = 1$) or not ($y_{ij} = 0$). We will denote by $\mathbf{Y}$ the $N \times N$ matrix with elements y_{ij}. Finally, $\mathbf{f}_i$ denotes the row vector corresponding to the i^{th} row of $\mathbf{F}$.

This model makes use of two real hyper-parameters, one for the IBP process (α), and one for the variance of the normal distribution underlying the weight matrix (σ_w). In the case of undirected networks, the matrices $\mathbf{Y}$ and $\mathbf{\Phi}$ are symmetric and only their upper (or lower) diagonal parts are generated. Lastly, both $\mathbf{F}$ and $\mathbf{\Phi}$ are infinite matrices. In practice however, one always deal with a finite number of latent features. A graphical representation of this model is given in Figure 1.

Standard Gibbs sampling and Metropolis-Hastings algorithms can be used for inference in this model. We do not detail them here and refer the interested reader to [**?**].

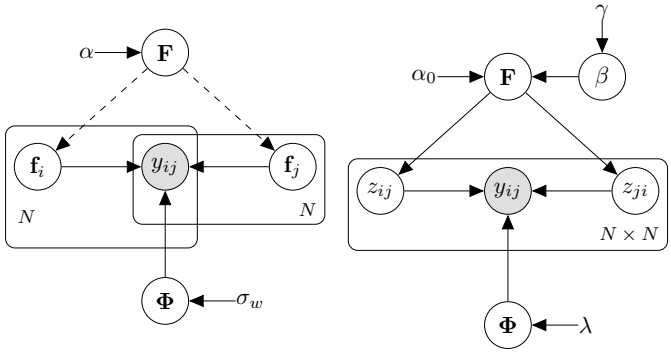

Fig. 1. The two graphical representations of (left) the latent feature model and (right) the latent class model. The difference between the two models lies in the way representations are associated to nodes: a fixed representation is used in the case of the latent feature model, whereas the representation in the latent class model varies according to the link considered.

B. Infinite Mixed-Membership Stochastic Block Model (IMMSB)

The MMSB model generates class membership distributions per node on the basis of a Dirichlet distribution. Then, for each connection between two nodes, a particular class for each node is first sampled from the class membership distribution, and the probability of connecting the two nodes is, as in the previous model, based on a Bernoulli distribution integrating the weight of the two classes.

The non-parametric version parallels this development but considers, in lieu of the Dirichlet distribution, a hierarchical Dirichlet process, leading to the following generative model:

1) Generate the class membership matrix $\mathbf{F}_{N\times\infty}$:

$$\begin{aligned} \beta &\sim \text{GEM}(\gamma) \\ \mathbf{f}_i &\sim \text{DP}(\alpha_0, \beta) \quad \text{for } i \in \{1, .., N\} \end{aligned}$$

2) Generate a weight matrix for each latent class:

$$\phi_{mn} \sim \text{Beta}(\lambda, 1),\ m, n \in \mathbb{N}^{+*}$$

3) For any node i and any node j, choose a class from their class membership distribution and generate or not a link according to:

$$\begin{aligned} z_{ij} &\sim \text{Cat}(\mathbf{f}_i) \\ z_{ji} &\sim \text{Cat}(\mathbf{f}_j) \\ y_{ij} &\sim \text{Bern}(\phi_{f_{ij} f_{ji}}) \end{aligned} \tag{5}$$

We have this time three real hyper-parameters, two for the hierarchical Dirichlet process (γ and α_0) and one for the weight matrix (λ). Adrien, peux-tu vérifier la paramétrisation de la loi Beta ? As for the previous model, in the case of undirected networks, the matrices $\mathbf{Y}$ and $\mathbf{\Phi}$ are symmetric and only their upper (or lower) diagonal parts are generated; as before again, both $\mathbf{F}$ and $\mathbf{\Phi}$ are infinite matrices

The updates can also be obtained via Gibbs sampling and Metropolis-Hastings procedures. Such updates can be found in [**?**] for the parameters β and $\mathbf{F}$, and in [**?**] for $\mathbf{\Phi}$. We provide below the update for f_{ij}, which we will use in the next sections and which is detailed in Appendix **??**:

1) If the class k has already been observed:

$$\mathbf{p}(z_{ij} = k \mid \mathbf{F}^{-ij}) \propto N_{ik}^{-ij} + \alpha_0 \beta_k \tag{6}$$

2) In case of a new class k_n:

$$\mathbf{p}(z_{ij} = k_n \mid \mathbf{F}^{-ij}) \propto \alpha_0 \beta_{k_n}$$

C. Model Comparison

Both ILFM and IMMSB are based on a latent representation of each node in the network (matrix $\mathbf{F}$). However, in the case of ILFM, this representation takes the form of a binary vector, whereas in IMMSB it is a vector of proportion. Interpreting the latent dimensions as characteristics or classes, one can view ILFM as performing a hard assignment on those classes whereas IMMSB performs a soft assignment.

Another distinction between the two lies in the use of the sigmoid function in ILFM to obtain the parameter of a Bernoulli distribution from the latent representations and their weights (or correlations). Because of that, the weight matrix $\mathbf{\Phi}$ can take on a very general form in this model and can easily be generalized to a multivariate distribution. This is not the case in IMMSB where the elements of $\mathbf{\Phi}$ should lie in the interval $[0; 1]$.

Lastly, a major difference lies in the fact that the complete latent representation of each node is used in ILFM to generate a link, whereas in IMMSB, for each link to be generated, one first selects one component from the latent representations of the nodes involved in the link. This allows one to capture the fact that different classes may explain different links for the same node. At the same time, it has an impact on the homophily effect, as shown in the next section. Note that ILFM is also able to capture the fact that different classes may explain different links for the same node, even though more implicitly, by relying on different dimensions of the latent representations.

Say a few words on the complexity and inference time of each model

IV. Homophily
Birds of a feather flock together

Homophily refers to the tendency of individuals to connect to similar others: two individuals (and thus their corresponding nodes in a social network) are more likely to be connected if they share common characteristics [**?**], [**?**]. The characteristics often considered are inherent to the individuals: they may represent their social status, their preferences, their interest, ... A related notion is the one of *assortativity*, which is slightly more general as it applies to any network, and not just social networks, and refers to the tendency of nodes in networks to be connected to others that are similar in some way.

A definition of homophily has been proposed in [**?**]. However, this definition, which relies on a single characteristic (as age or gender), does not allow one to assess whether latent models for link prediction capture the homophily effect or not. We thus introduce a new definition of homophily below, which directly aims at this:

Definition IV.1 (Homophily)**.** *Let $\mathcal{M} = \{\mathbf{F}, \mathbf{\Phi}\}$ be a link prediction model as defined above and s a similarity measure between nodes. We say that $\mathcal{M}$* captures the homophily effect *iff, $\forall (i, j, i', j') \in V^4$:*

$$s(i,j) > s(i',j') \implies \mathbf{p}(y_{ij} = 1 \mid \mathcal{M}) > \mathbf{p}(y_{i'j'} = 1 \mid \mathcal{M})$$

A model which verifies this condition is said to be strongly homophilic.

As one can note, this definition directly captures the effect "if two nodes are more similar, then they are more likely to be connected". The model $\mathcal{M}$ considered can either be regarded in a purely generative setting, or be learned from the data. The above definition encompasses both cases, the parameters $\mathbf{F}$ and $\mathbf{\Phi}$ being either generated or learned. Furthermore, in some cases, for example when one has access to users' interest or social information, one may use *a priori* representations of nodes as well as given weight matrix between them. This situation corresponds to the case where the matrices $\mathbf{F}$ and $\mathbf{\Phi}$ are given, the variables y_{ij} still being generated according to Eqs. 4 and 5. The definition of homophily given above is still applicable in this setting.

The similarity function assesses to which extent two nodes share the same characteristics. The parameters $\mathbf{F}$ and $\mathbf{\Phi}$ respectively capture latent characteristics of nodes and weights (or correlations) between the dimensions of these representations; a natural similarity between node characteristics is thus provided by:

$$s(i,j) = \mathbf{f}_i \mathbf{\Phi} \mathbf{f}_j^\top \tag{7}$$

From this, one can state the following property:

Proposition IV.1. *ILFM is strongly homophilic under the similarity measure defined by Eq. 7.*

The proof of this proposition is straightforward, and directly stems from the fact that the sigmoid function is strictly increasing.

The situation for IMMSB is slightly different inasmuch as the selection of classes from the latent class representations may lead to reversing the order between two pairs of nodes for the similarity and the probability of connecting the nodes. Indeed, IMMSB is *not* strongly homophilic. This said, IMMSB still captures an homophily effect in the following sense:

Proposition IV.2. *Let s be the similarity measure defined by Eq. 7. IMMSB is* weakly homophilic*, i.e. $\forall (i, j, i', j') \in V^4$:*

$$s(i,j) > s(i',j') \implies$$
$$\mathbb{E}_{\mathbf{p}(z_{ij}, z_{ji} \mid \mathcal{M})}[\mathbf{p}(y_{ij} = 1 \mid \mathcal{M})] > \mathbb{E}_{\mathbf{p}(z_{ij}, z_{ji} \mid \mathcal{M})}[\mathbf{p}(y_{i'j'} = 1 \mid \mathcal{M})]$$

The proof of this proposition is also straightforward and stems from the fact that:

$$\begin{aligned}
&\mathbb{E}_{\mathbf{p}(z_{ij}, z_{ji} \mid \mathcal{M})}[\mathbf{p}(y_{ij} = 1 \mid \mathcal{M})] \\
&= \textstyle\sum_{k,k'} \phi_{k,k'} \mathbf{p}(z_{ij} = k \mid \mathcal{M}) \mathbf{p}(z_{ji} = k' \mid \mathcal{M}) \\
&= \textstyle\sum_{k,k'} \phi_{k,k'} f_{ik} f_{jk'} = \mathbf{f}_i \mathbf{\Phi} \mathbf{f}_j^\top
\end{aligned}$$

where f_{ik} denotes, as before, the element of $\mathbf{F}$ at the i^{th} row and k^{th} column.

The above development shows that both ILFM and IMMSB can capture the homophily effect, with a strict adherence to it

in the case of ILFM, and a looser one for IMMSB. We now turn to the burstiness effect.

V. Burstiness

The rich get richer and the poor get poorer

The preferential attachment states that a node is more likely to create connections with nodes having a high degree. To take into account this behavior, in the BarabasiAlbert (BA) model, each node is connected to an existing node with a probability proportional to the number of links of the chosen node. This leads to scale-free networks, characterized by a degree distribution with a heavy tail which can be approximated by a power law distribution such that the fraction of nodes $\mathbf{p}(d)$ having a degree d follows $\mathbf{p}(d) \sim d^{-\gamma}$ where γ ranges typically between 2 and 3 [**?**]. An equivalent notion is the burstiness, studied by [**?**], which conveys the same idea : rich get richer or the more you have, the more you will get. In [**?**], a formalized definition has been proposed. According to the authors:

Definition V.1 (Burstiness)**.** *A discrete distribution* $\mathbf{p}$ *is bursty if and only if for all integers* (n, n')*,* $n \geq n'$ *:*

$$\mathbf{p}(d \geq n' + 1 \mid d \geq n') > \mathbf{p}(d \geq n + 1 \mid d \geq n) \tag{8}$$

A distribution which verifies this condition is said to be bursty.

In [**?**], this definition has been generalized to the continuous case but, in the sequel, we will retain this first definition since we focus on discrete distributions.

The burstiness can appear for different various variable in a model. In this paper we consider three different schemes at the node and feature level, that constitute some basic topology assumptions on networks:

Proposition V.1.
for all $i, j \in V^2$ *and* $k \in \{1, .., K\}$*, we have:*

- *Preferential Attachment: the distribution of degree* d_i *is bursty iff* f_i *is a stricly increasing function of n with:*

$$f_i(n) = \mathbf{p}(y_{ij} = 1 \mid d_i = n)$$

- *Local Preferential Attachment: Given a class couple* $c = \{k, k'\} \in \{1, .., K\}^2$*, and a degree restricted to nodes who belongs to this interaction couple* d_{ic}*, the distribution of degree* d_{ic} *is bursty iff* $f_{i,c}$ *is a stricly increasing function of n with:*

$$f_{i,c}(n) = \mathbf{p}(y_{ij} = 1 \mid d_i = n, c)$$

- *Feature burstiness (block/class burstiness ?:s): the distribution over the number of membership of each class* $\theta_{.k}^{-ik}$ *is bursty iff* f_k *is a stricly increasing function of n with:*

$$f_k(n) = \mathbf{p}(\theta_{ik} \mid \theta_{.k}^{-ik} = n)$$

We justify this approach in the supplementary materials **??**. One can see that the approach makes the link with the classical definition of preferential attachment. Furthermore one can see that the similaraty between th functions we track $(f_i, f_{i,c}, f_k)$ and the typical Gibbs updates. The difference is that we want to assess the generative model given the data and parameters of the model (ie the model has converged). Hence we are looking if the topological property of burstiness can be handle by the model once it learned from the data.

A. Burstiness for ILFM

In this model, the weight interaction matrix Φ are not conjugate of the likelihood. Thus it can not be integrated out into a closed form expression. As a matter of simplicity we consider this parameter as known, and omit it the following conditional distributions.

Let $F = \Theta$ and $W = \Phi$, each node i has a fixed feature vector noted F_i and a weighed interactions matrix W. In this case, the function f_i is:

$$\mathbf{p}(y_{ij} = 1 \mid d_i, F^{-i\cdot}) = \sum_{F_i} \mathbf{p}(y_{ij} = 1 \mid d_i, F^{-i\cdot} F_i,)\mathbf{p}(F_i \mid d_i, F^{-i\cdot}) \tag{9}$$

$$= \sum_{F_i} \sigma(F_i W F_j^T) \frac{\mathbf{p}(d_i \mid F^{-i\cdot}, F_i)\mathbf{p}(F_i \mid F^{-i\cdot})}{\mathbf{p}(d_i \mid F^{-i\cdot})} \tag{10}$$

$$\propto \sum_{F_i} \prod_{j' \in \mathcal{V}(i) \cup j} \sigma(F_i W F_{j'}^T) \prod_{j' \notin \mathcal{V}(i)} 1 - \sigma(F_i W F_{j'}^T) \prod_k \frac{m_{ik}}{N} \tag{11}$$

The term $\prod_k \frac{m_{ik}}{N}$ latter equation comes from the conditional probability of a feature f_{ik} for an IBP prior and applying a chain of product rule. The product is the only term that depend on d_i and we refer to it as $f(d_i)$. Under the assumption that all observed links have higher probability than the observed non-links to bind, we can choose an index dictionary g to reorder the terms such as:

$$\underbrace{\sigma(F_i W F_{g(1)}^T) \geq .. \geq \sigma(F_i W F_{g(p)}^T)}_{g(.) \in \mathcal{V}(i) \cup j} \geq \underbrace{.. \geq \sigma(F_i W F_{g(N)}^T)}_{g(.) \notin \mathcal{V}(i)} \tag{12}$$

We then have with some regularity:

$$f(d_i) \geq \sigma(F_i W F_{g(p)}^T)^{d_i + 1} (1 - \sigma(F_i W F_{g(p+1)}^T))^{N - d_i} \tag{13}$$

$$log(f(d_i)) \geq d_i log(\frac{\sigma(F_i W F_{g(p)}^T)}{1 - \sigma(F_i W F_{g(p+1)}^T)}) + cst \tag{14}$$

Reste a valider 2 point:

- Le passage la proportion
- Borner $log(f(d_i))$ entre deux fonction croissantes et montrer qu'on oscille pas l'intrieur ?!

Then a sufficient condition for the burstiness is to have: $\sigma(F_i W F_{g(p)}^T) > 1 - \sigma(F_i W F_{g(p+1)}^T)$. If σ is the sigmoid this is equivalent ot have $F_i W F_{g(p)}^T > -F_i W F_{g(p+1)}^T$.

In other term to enable burstiness in the ILFM, the model need to ensure some deterministic behavior when regarding the realization of outcomes and there actual distribution.

B. Burstiness for MMSB

In the latent class models each dyads has two underlying class assignments for each node of the couple. We note $Z \in N \times N \times 2$ the matrix that represents those class assignments. We seek for the following form of the likelihood, that we marginalize over all the possible couples classes $c = (k, k')$:

$$\mathbf{p}(y_{ij} = 1 \mid Y^{-i\cdot}, Z^{-ij}, d_i) = \sum_{c=(k,k')} \mathbf{p}(y_{ij} = 1 \mid Y^{-i\cdot}, d_i, c)\mathbf{p}(c \mid Z^{-ij}) \tag{15}$$

Here note that within the sum, the left hand term is conditionally independent of Z^{-ij}. And the right hand term is independent of the adjacency terms $Y^{-i\cdot}$ since it do not belongs to the Markov blanket of c random variable.

The first term is the likelihood for the links between (i, j) given the class of each node (k, k'). Due to the Beta-Bernoulli conjugacy of the model, ϕ and θ can be marginalized out, and it simplify to:

$$\mathbf{p}(y_{ij} = 1 \mid Y^{-i\cdot}, d_i, c) = \frac{C_{c1}^{-i\cdot} + d_{ic} + \lambda_1}{C_{c\cdot}^{-ij} + \lambda_0 + \lambda_1} \tag{16}$$

Where C_{c1} denotes the count matrix for all interactions having value 1 (link present) with the classes couple being $c = (k, k')$. Thus $C_{c1} = \sum_{i,j} \mathbf{1}(z_{i\to j} = k, z_{i\leftarrow j} = k', y_{ij} = 1)$ and $C_{c\cdot} = \sum_{i,j} \mathbf{1}(z_{i\to j} = k, z_{i\leftarrow j} = k')$

We recognize the likelihood form of the Gibbs update [**?**], except that we isolate the term depending of the degree on i, d_i. Hence the term d_{ic} is the element of the degree with a classes couple $c = (k, k')$ and $d_{ic} = \sum_{j' \neq j} \mathbf{1}(z_{i\to j'} = k, z_{i\leftarrow j'} = k', y_{ij'} = 1)$.

The second term of equation (15), can be rewrited by noting that the classes of the couple c are independent and that the term $Y^{-j\cdot}$ can be dropped because it is not present in the Markov blanket of the class assignment:

$$\mathbf{p}(c \mid Z^{-ij}) = \mathbf{p}(z_{i\to j} = k \mid Z^{-ij})\mathbf{p}(z_{i\leftarrow j} = k' \mid Z^{-ij}) \tag{17}$$

Again, the two members of the right hand equation (17) are the Gibbs updates for the topic assignments of nodes for the interaction (i, j). Both members reduce to simple form due to the conjugacy between the Dirichlet and Multinomial [**?**] or concurrently from the Chinese Restaurant Franchise [**?**]:

$$\mathbf{p}(z_{i\to j} = k \mid Z^{-ij}) = \frac{N_{ik}^{-ij} + \alpha_k}{N_{i\cdot}^{-ij} + \alpha_\cdot} \tag{18}$$

$$\mathbf{p}(z_{i\leftarrow j} = k' \mid Z^{-ij}) = \frac{N_{jk'}^{-ij} + \alpha_{k'}}{N_{j\cdot}^{-ij} + \alpha_\cdot} \tag{19}$$

Finally, one can see that the only term depending on the degree d_i is isolated, and we can rewrite equation (15), with term depending only on d_i, k, i and j:

$$\mathbf{p}(y_{ij} = 1 \mid Y^{-i\cdot}, Z^{-ij}, d_i) = \sum_{c=(k,k')} A_c(B_c + d_{ic}) \tag{20}$$

Where A_c and B_c are two positive function of c.

$$A_c = \frac{N_{ik}^{-ij} + \alpha_k}{N_{i\cdot}^{-ij} + \alpha_\cdot} \frac{N_{jk'}^{-ij} + \alpha_{k'}}{N_{j\cdot}^{-ij} + \alpha_\cdot} \frac{1}{C_{c\cdot}^{-ij} + \lambda_0 + \lambda_1} \tag{21}$$

$$B_c = C_{c1}^{-i\cdot} + \lambda_1 \tag{22}$$

As we sum over all possible couple classes, the probability to have a link will augment with the degree with the classes couple corresponding to the element of the degree with the same couple. Hence the probability to observe a link for node i is strictly crescent with his degree d_i.

a) Preferential Attachment:

The model is bursty hence it can handle the preferential attachment at the network level.

b) Local Preferential Attachment:

The Local preferential attachment is similar to the notion of burstiness but inside a community/class of the network. Assuming that we know the class of i $z_{i\to j}$ to be k, the probability to have a link becomes:

$$\mathbf{p}(y_{ij} = 1 \mid Y^{-i\cdot}, d_i, z_{i\to j}) = \sum_{k'} \mathbf{p}(y_{ij} = 1 \mid Y^{-i\cdot}, Z^{-ij}, d_i, c = (k, k'))\mathbf{p} \tag{23}$$

$$= \sum_{k'} \frac{C_{c1}^{-i\cdot} + d_{ic} + \lambda_1}{C_{c\cdot}^{-ij} + \lambda_0 + \lambda_1} \frac{N_{jk'}^{-ij} + \alpha_{k'}}{N_{j\cdot}^{-ij} + \alpha_\cdot} \tag{24}$$

$$= \sum_{k'} A'_{k'}(B'_{k'} + d_{i(k,k')}) \tag{25}$$

Here the probability increases with the degree independently of the interactions classes. This means that burstiness is possible inside but also between communities.

c) Communities Distribution:

....Need to count the table for each classes in Chinese Restaurant Franchise (CRF), to evaluate the distribution according to the hyperprior of HDP...

VI. Experimental study

To validate our theoretical results we fitted our models on synthetic networks and track how well we can reproduce the properties of interest on a generated network.

The synthetic network has 1000 nodes and 4 communities and a density of 0.05. [See the ref of the generator for the ground true on the preferential attachment effect...]

VII. Homophily Indicator

We consider a social network defined as an attributed graph G = (V,E), where V is a set of N nodes representing entities, $E \in V \times V$ is a set of m edges representing relationships between pairs of entities. Each node $i \in V$ is described by K features and s is a similarity function which allows to compare two vertices according to their features. We consider that two vertices are similar, denoted s(x,y), if s(x,y) is lower than a threshold.

Given a contingency table defined as follows:

$a = Card\{(x,y) \in V \times (V-1)/(x,y) \in E \wedge s(x,y)\}$
$b = Card\{(x,y) \in V \times (V-1)/(x,y) \in E \wedge \neg s(x,y)\}$
$c = Card\{(x,y) \in V \times (V-1)/(x,y) \notin E \wedge s(x,y)\}$
$d = Card\{(x,y) \in V \times (V-1)/(x,y) \notin E \wedge \neg s(x,y)\}$
$\frac{N*(N-1)}{2}$ is the total count of the cells in the contingency table.

The measure that we introduced to evaluate the homophily in the network is given by:

$Hobs(G) = \frac{2[(a+d)-(c+b)]}{N*(N-1)}$
This measure takes is value between -1 and 1. It is equal to 1 when all the pairs of similar vertices are linked and all the pairs of dissimilar verices are not linked. Otherwise, when all pairs of similar vertices are not linked and all pairs of dissimilar verices are linked, it is equal to -1.

This measure of observed homophily in the network can be compared with an expected value computed on a network having the same number of vertices and edges but where the probability of having a link beteween two vertices is independant of their similarity and consequently of their features, which does not respect the homophily property according which two vertices are more likely to be connected if they share common characteristics.

In order to compute the expected homophily indicator we estimate the probability for pairs of vertices of being linked and similar in the following way:
$PR = \frac{2M}{N*(N-1)}$
$PS = \frac{2*Card\{(x,y) \in V \times (V-1)/s(x,y)\}}{N*(N-1)}$
with $PNR = 1 - PR$ and $PNS = 1 - PS$

Then, with the following contingency table:
$a' = \frac{PR*PS*N*(N-1)}{2}$
$b' = \frac{PNR*PS*N*(N-1)}{2}$
$c' = \frac{PR*PNS*N*(N-1)}{2}$
$c' = \frac{PNR*PNS*N*(N-1)}{2}$
we compute the expected homophily as follows:
$Hexpect(G) = \frac{2[(a'+d')-(c'+b')]}{N*(N-1)}$

A social network exhibits homophily if $Hobs(G)$ is higher than $Hexpect(G)$.

VIII. Related Work

= Prop Burstiness on topic model: Modeling Word Burstiness Using the Dirichlet Distribution (DCM) Accounting for Burstiness in Topic Models (DCMLDA) LDA bursty on topics Proposal of a-MMSB in : Scalable Inference of Overlapping Communities with high diagonal only...

to read: Stochastic blockmodels and community structure in networks

= Model Recent work on MMSB and copula: Copula Mixed-Membership Stochastic Blockmodel with Subgr oup Correlation

IX. Conclusion

X. Sampling $beta$

According to [**?**], β is distributed as follows:

$$\beta = (\beta_1, ..., \beta_K, \beta_u) \sim Dir(m_{.1}, ..., m_{.K}, \gamma) \tag{26}$$

Where $m_{.k}$ represent the number of tables serving the dish k in all restaurants, in the chinese restaurant franchise. The sampling of the table configuration $\boldsymbol{m}$ can be done using the unsigned Stirling numbers of the first kind $s(n, m)$ [**?**]:

$$\mathbf{p}(m_{ik} = m \mid Z, \boldsymbol{m}^{-jk}, \beta) = \frac{\Gamma(\alpha_0 \beta_k)}{\Gamma(\alpha_0 \beta_k + N_{jk})} s(n_{jk}, m)(\alpha_0 \beta_k)^m \tag{27}$$

Appendix

A. Gibbs update for IMMSB

We provide here the derivation of the Gibbs update rules given in Section III for IMMSB.

From the definition of the model, one has: $\mathbf{p}(z_{ij} = k \mid \mathbf{f}_i) = f_{ik}$.

Adrien, peux-tu donner la dérivation ?; la forme actuelle n'est valable que pour MMSB.