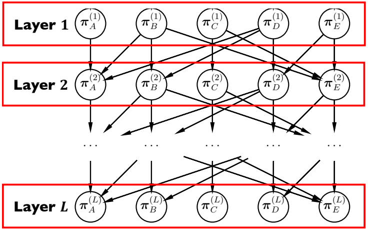
Dirichlet Belief Networks (DirBN)

Applications: topic modeling & relational modeling

$\boldsymbol{\pi}_i^{(l)}$: person i's categorical distribution at layer l; β_{ji}: coefficient from person j to i.

$$\boldsymbol{\pi}_i^{(l)} \sim \text{Dirichlet}(\sum_j \beta_{ji}^{(l-1)} \boldsymbol{\pi}_j^{(l-1)})$$

$$\boldsymbol{\pi}_i^{(l+1)} \sim \text{Dirichlet}(\sum_j \beta_{ji}^{(l)} \boldsymbol{\pi}_j^{(l)})$$



- interpretable units
- persons' dependencies

- $\boldsymbol{\pi}_i^{(l)}$ is involved in Gamma function $\Gamma(\cdot)$

$$P(\boldsymbol{\pi}_i^{(l+1)} | \{\boldsymbol{\pi}_j^{(l)}, \beta_{ji}^{(l)}\}_j) = \frac{\Gamma(\sum_{jk} \beta_{ji}^{(l)} \pi_{jk}^{(l)})}{\prod_k \Gamma(\sum_k \beta_{ji}^{(l)} \pi_{jk}^{(l)})} \prod_k [\pi_{ik}^{(l+1)}]^{\sum_j \beta_{ji}^{(l)} \pi_{jk}^{(l)} - 1}$$

- Forming conjugate models and enabling Gibbs sampling for $\boldsymbol{\pi}_i^{(l)}$ is difficult.

Previous inference method for DirBN

- Use Chinese Restaurant Table variable to augment the Gamma function ratio.
- Drawbacks:
  - information loss in generating Chinese Restaurant Table variable
  - Similar patterns in $\{\pi_i^{(l)}\}_l$

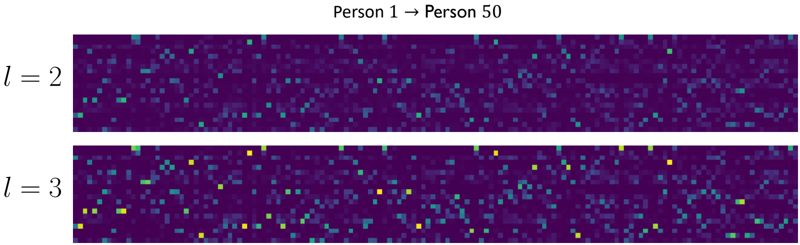


Figure: Visualisations on $\pi_{1:50}^{(2)}, \pi_{1:50}^{(3)}$.

DirBN does not benefit from the deep structure.

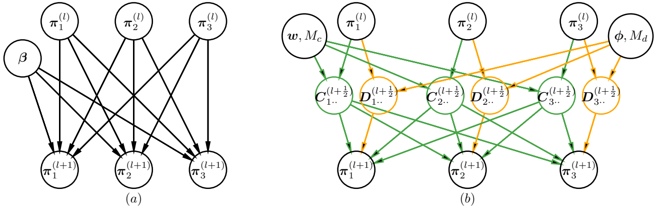
Poisson-randomised DirBN

Introduce augment variables C, D in the Dirichlet-Dirichlet connection.

$$C_{iki'}^{(l)} \sim \text{Poisson}(M_c \pi_{ik}^{(l)} \omega_{ii'}^{(l)}), D_{ikk'}^{(l)} \sim \text{Poisson}(M_d \pi_{ik}^{(l)} \phi_{kk'}^{(l)})$$

$$\boldsymbol{\pi}_i^{(l+1)} \sim \text{Dirichlet}(\boldsymbol{\alpha}^{(\pi)} + \sum_{i'} \boldsymbol{C}_{i' \cdot i}^{(l)} + \sum_{k'} \boldsymbol{D}_{ik' \cdot}^{(l)})$$

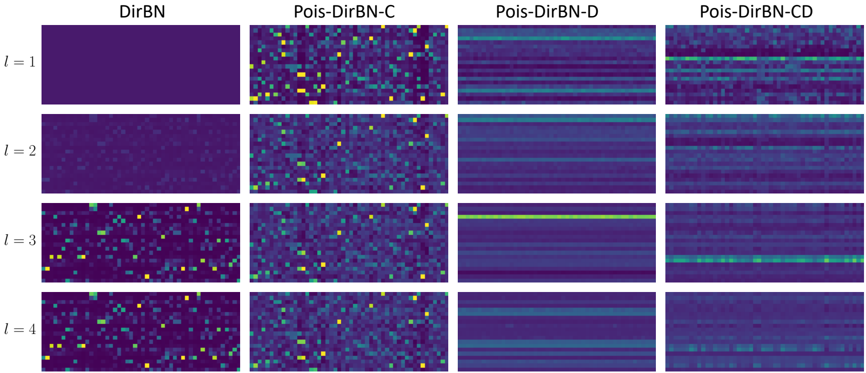
whereas $\omega_{ii'}^{(l)}$: coefficient from person i to person i', $\phi_{kk'}^{(l)}$: coefficient from component k to component k', $M_c, M_d, \boldsymbol{\alpha}^{(\pi)}$: hyper-parameters.



Benefits:

- Layer-wise Gibbs sampling without information loss
- Component-wise propagation variable $\boldsymbol{D}$ introduce larger mutations

Visualisation on our Pois-DirBN



DirBN: 1, similar patterns; 2, information loss

Pois-DirBN-C: 1, similar patterns across layers

Pois-DirBN-D: 1, similar patterns within same layer

Pois-DirBN-CD: None