

Sparse Graph Prior for Knowledge Graph

Dongwoo Kim
ANU

July 19, 2016

1 Completely Random Measure

A completely random measure (CRM) μ on \mathbb{R}_+ is a random measure such that for any countable number of disjoint measurable sets A_1, A_2, \dots of \mathbb{R}_+ , the random variable $\mu(A_1), \mu(A_2), \dots$ are independent and $\mu(\cup_i A_i) = \sum_i \mu(A_i)$. If one assumes that the distribution of $\mu([t, s])$ only depends on the difference $t - s$ then the CRM takes the form of $\mu = \sum_{i=1}^{\infty} w_i \delta_{\theta_i}$ where (w_i, θ_i) are the points of a Poisson point process on \mathbb{R}_+^2 with Lévy intensity measure $\nu(dw, d\theta) = \rho(dw)\lambda(d\theta)$ ¹. The Laplace transform of $\mu(A)$ on any measurable set A has a following representation: $\mathbb{E}[e^{-t\mu(A)}] = \exp(-\int_{\mathbb{R}_+ \times A} (1 - e^{-tw})\rho(dw)\lambda(d\theta))$ for any $t > 0$ and ρ such that $\int_{\mathbb{R}_+} (1 - e^{-w})\rho(dw) < \infty$. Laplace exponent is $\psi(t) = \int_{\mathbb{R}_+} (1 - e^{-tw})\rho(dw)$.

2 Caron and Fox Model

Caron and Fox (2015) propose a simple point process on \mathbb{R}^2 as a product measure of a complete random measure. They propose a hierarchical model for undirected graphs

$$\mu = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \quad \mu \sim \text{CRM}(\rho, \lambda) \quad (1)$$

$$D = \sum_{i,j} n_{ij} \delta_{(\theta_i, \theta_j)} \quad D|\mu \sim \text{PP}(\mu \times \mu) \quad (2)$$

$$Z = \sum_{i,j} \min(n_{ij} + n_{ji}, 1) \delta_{(\theta_i, \theta_j)}, \quad (3)$$

with intensity measure ν factorising as $\nu(dw, d\theta) = \rho(dw)\lambda(d\theta)$ for a jump part of the measure ρ and Lebesgue measure λ . D is simply generated from a Poisson process with a product measure as an intensity and can be interpreted as a directed multi-graph. Given μ , we can directly specify the undirected graph Z as

$$\text{Pr}(z_{ij} = 1|w) = \begin{cases} 1 - \exp(-2w_i w_j) & i \neq j \\ 1 - \exp(-w_i^2) & i = j. \end{cases}$$

¹Subordinator.

They show that the resulting graph is sparse, i.e. $\#$ of edges $= o(\# \text{ of nodes}^2)^2$, if the intensity measure³ is

$$\rho(dw) = \frac{1}{\Gamma(1-\sigma)} w^{-1-\sigma} e^{-\tau w} dw, \quad (4)$$

where the two parameters range

$$(\sigma, \tau) \in (0, 1) \times [0, +\infty) \quad (5)$$

and dense if the intensity measure is finite activity, i.e. $\int_0^\infty \rho(w) dw < \infty$.

The general construction of the sparse graph in Equation 3 results an infinite number of edges due to $\mu(\mathbb{R}_+) = \infty$. A restriction of Lebesgue measure λ on $[0, \alpha]$ is used to obtain a finite graph ($\lambda_\alpha = \lambda \delta_{[0, \alpha]}$). Therefore, restricted graph Z_α is defined on the box $[0, \alpha]^2$. We also denote the total mass on $[0, \alpha]^2$ by $Z_\alpha^* = Z_\alpha([0, \alpha]^2)$, and similarly for D_α^* and μ_α^* .

3 Sparse Prior for Knowledge Graph

A knowledge base consists of a set of triples (entity, entity, relation) such as (BarackObama, bornIn, Hawaii). The set of triples can be represented as a binary-valued three-way tensor where three dimensions represent entity, entity, and relation, respectively. Here, we directly extend the Caron and Fox's model for the three-way tensor based on two independent completely random measures.

$$\mu = \sum_{i=1}^{\infty} w_i \delta_{\theta_i} \quad \mu \sim \text{CRM}(\rho, \lambda) \quad (6)$$

$$\mu' = \sum_{k=1}^{\infty} w_k \delta_{\theta'_k} \quad \mu' \sim \text{CRM}(\rho', \lambda) \quad (7)$$

$$D = \sum_{i,j,k} n_{ijk} \delta_{(\theta_i, \theta_j, \theta'_k)} \quad D \sim \text{PP}(\mu \times \mu \times \mu') \quad (8)$$

$$Z = \sum_{i,j,k} \min(n_{ijk}, 1) \delta_{(\theta_i, \theta_j, \theta'_k)}, \quad (9)$$

where Z is asymmetric in i and j since the knowledge graph is a directed multi-graph. As done in the original model, we can also specify Z as

$$Pr(z_{ijk} = 1 | w, w') = \begin{cases} 1 - \exp(-w_i w_j w'_k) & i \neq j \\ 1 - \exp(-w_i^2 w'_k) & i = j. \end{cases}$$

If we consider θ_i , θ_j , and θ'_k as nodes in the graph, the above construction will generate a hypergraph where each edge connects three nodes. In the notion of knowledge graphs, it is more intuitive to consider a relation as a type of edge between two entities. In this case, we define two random measures on \mathbb{R}_+^2 :

$$\bar{D} = \sum_{i,j} \sum_k z_{ijk} \delta_{\theta_i, \theta_j} \quad (10)$$

$$\bar{Z} = \sum_{i,j} \min(\bar{D}(\{\theta_i, \theta_j\}), 1) \delta_{(\theta_i, \theta_j)}, \quad (11)$$

²only counts the nodes which has at least one edge

³This is the Lévy intensity of the generalised gamma process

where \bar{D} is a multigraph, and \bar{Z} is a binary graph of a knowledge base.

$$Pr(\bar{z}_{ij} = 1 | w, w') = \begin{cases} 1 - \exp(-w_i w_j \sum_k w'_k) & i \neq j \\ 1 - \exp(-w_i^2 \sum_k w'_k) & i = j. \end{cases}$$

To obtain a finite hypergraph (the number of edges is finite), we consider restrictions $D_{\alpha\beta}$ and $Z_{\alpha\beta}$ to the box $[0, \alpha]^2 \times [0, \beta]$. We denote by $Z_{\alpha\beta}^* = Z_{\alpha\beta}([0, \alpha]^2 \times [0, \beta])$ the total mass on the restricted area, and similar for $D_{\alpha\beta}^*$ and μ_α^* .

3.1 Generative Process through Urn approach

Given restriction α and β , the generative process of $D_{\alpha\beta}$ can be specified as follows:

1. $\mu_\alpha \sim \text{CRM}(\rho, \lambda_\alpha)$
2. $\mu'_\beta \sim \text{CRM}(\rho', \lambda_\beta)$
3. $D_{\alpha\beta}^* | \mu_\alpha, \mu'_\beta \sim \text{Poisson}(\mu_\alpha^{*2} \mu'_\beta)$
4. For $d = 1, \dots, D_{\alpha\beta}^*$:
 - (a) $\theta_{di} \sim \frac{\mu_\alpha}{\mu_\alpha^*}$
 - (b) $\theta_{dj} \sim \frac{\mu_\alpha}{\mu_\alpha^*}$
 - (c) $\theta'_{dk} \sim \frac{\mu_\beta}{\mu'_\beta}$
5. $D_{\alpha\beta} = \sum_{d=1}^{D_{\alpha\beta}^*} \delta_{(\theta_{di}, \theta_{dj}, \theta'_{dk})}$,

where we have used that the total mass of $D_{\alpha\beta}^*$ follows the Poisson distribution. Each node θ_i is drawn from the normalised CRM (NRM), $\frac{\mu_\alpha}{\mu_\alpha^*}$, which is discrete with probability 1. However, it is not possible to sample μ_α and μ'_β since these measures have infinite number of atoms. Instead we can simulate finite-dimensional generative process through the urn formulation. Let $\theta_1, \dots, \theta_n$ drawn from the normalised CRM $\frac{\mu_\alpha}{\mu_\alpha^*}$. Since NRM is discrete, variables $\theta_1, \dots, \theta_n$ takes $l \leq n$ distinct values ϕ_l , and m_l is the number of variables corresponding to ϕ_l . Given total mass μ_α^* and $\theta_1, \dots, \theta_n$, the conditional distribution of θ_{n+1} can be modelled in terms of exchangeable partition probability function (EPPF):

$$\theta_{n+1} | \mu_\alpha^*, \theta_1, \dots, \theta_n \sim \frac{\Pi_{n+1}^{l+1}(m_1, \dots, m_l, 1 | \mu_\alpha^*)}{\Pi_n^l(m_1, \dots, m_l | \mu_\alpha^*)} \frac{1}{\alpha} \lambda_\alpha + \sum_{i=1}^l \frac{\Pi_{n+1}^l(m_1, \dots, m_i + 1, \dots, m_l | \mu_\alpha^*)}{\Pi_n^l(m_1, \dots, m_l | \mu_\alpha^*)} \delta_{\phi_i} \quad (12)$$

where

$$\Pi_n^l(m_1, \dots, m_l | \mu_\alpha^*) = \frac{\sigma^l \mu_\alpha^{*-n}}{\Gamma(n-l\sigma) g_\sigma(\mu_\alpha^*)} \int_0^{\mu_\alpha^*} s^{n-l\sigma-1} g_\sigma(\mu_\alpha^* - s) ds \left(\prod_{i=1}^l \frac{\Gamma(m_i - \sigma)}{\Gamma(1 - \sigma)} \right), \quad (13)$$

and g_σ is the pdf of the positive stable distribution. Finally, the total mass of μ_α^* and μ'_β follows an exponentially tilted stable distribution where the exact sampler exists (Devroye, 2009; Hofert, 2011).

Using this urn representation, we can rewrite the generative process as

1. $\mu_\alpha^* \sim P_{\mu_\alpha^*}$
2. $\mu'_\beta \sim P_{\mu'_\beta}$
3. $D_{\alpha\beta}^* | \mu_\alpha, \mu'_\beta \sim \text{Poisson}(\mu_\alpha^{*2} \mu'_\beta)$
4. For $d = 1, \dots, D_{\alpha\beta}^*$:
 - (a) Sample θ_{di} , θ_{dj} , and θ'_{dk} with Urn process in Eqn 12
5. $D_{\alpha\beta} = \sum_{d=1}^{D_{\alpha\beta}^*} \delta_{(\theta_{di}, \theta_{dj}, \theta'_{dk})}$,

3.2 Characteristics of Random Graph in Gamma process case ($\sigma = 0$)

In case $\sigma = 0$, $\rho(dw)$ is an intensity of the Gamma process where the sum of the weights μ_α^* follows Gamma distribution with shape parameter α and scale parameter τ .

3.2.1 Expected number of triples

From the generative process of the random graph, the number of total edge follows the poisson distribution with mean intensity $\mu_\alpha^{*2} \mu'_\beta$.

$$\mathbb{E}[D_{\alpha\beta}^*] = \mathbb{E}[\mu_\alpha^{*2}] \mathbb{E}[\mu'_\beta] \quad (14)$$

$$= (\text{Var}(\mu_\alpha^*) + \mathbb{E}[\mu_\alpha^*]^2) \mathbb{E}[\mu'_\beta] \quad (15)$$

$$= \frac{\alpha(\alpha + 1)}{\tau} \frac{\beta}{\tau} \quad (16)$$

3.2.2 Expected number of entities and relations

From the generative process of the random graph, we can compute the expected number of entities N_α as

$$\mathbb{E}[N_\alpha | D_{\alpha\beta}^*] = \mathbb{E}\left[\sum_{i=1}^{2D_{\alpha\beta}^*} Y_i\right], \quad (17)$$

where

$$Y_i \sim \text{Ber}\left(\frac{\alpha}{\alpha + i - 1}\right). \quad (18)$$

So, the expected number of entities for the large number of $2D_{\alpha\beta}^*$ can be approximated as

$$\mathbb{E}[N_\alpha | D_{\alpha\beta}^*] = \sum_{i=1}^{2D_{\alpha\beta}^*} \frac{\alpha}{\alpha + i - 1} = \alpha(\Psi(\alpha + 2D_{\alpha\beta}^*) - \Psi(\alpha)) \approx \alpha \log(\alpha + 2D_{\alpha\beta}^*) \quad (19)$$

where Ψ is a digamma function (Arratia et al., 2003). By using Theorem 8 in (Caron and Fox, 2015), we can further show $\mathbb{E}[N_\alpha] = \Theta(\alpha \log \alpha)$ as $\alpha \rightarrow \infty$. The expected number of relations can be computed in a similar way:

$$\mathbb{E}[N_\beta | D_{\alpha\beta}^*] = \sum_{j=1}^{D_{\alpha\beta}^*} \frac{\beta}{\beta + j - 1} = \beta(\Psi(\beta + D_{\alpha\beta}^*) - \Psi(\beta)) \approx \beta \log(\beta + D_{\alpha\beta}^*) \quad (20)$$

Since N_α and N_β is independent,

$$\mathbb{E}[N_\alpha N_\beta | D_{\alpha\beta}^*] \approx \alpha \log(\alpha + 2D_{\alpha\beta}^*) \times \beta \log(\beta + D_{\alpha\beta}^*) \quad (21)$$

3.3 Posterior inference

We first characterise the posterior of μ_α given μ'_β and $D_{\alpha\beta}$. The conditional Laplace functional of μ_α given $D_{\alpha\beta}$ is $\mathbb{E}[e^{-\mu_\alpha(f)} | \mu'_\beta, D_{\alpha\beta}]$, for any non-negative measurable function f such that $\mu_\alpha(f) = \sum_{i=1}^\infty w_i f(\theta_i)$. We have $\mu_\alpha(f) = \Pi(\tilde{f})$ where $\Pi = \sum_{i=1}^\infty \delta_{w_i, \theta_i}$ is a Poisson random measure on $\mathcal{S} = (0, \infty) \times [0, \alpha]$ with mean measure $\rho \times \lambda$ and $\tilde{f}(w, \theta) = wf(\theta)$. Let $n_{i**} = \sum_{j=1}^{N_\alpha} \sum_{k=1}^{N_\beta} n_{ijk}$, $m_i = \sum_{j=1}^{N_\alpha} \sum_{k=1}^{N_\beta} n_{ijk} + n_{jik}$, and $m'_k = \sum_{i=1}^{N_\alpha} \sum_{j=1}^{N_\alpha} n_{ijk}$.

$$\mathbb{E}_{\mu_\alpha}[e^{-\mu_\alpha(f)} | D_{\alpha\beta}, \mu'_\beta] = \mathbb{E}_\Pi[e^{-\int \tilde{f}(w, \theta) \Pi(dw, d\theta)} | D_{\alpha\beta}, \mu'_\beta] \quad (22)$$

$$= \frac{\mathbb{E}_\Pi[e^{-\Pi(\tilde{f})} P(D_{\alpha\beta} | \Pi, \mu'_\beta)]}{\mathbb{E}_\Pi[P(D_{\alpha\beta} | \Pi, \mu'_\beta)]} \quad (23)$$

$$= \frac{\mathbb{E}_\Pi[e^{-\Pi(\tilde{f})} e^{-\Pi(h)^2 \mu'^*_{\beta}} \prod_{i=1}^{N_\alpha} w_i^{m_i}]}{\mathbb{E}_\Pi[e^{-\Pi(h)^2 \mu'^*_{\beta}} \prod_{i=1}^{N_\alpha} w_i^{m_i}]} \quad (24)$$

where $h(w, \theta) = w$ and

$$P(D_{\alpha\beta} | \Pi, \mu'_\beta) = P(D_{\alpha\beta} | \mu_\alpha, \mu'_\beta) \quad (25)$$

$$= \text{Poisson}(D_{\alpha\beta}^* | \mu_\alpha^{*2} \mu'^*_{\beta}) \prod_{i=1}^{N_\alpha} P(n_{i**} | \mu_\alpha) \prod_{j=1}^{N_\alpha} P(n_{*j*} | \mu_\alpha) \prod_{k=1}^{N_\beta} P(n_{**k} | \mu_\beta) \quad (26)$$

$$= \frac{(\mu_\alpha^{*2} \mu'^*_{\beta})^{D_{\alpha\beta}^*} e^{-\mu_\alpha^{*2} \mu'^*_{\beta}}}{D_{\alpha\beta}^*!} \prod_{i=1}^{N_\alpha} \left(\frac{w_i}{\mu_\alpha^*}\right)^{n_{i**}} \prod_{j=1}^{N_\alpha} \left(\frac{w_j}{\mu_\alpha^*}\right)^{n_{*j*}} \prod_{k=1}^{N_\beta} \left(\frac{w'_k}{\mu'^*_{\beta}}\right)^{n_{**k}} \quad (27)$$

$$= \frac{e^{-\mu_\alpha^{*2} \mu'^*_{\beta}}}{D_{\alpha\beta}^*!} \prod_{i=1}^{N_\alpha} w_i^{m_i} \prod_{k=1}^{N_\beta} w'_k{}^{m_k} = \frac{e^{-\Pi(h)^2 \mu'^*_{\beta}}}{D_{\alpha\beta}^*!} \prod_{i=1}^{N_\alpha} w_i^{m_i} \prod_{k=1}^{N_\beta} w'_k{}^{m_k} \quad (28)$$

$$(29)$$

$$\mu_\alpha^* = \sum_{i=1}^\infty w_i, \quad \mu'^*_{\beta} = \sum_{k=1}^\infty w'_k = \sum_{k=1}^{N_\beta} w'_k + w'^* \quad (30)$$

Applying the generalised Palm formula to the numerator yields

$$\mathbb{E}_{\Pi} \left[e^{-\Pi(\tilde{f})} e^{-\Pi(h)^2 \mu'_{\beta} *} \prod_{i=1}^{N_{\alpha}} w_i^{m_i} \right] \quad (31)$$

$$= \mathbb{E}_{\Pi} \left[e^{-\Pi(\tilde{f})} e^{-\Pi(h)^2 \mu'_{\beta} *} \prod_{i=1}^{N_{\alpha}} \sum_{w_j, \vartheta_j \in \Pi} w_j^{m_i} \mathbf{1}_{\theta_i}(\vartheta_j) \right] \quad (32)$$

$$= \mathbb{E}_{\Pi} \left[\int_{\mathcal{S}^{N_{\alpha}}} e^{-\Pi(\tilde{f})} e^{-\Pi(h)^2 \mu'_{\beta} *} \prod_{i=1}^{N_{\alpha}} w_j^{m_i} \mathbf{1}_{\theta_i}(\vartheta_j) \Pi(dw_j, d\vartheta_j) \right] \quad (33)$$

$$= \int_{\mathcal{S}^{N_{\alpha}}} \mathbb{E}_{\Pi} \left[e^{-(\Pi + \sum_{i=1}^{N_{\alpha}} \delta_{(w_i, \theta_i)})(\tilde{f})} e^{-(\Pi + \sum_{i=1}^{N_{\alpha}} \delta_{(w_i, \theta_i)})(h)^2 \mu'_{\beta} *} \right] \prod_{i=1}^{N_{\alpha}} w_j^{m_i} \mathbf{1}_{\theta_i}(\vartheta_j) \rho(dw_j) \lambda(d\vartheta_j) \quad (34)$$

$$= \int_{\mathcal{S}^{N_{\alpha}}} \mathbb{E}_{\mu_{\alpha}} \left[e^{-\mu_{\alpha}(f) - \sum_{i=1}^{N_{\alpha}} w_i f(\vartheta_j)} e^{-(\mu_{\alpha}(1) + \sum_{i=1}^{N_{\alpha}} w_i)^2 \mu'_{\beta} *} \right] \prod_{i=1}^{N_{\alpha}} w_j^{m_i} \mathbf{1}_{\theta_i}(\vartheta_j) \rho(dw_j) \lambda(d\vartheta_j) \quad (35)$$

$$= \int_{\mathcal{S}^{N_{\alpha}}} \mathbb{E}_{\mu_{\alpha}^*} \left[\mathbb{E}_{\mu_{\alpha}} \left[e^{-\mu_{\alpha}(f)} | \mu_{\alpha}^* \right] e^{-\sum_{i=1}^{N_{\alpha}} w_i f(\vartheta_j)} e^{-(\mu_{\alpha}^* + \sum_{i=1}^{N_{\alpha}} w_i)^2 \mu'_{\beta} *} \right] \prod_{i=1}^{N_{\alpha}} w_j^{m_i} \mathbf{1}_{\theta_i}(\vartheta_j) \rho(dw_j) \lambda(d\vartheta_j) \right] \quad (36)$$

The denominator is obtained by taking $f = 0$.

$$\mathbb{E}_{\mu_{\alpha}} [e^{-\mu_{\alpha}(f)} | D_{\alpha\beta}, \mu'_{\beta}] = \int_{\mathbb{R}^{N_{\alpha}+1}} E_{\mu_{\alpha}} [e^{-\mu_{\alpha}(f)} | \mu_{\alpha}^* = w^*] \quad (37)$$

$$\times e^{\sum_{i=1}^{N_{\alpha}} w_i f(\theta_i)} p(w_1, \dots, w_{N_{\alpha}}, w^* | D_{\alpha\beta}, \mu_{\beta}) dw_{1:N_{\alpha}} dw^* \quad (38)$$

where

$$p(w_1, \dots, w_{N_{\alpha}}, w^* | D_{\alpha\beta}, \mu_{\beta}) = \frac{\prod_{i=1}^{N_{\alpha}} w_j^{m_i} \rho(w_i) e^{-(w^* + \sum_{i=1}^{N_{\alpha}} w_i)^2 \mu'_{\beta} *} g_{\alpha}^*(w^*)}{\int_{\mathbb{R}^{N_{\alpha}+1}} \prod_{i=1}^{N_{\alpha}} \tilde{w}_j^{m_i} \rho(\tilde{w}_i) e^{-(\tilde{w}^* + \sum_{i=1}^{N_{\alpha}} \tilde{w}_i)^2 \mu'_{\beta} *} g_{\alpha}^*(\tilde{w}^*) d\tilde{w}_{1:N_{\alpha}} d\tilde{w}^*} \quad (39)$$

$$(40)$$

$g_{\alpha}^*(w^*)$ is a density function of random variable w^* of which Laplace transform is $\mathbb{E}[e^{tw^*}] = e^{\alpha\psi(t)}$. Therefore, the conditional of μ_{α} given $D_{\alpha\beta}, \mu'_{\beta}$ is

$$w^* \sum_{i=1}^{\infty} \tilde{P}_i \delta_{\tilde{\theta}_i} + \sum_{i=1}^{N_{\alpha}} w_i \delta_{\theta_i} \quad (41)$$

where (\tilde{P}) are distributed from a Poisson-Kingman distribution conditional on w^* , and the weights $w_1, \dots, w_{N_{\alpha}}, w^*$ are jointly dependent conditional on $D_{\alpha\beta}$ and μ'_{β} :

$$p(w_1, \dots, w_{N_{\alpha}}, w^* | D_{\alpha\beta}, \mu'_{\beta}) \propto \prod_{i=1}^{N_{\alpha}} w_i^{m_i} e^{(-w^* + \sum_{i=1}^{N_{\alpha}} w_i)^2 \mu'_{\beta} *} \prod_{i=1}^{N_{\alpha}} \rho(w_i) g_{\alpha}^*(w^*) \quad (42)$$

The conditional Laplace functional of μ'_{β} given μ_{α} and $D_{\alpha\beta}$ can be carried out in the same way as we've done in μ_{α} .

References

- Arratia, R., Barbour, A. D., and Tavaré, S. (2003). *Logarithmic combinatorial structures: a probabilistic approach*. European Mathematical Society.
- Caron, F. and Fox, E. B. (2015). Sparse graphs using exchangeable random measures. pages 1–64.
- Devroye, L. (2009). Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 19(4):18.
- Hofert, M. (2011). Sampling exponentially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 22(1):3.