Gamma Process Poisson Factorization for Joint Modeling of Network and Documents

Ayan Acharya[†]

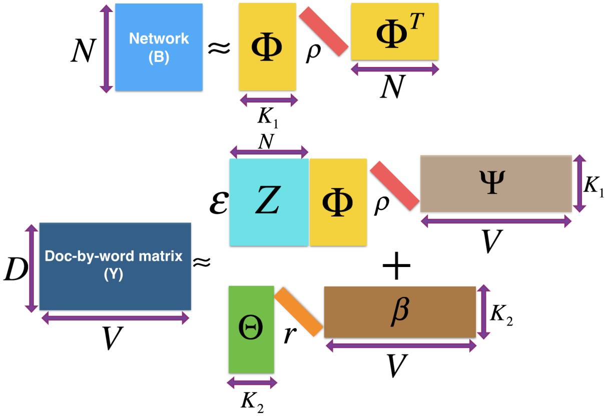
[†]UT Austin, Department of ECE; Cognitive Scale Inc., Austin, Texas, USA
Joint Work with Dean Teffer, Jette Henderson, Marcus Tyler, Mingyuan Zhou, Joydeep Ghosh

September 10, 2015

Motivation and Problem Setting

- Real world networks are sparse, but often come with side information.
- Real world dyadic data, such as a user-movie rating data, is often sparse, but is occasionally endowed with an interaction network of at least one type of entities, such as a user-user social network.

Gamma Process Poisson Factorization for Joint Network and Topic Modeling (J-GPPF)



Poisson Factor Analysis

- $Y \sim \text{Poisson}(\boldsymbol{\theta}\beta^{\dagger}),\ \boldsymbol{\theta} \sim *,\ \boldsymbol{\beta} \sim *.$
- When $\boldsymbol{\theta} \sim \text{Dir}(.),\ \boldsymbol{\beta} \sim \text{Dir}(.)$, one obtains smoothened LDA.
- Let $x_k \sim \text{Poisson}(\lambda_k)\ \forall k,\ X = \sum_{k=1}^{K} x_k,\ \lambda = \sum_{k=1}^{K} \lambda_k$. Suppose $(y_1, \cdots, y_K) \sim \text{mult}(X; \lambda_1/\lambda, \cdots, \lambda_K/\lambda)$. Then $P(x_1, \cdots, x_K) = P(y_1, \cdots, y_K; x)$.
- If $x_k \sim \text{Gam}(\alpha_k, 1/c)\ \forall k$, then $\boldsymbol{Y} = (y_k)_{k=1}^{K} \sim \text{Dir}(\alpha_1, \cdots, \alpha_K)$ where $y_k = x_k / \sum_{k=1}^{K} x_k$; Dirichlet = normalized Gamma

Gamma Process (ΓP)

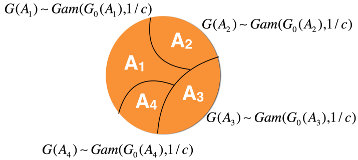


Figure: Illustration of Gamma Process

- The Gamma Process $G \sim \Gamma\mathrm{P}(G_0, c)$ is a completely random measure defined on the product space $\mathbb{R}_+ \times \Omega$ with concentration parameter c and a finite and continuous base measure G_0 over a complete separable metric space Ω, such that $G(A_i) \sim \mathrm{Gam}(G_0(A_i), 1/c)$ are independent gamma random variables for disjoint partition $\{A_i\}_i$ of Ω.

Gamma Process (ΓP)

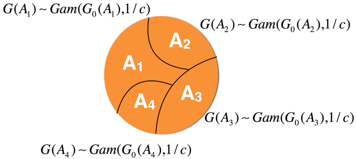


Figure: Illustration of Gamma Process

- The Gamma Process $G \sim \Gamma\mathrm{P}(G_0, c)$ is a completely random measure defined on the product space $\mathbb{R}_+ \times \Omega$ with concentration parameter c and a finite and continuous base measure G_0 over a complete separable metric space Ω, such that $G(A_i) \sim \mathrm{Gam}(G_0(A_i), 1/c)$ are independent gamma random variables for disjoint partition $\{A_i\}_i$ of Ω.
- $G = \sum_{k=1}^{\infty} r_k \delta_{\omega_k}, (r_k, \omega_k) \overset{iid}{\sim} r^{-1} e^{-cr} dr G_0(d\omega)$.

Gamma Process (ΓP)

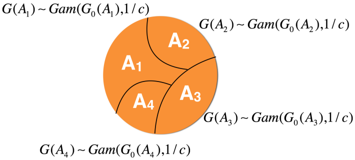
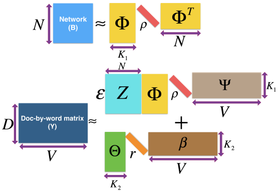


Figure: Illustration of Gamma Process

- The Gamma Process $G \sim \Gamma\mathrm{P}(G_0, c)$ is a completely random measure defined on the product space $\mathbb{R}_+ \times \Omega$ with concentration parameter c and a finite and continuous base measure G_0 over a complete separable metric space Ω, such that $G(A_i) \sim \mathrm{Gam}(G_0(A_i), 1/c)$ are independent gamma random variables for disjoint partition $\{A_i\}_i$ of Ω.
- $G = \sum_{k=1}^{\infty} r_k \delta_{\omega_k}, (r_k, \omega_k) \overset{iid}{\sim} r^{-1} e^{-cr} dr G_0(d\omega)$.
- Finite approximation of ΓP:

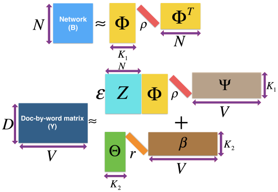
$$G = \sum_{k=1}^{K} r_k \delta_{\omega_k}, (r_k, \omega_k) \overset{iid}{\sim} r^{(\gamma_0/K-1)} e^{-cr} dr G_0(d\omega),\ \gamma_0 = G_0(\omega).$$

GPPF for Joint Network and Topic Modeling (J-GPPF)



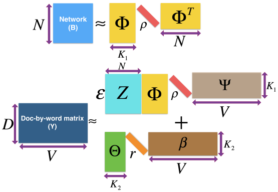
- $b_{nm} = I_{\{x_{nm} \geq 1\}}, x_{nm} \sim \text{Pois}(\sum_{k_B=1}^{K_1} \rho_{k_B} \phi_{nk_B} \phi_{mk_B})$, $\rho_{k_B} \sim \text{Gam}(\gamma_B / K_B, 1/c_B)$, $\phi_{k_B} \sim \prod_{n=1}^{N} \text{Gam}(a_B, 1/\sigma_n)$.

GPPF for Joint Network and Topic Modeling (J-GPPF)



- $b_{nm} = I_{\{x_{nm} \geq 1\}}, x_{nm} \sim \text{Pois}(\sum_{k_B=1}^{K_1} \rho_{k_B} \phi_{nk_B} \phi_{mk_B})$, $\rho_{k_B} \sim \text{Gam}(\gamma_B / K_B, 1/c_B)$, $\phi_{k_B} \sim \prod_{n=1}^{N} \text{Gam}(a_B, 1/\sigma_n)$.
- $y_{dw} \sim \text{Pois}(\sum_{k_Y=1}^{K_2} r_{k_Y} \theta_{dk_Y} \beta_{wk_Y} + \epsilon \sum_{k_B=1}^{K_1} \rho_{k_B} (\sum_n Z_{nd} \phi_{nk_B}) \psi_{wk_B})$,
- $r_{k_Y} \sim \text{Gam}(\gamma_Y / K_Y, 1/c_Y)$, $\theta_{k_Y} \sim \prod_{d=1}^{D} \text{Gam}(a_Y, 1/\varsigma_d)$, $\beta_{k_Y} \sim \prod_{w=1}^{V} \text{Gam}(\xi_Y, 1/\eta_w)$, $\psi_{k_B} \sim \prod_{w=1}^{V} \text{Gam}(\xi_B, 1/\zeta_w)$, $\epsilon \sim \text{Gam}(f_0, 1/g_0)$.

GPPF for Joint Network and Topic Modeling (J-GPPF)



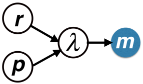
- $b_{nm} = I_{\{x_{nm} \geq 1\}}$, $x_{nm} \sim \text{Pois}(\sum_{k_B=1}^{K_1} \rho_{k_B} \phi_{nk_B} \phi_{mk_B})$, $\rho_{k_B} \sim \text{Gam}(\gamma_B / K_B, 1/c_B)$, $\phi_{k_B} \sim \prod_{n=1}^{N} \text{Gam}(a_B, 1/\sigma_n)$.
- $y_{dw} \sim \text{Pois}(\sum_{k_Y=1}^{K_2} r_{k_Y} \theta_{dk_Y} \beta_{wk_Y} + \epsilon \sum_{k_B=1}^{K_1} \rho_{k_B} (\sum_n Z_{nd} \phi_{nk_B}) \psi_{wk_B})$,
- $r_{k_Y} \sim \text{Gam}(\gamma_Y / K_Y, 1/c_Y)$, $\theta_{k_Y} \sim \prod_{d=1}^{D} \text{Gam}(a_Y, 1/\varsigma_d)$, $\beta_{k_Y} \sim \prod_{w=1}^{V} \text{Gam}(\xi_Y, 1/\eta_w)$, $\psi_{k_B} \sim \prod_{w=1}^{V} \text{Gam}(\xi_B, 1/\zeta_w)$, $\epsilon \sim \text{Gam}(f_0, 1/g_0)$.
- $\gamma_B \sim \text{Gam}(e_B, 1/f_B)$, $\gamma_Y \sim \text{Gam}(e_Y, 1/f_Y)$.

Characteristics of J-GPPF

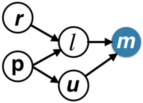
- Poisson factorization – samples latent counts corresponding to non-zeros only
- Computation Complexity: $O(S_1 K_1 + S_2 K_2)$, S_1 is the number of links in the network, S_2 is the number of non-zero entries in the count-valued matrix
- Joint Poisson factorization for imputing a graph
- Hierarchy of Gamma priors for less sensitivity towards initialization
- Non-parametric modeling with closed form inference updates

Negative Binomial Distribution (NB)

- Number of heads seen until r number of tails occurs while tossing a biased coin with probability of head p (or, number of successes before r failures in successive Bernoulli trials): $m \sim \text{NB}(r, p)$
- $m \sim \text{Poisson}(\lambda), \lambda \sim \text{Gam}(r, p/(1-p))$ – Gamma-Poisson Construction
- $m \sim \sum_{t=1}^{\ell} u_t,\ u_t \sim \text{Log}(p),\ \ell \sim \text{Poisson}(-r\log(1-p))$ – Compound Poisson Construction



Gamma-Poisson Construction



Compound Poisson Construction

Figure: Constructions of Negative Binomial Distribution

Chinese Restaurant Table Distribution (CRT)

- Chinese Restaurant Process: occupy an empty table w.p. γ_0 or occupy a table w.p. proportional to the number of customers in that table
- m : number of data points (number of customers)
- K : number of distinct atoms (number of tables)

$$\Pr(K = l|m, \gamma_0) = \frac{\Gamma(\gamma_0)}{\Gamma(m+\gamma_0)}|s(m,l)|\gamma_0^l, \;\; l = 0, 1, \cdots, m,$$

where, $s(m,l)$ is the Stirling number of the first kind

Figure: Illustration of Chinese Restaurant Table Distribution

Lemma

If $m \sim NB(r,p)$ is represented under its compound Poisson representation, then the conditional posterior of ℓ given m and r is given by $(\ell|m,r) \sim CRT(m,r)$, which can be generated via $\ell = \sum_{n=1}^{m} z_n, z_n \sim Bernoulli(r/(n-1+r))$.

Inference of Shape Parameter of Gamma Distribution

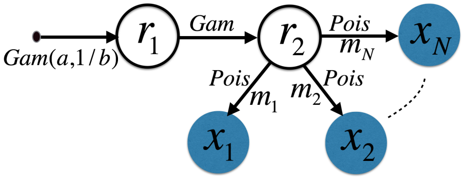


Figure: $x_i \sim \text{Pois}(m_i r_2)\ \forall i \in \{1, 2, \cdots, N\}$, $r_2 \sim \text{Gam}(r_1, 1/d)$, $r_1 \sim \text{Gam}(a, 1/b)$

Inference of Shape Parameter of Gamma Distribution

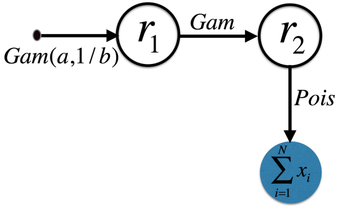


Figure: $\sum_{i=1}^{n} x_i \sim \mathrm{Pois}(r_2 \sum_{i=1}^{n} m_i)$, $r_2 \sim \mathrm{Gam}(r_1, 1/d)$, $r_1 \sim \mathrm{Gam}(a, 1/b)$

Lemma

If $x_i \sim Pois(m_i r_2)$ $\forall i$, $r_2 \sim Gam(r_1, 1/d)$, $r_1 \sim Gam(a, 1/b)$, then $(r_1|-) \sim Gam(a+\ell, 1/(b - log(1-p)))$ where $(\ell|\{x_i\}_i, r_1) \sim CRT(\sum_i x_i, r_1)$, $p = \sum_i m_i/(d + \sum_i m_i)$.

Inference of Shape Parameter of Gamma Distribution

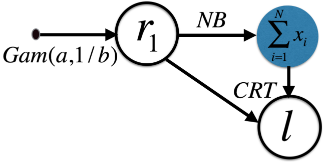


Figure:
$\sum_{i=1}^{n} x_i \sim \mathrm{NB}(r, p), \ell \sim \mathrm{CRT}(\sum_{i=1}^{n} x_i, r_1), p = \sum_i m_i/(d + \sum_i m_i)$

Lemma

If $x_i \sim Pois(m_i r_2)\ \forall i,\ r_2 \sim Gam(r_1, 1/d),\ r_1 \sim Gam(a, 1/b),$ then $(r_1|-) \sim Gam(a + \ell, 1/(b - log(1 - p)))$ where $(\ell|\{x_i\}_i, r_1) \sim CRT(\sum_i x_i, r_1), p = \sum_i m_i/(d + \sum_i m_i)$.

Inference of Shape Parameter of Gamma Distribution

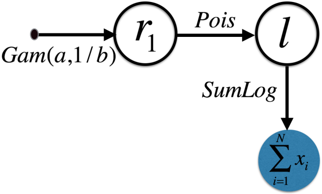


Figure: $\sum_{i=1}^{n} x_i \sim \sum_{\ell} \mathrm{Log}(p), \ell \sim \mathrm{Pois}(-r_1 \log(1-p))$

Lemma

If $x_i \sim Pois(m_i r_2)\ \forall i,\ r_2 \sim Gam(r_1, 1/d),\ r_1 \sim Gam(a, 1/b)$, then $(r_1|-) \sim Gam(a+\ell, 1/(b - log(1-p)))$ where $(\ell|\{x_i\}_i, r_1) \sim CRT(\sum_i x_i, r_1), p = \sum_i m_i/(d + \sum_i m_i)$.

J-GPPF Results: Synthetic Data

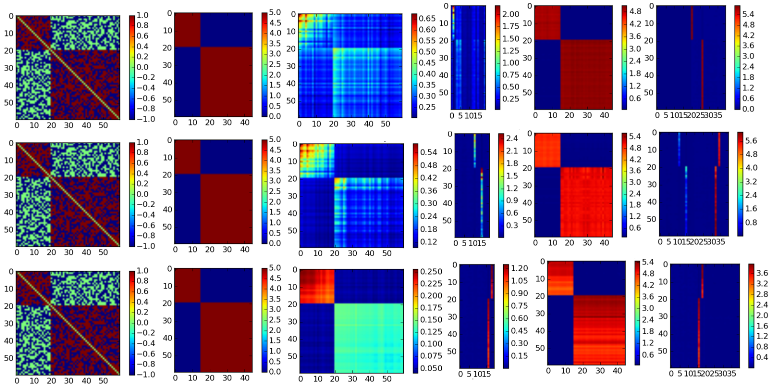


Figure: Row 1: $\epsilon = 0.01$, Row 2: $\epsilon = 1.00$, Row 3: $\epsilon = 10.0$, Column 1: Network with held-out links, Column 2: Count-valued side information, Column 3: network groups discovered by JGPPF, Column 4: Weights of the network factors, Column 5: groups of side-information matrix discovered by JGPPF, Column 6: Weights of the network factors and the topics

J-GPPF Results: Link Prediction in Real-world Network

Baselines:

- N-GPPF – network only GPPF
- MMSB – mixed membership stochastic block model
- IRM – Infinite relational model

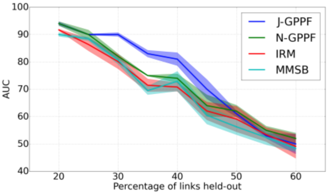


Figure 1: AUC NIPS

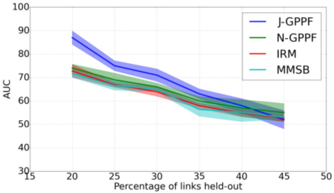


Figure 2: AUC GoodReads

J-GPPF Results: Ranking of Count-valued Real Data

Baselines:

- C-GPPF – corpus only GPPF
- HPMF – hierarchical Poisson matrix factorization

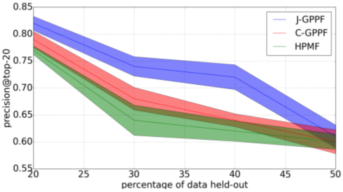


Figure 3: MAP NIPS

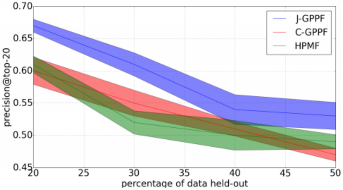


Figure 4: MAP GoodReads

Conclusion and Future Works

Conclusion:

- J-GPPF leverages from learning with multi-modal data.
- Scalable approach due to closed form inference updates and Poisson factorization
- Code available at `https://github.com/aacharya/JGPPF`

Future Works:

- Tractable Gibbs sampling inference with arbitrarily deep Gamma-Poisson model
- distributed Gamma Process Poisson factorization
- dynamic Gamma Process Poisson factorization
- deep Gamma Process Poisson factorization for count-valued data

Questions?