Scalable MCMC in degree corrected stochastic block model

Soumyasundar Pal

Dept. of Electrical and Computer Engineering
McGill University,
Montréal, Québec, Canada

February 11, 2019

Introduction

- Community detection from networks
- Academic collaboration, protein interaction, social networks
- Community: dense internal and sparse external connections
- Earlier approaches: hierarchical clustering, modularity optimization, spectral clustering, clique percolation
- Challenges: handling sparsity, scalability

Introduction

- Heuristic objective function, greedy optimization
- Plethora of techniques[1] [2]
- Numerous quality metrics[3]
- Principled approach: statistical modelling of community structures

[1] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, pp. 75–174, Feb. 2010.

[2] S. Parthasarathy, Y. Ruan, and V. Satuluri, "Community discovery in social networks: Applications, methods and emerging trends," in *Social Network Data Analytics*, pp. 79–113. Springer US, Boston, MA, Mar. 2011.

[3] T. Chakraborty, A. Dalmia, A. Mukherjee, and N. Ganguly,"Metrics for community analysis: a survey, *ACM Comput. Surv.*, vol. 50, no. 4, pp. 1–37, Aug. 2017.

Stochastic Block Model (SBM)

- Connectivity depends on community membership[4].

[4] E. Abbe, "Community detection and stochastic block models, *Found. and Trends Commun. and Inform. Theory*, vol. 14, no.1-2, pp. 1162, Jun. 2018.

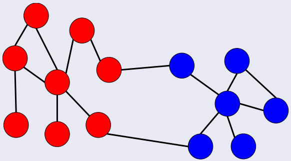
Stochastic Block Model (SBM)

- Connectivity depends on community membership[4].
- Stochastic equivalence of nodes within same community

[4] E. Abbe, "Community detection and stochastic block models, *Found. and Trends Commun. and Inform. Theory*, vol. 14, no.1-2, pp. 1162, Jun. 2018.

Stochastic Block Model (SBM)

- Connectivity depends on community membership[4].
- Stochastic equivalence of nodes within same community



N: no. nodes, K: no. communities

c_i: membership of node i

$c_i \in \{1, 2, ..., K\}$, $\mathcal{C} = \{c_i\}_{i=1}^N$

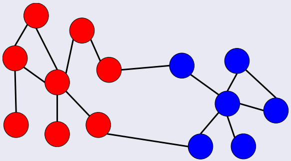
$y_{ab} \in \{0, 1\}$: (a, b)'th entry in adj. matrix

$\beta_{k\ell} \in (0, 1)$: link probability between two nodes in community k and ℓ

[4] E. Abbe, "Community detection and stochastic block models, *Found. and Trends Commun. and Inform. Theory*, vol. 14, no.1-2, pp. 1162, Jun. 2018.

Stochastic Block Model (SBM)

- Connectivity depends on community membership[4].
- Stochastic equivalence of nodes within same community



N: no. nodes, K: no. communities

c_i: membership of node i

$c_i \in \{1, 2, ..., K\}$, $\mathcal{C} = \{c_i\}_{i=1}^{N}$

$y_{ab} \in \{0, 1\}$: (a, b)'th entry in adj. matrix

$\beta_{k\ell} \in (0, 1)$: link probability between two nodes in community k and ℓ

$$y_{ab} | (c_a = k, c_b = \ell) \sim Bernoulli(\beta_{k\ell})$$

[4] E. Abbe, "Community detection and stochastic block models, *Found. and Trends Commun. and Inform. Theory*, vol. 14, no.1-2, pp. 1162, Jun. 2018.

Stochastic Block Model (SBM)[5]

$c_i \in \{1, 2, ..., K\}$, $\mathcal{C} = \{c_i\}_{i=1}^{N}$

$y_{ab} \in \{0, 1, 2, ...\}$: (a, b)'th entry in adj. matrix

$\omega_{k\ell} \in \mathbb{R}_+$: average number of links between two nodes in community k and ℓ

[5] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks, *Phys. Rev. E*, vol. 83,no. 1, pp. 016107, Jan. 2011.

Stochastic Block Model (SBM)[5]

$c_i \in \{1, 2, ..., K\}$, $\mathcal{C} = \{c_i\}_{i=1}^{N}$

$y_{ab} \in \{0, 1, 2, ...\}$: (a, b)'th entry in adj. matrix

$\omega_{k\ell} \in \mathbb{R}_+$: average number of links between two nodes in community k and ℓ

$$y_{ab}|(c_a = k, c_b = \ell) \sim Poisson(\omega_{k\ell}),$$
$$\mathcal{L}(\mathbf{y}|\mathcal{C}, \omega) = \sum_{k,\ell}(m_{k\ell} \log \omega_{k\ell} - n_k n_\ell \omega_{k\ell}),$$

[5] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks, *Phys. Rev. E*, vol. 83,no. 1, pp. 016107, Jan. 2011.

Stochastic Block Model (SBM)[5]

$c_i \in \{1, 2, ..., K\}$, $\mathcal{C} = \{c_i\}_{i=1}^N$

$y_{ab} \in \{0, 1, 2, ...\}$: (a, b)'th entry in adj. matrix

$\omega_{k\ell} \in \mathbb{R}_+$: average number of links between two nodes in community k and ℓ

$$y_{ab} | (c_a = k, c_b = \ell) \sim Poisson(\omega_{k\ell}) \,,$$
$$\mathcal{L}(\mathbf{y}|\mathcal{C}, \omega) = \sum_{k,\ell} (m_{k\ell} \log \omega_{k\ell} - n_k n_\ell \omega_{k\ell}) \,,$$

where, $m_{k\ell} = \sum_{a,b} y_{ab} \mathbf{1}_{\{c_a = k, c_b = \ell\}}$ and $n_k = \sum_a \mathbf{1}_{\{c_a = k\}}$

[5] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks, *Phys. Rev. E*, vol. 83,no. 1, pp. 016107, Jan. 2011.

Stochastic Block Model (SBM)

ML estimate:

$$\hat{\omega}_{k\ell} = \frac{m_{k\ell}}{n_k n_\ell}$$

Stochastic Block Model (SBM)

ML estimate:

$$\hat{\omega}_{k\ell} = \frac{m_{k\ell}}{n_k n_\ell}$$

$$\mathcal{L}(\mathbf{y}|\mathcal{C}) = \max_{\omega} \mathcal{L}(\mathbf{y}|\mathcal{C}, \omega) = \sum_{k,\ell} \left(\frac{m_{k\ell}}{m} \right) \log \frac{(m_{k\ell}/m)}{(n_k n_\ell / N^2)} ,$$

where, $m = \sum_{k,\ell} m_{k,\ell}$.

Stochastic Block Model (SBM)

ML estimate:

$$\hat{\omega}_{k\ell} = \frac{m_{k\ell}}{n_k n_\ell}$$

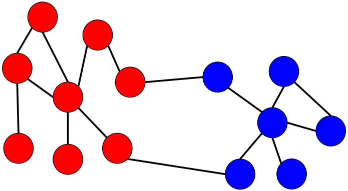
$$\mathcal{L}(\mathbf{y}|\mathcal{C}) = \max_{\omega} \mathcal{L}(\mathbf{y}|\mathcal{C}, \omega) = \sum_{k,\ell} \left(\frac{m_{k\ell}}{m} \right) \log \frac{(m_{k\ell}/m)}{(n_k n_\ell / N^2)} ,$$

where, $m = \sum_{k,\ell} m_{k,\ell}$.

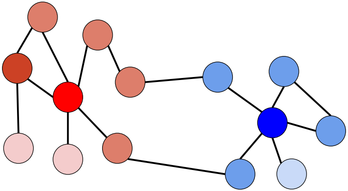
Greedy algorithm

pick a random node, place it in a community to maximally increase the objective.

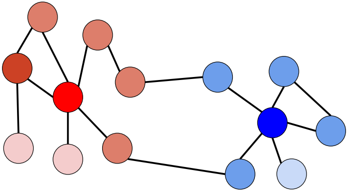
Degree Corrected Stochastic Block Model (DC-SBM)



Degree Corrected Stochastic Block Model (DC-SBM)

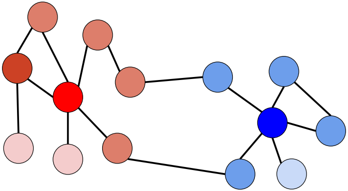


Degree Corrected Stochastic Block Model (DC-SBM)



- degree heterogeneity within community
- $\theta_a \in (0, 1)$: degree correction parameters

Degree Corrected Stochastic Block Model (DC-SBM)

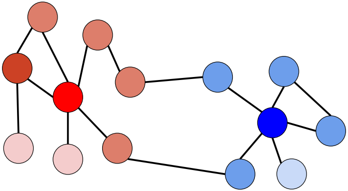


- degree heterogeneity within community
- $\theta_a \in (0, 1)$: degree correction parameters

$$y_{ab} | (c_a = k, c_b = \ell) \sim Poisson(\theta_a \theta_b \omega_{k\ell}) \,,$$

$$\sum_a \theta_a \mathbf{1}_{\{c_a = k\}} = 1 \,.$$

Degree Corrected Stochastic Block Model (DC-SBM)



- degree heterogeneity within community
- $\theta_a \in (0,1)$: degree correction parameters

$$y_{ab} | (c_a = k, c_b = \ell) \sim \mathit{Poisson}(\theta_a \theta_b \omega_{k\ell}) \,,$$

$$\sum_a \theta_a \mathbf{1}_{\{c_a = k\}} = 1 \,.$$

$$\mathcal{L}(\mathbf{y}|\mathcal{C}) = \max_{\omega,\theta} \mathcal{L}(\mathbf{y}|\mathcal{C},\omega,\theta) = \sum_{k,\ell} \left(\frac{m_{k\ell}}{m} \right) \log \frac{(m_{k\ell}/m)}{(\kappa_k \kappa_\ell / m^2)} \,,$$

where, $\kappa_k = \sum_a d_a \mathbf{1}_{\{c_a = k\}} \,.$

Mixed Membership Stochastic Blockmodel (MMSB)

- Overlapping communities[6]

[6]E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," in *J. Mach. Learn.Res.*, vol. 9, pp. 19812014, Jun. 2008.

Mixed Membership Stochastic Blockmodel (MMSB)

- Overlapping communities[6]
- Community membership probability: $\pi_a \in (0,1)^K, \sum_{k=1}^{K} \pi_{ak} = 1$

[6]E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," in *J. Mach. Learn.Res.*, vol. 9, pp. 19812014, Jun. 2008.

Mixed Membership Stochastic Blockmodel (MMSB)

- Overlapping communities[6]
- Community membership probability: $\pi_a \in (0,1)^K, \sum_{k=1}^{K} \pi_{ak} = 1$
- Prior distributions: $\beta_{k\ell} \sim Beta(\eta), \pi_a \sim Dir(\alpha)$

[6] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," in *J. Mach. Learn.Res.*, vol. 9, pp. 19812014, Jun. 2008.

Mixed Membership Stochastic Blockmodel (MMSB)

- Overlapping communities[6]
- Community membership probability: $\pi_a \in (0,1)^K, \sum_{k=1}^{K} \pi_{ak} = 1$
- Prior distributions: $\beta_{k\ell} \sim Beta(\eta), \pi_a \sim Dir(\alpha)$

Generative Model

for any two nodes a and b :

sample $Z_{ab} \sim \pi_a$ and $Z_{ba} \sim \pi_b$

sample $y_{ab} | (Z_{ab} = k, Z_{ba} = \ell) \sim Bernoulli(\beta_{k\ell})$

[6] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," in *J. Mach. Learn.Res.*, vol. 9, pp. 19812014, Jun. 2008.

Mixed Membership Stochastic Blockmodel (MMSB)

- Overlapping communities[6]
- Community membership probability: $\pi_a \in (0,1)^K, \sum_{k=1}^{K} \pi_{ak} = 1$
- Prior distributions: $\beta_{k\ell} \sim Beta(\eta), \pi_a \sim Dir(\alpha)$

Generative Model

for any two nodes a and b :

sample $Z_{ab} \sim \pi_a$ and $Z_{ba} \sim \pi_b$

sample $y_{ab} | (Z_{ab} = k, Z_{ba} = \ell) \sim Bernoulli(\beta_{k\ell})$

- Posterior inference of $p(\beta, \pi | \mathbf{y})$
- assortative MMSB (a-MMSB): $\beta_{k\ell} = \delta$ for $k \neq \ell$

[6] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, "Mixed membership stochastic blockmodels," in *J. Mach. Learn.Res.*, vol. 9, pp. 19812014, Jun. 2008.

Inference in a-MMSB

- Variational inference[7]

[7] P. K. Gopalan, S. Gerrish, M. Freedman, D. M. Blei, and D. M. Mimno, "Scalable inference of overlapping communities," in *Proc. Adv. Neural Inf. Proc. Systems*, Dec. 2012.

[8] W. Li, S. Ahn, and M. Welling, "Scalable MCMC for mixed membership stochastic blockmodels," in *Proc. Artificial Intell.and Statist.*, May 2016.

Inference in a-MMSB

- Variational inference[7]
  - Mean field approximation
  - Stochastic gradient optimization
  - Outperforms traditional techniques

[7] P. K. Gopalan, S. Gerrish, M. Freedman, D. M. Blei, and D. M. Mimno, "Scalable inference of overlapping communities," in *Proc. Adv. Neural Inf. Proc. Systems*, Dec. 2012.

[8] W. Li, S. Ahn, and M. Welling, "Scalable MCMC for mixed membership stochastic blockmodels," in *Proc. Artificial Intell.and Statist.*, May 2016.

Inference in a-MMSB

- Variational inference[7]
  - Mean field approximation
  - Stochastic gradient optimization
  - Outperforms traditional techniques
- Markov chain Monte Carlo[8]

[7] P. K. Gopalan, S. Gerrish, M. Freedman, D. M. Blei, and D. M. Mimno, "Scalable inference of overlapping communities," in *Proc. Adv. Neural Inf. Proc. Systems*, Dec. 2012.

[8] W. Li, S. Ahn, and M. Welling, "Scalable MCMC for mixed membership stochastic blockmodels," in *Proc. Artificial Intell.and Statist.*, May 2016.

Inference in a-MMSB

- Variational inference[7]
  - Mean field approximation
  - Stochastic gradient optimization
  - Outperforms traditional techniques
- Markov chain Monte Carlo[8]
  - Stochastic gradient Riemannian Langevin dynamics (SGRLD)
  - Faster convergence
  - Better approximation of posterior

[7] P. K. Gopalan, S. Gerrish, M. Freedman, D. M. Blei, and D. M. Mimno, "Scalable inference of overlapping communities," in *Proc. Adv. Neural Inf. Proc. Systems*, Dec. 2012.

[8] W. Li, S. Ahn, and M. Welling, "Scalable MCMC for mixed membership stochastic blockmodels," in *Proc. Artificial Intell.and Statist.*, May 2016.

Mixed Membership Degree Corrected Blockmodel (MMDCB)[9]

- Generalization of a-MMSB

[9] S. Pal and M. Coates, "Scalable MCMC in degree corrected stochastic block model," in *Proc. Intl. Conf. Acoust., Speech and Signal Proc*, May 2019 (Accepted).

Mixed Membership Degree Corrected Blockmodel (MMDCB)[9]

- Generalization of a-MMSB
- Node specific degree correction parameters: $r_a \in \mathbb{R}$

[9] S. Pal and M. Coates, "Scalable MCMC in degree corrected stochastic block model," in *Proc. Intl. Conf. Acoust., Speech and Signal Proc*, May 2019 (Accepted).

Mixed Membership Degree Corrected Blockmodel (MMDCB)[9]

- Generalization of a-MMSB
- Node specific degree correction parameters: $r_a \in \mathbb{R}$
- Community specific parameters: $q_k > 0$

[9] S. Pal and M. Coates, "Scalable MCMC in degree corrected stochastic block model," in *Proc. Intl. Conf. Acoust., Speech and Signal Proc*, May 2019 (Accepted).

Mixed Membership Degree Corrected Blockmodel (MMDCB)[9]

- Generalization of a-MMSB
- Node specific degree correction parameters: $r_a \in \mathbb{R}$
- Community specific parameters: $q_k > 0$

Generative Model

for any two nodes a and b :

sample $Z_{ab} \sim \pi_a$ and $Z_{ba} \sim \pi_b$

if $Z_{ab} = Z_{ba} = k$:

sample $y_{ab} \sim Bernoulli(logit^{-1}(q_k + r_a + r_b))$

else:

sample $y_{ab} \sim Bernoulli(logit^{-1}(r_a + r_b))$

[9] S. Pal and M. Coates, "Scalable MCMC in degree corrected stochastic block model," in *Proc. Intl. Conf. Acoust., Speech and Signal Proc*, May 2019 (Accepted).

Mixed Membership Degree Corrected Blockmodel (MMDCB)[9]

- Generalization of a-MMSB
- Node specific degree correction parameters: $r_a \in \mathbb{R}$
- Community specific parameters: $q_k > 0$

Generative Model

for any two nodes a and b :

  sample $Z_{ab} \sim \pi_a$ and $Z_{ba} \sim \pi_b$

  if $Z_{ab} = Z_{ba} = k$:

    sample $y_{ab} \sim Bernoulli(logit^{-1}(q_k + r_a + r_b))$

  else:

    sample $y_{ab} \sim Bernoulli(logit^{-1}(r_a + r_b))$

- Prior distributions: $r_a \sim \mathcal{N}(0, \sigma^2), q_k \sim \mathcal{N}(0, \sigma^2)\mathbf{1}_{\{q_k > 0\}}$
- Posterior inference of $p(\mathbf{q}, \mathbf{r}, \pi | \mathbf{y})$

[9] S. Pal and M. Coates, "Scalable MCMC in degree corrected stochastic block model," in *Proc. Intl. Conf. Acoust., Speech and Signal Proc*, May 2019 (Accepted).

Langevin Monte Carlo

Metropolis adjusted Langevin algorithm (MALA)

- parameter θ, observed data $\mathbf{X} = \{x_1, x_2, ..., x_N\}$

Langevin Monte Carlo

Metropolis adjusted Langevin algorithm (MALA)

- parameter θ, observed data $\mathbf{X} = \{x_1, x_2, ..., x_N\}$
- prior distribution $p(\theta)$, generative model $p(\mathbf{X}|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$

Langevin Monte Carlo

Metropolis adjusted Langevin algorithm (MALA)

- parameter θ, observed data $\mathbf{X} = \{x_1, x_2, ..., x_N\}$
- prior distribution $p(\theta)$, generative model $p(\mathbf{X}|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$
- posterior distribution: $p(\theta|\mathbf{X}) \propto p(\theta) \prod_{i=1}^{N} p(x_i|\theta)$

Langevin Monte Carlo

Metropolis adjusted Langevin algorithm (MALA)

- parameter θ, observed data $\mathbf{X} = \{x_1, x_2, ..., x_N\}$
- prior distribution $p(\theta)$, generative model $p(\mathbf{X}|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$
- posterior distribution: $p(\theta|\mathbf{X}) \propto p(\theta) \prod_{i=1}^{N} p(x_i|\theta)$
- $q(\theta^*|\theta) = \mathcal{N}(\theta^*|\theta + \frac{\epsilon}{2}\Big(\nabla_\theta \log p(\theta) + \sum_{i=1}^{N} \nabla_\theta \log p(x_i|\theta)\Big), \epsilon I)$

Langevin Monte Carlo

Metropolis adjusted Langevin algorithm (MALA)

- parameter θ, observed data $\mathbf{X} = \{x_1, x_2, ..., x_N\}$
- prior distribution $p(\theta)$, generative model $p(\mathbf{X}|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$
- posterior distribution: $p(\theta|\mathbf{X}) \propto p(\theta) \prod_{i=1}^{N} p(x_i|\theta)$
- $q(\theta^*|\theta) = \mathcal{N}\left(\theta^*|\theta + \frac{\epsilon}{2}\left(\nabla_\theta \log p(\theta) + \sum_{i=1}^{N} \nabla_\theta \log p(x_i|\theta)\right), \epsilon I\right)$
- acceptance probability: $\min\left(1, \frac{p(\theta^*|\mathbf{x})q(\theta|\theta^*)}{p(\theta|\mathbf{x})q(\theta^*|\theta)}\right)$

Langevin Monte Carlo

Metropolis adjusted Langevin algorithm (MALA)

- parameter θ, observed data $\mathbf{X} = \{x_1, x_2, ..., x_N\}$
- prior distribution $p(\theta)$, generative model $p(\mathbf{X}|\theta) = \prod_{i=1}^{N} p(x_i|\theta)$
- posterior distribution: $p(\theta|\mathbf{X}) \propto p(\theta) \prod_{i=1}^{N} p(x_i|\theta)$
- $q(\theta^*|\theta) = \mathcal{N}\left(\theta^*|\theta + \frac{\epsilon}{2}\left(\nabla_\theta \log p(\theta) + \sum_{i=1}^{N} \nabla_\theta \log p(x_i|\theta)\right), \epsilon I\right)$
- acceptance probability: $\min\left(1, \frac{p(\theta^*|\mathbf{x})q(\theta|\theta^*)}{p(\theta|\mathbf{x})q(\theta^*|\theta)}\right)$
- Preconditioning: Riemannian Langevin Dynamics (RLD)

Langevin Monte Carlo

Stochastic gradient Langevin dynamics (SGLD)

- complexity in LD: $\mathcal{O}(N)$

[10] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient Langevin dynamics," in *Proc. Intl. Conf. Machine Learning*, Bellevue, WA, USA, Jun. 2011.

Langevin Monte Carlo

Stochastic gradient Langevin dynamics (SGLD)

- complexity in LD: $\mathcal{O}(N)$
- stochastic gradient: $\mathcal{O}(n)$ complexity

[10] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient Langevin dynamics," in *Proc. Intl. Conf. Machine Learning*, Bellevue, WA, USA, Jun. 2011.

Langevin Monte Carlo

Stochastic gradient Langevin dynamics (SGLD)

- complexity in LD: $\mathcal{O}(N)$
- stochastic gradient: $\mathcal{O}(n)$ complexity
- $\nabla_\theta \log p(\mathbf{X}|\theta) \approx \frac{N}{n} \sum_{x_{ti} \in \mathbf{X_t}} \nabla_\theta \log p(x_{ti}|\theta)$

[10] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient Langevin dynamics," in *Proc. Intl. Conf. Machine Learning*, Bellevue, WA, USA, Jun. 2011.

Langevin Monte Carlo

Stochastic gradient Langevin dynamics (SGLD)

- complexity in LD: $\mathcal{O}(N)$
- stochastic gradient: $\mathcal{O}(n)$ complexity
- $\nabla_\theta \log p(\mathbf{X}|\theta) \approx \frac{N}{n} \sum_{x_{ti} \in \mathbf{X_t}} \nabla_\theta \log p(x_{ti}|\theta)$
- annealed step-size schedule $\sum_{t=1}^{\infty} \epsilon_t = \infty$ and $\sum_{t=1}^{\infty} \epsilon_t^2 < \infty$

[10] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient Langevin dynamics," in *Proc. Intl. Conf. Machine Learning*, Bellevue, WA, USA, Jun. 2011.

Langevin Monte Carlo

Stochastic gradient Langevin dynamics (SGLD)

- complexity in LD: $\mathcal{O}(N)$
- stochastic gradient: $\mathcal{O}(n)$ complexity
- $\nabla_\theta \log p(\mathbf{X}|\theta) \approx \frac{N}{n} \sum_{x_{ti} \in \mathbf{X_t}} \nabla_\theta \log p(x_{ti}|\theta)$
- annealed step-size schedule $\sum_{t=1}^{\infty} \epsilon_t = \infty$ and $\sum_{t=1}^{\infty} \epsilon_t^2 < \infty$
- no acceptance probability computation

[10] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient Langevin dynamics," in *Proc. Intl. Conf. Machine Learning*, Bellevue, WA, USA, Jun. 2011.

Langevin Monte Carlo

Stochastic gradient Langevin dynamics (SGLD)

- complexity in LD: $\mathcal{O}(N)$
- stochastic gradient: $\mathcal{O}(n)$ complexity
- $\nabla_\theta \log p(\mathbf{X}|\theta) \approx \frac{N}{n} \sum_{x_{ti} \in \mathbf{X_t}} \nabla_\theta \log p(x_{ti}|\theta)$
- annealed step-size schedule $\sum_{t=1}^{\infty} \epsilon_t = \infty$ and $\sum_{t=1}^{\infty} \epsilon_t^2 < \infty$
- no acceptance probability computation
- asymptotic convergence[10] to the posterior distribution

[10] M. Welling and Y. W. Teh, "Bayesian learning via stochastic gradient Langevin dynamics," in *Proc. Intl. Conf. Machine Learning*, Bellevue, WA, USA, Jun. 2011.

Numerical Experiments and Results

	NETSCIENCE	RELATIVITY	HEP-TH	HEP-PH[11]
Nodes	1589	5242	9877	12008
Edges	2742	14996	25998	118521

[11] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: densification and shrinking diameters," in *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, Mar. 2007

Numerical Experiments and Results

	NETSCIENCE	RELATIVITY	HEP-TH	HEP-PH[11]
Nodes	1589	5242	9877	12008
Edges	2742	14996	25998	118521

- held out test set: 10% of the links, same number of non-links

[11] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: densification and shrinking diameters," in *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, Mar. 2007

Numerical Experiments and Results

	NETSCIENCE	RELATIVITY	HEP-TH	HEP-PH[11]
Nodes	1589	5242	9877	12008
Edges	2742	14996	25998	118521

- held out test set: 10% of the links, same number of non-links
- evaluation metrics:
  - average perplexity:

$$perp_{avg}(\mathbf{Y}_{\mathbf{test}} | \{\pi^{(i)}, q^{(i)}, r^{(i)}\}_{i=1}^{T}) = \exp\left(- \frac{\sum_{y_{ab} \in \mathbf{Y}_{\mathbf{test}}} \log \left\{ \frac{1}{T} \sum_{i=1}^{T} p(y_{ab} | \pi^{(i)}, q^{(i)}, r^{(i)}) \right\}}{|\mathbf{Y}_{\mathbf{test}}|} \right).$$

  - area under ROC (AUC) for link prediction task

[11] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: densification and shrinking diameters," in *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, Mar. 2007

Convergence of Perplexity

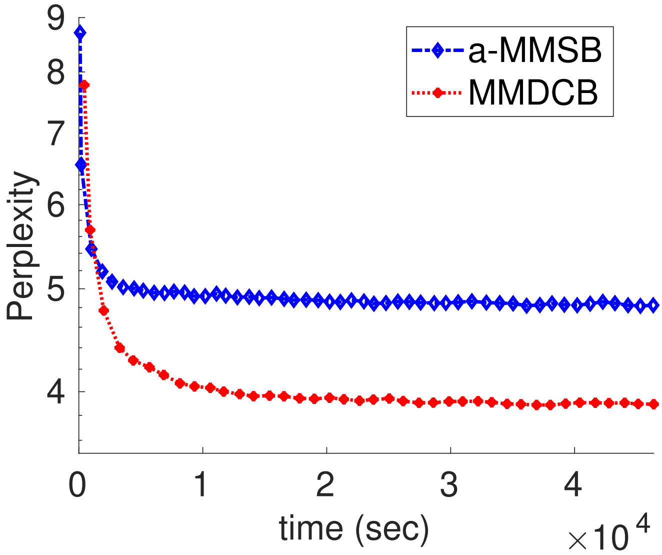


Figure: Convergence of perplexity for HEP-PH dataset, $K = 50$

Comparison of Perplexity at convergence

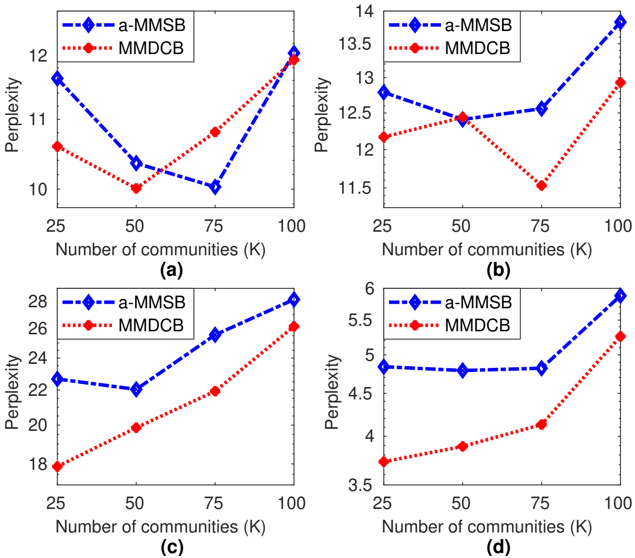


Figure: (a) NETSCIENCE, (b) RELATIVITY, (c) HEP-TH and (d) HEP-PH

Comparison of AUC at convergence

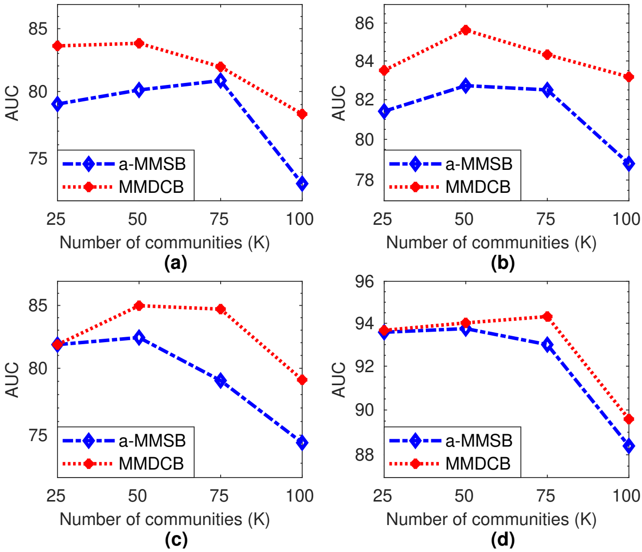


Figure: (a) NETSCIENCE, (b) RELATIVITY, (c) HEP-TH and (d) HEP-PH

Conclusion

- MMDCB models the observed graph better than a-MMSB.
- SG-MCMC algorithms scale well to large networks.
- Future work:
  - better generative models
  - efficient mini-batch sampling, variance reduction
  - more advanced SG-MCMC algorithms