On Dynamic Network Models and Application to Causal Impact

ELECTRICAL & COMPUTER ENGINEERING

Yu-Chia Chen [1] Avleen S. Bijral [2] Juan Lavista Ferres [2]

[1]University of Washington yuchaz@uw.edu [2]Microsoft {avbijral, jlavista}@microsoft.com

Introduction & Related Works

Introduction Stochastic block model (SBM) is a commonly used model for networks and can be estimated by various methods. Dynamic extensions of SBM are much harder to estimate, especially in the large scale setting. We propose a scalable method for this problem and introduce the novel problem of estimating impact of external events on networks.

Degree-corrected stochastic block model (DC-SBM) [2] Given the observed unweighted adjacency matrix $\mathbf{A}$ and the labels $\mathbf{c}$, edges $e = (i, j)$ are drawn i.i.d in Bernoulli trial with probability $P_{c_i c_j}$ corrected by degree heterogeneity parameter $\boldsymbol{\vartheta}$. In mathematical terms, the expected values of an edge given the labels $\mathbf{c}$ is

$$\mathbb{E}[\mathbf{A}_{ij}|\mathbf{c}] = \vartheta_i \vartheta_j P_{c_i, c_j} \quad (1)$$

MLE of (i) $\mathbf{P}$ is equal to the number success trial with appropriate normalization, i.e., $P_{lk} \propto \sum_{ij} A_{ij} 1(c_i = l, c_j = k)$, and (ii) $\boldsymbol{\vartheta}$ is proportional to the degree of node i, i.e., $\vartheta_i \propto \deg(i)$, with an identifiability constraint $\sum_i \vartheta_i 1(c_i = l) = 1 \, \forall \, l \in [K]$.

Related works – Dynamic network models

- Latent variable model (Sarkar & Moore, 2006; Sewell & Chen, 2015).
- State space model for dynamic SBM (Yang et al., 2011; Xu & Hero, 2014 [3]).
- Low rank sparse optimization (Bao & Michailidis, 2018).

The proposed method scales well because (i) model on pseudo-observation *block sums* $\mathbf{B}$, (ii) efficient relabeling step $\mathcal{O}(K^3)$ and (iii) using SGD in solving optimization problem which provides us with advantage over MCMC based approaches.

Related work – Causal impact on graphs

- Experimentation in networks (Basse & Airoldi, 2018; Gui et al., 2015; Sussman & Airoldi, 2017).
- Estimating causal impact of peer influence in networks (Toulis & Kao, 2013).
- Change point detection in networks data (De Ridder et al., 2016; Peel & Clauset, 2015).

We are interested in modeling *out* the impact of the exogenous event from the network generation dynamics. To the best of our knowledge, we are not aware of work that tackles this problem.

Conditional Pseudo-likelihood (SPL) [1]

The SPL method simplifies the combinatorial estimation of a block model by modeling instead the pseudo-observation *block sums* $\mathbf{B}(e) = \mathbf{A1}(e) \in \mathbb{R}^{N \times K}$ computed using an initialization of the community membership $\mathbf{e}$. The graphical representation of SPL model can be found in Figure 1, with the corresponding likelihood function is,

$$\ell(\boldsymbol{\pi}, \boldsymbol{\Theta}; \mathbf{B}) = \sum_{i=1}^{N} \log \left(\sum_{l=1}^{K} \pi_l \exp \left(\sum_{q=1}^{K} b_{iq} \log \theta_{lq} \right) \right) \quad (2)$$

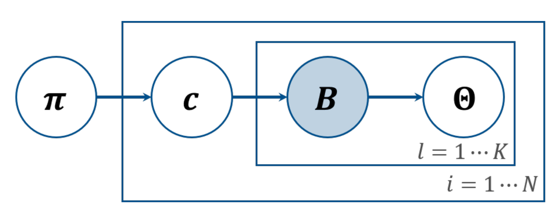


Figure 1. Graphical representation of SPL.

The optimal $\boldsymbol{\pi}, \boldsymbol{\Theta}$ parameter given $\mathbf{e}$ can be obtained by EM algorithm. Upon obtaining the MLE of $\boldsymbol{\pi}, \boldsymbol{\Theta}$, one can update $\hat{\mathbf{c}}$ by the posterior distribution $\psi_{il} = \Pr(c_i = l | \mathbf{b}_i, \boldsymbol{\pi}, \boldsymbol{\Theta})$ as in (3).

$$\psi_{il} = \frac{\pi_l \exp(\sum_q b_{iq} \log \theta_{lq})}{\sum_k \pi_k \exp(\sum_q b_{iq} \log \theta_{lq})} \quad (3)$$

The author proposed to initialize the parameters as in (4).

$$\boldsymbol{\pi} = \mathbf{n}/N, \quad \boldsymbol{\Lambda} = \mathrm{diag}(\mathbf{n})\mathbf{P}, \quad \boldsymbol{\Theta} = \mathbf{D}_{\boldsymbol{\Lambda}}^{-1}\boldsymbol{\Lambda} \quad (4)$$

where $[\mathrm{diag}(\mathbf{D}_{\boldsymbol{\Lambda}}^{-1})]_l = \sum_k \lambda_{lk}$

Algorithm 1: Conditional Pseudo-likelihood algorithm

```
CPL_static (A, e, M)
Set ĉ₀ ← e, initialize π, Λ using eq. (4) with e
for m = 1 : M do
    Update block sums B with estimated labels ĉ_{m−1}
    while not converge do
        E Step: Compute posterior Ψ using eq. (3)
        M step: π̂_l = Σ_i ψ_il / N;  θ̂_lk = Σ_i ψ_il b_ik / Σ_i ψ_il d_i
    end
    Update the labels: [ĉ_m]_i = arg max_k ψ_ik
end
Return: ĉ, Θ̂, π̂
```

Dynamic Pseudo-likelihood (DPL) Estimation

Parameters $\boldsymbol{\pi}$ and $\boldsymbol{\Lambda}$ govern the underlying structure of the observed network. A natural dynamic extension of SPL model is to propose stochastic processes on these two parameters, i.e., random walk priors on $\boldsymbol{\pi}$ and $\boldsymbol{\Lambda}$, that capture some underlying intuition of how the networks evolve.

The graphical representation of the proposed stochastic processes is shown in Figure 2, and the corresponding random walk priors on $\boldsymbol{\pi}^{1:T}$ and $\boldsymbol{\Lambda}^{1:T}$ are

$$\Pr\left(\zeta_l^t \,|\, \zeta_l^{t-1}; \sigma_1\right) = \mathcal{N}\left(\zeta_l^t; \zeta_l^{t-1}, \sigma_1\right) \; \forall \, l \in [K]$$
$$\Pr\left(\lambda_{lk}^t \,|\, \lambda_{lk}^{t-1}; \sigma_2\right) = \mathcal{N}_T\left(\lambda_{lk}^t; \lambda_{lk}^{t-1}, \sigma_2, 0, \infty\right) \; \forall \, l, k \in [K] \quad (5)$$

DPL (cont.) & Relabeling

MAP estimator $\hat{\boldsymbol{\pi}}^{1:T}, \hat{\boldsymbol{\Lambda}}^{1:T}$ is the minimizer of (6) with $\chi_1 = \frac{1}{2\sigma_1^2}$ and $\chi_2 = \frac{1}{2\sigma_2^2}$.

$$f(\boldsymbol{\zeta}^{1:T}, \boldsymbol{\Lambda}^{1:T} | \boldsymbol{B}^{1:T}; \chi_1, \chi_2) = -\sum_{t=1}^{T} \ell(\boldsymbol{\pi}^t, \boldsymbol{\Theta}^t; \boldsymbol{B}^t) + \chi_1 \sum_{t=2}^{T} \|\boldsymbol{\zeta}^t - \boldsymbol{\zeta}^{t-1}\|^2 + \chi_2 \sum_{t=2}^{T} \|\boldsymbol{\Lambda}^t - \boldsymbol{\Lambda}^{t-1}\|_F^2$$
$$\text{s.t. } \lambda_{lk} \geq 0 \text{ for all } l, k$$
with ℓ from eq. (2), $\boldsymbol{\pi} = \mathrm{SoftMax}(\boldsymbol{\zeta})$ and Θ from eq. (4). (6)

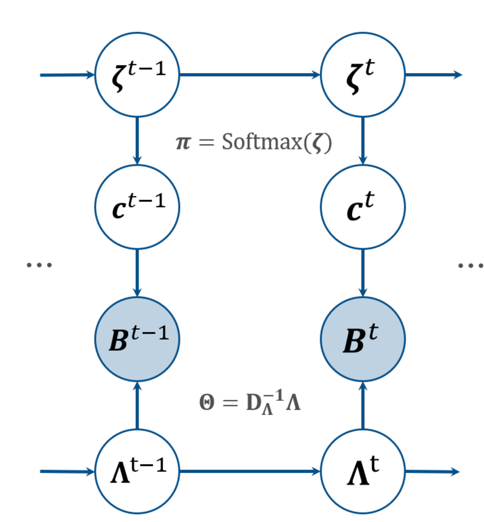


Figure 2. Graphical representation of the proposed DPL model.

Algorithm 2: Dynamic Pseudo-likelihood algorithm

```
DPL_estimate (A^{1:T}, e^{1:T}, M, χ1, χ2)
Set ĉ_0^{1:T} = e^{1:T}
for m = 1 : M do
    Initialize π^{1:T}, Λ^{1:T} with ĉ_{m−1}^{1:T} using eq. (4)
    Update B^{1:T} with ĉ_{m−1}^{1:T}
    Solve eq. (6): ζ̂^{1:T}, Λ̂^{1:T} = arg min f(ζ^{1:T}, Λ^{1:T} | B^{1:T}; χ1, χ2)
    Compute posterior Ψ^{1:T} using eq. (3)
    Ψ^{1:T} = relabel_with_posterior(Ψ^{1:T})
     ▷ algorithm 3
    Update the labels: [ĉ_m^t]_i = arg max_k ψ_ik^t
end
ĉ^{1:T} = ĉ_M^{1:T}; Update parameter π̂^{1:T}, Λ̂^{1:T} with eq. (4)
Return: ĉ^{1:T}, Λ̂^{1:T}, π̂^{1:T}
```

Permutations & relabeling

- *Global permutation*: labels of different time steps are defined consistently (up to a permutation of labels globally).
- *Local permutation*: labels of different time steps are not defined consistently.

$$\nu_*^t = \arg\min_{\nu} \sum_{i=1}^{N} \mathrm{KL}\left(\psi_{i,\nu(k)}^{t-1} \,\|\, \psi_{i,k}^t\right) \quad (7)$$

Can be cast into an assignment problem:

$$\min_{x} \sum_{l,k} C_{lk} x_{lk} \text{ s.t. } \sum_{l} x_{lk} = 1; \sum_{k} x_{lk} = 1 \quad (8)$$

Algorithm 3: Relabel with posterior

```
relabel_with_posterior (Ψ^{1:T})
for t = 2 : T do
    ν_*^t = arg min_ν Σ_i^N KL(ψ_{i,ν(k)}^{t−1} || ψ_{i,k}^t)
    Global mapping ν_g^t(k) = ν_*^t(ν_g^{t−1}(k))
     ▷ ν_g^1 is identity
    for k = 1 : K do
        [Ψ_*^t]_{·,k} = Ψ^t_{·,ν_g^t(k)}
    end
end
Return: Aligned posterior Ψ_*^{1:T}
```

Causal Impact

Level shift estimation An external event at $t = T_0$ causes a level shift in $\boldsymbol{\Lambda}$.

$$\Pr\left(\boldsymbol{\Lambda}^{1:T}\right) = \prod_{t \neq T_0} \Pr\left(\boldsymbol{\Lambda}^t \,|\, \boldsymbol{\Lambda}^{t-1}\right) \quad (9)$$

Drift estimation An external event at $t = T_0$ changes the increasing/decreasing trend in $\boldsymbol{\Lambda}$.

$$\Pr\left(\lambda_{ll}^t \,|\, \lambda_{ll}^{t-1}; \sigma_2, \delta_l\right) = \mathcal{N}_T\left(\lambda_{ll}^t; \lambda_{ll}^{t-1} + \delta_l, \sigma_2, 0, \infty\right) \; \forall \, l \in [K] \quad (10)$$

With the above modification, the objective shown in (6) becomes,

$$f(\boldsymbol{\zeta}^{1:T}, \boldsymbol{\Lambda}^{1:T} | \boldsymbol{B}^{1:T}; \chi_1, \chi_2) = -\sum_{t=1}^{T} \ell(\boldsymbol{\pi}^t, \boldsymbol{\Theta}^t; \boldsymbol{B}^t) + \chi_1 \sum_{t=2}^{T} \|\boldsymbol{\zeta}^t - \boldsymbol{\zeta}^{t-1}\|^2$$
$$+ \chi_2 \sum_{t=2}^{T_0 - 1} \|\boldsymbol{\Lambda}^t - \boldsymbol{\Lambda}^{t-1} - \mathrm{diag}(\boldsymbol{\delta}_1)\|_F^2 + \chi_2 \sum_{t=T_0+1}^{T} \|\boldsymbol{\Lambda}^t - \boldsymbol{\Lambda}^{t-1} - \mathrm{diag}(\boldsymbol{\delta}_2)\|_F^2 \quad (11)$$
s.t. $\lambda_{lk} \geq 0$ for all l, k; with $\boldsymbol{\pi} = \mathrm{SoftMax}(\boldsymbol{\zeta})$ and Θ from eq. (4).

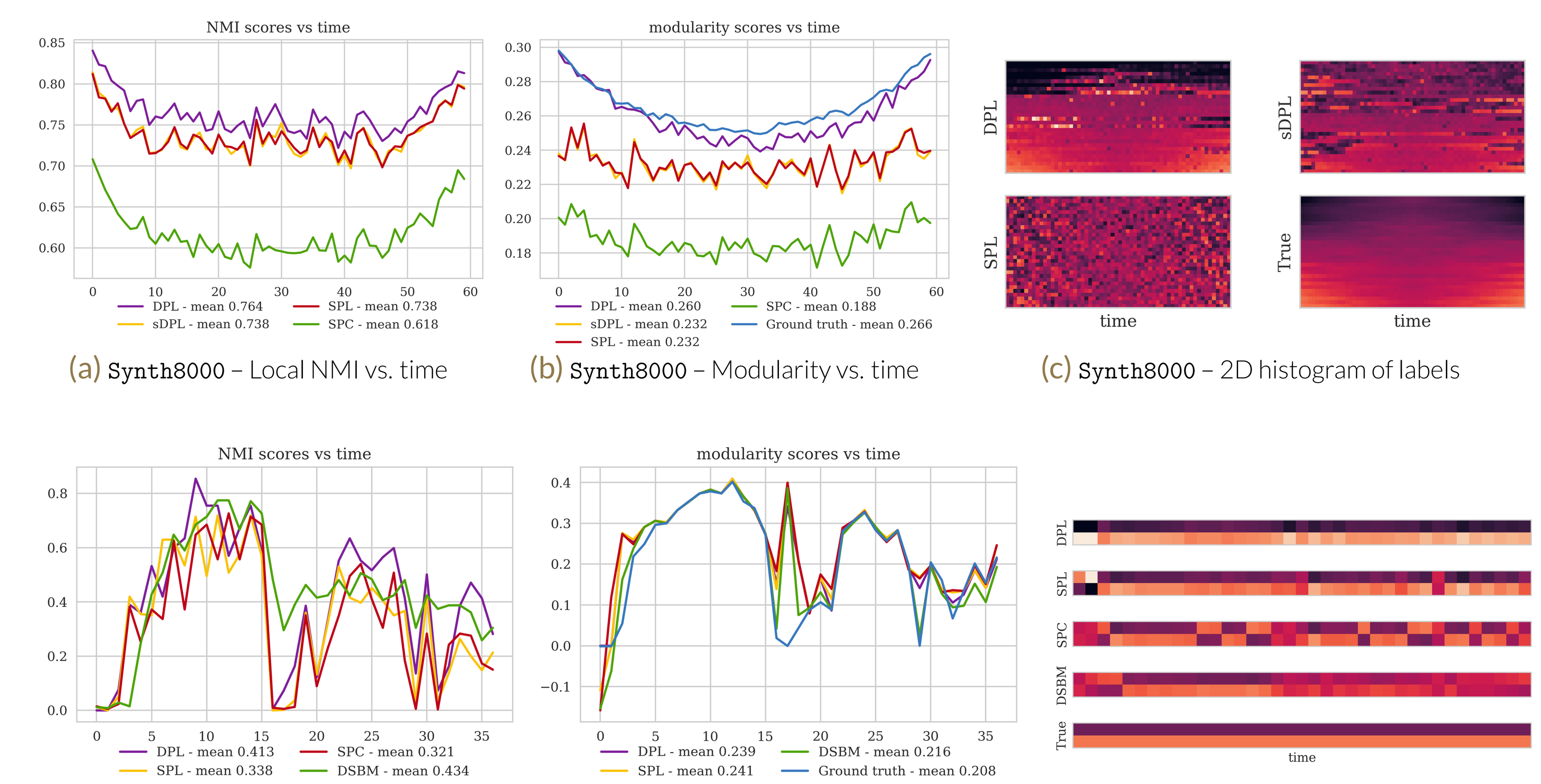
Impact assessment and re-sampling

- The uncertainty bands $\mathcal{B}_\alpha$ of the *counterfactual* model $\mathcal{W}$ with significance level α allows us to determine whether the change of the network parameters are significant.
- Based on the aforementioned DPL model, we propose a resample procedure to estimate the uncertainty band $\mathcal{B}_\alpha$ by repeatedly generating J resamples of series of networks corresponding to *impact model* $\mathcal{I}$ and the *counterfactual model* $\mathcal{W}$.
- A point estimate of the edge probability matrices of model $\mathcal{I}$ can be obtained by the average of the resampled data, while the uncertainty band of model $\mathcal{W}$ can be constructed from the $\alpha/2$-th to $1 - \alpha/2$-th quantile of the empirical distribution of the resampled data.

Experiments

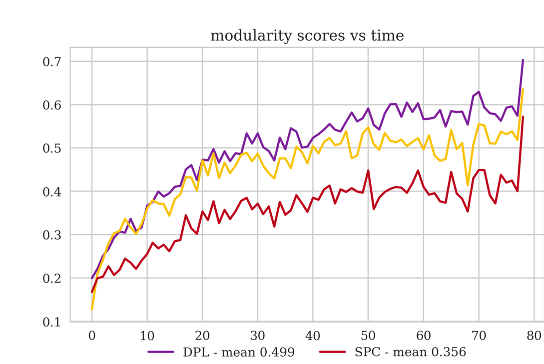
Dataset name	N	K	T	Dataset name	N	K	T	T_0
Synth8000	8,000	30	60	Synth8000Jump/Drift	8,000	2	30	20
MITReal (Eagle & Pentland, 2006)	94	2	37	MITRealFall	94	2	21	14
Math0 (Paranjape et al., 2017)	24,818	15	79	EnronMail (Klimt & Yang, 2004)	184	7	100	68

DPL model assessment



(a) Synth8000 – Local NMI vs. time (b) Synth8000 – Modularity vs. time (c) Synth8000 – 2D histogram of labels

(d) MITReal – Local NMI vs. time (e) MITReal – Modularity vs. time (f) MITReal – 2D histogram of labels



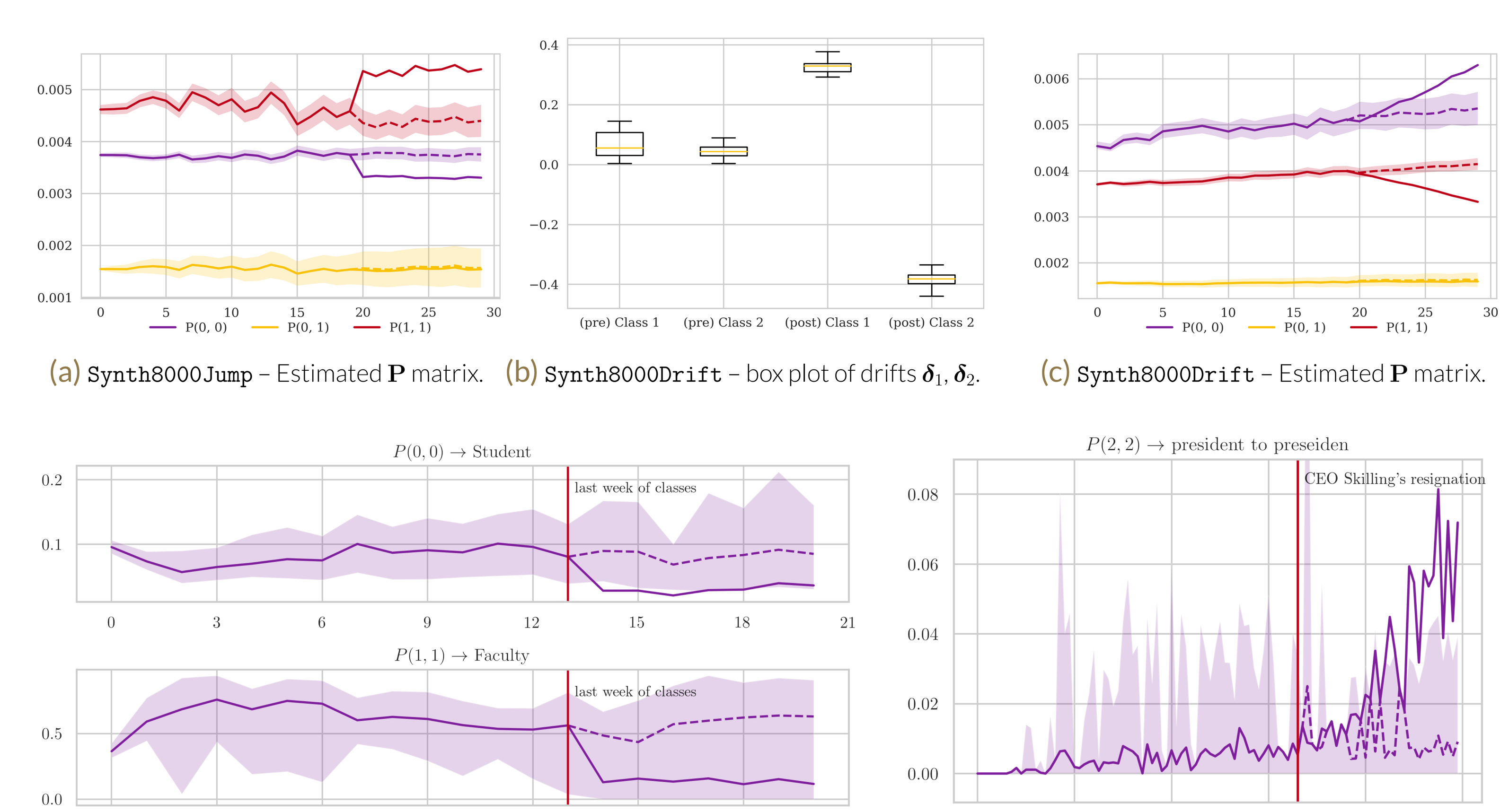
	precision	recall	F1	NMI
DPL	0.134	0.983	**0.236**	**0.744**
sDPL	0.107	0.986	0.194	0.552
SPL [1]	0.040	0.969	0.077	0.013
SPC	0.040	0.967	0.077	0.009

	Global NMI
DPL	**0.399**
SPL	0.327
SPC	0.032
DSBM [3]	0.377

(g) Math0 – Modularity vs. time. (h) Synth8000 – Global metrics. (i) MITReal – Global NMI scores.

Figure 3. Experiment result – DPL model.

Causal impact model



(a) Synth8000Jump – Estimated $\mathbf{P}$ matrix. (b) Synth8000Drift – box plot of drifts $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2$. (c) Synth8000Drift – Estimated $\mathbf{P}$ matrix.

(d) MITRealFall – estimated $\mathbf{P}$ matrix. (e) EnronMail – estimated $\mathbf{P}$ matrix.

Figure 4. Experiment results – causal impact.

References

[1] Arash A Amini, Aiyou Chen, Peter J Bickel, and Elizaveta Levina. Pseudo-likelihood methods for community detection in large sparse networks. *The Annals of Statistics*, 41(4):2097–2122, 2013.

[2] Brian Karrer and Mark E J Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):16107, 2011.

[3] Kevin S Xu and Alfred O Hero. Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):552–562, 2014.