Community Detection on Mixture Multi-layer Networks via Regularized Tensor Decomposition

张煜昭

School of Statistics, RUC

2021年11月09日

Multi-layer Networks

网络中的社群发现问题有着十分悠久的历史，但在绝大部分研究中人们所关注的是在单一网络中的社群发现问题，而现如今在很多情况下我们所收集到的是同一群对象在不同关系或不同时间时所形成的一系列网络结构，即Multi-layer Networks。

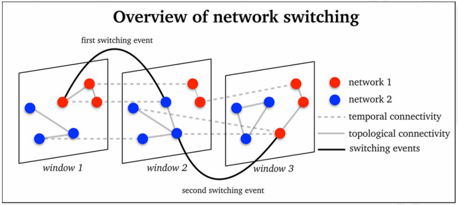


Figure: Multi-layer Network

MLSBM

当我们假设各layer之间独立时，可以很容易的写出在SBM模型下Multi-layer Network的概率分布，因此MLSBM模型在多层网络的社群分组研究中占据主要地位。从最早的Variational Inference方法[2]，到regularized Spectral Clustering方法[4]、Least Square Approach[3]等。研究者们从很多不同的角度对MLSBM模型进行了求解。

$$(\hat{g}, \hat{B}) = \arg\min_{h,\tilde{B}} \sum_{i=1}^{m} \omega_i \sum_{1\leq j\neq l\leq n} \left(Y_{ijl} - \tilde{B}_{ih_j h_l}\right)^2$$

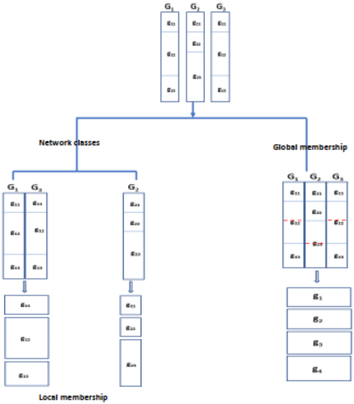
$$\hat{P} = \arg\max_{P^T P=I} \sum_{m=1}^{M} \left\| P^T A^{(m)} P \right\|_F^2, \quad \hat{\Lambda}^{(m)} = \hat{P}^T A^{(m)} \hat{P}, \quad m = \{1, \dots, M\}$$

Mixture Multi-layer SBM(MMSBM)

全部的上述模型均基于一个共同的假设：

- 所有Layer的社群分组情况相同。

在实际生活中上述假设显得没有那么合理，例如：Yelp数据中人们所形成的饮食网络和游玩网络的社群显然不应当是相同的。



Mixture Multi-layer SBM(MMSBM)

MMSBM模型的目标就是在不同的Layer的层面依据社群分组结构进行分类；在Node的层面从Global与Local两个视角给出各个node的社群分组。为了求解MMSBM模型文章设计了一套算法流程命名为TWIST。

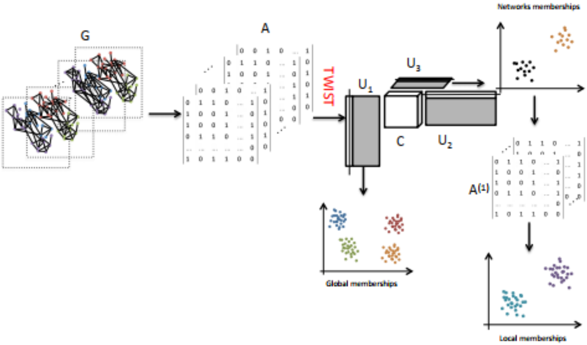


Figure: The general procedure of TWIST

Notations

The observed data contains L-layers of networks on the same set of vertices: $V = [n]$:

$$\mathcal{G} = \{\mathcal{G}_l : l = 1, \cdots, L\}$$

认为每一个$\mathcal{G}_l$是来自m类SBM模型中的一个，其中

- $\text{SBM}(Z_j, B_j) =$ the j-th SBM model's parameters.
- $V_k^j =$ the k-th community in the j-th SBM, and $\cup_{k=1}^{K_j} \mathcal{V}_k^j = \mathcal{V}$.
- $L_j = \# \{l : \ell_l = j, 1 \leq l \leq L\} =$ the number of layers generated by $\text{SBM}\,(Z_j, B_j)$.
- $\overset{\circ}{K} = K_1 + \cdots + K_m$ and $\mathbb{L} = \{\ell_l\}_{l=1}^{L}$ and $\mathbb{V}^j := \left\{\mathcal{V}_k^j\right\}_{k=1}^{K_j}$

The observed matrix of $\mathcal{G}_l$ is A_l follows: $A_l \sim_{iid} \text{Bern}\left(Z_{\ell_l} B_{\ell_l} Z_{\ell_l}^{\top}\right)$

Adjacency tensor and its decomposition

将L个A^l写成一个tensor $\mathbf{A} \in \mathbb{R}^{n\times n\times L}$，我们有：
$\mathbb{E}(\mathbf{A}) = \mathbf{B} \times_1 \bar{Z} \times_2 \bar{Z} \times_3 W$

- $\bar{Z} = (Z_1, Z_2, \cdots, Z_m) \in \{0,1\}^{n\times \dot{K}}$
- $W = (e_{\ell_1}, e_{\ell_2}, \cdots, e_{\ell_L})^\top \in \{0,1\}^{L\times m}$
- $\mathbf{B} \in \mathbb{R}^{\mathring{K}\times\mathring{K}\times m}$ is a 3-way probability tensor whose j-th frontal slide is

$$B(:,:,j) = \operatorname{diag}\left(0_{K_1}, \cdots, 0_{K_{j-1}}, B_j, 0_{K_{j+1}}, \cdots, 0_{K_m}\right), \quad 1 \le j \le m$$

Let $\bar{K}$ denote the number of global communities. 取$\overline{\mathbb{V}} = \{\overline{\mathcal{V}}_k\}_{k=1}^{\bar{K}}$代表global community clusters，对于两个点$i_1 \neq i_2$，

$$\{i_1, i_2\} \in \overline{\mathcal{V}}_k \quad \Longleftrightarrow \quad \bar{Z}(i_1,:) = \bar{Z}(i_2,:)$$

Tucker decomposition

有$r = \operatorname{rank}(\bar{Z})$，有$\bar{Z}$的SVD分解为：

$$\bar{Z} = \bar{U}\bar{D}\bar{R}^{\top}$$

其中$\bar{D} = \operatorname{diag}\left(\sigma_1(\bar{D}), \cdots, \sigma_r(\bar{D})\right) \in \mathbb{R}^{r\times r}, \quad \sigma_1(\bar{D}) \geq \cdots \geq \sigma_r(\bar{D}) > 0$
因此我们可以将tensor A的期望写成Tucker decomposition的形式：

$$\mathbb{E}(\mathbf{A} \mid \mathbb{L}) = \overline{\mathbf{C}} \times_1 \bar{U} \times_2 \bar{U} \times_3 \bar{W}$$

其中$\overline{\mathbf{C}} = \mathbf{B} \times_1 \left(\bar{D}\bar{R}^{\top}\right) \times_2 \left(\bar{D}\bar{R}^{\top}\right) \times_3 D_L^{1/2}$，$\bar{W} = WD_L^{-1/2} \in \mathbb{R}^{L\times m}$，其中$D_L$是W的行和的diagonal matrix。

Tucker decomposition with integrated SVD transformation(TWIST)

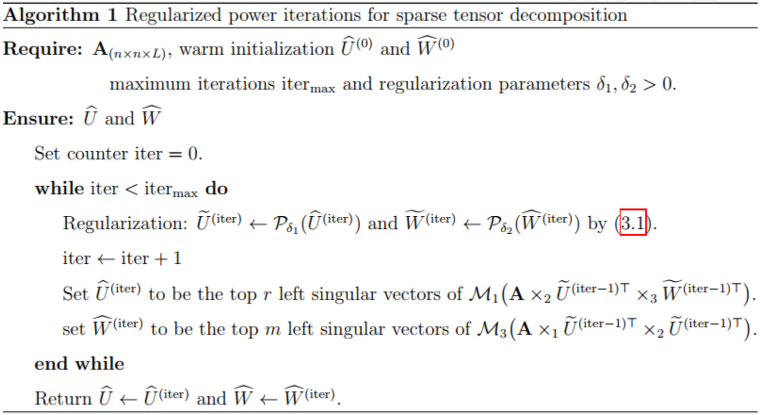
TWIST算法的步骤总结为：

- **Step 1: Decomposition of adjacency tensor**
- **Step 2: Global memberships**

  对上一步得到$\hat{U}$按行的方向做k-means即可得到$\widehat{\mathbb{V}} = \left\{\hat{\mathcal{V}}_k\right\}_{k=1}^{\bar{K}}$。
- **Step 3: Get Network Classes through sup-norm K-means**
- **Step 4:Local memberships** 在同一个label下所有的graph都share同一个SBM模型，因此直接对这些graph的和做k-means即可。

go

Regularized power iterations for sparse tensor decomposition



Algorithm 1 Regularized power iterations for sparse tensor decomposition

```
Require: $\mathbf{A}_{(n\times n\times L)}$, warm initialization $\widehat{U}^{(0)}$ and $\widehat{W}^{(0)}$
         maximum iterations $\text{iter}_{\max}$ and regularization parameters $\delta_1, \delta_2 > 0$.
Ensure: $\widehat{U}$ and $\widehat{W}$
  Set counter $\text{iter} = 0$.
  while $\text{iter} < \text{iter}_{\max}$ do
    Regularization: $\widetilde{U}^{(\text{iter})} \leftarrow \mathcal{P}_{\delta_1}(\widehat{U}^{(\text{iter})})$ and $\widetilde{W}^{(\text{iter})} \leftarrow \mathcal{P}_{\delta_2}(\widehat{W}^{(\text{iter})})$ by (3.1).
    $\text{iter} \leftarrow \text{iter} + 1$
    Set $\widehat{U}^{(\text{iter})}$ to be the top $r$ left singular vectors of $\mathcal{M}_1\big(\mathbf{A} \times_2 \widetilde{U}^{(\text{iter}-1)\top} \times_3 \widetilde{W}^{(\text{iter}-1)\top}\big)$.
    set $\widehat{W}^{(\text{iter})}$ to be the top $m$ left singular vectors of $\mathcal{M}_3\big(\mathbf{A} \times_1 \widetilde{U}^{(\text{iter}-1)\top} \times_2 \widetilde{U}^{(\text{iter}-1)\top}\big)$.
  end while
  Return $\widehat{U} \leftarrow \widehat{U}^{(\text{iter})}$ and $\widehat{W} \leftarrow \widehat{W}^{(\text{iter})}$.
```

Regularized power iterations for sparse tensor decomposition

在实际数据中我们得到的tensor A往往是非常稀疏的（这也是我们关注的情况），而如果直接采用Tucker Decomposition中最常用的high-order orthogonal iterations算法则往往无法得到很好的效果。事实上[5]中只是证明了HOOI在算法在dense tensor上的Statistical Optimality。
为了处理稀疏的情况，文章采用的是[?]中的算法，具体：

$$\mathcal{P}_\delta(U) = \mathrm{SVD}_r\left(U_\star\right) \quad \text{where} \quad U_\star(i,:) := U(i,:)\cdot\frac{\min\{\delta, \|U(i,:)\|\}}{\|U(i,:)\|} \quad i\in[n]$$

其背后的intuition在于，当各个global community的大小是balance的情况下，有

$$\max_j \left\|e_j^\top \bar{U}\right\| = O(\sqrt{r/n})$$

，因此$\hat{U}$的某一行信息量过大就说明其受到的stochastic errors影响很大需要被遏制。在实际操作中建议：

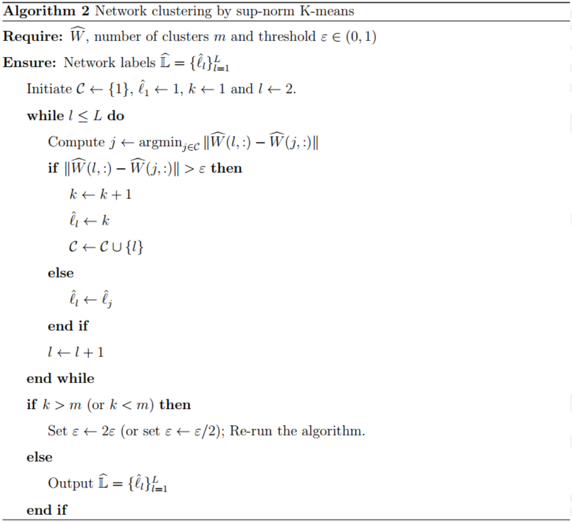
$$\hat\delta_1 = 2\sqrt{r}\cdot\frac{\max_{1\le i\le n}\deg_i}{\sqrt{\sum_i \deg_i^2}} \text{ and } \hat\delta_2 = 2\sqrt{m}\cdot\frac{\max_{1\le l\le L}\mathrm{neg}_l}{\sqrt{\sum_l \mathrm{neg}_l^2}}$$

Warm Start

事实上计算tensor A的low-rank approximation是一个NP-hard的问题，如果采用随机初值，算法往往收敛到一个没用的局部最小点[1]。具体到TWIST中，采用的对$\sum_{l=1}^{L} A_l$做SVD的方式得到$\hat{U}^0$，随后$\hat{W}^0$同样取自$\mathcal{M}_3(\mathbf{A})\left(\widetilde{U}^{(0)} \otimes \widetilde{U}^{(0)}\right)$的前m个left singular vectors，其中$\widetilde{U}^{(0)} = \mathcal{P}_{\delta_1}\left(\widehat{U}^{(0)}\right)$。在后续证明中会说明这种Initialization的合理性。

▸ back

Network Clustering by sup-norm K-means



Algorithm 2 Network clustering by sup-norm K-means

```
Require: $\widehat{W}$, number of clusters $m$ and threshold $\varepsilon \in (0,1)$
Ensure: Network labels $\widehat{\mathbb{L}} = \{\hat{\ell}_l\}_{l=1}^{L}$
   Initiate $\mathcal{C} \leftarrow \{1\}$, $\hat{\ell}_1 \leftarrow 1$, $k \leftarrow 1$ and $l \leftarrow 2$.
   while $l \le L$ do
      Compute $j \leftarrow \mathrm{argmin}_{j \in \mathcal{C}} \|\widehat{W}(l,:) - \widehat{W}(j,:)\|$
      if $\|\widehat{W}(l,:) - \widehat{W}(j,:)\| > \varepsilon$ then
         $k \leftarrow k + 1$
         $\hat{\ell}_l \leftarrow k$
         $\mathcal{C} \leftarrow \mathcal{C} \cup \{l\}$
      else
         $\hat{\ell}_l \leftarrow \hat{\ell}_j$
      end if
      $l \leftarrow l + 1$
   end while
   if $k > m$ (or $k < m$) then
      Set $\varepsilon \leftarrow 2\varepsilon$ (or set $\varepsilon \leftarrow \varepsilon/2$); Re-run the algorithm.
   else
      Output $\widehat{\mathbb{L}} = \{\hat{\ell}_l\}_{l=1}^{L}$
   end if
```

▸ back

Signal Strengths

对与core tensor有：

$$\bar{\mathbf{C}} = \mathbf{B} \times_1 \left(\bar{D}\bar{R}^\top\right) \times_2 \left(\bar{D}\bar{R}^\top\right) \times_3 D_L^{1/2} = \bar{\mathbf{B}} \times_1 \bar{D} \times_2 \bar{D} \times_3 D_L^{1/2}$$

其中$\bar{\mathbf{B}} = \mathbf{B} \times_1 \bar{R}^\top \times {}_2\bar{R}^\top \in \mathbb{R}^{r\times r\times m}$ 为衡量tensor的signal strengths我们定义：$\sigma_{\min}(\mathbf{T}) = \min\left\{\sigma_{r_j}\left(\mathcal{M}_j(\mathbf{T})\right) : j = 1, 2, 3\right\}$

Condition 1

- (A1):$\sigma_{\min}(\bar{\mathbf{B}}) \geq c_1 p_{\max}$, where $p_{\max} = \max_{i,j,l}\left[\mathbb{E}A_l\right]_{ij}$
- (A2):$\bar{D}$ is well-conditioned, $\sigma_{\min}(\bar{\mathbf{B}}) \geq c_1 p_{\max}$, where $\sigma_1(\bar{D}) \leq \kappa_0 \sigma_r(\bar{D})$
- (A3):Network balance:$L_{\min} \asymp L_{\max} \asymp L/m$

Lemma1

If conditions hold and m,r,k_0 are constants then $\sigma_{\min}(\bar{\mathbf{C}}) \geq C\sqrt{L} n p_{\max}$

Error Bound of Regularized Power Iteration

定义两个column space之间的距离：$\mathrm{d}(\hat{V}, V) := \inf_{O\in\mathbb{O}_{r,r}} \|\hat{V} - VO\|$，
基于此定义在算法第t个iteration时

$$\mathrm{Err}(t) = \max\left\{\mathrm{d}\left(\widehat{U}^{(t)}, \bar{U}\right), \mathrm{d}\left(\widehat{W}^{(t)}, \bar{W}\right)\right\} \tag{1}$$

Theorem2: General convergence results of regularized power iterations

在初值满足Err(0) $\leq 1/4$，上述conditions 满足，Lnp$_{\max} \geq C\log^4 n$时以概率趋近于1有： $\mathrm{Err}\left(t_{\max}\right) = O_p\left(\sqrt{\log(n)/\left(\mathrm{Ln}\, p_{\max}\right.}\Big)\right)$ 其中$t_{\max} = C\log\left(\sigma_{\min}(\bar{\mathbf{C}})/\left(\sqrt{np_{\max}} + \delta_1\delta_2\right)\right)$。

Consistency of recovering global memberships

定义global community分类的Hamming error：

$$\mathcal{L}(\widehat{\bar{V}}, \overline{\mathbb{V}}) = \min_{\tau\ :\ \text{a permutation on } [\bar{K}]} \sum_{i=1}^{n} \mathbf{1}\left(i \in \overline{\mathcal{V}}_k, i \notin \widehat{\overline{\mathcal{V}}}_{\tau(k)}\right)$$

Theorem 3: Consistency of global clustering

在所有conditions 满足且$\min_k \left|\overline{\mathcal{V}}_k\right| \asymp n/\bar{K}$ 以概率趋近于1有：

$$n^{-1} \cdot \mathcal{L}(\widehat{\overline{\mathbb{V}}}, \overline{\mathbb{V}}) \leq C_3 \kappa_0^6 \frac{r^2 \log n}{\mathrm{Lnp_{max}}}$$

Network classification

同样定义Hamming error：

$$\mathcal{L}(\hat{\mathbb{L}}, \mathbb{L}) = \min_{\tau\ :\ \text{permutation of } [m]} \sum_{l=1}^{L} \mathbf{1}\left(\ell_l \neq \tau\left(\hat{\ell}_l\right)\right)$$

Theorem 4: Consistency of network clustering

设$\widetilde{\mathbb{L}} = \left\{\tilde{\ell}_l\right\}_{l=1}^{L}$是对$\hat{W}$做标准k-means得到的结果。在所有conditions 满足时以概率趋近于1有：

$$L^{-1} \cdot \mathcal{L}(\widetilde{\mathbb{L}}, \mathbb{L}) \leq C_3 \kappa_0^4 \frac{\left(r^2/m\right) \log n}{L n p_{\max}}$$

Warm initialization for regularized power iteration

在Theorem 2中我们需要初值满足：

$$\mathrm{Err}(0)=\max\left\{\mathrm{d}\left(\widehat{U}^{(0)},\bar{U}\right),\mathrm{d}\left(\widehat{W}^{(0)},\bar{W}\right)\right\}\leq 1/4$$

，可以证明依据Warm Start所给出的步骤，同时保证core tensor的信息量满足：

$$\sigma_r\left(\overline{\mathbf{C}}\times_3(\mathbf{d}_L/L)^{1/2}\right)\geq 4C_3\sqrt{np_{\max}}\log^2 n$$

此时得到的初值满足Warm Start的条件。

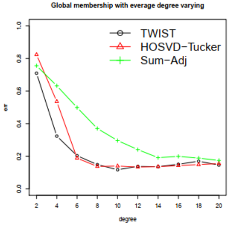
Simulation Settings

将L个layer分为m类，每一类通过等概率随机抽样产生。每一类的社群分组情况也由等概率的随机抽样产生。不同类的B始终相同：$B = pI + q\left(11^T - I\right)$。令$\alpha = q/p$为out-in ration。
模拟主要关心四个参数对结果的影响：Expected degree(network sparsity), "out-in" ratio(signal to noise), number of layers, number of nodes.
文章对三种方法进行了对比：

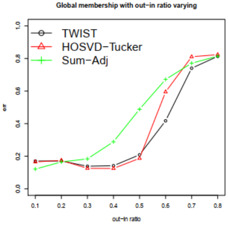
- **TWIST**
- **TWIST-HOSVD** 使用对A的HOSVD分解结果作为初值。
- **Sum-Adj** 直接对所有matrice做加和形成一个single-layer network。

Simulation results - Global memberships

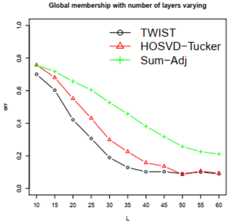
基准设定：L=20，n=600，K=2，m=3，d=10，$\alpha = 0.4$



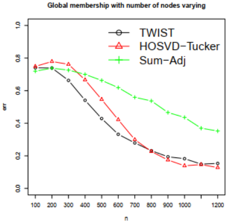
(a) The result of Simulation 1: $n = 600$, $K = 2$, $m = 3$, $L = 20$, $\alpha = 0.4$, varying d.



(b) The result of Simulation 2: $n = 600$, $K = 2$, $m = 3$, $L = 20$, $d = 10$, varying α.

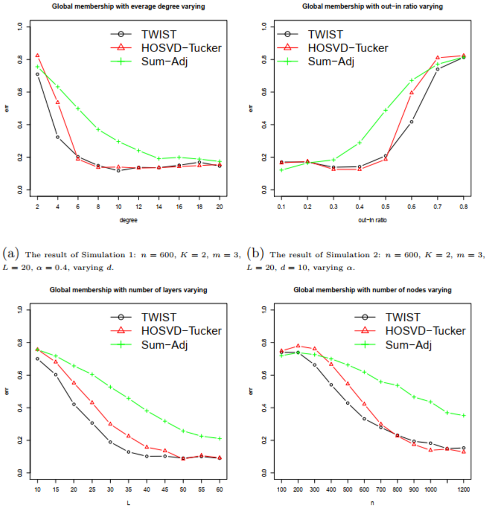


(c) The result of Simulation 3: $n = 600$, $K = 2$, $m = 3$,



(d) The result of Simulation 3: $K = 2$, $m = 3$, $d = 10$,

Simulation results - Network clustering



(a) The result of Simulation 1: $n = 600$, $K = 2$, $m = 3$, $L = 20$, $\alpha = 0.4$, varying d.

(b) The result of Simulation 2: $n = 600$, $K = 2$, $m = 3$, $L = 20$, $d = 10$, varying α.

(c) The result of Simulation 3: $n = 600$, $K = 2$, $m = 3$,

(d) The result of Simulation 3: $K = 2$, $m = 3$, $d = 10$,

Real data- Worldwide food trading networks

数据为：2010年99个国家在30种不同食物上的贸易往来网络。
计算参数：TWIST选择的core tensor为$20 \times 20 \times 2$。

some interesting results

- 网络所分成的两类可以概括为：未加工食品类别与加工食品类别。
- 在未加工食品类别中存在一定数量的经济大国与几乎全部国家都有交易往来；
- 在加工食品类别中地域上更加接近的国家更容易产生交易。

One important restriction

我认为MMSBM作为一个multi-layer SBM模型存在一个重要的缺陷：一般在multi-layer SBM模型当中我们认为$B_l{}_{l=1}^{L}$是随意变化或者服从某种规律在变化的，但MMSBM模型认为同类别的Layer服从一个完全相同的SBM模型显然大大简化了模型的估计难度。另一方面从intuition的角度，正如MMSBM模型文章本身所关注的，不同Layer之间的分类依据应当是社群分组情况是否不同，因此考虑只依据Z的变化进行分类的模型也是有现实意义的。

References

Gerard Ben Arous, Song Mei, Andrea Montanari, and Mihai Nica.
The landscape of the spiked tensor model.
Communications on Pure and Applied Mathematics, 72:2282–2330, 2019.

Qiuyi Han, Kevin S. Xu, , and Edoardo M. Airoldi.
Consistent estimation of dynamic and multi-layer block models.
PMLR, 9, 2015.

Jing Lei, Kehui Chen, and Brian Lynch.
Consistent community detection in multi-layer network data.
Biometrika, 36:5405–5414, 2020.

Subhadeep Paul and Yuguo Chen.
Spectral and matrix factorization methods for consistent community detection in multi-layer networks.
JASA, 2020.

Jingfei Zhang and Jiguo Cao.
Finding common modules in a time-varying network with application