

复杂网络社团结构的动力学方法研究

刘 健

北京大学数学科学学院

导师：鄂维南教授、李铁军教授

2011 年 6 月 3 日

目录

① 研究背景介绍

② 基于最优预测的动力学分区方法

- 基于最优预测的确定性分区方法
- 基于最优预测的概率性分区方法
- 基于有效性指标的确定性分区的自动模型选择

③ 总结和展望

目录

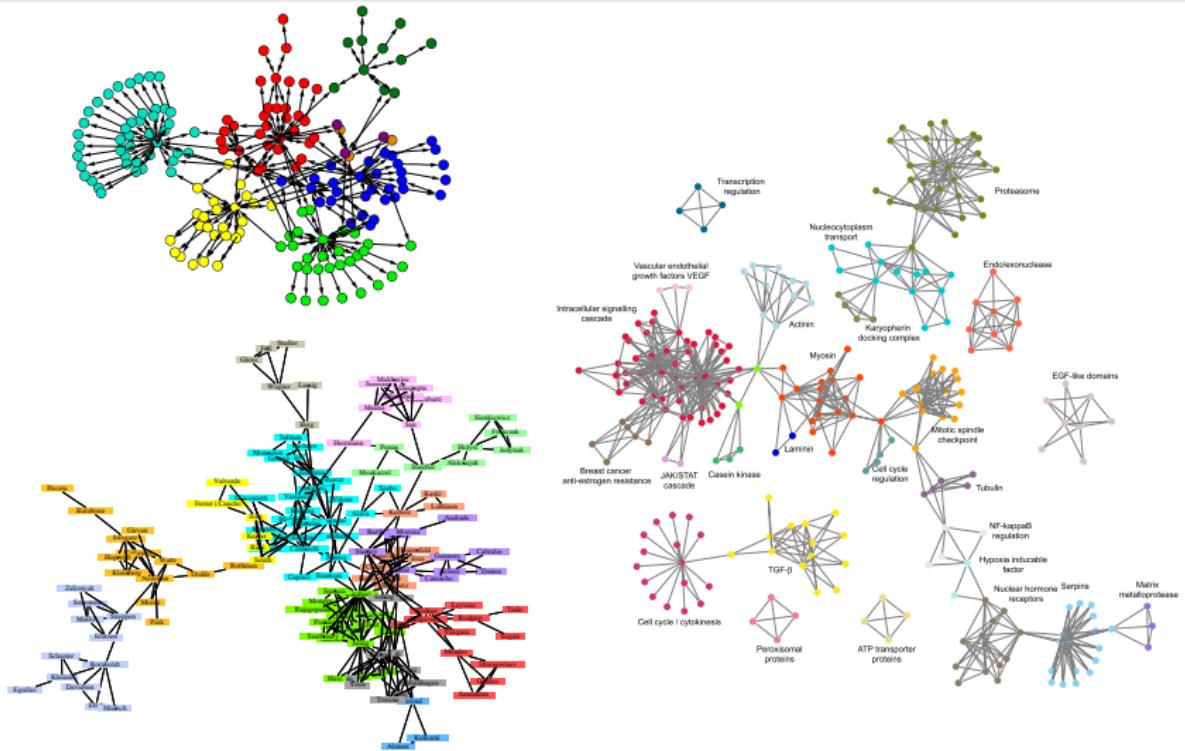
① 研究背景介绍

② 基于最优预测的动力学分区方法

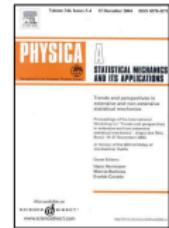
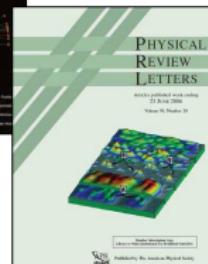
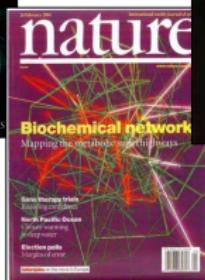
- 基于最优预测的确定性分区方法
- 基于最优预测的概率性分区方法
- 基于有效性指标的确定性分区的自动模型选择

③ 总结和展望

真实世界中的复杂网络



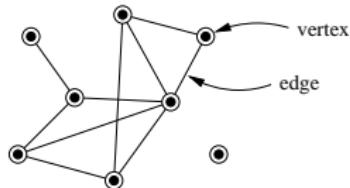
复杂网络方面的相关刊物



复杂网络的基本要素和主要性质

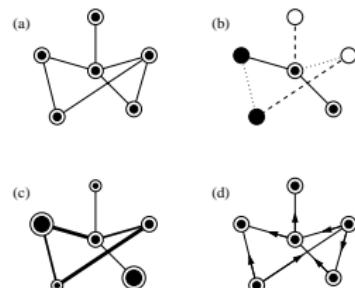
● 网络的基本元素

- 节点 (vertex, 或 node)
- 边 (edge)
- 无权/加权 (unweighted/weighted)
- 无向/有向 (undirected/directed)
- 度 (degree)
- 最短路径 (shortest path, 或 geodesic)
- 直径 (diameter)



● 复杂网络的主要性质:

- 小世界性质 (small-world effect)
- 无标度性质 (scale-free)
- 社团结构性质 (community structure)

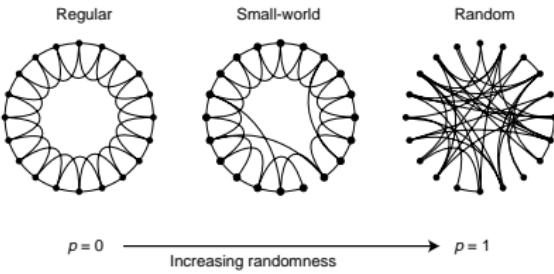


小世界性质和 Watts-Strogatz 小世界网络

- 小世界性质 (六度分离理论)

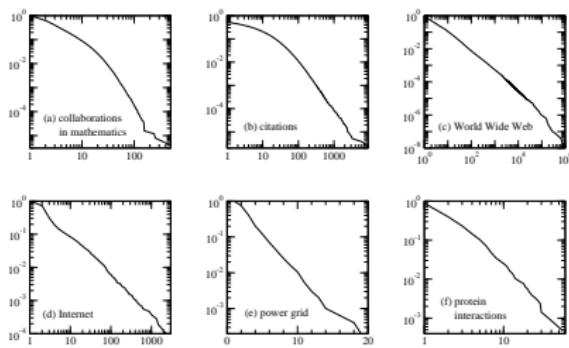


- Watts-Strogatz 小世界网络

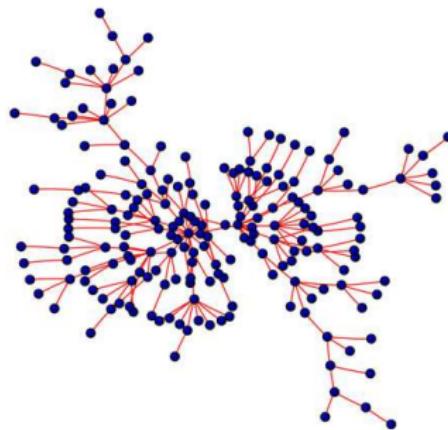


无标度性质和 Barabási-Albert 无标度网络

● 无标度性质 (幂律分布)

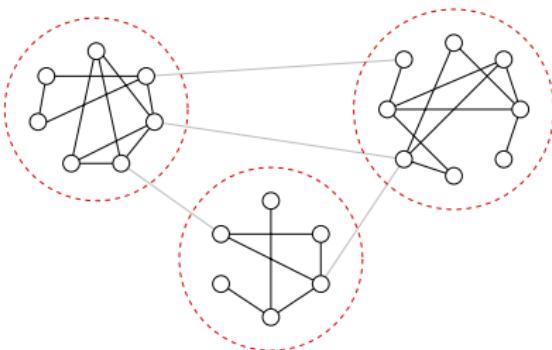


● Barabási-Albert 无标度网络



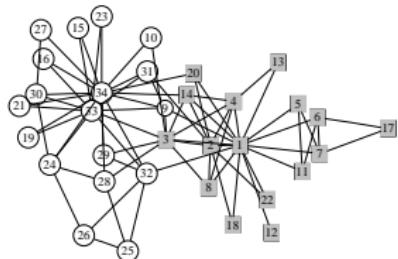
社团结构性质

- 整个网络是由若干个社团构成的，属于相同社团的节点之间具有许多边相连接，而属于不同社团的节点之间具有较少的边相连接。

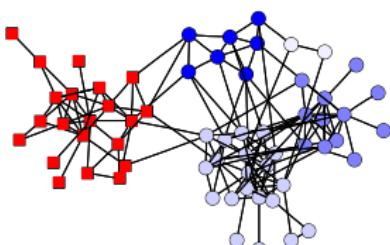


- 社团结构分析在生物学、物理学、计算机科学、社会学中都有广泛的应用。

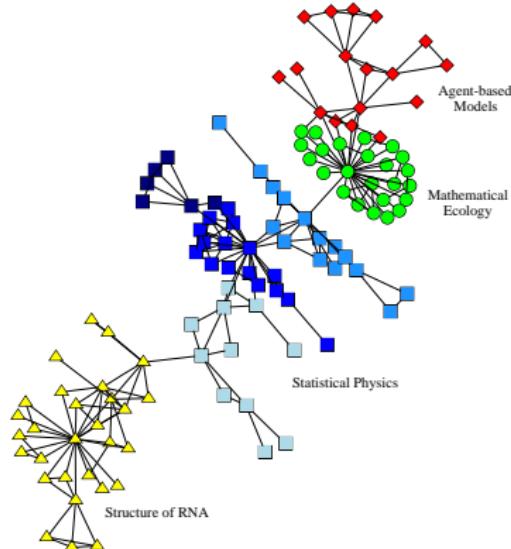
真实世界中的具有社团结构的网络



(a)



(b)



(c)

Newman 提出的模量 (modularity)

- Newman 提出的模量 (modularity) 可作为社团结构优良性的度量, 其值越大表示社团结构越强.

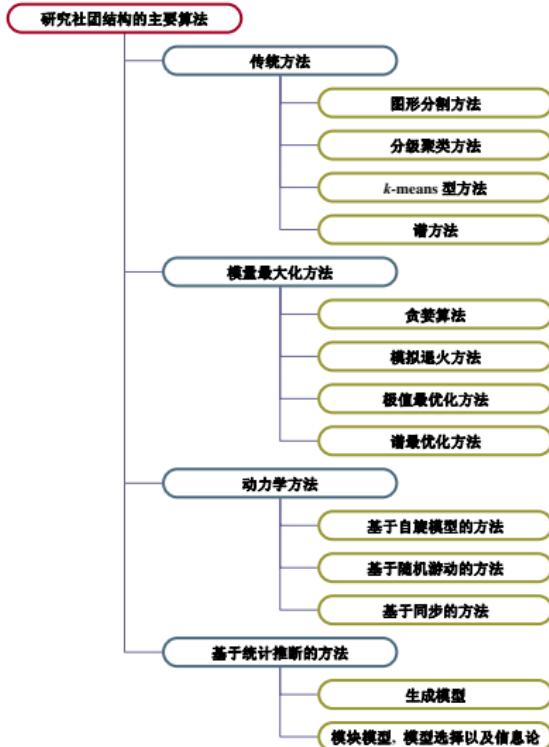
$$Q = \text{社团内部的边数} - \text{这些边数的期望}. \quad (1)$$

- 对于给定的分划 $\{S_k\}_{k=1}^N$, Q 可以写成

$$Q = \frac{1}{2m_e} \sum_{k=1}^N \sum_{x,y \in S_k} \left(e(x,y) - p^E(x,y) \right), \quad p^E(x,y) = \frac{d(x)d(y)}{2m_e}, \quad (2)$$

其中 m_e 是边上权重的和 $\sum_{x,y \in S} e(x,y)/2$.

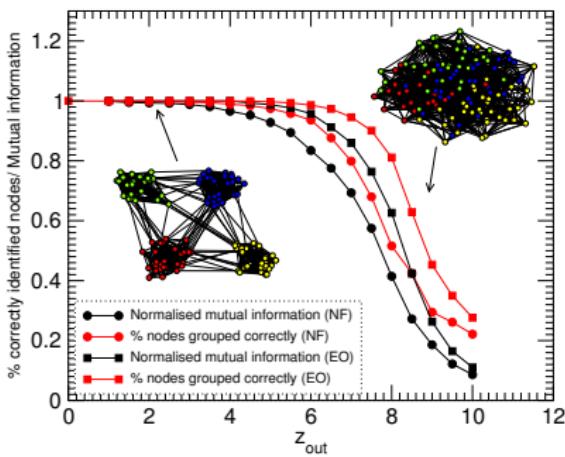
研究社团结构的主要方法



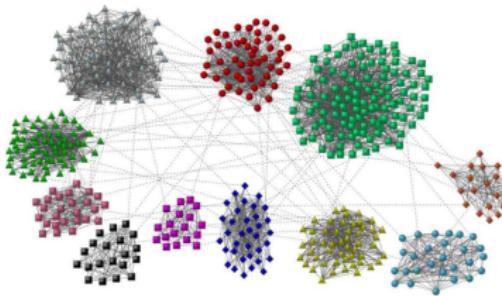
- 在社团结构检测中, 动力学方法是一类强而有力的方法.
- 在动力学方法这个分支中, 通常采用的模型是网络上的随机游动 (random walk).
- 我们研究的也是一类基于随机游动的动力学方法.

测试算法的主要人工网络

- GN 基准网络 (ad hoc 网络)



- LFR 基准网络



目录

1 研究背景介绍

2 基于最优预测的动力学分区方法

- 基于最优预测的确定性分区方法
- 基于最优预测的概率性分区方法
- 基于有效性指标的确定性分区的自动模型选择

3 总结和展望

目录

1 研究背景介绍

2 基于最优预测的动力学分区方法

- 基于最优预测的确定性分区方法
- 基于最优预测的概率性分区方法
- 基于有效性指标的确定性分区的自动模型选择

3 总结和展望

基本思想

- W. E, T. Li and E. Vanden-Eijnden, *Proc. Natl. Acad. Sci. USA*, **105** (2008), 7907–7912.
- 将网络与随机游动 Markov 动力学联系起来, 然后引入马氏链空间中的前向算子的 Hilbert-Schmidt 范数, 并且在这个度量下最优约化马氏链. 最终的极小化问题由类似于 k -means 的迭代算法来求解.

网络与马氏链

- 设 $G = (S, E)$ 为具有 n 个节点和 m 条边的网络, 其中 S 是节点集合, $E = \{e(x, y)\}_{x, y \in S}$ 是权重矩阵, $e(x, y)$ 是连接 x 和 y 的边的权重.
- 这对应于网络上各向同性的随机游动, 可以通过随机矩阵 $p = \{p(x, y)\}_{x, y \in S}$ 将这个网络与一个离散马氏链联系起来

$$p(x, y) = \frac{e(x, y)}{d(x)}, \quad d(x) = \sum_{z \in S} e(x, z). \quad (3)$$

- 该马氏链具有平稳分布

$$\mu(x) = \frac{d(x)}{\sum_{z \in S} d(z)}, \quad (4)$$

并且满足细致平衡条件 $\mu(x)p(x, y) = \mu(y)p(y, x).$

Hilbert-Schmidt 范数

- 若 p 是具有平稳分布 μ 的随机矩阵, 定义它的 μ 范数为

$$\|p\|_{\mu}^2 = \sum_{x,y \in S} \frac{\mu(x)}{\mu(y)} |p(x,y)|^2. \quad (5)$$

- μ 范数即为 $L_{\mu}^2(n)$ 空间中的 Hilbert-Schmidt 范数, 即

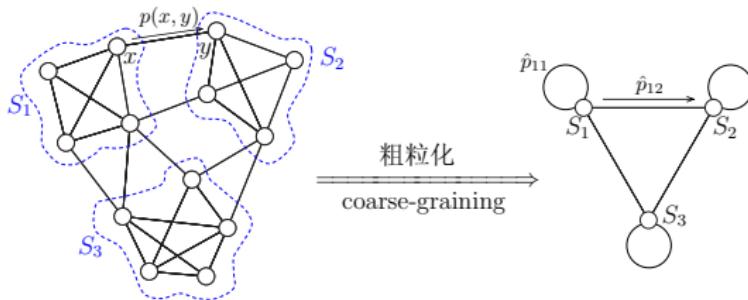
$$\|p\|_{\mu}^2 = \|p\|_{HS}^2. \quad (6)$$

- p 的 μ 范数为 p 的特征值平方和.

粗粒化

- 给定 S 的一个分区 $S = \bigcup_{k=1}^N S_k$, 且 $S_k \cap S_l = \emptyset$ 若 $k \neq l$.
 - 设 \hat{p}_{kl} 是状态空间 $\mathbb{S} = \{S_1, \dots, S_N\}$ 中从 S_k 到 S_l 的粗粒化转移概率, 满足

$$\hat{p}_{kl} \geq 0, \quad \sum_{l=1}^N \hat{p}_{kl} = 1. \quad (7)$$



- \hat{p}_{kl} 可以提升到原始状态空间上的随机矩阵空间

$$\tilde{p}(x, y) = \sum_{k,l=1}^N \mathbf{1}_{S_k}(x) \hat{p}_{kl} \mathbf{1}_{S_l}(y) \frac{\mu(y)}{\hat{\mu}_l}, \quad \hat{\mu}_k = \sum_{z \in S_k} \mu(z). \quad (8)$$

粗粒化

- 如果 \hat{p} 是 \mathbb{S} 上的随机矩阵, 具有平稳分布 $\hat{\mu}$, 则由 (8) 定义的 \tilde{p} 为 S 上的随机矩阵, 具有平稳分布 μ .
- 进一步, 如果 \hat{p} 满足关于 $\hat{\mu}$ 的细致平衡条件, 则 \tilde{p} 满足关于 μ 的细致平衡条件.

最优预测

- 给定社团数目 N , 我们考虑如下的极小化问题

$$\begin{aligned} \min_{\{S_1, \dots, S_N\}, \hat{p}_{kl}} E &= \|p - \tilde{p}\|_{\mu}^2 = \sum_{x,y \in S} \frac{\mu(x)}{\mu(y)} |p(x,y) - \tilde{p}(x,y)|^2 \\ &= \sum_{x,y \in S} \frac{\mu(x)}{\mu(y)} p^2(x,y) - \sum_{k,l=1}^N \frac{\hat{\mu}_k}{\hat{\mu}_l} \hat{p}_{kl}^2 \\ &= \sum_{k,l=1}^N \sum_{x \in S_k, y \in S_l} \mu(x)\mu(y) \left| \frac{p(x,y)}{\mu(y)} - \frac{\hat{p}_{kl}}{\hat{\mu}_l} \right|^2. \quad (9) \end{aligned}$$

最优预测

- 给定分区 $\{S_k\}_{k=1}^N, E$ 的极小值点 \hat{p} 中的 \hat{p} 有如下形式

$$\hat{p}_{kl}^* = \frac{1}{\hat{\mu}_k} \sum_{x \in S_k, y \in S_l} \mu(x)p(x, y). \quad (10)$$

- 上式中的 \hat{p}^* 是随机矩阵, $\hat{\mu}$ 是 \mathbb{S} 上转移矩阵为 \hat{p}^* 的马氏链的平稳分布, 且 \hat{p}^* 满足关于 $\hat{\mu}$ 的细致平衡条件.
- 给定社团数目 N , 则极小化问题 (9) 化为

$$\min_{\{S_1, \dots, S_N\}} E^* = \sum_{k,l=1}^N \sum_{x \in S_k, y \in S_l} \mu(x)\mu(y) \left| \frac{p(x, y)}{\mu(y)} - \frac{\hat{p}_{kl}^*}{\hat{\mu}_l} \right|^2. \quad (11)$$

变形 k -means 算法

- ① 随机初始化分区 $\{S_k^{(0)}\}_{k=1}^N$.
- ② 对于 $n \geq 0$, 计算 $\hat{p}_{kl}^{(n)}$ 如下

$$\hat{p}_{kl}^{(n)} = \frac{1}{\hat{\mu}_k} \sum_{x \in S_k^{(n)}, y \in S_l^{(n)}} \mu(x)p(x, y). \quad (12)$$

- ③ 对于 $n \geq 0$, 更新分区 $\{S_k^{(n+1)}\}$ 如下

$$S_k^{(n+1)} = \left\{ x : k = \arg \min_l \bar{E}(x, S_l^{(n)}) \right\}, \quad k = 1, \dots, N, \quad (13)$$

其中

$$\bar{E}(x, S_k) = \sum_{l=1}^N \sum_{y \in S_l} \mu(x)\mu(y) \left| \frac{p(x, y)}{\mu(y)} - \frac{\hat{p}_{kl}}{\hat{\mu}_l} \right|^2. \quad (14)$$

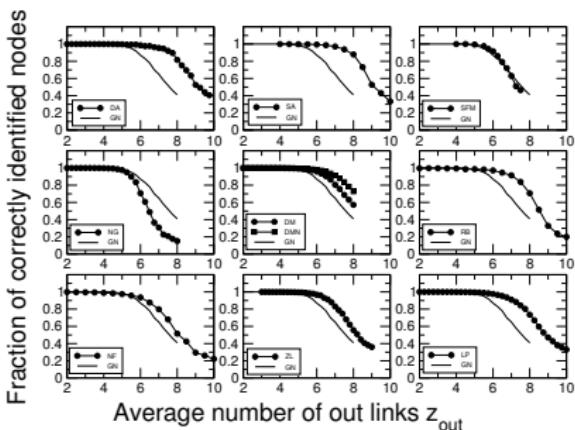
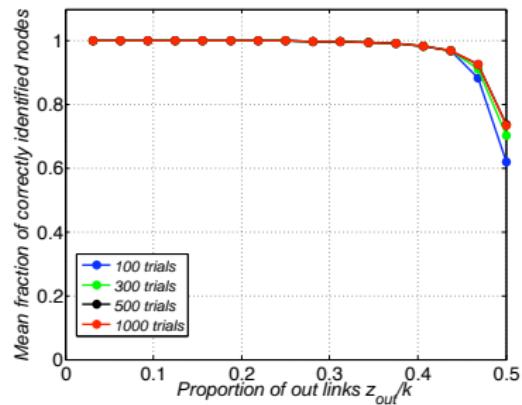
- ④ 重复步骤 (2) 到 (3) 直至目标函数不再下降.

算法的计算量

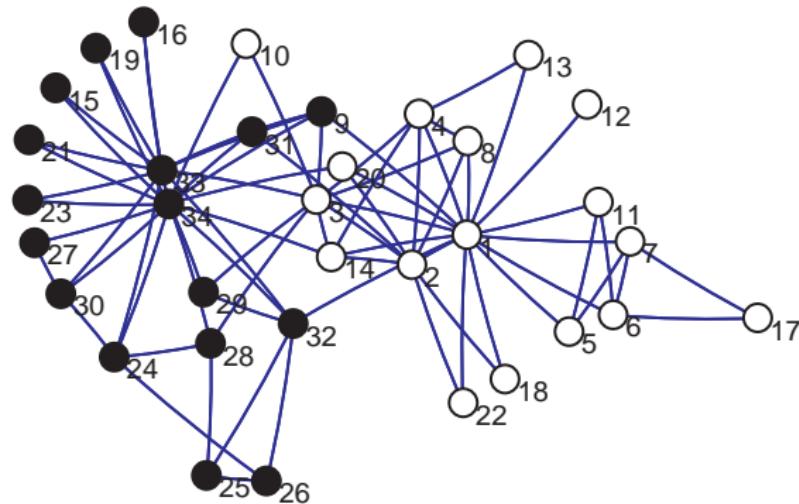
- 如果考虑平均迭代数 k_1 , 试验次数 k_2 , 则算法最终的计算量为 $O(k_1 k_2 N(m + n))$.
- Danon 等人的比较分析中所涉及到的算法的计算量.

作者	算法标记	阶数
Zhou & Lipowsky	ZL	$O(n^3)$
Latapy & Pons	LP	$O(n^3)$
Clauset 等	NF	$O(n \log^2 n)$
Newman & Girvan	NG	$O(nm^2)$
Girvan & Newman	GN	$O(n^2 m)$
Duch & Arenas	DA	$O(n^2 \log n)$
Fortunato 等	FLM	$O(m^3 n)$
Radicchi 等	RCCLP	$O(m^4 / n^2)$
Donetti & Muñoz	DM/DMN	$O(n^3)$
Bagrow & Bollt	BB	$O(n^3)$
Capocci 等	CSCC	$O(n^2)$
Wu & Huberman	WH	$O(n + m)$
Palla 等	PK	$O(\exp(n))$

算法的精度



数值实验: 空手道俱乐部网络



目录

1 研究背景介绍

2 基于最优预测的动力学分区方法

- 基于最优预测的确定性分区方法
- 基于最优预测的概率性分区方法
- 基于有效性指标的确定性分区的自动模型选择

3 总结和展望

基本问题和基本思想

- 在传统确定性分区中, 每个节点在分区后仅属于一个社团. 这在很多情形下有局限性, 比如网络中那些共享一个或几个社团的节点.
- T. Li, J. Liu and W. E, *Phys. Rev. E*, **80** (2009), 026106.
- 将网络与随机游动 Markov 动力学联系起来, 同样引入马氏链空间中的前向算子的 Hilbert-Schmidt 范数, 并在此度量下最优化马氏链. 最终的极小化问题由类似 fuzzy c -means 的迭代算法求解.
- 概率性分区方法得到网络中每个节点以某一概率从属于某一社团, 而不是将节点分配到确定的社团中.

粗粒化

- 在(8)中,用一般性的概率函数 $\rho_k(x)$ 代替示性函数 $\mathbf{1}_{S_k}(x)$, 这里 $\rho_k(x)$ 是节点 x 属于第 k 个社团 S_k 的概率, 满足

$$\rho_k(x) \geq 0, \quad \sum_{k=1}^N \rho_k(x) = 1, \quad x \in S. \quad (15)$$

- \hat{p}_{kl} 可以提升到原始状态空间上的随机矩阵空间

$$\tilde{p}(x, y) = \sum_{k,l=1}^N \rho_k(x) \hat{p}_{kl} \rho_l(y) \frac{\mu(y)}{\hat{\mu}_l}, \quad \hat{\mu}_k = \sum_{z \in S} \rho_k(z) \mu(z). \quad (16)$$

粗粒化

- 如果 \hat{p} 是 \mathbb{S} 上的随机矩阵, 具有平稳分布 $\hat{\mu}$, 则由 (16) 定义的 \tilde{p} 为 S 上的随机矩阵, 具有平稳分布 μ .
- 进一步, 如果 \hat{p} 满足关于 $\hat{\mu}$ 的细致平衡条件, 则 \tilde{p} 满足关于 μ 的细致平衡条件.

最优预测

- 给定社团数目 N , 我们通过考虑如下的极小化问题

$$\begin{aligned} \min_{\rho_k(x), \hat{p}_{kl}} J &= \|p - \tilde{p}\|_\mu^2 = \sum_{x,y \in S} \frac{\mu(x)}{\mu(y)} |p(x,y) - \tilde{p}(x,y)|^2 \\ &= \sum_{x,y \in S} \mu(x)\mu(y) \left(\frac{p(x,y)}{\mu(y)} - \sum_{k,l=1}^N \rho_k(x)\rho_l(y) \frac{\hat{p}_{kl}}{\hat{\mu}_l} \right)^2 \end{aligned} \quad (17)$$

服从于约束条件 (7) 和 (15), 并且从网络动力学的角度最优化地约化马氏链.

Euler-Lagrange 方程组

带有约束 $\sum_{k=1}^N \rho_k(x) = 1$ 的问题 (17) 的 Euler-Lagrange 方程组为

$$\left(I_{\hat{\mu}}^{-1} \cdot \hat{\mu} \right) \cdot \hat{p} \cdot \left(I_{\hat{\mu}}^{-1} \cdot \hat{\mu} \right) = \hat{p}^*, \quad (18a)$$

$$\rho = I_{\hat{\mu}} \hat{p}^{-1} \hat{\mu}^{-1} \rho p^T, \quad (18b)$$

这里

$$\hat{p}_{kl}^* = \frac{1}{\hat{\mu}_k} \sum_{x,y \in S} \mu(x) \rho_k(x) p(x,y) \rho_l(y), \quad (19)$$

$$\hat{\mu}_{kl} = \sum_{z \in S} \mu(z) \rho_k(z) \rho_l(z) = (\rho \cdot I_\mu \cdot \rho^T)_{kl}, \quad (20)$$

$$I_\mu(x, y) = \mu(x) \delta(x, y), \quad x, y \in S, \quad (21a)$$

$$(I_{\hat{\mu}})_{kl} = \hat{\mu}_k \delta_{kl}, \quad k, l = 1, \dots, N, \quad (21b)$$

其中 $\delta(x, y)$ 和 δ_{kl} 均为 Kronecker delta 符号.

带投影算子的交替迭代法—AIP

- ① 设置初始状态 $\rho^{(0)}$ 为变形 k -means 得到的关于网络中每个节点确定性分区的示性矩阵, $n = 0$.
- ② 执行下面的简单迭代, 直到 $\|\rho^{(n+1)} - \rho^{(n)}\| \leq E_{\text{tol}}$:

$$\hat{p}^{(n+1)} = \mathcal{P}[(\hat{\mu}^{-1} \cdot I_{\hat{\mu}} \cdot \hat{p}^* \cdot \hat{\mu}^{-1} \cdot I_{\hat{\mu}})^{(n)}], \quad (22a)$$

$$\rho^{(n+1)} = \mathcal{P}[(I_{\hat{\mu}} \hat{p}^{-1} \hat{\mu}^{-1} \rho p^T)^{(n)}]. \quad (22b)$$

这里 E_{tol} 是给定的精度.

- ③ 最终的 $\rho^{(n)}$ 给出每个节点的概率性分区.

投影算子 \mathcal{P} 的两种选择

设 $\mathbf{u} = (u_1, u_2, \dots, u_n) \in \mathbb{R}^n$, 且 $u_i < 0$ 当 $i \in \Lambda$.

- 选择1: 投影到边界的直接投影.

当 $i \in \Lambda$, 令 $\mathcal{P}u_i = 0$; 否则, 令 $\mathcal{P}u_i = u_i / \sum_{j \notin \Lambda} u_j$.

- 选择2: 反复投影.

首先将 \mathbf{u} 投影到超平面 $\sum_{i=1}^n u_i = 1$, 然后检查投影过的 \mathbf{u} 的每个分量. 如果 $u_{i_0} < 0$, 令 $\mathcal{P}u_{i_0} = 0$, 并向一个约化的超平面 $\sum_{i \neq i_0} u_i = 1$ 再次投影. 对于更低维的超平面重复此投影过程, 直到没有负的分量为止.

梯度下降方法

AIP 算法的构造基于 Euler-Lagrange 方程组, 然而作为非线性问题的迭代算法, 其收敛性不能保证, 这种情形下, 梯度下降的方法提供了一个合理的选择:

- 带投影算子的梯度下降方法 SDP, CGP.
- 指数变换的最速下降法 ETSD.

每次迭代中的计算量

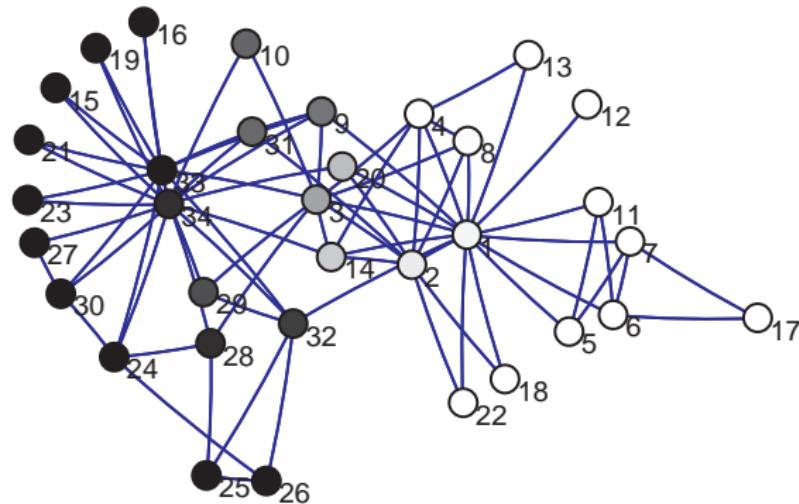
- \hat{p} 迭代步的计算量.

注意到在 \hat{p} 的迭代步中, 所有的矩阵都是 $N \times N$ 阶. 易知计算 $\hat{\mu}$ 花费 $O(Nn)$, 而计算 $\hat{\mu}$ 花费 $O(N^2n)$. 计算 \hat{p}^* 花费 $O(N^2m)$, 其中 m 表示边数, 其在真实网络中通常假定是 $O(n)$. 计算 $\hat{\mu}^{-1}$ 的花费是 $O(N^3)$. 因此, 计算 \hat{p} 的步骤中的总花费是 $O(N^2(m+n))$.

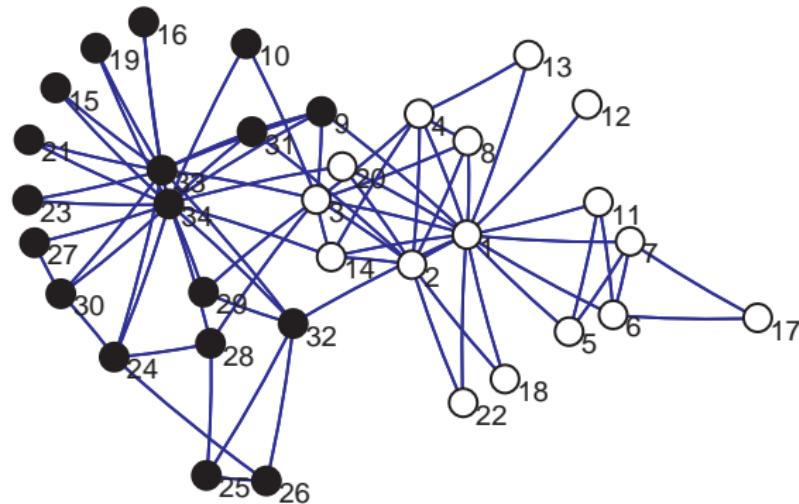
- ρ 迭代步的计算量.

ρp^T 花费 $O(Nm)$, $I_{\hat{\mu}} \hat{p}^{-1} \hat{\mu}^{-1}$ 花费 $O(N^3)$. 所以计算 ρ 的总花费是 $O(N^2n + Nm)$.

数值实验: 空手道俱乐部网络



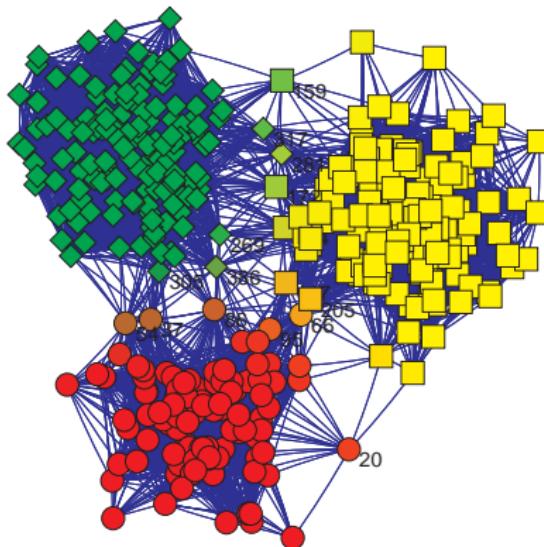
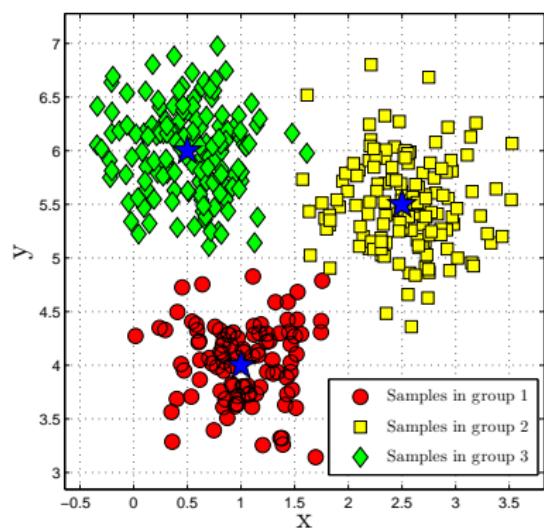
数值实验: 空手道俱乐部网络



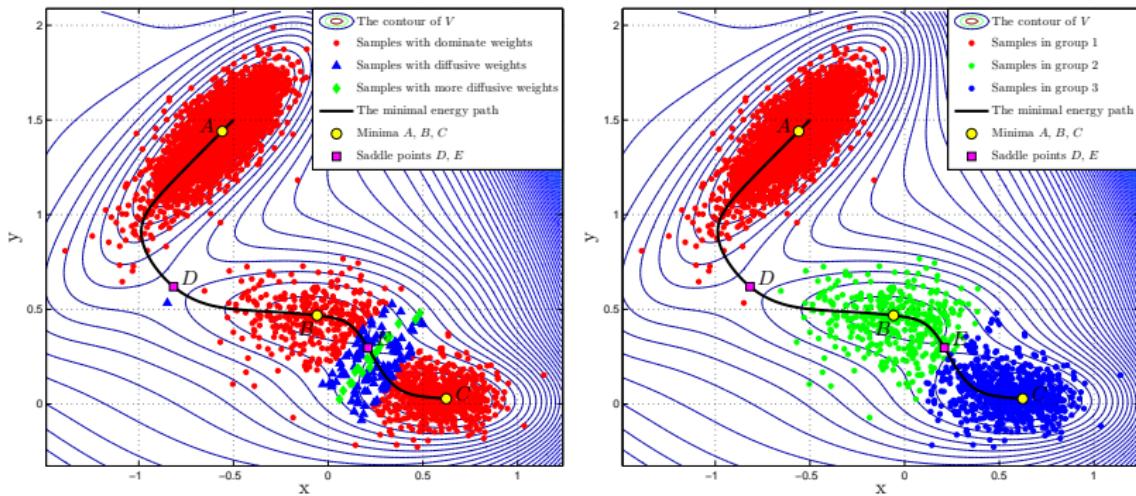
数值实验: 空手道俱乐部网络

Nodes	1	2	3	4	5	6	7	8	9	10	11	12
ρ_K	0.0427	0.0821	0.4314	0.0015	0	0	0	0.0111	0.6619	0.7430	0	0
ρ_W	0.9573	0.9179	0.5686	0.9985	1.0000	1.0000	1.0000	0.9889	0.3381	0.2570	1.0000	1.0000
Nodes	13	14	15	16	17	18	19	20	21	22	23	24
ρ_K	0	0.2262	1.0000	1.0000	0	0	1.0000	0.3012	1.0000	0	1.0000	1.0000
ρ_W	1.0000	0.7738	0	0	1.0000	1.0000	0	0.6988	0	1.0000	0	0
Nodes	25	26	27	28	29	30	31	32	33	34		
ρ_K	1.0000	1.0000	1.0000	0.9496	0.8344	1.0000	0.7210	0.8956	1.0000	0.9475		
ρ_W	0	0	0	0.0504	0.1656	0	0.2790	0.1044	0	0.0525		

数值实验: Gauss 混合模型生成的网络 ($n = 400, N = 3$)



数值实验: Mueller 势生成的样本网络 ($n = 5000, N = 3$)



目录

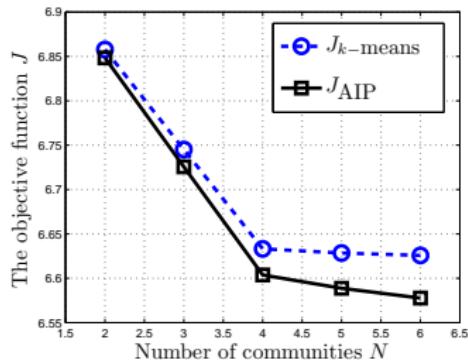
1 研究背景介绍

2 基于最优预测的动力学分区方法

- 基于最优预测的确定性分区方法
- 基于最优预测的概率性分区方法
- 基于有效性指标的确定性分区的自动模型选择

3 总结和展望

基本问题和基本思想



- 人们有时需要确定社团的数目，并遇到了 k -means 型算法的目标函数随社团数目增加而减少的困难。

- J. Liu and T. Li, *Physica A*, **390** (2011), 3579–3591.
- 提出网络分区的有效性指标 (validity index) 作为社团结构的优良程度提供一种度量，并设计了基于有效性指标的方法来实现确定性分区的自动模型选择。
- 算法 SAVI 可有效识别社团结构，并且无需任何关于社团结构的先验信息就可以自动确定社团的个数。

网络分区的有效性指标

- 在确定性分区的框架中,采用同时考虑**社团内部紧密程度 (compactness)**与**社团间分离程度 (separation)**的思想,并构造如下的对于网络分区的有效性指标

$$V_{\text{net}} = J(\hat{p}) \cdot K(\hat{p}). \quad (23)$$

- $J(\hat{p})$ 为(9)中的目标函数,它反映社团内部的紧密程度;因子

$$K(\hat{p}) = \frac{\frac{1}{N(N-1)} \sum_{k \neq l} \hat{p}_{kl}}{\frac{1}{N} \sum_{k=1}^N \hat{p}_{kk}} = \frac{N - \sum_{k=1}^N \hat{p}_{kk}}{\sum_{k=1}^N \hat{p}_{kk}} \cdot \frac{1}{N-1} \quad (24)$$

起到了度量社团之间分离程度的作用.

- 最优分区可通过求解下述而得到

$$\min_N \left\{ \min_{\{S_1, \dots, S_N\}} V_{\text{net}} \right\}. \quad (25)$$

网络分区的有效性指标

- 作为一个模型选择的框架, 在上述的设置中通常会遇到如何调整参数来控制 $J(\hat{p})$ 和 $K(\hat{p})$ 之间竞争力的问题, 定义

$$V_{\text{net}}^\lambda = J(\hat{p}) \cdot K(\hat{p})^\lambda, \quad \lambda \in \mathbb{R}. \quad (26)$$

- 当 $\lambda = 0$ 时, V_{net}^λ 退化成 $J(\hat{p})$. 在下述的所有数值实验中, 我们将仅考虑情形 $\lambda = 1$, 它已然给出了满意的结果.

极小化有效性指标的模拟退火算法—SAVI

- ① 设置参数 $T_{\max}, T_{\min}, \alpha$ 和 R . 在区间 $[N_{\min}, N_{\max}]$ 中随机选取 N , 并随机初始化分区 $\{S_k^{(0)}\}_{k=1}^N$. 设当前温度 $T = T_{\max}$.
- ② 根据 (12) 和 (13), 用变形 k -means 计算相应的 $\hat{p}_{kl}^{(0)}$ 和 $\{S_k^{(0)}\}_{k=1}^N$. 利用定义 (23) 计算能量 E^* .
- ③ 对于 $t = 0, 1, \dots, R$, 做如下迭代:
 - ① 根据下述提议产生新的分区 $\{S_k^{(t)}\}_{k=1}^{N'}$ 作为初始分区, 并设 $N = N'$.
 - ② 根据 (12) 和 (13), 用变形 k -means 更新相应的粗粒化转移概率 $\hat{p}_{kl}^{(t)}$ 和分区 $\{S_k^{(t+1)}\}_{k=1}^N$. 根据 (23) 更新新能量 $E^{(t+1)}$.
 - ③ 根据标准 Metropolis 准则接受新的分区, 即以概率 $\min\{1, \exp(-\frac{1}{T} \Delta E^{(t)})\}$ 接受; 令 $t = t + 1$.
 - ④ 更新最优状态. 若 $E^{(t)} < E^*$, 则 $E^* = E^{(t)}$, 并记录当前分区.
- ④ 降温 $T = \alpha \cdot T$. 如果 $T < T_{\min}$, 执行 (5); 否则重复 (3).
- ⑤ 输出最优分区 $\{S_k\}_{k=1}^N$ 和最小能量 E^* .

步骤 (3.1) 中的提议

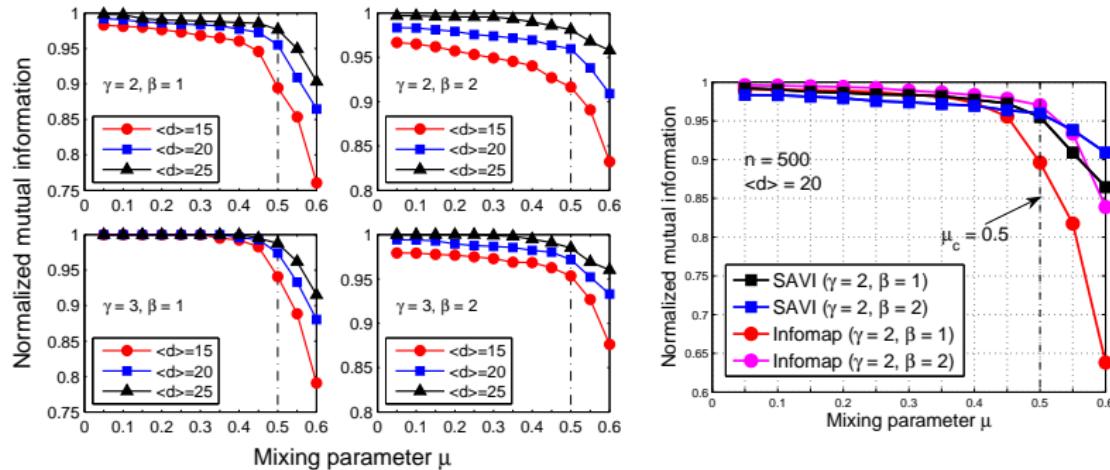
在每次变形 k -means 迭代时, 随机选择下述两个操作中的一个操作, 并且社团强度

$$M_k = \hat{p}_{kk}, \quad k = 1, \dots, N \quad (27)$$

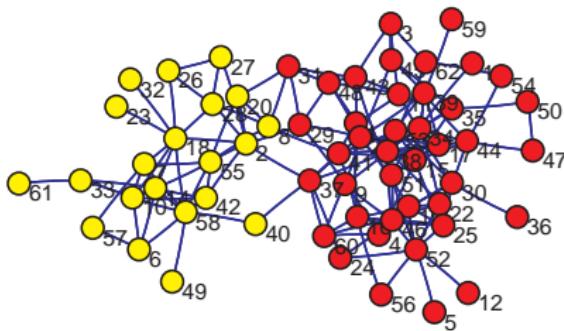
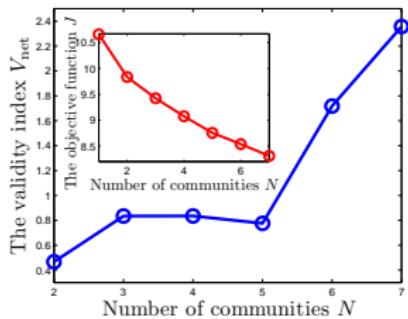
被用来选择一个社团, 它反映第 k 个社团呆在自身而不趋向于转移到其它社团中的概率. 显然, 如果社团强度越大, 则它的社团结构性越强. 这两个操作描述如下:

- **删除一个社团.** 具有最小社团强度 M_d 的社团被确定, 其中 $d = \arg \min_k M_k$. 将其从当前的分区中删除, 并将其中包含的节点合并到社团 S_k , 其中 $k = \arg \max_m \hat{p}_{dm}$.
- **分裂一个社团.** 具有最大社团强度 M_s 的社团被选择, 并将其随机分裂成两个相等大小的新社团. 如果节点数 n_s 是奇数, 两个子社团的大小则分别为 $(n_s + 1)/2$ 和 $(n_s - 1)/2$.

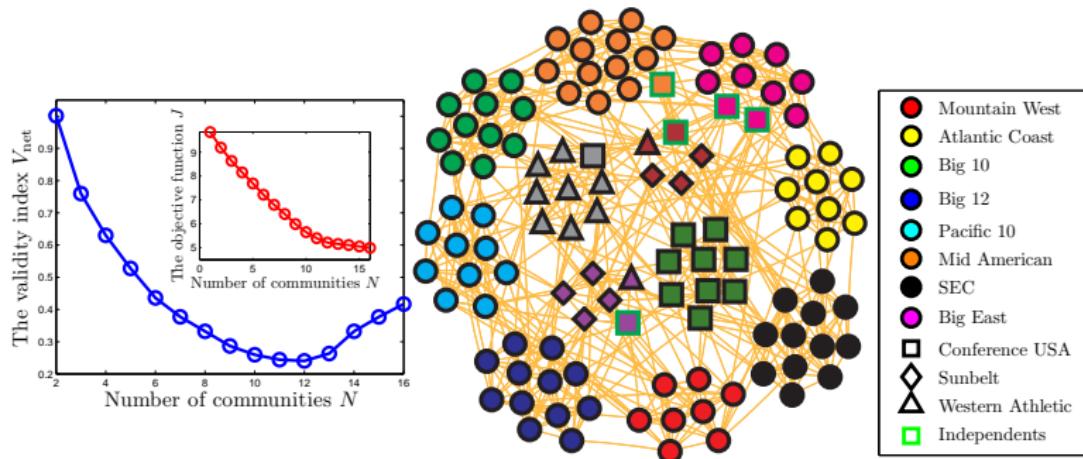
数值实验: LFR 基准网络



数值实验: 宽吻海豚网络



数值实验: 美国足球队网络



目录

1 研究背景介绍

2 基于最优预测的动力学分区方法

- 基于最优预测的确定性分区方法
- 基于最优预测的概率性分区方法
- 基于有效性指标的确定性分区的自动模型选择

3 总结和展望

本文研究的总结

- (a) 提出了复杂网络社团结构的一个概率性框架, 其中每个节点以某一概率从属于某一个社团. 这可以看作是统计中的 *fuzzy c-means* 算法向网络分区问题的自然扩展, 也可以看做是之前的网络最优分区的确定性框架的推广.
- (b) 设计了一个基于有效性指标 (validity index) 的方法来实现确定性分区的自动模型选择. 提出的有效性指标函数可以为社团结构的优良程度提供一种度量. 数值试验表明算法在降温过程中可以有效找出社团结构, 并且无需任何关于社团结构的先验信息就可以自动确定社团的个数. 算法的 matlab 程序可以从网上免费下载使用, 下载链接为:

<http://dsec.pku.edu.cn/~tieli/software/SAVI.zip>

本文研究的总结

- (c) 分别利用结合了两种 k -means 迭代的模拟退火方法来最大化模量 (modularity), 以实现确定性分区的自动模型选择. 这两种 k -means 分别基于相异性指标和扩散距离. 算法可以得到较之许多已有方法更大的模量的值, 从这个意义上来说胜过了许多已有的方法. 算法不仅可以确定社团的个数以及社团结构, 还可以给出每个社团的中心节点. (*Physica A, 389 (2010), 2300–2309*)
- (d) 构造了实现网络概率性分区的自动模型选择的方法. 提出了模糊模量 (fuzzy modularity) 函数, 它可以看作是传统模量的一个推广, 并为网络模糊社团结构的优良性提供了度量. 算法可以有效确定每个节点属于不同社团的概率, 并且初始的模糊分区可以随机选取, 社团的个数也可以自动确定而不再是将其固定为已知的模型参数. (*Eur. Phys. J. B, 77 (2010), 547–557*)

未来研究的展望

- **统计解释:** 本文关于复杂网络社团结构的研究基于动力学方法. 接下来我们希望从中寻找某种统计解释, 并在 Bayes 统计的框架下构造新的算法.
- **有向网络:** 文主要讨论的是无向网络中的社团结构. 下来我们将致力于把这些工作推广到有向网络中, 即推广到不一定满足细致平衡条件的更为一般的动力学中.
- **层次结构:** 接下来我们将研究类似于分级聚类算法这样的能够识别出复杂网络中层次结构的算法.
- **动态网络:** 如何定义动态网络, 即随时间演化的网络中的社团结构, 并将其检测出来将是我们下一步的工作, 这个方向在实际中具有深远意义, 例如 Internet 网络等等.
- **实际应用:** 将现有的算法应用到更大更实际的问题中去, 实现应用价值. 例如生物学网络, 社会学网络, 以及一些经济学系统, 银行系统, 和流行的推荐系统等等.

致谢

- 感谢我的导师鄂维南教授和李铁军教授对我的指导和培养!
- 感谢数学科学学院的领导和其他老师给予我的指导和帮助!
- 感谢论文评阅专家的宝贵意见!
- 感谢答辩委员会的各位专家评委!