软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2018,29(8):2448−2469 [doi: 10.13328/j.cnki.jos.005558]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

稀疏可交换图建模研究综述*

于千城[1,2,3], 於志文[1,2], 王 柱[1,2], 王晓峰[3]

[1](西北工业大学 计算机学院,陕西 西安 710072)

[2](陕西省嵌入式系统技术重点实验室(西北工业大学),陕西 西安 710072)

[3](北方民族大学 计算机学院,宁夏 银川 750021)

通讯作者: 於志文, E-mail: zhiwenyu@nwpu.edu.cn

摘 要: 可交换性假设是采用贝叶斯模型对网络数据建模的重要前提,基于 Aldous-Hoover 表示理论的可交换图不能生成稀疏网络.实证结果表明,真实世界中的很多复杂网络都具有节点度幂律分布的稀疏特征,基于 Kallenberg 表示理论的可交换图能够同时满足可交换性和稀疏性.以 Caron-Fox 模型和 Graphex 模型为例,对稀疏可交换图建模的相关概念、理论和方法的研究发展进行了综述.首先讨论了随机图、贝叶斯非参数混合模型、可交换表示理论、Poisson 点过程、离散非参数先验等理论的研究历程;然后介绍了 Caron-Fox 模型的表示;进而总结了进行稀疏可交换图的随机模拟所涉及的截断采样和边缘化采样方法;接下来综述了稀疏可交换图模型的后验推理技术;最后对稀疏可交换图建模的最新进展和研究前景做了介绍.

关键词: 稀疏可交换图建模;Caron-Fox 模型; Graphex 模型;Kallenberg 表示理论;完全随机测度

中图法分类号: TP311

中文引用格式: 于千城,於志文,王柱,王晓峰.稀疏可交换图建模研究综述.软件学报,2018,29(8):2448−2469. http://www.jos.org.cn/1000-9825/5558.htm

英文引用格式: Yu QC, Yu ZW, Wang Z, Wang XF. Survey of sparse exchangeable graph modeling. Ruan Jian Xue Bao/Journal of Software, 2018,29(8):2448−2469 (in Chinese). http://www.jos.org.cn/1000-9825/5558.htm

Survey of Sparse Exchangeable Graph Modeling

YU Qian-Cheng[1,2,3], YU Zhi-Wen[1,2], WANG Zhu[1,2], WANG Xiao-Feng[3]

[1](School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China)

[2](Shanxi Provincial Key Laboratory for embedded system (Northwestern Polytechnical University), Xi'an 710072, China)

[3](College of Computer, Beifang University of Nationalities, Yinchuan 750021, China)

Abstract: Exchangeability is a key to model network data with Bayesian model. The Aldous-Hoover representation theorem based exchangeable graph model can't generate sparse network, while empirical studies of networks indicate that many real-world complex networks have a power-law degree distribution. Kallenberg representation theorem based exchangeable graph model can admit power-law behavior while retaining desirable exchangeability. This article offers an overview of the emerging literature on concept, theory and methods related to the sparse exchangeable graph model with the Caron-Fox model and the Graphex model as examples. First, developments of random graph models, Bayesian non-parametric mixture models, exchangeability representation theorem, Poisson point process, discrete non-parametric prior etc. are discussed. Next, the Caron-Fox model is introduced. Then, simulation of the sparse

* 基金项目: 国家自然科学基金(61332005, 61725205, 61402369, 61462001, 61762002); 国家重点基础研究发展计划(973)(2015CB352401); "计算机应用技术"宁夏自治区重点学科项目; 北方民族大学校级科研项目(2014XBZ04)

Foundation item: National Natural Science Foundation of China (61332005, 61725205, 61402369, 61462001, 61762002); National Program on Key Basic Research Project of China (973) (2015CB352401); "Computer Application" Ningxia Provincial Key Discipline Project; Research Project of Beifang University of Nationalities (2014XBZ04)

收稿时间: 2017-03-16; 修改时间: 2017-12-11; 采用时间: 2018-01-18; jos 在线出版时间: 2018-02-08

CNKI 网络优先出版: 2018-02-08 15:24:28, http://kns.cnki.net/kcms/detail/11.2560.TP.20180208.1524.017.html

exchangeable graph model and related methods such as truncated sampler, and marginalized sampler are summarized. In addition, techniques of model posterior inference are viewed. Finally, state-of-the-art and the prospects for development of the sparse exchangeable graph model are demonstrated.

Key words: sparse exchangeable graph model; Caron-Fox model; Graphex model; Kallenberg representation theorem; complete random measure

社会学、信息科学与技术、生物学和统计物理学等领域的各种系统中存在大量的关系信息,例如,学术论文间的引用关系,万维网中网页间的超链接关系,生物细胞系统中蛋白质的物理交互关系,航空交通网络中城市间航班往来,社交网络中人和人的交互等.人们常常采用网络对这些复杂系统中的关系信息进行建模,称为复杂网络.随着移动互联网、普适计算[1]、生物计算、社会感知计算[2]、城市计算等应用技术的发展,对包含复杂关系信息的海量数据进行分析和挖掘已经成为很多行业和领域最迫切的需求,为复杂网络分析带来了新的机遇和挑战.

贝叶斯模型作为一种重要的统计建模方法因其具有完备的理论基础和可解释性,已经被广泛用于对网络数据进行建模和预测.可交换性假设是采用贝叶斯模型对网络数据进行建模的重要前提,网络数据的可交换性是指节点的出现顺序不影响统计网络模型的定义.de Finetti[3]提出了随机变量序列的可交换性,称作 de Finetti 表示理论;Aldous[4]和 Hoover[5]提出了随机数组的可交换性,称作 Aldous-Hoover 表示理论;Kallenberg[6]提出了随机测度(采用连续时间中的随机点过程表示网络数据)的联合可交换性和分部可交换性,称作 Kallenberg 表示理论.Aldous-Hoover 表示理论推动了贝叶斯统计模型用于网络数据建模的发展,很多著名的随机图模型如 ER 模型[7]、SBM 模型[8]、MMSB 模型[9]、IRM 模型[10]等都可以纳入这一框架(Aldous-Hoover 可交换图,简称 AHEG[11]),对这一类模型的性质和推理方法的研究也取得了很大的进展.然而,AHEG 模型要么只能生成空图,要么只能生成稠密图,不能产生稀疏图[12],即 AHEG 模型不能同时满足可交换性和稀疏性.

实证结果表明,真实世界中的很多复杂网络都具有节点度幂律分布的稀疏特征,Barabási 提出的 PA (preferential attachment)模型[13]为了生成具有幂律特征的稀疏网络,放弃了可交换性,因而 PA 模型只适用于解释网络整体结构属性(如节点度分布),不能用于进行社区发现和链路预测[14].Bollobás 等人、Wolfe 等人、Borgs 等人提出一个折衷的方向,基于重新调节标度的方法来产生有限可交换的稀疏图序列,然而这类方法生成的可交换稀疏网络却不满足可投影特性,也就不保证随机图随节点数增加产生的图序列具有一致的分布(可投影特性是采用流方式处理网络数据的重要前提),因而基于 Aldous-Hoover 表示理论的统计生成模型具有先天缺陷[11].

Caron 等人[15]利用随机测度和随机图间的联系提出了采用可交换随机测度来对网络数据进行建模,将网络表示成随机点过程而不是邻接矩阵,基于 Kallenberg 表示理论,Caron-Fox 模型既可以生成稠密可交换图也可以生成稀疏可交换图,解决了可交换性和稀疏性二者不可兼得的矛盾.稀疏可交换图建模由此成为学术界的研究热点,伴随大量的相关研究工作产生了很多重要的理论和方法.

本文对稀疏可交换图建模的研究进展进行综述.第 1 节讨论随机图模型、Graphon 模型、Graphex 模型、贝叶斯非参数混合模型、可交换性表示理论、泊松点过程、完全随机测度、带独立增量的归一化随机测度、可交换分区概率函数等理论基础.第 2 节介绍有向多图和带自边的无向图所对应的 Caron-Fox 模型.第 3 节总结进行稀疏可交换图模型的随机模拟所涉及的截断采样和边缘化采样方法的研究发展,重点描述通过条件采样构造 CRM 的 Adaptive Thinning 算法和通过边缘化采样构建稀疏可交换图的 Pólya Ura Scheme 方法.第 4 节对 Caron-Fox 模型后验推理过程采用的 MCMC 近似推理算法进行综述,详细阐述哈密顿 MCMC 算法,并对切片采样算法和指数倾斜稳定分布的采样进行说明.第 5 节对稀疏可交换图建模的最新研究进展和研究前景进行介绍.

1 相关理论

1.1 随机图模型(random graph model)

随机图是利用观测数据理解复杂网络结构的重要工具,其基本原理是:观测到的图结构是由一个潜在数据

生成过程(即随机图模型)产生的,从最大似然性思想出发,在模型空间中寻找使观测数据具有最大出现可能性的模型,其独特之处在于可以用概率的方法来证明所需要的图的存在性,而不必把图真实构造出来.

定义 1.1. 一个随机图模型是一族编了索引的图值(graph-valued)随机变量 $G_{s,\phi}$, s 定义了图的尺寸(取值自一个完全有序集合 S),ϕ是模型参数,决定了图 $G_{s,\phi}$的一些分布特性,$G_{s,\phi}$的分布记为 $g_{s,\phi}$[11].

例 1.1(ER 模型):经典随机图模型由埃尔德什和莱利[7],是 n 个节点上的一族简单随机图 $G_{n,p}$,$n\in\mathbf{N}$,节点间连边的概率为 $p\in(0,1)$,边的产生是相互独立的.可以将 $G_{n,p}$ 直观地表示成一个 $n\times n$ 随机邻接矩阵(一个对称的存放 0,1 值的 $n\times n$ 数组,其对角线全部是 0),统计分析的目标之一就是利用给定观测数据推理参数 p.

为了更准确地对真实网络进行建模,随机图模型的一个重要属性就是节点数和边数之间的关系.随机图模型 $G_{s,\phi}$,给定参数ϕ,记 $s_n\uparrow\infty$表示一个递增趋向于无穷大的尺寸参数序列.给定一个图 G,$v=|v(G)|$表示节点数,$e=|e(G)|$表示边数,假设 $n\to\infty$时 $v\to\infty$.若当 $n\to\infty$时 $\frac{\sqrt{e}}{v}\to 0$ 以概率 1 成立,则称随机图序列$(G_{s,\phi})$是稀疏的,节点数为 v 的稀疏图其边数为 $e=o(v^2)$.

例 1.2(KEG):Kallenberg 可交换图 $G_{\upsilon,\phi}$,尺寸参数 s 不是自然数 n(这是一种常用的随机图族编索引的方法),而是一个非负实数 v, $v\propto\sqrt{e}$, e 是图的边数的期望值;KEG 有 3 个可能的组成部分:隔离边 I、无限星形结构 S、包含图结构主要信息的有限部分 Θ,参数 ϕ是一个三元组 (I,S,Θ), $I\in\mathbf{R}_+$, $S:\mathbf{R}_+\to\mathbf{R}_+$, $\Theta:\mathbf{R}_+^2\to[0,1]$, 称作 graphex[11].通常只讨论没有隔离边和无限星形结构的 KEG,所以 $I=S=0$,不加区分地称 Θ 为 graphex.

统计网络分析的固定模式是:观测到的网络 g_s 被建模成 $G_{s,\phi}$的一个实例(参数 s 已知,ϕ未知),网络分析的目的就是推理出ϕ.某些随机图模型中,序列 $G_{s1,\phi}G_{s2,\phi}$,...是一个网络演化动力学模型,尺寸 s 与采样大小有关,随着采样到的观测数据的增多,图的尺寸也在增大,此时自然就必须考虑不同尺寸下的网络应该在分布上满足一致性(consistency).一致性可以通过要求随机图模型分布是可投影的(projective)来表述.可投影性定义在可投影系统(projective system)中,即一族可测映射$(f_{s,t};s\leqslant t\in S)$,$f_{s,t}$将尺寸为 t 的随机图映射到尺寸为 $s\leqslant t$ 的随机图,$f_{t,t}$是同等映射,$f_{r,t}=f_{r,s}\circ f_{s,t}$,$(r\leqslant s\leqslant t)$[11].

定义 1.2. 若存在可投影系统 $f_{s,t};s\leqslant t\in S$,使得对于任意 $s<t$ 和参数ϕ,都有 $G_{s,\phi}\underline{\underline{d}}f_{s,t}(G_{t,\phi})$ $f_{s,t}(G_{t,\phi})$,则称随机图模型是可投影的,$\underline{\underline{d}}$ 表示在分布上相等[11].

定义 1.3. Graphon 是定义在概率空间上的,在[0,1]上取值的可测对称函数Θ:$[0,1]^2\to[0,1]$.Graphon 是无限随机图序列$(G_{s,\phi})$的极限[11].

例 1.3(AHEG 模型):Aldous-Hoover 可交换图是 n 个节点上的一族简单随机图 $G_{n,\Theta}$,n 仍然表示节点数,但是参数ϕ不再是概率 p,而是一个对称可测随机函数Θ:$[0,1]^2\to[0,1]$,称作 graphon[11].ER 模型可以看做是 AHEG 的一个特例,即$\Theta(x,y)=p$ 是一个常 graphon.

对于无限图,Aldous-Hoover 表示理论断言:一个随机图具有可交换性当且仅当其分布是定义在特定的一族各态历经测度(ergodic measures)上的一个混合分布,每一个各态历经测度就是一个 graphon,由此导致的一个必然结果是 n 个节点间的期望连边数为 $\binom{n}{2}\|\Theta\|_1$ ($\|\Theta\|_1$ 表示Θ的 1 范数),因此 AHEG 生成的要么是空图,要么是稠密图[11].

随机图模型的一般结构:给定一个 Graphon,必定存在一个对应的具有可数无限可交换性的随机图 $G(n,\Theta)$,相应的随机邻接矩阵$(G_{ij})_{ij\in\mathbf{N}}$定义为

$$\begin{cases}\Theta\sim\mu\\ U_i\sim_{iid}Uniform[0,1]\quad \text{for}\, i\in\mathbf{N}\\ G_{ij}\,|\,\Theta,U_i,U_j\sim_{ind}Bernoulli(\Theta(U_i,U_j))\end{cases}\tag{1.1}$$

相应的随机邻接测度 $\{(\theta_i,\theta_j)\}_{\theta_i,\theta_j\in\mathbf{R}_+}$ 定义为

$$\begin{cases} N_\alpha \sim Poisson(c\alpha) \\ \{\theta_i\} \mid N_\alpha \sim_{ind} Uniform[0,\alpha] \\ \{\vartheta_i\} \mid N_\alpha \sim_{ind} Uniform[0,1] \\ (\theta_i,\theta_j) \mid \Theta,\vartheta_i,\vartheta_j \sim_{ind} Bernoulli(\Theta(\vartheta_i,\vartheta_j)) \end{cases} \tag{1.2}$$

$$\Theta(\vartheta_i,\vartheta_j)=\begin{cases} 1-\exp(-2\overline{\rho}^{-1}(\vartheta_i)\overline{\rho}^{-1}(\vartheta_j)), & \text{if } \vartheta_i \neq \vartheta_j \\ 1-\exp(-\overline{\rho}^{-1}(\vartheta_i)^2), & \text{if } \vartheta_i = \vartheta_j \end{cases} \tag{1.3}$$

将一个分布建模成由很多简单分布混合而成,既是一种有用的非参数密度估计方法,又是一种对解释变量间依赖关系的潜在类进行识别的重要方法.可以采用以某个先验分布作为混合比例的层次贝叶斯框架来处理由可数无限个成份组成的混合分布,混合比例有助于找到起决定作用的混合成份.

非参数贝叶斯估计的灵活性能够带来更好的预测性能,原因在于其学习能力不会饱和,从而使得预测性能可以随着观测数据的增多而持续提升[16].最大后验估计方法进行推理的目标只是找到某个特定的使后验最大的参数,进行预测的时候只用一个模型;而 NPB 估计方法进行推理的目标则是学习参数的分布(即考虑所有可能的参数),进行预测的时候把不同的模型都考虑进来,无穷多种模型按相应的重要性权重一起发挥作用.令 D 表示观测样本集,Θ表示模型的所有参数,x^*表示做预测的新样本,$\hat{y}$ 表示 x^*的预测值,MAP 和 NPB 的对比见表 1.

Table 1 Contrast between MAP and NPB

表 1 MAP 和 NPB 的对比

	MAP 估计	NPB 估计
推理的目标	找到使得后验 $\Pr(\Theta\mid D)$最大的一个参数Θ^*	得到后验分布 $\Pr(\Theta\mid D)=\dfrac{\Pr(D\mid\Theta)\Pr(\Theta)}{\int_\Theta \Pr(D,\Theta)}$
预测机制	求 $\Pr(\hat{y}\mid x^*,\Theta^*)$	求 $\Pr(y\mid x^*,D)=\int_\Theta \Pr(\hat{y}\mid x^*,\Theta)\Pr(\Theta\mid D)\mathrm{d}\Theta$
特点	1. 观测数据是由可重复随机采样得到的一个序列 2. 重复采样过程中所使用的参数是不变的 3. 参数是固定的	1. 数据是对某个采样的观测 2. 参数是未知的,采用概率描述 3. 数据是固定的

1.2 可交换性表示理论

模型空间中存在大量可选的随机图模型,然而大部分模型都只能建模出真实网络的一部分特性,换个角度可能就变成了病态模型,因此很难评估这些模型的统计可用性.想要找到既易于处理,又足够灵活(可以较准确地解释真实世界中各种现象)的随机图模型,就必须做出一定假设,可交换性假设是非常重要的一种假设,是指生成模型不依赖于观测样本出现的顺序,或者说变换节点的标签不会改变图模型的概率分布(即节点的标签不提供图结构的任何信息)[11].

可交换性是最常见的一种统计不变性(invariant),也称为概率对称(probability symmetry),具有统计不变性的分布族,其结构可以通过各态历经分解(ergodic decomposition)来解释,或者说采用表示理论来更直观地刻画[11].以下首先介绍最简单的随机序列的可交换性表示理论,然后将可交换性推广到更复杂的随机结构,包括随机数组的可交换性表示理论、随机测度的可交换性表示理论,并且对与可交换随机测度密切相关的可交换随机分区进行阐述.各种表示理论见表 2.

Table 2 Overview of representation theorems

表 2 各种表示理论

	离散随机结构	连续时间上的随机测度
可交换性	de Finetti	Bühlmann
联合(分部)可交换性	Aldous-Hoover	Kallenberg

定义 1.4. 一个随机变量序列,若对序列元素下标进行任意排列不改变序列的联合分布,称序列是可交换的 $(X_1,X_2,\ldots)\overset{d}{=}(X_{\pi(1)},X_{\pi(2)},\ldots)$,对于任意的下标置换$\pi$[11].

定理 1.1(de Finetti 表示理论). 随机序列$(X_i)_{i\in\mathbf{N}}$具有可交换性当且仅当存在X上的一个随机概率测度Θ,满足$X_1,X_2,...|\Theta \overset{\text{iid}}{:} \Theta$,即以$\Theta$为条件,观测变量独立并且服从$\Theta$-分布[11].

从概率角度来看,Θ表示了观测数据的公共结构,而统计推理出的条件概率 $\Pr(X_i|\Theta)$则捕获了每个观测变量独有的随机性.

定义 1.5. 二维随机数组 $X=(X_{ij})_{ij\in\mathbf{N}}$ 称作可交换数组,若满足$(X_{ij})(X_{\pi(i)\sigma(j)})(X_{\pi(i)\sigma(j)})$,对于任意下标置换$\pi$和$\sigma$.当$\pi=\sigma$时,称作联合可交换,否则称为分部可交换[11].

定理 1.2(Aldous-Hoover 表示理论). 二维随机数组$(X_{ij})_{ij\in\mathbf{N}}$具有可交换性当且仅当存在一个随机可测函数$F:[0,1]^3\to\mathbb{X}$,满足$(X_{ij})(F(U_i,U_j,U_{ij}))(F(U_i,U_j,U_{ij}))$,其中,$(U_i)_{i\in\mathbf{N}},(U_{ij})_{ij\in\mathbf{N}}$是取自均匀分布 $Uniform[0,1]$的随机变量构成的序列和矩阵,并且有 $U_{ij}=U_{ji}$[11].

定义 1.6. 空间$\mathbb{X}=\{0,1\}$时,二维随机数组 X 就是一个以 N 为节点集的随机图 G 的邻接矩阵,对于无向图,X是一个对称矩阵,若 X 满足定理 1.2,则称随机图 G 是可交换的.单一图对应联合可交换,二部图对应分部可交换[11].

此时,Aldous-Hoover 表示理论可以进一步简化为:无向图是可交换的,当且仅当存在一个参数对称的随机函数$\Theta:[0,1]^2\to[0,1]$,使定理 1.2 满足 $F(U_i,U_j,U_{ij})=\begin{cases}1, & \text{if } U_{ij}<\Theta(U_i,U_j)\\ 0, & \text{otherwise}\end{cases}$. U_i与节点相关联,U_{ij}与边相关联,将函数F分解成$\Theta:[0,1]^2\to[0,1]$和$H:[0,1]^2\to\mathbb{X}$,使得$(X_{ij}) \overset{d}{=} (F(U_i,U_j,U_{ij}))(H(U_{ij},\Theta(U_i,U_j)))(H(U_{ij},\Theta(U_i,U_j)))$[11].

Caron 等人建立了随机图和对称点过程的对应关系:将随机图中的节点看作任意实数x,将边看作 $\mathbf{R}_+^2$ 上的偶对(x,y),就可以将随机图与 $\mathbf{R}_+^2$ 上的简单点过程对应起来,若节点 i 和 j 有边相连则平面上的位置(θ_i,θ_j)或(θ_j,θ_i)处有一个点,如图 1 所示.

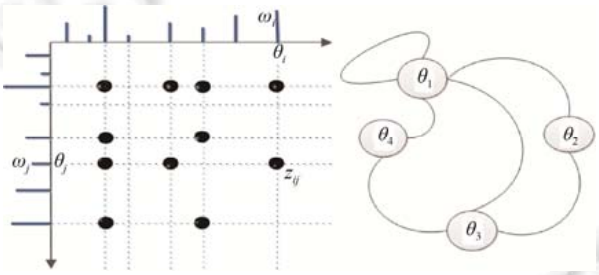


Fig.1 Point process representation of a random graph

图 1 随机图的点过程表示

定义 1.7. 一个简单无向图可以采用实平面 $\mathbf{R}_+^2$ 上的一个对称的点过程 $Z=\sum_{i,j} z_{ij}\delta(\theta_i,\theta_j)$ 来表示,节点 i 按出现的顺序嵌在实数轴的正半轴 $\mathbf{R}_+$的位置θ_i上,并且关联了一个参数ω_i(解释为该节点的社交能力(sociability)),若节点 i 和 j 有边相连则 $z_{ij}=1$,$\delta(\theta_i,\theta_j)$是一个 Dirac 测度,把单位质量集中在平面上$(\theta_i,\theta_j)$处[15].

定义 1.8. $\mathbf{R}_+^2$ 上的点过程 Z 是可交换的,当且仅当对于任意的 $h>0$,在任意下标置换π和σ下,都有:

$$(Z(A_i,A_j)) \overset{d}{=} (Z(A_{\pi(i)},A_{\sigma(j)}))(Z(A_{\pi(i)},A_{\sigma(j)})) \tag{1.4}$$

这里,区间 $A_i=[h(i-1),hi]$,$i=N$,考虑任意小的区间 A_i 以确保节点 i 和 j 不会落入同一个区间[15].

采用一个纯原子的简单随机测度来表示简单点过程更便于进行处理,将点过程的每个点用随机测度中的每个原子来表示,就得到了随机图的邻接测度(adjacent measure)表示,即 KEG(Kallenberg 可交换图).不同于邻接矩阵表示的 AHEG,KEG 在连续空间取节点,因此无限图 G 在某个有限区域截断得到受限图 $G_t=G(.I[0,t]^2)$,其节点个数是个随机量,此外,KEG 中节点至少要与 1 条边关联[11].

定义 1.9. $\mathbf{R}_+^2$ 上的随机测度 W,若对于 $\mathbf{R}_+$上任意的测度保持变换 f 都有 $W\circ(f\otimes f)^{-1}\overset{d}{=}W$,则称 W 是可交换的($\otimes$ 表示张量的内积运算)[11].

定理 1.3(Kallenberg 表示理论). $\mathbf{R}_+^2$ 上的随机测度 ζ 是联合可交换的当且仅当几乎确定满足:

$$\xi=\sum_{i,j}f(\alpha,\vartheta_i,\vartheta_j,\zeta_{\{i,j\}})\delta_{\theta_i,\theta_j}+ \tag{1.5}$$

$$\sum_{j,k}(g(\alpha,\vartheta_j,\varkappa_{jk})\delta_{\theta_j,\sigma_{jk}}+g'(\alpha,\vartheta_j,\varkappa_{jk})\delta_{\sigma_{jk},\theta_j})+ \tag{1.6}$$

$$\sum_{k}(l(\alpha,\eta_k)\delta_{\rho_k,\rho_k}+l'(\alpha,\eta_k)\delta_{\rho_k,\rho_k})+ \tag{1.7}$$

$$\sum_{j}(h(\alpha,\vartheta_j)(\delta_{\theta_j}\otimes\Lambda)+h'(\alpha,\vartheta_j)(\delta_{\theta_j}\otimes\Lambda))+\beta\Lambda_D+\gamma\Lambda^2 \tag{1.8}$$

$f:\mathbf{R}_+^4\to\mathbf{R}_+$ 是可测函数,$\alpha\geqslant 0$, $\tilde{N}_2=\{\{i,j\}\,|\,(i,j)\in N^2$, $\zeta_{(i,j)}$是 $\{i,j\}\in\tilde{N}_2$ 构成的 U-array(独立均匀分布随机变量组成的数组), $\{(\theta_i,\vartheta_j)\}$ 是 $\mathbf{R}_+^2$ 上的独立单位率(unit-rate)泊松过程[11].

因为邻接测度是纯原子的,所有带 Lebesgue 成分的项必须测度为 0,因此式(1.8)取 0,式(1.7)对应了 graphex 的 S 部分,式(1.6)对应了 graphex 的 I 部分,式(1.5)对应了 graphex 的Θ部分.

$$f(\alpha,\vartheta_i,\vartheta_j,\zeta_{\{i,j\}})=\begin{cases}1, & \zeta_{\{i,j\}}\leqslant\Theta(\vartheta_i,\vartheta_j)\\ 0, & \text{otherwise}\end{cases} \tag{1.9}$$

$$\Theta(\vartheta_i,\vartheta_j)=\begin{cases}1-\exp(-2\bar{\rho}^{-1}(\vartheta_i)\bar{\rho}^{-1}(\vartheta_j)), & \text{if } \vartheta_i\neq\vartheta_j\\ 1-\exp(-\bar{\rho}^{-1}(\vartheta_i)^2), & \text{if } \vartheta_i=\vartheta_j\end{cases} \tag{1.10}$$

1.3 泊松点过程(Poisson point process)

随机变量 X 是定义在样本空间Ω上的函数,当 x 是 X 的观测值时,存在Ω中的 ω 使得 $x=X(w)$.随机向量 $(X_1,...X_n)$是定义在样本空间Ω上的 n 元函数,同时要研究更多的随机变量时,就要引入随机过程的概念,设 T 是 $(-\infty,+\infty)$的子集,则称随机变量的集合$\{X_t|t\in T\}$是随机过程,称 T 为该随机过程的指标集(index set),称$\{x_t|x_t=X_t(w),t\in T\}$为$\{X_t|t\in T\}$的一次观测,当 $T=[0,+\infty)$或 $T=(-\infty,+\infty)$时,$\{X_t|t\in T\}$的一次观测是一条曲线,称作随机过程的一条轨迹(trajectory).泊松过程是一种时间连续、状态离散的随机过程[17].

定义 1.10(独立增量过程(dependent incremental process)). 用随机变量 $N(t)$表示时间段$[0,t]$内某类事件发生的个数,则$\{N(t);t\geqslant 0\}$是计数过程(counting process),满足如下条件:① 对 $t\geqslant 0,N(t)$是取非负整数值的随机变量;② 对 $t>s\geqslant 0,N(t)\geqslant N(s)$;③ 对 $t>s\geqslant 0,N(t)-N(s)$是时间段$(s,t]$内的事件发生数;④ $\{N(t)\}$的轨迹是单调不减右连续阶梯函数.若对于任意 n 和$0\leqslant t_1<t_2<...<t_n$都有随机变量$N(0),N(0,t_1],N(t_1,t_2),...,N(t_{n-1},t_n)$相互独立,则这个计数过程称为独立增量过程[17].

定义 1.11(平稳增量过程(stable incremental process)). 若在长度相等的时间段内,事件发生个数的概率分布是相同的,即对于任意 $s>0,t_2>t_1\geqslant 0$ 随机变量 $N(t_1,t_2)$和 $N(t_1+s,t_2+s)$同分布,则称计数过程为平稳增量过程[17].

定义 1.12(泊松过程(Poisson process)). 满足条件:① $N(0)=0$;② $\{N(t)\}$是独立增量过程;③ 对任意 $t,s\geqslant 0$ 都有 $N(s,t+s]$服从参数为λt 的泊松分布,即 $\Pr(N(s,t+s]=k)=\dfrac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t}$, $k=0,1,...$, 称计数过程为泊松过程[17].

常数λ取正值,称为泊松过程的强度(intensity),若λ不随时间变化,则泊松过程是齐次的,若λ随时间变化(或者说与时间有关),则泊松过程是非齐次的[17].

定理 1.4(叠加原理). 设$\Pi_1,\Pi_2,...$是可数个相互独立的泊松点过程,对应的计数过程分别为 $N_1,N_2,...$,强度分别为$\mu_1,\mu_2,...$,则 $\Pi=\bigcup_{i=1}^{\infty}\Pi_i$ 也是一个泊松过程,对应的计数过程为 $N=\sum_{i=1}^{\infty}N_i$, 强度为 $\mu=\sum_{i=1}^{\infty}\mu_i$ [17].

定理 1.5. 令$\Phi_1,\Phi_2,...$是一个率为λ的泊松过程中事件发生的时间(或者位置),给定 $N(t)=n$,随机变量$\Phi_1,\Phi_2,...,\Phi_n$的联合概率密度函数为 $f_{\Phi_1,\Phi_2,...,\Phi_n|N(t)=n}(\Phi_1,\Phi_2,...,\Phi_n)=n!t^{-n}$ [17].

定理 1.5 可以通俗地解释为给定区间内事件的总数量,事件发生的位置是某种方式的均匀分布.

定理 1.6(映射原理(mapping proposition)). $(\mathbb{S},\mathcal{S})$ 是一个可测空间,$\mathbb{S}$ 上的随机点集Π及其计数过程 N 是一个强度为μ的泊松点过程,$(\mathbb{T},\mathcal{T})$ 是另外一个可测空间,$f:\mathbb{S}\to\mathbb{T}$ 是一个可测函数,若集值函数μ与f的反函数复合 $\mu\circ f^{-1}$ 是非原子的(non-atomic),则$f(\Pi)=\{f(x):x\in\Pi\}$也是一个泊松点过程,强度为 $\mu\circ f^{-1}$ [17].

1.4 几类常用的离散非参数先验

任何贝叶斯模型都有一个重要的组成部分,被表示成一个随机可测函数Θ,Θ可以用于构建贝叶斯模型的自然参数,实现了对数据的解耦(或者说建立起数据间的联系)[11].de Finetti 理论刻画了与随机序列相关的模型参数Θ,Aldous-Hoover 表示理论刻画了与随机数组相关的模型参数Θ,Kallenberg 表示理论刻画了与随机测度相关的模型参数Θ[18].

通过为可测函数指定先验 Q,就可以为可交换图模型指定先验.Lijoi 等人 2013 年综述了几类非常适合作为混合模型的非参数先验 Q 的概率模型[19],包括:完全随机测度(completely random measure,简称 CRM)、带独立增量的规一化随机测度(normalized random measures with independent increment,简称 NRMI)、吉布斯型先验(Gibbs-Type prior)等.很多先验都是基于 CRM 对进行适当的变换得到的,CRM 的一个重要特性是其几乎确定的离散性,对 CRM 进行变换得到的随机测度也继承了这一特性,因此它们以概率 1 选择了离散分布.应用这些先验结构的方式主要有两种:① 直接用于对观测数据进行建模,当已知这些数据是由一个离散分布产生的;② 若数据是由连续分布(或者更复杂的混合分布)产生的,先验结构就是层次混合模型的某个的构造块.

1.4.1 CRMs(完全随机测度)

定义 1.13. $(\mathbb{T},\mathcal{T})$ 是一个完全并可分的度量空间(complete and separable metric space),$\mathcal{T}$ 是相应的 Borel σ-代数,W 是 $\mathbb{T}$ 上的随机测度,$A_1,A_2,...$是 $\mathcal{T}$ 中任意可数个互不相交的可测集,若随机变量 $W(A_1),W(A_2),...$相互独立并且满足 $W(\bigcup_i A_j)\ W\left(\bigcup_i A_j\right)=\sum_j W(A_j)$,则称 W 是一个 CRM[20].

Kingman[20]证明了一个 CRM W 可以分解成 3 个独立部分之和:不含原子的非随机的测度 W_d、不含原子的非负随机质量全体 W_f,含固定原子的非负随机质量全体 W_r,并且对于不包含固定原子部分 W_r 和确定部分 W_d 的任意 CRM W,存在一个泊松过程(随机点集$\Pi=(w_i,\theta_i)_{i\in\mathbf{N}}$及其计数过程 $N(\mathrm{d}w,\mathrm{d}\theta)$)使得 $W(\mathrm{d}\theta)=\int_{\mathbf{R}_+} wN(\mathrm{d}w,\mathrm{d}\theta)$ (一个 CRM 可以表示成泊松随机测度的线性函数)[17],W 可以等价地表示为 $W=\sum_{i=1}^{\infty} w_i\delta_{\theta_i}$,随机质量 ω_i 表示某种含义的权重(如,节点的社交能力),Dirac 测度 δ_{θ_i} 将单位质量集中在θ_i 处.

依据 Campbell 定理[20],一个 CRM 由其 Laplace 函数唯一刻画,$W(A)$的 Laplace 变换由以下公式给出:$\mathbb{E}[\exp(-tW(A))]=\exp\left(-\int_{\mathbf{R}_+\times A}[1-\exp(-tw)]\nu(\mathrm{d}\omega,\mathrm{d}\theta)\right),t\geqslant 0$, $\nu(\mathrm{d}\omega,\mathrm{d}\theta)$ 是泊松过程的列维强度(Lévy intensity);拉普拉斯指数(Laplace exponent)定义为 $\psi(t)=\int_0^{\infty}(1-\mathrm{e}^{-wt})\rho(w)\mathrm{d}w$. ν 包含了点过程的所有位置(location)和跳跃(jump)的全部信息,若 $\nu(\mathrm{d}\omega,\mathrm{d}\theta)=\rho(\mathrm{d}\omega)\mu_0(\mathrm{d}\theta)$,随机质量与随机位置相互独立,则这个泊松过程是齐次的;若 $\nu(\mathrm{d}\omega,\mathrm{d}\theta)=\rho(\mathrm{d}\omega|\mathrm{d}\theta)\mu_0(\mathrm{d}\theta)$,随机质量依赖于随机位置,则这个泊松过程是非齐次的(inhomogeneous)[20].

1.4.2 NRMIs(带独立增量的规一化随机测度)

定义 1.14. W 是空间 $(\mathbb{T},\mathcal{T})$ 上的 CRM,$T=W(\mathbb{T})$ 表示总的随机质量.并且几乎确定满足 $0<T<\infty$,则随机概率测度 $\tilde{W}=\frac{W}{T}=\sum_{i=1}^{\infty}\tilde{w}_i\delta_{\theta_i}$ 被定义为带独立增量的规一化随机测度[20].

为确保定义 1.14 中的规一化是良定的,T 必须几乎确定是有限正值,这一点可以由列维测度ρ的性质来保证

$$\int_{\mathbb{R}^+}\rho(\mathrm{d}w)=+\infty,\ \int_{\mathbf{R}^+}(1-\mathrm{e}^{-\omega})\rho(\mathrm{d}w)<+\infty \tag{1.11}$$

满足(1.11)表明 CRM 在任意区间$[0,S]$有无限多个跳(对于任意的 $T<\infty$),称这样的 CRM 为无限活动 CRM(infinite activity CRM)[20].

例 1.4:GP(伽马过程)的列维强度有如下形式$\rho_\alpha(\mathrm{d}\omega)\mu_0(\mathrm{d}\theta)=\alpha\omega^{-1}\mathrm{e}^{-\omega}\mathrm{d}\omega\mu_0(\mathrm{d}\theta)$,$\alpha>0$,记 GP 为 W_α,其总质量为

T_α,列维测度ρ_α满足(1.11),从而有$\tilde{W}_\alpha=\frac{W_\alpha}{T_\alpha}$是一个良定的随机概率测度,我们称这个 NRMI 为一个以α作为集中参数(concentration parameter),以μ_0作为基分布的 DP(狄利赫里过程)[11].

例 1.5:广义伽马过程的列维强度如下形式$\rho_{\alpha,\sigma,\tau}(\mathrm{d}\omega)\mu_0(\mathrm{d}\theta)=\frac{\alpha}{\Gamma(1-\sigma)}\omega^{-\sigma-1}\mathrm{e}^{-\tau\omega}\mathrm{d}\omega\mu_0(\mathrm{d}\theta)$, $\alpha>0,\sigma\in(0,1),\tau\geqslant 0$,记 GPP 为$W_{\alpha,\sigma,\tau}$,其总质量为$T_{\alpha,\sigma,\tau}$,列维测度$\rho_{\alpha,\sigma,\tau}$满足(1.11),从而有$\tilde{W}_{\alpha,\sigma,\tau}=\frac{W_{\alpha,\sigma,\tau}}{T_{\alpha,\sigma,\tau}}$是一个良定的随机概率测度,我们称这个 NRMI 是一个具有参数(α,σ,τ),以μ_0作为基分布的归一化广义伽马过程(NGPP)[15].

1.4.3 EPPF(可交换分区概率函数)

NRMI 的离散分布特性很自然地引发了对生成数据的划分结构的分析,事实上,给定取值于某个度量空间$(\mathbb{T},\mathcal{I})$上的 n 个观测量$X=(X_1,...,X_n)$,数据的离散分布意味着数据间应该是有联系的(ties),因此$\Pr(X_i=X_j)>0,(i\neq j)$,即 X 取了 $k\leqslant n$ 个不同的值,由此,可以用$[n]:=\{1,2,...,n\}$的随机分区π来作为 X 的一个等价的表示方式[10],可交换分区可以帮助我们更有效地构建关于 NRMI 的后验分析方法.

定义 1.15. π是$[n]$的一族随机子集,下标 i 和 j 同属于某个子集 c 当且仅当 $X_i=X_j$,则称π是$[n]$的随机分区(random partition)[11].

如将中任意子集A_i对应的唯一取值记作$X^*_{A_i}$(例如,当且仅当$X_2=X_3=X_6=X_9=X^*_{A_1}$,有$A_1=\{2,3,6,9\}$),在混合模型中,产生随机分区的过程可以看做是将观测变量指派到所隶属的成分,即聚类或者社区划分随机变量 X 采样自可交换序列$(X_1,X_2,...)$,相应的随机分区也是可交换的,其概率密度函数只依赖于分区数量$|\pi|=k$ 和每个子集的大小$n_i=|A_i|$ $(1\leqslant i\leqslant k, n_i\geqslant 1, \sum_{i=1}^k n_i=n)$.

定义 1.16. $\pi=\{A_1,...,A_k\}$是$[n]$的某个随机分区的概率$\Pr(\pi=\{A_1,...,A_k\})=\prod_n^k(n_1,...,n_k)$, $\prod_n^k$ $\Pi_n^k(n_1,...,n_k)$ 是一个对称函数且满足加性规则(addition rule)$\prod_n^k(n_1,...,n_k)=\prod_{n+1}^{k+1}(n_1,...,n_k,1)+\sum_{i=1}^k\prod_{n+1}^{k+1}(n_1,...,n_i+1,...,n_k)$,这样的函数称为 EPPF[11].

2 稀疏可交换图建模

Caron 和 Fox 利用随机测度和随机图间的联系提出了采用可交换随机测度来对网络数据进行建模,将网络表示成随机点过程而不是邻接矩阵,基于 Kallenberg 表示理论,Caron-Fox 模型既可以生成稠密可交换网络也可以生成稀疏可交换图.以下介绍 Caron 和 Fox 提出的有向多图、带自边的无向图所对应的稀疏可交换图模型[15].

2.1 有向多图(directed multigraphs)

可数无限多个节点的集合$V=(\theta_1,\theta_2,...),\theta_i\in\mathbf{R}_+$,其上的有向多图(两个节点$\theta_i,\theta_j$间可以有$n_{ij}\geqslant 1$ 条边,并且允许节点θ_i自己到自己有边,边是有方向的)可以表示成$\mathbf{R}_+$上原子随机测度 D[15].

$$D=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}n_{ij}\,\delta(\theta_i,\theta_j) \tag{2.1}$$

与每个节点θ_i相关联的参数 $\omega_i>0$(sociability)由一个原子随机测度 W 定义,W 的分布是一个列维强度为$\rho(\mathrm{d}\omega)\mu_0(\mathrm{d}\theta)$的齐次 CRM[15].

$$W=\sum_{i=1}^{\infty}w_i\delta_{\theta_i},\ W\sim CRM(\rho,\mu_0) \tag{2.2}$$

给定 W,则 D 由一个强度为乘积测度(product measure)$\tilde{W}=W\times W$ 的泊松过程(Poisson process,简称 PP)生成

$$D\,|\,W\sim PP(W\times W) \tag{2.3}$$

本质上,PP 是随机点集Π及其计数过程 N,式(2.1)可以非正式地解释为:两个节点θ_i,θ_j间有边,则位置(θ_i,θ_j)或(θ_j,θ_i)是Π中的点,每个点的计数n_{ij}由泊松分布 $Poisson(w_iw_j)$得到.

2.2 带自边的无向图(undirected graphs with self edge)

对有向图 $D=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}n_{ij}\,\delta(\theta_i,\theta_j)$ 做一个简单的变换就可以得到定义 1.7 中的无向图 $Z=\sum_{i,j}z_{ij}\delta(\theta_i,\theta_j)$, 可以令 z_{ij}=z_{ji}=1 若有 n_{ij}+n_{ji}>0,否则令 z_{ij}=z_{ji}=0,也就是说,如果有向图两个节点θ_i,θ_j间只要有 1 条有向边,那么对应的无向图中θ_i,θ_j 间就会有 1 条无向边,因此有向多图对应的无向图中允许节点自己和自己连边(一个现实的例子是在社交网络中,一个用户对自己的博文发表评论).无向图的构建过程可用层次结构模型(式 2.4)来表示[15].

$$\begin{cases} W=\sum_{i=1}^{\infty}w_i\delta_{\theta_i} & W\sim CRM(\rho,\mu_0)\\ D=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}n_{ij}\,\delta(\theta_i,\theta_j) & D\,|\,W\sim PP(W\times W)\\ Z=\sum_{i,j}\min(n_{ij}+n_{ji},1)\delta(\theta_i,\theta_j) & \end{cases} \tag{2.4}$$

因为 Pr(z_{ij}=1|w)=Pr(n_{ij}+n_{ji}>0|w)并且 n_{ij} 和 n_{ji} 是相互独立的随机变量(给定 W),由泊松过程的叠加原理(superposition principle),两个强度是 w_iw_j 的泊松随机变量相加,得到强度是 $2w_iw_j$ 的泊松随机变量.又因为 Pr(n_{ij}+n_{ji}>0|w)=1–Pr(n_{ij}+n_{ji}=0|w),所以当给出节点的社交能力 w=(w_i)后,也可以采用式(2.5)直接定义出无向图,可以看到式(2.5)和式(2.4)的构建过程是等价的[15].

$$\Pr(z_{ij}=1\,|\,w)=\begin{cases}1-\exp(-2w_iw_j), & i\neq j\\ 1-\exp(-w_i^2), & i=j\end{cases} \tag{2.5}$$

2.3 GPP(广义伽马过程)

Caron 等人证明了当 W 是一个 GPP,并且当$\sigma\geqslant 0$ 时,可以得到稀疏网络.GGP 具很多有非常好的特性,比如其参数具有非常好的可解释性,GGP 是条件共轭的.Hougaard[21]和 Caron 等人[22]对 GGP 进行了研究.

GGP 的定义在例 2.5 给出,$\sigma\geqslant 0$ 和$\sigma<0$ 的情况下,GGP 具有显著不同的特性,当$\sigma<0$ 时,GGP 是一个有限活动 CRM(finite activity CRM),在区间$[0,\alpha]$上的跳跃的个数 J 依概率 1 有限,并且 J 的分布是一个率为 $-\frac{\alpha}{\sigma}\tau^{\sigma}$ 的泊松分布,而跳跃 w_i 独立同分布于伽马分布 $Gamma(-\sigma,\tau)$.当$\sigma\geqslant 0$ 时,GGP 是一个无限活动 CRM,GGP 在任何区间$[s,t]$上的跳跃的个数 J 是无限的,这里有几个常见的特例:$\sigma=0,\tau>0$ 时对应 GP(伽马过程);$\sigma\in(0,1),\tau=0$ 时对应稳定过程(stable process); $\sigma=\frac{1}{2},\tau>0$ 时对应逆高斯过程(inverse-Gaussian process).GGP 的尾列维强度定义为

$$\bar{\rho}(w)=\int_x^{\infty}\frac{1}{\Gamma(1-\sigma)}\omega^{-\sigma-1}\mathrm{e}^{-\tau\omega}\mathrm{d}\omega=\begin{cases}\dfrac{\tau^{\sigma}\Gamma(-\sigma,\tau x)}{\Gamma(1-\sigma)}, & \tau>0\\ \dfrac{x^{-\sigma}}{\Gamma(1-\sigma)\sigma}, & \tau=0\end{cases},\ \Gamma(\alpha,x)\text{ 是一个不完全伽马函数[15].}$$

3 稀疏可交换图的随机模拟

稀疏可交换图的生成模型是一个非参数贝叶斯模型,非参数贝叶斯有无限多个参数,使得模型的复杂性随着采样尺寸的增大而增大,由于基于随机模拟的方法需要在有限维参数空间进行采样,因而需要通过边缘化或者截断方式对无限维参数空间进行采样.Caron 等人提出了采用截断方式对 GPP 进行模拟,采用边缘化对稀疏可交换图进行模拟[15].

3.1 对 GGP 进行随机模拟

截断方法也称条件采样方法(conditional sampler),通过合适的方法采样 W 的有限足够多个原子,采用一个有限维近似来替代无限维先验.这一类方法最早由 Muliere 等人[23]提出,他们证明了截断近似与无限维先验间的误差(采用总方差范式 total varation norm 表示)可以被选择小于某个特定值.Ishwaran 等人在采用截棍方式构建

NRMI $\tilde{u}$ 方面做了大量工作,他们[24]提出了适用于DP先验模型的截断方法.Ishwaran等人[25]研究了一类更广泛的截棍先验模型的截断采样方法,用总方差范式表示截断误差,并提出了一种简单的块 Gibbs 采样后验推理方法.Papaspiliopoulos等人[26]提出的MH采样通过移动相互改变一对原子来加速混合;Walker[27]提出了DP先验模型的切片采样方法;Kalli 等人[28]提出了 NRMI 先验模型的切片采样方法,通过使用自然无序表示避免了参数的弱可识;Favaro等人[29]提出了σ-稳定Poisson-Kingman先验模型的切片采样方法.在切片采样中,辅助变量被引入到后验分布中,使得 Gibbs 采样的所有完全条件具有有限维分布.重要的是,这一类随机截断方法可以对后验分布进行精确采样,避免了截断误差.然而,尚不清楚怎样将这类方法推广到更广泛的非参数先验模型.Orbanz 等人提出了采用单位率泊松过程构建 NRMI 的逆列维方法(inverse levy method).

3.1.1 通过条件采样构造 CRM 的 Adaptive Thinning 算法

NRMI 作为先验时,对应的 CRM W 通常只有不含原子的非负随机质量全体 W_f,只有当给出观测数据后才会将含固定原子的非负随机质量全体 W_r 引入到后验中.对于 W_f,因为 $\tilde{W}$ 是齐次NRMI,所以随机位置独立同分布采样自μ_0,随机质量与$[S,\infty)$上的具有指数倾斜强度测度$\rho'(\mathrm{d}s)=\mathrm{e}^{-Us}\rho(\mathrm{d}s)$的泊松随机测度同分布,可以采用 Adaptive Thinning 方法进行采样[30],见算法 1.

算法 1. Adaptive Thinning.

输入:强度为 ν' 的泊松随机测度,截断级别 S.

输出:对强度为 ν' 的泊松随机测度在$[S,\infty)$进行的有限采样 N.

算法步骤:

1. 令 $N:=\varnothing,t:=S$
2. 迭代一些操作一直到结束
   a) 从参数是 1 的指数分布采样得到 r
   b) 若 $r>W_t(\infty)$,结束;否则令 $t':=W_t^{-1}(r)$
   c) 以概率 $\nu'(t')/w_t(t')$接受采样,即令 $N:=N\cup(t')$
   d) 令 $t:=t'$,进行下一轮迭代
3. 返回 N(N 就是对强度为 ν' 的泊松随机测度在$[S,\infty)$进行的有限采样)

Thinning 是从一种泊松随机测度进行采样的方法,分为两步:首先从一个提议分布(一个比目标分布强度更高的泊松随机测度)采样出一些点,然后以提议分布和目标分布的强度之比作为概率接受或拒绝每个采样[31].

如图 2 所示,在 Adaptive Thinning 算法中,从提议分布中采样点时,从截断级别 S 出发从左向右迭代进行,令 $\nu'(s)$ 是$\rho'(\mathrm{d}s)$在 Lebesgue 测度下的密度,对于任意的 $t\in\mathbf{R}_+$,存在一个函数 $w_t(s)$满足 $w_t(t)=\nu'(t)$ 和 $w_t(s)\geqslant w_{t'}(s)\geqslant\nu'(s)$ (对于任意的 $s,t'\geqslant t$),对于 NGGP, $\nu'(s)=\dfrac{\alpha}{\Gamma(1-\sigma)}s^{-1-\sigma}\mathrm{e}^{-s(\tau+U)}$, 每次采样到一个点 t 时,引入一族渐进界限 $w_t(s)=\dfrac{\alpha}{\Gamma(1-\sigma)}t^{-1-\sigma}\mathrm{e}^{-s(\tau+U)}$. 令得到的界限 w_t 作为采样下一个点提议分布的强度,随着 t'的增加,界限被收紧,因此拒绝率不断被减小,注意 $w_t(s)$和 $W_t(s)=\int_t^s w_t(s')\mathrm{d}s'$ 的逆函数都可解析得到,通过对 $w_t(s)$其求积分然后求逆得到 $W_t^{-1}(r)=t-\dfrac{1}{\tau+U}\log\left(1-\dfrac{r(\tau+U)\Gamma(1-\sigma)}{at^{-1-\sigma}\mathrm{e}^{-t(\tau+U)}}\right)$, 并且有 $\int_t^\infty w_t(s')\mathrm{d}s'<\infty$, 所以,当有限数量的点被采样后迭代将结束[30].算法的描述如下.

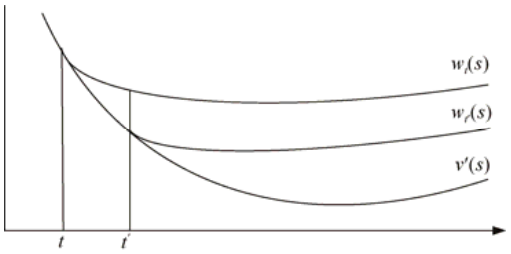


Fig.2 Asymptotic bounds used in sampling from a Poisson random measure[30]

图 2 从一个泊松随机测度进行采样时采用的渐进界限[30]

3.1.2 通过单位率泊松过程构造 CRM 的逆列维方法

随机变量序列$\{\omega_i\}$和$\{\theta_i\}$可以由截棍(stick breaking)过程构造[32],也可以由单位率泊松过程构造[33].逆列维方法利用泊松随机测度的映射理论按从大到小的顺序对随机质量进行采样.对于 $\mathbf{R}_+$上的任意单调函数 $f(x)$,记 f^{-1} 表示$f(x)$的右连续逆函数,因此 $f_{\neg 1}(y)=\begin{cases}Inf\{x\mid f(x)\geqslant y\}, & 若f是个非减函数\\ Inf\{x\mid f(x)\leqslant y\}, & 若f是个非增函数\end{cases}$.

定理 3.1(CRMs 的泊松采样). W 是一个 CRM,列维强度为 ν (dω,dθ),尾列维强度(tail Lévy intensity)定义为 $\bar{\rho}(w)=\int_x^\infty \rho(w)\mathrm{d}w$, 则 $W=W_f+W_r\sum_i J_i\delta(\theta_i)+\sum_k \bar{\rho}^{-1}(\vartheta_k)\delta(\theta_k)\sum_i J_i\delta(\theta_i)+\sum_k \bar{\rho}^{-1}(\vartheta_k)\delta(\theta_k)$, 这里,$\{\theta_1,\theta_2,...\}$是固定的跳跃位置[34].

给定 $\mathbf{R}_+^2$ 上一个单位率泊松过程(θ_i,ϑ_i),构造一个列维强度为$\rho(\mathrm{d}\omega)\mu_0(\mathrm{d}\theta)$的 CRM 的逆列维方法是直接令 $w_i=\bar{\rho}^{-1}(\vartheta_i)$, $\bar{\rho}(x)$ 是尾列维强度, $\bar{\rho}^{-1}$ 是一个单调函数,称逆列维强度(inverse Lévy intensity)[34].

3.2 对稀疏可交换图进行模拟

NRMI 作为先验时,若对应的 CRM 采用截棍过程表示时,网络的生成模型为(2.4),若对应的 CRM 采用 unit-rate poisson 表示时,网络的生成模型为(1.2).以下介绍 Caron 和 Fox 提出的对生成模型为(2.4)的稀疏可交换图所进行的模拟[15].

3.2.1 将图限制在某个区域(graph restriction)

采用邻接测度表示的随机图是一个无限图(因为 $W(\mathbf{R}_+)=\infty$),而实际应用中所研究的网络都是有有限数量的边,但是如何表达或者界定这个有限量是不知道的,所以只能考虑将 D 或 Z 限制在某个区域$[0,\alpha]^2$ 得到相应的有限图 D_α或 Z_α,α作为一个模型参数通过推理得到,相应的有 CRM W_α和 μ_0^α ,记 $Z_\alpha^*=Z_\alpha([0,\alpha])$ 为$[0,\alpha]^2$ 上的总质量,相应的有 $D_\alpha^*=D_\alpha([0,\alpha])$ 和 $W_\alpha^*=W_\alpha([0,\alpha])$, 令 $\bar{W}_\alpha=\frac{W_\alpha}{W_\alpha^*}$, 则有限有向图的模型为

$$\text{For}\,k=1,...,D_\alpha^*\ \text{and}\ j=1,2\ \begin{cases}W_\alpha\sim CRM(\rho,\mu_0^\alpha)\\ D_\alpha^*\mid W_\alpha^*\sim Poisson(W_\alpha^{*2})\ (\text{这里用到了泊松过程的叠加原理,定理 1.4})\\ U_{kj}\mid W_\alpha^*\sim_{ind}\bar{W}_\alpha\ (\text{这里用到了定理 1.5})\\ D_\alpha=\sum_{k=1}^{D_\alpha^*}\delta(U_{k1},U_{k2})\end{cases}\tag{3.1}$$

随机变量 $U_{kj}\in\mathbf{R}_+$对应图中的节点,节点偶对(U_{k1},U_{k2})对应 U_{k1} 到 U_{k2} 的一条边,有向边的数量 D_α^* 依赖于 CRM 的质量 W_α^*,对于每一条有向边,所关联的节点 U_{kj} 采样自 $\bar{W}_\alpha$,因为 $\bar{W}_\alpha$ 是 NRMI(以概率 1 离散分布),所以 U_{kj} 可以取 $N_\alpha\leqslant 2D_\alpha^*$ 个不同的值(N_α对应网络中度至少是 1 的节点的数量).

3.2.2 有限维生成过程(finite dimensional generative process)

由于 W 是一个无限活动,所以不能直接对 $W_\alpha\sim CRM(\rho,\mu_0^\alpha)$ 进行采样,通过引入罐子模型(Pólya ura scheme, 简称 PUS)[35]得到一个有限维过程,方法如下:令 $(U_1',...,U_{2D_\alpha^*}')=(U_{11},U_{12},...,U_{D_\alpha^*1},U_{D_\alpha^*2})$, $t=W_\alpha^*$,对于 GPP,可以积分

得到 $\bar{W}_\alpha$ 并且能够推理出 U'_{n+1} 在给定 $(W^*_\alpha,U'_1,...,U'_n)$ 后的条件分布,因为 $\bar{W}_\alpha$ 以概率 1 离散分布,所以 $U'_1,...,U'_n$ 取 $k<n$ 个不同的值 $\bar{U}'_i$ $(1\leqslant i\leqslant k)$, 相应的重数为 $1\leqslant m_i\leqslant n$ (即有 $U'_1,...,U'_n$ 中有 m_i 个取了同一个值 $\bar{U}'_i$),利用可交换分区表示,可以得到$[n]$的某个随机分区$\pi=\{A_1,...,A_k\}$,其 EPPF 为 $\prod_n^k(m_1,...,m_k\,|\,t)$, 给定 $(t,U'_1,...,U'_n)$ 后 U'_{n+1} 的预测分布可以由 EPPF 得到 $U'_{n+1}\,|\,(t,U'_1,...,U'_n)\sim\dfrac{\prod_{n+1}^{k+1}(m_1,...,m_k,1\,|\,t)}{\prod_n^k(m_1,...,m_k\,|\,t)}\dfrac{1}{\alpha}\mu_0^\alpha+\sum_{i=1}^{k}\dfrac{\prod_{n+1}^{k}(m_1,...,m_i+1,...,m_k,1\,|\,t)}{\prod_n^k(m_1,...,m_k\,|\,t)}\delta_{\bar{U}'_i}$, 采用这种 PUS 表示,我们可以将生成模型(3.1)重写为

$$\text{For}\,k=1,...,D^*_\alpha\,\text{and}\,j=1,2\quad\begin{cases}W_\alpha\sim P_{W^*_\alpha}\\ D^*_\alpha\,|\,W^*_\alpha\sim Poisson(W^{*2}_\alpha)\\ U_{kj}\,|\,W^*_\alpha\sim_{ind}\text{Urn process.}\\ D_\alpha=\sum_{k=1}^{D^*_\alpha}\delta(U_{k1},U_{k2})\end{cases}$$

$P_{W^*_\alpha}$ 是 W^*_α 的分布,在$\sigma\geqslant 0$ 的情况下, $P_{W^*_\alpha}$ 是一个指数倾斜的稳定分布,可以对 $P_{W^*_\alpha}$ 进行精确采样并且能够估计出 EPPF,就可以对图模型进行准确的采样.当 W^*_α 不能被精确采样时,在某些列维强度可以借助于 Adaptive Thinning,其他情况下采用逆列维方法.

边缘化采样(marginal sampler)通过求 W 的边缘分布移除了由于 W 是无限活动带来的麻烦,适用于预测分布体系显式给定的情形.边缘化得到的是先验的PUS表示,这种表示可以用来定义有效的MCMC算法.MacEachern 和 Neal 对在 DPM 模型上基于 PUS 表示的后验推理算法进行了总结,Favaro 等人将这一类算法推广到了 NRMI 作为先验的混合模型,这些方法的不足在于只有某些先验存在合适的 PUS,很多先验的 PUS 难以得到[36].

4 稀疏可交换图模型的推理

稀疏可交换图的生成模型是一个以 NRMI 作为先验的非参数贝叶斯混合模型,其后验分布的计算涉及到很复杂的积分,大多不可能精确计算得到,需要采用近似计算方法[37],比如马尔可夫链蒙特卡洛(Markov Chain Monte Carlo,简称 MCMC)采样方法.常用的 MCMC 方法包括:Gibbs 采样、Metropolis-Hastings 采样(MH)、切片采样、哈密顿蒙特卡罗(Hamiltonian Monte Carlo,简称 HMC)、朗之万动力学方法(Langevin dynamics).

4.1 切片采样算法

无限维先验不允许直接使用基于随机模拟的方法来进行后验分布推理,因此需要对 W 进行截断,仅对有限个原子进行处理.对给出 X 和 U 时 CRM W 的后验进行推理时,切片采样算法引入一个辅助变量 S_i,称为切片变量,其条件分布为 $S_i|X_i,W$:$Uniform(0,W(\{X_i\}))$,$W(\{X_i\}$表示原子 X_i 在 W 中的质量,以 S_i 为条件,X_i 的取值对应到在 W 中的原子的质量肯定是大于等于 S_i 的,因此如果只需要有限个质量大于等于 S_i 的原子就可以更新 X_i 时,S_i 实际上就作为了对 W 的截断级别[36].也就是说,在整个数据集上,采样的状态空间包括 X,切片变量 S,CRM W 和辅助变量 U,采样方法其实就是一个 Gibbs 采样,迭代更新 X,然后更新 U,然后同时更新 W 和 S.

这里只讨论对 W 和 S 的更新,W 的后验既包含 W_f,又包含由观测数据引入的 W_r 部分,切片变量只依赖于 W 中固定原子的质量,另外我们只需要 W 中有限个质量大于全局截断级别 $S=\min_{i\in[n]}S_i$ 的位置随机的原子,因此对 W 和 S 进行充分采样的方法是先采样 W_r,然后采样 S,最后采样 W_f.对于 W_r,定理 4.2 表明每个固定原子对应于 X 中的一个唯一值 $\{X^*_k:k\in\pi\}$,其质量相互独立且与位置无关,条件分布为 $\Pr(J'_k\in\mathrm{d}s\,|\,U,X)\propto s^{|k|}\mathrm{e}^{-Us}\rho(\mathrm{d}s)$, 当 $\tilde{W}$ 是 NGGP 时 $\Pr(J'_k\in\mathrm{d}s\,|\,U,X)\propto s^{|k|-\sigma-1}\mathrm{e}^{-(U+\tau)s}$, 对固定原子的位置的更新可以采样 Bush 和 MacEachern 提出的加速步骤进行.当 W_r 更新后,对 S 的更新可以通过从其分布进行独立采样得到每个 S_i.

4.2 哈密顿蒙特卡罗采样方法

4.2.1 哈密顿动力学方法

MH 算法的一个主要的局限是它具有随机游走的行为,而在状态空间中遍历的距离与步骤数量只是平方根

的关系,仅仅通过增加步长的方式是无法解决这个问题的,因为这会使得拒绝率变高.HMC 采样方法弥补了 MH 算法的缺陷,可以更有效地对状态空间进行搜索.HMC 起源于对哈密顿动力学(Hamiltonian dynamics)下进行变化的物理系统的行为的模拟,通过将概率仿真转化为哈密顿系统的形式,利用哈密顿动力学的框架能够让系统状态发生较大的改变,同时让拒绝概率较低[38].

哈密顿动力学对应于在连续时刻 T 下的状态变量 $z=\{z_i\}$ 的演化.经典的动力学由牛顿第二定律描述(物体的加速度正比于所受的力),是关于时间的二阶微分方程.通过引入中间的动量变量 r 可以将二阶微分方程分解为两个相互耦合的一阶方程,r 对应于状态变量 z 的变化率,元素为 $r_i=\frac{\mathrm{d}z_i}{\mathrm{d}\mathcal{T}}$, 从动力学的角度,$z_i$ 可以被看做位置变量,对于每个位置变量,都存在一个对应的动量变量,位置和动量组成的联合空间被称为相空间.

不失一般性,将概率分布 $\Pr(z)$ 写成下面的形式: $\Pr(z)=\frac{1}{Z_p}\exp(-E(z))$, 其中 $E(z)$ 可以看做状态 z 处的势能,系统的总能量是势能和动能之和 $H(z,r)=E(z)+K(r)$,其中 H 是哈密顿函数,加速度是动量的变化率,通过施加力的方式确定,是势能的负梯度,即 $\frac{\mathrm{d}r_i}{\mathrm{d}\mathcal{T}}=-\frac{\partial E(z)}{\partial z_i}$, 将动能定义为 $K(r)=\frac{1}{2}\|r\|^2=\frac{1}{2}\sum_i r_i^2$, 将系统的动力学用哈密顿方程的形式表示出来,形式为 $\begin{cases}\frac{\mathrm{d}z_i}{\mathrm{d}\mathcal{T}}=\frac{\partial H}{\partial r_i}\\ \frac{\mathrm{d}r_i}{\mathrm{d}\mathcal{T}}=-\frac{\partial H}{\partial z_i}\end{cases}$.

哈密顿动态系统的一个重要性质是在动态系统的变化过程中,哈密顿函数 H 的值是一个常数,这一点通过求微分的方式很容易看出来,即 $\frac{\mathrm{d}H}{\mathrm{d}\mathcal{T}}=\sum_i\left\{\frac{\partial H}{\partial z_i}\frac{\mathrm{d}z_i}{\mathrm{d}\mathcal{T}}+\frac{\partial H}{\partial r_i}\frac{\mathrm{d}r_i}{\mathrm{d}\mathcal{T}}\right\}=\sum_i\left\{\frac{\partial H}{\partial z_i}\frac{\partial H}{\partial r_i}-\frac{\partial H}{\partial r_i}\frac{\partial H}{\partial z_i}\right\}=0$. 哈密顿动态系统的第 2 个重要性质是动态系统在相空间中体积不变,这被称为 Liouville 定理,换句话说,如果我们考虑变量 (z,r) 空间中的某个区域,那么当这个区域在哈密顿动态过程下的变化时,它的形状可能会改变,但是它的体积不会改变.

考虑相空间上的联合概率分布,其总能量是哈密顿函数,概率分布的形式为 $\Pr(z,r)=\frac{1}{Z_H}\exp(-H(z,r))$, 利用体系的不变性和 H 的守恒性,可以看到哈密顿动态系统会使得 $\Pr(z,r)$ 保持不变.虽然 H 是不变的,但是 z 和 r 会发生变化因此通过在某个有限的时间间隔上对哈密顿动态系统积分,就可以让 z 以某种系统化的方式发生较大的变化,避免了随机游走的行为.

4.2.2 混合蒙特卡罗算法

HMC 也称混合蒙特卡罗(hybrid Monte Carlo),是一种对连续分布进行近似的 MCMC 方法:给定独立观测变量集 D,后验分布 $\Pr(z|D)\propto\exp(-U(z))$,势能函数 $U=-\sum_{x\in D}\log\Pr(x|z)-\log\Pr(z)$, 引入一组辅助动量变量 r 对参数空间进行扩展(r 与 z 具有相同的维度),先对联合分布 $\pi(z,r)\propto\exp\left(-U(z)-\frac{1}{2}r^TM^{-1}r\right)$ 进行采样,然后丢弃 r 的样本,就得了后验分布的样本,M 是一个质量矩阵(常被设置成单位矩阵 I),与 r 一起定义了动能.哈密尔顿函数 $H(z,r)=\frac{1}{2}r^TM^{-1}r+U(z)$, 系统的动力学方程式为 $\begin{cases}\frac{\mathrm{d}z_i}{\mathrm{d}\mathcal{T}}=\frac{\partial H}{\partial r_i}=M^{-1}r\mathrm{d}\mathcal{T}\\ \frac{\mathrm{d}r_i}{\mathrm{d}\mathcal{T}}=-\frac{\partial H}{\partial z_i}=-\nabla U(z)\mathrm{d}t\end{cases}$, 不能直接对连续系统进行采样,需要先将其离散化.常用的离散化方法是蛙跳积分(leapfrog integrator),利用下列公式对位置变量和动量变量的离散时间近似 $\hat{z}$ 和 $\hat{r}$ 进行交替的更新.

$$\hat{r}_i\left(\mathcal{T}+\frac{\varepsilon}{2}\right)=\hat{r}_i(\mathcal{T})-\frac{\varepsilon}{2}\frac{\partial E}{\partial z_i}(\hat{z}(r)),$$

$$\hat{z}_i(\mathcal{T}+\varepsilon)=\hat{z}_i(\mathcal{T})+\varepsilon\hat{r}_i\left(\mathcal{T}+\frac{\varepsilon}{2}\right),$$

$$\hat{r}_i(\mathcal{T}+\varepsilon)=\hat{r}_i\left(\mathcal{T}+\frac{\varepsilon}{2}\right)-\frac{\varepsilon}{2}\frac{\partial E}{\partial z_i}(\hat{z}(\mathcal{T}+\varepsilon)).$$

蛙跳积分对动量变量的更新形式是半步更新,步长为 $\varepsilon/2$,接着是对位置变量的整步更新,步长为 ε,然后是对动量变量的第 2 个半步更新.连续地使用几次蛙跳,对动量变量的半步更新就可以结合到步长为 ε 的整步更新中.于是,位置变量的更新和动量变量的更新互相之间以蛙跳的形式结合.为了将动态系统归进一个时间间隔 $\mathcal{T}$,我们需要进 $\mathcal{T}/\varepsilon$ 个步骤.因为蛙跳积分中的每一步对 z_i 或者 r_i 的更新都只是另一个变量的函数,即将相空间的一个区域进行形变而不改变它的体积,所以蛙跳积分精确地保持了相空间的体积不变性.

对于一个非零的步长 ε,蛙跳算法的离散化会在哈密顿动力学方程的积分过程中引入误差,为了消除与离散化过程关联的数值误差,HMC 将 Hamiltonian 动态系统框架与 Metropolis 算法结合在一起,构造了一个马尔可夫链,它由对动量 r 的随机更新和使用蛙跳算法对哈密顿动力系统的更新交替组成.每次应用蛙跳算法之后,基于哈密顿函数 H 的值,确定 Metropolis 准则,确定生成的候选状态被接受或者拒绝.HMC 完美地模拟了哈密顿动态系统,因此每个这种候选状态都会高概率地被接受(因为 H 的值会保持不变)[39].HMC 采样过程描述见算法 2.

算法 2. Hamiltonian Monte Carlo 采样.

输入:起始位置 $z^{(1)}$,步长 ε.

输出:后验分布 Pr(z|D)的有限采样 N.

算法步骤:

for t=1,2,...do

1. 对动量 r 进行重新采样: $r^{(\mathcal{T})}\sim Norm(0,M),(z_0,r_0)=(z^{(\mathcal{T})},r^{(\mathcal{T})})$
2. 对离散化的对哈密顿动力系统进行采样

   a) $z_0:=z_0-\frac{\varepsilon}{2}\nabla U(z_0)$

   b) for i=1 to m do

   $z_i:=z_{i-1}+\varepsilon M^{-1}r_{i-1}$

   $r_i:=r_{i-1}-\varepsilon\nabla U(z_i)$

   end

   c) $r_m:=r_m-\frac{\varepsilon}{2}\nabla U(z_m)$

   d) $(\hat{z},r):=(z_m,r_m)$
3. 进行 Metroplis-Hastings 修正,消除离散化带来的误差

   a) 计算接受率 $\rho:=\mathrm{e}^{H(\hat{z},r)-H(z^{(\mathcal{T})},r^{(\mathcal{T})})}$

   b) 从[0,1]上的均匀分布采样一个μ,若 min(1,ρ)>μ,则 $z^{(t+1)}:=\hat{z}$

   End

与基本的 MH 方法不同,HMC 能够利用到对数概率分布的梯度信息和概率分布本身的信息[40].在函数最优化领域有一个类似的情形,大多数可以得到梯度信息的情况下,使用哈密顿动力学方法是很有优势的.可以直观地解释为,这种现象是由于下列的事实造成的:在 n 维空间中,与计算函数本身的代价相比,计算梯度所带来的额外计算代价通常是一个与 n 无关的固定因子.而与函数本身只能传递一条信息相比,n 维梯度向量可以传递 n 条信息.

4.3 稀疏可交换图模型的后验分布的推理

首先对受限 CRM W_φ的条件分布 $W_\varphi|D_\varphi$进行分析,$W_\varphi|D_\varphi$可以被分解成两部分之和,一部分是全体带随机权值 w_i 的固定位置 θ_i 的测度,对应着观测到的网络数据中至少关联了 1 条边的节点全体,另一部分是权值随机且位置随机的测度,对应了其他节点,用 w^* 表示这些节点的总的随机质量.

定理 4.1. 令 $(\theta_1,...,\theta_{N_\alpha})$ 是有向图 D_α的支撑点, $D_\alpha=\sum_{1\leqslant i,j\leqslant N_\alpha} n_{ij}\delta(\theta_i,\theta_j)$,令 $m_i=\sum_{j=1}^{N_\alpha}(n_{ij}+n_{ji})>0$,i=1,...,N_α, 则有 $W_\alpha|D_\alpha$ $w_*\sum_{i=1}^{\infty}\tilde{P}_i\delta_{\tilde{\theta}_i}+\sum_{i=1}^{N_\alpha}w_i\delta_{\theta_i}$ $w_*\sum_{i=1}^{\infty}\tilde{P}_i\delta_{\tilde{\theta}_i}+\sum_{i=1}^{N_\alpha}w_i\delta_{\theta_i}$, 这里, $\theta_i\sim Unif([0,\alpha])$, 权值序列 $(\tilde{P}_i)_{i=1,2,...}$ 满足 $\tilde{P}_1>\tilde{P}_2>...$ 且 $\sum_{i=1}^{\infty}\tilde{P}_i=1$, 给定 w_*, $(\tilde{P}_i)$ 服从一个列维强度为ρ的 Poisson-Kingman 分布,即 $(\tilde{P}_i)\,|\,w_*\sim PK(\rho\,|\,w_*)$. 最终,给定 D_α,权值 (w,w_*) 的联合依赖条件分布为 $(w_1,...,w_{N_\varphi},w_*)\,|\,D_\alpha\propto\left[\prod_{i=1}^{N_\alpha}w_i^{m_i}\right]\mathrm{e}^{-\left(\sum_{i=1}^{N_\alpha}w_i+w_*\right)^2}\left[\prod_{i=1}^{N_\alpha}\rho(w_i)\right]\times g_\alpha^*(w_*)$, 这里, g_α^* 是随机变量 $W_\alpha^*=W_\alpha([0,\alpha])$ 的概率密度函数,其拉普拉斯变换为 $\mathbb{E}\left[\mathrm{e}^{-tW_\alpha^*}\right]=\mathrm{e}^{-\alpha\psi(t)}$ [15].

需要注意的是,规一化的权值 $(\tilde{P}_i)_{i=1,2,...}$ 和位置 $(\theta_i)_{i=1,2,...}$ 不是似然可识别的(likelihood identifiable),因为观测到的网络节点数据仅包含节点权值和 w_* 的信息,还需要注意的是 $(w_1,...,w_{N_\alpha},w_*)|D_\alpha$ 不依赖于节点位置 $(\theta_1,...,\theta_{N_\alpha})$, 我们考虑的是齐次 CRM,这一点非常重要,因为节点位置 $(\theta_1,...,\theta_{N_\alpha})$ 通常无法观测到的,推理过程不会考虑节点位置.

James 等人[41]通过引入一个辅助随机变量 U 刻画了 NRMI 所导出的 EFFP,令 Γ_n 是一个 Gamma 随机变量,其位置参数是 1,形状参数是 n,并且独立于总质量 T,取值为正的随机变量 U_n=Γ_n/T,容易证明对于任意的 $n\geqslant 1$ 可以给出的 U_n 密度函数 $f_{U_n}(W)=\dfrac{W^{n-1}}{\Gamma(n)}\int_{\mathbf{R}_+}t^n\mathrm{e}^{-Wt}f_T(t)\mathrm{d}t$.

特别地, $\Pr(\pi=\{A_1,...,A_k\},\{X_k^*\in\mathrm{d}x_k:k\in\pi\},U\in\mathrm{d}W\,|\,W)=\dfrac{1}{\Gamma(n)}W^{n-1}\mathrm{e}^{-TW}\mathrm{d}W\prod_{k\in\pi}W(\mathrm{d}x_k)^{|k|}$ (给定 W,X 和 U 的联合条件分布),利用 Palm 公式,可以由(w,w_*)的联合依赖条件分布推导出对 EPPF 和预测分布体系的刻画.

定理 4.2. 令 $\tilde{W}$ 是一个齐次 NRM,其列维测度为ρ,基分布为μ_0,拉普拉斯指数为 $\psi(W)$,边缘化 $\tilde{W}$ 后,X 和 U 的联合分布为 $\Pr(\pi=\{A_1,...,A_k\},\{X_k^*\in\mathrm{d}x_k:k\in\pi\},U\in\mathrm{d}W)=\dfrac{1}{\Gamma(n)}W^{n-1}\mathrm{e}^{-\psi(W)}\mathrm{d}W\prod_{k\in\pi}\mathcal{C}_{|k|}(W)\mu_0(\mathrm{d}x_k)$, $\mathcal{C}_m(W)$ 是指数倾斜列维测度 $\mathrm{e}^{-us}\rho(\mathrm{d}s)$ 的第 m 阶矩 $\mathcal{C}_m(W)=\int_0^\infty w^m\mathrm{e}^{-wt}\rho(w)\mathrm{d}w$. 特定地,当将辅助随机变量 U 边缘化后,EPPF 的表达式为 $\Pr(\pi=\{A_1,...,A_k\})=\int_{\mathbf{R}_+}\dfrac{1}{\Gamma(n)}u^{n-1}\mathrm{e}^{-\psi(W)}\mathrm{d}W\prod_{k\in\pi}\mathcal{C}_{|k|}(W)\mathrm{d}W$, 因为序列 $\{X_k^*\in\mathrm{d}x_k:k\in\pi\}$ 独立同分布于μ_0,相应的预测分布体系为 $\Pr(X_{n+1}\in\mathrm{d}x\,|\,U,X)\propto\mathcal{C}_{|k|}(W)\mu_0(\mathrm{d}x)+\sum_{k\in\pi}\dfrac{\mathcal{C}_{|k|+1}(U)}{\mathcal{C}_{|k|}(U)}\delta_{X_k^*(\mathrm{d}x)}$, 注意,$U$ 在给出 X 时的后验分布为 $\Pr(U\in\mathrm{d}W\,|\,X)\propto W^{n-1}\mathrm{e}^{-\psi(W)}\mathrm{d}W\prod_{k\in\pi}\mathcal{C}_{|k|}(W)$ [15].

定理 4.2 表明,在边缘采样中,CRM W 被边缘后,就可以采样得到代表 X 的分布的分区结构和聚类参数.

定理 4.3. 令 $\tilde{W}$ 是一个齐次 NRMI,列维测度为ρ,基分布为μ_0,对应的 CRM W 在给出 X 和 U 时的后验分布为 $W\,|\,U,X\sim W'+\sum_{k\in\pi}J_k'\delta_{X_k^*}$, W' 是一个齐次 CRM,具有指数倾斜列维强度 $\upsilon'(\mathrm{d}s,\mathrm{d}y)=\mathrm{e}^{-Us}\rho(\mathrm{d}s)\mu_0(\mathrm{d}y)$, 随机质量 $\{J_k':k\in\pi\}$ 相互独立并且与 W' 独立,其条件分布为 $\Pr(J_k'\in\mathrm{d}s\,|\,U,X)=\dfrac{1}{\mathcal{C}_{|k|}(U)}s^{|k|}\mathrm{e}^{-Us}\rho(\mathrm{d}s)$, $\tilde{W}$ 在给出 X 和 U 时的后验分布就是对 W,U,X 进行规一化[15].

定理 4.3 表明一个齐次 NRMI $\tilde{W}$ 在给出 X 和 U 时的后验分布仍然是一个 NRMI.

以下介绍 Caron 和 Fox 提出的对生成模型为(2.4)的稀疏可交换图所进行的后验推理[15].针对 CRM 是一个 GGP 的情况,其后验推理的 MCMC 采样过程,令 $\phi=(\alpha,\sigma,\tau)$ 是超参数集合,这些超参数也需要推理得到,为这些超参数假设合适的先验 $\Pr(\alpha)\propto\dfrac{1}{\alpha}$, $\Pr(\sigma)\propto\dfrac{1}{1-\sigma}$, $\Pr(\tau)\propto\dfrac{1}{\tau}$,为了强调列维测度和 w_*的概率密度函数对超参数的依赖,记列维强度为 $\rho(w|\sigma,\tau)$和 $g_{\alpha,\sigma,\tau}^*(w_*)$. 对于有向多图,需要近似推理的是后验分布 $\Pr(w_1,...,w_{N_\varphi},w_*,\phi\,|\,(n_{ij})_{1\leqslant i,j\leqslant N_\varphi})$, 对于无向图,需要要近似推理的是后验分布 $\Pr(w_1,...,w_{N_\varphi},w_*,\phi\,|\,(z_{ij})_{1\leqslant i,j\leqslant N_\varphi})$.

在观测数据是一个无向简单图的情况下,需要推算未知的有向边数 n_{ij},观测到 z_{ij}=1($i\leqslant j$)时,引入一个潜变量 $\overline{n}_{ij}=n_{ij}+n_{ji}$,其条件分布为

$$\overline{n}_{ij}\mid z,w\sim\begin{cases}\delta_0, & \text{if } z_{ij}=0\\ tPoisson(2w_iw_j), & \text{if } z_{ij}=0, i\neq j\\ tPoisson(w_i^2), & \text{if } z_{ij}=0, i=j\end{cases} \tag{4.1}$$

此处,$tPoisson(\lambda)$是一个零截断泊松分布,其概率分布密度函数为 $\frac{k^{\lambda}\exp(-\lambda)}{(1-\exp(-\lambda))k!}$,for $k=1,2,...$,方便起见,令 $\overline{n}_{ij}=\overline{n}_{ji}$, $m_i=\sum_{j=1}^{N_\varphi}\overline{n}_{ij}$. 为了更有效地得到所要推理的后验分布,使用 HMC 与 Gibbs 采样相结合对 $(w_1,w_2,...,w_{N_\varphi})$ 进行更新,HMC 需要计算对数后验的梯度 $\left[\nabla_{\omega_{1:N_\alpha}}\log(\Pr(w_{1:N_\alpha},w_*\mid D_\alpha)\right]_i=m_i-\sigma-w_i\left(\tau+2\sum_{j=1}^{N_\alpha}w_j+2w_*\right)$, $\omega_i=\log w_i$. 采用 Metropolis-Hastings 算法对总质量 w_*和超参数ϕ进行更新.需要注意的是,除了特定的当 $\sigma=0,\frac{1}{2}$ 时,其他情况下 w_*的概率密度函数 $g^*_{\alpha,\sigma,\tau}(w_*)$ 都不能被解析表示,需要通过对 $g^*_{\alpha,\sigma,\tau}(w_*)$ 的指数倾斜进行采样得到 w_*[15].

综上所述,当稀疏可交换图是一个有向图时,后验分布 $\Pr(w_1,...,w_{N_\alpha},w_*,\phi\mid(n_{ij})_{1\leqslant i,j\leqslant N_\alpha})$ 的 MCMC 推理过程包括以下两个步骤,若是一个无向图是,后验分布 $\Pr(w_1,...,w_{N_\alpha},w_*,\phi\mid(z_{ij})_{1\leqslant i,j\leqslant N_\alpha})$ 的 MCMC 推理还需要加上第 3 个步骤.

1. 给定 $w_*,\phi,(n_{ij})$,采用 HMC 方法对社交能力参数 $w=(w_1,w_2,...,w_{N_\alpha})$ 进行更新;
2. 给定 $(w_1,w_2,...,w_{N_\alpha})$ 和(n_{ij}),采用 MH 算法对总质量 w_*和超参数ϕ进行更新;
3. 给定 $(w_1,w_2,...,w_{N_\alpha})$,$w_*,\phi$,采用条件分布或 MH 算法对 $\overline{n}_{ij}$ 进行更新.

步骤 1:采用 HMC 对 $(w_1,w_2,...,w_{N_\alpha})$ 进行更新,令 $L\geqslant 1$ 是蛙跳步数,$\varepsilon>0$ 是步长,对数后验的梯度记作 $U'(w,w_*,\phi)=\nabla_{\omega_{1:N_\alpha}}\log(Pr(w_{1:N_\alpha},w_*\mid D_\alpha)$, HMC 算法过程如下.

(1) 首先对辅助动量变量 $r:N(0,I_{N_\alpha})$ 进行采样.

(2) (w,r)为哈密顿动力系统的初始状态,通过蛙跳离散化得到候选状态 $(\tilde{w},\tilde{r})$,步骤如下.

① $\tilde{r}^{(0)}=r+\frac{\varepsilon}{2}U'(w,w_*,\phi)$, $\tilde{w}^{(0)}=w$;

② for $l=1,...,L-1$

$$\log\tilde{w}^{(l)}=\log\tilde{w}^{(l-1)}+\varepsilon\tilde{r}^{(l-1)},$$
$$\tilde{r}^{(l)}=\tilde{r}^{(l)}+\varepsilon U'(\tilde{w}^{(l)},w_*,\phi);$$

③
$$\log\tilde{w}=\log\tilde{w}^{(L-1)}+\varepsilon\tilde{r}^{(L-1)},$$
$$\tilde{r}=-\left[\tilde{r}^{(L-1)}+\frac{\varepsilon}{2}U'(\tilde{w},w_*,\phi)\right],$$
$$\tilde{w}=\tilde{w}^{(L)}.$$

(3) 以概率 $accep$=min(1,ρ)接受 $(\tilde{w},\tilde{r})$,ρ为接受率,以下省略关于ρ的计算表达式,可参阅文献[15].

步骤 2:采用 MH 算法对 w_*,α,σ,τ 进行更新,从提议分布 $q(\tilde{\alpha},\tilde{\sigma},\tilde{\tau},\tilde{w}_*\mid\alpha,\sigma,\tau,w_*)$ 采样得到 $(\tilde{\alpha},\tilde{\sigma},\tilde{\tau},\tilde{w}_*)$ 并以概率 $accep$=min(1,mhr)接受它.可以对提议分布进行因子分解,得到:

$$q(\tilde{\alpha},\tilde{\sigma},\tilde{\tau},\tilde{w}_*\mid\alpha,\sigma,\tau,w_*)=q(\tilde{\tau}\mid\tau)q(\tilde{\sigma}\mid\sigma)q(\tilde{\alpha}\mid\tilde{\sigma},\tilde{\tau},w_*)q(\tilde{w}_*\mid\tilde{\alpha},\tilde{\sigma},\tilde{\tau},w_*).$$

这里,

$$q(\tilde{\tau}\mid\tau)=\log normal(\tilde{\tau};\log(\tau),\sigma_\tau^2),$$
$$q(\tilde{\sigma}\mid\sigma)=\log normal(1-\tilde{\sigma};\log(1-\sigma),\sigma_\tau^2),$$
$$q(\tilde{\alpha}\mid\tilde{\sigma},\tilde{\tau},w_*)=Gamma\left(\tilde{\alpha};N_\varphi,\frac{\left(\tau+2\sum w_i+w_*\right)^{\tilde{\sigma}}-\tau^{\tilde{\sigma}}}{\tilde{\sigma}}\right),$$
$$q(\tilde{w}_*\mid\tilde{\alpha},\tilde{\sigma},\tilde{\tau},w_*)=g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}+2\sum w_i+w_*}(\tilde{w}_*)=\frac{\exp\left(-2\sum w_i-w_*\right)g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}}(\tilde{w}_*)}{\exp(-\psi_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}}(\tilde{w}_*))}.$$

$\tilde{w}_*$的提议分布的选择基于$\tilde{w}_*$的分布$g^*_{\tilde{\alpha},\tilde{\sigma},\tilde{\tau}}(\tilde{w}_*)$的指数倾斜,这样就可以简化计算接受率$\rho$的表达式.指数倾斜也称埃舍尔变换,是对随机测度施加的一个指数式改变.若函数$\varphi:[0,\infty]\to\mathbf{R}$连续且各阶导数存在并满足$(-1)^k\varphi^{(k)}(t)\geqslant 0$(对于任意$t\in(0,\infty)$,$k\in\mathbf{N}$),则称$\varphi$完全单调.令$\Psi_\infty$表示满足$\varphi(0)=1$的全体完全单调函数$\varphi:[0,\infty]\to[0,1]$构成的集合,由 Bernstein 定理:$[0,\infty]$上的分布函数$F$的 Laplace-Stieltjes 变换是一个完全单调函数$\varphi:[0,\infty]\to[0,1]$当且仅当$\varphi\in\Psi_\infty$,即$\varphi=\mathcal{LS}(F)$或$F=\mathcal{LS}^{-1}(\varphi)$ iff $\varphi\in\Psi_\infty$,容易证明$\tilde{\varphi}(t):=\varphi(\lambda+t)/\varphi(t)$也在$\Psi_\infty$中,对应的分布函数$\tilde{F}=\mathcal{LS}^{-1}(\tilde{\varphi})$称作是分布函数$F=\mathcal{LS}^{-1}(\varphi)$的指数倾斜($\lambda$称作倾斜参数),因为若$F$的密度函数为$f$,则$\tilde{F}$的密度函数$\tilde{f}(x)=\exp(-\lambda x)f(x)/\varphi(\lambda),x\in[0,\infty)$是对$f$做了指数倾斜[42].

步骤 3:对潜变量$\overline{n}_{ij}$的更新,可以通过直接对条件分布(4.1)进行采样得到.当网络中的边数量众多的时候,一个更有效的更新方法是使用一个 Metropolis-Hastings 提议$q(\tilde{\overline{n}}_{ij}\mid\overline{n}_{ij})=\begin{cases}\frac{1}{2}, & 若\tilde{n}_{ij}=n_{ij}+1,n_{ij}>1\\ \frac{1}{2}, & 若\tilde{n}_{ij}=n_{ij}-1,n_{ij}>1\\ 1, & 若\tilde{n}_{ij}=n_{ij}+1,n_{ij}=1\\ 0, & 其他\end{cases}$,并且以概率$\min\left(1,\frac{\overline{n}_{ij}!}{\tilde{\overline{n}}_{ij}!}((1+\delta_{ij})w_iw_j)^{\tilde{n}_{ij}-n_{ij}}\frac{q(\overline{n}_{ij}\mid\tilde{\overline{n}}_{ij})}{q(\tilde{\overline{n}}_{ij}\mid\overline{n}_{ij})}\right)$接受从提议分布中得到的采样.

4.4 对指数倾斜稳定分布进行采样(sampling exponentially tilted stable distribution)

在$\sigma\geqslant 0$的情况下,W^*_φ是一个服从指数倾斜稳定分布(ETSD)的随机变量.稳定分布$S(\alpha,\beta,\gamma,\delta;1)$又称为非高斯稳定分布、重尾分布,由列维提出,是唯一满足广义中心极限定理的分布(即无限多个可能方差无限大的独立分布的随机变量之和,其极限分布是稳定分布)[43].高斯分布(α=2)、柯西分布(α=1,β=0)、列维分布(α=1/2,β=1)都仅是稳定分布的特例,很多不满足经典的中心极限定理的数据都可以用稳定分布来描述,因此具有更普遍的应用范围.同高斯分布具有指数衰减的拖尾不同,稳定分布的拖尾以平方律衰减,衰减的速度与指数α有关,α越小,分布的衰减就越缓慢,分布的拖尾越重.

指数倾斜稳定分布是严格稳定分布天然的指数族分布,采用指数倾斜可以在求解贝叶斯估计的近似公式时不涉及似然函数的条件最大值,求解过程更稳定,并且显著减少了所需要的计算时间.将严格稳定分布与指数倾斜相结合已经被证明在很多应用(例如,重要性采样、罕见事件的模拟、保险精算等)中非常有效[44].

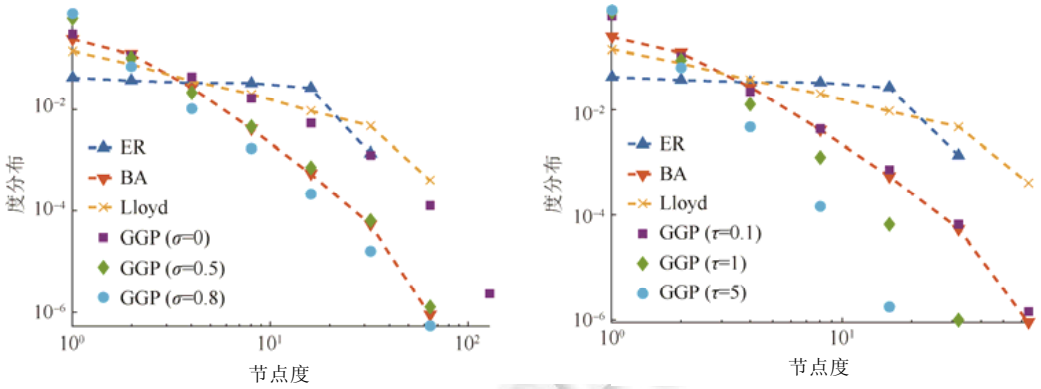
Brix、Rosinski、Ridout、Devroye、Hofert 对指数倾斜稳定分布采样算法进行了研究,针对分布为$\tilde{S}\sim\tilde{S}(\alpha,1,\cos(\alpha\pi/2)^{1/\alpha},1_{\{\alpha=1\}},\lambda 1_{\{\alpha\neq 1\}};1)$, $\alpha\in(0,1],\lambda\in[0,\infty)$的指数倾斜稳定分布,Hofert[44]提出了一种快速拒绝采样算法,Devroye[45]提出了一种双拒绝采样算法,这两种算法都属于精确采样算法.快速拒绝采样算法简单容易实现,当参数范围很大时采样速度也很快;双拒绝采样算法比较复杂,但是在所有参数组合上的复杂性界是均等的,非常适用于那些即使在线性复杂度λ^α下也很耗时的参数的采样问题.Hofert[42]修正了双拒绝采样算法中的两个小问题,并且解决了在α较小时存在的一些临界数值计算过程,使得双拒绝采样算法成为一种非常重要的指数倾斜稳定分布采样算法.

5 实验验证

Caron 等人[15]采用他们提出的生成模型模拟生成了一系列随机图(σ和τ取不同值情况下),并与采用 ER 模型[7]、PA 模型[13]、Lloyd[18]模拟生成的随机图进行了对比,结果表明 Caron-Fox 模型能够模拟生成节点度分布呈现幂律分布而且具有重尾特性的稀疏图,并且能够处理节点度分布尾部的指数截断(这一性质是进行网络数据分析时非常有用的一个重要性质),但是采用 3 个作对比的生成模型模拟生成的随机图则不能表现出这些性质.

如图 3(a)所示,在 log-log 坐标系下,采用 GGP 作为先验的 Caron-Fox 模型在σ分别取 0.2,0.5,0.8 的情况下模

拟生成的随机图,其节点度分布呈现出明显的幂律分布,而 p=0.05 的 ER 模型 $G(n,p)$和 Lloyd 模型模拟生成的随机图,其节点度分布则不能呈现出幂律分布,PA 模型模拟生成的随机图的节点度分布虽然呈现出了幂律分布,但不是可交换图,图 3(b)所示的是当τ分别取 0.1,1,5 时的情况,可以得到和图 3(a)类似的结论(图例中的 BA 表示 PA 模型,因为 PA 模型是由 Barabási 和 Albert 共同提出的,所以也常常被称作 BA 模型).



(a) τ=2,σ取不同值的情况 (b) σ=0.5,τ取不同值的情况

Fig.3 Degree distribution on a log-log scale

图 3 Log-Log 坐标系下的节点度分布图

如图 4 所示,采用 GGP 作为先验的 Caron-Fox 模型模拟生成的随机图表现出了这样一种特性:随着节点个数的增加,图中度是 1 的节点的个数是线性增加,PA 模型模拟生成的随机图有类似的性质,但是 ER 模型模拟生成的随机图中度是 1 的节点的个数是随着节点个数的增加呈现指数减少的,Lloyd 模型模拟生成的随机图则几乎没有度是 1 的节点.

如图 5 所示,采用 GGP 作为先验的 Caron-Fox 模型模拟生成的随机图表现出了这样一种特性:随着节点个数 n 的增加,图中边的个数呈现 $o(n^2)$增加(稀疏图的特征),PA 模型模拟生成的随机图有类似的性质,但是 ER 模型和 Lloyd 模型模拟生成的随机图中,边的个数呈现$\Theta(n^2)$增加(稠密图的特征).

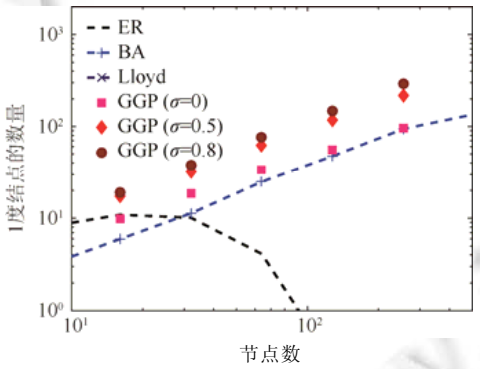


Fig.4 Number of nodes with degree 1 versus the number of nodes on a log-log scale

图 4 Log-Log 坐标系下图中度是 1 的节点的个数随着节点个数的增加而变化的情况

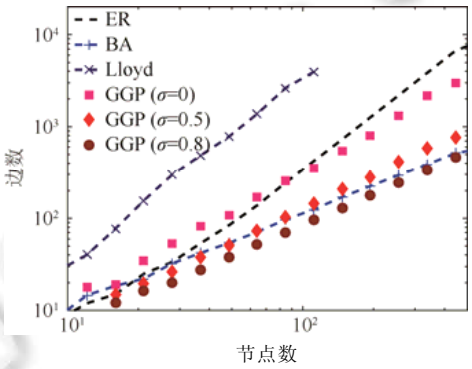


Fig.5 Number of edges versus the number of nodes on a log-log scale

图 5 Log-Log 坐标系下图中边的个数随着节点个数的增加而变化的情况

实证结果表明,当$\sigma \geqslant 0$ 时,采用 GGP 作为先验的 Caron-Fox 模型可以生成稀疏可交换图,而基于 Aldous-Hoover 表示理论的 ER 模型和 Lloyd 模型只能生成稠密可交换图,PA 模型虽然模拟生成稀疏图,但不是可交换

的.关于实证分析的更详细过程,可参阅文献[15].

6 最新进展与研究展望

在随机图模型方面,通过对经典 Graphon 的定义进行推广,Borgs 等人[46]提出了 Graphon process 模型,Veitch 等人[11]提出了 Graphex 模型,这两种更广义的稀疏可交换图模型都可以对一大类可交换图进行建模,并且将 Caron-Fox 模型和传统可交换稠密图作为特例纳入了一个统一的模型框架.Crane 等人[47]认为,尽管 Caron-Fox 模型能够生成满足可交换性的稀疏网络,但是尚不清楚可交换性是否在各种采样策略下都能得到保证,他们提出了一种边可交换的随机图模型(edge exchangeable model);Jacobs 和 Clauset[48]对各种网络生成模型从哲学思想、表示和统一性这 3 个方面进行了总结,提出了网络生成模型在可解释性、同质性、优化、模型可实现、模型选择、模型检验等方面所面临的挑战和机遇,并且指出 Caron-Fox 模型只能生成超线性密度(superlinear density)$O(n^{\alpha})$,$1<\alpha<2$ 的稀疏可交换图,还不能生成更稀疏的 $O(n)$数量级的节点数对应 $O(n)$数量级的边数的稀疏可交换图,还需要在理论创新和推理算法方面进行大量的研究才能使得这种基于连续空间的网络生成模型被应用到更广泛的真实网络数据分析中.

社区结构作为网络模型的重要特征刻画了网络的多种属性和功能,对人们准确理解复杂系统的特性有十分重要的意义.从社区结构这一中观层面对复杂网络进行研究对人们准确理解复杂系统的特性有十分重要的意义,既能够弥补宏观层面粒度过粗所造成的很多网络特性无法观察到的缺陷,又避免了微观层面粒度过细所带来的丢失共性并且计算复杂度高等问题.社区发现作为网络数据分析的重要内容已经得到了学术界的广泛关注和深入研究,Herlau 等人[14]将 SBM 与 Caron-Fox 模型相结合,为稀疏可交换图引入了社区结构,并提出了稀疏可交换图的非重叠静态社区发现方法;Todeschini 和 Caron[49]提出了稀疏可交换图的重叠静态社区发现方法,该方法受到了 Zhou[50]提出的无限边分区模型(Infinite edge partition model)的启发,假设网络中的节点 i 关联了一组权重参数 $w_{ik},k=1,...,p$(可以解释为节点的不同方面的社交能力),节点 i 关联的权重参数是有依赖关系的混合随机测度(compound random measure)[51],网络的生成模型为 $z_{ij}\sim Bern\left(1-\exp\left(-\sum_{k=1}^{p}\sum_{l=1}^{p}\eta_{kl}w_{ik}w_{jl}\right)\right)$,这样就可以同时发现网络中的同配结构和异配结构.Borgs 等人[46]提出了将 MMSB 与 Graphon process 模型相结合的重叠静态社区发现方法.

在稀疏可交换图的动态演化方面,Palla 等人[52]提出了采用动态泊松点过程对节点社交能力随时间演化的特性进行建模,可以同时捕获节点间边的平滑演化(smooth evolution)和长程演化(long term evolution).Matias[53]指出采用离散模型对网络进行建模,在研究网络和社区结构的动态演化时,需要将数据聚集到预先指导的时间段以得到网络快照序列,数据聚集会带来信息损失,并且时间段的选择会直接影响预测结果,但是采用随机过程作为模型参数的连续模型,则能够很自然地将时间信息引入网络模型,避免了上述问题.

在近似推理方法方面,现有的关于稀疏可交换图的研究工作几乎都采用的是 MCMC 推理方法.相对于随机变分推理(stochastic variational inference,简称 SVI)方法,MCMC 方法因为可扩展性和可并行性差从而导致其很难应用于大数据环境下的数据分析,因此如何采用 SVI 对稀疏可交换图进行推理是一个非常重要的研究课题.Roychowdhury[54]提出了伽马过程(GP)的变分推理方法;Tank 等人[55]提出了贝叶斯非参数混合模型的流式变分推理算法;缘于 MCMC 与 VI 在能量函数和信息熵层面颇有渊源,Salimans 等人[56]提出通过桥接二者之间的嫌隙将 MCMC 和 VI 结合起来进行推理;Wolf 等人[57]提出了一种将 HMC 和 VI 相结合的算法,把 HMC 步骤包含到变分下界,以加速后验推理过程的收敛.

Caron-Fox 模型中提出的“节点社交能力”这一概念可以为网络分析提供非常合理的社会学解释——节点间建立边的概率取决于两个节点潜在的社交能力,并且正是因为具有不同方面的社交能力(或者说不同方面的潜在兴趣),决定了一个节点可以同时隶属于不同的社区.“节点社交能力”与度纠正的随机块模型(degree-corrected SBM)[58]中的“度偏好”异曲同工,都能使网络生成模型很好地拟合具有各种特定度分布的真实网络,使得 Caron-Fox 模型与 SBM 模型相结合时,不需要再进行度纠正.

稀疏可交换图模型是对传统可交换稠密图模型的扩展.关于传统可交换稠密图的研究,从网络建模、网络

演化到社区发现、社区演化,再到链路预测、影响力分析、网络传播等各个方面的研究工作,都已经历时弥久并且成果丰硕.尽管传统可交换稠密图的先天不足导致其无法再继续成为对真实复杂网络进行数据分析的利器,但是很多已经被提出的研究问题和研究成果可以在进行稀疏可交换图研究时被广泛继承和发展.

综上所述,Caron 等人的开创性工作为复杂网络数据分析带来了新的转机,稀疏可交换图模型不仅解决了可交换性与稀疏性二者不可兼顾的矛盾,并且具有很多独特的优良基因.随着对稀疏可交换图模型以及贝叶斯非参数混合模型的深入研究,必将可以推动复杂网络数据分析,尤其是网络社区结构动态演化研究工作的进一步发展.

致谢 文中关于哈密顿动力学方法和混合蒙特卡罗算法的部分内容引自马毅未公开发表的对文献[39]的翻译手稿,在此向马毅表示感谢.

References:

[1] Yu ZW, Zhou XS. Relocation and discussion about pervasive computing. Communications of the CCF, 2011,7(7):49–56 (in Chinese).

[2] Yu ZW, Yu ZY, Zhou XS. Socially aware computing. Chinese Journal of Computers, 2012,35(1):16–26 (in Chinese with English abstract). [doi: 10.3724/SP.J.1016.2012.0001]

[3] Freer CE, Roy DM. Computable exchangeable sequences have computable de Finetti measures. In: Ambos-Spies K, Lőwe B, Merkle W, eds. Proc. of the CiE 2009. LNCS 5635, Berlin, Heidelberg: Springer-Verlag, 2009. 218–231.

[4] Aldous DJ. Representations for partially exchangeable arrays of random variables. Journal of Multivariate Analysis, 1981,11(4): 581–598. [doi: 10.1016/0047-259X(81)90099-3]

[5] Hoover DN, Keisler HJ. Adapted probability distributions. Trans. of the American Mathematical Society, 1984,286(1):159–201. [doi:10.1090/S0002-9947-1984-0756035-8]

[6] Kallenberg O. Exchangeable random measures in the plane. Journal of Theoretical Probability, 1990,3:81–136. [doi:10.1007/BF01063330]

[7] Erdős P, Rényi A. On the evolution of random graphs. Trans. of the American Mathematical Society, 2011,286(1):257–274.

[8] Airoldi EM, Costa TB, Chan SH. Stochastic blockmodel approximation of a graphon: Theory and consistent estimation. In: Advances in Neural Information Processing Systems, 2013,26:692–700.

[9] Airoldi EM, Blei D, Fienberg SE, Xing E. Mixed membership stochastic blockmodels. Journal of Machine Learning Research, 2007,9(5):1981–2014.

[10] Kemp C, Tenenbaum JB, Griffiths TL, Yamada T, Ueda N. Learning systems of concepts with an infinite relational model. In: Proc. of the National Conf. on Artificial Intelligence. 2006. 381–388.

[11] Veitch V, Roy DM. The class of random graphs arising from exchangeable random measures. Electronic pre-print, arXiv: 1512.03099, 2015. http://arxiv.org/abs/1512.03099

[12] Orbanz P, Roy DM. Bayesian models of graphs, arrays and other exchangeable random structures. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,37(2):437–461. [doi: 10.1109/TPAMI.2014.2334607]

[13] Barabási AR, Albert L. Statistical mechanics of complex networks. Reviews of Modern Physics, 2002,74(1):47–97.

[14] Herlau T, Schmidt MN, Mørup M. Completely random measures for modelling block-structured sparse networks. In: Advances in Neural Information Processing Systems (NIPS). Barcelona, 2016.

[15] Caron F, Fox E. Sparse graphs using exchangeable random measures. Journal of the Royal Statistical Society B, 2017,79:1295–1366.

[16] Borgs C, Chayes JT, Cohn H, Ganguly S. Consistent nonparametric estimation for heavy-tailed sparse graphs. Electronic pre-print, arXiv: 1508.06675, 2015. http://arxiv.org/abs/1508.06675

[17] Kingman JFC, Poisson processes. London: Oxford University Press, 1992. 79–98.

[18] Lloyd JR, Orbanz P, Ghahramani Z, Roy DM. Random function priors for exchangeable arrays. Advances in Neural Information Processing Systems (NIPS), 2012,25(6):1007–1015.

[19] Lijoi A, Prünster I. Models beyond the Dirichlet process. SSRN Electronic Journal, 2009,(23):80–136. [doi: 10.2139/ssrn.1526505]

[20] Kingman JFC. Completely random measures. Pacific Journal of Mathematics, 1967,21(1):59–78. [doi: 10.2140/ pjm.1967.21.59]

[21] Hougaard P. Survival models for heterogeneous populations derived from stable distributions. Biometrika, 1986,73(2):387–396. [doi: 10.1093/biomet/73.2.387]

[22] Caron F, Teh YW, Murphy TB. Bayesian nonparametric Plackett-Luce models for the analysis of preferences for college degree programmes. The Annals of Applied Statistics, 2014,8(2):1145–1181. [doi: 10.1214/14-AOAS717]

[23] Muliere P, Tardella L. Approximating distributions of random functionals of Ferguson-Dirichlet priors. The Canadian Journal of Statistics, 1998,26(2):283–297. [doi: 10.2307/3315511]

[24] Ishwaran H, Zarepour M. Exact and approximate sum represenations for the Dirichlet process. The Canadian Journal of Statistics, 2002,30(2):269–283. [doi: 10.2307/3315951]

[25] Ishwaran H, James L. Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association, 2001, 96(453):161–173. [doi: 10.2307/2670356]

[26] Papaspiliopoulos O, Roberts GO. Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models. Biometrika, 2008,95(1):169–186.

[27] Walker SG. Sampling the Dirichlet mixture model with slices. Communications in Statistics: Simulation and Computation, 2007, 36(1):45–54. [doi: 10.2139/ssrn.945330]

[28] Kalli M,Griffin JE,Walker SG. Slice sampling mixture models. Statistics and Computing, 2011,21(1):93–105. [doi: 10.1007/ s11222-009-9150-y]

[29] Favaro S, Walker SG. Slice sampling σ-stable Poisson-Kingman mixture models. Journal of Computational and Graphical Statistics, 2013,22(4):830–847. [doi: 10.1080/10618600.2012.681211]

[30] Favaro S, Teh YW. MCMC for normalized random measure mixture models. Statistical Science, 2013,28(3):335–359. [doi: 10. 1214/13-STS422]

[31] Lewis PAW, Shedler GS. Simulation of nonhomogeneous Poisson processes by thinning. Naval Research Logistics, 1979,26(3): 403–413. [doi: 10.1002/nav.3800260304]

[32] Favaro S, Lijoi A, Nava C, Nipoti B, Pr¨unster I, Teh YW. On the stick-breaking representation for homogeneous NRMIs. Bayesian Analysis, 2016,11(3):697–724. [doi: 10.1214/15-BA964]

[33] Orbanz P, Williamson S. Unit-Rate Poisson representations of completely random measures. Electronic Journal of Statistics, 2011,1–12. http://stat.columbia.edu/~porbanz/reports/OrbanzWilliamson2012.pdf

[34] Ferguson TS, Klass MJ. A representation of independent increment processes without Gaussian components. The Annals of Mathematical Statistics, 1972,43(5):1634–1643. [doi: 10.1214/aoms/1177692395]

[35] Blackwell D, MacQueen JB. Ferguson distributions via Pólya urn schemes. The Annals of Statistics, 1973,1(2):353–355.

[36] Griffin JE. An adaptive truncation method for inference in Bayesian nonparametric models. Statistics and Computing, 2016,26(1): 423–441. [doi: 10.1007/s11222-014-9519-4]

[37] Neal RM. Markov chain sampling methods for dirichlet process mixture models. Journal of Computational and Graphical Statistics, 2000,9(2):249–265. [doi: 10.1111/j.1467-9469.2008.00609.x]

[38] Bishop CM. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, 2006. 548–556.

[39] Neal RM. MCMC using Hamiltonian dynamics. In: Brooks S, Gelman A, Jones G, Meng XL, eds. Handbook of Markov Chain Monte Carlo, Vol.2. Chapman & Hall/CRC Press, 2011. 113–160.

[40] Chen T, Fox EB, Guestrin C. Stochastic gradient Hamiltonian Monte Carlo. In: Proc. of the Int'l Conf. on Machine Learning. 2014. 1683–1691.

[41] James L, Lijoi A, Prünster I. Posterior analysis for normalized random measures with independent increments. Scandinavian Journal of Statistics, 2009,36(1):76–97. [doi: 10.1111/j.1467-9469.2008.00609.x]

[42] Hofert M. Sampling exponentially tilted stable distributions. ACM Trans. on Modeling and Computer Simulation, 2011,22(1): Article No.3. [doi: 10.1145/2043635.2043638]

[43] Kharroubi SA, Sweeting TJ. Exponential tilting in Bayesian asymptotics. Biometrika, 2016,103(2):337–349.

[44] Hofert M. Sampling Archimedean copulas. Computational Statistics and Data Analysis, 2008,52(12):5163–5174. [doi: 10.1016/j.csda.2008.05.019]

[45] Devroye L. Random variate generation for exponentially and polynomially tilted stable distributions. ACM Trans. on Modeling and Computer Simulation, 2009,19(4):1–20. [doi: 10.1145/1596519.1596523]

[46] Borgs C, Chayes JT, Cohn H, Holden N. Sparse exchangeable graphs and their limits via graphon processes. Electronic pre-print, arXiv: 1601.07134, 2016. http://arxiv.org/abs/1601.07134

[47] Janson S. On edge exchangeable random graphs. Journal of Statistical Physics, 2017,8:1–37. [doi: 10.1007/s10955-017-1832-9]

[48] Jacobs AZ, Clauset A. A unified view of generative models for networks: Models, methods, opportunities and challenges. Electronic pre-print, arXiv:1411.4070, 2014. http://arxiv.org/abs/1411.4070

[49] Todeschini A, Miscouridou X, Caron F. Exchangeable random measures for sparse and modular graphs with overlapping communities. Electronic pre-print, arXiv: 1602.02114, 2017. http://arxiv.org/abs/1602.02114

[50] Zhou MY. Infinite edge partition models for overlapping community detection and link prediction. In: Proc. of the 18th Int'l Conf. on Artificial Intelligence and Statistics (AISTATS). 2015. 1135–1143.

[51] Griffin JE, Leisen F. Compound random measures and their use in Bayesian nonparametrics. Journal of the Royal Statistical Society—Series B, 2017,79:525–545. [doi: 10.1111/rssb.12176].

[52] Palla K,Caron F,Teh YW. Bayesian nonparametrics for sparse dynamic networks. Electronic pre-print, arXiv: 1607.01624, 2014. http://arxiv.org/abs/1607.01624

[53] Matias C, Rebafka T, Villers F. A semiparametric extension of the stochastic block model for longitudinal networks. Electronic pre-print, arXiv: 1512.07075v2, 2016. http://arxiv.org/abs/1512.07075v2

[54] Roychowdhury A, Kulis B. Gamma processes, stick-breaking, and variational inference. In: Proc. of the 18th Int'l Conf. on Artificial Intelligence and Statistics (AISTATS). San Diego, 2015. 800–809.

[55] Tank A, Foti N, Fox EB. Streaming variational inference for Bayesian nonparametric mixture models. In: Proc. of the 18th Int'l Conf. on Artificial Intelligence and Statistics. 2014. 968–976.

[56] Salimans T, Kingma DP, Welling M. Markov chain Monte Carlo and variational inference: Bridging the gap. In: Proc. of the 32nd Int'l Conf. on Machine Learning (ICML). Lille, 2015. 1218–1226.

[57] Wolf C, Karl M, Smagt PVD. Variational inference with Hamiltonian Monte Carlo. Electronic pre-print, arXiv: 1609. 08203, 2016. http://arxiv.org/abs/1609.08203

[58] Karrer B, Newman ME. Stochastic blockmodels and community structure in networks. Physical Review E, 2011,83(2):96–107. [doi: 10.1103/PhysRevE.83.016107]

附中文参考文献:

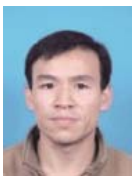
[1] 於志文,周兴社.普适计算的重定位与探讨.中国计算机学会通讯,2011,7(7):49–56.

[2] 於志文,於志勇,周兴社.社会感知计算:概念、问题及其研究进展.计算机学报,2012,35(1):16–26.

于千城(1976－),男,宁夏银川人,博士生,副教授,CCF 学生会员,主要研究领域为机器学习,复杂网络分析,社会感知计算.

王柱(1983－),男,博士,副教授,CCF 专业会员,主要研究领域为普适计算,移动社会网络,社会感知计算.

於志文(1977－),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为普适计算,移动社会网络,社会感知计算.



王晓峰(1981－),男,博士,副教授,CCF 专业会员,主要研究领域为算法分析与设计,智能计算,可计算性,计算复杂性.