arXiv:2302.02787v3 [cs.SI] 5 Oct 2023

网络中具有两种真实划分的生成模型

Lena Mangold[1,2,3,*] 和 Camille Roth[1,2,3]
[1] *Computational Social Science team, Centre Marc Bloch, Friedrichstr. 191, 10117 Berlin, Germany*
[2] *Centre national de la recherche scientifique (CNRS), 3 rue Michel-Ange, 75 016 Paris, France*
[3] *Centre d'Analyse et de Mathématique Sociales (CAMS), École des hautes études en sciences sociales (EHESS), 54 Bd Raspail, 75006 Paris, France*

（日期：2023年10月6日）

人们提出了无数种方法来刻画网络的中尺度结构——通常将其表示为一种划分，其模式被称为社区、块或簇。显然，旨在检测不同类型模式的不同方法，可能会对网络的中尺度结构给出各种各样的答案。然而，即使是对同一种方法多次运行，有时也会得到多样且相互矛盾的结果，从而形成完整的划分图景，其中可能包含对网络中尺度结构的多种（局部最优）解释。这种歧义促使我们进一步审视这些方法能否在网络中找到多个性质迥异的"真实标签"划分。在此，我们提出随机交叉块模型（SCBM），这是一种生成模型，能够将两个不同的划分纳入单个基准网络的中尺度结构中。我们通过评估随机块模型（SBM）检测隐式植入、强度各异的双社区和核心—边缘结构的能力，展示了这一基准模型的一种应用。根据我们的模型设计和实验设置，我们发现，各随机块模型变体单独检测这两种划分的能力各不相同，而只有在极少数情况下才能同时恢复这两种划分。我们的研究结果表明，在大多数情况下，即使存在其他划分，也只能检测到一种在某种程度上占主导地位的结构。这凸显了在存在相互竞争的解释时考察完整划分图景的必要性，并推动未来研究改进划分共存检测方法。我们的模型也为基准网络领域作出更广泛的贡献，使人们能够进一步探索新方法和现有方法检测网络中尺度结构歧义的能力。

一、引言

网络结构在中尺度层面上常通过一种通常称为"社区"的配置来刻画——即网络中连通性呈现某种相似性的节点群组。网络可能呈现多种中尺度结构，例如连接稠密或具有凝聚性的簇、同配或异配社群、核心—边缘结构、等价类，或它们的组合 [1, 2]。此外，划分现实世界网络中节点的方式往往不止一种具有科学合理性，例如，在多个案例中，凝聚簇与核心—边缘结构共存便证明了这一点 [3, 4]。

显然，旨在识别不同类型中尺度结构的方法，会为给定网络产生不同的划分。或许更有意思的是，针对识别某一种特定类型中尺度结构而设计的不同算法，即便应用于同一个网络，也可能产生相当大的差异。一个常被研究的实证案例是空手道俱乐部（KC）网络，这是一个由某体育俱乐部34名成员组成的友谊网络；由于成员之间发生矛盾，该俱乐部后来分裂成了两个新的俱乐部 [5]。尽管现有文献一再得出一个由两个凝聚群体构成的划分，其节点归属与俱乐部分裂所导致的划分类似 [6, 7], , 但对于该网络所检测出的"最优"划分，结果仍存在差异；这种差异不仅体现在节点的社区归属方面 [6] , 也体现在所识别出的社区总数方面 [8–10]。此外，还可以检测出其他类型的中尺度结构，作为对KC网络的合理解释 [11], , 其中包括由领导者和追随者构成的核心—边缘型结构。

真实网络中对中尺度结构的不同解释（例如 KC 示例）促使我们进一步探究这一尺度上的歧义；造成结果相互矛盾的原因，或许在于多个性质迥异的"真实结构"和划分共同决定了网络的生成过程及其中尺度构型 [12]。事实上，近年来随机块模型（SBM）在中尺度网络描述中的应用日益广泛，相关研究强调，应探究模型所返回的整个划分空间的变异性，而不是强行从划分分布中达成全局共识（即通过最大化某个目标函数，从众多划分中选出一个）[11]。在中尺度变异性这一背景下，另一个重要现象是可检测性极限：当信号强度不足时，已知结构将无法再被检测到。研究表明，由随机块模型生成的网络中存在相变，这正是可检测性极限出现的原因 [13, 14]。

* lena.mangold@cmb.hu-berlin.de，她/她/她的

总的来说，中尺度变异性既可能源于网络生成过程本身固有的歧义，也可能源于生成式块模型的随机歧义。本文主要关注在存在歧义的情况下评估中尺度结构检测的准确性。我们选择（1）从 KC 示例入手，因为它是同时具有核心—边缘结构和双社区结构（两个规模相同的社区）的最简单构型之一；（2）采用随机块模型作为理论框架，因为它能够容纳多种中尺度结构，不仅包括簇，也包括上述两种结构，而且不仅可用于生成中尺度结构，还可用于检测中尺度结构。更具体地说，我们希望探究：当我们在中尺度结构中隐式引入一定程度、可测量的歧义时，随机块模型检测特定结构的能力如何。我们提出一种生成式基准模型框架，即随机交叉块模型（SCBM）。该模型允许将两个性质迥异的划分植入同一网络，从而在其中尺度结构中内置这种歧义。借助这一框架，两个划分分别通过块矩阵定义，这些矩阵指定块内及块间的连通性。与植入划分模型 [15] 类似，这一设定便于分析不同植入中尺度结构强度下的随机块模型（或任何社区检测算法）。我们利用该框架将两个性质迥异的划分植入合成网络，并尝试用两种随机块模型变体将它们恢复出来。借此，我们分析随机块模型检测网络中两个相互竞争结构的能力，既评估分别恢复各个划分的程度，也评估成功检测到两种划分共存的能力：也就是说，在给定图的推断划分后验分布中，出现两个隐式植入结构。

本文结构如下。第二节介绍网络中对中尺度结构的相互矛盾的解释（II A）、块结构歧义与可检测性之间的相互作用（II B），并概述有关随机块模型（II C）和中尺度结构生成式基准模型（II D）的现有文献。随后，我们在第三节介绍模型框架，涵盖双分区情形下模型的推导，以及网络中边放置生成过程的两种变体。第四节通过一组模拟展示模型的一种应用场景，并在第五节讨论模拟结果。第六节总结主要结果，并简要探讨未来可能开展的工作，以扩展模拟和模型本身。

II. 背景

A. 社区检测与划分景观

社区检测通常采用聚类视角，关注凝聚型社区，即同一组中的节点彼此之间的连接比它们与其他组节点之间的连接更密集[16], ，而非更广泛意义上的其他类型的中尺度结构 [2]。用于识别社区结构的相应方法旨在针对特定类型的数据和网络发挥最佳效果 [12] ，许多作者也选择了不同的途径，以优化最可能的划分（不同方法的综述见 [17] ）。因此，现有算法都以多种度量为基础，例如模块度 [7], 、谱性质 [18], 、生成模型 [19], 、介数中心性 [6], 或信息论方法 [20], ；这只是采用不同方法的算法结果多样化的原因之一。

类似地，核心—边缘结构的检测——即将网络划分为连接紧密、凝聚力强的核心群体和连接稀疏的边缘群体，这一概念最初由 Borgatti 和 Everett [21]严谨地提出——也有多种研究方法 [22],，包括基于边密度 [3, 21, 23], 、路径长度 [23, 24]和生成式网络模型 [4, 25] 的方法。一些研究还探讨了社区结构与核心—边缘结构以嵌套方式共存的情形，尤其是社区内部呈现核心—边缘结构的情况 [3, 26–28]。

类似节点聚合任务所采用的方法多种多样，这些方法返回的划分自然也各不相同。尽管 20 世纪的大多数社区检测算法（从团到 k-核，再到 CONCOR [29–31]），以及在 2000 年代流行的 Girvan-Newman 算法 [32] ，都是确定性的，但 Louvain [9] 和随机块模型（SBM）[19, 33, 34]等方法的出现进一步加剧了这一挑战，因为这类方法的结果本质上具有非确定性。在这种情况下，即使同一种社区检测算法多次运行所得的结果，也未必能清楚地表明共识划分是什么；因此，除了上述模型选择问题之外，还会产生划分选择问题。现有研究通过在划分分布中寻找某种共识来确定“最优”划分，例如对同一算法多次运行的结果取平均 [35, 36]。

然而，当多个接近最优的划分在性质上彼此不同，这类寻求共识的方法就会遇到问题；这表明，我们需要考虑多个局部共识划分，因为它们可能为当前网络结构提供不同但可能性相近的解释 [11]。Peel 等人 [12], 也探讨了存在多个局部最优划分的问题。他们指出，许多现实世界网络都有多个合理（可能性较高的）

划分，不同的节点元数据集合可能与网络结构的不同方面相关。这些近期结果表明，如果在生成过程中容纳多种真实标签，那么某些社区检测方法所具有的随机性实际上不仅是需要应对的问题，也是这些方法的一项特性。在本研究中，我们提出一种模型，它可以接受两个划分的任意组合，包括与嵌套核心—边缘对有些相似的结构（如第 IV 节所示），同时又不局限于这些特定结构。我们还希望通过提出一种能够确保与两种不同植入结构相一致的模型，拓展现有研究，从而探索连接模式由多个块结构所诱导的网络（即多种结构共存），以及这两种结构被检测到和可检测到的程度。

B. 社群可检测性与歧义

近期研究开始关注分析划分分布变异性的重要性，这促使我们探索一种我们称之为中尺度模糊性的现象。Kirkley 和 Newman [37]旨在合成网络的中尺度上引入（并检测）他们所称的“歧义”，以识别能够代表整个划分景观中不同区域的典型网络划分。通过在随机块模型中的块之间指定特定的边概率来定义这种歧义（关于随机块模型的综述，见第二节 C），他们展示了其模型能够检测一组代表性划分，从而识别所引入歧义的不同方面。

这个关于歧义的具体例子引出了一个问题：一方面是真正存在歧义的块结构，另一方面是阻碍算法正确检测中尺度结构的微弱或含噪信号，两者之间有何区别？随机生成的随机块模型系综可能会表现出一定的块结构变异性，但如果只植入一个“真实标签”划分，那么定性上不同的恢复划分究竟源于真实的歧义，还是仅仅源于可检测性问题？此外，对于那些不可能由植入模型生成的划分（即落在具有指定参数的可能网络分布之外的划分），检测到它们应被视为划分算法的失败，还是成功恢复了歧义？

为了区分中尺度上某类歧义的正确恢复，与社区检测算法因噪声“模糊”了块结构信号而无法识别真实划分，我们需要了解社区检测中已被证明存在的可检测性相变。总体而言，研究表明，块结构在网络中的可检测性取决于网络的总体密度、各块连通性的差异以及块的数量（详见 [38, 39]中的综述）。当网络中存在结构信号，但信号过弱或噪声过大时，社区检测算法便无法识别此类结构。当潜在社区的结构痕迹在实际网络中不够明显时，算法会在某个相变点将网络误判为随机图。关于网络中模块的可检测性，先前的研究通过解析方法和启发式方法均表明，此类相变确实存在 [40]，并确定了其位置；其中尤以谱社区检测方法 [14] 和采用贝叶斯最大似然的方法 [41]为代表。早期关于相变的大部分研究都聚焦于传统随机块模型（度服从泊松分布）的对称情形。此后，其他研究者考察了节点度分布异质的网络，指出这类情形下也存在相变 [42] ，并证明在模块度最大化中，网络的异质性有助于社区检测 [43]。许多研究致力于评估具有明确结构（尽管有时含噪声）的社区检测相变；另有研究表明，核心—边缘结构中不存在此类可检测性阈值 [4], 。然而，鲜有研究关注随机块模型在特意引入一定程度歧义时的检测极限。进一步探索这一问题似乎尤为重要，因为社区检测方法主要用于现实世界网络，而这类网络往往具有更多“歧义”，并且已被反复证明可能存在多个局部最优解（见上文）。如前所述，实证网络还被证明可能同时具有不同类型的中尺度结构，这使问题更加复杂。

总体而言，区分某些块结构的可检测性问题与多种不同真实结构意义上的“真正”歧义，这一问题颇具挑战性；据我们所知，它仍是一个开放的研究问题，并由此引出如何描述这类歧义的问题。由于缺乏明确的定义，我们从现在起将歧义 界定为同一网络中多个植入划分同时存在的情况。我们将隐含结构同时存在的这种情况称为结构的共存 ，此时网络的连通性与上述结构的连通性始终同时保持一致。

C. 随机块模型

在这项工作中，我们从两个方面利用随机块模型的特点。一方面，我们使用随机块模型来生成具有预设中尺度结构的合成网络。另一方面

，我们通过将随机块模型拟合到合成图，探讨网络中尺度结构的歧义问题，从而将随机块模型用作检测中尺度结构的一种方法。随机块模型等贝叶斯推断方法具有随机性，因此尤其适合探索划分分布。

随机块模型起源于数学社会学，其建立在节点相似性这一概念之上：节点通过所属块所呈现的等价连接模式体现相似性。由此，人们创造了“块建模”一词，用来指将此类节点分组[44–46]。后来，这类等价关系早期较为严格的定义被放宽，形成了“随机等价”的概念：如果节点以相同的概率连接到其他节点集，那么这些节点便满足随机等价 [33]。这一后续研究也首次提出了随机块模型——这类生成模型先将节点划分为块，再依据各节点的块归属，以相应概率在节点对之间添加边，从而生成网络（或整个网络分布）。

因此，随机块模型是简单随机图模型的扩展：前者的边概率按块对确定且保持恒定，而非在整个网络中都相同。随机块模型的参数包括：(a) 一个块归属向量，其各项指明每个节点的块归属；(b) 一个方形连通性矩阵，其大小等于块的数量，矩阵元素表示相应块之间（或对角线元素所对应的块内部）存在连接的概率。自首次出现以来 [33, 44–46], ，随机块模型已被反复改造，作为基准模型，将社群检测问题视为推断问题来处理 [19, 34, 47]。它在理论和应用网络科学研究中日益流行，部分原因在于其灵活性；这种灵活性源于对节点相似性——即属于同一块或社群的含义——相对宽泛的定义。其基本思路是，为进行块结构检测，可以“逆向”运用随机块模型的生成过程：利用统计推断方法将随机块模型拟合到网络数据，从而恢复模型参数（本质上即块归属），以最可能地解释网络的生成过程。

众所周知，在“传统”的随机块模型中，两个节点之间边的放置仅取决于它们所属的块。因此，它无法体现许多真实网络的结构，因为这些网络往往具有异质的节点度分布。一种对度异质性进行建模的方法是在随机块模型中加入“度修正”，使边的放置还取决于每个节点各自的度。将度校正版本作为检测块结构过程所假设的生成模型，显著提升了随机块模型在真实网络中识别社群结构的能力 [19]。此后，人们开发了更多随机块模型扩展，包括层次型 [48], 、重叠型 [49, 50] 和多层型 [51] 变体。研究表明，与度校正版本相比，其中许多模型在某些类型的网络上具有更高的拟合优度，甚至优于度校正版本——版本。另一些研究则利用随机块模型的灵活性（即其检测不同类型中尺度结构的能力），展示了真实网络中核心—边缘结构的多样性 [25]。现有的随机块模型文献大多侧重于恢复单一划分，即通过优化某种模型选择准则来识别划分；不过，近期一些研究已不再局限于单一划分方法，并强调有必要考察随机块模型所得到的整个划分景观 [11, 12, 37]。在这项工作中，我们希望为随机块模型文献的这一特定分支作出贡献，着重探讨单个网络中存在多个植入结构及其可检测性。

在随机块模型文献中，有一类现有研究与植入多个真实结构这一情境尤为相关，即混合成员随机块模型（MMSBM） [49]。该模型允许节点以不同程度归属于多个块，这种归属关系通过分配给每个节点的混合成员向量来表示，其中的元素表示该节点属于不同块的概率。因此，在 MMSBM 中，节点可能同时体现多个块的连接模式，这使其与我们植入多个划分的情形相关。MMSBM 与我们的方法在概念上有一定相似之处，而且在某些情况下可以证明 MMSBM 等价于我们的方法；不过，后文将看到，基于 MMSBM 的解决方案存在显著局限，而我们的方法可以克服这些局限。

D. 生成式基准模型

与生成过程未知的现实世界网络不同，合成网络是植入特定结构并用作基准的自然选择，尤其适用于评估某种方法恢复各种中尺度结构（包括社区）的效果。此类基准框架能够将特定的中尺度结构“嵌入”合成网络中，并通过衡量预先定义的结构能否成功恢复，来测试算法的性能。通常，可以调整一个或多个参数，以探索算法的潜在极限，并模拟某些类型真实网络的特征。此类模型中最早的一种是 Girvan-Newman（GN）基准 [32], ：一个包含 128 个节点的网络，被划分为四个规模相同的组，并通过一个参数控制组间连通强度，该参数依据节点的外部（社区外）度设定。

为克服其一些缺点，例如整体缺乏灵活性、规模较小以及特征不切实际，Lancichinetti 和 Fortunato [52] 提出了一种基准模型，用于考虑异质的度分布和社群规模分布。其他现有基准模型则允许通过使用随机块模型来指定组内和组间的连通

性，例如已扩展至包括多层网络等其他特殊情形的植入划分模型 [15] [53]。一般而言，生成式基准模型旨在模拟实证网络的特征；尽管许多现有基准模型考虑了其中一种或几种特征，但中尺度结构中的歧义至今仍被忽视。我们的工作聚焦于这一特定方面，并通过一个能够生成具有多个内置真实结构的网络的框架，补充单一真实标签基准模型。

III. 模型框架

我们提出一种生成式网络模型，其参数旨在同时满足两个划分：通过在节点对之间连边，使生成的网络呈现出同时考虑这两个划分的块结构。我们将在后文讨论如何将这一框架扩展到包含两个以上划分的更复杂结构，但本文大部分内容聚焦于简单的双分区情形。

生成具有两种共存结构的网络，难点主要在于如何连接节点对，使其符合两种植入结构的连接模式。每对节点之间放置一条边的概率，必须取决于每个节点在每种植入结构中的块归属。因此，我们旨在设计一种生成过程，为所需的连接模式指定恰当的概率。一个显而易见的选择是受约束的混合成员随机块模型，其中节点以相同的概率属于两个块。然而，若要找到合适的归一化常数，使混合成员随机块模型的可加边概率与两种植入结构相一致，就需要额外计算；采用更简单的单成员随机块模型方法，将块重叠视为简单的块，则无须进行这些计算。下文首先概述该方法，解释基于混合成员随机块模型的方法的局限性，最后介绍随机交叉块模型（SCBM）作为替代方案。

如上所述，我们通常以 KC 示例为指导：在该示例中，既可以找到双社区（将网络划分为两个同配社群），也可以找到核心—边缘结构，而且这两种结构很可能共同解释网络的形成。图1a 和图1b 展示了同一个示例图，其中节点分别按照双社区划分和核心—边缘划分着色。

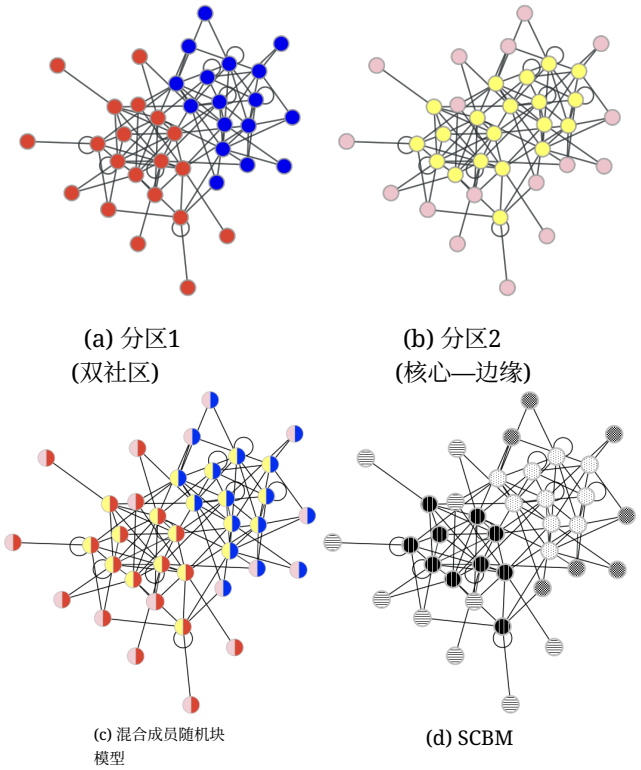
A. 正式概要

我们生成一个包含 N 个节点和 E 条边的网络，并在其中植入一组含有 K_p 个块的划分，记为 partition p。节点之间的边由大小为 $N \times N$ 的邻接矩阵 $\mathbf{Y}$ 描述。我们关注无向无权情形：当且仅当 $Y_{ij} = Y_{ji} = 1$ 和 j 之间存在一条边时，i；否则为 0。每个划分 p 中的块结构都定义为一个含有两个参数的随机块模型（SBM）：(a) 分配给每个节点 i、长度为 K_p 的块归属向量集合 $\mathbf{b}_i^p$；以及 (b) 大小为 $K_p \times K_p$ 的块矩阵 $\boldsymbol{\theta}_p$，其中对角元素（分别为非对角元素）表示块内（分别为块间）存在一条边的概率。注意，每个 $\mathbf{b}_i^p$ 都是一个独热向量，即仅有一个元素为 1（表示节点 i 在划分 p 中的块归属）、其余元素均为 0 的二元向量。我们用 $\mathbf{B}^p$ 表示期望边数矩阵，该矩阵对于含有自环的图，其元素为 $\theta_{p_{rs}} n_{p_r} n_{p_s}$，其中 n_{p_r} 是划分 p 中块 r 内的节点数。注意，由于我们生成的图是无向图，块矩阵是对称的；并且为便于计算和记号表示，对于 $r = s$，$\mathbf{B}^p$ 的元素表示期望边数的两倍。

为便于说明，我们聚焦于每个划分包含 $P = 2$ 个块的简单情形，并将两个隐式植入划分对应的块矩阵定义为 $\boldsymbol{\theta}_1 = \{\theta_{1_{rs}}\}$ 和 $\boldsymbol{\theta}_2 = \{\theta_{2_{rs}}\}$。在第三节 C 中，我们概述了如何将我们的框架扩展到更复杂的划分组合。

B. MMSBM 表述

给定植入两个不同划分所需的参数，我们需要定义一个生成过程，使其生成的网络满足两者的连接模式。一种可行的方法是通过适当的归一化，将双划分情形表述为混合成员随机块模型（MMSBM）的一个特例。在原始的混合成员随机块模型表述中 [49], ，每个节点都以一定概率属于所有潜在群组，该概率由每个节点特有的混合成员向量表示。两个节点之间是否存在一条边，取决于这两个节点所属的块；对于每一对节点，都会根据每个节点的混合成员向量反复抽取其块归属。因此，节点会继承多个块的连接模式，而多个块“重叠”部分的期望密度，则是各块单独密度的加权平均 [50]。为了将双真实划分问题表述为混合成员随机块模型，我们将两个植入划分中的块视为四个潜在块，但对生成过程施加约束，强制某些重叠部分为空。具体而言，我们需要：(a) 约束混合成员向量，使节点属于某些特定块组合的概率为零；(b) 以一种能使



(a) 分区1
(双社区)

(b) 分区2
(核心—边缘)

(c) 混合成员随机块模型

(d) SCBM

图1：一个示例图，节点根据其块归属着色，展示了不同的划分表述方式：植入式双社区分区（a）、植入式核心—边缘分区（b）、将节点以相等概率归入两个块的双划分问题 MMSBM 表述示意图（c），以及SCBM方法（d）。

生成网络的连通性与 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$ 给出的两个植入划分的连通性一致的方式，指定块内和块间的边概率。图 1c 中的示意图通过展示按节点在两个植入划分中的块归属着色的节点，对这些约束进行了可视化。

事实证明，对 MMSBM 引入的加性块概率进行必要的归一化并非易事。具体而言，需要为每种块对组合确定一个归一化常数，而这需要求解一个欠定方程组。正如我们在附录A中详细说明的，该方程组的最小范数解（可通过最小二乘求解器获得）在某些植入结构组合下为负值。这意味着，不同的植入结构需要采用不同的近似解方法，而这很可能会对植入结构的探索产生意想不到的影响。

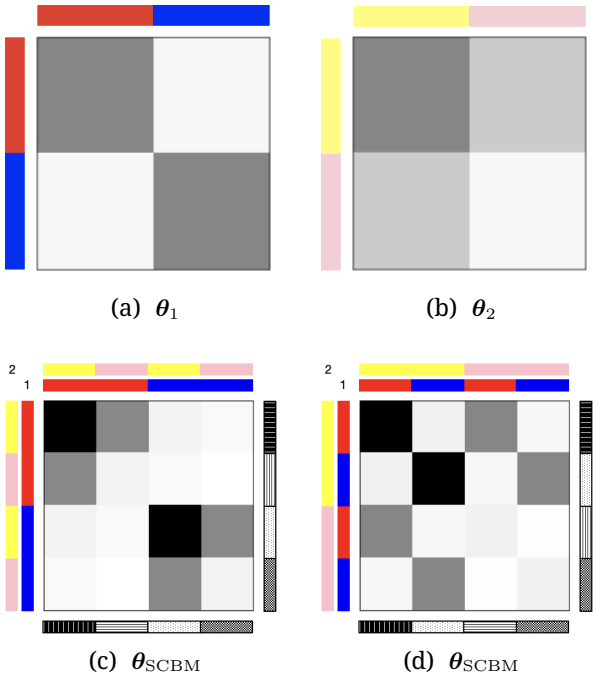
为了能够灵活探索足够广泛的结构组合，我们因此提出了一种替代基于混合成员随机块模型的方法：事实证明，我们可以绕过求解归一化常数所涉及的额外步骤：将双分区问题表述为单成员随机块模型，并以乘法概率取代混合成员随机块模型所设定的加性概率。我们提出的模型不再通过节点在两个块中的混合成员关系来生成网络，而是对问题进行简化，将各个植入划分中块之间的重叠视为随机块模型中的块。

C. 随机交叉块模型

在随机交叉块模型（SCBM）中，一个节点会被分配到多个划分中其所抑制的块的组合，这种组合称为交叉块。我们用 $\boldsymbol{\Gamma} = \{(r, r')\}$表示交叉块的集合，其中 r 和 r' 分别是分区1和分区2中的块。于是，我们可以将问题重新表述如下：为了在单个网络中植入两个划分，我们生成一个网络，并以符合块矩阵 $\boldsymbol{\theta}_1$和 $\boldsymbol{\theta}_2$预期密度的方式，在其中显式植入一个由$K_{\mathrm{SCBM}} = K_1 K_2$ 个交叉块构成的交叉划分。图1d展示了由图1a和图1b所示划分形成的交叉划分中的四个交叉块。请注意，交叉划分——即将网络显式划分为 K_{SCBM} 个交叉块——不同于分区共存；后者指网络与分别包含 K_1 个块和 K_2 个块的两个隐式植入结构的连通性相一致这一性质。

因此，我们的生成网络就简单地变成了一个标准随机块模型，其中节点 i 与节点 j 之间的边概率，完全由节点 i 所属交叉块 u 与节点 j所属交叉块 v 之间的节点概率决定。为了生成最终网络，我们直接根据向量$\mathbf{b}_i^p$，为每个节点 i 创建独热交叉块成员向量$\mathbf{b}_i^{\mathrm{SCBM}}$。在放置边时，我们需要通过定义$\boldsymbol{\theta}_{\mathrm{SCBM}}$ 和$\mathbf{B}_{\mathrm{SCBM}}$来确定交叉块之间及其内部的连通性。如上所述，元素 θ_{uv} 表示交叉块 u 与 v 之间的边概率，而B_{uv} 表示此类边的平均期望数量，其中我们省略了下标。块矩阵的构造方式确保边概率与每个植入分区 p的块矩阵 $\boldsymbol{\theta}_p$ 中的元素一致。这种交叉划分方法与我们问题的混合成员随机块模型表述之间的决定性差异在于：我们用归一化的乘法概率替代了混合成员随机块模型产生的加性概率。换言之，对于每个交叉块，边概率通过将导致重叠的原始块中的边概率相乘，并进行适当归一化来计算。

在图2中，我们展示了两个块矩阵示例 $\boldsymbol{\theta}_1$和 $\boldsymbol{\theta}_2$ 以及由此得到的交叉块矩阵 $\boldsymbol{\theta}_{\mathrm{SCBM}}$；请注意，图2c和图2d展示了同一交叉块矩阵的两种不同排序方式，



(a) $\boldsymbol{\theta}_1$ (b) $\boldsymbol{\theta}_2$ (c) $\boldsymbol{\theta}_{\text{SCBM}}$ (d) $\boldsymbol{\theta}_{\text{SCBM}}$

图 2：两个植入划分（a 和 b）的块矩阵可视化及其生成的交叉划分（c 和 d）。请注意，c 和 d 等价但排序不同，如彩色条带（划分 1 和 2 中的块）和图案条带（交叉块）所示。矩阵元素根据块矩阵中的预期边密度着色，其中黑色元素表示最大边概率，白色元素表示边概率为 0；请注意，颜色保持一致，因此可在这三张图之间进行比较。

以强调该交叉划分与 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$都一致。此处展示的 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$ 的具体选择也决定了图1a和图1b中展示的划分，因此这些交叉块矩阵与图1d中展示的图相对应。

区块大小相等 首先，我们关注植入划分和交叉划分中的区块大小都相等的情况，因此 $n_r = n$ 对所有区块 r和 $2n = N$均成立。我们允许存在自环，因此对于 $p \in \{1, 2\}$，有 $B_{p_{rs}} = n_r n_s \theta_{p_{rs}} = n^2 \theta_{p_{rs}}$。我们用 $\boldsymbol{\nu} = \{\nu_u\}$ 表示跨块大小向量，并将交叉块约束为大小相等：对于 $u \in \boldsymbol{\Gamma}$，$\nu_u = \nu$，因此 $2\nu = n$。我们通过将两个划分中各块的概率相乘，并进行适当归一化，来构造交叉块内部及交叉块之间的边概率。具体而言，我们将 SCBM 的块矩阵定义为 $\boldsymbol{\theta}_{\text{SCBM}} = \{\theta_{uv}\} = x\{\theta_{1_{rs}}\theta_{2_{r's'}}\}$，其交叉块为 $u = (r, r') \in \boldsymbol{\Gamma}$ 和 $v = (s, s') \in \boldsymbol{\Gamma}$，其中常数 x 需要以这样的方式选取：使得$\mathbf{B}_{\text{SCBM}}$ 中某些交叉块对内部及之间的期望边数之和，分别等于$\mathbf{B}_1$ 和$\mathbf{B}_2$，中的相应元素，从而使生成网络的连通性与划分 1 和 2 的连通性一致。在区块大小相等的情况下，对于$x = 1/\rho$，这一条件成立，其中 $\rho = 2E/N^2$ 是具有自环的网络的总体期望边密度（推导见附录 B）。期望边数矩阵的元素记为矩阵元素 $B_{uv} = \nu^2 \theta_{uv}$；其中我们省略了 B 和θ中的 SCBM 下标。

可变块大小 在块大小可变的其他更一般情形下，只有在引入其他约束（例如，两个植入划分的所有块之间边概率相等，即植入一个随机图；见附录 C）时，才存在满足两个植入划分的常数 x。要确定块大小可变的非平凡划分对应的 $\boldsymbol{\theta}_{\text{SCBM}}$，我们需要找到一组归一化常数 $\{x_{uv}\}$，每个交叉块对对应一个，使得 $\boldsymbol{\theta}_{\text{SCBM}} = \{\theta_{uv}\} = \{x_{uv}\theta_{1_{rs}}\theta_{2_{r's'}}\}$。为了以 $\boldsymbol{\theta}_{\text{SCBM}}$ 生成与 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$ 一致的网络的方式求得 $\{x_{uv}\}$，我们必须再次求解一个欠定方程组，并可计算其最小范数解。若该方程组存在解，则此解唯一且始终存在（见附录 D）。在这种情况下，在第 IV 节的整个参数空间内，最小范数解均给出一组完全为正的 $\{x_{uv}\}$；虽然我们的模拟考察的是块大小相等的情形，但我们在此表明，使用SCBM方法也可以采用更一般的设定。

多个块与划分 在第 IV 节的模拟中，我们重点研究上述简单的双分区情形，以展示两个“真实标签”划分的可检测性问题。不过，在某些情况下，探索包含更多共存划分和/或更多块的复杂结构也可能很有意义。尽管全面探索此类情形超出了本文的范围，我们仍简要概述对所述简单情形的一种扩展。事实上，保持划分数为 $P = 2$，但在一个或两个划分中植入两个以上的块，对于区块大小相等的情形，可以直接使用相同的归一化常数；而对于可变块大小的情形，则需要求解规模更大的方程组。由于植入两个以上的划分也可以重新表述为递归地植入一组组双分区，直到得到最终的交叉划分，因此这种做法同样可行。在区块大小相等的情形下，SCBM 块矩阵可推广为$\boldsymbol{\theta}_{\text{SCBM}} = 1/\rho^{p-1}\{\prod_p \theta_{p_{rs}}\}$。在可变块大小的情形下，则需要考虑划分数和/或块数较大时可能出现的可扩展性限制，因为交叉块的数量为 $K = \prod_p K_p$，而我们需要求出 $K(K+1)/2$ 个归一化常数。求最小范数解的最小二乘法需要计算系数矩阵的伪逆；对于大型 $n \times m$ 矩阵，这一计算的开销可能很高（计算复杂度约为 $O(n^2 m)$）。因此，对于非常复杂的植入划分组合，必须考虑计算时间和内存方面的限制。

D. 生成式随机块模型

我们根据连接矩阵生成最终网络，采用“传统”的随机块模型 [33]„ 该模型使用矩阵 $\boldsymbol{\theta}_{\mathrm{SCBM}}$ 以及交叉块成员向量$\mathbf{b}$ 来确定两个节点之间是否存在边。具体而言，我们将对每一对节点(i, j)独立随机地放置一条边，其概率为 θ_{uv}，其中 $u, v \in \mathbf{\Gamma}$分别是 i 和 j 的交叉块。因此，我们以 $Y_{ij} \sim$ 伯努利分布$(\mathbf{b_i}^T \boldsymbol{\theta}_{\mathrm{SCBM}} \mathbf{bj})$对 i 与 j 之间的交互取值进行抽样。在此版本的随机块模型中，$\mathbf{B}_{\mathrm{SCBM}}$ 中的期望边数平均而言得以满足。

我们还考虑微正则随机块模型 [34], ，在该模型中，给定矩阵 $\mathbf{B}_{\mathrm{SCBM}}$ 的各元素（取整后）都得到精确满足（而非仅在平均意义上满足），并且该模型基于构型模型 [54]。具体而言，我们考虑这种微正则随机块模型的度修正扩展；在该扩展中，两节点之间出现边的概率不仅取决于连接矩阵的元素，也取决于给定的度序列或度分布。已有研究表明，这种随机块模型变体生成的合成网络具有更接近实证网络的特征，其中块内度变异的类型也更可能出现在真实网络中[19]。给定度序列 $\{k_i\}$，其中 k_i 表示节点 i 的度，该方法会为节点 i 分配 k_i 条半边，然后在网络中随机选取两条半边（允许自环）并将它们连接起来，同时确保给定的块内和块间边数符合预期。不过，$\mathbf{B}_{\mathrm{SCBM}}$ 的元素可以是实数，因此在生成网络时必须进行取整。这会导致（隐式植入的）期望边数矩阵 $\mathbf{B}^p$ 与生成的网络之间出现细微差异，涉及边总数以及块内和交叉块之间的边数。这也意味着，只有在 $\mathbf{B}_{\mathrm{SCBM}} \in \mathbb{Z}^{N \times N}$ 时，给定的度序列才能得到精确满足；在我们的模拟中，我们从幂律分布中抽取节点度，而不要求度序列精确匹配。请注意，在下文的模拟中，我们以这种方式生成的图中，任意一对节点之间最多只有一条无权边；取消这一约束并改为生成多重图则很简单。

今后，我们将依据传统随机块模型生成网络的模型版本称为正则模型，以区别于后一个版本，后者称为微正则模型。这样可以避免在记号上混淆：我们使用哪些随机块模型变体来生成 网络，与使用哪些变体来推断划分。

IV. 模拟

现在，我们通过生成一组隐式植入了两个划分的网络，来探究随机块模型恢复两个内在真实结构的能力。显然，值得探索的双划分结构有很多；如前所述，我们关注的是动机示例 KC网络 中呈现的那类结构。对于该网络，推断出的划分的后验分布样本为中尺度结构 [11] 提供了多种合理解释；其中一个局部共识划分是著名的两派划分，即将网络划分为两个同配社群；另一个则是类似核心—边缘结构的领导者—追随者划分。我们关注网络中这两类结构的恢复极限，因此依据我们的生成框架，在一系列合成网络的系综中植入类似结构。在模拟中，我们在一组图中分别构建双社区结构和核心—边缘结构，并对这些网络拟合两种不同的随机块模型变体，以推断每个网络的划分后验分布。最后，我们计算恢复的划分与植入划分之间的相似度，并展示由每种模型变体植入、由每种随机块模型变体恢复的划分结果。

A. 参数

这里我们关注区块大小相等的情况。对于这种情况，$\boldsymbol{\theta}_{\mathbf{SCBM}}$ 中的乘性概率可以简单地通过常数 $x = 1/\rho$ 进行归一化，因此无需依赖最小范数解（见第 III C 节）。我们植入具有 $N = 400$ 的网络，并针对不同的平均度 $c = 5$，10，20 运行三组模拟。因此，三组模拟之间的总期望边密度不同，但在每组内部保持不变。这样可以确保我们所观察到的两个植入划分在可检测性方面的任何差异，都源于网络内部边的不同分布（我们将通过改变区块连接度来实现），而不是总边密度的差异。节点被均匀随机选取并分配到各个块中，以创建块归属向量 $\mathbf{b}_1$ 和 $\mathbf{b}_2$。为了植入这两个划分，我们在公式 1a 和公式 1b 中定义对称块矩阵 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$，其中 $\beta = 2E/n^2$，E 为网络中的边总数。分区1（$\boldsymbol{\theta}_1$）配置双社区结构，其中参数 μ控制两个植入的等大小社区之间预期的块内与块间连接强度。分区2（$\boldsymbol{\theta}_2$）配置核心—边缘（核心-边缘）划分，其中核心内部的期望边密度以及边缘节点之间的期望边密度由参数 λ控制。请注意，分区2中的块间边概率以这种方式固定，因此对于 $\lambda \in [0, 0.5)$，始终有 $\theta_{2_{11}} > \theta_{2_{12}} > \theta_{2_{22}}$，这与

核心—边缘结构的常见定义一致，并使 $\mu = 0.5$ 和 $\lambda = 0.5$在块内和块间具有相等的概率。

$$\begin{cases} \boldsymbol{\theta}_1 = \beta \begin{pmatrix} 1-\mu & \mu \\ \mu & 1-\mu \end{pmatrix} & \text{(1a)} \\ \boldsymbol{\theta}_2 = \beta \begin{pmatrix} 1-\lambda & \frac{1}{2} \\ \frac{1}{2} & \lambda \end{pmatrix} & \text{(1b)} \end{cases}$$

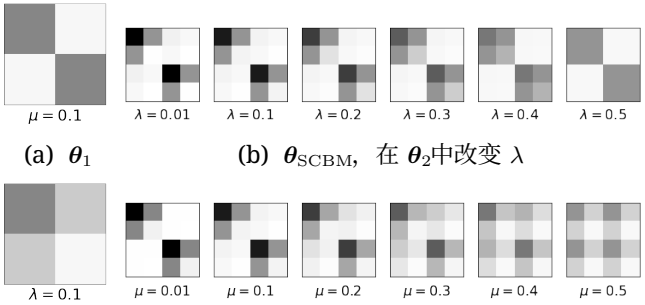
我们使用模型生成多组网络，分别将参数 μ 和 λ 从 0.01 扫描至 0.5，每次递增 0.01。注意，我们排除 $\mu = 0$，因为它会产生由两个连通分量组成的非连通图；同时，为了使两个维度保持对称，我们排除 $\lambda = 0$。

较小的 μ 会生成同配的社区结构（也就是说，大多数边连接块内的节点，块间连接较少），而当 μ 接近 0.5 时，生成的网络接近随机图。 λ接近零时，会生成“清晰”的核心—边缘结构：大多数边连接核心内的节点，边缘节点之间的连接较少；而 $\lambda = 0.5$ 则会生成随机图。图3展示了在固定 μ 并改变 λ，以及反过来固定 λ 并改变 μ 时， $\boldsymbol{\theta}_{\mathrm{SCBM}}$的表现。图3b 和图3d 中的块矩阵表明，当两个参数都较小时，交叉划分呈现出两个社区的嵌套结构，且各社区内部具有核心—边缘结构，这与网络中核心—边缘对的既有研究相似 [3, 26–28]。固定 μ 并增大 λ时，双社区划分的块密度保持不变，而核心—边缘结构逐渐减弱；当 $\lambda = 0.5$时，由于 $\boldsymbol{\theta}_2$ 此时定义了一个随机图，我们最终得到双社区划分。同样，固定 λ 并将 μ 增大至 $\mu = 0.5$时，我们最终得到核心—边缘结构（通过重新排列图3d 中 $\boldsymbol{\theta}_{\mathrm{SCBM}}$ 的行和列）。

请注意，我们使用 SCBM 显式生成的是一种简单的随机块模型，其交叉划分由交叉块矩阵 $\boldsymbol{\theta}_{\mathrm{SCBM}}$ 诱导——$\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$ 的连通模式均得到显式满足。当我们从生成矩阵中推断最可能的划分时，我们预期显式植入的交叉划分以及两个隐式植入划分 1 和 2 都能在一定程度上被恢复，其恢复程度可能随（λ, μ）空间而变化。

B. 生成图

为了评估度分布的差异，我们在一组模拟中使用正则模型，在另一组中使用微正则模型。因此，对于每个期望度 c，我们生成两组网络。在正则模型中，节点度遵循泊松度分布，块内和块间的边数在平均意义上得到满足。在微正则情形下，如第二节 D 所述，不同模型运行之间的边数不会发生波动，并且我们对节点度施加进一步的约束



(a) $\boldsymbol{\theta}_1$ (b) $\boldsymbol{\theta}_{\mathrm{SCBM}}$，在 $\boldsymbol{\theta}_2$中改变 λ

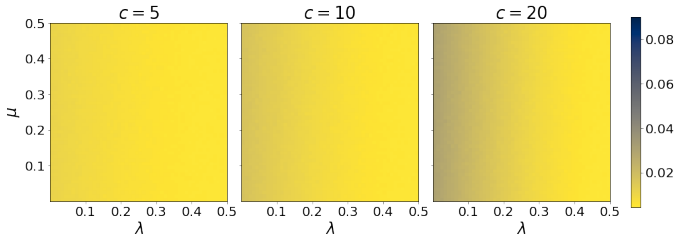
(c) $\boldsymbol{\theta}_2$ (d) $\boldsymbol{\theta}_{\mathrm{SCBM}}$，在 $\boldsymbol{\theta}_1$中改变 μ

图3：示意分块矩阵 $\boldsymbol{\theta}_{\mathrm{SCBM}}$ 以及在 a 和 b 中分别固定参数 μ 并增大 λ 时的情况，或在 c 和 d 中类似地固定 λ 并改变 μ 时的情况 $\boldsymbol{\theta}_2$ $\boldsymbol{\theta}_1$。

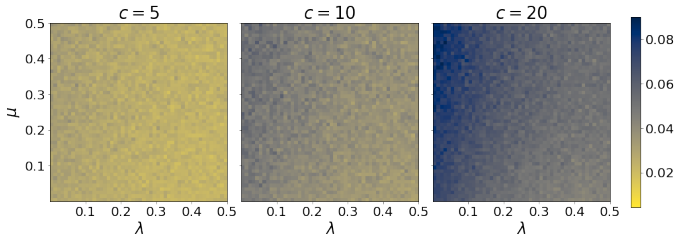
，从幂律分布中对节点度进行抽样，其指数为 $\gamma = 3$。我们采用软约束，即最终网络不必与给定的度序列完全匹配，而只需在平均意义上匹配。附录 F 总结了由于舍入误差，生成图中的边数相对于植入边计数矩阵所产生的小幅偏差。在这两种情况下，对于每个（λ, μ）对，我们都生成 8 个网络，以考虑生成过程中的潜在波动。

在尝试恢复两组图中的植入划分之前，我们先探究正则模型和微正则模型这两种不同生成过程，在不同的（λ, μ）参数对下，为网络引入的结构特征。图4显示，不出所料，在整个（λ, μ）取值范围内，微正则模型生成的图具有最高的度方差，因为我们对异质幂律度分布进行采样。正则版本生成的图，其度方差则要低得多。尽管如此，使用正则模型时，较低的 λ 会使网络呈现更高的度异质性，因为此时我们施加了强烈的核心—边缘结构（缩放后的图4顶行见附录中的图11）。

尽管微正则模型生成的图总体上呈现出最高的度异质性，但图 5 显示，与正则情形相比，其核心块和边缘块在节点度分布方面更为相似，这一相似程度通过核心节点与边缘节点的度分布之间的 Jensen–Shannon 距离 [55] 来衡量。在正则情形下，整个网络以及各块内节点的度均服从泊松分布。不过，为了适应核心—边缘结构，边缘中的分布均值低于核心中的均值；在泊松情形下，这会导致两种分布的重叠相对较小 [4]。在微正则情形下，我们通过度分布引入度异质性——



(a) 正则模型。



(b) 微正则模型。

图 4：具有不同期望度 c 的图的平均归一化度方差。

因此，植入 CP结构 并不会使块度分布产生同样的差异。

这意味着，如果我们简单地将在正则模型生成的图中度高于平均度的节点分配到核心，而将度低于平均度的节点分配到边缘，那么相比于微正则模型生成的网络，我们能够更准确地找回被标记的节点 [4], 其中“准确地”是指由 $\mathbf{B}_2$ 植入的核心—边缘区块，如（1b）所定义。微正则情形下度分布的异质性可能会产生不同于植入模式的其他类型的核心—边缘模式。（关于以这种方式对两个模型中正确分类的节点数量进行比较，以及两个示例图中核心节点和边缘节点的度分布，请参见附录 F。）

因此，值得进一步探讨在生成过程中将度分布指定为额外参数可能带来的影响，尤其是除植入的连接矩阵和节点块分配之外，额外指定度分布的影响。我们预计，正则模型生成的图会呈现出与我们明确植入的结构（即块连接矩阵）密切相关的结构；但在微正则情形下，对节点度施加的约束可能会隐式引入其他结构。例如，度分布的异质性可能导致一些节点群体在与网络其余部分的连接方面表现出相似性，具体而言，这些节点的连接数量相似。这些群体之间的分界线可能与我们植入的块结构所设定的分界线并不一致，从而可能使模型识别出植入结构以外的结构。这些由生成过程引入的度异质性差异

因此，很可能会影响随机块模型变体在 (λ, μ) 平面的不同区域中恢复植入划分（及其共存）的程度；第 V 节证实了这一点。

C. 相似性度量

为了量化植入划分与恢复划分之间的相似性，我们计算最大划分重叠 $\omega(p, q)$，即一个划分 p 中被分配到另一划分中同一假定正确区块的节点所占比例 q [11]。具体而言，我们寻找 q 的组标签的双射 $p' = \zeta(q)$，使得 p' 和 p 中具有相同区块标签的节点数量最大化，从而得到

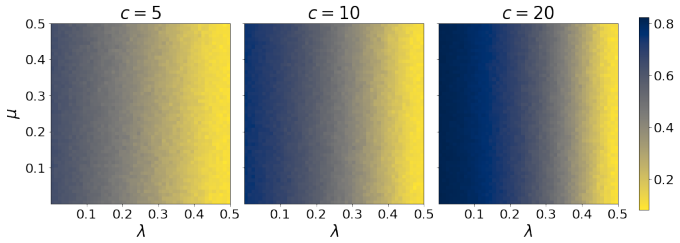
$$\omega(p, q) = \frac{1}{N} \max_{\zeta} \sum_i \delta_{p_i, \zeta(q_i)}. \tag{2}$$

具体而言，这是通过调用 graph-tool Python 库中的一个函数来求解两个划分的最大权二分图匹配问题 [56]，该函数基于 Kuhn-Munkres [57, 58]算法。请注意，此处我们使用的是归一化划分重叠度，其取值范围为 0 到 1。因此，$\omega = 1$表示两个划分中的所有节点都重合。请注意，如果比较的两个划分都包含两个块，$\omega = 0.5$ 就是下界，这意味着一半节点被正确分类，因此这两个划分不相关。如果一个划分包含两个以上的块，则可能出现 $\omega < 0.5$。

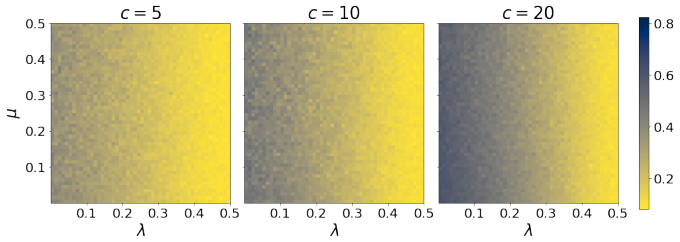
划分重叠度量是合适的相似性度量，因为与基于互信息等信息论度量相比，它更容易解释；而且它不受所比较的两个划分中块的数量和大小影响，这是某些成对计数方法（如兰德指数 [59]）存在的问题。我们通过使用约化互信息 [60]和信息变异计算一组模拟的划分相似度，来展示结果的稳健性；已有研究也表明，信息变异对于不平衡划分表现良好[61] ，相关结果见附录 E。两种相似性度量所得结果与使用划分重叠度量计算的结果相当。特别是，在 (λ, μ) 空间中，可检测性阈值的位置似乎完全相同（或至少极为接近）。

V. 结果

为推断生成图的划分，我们使用 graph-tool Python 库 [56]拟合了两种随机块模型变体（传统型和度修正型）。对于八个图中的每一个图，我们分别获得了 50 个划分的分布，因此对于 λ 和 μ的每种组合，总共获得了 400 个划分。为确保所选数量的

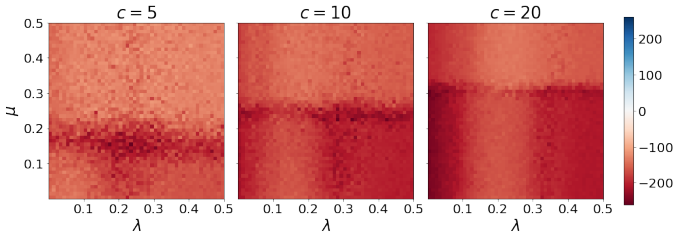


(a) 正则模型。

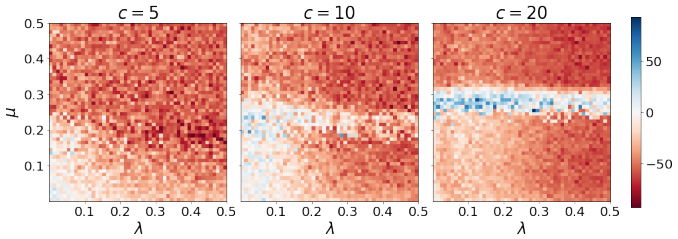


(b) 微正则模型。

图 5：期望度各异的图中，核心与边缘节点度分布之间的平均詹森-香农距离 $c_{\text{。}}$



(a) 正则模型。



（b）微正则模型。

图 6：对于具有不同期望度 c 的图，度修正和 NDC 模型类别的对数证据之差。负值（红色）表明 NDC 模型拟合得更好，正值（蓝色）表明度修正模型拟合得更好。

样本足以探索后验分布，我们使用每个（λ，μ）组合 4000 个划分运行了相同的模拟[1]（对四个图分别采样 1000 个划分），并发现没有定性差异。作为一种推断方法，以下我们将传统随机块模型称为 NDC（非度修正随机块模型），将度修正变体称为 DC，以避免与用于生成网络的模型（规范模型和微规范模型）混淆。最后，我们分别计算了两种植入结构的推断出的划分与植入划分之间的划分重叠，以及推断出的划分与植入的交叉划分之间的划分重叠。

A. 模型拟合

我们首先评估用于检测中尺度结构的两种随机块模型变体中，哪一种对我们生成的网络拟合得更好。我们计算模型证据的对数值，即对每次运行中的所有划分求和；其计算方法是用负的平均描述长度（对所有划分取平均）减去后验分布的熵 [34]。图 6a 表明，在整个（λ，μ）空间中，正则版本的 NDC 是更受青睐的模型；这并不意外，因为生成过程中边的放置与节点度无关。当我们使用微正则版本（在生成过程中确实考虑节点度）时

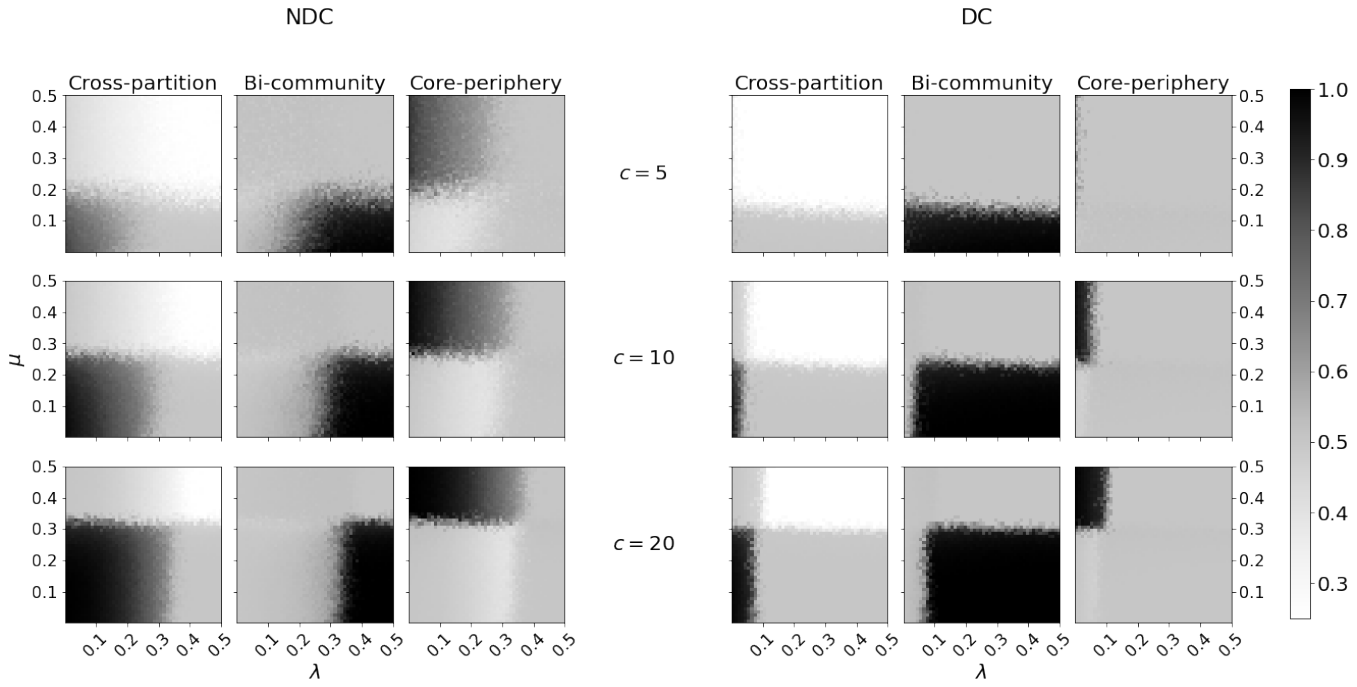
我们或许会预期，DC 能在整个（λ，μ）平面上更好地描述生成网络。然而，图 6b 表明事实并非如此：我们观察到一小片 λ和 μ 取值区域，在该区域内 DC 的模型证据更高；随着期望度增加，这一区域变得更加明显，并覆盖整个 λ 范围，但所对应的 μ取值范围却越来越窄。在其他所有区域，NDC 的模型拟合仍然更好。这表明，只有对于度分布高度异质、且具有一定强度双社区结构的网络，DC 更高的复杂度才是合理的；这种结构强度取决于 $c_{\text{。}}$就模型参数数量而言，在（λ，μ）平面上的其他所有位置，DC 似乎都给出了过于复杂的描述，尽管度异质性仍然很高。[2]

B. 植入结构的恢复

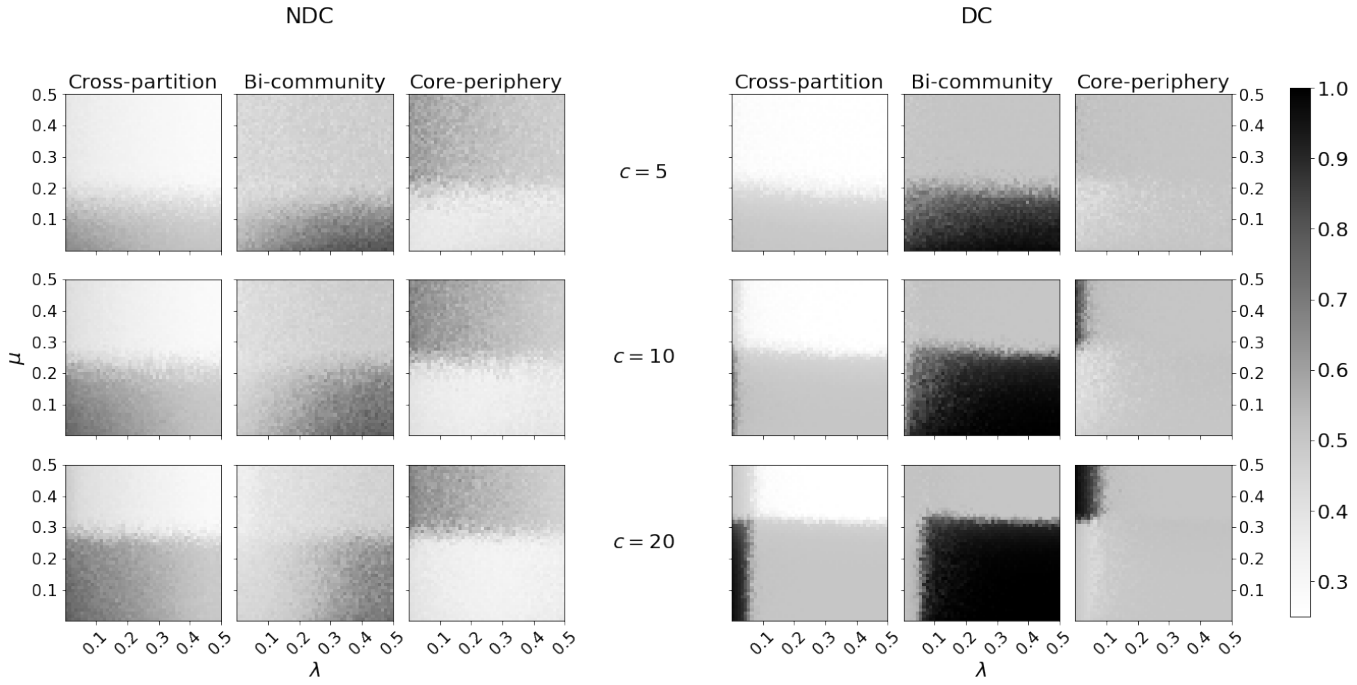
在图7中，我们展示了所有（显式和隐式）植入划分的平均划分重叠度 $\langle\omega\rangle$，这些划分位于 the

[1]对于由正则模型生成、具有 $N = 400$ 和 $c = 10$ 的图，以及使用非度修正随机块模型推断得到的划分。

[2]如果允许节点对之间存在多条边（使用 graph-tool Python 库 [56]），我们发现，在（λ，μ）平面的左下方还存在一个区域——大致对应于 $\lambda < 0.2$ 和 $\mu < 0.33$ 的取值——在该区域中，度修正模型更受青睐。一种可能的解释是，此情况下引入了更大的度方差，因此度修正模型较高的复杂度是合理的。由于这些多重图的划分恢复模式与我们在第 V 节讨论的模式相同，我们将分析限制于简单图。



(a) 正则模型。



(b) 微正则模型。

图 7：$\langle\omega\rangle$ 对于期望度各异的图，植入划分与后验分布中的划分之间的平均划分重叠度 c。

对于由两种模型生成的网络，每个（λ，μ）组合下推断出的划分的后验分布中，所有（显式和隐式）植入划分的平均划分重叠度。四个象限中最左侧的列显示推断出的划分与显式植入的交叉划分（$\boldsymbol{\theta}_{\mathrm{SCBM}}$）之间的 $\langle\omega\rangle$；中间和右侧的列显示对于隐式植入的双社区（$\boldsymbol{\theta}_1$）和核心—边缘（$\boldsymbol{\theta}_2$）划分，情况也分别相同。请注意，这里我们关注的是对各个划分的检测；关于分区共存的检测——即其程度——请参见第 VD 节的概述。

这两种隐式植入结构都会出现在给定生成图的推断出的划分的后验分布中。

正如预期，在（λ, μ）空间中，似乎存在一些明显的阈值，将恢复出交叉划分的区域与检测到两个隐式植入结构之一的区域区分开来。这些阈值的位置会随期望度、生成模型以及用于划分推断的随机块模型变体而变化。因此，关于两个隐式植入结构可检测性的问题，似乎与交叉划分的可检测性有关：由于这两个参数的值越低，交叉划分的信号就越强，因此交叉划分会一直被检测到，直到达到某个阈值。只有当交叉划分的信号弱到一定程度后，隐式植入划分才会更受青睐。

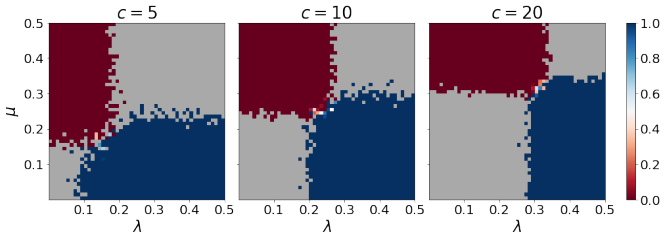
我们首先关注由正则模型生成的图中推断出的划分，在该模型中，节点度服从泊松分布。在图7a中，我们展示了植入划分与 NDC（左侧）和度修正（右侧）恢复出的划分之间的$\langle\omega\rangle$ 。这两种变体都经常能检测到双社区结构，直到 μ 达到某个阈值；随着 c 增大，该阈值也会升高。事实证明，这些阈值的位置大致符合现有文献中关于植入划分模型可检测性阈值的描述 [13, 14]。根据该研究，由边计数矩阵 $\mathbf{B}_1$ 定义的（1a）中的社区结构，当 $c > \frac{1}{(1-2\mu)^2}$ 时可检测；在我们的情形下，这意味着可检测性的阈值 μ_T 应位于 $\mu_T \approx 0.276$ ($c = 5$)、$\mu_T \approx 0.342$ ($c = 10$) 和 $\mu_T \approx 0.388$ ($c = 20$)。在我们的模拟中，两种变体都能检测到双社区，直到 μ 达到相近的值，该值略低于 μ_T。我们还观察到双社区检测的第二种阈值，这次沿着 λ 维度。对于较低的 λ 值，尽管 $\mu < \mu_T$，两种变体都无法检测到双社区划分，而是似乎找到了交叉划分。NDC 和度修正在 μ 维度上的阈值相近，但我们发现，NDC 的 λ 阈值远高于度修正。这意味着，与度修正相比，NDC 在 λ 的更高取值范围内都能恢复交叉划分，之后则会检测到双社区。 λ 阈值也会随着 c 增大而升高。

对植入式核心—边缘分区的检测也取决于网络的期望度。事实上，核心—边缘检测的阈值与上述双社区检测的阈值相对应：沿着 μ 维度，当交叉划分和双社区不再能够恢复时，就能检测到CP结构；沿着 λ 轴，只要CP结构尚未弱到无法检测，就能检测到它，而当其弱到无法检测时，则会检测到双社区。这与Zhang 等人 [4], 的研究结论有些矛盾，他们发现，在CP结构的情况下没有可检测性阈值。DC对CP结构的恢复范围比NDC窄，一个可能的解释是，度修正旨在通过考虑网络中的度异质性来帮助检测社群结构，而NDC则更倾向于将网络划分为度较低的块和度较高的块 [19], ，而在这里，这对应于隐式植入的CP结构。

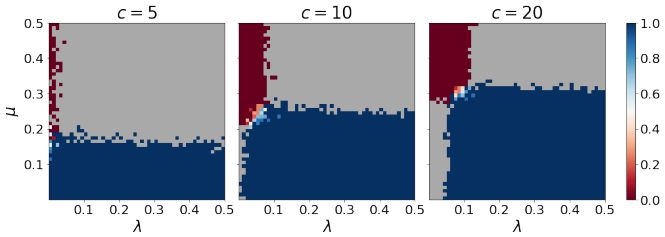
因此，在正则情形下，NDC 与 度修正之间的主要差异之一在于，沿 λ 轴检测到或未检测到结构的阈值。对于所有 c 的取值以及两种随机块模型变体，当双社区和 CP结构较强时，均可恢复交叉划分。随后，沿每个方向，一旦信号变弱，这两种变体都会开始识别相应的双块结构。由于双社区和 CP结构都仅在另一相应结构的信号较弱时才能被恢复，因此，在推断划分分布中两种结构的共存较为罕见；我们将在第V.D节重新讨论这一点。由于两个阈值（沿 λ 和 μ）都会随 c增大而提高，因此，更稠密的图中的交叉划分检测区域会扩大；这一现象在 NDC 中更为显著，且 NDC 在整个（λ， μ）平面上的模型拟合均优于 度修正。

C. 度分布的影响

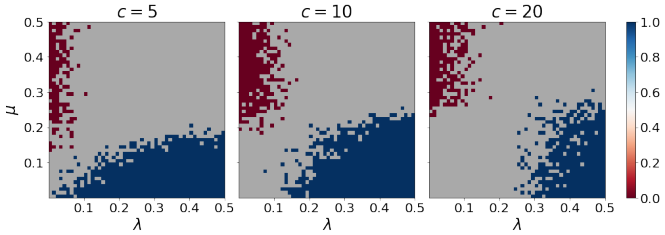
为探究度分布的异质性对划分恢复的影响，我们将两种 SBM 变体拟合到一组由微正则模型生成的网络上。图7b展示了在此情形下，NDC（左）和 DC（右）对划分的平均恢复率。总体检测模式与正则情形中讨论的模式相似。对于 DC，所有划分都在相近区域内以相近的阈值被恢复，尽管这些阈值的边界略显“模糊”。然而，我们观察到 NDC 变体的性能存在显著差异：对于所有植入划分，在整个（λ， μ）平面上， $\langle\omega\rangle$ 均明显更低。首先，这有些出人意料，因为我们在第V.A节已经看到，即使对于由微正则模型生成的图，NDC 也能提供更好的描述。对此现象的一个合理解释是，我们通过微正则模型中对节点度施加的额外约束，以及由此产生的度异质性，引入了额外的结构特征（见第IV.B节）。事实证明，对于相对较强的双社区结构，NDC 变体会恢复出嵌套于两个社群块各自内部的分层CP结构；这种情况不仅出现在植入的 CP 结构非常强时，也出现在完全没有显式植入 CP 结构时（见附录H中的示例图）。当显式植入具有强信号的 CP 结构时，NDC 恢复出的分层CP划分与交叉划分有些相似；当未显式植入 CP 结构时，两个同配块内部的分层CP结构则更类似于双社区划分（根据 ω 并结合对示例网络的目视检查）。如图7b所示，与 NDC 相比，DC 变体检测出的划分更接近植入的



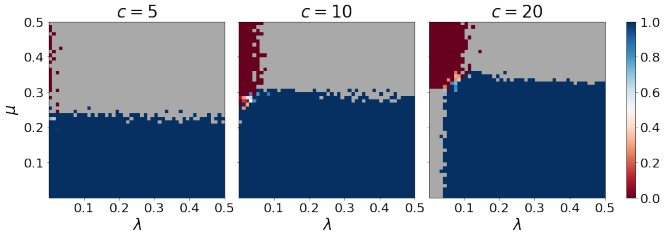
(a) NDC，正则模型。



(b) 度修正，正则模型。



(c) NDC，微正则模型。



(d) 度修正，微正则模型。

图8：在不同期望度 c 下，对于 $\omega_T = 0.75$，成功恢复的所有划分中恢复的双社区划分所占比例为 α；当 $\alpha = 1$ 时（深蓝色），仅检测到双社区结构；当 $\alpha = 0$ 时（深红色），仅检测到核心-边缘结构。

交叉划分（ λ值较低时）和双社区划分（ λ值较高时）。然而，NDC 变体具有更好的模型拟合；这意味着，通过强制施加异质度，我们可能在一定程度上向网络中“过度注入”了更多结构，而不仅仅是由块连接矩阵所定义的结构。

D. 结构共存

除了分别恢复两个植入划分之外，我们自然也关注在给定网络的划分景观的不同区域中，两种结构是否会同时出现；我们将这种现象称为共存。具体而言，我们想知道，对于任意特定的一组 (λ, μ) 参数对，SBM 推断出的划分后验分布是否同时包含两个植入划分（而不是只包含其中一个）。我们通过将划分重叠的阈值设为 $\omega_T = 0.75$ 来衡量共存恢复；在此阈值下，我们认为推断出的划分与植入划分足够接近，因而可视为“成功恢复”。

为说明共存检测，我们绘制了比例 $\alpha = \frac{q_1}{q_1+q_2}$，其中 q_1 表示在给定 ω_T 的情况下，后验分布中与双社区划分相似的划分所占比例，而 q_2 表示与核心-边缘划分相似的划分所占比例。图 8 展示了每种模型/变体组合的结果。对于 $\alpha = 1$ （深蓝色），我们只检测到双社区结构；对于 $\alpha = 0$ （深红色），后验分布中只存在核心-边缘结构；而接近 $\alpha = 0.5$ （白色）的值则表示后验分布更加均衡，即在一定程度上同时包含两种划分；当两种植入划分的检测区域似乎相接甚至重叠时，就会出现此类值。灰色区域表示 (λ, μ) 对，其中 α 未定义，因为两种结构都未能成功恢复。

从图 8c 可以看到，正如第 VC 节的结果所预期的那样，将 NDC 拟合到由微正则模型生成的图上，在我们的 (λ, μ) 参数空间中，任何地方都不会得到同时包含双社区和 核心-边缘划分的划分分布。在其他所有情况下，对于这个相对宽松的阈值 $\omega_T = 0.75$，都存在一些能够检测到共存的小区域。请注意，采用更严格的 ω_T 阈值时，所有情况下各个划分的恢复区域都会缩小，因此在给定后验分布中同时包含两种划分的重叠区域也会更小，甚至完全消失（见附录G）。显然，将 ω_T 取为“成功恢复”的标准有些任意；有人可能会认为，最多只有 75 % 的节点被分配到（隐式植入的）“正确”块中，并不能说明划分相似度高。事实上，我们在此选择展示采用这一较低阈值时的结果，因为这凸显出一个发现：即便阈值如此宽松，共存恢复的范围仍然非常小。

E. 讨论

总体而言，我们找到了每种单独植入结构的可检测性阈值，并发现只有在极少数情况下才能检测到两种结构共存。

我们还观察到，NDC 与度修正变体在成功恢复划分方面存在显著差异；与拟合由正则版本生成的图相比，拟合由微正则模型生成的图时，这种差异更加明显。

我们简要讨论通过微正则模型的生成过程引入网络的度异质性所产生的影响。我们发现，以这种方式约束度分布，会无意中引入超出明确植入的核心-边缘结构之外的CP划分。NDC会识别出这些额外的核心-边缘结构，因为它不对节点度进行校正，因此在检测单个植入划分以及多个划分的共存方面，表现相对较差。这些发现与现有的随机块模型文献一致，包括首次提出度校正版的原始研究 [19]。这些发现提醒我们，网络可能呈现多种相互混杂的结构属性，进而使某些类型划分的检测变得复杂，甚至导致检测出虚假的中尺度结构。一般而言，尤其是在已知网络可能具有哪些结构属性（例如较高的度方差），或关注哪些特定类型结构时，应仔细考虑适用的随机块模型变体。可以考虑采用旨在分离相互混杂的结构属性 [19, 62] 或恢复特定类型结构（如同配社群 [63],）的方法。

其次，我们聚焦于正则模型，考察两种随机块模型变体彼此之间如何恢复双社区划分和核心-边缘划分，以及它们如何联合恢复这两种划分（即它们在给定后验分布中共存时），并将这些结果与交叉划分进行比较。总体而言，我们发现，检测单个植入划分以及检测结构共存的阈值，取决于网络的期望度，也取决于用于检测结构的随机块模型变体。模型证据更高的 NDC 变体更擅长识别交叉划分，但代价是核心-边缘结构的恢复效果较差。与 DC 相比，它恢复双社区划分与核心-边缘划分共存的程度略低。由于我们明确地植入了交叉划分，而只是隐式地植入了双社区结构和核心-边缘结构——具体做法是确保相应交叉块内部和之间的边概率，满足两个双块划分中各块内部和块间的边概率条件——因此，模型拟合更好的变体偏向于交叉划分、以牺牲共存检测为代价，这或许并不令人意外。

无论使用 NDC 还是 DC，我们都发现，即使是在仅植入两个性质不同的真实划分这一相对简单的情况下，在我们的结构强度景观中能够检测到双社区与 CP结构共存的区域也仅限于极小的范围。这一点令人担忧，因为真实网络中的中尺度结构不太可能如此简单，其中可能存在数量更多的共存网络划分以及大量不同的结构。显然，还需要开展更多研究，以更好地理解共存恢复能力的缺失究竟是由于存在某个可检测性极限——超过该极限后，任何算法都无法检测共存（类似于已知的社区检测可检测性阈值 [13, 14]）——还是由于某种其他随机块模型变体能够更好地检测多个植入结构的共存。随后，还应通过改进现有方法或开发新工具，进一步拓展有关划分多样性的相对较新研究领域 [11, 37] 。目标应是使研究人员能够可靠地探索可能导致给定网络生成的多个共存真实划分；我们怀疑，真实图中正是如此 [12]。从这个意义上说，我们的框架和研究结果应被视为测试为此目的而开发的新随机块模型变体的动机。图 8 所示的共存发现区域可作为判断存在多个真实划分时可检测性阈值可能位置的参考。

更一般地，即使对于现有方法，在我们结构强度景观的某些区域中也检测到了共存这一事实，再次凸显了认识到划分分布具有多样性且可能存在分歧的重要性，也强调应用网络科学领域的更多研究者应考虑对网络中尺度结构的多种合理解释。

VI. 结论

我们提出了随机交叉块模型（SCBM），这是一种用于生成式网络模型的框架，可在其中预先设定中尺度结构的歧义。该框架补充了现有生成网络，可作为双真实划分基准，用于衡量中尺度结构检测算法在多大程度上能够恢复由同时植入的两个划分所引入的歧义。我们的工作也进一步强调了探究网络结构歧义问题的必要性：我们明确植入的交叉划分是无歧义的，而歧义则源自两个隐式植入的划分。虽然我们在模拟中聚焦于双分区情形，但也概述了如何将这一方法扩展到多划分情形，并鼓励未来研究更复杂的多重真实结构情形；这些情形或许更接近真实网络中可能出现的情况。

我们详细说明了如何将多重真实划分问题表述为混合成员随机块模型的一个特例，并解释了为什么我们的方法简化了混合成员随机块模型方法。我们发现，只有在“结构强度”空间中一个非常小的区域内，才能检测到两种定性上不同的划分（双社区和核心—边缘结构）的共存；对于我们模型的不同版本，以及用于中尺度结构检测的不同随机块模型变体而言，这一区域的大小和形状也各不相同。

只有当两种结构都足够强，且彼此没有一种占据主导地位时，我们才能恢复出两者共存这一事实。在大多数情况下，当其中一种结构的强度较弱时，我们能够恢复出两个植入划分中的另一个。因此，在多种类型的中尺度结构共同影响网络构建的情况下，我们揭示了一类可检测性阈值。由于在真实网络中，多种关于中尺度结构的合理解释共存似乎是一种普遍现象 [12], ，我们认为，揭示这种迄今仍鲜有研究的可检测性阈值，是对随机块模型文献的重要贡献，尤其是因为大多数社区检测方法仍然旨在发现单一划分，并将其与单一真实标签进行验证。未来还需要从解析角度探索这些可检测性阈值的性质，并评估其他类型共存结构所表现出的可检测性极限，例如包含多个社区划分或二部结构；鉴于某些随机块模型变体检测特定类型结构的能力（或缺乏这种能力），某些结构类型的组合可能比其他组合更容易或更难面临可检测性问题。未来的另一项工作可以是，将针对特定结构的随机块模型变体拟合到由SCBM基准生成的图上。例如，旨在检测同配结构 [63]或核心—边缘结构 [25],的随机块模型，可以与最小描述长度原理 [64]结合使用，以研究某些模型是否在特定参数选择下具有更好的拟合效果，并探索这些模型在检测特定隐式植入结构方面的表现。另一个值得未来研究、且可与SCBM结合的随机块模型变体是层次随机块模型，例如 [62]；已有研究表明，它能够防止随机块模型欠拟合。可以进一步探索它会将共存划分视为彼此嵌套的结构，还是视为出现在同一平坦划分后验分布中的划分。最后，我们认为，未来围绕网络中尺度结构检测方法理论的研究，应着重改进现有方法，使其能够识别共存结构。更广泛地说，与近期研究 [11, 37], 一致，我们认为，使用现有方法分析真实网络的研究者，应关注发现网络分割的多个维度，而不是接受单维解决方案；这些解决方案甚至可能是从“多峰”划分分布中取平均得到的。尤其是在计算社会科学领域——该领域研究在线公共空间中的互动动态——考虑多个合理的网络划分似乎尤为重要。评估社交媒体互动网络中多种结构类型的共存，例如社区/核心—边缘结构，甚至性质不同的社群划分（可能由不对齐的政治维度产生，正如近期对 Twitter 用户关系的研究所示 [65]）可能对研究在线对话动态大有裨益。在这一背景下，研究人员应考虑性质不同的结构共存的情况，并留意本文讨论的可检测性问题。在模拟中，我们聚焦于一种特殊情形：两个植入划分中的块大小均相等，交叉划分也具有相同特征，并且研究社群-CP情形。然而，该基准模型具有灵活性，可涵盖多种结构，包括块大小、度分布和预设中尺度结构各异的情况。未来的研究可以利用这一基准，分析在不同规模、期望度各异的网络中，或在其他中尺度结构组合下，歧义的恢复情况。也可以通过在该基准模型上测试其他类型的检测算法（除随机块模型外）来扩展本研究。进一步扩展基准模型本身，则可以允许每个植入划分包含两个以上的块、存在两个以上的植入划分，或采用模型的有向版本。

致谢

本研究获得欧盟“地平线2020”研究与创新计划下欧洲研究理事会（ERC）“Socsemics”整合者资助（资助协议编号772743）的支持。

附录A：详细的MMSBM 表述

在原始的 MMSBM 表述中， [49], 每个节点以一定概率属于所有潜在群组。对此，使用混合成员关系向量 $\boldsymbol{\pi}_i$ 表示节点i的归属，而非使用独热二进制块归属向量；其中，元素 π_i^k 表示节点 i 属于块 k 的概率，依此类推 $\sum_k \pi_i^k = 1$。对于每个节点 i，都从狄利克雷分布中抽取混合成员向量，且所有节点的参数相同并固定。对于每对节点 i 和 j，根据节点的混合成员向量所确定的多项分布，抽取其块归属 r 和 s；再根据预先设定的块 r 与 s之间的边概率，从伯努利分布中抽样，决定节点 i与 j 之间是否存在一条边。因此，网络的生成过程与伯努利随机块模型相似，但每个节点 i 都会针对与其配对的每个节点 j，反复抽取其块归属。这意味着，节点可能因所考察的配对不同而属于不同的块，并继承多个块的连接模式。由此产生一个独特特征：多个块重叠区域的密度是各块自身密度的加权平均值 [50]。

为便于说明，我们用标签 $\{a,b\}$ 表示划分1中的两个块，用标签 $\{c,d\}$ 表示划分2中的块。我们在方程 A1a 和方程 A1b 中分别用 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$ 定义块矩阵，以表示块内和块间的期望边密度。请注意，块矩阵是对称的，因为我们生成的图是无向的（例如 $\theta_{ab} = \theta_{ba}$）；并且按照无向网络中的惯例，为简化计算，对角线上的元素表示块内边密度的两倍。

$$\begin{cases} \boldsymbol{\theta}_1 = \begin{pmatrix} \theta_{aa} & \theta_{ab} \\ \theta_{ab} & \theta_{bb} \end{pmatrix} & \text{(A1a)} \\ \boldsymbol{\theta}_2 = \begin{pmatrix} \theta_{cc} & \theta_{cd} \\ \theta_{cd} & \theta_{dd} \end{pmatrix} & \text{(A1b)} \end{cases}$$

为了拟合 MMSBM 表述，我们将两个植入划分中的块视为四个潜在块$\{a,b,c,d\}$，但通过约束生成过程，确保某些重叠为空。具体来说，划分1中块 a 或 b 与划分2中块c 或 d 的重叠为空。图1c展示了这一约束，其中重叠区域中的节点按照其在两个植入划分中的块归属颜色着色。因此，我们的混合成员向量形式为 $\boldsymbol{\pi}_i = \frac{1}{2}(\mathbf{b}_i^1, \mathbf{b}_i^2)$。例如，如果 $\boldsymbol{\pi}_i = \frac{1}{2}(1,0,1,0)$，那么节点 i 以相同的概率 $\pi_i^a = \pi_i^c = \frac{1}{2}$ 属于块 a 和 c。因此，与一般的混合成员向量相比，该约束要求节点恰好以相同的概率属于两个块，并且禁止$\boldsymbol{\pi}_i = \frac{1}{2}(1,1,0,0)$和 $\pi_i = \frac{1}{2}(0,0,1,1)$，因为节点永远不可能属于块 a 和 b (或属于 c 和 d)。为每个节点 i抽取一个混合成员向量后，我们根据各自的混合成员向量，独立地为每个节点对 i 和 j 抽取块归属 r和 s。然后，根据取决于 r 和 s 的某个概率，在 i 和 j 之间添加一条边。

1. 归一化

一种直观的初步想法是：如果 $r,\ s \in \{a,b\}$（或 $r,\ s \in \{c,d\}$），则以概率 θ_{rs} 连接两个节点；对于其他任意组合的 r 和 s，连接概率为零。然而，由于每个节点分别根据两个块的连通性接收边，这一过程将无法与 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$ 的连通性保持一致。如果我们按照这些概率生成网络，那么分别属于块 r 和 s 的两个节点 i 和 j 之间的期望边密度 θ_{rs} 为

$$\hat{\theta}_{rs} = \frac{1}{4}\sum_{r's'}(\pi_i^r \pi_j^s \theta_{rs} + \pi_i^u \pi_j^v \theta_{r's'}) = \frac{1}{4}\theta_{rs} + \frac{1}{16}\sum_{r's'}\theta_{r's'}, \tag{A2}$$

对于 $r,\ s \in \{a,b\}$和 $r',\ s' \in \{c,d\}$（或 $r,\ s \in \{c,d\}$和$r',\ s' \in \{a,b\}$）。为了兼顾两个植入划分的连通性，我们因此需要对边概率进行适当归一化，使得 $\hat{}\theta_{rs} = \theta_{rs}$。可加边概率使我们无法找到单一的归一化常数 x。我们可以改为考虑为每个块对寻找 x_{rs}，使节点以概率 $x_{rs}\theta_{rs}$相连。为了保证 $\hat{}\,\theta_{rs} = \theta_{rs}$，我们因此需要

$$\theta_{rs} = \frac{1}{4}x_{rs}\theta_{rs} + \frac{1}{16}\sum_{r's'}x_{r's'}\theta_{r's'}. \tag{A3}$$

寻找合适的 x_{rs} 需要求解一个欠定的六个方程组成的方程组；只要 $\boldsymbol{\theta}_1$ 中概率之和等于 $\boldsymbol{\theta}_2$中概率之和，该方程组就是相容的，因此有无穷多个解。然而，事实证明，对于植入在 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$中的某些连接模式组合，该方程组不存在任何非负解；特别是当两个划分中的块密度差异较大时（证明见附录 A 2）。

另一种对可加边概率进行归一化的方法是：针对两个节点在两个划分中可能占据的每种块对组合，分别确定一个归一化常数。换言之，对于块归属分别为 r 和 s 的节点 i 和 j，若 i 和 j 也分别属于另外两个块 r' 和 s'，那么这两个节点就以概率 $x_{rsr's'}\theta_{rs}$ 相连。为确定 $\{x_{rsr's'}\}$ 的集合，我们将期望密度改写为

$$\hat{\theta}_{rs} = \frac{1}{4}\sum_{r's'}x_{rsr's'}(\theta_{rs} + \theta_{r's'}) \tag{A4}$$

ˆ 并建立一个包含六个方程的方程组，使其满足 $\theta_{rs} = \theta_{rs}$——每个方程对应 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$ 的一个上三角元素——再求解得到 $\mathbf{x} = \{x_{rsr's'}\}$。这同样是一个有无穷多个解的欠定方程组；对于我们在第 IV 节模拟中使用的所有 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$ 组合，都能找到非负解。我们可以使用最小二乘求解器来计算该方程组的最小范数解 [66]。不过，对于某些植入结构组合，最小范数解会使部分 $\{x_{rsr's'}\}$ 取负值。这意味着，对于两个植入结构中的某些边概率组合，我们必须使用非负最小二乘求解器（它可能会返回包含零值的解），或者使用线性规划求解器 [67, 68] 来寻找严格为正的解。

2. 非负解的存在性

对于 $\theta_{aa} = \theta_{bb}$ 和 $\theta_{cc} = \theta_{dd}$ 的简化情形，我们可以利用法卡斯引理 [69] 证明：对于植入这两个结构中的某种密度组合，该系统没有任何非负解；当且仅当 $|\theta_{aa} - \theta_{ab}| + |\theta_{cc} - \theta_{cd}| > \frac{E}{n^2} = 2\rho$ 时，情况如此，其中 ρ 是网络的整体密度。

为了求出归一化常数的集合，我们需要求解方程 $\theta_{rs} = \frac{1}{4}x_{rs}\theta_{rs} + \frac{1}{16}\sum_{uv} x_{uv}\theta_{uv}$，其中 $\mathbf{x} = \{x_{rs}\}$ 。不失一般性，我们假设 $\theta_{aa} = \theta_{bb}$ 并并令 $\theta_{cc} = \theta_{dd}$。我们还先假设 $\theta_{aa} > \theta_{ab}$ 和 $\theta_{cc} > \theta_{cd}$。该系统变为 $\mathbf{Ax} = \mathbf{y}$，其中

$$\mathbf{A} = \begin{pmatrix} \frac{1}{4}\theta_{aa} & 0 & \frac{1}{8}\theta_{cc} & \frac{1}{8}\theta_{cd} \\ 0 & \frac{1}{4}\theta_{ab} & \frac{1}{8}\theta_{cc} & \frac{1}{8}\theta_{cd} \\ \frac{1}{8}\theta_{aa} & \frac{1}{8}\theta_{ab} & \frac{1}{4}\theta_{cc} & 0 \\ \frac{1}{8}\theta_{aa} & \frac{1}{8}\theta_{ab} & 0 & \frac{1}{4}\theta_{cd} \end{pmatrix} \tag{A5}$$

以及 $\mathbf{y} = (\theta_{aa},\ \theta_{ab},\ \theta_{cc},\ \theta_{cd})$。

表述为择一性定理时，法卡斯引理指出，对于 $\mathbf{A} \in \mathbb{R}^{n\times m}$ 和 $\mathbf{x} \in \mathbb{R}^n$，以下两个陈述恰有一个成立：

1. $\exists \mathbf{x} \in \mathbb{R}^m : \mathbf{Ax} = \mathbf{y}$ and $\mathbf{x} \geq 0$
2. $\exists \mathbf{v} \in \mathbb{R}^n : \mathbf{v}^T\mathbf{A} \geq 0$ and $\mathbf{v}^T\mathbf{y} \leq 0$

这意味着，如果我们能够找到一个向量 $\mathbf{v}$，使得第二种情况始终成立，那么就说明无法找到该系统的非负解 $\mathbf{x}$。若选择 $\mathbf{v} = (-1,\ 1,\ 0,\ 2)$，则有 $\mathbf{v}^T\mathbf{A} = (0,\ \frac{1}{2}\theta_{ab},\ 0,\ \frac{1}{2}\theta_{cd})$，且 $\mathbf{v}^T\mathbf{y} = -\theta_{aa} + \theta_{ab} + 2\theta_{cd}$。因此，当且仅当 $\theta_{aa} - \theta_{ab} > 2\theta_{cd}$ 时，有 $\mathbf{v}^T\mathbf{y} < 0$。由于 $\frac{2E}{n^2} = 2\theta_{cc} + 2\theta_{cd}$，我们可以写成

$$\theta_{aa} - \theta_{ab} > 2\theta_{cd} \tag{A6a}$$

$$\theta_{aa} - \theta_{ab} > \frac{E}{n^2} - \theta_{cc} + \theta_{cd} \tag{A6b}$$

$$\theta_{aa} - \theta_{ab} + \theta_{cc} - \theta_{cd} > \frac{E}{n^2} = \frac{4E}{N^2} = 2\rho, \tag{A6c}$$

其中，ρ 是网络的整体密度。由于我们假设 $\theta_{aa} > \theta_{ab}$ 和 $\theta_{cc} > \theta_{cd}$，方程 A6c 表明，如果两个植入划分中的块内密度都足够大于块间密度，则该系统没有任何非负解。对于所有情况 $\theta_{aa} > \theta_{ab}$和 $\theta_{cc} < \theta_{cd}$、 $\theta_{aa} < \theta_{ab}$ 和 $\theta_{cc} < \theta_{cd}$ 以及 $\theta_{aa} < \theta_{ab}$和 $\theta_{cc} > \theta_{cd}$，都很容易找到一个向量 $\mathbf{v}$，因此最终可以说，当且仅当 $|\theta_{aa} - \theta_{ab}| + |\theta_{cc} - \theta_{cd}| > \frac{E}{n^2} = 2\rho$ 时，该系统没有任何非负解。

从更定性的角度来说，这意味着，当我们使两个划分中的块密度差异都变得很大时，就无法找到合适的归一化常数来求解方程组，从而无法构造出连接关系同时符合这两个植入划分的网络。这一限制会严重制约我们测试某些划分组合的可检测性的能力。事实上，这会使我们只能探索模拟中参数空间的一半；在第 IV 节中，我们通过调整两个决定各自结构强度的参数，植入了一个双社区划分和一个核心—边缘分区。

附录 B：常数乘法因子

在块分别标记为 $\{a, b\}$（分区1）和 $\{c, d\}$（分区2）的双划分 SCBM 中，每个节点都会被分配到交叉块 $\boldsymbol{\Gamma} = \{(a,\ c),\ (a,\ d),\ (b,\ c),\ (b,\ d)\}$之一。为便于说明，我们先构造对称矩阵 $\boldsymbol{\theta}'$，其元素是 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$中分别对应于组成 $\boldsymbol{\Gamma}$中交叉块的分区1和分区2中相应块的元素之积。注意，为便于阅读，我们省略跨划分矩阵 4×4 上的 SCBM 下标。因此，矩阵 $\boldsymbol{\theta}'$ 是 $\boldsymbol{\theta}_1$ 和 $\boldsymbol{\theta}_2$的克罗内克积，其顺序取决于指标集 J，

$$\boldsymbol{\theta}' = \boldsymbol{\theta}_1 \otimes \boldsymbol{\theta}_2 = \begin{array}{c} \\ (a,c) \\ (a,d) \\ (b,c) \\ (b,d) \end{array} \begin{array}{c} \begin{array}{cccc} (a,c) & (a,d) & (b,c) & (b,d) \end{array} \\ \begin{pmatrix} \theta_{aa}\theta_{cc} & \theta_{aa}\theta_{cd} & \theta_{ab}\theta_{cc} & \theta_{ab}\theta_{cd} \\ \theta_{aa}\theta_{cd} & \theta_{aa}\theta_{dd} & \theta_{ab}\theta_{cd} & \theta_{ab}\theta_{dd} \\ \theta_{ab}\theta_{cc} & \theta_{ab}\theta_{cd} & \theta_{bb}\theta_{cc} & \theta_{bb}\theta_{cd} \\ \theta_{ab}\theta_{cd} & \theta_{ab}\theta_{dd} & \theta_{bb}\theta_{cd} & \theta_{bb}\theta_{dd} \end{pmatrix} \end{array} \tag{B1}$$

定义乘法因子的最一般形式是为每一对交叉块 $u = (r,\ r')$ 和 $v = (s,\ s')$ 求出 x_{uv}，使得 $B_{uv} = x_{uv}\nu_u\nu_v\theta'_{uv}$。如果我们为所有交叉块固定 $\nu_u = \nu = \frac{n}{2}$（并允许存在自环），那么每个组内和组间可能的最大边数为 $\nu^2 = \frac{n^2}{4}$，因此我们有

$$B_{uv} = x_{uv}\nu^2\theta'_{uv}. \tag{B2}$$

我们希望找到归一化常数 $\{x_{uv}\}$，使得隐式植入划分 1 和 2 中各块内部及块之间的期望边数，等于构成每个原始块的重叠部分的期望边数之和。因此，我们要求

$$B_{1_{rs}} = n^2\theta_{1_{rs}} = \sum_{r's'} x_{uv}\nu^2\theta_{1_{rs}}\theta_{2_{r's'}} \tag{B3}$$

对于分区1中块 r 和 s，分区2中的块也同理。我们可以将其改写为

$$n^2 = \frac{1}{4}\sum_{r's'} x_{uv}n^2\theta_{2_{r's'}} \tag{B4}$$

因此，对于$\mathbf{B}_2$的元素，也有

$$n^2 = \frac{1}{4}\sum_{rs} x_{uv}n^2\theta_{1_{rs}}, \tag{B5}$$

由于 $x = x_{uv} = 2n^2/E = 1/\rho$ 为常数，因此满足

$$\sum_{rs} n^2\theta_{1_{rs}} = \sum_{r's'} n^2\theta_{2_{r's'}} = 2E \tag{B6}$$

附录 C：可变块大小

对于隐式植入的分区1和分区2中块大小相等的情况，令 $2n = N$，但允许跨块大小变化，则可写为 $\nu_1 = \nu_{ac} = \nu_{bd}$和 $\nu_2 = \nu_{ad} = \nu_{bc}$，其中 ν_{ac} 是块（a, c）中的节点数。在这种情况下，我们有 $\nu_1 + \nu_2 = n$。每个块对内部和块对之间的边数可能达到的最大值不再相同。因此，我们不再使用 (B2)，而是

$$B_{uv} = x_{uv}\nu_u\nu_v\theta'_{uv}, \tag{C1}$$

其中 $\nu_u\nu_v$ 是 $\{\nu_1^2, \nu_1\nu_2, \nu_2^2\}$ 之一。

假设对于 $n = \frac{N}{2}$ 和 $\nu_1 \neq \nu_2$，常数 $x = x_{uv}$ 也存在，并且我们正在植入非平凡划分，其中块之间的边放置概率在不同块对之间并不均匀。为了满足划分 1 和 2 中的连通性，我们现在有

$$B_{1_{rs}} = n^2 = x\sum_{r's'} \nu_u\nu_v\theta_{2_{r's'}}. \tag{C2}$$

具体来说，为了满足分区1的块内密度和块间密度，我们需要

$$\frac{n^2}{x} = \theta_{cc}\nu_1^2 + \theta_{dd}\nu_2^2 + 2\theta_{cd}\nu_1\nu_2 \tag{C3a}$$

$$\frac{n^2}{x} = \theta_{cd}\nu_1^2 + \theta_{cd}\nu_2^2 + \theta_{cc}\nu_1\nu_2 + \theta_{dd}\nu_1\nu_2 \tag{C3b}$$

$$\frac{n^2}{x} = \theta_{dd}\nu_1^2 + \theta_{cc}\nu_2^2 + 2\theta_{cd}\nu_1\nu_2, \tag{C3c}$$

为便于阅读，我们省略了 θ 上表示划分编号2的下标。令这些方程中的第一个和最后一个相等，可得

$$(\theta_{cc} - \theta_{dd})(\nu_1^2 - \nu_2^2) = 0. \tag{C4a}$$

由于 $\nu_1 \neq \nu_2$，我们有 $\theta_{cc} = \theta_{dd}$，因此（由方程 (C3a) 和 (C3b)）可得

$$\theta_{cc}\nu_1^2 + \theta_{cc}\nu_2^2 + 2\theta_{cd}\nu_1\nu_2 = \theta_{cd}\nu_1^2 + \theta_{cd}\nu_2^2 + 2\theta_{cc}\nu_1\nu_2 \tag{C5a}$$

$$(\theta_{cc} - \theta_{cd})(\nu_1 - \nu_2)^2 = 0, \tag{C5b}$$

于是 $\theta_{cc} = \theta_{cd} = \theta_{dd}$。显然，对 $\mathbf{B}_2$ 采取相同的步骤可得 $\theta_{aa} = \theta_{ab} = \theta_{bb}$。这与我们的假设相矛盾，因此这样的常数 x 不存在（除了植入随机图这一平凡情况）。

附录 D：求解欠定方程组

对于划分 1 和 2 中每个块都可以取任意大小的一般情况，我们需要确定一组归一化常数 $\{x_{uv}\}$，以构造最终的块矩阵 $\boldsymbol{\theta}_{\text{SCBM}} = \{\theta_{uv}\} = \{x_{uv}\theta_{1_{rs}}\theta_{2_{r's'}}\}$。在这种情况下，两个植入划分以及交叉划分中，每个块内和块之间可能存在的最大边数，都可能因块对而异。因此，期望边数矩阵的元素为 $B_{p_{rs}} = n_r r_s\theta_{p_{rs}}$，适用于 $p \in \{1, 2\}$，以及$B_{uv} = \nu_u\nu_v\theta_{uv} = \nu_u\nu_v x_{uv}\theta'$。常数集 $\{x_{uv}\}$的选取必须满足 $\mathbf{B}_1$和 $\mathbf{B}_2$ 的元素要求（对角线及上三角部分），如方程（D1a）至（D1f）所示。因此，我们共有 $2 \times \frac{K_s(K_s+1)}{2} = 6$ 个方程和$\frac{K(K+1)}{2} = 10$ 个未知数。左侧为$b_{rs} = n_r n_s\theta_{rs} = B_{p_{rs}}$；为便于阅读，我们省略了划分 1 和 2 的下标。右侧为 $\mathbf{B}'_{uv}$，其中 $B'_{uv} = \{\nu_u\nu_v\theta_{1_{rs}}\theta_{s_{r's'}}\}$，下标写作交叉块的索引；例如，元素 $x_{11} = x_{(a,c)(a,c)}$ 是块 (a, c) 中块内边的归一化常数（即划分 s_1 中的块 a 与划分 s_2 中的块 c 的重叠）。

$$\begin{cases} b_{aa} = x_{11}\hat{B}_{11} + 2x_{12}\hat{B}_{12} + x_{22}\hat{B}_{22} & \text{(D1a)} \\ b_{ab} = x_{13}\hat{B}_{13} + x_{14}\hat{B}_{14} + x_{23}\hat{B}_{23} + x_{24}\hat{B}_{24} & \text{(D1b)} \\ b_{bb} = x_{33}\hat{B}_{33} + 2x_{34}\hat{B}_{34} + x_{44}\hat{B}_{44} & \text{(D1c)} \\ b_{cc} = x_{11}\hat{B}_{11} + 2x_{13}\hat{B}_{13} + x_{33}\hat{B}_{33} & \text{(D1d)} \\ b_{cd} = x_{12}\hat{B}_{12} + x_{14}\hat{B}_{14} + x_{32}\hat{B}_{32} + x_{34}\hat{B}_{34} & \text{(D1e)} \\ b_{dd} = x_{22}\hat{B}_{22} + 2x_{24}\hat{B}_{24} + x_{44}\hat{B}_{44}. & \text{(D1f)} \end{cases}$$

方程（D1a）至（D1f）构成一个包含六个线性方程和十个未知数的欠定方程组。我们可以

将其写成矩阵形式：

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{D2}$$

with

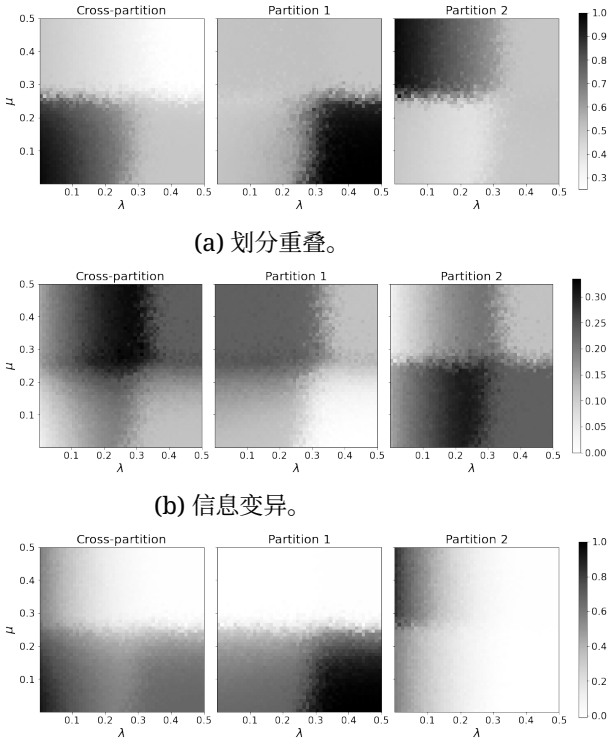
$$\mathbf{x} = (x_{11}, x_{12}, \ldots, x_{32}, x_{33}, x_{44}) \tag{D3}$$
$$\mathbf{b} = (b_{aa}, b_{ab}, b_{bb}, b_{cc}, b_{cd}, b_{dd}) \tag{D4}$$

使得 $\mathbf{x} \in \mathbb{R}^{10}, \mathbf{b} \in \mathbb{R}^{6}$，其中 $\mathbf{A} = \{a_{ij}\} \in \mathbb{R}^{6\times 10}$ˆ是系数矩阵，其元素由$\mathbf{B}$ 决定，并满足方程 (D1a)-(D1f)。根据 Rouch´e–Capelli 定理，我们知道，当且仅当系数矩阵的秩等于增广矩阵 $\mathbf{W} = [\mathbf{A}|\mathbf{b}] \in \mathbb{R}^{6\times 11}$的秩时，这样的欠定方程组才有无穷多个解。对于两个定义良好的连接矩阵$\mathbf{B}_1$和$\mathbf{B}_2$，这一条件总是成立；显然，两个植入划分中的边总数必须相同，并且我们可以证明，当且仅当 $b_{aa} + b_{ab} + b_{bb} = b_{cc} + b_{cd} + b_{dd}$ 时，rank(A) =rank(W)。

一般而言，欠定线性系统 Ax =b，其中 $\mathbf{A} \in \mathbb{R}^{m\times n}$ 且 $m < n$，没有唯一解 x。由于该系统的约束不足，如果它有解，就会有无穷多个解。求解欠定（或超定）线性方程组的一种常用方法称为最小二乘法。最小二乘法的思路是寻找一个解 x，使残差的欧几里得范数的平方 r(x) = b − Ax 最小。换句话说，我们要寻找使 ϕ(x) =$||r$(x)$||_2^2 = ||$b −Ax$||_2^2$ 最小的 x，这可以通过求得满足 $\nabla\phi$(x) = 0 的 x 来实现。由此可得所谓的正规方程 $\mathbf{A}^T$Ax = $\mathbf{A}^T$b；当 $\mathbf{A}^T$A 可逆时，该方程可以解析求解。在我们的情形中，$\mathbf{A}^T$A $\in \mathbb{R}^{n\times n}$的秩至多为 m，其中 $m < n$，因此它是奇异的。这意味着，在这种欠定情形下，正规方程无法解析求解。相反，我们可以求得一个特定的最小二乘解，使欧几里得范数 $||$x$||_2$ （或其平方）最小，并满足约束 Ax =b。当不存在其他约束时，可以通过计算奇异值分解（SVD）来求得最小范数解 ˆx，从而计算矩阵 A 的 Moore-Penrose 伪逆 $\mathbf{A}^+$。随后可将最小范数解计算为 ¯x =$\mathbf{A}^+$b；该解始终存在且唯一 [70]。

附录 E：其他相似性度量

为了展示划分重叠作为我们所用相似性度量的稳健性，我们将其与另外两种度量进行比较：信息变异 [61] 和约化互信息 [60]。我们使用这两种度量来计算划分相似度（对于信息变异而言，则是距离，因为划分差异越大，距离也越大），并针对一组特定模拟进行比较，其中 $N = 400$、 $c = 10$，且图由正则模型生成，划分则通过未进行度校正的随机块模型推断。我们展示结果



(a) 划分重叠。

(b) 信息变异。

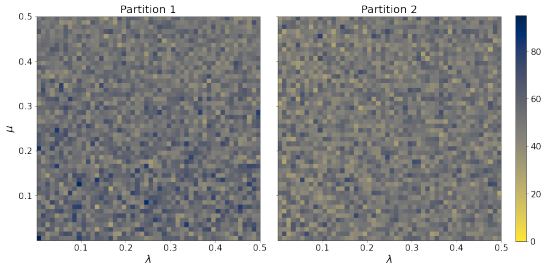
(c) 约化互信息。

图 9：对于具有 $c = 10$ 的图，植入划分与后验分布中的划分之间的平均划分重叠度、信息变异和互信息。

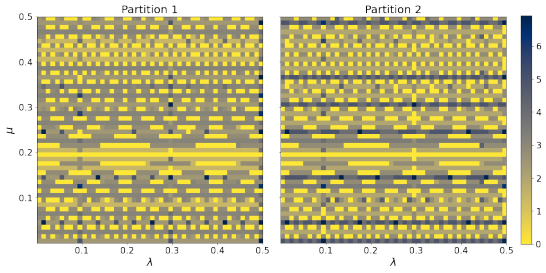
如图 9 所示。尽管平均相似度值存在一些细微差异，但不同划分的可检测性似乎发生变化的区域，位于（λ， μ）空间中的相同位置。

附录 F：生成图的特征

为确保恢复的划分中的变异实际上并非由我们生成的图所致，我们通过计算 $||\mathbf{B}_1 - \mathbf{M}_1||_F$ 和 $||\mathbf{B}_2 - \mathbf{M}_2||_F$，比较植入边计数矩阵 $\mathbf{B}_1$ 和 $\mathbf{B}_2$ 与生成图中的实际边数 $\mathbf{M}_1$ 和 $\mathbf{M}_2$，并遍历 μ 和 λ 的所有取值。图 10a 展示了正则模型生成的图中 $\mu \in [0.01, 0.5]$ 和 $\lambda \in [0.01, 0.5]$ 的平均弗罗贝尼乌斯范数。图 10b 展示了微正则随机块模型对应的图。图中给出了期望度为 $c = 5$ 的图的距离；我们注意到， c 取较高值时也呈现类似的模式。我们观察到，微正则情形下植入边计数矩阵与生成的边计数矩阵之间的距离，按定义，远低于传统情形。我们在微正则情形下观察到的非随机模式，是创建边计数矩阵 $\mathbf{B}$ 时所需的舍入所致；该矩阵用于生成此处的网络



(a) 正则模型。



(b) 微正则模型。

图 10：对于 $c = 5$，网络中植入的边计数矩阵与生成的边计数矩阵之间的平均差异（弗罗贝尼乌斯范数）。

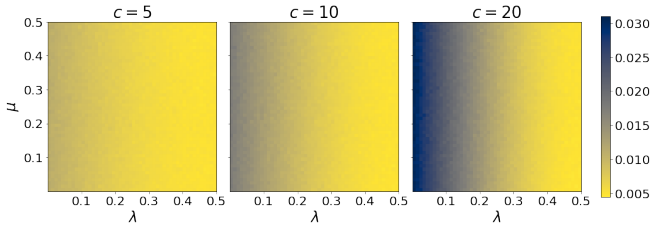


图 11：期望度为 c 的图的平均归一化度方差；这里只展示由正则模型生成的图，以说明 λ 取值较低时引入的度方差。

（而在正则情形下，使用的是边概率，而非边数）。

图 11 展示了正文中图 4 的重缩放版本。这里，我们绘制了由正则模型生成的图的平均归一化度方差，以说明 λ 取值较低时引入的较高方差，尤其是在 c 取值较高时。

在图 12 中，我们绘制了正确分类的节点比例（根据预设的核心与外围块确定），具体方法是将平均度以上的节点划分到核心，将平均度以下的节点划分到边缘，适用于期望度为 $c = 20$的图。在图 13 中，我们展示了 $\mu = 0.1$、 $\lambda = 0.1$ 和 $c = 20$的两个示例度分布。

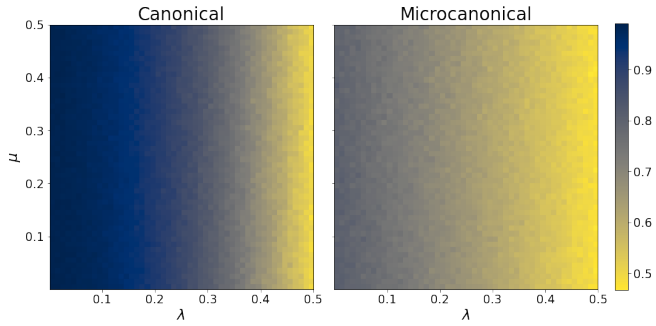


图 12：将度高于期望度 $c = 20$ 的节点归入核心、其余节点归入边缘，从而正确分类到核心与外围块的节点比例。

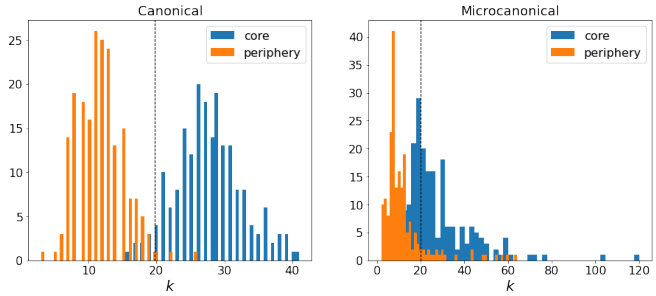


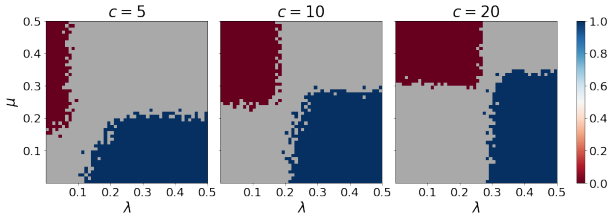
图 13：正则模型和微正则模型中核心节点与外围节点的度分布，以及总体期望度（虚线），对于 $\mu = 0.1, \lambda = 0.1$ 和 $c = 20$

附录G：更严格的划分重叠阈值

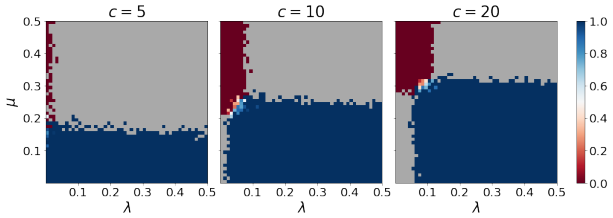
图 14 和图 15 分别展示了两种随机块模型变体中双社区结构和 CP结构的检测结果，以及二者的共存情况；对应的划分重叠分别为$\omega = 0.85$ 和 $\omega = 0.95$ 。

附录H：高度异质性的示例网络

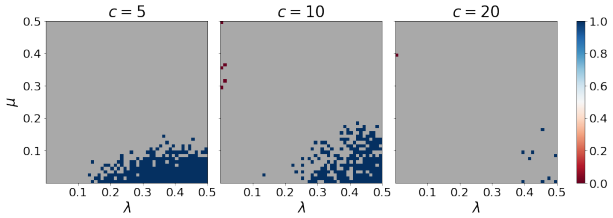
在图 16 中，我们绘制了两个由微正则模型生成的示例网络。网络可视化由 graph-tool Python 库[56]生成。我们固定两个网络的 $\mu = 0.1$，并创建了一张具有显著植入式 CP 结构的图（$\lambda = 0.01$），以及一张未通过边计数矩阵明确植入任何 CP 结构的图（$\lambda = 0.5$）。图 16a 和图 16c 分别展示了 NDC 对 $\lambda = 0.01$ 和 $\lambda = 0.5$ 经常恢复出的一类划分示例。图 16b 和图 16d 展示了 DC 变体对应的结果。我们观察到，DC 对 $\lambda = 0.01$ 恢复出交叉划分，对 $\lambda = 0.5$恢复出双社区划分，这与由微正则模型生成的图所得的对应结果一致，而且



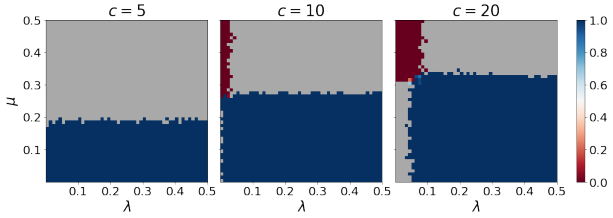
(a) NDC，正则模型。



(b) 度修正，正则模型。

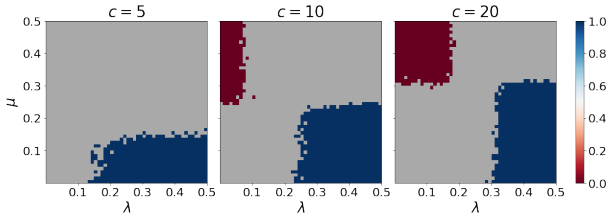


(c) NDC，微正则模型。

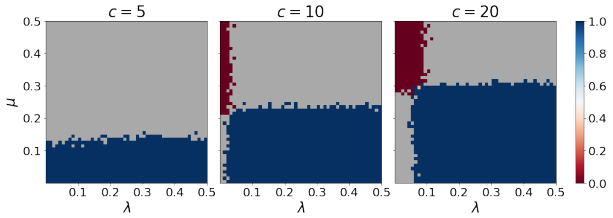


(d) 度修正，微正则模型。

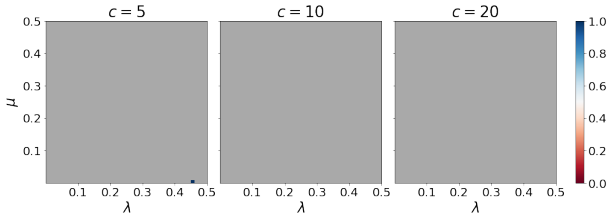
图14：在不同的 c 下， $\omega = 0.85$ 中成功恢复的所有划分里，恢复的双社区划分所占比例为 α ；在 $\alpha = 1$ （分别为 $\alpha = 0$）处，仅检测到双社区（分别为核心-边缘）结构。



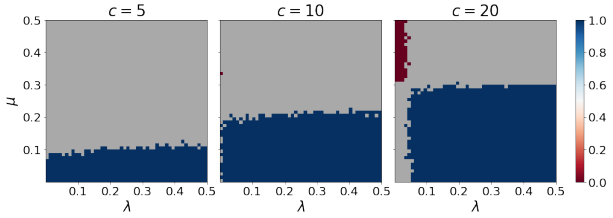
(a) NDC，正则模型。



(b) 度修正，正则模型。



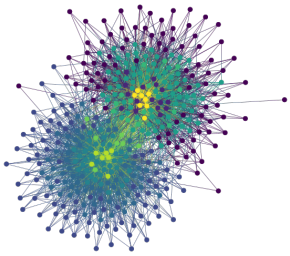
(c) NDC，微正则模型。



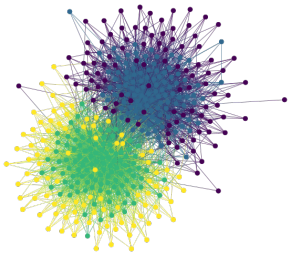
(d) 度修正，微正则模型。

图15：在所有成功恢复的划分中，恢复的双社区划分所占比例为 α ；对于 $\omega = 0.95$，当 c 变化时，在 $\alpha = 1$ （分别为 $\alpha = 0$）处仅检测到双社区（分别为核心-边缘）结构。

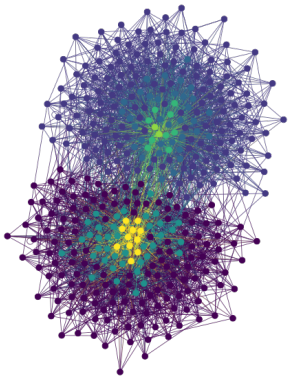
也与我们明确植入的结构一致。然而，NDC（其模型证据更高）对$\lambda = 0.01$ 和 $\lambda = 0.5$检测到相似的结构：一个双块划分，其中每个块都包含一个核心和多个分层边缘。检测到的两个划分之间的差异在于每个块内核心—边缘结构的层数，以及外层边缘的规模。由于这些差异，对 $\lambda = 0.01$ 检测到的划分与交叉划分更相似，而对 $\lambda = 0.5$检测到的划分与双社区划分更相似。



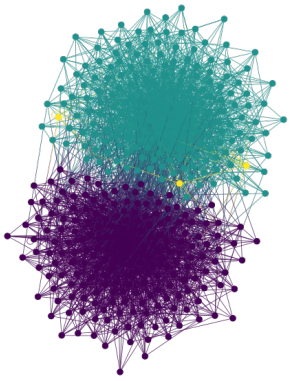
(a) $\mu = 0.1$, $\lambda = 0.01$, NDC



(b) $\mu = 0.1$, $\lambda = 0.01$, 度修正



(c) $\mu = 0.1$, $\lambda = 0.5$, NDC



(d) $\mu = 0.1$, $\lambda = 0.5$, 度修正

图16：固定 $\mu = 0.1$、变化 $\lambda = 0.01$ 时的示例网络；0。5。

[1] J. Leskovec、K. J. Lang、A. Dasgupta 和 M. W. Mahoney，Internet Mathematics 6，29（2009）。[2] M. Rosvall、J.-C. Delvenne、M. T. Schaub 和 R. Lam- biotte，载于*eling* 2019 4, 105（John Wiley & Sons, Ltd，）第 pp. – 119 章。[3] P. Rombach、M. A. Porter、J. H. Fowler 和 P. J. Mucha，SIAM Review 59，619（2017）。[4] X. Zhang、T. Martin 和 M. E. J. Newman，Physical Review E 91，032803（2015）。[5] W. W. Zachary，《人类学研究杂志》33，452（1977）。[6] M. E. J. Newman 和 M. Girvan，Physical Review E 69，026113（2004）。[7] M. E. J. Newman，《美国国家科学院院刊》103，8577（2006）。[8] J. Duch 和 A. Arenas，Physical Review E 72，027104（2005）。[9] V. D. Blondel、J.-L. Guillaume、R. Lambiotte 和 E. Lefebvre，《统计力学杂志：理论与实验》2008，P10008（2008）。[10] T. S. Evans，《统计力学杂志：理论与实验》2010，P12037（2010）。[11] T. P. Peixoto，Physical Review X 11，021003（2021）。[12] L. Peel、D. B. Larremore 和 A. Clauset，《科学进展》3，e1602548（2017）。[13] A. Decelle、F. Krzakala、C. Moore 和 L. Zdeborov´a，Physical Review E 84，066106（2011）。[14] A. Decelle、F. Krzakala、C. Moore 和 L. Zdeborov´a，Physical Review Letters 107，065701（2011）。[15] A. Condon 和 R. M. Karp，Random Structures & Algorithms 18，116（2001）。[16] R. D. Alba，《数学社会学杂志》3，113（1973）。[17] S. Fortunato，《物理学报告》486，75（2010）。[18] M. E. J. Newman，Physical Review E 88，042822（2013）。[19] B. Karrer 和 M. E. J. Newman，Physical Review E 83，016107（2011）。[20] M. Rosvall 和 C. T. Bergstrom，《美国国家科学院院刊》105，1118（2008）。[21] S. P. Borgatti 和 M. G. Everett，Social networks 21，375（2000）。[22] P. Holme，Physical Review E 72，046111（2005）。[23] S. H. Lee、M. Cucuringu 和 M. A. Porter，Physical Review E 89，032810（2014）。[24] M. Cucuringu、P. Rombach、S. H. Lee 和 M. A. Porter，《欧洲应用数学杂志》27，846（2016）。[25] R. J. Gallagher、J.-G. Young 和 B. F. Welles，《科学进展》7，eabc9800（2021）。[26] B. Yan 和 J. Luo，《网络科学》7，70（2019）。[27] B. Tun˛c 和 R. Verma，《公共科学图书馆·综合》10，e0143133（2015）。[28] S. Kojaku 和 N. Masuda，Physical Review E 96，052313（2017）。[29] R. D. Luce 和 A. D. Perry，《心理计量学》14，95（1949）。[30] S. B. Seidman，Social networks 5，269（1983）。[31] R. L. Breiger、S. A. Boorman 和 P. Arabie，《数学心理学杂志》12，328（1975）。[32] M. Girvan 和 M. E. Newman，《美国国家科学院院刊》99，7821（2002）。[33] P. W. Holland、K. B. Laskey 和 S. Leinhardt，《Social Networks》5，109（1983）。[34] T. P. Peixoto，《Physical Review E》95，012317（2017）。[35] A. Lancichinetti 和 S. Fortunato，《Scientific Reports》2，1（2012）。[36] A. Tandon、A. Albeshri、V. Thayananthan、W. Alhalabi 和 S. Fortunato，《Physical Review E》99，042301（2019）。[37] A. Kirkley 和 M. E. J. Newman，《Communications Physics》5，1（2022）。[38] C. Moore，《社群检测的计算机科学与物理学：景观、相变与计算困难性》（2017），arXiv:1702.00467。[39] E. Abbe，《机器学习研究期刊》18，6446（2017）。[40] J. Reichardt 和 M. Leone，《Physical Review Letters》101，078701（2008）。[41] R. R. Nadakuditi 和 M. E. J. Newman，《Physical Review Letters》108，188701（2012）。[42] X. Zhang、R. R. Nadakuditi 和 M. E. J. Newman，《Physical Review E》89，042816（2014）。[43] F. Radicchi，《Physical Review E》88，010801（2013）。[44] F. Lorrain 和 H. C. White，《数学社会学杂志》1，49（1971）。[45] H. C. White、S. A. Boorman 和 R. L. Breiger，《美国社会学杂志》81，730（1976）。[46] M. G. Everett 和 S. P. Borgatti，《数学社会学杂志》19，29（1994）。[47] M. B. Hastings，《Physical Review E》74，035102（2006）。[48] T. P. Peixoto，《Physical Review X》4，011047（2014）。[49] E. M. Airoldi、D. M. Blei、S. E. Fienberg 和 E. P. Xing，《神经信息处理系统进展》21（2008）。[50] T. P. Peixoto，《Physical Review X》5，011033（2015）。[51] T. P. Peixoto，《Physical Review E》92，042807（2015）。[52] A. Lancichinetti 和 S. Fortunato，《Physical Review E》80，056117（2009）。[53] M. Bazzi、L. G. S. Jeub、A. Arenas、S. D. Howison 和 M. A. Porter，《Physical Review Research》2，023100（2020）。[54] B. K. Fosdick、D. B. Larremore、J. Nishimura 和 J. Ugander，《SIAM Review》60，315（2018）。[55] J. Lin，《IEEE 信息论汇刊》37，145（1991）。[56] T. P. Peixoto，figshare 10.6084/m9.figshare.1164194（2014）。[57] H. W. Kuhn，《海军研究后勤季刊》2，83（1955）。[58] J. Munkres，《工业与应用数学学会期刊》5，32（1957）。[59] S. Wagner 和 D. Wagner，《聚类比较——综述》，https://publikationen.bibliothek.kit.edu/1000011477（2007）。[60] M. E. Newman、G. T. Cantwell 和 J.-G. Young，《Physical Review E》101，042304（2020）。[61] M. Meil˘a，《多元分析期刊》98，873（2007）。[62] T. P. Peixoto，《Physical Review X》12，011004（2022）。[63] L. Zhang 和 T. P. Peixoto，《Physical Review Research》2，043271（2020）。[64] T. P. Peixoto，《Physical Review Letters》110，148701（2013）。

[65] P. Ramaciotti Morales，载于 *Complex Networks and Their Applications XI*（Springer International Publishing，沙姆，2023）第176–189页。[66] C. L. Lawson 和 R. J. Hanson，*Solving least squares problems* 1995（SIAM，）。[67] G. Dantzig，*Linear programming and extensions*（普林斯顿大学出版社，1963）。[68] Q. Huangfu 和 J. J. Hall，《数学规划计算》**10**，119（2018）。[69] J. Farkas，《纯粹与应用数学杂志》（《克雷勒数学杂志》）**1902**，1（1902）。[70] S. Boyd、S. P. Boyd 和 L. Vandenberghe，*Convex Optimization* 2004（剑桥大学出版社，）。