arXiv:2602.03612v1 [stat.ML] 3 Feb 2026

基于生成器的热扩散图生成

Anthony Stephenson [1 2] **Ian Gallagher** [3] **Christopher Nemeth** [1]

摘要

图生成建模已成为一项重要任务，这得益于其在化学、生物学、社交网络和知识表示等领域的广泛应用。在本研究中，我们提出一种新颖的图生成框架，将生成元匹配（Holderrieth 等人，2024年）范式应用于图结构数据。我们利用图拉普拉斯算子及其对应的热核，为每个图定义连续时间扩散。拉普拉斯矩阵充当该扩散过程的无穷小生成元，其热核则提供一族以初始图为条件的扰动。我们通过最小化真实生成元与可学习代理生成元之间的布雷格曼散度，训练神经网络以匹配该生成元。训练完成后，我们使用代理生成元模拟时间反向的扩散过程，从而采样新的图结构。我们的框架统一并推广了现有的基于扩散的图生成模型，通过拉普拉斯矩阵注入领域特定的归纳偏置，同时保留神经近似器的灵活性。实验研究表明，我们的方法能够有效捕捉真实图和合成图的结构属性。

1. 引言

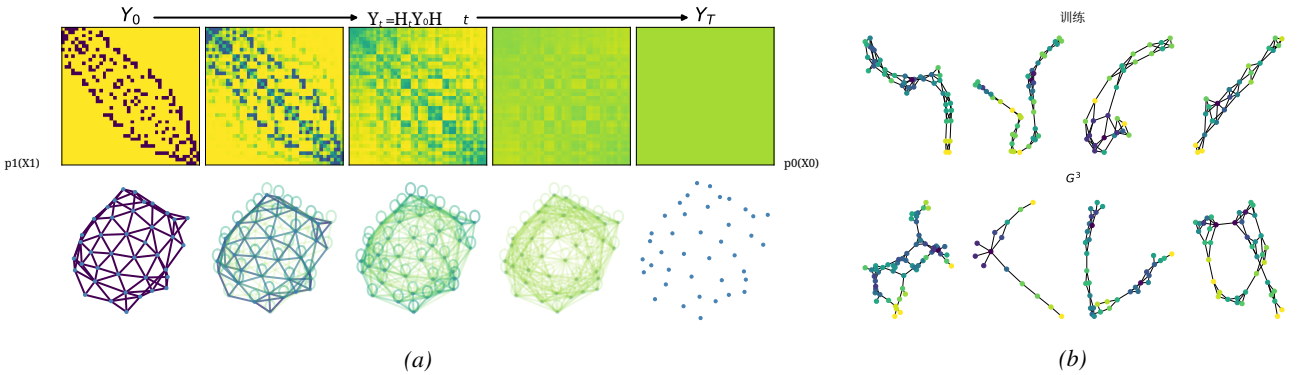
由于图在表示分子结构等复杂系统方面具有广泛的适用性（Sylvester，1878；DeCao& Kipf，2018年）、社交网络（Nettleton，2013）和知识图谱（Schneider，1973；Ji 等人，2021），图生成建模已成为一个重要的研究领域。图生成的目标是学习一组观测图的底层分布，并生成具有相似结构属性的新样本。其应用涵盖化学中的从头分子设计（Brown 等人，2019）到用于隐私保护数据共享的合成网络构建（Fu 等人，2023）。然而，图这类结构化数据给生成模型带来了独特挑战，原因在于其离散组合性质，以及节点标签之间的置换对称性。

图生成的早期方法包括基于图结构的变分自编码器（Kipf & Welling，2016年）、自回归模型（You 等人，2018年）以及生成对抗网络（Wang 等人，2018年）。尽管这些方法取得了一定成功，但它们通常难以捕捉高阶依赖关系，或难以确保置换对称性等基本不变性。近年来，基于分数的生成模型（也称为扩散模型）在图像等连续定义域中展现出卓越性能：通过前向扩散和学习得到的反向过程，逐步将噪声转换为数据（Song 等人，2020年）。这一成功促使人们越来越关注如何将扩散模型应用于图结构数据，从而产生了图扩散模型。现有方法通常作用于邻接矩阵或节点和边特征，并已在分子生成和网络补全等任务中取得了令人鼓舞的结果（Vignac 等人，2022年；Yang 等人，2023年）。值得注意的是，一些近期方法构造了置换不变随机微分方程，对节点和边属性进行联合扩散（Jo 等人，2022年）；另一些方法则在通过图自编码器获得的连续潜在空间中应用扩散。尽管取得了这些进展，如何在离散图结构上直接定义合理且有效的扩散过程仍然具有挑战性，这推动了更通用的以生成元为核心的框架的发展。

贡献。 本文提出基于生成器的图生成 (G^3)，这是一种用于图生成建模的生成器匹配方法。我们 (a) 选择一种有原则且考虑拓扑结构的前向 加噪机制，即在图的矩阵表示上进行拉普拉斯热扩散；(b) 利用其闭式无穷小生成元定义明确的学习目标；(c) 使用前向扩散样本训练神经网络代理生成元。在采样

[1]MARS：现实世界系统中的人工智能数学，兰卡斯特大学数学科学学院，Lancaster, LA1 4YF, UK [2]布里斯托大学数学系，Bristol BS8 1UG, UK。[3]墨尔本大学数学与统计学院，Parkville, VIC, 3010, Australia。通信作者：Anthony Stephenson <ant.stephenson@bristol.ac.uk>、Ian Gallagher <ian.gallagher@unimelb.edu.au>、Christopher Nemeth< c.nemeth@lancaster.ac.uk>。

预印本。*2026*年*2*月*4*日。



(a) *(b)*

图*1.* (a) 展示了对称热扩散过程对图（底行）及其邻接矩阵（顶行）的影响，随着 t 增大，概率质量会分散到所有可能的边上。颜色表示该数值，其中黄色和紫色分别对应区间 $[0,1]$ 的两端。边的透明度也由同一数值控制。 (b) 展示了从 Enzyme 数据集（顶行）的（训练）子集中随机抽取的一系列图，以及 G^3 的输出图（底行）。节点着色反映了该节点处的 Forman-Ricci 曲率（Forman, 2003）。

时，我们从简单的置换对称基分布出发，积分学习到的反向时间常微分方程，得到实值图矩阵，然后将其映射为离散邻接矩阵。图 1a展示了 G^3 框架的正向与反向行为。

2. 背景

连续时间生成模型可以表述为确定性或随机动力学，用以将简单参考分布中的样本变换为复杂数据分布中的样本。学习此类动力学最流行的方法之一是流匹配，及其扩展生成元匹配。除了在一般设定下回顾这些范式之外，我们还建立了图论预备知识——矩阵表示与拉普拉斯热扩散——以支撑我们的 G^3 框架，该框架在第3节中提出。

2.1. 图矩阵表示

我们考虑无向简单图 $G = (V,\ E)$，其具有$n = |V|$ 个节点 $V = \{1, \ldots, n\}$ 以及边集 $E \subseteq \{\{i, j\} :\ i,\ j \in V,\ i \neq j\}$。邻接矩阵 $A \in \{0,1\}^{n\times n}$表示图中的连通性，其中 $A_{ij} = 1$ 当且仅当$\{i, j\} \in E$，否则为 $A_{ii} = 0$。度矩阵 D =diag($d_1,\ \ldots,\ d_n$)是节点度$d_i = \sum_{j=1}^{n} A_{ij}$ 的对角线矩阵。

组合拉普拉斯矩阵 $L = D - A$ 是一个对称半正定矩阵，用于描述无向图上的热扩散。对于连通图，拉普拉斯矩阵的特征值 0 的重数为 1，其对应的特征向量为全一特征向量$\mathbf{1}$，$L\mathbf{1} = 0$。

拉普拉斯矩阵的其他版本以不同方式对图结构进行归一化，例如随机游走拉普拉斯矩阵 $L_{\text{rw}} = I - D^{-1}A$。以下方法适用于其他拉普拉斯矩阵，但除非另有说明，结果默认指组合拉普拉斯矩阵。

2.2. 图上的热方程与热核

我们研究无向图上的热扩散，并介绍其解，即图热核。这一构造为图上的扩散提供了一种概念，并将在我们的生成框架中使用。

定义 2.1（图上的热方程）。 设 $G = (V,\ E)$ 为一个无向图，具有 $n = |V|$ 个节点、图拉普拉斯矩阵 $L \in \mathbb{R}^{n\times n}$ 和热核

$$H_s := e^{-sL}, \qquad s \geq 0. \tag{1}$$

给定一个初始矩阵值图表示 $Y_0 \in \mathbb{R}^{n\times n}$（例如 $Y_0 = A$ 邻接矩阵），定义对称热扩散为

$$Y_s := H_s\, Y_0\, H_s, \qquad s \geq 0. \tag{2}$$

由于 $\frac{\mathrm{d}}{\mathrm{d}s} H_s = -LH_s = -H_s L$，对（2）求导可得矩阵值热方程

$$\frac{\mathrm{d}}{\mathrm{d}s} Y_s = -LY_s - Y_s L, \qquad Y_0 = A. \tag{3}$$

（证明见附录 A.1 中的命题 A.1。）由于L 是对称半正定的，算子 $Y \mapsto LY + YL$ 具有耗散性，而（3）定义了作用于图导出的矩阵的稳定线性扩散（示例见附录A.2 和A.4中的引理A.2 以及图 1a）。直观而言，左乘和右乘 H_s 会平滑 Y_0 中编码的结构，使其与图拓扑保持一致。

尽管（2）可以通过其关联半群和长时间极限来理解，我们将这一视角推迟到第3.1节，在那里它直接启发了我们的生成元匹配表述以及基分布的选择。

2.3. 流匹配

流匹配（Lipman 等人，2022年）是一种生成建模框架，人们在其中学习一个确定性流，将概率质量从单一且通常简单的初始分布传输到连续状态空间上的目标分布 $\mathcal{X} \subseteq \mathbb{R}^d$. 该方法可视为基于扩散的生成建模（Song 等人，2020年）的确定性类比，其中随机动力学被常微分方程（常微分方程）所取代。

向量场与概率路径。令 p_{init} 表示 $\mathcal{X}$ 上的参考分布（通常为高斯分布或均匀分布），令 p_{data} 表示目标数据分布。流匹配模型引入一个依赖时间的向量场 u：$\mathcal{X} \times [0,1] \to \mathcal{X}$，并通过常微分方程定义轨迹 $(X_t)_{t\in[0,1]}$

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = u_t(X_t), \qquad X_0 \sim p_{\text{init}}. \tag{4}$$

该常微分方程的解（在温和的正则性条件下）会诱导出一个流映射 $\psi_t : \mathcal{X} \to \mathcal{X}$，使得 $X_t = \psi_t(X_0)$。目标是学习一个向量场 u，使其终端分布 X_1 与目标数据分布p_{data}匹配。

流匹配并非直接通过似然或得分函数来指定 u，而是首先规定一条概率路径 $(p_t)_{t\in[0,1]}$，在 p_{init} 与 p_{data}之间进行插值。若依赖时间的向量场 $u^* : \mathcal{X} \times [0,1] \to \mathcal{X}$ 满足连续性方程，则称其与该路径一致（p_t）

$$\partial_t p_t(x) + \nabla \cdot \big(p_t(x), u_t^*(x)\big) = 0, \quad t \in (0,1). \tag{5}$$

代理向量场与目标函数。随后训练神经向量场 $u^\theta : \mathcal{X} \times [0,1] \to \mathcal{X}$通过最小化 u^*的期望差异来近似 u^*

$$\mathcal{L}_{\text{FM}}(\theta) = \mathbb{E}_{t\sim\text{Unif}, x\sim p_t}\Big[D\big(u_t^*(x), u_t^\theta(x)\big)\Big], \tag{6}$$

其中，$D(\cdot,\ \cdot)$ 是一种布雷格曼散度，在实际应用中通常采用均方误差或 KL散度。

流匹配避免了模拟随机微分方程时通常会遇到的数值问题（Song 等人，2020年），允许灵活选择概率路径，并且在采样时可以使用标准常微分方程求解器实现（Chen 等人，2018年）。然而，该框架从根本上是在向量场层面运行的，并未明确约束所学习的动力学必须对应于任何潜在的随机过程或无穷小生成元。

2.4. 生成元匹配

生成元匹配（Holderrieth 等人，2024年）是一种学习连续时间生成模型的框架

该框架将重点从匹配向量场转向匹配马尔可夫过程的无穷小生成元。设 $(X_t)_{t\in[0,1]}$ 是定义在状态空间 $\mathcal{X}$ 上的（可能非齐次）连续时间马尔可夫过程，其（无穷小）生成元为 $\mathcal{J}_t$。对于足够规则且$f : \mathcal{X} \to \mathbb{R}$ 属于 $\mathcal{J}_t$ 的定义域的测试函数，生成元定义为

$$(\mathcal{J}_t f)(x) := \lim_{h\downarrow 0} \frac{\mathbb{E}[f(X_{t+h}) \mid X_t = x] - f(x)}{h}. \tag{7}$$

半群与柯尔莫哥洛夫方程。生成元视角与描述期望和边缘分布如何演化的柯尔莫哥洛夫方程紧密相关。在时间齐次情形下（即当 $\mathcal{J}_t \equiv \mathcal{J}$ 不依赖于 t 时），该过程诱导出一个作用于可观测量 f 的一参数马尔可夫半群 $(S_t)_{t\geq 0}$： $\mathcal{X} \to \mathbb{R}$ 其形式为

$$(S_t f)(x) := \mathbb{E}[f(X_t) \mid X_0 = x], \qquad t \geq 0,$$

且对于 $f \in \text{Dom}(\mathcal{J})$，该半群满足 Kolmogorov 后向方程

$$\partial_t (S_t f) = \mathcal{J}(S_t f), \qquad S_0 f = f.$$

（等价地，$\partial_t(S_t f) = S_t(\mathcal{J} f)$。）对偶地，若 p_t表示 X_t 的边缘分布，且 $\mathcal{J}^*$ 是 $\mathcal{J}$ 在弱意义下的伴随算子，则 p_t 按照前向（Fokker–Planck／主）方程演化

$$\partial_t p_t = \mathcal{J}^* p_t.$$

在更一般的时间非齐次 情形下，我们使用一族 $(\mathcal{J}_t)_{t\in[0,1]}$，而 $(S_t f)$ 和 (p_t) 的相应演化仍由这些后向／前向方程支配，只需将 $\mathcal{J}$ 替换为 $\mathcal{J}_t$。这正是生成元匹配以生成元作用为对象的原因：匹配 $\mathcal{J}_t$（或 $\mathcal{J}$）可确保与可观测量和分布所诱导的演化保持一致。

代理生成元与目标函数。在生成元匹配中，假设可以访问一个前向过程，其（可能随时间变化的）生成元 $\mathcal{J}_t$ 诱导出一族边缘分布 $(p_t)_{t\in[0,1]}$，在 p_{init} 与 p_{data}之间进行插值。与流匹配中匹配向量场不同，目标是通过最小化在前向样本上评估的生成元作用之间的差异，用参数化的代理生成元 $\mathcal{J}_t^\theta$近似 $\mathcal{J}_t$。对于一组合适的测试泛函 $\mathcal{F}$，训练过程即是最小化损失

$$\mathcal{L}_{\text{GM}}(\theta) = \mathbb{E}_{t\sim\text{Unif},\, X_t\sim p_t}\Big[D\big((\mathcal{J}_t f)(X_t), (\mathcal{J}_t^\theta f)(X_t)\big)\Big], \tag{8}$$

对于 $f \in \mathcal{F}$ 以及其中 $D(\cdot,\ \cdot)$ 表示布雷格曼散度。

线性泛函参数化。对于我们针对 G^3 考虑的结构化状态空间，我们选择 $\mathcal{F}$ 为线性泛函 $f_A(X) = \langle A,\ X\rangle$ (例如 $\langle A,\ X\rangle =$ 矩阵状态的 $\mathrm{tr}(A^\top X)$)。当状态上存在一个诱导算子（为方便起见，仍记作 $\mathcal{J}_t$），使得二者均

$$(\mathcal{J}_t f_A)(X) = \langle A,\, \mathcal{J}_t(X)\rangle,$$
$$(\mathcal{J}_t^\theta f_A)(X) = \langle A,\, \mathcal{J}_t^\theta(X)\rangle,$$

目标函数可化简为状态空间形式

$$\mathcal{L}_{\mathrm{GM}}(\theta) = \mathbb{E}_{t\sim\mathrm{Unif},\, X_t\sim p_t}\Big[D\big(\mathcal{J}_t(X_t), \mathcal{J}_t^\theta(X_t)\big)\Big], \quad (9)$$

我们将在第3节中始终采用这一形式。

当生成元对应于确定性常微分方程，且散度的选取恰当时，流匹配可视为生成元匹配的一种特例。然而，以生成元为中心的视角表达能力更强：它能够捕捉随机动力学、跳跃过程和结构化状态空间，并自然地确保与支配底层过程的柯尔莫哥洛夫方程保持一致。因此，生成元匹配特别适用于图等结构化数据，在这类数据中，可以直接通过图算子定义有意义的前向过程。这一观察构成了我们 G^3 方法的基础。

3. 图生成

现在我们介绍基于生成器的图生成方法 G^3。该方法从一种图加噪机制出发：对矩阵值图表示进行对称拉普拉斯热扩散$(Y_s)^{s\in[0,T]}$ （见第2.2节），并通过时间重标度 $[0,1]$，将其转化为与生成元匹配（第$(X_t)^{t\in[0,1]}$ 2.4节）相容的参数化过程： X_t 。$= Y_{T(1-t)}$由此得到图状态上的显式真实生成元，我们使用生成元匹配目标函数，通过神经网络替代模型 $\mathcal{J}_t^\theta$对其进行近似。随后，我们从简单的基分布出发，通过数值积分学习到的反向时间常微分方程，对其采样生成新图。

3.1. 图扩散与生成元

我们将图热扩散嵌入生成元匹配框架，即第2.4节。关键在于，对称热扩散具有闭式生成元；通过对时间重新缩放，我们在 $[0,1]$ 上得到与 X记号一致的过程，该记号用于第2.3–2.4节。

半群视角与长时间极限。正向热扩散（2）具有一种方便的半群解释，自然地与生成元匹配以及简单基态的选择相联系。定义线性映射，对于$s \geq 0$,

$$\mathcal{P}_s(Y) := H_s\, Y\, H_s. \quad (10)$$

于是 $\{\mathcal{P}_s\}_{s\geq 0}$ 是 $\mathbb{R}^{n\times n}$上的强连续半群，且 $\mathcal{P}_0 = \mathrm{Id}$以及 $\mathcal{P}_{s+t} = \mathcal{P}_s \circ \mathcal{P}_t$。此外，其无穷小生成元为

$$(\mathcal{G}Y) = -\big(LY + YL\big),$$

从而可将（3）抽象地写为$\mathrm{d}\ _{\mathrm{d}s} Y_s = \mathcal{G}Y_s$ （见引理A.5，附录A.3）。

极限行为与概率解释。对于连通图，拉普拉斯矩阵在0处具有简单特征值，对应特征向量为$\mathbf{1}$，且其余特征值均为严格正数；因此

$$\lim_{s\to\infty} H_s = \Pi := \frac{1}{n}\mathbf{1}\mathbf{1}^\top,$$

到常数上的正交投影算子。因此，

$$\lim_{s\to\infty} Y_s = \Pi Y_0 \Pi = \frac{1}{n^2}\|Y_0\|_1\, \mathbf{1}\,\mathbf{1}^\top,$$

对于非负矩阵 Y_0。对于连通图，该收敛以指数速度进行，其速率由拉普拉斯矩阵的谱隙决定（见引理A.6 ，附录A.4）。此外，当使用随机游走拉普拉斯矩阵 L_{rw} 时， H_s 可解释为连续时间随机游走在 V 上的转移核（Chung，1997年）。

将时间重缩放为 $[0,1]$。 生成元匹配（以及流匹配）定义在 $t\in[0,1]$ 上，其中 X_0 从简单的参考分布中抽取，且 $X_1 \sim p_{\mathrm{data}}$。为与此约定保持一致，对于固定时域 $T>0$，我们定义：

$$X_t := Y_{T(1-t)}, \qquad t\in[0,1]. \quad (11)$$

因此， $X_0 = Y_T$ 是充分扩散后的（简单）状态，而 $X_1 = Y_0$ 则恢复原始图表示。对（11）求导可得

$$\frac{\mathrm{d}}{\mathrm{d}t}X_t = T\big(LX_t + X_tL\big), \qquad t\in[0,1]. \quad (12)$$

等价地，以上述前向生成元 $\mathcal{G}$ 表示时，反向时间算子满足 $\mathcal{J} = -T\mathcal{G}$。正向热扩散具有收缩性；其形式上的时间反演对应于一种反扩散演化，会指数级放大高频（拉普拉斯矩阵特征模态）分量，使精确的反向动力学成为病态问题且数值不稳定（Hadamard，1923年；Evans，2010年）（见附录A.7中的引理 A.4）。

我们的生成器匹配方法通过从前向扩散样本学习正则化代理生成元 $\mathcal{J}_t^\theta$来规避这种不稳定性，从而得到稳定且条件良好的反向时间生成采样器。

状态上的无穷小生成元。对于线性测试泛函 $f_A(X) = \langle A, X\rangle := \mathrm{tr}(A^\top X)$ 其中 $A \in \mathbb{R}^{n\times n}$,(12) 蕴含

$$\frac{\mathrm{d}}{\mathrm{d}t} f_A(X_t) = \left\langle A, \frac{\mathrm{d}}{\mathrm{d}t} X_t \right\rangle = \langle A,\ T(LX_t + X_t L)\rangle .$$

将关注范围限制在此类线性泛函上，会导出一个（状态空间）时间齐次线性算子，我们将其记为 $\mathcal{J}$，定义如下

$$(\mathcal{J} f_A)(X) = \langle A, \mathcal{J}(X)\rangle = \langle A, T\big(LX + XL\big)\rangle. \quad (13)$$

该 $\mathcal{J}$ 完全由图拉普拉斯矩阵决定，因此其天然具备置换等变性和图局部性。

3.2. 图生成元训练

训练数据和前向扩散。令 $\{Y_0^{(k)}\}_{k=1}^N$ 表示一组训练数据图，并将其表示为矩阵。定义对称热扩散

$$Y_s^{(k)} = e^{-sL^{(k)}}\, Y_0^{(k)}\, e^{-sL^{(k)}}, \qquad s \in [0, T],$$

以及 $[0, 1]$, 上经时间重缩放的过程

$$X_t^{(k)} := Y_{T(1-t)}^{(k)}, \qquad t \in [0, 1].$$

通过从$t \sim \mathrm{Unif}[0, 1]$ 中抽样并依据上述映射构造 $X_t^{(k)}$，即可获得加噪图训练样本。整个过程中，我们将初始图状态设为邻接矩阵 $Y_0^{(k)} = A^{(k)}$，其对应的拉普拉斯矩阵为 $L^{(k)}$。由于这些对象均为对称矩阵，我们在实现中只保留矩阵的下三角部分，这也隐式地去除了自环。

代理生成元与目标函数。根据（13），作用于状态的训练图真实生成元 $\{Y_0^{(k)}\}_{k=1}^N$ 为

$$\mathcal{J}^{(k)}(X) = T\big(L^{(k)}X + XL^{(k)}\big). \quad (14)$$

利用这些目标生成元，我们通过最小化以下目标来学习代理生成元$\mathcal{J}^\theta$： $\mathbb{R}^{n\times n} \times [0, 1] \to \mathbb{R}^{n\times n}$

$$\mathbb{E}_k\, \mathbb{E}_{t\sim\mathrm{Unif}} \left[D\big(\mathcal{J}^{(k)}(X_t^{(k)}),\ \mathcal{J}_t^\theta(X_t^{(k)})\big)\right], \quad (15)$$

其中 D 是布雷格曼散度。在第5节所述的实验中，我们只使用均方误差损失， $D(A,\ B) = \|A - B\|_F^2$。参见算法1。

3.3. 通过反向时间常微分方程采样

在学得代理生成元 $\mathcal{J}_t^\theta$（来自正向加噪图 $\{Y_{(k)\,0}, Y_{(k)\,1}, \ldots, Y_{(k)\,T}\}_{k=1}^N$）后，我们在时间重标定设定下通过对所学常微分方程自 $t = 0$ 处的简单基分布积分至 $t = 1$ 处的结构化图状态。注意， $t \mapsto X_t = Y_{T(1-t)}$ 已经反转了原始扩散时间 s，因此在 t 中正向积分对应于 s中热扩散的形式逆转。

基分布。我们将 p_{init} 设为一种置换对称分布，旨在近似由前向热扩散诱导的高度扩散状态 $X_0 = Y_T$（参见第2.2节）。具体而言，我们通过将各列独立同分布地采样为$\tilde{\boldsymbol{x}}_{0,i} \sim$狄利克雷分布$(\alpha\mathbf{1})$， $i = 1, \ldots, n$，再施加对称性并按平均度 $\langle d\rangle := \frac{2}{nN}\sum_{k=1}^N \|Y_{(k)\,0}\|_1$进行缩放，来采样辅助矩阵 $X_0 \in \mathbb{R}^{n\times n}$：

$$X_0 := \langle d\rangle\, (\widetilde{X}_0 + \widetilde{X}_0^\top), \quad (16)$$

其中 $\alpha > 0$ 控制均匀向量周围的浓度参数。尽管在数据分布下 Y_T 的精确分布难以求得， p_{init} 仍提供了一种简单的替代分布，能够捕捉置换对称性和强扩散特性。

在上模拟常微分方程。对于每个训练图k，经时间重缩放的热动力学满足$\frac{\mathrm{d}}{\mathrm{d}t} X_t^{(k)} = \mathcal{J}_{(k)}(X_t^{(k)})$，其中 $\mathcal{J}_{(k)}(X) = T(L^{(k)}X + XL^{(k)})$。在生成时，我们不以特定的拉普拉斯矩阵为条件；相反，我们对学得的替代动力学进行积分

$$\frac{\mathrm{d}}{\mathrm{d}t} X_t = \mathcal{J}_t^\theta(X_t), \qquad t \in [0, 1], \qquad X_0 \sim p_{\mathrm{init}}, \quad (17)$$

其中， $\mathcal{J}_t^\theta$经过训练，以匹配前向扩散状态的热生成元作用。

欧拉离散化与稳定化。取一个整数$M \geq 1$ 并设 $\delta := 1/M$，时间网格为 $t_m := m\delta$ 对于$m = 0, \ldots, M$. 将 X_{t_m}近似为 X^m，对（17）应用显式欧拉积分可得

$$X^{m+1} = X^m + \delta\, \mathcal{J}_t^\theta(X^m), \qquad m = 0, \ldots, M-1. \quad (18)$$

为满足基本有效性约束并提高数值稳定性，我们可以选择在每一步之后应用投影 Π：

$$X^{m+1} = \Pi\big(X^m + \delta\, \mathcal{J}_t^\theta(X^m)\big). \quad (19)$$

典型的投影包括对称化 $\Pi_{\mathrm{sym}}(Z) = \frac{1}{2}(Z + Z^\top)$、将对角线置零以及逐元素裁剪。最终的离散邻接矩阵通过阈值化操作 $\widehat{A}_{ij} = \mathbf{1}\{(\Pi_{\mathrm{sym}}X^M)_{ij} \geq c\}$ 得到。

4. 相关工作

除了第 1 节讨论的工作外，我们还介绍几种具有代表性的图生成方法。SPECTRE（Martinkus 等人，2022年）训练一系列生成对抗网络来采样特征值，然后重构特征向量和

算法 1 G^3 训练

输入： 图表示 $\{Y_0^{(j)}\}_{j=1}^N$，最大扩-
散时间 T，批大小 b、$p_{\text{iter}} = 0.0$、$\delta > 0$，学习率 η，
最小扩散时间 ，
重复
  对 批次**在** 各批次中**执行**
    采样扩散时间：$\{t_i\}_{i=}^{b}$ $_{1} \sim$均匀分布$[\tau, T]$
    扩散状态：$\{X_{s_i}^{(i)}\}_{i=1}^{b}$ 对于 s_i： $= 1 - t_i/T$
    生成器目标：$\{T(L_i X_{s_i}^{(i)} + X_{s_i}^{(i)} L_i)\}_{i=1}^{b}$
    预测生成器：$\{\mathcal{J}_{t_i}^{\theta}(X_{s_i}^{(i)})\}_{i=1}^{b}$
    损失 $= \frac{1}{b}\sum_{i=1}^{b} \left\| \mathcal{J}_{t_i}^{\theta}(X_{s_i}^{(i)}) - \mathcal{J}_{t_i}(X_{s_i}^{(i)}) \right\|_F^2$
    如果经过 10 次迭代后损失未改变**则**
      $\eta \leftarrow$ 最大值$\{\eta \times 0.99, 10^{-9}\}$
    结束如果
  结束循环
直到 损失 $< \delta$ 或 n_{iter} 达到

算法 2 G^3 采样

输入： 新图数量 N^*，最大扩散
时间 T，欧拉步数 M，基分布浓
度 $\alpha > 0$
$\delta \leftarrow 1/M$
$c \leftarrow \frac{1}{n^2 N}\sum_{j=1}^{N} \|A^{(j)}\|_1$
{基于平均度的稀疏性阈值}
对于 i=0 **到** N^*，**执行**
  采样 $X_0^{(i)} \sim p_{\text{init}}^{\alpha}$
  对于k=0 **到** M，**执行**
    $X_{k+1}^{(i)} \leftarrow \Pi_{\text{clip}}(X_k^{(i)} + \delta \mathcal{J}_t^{\theta}(X_k^{(i)}))$
    {采用元素裁剪的欧拉步}
  结束循环
  $\hat{A}_{ab}^{(i)} \leftarrow \mathbf{1}\{(\Pi_{\text{sym}}(X_M^{(i)})_{ab} \geq c\}$
  {对称化；对角线置零；阈值化}
结束循环

邻接矩阵；其创新之处在于，谱编码了显著的全局拓扑。

近期的进展也聚焦于扩散和流匹配思想的离散化改编，包括 DiGress（Vignac 等人，2022年）和 DeFoG（Qin 等人，2024年）。DeFoG 凭借其 Transformer 架构和相对随机游走概率特征（RRWP），在实证表现上十分出色，并提供理论保证，但计算开销较大，尤其是在较大的图上。

我们使用 SPECTRE 和 DeFoG 作为高质量、通用的基线方法。为使研究在计算上可行，我们将注意力集中于少量竞争方法，并报告自行重新运行其代码所得的结果，而非引用早期论文中的结果，以支持公平比较和可复现性。

Law等人（Law等人，2025）近期的研究探讨了在一种相关但本质上不同的设定中，使用热核进行图生成。他们的方法面向有向图，并依赖作用于学习得到的节点表示的随机热核，同时采用单独的边模型将这些表示映射回有效图。相比之下，我们的框架通过生成元匹配，直接定义并显式逆转作用于图导出矩阵状态上的扩散过程，而无需引入辅助的边重建机制。因此，我们的方法同样适用于无向图，原则上也可在同一生成元中心的表述框架下扩展到有向情形。此外，（Law等人，2025）采用的边扰动与重建步骤可能会改变或掩盖全局结构约束（例如平面性），而我们的基于拉普拉斯的扩散从构造上保留了这些约束。

5. 实验

5.1. 数据集

为评估 G^3 性能，我们考虑了三个合成数据集（随机块模型、度校正随机块模型、平面图）和三个真实世界图数据集（酶、蛋白质、QM9）；其中许多数据集常用于图生成文献中的基准评测。例外的是度校正随机块模型（DCSBM）数据集，它能够更合理地对社区结构网络进行建模（Karrer&Newman，2011）。详情见第B节。

5.2. 评估

为了评估 G^3 图生成性能，我们采用留出图样本与生成样本在各种图属性上的最大均值差异（MMD）进行评估，这也是文献中的标准做法。我们通过比较留出测试图与生成图在以下五种图特征上的属性分布来评估效果：聚类系数、度分布、4节点轨道计数、图拉普拉斯谱和三角形计数。更多详情请参见第C节。

5.3. 神经网络架构

该 G^3 生成元替代分布 $\mathcal{J}_t^{\theta}$使用算法1 训练得到，是一个简单的 4 层多层感知机神经网络，采用 ReLU 激活函数和层归一化。该网络作用于扩散后的网络输入状态的展平形式，并将其重塑为方阵以供输出。通常，每个隐藏层有 4096 个单元；第 5.4 节探讨了这一设计选择的敏感性。5.4。更多详情见第D节。

5.4. 图生成敏感性

为了了解 G^3 对图拓扑、训练动态和超参数的敏感程度，我们考察了系统地改变这些因素时该方法的有效性。我们重点关注谱 MMD，因为它是评估敏感性时最全面的单一指标。更多图表见第F节。

对最大扩散时间的敏感性。图2 展示了谱 MMD 对 T 的选择以及其与图数量 N 的关系有多敏感； 是前向训练过程中状态扩散所达到的最大时间。这为我们提供了一个简单的一般原则： T 应随训练集规模增大，从而无需将其作为可调超参数处理。

图5a 表明，对节点数量 n 的依赖程度没有预期的那么明显；总体而言，我们可以选择忽略它对这一参数选择的影响。曲线的最小值似乎出现在大致相同的 T 处，这表明图的数量 N 是更重要的因素，而且至少对于平面图而言，随着n增加，问题会变得“更难”。

对神经网络架构的敏感性。图3展示了谱 MMD 对 MLP 每层隐藏神经网络单元数量 w 的选择以及其与图数量 N 的关系有多敏感。最小值出现在隐藏单元数量约为 2^{12} =4096 时，且不受 N 的影响。

基于这种独立性，图6a 计算了一个用于固定 N 的小型表格，供我们后续实验中的所有其他数据集使用，以避免不必要的计算。换言之，我们不会针对每个数据集逐一调整神经网络架构。

对基分布的敏感性。图9 展示了基采样分布中的超参数 α 如何影响模型在随机块模型和平面图上的效果。参数 α 的取值不仅会影响谱性能，也会影响生成独特新图的能力（见图10）。

经验上，对于平面类图（包括生物分子数据集）， $\alpha < 1$ 更合适；而对于社区类图，例如随机块模型和度校正随机块模型， $\alpha = 1$ 更合适。

5.5. 性能

现在，我们在一系列合成数据和真实世界数据集上，将采用热扩散策略的对称形式和非对称形式的 G^3，与 SPECTRE 和 DeFoG 的性能进行比较。非对称版本

[1]使用 SciPy 软件包中的高斯一维滤波器函数。

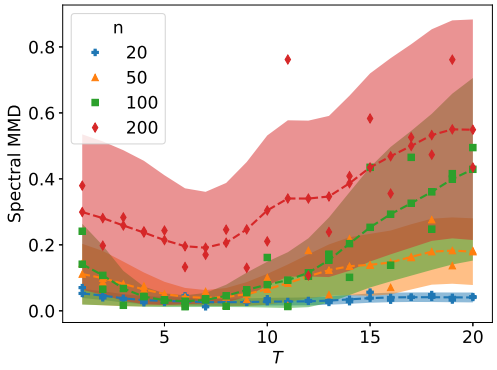


图 2。谱 MMD 随最大扩散时间 T和节点数为 $n = 64$ 的平面图数量 N变化。曲线是对所示观测值进行高斯平滑 [1] 后得到的函数；这些观测值本身是对 10 个随机种子的结果取平均所得。阴影区域表示基于中心极限定理假设的 66% 置信区间。训练时的学习率设为 1×10^{-4}。

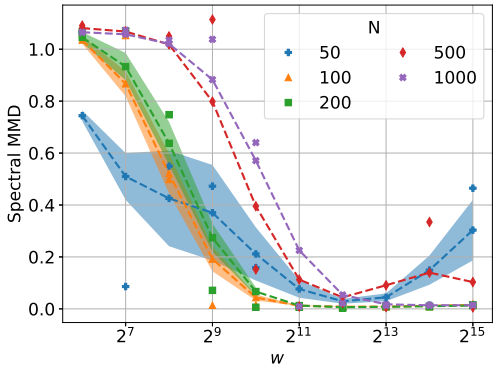


图 3。谱 MMD 随每层隐藏神经网络单元数量 w 以及具有 $n = 64$ 个节点的平面图数量 N 的变化。曲线为所示观测值经过高斯平滑[1] 后的函数；这些观测值本身是对 10 个随机种子的结果取平均所得。阴影区域表示在中心极限定理假设下的 66% 置信区间。训练时的学习率设为 1×10^{-4}。

的 G^3 详见第E.1节。

我们计算了 ±留出测试图与生成图之间最大均值差异（MMD）的中位数绝对偏差，结果见表1，数值越低越好。尽管在所有数据集和指标上都没有明显的赢家，G^3 不仅具有竞争力，而且运行速度显著更快。对于，每个样本图的训练和生成时间仅需数秒，而 DeFoG 和 SPECTRE 则需要 15 至 30 分钟（见图11）。

为突出这一点，我们在 DeFoG 和 SPECTRE 先前表现良好的数据集上重复了实验，但采用了更有限的计算预算。训练轮数减少了 10 到 100 倍，尽管它们的总训练时间通常仍长于 G^3。我们将这些受限变体分别记为 DeFoG* 和 SPECTRE*，并在

表 *1.* G^3、DeFoG 和 SPECTRE 在 $3-5$ 个随机种子划分上的最大均值差异指标的中位数（± 中位绝对偏差）。数值越低越好，**粗体** 表示最佳表现；斜体 表示次佳表现，基于中位数加一个中位绝对偏差确定。

数据集	模型	聚类	度	轨道	谱	三角形
SBM	G^3	**0.0356 ± 0.0005**	**0.0318 ± 0.0052**	0.0631 ± 0.0120	*0.0796 ± 0.0120*	*0.0270 ± 0.0078*
	非对称。G^3	*0.0398 ± 0.0010*	*0.0390 ± 0.0052*	0.0710 ± 0.0140	**0.0832 ± 0.0015**	0.0287 ± 0.0084
	DeFoG	0.0500 ± 0.0000	0.0258 ± 0.0190	*0.0500 ± 0.0000*	0.0884 ± 0.0740	**0.0097 ± 0.0083**
	SPECTRE	0.0588 ± 0.0089	0.2030 ± 0.1700	**0.0500 ± 0.0000**	0.7010 ± 0.4800	0.3320 ± 0.2300
度校正随机块模型	G^3	*0.1270 ± 0.0091*	*0.0344 ± 0.0062*	*0.0971 ± 0.0057*	0.1230 ± 0.0330	*0.0110 ± 0.0098*
	非对称。G^3	0.2340 ± 0.0160	0.0810 ± 0.0200	0.1420 ± 0.0170	*0.1010 ± 0.0220*	0.0578 ± 0.0250
	DeFoG	**0.0556 ± 0.0067**	**0.0063 ± 0.0050**	**0.0591 ± 0.0110**	**0.0499 ± 0.0240**	**0.0006 ± 0.0002**
	SPECTRE	1.0800 ± 0.0340	0.6640 ± 0.0007	1.0400 ± 0.0004	1.2500 ± 0.0250	0.7340 ± 0.0130
平面	G^3	0.3090 ± 0.0300	*0.0030 ± 0.0015*	*0.0035 ± 0.0018*	*0.0060 ± 0.0015*	0.0467 ± 0.0360
	非对称 G^3	1.0200 ± 0.0440	0.0074 ± 0.0050	0.0106 ± 0.0089	0.0113 ± 0.0021	0.1110 ± 0.0320
	DeFoG	**0.0413 ± 0.0047**	**0.0007 ± 0.0001**	**0.0002 ± 0.0000**	**0.0006 ± 0.0003**	**0.0056 ± 0.0013**
	SPECTRE	*0.1420 ± 0.0220*	0.0242 ± 0.0190	0.4240 ± 0.3800	0.0560 ± 0.0019	*0.0226 ± 0.0026*
酶	G^3	**0.0225 ± 0.0008**	*0.0854 ± 0.0550*	0.0902 ± 0.0390	*0.0755 ± 0.0220*	*0.0490 ± 0.0380*
	非对称。G^3	0.0803 ± 0.0200	0.0934 ± 0.0520	*0.1040 ± 0.0047*	0.1090 ± 0.0480	0.0752 ± 0.0120
	DeFoG	0.0765 ± 0.0036	0.7220 ± 0.0070	0.5530 ± 0.0140	1.2000 ± 0.0270	0.4650 ± 0.0033
	SPECTRE	*0.0275 ± 0.0002*	**0.0168 ± 0.0013**	**0.0796 ± 0.0120**	**0.0490 ± 0.0030**	**0.0236 ± 0.0029**
蛋白质	G^3	**0.0465 ± 0.0009**	**0.0220 ± 0.0012**	*0.5530 ± 0.0720*	*0.1600 ± 0.0074*	*0.0928 ± 0.018*
	非对称。G^3	*0.0750 ± 0.0073*	*0.0345 ± 0.0003*	**0.3480 ± 0.2000**	**0.0921 ± 0.0110**	**0.0419 ± 0.0071**
	DeFoG	0.1050 ± 0.0032	0.7500 ± 0.0003	1.2700 ± 0.0000	1.2000 ± 0.0110	0.5800 ± 0.0006
	SPECTRE	0.1070 ± 0.0001	0.0867 ± 0.0480	0.8400 ± 0.0085	0.1310 ± 0.0430	0.1240 ± 0.0110
QM9	G^3	0.2530 ± 0.0074	0.1870 ± 0.0400	0.1650 ± 0.0330	0.1100 ± 0.0027	0.1100 ± 0.0066
	非对称。G^3	*0.0834 ± 0.0170*	*0.0214 ± 0.0033*	0.0036 ± 0.0013	*0.0168 ± 0.0033*	*0.0521 ± 0.0130*
	DeFoG	0.5460 ± 0.3000	0.0449 ± 0.0170	*0.0036 ± 0.0001*	0.0640 ± 0.0049	0.1730 ± 0.0950
	SPECTRE	**0.0200 ± 0.0110**	**0.0027 ± 0.0001**	**0.0008 ± 0.0001**	**0.0046 ± 0.0003**	**0.0051 ± 0.0043**

表7中。在这些条件下，G^3 往往优于其他方法，尤其是在度分布和谱准确性方面。

尽管总体而言，对称热扩散的性能优于非对称版本，但优势并不大；在某些数据集（QM9、酶和蛋白质）上，非对称过程更受青睐。人们或许会认为生物分子结构更适合非对称方法，然而对称版本在平面合成数据集上的更优表现实际上排除了以平面性作为解释。

6. 讨论

在这项工作中，我们通过将对称热扩散与生成元匹配相结合，提出了一种新颖且基础扎实的图生成框架。将图拉普拉斯算子用作生成元，注入了领域特定的归纳偏置，从而确保加噪过程内在地遵循图拓扑（例如局部连通性与几何结构）。通过受控仿真研究和真实世界图生成基准，我们从经验上表明，我们的方法具有竞争力且高度可扩展。

这一方法的一个引人之处在于，它为进一步发展提供了许多途径。我们的通用框架在构造上具有置换不变性，但当前实现使用简单的多层感知机来匹配生成元。实验表明，即使在这种简单设置下也能取得成功，但我们认为，采用置换等变神经网络架构（例如图神经网络）有望进一步提升效果。

在第3.2–3.3 节中概述的理论框架依赖于与组合图拉普拉斯算子相关联的热核。也可以使用随机游走拉普拉斯算子等替代方案，因为底层的 G^3 框架更为通用（如第E.2节所述）。

类似地，我们关注无向图上的一种对称热扩散形式，但通过放宽对称性假设，我们可以将这项工作扩展到有向图。类似地，我们预计其他更为特殊的结构（包括超图）也可以通过适当的超图拉普拉斯算子采用 G^3 策略。

作者特别感兴趣的一项扩展，是纳入图和节点协变量信息以实现条件生成；在生成式文献中，这称为“引导”。第 E.3 节提供了一个示例，展示了这种方法可能如何运作：我们生成了组成员关系预先指定的随机块模型样本。

致谢

CN 衷心感谢英国工程与自然科学研究委员会（EPSRC）项目 EP/V022636/1 和 EP/Y028783/1 的资助。本研究还获得了英格兰研究机构通过英格兰卓越扩展（E3）资助渠道提供的支持；该资助授予了兰卡斯特大学数学科学学院的 MARS：现实世界系统中的人工智能数学项目。

影响声明

本文提出了一个广泛且适应性强的图生成框架，旨在支持包括药物发现、通用图合成、电路设计和网络安全在内的广泛应用。图生成方面的进展有望影响众多科学与技术领域，并可能带来既有益又具有挑战性的社会或伦理影响，尤其是在生物医学和化学研究中。然而，在现阶段，我们预计本文所介绍的方法不会引发任何直接的社会关切。

参考文献

Binkiewicz, N.、Vogelstein, J. T. 与 Rohe, K. 协变量辅助谱聚类。《*Biometrika*》，104(2):361–377，2017。

Brown, N., Fiscato, M., Segler, M. H. 和 Vaucher, A. C. GuacaMol：对用于从头分子设计的模型进行基准测试。《化学信息与建模杂志》，59(3):1096–1108，2019年。

Chen, R. T. Q., Rubanova, Y., Bettencourt, J. 和 Duvenaud, D. 神经常微分方程。*arXiv* 预印本 *arXiv:1806.07366*，2018年。

Chung, F. R. 谱图论，第92卷。美国数学学会，1997年。

De Cao, N. 和 Kipf, T. MolGAN：一种用于小型分子图的隐式生成模型。*arXiv* 预印本 *arXiv:1805.11973*，2018年。

Evans, L. C. 偏微分方程。美国数学学会，第2版，2010年。

Forman. 胞腔复形的 Bochner方法与组合 Ricci 曲率。《离散与计算几何》,29(3):323–374, 2003.

Fu, D., Bao, W., Maciejewski, R., Tong, H. 和 He, J. 从数据到计算的隐私保护图机器学习：一项综述。*ACMSIGKDD ExplorationsNewsletter*，25(1):54–72，2023年。

Hadamard, J. 线性偏微分方程中的柯西问题讲义。耶鲁大学出版社，1923年。Holderrieth, P., Havasi, M., Yim, J., Shaul, N., Gat, I., Jaakkola, T., Karrer, B., Chen, R. T. 和 Lipman, Y. 生成元匹配：基于任意马尔可夫过程的生成建模。*arXiv* 预印本 *arXiv:2410.20587*，2024年。

Ji, S., Pan, S., Cambria, E., Marttinen, P., 和 Yu, P. S. 知识图谱综述：表示、获取与应用。*IEEE* 神经网络与学习系统汇刊, 33(2):494–514, 2021.

Jo, J., Lee, S., 和 Hwang, S. J. 通过随机微分方程系统进行基于分数的图生成建模。载于国际机器学习会议, pp. 10362–10383. PMLR, 2022.

Karrer, B. 和 Newman, M. E. 网络中的随机块模型与社区结构。《物理评论*E*》—统计、非线性与软物质物理, 83(1): 016107, 2011.

Kipf, T. N. 与 Welling, M. 变分图自编码器。*arXiv* 预印本 *arXiv:1611.07308*，2016年。

Law, M. T.、Kreis, K. 与 Maron, H. 基于热核的有向图生成。《机器学习研究汇刊》，2025。

Lipman, Y.、Chen, R. T.、Ben-Hamu, H.、Nickel, M. 与 Le, M. 面向生成建模的流匹配。*arXiv* 预印本 *arXiv:2210.02747*，2022年。

Martinkus, K.、Loukas, A.、Perraudin, N. 与 Wattenhofer, R. SPECTRE：谱条件化有助于克服单步图生成器的表达能力限制。载于国际机器学习会议，第15159–15179页。PMLR，2022年。

Morris, C.、Kriege, N. M.、Bause, F.、Kersting, K.、Mutzel, P. 与 Neumann, M. TUDataset：用于图学习的基准数据集集合。*arXiv* 预印本 *arXiv:2007.08663*，2020年。

Nettleton, D. F. 对表示为图的社交网络进行数据挖掘。*Computer Science Review*，7:1–34，2013。

Qin, Y.、Madeira, M.、Thanou, D. 和 Frossard, P. De-FoG：用于图生成的离散流匹配。*arXiv* 预印本 *arXiv:2410.04263*，2024年。

Schneider, E. W. 应用课程模块化：界面系统及其对序列控制和数据分析的影响。1973。

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. 通过随机微分方程实现基于分数的生成建模。*arXiv* 预印本 *arXiv:2011.13456*，2020年。Sriperumbudur, B. K., Gretton, A., Fukumizu, K., Sch¨olkopf, B., and Lanckriet, G. R. 希尔伯特空间嵌入与概率测度上的度量。机器学习研究期刊，11:1517–1561，2010年。Sylvester, J. 化学与代数。《自然》，17:284–284, 1878。Thompson, R., Knyazev, B., Ghalebi, E., Kim, J., and Taylor, G. W. 图生成模型的评估指标。*arXiv* 预印本 *arXiv:2201.09871*，2022年。Tolstikhin, I. O., Sriperumbudur, B. K., and Sch¨olkopf, B. 使用径向核对最大均值差异进行极小极大估计。神经信息处理系统进展，29，2016年。Vignac, C., Krawczuk, I., Siraudin, A., Wang, B., Cevher, V., and Frossard, P. DiGress：用于图生成的离散去噪扩散。*arXiv* 预印本 *arXiv:2209.14734*，2022年。Wang, H., Wang, J., Wang, J., Zhao, M., Zhang, W., Zhang, F., Xie, X., and Guo, M. GraphGAN：利用生成对抗网络进行图表示学习。载于人工智能促进协会人工智能会议论文集，第32卷，2018年。Wills, P. and Meyer, F. G. 图比较度量：实践者指南。《公共科学图书馆·综合》，15(2):e0228728，2020年。Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., Leswing, K., and Pande, V. MoleculeNet：分子机器学习基准。《化学科学》，9(2):513–530，2018年。Yang, L., Zhang, Z., Zhang, W., and Hong, S. 采用自引导潜在扩散的基于分数的图生成建模。2023年。You, J., Ying, R., Ren, X., Hamilton, W., and Leskovec, J. GraphRNN：利用深度自回归模型生成真实图。载于国际机器学习会议，第5708–5717页。PMLR，2018年。Zheng, X., Zhang, M., Chen, C., Molaei, S., Zhou, C., and Pan, S. GNNEvaluator：无需标签即可评估未见图上的图神经网络性能。神经信息处理系统进展，36:33606–33623，2023年。

A. 证明与进一步细节

A.1. 热方程与热核

命题 A.1（通过热核求解对称热扩散）。 设 $L \in \mathbb{R}^{n\times n}$ 为无向图的（组合）拉普拉斯矩阵，并定义热核 H_s : $= e^{-sL}$ ，其中 $s \geq 0$。对于任意初始矩阵 $Y_0 \in \mathbb{R}^{n\times n}$，矩阵值对称热方程

$$\frac{\mathrm{d}}{\mathrm{d}s}Y_s \;=\; -\big(LY_s + Y_sL\big), \qquad s \geq 0, \tag{20}$$

存在唯一解

$$Y_s \;=\; H_s\, Y_0\, H_s \;=\; e^{-sL}\, Y_0\, e^{-sL}, \qquad s \geq 0. \tag{21}$$

特别地，由 $\mathcal{P}_s$（Y）：$= H_sYH_s$ 定义的线性映射族 （$\mathcal{P}_s$）$_{s\geq 0}$ 构成与生成元 $Y \mapsto LY + YL$相关联的（矩阵）热半群。

证明。 由于 L是实对称矩阵，因此它可正交对角化： $L \;=\; U\Lambda U^\top$，其中 U正交归一，且 Λ =为对角矩阵（λ_1，…， λ_n）， $\lambda_i \geq 0$。因此

$$H_s := e^{-sL} = Ue^{-s\Lambda}U^\top, \qquad \text{and} \qquad \frac{\mathrm{d}}{\mathrm{d}s}H_s = -LH_s = -H_sL.$$

定义 $Y_s := H_sY_0H_s$。求导并使用乘积法则，

$$\frac{\mathrm{d}}{\mathrm{d}s}Y_s = \Big(\frac{\mathrm{d}}{\mathrm{d}s}H_s\Big)Y_0H_s + H_sY_0\Big(\frac{\mathrm{d}}{\mathrm{d}s}H_s\Big) = (-LH_s)Y_0H_s + H_sY_0(-H_sL) = -(LY_s + Y_sL),$$

因此 Y_s 满足（3）。此外 $Y_0 \;=\; H_0Y_0H_0 \;=\; Y_0$。根据标准线性常微分方程理论，在 $\mathbb{R}^{n\times n}$上可得唯一性（等价地，将方程向量化为 $\mathbb{R}^{n^2}$上的线性系统）。 □

A.2. 对称热扩散的耗散性与稳定性

在第 2.2节中，我们考虑矩阵值热方程

$$\frac{\mathrm{d}}{\mathrm{d}s}Y_s \;=\; -\big(LY_s + Y_sL\big), \qquad Y_0 \in \mathbb{R}^{n\times n}, \tag{22}$$

其中 $L \in \mathbb{R}^{n\times n}$ 是 无向图的（组合）图拉普拉斯算子，因此是对称半正定的 e.

我们在 $\mathbb{R}^{n\times n}$ 上赋予 Frobenius内积

$$\langle A, B\rangle_F := \mathrm{tr}(A^\top B), \qquad \|A\|_F^2 := \langle A, A\rangle_F.$$

定义线性算子 $\mathcal{A}$： $\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ 如下

$$\mathcal{A}(Y) := LY + YL, \qquad \text{and} \qquad \mathcal{G}(Y) := -\mathcal{A}(Y) = -(LY + YL).$$

引理 A.2 (对称热算子的耗散性**)**。 若 L 是对称半正定的，则 $\mathcal{G}$ 相对于 $\langle\cdot,\cdot\rangle_F$是耗散的，即

$$\langle \mathcal{G}(Y),\, Y\rangle_F \;\leq\; 0 \qquad \textit{for all } Y \in \mathbb{R}^{n\times n}.$$

等价地， $\langle \mathcal{A}(Y)\ Y\rangle_F \geq 0$ 对所有 Y均成立。

证明。 我们计算

$$\langle \mathcal{A}(Y), Y\rangle_F = \langle LY, Y\rangle_F + \langle YL, Y\rangle_F = \mathrm{tr}(Y^\top LY) \;+\; \mathrm{tr}(Y^\top YL).$$

对于第一项，将 $Y = [y_1, \ldots, y_n]$ 写成列向量。然后

$$\mathrm{tr}(Y^\top LY) = \sum_{j=1}^{n} y_j^\top Ly_j \;\geq\; 0$$

由于 $L \succeq 0$（即半正定）。

对于第二项，注意 $Y^\top Y \succeq 0$ 和 $L \succeq 0$。利用恒等式

$$\operatorname{tr}(AB) = \|A^{1/2}B^{1/2}\|_F^2 \quad \text{for } A \succeq 0,\ B \succeq 0,$$

我们得到

$$\operatorname{tr}(Y^\top Y L) = \operatorname{tr}\big((Y^\top Y)L\big) \ \succeq\ 0.$$

因此 $\langle \mathcal{A}(Y)\ Y\rangle_F \geq 0$，从而 $\langle \mathcal{G}(Y)\ Y\rangle_F = -\langle \mathcal{A}(Y)\ Y\rangle_F \leq 0$。 □

推论 A.3（能量耗散与弗罗贝尼乌斯稳定性）。 设 $(Y_s)_{s\geq 0}$ 是 (22) 的解。那么弗罗贝尼乌斯范数单调不增：

$$\frac{\mathrm{d}}{\mathrm{d}s}\|Y_s\|_F^2 = 2\langle \dot{Y}_s, Y_s\rangle_F = 2\langle \mathcal{G}(Y_s), Y_s\rangle_F \leq 0, \qquad \textit{hence} \qquad \|Y_s\|_F \leq \|Y_0\|_F \ \ \forall s \geq 0.$$

特别是，(22) 在 $\mathbb{R}^{n\times n}$ 上定义了稳定线性扩散，其意义是在 $\|\cdot\|_F$ 中具有压缩性 .

证明。 对 $\|Y_s\|_F^2$ 求导并应用引理 A.2。 □

引理 A.4（闭式解与半群压缩）。 设 $H_s := e^{-sL}$。方程 (22) 的唯一解为

$$Y_s = H_s\, Y_0\, H_s, \qquad s \geq 0.$$

此外，映射 $\mathcal{P}_s(Y) \quad := H_s Y H_s$ 在 $\mathbb{R}^{n\times n}$ 上构成强连续半群，并满足压缩界 d

$$\|\mathcal{P}_s(Y)\|_F \leq \|Y\|_F, \qquad s \geq 0.$$

证明。 表示 $Y_s = H_s Y_0 H_s$ 可通过对 $H_s Y_0 H_s$ 求导并使用$\mathrm{d}\,\mathrm{d}s H_s = -LH_s = -H_s L$得到。对于压缩，由于 L 是对称的，且特征值为 $\lambda_i \geq 0$，因此 H_s 的特征值为 $e^{-s\lambda_i} \in (0, 1],$，故有 $\|H_s\|_2 \leq 1$。

回顾弗罗贝尼乌斯范数的次乘性结论。对于所有矩阵 A 和 B，$\|AB\|_F^2 = \operatorname{tr}(B^\top A^\top AB)$。现在，由于 $\|A\|_2^2 I \succeq A^\top A$，我们得到 $\operatorname{tr}(B^\top A^\top AB) \leq \|A\|_2^2 \operatorname{tr}(B^\top B) = \|A\|_2^2 \|B\|_F^2.$。两边开平方即可得到结论$\|AB\|_F \leq \|A\|_2 \|B\|_F.$。利用这一结果，我们因此有，

$$\|H_s Y H_s\|_F \leq \|H_s\|_2\, \|Y H_s\|_F \leq \|H_s\|_2^2\, \|Y\|_F \leq \|Y\|_F.$$ □

A.3. 半群与无穷小生成元

引理 A.5（半群结构与无穷小生成元）。 设 $\{H_s\}_{s\geq 0}$ 为图热核 $H_s = e^{-sL}$，并定义线性算子族

$$P_s : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}, \qquad P_s(Y) := H_s\, Y\, H_s.$$

则 $\{P_s\}_{s\geq 0}$ 在 $\mathbb{R}^{n\times n}$ 上构成强连续单参数半群，其中 $P_0 = \mathrm{Id}$ 且 $P_{s+t} = P_s \circ P_t$。其无穷小生成元 $\mathcal{A}$ 为线性算子

$$\mathcal{A}(Y) = LY + YL,$$

定义域为 $\mathrm{Dom}(\mathcal{A}) = \mathbb{R}^{n\times n}$，而对称热方程

$$\frac{\mathrm{d}}{\mathrm{d}s} Y_s = -\mathcal{A}(Y_s)$$

是与该半群相关联的抽象演化方程。

证明。 半群性质可直接由 $H_{s+t} = H_s H_t$ 得出：

$$P_{s+t}(Y) = H_{s+t} Y H_{s+t} = H_s(H_t Y H_t) H_s = (P_s \circ P_t)(Y),$$

and $P_0(Y) = IYI = Y$。由于 $s \mapsto H_s$ 在算子范数下连续，因此强连续性成立。

对于生成元，对任意 $Y \in \mathbb{R}^{n\times n}$ 进行计算，

$$\lim_{h\downarrow 0} \frac{P_h(Y) - Y}{h} = \lim_{h\downarrow 0} \frac{H_h Y H_h - Y}{h} = -LY - YL,$$

使用 $H_h = I - hL + o(h)$。因此， $\mathcal{A}(Y) = LY + YL$。 □

A.4. 正向和反向过程的结果

引理 A.6。 对于具有组合拉普拉斯矩阵 L 的连通图， 对称热扩散 $Y_T = H_T Y_0 H_T$ 与其前向过程极限 $\lim_{s\to\infty} Y_s = \Pi Y_0 \Pi$ 之间的弗罗贝尼乌斯范数满足

$$\|Y_T - \Pi Y_0 \Pi\|_F \le 2e^{-T\lambda_2}\|Y_0\|_F,$$

其中， λ_2 是 L的最小非零特征值。

证明。 考虑以下分解

$$\begin{aligned}\|Y_T - \Pi Y_0 \Pi\|_F &= \|H_T Y_0 H_T - \Pi Y_0 \Pi\|_F \\ &\le \|H_T Y_0 H_T - \Pi Y_0 H_T\|_F + \|\Pi Y_0 H_T - \Pi Y_0 \Pi\|_F \\ &= \|(H_T - \Pi) Y_0 H_T\|_F + \|\Pi Y_0 (H_T - \Pi)\|_F \\ &\le \|H_T - \Pi\|_2 \|Y_0\|_F \|H_T\|_2 + \|\Pi\|_2 \|Y_0\|_F \|H_T - \Pi\|_2 ,\end{aligned}$$

利用任意矩阵都满足的性质， $\|ABC\|_F \le \|A\|_2 \|B\|_F \|C\|_2$。
Π是投影到常数 $\frac{1}{n}\mathbf{1}\mathbf{1}^\top$上的秩一投影，其特征值为 $\mathbf{1}$，这意味着 $\|\Pi\|_2 = 1$。 $H_T = e^{-TL}$的特征值为 $e^{-T\lambda_i}$，这意味着 $\|H_T\|_2 = 1$。由于$\mathbf{1}$是 e^{-TL}和 Π的特征值为 $\mathbf{1}$ 的特征向量，$H_T - \Pi$ 可以同时对角化。因此，
$\|H_T - \Pi\|_2 = e^{-T\lambda_2}$ 对应于 L的最小非零特征值。将这些等式代入初始分解即可完成证明。 □

引理 A.7。 设 $X_0 = \Pi + X_\perp$ 是反向时间常微分方程的一个初始状态，其中 $X_\perp$ 的行和与列和均为 0，即 $\mathbf{1}^\top X_\perp = \mathbf{0}$ 和 $X_\perp \mathbf{1} = \mathbf{0}$。使用真实生成器 $\mathcal{J}X = T(LX + XL)$对热扩散进行逆向操作，会得到满足以下条件的 $X_1 = e^{TL} X_0 e^{TL}$

$$\|X_1\|_F \ge e^{2T\lambda_2}\|X_\perp\|_F.$$

证明。 注意到

$$X_1 = e^{TL}(\Pi + X_\perp)e^{TL} = \Pi + e^{TL} X_\perp e^{TL},$$

因为 $\mathbf{1}$是 e^{TL} 的特征向量，对应的特征值为 $\mathbf{1}$。因此， $\|X_1\|_F \ge \left\|e^{TL} X_\perp e^{TL}\right\|_F$。

设 $L = U\Lambda U^\top$ 是 L 的特征分解，即 $e^{TL} = U e^{T\Lambda} U^\top$。因此，

$$\left\|e^{TL} X_\perp e^{TL}\right\|_F^2 = \left\|e^{T\Lambda} U^\top X_\perp U e^{T\Lambda}\right\|_F^2 = \sum_{i,j} e^{2T(\lambda_i+\lambda_j)} (U^\top X_\perp U)_{ij}.$$

由于 $\mathbf{1}$是 L 对应于 $\lambda_1 = 0$ 的特征值，这意味着 $(U^\top X_\perp U)_{i1} = 0$ 和 $(U^\top X_\perp U)_{1j} = 0$ 均成立，也就是说， λ_1 项不会对求和产生贡献。对于其余非零项 $\lambda_i + \lambda_j \ge 2\lambda_2$，因此

$$\left\|e^{TL} X_\perp e^{TL}\right\|_F^2 \ge e^{4T\lambda_2} \sum_{i,j} (U^\top X_\perp U)_{ij} = e^{4T\lambda_2}\|X_\perp\|_F^2.$$

开平方即可完成证明。 □

B. 数据集

对于以下列出的所有数据集，在按照第5.4节所述运行实验或选择超参数时，我们使用范围为 $10 - 20$的随机种子。对于模型基准测试，我们使用种子 $1 - 5$。

B.1. 合成数据

随机块模型 我们生成随机块模型网络（随机块模型），其社区内和社区间边概率分别由 $p_{\text{inter}} = 0.05$ 和 $p_{\text{intra}} = 0.3$ 给定且保持一致，以确保社区具有合理的可区分性。社区数量从 $\{2, 3, 4, 5\}$中随机选取。在一般分析中，我们使用随机种子控制可复现性，并构造任意数量、节点数任意的此类图。对于基准测试，我们保留了种子 $1 - 5$，并将 $N = 200$个 200 节点图划分为 80% 的训练数据和 20% 的测试集。

度校正随机块模型 度校正随机块模型的总体构造与标准随机块模型相似，但能额外控制度分布，因此通常被认为是更合理的社区结构网络模型。我们使用与上述随机块模型相同的设置构造此类图的数据集，但边概率会根据从 Beta(1,1) 分布中抽取的值进行加权。

平面 除了上述介绍的随机块模型外，我们还将平面图作为一个重要的独立案例，以及一个在拓扑结构上截然不同的挑战纳入研究。在此情况下，我们通过均匀随机采样二维点，并利用 Delaunay 三角剖分将其连接起来，从而构造平面图。与上文一样，为进行一般分析，我们生成任意数量的 n 节点平面图，并使用与上文相同的随机种子和训练集/测试集划分，保留特定的数据集 $N = 200$、 $n = 64$ 。

B.2. 真实世界

为在真实世界图数据集上进行基准测试，我们使用了广受欢迎的 TUDataset。该数据集由（Morris 等人，2020年）引入，并通过 PyTorch Geometric 库使用，因为其接口便于操作。

蛋白质 大分子代表了一组丰富的可变大小图，其中包括平面网络这一子集。如果分子亚结构沿氨基酸序列彼此相邻，或在空间中属于三个最近邻之一，则在它们之间连接边，从而构造图。该数据集最初作为一个示例问题被提出，用于从一组大分子中识别酶。

酶 该数据集与蛋白质数据集非常相似，只不过其中所有元素都是酶（原始任务是对酶的类型进行分类）。

QM9 分子数据是现实世界平面图的一个重要例子，生成这类图也是该领域的一项常见目标。我们也采用这一做法，纳入 QM9（Wu 等人，2018年）数据集，其中包含小型有机分子（$n \leq 9$）以及其他分子属性。与 SPECTRE 和 DeFoG 类似，我们将数据随机划分为训练集、验证集和测试集，大小分别为 97,682、20,026 和 13,123。生成时，每个模型生成 512 个样本，以便与测试集进行比较。测试指标基于测试集中的最多 500 个图计算，以减少计算时间。

C. 评估

为了评估生成器匹配模型的性能，我们使用合成数据和真实世界图数据集。在定义好图训练集后，我们按照第 3.2 节所述的流程训练神经生成器，随后从适当的极限分布（由生成元的选择决定）中采样，并逆转扩散过程，以生成新图。

我们的目标是创建新的图集合，使其属性与训练集中的图高度相似。由于网络结构复杂，通常很难明确界定这些属性应当是什么。在众多使用这类数据结构的数学学科中，人们已投入大量精力来定义此类属性。在统计学和机器学习领域，我们重点介绍以下研究：（Wills & Meyer，2020年），作者探讨了各种常用指标及其与拓扑特征的关系；（Thompsonetal.，2022年），该研究提出了专门用于衡量图生成模型性能的指标；（Zhengetal.，2023），该研究着重评估图神经网络在节点分类任务中的表现。

沿用早期生成图模型的方法，我们采用最大均值差异（MMD）来对特征经验分布之间的差异进行数值评估。

最大均值差异的基础在于将概率测度嵌入再生核希尔伯特空间（再生核希尔伯特空间）。更精确地说，给定概率测度 P、定义在（可分）拓扑空间 $\mathcal{X}$上且对应再生核希尔伯特空间为 $\mathcal{H}$的连续正定实值核 k， P的核均值嵌入由 μ_P给出：$= \int k(\cdot, x)dP(x)$，前提是 k 为 P-可积。基于此，我们将测度 P与 Q 之间的最大均值差异定义为

$$\mathrm{MMD}(P, Q) := \|\mu_P - \mu_Q\|_{\mathcal{H}} .$$

在实践中，我们通常借助核技巧，用蒙特卡洛估计 $\mu_{P_n} = \frac{1}{n}\sum_{i=1}^{n} k(\cdot, X_i)$ 来近似 μ_P：

$$\mathrm{MMD}^2(P,Q) = \mathbb{E}_{X,X'}\{k(X,X')\} + \mathbb{E}_{Y,Y'}\{k(Y,Y')\} - 2\,\mathbb{E}_{X,Y}\{k(X,Y)\}$$

其中，期望通过蒙特卡洛估计进行近似，X，$X' \sim P$以及 Y，$Y' \sim Q$。在我们的情形中，蒙特卡洛估计的每一项都对应于在下文列出的图属性成对经验分布上计算核，而每个成对经验分布都由单个图计算得出。

关于该度量（及其估计量）的详细分析，可参见（Sriperumbudur等人，2010年）和（Tolstikhin等人，2016年），但只需指出，一般应选择一个通用核，使得 $\mathcal{H}$ 具有特征性，即在$\mathrm{MMD}(P,\ Q) = 0 \iff P = Q$；也就是说，$\mathrm{MMD}(\cdot,\ \cdot)$ 是一个有效的度量。

我们使用最大均值差异来衡量生成的图集合与留出图（来自相关数据集）在以下特征上的相似性：

度： 对于节点 i，其度 d_i 表示与之相连的边的数量。

三角形： T_i 表示以 i为顶点的三角形数量。

聚类（系数）： 节点 i 的聚类系数为 $c_i = 2T_i/d_i(d_i - 1)$.

轨道： 具体而言，4节点轨道计数：大小不超过4的图元中每种轨道出现的次数。

谱： （归一化）图拉普拉斯算子的特征值 λ_i定义了用于最大均值差异评估的谱。

D. 实现细节

D.1. 神经网络架构

我们的神经生成器是一个简单的 4 层 ReLU 多层感知机，每个隐藏层通常有 4096 个节点，并在第 2–3 层和第 3–4 层之间进行层归一化。该网络作用于展平后的输入状态，并输出一个展平状态，再将其重塑为方阵。我们使用 Adam 优化器，初始学习率为 10^{-4}，并采用调度器，在损失停止下降时将学习率缩小 0.99 倍，最低降至 10^{-9}。对于使用了特定于数据的参数设置的情况，这些设置列于表2中。

[为匿名处理已隐去]

数据集	参数 批大小*	训练轮数	α	T	w
平面	256	200	0.1	6	4096
SBM	256	20	1	15	1024
度校正随机块模型	8	100	1	20	4096

表 2. 超参数选择 。*：批大小是最大批大小；对于大小不同的图集合，批次通过 e
仅选择图 中相同大小的图来构成，最多不超过批大小。-

D.2. 节点数量可变

处理节点数量不同的图集合时，我们可以选择不同的处理方式。最朴素的方法是简单地用零行和零列填充邻接矩阵，直到它们都具有 $n = \max\{n_1, \ldots, n_N\}$ 个“节点”。使用这种方法时，热核也必须进行相应的填充，以确保维度匹配。

显然，这既低效又不可取。虽然填充有助于简化处理，但在训练期间（另外）屏蔽这些零值会更加有效；为此，可以对多层感知机架构进行适当修改。具体而言，我们定义了一个 MaskedLinear 模块，用作神经网络架构中标准 PyTorch 线性模块的直接替代品。该模块的功能是将给定输入的节点数量作为参数，并仅索引权重矩阵中与展平后的输入状态相应大小的部分（即 $\frac{1}{2}n(n-1)$）。 _

数据所在位置 数据集包含带有不连通节点的图，我们仅在最大连通子图上进行训练 .

D.3. 相关工作的架构

由于硬件限制，我们无法完全按照随附配置文件中的设定运行 DeFoG 模型，也无法按照 README文件中给出的详细信息运行 SPECTRE 模型。受时间限制，我们将 SPECTRE训练的训练轮数从 6000 减少到 5000。时间和内存限制共同迫使我们对 DeFoG SBM配置文件进行了以下修改（对于 DCSBM，该配置文件也进行了相同处理，因为当时没有预先指定的配置文件）：

参数	原始	修改后
批大小	32	8
训练轮数	50000	40000
层数	8	4
RRWP步数	20	10

表 *3.* DeFoG 随机块模型配置文件中修改后的参数选项表。

DeFoG 的平面图配置文件保持不变。我们注意到，尽管 SPECTRE 的 README文件为平面图提供了不同于随机块模型的训练参数，但使用随机块模型的设置所取得的效果优于平面图设置。对于 Enzymes数据集，由于 SPECTRE 和 DeFoG 均未记录该数据集的配置，我们采用了与 Proteins数据集所提供的参数设置非常相似的设置，因为这两个数据集中的图相对相似。具体差异如下：对于 SPECTRE，由于时间限制，我们将最大训练轮数从 1020 减少至 400；对于 DeFoG，虽然 Proteins数据集已有数据集基础设施，但没有配置文件。因此，我们采用了平面图设置，但将 Proteins数据集的训练轮数从 100000 减少至 1000，并将 Enzymes数据集的训练轮数减少至 2000。

DeFoG* 和 SPECTRE* 在第 5.5 节，我们介绍了 DeFoG* 和 SPECTRE*，它们分别是各自模型的计算受限版本。表 4 展示了我们如何约束相关参数以缩短计算时间。

模型	Data	原始	修改后
SPECTRE*	蛋白质	1020	10
SPECTRE*	酶 400		4
DeFoG*	SBM	40000	1000
DeFoG*	度校正随机块模型	40000	1000
DeFoG*	平面	100000	1000

表*4.* 按数据集列出的 DeFoG* 和 SPECTRE* 训练轮次缩减表。

E. 框架扩展

E.1. 非对称热核

在第 2 节中，我们考虑图上的热方程，并定义了一种对称热扩散，该过程将热核应用于某种图矩阵表示的左右两侧。接着，我们概述这一过程的行为，以及它如何融入生成器匹配方法。

在本节中，我们考虑更直观的非对称热扩散，其中 $Y_s := H_s Y_0$（沿用第 2 节中的记号），并主要通过表5简要总结相应的量以及对 G^3所做的修改。许多非对称量都很直观，但 X_0的选择取决于矩阵表示 Y_0的选择，因此需要进一步说明。例如，对于 $Y_0 = A$，即标准邻接矩阵，其极限矩阵为 $\frac{1}{n}\mathbf{1}\boldsymbol{d}^\top$，其中 $\boldsymbol{d}$是 A的度向量。由此，我们可以定义一个基分布，其元素为 $X_0 := \tilde{\boldsymbol{x}}_0 \boldsymbol{v}_0^\top$，其中 $\tilde{\boldsymbol{x}}_0 \sim$服从狄利克雷分布$(\alpha\mathbf{1})$，且 $\log \boldsymbol{v}_0 \sim \mathcal{N}(\hat{}\mu_d, \hat{}\sigma_d^2)$，其中 $\hat{}\mu_d$、$\hat{}\sigma_d^2$分别是训练集中缩放后的度向量均值和方差的对数，即 $\hat{}\mu_d := \log(n^{-1}\langle d\rangle)$。另一方面，选择 $Y_0 = D^{-1}A$ （

随机游走矩阵）会得到极限矩阵 $n^{-1}\mathbf{1}\mathbf{1}^\top$，自然的选择是纯狄利克雷基分布，与正文中所述的类似，只是不进行对称化和缩放。

数量	对称	非对称
$\mathcal{P}_s(Y)$	$H_s Y H_s$	$H_s Y$
$\frac{\mathrm{d}}{\mathrm{d}t} Y_t$	$-LY_t - Y_t L$	$-LY_t$
极限$_{t\to\infty} Y_t$	$\frac{1}{n}(\mathbf{1}^\top Y_0 \mathbf{1})\mathbf{1}\mathbf{1}^\top$	$\frac{1}{n}\mathbf{1}\mathbf{1}^\top Y_0$
$X_0(A)$	$\langle d\rangle(\tilde{X}_0 + \tilde{X}_0^\top)$	$\tilde{\boldsymbol{x}}_0 \boldsymbol{v}_0^\top$

表5. 对称和非对称热核过程下关键量的定义。

E.2. 替代图表示与热核

在第 3.2–3.3 节中，我们介绍了热核生成器匹配与采样的一般方案。要构建这一方案的实用实现，用户仍需做出多项选择。其中最重要的一项是选择图的矩阵表示，因为这决定了相应的基分布。如上一节所述，在对称和非对称形式下，选择 $Y_0 = A$ 会产生不同的样本 X_0。也可以采用其他表示，例如随机游走矩阵 $D^{-1}A$ 或对称化形式 $D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$。

为简单起见，正文重点介绍使用组合拉普拉斯矩阵的热核形式。不过，只要其他核满足适当的随机性性质，将这一方法推广到这些核上也很直接。一种替代方案是 随机游走核，定理 E.1 证明了它是有效的选择。若采用这种替代方案，则必须计算相应的极限矩阵和基分布。在表 6 中，我们考虑将组合拉普拉斯矩阵与随机游走拉普拉斯矩阵和归一化拉普拉斯矩阵，$L_{\text{rw}} = I - D^{-1}A$ 和 $L_N = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 以及随机游走核和惰性随机游走核。

引理 E.1. （惰性）随机游走核是有效的概率转移核。

证明。设 $Q\Lambda Q^{-1}$ 为 $D^{-1}A$的特征分解。我们有 $D^{-1}A\mathbf{1} = \mathbf{1}$，因此$\mathbf{1}$ 是特征值为1的特征向量。由特征向量正交性可知 $Q^{-1}\mathbf{1} = c^{-1}\boldsymbol{e}_1$。因此，$Q\Lambda^t Q^{-1}\mathbf{1} = Q\Lambda^t c^{-1}\boldsymbol{e}_1 = Q\lambda_1^t c^{-1}\boldsymbol{e}_1 = \boldsymbol{q}_1/c = \mathbf{1}$，因为 $\lambda_1 = 1$，并且我们已对特征值排序，使得 $\boldsymbol{q}_1 = Q_{:1} = c\mathbf{1}$是 Q的第一个列向量。同样的论证适用于惰性随机游走算子的情形，$\frac{1}{2}(I + D^{-1}A)$。 □

随机游走矩阵的极限形式为 $(D^{-1}A)^t \xrightarrow{t\to\infty} QJ^{11}Q^{-1}$，其中 J^{ij} 是一个矩阵，除第 ij 个元素为 1 外，其余元素均为 0。由此可简化为秩一矩阵 $\boldsymbol{q}_1\boldsymbol{q}_1^{-\top}$，其中 $\boldsymbol{q}_1^{-\top}$是 Q^{-1} 的第一行；进一步可简化为 $c\mathbf{1}\boldsymbol{q}_1^{-\top}$，因为第一个特征向量是全 $\mathbf{1}$ 向量的常数倍。惰性随机游走矩阵也适用同样的论证。实际上，这个量通常与热极限相近；扰动源于随机游走矩阵缺乏对称性，可通过交换子 $[D^{-1}, A] = D^{-1}A - AD^{-1}$ 来量化。

核	有效	极限
e^{-tL}	✓	$n^{-1}\mathbf{1}\mathbf{1}^\top$
$\mathrm{e}^{-tL_{\text{rw}}}$	✓	$n^{-1}\mathbf{1}\mathbf{1}^\top$
e^{-tL_N}	✗	$n^{-1}\mathbf{1}\mathbf{1}^\top$
$(D^{-1}A)^t$	✓	$q_{11}\mathbf{1}\boldsymbol{q}_1^{-\top}$
$\frac{1}{2^t}(I + D^{-1}A)^t$	✓	$q_{11}\mathbf{1}\boldsymbol{q}_1^{-\top}$

表6. 不同核的概率解释有效性与极限行为。

此外，任何此类核都可以被对称化，从而同时作用于主体矩阵的左右两侧，正如正文中对热核的描述那样。

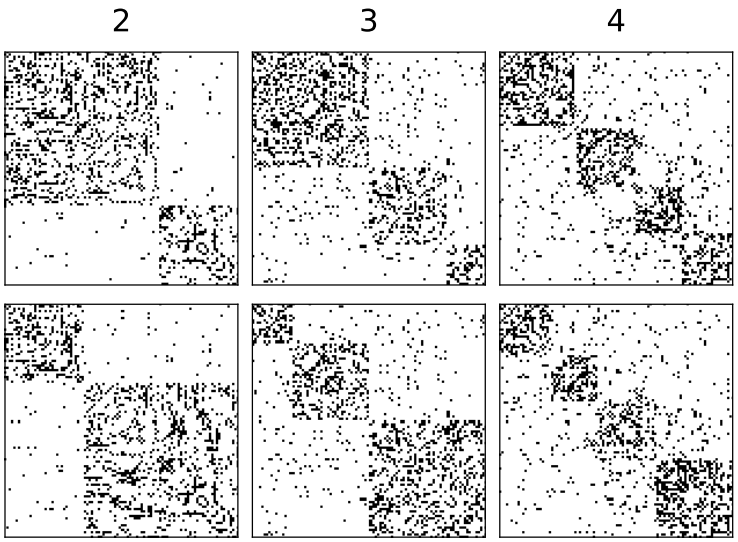


图 *4.* 通过预先向 G^3 模型的条件形式指定节点标签而生成的随机块模型邻接矩阵。列标题表示社区数量；顶行使用给定的成员规模比例，底行则按相反顺序使用这些规模。

E.3. 条件图生成

另一种矩阵表示的选择带来了一种额外的表述，为纳入协变量信息以实现条件生成（在生成文献中通常称为“引导”）提供了一种自然方式。协变量辅助的基于谱的聚类（Binkiewicz等人，2017）利用节点级协变量信息来辅助基于谱的聚类技术。该方法的关键在于按如下方式增广拉普拉斯矩阵： $L_\omega := L + \omega ZZ^\top$ 对于某个调节参数 ω 以及协变量 Z。我们将这一增广用于我们的矩阵表示（Y_0），并观察到，对于向量值的 Z（记为 $\boldsymbol{z}$），$\lim_{t\to\infty} H_t L_\omega = n^{-1}\omega \mathbf{1}(\mathbf{1}^\top \boldsymbol{z})\boldsymbol{z}^\top$；换言之，极限分布显式地依赖于节点协变量。由此产生一种自然方式，可通过构造类似于第E.2节中所述的合适极限矩阵，来实现以预先指定的协变量信息为条件的图生成。

在图 4中，我们通过随机块模型生成的一个简单示例，展示了该方法实际有效的证据。在图的每一列中，我们生成不同数量的社区，并采用预先选定的成员比例：这些信息被直接编码到向量 $(\boldsymbol{z}_i)_{i=1}^{N^*}$ 中，对应于 N^* 个新图，并用于构造形式为 $X_0 = n^{-1}\hat{}\omega(\tilde{\boldsymbol{x}}_0^\top \boldsymbol{z})\tilde{\boldsymbol{x}}_0\boldsymbol{z}^\top$ 的矩阵，其中$\tilde{\boldsymbol{x}}_0$ ∼服从狄利克雷分布$(\alpha\mathbf{1})$。

F. 补充实验结果

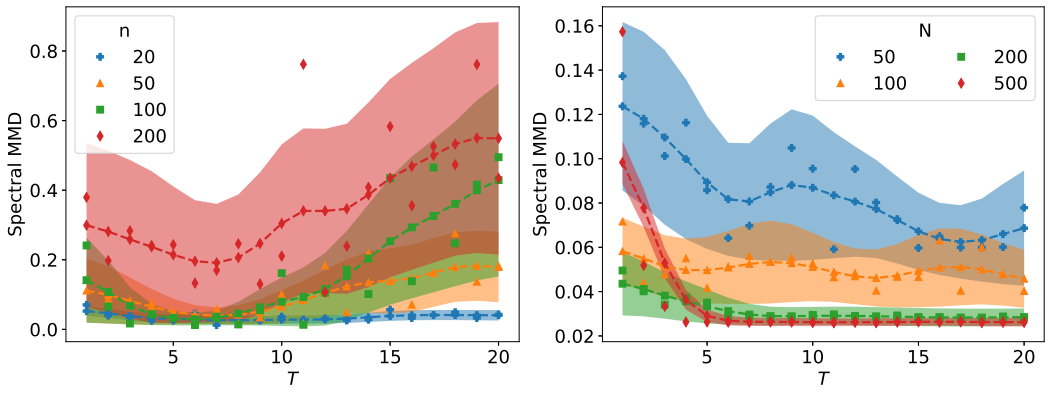
本节我们给出若干图，用以说明 G^3 相对于训练（及采样）变量的行为，包括图的数量、节点数量以及基分布的 α 参数选择。

图 5a 展示了谱 MMD 随 T 以及图中节点数量变化的情况。或许令人意外的是，曲线的表观极小值似乎都出现在大致相同的 T 处，这或许表明 N 是更重要的因素。曲线似乎还表明，至少对于此处展示的平面图而言，随着 n 增大，问题会变得更“难”。

图 7 补充了图 2 和图 5a 的内容，展示了非唯一输出图的比例如何随 n 和N（以及 T）变化。总体而言，谱性能与非唯一性之间存在一定的负相关，这表明需要谨慎优化图输出的整体质量。

表7。 ± 个最大均值差异指标的中位数（绝对偏差中位数），用于比较 G^3、DeFoG*、SPECTRE* 在 $3-5$ 个随机种子划分上的表现。数值越低越好，**粗体** 表示最佳表现，判定依据为中位数加一个绝对偏差中位数。另列出训练集与测试集比较，以展示最佳可能表现。

数据集	模型	聚类	度	轨道	谱	三角形
SBM	G^3	**0.0356 ± 0.0005**	**0.0318 ± 0.0052**	0.0631 ± 0.0120	0.0796 ± 0.0120	**0.0270 ± 0.0078**
	非对称 G^3	0.0398 ± 0.0010	0.0390 ± 0.0052	0.0710 ± 0.0140	**0.0832 ± 0.0015**	0.0287 ± 0.0084
	DeFoG*	0.0513 ± 0.0005	0.6270 ± 0.0027	**0.0500 ± 0.0000**	1.6600 ± 0.0003	0.0818 ± 0.0030
	训练	0.0312 ± 0.0000	0.0028 ± 0.0013	0.0312 ± 0.0000	0.0055 ± 0.0004	0.0017 ± 0.0002
度校正随机块模型	G^3	0.1270 ± 0.0091	**0.0344 ± 0.0062**	0.0971 ± 0.0057	0.1230 ± 0.0330	**0.0110 ± 0.0098**
	非对称。 G^3	0.2340 ± 0.0160	0.0810 ± 0.0200	0.1420 ± 0.0170	**0.1010 ± 0.0220**	0.0578 ± 0.0250
	DeFoG*	**0.1110 ± 0.0056**	0.7370 ± 0.0015	**0.0665 ± 0.0003**	1.2600 ± 0.0041	0.4330 ± 0.0210
	训练	0.0291 ± 0.0009	0.0017 ± 0.0007	0.0286 ± 0.0025	0.0277 ± 0.0069	0.0012 ± 0.0002
平面	G^3	0.3090 ± 0.0300	**0.0030 ± 0.0015**	**0.0035 ± 0.0018**	**0.0060 ± 0.0015**	0.0467 ± 0.0360
	非对称。 G^3	1.0200 ± 0.0440	0.0074 ± 0.0050	0.0106 ± 0.0089	0.0113 ± 0.0021	0.1110 ± 0.0320
	DeFoG*	**0.1200 ± 0.0043**	0.7350 ± 0.0010	1.0100 ± 0.0014	1.9500 ± 0.0160	**0.0296 ± 0.0005**
	训练	0.0224 ± 0.0021	0.0004 ± 0.0000	0.0002 ± 0.0001	0.0004 ± 0.0002	0.0050 ± 0.0016
酶	G^3	**0.0225 ± 0.0008**	**0.0854 ± 0.0550**	0.0902 ± 0.0390	**0.0755 ± 0.0220**	**0.0490 ± 0.0380**
	非对称 G^3	0.0803 ± 0.0200	0.0934 ± 0.0520	**0.1040 ± 0.0047**	0.1090 ± 0.0480	0.0752 ± 0.0120
	SPECTRE*	0.0598 ± 0.0110	0.7290 ± 0.0043	0.6140 ± 0.0230	1.1700 ± 0.0590	0.4700 ± 0.0096
	训练	0.0122 ± 0.0002	0.0004 ± 0.0004	0.0024 ± 0.0014	0.0114 ± 0.0024	0.0021 ± 0.0001
蛋白质	G^3	**0.0465 ± 0.0009**	**0.0220 ± 0.0012**	0.5530 ± 0.0720	0.1600 ± 0.0074	0.0928 ± 0.0180
	非对称。 G^3	0.0750 ± 0.0073	0.0345 ± 0.0003	**0.3480 ± 0.2000**	**0.0921 ± 0.0110**	**0.0419 ± 0.0071**
	SPECTRE*	0.1070 ± 0.0000	0.3600 ± 0.0650	0.7820 ± 0.0250	0.3750 ± 0.0044	0.0358 ± 0.0240
	训练	0.0504 ± 0.0001	0.0012 ± 0.0001	0.0093 ± 0.0010	0.0568 ± 0.0003	0.0011 ± 0.0003

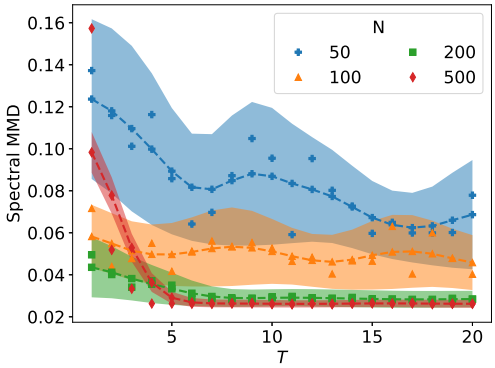


(a) *(b)*

图 *5.* 谱 MMD 随 T和节点数量的变化， n，对于（$N = 100$）（a）平面图和（b）随机块模型。曲线为观测值的高斯平滑[1] 函数，观测值本身取自 10 个随机种子的平均值。阴影区域表示在中心极限定理假设下的 66% 置信区间。训练时的学习率设为 1×10^{-4}。

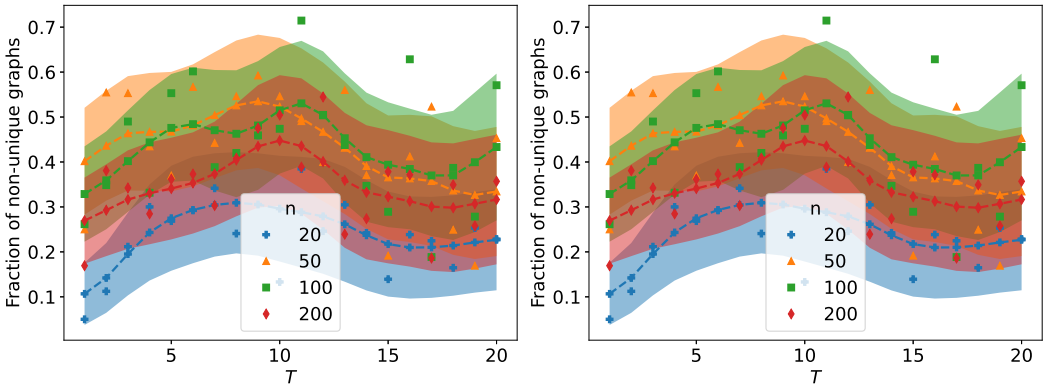
N	T	w	spec
100	15	1024	0.020 ± 0.019
		2048	0.027 ± 0.022
		4096	0.041 ± 0.043

(a)



(b)

图 6. (a) 对于不同的 w（每层隐藏节点数），在固定 T 和 N（针对随机块模型）时的谱 MMD。(b)：T 和图数量 N 对随机块模型的谱 MMD 的影响。曲线是对所示观测值进行高斯平滑[1] 得到的函数；这些观测值本身是对 10 个随机种子求平均的结果。阴影区域表示在中心极限定理假设下的 66% 置信区间。训练时的学习率设为 1×10^{-4}。



(a) *(b)*

图 7. 生成的非唯一图所占比例随 T 的变化。(a)：对于不同数据集大小 $n = 64$，平面图（N）。(b)：对于不同图规模的平面图（$N = 100$）。曲线是对所示观测值进行高斯平滑[1] 得到的函数；这些观测值本身是对 10 个随机种子求平均的结果。阴影区域表示在中心极限定理假设下的 66% 置信区间。训练时的学习率设为 1×10^{-4}。

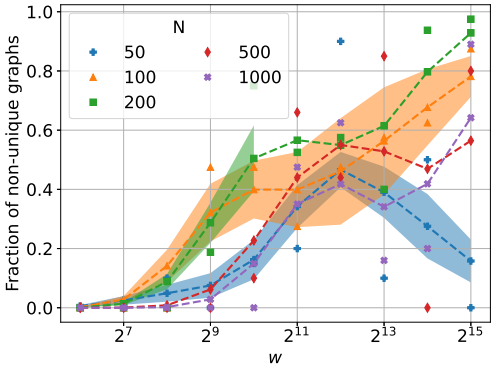
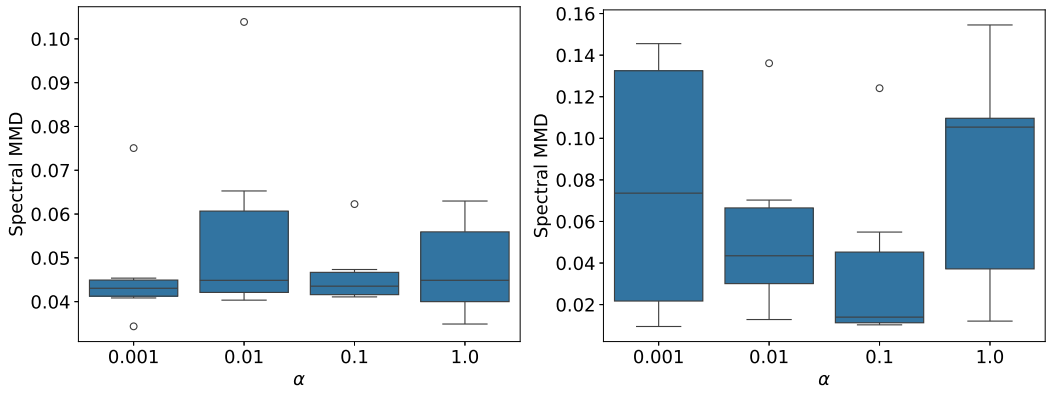


图 8. 生成的非唯一平面图所占比例随 w（每个隐藏层的节点数）和图的数量 N 的变化。曲线是对所示观测值进行高斯平滑[1] 后得到的函数；这些观测值本身是对 10 个随机种子的结果取平均所得。阴影区域表示基于中心极限定理假设的 66% 置信区间。训练时的学习率设为 1×10^{-4}。



(a) *(b)*

图*9.* 谱 MMD 作为 α（基分布的狄利克雷浓度参数）的函数。 (a)：200 节点随机块模型图的结果； (b)：64 节点平面图的结果。

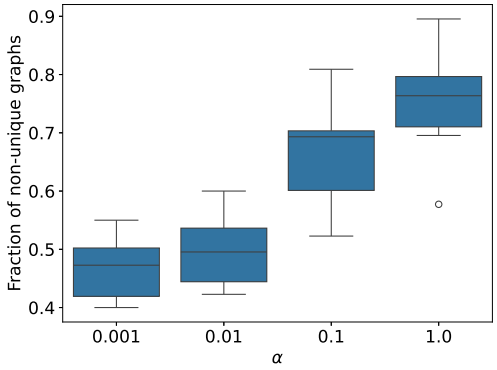


图*10.* 生成的（平面）图中非唯一图所占比例作为 α 的函数。

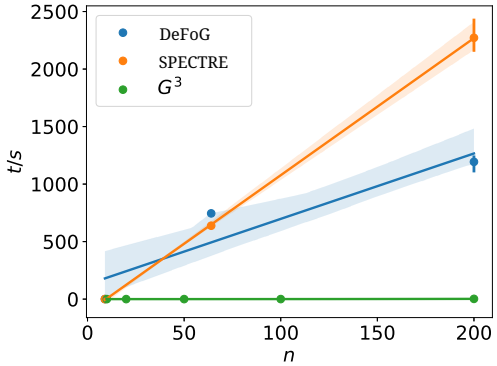


图*11.* DeFoG、SPECTRE 以及 G^3 的每个样本图的训练加生成时间。

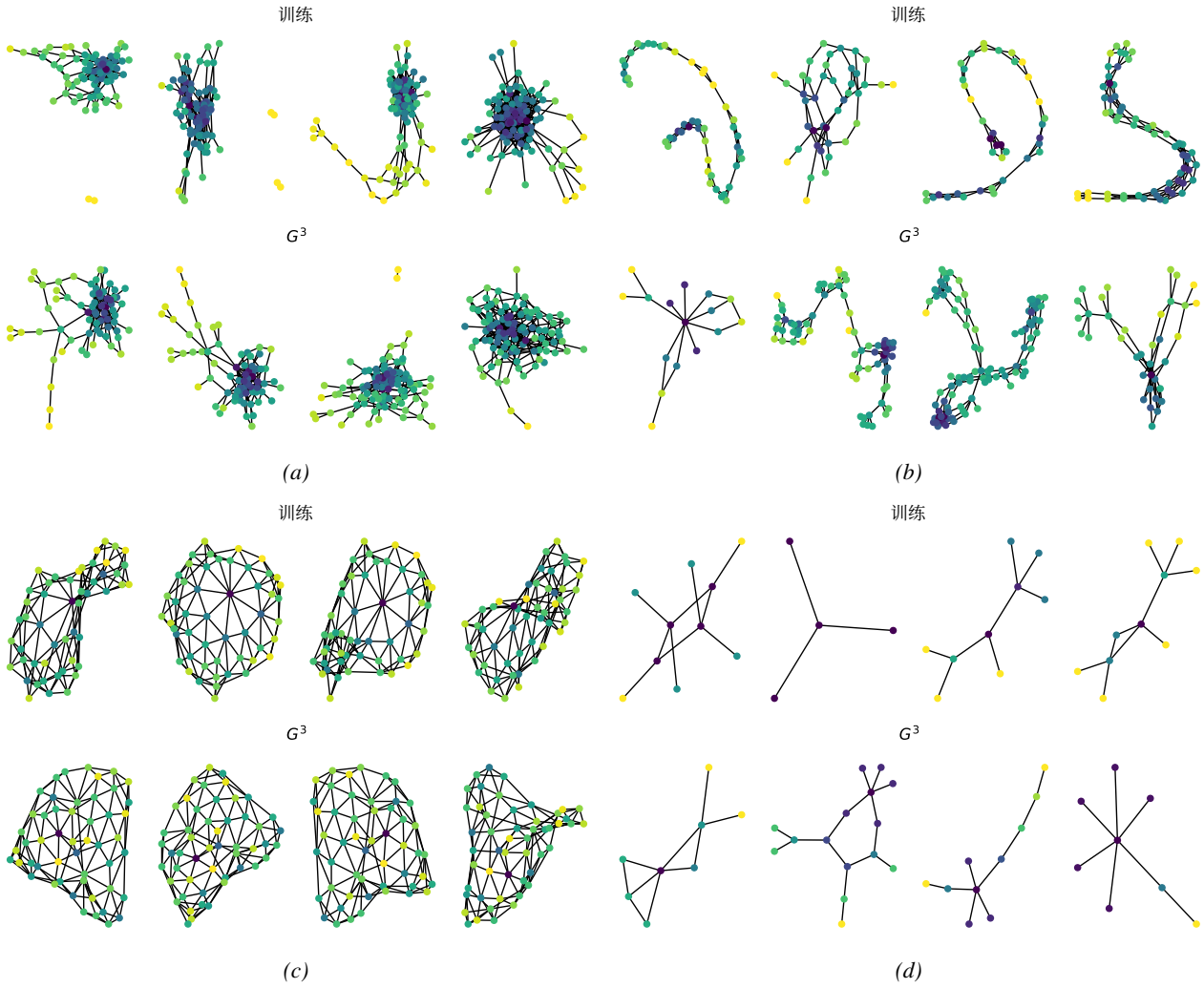


图*12.* 从训练集（上）中随机抽取的一系列示例图，以及 G^3 的输出（下）。(a)：度校正随机块模型，(b)：蛋白质，(c)：平面，(d)：QM9。节点着色反映该节点处的 Forman-Ricci 曲率。

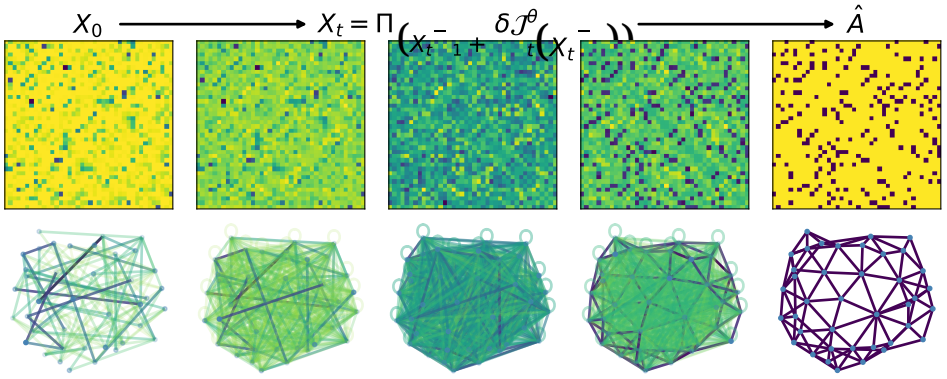


图 *13.* 从基分布中抽取的样本重构图的示例。