arXiv:2301.08364v1 [cs.SI] 20 Jan 2023

探索有序模式：通过邻接矩阵改进复杂网络上的机器学习

Mariane B.Neiva, OdemirM.Bruno*科学
计算组 圣卡洛斯物理研究所 圣保罗大学
marianeneiva@usp.br,
bruno@ifsc.usp.br

摘要

将复杂网络作为理解世界及其动力学的现代方法，已在文献中得到广泛认可。邻接矩阵能够一一表示复杂网络，也可用于计算图的多种指标。然而，这种表示并不总是唯一的，因为矩阵中行的置换可能表示同一个图。为解决这一问题，所提方法采用排序算法，将复杂图的邻接矩阵元素重新排列为特定顺序。随后，将排序后的邻接矩阵作为特征提取和机器学习算法的输入，以对网络进行分类。结果表明，所提方法在合成数据和现实世界数据上的表现均优于以往文献中的结果。

关键词 邻接矩阵 · 模式识别 · 复杂网络 · 深度学习

1 引言

大数据的兴起激发了人们的兴趣，希望尽可能贴近现实地构建数据结构，并对其进行评估以提取知识。传统的数据分析往往会将复杂现象简化为一个简单对象。然而，技术进步以及收集、处理和存储更多数据的能力，使我们能够从不同视角探索信息。复杂网络是能够基于特定方面连接各个元素的系统，填补了经典科学留下的空白。创建由元素及其关系构成的系统的能力，使研究从还原论方法转向整合方法。

此外，模式识别一直是数据科学的重要分支，借助技术视角帮助人们理解世界。如果要设想一种能够充分发挥人工智能技术和整合式数据分析优势的方法，复杂网络自然是一个选择。复杂网络中的模式识别包含多种算法，例如分类和聚类。尽管聚类在该领域发挥着重要作用，分类则能帮助我们识别疾病、物种、结构和城市等。如今，由于产生的数据量庞大，人工分析既耗时又昂贵，因此这项任务至关重要。此外，在大数据时代，复杂网络在模式识别方面具有显著优势，因为它们可用于对各种数据建模，从图像到生物系统，无所不包。多年来的研究表明，大多数真实系统都具有小世界网络和无标度网络的特征。前者指的是这样一种结构：其中的元素平均而言通过较短的最短路径相连；这类似于社交网络中的情形，即一个人的朋友很可能也是此人的朋友。后者指的是网络中经常被访问的元素，例如某一领域的知名研究人员或文本中的重要文章。后一示例说明了使用图分析特定组织中模式和结构的重要性。因此，本研究的定量分析侧重于合成网络和真实网络的分类。

基于上述优势，研究人员已成功将该模型用于各种应用中的模式识别，例如静态和动态纹理分类、 [1], 形状 [2], 作者身份 [3], 等。

近期，一些研究开展了相关工作，例如在 [4], 从网络构建多维深度嵌入中使用元胞自动机，以及在 [5], 分析形状图中形成的角度。与使用度和聚类系数等经典统计指标来创建一维图表示的传统分析相比，这些工作展现出了独特之处。一些研究人员近期致力于寻找新颖技术，以克服基于经典度量指标的描述符组合中存在的冗余信息，并在图分类中取得了良好结果。如 [8], 所示，连接一些相关度量指标有时并不利于模式识别。然而，我们是否已经穷尽了对复杂网络最直接的分析方法？本研究通过探究一种表示图的简单替代方法——邻接矩阵，旨在部分回答这一问题。

图的邻接矩阵与图本身一一对应。这种表示方式可以快速计算度和共被引等指标。然而，对邻接矩阵进行视觉检查只能提供很少的信息，因为对行或列进行置换并不会改变底层图。为解决这一问题，我们提出对矩阵的行进行排序，使矩阵中的模式保持一致，并能够区分合成网络和真实网络中的标签。此外，我们还评估了多种应用于排序后邻接矩阵的特征提取方法，以用于分类，包括数据投影、深度学习特征提取、CLBP分析、Hu矩和经典测量指标。我们在合成模型、代谢网络和社交网络上的结果表明，该方法能够以超过90%的准确率对网络进行分类，优于先前研究的准确率。

本文结构如下：第2节讨论邻接矩阵的构建、其对复杂网络分析的意义，以及所提出的排序方法。第3节介绍研究中评估的数据集，以及用于比较结果的文献方法。第4节中，我们进行两项评估：首先，对合成网络进行视觉分析，以考察所提出的排序方法是否突出了模型的重要特征（见第4.1节）。其次，我们采用各种签名方法作为网络的描述符进行分类，结果见第4.2节。最后，在第5节中，我们总结讨论内容，并强调本文的主要贡献。

2 所提出的方法

本节介绍一种图的替代表示形式——邻接矩阵，并描述若干复杂网络指标。

复杂网络是一个集合 G =(V,E)，其中 V 是系统的节点（或顶点），V =$\{v_1,\ v_2,\ v_i,\ v_j,\ v_N\}$；而 E 则是根据既定条件连接顶点的边，E =$\{e_{ij} = (v_i,\ v_j)|$ 当 v_i 与 v_j 相连时$\}$。如前所述，复杂网络源自图论；因此，这种结构的经典定义与其前身相同。不过，尽管一些研究人员认为复杂网络不只是图，因为某些拓扑特征能够将复杂性纳入模型，我们仍可以（并且将会）将这两种理论视为相同。因此，图的一种矩阵形式称为邻接矩阵。对于一个含有 N 个顶点的无向网络 G，矩阵 A 的计算方式如下：

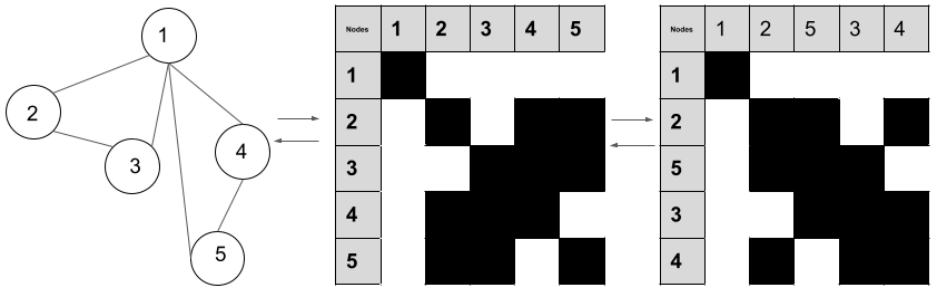
$$A(v_i, v_j) = \{\, 1\,,\ \text{if } v_i \text{ and } v_j \text{ are connected } 0,\ \text{otherwise} \tag{1}$$

这意味着每条边 e_{ij} 和所有节点 v_i 都表示在 A 中。此外，对于无向网络，邻接矩阵是对称的；如果没有自环，其对角线元素则全为零。矩阵 A 可用于计算一系列统计量，例如度、共被引（$C = AA^T$）和书目耦合（$B = A^T A$）。共被引和书目耦合尤其用于有向网络，以创建投影并将其转换为非对称形式。这些都是分析参考文献的重要指标，例如可用于检查网络中的关键节点。
然而可以注意到，在图表示中，节点通常没有顺序。因此，如图1所示，两个邻接矩阵可以表示同一个网络。鉴于矩阵缺乏空间顺序，本文提出一种矩阵排序技术，以寻找图中的模式。接下来一节旨在说明对 A 进行这种变换的方式及选择这种变换的原因。

2.1 复杂网络的邻接矩阵签名

图中可以分析的一项基本特征是顶点的度 i, k_i。该指标与顶点同图中其他节点的连接数量有关。此外，一种直接的分析方法是将度最高的节点视为系统中最重要的节点。因此，对矩阵 A 的第一种排序方式是按照节点的度从高到低排列所有节点。这意味着度较高的节点将位于 A 的第一行和第一列。

图1：一个图可以有多种正确的表示。第一幅图像以圆圈表示顶点，并用线连接；另外两幅则采用矩阵形式。可以看出，尽管两个矩阵对应相同的信息，其视觉模式却有所不同。利用网络中边的分布并不可靠。不过，排序算法可以让机器学习方法对这些模式进行分类。



然而，多个节点可能具有相同的度，这样一来，创建唯一网络表示的目标便无法实现。因此，我们提出以下规则：

1. 按节点度降序排列各行。(a) 如果第 i 行和第 j 行的两个节点度相同，则按中介中心性从高到低排序。

这一简单的排序将确保连接较多的节点占据新矩阵 A’ 的前几个位置。此外，由于第二条规则，对于给定的输入，单个网络对应的矩阵 A’ 将是唯一的。由于我们分析的是依赖于矩阵排列的统计量，因此需要确认，无论 A 最初如何初始化，矩阵图像始终相同。

A'（邻接矩阵 A 的有序版本）将在本研究中通过两种方式进行评估：视觉评估和定量评估。最后，将根据第4.2节中介绍的几种描述符方法评估最终签名。与实验结果相比，该模型的简洁性表明，对于邻接矩阵等系统的基本表示仍有探索空间。

3 实验

复杂网络数据集

在本研究中，我们将在四个数据集上评估基于邻接矩阵的签名。因此，本节介绍用于分类任务的¯两个合成网络和两个真实网络。在下文中，k 表示平均度，N 表示节点数。

• **合成数据集**：在复杂网络文献中，一些模型广为人知，有助于我们概括许多真实现象。因此，该数据集由四类不同的图组成：

¯– 通过按照概率 p = k/n 添加边生成的随机网络；– 先创建为正则图、再以概率 p = 0进行边重连的小世界网络；– 采用线性和非线性优先连接机制的无标度模型；– 地理网络。

¯每种模型均包含平均度不同的网络：k = {4, 6, 8, 10, 12, 14, 16}，以及规模 N 不同的网络：= {500, 1000, 1500, 2000}。对于每种模型组合、平均度和规模，均创建 100 个网络，共计 11200 个文件。

• **无标度网络**：在无标度模型中，若干研究人员提出了构建具有幂律分布的网络的方法。该数据集包含 Barabási 与 Albert 提出的图[9]，其中生成了线性和非线性网络（α = {0.5, 1.0, 1.5, 2.0}），以及 Dorogovtsev 与 Mendes 提出的图 [10]，共计 ¯500 个网络（分为五类），规模为 1000，且 k = 8。

• **代谢数据集**：代谢系统可以表示为图，这正是该数据集的情况。该代谢数据集以 43 个网络描述了三种不同物种的系统：6 种古菌、32 种细菌，以及

5 种真核生物 [11]。分类之前，使用 Weka 平台 [12] 中的重采样方法来评估分层数据。

• **社交网络**：为了区分 Google+和 Twitter 的社交网络结构，从 SNAP（斯坦福网络分析项目）平台 [13] 中提取了每个类别的 65 个图。每个网络都是通过获取若干用户的社交关系构建的。目标是检验一种简单的方法能否区分这两种结构。

与其他方法的比较

为将所提出的方法与文献中的结果进行比较，考虑了两种方法：LLNA（生命-生命网络元胞自动机） [4] 以及经典结构度量分析的性能。以下简要介绍这些方法：

生命-生命网络元胞自动机

这篇最新发表的论文利用了复杂网络和基于元胞自动机的统计量 [4]的优势。首先，该方法将原始网络转换为元胞自动机。然后，将每个细胞设置为存活或死亡，并应用动态演化。通过计算自动机演化过程中的熵和 Lempel-Ziv复杂度等统计量，生成特征向量。该方法还评估了合成系统和真实系统。

经典结构度量

复杂网络的经典分析包括度、聚类和接近中心性等结构度量 [14, 15]。为了将我们的方法与基础方法进行比较，我们计算了一组测量指标，并评估它们单独及组合时区分各类别的能力。所用的度量如下。此外，由于某些度量取决于网络中的节点数，因此我们计算了直方图，并将其归一化为大小为 500。

• **度同配性** (pp)：第一项度量根据图中邻居的度来分析它们之间的相似性。输出为一个大小为 N（节点数）的特征向量。
• **直径** (d)：这一针对整个网络的单一度量，表示图中节点对之间所有最短路径的最大距离。该方法首先计算所有最短路径，然后输出此度量，因此计算成本较高。

• **接近中心性** (cl)：第四项度量是大小为 N（节点数）的向量 cl，其中每个位置 i表示 i 与图中所有其他节点之间的最短距离之和的倒数。
• **离心率** (ecc)：与接近中心性不同，节点的离心率是该节点与网络中任意其他节点之间的最大测地距离。

• **中介中心性** (bet)：另一种基于最短路径的度量，介数中心性统计某个特定节点在图的所有最短路径中出现的次数。
• **度**：度是从网络中提取的最简单却很有力的指标。对于单个节点，它表示该节点包含的连接数量。在实验中，进行了两种类型的度分析： k是一个向量，其中对每个 k_i，返回节点 i的度。

• **聚类**：节点的聚类系数用于衡量一个顶点的邻居与完全图（所有邻居彼此相连）的接近程度。评估时使用一个向量（cc_i）来表示每个节点的聚类系数。

4 结果与讨论

4.1 合成网络有序邻接矩阵的可视化分析

首先，我们计算给定数据集中所有网络的邻接矩阵，并对其进行排序，使枢纽节点出现在图的起始位置。该排序可确保最终的 A’ 不受原始 A 中顶点排列方式的影响。本研究的主要分析方法是对 A’ 进行视觉检查。通过分析白色像素（A’ 中的连接）的分布，可以检查合成模型的特征是否在视觉检查中得到增强。图2展示了该方法论在合成数据集上的效果（所有图像均进行了少量膨胀处理，以改善可视化效果）。

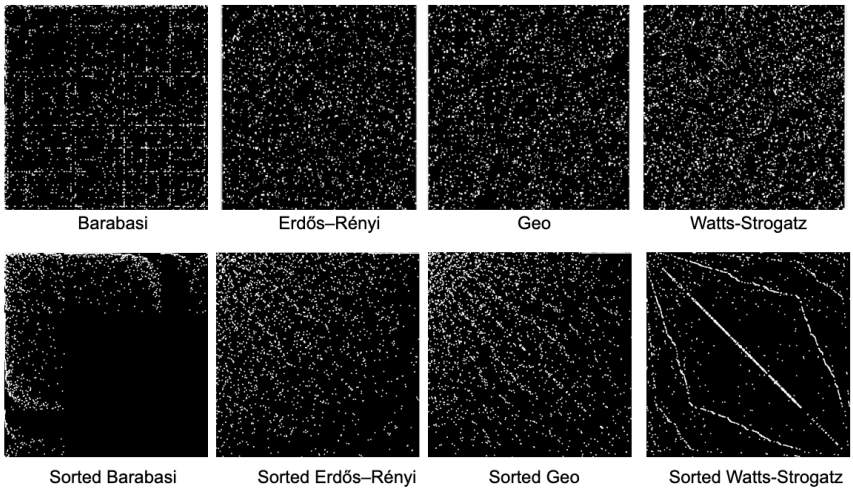
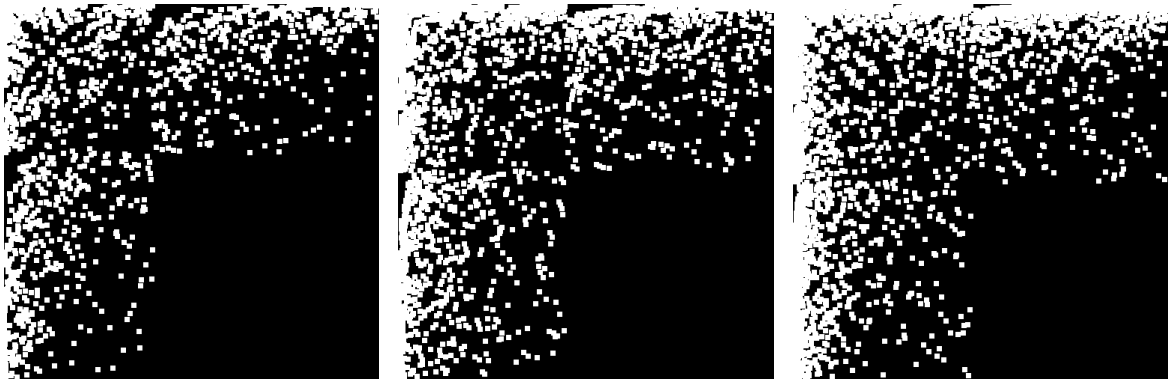


图2：第一行显示未排序的邻接矩阵，第二行显示所提方法的结果。与未排序矩阵相比，该方法增强了每个模型的特征。白色点表示边，图像像素经过少量膨胀处理，以改善可视化效果。



(a) Barabasi，n= 500, $\bar{k}$ =4　(b) Barabasi，n= 500, $\bar{k}$=6　(c) Barabasi，n= 500, $\bar{k}$ =8

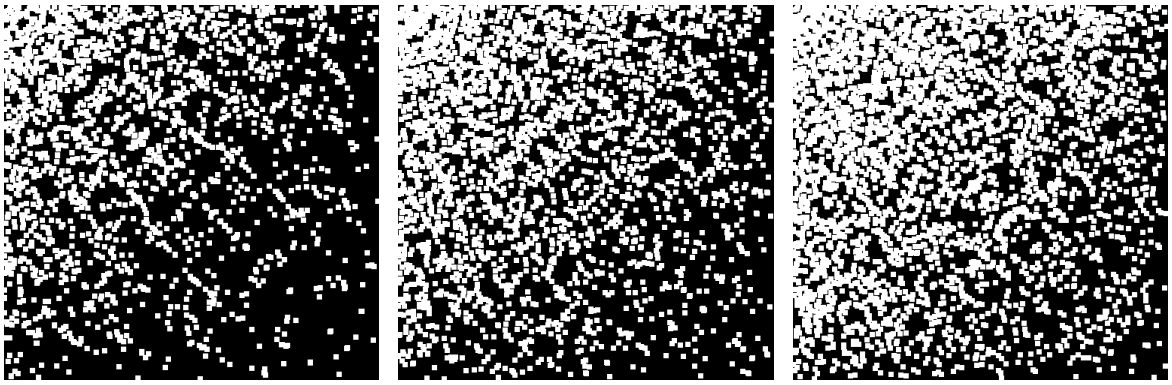
图3：应用所提方法后，Barabasi模型的模式变得可靠且更加突出

图3、4、5和6展示了每种合成类别中的一些样本，这些样本按照第2节所述进行了排序。通过视觉检查，我们可以在排序结果中看到这四种方法的一些理论特征。

首先，图3展示了 Barabási–Albert网络的示例，这是一种生成无标度图的方法 [9]。因此，要理解输出图像，我们必须先了解该模型。Barabási–Albert提出的无标度网络会生成这样的图：初始时包含一个由 N_0 个节点构成的任意网络。随后，不断添加节点，并将其连接到c（一个方法参数）以及其他顶点。新节点连接到哪些顶点，是根据概率 $p = \frac{k_i}{\sum_j k_j}$ 来决定的，其中 k_i 是顶点 i的度。这意味着，度较高的顶点与这个新节点相连的概率更大。

因此，可以说，新顶点与包含许多连接的顶点之间存在优先连接。因此，无标度网络的度分布将遵循幂律，形式为 P(k) k^{-3}。此外，万维网是这种网络的一个重要例子，其中 Google 等热门网站更有可能与新页面建立连接。这些连接密集的顶点称为 $hubs$，对于理解系统以及提取有关其动力学的统计量和信息非常重要。

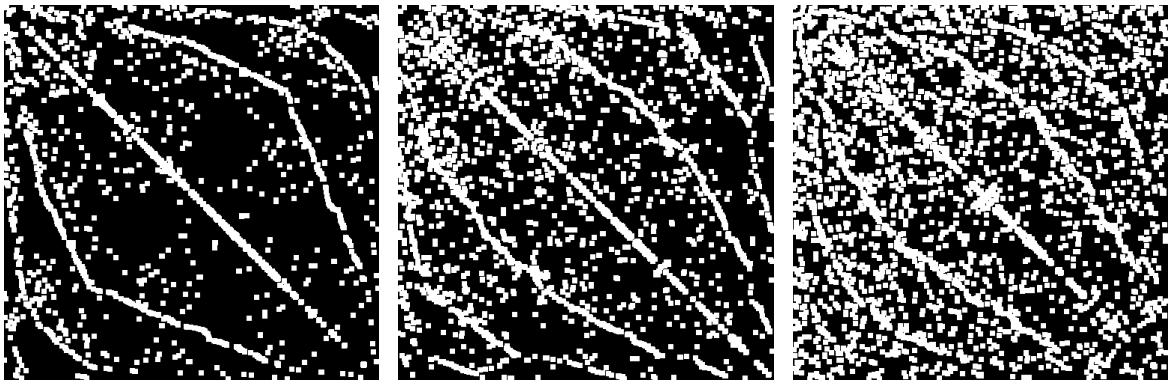
话虽如此，我们可以回到图3中的视觉检查。首先要注意的是，图像左上角的点非常密集。每个白点代表一条边。此外，排序确保连接较多的节点位于最前面的行和列中。右下角的点则较为稀疏，这表明图中度较低的节点占多数，这与该模型的幂律分布相符。



(a) Erdős，n= 500, $\bar{k}$ =4　(b) Erdős，n= 500, $\bar{k}$ =6　(c) Erdős，n= 500, $\bar{k}$ =8

图 4：Erdős模型图像显示，点在所有图像中广泛分布，凸显了网络的随机性特征。

我们的第二个模型是 Erdős–Rényi网络 [16]。创建随机网络最著名的方法是采用以下算法： N 节点按照概率 p相互连接。由于 p 是该方法的一个参数， p 较低时，图的连接较弱；而 p 较高时，生成的图则包含许多边。两个节点之间的连接以独立于节点本身的概率建立，这与前一种方法不同。这些特征在图4中清晰可见；在这种情况下，并没有像 Barabási–Albert模型中那样，点只聚集在左上方。点分布在整幅图像中，与其他合成模型相比，呈现出均匀分布的趋势。正如预期，左上方的点更多，这是排序造成的；但与无标度网络相比，随机网络中这种差异较小。

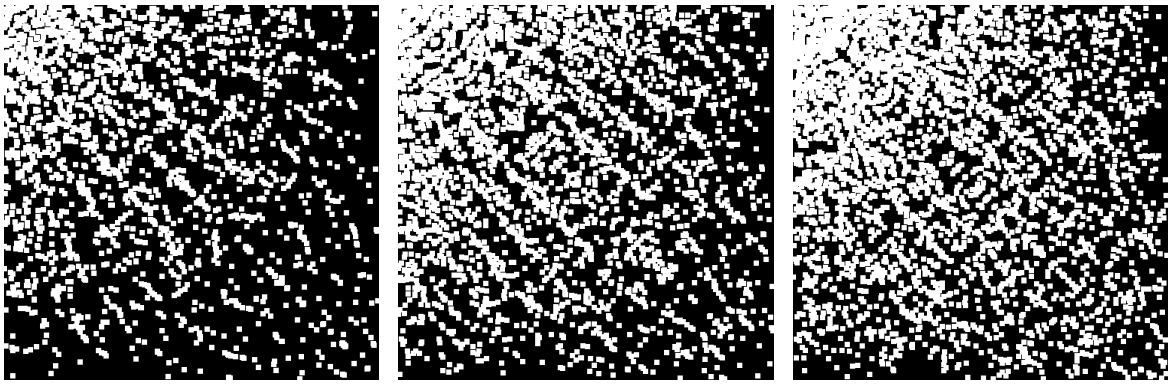


(a) Watts，n= 500, $\bar{k}$ =4　(b) Watts，n= 500, $\bar{k}$ =6　(c) Watts，n= 500, $\bar{k}$ =8

图 5：值得注意的是，枢纽节点的模式在 Watts 模型图像中的呈现方式。即使节点数量增加，基本模式仍保持不变。

自然系统中最常见的模型是小世界网络。**(author?)** 于**(year?)**提出了一种用于模拟这一类别的合成模型，该模型将图初始化为包含 N 个节点、平均度为 k 的规则网络：首先，每个节点与两侧各 k/2 个邻居相连。然后，以 $0 \leq \beta \leq 1$ 的概率重连每个节点右侧的边（β 是该方法的一个参数）。最终，算法生成彼此连接良好的节点，具有较高的聚类系数和较小的度量图直径。实际上，这意味着系统中任意两个顶点之间只需少量边即可连接。最后，让我们分析图 5 中展示的三个模型样本的排序邻接矩阵。这些图像呈现出一种有趣的模式；可以看到一种菱形形状，这可由应用于每个节点的 k/2 个邻居的重连方法以及所用的概率（$\beta = 0.1$）来解释。该模式的形状也有助于我们理解小世界特性。如果一个节点无法通过一步直接到达另一个节点，那么只需经过一条较短的路径即可相连。模式中的线性线条，以及网络初始设置为规则环这一事实，都体现了这一特性。

最后一个合成模型没有明确的统计特征，因此我们无法对图 6 中呈现的模式做出太多假设。地理网络是根据系统的空间属性构建的。



(a) Geo 模型，n= 500，$\bar{k}$ =4　(b) Geo 模型，n= 500，$\bar{k}$ =6　(c) Geo 模型，n= 500，$\bar{k}$ =8

图 6：Geo 模型是唯一没有固定生成流程的模型。从图像中可以看出，尽管缺乏严格规则，该模型仍保留了其中的模式。

该模型的重要现实案例包括机场和道路交通系统。尽管缺乏严格的规则，模型中的模式仍然存在。此外，我们可以注意到一种与图 4 中随机网络相似的模式。这种相似性可能表明难以区分这两类网络，而统计方法可以对此进行评估。

4.2 基于邻接矩阵特征的定量分析

视觉检查表明，第2.1节提出的方法适用于网络表征。不过，本节还使用所述签名进行了数值分析。在此背景下，我们不仅评估合成网络，也评估真实网络和无标度模型。我们将这一简单方法得到的结果与 LLNA [4] 以及经典结构度量所得到的结果进行比较。表1至表3展示了使用 K近邻（k = 1）和支持向量机得到的 CCR（正确分类率）。此外，还采用 k 折交叉验证（k = 10）分析泛化能力。值得注意的是，我们评估了从各复杂网络的邻接矩阵排序所生成的图像中提取的多种特征。本节分析并讨论了局部和全局方法，探索范围从最简单的图像分析逐步扩展到更复杂的层次。总之，我们进行了不同的实验：

1. **投影**：这种将二维图像转换为一维表示的简单方法，是对x 轴上各列的值求和。因此，由于较高的值位于前几列，所得特征向量呈下降趋势。无需进行定量分析，也可以认为，由于向量斜率系数不同，这种方法可能有助于区分随机图和无标度图。由于网络的规模各不相同，投影向量的长度设为 2500；当网络规模小于该值时，其余位置的值设为零。我们认为，补零不会影响结果，因为这表示度为零。

2. **CLBP**：完整局部二值 [18] 模式方法是经典局部二值模式（LBP [19]）的扩展。与其前身仅分析邻居之间的二值差异不同，CLBP 还利用差异幅值和信号信息来计算特征。事实证明，加入这些信息适用于许多应用，例如纹理识别 [18]。在实验中，用于评估邻居的中心像素周围窗口为 3x3。

3. **Hu矩**：Hu 图像矩被理解为特定区域内像素强度的加权平均值，或由前序矩构成的函数。这些方法已用于描述图像的区域、面积和质心。然而，已知某些统计量具有尺度、平移和旋转不变性。Hu 矩由（**作者?**）（（**年份?**））提出，是一组基于其他矩函数构建的七个特征，包含诸如图像质心周围的惯性等信息，并能够区分镜像图像。这组七个特征用作 K近邻和支持向量机分析的输入。

4. **VGG-19**：深度神经网络如今因具备图像分类能力而备受重视，是用于分类的强大特征来源。因此，本实验采用了使用来自 Imagenet 数据集的预训练权重的 VGG-19 模型 [21]。模型最后一个池化层输出的权重

表 1：使用 K近邻（k = 1）对排序后的矩阵进行分类。

	KNN, k = 1			
	投影	CLBP	Hu矩	VGG-19
合成	96.44 (0.56)	96.36(0.56)	99.68(0.20)	99.29(0.28)
合成无标度	99.80 (0.66)	99.36 (0.94)	100.00 (0.00)	99.80(0.63)
社交网络	84.00 (0.47)	92.31 (8.88)	47.69(15.72)	90.00(10.91)
代谢	92.03 (6.09)	91.95 (11.93)	97.95(6.20)	93.50(10.55)

表2：所提方法结合支持向量机对有序矩阵进行分类。

	SVM			
	投影	CLBP	Hu矩	VGG-19
合成	100.00 (0.00)	82.00(0.76)	53.10(1.15)	99.86(0.12)
合成无标度	100.00 (0.00)	98.74 (1.60)	88.08 (7.85)	100.00(0.00)
社交网络	87.69(6.49)	92.31 (8.88)	54.62(11.72)	91.47(9.21)
代谢	91.00(11.74)	93.00(11.46)	80.75(13.84)	91.50(11.07)

被用作实验中 K近邻和支持向量机的输入。 [22] 中的纹理分析方法论前景良好。

4.2.1 结果与讨论

在讨论表1至表3所呈现的结果之前，我们必须重申本研究提案的目标。尽管特征提取方法并非本研究的创新之处，但分析对象本身才是。第4.2节中介绍的所有方法，均根据其分析输入数据时的复杂程度进行排序。首先，投影是最基础的方法之一；由于所提出的排序依据节点的度值对节点进行排列，因此它能够捕捉网络的度结构。其次，文献中广泛用于表示节点的经典度量指标，主要分析图的结构以及各元素之间的连接方式。值得注意的是，实验中使用的大多数度量指标都依赖于对最短路径的研究，这可能会在综合分析中引入冗余 [8]。接下来，CLBP是一种纹理提取方法，它对图像进行局部分析；随着探索范围从图像左上角逐渐向外扩展，它能够提取重要特征，因为枢纽节点通常集中在该区域。Hu矩是一种对尺度、旋转和平移具有不变性的方法，用于分析图像区域的统计量，例如惯性和镜像特性。它能够取得重要结果，这得益于其量化所提出邻接矩阵中像素全局分散程度的能力。最后，VGG-19是唯一一种无法明确了解其从图像模式中提取了哪些特征的方法。不过，深度神经网络能够构建层次化特征，并为输入图像创建强大的表示，从而获得较高的分类率，结果中也体现了这一点。

在实验中，这两个合成数据集是基于理论约束构建的，而基本统计量能够很好地捕捉这些约束。结构统计量和投影在合成数据集及无标度数据集上取得了较高的分类率，体现了这一知识。结合结构度量后，分类正确率可以达到100%（见表2至表3）。由于度分布，A’在 x 轴上的投影也能够很好地区分这些模型。对于无标度合成网络，使用聚类系数进行分类取得了最高的可能结果，表明该数据集很容易被区分。最后，对于合成网络，七个Hu矩结合K近邻也能稳健地表示这些模式，并得到100%的正确分类率（CCR）。此外，深度神经网络在描述重要特征方面也表现出强大的能力

表3：对　比所提方法与文献中的结果　　文献

	SVM									
	VGG-19	LLNA	pp	d	cl	ecc	bet	$\vec{k}$	cc_i	组合
合成	99.86(0.12)	99.92	77.51(0.47)	50.67(1.21)	86.95(1.06)	73.18(1.43)	84.83(1.19)	81.97(0.43)	92.37(1.07)	100.00(0.00)
无标度	100.00(0.00)	98.3 (0.3)	93.25(5.28)	83.00(3.50)	100.00(0.00)	89.00(5.03)	100.00(0.00)	98.50(2.11	100.00(0.00)	100.00(0.00)
社交网络	91.47(9.21)	92.00 (1.00)	46.92(7.65)	44.62(4.87)	93.85(4.87)	93.85(4.87)	86.15(4.87)	89.23(6.49)	93.85(4.87)	92.31(5.13)
代谢	91.50(11.07)	87.00 (13)	75.00(18.26)	75.00(18.26)	81.00(17.76)	83.50(18.57)	91.00(15.06)	91.50(14.54)	84.50(17.23)	93.50(10.55)

模型的表现（支持向量机对特征进行分类时，准确率为99%）。因此，鉴于简单的度量指标能够轻易识别这些集合的结构，无需再投入太多精力。

图7：合成数据集的混淆矩阵和曲线下面积（AUC）。结果通过提取VGG-19神经网络的特征，并使用支持向量机进行分类获得。

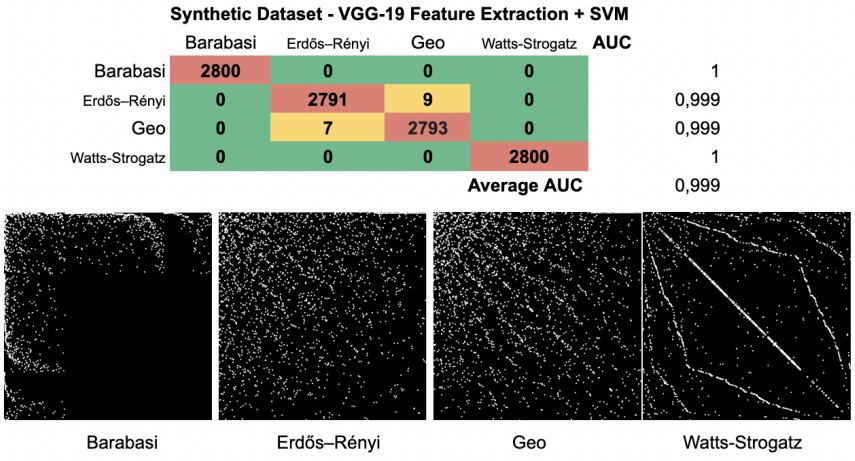
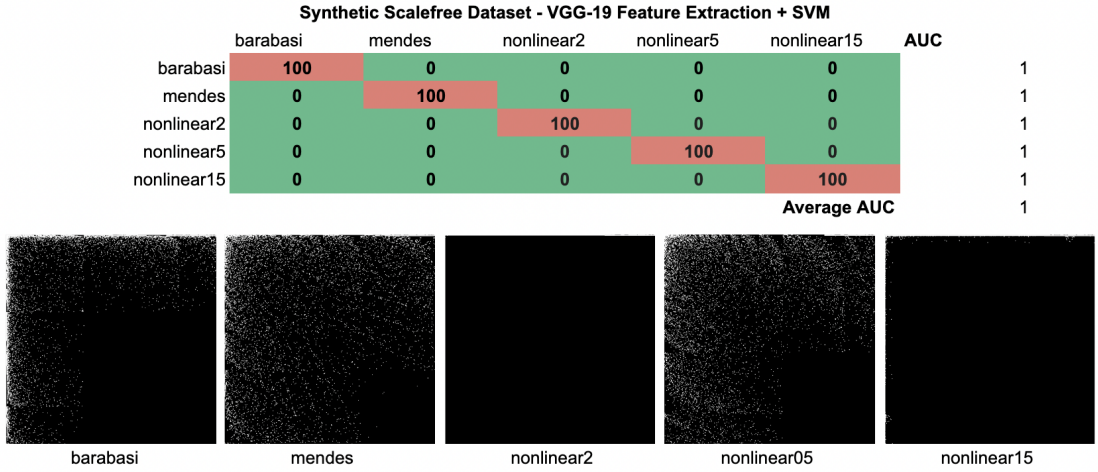


图8：VGG-19和支持向量机分类的无标度数据集混淆矩阵及曲线下面积。



关于图7和图8中所示方案的VGG-19特征提取与支持向量机分类的混淆矩阵，结果引出了对表3中合成数据集和无标度数据集准确率分别达到99.86%和100%的进一步探讨。正如视觉检查结果所预期的那样，Geo和Erdős–Rényi容易被混淆，因为它们的结构特征不够鲜明。Erdős–Rényi和地理网络共发生了16次混淆：7个地理网络被预测为Erdős–Rényi，另有9个被分类为随机图。另一方面，Barabasi和Watts–Strogatz的分类准确率均达到最高。对于无标度数据集，所有突出显示的模式都有助于分类。

然而，尽管评估结果表明这些方法适用于理论分析，探索真实世界仍然十分重要。本研究考察了两种性质不同的数据集：代谢网络和社交网络。可以发现，即使是简单的投影也能获得较高的分类率，这凸显了所提出的邻接矩阵利用流程的重要性。Hu矩的七个特征能够区分不同的代谢网络，正确率达到 97.95%（使用表 1 中的 K近邻）。此外，完整局部二值模式对两种网络的识别正确分类率均超过 90%，甚至基本的结构度量（无论是否组合使用）也能达到相近的识别率。

然而，简单的手工设计的描述符（如投影和局部模式提取器）的局限在于标准差非常高（实验中超过 6%）。这一问题不仅出现在所提出的方法中，也出现在结构

方法和 LLNA 方法中。此外，LLNA 的一个缺点是参数设置。该方法通过将原始网络转换为元胞自动机来实现。然后，该方法依据生命游戏规则演化镶嵌。然而，要找到能够实现良好分类效果的最佳规则，需要投入大量工作，并且计算成本很高。这意味着，尽管这种方法有望取得良好结果，但计算成本和参数设置方面的缺点使简单的邻接矩阵排序方法更具优势。

图9：VGG-19 和支持向量机对社交数据集进行分类的混淆矩阵和曲线下面积。

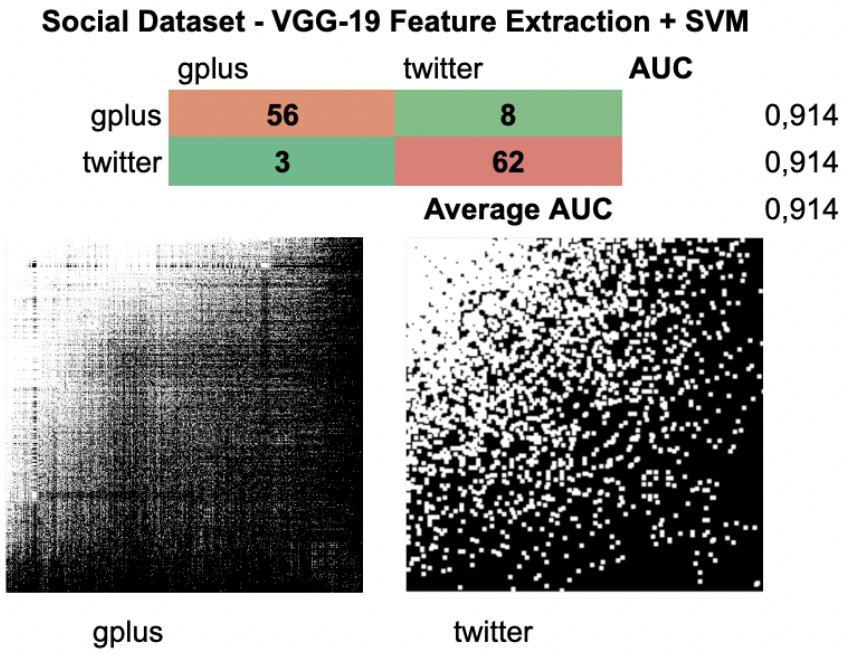
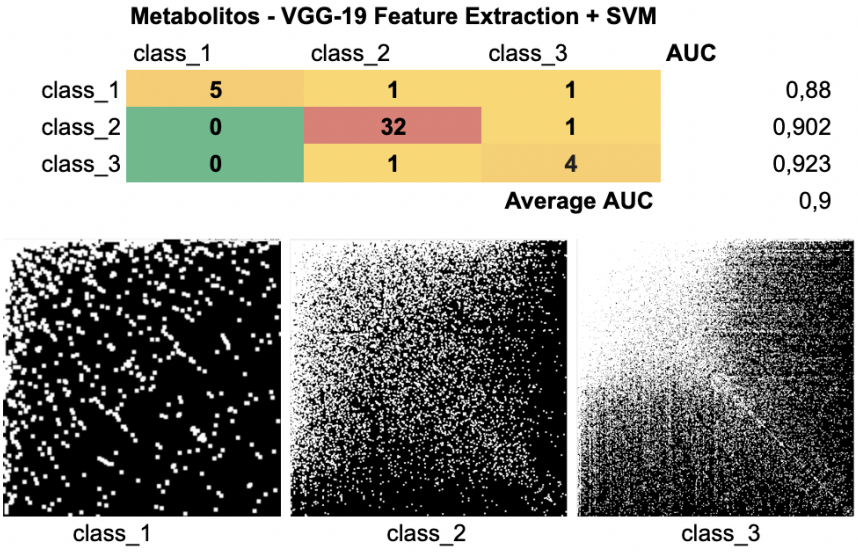


图10：VGG-19 和支持向量机对代谢数据集进行分类的混淆矩阵和曲线下面积。



VGG-19 与支持向量机结合后，能够对代谢网络进行分类，正确分类率为 91.50%；对社交图的正确分类率为 91.47%。此外，混淆矩阵显示了分类中的具体错误。Gplus 和 Twitter 的网络共被混淆 11 次：8 个 Gplus 网络被预测为 Twitter，另有 3 个被错误地归为相反类别。

对于代谢数据集，类别_2 和类别_3 各有一个错误输出。类别_1中有两张图像被错误标记。所有评估集的平均曲线下面积均超过0.9，体现了该方法论的重要性。

5 结论

本文研究了将图的一个简单特征作为分析主要输入的探索方法。邻接矩阵能够对网络进行一一表示，但对行进行简单置换会保留系统的所有属性，同时打乱其有序性。因此，本文提出了一种基于节点度的矩阵排序方法。通过这种方法，网络可以生成单一表示，并可将其作为图像进行分析。视觉分析显示，所提出的数据转换方法突出了模型中的模式。此外，本文还在四个不同的数据集上进行了定量分析，其中包括两个真实系统；结果表明，总体而言，所有提出的方法和对比方法都取得了良好的分类准确率。

然而，仍有三个方面需要分析：正确分类率（CCR）、标准差和方法的复杂度。显然，对于合成数据，基本指标足以表示网络。然而，对于真实数据集，由于数据的可变性，通常会得到较大的标准差。LLNA 的缺点是拟合参数所需的时间较长。因此，深度神经网络在图像分类方面的能力再次显现出优势（在本实验中，邻接矩阵可以理解为一幅图像），并展现出良好的结果，尤其是在代谢网络和社交网络这两个后续数据集上。总之，所提出的对基本网络表示进行排序和分析的方法适合在实际领域开展进一步研究。该方法表明，邻接矩阵这类简单表示是网络分类的良好信息来源。

致谢

O. M. Bruno 感谢 CNPq（资助编号 #307897/2018-4）和 FAPESP（资助编号 #2018/22214-6 和 #2021/08325-2）提供的支持。M. B. Neiva 感谢 CNPq（PROEX-9056169/D 资助编号）提供的支持。作者还感谢 NVIDIA GPU Grant Program。

参考文献

[1] Wesley Nunes Gonçalves、Bruno Brandoli Machado 和 Odemir Martinez Bruno。用于动态纹理识别的复杂网络方法。*Neurocomputing*，153:211–220，2015。[2] Lucas Correia Ribas、Mariane Barros Neiva 和 Odemir Martinez Bruno。用于形状分析的距离变换网络。*InformationSciences*，2018。[3] Jeaneth Machicao、Edilson A Corrêa Jr、Gisele HB Miranda、Diego R Amancio 和 Odemir M Bruno。基于类生命网络自动机的作者身份归属。*PloS one*，13(3):e0193703，2018。[4] Gisele Helena Barboni Miranda、Jeaneth Machicao 和 Odemir Martinez Bruno。探索元胞自动机的时空动力学，以实现网络中的模式识别。*Scientific Reports*，6:37329，2016。[5] Leonardo Scabini、Lucas Ribas、Eraldo Ribeiro 和 Odemir Bruno。利用卷积神经网络进行深度拓扑嵌入，实现复杂网络分类。发表于 *International Conference on Network Science*，第54–66页。Springer，2022。[6] Leonardo FS Scabini、Danilo O Fistarol、Sávio V Cantero、Wesley N Gonçalves、Bruno Brandoli Machado 和 Jose F Rodrigues Jr。复杂网络的角描述符：一种用于边界形状分析的新方法。*Expert Systems with Applications*，89:362–373，2017。[7] Lucas C Ribas、Leonardo Scabini 和 Odemir M Bruno。基于耳石形状的鱼类物种识别：一种复杂网络方法。发表于*2022 Eleventh InternationalConferenceon ImageProcessing Theory, ToolsandApplications (IPTA)*，第1–5页。IEEE，2022。[8] Luciano da Fontoura Costa。复杂网络，简单视觉。*arXivpreprint cond-mat/0403346*，2004。[9] Albert-László Barabási 和 Réka Albert。随机网络中尺度律的涌现。*science*，286(5439):509–512，1999。[10] Sergei N Dorogovtsev 和 José FF Mendes。网络的演化：从生物网络到互联网和万维网。OUP Oxford，2013。[11] Hawoong Jeong、Bálint Tombor、Réka Albert、Zoltan N Oltvai 和 A-L Barabási。代谢网络的大规模组织。*Nature*，407(6804):651，2000。[12] Mark Hall、Eibe Frank、Geoffrey Holmes、Bernhard Pfahringer、Peter Reutemann 和 Ian H Witten。Weka数据挖掘软件：更新。*ACM SIGKDD explorations newsletter*，11(1):10–18，2009。[13] Jure Leskovec、A Krevl 和 SNAP Datasets。SNAP数据集：斯坦福大型网络数据集合集。*URL:http://snap. stanford. edu/data/index. html*，2014。

[14] Mark EJ Newman. 网络中的混合模式。*Physical ReviewE*，67(2):026126，2003。[15] Mark Newman. 网络。牛津大学出版社，2018。[16] Paul Erdos 和 Alfréd Rényi. 随机图的演化。*Publ. Math. Inst. Hung. Acad. Sci*，5(1):17–60，1960。[17] Duncan J Watts 和 Steven H Strogatz. “小世界”网络的集体动力学。*Nature*，393(6684):440，1998。[18] Zhenhua Guo、Lei Zhang 和 David Zhang. 用于纹理分类的局部二值模式算子的完整建模。*IEEE*图像处理汇刊，19(6):1657–1663，2010。[19] Timo Ojala、Matti Pietikainen 和 Topi Maenpaa. 使用局部二值模式进行多分辨率灰度与旋转不变纹理分类。*IEEE*模式分析与机器智能汇刊，24(7):971–987，2002。[20] Ming-Kuei Hu. 基于矩不变量的视觉模式识别。*IRE*信息论汇刊，8(2):179–187，1962。[21] Karen Simonyan 和 Andrew Zisserman. 用于大规模图像识别的超深卷积网络。*arXiv*预印本 *arXiv:1409.1556*，2014。[22] Rayner HM Condori 和 Odemir M Bruno. 通过全局池化测量分析用于纹理分类的激活图。*Information Sciences*，555:260–279，2021。