



CoDAEN: 动态网络进化群落检测算法的基准与比较

都灵, 都灵, 意大利LUCA LUCA, 都灵, 都灵, 意大利

MARCO MELLIA, 都灵政治, , 意大利都灵, LUCA VASSIO,
都灵政治, 意大利都灵

贝洛奥里藏特联邦大学计算机科学系,

巴西

Web数据通常被建模为复杂的网络, 其中整体相互作用并形成社区。然而, web数据会随着时间的推移而发展, 网络社区也会随之变化。这使得动态图中的社区检测(CD)成为一个相关问题, 需要进化CD算法。这种算法性能的选择和评估是具有挑战性的, 因为缺乏一套全面的基准和具体的指标。为了解决这些挑战, 我们提出了进化网络中的CoDen-通信单元检测算法——动态网络中进化CD算法的基准框架, 我们作为开源提供给社区。CoDEN允许我们生成具有已知地面真相的综合社区结构图, 并设计结合9个修改边、节点和社区的基本图形转换的进化场景。我们提出了三个互补的指标。正确性、延迟性和稳定性)来比较没有进化的CD算法。

通过CoDEN, 我们考虑了三种基于进化模块化的CD方法, 解剖了它们的性能, 以衡量社区的稳定性和正确性之间的权衡。接下来, 我们比较了面向真实web的数据集中的算法, 确认了这种权衡。我们的研究表明, 在图中引入记忆的算法最大限度地提高了稳定性, 但在发生突变时增加了延迟。相反地, 通过用之前的解决方案初始化CD算法来引入内存的算法并不能识别新社区的分裂和诞生。这些观察结果强调了CoDENin的价值, 促进了替代进化群落检测算法的研究和比较

概念: ·计算理论→动态图形算法; ·信息系统→社交网络; 计算方法→仿真评估。

附加的关键词和短语: 社区检测, 动态网络, 在线网络, 模块化, 基准测试

1介绍

由于Web相关数据[37,43,44,65]的固有结构和复杂性, 图形和网络的研究在Web的背景下是高度相关的。网络远不仅仅是一个文档网络, 而是一个社交互动的重要平台。社会网络分析, 通常集中于动态复杂

这是早期动态网络[60]中的板凳进化社区检测算法的扩展版本。

作者的联系信息: 佐达诺·波利特, 都灵, 都灵, 皮埃蒙特, 意大利; 电子邮件: giordano.paoletti@polito.it; giordano.paoletti@polito.it; 都灵, 都灵, 都灵, 波灵, Italy; e-mail: lucagioacchini@polito.it; 梅尔拉, 都灵, 都灵, 意大利; 电子邮件: marco.melia@polito.it; 卢卡瓦西, 都灵, 皮埃蒙特, 意大利; 电子邮件: luca.vassio@polito.it; 阿尔梅达, 吉拉斯州联邦大学计算机科学系, 贝洛奥里藏特, MG, Brazi; e-mail: jussara@dc.ufmgbr。



本作品获得创意网络署名国际4.0许可。©2025版权由所有者/作者所有。

acm1559-114x/2025/2-艺术

<https://doi.org/10.1145/3718988>

ACM跨。蜘蛛网

网络[19,77]有助于研究用户之间的关系和影响模式，这对于理解在线社区和社会动态如何随着时间[33]演变至关重要。

与宽网络[28]的连接相比，检测表现出更强内部连接的节点的communities, i.e.groups, 并了解它们的时间动态，可以对它们所代表[17,63,78]的系统的底层结构和功能进行更深入的组合。在实际系统中，任何网络拓扑都会随着时间的推移而变化，因为网络拓扑的自然增长，他们的高度动态的联系，以及他们没有活力的性质。这种变化可能意味着也可能不意味着潜在群落自然进化，其进化速度可能比图中显示的要慢。在进化网络中找到解决CD问题的有效解决方案仍然是一个挑战，需要进化CD解决方案。

进化CD技术[7,23]的目标是在动态图的每个快照上识别高质量的社区，同时在连续的快照中保持社区分配的时间一致性。然而，没有明确的度量来评估和比较进化CD算法的结果。因此，它要求对进化CD算法在不同和可控场景下的（相对）性能进行深入研究，在这种情况下，地面真相的存在允许对性能的权衡进行清晰的理解。

在本文中，我们通过提出CoDEN来解决这些挑战，这是一个用于评估和基准测试具有已知地面真相的动态网络中的进化CD算法的框架。给定一个初始图，我们设计了9个影响边和节点的体外可控转换，反映了在Web上直观地发现的进化模式。我们定义了三个度量标准（即稳定性、正确性和延迟），它们捕获了进化CD算法的互补和理想的特性。稳定性量化的是社区随时间的变化，而正确性和延迟则是监督的指标，利用生成的地面真相来衡量社区的良好程度，以及算法需要多长时间来检测突然的变化。据我们所知，我们是第一个提出系统地使用专门针对这一目标而定制的这些度量标准的组合的人。CoDEN可以让我们了解哪些算法在哪些情况下表现得更好。

与以前的工作[15,69]相比，CoDEN允许生成受控基准，其中边缘和（新的）节点都随着时间的推移而演化，并提供定义良好和精心制作的度量，以彻底比较演化场景中的算法。

虽然CoDEN的设计并没有考虑到特定的进化CD算法家族，但我们在这里使用基于模块化的方法来验证它，因为它们从社会网络分析、生物系统和文献计量网络[18,29,34,80]的各个领域的广泛应用。这些都是基于众所周知的静态CD鲁汶，或贪婪的模块化算法（GMA）[14]，和莱顿算法[73]。模块化是对网络内部重叠社区划分质量的统计度量，它依赖于社区内的实际边缘与具有相同度分布[54]的随机网络中的期望边缘之间的差异。在静态图上，GMA迭代地合并社区以最大化这种度量。同样，Leiden通过合并细化步骤，改进了对最佳解的贪婪搜索。通过将内存合并到图中或从前一个快照继承初始化来适应动态场景（参见第4.2和4.3节）。在这里，我们提出了第三种方法来改进新节点的初始化，并估计co的一致性上一步中确定的材料（见第4.4节）。我们首先在CoDEN上比较了这些方法，然后验证了来自两个不同的web相关领域的真实世界数据集，即网络流量分析和社会网络分析，它们表现出非常不同的网络动态。

我们的结果表明，当进化的社区存在时，稳定性和正确性之间存在权衡。简而言之，最适合的算法依赖于图所呈现的实际变换。CoDEN有助于理解和精确地评估这种模式。为此，我们提供了CoDEN和基于主题的进化CD检测算法作为开源，以刺激设计和评估

其他的进化群落检测算法，包括不基于模块化的方法，如Facetnet[48]，PisCES[49]和DynGEM[30].¹

本文显著地扩展了我们在[60]中引入的初步工作，其中[60]的评估仅限于基于gma的进化CD算法。在这里，我们扩展了分析来包括莱顿算法。此外，我们通过对关键参数的深度敏感性分析来补充我们之前的评估，使我们能够评估这些参数对算法性能的影响。我们还进行了一系列额外的实验来评估可伸缩性，测量每种方法在更大的数据集上的效率。最后，我们在地面真实社区未知的真实数据集上验证了这些方法，并提供了对无监督的真实场景中算法性能的见解。

本文的其余部分组织如下。在第2节中，我们回顾了之前的工作，并在第3节中提供了对CoDEN框架的全面概述。接下来，在第4节中，我们将GMA和Leiden描述为参考CD算法，以及它们的三个进化扩展。我们分别在第5节和第6节中使用CoDEN和面向realweb的数据集来报告和讨论我们的结果。最后，在第7节中，我们提出了最后的考虑因素。

2相关工作

2.1静态图的社区检测

科学文献中有丰富的CD方法的静态图[39]。传统的方法依赖于标签传播算法[25,46]，该算法通过在相邻节点[62]之间的多数投票迭代更新节点标签，通过网络传播一组起始标签（即。社区标识符）。第二个家族是基于集团渗透方法[9,10]，它可以找到cliques，（ie.fully连接的子分支），并基于共享节点[58]连接它们。第三个家族将光谱聚类算法应用于图的拉普拉斯[20,35,55]。这种方法在大型网络中受到可伸缩性问题的影响。此外，它们需要先验了解要检测的社区数量，这在现实场景[38]中很少可用。除了这些方法之外，Infomap[67]优化了通过网络的随机游走集合，聚集了经常一起访问的节点，它提供了一种强调信息流的动态性而不是连通性的替代方法。随着深度学习技术的普及，recent工作开始探索基于人工神经网络的CD解决方案来嵌入节点属性和拓扑信息。因此，在生成的表示[4,22,51]之上应用了传统的聚类算法。

尽管有大量关于CD方法的文献，但基于模块化的算法，如鲁汶或贪婪模块化算法（GMA）[14]和Leiden[73]是现实世界中最常用的应用之一。作为可伸缩和无参数的算法，它们允许我们通过模块化的贪婪优化来找到图节点的层次划分，这是一种量化网络可以划分为不重叠和紧密连接的社区的度量。虽然我们承认它们可能并不总是能带来最好的结果，但它们在文献中的流行程度和在许多应用中的实际有效性证明了它们的选择不是我们研究的重点。

2.2进化群落的检测

在演化图中寻找解决CD问题的有效解决方案仍然是一个人工挑战[19,66]。一种简单的方法是在动态图[15]的连续快照上独立地运行一个静态CD算法m。这将在局部优化每个快照中使用的静态算法的目标（如GMA和Leiden中的模块化）。然而，它还需要一个额外的步骤来跨时间调整社区，通常是通过比较成员重叠的[72]，而没有保证任何临时的口头一致性。因此，有必要

¹the CoDEN代码，以及重现本文所述实验的算法，在https://github.com/SmartData-Polito/Dynamic_CommunityDetection_Benchmark上可用

Paoletti等人。

对于进化CD算法，它明确地考虑了在寻找社区时graph随着时间的进化。

特别是基于模块化的算法，这两个主要的进化扩展是专门在GMA上设计的。它们包括(i)在应用标准CD方法（例如 α GMA）[23,31]之前，将内存项合并到图边的权值中；（ii）通过将节点初始化为之前分配的communities(e.g.sGMA)的成员，将内存组件引入算法。[7]。在这项工作中，我们考虑了这两种方法，并提出了sGMA方法的泛化，该方法明确地考虑了新节点的到来（以前没有讨论过）以及对前一步中确定的社区一致性的局部评估。我们也将它们都扩展到莱顿算法中。

虽然基于模块化的方法很流行，但其他技术提供了明显的优势，在某些情况下可能表现得更好。例如，一些方法，通过只重新计算改变的网络段，保持不变的部分来保持稳定。这一方法支持像面部网络[48]这样的算法，它利用非负矩阵分解，ESPRA[76]，基于倾向的方法，利用拓扑特征和结构扰动来评估网络的相似性，以及PiscES[49]，它使用特征向量平滑的进化谱聚类。其他模式则使用节点或子图嵌入来捕获图的动态属性，并对这些表示进行聚类以揭示社区结构。例如，诸如自动编码器（例如，DynGEM[30]）、图神经网络（roland[81]）和图卷积网络（CTGCN [52]）等架构。

如前所述，我们在这里选择测试我们的CoDEN b实质标记的基于模块化的方法，因为它们在各种场景中被广泛采用。Yet，CoDEN并不是专门为这一算法家族设计的，而且确实可以用于支持在各种方法之间的比较评估。

最近，研究人员通过引入比较框架[15,69]，一直致力于优化进化CD算法。然而，它们的重点仍然局限于网络，其中图的进化仅仅基于边缘的变化，忽略了节点从社区中到达或消失的场景。这些情况在电子邮件网络中不活跃、用户加入社交网络或帖子在转发网络中出现。此外，以前的比较是基于反映图属性的标准度量（如切割比和节点度），但缺乏一套专门为评估算法性能而定制的综合指标（找到的解决方案的稳定性、检测变化的延迟等）。为了解决这些问题，CoDEN处理节点的演化（包括它们的消失、零星存在或添加），并使用稳定性和变化检测延迟等定制指标对算法性能进行全面的综合评估（见3.3节）。

2.3用于Web分析的社区检测

由于Web是一个大型的复杂的连接的entities（e.g.documents网络，使用rs，网页），所以它可以直接将与Web相关的数据表示为图表。因此，采用CD算法来分析社区和识别它们的模式是一种公认的实践[13,32,70]。常见的应用程序是提高终端用户的体验，在识别的网页托管类似内容快速索引[56]和识别社区表现出类似的兴趣边缘缓存在线服务[3,68]，或检测恶意社区商务平台防止在线欺诈[16,47,50]。

在众多不同的应用中，CD技术特别适合于分析与网络相关的数据，特别关注社交网络分析[12,33]。社交媒体通常被建模为动态复杂网络或图，其中实体（节点）相互作用，形成连接（边缘），随着[19,77]的推移而进化，通常表现出强大的社区结构。分析这样的社区

ACM跨。蜘蛛网

让我们了解一些现实世界中的事件，比如[21,53]的传播，[59]-19大流行[59]期间的疫苗辩论和政治两极化[11]，如2020年U.S.election欺诈辩论[2]。

人们对网络内部安全和隐私问题的日益担忧，促使研究人员在网络安全应用程序中使用CD算法。值得注意的例子是，建立一个相似的网络的软件包使用的安卓移动应用程序指纹熟悉的恶意软件[6,41,42]。在交通分析应用中已经获得了成功的结果，其中将网络交通建模为一个图，允许识别在面板中表现出类似行为的主机社区，并检测新出现的threats（e.g.botnets）[26,27,40]。

在这项工作中，我们考虑了两个世界的数据集，它们是建模网络不同现象的进化图的例子：2020年美国在推特[2]上的选举欺诈辩论，以及僵尸网络或互联网扫描仪[27]的新攻击。

3CoDEN：框架概述

进化网络中的通信单位检测算法，或CoDEN，由(i)综合性的组合组成

生成具有已知社区结构的动态图；(ii)一组涉及节点和边缘集的时间变化的转换，涵盖了不同的现实场景；(iii)三个互补的度量，能够评估替代进化CD算法的性能。

我们定义了一个动态图 $\{\mathcal{G}^t\}_{t=1}^N$ ，as N 个无向图 $G^t = (V, E^t)$ 序列，其中 V_t 和 E_t 分别是 t 快照上的节点和边集。即， $edge = (u, v, w) \in E_t$ 表示节点 u, v 之间的连接，权重为 $w \in (0, 1)$ 的 w 。以 $singlesnapshot$ 为重点，我们将社区 C_s 定义为与网络其他部分相比具有更高内部连接密度的节点。因此， $U_{eccc} = V$ ，其中 C_s 是检测到的社区集合。在这项工作中，我们将重点关注不重叠的社区。

3.1生成合成的graphs

对进化CD算法的评价往往因缺乏基本真理而受到挑战。因此，我们提出了反映拓扑图转换的综合场景，类似于真实世界的进化网络，并提供了一个已知的地面真相，可以与替代解决方案进行比较。

具体来说，我们通过兰开奇内蒂-福尔图纳托-拉基（LFR）基准[45]运行体外实验生成合成图，这是一种被广泛认可的工具，用于创建展示社区结构的复杂网络，并提供已知的ground-truth（i.e.the指定的社区）。虽然可以使用任何基于综合社区的图生成算法，但我们在这里选择了LFR，因为它可以生成真实的图，以幂律度分布和异构社区规模为特征。我们将权值随机初始化到边值均匀分布在 $(0, 1)$ 范围内的边。

3.2图形转换

我们提出了9种转换，分为3个场景。(i)噪音，(ii)变形性噪音和(iii)破坏性噪音。这些转换可以应用于不同的方法。我们在表1中提供了对设计场景的简要概述。尽管很简单，但这些转换将真实世界图的复杂动态分解为单个事件，允许识别算法的弱点，并且它们受到了之前的工作[15,57,66]的启发。作为未来的工作，我们的框架可以扩展到探索它们的组合，以创建更复杂的转换序列

[†]Each快照可以用连续时间模型表示单个边或节点变化，也可以定期分批分组。
3为了清楚起见，我们省略了后缀 t 。

Paoletti等人。

表1.CoDAN基本间隙h转换的概述。

场景	转换	描述
噪音	膨胀 间歇开关	在社区内添加节点 为一个快照的节点 关闭/打开了可针对多个快照的节点
变形的	合 并 分 裂 死 亡 出 生	合并社区对形成子社区 解散社区 聚合新社区中的节点
破坏性的	混合去 除	社区结构的衰减，删除节点

3.2.1噪声场景。在这种情况下，我们通过3个变换来扰乱图，而不影响LFR图提供的初始地面真相：

(i)Expansion.在每个快照中，我们会逐步向现有的社区添加新的节点。我们像初始LFR网络一样，通过优先连接[8]随机选择连接到社区内的节点，以保持节点度的异质性。新节点在整个模拟过程中持续存在。

(ii)Intermittence.在每个快照 t 上，我们删除一个节点，并添加+1。为了保证表征社区的枢纽节点的持久性，每个节点 k 具有被选择的概率 $p_{k \times \delta t}$ ，where δk 为节点度。

(ii)Switch.在每个快照 t 上，我们删除了一个fraction ϕ_{swi} 个节点。每个节点 k 以概率 $P_{k \times}$ 被移除。非活动节点以概率 $p_{backx} \exp^{2off-Y}$ 重新出现，其中 t_{off} 是节点已经处于非活动的快照数量， $y \in N$ 是控制非活动平均持续时间的参数

这些转换反映了世界场景，例如社交媒体平台上的用户，其中活动表现出时断断续的模式，新用户可以加入系统，而不需要改变底层的社区结构。

3.2.2变形。在这种情况下，我们会随着时间的推移逐步改变社区结构和网络拓扑结构。我们从最初的地面真相 GT^0 for 社区开始，最后以最终的地面真相社区 GT_N 结束。

i)Split.我们选择了 n 个群落，并通过在每个快照上（数量 t ）减少连接子群落的边的权重，逐步将它们划分为两个子群落。

(ii)Merge.我们选择了 n 个群落对，并通过在每个快照上增加连接它们节点的边的权重，逐步将它们合并为一个更大的群落对。

(ii)Death.我们选择 n 个群落，通过减少每个快照的 t 个群落内边缘权值来逐步溶解它们；同时，我们通过对节点程度的优先附着来加强这些节点指向其他现有群落的边缘；

(iv)Birth.我们减少了每个社区中某些节点的边权值（数量为 r ），并在它们之间建立连接，形成 n 个新的社区。

我们通过参数 $t \in [0,1]$ 来控制转换发生的速度，它表示每个快照的边缘权值的变化。在快照 t 时，我们通过修改前一个快照 w^{t-1} 的权重来获得一条边 w^t 的权重。正式上， $w^t = w^{t-1} \pm t$ ，我们剪辑 $w^t \in [0,1]$ 。

这些转变反映了在以特定主题为中心的社交网络中新社区的兴起过程中可能发生的转变，或者相反，对特定主题的广泛兴趣

3.2.3破坏性。在这种情况下，我们会逐步破坏该网络的社区结构。我们定义了两个转换：

(i)Mixing.在每个快照中，我们通过从节点的fraction ϕ dis中重新分配一条边的目标来关闭初始社区，同时保持加权重度分布。(ii)Removal.在每个快照中，我们逐步随机删除一个fraction ϕ dis的节点。

这些转换反映了现实的场景，即向节点的无序排列转变，如社会混合，当来自不同社区的节点彼此随机连接，或者当一些节点离开网络，如用户不再对网络感兴趣。

3.3质量指标

静态CD评估中的指标（例如，电导[61,79]-跨越聚类边界的年龄边界百分比；内部密度[61]-集群内的边缘比率，所有可能的边缘；）都是无监督的metrics（i.e.they，不需要对已发现的群落的先验知识）。因此，它们不提供关于所检测到的分区正确性的见解。此外，它们反映了图的特定属性，而没有提供关于通过不同快照演化图中CD算法的性能的信息。在这里，我们定义了三个捕获互补因素的度量标准：稳定性（未受监督）、正确性和延迟（有监督）4。这些指标依赖于调整后的互信息（AMI）[75]，i.e.a相似性

在标签集之间的测量调整为随机机会，归一化到范围从0到1

给定在快照t中找到的社区Ct，我们定义的度量如下：

稳定性(S)。形式上， $st=AMI(C_t-1, c)$ 。它是社区分配中时间一致性的一个代理。稳定性越高，表示两个快照之间的社区成员变化越少。我们还通过平均N个模拟模型ts上的St来计算总体稳定性

正确性(K)。给定一个地面事实（GT），我们评估了所考虑的进化CD算法在在快照 $K'=AMI(GT, C)$ 检测转换时的正确性。在噪声场景中，变换不影响初始地面真实值，我们设置 $GT=GT^0$ 。相反，在变形场景中，由于我们不能在转换结束之前定义最终的地面真相，所以我们考虑正确性Kat作为最终的快照，设置 $GT=GTN$ 。反过来，在破坏性的转换中，最终的地面真相GTN是不可用的，我们也不考虑其正确性。

“延迟(D).”它是算法检测新社区结构的变化所采取的快照数量。我们只在变形变换中定义这个度量，并使用它来评估每个算法的响应性。也就是说，我们将交叉点CP定义为社区分配比 GT^0 更接近GTN的第一个快照。形式上， $CP=\min\{t: AMI(GTN, c) > AMI(GT^0, c)\}$ 。我们认为延迟D是交叉点和转换开始时的快照之间的差值。如果没有检测到exist，i.e.the变化，我们将延迟设置为N。

4个基于模块化的进化CD算法

我们设计CoDEN作为一个工具来评估任何CD算法的进化。然而，鉴于它们在现实世界中的广泛应用，我们将评估重点放在基于模块化的算法上。

给定一组不重叠的社区C，图模块化[54]量化了定义明确的社区的数量。简而言之，模块化将社区内连接的密度与随机网络中期望的连接进行了比较。形式上，对于社区 $c \in C$ ，其模块化定义为

$Q_c = \frac{w_{in,c}}{w^*} - \left(\frac{w_c}{2w^*}\right)^2$ ，其中 w^* 是图的所有边权值的和， $w_{in,c}$ 是社区内*Notice的和，后者二需要地面真实的信息。这意味着合成数据。

Paoletti等人。

边权重，我们是社区 c 中节点加权度的和。图的模块化被定义为其communities，i.e. $Q = \sum_c Q_c$ 的模块化之和。

4.1 鲁汶（GMA）和莱顿公司的静态图

鲁汶方法，被称为贪婪模块化算法（GMA）[14]，通过贪婪地和迭代地最大化图的模块化性来操作。在第一次迭代中，GMA将每个节点分配给一个唯一的社区。在每次后续迭代中，它通过评估将一个节点移动到相邻社区或合并两个相邻社区所获得的模块化增益来合并社区。如果模块增益为正，则接受更改，并且过程继续。详情请参考[14]。

莱顿算法[73]通过结合了细化步骤，改进了GMA。在GMA的每一次迭代中——社区被合并以优化图的模块化——Leiden引入了两个关键的修改：首先，它放宽了模块化增加的约束，允许节点与任何显示性改进的社区合并；其次，它通过选择性地分割社区来细化检测到的分区，这有助于降低图的模块化。

GMA或Leiden可以天真地用于进化图中，通过在每个快照[15]上独立应用任何一种算法。我们将这种方法分别称为独立的GMA方法和独立的莱顿方法。这种策略可能会导致社区中的时间不一致，例如，由于在第一次迭代时对标签的随机初始化。研究人员提出了两种进化方法，通过向图本身添加内存或通过向算法中引入内存组件来最小化这种冲突。虽然这些方法最初是作为GMA[7,23]的扩展而提出和评估的，但我们也认为它们是对莱顿的扩展。

4.2 向图形中引入内存（ α GMA and α Lei）

这种方法[23]包括引入随着时间推移的边的持久性，增加社区的稳定性。该方法在边缘权重中引入了一个内存项 α 。嗯，一个以前存在的编辑者

拍照期间的重量为 w_t ， w^{t-1} 是 $t-1$ 的实际重量更新了 $w^t = (1-\alpha) \cdot w^* + \alpha \cdot w_{t-1}$ 。

If $w^{t-1} = 0$ ，我们设置 $w^t = w$ 。在这样的图上应用独立的GMA或Leiden，会随着新信息的可用而逐步进行模块化调整。我们将这些方法称为 α GMA and α Lei。

4.3 向算法引入内存（sGMA和sLei）

这种方法改变了GMA和Leiden的初始化过程。虽然原始方法在快照上将每个节点分配给一个独特的社区，但这种方法维护 $t-1$ 活动节点以前的社区成员关系。每个新节点都被分配给一个新的不同的社区。给定这种初始化，我们运行原始的CD算法，允许节点在初始分配不合适时调整它们的社区成员身份。我们将这种方法在GMA和Leiden上的应用分别称为sGMA和sLei。

4.4 基于邻里的新方法（NeGMA和NeLei）

sGMA和sLei都继承了前一个快照的社区结构，而新出现的节点则被分配给一个新的社区。为了改进这种选择，我们建议根据当前快照中新节点附近的大多数类别来初始化新节点的标签。接下来，我们通过评估每个社区的一致性来细化初始化，只有当其局部模块化超过 threshold_{0Q} 时才保留它。

为了说明，我们提供了图1中我们的方法如何工作的错误视图。也就是说，(i)我们从前一个 snapshot_{t-1} 中检测到的社区开始；(i)在快照 t 时，图演变为一些新的边（红线）和节点（红色方块）出现，并且一些 d 出现（偏差）；(i i) 我们已经分配了

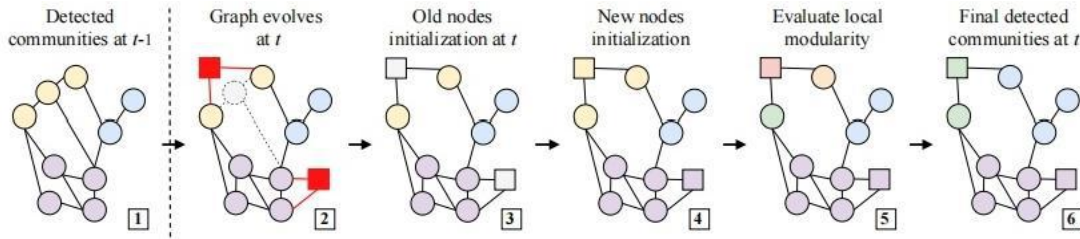


Fig.1.Neighbourhood-based方法：从快照1开始的快照t节点初始化。不同的颜色代表不同的社区；广场表示新节点；虚线元素表示消失的n个节点和边。

将att-1的现有节点（彩色ci方圈）检测到的社区（asinsGMA/sLei）；（iv）我们通过对其邻居的标签进行多数投票，将新节点分配给一个社区（彩色方块）；（v）对于每个社区c，我们计算相对于前一个快照的局部模块化的增益： $\Delta Q_c^t = Q_c^t - Q_c^{t-1}$ 。如果 ΔQ_c^t 低于阈值 θ_g ，我们将其每个节点分配给一个独特的社区，如传统的GMA/Leiden（图左上角的前黄色社区）；最后，（vi）我们应用独立的GMA/Leiden与最终的社区初始化。我们参考了推导出的算法从将这种方法应用于GMA和Leiden，分别作为基于社区的GMA或基于社区的Leiden（或简称为NeGMA或NeLei）。

我们提出了基于遗传的方法的伪编码，for α GMA/aLei和sGMA/sLeiin附录A。

5.在合成图中的评价

我们使用CoDEN基准测试来评估进化CD算法。我们重复每个实验，生成100个独立的动态图。我们设置了LFR参数 $\mu=0.2$ [45]。对于每个动态图，我们生成了 $N=150$ 个时间快照。我们在快照25处触发tr转换的开始，并在快照125处结束它。对于每个动态图，我们对每个快照执行每个算法的10次独立运行。为了解决(i)图生成阶段和(ii)算法的随机性所引入的可变性，我们对每个动态图上10次运行的结果进行了平均值。随后，我们评估了100个动态图中这些中位数的中位数和自举的99%置信区间。

对于转换，我们将间歇性节点的比例设为 $p_{int}=0.2$ ；对于 $switchP_{swi}=0.005$ and $x=10$ ；对于混合和混合，我们均匀地采样 p_{dis} between $[0.005,0.02]$ 。当没有明确说明时，我们设置转换速度 $t=1/100$ ，受影响的社区数量为 $=3$ 。

关于算法，正如Elgazzar等人的[23]所建议的，我们将 $set_{\alpha}=0.8$ in α GMA and α Lei，而对于NeGMA和NeLei，我们将 $set_{\theta_g}=0$ 作为一个通用的中间值。

5.1本设备的正确性和稳定性

5.1.1GMA-based算法。在表2中，我们总结了基于gma的算法在所有模拟算法中的性能

情节这些符号指的是每个算法通过使用独立的GMA获得的相对性能/损失，作为基线。没有一种算法在所有场景中都更好；每一种都根据转换提供不同的权衡。在稳定性方面，所有的进化算法的表现都与基线一样好或更好。然而，这通常是以正确性的降低为代价而获得的。

Paoletti等人。

表2。测试的基于GMA的进化算法与独立GMA（基线）相比的正确性K和St能力S概述。‘=’表示性能变化 $\leq 5\%$ ，‘+’表示在（ $5n\%$ ， $5(n+1)\%$ ）范围内有改善。“-”数n”表示 $[-5(n+1)\%$ ， $-5n\%$ ）范围内的减少。

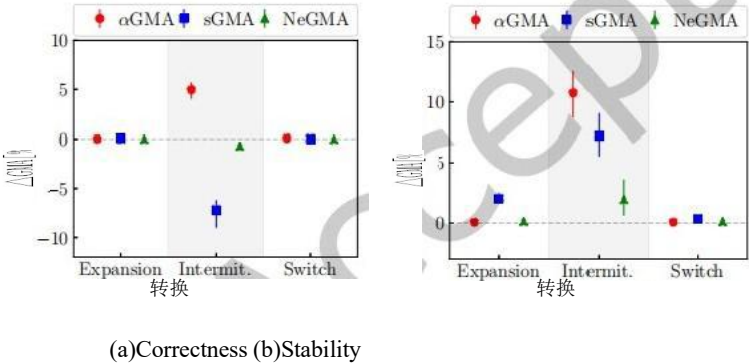
	噪声场景扩展接口。			变形场景合并分裂出生死亡				破坏性的场景，混合去除	
GMA	0.84	0.95	0.99	1.00	0.86	0.91	1.00	n. a.	n. a.
Ki	=	=	=	=	=	=	=	n. a.	n. a.
α GMA	=	=	=	=	=	=	=	n. a.	n. a.
sGMA	=	=	=	=	=	=	=	n. a.	n. a.
内格马	=	=	=	=	=	=	=	n. a.	n. a.
GMA	0.98	0.90	1.00	0.98	0.93	0.88	0.87	0.99	0.99

S α GMA=++ = = = = =

sGMA=+=+++++=

NeGMA= === = + + ==

记住，我们对噪声和变形场景的相关计算方法不同（见第3.3节）



基于Fig.2.Noise和lgma的算法。中值增益和引导带99%置信区间的正确性和稳定性。

在图2中，我们报告了在噪声场景中与独立GMA相比的中值正确性和稳定增益/损失，其范围从快照25到125。在扩展和开关转换方面，所有算法的性能都与GMA相同，增益高达2%

内存terminαGMA通过提高正确正确性（+5%）和稳定性（+11%）增强了对间歇性噪声的鲁棒性。事实上，图3b定性显示，社区检测到αGMA快照15保持一致快照90，与独立GMA，而不是波动和未能维护原始设置，如图3所示。sGMA初始化，反过来，提高了7%的稳定性间歇性转换，但失去了7%的正确性相比独立GMA。在图3c的展示中，我们可以看到间歇性噪声导致sGMA在快照60处合并社区，而在后续快照中算法的初始化阻碍了对原始社区结构的进一步检测和重建。相反，NeGMA克服了sGMA的限制，平衡了正确性和稳定性，正确性降低了<1%，稳定性增加了≈3%。为了支持这一点，我们在图3d中的示例中看到了原始社区的部分恢复 快照45和75之间的联系。

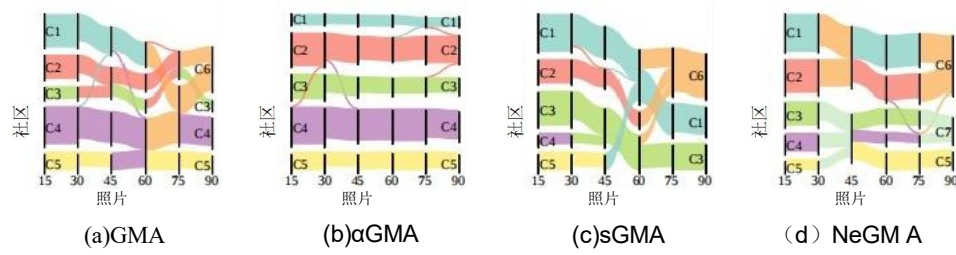
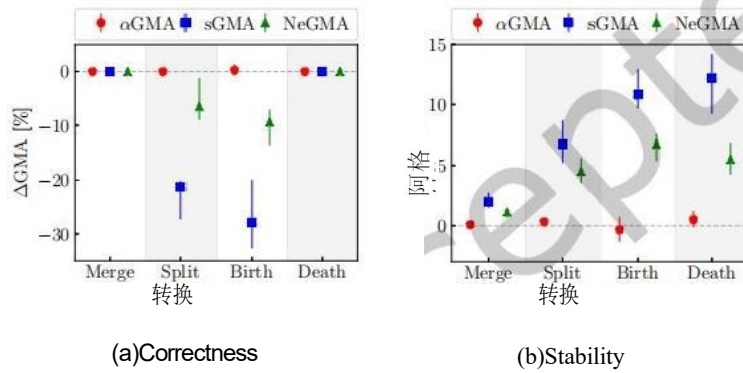


Fig.3.GMA-based算法。由区间变换扰动的图的动态演化的例子，展示社区分配网络。使用步骤15将发生变化的快照样本从15放大到90。



基于Fig.4.Morphing和gma的算法。中值增益和引导了99%的置信度在正确性和稳定性

变形变换图4报告了由变形变换产生的中值校正率和稳定性增益/损失。值得注意的是，αGMA在所有设计的转换中执行的性能与独立的GMA相当。反过来，sGMA在稳定性方面领先，从合并+3%到死亡+13%。然而，它的正确性受到了影响，分裂和出生转换的损失超过了20%，这反映出它难以识别新的社区。NeGMA再次证明了在正确性和稳定性之间的最佳平衡，总体7%的最大显著性增加，同时将正确性不超过~10%。

我们还测试了破坏性的场景（表2，为了简洁起见，省略了完整的结果）：与独立的GMA相比，没有一个被测试的算法能导致实质性的改进，达到0.99的稳定性。

5.1.2莱顿算法。在大多数转换中，Leiden引入的社区细化阶段在转换后检测正确的社区方面非常有效。因此，与独立执行相比，莱顿蛋白的进化扩展没有一个实质性的改进，除了以下三个转变：间歇性、出生和死亡。图5报告了这三种转换导致的中值正确性和稳定性增益/损失。

Paoletti等人。

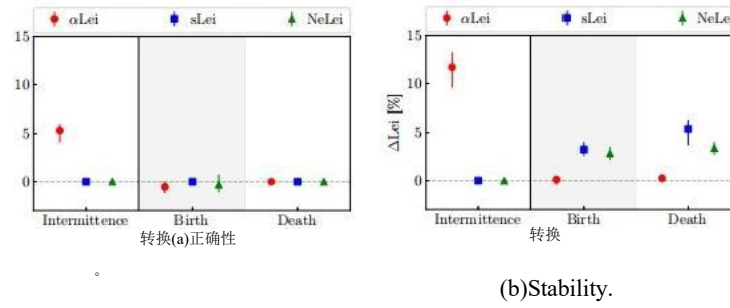


图5。基于莱顿的算法。中值增益和引导的99%的正确性和稳定性置信区间三个差异较大的转换。垂直的实线区分了噪声和变形的场景。

包含内存三元 α Lei图生成过程忽略了间歇变换中引入的拓扑变化。在 α GMA中，与独立的莱顿相比，这使得正确性显著提高了5%，稳定了12%（事实上，莱顿不断修改社区以反映节点的进出）

在变形场景（出生和死亡）中，所有算法具有相似的correctness.sLei，被确认为最稳定算法，+分别为4%和+7%。 α Lei的性能与独立Leiden和NeGMA的稳定性略低于sLei。

鉴于Leiden似乎对主体变换中采用的进化方法不敏感，在本文的其余部分中，我们重点比较基于gma的进化y算法外卖：(i)与独立的GMA和Leiden相比，aGMA和aLei中的记忆项引起最小的变化，除了间歇性转换；(ii)sGMA提高了稳定性，但未能检测新兴社区，而NeGMA在变形转换方面证明了平衡，在正确性和稳定性方面实现了竞争性能；(ii)对于大多数转换，莱顿i不受进化初始化的影响，除了变形（出生和死亡），sLei和NeLei都提高了独立莱顿稳定性，高达~5%，而不损失正确性。

5.2响应能力

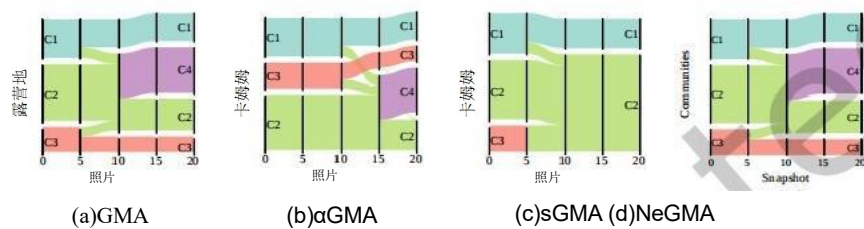
我们通过监测变形转换中的延迟度量来评估基于进化gma的解决方案在检测变化方面的响应性。我们为每个转换生成100个独立的图，并为N个=20个模拟快照运行每个算法的10个实例。我们在快照=10上执行一个instantaneous (i.e. $t=1$) 变换，并测量延迟D来检测它。给定这个设置，D在[0,10]范围内。延迟时间越长，响应能力就越低。

表3报告了每个转换的中位数延迟和检测到的转换的分数（即交叉点CP为defined⁵）。正如预期的那样，独立GMA使用D=0检测精确快照的转换。 α GMA的记忆项反过来迫使整个图以较慢的速度进化，导致系统延迟和较低的反应性（从mD=1分裂到D=4.6出生）。图6b中可视化的示例证实了这一点，与图6a中的独立GMA相比，第四个社区的延迟为5个快照。

⁵我们认为，交叉点（CP）是第一个社区分配与最终的地面真相GTN比与初始的地面真相GT更接近的快照。

表3. 基于GMA的算法。在变形场景中，中间延迟和到达交叉点。最佳结果用粗体突出显示，关键结果用红色突出显示。

	延迟D			达到CP[%]		
	合并	分裂出生	死亡	合并	分裂出生	死亡
GMA	0.0	0.0 0.0	0.0	0.98	0.88 0.90	1.00
α GMA	3.0	1.0 4.6	2.0	0.98	0.90 0.90	0.96
sGMA	0.0	最大0.0	0.0	0.98	0.04 0.98	0.82
内格马	0.0	0.0 0.0	0.0	0.98	0.94 0.88	1.00



时间10时刻即时杂出生转换的图的动态演化的Fig.6.Example，显示社区分配（基于gma的算法thms）。

在合并、出生和死亡方面，sGMA匹配独立GMA的延迟，但mi在分裂中96%的CPs。通过保留以前的快照分配，sGMA无法检测到其他社区。这种限制也会影响出生转换：虽然sGMA可以检测到98%的CPs，但它仍然与theGT^{as}分配相关（如图6c所示），因此未能捕获新的社区，并导致GTN上的ina-37.8%correctness下降。

最后，NeGMA检测所有情况下的中值延迟D=0的瞬时转换。它对局部模块化 ΔQ 期间节点初始化的评估解决了sGMA的限制，细化了初始分配，并成功检测到了94%的分割。值得注意的是，尽管NeGMA的Cp比sGMA少10%，但NeGMA成功地检测到了转换，如图6d所示，其正确性为0.97，而GTN上sGMA的正确性为0.60。

要点：(i) α GMA的记忆项在检测瞬时变换时引入了系统的延迟；(ii) sGMA无法捕捉新社区的突然出现；而(ii) NeGMA在大多数瞬时变形转换中表现出较高的响应性和检出率。

5.3敏感性分析

在本节中，我们将研究关键算法参数在正确性和稳定性方面的影响。即，我们评估了NeGMA的模块化阈值 θ_o 和 α GMA的记忆项 α 的影响。

5.3.1 NeGMA：模块化对threshold θ_g 的影响。我们依赖于相同的100个合成图，通过生成

如第5.1节。对于每个图，我们系统地改变了参数 $\theta_g \in [-0.1, 0.1]$ 。然后，我们将对每个值执行10次NeGMA运行，平均结果。

回想一下，参数 θ_g 作为第一次NeGMA初始化后社区内最大可容忍模块化下降的指标。较高的正阈值使NeGMA对这种减少的响应较差，将现有节点的初始化与sGMA对齐，并保持社区从

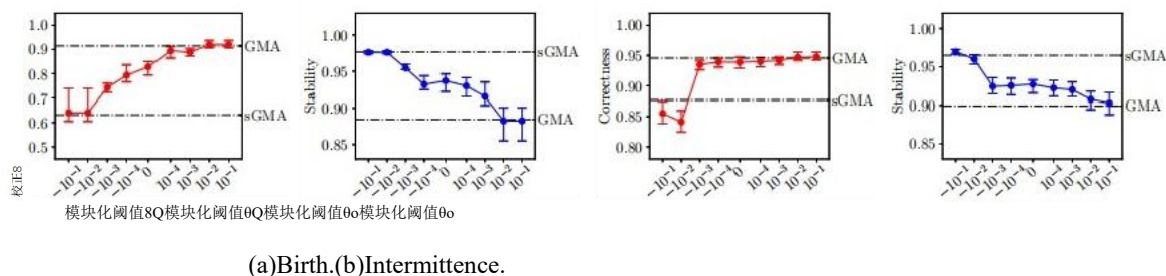


Fig.7.Impact对出生（变形）和间歇（噪声）转换的 θ on的正确性和稳定性。

之前的快照。相反，一个更多的negative θ g放大了社区内对模块化下降的敏感性，这意味着所有现有节点的初始化。在这种情况下，NeGMA表现为一个独立的GMA。

这些观察结果在图7中得到了证实，它报告了impactof θ g关于出生（变形）和间歇性（噪声）转换的正确性和稳定性的结果。6为了比较，还显示了独立GMA和sGMA的结果（水平线）。

让我们首先关注出生转换，结果如图7a所示。这里，when $\theta \leq -0.1$ ，NeGMA与sGMAachieving ≈ 0.60 of的正确性和 ≈ 0.98 的稳定性一致。相反，在 $\theta \geq 0.01$ 中，NeGMA对模块化下降变得不那么敏感，其表现类似于独立的GMA（正确性and ≈ 0.89 of稳定性的 ≈ 0.90 ）。当 θ 下降 $[-0.1, 0.01]$ 范围时，正确性逐步提高30%，稳定性降低9%。在噪声场景中也出现了类似的考虑，如图7b所示。这里，off θ g对算法稳定性的影响与出生转换的趋势一致。然而，与sGMA相比，correctness， $\theta \leq -0.01$ 诱导的边际性能下降为2%。值得注意的是，更剧烈的性能变化体现了for $\theta > -0.01$ 。

总而言之，这些发现重申了NeGMA作为一种通用模块化的算法。Fine-tuningthe θ e参数允许在稳定性和正确性之间取得平衡，并考虑特定图演化的尾部性能。

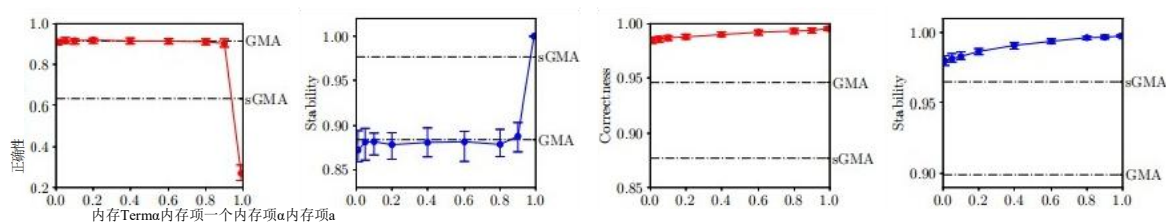
5.3.2aGMA: 内存term α 的影响。aGMA上的记忆术语加权了h存储的重要性

更新图形时针对最近数据的信息。For $\alpha=0$ 边缘权值的设置只考虑当前快照的权重（如独立的GMA），而对于 $\alpha=1$ ，边缘权值被设置为之前快照的权重。我们依赖于通过CoDEN和vary $\alpha \in [0.05, 0.99]$ 生成的相同的100个图。对于每个值和每个图，我们运行10个aGMA算法实例，对得到的结果取平均值。

图8显示了 α 对出生和间歇性转换的正确性和稳定性的影响，如变形和噪声场景，respectively⁷。首先，对于变形ns（图8a），impactof α 可以忽略不计。For $\alpha \leq 0.90$ ，aGMA在稳定性和正确性方面都与独立的GMA一致。相反，for $\alpha=0.99$ ，在当前快照上的边缘权值的对比离子是边缘的（i.e.0.01），因此该算法阻止了图的演化，保留了过去结构。因此，正确性从0.9大幅下降到0.3，而稳定性从0.88提高到1。

当关注图8b中的间歇性变换时，结果与第5节的考虑因素一致。值得注意的是，与独立的GMA和sGMA相比，aGMA在正确性和稳定性方面都有改进。由于所考虑的转换导致了节点的零星存在，即使是一个很小的

6我们评估了所有9个获得相似的质量结果7我们测试了所有9个转换，得到了相似的结果。



(a)Birth.(b)Intermittence.

记忆项 α 对aGMA对出生（变形）和间歇（噪声）的转换的正确性和稳定性的Fig.8.Impact。

值为 $\alpha=0.05$ 允许平滑图拓扑的变化，保持节点之间过去的连接，这导致了 $\approx 4\%$ 的正确性增益（aGMA compared to ≈ 0.94 of 独立GMA的0.99）和 $\approx 2\%$ 的稳定性增益（0.98比sGMA的 ≈ 0.96 ）。值为 $\alpha=0.99$ 得到一个几乎静态的图，实现了100%的正确性和稳定性。

要点：(i) 较低的threshold θ_g 值使NeGMA对模块化的响应降低（与sGMA对齐）；(ii) 较高的值of θ_g 提高了NeGMA对模块化的敏感性，从而初始化更多的节点（与独立的GMA对齐）；(ii) 内存term α 对of α GMA的正确性和稳定性有疏忽或有限的影响

5.4可伸缩性

最后，我们评估了随着图的发展，该算法的执行时间。 α GMA的收敛时间 \approx 比独立的GMA慢20%。通过 α 保留节点之间过去的连接会导致算法在每个快照上处理一个不断增长的图（更多的节点和边）。这影响了过度的可伸缩性，使得 α GMA不太适合不断进化的动态图。相反，与算法相比，sGMA是最可扩展的算法，执行速度是前者的两倍。通过为现有节点保留先前检测到的社区的成员关系，它有效地引导算法，生成一个次优解作为初始化。最后，NeGMA实现了与基线相当的收敛时间。

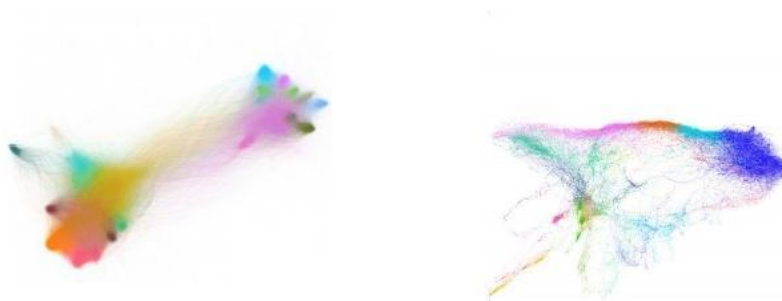
对于莱顿算法，值得注意的是，与独立的莱顿算法相比，只有sLei在收敛时间上显示出增强，因此在可伸缩性方面具有显著的优势。然而，对于莱顿来说，执行时间的改进较小，约为18%。相比之下，NeLei的收敛时间明显慢于基线：与GMA不同，平均下降速度为29%。在诱导记忆白蚁时， α Lei持续延长了收敛时间，平均延迟约为30%。

外卖：(i) 由于增加了内存，aGMA比GMA慢20%，使得动态图形hs的可伸缩性降低 (ii) sGMA的速度是基线的两倍，而NeGMA与基线相似。(ii) sLei使莱顿的收敛率提高了18%，而NeLei则降低了29%

6.在现实世界的图形中的评估

在本节中，我们将测试GMA算法对动态真实和真实的web相关场景的进化扩展。我们关注基于GMA的算法由于方法的性能权衡中分析的各种合成场景，以及由于大量使用GMA作为CD方法应用于各个层收集的数据网络（例如，[5,71,74]等）。

Paoletti等人。



(a)选民欺诈数据集。(b)网络望远镜数据集。

使用ForceAtlas2[36]从考虑的数据集生成的快照图的Fig.9.Visualisation。不同的颜色表示由一次独立的GMA检测到的群落。

在这里，我们依赖于两个著名的开放数据集，即网络欺诈和网络望远镜，它们分别捕获推特用户之间的交互和网络扫描，使用它们来分析在无监督场景中替代算法的相对运行。我们选择这两个真实的数据集，是因为它们代表了Web的两个不同的通信层，即较低的网络流量层和应用程序（社交网络）层，它们往往表现出非常不同的进化特征。一方面，通过构造，在网络望远镜数据集中，节点的嵌入逐渐走向平衡状态，导致图随着时间的急剧变化。因此，该数据集可以作为社区结构发生变化的情况的代表——这些特征通常不能事先知道，只能由于任务的无监督性质而进行假设。另一方面，选民欺诈数据集是社会案例的例证 活动高度异构的媒体用户分析，在图结构中产生大量的噪声。此外，与网络望远镜相比，我们预计社区的进化过程要平稳得多。

请注意，在数据集中没有基本事实。因此，与合成案例不同，这里我们不能评估正确性或响应性。因此，除了稳定性（即使在没有地面事实的情况下也可以计算）之外，我们还将模块化作为每个快照中检测到的社区质量的度量。

6.1数据集

6.1.1选民欺诈数据集。这个数据集[1,2]包含了一个推特网站，包含与美国相关的关键词和标签。从2020年10月23日到2020年12月16日。我们建立了一个加权时间转发图，其中节点为Twitter用户，边表示用户之间的转发。边缘权重表示两个用户之间的转发次数。我们每周汇总一次数据，获得9个快照。在每个快照中，我们过滤掉了孤立的节点，只保留了巨大的弱连接组件（84%的节点）。我们在Tab中显示了结果图的摘要。 4.

图9a提供了一个定性的概述，显示了一个快照中提取的子图，其颜色代表在独立GMA的一次运行中检测到的社区。虽然有两个更大的边占主导地位，但GMA确定了一个更颗粒状的群落结构。

该数据集例证了社交媒体用户分析的案例，其中活动高度异构，在图结构中产生大量噪声，群体的极化可能随时间变化，使其成为社交网络图的代表。

表4.真实数据图的图形和每个快照的度量和平均值汇总：|E|：边总数IVI：总节点数N：快照数（|E*|）t：快照的平均边数（|E|/t）：快照的平均节点数（N/t）$\langle \text{deg}(0) \rangle_{a,t}$：快照的平均点头度

	E	重要 值指数	N	$\langle \text{deg}(0) \rangle_{a,t}$	$\langle \text{IVI} \rangle$	(deg (0)) _{a, t}
选民欺诈	~24M	~158k	9	~2.7M	~27k	437
网络望远镜	~900k	~1M	30	~35k	~13k	5

6.1.2网络望远镜数据集。我们在这里依赖于由网络望远镜收集的Web，i.e.raw包跟踪的一个不同方面。网络望远镜由一组不承载任何服务的被动ip地址组成，并根据定义为[64]接收未经请求的流量。这种生产流量的缺乏使得这样的网络可以用来观察主机在Internet（e.g.benign例行扫描上运行协调扫描或由僵尸网络[71]执行的分布式攻击）。该数据集[26,27]包含一个月的原始包跟踪，由2022年10月我们校园网络中的24网络望远镜收集。我们依赖于在考虑的网络中活跃的主机的主机embeddings，i.e.numerical表示，以突出类似的扫描行为（我们参考我们之前的方法几何细节[26,27]）。我们在嵌入数据之上构建一个加权k-最近邻图[24]8，并运行CD算法来检测协调主机组。我们在表4中显示了结果的动态图的一个总和。通过构造，嵌入逐渐演化为eq平衡状态，r随着时间的推移，这导致了图的快速转换。因此，该数据集代表了社区结构转移的情况——这种特征通常不能事先知道，由于任务的无监督性质，只能进行假设。

在图9b中，我们提供了网络望远镜图的一个snapshot的定性概述。在这种情况下，不同的部分并不像VoterFraud数据集那样清晰，尽管每snapshot（i.e.tens数千个节点的数量相当），边的数量较少（每个快照与3M相比）导致了更不密集的图和较弱的社区凝聚力。

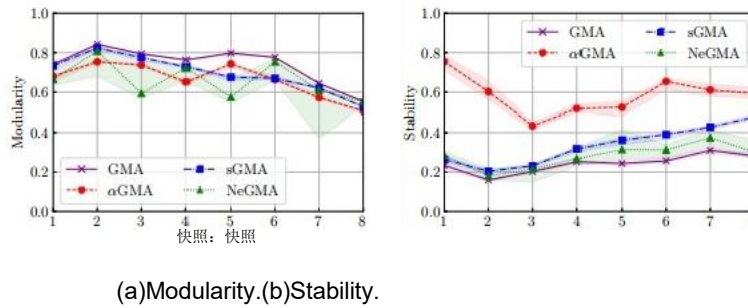
我们强调，这两个数据集在我们的有效工作中提供了多样性，捕获社交网络上用户Web（i.e.interactions的不同方面，并在包级别上具体分析）。网络望远镜数据集显示了图表随着时间的快速和显著的变化，反映了社区结构动态地向平衡发展的情景。相反，VoterFraud数据集呈现出一个更平滑的进化过程，典型的社交媒体用户活动，其特征是高异质性和图结构中的大量噪声。这也影响了社区的意义：在选民欺诈数据集中，社区表明了政治两极化（两个更大的社区），而在网络望远镜中，它们突出了扫描相同或互补目标的协调主机。

6.2实验结果

我们现在讨论了在两个世界数据集上运行50个独立GMA实例及其三个进化扩展所获得的模块化和稳定性结果。为了进行这些实验，我们遵循对第5.3节中讨论的合成图进行的灵敏度分析的结果，并将NeGMA as0g=-0.01的模块化阈值和αGMA的记忆项设置为α=0.8。

6.2.1选民欺诈数据集的结果。我们在图10a和图10b中报告了模块化和稳定性。独立GMA被设计用来生成模块化的独立解决方案，实现了最高质量的社区（从快照1到6的中值≈0.8）。然而，这三种进化算法表现出了相当的模块化性能。

和我们之前的作品一致，我们设置了=3。



不同算法的Fig.10.Performance与网络欺诈数据集。

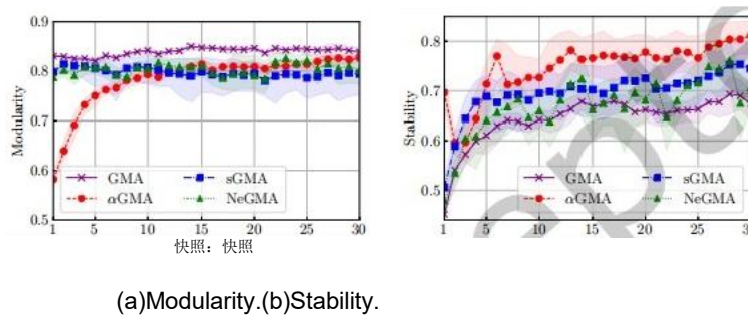


Fig.11.Performance的不同算法与网络环境处理数据集。

正如预期的那样，GMA实现的良好模块化性能是以时间相干性为代价的。它显示出显著的不稳定性（所有快照的 <0.4 ，如图10b所示）。与独立GMA相比，sGMA略微提高了稳定性，导致上次快照中 \approx 最大增加20%。同时，NeGMA在合成图上的参数调整允许它找到独立GMA和sGMA之间的匹配。同样，aGMA在稳定性方面表现出色，展示了每个快照的改进，从快照1的 $\approx 50\%$ 到上次快照的 $\approx 10\%$ ，优于其他测试算法。

我们注意到，在多个快照中，只有28%的节点仍然保持活跃状态，这表明了零星的用户活动，并可能表明社区动态受到类似间歇性转换的高度影响。事实上，在VoterFraud数据集上获得的aGMA的良好结果与第5节中讨论的和表2中总结的结果一致，这表明aGMA在这种转换中表现最好。然而，正如第5.4节中所讨论的，aGMA稳定性的实质性提高极大地影响了大图中的可伸缩性，每个快照的平均执行时间为892秒，而其他算法仅为72秒。

我们强调，CoDEN能够更快地控制社区检测算法的比较。它避免了大型真实图带来的可伸缩性挑战，简化了针对特定需求的最佳算法的选择，并有助于确定最终应用的最优参数。

6.2.2 网络望远镜数据集的结果。我们在图11a和图11b中报告了每个快照的模块化和稳定性。我们记得，通过构造，这是一个更动态的图，特别是在第一个快照（ $t \in [1, 10]$ ），相比于语音欺诈和合成图。

根据这一观察，而内存引入aGMA导致整体更高的稳定性比其他算法（图11b），它hur ts识别图的变化，影响模块化——即需要10快照成为与sGMA和NeGMA（图11）。另一方面，独立GMA考虑了第5节的考虑，即它达到了较差的稳定性，同时导致最高的modularity.sGMA，反过来，在独立GMA和αGMAboth的模块化（always>0.75）和stability（ ≥ 0.70 for ≥ 10 ）之间达到了平衡。

有趣的是，与类间歇性变换似乎占主导地位的语音欺诈图相比，在网络望远镜中，图的进化更快、更复杂，导致了一个高度动态的节点邻域。这影响了NeGMA的初始化，导致独立GMA和sGMA的稳定性之间的广泛振荡。

总而言之，网络望远镜数据集提供了一个特别的案例研究图动态的复杂性有很大的影响，晦照明的限制引入内存项图建设的社区质量（aGMA）和努力评估一个快速变化的社区的稳定（NeGMA）。

要点：(i)在没有一致地面真实的情况下评估真实数据集是具有挑战性的，对于稳定性和社区质量没有明确的算法赢家；(i)在间歇变换占优势但在识别突然变化方面响应性较低的情况下，真实数据与合成graphs，i.e.aGMA的结果一致是最稳定的算法。反过来，sGMA提高了独立GMA的稳定性，但检测次优社区，而sGMA平衡了两者；(ii) CoDEN解决了现实动态图的可伸缩性挑战，能够更快地比较替代算法，并指导参数选择。

7结论

在本文中，我们解决了在动态网络中评估进化CD算法的挑战。我们提出了一个基准测试框架（CoDEN），依赖于组图转换反映现实场景。我们提供了缺乏一套全面的工具来评估测试社区的质量，提出了三个新的度量标准（有监督和无监督）来补充广泛使用的模块化。

我们采用了所提出的框架来广泛地测试和比较基于模块化的进化CD算法的性能。我们测试了GMA、莱顿和它们的进化变体。我们还提出了NeGMA（NeLei），一种基于GMA（Leiden）和节点边界的广义模块化的进化CD方法。此外，我们还对测试解决方案中所涉及的参数的影响进行了调查。最后，我们在与Web相关的真实数据集上测试了这些算法。

简而言之，我们的实验结果表明(i)所有算法对干扰和噪声场景表现出鲁棒性；(ii) αGMA在检测间歇转换方面表现出较高的稳定性和正确性，但在检测瞬时转换方面引入了高延迟和较强的收敛时间；(ii) sGMA具有尺度性，随着时间的推移保持高稳定性（特别是在影响网络社区结构的渐进转换中），但在检测突发社区方面表现出低响应性和困难；(iv) NeGMA是在正确性和稳定性方面最先进的算法。此外，当发生突变时，它的性能优于现有进化解决方案的响应性和处理率。(v)以简单转换为特征的真实数据集上的性能反映了合成图上的性能，而(iv)更复杂的tr转换突出了进化CD算法的局限性。

这项工作的未来发展包括将CoDEN扩展到依赖于单个转换组合的更复杂的场景，以及将CoDEN应用于其他进化CD算法家族，而不选择实现模块化。最后，我们相信与CoDEN共享

20-G. Paoletti等人。

研究界将推广其在其他数据集上的应用，可能确定策略，以给定特征场景选择最佳方法。

确认信息

这项工作由“惠普C、大数据和量子计算国家中心”项目资助，CNOO00013（Nazionali-D.D.n.3138 M42 Nazionali-D.D.n.3138-1.4-16.12.2021的Avviso中心，由17.06.2022的1031法令资助），由CNPq资助和巴西资助机构（米纳斯吉拉斯州移民基金会）的资助机构。

参考文献

- [1]Anton Abilov and 一清Hua.2021.Voter欺诈2020. (2021) .doi: 10.6084/m9.figshare.13571084.v2
- [2]Anton Abilov, 一清华一清, 哈娜·马塔托夫, 奥弗拉·阿米尔和Mor Naaman.2021.VoterFraud2020: a多模式的选举方式在推特上声称。网络和社交媒体的AAAI会议论文集 (2021) .doi: 10.1609/icwsmv15i1.1813
- [3]Huma Aftab, 朱纳德Shuja, 瓦利德阿拉斯玛丽, 和Eisa Alanazi.2021.Hybrid DBSCAN基于社区检测前边缘缓存在社交媒体应用程序。2021年, 国际无线通信和移动计算 (WCMC)。
- [4]Mohsan阿里, 迈赫迪哈桑, 卡希夫基法亚特, 金扬金, 萨奇布哈卡克, 和穆罕默德库拉姆汗。2023.使用深度学习和图形分析的社交媒体内容分类和社区检测。技术逻辑预测和社会变化 (2023年)。
- [5]詹巴蒂斯塔·阿玛蒂、西蒙尼·安吉利尼、安东尼奥·克鲁西亚尼、吉安马尔科·福斯科、吉安卡洛·高迪诺、丹尼尔·帕斯基·尼和保拉·沃卡。2021.通过社区检测算法进行主题建模。在2021年在线社交网络开放挑战的过程中。
- [6]阿米拉、阿米拉、比利卡巴布和奥马尔Nouali.2023.A使用社区检测算法进行马氏分析。压缩。调查 (2023年)
- [7]Thomas Aynaud和Jean-Loup Guillaume.2010.Static社区检测算法。8th International 移动、特别和有线网络的建模和优化研讨会
- 随机网络缩放的[8]-巴拉巴西和Reka Albert.1999.Emergence。科学(1999.doi: 10.1126/science.286.5439.509
- [9]亚历克西斯·鲍丁, 最大西米利安, 谢尔吉·基尔吉佐夫, 克莱门斯·马格宁, 和马尔万Ghanem.2022.Clique渗透方法: 记忆效率几乎精确的社区。参加高级数据挖掘和应用国际会议。
- [10]亚历克西斯·鲍丁, 莱昂内尔·塔布里尔, 克莱门斯Magnien.2023.LSCPM: Communities在大规模现实世界链接流由集团渗透方法。第30届时间与推理国际研讨会, 时间2023年, 2023年9月25-26日
- NCSR德莫克里托斯, 希腊雅典 (LIPiS, 第278卷), 亚历山大·阿蒂基斯, 弗洛里安·布鲁斯和卢克·亨斯伯格 (Eds,)。大格图尔莱布尼茨-中心毛皮Informatik, 3: 1-3: 18.doi: 10.4230/LIPiCS.TIME.2023.3
- [11]法比安·鲍曼, 菲利普·洛伦茨·斯普伦, 伊戈尔·Stanini.2020.Mod·索科洛夫, 以及米歇尔·Stanini.2020.Mod在社交网络中的回音室和极化动力学。物理评论信 (2020) .doi: 10.1103/PhysRevLett.124.048301
- [12]Punam Bedi和Chhavi Sharma.2016.Community检测。Wiley跨学科综述: 数据挖掘和知识发现 (2016)。
- [13]罗恩·贝克曼, 什洛莫·齐伯斯坦和詹姆斯 Allan.2007.Web页面使用启发式搜索网络图。在诉讼中
- 在第20届国际人工情报联席会议上。
- [14]文森特·布朗德尔, 让-卢普·纪尧姆, 雷诺·兰比奥特和艾蒂安·Lefebvre.2008.Fast在大社区
- 网络统计力学杂志: 理论与实验》(2008)。doi:10.1088/1742-5468/2008/10/P10008
- [15]雷米卡扎贝特, 苏阿德布德布扎, 和Giulio Rossetti.2020.Evaluating社区检测算法的逐步进化图。复杂网络的联合网络 (2020) .doi: 10.1093/comnet/cnaa027
- [16], 查塔帕海亚, 坦梅巴苏, 克达斯, 昆塔尔高什和晚期CA Murthy.2021.Towards在复杂网络中的发现和电子商务的影响。电子商务研究中心 (2021年)。
- 节点归属社交网络中的[17]Petr Chunaev.2020.Community检测: 一项调查。《计算机科学评论》(2020) .doi: 10.1016/j.cosrev.2020.100286
- [18]Matteo·奇内利, 方济各·莫拉莱斯, 亚历山德罗·加莱亚齐, 沃尔特·夸特罗乔奇和米歇尔·Starnini.2021.The在社交媒体上的回音室效应。《美国国家科学院院刊》118,9 (2021), e2023301118。
- [19], [19], 和Itchba.2019.Tracking社交网络社区演变: 一项调查。信息处理与管理 (2019) .doi: 10.1016/j.ipm.2018.03.005
- [20]洛伦佐·达尔·阿米科, 罗曼·库利特和尼古拉斯·Tremblay.2020.Optimal拉普拉斯正则化稀疏光谱社区检测。在IEEE国际声学、演讲和信号会议上
- [21]Nur Nasuha Daud, 哈菲扎, 哈米德, 蒙塔赫萨杜恩, 萨赫兰, 和非巴德鲁尔Anuar.2020.Applications的社交网络链接预测: 回顾。网络和计算机应用的杂志 (2020) .doi: 10.1016/j.jnca.2020.102716

ACM跨。蜘蛛网

- [22]Chi Thang杜昂.阮清忠、阮清忠、尹洪志、魏德利克, 以及通过检测社区进行的Nguyen.2023.DeepMinCut: Learning节点嵌入。模式识别(2023)。
- [23]Heba Elgazar, 凯尔·斯普尔洛克, 和坦纳Bogart.2021.Evolutionary俱乐部环和社区检测算法的社会媒体健康监测算法。机器泄漏与应用程序(2021).doi: 10.1016/j.mlwa.2021.100084
- [24]大卫·埃普斯坦, 迈克尔·帕特森, F·弗朗西斯Yao.1997.On最近邻图。离散与计算几何学(1997年)。
- [25]Sara E Garza和Satu Elisa Schaeffer.2019.Community检测与标签传播算法: 一个调查。物理学A: 统计力学及其应用(2019年)。
- [26]卢卡·乔阿奇尼, 卢卡·瓦西奥, 马尔科·米利亚, 尔迪利奥·德拉戈, 齐德·本·豪迪和达里奥Rossi.2021.DarkVec: Automatic分析暗网流量与文字嵌入。在第十七届新兴网络实验和技术国际会议的记录上
- [27]卢卡·乔阿奇尼, 卢卡·瓦西奥, 马尔科·梅利亚, 尔迪利奥·德拉戈, 齐德·本·豪迪和达里奥Rossi.2023.i-DarkVec: 公司负责暗网交通分析。ACM互联网技术学报(2023年)。
- [28]米歇尔·Girvan和Mark EJ Newman.2002.Community的社会和生物网络结构。美国国家科学院的程序学研究中心(2002).doi: 10.1073/pnas.122653799
- [29]约翰·Pattanaik.2021.Artificial·古德尔, 萨蒂什·库马尔, 翁马克·林, 和德比杜塔Pattanaik.2021.Artificial在金融领域的智慧和机器学习: 从文献计量学分析中识别基础、主题和研究集群。行为和实验财务的判断32(2021),100577.
- [30]Palash Goyal, Nitin Kamra, 何欣然, 和Yan Liu.2018.Dyngem: Deep动态graphs.arXiv预印本嵌入方法arXiv:1805.11273(2018)
- 郭俊辉、王佳佳、Zhen Zhang.2014.Evolutionary社区结构发现动态加权网络。
物理学A: 统计力学及其应用(2014)。
- [32]Khaled M哈穆达和穆罕默德S Kamel.2004.Eficient基于短语的文档索引g用于web文档集群。埃尔
知识与数据工程交易(2004年)。
- [33]YiHan, 莎妮卡·卡鲁纳塞克拉, 和克里斯托弗·Leckie.2020.Graph神经网络, 从社交Media.arXiv: 2007.03 316中不断检测假新闻
- [34]卢卡斯·休莫斯、安娜克·沙尔、克里斯托弗·兰斯、阿纳斯塔西娅·德罗斯特、卢克·扎皮亚、幸运肯、丹尼尔·C·斯特罗布尔、胡安·赫纳奥、法比奥·拉·库恩等al.2023.Best实践在莫达利特进行单细胞分析。《自然评论遗传学》24,8(2023), 550-572
- 其中[35]黄思翰、翁浩雷和Yang Feng.2023.Spectral通过自适应层聚合进行聚类。杂志的
计算和图形统计学(2023年)
- [36]马修·雅各米, 托马索·文丘里尼, 塞巴斯蒂安·海曼, 和马修·Bastian.2014.ForceAtlas2, 一种为Gephi软件设计的方便的网络可视化的连续图布局算法。美国公共科学图书馆一(2014).doi: 10.1371/journal.pone.0098679
- [37]Nikhil Jha, 马蒂诺·特雷维索, 卢卡·瓦西奥, 和Marco Mellia.2022.The互联网与隐私政策: 测量ng网络一致同意。ACM交易(2022).doi: 10.1145/3555352
- [38]盛明·罗嘉顺·金, 郑特蕾西·柯和敏哲Wang.2023.OptimalEstimation的网络社区的数量。天Amer: 统计者。使发生联系(2023).doi:10.1080/01621459.2022.203 5736
- [39]Di Jin、余治芝、焦鹏飞、潘石瑞、何东晓、贾昊、余Philip、伟雄Zhang.2021.A社区检测调查方法: 从统计建模到深度学习。IEEE知识与数据工程学报(2021年)。
- [40], 印度教授, 牧师, 吴定浩, 甄子强。2022.检测和解释大型网络望远镜中扫描行为的变化。IEEE信息取证与安全交易(2022年)。
- [41]EI穆阿特兹比利拉卡巴布, 莫拉德德巴比, 阿卜德鲁希德德哈布和德杰吉加Mouheb.2020.Android恶意软件集群使用社区检测机器人包相似性network.arXiv预印本arXiv: 2005.06075(2020)。
- [42]EIMouatez比利拉卡巴布, 哀悼者, 德哈布, Mouheb.2020.Scalable和抢劫使用基于社区的网络分区。计算机与安全(2020)。
- [43]Moaiad Ahmad Khder.2021.Web刮或网络爬行: 艺术, 技术, 方法和应用阶段
软计算及其应用的进展杂志(2021).doi: 10.15849/ijasca.211128.11
- [44]Santosh库马拉什和Ravi Kumar.2021.A研究网络挖掘的不同方面和研究问题。IOP会议系列: 材料、科学和Engineering.doi: 10.1088/1757-899X/1022/1/012018
- [45]安德里亚兰开奇尼蒂, 圣福尔托纳托, 和菲利波Radicchi.2008.Benchmark图的测试和社区检测算法。物理回顾版(2008年)。doi:10.1103/PhysRevE.78.046110
- [46]李春莹、唐勇、唐志康、曹锦里、彦春Zhang.2022.Motifbased嵌入实验室电子传播算法进行社区检测。国际智能系统杂志(2022年)。
- [47]赵李, 潘辉、张鹏、黄家明、王彪、凌田、吉张、高建力、星唐2021。在幕后发生了什么事? 从在线到在线电子商务中的欺诈社区检测。在网络
的伴随程序中
会议2021。

Paoletti等人。

林[48]玉茹、云志、朱圣福、比利时L Tseng.2008.Facetnet: a框架分析动态网络中的社区及其演变。在第17届万维网国际会议的论文集上。685-694

[49]Fuchen刘, 崔, 吕解, 凯瑟琳Roeder.2018.Global动态网络光谱聚类。美国国家科学院的研究程序115,5 (2018), 927-932。

刘[50]范珍、赵李、王宝坤、贾武、吉安、杨家明、张一清、王伟强、山雪、尼泊尔等al2022.eRiskCom: 一个电子商务风险社区检测平台。VLDB 杂志 (2022)。

[51]刘范珍, 雪, 贾吴, 周川, 胡文斌, 塞西尔巴黎, 尼泊尔吉安杨, 菲利普Yu.2021.Deep学习

社区检测: 进步、挑战和机遇。第二十九届国际国际会议的进展

关于人工智力的联合会议。4981-4987.

刘[52]京信、徐张欣、张殷、吴伟强、基于尤Song.2020.K-core的动态模板图卷积网络

图表IEEE 交易与知识与数据工程34,8 (2020), 3841-3853。

[53]Yi-Ju Lu and Cheng-Te Li.2020.GCAN: Graph-aware Co-Attention网络, 为社交媒体上的假新闻检测。在

计算语言学协会第58届年度会议的程序, 丹-尤拉夫斯基, 乔伊斯柴, 娜塔莉施卢特, 和

乔尔特雷奥 (版)。计算语言学协会, Online, 505-514.doi: 10.18653/v1/2020.acl-main.48

[54]Mark EJ Newman.2006.Modularity和网络中的社区结构。美国国家科学院 (2006).doi: 10.1073/pnas.0601602103

[55]Andrew Ng, Weiss.2001.On光谱聚类: 分析和算法。神经信息处理系统 (2001).https://dl.acm.org/doi/10.5555/2980539.2980649的研究进展

[56]Ahmed/Obaide, 塔努斯里·查特吉, 和阿比谢克·Bhattacharya.2021.Semantic网络和网页聚类算法: 一个景观

看法EAI认可的能源网络交易 (2021年)。

[57]活力巴拉拉, 艾伯特·拉兹洛巴拉巴西, 和Tamas Vicsek.2007.Quantifying社会群体进化。《自然》杂志446,7136 (2007), 664-667。

[58], 德雷尼, 勒斯特法卡斯和塔马斯Vicsek.2005.Uncovering在自然和社会自然复杂网络的过度社区结构 (2005)。

[59]佐丹诺·保莱蒂, 洛伦佐·达拉米科, 卡里梅里, 雅格布·伦蒂, 叶莲娜·梅约瓦, 丹尼尔·保洛蒂, 米歇尔·斯塔尼尼和米歇尔·Tizzani.2024.Political 在推特上关于欧洲疫苗辩论的背景。科学报告ts14,1 (2024), 4397。

[60]佐丹诺保莱蒂, 卢卡乔阿奇尼, 马可米利亚, 卢卡瓦西奥, 和尤萨拉M.Almeida.2023.g进化群落的基准

动态Networks.https://doi.org/10.48550/arXiv.2312.13784中的检测算法在第四届研讨会上提出

更复杂的结构, 为挖掘和推理 (GCLR) 在AAAI 2024。

[61]Filipo电台, 克劳迪奥·卡斯特拉诺, 费德里科·塞科尼, 维托里奥·洛雷托, 和多梅尼科·帕里西.2004.Defining, 并在网络中识别社区。《国家科学院院报》101,9 (2004), 2658-2663。

[62]Usha·南迪尼·拉加万、里卡·阿尔伯特和声音Kumara.2007.Near线性时间算法来检测大规模网络中的社区结构。物理版本ewE (2007)

[63]Ebin Deni Raj, 古纳斯卡兰马诺加兰, 高塔姆斯里瓦斯塔瓦, 和基于Wu.2020.Information颗粒的社交网络社区检测。IEEE的计算社会系统交易 (2020).doi: 10.1109/TCSS.2019.2963247

[64]菲利普·里希特和Arthur Berger.2019.Scanningthe扫描仪: 从一个大规模分布式网络望远镜扫描互联网

在联合测量会议的会议记录中。doi:10.1145/3355369.3355595

[65]P在数据挖掘和知识发现中的Paulheim.2016.Semantic网络: 一个全面的调查。fweb语义学期刊 (2016).doi: 10.1016/j.websem.2016.01.001

[66]Giulio·罗塞蒂和Remy Cazabet.2018.Community在动态网络中的发现: 一项调查。汇编调查 (2018).doi: 10.1145/3172867

[67]复杂网络随机漫步的MartinRosvall和Carl T Bergstrom.2008.Maps揭示了社区结构科学学术程序105,4 (2008), 1118-1123。

[68]雷扎萨克迪尼亚, 奥米德·法特米, 丹尼尔·洛伦齐, 法扎德·塔什特利克, 和克里斯蒂安Timmerer.2023.Community-Based QoE增强用户生成的内容直播。2023年第13届全国计算机与知识工程会议 (ICCKE)

[69]诺, 萨芬林萨塔, 艾丁·布鲁克, 哈勒兹。易卜拉欣和Shaikh Arifuzaman.2023.Exploring时间社区进化: 动态社区检测的算法方法和并行优化。应用网络科学 (2023).doi: 10.1007/s41109-023-00592-1

[70]亚当·申克, 马克·最后, 霍斯特·邦克, 和亚伯拉罕·Kandel.2003.Clustering的网络文档使用地图模型。在网络中

文件分析: 挑战和机遇。

[71], 弗朗西斯卡·索罗, 毛罗·阿列格雷塔, 马可·梅利亚, 杰迪利奥·德拉戈, 和莱安德罗·MBertholdo.2020.Sensing, 噪音: 在暗网交通中发现社区。在2020年地中海通信和计算机网络会议 (MedComNet)。

[72]Myra, 尤尼斯, 和Rene Schult.2013.MONC, 以及对建模和监测的跟踪

集群转换。在机器泄漏和知识发现中的数据库: 欧洲Conference.doi: 10.1007/978-3-642-40994-3_41

[73]文森特陷阱, 卢多沃特曼和Nees Jan Van Eck.2019.From卢汶莱顿: 保证连接社区。科学报告 (2019).doi: 10.1038/s41598-019-41695-z

ACM跨。蜘蛛网

- [74]Hai Van Pham和Dong阮Tien.2021.Hybridlouvain-clustering模型利用知识图改进用户在社交网络上的聚类行为。在智能系统和网络中：从ICISN中获得的精选文章
- [75]阮XuyenVinh, JulienEpps, 和James Bailey.2009.Info信息理论措施聚类比较：计算机会是必要的吗？。在第26届机器Learning.doi: 10.1145/1553374.1553511临时会议上
- 基于网络结构扰动和拓扑的王[76]、高林、小Ma.2017.Dynamic社区检测
相似度《统计力学期刊：理论与实验》，2017,1（2017），013401。
- 王一书、叶元、马玉良、歌仁Wang.2019.Time-dependent图：定义、应用和算法。数据科学和工程学（2019）.doi: 10.1007/s41019-019-00105-0
- 吴[78]，岐山张志华，郭坤，德钦Wang.2020.Deep社交网络社区检测技术。IEEE访问（2020）.doi: 10.1109/ACCESS.2020.2996001
- [79]Jaewon Yang和Jure Leskovec.2012.Defining和评估基于地面真相的网络社区。在ACM的P中
SIGKDD工作组正在挖掘数据语义。1-8.
- [80]阿依江、依西玛伊、宋维良、王静、方贞吉安、余园陵、陈晓苏、杨立旭、杨西杰、小牛、小天河、al.2024.Repeated欧米龙暴露于祖先SARS-CoV-2免疫印迹上。自然的625,7993（2024）.148-156。
- [81]家家璇，杜天宇，JureLeskove.2022.ROLAND：graph动态图框架。在28号的程序中
ACM SIGKDD知识发现和数据挖掘会议。2358-2366.

Paoletti等人。

补充材料

算法1将内存引入图（ α GMA and α Lei）	
要求：加权动态图，记忆项 $\mathcal{G}^t = (\mathcal{V}^t, \mathcal{E}^t)_{t=1}^N$	
确保：每个快照t的社区集：从snapshot- $\{C^t\}_{t=1}^N$ 到N do	
2: 为edge $e = (u, v, w) \in$ 等人做	3: if $w_t = 0$, 那么D, 如果边缘e在前一个快照中不存在
	4: 保留边缘重量
5: 其他的	6: $w^4 = (1 - \alpha) \cdot w_t + \alpha \cdot w_{t-1}$ 嵌入内存的边缘权重
	7: 更新 $= (u, v, w^2) \in$ 和更新的重量
	8: Ct独立算法（Vt, 8t, 0）►在具有未初始化社区的新图上运行CD算法

算法2将算法引入内存（sGMA和sLei）	
要求：加权动态图	$\{\mathcal{G}^t = (\mathcal{V}^t, \mathcal{E}^t)\}_{t=1}^N$
确保：每个管理程序节点的社区集	$\{C^t\}_{t=1}^N$
1: 从t-1到N do	
2: 然后ift=1	
3:Ct=0	►将初始化设置为一个空集
4: elsc-	
5:Ct-INDEPENDENTALG（Vt, 8t, C）	D将GMA初始化设置为之前定义的社区D使用自定义初始化运行CD算法

为了便于理解和增强the研究的再现性，本补充部分提供了三个进化扩展的伪编码表示，以将内存集成到GMA和Leiden算法中。

- 算法1-边缘权重中的内存（ α GMA and α Lei）：该算法集成了一个内存参数，a，根据它们在快照之间的持久性来调制边缘权值。在每个快照t中，算法保留了前一步中缺失的边的当前权值，同时为现有边分配当前和历史权值的加权平均值。通过将调整后的图输入GMA或Leiden中，社区建筑的时间一致性随着时间的推移而保持。
- 算法2-初始化中的内存（sGMA和sLei）：我们修改了算法的初始化，使用社区从之前的快照中初始化节点，从而向前携带社区标签。新出现的节点被分配了唯一的社区，而来自前一个快照的活动节点则保持其社区成员身份，允许每个算法根据模块化优化的需要进行调整。
- 算法3-基于邻域的初始化与局部模型的价值评估：第三种方法这里使用的节点涉及到根据历史数据和邻域上下文的节点初始化。持久节点被分配给它们的先例社区，而新节点采用来自邻居的社区成员。随后，我们进行了模块化评估，以确定这些新形成的社区增加了它们的局部模块化。如果增强值低于阈值

算法3对所提出的基于邻域的方法的概述

- 1: 功能EVALUATEModULARITY (C, ct-1)
- 2: 社区 $c \in C_{t-1}$ 做

为每个社区所提供的初始化 d在局部模块化退化时得到局部模块化增益
- 3: $\Delta Q_c^t \leftarrow Q_c^t - Q_c^{t-1}$
- 4: 如果 $\Delta Q_c^t \leq 0$, 那么

5: 解除社区的绑定

6: 将c的每个节点分配给一个不同的社区

7: 返回最终初始化n
- 要求: 加权动态图 $\{\mathcal{G}^t = (\mathcal{V}^t, \mathcal{E}^t)\}_{t=1}^N$, 模块化阈值 θ
- 确保: 每个快照中的社区集 $\{C^t\}_{t=1}^N$
- 8: 从1到N

9: 将初始设置设置为空集
- 10: 然后 if t>1
- 11: 为 $v \in V_t$ 做

12: if $v \in V_{t-1}$, 那么如果节点v是活动的, 则是上一个快照

13: $C_{v \in V_{t-1}} \leftarrow C_{v \in V_{t-1}} \cup \{v\}$
- 14: 其他的
- 15: 提取活动节点邻居 $N_t(v)$
- 16: 获取 $C_{v \in N_t(v)}$ 中节点的先前社区

17: $C_t \leftarrow C_t \cup \{v\}$

18: $C_t \leftarrow \text{EVALUATEModULARITY}(C_t, ct-1)$ 评估局部模块化
- 19: ct独立标记 (V_t, δ_t, C) 使用自定义初始化运行独立的GMA (或Leiden)

established in θ g, 在最终应用社区检测算法之前, 节点被重新分配到不同的社区。

Paoletti等人。

2024年1月29日收到；2024年11月26日修订；2025年2月6日接受

ACM跨。蜘蛛网