arXiv:2301.00310v2 [cs.SI] 3 Jan 2023

随时间变化的图小片段：时序网络分析的新视角

Deukryeol Yoon 韩国科学技术院 AI 首尔，韩国
deukryeol.yoon@kaist.ac.kr

Dongjin Lee 韩国科学技术院 EE 大田，韩国
dongjin.lee@kaist.ac.kr

Minyoung Choe 韩国科学技术院 AI 首尔，韩国
minyoung.choe@kaist.ac.kr

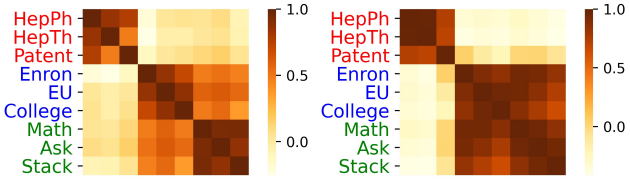
Kijung Shin
韩国科学技术院 AI 与 EE
韩国首尔
kijungs@kaist.ac.kr

摘要

图被广泛用于对各种类型的交互进行建模，例如电子邮件通信和在线讨论。许多这类真实世界图都是时序图，具体而言，它们会随着时间推移不断增长，加入新的节点和边。

统计每种图元（即诱导子图的一个同构类别）的实例，已成功用于刻画图的局部结构，并具有广泛的应用。尽管图元已扩展至时序图，但这些扩展旨在考察由到达时间相近的边构成的时间局部子图，而非局部结构的长期变化。

在本文中，作为时序图分析的一种全新视角，我们研究真实世界图中图元实例分布随时间的演化，并从三个不同层级（图、节点和边）进行考察。在图层级，我们首先发现，其演化模式与随机图中的模式显著不同。随后，我们提出一种图元转移图，用于衡量图的演化模式之间的相似性，并发现来自同一领域的图之间存在出人意料的相似性。在节点和边层级，我们证明，节点和边早期阶段周围的局部结构能够强烈预示其未来重要性。具体而言，利用节点和边在图元中所扮演角色（亦称轨道）的计数，我们显著提高了对节点和边未来重要性的可预测性。



(a) 图之间的相似性（关于**图元转移**（分类准确率 = **97.2%**）

(b)图之间的相似性（关于**图元出现**（分类准确率 = **83.3%**）

图1：来自同一领域的现实世界时间图具有相似的演化模式，这些模式可通过图元之间的转移来捕捉。图中展示了3个领域（以文字颜色区分）的9个图之间的成对相似性，分别针对图元之间的转移（见(a)）和图元的出现（见(b)）。与(b)相比，在(a)中可以更准确地对图所属的领域进行分类。具体而言，在最佳相似性阈值下，(a)中的分类准确率为97.2%，而(b)中为83.3%。有关相似性度量的详细信息，请参见第 3.2 节。

1 引言

图是一种简单而强大的工具，因此被用于表示各种类型的交互，例如电子邮件通信、在线问答、科研合作等。由于新的交互不断形成，这类真实世界图具有时序性，也就是说，随着新节点和边的出现，它们会随时间演化。许多研究考察了现实世界时间图的动态变化，并揭示了有趣的模式，包括致密化 [26], 直径收缩 [26], 以及三角形形成中的时序局部性 [23]。

图元已广泛用于分析图的局部结构。图元 [36] 定义为由预先指定数量的节点组成的同构小子图集合。具体而言，不同图元实例的相对计数能够有效刻画图的局部结构，并已成功应用于……允许免费制作本作品全部或部分内容的数字副本或纸质副本，供个人或课堂使用，但不得为营利或商业利益而制作或分发副本，且副本必须保留本声明以及首页上的完整引用。必须尊重本作品中由ACM 以外的其他方拥有版权的组件。注明出处后可进行摘要。除此之外，若要复制、再版、发布到服务器或分发至列表，须事先取得具体许可和/或支付费用。请通过 permissions@acm.org 申请许可。*Conference'17*，*2017*年7月，美国华盛顿特区 © 2023 Association for Computing Machinery。ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

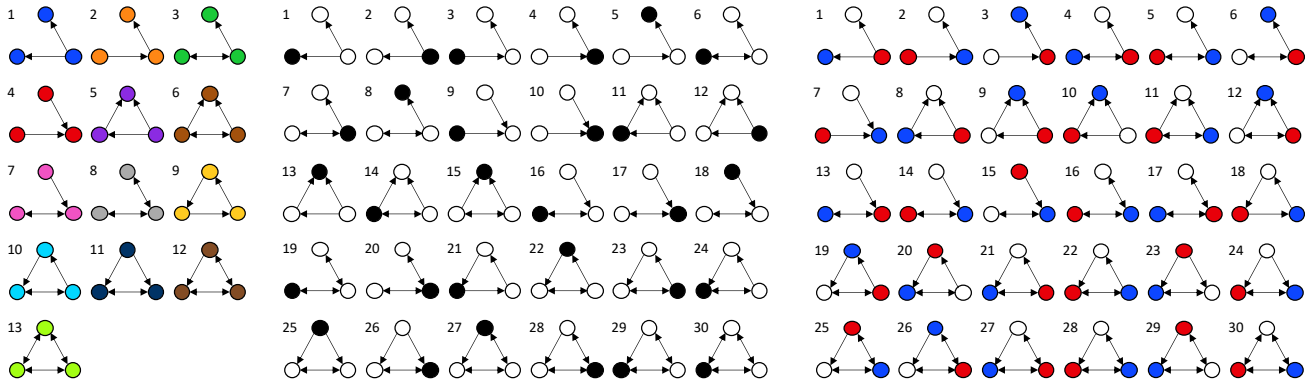
应用包括图分类 [31, 32], 社区检测 [5, 10, 39], 异常检测 [20], 和节点嵌入 [24, 28, 42]。

由于时序图无处不在，图元的概念已通过多种方式推广，用于时序图分析。时序网络模体 [21, 35] 是由时序子图组成的集合，这些时序子图：(a)不仅在拓扑结构上相同，在时间上也相同；(b)由固定数量的节点组成；(c)具有时序局部性，即由到达时间足够接近的边组成（详细信息请参见第6节）。由于最后一个条件，它们适用于分析图的短期变化，但不适用于分析局部结构的长期变化，而后者正是本文的研究重点。

本文考察了来自三个不同领域的九个现实世界时间图中，由图元捕捉的局部结构的长期演化，并将其作为分析时序图的一种新视角。我们的分析涵盖三个层级：图、节点和边。

在图层级，我们首先考察图元实例分布随时间的变化。我们发现，其演化模式不同于通过随机打乱边得到的随机化图中的演化模式。此外，来自同一领域的图具有一些共同的演化特征。为了系统地比较演化模式，我们引入了图元转移图，用于编码图中变化所导致的图元之间的转移。如图1(a)所示，来自同一领域的图具有相似的图元转移模式，这有助于实现准确的图分类，尽管这些图的规模各不相同。

在节点和边层级，我们研究每个节点和边在早期阶段周围的局部结构如何预示其未来重要性。具体而言，作为局部结构，我们考虑节点角色



(a) 13 个图元　(b) 30 种节点角色（也称为节点轨道）　(c) 30 种边角色（也称为边轨道）

图2：(a) 由三个节点组成的13个图元 [36]。(b) 图元中的30种节点角色 [36]（参见黑色节点的位置）。(c) 图元中的30种边角色（参见从红色节点指向蓝色节点的边的位置）。

（形式上称为节点自同构轨道 [36]）和边角色 [17]，它们大致是图元内节点和边的对称位置集合。我们还证明，对于预测未来中心性（具体而言，包括入度、中介中心性 [15], 接近度 [9], 和 PageRank [34]）这一任务，每个节点和边在早期阶段所承担的角色计数，比此前使用的特征 [41], 更具信息量；同时，它们还与简单的全局特征（例如节点和边的总计数）互为补充。

我们的贡献总结如下：

- **模式：** 我们对图元的时间演化做出了一些有趣的观察：同一领域的图之间存在令人惊讶的相似性，以及关于节点和边未来重要性的局部结构信号。
- **工具：** 我们提出图元转移图，这是一种有效的工具，可用于衡量不同规模时序图中局部动态的相似性。
- **预测：** 我们引入基于角色的局部特征，以提升对节点和边未来重要性的预测准确率；这些特征与全局特征互为补充。

可复现性： 代码和数据集可在 https://github.com/deukryeol-yoon/graphlets-over-time 获取。

在第2节中，我们介绍基本概念、符号和数据集。在第3节中，我们介绍图级分析。在第 4 节和第5节中，我们介绍节点级和边级分析。在第6节中，我们简要综述相关工作。在第7节中，我们总结本文。

2 基本概念、符号与数据

本节首先介绍一些基本概念和符号。然后，我们介绍本文使用的九个数据集。

2.1 基本概念与符号

时序图： 一个时序图 $G=(V, E, T)$ 由节点集合 V、有向边集合 $E:=\{e_1, \cdots e_{|\mathcal{E}|}\}$，以及边到达时间的多重集合 $T:=[t_1, \cdots t_{|\mathcal{E}|}]$ 组成。对于每条有向边 $e_i \in E$，我们用 $t_i \in T$ 表示 e_i 的到达时间。我们用 $u \to v$ 表示从节点 u 到节点 v 的有向边；如果 $u \to v$ 或 $v \to u$ 存在，则节点 u 和 v 是相邻的。从现在起，在不产生歧义的情况下，我们将使用边 一词来表示有向边。

表1：符号表。

符号	定义
$G=(V, E, T)$ $G(t)=(V(t), E(t))$	节点为 V、边为 E、时间为 T 的时序图 G 在时间 t 的快照
$\tilde{G}=(V, E, \tilde{T})$ $\tilde{G}(t)=(\tilde{V}(t), \tilde{E}(t))$	从 G 随机生成的时序图 快照 $\tilde{G}$ 在时间 t
$m_i^{(t)}(v)$	节点角色计数 i 在节点处 v 于 $G(t)$

随机图： 随机图 $\tilde{G}=(V, E, \tilde{T})$ 是通过以均匀随机的一一对应方式，将 T 中的到达时间分配给 $G=(V, E, T)$ 中的边而得到的。对于每条边 $e_i \in E$，我们用 $\tilde{t}_i \in \tilde{T}$ 表示分配给它的到达时间。

快照：我们将 $G=(V, E, T)$ 在时间 t 的快照定义为 $G(t)=(V(t), E(t))$，其中 $E(t):=\{e_i \in E: t_i \le t\}$，且 $V(t) \subseteq V$ 是 $E(t)$ 中任意一条边的端点。也就是说，$G(t)$ 由在时间 t 或更早到达的节点和边组成。类似地，我们将 $\tilde{G}=(V, E, \tilde{T})$ 在时间 t 的快照定义为 $\tilde{G}(t)=(\tilde{V}(t), \tilde{E}(t))$，其中 $\tilde{E}(t):=\{e_i \in E: \tilde{t}_i \le t\}$，且 $\tilde{V}(t)$ 是 $\tilde{E}(t)$ 中任意一条边的端点。我们将快照 $G(t)$ 中节点 $v \in V(t)$ 的邻居定义为 $G(t)$ 中与 v 相邻的节点。我们将快照 $G(t)$ 中节点 $v \in V(t)$ 的度定义为端点包含 v 的有向边数量，并记作 $d(t)(v)$。我们用 $d(v)$ 表示节点 v 在最后一个快照 $G(t_{|\mathcal{E}|})$ 中的度。

诱导子图： 快照 $G(t)=(V(t), E(t))$ 是诱导 子图，当且仅当它由 $V(t)$ 的一个子集以及该子集中节点两两相连的所有边组成。若存在一个一一映射 f 将两个图中的节点相互对应，并且 H 中存在一条从节点 u 到节点 v 的边，当且仅当 H′ 中存在一条从节点 $f(u)$ 到节点 $f(v)$ 的边，则子图 H 和 H′ 同构。

图元： 一个图元是由彼此同构的诱导子图组成的集合。本文将研究范围限定在由三个连通节点组成的 13 种图元上。一个诱导

表2：本文使用的九个现实世界时间图概述。

领域	数据集	$\|V\|$	$\|E_T\|$	时间跨度
引用	HepPh	34，565	346，849	9年
	HepTh	18，477	136，190	10年
	专利	3，774，362	16,512,782	25年
电子邮件/消息	安然	55，655	209，203	24年
	EU	986	24,929	1.5年
	学院	1，899	20,296	0.5年
在线问答	提问ubuntu	159，316	262，106	6年
	数学溢出	24，818	90，489	7年
	堆栈溢出	2，601，977	16,266,395	8年

子图若与图2(a)中的第 k个图同构，则称为图元 k 的一个实例。

节点角色： 考虑一个节点集为 V′ 的诱导子图 H。H 的自同构是 H 到其自身的同构映射。也就是说，H 的自同构是 H 的节点之间的一一映射，并且在 H 中，节点 u 到节点 v 之间存在一条边，当且仅当在 H 中，对应于 u 的节点与对应于 v 的节点之间存在一条边。若用 Aut(H) 表示 H 的自同构集合，则节点 $u \in$ V′ 的自同构轨道是节点集 $\{y \in$ V′ : $\exists g \in Aut$(H) s.t. $y = g(u)\}$。形式上，节点角色是节点自同构轨道；粗略地说，它们是图元中节点对称位置的集合。图2(b)（见黑色节点的位置）展示了我们所考虑的 13 个图元中的全部 30 种节点角色。如果存在图元实例与图2(b)中的第 i 个图之间的同构映射，且该映射将 v 映射到图中的黑色节点，那么我们就说节点 v 在一个图元实例中“占据”节点角色 i。我们将节点 v 的节点角色计数 i 定义为满足 v 占据 i的图元实例数量，而 $m(t)\,i(v)$ 表示快照 G(t) 中的计数。

边角色： 考虑一个边集为 E′ 的诱导子图 H。基于上述定义的概念，我们将边

$u \to v$ $x \to y \in$ $\exists g \in Aut$ $x =$ 的边 $g\ u\ \wedge y = g\ v$

角色 定义为边集 { E′ : (H) s.t. ( ) ( )}。粗略地说，边角色是图元内部边的对称位置集合。图2(c)（参见从红色节点指向蓝色节点的边的位置）展示了所考虑的13个图元中的全部30种边角色。如果某个图元实例与图2(c)中的第 j 个图同构映射，并且该同构映射分别将 u 和 v 映射到图中的红色节点和蓝色节点，则称边 $u \to v$ 在该图元实例中“承担”边角色 j。我们将边 j 处的边角色计数 e 定义为 e 承担 j 的图元实例数量。

2.2 数据集

本文使用来自三个领域的九个现实世界时间图，这些图总结于表2中。

引文图： 每个节点代表一篇论文或一项专利。节点 u 到节点 v 的每条有向边都表示 u 引用了 v。

电子邮件/消息图： 每个节点代表一位用户。节点 u 到节点 v 的每条有向边都表示 u 向 v 发送了电子邮件（消息）。

在线问答图： 每个节点代表一位用户。从节点 u 指向节点 v 的每条有向边都表示 u 回答了 v提出的问题。

算法 1： 统计时序图中每种图元的实例的数量

输入： 时序图 G = (V，E，T)
输出： G 中每种图元实例的计数

```
将每种图元实例的计数初始化为零
将 E 初始化为空集
按到达顺序遍历每条边 e_i = u → v，执行
    N ← 与 u 的邻居和 v 的邻居的并集（u 和 v除外）
    对于每个 w ∈ N，执行
        如果 u、v 和 w 构成一个图元实例，则
            将该实例所对应图元的计数减一
            由 u、v 和 w构成
    将 u → v 添加到 E
    对于每个 w ∈ N，执行
        增加由该实例构成的图元的计数
        通过 u、v 和 w
返回每种图元实例的计数
```

3 图级分析

本节研究真实世界图中局部结构的演化。我们考察图元实例的分布以及图元之间转移的动态变化。

3.1 全局层级 1. 随时间变化的图小片段

我们跟踪所考察的真实世界图随时间演化时，各类图元实例的比率如何变化。我们的跟踪算法见算法 1，该算法改编自 StreaM[38], 。StreaM 在全动态图流中维护 4 节点无向图元的实例计数，其中边不仅会随时间添加，也会被删除。算法 1 的时间复杂度为 $\Theta(\Sigma_{v\in\mathcal{V}}(d(v))^2)$，证明见定理 1。值得注意的是，根据引理 1，时间复杂度为 Θ（最后一个快照中所有图元实例的数量），这是任何通过枚举图元实例来计数的算法所能达到的最优时间复杂度。

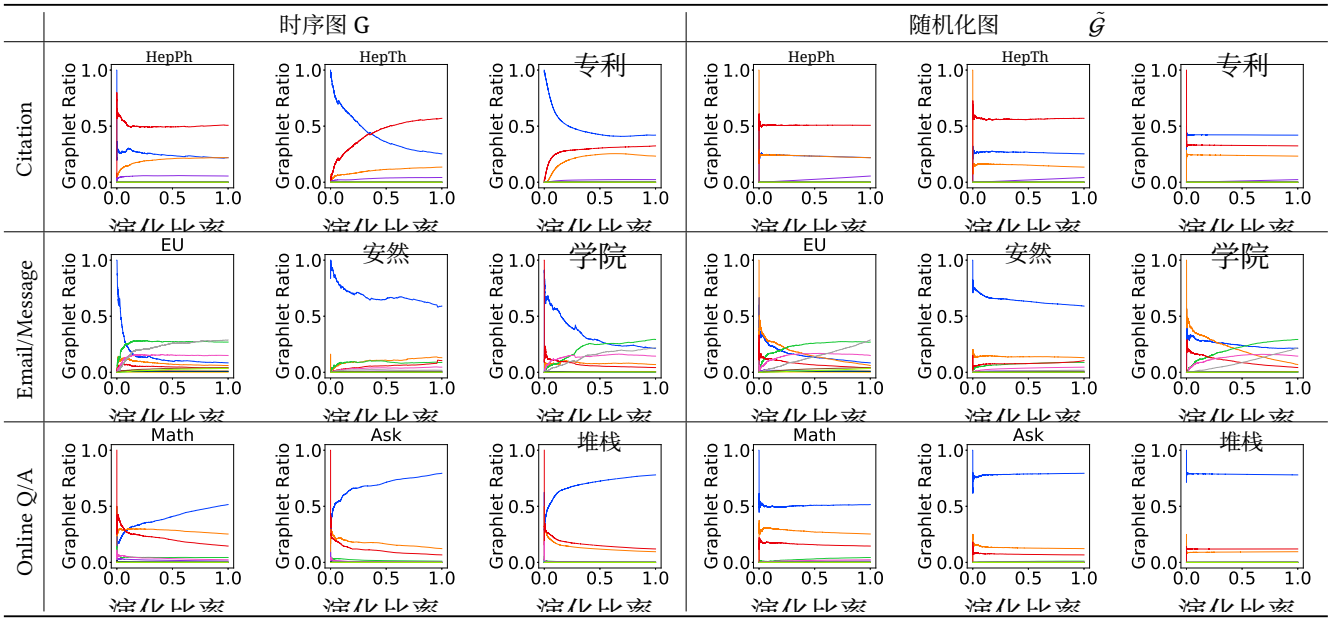
定理 1. 算法 *1* 的时间复杂度为 $\Theta(\Sigma_{v\in\mathcal{V}}(d(v))^2)$。

证明。由于构成每个图元实例的节点数是常数，找到与给定实例对应的图元并更新相应的计数（第 6–7 行和第 10 行）需要 $O(1)$ 时间。因此，处理每条传入边 $e_i = u \to v$所需的时间，取决于计算 u和 v的邻居并集（第 4 行），其复杂度为 $\Theta(d(t_i-1)(u) + d(t_i-1)(v))$. 因此，总复杂度为 $\Theta(\Sigma_{e_i=u\to v\in E}(d(t_i-1)($ u v = Θ $d\ v\ ^2$ □) + $d(t_i-1)$ ( )) $(\Sigma_{v\in\mathcal{V}}($ ( ))).

引理 1. 快照$\mathcal{G}(t)$ 中全部图元的实例数为 $\Theta(\Sigma_{v\in\mathcal{V}^{(t)}}(d(t)(v))^2)$。

证明。给定一个快照 G(t) = (V (t), E (t))，对于每个节点$v \in$V (t)，如果我们对由 v 及其两个邻居组成的全部图元实例进行计数，那么对于每个节点 v，此类实例的计数为$\Theta((d(t)(v))^2)$；由于对于每个节点 v都有 $d(t)(v) \geq 1$，因此总计数 C为 $\Theta(\Sigma_{v\in\mathrm{V}^{(t)}}(d(t)(v))^2)$。

表 3：随时间变化的图元比率。图中的颜色与图2中图元的颜色相对应，演化比率表示添加到图中的边所占的比例。真实世界图中的演化模式因领域而异（观察1），且与随机化图中的演化模式明显不同（观察2）。



下界： 由于每个由三个节点组成的图元实例至多被计数三次，因此 C 至多是 $G_{(t)}$中全部图元实例数量的三倍。换言之，全部图元实例的数量至少是 C 的 1/3，因此为 $\Omega(\sum_{v\in V^{(t)}}(d_{(t)}(v))^2)$。

上界： 在每个图元实例中，至少存在一个中心节点，它与其邻居共同构成该图元。因此，每个实例至少会被计数一次，因此 C 至少等于 $G_{(t)}$中所有图元的实例数。换句话说，所有图元的实例数至多为 C，因此其为$O(\sum_{v\in V^{(t)}}(d_{(t)}(v))^2)$。□

如表 3 所示，比率的变化取决于图所属的领域，如观察1所总结。

观察1 . 同一领域的图中，图元实例分布的动态变化具有一些共同点。
- 图元*4*的实例在引文图中占比更高而非其他图。
- 边较多的图元（例如图元*8*、*12*和*13*）占据在电子邮件/消息网络中所占比例高于其他网络。
- 图元 *1* 的比例仅在在线问答图中随时间增加。

然而，领域内部的动态并不完全相同。例如，虽然图元 1、2 和 4 在所有引文图中都比其他图元占主导地位，但它们之间的比率在不同图中差异很大。

我们还注意到，真实世界图与随机化图中的动态存在一致的差异（见第2.1节），如观察2所总结。

观察2. 随机化图中图元实例的比率变化比真实世界图中更趋于线性。

表4：真实世界图和随机化图中图元实例比率随时间变化的非线性度。我们在第 3.1 节介绍如何度量非线性度。非线性度越低，相应图元实例比率的变化就越趋于线性。请注意，随机化图中图元实例的比率变化比真实世界图中更趋于线性。

数据集	HepPh	Hepth	专利	EU	安然	学院	Math	Ask	堆栈
real	0.0027	0.0080	0.0093	0.0107	0.0042	0.0095	0.0028	0.0038	0.0047
随机	0.0003	0.0011	0.0000	0.0081	0.0017	0.0058	0.0007	0.0005	0.0001

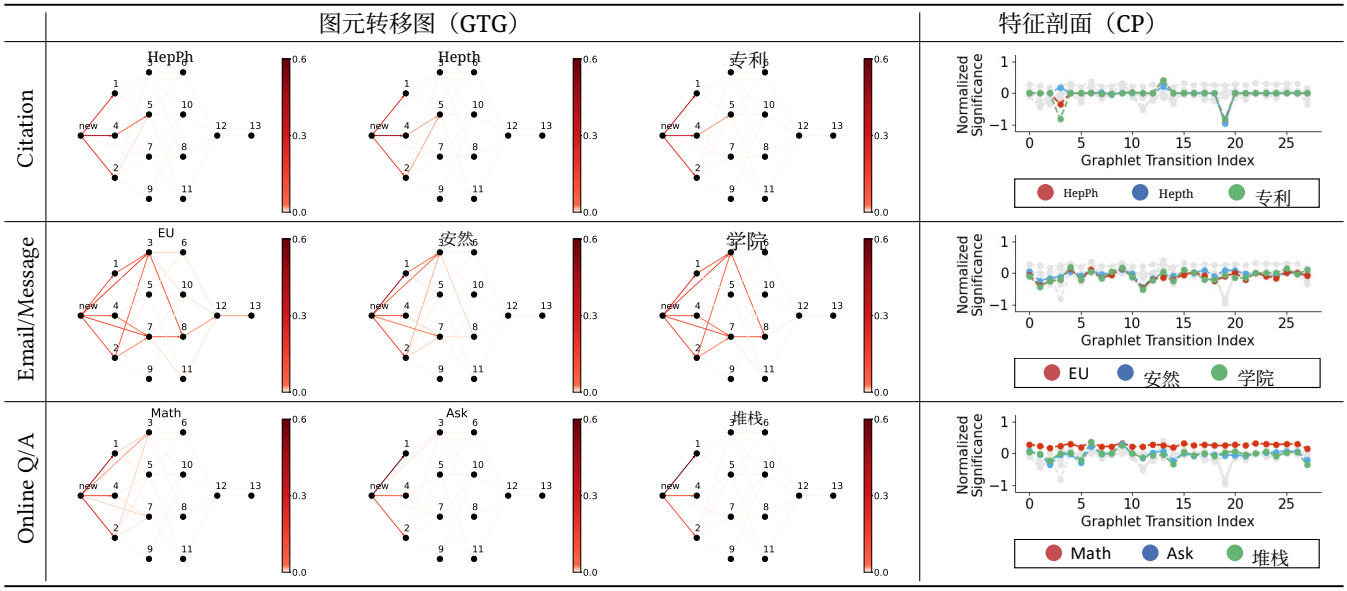
为了用数值支持这一观察，我们度量图元实例比率随时间变化的非线性程度 [18, 22] 。具体而言，我们对表 3 中的每个时间序列分别拟合线性回归模型和非线性多项式回归模型，然后将两个模型预测值之间的平均绝对差作为该时间序列的非线性程度。[1]最后，我们对每个图的所有时间序列的非线性程度取平均值，并在表 4 中报告结果。请注意，真实世界图的非线性程度显著高于相应的随机化图。也就是说，与真实世界图相比，随机化图中的图元实例比率变化更加线性。

3.2 全局层级 2. 图元转移

在时序图中，由于添加了新边，一个图元实例可能会转移为另一个图元的实例。在本小节中，我们考察图元之间此类转移的计数，以刻画时序图中的局部动态，并对它们进行比较。

[1]我们使用 Analyse-it（版本 5.65）中实现的线性度检验，并按照程序建议，选用三次模型作为非线性多项式模型。为提高计算效率，我们在 1,000 个等间隔均匀采样的演化比率处测量绝对差值。

表5：利用图元转移图（GTG）以及由图元转移图得到的特征剖面（CP），我们可以准确刻画真实世界图中局部结构的动态变化。图元转移图中边的颜色表示其归一化权重。请注意，来自相同领域的真实世界图，其图元转移图和特征剖面尤为相似（观察3）。



图元转移图： 我们将图元转移图（GTG）定义为对图元之间的转移进行编码的图。

定义1（图元转移图）。时序图$\mathcal{G}$的图元转移图（*GTG*）$G=(V,E,W)$是一个静态有向加权图，其中节点为图元，每条边表示源图元通过向$\mathcal{G}$添加一条边而转变为目标图元。边的权重是相应转移发生的次数。我们用$W=\{w_1,\cdots w_{|E|}\}$表示边权重。

由于我们关注图2(a)中的13种图元，图元转移图由这些图元之间的28种转移构成。表5展示了真实世界图的图元转移图。算法2描述了图元转移图边权重的计算方法。简而言之，我们按照到达顺序遍历每条边，并统计由其引起的转移。其时间复杂度由定理2形式化给出。

定理2. 算法*2*的时间复杂度为 $\Theta(\Sigma_{v\in\mathcal{V}}(d(v))^2)=\Theta$(最后一个快照中所有图元的实例数量)。

证明。我们可以类似于定理 1 证明 $\Theta(\Sigma_{v\in\mathcal{V}}(d(v))^2)$ 的复杂度；根据引理 1，该复杂度为 Θ(最后一个快照中所有图元的实例数量)。□

特征剖面（CP）： 我们使用图 G 的 GTG $G=(V、E、W)$ 中边权重的显著性来刻画其局部结构的演化。为了度量显著性，我们遵循 [31] 中度量每个图元自身显著性的步骤。为此，我们构建随机图 ˜G 的图元转移图 $\tilde{G}$。然后，我们按如下方式度量 G 中每个边权重 w_i 的显著性 SP_i：

$$SP_i := \frac{w_i - \tilde{w}_i}{w_i + \tilde{w}_i + \epsilon}, \tag{1}$$

算法2： 计算图元转移图的边权重

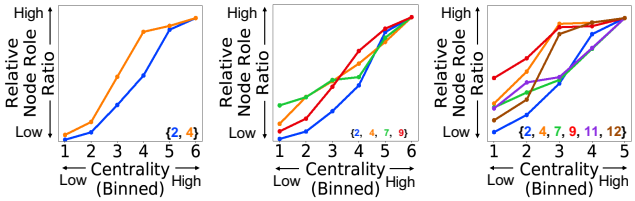
输入：时序图 G =(V, E, T) 输出：G 的图元转移图的边权重 1 将所有边权重初始化为零 2 将 E 初始化为空集 3 按到达顺序遍历每条边 $e_i = u \to v$ 4 对于每个 $w_1 \in$ 邻居$(u)\setminus\{v\}$，执行 5 UPDATE(u, v, w_1) 6 对于每个 $w_2 \in$ 邻居$(v)\setminus\{$邻居$(u) \cup u\}$，执行 7 UPDATE(u, v, w_2) 8 将 $u \to v$ 添加到 E 9 返回边权重 10 过程 UPDATE(u, v, w) 11 如果 u、v和 w 构成一个图元实例，则 12 该实例 (u, v, w) 去掉 $u \to v$后的前一个图元 ← 13 该实例 (u, v, w) 加上 $u \to v$后的下一个图元 ← 14 $i \leftarrow$ 图元转移的索引，从 prev 到 next 15 将边 i（即 w_i）的权重增加 1

˜其中，$\tilde{w}_i$是 G中的相应边权重，ϵ是一个常数，我们将其固定为 4。对于 $\tilde{w}_i$，我们生成 50 个随机化图实例，并使用其中的平均边权重。最后，我们按如下方式对每个显著性进行归一化：

$$CP_i := SP_i / \sqrt{\Sigma_{i=1}^{|E|} SP_i^2}. \tag{2}$$

我们使用归一化显著性向量（即 $[CP_1,\cdots\ CP_{|E|}]$）来刻画 G 中局部结构的演化，并将其称为特征剖面（*CP*）。

CP 之间的比较： 我们绘制了表5中所考察真实世界图的 CP，可以观察到同一领域内具有较高的相似性。我们通过数值计算相似性



(a) $d_\theta = 2$ **(b)** $d_\theta = 4$ **(c)** $d_\theta = 8$

图3：节点局部结构中关于其未来重要性的示例信号。一些节点角色（例如，节点角色2和4）的比率会随着节点未来入度的增加而单调上升。对这些比率进行了重新缩放，使其最大值相同。

，方法是使用皮尔逊相关系数比较 CP，结果如图1(a)所示。同一领域内的图之间相关系数尤其高；具体而言，若采用相关系数的最佳阈值（0.58），则领域的分类准确率可达 97.2%。结果表明，CP 能准确刻画局部结构的演化。我们的观察总结于观察3。

观察3。 来自相同领域的真实世界图，其局部结构的演化模式相似。

与其他方法的比较： 我们评估了另外三种图表征方法，评估方式与上文右侧段落中评估我们的方法相同。在图1(b)中，我们给出了根据各图元实例计数所得 CP 之间的相关系数 [31]。请注意，电子邮件/消息图（蓝色）和在线问答图（绿色）无法清晰区分。从数值上看，采用最佳相关系数阈值（0.95）时，分类准确率为83.3%。

我们还使用图元轨道转移（GoT） [4] 和轨道时间一致性（OTA） [4], 计算所考虑的真实世界图之间的相似性；这两种方法也基于图元之间的转移（详情见第6节）。与这些方法相比，我们的表征方法具有以下主要优势：

- **(1) 速度：** 如附录A所示，GoT 和 OTA 的速度最多比我们的方法慢 10×。它们的时间复杂度与所有已使用快照中图元实例计数之和成正比，而算法2的时间复杂度仅与最后一个快照中的图元实例计数成正比（定理2）。

- **(2) 空间效率：** 如附录A所示，GoT 和 OTA 在最大的两个图（专利和 Stackoverflow）上会耗尽内存，而我们的方法不会。它们需要存储每个待比较快照中的所有图元实例，以便与下一个快照进行比较；而算法2只维护最新的快照，无需存储图元实例。
- **(3) 分类准确率：** 使用所考虑的真实世界图计算出的最佳分类准确率（专利和 Stackoverflow 除外，因为 GoT 和 OTA 会耗尽内存）分别为 81.0%（GoT）和 85.7%（OTA），低于我们的分类准确率（97.2%）。详细结果见附录A。请注意，GoT 和 OTA 根据少量快照估算图元之间的转移计数，而算法2则对转移进行精确计数。

表 6：节点角色比例与未来中心性之间的 Spearman 秩相关系数绝对值（针对每种中心性度量，对所有节点角色和所有数据集取平均值），以及阈值 d_θ 的各个取值。随着节点邻居数量增加（即 d_θ 增大），有关未来中心性的局部结构信号会变得更强（即绝对值增大）。

d_θ	度	中介中心性	接近度	PageRank	边介数
2	0.640	0.697	0.682	0.663	0.546
4	0.721	0.723	0.712	0.704	0.558
8	0.816	0.793	0.759	0.701	0.599

总而言之，**我们使用图元转移图（GTG）表征时序图的方法，以**97.2%**的准确率最准确地区分时序图的领域**。其他方法的准确率分别为83.3%、81.0%和85.7%。

4 节点级分析

在本节中，我们研究节点周围的局部结构与其未来重要性之间的关系。然后，我们利用这些关系提高未来节点中心性的可预测性。

4.1 模式

我们使用节点角色来表征节点的局部结构，并考察这些结构与节点未来中心性之间的关系。

节点的局部结构： 给定一个时序图 G，我们通过测量每个节点角色在时间 t 的快照中所占的比率，来刻画每个节点 v 在其早期阶段的局部结构；此时，v 的入度首次达到阈值 d_θ。也就是说，每个节点 v 都表示为一个 30 维向量，其 i维为$m(t)_i(v)/(\sum_{j=1}^{30} m(t)_j(v))$（关于 $m(t)_i(v)$，参见第2.1节）。

节点的未来重要性： 给定一个时序图 G，我们用每个节点在 G 的最后一个快照中的入度、节点介数中心性 [15], 接近中心性 [9], 和 PageRank [34] 来衡量其未来重要性。根据每种中心性度量，我们将 G 中的节点分为六组（第1组：前 50–100 %；第2组：前 30–50%；第3组：前 10–30%；第4组：前 5–10%；第5组：前 1–5%；第6组：前 0–1%）。

发现信号： 对于每个组，我们对组内节点的比率向量取平均值。图3展示了使用入度作为中心性度量时的一些平均比率。请注意，无论 d_θ 取何值，节点角色2和4的比率都会随着未来中心性的增加而单调上升。也就是说，节点角色2和4的比率能够为节点未来的入度提供一致的信号。

图4报告了每个平均比率与节点未来中心性之间的斯皮尔曼秩相关系数 [43]（具体而言，是上述1至6的组编号）。表6还报告了针对每种中心性度量和每个阈值 d_θ，系数绝对值的平均值（对所有节点角色和所有数据集取平均）。请注意，这些平均值显著大于0，具体约为0.7；并且随着 d_θ 增加而增大，正如观察4所总结的那样。

观察4。 在真实世界图中，节点早期阶段的局部结构能够提供有关其未来重要性的信号。节点的邻居越多，这些信号就越强。

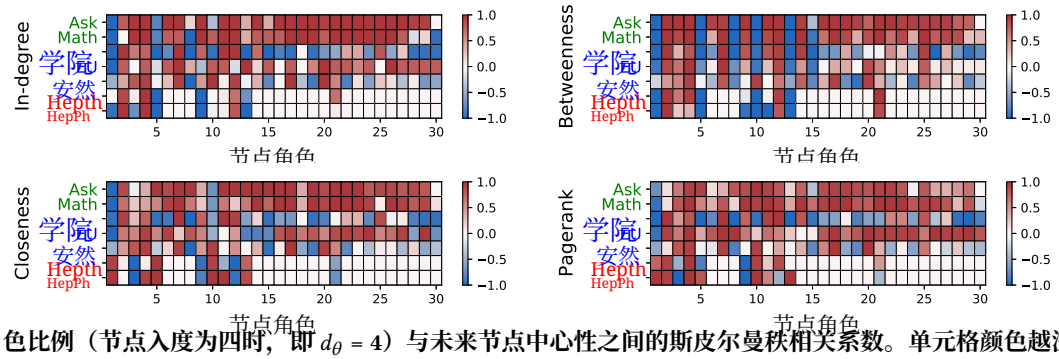


图4：节点角色比例（节点入度为四时，即 $d_\theta = 4$）与未来节点中心性之间的斯皮尔曼秩相关系数。单元格颜色越深，对应系数的绝对值越大。请注意，大多数系数的绝对值都显著大于0。

4.2 预测

基于上述观察，我们利用节点在早期阶段所具有的角色计数来预测其未来中心性。

问题形式化： 我们将预测问题表述为一个分类问题，如问题1所述。

问题1（节点中心性预测）

- 给定：输入图的快照 $\mathcal{G}(t_v, d_\theta)$，其中节点 v 的入度首次达到 d_θ，
- 预测：节点 v 的中心性是否属于20%，即在 $\mathcal{G}$ 的最后一个快照中排名前 *20%*。

我们使用入度、介数中心性、接近中心性和 PageRank 作为中心性度量。对于 d_θ，我们使用 2、4 或 8。

输入特征： 对于每个节点 v，我们考虑输入图 G 的快照 $G(t)$，此时 v 的入度首次达到 d_θ。也就是说，$t = t_{v,\ d_\theta}$ 以及 $G(t) = G(t_{v,\ d_\theta})$。然后，我们为 v 提取以下输入特征集合：

- **局部-NR：** 在 $G(t)$ 中，节点角色在 v 处的计数。也就是说，$[m(t)1(v)\ \ m(t)2(v),\ \cdots\ \ m(t)30(v)]$（关于 $m(t)i(v)$，参见第2.1节）。
- **局部 NPP [41]：** </style}在 **G(**t**)** 中，我们计算：(1) v处的三角形计数；(2) 以 v 为中心的楔形计数；(3) 以 v 为端点的楔形计数。

- **全局基础：** 快照中的节点和边的计数。
- **全局-NR：** 我们计算一个 30 维向量，其中第 i 个分量 $m(t)i(v)/(\sum_{j=1}^{30} m(t)j(v))$ 表示 $G(t)$ 中 v 处各节点角色的比率。然后，我们使用 $G(t)$ 中度为 d_θ 的所有节点的角色比例向量的均值和标准差，对该角色比例向量进行标准化（即计算其 z-分数）。局部-NR 中的特征也包含在内。
- **全局-NPP [41]：** </style}在 **G(**t**)** 中，我们计算：(1) 与 v 不关联的边数；(2) 非邻接节点对的数量，其中一个节点是 v 的邻居，另一个既不是 v 的邻居，也不是 v 本身。局部-NPP 中的特征也包含在内。

- **全部：** 全部 **全局-NR、全局-NPP**和**全局基础**。

请注意，我们根据是否使用 $G(t)$ 中的全局信息（即 $G(t)$中的所有节点数），还是仅使用 v 处的局部信息，将上述集合分为全局集合和局部集合。

预测方法： 作为分类器，我们使用 Scikit-learn 库中的随机森林 模型。该模型包含 30 棵决策树，最大深度为 10。

表 7：在 $d_\theta = 2$ 对所考察的 7 个真实世界图取平均后，预测未来节点重要性任务的 F1 分数、准确率和 AUROC。在局部特征中，在所有设置下，使用局部-NR 的表现都优于使用 Local-NPP。大多数情况下，使用全部特征的表现最佳，这表明所考察的特征集彼此互补。各数据集的详细结果见附录 C。

目标	度			中介中心性		
度量	F1 分数	准确率	AUROC	F1 分数	准确率	AUROC
局部-NR	0.39	0.69	0.68	0.59	0.83	0.82
局部 NPP	0.38	0.68	0.64	0.58	0.81	0.79
全局-NR	**0.57**	**0.74**	**0.78**	0.64	0.84	0.85
全局-NPP	**0.57**	0.73	0.77	0.64	0.84	0.85
全局基础	0.50	0.72	0.73	0.24	0.73	0.67
ALL	**0.57**	**0.74**	**0.78**	**0.65**	**0.85**	**0.86**

目标	接近度			PageRank		
度量	F1 分数	准确率	AUROC	F1 分数	准确率	AUROC
局部-NR	0.51	0.76	0.78	0.42	0.73	0.73
局部 NPP	0.43	0.70	0.69	0.37	0.69	0.67
全局-NR	0.68	0.82	0.87	0.54	**0.75**	**0.79**
全局-NPP	0.66	0.80	0.85	0.54	0.74	0.78
全局基础	0.59	0.75	0.79	0.47	0.71	0.74
ALL	**0.69**	**0.83**	**0.88**	**0.56**	**0.75**	**0.79**

评估方法： 我们使用 80% 的节点进行训练，其余 20% 用于测试。我们根据 F1 分数、准确率和*ROC* 曲线下面积 （AUROC）评估预测性能。数值越高，表示预测性能越好。

结果： 表 7 展示了 $d_\theta = 2$ 时每组输入特征的预测性能，表 8 展示了性能如何取决于入度阈值 d_θ。在这些表中，我们报告了第 2.2 节中 7 个数据集（最大的两个数据集，即专利和 Stackoverflow 除外）上每种预测性能经过 10 次运行后的均值。根据结果，我们得出以下观察。

观察 5。 在局部特征中，每个节点的节点角色计数（局部-***NR***）相比（局部 ***NPP***）包含更多关于未来重要性预测的信息。

观察 6。 所考虑的特征集彼此互补。在大多数情况下，使用全部特征（全部）可获得最佳预测性能。

观察 7。 节点拥有的邻居越多，就越能准确地预测其未来重要性。

表 8：根据 d_θ（即提取输入特征时节点的入度），预测未来节点重要性任务中的平均 F1 分数、准确率和 AUROC。大多数情况下，整体性能随着d_θ 的增加而提升。也就是说，节点拥有的邻居越多，就越能准确地预测其未来重要性。每个数据集的详细结果见附录 C。

目标	度			中介中心性		
度量	F1 分数	准确率	AUROC	F1 分数	准确率	AUROC
全部 ($d_\theta = 2$)	0.59	0.74	0.78	0.65	0.85	0.86
全部 ($d_\theta = 4$)	0.69	0.78	0.79	0.73	0.83	0.87
全部 ($d_\theta = 8$)	0.80	0.81	0.86	0.82	0.85	0.90
目标	接近度			PageRank		
度量	F1 分数	准确率	AUROC	F1 分数	准确率	AUROC
全部 ($d_\theta = 2$)	0.69	0.83	0.88	0.55	0.75	0.79
全部 ($d_\theta = 4$)	0.78	0.83	0.89	0.73	0.77	0.80
全部 ($d_\theta = 8$)	0.86	0.88	0.92	0.85	0.85	0.83

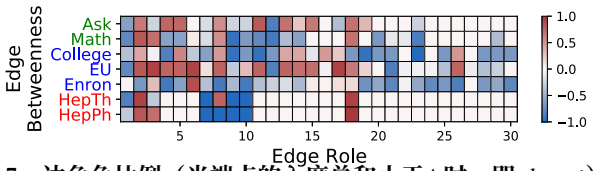


图 5：边角色比例（当端点的入度总和小于4 时，即 $d_\theta = 4$）与未来边中心性之间的斯皮尔曼秩相关系数。单元格颜色越深，对应系数的绝对值越大。请注意，许多系数的绝对值显著大于0，但它们往往小于图4中的系数。

特征重要性：此外，我们使用基尼重要性 [29], 度量集合**全部**中各项特征的重要性，并在附录 B 的表 10 中报告最重要的五项特征。

观察 8。较强的预测因子因中心性度量而异。例如，对于介数中心性，作为桥的节点角色的计数（即***Local NR-4*** 和***Global NR-4***）具有较强的预测能力。

5 边级分析

在本节中，我们研究每条边的局部结构对其未来中心性的信号，并基于该信号预测边的未来重要性。

我们总体上遵循第 4 节中的流程，但有以下区别：(a) 当每条边 $u \to v$ 的端点 u 和 v 的入度之和变为d_θ时，我们考察边角色的比率；(b) 我们使用边介数中心性 [15] 作为重要性度量；(c) 我们按照问题 2 中的描述，将预测未来边重要性的问题形式化；(d) 我们使用边上边角色的（相对）计数提取特征集**Local-ER** 和 **Global-ER**，方式与提取**Local-NR** 和**Global-NR**相同；(e) 对于**全部**，我们取**Global-ER** 和**Global-Basic** 的并集。

问题 2（边中心性预测）。
- 给定：当每条边的端点入度之和首次达到 d_θ 时，输入图的快照$\mathcal{G}(t_e,\ ^{d_\theta})$，
- 预测：判断每条边的中心性是否属于20%，在$\mathcal{G}$的最后一个快照中。

根据图 5、表 6 和表 9，我们得出以下观察结果。

表 9：在所考虑的 7 个真实世界图上，对未来边重要性进行预测任务的 F1 分数、准确率和 AUROC 的平均值。在大多数设置下，使用基于边角色的特征（Local-ER 和 Global-ER）比使用全局基础（Global-Basic）表现更好。总体表现相对于 d_θ 有所提升。也就是说，边之间连接得越好，对其未来重要性的预测就越准确。各数据集的详细结果见附录 D

目标	边介数		
度量	F1 分数	准确率	AUROC
Local-ER ($d_\theta = 2$)	0.45	0.78	0.76
Global-ER ($d_\theta = 2$)	0.47	0.81	0.78
全局基础 ($d_\theta = 2$)	0.42	0.79	0.73
全部 ($d_\theta = 2$)	0.50	0.80	0.75
全部 ($d_\theta = 2$)	0.50	0.80	0.75
全部 ($d_\theta = 4$)	0.53	0.82	0.84
全部 ($d_\theta = 8$)	0.52	0.85	0.85

观察9. 在真实世界图中，与边的未来重要性相关的边的局部结构信号，在早期阶段较弱，相比之下，来自节点局部结构的信号更强（见图*5*）。

观察10. 然而，随着边的连接变得更加紧密，这些信号会增强，从而带来更好的预测性能（见表*6*和表 *9*）。

观察 11. 对于未来重要性预测，边角色（***Local-ER*** 和***Global-ER***）的特征比简单的全局统计量（***Global-Basic***）包含更多信息（见表 *9*）。

6 相关工作

以往关于时序图分析的研究大致可分为三类：（a）为流图设计算法 [13, 23, 27, 30],；（b）发现图中的时序模式 [2, 3, 7, 11, 26], ；（c）生成具有真实动态的图 [3, 8, 25]。本研究属于第二类。

这一类研究揭示了：（a）普适时序模式，例如致密化 [26], 、直径收缩 [26],，以及主特征值与边数之间的幂律关系 [3]；（b）特定领域的模式，包括超链接网络 [12], 、代谢网络（例如生化反应和蛋白质相互作用） [11], 、通信网络（例如电话通话和短信） [2, 16], 以及友谊网络 [7]中的模式。

特别是，对于局部结构的分析，图元这一概念 [36]（即由预先设定数量的节点构成的同构小子图集合）已扩展到时序图中。这些扩展称为时序网络模体，且有多个变体。Kovanen 等人 [21] 将其定义为由固定数量节点构成的时序子图集合，这些子图满足：(a) 拓扑等价；(b) 时序等价（具体而言，组成边的相对顺序相同）；(c) 连续（具体而言，对于每个节点，组成边都是连续的）；以及 (d) 时序局部（具体而言，连续边的到达时间足够接近）。Hulovaty 等人 [19] 忽略了条件 (c)；Paranjape 等人 [35] 则忽略了条件 (c)，并通过仅限制第一条边与最后一条边之间的时间差来放宽条件 (d)。请注意，所有这些概念都聚焦于时序局部子图，因此仅适用于分析短期动态。

对于局部结构中的长期动态，David 等人 [4] 提出了图轨道转移（GoT）和轨道时间一致性（OTA），通过近似计数节点角色之间的转移来刻画时序图的动态。然而，由于计算开销高，在估算转移计数时只能比较少量快照，因此其刻画能力显著弱于我们使用图元转移图（GTG）的刻画方法（见第 3.2 节）。请记住，我们的方法会计数图元之间的“每一次”转移，而且速度仍显著快于 GoT 和 OTA（见附录 A 节）。

为了预测节点未来的入度，Yang 等人 [41]提出使用从包含三个节点的图元中提取的五种特征（描述见第 4.2 节）。实证结果表明，我们提出的特征往往比这五种特征具有更好的预测性能；更重要的是，这些特征彼此互补。Faisal 和 Milenković [14] 旨在检测衰老相关节点，即在基因表达过程中，其拓扑属性（例如图元计数）随时间发生显著变化的节点。

在算法方面，针对统计每种图元实例数量的问题（也称为子图计数问题），已有大量算法被提出。子图计数综述 [37], 指出，子图计数算法大致分为精确计数 [1, 32, 33, 38] [6, 40]

和近似计数两类。前一类又分为基于枚举的方法[32, 38], 、基于矩阵的方法 [33], 和基于分解的方法 [1]。算法 1 属于第一种子类，正如第 3.1 节开头所述，它达到了该子类算法所能达到的最优时间复杂度。该算法改编自 StreaM [38], ; StreaM 用于在全动态图流中维护 4 节点无向图元实例的计数（即边插入与删除流）。

7 结论

在这项工作中，我们考察了图在图元、节点和边层级上所捕捉的局部结构的长期演化。我们的贡献总结如下：

• **模式：**我们考察了时序图中局部结构动态变化的多种模式。例如，真实世界图中的图元随时间变化的分布与随机图中的分布存在显著差异，而来自同一领域的图中，图元之间的转移却惊人地相似。此外，节点和边在早期阶段的局部结构能够为其未来重要性提供强有力的信号。

• **工具：**我们提出了图元转移图，并证明它是一种有效工具，可用于衡量不同规模时序图之间的相似性。

• **可预测性：**通过引入基于节点角色和边角色的新特征，我们提高了对节点和边未来重要性的预测准确率。这些特征也与全局图统计量互为补充。

可复现性：代码和数据集可在 https: // github.com/deukryeol-yoon/graphlets-over-time 获取。

参考文献

[1] Nesreen K Ahmed、Jennifer Neville、Ryan A Rossi、Nick G Duffield 和 Theodore L Willke。2017。图元分解：框架、算法与应用。KAIS 50, 3（2017），689–722。[2] Leman Akoglu 和 Bhavana Dalvi。2010。大型电话与短信网络中的结构、关系持久性与事件检测。载于 MLG。[3] Leman Akoglu、Mary McGlohon 和 Christos Faloutsos。2008。RTM：加权时变图的定律与递归生成器。载于 ICDM。[4] David Aparício、Pedro Ribeiro 和 Fernando Silva。2018。图元轨道转移（GoT）：用于时序网络比较的指纹。PloS one 13, 10（2018），e0205497。[5] Alex Arenas、Alberto Fernandez、Santo Fortunato 和 Sergio Gomez。2008。复杂网络中基于模体的社群。Journal of Physics A: Mathematical and Theoretical 41, 22（2008），224001。[6] Çigdem Aslay、Muhammad Anis Uddin Nasir、Gianmarco De Francisci Morales 和 Aristides Gionis。2018。在演化图中挖掘频繁模式。载于 CIKM。[7] Ashwin Bahulkar、Boleslaw K Szymanski、Omar Lizardo、Yuxiao Dong、Yang Yang 和 Nitesh V Chawla。2016。分析 NetSense 数据中的链接形成、持久性与消亡。载于 ASONAM。[8] Albert-László Barabási 和 Réka Albert。1999。随机网络中标度规律的涌现。《科学》286, 5439（1999），509–512。[9] Alex Bavelas。1950。面向任务的群体中的通信模式。The journal of the acoustical society of America 22, 6（1950），725–730。[10] Austin R Benson、David F Gleich 和 Jure Leskovec。2016。复杂网络的高阶组织。《科学》353, 6295（2016），163–166。[11] Antje Beyer、Peter Thomason、Xinzhong Li、James Scott 和 Jasmin Fisher。2010。使用定性网络深入了解 2 型糖尿病和肥胖期间的代谢紊乱机制。载于 Transactions on Computational Systems Biology XII。Springer，146–162。[12] Andrei Broder、Ravi Kumar、Farzin Maghoul、Prabhakar Raghavan、Sridhar Rajagopalan、Raymie Stata、Andrew Tomkins 和 Janet Wiener。2011。Web 的图结构。载于 The Structure and Dynamics of Networks。Princeton University Press，183–194。[13] Dhivya Eswaran、Christos Faloutsos、Sudipto Guha 和 Nina Mishra。2018。Spotlight：检测流图中的异常。载于 KDD。[14] Fazle E Faisal 和 Tijana Milenković。2014。动态网络揭示衰老过程中的关键参与者。《生物信息学》30, 12（2014），1721–1729。[15] Linton C Freeman。1977。基于中介中心性的中心性度量集。Sociometry（1977），35–41。[16] Cesar A Hidalgo 和 Carlos Rodríguez-Sickert。2008。移动电话网络的动态。Physica A: Statistical Mechanics and its Applications 387, 12（2008），3017–3024。[17] Tomaž Hočevar 和 Janez Demšar。2016。稀疏图中节点和边的图元轨道计算。Journal of Statistical Software 71, 1（2016），1–24。[18] Eric Hsieh 和 Jen-pei Liu。2008。测定验证中线性度的统计评估。Journal of biopharmaceutical statistics 18, 4（2008），677–690。[19] Yuriy Hulovatyy、Huili Chen 和 Tijana Milenković。2015。利用动态图元探索时序网络的结构与功能。《生物信息学》31, 12（2015），i171–i180。[20] Krzysztof Juszczyszyn 和 Grzegorz Kołaczek。2011。网络通信图中基于模体的攻击检测。载于 CMS。[21] Lauri Kovanen、Márton Karsai、Kimmo Kaski、János Kertész 和 Jari Saramäki。2011。时间依赖网络中的时序模体。Journal of Statistical Mechanics: Theory and Experiment 2011, 11（2011），P11005。[22] Martin H Kroll 和 Kenneth Emancipator。1993。线性度的理论评估。Clinical chemistry 39, 3（1993），405–413。[23] Dongjin Lee、Kijung Shin 和 Christos Faloutsos。2020。具备时序局部性感知能力的采样方法，可在真实图流中准确进行三角形计数。VLDB 29, 6（2020），1501–1525。[24] John Boaz Lee、Ryan A Rossi、Xiangnan Kong、Sungchul Kim、Eunyee Koh 和 Anup Rao。2019。采用基于模体的注意力机制的图卷积网络。载于 CIKM。[25] Jure Leskovec、Deepayan Chakrabarti、Jon Kleinberg、Christos Faloutsos 和 Zoubin Ghahramani。2010。克罗内克图：一种网络建模方法。JMLR 11, 2（2010）。[26] Jure Leskovec、Jon Kleinberg 和 Christos Faloutsos。2005。图随时间演化：致密化定律、收缩直径及可能的解释。载于 KDD。[27] David Liben-Nowell 和 Jon Kleinberg。2007。社交网络中的链接预测问题。JASIST 58, 7（2007），1019–1031。[28] Zhijun Liu、Chao Huang、Yanwei Yu 和 Junyu Dong。2021。保模体动态图属性网络嵌入。载于 WWW。[29] Wei-Yin Loh。2011。分类回归树。WIREs: data mining and knowledge discovery 1, 1（2011），14–23。[30] Andrew McGregor。2014。图流算法综述。ACM SIGMOD Record 43, 1（2014），9–20。

表 10：特征重要性分析的结果。我们报告了最强的五个预测因子及其基尼重要性。

中心性	排名 1		排名 2		排名 3		排名 4		排名 5	
度	边数	0.07	全局-NPP 2	0.06	节点数	0.06	全局-NR 3	0.04	全局-NR 10	0.03
中介中心性	局部 NPP 2	0.10	局部-NR 4	0.09	全局-NR 4	0.08	局部-NR 9	0.06	全局-NR 3	0.05
接近度	全局-NR 5	0.09	局部-NR 5	0.07	边数	0.07	全局-NPP 2	0.06	节点数	0.06
PageRank	局部-NR 1	0.07	边数	0.06	全局-NPP 2	0.06	节点数	0.05	全局-NR 1	0.05
边介数	Global-ER 7	0.11	Global-ER 2	0.09	Global-ER 3	0.09	节点数	0.07	Local-ER 7	0.07

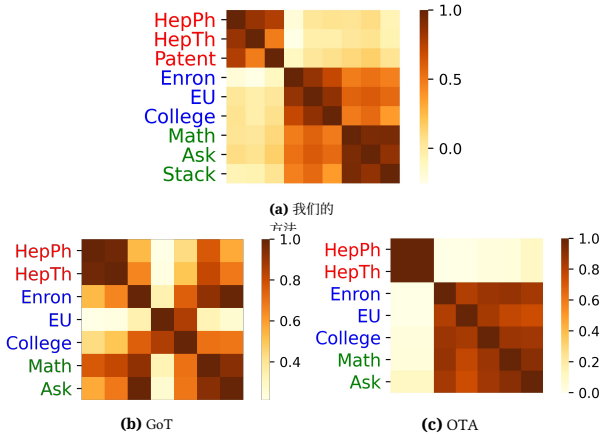


图6：我们的方法、GoT 和 OTA 得出的相似度矩阵。与 GoT 或 OTA 相比，我们的方法对图的领域（以颜色区分）分类得更准确。

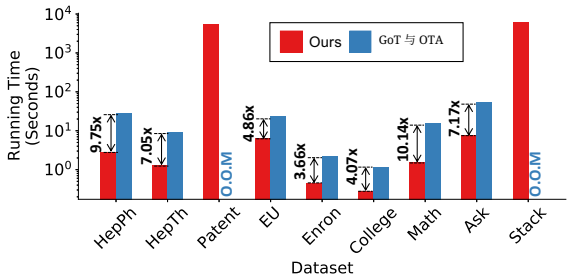


图 7：我们的方法、GoT 和 OTA 的运行时间。我们的方法始终显著快于另外两种方法；后两者在最大的两个数据集上会因内存耗尽而无法运行。

使用 Java 基于图元转移图实现了我们的表征方法，并采用了作者提供的 GoT 和 OTA 官方实现，这些实现使用 C++ 编写。在每个数据集中，我们为 GoT 和 OTA 使用了时间间隔相同的 12 个快照。

输出相似矩阵：图6展示了我们基于图元转移图、GoT 和 OTA 的特征描述方法生成的输出相似矩阵。GoT 和 OTA 在两个最大的数据集（专利和 Stackoverflow）上耗尽了内存。GoT 和 OTA 都无法清晰地区分电子邮件/消息图（蓝色）和在线问答图（绿色）。从数值上看，在最佳相似性阈值下，分类准确率分别为 81.0%（GoT）和 85.7%（OTA），而我们的方法的准确率为 97.2%。

速度比较：如图 7 所示，我们的方法在所有图上都比 GoT 和 OTA 快。具体而言，我们的方法的平均速度比其他方法快 6.68×。

B 特征重要性分析

我们使用 基尼重要性 [29], 来度量集合 **全部** 中的每个特征的重要性（参见主论文第 4.2 节），并在表 10 中报告排名前五的重要特征。

C 未来节点重要性预测的详细结果

在表11-16中，我们报告了在每个真实世界图上预测未来节点中心性任务的各项评估指标的平均预测性能和标准差，结果基于10次运行。详细的实验设置见论文正文第 4.2 节。

D 未来边重要性预测的详细结果

在表17-19中，我们报告了在每个真实世界图上预测未来节点中心性任务的各项评估指标的平均预测性能和标准差，结果基于10次运行。详细的实验设置见论文正文第5节。

[31] Ron Milo、Shalev Itzkovitz、Nadav Kashtan、Reuven Levitt、Shai Shen-Orr、Inbal Ayzenshtat、Michal Sheffer 和 Uri Alon。2004。进化和设计网络的超家族。《科学》303, 5663（2004），1538–1542。[32] Ron Milo、Shai Shen-Orr、Shalev Itzkovitz、Nadav Kashtan、Dmitri Chklovskii 和 Uri Alon。2002。网络模体：复杂网络的简单构件。《科学》298, 5594（2002），824–827。[33] Mark Ortmann 和 Ulrik Brandes。2017。有向图和无向图中高效的轨道感知三元组和四元组计数。《应用网络科学》2, 1（2017），1–17。[34] Lawrence Page、Sergey Brin、Rajeev Motwani 和 Terry Winograd。1999。*PageRank* 引用排名：为 *Web* 建立秩序。技术报告。斯坦福信息实验室。[35] Ashwin Paranjape、Austin R Benson 和 Jure Leskovec。2017。时序网络中的模体。载于 *WSDM*。[36] Nataša Pržulj。2007。使用图元度分布进行生物网络比较。《生物信息学》23, 2（2007），e177–e183。[37] Pedro Ribeiro、Pedro Paredes、Miguel EP Silva、David Aparicio 和 Fernando Silva。2021。子图计数综述：概念、算法及其在网络模体和图元中的应用。*CSUR* 54, 2（2021），1–36。[38] Benjamin Schiller、Sven Jager、Kay Hamacher 和 Thorsten Strufe。2015。StreaM：一种用于动态图中模体计数的基于流的算法。载于 *AICoB*。[39] Charalampos E Tsourakakis、Jakub Pachocki 和 Michael Mitzenmacher。2017。可扩展的模体感知图聚类。载于 *WWW*。[40] Sebastian Wernicke。2005。一种更快的网络模体检测算法。载于

WABI。Springer，165–177。[41] Yang Yang、Yuxiao Dong 和 Nitesh V Chawla。2014。使用节点显著性概况预测节点度中心性。《*Scientific Reports*》4，1（2014），1–7。[42] Yanlei Yu、Zhiwu Lu、Jiajun Liu、Guoping Zhao 和 Ji-rong Wen。2019。Rum：使用模体进行网络表示学习。发表于*ICDE*。[43] Daniel Zwillinger 和 Stephen Kokoska。1999。*CRC*标准概率与统计表及公式。CRC出版社。

与 GoT 和 OTA 的比较

我们进一步介绍基于图元转移图（GTG）的刻画方法，并补充说明其与图元轨道转移（GoT）和轨道时间一致性（OTA）的比较。

详细设置：我们的实验在一台配备 3.8 GHz AMD Ryzen 3900x CPU 和 128GB 内存的台式机上进行。我们

表 11： $d_\theta = 2$**时预测未来节点重要性任务的 F1 分数。**

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
度	局部-NR	0.11±0.008	0.20±0.014	0.36±0.092	0.68±0.007	0.27±0.027	0.38±0.017	**0.73**±0.002	0.39
	局部 NPP	0.12±0.013	0.19±0.016	0.40±0.102	0.60±0.011	0.27±0.040	0.36±0.016	0.72±0.003	0.38
	全局-NR	0.52±0.013	0.56±0.013	0.51±0.077	**0.79**±0.004	0.37±0.025	**0.52**±0.021	0.70±0.004	**0.57**
	全局-NPP	0.51±0.010	0.57±0.018	0.51±0.063	0.76±0.005	**0.40**±0.030	**0.52**±0.022	0.70±0.006	**0.57**
	全局基础	0.36±0.014	0.38±0.024	0.44±0.073	0.77±0.008	0.34±0.050	0.51±0.022	0.71±0.003	0.50
	ALL	**0.53**±0.010	**0.58**±0.013	**0.52**±0.043	**0.79**±0.008	0.38±0.041	**0.52**±0.026	0.70±0.005	**0.57**
中介中心性	局部-NR	0.59±0.011	0.88±0.009	0.34±0.063	0.49±0.011	0.34±0.033	0.74±0.011	0.73±0.007	0.59
	局部 NPP	0.58±0.010	0.87±0.006	0.35±0.076	0.45±0.010	0.36±0.073	0.74±0.011	0.73±0.007	0.58
	全局-NR	**0.64**±0.007	**0.90**±0.005	0.48±0.089	**0.62**±0.014	0.38±0.047	**0.75**±0.008	**0.74**±0.006	0.64
	全局-NPP	0.62±0.010	0.89±0.006	0.49±0.037	0.58±0.019	**0.40**±0.034	**0.75**±0.010	**0.74**±0.006	0.64
	全局基础	0.01±0.003	0.27±0.028	0.40±0.079	0.36±0.017	0.25±0.038	0.32±0.013	0.10±0.013	0.24
	ALL	**0.64**±0.007	**0.90**±0.007	**0.53**±0.052	**0.62**±0.016	0.38±0.045	**0.75**±0.010	**0.74**±0.007	**0.65**
接近度	局部-NR	0.49±0.010	0.53±0.010	0.28±0.077	0.69±0.008	0.24±0.034	0.58±0.014	0.75±0.005	0.51
	局部 NPP	0.37±0.015	0.51±0.014	0.31±0.067	0.46±0.010	0.25±0.038	0.43±0.017	0.66±0.006	0.43
	全局-NR	0.84±0.006	0.75±0.007	0.47±0.046	**0.83**±0.008	0.38±0.055	**0.69**±0.024	**0.81**±0.002	0.68
	全局-NPP	0.83±0.008	0.74±0.007	0.52±0.047	0.76±0.008	**0.39**±0.022	0.64±0.019	0.71±0.005	0.66
	全局基础	0.82±0.004	0.70±0.010	0.44±0.074	0.72±0.010	0.33±0.010	0.51±0.015	0.60±0.004	0.59
	ALL	**0.85**±0.008	**0.76**±0.008	**0.53**±0.043	**0.83**±0.007	0.36±0.051	**0.69**±0.022	**0.81**±0.003	**0.69**
PageRank	局部-NR	0.44±0.013	0.15±0.018	0.42±0.069	0.64±0.008	0.25±0.038	0.46±0.016	0.58±0.009	0.42
	局部 NPP	0.41±0.012	0.18±0.017	0.43±0.086	0.39±0.009	0.25±0.040	0.41±0.013	0.53±0.008	0.37
	全局-NR	**0.64**±0.014	0.41±0.015	0.49±0.078	**0.74**±0.006	0.35±0.056	**0.53**±0.019	**0.63**±0.009	0.54
	全局-NPP	**0.64**±0.012	0.43±0.015	**0.55**±0.046	0.65±0.011	**0.38**±0.047	0.52±0.017	0.61±0.007	0.54
	全局基础	0.63±0.010	0.31±0.023	0.41±0.054	0.65±0.007	0.28±0.035	0.49±0.010	0.54±0.008	0.47
	ALL	**0.64**±0.009	**0.44**±0.013	**0.55**±0.035	**0.74**±0.006	0.37±0.030	**0.53**±0.020	**0.63**±0.008	**0.56**

表12：在 $d_\theta = 2$**时预测未来节点重要性任务的准确率。**

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
度	局部-NR	0.71±0.008	0.72±0.006	0.80±0.024	0.62±0.006	**0.77**±0.019	0.66±0.008	0.58±0.003	0.69
	局部 NPP	0.71±0.010	0.72±0.007	0.79±0.031	0.55±0.006	0.75±0.020	0.65±0.016	0.58±0.002	0.68
	全局-NR	**0.76**±0.007	0.77±0.007	**0.82**±0.024	**0.77**±0.004	0.76±0.017	**0.68**±0.014	**0.61**±0.004	**0.74**
	全局-NPP	0.75±0.003	0.77±0.007	0.80±0.023	0.75±0.004	**0.77**±0.019	**0.68**±0.012	**0.61**±0.004	0.73
	全局基础	0.73±0.006	0.74±0.008	0.77±0.028	0.75±0.006	0.74±0.019	0.67±0.011	0.60±0.003	0.72
	ALL	**0.76**±0.008	**0.78**±0.006	**0.82**±0.019	**0.77**±0.005	0.76±0.016	**0.68**±0.014	**0.61**±0.004	**0.74**
中介中心性	局部-NR	0.79±0.005	0.93±0.005	0.78±0.017	0.82±0.006	**0.76**±0.019	**0.86**±0.005	**0.90**±0.003	0.83
	局部 NPP	0.79±0.004	0.92±0.004	0.75±0.028	0.81±0.005	0.65±0.032	**0.86**±0.005	**0.90**±0.003	0.81
	全局-NR	**0.81**±0.004	**0.94**±0.003	0.81±0.017	**0.84**±0.007	0.75±0.016	**0.86**±0.004	**0.90**±0.002	0.84
	全局-NPP	0.80±0.005	**0.94**±0.003	0.79±0.021	0.83±0.007	**0.76**±0.020	**0.86**±0.004	**0.90**±0.003	0.84
	全局基础	0.71±0.005	0.73±0.010	0.75±0.033	0.77±0.007	0.70±0.019	0.67±0.008	0.77±0.004	0.73
	ALL	**0.81**±0.003	**0.94**±0.004	**0.82**±0.018	**0.84**±0.008	0.75±0.019	**0.86**±0.004	**0.90**±0.003	**0.85**
接近度	局部-NR	0.76±0.003	0.76±0.005	0.78±0.027	0.75±0.005	**0.76**±0.015	0.72±0.006	0.78±0.003	0.76
	局部 NPP	0.73±0.003	0.75±0.006	0.75±0.032	0.63±0.007	0.74±0.016	0.65±0.010	0.64±0.004	0.70
	全局-NR	**0.91**±0.003	**0.86**±0.004	0.80±0.018	**0.85**±0.006	0.75±0.019	**0.77**±0.011	**0.82**±0.002	0.82
	全局-NPP	0.90±0.005	0.85±0.002	0.81±0.013	0.80±0.007	**0.76**±0.011	0.73±0.010	0.73±0.003	0.80
	全局基础	0.90±0.003	0.82±0.004	0.76±0.033	0.77±0.008	0.73±0.024	0.66±0.009	0.64±0.003	0.75
	ALL	**0.91**±0.004	**0.86**±0.004	**0.82**±0.020	**0.85**±0.005	0.75±0.017	**0.77**±0.011	**0.82**±0.002	**0.83**
PageRank	局部-NR	0.76±0.007	0.72±0.009	0.80±0.020	0.74±0.006	0.75±0.017	0.66±0.011	0.65±0.006	0.73
	局部 NPP	0.74±0.007	0.72±0.007	0.77±0.023	0.65±0.018	0.73±0.023	0.63±0.005	0.62±0.006	0.69
	全局-NR	**0.81**±0.005	**0.75**±0.007	0.81±0.022	**0.80**±0.005	0.75±0.015	**0.68**±0.010	**0.67**±0.006	**0.75**
	全局-NPP	**0.81**±0.005	0.74±0.007	**0.82**±0.023	0.74±0.011	**0.76**±0.019	0.67±0.008	0.66±0.006	0.74
	全局基础	**0.81**±0.004	0.73±0.008	0.75±0.021	0.74±0.006	0.70±0.025	0.65±0.007	0.59±0.006	0.71
	ALL	**0.81**±0.005	**0.75**±0.004	0.81±0.017	**0.80**±0.006	0.75±0.014	**0.68**±0.010	**0.67**±0.005	**0.75**

表 13：在 $d_\theta = 2$时预测未来节点重要性任务的 AUROC。

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
度	局部-NR	0.69±0.006	0.70±0.006	0.80±0.037	0.67±0.007	0.69±0.020	0.65±0.012	0.58±0.006	0.68
	局部 NPP	0.64±0.005	0.67±0.007	0.75±0.032	0.56±0.007	0.64±0.025	0.63±0.012	0.56±0.004	0.64
	全局-NR	**0.81**±0.004	0.83±0.005	**0.85**±0.037	**0.86**±0.005	0.73±0.026	**0.71**±0.013	**0.65**±0.006	**0.78**
	全局-NPP	0.80±0.002	0.83±0.005	0.83±0.032	0.83±0.005	**0.74**±0.026	**0.71**±0.012	**0.65**±0.005	0.77
	全局基础	0.74±0.004	0.74±0.007	0.82±0.035	0.84±0.006	0.68±0.037	0.68±0.009	0.62±0.005	0.73
	ALL	**0.81**±0.004	**0.84**±0.006	**0.85**±0.036	**0.86**±0.005	0.73±0.025	**0.71**±0.014	**0.65**±0.006	**0.78**
中介中心性	局部-NR	0.84±0.006	0.98±0.002	0.79±0.033	0.80±0.007	0.70±0.029	0.83±0.010	0.82±0.005	0.82
	局部 NPP	0.83±0.006	0.98±0.003	0.73±0.031	0.68±0.008	0.65±0.041	0.82±0.010	0.81±0.006	0.79
	全局-NR	**0.87**±0.005	**0.99**±0.002	0.81±0.030	**0.87**±0.007	0.71±0.032	**0.86**±0.008	**0.86**±0.005	0.85
	全局-NPP	0.85±0.007	0.98±0.002	0.81±0.032	0.84±0.008	**0.72**±0.033	**0.86**±0.008	**0.86**±0.005	0.85
	全局基础	0.62±0.008	0.74±0.013	0.76±0.042	0.74±0.005	0.63±0.038	0.63±0.014	0.60±0.007	0.67
	ALL	**0.87**±0.006	**0.99**±0.001	**0.83**±0.024	**0.87**±0.007	0.71±0.033	**0.86**±0.009	**0.86**±0.005	**0.86**
接近度	局部-NR	0.79±0.005	0.79±0.006	0.76±0.047	0.84±0.006	0.68±0.026	0.76±0.009	0.84±0.002	0.78
	局部 NPP	0.74±0.005	0.78±0.008	0.72±0.042	0.65±0.009	0.61±0.022	0.63±0.007	0.70±0.005	0.69
	全局-NR	**0.97**±0.002	0.93±0.005	0.82±0.035	**0.93**±0.003	0.73±0.024	**0.83**±0.011	0.89±0.002	0.87
	全局-NPP	0.96±0.003	0.92±0.004	**0.83**±0.028	0.88±0.004	0.73±0.026	0.79±0.011	0.81±0.003	0.85
	全局基础	0.95±0.003	0.89±0.005	0.79±0.044	0.85±0.007	0.69±0.035	0.68±0.012	0.70±0.003	0.79
	ALL	**0.97**±0.002	**0.94**±0.005	0.82±0.027	**0.93**±0.004	**0.75**±0.028	**0.83**±0.010	**0.90**±0.001	**0.88**
PageRank	局部-NR	0.77±0.006	0.65±0.010	0.80±0.031	0.81±0.004	0.67±0.011	0.69±0.015	0.70±0.006	0.73
	局部 NPP	0.75±0.004	0.63±0.008	0.77±0.031	0.63±0.010	0.62±0.024	0.64±0.009	0.67±0.006	0.67
	全局-NR	**0.87**±0.005	0.78±0.006	0.84±0.035	**0.88**±0.003	**0.72**±0.029	**0.71**±0.012	**0.73**±0.005	**0.79**
	全局-NPP	0.86±0.005	0.77±0.007	**0.85**±0.028	0.82±0.008	**0.72**±0.018	0.70±0.011	0.71±0.006	0.78
	全局基础	0.86±0.005	0.73±0.010	0.81±0.033	0.81±0.008	0.66±0.026	0.66±0.011	0.62±0.004	0.74
	ALL	**0.87**±0.005	**0.79**±0.007	**0.85**±0.035	**0.88**±0.004	**0.72**±0.031	**0.71**±0.014	**0.73**±0.005	**0.79**

表 14：根据 d_θ（即提取节点输入特征时的节点入度），预测未来节点重要性任务的 F1 分数。

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
度	全部 ($d_\theta = 2$)	0.53±0.010	0.58±0.013	0.52±0.043	**0.19±0.005**	0.38±0.041	0.52±0.026	**0.70±0.005**	0.59
	全部（$d_\theta = 4$）	0.67±0.007	0.78±0.008	0.62±0.062	1.00*	0.49±0.045	**0.89**±0.006	1.00*	0.69
	全部 ($d_\theta = 8$)	**0.82**±0.006	**0.93**±0.006	**0.74**±0.036	1.00*	**0.71**±0.022	1.00*	1.00*	**0.80**
中介中心性	全部 ($d_\theta = 2$)	0.64±0.007	0.90±0.007	0.53±0.052	0.62±0.016	0.38±0.045	0.75±0.010	0.74±0.007	0.65
	全部 ($d_\theta = 4$)	0.72±0.012	0.94±0.007	0.52±0.066	0.75±0.006	0.50±0.042	0.84±0.009	0.84±0.008	0.73
	全部（$d_\theta = 8$）	**0.77**±0.008	**0.97**±0.005	**0.65**±0.045	**0.86**±0.009	**0.69**±0.057	**0.91**±0.009	**0.89**±0.007	**0.82**
接近度	全部 ($d_\theta = 2$)	0.85±0.008	0.76±0.008	0.53±0.043	0.83±0.007	0.36±0.051	0.69±0.022	0.81±0.003	0.69
	全部（$d_\theta = 4$）	0.87±0.008	0.85±0.010	0.55±0.072	0.91±0.006	0.54±0.032	0.85±0.013	0.91±0.004	0.78
	全部（$d_\theta = 8$）	**0.88**±0.007	**0.90**±0.009	**0.65**±0.061	**0.97**±0.004	**0.74**±0.045	**0.95**±0.007	**0.98**±0.003	**0.86**
PageRank	全部（$d_\theta = 2$）	0.64±0.009	0.44±0.013	0.52±0.035	0.74±0.006	0.37±0.030	0.53±0.020	0.63±0.006	0.55
	全部 ($d_\theta = 4$)	0.74±0.008	0.71±0.010	0.62±0.037	0.87±0.006	0.48±0.040	0.79±0.008	0.89±0.003	0.73
	全部 ($d_\theta = 8$)	**0.83**±0.007	**0.85**±0.012	**0.68**±0.049	**0.95**±0.003	**0.72**±0.033	**0.95**±0.006	**0.98**±0.003	**0.85**

* 满足 d_θ 条件的所有节点类别相同，且在所考察的中心性度量方面位于前 20%。

表 15：准确率　　在预测未来节点重要性的任务中，　　其表现取决于 d_θ。

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
度	全部 ($d_\theta = 2$)	0.76±0.008	0.78±0.006	0.82±0.019	**0.77±0.005**	0.76±0.016	0.68±0.014	**0.61±0.004**	0.74
	全部 ($d_\theta = 4$)	0.76±0.006	0.80±0.009	0.83±0.028	1.00*	0.70±0.046	**0.81**±0.009	1.00*	0.78
	全部 ($d_\theta = 8$)	**0.79**±0.006	**0.88**±0.010	**0.86**±0.021	1.00*	**0.72**±0.025	1.00*	1.00*	**0.81**
中介中心性	全部 ($d_\theta = 2$)	0.81±0.003	0.94±0.004	**0.82**±0.018	0.84±0.008	**0.75**±0.019	0.86±0.004	**0.90**±0.003	**0.85**
	全部 ($d_\theta = 4$)	0.81±0.008	0.96±0.004	0.80±0.023	0.82±0.004	0.70±0.023	0.86±0.009	0.89±0.006	0.83
	全部 ($d_\theta = 8$)	0.81±0.004	**0.98**±0.004	**0.82**±0.019	**0.85**±0.009	0.70±0.049	**0.88**±0.010	0.88±0.006	**0.85**
接近度	全部 ($d_\theta = 2$)	0.91±0.004	0.86±0.004	**0.82**±0.020	0.85±0.005	**0.75**±0.017	0.77±0.011	0.82±0.002	0.83
	全部 ($d_\theta = 4$)	0.91±0.005	0.88±0.007	0.80±0.022	0.89±0.007	0.70±0.021	0.80±0.015	0.86±0.006	0.83
	全部（$d_\theta = 8$）	0.91±0.006	**0.89**±0.009	**0.82**±0.020	**0.94**±0.006	0.73±0.046	**0.91**±0.013	**0.95**±0.006	**0.88**
PageRank	全部（$d_\theta = 2$）	0.81±0.005	0.75±0.004	0.81±0.017	0.80±0.006	**0.75**±0.014	0.68±0.010	0.67±0.006	0.75
	全部（$d_\theta = 4$）	0.81±0.006	0.75±0.007	**0.83**±0.017	0.83±0.007	0.68±0.025	0.68±0.011	0.81±0.003	0.77
	全部（$d_\theta = 8$）	**0.83**±0.005	**0.81**±0.012	0.82±0.023	**0.92**±0.004	0.72±0.025	**0.91**±0.011	**0.96**±0.003	**0.85**

* 全部节点 satis　满足条件的 d_θ 具有相同的类别，在所考虑的中心性度量中属于前 20%　　e.

表 16：根据 d_θ 预测未来节点重要性任务的 AUROC。

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
度	全部 ($d_\theta = 2$)	0.81±0.004	0.84±0.005	0.85±0.035	**0.86±0.005**	0.73±0.025	**0.71**±0.014	**0.65±0.005**	0.78
	全部（$d_\theta = 4$）	0.83±0.005	0.87±0.006	0.85±0.036	1.00*	0.72±0.027	0.68±0.018	1.00*	0.79
	全部 ($d_\theta = 8$)	**0.87**±0.007	**0.90**±0.013	**0.88**±0.027	1.00*	**0.78**±0.031	1.00*	1.00*	**0.86**
中介中心性	全部 ($d_\theta = 2$)	0.87±0.005	0.99±0.001	0.83±0.024	0.87±0.007	0.71±0.033	0.86±0.009	0.86±0.005	0.86
	全部 ($d_\theta = 4$)	0.89±0.006	0.99±0.001	0.81±0.040	0.89±0.004	0.73±0.026	0.90±0.007	0.91±0.004	0.87
	全部 ($d_\theta = 8$)	**0.90**±0.003	**1.00**±0.001	**0.84**±0.026	**0.93**±0.006	**0.77**±0.044	**0.94**±0.009	**0.94**±0.006	**0.90**
接近度	全部（$d_\theta = 2$）	0.97±0.002	0.94±0.005	0.84±0.033	0.93±0.004	0.73±0.028	0.83±0.010	0.90±0.002	0.88
	全部 ($d_\theta = 4$)	0.97±0.002	0.95±0.004	0.82±0.027	0.95±0.004	0.75±0.030	0.88±0.012	0.93±0.004	0.89
	全部 ($d_\theta = 8$)	0.97±0.003	**0.96**±0.006	**0.88**±0.024	**0.98**±0.004	**0.79**±0.043	**0.92**±0.016	**0.95**±0.013	**0.92**
PageRank	全部 ($d_\theta = 2$)	0.87±0.005	0.79±0.008	0.85±0.035	0.88±0.004	0.72±0.031	0.71±0.014	**0.73**±0.005	0.79
	全部（$d_\theta = 4$）	0.89±0.006	0.83±0.008	**0.87**±0.018	0.90±0.006	0.71±0.028	0.69±0.009	0.70±0.007	0.80
	全部 ($d_\theta = 8$)	**0.91**±0.005	**0.87**±0.013	**0.87**±0.034	**0.95**±0.009	**0.79**±0.018	**0.73**±0.040	0.71±0.049	**0.83**

* 所有节点均满足条件 d_θ 的节点属于同一类别，且在所考虑的中心性指标中位居前 20%

e.

表 17：预测未来边重要性任务的 F1 分数。

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
边介数	Local-ER ($d_\theta = 2$)	0.68 ± 0.004	0.59 ± 0.011	0.14 ± 0.094	0.74 ± 0.013	**0.41** ± 0.042	0.21 ± 0.038	**0.40** ± 0.013	0.45
	Global-ER（$d_\theta = 2$）	0.69 ± 0.015	0.63 ± 0.041	0.18 ± 0.122	0.79 ± 0.051	0.38 ± 0.060	**0.23** ± 0.058	0.39 ± 0.022	**0.47**
	全局基础 ($d_\theta = 2$)	0.69 ± 0.013	0.51 ± 0.168	0.22 ± 0.132	0.75 ± 0.072	0.37 ± 0.064	0.15 ± 0.116	0.26 ± 0.180	0.42
	全部（$d_\theta = 2$）	**0.71** ± 0.005	**0.68** ± 0.009	**0.25** ± 0.186	**0.84** ± 0.005	0.40 ± 0.062	**0.23** ± 0.060	0.36 ± 0.018	0.50
	全部 ($d_\theta = 2$)	**0.71** ± 0.005	0.68 ± 0.009	0.25 ± 0.186	**0.84** ± 0.005	0.40 ± 0.062	0.23 ± 0.060	0.36 ± 0.018	0.50
	全部 ($d_\theta = 4$)	**0.71** ± 0.007	0.72 ± 0.009	**0.33** ± 0.104	0.77 ± 0.006	**0.43** ± 0.086	0.29 ± 0.071	0.46 ± 0.014	**0.53**
	全部 ($d_\theta = 8$)	0.69 ± 0.004	**0.75** ± 0.009	0.17 ± 0.079	0.72 ± 0.011	0.39 ± 0.052	**0.31** ± 0.055	**0.53** ± 0.023	0.52

表 18：预测未来边重要性任务的准确率。

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
边介数	Local-ER ($d_\theta = 2$)	0.66 ± 0.003	0.72 ± 0.005	0.87 ± 0.041	0.75 ± 0.012	**0.70** ± 0.030	0.91 ± 0.008	0.91 ± 0.003	0.78
	Global-ER ($d_\theta = 2$)	**0.68** ± 0.020	**0.75** ± 0.023	**0.88** ± 0.041	**0.81** ± 0.058	**0.70** ± 0.035	0.91 ± 0.009	0.91 ± 0.003	**0.81**
	全局基础 ($d_\theta = 2$)	0.66 ± 0.036	0.72 ± 0.046	0.87 ± 0.040	0.79 ± 0.052	0.67 ± 0.053	0.91 ± 0.009	0.91 ± 0.008	0.79
	全部（$d_\theta = 2$）	0.67 ± 0.037	0.73 ± 0.047	0.87 ± 0.042	**0.81** ± 0.055	0.68 ± 0.052	0.91 ± 0.009	0.91 ± 0.007	0.80
	全部 ($d_\theta = 2$)	0.67 ± 0.037	0.73 ± 0.005	0.87 ± 0.042	0.81 ± 0.055	0.68 ± 0.052	0.91 ± 0.009	**0.91** ± 0.007	0.80
	全部 ($d_\theta = 4$)	0.73 ± 0.006	0.79 ± 0.005	0.86 ± 0.050	0.78 ± 0.024	0.78 ± 0.024	0.93 ± 0.007	0.90 ± 0.004	0.82
	全部 ($d_\theta = 8$)	**0.75** ± 0.003	**0.81** ± 0.005	**0.88** ± 0.046	**0.82** ± 0.017	**0.82** ± 0.017	**0.94** ± 0.005	**0.91** ± 0.002	**0.85**

表 19：预测未来边重要性任务的 AUROC。

中心性	特征	引文网络		电子邮件/消息网络			在线问答网络		平均值
		HepPh	Hepth	Email-EU	Email-Enron	Message-College	Mathoverflow	Askubuntu	
边介数	Local-ER ($d_\theta = 2$)	0.71 ± 0.003	0.77 ± 0.007	**0.64** ± 0.080	0.82 ± 0.009	**0.68** ± 0.035	**0.85** ± 0.013	**0.86** ± 0.006	0.76
	Global-ER（$d_\theta = 2$）	**0.74** ± 0.027	**0.80** ± 0.033	0.63 ± 0.092	0.88 ± 0.058	**0.68** ± 0.032	**0.85** ± 0.015	**0.86** ± 0.007	**0.78**
	全局基础 ($d_\theta = 2$)	0.71 ± 0.053	0.77 ± 0.055	**0.64** ± 0.093	0.87 ± 0.049	0.65 ± 0.058	0.73 ± 0.164	0.76 ± 0.150	0.73
	全部 ($d_\theta = 2$)	0.72 ± 0.054	0.79 ± 0.057	**0.64** ± 0.099	**0.89** ± 0.051	0.66 ± 0.056	0.76 ± 0.150	0.78 ± 0.139	0.75
	全部 ($d_\theta = 2$)	0.72 ± 0.054	0.79 ± 0.057	0.64 ± 0.099	0.89 ± 0.051	0.66 ± 0.056	0.76 ± 0.150	0.78 ± 0.139	0.75
	全部 ($d_\theta = 4$)	0.80 ± 0.004	0.87 ± 0.004	**0.75** ± 0.050	**0.94** ± 0.002	0.74 ± 0.026	0.89 ± 0.016	0.89 ± 0.009	0.84
	全部 ($d_\theta = 8$)	**0.82** ± 0.005	**0.89** ± 0.005	0.70 ± 0.046	0.93 ± 0.002	**0.79** ± 0.021	**0.90** ± 0.011	**0.90** ± 0.008	**0.85**