流式模型中子图查找问题的空间复杂度二分性

Yu-Sheng Shih 国立台湾大学，台湾

Meng-TsungTsai 中央研究院，台湾

Yen-Chu Tsai 国立台湾大学，台湾

Ying-Sian Wu 国立台湾大学，台湾

arXiv:2602.08002v1 [cs.DS] 8 Feb 2026

摘要

我们研究流式模型中标准子图查找问题四种变体的空间复杂度。具体而言，给定一个含有 n 个顶点的输入图和一个固定大小的模式图，我们考虑两种情形：无向简单图，记作 G 和 H；以及定向图，记作 $\vec{G}$ 和 $\vec{H}$。根据具体情形，任务是判断 G 是否包含 H 作为子图或导出子图，或者判断 $\vec{G}$ 是否包含 $\vec{H}$ 作为子图或导出子图。分别用 Sub(H)、IndSub(H)、Sub($\vec{H}$) 和 IndSub($\vec{H}$) 表示这四种变体。

如果一个定向图存在一种二部划分，使得每条弧都从其中一部指向另一部，则称其为良定向；如果一个顶点的入度和出度均不为零，则称其为非良定向。对于每种变体，我们都得到了一个完整的二分定理，简要概括如下。

- 当且仅当 H 是二分图时，Sub(H)才能由 $\tilde{O}$遍、 $n^{2-\Omega(1)}$空间的算法求解。当且仅当 $H \in \{P_3, P_4,$
- co-P_3}时，IndSub(H)才能由 $\tilde{O}$遍、 $n^{2-\Omega(1)}$空间的算法求解。当且仅当 H 的每个连通$\vec{}$分量要么是
- 良定向二分图，要么是至多包含一个非良定向顶点的树时，Sub($\vec{H}$)才能由单遍 $n^{2-\Omega(1)}$空间的算法求解。当且仅当底层无向简单图 H 是 co-P_3时，IndSub($\vec{H}$)才能由 $\tilde{O}$遍、 $n^{2-\Omega(1)}$空间的算法求解。
-

2012 年 ACM 学科分类 计算理论 → 流式计算、次线性与近线性时间算法

关键词与短语 图兰数、通信复杂度、多方集合不交、稀疏证书、余图

资助*Meng-Tsung Tsai*: 本研究部分经费由国家科学及技术委员会根据合同 NSTC 114-2221-E-001-023 资助。

1 引言

对于固定的无向简单图 H 和整数 n，图兰数 $ex(n, H)$ 是不包含 H 作为子图的 n 顶点无向图中边数的最大值。因此，若要在 n 顶点图 G 中寻找 H 的副本，只需保留 G 的任意一个至少含有 $ex(n, H)$ 条边的子图，因为这样的子图必定包含 H。由此可立即得到一种节省空间的单遍流式算法：只要 $ex(n, H)$ 足够小，就能找到 H。同一思路也可扩展到寻找定向图[1] $\vec{H}$。在相关文献中，Feigenbaum等人 [10] 曾指出，这种方法可用于判断一个图的围长是否大于 k。

Erdős–Stone定理 [9] 表明，$ex(n, H) = o(n^2)$ 当且仅当 H 是二分图。Alon、Krivelevich和Sudakov [1], 进一步加强了这一界，他们证明：对于每个二分图 $H = (V_1 \sqcup V_2, E)$，令 $\Delta'(H)$ 表示 V_1 中顶点的最大度与 V_2中顶点的最大度的较小值，则有 $ex(n, H) = O(n^{2-1/\Delta'(H)})$。

对于定向图 $\vec{H}$，Valadkhan [27] 证明了 $ex(n, \vec{H}) = o(n^2)$ 当且仅当 $\vec{H}$ 是二分图且良定向，即在 H 的某个二部划分下，每条弧都从一个部定向到另一个部。我们将Alon等人的证明推广到良定向情形，并在推论 11 中证明，对于每个良定向二分图 $\vec{H}$，其中 H 是 H的底层无向图，都有 $ex(n, \vec{H}) = O(n^{2-1/\Delta'(H)})$。

因此，应用上述朴素算法，在一个含有 n 个顶点的图中寻找二分图无向简单图 H 或良定向二分图 $\vec{H}$ 需要 $n^{2-\Omega(1)}$ 空间，而寻找非二分图 H、定向非二分图 H 或非良定向二分图 $\vec{H}$ 则需要 $O(n^2)$ 空间。自然会产生这样的问题：

> **问题：** 在一个含有 n 个顶点的图中寻找固定图 H（分别为 $\vec{H}$）G（分别为 $\vec{G}$），是否必然需要 $\Omega(ex(n, H))$（分别为 $\Omega(ex(n, \vec{H}))$）空间，还是可以设计出更巧妙的算法?

在回答上述问题之前，我们首先在问题 1中正式定义我们要研究的问题，随后定义我们的计算模型。本文中，所有图均假设为简单图；对于给定的图是否为定向图，我们将明确说明。

▶**问题 1**（子图查找）。

- **输入：** 一个含 n 个顶点的图和一个固定大小的模式图：两者要么都是无向图，$\vec{\ }$ $\vec{\ }$分别记为 G 和 H；要么都是定向图，分别记为 G 和 H。模式图至少有三个顶点，且可以是非连通的。
- **问题：** 根据输入类型，判断 G 是否包含 H 作为子图，$\vec{\ }$ $\vec{\ }$或作为导出子图；或者判断 G 是否包含 H 作为子图或作为导出子图。如果存在这样的副本，则输出 G（分别为 $\vec{G}$）中构成 H（分别为 $\vec{H}$）副本的相应顶点集；否则，报告不存在这样的副本。

分别用 Sub(H)、IndSub(H)、Sub($\vec{H}$)、IndSub($\vec{H}$) 表示这四种变体。

▶ **定义 2**（仅插入图流模型 [22]）。 输入图的边集以任意顺序排列为数据流，称为输入数据流。算法预先得知有 n 个顶点，并可从开头扫描数据流直至

[1] 定向图是一种有向图，其中任意一对顶点之间至多有一条弧。

末尾 p 遍，且不能反向读取。该算法仅有 S 空间（以比特计），通常远小于输入规模，因此必须丢弃对完成任务并非必要的信息，例如 $Sub(H)$。我们将此类算法称为一种 p遍次 S空间流式算法。由于数据流仅由边插入构成，该模型称为仅插入流式模型。

我们在表 1 中总结了子图查找问题四种变体各自的二分结果。这里，P_ℓ 表示具有 ℓ 个顶点的路径图，co-P_ℓ 表示$\vec{}$ $\vec{}$ $\vec{}$的补图 P_ℓ。“H 是 WO”表示 H 是良定向的。如果底层图 H ∋NWOC$_\ell$ 存在一个连通分量，其中包含 H 个 ℓ 非良定向（NWO）$\vec{}$顶点，则记为 H；NWO 顶点是指入度和出度均非零的$\vec{}$的顶点。若 H ∋NWOC$_{\ell+C}$ 且 H 至少包含一个环，则记为

图兰数 $ex(n,\ H)$ 实际上给出了SUB(H) 问题空间复杂度的一个二分结果；也就是说，存在一个 $\tilde{O}$(1) 遍次、 $n^{2-\Omega(1)}$ 空间的算法来求解SUB(H)，当且仅当 H 是二分图。该结论在定理 3 中得到证明。

对于诱导子图查找问题 IndSub(H) 和 IndSub($\vec{H}$)， n顶点完全图中除了团之外不含其他诱导子图，因此朴素方法可能需要保留整个输入图。不过，也有一些 H、$\tilde{}$的例子，即 P_3、 P_4 和 co-P_3，可以通过更巧妙的算法，使用 $O(n)$ 空间和 $\tilde{O}$(1) 遍次算法来处理；其余所有图 H 则要求任何 $\tilde{O}$(1) 遍次算法都使用 Ω(n^2) 空间。这一点由定理 5 和定理 8 证明。

	H（分别为 $\vec{H}$）	$\tilde{O}$(1) 遍次算法。	单遍算法。	参考文献
Sub(H)	如果 H 是非二分图	$\tilde{\Omega}(n^2)$		定理 3
	否则		$\tilde{O}(n^{2-1/\Delta'(H)})$	定理 3
诱导子图(H)	if $H \notin \{P_3, P_4, co\text{-}P_3\}$	$\tilde{\Omega}(n^2)$		定理 5
	否则	$\tilde{O}(n)$		定理 5
SUB($\vec{H}$)	如果 H 是非二分图	$\tilde{\Omega}(n^2)$		推论 33
	$\tilde{}$否则如果 H 是 WO		$\tilde{O}(n^{2-1/\Delta'(H)})$	推论 33
	否则如果 $\vec{H}$ ∋NWOC$_2$		$\Omega(n^2)$	定理 7
	否则如果 $\vec{H}$ ∋NWOC$_{1+C}$		$\Omega(n^2)$	定理 7
	否则如果 H 是森林		$\tilde{O}(n)$	定理 7
	else		$\tilde{O}(n^{2-1/\Delta'(H)})$	推论 45
诱导子图($\vec{H}$)	如果 $H \neq co\text{-}P_3$	$\tilde{\Omega}(n^2)$		定理 8
	否则	$\tilde{O}(n)$		定理 8

表 1 问题四种变体的二分结果1。

对于 Sub($\vec{H}$)，存在无穷多个定向图 $\vec{H}$，其图兰数满足$ex(n, \vec{H}) = \Omega(n^2)$，但寻找 $\vec{H}$ 的一个副本可由一个 $\tilde{O}$(1) 遍次 $\tilde{O}(n^{2-\Omega(1)})$ 空间算法完成。我们的二分结果如下：存在用于 Sub($\vec{H}$) 的单遍 $n^{2-\Omega(1)}$ 空间算法，当且仅当 H 的每个连通分量要么是 WO，要么是恰好包含一个 NWO 顶点的树。注意，若 H 的某个连通分量是恰好含有一个 NWO 顶点的树，则 $ex(n, \vec{H}) = \Omega(n^2)$。这为上述问题给出了一个无穷例外族。特别地，若 H 是一个森林，且其每个连通分量至多包含一个$\tilde{}$NWO 顶点，则存在一个基于新型稀疏证书的单遍 $O(n)$ 空间算法；我们称该证书为小森林保持证书，它由 $O(|H|)$ 个 $(c|H|)$-核组成，这些核能保持以指定顶点为根的任意指定森林。这是回答我们问题的较为巧妙的算法之一。

1.1 相关工作

在文献中，Sub(C_ℓ) 的空间复杂度对于 $\ell \in \{3, 4\}$ 已得到充分理解。Braverman 与 Ostrovsky 在 [5] 中证明，对任意 $O(1)$ 遍次算法，Sub(C_3) 需要 $\Omega(ex(n, C_3))$ 空间；McGregor 与 Vorotnikova 在 [23] 中证明，对任意 $O(1)$ 遍次算法，Sub(C_4) 需要 $\Omega(ex(n, C_4))$ 空间。我们的定理 3 和定理 4 推广了这些结果，证明：对每个固定整数 $\ell \geq 3$，对任意 $O(1)$ 遍次算法，Sub(C_ℓ) 需要 $\Omega(ex(n, C_\ell))$ 空间。

Oostveen 和 van Leeuwen [25] 研究了基础图问题流式复杂度的参数化方面，重点关注如何在算法上利用有关输入的额外结构信息。具体而言，他们探讨了当存在一个较小的 H-自由调节子，即一个顶点集 X 满足 $G - X$ 是 H-自由时，它会如何影响流式复杂度。在算法中利用这一假设时，必须首先验证所给集合 X 确实能使剩余部分成为 H-自由图。因此，我们关于 诱导子图(H) 乃至诱导子图($\vec{H}$) 的结果，可视为对他们提出的基于调节子的视角所需预处理原语的支撑。

鉴于即使寻找 H 的单个副本也很困难，现有关于在流式图 G 中统计 H 出现次数的研究，通常假设 H 出现频率足够高，以便将空间复杂度降低到可接受的水平；例如，Kane 等人 [19] 以及 Fichtenberger 和 Peng [11]，的工作适用于任意固定图 H。此外，大量研究针对特定的 H 给出了更强的界。

Halldórsson 等人和 Braverman 等人 [13, 4] 研究了一个承诺问题：判定输入图的团数至少为 r 还是至多为 s。这可以看作是在承诺条件下的团检测问题：判定图中是否包含 K_r 的个副本，还是不包含 K_{s+1}的任何副本。他们证明，对于这项任务，任何 $O(1)$遍流式算法都需要 $\Omega(n^2)$ 的空间。

1.2 概述

我们将 $R^p_{2/3}$(Sub(H)) 定义为：成功概率至少为 $2/3$的随机化 p遍流式算法求解 Sub(H) 所需的最小空间复杂度；将 D^p(Sub(H)) 定义为：确定性 p遍流式算法求解 Sub(H) 所需的最小空间复杂度。其他 3 类子图问题也同理。我们的下界总是针对随机算法给出。请注意，尽管问题 1 要求我们在模式图存在时找出其一个副本，但由于我们所用通信博弈的性质，即使是仅检测模式图是否存在的更弱算法，也同样适用这些下界。出于同样的原因，尽管表 1 中我们只给出了 $O(1)$遍和单遍算法的界，事实上，即使 p 关于 n 是多项式级的，我们针对多遍（p遍）算法的下界仍然成立。

1.2.1 无向情形：Sub(H)

我们首先研究无向情形。我们通过从通信博弈多重不相交性进行归约，证明对于所有非二分图H， $R^p_{2/3}$(Sub(H)) $= \Omega(n^2/p)$，由此得到 Sub(H) 的二分定理，如下所示，正如图兰数的情形一样。

▶**定理 3.** 对于固定图 H，*Sub*(H) 满足以下二分定理：

- 如果 H 是非二分图，那么 $R^p_{2/3}(\textit{Sub}(H)) = \Omega(n^2/p)$。
- 否则， $D^1(\textit{Sub}(H)) = O(n^{2-1/\Delta'(H)})$。

我们指出，对于少数非二分图 H，特别是所有奇环，只要$p = O(n)$，就可以基于 p遍次 $\tilde{O}(n^2/p)$空间算法；该算法采用了 [2] 中的颜色编码思想以及 [26]中的着色结果。

接下来，我们研究二分图的下界。对于某些特定类别的 H，我们确实有 $R^p_{2/3}(\mathsf{Sub}(H)) = \Omega(\mathsf{ex}(n, H)/p)$；而一般而言，只要 H 的某个连通分量的连通度为 > 2，我们就有单遍下界 $R^1_{2/3}(\mathsf{Sub}(H)) = \Omega(\mathsf{ex}(n, H))$。

▶**定理 4**。设 H 是一个固定的二分图。

- 如果 H 是一个偶圈、一个非匹配的森林，或者其某个连通分量的直径小于4，那么 $R^p_{2/3}(\mathit{Sub}(H)) = \Omega(\mathsf{ex}(n, H)/p)$。如果 H 的某个连通分量的连通度为 > 2，那么 $R^1_{2/3}(\mathit{Sub}(H)) = \Omega(\mathsf{ex}(n, H))$。

1.2.2 无向诱导情形：Indsub(H)

接下来，我们研究子图查找问题的导出变体。借助定理3 以及从通信博弈集合不相交问题得到的进一步归约，我们发现，对于除少数例外图 H 之外的所有图，$R^p_{2/3}(\textsc{Indsub}(H)) = \Omega(n^2/p)$，这符合我们的直觉：导出检测应该要困难得多。此外，这些例外情形存在只需 $O(1)$ 遍且占用 $O(n)$ 空间的算法，其主要依据文献 [24], 所提出的方法；我们采用了其中关于余图结构的观察，但在流式模型下使用了不同的算法策略。

▶**定理 5**。固定 H 的 $\mathit{Indsub}(H)$ 满足以下二分定理：

- 若 H 是 P_3、共-P_3 或 P_4，则对某个 $q = O(1)$，有 $D^q(\mathit{Indsub}(H)) = O(n)$。否则，$R^p_{2/3}(\mathit{Indsub}(H)) = \Omega(n^2/p)$。

1.2.3 定向情形：Sub$(\vec{H})$

接下来我们研究定向情形。作为定理 3 的一个简单应用，我们可以看到，任何非二分图 $\vec{H}$ 都有 $R^p_{2/3}(\mathsf{Sub}(\vec{H})) = \Omega(n^2/p)$。另一方面，WO 二分图 $\vec{H}$有 $\mathsf{ex}(n, \vec{H}) = O(n^{2-1/\Delta'(H)})$，因此我们立即得到 $D^1(\mathsf{Sub}(\vec{H})) = \tilde{O}(n^{2-1/\Delta'(H)})$。到目前为止，情况与定理 3 中类似。

有趣的情况是 NWO 二分图。尽管它们的 $\mathsf{ex}(n, \vec{H}) = \Theta(n^2)$，全都如此，但以下结果表明，并非所有这类 H 都是困难实例。事实上，有些 H 满足$R^p_{2/3}(\mathsf{Sub}(\vec{H})) = \Omega(n^2/p)$，这意味着 $\vec{H}$ 的非良定向决定了空间复杂度；但据我们所知，当底层 H 是森林时，H 可以使用常数遍次和 $O(\mathsf{ex}(n, H)) = O(n)$ 空间实现高效算法，这意味着在这种情况下，无向森林结构反而起主导作用。以下总结了这些发现。

▶**定理 6**。设 H 为 *NWO* 二分图。

- 如果 H 是偶圈或完全二分图，则 $R^p_{2/3}(\mathit{Sub}(\vec{H})) = \Omega(n^2/p)$。如果 H 是森林，且 r 是其连通分量中的最大半径，则 $D^{2r}(\mathit{Sub}(\vec{H})) = \tilde{O}(n)$。

尽管我们相信，任何具有 NWO 连通分量的二分图 H 都应是一个多遍困难实例，其$1+c$ 空间下界为 $R^p_{2/3}(\textsc{Sub}(\vec{H})) = \Omega(n^2/p)$，但这一一般性的多遍下界仍未解决。尽管如此，对于单遍算法，我们已知的要多得多。根据定理 $D^1(\textsc{Sub}(\vec{H})) = \Omega(n^2)$，只要 $\vec{H}$ 包含一个 NWO 连通分量$_2$ 或 NWO 连通分量$_{1+c}$，就可以得出这一结论，如定理 7的第一部分所示。

剩下的情况是，H 的每个连通分量要么是 WO，要么是一棵恰好包含一个 NWO 顶点的树。在这种情况下，如果 H 是森林，那么我们就有一个单遍 $O(n)$-空间算法，如定理7的第二部分所示。这立即推出了定理7的第三部分，因为 $\vec{H}$ 中的 NWO 树只需 $\tilde{O}(n)$ 空间，而 WO 部分可以用朴素方法处理。

▶**定理 7**。设 H 是一个 *NWO* 二分图。

- 如果 $\vec{H} \ni \mathit{NWOC}_2$ 或 $\vec{H} \ni \mathit{NWOC}_{1+c}$，那么 $R^1_{2/3}(\mathit{Sub}(\vec{H})) = \Omega(n^2)$。否则，如果 $\vec{H}$
- 是森林，那么 $R^1_{2/3}(\mathit{Sub}(\vec{H})) = \tilde{O}(n)$。否则，$H$ 的每个连通分量要么是 *WO*，要么是
- 一棵恰有一个 *NWO* 顶点的树，并且 $R^1(\mathit{Sub}(\vec{H})) = \tilde{O}(n^{2-1/\Delta'(H)})$。

1.2.4 有向诱导情形：诱导子图$(\vec{H})$

最后，我们将结果扩展到有向诱导子图问题。通过一个简单的归约，H 在 H 是定理5中的困难实例时，仍然是困难实例；而从集合不相交问题可知，H 即使对于 $H = P_3$，P_4也仍是困难实例。因此，唯一的例外情形是 co-P_3，此时存在高效算法，与定理5中的情况类似。

▶ **定理 8.** $\mathit{Indsub}(\vec{H})$ 对于固定的 $\vec{H}$，满足以下二分定理：

- 若 $\vec{H}$ 是一个共-P_3，则存在一个 $p = \tilde{O}(1)$ 使得 $D^p(\mathit{Indsub}(\vec{H})) = \tilde{O}(n)$。
- 否则，$R^p_{2/3}(\mathit{IndSub}(\vec{H})) = \Omega(n^2/p)$。

1.3 论文结构

第 2 节介绍极值图论和通信复杂度中的记号与背景。第 3 节研究 Sub(H)，并证明定理 3 和定理 4。第 4 节转向诱导情形，并证明关于 IndSub(H) 的定理 5。第 5 节考察有向变体 Sub($\vec{H}$)，并证明定理 6 和 7。第 6 节分析有向诱导子图检测 IndSub($\vec{H}$)，并证明定理 8。最后，我们在第 7 节提出若干开放问题。

2 定义与预备知识

2.1 图论

我们的记号大多是标准的。我们将讨论无向图和定向图，并假设它们都是简单图。对于无向图 G，一条边记作顶点的无序对 $\{u, v\}$。另一方面，对于定向图，我们使用向量记号 $\vec{H}$；从顶点 u 指向顶点 v 的一条弧记作有序对 (u, v)。每当给定一个定向图H 时，H 表示其底层无向图；每当我们提到 H中的一个环 C_ℓ 时，指的是 H 中一个边的定向与 H一致的环 C_ℓ。下面给出关于定向的定义。

▶**定义 9** (良定向)。设 $\vec{H}$ 是一个定向二部图。如果存在一种将 H 划分为两部 $A(H)$、$B(H)$ 的二部划分，使得所有弧都从 $A(H)$ 指向 $B(H)$，则称 H 为良定向；否则称为非良定向。

此外，在定向图 H 中，如果顶点在 H 中的入度或在 H 中的出度为 0，则称其在 H 中是良定向的；否则称其在 H 中是非良定向的。我们始终用 *WO* 表示良定向，用 *NWO* 表示非良定向。

给定子图查找问题中的模式图，我们的空间复杂度界与其图兰数密切相关，定义及性质如下。

▶ **定义 10**（图兰数）。设 H 为一个固定图，则图兰数 $\mathrm{ex}(n, H)$ 是一个有 n 个顶点且不含 H 的图所能拥有的最大边数。同理，设 H 为一个固定定向图，则图兰数 $\mathrm{ex}(n, \vec{H})$ 是一个有 n 个顶点且不含 H 的定向图所能拥有的最大边数。

▶**定理**（Erdős-Stone [9]）。设 H 为一个色数为 $\chi(H)$ 的固定图，则

$$\mathrm{ex}(n, H) = \left(\frac{\chi(H)-2}{\chi(H)-1} + o(1)\right)\binom{n}{2}.$$

▶**定理**（Valadkhan [27]）。设 $\vec{H}$ 为一个固定定向图。设 $z(H)$ 为满足以下条件的最小整数 k：$\vec{H}$ 可同态映射到任意一个具有 k 个顶点的竞赛图；否则，$z(H) = \infty$ 则

$$\mathrm{ex}(n, \vec{H}) = \left(\frac{z(H)-2}{z(H)-1} + o(1)\right)\binom{n}{2}.$$

特别地，$\mathrm{ex}(n, H) = o(n^2)$ 当且仅当 H 是二分图。此外，$z(\vec{H}) = 2$ 当且仅当 $\vec{H}$ 是二分图且 WO，此时 $\mathrm{ex}(n, \vec{H}) = o(n^2)$。事实上，对于二分图 H，$\mathrm{ex}(n, H)$ 与 n^2 之间存在多项式间隔，如下所示。

▶ **定理** (Alon,Krivelevich,Sudakov [1])。设 H 是一个二分图，其部为 $A(H)$ $B(H)$，并且

$$\Delta'(H) = \min\left\{\max_{v\in A(H)} \deg(v), \max_{v\in B(H)} \deg(v)\right\},$$

则 $\mathrm{ex}(n, H) = O(n^{2-1/\Delta'(H)})$。

上述结果可由依赖随机选择证明，并且该证明很容易推广到适用于 WO 二分图 $\vec{H}$ 的 $\mathrm{ex}(n, \vec{H})$。

▶ **推论 11**。设 $\vec{H}$ 是关于二部划分 $V(H) = A(H) \sqcup B(H)$ 的 *WO*，并且

$$\Delta'(H) = \min\left\{\max_{v\in A(H)} \deg_H(v), \max_{v\in B(H)} \deg_H(v)\right\},$$

则 $\mathrm{ex}(n, \vec{H}) = O(n^{2-1/\Delta'(H)})$。

2.2 通信复杂度

我们的流式下界是通过从标准通信问题进行归约来证明的。具体而言，我们将数据流划分给各参与者，并在他们之间传递算法的内存状态，从而用一个 p 遍次的流式算法模拟任意 p 轮通信协议。因此，底层问题的通信成本下界意味着流式算法的空间下界。

我们将以下经典问题及其已知下界作为归约中的基本问题：Index、集合不相交问题，以及多重不相交性（消息传递模型中的 k 方版本）。第一个问题仅适用于推导单遍流式下界，而其他问题则可导出多遍下界。

2.2.1 INDEX 问题。

在 INDEX$_n$ 通信问题中，Alice 收到一个字符串 $x \in \{0,1\}^n$，Bob 收到一个索引 $i \in [n]$；Alice 向 Bob 发送一条消息，Bob 必须输出 x_i（等价地，Alice 持有一个集合 $T \subseteq [n]$，Bob 必须判断 $i \in T$）。一项经典结果表明，对于 INDEX$_n$，任何正确率至少为 $_{2\ 3}$ 的随机单向协议都需要 $\Omega(n)$ 通信量，即使使用共享随机性也是如此 [20]。另请参见 [14]，其中给出了初等证明；背景介绍见 [21]。

2.2.2（两方）集合不相交问题

在两方问题不相交性问题$_n$中，Alice 和 Bob 收到由集合表示的 x，$y \in \{0,1\}^n A$，$B \subseteq [n]$。目标是当且仅当 $A \cap B \neq \emptyset$ 时输出 1。该问题的任何有界错误随机协议都需要 $\Omega(n)$ 比特的通信；Kalyanasundaram 和 Schnitger 证明了这一通信下界 [18]（另见他们较早的会议版本 [17]）。

2.2.3 多重不相交性。

在 k 方（手持输入）版本中，参与者 $t \in [k]$ 收到 $x^{(t)} \in \{0,1\}^n$，其由集合 $S_t \subseteq [n]$ 表示；目标是当且仅当 $\bigcap_{t=1}^{k} S_t \neq \emptyset$ 时输出 1。我们主要使用该问题的承诺版本，记为多方不相交性问题$_{n,\ k}$，其中输入集合被承诺为属于以下两种情形之一。任务是区分这两种情形，在 YES 情形下输出 1，在 NO 情形下输出 0。

(是) $\bigcap_{t=1}^{k} S_t = \{i^\star\}$ 对于某个 $i^\star \in [n]$。
(否) $S_1, \cdots S_k$ 两两不相交

在黑板（广播）模型中，每个写下的比特都对所有参与者可见，通信成本是写下的比特总数。Chakrabarti–Khot–Sun [6]证明了一般黑板协议的 $\Omega\left(_{k}{}^{n}{}_{\text{对数}\ k}\right)$ 下界，并为受限的单向（预先确定顺序）模型证明了最优的 $\Omega\binom{n}{k}$下界。Gronemeier [12]后来将其加强为一般黑板协议的有界错误随机化下界 $\Omega\binom{n}{k}$。当只有 $k = O(1)$ 个参与者时，这给出了 $\Omega(n)$ 的空间下界。由于通信到流处理的归约依赖于弱于黑板模型的协议，因此这些黑板下界足以推出我们的流处理空间下界。

3 无向简单图

3.1 非二分图：困难实例

我们首先研究无向子图查找问题。Erdős-Stone定理[9] 表明，任意二分图与非二分图的图兰数之间存在多项式间隔。现在我们证明定理 3，该定理表明相应子图查找流算法的空间复杂度之间也存在类似的间隔。

二分图的上界 H 由朴素算法得出，因为我们有$ex(n,\ H) = O(n^{2-1/\Delta'(H)})$，如 [1]所示。另一方面，非二分图的下界$H$ 通过从 多重不相交性归约得到。为了构造归约图，H 中的最短奇环是关键结构；为此，我们定义了最短奇环分量：H。

▶**定义 12** (SOCC)。设 H 为非二分图，ℓ 为 H中最短奇环的长度。若 $\{v_1,\ldots,v_k\} \subseteq V(H)$ 构成 H 的一个最短奇环分量（*SOCC*），则

- 对于所有 $x,y \in \{v_1,\ldots,v_k\}$，都存在一条从 x 到 y的路径 $e_1e_2\ldots e_j$，其中每个 $e_i \in E(H)$ 都属于 C_ℓ的某个副本；并且不存在 $v_{k+1} \notin \{v_1,\ldots,v_k\}$，使得上述性质对 $\{v_1,\ldots,v_{k+1}\}$成立。

在这种情况下，我们将 $H[\{v_1,\ldots,v_k\}]$ 定义为一个 *SOCC*。此外，我们将 *SOCC* 的大小定义为其中边的数量。

请注意，对于每个最短奇圈，SOCC 要么包含它，要么与它不相交。现在，令 ℓ 为 H中的最短奇圈长度。以下性质说明了为什么˜最短奇圈是值得考虑的自然结构。若我们将图 H 定义为

$$V_1 = \{v^{(1)} : v \in V(H)\}, \ldots, V_n = \{v^{(n)} : v \in V(H)\},$$
$$V(\tilde{H}) = V_1 \sqcup \ldots \sqcup V_n, \quad E(\tilde{H}) = \left\{\{u^{(i)}, v^{(j)}\} : \{u,v\} \in E(H), i,j \in [n]\right\},$$

˜那么，对于由以下方式给出的 H 中的任意 C_ℓ

$$C : v_1^{(i_1)} - v_2^{(i_2)} - \ldots - v_\ell^{(i_\ell)} - v_1^{(i_1)}, \quad i_1,\ldots,i_\ell \in [n], \quad v_1,\ldots,v_\ell \in V(H),$$

回到 H，$v_1 - v_2 - \ldots - v_\ell - v_1$ 将是一个长度为 ℓ的闭游走，因此由于 ℓ 是奇数，它包含一个奇圈，所以根据 ℓ的最小性，它本身也必定是一个 C_ℓ。请注意，如果我们只考虑 H中的最短圈，情况可能就不一样了，因为我们需要 ℓ 是奇数。

这有助于证明定理 3的以下下界。虽然没有直接使用上述H 结论，但我们利用了连接不同顶点副本对的交叉边这一思路。我们将通过复制 H 中的一个最大 SOCC 来构建归约图，然后将 多重不相交性问题的输入编码为一些交叉边。接着，通过分析归约图中的 SOCC，利用上述性质，我们可以证明，如果多重不相交性问题的输入是一个否定实例，那么就不会出现 H 的任何副本。

▶ **命题13**。设 H 是一个固定的非二分图，则 $R^p_{2/3}(\mathit{Sub}(H)) = \Omega(n^2/p)$。

证明。令 ℓ 为 H 中最短奇圈的长度。令 A_1、A_2、$\ldots A_k \subseteq V$（H）为两两不相交的顶点集，使得 $H[A_1]$、.....、$H[A_k]$ 是 H 中的所有 SOCC，并且不失一般性地假设 $H[A_1]$ 是最大的 SOCC。我们构造 n 个 A_1 的副本 $A_1^{(1)}$、.....、$A_1^{(n)}$，即对于 $1 \le j \le n$，$A_1^{(j)} = \{x^{(j)} : x \in A_1\}$。此外，我们固定一个顶点 $v \in A_1$，并定义

$$\{u_1,\ldots,u_r\} = \{x \in A_1 : \{v,x\} \in E(H) \text{ and } \{v,x\} \text{is contained in some } C_\ell\},$$
$$\{w_1,\ldots,w_s\} = A_1 \setminus \{v,u_1,\ldots,u_r\}.$$

给定一个 MULTIDISJ$_{n^2,r}$ 实例，其中对每个 $q \in [r]$，第 q 个参与方持有输入$S_q \subseteq [n^2]$，我们定义归约图 G_H 的顶点集为

$$V(G_H) = A_1^{(1)} \sqcup \ldots \sqcup A_1^{(n)} \sqcup (V(H) \setminus A_1)$$

其边集为

$$E(G_H) = E_1 \sqcup E_2 \sqcup E_3 \sqcup E_4 \sqcup E_5,$$

$$E_1 = \bigsqcup_{j=1}^{n} \Big\{ \{x^{(j)}, y^{(j)}\} : \{x, y\} \in E(A_1), \text{ and } \{x, y\} \neq \{v, u_q\} \text{ for all } q \in [r] \Big\},$$

$$E_2 = \bigsqcup_{j=1}^{n} \Big\{ \{v^{(j)}, u_q^{(i)}\} : q \in [r], (j-1)n + i \in S_q \Big\},$$

$$E_3 = \bigsqcup_{i \neq j} \Big\{ \{v^{(j)}, w_t^{(i)}\} : t \in [s], \{v, w_t\} \in E(H) \Big\},$$

$$E_4 = \Big\{ \{x, y\} \in E(H) : x, y \in V(H) \setminus A_1 \Big\},$$

$$E_5 = \bigsqcup_{j=1}^{n} \Big\{ \{x^{(j)}, y\} : x \in A_1,\ y \in V(H) \setminus A_1,\ \{x, y\} \in E(H) \Big\}.$$

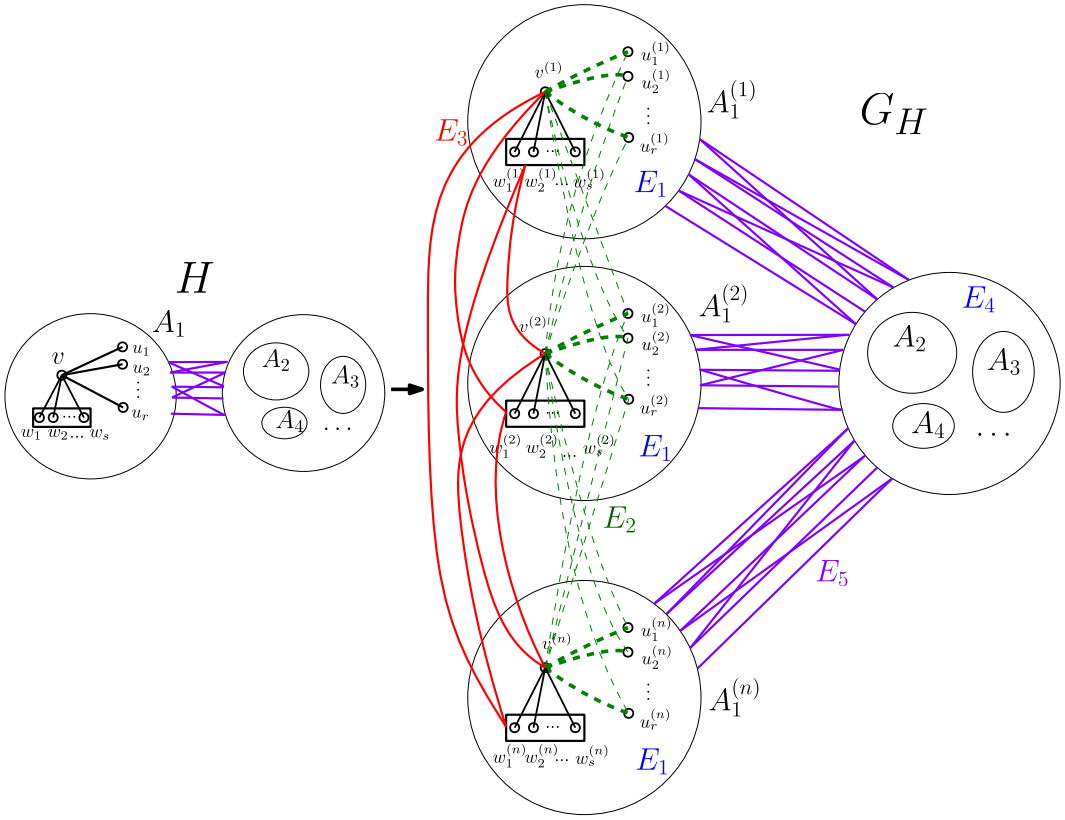

图1 归约图 G_H 见命题 13。 E_2 中的边以虚线表示，因为它们取决于 MULTIDISJ$_{n^2,\ r}$的输入。

下面的断言表明，MULTIDISJ$_{n^2,\ r}$ 可以归约到 G_H中的 Sub(H)。但是

$$|V(G_H)| = |V(H)| + (r + s + 1)(n - 1) = \Theta(n),$$

因此，任何能够在任意具有 n 个顶点的输入图中解决 H-查找问题的流式算法，在 p 遍的情况下都需要 $\Omega(n^2/p)$ 的空间；否则，我们就可以将其应用于 G_H

10 子图查找问题的流式空间复杂度二分性

并打破 MULTIDISJ$_{n^2,\ r}$ 的复杂度。

断言： 当且仅当 $\bigcap_{q=1}^{r} S_q \neq \emptyset$ 时，我们有 $H \subseteq G_H$。

证明。 如果 $\bigcap_{q=1}^{r} S_q \neq \emptyset$，那么 $(j-1)n+i \in \bigcap_{q=1}^{r} S_q$ 对于某个 $i,\ j \in [n]$。根据 $E(G_H)$ 的定义，我们有 $\{v^{(j)},\ u_q^{(i)}\} \in E_2$ 对于所有 $q \in [r]$，但在 G_H 中还有边 $\{v^{(j)},\ w_t^{(i)}\}$ 来自 E_3，因此我们得到一个 H 副本，由以下部分构成

$$V(H) \setminus A_1,\ A_1^{(i)} \setminus \{v^{(i)}\},\ \{v^{(j)}\}.$$

反过来，如果 $\bigcap_{q=1}^{r} S_q = \emptyset$，那么根据 MULTIDISJ$_{n^2,\ r}$ 的假设，对于所有 $q_1 \neq q_2$，我们都有 $S_{q_1} \cap S_{q_2} = \emptyset$。于是根据定义，对于所有 $i,\ j \in [n]$，至多存在一个 $q \in [r]$，使得 $\{v^{(j)}, u_q^{(i)}\} \in E_2$。在这种情况下，我们证明每个 $e \in E_2 \cup E_3 \cup E_5$ 都不能包含在 C_ℓ 的任何副本中。

首先，假设某个 $\{v^{(j_1)}\ w^{(j_2)}\ t\} \in E_3$ 包含在有向图 C_ℓ 的一个副本中，设该副本为

$$v^{(j_1)} - w_t^{(j_2)} - \ldots - x^{(i)} - \ldots - z - \ldots - v^{(j_1)},$$

这里的意思是，每个内部顶点都可以是某个 $i \in [n]$ 对应的 $x^{(i)} \in A_1^{(i)}$，或者是某个 $z \in V(H) \setminus A_1$。然后回到 H 中，我们得到长度为 ℓ 的闭游走 $v - w_t - \ldots - x - \ldots - z - \ldots - v$。根据 ℓ 的最小性，它必定是一个 C_ℓ，这就导致矛盾，因为按定义，边 $\{v, w_t\}$ 不应属于任何 C_ℓ。同理，我们可以证明每个 $e \in E_5$ 都不包含在任何 C_ℓ 中。

另一方面，假设某个 $\{v^{(j_1)}\ u^{(j_2)q}\} \in E_2$ 包含在一个有向图 C_ℓ 中，那么这个 C_ℓ 必须落在 $A_1^{(1)} \sqcup \ldots \sqcup A_1^{(n)}$ 中，因为没有 $e \in E_5$ 能包含在其中，因此我们可以假设它为

$$v^{(j_1)} - u_q^{(j_2)} - x_1^{(j_3)} - \ldots - x_{\ell-2}^{(j_\ell)} - v^{(j_1)}, \quad j_1, \ldots, j_\ell \in [n],$$

其中每个 $x_i \in A_1$，且最后一条边 $\{x^{(j_\ell)}{}_{\ell-2}, v^{(j_1)}\}$ 都属于 E_2，因为该 C_ℓ 中不能包含任何 $e \in E_3$。然而，对于所有 i、$j \in [n]$，我们至多可以有一个 $q \in [r]$ 使得 $\{v^{(j)}, u_q^{(i)}\} \in E_2$，因此特别地，$\{v^{(j_1)}\ u^{(j_2)q}\}$ 是 E_2 中端点位于 $A^{(j_1)}\mathbf{1}$ 和 $A^{(j_2)}\mathbf{1}$ 的唯一边，从而有 $j_\ell \neq j_2$。因此，从 $v^{(j_1)}\ u^{(j_2)q}$ 出发，该 C_ℓ 必须在到达 $x^{(j_\ell)}{}_{\ell-2}$ 之前的某个时刻离开 $A^{(j_2)}\mathbf{1}$，所以它必须使用 E_2 中除 $\{v^{(j_1)}\ u^{(j_2)q}\}$ 和 $\{x^{(j_\ell)}{}_{\ell-2}, v^{(j_1)}\}$ 之外的某条边。这意味着该 C_ℓ 的某个内部顶点 $x^{(j_{i+2})}{}_i$ 实际上是 $v^{(j_{i+2})}$，但回到 A_1 中，我们有长度为 ℓ 的闭游走 $v - u_q - x_1 - \ldots - x_{\ell-2} - v$，且它两次使用了 v，因此其中包含一个比 C_ℓ 更短的奇圈，矛盾。

因此，确实没有 $e \in E_2 \cup E_3 \cup E_5$ 包含在 C_ℓ的副本中，所以G_H中的一个 SOCC 要么是 $G_H[A_2]$，$G_H[A_3]$，……，$G_H[A_k]$，要么包含在某个 $i \in [n]$的 $A_1^{(i)}$中。然而，现在每个 $G_H[A_1^{(i)}]$ 中的边数都少于 $H[A_1]$中的边数，因为$\{v^{(i)},\ u_1^{(i)}\}$……，$\{v^{(i)},\ u_r^{(i)}\}$中至多只有一个能出现在 $E(G_H)$中。因此，G_H中大小与 $H[A_1]$ 相同的 SOCC 数量少于 H中的相应数量，从而 $H \not\subseteq G_H$。◁

◀

当 $p \leq \text{polylog}(n)$时，p-遍次 $\Omega(n^2/p)$-空间下界在至多相差一个多重对数因子的意义下是紧的，因为我们有朴素的单遍 $O(n^2)$-空间算法。至于更大的p，利用颜色编码 来自 [2]，的思想，我们可以为某些无向 H （结构简单）找到相应的上界。特别地，我们可以对所有奇圈$H = C_{2\ell+1}$证明这一点。为此我们需要如下着色引理。

▶ **引理 14**（Schimdt-Siegal [26]）。设 $\ell \in \mathbb{N}$ 与 V 为一个满足 $|V| = n$ 的集合，则存在一个由 $(2\ell+1)$-着色 χ： $V \to [2\ell+1]$ 组成的族 F，其大小为 $|F| = O(\log^2 n)$，使得对所有 $V' \subseteq V$，在满足 $|V'| = 2\ell+1$ 时，我们有 $\chi(V') = [2\ell+1]$，对某个 $\chi \in F$。

特别地，对于任意一个 n 顶点图 G，我们可以对 $V(G)$ 进行 $L = O(\log^2 n)$ 次着色，且与 G 无关，使得如果 G 包含 $C_{2\ell+1}$ 的一个副本，那么该副本在其中一种着色下是彩色的。然后，我们可以通过动态规划找到一个 $C_{2\ell+1}$ 的彩色副本。粗略地说，我们将 $V(G)$ 划分为不同的部，每一部都是某个区段中的锚点顶点，而一个区段由若干次数据流扫描组成。然后，对于每个锚点顶点，我们记录在每个指定距离和颜色下从它出发的可能目的地；如果存在彩色 $C_{2\ell+1}$，就能检测到它。

▶ **命题15**。固定 $\ell \geq 1$，则对于所有 $p = O(n)$，有 $D^p(\mathit{Sub}(C_{2\ell+1})) = \tilde{O}(n^2/p)$。

证明。令 $r = p/2\ell - 1$ 并将 $V(G) = V_1 \sqcup \ldots \sqcup V_r$ 划分为若干近似平衡的部分，任意两部分的大小相差至多 1。取 $L = O(\log^2 n)$ 种着色 $\chi_1, \ldots, \chi_L$： $V(G) \to [2\ell+1]$如上所述，使用空间 $O(Ln) = \tilde{O}(n^2/p)$。令 $j = 1$。

(1) 写出 $V_j = \{v_1^{(j)}, \ldots, v_{n/r}^{(j)}\}$。对于所有满足条件的 $i \in [n/r]$、$c \in [L]$、$0 \leq d \leq 2\ell$ 和 $S \subseteq [2\ell+1] |S| = d+1$，我们声明一个数组 $D_d^{(i,c)}[S]$： $V(G) \to \{0,1\}$，使用空间

$$\tilde{O}((n/r)L(2\ell)2^{2\ell+1}n) = \tilde{O}(n^2/p)$$

总计，并为所有 $x \in V(G)$ 和 $d \geq 1$ 初始化 $D_d^{(i,c)}[S](x) = 0$。我们的目标是使其满足

$$D_d^{(i,c)}(x) = 1 \Leftrightarrow \text{under } \chi_c, \exists \text{ a length-}d \text{ path } v_i^{(j)} \rightsquigarrow x \text{ using exactly the colors in } S.$$

为此，对于 $d = 0$，我们需要通过以下方式进行初始化

$$D_0^{(i,c)}[S](x) = \begin{cases} 1, & \text{if } x = v_i^{(j)} \text{ and } S = \{\chi_c(v_i^{(j)})\}, \\ 0, & \text{otherwise.} \end{cases}$$

（2）对于 $d \geq 1$，假设前 $d-1$ 遍中已确定所有 $D_{d-1}^{(i,c)}[S]$，那么在第 d 遍中，每当一条边 $\{x, y\}$ 到达时，我们检查 $D_{d-1}^{(i,c)}[S \setminus \{\chi_c(x)\}](y) = 1$ 是否成立；若成立，则设置 $D_d^{(i,c)}[S](x) = 1$。类似地，如果 $D_{d-1}^{(i,c)}[S \setminus \{\chi_c(y)\}](x) = 1$，则设置 $D_d^{(i,c)}[S](y) = 1$。这样，在第 (2ℓ) 遍之后，对于所有 i, c, S， $D_{2\ell}^{(i,c)}[S]$ 都已确定。随后，在第 $(2\ell+1)$ 遍中，我们可以检查是否存在某个满足 $D_{2\ell}^{(i,c)}[S](w) = 1$ 的 w 与 $v_i^{(j)}$ 相邻。若存在，则转到步骤（3）；否则转到步骤（4）。

(3) 现在我们知道 w 和 $v_i^{(j)}$ 位于一个 $C_{2\ell+1}$上，并且我们存储 $v_i^{(j)}$， w。在第 $(2\ell+2)$ 遍中，每当一条与 w 关联的边 $\{w, y\}$ 到达时，我们就检查是否存在某个$S' \not\ni \chi_c(w)$，使得 $D_{2\ell-1}^{(i,c)}[S'](y) = 1$。若存在，则 $v_i^{(j)}$， w， y 是 χ_c下一个彩色 $C_{2\ell+1}$的组成部分，我们便将它们存储起来。经过迭代，我们保证在第 4ℓ 遍后找到一个 χ_c下的彩色 $C_{2\ell+1}$，并输出这个副本。

(4) 现在，没有满足 $D_{2\ell}^{(i,c)}[S](w) = 1$ 的顶点 w 与 $v_i^{(j)}$ 相邻。若 $j = r$，则 $C_{2\ell+1} \not\subseteq G$。否则，我们将 j 加 1，并从步骤 (1) 重新开始。注意，所有数组 $D_d^{(i,c)}[S]$ 都可以重新初始化并重复使用，因此总空间仍为 $\tilde{O}(n^2/p)$。

该算法是正确的，因为如果存在一个 $C_{2\ell+1}$的副本，那么在某种 χ_c 下它会是彩色的，因此对于某个 j，它会包含一个位于 V_j 中的顶点，并被找到。此外，步骤（1）和（2）总共至多使用 2ℓ 遍，步骤（3）使用 $2\ell - 1$ 遍，因此遍数总计至多为 $2\ell(r+1) = p$。◀

3.2 二分图的下界

除了我们刚刚确立的二分定理之外，我们还可以研究二分图的下界，如定理 4 所总结的。我们可能会问，对于二分图 H，是否也有 $R^p_{2/3}(\textsc{Sub}(H)) = \Omega(\mathrm{ex}(n,H)/p)$；定理 4 的第一部分证明了偶圈以及非匹配森林的情形。证明通过从集合不相交问题进行归约得到。

对于偶圈 $C_{2\ell}$，我们将双方的输入编码为两个相同的稠密 $C_{2\ell}$-自由图的边，然后用长度为 $\ell-1$的额外路径连接相应的顶点，从而完成归约。

▶ **命题 16.** 固定 $\ell \geq 2$，则 $R^p_{2/3}(\textsc{Sub}(C_{2\ell})) = \Omega(\mathrm{ex}(n,C_{2\ell})/p)$。

证明。 取一个 n-顶点 $C_{2\ell}$-自由图 G_1，其 $|E(G_1)| = \mathrm{ex}(n,C_{2\ell})$。令

$$V(G_1) = \{v_1,\dots,v_n\}, \quad E(G_1) = \{e_1,\dots,e_{\mathrm{ex}(n,C_{2\ell})}\}.$$

令 G_2 为 G_1 的一个副本，包含对应的顶点和边

$$V(G_2) = \{v'_1,\dots,v'_n\}, \quad E(G_2) = \{e'_1,\dots,e'_{\mathrm{ex}(n,C_{2\ell})}\}.$$

此外，考虑所有 $j \in [n]$ 的新顶点 $u^{(j)}_1,\dots,u^{(j)}_{\ell-2}$。给定一个 $\mathrm{DISJ}_{\mathrm{ex}(n,C_{2\ell})}$ 实例，其中 Alice 和 Bob 分别持有 $S_1, S_2 \subseteq [\mathrm{ex}(n,H)],$，我们将归约图 $G_{C_{2\ell}}$ 定义为其顶点集为

$$V(G_{C_{2\ell}}) = V(G_1) \sqcup V(G_2) \sqcup \bigsqcup_{j=1}^{n} \{u^{(j)}_1,\dots,u^{(j)}_{\ell-2}\}$$

且边集为 $E(G_{C_{2\ell}}) = E_1 \sqcup E_2$，其中

$$E_1 = \bigsqcup_{j=1}^{n} \Big\{\{v_j,u^{(j)}_1\},\{u^{(j)}_1,u^{(j)}_2\},\dots,\{u^{(j)}_{\ell-2},v'_j\}\Big\}, \quad E_2 = \{e_i : i \in S_1\} \cup \{e'_i : i \in S_2\}.$$

以下断言表明，$\mathrm{DISJ}_{\mathrm{ex}(n,C_{2\ell})}$ 可以归约为 $G_{C_{2\ell}}$ 中的 $\mathrm{Sub}(C_{2\ell})$，由此证明即告完成，因为我们有 $|V(G_{C_{2\ell}})| = \Theta(n)$。

断言：</style>当且仅当 $S_1 \cap S_2 \neq \emptyset$ 时，$C_{2\ell} \subseteq G_{C_{2\ell}}$。

证明。若 $i \in S_1 \cap S_2$，则 $e_i, e'_i \in E(G_{C_{2\ell}})$。令 $e_i = \{v_a, v_b\}$ 其中某个 $a,b \in [n]$，因此$e'_i = \{v'_a, v'_b\}$，则存在一个 $C_{2\ell}$，使用以下边

$$e'_i, \{v_a,u^{(a)}_1\},\{u^{(a)}_1,u^{(a)}_2\},\dots,\{u^{(a)}_{\ell-2},v'_a\}, e'_i, \{v'_b,u^{(b)}_{\ell-2}\},\{u^{(b)}_{\ell-2},u^{(b)}_{\ell-3}\},\dots,\{u^{(b)}_1,v_b\}.$$

反过来，由于 G_1 和 G_2 都不含 $C_{2\ell}$，若 $G_{C_{2\ell}}$ 中存在 $C_{2\ell}$ 的副本，则该副本会使用 E_1 的某些边，因此它包含路径 $\{v_a, u^{(a)}_1\}$、$\{u^{(a)}_1, u^{(a)}_2\}$.....$\{u^{(a)}_{\ell-2}, v'_a\}$ 以及 $\{v_b, u^{(b)}_1\}$、$\{u^{(b)}_1, u^{(b)}_2\}$.....$\{u^{(b)}_{\ell-2}, v'_b\}$，其中 a、$b \in [n]$、$a \neq b$ 为某些值。但这些路径的长度为 $\ell-1$，因此要构成环 $C_{2\ell}$，必须有 $\{v_a, v_b\}$、$\{v'_a, v'_b\} \in E(G_{C_{2\ell}})$，这意味着 $S_1 \cap S_2 \neq \emptyset$。◁

◀

对于非匹配的森林，归约很简单，因为图兰数就是 $\Theta(n)$。注意，如果 H 是一个匹配，那么就存在一种非常简单的算法，正如区段 7中所述。

▶**命题 17**。 设 H 是一个固定的非匹配森林，则

$$R^p_{2/3}(\textsc{Sub}(H)) = \Omega(n/p).$$

证明。 令 H_1 为 H的一个最大连通分量，并固定两条边 e_1， $e_2 \in H_1$。取 n 个H_1的副本，记为 $H_1^{(1)}$，……，$H_1^{(n)}$，其中

$$E(H_1^{(1)}) = \{e^{(1)} : e \in E(H)\}, \ldots, E(H_1^{(n)}) = \{e^{(n)} : e \in E(H)\}.$$

给定一个 DISJ$_n$ 实例，其中 Alice 和 Bob 分别持有 S_1， $S_2 \subseteq [n]$，我们将归约图 G_H 定义为顶点集上的图

$$V(G_H) = V(H \setminus H_1) \sqcup V(H_1^{(1)}) \sqcup \ldots \sqcup V(H_1^{(n)})$$

其边集为 $E(G_H) = E_1 \sqcup E_2$，其中

$$E_1 = E(H \setminus H_1) \sqcup \bigsqcup_{j=1}^{n} E(H_1^{(j)} \setminus \{e_1^{(j)}, e_2^{(j)}\}), \quad E_2 = \{e_1^{(j)} : j \in S_1\} \cup \{e_2^{(j)} : j \in S_2\}.$$

若 $j \in S_1 \cap S_2$，则 $V(H_1^{(j)})$ 和 $V(H \setminus H_1)$ 共同构成 G_H 中 H 的一个副本；而若 $S_1 \cap S_2 = \emptyset$，则 G_H 中大小为 $|H_1|$ 的连通分量数量少于 H 中的数量，因此 $H \not\subseteq G_H$。因此，DISJ$_{\mathrm{ex}(n,C_{2\ell})}$可归约为 G_H 中的 Sub(H)，证明至此完成，因为 $|V(G_H)| = \Theta(n)$。◀

仿照上面的证明，我们可以将这一 p遍$\Omega(\mathrm{ex}(n,H)/p)$ 下界推广到其他一些二分图 H，但我们不知道这对所有一般二分图 H是否成立。与定理4一样，我们要么需要对直径附加条件，要么只能推出单遍下界，而仍有一些情况尚未覆盖。我们将在区段7中进一步讨论这些开放问题。

为了推出多轮下界，我们将固定一个无 H的二分稠密 G' 作为骨架，然后复制 $V(H)$，根据 G'中的边，将多重不相交性问题的输入编码到交叉对中。要使这一归约有效，我们需要的附加条件是： H 的某个连通分量的直径小于 4。这等价于说，该连通分量中任意两个属于同一部的节点都有一个公共邻点，因为 H 是二分图。

▶**命题 18.** 设 H 是一个固定的二分图。若 H 的某个连通分量的直径小于 4，则 $R^p_{2/3}(\textsc{Sub}(H)) = \Omega(\mathrm{ex}(n,H)/p)$。

P证明。 由于我们通过归约证明下界，因此可以将 H 替换为
c连通分量，因此不失一般性地假设 H 的直径 diam(H) < 4 .
设 G' 是一个具有平衡部的 $(2n)$-顶点 H-自由二分图，其边数为 $\Theta(\mathrm{ex}(n,H))$。
设 $V(G')$ 的两个部为 $\{g_1, \ldots, g_n\}$ 和 $\{g'_1, \ldots, g'_n\}$，记

$$E(G') = \{f_1, \ldots, f_{|E(G')|}\}, \quad f_k = \{g_{i_k}, g'_{j_k}\} \quad \forall 1 \le k \le |E(G')|.$$

同样，设 $V(H)$ 的两个部为 $\{v_1, \ldots, v_a\}$ 和 $\{u_1, \ldots, u_b\}$，记

$$E(H) = \{e_1, \ldots, e_{|E(H)|}\}, \quad e_q = \{v_{a_q}, u_{b_q}\} \quad \forall 1 \le q \le |E(H)|, \quad a_q \in [a], \quad b_q \in [b].$$

现在我们构造 n 个 V_1、…、V_n，即 $V(H)$

$$V_j = \{v_1^{(j)}, \ldots, v_a^{(j)}\} \sqcup \{u_1^{(j)}, \ldots, u_b^{(j)}\}, \quad 1 \le j \le n.$$

给定一个 MULTIDISJ$_{|E(G')|,\ |E(H)|}$ 的实例，其中对于每个 $q \in [|E(H)|]$，第 q 个参与方持有输入 $S_q \subseteq [|E(G')|]$，我们将归约图 G_H 定义在顶点集上

$$V(G_H) = V_1 \sqcup \ldots \sqcup V_n$$

并且边集为

$$E(G_H) = \left\{ \{v_{a_q}^{(i_k)}, u_{b_q}^{(j_k)}\} : q \in [|E(H)|],\ k \in S_q \right\}.$$

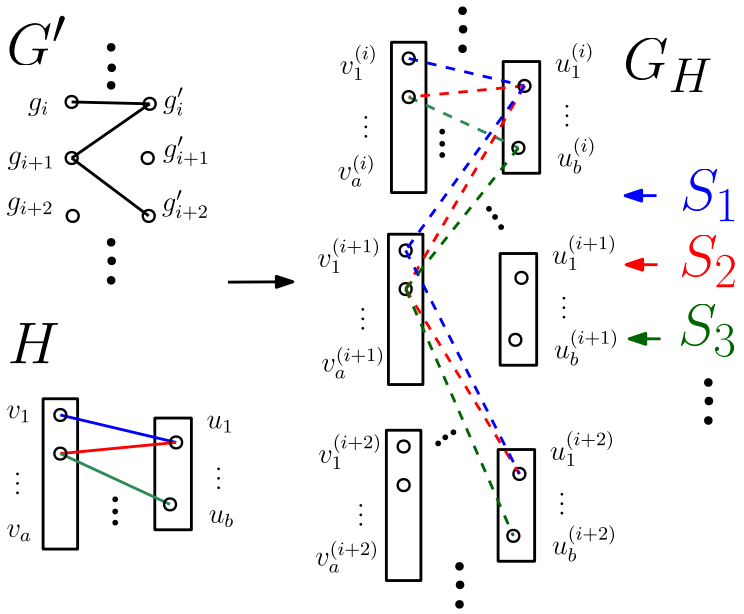

图2 归约图 G_H 命题 18。一种颜色的虚线边对应于 MULTIDISJ$_{|E(G')|,|E(H)|}$ 中某个参与方的输入。

直观地说，H 中的边 $e_q = \{v_{a_q}, u_{b_q}\}$ 被分配给第 q 个参与方；只有对于这样的顶点对 $\{v^{(i_k)_{a_q}}, u^{(j_k)_{b_q}}\}$，其中 (i_k, j_k) 对应于 G' 中的边 $f_k = \{g_{i_k}, g'_{j_k}\}$，参与方才可以编码其输入 S_q，也就是说，根据 $k \in S_q$ 来添加该边。接下来的断言完成了证明，因为我们有 $|V(G_H)| = \Theta(n)$，并且 $|E(G')| = \Theta(\mathrm{ex}(n, H))$。

断言： 当且仅当 $\bigcap_{q=1}^{|E(H)|} S_q \neq \emptyset$ 时，我们有 $H \subseteq G_H$。

证明。 如果 $k \in \bigcap_{q=1}^{|E(H)|} S_q$，那么 $\{v^{(i_k)_{a_q}}, u^{(j_k)_{b_q}}\} \in E(G_H)$ 对所有 $q \in [|E(H)|]$ 都成立，但根据定义

$$\{\{v_{a_q}, u_{b_q}\} : q \in [|E(H)|]\} = E(H),$$

因此，我们得到一个由这些部构成的 H 副本

$$\{v_1^{(i_k)}, \ldots, v_a^{(i_k)}\},\ \{u_1^{(j_k)}, \ldots, u_b^{(j_k)}\}.$$

反过来，如果 $\bigcap_{q=1}^{|E(H)|} S_q = \emptyset$，那么根据 MULTIDISJ$_{|E(G')|,\ |E(H)|}$ 的假设，我们有 $S_{q_1} \cap S_{q_2} = \emptyset$ 对所有 $q_1 \neq q_2$ 均成立。这意味着，对于所有 $1 \leq i \leq n$，来自

$\{v_1^{(i)}$，……，$v_a^{(i)}\}$ 的两个不同顶点不能有公共邻点；否则，根据我们对 $E(G_H)$ 的构造，存在一些 $1 \le j \le n$ 和 $q_1 \ne q_2$，使得 $b_{q_1} = b_{q_2}$，从而

$$\{v_{a_{q_1}}^{(i)}, u_{b_{q_1}}^{(j)}\}, \{v_{a_{q_2}}^{(i)}, u_{b_{q_2}}^{(j)}\} \in E(G_H)$$

且 $(i,j) = (i_k, j_k)$，对于某个 $k \in [|E(G')|]$，从而得到 $k \in S_{q_1} \cap S_{q_2} \ne \emptyset$，这与 $k \in S_{q_1} \cap S_{q_2} \ne \emptyset$ 矛盾。

现在假设 G_H 中存在 H 的一个副本 H'，那么由 diam(H') < 4 可知，来自 H' 同一部的两个顶点会共享一个公共邻点；但如上所述，$\{v_1^{(i)}, \ldots, v_a^{(i)}\}$ 中的两个不同顶点在 G_H 中不可能有公共邻点。因此，对于所有 $1 \le j \le n$，我们有

$$|V(H') \cap \{v_1^{(j)}, \ldots, v_a^{(j)}\}|, |V(H') \cap \{u_1^{(j)}, \ldots, u_a^{(j)}\}| \le 1.$$

因此，存在互不相同的 $i_1, \ldots, i_a \in [n]$ 和互不相同的 $j_1, \ldots, j_b \in [n]$，使得以下条件之一成立

$$V(H') = \{v_{\alpha_1}^{(i_1)}, \ldots, v_{\alpha_a}^{(i_a)}\} \cup \{u_{\beta_1}^{(j_1)}, \ldots, v_{\beta_b}^{(j_b)}\}$$

对于某些 $\alpha_1, \ldots, \alpha_a \in [a]$ 和 $\beta_1, \ldots, \beta_b \in [b]$，或者

$$V(H') = \{v_{\alpha_1}^{(j_1)}, \ldots, v_{\alpha_b}^{(j_b)}\} \cup \{u_{\beta_1}^{(i_1)}, \ldots, v_{\beta_a}^{(i_a)}\}$$

对于某些 $\alpha_1, \ldots, \alpha_b \in [a]$ 和 $\beta_1, \ldots, \beta_a \in [b]$。前者表明，回到 G' 时，部$\{g_{i_1}, \ldots, g_{i_a}\}$ 与 $\{g'_{j_1}, \ldots, g'_{j_b}\}$ 构成 H 的一个副本，而后者表明 $\{g_{j_1}, \ldots, g_{j_b}\}$与 $\{g'_{i_1}, \ldots, g'_{i_a}\}$ 构成 H 的一个副本，但这两者相互矛盾，因为 G' 是 H-free 的。 ◁

◀

如果我们只要求单遍 Ω(ex(n, H)) 空间下界，那么只要 H 的某个连通分量的连通度大于 2，从 INDEX 出发的归约就可行。为证明这一点，我们同样可以不失一般性地假设 H 本身的连通度为 $\kappa(H) > 2$。然后，我们将固定一个 H-free 的稠密 G' 作为骨架，并用它的边来编码 Alice 的输入。将选择 H 的一条边，使其与 G' 中的索引边重叠，而 H 的剩余部将相应地附接。

▶**命题 19.** 设 H 是一个固定的二分图。如果 H 的某个连通分量的连通度大于 2，那么 $R^1_{2/3}(Sub(H)) = \Omega(\text{ex}(n, H))$。

证明。不失一般性，假设 $\kappa(H) > 2$。设 $H_0 = H$ 定义在顶点集$V(H_0) = \{v_1, \ldots, v_{|V(H)|}\}$ 上，并固定任意 $\{v_i, v_j\} \in E(H_0)$。设 G'是一个含有 n个顶点且无 H的图，有 ex(n, H) 条边，记作

$$E(G') = \{e_1, \ldots, e_{\text{ex}(n,H)}\}, \quad e_i = \{v_{i,1}, v_{i,2}\}, \quad v_{i,1}, v_{i,2} \in V(G').$$

给定一个 INDEX$^{\text{ex}(n,H)}$实例，其中 Alice 持有 $S \subseteq [\text{ex}(n, H)]$，Bob 持有$k \in [\text{ex}(n, H)]$, ，我们定义归约图 G_H 如下：

$$V(G_H) = V(G') \sqcup V(H_0) \setminus \{v_i, v_j\}, \quad E(G_H) = E_1 \sqcup E_2 \sqcup E_3,$$
$$E_1 = \{e_\ell : \ell \in S\}, \quad E_2 = \{e \in E(H_0) : \{v_i, v_j\} \cap e = \emptyset\},$$
$$E_3 = \Big\{\{v_{k,1}, v_\ell\} : \{v_i, v_\ell\} \in E(H_0)\Big\} \sqcup \Big\{\{v_{k,2}, v_\ell\} : \{v_j, v_\ell\} \in E(H_0)\Big\}.$$

也就是说，我们使用 $E(G')$ 对 S 进行编码，并通过重叠 $\{v_i, v_j\}$ 与$\{v_{k,1}, v_{k,2}\}$来附加一个 H 的副本，其中 k 是目标索引。下面的断言完成证明，因为$|V(G_H)| = \Theta(n)$。

断言： </style当且仅当 $k \in S$，我们有 $H \subseteq G_H$。

证明。若 $k \in S$，则 $\{v_k,1, v_k,2\} \in E(G_H)$，因此 $\{v_k,1, v_k,2\} \sqcup V(H_0) \setminus \{v_i, v_j\}$ 构成有向图 H 的一个副本。反之，假设 G_H 中存在有向图 H 的一个副本 H'，但 $k \notin S$，则 $\{v_k,1, v_k,2\} \notin G_H$，因此

$$H \not\subseteq G_H[V(G')], \quad H \not\subseteq G_H[\{v_{k,1}, v_{k,2}\} \sqcup V(H) \setminus \{v_i, v_j\}],$$
$$V(H') \cap (V(G') \setminus \{v_{k,1}, v_{k,2}\}) \neq \emptyset, \quad V(H') \cap (V(H_0) \setminus \{v_i, v_j\}) \neq \emptyset,$$

因此 $V(H') \cap \{v_{k,1}, v_{k,2}\}$ 是 H' 中大小为 ≤ 2 的顶点割，这与 $\kappa(H) > 2$ 相矛盾。 ◁

◀

在整个~区段中，我们一直致力于证明 p遍 $\Omega(\mathrm{ex}(n, H)/p)$空间下界，对于任意目标图 H，我们确实有朴素的单遍 $O(\mathrm{ex}(n, H))$空间确定性算法。然而，当允许多遍处理时，尤其是遍数为 n的多项式时，目前尚不清楚一般情况下是否存在空间效率更高的算法。命题 15 是存在此类算法的一个特例。关于上界的更多讨论将在第7区段中给出。

4 诱导简单图

接下来我们研究流式模型中子图查找问题的诱导变体，其目标是检测 H 的诱导副本，而不是任意的（不一定是诱导的）副本。一般而言，导出检测要困难得多：即使是非常稠密的图，也可能不包含给定的诱导模式。我们的结果在定理5中得出：除了一小组例外图 H外，我们有 $R^p_{2/3}(\textsc{IndSub}(H)) = \Omega(n^2/p)$，

~而特殊情形则存在确定性算法，仅需使用 $O(n)$ 空间和~$O(1)$ 遍。

4.1 困难实例

我们首先证明，在诱导情形下，非二分图情形依然同样困难。

▶ **命题20。** 设 H 为非二分图，则 $R^p_{2/3}(\mathit{IndSub}(H)) = \Omega(n^2/p)$。

证明。 在定理非二分图情形所用的归约中3，每个是实例都包含一个 H的副本，而且这个副本还是诱导的。另一方面，每个否实例都完全不包含 H 的副本（甚至不包含诱导副本）。因此，针对$\textsc{IndSub}(H)$ 的算法也能解决上一节中的通信博弈，从而得到所述下界。 ◀

此外，通过一个简单的取补图论证，同样的困难性也可推广到补图为非二分图的图。

▶**命题 21。** 设 H 为一个图，使得 H 是二分图，¯则

$$R^p_{2/3}(\textsc{IndSub}(H)) = \Omega(n^2/p).$$

证明。 给定一个图 G，注意到

$$G \text{ contains an induced copy of } \overline{H} \iff \overline{G} \text{ contains an induced copy of } H.$$

如果 H <style id='4}>是非二分图，则考虑定理 多重不相交性3 中使用的 H所对应的多方归约，该归约构造了归约图 G_H。参与者不对 $\overline{G_H}$进行数据流处理，¯而是可以对其补图 G_H 进行数据流处理，方法是对其

输入集合所贡献的边取补，并对构造中的其余固定边取补。在 G_H 上运行一个检测诱导-H-的流式算法，便可以解决多重不相交性问题。因此，这样的算法在进行 p 遍时，必须使用 $\Omega(n^2/p)$ 的空间。◀

▶**推论 22.** 设 H 是一个至少有 5 个顶点的图。那么 $R^p_{2/3}(\text{诱导子图}(H)) = \Omega(n^2/p)$。

证明。假设 H 有 t 个顶点，且 H 及其补图 $\overline{H}$ 都是二分图。设 $\phi\colon V(H) \to \{0,1\}$ 和 $\psi\colon V(\overline{H}) \to \{0,1\}$ 分别是 H和 $\overline{H}$的正常 2-着色。那么映射 $v \mapsto (\phi(v)\ \psi(v)) \in \{0,1\}^2$ 是 K_t的正常 4-着色。由此可得

$$t = \chi(K_t) \le 4.$$

等价地，每个图 H 至少有 5 个顶点的图，要么是非二分图，要么其补图是非二分图。命题 20和 21共同证明了诱导子图(H)的困难性。◀

对于具有 $|V(H)| \in \{3,4\}$ 的图，H 和 $\overline{H}$ 均为二分图的情况仅有 P_3、co-P_3、C_4、$2K_2$ 和 P_4；其余图都属于困难情形。此外，C_4 和 $2K_2$ 也很难。我们仅证明 IndSub(C_4) 的困难性；由于 $2K_2 = C_4$，$2K_2$ 的相应下界可通过与命题 21 相同的取补图论证得出。

▶ **命题 23。** $R^p_{2/3}(\text{诱导子图}(C_4)) = \Omega(n^2/p)$

证明。 我们将 不相交性问题$_{n^2}$ 归约到 诱导子图(C_4)。令 U_1，U_2，V_1，V_2 为四个两两不相交的顶点集，每个顶点集的大小均为 n。对于 $i \in \{1,2\}$，记

$$U_i = \{u_{i,1}, u_{i,2}, \ldots, u_{i,n}\}, \qquad V_i = \{v_{i,1}, v_{i,2}, \ldots, v_{i,n}\}.$$

给定 S_1，$S_2 \subseteq [n^2]$，定义

$$E_1 = \big\{\{u_{1,i}, v_{1,j}\} : (i-1)n + j \in S_1\big\}, \qquad E_2 = \big\{\{u_{2,i}, v_{2,j}\} : (i-1)n + j \in S_2\big\}.$$

此外，令

$$E_3 = \big\{\{u,u'\} : u,u' \in U_1,\ u \ne u'\big\} \sqcup \big\{\{v,v'\} : v,v' \in V_2,\ v \ne v'\big\},$$

and

$$E_4 = \left(\bigsqcup_{i\in[n]} \big\{\{u_{1,i}, u_{2,i}\}\big\}\right) \sqcup \left(\bigsqcup_{j\in[n]} \big\{\{v_{1,j}, v_{2,j}\}\big\}\right).$$

我们构造一个图 G，它具有

$$V(G) = U_1 \sqcup U_2 \sqcup V_1 \sqcup V_2 \qquad \text{and} \qquad E(G) = E_1 \sqcup E_2 \sqcup E_3 \sqcup E_4.$$

断言。 $S_1 \cap S_2 \ne \emptyset$ 当且仅当 G 包含 C_4的一个诱导副本。

证明。若 $(i-1)n+j \in S_1 \cap S_2$，那么根据 E_1和 E_2的定义，我们有 $\{u_{1,i}, v_{1,j}\} \in E_1$以及 $\{u_{2,i},v_{2,j}\} \in E_2$。此外，$\{u_{1,i}, u_{2,i}\} \in E_4$ 以及 $\{v_{1,j}, v_{2,j}\} \in E_4$。因此，这四个顶点 $\{u_{1,a},v_{1,b},v_{2,b}, u_{2,a}\}$ 构成环

$$u_{1,a} - v_{1,b} - v_{2,b} - u_{2,a} - u_{1,a}.$$

这四个顶点之间没有其他边，因此它们诱导出一个 C_4。

反之，假设 G 具有一个导出子图 $H \cong C_4$，其中 $V(H) = \{w_1, w_2, w_3, w_4\}$，且 $E(H) = \{\{w_i, w_{i+1}\} : i \in [4]\}$，其中 $w_5 = w_1$。我们证明 $S_1 \cap S_2 \ne \emptyset$。

情形 1： H 在每个 U_1、U_2、V_1、V_2中各有一个顶点。

注意， U_1 与 V_2 之间没有边， U_2 与 V_1之间也没有边，而 U_1， U_2 或 V_1， V_2 之间的边只连接具有相同索引的顶点。使用每一部中的一个顶点构成一个 4-环的唯一方式是

$$V(H) = \{u_{1,s}, v_{1,t}, v_{2,t}, u_{2,s}\}$$

对于某些 s、$t \in [n]$，其边为$\{u_{1,s}, v_{1,t}\} \in E_1$、$\{u_{2,s}, v_{2,t}\} \in E_2$，以及匹配边$\{u_{1,s}, u_{2,s}\}$,$\{v_{1,t}, v_{2,t}\} \in E_4$。因此，$(s-1)n+t \in S_1$和$(s-1)n+t \in S_2$，这意味着 $S_1 \cap S_2 \neq \emptyset$。

情况2： </style}某一部至少包含 H中的两个顶点。

我们导出一个矛盾。若 $|U_1 \cap V(H)| \geq 2$，取互不相同的 x、$y \in U_1 \cap V(H)$。由于根据 E_3，U_1是一个团，因此有$\{x, y\} \in E(G)$。由于 H是一个诱导 C_4，边$\{x, y\}$必须属于该环，因此另外两个顶点不能在 U_1 中（否则会形成一个三角形）。所以，它们位于 $N(U_1) \subseteq U_2 \cup V_1$中。

然而，它们也不能同时位于 $U_2 \cup V_1$ 中，因为 $U_2 \cup V_1$ 是一个独立集，而 4-环中剩余的两个顶点必须相邻。这就产生了矛盾，因此 $|U_1 \cap V(H)| \leq 1$。对称地可得 $|V_2 \cap V(H)| \leq 1$。

若 $|U_2 \cap V(H)| \geq 2$，则 U_2 中的两个顶点不相邻。在诱导 4-环中，对顶点恰有两个公共邻居；但在 G 中，公共邻域结构迫使这两个邻居都位于 V_2 （通过 E_2 中的边），这会使它们相邻（因为 V_2 诱导出一个团），从而形成一条弦。这与 H 是诱导的相矛盾。对称地，也可排除 $|V_1 \cap V(H)| \geq 2$。因此，情况2不可能成立，情形 1 必须成立。
$S_1 \cap S_2 \neq \emptyset$。$\lhd$

根据该断言，流式算法的存在意味着存在一个不相交性问题协议：Alice 和 Bob 可以在构造的数据流上模拟该算法，以判断两个集合是否不相交。◀

4.2 简单实例

上一小节剩余的 3 个图很容易用流式算法判定。~ ~事实上，它们都可以用确定性算法在 $O(1)$ 遍次和 $O(n)$ 空间内求解。

▶ **命题 24。** D^1（诱导子图（P_3）$= \widetilde{O}$（n）。

证明。 根据一个简单的观察，一个图不包含诱导 P_3 当且仅当它是若干个团的并。

在流式模型中，可以保留一个生成森林，并在单遍扫描和 $O(n)$ 空间内计算~每个顶点的度。要检查该图是否是若干个团的并，只需检查每个顶点的度是否恰好等于其所在连通分量的大小减一。◀

对于接下来的两种情况，我们将经常需要推理补图 G，特别是将 $V(G)$ 划分为 G 的¯连通分量。然而，在不对 G 作任何结构性假设的情况下，直接在~流式模型中执行这项任务并非易事。~为此，我们使用一种在 $O(n)$ 空间内运行、仅需对数多项式遍数的例程；它要么输出所需的划分，要么在无法做到时，输出 G 中的特定证书结构。

▶ **引理 25.** 给定一个图 G 和一个整数 k，存在一种 $O(k \log n)$ 遍确定性~流式算法，使用 $O(kn)$ 空间；该算法要么输出将 $V(G)$ 划分为 G 的连通分量的划分，要么在 G 中找到一个诱导的补图-P_{k+2} 副本。

证明。 我们首先固定一个任意的严格线性序 $\prec$ 在 $V(G)$。对于每个顶点 $v \in V(G)$以及每个 $0 \le i \le k$，我们定义

$$N(v,i) := \min_{\prec}\{u \in V(G) \mid \mathrm{dist}_{\overline{G}}(u,v) \le i\},$$

其中，最小值是相对于序 $\prec$而取的。换言之，$N(v, i)$ 表示从 v 在 G 中通过长度至多为 i 的路径可到达的最小顶点。特别地，我们对所有 $v \in V(G)$ 初始化 $N(v, 0) = v$。

我们采用动态规划方法计算所有 $v \in V(G)$ 对应的 $N(v, k)$。假设对于所有 $v \in V(G)$，$N(v, i)$ 的值都已计算完毕。下面说明如何使用额外的 $O(\log n)$ 遍次计算 $N(v, i+1)$。根据定义，$N(v, i+1)$ 是 G 中 v 的所有非邻点 w 对应的 $N(w, i)$ 的最小值，以及 v 本身的值，两者取最小。

为此，我们通过设定以下方式，在 $V(G)$ 上定义一个新的严格线性序 $\prec_i$

$$u \prec_i w \iff N(u,i) \prec N(w,i) \text{ or } (N(u,i) = N(w,i) \wedge u \prec w).$$

Let

$$x_{1,i} \prec_i x_{2,i} \prec_i \cdots \prec_i x_{n,i}$$

作为顶点的所得排序。然后，只需为每个 v 找到满足 $x_{s_v,i} \not\sim v$ 在 G 中的最小索引 s_v，并设定

$$N(v, i+1) = \min_{\prec}\big(v,\, N(x_{s_v,i}, i)\big).$$

值 s_v 可以通过二分搜索过程确定。初始时，为每个顶点 v 分配区间 $[1, n]$，其对应的范围是从 $x_{1,i}$ 到 $x_{n,i}$。每一步中，我们将当前区间分成两半，并进行单遍扫描，统计对于每个 v，v 的邻居中有多少个落在左半区间内。如果左半区间中的所有顶点都是 v 的邻居，我们就舍弃左半区间，继续处理右半区间；否则，舍弃右半区间，继续处理左半区间。重复此过程 $O(\log n)$ 轮次，即可得到所有 v 对应的 s_v。

~总计，使用 $O(kn)$ 的空间，可通过 $O(k \log n)$ 遍次计算出值 $N(v, k)$。

计算出值 $N(v, k)$ 后，我们根据这些值将 $V(G)$ 划分为块 B_1、.....、B_ℓ：当且仅当 $N(u, k) = N(v, k)$ 时，顶点 u 和 v 才属于同一块。然后，我们再进行一个遍次，统计不同块之间跨越的边总数。将这个数记为 F。

- 情形 1：$F = \sum_{1 \le i < j \le \ell} |B_i| \cdot |B_j|$。在这种情况下，$G$ 中任意一对不同的块之间都完全相连，这意味着来自不同块的顶点在 G 中属于不同的连通分量。此外，由于 $N(u, k) = N(v, k)$ 意味着 u 和 v 在 G 中相连，因此每个块 B_i 恰好对应于 G 的一个连通分量。情形 2：~$F < \sum_{1 \le i < j \le \ell} |B_i| \cdot |B_j|$。此时，存在顶点 $u \in B_i$ 和 $v \in B_j$，满足 $i \ne j$，使得 $\{u, v\} \notin E(G)$，等价地，$\{u, v\} \in E(\overline{G})$。通过先计算每个顶点在其所属块之外的邻居数，选取一个未与其他所有块中的顶点完全相邻的顶点，再显式检查其邻居，即可使用 $O(1)$ 个额外遍次找到这样一对顶点。

不妨假设 $w = N(u,k) \prec N(v,k)$。那么我们必有

$$\mathrm{dist}_{\overline{G}}(w,v) \geq k+1,$$

否则 $N(v,k) \preceq w$，这与假设 $w \prec N(v,k)$ 矛盾。另一方面，

$$\mathrm{dist}_{\overline{G}}(w,v) \leq \mathrm{dist}_{\overline{G}}(w,u) + 1 \leq k+1,$$

其中，最后一个不等式由 $N(u,k)$ 的定义得出。因此，

$$\mathrm{dist}_{\overline{G}}(w,u) = k \quad \text{and} \quad \mathrm{dist}_{\overline{G}}(w,v) = k+1.$$

因此，在 G 中，从 w 到 v 的一条最短路径上的顶点构成一条包含$k+2$ 个顶点的诱导路径，它对应于 G 中的一个诱导 co-P_{k+2}。此外，在计算过程中，通过为每个 v 维护相应的前驱索引 s_v，可以从表 N 中恢复这条路径。

◀

此算法可直接用于求解 IndSub(co-P_3)。

推论 26。 D^p(诱导子图$(co\text{-}P_3)) = O(n)$ 对于某个 $p = O(\log n)$。

证明。 注意到 G 不包含诱导的 余图-P_3 当且仅当 G 是团的不交并。根据引理 25，经过 $O(\log n)$ 遍次后，我们要么得到一个诱导的 余图-P_3，要么得到 $V(G)$ 划分为 G的连通分量。

在后一种情况下，只需再对数据流进行一遍，检查 G 的边是否有任何一条位于 G的某个连通分量内部。如果存在这样的边，它就能直接给出一个诱导的余图-P_3；否则， G 不包含诱导的余图-P_3。
◀

对于最后一种情况 P_4，我们的方法主要基于 [24], 提出的方法，该方法研究在并行计算模型下对诱导P_4-自由图或余图的识别。我们采用了关于诱导P_4-自由图结构性质的类似观察，同时采用不同的算法策略，以应对计算模型的不同限制。

▶ **定理 27**。 D^p(诱导子图$(P_4)) = \widetilde{O}(n)$ 对于某个 $p = O(\log^2 n)$。

在介绍我们的算法之前，我们首先推导它所依赖的余图结构性质。注意，任何诱导 P_4 都必须位于 G 的某个连通分量内，并且位于 G的某个连通分量内。根据这一观察以及引理 25，我们可以使用 $O(1)$ 遍次进行初始预处理，将 $V(G)$ 划分为 G 和 G的连通分量。因此，从现在起，我们可以只关注 G和G均为连通图的实例。

▶**引理 28** ([24], 引理 2.1 和推论 2.1)。 连通图 G 是诱导P_4-自由的，当且仅当对于任意 $v \in V(G)$，都存在 $k \in \mathbb{N}$ 以及对 $V(G)$ 的一个划分，将其分为

$$A_1, \ldots, A_k,\ B_1, \ldots, B_k$$

使得：

- $N(v) = A_1 \sqcup \cdots \sqcup A_k$以及 $V(G) \setminus N(v) = B_1 \sqcup \cdots \sqcup B_k$。对于所有 $1 \leq i < j \leq k$,
- $A_i \cup B_i \cup B_j$ 中的每个顶点都与 A_j中的每个顶点相邻。

对于所有 $1 \le i < j \le k$， $A_i \cup B_i$ 与 B_j之间没有边。对于每个 $i \in [k]$，导出子图 $G[A_i]$ 和 $G[B_i]$ 都是导出-P_4-无的。

我们将这样的分解称为关于根的 G 分量划分 v。此外，$v \in B_1$。

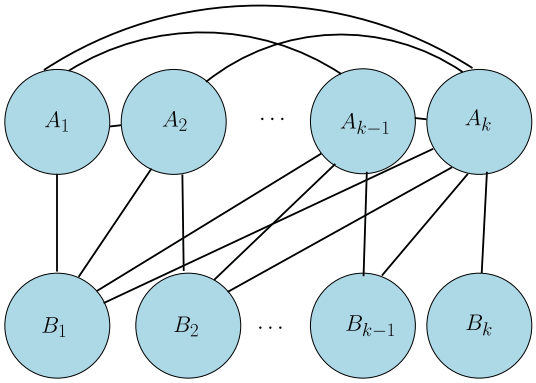

图 3 诱导无P_4图的结构。两组之间的一条边表示其中一组的所有顶点都与另一组的所有顶点相邻。

证明。容易验证，若 G 具有所述结构，则 G 不包含诱导 P_4。反过来，假设 G是诱导无P_4图。我们逐步推导出该结构。

观察1。 G的直径至多为 2。
事实上，若存在 u、$w \in V(G)$，使得 dist$(u, w) \ge 3$，那么从 u到 w的任意最短路径都包含一个诱导 P_4。特别地，对于任意固定的根 v，我们有 $V(G) \setminus N(v) \subseteq N(N(v))$。记 $A := N(v)$ 和 $B := V(G) \setminus N(v)$。

观察 2. 如果 $u, w \in A$ 不相邻，则 $N(u) \cap B = N(w) \cap B$。
否则，假设存在 $x \in (N(u) \cap B) \setminus (N(w) \cap B)$。由于 $x \in B$，我们有 $x \not\sim v$，而 $x \sim u$ 和 $v \sim w$。此外，根据假设有 $u \not\sim w$，并且根据对 x 的选择有 $x \not\sim w$。因此 $x - u - v - w$ 是一个导出 P_4，矛盾。从而，$G[A]$ 的同一连通分量中的所有顶点在 B 中具有相同的邻域。

观察 3. 如果 $u, w \in B$ 相邻，则 $N(u) \cap A = N(w) \cap A$。
否则，假设存在 $x \in (N(u) \cap A) \setminus (N(w) \cap A)$。则 $v - x - u - w$是一个导出 P_4：我们有 $v \sim x \sim u \sim w$，而 $v \not\sim u$ 和 $v \not\sim w$ （因为 $u, w \in B$），并且根据对 x 的选择还有 $x \not\sim w$。因此， $G[B]$ 的同一连通分量中的顶点在 A中具有相同的邻域。

观察4。 对于任意 u， $w \in A$，集合 $N(u) \cap B$和 $N(w) \cap B$之间存在包含关系。

我们只关心 u 和 v 相邻的情况，因为观察2处理了另一种情况。反设不然，则存在 $x \in (N(u) \cap B) \setminus (N(w) \cap B)$ 和 $y \in (N(w) \cap B) \setminus (N(u) \cap B)$。若 $x \sim y$，则 x和 y是 B中邻域不同的相邻顶点 A （一个与 u 相邻但与 w不相邻，另一个与 w 相邻但与 u不相邻），这与观察3矛盾。因此 $x \not\sim y$，且 $x - u - w - y$是一个诱导的 P_4，矛盾。

结论。 按照集合 $N(\cdot) \cap B$将 A 划分为若干类；将这些类记为A_1，……， A_k，并按如下次序排列，使得

$$N(A_1) \cap B \subsetneq N(A_2) \cap B \subsetneq \cdots \subsetneq N(A_k) \cap B = B.$$

（其严格性由观察4以及这些类的定义得出，最后一个等式来自观察1）令 $A_0 := \emptyset$，并定义

$$B_i := \big(N(A_i) \setminus N(A_{i-1})\big) \cap B \qquad \text{for } i = 1, \dots, k.$$

于是，得到的划分 $\{A_i\}_{i=1}^{k}$,$\{B_i\}_{i=1}^{k}$ 满足所述结构。 ◀

此外，对于任意图 G 和任意根顶点 v，存在一种 $O(1)$遍次流式算法~，使用 $O(n)$ 空间；该算法要么输出相对于 v 的分量划分，要么找到一个诱导 P_4，从而证明 G 不是余图。第一遍中，我们确定 $N(v)$ 和$V(G) \setminus N(v)$。第二遍中，对于每个顶点 u，我们计算 u在对侧的邻居数量（即，对于 $u \in N(v)$ 和 $|N(u) \cap N(v)|$，为 $|N(u) \cap (V(G) \setminus N(v))|$；对于 $u \notin N(v)$，则为.....）；当 G 是余图时，这些值便能确定划分为集合 A_i 和B_i。第三遍随后验证不同的 A_i、A_j、B_i和 B_j 之间的所有边是否符合分量划分所规定的邻接模式；任何违反规定的情况都会产生一个诱导 P_4。

最后请注意，即使 G 不是余图，该过程也可能返回一个分量划分。不过在这种情况下，P_4 的任何诱导副本都必须完全包含在单个部 A_i 或 B_i中。受此启发，我们的算法迭代地将 $V(G)$划分为部 A_i 和 B_j，并递归检查每个部是否诱导出一个余图。由于这些部在顶点上互不相交，因此可以在同一遍次中独立检查每个部，并且在任意阶段，所有活跃部所占的总空间都保持线性。

主要问题在于轮次数。如果最大部的大小接近 n，递归最多可能需要 $\Omega(n)$轮。为避免这种情况，我们修改该过程，使得每一轮中最大的剩余部至少缩小一个常数因子，从而使轮数的总体上界为 $O(\log n)$。此外，每一轮最多可以用 $O(\log n)$遍实现，因此总遍数最多为 $O(\log^2 n)$。

我们根据顶点的度将顶点划分为三个集合：

$$\mathrm{Low}(G) = \{v \in V(G) : \deg(v) < n/4\}$$

$$\mathrm{High}(G) = \{v \in V(G) : \deg(v) > 3n/4\}$$

$$\mathrm{Mid}(G) = \{v \in V(G) : n/4 \le \deg(v) \le 3n/4\}.$$

这种分类可以通过额外的一次遍次计算得出。如果 $\mathrm{Mid}(G) \neq \emptyset$，我们任取一个 $v \in \mathrm{Mid}(G)$作为根；那么每个 A_i 和 B_j 的大小都至多为 $3n/4$。为处理剩余情况，我们先证明 $\mathrm{Low}(G)$ 中的顶点无法构成一个较大的连通子图。

▶ **引理 29** ([24] 引理 2.3)。设 G 是一个余图。如果 F 是 G的一个连通子图，且 $V(F) \subseteq \mathrm{Low}(G)$，那么 $|V(F)| \le n/2$。

证明。 任取一个顶点 $x \in V(F)$ 作为根，并构造其分量划分。令 ℓ为满足 $A_\ell \cap V(F) \neq \emptyset$ 的最大索引，并固定一个顶点 $y \in A_\ell \cap V(F)$。由引理 28，不难看出

$$V(F) \subseteq \bigcup_{i=1}^{\ell} A_i \;\cup\; \bigcup_{i=1}^{\ell} B_i.$$

此外，根据划分的定义，$\bigcup_{i=1}^{\ell} A_i$ 中的每个顶点都与 x相邻，且 $\bigcup_{i=1}^{\ell} B_i$ 中的每个顶点都与 y相邻。因此，

$$|V(F)| \le \left|\bigcup_{i=1}^{\ell} A_i\right| + \left|\bigcup_{i=1}^{\ell} B_i\right| \le \deg(x) + \deg(y).$$

由于 $V(F) \subseteq \text{Low}(G)$，我们有 $\deg(x)$，$\deg(y) < \frac{n}{4}$，因此 $|V(F)| < n/2$。 ◀

因此，若 $\text{Low}(G) = V(G)$，则G 的每个连通分量大小至多为 $n/2$，这与 G 连通的假设矛盾。类似地，若 $\text{High}(G) = V(G)$，则 $V(G) = \text{High}(G) = \text{Low}(G)$。此外，$G$ 是余图当且仅当G 是余图。G 的一个连通分量大小至多为 $n/2$，这与G连通相矛盾。

—

剩下需要考虑的是 $\text{Mid}(G) = \emptyset$ 且 $\text{Low}(G)$ 和 $\text{High}(G)$ 均非空的情形。为处理这种情形，我们首先需要找到一个顶点 $u \in \text{Low}(G)$，使$|N(u) \cap \text{High}(G)|$ 最大化，并找到其分量划分；这可以通过 $O(1)$ 遍额外遍历完成。

▶**引理 30** ([24] 引理2.4)。假设存在一个索引 i 以及一个连通分量F属于导出子图 $G[B_i]$，使得 $|V(F)| > 3n/4$。那么 $V(F) \subseteq \text{High}(G)$。

证明。固定这样一个连通分量 $F \subseteq G[B_i]$，对于任意顶点 $v \in A_i$、$F \subseteq B_i \subseteq N(v)$，我们可以看到 $A_i \subseteq \text{High}(G)$。任取顶点 $x \in \text{High}(G) \setminus V(F)$。由于 $\deg(x) > 3n/4$ and$|V(F)| > 3n/4$，我们必有 $N(x) \cap V(F) \neq \emptyset$。根据假设，$x \notin B_i$。此外，根据引理 28 所保证的结构，$x$ 必须位于某个部 A_j 中，且满足 $j \geq i$。特别地，F 中的每个顶点都与 x 相邻，因此

$$\text{High}(G) \setminus V(F) \subseteq N(w) \qquad \forall w \in V(F).$$

反设 $V(F)$ 包含一个来自 $\text{Low}(G)$ 的顶点。由于 $|V(F)| > 3n/4$且 F 是连通的，引理 29 蕴含 $V(F) \not\subseteq \text{Low}(G)$。于是存在相邻顶点 $s \in V(F) \cap \text{Low}(G)$ 和 $t \in V(F) \cap \text{High}(G)$。根据上一段，$s$与 $\text{High}(G) \setminus V(F)$中的每个顶点都相邻，并且还与 t 相邻。因此，

$$|N(s) \cap \text{High}(G)| \geq |\text{High}(G) \setminus V(F)| + 1.$$

另一方面，根据以 u 为根的划分的构造，我们有 $V(F) \cap N(u) = \emptyset$。因此，$u$ 在 $\text{High}(G) \cap V(F)$ 内没有邻居，于是

$$|N(u) \cap \text{High}(G)| \leq |\text{High}(G) \setminus V(F)|.$$

这与选择使 $|N(\cdot) \cap \text{High}(G)|$ 相对于 $\text{Low}(G)$ 最大化的 u 矛盾。因此，$V(F) \cap \text{Low}(G) = \emptyset$，由此可得 $V(F) \subseteq \text{High}(G)$。◀

在获得相对于 v 的分量划分后，我们进一步将每个A_i 和 B_i 分解为导出子图 $G[A_i]$ 和 $G[B_i]$的连通分量。由于$|A_i| \leq n/4$ 对所有 i均成立，若不存在引理 (F, B_i) 30 所述的数对，则每个由此得到的部的大小至多为 $3n/4$。

否则，令 F如引理 30所述。则 $V(F) \subseteq \text{High}(G) = \text{Low}(G)$。根据引理 29，$\text{Low}(G)$ 中的顶点所构成的连通子图 G 不能超过 $n/2$ 个顶点。因此，考虑导出子图 $G[V(F)]$（等价地，补图 F）时，每个连通分量的大小至多为 $n/2$。我们可以再应用一次引理 25 中的过程来得到这一分解。

定理 27 的证明。 为简便起见，我们只证明该问题的判定版本；若 G 包含一个诱导 P_4，则可通过追踪算法返回FALSE时所执行的步骤发现它。所述算法总结于算法1中，用于判断一个图是否为余图，也就是说，是否不含诱导 P_4。如上文讨论所示，在所有情况下，该过程都会在 $O(\log n)$ 遍数内，将

$V(G)$ 划分为大小至多为 $3n/4$ 的部，并具有这样的性质：任何诱导的 co-P_4 副本（如果存在）都完全包含在单个部中。对所有部递归地应用相同的过程（每一遍并行执行），可得到一个使用 $\widetilde{O}(n)$ 空间的 $O(\log^2 n)$ 遍算法，用于判定 G 是否为余图。◀

▶ 备注 31。尽管引理 25 对查找诱导的余-P_3 和余-P_4有效，但当 $k \geq 5$ 时，它无法用于检测诱导的余-P_k 。这是因为，对于更大的 k ，排除诱导的余-P_k 的图，可能不具备与情形 $k = 3$、4 同样强的结构性质。即使经过可能的预处理步骤，算法仍可能以一个大小接近 n 的 G 连通分量结束。事实上，如命题 21 所示，判定是否存在诱导的余 -P_k ，对于 $k \geq 5$ 而言空间开销很大。

5 定向图

我们转向定向子图查找问题。尽管有些结果与无向情形相似，但也出现了一些有趣的新现象和开放问题，尤其是对于 NWO$^{\rightarrow}$二分图 H，正如定理 6 和 7所述，尽管它们的 ex 均为 $(n, \vec{H}) = \Theta(n^2)$。

5.1 从无向版本归约

首先，给定一个固定定向图 $\vec{H}$，子图查找问题 $\textsc{Sub}(H)$ 可以归约为 $\textsc{Sub}(\vec{H})$。从 $\textsc{Indsub}(H)$ 到 $\textsc{Indsub}(\vec{H})$ 也成立相同的结果，因此我们在此陈述并证明一般版本，该版本稍后也将应用于区段 6。

命题 32。 设 H 是一个固定无向图， H 是 H的任意定向。假设 $R^p_{2/3}(Sub(\vec{H})) = O(f(n,p))$，令 $m = |E(H)| = O(1)$，则

$$R^{pm2^m}_{2/3}(\textsc{Sub}(H)) = O(f(n,p) + m \log n)$$

从 $Indsub(H)$ 到 $Indsub(\vec{H})$ 的相同归约也成立。

证明。设 $R^p_{2/3}(\text{Sub}(\vec{H})) = O(f(n,p))$ 由算法 A 达到。给定一个无向$^{\rightarrow}$输入图 G ，其具有 n 个顶点，我们通过为每条边随机指定定向来生成一个定向图 G 。为了在数据流模型的不同遍次中保持相同的定向，我们采样一个 m-独立哈希函数；对于每条边 $\{a,b\}$，该函数决定将其定向为 $a \to b$ 还是 $b \to a$。该哈希函数可以使用有限域上 m-独立哈希的经典多项式方法构造（参见例如 [7]，其中将该结果归因于 [15, 28]），所使用的随机种子大小为 $O(m\log\ n)$。

步骤 1：降低$\mathcal{A}$的错误

$^{\rightarrow}$固定一个特定的定向 G。通过标准放大，我们可以在同一数据流上独立运行$\mathcal{A}$ $r = O(m)$ 次，并进行多数表决。将所得算法记为 $\mathcal{A}'$。选择 $r = O(m)$ 可将$\mathcal{A}'$ 的错误概率降低至最多

$$\delta := \frac{2^{-m}}{12}.$$

这会使遍数增加 $O(m)$，同时将空间使用量保持在 $f(n,p)$。

步骤 2：重复进行随机定向。

我们重复以下实验

$$T := 2^{m+1}$$

独立试验。在每次试验中，我们抽取一个新的随机定向 $\vec{G}$ 并在 $\vec{G}$ 上运行 $\mathcal{A}'$ 。当且仅当至少有一次试验输出 YES 时，我们才输出 YES。

正确性： 首先假设 G 中包含 H 的一个（诱导）副本。在 G 中固定这样一个副本。由于该副本有 m 条边，在 m 阶独立的随机定向下，其所有边的定向都与 H 一致的概率恰好为 2^{-m}。因此，每次试验中，$\vec{G}$ 包含 $\vec{H}$ 的一个（诱导）副本的概率至少为 2^{-m}。于是，T 次试验中没有任何一次产生包含 H 的定向的概率至多为

$$(1-2^{-m})^T \leq e^{-2} < \frac{1}{6}.$$

此外，根据并集界，$\mathcal{A}'$ 在 T 次试验中的任意一次出错的概率至多为

$$T\delta = 2^{m+1} \cdot \frac{2^{-m}}{12} = \frac{1}{6}.$$

因此，YES 实例上的总体失败概率至多为 $1/6 + 1/6 = 1/3$。

接下来，假设 G 不包含 H 的（诱导）副本。那么对于每个定向$\vec{G}$，图 $\vec{G}$ 都不可能包含 $\vec{H}$ 的（诱导）副本。因此，我们的算法输出 YES 的唯一可能情形是 $\mathcal{A}'$ 在某次试验中出错，其发生概率至多为 $T\delta = 1/6$。特别地，在 NO 实例上的成功概率至少为 $5/6$ 。

复杂度： 每次试验运行 $\mathcal{A}'$ ，进行 p 遍次，并重复 $O(m)$ 次，即每次试验进行 $O(pm)$ 遍次。共进行 $T = 2^{m+1}$ 次试验，遍数总计为 $O(pm2^m)$。运行 $\mathcal{A}$ 所需空间为 $f(n, p)$，此外还需 $O(m \log n)$ 空间来存储定义随机定向的哈希种子。◀

结合命题 32 和定理 1 可知，所有非二分图 H 都是困难实例，其 $R^p_{2/3}(\text{Sub}(\vec{H})) = \Omega(\text{ex}(n, \vec{H})/p)$。另一方面，对于 WO 二分图 $\vec{H}$，根据推论 11，$\text{ex}(n, \vec{H}) = O(n^{2-1/\Delta'(H)})$；因此，与非二分定向图相比，确实存在多项式间隔，这与定理 3 中的二部图情形相同。以下总结这些结果。

推论 33. 令 H 为一个固定定向图。

- 如果 $\vec{H}$ 是非二分图，那么 $R^p_{2/3}(\textit{Sub}(\vec{H})) = \Omega(n^2/p)$。否
- 则，如果 $\vec{H}$ 是 *WO*，那么 $D^1(\textit{Sub}(\vec{H})) = \tilde{O}(n^{2-1/\Delta'(H)})$。

然而，NWO 二分图 $\vec{H}$ 的 $\text{ex}(n, \vec{H}) = \Theta(n^2)$ 则有所不同。这类图才是有趣的情形。已知其图兰数，我们可能会问 $R^p_{2/3}(\text{Sub}(\vec{H})) = \Omega(n^2/p)$ 是否成立，或者至少 $D^1(\text{Sub}(\vec{H})) = \Omega(n^2)$ 是否成立。

根据定理6 和7，这些下界确实适用于某些 NWO 二分图 H，但也存在反例。例如，任何 NWO 森林都存在高效的常数遍算法，而能够通过高效单遍算法处理的 NWO 二分图，恰好是每个连通分量要么为 WO，要么为恰有一个 NWO 顶点的树的那些图。

5.2 非良定向二分图：多遍

我们首先考虑多遍设置，逐步证明定理6 。首先，我们证明 NWO 偶圈是困难实例，其空间下界为 p-遍 $\Omega(n^2/p)$，与非二分图的情况相同。证明通过从集合不相交问题进行归约来完成，与命题16的证明类似，但这次我们可以取两个 WO $K_{n,\ n}$ 副本来编码输入，这使我们能够达到 $\Omega(n^2/p)$ 的空间下界。

命题 34. 设 H 为固定的 *NWO* 二分图。若对某个 $\ell \geq 2$ 有 $H = C_{2\ell}$，则

$$R^p_{2/3}(SUB(\vec{H})) = \Omega(n^2/p).$$

证明。 将 H 视为平面图，不失一般性，假设至少有三条边在 H 中具有逆时针定向。将 H 的顶点标记为

$$v_1,\ v_2,\ \ldots\ldots,\ v_{2\ell},\ v_{2\ell+1} = v_1$$

按逆时针顺序。不失一般性，假设对某个 $3 \leq k \leq 2\ell-1$，$(v_1,\ v_2)$，$(v_k,\ v_{k+1}) \in E(\vec{H})$。给定不相交性问题$_{n^2}$的一个实例，其中 Alice 与 Bob 分别持有 $S_1,\ S_2 \subseteq [n^2]$，我们将有向归约图 $G_{\vec{H}}$ 定义在顶点集

$$V(\vec{G}_{\vec{H}}) = \bigsqcup_{i=1}^{n} \left\{x_i, y_i, z_i, w_i\right\} \sqcup \bigsqcup_{i=1}^{n} \left\{v_3^{(i)}, v_4^{(i)}, \ldots, v_{k-1}^{(i)}, v_{k+2}^{(i)}, v_{k+3}^{(i)}, \ldots, v_{2\ell}^{(i)}\right\}$$

并且弧集为 $E(\vec{G}_{\vec{H}}) = E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$，其中

$$E_1 = \{(x_i, y_j) : i, j \in [n], (j-1)n + i \in S_1\} \cup \{(w_j, z_i) : i, j \in [n], (j-1)n + i \in S_2\},$$

$$E_2 = \bigsqcup_{i=1}^{n} \left\{(v_j^{(i)}, v_{j+1}^{(i)}) : j \in \{3, 4, \ldots, k-2, k+2, k+3, \ldots, 2\ell\}, (v_j, v_{j+1}) \in E(\vec{H})\right\},$$

$$E_3 = \bigsqcup_{i=1}^{n} \left\{(v_{j+1}^{(i)}, v_j^{(i)}) : j \in \{4, 5, \ldots, k-1, k+3, k+4, \ldots, 2\ell+1\}, (v_{j+1}, v_j) \in E(\vec{H})\right\},$$

$$E_4 = \begin{cases} \{(y_i, w_i) : 1 \leq i \leq n\}, \text{ if } k = 3, \\ \{(y_i, v_3^{(i)}), (v_{k-1}^{(i)}, w_i) : 1 \leq i \leq n\}, \text{ if } k > 3, \end{cases}$$

$$E_5 = \begin{cases} \{(z_i, x_i) : 1 \leq i \leq n\}, \text{ if } k = 2\ell - 1, \\ \{(z_i, v_{k+2}^{(i)}), (v_{2\ell}^{(i)}, x_i) : 1 \leq i \leq n\}, \text{ if } k < 2\ell - 1. \end{cases}$$

由于我们有 $|V(\vec{G}_{\vec{H}})| = \Theta(n)$，以下断言即可完成证明。

断言： 当且仅当 $S_1 \cap S_2 \neq \emptyset$ 时，我们有 $\vec{H} \subseteq \vec{G}_{\vec{H}}$。

证明。 若 $(j-1)n + i \in S_1 \cap S_2$ 对于某个 $i,\ j \in [n]$，那么 $(x_i,\ y_j)$，$(w_j,\ z_i) \in E(\vec{G}_{\vec{H}})$，因此分别对应顶点 $v_1,\ v_2,\ \ldots,\ v_{2\ell-1}$，一个 H 副本由以下方式形成

$$x_i, y_j, v_3^{(j)}, v_4^{(j)}, \ldots, v_{k-1}^{(j)}, w_j, z_i, v_{k+1}^{(i)}, \ldots, v_{2\ell-1}^{(i)}.$$

反过来，由于导出子图 $\vec{G}_{\vec{H}}[E_1]$ 是 WO，而 $\vec{H}$ 不是，若 $\vec{G}_{\vec{H}}$ 中存在一个 $\vec{H}$ 的副本 $\vec{H}'$，那么 $\vec{H}'$ 会使用一些不属于 E_1 的弧，因此要闭合成一个环，它包含路径 $y_j - v_3^{(j)} - v_4^{(j)} - \ldots - v_{k-1}^{(j)} - w_j$ 和 $z_i - v_{k+1}^{(i)} - \ldots - v_{2\ell-1}^{(i)} - x_i$，其中 i 和 j 取某些值。但这两条路径的长度之和为 $2\ell - 2$，因此要构成 $\vec{H}'$，我们得到 $(x_i,\ y_j)$，$(w_j,\ z_i) \in E(\vec{G}_{\vec{H}})$，从而得到 $S_1 \cap S_2 \neq \emptyset$。 ◁

接下来，我们通过从集合不相交问题进行归约，证明 NWO 完全二分图也是困难实例。粗略地说，我们让两个 WO $K_{n,\ n}$ 在其两个部中的一个部重叠，形成一个 WO $K_{n,\ 2n}$。其中一个 $K_{n,\ n}$ 的边集将用于编码 Alice 的输入，另一个则用于编码 Bob 的输入，使得同一个元素对应一对对称边。接着，我们在每个水平层上附加一个 H 的额外副本，并选择某些 $v \in V(\vec{H})$，使其两条关联弧与该层上的水平索引弧重叠。为了使归约有效，必须谨慎选择 v，并且还需要一些交叉弧。

命题 35。 固定 $2 \le s \le t$。设 H 是一个 *NWO* 二分图，其中 $H = K_{s,\ t}$，则

$$R^{p}_{2/3}(\textsc{Sub}(\vec{H})) = \Omega(n^2/p)$$

证明。 令 $\vec{H}$ 的两个部为 $A = A(\vec{H})$ 和 $B = B(\vec{H})$，其中 $|A| = s$ 和 $|B| = t$。由于情形 $s = t$ 较为简单，我们不妨假设 $s < t$。我们始终给定一个 不相交性问题 $_{n^2}$ 实例，其中 Alice 和 Bob 分别持有 $S_1,\ S_2 \subseteq [n^2]$并构造有向归约图 $\vec{G}_{\vec{H}}$，使得 $|V(\vec{G}_{\vec{H}})| = \Theta(n)$。

情形 1： 在 H 中有一些 $v \in A$ 是 WO，且 B 中没有顶点是 WO。

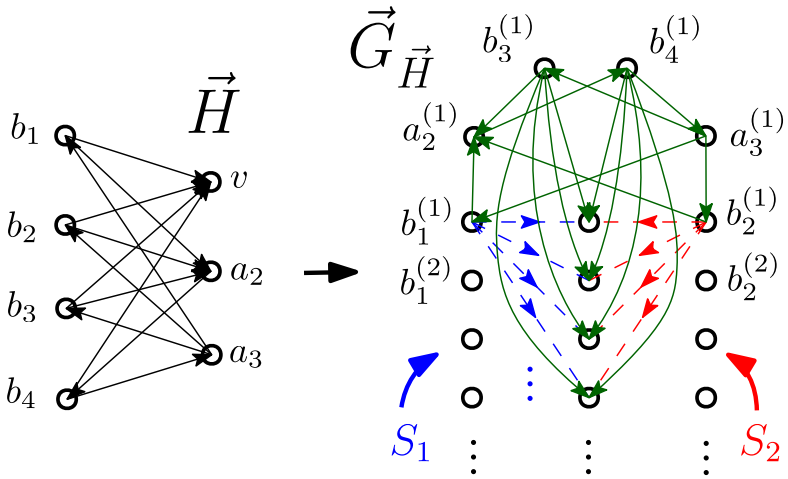

图 4 归约图 $\vec{G}_{\vec{H}}$，对应命题 35 中的情形 1。图中仅展示了 $a_2^{(i)}$、$a_3^{(i)}$、$b_3^{(i)}$、$b_4^{(i)}$ 对应的 $i = 1$，但所有 $i \in [n]$也应存在相应的结构。

不失一般性地，假设 v的出度为 0。令 $A = \{v, a_2, \ldots, a_s\}$和 $B = \{b_1, \ldots, b_t\}$。定义 $\vec{G}_{\vec{H}}$的顶点集为

$$V(\vec{G}_{\vec{H}}) = \bigsqcup_{i=1}^{n} \left\{ b_1^{(i)}, v^{(i)}, b_2^{(i)} \right\} \sqcup \bigsqcup_{i=1}^{n} \left\{ a_2^{(i)}, a_3^{(i)}, \ldots, a_s^{(i)}, b_3^{(i)}, b_4^{(i)}, \ldots, b_t^{(i)} \right\}$$

以及弧集 $E(\vec{G}_{\vec{H}}) = E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$，其中

$$E_1 = \{(b_1^{(i)}, v^{(j)}) : i,j \in [n], (j-1)n+i \in S_1\},$$
$$E_2 = \{(b_2^{(i)}, v^{(j)}) : i,j \in [n], (j-1)n+i \in S_2\},$$
$$E_3 = \bigsqcup_{i=1}^{n} \left\{(a_k^{(i)}, b_\ell^{(i)}) : 2 \le k \le s, 1 \le \ell \le t, (a_k, b_\ell) \in E(\vec{H})\right\},$$
$$E_4 = \bigsqcup_{i=1}^{n} \left\{(b_\ell^{(i)}, a_k^{(i)}) : 2 \le k \le s, 1 \le \ell \le t, (b_\ell, a_k) \in E(\vec{H})\right\},$$
$$E_5 = \bigsqcup_{i,j\in[n]} \left\{(b_3^{(i)}, v^{(j)}), \ldots, (b_t^{(i)}, v^{(j)})\right\}.$$

与之前一样，若 $S_1 \cap S_2 \neq \emptyset$，则对于某个 i，$j \in [n]$，$(b_1^{(i)}, v^{(j)})$，$(b_2^{(i)}, v^{(j)}) \in E(\vec{G}_{\vec{H}})$，因此，分别对应于顶点 a_1，a_2、…、a_s，b_1、…、b_t的¯构成了 H的一个副本，其构造方式为

$$v^{(j)}, a_2^{(i)}, \ldots, a_s^{(i)}, b_1^{(i)}, \ldots, b_t^{(i)}.$$

反过来，由于导出子图 $\vec{G}_{\vec{H}}[E_1 \cup E_2]$ 是WO，而 $\vec{H}$ 不是，若 $\vec{G}_{\vec{H}}$ 中存在 $\vec{H}$ 的一个副本 $\vec{H}'$，则 $\vec{H}'$ 会从$\{a_2^{(i)}, a_3^{(i)}、\ldots、a_s^{(i)}, b_3^{(i)}, b_4^{(i)}、\ldots、b_t^{(i)}\}$中恰好取用一个 的某个顶点，而 的唯一性来自，$t$。

首先假设所有 $a_2^{(i)}, \ldots, a_s^{(i)} \notin V(\vec{H}')$，则 $\vec{H}'$包含在由$\{b_3^{(i)}, \ldots, b_t^{(i)}\}$与$\{b_1^{(j)}, v^{(j)}, b_2^{(j)}\}$对所有 j所诱导的 $\vec{G}_{\vec{H}}$的子图中。但后者是WO，而 $\vec{H}'$不是，这是一个矛盾。因此，对某个 $2 \le k \le s$，我们有 $a_k^{(i)} \in V(\vec{H}')$。

如果 $a_k^{(i)} \in B(\vec{H}')$，则由于 $s < t = |B(\vec{H}')|$，我们需要某个不同于 $a_2^{(i)}, \ldots, a_s^{(i)}$的顶点属于 $B(\vec{H}')$，而这只能是某个 $v^{(j)}$。然而，$v^{(j)}$在 $\vec{G}_{\vec{H}}$中是WO，而根据假设，$B(\vec{H}')$中的任何顶点在 $\vec{H}'$中都不应是WO，这是一个矛盾。因此 $a_k^{(i)} \in A(\vec{H}')$，进而由于 $d_{\vec{G}_{\vec{H}}}(a_k^{(i)}) = t = |B(\vec{H}')|$，有 $N_{\vec{G}_{\vec{H}}}(a_k^{(i)}) = B(\vec{H}')$。特别地，$b_1^{(i)}, b_2^{(i)} \in B(\vec{H}')$，因此除了 $a_2^{(i)}, \ldots, a_s^{(i)}$之外，$b_1^{(i)}, b_2^{(i)}$还有一个公共邻点，它必定是某个 $v^{(j)}$，由此可得 $S_1 \cap S_2 \neq \emptyset$。

¯**情况2：** 某个 $v \in B$ 在 H中是良好定向的。

令 $A = \{a_1, a_2, \ldots, a_s\}$和 $B = \{v, b_2, \ldots, b_t\}$。定义 $\vec{G}_{\vec{H}}$的顶点集为

$$V(\vec{G}_{\vec{H}}) = \bigsqcup_{i=1}^{n} \left\{a_1^{(i)}, v^{(i)}, a_2^{(i)}\right\} \sqcup \bigsqcup_{i=1}^{n} \left\{a_3^{(i)}, \ldots, a_s^{(i)}, b_2^{(i)}, \ldots, b_t^{(i)}\right\}$$

以及弧集 $E(\vec{G}_{\vec{H}}) = E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$，其中

$$E_1 = \{(a_1^{(i)}, v^{(j)}) : i,j \in [n], (j-1)n+i \in S_1\},$$
$$E_2 = \{(a_2^{(i)}, v^{(j)}) : i,j \in [n], (j-1)n+i \in S_2\},$$
$$E_3 = \bigsqcup_{i=1}^{n} \left\{(a_k^{(i)}, b_\ell^{(i)}) : 1 \le k \le s,\ 2 \le \ell \le t, (a_k, b_\ell) \in E(\vec{H})\right\},$$
$$E_4 = \bigsqcup_{i=1}^{n} \left\{(b_\ell^{(i)}, a_k^{(i)}) : 1 \le k \le s,\ 2 \le \ell \le t, (b_\ell, a_k) \in E(\vec{H})\right\},$$
$$E_5 = \{(a_3^{(i)}, v^{(j)}), \ldots, (a_s^{(i)}, v^{(j)}) : i \neq j\}.$$

与情形 1 中类似的论证表明，$\vec{H} \subseteq \vec{G}_{\vec{H}}$ 当且仅当 $S_1 \cap S_2 \neq \emptyset$。也就是说，如果 $\vec{G}_{\vec{H}}$ 中存在 $\vec{H}$ 的一个副本 $\vec{H}'$，那么存在唯一的 $i \in [n]$，使得 H' 使用某个顶点$b_\ell^{(i)}$。于是 $b_\ell^{(i)} \in B(\vec{H}')$，因此 $a_1^{(i)}$，$a_2^{(i)} \in A(\vec{H}')$，并且它们必定有某个 $v^{(j)}$ 作为公共邻点。

情形 3： H中没有 WO 顶点。

任取 $v \in B$，并令 $A = \{a_1, \ldots, a_s\}$、$B = \{v, b_2, \ldots, b_t\}$，其中 $(a_1, v), (v, a_2) \in E(\vec{H})$。定义 $\vec{G}_{\vec{H}}$，方式与情况2相同，但在 E_2 中，我们将 $(a_2^{(i)}, v^{(j)})$ 替换为 $(v^{(j)}, a_2^{(i)})$。

同样，若 $S_1 \cap S_2 \neq \emptyset$，则 $\vec{H} \subseteq \vec{G}_{\vec{H}}$。反过来，设 $\vec{G}_{\vec{H}}$ 中有一个 $\vec{H}$ 的副本 $\vec{H}'$。在导出子图 $\vec{G}_{\vec{H}}[E_1 \cup E_2]$ 中，顶点 $a_1^{(i)}$、$a_2^{(i)}$ 是 WO 顶点，而 $\vec{H}$ 中没有顶点是 WO 顶点，因此 $\vec{H}'$ 再次使用了 $\{a_3^{(i)}, \ldots, a_s^{(i)}, b_2^{(i)}, \ldots, b_t^{(i)}\}$ 中的某个顶点。若所有 $b_2^{(i)}, \ldots, b_t^{(i)} \notin V(\vec{H}')$ 都成立，则 $\{a_3^{(i)}, \ldots, a_s^{(i)}\}$ 中的任意顶点 $\cap V(\vec{H}')$ 都会是 $\vec{H}'$ 中的 WO 顶点，这与假设矛盾。因此，对某个 $2 \leq k \leq t$，$b_k^{(i)} \in V(\vec{H}')$，于是 $N_{\vec{G}_{\vec{H}}}(b_k^{(i)}) = A(\vec{H}')$，因为 $d_{\vec{G}_{\vec{H}}}(b_k^{(i)}) = s = |A(\vec{H}')|$。这会导出 $a_1^{(i)}$、$a_2^{(i)}$ 的一个公共邻点，且该邻点不属于 $b_2^{(i)}, \ldots, b_t^{(i)}$，因此只能是某个 $v^{(j)}$，从而证明 $S_1 \cap S_2 \neq \emptyset$。◀

上面两个命题的结果都表明，$\mathsf{ex}(n, \vec{H})$ 决定了空间复杂度；但正如定理6的后一部分所示，如果 H 是森林，那么无论 $\vec{H}$如何定向，总存在一个常数遍次的 $\tilde{O}(n)$空间算法。这意味着，尽管 $\mathsf{ex}(n, \vec{H}) = \Theta(n^2)$，$H$ 的结构非常简单，因此即使定向任意，我们仍然可以在只存储少量信息的情况下检测出 H 。

证明仍然采用颜色编码。粗略地说，我们像引理$V(\vec{G})$ 14中那样，选择一组的着色，使得如果 $\vec{H}$的副本存在，那么在其中一种着色下必定是彩色的。然后，我们可以使用动态规划，迭代地记录一个顶点$v \in V(\vec{G})$ 是否能作为 $\vec{H}$ 的一棵子树的根，且该子树的颜色属于指定颜色集，从而最终检测出。 H

命题 36. 设 H 为一片固定的森林， H 为 H的任意定向。设 r 为 H各连通分量中最大的半径，则 $D^{2r}(\mathit{Sub}(\vec{H})) = \tilde{O}(n)$。

证明。设 $H_1, \ldots, H_k$为 H的连通分量，其半径分别为 $r_1, \ldots, r_k$。对于每个 $\vec{H}_i$，从 H_i 的中心中选择 v_i 作为根。于是，对所有 $x \in V(\vec{H}_i)$，定义 $h(x) = r_i - \mathsf{dist}_{H_i}(v_i, x)$。此外，记 $\vec{T}_x$ 为以 x为根的 $\vec{H}_i$ 子树。

根据引理 14，我们可以取 $(\log^2 n)$ 个着色 $\chi_1, \ldots, \chi_L \colon V(\vec{G}) \to [|V(\vec{H})|]$，使得若 G 包含 H的一个副本，则该副本在其中一种着色下是彩色的。这需要空间 $O(Ln) = \tilde{O}(n)$。现在，对所有 $x \in V(\vec{H})$， $c \in [L]$ 以及满足 $|S| = |\vec{T}_x|$的所有 $S \subseteq [|V(\vec{H})|]$，定义数组 $D_x^c[S] \colon V(\vec{G}) \to \{0, 1\}$，并将其初始化为 $D_x^c[S](v) = 0$，其中所有 $v \in V(\vec{G})$ 均如此。这需要空间 $\tilde{O}(|V(\vec{H})| L 2^{|V(\vec{H})|} n) = \tilde{O}(n)$，我们的目标是使其满足

$$D_x^c[S](v) = 1 \iff \exists \text{ embedding } \varphi : \vec{T}_x \hookrightarrow \vec{G},\ x \mapsto v,\ \varphi(\vec{T}_x) \text{ uses color } S \text{ under } \chi_c.$$

另一方面，对于所有如前所述的 x, c, S，如果 x 有父节点 y，那么对于所有 $s \in [|V(\vec{H})|] \setminus S$，我们还声明一个数组 $W_{y,x}^c[S, s] \colon V(\vec{G}) \to \{0, 1\}$，将其初始化为 $W_{y,x}^c[S, s](v) = 0$，对所有$v \in V(\vec{G})$均如此。此操作使用空间 $\tilde{O}(n)$，我们的目标是使 $W_{y,x}^c[S, s](u) = 1$ 当且仅当

$$\chi_c(u) = s, \text{ and}$$

$$\exists v \in V(\vec{G}) \text{ such that } D_x^c[S](v) = 1 \text{ with } \begin{cases} (v, u) \in E(\vec{G}), & \text{if } (x, y) \in E(\vec{H}), \\ (u, v) \in E(\vec{G}), & \text{if } (y, x) \in E(\vec{H}). \end{cases}$$

然后我们开始进行动态规划，从 $j=1$ 开始。

(1) 对于每个满足 $|S|=1$ 的叶节点 $x\in V(\vec{H})$、$c\in[L]$ 和 $S\subseteq[|V(\vec{H})|]$，当且仅当 $S=\{\chi_c(u)\}$ 时，我们设定 $D_x^c[S](u)=1$。这就确定了所有满足 $h(x)=0$ 的 x 的 $D_x^c[S]$。

(2) 在第 j 遍之前，对于所有满足 $h(x)\le j-1$ 的 x，$D_x^c[S]$ 已经确定。于是，在第 j 遍中，每当一条弧 $\vec{e}=(u,v)$ 到达时，我们检查所有满足 $h(x)=j-1$ 的 $x\in V(\vec{H})$、$c\in[L]$、满足 $|S|=|\vec{T}_x|$ 的 $S\subseteq[|V(\vec{H})|]$，以及 $s\in[|V(\vec{H})|]\setminus S$。令 y 为 x 的父节点，我们检查是否

$$D_x^c[S](u)=1,\quad \chi_c(v)=s,\quad (x,y)\in E(\vec{H}).$$

若情况如此，则我们设置 $W^c_{y,\,x}[S,s](v)=1$。同样，我们也检查是否

$$D_x^c[S](v)=1,\quad \chi_c(u)=s,\quad (y,x)\in E(\vec{H}).$$

若情况如此，则我们设置 $W^c_{y,\,x}[S,s](u)=1$。

(3) 在第 j 遍次之后，对于所有满足 $h(x)=j-1$ 且 y 是 x 的父节点的 x，$W^c_{y,\,x}[S,s]$ 已确定。现在，对于每个满足 $h(y)=j$ 的 $y\in V(\vec{H})$，令 y 的子节点为 x_1、……、x_i。对于所有 $u\in V(\vec{G})$，如果存在 $c\in[L]$、$s\in[|V(\vec{H})|]$，以及两两不相交的 S_1、……、$S_i\subseteq[|V(\vec{H})|]\setminus\{s\}$，使得

$$W^c_{y,x_1}[S_1,s](u)=\ldots=W^c_{y,x_i}[S_i,s](u)=1,$$

那么我们设置 $D_y^c[S_1\cup\ldots\cup S_i\cup\{s\}](u)=1$。这样便确定了所有满足 $h(x)=j$ 的 x 的 $D_x^c[S]$。现在，如果 $j=r$，则继续下一步。否则，将 j 加 1，并从步骤 (2) 开始重复。

（4）现在，对于所有 $x\in V(\vec{H})$，都已知 $D_x^c[S]$。特别地，我们知道 $D^c_{v_1}[S]$、……、$D^c_{v_k}[S]$。如果 $\vec{H}\subseteq G$，那么在某个 χ_c 下会存在一个 H 的彩色副本。因此，如果存在一些$c\in[L]$ 以及互不相交的 S_1、……、$S_k\subseteq[|V(\vec{H})|]$，使得

$$D^c_{v_1}[S_1](u_1)=\ldots=D^c_{v_k}[S_k](u_k)=1$$

对于某些 u_1、……、$u_k\in V(\vec{G})$，那么我们便得出 $\vec{H}\subseteq\vec{G}$。否则 $\vec{H}\nsubseteq G$。

（5）如果 $\vec{H}\subseteq\vec{G}$，那么我们可以找到一个 $\vec{H}$，使用 r 额外遍次和 $\tilde{O}(|V(H)|n)=\tilde{O}(n)$ 额外空间；这与命题 15 证明中的步骤（3）类似，因此我们仅简要说明$\vec{H}\subseteq\vec{G}$证明如下。现在，我们已知存在一个副本 $H\subseteq G$ 由以下部分给出

$$\vec{H}\hookrightarrow\vec{G},\quad v_1\mapsto u_1,\ldots,v_k\mapsto u_k,$$

在第一个额外遍次中，我们存储与 $u_1,\ldots,u_k$ 相 incident 的所有弧。对于所有 $i\in[k]$，设 v_i 的子节点为 $v_{i,1},\ldots,v_{i,k_i}$，那么我们可以找到一些不相交的

$$Z_{i,1},\ldots,Z_{i,k_i}\subseteq[|V(\vec{H})|],\ Z_{i,1}\sqcup\ldots\sqcup Z_{i,k_i}=S_i\setminus\{\chi_c(u_i)\},\ |Z_{i,j}|=|\vec{T}_{v_{i,j}}|\ \forall j\in[k_i]$$

以及某些 $u_{i,\,1}$，……，$u_{i,\,k_i}\in V(\vec{G})$，使得

$$D^c_{v_{i,1}}[Z_{i,1}](u_{i,1})=\ldots=D^c_{v_{i,k_i}}[Z_{i,k_i}](u_{i,k_i})=1\quad\forall i\in[k],$$

因此我们知道存在一个由以下给出的副本 $H \subseteq G$

$$\vec{H} \hookrightarrow \vec{G}, \quad v_1 \mapsto u_1, \ldots, v_k \mapsto u_k, \quad v_{i,1} \mapsto u_{i,1}, \ldots, v_{i,k_i} \mapsto u_{i,k_i} \quad \forall i \in [k].$$

接着，在下一遍中，我们存储与 $v_{i,1}$、……、v_{i,k_i} 相连的所有弧；对 $i \in [k]$、等也依此处理。这样便能在 r 遍额外遍历内恢复一个 H 的副本。◀

一般来说，给定任意 NWO 二分图 H，我们尚无法确定它究竟是满足 $R^p_{2/3}(\text{Sub}(\vec{H})) = \Omega(n^2/p)$ 的困难实例，还是存在空间高效的多遍算法，尽管我们倾向于认为，只要 H 的某个 NWO 分量包含一个环，它就是困难实例。不过，对于单遍情形，我们了解得更多，并且在下一小节中给出了完整的二分结果。

5.3 非良定向二分图：单遍

我们研究了 NWO 二分图的单轮下界，所得结果由定理 7 总结。结果表明，大多数非良定向二分图 H 都是满足 $R^1_{2/3}(\text{Sub}(\vec{H})) = \Omega(n^2)$ 的困难实例。唯一的简单实例是 $\vec{H}$ 其中每个连通分量要么是 WO 二分图，要么是一棵恰好包含一个 NWO 顶点的树。

定理7 中的下界将分步证明。首先，我们证明 H 只要其中一个连通分量包含至少 2 个 NWO 顶点，就是一个困难实例。为此，我们需要以下关于临界路径的定义。

▶ **定义 37**（临界路径）。设 P：$v_1 v_2 \ldots v_\ell$ 是一个包含 $\ell \geq 5$ 个顶点的路径，$\vec{P}$ 是 P 的一个定向。若在 v_1、v_2、……、v_ℓ 中，只有 v_2 和 $v_{\ell-1}$ 在 P 中是 *NWO* 顶点，则称 P 为一条临界路径。例如，以下两者都是临界路径。

$$\vec{P} : v_1 \to v_2 \to v_3 \leftarrow v_4 \to v_5 \leftarrow v_6 \to v_7 \leftarrow v_8 \to v_9 \to v_{10},$$
$$\vec{P'} : v_1 \leftarrow v_2 \leftarrow v_3 \to v_4 \to v_5$$

请注意，在临界路径 P 的定义中，v_1，v_3，v_4，……，$v_{\ell-2}$，v_ℓ 相对于 P 是良定向的。因此，当我们说 H 包含一条临界路径 P：$v_1 v_2 \ldots v_\ell$ 时，顶点 v_1，v_3，v_4，……，$v_{\ell-2}$，v_ℓ 在 H 中可能不是 WO 顶点，但在 P 中一定是 WO 顶点。

接下来我们需要引理 38，该引理指出，如果 H 的某个连通分量至少包含 2 个 NWO 顶点，那么该连通分量包含一个 NWO 圈、一条有向路径 P_4 或一条临界路径。然后在命题 39 中，我们证明 H 一旦包含这三种结构中的一种，就是一个困难实例。

▶**引理 38**。设 H 是一个定向图。如果 H 有一个连通分量至少包含2 个 *NWO* 顶点，那么 H 包含一个 *NWO* 圈、一条有向 P_4 或一条临界路径。

证明。设 u、$v \in V(\vec{H})$ 是 $\vec{H}$ 的同一连通分量中的不同 NWO 顶点。根据定义，存在 u_1、u_2、v_1、$v_2 \in V(\vec{H})$ 使得 $\vec{H}$ 包含这些有向路径

$$u_1 \to u \to u_2, \quad v_1 \to v \to v_2.$$

如果$\{u_1, u_2\} = \{v_1, v_2\}$，那么我们就得到一个 NWO 圈。否则，如果 $u_1 = v_2$ 或 $u_2 = v_1$，那么我们就得到一条有向 P_4；如果 $u_1 = v_2$ 或 $u_2 = v_2$，那么我们就得到一条包含 5 个顶点的临界路径。因此，我们可以假设 u_1，u_2，v_1，v_2 两两不同。在这种情况下，由于 u 和 v 位于同一个连通分量中，存在一条定向路径 P，其底层路径是

$$P : u - w_1 - w_2 - \ldots - w_\ell - v, \quad w_1, w_2, \ldots, w_\ell \in V(\vec{H}).$$

如果$\{u_1,\ u_2\}\cap\{w_2,\ \ldots,\ w_\ell\}\neq\emptyset$或$\{v_1,\ v_2\}\cap\{w_1,\ \ldots,\ w_\ell\}\neq\emptyset$，那么我们就得到一个 NWO 圈，因此可以假设

$$\{u_1,u_2\}\cap\{w_2,\ldots,w_\ell\}=\{v_1,v_2\}\cap\{w_1,\ldots,w_{\ell-1}\}=\emptyset.$$

现在定义

$$z_0=\begin{cases}u_1,\ \text{if } (u,w_1)\in E(\vec{H}),\\ u_2,\ \text{if } (w_1,u)\in E(\vec{H}),\end{cases}\qquad z_{\ell+1}=\begin{cases}v_1,\ \text{if } (v,w_\ell)\in E(\vec{H}),\\ v_2,\ \text{if } (w_\ell,v)\in E(\vec{H}),\end{cases}$$

那么在扩展路径 $\vec{P}'$ 中，其底层路径为

$$P':z_0-u-w_1-w_2-\ldots-w_\ell-v-z_{\ell+1},$$

顶点 u 和 v 在 $\vec{P}'$ 中是 NWO。这意味着 $\vec{P}'$ 本身是临界路径或包含临界子路径。◀

命题的证明39 通过从 INDEX 进行归约得出。同样，我们使用一个 WO $K_{n,\ n}$来编码 Alice 的输入，并仔细选择某个 $\vec{e}\in E(\vec{H})$，然后通过将 $\vec{e}$ 与索引弧重叠，附加一个 H 的副本。通过计算归约图中的最短NWO环、最长有向路径或最短临界路径的数量，可以证明该归约是有效的。

命题39。 设 H 是一个 *NWO* 二分图。如果 H 有一个连通分量包含至少 2 个 *NWO* 顶点，则 $R^1_{2/3}(Sub(\vec{H}))=\Omega(n^2)$。

证明。 根据引理 38， H 包含一个 NWO 圈、一条有向 P_4或一条临界路径。若 H 包含一个 NWO 圈，则令 ℓ 为 H中最短 NWO 圈的长度。取 C 为 C_ℓ 中的一个 NWO H，其底层环为 C： $v_1-v_2-\ldots-v_\ell-v_1$，并令 $v=v_1$，$u=v_2$。令 $\vec{e}\in E(\vec{H})$ 的端点为 u，v。

否则，如果 H 包含一条有向 P_4，则令 ℓ 为 H中最长有向路径的长度，因此 $\ell\geq 3$。选择 H 中的一条有向路径 $v_1\to v_2\to\ldots\to v_{\ell+1}$，并令 $v=v_3$、$u=v_2$。令 $\vec{e}\in E(\vec{H})$ 的端点为 u，v。

否则 H 包含一条临界路径，并令 ℓ 为 H中最短临界路径 的长度。选择一条最短临界路径 P，其底层路径为 P: $v_1-v_2-v_3-\ldots-v_\ell$，并令 $v=v_3$, $u=v_2$。令 $\vec{e}\in E(\vec{H})$ 的端点为 u,v。

给定一个 INDEX$_{n^2}$ 实例，其中 Alice 持有 $S\subseteq[n^2]$，而 Bob 持有 $t(n-1)+s\in[n^2]$，且 $s,\ t\in[n]$，我们将有向归约图 $\vec{G}_{\vec{H}}$ 定义为

$$V(\vec{G}_{\vec{H}})=\Big(V(\vec{H})\setminus\{v,u\}\Big)\sqcup\bigsqcup_{i=1}^{n}\{v^{(i)},u^{(i)}\},\quad E(\vec{G}_{\vec{H}})=E_1\cup E_2\cup E_3,$$

$$E_1=\begin{cases}\{(u^{(i)},v^{(j)}):i,j\in[n],(j-1)n+i\in S\},\ \text{if } (u,v)\in E(\vec{H}),\\ \{(v^{(i)},u^{(j)}):i,j\in[n],(j-1)n+i\in S\},\ \text{if } (v,u)\in E(\vec{H}),\end{cases}$$

$$E_2=\{(v^{(t)},w):(v,w)\in E(\vec{H})\}\cup\{(w,v^{(t)}):(w,v)\in E(\vec{H})\},$$

$$E_3=\{(u^{(s)},w):(u,w)\in E(\vec{H})\}\cup\{(w,u^{(s)}):(w,u)\in E(\vec{H})\}\cup E(\vec{H}-\{v\}-\{u\}).$$

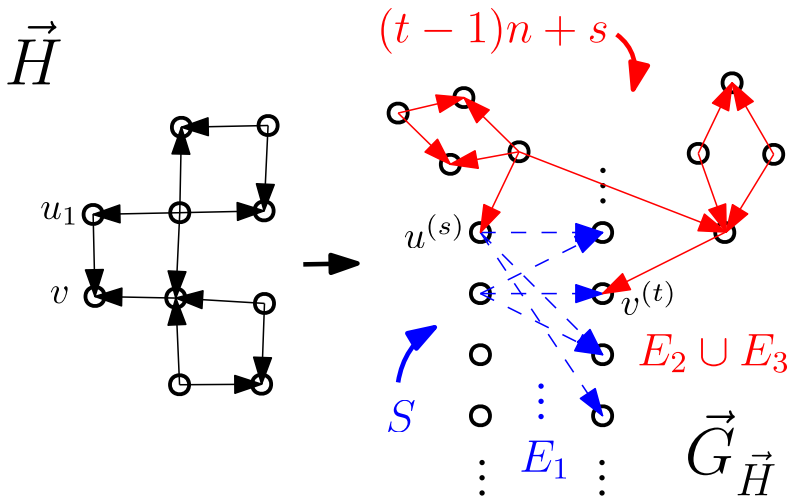

图5 命题 $\vec{G}_{\vec{H}}$ 的归约图，39。在此例中， $\vec{H}$ 包含一个 NWO 圈和临界路径。

下面的断言完成了证明，因为 $|V(\vec{G}_{\vec{H}})| = \Theta(n)$。

断言： 当且仅当 $(t-1)n+s \in S$ 时， $\vec{H} \subseteq \vec{G}_{\vec{H}}$。

证明。 若 $(t-1)n+s \in S$，则可通过以下方式形成 $\vec{H}$ 的一个副本：

$$v^{(t)}, u^{(s)}, V(\vec{H}) \setminus \{u, v\}.$$

Conversely, if $(t-1)n+s \notin S$, then $(u^{(s)}, v^{(t)}), (v^{(t)}, u^{(s)}) \notin E(\vec{G}_{\vec{H}})$.

情形 1： H 包含一个 NWO 圈。

假设 $\vec{G}$ 中有 $\vec{H}$ 的一个副本 $\vec{H}'$。注意， $\vec{G}_{\vec{H}} - E_1$ 只是 $\vec{H} - \vec{e}$ 的一个副本与若干孤立顶点的并，而在 $\vec{H}$ 中，弧 $\vec{e}$ 位于一个 NWO $\vec{C}_\ell$ 上。因此，若要使 $\vec{H}'$ 包含足够多的 NWO $\vec{C}_\ell$，它必须包含一个与 E_1 相交的 NWO C_ℓ 的副本 $\vec{C}'$。但是 $\vec{G}_{\vec{H}}[E_1]$ 是 WO，所以 $\vec{C}'$ 也与 $E_2 \cup E_3$ 相交。由于 $V(E_2 \cup E_3)$ 与 $V(E_1)$ 至多在$\{u^{(s)}, v^{(t)}\}$中相交，我们可以令

$$\vec{C}' = \vec{e}_1, \ldots, \vec{e}_{\ell'}, \vec{f}_{\ell'+1}, \ldots, \vec{f}_\ell,$$
$$e_1, \ldots, \vec{e}_{\ell'} \in E_1, \quad \vec{f}_{\ell'+1}, \ldots, \vec{f}_\ell \in E_2 \cup E_3,$$

其中 $u^{(s)}$ 是 $\vec{e}_1$ 的一个端点， $v^{(t)}$ 是 $\vec{e}_\ell$的一个端点。然而，这意味着，如果我们用 u 替换 $u^{(s)}$ ，并用 v 替换上述边中的 $v^{(t)}$ ，那么在 $\vec{H}$ 中， $\vec{}$ $\vec{}$就会出现一个由边 $\vec{e}$、$f_{\ell'+1}$，…， f_ℓ构成的 NWO 圈，其长度小于 ℓ，这就产生了矛盾。因此 $\vec{H} \not\subseteq \vec{G}_{\vec{H}}$。

情况2： H 不包含 NWO 圈，但包含一个有向 P_4。

假设 $\vec{H}'$ 中存在 $\vec{G}$的一个副本 $\vec{H}$ 。与情形 1 类似，由于 $\vec{e}$ 位于一条有向 $P_{\ell+1}$上，为了包含足够多的有向 $P_{\ell+1}$， $\vec{H}'$ 必须包含一个与E_1相交的有向 $P_{\ell+1}$ 的副本 $\vec{Q}$ 。然而， $\vec{G}_{\vec{H}}[E_1]$是 WO，而 $V(E_2 \cup E_3)$与 $V(E_1)$至多在$\{u^{(s)}, v^{(t)}\}$中相交，因此 $\vec{Q}$

与 E_1 相交于唯一的 $\vec{f}$，它在$\{u^{(s)},\ v^{(t)}\}$中恰有一个端点。若对某个 $s' \neq s$，$\vec{f} = (u^{(s')}v^{(t)})$，则我们可以令

$$\vec{Q} = u^{(s')} \to v^{(t)} \to w_1 \to w_2 \to \ldots \to w_{\ell-1},\quad w_1, \ldots, w_{\ell-1} \in V(\vec{H}).$$

根据假设，H 包含由 $v_1 \to u \to v \to v_4 \to \ldots \to v_{\ell+1}$ 给出的有向 $P_{\ell+1}$。此外，我们有 $v_1 \notin \{w_1, \ldots\ldots, w_{\ell-1}\}$；否则，$\vec{H}$ 中将存在一个由 $v_1 \to u \to v \to w_1 \to \ldots \to v_1$ 给出的 NWO 圈。然而，这会在 H 中形成一条有向 $P_{\ell+2}$，由以下内容给出：

$$v_1 \to u \to v \to w_1 \to \ldots \to w_{\ell-1},$$

这与 ℓ 是 H 中最长有向路径的长度矛盾。同理，若对某个 $t' \neq t$ 有 $\vec{f} = (u^{(s)},\ v^{(t')})$，也会得到矛盾。因此，结论是 $\vec{H} \nsubseteq \vec{G}_{\vec{H}}$。

情形 3： H 不含有向 P_4，但含有一条临界路径。

假设 $\vec{G}$ 中有一个 $\vec{H}$ 的副本 $\vec{H}'$。与前几种情形一样，由于 $\vec{e}$ 位于 $\vec{H}$ 中一条最短临界路径上，为了包含足够多的最短临界路径，$\vec{H}'$ 必须包含一个与 E_1 相交、长度为 ℓ 的临界路径副本 $\vec{Q}$。然而，在 $V(E_1) = \{u^{(1)},\ v^{(1)},\ \ldots\ldots,\ u^{(n)},\ v^{(n)}\}$ 中，只有 $u^{(s)}$ 和 $v^{(t)}$ 在 $\vec{G}_{\vec{H}}$ 中可能是 NWO，因此 $u^{(s)},\ v^{(t)} \in V(\vec{Q})$，并且我们可以令 $\vec{Q}$ 为

$$\vec{Q} : \vec{e}_1, \ldots, \vec{e}_a, \vec{f}_1, \ldots, \vec{f}_b, \vec{g}_1, \ldots, \vec{g}_c,$$
$$\vec{e}_1, \ldots, \vec{e}_a, \vec{g}_1, \ldots, \vec{g}_c \in E_2 \cup E_3,\quad \vec{f}_1, \ldots, \vec{f}_b \in E_1,\quad a, c \geq 1,\quad b \geq 1,\quad a + b + c = \ell,$$

其中 $u^{(s)}$ 是 $\vec{f}_1$ 的一个端点，$v^{(t)}$ 是 $\vec{f}_b$的一个端点。然而，这意味着，如果我们将上述边中的 $u^{(s)}$ 替换为 u，并将 $v^{(t)}$ 替换为 v，那么在 $\vec{H}$ 中打包后，我们会得到由以下弧构成的定向路径

$$\vec{e}_1, \ldots, \vec{e}_a, \vec{e}, \vec{g}_1, \ldots, \vec{g}_c,$$

若 $a = c = 1$，则它将是一个有向 P_4；否则，它将是长度小于 ℓ的临界路径。两者都构成矛盾，因此我们得出结论：$\vec{H} \nsubseteq \vec{G}_{\vec{H}}$。◁

◀

现在我们证明，对于一个 NWO 二分图 H，如果 H 有一个包含环且恰好含有一个 NWO 顶点的连通分量，那么 H 也是一个单遍困难实例，由此即可完成定理7中下界的证明。困难的情形是 H 中的所有环都是 WO；否则，命题39 的证明 H 已经表明是一个困难实例。此外，H 由于我们要证明下界，将其替换为目标连通分量后，不失一般性地，我们可以假设 H 是连通的。

在命题39中，构造归约图时，我们考虑的三种特殊结构不可能包含在 E_1（Alice 的弧集）中，因此通过计算这些特殊结构的数量，可以相当容易地证明该归约。但现在没有这些特殊结构，人们可能会担心，来自 E_1 的某些环或结构与Bob 添加的弧结合后，即使 INDEX 的答案为假，是否也会意外地产生 H 的副本。

为解决这个问题，我们将记录 NWO 顶点到每个环的距离。在所有使该距离最小的环中，我们仔细选取一个与索引边重叠的环，随后通过归约图中的距离论证完成证明。

命题40。 设 H 是一个 *NWO* 二分图。若 H 有一个包含环且恰好有一个 *NWO* 顶点的连通分量，则 $R^1_{2/3}(\mathit{Sub}(\vec{H})) = \Omega(n^2)$。

证明。假设 $\vec{H}$ 是连通的，不失一般性。令 $w \in V(\vec{H})$ 为 H中唯一的 NWO 顶点，并令 ℓ 为 H中最短环的长度。对于 $\vec{H}$中的每个最短环$\vec{C}$，我们用 $d_{\vec{H}}(\vec{C}) = \mathrm{dist}_H(w, C)$ 表示在 H中从 w到 C的距离。如果 $w \in V(\vec{C})$，则 $d_{\vec{H}}(\vec{C}) = 0$是可能的。现在定义

$$m = \min\{d_{\vec{H}}(\vec{C}) : \vec{C} \text{ is a } \vec{C}_\ell \text{ in } \vec{H}\}, \quad \mathcal{C}_{\vec{H}} = \{\vec{C} : \vec{C} \text{ is a } \vec{C}_\ell \text{ in } \vec{H} \text{ with } d_{\vec{H}}(\vec{C}) = m\}.$$

此外，对于所有 $C \in \mathcal{C}_{\vec{H}}$，我们定义

$$A_{\vec{H}}(\vec{C}) = \{z \in V(\vec{C}) : \exists \text{ a path } w \rightsquigarrow z \text{ of length } \mathrm{dist}_H(w, C) \text{ in } H\},$$

V在 H中，从 w到 C的最短路径的端点所构成的$\vec{C}$（V）顶点集合。选取$\vec{C}^*$，使其从$\mathcal{C}_{\vec{H}}$中令 $|A_{\vec{H}}(\vec{C}^*)|$最大，并令 $M = |A_{\vec{H}}(\vec{C}^*)|$，再从 E中选取 $\vec{e} = (u, v)$，使得 $A(\vec{C}^*) \not\subseteq \{u, v\}$。

给定一个 INDEX$_{n^2}$ 实例，其中 Alice 持有 $S \subseteq [n^2]$，Bob 持有 $t(n-1)+s \in [n^2]$，且 $s, t \in [n]$，我们将有向归约图 $\vec{G}_{\vec{H}}$ 定义为

$$V(\vec{G}_{\vec{H}}) = \Big(V(\vec{H}) \setminus \{u, v\}\Big) \sqcup \bigsqcup_{i=1}^{n} \{u^{(i)}, v^{(i)}\}, \quad E(\vec{G}_{\vec{H}}) = E_1 \cup E_2 \cup E_3,$$

$$E_1 = \{(u^{(i)}, v^{(j)}) : i, j \in [n], (j-1)n + i \in S\}$$

$$E_2 = \{(w, v^{(t)}) : (w, v) \in E(\vec{H})\} \cup \{(v^{(t)}, w) : (v, w) \in E(\vec{H})\}$$

$$E_3 = \{(w, u^{(s)}) : (w, u) \in E(\vec{H})\} \cup \{(u^{(s)}, w) : (u, w) \in E(\vec{H})\} \cup E(\vec{H} - \{u\} - \{v\}).$$

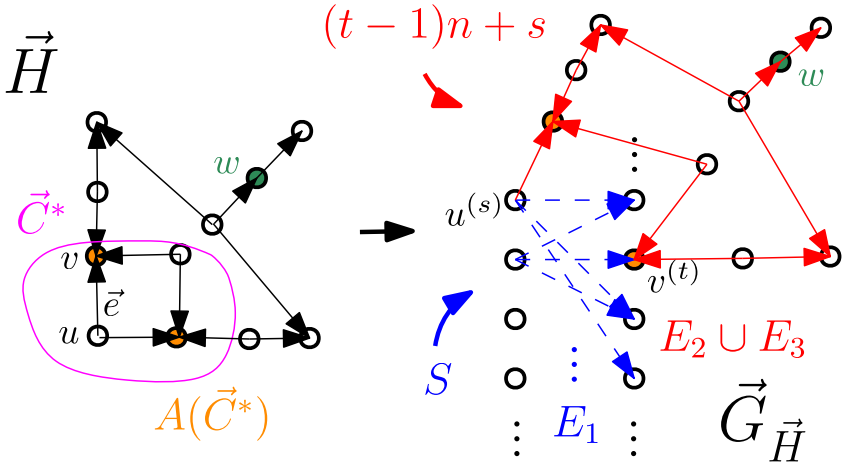

图 6 命题 $\vec{G}_{\vec{H}}$ 中的归约图 40。

若 $(t-1)n + s \in S$，则通过以下方式构造 $\vec{H}$ 的一个副本

$$u^{(s)}, v^{(t)}, V(\vec{H}) \setminus \{u, v\}.$$

反之，如果 $(t-1)n + s \in S$，那么 $(u^{(s)}, v^{(t)}) \notin E(\vec{G}_{\vec{H}})$。假设 $\vec{G}_{\vec{H}}$ 中有一个 $\vec{H}$ 的副本 $\vec{H}'$。注意，$\vec{G}_{\vec{H}}$ 中唯一的 NWO 顶点是 w，因此 $\vec{H}'$ 中唯一的 NWO 顶点

也是 w。因此，如果 $\vec{C}'$ 是 $\vec{H}'$ 中包含于 $\vec{G}_{\vec{H}}[E_2 \cup E_3]$ 的一个 $\vec{C}_\ell$，并且 $\vec{C}$ 是由 $\vec{C}'$ 将顶点 $u^{(s)}$，$v^{(t)}$ 分别替换为 u，v 后得到的 $\vec{H}$ 中的 $\vec{C}_\ell$，那么

$$d_{\vec{H}}(\vec{C}) \le d_{\vec{H}'}(\vec{C}'),$$

并且等式 $d_{\vec{H}}(\vec{C}) = d_{\vec{H}'}(\vec{C}')$ 将意味着

$$|A_{\vec{H}}(\vec{C})| \ge |A_{\vec{H}'}(\vec{C}')|,$$

因为在 H' 中从 w 到 C' 的最短路径必然包含于 $E_2 \cup E_3$ 中，而后者又包含于 $E(\vec{H})$中，所以这条路径也将是在 H中从 w 到 C 的最短路径。因此，由于$(u^{(s)}, v^{(t)})$不属于 $E(\vec{G}_{\vec{H}})$，$\vec{H}'$中包含于 $\vec{G}_{\vec{H}}[E_2 \cup E_3]$且满足$d_{\vec{H}'}(\vec{C}') = m$和$|A_{\vec{H}'}(\vec{C}')| = M$的 $\vec{C}_\ell$数量少于 $\vec{H}$中的相应数量。

另一方面，如果 $\vec{H}'$中包含于 $\vec{G}_{\vec{H}}[E_1]$且满足 $d_{\vec{G}_{\vec{H}}}(\vec{C}') = m$和$|A_{\vec{G}_{\vec{H}}}(\vec{C}')| = M$的 $\vec{C}_\ell$ 是 $\vec{C}'$，那么由于在 H'中从 w到 C'的最短路径必然包含于 $E_2 \cup E_3$中，

$$A_{\vec{H}'}(\vec{C}') \subseteq \{u^{(s)}, v^{(t)}\} \subseteq V(\vec{C}^*),$$

因此 $d_{\vec{H}}(\vec{C}^*) \le d_{\vec{H}'}(\vec{C}')$，这就迫使 $d_{\vec{H}}(\vec{C}^*) = d_{\vec{H}'}(\vec{C}') = m$，因为我们还有 $d_{\vec{H}}(\vec{C}^*) = m$。然而，这意味着

$$u^{(s)} \in A_{\vec{H}'}(\vec{C}') \Rightarrow \exists \text{ path } P : w \rightsquigarrow u^{(s)} \text{ of length } m \text{ in } H' \text{ and } P \subseteq G_{\vec{H}}[E_2 \cup E_3]$$
$$\Rightarrow \exists \text{ path } w \rightsquigarrow u \text{ of length } m \text{ in } H \Rightarrow u \in A_{\vec{H}}(\vec{C}^*)$$

同理

$$v^{(s)} \in A_{\vec{H}'}(\vec{C}') \Rightarrow v \in A_{\vec{H}}(\vec{C}^*).$$

但是 $A(\vec{C}^*) \not\subseteq \{u, v\}$，而 $|A_{\vec{H}}(\vec{C}')| = M$，并且 $A_{\vec{H}'}(\vec{C}') \subseteq \{u^{(s)}, v^{(t)}\}$，所以 $|A_{\vec{H}}(\vec{C}^*)| > M$，这与假设矛盾。

最后，与命题39的证明类似，在 $\vec{H}'$ 中不可能存在同时使用来自 E_1 的两条弧和来自 $E_2 \cup E_3$ 的弧的 $\vec{C}_\ell$；否则，将这个 $\vec{C}_\ell$ 的 E_1 部替换为弧 (u, v)，我们便会在 H 中再次得到一个长度短于 ℓ 的环，这与假设矛盾。因此，总而言之，$\vec{H}'$ 中满足 $d_{\vec{H}'}(\vec{C}') = m$ 且 $|A_{\vec{H}'}(\vec{C}')| = M$ 的 $\vec{C}_\ell$ 数量少于 $\vec{H}$ 中的数量，所以 $\vec{H} \not\subseteq \vec{G}_{\vec{H}}$。◀

接下来我们证明定理 7 的第二部分，该部分由命题 41 推出，也是我们单遍二分定理中关于上界的主要结果。

▶**命题 41**。设 H 是一个至多包含一个 *NWO* 顶点的定向树。则

$$R^1_{2/3}\big(\textsc{Sub}(\vec{H})\big) = \widetilde{O}(n).$$

在证明命题 41 之前，我们先介绍一个单遍流式例程，该例程为输入图 G 构造一个~稀疏证书。该证书是一个子图 $G' \subseteq G$，其大小为 $O(n)$，满足是否存在 $\vec{H}$ 于 G 中可以仅通过检查 G' 来判定。我们的构造依赖于标准采样原语；我们在此陈述这些原语，并将证明留待文献。

▶ **定义 42** (ℓ_0-采样器)。设 $\mathbf{x} \in \mathbb{Z}^n$ 是在（一般）转盘模型中，在执行以下更新时维护的频率向量 (i, Δ) 执行 $x_i \leftarrow x_i + \Delta$ 对于 $\Delta \in \mathbb{Z}$。固定参数$\varepsilon, \delta \in (0,1)$。一个 (ε, δ)-ℓ_0-采样器 是一种随机化流式算法，在处理完数据流后，输出一个特殊符号 $\perp$ 或一个索引 $i \in [n]$，满足：

1. (无假阳性) 如果算法输出 $i \in [n]$，那么 $x_i \neq 0$。**2.** (成功概率) 如果 $\|\mathbf{x}\|_0 > 0$，那么 $\mathbb{P}[\text{输出 } \perp] \leq \delta$。**3.** (近似均匀性) 对于所有 $i \in \mathrm{supp}(\mathbf{x}) = \{j \in [n] : x_j \neq 0\}$，

$$\mathbb{P}[\text{output } i] = (1 \pm \varepsilon) \cdot \frac{1}{|\mathrm{supp}(\mathbf{x})|} \pm O(\delta).$$

▶ **引理 43** ([16], 定理 2)。 存在一个 $(0,\delta)$-ℓ_0-采样器，使用 $O\big(\log^2 n \cdot \log(1/\delta)\big)$ 空间。

我们还使用 k-核的概念。对于一个无向图 G，k-核是 C 的唯一极大诱导子图，G 其最小度至少为 k。通过反复删除度为 $< k$ 的任意顶点，直到不再有此类顶点，即可得到该子图。对于适当选取的常数 k，我们的证书构造使用 ℓ_0-采样器对 k-核内部的边进行稀疏化，同时保留其外部的所有内容。

▶ **引理 44**。 令 G 为一个 n-顶点无向图，令 $k = O(1)$，并令 K 为 G的 k-核。存在一个单遍随机流算法，使用 $O(kn)$空间；该算法以高概率输出并存储一个子图 $G' \subseteq G$，满足：

1. 对于每个顶点 $v \notin V(K)$，所有与 v 相 *incident* 的 G 中的边都包含在 G' 中。*2.* 对于每个顶点 $v \in V(K)$，有 $\deg_{G'[V(K)]}(v) \geq k$。*3.* $|E(G')| = O(kn)$。

我们将其称为 G' 小森林保持证书。

证明。 我们为每个顶点 v 维护 k 个相互独立的 $(0,\delta)$-ℓ_0-采样器 $S_{v,1}$、……、$S_{v,k}$，并维护度计数器 $d[v]$。遇到一条边 $\{u, v\}$ 时，我们将 $d[u]$，$d[v]$ 加一，并将 u 插入每个$S_{v,i}$，再将 v 插入每个$S_{u,i}$。

数据流结束后，我们离线运行标准的 k-核剥离过程，并使用采样器恢复邻居。每当一个顶点 v 的当前度为 $d[v] < k$ 时，我们查询 $S_{v,1}$、……、$S_{v,d[v]}$（从后续采样器中删除已恢复的邻居，以确保邻居互不相同），将与 v 相 incident 的相应边加入 G'，并移除 v，同时将其已恢复邻居的度递减。这恰好模拟了核剥离，因此剩余的活跃集合等于 k-核 K。

因此，每个顶点 $v \notin K$ 都将其在剩余图中的所有关联边加入 G'，从而证明了 (1)。过程终止后，对于每个 $v \in C$，我们使用 $S_{v,1}$，……，$S_{v,k}$在 K 中恢复 k 个（互不相同的）邻居，并将这些边加入 G'，由此得到 $\deg_{G'[V(K)]}(v) \geq k$，并证明了 (2)。由于每个剥离顶点添加的边少于 k 条，而每个核心顶点至多添加 k 条边，因此 $|E(G')| = O(kn)$，证明了 (3)。

空间开销利用 $kn\delta$ 个采样器和上述计数器，即 $O(kn)$。对 n 个采样器的 $O(kn)$ 次采样器查询，因此根据并集界，失败概率至多为 $O(kn)\delta = o(1)$。◀

命题 41 的证明。 若 H 不含 NWO 顶点，则每个 H-自由定向图都有 $O(n)$ 条弧，即 $\mathrm{ex}(n, \vec{H}) = O(n)$。因此，我们可以通过存储所有弧$(\vec{H})$，在单遍扫描中求解，空间开销为 $O(n)$。

现在假设 H 恰有一个 NWO 顶点。我们使用一个常数规模的颜色编码步骤。对输入定向图 G 中的每个顶点，独立地从 $\{c_1, c_2, c_3\}$中均匀随机选取一种颜色。令 $\vec{G}_1$ 为由颜色 c_1 指向颜色 c_2的弧所构成的子图，令 G_2 为由颜色 c_2 指向颜色 c_3的弧所构成的子图。忽略定向后，我们将引理44应用于 G_1 和 G_2，并令

$$k := 2|V(\vec{H})|$$

并得到小森林保持证书 $\vec{G}'_1 \subseteq \vec{G}_1$ 和 $\vec{G}'_2 \subseteq \vec{G}_2$。当且仅当$\vec{G}' := \vec{G}'_1 \cup \vec{G}'_2$ 包含 $\vec{H}$ 的副本时，我们输出真。由于 $\vec{G}' \subseteq \vec{G}$，该过程不会产生假阳性。独立地重复整个实验 $O(1)$ 次（在单遍扫描中并行执行），可将成功概率提升至至少 $2/3$。

接下来需要证明完备性。令 x 为 H的唯一 NWO 顶点。根据此类的定义，$V(\vec{H})$ 可划分为三个集合 V_1、V_2、V_3，其中 $x \in V_2$，且 H 的每条弧都从 V_i 定向到 V_j，并满足 $i < j$。等价地，H 可分解为 $H = T_1 \cup T_2$、，其中 T_1 由从 V_1 到 V_2 的弧组成，而 T_2 由从 V_2 到 V_3的弧组成。此外，T_1和 T_2 都是 WO 树，且仅在顶点 x处相交。

固定任意一个与 $\vec{H}$同构的副本 $\vec{H}^* \subseteq \vec{G}$。随机着色将该副本中 V_i 的每个顶点都赋予颜色 c_i（对于 $i = 1$、2、3）的概率为 $3^{-|V(\vec{H})|}$。以此事件为条件。则 $\vec{T}_1^*:= \vec{H}^*[V_1 \cup V_2]$ 是包含在 $\vec{G}_1$中的 $\vec{T}_1$ 的一个副本，而 $\vec{T}_2^*:= \vec{H}^*[V_2 \cup V_3]$ 是包含在 G_2 中的 T_2 的一个副本；两者共享 x 的同一像 x^*。

我们断言，$\vec{G}'_1$ 包含一个以 x^* 为根的 $\vec{T}_1$ 副本，且 $\vec{G}'_2$ 包含一个以 x^*为根的 $\vec{T}_2$ 副本，并且这两个副本可以选取为在 x^*之外顶点不相交。首先注意到，$\vec{T}_i$ 是一棵 WO 树，而 $\vec{G}_i$ 是 $\vec{G}$ 的一个 WO 子图，二者具有相同的定向模式（所有弧都从颜色 c_i 指向颜色 c_{i+1}）。因此，一旦我们在 $\vec{G}'_i$中找到 $\vec{T}_i$ 的底层无向嵌入，定向就会自动匹配，所以在嵌入过程中可以忽略方向。

固定 $i \in \{1, 2\}$，并考虑 $\vec{G}_i$ 及其证书 $\vec{G}'_i$。引理 44 保留了所有与 k-核之外的顶点相关联的边，并保证 k-核内的每个顶点在核内的度至少为 k。由于 $k = 2|V(\vec{H})|$ 且 $|\vec{T}_i| \le |V(\vec{H})|$，我们可以从 x^*开始在 $\vec{G}'_i$ 中贪心地嵌入 $\vec{T}_i$：在核之外，所有所需的边都由构造保证存在；一旦嵌入进入核内，最小度条件确保在每一步中，即使已有顶点被占用，仍有足够的未占用邻居来扩展嵌入，因此这两个嵌入除了共享根 x^*之外可以选取为不相交。

因此，在 $\vec{H}^*$ 着色正确的条件下，证书图 $\vec{G}' = \vec{G}'_1 \cup \vec{G}'_2$ 包含 $\vec{H}$ 的一个副本。因此，单次试验的成功概率至少为 $3^{-|V(\vec{H})|}$；重复进行 $O(3^{|V(\vec{H})|})$ 次独立试验，可达到常数成功概率；再通过 $O(1)$ 次额外的并行重复，成功概率至少可达到 $2/3$。总空间仍为 $\widetilde{O}(n)$，因为 $|V(\vec{H})|$ 是常数。◀

命题 41 也适用于每个连通分量至多包含个 NWO 顶点的有向森林。将 G 中的顶点独立且均匀地着以 r 种颜色，其中 r是 H 的连通分量数。以常数概率，H 的每个连通分量都会成为单色，且不同连通分量获得不同颜色。在此事件发生的条件下，只需在该次试验中针对每个 $i \in [r]$，检查由颜色 c_i诱导的子图是否包含 i的第 H个连通分量。并行重复进行常数次独立测试，可使成功概率至少达到 $2/3$，同时保持单遍扫描和 $O(n)$ 空间。

最后，唯一尚未涵盖的 NWO 二分图是这样的图：其中一些连通分量是 WO 且含有一个环，而另一些是恰好有一个 NWO 顶点的树，正如定理7。

设 H_1 为一个 NWO 森林，其中每个连通分量恰好包含一个 NWO 顶点，并设 H_2 为一个 WO 二分图。若 $H = H_1 \sqcup H_2$，则我们可以执行命题 41 中的算法，同时再存储 $\mathsf{ex}(n, \vec{H}_2) = \tilde{O}(n^{2-1/\Delta'(H_2)})$ 条额外边。若所有输入边都被存储，则我们直接搜索 H 的一个副本。否则，边数多于$n + \mathsf{ex}(n, \vec{H}_2)$，因此 $\vec{H}_1$ 的一个副本的存在等价于

$\vec{H}$ 的一个副本。这证明了下述结论，从而得到完整的单遍二分定理。

推论 45。 设 H 是一个 *NWO* 二分图。若 H 的每个连通分量要么是 *WO*，要么是恰有一个 *NWO* 顶点的树，则 $R^1_{2/3}(\textit{Sub}(\vec{H})) = \tilde{O}(n^{2-\Delta'(H)})$。

6 诱导定向图

我们的论证也可推广到诱导有向子图问题，并且我们还得到了完整的二分结果。

显然，诱导有向 co-P_3 是一个可处理情形：一个定向图 G 包含诱导有向 co-P_3 当且仅当 G 包含一个诱导 co-P_3，而后者可由推论26高效检测。事实上，在具有 H 且 $|V(H)| \geq 3$ 的图中，这是唯一可处理的实例。

结合命题32 与第 4 区段中的分类，我们首先得出结论：大多数有向诱导模式都是困难实例。

推论46. 如果 H 是一个至少有3 个顶点的定向图，且 H 不是 P_3、co-P_3或 P_4，那么 $R^p_{2/3}(\text{诱导子图}(\vec{H})) = \Omega(n^2/p)$。

进一步证明，即使底层无向图H 是 P_3 或 P_4， H 仍是困难实例，因此 co-P_3 是唯一的简单实例。

命题 47. 设 H 是一个定向图，且 $H = P_3$，则

$$R^p_{2/3}(\textsc{IndSub}(\vec{H})) = \Omega(n^2/p).$$

证明. 在同构和反向意义下， P_3有两种定向：

the directed path $a \to b \to c$ the "in-star" $a \to b \leftarrow c$

对于两种情形，我们都给出一个从不相交性问题$_{n^2}$的归约。Alice 和 Bob 分别持有集合 S_1、$S_2 \subseteq [n^2]$。我们构造归约图 $G_{\vec{H}}$，其顶点集为

$$V(\vec{G}) = X_1 \sqcup X_2 \sqcup X_3, \qquad X_i = \{x_1^{(i)}, \ldots, x_n^{(i)}\}.$$

情形 1： $\vec{H} \cong (a \to b \to c)$

将弧集定义为 $E(\vec{G}) = E_1 \sqcup E_2 \sqcup E_3$，其中

$$E_1 = \{(x_i^{(1)}, x_j^{(2)}) : (i-1)n + j \in S_1\}, \qquad E_2 = \{(x_j^{(2)}, x_i^{(3)}) : (i-1)n + j \in S_2\},$$

and

$$E_3 = \{(x_i^{(1)}, x_j^{(3)}) : i \neq j\}.$$

断言： $S_1 \cap S_2 \neq \emptyset$ 当且仅当 $\vec{G}$ 包含 $a \to b \to c$ 的诱导副本。

证明。 若 $n\,(i-1)\, + j \in S_1 \cap S_2$，则 $\{x_i^{(1)}, x_j^{(2)}, x_i^{(3)}\}$ 诱导出 $a \to b \to c$。反之，任何诱导的 $a \to b \to c$ 都必须从每一层中选取一个顶点 X_1， X_2， X_3 （因为所有弧都从较低索引指向较高索引），而端点之间不存在弧，这迫使两个端点在 X_1 和 X_3 中具有相同的索引（根据 E_3的定义），这意味着相应元素属于 $S_1 \cap S_2$。 $\triangleleft$

情况2： $\vec{H} \cong (a \to b \leftarrow c)$

我们保留相同的顶点集 $X_1 \sqcup X_2 \sqcup X_3$，并将指向 X_3 的弧定义为

$$E_1 = \{(x_i^{(1)}, x_j^{(3)}) : (i-1)n + j \in S_1\}, \qquad E_2 = \{(x_i^{(2)}, x_j^{(3)}) : (i-1)n + j \in S_2\}.$$

为确保在任何诱导副本中，两个源点 a、c 都不相邻，我们将 X_1 内部和 X_2 内部的所有弧都定向为传递竞赛图，并将 X_1 与 X_2 之间的二分图定向，使得除“匹配”对 $(x_i^{(1)}$、$x_i^{(2)})$ 以外的每一对顶点都相邻。具体来说，添加

$$E_3 = \{(x_i^{(1)}, x_j^{(2)}) : i > j\}, \qquad E_4 = \{(x_i^{(2)}, x_j^{(1)}) : i > j\},$$

$$E_5 = \{(x_i^{(1)}, x_j^{(1)}) : i < j\}, \qquad E_6 = \{(x_i^{(2)}, x_j^{(2)}) : i < j\}.$$

断言： $S_1 \cap S_2 \neq \emptyset$ 当且仅当 $\vec{G}$ 包含 $a \to b \leftarrow c$ 的一个诱导副本。

证明。若 $n(i-1) + j \in S_1 \cap S_2$，则 $\{x_i^{(1)}, x_i^{(2)}, x_j^{(3)}\}$ 诱导出 $a \to b \leftarrow c$。反过来，在任意诱导 $a \to b \leftarrow c$ 中，这两个源点必须是 $X_1 \cup X_2$ 中一对不相邻的顶点，因为 X_3 中所有节点都没有出邻点；根据构造，这迫使它们为某个 i 对应的 $(x_i^{(1)}, x_i^{(2)})$。根据我们的构造，它们的公共出邻点只能位于 X_3 中，由此确定一个指标 j，使得 $(i-1)n + j \in S_1 \cap S_2$。◁

无论哪种情形，我们都得到了从不相交性问题$_{n^2}$ 到诱导子图查找$(\vec{H})$的归约。这意味着，任何 p遍次流式算法都需要 $\Omega(n^2/p)$ 的空间下界。◀

命题 48. 设 H 是一个定向图，且 $H = P_4$，那么

$$R^p_{2/3}(\textsc{IndSub}(\vec{H})) = \Omega(n^2/p).$$

证明. 在同构和反向意义下，P_4 有三种定向：

$$\vec{H}_1 : a \to b \to c \to d, \qquad \vec{H}_2 : a \to b \to c \leftarrow d, \qquad \vec{H}_3 : a \to b \leftarrow c \to d.$$

我们将 不相交性问题$_{n^2}$ 归约到 诱导子图($\vec{H}_i$)，对每个 $i \in \{1,2,3\}$ 直接扩展命题 47中的构造即可。

对于 $\vec{H}_1$和 $\vec{H}_2$，从命题 47 情形 1 中的归约图开始（模式为 $a \to b \to c$）。添加第四层 $X_4 = \{x_1^{(4)}, \ldots, x_n^{(4)}\}$ 以及匹配弧

$$\left\{(x_i^{(3)}, x_i^{(4)}) : i \in [n]\right\} \quad \text{or} \quad \left\{(x_i^{(4)}, x_i^{(3)}) : i \in [n]\right\},$$

→　　→选择方向，使得导出模式分别变为 H_1 或 H_2。与命题 47 中相同的论证表明，$S_1 \cap S_2 \neq \emptyset$ 当且仅当所得图包含 $\vec{H}_1$（分别为 $\vec{H}_2$）的诱导副本。

对于 $\vec{H}_3$，从命题47（针对模式$a \to b \leftarrow c$）中的情况2的归约图开始。添加 X_4 以及匹配弧$\{(x_i^{(2)}, x_i^{(4)}) : i \in [n]\}$。同样，利用相同的结构性论证可得，当且仅当所得图包含 $S_1 \cap S_2 \neq \emptyset$ 一个诱导副本→，且该副本是 H_3。

在这三种情况下，Indsub(H_i) 的难度至少与 DISJ$_{n^2}$ 相当，因此可推出所述的 $\Omega(n^2/p)$ 空间下界。◀

7 备注与开放问题

我们列出并讨论前面各区段中提到的一些开放问题。

7.1 二分图的下界

▶ **问题 1.** 对于各连通分量的直径均为 ≥ 4 的一般二分图 H， $R^p_{2/3}(\mathit{Sub}(H))$ 是多少？我们是否有 $\Omega(\mathrm{ex}(n,H)/p)$ 下界？

尽管我们在命题 18 中证明了，当有向图 H 的某个连通分量直径为 < 4 时，$R^p_{2/3}(\mathrm{Sub}(H)) = \Omega(\mathrm{ex}(n,H)/p)$，但该证明不适用于问题 1。对于问题 1 中的 H，应用命题 18 的证明时，尽管 G' 是 H-自由的，但不再有这样的限制：来自有向图 H 同一部分的两个顶点必须有一个公共邻点。因此，在证明中，当 $\bigcap_{q=1}^{|E(H)|} S_q = \emptyset$ 时，假设有一个 H 的副本 H'，我们不再能推出

$$|V(H') \cap \{v_1^{(j)}, \ldots, v_a^{(j)}\}|, |V(H') \cap \{u_1^{(j)}, \ldots, u_a^{(j)}\}| \leq 1,$$

因此可能并不存在矛盾。例如，图7展示了在 $H = C_8$ 时证明可能出错的情况。

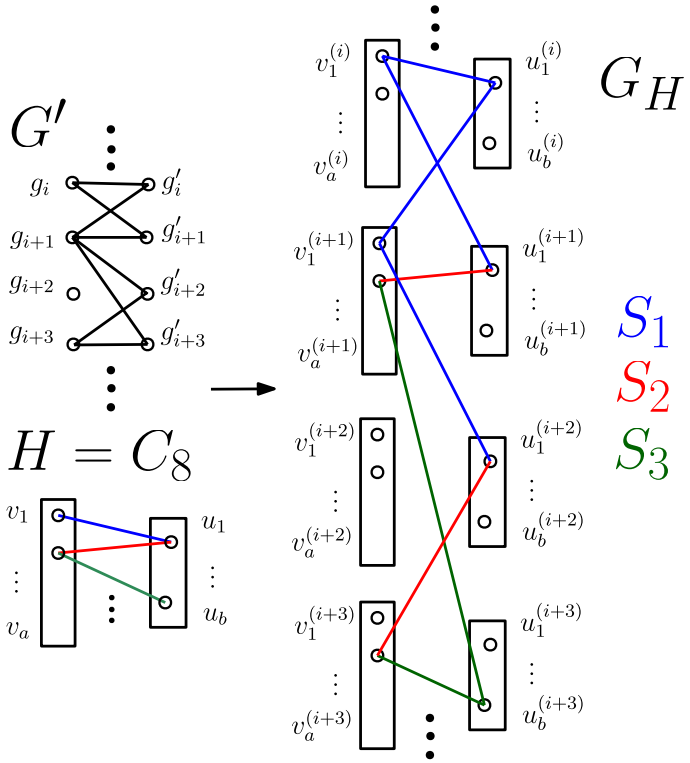

图 7 一个 $H \subseteq G_H$，但双方没有共同元素的例子。

▶ **问题 2。** 对于一般的二分图 H，若其所有连通分量的连通度均为 ≤ 2，那么 $R^1_{2/3}(\mathit{Sub}(H))$ 是多少？我们是否有 $\Omega(\mathrm{ex}(n,H))$ 下界？

我们在命题 19 中证明了，对于某个连通分量的连通度为 > 2 的二分图 H，有 $= \Omega(\mathrm{ex}(n,H))$ $R^1_{2/3}(\mathrm{Sub}(H))$。在证明中，我们通过将一个 H_0 的副本与目标索引边 e_k 重叠，附加了一个 H 的副本 H_0，其中重叠部分为一个 $\{v_i, v_j\} \in E(H_0)$。接着，在 G_H 中称一个 H 的副本 H'

如果它同时使用来自 $V(G')$ 的两个顶点以及来自 $V(H_0) \setminus \{v_i, v_j\}$ 的顶点，我们就称其为“交叉”的。通过 $\kappa(H) > 2$，我们证明了若 $e_k \notin E(G_H)$，则不存在交叉的 H 副本，因此 $H \not\subseteq G_H$。然而，当转向 $\kappa(H) = 2$ 的情形时，我们遇到了障碍。为描述这些障碍，我们首先定义 H 的二顶点割分量。

▶ 定义 *49*（*TVCC*）。设 H 是具有 $\kappa(H) = 2$ 的二分图，且 $\{u, v\}$ 是 H 的一个二顶点割，其中 u, v 属于不同的部，并且 $\{u, v\} \notin E(H)$。设 A' 是 $H - u - v$ 中的一个连通分量，其中 $H - u - v$ 是移除 u, v 后剩余的部分，且 $A(u, v) = H[\{u, v\} \cup V(A')]$。我们称 $A(u, v)$ 是 H 的一个二顶点割分量（*TVCC*）。

如果 $\kappa(H) = 2$，那么在命题 19 的证明中，当索引边 $e_k \notin E(G_H)$ 时，尽管 $G_H[V(G')]$ 和 $G_H[V(H_0)]$ 都是 H-free，仍可能存在交叉的 H 副本，因为 $G_H[V(G')]$ 和 $G_H[V(H_0)]$ 中的一个可能包含 H 的 TVCC，而另一个可能包含剩余的部分。

有人可能会问，能否通过精心选择要与索引边 e_k 重叠的$\{v_i, v_j\} \in E(H)$来避免这种情况。事实上，如果存在$\{v_i, v_j\} \in E(H)$，使得有向图 H 中的任何 TVCC$A(u, v)$都无法以顶点对$\{u, v\} \mapsto \{v_i, v_j\}$嵌入有向图 H，那么在证明中就可以排除任何交叉副本。然而，以下结果构成了阻碍。

▶ **命题 50.** 存在一个二分图 H，其具有 $\kappa(H) = 2$，使得对于所有 $e \in E(H)$，都有一个有向图 H 的 *TVCC* $A(u, v)$ 以及一个嵌入 $\varphi\colon A(u, v) \to H$，满足 $\varphi : \{u, v\} \mapsto e$。

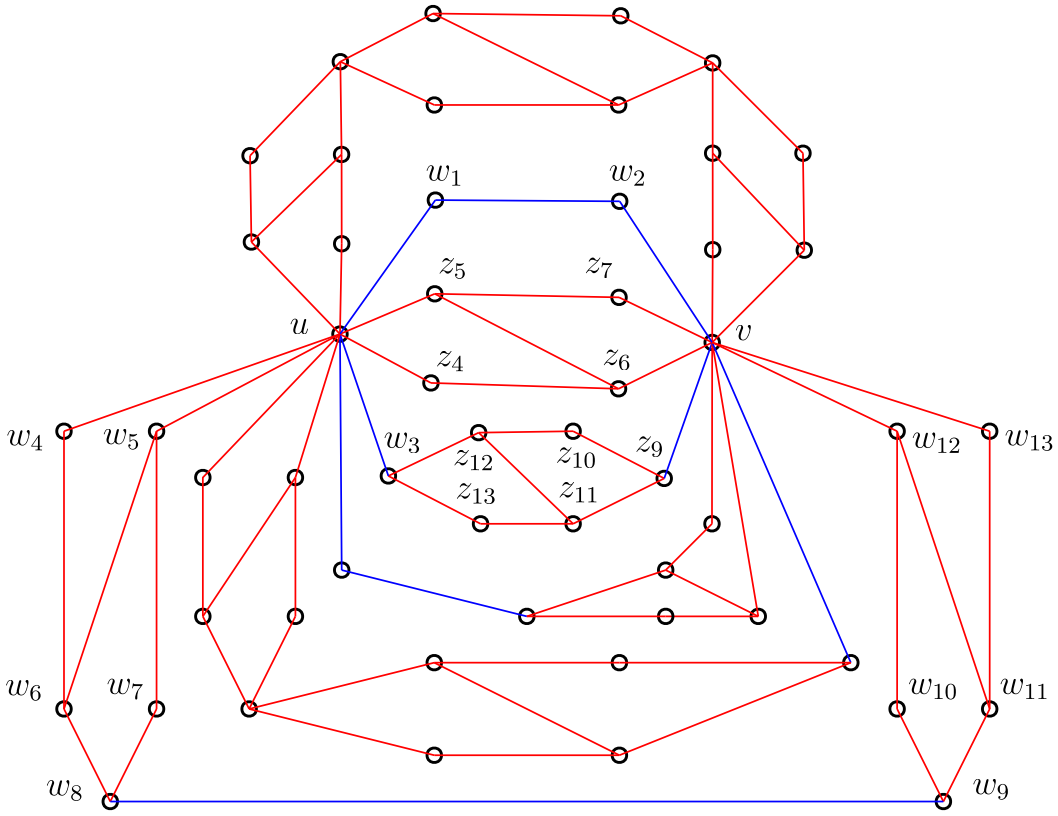

图 8 命题 50 的一个示例。

证明。 图 8给出了一个例子。在命题 50 的表述中，如果 e 选取任意一条红色边，那么 u、w_1、w_2、v 就构成所需的 TVCC。另一方面，无论

我们选择哪条蓝色边作为 e，总能嵌入某个 TVCC $A(u、v)$，使得 $\{u、v\} \mapsto e$。例如，若取 $e = \{u、w_3\}$，那么由$u、w_4、w_5、\ldots\ldots、w_{12}、w_{13}、v$构成的 TVCC 可以通过

$$u \mapsto u, \quad v \mapsto w_3, \quad w_8 \mapsto v, \quad w_i \mapsto z_i \quad \forall 4 \le i \le 13,\ i \neq 8.$$ ◀

7.2 无向图的上界

在命题 17 中，我们证明了 $R^p_{2/3}(\textsc{Sub}(H)) = \Omega(\mathrm{ex}(n,H)/p)$ 当 H 是一个非匹配森林时成立。不过，如果 H 是一个匹配，则存在一个非常简单的算法。

▶ **命题51.** 设 H 是一个含有 ℓ 条边的匹配，则 $D^1(\textit{Sub}(H)) = \tilde{O}(1)$。

证明。 如果某个 $v \in V(G)$ 的度至少为 2ℓ，那么移除它的一条关联边不会影响 H 在 G中存在的可能性。实际上，如果 H' 的一个副本 H 存在于 G 中，并且 $v \in V(H')$，那么 $N(v) \setminus V(H') \neq \emptyset$，因此无论移除哪条关联边，仍然会存在一个 H 的副本。

另一方面，最大度为 $2\ell - 1$ 的图若不含 H，则至多有 $(\ell-1)(4\ell-3)$ 条边。这是因为在这样的图中，每条边至多与 $(4\ell-4)$ 条其他边关联，所以如果边数超过 $(\ell-1)(4\ell-3)$，我们就可以通过反复选择一条边并将与其关联的边标记为不可选，从而找到 H 的一个副本。

以下算法基于上述两个事实。我们保留一个图 G'，并将其初始化为 $\emptyset$。对于每条输入边 $e \in E(G)$，当且仅当将 e 存入 G' 后，G' 中的所有顶点的度仍小于 2ℓ 时，我们才将其存入。在此过程中，如果某一时刻 $|E(G')| > (\ell-1)(4\ell-3)$，那么我们就知道 $H \subseteq G'$，并且可以找到一个。否则，在数据流结束时，我们仍有$|E(G')| \le (\ell-1)(4\ell-3)$，接着检查 $H \subseteq G'$，这等价于检查 $H \subseteq G$。◀

如区段 3末尾所述，我们有以下关于上界的开放问题。

▶ **问题 3。** 设 H 是一个固定的图。对于 $p = \mathrm{poly}(n)$，$R^p_{2/3}(\textit{Sub}(H)) = \tilde{O}(\mathrm{ex}(n,H)/p)$ 是否成立？

允许随机性和 $p = \mathrm{poly}(n)$ 遍次，下面给出一个简单的 $o(\mathrm{ex}(n,H))$-空间随机化算法，但其空间复杂度仍远未达到 $\Omega(\mathrm{ex}(n,H)/p)$。

▶ **命题 52。** 设 H 是一个固定的图，其最小顶点覆盖的大小为 $\tau(H)$，则对于所有 $p = \omega(1)$，我们有

$$R^{p^{\tau(H)}}_{2/3}(\textsc{Sub}(H)) = \tilde{O}(\mathrm{ex}(n,H)/p + n).$$

证明。设 $\{v_1,\ldots,v_{\tau(H)}\}$ 是 H的一个顶点覆盖。不妨假设 $\mathrm{ex}(n,H)/p = \Omega(n)$。我们声明一个集合 E，用于存储至多 $4\mathrm{ex}(n,H)/p$ 条边，并声明一个顶点集 S，二者均在每一遍开始时初始化为 $\emptyset$。在每一遍中，我们首先通过以概率 $1/p$ 独立地选择每个$v \in V(G)$来构造 S。随后，当每个 $e \in E(G)$ 到达时，当且仅当 $e \cap S \neq \emptyset$ 且 E中仍有空间时，我们才将 e存储到 E 中。在这一遍结束时，如果我们在 E中找到 H的一个副本，则输出它并停止。否则，我们继续进行下一遍。算法结束时，如果我们从未找到 $H \subseteq E$，则输出 $H \not\subseteq G$。

如果 $H \not\subseteq G$，那么该算法始终正确。反过来，令 X 表示与 S关联的边数，则在每一遍次中，E 捕获一条 H 的概率至少为

$$\mathbb{P}(v_1,\ldots,v_{\tau(H)} \in S)\mathbb{P}\Big(X \le 4\mathrm{ex}(n,H)/p \Big| v_1,\ldots,v_{\tau(H)} \in S\Big).$$

X 的条件期望为

$$\mathbb{E}\Big[X\Big|v_1,\dots,v_{\tau(H)}\in S\Big]\le n\tau(H)+\sum_{e\in E(G)}\mathbb{P}(e\cap S\neq\emptyset)\le n\tau(H)+\frac{2\mathrm{ex}(n,H)}{p},$$

根据马尔可夫不等式，成功概率至少为

$$\begin{aligned}&\mathbb{P}(v_1,\dots,v_{\tau(H)}\in S)\Big[1-\mathbb{P}\Big(X>4\mathrm{ex}(n,H)/p\Big|v_1,\dots,v_{\tau(H)}\in S\Big)\Big]\\&\ge\frac{1}{p^{\tau(H)}}\Big[1-\frac{\mathbb{E}[X|v_1,\dots,v_{\tau(H)}\in S]}{4\mathrm{ex}(n,H)/p}\Big]\ge\frac{1}{4p^{\tau(H)}},\end{aligned}$$

因此，使用 $\tilde{O}(p^{\tau(H)})$ 遍数，我们以高概率成功。 ◀

特别地，当 H 是偶圈 $C_{2\ell}$ 时，存在一个随机化流式算法，它使用 $\tilde{O}(p^\ell)$ 遍数和 $\tilde{O}\Big(\frac{\mathrm{ex}(n,C_{2\ell})}{p}+n\Big)$ 空间，并以高概率要么在输入的 n-顶点图 G 中找到 $C_{2\ell}$ 的一个副本，要么正确报告不存在这样的副本。如果我们令 $p:=\mathrm{ex}(n,C_{2\ell})/n$，则空间界为 $O(n)$，而遍数变为

$$\tilde{O}\left(\left(\frac{\mathrm{ex}(n,C_{2\ell})}{n}\right)^\ell\right)=\tilde{O}\left(\frac{\mathrm{ex}(n,C_{2\ell})^\ell}{n^\ell}\right).$$

根据 [3] 我们有 $\mathrm{ex}(n,C_{2\ell})=O\big(n^{1+1/\ell}\big)$，因此上述是一个 $\tilde{O}(n)$-遍次，$\tilde{O}(n)$-空间随机化流式算法，用于查找 $C_{2\ell}$。这自然引出一个问题：这些权衡是否是紧的（在多对数因子范围内）。

▶ **问题 4。** 命题 52 对于 $H=C_{2\ell}$是否是紧的 $p=O\left(\mathrm{ex}(n,\ C_{2\ell})/n\right)$？等价地，对于 $R^{p^\ell}_{2/3}\left(\mathrm{SUB}\left(C_{2\ell}\right)\right)\ =\Theta\left(\mathrm{ex}(n,\ C_{2\ell})/p\right)$，是否有此结论?

Erdős 的一个著名猜想 [8] 询问：对于所有 ℓ，$\mathrm{ex}(n,C_{2\ell})=\Theta(n^{1+1/\ell})$。除了 $\ell=2$、3、5 之外，对于所有 ℓ，这个问题仍未解决。假设存在一个数 ℓ，使得 $\mathrm{ex}(n,C_{2\ell})=\tilde{o}(n^{1+1/\ell})$，那么命题 52 蕴含：对于该问题，存在一个使用 n空间、运行 n遍次的 $\tilde{o}$算法。因此，若以下问题的答案为肯定，则令人意外的是，这将意味着 $\mathrm{ex}(n,C_{2\ell})$ 的猜想数量级成立，误差至多相差多对数因子。

▶**问题 5。** 对于每个随机化的数据流 $Sub\left(C_{2k}\right)$ 算法，若其使用 $O\left(n\right)$ 空间，是否都需要 $\Theta\left(n\right)$ 遍次才能以至少 $2/3$ 的概率成功?

特别是，如果问题4 成立，那么问题5 就变得（相差多对数因子）等价于关于 $\mathrm{ex}(n,C_{2k})$ 的 Erdős 下界猜想。还值得注意的是，即使问题5 的答案为肯定，问题4的答案也可能不是肯定的。此外，也可能两个问题的答案都是否定的，而 Erdős 猜想仍然成立。

7.3 NWO 二分图的多遍行为

▶**问题6。** 设 H 是一个 *NWO* 二分图。如果 H 有一个包含环的 *NWO* 连通分量，那么是否有 $R^p_{2/3}(Sub(\vec{H}))\ =\Omega(n^2/p)$?

在多遍设置下，定理6 表明，NWO 二分图要么是一个困难实例，其空间复杂度为 $R^p_{2/3}(\mathrm{SUB}(\vec{H}))=\Omega(n^2/p)$；要么是一个简单实例，其空间复杂度为 $D^p(\mathrm{SUB}(\vec{H}))=\tilde{O}(n)$，其中 p 为某个常数；所有 NWO 森林都属于后一类，而高效算法由命题36给出。自然会问，多遍设置下是否也存在二分定理。

类似于我们证明单遍推论45的方式，如果 H 是一个 NWO 森林与 WO 二分图的并集，那么通过为命题 $\vec{H}$ 36中的算法增加 $\tilde{O}(n^{2-1/\Delta'(H)})$额外空间，36可知它在多遍设置下是一个简单实例。我们猜想，所有简单实例都可以通过这种方式得到。也就是说，只要 H 含有一个包含环的 NWO 连通分量， H就应该是一个困难实例。例如，以下就是一个困难实例。

命题53。设 H 为一个 *NWO* 二分图，其定义如下

$$V(\vec{H}) = \{v_1, v_2, v_3, v_4, v_5\}, \quad E(\vec{H}) = \{(v_1, v_2), (v_3, v_2), (v_1, v_4), (v_3, v_4), (v_4, v_5)\},$$

则 $R^p_{2/3}(\textsc{Sub}(\vec{H})) = \Omega(n^2/p)$。

证明。给定一个 不相交性问题$_{n^2}$实例，其中 Alice 和 Bob 分别持有 S_1， $S_2 \subseteq [n^2]$， 我们将有向归约图 $G_{\vec{H}}$ 定义在顶点集

$$V(\vec{G}_{\vec{H}}) = \bigsqcup_{i=1}^{n} \left\{v_1^{(i)}, v_2^{(i)}, v_3^{(i)}, v_4^{(i)}, v_5^{(i)}\right\}$$

并具有弧集 $E(\vec{G}_{\vec{H}}) = E_1 \cup E_2 \cup E_3$，其中

$$E_1 = \{(v_1^{(i)}, v_2^{(j)}) : i, j \in [n], (j-1)n + i \in S_1\},$$
$$E_2 = \{(v_3^{(i)}, v_2^{(j)}) : i, j \in [n], (j-1)n + i \in S_2\},$$
$$E_3 = \bigsqcup_{i=1}^{n} \left\{(v_1^{(i)}, v_4^{(i)}), (v_3^{(i)}, v_4^{(i)}), (v_4^{(i)}, v_5^{(i)})\right\}.$$

若 $(j-1)n + i \in S_1 \cap S_2$ 对某个 i， $j \in [n]$成立，则由以下方式构造出 $\vec{H}$ 的一个副本

$$v_1^{(i)}, v_2^{(j)}, v_3^{(i)}, v_4^{(i)}, v_5^{(i)}.$$

反过来，如果 $\vec{G}_{\vec{H}}$ 中存在 $\vec{H}$ 的某个副本 $\vec{H}'$，那么 $\vec{H}'$ 中的 NWO 顶点必定是某个 $i \in [n]$，因为这些是 $\vec{G}_{\vec{H}}$ 中仅有的 NWO 顶点。这个 $v_4^{(i)}$ 必须位于 $\vec{H}'$ 中的一条 $\vec{C}_4$ 上，但它只能由某个 $j \in [n]$ 的 $v_4^{(i)}$ 、 $v_1^{(i)}$ 、 $v_2^{(j)}$ 、 $v_3^{(i)}$ 构成，因此 $(j-1)n + i \in S_1 \cap S_2$。◀

一般来说，我们猜想，类似命题40中的证明或许能解决问题6。也就是说，或许对于每个环，我们都可以记录每个 NWO 顶点到该环的距离之和，然后谨慎地选取一个使此距离最小的环，让它的某些弧编码 集合不相交问题或甚至多重不相交性问题的输入。然而，障碍在于，这类归约往往需要在不同顶点副本之间引入交叉弧，如命题13的证明中那样，但这会影响命题40的证明；在该证明中，我们要计算一个环上有多少个顶点可以作为从 NWO 顶点出发的目标最短路径的端点。也许还需要更多思路。

7.4 NWO 二分图的确定性单遍上界

问题 7。 设 H 为一棵至多含有一个 *NWO* 顶点的定向树。我们是否有

$$D^1(\textsc{Sub}(\vec{H})) = \tilde{O}(n)?$$

46 流式空间复杂度：子图查找问题的二分性

在命题41中，我们使用 ℓ_0采样器来证明 $R^1_{2/3}(\textsc{Sub}(\vec{H})) = \tilde{O}(n)$ 对于至多含有一个 NWO 顶点的定向树成立，从而得出$\textsc{Sub}(\vec{H})$的单遍二分定理。然而，这意味着我们的算法无法直接去随机化，因此人们可能会问，问题7 中提到的确定性版本 是否确实存在。我们有一些此类 H 的例子，其中$D^1(\textsc{Sub}(\vec{H})) = \tilde{O}(n)$的算法基于颜色编码。下面给出一个最简单的例子。不过，一般情形仍未解决。

▶**命题 54**。 设 $\vec{H}$ 为定向路径 $v_1 \to v_2 \to v_3 \leftarrow v_4$，则 $D^1(\mathit{Sub}(H)) = \tilde{O}(n)$。

证明。根据引理 14，我们可以取 $L = O(\log^2 n)$ 个着色 χ_1、…、χ_L： $V(\vec{G}) \to [4]$ 使得 $\vec{\ }$，如果 G 包含 H 的一个副本，那么该副本在其中一个着色下是彩色的。特别地，如果我们考虑四种颜色的所有可能排列 σ： $[4] \to [4]$，并将$\sigma \circ \chi_1$、…、$\sigma \circ \chi_L$复合起来，那么我们得到 $24L$ 个着色 χ_1、…、χ_{24L}，使得如果 G 包含H的一个副本，那么对于该副本，在某个着色 χ_c 下有

$$\chi_c(v_1) = 1, \quad \chi_c(v_2) = 2, \quad \chi_c(v_3) = 3, \quad \chi_c(v_4) = 4.$$

考虑每个 $c \in [24L]$ 和 $w \in V(\vec{G})$。如果 $\chi_c(w) \in \{2, 4\}$，则我们声明两个变量 $\mathrm{IN}_c(w)$，$\mathrm{ON}_c(w) \in V(\vec{G})$，并将其初始化为 w；而如果 $\chi_c(w) = 3$，则我们声明一个列表 $\mathrm{LN}_c(w) \subseteq V(\vec{G})$，其大小至多为 2，并将其初始化为 $\emptyset$。这使用的空间为 $\tilde{O}(n)$。

现在，对于数据流中的每条输入弧 $\vec{e} = (u, v)$，我们对所有 $c \in [24L]$ 执行以下操作：

（1）如果 $\chi_c(u) = 1$，$\chi_c(v) = 2$，并且 $\mathrm{IN}_c(v) = v$，则将 $\mathrm{IN}_c(v)$ 设置为 $= u$。（2）如果 $\chi_c(u) \in \{2, 4\}$，$\chi_c(v) = 3$ 且 $|\mathrm{LN}_c(v)| \le 1$，则将 u 添加到 $\mathrm{LN}_c(v)$ 中。此外，如果这使得 $|\mathrm{LN}_c(v)| = 2$，则将 $\mathrm{ON}_c(u)$ 设置为 $= v$。（3）如果以上条件均不满足，则不执行任何操作。

最后，我们检查是否存在某些 $c \in [24L]$ 和 $w \in V(\vec{G})$，使得 $\chi_c(w) \in \{2, 4\}$ 满足 $\mathrm{IN}_c(w), \mathrm{ON}_c(w) \neq w$。若存在，则必定有一个 $\vec{H}$ 的副本，如下所示

$$\mathrm{IN}_c(w) \to w \to \mathrm{ON}_c(w) \leftarrow v, \quad v \in \mathrm{LN}_c(\mathrm{ON}_c(w)) \setminus \{w\},$$

因此我们输出 $\vec{H} \subseteq \vec{G}$。若不存在，则我们得出结论 $\vec{H} \not\subseteq \vec{G}$；这是正确的，因为如果存在 H 的一个 $\vec{\ }$副本，那么在某种 c 下，它会看到所需的颜色模式，因此我们会得到上述证书 $w \in V(\vec{G})$。◀

参考文献

1 Noga Alon、Michael Krivelevich 和 Benny Sudakov。二分图的 Turán 数及相关的拉姆齐型问题。 组合概率与计算，12(5-6):477–494，2003。2 Noga Alon、Raphael Yuster 和 Uri Zwick。颜色编码。 *ACM*学报，42(4):844–856，1995。3 J. A. Bondy 和 M. Simonovits。图中的偶长度圈。组合理论杂志 *B* 辑，16(2):97–105，1974。4 Vladimir Braverman 、 Zaoxing Liu 、 Tejasvam Singh、N. V. Vinodchandran 和 Lin F. Yang。利用图分解理论为 CLIQUE-GAP问题提供新界。 *Algorithmica*，80(2):652–667，2018。5 Vladimir Braverman、Rafail Ostrovsky 和 Dan Vilenchik。流式模型中的三角形计数有多难？载于 Fedor V. Fomin、Rusins Freivalds、Marta Z. Kwiatkowska 和 David Peleg（编）， 自动机、语言与编程——第*40*届国际研讨会（*ICALP*），计算机科学讲义第7965卷， 计算机科学讲义，第244–254页。施普林格，2013。

6 Amit Chakrabarti、Subhash Khot 和 Xiaodong Sun。集合不相交问题多方通信复杂度的近乎最优下界。载于 IEEE计算复杂性会议（CCC）论文集，2003年。 7 Tobias Lybecker Christiani 和 Rasmus Pagh。在常数时间内生成 k-独立变量。载于第55届 IEEE计算机科学基础研讨会（FOCS），第196–205页。IEEE，2014年。 arXiv:1408.2157, doi:10.1109/FOCS.2014.29。 8 Paul Erdős 和 Miklós Simonovits。极值图论中的紧性结果。组合数学，2(3):275–288，1982年。 doi:10.1007/BF02579234。 9 P. Erdös 和 A. H. Stone。论线性图的结构。美国数学会公报，52(12):1087–1091，1946年。 10 Joan Feigenbaum、Sampath Kannan、Andrew McGregor、Siddharth Suri 和 Jian Zhang。半流式模型中的图问题。载于 Josep Díaz、Juhani Karhumäki、Arto Lepistö 和 Donald Sannella（编），自动机、语言与编程：第31届国际研讨会（ICALP），计算机科学讲义第3142卷，第531–543页。施普林格，2004年。 11 Hendrik Fichtenberger 和 Pan Peng。数据流中子图的近似计数。载于 Leonid Libkin 和 Pablo Barceló（编），（PODS）'22：数据管理国际会议，第413–425页。ACM，2022年。 12 Andre Gronemeier。与函数和不相交性问题的 nih 多方信息复杂度的渐近最优下界。载于 STACS论文集，2009年。 13 Magnús M. Halldórsson、Xiaoming Sun、Mario Szegedy 和 Chengu Wang。团近似的流式与通信复杂度。载于 Artur Czumaj、Kurt Mehlhorn、Andrew M. Pitts 和 Roger Wattenhofer（编），自动机、语言与编程——第39届国际研讨会（ICALP），计算机科学讲义第7391卷，第449–460页。施普林格，2012年。 14 T. S. Jayram、Ravi Kumar 和 D. Sivakumar。汉明距离的单向通信复杂度。计算理论，4:129–135，2008年。 15 A. Joffe。关于一组几乎确定性的 k-独立随机变量。概率年刊，2(1):161–162，1974年。 doi:10.1214/aop/1176996762。 16 Hossein Jowhari、Mert Sağlam 和 Gábor Tardos。lp 采样器、数据流中重复项查找及相关问题的紧确界，2010年。URL: https://arxiv.org/abs/1012.4889,arXiv:1012.4889。 17 Bala Kalyanasundaram 和 Georg Schnitger。集合交的概率通信复杂度。载于第2届年度复杂性理论结构会议论文集，第41–49页，1987年。 18 Bala Kalyanasundaram 和 Georg Schnitger。集合交的概率通信复杂度。SIAM离散数学杂志，5(4):545–557，1992年。 doi:10.1137/0405044。 19 Daniel M. Kane、Kurt Mehlhorn、Thomas Sauerwald 和 He Sun。数据流中任意子图的计数。载于 Artur Czumaj、Kurt Mehlhorn、Andrew M. Pitts 和 Roger Wattenhofer（编），自动机、语言与编程——第39届国际研讨会（ICALP），计算机科学讲义第7392卷，第598–609页。施普林格，2012年。 20 Il an Kremer、Noam Nisan 和 Dana Ron。随机化单轮通信复杂度。计算复杂性，8(1):21–49，1999年。 21 Eyal Kushilevitz 和 Noam Nisan。通信复杂度。剑桥大学出版社，1997年。 22 Andrew McGregor。图流算法：综述。SIGMOD Rec.，43(1):9–20，2014年。 23 Andrew McGregor 和 Sofya Vorotnikova。数据流模型中的三角形与四环计数。载于 Dan Suciu、Yufei Tao 和 Zhewei Wei（编），第39届 ACM SIGMOD-SIGACT-SIGAI数据库系统原理研讨会（PODS）论文集，第445–456页。ACM，2020年。

24 Stavros D. Nikolopoulos 和 Leonidas Palios。高效并行识别余图。离散应用数学，150(1–3):182–215，2005年9月。25 Jelle J. Oostveen 和 Erik Jan van Leeuwen。流式直径与连通性问题的参数化复杂度。*Algorithmica*，86(9):2885–2928，2024年。26 Jeanette P. Schmidt 和 Alan Siegel。无偏 k探测哈希函数的空间复杂度。*SIAM*计算杂志，19(5):775–786，1990年。27 Payam Valadkhan。极值有向图与 Erdős–Hajnal 猜想。硕士论文，沙里夫理工大学，德黑兰，伊朗，2007年。28 Mark N. Wegman 和 J. Lawrence Carter。新型哈希函数及其在身份验证和集合相等性中的应用。计算机与系统科学杂志，22(3):265–279，1981年。 doi:10.1016/0022-0000(81)90033-7。

算法 1 流式模型中的余图识别 **1:** 过程 **IsCograph(**G**) 2:** $n \leftarrow |V(G)|$**3:** 若 $n \le 3$ 则返回 真 **4:** 结束 **if** 预处理。 **5:** $(\mathbf{P}$ `foundP4`$) \leftarrow$ **PreprocessComponents(**G**)** 引理 25 **6:** 若 `foundP4` 则返回 **False 7:** 结束 **if 8:** 若 $|\mathbf{P}| > 1$ 则返回 **AllCographs(**G**, P)** 每遍次并行处理所有部 **9:** 结束 **if** 选择一个根并构造分量划分。 **10:** $(\mathbf{Low}, \mathbf{Mid}, \mathbf{High}) \leftarrow$ **DegreeClasses(**G**) 11:** 若 **Mid** $\neq \emptyset$ 则 **12:** 任取 $u \in$ **Mid 13:** 否则 **14:** 若 **Low** $= \emptyset$ 或 **High** $= \emptyset$ 则返回 **False 15:** 结束 **if 16:** $u \leftarrow$ **arg max**$_{x \in \mathrm{Low}} |N($**x**$) \cap$ **High**$|$**17:** 结束 **if 18:** $(\{A_i\}_{i=1}^{k}, \{B_i\}_{i=1}^{k},$ `ok`$) \leftarrow$ **ComponentPartition(**G, u**)** 引理 28 **19:** 若 `ok`=**False** 则返回 **False 20:** 结束 **if** 细化各部并递归。 **21:** Q $\leftarrow$ **RefineParts(**G**,** $\{A_i\}$**,** $\{B_i\}$**) 22:** 若 **max**$_{S \in \mathcal{Q}} |S| \le 3n/4$ 则返回 **AllCographs(**G**,** Q**)** 每遍次并行处理 **23:** 结束 **if 24:** 选择一个 $F \in$ **Q**，满足 $|F| > 3n/4$**25:** 若 $F \cap$**Low** $\neq \emptyset$ 则返回 **False 26:** 结束 **if** $F \subseteq B_i$，引理 30 **27:** $(\mathbf{R}$ `foundP4`$) \leftarrow$ **CoComponentsOn(**G, F**)** 在 $G[F]$ 上应用引理 25 **28:** 若 `foundP4` 则返回 **False 29:** 结束 **if 30:** Q $\leftarrow$ **(Q** $\setminus \{F\}$**)** $\cup$ **R** 返回 **AllCographs(**G**,** Q**)**，每遍次并行处理 **31:** 结束过程 **32:** 函数 **AllCographs(**G**,** S**) 33:** 对所有 $S \in$**S**，在每一遍次中同时维护所有子实例 **34:** 若 **IsCograph(**$G[S]$**)**=**False** 则返回 **False 35:** 结束 **if 36:** 结束 **for**返回 真 **37:** 结束函数 **38:** 函数 **RefineParts(**G**,** $\{A_i\}$**,** $\{B_i\}$**) 39:** Q $\leftarrow \emptyset$**40:** 对 $i = 1$ 至 k 执行 **41:** C $\leftarrow$**ConnComponents(**$G[A_i]$**) 42:** Q $\leftarrow$ Q $\cup$**C 43:** C $\leftarrow$**ConnComponents(**$G[B_i]$**) 44:** Q $\leftarrow$ Q $\cup$**C 45:** 结束 **for 46:** 返回 Q **47:** 结束函数

辅助引理

为完整起见，我们在此证明推论 11 ，这是对 [1] 的直接推广。

▶ **引理 55** (依赖随机选择)。 设 G 是一个二分图，其两部为 $V(G) = V_1 \sqcup V_2$，其中 $|V_1| = |V_2| = n/2$ 且 $d = 2|E(\vec{G})|/n$。若 a、b、$r \in \mathbb{N}$ 且

$$\frac{d^r}{n^{r-1}} - \binom{n}{r}\left(\frac{b-1}{n}\right)^r \geq a,$$

那么存在某个 $A_0 \subseteq V_1$、$|A_0| \geq a$ ，使得从 A_0 中任取 r 个顶点，它们都至少有 b个公共邻居。

证明。 令 T 为 V_2中的 r 个随机顶点组成的集合，按均匀分布并允许重复抽取。令 A为所有 T中顶点的公共邻居集合，根据詹森不等式，

$$\mathbb{E}[|A|] = \sum_{v\in V_1}\left(\frac{\deg(v)}{n/2}\right)^r = \frac{2^r}{n^r}\sum_{v\in V_2}\deg(v)^r \geq \frac{2^r}{n^{r-1}}\left(\frac{\sum_{v\in V_1}\deg(v)}{n}\right)^r = \frac{d^r}{n^{r-1}}.$$

另一方面，定义

$$Y = \Big\{W \subseteq A : |W| = r \text{ but the vertices in } W \text{ share } \leq b-1 \text{ common neighbors}\Big\}.$$

若 $U \subseteq V_1$ 具有 $|U| = r$ ，且 U 中的顶点共享 $\leq b-1$ 个公共邻居，则 $U \subseteq A$ 是 $(b-1/n)^r$的概率为，因此

$$\mathbb{E}[|Y|] \leq \binom{n}{r}\left(\frac{b-1}{n}\right)^r.$$

特别地，某个 T 的取值使得

$$|A| \geq \frac{d^r}{n^{r-1}}, \quad |Y| \leq \binom{n}{r}\left(\frac{b-1}{n}\right)^r.$$

从 A中，我们从每个 $W \in Y$移除一个顶点，得到 A_0，于是

$$|A_0| \geq |A| - |Y| = \frac{d^r}{n^{r-1}} - \binom{n}{r}\left(\frac{b-1}{n}\right)^r \geq a$$

并且 A_0 中的每个 r 个顶点都至少共享 b个公共邻居。 ◀

▶ **推论 11**。 设 $\vec{H}$ 关于二部划分 $V(H) = A(H) \sqcup B(H)$是 *WO*，且

$$\Delta'(H) = \min\left\{\max_{v\in A(H)}\deg_H(v), \max_{v\in B(H)}\deg_H(v)\right\},$$

则 $\mathrm{ex}(n, \vec{H}) = O(n^{2-1/\Delta'(H)})$。

证明。 不失一般性地假设 $\Delta'(H) = \max^{v\in B(H)}\deg_H(v)$。令 $a = |A(H)|$，$b = |B(H)|$，且 $C > 0$ 为一个常数，使得

$$(2C)^{\Delta'(H)} - \binom{n}{\Delta'(H)}\left(\frac{b-1}{n}\right)^{\Delta'(H)} \geq a.$$

我们证明，每个具有 $|E(G_0)| \geq 5Cn^{2-1/\Delta'(H)}$ 的 n 个顶点定向图 $\vec{G}_0$ 都包含一个 $\vec{H}$ 的副本。事实上，这样的有向图 $\vec{G}_0$ 含有一个二部子图 $\vec{G}_1$，其平衡部 $V(\vec{G}_1) = V_1 \sqcup V_2$ 满足

$|E(\vec{G}_1)| \geq 2Cn^{2-1/\Delta'(H)}$，于是通过选取两部之间的主要定向，我们可以进一步得到一个具有 $V(\vec{G}) = V_1 \sqcup V_2$ 和 $|E(\vec{G})| \geq Cn^{2-1/\Delta'(H)}$的良定向子图 $\vec{G}$。注意，对于 $d = 2|E(\vec{G})|/n$，我们有

$$\frac{d^r}{n^{r-1}} - \binom{n}{r}\left(\frac{b-1}{n}\right)^r \geq (2C)^r - \binom{n}{\Delta'(H)}\left(\frac{b-1}{n}\right)^{\Delta'(H)} \geq a.$$

不失一般性，假设 $\vec{H}$ 上的定向是从 $A(H)$指向 $B(H)$，且 $\vec{G}$ 上的定向是从 V_1 指向 V_2。根据引理 55，选取 $A_0 \subseteq V_1$，其中每个 $\Delta'(H)$顶点至少有 b 个公共邻居，且这些公共邻居属于 V_2，并任取一个双射

$$\varphi : A(H) \to A_0.$$

令 $B(H) = \{v_1, \ldots, v_b\}$。首先考虑 $N_H(v_1) \subseteq A(H)$。由于 $|N_H(v_1)| \leq \Delta'(H)$，根据 A_0 的选取，我们知道 $\varphi(N_H(v_1))$ 在 V_2中有多于 b 个公共邻居。从这些公共邻居中任取一个 u_1，并将双射扩展为

$$\varphi : A(H) \cup \{v_1\} \to A_0 \cup \{u_1\}.$$

归纳地，若我们已有双射

$$\varphi : A(H) \cup \{v_1, \ldots, v_k\} \to A_0 \cup \{u_1, \ldots, u_k\}$$

对于某个 $k < b$，接下来考虑 $N_H(v_k) \subseteq A(H)$。由于 $|N_H(v_k)| \leq \Delta'(H)$，我们再次知道 $\varphi(N_H(v_1k))$ 在 V_2中有多于 b 个公共邻居；特别地，这些公共邻居中至少有一个不在 $\{u_1, \ldots, u_k\}$ 中，因此可以令其为 u_{k+1}，然后通过映射 $v_{k+1} \mapsto u_{k+1}$ 来扩展 φ。最后，这个双射

$$\varphi : A(H) \cup B(H) \to A_0 \cup \{u_1, \ldots, u_b\}$$

$\vec{\ }$ $\vec{\ }$给出了 H 在 G中的一个副本。 ◀