arXiv:2304.06232v1 [cs.DB] 13 Apr 2023

单射语义下的合取正则路径查询

Diego Figueira
diego.figueira@cnrs.fr 波尔多大学、法国国家科学研究中心、波尔多国立高等工程学院、波尔多计算机科学研究实验室（UMR 5800），F-33400，塔朗斯，法国

Miguel Romero
miguel.romero.o@uai.cl 工程与科学学院，阿道夫·伊瓦涅斯大学，智利

摘要

我们为合取正则路径查询（CRPQ）引入单射语义，并研究其基本性质。我们确定了两种此类语义：原子单射语义和查询单射语义，它们均以单射同态定义。这些语义将研究充分的简单路径语义下的 RPQ 类自然推广到了 CRPQ 类。我们研究了这两种语义下的求值问题和包含问题，为其提供了有用的刻画，并明确了这些问题的复杂性。或许出人意料的是，我们证明，在原子单射语义下，CRPQ 的包含问题变得不可判定；而在查询单射语义下，该问题为 PSPACE 完全。这与标准语义下已知的指数空间完全性结果形成对比。所用技术与标准语义下的已知方法有显著差异，因此需要专为单射语义设计的新工具。为完整呈现复杂性图景，我们针对每种语义，研究了 CRPQ 主要子类的包含问题，即合取查询以及具有有限语言的 CRPQ。

CCS 概念

• **信息系统** → **非关系型引擎查询语言**；• **计算理论** → 数据库查询处理与优化（理论）。

关键词

图数据库、正则路径查询（RPQ）、包含、求值、简单路径、单射同态

1 引言

如今，图数据库在许多应用中都十分重要[2, 29]。简而言之，图数据库被抽象为边标记有向图，其中节点表示实体，标记边表示这些实体之间的关系。查询图数据库的一种基本方式是查找关于实体间相互关系的模式。在这方面，现代图查询语言的一种核心查询机制是正则路径查询（RPQ）。RPQ 提供了一种简单形式的递归，专门用于发现通过具有特定属性的路径相连的实体。这些查询

形式为 $x \xrightarrow{L} y$，其中 L 是数据库边标签字母表上的一个正则表达式。此类查询返回数据库中所有节点对（u，v），使得存在一条从 u 到 v 的（有向）路径，其标签与 L 匹配。

Diego Figueira 部分获得 ANR QUID 资助，项目编号 ANR-18-CE400031。Miguel Romero 的资助来源包括 Fondecyt 项目 11200956、数据观测基金会以及国家人工智能中心 CENIA FB210017 和 ANID 基础资助。

对 RPQ 进行合取和存在量化闭包运算，得到所谓的合取正则路径查询（CRPQ）。实际上，CR-PQ 可以理解为合取查询在这种简单递归形式下的推广。CRPQ 是 SPARQL 的一部分 [20]，SPARQL 是 W3C 查询 RDF 数据（图数据库的一种广泛使用的格式）的标准。DBpedia 和 Wikidata 等知名知识库就是 RDF 数据库的例子。例如，（C）RPQ 常用于查询 Wikidata [7, 23]。更一般地说，CRPQ 是查询图结构数据库的基本构件 [2, 4]。它们也是 G-core [1]和 Cypher[17], 的一部分；后者是 Neo4j 的查询语言，而 Neo4j 目前是最流行的商业图数据库之一。它们还属于面向图查询语言的 GQL 标准化工作 [13, 19]。

替代语义。 如上所述，评估 RPQ $x \xrightarrow{L} y$ 的语义通常假设，只要满足正则性质 L，就允许采用从 u 到 v 的任意（有向）路径。然而，也存在其他语义，允许将路径限制为不重复经过顶点（又称简单路径），或不重复经过边（又称迹）。这样一来，只需考虑有限数量的路径。事实上，这些替代语义在实践中以及数据库理论领域都受到了广泛关注。它们是 Neo4j 的 Cypher 查询语言的一部分，也被纳入 GQL，作为评估 RPQ 的可选方式。此外，自 80 年代末以来，文献已对这些替代语义进行了广泛研究 [3, 12, 22, 24–26]。然而，颇令人意外的是，这些研究主要聚焦于 RPQ，而基本未曾探讨替代语义下的 CRPQ。

贡献。 我们为 CRPQ 引入单射 语义，它推广了 RPQ 的简单路径语义，并研究了 CRPQ 在这些语义下的基本性质。具体而言，我们确定了两种可能的自然语义：

(1) 原子单射。 第一种语义要求每个 CRPQ原子 都采用 RPQ 的简单路径语义进行解释，也就是说，形如 $x \xrightarrow{L} y$ 的原子必须映射到简单路径，而形如 $x \xrightarrow{L} x$ 的原子必须映射到简单环。尤其是，不要求不同原子的路径彼此不相交。在这种语义下，像 $Q = \exists x, y, z(x \xrightarrow{(a+b)^+} y \wedge x \xrightarrow{(b+c)^+} z)$这样的布尔合取正则路径查询，在由一条包含 b 个节点的有向路径构成的图数据库中成立：将 x 映射到路径的第一个节点，并将 y 和 z 映射到最后一个节点。

（2）查询单射。 另一种选择是考虑一种限制性更强的语义，在该语义下，与不同原子对应的路径必须映射到不同的节点，因此路径内部以及路径之间都不能出现重复节点。这种语义

同时推广了简单路径语义下的正则路径查询和单射语义下的合取查询。在这种情况下，只有当数据库中包含两条从同一节点出发、具有相应语言且彼此不相交（起点除外）的简单路径时，上述查询 Q才可能为真。

我们将前者称为原子单射语义，因为“节点不得重复”的条件需要分别对每个原子施加；将后者称为查询单射语义，因为单射性要求适用于整个查询。这三种语义（标准、原子单射、查询单射）构成一个层次结构，其中查询单射语义限制最严格，标准语义限制最宽松。

我们首先考察 CRPQ 的求值问题，即检查元组 $\bar{v}$ 是否属于某个特定图数据库 G 上查询 Q 的结果。对于标准语义，该问题在组合复杂度下是 NP 完全的，也就是说，查询和数据库都是输入的一部分；而在数据复杂度下是 NL 完全的，也就是说，查询被视为固定不变。在两种单射语义下，求值问题在组合复杂度下仍是 NP 完全的，而在数据复杂度下则变为 NP 完全。这直接源于以下事实：即使对于非常简单的 RPQ，在简单路径语义下，RPQ 求值问题也是 NP 完全的 [4, 26]。

因此，我们将注意力转向包含问题，这是本文的主要研究重点。该问题询问，对于任意底层数据库，查询 Q_1 返回的每个结果是否也会由查询 Q_2返回。包含性检查是最基本的静态分析任务之一，也可用于查询优化。已知 CRPQ 在标准语义下的包含问题是指数空间完全的 [9, 16]，而针对形状受限 [14] 或正则表达式受限 [15]的查询，对包含问题的研究也已得到拓展。

令人相当意外的是，正如我们将要展示的，三种语义（标准语义、原子单射语义、查询单射语义）的层次结构并未体现在包含问题的复杂性上：限制最严格的语义（查询单射语义）是 PSPACE 完全的，限制最宽松的语义（标准语义）是 ExpSpace 完全的，而居于两者之间的语义（原子单射语义）则是不可判定的。关于原子单射语义和查询单射语义的这些结果，是本文的主要技术贡献。我们还通过全面研究两类主要 CRPQ 子类的包含问题，补全了整体图景：合取查询，以及不含克莱尼星号运算符的 CRPQ。我们给出了所有可能组合的复杂性完全性结果（参见图 1）。我们的主要结果需要发展新的技术，这些技术有助于深入理解在这些语义下处理静态分析时面临的微妙难点。

结构安排。在第 2 节介绍预备知识之后，我们将在第 3 节和第 4 节分别定义并刻画求值问题和包含问题。第 5 节研究一般 CRPQ 的包含问题，第 6 节研究 CRPQ 的若干子类。最后在第 7 节总结。省略或略述的证明见附录。

2 预备知识

我们假设读者熟悉正则语言、正则表达式和非确定有限自动机（非确定有限自动机）。我们常常不区分正则表达式及其定义的语言；非确定有限自动机也是如此。

图数据库和路径。字母表 $\mathbb{A}$ 上的图数据库是有限的边标记图 $G=(V,E)$，其字母表为 $\mathbb{A}$，其中 V 是有限顶点集，$E \subseteq V \times \mathbb{A} \times V$ 是标记边集（或简称边）。我们用 $u \xrightarrow{a} v$ 表示一条边 $(u,a,v) \in E$。图数据库 $G=(V,E)$ 中从 u 到 v 的一条路径，是由图中的边组成的（可能为空的）序列

$$\pi = v_0 \xrightarrow{a_1} v_1 \ v_1 \xrightarrow{a_2} v_2 \ \cdots \ v_{k-1} \xrightarrow{a_k} v_k$$

其中，G 且 $k \geq 0$。此类路径的内部节点是任何满足 $u = v_0$ $v = v_k$ $0 < i < k$ 的节点 v_i。π 的标签是词 $a_1 \ldots a_k \in \mathbb{A}^*$。当 $k = 0$ 时，π 的标签为空词 ε。如果所有节点 v_i 两两不同，我们称 π 为简单路径；如果 $v_0 = v_k$ 且所有节点 v_i（对于 $i < k$）两两不同，则称其为简单环。

合取查询与同态。在图数据库的语境中，有限字母表 $\mathbb{A}$ 上的合取查询（CQ）Q是一种表达式 $Q(x_1, \ldots\ldots, x_n) = A_1 \wedge \cdots \wedge A_m$，其中 $m \geq 0$，$(x_1, \ldots\ldots, x_n)$ 是一个变量元组，并且每个 A_i 都是形如 $x \xrightarrow{a} y$ 的原子，其中 x 和 y 是变量，且 $a \in \mathbb{A}$。我们用$vars(Q)$ 表示出现在 Q中的变量集合。为了强调 Q的自由变量 元组 $\bar{x} = (x_1, \ldots\ldots, x_n)$，我们常写作 $Q(\bar{x})$，而不写 Q。我们假设自由变量 x_i不一定互不相同。Q中不属于 $\{x_1, \ldots\ldots, x_n\}$ 的变量均为（隐式）存在量化变量。与通常做法一样，如果 $\bar{x}$为空，我们称 CQ Q为布尔查询。注意，每个 CQ 都可以看作一个图数据库（每个原子都是一条边），因此，稍微滥用一下术语时，我们有时也会对 CQ 使用图数据库术语。

从合取查询 $Q(\bar{x})$ 到图数据库 $G=(V,E)$ 的同态 h 是一个从 $\mathrm{vars}(Q)$ 到 V 的映射，使得对于 Q 的每个原子 $x \xrightarrow{a} y$，$h(x) \xrightarrow{a} h(y)$ 属于 E。如果此外对于任意一对不同的变量 x 和 y，我们都有 $h(x) \neq h(y)$，则称 h 是单射的。若存在从 Q 到 G 的同态，我们记作 $Q \to G$；若 h是这样的同态，则记作 $h\colon Q \to G$。类似地，对于元组 $\bar{v}$，若存在从 Q 到G 的同态 h，使得 $h(\bar{x}) = \bar{v}$，则记作 $Q \to (G, \bar{v})$；我们用 $Q \to (G, \bar{v})$ 明确表示这样的 h。对于单射同态，我们使用类似的记法，将 $\to$ 替换为 $\mathrm{inj} \longrightarrow$。合取查询之间的同态基本上按前述方式定义，区别在于自由变量映射到自由变量。也就是说，给定两个合取查询 $Q_1(\bar{x}_1)$, $Q_2(\bar{x}_2)$，若 $h\colon Q_1 \to (G, \bar{x}_2)$，则有 $h\colon Q_1 \to Q_2$；若 $h\colon Q_1 \mathrm{inj}\longrightarrow (G, \bar{x}_2)$，则有 $h\colon Q_1 \mathrm{inj}\longrightarrow Q_2$，其中 G 是记作 Q_2 的图数据库。

我们还研究带等式原子的合取查询，其形式为 $Q(\bar{x}) = P \wedge I$，其中 P是不带等号原子的合取查询，而I是形如 $x = y$的等式原子的合取（变量 x和 y可能不属于 $\mathrm{vars}(P)$）。同样地，我们用 $\mathrm{vars}(Q)$表示出现在 Q中的变量集合。我们将 $\mathrm{vars}(Q)$上的二元关系$=_Q$定义为二元关系 $\{(x,y)\colon x = y$ 是 Q中的等式原子$\}$的自反对称传递闭包。换言之，如果等式 $x = y$ 由 Q的等式原子强制推出，则有 $x =_Q y$。注意，每个形式为$Q(\bar{x}) = P \wedge I$的带等式原子的合取查询都等价于一个不带等号原子的合取查询 $Q^{=}$；后者由 Q 通过将关系 $=_Q$ 的每个等价类合并为单个变量得到。该变换给出了一个典范重命名，我们始终将其记为 Φ，它从 $\mathrm{vars}(Q)$映射到 $\mathrm{vars}(Q^{=})$，定义为 $\Phi(x) = C$，其中 C是包含 x的等价类。特别地，$Q^{=}$ 的自由变量元组为 $\Phi(\bar{x})$。

	合取查询/合取查询	合取查询/CRPQ	CRPQ/合取查询	合取查询/CRPQfin	CRPQfin/合取查询
标准	NP-困难 [10]	NP-困难 (†)	Π_2^p-c (‡)	NP-困难 (†)	Π_2^p-c (§,‡)
查询单射	NP-困难 (F.2)	NP-困难 (F.2)	Π_2^p-c (6.1,F.7)	NP-困难 (F.2)	Π_2^p-c (6.1,F.7)
原子单射	NP-困难 (F.4)	Π_2^p-c (6.2,F.10)	Π_2^p-c (F.6,F.7)	Π_2^p-c (6.2,F.10)	Π_2^p-c (F.6,F.7)

	CRPQ/CRPQfin	CRPQfin/CRPQ	CRPQfin/CRPQfin	CRPQ/CRPQ
标准	PSPACE 完全 (F.8,F.9)	Π_2^p-c (§,F.10)	Π_2^p-c（§，F.10）	指数空间完全 ($,¢)
查询单射	PSPACE 完全 (F.8,5.1)	Π_2^p-c (6.1,F.10)	Π_2^p-c (6.1,F.10)	PSPACE 完全 (F.8,5.1)
原子单射	不可判定 (5.2)	Π_2^p-c (6.2,F.10)	Π_2^p-c (6.2,F.10)	不可判定 (5.2)

†： [15, 定理 4.2] §： [15, 定理 4.3]‡： [15, 定理 4.4] $： [9, 定理 6] ¢： [16, 定理 4.8]**图 1：标准、查询单射和原子单射语义下包含问题的复杂性。方括号中的数字指向命题/定理编号（其中一些见附录 F）。**

合取正则路径查询。 合取正则路径查询 （CRPQ） Q是在有限字母表 $\mathbb{A}$上的一个表达式 $Q($ x_1 x_n $=$ $A_1 \wedge$ $\wedge A_m$ $m \geq 0$ A_i 、……、$)$ $\cdots$，其中每个都是形如 $x \xrightarrow{L} y$的原子，变量为 x 和 y，而 L 是 $\mathbb{A}$ 上的一个正则表达式。同前面一样，我们用 $vars(Q)$表示 Q的变量集合，并且通常将 $Q(\bar{x})$ 写作 Q，其中 $\bar{x} = (x_1$、……、$x_n)$ 是 Q的自由变量元组（其中的变量不一定互不相同）。如果元组 $\bar{x}$为空，我们称 Q为布尔的。CRPQ类扩展了CQ类以及广为研究的正则路径查询（RPQ）类。事实上，每个CQ都可以看作一个CRPQ，其中正则表达式是来自 $\mathbb{A}$的单个标签。另一方面，一个RPQ对应于形如 $Q(x$、$y) = x \xrightarrow{L} y$的CRPQ。

本文将考虑三个基本类别：包含所有合取查询的类别 CQ、包含所有 CRPQ 的类别 CRPQ，以及 CRPQfin，即使用不含克莱尼星号的正则表达式（表示有限语言）的 CRPQ。注意，后者对应于不含递归的 CRPQ 子类。

2.1 标准、原子单射和查询单射语义

现在我们定义 CRPQ 的标准语义 （即数据库理论文献中的通常语义），并介绍两种新的单射语义。

为便于说明，我们首先给出不含 ε的 CRPQ 的语义，然后展示如何将语义扩展到包含 ε的语言。设 Q 是形如$Q(\bar{z}) = x_1 \xrightarrow{L_1} y_1 \wedge \cdots \wedge x_n \xrightarrow{L_n} y_n$的 CRPQ，并假设没有语言 L_i包含 ε（空字）。给定一个图数据库

G，在标准语义下（简称 st-语义）， $G = (V, E)$ 上 Q 的求值记为 $Q(G)^{st}$，它是满足以下条件的节点元组 $\bar{v}$ 的集合：存在映射 μ: vars(Q) $\to V$，使得 $\mu(\bar{z})$ $= \bar{v}$，并且对于每个 i，在 G 中都存在一条从 $\mu(x_i)$ 到 $\mu(y_i)$ 的路径 π_i，其标签属于 L_i。在原子单射语义下（简称 a-inj 语义）的求值记为 $Q(G)$ a-inj，定义方式类似，但还要求每个 π_i 都是一条简单路径（若 $x_i \neq y_i$）或一个简单环（若 $x_i = y_i$）。最后，在查询单射语义下（简称 q-inj 语义）的求值记为 $Q(G)$q-inj，它与原子单射语义类似（即每个 π_i 都必须是简单的），但还要求 μ 是单射，并且对于每个 $i \neq j$， π_i 和 π_j 不共享任何内部节点。

含有 ε-词的 CRPQ Q 的语义按预期定义：查询 Q 等价于若干不含 ε 的 CRPQ 的并，因此

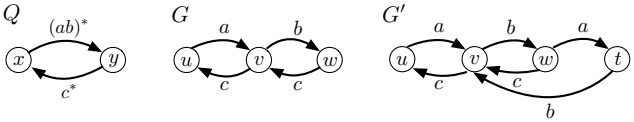

图 2：CRPQ $Q(x$，$y)$ 以及示例 2.1 中的图数据库 G和 G'。

其求值是这些不含 ε 的查询的求值之并。更形式化地，对于 $\star \in \{st$,原子单射, 查询单射$\}$， $Q(\bar{z})$ $= x \xrightarrow{L} y \wedge Q'$ 在 $\star$-语义下的语义，其中 L 含有 ε，且 Q' 是一个 CRPQ，是由$\star$-求值 $Q(\bar{z})$ $= x \xrightarrow{L} y \wedge Q'$ 得到的元组集合与$\star$-求值 $Q(\bar{z}[x/y])$ $= Q'[x/y]$ 得到的元组集合的并，其中 $X[x/y]$ 是将 X 中每个变量 x 的出现都替换为变量 y 后的结果。

注 2.1。 这三种语义构成一个层次结构。特别地，对于任意 *CRPQ* Q G Q G $\subseteq$以及任意图数据库，Q G $\subseteq Q$ G st都有 ()查询单射 ()原子单射 ()。反向包含一般情况下并不成立。

示例 *2.1*。 考虑 CRPQ $Q(x$，$y)$ $= x \xrightarrow{(ab)^*} y \wedge y \xrightarrow{c^*} x$ $)$以及图 中的图数据库 G u w $\in$。注意，Q G u w $\notin Q$ G (，) ()原子单射，但 (，) ()查询单射。另一方面，容易验证 $Q(G)^{st}$ $= Q(G)$原子单射。图 2 中的图数据库 G'可将这三种语义区分开来。事实上，与前面一样，Q G $\not\subseteq Q$ G u v $\in$(′)原子单射 (′)查询单射，此外还有 Q G st u v $\notin Q$ G (，) (′) 和 (，) (′)原子单射。

2.2 刻画求值

我们用基于单射性的限制性同态概念来表述上述定义的语义。这基于 CRPQ 的展开（又称规范数据库）这一关键概念，这将在后续章节的技术推导中发挥作用。

对于 CRPQ Q和 $w \in L$的任意原子 $x \xrightarrow{L} y$， w-展开是一个带等式原子的布尔合取查询，其形式为 (i) $P = x \xrightarrow{a_1} z_1 \wedge z_1 \xrightarrow{a_2} z_2 \wedge \cdots \wedge z_{k-1} \xrightarrow{a_k} y$，若 $w \neq \varepsilon$，其中 z_i 是新引入的变量；或者形式为 (ii) $P = (y = z)$，若$w = \varepsilon$。通常用 $x \xrightarrow{w} y$ 表示这样的 w-展开，其中 $w = a_1 \cdots a_k$。$x \xrightarrow{L} y$的一个展开是某个 $w \in L$对应的 w-展开。CRPQ Q的一个展开配置是任意一个函数

φ，将 Q的每个原子映射到其一个展开。CRPQ $Q(\bar{x}) = A_1 \wedge \cdots \wedge A_m$ 的一个展开是一个 CQ $E(\bar{y})$，其中存在 $\widetilde{\quad}$ $\widetilde{\quad}$，它是 Q的一个展开配置 φ，使得 $E = E^{\equiv}$；这里，E是由 $\widetilde{E}(\bar{x}) = \varphi(A_1) \wedge \cdots \wedge \varphi(A_m)$定义的带等式原子的 CQ。我们用 $\mathrm{Exp}(Q)$表示 CRPQ Q的所有展开构成的集合。直观来说，Q 的一个展开是通过展开 Q 的每个原子，再合并等价变量得到的。例如，图 2 中查询 $Q(x,y) = x(\xrightarrow{ab^*} \wedge y \xrightarrow{c^*} x)$的一种可能展开是 $E_1(x,x) = x \xrightarrow{a} z \wedge z \xrightarrow{b} x$，其展开配置将 x ($\xrightarrow{ab^*} y$)映射到 ab，并将 $y \xrightarrow{c^*} x$ 映射到 ε；另一种展开是 $E_2(x,y) = x \xrightarrow{a} z \wedge z \xrightarrow{b} y \wedge y \xrightarrow{c} x$，它将各原子分别映射到 ab和 c。

标准语义与查询单射语义。标准语义和查询单射语义可改述如下。

命题 2.2。设 Q是一个 *CRPQ*，G是一个图数据库。那么，$Q(G)^{st}$ [（分别地，$Q(G)$查询单射]）是由节点构成的元组 $\bar{v}$ 的集合，其中存在一个 $E \in \mathrm{Exp}(Q)$，使得 $E \to (G,\bar{v})$ [（分别地，E *inj*$\longrightarrow (G,\bar{v})$]）。

原子单射语义。原子单射语义对应于一种限制较少的替代方案：允许任意同态，只要其限制在每个原子的展开上时是单射。设 E（$\bar{y}$）是 CRPQ Q（$\bar{x}$）的一个展开，φ A_1 ∧ ∧ φ A_m = 是产生 E 的展开配置，而 $\widetilde{E}$（$\bar{x}$）（）（）（）是带有等式原子的关联合取查询（特别是 $E = \widetilde{E}^{\equiv}$）。设 Φ: 变量($\widetilde{E}$) → 变量(E) 是规范重命名。若存在某个原子展开 $\varphi(A_i)$，其中包含某个 $x_\prime y_\prime$，使得 $\Phi(x_\prime) = x$ 且 $\Phi(y_\prime) = y$，则称两个变量 x、$y \in$ 变量(E) 之间存在 φ原子相关关系。

现在我们定义一种专为 CRPQ 展开设计的单射同态概念。若存在一个产生 E 的展开配置φ，使得 h 是从 CRPQ Q的展开 E到图数据库 G、将自由变量映射到 $\bar{v}$的原子单射同态，并且 h： $E \to (G,\bar{v})$，且对于任意一对不同的 φ-原子相关变量 x和 y，都有 $h(x) \neq h(y)$，那么我们称 h 为这样的原子单射同态。若存在这样的 h，则记作 E 原子单射$\longrightarrow (G,\bar{v})$。从 E到合取查询的原子单射同态，其定义方式显而易见。

命题 2.3。设 Q是一个 *CRPQ*，G是一个图数据库。那么，$Q(G)$原子单射是由满足以下条件的节点元组 $\bar{v}$构成的集合：存在 $E \in$ 展开(Q)，使得 E 原子单射$\longrightarrow (G,\bar{v})$。

3 求值问题

与求值相关的判定问题是针对 CRPQ 类 C 和语义$\star \in \{st$，原子单射，查询单射$\}$的求值问题 。

问题	C 在$\star$-语义下的求值问题
给定	一个图数据库 G、一个查询 $Q(\bar{x}) \in$C，以及一个元组 $\bar{v}$ 个节点。
问题	$\bar{v} \in Q(G)\star$?

与标准语义下的情况一样，单射语义下的求值问题是 NP 完全的。

命题 3.1。 *CRPQ* 和合取查询的求值问题在组合复杂度下，对于所有语义都是NP 完全的。

证明。下界可通过从单射同态测试问题（也称为子图同构问题）进行一个简单归约得到；这是一个众所周知的 NP 完全问题。[11, 18]。事实上，布尔合取查询 Q 单射映射到 G 当且仅当 $Q(G)$查询单射$\neq \emptyset$ 当且仅当 $Q^+(G^+)$原子单射$\neq \emptyset$，其中 G^+ [（分别为 Q^+]）是这样得到的：对于一个新符号 R，在每一对顶点 [之间添加一条 R-边（分别在每一对变量] 之间添加一个 R-原子。

上界是多项式规模见证性质的推论。也就是说，若 $Q \in$ CRPQ，且 $\bar{v} \in Q(G)\star$，则存在 Q 的一个展开 E，使得当 $\star$ = 查询单射 时，E *inj*$\longrightarrow (G,\bar{v})$；当 $\star$ = 原子单射 时，E *a-inj* $\longrightarrow (G,\bar{v})$。两种情况下，$E$ 都关于 G 和 Q 呈线性。随后可以猜测这样一个展开，并检查相应同态是否存在。□

替代语义下的数据复杂度（即，当查询被视为常数规模时）同样是 NP 完全的，因为即使对于非常简单的正则表达式 [26]，简单路径语义下 RPQ 的求值也是 NP 完全的：

命题 3.2. 在原子单射和查询单射语义下，*CRPQ* 的求值问题在数据复杂度方面是NP完全的。

通过一个三分结果，人们刻画了哪些 RPQ 能够在数据复杂度方面得到高效求值：它们的复杂性可能是 NP 完全、NL 完全，或属于 AC^0 [3,Theorem 2]。将这一结果推广到单射语义下的 CRPQ 似乎相当棘手；特别是，这需要全面理解查询等价问题，而这正是下一节的重点。

4 包含问题

如果对于每个图数据库 G，都有 $Q_1(G)\star \subseteq Q_2(G)\star$，则称 CRPQ Q_1 在包含 于 CRPQ Q_2（在$\star$语义下），记作 $Q_1 \subseteq_\star Q_2$。我们定义包含问题，其参数包括 CRPQ 类 C_1 和 C_2，以及所采用的语义（标准语义、查询单射语义或原子单射语义）。

问题	$\mathrm{C}_1/\mathrm{C}_2$ 在$\star$-语义下的包含问题
给定	CRPQ $Q_1 \in \mathrm{C}_1$ 和 $Q_2 \in \mathrm{C}_2$。
问题	是否有$Q_1 \subseteq_\star Q_2$ 成立?

在标准语义下，CQ、CRPQ 和 $\mathrm{CRPQ}_{\mathrm{fin}}$ 之间的所有组合都已得到研究，且均为可判定的。具体而言：

定理 4.1. *[9, 16]* 标准语义下的 *CRPQ/CRPQ* 包含问题是指数空间完全的。

本文余下部分将专门研究单射语义下的情形。我们将证明，一种单射语义变得不可判定，而另一种在标准复杂性理论假设下的计算表现则更好（参见 图 1 中 CRPQ/CRPQ 一列）。

4.1 刻画包含

对于标准语义，众所周知，CRPQ 的包含可以用展开来刻画：

命题 4.2。 *[9]* 设 Q_1 和 Q_2 是 *CRPQ*。那么 $Q_1 \subseteq_{st} Q_2$ 当且仅当对于每个 $E_1 \in$ 指数(Q_1) 都存在 $E_2 \in$ 指数(Q_2) 使得 $E_2 \to E_1$。

查询单射语义也有类似的刻画：

命题 4.3。设 Q_1 和 Q_2 是 *CRPQ*。那么$Q_1 \subseteq$查询单射 Q_2 当且仅当对于每个 $E_1 \in$ 指数(Q_1) 都存在 $E_2 \in$ 指数(Q_2) 使得 E_2 单射$\longrightarrow E_1$。

事实证明，先前的刻画不适用于原子单射语义（将 $inj \longrightarrow$ 替换为 $a\text{-}inj \longrightarrow$）。在这种情况下，$Q_1$的展开空间并不足够，我们需要在下文定义的一个更大的 Q_1 展开空间中检查 Q_2。

CRPQ Q（$\bar{x}$）的一个原子单射展开 （简称*a-inj*-展开）是一个 CQ F（$\bar{y}$），且存在一个带等式原子的 CQ$\widetilde{F}$（$\bar{z}$）$= E$（$\bar{z}$）$\wedge J$，满足：(i) $F = \widetilde{F}^{\equiv}$；(ii) E（$\bar{z}$）是由某个展开配置 φ 生成的 Q的展开；(iii) J 是由等式原子 $x\prime = y\prime$构成的合取，其中变量 $x\prime, y\prime \in$属于 E 的变量集，且对于 E中每一对不同的 φ-原子相关变量 x, y，都有$x \neq_{\widetilde{F}} y$。我们用 Exp$_{a\text{-}inj}$（Q）表示Q所有*a-inj*-展开的集合。直观而言，这类展开是通过合并 Q的普通展开中某些不具原子相关性的变量对得到的（这些合并由 J 给出）。我们将用到以下一个有用的结果：

引理 4.4. 设 Q 为一个 *CRPQ*，$E\prime$ 为一个合取查询，G 为一个图数据库，且 $\bar{v}$ 为一个节点元组。以下条件等价：

(1) There is $E \in \text{Exp}(Q)$ *s.t.* $E \xrightarrow{a\text{-}inj} (G, \bar{v})$ *[resp.* $E \xrightarrow{a\text{-}inj} E'$*].*

(2) There is $F \in \text{Exp}^{a\text{-}inj}(Q)$ *s.t.* $F \xrightarrow{inj} (G, \bar{v})$ *[resp.* $F \xrightarrow{inj} E'$*].*

作为引理 4.4 的推论，我们得到原子单射语义的另一种定义：

推论 4.5. 设 Q 为一个 *CRPQ*，G 为一个图数据库，且 $\bar{v}$ 为一个节点元组。那么 $\bar{v} \in Q(G)$是原子单射的，当且仅当存在$F \in$ Exp是原子单射的(Q) 使得 F 是单射的$\longrightarrow (G, \bar{v})$。

现在我们给出原子单射包含的刻画：

命题 4.6. 对于任意一对 Q_1, Q_2 *CRPQ*，以下条件等价：

(1) $Q_1 \subseteq$是原子单射的 Q_2。*(2)* 对每个 $F_1 \in$ Exp是原子单射的(Q_1)，都存在 $E_2 \in \text{Exp}(Q_2)$ 使得 E_2 是原子单射的$\longrightarrow F_1$。*(3)* 对每个 $F_1 \in$ Exp是原子单射的(Q_1)，都存在 $F_2 \in$ Exp是原子单射的(Q_2) 使得 F_2 是单射的$\longrightarrow F_1$。

根据上述刻画，对于任意一对 CRPQ Q_1, Q_2，我们有$Q_1 \subseteq$查询单射 Q_2 蕴含 $Q_1 \subseteq_{st} Q_2$，且 $Q_1 \subseteq$原子单射 Q_2蕴含 $Q_1 \subseteq_{st} Q_2$（尽管反向蕴含一般情况下并不成立）。相比之下，尽管这些语义之间存在层次结构，但查询单射包含与原子单射包含之间并不存在这样的蕴含关系，以下示例将对此进行说明。

例4.7。考虑布尔 CRPQ $Q_1 = x \xrightarrow{a} y \wedge y \xrightarrow{b} z$、$Q_2 = x \xrightarrow{ab} y$、$Q'_1 = x \xrightarrow{a} y \wedge x \xrightarrow{b} y$ 和 $Q'_2 = x \xrightarrow{a} y \wedge x' \xrightarrow{b} y'$。我们有 $Q'_1 \subseteq$ 原子单射 $Q'2$（以及 $Q'_1 \subseteq_{st} Q'2$），但 $Q'_1 \not\subseteq$查询单射 $Q'2$，因为不存在从 $Q'2$ 的唯一展开到 $Q'1$ 的唯一展开的单射同态。另一方面，我们有 $Q_1 \subseteq$查询单射 Q_2（以及 $Q_1 \subseteq_{st} Q_2$），但 $Q_1 \not\subseteq$原子单射 Q_2。实际上，我们可以取 Q_1 的 a-inj-扩展 F，它由 $x \xrightarrow{a} y \wedge y \xrightarrow{b} z$ 得到通过识别 x 和 z。于是，不存在从 Q_2 的唯一展开到 F 的原子单射同态。

鉴于上述刻画，在下文中，我们有时会使用 st-展开或 q-inj-展开来表示一个（普通）展开。对于 ★ $\in$ {*st*, q-inj, a-inj}，如果 E_1 是 Q_1 的一个 ★-展开，且满足 $\bar{y} \notin Q_2(E_1)$★，那么我们称 $E_1(\bar{y})$ 是针对 ★- 语义的一个反例（回顾一下，任何 CQ，尤其是 E_1，都可以被视为一个图数据库）。注意，后一条件 $\bar{y} \notin Q_2(E_1)$★ 等价于不存在 Q_2 的某个（普通）展开 E_2，使得以下条件之一成立：(a) 如果 ★ $= st$，则 $E_2 \to E_1$；(b) 如果 ★ = q-inj，则 E_2 inj$\longrightarrow E_1$；或 (c) 如果 ★ = a-inj，则 E_2 a-inj$\longrightarrow E_1$。因此，当且仅当存在一个针对 ★-语义的反例时，$Q_1 \not\subseteq$★ Q_2。

5 无限制CRPQ的包含

本节研究查询单射语义和原子单射语义下的 CRPQ/CRPQ 包含问题。我们证明，前者属于多项式空间，而后者不可判定。这两项证明都并非显而易见，并为如何利用语义解决静态分析问题提供了新的见解：在第一种情况下，通过将所需空间从指数级降至多项式级；在第二种情况下，则通过精巧的编码，强制要求反例见证 PCP问题的解。

标准语义下的 CRPQ/CRPQ 包含问题属于指数空间 [16]，而我们所知的所有证明 [9, 14, 16] 都将该问题归约为检验规模呈指数级的非确定有限自动机上的包含性或全性，这些自动机对展开集合或（非）反例集合进行编码。然而，查询单射语义下的多项式空间界在本质上截然不同，它利用数量呈指数级、规模为多项式的证书来确保包含成立。

定理 5.1。在查询单射语义下，CRPQ/CRPQ 包含问题属于多项式空间。

证明概要（完整证明见附录 C）。设 Q_1，Q_2 为 CRPQ；我们要检验 $Q_1 \subseteq$查询单射 Q_2。由于篇幅有限，我们仅给出证明的高层次描述。我们将使用 Q_1 的展开的多项式大小“抽象”。对于 A Q_1 的每个原子，这些抽象都包含有关如何将 Q_2 的语言映射到该原子中的所有信息。例如，其中包括如下信息：“存在一条从状态 q 到状态 $q\prime$ 的非确定有限自动机 A_L，其语言为 L，从 Q_2 读取 A 的展开词时的部分运行”，或者“存在一条从 A_L 的初始状态到 q，读取 A 的展开词某个后缀时的部分运行”。这样的抽象包含了从某个展开中保留必要信息的全部内容，以便检验它是否为反例。事实上，任何与某个反例具有相同抽象的展开也都是反例。

为了测试一个抽象 α 是否抽象了某个反例，我们需要考虑将 Q_2 的一个展开单射映射到 Q_1的一个展开的所有可能方式。我们将其称为态射类型，它包含了 Q_2 的每个原子展开被映射到何处的信息。例如，我们可以获得这样的信息：Q_2的原子 A 的展开所对应的路径，从 Q_1 的原子$A_1 = x \xrightarrow{L_1} y$ 展开的某个内部节点出发，然后到达状态为 q 的变量 y，并且

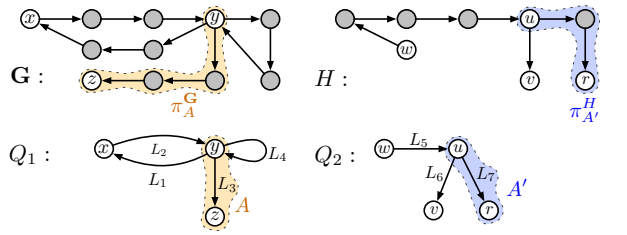

图3： G的定义示例，以及从 Q_1, Q_2到 Q_1, Q_2的态射类型（H, h）。在此例中，从 H到G的单射态射 h将 H的每个节点映射到G上相同位置的节点（例如， H的右下角节点 r映射到G的右下角节点）。

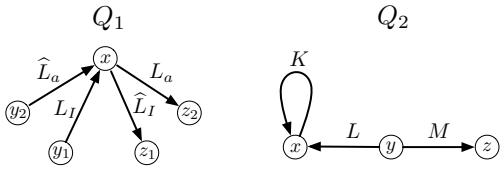

图4： 归约中布尔 CRPQ Q_1 和 Q_2 的一般结构。

继续读取 $A_2 = y \xrightarrow{L_2} z$ 的完整原子展开，到达状态为 q_r的变量 z，然后读取 $A_3 = z \xrightarrow{L_3} t$ 的原子展开的一个前缀，最终在某个内部节点到达终态，从而结束其行程。对于每种态射类型，我们都可以通过检查它是否兼容某个抽象来进行验证，例如检查 α是否确实包含了这样的信息：存在一个从 q_r 到终态的部分运行，该运行读取 A_3展开的一个前缀。

更具体地说，考虑有向图G，它由将 Q_1 的每个原子 $A = x \xrightarrow{L} y$ 替换为长度为3的路径G$_A$ 构成（即，添加两个新的内部顶点）。从 Q_2 到Q_1 的态射类型 是一个对（H, h），满足 h： H 单射⟶ G，且 H是通过将 Q_2 的每个原子 $A = x \xrightarrow{L} y$ 替换为从 x 到 y的路径 π_A^H而得到的图。图3给出了一个态射类型的示例。根据单射性，任何态射类型中 H的大小都以线性方式受 Q_1限制。

态射类型（H, h）与抽象 α 相兼容，当且仅当存在一个映射 λ，将路径 π_A^H 的内部节点映射到 A 的状态，使得对于 Q_1 的每个原子 $A = x_1 \xrightarrow{L} x_2$，所有预期属性都成立。例如，若存在 Q_2 的一个原子 A'，以及路径 $\pi_{A'}^H$ 的一个中缀 π，满足 $h(\pi) = \pi$ G$_A$，则抽象α保证存在一条从 $\lambda(src(\pi))$ 到 $\lambda(tgt(\pi))$ 的 A$_{A'}$运行，其中 A$_{A'}$是 A' 的语言对应的非确定有限自动机，而 $src(\pi)$ 和 $tgt(\pi)$ 分别是 π 的首节点和末节点。再举一例，还必须检查：若存在 Q_2 的一个原子 A'，以及 $\pi_{A'}^H$ 的一个后缀π，满足 $h(\pi)$ 是 π G$_A$ 的前缀，则 α保证在 A 的展开的前缀上，存在一条从 $\lambda(src(\pi))$ 出发并到达 A$_{A'}$ 中某个终态的运行。实际上还有许多其他可能的情况（总共 17 种），但所有这些情况都可以利用抽象中汇总的信息轻松检查。

兼容性的关键性质在于，它能够刻画一个抽象是否包含一个作为反例的 Q_1-展开：

断言5.1。 以下命题等价：

(1) 存在一种与抽象 α兼容的态射类型；*(2)* 对于每个具有抽象 α 的展开 $E_1 \in \text{Exp}(Q_1)$，都存在某个展开 $E_2 \in \text{Exp}(Q_2)$ 使得 $E_2 \xrightarrow{inj} E_1$；*(3)* 存在一个具有抽象 α 的展开 $E_1 \in \text{Exp}(Q_1)$，以及一个展开 $E_2 \in \text{Exp}(Q_2)$，使得 $E_2 \xrightarrow{inj} E_1$。

最后，PSPACE 算法猜测一个映射 α，将 Q_1 的原子映射到 P 的子集，检查 α 是否是 Q_1的抽象，并检查是否不存在与 α 兼容的态射类型 (H, h)。根据上述断言，如果算法成功，那么 Q_1 的任意展开 E_1 都是反例，因此$Q_1 \not\subseteq_{\text{查询单射}} Q_2$；否则，对于每个具有抽象 α 的展开 E_1，都存在一种兼容的态射类型，这意味着 E_1 不是反例，因此$Q_1 \subseteq_{\text{查询单射}} Q_2$。□

另一方面，原子单射语义下的 CRPQ/CRPQ 包含问题变得不可判定。值得注意的是，即使右侧查询不含无限语言，且两个查询的形式都非常简单（参见 图4），这一界限仍然成立。

定理 5.2。 原子单射语义下的 CRPQ/CRPQ 和 CRPQ/CRPQ$_{\text{fin}}$ 包含问题不可判定。

证明概要（完整证明见附录 D）。我们从 PCP（Post对应问题）归约而来；这是一类众所周知的不可判定问题。PCP 的一个实例是一个对序列 $(u_1, v_1)\ldots,(u_\ell, v_\ell)$，其中 u_i和 v_i是字母表 Σ上的非空词。目标是判定是否存在一个解，即一个由 $\{1,\ldots,\ell\}$中的索引组成的序列 $i_1,\ldots,i_k$，满足 $k \geq 1$，使得词 $u_{i_1}\cdots u_{i_k}$和 $v_{i_1}\cdots v_{i_k}$相同。

我们给出该归约的高层次描述。其思路是构造布尔合取正则路径查询 Q_1和 Q_2，使得 PCP 实例 $(u_1, v_1)\ldots,(u_\ell, v_\ell)$ 有解，当且仅当 $Q_1 \not\subseteq_{\text{原子单射}} Q_2$。特别地，PCP 实例有解，当且仅当存在一个针对原子单射语义的反例，即，一个Q_1的原子单射展开 F，使得不存在满足 $E \xrightarrow{\text{原子单射}} F$ 的展开 $\text{Exp}(Q_2)$ $E \in$。 Q_1的一般结构如图4所示。我们有一个“中间”变量 x、两个“输入”原子和两个“输出”原子：

$$y_1 \xrightarrow{L_I} x \wedge y_2 \xrightarrow{\widehat{L}_a} x \wedge x \xrightarrow{\widehat{L}_I} z_1 \wedge x \xrightarrow{L_a} z_2$$

语言 L_I和 L_I中的词编码来自 $\{1, \ldots\ldots, \ell\}$ 的索引序列，并分别使用来自 $\mathbb{I} := \{I_1, \ldots\ldots, I_\ell\}$ 和$\widehat{\mathbb{I}} := \{\widehat{I_1}, \ldots\ldots, \widehat{I_\ell}\}$ 的特殊符号。另一方面，来自 L_a和$\widehat{L_a}$的词编码来自 $\{u_1, \ldots\ldots, u_\ell\}$ 和 $\{v_1, \ldots\ldots, v_\ell\}$ 的词序列，并分别使用 PCP 字母表 Σ 以及 $\widehat{\Sigma} := \{\widehat{a} : a \in \Sigma\}$ 中的符号。在这四种语言中，我们加入了一些额外符号以使归约成立。需要强调的是，CRPQ Q_1（以及 Q_2）所用的有限字母表取决于 PCP 实例。

我们关注 Q_1 的一种特殊类型的原子单射-展开，称为良构展开。其思路是，良构的原子单射-展开对应 PCP 实例的解。具体而言，如果存在 Q_1 的良构原子单射-展开，那么就存在一个

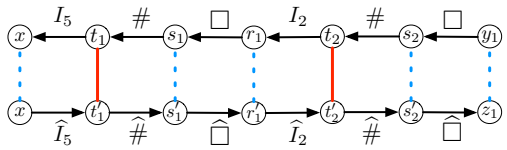

图 5： I-$\widehat{I}$-良构展开条件示例。针对 Q_1。我们展示原子 $y_1 \xrightarrow{L_I} x$ 和 $x \xrightarrow{L_I} z_1$ 的展开 $y_1 \xrightarrow{w_I} x$ 和 $x \xrightarrow{w_I} z_1$。词 $w_I = \square\#I_2\square\#I_5$ 和 $\widehat{w}_I = \widehat{I}_5\widehat{\#}\widehat{\square}\widehat{I}_2\widehat{\#}\widehat{\square}$ 编码了指标序列 5、2。蓝色虚线表示相等变量对，红线表示不同变量。我们还包含了一些额外符号 $\#\,\widehat{\#}\,\square,\ \widehat{\square}$。

PCP 实例的解，反之亦然。接着，我们构造 Q_2，使得 Q_1 的原子单射-展开是良构的，当且仅当它是 $Q_1 \subseteq_{\text{原子单射}} Q_2$ 的反例。

令 F 为 Q_1 的一个 $E \wedge J$ a-inj- 扩展，使得对于 $E \in$（这里 Exp( ) 和 是等式原子）。若 a-inj- 扩展 F 满足我们称为 I-$\widehat{I}$-、I-a-、$\widehat{a}$-$\widehat{I}$-和 $\widehat{a}$-a-条件的四个结构条件，则称其为良构。直观而言，

I-I-条件要求在展开 E 中选取的词 $w_I \in L_I$ 和 $I \in L_I$ 编码相同的指标序列，且该序列来自 $\{1,\ldots,\ell\}$。另一方面，I-a-条件确保在展开 E 中选取的词 $w_a \in L_a$，按照 $w_I \in L_I$ 所编码的序列，编码一个由 $\{u_1,\ldots,u_\ell\}$ 中的词构成的序列。同理，$\widehat{a}$-$\widehat{I}$-条件要求所选词 $\widehat{w}_a \in \widehat{L}_a$ 按照 $\widehat{w}_I \in \widehat{L}_I$ 编码一个由 $\{v_1,\ldots,v_\ell\}$ 中的词构成的序列。最后，$\widehat{a}$-a-条件要求所选词 $w_a \in L_a$ 和 $\widehat{w}_a \in \widehat{L}_a$ 在去除 $\widehat{w}_a$ 中的 $\widehat{\cdot}$ 上标并仅关注来自 Σ 的符号后“相同”。换言之，$\widehat{a}$-a-条件确保由 w_I（以及 $\widehat{w}_I$）选取的指标序列实际上是 PCP 的一个解。在这四种情况下，我们还需要对等式原子 J 施加一些条件，才能使归约成立（示例见图 5）。

良构性的关键性质是，可以通过不存在具有特定标签的有限多个简单环和简单路径来刻画。下面以图 5 中的 I-$\widehat{I}$-条件和展开 F 为例说明这一点。简单环的禁用标签由有限语言 $K_{I\widehat{I}} = \mathbb{I}\widehat{\mathbb{I}}$ 给出。对于简单路径，其禁用标签则由 $M_{I\widehat{I}} = \sum_{i\neq j} I_i\widehat{I}_j + \widehat{\mathbb{I}}\#\ +\widehat{\#}\,\mathbb{I} + \#\,\mathbb{I}\widehat{\mathbb{I}}\widehat{\#}\ + \square\widehat{\square}$ 给出。我们有，一个原子单射-$\widehat{\ }$展开 F 满足 I-I-条件，当且仅当 F 不包含标签属于 $K_{I\widehat{I}}$ 的简单环，也不包含标签属于 $M_{I\widehat{I}}$ 的简单路径。

要理解反向情况，注意到 t_1 和 $t_{\prime 1}$ 不能被合并，否则我们就会得到一个从 t_1 到其自身、标签属于 $\mathbb{I}\widehat{\mathbb{I}} \subseteq K_{I\widehat{I}}$ 的简单环。现在，符号 I_5 和 $\widehat{I}_5$ 必须对应于 $\{1,\ldots,\ell\}$ 中的同一个索引；否则，我们就能找到一条从 t_1 到 $t_{\prime 1}$、标签属于 $\sum_{i\neq j} I_i\widehat{I}_j \subseteq M_{I\widehat{I}}$ 的简单路径。为了理解 s_1 与 $s_{\prime 1}$ 之间的合并，首先注意到 t_1 和 $s_{\prime 1}$ 不能被合并，因为这将意味着存在一条从 $t_{\prime 1}$ 到 x、标签属于 $\widehat{\ }\#\,\mathbb{I} \subseteq M_{I\widehat{I}}$ 的简单路径。类似地，我们有 $t_{\prime 1}$ 和 s_1 不能被合并。这意味着 s_1 和 $s_{\prime 1}$ 实际上被合并了，否则我们就会得到一个从 s_1 到 $s_{\prime 1}$ 的简单路径，其标签属于 $\#\,\widehat{\mathbb{I}\mathbb{I}}\widehat{\ }\#\ \subseteq M_{I\widehat{I}}$。最后，$r_1$ 和 $r_{\prime 1}$ 被视为同一节点，否则就会存在一条从 r_1 到 $r_{\prime 1}$ 的简单路径，其标签属于 $\square\widehat{\square} \subseteq M_{I\widehat{I}}$。注意，我们可以从 $r_i = r_{\prime i}$ 出发，而不是从 x 出发，重复这一论证，但方向改为“从左到右”，从而得到 I-$\widehat{I}$-条件。为了确保词 w_I 和 $\widetilde{w}_I$ 具有相同的长度，我们需要稍微修改 Q_1、$K_{I\widehat{I}}$ 和 $M_{I\widehat{I}}$ 的构造（详情见附录 D）。正向结论直接由 I-$\widehat{I}$-条件的定义得出。

由于这四个条件都可以通过简单环和路径的有限禁用集合来刻画，因此可以写出两个 CRPQ有限，形式为 $Q_2^{\circlearrowleft} = x \xrightarrow{K^{\circlearrowleft}} x$ 和 $Q_2^{\rightarrow} = y \xrightarrow{M^{\rightarrow}} z$，使得对于 Q_1 的每个原子单射展开 F，F 是良构的，当且仅当 $Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}(F)$原子单射 $= \emptyset$，其中 $Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}$ 是这两个 CRPQ 的并。特别地，PCP 实例有解，当且仅当 $Q_1 \not\subseteq_{\text{原子单射}} Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}$。最后，我们说明如何用单个查询 $Q_2 \in$ CRPQ有限 模拟并 $Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}$，如图4所示。□

6 CRPQ子类的包含

在前述两个关于无约束 CRPQ 包含的结果基础上，我们现在探讨所有可能语义下 C1/C2 的包含问题，其中 C1 和/或 C2 属于较简单的查询类别，即合取查询，或不含 Kleene星号的 CPRQ（因此其语言有限）。

在许多情况下，可以应用或改编已有的技术或归约。不过，有两个显著的例外：查询单射语义下 CRPQfin/CQ 的下界，以及原子单射语义下的 CQ/CRPQfin 下界。我们仅重点介绍这两个结果。其余证明见附录 F。特别是，如第 4 节所述，标准语义下几乎所有结果都可由先前的工作推出（尤其是 [9, 15, 16]）。

定理 6.1. 查询单射语义下 CRPQfin/CQ 包含问题是 Π_2^p-困难的。

证明。我们证明，即使左侧查询的语言是字母表符号的并，Π_2^p 困难性仍然成立。我们将以下图上的问题归约到非包含问题；已知该问题是 Σ_2^p-完全的 [28, 定理 5],。对于图 G，令 $V(G)$ 表示其顶点集合，并令 $G|_{V'}$ 表示由 $V' \subseteq V(G)$ 诱导的子图。

问题	广义二染色问题（GCP2）
给定问题	一个无向图 G，一个数值 $n \in \mathbb{N}$（一元表示）。是否存在划分 $V_1 \cup V_2 = V(G)$，使得两者均不满足 $G\|_{V_1}$ nor $G\|_{V_2}$ 是否包含一个作为子图的 n-顶点团?

我们将构造两个布尔查询 Q_1，Q_2，其字母表为 $\mathbb{A} = \{E, 1, 2, \#\}$，使得：(1) $Q_1 \not\subseteq_{\text{查询单射}} Q_2$ 当且仅当 GCP2 实例为正；(2) Q_2 是一个合取查询，且 Q_1 的每种语言都是单字母词的集合。考虑输入图 G，以及其在字母表 $\{E\}$ 上的关联合取查询 Q_G；对于 G 中的每条边 $\{u, v\}$，我们在 Q_G 中都有原子 $u \xrightarrow{E} v \wedge v \xrightarrow{E} u$。类似地，令 K_n 为与 n 顶点团关联的合取查询。对于合取查询 Q 和 $i \in \{1, 2\}$，令我们

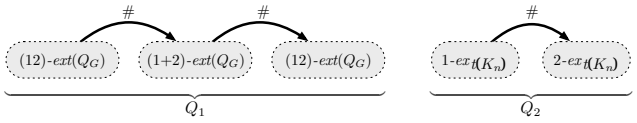

图 6： 定理 6.1 的归约中，Q_1 和 Q_2 由 G 和 n 定义。带 # 标签的粗箭头表示，对于从源查询到目标查询的每个变量 x 以及目标查询中的每个变量 y，都存在一个原子 $x \xrightarrow{\#} y$。

定义 i-扩展(Q) [以及 $(1+2)$-扩展(Q)；(12)-扩展(Q)] 是指通过添加一个原子 $x \xrightarrow{i} x$ [或一个原子 $x \xrightarrow{1+2} x$ 来扩展 Q；对于变量 $x \in \mathrm{vars}(Q)$中的每个变量，添加两个原子 $x \xrightarrow{1} x \wedge x \xrightarrow{2} x$]。我们按图 6 中的方式定义 Q_1, Q_2。一方面，如果 $Q_1 \not\subseteq$查询单射 Q_2，那么必定存在 Q_1 的某个展开 E，它是一个反例。由 E 可推出 $V(G)$的划分 $V_1 \dot\cup V_2$，其中 V_i 是中间构件中带有 i-自环标签的顶点集合。现在注意，对于每个 i, K_n 都不能单射映射到$G|_{V_i}$；否则，我们便有 i-扩展(K_n) inj $\longrightarrow E$，从而推出$Q_2 \xrightarrow{\text{inj}} E$，这与 E 是反例相矛盾。这意味着 GCP2 实例为正。另一方面，如果存在 $V(G)$的一个划分 $V_1 \dot\cup V_2$，使其不含作为子图的 n-团，那么相应的 Q_1 的展开 E（为每个节点 $x \in V_i$选择一个i-自环）就无法将 Q_2 单射映射到 E；换句话说，这表明 E 是一个反例，因此$Q_1 \not\subseteq$查询单射 Q_2。□

定理 6.2。在原子单射语义下，合取查询/CRPQfin 包含问题是 Π_2^p-困难的。

证明概要（完整证明见附录 E）。我们证明，即使右侧的所有语言都具有 $\{w\}$ 的形式，且$|w| \le 2$，包含问题仍然是 Π_2^p-困难的。为此，我们展示如何改写定理 4.3], 中 [15, 的 Π_2^p-困难性证明；该定理证明了在标准语义下 CRPQfin/CQ 包含问题是 Π_2^p-困难的。我们从 ∀∃-QBF（即，Π_2-量化布尔公式）归约。设 $\Phi = \forall x_1, \ldots, x_n \, \exists y_1, \ldots, y_\ell \; \varphi$（$x_1, \ldots, x_n, y_1, \ldots, y_\ell$）是一个 ∀∃-QBF 实例，其中 φ是无量词且为 3-CNF。我们构造布尔查询 Q_1 和 Q_2，使得$Q_1 \subseteq_{\text{原子单射}} Q_2$ 当且仅当 Φ 可满足。

查询 Q_1 定义于图 7 中，其所基于的标签字母表为 $\{a, x_1, \ldots, x_n, y_1, \ldots, y_\ell, t, f, r\}$。现在我们解释如何在相同的字母表上定义 Q_2。Φ 的每个子句都由 Q_2 中的一个子查询表示，如图 7 所示。图中所有具有相同标签的节点（构件 D, E 中的 $y_{1,t}$ 和 $y_{1,f}$）实际上是同一个节点。注意，对于每个子句及其中的每个存在量化文字 y_i，我们在 Q_2 中都有一个名为 $y_{i,tf}$ 的节点。 E-构件的设计使得每个表示的文字都可以通过同态嵌入，同时必须恰好嵌入一个文字。

D构件。

直观的想法是， x-变量的赋值由 Q_1的原子单射展开 E_1 给出，无论 x下与 t 相邻的两个节点是否相等：若它们相等，则对应于假 赋值；否则对应于真 赋值。另一方面，

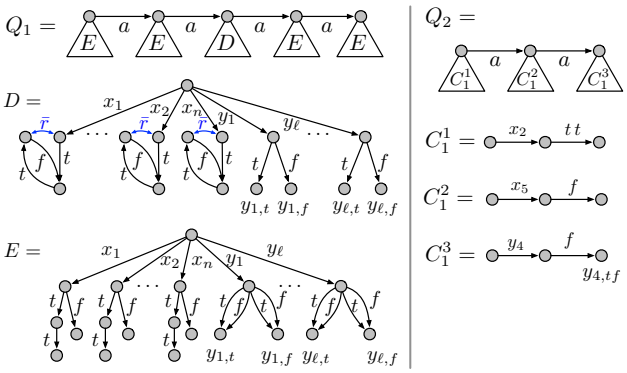

图 7：左： 在定理 6.2 中使用的查询 Q_1 以及构件 D，还有其定义中使用的构件 E。为清晰起见， $\bar r$蓝色边表示 r的补关系中的边（即不属于关系 r的边）。**右：** $\varphi = (x_2 \vee \neg x_5 \vee \neg y_4)$的 Q_2 示例。

y变量的赋值由从 Q_2的某个展开到 E_1 的同态给出（即，相应节点是否映射到节点$y_{\lrcorner}$、t或$y_{\lrcorner}$、f）。跨多个子句的 y变量的同态必须保持一致，因为所有子句共享相同的节点$y_{\lrcorner}$、tf，而这些节点会被唯一地映射到$y_{\lrcorner}$、t或$y_{\lrcorner}$、f。因此，当公式 Φ可满足时，对于变量$\{x_i\}$的任意赋值（由 D中选择 t/f边给出），都存在从$y_{\lrcorner}$、tf到$y_{\lrcorner}$、f或$y_{\lrcorner}$、t之一的映射。这给出了$Q_1 \subseteq$原子单射 Q_2。反过来，如果 Q_2 的某个展开可以映射到 K，那么对于 D中 t/f边的一种选择，我们就能将 Q_2 的每个子句构件嵌入到 K中。特别地，我们总能将 Q_2 每个子句中的一个文字映射到 D，从而确保 φ得到满足。由于对于通过对$\{x_i\}$进行任意 t/f赋值而得到的任何展开 K，这一点都成立，因此我们可以得出 Φ 可满足。□

其他结果。其余结果总结于图 1 和以下定理中，相关证明见附录 F。

定理 6.3。

(1) 在查询单射语义下，*CQ/CRPQ* 和 *CQ/CQ* 包含问题是 *NP* 完全的。（见附录 *F* 中的命题 *F.2*。）*(2)* 在原子单射语义下，*CQ/CQ* 包含问题是 *NP* 完全的。（见附录 *F* 中的推论 *F.4*。）*(3)* 在标准语义和原子单射语义下，*CRPQ/CQ* 和 *CRPQfin/CQ* 包含问题是Π_2^p-困难的。（见附录 *F* 中的命题 *F.6*。）*(4)* 在所有语义下，*CRPQ/CQ* 和 *CRPQfin/CQ* 包含问题均属于 Π_2^p。（见附录 *F* 中的命题 *F.7*。）*(5)* 在所有语义下，*CRPQ/CRPQfin* 包含问题是 *PSPACE*-困难的。（见附录 *F* 中的命题 *F.8*。）*(6)* 在标准语义下，*CRPQ/CRPQfin* 包含问题属于多项式空间。（见附录 *F* 中的命题 *F.9*。）*(7)* 在所有语义下，*CRPQfin/CRPQ* 包含问题属于 Π_2^p。（见附录 *F* 中的命题 *F.10*。）

7 讨论与展望

我们为 CRPQ 定义了两种可能的单射语义，为将 RPQ 的简单路径语义扩展到 CRPQ 领域提供了两种方式。在这些语义下，我们证明了包含问题与标准语义下的情况截然不同：某些情况下复杂性有所降低，而另一些情况下问题则直接变得不可判定。

这两种语义都是 RPQ 简单路径语义的自然推广。例如，如果互换边和节点的角色，那么原子单射语义下的 CRPQ 已经应用于流行的图数据库 Neo4j 中。虽然查询单射语义在实践中并不常见，但我们仍然认为，这种语义，尤其是寻找不相交路径，可能对用户有所帮助，也可能成为图查询语言的一项有趣功能。要评估这两种语义在实践中的实用性，还需要进一步的实证研究。

尽管我们研究的 CRPQ 类中的这些片段可能是最基础的三个子类，但还有一些更细致的片段，它们基于 CRPQ 中所用正则表达式的形式划分，在实践中也非常重要 [7, 8]。这些片段已在标准语义下得到研究 [15],，探究它们在单射语义下的表现将很有意义。另一个研究方向是考虑更大的查询类，了解单射语义如何扩展，以及这对包含界限有何影响——例如，具有双向导航和并的 CRPQ（UC2RPQ） [9],、扩展 CRPQ（ECRPQ） [6], 或正则查询 [27]。

我们的研究仅限于简单路径语义（及其扩展）。另一种可能是考虑迹语义，并以类似的方式将其扩展到 CRPQ，从而再次得到两种替代语义：基于边单射同态概念的查询边单射和原子边单射。我们的许多结果都可以扩展到这些语义；我们怀疑，在所有已研究的片段中，查询边单射与查询单射的复杂度相同，原子边单射与原子单射的复杂度也相同。具体而言，虽然简单地交换节点与边的角色似乎无法得到不可判定性结论，也无法得到多项式空间上界，但我们相信这两种证明都可以加以改造。

一个可能的研究方向是考察 CRPQ 的另一个基本静态分析问题，即有界性问题；该问题用于判断一个 CRPQ 是否等价于有限个合取查询的并。对于标准语义，该问题是可判定的 [5]。

参考文献

[1] Renzo Angles、Marcelo Arenas、Pablo Barceló、Peter A. Boncz、George H. L. Fletcher、Claudio Gutiérrez、Tobias Lindaaker、Marcus Paradies、Stefan Plantikow、Juan F. Sequeda、Oskar van Rest 和 Hannes Voigt。2018年。G-CORE：面向未来图查询语言的核心。载于*2018*年数据管理国际会议论文集，*2018* 年 *SIGMOD* 会议，*15*，*2018*，*10* 得克萨斯州休斯敦，美国，*6*月 -Gautam Das、Christopher M. Jermaine 和 Philip A. Bernstein（编）。ACM，1421–1432。https://doi.org/10.1145/3183713.3190654[2] Renzo Angles、Marcelo Arenas、Pablo Barceló、Aidan Hogan、Juan Reutter 和 Domagoj Vrgoč。2017年。图数据库现代查询语言基础。*ACM* 计算调查 50，5，文章 68（2017年9月），40页。https: //doi.org/10.1145 3104031 。 / Guillaume Bagan、Angela Bonifati 和 Benoît Groz。[3] 2020年。图上正则简单路径查询的三分法。计算机系统科学杂志 108（2020年），29–48。https://doi.org/10.1016 2019 08 006 。/j.jcss. . . Pablo Barceló。. 图数据库查询。载于第*32*届 *ACM* 数据库原理研讨会（*ACM SIGMOD-SIGACT-SIGART* 联合举办，*PODS*）论文集[4] 2013年，美国纽约州纽约市，2013年6月22日至27日，Richard Hull 和 Wenfei Fan（编）。ACM，175 188 10 1145 2463664 2465216 – 。https://doi.org/ 。 / 。[5] 2019 Pablo Barceló、Diego Figueira 和 Miguel Romero。 。合取正则路径查询的有界性。载于自动机、语言与编程国际研讨会（ICALP）（《莱布尼茨国际信息学论文集》（LIPIcs），第132卷）。莱布尼茨信息学中心，104:1–104:15。https://doi.org/10 4230 2019 104 。/LIPIcs.ICALP. 。Pablo Barceló、Leonid Libkin、Anthony Widjaja Lin 和 Peter T. Wood。2012年。图结构数据路径查询的表达语言。ACM数据库系统汇刊[6] 37, 4 2012 31 1 31 46 10 1145 2389241 2389250 （ ），：–：。https://doi.org/ 。 / 。[7] 2019 Angela Bonifati、Wim Martens 和 Thomas Timm。 。探索 Wikidata 查询日志的迷宫。载于万维网大会（WWW）。127–138。[8] Angela Bonifati、Wim Martens 和 Thomas Timm。2020年。大型 SPARQL 查询日志的分析研究。VLDB期刊（2020）。即将发表，https://doi.org/10.1007/s00778-019-00558-9。[9] Diego Calvanese、Giuseppe De Giacomo、Maurizio Lenzerini 和 Moshe Y. Vardi。2000年。带逆运算的合取正则路径查询包含问题。载于知识表示与推理原理（KR）。176–185。[10] Ashok K. Chandra 和 Philip M. Merlin。1977年。关系数据库中合取查询的最优实现。载于1977年5月4日至6日于科罗拉多州博尔德举行的第9届 ACM 年度计算理论研讨会论文集，John E. Hopcroft、Emily P. Friedman 和 Michael A. Harrison（编）。ACM，77–90。https://doi.org/10 1145 800105 803397 。 / 。Stephen A. Cook。 。定理证明过程的复杂性。载于1971年5月3日至5日于俄亥俄州沙克高地举行的第3届 ACM 年度计算理论研讨会论文集，Michael A. Harrison、Ranan B. Banerji 和 Jeffrey D. Ullman（编）。ACM，151 158 10 1145 800157 805047 – 。[11] 1971 https://doi.org/ 。 / 。Isabel F. Cruz、Alberto O. Mendelzon 和 Peter T. Wood。 。支持递归的图形查询语言。载于美国计算机协会数据管理特别兴趣小组1987年度会议论文集，1987年5月27日至29日于美国加利福尼亚州旧金山举行，Umeshwar Dayal 和 Irving L. Traiger（编）。ACM出版社，323–330。https://doi.org/10.1145/38713.38749[13] Alin Deutsch、Nadime Francis、Alastair Green、Keith Hare、Bei Li、Leonid Libkin、Tobias Lindaaker、Victor Marsault、Wim Martens、Jan Michels、Filip Murlak、Stefan Plantikow、Petra Selmer、Oskar van Rest、Hannes Voigt、Domagoj Vrgoč、Mingxi Wu 和 Fred Zemke。2022年。GQL 和 SQL/PGQ 中的图模式匹配。载于2022年数据管理国际会议（SIGMOD ’ 22 2246 2258 10 1145 3514221 3526057 ）论文集。– 。[12] 1987 https://doi.org/ 。 / 。[14] 2020 Diego Figueira。 。UC RPQ 的包含问题：难易情形。载于数据库理论国际会议（ICDT）（《莱布尼茨国际信息学论文集》）。莱布尼茨信息学中心。[15] Diego Figueira、Adwait Godbole、S. Krishna、Wim Martens、Matthias Niewerth 和 Tina Trautner。2020年。简单合取正则路径查询包含问题。载于知识表示与推理原理（KR）。https://hal.archives-ouvertes.fr/hal-02505244[16] Daniela Florescu、Alon Levy 和 Dan Suciu。1998年。带正则表达式的合取查询的查询包含问题。载于美国计算机协会数据库系统原理研讨会（PODS）。ACM出版社，139–148。https://doi.org/10.1145/275487。275503[17] Nadime Francis、Alastair Green、Paolo Guagliardo、Leonid Libkin、Tobias Lindaaker、Victor Marsault、Stefan Plantikow、Mats Rydberg、Petra Selmer 和 Andrés Taylor。2018年。Cypher：一种不断发展的属性图查询语言。载于2018年数据管理国际会议论文集，2018 年 SIGMOD 会议，2018年6月10日至15日于得克萨斯州休斯敦举行，Gautam Das、Christopher M. Jermaine 和 Philip A. Bernstein（编）。ACM，1433–1445。https://doi.org/10 1145 3183713 3190657 。 / 。M. R. Garey 和 David S. Johnson。 。计算机与难解性：NP完全性理论指南。W. H. Freeman。[19] GQL 2021年。GQL标准网站。https://www.gqlstandards.org/。[20] Steve Harris 和 Andy Seaborne。2013年。SPARQL 1.1查询语言。http://www.w3.org/TR/sparql11-query。[21] Dexter Kozen。1977年。自然证明系统的下界。载于计算机科学基础年会研讨会（FOCS）。IEEE计算机学会出版社，254 266 10 1109 1977 16 – 。https://doi.org/ 。/SFCS. 。Katja Losemann 和 Wim Martens。 。SPARQL中正则表达式和属性路径的复杂性。ACM数据库系统汇刊 38, 4（2013），24:1–24:39。https://doi.org/10.1145/2494529[23] Stanislav Malyshev、Markus Krötzsch、Larry González、Julius Gonsior 和 Adrian Bielefeldt。2018年。充分利用 Wikidata：维基百科知识图谱中的语义技术应用。载于国际语义网会议（ISWC）。376–394。[24] Wim Martens 和 Tina Popp。2022年。即将发表。无向图上的正则迹查询和简单路径查询的复杂性。载于美国计算机协会数据库系统原理研讨会（PODS）。ACM。[25] Wim Martens 和 Tina Trautner。2019年。简单正则路径查询求值的二分性。ACM数据库系统汇刊 44, 4（2019），16:1–16:46。https://doi.org/10.1145/3331446[22] 2013

[26] Alberto O. Mendelzon 和 Peter T. Wood。1995年。图数据库中的正则简单路径查找。*SIAM*计算机杂志24, 6（1995），1235–1258。https://doi.org/10. 1137/S009753979122370X[27] Juan L. Reutter、Miguel Romero 和 Moshe Y. Vardi。2017年。图数据库上的正则查询。理论计算机系统61, 1（2017），31–83。https://doi.org/10. 1007/s00224-016-9676-2[28] Vladislav Rutenburg。1986年。广义图着色的复杂性。载于*1986*年计算机科学基础数学会议，布拉迪斯拉发，捷克斯洛伐克，*1996*年*8*月*25*日至*29*日，论文集（《计算机科学讲义》，第*233*卷），Jozef Gruska、Branislav Rovan 和 Juraj Wiedermann（编）。施普林格，573–581。https://doi.org/10.1007/BFb0016284

[29] Sherif Sakr、Angela Bonifati、Hannes Voigt、Alexandru Iosup、Khaled Ammar、Renzo Angles、Walid Aref、Marcelo Arenas、Maciej Besta、Peter A. Boncz、Khuza- ima Daudjee、Emanuele Della Valle、Stefania Dumbrava、Olaf Hartig、Bernhard Haslhofer、TimHegeman、Jan Hidders、Katja Hose、Adriana Iamnitchi、Vasiliki Kalavri、Hugo Kapp、Wim Martens、M. Tamer Özsu、Eric Peukert、Stefan Plan- tikow、Mohamed Ragab、Matei R. Ripeanu、Semih Salihoglu、Christian Schulz、Petra Selmer、Juan F. Sequeda、Joshua Shinavier、Gábor Szárnyas、Riccardo Tom- masini、Antonino Tumeo、Alexandru Uta、Ana Lucia Varbanescu、Hsiang-Yun Wu、Nikolay Yakovets、Da Yan 和 Eiko Yoneki。2021年。《未来属于大图：图处理系统的社区视角》。《美国计算机协会通讯》64, 9（2021年8月），62–71。https://doi.org/10.1145/3434642

A 节第 2 节的附录

命题 2.2 的证明。标准语义情形下的结果众所周知，因此我们聚焦于查询单射语义。假设 $G=(V,E)$ 和 Q 的形式为 $Q(\bar{x}) = A_1 \wedge \cdots \wedge A_m$。回忆一下，$Q(\bar{x})$ 等价于一个由 ε-自由 CRPQ 构成的并 $\mathrm{Q}^{\varepsilon-free}$。首先假设 $\bar{v} \in Q(G)$ 满足查询单射语义。那么，对于某个形式为 $Q'(\bar{z}) = A'_1 \wedge \cdots \wedge A'k$ 的 $Q' \in \mathrm{Q}^{\varepsilon-free}$，$\bar{v} \in Q'(G)$ 满足查询单射语义。不妨假设 A_{k+1}，……，A_m 恰好是构造 Q' 时被合并的 Q 的原子（也就是说，我们取这些原子上的 ε- 词）。特别地，原子 $A'i$ 的语言是 A_i 的语言去掉 ε 后所得，其中 $i \in \{1,\ldots,k\}$。于是存在一个从 $\mathrm{vars}(Q')$ 到 V 的单射映射 μ（满足 $\mu(\bar{z}) = \bar{v}$）；并且对于每个原子 $A'_i = x_i \xrightarrow{L'_i} y_i$，都存在一条从 $\mu(x_i)$ 到 $\mu(y_i)$ 的简单路径 π_i，且不同路径不共享内部节点。我们可以取展开 $E \in \mathrm{Exp}(Q)$，它由 Q 的展开配置生成：当 $i \in \{k+1,\ldots,m\}$ 时，将 A_i 映射到词 ε；对于 $i \in \{1,\ldots,k\}$，则将 A_i 映射到 π_i 的标签。我们可以定义一个同态 $h\colon E \to (G,\bar{v})$，将 E 的非内部变量按照 μ 进行映射，并将每个原子展开映射到 G 中对应的简单路径。由于这些路径不共享内部节点，这是一个单射同态，因此 $E \xrightarrow{\mathrm{inj}} (G,\bar{v})$。

对于另一个方向，假设 $h\colon E \xrightarrow{\mathrm{inj}} (G,\bar{v})$ 对某个展开 $E \in \mathrm{Exp}(Q)$ 成立。令 φ 为生成 E 的展开配置。我们按如下方式构造一个无 ε 的 CRPQ $Q' \in \mathrm{Q}^{\varepsilon-free}$：若 φ 将词 ε 赋给 A_i，则将 A_i 折叠；否则，从其语言中移除 ε。假设 Q $Q = A_1 \wedge \wedge'$ 的 A $\bar{v} \in Q$ G μ 形式为 $'(\bar{z}) ' \cdots 'k$。我们有 $'(\,)$q-inj。事实上，定义映射 $:\mathrm{vars}(Q') \to V$ 为 h 在 E 的非内部变量上的限制，并将路径 π_i 定义为 h 对 E 中原子 $A'i$ 的展开所取的像。由于 h 是单射，π_i 都是简单路径，且不共享内部节点。此外，映射 μ 是单射。因此，$\bar{v} \in Q'(G)$q-inj，这意味着所需的 $\bar{v} \in Q(G)$q-inj。□

命题 2.3 的证明。该证明与命题 2.2 的证明类似，只需将单射同态替换为原子单射同态。□

B 第 4 节附录

命题 4.3 的证明。该证明与命题 4.2 的证明完全相同，只需将同态替换为单射同态。为完整起见，我们在下文给出证明。

假设 $Q_1 \subseteq$q-inj Q_2，并取 $E_1(\bar{y}) \in \mathrm{Exp}(Q_1)$。回想一下，我们可以将查询 $E_1(\bar{y})$ 看作一个图数据库（同名），其中每个原子都被解释为一条边。我们有 $E_1 \,\mathrm{inj}\!\longrightarrow (E_1,\bar{y})$，因此 $\bar{y} \in Q_1(E_1)$q-inj。根据假设，$\bar{y} \in Q_2(E_1)$q-inj，也就是说，存在 $E_2 \in \mathrm{Exp}(Q_2)$，使得 $E_2 \,\mathrm{inj}\!\longrightarrow (E_1,\bar{y})$，即 $E_2 \,\mathrm{inj}\!\longrightarrow E_1$。反过来，假设对于某个图数据库 G 和节点元组 $\bar{v}$，有 $\bar{v} \in Q_1(G)$q-inj。存在一个展开 $E_1 \in \mathrm{Exp}(Q_1)$，使得 E_1 inj ~~$E_2 \to (G,\bar{v})$~~，根据假设，存在 $E_2 \in \mathrm{Exp}(Q_2)$，满足 $\bar{v}$ $\bar{v} \in$ Q_2 G □ inj。

通过复合，我们得到 inj $(\,,\,)$，因此 $(\,)$q-inj。

引理 4.4 的证明。我们从 (1)⇔(2) 开始。令 h 为 Ea-inj $\longrightarrow (G,\bar{v})$ 的一个见证。定义查询 $\widetilde{F} = E \wedge J$，其中 J 是所有等式原子 $x = y$ 的合取，且 $x,y \in \mathrm{vars}(E)$，以及 $h(x) = h(y)$。由于 h 是原子单射，我们有 $F := \widetilde{F}^{\equiv} \in$Expa-inj$(Q)$。此外，存在 $g\colon F \,\mathrm{inj}\!\longrightarrow (G,\bar{v})$，符合要求。事实上，令 $\Phi\colon \mathrm{vars}(E) \to \mathrm{vars}(F)$ 为典范重命名。对于每个 $x \in \mathrm{vars}(F)$，我们令 $g(x) = h(x')$，其中 x' 是 E 中任意满足 $\Phi(x') = x$ 的变量。根据构造，g 是一个单射同态。反过来，假设 $\sim \quad \sim F = F^{\equiv}$ 对于 $F = E \wedge J$ 成立，其展开为 $E \in \mathrm{Exp}(Q)$，并令 g 为 $F \,\mathrm{inj}\!\longrightarrow (G,\bar{v})$ 的一个见证。令 $\Phi\colon \mathrm{vars}(E) \to \mathrm{vars}(F)$ 为典范重命名。注意，Φ 实际上是从 E 到 F 的一个同态。根据 a-inj-扩展的定义，Φ 是从 E 到 F 的原子单射同态。将 Φ 与 g 复合，我们得到 E a-inj$\longrightarrow (G,\bar{v})$。用 G 代替 E' 的情况类似。□

命题 4.6 的证明。根据引理 4.4，只需考虑 (1) 和 (2) 两个条目之间的等价性。首先，假设 $Q_1 \subseteq$a-inj Q_2，并取 $F_1 \in$Exp 原子单射(Q_1)。我们可以将 $F_1(\bar{y})$ 视为一个图数据库（同名），其中每个原子都被解释为一条边。我们有 $F_1 \,\mathrm{inj}\!\longrightarrow (F_1,\bar{y})$，并且根据推论 4.5，得到 $\bar{y} \in Q_1(F_1)$ 原子单射。根据假设，$\bar{y} \in Q_2(F_1)$ 原子单射，即存在 $E_2 \in \mathrm{Exp}(Q_2)$，使得 E_2 a-inj$\longrightarrow (F_1,\bar{y})$，也就是说，$E_2$ a-inj$\longrightarrow F_1$。对于另一个方向，假设对于某个图数据库 G 和节点元组 $\bar{v}$，有 $\bar{v} \in Q_1(G)$ 原子单射。根据推论 4.5，存在 $F_1 \in$Exp 原子单射(Q_1)，以及 $g\colon F_1 \,\mathrm{inj}\!\longrightarrow (G,\bar{v})$。根据假设，存在 $E_2 \in \mathrm{Exp}(Q_2)$ 和 $f\colon E_2$ a-inj$\longrightarrow F_1$。通过将 f 与 g 组合，我们得到 E_2 a-inj$\longrightarrow (G,\bar{v})$。由此得出，$\bar{v} \in Q_2(G)$ 原子单射。□

C 定理 5.1 的完整证明

设 $Q_1(\bar{x_1})$ $Q_2(\bar{x_2})$ 为 CRPQ；我们要检验 $Q_1 \subseteq_{\text{查询单射}} Q_2$。

We then obtain the desired 边带标签的图，其中边上的标签是正则表达式。d因此，变量的度 [（分别为入度和出度）] 是包含该变量的原子数量 [（分别作为第二个变量和第一个变量）]。

高层思路。我们首先对证明作高层次的描述。我们将使用 Q_1 展开的多项式大小的“抽象”。对于 Q_1 的每个原子 A，这些抽象都包含语言如何映射到该原子中的全部信息 Q_2。例如，其中包括这样一条信息：“存在一条部分运行，从非确定有限自动机 A_L（其语言为 L）的状态 q 转移到状态 $q_\prime$，并且 Q_2 读取 A 的展开词”，或者“存在一条部分运行，从 A_L 的初始状态转移到 q，并读取 A 展开词的某个后缀”。这样的抽象包含了从展开中保留的所有必要信息，以便判断它是否为反例。确实，任何具有与某个反例相同抽象的展开也将是反例。

为了检验某个抽象 α 是否抽象了一个反例，我们需要考虑将 Q_2 的一个展开单射映射到 Q_1 的一个展开的所有可能方式。我们称之为同态类型，其中包含了 Q_2 的每个原子展开被映射到何处的信息。例如，我们可以知道 Q_2 的原子 A 的展开所映射到的路径

始于原子 $A_1 = x$ 的展开中的某个内部节点 $\xrightarrow{L_1}$ y 的 Q_1 随后到达变量 y，其状态为 q，并继续读取

原子 $A_2 = y \xrightarrow{L_2} z$ 的完整展开到达变量 z，状态为 $q_\prime$，并通过读取原子 $A_3 = z \xrightarrow{L_3} t$ 的展开的一个前缀，在某个内部节点到达终态，从而结束其路径。对于每种态射类型，我们可以通过检查它是否与某个抽象兼容来进行判断；例如，检查 α 是否确实包含以下信息：存在一个部分运行，从 $q_\prime$ 出发，读取 A_3 的展开的一个前缀后到达终态。

重要性质是，抽象 α 与态射类型 τ 兼容，当且仅当对于 Q_1 的每个带有抽象 α 的展开 E_1，都存在 Q_2 的一个展开 E_2，其态射类型为 τ，使得 $E_2 \xrightarrow{inj} E_1$。因此，多项式空间算法只需猜测 α，并检查对于每一种可能的态射类型 τ，α, τ 都不兼容。现在我们进一步阐述这些想法。

注 C.1. 任意 *CRPQ* 都与一个 *q*-注入等价的 *CRPQ* 等价，其中不存在同时与两个原子关联且入度和出度均为一的变量 y。这是因为 $x \xrightarrow{L} y \wedge y \xrightarrow{L'} x_\prime$（在查询单射语义或标准语义下）等价于 $x \xrightarrow{L \cdot L'} x_\prime$（假设 $y \notin \{x, x_\prime\}$）。

根据注 C.1，我们可以假设从 Q_2 的展开到 Q_1 的展开的映射满足：任意两个变量都不会被映射到某个原子展开的两个内部节点。

注释 C.2. 对于每个 *CRPQ* Q，都可以构造一个与之等价的 *CRPQ* 并 $Q_\prime$，使得 *(i)* $Q_\prime$ 中的任何语言都不包含 ε，且 *(ii)* $Q_\prime$ 中不存在两个不同的原子 $x \xrightarrow{L} y$ 和 $x \xrightarrow{L'} y$，使得某个单字母词 $a \in \mathbb{A}$ 属于 $L \cap L_\prime$。此外，$Q_\prime$ 是由多项式大小 *CRPQ* 构成的指数规模并，且判断一个 *CRPQ* 是否属于该并的问题属于多项式空间。

术语。路径（展开或有向图中的路径）指有向路径，也就是形如 $\pi = (v_0, v_1)(v_1, v_2) \cdots (v_{n-1}, v_n)$ 的边序列。路径的内部节点是指除起点和终点之外的任意节点（即 v_0 和 v_n）。对于路径 π 和态射 h，我们用 $h(\pi)$ 表示将每个顶点 v 替换为 $h(v)$ 后得到的路径。对于一条（有向）路径 π，我们分别用 $src(\pi)$ 和 $tgt(\pi)$ [resp.] 表示第一个 [和最后一个] 顶点。路径 π 的子路径是形如 $(v_i, v_{i+1}) \cdots (v_j, v_{j+1})$ 的路径，其中 $0 \le i \le j < n$（特别地，长度至少为 1）。路径 π 的中缀 [resp. 前缀、后缀] 是不包含 $src(\pi)$ 或 $tgt(\pi)$ [resp. 包含 $tgt(\pi)$ 且不包含 $src(\pi)$，包含 $src(\pi)$ 且不包含 $tgt(\pi)$] 的子路径。我们经常不区分正则语言、正则表达式、非确定有限自动机和包含正则语言的 CRPQ 原子。例如，我们可能会写“q 是原子 $x \xrightarrow{L} y$ 的终态”，意思是它是表示 L 的非确定有限自动机的终态。

查询的限制。为简化证明，我们假设 Q_1，Q_2 具有以下性质：

- 所有语言中都不存在 ε；
- 不存在两个原子 $x \xrightarrow{L} y$ 和 $x \xrightarrow{L'} y$，使得 $a \in \mathbb{A}$ 中存在某个单字母词 $L \cap L_\prime$；
- 这些查询是连通的。

稍后我们将说明如何消除这些假设 .

备注 C.3。根据备注 *C.1*，对于在 *q-inj* 语义下的 $Q_1 \subseteq$ *q-inj* Q_2 *CRPQ/CRPQ* 包含问题，并假设上述性质成立，我们可以将关注范围限制为单射同态 $E_1 \xrightarrow{inj} E_1$（其中 $E_i \in \text{Exp}(Q_i)$）满足：若两个不同的变量 $x, y \in vars(Q_1)$ 被映射到原子展开的不同内部节点，那么它们的度都必须为 *1*。也就是说，根据命题 *4.3* 的刻画，$Q_1 \subseteq$查询单射 Q_2 当且仅当对于每个 $E_1 \in \text{Exp}(Q_1)$，都存在 $E_2 \in \text{Exp}(Q_2)$，使得 $h: E_2 \xrightarrow{inj} E_1$，其中 h 具有上述性质。

不失一般性，假设 Q_2 各语言的所有非确定有限自动机的状态集两两不相交，并且它们都是完全且余完全的（即，对于每个字母和状态，都存在一条以该字母为标记的入转移和出转移）。考虑 A_{Q_2}，将 Q_2 各自动机的所有转移作（不相交）并，作为其转移。在此情形下，我们将 [中一个在其所属自动机中为初始的 [状态称为初始状态，将] 中一个在其所属自动机中为终态的] 状态称为终态。

展开的抽象 E_1 Q_1 是一个映射：α 将 Q_1 的原子映射到 P 的子集，其中

$$P = \{\langle q\text{-}q' \rangle : q, q' \text{ states of } \mathcal{A}_{Q_2}\} \cup \{\langle q \dashv q' \rangle : q, q' \text{ states of } \mathcal{A}_{Q_2}\} \cup \{\langle q \dashv\cdot\cdot \dashv q' \rangle : q, q' \text{ states of } \mathcal{A}_{Q_2}\} \cup \{\langle \cdot\cdot q\text{-}q' \cdot\cdot \rangle : q, q' \text{ states of } \mathcal{A}_{Q_2}\}$$

对于 $x \xrightarrow{w} y$ 的每个原子 A 的展开，我们都有：

- $q\text{-}q' \in \alpha(A)$：如果存在一条部分运行，使 A_{Q_2} 从 q 读取 w 到达 q'；
- $q \dashv q' \in \alpha(A)$：如果对于某个 $w = u \cdot v$（满足 $u, v \neq \varepsilon$），存在一条部分运行，使 A_{Q_2} 从 q 读取 u 到达终态，并且存在一条部分运行，从初始状态读取 v 到达 q'；
- $q \dashv\cdot\cdot \dashv q' \in \alpha(A)$：如果对于某个 $w = u \cdot s \cdot v$（满足 $s, u, v \neq \varepsilon$），存在一条部分运行，使 A_{Q_2} 从 q 读取 u 到达终态，并且存在一条部分运行，使 A_{Q_2} 从初始状态读取 v 到达 q'；
- $\cdot\cdot q\text{-}q' \cdot\cdot \in \alpha(A)$：如果对于某个 $w = u \cdot s \cdot v$（满足 $s, u, v \neq \varepsilon$），存在一条部分运行，使 A_{Q_2} 从 q 读取 s 到达 q'。

注意，任意抽象的大小都关于 Q_1，Q_2 呈多项式级。

的抽象构成其所有展开的抽象的集合。

命题 C.1. 判断一个映射是否是 Q_1 的抽象，属于 多项式空间。

事实上，标准的抽水论证表明，如果 α 是 Q_1 的抽象，那么它存在一个大小至多为指数级的见证展开。利用这个界，即时多项式空间算法可以为每个原子猜测其展开，并检查每个原子 A 是否具有抽象 $\alpha(A)$。具体做法是逐个猜测字母，同时跟踪其中包含的所有可能部分运行。该过程还会使用一个多项式大小的计数器，跟踪当前生成的展开的大小；一旦大小超过指数级界限，就拒绝该计算。

考虑有向图 G，它由将 Q_1 中的每个原子 $A = x \xrightarrow{L} y$ 替换为一条长度为 3 的路径 π_{GA}（即，添加两个新的内部顶点）得到。从 $Q_2(\bar{x}_2)$ 到 $Q_1(\bar{x}_1)$ 的一个态射类型 是一个对 (H, h)，满足 $h: H \xrightarrow{inj} \mathbf{G}$，且 H 是通过将 Q_2 中的每个原子 $A = x \xrightarrow{L} y$ 替换为一条从 x 到 y 的（非空）路径 π_A^H 得到的图。此外，我们还要求自由变量按相应方式映射，也就是说，$h(\bar{x}_2) = \bar{x}_1$。图 8 给出了一个态射类型 (H, h) 的示例。因此，由单射性可知，任何态射类型中 H 的大小都以线性方式受 Q_1 限制。

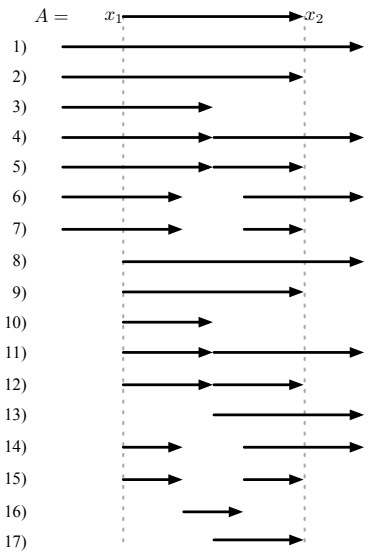

图 9：兼容性的所有可能情况。注意，由于备注 C.3，不存在其他情况。

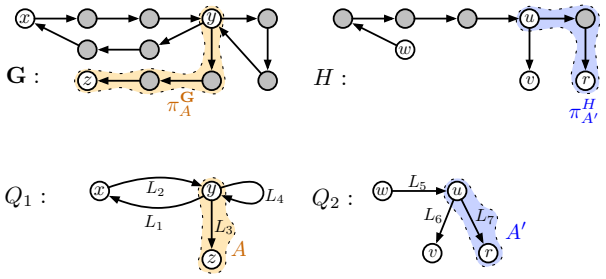

图 8：G 的定义和态射类型（H，h）的示例，来自 Q_1，Q_2。在此情况下，从 H 到 G 的单射态射 h 将 H 的每个节点映射到 G 上相同位置的节点（例如，H 的右下节点 r 映射到 G 的右下节点）。

命题 C.2. 判断一对 (H, h) 是否为态射类型属于多项式空间。

态射类型（H，h）与抽象 α 兼容，当且仅当存在一个映射 λ，将路径 π_A^H 的内部节点映射到 A 的状态，使得对于 Q_1 的每个原子 $A = x_1 \xrightarrow{L} x_2$，

- 如果存在 Q_2 的一个原子 A' 和 $\pi_{A'}^H$ 的一个中间子路径 π，满足 $h(\pi) = \pi_G A$，那么 $\lambda(src(\pi))\text{-}\lambda(tgt(\pi)) \in \alpha(A)$（对应图 9 中的情况 1）；
- 如果存在 Q_2 的一个原子 A'，以及 $\pi_{A'}^H$ 的一个后缀子路径 π，满足 $h(\pi) = \pi_G A$，那么 $\lambda(src(\pi))\text{-}q_F \in \alpha(A)$，其中 q_F 是 A' 的一个终态（图 9 中的情形 2）；
- 如果存在原子 A' of Q_2 和后缀子路径 π of $\pi_{A'}^H$，且 $h(\pi)$ 是 $\pi_G A$ 的前缀，那么对于某个 q，有 $\lambda(src(\pi))\text{+}q \in \alpha(A)$（图 9 中的情形 3）；
- 如果存在原子 $A'1, A'2$ of Q_2、π_1 of $\pi_{A'1}^H$ 的后缀子路径，以及 π_2 of $\pi_{A'2}^H$ 的前缀子路径，且 $tgt(\pi_1) = src(\pi_2)$ 和 $h(\pi_1\pi_2) = \pi_G A$，那么 $src(\pi_1)\text{+}tgt(\pi_2) \in \alpha(A)$（图 9 中的情形 4）；
- 如果存在 Q_2 的原子 $A'1, A'2$，以及 $\pi_{A'1}^H$ 的一个后缀子路径 π，满足 $tgt(\pi) = src(\pi_{A_2}^H)$ $h(\pi\pi_{A_2}^H) = \pi_A$ $\lambda(src(\pi)) \mid q_F \in \alpha(A)$ 和 $\alpha(A) \in G$，那么 $\alpha(A)\text{-}\alpha(A)$，其中 q_F 是 $A'2$ 的一个终态（图中的情形）；
- 如果存在 Q_2 的原子 $A'1$、$A'2$，$\pi_{A'1}^H$ 的一个后缀子路径 π_1，以及 $\pi_{A'2}^H$ 的一个前缀子路径 π_2，满足 $\neq src(\pi_2)$ 以 $h(\pi_1)$ 为前缀且属于 $\pi_G A$，而 $h(\pi_2)$ 是 $\pi_G A$ 的后缀，那么 $\lambda(src(\pi_1))\text{+}\cdot\cdot\text{+}\lambda(tgt(\pi_2)) \in \alpha(A)$（图 9 中的情形 6）；
- 如果 Q_2 中存在原子 $A'1$、$A'2$，且存在 $\pi_{A'1}^H$ 的一个后缀子路径 π，其中 $\pi(h)$ 是 A G π 的前缀，$\pi_A^H(h'2)$ 是 A G π 的后缀，并且 $\pi(tgt)\,\pi_A^H(\neq src'2)$，那么对于 $A'2$ 的某个终态 q_F，$src(\pi(\lambda))\text{+}\cdot\cdot\text{+}q_F\,A(\in\alpha)$（图 9 中的情形 7）；
- 如果 Q_2 中存在一个原子 A'，且存在 $\pi_{A'}^H$ 的一个前缀子路径 π，满足 $\pi(h) = \pi_G A$，那么 $q_0\text{-}\lambda(tgt(\pi))\,A(\in\alpha)$，其中 q_0 是 A' 的一个初始状态（图 9 中的情形 8）；
- 如果 Q_2 中存在一个原子 A'，满足 $\pi_A^H(h) = \pi_G A$，那么 $q_0\text{-}q_F\,\alpha A(\in)$，其中 q_0/q_F 是 A' 的一个初始状态/终态（图 9 中的情形 9）；
- 如果 Q_2 的原子 A' 是 $\pi_G A$ 的前缀，且 $h(\pi_{A'}^H)$，那么 $q_0\text{+}q \in \alpha(A)$，其中 q_0 是 A' 的初始状态，而 q 是任意状态。（图 9 中的情形10）；
- 如果 Q_2 有原子 $A'1, A'2$，且有一个前缀子路径 π_2，它是 $\pi_{A'2}^H$ 的前缀，并且 $tgt\,\pi_{A_1}^H = src\,\pi_2$ $h\,\pi_{A_1}^H\pi_2 = \pi_A$ $q_0 \mid \lambda\ tgt\ \pi_2 \in \alpha(A) \cdot\cdot\cdot \alpha(A) \in G$，那么 $\alpha(A) - \alpha(A)$，其中 q_0 是 $A'1$ 的初始状态（图中的情形）；
- 如果 Q_2 有原子 $A'1, A'2$，且 $tgt(\pi_{A'1}^H) = src(\pi_{A'2}^H)$，$h(\pi_{A'1}^H\pi_{A'2}^H) = \pi_G A$，那么 $q_0\text{+}q_F \in \alpha(A)$，其中 q_0 是 $A'1$ 的初始状态，q_F 是 $A'2$ 的终态（图 9 中的情形12）；
- 如果存在一个原子 A' 属于 Q_2，且 π 是 $\pi_{A'}^H$ 的前缀，其中 $h(\pi)$ 是 $\pi_G A$ 的后缀，那么对于某个 q，有 $q\text{+}\cdot\cdot\text{+}\lambda(tgt(\pi)) \in \alpha(A)$（图 9 中的情况 13）；
- 如果存在 $A'1$、$A'2$ 这两个 Q_2 的原子，以及 π_A^H'2 的前缀子路径 π_2，满足 $\neq src(\pi_2)\,\pi_A^H(tgt'1)$，$h(\pi_A^H'1)$ 是 $\pi_G A$ 的前缀，并且 $h(\pi_2)$ 是 $\pi_G A$ 的后缀，那么 $q_0\text{+}\cdot\cdot\text{+}\lambda(tgt(\pi_2)) \in \alpha(A)$，其中 q_0 是 $A'1$ 的初始状态（图 9 中的情形 14）；
- 如果 A 中存在原子 $A'1$、$Q_2'2$，且 $tgt(\pi_A^H'1) \neq src(\pi_A^H'2)$，$h(\pi_A^H'1)$ 是 $\pi_G A$ 的前缀，而 $h(\pi_A^H'2)$ 是 $\pi_G A$ 的后缀，那么 $q_0\text{+}\cdot\cdot\text{+}q_F \in \alpha(A)$，其中 q_0 是 $A'1$ 的初始状态，q_F 是 $A'2$ 的终态（图 9 中的情形 15）；
- 如果存在一个原子 A'，它属于 Q_2，且 $h(\pi_{A'}^H)$ 是 $\pi_G A$ 的中缀，那么，对于 q_0 和 q_F（它们是 A' 的初始状态和终态），有 $\cdot\cdot q_0\text{-}q_F\cdot\cdot \in \alpha(A)$（图 9 中的情况 16）；
- 如果存在一个原子 A' 属于 Q_2，且 $h(\pi_{A'}^H)$ 是 $\pi_G A$ 的后缀，那么对于某个状态 q 和 A' 的某个终态 q_F，有 $q\text{+}q_F \in \alpha(A)$（图 9 中的情形 17）。

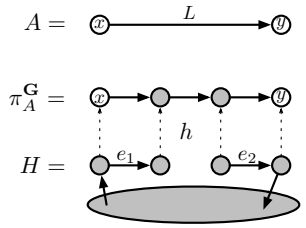

图 10：命题 C.4 的证明示例

以下结论可直接由 H, h, λ 是多项式有界，以及上述每个条件都可在多项式时间内测试得出。

命题 C.3。 测试态射类型是否与抽象兼容属于多项式空间。

兼容抽象的关键性质是，它们能够判定在给定抽象下对 Q_1 的展开是否构成包含问题$Q_1 \subseteq_{\text{查询单射}} Q_2$的反例。

命题 C.4. 以下各项等价：

(1) 存在一个与抽象 α相容的态射类型；*(2)* 对于每个具有抽象 α 的展开 $E_1 \in \mathrm{Exp}(Q_1)$，都存在某个展开 $E_2 \in \mathrm{Exp}(Q_2)$，使得 $E_2\ inj \longrightarrow E_1$；*(3)* 存在一个具有抽象 α 的展开 $E_1 \in \mathrm{Exp}(Q_1)$，以及一个展开 $E_2 \in \mathrm{Exp}(Q_2)$，使得 $E_2\ inj \longrightarrow E_1$。

证明。1 $E_1 \in \mathrm{Exp}(Q_1)$ 具有抽象 α，且 (H, h) 是通过映射 λ 的兼容态射类型。对于 Q_1 的每个原子 A，我们将满足 $h(\ )\ \mathrm{G}A$ 的πH中的每条路径 π A E_1 H 替换为 中的展开。 的其余边都属于映射部分 到某个 $\pi\mathrm{G}A$ 的路径；根据每个 $\alpha(A)$ 中形如 q+⋅+q' 和 ⋅⋅q-q'⋅⋅ 的元素所对应的见证词，将这些边替换为相应的路径。

例如，在图 10 所示的情形中，我们知道 $w \in L$ of A 的展开具有 $w = u \cdot s \cdot v$ 的形式，其中 AQ_2 存在一条从 q 到某个终态、读取 u 的部分运行，而 AQ_2 存在一条从某个初始状态到 q'、读取 v 的部分运行。因此，我们将边 e_1 替换为一条读取 u 的路径，并将边 e_2 替换为一条读取 v 的路径。由抽象的定义可知，所得合取查询（即，由所得边标记图表示的查询）是 Q_2 的一个展开，并且存在一个单射同态将其映射到 E_1。

$2 \Rightarrow 3)$ 这显然成立，因为 Q_1 的一个抽象也是其某个展开的抽象。

$3 \Rightarrow 1)$ 任取一个具有抽象 α 和 $E_2 \in$的 $E_1 \in \mathrm{Exp}(Q_1)$ $\mathrm{Exp}(Q_2)$，使得 g：$E_2\ inj \longrightarrow E_1$。现在，我们按照以下方式从 E_2 构造图 H。下文中，每当我们说将一条路径π 替换为一条长度为 n的路径时，意思是：(1) 删除 π 的所有内部节点及与这些节点关联的所有边；(2) 添加 $n-1$个新节点和 n 条边，使得存在一条长度为 n 的路径，从 src（π）到 tgt（π）。对于 E_1 中对应于原子 A 的展开的每条路径 π：

- 如果存在 π 的一个前缀 [（分别地，后缀]）没有 g-原像，我们就用一条边替换路径 $g^{-1}(\pi)$，并将该变量赋给 $\pi\mathrm{G}A$ 的第一个内部节点。对于其余情况，假设 π 的每个节点都有一个 g-原像。
- 如果 $g^{-1}(\pi)$ 是一条内部节点不包含 Q_2-变量的路径，那么我们将其替换为一条长度为 3 的无标记路径 π'。我们定义 h将 π'的第一个（分别为第二个）内部节点映射到 $\pi\mathrm{G}A$的第一个（分别为第二个）内部节点。
- 如果 $g^{-1}(\pi)$ 是一条包含一个变量 $x \in$ 变量(Q_2) 作为内部节点的路径，我们将 $g^{-1}(\pi)$截至x 的前半段路径替换为一条无标记边，并将后半段替换为一条长度为 2 的路径 π'。我们令 h 将变量 x 映射到 $\pi\mathrm{G}A$ 的第一个内部节点，并将 π 的内部节点映射到 $\pi\mathrm{G}A$ 的第二个内部节点。
- 如果 $g^{-1}(\pi)$ 包含两个变量和两条不相交路径，我们将它们分别替换为一条无标记边。我们将这些变量（即，路径的端点）相应地映射到 $\pi\mathrm{G}A$的两个内部顶点。

由此得到的图 H和映射 h 构成一种与 α兼容的态射类型。□

最后，多项式空间算法猜测一个映射 α，将 Q_1 的原子映射到 P 的子集，检查 α是否是 Q_1 的抽象（根据命题 C.1，此检查可在多项式空间内完成），并检查是否不存在与 α兼容的态射类型（H，h）（根据命题 C.2 和 C.3，以及多项式空间对补运算的封闭性，此检查可在多项式空间内完成）。根据命题 C.4，如果算法成功，则 Q_1 的任意展开 E_1 都是反例，因此$Q_1 \not\subseteq_{\text{查询单射}} Q_2$；否则，对于每个具有抽象 α 的展开E_1，都存在一种兼容的态射类型，这意味着 E_1 不是反例，因此$Q_1 \subseteq_{\text{查询单射}} Q_2$。

首先注意，如果 Q_2 不连通，我们可以调整多项式空间算法，检查 Q_2 的各个连通分量是否不存在这样的态射类型，使得它们全都与猜测的 Q_1的抽象兼容。此外，注意该过程可以扩展到由多项式大小的 CRPQ 组成的指数大小并：多项式空间算法首先从左侧的并中选择一个 CRPQ Q_1，猜测 Q_1 的一个抽象，并检查右侧并中的 CRPQ 是否都不存在兼容的态射类型。因此，结合注释 C.2，同样的论证可以扩展到任意 CRPQ（的并）。□

D 定理 5.2 的完整证明

我们从$Post$对应问题 (PCP)归约；这是一个众所周知的不可判定问题。PCP 的一个实例是由若干对组成的序列 (u_1, v_1) ……，(u_ℓ, v_ℓ)，其中 u_i和 v_i是字母表 Σ上的非空词。目标是判定是否存在一个解，也就是说，是否存在一个由 {1，……，ℓ}中的索引组成的序列 i_1，……，i_k，满足 $k \geq 1$，使得词 $u_{i_1} \cdots u_{i_k}$和 $v_{i_1} \cdots v_{i_k}$相同。

对于字母表 $\mathbb{A}$，我们将字母表 $\widehat{\mathbb{A}} = \{\widehat{a}\colon a \in \mathbb{A}\}$ 记作 $\widehat{\mathbb{A}}$。设 $(u_1, v_1) \ldots, (u_\ell, v_\ell)$ 是一个 PCP 实例，其底层字母表为 Σ。设字母表 $\mathbb{I} = \{I_1, \ldots, I_\ell\}$ 和 $\mathbb{A} = \Sigma \cup \mathbb{I} \cup \{\#, \#_\infty, \square, \$, \$', \$_\infty, \blacksquare, \blacksquare'\}$ 分别为 $\mathbb{I}$ 和 $\mathbb{A}$。我们在字母表 $\mathbb{A} \cup \widehat{\mathbb{A}}$ 上构造布尔合取正则路径查询 Q_1 和 Q_2，使得 PCP 实例 $(u_1, v_1) \ldots, (u_\ell, v_\ell)$ 有解，当且仅当 $Q_1 \not\subseteq$a-inj Q_2。特别地，该 PCP 实例有解，当且仅当存在一个针对 a-inj 语义的反例，即一个 a-inj-扩展 F，使得不存在满足 E a-inj$\longrightarrow F$ 的扩展 $E \in \mathrm{Exp}(Q_2)$。

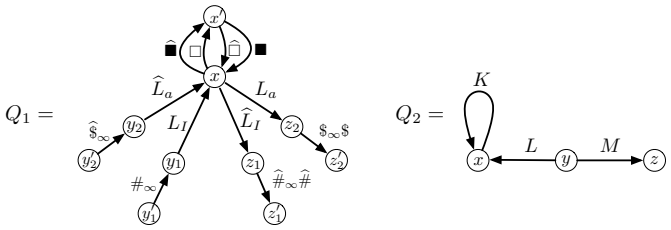

图 11：归约中的布尔 CRPQ Q_1 和 Q_2。

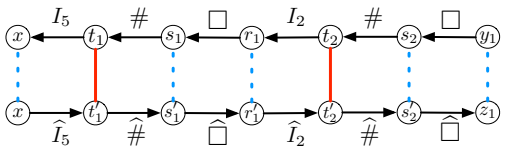

图 12：Q_1的良构展开的 I-$\widehat{I}$条件示例。我们分别展示原子 $y_1 \xrightarrow{L_I} x$ 和 $x \xrightarrow{L_I} z_1$的展开 $y_1 \xrightarrow{w_I} x$ 和 $x \xrightarrow{w_I} z_1$。词 $w_I = \square\#\, I_2\, \square\#\, I_5$ 和 $\widehat{w}_I = \widehat{I}_5\,\widehat{\ }\#\widehat{\square}\,\widehat{I}_2\,\widehat{\ }\#\,\widehat{\square}$ 编码了指标序列5，2。蓝色虚线表示相等变量对，红线表示不同变量。我们还有一些额外符号$\#\,\widehat{\ }\#\,\square,\ \widehat{\square}$。

$\mathbb{A}$ 中的符号与词 u_i 相关联，而 $\widehat{\mathbb{A}}$ 中的符号与词 v_i 相关联。对于每个 $u_i = a_1 \cdots a_k$，我们定义词 $U_i = a_1\$\,\blacksquare\, a_2\$\,\blacksquare \cdots a_k\$,\ \blacksquare$。类似地，对于每个 $v_i = a_1 \cdots a_k$，我们定义 $V_i = \widehat{\blacksquare}, \widehat{a}_k\,\widehat{\blacksquare}\,\widehat{\ }\$\ \widehat{a}_{k-1} \cdots \widehat{\blacksquare}\,\widehat{\ }\$\ \widehat{a}_1$。CRPQ Q_1 定义如下（图 1 展示了示意图）：

$$
\begin{aligned}
Q_1 = y_1 \xrightarrow{L_I} x \wedge y_2 \xrightarrow{\widehat{L}_a} x \wedge x \xrightarrow{\widehat{L}_I} z_1 \wedge x \xrightarrow{L_a} z_2 \\
\wedge\, x \xrightarrow{\square} x' \wedge x \xrightarrow{\widehat{\blacksquare}} x' \wedge x' \xrightarrow{\widehat{\square}} x \wedge x' \xrightarrow{\blacksquare} x \\
\wedge\, y_1' \xrightarrow{\#_\infty} y_1 \wedge y_2' \xrightarrow{\widehat{\$}_\infty} y_2 \wedge z_1 \xrightarrow{\widehat{\#}_\infty\widehat{\#}} z_1' \wedge z_2 \xrightarrow{\$_\infty\$} z_2'
\end{aligned}
$$

其中：

$$
L_I = (\square\#\mathbb{I})^+ \qquad \widehat{L}_I = (\widehat{\mathbb{I}}\widehat{\#}\widehat{\square})^+
$$

$$
L_a = (U_1 + \cdots + U_\ell)^+ \qquad \widehat{L}_a = (V_1 + \cdots + V_\ell)^+
$$

直观而言，来自 L_I 的一个词对应于从 $\{1,\ldots,\ell\}$ 中进行一次索引选择；对于 $\widehat{L}_I$，情况类似。另一方面，来自L_a [（分别是 $\widehat{L}_a$]）的一个词对应于从 $\{u_1,\ldots,u_\ell\}$（分别是 $\{v_1,\ldots,v_\ell\}$）中选择一个词[]。

我们关注 Q_1 的一种特定类型的原子单射展开，称为良构 展开，定义如下。其思路是，良构的原子单射展开对应于 PCP 实例的解。具体而言，若 Q_1 存在良构的原子单射展开，则 PCP 实例存在解，反之亦然。接着，我们将展示如何构造 Q_2，使得 Q_1 的原子单射展开是良构的，当且仅当它是$Q_1 \subseteq_{\text{原子单射}} Q_2$的反例。

~ ~令 F为 Q_1 的原子单射展开，并满足 $F = F^{=}$，其中 $F = E \wedge J$（这里 $E \in \mathrm{Exp}(Q_1)$ 和 J是等式原子）。如果 F满足以下四个条件，我们称其为良构：

(1) I-$\widehat{I}$-条件：此条件适用于 Q_1的原子 $y_1 \xrightarrow{L_I} x$和 $x \xrightarrow{\widehat{L}_I} z_1$。设 $y_1 \xrightarrow{L_I} x \quad x \xrightarrow{L_I} z_1$ 及展开中的原子所关联的展开分别为 $y_1 \xrightarrow{w_I} x$ 和 $x \xrightarrow{\widehat{w}_I} z_1$。此条件要求词$w_I$和 $\widehat{w}_I$分别具有形式 $w_I = \square\# I_{i_k} \cdots \square\# I_{i_1}$和$\widehat{w}_I = \widehat{I}_{i_1}\,\widehat{\ }\#\,\widehat{\square} \cdots \widehat{I}_{i_k}\,\widehat{\ }\#\,\widehat{\square}$，其中索引序列为 $i_1, \ldots, i_k \in \{1,\ldots,\ell\}$。换言之，原子 $y_1 \xrightarrow{L_I} x$ 和 $x \xrightarrow{L_I} z_1$ 的展开对应于相同的索引序列。~ I-I-条件还要求等式原子 J 具有特定行为（见图 12）。假设展开$y_1 \xrightarrow{w_I} x$ 具有如下形式：

$$
\begin{aligned}
y_1 \xrightarrow{\square} s_k \xrightarrow{\#} t_k \xrightarrow{I_{i_k}} r_{k-1} \xrightarrow{\square} s_{k-1} \xrightarrow{\#} t_{k-1} \xrightarrow{I_{i_{k-1}}} r_{k-2} \cdots \\
\cdots r_1 \xrightarrow{\square} s_1 \xrightarrow{\#} t_1 \xrightarrow{I_{i_1}} x
\end{aligned}
$$

~和 $x \xrightarrow{w_I} z_1$ 具有以下形式：

$$
\begin{aligned}
x \xrightarrow{\widehat{I}_{i_1}} t_1' \xrightarrow{\widehat{\#}} s_1' \xrightarrow{\widehat{\square}} r_1' \xrightarrow{\widehat{I}_{i_2}} t_2' \xrightarrow{\widehat{\#}} s_2' \xrightarrow{\widehat{\square}} r_2' \cdots \\
\cdots r_{k-1}' \xrightarrow{\widehat{I}_{i_k}} t_k' \xrightarrow{\widehat{\#}} s_k' \xrightarrow{\widehat{\square}} z_1
\end{aligned}
$$

那么，由等式原子 J 产生的关系$=_{\widetilde{F}}$满足：

(a) $t_1 \neq_{\widetilde{F}} t_1', \cdots, t_k \neq_{\widetilde{F}} t_k'$
(b) $s_1 =_{\widetilde{F}} s_1', \cdots, s_k =_{\widetilde{F}} s_k'$
(c) $r_1 =_{\widetilde{F}} r_1', \cdots, r_k =_{\widetilde{F}} r_k'$
where $r_k := y_1$ and $r_k' := z_1$.

(2) I-a-条件：此条件适用于 Q_1中的原子 $y_1 \xrightarrow{L_I} x$和 $x \xrightarrow{L_a} z_2$。设 $y_1 \xrightarrow{w_I} x$ 和 $x \xrightarrow{w_a} z_2$ 分别为展开 E中与原子 $y_1 \xrightarrow{L_I} x$和 $x \xrightarrow{L_a} z_2$关联的展开。该条件要求词 w_I和 w_a 具有以下形式：$w_I = \square\# I_{i_k} \cdots \square\# I_{i_1}$和$w_a = U_{i_1} \cdots U_{i_k}$，其中指标序列为 $i_1,\ldots,i_k \in \{1,\ldots,\ell\}$。直观地说，词 w_a根据序列 $i_1,\ldots,i_k$从$\{u_1,\ldots,u_\ell\}$中选择词。 I-a-条件还要求等式原子 J具有特定行为。假设展开 $y_1 \xrightarrow{w_I} x$ 具有以下形式：

$$
\begin{aligned}
y_1 \xrightarrow{\square} s_k \xrightarrow{\#} t_k \xrightarrow{I_{i_k}} r_{k-1} \xrightarrow{\square} s_{k-1} \xrightarrow{\#} t_{k-1} \xrightarrow{I_{i_{k-1}}} r_{k-2} \cdots \\
\cdots r_1 \xrightarrow{\square} s_1 \xrightarrow{\#} t_1 \xrightarrow{I_{i_1}} x
\end{aligned}
$$

且展开 $x \xrightarrow{w_a} z_2$ 的形式为：

$$
x \xrightarrow{\widetilde{U}_{i_1}} s_1' \xrightarrow{\blacksquare'} r_1' \xrightarrow{\widetilde{U}_{i_2}} s_2' \xrightarrow{\blacksquare'} r_2' \cdots r_{k-1}' \xrightarrow{\widetilde{U}_{i_k}} s_k' \xrightarrow{\blacksquare'} z_2
$$

~其中 U_i是从 U_i中去掉最后一个符号 $\blacksquare'$后得到的词。然后，由等式原子 J 生成的关系 $=_{\widetilde{F}}$满足：

（a）对于展开中的每个内部变量 t，$t_j \neq_{\widetilde{F}} t$

$r_{j-1}' \xrightarrow{\widetilde{U}_{i_j}} s_j'$ (here $r_0' := x$)
(b) $s_1 =_{\widetilde{F}} s_1', \cdots, s_k =_{\widetilde{F}} s_k'$

(c) $r_1 =_{\widetilde{F}} r_{\prime 1}, \ldots r_k =_{\widetilde{F}} r_{\prime k}$其中 $r_k := y_1$且 $r_{\prime k} := z_2$。

(3) $\widehat{a}$-*I*-条件：这类似于 *I*-*a*-条件，并且适用于 Q_1的原子 $y_2 \xrightarrow{L_a} x$ 和 $x \xrightarrow{L_I} z_1$。令 $y_2 \xrightarrow{w_a} x$ 和 $x \xrightarrow{w_I} z_1$ 分别为与展开 E中的原子 $y_2 \xrightarrow{L_a} x$和 $x \xrightarrow{L_I} z_1$ 相关联的展开。该条件要求词 $\widehat{w}_a$ 和 $\widehat{w}_I$ 具有$\widehat{w}_I = \widehat{I}_{i_1} \widehat{\#}\, \widehat{\square} \cdots \widehat{I}_{i_k} \widehat{\#}\, \widehat{\square}$以及 $\widehat{w}_a = V_{i_k} \cdots V_{i_1}$的形式，其中指标序列为 i_1、…、 $i_k \in \{1, \ldots, \ell\}$。也就是说，词 $\widehat{w}_a$根据序列 i_1、…、 i_k从$\{v_1$、…、 $v_\ell\}$中选取词。我们还对等式原子 J提出一些条件。假设展开 $y_2 \xrightarrow{w_a} x$ 具有以下形式：

$$y_2 \xrightarrow{\widetilde{\blacksquare}'} s_k \xrightarrow{V_{i_k}} r_{k-1} \xrightarrow{\widetilde{\blacksquare}'} s_{k-1} \xrightarrow{V_{i_{k-1}}} r_{k-2} \cdots r_1 \xrightarrow{\widetilde{\blacksquare}'} s_1 \xrightarrow{V_{i_1}} x$$

其中 V_i是从 V_i中移除第一个符号 $\widehat{\blacksquare}\prime$后得到的词。还假设展开 $x \xrightarrow{w_I} z_1$ 具有以下形式：

$$x \xrightarrow{\widehat{I}_{i_1}} t'_1 \xrightarrow{\widehat{\#}} s'_1 \xrightarrow{\widehat{\square}} r'_1 \xrightarrow{\widehat{I}_{i_2}} t'_2 \xrightarrow{\widehat{\#}} s'_2 \xrightarrow{\widehat{\square}} r'_2 \cdots \cdots r'_{k-1} \xrightarrow{\widehat{I}_{i_k}} t'_k \xrightarrow{\widehat{\#}} s'_k \xrightarrow{\widehat{\square}} z_1$$

那么，由等式原子 J产生的关系 $=_{\widetilde{F}}$满足：

(a) $t'_j \neq_{\widetilde{F}} t$, for every internal variable t of the expansion $s_j \xrightarrow{\widetilde{V}_{i_j}} r_{j-1}$ (here $r_0 := x$)
(b) $s_1 =_{\widetilde{F}} s'_1, \cdots, s_k =_{\widetilde{F}} s'_k$
(c) $r_1 =_{\widetilde{F}} r'_1, \cdots, r_k =_{\widetilde{F}} r'_k$
where $r_k := y_2$ and $r'_k := z_1$.

(4) $\widehat{a}$-*a*-条件：此条件适用于 Q_1中的原子 $y_2 \xrightarrow{L_a} x$和 $x \xrightarrow{L_a} z_2$。设 $y_2 \xrightarrow{\widehat{w}_a} x$ 和 $x \xrightarrow{w_a} z_2$ 分别是与原子 $y_2 \xrightarrow{L_a} x$ $x \xrightarrow{L_a} z_2$及展开中的原子相关联的展开。该条件要求词$\widehat{w}_a$和 w_a具有以下形式： $\widehat{w}_a = \clubsuit\, \clubsuit\, \widehat{a}_n \cdots \clubsuit\, \clubsuit\, \widehat{a}_1$和$w_a = a_1 \clubsuit\, \clubsuit \cdots a_n \clubsuit\, \clubsuit$，其中 $a_1 \cdots a_n \in \Sigma^*$是一个词（回想一下， Σ是 PCP 实例的字母表）。此处，$\clubsuit$ 是表示某个符号的占位符。直观而言，词 $\widehat{w}_a$和 w_a表示来自 Σ^*的同一个词。 $\widehat{a}$-*a*-条件还对等式原子 J提出了一些要求。假设展开 $y_2 \xrightarrow{w_a} x$ 具有以下形式：

$$y_2 \xrightarrow{\clubsuit} s_n \xrightarrow{\clubsuit} t_n \xrightarrow{\widehat{a}_n} r_{n-1} \xrightarrow{\clubsuit} s_{n-1} \xrightarrow{\clubsuit} t_{n-1} \xrightarrow{\widehat{a}_{n-1}} r_{n-2} \cdots \cdots r_1 \xrightarrow{\clubsuit} s_1 \xrightarrow{\clubsuit} t_1 \xrightarrow{\widehat{a}_1} x$$

且展开 $x \xrightarrow{w_a} z_2$ 的形式如下：

$$x \xrightarrow{a_1} t'_1 \xrightarrow{\clubsuit} s'_1 \xrightarrow{\clubsuit} r'_1 \xrightarrow{a_2} t'_2 \xrightarrow{\clubsuit} s'_2 \xrightarrow{\clubsuit} r'_2 \cdots \cdots r'_{n-1} \xrightarrow{a_n} t'_n \xrightarrow{\clubsuit} s'_n \xrightarrow{\clubsuit} z_2$$

那么，由等式原子 J 产生的关系 $=_{\widetilde{F}}$ 满足：

(a) $t_1 \neq_{\widetilde{F}} t'_1, \cdots, t_n \neq_{\widetilde{F}} t'_n$
(b) $s_1 =_{\widetilde{F}} s'_1, \cdots, s_n =_{\widetilde{F}} s'_n$
(c) $r_1 =_{\widetilde{F}} r'_1, \cdots, r_n =_{\widetilde{F}} r'_n$
where $r_n := y_2$ and $r'_n := z_2$.

良构性的关键性质在于，它可以通过不存在具有特定标签的有限个简单环和简单路径来刻画。为此，我们需要定义一些有限语言。回顾一下， $\widetilde{U}_i$ 是通过从U_i中移除最后一个符号 $\blacksquare\prime$得到的，而 V_i是通过从 V_i中移除第一个符号 $\widehat{\blacksquare}\prime$得到的。我们将 N定义为词 U_i的最大长度。对于 $i \leq j$，我们用 $e^{i,j}$表示正则表达式$(e^i + e^{i+1} + \cdots + e^j)$。我们有：

$$K_{I\widehat{I}} = \mathbb{I}\widehat{\mathbb{I}} + \#_\infty \widehat{\mathbb{I}} + \mathbb{I}\widehat{\#}_\infty$$

$$M_{I\widehat{I}} = \sum_{i \neq j} I_i \widehat{I}_j + \widehat{\mathbb{I}}\# + \widehat{\#}\mathbb{I} + \#\mathbb{I}\widehat{\mathbb{I}}\widehat{\#} + \square\widehat{\square} + \#_\infty \widehat{\mathbb{I}} + \mathbb{I}\widehat{\#}_\infty$$

$$K_{Ia} = \mathbb{I}\Sigma + \#_\infty \Sigma + \mathbb{I}\$_\infty$$

$$M_{Ia} = (\Sigma + \$ + \$' + \blacksquare)\mathbb{I} + (\Sigma + \$ + \blacksquare)^{1,N} \# + \sum_i \sum_{j \neq i} I_i \widetilde{U}_j + \\ + \#\mathbb{I}(\widetilde{U}_1 + \cdots + \widetilde{U}_\ell) + \square\blacksquare' + \#_\infty \Sigma + \mathbb{I}\$_\infty$$

$$K_{\widehat{a}\widehat{I}} = \widehat{\Sigma}\widehat{\mathbb{I}} + \widehat{\$}_\infty \widehat{\mathbb{I}} + \widehat{\Sigma}\widehat{\#}_\infty$$

$$M_{\widehat{a}\widehat{I}} = \widehat{\mathbb{I}}(\widehat{\Sigma} + \widehat{\$} + \widehat{\$'} + \widehat{\blacksquare}) + \widehat{\#}\widehat{\Sigma} + \widehat{\mathbb{I}}\widehat{\#}(\widehat{\Sigma} + \widehat{\$} + \widehat{\blacksquare}) + \sum_i \sum_{j \neq i} \widetilde{V}_j \widehat{I}_i + \\ + (\widetilde{V}_1 + \cdots + \widetilde{V}_\ell)\widehat{\mathbb{I}}\widehat{\#} + \widehat{\blacksquare}'\widehat{\square} + \widehat{\$}_\infty \widehat{\mathbb{I}} + \widehat{\Sigma}\widehat{\#}_\infty$$

$$K_{\widehat{a}a} = \widehat{\Sigma}\Sigma + \widehat{\$}_\infty \Sigma + \widehat{\Sigma}\$_\infty$$

$$M_{\widehat{a}a} = \sum_{a \neq b} \widehat{a}b + \Sigma(\widehat{\$} + \widehat{\$'}) + (\$ + \$')\widehat{\Sigma} + (\widehat{\$} + \widehat{\$'})\widehat{\Sigma}\Sigma(\$ + \$') + \\ + (\widehat{\blacksquare} + \widehat{\blacksquare}')(\blacksquare + \blacksquare') + \widehat{\$}_\infty \Sigma + \widehat{\Sigma}\$_\infty$$

命题 D.1. 设 F 是 Q_1 的原子单射展开。那么：

(1) F满足 *I*-*I*-条件，当且仅当 F不包含标签属于 $K_{I\widehat{I}}$的简单环，也不包含标签属于 $M_{I\widehat{I}}$的简单路径。(2) F满足 *I*-*a*-条件，当且仅当 F不包含标签属于 K_{Ia} 的简单环，也不包含标签属于 M_{Ia}的简单路径。(3) F满足 $\widehat{a}$-*I*-条件，当且仅当 F不包含标签属于 $K_{\widehat{a}\widehat{I}}$的简单环，也不包含标签属于 $M_{\widehat{a}\widehat{I}}$的简单路径。(4) F满足 $\widehat{a}$-*a*-条件，当且仅当 F不包含标签属于 $K_{\widehat{a}a}$ 的简单环，也不包含标签属于 $M_{\widehat{a}a}$的简单路径。

证明。设 $F = \widetilde{F}^{=}$for $\widetilde{F} = E \wedge J$，其中 $E \in \mathrm{Exp}(Q_1)$，且 J是等式原子。我们从第 (1) 项开始。令 $y_1 \xrightarrow{w_I} x$和 $x \xrightarrow{w_I} z_1$为与原子 $y_1 \xrightarrow{L_I} x$ $x \xrightarrow{L_I} z_1$以及展开中的原子相关联的展开。假设 E $w_I = \square\, I_{i_k} \cdots \square\, I_{i_1}$ #⋯# and$\widehat{w}_I = \widehat{I}_{j_1} \widehat{\#}\, \widehat{\square} \cdots \widehat{I}_{j_p} \widehat{\#}\, \widehat{\square}$，其中索引为 $i_1, \ldots, i_k, j_1, \ldots, j_p \in \{1, \ldots, \ell\}$。另假设 $y_1 \xrightarrow{w_I} x$ 具有如下形式：

$$y_1 \xrightarrow{\square} s_k \xrightarrow{\#} t_k \xrightarrow{I_{i_k}} r_{k-1} \xrightarrow{\square} s_{k-1} \xrightarrow{\#} t_{k-1} \xrightarrow{I_{i_{k-1}}} r_{k-2} \cdots \cdots r_1 \xrightarrow{\square} s_1 \xrightarrow{\#} t_1 \xrightarrow{I_{i_1}} x$$

and $x \xrightarrow{w_I} z_1$ 具有如下形式：

$$x \xrightarrow{\widehat{I}_{j_1}} t'_1 \xrightarrow{\widehat{\#}} s'_1 \xrightarrow{\widehat{\square}} r'_1 \xrightarrow{\widehat{I}_{i_2}} t'_2 \xrightarrow{\widehat{\#}} s'_2 \xrightarrow{\widehat{\square}} r'_2 \cdots \cdots r'_{p-1} \xrightarrow{\widehat{I}_{j_p}} t'_p \xrightarrow{\widehat{\#}} s'_p \xrightarrow{\widehat{\square}} z_1$$

我们首先考虑第 (1) 项的反向。我们有 $t_1 \neq_{\widetilde{F}} t'1$，否则就会存在一个标签属于 $\mathbb{I}\widehat{\mathbb{I}} \subseteq K_{I\widehat{I}}$ 的简单环 $t_1 \xrightarrow{I_{i_1}} x \xrightarrow{I_{j_1}} t'1$。这意味着 $i_1 = j_1$。如果情况并非如此，就会存在一条标签属于 $\sum_{i\neq j} I_i\widehat{I}_j \subseteq M_{I\widehat{I}}$ 的简单路径 $t_1 \xrightarrow{I_{i_1}} x \xrightarrow{I_{j_1}} t'1$。我们断言 $s_1 =_{\widetilde{F}} s'1$。首先注意，$t_1 \neq_{\widetilde{F}} s'1$ 且 $t'1 \neq_{\widetilde{F}} s_1$。事实上，如果 $t_1 =_{\widetilde{F}} s'1$，那么就会存在一条简单路径 $t'_1 \widehat{\#}\!\!\rightarrow s'_1 \xrightarrow{I_{i_1}} x$，其标签属于 $\widehat{\#}\,\mathbb{I} \subseteq M_{I\widehat{I}}$。如果 $t'1 =_{\widetilde{F}} s_1$，那么就会存在一条标签属于 $\widehat{\mathbb{I}}\# \subseteq M_{I\widehat{I}}$ 的简单路径 $x \xrightarrow{I_{j_1}} t'1 \#\!\!\rightarrow t_1$。由此可知 $s_1 =_{\widetilde{F}} s'1$，否则路径 $s_1 \#\!\!\rightarrow t_1 \xrightarrow{I_{i_1}} x \xrightarrow{I_{j_1}} t'_1 \widehat{\#}\!\!\rightarrow s'1$ 就会是一条标签属于 $\#\,\mathbb{I}\widehat{\mathbb{I}}\widehat{\#} \subseteq M_{I\widehat{I}}$ 的简单路径。最后，我们有 $r_1 =_{\widetilde{F}} r'1$。如果这不为真，那么 $r_1 \xrightarrow{\square} s_1 \xrightarrow{\square} r'1$ 就会是一条标签属于 $\square\widehat{\square} \subseteq M_{I\widehat{I}}$ 的简单路径。

我们可以重复这一论证，在每一步中将“中间”变量 x 替换为相应的新“中间”变量 r_i（见图 12）。我们得到如下结果（$\alpha = \min\{k, p\}$）：

- $i_1 = j_1, i_2 = j_2, \ldots, i_\alpha = j_\alpha$
- $t_1 \neq_{\widetilde{F}} t'_1, \cdots, t_\alpha \neq_{\widetilde{F}} t'_\alpha$
- $s_1 =_{\widetilde{F}} s'_1, \cdots, s_\alpha =_{\widetilde{F}} s'_\alpha$
- $r_1 =_{\widetilde{F}} r'_1, \cdots, r_\alpha =_{\widetilde{F}} r'_\alpha$

其中 $r_k := y_1$ 且 $r'p := z_1$ 注意，要得出 *I-I*-条件，只需证明 $\alpha = k = p$。为导出矛盾，首先假设 $k < p$。我们知道 $y_1 =_{\widetilde{F}} r'k$，且 $r'k \neq z_1$。特别地，我们至少有以下原子：

$$y'_1 \xrightarrow{\#_\infty} y_1 \qquad r'_k \xrightarrow{\widehat{I}_j} t'_{k+1}$$

我们分两种情况。如果 $y'1 =_{\widetilde{F}} t'k+1$，那么我们有一个简单环 $y'1 \#\xrightarrow{\infty} y_1 \xrightarrow{I_j} t'k+1$，其标签属于 $\#_\infty\widehat{\mathbb{I}} \subseteq K_{I\widehat{I}}$。另一方面，如果 $y'1 \neq_{\widetilde{F}} t'k+1$，那么我们有一条简单路径 $y'1 \#\xrightarrow{\infty} y_1 \xrightarrow{I_j} t'k+1$，其标签属于 $\#_\infty\widehat{\mathbb{I}} \subseteq M_{I\widehat{I}}$。无论哪种情况，我们都会得到矛盾。现在假设 $k > p$。我们知道 $r_p =_{\widetilde{F}} z_1$，并且 $r_p \neq y_1$。至少有以下原子：

$$t_{p+1} \xrightarrow{I_i} r_p \qquad z_1 \xrightarrow{\widehat{\#}_\infty} z''_1$$

同样，我们分两种情况。如果 $t_{p+1} =_{\widetilde{F}} z''1$，那么我们有一个简单环 $t_{p+1} \xrightarrow{I_i} r_p \#\xrightarrow{\infty} z''1$，其标签属于 $\mathbb{I}\widehat{\#}_\infty \subseteq K_{I\widehat{I}}$。另一方面，如果 $t_{p+1} \neq_{\widetilde{F}} z''1$，那么我们有一条简单路径 $t_{p+1} \xrightarrow{I_i} r_p \#\xrightarrow{\infty} z''1$，其标签属于 $\mathbb{I}\widehat{\#}_\infty \subseteq M_{I\widehat{I}}$。无论哪种情况，都会得出矛盾。由此可知 $k = p$，因此 *I-Î*-条件成立。

第（1）项的正向部分可直接由 *I-I*-条件的定义以及对语言 $K_{I\widehat{I}}$ 和 $M_{I\widehat{I}}$ 的检查得出。

现在我们转向第（2）项。设 $y_1 \xrightarrow{w_I} x$ 和 $x \xrightarrow{w_a} z_2$ 分别是展开 E 中 $w_I = \square\ I_{i_k}\ \square\ I_{i_1}$ 和 $w_a = L_a^{i_1 \ldots i_k}$ 原子 $y_1 \xrightarrow{U_{j_1} \ldots U_{j_p}} x$ 和 $x \xrightarrow{} z_2$ 所对应的展开。假设 $j_1 \ldots j_p \in 1 \ldots \ell$ 引入了 $\#\cdots\#$ 和 $\cdots$ 对索引，且 $\{,\}$ 成立。设展开 $y_1 \xrightarrow{w_I} x$ 具有如下形式：

$$y_1 \xrightarrow{\square} s_k \xrightarrow{\#} t_k \xrightarrow{I_{i_k}} r_{k-1} \xrightarrow{\square} s_{k-1} \xrightarrow{\#} t_{k-1} \xrightarrow{I_{i_{k-1}}} r_{k-2} \cdots \cdots r_1 \xrightarrow{\square} s_1 \xrightarrow{\#} t_1 \xrightarrow{I_{i_1}} x$$

并且展开 $x \xrightarrow{w_a} z_2$ 的形式如下：

$$x \xrightarrow{\widetilde{U}_{j_1}} s'_1 \xrightarrow{\blacksquare'} r'_1 \xrightarrow{\widetilde{U}_{i_2}} s'_2 \xrightarrow{\blacksquare'} r'_2 \cdots r'_{p-1} \xrightarrow{\widetilde{U}_{j_p}} s'_p \xrightarrow{\blacksquare'} z_2$$

其中，U_j 是从 U_j 中去掉最后一个符号 $\blacksquare'$ 后得到的词。

我们首先考虑第 (2) 项的反向。假设展开 $x \xrightarrow{U_{j_1}} s'1$ 的形式如下：

$$x \xrightarrow{\clubsuit} o_1 \xrightarrow{\clubsuit} o_2 \cdots o_m \xrightarrow{\clubsuit} s'_1$$

其中 $\clubsuit$ 是一个代表某个符号的占位符。我们声称，对于所有 $t \in \{o_1, \ldots, o_{m+1}\}$，$o_{m+1} := s'1$。首先注意到 $o_1 \in \Sigma$ 且 $\mathbb{I}\Sigma \subseteq K_{Ia}$，否则就会存在一个简单环 $t_1 \xrightarrow{I_{i_1}} x\clubsuit$，其中 $\clubsuit$。特别是，该标签将属于；矛盾。现在我们用归纳法论证。假设对于某个 $h \in \{1, \ldots, m\}$，$t_1 \neq_{\widetilde{F}} o_h$。反设 $t_1 =_{\widetilde{F}} o_{h+1}$。我们有一条简单路径 $o_h\clubsuit \rightarrow o_{h+1} \xrightarrow{I_{i_1}} x$，其中 $\clubsuit \in \Sigma \cup \{\$\ \blacksquare,\ \$'\}$（注意，$\clubsuit$ 不可能是 $\blacksquare'$）。那么该标签属于 $(\Sigma + \$ + \$' + \blacksquare)\,\mathbb{I} \subseteq M_{Ia}$；矛盾。

现在我们断言，对于所有 $t \in \{o_1, \ldots, o_m\}$，$s_1 \neq_{\widetilde{F}} t$。反证，假设对于某个 $t \in \{o_1, \ldots, o_m\}$，$s_1 =_{\widetilde{F}} t$。那么存在一条简单路径 $x \xrightarrow{U} t\# \rightarrow t_1$，其中 $U \in (\Sigma + \$ + \blacksquare)^{1,N}$（请注意，我们利用了这样一个事实：对于所有 $t \in \{o_1, \ldots, o_m\}$，$t_1 \neq_{\widetilde{F}} t$；否则该路径不一定是简单路径）。这条简单路径的标签属于 $(\Sigma + \$ + \blacksquare)^{1,N}\# \subseteq M_{Ia}$；矛盾。

我们有 $i_1 = j_1$。若情况并非如此，则会有一条简单路径 $t_1 \xrightarrow{I_{i_1}} x \xrightarrow{U_{j_1}} s'1$，其标签属于 $\sum_i \sum_{j\neq i} I_i\widetilde{U}_j \subseteq M_{Ia}$。此外，我们有 $s_1 =_{\widetilde{F}} s'1$，否则就会有一条简单路径 $s_1 \#\!\!\rightarrow t_1 \xrightarrow{I_{i_1}} x \xrightarrow{U_{j_1}} s'1$，其标签属于 $\#\,\mathbb{I}(\widetilde{U}_1 + \cdots + \widetilde{U}_\ell) \subseteq M_{Ia}$。最后，我们有 $r_1 =_{\widetilde{F}} r'1$。若情况并非如此，则存在一条简单路径 $r_1 \xrightarrow{\square} s_1 \xrightarrow{\blacksquare'} r'1$，其标签属于 $\square\blacksquare' \subseteq M_{Ia}$。

与条目 (1) 的情况一样，我们可以重复这一论证，在每一步中将“中间”变量 x 替换为相应的新“中间”变量 r_i。我们得到如下结果（$\alpha = \min\{k, p\}$）：

- $i_1 = j_1, i_2 = j_2, \ldots, i_\alpha = j_\alpha$
- For every $j \in \{1, \ldots, \alpha\}$, we have $t_j \neq_{\widetilde{F}} t$, for every internal variable t of the expansion $r'_{j-1} \xrightarrow{\widetilde{U}_{i_j}} s'_j$ (here $r'_0 := x$)
- $s_1 =_{\widetilde{F}} s'_1, \cdots, s_\alpha =_{\widetilde{F}} s'_\alpha$
- $r_1 =_{\widetilde{F}} r'_1, \cdots, r_\alpha =_{\widetilde{F}} r'_\alpha$

其中 $r_k := y_1$ 且 $r_{\prime p} := z_2$。通过使用与条目 (1) 的情况相同的论证，我们得到 $k = p$，因此 I-a-条件成立。

第 (2) 项的正向直接由 I-a-条件的定义以及对语言 K_{Ia} 和 M_{Ia}的检查得出。

第 (3) 项和第 (4) 项的情形与第 (1) 项和第 (2) 项类似。 □

令 $Q_2^{\circlearrowleft}$和 $Q_2^{\rightarrow}$为以下 CRPQfin中的布尔合取正则路径查询：

$$Q_2^{\circlearrowleft} = x \xrightarrow{K^{\circlearrowleft}} x \qquad Q_2^{\rightarrow} = y \xrightarrow{M^{\rightarrow}} z$$

其中 $K^{\circlearrowleft}$：

$$= K_{I\widehat{I}} + K_{Ia} + K_{\widehat{a}\widehat{I}} + K_{\widehat{a}a} \qquad M^{\rightarrow} = M_{I\widehat{I}} + M_{Ia} + M_{\widehat{a}\widehat{I}} + \text{且：。}$$

$M_{\widehat{a}a}$ 根据命题 D.1，我们得到了右侧查询为 CRPQ $Q_2^{\circlearrowleft}$和 $Q_2^{\rightarrow}$ 的并 时的归约，并将其记为 $Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}$。

断言 D.2. 令 F为 Q_1的原子单射-扩展。当且仅当 $Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}(F)$ 原子单射 = ∅时， F是良构的。此外，当且仅当 $Q_1 \not\subseteq$原子单射 $Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}$时，*PCP* 实例存在解。

根据断言 D.2，我们得到：在原子单射语义下，来自 CRPQfin 的一个 CRPQ 包含于两个 CRPQ 的并这一问题是不可判定的。我们通过说明如何像图 11 中那样用单个查询 $Q_2 \in$ CRPQfin 模拟并 $Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}$，来完成证明。

我们定义以下语言：

$$K_{dummy} = (\square + \widehat{\blacksquare} + \widehat{\blacksquare}')(\widehat{\square} + \blacksquare + \blacksquare')$$

$$M_{dummy} = \widehat{\#} + \$ + \$'$$

$$L = \varepsilon + \mathbb{I} + \#\mathbb{I} + \widehat{\#}\mathbb{I} + \square\#\mathbb{I} + \#_\infty + (\Sigma + \$ + \$' + \blacksquare)\,\mathbb{I} +$$
$$\widehat{\Sigma} + \widehat{\#}\widehat{\Sigma} + (\widetilde{V}_1 + \cdots + \widetilde{V}_\ell) + \widehat{\blacksquare}'\,(\widetilde{V}_1 + \cdots + \widetilde{V}_\ell) + \widehat{\$}_\infty +$$
$$+ (\$ + \$')\,\widehat{\Sigma} + (\widehat{\$} + \widehat{\$}')\,\widehat{\Sigma} + (\widehat{\blacksquare} + \widehat{\blacksquare}')(\widehat{\$} + \widehat{\$}')\,\widehat{\Sigma}$$

令 Q_2 为如下定义的 CRPQ：

$$Q_2 = x \xrightarrow{K} x \wedge y \xrightarrow{L} x \wedge y \xrightarrow{M} z$$

其中 $K := K^{\circlearrowleft} + K_{dummy}$ 且 $M := M^{\rightarrow} + M_{dummy}$。我们以下述断言结束：

断言 D. 3 F Q_1 F ≠.设 为 ∅ Q_2 F ≠ ∅ 的一个原子单射扩展。那么，$Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}()$ 是原子单射，当且仅当 () 是原子单射。

证明。对于正向，假设 $Q_2^{\circlearrowleft} \vee Q_2^{\rightarrow}(F)$ 是原子单射 ≠ ∅。首先假设 $Q_2^{\circlearrowleft}(F)$ 是原子单射 ≠ ∅。对于映射到 F 的简单环的标签 w，我们分两种情况讨论，并给出一个映射到 F 的 Q_2 的展开：

- $w \in K_{I\widehat{I}} \cup K_{\widehat{a}\widehat{I}}$：take expansion $x \xrightarrow{w} x \wedge y \xrightarrow{\varepsilon} x \wedge y \xrightarrow{\widehat{\#}} z$.
- $w \in K_{Ia} \cup K_{\widehat{a}a}$：take either expansion $x \xrightarrow{w} x \wedge y \xrightarrow{\varepsilon} x \wedge y \xrightarrow{\$} z$ or $x \xrightarrow{w} x \wedge y \xrightarrow{\varepsilon} x \wedge y \xrightarrow{\$'} z$.

注意上文中的 $\varepsilon \in L$和 $\widehat{\ }\#,\ \$,\ \$' \in M_{dummy} \subseteq M$.

现在假设 $Q_2^{\rightarrow}(F)$原子单射 ≠ ∅。同样，我们考虑映射到 F 的简单路径的标签 w 的所有可能情况，并且给出映射到 F 的 Q_2 的一种展开。我们从情况$w \in M_{I\widehat{I}}$开始：

- $w = I_p\widehat{I_q} \in \sum_{i\neq j} I_i\widehat{I_j}$：take expansion $x \xrightarrow{\square\widehat{\square}} x \wedge y \xrightarrow{I_p} x \wedge y \xrightarrow{w} z$.
- $w \in \widehat{\mathbb{I}}\#$：take expansion $x \xrightarrow{\square\widehat{\square}} x \wedge y \xrightarrow{\varepsilon} x \wedge y \xrightarrow{w} z$.
- $w = \widehat{\#}I_p \in \widehat{\#}\mathbb{I}$：take expansion $x \xrightarrow{\square\widehat{\square}} x \wedge y \xrightarrow{\widehat{\#}I_p} x \wedge y \xrightarrow{w} z$.
- $w = \#I_p\widehat{I_q\#} \in \#\mathbb{I}\widehat{\mathbb{I}\#}$：take expansion $x \xrightarrow{\square\widehat{\square}} x \wedge y \xrightarrow{\#I_p} x \wedge y \xrightarrow{w} z$.
- $w \in \square\,\widehat{\square}$：take expansion $x \xrightarrow{\square\widehat{\square}} x \wedge y \xrightarrow{\square\#I_i} x \wedge y \xrightarrow{w} z$, for a suitable $I_i \in \mathbb{I}$.
- $w \in \#_\infty\widehat{\mathbb{I}}$：take expansion $x \xrightarrow{\square\widehat{\square}} x \wedge y \xrightarrow{\#_\infty} x \wedge y \xrightarrow{w} z$.
- $w = I_p\widehat{\#_\infty} \in \mathbb{I}\widehat{\#_\infty}$：take expansion $x \xrightarrow{\square\widehat{\square}} x \wedge y \xrightarrow{I_p} x \wedge y \xrightarrow{w} z$.

注意 $\square\widehat{\square} \in K_{dummy} \subseteq K$和 I_p， ε， $\widehat{\ }\#I_p$， $\#I_p$， $\square\#I_i$， $\#_\infty \in L$对于情形 $w \in M_{Ia}$，我们有如下结论：

- $w = \clubsuit I_p \in (\Sigma + \$ + \$' + \blacksquare)\,\mathbb{I}$：take expansion $x \xrightarrow{\square\blacksquare'} x \wedge y \xrightarrow{\clubsuit I_p} x \wedge y \xrightarrow{w} z$.
- $w \in (\Sigma + \$ + \blacksquare)^{1,N}\,\#$：take expansion $x \xrightarrow{\square\blacksquare'} x \wedge y \xrightarrow{\varepsilon} x \wedge y \xrightarrow{w} z$.
- $w = I_p\widetilde{U_q} \in \sum_i \sum_{j\neq i} I_i\widetilde{U_j}$：take expansion $x \xrightarrow{\square\blacksquare'} x \wedge y \xrightarrow{I_p} x \wedge y \xrightarrow{w} z$.
- $w = \#I_p\widetilde{U_q} \in \#\mathbb{I}(\widetilde{U_1} + \cdots + \widetilde{U_\ell})$：take expansion $x \xrightarrow{\square\blacksquare'} x \wedge y \xrightarrow{\#I_p} x \wedge y \xrightarrow{w} z$.
- $w \in \square\,\blacksquare'$：take expansion $x \xrightarrow{\square\blacksquare'} x \wedge y \xrightarrow{\square\#I_i} x \wedge y \xrightarrow{w} z$, for a suitable $I_i \in \mathbb{I}$.
- $w \in \#_\infty\Sigma$：take expansion $x \xrightarrow{\square\blacksquare'} x \wedge y \xrightarrow{\#_\infty} x \wedge y \xrightarrow{w} z$.
- $w = I_p\$_\infty \in \mathbb{I}\$_\infty$：take expansion $x \xrightarrow{\square\blacksquare'} x \wedge y \xrightarrow{I_p} x \wedge y \xrightarrow{w} z$.

注意， $\square\blacksquare' \in K_{dummy} \subseteq K$和 $\clubsuit I_p \in L$，对于 $\clubsuit \in \Sigma + \$ + \$' + \blacksquare$，以及 ε， I_p， $\#I_p$， $\square\#I_i$， $\#_\infty \in L$。对于情形 $w \in M_{\widehat{a}\widehat{I}}$，我们有：

- $w \in \widehat{\mathbb{I}}(\widehat{\Sigma} + \widehat{\$} + \widehat{\$}' + \widehat{\blacksquare})$：take expansion $x \xrightarrow{\widehat{\blacksquare}'\widehat{\square}} x \wedge y \xrightarrow{\varepsilon} x \wedge y \xrightarrow{w} z$.
- $w = \widehat{\#a} \in \widehat{\#}\widehat{\Sigma}$：take expansion $x \xrightarrow{\widehat{\blacksquare}'\widehat{\square}} x \wedge y \xrightarrow{\widehat{\#a}} x \wedge y \xrightarrow{w} z$.
- $w \in \widehat{\mathbb{I}\#}\,(\widehat{\Sigma} + \widehat{\$} + \widehat{\blacksquare})$：take expansion $x \xrightarrow{\widehat{\blacksquare}'\widehat{\square}} x \wedge y \xrightarrow{\varepsilon} x \wedge y \xrightarrow{w} z$.
- $w = \widetilde{V_p}\widehat{I_q} \in \sum_i \sum_{j\neq i} \widetilde{V_j}\widehat{I_i}$：take expansion $x \xrightarrow{\widehat{\blacksquare}'\widehat{\square}} x \wedge y \xrightarrow{\widetilde{V_p}} x \wedge y \xrightarrow{w} z$.
- $w = \widetilde{V_p}\widehat{I_q\#} \in (\widetilde{V_1} + \cdots + \widetilde{V_\ell})\,\widehat{\mathbb{I}\#}$：take expansion $x \xrightarrow{\widehat{\blacksquare}'\widehat{\square}} x \wedge y \xrightarrow{\widetilde{V_p}} x \wedge y \xrightarrow{w} z$.
- $w \in \widehat{\blacksquare}'\widehat{\square}$：take expansion $x \xrightarrow{\widehat{\blacksquare}'\widehat{\square}} x \wedge y \xrightarrow{\widehat{\blacksquare}'\widetilde{V_i}} x \wedge y \xrightarrow{w} z$, for a suitable $\widetilde{V_i} \in (\widetilde{V_1} + \cdots + \widetilde{V_\ell})$.
- $w \in \widehat{\$_\infty}\,\widehat{\mathbb{I}}$：take expansion $x \xrightarrow{\widehat{\blacksquare}'\widehat{\square}} x \wedge y \xrightarrow{\widehat{\$_\infty}} x \wedge y \xrightarrow{w} z$.
- $w = \widehat{a}\widehat{\#_\infty} \in \widehat{\Sigma}\,\widehat{\#_\infty}$：take expansion $x \xrightarrow{\widehat{\blacksquare}'\widehat{\square}} x \wedge y \xrightarrow{\widehat{a}} x \wedge y \xrightarrow{w} z$.

注意，$\widehat{\blacksquare}, \widehat{\square} \in K_{dummy} \subseteq K$和 ε，$\widehat{\ }\#\widehat{a}$，$\widetilde{V_p}$，$\widehat{\blacksquare}\widetilde{V_i}$，$\widehat{\ }\$_\infty$，$\widehat{a} \in L$。对于情形 $w \in M_{\widehat{aa}}$，我们有（$\clubsuit \in \{\widehat{\blacksquare}, \widehat{\blacksquare'}\}$ 和 $\spadesuit \in \{\blacksquare, \blacksquare'\}$分别是在各自情形下合适的符号）：

- $w = \widehat{p}q \in \sum_{a \neq b} \widehat{a}b$: take expansion $x \xrightarrow{\clubsuit\spadesuit} x \wedge y \xrightarrow{\widehat{p}} x \wedge y \xrightarrow{w} z$.
- $w \in \Sigma\,(\widehat{\$} + \widehat{\$'})$: take expansion $x \xrightarrow{\clubsuit\spadesuit} x \wedge y \xrightarrow{\varepsilon} x \wedge y \xrightarrow{w} z$.
- $w = \diamond\widehat{p} \in (\$+\$')\,\widehat{\Sigma}$: take expansion $x \xrightarrow{\clubsuit\spadesuit} x \wedge y \xrightarrow{\diamond\widehat{p}} x \wedge y \xrightarrow{w} z$.
- $w = \widehat{\diamond\widehat{p}}q\diamond \in (\widehat{\$} + \widehat{\$'})\,\widehat{\Sigma}\,\Sigma\,(\$ + \$')$: take expansion $x \xrightarrow{\clubsuit\spadesuit} x \wedge y \xrightarrow{\widehat{\diamond\widehat{p}}} x \wedge y \xrightarrow{w} z$.
- $w = \widehat{\bigstar\bigstar} \in (\widehat{\blacksquare} + \widehat{\blacksquare'})(\blacksquare + \blacksquare')$: take expansion $x \xrightarrow{\clubsuit\spadesuit} x \wedge y \xrightarrow{\widehat{\bigstar\diamond\widehat{a}}} x \wedge y \xrightarrow{w} z$, for suitable $\widehat{\diamond} \in (\widehat{\$} + \widehat{\$'})$ and $\widehat{a} \in \widehat{\Sigma}$.
- $w \in \widehat{\$_\infty}\,\Sigma$: take expansion $x \xrightarrow{\clubsuit\spadesuit} x \wedge y \xrightarrow{\widehat{\$_\infty}} x \wedge y \xrightarrow{w} z$.
- $w = \widehat{p}\widehat{\$_\infty} \in \widehat{\Sigma}\,\widehat{\$_\infty}$: take expansion $x \xrightarrow{\clubsuit\spadesuit} x \wedge y \xrightarrow{\widehat{p}} x \wedge y \xrightarrow{w} z$.

没有 $\widehat{\ }$ 的 $\clubsuit\spadesuit$ 对应于 K_{dummy} 中的词。$\widehat{p}$ 对应于 K 中的词，$\widehat{\$}$ 对应于 L 中的词 $\widehat{\diamond p}$，$\widehat{\bigstar\diamond a}$，$L$ 中的词 $\diamond a$，对于 $\diamond \in (\$ + \$')$、$\widehat{\diamond} \in (\widehat{\ }\$ + \widehat{\ }\$')$，以及 $\widehat{\bigstar} \in (\widehat{\blacksquare} + \widehat{\blacksquare'})$。

对于反向，假设 $Q_2(F)_{\text{原子单射}} \neq \emptyset$。只需证明映射到 F 的 Q_2 的展开不能同时使用 K_{dummy} 和 M_{dummy} 中的词。可以验证，任意展开 $x \xrightarrow{w} x \wedge x \xleftarrow{w'} y \wedge y \xrightarrow{w''} z$ 的映射，其中 $w \in K_{dummy}$、$w' \in L$ 且 $w'' \in M_{dummy}$，都会将 y 映射到一个变量 $\bullet$，使得 $\bullet$ 的所有出边的标签都属于集合

$$\{\square, \mathbb{I}, \#, \widehat{\blacksquare}, \widehat{\blacksquare'}, \widehat{\Sigma}, \widehat{\$}, \widehat{\$'}, \widehat{\square}, \widehat{\mathbb{I}}, \blacksquare, \blacksquare', \Sigma, \#_\infty, \widehat{\#_\infty}, \$_\infty, \widehat{\$_\infty}\}$$

然而，这个符号集合与 M_{dummy} 不相交，因此 z 无法映射到任何变量。□

E 定理 6.2 的完整证明

我们证明，即使右侧的所有语言都具有 $\{w\}$ 且 $|w| \leq 2$ 的形式，包含问题仍然是 Π_2^p-困难的。为此，我们说明如何调整 [15, 定理 4.3]，中 Π_2^p-困难性的证明；该证明表明，在标准语义下，CRPQfin/CQ 包含问题是 Π_2^p-困难的。[1]

我们使用从 ∀∃-QBF 出发的归约。核心思路是使用 Q_1 中的集合 $\{t, f\}$ 来编码真或假。

更准确地说，令

$$\Phi \quad = \quad \forall x_1, \ldots, x_n\ \exists y_1, \ldots, y_\ell\ \varphi(x_1, \ldots, x_n, y_1, \ldots, y_\ell)$$

为一个 ∀∃-QBF 实例，其中 φ 是无量词的且属于 3-CNF。我们构造布尔查询 Q_1 和 Q_2，使得 $Q_1 \subseteq_{\text{原子单射}} Q_2$ 当且仅当 Φ 可满足。

该**查询** Q_1 在图13中进行了示意，构造如下：基础部分是一条长度为 4 的 a-路径。我们在路径的外侧节点上添加 4 个构件 E，并在最内侧添加一个构件 D。围绕 D 构件的 4 个 E 构件为何这样选择，将在讨论 Q_2 时说明。基本上，E-构件会接受所有内容，而 D-构件将确保所选文字的取值为真。

[1] 实际上，这表明即使在左侧只能包含形如 $a_1 + \cdots + a_n$ 的正则表达式的片段中，问题仍然是困难的。从某种意义上说，我们通过为合取查询选择一个原子单射展开来模拟析取。

这些构件也绘于图13中。构件的构造如下。

构造**构件** D，使得根节点对 Φ 中的每个变量都有一条出边，也就是说，共有 $n+\ell$ 条。每条边的标签各不相同，即 $x_1, \ldots, x_n, y_1, \ldots, y_\ell$。在每条 x_i 边之后，我们添加一条 t 边，指向一个不同的节点。从这些节点中的每一个出发，都有一个长度为 2、读取 tf 的环。对于每个 $i \in \{1, \ldots, \ell\}$，我们执行如下操作：添加一条 t 边，指向一个命名为 $y_{i,t}$ 的节点，该节点位于 y_i 边之后；再添加一条标记为 f 的边，指向一个命名为 $y_{i,f}$ 的节点。我们给这些节点命名，是因为在 E 构件中也需要用到它们。不同构件中名称相同的节点实际上是同一个节点。

每个**构件** E的构造方式与 D构件类似。根节点针对 Φ中的每个变量都有一条出边，也就是 $n+\ell$条。每条边的标签各不相同，即 x_1、……、x_n、y_1、……、y_ℓ。在每条 x_i边之后，我们添加一条长度为 2 的路径，其标签依次为 tt-和一条 f边。这些边分别通向不同的节点。在每条 y_i边之后，我们添加一条 t边和一条 f边，分别指向$y_{i,\ t}$和$y_{i,\ f}$。

现在我们解释 Q_2的构造。图13给出了一个示例。对于每个子句 i，**查询** Q_2 都有一个小型有向无环图，它可能与为其他子句构造的有向无环图共享节点（y_k、$_{tf}$）。对于子句 i，我们构造 C_i^1，并添加一条 a边指向构件C_i^2，再从那里添加一条 a边指向构件 C_i^3。

构件 C_i^j表示第 i个子句中的第 j个文字。由于量化布尔公式是 3-CNF，我们有 $j \in \{1、2、3\}$。如果该文字是正变量 x_k，那么 C_i^j是一条标签为 $x_k\,t\,t$的路径。如果它是负变量$\neg x_k$，那么 C_i^j是一条标签为 $x_k f$的路径。如果该文字是正变量y_k，那么 C_i^j是一条标签为 $y_k t$的路径，并且终止于一个我们称为$y_{k、\ tf}$的节点；如果它是负变量 $\neg y_k$，那么 C_i^j也是一条标签为 $y_k f$的路径，并且终止于$y_{k、\ tf}$。

至此，构造完成。下面给出一些直观解释。构件 D通过 $\{t, f\}$（简单）路径来控制哪些变量 x_i被设为真，哪些被设为假。当存在一条 $(x_i\ f)$-路径时，我们就认为它为假，这意味着在 Q_1的 a-inj-扩展中，两个不通过 r相连的节点相等。否则，x_i就设为真。注意，每当 x_i为假时，就无法以单射方式映射任何来自 Q_2编码的子句的路径 $v \xrightarrow{x_i} v' \xrightarrow{t\,t} v''$。反过来，每当 x_i为真时，就无法映射一条 $(x_i\ f)$-路径。因此，视情况而定，我们可以将 C_i^j映射进去，也可以不映射。E构件的构造方式确保每个 C_i^j都能映射进去。Q_2 的 a-inj-扩展决定了应将哪条路径映射到 D中，因此也决定了应验证哪个文字。Q_1 的结构中，两个 E构件各自围绕着 D构件，这有助于将 C_i^1、C_i^2、C_i^3这些子句嵌入 E_1的 a-inj-扩展中每个 i。如果第 i 个子句是 $(x_2 \vee \neg y_1 \vee \neg x_3)$，且我们将 f赋值为 x_2、将 t 赋值为 x_3，那么就可以将 C_i^1、C_i^3嵌入第二个和第三个 E 中，并将 $\neg y_1$嵌入 D中的 $y_1 f$。将 $\neg y_1$嵌入 $y_1 f$，就能在所有构件 E、D 以及 Q_2中的所有子句里，固定 f为 y_1的赋值。同理，对于 Φ中的一个子句 $(x_1 \vee \neg x_4 \vee y_5)$，以及典范模型 G中将 f赋值为 x_1、将 t 赋值为 x_4的赋值，我们可以将 x_1、$\neg x_4$ 嵌入第一个和第二个 E 中，并将 y_5 映射到节点 y_{5t}。

现在我们将证明正确性，即：$Q_1 \subseteq_{\text{原子单射}} Q_2$ 当且仅当 Φ 是可满足的。设 $Q_1 \subseteq_{\text{原子单射}} Q_2$。那么存在一个单射

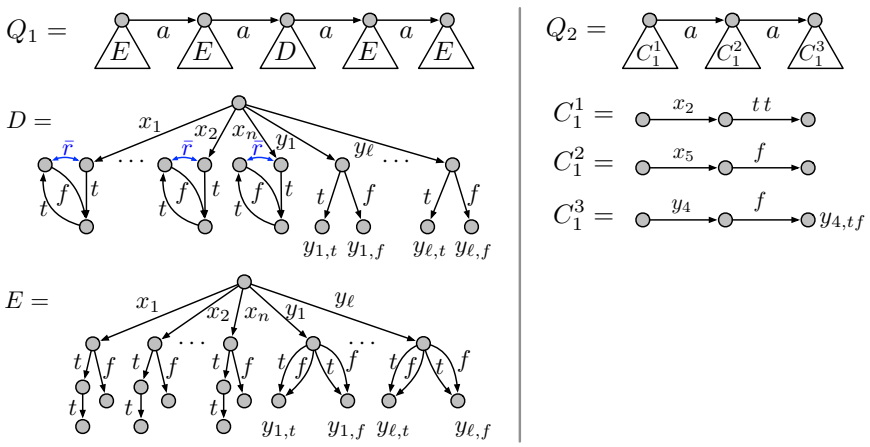

图13：定理 6.2 中使用的查询 Q_1 及其定义中使用的构件 D 和 E。对于 r 边关系（蓝色），我们展示其补关系的边（即，与关系 r 不相关的边）。公式 $\varphi = (x_2 \vee \neg x_5 \vee \neg y_4)$ 的 Q_2 示例。

同态，从 Q_2 到 Q_1的每个 原子单射-展开。 Q_1 的原子单射-展开看起来与 Q_1 完全相同，只是其中一些通过 r 不相关的顶点对可能已被合并。

设 B为 Q_1的任意一个原子单射-扩展， D_B为 B中的构件D。如果存在两个不同的顶点，它们不通过 r关联，且可通过 x_i访问，则定义 $\theta_B(x_i)$为 = 1；否则定义为 $\theta_B(x_i) = 0$。设 h为一个单射同态，将 Q_2的一个原子单射-扩展映射到B。此外，如果 h将y_i、tf 映射到y_i、t，则定义 $\theta_B(y_i)$为 = 1；否则定义为$\theta_B(y_i) = 0$，也就是说，若y_i、tf被映射到y_i、f。现在我们证明 θ_B定义良好且满足 φ。显然，每个 C_i^j都会被映射到构件 D_B或 E，并且对于每个$i \in \{1,\dots,m\}$，恰有一个 C_i^j被映射到 D_B。若 C_i^j对应于 x_k，即它是一个标记为 $x_k t$的路径，那么只有当 $\theta_B(x_k) = 1$时，它才能被映射到 D_B。类似地，若 C_i^j对应于 $\neg x_k$，它是一个标记为 $x_k f$的路径，因此只有当 $\theta_B(x_k) = 0$时，它才能被映射到 D_B。若 C_i^j对应于 y_k 或 $\neg y_k$，它总能被映射到 D_B；但由于y_k、tf只能分别被映射到y_k、t或y_k、f，因此我们可以将正的 y_k映射到 D_B，或将负的映射到 D_B，但不能两者兼得。因此， $\theta_B(y_k)$的定义没有歧义，并且确实满足 φ。

由于 B 是任意的，我们可以据此为 $x_1,\dots,x_n$ 的所有可能真值赋值选取 $y_1,\dots,y_\ell$。因此， Φ 是可满足的。

对于“仅当”方向，设 Φ 是可满足的。于是，对于 $x_1,\dots,x_n$ 的每个真值赋值，我们都能找到 $y_1,\dots,y_\ell$ 的一个赋值，使得 $\varphi(x_1,\dots,x_n,y_1,\dots,y_\ell)$ 为真。设 θ 是一个函数，它根据给定的 x_i，返回所有 y_i 的一个赋值，使公式的求值结果为真。我们将展示如何把 Q_2 映射到 Q_1 的任意一个 原子单射展开 B 中。

给定 B和 θ。再次令 D_B为 B中的构件 D。我们使用 θ为 $y_1,\dots,y_\ell$ 获取真值，方法如下。由于该赋值可满足，每个子句中都有一个文字的取值为真 。我们将这个文字映射到 D_B，并将该子句中的其他文字映射到构件 E。如果这个文字是 x_i，那么我们可以映射到 D_B中的 $(x_i\ t\ t)$-路径。如果它是 $\neg x_i$，那么我们可以映射到 D_B中的 $(x_i\ f)$-路径。如果该文字是 y_i，我们可以将以 $y_{i,tf}$结尾的 $(y_i\ t)$-路径映射到 D_B。这也意味着， Q_2中的每个 $y_{i,tf}$都映射到 $y_{i,t}$，这没有问题，因为映射到 E的每条路径都可以在 $y_{i,t}$之间自由选择以及 $y_{i,f}$；并且由于 θ是一个函数，我们只有 $\theta(y_i) = 1$或 $\theta(y_i) = 0$。类似地，如果该文字是 $\neg y_i$，我们可以将以 $y_{i,tf}$结尾的 $(y_i\ f)$-路径映射到 D_B，这意味着 Q_2中的每个 $y_{i,tf}$都映射到 $y_{i,f}$。□

请注意，在标准语义下，上述定理 6.2 的归约中的右侧查询 Q_2 实际上等价于一个合取查询；但在 原子单射语义下，情况并非如此。

F 其他结果

下述定理总结了其余所有复杂性结果，从而补全了图 1 中的复杂性结果全貌；其证明见下文。

定理 F.1（定理 6.3 的重述）。
（*1*）在查询单射语义下，合取查询/*CRPQ* 和合取查询/合取查询包含问题均为 *NP* 完全。（命题 *F.2*）（*2*）在原子单射语义下，合取查询/合取查询包含问题为 *NP* 完全。（推论 *F.4*）（*3*）在标准语义和原子单射语义下，*CRPQ*/合取查询和 *CRPQfin*/合取查询包含问题均为Π_2^p-困难。（命题 *F.6*）（*4*）在所有语义下，*CRPQ*/合取查询和 *CRPQfin*/合取查询包含问题均属于 Π_2^p。（命题 *F.7*）（*5*）在所有语义下，*CRPQ/CRPQfin* 包含问题为多项式空间-困难。（命题 *F.8*）（*6*）在标准语义下，*CRPQ/CRPQfin* 包含问题属于多项式空间。（命题 *F.9*）（*7*）在所有语义下，*CRPQfin/CRPQ* 包含问题属于 Π_2^p。（命题 *F.10*）

命题 F.2。 合取查询/CRPQ 和合取查询/合取查询 包含问题在查询单射语义下是NP-完全的。

证明。对于上界，设 $Q_1(\bar{x})$ 是一个合取查询， $Q_2(\bar{y})$ 是一个 CRPQ。请记住，对于任意 $\star \in$ {原子单射、查询单射}，都有 $Q_1 \subseteq_\star Q_2$ 当且仅当 $\bar{x} \in Q_2(Q_1)_\star$，其中 Q_1 被视为一个图数据库。因此，根据命题 3.1，我们得到 NP 隶属。

下界可通过从合取查询对应的求值问题直接归约得到，再结合命题 3.1。 □

如果存在某个 $x \xrightarrow{a} y$，使得 A 中的 $x \neq y$ 满足 $h(x) = h(y)$，则称同态 $h\colon A \to B$ 是收缩的。注意，两个非收缩同态的复合仍是非收缩的。

引理 F.3。 对于任意两个合取查询 Q_1、Q_2，以下条件等价：*(1)* 存在一个非收缩同态 $Q_2 \to Q_1$，*(2)* $Q_1 \subseteq$原子单射 Q_2。

证明。从上到下，设 $h\colon Q_2 \to Q_1$ 是一个非收缩同态，也就是说，对于 Q_2 的每个原子 $x \xrightarrow{a} y$，都有 $h(x) \neq h(y)$。令 $E_1 \in \text{Expa-inj}(Q_1)$。注意，存在一个非收缩的 $g\colon Q_1 \to E_1$，因此它们的复合 $g(h)\colon Q_2 \to E_1$ 也非收缩。令 U 为所有满足 $g(h(x)) = g(h(y))$ 的原子 $x = y$ 的合取，并注意到 $E_2 = (Q_2 \wedge U)^{\equiv} \in \text{Expa-inj}(Q_2)$。此外，我们有 $g(h)\colon E_2 \xrightarrow{\text{inj}} E_1$。综上，对于每个 $E_1 \in \text{Expa-inj}(Q_1)$，都存在 $E_2 \in \text{Expa-inj}(Q_2)$，使得 $E_2 \xrightarrow{\text{inj}} E_1$；根据命题 4.6 的刻画，这证明了 $Q_1 \subseteq$原子单射 Q_2。

从下往上观察，$Q_1 \in \text{Exp}_{\text{原子单射}}(Q_1)$，因此根据命题 4.6，存在某个 $E_2 \in \text{Exp}_{\text{原子单射}}(Q_2)$，使得 $h\colon E_2 \xrightarrow{inj} Q_1$，这尤其意味着 h 是非收缩的。另一方面，如前所述，必定存在一个非收缩的同态 $g\colon Q_2 \to E_2$。由于非收缩同态的复合仍是非收缩同态，我们得到 $h(g)\colon Q_2 \to Q_1$ 是非收缩的。□

推论 F.4。 在原子单射语义下，合取查询/合取查询 包含问题是NP完全的。

证明。上界由上述引理 F.3 得出。对于下界，很容易看出，在标准语义下针对合取查询的 3-可着色性标准归约 [10]在此情形下仍然适用。□

推论 F.5. 在简单路径语义下，合取查询/CRPQ(A) 包含问题是NP-完全的。[2]

命题 F.6. 在标准语义和原子单射语义下，CRPQ/合取查询 以及CRPQ$_{\text{fin}}$/合取查询 包含问题是 Π_2^p-困难的。

证明。标准语义下的下界由[15, 定理 4.3]得出。此外，很容易看出，[15, 定理 4.3] 中的 Π_2^p-困难性证明对于原子单射语义同样成立。这是因为，在该归约中，查询的所有原子都包含长度为 1 的词所构成的语言，并且它们没有自环。事实上，根据引理 F.3，在如此严格的条件下，原子单射 与标准包含问题等价。□

命题 F.7. 在所有语义下，CRPQ/合取查询 以及CRPQ$_{\text{fin}}$/合取查询 包含问题属于 Π_2^p。

[2]在 [15], CRPQ(A) 的术语中，是指其正则表达式形式为 $a_1 + \cdots + a_n$ 的 CRPQ。

证明。令 $\star \in \{st, \text{查询单射}, \text{原子单射}\}$。给定 Q_1, Q_2，令 N 为 Q_2中原子的数量。考虑集合 S，其中包含 Q_1的所有展开，但每个大小大于 $2N$的原子展开 $x \xrightarrow{w} y$都替换为 $x \xrightarrow{u} \xrightarrow{\#} \xrightarrow{v} y$，其中 # 是一个新符号，而 u [和 v] 分别是 w的 N 前缀 [和 N 后缀]。注意，S中的每个元素大小都是多项式级别，并且我们可以在多项式时间内检查给定的任意多项式级别 CQ 是否属于 S。考虑以下用于判定非包含的 Σ_2^p 算法。我们检查是否存在 Q_2 的某个连通分量 $\hat{}\, Q_2$ 以及元素 $E_{\#_1} \in S$，使得以下两个条件成立：

(i) $E_{\#_1} \not\subseteq_\star \hat{}\, Q_2$（对于 查询单射，根据命题 F.2，该问题属于 co-NP；对于标准情形，则根据 [15, 定理 4.2] ）；
(ii) 对于与 $E_{\#1}$ 的某个原子 $x \xrightarrow{L} y$相关联的每个原子 $x \xrightarrow{u} \xrightarrow{\#} \xrightarrow{v} y$，不存在满足以下条件的 w： (a) $uwv \in L$，且 (b) $E_1^w \subseteq_\star \hat{}\, Q_2$，其中 $E_1^w() = x \xrightarrow{uwv} y$。

$\hat{}$由于 Q_2 是连通的，因此它必须通过同态[（分别是单射同态、原子单射 同态]）映射到N的某个变量的邻域中，或者完全位于某个原子展开内。上述两项确保这些情况均不可能发生，因此存在一个关于包含$Q_1 \subseteq_\star Q_2$的反例。注意，在第 (ii) 项中，如果 Q_2 的某个 ⋆展开 $\hat{}\,\hat{}$ 映射到一条有向路径上，这意味着 Q_2 与一条有向路径在 ⋆意义下等价，因此 w可以取为多项式大小。进而，这意味着 (ii) 可以在 co-NP 中完成。□

命题 F.8. 在所有语义下，CRPQ/CRPQ$_{\text{fin}}$ 包含问题是 *PSPACE*-困难的。

证明。首先注意，对于形如 $Q_i() = x \xrightarrow{L_i} y$ 的布尔查询 Q_1, Q_2，有 $Q_1 \subseteq_{st} Q_2$ 当且仅当 $Q_1 \subseteq_{\text{原子单射}} Q_2$ 当且仅当 $Q_1 \subseteq_{\text{查询单射}} Q_2$。在 [15, 定理 4.5] 中已证明，在标准语义下，此类查询的包含问题是 PSPACE-困难的，即使 L_2 是无星表达式且字母表大小固定也是如此。□

命题 F.9. 在标准语义下，CRPQ/CRPQ$_{\text{fin}}$ 包含问题属于 多项式空间 。

证明。给定 Q_1, Q_2，令 N为 Q_2的一个展开中原子的最大数量。考虑 Q_1的所有展开 S构成的集合，其中，对于其中每个原子 $x \xrightarrow{L} y$的展开 $x \xrightarrow{w} y$，若$|w| > 2N$，则将其替换为 $u \cdot \# \cdot v$；其中 u [分别是 w的 N 前缀 [和 N 后缀]，# 是一个新符号。随后，PSPACE 算法猜测一个 S 的元素 $E_{\#1}$，并检查是否存在 Q_1 的某个展开 E_1，使得 E可由其得到，且 Q_2 的任何展开 E_2 都不能同态映射到 E_1。$\hat{}$ 为此，我们检查是否存在 Q_2 的某个连通分量 Q_2满足以下两个条件：

(i) $E_{\#1} \not\subseteq \hat{}\, Q_2$（属于 co-NP [15, 定理 4.2]）；
(ii) 对于与 Q_1 的某个原子 $x \xrightarrow{L} y$ 的展开相关联的 $E_{\#1}$ 中每条形如 $x \xrightarrow{u} \xrightarrow{\#} \xrightarrow{v} y$ 的路径，都不存在满足以下条件的 w：(a) $uwv \in L$，且 (b) $E_1^w \subseteq \hat{}\, Q_2$，其中 $E_1^w() = x \xrightarrow{uwv} y$。

ˆ由于 Q_2 是连通的，它必须被映射到 Q_1 的某个变量的 N-邻域中（由条目 i 排除），或者完全位于某个原子展开内部（由条目 ii 排除）。另一方面，对于某个分量 ˆ Q_2，$E_1 \not\subseteq Q_2$ 当且仅当 $E_1 \not\subseteq \hat{Q}_2$。因此，这两个条目确保这些情况均不会发生，并且存在一个针对包含关系 $Q_1 \subseteq Q_2$ 的反例。注意，条目 (ii) 可以看作是正则语言的交集判空问题的一个实例，也就是判断是否 $\bigcap_{i\in I} L_i = \emptyset$ 对于给定的一组正则表达式，这是一个 PSPACE 完全问题 [21]。□

命题 F.10。 在所有语义下，CRPQfin/CRPQ 包含问题属于 Π_2^p。

证明。令 $\star \in \{st$，原子单射，查询单射$\}$，并令 $Q_1(\bar{x})$ $Q_2(\bar{x})$ 为一个实例。可以通过猜测一个 $\star$-展开 $E_1(\bar{x}) \in \mathrm{Exp}_\star(Q_1(\bar{x}))$（线性大小）来测试非包含，并测试 $\bar{x} \notin Q_2(E_1)$ 在 $\star$ 语义下是否成立；根据命题 3.1，这属于 co-NP。□