arXiv:2512.18861v1 [math.DS] 21 Dec 2025

合并 关于将工作空间视为霍普夫代数马尔可夫链

Matilde Marcolli 和 David Skigin

Abstract. 我们研究了一个霍普夫代数马尔可夫链的动力学性质，其状态空间是叶子带有标号的二叉有根森林。根据乔姆斯基生成语言学最简主义模型，这一马尔可夫动力系统通过合并操作，描述了句法中结构形成与转换的核心计算过程。该动力学可分解为一个由内部合并作用所决定、具有均匀平稳分布的遍历动力系统；而外部合并和（最简形式的）侧向合并所带来的贡献，则可化简为一个以划分集合为状态空间、带有组合权重的马尔可夫链。侧向合并部分的动力学会阻碍系统收敛到完全形成的连通结构（树），除非通过成本函数为不同形式的合并赋予权重；这与语言学理论的预测相符。对于加权情形，我们借助热带半环中一个相关的 Perron-Frobenius 问题，得到了关于 Perron-Frobenius 特征值和特征向量渐近行为的结果。这些结果表明，语言学文献中提出的常见成本函数（最小搜索和资源限制）不足以使系统收敛到树结构，而基于香农熵的额外优化性质则能使动力学达到预期结果。我们还讨论了引入与语义嵌入及其他计算模型相关的连续参数，以及通过着色规则对动力学进行某种过滤；这些规则模拟了由 θ角色和阶段结构所施加的语言学过滤。最后，我们也讨论了外化中的参数变异与参数设定过程。

Contents

1. 引言 2 1.1. 工作空间上的合并作用概述 3 2. 合并图 6 2.1. 森林计数 6 2.2. 合并图中的入度和出度 9 2.3. 稀疏性 13 2.4. 强连通性 13 3. 随机游走与霍普夫代数马尔可夫链 14 3.1. 合并图上的随机游走 15 3.2. 霍普夫代数马尔可夫链 16 3.3. 作为最大熵随机游走的霍普夫代数马尔可夫链 17 3.4. 自由能优化 19 3.5. 熵优化的语言学意义 23 4. 霍普夫代数马尔可夫链的分解 24 5. 内部合并动力学 26 5.1. 连通分量 26 5.2. 分量的强连通性 27 5.3. 均匀分布 29 6. 内部与外部合并动力学 30 6.1. 可约性与暂态 30

日期：2025 年。

6.2. 通信类 31 6.3. 闭通信类上的平稳分布 32 6.4. IM–EM结构形成模型：语言学方面 32 7. 内部与侧向合并动力学 33 8. 通过杨图研究外部与侧向合并动力学 35 8.1. 杨图图 35 8.2. 强连通性 36 8.3. 杨氏表图 38 9. 带最小侧向合并的完整合并动力学 39 9.1. 马尔可夫链投影 40 9.2. 工作空间上合并动力学的平稳分布 42 9.3. $n = 4$ 43的例子9.4. 侧向合并问题 46 10. 带收缩余乘的合并动力学 46 10.1. 移位轨迹的标签 47 10.2. 带收缩余乘的合并操作 48 10.3. 带收缩余乘的合并图的边 49 10.4. 瞬态动力学与投影 50 11. 加权合并动力学：代价函数 52 11.1. 组合代价函数 54 12. 热带 Perron-Frobenius 与特征向量估计 56 12.1. 热带半环与 Perron-Frobenius 问题 57 12.2. 最优树形图与佩龙–弗罗贝尼乌斯特征值 58 12.3. 热带 Perron-Frobenius 特征空间 60 12.4. 关于代价函数的优化 61 12.5. 加权动力学中的侧向合并问题 61 13. 香农熵优化与合并动力学 62 13.1. 香农熵代价函数 63 14. 立方体与动力系统的参数化 66 15. I-语言约束与彩色Merge动力学 72 15.1. 彩色歌剧与芽生成系统 73 15.2. 通过过滤构建有色合并图 73 15.3. 通过动力学构建构建有色合并图 74 16. 参数设定与外化 76 致谢 79 参考文献 79

1. Introduction

在 Diaconis、Pang 和 Ram 于 [14] 以及 Pang 于 [27], [28], 的研究中，发展出了一套霍普夫代数马尔可夫链理论。这些离散动力系统与组合霍普夫代数相关。这些霍普夫代数（见 [19]）是分次连通交换霍普夫代数$\mathcal{H}$，其基由组合对象构成（在本文中，这些对象将是二叉森林）。称为霍普夫代数马尔可夫链的动力系统作用于该霍普夫代数的向量空间，其形式为以下复合

$$\mathcal{K} = \mu \circ \mathcal{L} \circ \Delta \,, \tag{1.1}$$

其中， μ 是霍普夫代数的乘法， Δ 是霍普夫代数的余积，而 $\mathcal{L}$ 是 $\mathcal{H} \otimes \mathcal{H}$的线性自同态。换言之，先利用余积Δ将组合对象分解，再对分解得到的各部分施加 $\mathcal{L}$ ，最后利用乘法 μ重新组合出一个对象。迄今研究的主要例子中， $\mathcal{L}$ 是一个投影。文献 [14], [27], [28] 还研究了更一般的形式：在复合（1.1）中使用幂 Δ^k 和 μ^k ，也就是将对象分解为多个部分。就本文而言，我们只考虑（1.1）这一形式，其中 $\mathcal{L}$ 不仅是一个投影，还是投影与“嫁接”操作的复合。虽然在文献 [14], [27], [28] 研究的情形中， $\mathcal{L}$的特殊形式使得平稳分布可以用组合方式描述，但我们考虑的情形并不具备这种结构；不过，基于佩龙–弗罗贝尼乌斯定理以及对组合对象的显式计数（涉及划分和广义多项式系数），我们仍能推导出关于平稳分布和动力学的一些一般结果。

我们所考察的霍普夫代数马尔可夫链的形式，受到语言学中合并操作数学模型的启发；该模型由第一作者、乔姆斯基和 Berwick 在 [22]的研究中发展而来。

生成语言学中的句法最简方案由诺姆·乔姆斯基于20世纪90年代初首次提出， [6]。此后，该方案经历了重大发展，并最终形成了近期的表述：自2013年前后开始，乔姆斯基将“自由对称合并”引入为句法的基本结构构建操作，并将这一普遍的核心计算机制单独提取出来，使其受到最优性约束的制约， [7], [8], [9]。在这一模型中，核心合并作用解释了句法的全部结构形成过程和关键计算方面（组合性和转换性），而“外化接口”则解释了不同语言之间的参数化句法变异以及具体语言中的实现；“句法-语义接口”则解释了语义解析。2023年，Marcolli、乔姆斯基和 Berwick 的合作研究表明，自由对称合并的核心计算机制可以通过霍普夫代数和霍普夫代数马尔可夫链动力系统得到精确的数学形式化， [22]。该研究 [22], 以及后续研究 [26], [24], [25], [21], [31] 和正在开展的进一步研究表明，两个接口（外化接口和句法-语义接口）同样可以得到良好的数学建模，不过，要对其运作方式作出完整的数学说明，仍需进行大量进一步研究。

本文聚焦于 [22], 所提出的自由对称合并的核心计算结构，并进一步探究其作为作用于句法工作空间的动力系统所具有的数学性质。具体而言，我们将霍普夫代数马尔可夫链视角下的句法形成与转换的普适动力学性质作为研究对象。

目前，人们特别关注语言结构形成的精确数学模型，因为模拟人类语言的现有人工系统（例如大语言模型）在训练和计算资源方面都极为低效，而且完全基于统计数据（计算文本字符串扩展的概率），却完全忽视了我们对语言结构及其底层计算机制所掌握的大量细致知识。人工智能领域近期发展中显而易见的惊人理论鸿沟，使得用精确的数学语言表述现有的语言生成科学理论——广为人知的“生成语法”——成为一项紧迫任务。尽管应用并非本文及其他理论论文的重点，但它们是发展句法作为计算系统的数学模型的自然动机之一。

1.1. 工作空间上合并作用的概述。我们在此回顾一些基本结果和记号，这些内容将在全文中使用。

• $\mathcal{SO}_0$ 是一个有限集合。元素 $\alpha \in \mathcal{SO}_0$ 称为词汇项和句法特征（为简洁起见，我们在此统称为“词汇数据”）。

• $\mathcal{SO} = \mathfrak{T}_{\mathcal{SO}_0}$ 是由满二叉有根树组成的可数无限集（非平面的，即没有指定的平面嵌入），其叶子由集合 $\mathcal{SO}_0$中的元素装饰。元素 $T \in \mathcal{SO}$ 称为句法对象。我们将 $\mathfrak{T}_{\mathcal{SO}_0,\, n}$ 记作具有 $n = \#L(T)$ 个叶子的句法对象 $T \in \mathcal{SO}$ 的集合。

• $\mathcal{SO}$ 是由集合 $\mathcal{SO}$生成的自由非结合交换幺半群

$$\mathcal{SO} = \mathrm{Magma}_{c,na}(\mathcal{SO}_0, \mathfrak{M}) \cong \mathfrak{T}_{\mathcal{SO}_0},$$

其二元运算 $\mathfrak{M}$ 作用于一对树上，将它们连接到一个公共根（作为非平面树），

$$\mathfrak{M}(T, T') := \widehat{T\ \ T'} = \widehat{T'\ \ T}.$$

• $\mathfrak{F}_{\mathcal{SO}_0}$ 是由二叉有根森林组成的可数无限集，即其连通分量是 $\mathfrak{T}_{\mathcal{SO}_0}$中的非平面满二叉有根树的森林。元素 $F \in \mathfrak{F}_{\mathcal{SO}_0}$ 称为工作空间。我们也将 $\mathfrak{F}_{\mathcal{SO}_0,\, n}$ 记作具有$n = \#L(F) = \#L(T_1) + \cdots + \#L(T_r)$ 个叶子的工作空间 $F = T_1 \sqcup \cdots \sqcup T_r \in \mathfrak{F}_{\mathcal{SO}_0}$的集合。

• 句法对象的可访问项 $T_v \subset T$ 是由一个非根顶点 v 及其所有后代构成的子树。因此， T 的可访问项集合可与 T的非根顶点集合相对应。

• 向量空间 $\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})$ 配备由不交并 $\sqcup$给出的乘法后，便是由$\mathfrak{T}_{\mathcal{SO}_0}$生成的自由交换代数（多项式代数）。

• $\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})$上存在不同形式的余积，它们具有略微不同的代数性质（参见 [22]第 1.2 节中的讨论）。这些余积的形式均为

$$\Delta(T) = \sum_{\underline{v}} F_{\underline{v}} \otimes T/F_{\underline{v}}, \tag{1.2}$$

其中，$F_{\underline{v}} = T_{v_1} \sqcup \cdots \sqcup T_{v_n}$是 T的一组两两不交的可访问项，而 T/F_v是由此得到的商。它们的区别在于获得商项 T/F_v的方式不同。这里我们将重点讨论余积的一种形式，称为 [22],中的“删除余乘” Δ^d ：其中，商 T/F_v 取为通过边收缩（消去非分支顶点）从 T 沿每个非根顶点 v_i上方的边切开后得到的（非满）二叉树所获得的最大满二叉树。余积的另一种形式称为 [22] 中的“收缩余乘”，它采用一种商 T/F_v ，其中森林 F_v 的每个分量都收缩到其根顶点。（我们只会在第 10 节讨论收缩余乘。）

除了商 T/F_v 的不同形式外，余积的不同形式也可能因提取可访问项 T_{v_i} 时所采用的条件（可容许切割的类型）而异。本文重点讨论的删除余乘中，余积的某些代数性质会随我们是否允许任意可容许切割而改变：这类切割提取出森林 F_v （这里的“可容许”是指切割中的任意两条边都不位于从根到某个叶子的同一路径上，因此 F_v 确实由可访问项组成）；另一种做法则要求满足更严格的条件，即列表$v = (v_1, \ldots\ldots, v_n)$ 中不包含任何一对姐妹顶点（绝不切割同一顶点下方的两条边）。本文不讨论这两种选择在代数性质层面的差异，因为这一问题将在其他地方分析。不过，我们会简要讨论这两种不同的可容许切割选择如何影响动力系统的性质，以及从这一角度看，为何最好允许同时切割同一顶点下方的两条边。

- 工作空间上的合并操作是由作用于基元素 F 的线性算子生成、定义在 $\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})$ 上的动力系统

$$\mathcal{K}_2(F) = \sqcup \circ (\mathfrak{B} \otimes \mathrm{id}) \circ \Pi^{(2)} \circ \Delta \quad \text{and} \quad \mathcal{K}_1(F) = \sqcup \circ \Pi^{(1)} \circ \Delta\,, \tag{1.3}$$

其中，$\mathfrak{B}$ 是从森林到树的嫁接算子 $\mathfrak{B}(T \sqcup T') = \mathfrak{M}(T,\ T')$，而 $\Pi^{(k)}$ 是作用于 $\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0}) \otimes \mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})$ 上的线性投影：它在基元素 $F \otimes F'$ 上等于恒等映射，其中 $F = T_1 \sqcup \cdots \sqcup T_k$ 有 k 个分量，其他情况下为零。我们可以将其写成一个单独的算子

$$\tilde{\mathcal{K}}(F) = \mathcal{K}_2(F) + \mathcal{K}_1(F) = \sqcup \circ (\mathfrak{B} \otimes \mathrm{id}) \circ \tilde{\Pi}^{(2)} \circ \Delta\,, \tag{1.4}$$

其中 $\tilde{\Pi}^{(k)} = \oplus_{j=1,\ \ldots,\ k} \Pi^{(j)}$；对于单棵树 T，我们将 $\mathcal{B}$ 扩展为 $\mathcal{B}(T) = T$，因此（1.4）中的最后一个表达式确实等于 $\mathcal{K}_2(F) + \mathcal{K}_1(F)$。

虽然从数学视角来看，（1.4）是表示 Merge 动力系统最自然的方式，因为它恰好具有霍普夫代数马尔可夫链的形式（1.1），但为了更贴切地体现其语言学解释，最好将（1.3）中的算子 $\mathcal{K}$ 拆分为一个和

$$\mathcal{K}_2 = \sum_{S,S'} \mathfrak{M}_{S,S'}$$

关于 $S,\ S' \in \mathcal{SO}$，其中

$$\mathfrak{M}_{S,S'} = \sqcup \circ (\mathfrak{B} \otimes \mathrm{id}) \circ \delta_{S,S'} \circ \Delta\,, \tag{1.5}$$

其中 $\delta_{S,\ S'}$ 克罗内克δ算子在 Δ 取值范围内、且在余积左侧具有 $S \sqcup S'$ 的项上取恒等映射（否则取零）。对工作空间 F 应用时，关于 $S,\ S'$ 的求和总会归结为有限求和。这些运算 $\mathfrak{M}_{S,\ S'}$ 分别对应语言学上的外部合并（$\mathfrak{M}_{S,\ S'}$ 工作空间的两个连通分量）和侧向合并（$\mathfrak{M}_{S,\ S},\ S'$ 中至少有一个是可访问项）。

式 (1.3) 中的算子 $\mathcal{K}_1$ 也可以类似地写成 $\mathcal{K}_1 = \sum_S \mathfrak{M}_{S,\ 1}$ 的形式，其中 1 是幺半群 $\mathcal{SO}$（形式上的空树）的单位元。这些算子 $\mathfrak{M}_{S,\ 1}$ 本身没有直接的语言学解释（可以称它们为虚拟合并），但它们引出了一种非常重要的运算，用来解释句法中的“移位”或“转换”，即内部合并；这里它被表示为复合 $\mathfrak{M}_{S,\ T/S} \circ \mathfrak{M}_{S,\ 1}$。

为了使我们在此处的表述尽可能贴近语言学解释，我们将用外部合并（EM）、内部合并（IM）和侧向合并（SM）操作来描述合并马尔可夫链 $\mathcal{K}$，而不是像（1.3）和（1.4）那样用 EM、SM 和虚拟合并（vM）来描述，从而确保合并图中描述马尔可夫链动力学的每个箭头都有语言学解释。我们提出的所有论证都可以轻松推广到（1.4）的情形。

- 式 (1.4) 所示 Merge 动力系统一条霍普夫代数马尔可夫链，其含义见 [14], [27], [28]。

- 各个合并操作 $\mathfrak{M}_{S,\ S'}$ 和 $\mathfrak{M}_{S,\ T/S} \circ \mathfrak{M}_{S,\ 1}$ 可以根据特定的成本函数（最小搜索、资源限制）赋予权重，具体讨论见 [22] 和 [26]。
- 除了由合并操作 $\mathfrak{M}_{S,S'}$ 和 $\mathfrak{M}_{S,T/S} \circ \mathfrak{M}_{S,1}$所描述的句法核心计算机制外，还可以考虑一种过滤系统，作用于工作空间上的合并作用所构造的句法对象，从而消除不适用于语义解释的结构。主要有两种此类过滤器，分别在 [25] 和 [24], 中进行分析：一种处理 θ角色指派，另一种处理阶段结构。正如 [25] 和 [24] 所示，这些过滤器可以用以下方式建模：

彩色歌剧配备了 [17], 意义上的芽生成系统，也可以等价地通过根据着色规则剪除合并操作来实现。从我们在此考察的动力系统的角度来看，这种剪除相当于从马尔可夫链动力学的有向图中删除一些箭头，这些箭头对应于与着色限制不兼容的合并操作。我们将在§15中讨论这一点。

我们将其视为本文所分析的动力系统的基本结构。从仅由带标签的叶顶点组成且没有边的森林出发，Merge 在工作空间上的作用轨道对应于句子构造的语言推导：通过反复应用合并操作，从无结构的词汇项集合构建句法结构。

2. The Merge graphs

Merge 在工作空间上的作用所定义的动力系统，可以用一个有向图 $\mathcal{G}$ 来描述，其顶点是边集非空的森林 $F \in \mathfrak{F}_{\mathcal{SO}_0}$， $E(F) \neq \emptyset$。图 $\mathcal{G}$ 在顶点 F 与顶点 F' 之间有一条边，当且仅当存在一个合并操作，其作用为$F' = \mathfrak{M}_{S,\ S'}(F)$ 或 $F' = \mathfrak{M}_{S,\ T/S}(\mathfrak{M}_{S,\ 1}(F))$。

我们只考虑边集非平凡的森林 F，原因在于每个合并操作（即使作用于由没有边的顶点构成的不交并）都会产生一棵含有若干边的森林。因此，仅由叶子组成且没有边的森林对于这种动力学而言是暂态的，动力学不会返回这些森林，所以它们不会影响动力学的主要性质，例如平稳分布、环等。

由于所有合并操作（采用这种余积形式）都保持森林的叶子数量不变，因此无限图 $\mathcal{G}$ 可分解为有限图的不交并

$$\mathcal{G} = \bigsqcup_{n\geq 2} \mathcal{G}_n \,.$$

（此处不包括 $n = 1$ 这种情况，因为那里没有动力学。）此外，合并操作不会改变叶子的标记，因此每个图 $\mathcal{G}_n$ 本身都是所有此类标记的不交并

$$\mathcal{G}_n = \bigsqcup_{A\in \mathrm{Sym}^n(\mathcal{SO}_0)} \mathcal{G}_{n,A} \,.$$

集合 $\mathcal{G}_{n,\ A}$ 的顶点是森林 $F \in \mathfrak{F}_{\mathcal{SO}_0}$ 的有限子集 $\mathfrak{F}_{A,\ n} \subset \mathfrak{F}_{\mathcal{SO}_0}$，其中 $n = \#L(F)$，对于集合 $L(F)$ 中的叶子，标签集合为 $\{\alpha_\ell\}_{\ell\in L(F)} = A$，且边的集合 $E(F)$ 非空。

在我们的分析中，我们将限定于一个固定的 $n \geq 2$ 和一个固定的词汇项集合 $A = \{\alpha_1, \dots, \alpha_n\}$，因此每次只研究一个 $\mathcal{G}_{n,\ A}$ 图。

2.1. 森林计数。 要理解合并图 $\mathcal{G}_{n,A}$的结构，第一步是找到一个好的计数公式来计算顶点数。顶点就是工作空间，也就是由非平面满二叉根树构成的森林，这些树的叶子由 α_i 的 A元素装饰。

为了计算顶点数，也就是计算森林 $F \in \mathfrak{F}_{A,n}$的数量，我们首先根据森林 $F = T_1 \sqcup \cdots \sqcup T_r$ 的叶子集合 $L(F)$ 如何划分为 F中树分量 T_i 的叶子集合 $L(T_i)$，将集合 $\mathfrak{F}_{A,n}$ 划分为若干子集。

令 $\mathcal{P}(n)$ 表示将 n 表示为若干个满足 $k_i \geq 1$ 的整数之和所得的所有划分的集合。对于一个划分 $n = k_1 + \cdots + k_r$，我们将其记作 $\wp = \{k_1, \dots, k_r\}$ （其中 k_i 可以有重数）。

我们将 $n=1+\cdot\dot{.}\cdot+1$ 划分为若干个 **1** 之和的划分记作 $\wp_{1^n}=\{1^n\}=\{1,\ldots,1\}$，并令

$$\mathcal{P}'(n):=\mathcal{P}(n)\smallsetminus\{\wp_{1^n}\}\,. \tag{2.1}$$

是所有满足以下条件的划分 $n=k_1+\cdots+k_r$ 的集合：至少有一个 k_i 满足 $k_i\geq 2$。

引理 2.1. 有向图 $\mathcal{G}_{n,\ A}$ 的顶点集 $V(\mathcal{G}_{n,\ A})=\mathfrak{F}_{A,\ n}$ 可划分为

$$V(\mathcal{G}_{n,A})=\bigsqcup_{\wp\in\mathcal{P}'(n)}V_{\wp,n,A}\,, \tag{2.2}$$

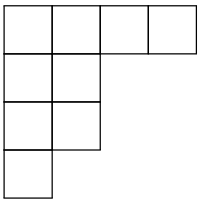
其中，对于 $\wp=\{k_1,\ldots,k_r\}$，若 $n=k_1+\cdots+k_r$，集合 $V_{\wp,\ n,\ A}$ 由 $\mathcal{G}_{n,\ A}$ 的所有顶点 F 构成，这些顶点是形如 $F=T_1\sqcup\cdots\sqcup T_r$ 且 $T_i\in\mathfrak{T}_{\mathcal{SO}_{0,\ k_i}}$ 的森林，并且边集非空。

证明。给定 $\mathfrak{F}_{A,\ n}$ 中的一个工作空间 $F=T_1\sqcup\cdots\sqcup T_r$，其有 n 个叶子，标签为 $A=\{\alpha_1,\ldots,\alpha_n\}$，叶子集合 $L(F)$ 也可分解为一个不交并

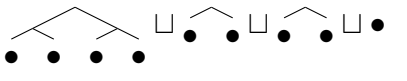
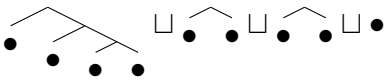
$$L(F)=L(T_1)\sqcup\cdots\sqcup L(T_r) \tag{2.3}$$

其中 $L(T_i)$ 是由树 T_i 给出的 F 的连通分量的叶子集合。若 $k_i=\#L(T_i)$，则得到一个形如 $\wp\in\mathcal{P}(n)$ 的划分，其中 $n=\#L(F)=k_1+\cdots+k_r$。因此，我们可以将集合 $\mathfrak{F}_{A,\ n}$ 分成互不相交的子集 $V_{\wp,\ n,\ A}$，对应于 $\wp\in\mathcal{P}(n)$，其中 $F\in V_{\wp,\ n,\ A}$ 是森林，其叶子按照划分 $\wp$ 在 (2.3) 中分开。边集非空的森林 $E(F)\neq\emptyset$ 对应于 $\mathcal{P}'(n)$ 中的划分。□

划分 $\wp=\{k_1,\ldots,k_r\}\in\mathcal{P}(n)$ 可以表示为杨图，由 n 个方格组成，这些方格排列成长度为 k_i 的 r 行（按非递减顺序绘制）。例如，我们有

$\wp=\{4,2,2,1\}\in\mathcal{P}(9)\Leftrightarrow D(\wp)=$  .

对于不同的工作空间，此划分是图像 $\wp=p(F)$

$F=$  or $F=$ 

其中叶部的 $\bullet$ 个槽位可以用 $A=\{\alpha_1,\ldots,\alpha_9\}$中各元素的一个排列来填充。特别地，杨图描述了叶子在工作空间不同分量之间的划分，但无法表示这些分量的不同树拓扑。为了跟踪叶部标签 α_i 的各种可能分配，我们还可以考虑以 A为字母表的杨表，例如

α_{i_1}	α_{i_2}	α_{i_3}	α_{i_4}
α_{i_5}	α_{i_6}		
α_{i_7}	α_{i_8}		
α_{i_9}			

其中 α_{i_1}，……，α_{i_9}是 $A=\{\alpha_1,\ldots,\alpha_9\}$的一个排列。

对于划分 $\wp=\{k_1,\ldots,k_m\}\in\mathcal{P}(n)$，多项式系数

$$\mu_{\wp,n}:=\binom{n}{k_1,\ldots,k_r}=\frac{n!}{k_1!\cdots k_m!} \tag{2.4}$$

是对称群 S_n 中由该划分对应的杨群所确定的陪集 $S_n/S_\wp$ 的大小，

$$S_\wp := S_{k_1} \times \cdots \times S_{k_r} \tag{2.5}$$

其中 k_i 按照杨图 $D(\wp)$中的顺序排列。注意，在（2.4）中，我们列出了带有重数的 k_i 。明确列出重数为 $n = a_1 k_1 + \cdots a_r k_r$ 时，我们写作

$$\mu_{\wp,n} := \binom{n}{\underbrace{k_1,\dots,k_1}_{a_1\text{-times}} \dots \underbrace{k_r,\dots,k_r}_{a_r\text{-times}}} = \frac{n!}{(k_1!)^{a_1} \cdots (k_r!)^{a_r}} . \tag{2.6}$$

该划分的广义多项式系数为

$$\wp = \{\underbrace{k_1,\dots,k_1}_{a_1},\dots,\underbrace{k_r,\dots,k_r}_{a_r}\} \quad \text{with} \quad n = \underbrace{k_1+\cdots+k_1}_{a_1\text{-times}} + \cdots + \underbrace{k_r+\cdots+k_r}_{a_r\text{-times}} \tag{2.7}$$

其表示为

$$\Upsilon_{\wp,n} := \frac{1}{a_1! \cdots a_r!} \binom{n}{\underbrace{k_1,\dots,k_1}_{a_1\text{-times}} \dots \underbrace{k_r,\dots,k_r}_{a_r\text{-times}}} = \frac{1}{a_1! \cdots a_r!} \ \frac{n!}{(k_1!)^{a_1} \cdots (k_r!)^{a_r}} . \tag{2.8}$$

我们还考虑以下量

$$\Lambda_{\wp,n} := \frac{1}{a_1! \cdots a_r!} \binom{n}{\underbrace{k_1,\dots,k_1}_{a_1\text{-times}} \dots \underbrace{k_r,\dots,k_r}_{a_r\text{-times}}} \prod_{i=1}^{r} ((2k_i-3)!!)^{a_i} \tag{2.9}$$

我们也可以将（2.9）等价地写成

$$\Lambda_{\wp,n} = \mu_{\wp,n} \cdot \Gamma_{\wp,n} \tag{2.10}$$

其中 $\mu_{\wp,n}$ 是多项式系数 (2.6)，并且

$$\Gamma_{\wp,n} := \frac{1}{a_1! \cdots a_r!} \prod_{i=1}^{r} ((2k_i-3)!!)^{a_i} . \tag{2.11}$$

引理 2.2. $V(\mathcal{G}_{n,A}) = \mathfrak{F}_{A,n}$ 中的顶点数由下式给出

$$\Lambda_n := \#\mathfrak{F}_{A,n} = \sum_{\wp=\{k_1,\dots,k_r\}\in\mathcal{P}'(n)} \Lambda_{\wp,n} , \tag{2.12}$$

其中 $\Lambda_{\wp,n}$ 如 (2.9) 所示，对 *(2.1)* 的集合 $\mathcal{P}'(n)$ 求和。

证明. </style} n 个带标签叶子的非平面满二叉根树的数量由奇双阶乘给出

$$\#\mathfrak{T}_{A,n} = (2n-3)!! . \tag{2.13}$$

因此，形如 $F = T_1 \sqcup \cdots \sqcup T_r$ 的森林的数量，其中 $T_i \in \mathfrak{T}_{A,k_i}$ 和 $k_1 + \cdots + k_r = n$，由下式给出

$$\#\{F \in \mathfrak{F}_{A,n} \,|\, F = T_1 \sqcup \cdots \sqcup T_r,\ T_i \in \mathfrak{T}_{A,k_i}\} = \Upsilon_{\wp,n} \prod_{i=1}^{r} ((2k_i-3)!!)^{a_i} \tag{2.14}$$

其中，多项式系数反映了标签在各分量之间的分配方式，而除以因子 $a_1! \cdots a_r!$ 的归一化则考虑了

划分的杨图中相同长度的行互换所产生的对称性。因此，$\mathfrak{F}_{A,\ n}$ 中的森林总数是一个和

$$\#\mathfrak{F}_{A,n} = \Lambda_n = \sum_{\wp\in\mathcal{P}'(n)} \Lambda_{\wp,n}\,, \tag{2.15}$$

与（2.9）中一样，$\Lambda_{\wp,\ n}$，其中 $\Lambda_{\wp,\ n}$ =#$V_{\wp,\ n,\ A}$。 □

注意，作为比较，不带装饰叶子的非平面满二叉有根树集合的计数要复杂得多。它由第 n 个韦德伯恩–埃瑟林顿数W_n给出，该数没有闭式公式，但其生成函数是以下方程的解

$$\mathcal{T}(x) = x + \frac{1}{2}\mathcal{T}(x)^2 + \frac{1}{2}\mathcal{T}(x^2)\,.$$

相应地，二叉有根森林（叶子无标记）的生成函数为

$$\mathcal{F}(x) = \exp(\mathcal{T}(x))\,.$$

2.2. 合并图中的入度和出度。 我们将把图 $\mathcal{G}$ 及其分量 $\mathcal{G}_{n,\ A}$ 的边视为合并变换，并将其细分为外部合并、内部合并和侧向合并箭头。对于内部合并，我们不考虑提取的可访问项是分量树 $T = \mathfrak{M}(T_1,\ T_2)$ 根下的两棵树 T_1，T_2 之一的情况，因为按照我们采用的余积形式，对 T_1 或 T_2 进行内部合并只会重现 T，而我们不将这类恒等映射纳入动力学中。对于侧向合并，我们将只考虑“最小”的侧向合并变换（其含义见 [26]），即仅提取并合并原子分量（单叶）的变换，因为按照 [22], [26]中的代价计数，这些是代价最低的侧向合并。

我们首先为方便起见，引入一些记号。给定一个工作空间 F，记：

- $\pi_0(F)$ 表示 F 的连通分量集合，其中 $\#\pi_0(F) = b_0(F)$,
- $\pi_{0,\ E}(F)$ 表示 F 的连通分量子集，其中每个分量都是边集非平凡的树；并且 $c(F) := \#\pi_{0,\ E}(F)$,
- $\mathcal{C}(F) \subset L(F) \times L(F)$ 表示 F 中属于 F 内某棵樱桃树的叶子对 $(\ell,\ \ell')$ 的子集（子树 $T_v \subset F$ 由两片姐妹叶子 ℓ，ℓ'、位于它们上方的一个顶点 v 以及两条边组成），其中 $d(F) = \#\mathcal{C}(F)$。
- $\mathcal{C}'(F) = \mathcal{C}(F) \cap \pi_0(F)$ 表示 F 中作为连通分量的樱桃集合（而非分量的可访问项），其中 $d'(F) = \#\mathcal{C}'(F)$。
- $\tilde{c}(F) = c(F) - d'(F)$ 表示具有至少 3 个叶子的 F 连通分量的数量。
- $d''(F) = d(F) - d'(F)$ 表示作为分量（而非完整分量）可访问项的樱桃数量,
- $\mathcal{C}(T_i)$ 表示分量 T_i 的樱桃子树集合，其中 $d(T_i) = \#\mathcal{C}(T_i)$。

我们可以按以下方式描述图 $\mathcal{G}_{n,\ A}$ 中的边。

命题 2.3。 图$\mathcal{G}_{n\text{的边}A}$（顶点集为$\mathfrak{F}_{A,\ n}$）可按以下方式描述。考虑顶点 F，它属于 $\mathcal{G}_{n,\ A}$ 其中有 $b_0(F) = r$ 个连通分量，分别为 $F = T_1 \sqcup \cdots \sqcup T_r$；这些分量分别有 $k_i = \#L(T_i)$ 个叶子，且满足 $\sum_i k_i = n$。不失一般性，假设T_1、……、$T_{c(F)}$是边集非空的分量，其中 T_1、……、$T_{d'(F)}$是樱桃分量。

(1) 顶点 F 处的出射外部合并（*EM*）箭头数等于

$$N^{\rm out}_{\rm EM}(F) = \begin{cases} \binom{r}{2} & r = b_0(F) \geq 2 \\ 0 & r = 1\,. \end{cases} \tag{2.16}$$

(2) 顶点 F 处入射外部合并（*EM*）箭头的数量等于

$$N^{\rm in}_{\rm EM}(F) = c(F)\,. \tag{2.17}$$

(3) 顶点 $F = T_1 \sqcup \cdots \sqcup T_r$ 与$k_i = \#L(T_i)$ 处出射内部合并（*IM*）箭头的数量等于

$$N^{\rm out}_{\rm IM}(F) = \sum_{i=1}^{c(F)} (2k_i - 4)\,. \tag{2.18}$$

(4) F 处入射 *IM* 箭头的数量等于

$$N^{\rm in}_{\rm IM}(F) = \sum_{i=1}^{c(F)} (2k_i - 4)\,. \tag{2.19}$$

(5) F 处出射最小侧向合并箭头的数量为

$$N^{\rm out}_{\rm SM,\,min}(F) = \sum_{i=d'(F)+1}^{c(F)} \binom{k_i}{2} + \sum_{i=1}^{c(F)} k_i(r - c(F)) + \sum_{i\neq j,\ i,j=1}^{c(F)} k_i k_j\,, \tag{2.20}$$

如果允许从具有至少3 个叶子的分量树中提取一对姊妹叶节点，并且

$$N^{\rm out}_{\rm SM,\,min}(F) - d''(F) \tag{2.21}$$

如果不允许提取同一顶点下的两片叶子，且$N^{\rm out}_{\rm SM,\ min}(F)$ 如(2.20)中所示。

(6) F 处入射的最小侧向合并箭头数量为零，当 $d'(F) = 0$ 时；否则

$$N^{\rm in}_{\rm SM,\,min}(F) = 6d'(F)(n - c(F)) + 2\sum_{i=1}^{d'(F)} \sum_{a\neq b\ a,b\neq i\ a,b=1}^{c(F)} (2k_a - 2)(2k_b - 2) + \tag{2.22}$$

$$2\sum_{i=1}^{d'(F)} \sum_{j\neq i,\ j=1}^{c(F)} (2k_j - 2)(2k_j + 2 + n - c(F))$$

当允许在含有至少3 片叶子的分量中切除姊妹叶子时，而在不允许切除同一顶点下两片叶子的情况下，因子 $2k_j + 2 + n - c(F)$将替换为 $2k_j + 1 + n - c(F)$。

证明。(1) 若 $F = T$ 只有一个连通分量，则不存在出射的 EM 箭头，因为 EM 会合并工作空间的两个分量。对于 $r \geq 2$ 个连通分量，可能的出射 EM 箭头对应于 F 中所有可由 EM 合并的两个分量的选择，因此共有 $\binom{r}{2}$ 个箭头。

(2) 任意具有非空边集的非平面完全二叉树 T 都能以唯一的方式写成 $T = \mathfrak{M}(T_1, T_2)$，因此它是形式为 $T_1 \sqcup T_2$ 的某个工作空间在 EM 下的像。设给定一个工作空间 $F = T_1 \sqcup \cdots \sqcup T_r$，其中 $T_1, \ldots, T_{c(F)}$ 具有非空边集，ˆ ˆ 因而对于 $i = 1, \ldots, c(F)$，有 $T_i = \mathfrak{M}(T_{i,1}, T_{i,2})$。我们将 $F = T_i \sqcup F_i = \mathfrak{M}(T_{i,1}, T_{i,2}) \sqcup F_i$ 记作 where

ˆ ˆ$F_i = \sqcup_{j\neq i} T_j$。那么 F是 $c(F)$ 个 EM 箭头的目标，这些箭头的源点是工作空间 $T_{i,1} \sqcup T_{i,2} \sqcup F_i$，其中 $i = 1, \ldots, c(F)$。

（3）每个具有两个以上边的分量 T_i of F 都是若干非平凡（非恒等映射）IM 箭头的源点；对于每个可访问项 $T_v \subset T_i$，都有一个箭头，其中非根顶点 v 不是 T_i 的根正下方的两个顶点之一（按照我们所考虑的余积形式，对这两个可访问项之一应用 IM 将只是恒等映射）。如果分量树 T_i 有 k_i 个叶子，那么它有 $2k_i - 2$ 个非根顶点，因此除去根下方的两个顶点后还有 $2k_i - 4$ 个，所以得到（2.18）。

（4）要使 F 成为一个 IM 箭头的目标，具有非平凡边集的 F 的某个分量 $T_i = \mathfrak{M}$（T_i,1,T_i,2）必须通过 IM 从形如 $T_{i,2} \lhd_e T_I$,1 或 $T_{i,1} \lhd_{e'} T_I$,2 的树得到，其中 $\lhd_e$ 是在一条边 e 上的插入前李运算。这里，我们将工作空间视为这些箭头的源点：其中 F 的 T_i 分量被其中一种边插入所替换，而 F 的所有其他分量保持不变。如果 $k_i = k_{i,1} + k_{i,2} = \#L(T_i)$，且 $k_{i,1} = \#L(T_i,1)$，$k_{i,2} = \#L(T_i,2)$，那么在 T_i,1 中有 $2k_{i,1} - 2$ 条边可以进行插入，在 T_i,2 中有 $2k_{i,2} - 2$ 条边可以进行插入，总计有 $2k_i - 4$ 种可能的边插入，也就是说，对于 T 的每个分量 T_i，有 $2k_i - 4$ 个 IM 箭头的源点。这给出了如（2.19）所示的入射 IM 箭头总数。

(5) 最小侧向合并运算选择作为单叶的可访问项或分量，并将它们合并成一个新分量。可能的情况包括：属于同一非樱桃分量的两片单叶；属于 $\pi_{0,\ E}(F)$ 中两个不同分量的两片叶子；以及 F 中由单叶构成的一个分量与某个边集非空分量中的另一片叶子。（两个分量均为单叶的情况已计入 EM 箭头，不属于 SM 的情况。）因此，最小 SM 出箭头的数量是这三种情况之和。

第一种情况，将 所有非樱桃的 $\tilde{c}(F) = c(F) - d'(F)$ 分量相加，得到

$$\sum_{i=d'(F)+1}^{c(F)} \binom{k_i}{2},$$

而在不允许提取樱桃中的一对叶子（即使该樱桃是更大的一棵树的一部分）的情况下，则需要减去 $d''(F)$。

第二种情况的总贡献为

$$\sum_{T_i \neq T_j \in \pi_{0,E}(F)} k_i \cdot k_j \,.$$

第三种情形的总和为

$$(r - c(F)) \cdot \sum_{i=1}^{c(F)} k_i$$

其中，第一个因子表示选择一个单叶分量，第二个因子表示从某个边集非空的分量中选择一片叶子。因此，出射最小 SM 箭头的总数由式（2.20）给出。

（6）由于每个最小 SM 变换都会生成一个樱桃分量，如果 F 有 $d'(F) = 0$，则不会有入射 SM 箭头。如果 $d'(F) \neq 0$，那么每个 $d'(F)$ 樱桃分量都可以由一个 SM 变换生成。对于 F 的一个给定樱桃分量，例如 $T_i = \{\alpha, \beta\}$，生成 T_i 的 SM 箭头的源工作区中，叶子 α 和 β 有以下几种可能：

(1) α 或 β 中一个是单个分量，另一个则插入$\pi_{0,E}(F)$中某个分量的一条边上，

(2) α 或 β 中一个是单个分量，另一个与F 中的一个单叶分量连接，形成一个樱桃分量；(3) α 和 β 分别插入$\pi_{0,E}(F)$中两个不同分量的边上；(4) α 或 β 中一个插入$\pi_{0,E}(F)$中的一个分量，另一个与一片单叶连接；(5) α 和 β 都插入$\pi_{0,E}(F)$中同一个分量的边上（不形成樱桃子树；若不允许切断树内同一顶点下方的两条边，则此情况除外）；(6) α 或 β 中一个与一片单叶连接，形成一个樱桃，另一个则插入该樱桃的一条边上。

第一种情况对应于

$$2\sum_{i=1}^{d'(F)}\sum_{j\neq i,\, j=1}^{c(F)}(2k_j-2)\,.$$

第二种情况对应于

$$2\cdot d'(F)\cdot(n-c(F))\,.$$

第三种情况对应于

$$2\sum_{i=1}^{d'(F)}\sum_{a\neq b\ a,b\neq i}^{c(F)}\sum_{a,b=1}(2k_a-2)(2k_b-2)\,.$$

第四种情况对应于

$$2\cdot(n-c(F))\sum_{i=1}^{d'(F)}\sum_{j\neq i,\, j=1}^{c(F)}(2k_j-2)\,.$$

第五种情况对应于

$$2\sum_{i=1}^{d'(F)}\sum_{j\neq i\, j=1}^{c(F)}(2k_j-2)(2k_j+1)$$

其中，因子 $2k_j-2$ 计入了在 T_j 中选择首次插入所在边的情况。随后，在 $T_j\lhd_e\alpha$ 中有 $2k_j+1$ 条边可供选择，以便插入 β。如果不允许在树分量内剪开樱桃的两片叶子，那么第二条边只有 $2k_j$种选择，因此计数式改为

$$2\sum_{i=1}^{d'(F)}\sum_{j\neq i\, j=1}^{c(F)}(2k_j-2)2k_j\,.$$

第六种情况计入了

$$4\cdot d'(F)\cdot(n-c(F))\,.$$

由此得到式 (2.22)。 □

正如我们将进一步讨论的，从 Merge 动力系统的性质来看，纳入那些同时移除某个顶点下方两条边的可容许切割之所以更可取，其中一个原因是，这样一来，SM 箭头入度/出度的计数就只取决于划分 $\wp=p(F)$ 的数据，而不再取决于森林F的具体拓扑；而区分是否切断同一顶点下方的两条边，则取决于关于 F 的具体信息，而不仅仅是划分 $\wp$。正如我们将在第4节讨论的，入度/出度仅取决于 $\wp$ 这一事实极大地简化了动力学。

2.3. 稀疏性。 命题 2.3 还展示了这些图的稀疏性 性质。也就是说，随着 n规模增大，图$\mathcal{G}_{n,\ A}$会以极快的速度变得越来越稀疏。有向图的密度定义为

$$D(\mathcal{G}) = \frac{\#E(G)}{\#V(G)\cdot(\#V(G)-1)}\,, \tag{2.23}$$

即与边的最大数量之比。如果图族 $\mathcal{G}_n$ 日益变得稀疏，则

$$\lim_{n\to\infty} D(\mathcal{G}_n) = 0\,.$$

推论 2.4。 密度 D（$\mathcal{G}_{n,\ A}$）满足

$$D(\mathcal{G}_{n,A}) \le \frac{P(n)}{\Lambda_n}\,, \tag{2.24}$$

其中 P（n） 是一个关于 n 次数至多为三次的多项式，并且 Λ_n 如 （2.12） 中所示。

证明。 边数 #E($\mathcal{G}_{n,\ A}$) 可以由命题 2.3 中计算出的入射边和出射边数量得到，因为令 deg 表示顶点的度，有

$$\#E(\mathcal{G}_{n,A}) = \frac{1}{2}\sum_{F\in V(\mathcal{G}_{n,A})} \deg(F)$$

$$= \frac{1}{2}\sum_{F\in V(\mathcal{G}_{n,A})} \left(N^{\mathrm{out}}_{\mathrm{EM}}(F) + N^{\mathrm{in}}_{\mathrm{EM}}(F) + N^{\mathrm{out}}_{\mathrm{IM}}(F) + N^{\mathrm{in}}_{\mathrm{IM}}(F) + N^{\mathrm{out}}_{\mathrm{SM\,min}}(F) + N^{\mathrm{in}}_{\mathrm{SM\,min}}(F)\right).$$

我们可以将这些项界定为

$$N^{\mathrm{out}}_{\mathrm{EM}}(F) \le \binom{n}{2} = O(n^2), \quad N^{\mathrm{in}}_{\mathrm{EM}}(F) \le n$$

$$N^{\mathrm{out}}_{\mathrm{IM}}(F) \le 2n(n-2) = O(n^2), \quad N^{\mathrm{in}}_{\mathrm{IM}}(F) \le 2n(n-1) = O(n^2)$$

$$N^{\mathrm{out}}_{\mathrm{SM\,min}}(F) \le 2n(n-1) = O(n^2), \quad N^{\mathrm{in}}_{\mathrm{SM\,min}}(F) \le 2n^2(3n+1) = O(n^3)\,.$$

接着，我们得 到一个总体多项式增长估计，再乘以顶点数

$$\#E(\mathcal{G}_{n,A}) \le P(n)\cdot \#V(\mathcal{G}_{n,A})\,.$$

因此，密度满足

$$D(\mathcal{G}_{n,A}) \le \frac{P(n)}{\#V(\mathcal{G}_{n,A})}\,.$$

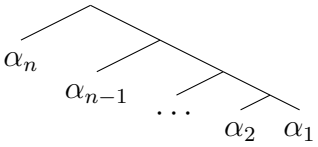
顶点数 #V($\mathcal{G}_{n,\ A}$) 等于森林 $F \in \mathfrak{F}_{A,\ n}$ 的数量。这可以按照引理 2.2 计算。□

2.4. 强连通性。 回顾一下，如果图 G 中任意一对顶点之间都存在一条由边构成的有向路径，则该有向图是强连通的。若图 G 中所有循环的环长的最大公约数都大于 1，则该图是周期的；若环长的最大公约数等于 1，则该图是非周期的。

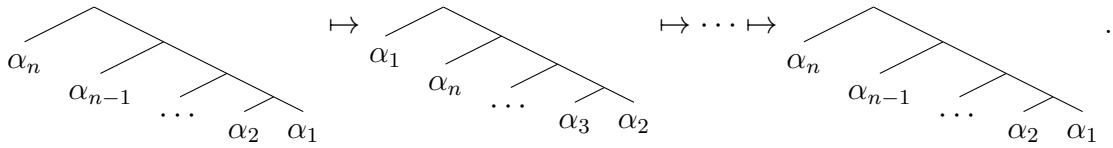
命题 2.5。 对于任意选择的(n, A)，图 $\mathcal{G}_{n,\ A}$ 都是强连通且非周期的。

证明。 图$\mathcal{G}_{n,\ A}$的强连通性已在 [22] 和 [26], 中证明，其依据是存在最小 SM 箭头。为证明非周期性，在 $n = 3$ 的情况下，直接检查可知存在长度为 2 和 3 的循环，因此环长的最大公约数为 1。对于 $n \geq 4$，注意到每个图$\mathcal{G}_{n,\ A}$ 都包含一个顶点 F，它对应于工作空间$F = T_n$，该工作空间只有一个分量，且该分量由梳状树给出

$$T_n = \quad .$$



作用于该 F 的可能 IM 变换中，有一种会提取叶子α_1，并生成一个具有相同梳状拓扑、但标签发生循环置换的树。随后，可以对所得的树应用同类型的 IM 操作，提取 α_2 叶子，如此继续。这样的内部合并操作的复合构成了一个长度为 $n - 1$的循环



T图 $\mathcal{G}_{n,\ A}$ 还包含，对于每个 $2 < k < n$，一个该形式的工作空间 m

$$F = T_k \sqcup \alpha_{k+1} \sqcup \cdots \sqcup \alpha_n \, .$$

内部合并通过作用于 T_k 分量而作用于这些工作空间，我们再次得到一个长度为 k 的循环，如上所述。因此，$\mathcal{G}_{n,\ A}$ 包含所有长度为 $k - 1 = 2$，....., $n - 1$的循环，所以循环环长的最大公约数再次为 1。□

3. Random walks versus Hopf algebra Markov chains

鉴于上文所述的合并图，研究合并作用于工作空间上的动力学有两种不尽相同的方法。一种可能的方法是将合并作用的动力系统描述为合并图上的随机游走 $\mathcal{G}$（即在每个$\mathcal{G}_{n,\ A}$ 图上的随机游走）。另一种方法是将合并算子作为霍普夫代数马尔可夫链（HAMC）进行研究，正如 [22] 和 [26]中已经讨论的那样。

在本节中，我们将重点说明这两种方法有何不同，以及为什么在我们的模型中应当采用第二种方法。为此，我们分别描述两种情况下构造相关动力系统的主要步骤，并突出介绍它们的主要性质。在本文其余部分，我们将只关注霍普夫代数马尔可夫链的构造。

描述两者差异的关键在于：在第一种情况下，保留工作空间组合霍普夫代数的同一组基（森林 F），并对该基下 Merge 的矩阵表示（即 Merge 图的邻接矩阵）的各项重新缩放，使其归一化为随机矩阵；而在第二种情况下，则寻找一个合适的基，使 Merge 算子本身成为马尔可夫链：通过用 Merge 图邻接矩阵的佩龙–弗罗贝尼乌斯特征向量的各分量对基元素 F 重新缩放，即可得到所需的基。因此，主要区别在于，我们是像 HAMC 方法那样，将 Merge 算子本身作为马尔可夫链来研究（这反映了语言学情境中对 Merge 的马尔可夫性质的考察，见 [8], [9]）；还是根据 Merge 图顶点的度（即可以应用于给定工作空间的可能 Merge 变换的数量）为 Merge 算子加权，从而在相同的组合基下得到另一个马尔可夫链算子，如随机游走算子的情形。

考虑这两种可能性都有充分理由：语言学文献中通常认为 Merge 算子本身具有马尔可夫性的说法，通过此处介绍的 HAMC 方法得到了严格的形式化；而随机游走方法则可以纳入这样一个事实：无论如何，都需要对 Merge 操作加权，以描述成本函数和最优性性质，我们将在 §11 中进一步详细讨论这一点。此外，第一种方法中使用的重新缩放也可以结合到一个总体成本函数中（尽管其在语言学模型中的解释不那么直观）。

然而，HAMC 方法还有一项额外优势：HAMC 动力系统满足熵优化性质，正如我们在 §3.3 中讨论的那样。事实上，正如我们将在 §13 中进一步讨论的，关于香农熵的其他形式的优化，也将在合并作用的动力学性质中发挥重要作用。

基于这些原因，我们将在此讨论这两种方法，但只着重于 HAMC 表述。

3.1. 在合并图上的随机游走。 对于无向图 G，随机游走的行为很简单。它是一个转移概率为 $v, v' \in V(G)$ 的马尔可夫链，

$$\mathbb{P}(v,v') = \begin{cases} \frac{1}{\deg(v)} & \exists e \in E(G) : \partial e = \{v, v'\} \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

平稳分布

$$\pi(v') = \sum_{v \in V(G)} \pi(v)\mathbb{P}(v,v') \tag{3.2}$$

则可直接写为

$$\pi(v) = \frac{\deg(v)}{2\#E(G)},$$

这里我们再次利用了这一事实：$\sum_{v\in V(G)} \deg(v) = 2\#E(G)$。

然而，有向图上的随机游走更为微妙。它由具有如下转移概率的马尔可夫链给出，对于 $v,\ v' \in V(G)$，

$$\mathbb{P}(v,v') = \begin{cases} \frac{1}{\deg^{\text{out}}(v)} & \exists e \in E(G) : v = s(e), v' = t(e) \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

此时，每个顶点处只选择出向方向。在这种情况下，无法像无向图那样用度来简洁地表示平稳分布，因为此时我们会有 $\sum_v \deg^{\text{out}}(v)\mathbb{P}(v,\ v') = \deg^{\text{in}}(v')$。然而，如果图 G 是强连通的，则可以通过佩龙–弗罗贝尼乌斯定理求得平稳分布。

收敛到平稳分布的过程取决于图的非周期性。环长的最大公约数等于 $k > 1$ 的周期性，等价于存在一个态射 $G \to C_k$，其中 C_k 是长度为 k的循环有向多边形图；这又等价于转移矩阵的谱中包含 k 次单位根，并进而意味着通过迭代无法收敛到平稳分布，而是会出现振荡行为。另一方面，在非周期的情况下，Perron-Frobenius 定理保证所有其他特征值的绝对值都更小，因此随机游走总会收敛到由 Perron-Frobenius 特征向量（3.2）给出的平稳分布。

对于合并图，这种方法考虑的是上述有向图$\mathcal{G}_{n,\ A}$ 上的随机游走。根据命题 2.5，我们知道合并图 $\mathcal{G}_{n,\ A}$ 都是强连通且非周期的。原始合并算子是在$\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})$上的线性算子；对于标准基 $\{F \in \mathfrak{F}_{\mathcal{SO}_0}\}$，它在不变子空间 $\mathcal{V}(\mathfrak{F}_{A,\ n})$上的矩阵表示为有向图$\mathcal{G}_{n,\ A}$的邻接矩阵。随机

游走算子则是在同一向量空间上的线性算子，并在同一基下通过缩放各项 $\mathcal{K}_{F,\,F'} \mapsto \deg^{out}(F)^{-1}\,\mathcal{K}_{F,\,F'}$ 来表示

$$\deg^{out}(F) = N^{\rm out}_{\rm EM}(F) + N^{\rm out}_{\rm IM}(F) + N^{\rm out}_{\rm SM\,min}(F)\,.$$

我们使用记号 $\mathcal{RW}$ 和

$$\mathcal{RW}_{F,F'} := \deg^{out}(F)^{-1}\,\mathcal{K}_{F,F'} \tag{3.4}$$

表示合并图上的这个随机游走算子；当我们希望明确追踪图 $\mathcal{G}_{n,A}$ 的数据（A, n）时，记作 $\mathcal{RW}^{(A,n)}$ 和 $\mathcal{RW}^{(A,n)}_{F\,F'}$。

从促使我们进行这一构造的语言学模型的角度来看，使用随机游走算子 $\mathcal{RW}^{(A,n)}$ 的问题在于，人们无法对将 $\mathcal{K} = (\mathcal{K}_{F,F'})$（语言学上的合并操作）的各项乘以出度 $\deg^{out}(F)^{-1}$ 的倒数这一重新缩放操作给出自然的理论依据：从理论语言学的视角看，具有马尔可夫性的是合并作用本身。因此，采用以下方法能得到更好的表述。

3.2. 霍普夫代数马尔可夫链。 上文提到的另一种方法并不对 $\mathcal{K}_{F,F'}$ 矩阵的各项进行归一化以使其成为随机矩阵，而是寻找恰当的基变换，使原线性算子在该基下表示为随机矩阵。如 [14], [27] 中所述，这可以做到。由于图 $\mathcal{G}_{n,A}$ 是强连通的，我们可以应用佩龙-弗罗贝尼乌斯定理；该定理保证存在右佩龙-弗罗贝尼乌斯特征向量 $\eta(F) = \eta_{n,A}(F) > 0$，

$$\sum_{F'} \mathcal{K}^{(A,n)}_{F,F'}\ \eta(F') = \lambda\,\eta(F)\,, \tag{3.5}$$

其中 $\lambda = \lambda_{n,\,A}$ 是图 $\mathcal{G}_{n,\,A}$ 的邻接矩阵 $\mathcal{K}^{(A,n)}$ 的佩龙-弗罗贝尼乌斯特征值。随后，我们可以将基元素按 $F \mapsto \eta(F)^{-1}F$ 重新缩放。在这个新基 $\{\eta(F)^{-1}F\}$ 下，同一个合并算子由矩阵 $\hat{\mathcal{K}}^{(A,n)}_{F\,F'}$ 表示，其形式为

$$\hat{\mathcal{K}}^{(A,n)}_{F,F'} = \lambda^{-1}\frac{\eta(F')}{\eta(F)}\mathcal{K}^{(A,n)}_{F,F'}\,. \tag{3.6}$$

按（3.6）构造得到的矩阵 $\hat{\mathcal{K}}^{(A,n)} = (\hat{\mathcal{K}}^{(A,n)}_{F\,F'})$ 按构造即为随机矩阵，因此它定义了一个马尔可夫链的转移矩阵。这就是“合并具有马尔可夫性”这一性质的严格表述，而（3.6）则是实际对应的霍普夫代数马尔可夫链。

随后，可以研究平稳分布（3.2）与 $\mathbb{P}(F,\ F') = \hat{\mathcal{K}}^{(A,n)}_{F\,F'}$，以及这个马尔可夫链的其他性质；这些性质描述了合并在工作空间上的动力学。命题 2.5 已经提供了关于合并作用于工作空间的一些重要信息。

命题 3.1。对于一个强连通图 $\mathcal{G}$，其邻接矩阵为 $\mathcal{K}$，佩龙-弗罗贝尼乌斯特征值为 λ，且（右）特征向量为 η。设 $\mathcal{K}$ 是按照 (3.6) 所定义的相应马尔可夫链。$\mathcal{K}$ 的平稳分布 π 为

$$\sum_x \pi(x)\,\hat{\mathcal{K}}_{x,y} = \pi(y)\,, \tag{3.7}$$

给定为

$$\pi(x) = \psi(x)\eta(x)\,, \tag{3.8}$$

其中 ψ、η 分别是 $\mathcal{K}$ 的左、右 *Perron-Frobenius* 特征向量。

ˆ证明。 由于图 $\mathcal{G}$ 是强连通的，马尔可夫链 $\mathcal{K}$ 是不可约的，因此其（归一化）平稳分布是唯一的。因此，只需验证 $\psi(x)\eta(x)$满足 (3.7)。我们有

$$\sum_x \psi(x)\eta(x)\lambda^{-1}\frac{\eta(y)}{\eta(x)}\mathcal{K}_{x,y} = \sum_x \psi(x)\mathcal{K}_{x,y}\lambda^{-1}\eta(y) = \psi(y)\eta(y)\,.$$

□

我们将在命题 9.1 中证明这一简单观察的一个更精细版本，并将其应用于马尔可夫链之间的投影。

注 3.2。 转移矩阵为 $\mathcal{K}$ 的马尔可夫链，若平稳分布唯一，且底层图 G 的顶点集上的任意给定初始分布经过 $\mathcal{K}$ 的迭代后都收敛到该平稳分布，则称其为遍历的 。

如上文注 3.2 所述，遍历性源于 G 的强连通性和非周期性，因此命题 2.5 有以下直接推论。

命题 3.3。 合并霍普夫代数马尔可夫链ˆ$\mathcal{K}^{(A,n)}$ 是遍历的，合并随机游走 $\mathcal{RW}^{(A,n)}$ 也是如此。

这一结果为句法极简主义模型提供了有趣的理论信息。它表明，句法形成与转换（合并算子所模拟的过程）是一个遍历的动力系统。这意味着该动力学具有唯一的平稳分布，能够刻画其长期行为（处于某个可能状态——即工作空间、句法结构——的长期概率）。遍历马尔可夫链的长期行为是可预测的：对于动力系统，遍历性通常被视为时间平均与空间平均相同这一性质。对于马尔可夫链，这可以更精确地表述为注 3.2 中的性质：处于给定状态的长期概率不取决于初始状态的选择。平稳分布的值 $\pi(F)$ 也可以计算该动力学的期望返回时间，其值为 $1/\pi(F)$。

不过请注意，当我们将侧向合并纳入动力学时，这些良好的动力学性质才成立（因为图$\mathcal{G}_{n,A}$的强连通性要求存在 SM 箭头）。我们将在第 6 节更详细地讨论：缺少 SM 箭头时会发生什么变化，以及如何通过适当的代价函数为不同类型的箭头——EM、IM 和 SM——赋予不同的权重。

3.3. **作为最大熵随机游走的霍普夫代数马尔可夫链**。 我们还要指出，合并的 HAMC 表述（如式 (3.6) 所示）具有一项随机游走算子 $\mathcal{RW}$ 所不具备的重要性质：最大熵优化。

设有一个有向图 $\mathcal{G}$，其邻接矩阵为 $\mathcal{K}_{\mathcal{G}}$。再设 $\mathcal{S}$ 是定义在同一底层有向图 $\mathcal{G}$上的马尔可夫链，即邻接矩阵 $\mathcal{K}_{\mathcal{G}}$的一个加权版本，且该版本是随机矩阵。

转移概率为 $\mathcal{S}(x,\ y)$且平稳分布为 $\pi(x)$的马尔可夫链，其熵定义为

$$\mathrm{Sh}(\mathcal{S}) = -\sum_x \pi(x)\sum_y \mathcal{S}(x,y)\log\mathcal{S}(x,y)\,. \tag{3.9}$$

这也称为 $\mathcal{S}$ 的熵率。

熵率 $\mathrm{Sh}(\mathcal{S})$ 等于路径长度趋于无穷时，路径空间上的概率分布的平均熵的渐近行为。也就是说，假设

给定图 $\mathcal{G}$ 中的一条有向路径 $\gamma = e_0 e_1 \dots e_{\ell-1}$，其边$s(e_i) = x_i$ 和 $t(e_i) = x_{i+1}$ 的源点和目标顶点为.....长度为 ℓ 的路径上的概率分布由下式给出

$$\mathbb{P}_\ell(\gamma) = \pi(x_0)\, \mathcal{S}(x_0, x_1) \cdots \mathcal{S}(x_{\ell-1}, x_\ell)\,. \tag{3.10}$$

熵率（3.9）是以下极限

$$\mathrm{Sh}(\mathcal{S}) = \lim_{\ell\to\infty} \frac{\mathrm{Sh}(\mathbb{P}_\ell)}{\ell}\,, \tag{3.11}$$

其中，$\mathrm{Sh}(\mathbb{P}_\ell)$ 是概率（3.10）的香农熵。

当 $\mathcal{S}$ 遍历所有具有相同底层有向图 $\mathcal{G}$ 的马尔可夫链时，已知（见 [29], [30]），可能达到的最大熵率 $\mathrm{Sh}(\mathcal{S})$ 就是有向图 $\mathcal{G}$ 的拓扑熵，其表达式为

$$S_{\mathrm{top}}(\mathcal{G}) := \log \lambda_{\mathcal{G}}\,, \tag{3.12}$$

其中，$\lambda_{\mathcal{G}}$ 是邻接矩阵 $\mathcal{K}_{\mathcal{G}}$ 的佩龙-弗罗贝尼乌斯特征值 .

以下事实广为人知（例如参见 [5], [11], [29], [30]），但为了方便读者，我们在此重述。

命题 3.4。设给定一个强连通有向图 $\mathcal{G}$。令 $\mathcal{K}_{\mathcal{G}}$ 为 $\mathcal{G}$的邻接矩阵，其佩龙-弗罗贝尼乌斯特征值为 λ，右、左佩龙-弗罗贝尼乌斯特征向量分别为 η 和 ψ，并将它们归一化，使得

$$\sum_{x,y} (\psi(x)\eta(y))^2 = 1\,. \tag{3.13}$$

最大熵随机游走（*MERW*）是如下马尔可夫链

$$\mathcal{S}_{\mathrm{MERW}}(x,y) = \hat{\mathcal{K}}_{\mathcal{G}}(x,y) = \frac{1}{\lambda}\frac{\eta(y)}{\eta(x)} \mathcal{K}_{\mathcal{G}}(x,y)\,. \tag{3.14}$$

$\mathcal{S}_{\mathrm{MERW}}$ 满足以下性质。

(1) 概率分布(3.10)（其中 $\mathcal{S} = \mathcal{S}_{\mathrm{MERW}}$）在所有长度和端点相同的路径上均匀分布。(2) 熵率达到最大值，

$$\mathrm{Sh}(\mathcal{S}_{\mathrm{MERW}}) = S_{\mathrm{top}}(\mathcal{G})\,.$$

证明。将 MERW 马尔可夫链定义为 (3.14)。由 (3.10) 可知，它在端点和长度固定的路径空间上确定了一个均匀分布，因为

$$\mathbb{P}_{\mathrm{MERW},\ell}(\gamma) = \frac{1}{\lambda^\ell}\pi(x_0)\frac{\eta(x_\ell)}{\eta(x_0)} \mathcal{K}_{\mathcal{G}}(x_0,x_1)\cdots\mathcal{K}_{\mathcal{G}}(x_{\ell-1},x_\ell) = \frac{\psi(x_0)\eta(x_\ell)}{\lambda^\ell}\,, \tag{3.15}$$

其中，在第二个等式中，我们用到了命题 3.1，以及以下事实：对于$\mathcal{G}$ 中的有向边 e_i，若 $s(e_i) = x_i$ 且 $t(e_i) = x_{i+1}$，则有 $\mathcal{K}_{\mathcal{G}}(x_i,\ x_{i+1}) = 1$。为说明 MERW 也使熵率最大，我们使用极限 (3.11)。我们有

$$\mathrm{Sh}(\mathbb{P}_{\mathrm{MERW},\ell}) = -\sum_{\gamma} \mathbb{P}_{\mathrm{MERW},\ell}(\gamma)\log \mathbb{P}_{\mathrm{MERW},\ell}(\gamma)\,,$$

对所有长度为 ℓ 的路径 γ 求和。这就是

$$\mathrm{Sh}(\mathbb{P}_{\mathrm{MERW},\ell}) = -\sum_{x_0,x_\ell} \mathcal{K}^\ell_{\mathcal{G}}(x_0,x_\ell)\frac{\psi(x_0)\eta(x_\ell)}{\lambda^\ell}\log\left(\frac{\psi(x_0)\eta(x_\ell)}{\lambda^\ell}\right)$$

$$= -\sum_{x_0,x_\ell} \frac{\mathcal{K}_{\mathcal{G}}^{\ell}(x_0,x_\ell)}{\lambda^{\ell}}\,\psi(x_0)\eta(x_\ell)\log(\psi(x_0)\eta(x_\ell)) + \sum_{x_0,x_\ell} \frac{\mathcal{K}_{\mathcal{G}}^{\ell}(x_0,x_\ell)}{\lambda^{\ell}}\,\psi(x_0)\eta(x_\ell)\log(\lambda^{\ell})\,,$$

其中 $\mathcal{K}_{\mathcal{G}}^{\ell}(x,\ y) = \#\{\gamma \mid s(\gamma) = x,\ \ t(\gamma) = y\,,\ \text{长度}(\gamma) = \ell\}$。因此，我们有

$$\mathrm{Sh}(\mathcal{S}_{\mathrm{MERW}}) = \lim_{\ell\to\infty} \frac{\mathrm{Sh}(\mathbb{P}_{\mathrm{MERW},\ell})}{\ell} = \lim_{\ell\to\infty} \sum_{x,y} \frac{\mathcal{K}_{\mathcal{G}}^{\ell}(x,y)}{\lambda^{\ell}}\,\psi(x)\eta(y)\log(\lambda)\,.$$

佩龙-弗罗贝尼乌斯定理表明，当 $\ell \to \infty$

$$\frac{\mathcal{K}_{\mathcal{G}}^{\ell}(x,y)}{\lambda^{\ell}} \sim \psi(x)\eta(y)\,,$$

因此，上述极限和归一化 (3.13) 给出

$$\mathrm{Sh}(\mathcal{S}_{\mathrm{MERW}}) = \sum_{x,y}(\psi(x)\eta(y))^2\ \log(\lambda) = \log(\lambda) = S_{\mathrm{top}}(\mathcal{G})\,.$$

□

请注意，最大熵随机游走的转移概率是非局部的（不像随机游走$\mathcal{RW}$ 的转移概率只取决于各顶点的（出）次数），而且它们通过邻接矩阵的佩龙-弗罗贝尼乌斯特征值和特征向量，取决于整个图拓扑。

3.4. 自由能优化。 我们还回顾一个一般化的命题 3.4，以备后用：其中使用加权版本代替邻接矩阵，而最大熵随机游走会在路径上诱导玻尔兹曼分布，并相应地优化自由能，而非熵率（见 [30]）。

设 $\mathcal{G}$ 为一个强连通有向图，并设 $\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}$ 为一个加权邻接矩阵，即

$$\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x,y) = e^{-\beta\,\omega(x,y)}\ \mathcal{K}_{\mathcal{G}}(x,y)\,, \tag{3.16}$$

其中 $\omega(x,\ y) \in \mathbb{R}$ 是实权重，而 $\mathcal{K}_{\mathcal{G}}$ 是 $\mathcal{G}$的邻接矩阵。我们将 $\mathcal{G}$ 的边上的能量函数 $\mathbb{E}: E(\mathcal{G}) \to \mathbb{R}_+$ 定义为

$$\mathbb{E}(e) = \omega(x,y)\,, \quad \text{for } x = s(e) \text{ and } y = t(e)\,.$$

尽管权重不一定为正，但由于图是有限的，能量有下界，因此经过整体平移后能量非负。设 $\mathcal{S}$ 是一个底层有向图为 $\mathcal{G}$的马尔可夫链。该马尔可夫链的自由能 $\mathbb{F}(\mathcal{S})$定义为

$$\mathbb{F}(\mathcal{S}) = \overline{\mathbb{E}}(\mathcal{S}) - \beta^{-1}\mathrm{Sh}(\mathcal{S}) \tag{3.17}$$

¯其中 $\mathbb{E}(\mathcal{S})$是期望能量

$$\overline{\mathbb{E}}(\mathcal{S}) := \sum_{x} \pi^{\mathcal{S}}(x) \sum_{y} \mathcal{S}(x,y)\,\omega(x,y)\,, \tag{3.18}$$

$\pi^{\mathcal{S}}$ 其中 $\mathcal{S}$是 $\mathcal{S}$的平稳分布。我们寻找使 $\mathcal{S}(\mathcal{S})$达到最小值的马尔可夫链。注意，在没有权重的情况下，$\omega = 0$，这等价于最大化熵 $\mathrm{Sh}(\mathcal{S})$，因此可以得到前面的情形。有时也会用等价的方式，将这个问题表述为最大化 自由能 $\mathbb{F}$，即

$$\tilde{\mathbb{F}}(\mathcal{S}) = \mathrm{Sh}(\mathcal{S}) - \beta\overline{\mathbb{E}}(\mathcal{S}) \tag{3.19}$$

与 [30], 中的表述一样，而我们在此采用的对（3.17）进行最小化的方法，更适合直接与 [21] 中使用的热力学半环进行比较。

定义 3.5。 $\mathcal{G}$ 的边集上的玻尔兹曼分布 定义为(3.20
)

$$\mathbb{P}_{(\omega,\beta)}(e) := \frac{\psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))e^{-\beta\mathbb{E}(e)}}{Z_\omega(\beta)} \qquad with\ Z_\omega(\beta) = \sum_e \psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))\ e^{-\beta\mathbb{E}(e)}\,,$$

其中 $\psi^{(\omega,\beta)}$ 和 $\eta^{(\omega,\beta)}$ 分别是 $\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}$ 的左、右 *Perron-Frobenius* 特征向量，其 *Perron-Frobenius* 特征值为 $\lambda_{(\omega,\beta)}$。我们采用如下归一化

$$\sum_x \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(x) = 1\,. \tag{3.21}$$

引理 3.6。 玻尔兹曼分布 (3.20)的配分函数满足

$$Z_\omega(\beta) = \lambda_{\omega,\beta} \tag{3.22}$$

证明。我们有

$$\sum_e \psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))\ e^{-\beta\mathbb{E}(e)} = \sum_{x,y} \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)e^{-\beta\omega(x,y)}\mathcal{K}_{\mathcal{G}}(x,y) =$$

$$\lambda_{\omega,\beta}\sum_x \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(x) = \lambda_{\omega,\beta}$$

采用归一化 (3.21)。 □

我们将由加权邻接矩阵所关联的马尔可夫链记为 $\hat{}\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}$

$$\hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x,y) := \frac{1}{\lambda_{(\omega,\beta)}}\frac{\eta^{(\omega,\beta)}(y)}{\eta^{(\omega,\beta)}(x)}\ \mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x,y)\,. \tag{3.23}$$

它具有以下性质。

引理 3.7。 对于马尔可夫链 $\hat{}\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}$ ， 期望能量满足

$$\mathbb{F}(\hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}) = \overline{\mathbb{E}}_{\mathbb{P}_{(\omega,\beta)}} - \beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\omega,\beta)})\,, \tag{3.24}$$

其中 $\mathbb{P}_{(\omega,\beta)}$ 为玻尔兹曼分布(3.20) ， 且

$$\overline{\mathbb{E}}_{\mathbb{P}_{(\omega,\beta)}} = \sum_e \mathbb{P}_{(\omega,\beta)}(e)\,\mathbb{E}(e)\,.$$

证明。我们有

$$\mathbb{F}(\hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}) = \sum_x \pi^{(\omega,\beta)}(x)\sum_y \hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x,y)(\omega(x,y) + \beta^{-1}\log(\hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x,y)))$$

$$= \sum_x \pi^{(\omega,\beta)}(x)\sum_y \hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x,y)(\omega(x,y) + \beta^{-1}\log(e^{-\beta\omega(x,y)}) + \beta^{-1}\log(\hat{\mathcal{K}}_{\mathcal{G}}(x,y)))$$

$$= \beta^{-1}\sum_x \pi^{(\omega,\beta)}(x)\sum_y e^{-\beta\omega(x,y)}\hat{\mathcal{K}}_{\mathcal{G}}(x,y)\log(\hat{\mathcal{K}}_{\mathcal{G}}(x,y)))\,.$$

由引理 3.6 和命题 3.1 应用于 $\hat{}\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}$ ， 我们还得到

$$\overline{\mathbb{E}}_{\mathbb{P}_{(\omega,\beta)}} = \sum_{x,y}\frac{\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)}{\lambda_{(\omega,\beta)}}e^{-\beta\omega(x,y)}\,\mathcal{K}_{\mathcal{G}}(x,y)\,\omega(x,y) = \sum_x \pi^{(\omega,\beta)}(x)\sum_y e^{-\beta\omega(x,y)}\,\hat{\mathcal{K}}_{\mathcal{G}}(x,y)\,\omega(x,y)$$

and

$$-\beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\omega,\beta)}) = \beta^{-1}\sum_x \pi^{(\omega,\beta)}(x)\sum_y \hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x,y)\log(\hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x,y))$$

$$= \beta^{-1}\sum_x \pi^{(\omega,\beta)}(x)\sum_y \hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x,y)\log(e^{-\beta\omega(x,y)}) + \beta^{-1}\sum_x \pi^{(\omega,\beta)}(x)\sum_y e^{-\beta\omega(x,y)}\hat{\mathcal{K}}_{\mathcal{G}}(x,y)\log(\hat{\mathcal{K}}_{\mathcal{G}}(x,y))$$

$$= -\sum_x \pi^{(\omega,\beta)}(x)\sum_y e^{-\beta\omega(x,y)}\,\hat{\mathcal{K}}_{\mathcal{G}}(x,y)\,\omega(x,y) + \beta^{-1}\sum_x \pi^{(\omega,\beta)}(x)\sum_y e^{-\beta\omega(x,y)}\hat{\mathcal{K}}_{\mathcal{G}}(x,y)\log(\hat{\mathcal{K}}_{\mathcal{G}}(x,y))\,,$$

因此

$$\overline{\mathbb{E}}_{\mathbb{P}_{(\omega,\beta)}} - \beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\omega,\beta)}) = \beta^{-1}\sum_x \pi^{(\omega,\beta)}(x)\sum_y e^{-\beta\omega(x,y)}\hat{\mathcal{K}}_{\mathcal{G}}(x,y)\log(\hat{\mathcal{K}}_{\mathcal{G}}(x,y))\,.$$

□

我们还可以得到玻尔兹曼分布的以下性质。

引理 3.8。 式 (3.20) 中的玻尔兹曼分布 $\mathbb{P}_{(\omega,\beta)}$ 满足

$$\overline{\mathbb{E}}_{\mathbb{P}_{(\omega,\beta)}} = -\frac{d}{d\beta}\log Z_\omega(\beta)\,, \tag{3.25}$$

$$\mathrm{Sh}(\mathbb{P}_{(\omega,\beta)}) = \log Z_\omega(\beta) - \beta\frac{d}{d\beta}\log Z_\omega(\beta) + \mathrm{Sh}(\pi^{(\omega,\beta)})\,, \tag{3.26}$$

$$\mathbb{F}_{\mathbb{P}_{(\omega,\beta)}} = \overline{\mathbb{E}}_{\mathbb{P}_{(\omega,\beta)}} - \beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\omega,\beta)}) = -\beta^{-1}(\mathrm{Sh}(\pi^{(\omega,\beta)}) + \log Z_\omega(\beta))\,. \tag{3.27}$$

证明。 我们将 (3.25) 写成如下恒等式

$$\overline{\mathbb{E}}_{\mathbb{P}_{(\omega,\beta)}} = \sum_e \mathbb{P}_{(\omega,\beta)}(e)\,\mathbb{E}(e) = \frac{\sum_e \psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))e^{-\beta\mathbb{E}(e)}\mathbb{E}(e)}{Z_\omega(\beta)}$$

$$= \frac{-\frac{d}{d\beta}\sum_e \psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))e^{-\beta\mathbb{E}(e)}}{Z_\omega(\beta)} = -\frac{d}{d\beta}\log Z_\omega(\beta)\,.$$

我们还有

$$\mathrm{Sh}(\mathbb{P}_{(\omega,\beta)}) = -\sum_e \frac{\psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))e^{-\beta\omega(e)}}{Z_\omega(\beta)}\log(\frac{\psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))e^{-\beta\omega(e)}}{Z_\omega(\beta)})$$

$$= \beta\sum_e \frac{\psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))e^{-\beta\omega(e)}}{Z_\omega(\beta)}\omega(e)$$

$$+\sum_e \frac{\psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))e^{-\beta\omega(e)}}{Z_\omega(\beta)}\log Z_\omega(\beta)$$

$$-\sum_e \frac{\psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e))e^{-\beta\omega(e)}}{Z_\omega(\beta)}\log(\psi^{(\omega,\beta)}(s(e))\eta^{(\omega,\beta)}(t(e)))$$

$$= \beta\overline{\mathbb{E}}_{\mathbb{P}_{(\omega,\beta)}} + \log Z_\omega(\beta) - \sum_{x,y}\frac{\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x,y)}{Z_\omega(\beta)}\log(\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y))$$

$$= \log Z_\omega(\beta) - \beta\frac{d}{d\beta}\log Z_\omega(\beta) - \sum_{x,y}\frac{\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x,y)}{\lambda_{(\omega,\beta)}}\log(\psi^{(\omega,\beta)}(x))$$

$$-\sum_{x,y}\frac{\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x,y)}{\lambda_{(\omega,\beta)}}\log(\eta^{(\omega,\beta)}(y))=\log Z_\omega(\beta)-\beta\frac{d}{d\beta}\log Z_\omega(\beta)$$
$$-\sum_x\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(x)\log(\psi^{(\omega,\beta)}(x))-\sum_y\psi^{(\omega,\beta)}(y)\eta^{(\omega,\beta)}(y)\log(\eta^{(\omega,\beta)}(y))$$
$$=\log Z_\omega(\beta)-\beta\frac{d}{d\beta}\log Z_\omega(\beta)-\sum_x\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(x)\log(\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(x))\,,$$

由此得到 (3.26)，进而推出 (3.27)。 □

我们还有命题 3.4 的如下推广。

命题 3.9。设有一个强连通有向图 $\mathcal{G}$，其加权邻接矩阵为 $\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}$，如上所述；其佩龙-弗罗贝尼乌斯特征值为 $\lambda_{(\omega,\beta)}$，左特征向量和右特征向量分别为$\psi^{(\omega,\beta)}$ 和 $\eta^{(\omega,\beta)}$。那么，$\mathcal{G}$ 上的加权最大熵随机游走是马尔可夫链 (3.23)，

$$\mathcal{S}_{\mathrm{MERW}}^{(\omega,\beta)}(x,y)=\hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x,y)=\frac{1}{\lambda_{(\omega,\beta)}}\frac{\eta^{(\omega,\beta)}(y)}{\eta^{(\omega,\beta)}(x)}\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x,y)\,. \tag{3.28}$$

加权最大熵随机游走具有以下性质。

(1) 由 $\hat{}\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}$ 诱导的路径空间上的概率是玻尔兹曼分布

$$\mathbb{P}_{(\omega,\beta),\ell}(\gamma)=\frac{1}{Z_{\omega,\ell}(\beta)}\psi^{(\omega,\beta)}(x_0)\eta^{(\omega,\beta)}(x_\ell)\exp(-\beta\,\mathbb{E}(\gamma))\,, \tag{3.29}$$

对于路径 $\gamma=e_0e_1\dots e_\ell$，其中 $s(e_i)=x_i$ 且 $t(e_i)=x_{i+1}$，其能量为

$$\mathbb{E}(\gamma):=\sum_{i=0}^{\ell}\omega(x_i,x_{i+1})\,,$$

配分函数为

$$Z_{\omega,\ell}(\beta)=\sum_\gamma\psi^{(\omega,\beta)}(x_0)\eta^{(\omega,\beta)}(x_\ell)\exp(-\beta\,\mathbb{E}(\gamma))=\lambda_{(\omega,\beta)}^{\ell}\,,$$

其中求和遍历长度为 ℓ 的路径 γ。

(2) 长度为 ℓ 的路径上玻尔兹曼分布的自由能 $\mathbb{F}_{\mathbb{P}_{(\omega,\beta),\,\ell}}$ 满足

$$\lim_{\ell\to\infty}\frac{\mathbb{F}_{\mathbb{P}_{(\omega,\beta),\ell}}}{\ell}=\mathbb{F}(\mathcal{S}_{\mathrm{MERW}}^{(\omega,\beta)})+\beta^{-1}\mathrm{Sh}(\pi^{(\omega,\beta)})\,, \tag{3.30}$$

其中

$$\mathbb{F}_{\mathbb{P}_{(\omega,\beta),\ell}}:=-\beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\beta,\omega),\ell})+\sum_\gamma\mathbb{P}_{(\beta,\omega),\ell}(\gamma)\,\mathbb{E}(\gamma)\,, \tag{3.31}$$

且极限 (3.30) 给出最优值

$$\lim_{\ell\to\infty}\frac{\mathbb{F}_{\mathbb{P}_{(\omega,\beta),\ell}}}{\ell}=-\beta^{-1}\log\lambda_{(\omega,\beta)}\,. \tag{3.32}$$

证明。论证步骤与命题 3.4 相同，只需将邻接矩阵 $\mathcal{K}_{\mathcal{G}}$ 替换为加权版本 $\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x,y)=\exp(-\beta\,\omega(x,y))\mathcal{K}_{\mathcal{G}}(x,y)$。$\hat{}\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}$ 在路径上诱导的分布为

$$\mathbb{P}_{(\omega,\beta),\ell}(\gamma)=\pi^{(\omega,\beta)}(x_0)\hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x_0,x_1)\cdots\hat{\mathcal{K}}_{\mathcal{G}}^{(\omega,\beta)}(x_{\ell-1},x_\ell)$$

$$= \frac{1}{\lambda_{(\omega,\beta)}^{\ell}} \psi^{(\omega,\beta)}(x_0)\eta^{(\omega,\beta)}(x_\ell)\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x_0,x_1)\cdots\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)}(x_{\ell-1},x_\ell) = \frac{1}{\lambda_{(\omega,\beta)}^{\ell}} \psi^{(\omega,\beta)}(x_0)\eta^{(\omega,\beta)}(x_\ell)e^{-\beta\mathbb{E}(\gamma)}\,.$$

要证明这与 (3.29) 一致，只需证明 $Z_{\omega,\ell}(\beta) = \lambda_{(\omega,\beta)}^{\ell}$。如引理 3.6 中所示，这可由以下内容推出：

$$Z_{\omega,\ell}(\beta) = \sum_{x,y} \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)(\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)})^{\ell}(x,y) = \lambda_{(\omega,\beta)}^{\ell} \sum_{x} \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(x) = \lambda_{(\omega,\beta)}^{\ell}\,.$$

长度为 ℓ 的路径集合上分布$\mathbb{P}_{(\omega,\beta),\ell}$ 的自由能为

$$\mathbb{F}_{\mathbb{P}_{(\omega,\beta),\ell}} = -\beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\beta,\omega),\ell}) + \sum_{\gamma} \mathbb{P}_{(\beta,\omega),\ell}(\gamma)\,\mathbb{E}(\gamma)$$

其中

$$-\beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\beta,\omega),\ell}) = \beta^{-1}\sum_{x,y} \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)\frac{(\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)})^{\ell}(x,y)}{\lambda_{(\omega,\beta)}^{\ell}}\,\log(\frac{\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)e^{-\beta\mathbb{E}(\gamma)}}{\lambda_{(\omega,\beta)}^{\ell}})$$

利用命题 3.4 中的结论，沿路径 γ 的边 e_i，每个 $\mathcal{K}_{\mathcal{G}}(x_i,x_{i+1}) = 1$，因此得到

$$-\beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\beta,\omega),\ell}) = \beta^{-1}\sum_{x,y} \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)\frac{(\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)})^{\ell}(x,y)}{\lambda_{(\omega,\beta)}^{\ell}}\,(-\beta\mathbb{E}(\gamma) + \log(\frac{\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)}{\lambda_{(\omega,\beta)}^{\ell}}))\,,$$

而

$$\sum_{\gamma} \mathbb{P}_{(\beta,\omega),\ell}(\gamma)\,\mathbb{E}(\gamma) = \sum_{x,y} \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)\frac{(\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)})^{\ell}(x,y)}{\lambda_{(\omega,\beta)}^{\ell}}\,\mathbb{E}(\gamma)\,.$$

由此得到

$$\mathbb{F}_{\mathbb{P}_{(\omega,\beta),\ell}} = \beta^{-1}\sum_{x,y} \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)\frac{(\mathcal{K}_{\mathcal{G}}^{(\omega,\beta)})^{\ell}(x,y)}{\lambda_{(\omega,\beta)}^{\ell}}\log(\frac{\psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(y)}{\lambda_{(\omega,\beta)}^{\ell}})\,.$$

$$= \beta^{-1}(\sum_{x} \psi^{(\omega,\beta)}(x)\eta^{(\omega,\beta)}(x)\log(\psi^{(\omega,\beta)}(x)) + \sum_{y} \psi^{(\omega,\beta)}(y)\eta^{(\omega,\beta)}(y)\log(\eta^{(\omega,\beta)}(y)) - \beta^{-1}\log(\lambda_{(\omega,\beta)}^{\ell})$$

$$= -\beta^{-1}(\mathrm{Sh}(\pi^{(\omega,\beta)}) + \ell\log\lambda_{(\omega,\beta)})\,,$$

使得 $\mathbb{F}_{\mathbb{P}_{(\omega,\beta),\ell}}/\ell$ 的极限给出最优值 $\min_{\mathcal{S}} \mathbb{F}(\mathcal{S}) = -\beta^{-1}\log(\lambda_{(\omega,\beta)})$。 □

3.5. 熵优化的语言学含义。对于 $\mathcal{G}_{n,A}$ 上的合并动力学，我们考察合适的基下合并操作的矩阵表示 $\hat{\mathcal{K}}^{(A,n)}$；如上所述，在该基下它是一个霍普夫代数马尔可夫链。正如本节所见，合并动力学的形式 $\hat{\mathcal{K}}^{(A,n)}$ 恰好是合并图 $\mathcal{G}_{n,A}$ 上的最大熵随机游走。

因此，这表明合并作用于工作空间时具有一个特别有趣的性质；这种作用被建模为霍普夫代数马尔可夫链：合并是信息最大化的，也就是说，它使同一有向图上所有可能的马尔可夫链的熵率 $\mathrm{Sh}(\mathcal{S})$ 达到最大。

正如本节所示，霍普夫代数马尔可夫链 $\hat{\mathcal{K}}^{(A,n)}$ 会在图 $\mathcal{G}_{n,A}$ 的有向路径集合上诱导出均匀分布。根据命题 3.4，我们知道，工作空间 F 和 F' 之间长度为 ℓ 的路径 γ 上的概率分布等于

$$\frac{\psi(F)\eta(F')}{\lambda^{\ell}}\,.$$

这意味着，例如，如果我们从给定的词汇材料（初始工作空间 F）开始进行合并推导，并得到一个结果工作空间 F'（比如，一个完整成形的句子

$F' = T$ 由单个句法对象构成），那么在马尔可夫链中，所有长度相同的可行推导（图 $\mathcal{G}_{n,A}$ 中从 F 到 F' 的有向边路径）具有相同的概率。这是一种比 “ 合并具有马尔可夫性 ” 这一通常观察更强的马尔可夫性质。这意味着推导的历史并不重要，只有起点、终点和长度决定该推导的可能性，而具体历史在任何方面都不起作用。

事实上，合并应满足这种更强的马尔可夫性质，这一点已由乔姆斯基在 [8], 第20页指出；他在那里谈到“严格意义上具有马尔可夫性的推导，超越通常的马尔可夫性质”。这一性质被描述为：在句法结构形成的过程中，推导的当前状态不包含推导的历史。我们在此将其理解为这样一种性质：推导（路径）的概率只取决于其结果，而不取决于路径本身（即历史）。

我们将在§11中讨论最小产出及其他最优性约束的影响，但这里首先要指出，动力学 $\hat{\mathcal{K}}^{(A,n)}$ 所满足的路径上的均匀分布，已经确保了这种更严格的马尔可夫性质成立；按照这一性质，推导的历史不会对动力学产生影响（除了其长度之外，这是不可避免的，因为出于最优性和收敛性的考虑，应当优先选择最小长度的推导，而不是较长的推导）。

还有一种与香农熵相关的优化性质，在评估语言学中通常考虑的成本函数（最小搜索和资源限制）及其对动力学的影响时，它将提供一条重要的补充信息。我们将说明，只有同时纳入一种信息优化性质，成本函数才能发挥预期作用，抑制侧向合并对动力学的影响，并确保句法结构形成过程通过外部合并收敛至连通性最高的工作空间（单棵树）。我们将在第 13 节回到这一问题。

本节讨论的自由能优化（而非熵优化），将在我们探讨结合成本函数优化的合并动力学时发挥作用；届时，最大化熵的马尔可夫链 $\hat{\mathcal{K}}^{(A,n)}$ 会导出最小化自由能的马尔可夫链。

4. Decomposing the Hopf algebra Markov chain

我们将分别分析合并图的不同部分以及不同的合并操作，以此考察合并霍普夫代数马尔可夫链的动力学。在语言学模型中，外部合并负责结构形成（包括 θ 角色指派；我们将在第 15 节回到这一问题），而内部合并则通过移位实现转换。侧向合并被用来解释某些特殊的语言现象（例如头对头移位），相关讨论见 [26]。侧向合并作为最简主义组成部分的存在及必要性颇具争议。我们将在第 6 节详细讨论它在合并动力学中的作用，以及仅考虑外部合并和内部合并时会发生哪些变化。

我们分析动力学不同部分的方法，是将图 $\mathcal{G}_{n,A}$ 的顶点集合 $V(\mathcal{G}_{n,A})$ F 分成若干子集 $V_{\wp,n,A}$，并以整数 n 的一个划分 $\wp \in \mathcal{P}'(n)$ 为这些子集命名，其中 $\mathcal{P}'(n)$ 是式 (2.1) 中所述的划分子集。接着，对于每个 $\wp \in \mathcal{P}'(n)$，我们考察顶点集 $V_{\wp,n,A}$ 上仅包含 IM 箭头所对应边的子图 $\mathcal{G}_{n,A,\wp}$。我们分别研究每个图 $\mathcal{G}_{n,A,\wp}$ 内部的动力学。我们还考察图 $\mathcal{G}_{\mathcal{P}'(n)}$：其顶点为划分 $\wp \in \mathcal{P}'(n)$，并且当 $\mathcal{G}_{n,A}$ 中某个属于 $V_{\wp,n,A}$ 的顶点与某个属于 $V_{\wp',n,A}$ 的顶点之间存在有向边时，图中就有一条有向边 $\wp \to \wp'$。图 $\mathcal{G}_{\mathcal{P}'(n)}$ 上的动力学将提供剩余的信息，帮助我们理解图 $\mathcal{G}_{n,A,\wp}$ 上的动力系统如何在图 $\mathcal{G}_{n,A}$ 上的完整合并动力学中相互交织。我们将看到

这一策略以自然的方式区分了合并的不同语言形式，并解释了它们相互作用对动力学的影响。

命题 *4.1*。令 $\mathfrak{T}_{\mathcal{SO}_0,k}$ 表示以 $\mathcal{SO}_0$ 中的元素为叶子装饰、具有 k 个叶子的非平面完全二叉有根树集合。令 $\mathcal{P}(n)$ 表示所有划分的集合，并令 $\mathcal{P}'(n) = \mathcal{P}(n) \setminus \{1, 1, \ldots, 1\}$，如 *(2.1)* 所示。考虑顶点集 $V(\mathcal{G}_{n,A}) = \sqcup_{\wp \in \mathcal{P}'(n)} V_{\wp,n,A}$ 的分解 *(2.2)*。令 $\mathcal{G}_{n,A,\wp}$ 表示 $\mathcal{G}_{n,A}$ 的子图，其顶点集为 $V_{\wp,n,A}$，且仅包含由 *IM* 箭头构成的边。*EM* 箭头将 $\mathcal{G}_{n,A,\wp}$ 的顶点映射到 $\mathcal{G}_{n,A,\wp'}$ 的顶点，其中 $\wp'$ 是形如 $\wp' = \{k_i + k_j\} \cup \{k_\ell\}_{\ell \neq i,j}$ 的划分，某个 $i \neq j \in \{1, \ldots, r\}$。如果划分 $\wp$ 包含一个 *2* 和一个 *1*，则存在一些极小 *SM* 箭头，将集合 $V_{\wp,n,A}$ 映射到其自身。所有其他极小 *SM* 箭头都将 $\mathcal{G}_{n,A,\wp}$ 的顶点映射到 $\mathcal{G}_{n,A,\wp'}$ 的顶点，其中 $\wp' \neq \wp$ 是形如 $\wp' = \{2, k_i - 1, k_j - 1\} \cup \{k_\ell\}_{\ell \neq i,j}$ 或形如 $\wp' = \{2, k_i - 2\} \cup \{k_\ell\}_{\ell \neq i}$ 的划分，分别对应某个 $i \neq j \in \{1, \ldots, r\}$ 或 $i \in \{1, \ldots, r\}$。因此，在所有这些情况下，$\mathcal{G}_{n,A,\wp}$ 都是顶点集 $V_{\wp,n,A}$ 上的诱导子图。

证明。给定引理 2.1 中的顶点集分解，我们得到一个映射 $p : \mathfrak{F}_{A,n} \to \mathcal{P}'(n)$，它将每个 F 映射到相应的划分：将 n 叶子划分为各连通分量中的叶子。该映射的值域由 $\mathcal{P}'(n)$ 构成，其中不计入划分 $n = 1 + 1 + \cdots + 1$，因为如前所述，我们已将仅由叶子构成且没有边的森林从动力学中移除，因为它始终是一个暂态状态。顶点集 $V_{\wp,n,A}$ 是该映射下的原像 $V_{\wp,n,A} = p^{-1}(\wp)$。在图 $\mathcal{G}_{n,A}$ 的顶点 F 处，每个 IM 箭头都作用于森林 F 的一个连通分量 T_i 上（在其他分量上作用为恒等映射），并保持每个分量的叶子数 $k_i = \#L(T_i)$ 不变。因此，IM 变换保持该划分不变，即有 $p \circ \mathrm{IM} = p$，所以集合 $V_{\wp,n,A}$ 在 IM 作用下保持不变，换言之，IM 箭头是图 $\mathcal{G}_{n,A,\wp}$ 的边。若工作空间的连通分量不包含樱桃树与单叶的并集 $\mathfrak{M}(\alpha_i, \alpha_j) \sqcup \alpha_k$（该划分 $\wp$ 不含 2 和 1），则图 $\mathcal{G}_{n,A,\wp}$ 与图 $\mathcal{G}_{n,A}$ 在同一顶点集 $V_{\wp,n,A}$ 上的诱导子图一致。事实上，在这种情况下，只有 IM 箭头能够保持 $\wp$ 不变。EM 箭头总是通过将两个连通分量嫁接到一个公共根上，使连通分量的数量减少一个，从而将 n 的划分 $\wp = \{k_1, \cdots, k_r\}$ 改为 n 的划分，其中两项 k_i, k_j 被一项 $k_i + k_j$ 取代。能够保持 $\wp$ 不变的最小 SM 箭头，仅有作用于两个分量上的那些箭头，这两个分量分别是一个樱桃和一个单叶，形式可以是以下两种之一

$$\overset{\frown}{\alpha_i \quad \alpha_j} \sqcup \alpha_k \mapsto \overset{\frown}{\alpha_i \quad \alpha_k} \sqcup \alpha_j$$

$$\overset{\frown}{\alpha_i \quad \alpha_j} \sqcup \alpha_k \mapsto \overset{\frown}{\alpha_j \quad \alpha_k} \sqcup \alpha_i \,.$$

所有最小 SM 箭头都会生成一棵新的樱桃树，因此会向划分中添加一项 2 $\wp$；所以，除非某个先前的单叶分量通过与另一个叶子合并到新的樱桃中而被消除，否则连通分量的数量将减少：要么两个项 k_i，k_j 减少为 $k_i - 1$，$k_j - 1$（包括 k_i 或 k_j 中恰有一个等于 1 的情况），要么单个项 $k_i \geq 3$ 减少为 $k_i - 2$。因此，只有当森林 $F \in V_{\wp,\ n,\ A}$ 的分量包含一棵樱桃树和一个单叶时，IM 图 $\mathcal{G}_{n,\ A,\ \wp}$ 才小于顶点集为 $V_{\wp,\ n,\ A}$ 的诱导子图。另一方面，所有 IM 箭头都会作为某个 $\mathcal{G}_{n,\ A,\ \wp}$ 的边出现。□

这一分解尤其意味着，如果我们想研究内部合并动力学，就可以每次考察一张图 $\mathcal{G}_{n,A,\wp}$。

5. The Internal Merge dynamics

我们在此讨论内部合并动力学的长期行为及其向平稳分布的收敛。

定义 5.1. 令 $\mathcal{G}^{\rm IM}_{n,A}$ 为顶点集为 $\mathfrak{F}_{A,n}$ 且边为 *IM* 箭头的图（即移除 *EM* 和 *SM* 箭头后的图 $\mathcal{G}_{n,A}$ ）。我们将 $\mathcal{G}^{\rm IM}_{n,A,\wp}$ 记作 $\mathcal{G}^{\rm IM}_{n,A}$ 的诱导子图，其顶点集为 $V_{\wp,n,A}$。

5.1. 连通分量。 我们首先考察由单棵树构成的工作空间上的内部合并动力学。

命题 5.2. 图 $\mathcal{G}^{\rm IM}_{n,A,\{n\}}$ （其中 $\wp=\{n\}$，顶点为由单棵树构成的工作空间 $F=T\in\mathfrak{F}_{A,n}$）是连通的。

证明。 该图有 $(2n-3)!!$ 个顶点，分别对应 T 的所有可能树拓扑（即带标号叶子的非平面树）。我们需要证明它们都属于同一个连通分量。满二叉根树 T 是一棵梳状树 当且仅当它只含有一个樱桃。在这种情况下，如果 IM 提取任意一个叶子，就会生成另一棵梳状树，其叶子标签经过置换；事实上，通过这种方式可以得到标签的所有排列。因此，所有梳状树都属于同一个连通分量。由此，要证明 $\mathcal{G}^{\rm IM}_{n,A,\{n\}}$ 连通，只需证明：对于任意一棵 T，其 n 个叶子由 A 中的元素标记，且它不是梳状树，都存在一条由 IM 箭头组成的路径，每次提取一个叶子，将 T 连接到某棵具有 n 个叶子的梳状树，其标签由 A 中的元素构成。我们可以用归纳法证明，从 $n=4$ 开始（对于 $n=3$，只有梳状树），此时对非梳状拓扑中的某个叶子执行一次 IM 操作即可生成一棵梳状树。假设对于 $m<n$ 个叶子的情况，结论成立。如果 T 不是梳状树，那么它就有多个樱桃。执行一次 IM 操作，提取其中一个樱桃的两个叶子之一，记作 α。所得树中，该叶子 α 直接连接到根，而原本是 T 根的顶点 v 成为它的姐妹顶点，现在则是 $T/^d\alpha$ 的根顶点。树 $T/^d\alpha$ 有 $n-1$ 个叶子，因此存在一条 IM 箭头路径，每次提取一个叶子，将 $T/^d\alpha$ 连接到一棵梳状树。于是，在树 $\mathfrak{M}(\alpha,T/^d\alpha)$ 上按相同顺序对相同叶子执行同一系列内部合并操作，也会生成一棵梳状树，因为被提取并在根处合并的叶子序列，会延续从 α 开始的梳状序列。□

注意，这个论证证明了连通性，但没有证明强连通性。事实上，图$\mathcal{G}^{\rm IM}_{n,A,\{n\}}$ 也是强连通的：这将由命题 5.5 中给出的更一般论证得出。

命题 5.3。 对于具有 $\wp=\{k_1,\dots,k_1,\dots,k_r,\dots k_r\}$ *with*$n=a_1k_1+\cdots+a_rk_r$形式的划分 $\wp\in\mathcal{P}'(n)$，图 $\mathcal{G}^{\rm IM}_{n,A,\wp}$ 由若干彼此分离的连通分量组成。连通分量的数量等于式 (2.8) 的广义多项式系数，

$$b_0(\mathcal{G}^{\rm IM}_{n,A,\wp})=\Upsilon_{\wp,n}=\frac{1}{a_1!\cdots a_r!}\binom{n}{\underbrace{k_1,\dots,k_1}_{a_1\text{-}times}\dots,\underbrace{k_r,\dots,k_r}_{a_r\text{-}times}}.$$

我们将图 $\mathcal{G}^{\rm IM}_{n,\,A,\,\wp}$ 的连通分量集合记为 $C_\wp$，并将分量 $\sigma\in C_\wp$记为$\mathcal{G}^{\rm IM}_{n,\,A,\,\wp,\,\sigma}$。集合 $C_\wp$ 可以通过对称群 S_n 中循环类型为 $\wp$的共轭类 $\mathrm{Conj}_\wp$ 取商得到，其关系为

$$\Upsilon_{\wp,n}\cdot\prod_{i=1}^r((k_i-1)!)^{a_i}=\#\mathrm{Conj}_\wp=\frac{n!}{\prod_{i=1}^r a_i!\;k_i^{a_j}}.$$

每个连通分量 $\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp,\,\sigma}$ 所包含的顶点数等于

$$\prod_{i=1}^{r}((2k_i-3)!!)^{a_i}\ .$$

证明。内部合并作用于一个工作空间 $F=T_1\sqcup\cdots\sqcup T_r$ 时，每次分别作用于工作空间的一个连通分量 T_i，因此，对工作空间的作用完全由对树的作用决定。具体而言，该作用通过保留每个集合 $L(T_i)=L(\mathfrak{M}(T_v,T_i/T_v))$，从而保留划分 $L(F)=\sqcup_i L(T_i)$。这意味着，森林 $F\in V_{\wp,n,A}$ 经 IM 箭头作用后，总会映射为 $V_{\wp,n,A}$ 中具有相同 $\wp$的其他森林。叶子集合 $L(T_i)$ 上标签集的划分也受 IM 保持（与侧向合并和外部合并不同）：假设给定的 $F\in V_{\wp,n,A}$ 对于 $\wp=\{k_1,\dots,k_r\}$，将标签集 $A=\{\alpha_1,\dots,\alpha_n\}$ 划分为若干子集，各子集的大小为 k_i，并分别指派给 $L(T_i)$的叶子集合，这些叶子集合属于 F的分量 T_i。那么，内部合并会保留这一标签划分，不会混合指派给工作空间不同分量的叶标签。因此，这些集合 A 的划分沿着集合 $V_{\wp,n,A}$上内部合并箭头的每条有向路径保持不变。这意味着，由标签指派决定的每个此类轨道，都是图 $\mathcal{G}_{n,A,\wp}$的一个不同连通分量。

如果划分 $\wp$ 是 $n=k_1+\cdots+k_r$ 这种形式，且所有 $k_i\le 2$ 都如此，则不存在内部合并箭头（我们不考虑提取紧邻根下方顶点处某个可访问项的 IM 箭头，因为那只是恒等映射，所以单叶和樱桃树都没有 IM动力学）。因此，在这种情况下，$\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp}$ 的每个顶点都是一个连通分量。

当划分 $\wp$ 中至少有一个 k_i 满足 $k_i\ge 3$ 时，$\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp}$ 中就存在非平凡 IM箭头。对于工作空间各分量上标签集 A 的任意划分，根据命题 5.2，$\prod_i(2k_i-3)!!$ 种不同的树拓扑都属于同一个连通分量。

因此，$\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp}$ 的连通分量与工作空间各分量上的标签集 A 的划分一一对应，因此共有 $\Upsilon_{\wp,\,n}$ 个分量。

考虑具有由划分 $\wp$ 确定的循环结构的排列 $\sigma\in S_n$，即包含数量为 a_i 的 k_i 阶循环。循环类型 $\wp$ 确定了对称群 S_n 中的一个共轭类 $\mathrm{Conj}_\wp$。在计算循环类型为 $\wp$ 的共轭类 $\mathrm{Conj}_\wp$ 中的元素数量时，如式（2.4）中的多项式系数 $\mu_{\wp,\,n}$ 这一因子，计算了分配给各个循环的元素的选择数。每种选择都会产生 $k_i!/k_i=(k_i-1)!$ 种标签分配，并按循环置换取模计数。计数时还要对大小为 k_i 的 a_i 个循环的置换取模，因此得到 $\#\mathrm{Conj}_\wp=\Upsilon_{\wp,\,n}\cdot\prod_i((k_i-1)!)^{a_i}=\#C_\wp\cdot\prod_i((k_i-1)!)^{a_i}$。

对于每个分量 $\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp,\,\sigma}$，都可以从工作空间 $F=T_1\sqcup\cdots\sqcup T_r$ 中的树 T_i 的所有可能树拓扑中进行选择。每个分量 T_i 对应一个乘积 $((2k_i-3)!!)^{a_i}$，它确定了每个连通分量中的顶点数。□

5.2. **分量的强连通性**。随后，我们可以证明每个连通分量上的内部合并动力学具有更强的结果 $\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp,\,\sigma},\sigma\in C_\wp$。

命题 5.4。设 $\wp\in\mathcal{P}'(n)$ 是一个划分 $\wp=\{k_1,\dots,k_r\}$，其中至少有一个 $k_i\ge 3$，因此存在非平凡的 *IM* 箭头。于是，图 $\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp}$的每个连通分量 $\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp,\,\sigma}$ 都是强连通且非周期的，因此内部合并霍普夫代数马尔可夫链 $\hat{}\mathcal{K}^{(A,n,\wp,\sigma)}$在每个连通分量上都是遍历的。

证明。正如我们在命题 3.3 中提到的，遍历性由强连通性和非周期性推出。为了证明内部合并的霍普夫代数马尔可夫链 $\hat{}\mathcal{K}^{(A,n,\wp,\sigma)}$ 是遍历的，我们需要对命题 2.5 作进一步细化，以证明不仅图 $\mathcal{G}_{n,\,A}$ 是强连通且非周期的，

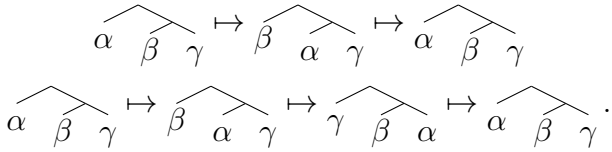
而且子图 $\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp}$ 的每个连通分量 $\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp,\,\sigma}$ 也都具有相同的性质。

对于给定的分量 $\mathcal{G}^{\mathrm{IM}}_{n,\,A,\,\wp,\,\sigma}$，我们考察动力学 $\hat{\mathcal{K}}^{(A,n,\wp)}$ 在该分量上的限制 $\hat{\mathcal{K}}^{(A,n,\wp,\sigma)}$。根据命题 2.3，我们知道，在每个$F \in V_{\wp,\,n,\,A,\,\sigma}$ 处，出射的 IM 边数量与入射的 IM 边数量相同，

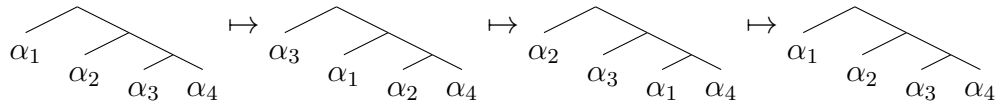
$$N^{\mathrm{out}}_{\mathrm{IM}}(F) = \sum_{i=1}^{c(F)} (2k_i - 4) = N^{\mathrm{in}}_{\mathrm{IM}}(F)\,, \tag{5.1}$$

其中 $F = T_1 \sqcup \cdots \sqcup T_r$with $\wp$ 的划分 $n = k_1 + \cdots + k_r$ 对应于 $k_i = \#L(T_i)$，并且为简化记号，我们假设 $T_1,\ \ldots,\ T_{c(F)}$ 是边集非空的分量，且$c(F) \leq r$。数值 k_i 以及 $c(F) = \#\{i \,|\, k_i > 1\}$ 由划分 $\wp$ 确定，因此对于固定的 $\wp$，它们对所有顶点 $F \in V_{\wp,\,n,\,A}$ 都相同。

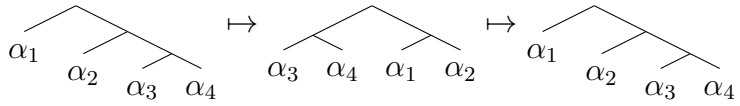
强连通性随后由以下事实推出：$\mathcal{G}_{n、\,A、\,\wp、\,\sigma}$ 中的每个顶点都满足 $\deg^{\mathrm{out}}(F) = \deg^{\mathrm{in}}(F)$，这意味着有向图 $\mathcal{G}_{n、\,A、\,\wp、\,\sigma}$ 是欧拉图，即存在一条有向环，恰好访问$\mathcal{G}_{n、\,A、\,\wp、\,\sigma}$ 的每条边一次。由此可推出强连通性，因为从任意顶点 F 出发，都可以沿着欧拉回路到达任意其他顶点 F'。通过细化命题 2.5 中用于确定内部合并操作显式环的论证，我们还可以看出，$\mathcal{G}_{n、\,A、\,\wp、\,\sigma}$ 必然是非周期的。只需考虑 $\wp = \{n\}$的情形，即由单个连通分量$T \in \mathfrak{T}_{\mathcal{SO}_0、\,n}$组成的工作空间 $F = T$。图 $\mathcal{G}_{n\,、\ A\,、\ \wp=\{n\}}$ 只有一个连通分量。如果我们证明这些图是非周期的，那么其他所有图 $\mathcal{G}_{n、\,A、\,\wp、\,\sigma}$ 也都将是非周期的，因为对于划分$\wp = \{k_1, \ldots, k_r\}$ 中的每个 $k_i \geq 3$，它们都包含我们在 $\mathcal{G}_{n、\,A、\,\{k_i\}}$中展示的环的副本。我们分别考虑 $n = 3$ 和 $n = 4$ 个叶子的情形，以及一般的 $n \geq 5$ 个叶子情形。我们证明，在所有这些情形中，总是存在一个阶数为 2 的内部合并环和一个阶数为 3 的内部合并环，因此环长的最大公约数等于 1。对于 $n = 3$，通过提取樱桃中的另一个叶子之一，我们得到环



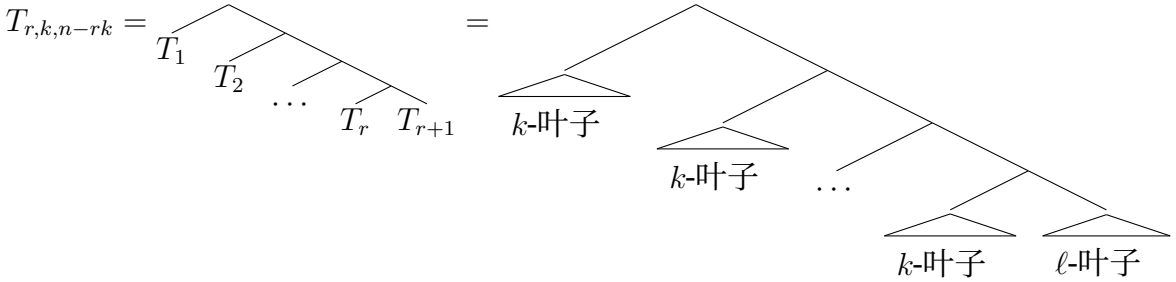
对于 $n = 4$ 的情形，我们同样有一个长度为 3 的循环



还有一个长度为 2 的循环（先提取 $\{\alpha_3, \alpha_4\}$ 樱桃，再提取单叶 α_1）



因此，$\mathcal{G}_{2,\,A\{2\}}$ 和 $\mathcal{G}_{3,\,A\{3\}}$ 都是非周期的。现在假设 $n \geq 5$。考虑以下形式的树：



其子树 T_i 对应于 $i=1$、.....、r，这些子树各有 k 个叶子，且树拓扑相同；此外还有一个子树 T_{r+1}，含有 ℓ 个叶子，其中 $n=rk+\ell$ 和 $\ell\geq 1$、$r\geq 2$ 和 $k\geq 1$。命题 2.5 中所考虑的树是特殊情形 $T_{n-1,\,1,\,1}$。接着，依次提取可访问项 T_r、T_{r-1}，以此类推，直到 T_1 的内部合并操作迭代会产生一个长度为 r 的循环。对于 $n\geq 5$，这些循环总是包含一个长度为 $r=2$ 的循环和一个长度为 3 的循环，因此循环长度的最大公约数仍为 1，非周期性总是成立。□

5.3. 均匀分布。 内部合并动力学的入度/出度具有相同的性质 (5.1)，这也给出了 $\mathcal{G}_{n,A,\wp,\sigma}^{\mathrm{IM}}$ 图上的平稳分布形式。

命题 5.5。 对于至少有一个 $k_i\geq 3$ 的划分，图 $\mathcal{G}_{n,A,\wp}^{\mathrm{IM}}$ 的连通分量 $\mathcal{G}_{n,A,\wp,\sigma}^{\mathrm{IM}}$ 是强连通的；对于任意 $\sigma\in C_\wp$，由内部合并在 $\mathcal{G}_{n,A,\wp,\sigma}^{\mathrm{IM}}$ 上的作用所给出的霍普夫代数马尔可夫链的部分 $\hat{\mathcal{K}}^{(A,n,\wp,\sigma)}$，其平稳分布是均匀分布。当所有 $k_i\leq 2$ 时，不存在 *IM* 箭头，且每个顶点都是一个分量。

证明。如命题 2.3 和 5.4 所示，在每个顶点 $F\in V_{\wp,n,A,\sigma}$ 处，出射 IM 边与入射 IM 边数量相同，如式 (5.1) 所示：$\deg^{\mathrm{out}}(F)=\deg^{\mathrm{in}}(F)$。注意，式 (5.1) 中的值对所有顶点 $F\in V_{\wp,n,A}$都相同，且仅取决于划分 $\wp$，

$$N_{\mathrm{IM}}^{\mathrm{out}}(F)=N_{\mathrm{IM}}^{\mathrm{in}}(F)=:d_\wp\,.$$

令 $\mathcal{K}^{(A,n,\wp,\sigma)}=(\mathcal{K}_{F\,F'}^{(A,n,\wp,\sigma)})$ 表示图 $\mathcal{G}_{n\,A\quad\sigma}^{|rmIM}{}_{,,\wp,}$ 的邻接矩阵。

对于一个所有顶点的入度和出度都相同、且均等于 $d_\wp$ 的有向图，将佩龙-弗罗贝尼乌斯定理应用于邻接矩阵，可得佩龙-弗罗贝尼乌斯特征值满足

$$\min_F\sum_{F'}\mathcal{K}_{F,F'}^{(A,n,\wp,\sigma)}\leq\lambda\leq\max_F\sum_{F'}\mathcal{K}_{F,F'}^{(A,n,\wp,\sigma)}\,,$$

其中 $\sum_{F'}\mathcal{K}_{F}^{(A,n,\wp,\sigma)}{}_{,F'}=\deg^{\mathrm{out}}(F)$，因此若对于所有 $F\in V_{\wp,n,A,\sigma}$，$\deg^{\mathrm{out}}(F)=d_\wp$，则有 $\lambda=d_\wp$ 是佩龙-弗罗贝尼乌斯特征向量，相应地，均匀分布 $\eta(F)=1$ 是佩龙-弗罗贝尼乌斯特征值

$$\sum_{F'}\mathcal{K}_{F,F'}^{(A,n,\wp,\sigma)}\eta(F')=d_\wp\,\eta(F)\,.$$

因此，霍普夫代数马尔可夫链中由内部合并给出的部分，其转移矩阵仅为

$$\hat{\mathcal{K}}_{F,F'}^{(A,n,\wp,\sigma)}=d_\wp^{-1}\cdot\mathcal{K}_{F,F'}^{(A,n,\wp)}\,,$$

由于无需重新缩放基，因为对所有 $F\in V_\wp$ 都有 $\eta(F)=1$ $_{,n,A}$。该马尔可夫链具有平稳分布

$$\sum_F\pi(F)\hat{\mathcal{K}}_{F,F'}^{(A,n,\wp,\sigma)}=\pi(F')$$

由均匀分布 $\pi(F)=1$ 给出（归一化为 $\pi(F)=1/\,\#V_{n,A,\wp}$），因为

$$d_\wp^{-1}\sum_F\mathcal{K}_{F,F'}^{(A,n,\wp,\sigma)}=\frac{\deg^-(F')}{d_\wp}=1\,,$$

因为对于所有 $F\in V_{\wp,n,A,\sigma}$，$\deg^+(F)=\deg^-(F)=d_\wp$。 □

命题 5.5 和 5.4 可以改述如下。

推论 5.6。均匀分布是内部合并霍普夫代数马尔可夫链$\hat{\mathcal{K}}^{(A,n,\wp,\sigma)}$在每个连通分量$\mathcal{G}_{n,A,\wp,\sigma}$ 的唯一平稳分布$\mathcal{G}_{n,A,\wp}$，并且工作空间集合$V_{\wp,n,A,\sigma}$ 上的任意初始分布，在$\hat{\mathcal{K}}^{(A,n,\wp,\sigma)}$ 的迭代作用下都会收敛到均匀分布。

这表明，单独的内部合并构成一个具有均匀平稳分布的遍历动力系统。这一事实凸显了内部合并具有尤为良好的动力学性质。

6. Internal and External Merge dynamics

本节和下一节将通过展示以下情形的结果，阐述最小侧向合并与外部合并箭头在动力学中所起的相反作用：一种情形是移除所有 SM 箭头，同时保留内部合并和外部合并箭头；另一种情形（将在下一节讨论）是移除外部合并箭头，同时保留内部合并和 SM 箭头。在这两种情况下，强连通性都会被破坏，图的顶点会分化为暂态和常返状态，而平稳分布仅支撑在常返状态上。两种情况下常返状态的性质互为补充：可以预见，没有 SM 时，动力学会转向连通性最强的结构；没有外部合并时，则会转向连通性最弱的结构。

仅涉及外部合并和内部合并的动力系统，正是语言学文献中通常考察的系统；相关文献认为（见 [7], [8], [9]），基于对某些成本函数的优化，应排除动力学中的侧向合并部分，这些成本函数将在 §11 中讨论。然而，从数学视角来看，并结合物理模型中的类似情形，可以认为（见 [26]），代价优化并非一种会排除 SM 的硬性结构约束，而是一种软约束：偏离最优状态越多，发生的可能性就越低，或者其作用越不显著。[26] 中指出，某些语言现象可以通过一种仅轻微违背最优性的 SM 形式来解释；这正是我们在此处的图$\mathcal{G}_{n,\ A}$ 中所讨论的最小 SM。

本节中，我们按照 [7], [8], [9], 的原始提案介绍合并动力学，其中图 $\mathcal{G}_{n,\ A}$ 中的 SM 箭头被完全移除。我们将展示，尽管在这种情况下，动力学确实会通过外部合并收敛到连通结构（由单棵树构成的工作空间），而树上的内部合并作用则是动力学中唯一的常返部分，但强连通性会遭到破坏，从而消除遍历性这一理想的动力学性质，以及动力学与初始位置无关这一性质；此外，我们前面讨论过的熵优化性质也会消失，因为它同样依赖于强连通性。我们将在 §11 和 §13 中进一步讨论，如何利用 [7], [8], [9] 和 [22] 提出的这类代价函数以及熵优化性质，通过纳入能够抑制侧向合并不良影响的代价权重来修正动力学，同时保留其理想的强连通性、遍历性和熵优化性质。

我们在此考虑子图 $\mathcal{G}_{n\ A}^{\mathrm{IM-EM}}{}_{,\ A} \subset \mathcal{G}_{n,\ A}$，其顶点集合相同，即 $V(\mathcal{G}_{n\ A}^{\mathrm{IM-EM}}{}_{,\ A}) = V(\mathcal{G}_{n,\ A}) = \mathfrak{S}_{A,\ n}$，但其边仅包含内部合并和外部合并箭头，不包含$\mathcal{G}_{n,\ A}$的侧向合并箭头。因此显然，这个子图 $\mathcal{G}_{n\ A}^{\mathrm{IM-EM}}{}_{,}$ 不满足强连通性。我们将由此得到的马尔可夫链称为 $\hat{\mathcal{K}}^{\mathrm{IM-EM},\ (A,n)}$。

6.1. 可约性与暂态。马尔可夫链中的暂态是指具有永不返回的非零概率的状态。

引理 6.1。由具有多个连通分量的工作空间给出的每个顶点 $F \in V(\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}_{,\,,\,})$（即，其像为 $\wp = p(F) \in \mathcal{P}'(n) \quad \wp \neq \{n\}$）都是 $\hat{}\mathcal{K}^{\mathrm{IM-EM},\,(A,n)}$ 的暂态。

证明。由于外部合并总会使工作空间的连通分量数减少一个，因此显然，状态 $F = T_1 \sqcup \cdots \sqcup T_r \in \mathfrak{F}_{A,\,n}$ 且满足 $r > 1$，也就是说，$\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}_{,\,,\,\wp=\{k_1,\ldots,k_r\}}$中的一个顶点，有非零概率通过某个外部合并箭头离开图 $\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}_{,\,,\,\wp=\{k_1,\ldots,k_r\}}$，并进入某个满足 $\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}_{,\,,\,\wp'=\{k_i+k_j,k_1,\ldots,\hat{}k_i,\ldots,\hat{}k_j,\,\ldots k_r\}}$的图，其中记号 $\hat{}\,k_i$ 表示条目 k_i 被移除。由于没有侧向合并箭头时，图 $\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}_{,}$ 中的每个箭头要么保持工作空间的连通分量数不变（内部合并箭头），要么使其减少一个（外部合并箭头），因此无法从 $\mathcal{G}^{\mathrm{IM-EM}}$中的任何顶点返回到 $\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}_{,,\wp=\{k_1,\ldots,k_r\}}$ 的一个顶点，因此 $\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}^{n,A,\wp'=\{k_i+k_j,k_1,\ldots,\hat{k}_i,\ldots,\hat{k}_j,\ldots k_r\}}_{,,\wp=\{k_1,\ldots,k_r\}}$ 的顶点是暂态的。对于 $\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}_{,,\wp=\{n\}}$ 的顶点，不存在出射 EM 箭头，因此该子图的顶点会通过 IM 箭头留在同一个 $\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}{}_{,,\wp=\{n\}}$ 内，而这些正是马尔可夫链 $\hat{}\mathcal{K}^{\mathrm{IM-EM},(A,n)}$ 唯一的常返状态。□

推论 6.2。霍普夫代数马尔可夫链 $\hat{}\mathcal{K}^{\mathrm{IM-EM},(A,n)}$ 是可约的。等价地，图$\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}$,不是强连通的。

证明。有限不可约马尔可夫链不可能有任何暂态，因此引理 6.1 蕴含 $\hat{}\mathcal{K}^{\mathrm{IM-EM},(A,n)}$ 是可约的。□

6.2. **通信类**。可约马尔可夫链可以分解为通信类：这与将底层有向图分解为强连通分量相同。

给定一个有向图 G，若两个顶点 $v, v' \in V(G)$ 满足以下条件之一：$v = v'$，或者 $v \neq v'$，并且存在一条从 v 到 v' 的边组成的有向路径，以及一条从 v' 到 v 的边组成的有向路径，则称它们是强连通的，或称彼此通信。如此定义的强连通性是 $V(G)$ 上的一种等价关系，其等价类就是 G 的强连通分量（或通信类）。

¯凝聚图 G 是通过将 G 的每个强¯连通分量收缩为一个顶点而得到的有向图。图 G 是一个有向无环图。一般而言，由此得到的 $\_G$ 是一个多重图，同一对 ¯ 顶点之间可能存在多条平行边。我们还可以将 G 中任意一组平行边合并为一条加权单边。

一个闭通信类是一个不含指向其他强连通分量的箭头的强连通分量，也就是说，其强连通性等价类是凝聚有向无环图的一个汇。

再次考虑图 $\mathcal{G}^{\mathrm{IM}}_{n}{}_{、A、\wp、\sigma}$以及 $\sigma \in C_\wp$，它们是 IM 图 $\mathcal{G}^{\mathrm{IM}}_{n}{}_{、A}$的连通分量。

下面关于 $\hat{}\mathcal{K}^{\mathrm{IM-EM}、\,(A,n)}$ 的通信类的描述，直接来源于对图 $\mathcal{G}^{\mathrm{IM}}_{n}{}_{、A、\,\wp、\,\sigma}$ 上的 IM动力学以及 EM 箭头的描述。

命题 6.3。图 $\mathcal{G}^{\mathrm{IM}}_{n}{}_{、A、\wp、\sigma}$，对于 $\wp \in \mathcal{P}'(n)$和 $\sigma \in C_\wp$，是图 $\mathcal{G}^{\mathrm{IM-EM}}_{n\,A}$ 、的强连通分量（马尔可夫链 $\hat{}\mathcal{K}^{\mathrm{IM-EM}、\,(A,n)}$ 的通信类）。凝聚图 $\bar{\mathcal{G}}^{\mathrm{IM-EM}}_{n\,A}$ 、是图 $\tilde{\mathcal{G}}^{\mathrm{EM}}_{\mathcal{P}'\,(n)}$，其顶点集为

$$V(\tilde{\mathcal{G}}^{\mathrm{EM}}_{\mathcal{P}'(n)}) = \{(\wp, \sigma) \,|\, \wp \in \mathcal{P}'(n)\,,\ \sigma \in C_\wp\}$$

并且从 $(\wp, \sigma)$ 到 $(\wp', \sigma')$ 有一条有向边

$$\wp = \{k_1, \dots, k_r\} \quad and \quad \wp' = \{k_i + k_j, k_1, \dots, \hat{k}_i, \dots, \hat{k}_j, \dots k_r\}$$

（记号同引理 *6.1*）对于某些 k_i、$k_j \in \wp$，且 σ' 是 σ 在诱导映射 $C_\wp \to C_{\wp'}$下的像。图 $\mathcal{G}_{n,\, A,\, \wp}^{\mathrm{IM}}$ 其中 $\wp = \{n\}$，顶点为由含有 n 个叶子的单棵树组成的工作空间，是唯一的闭通信类。

诱导映射 $C_\wp \to C_{\wp'}$ 由这样一个事实直接确定：A到 D（$\wp$）各方格的每种标签分配，都确定了到 D（$\wp'$）各方格的相应标签分配。

6.3. **闭通信类上的平稳分布**。这意味着，当我们不将任何侧向合并视为动力学的一部分时，$\hat{\mathcal{K}}^{\mathrm{IM-EM},\,(A,n)}$ 的极限分布仅支撑在连通结构上。这描述了合并作为一个结构形成过程：动力学通过外部合并朝连通结构演化，并远离具有多个连通分量的工作空间；而一旦达到连通结构，剩下的唯一动力学便是通过内部合并实现的移位转换。这确实与乔姆斯基在 [7], [8], [9], 中仅以外部合并和内部合并来描述合并动力学的通常说法一致。

命题 6.4。可约马尔可夫链 $\hat{\mathcal{K}}^{\mathrm{IM-EM},\,(A,n)}$ 的平稳分布是支撑在图 $\mathcal{G}_{n\,,\,A,\,\wp=\{n\}}^{\mathrm{IM}}$ 顶点上的均匀分布（工作空间由一棵含有 n 个叶子的树构成）。

证明。如果平稳分布 π 存在，它具有如下性质：

$$\pi(F) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} (\hat{\mathcal{K}}^{\mathrm{IM-EM},(A,n)})^N_{F,F'} \,, \tag{6.1}$$

对于同一通信类中的任意初始状态 F'。在可约情形下，$\pi(F) = 0$ 当且仅当F 是 $\hat{\mathcal{K}}^{\mathrm{IM-EM},(A,n)}$ 的暂态。因此，π 只能支撑在常返状态上；在此情形下，这些状态是子图 $\mathcal{G}_{n\,,A,\wp=\{n\}}^{\mathrm{IM}}$的顶点。事实上，可约马尔可夫链的平稳分布总是支撑在闭通信类上。在此情形下，这样的类只有一个，由 $\mathcal{G}_{n\,,A,\wp=\{n\}}^{\mathrm{IM}}$的顶点构成。此外，在这些顶点上，$\pi$ 应与图 $\mathcal{G}_{n\,,A,\wp=\{n\}}^{\mathrm{IM}}$ 上的内部合并动力学的平稳分布一致，因为在式（6.1）中同时选择 F, F' 为 $\mathcal{G}_{n\,,A,\wp=\{n\}}^{\mathrm{IM}}$ 中的顶点，会使 $(\hat{\mathcal{K}}^{\mathrm{IM-EM},(A,n)})^N_{F,F'}$ 中的所有非平凡项都成为 IM 箭头，因此极限必定是 $\mathcal{G}_{n\,,A,\wp=\{n\}}^{\mathrm{IM}}$上 IM动力学的平稳分布，也就是均匀分布。□

6.4. **IM–EM结构形成模型：语言学方面**。从语言学角度来看，本节所述的动力学及命题 6.4 的结果总结，为句法中的结构形成提供了一种令人十分满意的观点。该观点通过自由对称合并实现，显示出系统会收敛到连通结构（完全形成的句子），而内部合并动力学则以遍历方式作用于这些连通结构，不偏袒任何特定的可用结构子集（均匀平稳分布）。这确实反映了乔姆斯基的模型 [7] 和 [8], 中所述的内容；在该模型中，侧向合并因不符合最优性而被排除。

然而，正如我们在本节开头指出的，这存在一个问题：消除侧向合并箭头显得不自然，因为这些合并操作显然是霍普夫代数马尔可夫链中合并的自然定义的一部分，即作为组合

$$\mathfrak{M}_{S,S'} = \sqcup \circ (\mathcal{B} \otimes \mathrm{id}) \circ \delta_{S,S'} \circ \Delta \,,$$

其中使用了余积的非本原项，因此没有 代数上的 理由来消除它们；代价计数只能使这些合并操作相较于外部合并和内部合并处于次要地位，但它们仍然存在。

如 [22], 中所述，可以用本原部分 Δ^{P}替换余积 Δ ；本原部分只对工作空间进行划分，而不提取可访问项。如此一来，同一公式便只能给出外部合并，而不能给出侧向合并。另一方面，这样做又无法与内部合并中保留对可访问项的提取相一致，也无法继续将内部合并和外部合并视为同一操作的不同情形（内部合并涉及一个$\mathfrak{M}_{S,\ 1}$操作）。因此，仅从霍普夫代数合并马尔可夫链 $\mathcal{K}$ 中移除 SM 箭头，并不能直接由其底层代数结构得到合理依据。

在乔姆斯基的最简主义表述中，选择外部合并和内部合并而非侧向合并，被表述为最优性（即最小化某种自然的成本函数），其中外部合并和内部合并是最优的，而侧向合并则不是。在关于合并的语言学文献中，本文语境下讨论了不同类型的最优性：最小搜索，以及诸如最小产出之类的资源限制。这些内容在 [22]中通过数学模型进行了分析。 [26], 还进一步详细讨论了这些不同的成本函数与侧向合并，并更精确地说明了最小搜索和资源限制在约束侧向合并可能形式方面所起的作用。

在我们的数学框架中，要通过代价函数处理不同权重，方法是将合并图 $\mathcal{G}_{n,\ A}$ 的邻接矩阵替换为加权版本，其中的权重纳入代价函数；做法与我们在 §3.4 中分析的情形类似。随后，我们可以论证，这种加权形式的合并动力学能更好地近似本节所述的外部合并–内部合并动力学，同时代价函数会削弱侧向合并的影响。这个想法看起来很自然，而且在 [26], 中计算的简单 $n = 3$ 叶子示例中也很容易得到验证；但对于任意 n 的一般情形，问题要微妙得多。我们将在 §11 中再讨论代价函数，以及其对 §11 和 §13 中合并动力学的影响。

7. Internal and Sideward Merge dynamics

现在，为了进行比较，我们来考察一种情形：如果只有内部合并和侧向合并箭头，而没有外部合并，会发生什么。当然，这纯属假设，并不对应任何可行的语言学模型，因为它显然无论如何都无法形成结构。不过，简要讨论这一情形仍有助于为下一节中对外部合并与侧向合并的直接比较做准备。

关于 Merge 动力系统中的侧向合并部分，我们首先可以观察到，它的作用类似于外部合并，但方向“相反”，也就是推动动力学朝着结构更加不连通的方向发展。因此，如果保持动力学中的内部合并部分不变，并且只考虑 SM 箭头而不考虑 EM 箭头，那么整体图景就与我们在 IM 和 EM 中看到的相似，但方向相反。与 §2 中引入的完整合并动力学 $\mathcal{K}$ 一样，我们假设只包含极小侧向合并箭头（也就是只提取两个单叶的箭头）。

命题 7.1. 令 $\mathcal{G}_{n}^{\mathrm{IM-SM}}{}_{,A} \subset \mathcal{G}_{n,A}$ 为顶点集相同的子图，其边仅由 *IM* 箭头和最小 *SM* 箭头构成（不包含任何 *EM* 箭头）。相应的马尔可夫链 $\hat{}\mathcal{K}^{\mathrm{IM-SM},(\mathrm{A,n})}$ 是可约的。其强连通分量（通信类）由以下情形给出。

- 对于不包含任何一对 $\wp \in \mathcal{P}'(n)$ 的划分，其强连通分量为图 $\mathcal{G}_{\text{内部合并}n,A,\wp,\sigma}$，其中有 $\sigma \in C_{\wp}$;

• 对于包含若干 $\{2,1\}$ 对的划分，其强连通分量为图 $\mathcal{G}_{\text{内部合并–侧向合并}\ n,A,\wp,\tilde{\sigma}}$ 其中 $\tilde{\sigma} \in C_{\tilde{\wp}}$ 适用于从 $\wp$ 中移除所有 $\{2,1\}$ 对后得到的划分（某个 $\mathcal{P}'(m)$ 中有 $m < n$）。

唯一的闭通信类由 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ \wp=\{2,1,\dots,1\}}$ 上的诱导子图给出。该图没有 *IM* 箭头，且只有最小 *SM* 箭头。平稳分布是在子图 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ \wp=\{2,1,\dots,1\}}$ 的顶点上支撑的均匀分布。凝聚图 $\bar{\mathcal{G}}^{\mathrm{IM-SM}}_{n\,A}$ 是图 $\tilde{\mathcal{G}}^{\mathrm{SM}}_{\mathcal{P}'(n)}$，其顶点集为

$$V(\tilde{\mathcal{G}}^{\mathrm{SM}}_{\mathcal{P}'(n)}) = \{(\wp,\sigma)\,|\,\wp \in \mathcal{P}'_{2,1}(n)^c\,\sigma \in C_\wp\} \cup \{(\wp,\sigma)\,|\,\wp \in \mathcal{P}'_{2,1}(n)\,\sigma \in C_{\tilde{\wp}}\}$$

其中 $\mathcal{P}'_{2,1}(n)$ 是含有 $\{2,1\}$ 对的划分集，$\mathcal{P}'_{2,1}(n)^c$ 是其补集；对于 $\wp = \{k_1,\dots,k_r\}$ 和 $\wp' = \{2,k_1,\dots,k_i-2,\dots,k_r\}$，或 $\wp' = \{2,k_1,\dots,k_i-1,\dots,k_j-1,\dots k_r\}$，*SM* 箭头 $(\wp,\sigma) \to (\wp',\sigma')$，其中 σ' 是 σ 在诱导映射 $C_\wp \to C_{\wp'}$ 下的像。该凝聚图的源顶点是这样的点对 $(\wp,\sigma)$：其划分为 $\wp \in \mathcal{P}'(n)$，且其中所有 $k_i \neq 2$。

证明。任何工作空间 $F \in \mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ \wp}$ 只要具有 $\wp \neq \{2,1,\dots,1\}$，就是一个暂态状态。事实上，任何包含多个樱桃分量，或至少有一个分量具有 $k_i \geq 3$ 叶子的工作空间 F，都至少有一个极小 SM 箭头，其目标 F' 位于不同的 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ \wp'}$ 中。对于具有某些 $k_i \geq 3$的情况，存在一些 SM 箭头，其中划分 $\wp'$ 用一对 $\{2,k_i-2\}$替换一个 k_i；而对于包含多个樱桃的情况，$\wp'$用一个三元组 $\{2,1,1\}$替换一对 $\{k_i=2,k_j=2\}$，从而增加连通分量的数量，并且不存在能够反方向移动、减少分量数量的 SM（或 IM）箭头。另一方面，在诱导子图 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ \wp=\{2,1,\dots,1\}}$ 中没有 IM 箭头，而其中唯一从 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ \wp=\{2,1,\dots,1\}}$ 的顶点出发的箭头是形如以下形式的 SM 箭头

$$\widehat{\alpha_i\ \alpha_j} \sqcup \alpha_k \bigsqcup_{\ell\neq i,j,k} \alpha_\ell \mapsto \widehat{\alpha_i\ \alpha_k} \sqcup \alpha_j \bigsqcup_{\ell\neq i,j,k} \alpha_\ell\,. \tag{7.1}$$

这些 箭头保持 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ ,\ \wp=\{2,1,\dots,1\}}$ sothis 子图由常返状态组成。

对于 $\wp$ 其中 k_i's 中没有 2 和 1 的情况（这些具有 $k_i \geq 3$），与 $\mathcal{G}^{\mathrm{IM}}_{n,\ A,\ \wp,\ \sigma}$ 具有相同顶点的 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ ,\ \wp}$子图 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A\ \sigma\ ,\ ,\ \wp,}$ 没有将顶点集映射到自身的最小 SM 箭头，因此这些图与 $\mathcal{G}^{\mathrm{IM}}_{n,\ A,\ \wp,\ \sigma}$ 这些强连通且非周期的 IM 图一致。

另一方面，当划分中包含一些 2 和一些 1 时，$\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,\ ,\ \wp}$ 包含极小 SM 边，它们通过置换一些叶标签（如 (7.1) 所示），将子图 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A\ \sigma\ ,\ ,\ \wp,}$ 映射到另一个 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A\ \sigma'\ ,\ ,\ \wp,}$。因此，在这种情况下，一些不同的 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A\ \sigma\ ,\ ,\ \wp,}$ 成为同一强连通分量的一部分，而由此得到的强连通分量仅取决于图 $D(\tilde{\wp})$ 各行上的标签分配。事实上，通过反复应用类似 (7.1) 的箭头，可以置换这些工作空间中樱桃和单叶分量的所有标签。凝聚图 $\bar{\mathcal{G}}^{\mathrm{IM-SM}}_{n\,A}{}_{,}$ 的顶点是列出 $\mathcal{G}^{\mathrm{IM-SM}}_{n\,A}{}_{,}$ 强连通分量的对 $(\wp,\ \sigma)$。其中，源点是没有入射 SM 箭头的顶点，也就是说，其工作空间不含樱桃分量。□

正如我们上文所观察到的，对于至少包含一对 $k_i = 2$、$k_j = 1$ 的划分 $\wp$，每一对这样的元素都对应一个将 $\mathcal{G}_{n,\ A,\ \wp}$ 映射到自身的 SM 变换，它作用于分量 $T_i = \mathfrak{M}(\alpha_a,\ \alpha_b)$ 和 $T_j = \alpha_c$，其形式为

$$\widehat{\alpha_a\ \alpha_b} \sqcup \alpha_c \mapsto \widehat{\alpha_a\ \alpha_c} \sqcup \alpha_b$$

并在工作空间 $F = T_1 \sqcup \cdots \sqcup T_r$ 的其他分量上为恒等映射。这些是唯一能够保持 $\wp$ 不变的 SM 箭头。我们在此将它们记作 $\mathrm{SM}_{2,1}$。对于任意 $F \in V(\mathcal{G}_{n,A,\wp})$

我们有

$$N^{\mathrm{out}}_{\mathrm{SM}_{2,1}}(F) = N^{\mathrm{in}}_{\mathrm{SM}_{2,1}}(F) = \#\{\text{ pairs } \{2,1\} \text{ in } \wp\}\,. \tag{7.2}$$

8. External and Sideward Merge dynamics via Young diagrams

迄今为止，我们的分析表明，为了理解 $\mathcal{G}_{n,\ A}$ 上完整合并动力学的性质——其中包括所有 EM、IM 和最小 SM 箭头——我们需要理解 SM 箭头与 EM 箭头如何相互作用并驱动动力学的行为。与之前一样，我们仅限于最小 SM 箭头，因为这些 SM 箭头已足以使霍普夫代数马尔可夫链具有遍历性，而 [26] 的结果也表明，这些箭头足以解释任何适合采用 SM 作为更优解释的语言现象。

正如我们此前在 §2.2（以及 [26]）中所讨论的，还可以选择只允许那些不会切断樱桃两条边的最小 SM 箭头。这一选择很自然，因为它能带来稍好的代数性质，尤其是对于删除余乘 Δ^d（其他地方将对此进行讨论）。另一方面，我们将看到，纳入这些 SM 箭头会使合并 Hopf 代数马尔可夫链的动力学性质更加理想。因此，鉴于这是我们研究的重点，我们在此始终作出如下假设：在最小 SM 箭头中，也包括那些提取樱桃的两片叶子的箭头。我们可以将其限制在樱桃是更大树的子树这一情形，因为当樱桃是一个分量时，所得的 SM 只是恒等映射，因此可以忽略。

由于我们已经了解 IM 动力学的行为，进一步研究带有 IM、外部合并和最小 SM 箭头的霍普夫代数马尔可夫链 $\mathcal{K}$ 的性质，第一步是研究仅含外部合并和最小 SM 箭头的图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}{}_{()}$。注意，图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}{}_{()}$（不同于上文讨论的 $\tilde{\mathcal{G}}^{\mathrm{EM}}_{\mathcal{P}'(n)}$ 和 $\tilde{\mathcal{G}}^{\mathrm{SM}}_{\mathcal{P}'(n)}$）不再是 $\mathcal{G}_{n,A}$ 的凝聚图，因为$\mathcal{G}_{n,A}$本身是强连通的，因此只有一个强连通分量，没有凝聚图。

我们分步骤分析外部合并–SM 动力学。首先，我们引入图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}{}_{()}$，其顶点集为 $\mathcal{P}'(n)$，箭头描述了在工作空间中，外部合并和最小 SM 对叶子划分的影响。该图不包含树拓扑，也不考虑叶子处标签在 $A = \{\alpha_1, \ldots, \alpha_n\}$ 中的分配。

其次，我们介绍一种追踪叶子处标签 α_i 的方法：将描述 $\mathcal{P}'(n)$ 中划分的杨图替换为字母表 A 上的杨表。最后，我们讨论如何通过适当的重数来考虑不同的树拓扑。

我们将在 §9 讨论，$\mathcal{G}_{n,\ A}$ 上的完整合并动力学如何通过一个中间图 $\mathcal{G}_{\mathcal{P}'(n)}$ 与本节讨论的 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}{}_{()}$ 上的动力学相关联；该中间图也纳入了动力学中的内部合并部分。

8.1. 杨图图。杨图通常构成一个格：可以依据 $\mathcal{P}(n)$ 上的支配序来组织，也可以依据 $\mathcal{P} = \sqcup_n \mathcal{P}(n)$ 上的图的包含关系（亦称覆盖关系）来组织（杨格）。这些序关系及其产生的图可以解释为离散动力系统，并可应用于沙堆模型等模型（这是一种表现出自组织临界性的动力系统）。

这里我们考虑一种作用于杨图上的不同类型的离散动力系统（并非源自格结构），它是通过投影映射，由作用于$\mathcal{G}_{n,A}$ 上的合并作用$\hat{}\mathcal{K}^{(A,n)}$ 在划分集合 $\mathcal{P}'(n)$上诱导出的马尔可夫链$\mathcal{K}_{\mathcal{P}'(n)}$，由此得到一个图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}{}_{()}$。这里我们只考虑其中的外部合并和侧向合并箭头，并在第 9 节中通过添加 IM 自环加以扩展。

定义 8.1。 令 $p\colon \mathfrak{F}_{A,n} \to \mathcal{P}'(n)$ 表示一个投影映射，它将 $\mathfrak{F}_{A,n}$ 中的森林 $F = T_1 \sqcup \cdots \sqcup T_r$ 映射到划分 $\wp = p(F)$，$n = k_1 + \cdots + k_r$。该划分由 n 的划分方式定义

将叶子划分到工作空间的不同分量中，

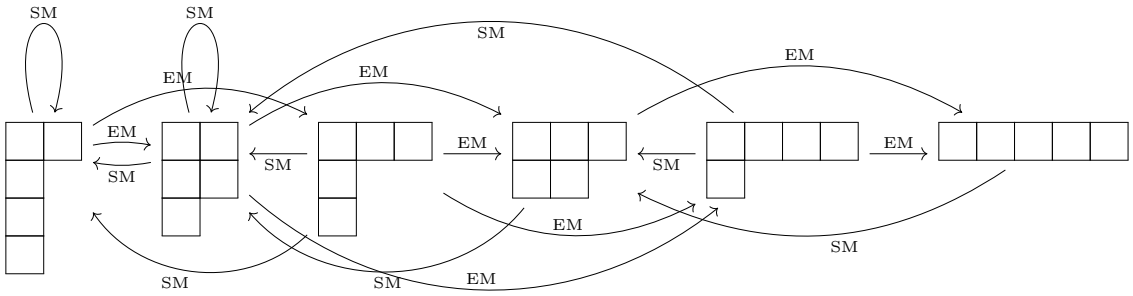
$$L(F) = L(T_1) \sqcup \cdots \sqcup L(T_r)$$

进入每个分量 T_i 的 $k_i = L(T_i)$ 叶子。

这里考虑一个图，其顶点为划分，边则对应于工作空间上 EM 和 SM 箭头对划分产生的作用。

定义 *8.2*。令 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()$ 表示一个图，其顶点集为 $\mathcal{P}'(n)$ 的划分集合，其中排除划分 $\wp_1 = \{1, 1, \ldots, 1\}$；这些划分 $\wp$ 以其杨图 $D(\wp)$ 表示。边由 $\mathcal{G}_{n,A}$ 的 *EM* 和 *SM* 箭头诱导：当图 $\mathcal{G}_{n,A}$ 中存在某个 $F \in p^{-1}(\wp)$ 和某个 $F' \in p^{-1}(\wp')$，使得在投影 $p : \mathfrak{F}_{A,n} \to \mathcal{P}'(n)$ 下，对某个 $F \to F'$ 存在 *EM* 或 *SM* 箭头时，$\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()$ 中就有一条 *EM* 或 *SM* 边 $\wp \to \wp'$。$\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()$ 中的 *EM* 箭头将图中的两行合并成一行，而 *SM* 箭头则从同一行或不同行中取出两个方格，并用它们组成一个长度为 *2* 的新行。

例如，在 $n = 5$ 的情形下，图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}\,5}()$ 呈现如下形式



现在我们来考察图$\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()$ 和马尔可夫链 $\hat{}\mathcal{K}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()$ 的一些一般性质。

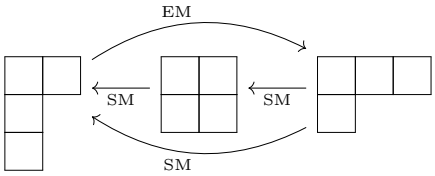
8.2. **强连通性**。 事实上，图$\mathcal{G}_{n,A}$ 的强连通性可由其投影图的强连通性推出，但我们还是给出一个直接证明（这反映了$\mathcal{G}_{n,A}$中给出的证明），因为稍后会用到。

引理 8.3。 对于所有 $n \geq 4$，图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()$ 都是强连通且非周期的。因此，马尔可夫链 $\hat{}\mathcal{K}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()$ 是遍历的：它具有唯一的平稳分布（通常并非均匀分布）。在$\hat{}\mathcal{K}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()$ 的迭代作用下，任意初始分布都会收敛到平稳分布。

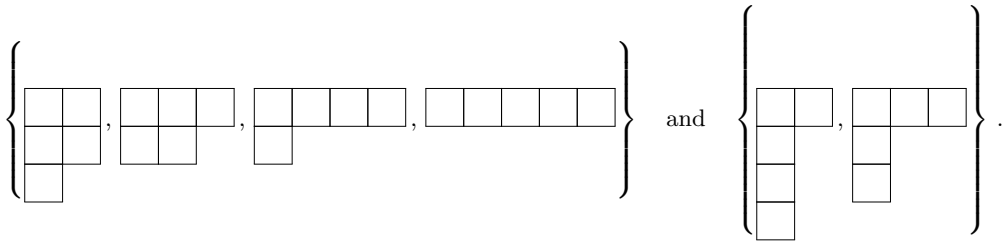
证明。 要证明强连通性，只需证明：对于每个 $\wp, \wp' \in V(\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}()) = \mathcal{P}'(n)$，都存在一条由极小 SM 箭头构成的有向路径，从 $\wp$ 通向划分 $\{2, 1, \ldots, 1\}$；并且存在一条由 EM 箭头构成的有向路径，从 $\{2, 1, \ldots, 1\}$ 通向任意其他 $\wp'$。对于第一个结论，假设 $\wp = \{k_1, \ldots, k_r\}$。对于任意 $k_i \geq 3$，都存在一串 SM 箭头，从长度为 k_i 的一行中取出两个方格，并得到 $\{2, k_1, \ldots, k_i - 2, \ldots, k_r\}$；如此反复，直到每个这样的 k_i 都缩减为 2 或 1。应用这一串 SM 变换后，得到的划分只包含 2 和 1。通过反复从两条长度为 2 的行中各取出一个方格，便可用极小 SM 箭头将任何这样的划分化为仅有一个 2、其余全为 1 的划分；这会新生成一条长度为 2 的行和两条长度为 1 的行（其他各行不变），使长度为 2 的行数减少一行、长度为 1 的行数增加两行，直到得到 $\{2, 1, \ldots, 1\}$。要得到从 $\{2, 1, \ldots, 1\}$ 通向另一个顶点 $\wp'$的 EM 路径，注意到 $\wp'$至少包含一条长度为 $k_i \geq 3$ 的行，和/或其他长度为 2 的行。这些行中的每一行

长度为 $k_i \geq 2$ 的行都可以通过连续的 EM 箭头得到：从合并两行$\{2,1,\ldots,1\}$开始，利用剩余的长度为 1 的行$\{2,1,\ldots,1\}$，每次向行中添加一个方格。由于 $\{2,1,\ldots,1\}$ 处及其他仅由 2 和 1 构成的划分处存在 SM 自环箭头，非周期性显然成立。因此，由强连通性和非周期性可得遍历性。□

请注意，即使我们将范围限制在不含 SM 自环的子图上，对于所有 $n \geq 4$，我们仍然具有非周期性，因为任何包含下方所示图的三个 $\wp$，$\wp'$，$\wp''$，都存在一个来自该 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}\ 4}{}_{()}$ 子图的长度为 2 的循环和一个长度为 3 的循环：



注8.4。 尽管对于所有图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}{}_{()}$，强连通性都成立，但直接检查上文 $n = 5$ 的示例可知，这并非图的一种“稳健”性质：只需移除一条边（连接 $\{2,2,1\}$ 与 $\{2,1,1,1\}$划分的 SM 箭头），就会失去强连通性。事实上，移除该箭头后，图会分裂为两个强连通分量（第一个是常返的，第二个是暂态的）：



这种“鲁棒性”可通过强 k边连通性来衡量。若移除至多 $k-1$ 条边后，图k仍然连通，则称其为k-边连通；若移除至多 $k-1$ 条边后，图$\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}\ 5}$ 仍然强连通，则称其为强k-边连通。图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}\ 5}{}_{()}$ 是强 1-边连通的（也就是说，它是强连通的），但不是强 2-边连通的。

事实上，随着 n 增大，强边连通性永远不会提高。

命题 8.5. 对于所有 $n \geq 3$，图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}{}_{()}$ 都是强1-边连通的，但从来都不是强2-边连通的。

证明。引理 8.3 证明了强 1-边连通性（强连通性）。$n = 3$ 的情形仅包含两个顶点，且两者之间各有一条箭头，因此显然，移除一条边就会导致强连通性不再成立。对于任意 $n \geq 4$，考虑两个划分 $\wp = \{2,1,\ldots,1\}$ 和 $\wp' = \{3,1,\ldots,1\}$。划分 $\wp'$ 不可能是任何最小 SM 箭头的目标，因为它不包含任何 $k_i = 2$，所以只能通过一条外部合并箭头到达。这样的箭头只有一条，它来自标记为 $\wp$ 的顶点。因此，移除这条外部合并箭头所对应的单条边 $\wp \longrightarrow \wp'$，会导致 $\wp'$ 成为一个暂态状态，实际上是一个源顶点，因此剩余的图不再强连通。□

这表明，$\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}{}_{()}$ 网络强连通性的“脆弱性”源于某些顶点无法通过最小 SM 箭头直接到达。这些顶点只能通过 EM 箭头到达。因此，某些 EM 箭头对图的强连通性尤为关键。请注意，命题 8.5 的论证指出，在结构形成的早期阶段，一类关键的

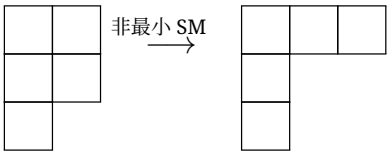
EM 箭头是从一个樱桃和一片单叶指向一个三叶结构的箭头

$$\mathrm{EM}:\ \alpha \sqcup \overset{\frown}{\beta\ \ \gamma} \longmapsto \overset{\frown}{\alpha\ \ \overset{\frown}{\beta\ \ \gamma}}\,.$$

从语言学的角度来看，可以认为这些 EM 箭头确实发挥着特殊作用，因为正是它们使得从两个基本结构的简单二元关联 $\{\alpha,\beta\}$过渡到诸如 $\{\{\alpha,\beta\},\gamma\}$ 这样的实际层级结构成为可能。

这一性质还体现了 EM 与 SM 之间的根本不对称性，其中 EM（尤其是这些特定的 EM 箭头）在决定动力学的性质方面发挥着更为关键的作用。

不过请注意，如果包括所有 SM 箭头，而不仅仅是最小的箭头，那么就会存在一些目标为不包含任何 $k_i = 2$ 的划分的 SM 箭头，例如



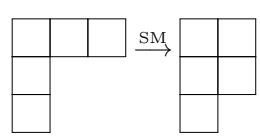
对应于如下这样的非最小 SM 变换

$$\text{non-min SM}:\ \overset{\frown}{\alpha_1\ \ \alpha_2} \sqcup \overset{\frown}{\alpha_3\ \ \alpha_4} \sqcup \alpha_5 \longmapsto \overset{\frown}{\alpha_2\ \ \overset{\frown}{\alpha_3\ \ \alpha_4}} \sqcup \alpha_1 \sqcup \alpha_5\,.$$

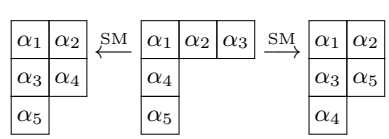
因此，保持 SM 的极小性（如 [26], 所讨论的，另见下文 §11）会导致强连通性的“脆弱性”，以及某些 EM 箭头的特殊作用。

8.3. 杨氏表图。 作为一个中间步骤，在回过头来讨论工作空间上的完全合并作用和图 $\mathcal{G}^{\mathrm{EM-SM}}_{n\ A}$，之前，我们通过纳入叶子处的标签分配 α_i（按划分 $\wp \in \mathcal{P}'(n)$ 分类），对图$\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}$ $()$ 进行细化。因此，对于 $A = \{\alpha_1,\ldots,\alpha_n\}$，我们考虑图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n\ A}$ $()$，其中顶点是形如 $(\wp,\ \alpha)$ 的配对，且$\alpha:\ A \to Y(\wp)$ 是一个映射，它将元素 $\alpha_i \in A$ 双射地分配给划分 $\wp \in \mathcal{P}'(n)$ 的杨图 $Y(\wp)$ 中的方格；这些分配视为在杨群 $S_\wp$（行内置换）作用下等价，并且相同行长的行可以互换。因此，此类分配的总数 $\alpha:\ A \to Y(\wp)$ 由式 (2.8) 的广义多项式系数 $\Upsilon_{\wp,\ n}$ 给出。

同样，我们不将划分 $\{1,\ldots,1\}$ 包含在 $V(\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n\ A}$ $()$ $)$ 中。于是，这些映射与 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}$ $()$ 中划分上的映射、外部合并和 SM 箭头相同，但现在还会追踪标记，因此，例如，一个像这样的箭头

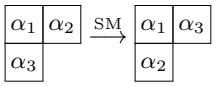


会分裂成两个箭头



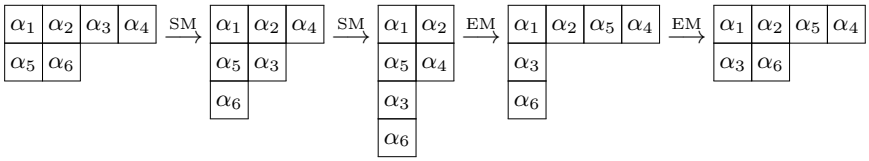
对于 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}$ $()$ 的其他箭头也同样如此：它们在遗忘标签 $\alpha_i \in A$ 的投影映射 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}$ $()$ $_A$ $\to \mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\ n}$ $()$ 下具有不同的可能原像。

注释 8.6. 请注意，在图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}$ () 中，位于包含 2 和 1 的划分 $\wp$ 所对应顶点上的自环 SM 映射，在 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ () 中不再是自环，因为它们现在将一个顶点（$\wp$，α）映射到另一个不同的顶点（$\wp$，α'），其中 $\alpha \neq \alpha'$，例如



引理 8.7。 图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ ()是强连通且非周期的，因此马尔可夫链 $\hat{}\mathcal{K}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ ()是遍历的。

证明。由 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}$ () 的强连通性可知，只需证明固定划分 $\wp$ 上的任意标签排列都能通过 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ ()中的箭头串接实现，即可说明 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ ()也是强连通的。下面以一个例子说明这一通用方法。假设我们要在下面的杨表中交换标签 α_3 和 α_5。这可以通过以下一串侧向合并和外部合并箭头实现：



同样的步骤可以用来交换图中不同行的任意两个标签。由于标签在图中的分配只需确定到行内置换，因此组合这些操作就足以实现所有可能的置换。用于 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}$ ()的同一论证表明 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ ()具有非周期性。

□

因此，我们得到以下结果。

推论 8.8。 由作用在 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ () 上的合并霍普夫代数马尔可夫链 $\hat{}\mathcal{K}^{(A,n)}$ 在 $\mathcal{G}_{n,A}$ 上诱导的马尔可夫链 $\hat{}\mathcal{K}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ ()具有唯一的渐近分布

$$\pi_{\mathcal{P}'(n),A}\;\hat{\mathcal{K}}^{\mathrm{EM-SM}}_{\mathcal{P}'(n),A} = \pi_{\mathcal{P}'(n),A}\,,$$

其中，对于所有 $(\wp,\alpha) \in V(\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}())$，以及任意初始分布，$\pi_{\mathcal{P}'(n)\,A}(\wp,\alpha) > 0$ 均收敛于 $\pi_{\mathcal{P}'(n)\,A}$，这是在 $\hat{}\mathcal{K}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ () 的迭代作用下成立的。

请注意，与内部合并动力学不同，外部合并和侧向合并动力学在 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n\;A}$ () 上的平稳分布 $\pi_{\mathcal{P}'(n)\,A}$ 通常并不均匀。

9. Full Merge dynamics with minimal Sideward Merge

现在，我们将分别研究图 $\mathcal{G}_{A,\,n,\,\wp,\,\sigma}$ 上的内部合并动力学，以及图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}$ () 上由侧向合并和外部合并所诱导的合并动力学所得的全部结果结合起来，以理解 $\mathcal{G}_{n,\,A}$ 上的完整合并动力学。为此，我们将在此考虑对图 $\mathcal{G}^{\mathrm{EM-SM}}_{\mathcal{P}'\,n}$ () 的扩展，将其扩展为图 $\mathcal{G}_{\mathcal{P}'\,(n)}$，该图还包括作为 $\mathcal{G}_{n,\,A}$ 的 IM 箭头的像的自环。为了研究该图 $\mathcal{G}_{\mathcal{P}'\,(n)}$ 上的动力学与完整合并图 $\mathcal{G}_{n,\,A}$ 上的动力学之间的关系，我们首先在第 9.1 节证明关于马尔可夫链投影的一般性结果。

9.1. 马尔可夫链投影。 我们将在此说明，当一个有限马尔可夫链相对于将状态空间投影到较小状态空间的投影具有某种特定对称性时，可以在较小的图上计算平稳分布。

定义 9.1。 设 X 是一个有限集，且 $p\colon X \to Y$ 是一个满射。设 G_X 是一个具有 $V(G_X) = X$ 的强连通非周期有向图，并设 $\mathcal{K}^X$ 是 G_X 的邻接矩阵。若图 G_X 具有如下性质，则称其为 p-对称的：从顶点 x 出发、指向纤维 $p^{-1}(y')$ 中顶点的边数，在 $x \in p^{-1}(y)$ 上为常数

$$\sum_{x' \in p^{-1}(y')} \mathcal{K}^X_{x,x'} = \mathcal{K}^{Y,R}_{y,y'}, \quad \forall x \in p^{-1}(y) \tag{9.1}$$

并且从纤维 $p^{-1}(y)$ 中顶点出发、指向顶点 x' 的边数，在 $x' \in p^{-1}(y')$ 上为常数，

$$\sum_{x \in p^{-1}(y)} \mathcal{K}^X_{x,x'} = \mathcal{K}^{Y,L}_{y,y'}, \quad \forall x' \in p^{-1}(y'), \tag{9.2}$$

with

$$\#p^{-1}(y) \cdot \mathcal{K}^{Y,R}_{y,y'} = \#p^{-1}(y') \cdot \mathcal{K}^{Y,L}_{y,y'}. \tag{9.3}$$

其中计算的是 G_X 中从纤维 $p^{-1}(y)$ 到纤维 $p^{-1}(y')$ 的有向边总数。

(9.3) 所计算的数可以作如下几何解释。设 G_Y 是一个顶点集为 Y 的图；当存在某些 $F \in p^{-1}(y)$ 和 $F' \in p^{-1}(y')$，使得它们之间在 G_X 中存在有向边 $F \to F'$ 时，就在其中加入一条有向边 $y \to y'$。这样，映射 $p\colon X \to Y$ 可扩展为映射 $p\colon G_X \to G_Y$。我们将 G_Y 称为由 G_X 和 p 诱导的图。于是，(9.3) 两边所计算的都是 G_X 中位于 G_Y 的一条边 $y \to y'$ 之上的边数，其中相对于投影 p 而言。

命题 9.2。 设 X 为一个有限集，且 $p\colon X \to Y$ 为满射，G_X 为定义在顶点集 X 上的一个强连通、非周期的 p-对称图。那么诱导图 G_Y 是强连通的。假设它也是非周期的。令 $c_y = \#p^{-1}(y)$ 表示点 $y \in Y$ 上方纤维的大小。令 λ 和 η_X 分别为图 G_X 的邻接矩阵 $\mathcal{K}^X$ 的佩龙-弗罗贝尼乌斯特征值和（右）特征向量，

$$\sum_{x' \in X} \mathcal{K}^X_{x,x'}\, \eta_X(x') = \lambda\, \eta_X(x), \tag{9.4}$$

并令 $\hat{}\mathcal{K}^X$ 为相应的马尔可夫链转移矩阵

$$\hat{\mathcal{K}}^X_{x,x'} := \lambda^{-1} \frac{\eta_X(x')}{\eta_X(x)} \mathcal{K}^X_{x,x'}.$$

(1) $\mathcal{K}^X$ 的右佩龙-弗罗贝尼乌斯特征函数 $\eta_X(x) = \eta_Y(y)$ 在纤维 $p^{-1}(y)$ 上为常数（对 $y \in Y$ 而言），其值 $\eta_Y(y)$ 由 $\mathcal{K}^{Y,R}$ 的右佩龙-弗罗贝尼乌斯特征函数给出

$$\sum_{y' \in Y} \mathcal{K}^{Y,R}_{y,y'}\, \eta_Y(y') = \lambda\, \eta_Y(y). \tag{9.5}$$

(2) $\xi_X(x) = \xi_Y(y)$ $\mathcal{K}^X$ 的左佩龙-弗罗贝尼乌斯特征函数在纤维$p^{-1}(y)$上也为常数，并由 $\mathcal{K}^{Y,L}$的左佩龙-弗罗贝尼乌斯特征函数给出

$$\sum_{y \in Y} \xi_Y(y)\, \mathcal{K}^{Y,L}_{y,y'} = \lambda\, \xi_Y(y). \tag{9.6}$$

(3) $\pi_X(x)$的平稳分布$\hat{\mathcal{K}}^X$,

$$\sum_{x\in X}\pi_X(x)\,\hat{\mathcal{K}}^X_{x,x'}=\pi_X(x') \tag{9.7}$$

也为常数，$\pi_X(x)=\pi_Y(y)$，在纤维 $x\in p^{-1}(y)$ 上，其中 $\pi_Y(y)$ 是平稳分布n

$$\sum_{y\in y}\pi_Y(y)\;\hat{\mathcal{K}}^Y_{y,y'}=\pi_Y(y')\,, \tag{9.8}$$

在具有转移矩阵的 Y 上的马尔可夫链

$$\hat{\mathcal{K}}^Y_{y,y'}=\lambda^{-1}\,\frac{\eta_Y(y')\cdot c_y}{\eta_Y(y)\cdot c_{y'}}\,\mathcal{K}^{Y,R}_{y,y'}\,.$$

（4）等价地，函数

$$\psi_Y(y):=\frac{\pi(y)\,c_y}{\eta_Y(y)} \tag{9.9}$$

是 $\mathcal{K}^{Y,\,R}$ 的（左）*Perron-Frobenius* 特征向量，满足

$$\sum_y\psi_Y(y)\,\mathcal{K}^{Y,R}_{y,y'}=\lambda\,\psi_Y(y')\,, \tag{9.10}$$

(5) 马尔可夫链$\hat{\mathcal{K}}^X$ 的平稳分布为

$$\pi_X(x)=\frac{\eta_Y(p(x))\cdot\psi_Y(p(x))}{c_{p(x)}}\,, \tag{9.11}$$

可用 $\mathcal{K}^{Y,R}$ 的左、右 *PF* 特征向量 ψ_Y 和 η_Y，以及投影 $p:\,X\to Y$的纤维大小 c_y 来表示。

(6) $\mathcal{K}^{Y,L}$ 的左佩龙-弗罗贝尼乌斯特征向量 ξ_Y 与 $\mathcal{K}^{Y,R}$ 的 ψ_Y（经归一化后）满足如下关系：

$$\xi_Y(y)=\frac{\psi_Y(y)}{c_y}\,. \tag{9.12}$$

证明。图 G_Y是强连通的，因为 G_X是强连通的。事实上，给定 $y,y'\in Y$，任意两个顶点 $F\in p^{-1}(y)$ 和 $F'\in p^{-1}(y')$ 在 G_X 中双向路径相连，而这些路径中的后续边在 G_X 中的投影是 G_Y中的后续边。

(1) 根据假设，图 G_Y 也是非周期的，因此矩阵 $\mathcal{K}^{Y,R}$ 有唯一的（相差一个缩放因子）右佩龙-弗罗贝尼乌斯特征函数，其值为 $\eta_Y(y)>0$，对所有 $y\in Y$均成立，并满足 (9.6)，其中$\lambda>0$ 是 Y的 PF 特征值。于是我们有

$$\sum_{x'\in X}\mathcal{K}^X_{x,x'}\,\eta_Y(y')=\sum_{y'}\sum_{x'\in p^{-1}(y')}\mathcal{K}^X_{x,x'}\,\eta_Y(y')=\sum_{y'}\mathcal{K}^{Y,R}_{y,y'}\,\eta_Y(y')=\lambda\eta_Y(y)\,,$$

因此，$\eta_X(x):=\eta_Y(p(x))$ 是具有相同 λ 的 (9.4) 的一个解。$\mathcal{K}^X$ 的佩龙-弗罗贝尼乌斯定理保证，左/右 PF 特征向量是 $\mathcal{K}^X$ 中所有分量均为正实数的唯一特征向量。因此，η_X 必须是 $\mathcal{K}^X$ 的右佩龙-弗罗贝尼乌斯特征向量，且除比例因子外唯一。

(2) T $\mathcal{K}^X$ 的左佩龙-弗罗贝尼乌斯特征函数情形与

$$\sum_{x\in X}\xi_X(x)\mathcal{K}^X_{x,x'}=\sum_y\sum_{x\in p^{-1}(y)}\xi_Y(y)\mathcal{K}^X_{x,x'}=\sum_y\xi_Y(y)\mathcal{K}^{Y,L}_{y,y'}=\lambda\xi_Y(y)\,.$$

(3) 类似地，鉴于 η_X 在 p 的原像上为常数，若 $\pi_Y(y)$ 是满足（9.8）的 $\hat{\mathcal{K}}^Y$ 的平稳分布，则 $\pi_X(x) := \pi_Y(p(x))$ 对于 $x' \in p^{-1}(y')$ 满足，

$$\sum_x \pi_X(x)\,\hat{\mathcal{K}}_{x,x'} = \sum_y \pi_Y(y)\,\lambda^{-1}\frac{\eta_Y(y')}{\eta_Y(y)} \sum_{x\in p^{-1}(y)} \mathcal{K}^X_{x,x'} = \sum_y \pi_Y(y)\,\lambda^{-1}\frac{\eta_Y(y')}{\eta_Y(y)}\mathcal{K}^{Y,L}_{y,y'}$$

$$= \sum_y \pi_Y(y)\,\lambda^{-1}\frac{\eta_Y(y')\cdot c_y}{\eta_Y(y)\cdot c_{y'}}\mathcal{K}^{Y,R}_{y,y'} = \sum_y \pi_Y(y)\,\hat{\mathcal{K}}^Y_{y,y'} = \pi_Y(y') = \pi_X(x')\,.$$

由于图 G_X 是强连通且非周期的，平稳分布 π_X 是唯一的，因此它必定具有这种形式。

(4) 我们也可以等价地将方程写成

$$\sum_{y\in y} \pi_Y(y)\ \lambda^{-1}\,\frac{\eta_Y(y')\cdot c_y}{\eta_Y(y)\cdot c_{y'}}\,\mathcal{K}^{Y,R}_{y,y'} = \pi_Y(y')$$

如下形式

$$\sum_{y\in y} \frac{\pi_Y(y)c_y}{\eta_Y(y)}\,\mathcal{K}^{Y,R}_{y,y'} = \lambda\,\frac{\pi_Y(y')c_{y'}}{\eta_Y(y')}\,,$$

这说明 ψ（y）如 (9.9) 所示，是 $\mathcal{K}^{Y,\,R}$ 的左 Perron-Frobenius 特征向量方程，其 PF 特征值为 λ。

(5) 这直接由 (4) 以及如 (9.11) 所示的平稳分布 π（x）的形式得出。(6) 由于 ψ_Y 是 $\mathcal{K}^{Y,\,R}$ 的左 PF 特征向量，我们有

$$\sum_y \frac{\psi_Y(y)}{c_y}\ \frac{c_y}{c_{y'}}\mathcal{K}^{Y,R}_{y,y'} = \lambda\frac{\psi_Y(y')}{c_{y'}}\,,$$

由（9.3）以及左 PF 特征向量（模归一化）的唯一性可得关系（9.12）。

□

9.2. **工作空间上合并动力学的平稳分布**。这一结果适用于我们在此讨论的合并霍普夫代数马尔可夫链情形。

命题 *9.3.* 设 $X = V(\mathcal{G}_{n,A}) = \mathfrak{F}_{A,n} \subset \mathfrak{F}_{SO_0}$ 为叶子数为 n、标签集为 $A = \{\alpha_1, \dots, \alpha_n\}$且边集非空的非平面满二叉森林集合。设 $Y = \mathcal{P}'(n)$ 为 n 的划分集合（不包括划分 $\{1,\dots,1\}$）。设 $p: \mathfrak{F}_{A,n} \to \mathcal{P}'(n)$ 为映射，它将具有 n 个叶子的森林 F 映射到 $n = k_1 + \cdots + k_r$ 的划分 $\wp = p(F)$，其中各部分表示森林中每个树分量 $F = T_1 \sqcup \cdots \sqcup T_r$ 的叶子数，$k_i = \#L(T_i)$。图 $\mathcal{G}_{n,A}$ 相对于此映射是 p-对称的。投影 $p: \mathfrak{F}_{A,n} \to \mathcal{P}'(n)$ 诱导出图的投影 $\mathcal{G}_{n,A} \to \mathcal{G}_{\mathcal{P}'(n)}$，其中图 $\mathcal{G}_{\mathcal{P}'(n)}$ 的顶点集为 $\mathcal{P}'(n)$，并具有定义 *8.2* 中的 *EM*、*SM* 箭头，以及在某个顶点 $\wp$ 处的一个额外 *IM* 自环；当某些顶点 $F, F' \in p^{-1}(\wp) = \mathcal{G}_{n,A,\wp}$ 之间存在 *IM* 箭头 $F \to F'$ 时，就有这个自环。

证明。图 $\mathcal{G}_{n,\,A}$ 之所以关于映射 $p: \mathfrak{F}_{A,\,n} \to \mathcal{P}'(n)$ 具有 p-对称性，关键在于，从（或指向）$\mathcal{G}_{n,\,A}$中的顶点 F 出发的边数仅取决于 $p(F)$（假设我们还包括侧向合并箭头：当樱桃属于更大的树时，这些箭头会提取该樱桃的两片叶子）。这由命题 2.3 得出。事实上，这一性质还可以进一步细化：从（或指向）$\mathcal{G}_{n,\,A}$ 上一个顶点 F 的边的计数，可以按照目标（分别为源点）顶点位于各个纤维 $p^{-1}(\wp')$ 中的边的子集分别计数，而对每个此类集合的计数也只取决于 $\wp = p(F)$，而不取决于 F 本身。为此，注意对于每一种箭头类型（EM/IM/SM），

该类型出射（或入射）边的计数取决于在划分 $\wp = p(F)$ 中选择适当方格的方式。任意这样的方格选择，连同操作类型（EM/IM/SM），足以确定与该箭头目标顶点相关的结果划分 $\wp'$ 的形状。因此，源点为 F 且另一端位于纤维 $p^{-1}(\wp')$ 中的每个箭头子集，都由已知划分 $\wp$ 所确定。这意味着（9.2）和（9.1）均成立。按照构造，映射 $\mathcal{G}_{n,\ A} \to \mathcal{G}_{\mathcal{P}'(n)}$ 的纤维为 $p^{-1}(\wp) = \mathcal{G}_{n,\ A,\ \wp}$，因此 $\mathcal{G}_{n,\ A}$ 的 EM 和 SM 箭头产生了 $\mathcal{G}_{\mathcal{P}'(n)}$ 中的边，如定义 8.2 所述；而 IM 箭头始终沿着纤维的连通分支，因此在 $\mathcal{G}_{\mathcal{P}'(n)}$ 中产生自环边。□

因此，我们得到了工作空间集合上完整合并动力学的平稳分布形式，该分布由图$\mathcal{G}_{n,A}$上的霍普夫代数马尔可夫链 $\hat{}_{\mathcal{K}^{A,n}}$ 给出。

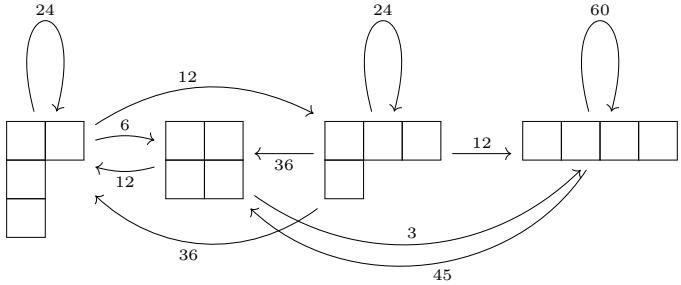
定理 *9.1*。图$\mathcal{G}_{n,A}$ 上的合并霍普夫代数马尔可夫链 $\hat{}_{\mathcal{K}^{A,n}}$ 具有平稳分布，该分布在子集 $V_{n,A,\ \wp} = p^{-1}(\wp)$ 上为常数，这些子集是投影$p : \mathcal{G}_{n,A} \to \mathcal{G}_{\mathcal{P}'(n)}$的纤维。除一个整体乘法常数外，合并霍普夫代数马尔可夫链 $\hat{}_{\mathcal{K}^{A,n}}$ 的平稳分布 $\pi(F)$由图 $\mathcal{G}_{\mathcal{P}'(n)}$的邻接矩阵 $\mathcal{K}^{\mathcal{P}'(n)R}$ 的左、右 *Perron-Frobenius* 特征向量 $\eta_{\mathcal{P}'(n)}$ $\psi_{\mathcal{P}'(n)}$以及纤维的重数 $\#V_{n,A,\wp} = \Lambda_{\wp,n}$ （如式 *(2.9)* 所示）得到，其中包括 *IM* 自环边，形式如下：

$$\pi(F) = \frac{\eta_{\mathcal{P}'(n)}(\wp)\ \psi_{\mathcal{P}'(n)}(\wp)}{\Lambda_{\wp,n}} \qquad for\ \wp = p(F)\,. \tag{9.13}$$

证明。分布的形式 (9.13) 由命题 9.2 得出；分布在 $V_{n,A,\wp}$ 上均匀这一事实则由 (5.1) 和 (7.2) 得出，与命题 5.5 中的情况相同。□

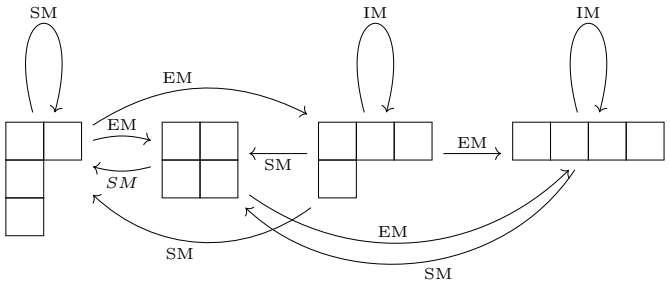
9.3. 的例子 $n = 4$。 这里我们展示一个具体例子。这个例子足够简单，因此所有内容都可以明确计算，同时又足以说明上述步骤的全部过程。 $n = 3$ 片叶子的情况过于简单（且已在 [26] 中完整讨论），因此我们考察 $n = 4$ 片叶子的情况。我们首先描述投影 p 的几何结构：$\mathcal{G}_{4,\ A} \to \mathcal{G}_{\mathcal{P}(4)}$。

引理 9.4。 在 $n = 4$ 的情况下，投影 p： $\mathcal{G}_{4,\ A} \to \mathcal{G}_{\mathcal{P}(4)}$ 在每条边上方的重数为



而顶点上方的重数分别为：第一个顶点6 ， 第二个顶点3 ， 第三个顶点12 ， 第四个顶点15 。

证明。 带有外部合并和侧向合并箭头以及自环 IM 箭头的图 $\mathcal{G}_{\mathcal{P}(4)}$ 具有如下形式



要计算箭头的权重，我们需要通过投影 $\mathcal{G}_{n,\ A} \to \mathcal{G}_{\mathcal{P}'(n)\ A} \to \mathcal{G}_{\mathcal{P}'(n)}$，计算图中 $\mathcal{G}_{n,\ A}$ 每条边上方的边数。由于 F 在 $\mathcal{G}_{n,\ A}$ 中每个顶点处的入边和出边数量仅取决于划分$\wp = p(F)$（因为这里也允许 SM 箭头切分较大树内部樱桃的两片叶子），边 $\wp \to \wp'$ 上方的边数可以通过以下乘积计算： $\wp$ 在 $\mathcal{G}_{n,\ A}$ 中上方的顶点数，乘以从 $\wp$ 上方的某个顶点到 $\wp'$上方纤维的边数。因此，我们得到：

- $\mathcal{G}_{4,\ A}$中位于 $\mathcal{G}_{\mathcal{P}(4)}$的顶点 $\wp = \{2,1,1\}$ 上方共有 6 个顶点，也就是叶子标签的各种选择

  $$\overset{\frown}{\alpha_i \quad \alpha_j} \sqcup \alpha_k \sqcup \alpha_\ell$$

  也就是从 $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ 中任选两个来作为樱桃叶的所有选择。
- 在$\mathcal{G}_{4,\ A}$ 顶点 $\wp = \{2,2\}$ 的$\mathcal{G}_{\mathcal{P}\ (4)}$ 上方有 3 个顶点，再次由标签的所有选择给出

  $$\overset{\frown}{\alpha_i \quad \alpha_j} \sqcup \overset{\frown}{\alpha_k \quad \alpha_\ell}$$

  这可以看作选择两个标签形成第一个樱桃的 6 种方式；考虑两个樱桃之间的对称性后，计数降为 3 种。
- 在$\mathcal{G}_{4,\ A}$ 顶点 $\wp = \{3,1\}$ 的$\mathcal{G}_{\mathcal{P}\ (4)}$ 上方有 12 个顶点，由标签分配给

  $$\overset{\frown}{\alpha_i \quad \alpha_j \quad \alpha_k} \sqcup \alpha_\ell$$

  因此，为孤立顶点选择标签有 4 种，而在这棵树中，为连接到根的叶子选择标签还有 3 种。

- 在$\mathcal{G}_{4,\ A}$ 顶点 $\wp = \{4\}$ 的$\mathcal{G}_{\mathcal{P}\ (4)}$ 上方有 15 个顶点，这一数量是对两种树拓扑中叶子的不同可能标记求和得到的，

  $$\overset{\frown}{\alpha_i \quad \alpha_j \quad \alpha_k \quad \alpha_\ell} \quad \text{and} \quad \overset{\frown}{\alpha_i \quad \alpha_j \quad \alpha_k \quad \alpha_\ell}\ .$$

  第一种情况下，叶子有 3 种可能的分配方式（如 $\wp = \{2,2\}$的情况）；第二种情况下，有 12 种（为连接到根的顶点选择标签有 4 种，而在第二个顶点处剩下 3 种选择）。

接下来，我们只需统计从一个高于 $\wp$ 且目标高于 $\wp'$的顶点出发的边数，因为所有高于 $\wp$的顶点的这个数都相同。我们通过对 $\mathcal{G}_{n,\ A}$ 的入度和出度进行一般性计数来统计这些边。

- 有 4 个侧向合并箭头从 F上方 $\wp = \{2,1,1\}$ 指向同一个纤维，因此上方有 $24 = 6 \cdot 4$ 条边 $\{2,1,1\} \to \{2,1,1\}$。

- 有 1 个外部合并箭头从 F 上方 $\wp=\{2,1,1\}$ 指向上方的顶点 $\wp'=\{2,2\}$，因此上方有 6 条边 $\{2,1,1\}\to\{2,2\}$。
- 从 F 上方到 $\wp'=\{3,1\}$上方的顶点有 2 个外部合并箭头，因此 $\{2,1,1\}\to\{3,1\}$上方有 12 条边。
- 从 F 上方到 $\wp'=\{2,1,1\}$上方的顶点有 4 个侧向合并箭头，因此 $\{2,2\}\to\{2,1,1\}$上方有 12 条边。
- 从 F 上方到 $\wp'=\{4\}$上方的顶点有 1 个外部合并箭头，因此 $\{2,2\}\to\{4\}$上方有 3 条边。
- 从 F 上方的一个顶点到同一纤维有 2 个内部合并箭头，因此 $\{3,1\}\to\{3,1\}$上方有 $24=12\cdot 2$ 条边。
- 从 F 上方到 $\wp'=\{2,2\}$上方的顶点有 3 个侧向合并箭头，因此有 $36=12\cdot 3$ 条上方的边 $\{3,1\}\to\{2,2\}$。
- 从 F 上方到 $\wp'=\{2,1,1\}$上方的顶点有 3 个侧向合并箭头，因此有 $36=12\cdot 3$ 条上方的边 $\{3,1\}\to\{2,1,1\}$。
- 从 F 上方到 $\wp'=\{4\}$上方的顶点有 1 个外部合并箭头，因此上方有 12 条边 $\{3,1\}\to\{4\}$。
- 从 F 上方到 $\wp'=\{2,2\}$上方的顶点有 3 个侧向合并箭头，因此有 $45=15\cdot 3$ 条上方的边 $\{4\}\to\{2,2\}$。
- 从顶点 F 上方 $\wp=\{4\}$ 指向同一纤维有 4 个内部合并箭头，因此有$60=15\cdot 4$ 条上方的边 $\{4\}\to\{4\}$。

□

接下来我们计算命题 9.2 的矩阵 $\mathcal{K}^{\mathcal{P}(4),R}$。

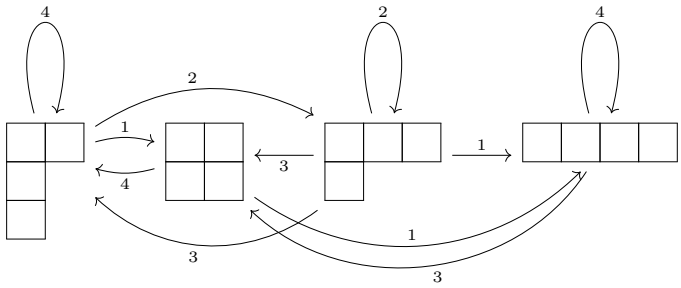
引理 9.5。对于 $n=4$，满足条件的矩阵 $\mathcal{K}^{\mathcal{P}(4),R}$

$$\mathcal{K}^{\mathcal{P}(4),R}_{\wp,\wp'}=\sum_{F'\in p^{-1}(\wp')}\mathcal{K}^{(A,4)}_{F,F'},$$

其中 $\mathcal{K}^{(A,4)}_{F\,F'}$，$\mathcal{G}_{4,A}$的邻接矩阵为

$$\mathcal{K}^{\mathcal{P}(4),R}=\begin{pmatrix}4&1&2&0\\4&0&0&1\\3&3&2&1\\0&3&0&4\end{pmatrix}\tag{9.14}$$

并对应于加权图（顶点从左到右列出）



$\mathcal{K}^{\hat{A}}$在$\mathcal{G}_{4,A}$ 上的，4 平稳分布具有如下形式

$$\pi_{4,A}(F)\sim\begin{cases}0.5267 & F\in p^{-1}(\{2,1,1,1\})\\0.3109 & F\in p^{-1}(\{2,2\})\\0.1217 & F\in p^{-1}(\{3,1\})\\0.0406 & F\in p^{-1}(\{4\})\,.\end{cases}\tag{9.15}$$

证明。 (9.14) 的矩阵 $\mathcal{K}^{\mathcal{P}(4),\,R}$ 的佩龙–弗罗贝尼乌斯特征值为 $\lambda \sim 6.9656$，右佩龙–弗罗贝尼乌斯特征向量为 $\eta_{\mathcal{P}(4)}(\{2,1,1,1\}) \sim 0.5565$、$\eta_{\mathcal{P}(4)}(\{2,2\}) \sim 0.3739$、$\eta_{\mathcal{P}(4)}(\{3,1\}) \sim 0.6383$、$\eta_{\mathcal{P}(4)}(\{4\}) \sim 0.3782$，左佩龙–弗罗贝尼乌斯特征向量为 $\psi_{\mathcal{P}(4)}(\{2,1,1,1\}) \sim 0.8345$、$\psi_{\mathcal{P}(4)}(\{2,2\}) \sim 0.3666$、$\psi_{\mathcal{P}(4)}(\{3,1\}) \sim 0.3361$、$\psi_{\mathcal{P}(4)}(\{4\}) \sim 0.2369$。纤维的大小 $c_\wp = \#p^{-1}(\wp)$ 已在上面的引理 9.4 中计算得出，因此使用平稳分布的表达式 (9.11)，并将结果归一化为概率分布，我们得到 (9.15)。□

9.4. 侧向合并问题。 仅从$n = 4$的显式示例，就很容易看出侧向合并箭头所引发的一般性问题，即便只考虑最小的侧向合并箭头也是如此：它们倾向于将动力学推向更加不连通的结构，从而完全阻碍结构形成过程。我们可以从平稳分布中各结构的权重看出这一点：完全连通结构（$\wp = \{4\}$的原像中的树）的权重很低，而最不连通的结构（$\wp = \{2,1,1\}$的原像中的结构）则获得最高权重。

请注意，在 $n = 4$的情形下，其行为就已不同于 [26]中分析的 $n = 3$ 情形。在 $n = 3$的情形下，考虑外部合并、侧向合并和内部合并箭头时，动力学具有均匀平稳分布，所有工作空间的权重都相同。这已经表明，侧向合并带来了一个问题：动力学不会收敛到完全形成的句子（即由单棵树构成的工作空间）。通过引入成本函数可以解决这个问题，见§11。

事实上，这种行为显然无法反映语言结构形成的现实，这意味着在图 $\mathcal{G}_{n,\,A}$ 中，不同类型的箭头——外部合并、内部合并和侧向合并——必须具有不同的权重，并由某种代价函数决定。如前文所述，侧向合并相对于语言学文献中采用的某些最优性衡量标准，确实不是最优的。此外，只有少数语言现象可能需要借助侧向合并来解释，因此它在合并推导中应该极少出现；然而，我们在此讨论的马尔可夫链的动力学 $\hat{\mathcal{K}}^{A,\,n}$ 却与此相反，其中侧向合并箭头最终在动力学中占据主导地位。这意味着，正如 [26], 中所讨论的，一个现实的模型必须用加权版本取代 $\hat{\mathcal{K}}^{A,\,n}$，以计入不同的代价。我们将在第 11 节讨论这一点。

10. Merge dynamics with the contraction coproduct

在引入成本函数并分析它们对动力学不同部分（EM、IM、SM）的影响之前，本节先讨论这里所述动力学的可能变体。文献 [22] 指出，工作空间上的余积有多种可能形式，可用于定义合并操作。前几节的所有讨论和结果都基于使用文献 [22] 所称的“删除”余积：对于 $F_v = T_{v_1} \sqcup \cdots \sqcup T_{v_k}$中由互不相交的可访问项组成的集合，$T/^d F_v$被定义为通过边收缩，从由可容许切割（即移除各顶点 v_i （可访问项 T_{v_i}的根）上方的边）得到的非满二叉树（包括非分支顶点）中，所获得的最大满二叉根树。我们在分析动力学时使用的这种余积 Δ^d的关键性质是，叶子的数量在这种分解下具有可加性：

$$\#L(T) = \#L(F_{\underline{v}}) + \#L(T/^d F_{\underline{v}})\,.$$

就语言学模型而言，这个余积适用于核心合并动态与外化之间的接口，其中移位的移动痕迹并未表达出来。

[22] 中使用的另一种余积是“收缩”余积，其中商 $T/^c F_v$ 是通过将 $F_v = T_{v_1} \sqcup \cdots \sqcup T_{v_k} \subset T$ 的每个分量 T_{v_i} 收缩到其顶点 v_i而得到的。其中每个

这些顶点 v_i 由此成为 $T/^cF_v$ 的新叶子，其标签是 $\mathcal{SO}_0$ 的一个元素，用来保留被提取的可访问项移位的“轨迹”。这种标签用符号 T_{v_i} 表示，以标明它对应于移位时下方副本 T_{v_i} 的消去。这种余积形式 $\overline{\Delta^c}$ 不再满足叶子数量的可加性，因为现在我们有

$$\#L(T) + k = \#L(F_{\underline{v}}) + \#L(T/^cF_{\underline{v}})\,,$$

用于 $F_v = T_{v_1} \sqcup \cdots \sqcup T_{v_k}$。

10.1. 移位痕迹的标签。 在进一步讨论利用这一余积构造的合并操作（如 [22]§1.3 中所述）之前，有必要澄清用于标记商 $T/^cT_v$ 的新叶 v 的标记系统 T_v 的含义，其中子树 $T_v \subset T$ 被收缩到其根顶点 v。本节将对 [22]§1.3 中的简要讨论作出更详细的说明和扩展。

商 $T/^cT_v$ 的形式需要加以解释：由于由 T_v 收缩所产生的叶的标签通常写作 T_v，这看起来似乎无法与具有固定有限叶标签集相容。事实上，如果标签 T_v 确实包含关于 T_v 的完整信息，并加上一个“消去符号”以将其标记为移位的“轨迹”，那么它将产生一个可数无限的标签集，因为 T_v 可以是任意句法对象（因为 T 本身可以是任意的），而所有可能的句法对象的集合是可数无限的。然而，标签 T_v 实际上不必保留 T_v 作为句法对象（即完整的非平面二叉有根树）的全部结构。毕竟，被提取的项 T_v 仍然存在于余积的另一侧，因此信息并未丢失。所需要的只是 T_v 记住 T_v 的某个足以进行解释的特征：也就是说，它要记住移位元素 T_v 在其原始位置所扮演的角色。我们将在此说明，这可以通过一种标记算法实现：该算法使用头函数（其含义见 [22], ，我们将在下文回顾）以及 T_v 的结构，为非叶顶点（包括根 v）分配标签。

首先，我们对词汇项和句法特征的初始集合 $\mathcal{SO}_0$ 作出更精细的选择。在前几节给出的表述中并不需要这样做，因为只有涉及余积 Δ^c 时，这种选择才相关。

定义 10.1。 我们将 $\mathcal{SO}_0$ 替换为一个形如 $\mathcal{SO}_0 \sqcup \mathcal{SO}_0$ 的集合，其中 $\mathcal{SO}_0$ 表示形如以下形式的符号集合

$$\mathcal{SO}_0 := \{\alpha \,|\, \alpha \in \mathcal{SO}_0\}\,. \tag{10.1}$$

我们将 $\mathcal{SO}_0$ 的元素称为“迹标签”。

我们还回顾一下 [22] 第 1.13 节中给出的头函数定义。

定义 10.2。 完整非平面二叉根树 T 上的头函数是一个函数 $h_T : V^o(T) \to L(T)$，它从 T 的非叶顶点集合映射到叶子集合，并满足 $h_T(v) \in L(T_v)$；此外，若 $T_w \subset T_v$ 且 $h_T(v) \in L(T_w)$，则 $h_T(v) = h_T(w)$。

现在我们可以更精确地定义标签 T_v 的含义如下。

定义 10.3。 设 $T \in \mathfrak{T}_{\mathcal{SO}_0}$ 是一个配有头函数 h_T 的句法对象，且 $T_v \subset T$ 是一个可访问项。若 $T_v = \ell$ 是一个由元素 $\alpha_\ell \in \mathcal{SO}_0$ 装饰的单叶，则我们直接将 T_v 定义为标签 $\alpha_\ell \in \mathcal{SO}_0$。若 T_v 是一个包含多个叶子的子树，则标签 T_v 定义为 $\alpha_{h_T(v)}$ 中的元素，其中 $\ell = h_T(v)$ 是头函数 h_T 在顶点 $v \in V^o(T)$ 处的取值，而 $\alpha_{h(T_v)}$ 是该叶子在 $\mathcal{SO}_0$ 中的标记。

我们需要验证，这种标记 T_v 的定义与余积 Δ^c 的余结合性相容。

引理 10.4。 由 $\alpha_{h(T_v)}$ 给出的、以 $h = h_T$ 为头函数的收缩的可达项 T_{v_i} 的叶子标签 T_{v_i} 余积 Δ^c 是余结合的。

证明。如 [22] 所示，头函数将 T 的顶点划分为互不相交的路径 γ_ℓ，其中对于$\ell \in L(T)$，凡是包含多个顶点的 γ_ℓ，其另一端顶点 v_ℓ 称为 ℓ的“最大投影”。对于 γ_ℓ 中的每个顶点 v，都有 $h_T(v) = \ell$。若叶子 ℓ 的最大投影是 T 的根，则称其为 T的头，记作 $\ell = h(T)$。将顶点划分为路径 γ_ℓ，即可确定 T所有顶点的标记：路径 γ_ℓ 上的顶点获得最初分配给叶子 ℓ的相同标签 α_ℓ。因此，标记$\alpha_{h(T_v)}$ 将标签$\alpha_{h_T}(v)$ 分配给 T_v 的根顶点 v，并附加消去符号$\alpha_{h_T}(v)$，以表示新叶子 v 是移位的“轨迹”。我们需要检验的是，这种标签分配是否保持余积 Δ^c 的余结合性。余结合关系 $(\Delta^c \otimes \mathrm{id}) \circ \Delta^c = (\mathrm{id} \otimes \Delta^c) \circ \Delta^c$ 是一个恒等式，涉及两个连续的可容许切割 C、C'：第一次切割 C 对应于首次应用余积 Δ^c，而第二次切割 C' 可以发生在第一次 Δ^c输出的左侧或右侧（$\Delta^c \otimes \mathrm{id}$或$\mathrm{id} \otimes \Delta^c$）。不带叶子标签时，这种形式的余积的余结合性早已众所周知，因此问题仅在于标记在余结合恒等式两侧是否一致。由于余积是乘法的，我们只需在树 T上检验这一点，而不必在任意森林 F上检验。对于可容许切割 C，我们用记号 $\pi_C(T) = F_v$表示由该切割提取的森林，它出现在余积相应项的左侧；并用$\rho^c_C(T) = T/^c F_v$表示右侧的商项。因此，由切割 C确定的余积项为 $\pi_C(T) \otimes \rho^c_C(T) = F_v \otimes T/^c F_v$，而完整的余积为 $\Delta^c(T) = \sum_C \pi_C(T) \otimes \rho^c_C(T)$，其中包含本原部分 $T \otimes 1 + 1 \otimes T$，其形式为两个平凡切割（根上方、叶子下方）。余结合关系比较如下形式的项

$$\pi_{C'}(\pi_C(T)) \otimes \rho^c_{C'}(\pi_C(T)) \otimes \rho^c_C(T)$$

在 $(\Delta^c \otimes \mathrm{id})(T)$ 中，含有形如

$$\pi_{C'}(T) \otimes \pi_C(\rho^c_{C'}(T)) \otimes \rho^c_C(\rho^c_{C'}(T))$$

在 $(\mathrm{id} \otimes \Delta^c)(T)$ 中，其中在第二项里，C' 是第一处切割，C 是第二处切割。项 $\pi_{C'}(\pi_C(T)) = \pi_{C'}(T)$（因为切割 C'位于 C下方）不涉及任何迹标记，因此无需检查。中间项 $\rho^c_{C'}(\pi_C(T)) = \pi_C(\rho^c_{C'}(T))$ 在顶点 w_j 处有迹标签，这些顶点是 $\pi_{C'}(T) = \sqcup_j T_{w_j}$ 中各树的根。在 $\rho^c_{C'}(T))$ 中，这些顶点被标记为 $\alpha_{h_T}(w_j)$，并且在 $\pi_C(\rho^c_{C'}(T))$ 中保持相同的标签；而在 $\rho^c_{C'}(\pi_C(T))$ 中，它们被标记为 $\alpha_{h_{T_{v_i}}}(w_j)$，其中 $\pi_C(T) = \sqcup_i T_{v_i}$。但头函数的性质保证，只要 $T_{w_j} \subset T_{v_i} \subset T$，就有 $h_{T_{v_i}}(w_j) = h_T(w_j)$，因此标签相匹配。类似地，对于项 $\rho^c_C(T) = \rho^c_C(\rho^c_{C'}(T))$，迹标签位于顶点 v_i 处，这些顶点是 T_{v_i}的根顶点，且带有 $\pi_C(T) = \sqcup_i T_{v_i}$。在 $\rho^c_C(T)$ 和 $\rho^c_C(\rho^c_{C'}(T))$ 两种情况下，这些顶点都被标记为 $\alpha_{h_T}(v_i)$，因此同样不存在标记不匹配。□

这种移位痕迹标记方法要求，所有工作空间 $F = \sqcup_i T_i$ 中的所有分量 T_i 都必须指定一个头函数 h_{T_i}。对于一棵有 $k_i = \#L(T_i)$ 个叶子的树 T_i，头函数有 $2^{k_i - 1}$ 种可能的选择。这里我们简单地假设已经做出了这样一种选择。更切合实际的做法是，采用一个能够正确模拟句法头和扩展投射结构的头函数，详见 [24]。我们将在 §15 中回到这一问题并展开讨论。

10.2. **带有收缩余乘的合并操作**。那么，自然会产生这样的问题：如果我们定义

$\mathfrak{M}^c_{S,\ S'}$ 为

$$\mathfrak{M}^c_{S,S'} = \sqcup \circ (\mathcal{B} \otimes \mathrm{id}) \circ \delta^c_{S,S'} \circ \Delta^c \tag{10.2}$$

而不是前几节讨论的合并算子，wit h

$$\mathfrak{M}_{S,S'} = \sqcup \circ (\mathcal{B} \otimes \mathrm{id}) \circ \delta_{S,S'} \circ \Delta^d \,.$$

需要指出的是，从余积 Δ^d 转到 Δ^c 时，我们也改变了合并所使用的余积项的选择 $\delta^c_{S,\ S'}$。（这一点反映在对合并影响可访问项数量的计算中，尽管文中没有明确说明，因此我们在这里将其阐明。）在语言学中，先前移位的轨迹不再通过推导中连续应用合并而被提取出来，以用于进一步移位：换言之，轨迹不再移动。为了反映这一事实，可访问项的选择 $\delta_{S,\ S'}$（在余积 Δ^d 的情况下），在 [22] 中定义为 $\delta_{S,\ S'} = \gamma_{S,\ S'} \otimes$id，其中

$$\gamma_{S,S'}(F) = \begin{cases} F & F = S \sqcup S' \\ 0 & \text{otherwise,} \end{cases}$$

可以用稍作修改的形式定义为

$$\delta^c_{S,S'} = \gamma^c_{S,S'} \otimes \tilde{\gamma}^c_{S,S'} \,,$$

其中

$$\gamma^c_{S,S'}(F) = \begin{cases} F & F = S \sqcup S' \ \text{ and } S, S' \notin \mathfrak{T}_{\mathcal{SO}_0} \\ 0 & \text{otherwise,} \end{cases} \tag{10.3}$$

并且，对于 $F = T_1 \sqcup \cdots \sqcup T_r$,

$$\tilde{\gamma}^c_{S,S'}(F) = \begin{cases} F & T_i \notin \mathfrak{T}_{\mathcal{SO}_0},\ \forall i = 1, \ldots, r \\ 0 & \text{otherwise.} \end{cases} \tag{10.4}$$

请注意，$\gamma^c_{S,S'}$ 并不会阻止提取任何术语 $S = T_w$，其中 $T_w \in \mathfrak{T}_{\mathcal{SO}_0 \sqcup\, \mathcal{SO}_0}$ 的一些叶子由 $\mathcal{SO}_0$ 中的元素装饰：它只会阻止提取叶子全部 标记为 $\mathcal{SO}_0$ 中的迹标签的可访问项。

相应地，算子 $\tilde{\gamma}^c_{S,S'}$ 会阻止在余积右侧创建仅含迹叶的分量。这里还纳入约束 $\tilde{\gamma}^c_{S,S'}$，是为了更好地将合并的作用与余积的两种形式联系起来。例如，涉及术语 $T/^c T_v \in \mathcal{T}_{\mathcal{SO}_0}$ 的内部合并转换，在余积 Δ^d中没有对应形式，因此从语言学角度来说，它在外化阶段不可行。

同时满足约束 $\gamma^c_{S,S'}$ 和 $\tilde{\gamma}^c_{S,S'}$，可确保在 $\mathcal{K}^c = \sum_{S,S'} \mathfrak{M}^c_{S,S'}$ 中能够发生的所有合并转换，在 $\mathcal{K} = \sum_{S,S'} \mathfrak{M}_{S,S'}$中也都有对应形式；这表达了这样一种观点：同一结构形成机制总是同时映射到外化接口（其中不表达迹）和概念—意向接口（也称句法—语义接口，其中保留迹以供解释）。

在本节余下部分，我们将讨论基于余积 Δ^c 的形式下合并动力学的行为。

10.3. 收缩余乘下合并图的边。 使用（10.2）中定义的合并算子得到的动力系统图中的箭头，同样对应于外部合并、内部合并和侧向合并操作。

(1) 外部合并箭头与前几节所讨论的完全相同：无论选择 Δ^d 还是 Δ^c 余积，它们都保持不变，因为它们只

涉及余积的本原部分，该部分将工作空间划分为若干组成部分，而不提取可访问项。由于此处不涉及商，两种余积具有相同的本原部分。

(2) 内部合并箭头对应于复合 $\mathfrak{M}^c_{T_v,\ T/^cT_v} \circ \mathfrak{M}_{T_v,\ 1}$，因此不同于余积为 Δ^d 的情况，这是由于存在一个标签为T_v 的额外叶子 $T/^cT_v$

(3) 侧向合并箭头也同样不同于余积为 Δ^d 的情况。由于我们再次只考虑仅提取叶子的最小 SM 箭头，因此有以下情形

- $\mathfrak{M}(\beta,\ \alpha) \sqcup T/^c\alpha$（当 β 和 T 是该工作空间中侧向合并作用对象的分量时），此时得到的分量 $T/^c\alpha$ 仍保留着 T 的叶子 α，但其标签已改为 α,
- $\mathfrak{M}(\beta,\ \alpha) \sqcup T/^c$（$\alpha \sqcup \beta$），其中商 $T/^c$（$\alpha \sqcup \beta$）与 T 具有相同的叶子，但原先分别标记为 α 和 β 的叶子现在被重新标记为 α 和 β,
- $\mathfrak{M}(\beta,\ \alpha) \sqcup T/^c\alpha \sqcup T'/^c\beta$其中，类似地，所得分量 $T/^c\alpha$和 $T'/^c\beta$的叶子数量与 T 和 T'相同，而 α 和 β 的标签则分别改为 α和 β。

在所有涉及商 $T/^cT_v$（或 $T/^cF_v$）的情形中，应按照定义 10.3 来理解标签 T_v。

定义10.5。 我们将收缩合并图 $\mathcal{G}_{\mathfrak{F}_{\mathcal{SO}_0 \sqcup\ \mathcal{SO}_0}}$ 定义为一个无限、局部有限的图，它有可数无限个顶点 $\mathfrak{F}_{\mathcal{SO}_0 \sqcup\ \mathcal{SO}_0}$，每个顶点有有限条入边和出边，这些边由上述所有可能的外部合并、内部合并和侧向合并变换构成。

10.4. 瞬态动力学与投影。 我们在此说明，从余积 Δ^c 的角度考察合并动力学并不能提供任何有用信息，因为此时动力学涉及可数无限个状态（它不再是有限马尔可夫链），且所有状态都会变为暂态，因此不存在描述该动力学长期行为的平稳分布。不过，我们还将说明，Δ^d 情形下的合并动力学提供了完整信息，也可据此确定 Δ^c 情形下的动力学。其关键原因在于，在语言学中，先前移位的痕迹不会再被提取出来以进行进一步移位。

命题10.6。 设$\mathcal{K}^c = \sum_{S,S'} \mathfrak{M}^c_{S,S'}$ 为合并在工作空间线性张成空间$\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0 \sqcup\ \mathcal{SO}_0})$上的作用，其中 $\mathfrak{M}^c_{S,S'}$ 如(10.2)所述。此作用不保持由有限集$\mathfrak{F}_{A,n}$张成的子空间，并且所有满足 $F \notin \mathfrak{F}_{\mathcal{SO}_0}$ 的 $F \in \mathfrak{F}_{\mathcal{SO}_0 \sqcup\ \mathcal{SO}_0}$ 都是由 $\mathcal{K}^c$定义的动力学的暂态状态。

证明。 如§10.3所述，$\mathcal{G}_{\mathfrak{F}_{\mathcal{SO}_0}}$ 中的 EM 箭头与$\sqcup_n \mathcal{G}_{n,A}$ 中的相同，因此它们确实保持$\mathfrak{F}_{A,n} = V(\mathcal{G}_{n,A})$。然而，IM 箭头和侧向合并箭头并非如此。事实上，对于ˆ IM 箭头的情形，若$F = T \sqcup F \in \mathfrak{F}_{A,n}$，则其在一个使用余积的 Δ^c 形式、并作用于 T 分量的 IM 变换下的像是一个新的工作空间

$$F' = \mathfrak{M}(T_v, T/^cT_v) \sqcup \hat{F} \in \mathfrak{F}_{B,n+1}\,, \qquad \text{with} \qquad B = A \cup \{\cancel{T_v}\}$$

并带有定义 10.3 中的标签T_v。类似地，在最小 SM 的情形下，如 §10.3 所示，我们可以从 $F \in \mathfrak{F}_{A,\ n}$ 得到一个新的工作空间

$$F' = \begin{cases} \mathfrak{M}(\alpha,\beta) \sqcup T/^c\alpha \sqcup \hat{F} \ \in \mathfrak{F}_{A\cup\{\cancel{\alpha}\},n+1} & \text{for } F = \beta \sqcup T \sqcup \hat{F} \\ \mathfrak{M}(\alpha,\beta) \sqcup T/^c(\alpha \sqcup \beta) \sqcup \hat{F} \ \in \mathfrak{F}_{A\cup\{\cancel{\alpha},\cancel{\beta}\},n+2} & \text{for } F = T \sqcup \hat{F} \\ \mathfrak{M}(\alpha,\beta) \sqcup T/^c\alpha \sqcup T'/^c\beta \sqcup \hat{F} \ \in \mathfrak{F}_{A\cup\{\cancel{\alpha},\cancel{\beta},\ \mathrm{n}+2\}} & F = T \sqcup T' \sqcup \hat{F}\,. \end{cases}$$

所得的图 $\mathcal{G}_{\mathfrak{F}_{\mathcal{SO}_0 \sqcup \mathcal{SO}_0}}$ 不可能是强连通的，因为 IM 和 SM 箭头总是将任意顶点子集 $\mathfrak{F}_{A,n}$ 映射到 $\mathfrak{F}_{B,n+2}$ 的顶点子集 $\mathfrak{F}_{B,n+1}$，而 EM 箭头将$\mathfrak{F}_{A,n}$ 映射到 $\mathfrak{F}_{A,n}$；同时，不存在能减少叶子数量的有向箭头。由于每个 $F \in \mathfrak{F}_{A,n}$ 都有一些出射的 IM 或 SM 箭头，这意味着每个$F \in \mathfrak{F}_{A,n}$ 都是动力系统 $\mathcal{K}^c$的暂态状态。□

这表明，使用余积 Δ^c 而非 Δ^d时，这里没有与我们在前几节研究过的霍普夫代数马尔可夫链 $\mathcal{K}$ 及其平稳分布相类似的对象；该平稳分布支配长期动力学，并提供有关动力学行为的信息。

另一方面，这并不意味着我们对$\mathcal{K}^c$ 的行为一无所知。事实上，下面的结果表明，作用于工作空间$F \in \mathfrak{F}_{\mathcal{SO}_0}$ 上的 $\mathcal{K}^c$的动力学，完全由其在 Π_d 上的投影像$_c$（$\mathcal{V}(\mathfrak{T}_{\mathcal{SO}_0})$）决定（参见 [22] 的 §1.3）。

定义 10.7。设 $\Pi_{d,\ c}$：$\mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0 \sqcup \mathcal{SO}_0}) \twoheadrightarrow \mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0})$ 为如下投影：从 F 的每个分量 T 中删去所有带迹标记的叶子，然后收缩边，从而得到由剩余部分确定的最大满二叉根树。

基于余积的两种形式 Δ^d 和 Δ^c，动力系统 $\mathcal{K} = \sum_{S,S'} \mathfrak{M}_{S,S'}$ 以及 $\mathcal{K}^c = \sum_{S,S'} \mathfrak{M}^c_{S,S'}$的迭代满足以下关系。

命题 10.8。对于 $F \in \mathfrak{F}_{A,n} \subset \mathfrak{F}_{\mathcal{SO}_0}$，在 $\mathcal{K}^c$ 的迭代下，轨道 $(\mathcal{K}^c)^m(F)$ 包含于纤维 $\Pi^{-1}_{d\,c}(\mathcal{K}^m(F)) \subset \mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0 \sqcup \mathcal{SO}_0})$ 中，并且完全由 F 在 $\mathcal{K}$作用下的轨道决定。

证明。我们首先证明，对于 $F \in \mathfrak{F}_{\mathcal{SO}_0}$，有

$$\Pi_{d,c} \circ \mathcal{K}^c = \mathcal{K}\,. \tag{10.5}$$

对于森林 $F \in \mathfrak{F}_{\mathcal{SO}_0}$，有

$$(\mathrm{id} \otimes \Pi_{d,c}) \circ \Delta^c(F) = \Delta^d(F)\,, \tag{10.6}$$

或者等价地

$$\pi_C(T) \otimes \Pi_{d,c}(\rho^c_C(T)) = \pi_C(T) \otimes \rho_C(T))\,.$$

注意，在此情形下，由于 F没有迹标签，从 $F \in \mathfrak{F}_{A,n}$中提取出的项 $F_v = \pi_C(F)$ 的各个分量都不可能属于 $\mathfrak{T}_{\mathcal{SO}_0}$。因此，我们有 $\delta^c_{S,S'}\Delta^c(T) = \delta_{S,S'}\Delta^d(F)$，并在此情形下得到

$$\mathfrak{M}_{S,S'}(F) = \sqcup \circ (\mathcal{B} \otimes \mathrm{id}) \circ (\gamma_{S,S'} \otimes \Pi_{d,c})\Delta^c(F) = \sqcup \circ (\mathrm{id} \otimes \Pi_{d,c}) \circ (\mathcal{B} \otimes \mathrm{id}) \circ \delta^c_{S,S'}\Delta^c(F) =$$
$$\sqcup \circ (\Pi_{d,c} \otimes \Pi_{d,c}) \circ (\mathcal{B} \otimes \mathrm{id}) \circ \delta^c_{S,S'} \circ \Delta^c(F) = \Pi_{d,c} \circ \mathfrak{M}^c_{S,S'}(F)$$

因此我们得到（10.5）。现在我们归纳地进行。假设从一个 $F \in \mathfrak{F}_{\mathcal{SO}_0}$ 和~开始，应用 $\mathcal{K}^c$ 作用的 m 次迭代后，我们得到 $\Pi_{d,\ c} \circ (\mathcal{K}^c)^m(F) = \mathcal{K}^m(F)$。设 F_i be ~ ~为 $(\mathcal{K}^c)^m(F) = \sum_i c_i F_i \in \mathcal{V}(\mathfrak{F}_{\mathcal{SO}_0 \sqcup \mathcal{SO}_0})$ 的任意一个分量。当我们取 $\mathcal{K}^c(F_i) = \sum_j c_{ij} F_{ij}$~时，每个分量 F_{ij} 的迹叶要么已经是 F_i 的迹叶，要么是由 $\mathcal{K}^c$ 的内部合并和侧向合并箭头添加的，因此我们有

$$\Pi_{d,c}(\mathcal{K}^c(\tilde{F}_i)) = \Pi_{d,c}(\mathcal{K}^c(\Pi_{d,c}(\tilde{F}_i))\,.$$

利用（10.5），我们进而看到

$$\Pi_{d,c}(\mathcal{K}^c(\Pi_{d,c}(\tilde{F}_i)) = \mathcal{K}(\Pi_{d,c}(\tilde{F}_i))$$

因此，由线性性可得

$$\Pi_{d,c}((\mathcal{K}^c)^{m+1}(F)) = \Pi_{d,c}(\mathcal{K}^c(\Pi_{d,c}((\mathcal{K}^c)^m(F))) = \mathcal{K}(\Pi_{d,c}((\mathcal{K}^c)^m(F)))$$

并根据归纳假设，我们进而得到

$$\Pi_{d,c}((\mathcal{K}^c)^{m+1}(F)) = \mathcal{K}^{m+1}(F)\,. \tag{10.7}$$

这意味着，$F \in \mathfrak{F}_{A,n}$ 的轨道 $(\mathcal{K}^c)^m(F)$ 包含在 $\Pi_{d\,c}^{-1}$, $(\mathcal{K}^m(F))$ 中。还可以看出，对于 $F \in \mathfrak{F}_{\mathcal{SO}_0}$，轨道 $(\mathcal{K}^c)^m(F)$ 完全由轨道 $\mathcal{K}^m(F)$ 确定，因为每次应用 $\mathcal{K}^c$ 时，确定像所需的唯一额外信息，就是余积 Δ^c 所产生的迹叶。只要知道提取可访问项 T_v 的顶点 v，以及分量 $T \supset T_v$ 中头函数 $h_T(v)$ 的值，就能确定这些信息。和上文一样，也可以通过归纳证明这一点。在第一步中，对于 $\mathcal{K}^c$ 和 $\mathcal{K}$，提取出的 T_v 相同，因为轨道始于一点 $F \in \mathfrak{F}_{\mathcal{SO}_0}$。因此，将 F 与 $\mathcal{K}(F)$ 的各分量进行比较，就能确定每个箭头中涉及的项 T_v。这里我们假设所涉及工作空间的所有分量都具有头函数（我们将在 §15 中再讨论这一点）。因此，对于 $T_v \subset T$，根据头函数的性质，它满足 $h_T(v) = h_{T_v}(v) = h_{T'}(v)$；其中 T' 是 $\mathcal{K}(F)$ 中包含顶点 v 的分量，因此可以由 $\mathcal{K}(F)$ 计算得出。归纳地，假设在 $(\mathcal{K}^c)^m(F)$ 中，通过头函数计算出的迹叶及其标签，都由可从轨道 $\{\mathcal{K}^k(F)\}_{k\le m}$ 计算得出的数据所确定。下一个迭代 $\mathcal{K}^c((\mathcal{K}^c)^m(F))$ 不会改变先前的迹叶及标签，但会通过 $\mathcal{K}^c$ 中的内部合并和侧向合并箭头添加新的迹叶及标签。同样，这些取决于提取出的可访问项，以及其根顶点处头函数的值。与单次迭代的情况一样，这些只取决于 $(\mathcal{K}^c)^m(F)$ 中的数据，而这些数据由轨道 $\{\mathcal{K}^k(F)\}_{k\le m}$ 确定。□

因此，结论是，合并动力学已由我们在前几节研究过的霍普夫代数马尔可夫链 $\hat{}_{\mathcal{K}^{A,\,n}}$ 完全描述，而将余积从 Δ^d 改为Δ^c 并不会提供任何额外信息。

从语言学模型的角度来看，命题 10.8 的结果可以作如下解释。$\mathcal{K}^c$ 的动力学决定了（通过投影 $\Pi_{d,\,c}$），而反过来，$\mathcal{K}$ 的动力学也决定了它（依据命题 10.8 中给出的论证）。这意味着，尽管迹被认为是语义解析所必需的，但结构构建动力学本身并不需要迹；无需使用迹叶，也能完全确定结构构建动力学。正如我们所论证的，这是因为在语言学中，移位不再提取迹。

11. Weighted Merge dynamics: cost functions

我们从前几节得出的结论，与语言学文献中关于合并在 [7] 和 [8]的工作空间上所起作用的表述一致。具体而言，某个自然的代价函数必须对合并的不同形式赋予不同权重，其中外部合并和内部合并是最优的，而侧向合并则是次优的。

正如 [7] 和 [8], 中提出的，Merge 预期满足两种主要的最优性形式。一种称为最小搜索，它优化在工作空间中搜索 Merge 所作用的可访问项的代价。另一种称为资源限制，尤其是最小产出，其表述是在 Merge 转换对工作空间连通分量数量和可访问项数量的影响之间取得平衡。有了这两种最优性形式，还可以加入另一种形式，[22]将其称为“无复杂度损失”。在语言学文献中，它通常与“扩展条件”归为一类：该条件要求结构通过 Merge逐步生成（扩展条件的第二部分还要求在根处进行这种生成，[26]对此有详细讨论）。

最小产出条件在 [22]（以及 [26]）中的表述，与 [7] 和 [8]的原始表述略有不同。我们将在此简要回顾这些代价函数，具体见

§11.1，采用 [26]中的表述。使用代价函数的原始形式并不会影响此处陈述的主要结果，因此我们采用 [22] 和 [26], 的表述，因为它能更直观地体现树拓扑。

接着，我们将在 §12 中介绍一种技术，用于估计代价加权合并动力学的佩龙–弗罗贝尼乌斯特征值和特征向量关于权重参数的首阶项，以及相关霍普夫代数马尔可夫链的平稳分布。

量级代数由 min-plus 半环（也称为热带半环）支配。因此，代价加权合并的佩龙-弗罗贝尼乌斯数据的首阶项估计，可表述为热带半环中的佩龙-弗罗贝尼乌斯问题。

这类热带佩龙–弗罗贝尼乌斯问题的一般方法，是通过在合并图中识别临界回路（代价最小的回路）来求出佩龙-弗罗贝尼乌斯特征值的阶，并通过为合并图构造最优树形图来构造佩龙-弗罗贝尼乌斯特征空间的一组基。我们将在 §12 中讨论这一方法的细节及其结果。

这项分析的结论是，由于临界回路并不唯一，我们识别出的不是唯一的 Perron-Frobenius 特征向量，而是特征空间的一组基。这组基的形式表明，在主导阶上，加权动力学会远离那些在非加权动力学中因侧向合并（SM）的影响而占主导地位的较不连通结构。另一方面，与 [26] 中针对 $n = 3$ 的加权动力学所观察到的收敛至树结构（即划分 $\wp = \{n\}$）不同，我们发现，任何包含至少某些 $k_i \geq 3$（因而具有非平凡 IM 自环）的划分 $\wp \in \mathcal{P}'(n)$ 都会对 Perron-Frobenius 特征空间有所贡献。

这表明，尽管 [26] 中描述的较简单情形 $n = 3$ 似乎说明了相反的情况，但结合了最小搜索（采用 [26] 中的自底向上表述）、最小产出（按 [22] 中的表述）和复杂度损失的总成本函数，并不足以完全消除侧向合并对动力学的不利影响（同时保持所需的强连通性），也无法确保收敛至树结构。

我们通过修正代价函数，并在其中引入一项依赖于与划分$\wp \in \mathcal{P}'(n)$相关联的分布 $\mathbb{P}_\wp$ 的香农熵的项，来说明如何解决这一问题。这一项仅刻画了组合多项式系数的主阶，而该系数描述了如何将工作空间的 n 叶子划分到不同的树分量中。我们证明，代价函数中加入这一额外量 $\mathrm{Sh}(\mathbb{P}_\wp)$，足以推动动力学朝连通工作空间（由单棵树构成）演化。具体而言，我们证明，采用这一修正后的代价函数加权的合并马尔可夫链的平稳分布 $t^{\mathfrak{c}}$ 在极限 $t \to 0$ 下收敛于树上的均匀分布。换言之，在$\mathcal{G}_{n,A}$上的$\mathcal{K}^{(A,n)}$ 动力学，采用这一修正后的代价函数加权后，会收敛到$\mathcal{G}^{\mathrm{IM}}_{n,A,\{n\}}$上的 IM 动力学。平稳分布中所有分量的主阶 t 都可以用图$\mathcal{G}_{n,A}$ 中沿着$\mathcal{G}_{n,A}$最优树形图的某些边回路的代价显式表示。

事实上，我们证明了一个更强的结论：只使用香农熵 $\mathrm{Sh}(\mathbb{P}_\wp)$给出的代价函数，也能得到相同的结果，完全不必纳入最小搜索、最小产出和复杂度损失代价函数。这些通常在语言学中被讨论的代价函数，仅在初始阶段用于筛选涉及原子分量的“最小”侧向合并箭头；动力学则完全由信息论成本函数所驱动。

11.1. 组合代价函数。这里我们考虑 [26], 中讨论的三种主要代价函数类型：自下而上的最小搜索，以及两种资源限制代价：最小产出和复杂度损失。

11.1.1. 最小搜索代价。与 [26], 中一样，我们考虑最小搜索代价的“自底向上”版本（而不是 [22] 中使用的 “ 自顶向下 ” 版本）。在这种最小搜索方法中，我们假定，句法结构是自底向上构建的，因此搜索并非如有根树上的搜索算法通常那样从根开始向树下方进行，而是从叶子处的词汇项开始，搜索能够构成成分结构的叶子子集，也就是说，搜索能够确定一个可访问项，该项可被提取出来供合并操作使用。从这一角度看，搜索代价取决于需要识别的叶子集合的（相对）大小。这便导出了如下代价函数定义：

$$\mathfrak{c}_{MS}(T_v) = \frac{\ell(T_v)}{\ell(T)},$$

其中，$\ell(T) = \#L(T)$ 表示叶子的数量（即工作空间霍普夫代数中的分次）。这为提取可访问项赋予了代价。相应商运算 T/T_v 的代价则设为

$$\mathfrak{c}_{MS}(T/T_v) = \frac{\ell(T/T_v)}{\ell(T)} = 1 - \frac{\ell(T_v)}{\ell(T)} = 1 - \mathfrak{c}_{MS}(T_v).$$

与 [22] 中一样，我们也为两个分量或可访问项的合并A, B 赋予相应的代价，其形式为

$$\mathfrak{c}_{MS}(\mathfrak{M}(A,B)) = \mathfrak{b}(A,B) - \mathfrak{c}(A) - \mathfrak{c}(B),$$

其中，若 A 和 B 来自工作空间中的同一分量，则 $\mathfrak{b}(A, B) = 1$；若它们来自不同分量，则 $\mathfrak{b}(A, B) = 2$。这一选择体现了这样的理念：最小搜索会先在固定的分量内搜索项（代价较低），然后再跨不同分量搜索项（代价较高）。利用这个代价函数，我们可以看到，在合并的不同情形下，外部合并 $\mathfrak{M}(T, T')$、内部合并 $\mathfrak{M}(T_v, T/T_v)$、SM(1) $\mathfrak{M}(T_v, T')\sqcup T/T_v$、SM(2) $\mathfrak{M}(T_w, T'_w)\sqcup T/T_v\sqcup T'/T'_w$ 和 SM(3) $\mathfrak{M}(T_v, T_w) \sqcup T/(T_v \sqcup T_w)$ 的相应代价为

	外部合并	内部合并	SM(1)	SM(2)	SM(3)
$\mathfrak{c}_{MS}$	0	0	$2 - \mathfrak{c}_{MS}(T_v)$	$2 - \mathfrak{c}_{MS}(T_v) - \mathfrak{c}_{MS}(T'_w)$	$1 - \mathfrak{c}_{MS}(T_v) - \mathfrak{c}_{MS}(T_w)$

特别地，对于一棵树 T，若其满足 $k = \#L(T)$，并且我们只考虑最小 SM 变换，其中提取和合并操作涉及的项 T_v、T_w、T'_w、T' 均为由单个叶子构成的原子分量，则可得到代价

	外部合并	内部合并	最小 SM(1)	最小 SM(2)	最小 SM(3)
$\mathfrak{c}_{MS}$	0	0	$2 - 1/k_i$	$2 - 1/k_i - 1/k_j$	$1 - 2/k_i$

这意味着，对于工作空间 F，其中 $n = \#L(F)$ 被划分到分量 $F = T_1 \sqcup \cdots \sqcup T_r$ 中，且 $n = k_1 + \cdots + k_r$, $k_i = \#T_i$, 一种从某个分量 T_i 中提取（对于 SM(1) 和 SM(3)）或提取到两个分量 T_i 和 T_j（对于 SM(2)）的 SM 变换，其代价分别为 $2 - 1/k_i$、$2 - 1/k_i - 1/k_j$ 和 $1 - 2/k_i$，对应 SM(1)、SM(2) 和 SM(3)。

11.1.2. 最小产率成本。计算最小产率成本时，我们采用与 [22]和 [26] 相同的方法，并评估不同合并操作（外部合并、内部合并、SM(1)、SM(2)、SM(3)）对工作空间中连通分量的数量 $b_0(F)$ 以及可访问项的数量 $\alpha(F)$ 的影响，计算总数 $\sigma(F) = b_0(F) + \alpha(F)$ 的变化。这里我们只考虑余积 $\Delta = \Delta^d$ 的情况，因为我们已经论证，这足以研究合并作为动力系统的性质。我们分别统计每种合并操作的

结果工作空间与初始工作空间的分量数量之差。对于外部合并，该差值为 -1（它会减少连通分量的数量）；对于内部合并和 SM(1)，差值为 0；对于 SM(2) 和 SM(3)，差值为 $+1$，因为它们会使分量增加一个。我们还统计可访问项数量的变化：外部合并使其增加 2，内部合并和 SM(1)使其保持不变，而 SM(2) 和 SM(3)会使其减少，变化量为 -2（每个商 $T/^dT_v$ 都需要移除一个顶点，才能得到一棵满二叉树）。因此，分量数量与可访问项数量之和的变化总计给出了最小产出计数：

	外部合并	内部合并	SM(1)	SM(2)	SM(3)
$\mathfrak{c}_{MY}$	1	0	0	-1	-1

这种计数与 SM 是否为最小形式（仅提取并合并原子分量）无关。

11.1.3. 复杂度损失代价。复杂度损失是另一种可能的代价度量，文献 [22] 对此有所讨论。它体现了语言学模型的一项性质。在语言学文献中，这有时会与“扩展条件”归为一类，但最好将二者分开处理，因为正如 [26]，所论证的，扩展条件实际上是模型的一项“代数”约束（硬约束），而这里讨论的是一项最优性条件（软约束）。我们可以将其表述为如下要求：“句法组合总会使结构增长或扩展”。更准确地说，如果在变换后的工作空间中，经过合并操作的作用，原工作空间中的每个分量最终都成为一个分量的一部分，且该分量的复杂度至少与原分量相同，那么就不存在复杂度损失。衡量工作空间中分量复杂度的一种简单方法是看其大小，即该分量在工作空间霍普夫代数中的次数（叶子的数量）。因此，复杂度损失以工作空间各分量次数下降的最大值来衡量。于是，对于形式为 $\mathfrak{M}(T, T')$ 的 EM，分量 T、T' 最终成为次数最大的分量 $\mathfrak{M}(T, T')$ 的一部分（其他任何分量均保持不变），因此不存在损失。对于形式为 $\mathfrak{M}(T_v, T/T_v)$ 的 IM，原分量 T 与所得分量 $\mathfrak{M}(T_v, T/T_v)$ 的次数相同，且其他分量保持不变，因此也不存在复杂度损失。对于形式为 $\mathfrak{M}(T', T_v)$ 的 SM(1)，分量 T' 最终进入一个更大的分量（因此不会造成复杂度损失），但提取出 T_v 的分量 T 会变为次数更低的分量 T/T_v，复杂度损失为 $\ell(T_v)$。对于 SM(2)，有两个分量 T 和 T' 分别变为分量 T/T_v 和 T'/T'_w，损失分别为 $\ell(T_v)$、$+\ell(T'_w)$；对于 SM(3)，分量 T 变为 $T/(T_v \sqcup T_w)$，损失为 $\ell(T_v) + \ell(T_w)$。因此，复杂度损失代价函数定义如下：

	外部合并	内部合并	SM(1)	SM(2)	SM(3)
$\mathfrak{c}_{CL}$	0	0	$\ell(T_v)$	$\ell(T_v) + \ell(T'_w)$	$\ell(T_v) + \ell(T_w)$

对于仅涉及微型　　型 SM 操作且操作对象仅为原子（单叶）元素的情况，这给出

	外部合并	内部合并	微型 SM(1)	微型 SM(2)	微型 SM(3)
$\mathfrak{c}_{CL}$	0	0	1	2	2

11.1.4. 组合代价。因此，按照这些不同的代价函数计算合并操作的代价，可总结如下。对于作用于森林中分量 T_i，或 T_i和 T_j的合并操作（该森林有 $n = k_1 + \cdots + k_r$ 个叶子），权重 $\mathfrak{c}_{\mathrm{tot}}$ 为：

	MS	MY	CL 总权重 $\mathfrak{c}_{\mathrm{tot}}$	
EM	0	1	0	1
IM	0	0	0	0
SM(1)	$2-1/k_i$	0	1	$3-1/k_i$
SM(2)	$2-1/k_i-1/k_j$	-1	2	$3-1/k_i-1/k_j$
SM(3)	$1-2/k_i$	-1	2	$2-2/k_i$

根据最小搜索、最小产出和复杂度损失的总成本函数计数方式，我们来考虑合并动力学的加权版本。我们将此前讨论的无权合并动力学视为依赖于变形参数 $0<t\leq 1$的变形动力学在 $t=1$ 处的取值；在 $t\to 0$ 时，其首阶项由动力学中代价最低的部分给出。实现这一点的方法是，使用代价函数 $t^{\mathfrak{c}_{\mathrm{tot}}}$对合并图邻接矩阵的每个分量进行加权。下一节将分析这种变形对所得动力系统性质的影响。

12. Tropical Perron-Frobenius and eigenvector estimates

我们在定理 9.1 中已经看到，合并霍普夫代数马尔可夫链的平稳分布由下式给出

$$\pi(F)=\frac{\eta_{\mathcal{P}'(n)}(\wp)\ \psi_{\mathcal{P}'(n)}(\wp)}{\Lambda_{\wp,n}}\qquad \text{for } \wp=p(F)$$

其中，$\eta_{\mathcal{P}'(n)}$ 和 $\psi_{\mathcal{P}'(n)}$ 分别是图 $\mathcal{G}_{\mathcal{P}'(n)}$的邻接矩阵的左、右佩龙–弗罗贝尼乌斯特征向量。在考虑按代价函数加权的情形下，加权合并霍普夫代数马尔可夫链动力学也有相同的结果。具体而言，有如下命题。

命题 12.1。 考虑有向合并图 $\mathcal{G}_{n,A}$，其中箭头按照 *§11.1.4* 中计算的总成本 $\mathfrak{c}_{\mathrm{tot}}$（结合 *MS*、*MY*、*CL*）赋予权重 $t^{\mathfrak{c}_{\mathrm{tot}}}$，并设 $t\in\mathbb{R}^*_+$ 为参数。我们用 $\mathcal{K}^{A,n,w}(t)$ 表示由这些权重构成的非负矩阵，

$$\mathcal{K}^{A,n,w}(t)_{F,F'}=t^{\mathfrak{c}_{\mathrm{tot},F,F'}}\ \mathcal{K}^{A,n}_{F,F'} \tag{12.1}$$

将其视为参数 t 的函数。再次考虑投影映射 $p:\mathcal{G}_{n,A}\to\mathcal{G}_{\mathcal{P}'(n)}$；在图 $\mathcal{G}_{\mathcal{P}'(n)}$ 中，我们将 *EM* 和 *SM* 箭头按 $t^{\mathfrak{c}_{\mathrm{tot}}}$ 加权，并将 *IM* 回路按 $t^{\mathfrak{c}_{\mathrm{tot}}}=1$ 加权。我们用 $\mathcal{K}^{\mathcal{P}'(n),R,w}(t)$ 表示由此得到的 $\mathcal{G}_{\mathcal{P}'(n)}$ 的加权邻接矩阵。那么，由该加权 $\mathcal{K}^{A,n,w}(t)$ 确定的加权合并霍普夫代数马尔可夫链 $\hat{\mathcal{K}}^{A,n,w}(t)$ 的平稳分布为

$$\pi^w(F,t)=\frac{\eta_{\mathcal{P}'(n),w}(\wp,t)\ \psi_{\mathcal{P}'(n),w}(\wp,t)}{\Lambda_{\wp,n}}\qquad \text{for } \wp=p(F)\,, \tag{12.2}$$

其中 $\eta_{\mathcal{P}'(n),w}$ 和 $\psi_{\mathcal{P}'(n),w}$ 分别是 $\mathcal{K}^{\mathcal{P}'(n),R,w}(t)$ 的左、右 *Perron–Frobenius* 特征函数。

证明。 论证与命题 9.2 相同，但需考虑边上的权重。$\mathcal{G}_{n,A}$ 中的所有 IM 边都具有 $\mathfrak{c}_{\mathrm{tot}}=0$，因此投影 p 的纤维中的所有边都不带任何权重。顶点处的纤维重数中具有相同的数值因子 $\Lambda_{\wp,n}=\#V_{n,A,\wp}$。□

请注意，形式 (12.1) 与我们在 §3.4 中讨论的加权邻接矩阵类似，只需将 $t=e^{-\beta}$ 设为玻尔兹曼分布的热力学参数即可。

与无权情形一样，平稳分布的形式 (12.2) 将问题归结为计算加权图 $\mathcal{G}_{\mathcal{P}'(n)}$ 的矩阵 $\hat{\ }\mathcal{K}^{\mathcal{P}'(n)\,R,w}(t)$ 的左右佩龙–弗罗贝尼乌斯特征函数。

尽管图 $\mathcal{G}_{\mathcal{P}'\,(n)}$ 比合并图 $\mathcal{G}_{n,\ A}$小得多，但要像在 [26] 中的简单情形 $n=3$ 那样，求出佩龙-弗罗贝尼乌斯特征值和特征向量关于参数 t 的闭式表达式，计算上仍然很困难。这意味着，在任意 $n\geq 4$ 的一般情形下，需要估计佩龙-弗罗贝尼乌斯特征值和左右特征向量在t中的主阶行为。

对于依赖于实参数的非负矩阵，有一些通用方法（例如参见 [1]）可用于求其佩龙-弗罗贝尼乌斯特征值和特征向量的主阶；这些方法基于在一个半环中求解相应的佩龙-弗罗贝尼乌斯问题，而该半环提供了恰当的“数量级代数”。

我们采用 [1], 中针对本问题调整后的框架。具体而言， [1] 中的大实参数 $p>0$ 在这里将是参数 $-\log t$，其中 $0<t<1$。我们的参数 t经过设计，使得当 $t=1$ 时，能够回到前几节讨论的无权情形；而当 $t\to 0$时，则给出动力学的极限形式，这正是本节要讨论的内容，如同 [26] 中分析的 $n=3$ 示例。因此，我们的 $t\to 0$ 极限确实对应于[1]的 $p=-\log t\to\infty$ 极限。 [1] 中记作 $\mathcal{A}_p$ 的矩阵将是依赖于 t 的矩阵 $\mathcal{K}^{\mathcal{P}'\,(n)\ \ R,\ w}(t)$，按构造，它具有以下形式。

引理 12.2。 矩阵 $\mathcal{K}^{\mathcal{P}'(n)\,R,\ w}(t)$ 的元素具有如下形式

$$\mathcal{K}^{\mathcal{P}'(n),R,w}_{\wp,\wp'}(t)=\mathcal{K}^{\mathcal{P}'(n),R}_{\wp,\wp'}\,t^{\mathfrak{c}_{\wp,\wp'}}\,, \tag{12.3}$$

其中 $\mathcal{K}^{\mathcal{P}'}_{\wp\ \wp'}{}^{(\quad n\ \ R\ ),}$ 是未加权矩阵 $\mathcal{K}^{\mathcal{P}'(n)R}$ 的元素，并且

$$\mathfrak{c}_{\wp,\wp'}:=\mathfrak{c}_{\mathrm{tot}}(\upsilon_{\wp\wp'})\,,$$

其中标签 $\upsilon_{\wp,\ \wp'}$ 表示相应元素 $\mathcal{K}^{\mathcal{P}'}_{\wp\ \wp'}{}^{(\quad n\ \ R\ ),}$ 对应的是外部合并、*SM(1)*、*SM(2)* 还是 *SM(3)* 箭头，而 $\mathfrak{c}_{\mathrm{tot}}(\upsilon_{\wp\wp'})$ 则是根据 *§11.1.4* 的表格给出的相应代价。

12.1. **热带半环与 Perron-Frobenius 问题**。 我们希望在 t 中提取 Perron–Frobenius 特征值和特征向量的主导项，其中涉及 $\mathcal{K}^{\mathcal{P}'(n)\,R,\ w}(t)$，并考察 $t\to 0$ 时的极限，如 [26] 所示。为此，我们将关于 $\mathcal{K}^{\mathcal{P}'(n)\,R,\ w}(t)$ 的 Perron-Frobenius 问题转化为最小加半环 $S=(\mathbb{R}\sqcup\{\infty\},\ \min\ +,\ \infty,\ 0)$ 上相应的 Perron-Frobenius 问题，以计算 $t\to 0$ 所示的数量级。

引理 12.3。 对于 $\mathcal{K}^{\mathcal{P}'(n)\,R,\ w}(t)$ 如(12.3)中所示，佩龙-弗罗贝尼乌斯问题

$$\sum_{\wp'}\mathcal{K}^{\mathcal{P}'(n),R,w}_{\wp,\wp'}(t)\,\eta_{\wp'}(t)=\lambda_t\,\eta_{\wp}(t) \tag{12.4}$$

$$\sum_{\wp}\psi_{\wp}(t)\,\mathcal{K}^{\mathcal{P}'(n),R,w}_{\wp,\wp'}(t)=\lambda_t\,\psi_{\wp'}(t) \tag{12.5}$$

其首阶为 $\lambda_t=\lambda\,t^{\ell}+o(t^{\ell})$、 $\eta_{\wp}(t)=\eta_{\wp}\,t^{u_{\wp}}+o(t^{u_{\wp}})$ 和 $\psi_{\wp}(t)=\psi_{\wp}\,t^{q_{\wp}}+o(t^{q_{\wp}})$，它确定了形式为……的 *min-plus* 半环中关于数量级 $u_{\wp}$、 $q_{\wp}$、 ℓ 的 *Perron-Frobenius* 问题

$$\min_{\wp'}\{\mathfrak{c}_{\wp,\wp'}+u_{\wp'}\}=\ell+u_{\wp} \tag{12.6}$$

$$\min_{\wp}\{\mathfrak{c}_{\wp,\wp'}+q_{\wp}\}=\ell+q_{\wp'}\,. \tag{12.7}$$

证明。 我们将 (12.4) 写为

$$\sum_{\wp'} \left(\mathcal{K}^{\mathcal{P}'(n),R}_{\wp,\wp'} t^{\mathfrak{c}_{\wp,\wp'}} \eta_{\wp'}\, t^{u_{\wp'}} + o(t^{u_{\wp'}+\mathfrak{c}_{\wp,\wp'}})\right) = \lambda\, \eta_{\wp}\, t^{\ell+u_{\wp}} + O(t^{\ell+u_{\wp}})\,.$$

由于矩阵和 Perron–Frobenius 数据均为正，因此不会发生抵消。左侧求和中当 $t \to 0$ 时的首项等于 $\min\{\mathfrak{c}_{\wp,\wp'} + u_{\wp'}\}$，而这必须与右侧的首项 $\ell + u_{\wp}$ 相匹配，从而得到 (12.6)。(12.5) 和 (12.7) 的情形类似。□

12.2. **最优树形图与 Perron–Frobenius 特征值**。为了明确描述极小加 Perron-Frobenius 问题 (12.6) 和 (12.7) 的 Perron–Frobenius 特征向量，我们先回顾加权有向图中最优树形图的概念。

定义 12.4。设 G 是一个有向图，其边由代价函数 $\mathfrak{c}\colon E(G) \to \mathbb{R}$ 赋权。子图 $G' \subset G$ 的权重定义为其各边权重之和，

$$\mathfrak{c}(G') = \sum_{e \in E(G')} \mathfrak{c}(e)\,.$$

对于选定的顶点 $v_r \in V(G)$，以 v_r 为汇的树形图 T 是满足以下条件的子图 $T \subset G$

（1） T 是 G的底层无向图的一棵生成树，

（2）对于每个 $v \in V(G)$，在 $E(T)$ 中，都存在一条从 v 到根顶点 v_r的有向路径。

˜以 v_r 为源点的树形图 T 满足上述条件（*1*），并且

（$2'$）对于每个 $v \in V(G)$，在 $E(T)$ 中，都存在一条从根顶点 v_r 到 v的有向路径。

˜如果树形图 T 或 T 还满足第三个性质，则称其为最优树形图：

(3) T （或 $\tilde{T}$)具有 t 所有以 v_r为汇（或为源点）的树形图中权重最小的一个 .

如果图 G 是强连通的，那么对于任意选择的顶点 v_r，都存在一个以 v_r 为汇的树形图。在这种情况下，可以通过 Chu-Liu/Edmonds/Bock 算法构造最优树形图：首先构造 G 的子图 G'，方法是在 G 的每个顶点保留所有入边中代价最小的边，然后按照一定步骤消除 G' 中的回路，从而得到 T。以 v_r 为源点的情况类似，只需选择代价最小的出边。（有关寻找最小树形图的算法讨论，请参见 [16]。）

对于采用总成本函数 $\mathfrak{c}_{\rm tot}$加权的合并动力学，在 $t \to 0$ 时，佩龙–弗罗贝尼乌斯特征值和特征空间的渐近数量级描述如下。

定理12.1. 考虑具有矩阵 $\mathcal{K}^{\mathcal{P}'(n)\,R,w}(t)$ 的 $\mathfrak{c}_{\rm tot}$ 加权合并动力学，如 (12.3) 所示，其渐近行为由极小加 *Perron-Frobenius* 问题 (12.6) 和(12.7)的解给出。*min-plus Perron–Frobenius*特征值为 $\ell = 0$。

证明. 我们计算佩龙–弗罗贝尼乌斯特征值 ℓ 应具有的数量级，使得 Perron–Frobenius方程组 (12.6) 和 (12.7) 成为关于数量级 $u_{\wp}$ 和$q_{\wp}$的方程，其中 $\mathfrak{c}_{\wp,\wp'}$ 和 ℓ 为已知量。

有一种通用方法可以计算半环上的佩龙–弗罗贝尼乌斯特征值，例如该方法就用于 [1], [2], [10], [15]。对于一个半环 $(S, \oplus, \odot, e, \epsilon)$（特别是我们在此关注的最小加半环

），定义一个 S 值 $n\times n$ 矩阵的积和式 为

$$\operatorname{per}(K)=\bigoplus_{\sigma\in S_n}\bigodot_{i=1}^{n}K_{i\sigma(i)}\,.$$

在最小加半环的情形下，这给出

$$\operatorname{per}(K)=\min_{\sigma\in S_n}\left(K_{1,\sigma(1)}+\cdots+K_{n,\sigma(n)}\right).$$

可以据此定义特征多项式的一个类比

$$\operatorname{per}(Y\,\mathrm{Id}\oplus K)=\bigoplus_{\sigma\in S_n}\bigodot_{i=1}^{n}(Y\delta_{i,\sigma(i)}\oplus K_{i,\sigma(i)})\in S[Y]\,,$$

其系数为半环 k 阶迹

$$\operatorname{Tr}_k^S(K):=\bigoplus_{J\subset\{1,\dots,n\},\,\#J=k}\left(\bigoplus_{\sigma\in S_J}\bigodot_{j\in J}A_{j,\sigma(j)}\right)$$

其中 $\sigma\in S_J$ 是子集 J的置换。这个特征多项式的根称为 K 的“代数特征值”（参见 [2]）。请注意，在这种半环情形下，这些代数特征值一般并非全都有特征向量；其中确有特征向量的称为“几何特征值”。特别地，对于最小加半环，最小代数特征值为

$$\rho_{\min}(K)=\bigoplus_{k=1}^{n}\bigoplus_{i_1,\dots,i_k}(K_{i_1i_2}\odot\cdots\odot K_{i_ki_1})^{1/k}=\min_{k=1,\dots,n}\min_{c=(i_1,\dots,i_k)}\frac{1}{k}(K_{i_1i_2}+\cdots+K_{i_ki_1})\,.\tag{12.8}$$

其中 $c=(i_1,\ldots\ldots,i_k)$ 遍历图 $G(K)$ 中长度为 k 的所有回路；该图有 n 个顶点，并且每个非平凡项 K_{ij}对应一条边 e_{ij}。达到最小值的回路 $c=(i_1,\ldots\ldots,i_k)$ 称为“临界回路”。临界图 $G_{\mathrm{crit}}(K)\subset G(K)$由属于临界回路的顶点和边组成。如果 $G(K)$是强连通的，那么 $\rho_{\min}(K)$是唯一的几何特征值，其重数等于 $G_{\mathrm{crit}}(K)$的强连通分量数。

.

带有边 $e_{\wp、\wp'}$ 且由矩阵 $\mathcal{K}^{\mathcal{P}'(n)R}(t)$ 的条目 $\mathcal{K}^{\mathcal{P}'(\ n\ R)、\wp'}\ \ t^{\mathfrak{c}_{\wp、\wp'}}$ 赋予权重的图 $\mathcal{G}_{\mathcal{P}'(n)}$ 是强连通的，并且在所有包含至少一个大小为 $k_i\geq 3$ 的行的划分 $\wp$ 处都有 IM 自环。由于代价 $\mathfrak{c}_{\mathrm{tot}}(\mathrm{IM})=0$，$\mathcal{K}^{\mathcal{P}'(n)R}(t)$ 的这些 IM 对角条目与 t 无关，并且等于相应的条目 $\mathcal{K}^{\mathcal{P}'(\ \ n\ R)、}_{\wp\ \wp}$。

令 K 为 min-plus 值矩阵，其条目是与各边相关的合并变换的代价 $K_{\wp、\wp'}=\mathfrak{c}_{\wp、\wp'}$，如 §11.1.4 中所述。令 $G(K)$ 为与上述 min-plus 矩阵 K 相关联的图。图 $G(K)$ 与 $\mathcal{G}_{\mathcal{P}'(n)}$ 具有相同的顶点和边，其边的权重为 $\mathfrak{c}_{\wp、\wp'}$，因此它也是强连通的，因为所有 $\mathfrak{c}_{\wp、\wp'}$ 都不等于∞（min-plus 半环的加法单位元）。$\mathcal{G}_{\mathcal{P}'(n)}$ 的 IM 自环具有权重 $\mathfrak{c}_{\wp、\wp'}=0$（最大加半环的乘法单位元）。由于来自非 IM（EM 或 SM）箭头的所有其他代价都更大，$\mathfrak{c}_{\wp、\wp'}>0$，使得实现 $(\mathfrak{c}_{\wp_1,\wp_2}+\cdots+\mathfrak{c}_{\wp_k,\wp_1})/k$ 最小值的回路 $c=(\wp_1,\ldots\ldots,\wp_k)$ 是单边 IM 环 $\mathfrak{c}_{\wp、\wp}=0$（带有 $k=1$）。因此，临界回路有许多，每个都由一条边和一个顶点组成（任何至少包含一个长度至少为三的行的划分 $\wp$ 都是这样的顶点）。于是，临界图 $G_{\mathrm{crit}}(K)$ 由单边环的不交并组成，每个这样的划分 $\wp$ 对应一个单边环。

因此，我们在 (12.6) 和 (12.7) 中得到 $\ell=0$，因为这个 ℓ 对应于唯一的几何特征值 (12.8)。□

12.3. **热带 Perron-Frobenius 特征空间**。接下来，我们分析带有 $\ell = 0$ 的热带佩龙–弗罗贝尼乌斯问题 (12.6) 和 (12.7) 的佩龙–弗罗贝尼乌斯特征空间。

由于存在多个临界回路，我们无法使用 [1], [2] 的特征值渐近性；该渐近性依赖于临界回路的唯一性，从而将 min-plus（或 max-×）Perron-Frobenius 特征向量与原问题（12.4）、（12.5）的 Perron-Frobenius 特征向量联系起来。不过，我们可以像 [3] 中那样，用其生成元来描述 Perron-Frobenius 特征空间。

定理 *12.2.* 考虑动力学 $\mathcal{K}^{\mathcal{P}'(n)}{}_R$、$^w(t)$ 的代价函数 $\mathfrak{c}_{\mathrm{tot}}$，如（*12.3*）所示，其渐近行为由极小加 *Perron-Frobenius* 问题（*12.6*）和（*12.7*）决定。右 *Perron-Frobenius* 特征空间由一族 $\{u^{(\wp)}\}$ *min-plus* 特征向量生成，如（*12.6*）所示，其中 $\wp$ 遍历这样的划分 $\wp \in \mathcal{P}'(n)$：其杨图至少有一行的长度不小于 *3*。基特征向量 $u^{(\wp)} = (u^{(\wp)}_{\wp'})_{\wp' \in \mathcal{P}'(n)}$ 的元素 $u^{(\wp)}_{\wp'}$ 等于最优树形图 $T_\wp$ 中从 $\mathcal{G}_{\mathcal{P}'(n)}$ 到汇 $\wp$ 的有向路径的代价 $\mathfrak{c}(\gamma_{\wp'\wp})$。类似地，左 *Perron-Frobenius* 特征空间由一族 $\{q^{(\wp)}\}$（*12.7*）的解生成，其中 $\wp$ 遍历同一组划分 $\wp$，且这些划分至少有一行的长度不小于 *3*；对于 $\wp' \in \mathcal{P}'(n)$，元素 $q^{(\wp)}_{\wp'}$ 等于最优树形图 $\tilde{T}_\wp$ 中从源点 $\wp$ 出发的有向路径 $\tilde{\gamma}_{\wp\wp'}$ 的代价 $\mathfrak{c}(\tilde{\gamma}_{\wp\wp'})$，该最优树形图位于 $\mathcal{G}_{\mathcal{P}'(n)}$ 中。

证明。对于 min-plus 值矩阵 K，Kleene 星 $K^\star$ 定义为

$$K^\star := \mathcal{I} \oplus K \oplus K^2 \oplus \cdots \oplus K^{n-1}, \tag{12.9}$$

其中，$\mathcal{I}$ 是单位矩阵（在 min-plus 意义下）。各元素为

$$(K^\star)_{ij} = K_{i=i_0,i_1} \odot \cdots \odot K_{i_{k-1},i_k=j}$$

在 $G(K)$ 中从 i 到 j 的最小权重有向路径上，该路径的长度为某个 $1 \le k \le n-1$。如 [3] 中的命题 2.6 所示（针对我们的情形，即 $\rho_{\min}(K) = 0$，且每个临界环 C_j 都由单个顶点 j 和一条自环边组成），对于每个临界环 C_j，K都有一个特征向量 u_j，它与矩阵 $K^\star$的第 j 列成比例。这些特征向量是 $\rho_{\min}(K)$ 的特征空间的生成元。

在我们的情形中，这意味着，对于每个至少有一行长度为 $k_i \ge 3$ 的划分 $\wp$，都有一个特征向量 $u^{(\wp)} = (u^{(\wp)}_{\wp'})_{\wp' \in \mathcal{P}'(n)}$，它满足带有 $\ell = 0$的（12.6）式。相应地，将同样的论证应用于转置矩阵可知，对于每个这样的 $\wp$，都有一个特征向量$q^{(\wp)} = (q^{(\wp)}_{\wp'})_{\wp' \in \mathcal{P}'(n)}$，它满足带有 $\ell = 0$的（12.7）式。这些特征向量的集合 $\{u^{(\wp)}\}$（分别为 $\{q^{(\wp)}\}$）张成该特征空间。

此外，在我们的情形下，可以用图 $\mathcal{G}_{\mathcal{P}'(n)}$，更明确地从几何角度描述 Kleene 星矩阵的各列所对应的生成元。为此，我们使用加权有向图中最优树形图的概念，相关内容已在上文定义 12.4 中回顾。

设 $\wp$ 是 $\mathcal{P}'(n)$ 中满足以下条件的一个划分：至少有一行包含 $k_i \ge 3$，因此它们是 $\mathcal{G}_{\mathcal{P}'(n)}$ 中非平凡的 IM 自环边。再设 $T_\wp$ 表示一个以 $\wp$ 为汇、相对于代价函数 $\mathfrak{c}_{\mathrm{tot}}$ 的最优树形图。这是 $\mathcal{G}_{\mathcal{P}'(n)}$ 的一棵有向生成树，其中每个顶点 $\wp' \neq \wp$ 在 $T_\wp$ 中都有一条出边，且该出边在 $\wp'$ 处所有出边中代价最小；而顶点 $\wp$ 在 $T_\wp$ 中没有出边。

在构造 Kleene 星矩阵时，我们有

$$(K^\star)_{\wp',\wp} = \mathfrak{c}(e_{\wp',\wp_1}) + \cdots + \mathfrak{c}(e_{\wp_{k-1},\wp})$$

其中，$e_{\wp',\,\wp_1} \cdots e_{\wp_{k-1},\,\wp}$ 是从 $\wp'$ 到 $\wp$ 的一条最小代价路径。特别地，这意味着（贝尔曼最优性原理）：对于任意 $0 \le i < j \le k$，子路径 $e_{\wp_i,\,\wp_{i+1}} \cdots e_{\wp_{j-1},\,\wp_j}$ 也是从 $\wp_i$ 到 $\wp_j$ 的代价最小路径，因此，特别是对于每个顶点 $\wp_i$，路径中的下一条出边可以取自

路径可以取为 $T_{\wp}$ 中的一条边，并且我们可以将（$K^{\star}$）$_{\wp',\ \wp}$ 识别为 $\mathfrak{c}$代价，即 $T_{\wp}$ 中从 $\wp'$ 到 $\wp$的唯一一条路径$\gamma_{\wp',\ \wp}$ 的代价。因此，满足（12.6）且张成该特征空间的特征向量 $u^{(\wp)}$ 与最优树形图 $T_{\wp}$ 相关，其分量为 $u^{(\wp)} = (u^{(\wp)}_{\wp'})_{\wp' \in \mathcal{P}'(n)}$，由树形图中各条路径的代价 $u^{(\wp)}_{\wp'} = \mathfrak{c}(\gamma_{\wp',\ \wp})$ 给出。

对于 (12.7) 极小加 Perron-Frobenius 问题的特征值 $q^{(\wp)}$，我们考虑相同的设定，但使用矩阵的转置 K^{τ} K。因此，向量 $q^{(\wp)}$是 Kleene 星矩阵 $(K^{\tau})^{\star}$ 的转置 K^{τ}的 $\wp$列。该 Kleene 星矩阵的元素为

$$(K^{\tau})^{\star}_{\wp',\wp} = \mathfrak{c}(e_{\wp,\wp_1}) + \cdots + \mathfrak{c}(e_{\wp_{k-1},\wp'}) = \mathfrak{c}(\tilde{\gamma}_{\wp,\wp'}),$$

其中 $\tilde{\gamma}_{\wp,\ \wp'} = e_{\wp,\ \wp_1} \cdots e_{\wp_{k-1},\ \wp'}$是从 $\wp$到 $\wp'$的最小代价路径。因此，在这种情况下，特征向量 $q^{(\wp)}$ 对应于源点位于 $\wp$的最优树形图 $\tilde{T}_{\wp}$，其中分量$q^{(\wp)}_{\wp'} = \mathfrak{c}(\tilde{\gamma}_{\wp',\ \wp})$由树形图中路径的代价给出。□

此外，[3] 的结果表明，在特征空间中，$\mathcal{K}^{\mathcal{P}'(n)\ R,w}(t)$ 的左/右 Perron-Frobenius 特征向量的渐近行为分别为 $u = (u_{\wp})_{\wp \in \mathcal{P}'(n)}$ 和 $q = (q_{\wp})_{\wp \in \mathcal{P}'(n)}$，尽管它们不一定等于上述特征空间生成元中的某一个。事实上，[18] 中计算出的具体例子表明，一般而言，Perron-Frobenius 特征向量的渐近行为并非极小加 Perron-Frobenius 问题的特征空间生成元之一。

12.4. **关于成本函数的优化**。在我们此处所考虑的 Merge Hopf 代数马尔可夫链动力系统这一框架下，[7], [8], [22], 所阐述的极简主义语言学模型中相对于最小搜索、最小产出和复杂度损失的最优性概念，成为马尔可夫链的最小化性质

$$\hat{\mathcal{K}}^{(A,n,(\mathfrak{c}_{\mathrm{tot}},\beta))} = \hat{\mathcal{K}}^{(A,n,\mathfrak{c}_{\mathrm{tot}})}(t) \quad \text{for } t = e^{-\beta},$$

with $\mathfrak{c}_{\mathrm{tot}}$ §11.1.4 的代价函数，相对于自由能函数

$$\mathbb{F}(\hat{\mathcal{K}}^{(A,n,(\mathfrak{c}_{\mathrm{tot}},\beta))}) = \bar{\mathbb{E}}_{\mathbb{P}_{(\mathfrak{c}_{\mathrm{tot}},\beta)}} - \beta^{-1}\mathrm{Sh}(\mathbb{P}_{(\mathfrak{c}_{\mathrm{tot}},\beta)})$$

其中 $\mathbb{P}_{(\mathfrak{c}_{\mathrm{tot}},\beta)}$ 是玻尔兹曼分布，其中 $\mathbb{E} = \mathfrak{c}_{\mathrm{tot}}$，如 §3.4 所述。

我们将在 §12.5 讨论，使用代价函数 $\mathfrak{c}_{\mathrm{tot}}$ 在多大程度上实现了消除侧向合并箭头的不良特性这一目标，并使系统能够通过外部合并收敛到完全形成的句子（树结构）。随后，我们将在 §13 提出另一种简单代价函数的选择，以修正剩余问题，并完全用信息论术语来表述优化。

12.5. **加权动力学中的侧向合并问题**。我们对加权动力学的讨论涉及代价函数 $\mathfrak{c}_{\mathrm{tot}}$ 以及合并图邻接矩阵上的权重 $t^{\mathfrak{c}_{\mathrm{tot}}}$ $\mathcal{G}_{n,\ A}$ 讨论结果表明，该代价函数修正了非加权动力学中 SM 箭头的一些不良特性。然而，与 [26] 中讨论的简单情形$n = 3$ 相比，令人意外的是，仅靠代价计数 $\mathfrak{c}_{\mathrm{tot}}$ 并不足以彻底纠正 SM 将动力学拉向连通性较低结构的趋势。

更准确地说，我们发现，当 $t \to 0$ 时，在主导阶上，加权动力学的平稳分布位于上述佩龙–弗罗贝尼乌斯特征空间中。由于该特征空间的基只涉及至少有一个 $k_i \geq 3$ 的划分 $\wp$，我们不再看到非加权动力学中典型的现象：由于 SM 占主导，平稳分布会将最高概率赋予最不连通的结构 $\wp_{1^n}$。然而，

现在，平稳分布的主导阶可以包含所有至少有一个 $k_i \geq 3$ 的 $\wp$ 的贡献（即所有具有 IM 动力学的划分），而不再仅仅支撑于 $\wp = \{n\}$（在 IM 动力学下收敛到连通结构，即由单棵树构成的工作空间）；这正如 EM 在动力学中压过 SM–EM 张力时所预期的，也如在 [26] 具有相同权重和代价函数的 $n = 3$ 情形中所发生的那样。

这表明，语言学文献中通常讨论的最小搜索、最小产出和复杂度损失（ [7], [8] 稍有不同的 MY 表述并无影响）并不足以作为代价函数来消除侧向合并的不良影响。

正如我们将在下一节讨论的，缺失的是信息优化的作用；我们已经在 §3，尤其是 §3.3 中以另一种形式指出，合并霍普夫代数马尔可夫链满足这一点。我们在此证明，在代价函数中加入香农信息计数，便能使平稳分布的主导阶部分中的动力学集中于仅由一棵树构成、且遵循 IM 动力学的工作空间。

13. Shannon entropy optimization and the Merge dynamics

定义 13.1. 对于形如 $n = a_1 k_1 + \cdots + a_r k_r$ 的划分 $\wp \in \mathcal{P}'(n)$，我们赋予其概率分布

$$\mathbb{P}_\wp := (\underbrace{\frac{k_1}{n}, \dots, \frac{k_1}{n}}_{a_1\text{-}times}, \dots, \underbrace{\frac{k_r}{n}, \dots, \frac{k_r}{n}}_{a_r\text{-}times}) . \tag{13.1}$$

我们扩展上一节中考虑的代价函数 $\mathfrak{c}_{\rm tot}$，加入一个取决于分布（13.1）的香农熵 $\mathrm{Sh}(\mathbb{P}_\wp)$ 的项。

定义 13.2。关于 图 $\mathcal{G}_{\mathcal{P}'(n)}$，我们考虑如下形式的代价函数 $\mathfrak{c}$： $E(\mathcal{G}_{\mathcal{P}'(n)}) \to \mathbb{R}$

$$\tilde{\mathfrak{c}}_{\rm tot}(e) = \mathfrak{c}_{\rm tot}(e) + \mathrm{Sh}(\mathbb{P}_{s(e)}) , \tag{13.2}$$

其中， $\mathfrak{c}_{\rm tot}$ 是 *MS*、*MY*、*CL* 的总成本函数，如 *§11.1.4* 所述； $\mathbb{P}_{s(e)}$ 是在划分 $\wp = s(e) \in \mathcal{P}'(n)$处的分布 (13.1)，也是有向图 $\mathcal{G}_{\mathcal{P}'(n)}$中边 e 的源顶点。

将定理12.1和定理 12.2 中使用的相同论证应用于修正后的代价函数（13.2），可得到以下结果。

命题 13.3。 对于 (13.2) 的修正后的代价函数 $\tilde{\mathfrak{c}}_{\rm tot}$，动力学 $\mathcal{K}^{\mathcal{P}'(n)\, R,\, w}(t)$ 如 (12.3) 所示，其渐近行为由极小加 *Perron-Frobenius* 问题 (12.6)和(12.7) 的解给出，其中 $\ell = 0$，且 *PF* 左/右特征向量为 $\upsilon = \upsilon^{(\wp=\{n\})}$，其中 $\upsilon_{\wp'} = \tilde{\mathfrak{c}}(\gamma_{\wp'\{n\}})$ 是以 $\wp = \{n\}$ 为汇的最优树形图 $T_{\{n\}}$ 中有向路径的代价；而 $q = q^{(\wp=\{n\})}$、$\tilde{}$ 满足 $q_{\wp'} = \tilde{\mathfrak{c}}(\tilde{\gamma}_{\{n\}\wp'})$，是以$\wp = \{n\}$为源点的最优树形图 $T_{\{n\}}$ 中有向路径的代价。

证明。与定理12.1和定理 12.2相比，主要区别在于，使用修正后的代价函数 (13.2) 时，只有一个临界回路，即划分 $\wp = \{n\}$ 处的 IM 自环。事实上，代价函数中新增的项 $\mathrm{Sh}(\mathbb{P}_{s(e)})$ 在 $\mathbb{P}_\wp = 0$ iff $\wp = \{n\}$ 处取最小值，在所有其他 $\wp \in \mathcal{P}'(n)$ 处都严格为正。因此，唯一的最小代价回路（其 $\tilde{\mathfrak{c}}_{\rm tot}(\gamma)$ 为 $= 0$ ）是源点和目标点均为 $\wp = \{n\}$ 的单条 IM 边。这再次得到极小加 PF 特征值的 $\ell = 0$，与定理12.1中的情况相同。由于此处确实只有一个临界回路，我们不仅可以得到极小加 Perron-Frobenius 问题的 PF 特征空间的一组基，还可以直接得到 PF 特征向量，它给出了原 Perron-Frobenius 问题的渐近行为。具体而言，这给出了极小加问题的右/左特征向量的形式 $\upsilon_{\wp'} = \tilde{\mathfrak{c}}(\gamma_{\wp'\{n\}})$ 和 $q_{\wp'} = \tilde{\mathfrak{c}}(\tilde{\gamma}_{\{n\}\wp'})$。□

因此，我们得到加权合并动力学的平稳分布的渐近行为，其中 $\mathcal{G}_{n,\ A}$。

定理 13.1。 在工作空间上作用的合并霍普夫代数马尔可夫链中，带有权重 $t^{\tilde{\mathfrak{c}}}$ 和代价函数 $\tilde{\mathfrak{c}}$（式 (13.2)）的平稳分布 $\hat{}\mathcal{K}^{A,\ n,\ w}(t)$ 的 $t \to 0$ 渐近行为，由沿着最优树形图 $T_{\{n\}}$ 和 $T_{\{n\}}$的环路总成本给出，

$$\operatorname{order}_{t\to 0}(\pi^{w}(F,t)) = \tilde{\mathfrak{c}}(\gamma_{p(F)\,\{n\}} \cup \tilde{\gamma}_{\{n\}\,p(F)})\,. \tag{13.3}$$

这意味着，当 $t \to 0$ 趋于极限时，$\hat{}\mathcal{K}^{A,\ n,\ w}(t)$ 的平稳分布是支撑在由单棵树组成的工作空间上的均匀分布，$F = T \in p^{-1}(\{n\})$。因此，由代价函数 $\tilde{\mathfrak{c}}$ 加权的合并动力学会收敛到完全形成的树，同时保留内部合并的作用。

证明。 根据命题 12.1，对于在$\mathcal{G}_{n,\ A}$上进行合并动力学的霍普夫代数马尔可夫链 $\hat{}\mathcal{K}^{A,\ n,\ w}(t)$，其加权邻接矩阵为

$$\mathcal{K}^{A,n,w}_{F,F'}(t) = t^{\tilde{\mathfrak{c}}_{\mathrm{tot},F,F'}}\ \mathcal{K}^{A,n}_{F,F'}\,,$$

具有如 (12.2) 所示的平稳分布，其形式为

$$\pi^{w}(F,t) = \frac{\eta_{\mathcal{P}'(n),w}(\wp,t)\ \psi_{\mathcal{P}'(n),w}(\wp,t)}{\Lambda_{\wp,n}}\,,$$

其中，在投影映射 $p : \mathcal{G}_{n,A} \to \mathcal{G}_{\mathcal{P}'(n)}$ 下，$\wp = p(F)$ 为 $\eta_{\mathcal{P}'(n)\,w}$，而 $\psi_{\mathcal{P}'(n)\,w}$ 分别是 $\mathcal{K}^{\mathcal{P}'}{}_{(n)\,R,}{}^{w}(t)$ 在 $\mathcal{G}_{\mathcal{P}'(n)}$ 上的左、右 Perron–Frobenius 特征函数。因此，考察 $t \to 0$ 的 $\pi^{w}(F,t)$ 的渐近行为时，由命题 13.3 可知，其形式为

$$\operatorname{order}_{t\to 0}(\pi^{w}(F,t)) = \upsilon_{p(F)} + q_{p(F)}$$

其中 $\upsilon = \upsilon^{(\wp=\{n\})}$ 和 $q = q^{(\wp=\{n\})}$ 如命题 13.3 所示。由此得到 (13.3)，其中 $\upsilon_{p(F)} + q_{p(F)}$是回路 $\gamma_{p(F)\,\{n\}} \cup \tilde{\gamma}_{\{n\}\,p(F)}$ 的 $\tilde{\mathfrak{c}}$ 代价。这些回路的代价均严格为正，除非 $p(F) = \{n\}$，此时代价为零。因此，在极限 $t \to 0$ 下，$\pi^{w}(F,t)$ 将收敛至 $p^{-1}(\{n\})$ 上的均匀分布，即由单个连通分量构成的工作空间。□

在这种形式下，加权动力学恰好产生预期的效果：它弥补了侧向合并的存在往往会使动力学趋向连通性较弱的结构，从而阻碍结构形成过程收敛到单棵树这一事实。与此同时，侧向合并的良好性质得以保留，即图的强连通性（动力学的不可约性）。

13.1. **香农熵代价函数**。事实上，虽然我们在命题 13.3 和定理 13.1 中，按照定义 13.2，结合了语言学文献中提出的最小搜索、最小产出和资源限制等通常概念所构成的代价函数 $\mathfrak{c}_{\mathrm{tot}}$（见 [7], [8]），但在我们的加权动力学中，真正发挥预期作用的是代价函数中的香农熵部分：它促使动力学收敛到由单棵树构成的工作空间，这正是句法结构形成过程所需要的。下面的结果可完全仿照定理 13.1 证明。

定理 13.2。 考虑合并动力学$\hat{}\mathcal{K}^{A,\ n,\ \mathrm{Sh}}(t)$，其权重仅由香农熵代价决定，即使用形式如下的加权邻接矩阵 $\mathcal{G}_{n,\ A}$

$$\mathcal{K}^{A,n,\mathrm{Sh}}_{F,F'}(t) := t^{\mathrm{Sh}(\mathbb{P}_{p(F)})}\ \mathcal{K}^{A,n}_{F,F'}\,. \tag{13.4}$$

那么，平稳分布的渐近行为为

$$\operatorname{order}_{t\to 0}(\pi^{\mathrm{Sh}}(F,t)) = \sum_{e\in\gamma_{p(F)}\{n\}} \mathrm{Sh}(\mathbb{P}_{s(e)}) + \sum_{e'\in\tilde{\gamma}_{\{n\}\,p(F)}} \mathrm{Sh}(\mathbb{P}_{s(e')})\,. \tag{13.5}$$

特别地，对于所有 $F\in p^{-1}(\wp)$ 且满足 $\wp\neq\{n\}$ 的情形，阶$_{t\to 0}(\pi^{\mathrm{Sh}}(F,t))>0$，而对于 $F=T\in p^{-1}(\{n\})$，阶$_{t\to 0}(\pi^{\mathrm{Sh}}(F,t))=0$。因此，在极限 $t\to 0$ 下，平稳分布收敛到树 $T\in p^{-1}(\{n\})$ 上的均匀分布，此时只有内部合并在起作用。

注 13.4. 定理 13.2 的结果表明，最小搜索和资源限制成本函数在简化图 $\mathcal{G}_{n,A}$ 方面发挥作用：它们通过识别足以实现强连通性的“最小”侧向合并箭头，显著缩小图的规模，使其变得稀疏。随后，对信息论数据的优化推动动力学朝连通工作空间演进，从而弥补侧向合并的存在。

这一观察引出了一个自然的问题：在投影映射 $p:\mathcal{G}_{n,A}\to\mathcal{G}_{\mathcal{P}'(n)}$ 的几何背景下，所添加的代价函数$\mathfrak{c}_{\mathrm{Sh}}(e)=\mathrm{Sh}(\mathbb{P}_{s(e)})$ 有何含义？

我们在 §3.3 中看到，合并霍普夫代数马尔可夫链 $\hat{}_{\mathcal{K}^{A,n}}$ 具有香农熵最大化性质：该香农熵是由边上的概率所确定的路径上的概率的香农熵，而边上的概率则由马尔可夫链的转移概率定义。我们还在 §3.4 中看到，在边上引入由代价函数赋予的权重后，可以将这一优化问题重新表述为自由能泛函 $\bar{\mathbb{F}}=\bar{\mathbb{E}}-\beta^{-1}\mathrm{Sh}$ 的最小化问题，其中能量 $\mathbb{E}$ 由代价函数决定，而边上的概率分布则由（加权）马尔可夫链的转移概率决定。

这里我们考虑一个附加数据：与每个顶点 $\wp\in\mathcal{P}'(n)$ 相关联的概率分布 $\mathbb{P}_\wp$。量 $\mathrm{Sh}(\mathbb{P}_\wp)$ 的含义可描述如下。

注 13.5。 根据定义，$\mathrm{Sh}(\mathbb{P}_\wp)$度量了观察一个按照 $\mathbb{P}_\wp$分布的事件结果时的预期信息量（或预期惊奇度）。如果我们固定工作空间 $F=T_1\sqcup\cdots\sqcup T_r$ in $p^{-1}(\wp)\subset V(\mathcal{G}_{n,A})$，就可以观察叶子处的词汇材料（集合 A的元素）如何分布在工作空间 F的分量 T_i 之间。由于这里处理的仅仅是自由结构形成（自由对称合并），因此对于 A 的元素如何分布在 T_i的叶子之间没有任何约束，这意味着支配可能的分配的唯一规律就是概率$\mathbb{P}_\wp$。（我们将在§15和§16中看到，对θ角色分配、阶段结构和句法参数进行过滤，将施加约束并改变这一概率分布。）

注 13.6。 概率分布 $\mathbb{P}_\wp$ 与香农熵 $\mathrm{Sh}(\mathbb{P}_\wp)$直接关联于将 A 的元素分配给工作空间各分量的组合计数，这通过$\mathrm{Sh}(\mathbb{P}_\wp)$与(2.4)、(2.6)中的多项式系数 $\mu_{\wp,\,n}$ 之间的众所周知的关系体现出来，即渐近行为

$$\mu_{n,\wp}\sim e^{n\,\mathrm{Sh}(\mathbb{P}_\wp)+o(n)}\,. \tag{13.6}$$

这可以直接由阶乘的斯特林近似推导出来，它还表明低阶项的表现类似于 $o(n)=O(\log(n))$。式 (13.6) 中的行为可以从以下事实来理解：$\mu_{n,\,\wp}$ 统计的是长度为 n 的词，这些词由含有 r 个符号的字母表构成，其中第 i 个符号出现了 k_i 次。

注 13.7。 在计算投影映射 $p:\mathcal{G}_{n,A}\to\mathcal{G}_{\mathcal{P}'(n)}$ 的重数时，也就是计算纤维的大小 $\#p^{-1}(\wp)=\Lambda_{\wp,n}$ 我们有 $\Lambda_{\wp,\,n}=\mu_{n,\wp}\cdot\Gamma_{\wp,n}$ 其中 $\Gamma_{\wp,n}$ 如 (2.11) 所示，因子$\Gamma_{\wp,n}$ 用于考虑划分 $\wp$ 中长度相同的行的排列，以及

计算具有 k_i 个叶子的树 T_i 的所有不同可能树拓扑。不过，与命题 5.3 中的计数不同，当我们固定了 $p^{-1}(\wp)$ 中的一个工作空间 $F = T_1 \sqcup \cdots \sqcup T_r$，并考察词汇材料 A 在 T_i之间的分布时，只有重数 $\Lambda_{\wp,n}$ 中的 $\mu_{n,\wp}$ 部分仍然相关，因为因子$\Gamma_{\wp,n}$ 所编码的其他选择都已固定。

各种合并变换对划分的影响，相应地也会影响概率分布 $\mathbb{P}_\wp$ 及其香农熵；这种影响可以用熵泛函的自然性质来表述。例如，我们有如下观察。

命题 13.8. 外部合并对划分的作用如下：

$$\wp = \{k_1, \dots, k_r\} \mapsto \wp' = \{k_1, \dots, k_i + k_j, \dots, k_r\}$$

对 $\mathrm{Sh}(\mathbb{P}_\wp)$ 的作用，遵循香农熵的链式法则：

$$\mathrm{EM} : \mathrm{Sh}(\mathbb{P}_\wp) \mapsto \mathrm{Sh}(\mathbb{P}_{\wp'}) = \mathrm{Sh}(\mathbb{P}_\wp) - \frac{k_i + k_j}{n}\mathrm{Sh}(\frac{k_i}{k_i + k_j}, \frac{k_j}{k_i + k_j}) \tag{13.7}$$

证明。 根据香农熵的对称性，不失一般性，我们可以假设这对 k_i, k_j 是 k_1, k_2。香农熵的链式法则（也称广延性条件）给出：

$$\mathrm{Sh}(\frac{k_1}{n}, \dots, \frac{k_r}{n}) = \mathrm{Sh}(\frac{k_1 + k_2}{n}, \frac{k_3}{n}, \dots, \frac{k_r}{n}) + \frac{k_1 + k_2}{n}\mathrm{Sh}(\frac{k_1}{k_1 + k_2}, \frac{k_2}{k_1 + k_2}) \,, \tag{13.8}$$

根据 (13.7)，这对应于一个外部合并变换对香农熵的作用：该变换将工作空间 T_i 和 T_j 两个分量合并为 $\mathfrak{M}(T_i, T_j)$。□

注 13.9。 多项式系数的恒等式

$$\binom{n}{m, k_3, \dots, k_r} \cdot \binom{m}{k_1, k_2} = \binom{n}{k_1, k_2, k_3, \dots, k_r}$$

在渐近行为中推出香农熵的链式法则恒等式（13.8）。

注 13.10。 请注意，这表明外部合并与香农熵的链式法则相关联，而链式法则正是信息公理化刻画（Khinchin 公理）中的关键性质。探究内部合并和侧向合并是否也具有信息论性质方面的含义，似乎很有意思。

当我们使用 $\mathrm{Sh}(\mathbb{P}_\wp)$ 作为能量/代价函数，并令 $\mathfrak{c}_{\mathrm{Sh}}(e) = \mathrm{Sh}(\mathbb{P}_{p(s(e))})$ 时，就将邻接矩阵 $\mathcal{K}^{A,n}$ 的 $\mathcal{G}_{n,A}$ 加权为

$$\mathcal{K}^{A,n,\mathrm{Sh}}_{F,F'} = t^{\mathrm{Sh}_{p(F)}}\, \mathcal{K}^{A,n}_{F,F'} = e^{-\beta \mathrm{Sh}_{p(F)}}\, \mathcal{K}^{A,n}_{F,F'} \,,$$

其中 $t = e^{-\beta}$，β 是玻尔兹曼分布的逆温度热力学参数，因此范围 $0 < t < 1$ 对应于逆温度的通常范围 $0 < \beta < \infty$，其中 $t \to 0$ 对应于零温极限 $\beta \to \infty$。使用这一记法，为了与 §3.4 保持一致，我们在此写作 $\mathcal{K}^{A,\ n\,\mathrm{Sh}\,\beta}_{F\ F'}$，而不是像之前那样写作 $\mathcal{K}^{A,\ n\,\mathrm{Sh}}_{F\ F'}(t)$。

命题 13.11。 在零温度的热力学极限下，$\beta \to \infty$ $\hat{\mathcal{K}}^{A,\ n\,\mathrm{Sh}}_{F\ F'}$，的自由能优化由内部合并图 $\mathcal{G}^{\mathrm{IM}}_{n,\ A,\ \{n\}}$ 的边上的均匀分布实现，该分布使香农熵最大化。

证明。如 §3.4 所示，相应的马尔可夫链 $\hat{\mathcal{K}}^{A,\ n,\ \mathrm{Sh},\ \beta}$ 使自由能函数最小化

$$\mathbb{F}(\hat{\mathcal{K}}^{A,n,\mathrm{Sh},\beta}) = \bar{\mathbb{E}}(\hat{\mathcal{K}}^{A,n,\mathrm{Sh},\beta}) - \beta^{-1}\mathrm{Sh}(\hat{\mathcal{K}}^{A,n,\mathrm{Sh},\beta}) \,.$$

为了考察 $\beta \to \infty$ 的行为，更方便的做法是考虑如 (3.19) 所示的能量函数形式的最大化，

$$\tilde{\mathbb{F}}(\mathcal{S}) = \mathrm{Sh}(\mathcal{S}) - \beta \overline{\mathbb{E}}(\mathcal{S}) .$$

如引理 3.7 所示，该自由能由式 (3.24) 给出：

$$\tilde{\mathbb{F}}(\hat{\mathcal{K}}^{A,n,\mathrm{Sh},\beta}) = -\beta \overline{\mathbb{E}}_{\mathbb{P}_{(\mathrm{Sh},\beta)}} + \mathrm{Sh}(\mathbb{P}_{(\mathrm{Sh},\beta)}) ,$$

其中

$$\mathbb{P}_{(\mathrm{Sh},\beta)}(e) = \frac{1}{\lambda} \psi(s(e)) e^{-\beta \mathrm{Sh}(\mathbb{P}_{p(s(e))})} \eta(t(e))$$

其中 λ 以及 ψ、η 分别是 $\mathcal{K}^{A,n,\mathrm{Sh},\beta}$的 PF 特征值和左右 PF 特征向量。因此，我们有 e

$$\bar{\mathbb{E}}(\hat{\mathcal{K}}^{A,n,\mathrm{Sh},\beta}) = \sum_F \pi(F) \mathrm{Sh}(\mathbb{P}_{p(F)})$$

其中 π 是$\hat{}_{\mathcal{K}^{A,n,\mathrm{Sh},\beta}}$的平稳分布。如 §3.4 中所述，我们有 $\mathbb{F}(\hat{}_{\mathcal{K}^{A,n,\mathrm{Sh},\beta}}) = -\beta^{-1}(\log(\lambda) + \mathrm{Sh}(\pi))$，如式 (3.27) 所示，其中 $\lambda = \lambda_{(\mathrm{Sh},\beta)}$ 是 $\mathcal{K}^{A,n,\mathrm{Sh},\beta}$的 PF 特征值。

定理 13.2 进一步表明，在零温极限 $\beta \to \infty$ ($t \to 0$) 下，平稳分布 $\pi(F) = \pi^{(\mathrm{Sh},\beta)}(F)$ 收敛到纤维 $p^{-1}(\{n\})$ 上的均匀分布。因此，求和中的各项

$$\sum_F \pi^{(\mathrm{Sh},\beta)}(F) \ \mathrm{Sh}(\mathbb{P}_{p(F)})$$

满足 $\pi^{(\mathrm{Sh},\beta)}(F) \to 0$ 对于 $\beta \to \infty$，对所有 $F \notin p^{-1}(\{n\})$ 均成立，其中 $\mathrm{Sh}(\mathbb{P}_{p(F)}) \neq 0$ 且 $\pi^{(\mathrm{Sh},\beta)}(F) \to 1$ 对于$F \in p^{-1}(\{n\})$，其中 $\mathrm{Sh}(\mathbb{P}_{p(F)}) = 0$，因此该和在极限下趋于零。

此外，在固定 $T = s(e)$ 且满足 $T' = t(e)$ 的情况下，$\mathcal{G}_{n,A}$ 上的分布 $\mathbb{P}_{(\mathrm{Sh},\beta)}(e)$ 化简为内部合并图 $\mathcal{G}^{\mathrm{IM}}_{n\ ,A,\{n\}}$ 的边集上的分布 $\mathbb{P}_{\mathrm{IM}}(e) = \hat{}_{\mathcal{K}^{\mathrm{IM}}_{T\ ,T'}}$。根据命题 5.5，该分布是均匀分布

$$\mathbb{P}_{\mathrm{IM}}(e) = \frac{1}{\lambda_{\mathrm{IM}}} \frac{\eta^{\mathrm{IM}}(t(e))}{\eta^{\mathrm{IM}}(s(e))} = \frac{1}{2n-4}$$

其中，对于所有 $T \in V_{n,\ A,\ \{n\}}$ 以及 $\lambda_{\mathrm{IM}} = d_{\{n\}} = 2n-4$，$\eta^{\mathrm{IM}}(T) = 1$。因此，在 $\beta \to \infty$ 极限下，最大化问题归结为对 $\mathcal{G}^{\mathrm{IM}}_{n\ ,\ A,\ \{n\}}$ 的边上的概率分布的香农熵进行最大化，并在均匀分布处达到最大值。□

14. Parameterizing cube and dynamics

在本节及接下来的几节中，我们将介绍一些进一步的研究方向，以探讨合并在工作空间上的作用的动力学性质。[22] 提出的一个问题是：合并的结构形成过程在多大程度上能够通过句法对象在某些语义空间中的映像，在句法-语义接口处加以编码。

[22] 中提出的句法-语义接口模型基于这样一种观点：语言的计算结构完全位于句法一侧（句法优先模型），而语义一侧则可以简化为一种最基本的拓扑结构，用来刻画某个背景“语义空间”中的邻近性概念（例如流形或分层空间等拓扑空间或度量空间）。由于这种语义模型本身并未赋予类似合并的计算结构形成过程，因此自然会引出这样一个问题：合并是否仍会以某种方式编码在句法对象于语义空间中的映像里，并至少能够部分地从中重构出来？这个问题对于当前在大语言模型等机器架构中模拟人类语言的研究也具有现实意义：我们可以探究是否存在某些回路

（按机制可解释性的含义），涉及 Transformer 架构中注意力模块的权重，并且对应于通过工作空间上的合并作用来计算句法。

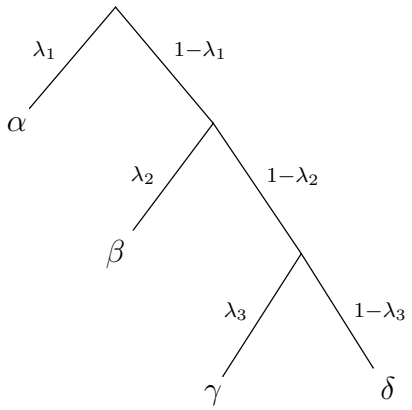
另一个相关问题是，通过合并在神经计算架构中实现结构形成过程：也就是句法中结构形成的神经科学相关性这一广泛问题。最近一项关于在函数空间（如小波）中编码合并操作的可能方式的提案 [21] 表明，在由 R´ényi 熵确定的热带半环变形中，可以实现对合并操作的良好编码。

无论是在这一设定中，还是在句法-语义接口模型中 [22], ，人们都考虑一个由带标号叶子的非平面满二叉根树构成的参数化空间，并在其中引入连续参数，以反映此类树在某个环境语义空间中的度量嵌入。在句法-语义接口模型中 [22]，人们采用非平面树的 BHV 模空间 [4]来获得这种参数化，遵循 [13]。尽管 BHV 模空间似乎足以描述句法对象在语义空间中的嵌入，以及外化过程中与树的平面化之间的相互作用（通过其与结合子多面体和零亏格实曲线模空间的关系，如 [13]中所述），但这种几何设定似乎并不理想，无法捕捉合并操作可能具有的图像。BHV 模空间具有一种有趣的相关 operad 结构，即 [12]中讨论的马赛克歌剧。然而，BHV 模空间的边界分层与产生合并作用的余乘运算之间并不存在直接而明显的关系：前者对应于将边收缩为零长度的操作，而后者对应于通过容许切割提取可访问项。

另一方面，在 [21], 中，连续参数也与句法对象相关联，具体表现为与树的非叶顶点相关联的参数 $\lambda_v \in [0,1]$ 。这些参数决定了树叶子上的概率分布，该分布是关于这些变量的多项式函数，它编码了树结构，并在变形半环中表示为求和操作的合并操作中进行优化。由于在这种情形下，合并操作被表示为对连续参数进行优化，这也引出了一个问题：描述合并对工作空间作用的离散霍普夫代数马尔可夫链，如何映射到涉及连续参数化（度量树）的情形中。

本文不讨论这个问题，但本节将对度量树的连续参数化作一些一般性考量，以此说明该问题及几种可能的处理方法。

已在 [21] 中证明，给定一个句法对象 $T \in \mathfrak{T}_{\mathcal{SO}_0}$，并赋予如定义 10.2（另见 [22], §1.13.3）中的头函数 h_T，可以通过一组参数 $\lambda = (\lambda_v)_{v \in V^o(T)}$ 为 T 的边赋予度量结构，其中每个变量 λ_v 的取值于 $[0,1],$，使得 $\lambda \in [0,1]^{n-1}$，而 $n = \#L(T)$， $V^o(T)$则表示 T的非叶顶点集合。边长的赋予方式如下：给定 $v \in V^o(T)$后，其下方的两条边分别取得长度 λ_v 和 $1-\lambda_v$，例如：



引理 14.1. 给定一个头函数，等价于选择两条边中哪一条的权重为 λ_v，哪一条的权重为 $1-\lambda_v$。

证明。 如定义 10.2 中所回顾的，头函数 $h_T: V^o(T) \to L(T)$ 为每个内部（非叶）顶点指定一个叶子，并满足以下规则：如果 $T_w \subset T_v$ 且 $h_T(v) \in L(T_w)$，那么 $h_T(v) = h_T(w)$。因此，若 $h_T(v)$ 是头函数 h_T 为 $v \in V^o(T)$ 指定的叶子，而 $\gamma_{h_T(v)}$ 是 T 中从 v 到叶子 $h_T(v)$ 的路径，则可按照以下规则赋予权重 λ_v 和 $1-\lambda_v$：对于顶点 $v \in V^o(T)$ 下方的两条边之一，我们赋予权重

$$\epsilon_{e_v} = \begin{cases} \lambda_v & e_v \in \gamma_{h_T(v)} \\ 1-\lambda_v & e_v \notin \gamma_{h_T(v)} \end{cases} \tag{14.1}$$

反过来，给定一组对 v 下方的每对边分别赋予 λ_v 和 $1-\lambda_v$ 的标记，对于每个 $v \in V^o(T)$，函数 $h_T: V^o(T) \to L(T)$ 将 v 映射到这样一片叶子：沿着从 v 出发、由标记为变量 λ_v（而非 $1-\lambda_v$）的边组成的路径前进即可到达该叶子；这样的函数就是一个头函数。□

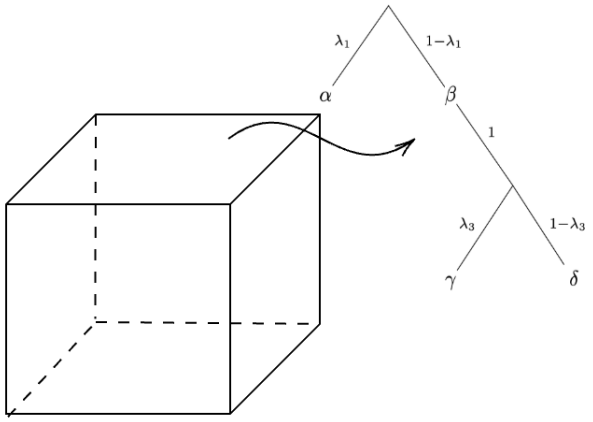
用这些连续变量对边进行标记，会在 T 的叶子上诱导出一个概率分布 $A^T = (a_\ell^T)_{\ell \in L(T)}$，如 [22], 所示，其中每个 a_ℓ^T 都是变量 λ_v 的多项式函数 $a_\ell^T = P_\ell^T(\lambda)$，其形式为

$$a_\ell^T = \prod_{v \in \gamma_{v_0,\ell}} \epsilon_{e_v}, \tag{14.2}$$

其中 $\gamma_{v_0,\ell}$ 是 T 中从根顶点 v_0 到叶子 ℓ 的路径。

概率 A 用于 [21]，在适当的函数空间中组合词汇数据 $\alpha_\ell \in \mathcal{SO}_0$ 的表示，以获得句法对象 T的表示。

现在我们考虑将点 $\lambda = (\lambda_v)$ 限制到超立方体 $[0,1]^{n-1}$ 边界中的某个面上这一操作，也就是将变量 λ_v 之一限制为 0 或 1，如下图所示。



引理 14.2. 将变量 λ_v 之一取为0 或1 ，对应于对 $T/^dT_w$ 的收缩；对于 w ，它是下方的顶点 v ，通过边 e_v 与 $\epsilon_{e_v} \to 0$ 相连。

证明。考虑概率分布 $A^T = (a_\ell^T)_{\ell \in L(T)}$ ，以及将其中一个 λ_v取为 0 或 1 所产生的影响。如果 $\lambda_v = 0$，那么边 $e_{v,\,1}$ （其 $e_{v,\,1} \in \gamma_{h_T\,(v)}$ ）的长度为 $\lambda_v = 0$ ，而下方的另一条边$e_{v,\,2}$ 在 v 处的长度为 $1 - \lambda_v = 1$。因此，概率分布 A^T 会发生如下变化

$$a_\ell^T|_{\lambda_v=0} = \prod_{w \neq v \in \gamma_{v_0,\ell}} \epsilon_{e_w}\,, \tag{14.3}$$

如果路径 $\gamma_{v_0,\,\ell}$ 包含边 $e_{v,\,2}$，并且

$$a_\ell^T|_{\lambda_v=0} = 0\,, \tag{14.4}$$

如果路径 $\gamma_{v_0,\,\ell}$ 包含边 $e_{v,\,1}$，且保持不变，

$$a_\ell^T|_{\lambda_v=0} = a_\ell\,, \tag{14.5}$$

如果 $\gamma_{v_0,\,\ell}$ 既不包含 $e_{v,\,1}$ ，也不包含 $e_{v,\,2}$。那么注意，(14.3) 与 (14.5)（以及 (14.4)，因为这些叶子已不再存在）恰好给出了商树 $T/^dT_w$ 上的结果概率分布 $A^{T/^dT_w}$。对于沿着 $e_{v,\,1}$、位于 v 下方的顶点 $w = t(e_{v,\,1})$，如示例 □ 所示

这一结果可从单条边推广到如下可容许切割。

推论 14.3。 设 C 是 T 上的一个可容许切割，$F_v = \pi_C(T)$ 是通过该可容许切割提取出的森林，$\rho_C^{(d)}(T) = T/^dF_v$ 是对应的商，出现在余积项 $\pi_C(T) \otimes \rho_C^{(d)}(T)$ 中。于是，对切割中的每条边 $e \in C$，令 $\epsilon_e = 0$，可得

$$A^T|_{\epsilon_e=0\,,e\in C} = A^{T/^dF_v}\,. \tag{14.6}$$

证明。如引理 14.2 所述，我们有 $A^T|_{\epsilon_e=0,\ e\in C}$在 F_v中的所有叶子上均为零；在通往根的路径不包含切割 C的边上方任何顶点的所有叶子上保持不变；而对于路径包含与 C的一条边互为姐妹边的叶子，它等于路径中所有其他不与切割边互为姐妹边的边的 ϵ_e 之乘积。一条从叶子到根的路径不能包含多于一条 C的边，因为 C是一个可容许切割。因此，所得的 $A^T|_{\epsilon_e=0,\ e\in C}$恰好就是 $A^{T/^dF_v}$。□

通过将 A^T 特化到参数化立方体 $[0,1]^{n-1}$边界的某些更深层分层，我们还可以得到所提取的可访问项 T_v 的概率分布。

引理 14.4。 设 $T_v \subset T$ 是句法对象 T 的一个可访问项，并设 $\gamma_{v_0,\, v}$ 是从根顶点 v_0 到 v 的 T 中的路径。那么，将所有边 $e \in \gamma_{v_0,\, v}$ 的变量 ϵ_e 特化为 $\epsilon_e = 1$ ，得到

$$A^T|_{\epsilon_e=1\,, e\in\gamma_{v_0,v}} = A^{T_v}\,. \tag{14.7}$$

证明。 对于 $\ell \in L(T_v)$ ，我们有

$$a_\ell^T|_{\epsilon_e=1\,, e\in\gamma_{v_0,v}} = \prod_{w\in\gamma_{v,\ell}} \epsilon_w\,, \tag{14.8}$$

而对于 $\ell \in L(T) \smallsetminus L(T_v)$ ，路径 $\gamma_{v_0,\, \ell}$ 将包含 $\gamma_{v_0,\, v}$ 中某条边的姐妹边。由于 $\gamma_{v_0,\, v}$ 中各边的权重 ϵ_e 被特化为 $\epsilon_e = 1$，因此它们的姐妹边被特化为 0。这意味着，对于所有 $\ell \in L(T) \smallsetminus L(T_v)$

$$a_\ell^T|_{\epsilon_e=1\,, e\in\gamma_{v_0,v}} = 0\,. \tag{14.9}$$

这就恰好给出了概率分布 A^{T_v}。 □

于是，我们可以在参数化立方体中关联以下几何轨迹，该轨迹对应于可访问项 $T_v \subset T$ 的提取及相应的商 $T/^d T_v$。

命题 *14.5*。设 $T_v \subset T$ 是一个可访问项。对于余积$\Delta^d(T)$中的项 $T_v \otimes T/^d T_v$，我们可以分别关联一对定义在 $L(T_v)$ 和 $L(T/^d T_v)$ 上的概率分布 $(A^{T_v}, A^{T/^d T_v}) = L(T) \smallsetminus L(T_v)$。用变量 $\lambda = (\lambda_v)_{v\in V^o(T)}$ 表示时，这对概率分布对应于乘法 $\mathcal{Q}_{T_v} \times \mathcal{Q}_{T/^d T_v}$ 中的一个点 $(\lambda^{T_v}, \lambda^{T/^d T_v})$，它位于参数化立方体 $[0,1]^{n-1}$ 的两个子立方体中，其维数分别为 $\dim \mathcal{Q}_{T_v} + \dim \mathcal{Q}_{T/^d T_v} = n-2$，其中 $\mathcal{Q}_{T_v} \simeq [0,1]^{V^o(T_v)} \subset [0,1]^{V^o(T)}$，且 $\mathcal{Q}_{T/^d T_v} \simeq [0,1]^{V^o(T/^d T_v)} \subset [0,1]^{V^o(T)}$。

证明。在引理 14.2 中，概率 $A^{T/^d T_v}$ 是维数为 $\#L(T/^d T_v) - 1$的分层 $\Delta_{L(T/^d T_v)}$ 中的一个点，该分层位于 $(n-1)$-单纯形 $\Delta_{L(T)}$内。另一方面，超立方体$[0,1]^{n-1}$中由条件 $\lambda_v = 0$ (或 $\lambda_v = 1$) 切出的轨迹是一个余维为 1 的面 $[0,1]^{n-2}$。区别在于，顶点 $w \in V^0(T_v)$ 的权重 λ_w 不会对 $A^{T/^d T_v}$产生贡献，这是由 (14.4) 所致；而它们是 $[0,1]^{n-2}$中的坐标。因此，通过固定变量 λ_w 的值，可以得到面 $[0,1]^{n-2}$ 的一个子轨迹，它对应于点 $A^{T/^d T_v}$ ；这些变量满足$w \in V^0(T_v)$（由 (14.4) 可知，固定为哪个值并不重要：我们将使用 $\lambda_w = 0$）。我们将所得的 $(\#L(T/^d T_v) - 1)$ 维子立方体记作 $\mathcal{Q}_{T/^d T_v} \subset [0,1]^{n-1}$，并可将其识别为 $\mathcal{Q}_{T/^d T_v} \simeq [0,1]^{V^o(T/^d T_v)} \subset [0,1]^{V^o(T)}$。

同样，对于引理 14.4，概率 A^{T_v} 对应于维数为 $\#L(T_v) - 1$ 的分层$\Delta_{L(T_v)}$中的一个点，该分层位于 $n-1$ 维单纯形 $\Delta_{L(T)}$内。另一方面，轨迹 $\{\epsilon_e = 1\,, e \in \gamma_{v_0,v}\}$ 确定了参数化立方体边界上的一个子立方体 $[0,1]^{n-r-1} \subset \partial[0,1]^{n-1}$，其中 $r = \mathsf{length}(\gamma_{v_0,v})$，其余维更大，因为在 $A^{T_v} \in \Delta_{L(T_v)}$ 中，所有位于从 $L(T) \smallsetminus L(T_v)$ 到路径 γ_{v_0}的路径上的顶点 w 的变量 ϵ_w 不再起作用。因此，我们可以将范围限制在立方体 $[0,1]^{\#L(T_v)-1} \subset [0,1]^{n-r-1}$ 上，它是通过将这些顶点的所有λ_w 设为一个固定值而得到的。再次强调，取哪个值并不重要，因为根据 (14.9)，这些权重不会影响 A^{T_v} 。为方便后续讨论，我们将这些值设为 0。我们将所得的（$\#L(T_v) - 1$）维子立方体记为 $\mathcal{Q}_{T_v} \subset [0,1]^{n-1}$，并将其视为 $\mathcal{Q}_{T_v} \simeq [0,1]^{V^o(T_v)} \subset [0,1]^{V^o(T)}$。

因此，对于任意可访问项 $T_v \subset T$，通过对参数化立方体 $[0,1]^{n-1}$中的变量进行特化，我们在边界 $\partial[0,1]^{n-1}$内得到了两个低维立方体 $\mathcal{Q}_{T_v}$ 和 $\mathcal{Q}_{T/{}^dT_v}$，且 $\dim \mathcal{Q}_{T_v} + \dim \mathcal{Q}_{T/{}^dT_v} = n-2$。□

推论 14.6。在内部合并操作 $\mathfrak{M}(T_v, T/{}^dT_v)$下，叶子处的概率分布 $A^{\mathfrak{M}(T_v,T/{}^dT_v)}$ 是连接中的一个点

$$\Delta_{L(T_v)} \star \Delta_{L(T/{}^dT_v)} \tag{14.10}$$

两个单纯形$\Delta_{L(T_v)}$ 和 $\Delta_{L(T/{}^dT_v)}$，其中 $A^{\mathfrak{M}(T_v,T/{}^dT_v)}$ 是如 (14.2) 所示的关于 $\lambda \in [0,1] \times \mathcal{Q}_{T_v} \times \mathcal{Q}_{T/{}^dT_v}$的多项式函数。

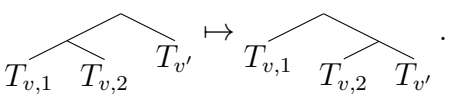
证明。当我们构造内部合并 $\mathfrak{M}(T_v, T/{}^dT_v)$ 时，会添加一个新的根顶点 u，并相应地添加一个新变量 $\lambda_u \in [0,1]$。此时叶子处的概率分布满足

$$A^{\mathfrak{M}(T_v,T/{}^dT_v)} = \begin{cases} \lambda_u A^{T_v} + (1-\lambda_u) A^{T/{}^dT_v} & h_T(v_0) \in L(T_v) \\ (1-\lambda_u) A^{T_v} + \lambda_u A^{T/{}^dT_v} & h_T(v_0) \notin L(T_v)\,, \end{cases} \tag{14.11}$$

其中 v_0 是 T的根。也就是说，$A^{\mathfrak{M}(T_v,T/{}^dT_v)}$ 是连接 (14.10) 中的一个点，它又可以与单纯形 $\Delta_{L(T)}$ 对应起来。根据命题 14.5，对于参数化立方体，概率 $A^{\mathfrak{M}(T_v,T/{}^dT_v)} \in \Delta_{L(T_v)} \star \Delta_{L(T/{}^dT_v)}$ 可以写成形式为 (14.2) 的多项式函数 $a^{\mathfrak{M}(T_v,T/{}^dT_v)}{}_\ell = \mathbb{P}_\ell(\lambda)$，其中 $\lambda \in [0,1] \times \mathcal{Q}_{T_v} \times \mathcal{Q}_{T/{}^dT_v}$。□

上文已通过一个参数化立方体 $[0,1]^n = [0,1] \times [0,1]^{n-1}$ 描述了内部合并操作。从一个点 $\lambda^T \in \{0\} \times [0,1]^{n-1}$ 出发，其中 $A^T = P(\lambda^T) \in \Delta_{L(T)}$，首先按照上文所述通过变量特化，将其映射为一对点 $(\lambda^{T_v}, \lambda^{T/{}^dT_v}) \in \mathcal{Q}_{T_v} \times \mathcal{Q}_{T/{}^dT_v}$，其中 $(A^{T_v} = P(\lambda^{T_v})\ A^{T/{}^dT_v} = P(\lambda^{T/{}^dT_v})) \in \Delta_{L(T_v)} \times \Delta_{L(T/{}^dT_v)}$。然后将这两个点合并，得到一个如 (14.11) 所示的点，其中 $A^{\mathfrak{M}(T_v,T/{}^dT_v)} = P(\lambda_u, \lambda^{T_v}, \lambda^{T/{}^dT_v}) \in \Delta_{L(T_v)} \star \Delta_{L(T/{}^dT_v)}$，适用于 $(\lambda_u, \lambda^{T_v}, \lambda^{T/{}^dT_v}) \in [0,1] \times [0,1]^{V^o(T_v)} \times [0,1]^{V^o(T/{}^dT_v)} \subset [0,1]^n$。因此，我们可以将内部合并视为 $[0,1]^n$ 内部的一种连续流的结果：它从子立方体 $\{0\} \times [0,1]^{V^o(T)}$ 中的一个点流向子立方体 $[0,1] \times [0,1]^{V^o(T_v)} \times [0,1]^{V^o(T/{}^dT_v)}$ 中的一个点。接下来可以探讨，这些内部合并变换能否作为参数化立方体 $[0,1]^n$ 上某种连续流的一部分来实现。类似地，也可以考察外部合并和最小 SM 变换对概率 $A^T = P(\lambda^T)$ 和参数 $\lambda^T \in [0,1]^{V^o(T)}$ 的影响；这或许能使合并动力学嵌入这样的实现之中：在某个语义空间（如 [22] 第 3 章所述）中将句法对象表示为度量树，或采用 [21] 中所述的函数空间表示。任何这样的连续流都必须满足某种相容性条件，以与前文各节所描述的、合并对工作空间的作用所对应的离散马尔可夫链相容，从而能够表示合并动力学的一种实现。

当把句法对象视为 [4] 的 BHV 模空间中的点，并为沿边分配连续参数时，会引出一个相关问题。BHV 模空间各开胞腔之间的墙对应于这样的树：其中一个边变量（或在高余维分层中，多个边变量）等于 0。由余维一边界分层分隔的两个相邻开胞腔之间的墙穿越，对应于在 T 的某个内部顶点处进行如下形式的转换



由于BHV中的墙穿越是以这种方式发生的，即将一条边穿过某个非叶顶点，因此像IM这样提取一个可访问项（可能位于树的深处）的操作，

并将其在根处重新合并，似乎并不直接符合BHV边界分层的形式，也就是说，项 $T_v \otimes T/^d T_v$ 似乎并不自然地对应于余维一分层（或 $F_v \otimes T/^d F_v$ 对应于高余维分层）。

问题14.7。 离散合并霍普夫代数马尔可夫链 $\hat{\mathcal{K}}^{(A,n)}$ （及其加权版本）能否嵌入参数化立方体 $[0,1]^n$或模空间 BHV_n中的连续马尔可夫链，或者嵌入语义空间的拓扑模型中？

关于连续参数、参数化空间以及与离散马尔可夫链相容的流，我们将这些考量留作未来研究的问题。

15. I-language constraints and Colored Merge dynamics

合并作用于工作空间所产生的霍普夫代数马尔可夫链，是通过自由对称合并在句法中形成自由结构的动力学模型。

在句法极简主义模型中，自由对称合并生成的自由形成结构，随后会经过过滤，以检验其是否符合θ角色指派（θ准则）以及结构良构的相位（中心语与补足语结构、扩展投射）。正如 [25] 和 [24]中所解释的，这两种过滤系统都可以用彩色歌剧代数的芽生成系统来进行数学建模。

此外， [25] 和 [24] 还表明，按照这些着色规则过滤自由构成的结构，等同于对合并作用施加约束，并且只形成与这些约束相容的结构。这种等价性的原因在于，相关彩色歌剧（定义见§15.1）的所有芽生成元都是只有一个节点、两个输入叶和一个输出根的花冠。人们将这种受约束形式的合并称为“彩色合并”。

如此解读便可看出，只选取具有结构良构的相位和θ角色的句法对象，其效果相当于对霍普夫代数马尔可夫链 $\hat{\mathcal{K}}^{A,\,n}$ 在合并图$\mathcal{G}_{n,\,A}$上给出的自由合并动力学进行剪枝；剪枝后仅保留着色与某个芽生成元相对应的箭头，以及通过这些彩色合并运算并带有相应着色而得到的工作空间所对应的顶点，从而使该工作空间的所有分量树都是给定彩色歌剧中的元素。

在这种彩色合并表述中，人们不再采用上文所述的自由合并操作形式

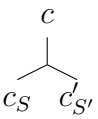
$$\mathfrak{M}_{S,S'} = \sqcup \circ (\mathcal{B} \otimes \mathrm{id}) \circ \delta_{S,S'} \circ \Delta \,,$$

而是考虑如下形式的版本

$$\mathfrak{M}^c_{S,S'} = \sqcup \circ (\mathcal{B}^c \otimes \mathrm{id}) \circ \delta^c_{c_S,c_{S'}} \delta_{S,S'} \circ \Delta \,, \tag{15.1}$$

其中 c， c_S， $c_{S'} \in \Omega$ 是颜色，而

(15.2)



属于彩色歌剧 $\mathfrak{O}_{\Omega,\,\mathbb{B}}$ 的芽生成元集合 $\mathbb{B}$ （更多细节见 §15.1），并且S， S' 是属于该彩色歌剧的句法对象，其中 c_S 和 $c_{S'}$ 分别是分配给 S 和 S' 的根顶点的颜色。由此得到的 $\mathfrak{M}^c$（S， S'）的根顶点着色为颜色 c，其余顶点则根据 S 和 S'的着色进行着色。

令 Ω 为颜色的集合（Ω_Θ、Ω_Φ 分别表示 θ角色和阶段，或者如 [24]中所述的组合情形 $\Omega_{\Theta,\,\Phi}$ ）。

有色合并图 $\mathcal{G}^{\Omega}_{n,\,A}$（特别是 $\mathcal{G}^{\Theta}_{n,\,A}$ 或 $\mathcal{G}^{\Phi}_{n,\,A}$，或者 $\mathcal{G}^{\Theta,\,n}{}^{\Phi}_{A\,,}$）可以通过两种可能的方式从 $\mathcal{G}_{n,\,A}$得到，这反映了将着色视为对已经形成的结构进行过滤，或视为通过有色合并进行构建这两种等价的思考方式。我们将在第 15.1 节简要回顾有色歌剧和芽生成系统的概念后，于第 15.2 节和第 15.3 节介绍这两种构造。

15.1. 彩色歌剧与芽生成系统。（集合范畴中的）彩色歌剧是由一族集合构成的

$$\mathfrak{O} = \{\mathfrak{O}(c, \underline{c}) \,|\, \underline{c} = (c_1, \dots, c_n),\ n \geq 1,\ c, c_i \in \Omega\}$$

其中 Ω 是一个（有限）颜色集合，并配有复合运算

$$\begin{aligned} \gamma:\ & \mathfrak{O}(c, (c_1, \dots, c_n)) \times \mathfrak{O}(c_1, (c_{1,1}, \dots, c_{1,k_1})) \times \cdots \times \mathfrak{O}(c_n, (c_{n,1}, \dots, c_{n,k_n})) \\ & \to \mathfrak{O}(c, (c_{1,1}, \dots, c_{1,k_1}, \dots, c_{n,1}, \dots, c_{n,k_n})) \end{aligned} \tag{15.3}$$

这些运算满足结合性（此处不明确写出）。彩色算子上的代数是一族集合 $A = \{A_c\}_{c\in\Omega}$，配有一个作用

$$\gamma_A : \mathfrak{O}(c, (c_1, \dots, c_n)) \times A_{c_1} \times \cdots \times A_{c_n} \to A_c$$

满足与复合 γ 相容的条件（此处同样不明确写出）。

构造彩色歌剧的一种便捷方法，也是形式化 [25] 和 [24], 中 θ角色与阶段结构的基础，是采用 芽生成系统, 如 [17] 中所述。

首先从一个普通算子 $\mathfrak{O}$（无颜色）和一个有限颜色集 Ω 开始。首先，通过所有可能的颜色分配构造一个彩色算子,

$$\mathbb{B}_\Omega(\mathfrak{O})(n) := \Omega \times \mathfrak{O}(n) \times \Omega^n\,. \tag{15.4}$$

然后在 $\mathbb{B}_\Omega(\mathfrak{O})(n)$中选取一个有限集合 $\mathcal{R} \subset \mathbb{B}_\Omega(\mathfrak{O})$，作为希望得到的实际（限制性更强的）彩色算子的生成元。

定义一个 芽生成系统 $\mathbb{B} = (\mathfrak{O},\ \Omega\ \mathcal{R},\ \mathcal{I},\ \mathcal{T})$，其数据包括（无颜色）算子$\mathfrak{O}$、有限颜色集 Ω（划分为初始颜色集 $\mathcal{I}$ 和终端颜色集 $\mathcal{T}$），以及生成元集合 $\mathcal{R} \subset \mathbb{B}_\Omega(\mathfrak{O})$（即局部着色规则）的选择。

然后得到一个由芽系统 $\mathbb{B} = (\mathfrak{O},\ \Omega\ \mathcal{R},\ \mathcal{I},\ \mathcal{T})$ 生成的彩色算子$\mathfrak{O}_{\Omega,\,\mathbb{B}}$，其形式为

$$\mathfrak{O}_{\Omega,\mathbb{B}}(n) := \{x = (c, T, \underline{c}) \in \mathbb{B}_\Omega(\mathfrak{O})(n) \,|\, \mathbf{1}_c \to_{\mathbb{B}} x,\ c \in \Omega \smallsetminus \mathcal{T},\ \underline{c} \in (\Omega \smallsetminus \mathcal{T})^n\}\,, \tag{15.5}$$

其中 $\mathbf{1}_c \to_{\mathbb{B}} x$ 表示：x 是通过对生成元 r_1、.....、$r_N \in \mathcal{R}$ 进行一系列算子复合，从彩色单位元 $\mathbf{1}_c$ 得到的。

由萌芽系统 $\mathbb{B} = (\mathfrak{O}、\ \Omega\ \mathcal{R}、\ \mathcal{I}、\ \mathcal{T})$ 生成的语言 $\mathcal{L}(\mathbb{B})$ 是彩色歌剧 $\mathfrak{O}_{\Omega、\ \mathbb{B}}$ 上的一个代数，其定义如下

$$\mathcal{L}(\mathbb{B}) = \{x = (c, T, \underline{c}) \in \mathbb{B}_\Omega(\mathfrak{O}) \,|\, \mathbf{1}_c \to_{\mathbb{B}} x,\ c \in \mathcal{I},\ \underline{c} \in \mathcal{T}^n\}\,. \tag{15.6}$$

15.2. 通过过滤构造有色合并图。对于每个 $\alpha \in \mathcal{SO}_0$，都存在一个颜色子集 $\Omega_\alpha \subset \Omega$ 与 α 兼容。因此，对于 $A = \{\alpha_1, \dots, \alpha_n\}$ 叶子上的词汇项，我们有相应的着色选择

$$\Omega_A := \Omega_{\alpha_1} \times \cdots \times \Omega_{\alpha_n}. \tag{15.7}$$

给定叶子 α_1、α_n 的一种着色选择 $(c_1,\ .....,\ c_n)\ \in \Omega_A$，对于图 $\mathcal{G}_{n,\,A}$ 的每个顶点$F = T_1 \sqcup \cdots \sqcup T_r \in \mathfrak{F}_{A,\,n}$，以及 F 的每个分量 T_i，考虑将 T_i 着色为 $\mathcal{L}(\mathbb{B})(k_i) \subset \mathfrak{O}_{\Omega,\,\mathbb{B}}(k_i)$ 中的一个元素的所有可能方式，其中 $k_i = \#L(T_i)$。

如果某个分量在 $\mathcal{L}(\mathbb{B})$ 中没有任何可能的着色（对于叶子的任意可用着色选择），那么顶点 F 不会向 $\mathcal{G}^{\Omega}_{n,\,A}$贡献任何顶点。如果存在着色，那么 F 会贡献若干个“位于 F之上”的顶点，其数量等于不同可能着色的数量。

对于 $F \in \mathfrak{F}_{A,\,n}$ 的每一种着色，以及每个 Merge 箭头（即 $\mathcal{K}^{A,\,F^{n},\,F'} = 1$ 中 $\mathcal{G}_{n,\,A}$的邻接矩阵的每个条目），$\mathcal{G}^{\Omega}_{n,\,A}$ 中都存在一条指向 F'某种着色的箭头，因为我们知道，着色可以通过彩色Merge运算来描述，并且 F 的每种着色都是某个 F''在 $\mathcal{K}^{A,\,F''^{n},\,F} = 1$ 条件下的一种着色所对应箭头的目标。

由这些顶点和箭头组成的图是有色合并图 $\mathcal{G}^{\Omega}_{n,\,A}$。相应地，这给出了一个着色 Merge 马尔可夫链 $\hat{\mathcal{K}}^{A,\,n,\,\Omega}$，其中 $\mathcal{K}^{A,\,n,\,\Omega}$ 是 $\mathcal{G}^{\Omega}_{n,\,A}$ 的邻接矩阵，而 $\hat{\mathcal{K}}^{A,\,n,\,\Omega}$ 则按照（3.6）中的方式得到。

注意，$\mathcal{G}^{\Omega}_{n,\,A}$ 与 $\mathcal{G}_{n,\,A}$ 之间的关系并不直接，因为它们之间不存在投影映射，无法对其应用我们此前发展的技术。实际存在的是一个投影 $p^{\Omega}: \mathcal{G}^{\Omega}_{n,\,A} \to \mathcal{G}^{\mathrm{pruned}}_{n\,A}$，它将前者投影到 $\mathcal{G}_{n,\,A}$ 的剪枝版本，其中移除了所有顶点 F：这些顶点在 $\mathcal{L}(\mathbb{B})$ 中不存在与叶子及其相邻顶点的可用着色相容的着色。

产生 $\mathcal{G}^{\mathrm{pruned}}_{n\,A}$ 的剪枝通常会破坏 $\mathcal{G}_{n,A}$ 关于投影 $p: \mathcal{G}_{n,A} \to \mathcal{G}_{\mathcal{P}'(n)}$ 的对称性，从而使着色 Merge 马尔可夫链 $\hat{\mathcal{K}}^{A,n,\Omega}$ 的动力学比自由 Merge 作用的霍普夫代数马尔可夫链 $\hat{\mathcal{K}}^{A,n}$ 更难分析。此外，尽管 $\hat{\mathcal{K}}^{A,n}$ 的行为与叶子处具体选择的词汇数据 $A = \{\alpha_1, \dots, \alpha_n\}$ 无关，但 $\hat{\mathcal{K}}^{A,n,\Omega}$ 的动力学会对这些数据敏感，并且会因 A 的不同选择而不同，因为叶子处可用着色的集合 $\Omega_A \subset \Omega^n$ 会随 A 改变，而集合 Ω_A 决定了在 $\mathcal{G}^{\mathrm{pruned}}_{n\,A}$ 中哪些顶点会被剪枝，以及 $p^{\Omega}: \mathcal{G}^{\Omega}_{n,\,A} \to \mathcal{G}^{\mathrm{pruned}}_{n\,A}$ 中的重数（纤维的大小）。

这对语言学模型有着重要的启示。原则上，我们知道，可以认为句法中的结构形成是由自由对称合并完成的（动力系统 $\hat{\mathcal{K}}^{A,\,n}$），随后对由此形成的完整结构 $T \in p^{-1}(\{n\})$ 按阶段和θ角色进行着色过滤；也可以只通过彩色Merge运算来形成结构，使得每一步都进行可着色性过滤，且动力学由 $\hat{\mathcal{K}}^{A,\,n,\,\Omega}$ 支配。这些过程是等价的，因为它们在 $\mathcal{L}(\mathbb{B})$ 中产生相同的一组良着色对象，其底层树为 $T \in p^{-1}(\{n\})$。

然而，$\hat{\mathcal{K}}^{A,\,n,\,\Omega}$ 并非等价的动力系统。这一区别可以类比形式语言领域中弱生成能力相同与强生成能力相同之间的区别。无论是 $\hat{\mathcal{K}}^{A,\,n,\,\Omega}$，还是 $\hat{\mathcal{K}}^{A,\,n}$，再经过对 $T \in p^{-1}(\{n\})$ 的可着色性筛选，都会在 $\mathcal{L}(\mathbb{B})$ 中产生相同的元素；这可以类比为形式语言具有相同的弱生成能力。然而，其中涉及的结构构建动力过程本身并不等价。

从动力系统的角度来看，更自然的理解是：结构形成由自由合并通过 $\hat{\mathcal{K}}^{(A,n)}$（以及加权版本）实现，随后进行过滤，因此过滤可以理解为以加权自由合并动力学的平稳分布 π 为条件。

15.3. **通过动力学构建有色合并图**。我们还将在此说明，如何将$\mathcal{G}^{\Omega}_{n,A}$ 的构造等价地视为一个递归过程：将 Merge 箭头$\mathfrak{M}_{S,S'}$ 替换为有色 Merge 箭头$\mathfrak{M}^{c}_{S,S'}$。首先需要假设 S 和 S' 已经着色，特别是需要知道分配给它们根顶点的颜色 c_S 和 $c_{S'}$，然后这便决定了原始箭头$\mathfrak{M}_{S,S'}$ 将被一组箭头$\mathfrak{M}^{c}_{S,S'}$替换；对于每个满足条件（15.2）属于有色歌剧代数生成元集合的 c，都有一个对应的箭头。以同样的方式分析从顶点$\mathfrak{M}^{c}_{S,S'}$ 出发的新箭头时，SM 和 IM 箭头会为原来的 S 和 S'添加新的着色，如此继续，直到得到所有允许的着色和所有允许的彩色合并操作。

可以从工作空间 $F \in p^{-1}(\wp = \{2, 1^{n-2}\})$开始，也就是包含一个樱桃分量、其余分量均为单叶的工作空间。不失一般性，设 $F = \mathfrak{M}(\alpha_1, \alpha_2) \sqcup \alpha_3 \sqcup \cdots \sqcup \alpha_n$。

对于单叶分量 $\alpha_3 \sqcup \cdots \sqcup \alpha_n$，我们有 $\#(\Omega_{\alpha_3} \times \cdots \times \Omega_{\alpha_n})$ 种可能的着色，而对于 $\mathfrak{M}(\alpha_1,\ \alpha_2)$，可能的着色集合

$$\Omega_{\alpha_1,\alpha_2} \subset \Omega \times \Omega_{\alpha_1} \times \Omega_{\alpha_2}$$

是子集

$$\Omega_{\alpha_1,\alpha_2} = \{(c, c_1, c_2) \mid \begin{array}{c} c \\ \diagup\ \diagdown \\ c_1 \quad c_2 \end{array} \in \mathbb{B}\}.$$

因此，对于 $\mathcal{G}_{n,\ A}$ 中的每个顶点 $F \in p^{-1}(\wp = \{2, 1^{n-2}\})$，在 $\mathcal{G}^{\Omega}_{n,\ A}$ 中将有 $N_{\alpha_1,\ \alpha_2} \cdot \prod_{i=3}^{n} N_{\alpha_n}$ 个顶点，其中 $N_{\alpha_i} = \#\Omega_{\alpha_i}$ 且 $N_{\alpha_i,\ \alpha_j} = \#\Omega_{\alpha_i,\ \alpha_j}$。

接下来，可以考虑从 $\mathcal{G}^{\Omega}_{n,\ A}$ 的顶点指向 $\mathcal{G}_{n,\ A}$ 中 $p^{-1}(\wp = \{2, 1^{n-2}\})$ 内顶点上方顶点的 EM 箭头，并通过以下方式为所得分量 $\mathfrak{M}(\mathfrak{M}(\alpha_i.\alpha_j)\ \alpha_k)$ 着色：如果 $\mathfrak{M}(\alpha_i,\ \alpha_j)$ 的根与 α_k 的颜色 c 和 c_k 满足.....，则将颜色 c' 指派给 $\mathfrak{M}(\mathfrak{M}(\alpha_i.\alpha_j)\ \alpha_k)$ 的根顶点

$$\begin{array}{c} c' \\ \diagup\ \diagdown \\ c \quad c_k \end{array} \in \mathbb{B},$$

当不存在具有此性质的 $c' \in \Omega$ 时，就不存在 EM 箭头。

对给定的 $\wp \in \mathcal{P}'(n)$，重复这一过程，将所有工作空间 $F \in p^{-1}(\wp)$ 视为 $\mathcal{G}_{n,A}$ 中的顶点。由于每个这样的 F 都是某个 EM 映射的目标，我们便得到 $\mathcal{G}^{\Omega}_{n,A}$ 中位于 F 上方的一组顶点；这些顶点是通过上述方法对 EM 映射的目标进行着色得到的，前提是这些 EM 箭头的源点已用同样的方法着色。

通过着色 EM 得到所有可能的着色后，对于所有 $\wp \in \mathcal{P}'(n)$，在每个 $F \in p^{-1}(\wp)$ 上方，我们考察每个 $F \in p^{-1}(\wp)$ 处的 IM 箭头，其中 $\wp$ 至少有一个 $k_i \geq 3$（因此存在 IM 箭头）。对于给定 $F \in p^{-1}(\wp)$ 上方的每个着色顶点，我们考察 IM 箭头的所有可能着色。

如 [25] 和 [24], 中所述，在这种情况下，树 $\mathfrak{M}(T_v, T/^d T_v)$ 的着色通过两次彩色Merge运算得到：$\mathfrak{M}^c_{T_v,1}$ 其着色为

$$\begin{array}{c} c_1 \\ \diagup\ \diagdown \\ c_v \quad c_0 \end{array} \in \mathbb{B} \tag{15.8}$$

c_v 顶点 v 的颜色为 T，c_0 单位元 1 所赋的颜色为，c_1 是为此类生成元指定的另一种颜色（例如，在阶段的情况下为相位边颜色 $\mathfrak{s}^{\downarrow}$，或θ角色的非θ位置颜色）。随后进行一次彩色 Merge$\mathfrak{M}^{c'}(\mathfrak{M}^c_{T_v,1}, T/^d T_v)$，其中

$$\begin{array}{c} c' \\ \diagup\ \diagdown \\ c_1 \quad c \end{array} \in \mathbb{B} \tag{15.9}$$

c根 T 的颜色为。由于标记为 1 的叶子在 $\mathfrak{M}(T_v, T/^d T_v)$ 中不计作叶子，因此这种着色等价于仅满足条件 (15.9)，其中 c根 T 的颜色为，并且将新颜色赋予 v （以替换其原有颜色 c_v）。

每当不存在满足所需条件的颜色三元组时，$\mathcal{G}_{n,A}$ 所对应的 IM 箭头就会被剪除，不再作为$\mathcal{G}^{\Omega}_{n,A}$的一条边存在。这些 IM 箭头将添加

新的着色顶点到 $F \in p^{-1}(\wp)$之上，因为它们可能产生 EM 箭头尚未实现的着色。这些顶点会被添加到$\mathcal{G}^{\Omega}_{n,A}$。

类似地，可以从 $\mathcal{G}^{\Omega}_{n,A}$ 中所有已获得的顶点出发，考虑所有可能的最小 SM 箭头，并考察相应的着色条件。如 [26], 所示，这些箭头会用到 $\mathbb{B}$ 中 EM 和 IM 未使用的额外生成元，因此会产生新的可能着色，进而为 $\mathcal{G}^{\Omega}_{n,A}$ 添加新的顶点。

接着，可以再次考察从所有已获得的着色结构出发的全部 EM、IM、SM 箭头，看看它们是否会产生新的着色。由于底层树结构保持不变，因此可能的着色数量存在上限，整个过程最终会稳定下来。由此便可从合并图 $\mathcal{G}_{n,A}$ 得到有色合并图 $\mathcal{G}^{\Omega}_{n,A}$ 。

16. Parameter setting and Externalization

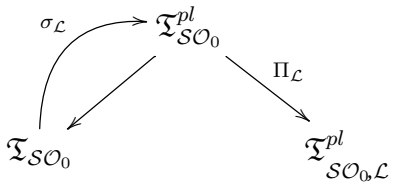
除了针对阶段和 θ角色的“普遍”过滤（即内在语言过滤）之外，对自由对称合并所形成的句法对象还有另一种极为重要的过滤方式，即依据语言特定句法参数进行过滤。自由对称合并被视为普遍语法的核心计算结构，而各个语言之间的参数变异（自原则与参数理论最初提出以来）则通过参数设定程序来建模。该程序发生在语言习得过程中，并使普遍计算机制适应某一种特定语言的具体特点。

这一过程可以看作是依据语言特定过滤器，对通过 $\hat{}_{\mathcal{K}^{A,\,n,\,\Omega}}$动力学（或先经过 $\hat{}_{\mathcal{K}^{A,\,n}}$ 动力学，再进行可着色性筛选）得到的 $\mathcal{L}(\mathbb{B})$ 中的结构进一步进行过滤，同时为树指定平面结构（这也取决于语言特定的语序参数）。这一过程称为外化。

《 [22], 》第 1.12 节对最简主义数学模型语境下的外化作了一些讨论，另有一篇更为详尽的讨论将另行发表（[23], [32]），因此我们在此不讨论外化和句法参数的一般模型。不过，我们会简要谈谈如何从马尔可夫链 $\hat{}_{\mathcal{K}^{A,\,n}}$ 和 $\hat{}_{\mathcal{K}^{A,\,n,\,\Omega}}$的视角看待参数设定这一过程。

这里我们仅从这一动力系统的视角，对如何理解参数设定过程作非常概略的介绍。全面的探究必须纳入更为详细的句法参数模型以及参数设定过程，其中一部分正在 [23], [32], 中开发，因此留待未来研究。

这里我们采用 [22], §1.12 中提出的外化模型，将外化视为一个两步过程：一个截面 $\sigma_{\mathcal{L}}$ 以依赖于语言的方式为树赋予平面结构；以及一个投影 $\Pi_{\mathcal{L}}$ 以同样依赖于语言的方式过滤掉某些结构，从而得到如下形式的图



$\mathfrak{T}^{pl}_{\mathcal{SO}_0}$ 表示平面二叉有根树的集合，而 $\mathfrak{T}^{pl}_{\mathcal{SO}_0,\,\mathcal{L}}$ 表示未被语言 $\mathcal{L}$ 的句法参数过滤掉的那些平面二叉有根树的集合。

在这幅图景背后，我们假设存在一个句法参数空间（为简便起见，这里将其取为二值的，尽管允许第三个“未定义”值通常也很有用）。我们将所有可能的参数值组成的空间与二值超立方体 $\mathbb{F}_2^N = \{0,1\}^N$ 对应，其中

N 是参数的总数。为便于讨论，我们还假设这些 N 参数是相互独立的二元变量，尽管实际上，句法参数之间的关系问题是外化建模中的关键问题之一；但就我们这里的目的而言，可以暂时忽略这一点，并假设能够确定一组实际足够的独立参数。在这些 N 参数中，有一定数量的参数 $M < N$涉及语序属性，其作用是确定截面 $\sigma_{\mathcal{L}}$。其余参数则确定投影 $\Pi_{\mathcal{L}}$。（就我们这里的目的而言，$\sigma_{\mathcal{L}}$ 和 $\Pi_{\mathcal{L}}$ 究竟如何从参数值中获得并不重要，除了下面将要讨论的少数几个问题之外。）

作为模型的一部分，我们假设句法对象叶子上的词汇项 $\alpha_\ell \in \mathcal{SO}_0$ 携带与之关联的数据 φ_ℓ，这些数据来自有限的句法特征集 $\mathbf{\Phi}$。句法参数$\mathfrak{p} = \mathfrak{p}(\varphi \in \mathbf{\Phi})$ 是特征 $\varphi \in \mathbf{\Phi}$ 的函数，其中通常每个句法参数只取决于某个特定的特征子集。因此，句法参数 $\mathfrak{p}$ 会影响（作为平面化或过滤器）某个结构 F^Ω（$\mathcal{G}^\Omega_{n,\ A}$ 的一个顶点），当且仅当其所依赖的特征 $\varphi \in \mathbf{\Phi}$ 存在于 F^Ω 各分量的叶子上。我们用 $\mathbf{\Phi}_\mathfrak{p} \subset \mathbf{\Phi}$ 表示参数 $\mathfrak{p}$ 所依赖的特征子集。

参数值是在语言习得过程中设定的。正如 [20], 所指出的，这两类参数在这方面的表现不同。对于我们认为决定投影 $\Pi_{\mathcal{L}}$ 的那类句法参数，有经验证据表明其默认值为$-$（或 $0 \in \mathbb{F}_2$），而要将该值切换为 $+$（$1 \in \mathbb{F}_2$），语言习得过程中就需要正面证据。因此，我们可以假设子空间 $\mathbb{F}_2^{N-M}$ 初始化为零向量。另一方面，语序参数似乎没有默认状态。因此，[20] 提出，这些参数应具有“未指定”的初始状态，随后两种可能的 $-/+$ 状态作为对称性破缺现象由此产生。在此处提出的模型中，构造截面 $\sigma_{\mathcal{L}}$ 可以表述为这样一种设定：语序参数被初始化为未定义状态，而正面证据则为其赋值（等价地，递归地构造 $\sigma_{\mathcal{L}}$）。另一方面，[23] 对最终语在最终语条件的分析表明，可以假设存在一种由头函数决定的默认和谐中心语后置平面嵌入；其他嵌入则通过群作用获得，而最终语在最终语条件体现了相对于该群作用生成元的极小性性质。请注意，所有在阶段中可着色的句法对象都有头函数，因此，在经过结构良好的阶段过滤后，就可以采用嵌入的默认假设。为简单起见，我们在此假设语序参数没有默认值（因此也没有平面嵌入的默认选择）。

假设我们从图 $\mathcal{G}^\Omega_{n,\ A}$ 开始，那么基于 θ角色和阶段的过滤就已经完成。

首先，我们需要对工作空间 $F\Omega$ 中的分量树 T_i^Ω（$\mathcal{G}^\Omega_{n,\ A}$的顶点，按上一节所述构造）进行特定于语言的平面嵌入选择（截面$\sigma_{\mathcal{L}}$）。这里使用 Ω 上标，以便记住这些树带有颜色，$T_i^\Omega \in \mathcal{L}(\mathbb{B})$。

为此，我们可以在 $\mathcal{G}^\Omega_{n,\ A}$ 的每个顶点 $F\Omega = T_1^\Omega \sqcup \cdots \sqcup T^{\Omega r}$ 上方，取一组

$$2^{n-r} = \prod_{i=1}^{r} 2^{k_i - 1}$$

顶点由各个分量所有可能的平面嵌入选择给出 T_i^Ω。（一个带有标号叶子的满二叉根树 T 有 $2^{\#L(T)-1}$ 个平面嵌入。）

先验地说，这些顶点都未被选中。然而，在语言习得过程中接触正例会选中其中一些 $\sigma_{\mathcal{L}}(T^{\Omega})$，也就是为某些句法对象 $T^{\Omega} \in \mathcal{G}^{\Omega}_{n,\,A}$ 选定一个特定的平面嵌入。这一选择会自动导致在 $\sigma_{\mathcal{L}}(T_{\Omega})$ 出现的所有地方进一步选定相同的平面嵌入：作为某个 F^{Ω} 的分量，出现在某个具有另一组词汇数据的其他 $\mathcal{G}^{\Omega}_{m,\,A'}$ 中；这组数据与 A 共享相同特征，而语序参数取决于这些特征。当 T_{Ω} 作为可访问项出现在某个更大的 $\tilde{T}^{\Omega}$ 内部时，它也会选定 $\sigma_{\mathcal{L}}(T_{\Omega})$，除非其他语序参数因 $\tilde{T}^{\Omega}$ 其余部分中存在的其他数据而导致 T^{Ω} 的平面嵌入发生变化。对句法参数作局部性与有限性假设（进一步讨论见 [23]，[32] ）意味着，任意大小的所有 $\sigma_{\mathcal{L}}(T^{\Omega})$，都可通过了解大小较小的 $\mathcal{G}^{\Omega}_{n,\,A}$ 图上的嵌入 $\sigma_{\mathcal{L}}$ 来确定；具体做法是识别小规模可访问项与商项的平面化，如下文所述。（关于所需大小 n_0 的具体界限，见 [23], [32]。）

此过程会选出一组顶点 $\sigma_{\mathcal{L}}(F^{\Omega})$，其基数与 $\mathcal{G}^{\Omega}_{n,A}$ 的顶点集 F^{Ω} 相同，因为 $\sigma_{\mathcal{L}}$ 为每个 F_{Ω} 指定了唯一的平面结构（此处我们不讨论移位现象，以及如何处理自由语序相对灵活的语言）。等价地，我们可以将这些顶点写成对 $(F^{\Omega}, \pi_{\mathcal{L}})$，其中 $\pi_{\mathcal{L}}$ 是从平面嵌入中选出的一个。

用于确定投影 $\Pi_{\mathcal{L}}$ 的其余参数 $N-M$ 初始化为 $0 \in \mathbb{F}_2^{N-M}$。除非语言习得过程中的正例将其中某些参数的值翻转为 1，否则它们将保持默认值 0。句法参数 $\mathfrak{p}$ 的默认值为 0，意味着所有结构 $F^{\Omega} \in \mathcal{G}^{\Omega}_{n,A}$ 中，与叶子 $\ell \in L(T)$ 处词汇数据 α_{ℓ} 相关联、且参与确定参数 $\mathfrak{p}$ 的句法特征 φ_{ℓ}，都会在投影 $\Pi_{\mathcal{L}}$ 中被过滤掉。

我们假设投影 $\Pi_{\mathcal{L}}$具有两个基本性质。

- 投影 $\Pi_{\mathcal{L}}$ as 可分解为乘法 $\Pi_{\mathcal{L}} = \prod_{\mathfrak{p}} \Pi_{\mathcal{L},\,\mathfrak{p}}$。因此，结构 F_{Ω}通过过滤器，当且仅当它相对于每个参数都通过过滤器，$\Pi_{\mathcal{L},\,\mathfrak{p}}(F_{\Omega}) = F_{\Omega}$ 对所有 $\mathfrak{p}$均成立。
- 对于 $F_{\Omega} = T_{\Omega_1} \sqcup \cdots \sqcup T_{\Omega_r}$，我们有 $\Pi_{\mathcal{L}}(F_{\Omega}) = 0$ 当且仅当存在某个 $1 \le i \le r$，使得 $\Pi_{\mathcal{L}}(T_{\Omega_i}) = 0$。

将所有 $N-M$ 参数初始化为 0，意味着在过程开始时，所有句法对象都会被过滤掉，因此没有任何对象能通过外化过滤器。接着，假设在语言习得过程中遇到正例，导致其中一个参数 $\mathfrak{p}$切换为值 1。这一变化只会影响那些满足以下条件的结构：$F^{\Omega} \in \mathcal{G}^{\Omega}_{n,\,A}$ 叶子处的特征满足 $\{\varphi_{\ell}\}_{\ell \in L(F^{\Omega})} \supset \mathbf{\Phi}_{\mathfrak{p}}$。

根据上述关于投影 $\Pi_{\mathcal{L}}$ 的考量，我们可以假设 $F_{\Omega} = T^{\Omega}$ 只有单个分量，并且特征集 $\{\varphi_{\ell}\}_{\ell \in L(T^{\Omega})}$ 包含集合 $\mathbf{\Phi}_{\mathfrak{p}}$。

为了对将 $\mathfrak{p}$ 的值改为 1 如何改变 $\Pi_{\mathcal{L},\mathfrak{p}}$进行建模，我们可以先假设，正例足以确定在树 T^{Ω} 中哪些对象（其大小 $\#L(T_{\Omega}) \le n_0$ 足够小，且具有 $\{\varphi_{\ell}\}_{\ell \in L(T^{\Omega})} \supset \mathbf{\Phi}_{\mathfrak{p}}$）现在属于 $\Pi_{\mathcal{L},\mathfrak{p}}$ 的值域，而不再属于其核。换言之，集合

$$\mathcal{T}_{\mathfrak{p}} := \{T^{\Omega} \,|\, \#L(T^{\Omega}) \le n_0 \text{ and } \{\varphi_{\ell}\}_{\ell \in L(F^{\Omega})} \supset \mathbf{\Phi}_{\mathfrak{p}}\} \tag{16.1}$$

被分离为两个子集

$$\mathcal{T}_{\mathfrak{p}} = \mathcal{T}_{\mathfrak{p}}^{-} \sqcup \mathcal{T}_{\mathfrak{p}}^{+} \tag{16.2}$$

其中 $\Pi_{\mathcal{L},\mathfrak{p}}|_{\mathcal{T}_{\mathfrak{p}}^{+}} = \mathsf{id}$ 和 $\Pi_{\mathcal{L},\mathfrak{p}}|_{\mathcal{T}_{\mathfrak{p}}^{-}} = 0$。

对于规模更大的树 T^{Ω}，参数 $\mathfrak{p}$（见 [23]）的局部性条件将确保：如果所有可访问项 $T_{\Omega_v} \subset T^{\Omega}$ 以及相应的商 $T_{\Omega} /^{d} T^{\Omega_v}$（其规模至多为 n_0，且其叶子包含相关特征）都在投影 $\Pi_{\mathcal{L},\mathfrak{p}}$的值域内，那么 T^{Ω}也在其中。（注意

，T^{Ω} 仍可能属于其他参数 $\mathfrak{p}' \neq \mathfrak{p}$ 对应的其他投影 $\Pi_{\mathcal{L},\mathfrak{p}'}$ 的核，因此属于 $\Pi_{\mathcal{L}}$的核。）

通过这种参数设定过程，只需少量正例，就足以借助局部性条件“激活”任意规模的结构F^{Ω} 。如果 $\mathcal{T}_{\mathfrak{p}}^{+}$中包含 F^{Ω}的所有大小不超过 n_0的分量的可访问项和商，那么 $F\Omega \in \mathcal{T}_{\mathfrak{p}}^{+}$，因此$\Pi_{\mathcal{L},\,\mathfrak{p}}(F\Omega) = F\Omega$，否则$\Pi_{\mathcal{L},\,\mathfrak{p}}(F^{\Omega}) = 0$ 。

在语言学习过程中为所有参数设定新向量$\mathfrak{p} \in \mathbb{F}_2^N$后，上述过程会激活图$\mathcal{G}^{\Omega}_{n,\,A}$ 中某个顶点子集（F^{Ω}，$\pi_{\mathcal{L}}$），无论如何选择 n 和 A。我们将这个顶点集记为 $V(\mathcal{G}^{\Omega,\,\mathcal{L}}_{n,\,A,\,})$。图 $\mathcal{G}^{\Omega,\,\mathcal{L}}_{n,\,A,\,}$ 是$\mathcal{G}^{\Omega}_{n,\,A}$ 在顶点集 $V(\mathcal{G}^{\Omega,\,\mathcal{L}}_{n,\,A,\,})$上诱导出的图。对 $\mathcal{T}_{\mathfrak{p}}^{+}$和 $\mathcal{T}_{\mathfrak{p}}^{-}$的选择，以及由此构造 $\mathcal{G}^{\Omega,\,\mathcal{L}}_{n,\,A,\,}$，标志着参数设定过程的完成。

对于给定的 $\mathcal{L}$，图 $\mathcal{G}^{\Omega,\,\mathcal{L}}_{n,\,A}$、及其各自关联的最大熵随机游走马尔可夫链，可以被视为语言 $\mathcal{L}$ 的生成语法模型。

致谢。第一作者的本项目研究得到美国国家科学基金会资助，项目编号为 DMS-2104330 和 DMS-2506176，并得到加州理工学院 T&C Chen系统神经科学中心的支持。第二作者的本项目研究得到 Epistea 奖学金资助，并由加州理工学院本科生访问研究项目接待。我们感谢 Paolo Aluffi、Yassine El Maazouz、Sita Gakkhar、Henry Gustafson、Riny Huijbregts、Amy Pang、Juan Pablo Vigneaux 和 Elizabeth Xiao 的宝贵交流。

References

[1] M. Akian, R. Bapat, S. Gaubert, 利用极大加代数研究佩龙-弗罗贝尼乌斯特征值和特征向量的渐近行为，《巴黎科学院院刊》，第327卷，I辑（1998），927–932页。[2] M. Akian, R. Bapat, S. Gaubert, 矩阵多项式特征值的非阿基米德赋值，《线性代数及其应用》498（2016），592–627页。[3] M. Akian, R. Bapat, S. Gaubert, 特征值摄动理论中的极小加方法及广义 *Lidskiĭ-Viĭsik-Ljusternik* 定理，arXiv:math/0402090v3。[4] L.J. Billera, S.P. Holmes, K.Vogtmann, 系统发育树空间的几何，《应用数学进展》27卷第4期（2001），733–767页。[5] T. M. Cover and J. A. Thomas, 信息论基础，第2版。Wiley，2006。[6] N. Chomsky,《极简主义纲领》，麻省理工学院出版社，1995/2015。[7] N. Chomsky,《加州大学洛杉矶分校讲座》，2019。lingbuzz/005485[8] N. Chomsky, 最简主义：我们现处何处，又希望走向何方，《言语研究》第160卷（2021），1–41页。[9] N. Chomsky, T.D. Seely, R.C. Berwick, S. Fong, M.A.C. Huybregts, H. Kitahara, A. McInnerney, Y. Sugimoto,《合并与强极简主义论题》，剑桥专题丛书，剑桥大学出版社，2023。[10] R.A. Cuninghame-Green, 矩阵的特征极大多项式，《数学分析及其应用杂志》95（1983），110–116页。[11] J.C. Delvenne, A.S. Libert, 复杂网络的中心性度量与热力学形式主义，《物理评论E》83（2011），046117。[12] S.L. Devadoss, 模空间的镶嵌与马赛克歌剧，载于《同伦不变代数结构》，第91–114页，《当代数学》第239卷。美国数学会，1999。[13] S.L. Devadoss, J. Morava.《树空间中的导航》，《应用数学进展》67（2015），75–95页。[14] Persi Diaconis, C. Y. Amy Pang, and Arun Ram.《霍普夫代数与马尔可夫链：两个例子与一个理论》。《代数组合学杂志》，39卷第3期（2014）：527–585。[15] S. Friedland, 非负矩阵的极限特征值。《线性代数及其应用》74（1986），173–178页。[16] H.N. Gabow, Z. Galil, T.H. Spencer, R.E. Tarjan. 在无向图和有向图中寻找最小生成树的高效算法。《Comb.》6（1986）第2期，109–122页。

[17] S. Giraudo, 彩色歌剧、彩色歌剧上的级数与组合生成系统，《离散数学》342（2019）第6期，1624–1657页。[18] B. Kustar, 佩龙-弗罗贝尼乌斯特征向量渐近行为的热带分析，arXiv:1908.08234。[19] J.L. Loday, M. Ronco, 组合霍普夫代数，《数学量子》，347–383页，《克莱数学论文集》11，美国数学会，2010。[20] G. Longobardi, A. Treves, 从类基因编码到自组织吸引子的语法参数：一项研究计划，预印本，2023。[21] M. Marcolli, R.C. Berwick, 在函数空间中编码句法对象与合并操作，预印本 arXiv:2507.13501。[22] M. Marcolli, N. Chomsky, R.C. Berwick, 句法合并的数学结构，麻省理工学院出版社，2025。[23] M. Marcolli, M.A.C. Huijbregts, R.C. Berwick, 外化建模，编写中。[24] M. Marcolli, M.A.C. Huijbregts, R.K. Larson, 超岩浆与彩色歌剧：中心语、阶段与θ角色，arXiv:2507.06393。[25] M. Marcolli, R.K. Larson, θ理论：歌剧代数与着色，arXiv:2503.06091。[26] M. Marcolli, R.K. Larson, M.A.C. Huijbregts, 扩展条件“违规”与合并最优性约束，预印本，2025。[27]C. Y. Amy Pang. 组合霍普夫代数上下降算子导出的马尔可夫链。预印本，2016, arXiv:1609.04312。[28] C. Y. Amy Pang.《超八面体下降算子的特征值及其在洗牌中的应用》。《组合学电子期刊》29卷第1期（2022）：论文第1号。32 [50页数]。[29] W. Parry, 内在马尔可夫链，《美国数学会汇刊》112（1964），55–66页。[30] D. Ruelle, 热力学形式主义，艾迪生-韦斯利，1978。[31] I. Senturia, M. Marcolli, 形态句法的代数结构，arXiv:2507.00244。[32] I. Senturia, M. Marcolli, 外化：词汇插入、平面化与参数过滤，编写中。

Department of Mathematics and Department of Computing and Mathematical Sciences, Cali-fornia Institute of Technology, CA 91125, USA电子邮箱地址: matilde@caltech.edu

Mathematical & Physical Sciences, University College London, London, WC1E 6BT, UK电子邮箱地址: david.skigin.24@ucl.ac.uk