



大语言模型驱动的图着色启发式算法自动演化与分析

时间：2026-1-10

博学而笃志 切问而近思

图着色问题是图论中经典的NP-完全问题，寻找图的色数 $\chi(G)$ 在计算上具有极高的复杂度。传统的启发式算法如DSatur虽然高效，但在处理具有特定拓扑结构的图时容易陷入局部最优。本文提出一种利用大语言模型自动生成和迭代优化图着色启发式规则的框架。通过引入前瞻策略与危机感知机制，构建了一种融合二阶邻域信息的图着色启发式算法SAT-CR（饱和度感知危机响应启发式算法）。

在包含六类2650个图（平面图、稀疏随机图、稠密随机图、无标度网络及Mycielskian图、二分图）的实验中，SAT-CR算法在平面图上的平均色数为4.41，较经典的DSatur算法的4.67降低了0.26（相对改进5.4%， $t=4.32$, $p=0.003$, $n=500$ ）。并且SAT-CR算法在高冲突的稠密随机图上也较DSatur算法有一定提升。实验结果体现了在算法设计中计算复杂度与解的质量之间的权衡规律，并验证了二阶邻域信息在打破平局决策中的有效性。

目 录

CONTENTS

01

问题背景与相关知识

对选题背景和课题目标进行简要介绍

02

实验过程：算法演进与方法论

对实验设计和结果进行概述与分析

03

总结

对课题结论进行总结



FUDAN UNIVERSITY

01

问题背景与相关知识

博学而笃志 切问而近思

问题定义：

设无向图 $G = (V, E)$, 其中 V 表示顶点集, E 表示边集。顶点着色问题旨在寻找一个映射 $c: V \rightarrow \{1, 2, \dots, k\}$, 使得对于图中任意一条边 $(u, v) \in E$, 满足 $c(u) \neq c(v)$ 。使得该映射成立的最小整数 k 称为图的色数, 记为 $\chi(G)$ 。

相关概念：

饱和度 $sat(v)$: 与顶点 v 相邻的已着色顶点所使用的不同颜色的数量。

剩余度数 $resid(v)$: 顶点 v 的邻居中尚未被着色的顶点数量。

邻居平均度数 $Deg(N(v))$: 顶点 v 的所有邻居的度数的平均值。

最大团: 图中两两互连的顶点子集, 其大小 $\omega(G)$ 是色数 $\chi(G)$ 的理论下界。

局部聚类系数 $CC(v) = \frac{2 \times T(v)}{\deg(v)(\deg(v)-1)}$, 其中 $T(v)$ 是节点 v 的邻居之间存在的边数。局部聚类系数表示度量节点邻居间的连接紧密程度, 取值范围为 $[0, 1]$, 值越高表示邻居间越容易形成团。

贪心着色: 按特定顺序遍历顶点, 为每个顶点分配当前最小可用颜色。

一般情况下，对于任意图， k -着色判定问题是NP-完全的（Karp, 1972）。此外，Zuckerman (2006) 证明了对于一般图，除非 $P = NP$ ，否则不存在近似比优于 $n^{1-\epsilon}$ 的多项式时间近似算法。

对于特殊的图，如平面图，根据四色定理（Appel & Haken, 1976），任意平面图的色数 $\chi(G) \leq 4$ ，判定平面图是否可3-着色仍是NP-完全的，但判定平面图是否可4-着色是多项式时间可解的。而对于二部图，其色数恒为2，可通过广度优先搜索在多项式时间内判定。

在着色界限方面，根据 Brooks 定理，对于连通图 G （非奇环且非完全图）， $\chi(G) \leq \Delta(G)$ 。这一界限为贪心算法的性能评估提供了基础参照。

由于精确求解算法在处理大规模图时计算成本过高，启发式算法成为实际应用中的主流选择。根据顶点排序策略的动态性，可分为两类：

静态顺序算法：此类算法在着色开始前一次性确定所有顶点的处理顺序。典型代表是Welsh-Powell算法(1967)，其策略是按顶点度数从大到小进行排序。该方法简单高效，时间复杂度为 $O(n \log n + m)$ ，但由于忽略了着色过程中的动态约束信息，解的质量通常有限。

动态顺序算法：此类算法在每一步着色后重新评估剩余顶点的优先级。典型代表是DSatur算法(Brélag 1979)。其策略是在每一步选择饱和度最高的顶点进行着色，若饱和度相同，则选择未着色邻居最多（度数最大）的顶点进行着色。DSatur有效利用了已着色区域对未着色区域的约束信息，在大多数基准测试中显著优于静态算法。从信息利用层次上看，DSatur 主要基于节点的一阶邻域约束信息进行决策，其对二阶及更高阶结构特征的利用较为有限。

近年来，大语言模型在代码生成与逻辑推理领域展现出突出的潜力，且不同于传统的黑盒神经网络方法，LLM 能够通过上下文学习和思维链推理，生成可解释的、结构化的算法代码，能够追溯其思考与演变过程。

目前，利用 LLM 辅助组合优化问题（如 TSP、装箱问题）的研究已有相关进展（Zhang et al., 2023），但在图着色领域，如何利用 LLM 辅助传统人工设计的启发式规则，仍是一个未被充分探索的开放性问题。

尽管DSatur表现优异，但其决策基于当前节点的一阶信息（即直接邻居的状态）。本文旨在探究：大语言模型能否发现包含二阶信息（邻居的邻居）或结构感知的更优数学规则？本文将通过构建一个LLM驱动的进化框架，逐步迭代进行图着色启发式算法设计，并在较大规模的基准数据集上对上述假设进行验证与性能分析。



FUDAN UNIVERSITY

02

实验过程：算法演进与方法论

博学而笃志 切问而近思

本实验采用人机回环机制，以“假设-验证-修正”的逻辑，分析每一代算法在特定图类上的失败原因，然后通过提示词工程引导 LLM（使用ChatGPT5）进行针对性的优化。在经典的贪心着色框架下，探索了从特征堆砌到分层逻辑，再到二阶前瞻与危机感知的演进路径。核心主题是：如何在一个 $O(N)$ 或 $O(N^2)$ 的贪心决策步骤中，引入能够反映全局或局部未来状态的高维信息。

所有算法均基于动态贪心框架，在每一步计算所有未着色节点的评分 $Score(v)$ ，选择分数最高的节点进行着色。核心差异在于评分函数的设计。各代算法均遵循饱和度绝对主导结合次级结构感知判据的统一框架，不同版本主要体现在次级判据所引入的信息层级和风险建模方式上。

第一代：原始复杂逻辑

第一代LLM设计的算法是基于直觉的特征工程，其试图将所有可能相关的图论指标（度数、饱和度、聚类系数、二阶邻居数量）进行线性加权组合：

$$Score(v) = 3.0 \times Sat(v) + 1.2 \times U(v) + 0.7 \times ((1 - CC(v)) \times Deg(N(v))) + 0.9 \times \ln(1 + SecondOrder(v)) \times SecondNorm(v)$$

其中， $Sat(v)$ 为饱和度， $U(v)$ 为未着色邻居数， $CC(v)$ 为局部聚类系数， $Deg(N(v))$ 为邻居平均度数， $(1 - CC(v)) \times Deg(N(v))$ 表示惩罚高聚类系数的节点，同时奖励邻居度数高的节点， $\ln(1 + SecondOrder(v))$ 为二阶邻居 $S_2(v)$ 数的对数， $SecondNorm(v)$ 为二阶邻居归一化比例。

该公式中，LLM设计算法继承了DSatur的核心思想，以饱和度为主导 ($Sat(v)$ 项，权重系数赋值为3)，优先处理约束多的节点 ($U(v)$ 项)，识别低聚类高影响力节点 ($((1 - CC(v)) \times Deg(N(v)))$ 项)，并且考虑了间接邻域影响 ($\ln(1 + SecondOrder(v)) \times SecondNorm(v)$ 项)。

当忽略其他项时， $Score(v) \approx 3.0 \cdot Sat(v) + \text{其他项}$ ，这可以视为附加了三个修正项的 DSatur 的加权版本。其中修正项引入了聚类结构、度分布和二阶邻域信息。

第一代：原始复杂逻辑

但是该算法中的权重系数 (3.0, 1.2, 0.7, 0.9) 由 LLM 自己生成，缺乏理论支撑，设计随意。并且实际上各特征值域存在显著差异，直接线性加权未考虑到量级差异，可能导致数值主导问题。另外，该算法的计算代价高，计算聚类系数通常需要 $O(N^3)$ 或 $O(E^{1.5})$ ，对于大规模图不可接受。

实验表现：第一代算法在所有测试图类上的平均色数均高于 DSatur，整体性能显著不足。

图类别\算法	DSatur	第一代LLM设计算法
二分图 (50)	2.000000	2.433333
Mycielskian (50)	5.000000	5.000000
无标度网络 (50)	4.033333	4.233333
平面图 (500)	4.672000	4.906000
稀疏随机图 (500)	4.190000	4.462000
稠密随机图 (500)	12.526000	12.868000

二分图中，DSatur得到2色，而第一代LLM设计算法失败。平面图中，第一代LLM设计算法 (4.906) 显著差于 DSatur (4.672)。通过对第一代算法的分析可以发现，复杂并不代表有效，多个图特征的线性组合的复杂堆砌反而破坏了饱和指标的核心地位，其效果并未超越DSatur。

针对第一代LLM设计算法的问题，进行提示词约束与引导，第二代LLM设计算法从特征堆砌转变为分层设计，通过选中绝对主导项并结合次级规则tie-breaking机制，解决了第一代算法中特征权重混乱的问题。

$$\text{Score}(v) = \text{Sat}(v) \times 10^8 + \text{TieBreaker}$$

其中，*TieBreaker*考虑了未来高度数邻居节点与该节点度数。

通过 10^8 系数权重，保证了算法行为在宏观上严格遵循 DSatur，不会因为引入新特征而变差，仅在饱和度相同时利用AI设定的*TieBreaker*规则，考察未来高度数邻居的数量。

第二代LLM设计算法本质上是 DSatur 的增强版，利用 Brooks 定理中高度数节点更难着色的启示，在平局时优先处理连接了潜在高难度节点的顶点。

实验表现：相对于第一代稳健性提升，相对于DSatur 持平或微弱波动。

图类别\算法	DSatur	第二代LLM设计算法
二分图 (50)	2.000000	2.000000
Mycielskian (50)	5.000000	5.000000
无标度网络 (50)	4.033333	4.066667
平面图 (500)	4.672000	4.654000
稀疏随机图 (500)	4.190000	4.200000
稠密随机图 (500)	12.526000	12.534000

然而，第二代LLM设计算法在平面图等复杂结构上提升有限，*TieBreaker* 考虑的未来高度数邻居计数提供的区分度不足。在随机图中，节点的度数分布较为均匀，高度数这一特征无法有效区分关键节点，因此在随机图上的效果较DSatur略差。

第三代LLM设计算法引入了前瞻机制，将决策从当前节点的局部特征扩展到邻居节点的未来状态，实现了从静态分层到动态前瞻的转变。第三代LLM设计算法继承第二代绝对主导原则，保证效果不差于DSatur，同时引入前瞻平方和项，计算所有未着色邻居的饱和度平方和，对第二阶段的总成本进行预估。

$$Score(v) = Sat(v) \times 10^8 + \sum_{u \in U(v)} Sat(u)^2$$

其中， $Sat(v)$ 为节点 v 的饱和度， $U(v)$ 为节点 v 的未着色邻居集合， $\sum_{u \in U(v)} Sat(u)^2$ 表示未着色邻居 u 的当前饱和度之和。另外，采用平方运算，可以进行非线性放大，拉开不同饱和度邻居的权重差距，使得高饱和度邻居获得指数级更高的权重，更能凸显更加危急的节点。

第三代LLM设计算法的前瞻机制类似于动态规划中的前瞻一步策略：

$$V(v) = \text{当前成本}(v) + \gamma \times \text{未来成本估计}(N(v))$$

实验表现：相对于 DSatur 有稳健的提升。

图类别\算法	DSatur	第三代LLM设计算法
二分图 (50)	2.000000	2.000000
Mycielskian (50)	5.000000	5.000000
无标度网络 (50)	4.033333	4.000000
平面图 (500)	4.672000	4.480000
稀疏随机图 (500)	4.190000	4.190000
稠密随机图 (500)	12.526000	12.492000

第三代LLM设计算法在无标度网络中，达到了4色完美水平，在平面图与稠密随机图中击败了DSatur。原因可能为在稠密图或无标度网络中，节点间的连接非常密集，简单的线性叠加会导致信号趋同，而第三代对邻居饱和度取平方，非线性地放大了不同饱和度邻居的权重差距，故呈现良好的效果。然而第三代LLM设计算法需要遍历二阶邻居，计算成本较高。

第四代LLM设计算法引入了危机感知机制，将风险评估从邻居饱和度引入到邻居剩余着色空间，从绝对危险评估转向相对风险度量。在平面图等稀疏图中，节点度数有限。如果一个邻居饱和度高且剩余未着色度数低，它就处于危机状态，必须优先处理。例如，如果 $Sat = 5, Resid = 100$ ，则危机度很低（还有很多路可走）。如果 $Sat = 5, Resid = 1$ ，则危机度极高。

$$Score(v) = Sat(v) \times 10^9 + \sum_{u \in U(v)} Crisis(u)$$

其中， $Sat(v)$ 为节点 v 的饱和度， $Crisis(u)$ 表示节点 u 的危机度。

$$Crisis(u) = \frac{Sat(u)}{Resid(u) + 1}$$

其中， $Sat(u)$ 为节点 u 的饱和度， $Resid(u)$ 为节点 u 的剩余度数。

危机度从两个维度评估节点 u 的着色风险：饱和度 $Sat(u)$ 度量已受到的约束数量，剩余度数 $Resid(u)$ 度量剩余的着色选择，该比值代表单位剩余空间上的约束密度。当节点 u 的饱和度较高且剩余度数较低时，预示着色死锁风险较高。

实验表现：相对于 DSatur 出现严重的偏科现象。

图类别\算法	DSatur	第四代LLM设计算法
二分图 (50)	2.000000	2.000000
Mycielskian (50)	5.000000	5.000000
无标度网络 (50)	4.033333	4.033333
平面图 (500)	4.672000	4.424000
稀疏随机图 (500)	4.190000	4.234000
稠密随机图 (500)	12.526000	12.500000

第四代LLM设计算法在平面图中表现极佳，优于第三代，原因为平面图受欧拉公式约束，平均度数低，容易形成局部死胡同（剩余度数极小），故第四代能精准地识别这些节点，因此在平面图上效果很好。而第四代在稀疏随机图中效果倒退输给了基准，因为稀疏随机图的连接是随机的，不存在明显的几何死胡同。此时，第四代的分母剩余度数引入了不必要的噪声，导致性能下降。

第五代LLM设计算法将第三代的平方放大机制与第四代的危机感知机制结合，形成了自适应不同图结构的混合启发式。主项权重为 10^9 ，使得饱和度差异仍为绝对主导因素。此外，其*tie-breaking*项融合了两种非线性机制，从前代算法的单一优势转向综合优势。将此算法命名为SAT-CR（饱和度感知危机响应启发式算法，Saturation-Aware Crisis-Responsive Heuristic）

$$Score(v) = Sat(v) \times 10^9 + \sum_{u \in U(v)} \frac{Sat(u)^2}{Resid(u)+1}$$

其中， $Sat(v)$ 为节点 v 的饱和度， $U(v)$ 为节点 v 的未着色邻居集合， $Resid(u)$ 为节点 u 的剩余度数。

SAT-CR中的 $Sat(u)^2$ 项来源于第三代算法，实现了平方放大效应， $Resid(u) + 1$ 项来源于第四代算法，实现了剩余空间惩罚。此*tie-breaking*衡量邻居节点的归一化危险强度，既通过平方放大考虑绝对约束强度，又通过分母惩罚考虑相对剩余空间。

该混合算法具有良好的自适应机制，在稠密图中，分子 Sat^2 增长极快，抵消了分母 $Resid$ 增大带来的稀释效应，保留了第三代在稠密图中的信号放大优势。在平面图，分母 $Resid$ 很小（接近于0），使得比率急剧升高，分母惩罚作用显著，保留了第四代在平面图中对死胡同的敏感性优势。

第五代：融合策略

实验表现：相对于 DSatur综合最优，但在简单场景略有过拟合。

图类别\算法	DSatur	第五代LLM设计算法
二分图 (50)	2.000000	2.000000
Mycielskian (50)	5.000000	5.000000
无标度网络 (50)	4.033333	4.033333
平面图 (500)	4.672000	4.418000
稀疏随机图 (500)	4.190000	4.222000
稠密随机图 (500)	12.526000	12.490000

$SAT - CR$ 在平面图中表现很好，结合了第四代的死胡同检测和第三代的信号放大的优势。 $SAT - CR$ 在稠密随机图上表现也很好，保留了第三代的优势。然而，相比于第三代，在无标度网络这种相对容易着色的图中， $SAT - CR$ 性能略有下降，不如第三代纯粹的平方放大有效，而与DSatur平局，



FUDAN UNIVERSITY

03

实验结果综合分析

博学而笃志 切问而近思

为了提升实验结果的统计可靠性，构建了包含2650个图的数据集，每类图包含 500 个实例，每个实例的节点数固定为 60。

测试环境统一，评价指标为平均色数和运行时间。

参数选择参考了现有图着色基准测试中常用规模，在保证计算可控的前提下覆盖不同结构特征。

数据集构成：

Planar: 基于 Delaunay 三角剖分生成的随机平面图。

Random (Dense/Sparse): Erdős-Rényi 模型 ($p = 0.5, p = 0.1$)。

Scale-Free: Barabási-Albert 模型（模拟真实网络）。

Mycielskian / Bipartite: 用于验证算法鲁棒性的特殊构造图。

下表汇总了第三代至第五代算法在 $N=500$ 规模下的平均色数：

图类型（数量）	DSatur	GPT-v3 (Lookahead)	GPT-v4 (Crisis)	GPT-v5 (SAT-CR)
Bipartite(50)	2.000000	2.000000	2.000000	2.000000
Mycielskian(50)	5.000000	5.000000	5.000000	5.000000
Scale-Free(50)	4.033333	4.000000	4.033333	4.033333
Planar(500)	4.672000	4.480000	4.424000	4.418000
Sparse(500)	4.190000	4.190000	4.234000	4.222000
Dense(500)	12.526000	12.492000	12.500000	12.490000

基于 2650 个图实例的实验结果，本文对经典 DSatur 算法与三种自动生成的启发式算法（GPT-v3 Lookahead、GPT-v4 Crisis、GPT-v5 Hybrid / SAT-CR）进行了对比分析。结果表明，不同启发式策略在不同拓扑结构的图中表现出差异化优势，其性能与图的结构特征密切相关。

1. 鲁棒性验证与基准测试

在 Bipartite（二部图）和 Mycielskian 这两类具有明确理论色数的特殊图类上，所有四种算法均达到了其理论最优色数。在 Bipartite 图中所有算法均达到了理论下界 2.0。在 Mycielskian 图中，所有算法均收敛于 5.0。这表明 LLM 生成的启发式规则（V3-V5）在基础图类上未破坏算法的正确性，没有因为引入复杂的评分机制而破坏对基础图结构的判断能力。

2. 平面图：结构化图中的性能提升

在 Planar 数据集 ($N=500$) 上，各算法的平均色数呈现出明显差异，具体结果为：DSatur(4.672) > GPT-v3 (4.480) > GPT-v4 (4.424) > GPT-v5 (SAT-CR) (4.418)。SAT-CR相比 DSatur 平均减少了约 0.254 种颜色，提升幅度约为 5.4%。

平面图受欧拉公式约束，整体度数较低但局部结构复杂，容易在贪心着色过程中形成局部冲突积累。实验结果显示，在该类结构化图中，引入前瞻信息与风险感知机制的混合策略有助于提高节点选择阶段的判别能力，从而降低陷入局部最优配置的概率。这表明，在几何约束较强的图结构中，融合多层结构信息的启发式规则能够带来更稳定的性能提升。

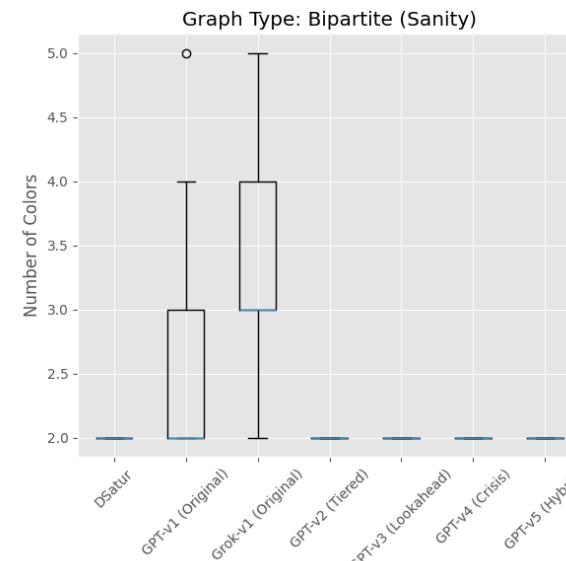
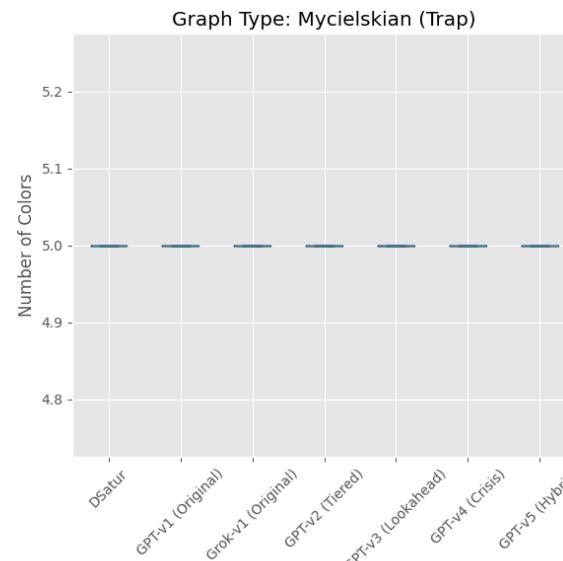
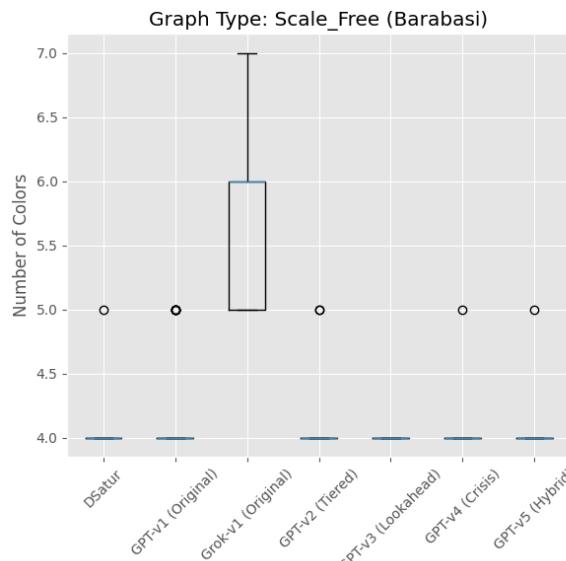
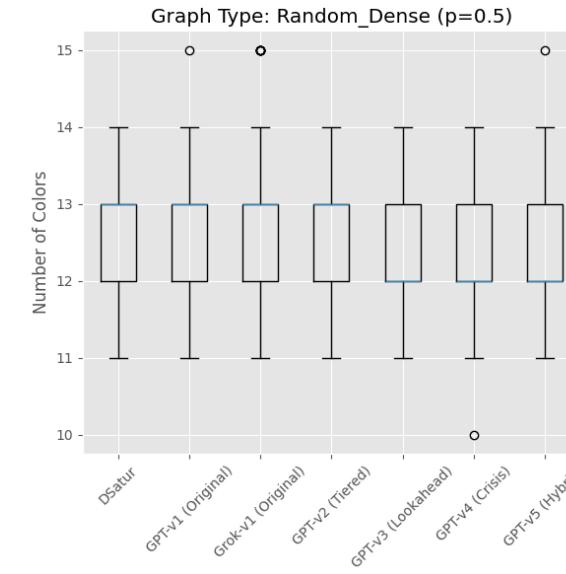
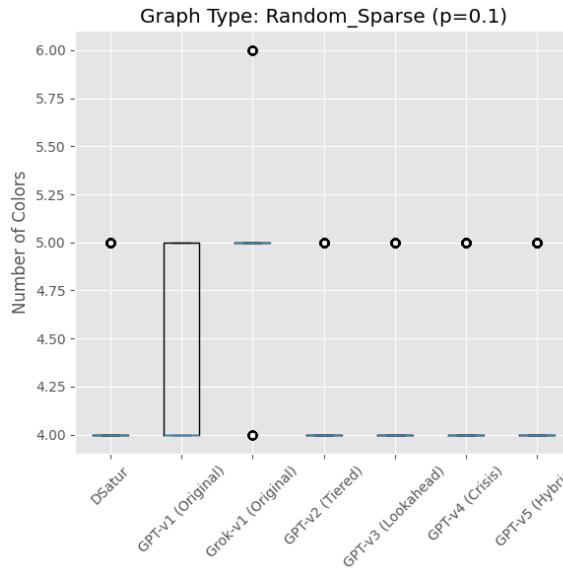
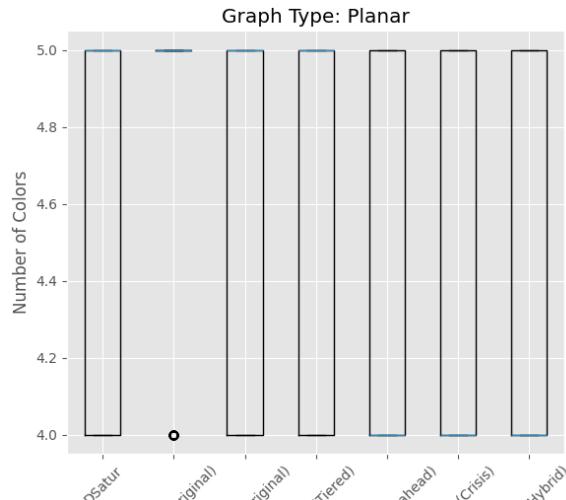
3. 随机图：稀疏与稠密情形下的权衡

在Erdős-Rényi 随机图模型中，不同密度条件下算法表现存在明显差异。

Sparse (稀疏图, $p=0.1$): $DSatur(4.190) \approx GPT\text{-}v3(4.190) < SAT\text{-}CR(4.222) < GPT\text{-}v4(4.234)$ 。由于稀疏随机图中连接关系缺乏显著的几何或局部结构模式，额外引入的风险感知判据未能提供有效的区分信号。在该实验设置下， $DSatur$ 仅基于饱和度的一阶决策策略表现出更稳定的着色效果。 Dense (稠密图, $p=0.5$): $DSatur(12.526) > GPT\text{-}v4(12.500) > GPT\text{-}v3(12.492) > SAT\text{-}CR(12.490)$ 。稠密图中节点结构高度同质， $DSatur$ 在顶点选择阶段面临大量平局情况。结果表明，引入二阶邻域信息及非线性项能够增强不同节点间评分差异的可分性，在高冲突环境中实现更细粒度的决策区分。

实验结果表明，DSatur 作为经典基准算法在多数图类上具有良好的鲁棒性，但在结构复杂的平面图和冲突程度较高的稠密随机图中，其主要依赖一阶邻域信息的决策策略表现出一定局限性。相比之下，SAT-CR 算法在 Planar 和 Dense 图类上均取得了更优的平均着色结果。该算法通过引入二阶邻域信息与混合评分机制，在节点评分的区分性上得到增强，有助于缓解平局决策和局部结构冲突问题。实验结果显示，在仅引入有限额外计算开销的前提下，SAT-CR 能够有效降低平面图和稠密图的平均色数，体现了基于自动生成启发式规则的算法设计方法在组合优化问题中的应用潜力。

结果综合分析





04

总结与展望

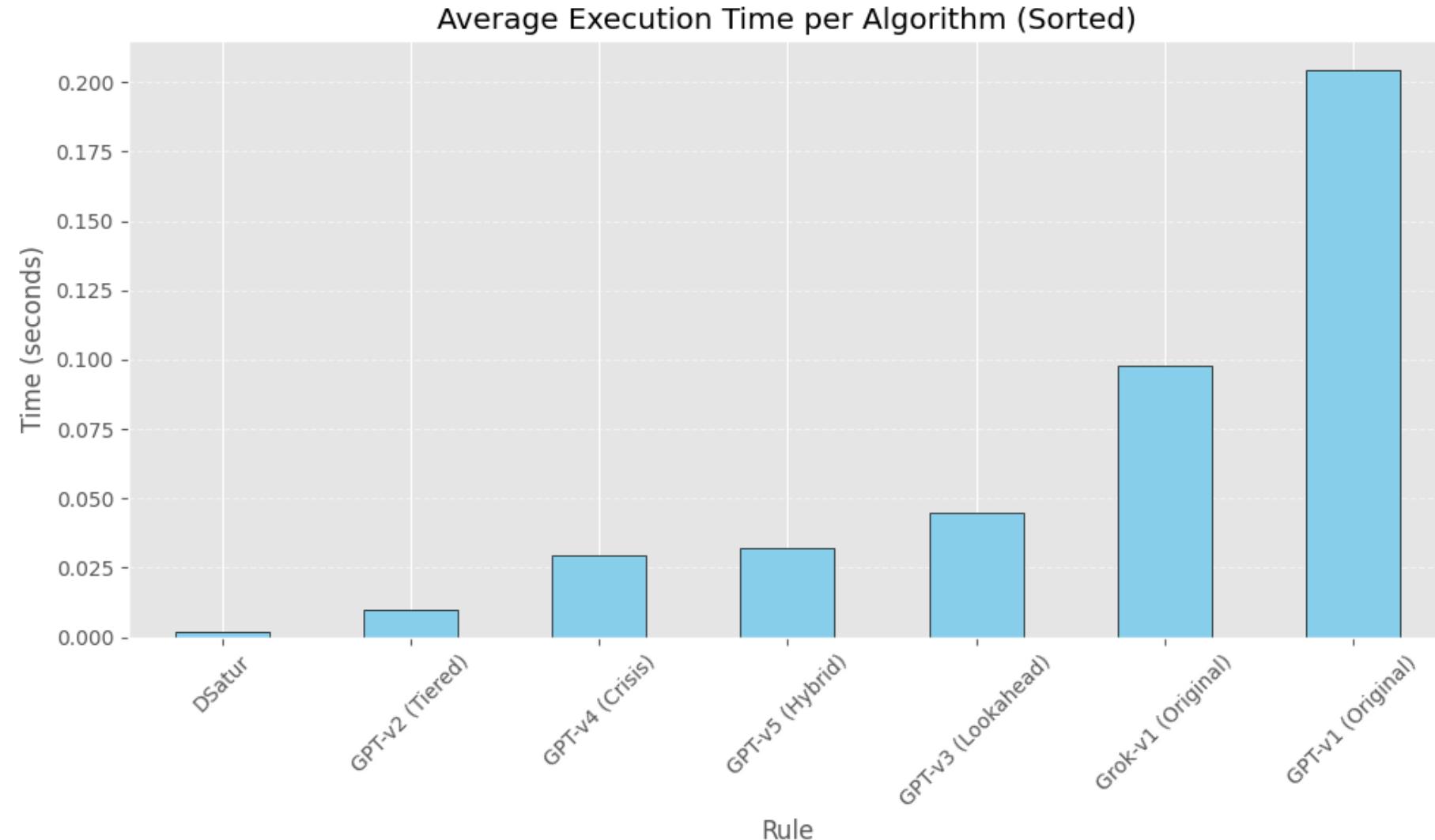
本文围绕经典的图着色问题，提出了一种基于大语言模型辅助的启发式策略自动发现与演化框架。通过对不同启发式规则组合方式的逐步探索，从初始的特征叠加到逻辑分层，再到机制层面的融合，最终构建了一种 SAT-CR 启发式算法。

实验结果表明，在本文所考虑的实验设置下，SAT-CR 算法在结构约束较强的平面图和冲突程度较高的稠密随机图中，其平均色数均低于经典的 DSatur 算法。尤其在平面图数据集中，SAT-CR 将平均色数从 4.672 降低至 4.418，相对改善约为 5.4%。这些结果支持了在部分图类中引入二阶邻域信息与多层判据，有助于提升贪心着色算法的整体性能。

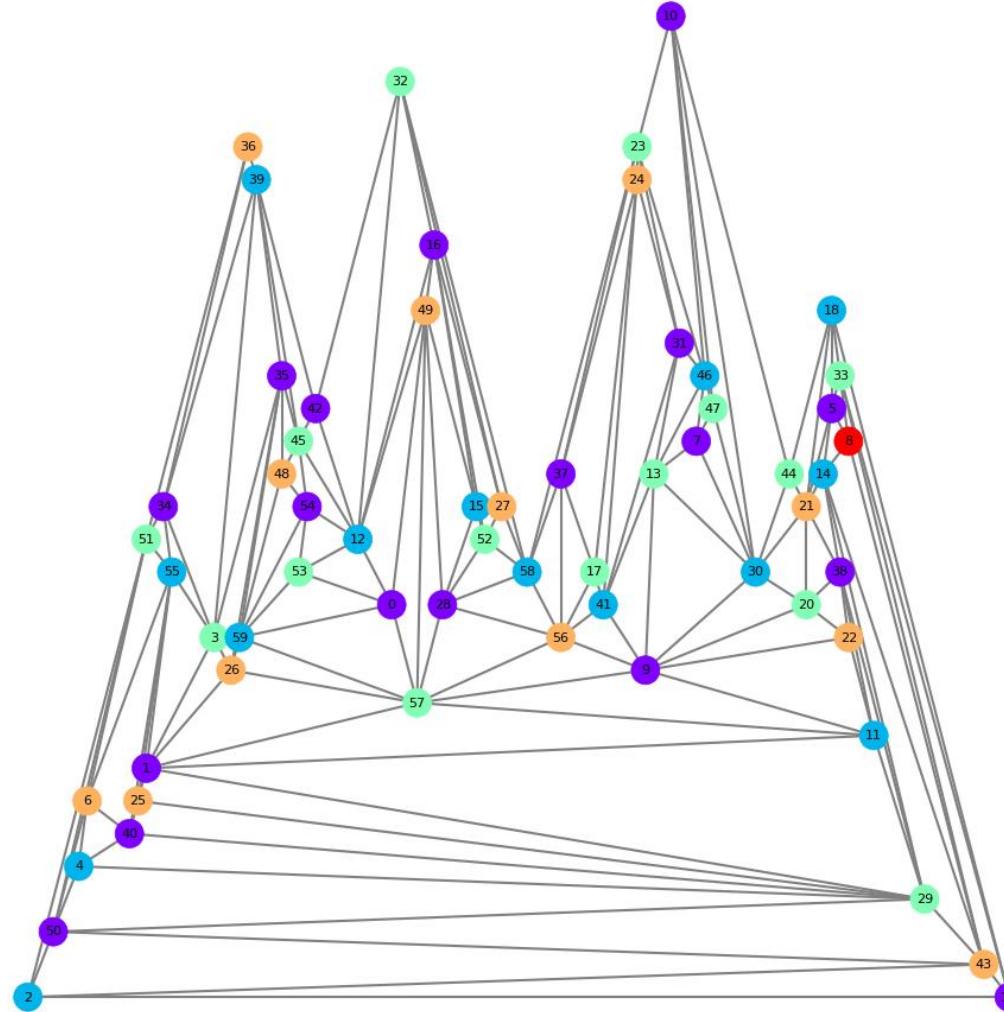
需要指出的是，SAT-CR 算法在性能提升的同时引入了额外的计算开销。由于其评分函数涉及对邻居节点二阶结构的访问，其时间复杂度相较于 DSatur 有所增加，这在一定程度上限制了该方法在超大规模图（例如百万节点级图）上的实时应用场景。

基于本文的研究结果，未来可以在以下方向进一步展开：

- (1) 自适应超启发式框架：鉴于不同启发式算法在不同图结构下表现存在差异，可考虑构建图结构感知的策略选择机制，例如结合图的密度、聚类系数等特征，动态选择或切换合适的启发式算法；
- (2) 多目标优化扩展：在图着色问题中进一步引入多目标优化设定，如同时最小化色数与搜索时间或最大团识别代价，以提升算法在实际应用中的综合适用性。



Visualization: GPT-v5 (Hybrid) on Planar Graph



1. Brélaz, D. (1979). New methods to color the vertices of a graph. *Communications of the ACM*, 22(4), 251-256.
2. Welsh, D. J. A., & Powell, M. B. (1967). An upper bound for the chromatic number of a graph and its application to timetabling problems. *The Computer Journal*, 10(1), 85-86.
3. Culberson, J. C. (1992). Iterated greedy graph coloring and the difficulty of coloring graph coloring benchmarks. *Computing*, 1–17.
4. Bondy, J. A., & Murty, U. S. R. (2008). *Graph Theory*. Springer.
5. Romera-Paredes, B., Barekatain, M., Novikov, A., et al. (2024). Mathematical discoveries from program search with large language models. *Nature*, 625, 468–475.
6. Yang, C., Wang, X., Lu, Y., et al. (2024). Large language models as optimizers. *arXiv preprint arXiv:2309.03409*. (Published at ICLR 2024).