

非常に難しいグラフ3彩色問題の組織的生成法と考察

西原 清一 水野 一徳

筑波大学 電子・情報工学系

〒305-8573 つくば市天王台1-1-1

要旨：解を求めるのに非常に計算手間のかかるグラフ3彩色問題を組織的に生成する方法を提案する。ランダムに生成した中から困難な問題を選び取るという従来のアプローチと異なり、困難な問題を確実に生成することを目指すものである。生成されるグラフ3彩色問題は、従来、指摘されてきたように、解探索が非常に困難となる第2相転移の領域に存在し、平均次数やグラフの均質性などの条件に合致している。また、頂点数の指数オーダーの複雑性を有することを実験的に示す。

Systematic Generation of Exceptionally Hard Instances for Graph 3-Colorability with Considerations

Seiichi Nishihara

nishihar@is.tsukuba.ac.jp

Kazunori Mizuno

mizuno@algor.is.tsukuba.ac.jp

Institute of Information Sciences and Electronics, University of Tsukuba

1-1-1 Ten-no-dai, Tsukuba, Ibaraki 305-8573, Japan

Abstract: Graph coloring problem(COL) is an interesting NP-complete problem, on which many ideas and conjectures have been proposed to explain the mechanism that may cause the phase transition phenomena and make COL instances exceptionally hard. We propose an ordered method to generate exceptionally hard 3COLs by embedding rigid pairs, similar to frozen pairs, one after another. Experimental data resulted by using Brélaz heuristics show that our method always produces very hard problems, whose complexity is like an exponential order of the number of vertices.

Keywords: graph coloring, phase transition, NP-complete, complexity, hard problems, heuristics

1 まえがき

グラフ彩色問題 (COL) や命題論理式の充足問題 (SAT) は、組合せ探索を含む代表的な問題であり、計算複雑さや解探索アルゴリズムの観点から研究されている。NP 完全であるこれらの問題は、ヒューリスティクスの研究の材料としても興味深い [1]。COL の計算複雑さについては、多くの研究がある [1-11]。とくに、例外的に困難な問題の存在条件とその原因さらにその生成法について、Vlasie[9,10] の 3-path,

Mammen ら [7] の極小非可解部分問題, Culberson ら [3] の frozen development など、多くの有効なパラメータが提案されている。

本稿では、色数 3 の COL について、非常に計算手間のかかる問題 (インスタンス) を組織的ないしは系統的に生成する手法について報告する。これは、同じ色を持つ 2 頂点を含む微小グラフ (rigid subgraph[8]) の埋め込み操作を繰り返すもので、生成されたグラフは頂点数の指数オーダーの複雑さを持つことを実験的に示す。

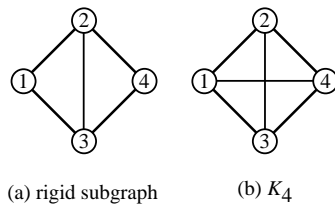


図 1: 特徴的なグラフ構造

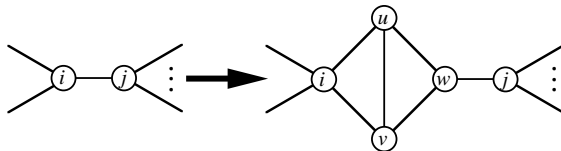


図 2: 埋め込み操作 $embed(i, j)$ (ただし, $deg(i) \leq 3$)

2 生成アルゴリズム

2.1 難しいグラフ 3 彩色問題の性質

グラフ彩色問題は制約充足問題 (CSP, Constraint Satisfaction Problem) の一種と見ることができる。実際、2 頂点 i, j 間に辺がある場合は、頂点 (すなわち変数) i, j には同じ色 (すなわち値) を与えてはならないという制約が存在することを意味する。色の個数を 3 に限定したグラフ 3 彩色問題 (3COL) は NP 完全であり、多くの研究報告がある [1-4,6-10]。計算複雑さの観点から、ランダムに生成した CSP 集合について調べると、各問題を解くのに要する手間は、頂点数を固定したとき、制約密度が粗から密へ変化するにつれて易-難-易 (easy-hard-easy) のパターンを示すことが知られている [6,7]。計算量の中央値が最も大きく (つまり難に) なるのは、可解 / 非可解の問題が共に 50% の確率で生成される極端に狭い制約密度領域の付近である [6,8]。この現象を第 1PT (相転移, phase transition) という。

一方、第 1PT よりも小さな制約密度のところにおいて、「例外的に非常に手間のかかる問題 (EHI)」が発生する傾向があることも知られている。これを第 2PT という。3COL では (少なくとも頂点数 $n \leq 100$ においては)、頂点の平均次数 γ が 4.5~4.6 および 3.3~3.4 付近においてそれぞれ第 1PT と第 2PT が観察されるという実験結果が報告されている [3,6,8]。ここで、制約密度とは、変数に対する (2 項) 制約の割合を指し、 γ の 2 分の 1 に等しい。また、ランダムな 3COL の実験では、第 2PT における EHI は殆

procedure graph-generator(k)

begin

for $w := 1$ to k **do**

 choose randomly an edge (i, j) where $deg(i) \leq 3$;

 embed (i, j) ;

end for;

end.

図 3: 3COL のグラフを生成する手続き

どが非可解な問題である [6]。Vlasie[10] は、頂点次数の分散が小さいいわゆる regular なグラフが困難となる傾向があると指摘している。

2.2 生成法とそのアルゴリズム

図 1(a) が 3COL の部分グラフとして含まれているとき、頂点 1 と 4 は同じ色となる (以下、3COL の具体例 (instance) も 3COL と略記する)。「同じ色」という制約の下にある 2 頂点を「rigid pair」と呼ぶ。辺で結ばれていない 2 頂点の間にこのような強い制約が存在することは、辺によって陽に示された 2 頂点間の制約が「異なる色を割り当てること」という弱い制約であることを考慮すると、興味深い性質である。図 1(b) は完全グラフ K_4 であり、3 色では解けない問題である。しかし、任意の辺を削除すると同図 (a) と同型になり、可解となる。この K_4 のように、任意の 1 辺を除去すると可解となるようなグラフを極小非可解グラフ (mug, minimal unsolvable graph) と呼ぶことにする。

いま 2 頂点 i, j 間に辺 (i, j) が存在し、かつ i の次数 $deg(i)$ が 3 以下であるとしよう。ここで、新たに頂点 u, v, w からなる K_3 を作り、3 辺 $(i, u), (i, v), (w, j)$ を追加し、従来の辺 (i, j) を削除する操作 $embed(i, j)$ を導入しよう。これは図 2 に示すように、辺 (i, j) の頂点 i に接する場所に K_3 を埋め込む操作である。一般に、 $embed(i, j)$ と $embed(j, i)$ は異なる処理である。

図 3 は、与えられた初期グラフに対して embed 操作を繰り返しながら大きいグラフを生成する手続き graph-generator である。われわれの方法では、初期グラフとして頂点数が最も少ない mug である K_4 を入力するものとする。引数 k は、埋め込み操作の回数である。図中、 $embed(i, j)$ は、図 2 に示した埋め

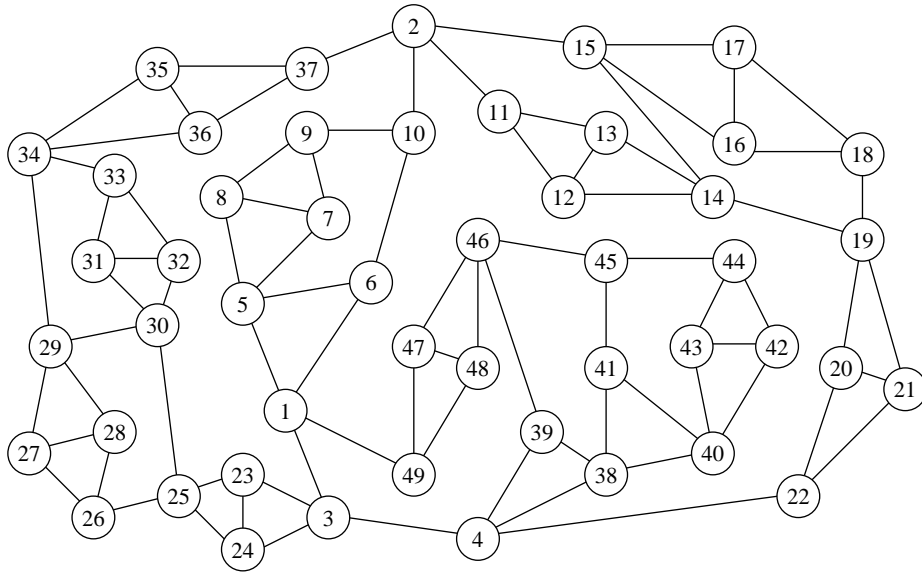


図 4: 生成した極小非可解グラフ mug の例 ($n = 49, m = 81$)

込み操作である。

初期グラフ K_4 については，任意の頂点 i について $\deg(i) = 3$ である．また， $\text{embed}(i, j)$ を施すと，頂点 i の次数は 1 増えてしまうが，新たに次数 3 の頂点が 3 個 (図 2 の u, v, w) 追加されるので，次数が 3 以下の頂点が尽きることはない．したがって，埋め込み回数 k は任意に指定することができる．このため，グラフは (頂点数に関して，間断はあるものの) 任意に大きいものを生成することができる．一般に， embed を 1 回施すたびに，グラフの頂点数 n と辺数 m はそれぞれ 3, 5 ずつ増える．また，頂点 i の次数は 4 に増えるが，既述のように，次数 3 の頂点 u, v, w が追加されるので，都合，次数 3 の頂点は 2 個増えることになる．次数 3 以外の頂点の次数は全て 4 である．開始グラフ K_4 は $n = 4, m = 6$ であるので， embed を k 回施すと，頂点数は $4 + 3k$ ，辺数は $6 + 5k$ となる．また，次数 3 の頂点は $4 + 2k$ 個，4 の頂点は k 個となる．

図 4 に， k を 15 としたときに得られるグラフの例を示す．頂点数は 49，辺数は 81 である．また，頂点の次数は 3 または 4 のみであり，それぞれの頂点数は 34 個，15 個である．

2.3 諸性質

前節で提案したアルゴリズムおよび生成される 3COL の性質について述べる．

まずアルゴリズムによって多様なグラフができることを許しているのは， embed を施すべき場所，すなわち頂点および，それに接続する辺をランダムに選べるようにしているからである．実際，まず次数 3 の頂点をランダムに選び，さらに，その接続 3 辺のうち一つをランダムに選ぶ．いずれにしても， embed を 1 回施すと，その頂点は次数 4 になり選択枝は 3 減るが，新たに追加された 3 頂点についてそれぞれ 3 つの選択枝が追加されるので，結局，選択枝は 6 増えることになる．初期の K_4 には embed を施すことのできる候補が 12カ所あるので， k 回の埋め込みの後では， $12 + 6k$ になる．

次に，得られる 3COL について述べる．頂点の次数は 3 または 4 の 2 種類のみである．また， k が大きくなると，頂点数 n も正比例して大きくなり，それにつれて平均次数 γ は， $k \rightarrow \infty : \gamma = 2(6 + 5k) / (4 + 3k) \rightarrow 3.33 \dots$ となる．本稿の生成方法は組織的であるにもかかわらず，結果的には，多くの文献 [6,10,11] で指摘されている EHI の性質：(1) 頂点次数が正則 (regular) に近いこと，(2) 第 2PT の領域付近に発生しやすいこと，に合致している．とくに，われわれの方法で生成する 3COL の γ が，シミュレーション実験において第 2PT が観測される 3.3~3.4 に結果的に一致するという事は興味深い．

われわれの方法により生成されるグラフは次の性質を持つ．

[定理] K_4 を初期グラフとして，手続き graph-generator

表 1: 実験結果 (非可解グラフ mug の探索コスト) (斜体は実験未完了)

ケース	n	問題数	γ	max	mean	min
1	10	50/50	3.20	25	18 (71%)	16 (64%)
2	19	50/50	3.26	297	199 (67%)	92 (31%)
3	28	50/50	3.29	2,930	1,784 (61%)	800 (27%)
4	40	50/50	3.30	92,031	33,997 (37%)	7,696 (8.3%)
5	49	50/50	3.31	746,396	273,900 (37%)	31,014 (4.2%)
6	58	50/50	3.31	9,064,868	2,125,440 (23%)	405,280 (4.5%)
7	70	50/50	3.31	174,649,636	42,082,268 (24%)	1,410,310 (0.8%)
8	79	20/20	3.32	1,734,145,753	466,854,997 (27%)	45,846,569 (2.6%)
9	88	10/10	3.32	5,472,177,992	1,964,045,777 (36%)	197,706,601 (3.6%)
10	100	4/10	3.32	10,702,506,506	5,553,271,146 (52%)	2,920,643,478 (27%)

() 内は max との比

によって生成される 3COL は mug である。(証明) embed は、従来から存在していた頂点の色割り当てには影響しないから、詳細省略。■

ところで、Mammenら [7] の実験によると、第 2PT 付近において見出される EHI は非可解である場合が多く、その困難さは、部分問題として含んでいる mug のうち最小のものに関連していることが指摘されている。この性質が普遍的なものであるならば、graph-generator で生成する 3COL は、それ自身が全体で一つの mug であるので、EHI であることになる。

3 実験

3.1 問題生成と Bréaz ヒューリスティクス

K_4 を初期グラフとして、graph-generator(k) の引数 k を $2(n=10), 5(19), 8(28), 12(40), 15(49), 18(58), 22(70), 25(79), 28(88), 32(100)$ の 10 個のケースに設定して、各々 50 題ずつ 3COL を生成した。カッコ内は、頂点の個数を示しており、10 から 100 までなるべく 10 ずつ増えるようにとってある。また、生成した各 3COL について、次数 4 の頂点同士を繋ぐ任意の辺を一つ除去することによって可解な問題を生成した。ただし、 $n = 10, 19$ のケースではそのような辺が存在しない mug もあった。実験では、上のようにして生成した可解な問題についても探索コストを調べた。実験は、PC/AT 互換機 (CPU: Pentium III 1GHz, メモリ: 512MByte) を使用し、プログラ

ムは C 言語で記述した。解探索アルゴリズムとしては、他の研究との比較ができる標準的なものとして、Bréaz ヒューリスティクス [1] を組み込んだ木探索法を用いた。探索コストとしては、頂点への色の割り付けを行った回数を計数した。なお、最初の 2 色へのバックトラックは行わないようにした。また、Vlasie[9] のランダム生成による方法についても追試実験を行い、われわれの方法と比較した。

3.2 実験結果

10 通りの n についてそれぞれ非可解な問題 (mug) および可解な問題 (sol) の 2 通りに分けて実験を行った。 $n \leq 70$ の 7 ケースについては生成した全問題について解探索を行なった。 $n = 79, 88$ の 2 ケースについてはそれぞれ 20, 10 問題を無作為に選んだ。 $n = 100$ については設定時間内に処理が終了しなかった。計測データとしては、最大コスト (max)、平均コスト (mean)、最小コスト (min) を求めた。表 1、表 2 はそれぞれ mug, sol の場合の実験結果である。本実験に関する限り、mug の場合、max に対する mean の値は約 20 ~ 70%、同じく min は 1 ~ 60% となっている。また、mug に対する sol のコストは、max, mean に関して、約 40 ~ 130% となっている。

これらの実験結果をグラフ表示したものを図 5 に示す。同図には、文献 [8] のデータも併記してある。非可解 (mug) の max と mean については、ほぼ n の指数オーダの結果が示されている。mug_min についてもほぼ同様の傾向が見られることは注目に値する。

表 2: 実験結果 (可解グラフ sol の探索コスト) (斜体は実験未完了)

ケース	n	問題数	γ	max	mean	min
1	10	17/17	3.00	33 (132%)	17 (93%) (50%)	10 (63%) (30%)
2	19	48/48	3.16	370 (125%)	125 (63%) (34%)	19 (21%) (5.1%)
3	28	50/50	3.21	3,308 (113%)	985 (55%) (30%)	99 (12%) (3.0%)
4	40	50/50	3.25	100,529 (109%)	15,830 (47%) (16%)	461 (6.0%) (0.45%)
5	49	50/50	3.27	690,953 (93%)	138,713 (51%) (20%)	619 (2.0%) (0.09%)
6	58	50/50	3.28	3,924,568 (43%)	921,541 (43%) (23%)	254 (0.06%) (0.00%)
7	70	50/50	3.29	82,697,475 (47%)	16,958,787 (40%) (21%)	228,713 (16%) (0.28%)
8	79	20/20	3.29	1,529,132,138 (88%)	321,077,377 (69%) (21%)	1,863,348 (4.1%) (0.12%)
9	88	10/10	3.30	2,307,095,600 (42%)	913,053,941 (46%) (40%)	4,863,921 (2.5%) (0.21%)
10	100	5/10	3.30	3,258,872,519 (30%)	1,574,732,751 (28%) (48%)	104,002,630 (3.6%) (3.2%)

() 内は mug との比

() 内は max との比

西原 [8] や Vlasie[9] においては, EHI は試行錯誤的に発生するインスタンスであったのに対して, 本方法では安定して EHI が生成されている.

われわれは Vlasie[9] の追試実験を行い, $n = 50$, $m = 80$ (辺数), $\gamma = 3.2$ の場合について 100 個の問題を生成して調べてみた. 内訳は, 非可解の問題が 17 問, 可解が 83 問であった. 表 3 に結果を示す. 表中, median は中央値である. 非可解 17 問中, コストが 10 万以上の EHI は 7 問あった. 続く第 8 番目は 2,174 と劇的に小さくなる. また, 可解な問題には EHI は含まれていない. これに対して, 提案手法では, 可解な場合でも, max, mean について EHI と同程度の困難さの sol が生成されている (表 2).

4 考察

本稿の方法によると, EHI が安定的に生成されること, また, その EHI の処理コストはほぼ頂点数の

指数オーダーとなることを実験的に示した. ランダムに生成する接近法をとっている他の研究の報告によると, EHI が発生し易い条件として, (1) 第 2PT 付近, すなわち $\gamma = 3.3 \sim 3.4$ のとき, (2) 含まれる極小非可解部分問題 (3COL では mug) の最小のものサイズが大きいこと, (3) グラフ構造が一様, 例えば正則 (regular) であること, などが指摘されている. 本方法は組織的方法であると同時に, これらの諸条件を (結果的に) 満たしているといえる.

しかし, 本方法による 3COL が, EHI である理由や, アルゴリズムに依存せずに困難であるのかどうか, などについては未知である. ただ, われわれの 3COL の mug は任意の辺を一つ除去すると可解になってしまうので, 頂点数 n が大きくなるほどその 3COL は非可解と可解の相対的に狭い境界に面している問題であるといえる. また, 準正則であるため, 対象 3COL の構造的弱点を見つけるヒューリスティクスは存在し難いのではないと思われる. なお,

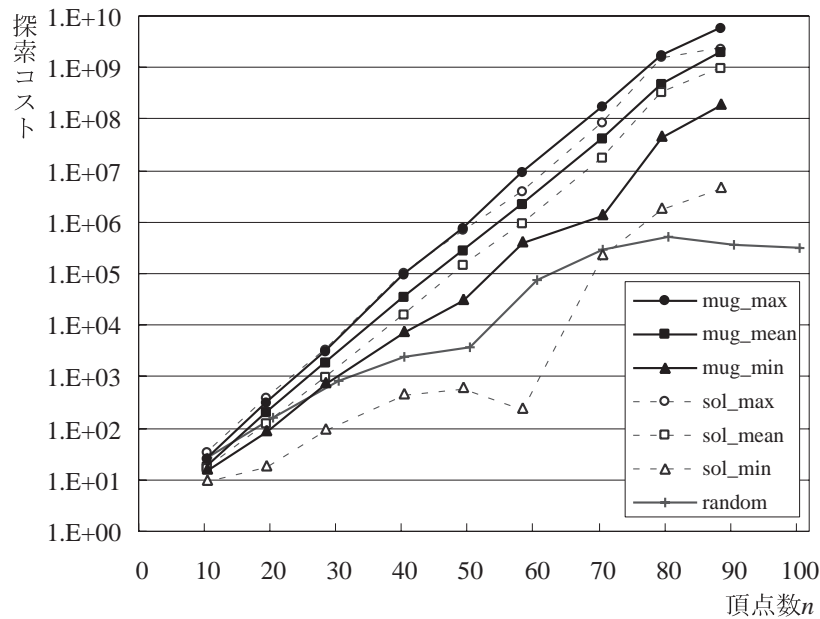


図 5: 実験結果 (random は文献 [8] のデータ)

表 3: Vlasie[9] の 100 問 ($n = 50, m = 80$)

	非可解な問題 (17 問)	可解な問題 (83 問)
max	6,490,670	1,082
mean	711,330	85
median	1382	55
min	5	50

Vlasie[9] では, EHI の存在条件として '3-path が少ないこと' を挙げており, 不安定ながら強力な EHI の生成に実験的に成功している. しかし, その理由については依然として不明である.

5 むすび

本稿では, グラフ 3 彩色問題について, 解決に非常に手間のかかる問題を組織的に生成する方法を提案し, 頂点が 100 個以下の場合について実験的に確認した. また, その解決コストは, 頂点数の指数オーダーで増加することを確認した. 今後は, K_4 以外の mug を初期グラフとしたり, 埋め込み操作を改良した, より困難な問題の生成法の開発, およびヒューリスティクスについて検討する予定である.

参考文献

- [1] D. Brélaz: New Methods to Color the Vertices of a Graph, *Comm.ACM*, 22,4(1979), pp.251-256.
- [2] P. Cheeseman, B. Kanefsky and W.M. Taylor: Where Really Hard Problems Are, *Proc. 12th IJCAI* (1991), pp.331-337.
- [3] J. Culberson and I. Gent: Frozen development in graph coloring, *Theor. Computer Sci.*, 265 (2001), pp.227-264.
- [4] D. Eppstein: Improved Algorithms for 3-Coloring, 3-Edge-Coloring, and Constraint Satisfaction, *12th ACM-SIAM Symp. Discrete Algorithms*, (2001), pp.329-337.
- [5] A. Herz, B. Jaumard and M.P. de Aragão: Local optima topology for the k -coloring problem, *Discrete Applied Mathematics*, 49 (1994), pp.257-280.
- [6] T. Hogg and C.P. Williams: The hardest constraint problems: a double phase transition, *Artif. Intell.*, 69 (1994), pp.359-377.
- [7] D.L. Mammen and T. Hogg: A New Look at Easy-Hard-Easy Pattern of Combinatorial Search Difficulty, *Jour. Artif. Intell. Research*, 7 (1997), pp.47-66.
- [8] 西原, 林元, 水野: グラフ色塗り問題の制約構造に基づく相転移現象の解析, 情処大会, 第 60 回, 3H-6 (2000), pp.2-75 ~ 2-76.
- [9] D.R. Vlasie: Systematic Generation of Very Hard Cases for Graph 3-Colorability, *Proc. 7th IEEE ICTAI* (1995), pp.114-119.
- [10] D.R. Vlasie: The Very Particular Structure of the Very Hard Instances, *Proc. 13th AAAI* (1996), pp.266-270.
- [11] C.P. Williams and T. Hogg: Exploiting the deep structure of constraint problems, *Artif. Intell.*, 70 (1994), pp.73-117.