НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Алгоритмы решения задачи о максимальной клике

аспирант второго года обучения,
стажер-исследователь Лаборатории Алгоритмов и Технологий Анализа Сетевых структур,
старший преподаватель кафедры Прикладной математики и информатики Факультета информатики, математики и компьютерных наук

Николаев А. И.

Публикации

2016

- Николаев А. И. Эффективный подход на основе машинного обучения для решения **задачи о максимальной клике** // Информационные технологии, Т. 22, № 4, 2016, С. 249-254.
- San Segundo P., Nikolaev A., Batsyn M., Pardalos P.M. Improved Infra-Chromatic Bound for Exact **Maximum Clique** Search // Informatica. 2016. Vol. 27. No. 2. P. 463-487.
- San Segundo P., Lopez A., Batsyn M. V., Nikolaev A. I., Pardalos P. M. Improved initial vertex ordering for exact **maximum clique** search // Applied Intelligence. 2016. Vol. 45. No. 3. P. 868-880.

2015

- San Segundo P., Nikolaev A., Batsyn M. Infra-chromatic bound for exact **maximum clique** search // Computers & Operations Research. 2015. Vol. 64. P. 293-303.
- Nikolaev A., Batsyn M., San Segundo P. Reusing the Same Coloring in the Child Nodes of the Search Tree for the **Maximum Clique Problem** // Lecture Notes in Computer Science. 2015. Vol. 8994. P. 275-280.

Определения

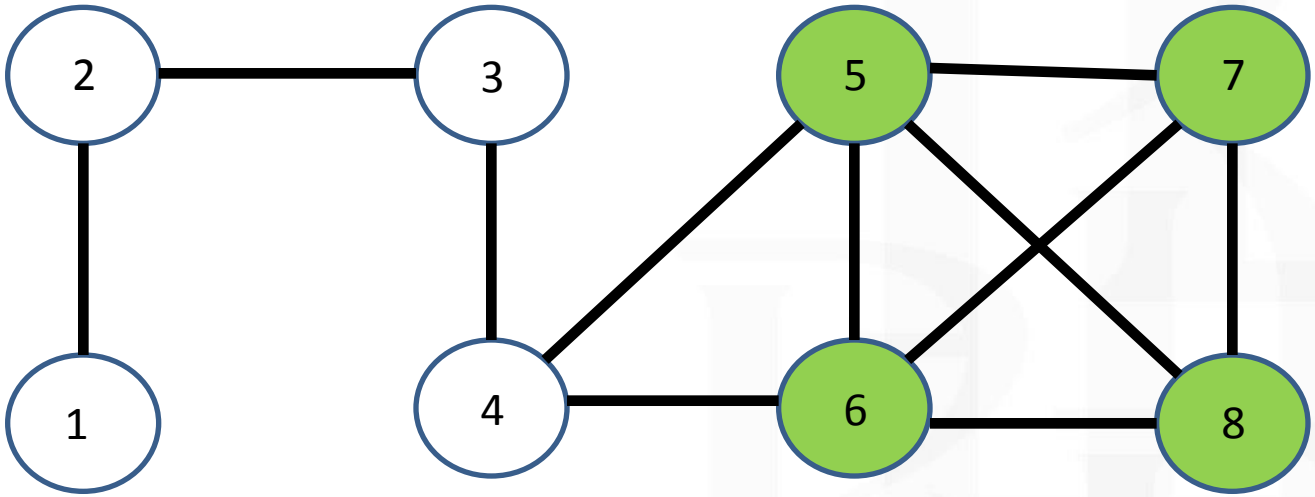
- Будем рассматривать простые (без петель и кратных ребер) неориентированные графы $G = (V, E)$.

- ***Кликой Q*** (или ***полным подграфом***) графа G называется такое подмножество его вершин, в котором любые две вершины соединены ребром.

- Клика, которая не содержится в клике большего размера, называется ***максимальной по включению*** (***maximal clique***).

Определения

- **Задача о Максимальной Клике** (**MCP**) состоит в том, чтобы для заданного графа *G* найти клику максимального размера (***maximum clique***).



Edge formulation

$$\max \sum_{i=1}^{n} x_i$$

$$x_i + x_j \leq 1, \forall (i,j) \in \bar{E}$$
$$x_i \in \{0,1\}, i = 1, \dots, n$$

$$\bar{E} = \{(i,j) | i,j \in \{1, \dots, n\}, i \neq j \text{ и } (i,j) \notin E\}$$

«Трудность» NP-трудной задачи

Type	Name	$\|V\|$	$\|E\|$	d_{max}	d_{avg}	$K(G)+1$	ω	BBMCSP ω_o	BBMCSP heur	BBMCSP search
Affiliation	*aff-orkut-user2groups*	8730857	327037420	318268	74.92	472	6	5	852	1377
Infrastructure	*inf-germany_osm*	11548845	12369181	13	2.14	4	3	3	< 0.001	< 0.001
Infrastructure	*inf-great-britain_osm*	7733822	8156517	8	2.11	4	3	3	< 0.001	< 0.001
Infrastructure	*inf-italy_osm*	6686493	7013978	9	2.10	4	3	3	< 0.001	< 0.001
Scientific comp.	*adaptive*	6815744	13624320	4	4.00	3	2	2	1.35	1.15
Scientific comp.	*channel-500x100x100-b050*	4802000	42681372	18	17.78	10	4	4	4.47	5.22
Scientific comp.	*delaunay_n22*	4194304	12582869	23	6.00	5	4	4	2.14	1.58
Scientific comp.	*delaunay_n23*	8388608	25165784	28	6.00	5	4	4	4.46	3.29
Scientific comp.	*delaunay_n24*	16777216	50331601	26	6.00	5	4	4	8.96	6.84
Scientific comp.	*hugebubbles-00020*	21198119	31790179	3	3.00	3	2	2	6.42	3.62

San Segundo P., Lopez A., Pardalos P. M. A new exact maximum clique algorithm for large and massive sparse graphs //Computers & Operations Research. – 2016. – Т. 66. – С. 81-94.

«Трудность» NP-трудной задачи

Social	*soc-friendster*	65608366	1806067135	5214	55.06	305	129	119	692	1027
Social	*soc-livejournal07*	5204176	48709773	15017	18.72	375	358	356	0.052	0.007
Social	*soc-livejournal-user-groups*	7489073	112307315	1053749	29.99	117	9	8	771	718
Social	*soc-ljournal-2008*	5363260	49514271	19432	18.46	426	400	400	0.033	0.001
Social	*soc-orkut-dir*	3072441	117185083	33313	76.28	254	51	50	38.7	55.8
Social	*soc-sinaweibo*	58655849	261321033	278489	8.91	194	44	41	662	94.8
Social (facebook)	*socfb-A-anon*	3097165	23667394	4915	15.28	75	25	24	7.18	9.01
Social (facebook)	*socfb-konect. edges*	46009640	72040814	4960	3.13	16	6	6	4.77	0.638
Social (facebook)	*socfb-uci-un*	58790782	92208195	4960	3.14	17	6	6	8.95	0.916
Web links	*web-ClueWeb09-50 m*	428136612	446534058	308477	2.09	193	56	53	267	4.77
Web links	*web-indochina-2004-all*	7414866	150984819	256425	40.72	6870	6848	6848	14.7	0.282
Web links	*web-it-2004-all*	41291318	1027474947	1326744	49.77	3225	3222	3222	3.59	0.195
Web links	*web-wikipedia_link_de*	3930109	68714064	437732	34.97	461	339	339	14.4	0.007
Web links	*web-wikipedia_link_fr*	5115915	104591689	1274642	40.89	818	332	332	253	0.115

Точные алгоритмы для решения MCP

1957 – Harary and Ross
1973 – Bron and Kerbosch
1977 – Tarjan and Trojanowski
1990 – Carraghan and Pardalos
1986, 2001 – Robson
2002 – P. R. J. Ostergard
…

Эффективные точные алгоритмы для решения MCP

2010 – MaxCLQ (Li and Quan)
2010 – MCS (Tomita et al.)
2011 – BBMCI (Segundo et al.)
2013 – IncMaxCLQ (Li et al.)
2015 – BBMCX (Segundo et al.)
2015 – RPC (Nikolaev et al.)
2016 – BBMCSP (Segundo et al.) (for large and massive sparse graphs)

Эффективные точные алгоритмы для решения MCP

Ограничение по времени: 24 часа

name	n	d	ω	Cliquer	Regin	MCR	MCQdyn	MaxCLQ$^-$	MaxCLQ	s	Rate
brock200_1	200	0.74	21	6.37	4.60	1.13	0.96	2.28	**0.67**	2.66	0.86
brock400_1	400	0.75	27	22182	4867	1137	703.5	1447	**370.84**	2.92	0.86
brock400_2	400	0.75	29	5617	3395	465.10	309.0	664.9	**178.70**	2.86	0.86
brock400_3	400	0.75	31	1667	1922	766.51	565.0	971.3	**290.06**	2.81	0.85
brock400_4	400	0.75	33	247.7	2597	409.43	320.4	605.6	**167.30**	3.15	0.85
brock800_1	800	0.65	23	-	-	10712	8821	22821	**8815**	2.92	0.80
brock800_2	800	0.65	24	-	-	9679	8125	21001	**7690**	2.77	0.81
brock800_3	800	0.65	25	26014	-	6546	5565	13559	**5285**	2.73	0.80
brock800_4	800	0.65	26	6108	-	4561	4240	9625	**3880**	2.71	0.80
MANN_a27	378	0.99	126	-	7.93	1.98	3.10	6.86	**0.66**	3.21	0.75
MANN_a45	1035	0.996	345	-	-	2931	2006	8965	**255.67**	6.30	0.91

Chu Min Li and Zhe Quan.: An efficient branch-and-bound algorithm based on maxsat for the maximum clique problem. In: Proc. of the 24th AAAI, pages 128–133, 2010.

Почему разрабатываются новые точные алгоритмы для MCP?

- Chu Min Li et al.
- Pablo San Segundo et al.
- Etsuji Tomita et al.
- Mikhail Batsyn, Panos M. Pardalos et al.
- Patrick Prosser & Ciaran McCreesh

Почему разрабатываются новые точные алгоритмы для MCP?

Chu Min Li published new research

ResearchGate [no-reply@researchgatemail.net]

Действия

Кому: Николаев Алексей Игоревич

28 февраля 2017 г. 10:00

- Для защиты конфиденциальности некоторое содержимое в данном сообщении заблокировано. При уверенности, что это сообщение пришло от надежного отправителя и при желании включить заблокированные функции, щелкните здесь.

4

Chu Min Li published this article:

Article · **February 2017**

On Minimization of the Number of Branches in Branch-and-Bound Algorithms for the Maximum C...

Computers & Operations Research

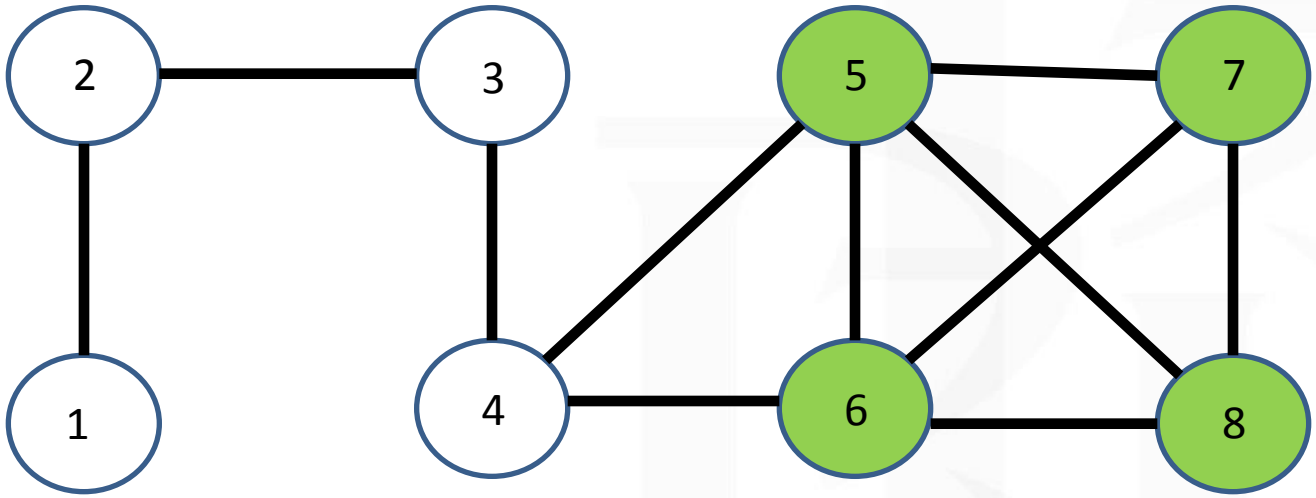
Определения

- ***Вершинной раскраской графа*** $G = (V, E)$ называется такое отображение $c: V \to \mathbb{N}$, что $c(v) \neq c(w)$ для любых двух смежных вершин v, w.

- **Утверждение.** Если граф $G = (V, E)$ может быть раскрашен в k цветов, то $\omega(G) \leq k$, причем в максимальную клику входит не более одной вершины из каждого цвета.

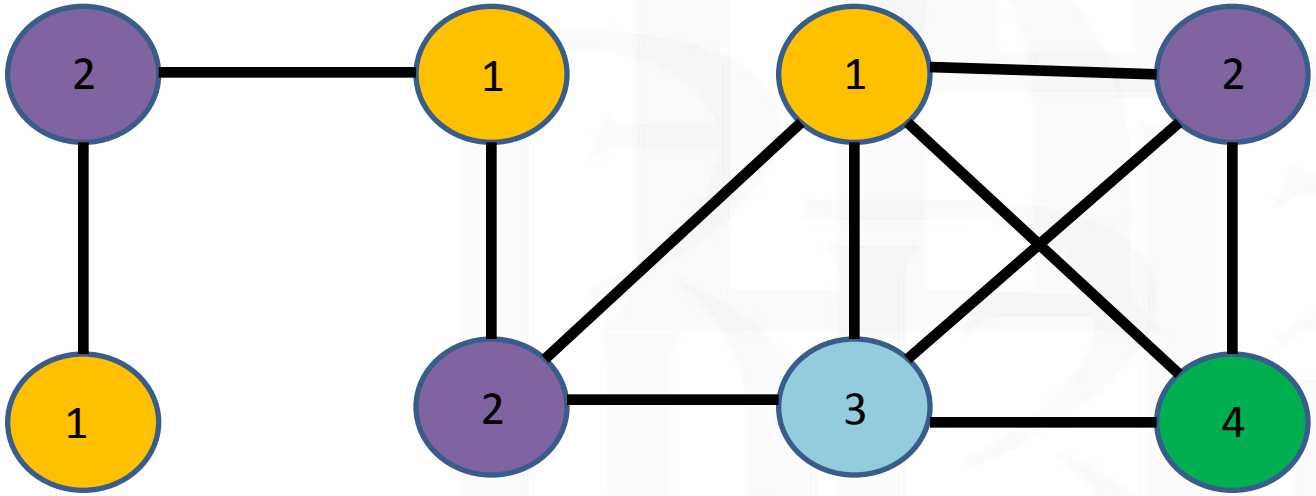
Пример

$\omega(G) \geq 4$



$C_1 = \{1,3,5\}$
$C_2 = \{2,4,7\}$
$C_3 = \{6\}$
$C_4 = \{4\}$

$\omega(G) \leq 4$



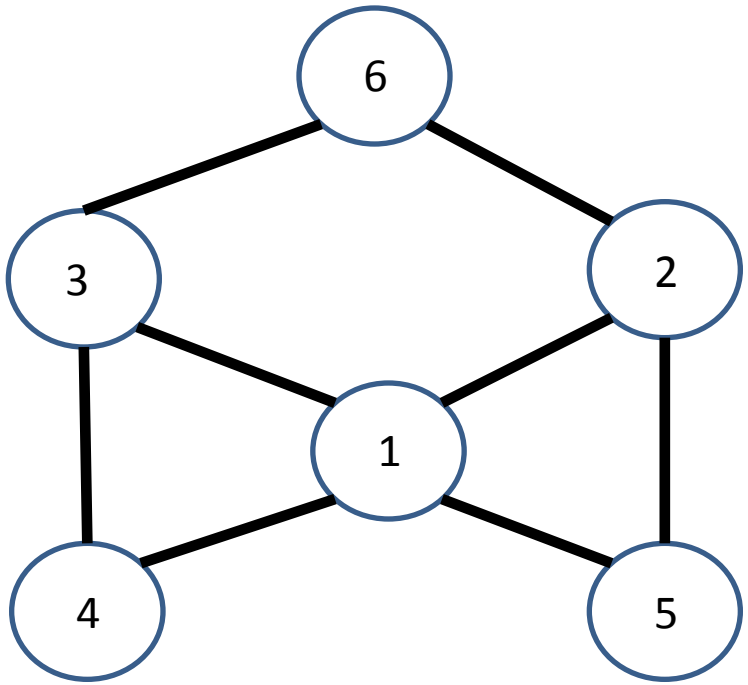
Алгоритм MCS

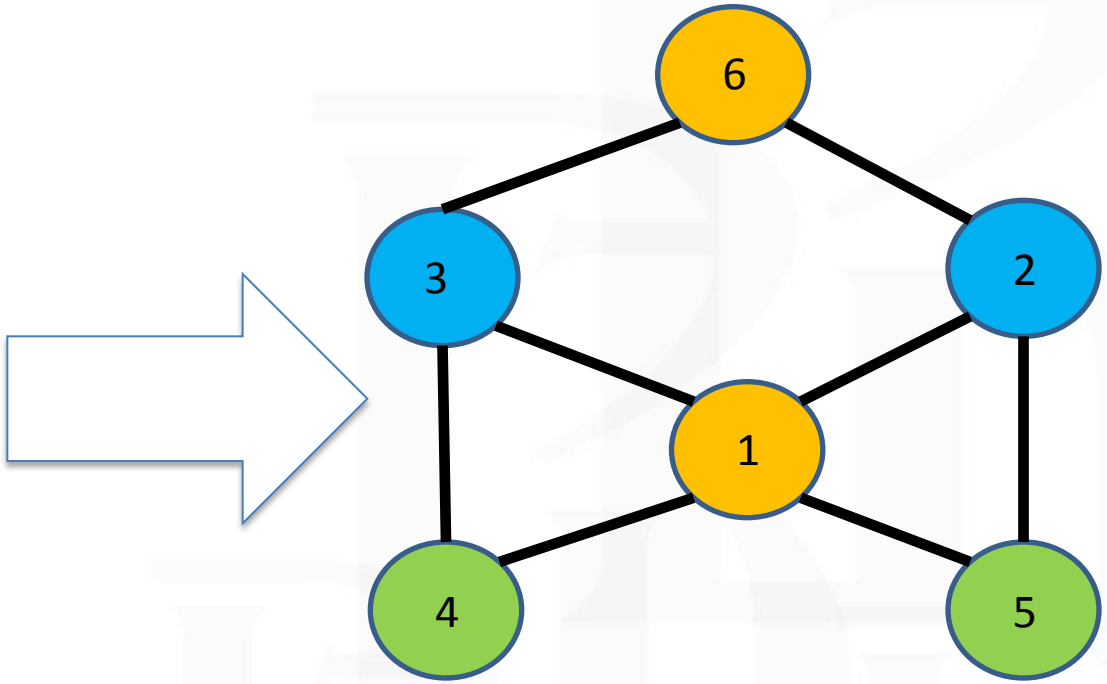
- Алгоритм MCS реализует **метод ветвей и границ**.

- В начале работы алгоритма происходит перенумерация вершин и запускается эвристика для нахождения начального решения.
- В первую очередь алгоритм ветвиться по вершинам, покрашенным в наибольший цвет.
- Для нахождения верхней границы используется жадный алгоритм раскраски с перекраской (процедура Re-NUMBER).

Tomita E., Sutani Y., Higashi T., Takahashi S., Wakatsuki M. *A Simple and Faster Branch-and-Bound Algorithm for Finding a Maximum Clique*. Lecture Notes in Computer Science, 5942, 2010, 191-203

Пример работы MCS (без нахождения начального решения)

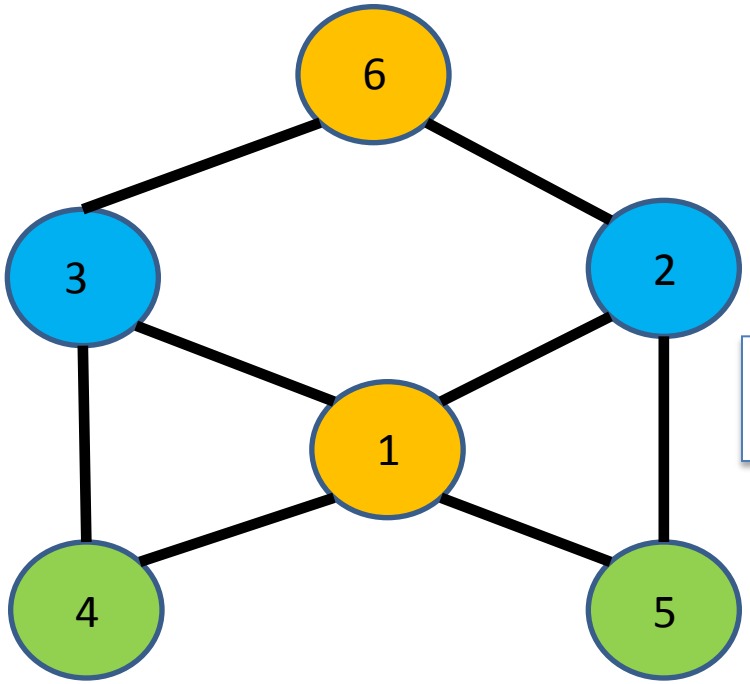


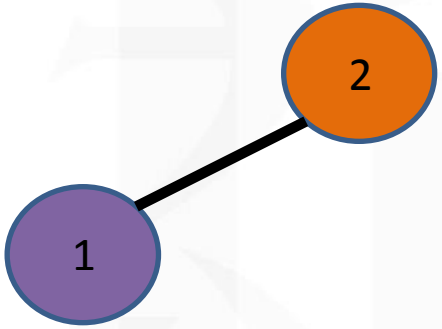


Q – текущая клика
Q_{max} – найденная максимальная клика
U – список вершин кандидатов

$Q = \emptyset, Q_{max} = \emptyset$
$U = \{1,2,3,4,5,6\}$
$C = \{\{1,6\},\{2,3\},\{4,5\}\}$

Пример работы MCS

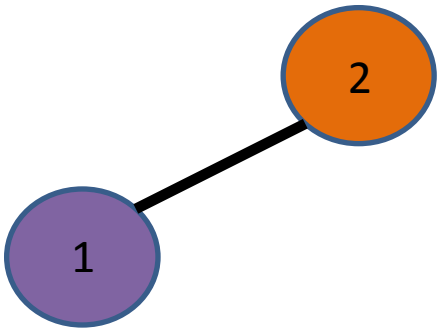


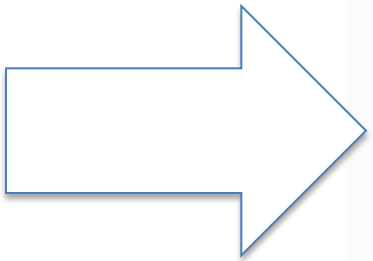


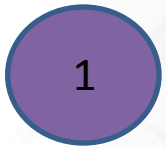
$Q = \emptyset, Q_{max} = \emptyset$
$U = \{1,2,3,4,5,6\}$
$C = \{\{1,6\},\{2,3\},\{4,5\}\}$

$Q = \{5\}, Q_{max} = \emptyset$
$U = \{1,2\}$
$C = \{\{1\},\{2\}\}$

Пример работы MCS







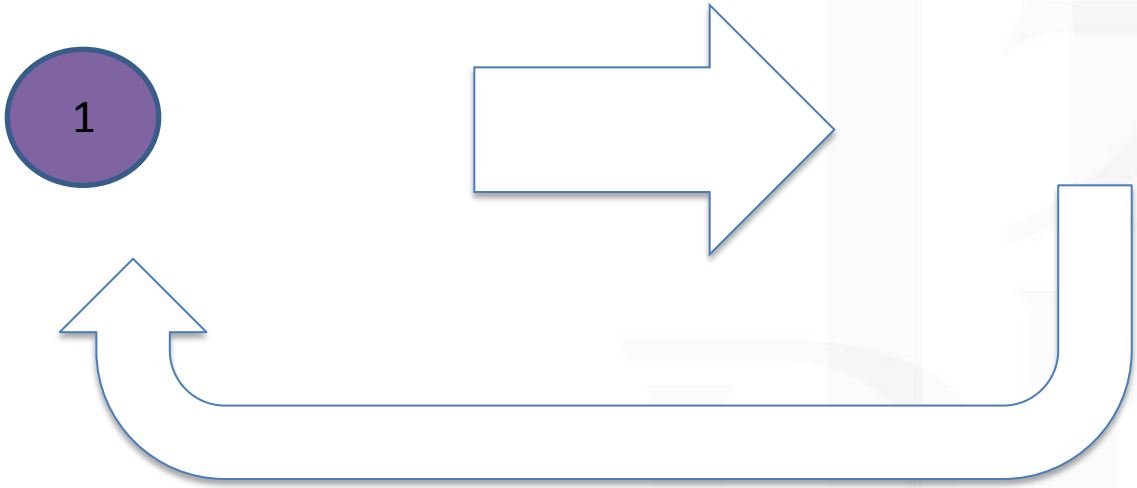
$Q = \{5\}, Q_{max} = \emptyset$
$U = \{1,2\}$
$C = \{\{1\},\{2\}\}$

$Q = \{5,2\}, Q_{max} = \emptyset$
$U = \{1\}$
$C = \{\{1\}\}$

Пример работы MCS

$Q = \{5,2\}, Q_{max} = \emptyset$
$U = \{1\}$
$C = \{\{1\}\}$

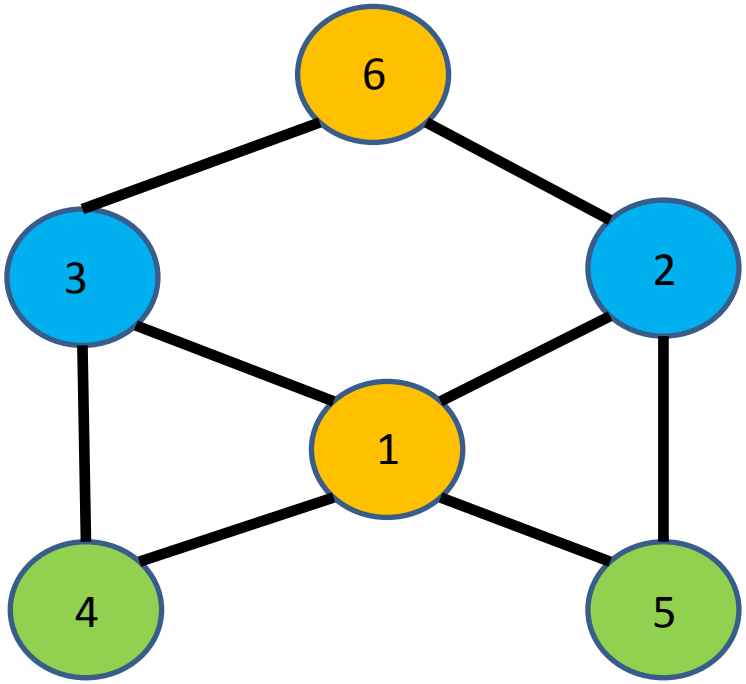
$Q = \{5,2,1\}, Q_{max} = \emptyset$
$U = \emptyset$
$|Q|>|Q_{max}| \Rightarrow Q_{max} = \{5,2,1\}$



Удалить вершину 1 из U

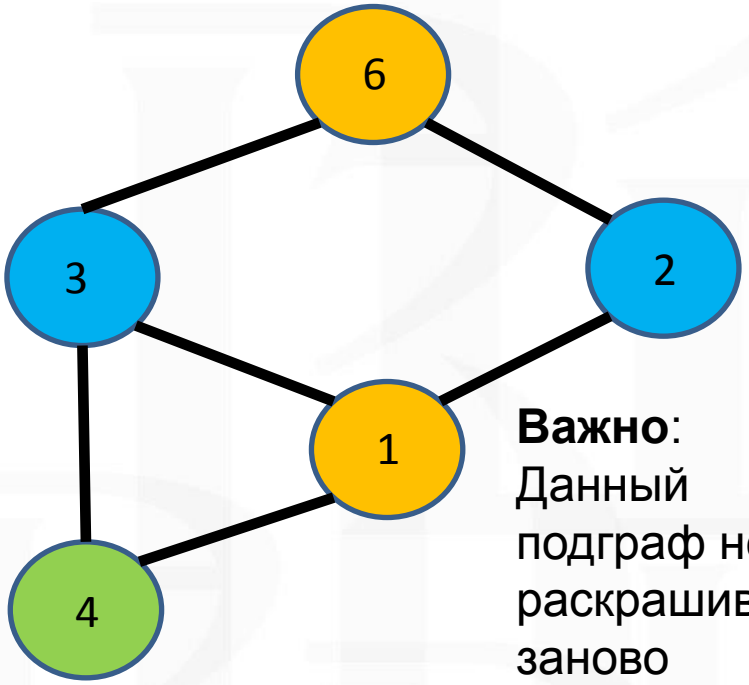
После нескольких пропущенных шагов

Было:



$Q = \emptyset, Q_{max} = \emptyset$
$U = \{1,2,3,4,5,6\}$
$C = \{\{1,6\},\{2,3\},\{4,5\}\}$

Стало:



Важно: Данный подграф не раскрашивается заново

$Q = \emptyset, Q_{max} = \{5,2,1\}$
$U = \{1,2,3,4,6\}$
$C = \{\{1,6\},\{2,3\},\{4\}\}$

Т.к. $|Q| + |C| \leq |Q_{max}|$, то алгоритм останавливается

Жадный алгоритм раскраски с перекраской

Сложность алгоритма жадной раскраски с перекраской: $0(n^3)$

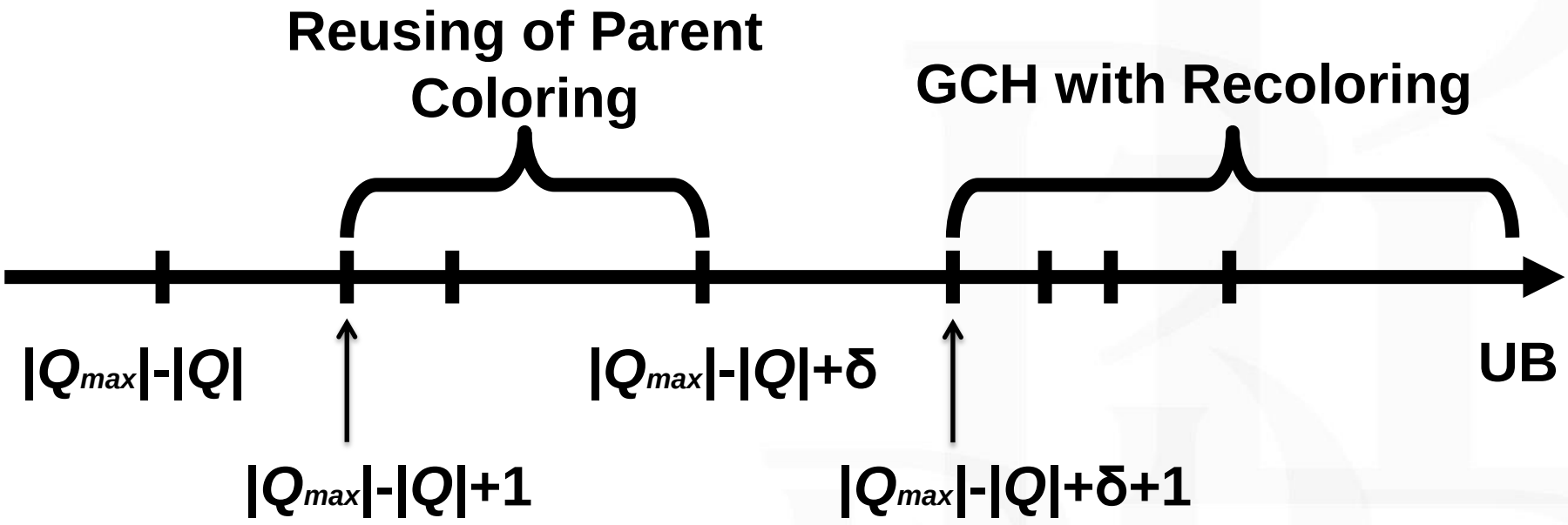
Tomita E., Sutani Y., Higashi T., Takahashi S., Wakatsuki M. *A Simple and Faster Branch-and-Bound Algorithm for Finding a Maximum Clique*. Lecture Notes in Computer Science, 5942, 2010, 191-203

Алгоритм RPC

- Алгоритм RPC - параметрический алгоритм с целым неотрицательным параметром δ.
- Основная идея алгоритма RPC заключается в повторном использовании «родительской» раскраски для некоторых подзадач в дереве поиска вместо трудоемкого вычисления жадной раскраски с перекрашиванием для каждой подзадачи.
- При $\delta = 0$ алгоритм RPC совпадает с алгоритмом MCS.

Nikolaev A., Batsyn M., San Segundo P. *Reusing the same coloring in the child nodes of the search tree for the maximum clique problem*. Lecture Notes in Computer Science, 8994, 2015, 275-280

Алгоритм RPC



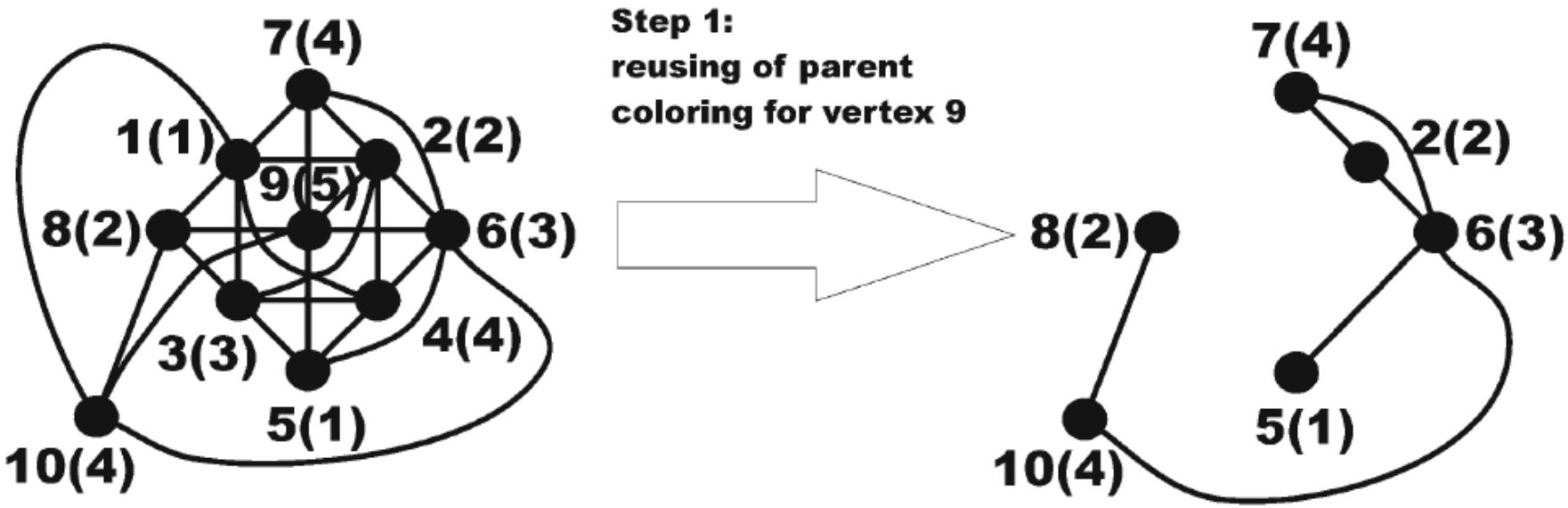
Процедура Reusing of Parent Coloring

1. Повторное использовании родительской раскраски
2. Перекрашивание «вершин-цветов».
3. Сортировка и перенумерация цветов.

Сложность этой процедуры: $0(n^2)$

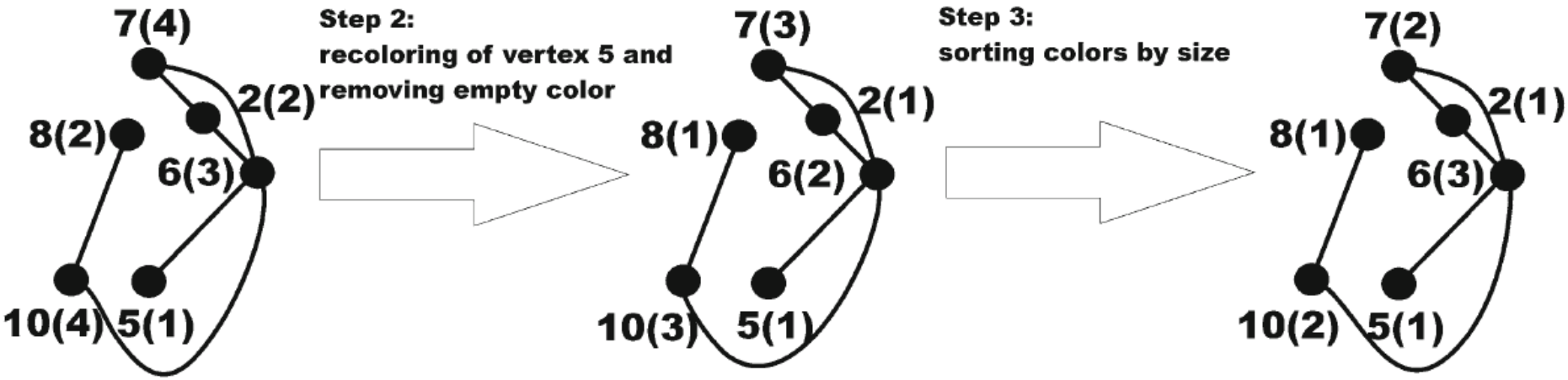
Шаг 1

$$\delta = 2, |Q| = 1, |Q_{max}| = 4$$



Шаг 2 и 3

$$\delta = 2, |Q| = 2, |Q_{max}| = 4$$



Проблема

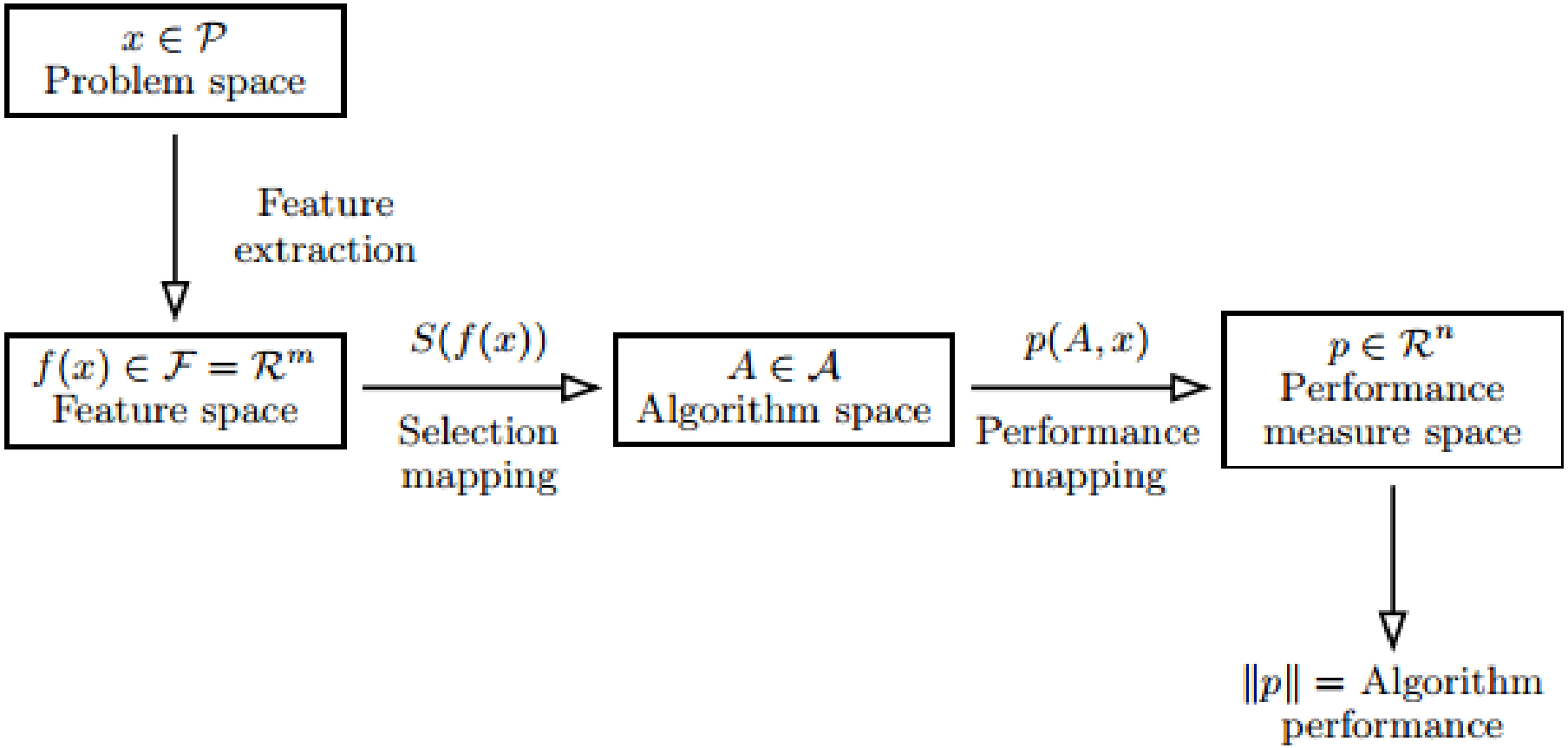
Как выбрать δ?

Algorithm Selection Problem

- ASP состоит в том, что из заданного множества алгоритмов необходимо выбрать самый быстрый алгоритм для решения конкретной задачи.

Rice, J. R. (1976). The algorithm selection problem. Advances in Computers, 15, 65–118.

Algorithm Selection Problem



Rice, J. R. (1976). The algorithm selection problem. Advances in Computers, 15, 65–118.

Selection mapping (Algorithm Selectors)

- Case-based reasoning
  k-NN,…
- Classification
  SVM, decision tree, random forest,…
- Regression
  linear regression, nonlinear regression,…

* Kotthoff, L., Gent, I., Miguel, I.: A Preliminary Evaluation of Machine Learning in Algorithm Selection for Search Problems. In Borrajo, D., Likhachev, M., Lopez, C., eds.: Procs. SoCS'11, AAAI Press (2011) 84–91

Признаки

Номер признака	Описание признака
1	Число вершин
2	Число ребер
3	Плотность графа
4	Минимальная степень вершин
5	Максимальная степень вершин
6	Среднее значение степени вершины
7	Среднее квадратичное отклонение степени вершины

Николаев А. И. Эффективный подход на основе машинного обучения для решения задачи о максимальной клике // Информационные технологии, Т. 22, № 4, 2016, С. 249-254.

Признаки

Номер признака	Описание признака
8	Минимальная сумма степеней смежных вершин
9	Максимальная сумма степеней смежных вершин
10	Среднее значение суммы степеней смежных вершин
11	Среднее квадратичное отклонение суммы степеней смежных вершин
12	Размер клики, найденной жадной эвристикой (нижняя граница)
13	Число цветов в жадной раскраске графа (верхняя граница)
14	Число вершин, которые нужно рассмотреть на первом уровне дерева поиска (ширина дерева поиска)
15	Доля вершин в графе, которые нужно рассмотреть на первом уровне дерева поиска (относительная ширина дерева поиска)

Сгенерированная выборка для обучения (950 графов)

n	*p*
150	**0.05-0.95**
200	**0.05-0.95**
300	**0.05-0.85**
400	**0.05-0.6**
500	**0.05-0.6**
1000	**0.05-0.4**
1500	**0.05-0.4**

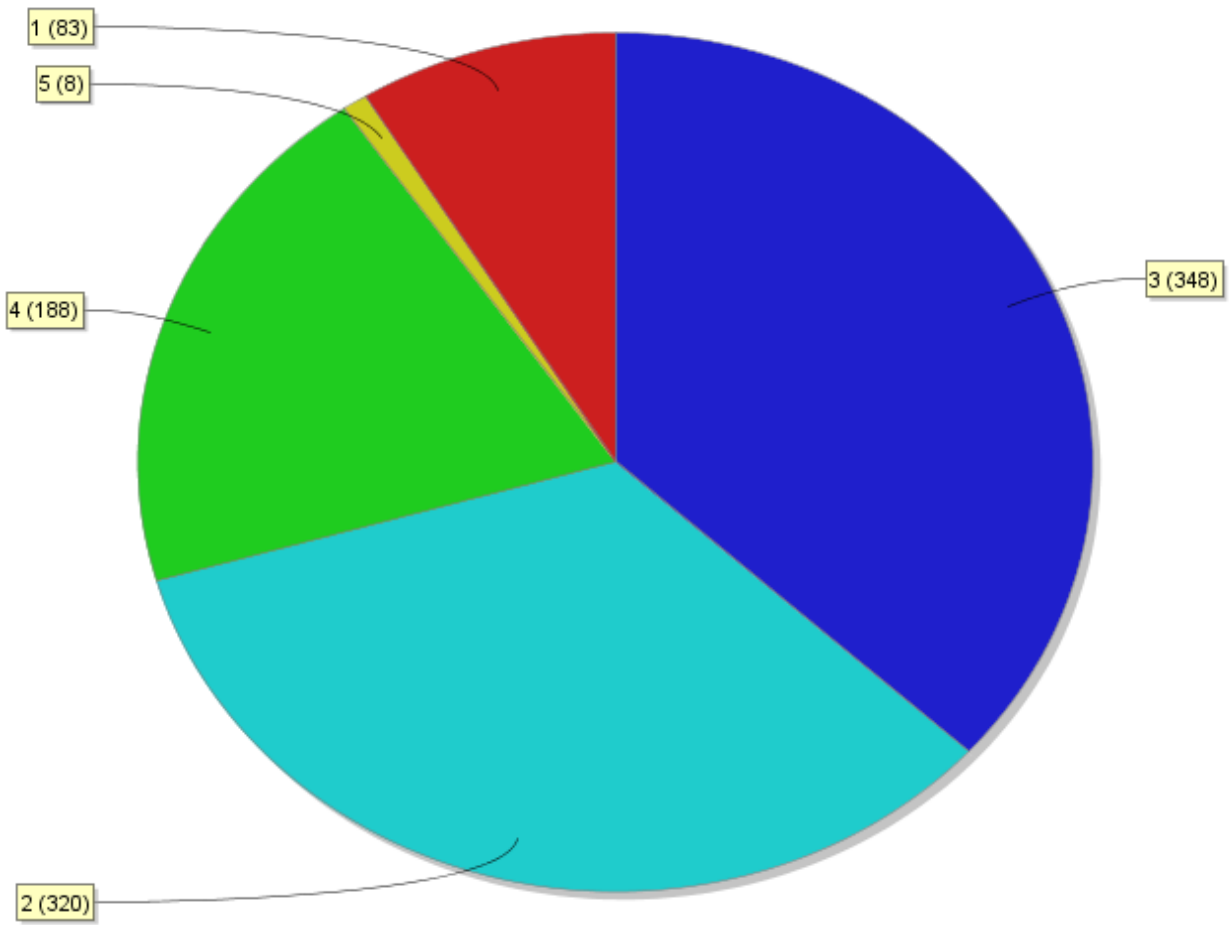
Algorithm Selection Problem

Algorithm portfolio:

1. MaxCLQ

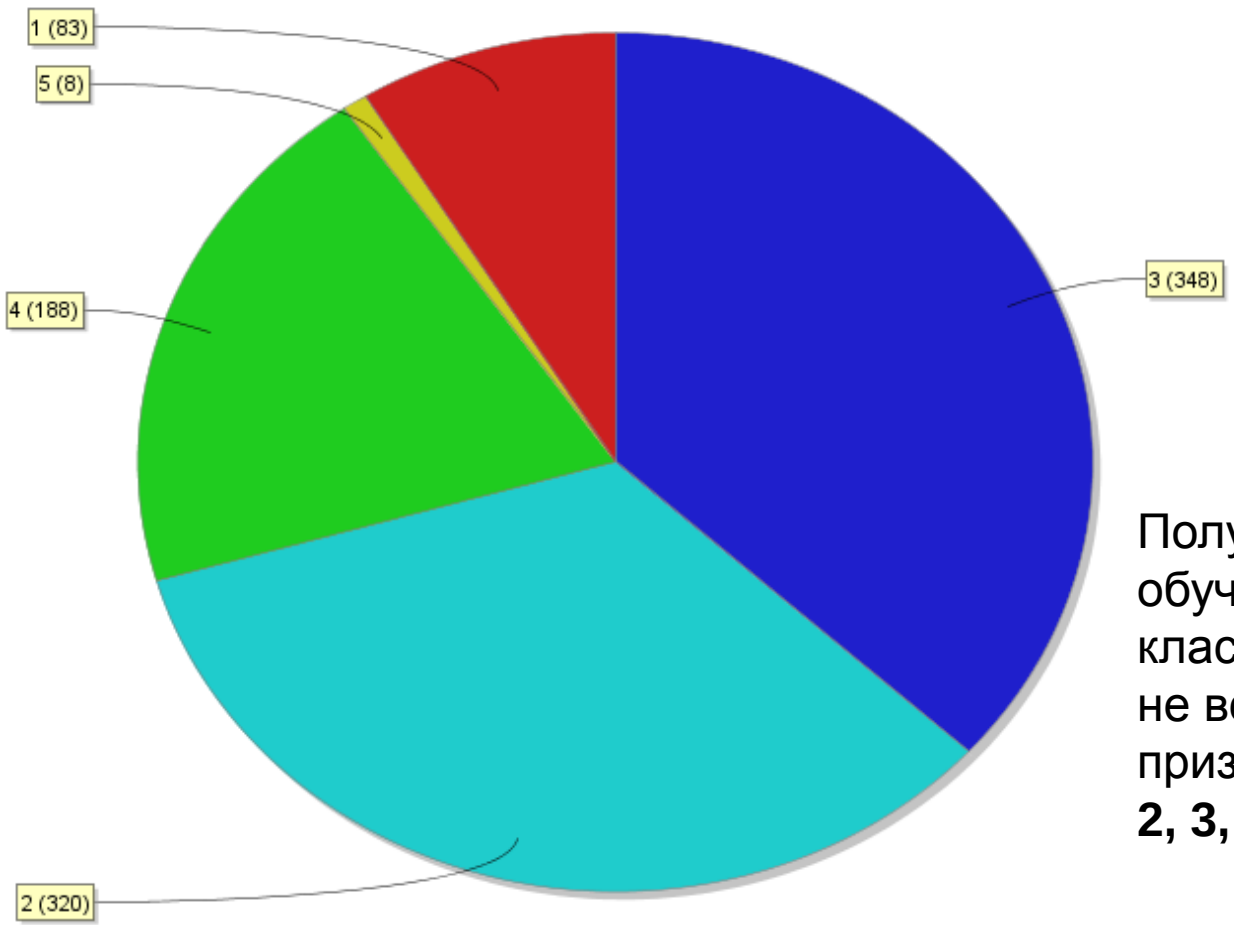
   Li C.M., Quan Z. *Combining graph structure exploitation and propositional reasoning for the maximum clique problem*. Proceedings of the 2010 22nd IEEE International Conference on Tools with Artificial Intelligence, Volume 01 (ICTAI'10), 2010, 344-351
2. RPC, δ=0
3. RPC, δ=1
4. RPC, δ=2
5. RPC, δ=3

Обучающая выборка (947 графов)



1 класс – MaxCLQ
2 класс – RPC, δ=0
3 класс – RPC, δ=1
4 класс – RPC, δ=2
5 класс – RPC, δ=3

Decision tree



Полученные в результате обучения деревья классификации использовали не все признаки, а только признаки
2, 3, 6, 8, 10, 12, 13, 14.

Вычислительные результаты (DIMACS)

Confusion matrix

	true 1	true 3	true 4	true 5	true 2	class precision
pred. 1	**2**	1	0	0	0	66.67%
pred. 3	0	**1**	0	0	0	100.00%
pred. 4	0	0	**13**	8	0	61.90%
pred. 5	0	0	0	**0**	0	0.00%
pred. 2	0	0	0	0	**0**	0.00%
class recall	100.00%	50.00%	100.00%	0.00%	0.00%	

Точность предсказания на тестовой выборке - 64% (16 из 25). В случае неправильного предсказания выбирался второй по быстроте алгоритм.

Вычислительные результаты (DIMACS)

Граф DIMACS	MaxCLQ	MCS	RPC, $\delta = 1$	RPC, $\delta = 2$	RPC, $\delta = 3$	Наиболее быстрый алгоритм	Предложенный подход
C250.9	344516	1361335	1041168	987836	971229	**344516**	344517
MANN_a45	34148	43230	31159	34574	85558	**31159**	34154
brock400_1	259708	284586	244209	234715	235103	**234715**	234716
brock400_2	118908	125398	107746	103844	103952	**103844**	103845
brock400_3	204222	193386	164518	158995	159121	**158995**	158996
brock400_4	130754	92893	79692	76856	77071	**76856**	76858
brock800_1	5606592	3914626	3430660	3294158	3300223	**3294158**	3294165
brock800_2	4889039	3395345	2971610	2845765	2842208	**2842208**	2845772
brock800_3	3222601	3461488	3020965	2899545	2898460	**2898460**	2899552
brock800_4	2438408	1602096	1405760	1341436	1333342	**1333342**	1341442
dsjc500.5	3532	1555	1412	1426	1487	**1412**	1415
dsjc1000.5	317877	140509	127070	123663	127732	**123663**	123675
frb30-15-1	655244	721617	440375	339713	289979	**289979**	339715
frb30-15-2	951654	533285	296789	208696	160926	**160926**	208698
frb30-15-3	580959	473354	278528	210579	176525	**176525**	210581
frb30-15-4	1155555	1327562	765939	574914	481628	**481628**	574915
frb30-15-5	873662	489109	266500	187442	147959	**147959**	187443

Вычислительные результаты (DIMACS)

Граф DIMACS	MaxCLQ	MCS	RPC, $\delta = 1$	RPC, $\delta = 2$	RPC, $\delta = 3$	Наиболее быстрый алгоритм	Предложенный подход
p_hat300-3	1387	1245	1051	1044	1089	**1044**	1045
p_hat500-3	49829	60698	51642	49310	49779	**49310**	49313
p_hat700-2	3586	2726	2319	2250	2371	**2250**	2255
p_hat700-3	1082242	1063763	894519	841329	854369	**841329**	841333
p_hat1000-2	117828	106392	87986	82428	83927	**82428**	82439
p_hat1500-1	11408	2257	1993	1916	1922	**1916**	1936
sanr200_0.9	5604	13855	10712	10107	10122	**5604**	5604
sanr400_0.7	97663	79299	68261	66440	68368	**66440**	66442
Общее время	23156926	19491609	15792583	14678981	14464450	**13750666**	14030826

Вычислительные результаты (DIMACS)

Среднее сокращение времени вычислений предложенного подхода относительно каждого из рассматриваемых алгоритмов в отдельности для графов библиотеки DIMACS

MaxCLQ	MCS	RPC, $\delta = 1$	RPC, $\delta = 2$	RPC, $\delta = 3$	Наиболее быстрый алгоритм	Предложенный подход
35,74%	28,97%	11,83%	4,41%	3,21%	-4,97%	0,00%

Вычислительные результаты (protein alignment graphs*)

Confusion matrix

	true 5	true 4	true 3	true 2	true 1	class precision
pred. 5	382	11	0	0	0	97.20%
pred. 4	0	0	0	0	0	0.00%
pred. 3	0	0	0	0	0	0.00%
pred. 2	0	0	0	0	0	0.00%
pred. 1	0	0	0	0	0	0.00%
class recall	100.00%	0.00%	0.00%	0.00%	0.00%	

Точность предсказания на тестовой выборке - 97.2% (382 из 393).

*N. Malod-Dognin, R. Andonov, and N. Yanev, Maximum Cliques in Protein Structure Comparison, in Lecture Notes in Computer Science, 6049, SEA 2010, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 106-117.

Вычислительные результаты (protein alignment graphs)

Среднее сокращение времени вычислений предложенного подхода относительно каждого из рассматриваемых алгоритмов в отдельности

MaxCLQ	MCS	RPC, $\delta = 1$	RPC, $\delta = 2$	RPC, $\delta = 3$	Наиболее быстрый алгоритм	Предложенный подход
96,08%	32,69%	13,18%	0,56%	-4,46%	-4,47%	0,00%

Summer School on Operational Research and Applications

10 – 14 мая, 2017, Н. Новгород

https://nnov.hse.ru/en/latna/conferences/school2017

http://intelligent-optimization.org/lion11/

22 – 24 июня, 2017,
Н. Новгород
https://nnov.hse.ru/en/latna/conferences/net2017

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

Спасибо за внимание!