

Evolučné algoritmy pre optimalizáciu tried

David Chalupa

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií

Obsah

- problém optimalizácie tried
- motivácia a východiská
- nové algoritmy
 - multiagentový evolučný algoritmus (MEA)
 - pseudo-reaktívne zakázané hľadanie (PRTS)
- nové aplikácie optimalizácie tried
 - hľadanie podštruktúr v sociálnych sieťach
 - tvorba heterogénnych tímov
- zhodnotenie a otvorené otázky

Problém optimalizácie tried

Problém optimalizácie tried

- ofarbenie grafu

- priradenie $f: V \rightarrow C$

- *dekompozícia do tried* $S = \{V_1, V_2, \dots, V_{|C|}\}$

- aplikácie

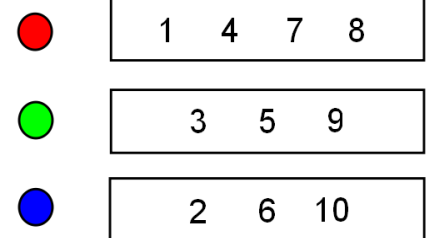
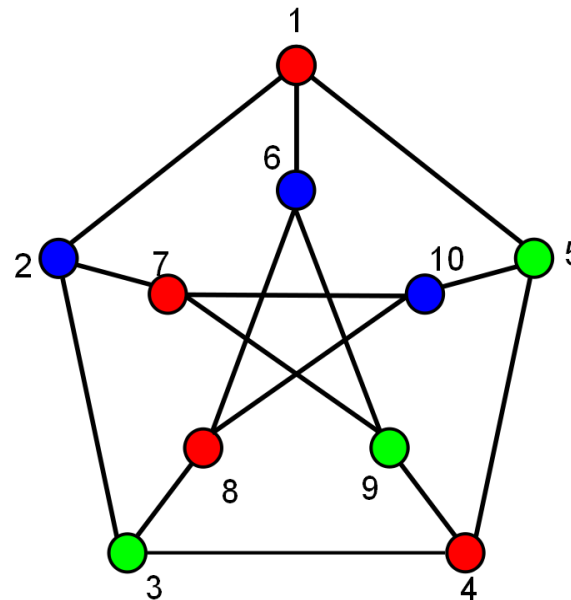
- Sudoku, n dām

- plánovanie

- priradovanie

- tvorba tímov

- klasifikácia na základe závislostí



Veta o ofarbení a triedach

Veta Pre ofarbenie grafu G s vrcholovou množinou V a hranovou množinou E platí, že počet kolízií v ofarbení f , predstavujúceho množinu tried $f = \{V_1, V_2, \dots, V_c\}$, je rovný súčtu počtov kolízií v jednotlivých triedach:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{confl}(v_i, v_j) = \sum_{V_k \in f} \sum_{i=1}^{|V_k|-1} \sum_{j=i+1}^{|V_k|} \text{confl}(v_{ki}, v_{kj}),$$

kde v_{ki} a v_{kj} predstavujú i -ty, resp. j -ty vrchol v triede V_k . Ľavá strana rovnosti reprezentuje pohľad na ofarbenie ako celok, pravá strana dekomponuje kolízie na jednotlivé triedy.

- dôkaz: veta vyplýva z disjunktnosti tried
- dôsledok: optimalizáciou ofarbenia optimalizujeme aj triedy

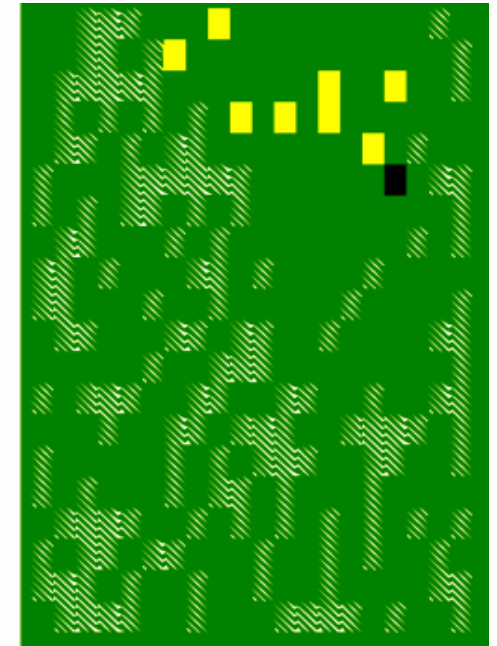
Motivácia a východiská

Motivácia a východiská

- inteligencia roja - moderná oblasť
 - kolónia mravcov, včelí úl'
 - Reynoldsov model, PSO
 - *čmelí algoritmus*
- primárne ciele projektu
 - skúmanie čmelieho algoritmu a jeho vylepšenie
 - skúmanie tradičnejších (meta)heuristík, najmä tzv. zakázaného hľadania a jeho modifikácia

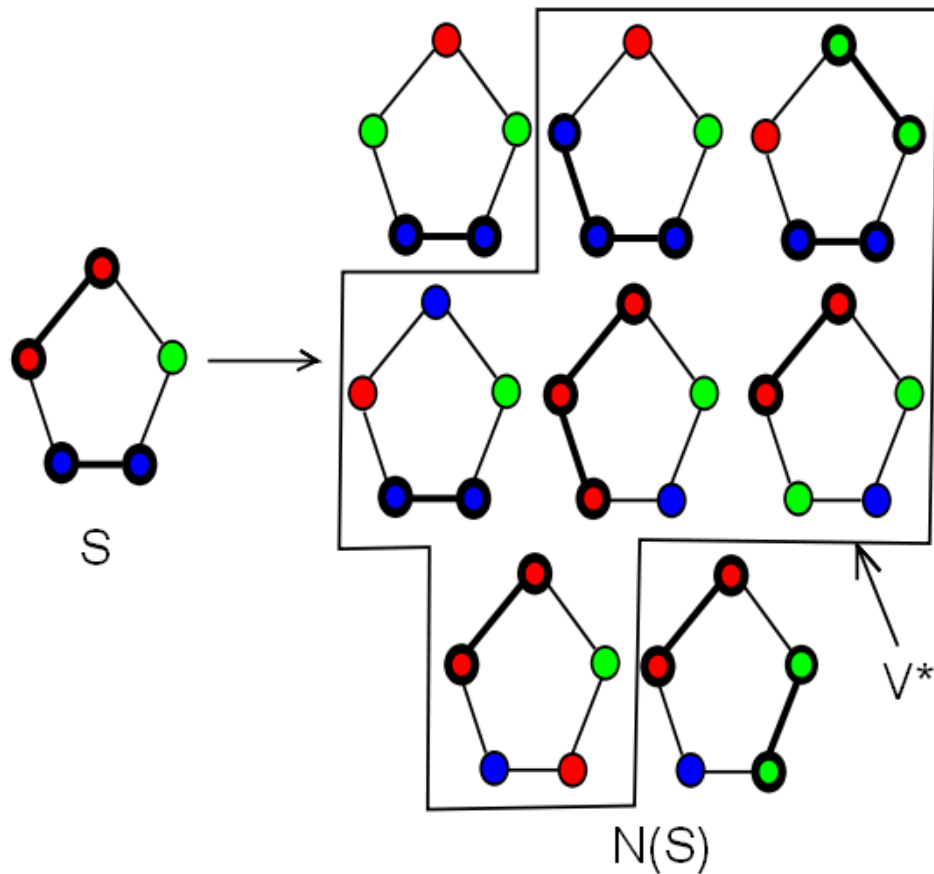
Simulácia správania sa čmeliakov

- čmeliak = stav = ofarbenie grafu
- umelý svet - hniezdo, čmeliaky, potrava
- čmeliaky sa pohybujú kvázináhodne
- zbierajú potravu
- učenie hľadania
- životné body
- potrava zvyšuje ŽB¹



¹ Comellas, F., Martinez-Navarro, J.: Bumblebees: A Multiagent Combinatorial Optimization Algorithm Inspired by Social Insect Behaviour. In: Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation (2009), pp. 811-814.

Lokálne hľadanie



- S - stav (ofarbenie)
- $N(S)$ - úplné okolie²
- V^* - „trénovacie“ okolie
- lokálne hľadanie
 - začína náhodným stavom
 - posúva sa vždy k stavu v okolí s najvyššou fitness
 - „remízy“ rieši náhodne

² Hertz, A., de Werra, D: Using tabu search techniques for graph coloring. In: Computing, vol. 39, 4 (1987), pp. 345-351.

Zakázané hľadanie

- zakázané hľadanie (tabu search)
 - lokálne hľadanie so zákazom „spätných mutácií“
 - v literatúre známe ako *TabuCol* (skr. TCol)
- statický interval zakázanosti
 - konštanta
- dynamický interval zakázanosti
 - $t_i = \alpha f(S) + \text{random}(0, A-1)$
 - $f(S)$: počet konfliktných vrcholov, $\alpha = 0.6$, $A = 10$.³

³ Galinier, P., Hao, J.: Hybrid Evolutionary Algorithms for Graph Coloring. In: Journal of Combinatorial Optimization, vol. 3 (1999), pp. 379-397.

Kvalita základných algoritmov

- tabuľka - počty farieb
- BRE - Brélaz (greedy alg.)
- HC - hillclimbing (bez zakázanosti)
- ČA - čmelí alg.
- TCol - TabuCol (dyn. zakázanosť)
- PCol - PartialCol

G	$ V , \chi$	BRE	HC	ČA	TCol	PCol
<i>le450_15c</i>	450, 15	24	26	30	16	15
<i>le450_15d</i>	450, 15	24	27	30	16	15
<i>le450_25c</i>	450, 25	28	33	37	26	27
<i>le450_25d</i>	450, 25	29	33	37	26	27
<i>flat300_26_0</i>	300, 26	42	39	48	26	26
<i>flat300_28_0</i>	300, 28	41	37	47	31	28
<i>flat1000_76_0</i>	1000, 76	-	106	-	88	88
<i>dsjc500.1</i>	500, ?	-	15	19	12	12
<i>dsjc500.5</i>	500, ?	-	60	76	49	49
<i>dsjc500.9</i>	500, ?	-	138	186	127	127

Multiagentový evoluční algoritmus

Multiagentový evolučný algoritmus

Multiagentový evolučný algoritmus

Vstup: problém, začiatočná veľkosť populácie $|P_0|$,
veľkosť elitného zoznamu $|L_e|$,
maximálny počet životných bodov q , perióda narodenia T_b
podiel precíznych agentov p , odmena/trest r
maximálny počet generácií g_{max}

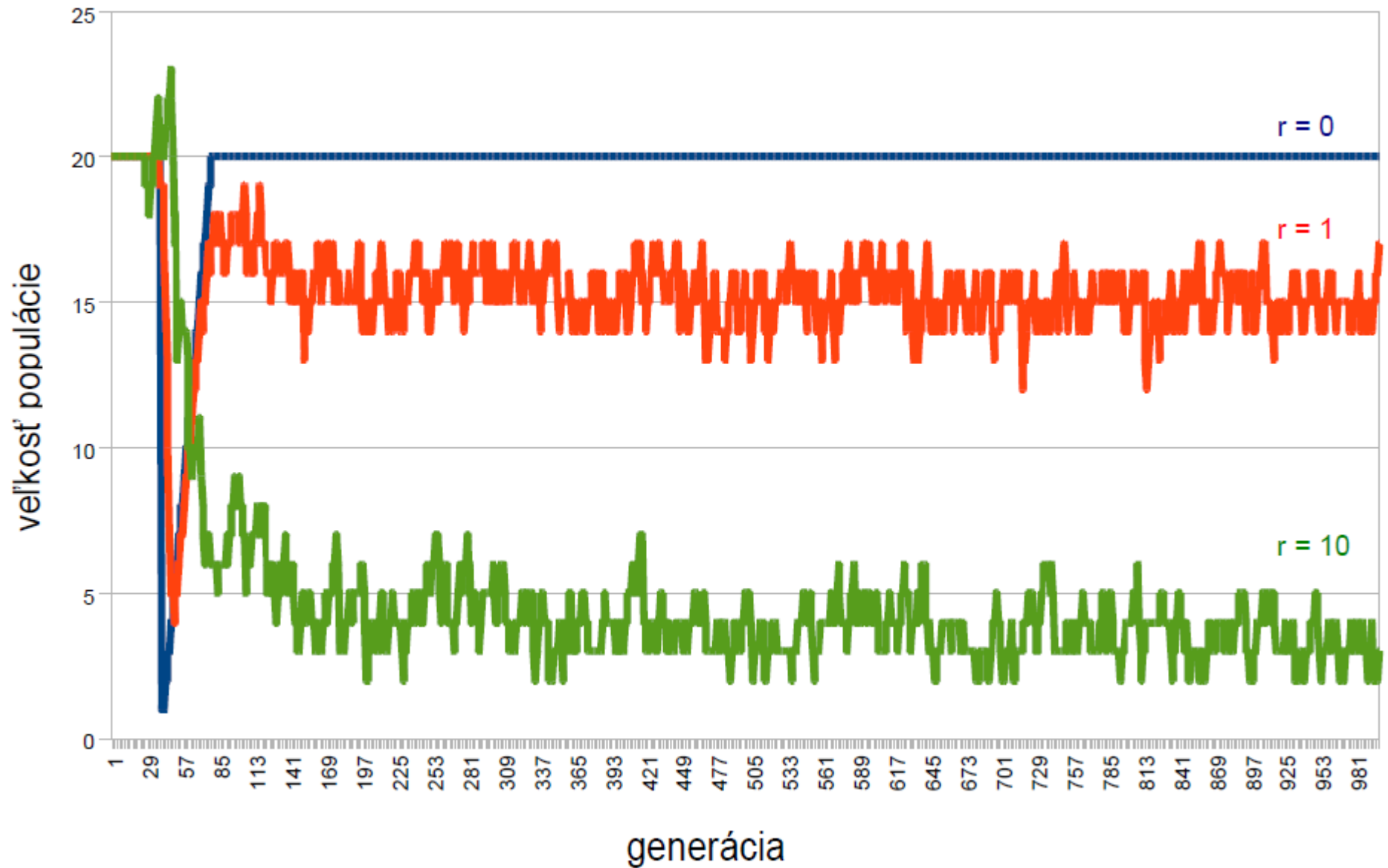
Výstup: koncový stav S

```
1  generuj_agentov()
2  pre  $g = 1..g_{max}$ 
3      ak najlepší_agent.fitness = optimálna_fitness
4          vráť  $S =$  najlepší_agent.stav
5      zníž_životné_body()
6      eliminácia()
7      narodenie()
8      lokálne_hľadanie()
9      vyhodnotenie()
10 vráť  $S =$  najlepší_nájdenny_stav()
```

Natalita a mortalita

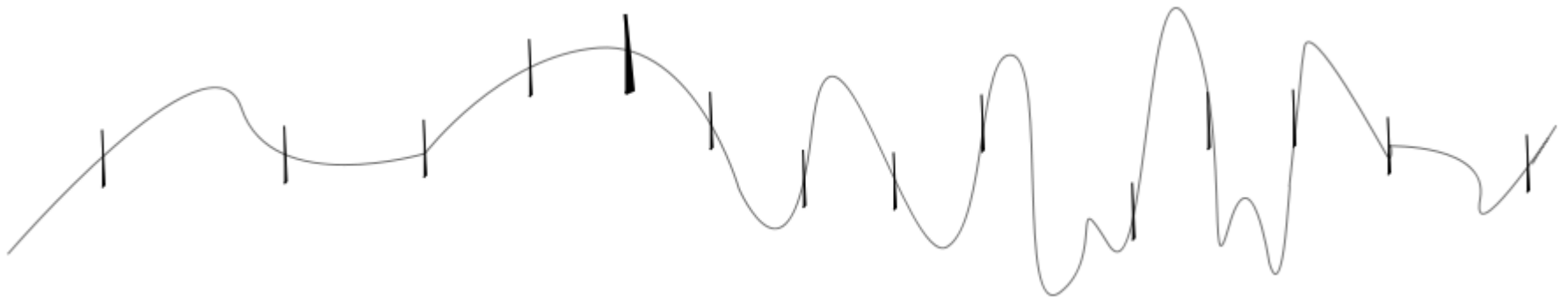
- určujú, ako sa mení populácia v čase
- parametre
 - q - začiatočný počet životov agenta
 - T_b - perióda narodenia agenta
 - r - odmena / trest - vplyv inštancie problému
- veľkosť populácie
 - ak $r \neq 0$, má oscilačný charakter
 - ak $r = 0$, ustáli sa na hodnote q / T_b

Dynamika populácie



Riadenie elitného zoznamu

- *elitný zoznam* - nie je to iba dátová štruktúra na pamätanie si najlepších stavov
- 3 pravidlá
 - agent môže za život odovzdať iba jeden stav
 - tento stav musí byť výsledkom zakázaného hľadania v danej generácii (tzv. miľníky)
 - nahrádza sa najhorší stav v elitnom zozname

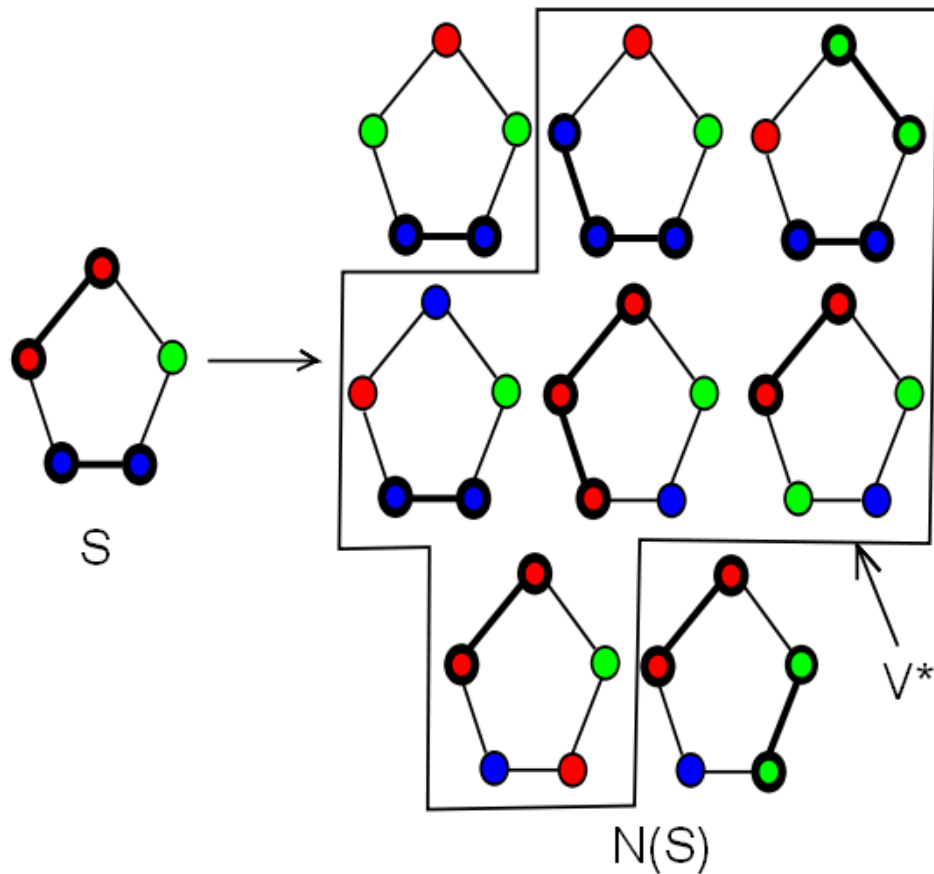


Lokálne hľadanie

- zakázané hľadanie (tabu search)
 - lokálne hľadanie so zákazom „spätných mutácií“
 - interval zakázanosti: $t_i = \alpha f(S) + \text{random}(0, A-1)$
 - $f(S)$: počet konfliktných vrcholov, $\alpha = 0.6$, $A = 10$.²
- počet iterácií:
 - $T_{\max} = \beta |V(G)||C| / |V^*|$
 - β je parameter, $|V^*|$ veľkosť okolia
- ŽB agenta modifikované podľa zmeny fitness

² Galinier, P., Hao, J.: Hybrid Evolutionary Algorithms for Graph Coloring. In: Journal of Combinatorial Optimization, vol. 3 (1999), pp. 379-397.

Definícia okolia



- S - stav (agent)
- $N(S)$ - úplné okolie³
- V^* - „trénovacie“ okolie
- ak $V^* = N(S)$
 - precízny agent
- ak $V^* \neq N(S)$
 - lenivý agent

³ Hertz, A., de Werra, D: Using tabu search techniques for graph coloring. In: Computing, vol. 39, 4 (1987), pp. 345-351.

Parametre algoritmu

- $|P_0| = |L_e|$ - začiatočná veľkosť populácie a veľkosť elitného zoznamu
- q, T_b - ovplyvňujú natalitu a mortalitu
- β - ovplyvňuje počet iterácií zakázaného hľadania
- p - percentuálny podiel precíznych agentov
- r - miera vplyvania inštancie na dynamiku populácie

Experimentálne výsledky

		Metódy lokálneho hľadania			Hybridné algoritmy		
G	MEA	TCol	PCol	VSS	DCNS	GH	EvoCol
<i>le450_15c</i>	15	16	15	15	15	15	-
<i>le450_15d</i>	15	16	15	15	15	-	-
<i>le450_25c</i>	25	26	27	26	25	26	25
<i>le450_25d</i>	25	26	27	26	25	-	25
<i>flat300_26_0</i>	26	26	26	26	26	-	-
<i>flat300_28_0</i>	31	31	28	29	31	31	31
<i>flat1000_76_0</i>	87	88	88	87	89	83	82
<i>dsjc500.1</i>	12	12	12	12	12	-	12
<i>dsjc500.5</i>	48	49	49	48	49	48	48
<i>dsjc500.9</i>	126	127	127	126	128	-	126
<i>artif_100_20_c</i>	20	20	-	-	-	-	-
<i>artif_500_100_c</i>	97	99	-	-	-	-	-
<i>artif_2000_400_c</i>	397	398	-	-	-	-	-
<i>facebook_c</i>	16	16	-	-	-	-	-

Pseudo-reaktívne zakázané hľadanie

Pseudo-reaktívne zakázané hľadanie

- myšlienka - nepoužívať „zázračné konštanty“, ale naučiť sa parametre počas výpočtu
- dynamický interval zakázanosti
 - $t_i = \alpha f(S) + \text{random}(0, A-1)$
- reaktívny interval zakázanosti
 - t_i sa mení podľa fluktuácie fitness
- pseudo-reaktívny interval zakázanosti
 - $t_i = \alpha_T/10 f(S) + A$
 - α_T a A sa menia podľa fluktuácie fitness

Parametre PRTS

- parametre pseudo-reaktívneho hľadania
 - α_T a A – parametre výpočtu intervalu zakázanosti
 - Φ – perióda sledovania fluktuácie fitness
 - b – prah zacyklenia, c – prah prílišnej diverzifikácie
 - bližšie informácie v publikácii ⁴

parameter	Intervaly		Rozsah mutácie	
	min	max	min	max
α_T	0	10	0	1
A	0	20	0	2
ϕ	500	5000	0	500
b	0	10	0	1
c	0	10	0	1

⁴ Chalupa, D.: **Population-based and Learning-based Metaheuristic Algorithms for the Graph Coloring Problem**. In: *Proceedings of Genetic and Evolutionary Conference (2011), Dublin, to appear.*

Porovnanie PRTS a algoritmu TabuCol

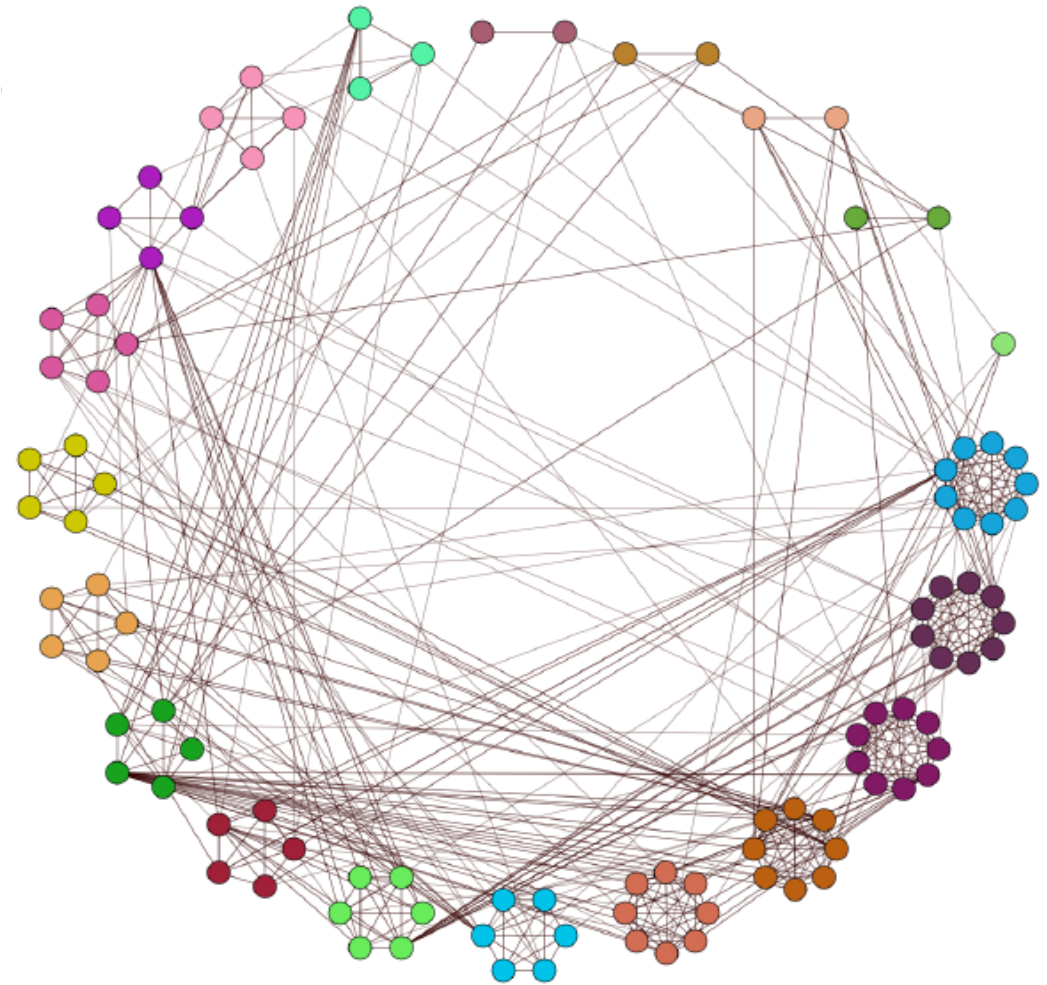
- pseudo-reaktívny (PRTS) a dynamický (TabuCol) interval

Inštancia			PRTS			TabuCol		
G	$ V , \chi$	c	úsp.	st. $\times 10^6$	CPU	úsp.	st. $\times 10^6$	CPU
<i>le450_15c</i>	450, 15	15	10/10	6606	4 m	0/10		
		16	10/10	660	20 s	10/10	54	3 s
<i>le450_15d</i>	450, 15	15	7/10	33316	27 m	0/10		
		16	10/10	890	28 s	10/10	129	4 s
<i>le450_25c</i>	450, 25	26	10/10	1894	70 s	10/10	33	2 s
<i>le450_25d</i>	450, 25	26	10/10	1872	64 s	10/10	27	1 s
<i>flat_300_26_0</i>	300, 26	26	10/10	651	17 s	10/10	2226	63 s
<i>flat_300_28_0</i>	300, 28	30	1/10	93873	58 m	0/10		
		31	10/10	10490	7 m	10/10	10236	7 m
<i>flat_1000_76_0</i>	1000, 76	88	1/5	83761	51 m	4/5	53696	34 m
<i>dsjc500.1</i>	500, ?	12	10/10	5523	5 m	10/10	3368	3 m
<i>dsjc500.5</i>	500, ?	49	6/10	51392	35 m	3/10	91285	49 m
<i>dsjc500.9</i>	500, ?	127	9/10	33119	18 m	10/10	8568	6 m

Nové aplikácie optimalizácie tried

Hľadanie podštruktúr v sociálnych sieťach

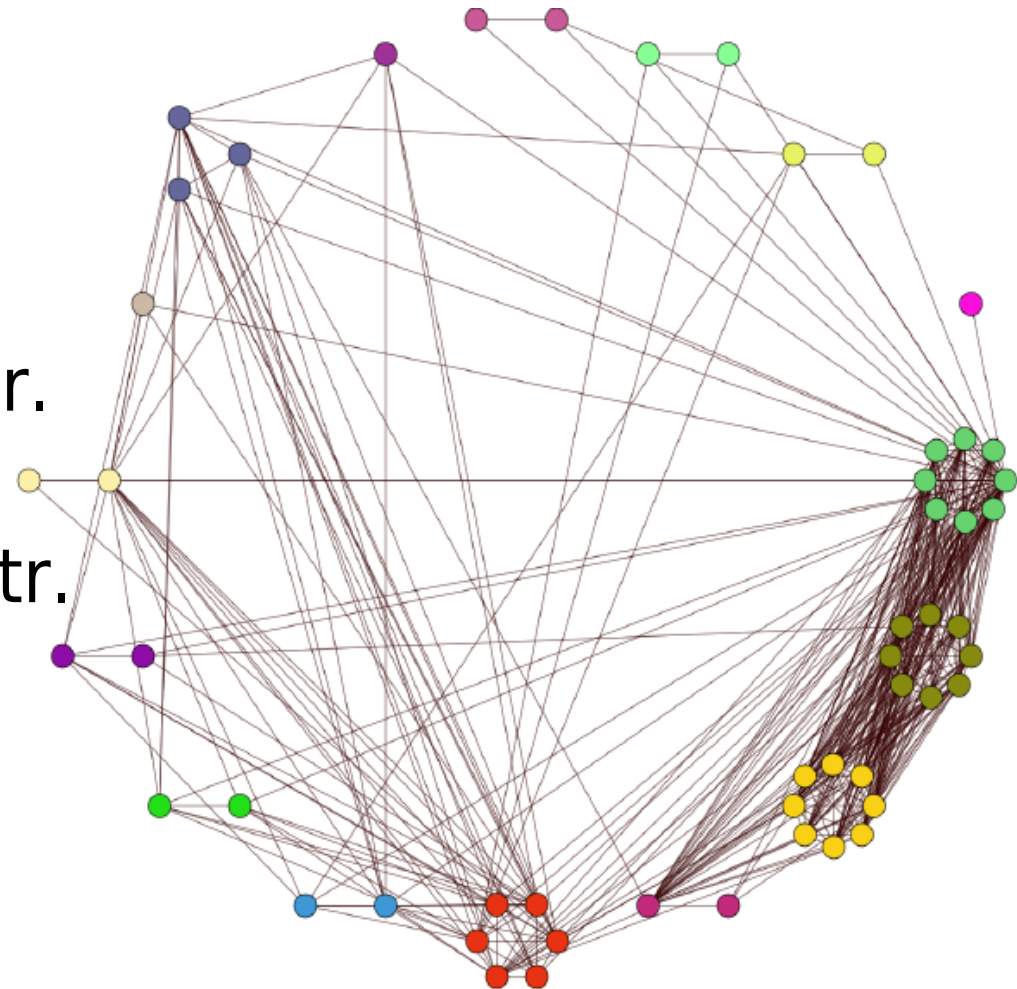
- prirodzený problém
 - hľadáme kompletné subgrafy
- optimalizácia tried
 - hľadá nezávislé triedy
- obrázok
 - umelá sieť
 - 100 vrcholov, 20 podštruktúr



Hľadanie podštruktúr v sociálnych sieťach

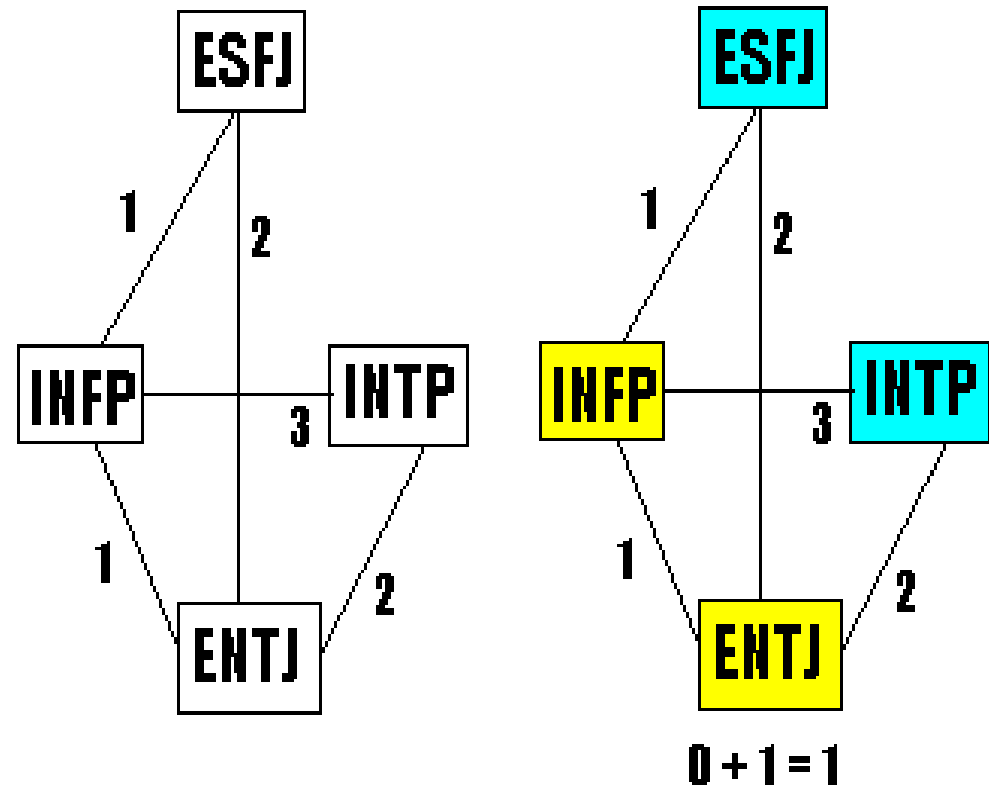
- obrázok

- reálna sieť 52 ľudí
- optimum: 16 podštruktúr
- silne zviazané podštr. (vpravo dole)
- slabo zviazané podštr. (vľavo)
- izolované podštr. (hore)



Tvorba heterogénnych tímov

- Myers-Briggsovej typológia
 - 4 dichotómie
- optimalizácia tried
 - hrana vyjadruje podobnosť osobností
 - hľadanie optimálneho rozdelenia do tímov
 - optimálne rozdelenie je rovnomerné



Zhodnotenie a otvorené otázky

Vedecký prínos

- multiagentový evolučný algoritmus
 - zlepšenie výsledkov čmelieho algoritmu
 - zlepšenie výsledkov zakázaného hľadania
 - „lokálne hľadanie riadené pomocou roja“
 - využitie umelého života v optimalizácii
 - všeobecné myšlienky - potenciál aj pre iné problémy
- pseudo-reaktívne zakázané hľadanie
 - zlepšenie výsledkov zakázaného hľadania
 - interval zakázanosti už nie je viazaný na „zázračné konštanty“

Otvorené problémy

- spojenie algoritmov MEA a PRTS
 - komunikácia parametrov v rámci modelu
- alternatívy k zakázanému hľadaniu
 - „inteligentné“ prekonávanie lokálnych extrémov
- paralelné a distribuované verzie algoritmov
 - paralelné lokálne hľadanie, distribuovaná verzia MEA
- použitie operátorov kríženia
 - inovácie v samotnom krížení, nové využitie
- nové aplikácie, topológie a inštancie
 - sociálne siete, dokumenty, vizualizácia, ...

Ďalší výskum?

- farbenie a klasifikácia
 - farbenie je určitá forma „klasifikácie“
 - premostenie farbenia a klasifikačných problémov
 - nové aplikácie, nové inštancie
- algoritmy
 - nameraná kvalita algoritmov závisí od testovacích dát
 - súčasný výskum vo svete – optimálne farbenie „stredne“ veľkých grafov (1000-2000 vrcholov)
 - otvorená otázka – suboptimálne farbenie obrovských grafov (2000-10000 vrcholov)

Publikácie

- SOFSEM 2011 SRF
 - Chalupa, D.: **An optimization Strategy Based on Social Insect Behavior**. In: *SOFSEM 2011: Student Research Forum (2011)*, pp. 35-50.
- IIT.SRC 2011
 - Chalupa, D.: **Discovering the ability of graph coloring heuristics to find substructures in social networks**. In: *Proceedings in Informatics and Information Technologies: Student Research Conference (2011)*, to appear.
- GECCO 2011
 - Chalupa, D.: **Population-based and Learning-based Metaheuristic Algorithms for the Graph Coloring Problem**. In: *Proceedings of Genetic and Evolutionary Conference (2011)*, Dublin, to appear.

Ďakujem za pozornosť.

xchalupa@stuba.sk