Meta-optimization for combinatorial problems

Oleg Kovářík
oleg.kovarik@gmail.com

SUI 31.3.2011

Computational Intelligence Group
Faculty of Electrical Engineering
Faculty of Information Technology
Czech Technical University in Prague

http://cig.felk.cvut.cz

Content

- Part 1: Combinatorial optimization
  - TSP and improvements
  - Algorithm combination
- Part 2: Meta-optimization
  - TSP Meta-data extraction
  - on-line database
  - GPTSP: evolving data

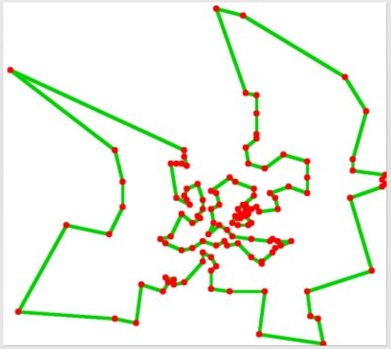
Part 1: Combinatorial optimization

Combinatorial optimization

- NP-complete problems
  - TSP - Traveling Salesman Problem
  - SAT - Boolean Satisfiability Problem
  - JSP - Job-shop Problem
  - QAP - Quadratic Assignment Problem
  - ...
- Algorithms
  - Exact
  - Heuristics
  - Local search algorithms

Traveling Salesman Problem (TSP)

- Find shortest possible tour that visits each city from given set exactly once.



- Exact algorithms
  - Branch & Bound, solution of LP relaxation, ...
  - Best: Concorde http://www.tsp.gatech.edu/concorde.html
- Heuristics
  - Simulate Annealing, Evolutionary Algorithms, Tabu Search, GRASP, Ant Colony Optimization, … often combined with
- Local search algorithms
  - 2-OPT, 3-OPT, k-OPT, Lin-Kernighan (best), …
- Construction heuristics

Problems

- Exact algorithms: non-standard (dynamic) problems, very large instances, complicated
- Local search: is local
- Heuristics: many possibilities

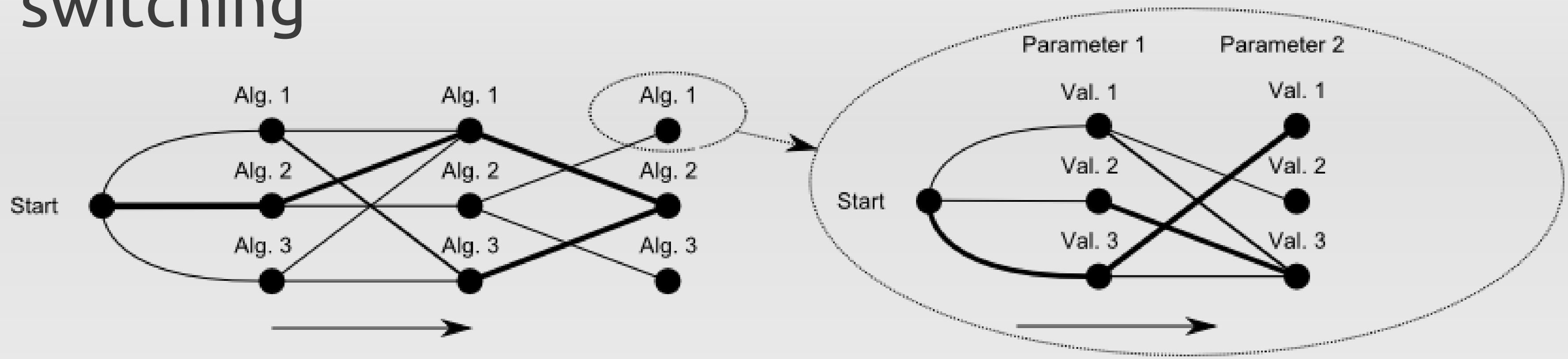
- All: parameters, different implementations, unknown performance for given data
- Improvements: adaptivity, parallelization, combination of various algorithms

Previous improvements

- Ant Colony Optimization with Castes
  - diversification of parameters inside one algorithm
- Ant Colony Optimization with Parallel Subsolutions Heuristic (2008)
  - decomposition & parallelization on PS3
- described in previous presentation

Algorithm combination - serial

- switching



- by ants

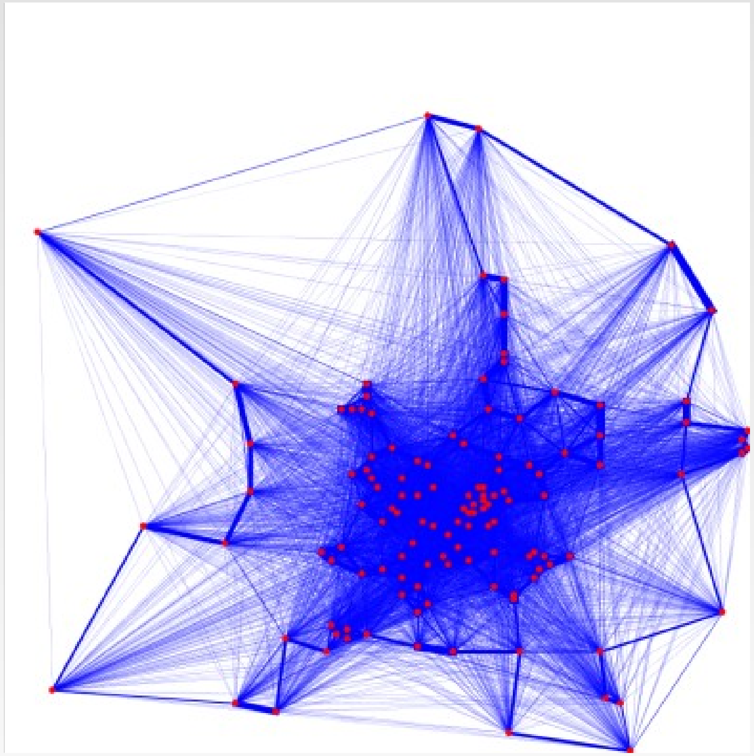
```
root (global prob = 10)
 Tabu1 (238 / 37199)
   Tabu2 (155 / 37066)
     GA9 (60 / 37021)
       GRASP25 (17 / 36986)
     NM14 (22 / 36983)
 GA1 (63 / 37296)
```

```
best: 679063 (Final1)
best: 37120 (Tabu1Final3)
best: 37056 (GA1Final8) avg best: 37225 (Tabu1)
best: 37006 (Tabu1Tabu2Final44) avg best: 37149 (ILS5SA6)
best: 36918 (Tabu1Tabu2ILS15Final137) avg best: 37006 (Tabu1Tabu2GA9)
```

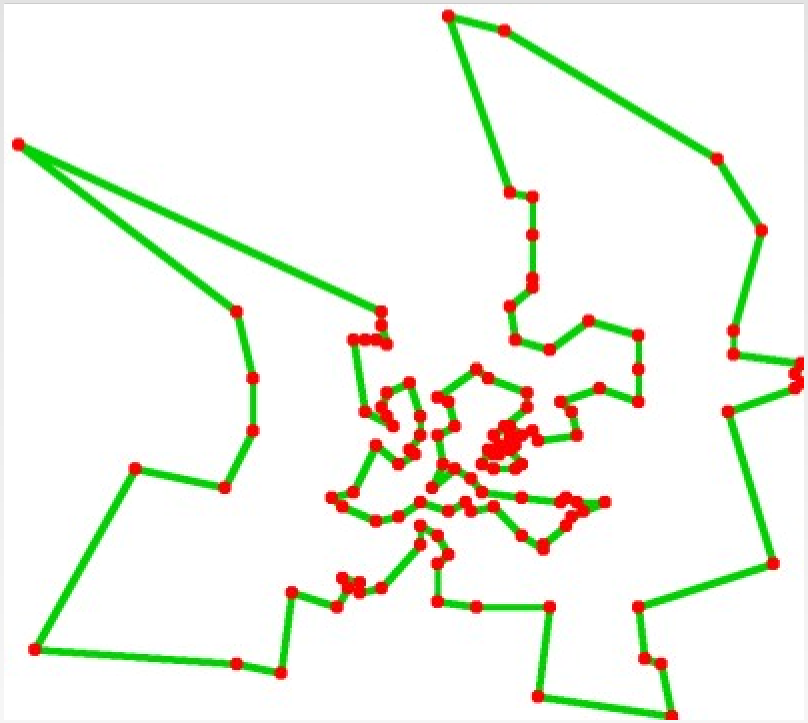
Best average: Tabu1 Tabu2 NM14

Algorithm combination - parallel

- Ant Colony Optimization
  - probabilistic representation of solution space - pheromone



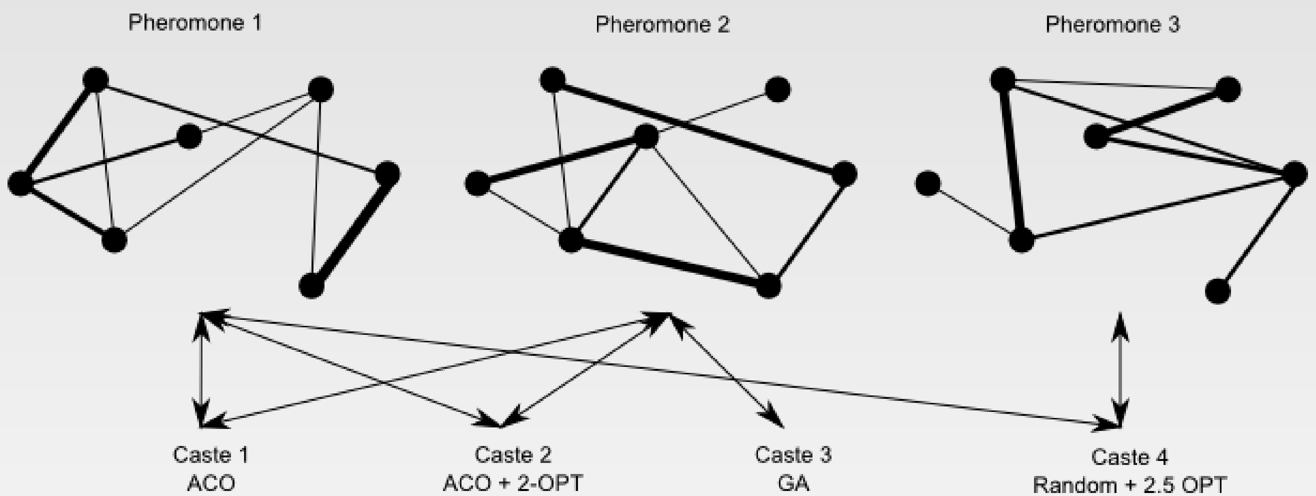
pheromone



best solution

Algorithm combination - parallel

- exchanges (ants on TSP)



3 instances of solution space representations

4 algorithms working with representations

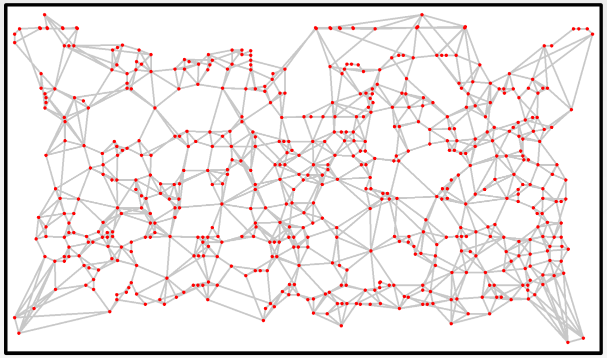
Parallel combination - example

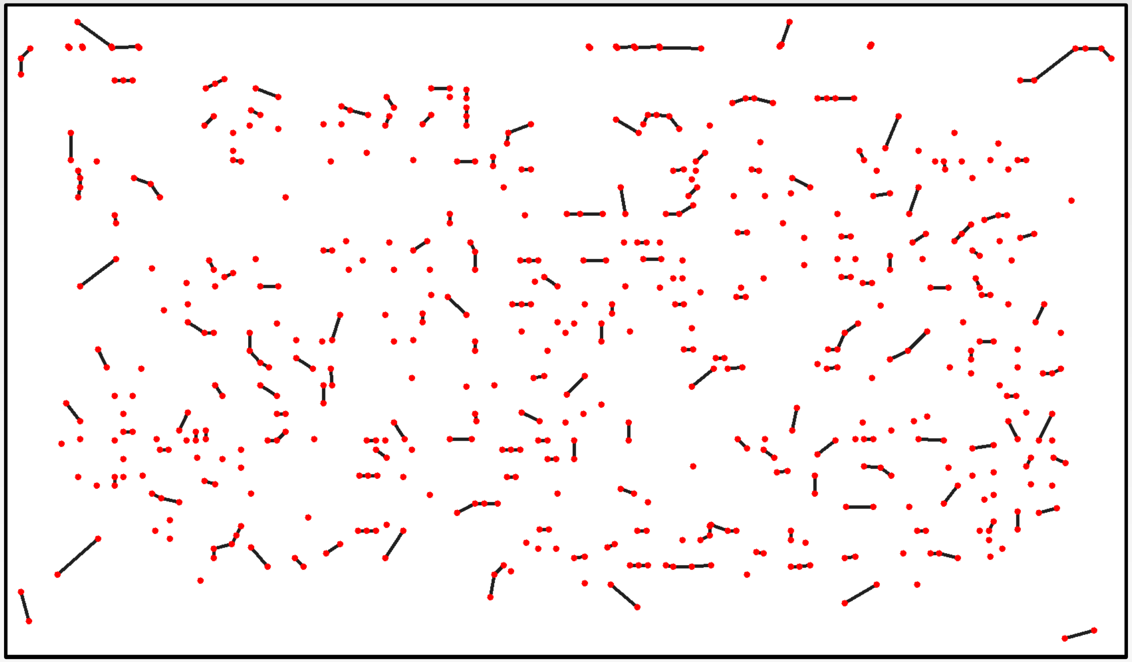
- Kovářík, Skrbek. Simulation of Ant Colonies with Hints Generated by Parallel Heuristics, Eurosim 2010

pheromone – probability of each edge to be present in good solution

Intersection of solutions
from 4 Simulated Annealing algorithms

Main
ant algorithm
pheromone matrix

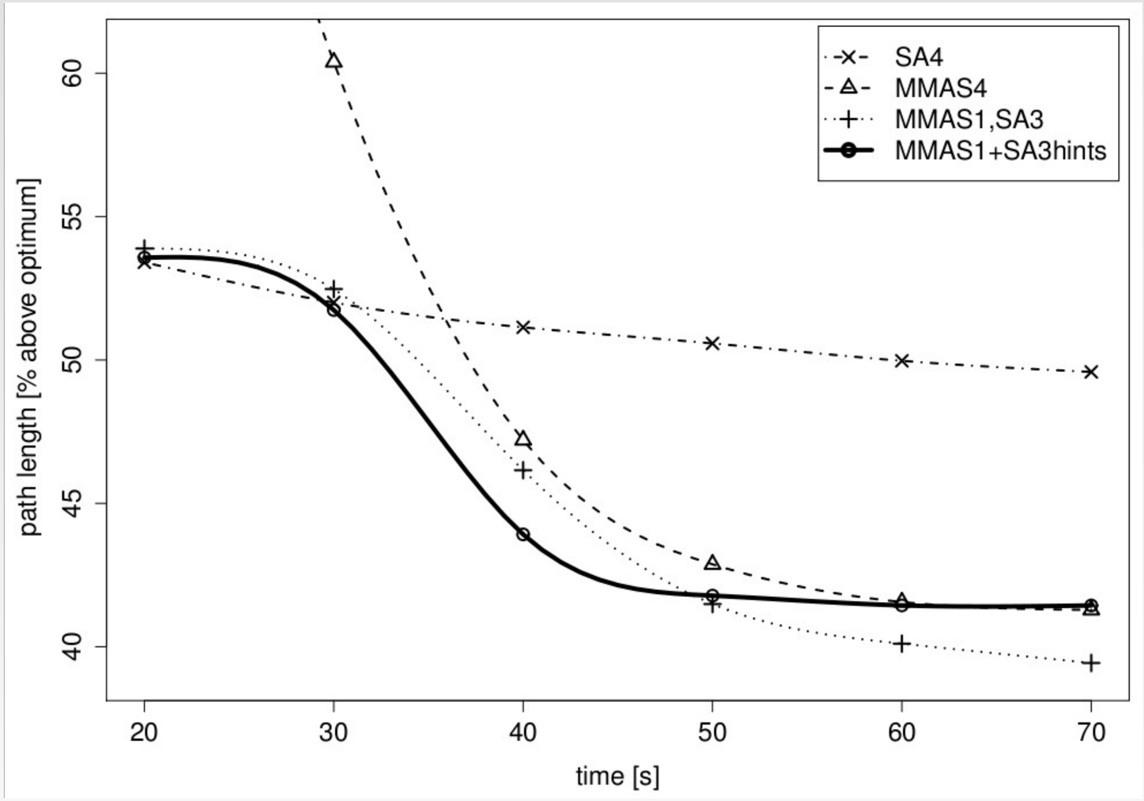




similar to "tour merging"

Parallel combination - example

- only for random distance matrix instances
- only in limited time interval



How to improve results?

- find best performing algorithm
  - can be temporarily achieved e.g. by new speed-up technique
  - difficulties on theoretic level: No Free Lunch Theorem for Search/Optimization (Wolpert and Macready 1997)
- specialization
  - learn from previous experiments
  - recommend optimization settings

Part 2: Meta-optimization

Meta-learning

- Meta-learning: Automatic learning from meta-data
- Meta-data:
  - data about experiments
  - characteristics of data
- Target: improving results of learning algorithms

Meta-optimization

- Meta-learning applied to optimization problems/experiments
- Target: for given data recommend
  - algorithm
  - parameters
  - combination of algorithms
  - strategy of data exchange

Meta-data

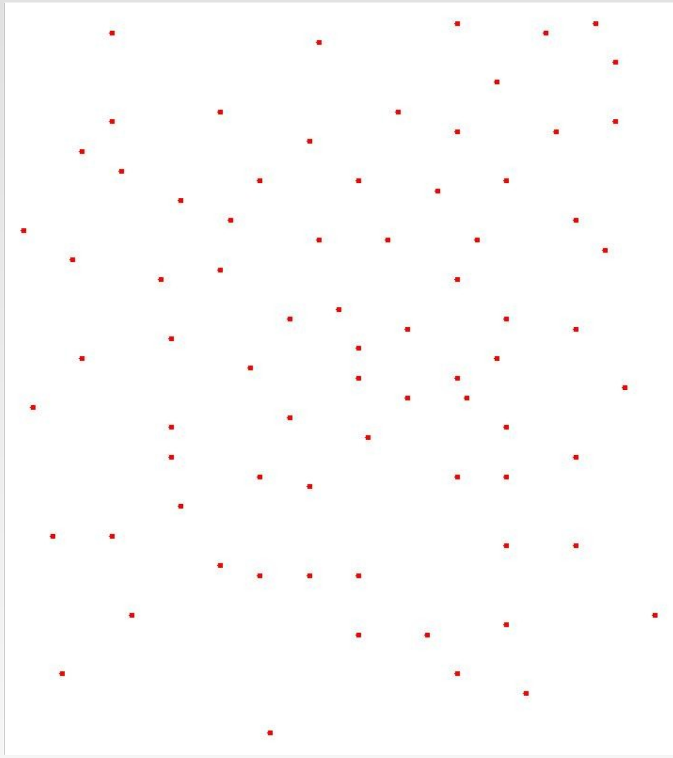
- static
  - e.g. statistical descriptors of input or results
  - helps decide what to do with given data
- dynamic
  - information from optimization in progress
  - helps decide how to adjust optimization
    - parameter adjustment
    - switch to more suitable algorithm
    - stop optimization

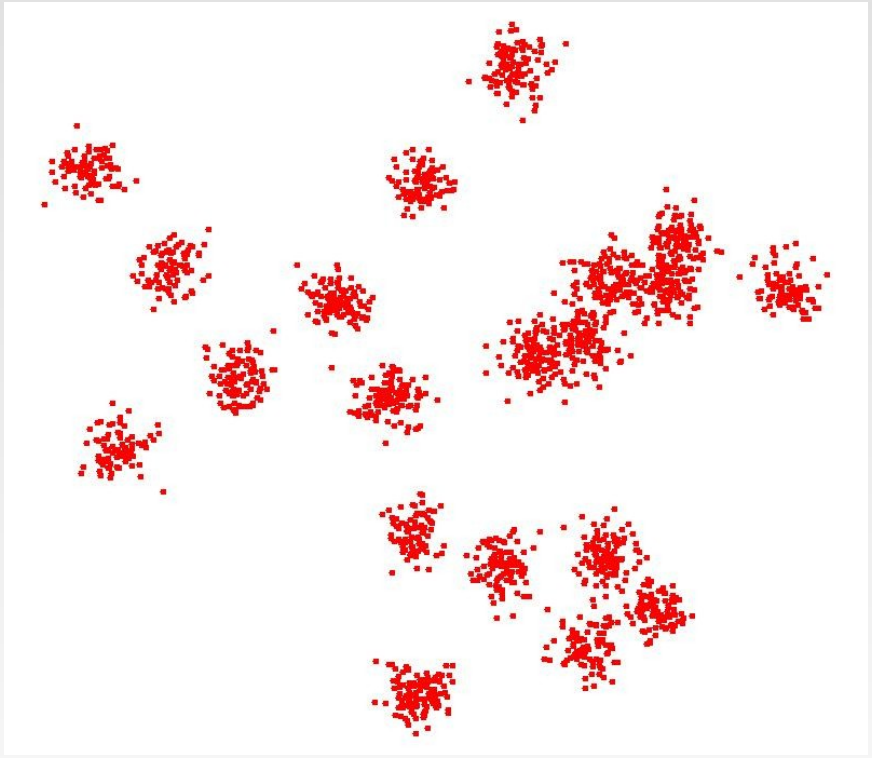
Meta-data: TSP

- Features available for TSP (general + Euclidean)
  - size
  - symmetric, geometric
  - average & std. deviation of edge lengths
  - edge length diversity
  - nearest neighbor distance, direction distribution
  - analytic & empiric lower bounds
  - max & min edge length, diameter
  - Held-Karp Lower Bound estimate and its estimate
  - construction heuristics (e.g. greedy) solution length
  - results of heuristics with fast configuration sampled after 10s
  - determinant of distance matrix
  - height/width ratio
  - ...

TSP can vary in city distribution

- City distribution can be uniform vs. clustered

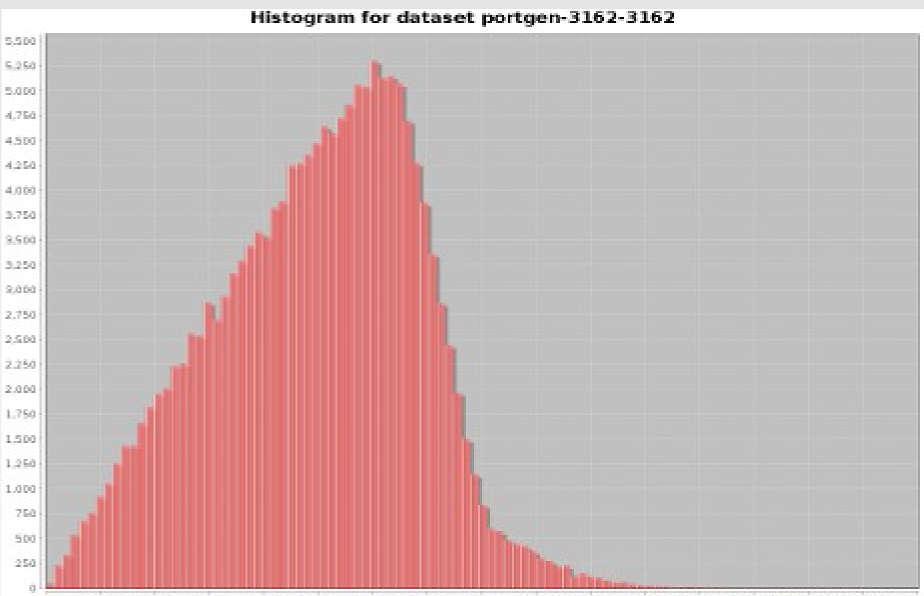




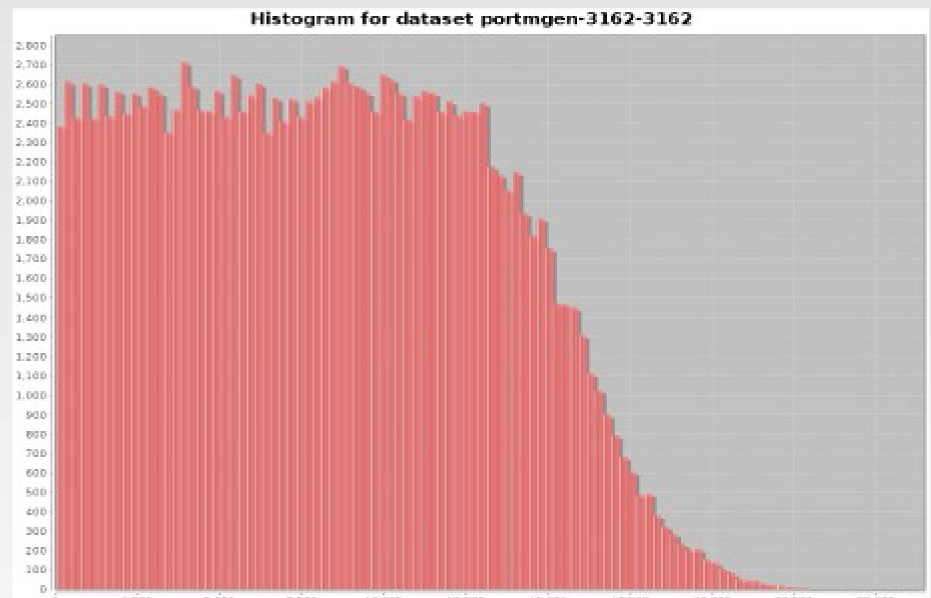
TSP can vary in distance distribution

- Distance matrix can be euclidean or random

Histograms of city distances (20 nearest neighbors)



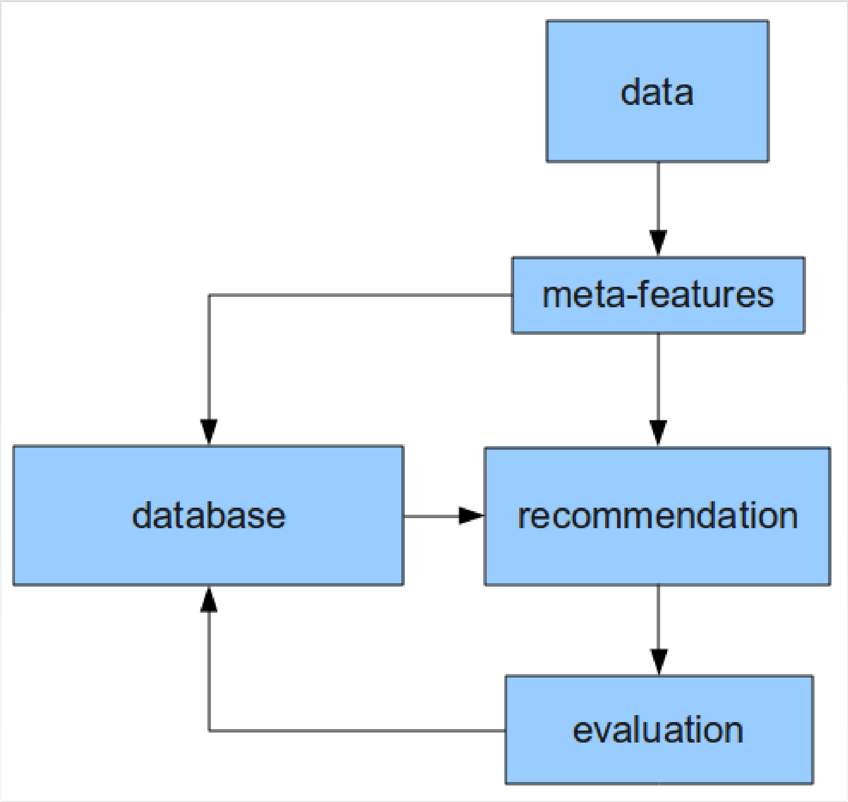
Random euclidian instance



Random instance
(random matrix)

Meta-data extraction

- Perform automatic experiments to fill database



On-line TSP meta-data

- http://kovarik.felk.cvut.cz/pheromone/data.php

Home Meta-data

Data

Results for heuristics are average values from 30 runs.

label	size	source	symm	geom	avgEdge	nNNd	Analytic Lower Bound	Empiric Lower Bound	Edge Diversity	edgeStdDev	diameterMin	diameterMax	HKBound	LK4-HKEstimate	Greedy solution	Greedy solution 3-OPT	SA150	MMASC150	GA150	NM150	TABU150	optimum
swiss42	42	TSPLIB	T	F	112	0.6739	4.05	5.14	0.1347	375.42	1.94	4.54			1437 (12.88%)	1273 (0%)						1273
eil51	51	TSPLIB	T	T	32	0.6283	4.46	5.61	0.029	106.63	0.46	0.95			482 (13.15%)	427 (0.23%)	426 (0%)	426 (0%)	426 (0%)			426
berlin52	52	TSPLIB	T	T	575	0.6182	4.51	5.66	0.3058	2427.07	7.53	20.93			8181 (8.47%)	7542 (0%)	7542 (0%)	7542 (0%)	7542 (0%)			7542
st70	70	TSPLIB	T	T	52	0.6947	5.23	6.47	0.0242	202.59	0.56	1.05			796 (17.93%)	675 (0%)	675 (0%)	675 (0%)	675 (0%)			675
eil76	76	TSPLIB	T	T	33	0.5391	5.45	6.72	0.014	136.51	0.32	0.64			608 (13.01%)	541 (0.56%)	538 (0%)	538 (0%)	538 (0%)			538
pr76	76	TSPLIB	T	T	7558	0.5716	5.45	6.72	0.2628	33896.97	71.98	167.15			130921 (21.04%)	108159 (0%)	108159 (0%)	108159 (0%)	108159 (0%)			108159
rat99	99	TSPLIB	T	T	84	0.5843	6.22	7.59	0.022	463.34	0.63	1.22			1437 (18.66%)	1212 (0.08%)						1211
kroA100	100	TSPLIB	T	T	1710	0.6078	6.25	7.63	0.2634	9114.44	12.87	23.01		18993 (-10.76%)	24698 (16.05%)	21282 (0%)	21282 (0%)	21282 (0%)	21282 (0%)			21282
kroB100	100	TSPLIB	T	T	1687	0.6504	6.25	7.63	0.2629	9083.24	12.65	24.67			25884 (16.91%)	22199 (0.26%)	22141 (0%)	22142 (0%)	22141 (0%)			22141
kroC100	100	TSPLIB	T	T	1700	0.6283	6.25	7.63	0.2664	9061.74	13.08	22.87		18557 (-10.56%)	23660 (14.03%)	20749 (0%)	20749 (0%)	20749 (0%)	20749 (0%)			20749
kroD100	100	TSPLIB	T	T	1631	0.5701	6.25	7.63	0.2558	8628.63	12.04	22.77			24852 (16.71%)	21330 (0.17%)	21294 (0%)	21295 (0%)	21294 (0%)			21294
kroE100	100	TSPLIB	T	T	1732	0.5714	6.25	7.63	0.2664	9288.75	13.02	23.84			24782 (12.3%)	22068 (0%)	22068 (0%)	22068 (0%)	22068 (0%)			22068
rd100	100	TSPLIB	T	T	555	0.5606	6.25	7.63	0.109	2615.01	4.22	7.74		7056 (-10.8%)	9423 (19.13%)	7910 (0%)	7910 (0%)	7910 (0%)	7910 (0%)			7910
eil101	101	TSPLIB	T	T	33	0.6119	6.28	7.66	0.0087	163.78	0.25	0.53		559 (-11.13%)	746 (18.6%)	630 (0.16%)	629 (0%)	629 (0%)	629 (0%)			629
lin105	105	TSPLIB	T	T	1177	0.5711	6.4	7.8	0.1895	6841.39	8.22	16.89		13250 (-7.85%)	16935 (17.78%)	14379 (0%)	14379 (0%)	14379 (0%)	14379 (0%)			14379
pr107	107	TSPLIB	T	T	5404	0.5921	6.47	7.87	0.0788	31970.58	45.04	58.49		35157 (-20.64%)	46680 (5.37%)	44303 (0%)	44303 (0%)	44303 (0%)	44303 (0%)			44303
pr124	124	TSPLIB	T	T	5623	0.5887	6.96	8.43	0.3204	31590.84	34.17	62.3			67055 (13.59%)	59076 (0.08%)						59030
bier127	127	TSPLIB	T	T	4952	0.5872	7.04	8.53	0.1324	34595.92	26.31	98.97		97313 (-17.73%)	133953 (13.25%)	118326 (0.04%)	118282 (0%)	118282 (0%)	118282 (0%)			118282
ch130	130	TSPLIB	T	T	356	0.5571	7.13	8.62	0.0466	1930.62	2.01	3.95			7129 (16.68%)	6150 (0.65%)	6110 (0%)	6112 (0.03%)	6110 (0%)			6110
xqf131	131	VLSI	T	T	35	0.5763	7.15	8.65	0.006	235.88	0.19	0.53		497 (-11.88%)	629 (11.52%)	564 (0%)	564 (0%)	564 (0%)	564 (0%)	564 (0%)	564 (0%)	564
pr136	136	TSPLIB	T	T	6073	0.5768	7.29	8.81	0.0582	34222.83	35.23	58.73			114553 (18.37%)	100326 (3.67%)						96772
pr144	144	TSPLIB	T	T	5639	0.5862	7.5	9.04	0.1555	33643.89	30.29	49.31			60964 (4.15%)	58537 (0%)	58537 (0%)	58537 (0%)	58537 (0%)			58537
ch150	150	TSPLIB	T	T	359	0.5895	7.65	9.22	0.035	2067.14	1.78	3.18		5933 (-9.11%)	7113 (8.96%)	6549 (0.32%)	6528 (0%)	6533 (0.08%)	6528 (0%)			6528
kroA150	150	TSPLIB	T	T	1717	0.6046	7.65	9.22	0.1553	11218.19	8.57	15.57		23865 (-10.02%)	31479 (18.68%)	26624 (0.38%)	26524 (0%)	26542 (0.07%)	26524 (0%)			26524
kroB150	150	TSPLIB	T	T	1711	0.6064	7.65	9.22	0.1554	11259.33	8.64	15.65			31611 (20.98%)	26176 (0.18%)	26130 (0%)	26131 (0%)	26130 (0%)			26130
pr152	152	TSPLIB	T	T	6914	0.5451	7.71	9.28	0.1313	45078.41	32.76	60.15			79564 (7.98%)	74254 (0.78%)						73682
u159	159	TSPLIB	T	T	2827	0.5873	7.88	9.48	0.0382	18571.86	13.31	24.56			48589 (15.47%)	42080 (0%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	42080
si175	175	TSPLIB	T	F	273	0.6008	8.27	9.91	0.0092	859.39	1.38	1.84			22000 (2.77%)	21411 (0.02%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
brg180	180	TSPLIB	T	F	5028	0.5806	8.39	10.05	0.0002	49983.57	2.95	29.89			8890 (355.9%)	1950 (0%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
rat195	195	TSPLIB	T	T	116	0.5859	8.73	10.43	0.0079	890.74	0.44	0.84			2612 (12.44%)	2346 (0.99%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
d198	198	TSPLIB	T	T	962	0.5885	8.79	10.51	0.0653	11304.24	3.34	12.43		11814 (-25.13%)	17620 (11.66%)	15835 (0.35%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
kroA200	200	TSPLIB	T	T	1701	0.5791	8.84	10.56	0.0951	12941.01	6.43	11.77			34543 (17.62%)	29451 (0.28%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
kroB200	200	TSPLIB	T	T	1664	0.5693	8.84	10.56	0.0931	12588.27	6.19	11.74			35389 (20.22%)	29713 (0.94%)	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ts225	225	TSPLIB	T	T	7080	0.6564	9.38	11.17	0.0054	49715.49	23.09	41.86		112500 (-11.17%)	140486 (10.93%)	126643 (0%)	126643	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
tsp225	225	TSPLIB	T	T	183	0.5628	9.38	11.17	0.0097	1424.94	0.59	1.3			4581 (16.98%)	3934 (0.46%)	3919 (0.08%)	3924 (0.2%)	3919 (0.08%)			3916
pr226	226	TSPLIB	T	T	7503	0.5894	9.4	11.19	0.1306	55633.73	25.93	44.59			92552 (15.16%)	80867 (0.62%)						80369
xqg237	237	VLSI	T	T	52	0.5853	9.62	11.45	0.0025	448.56	0.16	0.34		902 (-11.48%)	1221 (19.82%)	1028 (0.88%)	1019 (0%)	1021 (0.2%)	1019 (0%)	1019 (0%)	1019 (0%)	1019
gil262	262	TSPLIB	T	T	101	0.5935	10.12	12.01	0.0037	779.65	0.29	0.59			2823 (18.71%)	2397 (0.8%)						2378

live demo…
“easy” problem size

On-line TSP heuristics results

Time: 80

Instance name: dcc1911

Go

results for instance: dcc1911

configuration	value	% above optimum
SimulatedAnneling, 3-OPT=true, factor=0.95 (1)	6416	0.313
TabuSearch, 3-OPT=true, tabuDuration=10 (6)	6419	0.36
MMASC, 3-OPT=true, tabuDuration=10, cstFile=01.cst (8)	6473	1.204
GeneticAlgorithm, 3-OPT=true, populationSize=10, elitistCount=1, mutationProbability=0.1, crossOverProbability=0.8 (10)	6425	0.453
NoisingMethod, 3-OPT=true, factor=0.9 (11)	6421	0.391
optimum	6396	

Results for heuristics are average values from 30 runs.

live demo...
ants on DIMACS "M" instances

Further development

- Automatic experimenter
- Generate experiment configurations from descriptions of parameters
- Target: automatic meta-data collection

- General task, will be applied in our other projects:
  - Continuous optimization
  - FAKE GAME – heterogeneous inductive neural network
  - Data preprocessing framework

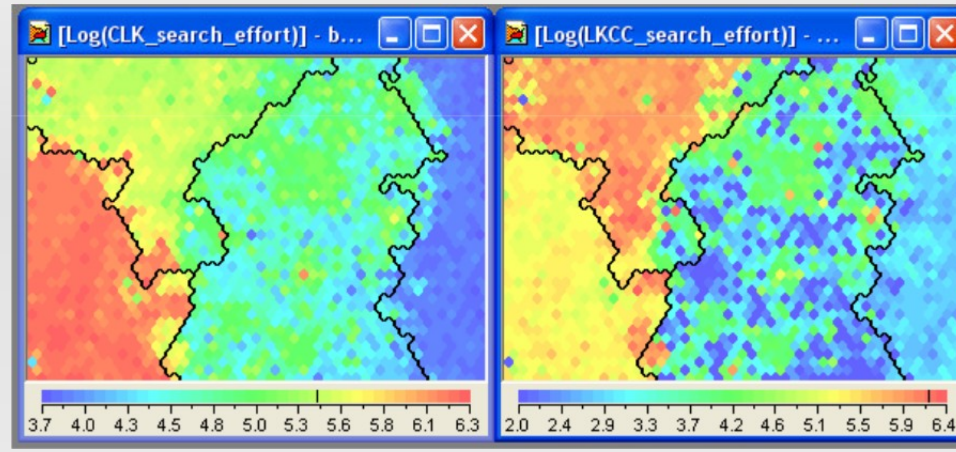
Meta-data usage: quick view

- correlation matrix

Attributes	size	geom	avgEdge	nNNd	Analy...	Empiri...	Edge ...	edgeS...	diame...	diame...	Greedy...	Gree...	SA150	MMASC150	GA150	optimum
nNNd	-0.203	0.183	-0.255	1	-0.271	-0.271	-0.022	-0.264	-0.224	-0.194	-0.159	-0.149	-0.147	-0.147	-0.148	-0.142
geom	-0.269	1	-0.460	0.183	-0.283	-0.283	-0.271	-0.540	-0.547	-0.231	0.103	0.148	0.153	0.158	0.151	0.175
Edge Diversity	-0.003	-0.271	0.630	-0.022	-0.009	-0.009	1	0.552	0.757	0.764	0.443	0.428	0.426	0.424	0.425	0.419
size	1	-0.269	0.556	-0.203	0.972	0.972	-0.003	0.599	0.387	0.351	0.452	0.435	0.432	0.432	0.434	0.423
Analytic Lower Bound	0.972	-0.283	0.588	-0.271	1	1.000	-0.009	0.616	0.438	0.401	0.467	0.450	0.446	0.446	0.448	0.437
Empiric Lower Bound	0.972	-0.283	0.588	-0.271	1.000	1	-0.009	0.616	0.438	0.401	0.467	0.450	0.446	0.446	0.448	0.437
diameterMin	0.387	-0.547	0.923	-0.224	0.438	0.438	0.757	0.869	1	0.934	0.602	0.572	0.568	0.566	0.568	0.554
edgeStdDev	0.599	-0.540	0.982	-0.264	0.616	0.616	0.552	1	0.869	0.797	0.707	0.675	0.670	0.669	0.672	0.654
diameterMax	0.351	-0.231	0.886	-0.194	0.401	0.401	0.764	0.797	0.934	1	0.735	0.720	0.717	0.716	0.716	0.709
avgEdge	0.556	-0.460	1	-0.255	0.588	0.588	0.630	0.982	0.923	0.886	0.770	0.743	0.738	0.737	0.739	0.724
Greedy solution	0.452	0.103	0.770	-0.159	0.467	0.467	0.443	0.707	0.602	0.735	1	0.999	0.999	0.998	0.999	0.997
Greedy solution 3-OP	0.435	0.148	0.743	-0.149	0.450	0.450	0.428	0.675	0.572	0.720	0.999	1	1.000	1.000	1.000	1.000
GA150	0.434	0.151	0.739	-0.148	0.448	0.448	0.425	0.672	0.568	0.716	0.999	1.000	1.000	1.000	1	1.000
SA150	0.432	0.153	0.738	-0.147	0.446	0.446	0.426	0.670	0.568	0.717	0.999	1.000	1	1.000	1.000	1.000
MMASC150	0.432	0.158	0.737	-0.147	0.446	0.446	0.424	0.669	0.566	0.716	0.998	1.000	1.000	1	1.000	1.000
optimum	0.423	0.175	0.724	-0.142	0.437	0.437	0.419	0.654	0.554	0.709	0.997	1.000	1.000	1.000	1.000	1

Meta-data usage: extracting rules

- **Smith-Miles** BAO2010
  - SOM for meta-data



  - Decision trees (C4.5) for classification to classes Easy and Hard for 2 algorithms

Rule 1	IF coeff. variation < 45.34 THEN Easy_CLK
Rule 2	IF coeff. variation >= 45.34 AND fraction distinct < 0.014 THEN Easy_LKCC
Rule 3	IF cluster ratio >= 0.05 THEN Class = Hard_CLK
Rule 4	IF fraction distinct >= 0.014 AND mean cluster radius >= 63.91 THEN Hard_LKCC
Rule 5	IF cluster ratio < 0.05 AND outlier ratio < 0.12 THEN Hard_CLK
Rule 6	IF cluster ratio < 0.05 AND outlier ratio >= 0.12 THEN Hard_LKCC

Diversity is necessary

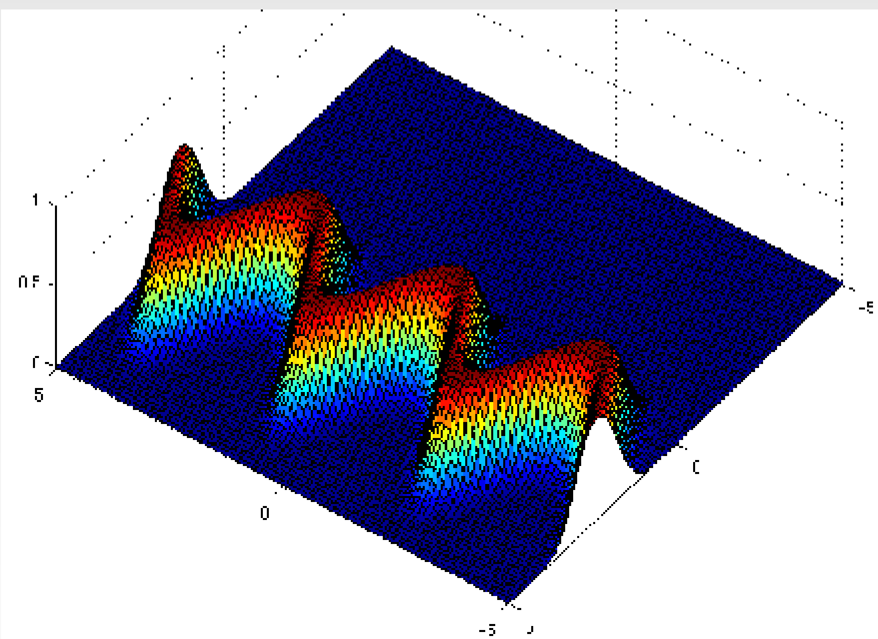
- Combining algorithms make sense only for set of algorithms which differ in exploration/exploitation strategy

- Benchmark data with low number of instances (TSPLIB, VLSI, WORLD) or only simple artificial data (DIMACS)

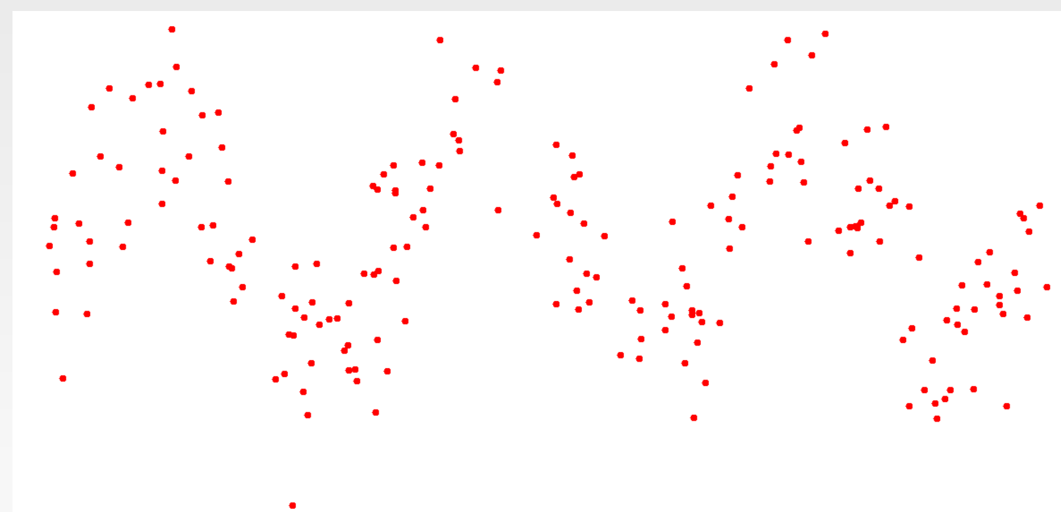
=> evolve data to examine differences between algorithms and generate large number of instances for meta-data mining

GPTSP: evolving interesting data

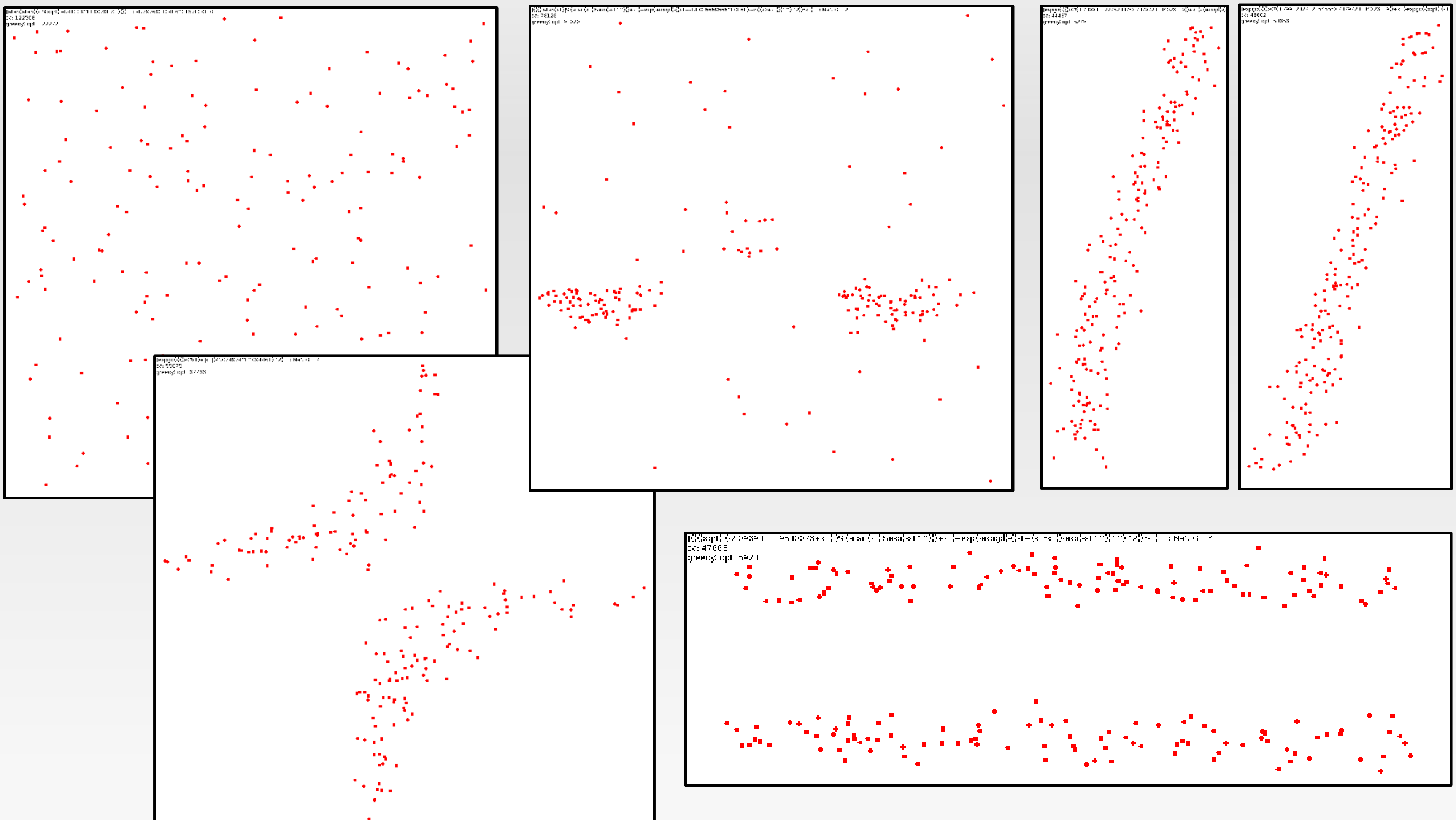
- Genetic Programming (Jan Drchal @ FEE CTU)
- evolve 2D functions = probability density functions for cities
- fitness average difference of results for 2 algorithms – competition



Greedy vs. Simulated Annealing champion
z = (((-1) + exp(-exp(-((y-1.84073)-sin((x+x))).^2).^2))*-1)



GPTSP: other evolved examples

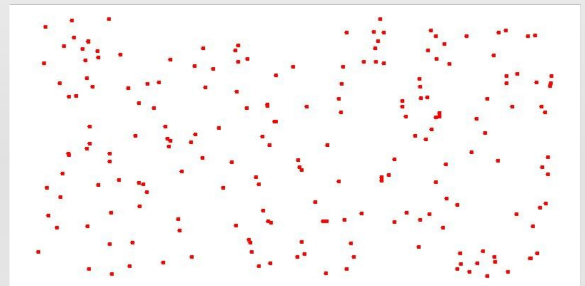


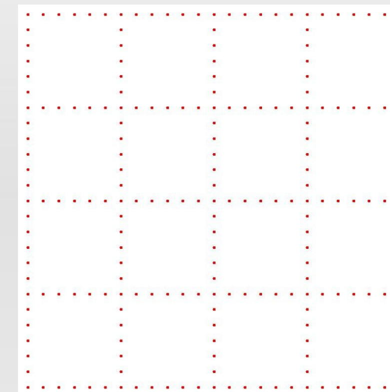
Visualizing solution space

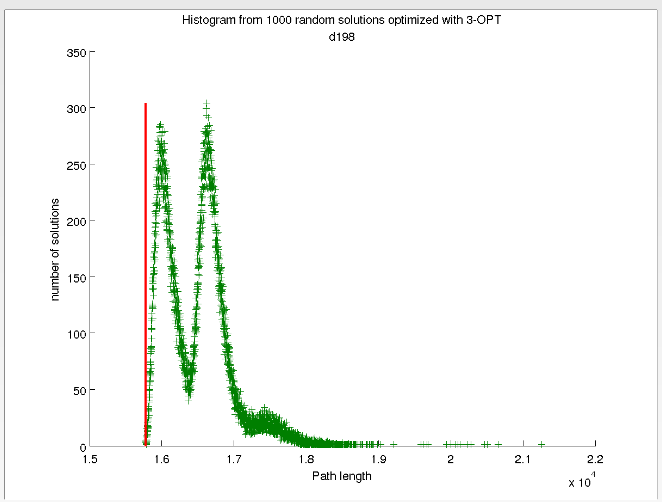
- continuous optimization
  - direct visualization or dimension reduction techniques
- combinatorial optimization
  - large neighborhoods, visualization problematic

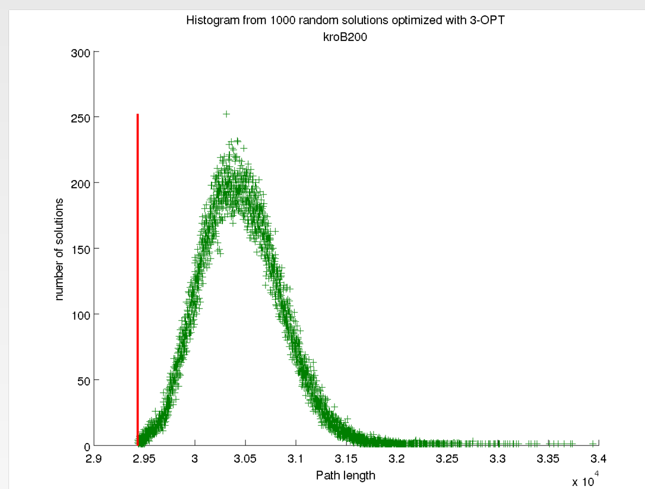
- indirect visualization
  - sample random solutions
  - examine distribution of solution quality

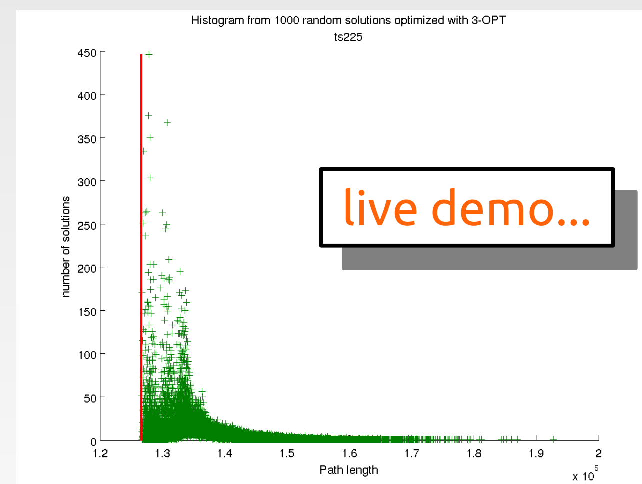
Visualizing solution space











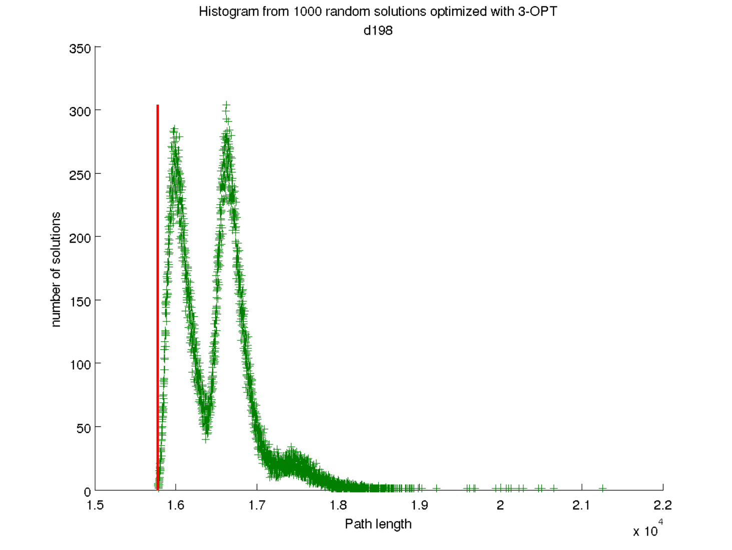
live demo...

Distribution of solutions

- For random geometric instances: Rayleigh distribution (Gremlich et al. 2006)
- 404 paper

Usage of distribution info

- more meta-data
- prediction of optima, possibly useful for setting bounds for branch&bound
- comparison of algorithms

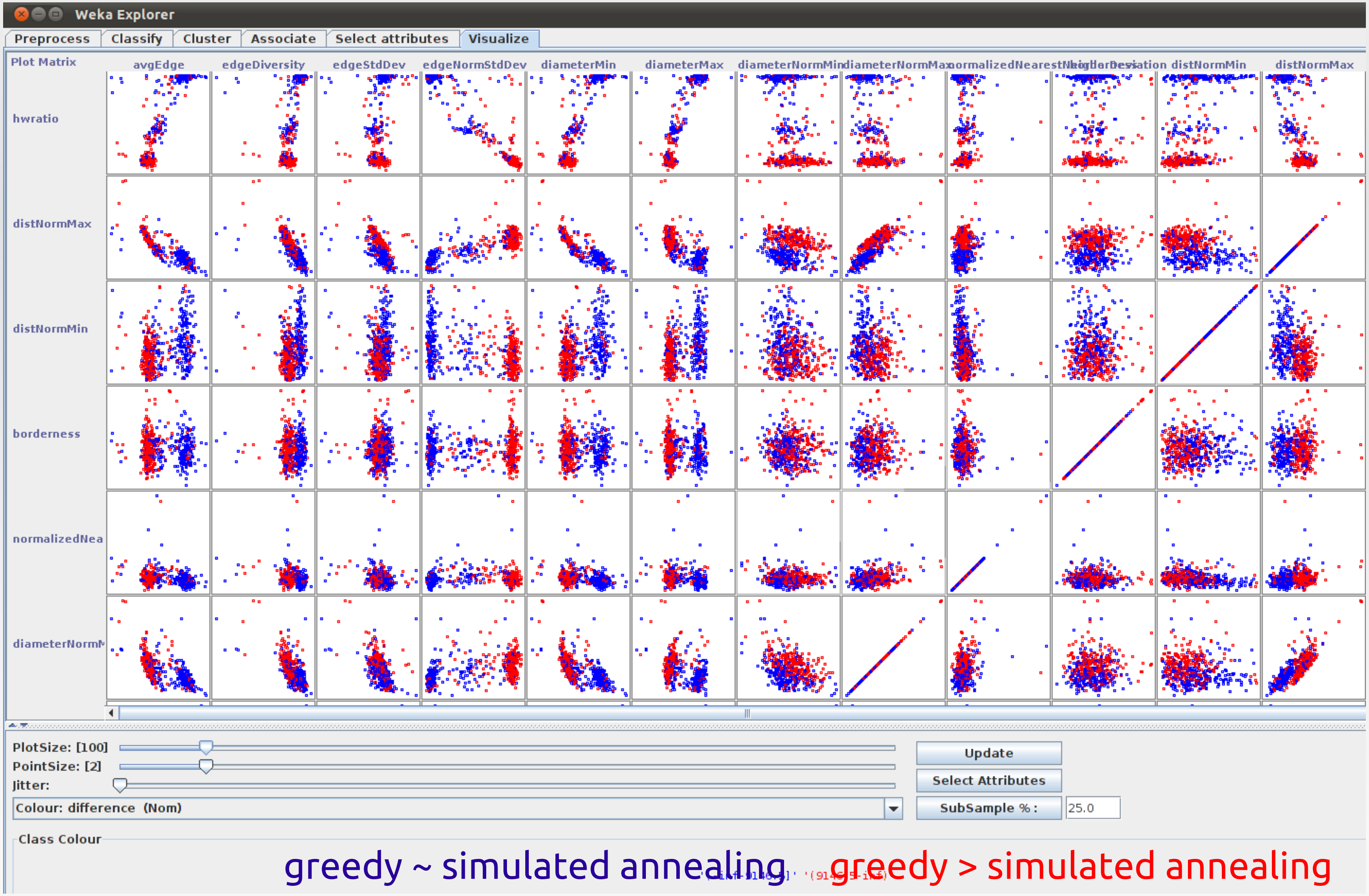


GPTSP meta-data ranking example

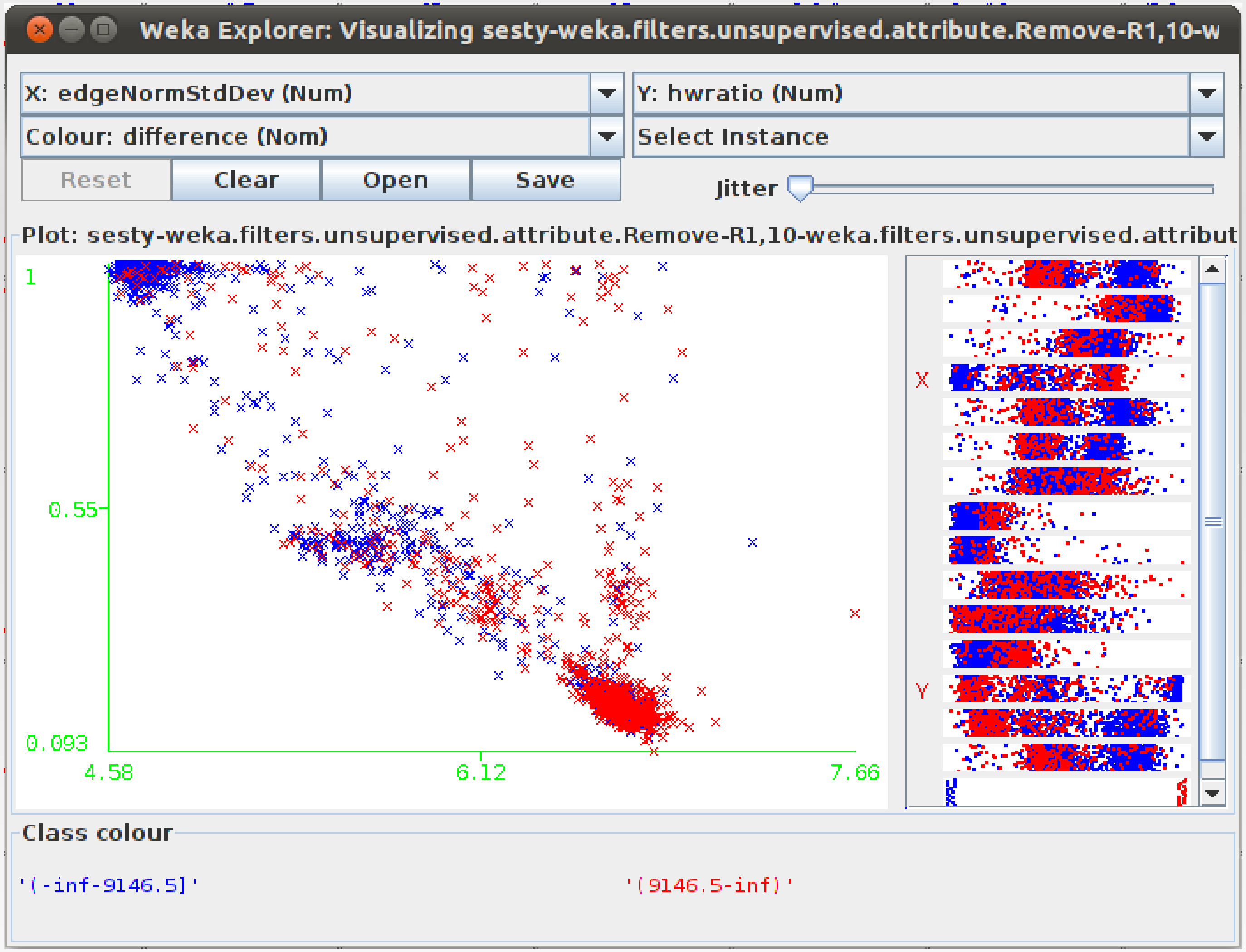
- Greedy algorithm vs. Simulated Annealing
- Weka: CfsSubsetEval – Exhaustive
  - Selected attributes: edgeNormStdDev, diameterMin, diameterMax, distNormMax
- Weka: GainRatioAttributeEval – Ranker

```
Ranked attributes:
 0.2082    4 edgeNormStdDev
 0.2057    1 avgEdge
 0.1967   13 hwratio
 0.195    12 distNormMax
 0.1936    6 diameterMax
 0.1914    5 diameterMin
 0.124     2 edgeDiversity
 0.0534    8 diameterNormMax
 0.0516    3 edgeStdDev
 0.034    11 distNormMin
 0        10 borderness
 0         7 diameterNormMin
 0         9 normalizedNearestNeigborDeviation
```

GPTSP meta-data



GPTSP meta-data



Attributes	difference
greedy3OptSolutionLength	-0.567
hwratio	-0.541
GRSolutionLength	-0.538
avgEdge	-0.492
greedySolutionLength	-0.491
diameterMax	-0.483
diameterMin	-0.478
AntiGRSolutionLength	-0.472
edgeDiversity	-0.300
distNormMin	-0.232
edgeStdDev	-0.144
normalizedNearestNeigborD	0.031
determinant	0.037
borderness	0.046
diameterNormMin	0.129
diameterNormMax	0.200
distNormMax	0.415
edgeNormStdDev	0.572
difference	1

correlations

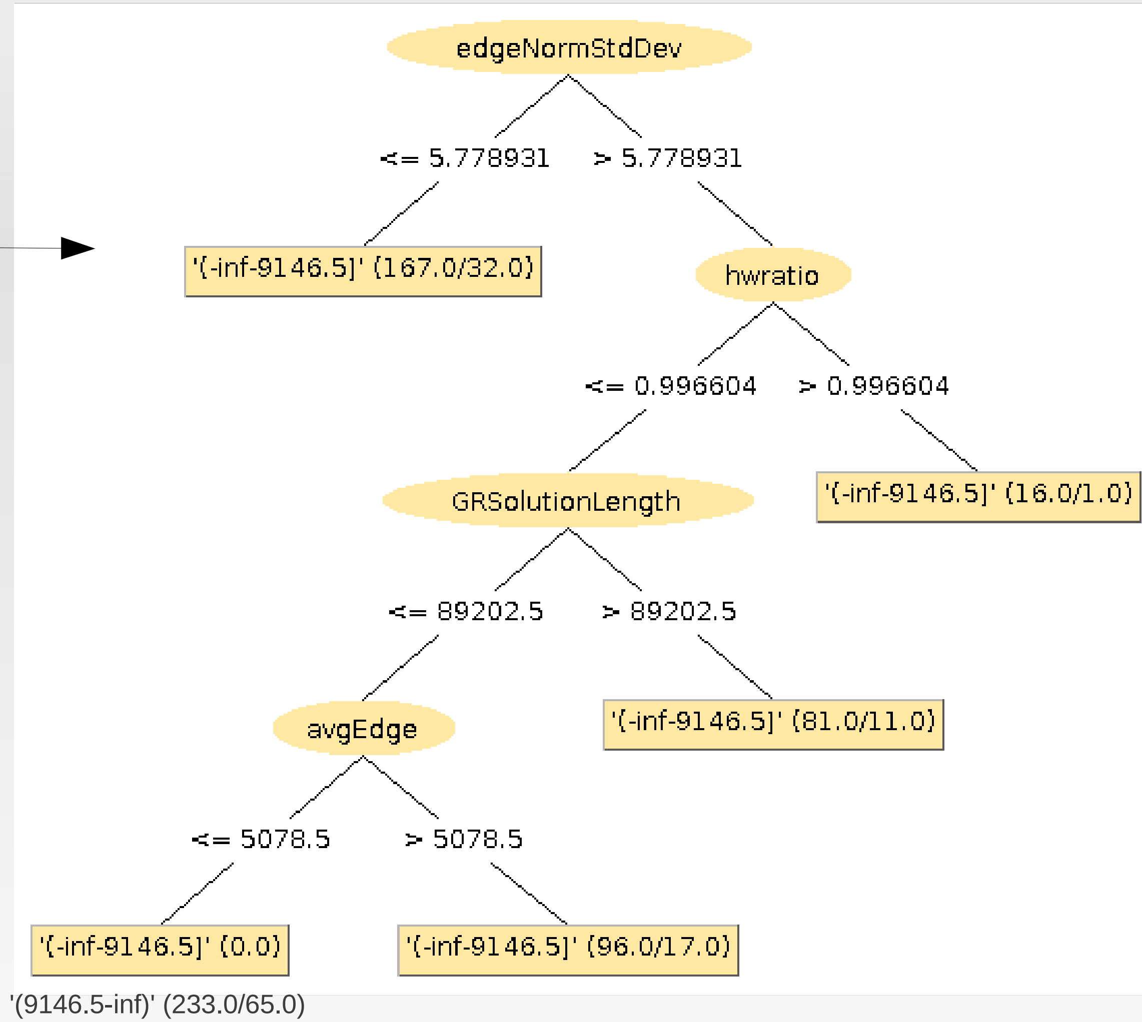
greedy ~ simulated annealing

greedy > simulated annealing

GPTSP meta-data trees & rules

- Weka:

J48graft pruned tree



Single conjunctive

rule learner:

(edgeNormStdDev <= 5.779428)

and (diameterMin <= 42.134596)

and (avgEdge > 3811)

=> difference = '(-inf-9146.5]'

Conclusion

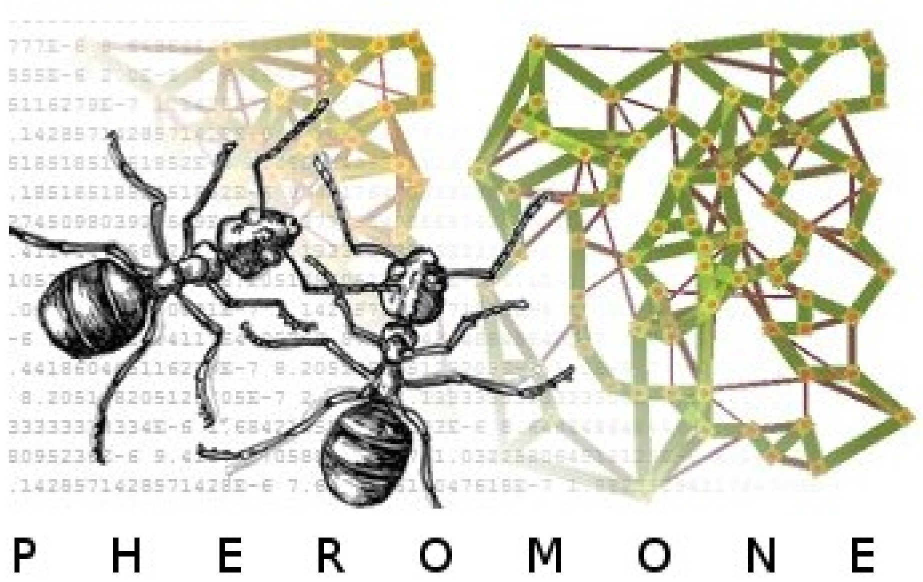
- Large amount of Ad-hoc experiments/comparisons for new algorithms – not very useful
- We need automatic experimenter to explore performance of algorithms/implementations
- We want to collect meta-data and use them to improve future optimization by specialization

Conclusion - Goal of project

- Framework for combinatorial optimization
  - JCOP (general) - http://jcop.sourceforge.net/
  - Pheromone (TSP) - http://kovarik.felk.cvut.cz/pheromone
- Module for automatic meta-data extraction

- Rules extraction =>
  - algorithm recommendation
  - parameter recommendation
  - serial/parallel algorithm combination scenarios recommendation

Thank you



http://kovarik.felk.cvut.cz/pheromone