Distance-based exponential probability models for constrained combinatorial optimization problems.

Josu Ceberio, Alexander Mendiburu, Jose A. Lozano

Intelligent Systems Group

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

XVIII Conference of the Spanish Association for Artificial Intelligence

23-26 October 2018, Granada, Spain

The Question

How do we approach problems with constrains by means of EDAs?
(taking into account their nature…)

Constrained Optimization Problems

Definition

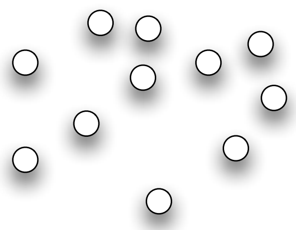
$$\begin{aligned} \text{minimising} \quad & f(\mathbf{x}), & \mathbf{x} = (x_1, \dots, x_n) \\ \text{subject to,} \quad & g_i(\mathbf{x}) \leq 0, & i = 1, \dots, r \\ & h_j(\mathbf{x}) = 0, & j = r+1, \dots, m \end{aligned}$$

Some examples

- → Knapsack Problem
- → Graph Colouring Problem
- → Maximum Satisfiability Problem
- → Capacitated Arc Routing Problem
- → …

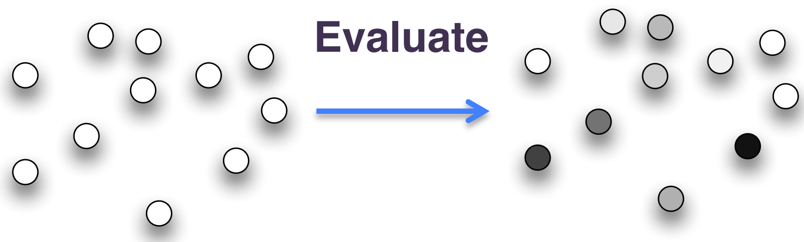
Estimation of distribution algorithms

Generate a set of solutions



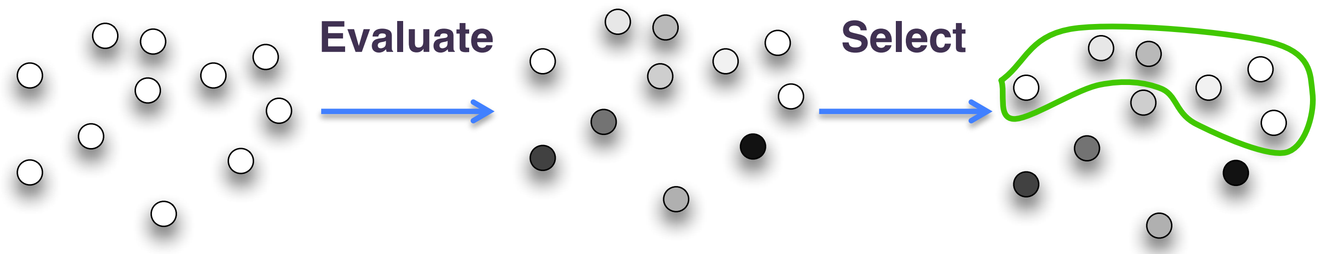
Estimation of distribution algorithms

Generate a set of solutions

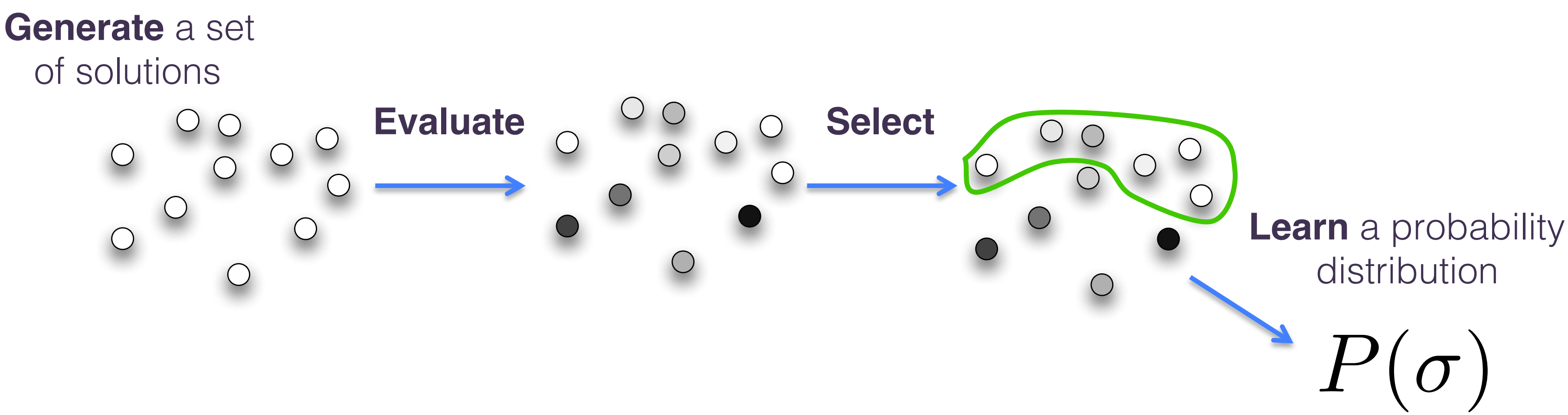


Estimation of distribution algorithms

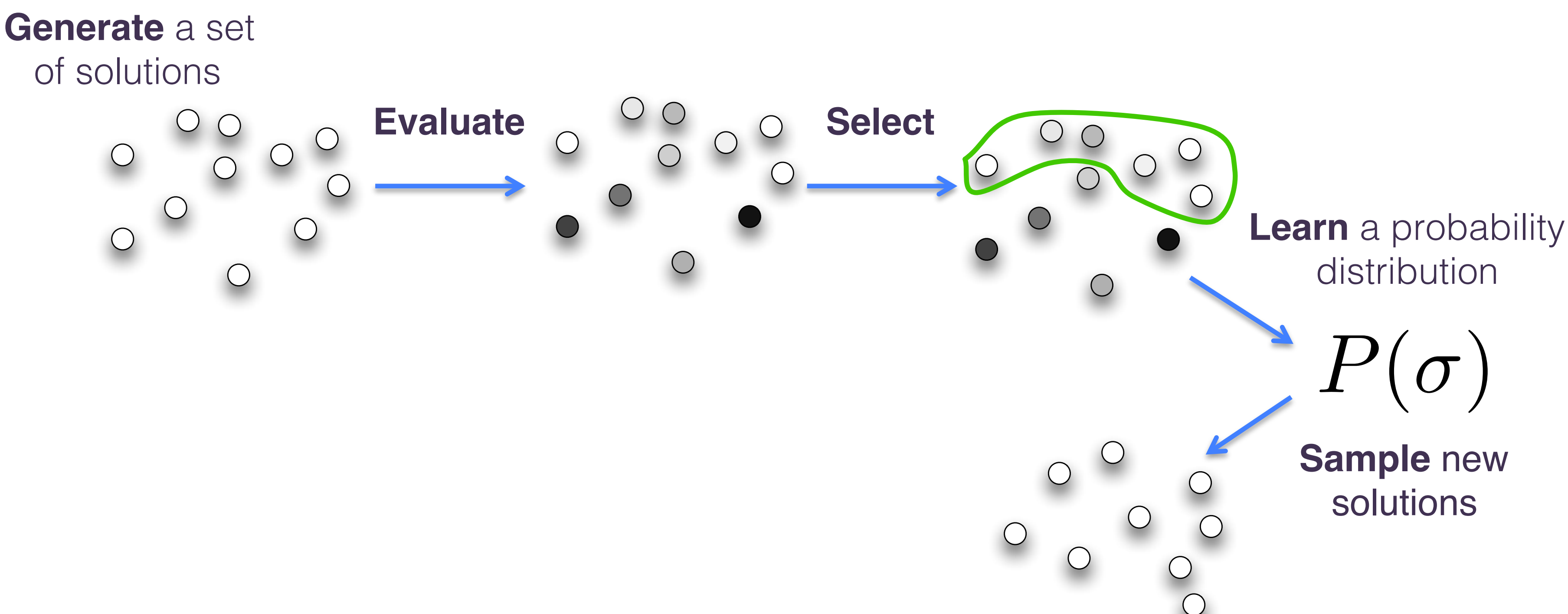
Generate a set of solutions



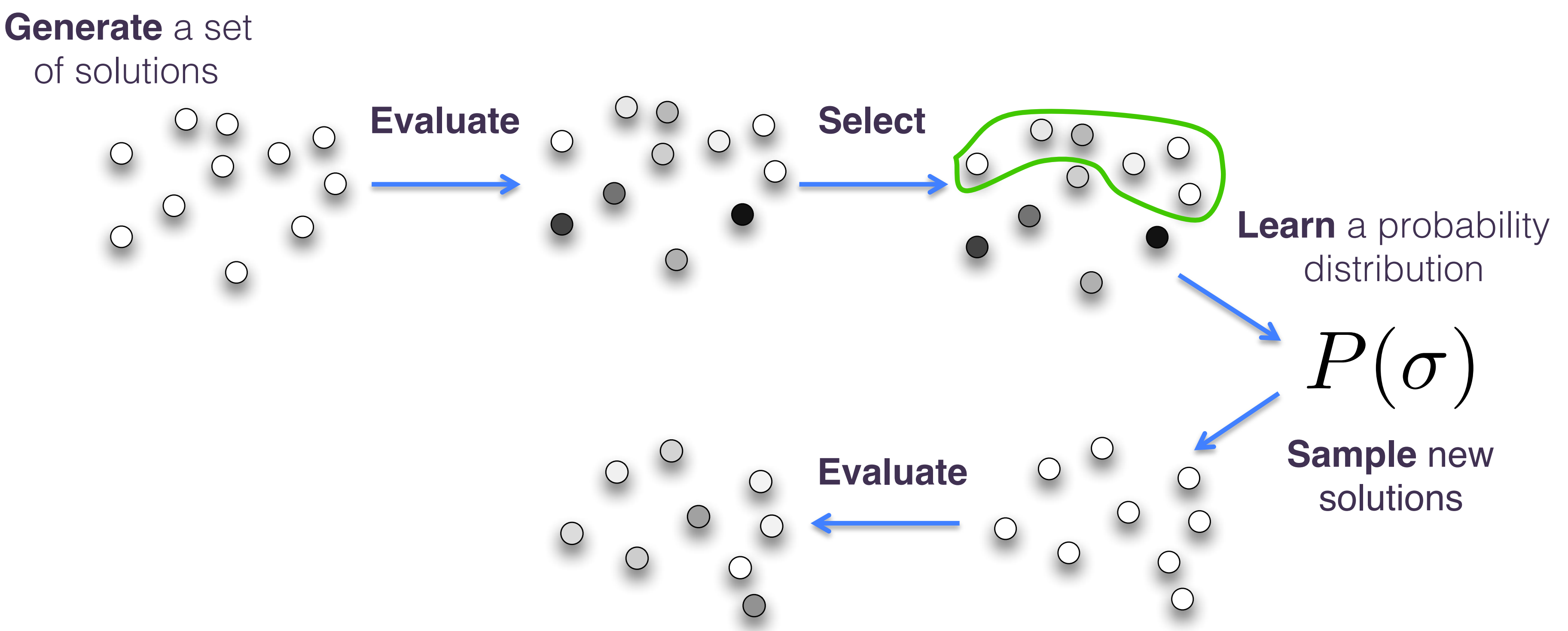
Estimation of distribution algorithms



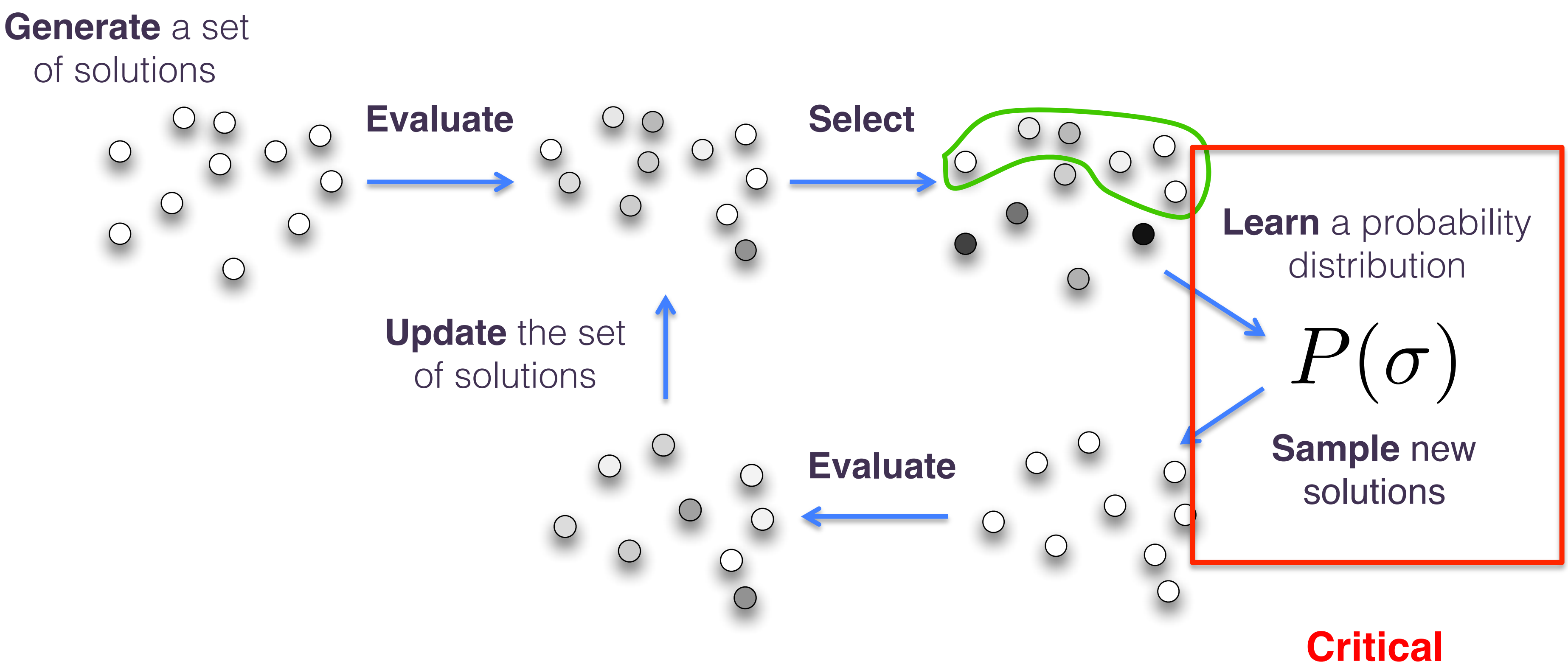
Estimation of distribution algorithms



Estimation of distribution algorithms



Estimation of distribution algorithms



EDAs reported
in the literature

Combinatorial Problems

UMDA [Mühlenbein, 1998]
MIMIC [DeBonet, 1997]
FDA [Mühlenbein, 1999]
EBNA [Etxeberria, 1999]
BOA [Pelikan, 2000]
EHBSA [Tsutsui, 2003]
NHBSA [Tsutsui, 2006]
TREE [Pelikan, 2007]
REDA [Romero, 2009]

Ω

Continuous Problems

UMDA_c [Larrañaga, 2000]
MIMIC_c [Larrañaga, 2000]
EGNA [Larrañaga, 2000]
EMNA [Larrañaga, 2001]
IDEA [Bosman, 2000]

$\mathbb{R}^n$

$\mathbb{S}_n$

Permutation Problems

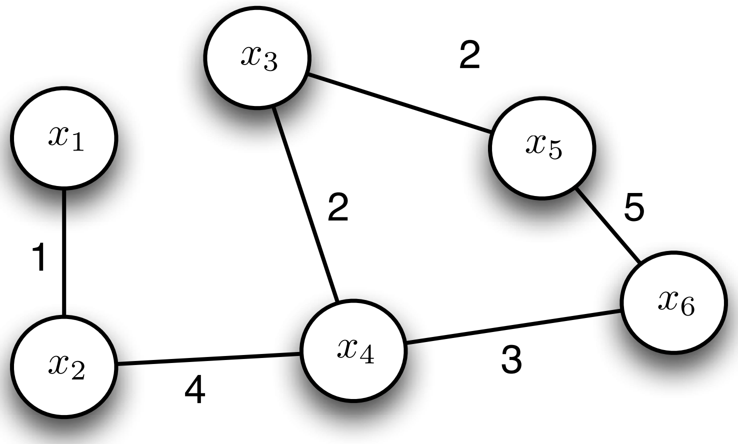
IDEA-ICE [Bosman, 2001]
MEDA [Ceberio, 2011]
PLEDA [Ceberio, 2013]
GMEDA [Ceberio, 2014]
RKEDA [Ayodele, 2016]

Constrained Problems

?

Graph Partitioning Problem

Definition



Find a *k*-partition of vertices minimising the weight of edges between sets: **the cut size**

We considered the balanced *2*-partition GPP.

Solutions are codified as…

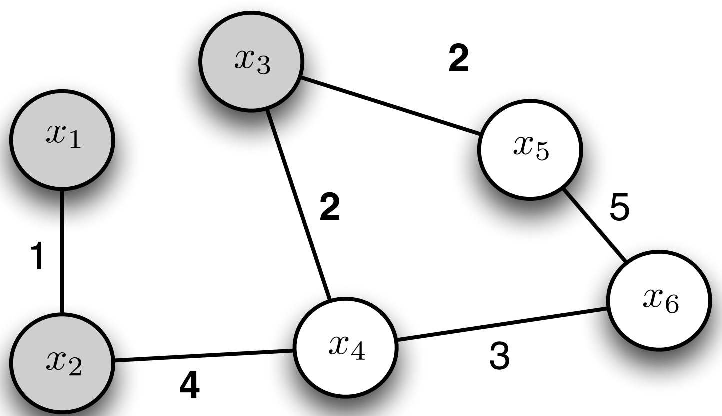
$$\mathbf{x} \in \{0,1\}^n$$

$$f(\mathbf{x}) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i(1 - x_j) w_{ij}$$

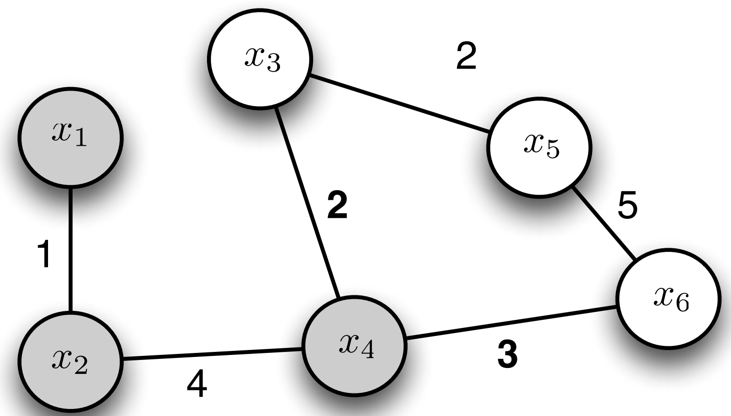
Objective Function

Graph Partitioning Problem

Example



$\mathbf{x}^1$	0	0	0	1	1	1	$f(\mathbf{x}^1) = 8$
	x_1	x_2	x_3	x_4	x_5	x_6	



$\mathbf{x}^2$	0	0	1	0	1	1	$f(\mathbf{x}^2) = 5$
	x_1	x_2	x_3	x_4	x_5	x_6	

The constraint: equal number of zeros as ones

$$\sum_{i=1}^{n} x_i = n/2$$

Graph Partitioning Problem

Why are they challenging?

The search space of solutions induced by the codification is…

000000	001000	010000	011000	100000	101000	110000	111000
000001	001001	010001	011001	100001	101001	110001	111001
000010	001010	010010	011010	100010	101010	110010	111010
000011	001011	010011	011011	100011	101011	110011	111011
000100	001100	010100	011100	100100	101100	110100	111100
000101	001101	010101	011101	100101	101101	110101	111101
000110	001110	010110	011110	100110	101110	110110	111110
000111	001111	010111	011111	100111	101111	110111	111111

Graph Partitioning Problem

Why are they challenging?

The search space of solutions induced by the codification is…

000000	001000	010000	011000	100000	101000	110000	111000
000001	001001	010001	011001	100001	101001	110001	111001
000010	001010	010010	011010	100010	101010	110010	111010
000011	001011	010011	011011	100011	101011	110011	111011
000100	001100	010100	011100	100100	101100	110100	111100
000101	001101	010101	011101	100101	101101	110101	111101
000110	001110	010110	011110	100110	101110	110110	111110
000111	001111	010111	011111	100111	101111	110111	111111

The majority of the solutions are not feasible !

What happens if we run a UMDA?

Univariate Marginal Distribution Algorithm (UMDA)

$$P(x) = \prod_{i=1}^{n} P(x_i)$$

First order marginals
No dependencies are considered

What happens if we run a UMDA?

	x_1	x_2	x_3	x_4	x_5	x_6
0	0.6	0.5	0.25	0.66	0.9	0.5
1	0.4	0.5	0.75	0.33	0.1	0.5

$\downarrow$

0 0 1 1 0 0

1 1 0 0 0 0

0 0 1 1 0 1

0 1 0 1 0 1

0 1 0 1 1 1

What happens if we run a UMDA?

	x_1	x_2	x_3	x_4	x_5	x_6
0	0.6	0.5	0.25	0.66	0.9	0.5
1	0.4	0.5	0.75	0.33	0.1	0.5

0 0 1 1 0 0

1 1 0 0 0 0

0 0 1 1 0 1

0 1 0 1 0 1

0 1 0 1 1 1

Unfeasible solutions are generated…

Different approaches

Literature review

1. Repair solutions

- Modify solutions to hold the constraints

2. Penalty functions

- Punish solutions to be discarded when selection

3. Guarantee feasibility when sampling

- In EDAs, adapt the sampling to create feasible solutions

The role of the probability model is somehow **denaturalized**

The Idea

How do we approach problems with constrains by means of EDAs?
(taking into account their nature…)

Conduct the optimisation entirely
on the set of feasible solutions…

4. Use probability distributions defined on this set

Motivation

Permutation-based Problems

Combinatorial Optimization Problems

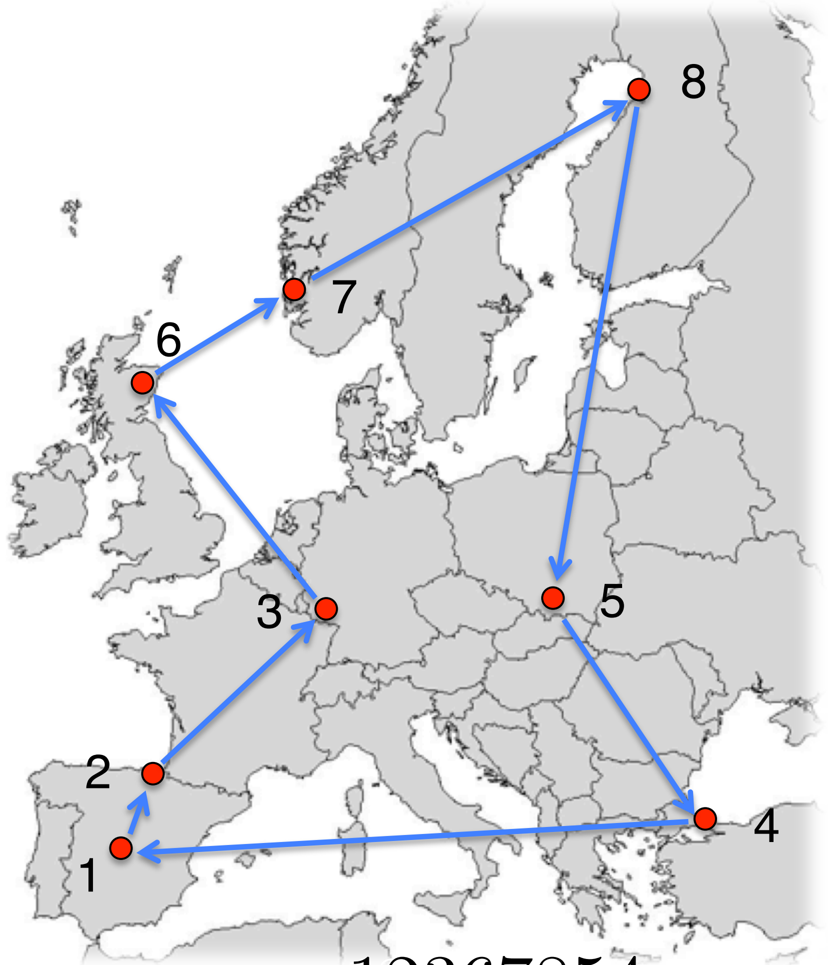
Whose solutions are represented as permutations

The search space consist of $n!$ solutions

$$8! = 40320$$

$$20! = 2.43 \times 10^{18}$$

Travelling Salesman Problem (TSP)



$$\sigma = 12367854$$

Motivation

Permutation-based Problems

The space of permutations can be seen as a constrained space of the integers space

$n = 3$

111	211	311
112	212	312
113	213	313
121	221	321
122	222	322
123	223	323
131	231	331
132	232	332
133	233	333

Motivation

Permutation-based Problems

The space of permutations can be seen as a constrained space of the integers space

$n = 3$

111	211	311
112	212	312
113	213	313
121	221	321
122	222	322
123	223	323
131	231	331
132	232	332
133	233	333

Motivation

Probability Models on Rankings

Bibliography

- M. A. Fligner and J. S. Verducci (1998), Multistage Ranking Models, *Journal of the American Statistical Association,* vol. 83, no. 403, pp. 892-901.
- D. E. Critchlow, M. A. Fligner, and J. S. Verducci (1991), Probability Models on Rankings, *Journal of Mathematical Psychology*, vol. 35, no. 3, pp. 294-318.
- P. Diaconis (1988), Group Representations in Probability and Statistics, *Institute of Mathematical Statistics.*
- M. A. Fligner and J. S. Verducci (1986), Distance based Ranking Models, *Journal of Royal Statistical Society, Series B*, vol. 48, no. 3, pp. 359-369.
- R. L. Plackett (1975), The Analysis of Permutations, *Applied Statistics*, vol. 24, no. 10, pp. 193-202.
- D. R. Luce (1959), Individual Choice Behaviour, *Wiley.*
- R. A. Bradley AND M. E. Terry (1952), Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons, *Biometrika*, vol. 39, no. 3, pp. 324-345.
- L. L. Thurstone (1927), A law of comparative judgment, *Psychological Review*, vol 34, no. 4, pp. 273-286.

Motivation

Probability Models on Rankings

$$P(\sigma) = \frac{1}{\psi(\theta)} e^{-\theta D(\sigma,\sigma_0)}$$

Mallows

Distance-based

$$P(\sigma) = \frac{1}{\psi(\theta)} e^{-\sum_{j=1}^{n-1} \theta_j S_j(\sigma,\sigma_0)}$$

Generalized Mallows

$$P(\sigma) = \prod_{i=1}^{n-1} \frac{w_{\sigma(i)}}{\sum_{j=i}^{n} w_{\sigma(j)}}$$

Plackett-Luce

Order statistics

$$P(\sigma) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \frac{w_{\sigma(i)}}{w_{\sigma(i)} + w_{\sigma(j)}}$$

Bradley-Terry

The Idea

Do probability models for constrained spaces exist?

No…

at least, we do not know them…

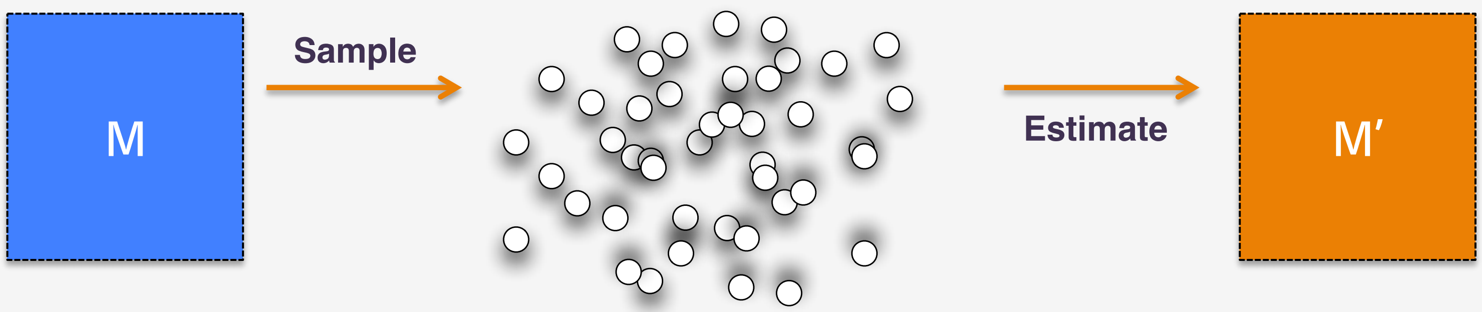
Designing probability models

Requirements

1 $\forall x \in \Omega, \quad 0 \leq P(x) \leq 1$

2 $\sum_{i=1}^{|\Omega|} P(x_i) = 1$

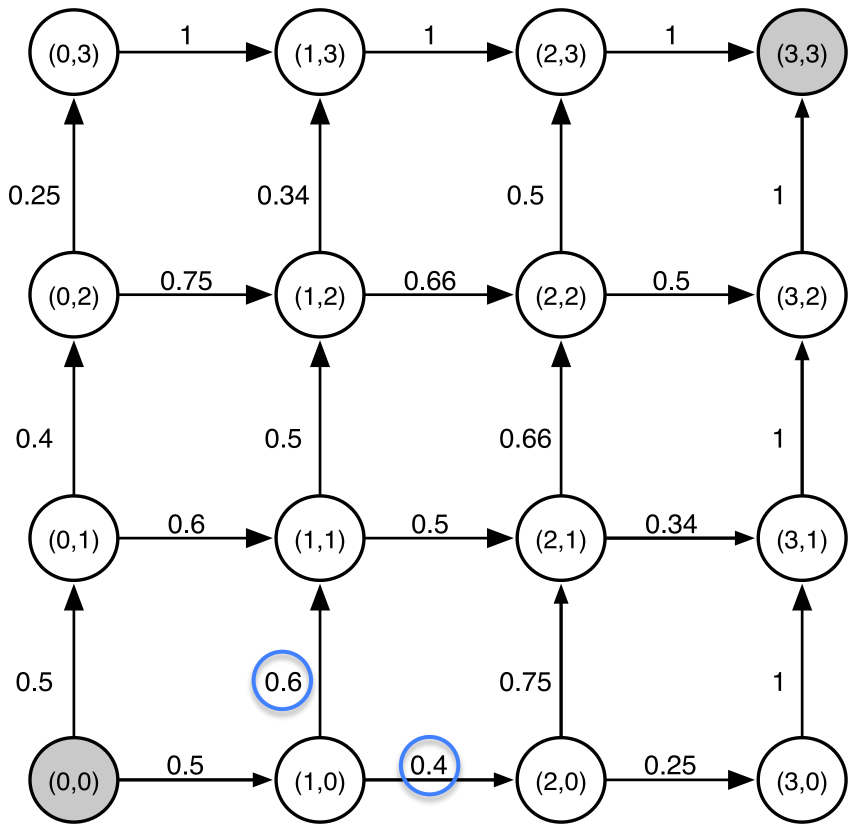
3 Develop efficient **learning** and **sampling** methods



$$M = M'$$

Previously: A Square Lattice

The Probability Model



Square Lattice for n=6 GPP instance

Solutions are modelled as paths on a square lattice of $(n/2+1)^2$ vertices

Vertices: the number of ones and zeros at that stage

Edges: the probability of moving from one vertex to another.

$$P(\pi) = \prod_{i=1}^{n} \underbrace{P^{(h^{i+1}-h^{i})}_{(h^{i},k^{i}),(h^{i}+1,k^{i})}}_{\text{Horizontal step}} \underbrace{P^{(k^{i+1}-k^{i})}_{(h^{i},k^{i}),(h^{i},k^{i}+1)}}_{\text{Vertical step}}$$

$$\mathbf{x} = (1,0,0,1,0,1)$$

$$\pi = ((0,1),(1,1),(2,1),(2,2),(3,2),(3,3))$$

Previously: A Square Lattice

The Probability Model

Suffer from...

1. Redundant model
2. The order?
3. Too many parameters
4. Difficult to analyze

$\pi = ((0,1),(1,1),(2,1),(2,2),(3,2),(3,3))$

The Idea

A different model

Distance-based exponential probability models

A distance-metric

Spread parameter

Central solution

$$P(x) = \frac{e^{-\theta d(x,\bar{x})}}{\sum_{y \in \Omega} e^{-\theta d(y,\bar{x})}}$$

The Challenges

A distance-metric is need

$$x, y \in \{0, 1\}^n$$

$$d(x, y)?$$

000000	001000	010000	011000	100000	101000	110000	111000
000001	001001	010001	011001	100001	101001	110001	111001
000010	001010	010010	011010	100010	101010	110010	111010
000011	001011	010011	011011	100011	101011	110011	111011
000100	001100	010100	011100	100100	101100	110100	111100
000101	001101	010101	011101	100101	101101	110101	111101
000110	001110	010110	011110	100110	101110	110110	111110
000111	001111	010111	011111	100111	101111	110111	111111

Hamming Distance

$$d_H(x, \bar{x}) = \sum_{i=1}^{n} \mathbf{1}_{[x(i) \neq x(j)]}$$

x	1	1	0	0	1	0
$\bar{x}$	1	0	0	0	1	1

$$d_H(x, \bar{x}) = ?$$

The Challenges

A distance-metric is need

$$x, y \in \{0, 1\}^n$$

$$d(x, y)?$$

000000	001000	010000	011000	100000	101000	110000	111000
000001	001001	010001	011001	100001	101001	110001	111001
000010	001010	010010	011010	100010	101010	110010	111010
000011	001011	010011	011011	100011	101011	110011	111011
000100	001100	010100	011100	100100	101100	110100	111100
000101	001101	010101	011101	100101	101101	110101	111101
000110	001110	010110	011110	100110	101110	110110	111110
000111	001111	010111	011111	100111	101111	110111	111111

Hamming Distance

$$d_H(x, \bar{x}) = \sum_{i=1}^{n} \mathbf{1}_{[x(i) \neq x(j)]}$$

x	1	1	0	0	1	0
$\bar{x}$	1	0	0	0	1	1

$$d_H(x, \bar{x}) = 2$$

The Challenges

A distance-metric is need

$$x, y \in \{0, 1\}^n$$

$$d(x, y)?$$

000000	001000	010000	011000	100000	101000	110000	111000
000001	001001	010001	011001	100001	101001	110001	111001
000010	001010	010010	011010	100010	101010	110010	111010
000011	001011	010011	011011	100011	101011	110011	111011
000100	001100	010100	011100	100100	101100	110100	111100
000101	001101	010101	011101	100101	101101	110101	111101
000110	001110	010110	011110	100110	101110	110110	111110
000111	001111	010111	011111	100111	101111	110111	111111

Hamming Distance

$$d_H(x, \bar{x}) = \sum_{i=1}^{n} \mathbf{1}_{[x(i) \neq x(j)]}$$

x	1	1	0	0	1	0
$\bar{x}$	0	0	1	1	0	1

$$d_H(x, \bar{x}) = ?$$

The Challenges

A distance-metric is need

$$x, y \in \{0, 1\}^n$$

$$d(x, y)?$$

000000	001000	010000	011000	100000	101000	110000	111000
000001	001001	010001	011001	100001	101001	110001	111001
000010	001010	010010	011010	100010	101010	110010	111010
000011	001011	010011	011011	100011	101011	110011	111011
000100	001100	010100	011100	100100	101100	110100	111100
000101	001101	010101	011101	100101	101101	110101	111101
000110	001110	010110	011110	100110	101110	110110	111110
000111	001111	010111	011111	100111	101111	110111	111111

Hamming Distance

$$d_H(x, \bar{x}) = \sum_{i=1}^{n} \mathbf{1}_{[x(i) \neq x(j)]}$$

x	1	1	0	0	1	0
$\bar{x}$	0	0	1	1	0	1

$$\cancel{d_H(x, \bar{x}) = 6}$$

$$d(x, \bar{x}) = 0$$

The Challenges

A distance-metric is need

$$x, y \in \{0, 1\}^n$$

$$d(x, y)?$$

000000	001000	010000	011000	100000	101000	110000	111000
000001	001001	010001	011001	100001	101001	110001	111001
000010	001010	010010	011010	100010	101010	110010	111010
000011	001011	010011	011011	100011	101011	110011	111011
000100	001100	010100	011100	100100	101100	110100	111100
000101	001101	010101	011101	100101	101101	110101	111101
000110	001110	010110	011110	100110	101110	110110	111110
000111	001111	010111	011111	100111	101111	110111	111111

New Distance (based on Hamming)

$$d(x, \bar{x}) = \min\{d_H(x, \bar{x}), d_H(\neg x, \bar{x})\}$$

Max distance

$$\lfloor n/2 \rfloor$$

$$\binom{n/2}{k/2}^2 \leftarrow$$

The number of solutions at distance ***k***

The Challenges

What about the normalization constant?

Distance-based exponential probability models

Spread parameter

A distance-metric

Central solution

$$P(x) = \frac{e^{-\theta d(x,\bar{x})}}{\underbrace{\sum_{y\in\Omega} e^{-\theta d(y,\bar{x})}}}$$

Normalization function

$$\sum_{l=0}^{K/2} \binom{n/2}{l}^2 e^{-\theta 2l}$$

The size of the search space. GPP:

$$K = \binom{n}{n/2}$$

The Challenges

How to estimate MLE parameters? In two steps…

$$L(\theta, \bar{x}|\mathbf{x}) = \prod_{i=1}^{N} \frac{e^{-\theta d(x_i, \bar{x})}}{\psi(\theta)}$$

← Find the parameters that maximize it

$$\log L(\theta, \bar{x}|\mathbf{x}) = -\theta N \bar{d} - N \log \left[\sum_{l=0}^{K/2} \binom{n/2}{l}^2 e^{-\theta 2l} \right]$$

1

Calculate $\bar{x}$ as average solution

1	1	0	0	1	0
0	1	0	1	1	0
1	0	0	0	1	1
1	1	0	0	1	0

The Challenges

How to estimate MLE parameters? In two steps…

$$L(\theta, \bar{x}|\mathbf{x}) = \prod_{i=1}^{N} \frac{e^{-\theta d(x_i, \bar{x})}}{\psi(\theta)}$$

← Find the parameters that maximize it

$$\log L(\theta, \bar{x}|\mathbf{x}) = -\theta N \bar{d} - N \log \left[\sum_{l=0}^{K/2} \binom{n/2}{l}^2 e^{-\theta 2l} \right]$$

1

Calculate $\bar{x}$ as average solution

1	1	0	0	1	0
0	1	0	1	1	0
1	0	0	0	1	1
1	1	0	0	1	0

0.75	0.75	0	0.25	1	0.25

The Challenges

How to estimate MLE parameters? In two steps…

$$L(\theta, \bar{x}|\mathbf{x}) = \prod_{i=1}^{N} \frac{e^{-\theta d(x_i, \bar{x})}}{\psi(\theta)}$$

← Find the parameters that maximize it

$$\log L(\theta, \bar{x}|\mathbf{x}) = -\theta N \bar{d} - N \log \left[\sum_{l=0}^{K/2} \binom{n/2}{l}^2 e^{-\theta 2l} \right]$$

1

Calculate $\bar{x}$ as average solution

1	1	0	0	1	0
0	1	0	1	1	0
1	0	0	0	1	1
1	1	0	0	1	0

1	1	0	0	1	0

2 Approximate θ numerically

$$\sum_{l=0}^{K/2} \binom{n/2}{l}^2 (2l - \bar{d}) e^{-\theta 2l} = 0$$

Newton-Raphson…

The Challenges

How to sample new solutions? Also in two steps…

1. A distance ***k*** at which generate a solution is sampled from

$$P(k) = \sum_{x|d(x,\bar{x})=k} P(x) = \binom{n/2}{k/2}^2 \frac{e^{-\theta k}}{\psi(\theta)}$$

2. Generate a solution at distance ***k***

$k = 4$

$\bar{x}$						
	1	1	0	0	1	0

x						
	1	1	0	0	1	0

The Challenges

How to sample new solutions? Also in two steps…

1 A distance ***k*** at which generate a solution is sampled from

$$P(k) = \sum_{x|d(x,\bar{x})=k} P(x) = \binom{n/2}{k/2}^2 \frac{e^{-\theta k}}{\psi(\theta)}$$

2 Generate a solution at distance ***k***

$$k = 4$$

$\bar{x}$	1	1	0	0	1	0
x	1	1	0	0	1	0

The Challenges

How to sample new solutions? Also in two steps…

1 A distance ***k*** at which generate a solution is sampled from

$$P(k) = \sum_{x \mid d(x,\bar{x})=k} P(x) = \binom{n/2}{k/2}^2 \frac{e^{-\theta k}}{\psi(\theta)}$$

2 Generate a solution at distance ***k***

$k = 4$

$\bar{x}$	1	1	0	0	1	0
x	1	0	1	0	0	1

Experimental Study

Experimental Setting

UMDA *, TREE*[1] ,
Lattice, Exp
(*) adapted sampling

Algorithms

Pop-size: 10n
Sel-size: 5n
Off-size: 10n
Max evals.: $100n^2$
10 repetitions

Parameters

22 Instances
(Johnson et al.)

G-type and
U-type

n=124, 250, 500,
1000

Benchmarks

[1]M. Pelikan, S. Tsutsui, and R. Kalapala, *Dependency Trees, Permutations and Quadratic Assignment Problem,* Medal Report No. 2007003 Tech. Rep. 2007.

Experimental Study

Results - Performance

Instance	Best Fitness	ARPD Exp	ARPD Lattice	ARPD UMDA	ARPD Tree
G124.02	13	0,32	0,32	0,61	0,19
G124.16	449	0,04	0,02	0,05	0,01
G250.01	31	0,40	0,33	0,49	0,20
G250.02	118	0,10	0,07	0,14	0,06
G250.04	360	0,04	0,04	0,10	0,03
G250.08	830	0,05	0,01	0,05	0,01
G500.005	61	0,38	0,30	0,40	0,08
G500.01	234	0,14	0,09	0,21	0,07
G500.02	642	0,06	0,03	0,11	0,03
G500.04	1754	0,05	0,02	0,06	0,02
G1000.0025	131	1,16	2,96	3,20	0,74
G1000.005	496	0,52	1,22	1,28	0,88
G1000.01	1420	0,28	0,56	0,66	0,62
G1000.02	3450	0,18	0,35	0,40	0,39
U500.05	23	1,34	1,17	1,89	0,57
U500.10	61	1,50	1,05	1,12	0,57
U500.20	185	0,89	0,56	0,87	0,44
U500.40	412	0,81	0,41	0,38	0,28
U1000.05	77	4,82	1,62	12,83	2,39
U1000.10	170	5,09	1,67	11,67	3,73
U1000.20	352	5,23	1,67	10,58	4,94
U1000.40	862	4,09	1,53	3,24	2,29

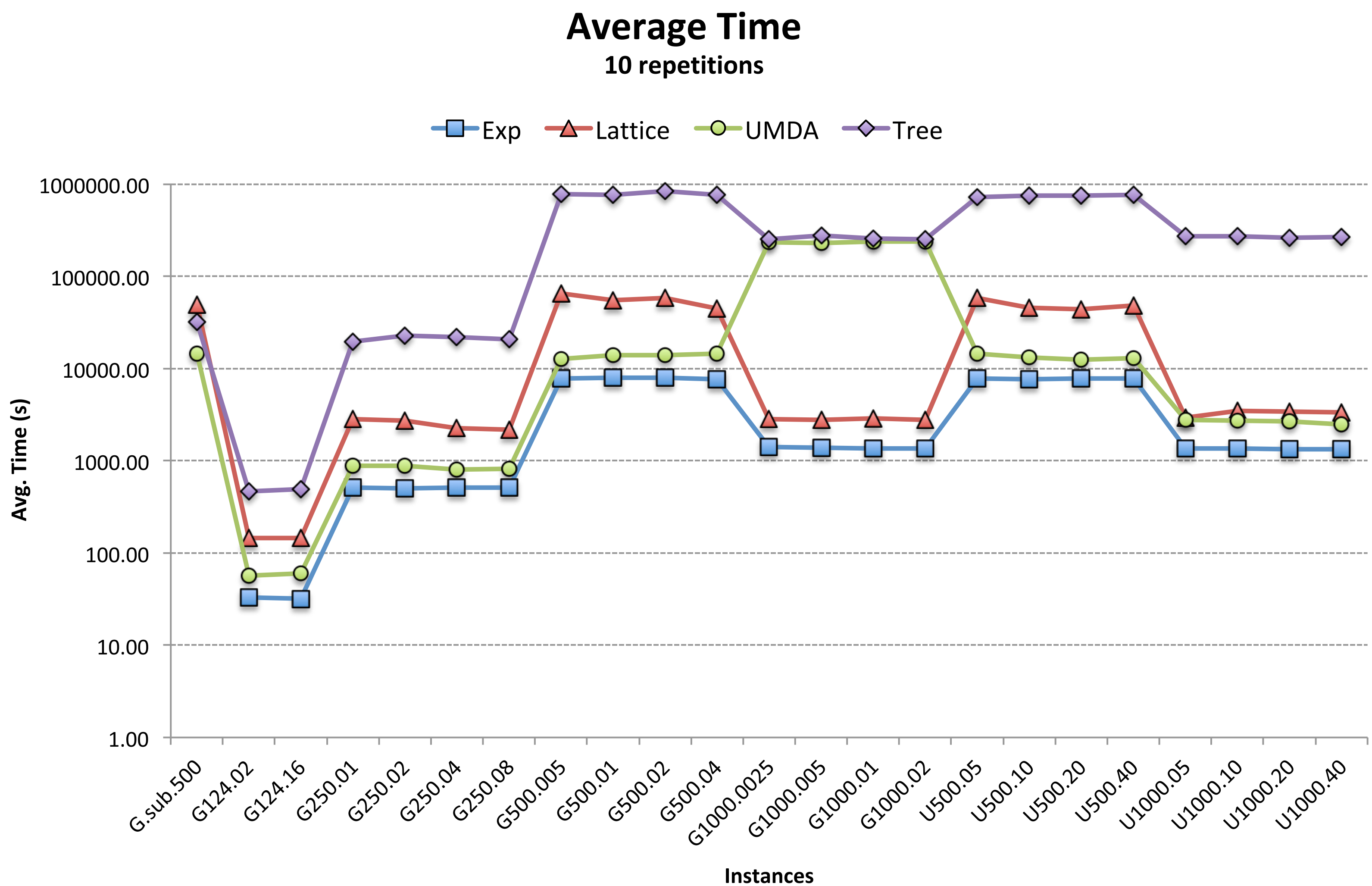
Experimental Study

Results - Performance

Instance	Best Fitness	ARPD			
		Exp	Lattice	UMDA	Tree
G124.02	13	0,32	0,32	0,61	**0,19**
G124.16	449	0,04	0,02	0,05	**0,01**
G250.01	31	0,40	0,33	0,49	**0,20**
G250.02	118	0,10	0,07	0,14	**0,06**
G250.04	360	0,04	0,04	0,10	**0,03**
G250.08	830	0,05	**0,01**	0,05	**0,01**
G500.005	61	0,38	0,30	0,40	**0,08**
G500.01	234	0,14	0,09	0,21	**0,07**
G500.02	642	0,06	**0,03**	0,11	**0,03**
G500.04	1754	0,05	**0,02**	0,06	**0,02**
G1000.0025	131	1,16	2,96	3,20	**0,74**
G1000.005	496	**0,52**	1,22	1,28	0,88
G1000.01	1420	**0,28**	0,56	0,66	0,62
G1000.02	3450	**0,18**	0,35	0,40	0,39
U500.05	23	1,34	1,17	1,89	**0,57**
U500.10	61	1,50	1,05	1,12	**0,57**
U500.20	185	0,89	0,56	0,87	**0,44**
U500.40	412	0,81	0,41	0,38	**0,28**
U1000.05	77	4,82	**1,62**	12,83	2,39
U1000.10	170	5,09	**1,67**	11,67	3,73
U1000.20	352	5,23	**1,67**	10,58	4,94
U1000.40	862	4,09	**1,53**	3,24	2,29

Experimental Study

Results - Performance



Conclusions

The experiments support the validity of our research line:

Designing probability models exclusively on the set of feasible solutions

However, many issues still pending…

Future Work

Many aspects to be faced!

Consider other methods to estimate $\bar{x}$

Analyze regularity of the solutions regarding the distance

Propose other distances

The probability model

Larger benchmarks

Understand the dynamics of the EDA for different problem sizes

Experimentation

New probability models

EDA

Distance-based exponential probability models for constrained combinatorial optimization problems.

Thank you for your attention!!!

Intelligent Systems Group

eman ta zabal zazu

Universidad del País Vasco Euskal Herriko Unibertsitatea

XVIII Conference of the Spanish Association for Artificial Intelligence
23-26 October 2018, Granada, Spain