ProDiGen: a genetic algorithm for process discovery guided by completeness, precision and simplicity

Borja Vázquez Barreiros

Supervisors **: Manuel Mucientes, Manuel Lama**

Centro Singular de Investigación en Tecnoloxías da Información

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Process

WHAT IS A PROCESS?

“[…] a collection of tasks ---or activities--- with coordination requirements among them.”
- Wil van der Aalst

EXAMPLE

1. A teacher uploads an exercise
2. A learner makes the exercise
3. The teacher grades it

CiTIUS

Workflow

And what is a workflow?

"The automation of a business **process**, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules."

-Workflow Management Coalition

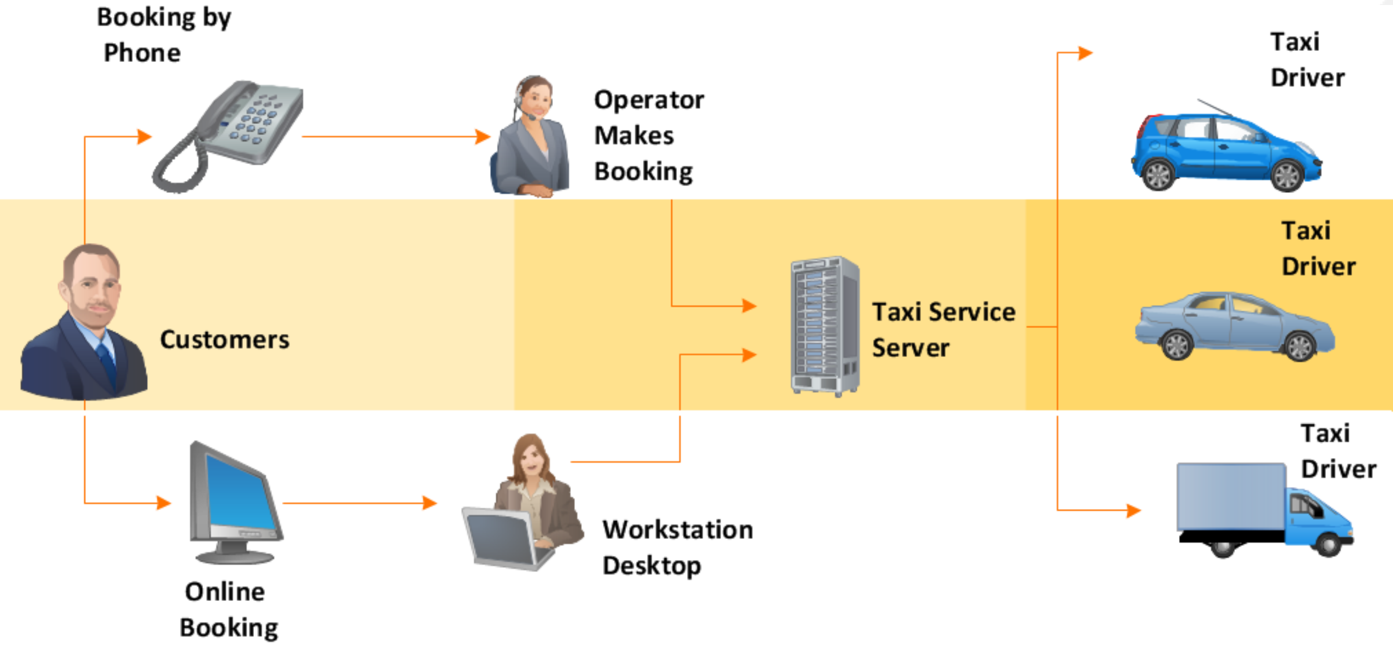
Example

1. A teacher uploads an exercise
2. A learner makes the exercise
3. The teacher grades it

Workflow

Provides useful information:

- **What** is happening on a process.
- **How** is that process being executed.
- **Who** is involved in the execution of the process.



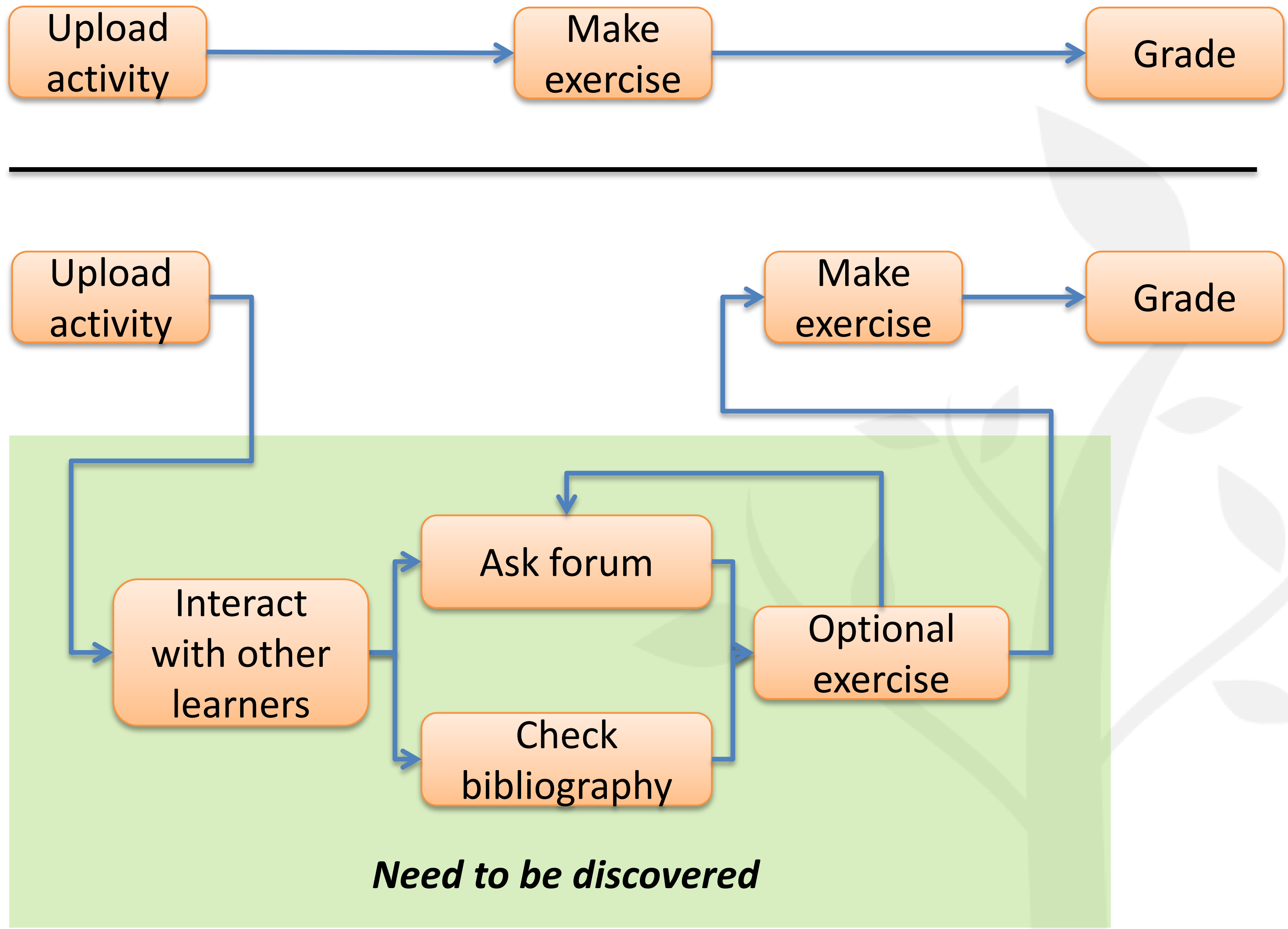
CiTIUS

Workflow

Issues

What we think is going on

What is actually happening



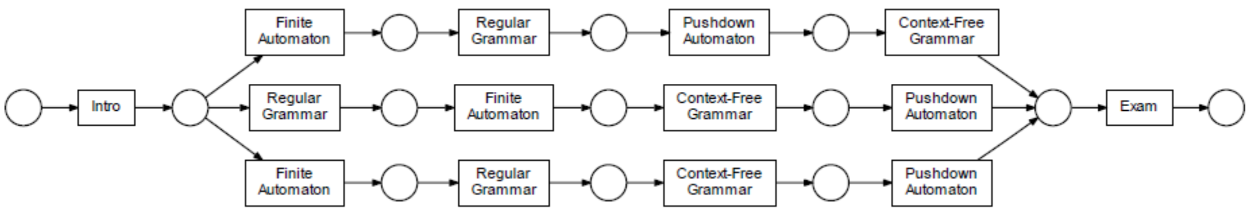
CiTIUS

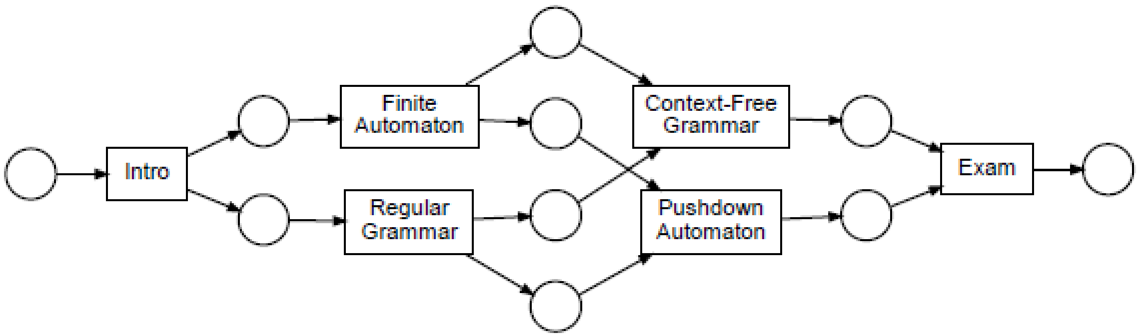
Process mining

Definition

The goal of **process mining** is to automatically discover the models that better fit the process, *taking as a starting point the logs*.

case ID	task
Saul	Intro
Walter	Intro
Saul	Finite Automaton
Saul	Regular Grammar
Walter	Finite Automaton
Walter	Regular Grammar
Saul	Context-Free Grammar
Walter	Pushdown Automaton
Saul	Pushdown Automaton
Walter	Context-Free Grammar
Saul	Exam
Walter	Exam





What is actually happening, and not what the people think it is

CiTIUS

Process mining

We want models with some characteristics…

- **Completeness:** model all the behavior shown in the log**.**
- **Precision:** avoid overly general models**.**
- **Generalization:** avoid overly precise models**.**
- **Simplicity:** Occam's razor.

CiTIUS

Process mining

We want models with some characteristics…

- **Completeness:** model all the behavior shown in the log**.**
- **Precision:** avoid overly general models**.**
- **Generalization:** avoid overly precise models**.**

Bias-variance tradeoff

- **Simplicity:** Occam's razor.

CiTIUS

WE WANT MODELS WITH SOME CHARACTERISTICS…

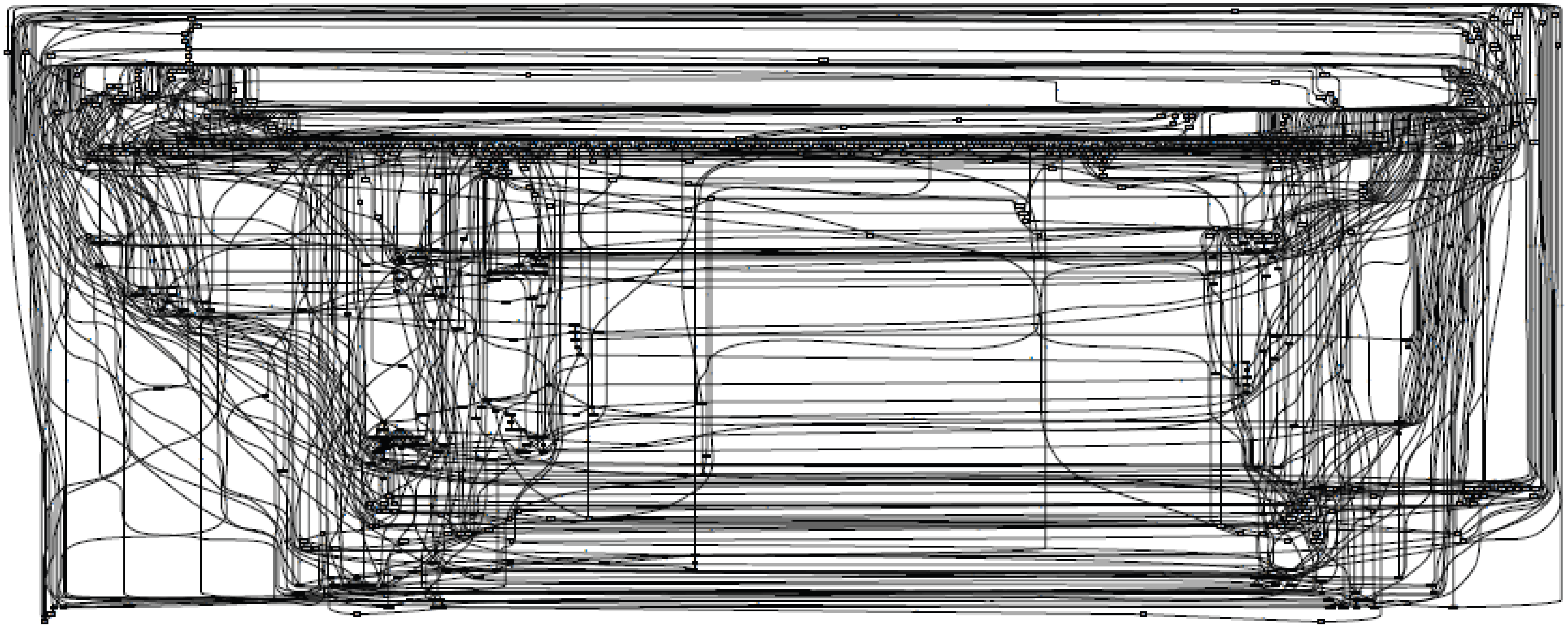
- **Completeness:** model all the behavior shown in the log**.**
- **Precision:** avoid overly general models**.**
- **Generalization:** avoid overly precise models**.**
- **Simplicity:** Occam's razor.

Bias-variance tradeoff (Precision, Generalization)

CiTIUS

Process mining

Spaghetti models



We want models with some characteristics…

- **Completeness:** model all the behavior shown in the log.
- **Precision:** avoid overly general models.
- **Generalization:** avoid overly precise models.
- **Simplicity:** Occam's razor.
- **Robust to noise**

Bias-variance tradeoff (Precision, Generalization)

CiTIUS

State of the art

- ***Abstraction based*** : Poor completeness
  - *α-algorithm (and extensions)*
- ***Heuristics based***: Cannot handle all the constructs at once
  - *Heuristics Miner*
- ***Search based***: Do not consider simplicity
  - *Genetic Miner*
- **Based theory of regions**: Cannot handle noise and infrequent behavior
  - *ILP*

CiTIUS

Objectives

1. An algorithm that retrieves ***complete, precise*** and ***simple*** models.
2. Robust to noise
3. Application in different domains

CiTIUS

Objectives

1. An algorithm that retrieves ***complete, precise*** and ***simple*** models***.***
2. Robust to noise
3. Application in different domains

Objectives

1. An algorithm that retrieves ***complete, precise*** and ***simple*** models***.***
2. Robust to noise
3. Applicable in different domains

CiTIUS

Index

1. ProDiGen
2. SoftLearn
3. Future work
4. Publications

CiTIUS

ProDiGen

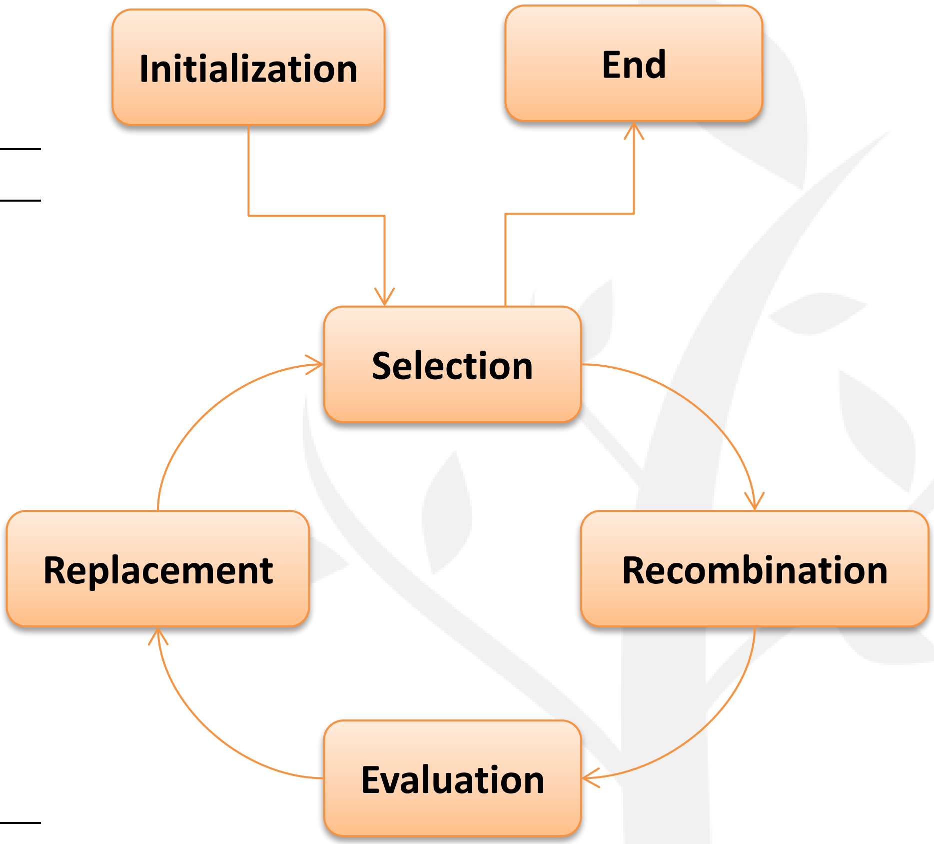
Genetic algorithm

- Components:
  - **Individuals**: solutions.
  - **Population**: set of individuals.

Algorithm 1: Genetic algorithm for process discovery.

```
Initialize population;
Evaluate population;
t = 1, timesRun = initialTimesRun, restarts = 0;
while t ≤ maxGenerations && restarts < maxRestarts do
    Selection;
    Crossover;
    Mutation;
    Evaluate new individuals;
    Replace population;
    t = t + 1;
    if bestInd(t) == bestInd(t − 1) then
        timesRun = timesRun − 1;
    if none of the individuals of the population have been replaced then
        timesRun = timesRun − 1;
    if timesRun < 0 then
        Reinitialize population;
        Evaluate population;
        timesRun = initialTimesRun, restarts = restarts + 1;
```

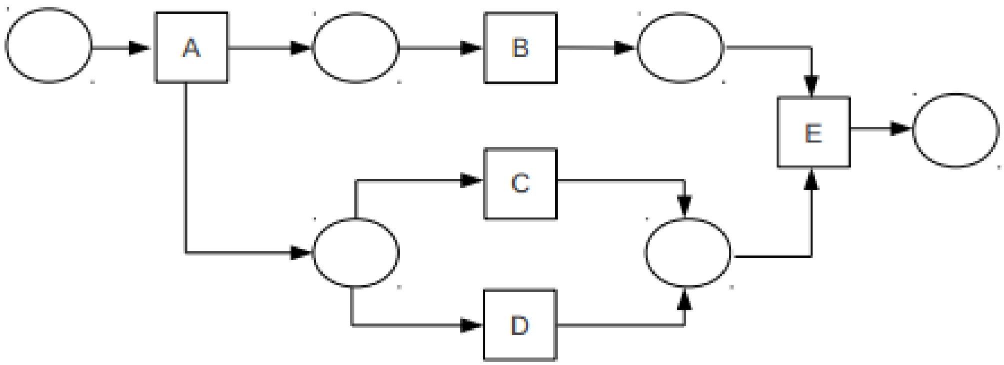
Evolutionary cycle



ProDiGen

Internal representation

- **Causal matrix:** maps the input and output dependencies of each task.



Tarea	I(Tarea)	O(Tarea)
A	{}	{{B},{C,D}}
B	{A}	{E}
C	{A}	{E}
D	{A}	{E}
E	{{B},{C,D}}	{}

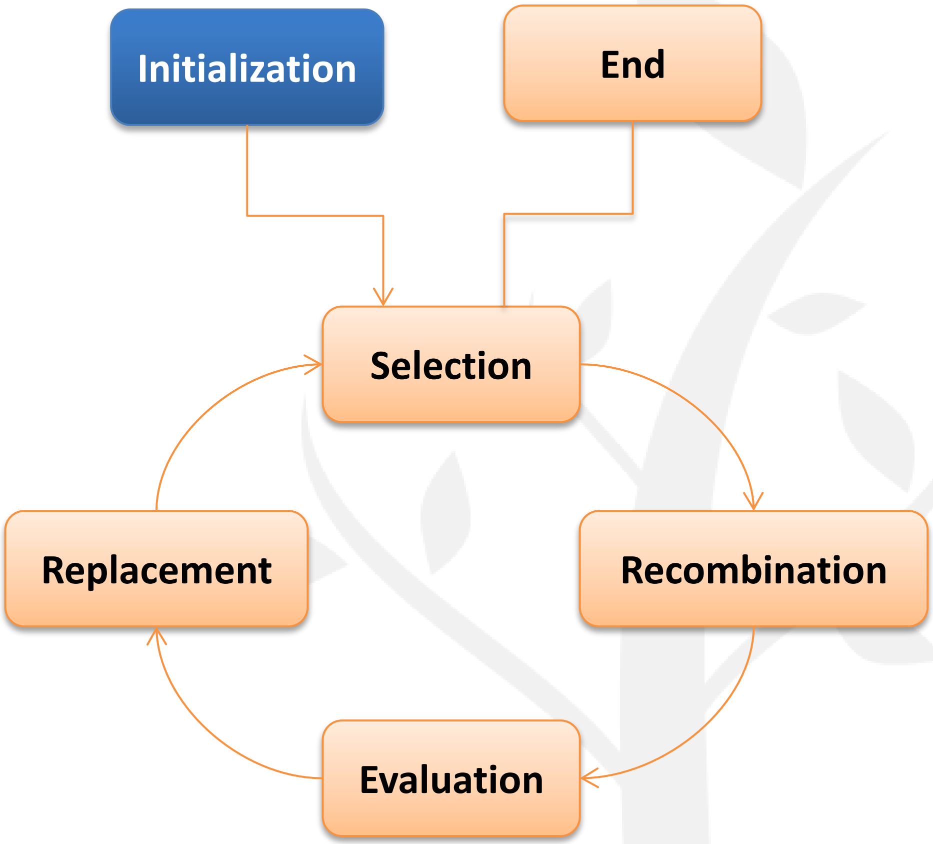
CiTIUS

ProDiGen

Genetic algorithm

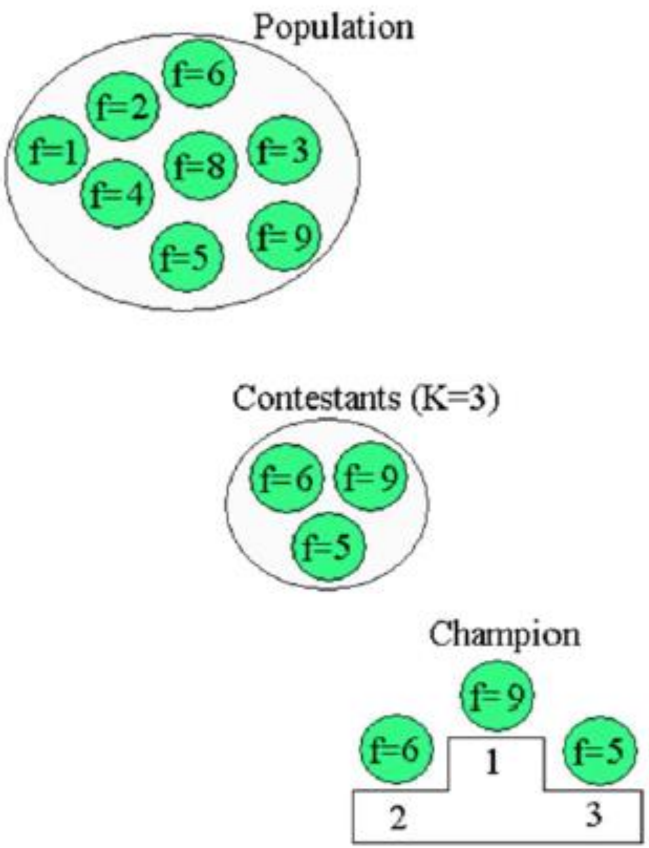
- Initialization of each individual of the population
- Based on a heuristics approach
  - Using the local information of the log.
  - Dependencies between tasks.

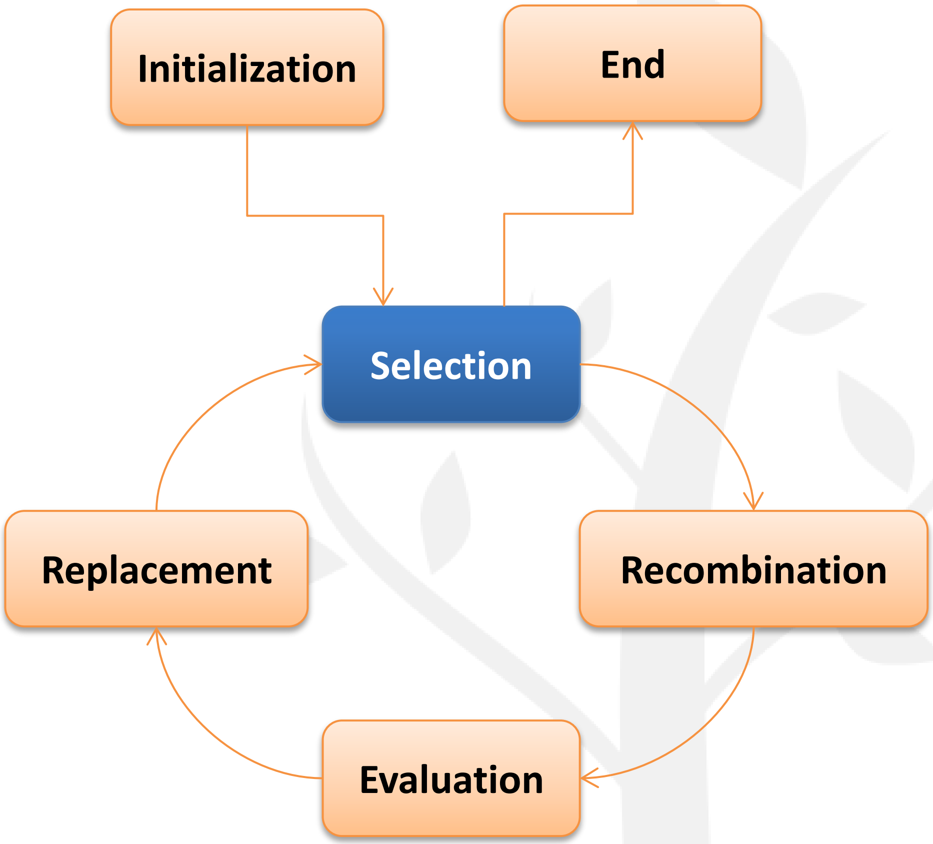
$$\begin{cases} \frac{\#aba+\#bab}{\#aba+\#bab+1} & \text{if } a \neq b \text{ and } \#aba > 0; \\ \frac{\#ab-\#ba}{\#ab+\#ba+1} & \text{if } a \neq b \text{ and } \#aba = 0; \\ \frac{\#ab}{\#ab+1} & \text{if } a = b. \end{cases}$$



CiTIUS

- Binary tournament



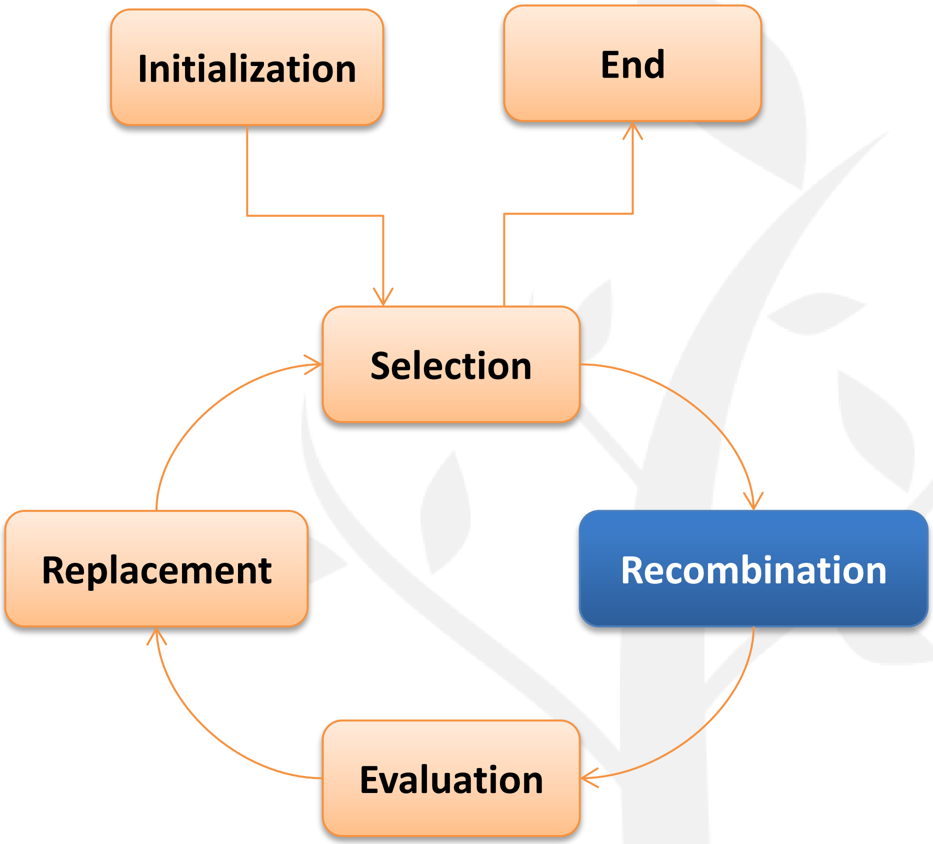


ProDiGen

Genetic algorithm

Generate new individuals.

- **Crossover**:
  - Combine the characteristics of two parents into two offspring
- **Mutation**:
  - Add or remove material form an individual.



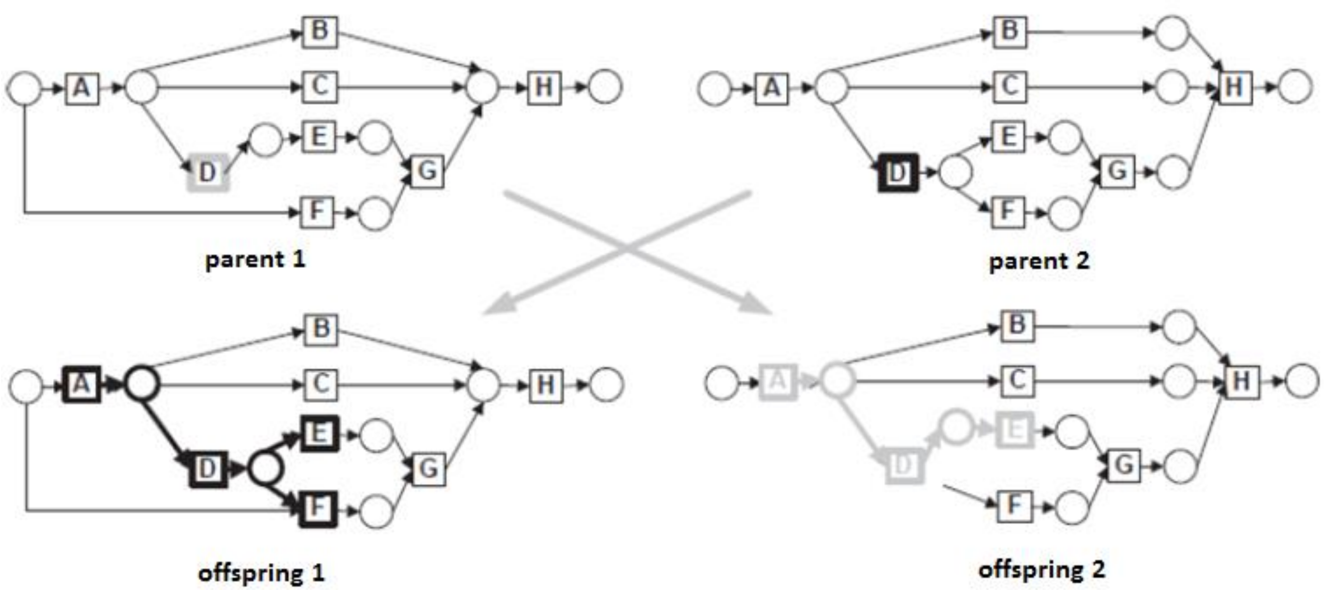
CiTIUS

Crossover

- Guided by a *Probability Density Function* generated from the errors.

Algorithm 2: Pseudo-code for the crossover operator.

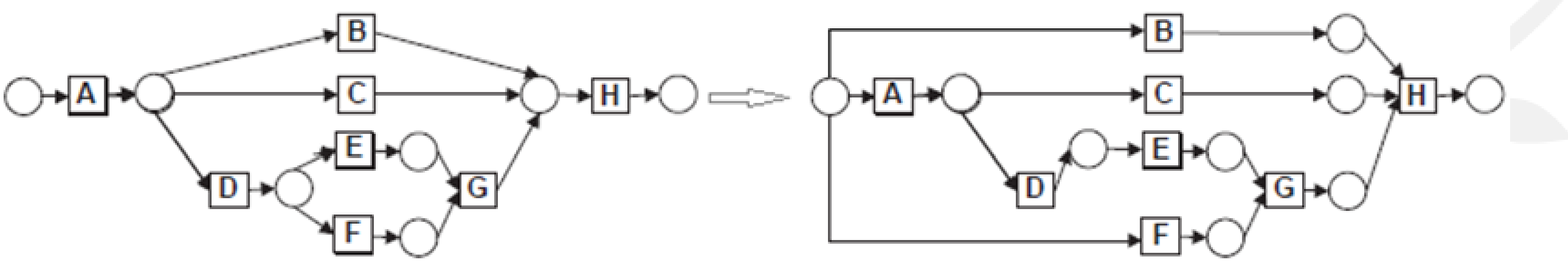
```
r ← getRandomNumber() // returns a random number between [0,1);
if r < crossoverRate then
    incorrectlyFiredActivities ← ∅;
    if fitness(parent_1) >= fitness(parent_2) then
        incorrectlyFiredActivities ← set of incorrectly fired activities of
        parent_1;
    else
        incorrectlyFiredActivities ← set of incorrectly fired activities of
        parent_2;
    if incorrectlyFiredActivities ≠ ∅ then
        crossoverPoint ← randomly select an activity t from
        incorrectlyFiredActivities;
    else
        crossoverPoint ← randomly select an activity t from the bag of all
        possible tasks in the log;
    offspring_1, offspring_2 ← doCrossover(parent_1, parent_2, crossoverPoint);
    Repair offspring_1 and offspring_2;
```



- Guided by the causal dependencies of the log.

Algorithm 3: Pseudo-code for the mutation operator.

```
1  while the individual does not change do
2      Randomly choose one task t in the individual;
3      mutationType ← getRandomNumber() // returns a random number
       between [0,1) ;
4      if mutationType < 1/3 then
5          Randomly add a new task t' to I(t), being t' a task from
           inputDependencies(t);
6          if getRandomNumber() < 1/2 then
7              Randomly choose one subset X ∈ I(t) and add the task t' to X;
8          else
9              Create a new subset X, add the task t' to X, and add X to I(t);
10     else if mutationType < 2/3 then
11         Randomly choose one subset X ∈ I(t) and remove a task t' from X,
           where t' ∈ X. If X is empty after this operation, exclude X from I(t);
12     else
13         Randomly redistribute the elements from I(t);
14     Repeat from line 3, but using O(t) instead of I(t) and outputDependencies(t)
       instead of inputDependencies(t);
15     Repair the individual;
```



Individual

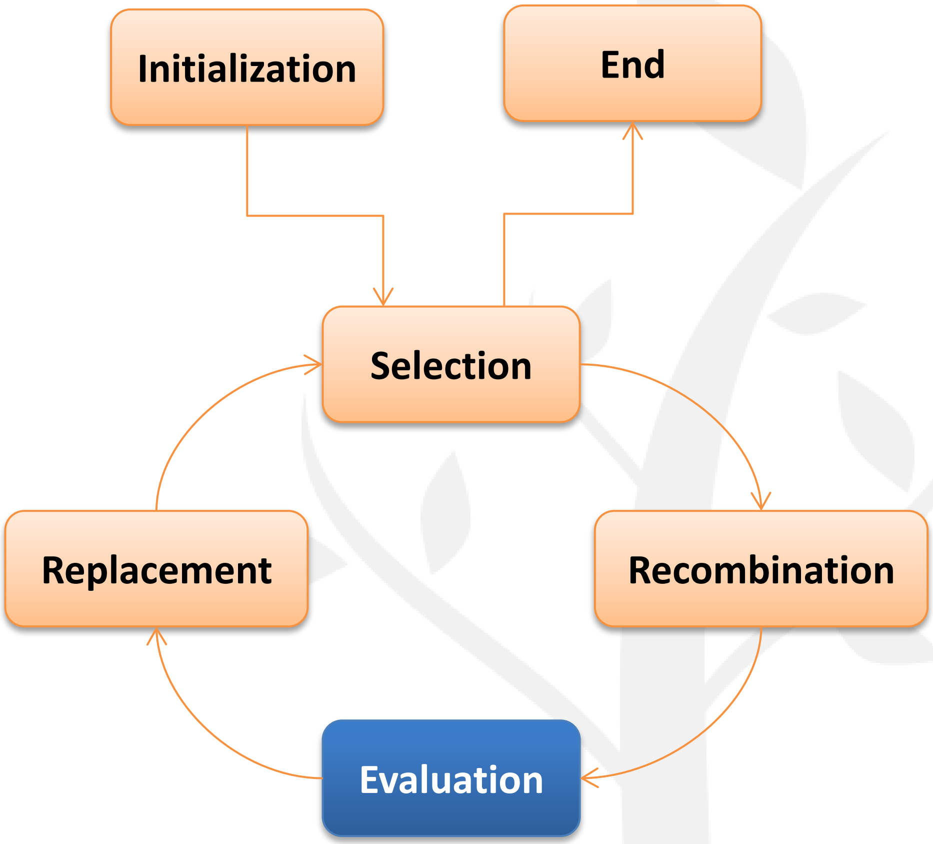
Mutated individual

- Each individual is evaluated with **three** objectives:

Completeness

Precision

Simplicity



CiTIUS

- **Completeness**

$$C_f\,(L,\ CM) = \frac{allParsedActivities(L,CM) - punishment}{numActivitiesLog(L)}$$

where

$$punishment = \frac{allMissingTokens(L,CM)}{numTracesLog(L) - numTracesMissingTokens(L,CM) + 1} + \frac{allExtraTokensLeftBehind(L,CM)}{numTracesLog(L) - numTracesExtraTokensLeftBehind(L,CM) + 1}$$

CiTIUS

- **Precision**:

$$P_f\left(L,\ CM\right) = \frac{1}{allEnabledActivities\left(L,\ CM\right)}$$

- **Simplicity**:

$$S_f\left(CM\right) = \frac{1}{\sum_{t \in CM}\left(\sum_{\Phi \in I(t)} |\Phi| + \sum_{\Psi \in O(t)} |\Psi|\right)}$$

Hierarchical fitness function

$$F(a) > F(b) \iff \{C_f(a) > C_f(b)\} \vee \{C_f(a) = C_f(b) \wedge P_f(a) > P_f(b)\} \vee \{C_f(a) = C_f(b) \wedge P_f(a) = P_f(b) \wedge S_f(a) > S_f(b)\}$$

Completeness → ***Precision*** → ***Simplicity***

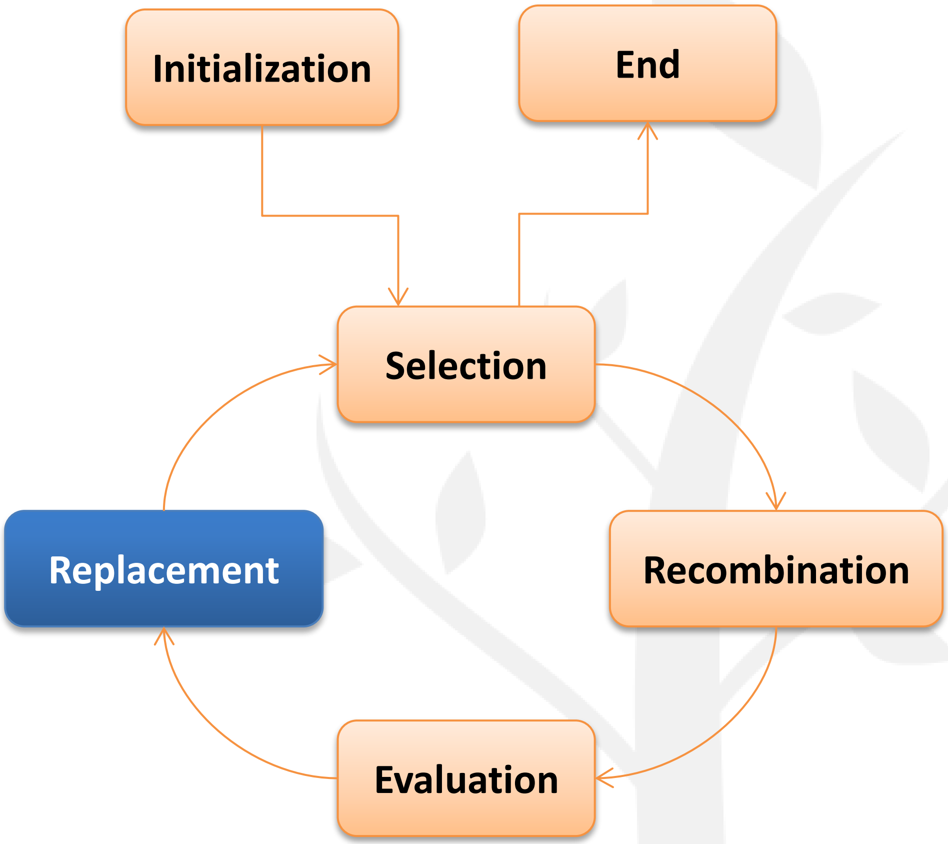
CiTIUS

ProDiGen

Genetic algorithm

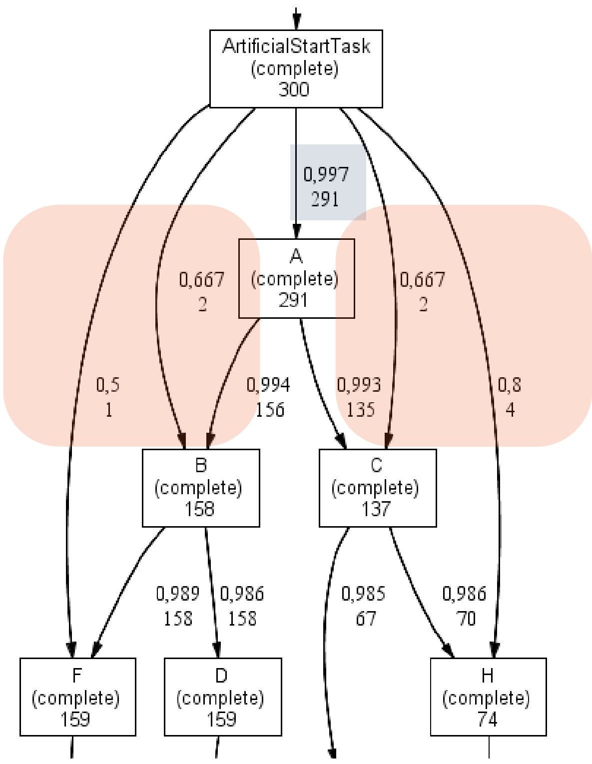
Update of the population

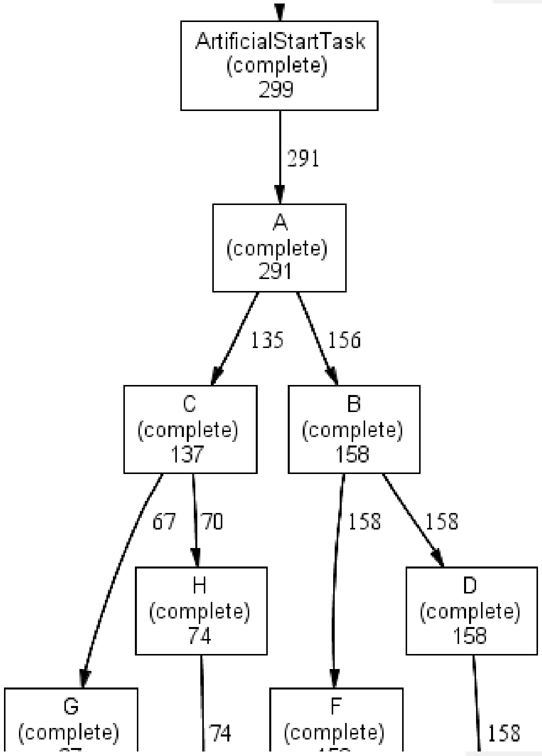
- **Steady-state** process…
  - Combine and sort parents and offspring (2N)
  - The N best survive to the next cycle.
- …with **reinitialization**:
  - If the best solution does not change
  - If there are not new individuals in the population



WAIT…, AND WHAT ABOUT NOISE?

- Post-processing of the best individual: **arc-pruning**
  - Remove those arcs used fewer than a certain threshold





ProDiGen

Experimentation - Balanced logs

- 18 different models
  - Different degrees of complexity
  - Logs with several levels of noise:
    - 0% of noise
    - 1% of noise
    - 5% of noise
    - 10% of noise.
    - 20% of noise.
  - In total: 90 different logs

		Activity structures										Log content	
Model	#Tasks	Sequence	Choice	Parallelism	Length-One Loop	Length-Two Loop	Arbitrary Loop	Structured Loop	Non-local NFC	Local NFC	Invisible tasks	#traces	#events
Caminatas	12	✓	✓	✓								700	4,200
A8	7	✓	✓	✓								300	1,200
D2	6	✓		✓								300	1,200
M11Skip [11]	6	✓	✓	✓	✓							500	4,757
Ma5 [11]	7	✓	✓	✓	✓							300	2,178
Ml2l [11]	6	✓	✓			✓						300	4,668
MDriverLL [11]	11	✓	✓	✓				✓	✓		✓	700	13,303
allLoops	5	✓	✓		✓	✓		✓			✓	300	1,035
l2la	6	✓	✓	✓			✓					300	2,264
Ma7 [11]	9	✓	✓	✓								500	2,427
Herbst6p37 [11]	16	✓		✓								700	12,600
MexampleL [11]	8	✓	✓	✓								300	1,645
Ma6nfc [11]	8	✓	✓							✓		300	2,006
MParallel5 [11]	10	✓		✓								700	12,600
NC	7	✓	✓	✓						✓		300	1,704
L2LP	7	✓	✓	✓	✓	✓		✓			✓	300	5,476
NCB	7	✓	✓	✓			✓			✓		300	2,950
DWS [23]	12	✓	✓	✓							✓	500	4,033

(a) *Balanced logs*.

CiTIUS

ProDiGen

Experimentation - Unbalanced logs

- 21 different models
  - Models with many interleaving situations.
  - Models with many different traces and frequencies
  - In total: 21 different logs

	Activity structures										Log content	
Model	#Tasks	Sequence	Choice	Parallelism	Length-One Loop	Length-Two Loop	Arbitrary Loop	Structured Loop	Invisible tasks	Unbalance	#traces	#events
g2 [11]	22	✓	✓	✓		✓	✓		✓		300	4,501
g3 [11]	29	✓	✓	✓			✓	✓	✓		300	14,599
g4 [11]	29	✓	✓	✓		✓				✓	300	5,975
g5 [11]	20	✓	✓	✓				✓	✓		300	6,172
g6 [11]	23	✓	✓	✓		✓			✓		300	5,419
g7 [11]	29	✓	✓	✓			✓		✓		300	14,451
g8 [11]	30	✓	✓	✓		✓	✓		✓	✓	300	5,133
g9 [11]	26	✓	✓	✓		✓	✓		✓		300	5,679
g10 [11]	23	✓	✓	✓				✓	✓		300	4,117
g12 [11]	26	✓	✓	✓		✓		✓	✓		300	4,841
g13 [11]	22	✓	✓	✓	✓	✓			✓	✓	300	5,007
g14 [11]	24	✓	✓	✓			✓		✓	✓	300	11,340
g15 [11]	25	✓	✓		✓	✓			✓		300	3,978
g19 [11]	23	✓	✓	✓		✓			✓	✓	300	4,107
g20 [11]	21	✓	✓		✓	✓		✓	✓		300	6,193
g21 [11]	22	✓	✓					✓	✓		300	3,882
g22 [11]	24	✓	✓	✓			✓		✓	✓	300	3,095
g23 [11]	25	✓	✓	✓		✓				✓	300	9,654
g24 [11]	21	✓	✓	✓				✓	✓	✓	300	4,130
g25 [11]	20	✓	✓	✓	✓				✓		300	6,312
EMT [9]	7	✓	✓	✓					✓	✓	100	790

CiTIUS

ProDiGen

Experimentation - Unbalanced logs

- 21 different models
  - ▷ Models with many interleaving situations.
  - ▷ Models with many different traces and frequencies
  - ▷ In total: 21 different logs

		Activity structures									Log content	
Model	#Tasks	Sequence	Choice	Parallelism	Length-One Loop	Length-Two Loop	Arbitrary Loop	Structured Loop	Invisible tasks	Unbalance	#traces	#events
g2 [11]	22	✓	✓	✓		✓	✓		✓		300	4,501
g3 [11]	29	✓	✓	✓			✓	✓	✓		300	14,599
g4 [11]	29	✓	✓	✓		✓				✓	300	5,975
g5 [11]	20	✓	✓	✓				✓	✓		300	6,172
g6 [11]	23	✓	✓	✓		✓			✓		300	5,419
g7 [11]	29	✓	✓	✓			✓		✓		300	14,451
g8 [11]	30	✓	✓	✓		✓	✓		✓	✓	300	5,133
g9 [11]	26	✓	✓	✓		✓	✓		✓		300	5,679
g10 [11]	23	✓	✓	✓				✓	✓		300	4,117
g12 [11]	26	✓	✓	✓		✓		✓	✓		300	4,841
g13 [11]	22	✓	✓	✓	✓	✓			✓	✓	300	5,007
g14 [11]	24	✓	✓	✓			✓		✓	✓	300	11,340
g15 [11]	25	✓	✓		✓	✓			✓		300	3,978
g19 [11]	23	✓	✓	✓		✓			✓	✓	300	4,107
g20 [11]	21	✓	✓		✓	✓		✓	✓		300	6,193
g21 [11]	22	✓	✓					✓	✓		300	3,882
g22 [11]	24	✓	✓	✓			✓		✓	✓	300	3,095
g23 [11]	25	✓	✓	✓		✓				✓	300	9,654
g24 [11]	21	✓	✓	✓				✓	✓	✓	300	4,130
g25 [11]	20	✓	✓	✓	✓				✓		300	6,312
EMT [9]	7	✓	✓	✓					✓	✓	100	790

PRODIGEN HAS BEEN TESTED WITH 111 DIFFERENT LOGS

CiTIUS

- Based on the original model:
  - Behavior similarity:

$$B_p(L, CM_o, CM_m) = \frac{\sum_{\sigma \in L} \left(\frac{L(\sigma)}{|\sigma|} \times \sum_{i=1}^{|\sigma|} \frac{|Enabled(CM_o, \sigma, i) \cap Enabled(CM_m, \sigma, i)|}{|Enabled(CM_m, \sigma, i)|} \right)}{\sum_{\sigma \in L} L(\sigma)}$$

$$B_r(L, CM_o, CM_m) = \frac{\sum_{\sigma \in L} \left(\frac{L(\sigma)}{|\sigma|} \times \sum_{i=1}^{|\sigma|} \frac{|Enabled(CM_o, \sigma, i) \cap Enabled(CM_m, \sigma, i)|}{|Enabled(CM_o, \sigma, i)|} \right)}{\sum_{\sigma \in L} L(\sigma)}$$

  - Similarity from the structural point of view:

$$S_p(CM_o, CM_m) = \frac{|C_o \cap C_m|}{|C_m|}$$

$$S_r(CM_o, CM_m) = \frac{|C_o \cap C_m|}{|C_o|}$$

CiTIUS

- Based on the log:
  - **Proper completion**: percentage of correctly parsed traces.
  - **Precision**: how much behavior of the log is allowed by the model

$$precision(L, M) = \frac{1}{|\mathcal{E}|} \sum_{e \in \mathcal{E}} \frac{|en_L(e)|}{|en_M(e)|}$$

  - **Simplicity:** Weighted place/transition node arc degree

CiTIUS

ProDiGen

Results on balanced logs (i)

Logs with 0% of noise

Method	Metrics		Caminatas	A8	D2	M11Skip	MaS	M21	MDriveLL	allLoops	Ma7	l2la	MexampleL	Herbst6p37	Ma6nfc	MParallel5	NC	L2LP	NCB	DWS
ProDiGen	Model metrics	B_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		B_r	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S_r	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Log metrics	P	1.0	1.0	1.0	0.78	0.89	0.84	0.89	0.77	1.0	0.94	0.94	0.95	1.0	1.0	1.0	0.56	0.87	0.78
		C	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S	0.25	0.32	0.3	0.31	0.3	0.34	0.28	0.29	0.28	0.29	0.29	0.34	0.31	0.31	0.29	0.31	0.31	0.3
GM	Model metrics	B_p	0.84	1.0	1.0	1.0	0.96	1.0	0.9	0.99	0.95	1.0	0.94	1.0	0.98	0.74	1.0	1.0	1.0	1.0
		B_r	0.98	1.0	1.0	1.0	1.0	1.0	1.0	0.99	0.99	1.0	0.97	1.0	1.0	0.97	1.0	1.0	1.0	1.0
		S_p	0.85	1.0	1.0	1.0	1.0	1.0	0.93	0.98	0.94	1.0	0.86	1.0	0.96	0.08	1.0	1.0	1.0	1.0
		S_r	0.73	1.0	1.0	1.0	0.91	1.0	0.84	0.96	0.97	1.0	0.88	1.0	0.99	0.98	1.0	1.0	1.0	1.0
	Log metrics	P	0.81	1.0	1.0	0.78	0.86	0.84	0.81	0.68	0.68	0.94	0.91	0.95	0.81	0.89	1.0	0.56	0.87	0.78
		C	0.36	1.0	1.0	1.0	1.0	1.0	1.0	0.48	0.15	1.0	0.52	1.0	1.0	0.81	1.0	1.0	1.0	1.0
		S	0.29	0.32	0.3	0.31	0.32	0.34	0.3	0.28	0.28	0.29	0.31	0.34	0.31	0.29	0.29	0.31	0.31	0.3
HM	Model metrics	B_p	1.0	1.0	1.0	0.97	1.0	1.0	1.0	0.96	1.0	1.0	1.0	1.0	0.91	1.0	0.44	0.82	0.88	0.89
		B_r	1.0	1.0	1.0	0.95	1.0	1.0	1.0	0.92	1.0	1.0	1.0	1.0	1.0	1.0	0.82	0.93	1.0	0.94
		S_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	0.83
		S_r	1.0	1.0	1.0	0.82	1.0	1.0	1.0	0.83	1.0	1.0	1.0	1.0	0.9	1.0	0.91	0.9	0.88	1.0
	Log metrics	P	1.0	1.0	1.0	0.71	0.89	0.84	0.89	0.8	1.0	0.94	0.94	0.95	0.89	1.0	0.84	0.65	0.0	0.82
		C	1.0	1.0	1.0	0.84	1.0	1.0	1.0	0.73	1.0	1.0	1.0	1.0	0.69	1.0	0.84	0.24	0.0	0.51
		S	0.25	0.32	0.3	0.32	0.3	0.34	0.28	0.27	0.28	0.29	0.29	0.34	0.33	0.31	0.31	0.3	0.32	0.3
α^{++}	Model metrics	B_p	1.0	1.0	1.0	1.0	1.0	1.0	0.87	0.77	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.59	0.88	0.75
		B_r	1.0	1.0	1.0	1.0	1.0	1.0	0.89	0.83	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.72	1.0	0.63
		S_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.72	1.0	0.81
		S_r	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.83	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.88	1.0
	Log metrics	P	1.0	1.0	1.0	0.78	0.89	0.84	0.95	0.74	1.0	0.94	0.94	0.95	1.0	1.0	1.0	0.0	0.0	0.0
		C	1.0	1.0	1.0	1.0	1.0	1.0	0.63	0.25	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.0
		S	0.25	0.32	0.3	0.31	0.3	0.34	0.26	0.3	0.28	0.29	0.29	0.34	0.31	0.31	0.29	0.24	0.32	0.27
ILP	Model metrics	B_p	1.0	1.0	1.0	1.0	1.0	1.0	0.86	0.81	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.63	1.0	0.82
		B_r	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.92	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	0.97
		S_p	1.0	1.0	1.0	1.0	1.0	1.0	0.86	0.54	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.66	1.0	0.66
		S_r	1.0	1.0	1.0	1.0	1.0	1.0	0.93	0.71	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Log metrics	P	1.0	1.0	1.0	0.78	0.89	0.84	0.81	0.77	1.0	0.94	0.94	0.95	1.0	1.0	1.0	0.48	0.87	0.72
		C	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S	0.25	0.32	0.3	0.31	0.3	0.34	0.28	0.25	0.28	0.29	0.29	0.34	0.31	0.31	0.29	0.3	0.31	0.25

Logs with 1% of noise

Method	Metrics		Caminatas	A8	D2	M11Skip	MaS	M21	MDriveLL	allLoops	Ma7	l2la	MexampleL	Herbst6p37	Ma6nfc	MParallel5	NC	L2LP	NCB	DWS
ProDiGen	Model metrics	B_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		B_r	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S_r	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Log metrics	P	1.0	1.0	1.0	0.82	0.91	0.85	0.9	0.86	1.0	0.96	0.96	0.96	1.0	1.0	1.0	0.56	0.87	0.78
		C	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S	0.25	0.32	0.3	0.32	0.31	0.34	0.29	0.3	0.29	0.3	0.3	0.34	0.32	0.31	0.3	0.31	0.31	0.3
GM	Model metrics	B_p	0.68	1.0	0.95	0.99	0.97	1.0	0.92	0.99	0.82	1.0	1.0	1.0	0.98	0.8	1.0	1.0	1.0	0.86
		B_r	0.99	0.99	1.0	0.96	0.99	0.99	1.0	1.0	0.99	1.0	1.0	0.99	1.0	0.98	1.0	1.0	1.0	0.81
		S_p	0.79	0.8	1.0	1.0	0.91	0.88	0.94	1.0	0.75	1.0	1.0	0.95	0.96	0.85	1.0	1.0	1.0	0.83
		S_r	0.67	1.0	0.87	0.92	0.91	1.0	0.85	0.96	1.0	1.0	1.0	1.0	0.99	0.93	1.0	1.0	1.0	0.95
	Log metrics	P	0.0	1.0	0.64	0.78	0.89	0.73	0.78	0.68	0.77	0.96	0.96	0.9	0.93	0.75	1.0	0.56	0.87	0.71
		C	0.23	1.0	1.0	1.0	1.0	1.0	1.0	0.49	1.0	1.0	1.0	1.0	1.0	0.14	1.0	1.0	1.0	0.82
		S	0.29	0.29	0.29	0.29	0.32	0.32	0.3	0.29	0.29	0.3	0.3	0.33	0.31	0.27	0.3	0.31	0.31	0.29
HM	Model metrics	B_p	1.0	1.0	1.0	0.97	1.0	1.0	0.93	0.96	1.0	1.0	1.0	1.0	0.9	1.0	0.87	0.82	0.88	0.89
		B_r	1.0	1.0	1.0	0.92	1.0	1.0	1.0	0.92	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.93	1.0	0.94
		S_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	0.81
		S_r	1.0	1.0	1.0	0.76	1.0	1.0	0.9	0.83	1.0	1.0	1.0	1.0	0.91	1.0	0.91	0.9	0.88	1.0
	Log metrics	P	1.0	1.0	1.0	0.71	0.91	0.85	0.81	0.87	1.0	0.96	0.96	0.96	0.91	1.0	0.88	0.65	0.0	0.82
		C	1.0	1.0	1.0	0.84	1.0	1.0	1.0	0.73	1.0	1.0	1.0	1.0	0.69	1.0	0.84	0.24	0.0	0.51
		S	0.25	0.32	0.3	0.32	0.31	0.34	0.3	0.28	0.29	0.3	0.3	0.34	0.33	0.31	0.32	0.3	0.32	0.3
α^{++}	Model metrics	B_p	0.76	0.82	1.0	0.73	0.84	0.76	0.77	0.94	0.83	0.73	0.4	0.41	0.6	0.87	0.76	0.77	0.88	0.46
		B_r	0.87	0.87	1.0	0.63	0.82	0.69	0.75	0.83	0.85	0.6	0.7	0.52	0.78	0.99	0.86	0.56	1.0	0.45
		S_p	0.93	0.77	1.0	0.81	0.83	1.0	0.79	0.81	0.83	0.69	0.69	0.46	0.68	0.68	0.84	1.0	1.0	0.74
		S_r	1.0	0.87	1.0	0.69	0.83	0.75	0.9	0.75	0.83	0.81	0.75	0.61	0.91	0.86	0.91	0.6	0.88	0.9
	Log metrics	P	0.0	1.0	1.0	0.0	0.87	0.68	0.0	0.82	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C	0.0	0.5	1.0	1.0	1.0	1.0	0.0	0.25	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
		S	0.24	0.31	0.3	0.29	0.26	0.35	0.13	0.27	0.3	0.28	0.28	0.3	0.24	0.28	0.28	0.24	0.29	0.19
ILP	Model metrics	B_p	0.3	0.39	0.64	0.58	0.84	0.63	0.32	0.31	0.42	0.48	0.47	0.33	0.47	0.3	0.52	0.62	0.63	0.42
		B_r	0.68	0.56	0.83	0.75	0.89	0.9	0.68	0.53	0.69	0.7	0.75	0.89	0.78	0.74	0.84	0.9	0.99	0.56
		S_p	0.28	0.53	0.63	0.69	0.75	0.76	0.32	0.53	0.43	0.34	0.41	0.87	0.49	0.45	0.45	0.66	0.66	0.29
		S_r	1.0	0.83	1.0	1.0	0.87	1.0	1.0	1.0	0.95	0.92	0.88	0.96	1.0	0.98	1.0	1.0	1.0	1.0
	Log metrics	P	0.48	0.51	0.73	0.64	0.87	0.55	0.54	0.65	0.58	0.0	0.75	0.56	0.63	0.74	0.64	0.48	0.54	0.52
		C	1.0	0.5	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0	0.49	1.0	1.0	1.0	1.0	1.0	1.0	0.82
		S	0.16	0.27	0.25	0.26	0.26	0.32	0.15	0.19	0.21	0.19	0.24	0.27	0.24	0.21	0.2	0.3	0.3	0.14

CiTIUS

ProDiGen

Experimentation - Results on balanced logs (i)

Logs with 5% of noise

Method	Metrics		Caminate	A8	D2	M11Skip	MaS	M21	MDriverLicensed	allLoops	Ma7	l2la	MexampleLog	MherbstFig6p37	Ma6nfc	MParallel5	NC	L21P	NCB	DWS
ProDiGen	Model metrics	B_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		B_r	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S_r	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Log metrics	P	1.0	1.0	1.0	0.82	0.91	0.85	0.9	0.85	1.0	0.96	0.96	0.96	1.0	1.0	1.0	0.56	0.87	0.78
		C	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S	0.25	0.32	0.3	0.32	0.31	0.34	0.29	0.3	0.29	0.3	0.3	0.34	0.32	0.31	0.3	0.31	0.31	0.3
GM	Model metrics	B_p	0.8	0.91	0.94	0.99	0.84	1.0	0.88	0.99	0.94	0.94	0.87	0.79	1.0	0.86	0.85	0.95	1.0	0.86
		B_r	0.99	0.99	0.99	0.96	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99	1.0	1.0	0.81
		S_p	0.66	0.77	0.77	0.92	0.66	0.88	0.8	1.0	0.8	0.9	0.68	0.76	0.92	0.66	0.69	0.9	1.0	0.83
		S_r	0.5	0.87	0.87	0.92	0.66	1.0	0.76	0.96	1.0	0.9	0.91	0.95	1.0	0.66	0.75	0.9	1.0	0.95
	Log metrics	P	0.68	0.64	0.87	0.68	0.66	0.66	0.54	0.6	0.78	0.86	0.54	0.61	0.92	0.47	0.63	0.55	0.87	0.6
		C	0.22	1.0	1.0	1.0	0.69	1.0	0.38	1.0	0.86	1.0	1.0	0.76	1.0	1.0	0.55	0.77	1.0	0.27
		S	0.29	0.28	0.29	0.27	0.3	0.32	0.28	0.28	0.26	0.3	0.28	0.31	0.29	0.29	0.28	0.31	0.31	0.29
HM	Model metrics	B_p	0.96	1.0	1.0	0.97	0.92	0.91	0.92	0.93	1.0	1.0	1.0	1.0	0.9	1.0	0.87	0.77	0.88	0.77
		B_r	1.0	1.0	1.0	0.92	0.79	0.99	0.95	0.83	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.95	1.0	0.94
		S_p	1.0	1.0	1.0	1.0	1.0	0.88	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	1.0	0.83
		S_r	0.92	1.0	1.0	0.76	0.91	1.0	0.9	0.83	1.0	1.0	1.0	1.0	0.91	1.0	0.91	0.8	0.88	1.0
	Log metrics	P	0.92	1.0	1.0	0.0	0.0	0.57	0.88	0.89	1.0	0.96	0.96	0.96	0.91	1.0	0.88	0.55	0.0	0.0
		C	0.66	1.0	1.0	0.0	0.0	1.0	0.63	0.26	1.0	1.0	1.0	1.0	0.69	1.0	0.85	0.77	0.0	0.0
		S	0.29	0.32	0.3	0.31	0.32	0.33	0.3	0.28	0.29	0.3	0.3	0.34	0.33	0.31	0.32	0.31	0.32	0.29
α^{++}	Model metrics	B_p	0.73	0.81	0.69	0.34	0.75	0.63	0.38	0.68	0.83	0.72	0.55	0.38	0.8	0.45	0.75	0.76	0.23	0.57
		B_r	0.69	0.99	0.83	0.32	0.88	0.68	0.49	0.67	0.97	0.65	0.6	0.47	0.88	0.54	0.99	0.56	0.24	0.66
		S_p	0.62	0.41	0.57	0.31	0.46	0.3	0.41	0.5	0.47	0.43	0.36	0.4	0.55	0.37	0.5	0.6	0.5	0.48
		S_r	0.82	0.62	0.5	0.38	0.5	0.37	0.57	0.58	0.66	0.63	0.66	0.56	0.91	0.6	0.83	0.6	0.66	0.72
	Log metrics	P	0.0	0.69	0.95	0.0	0.75	0.0	0.0	0.0	0.85	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		C	0.0	0.49	1.0	0.0	1.0	0.0	0.0	0.0	0.86	0.0	0.0	0.0	0.69	0.0	0.0	0.0	0.0	0.0
		S	0.18	0.27	0.32	0.18	0.29	0.21	0.11	0.16	0.25	0.17	0.2	0.34	0.25	0.25	0.21	0.21	0.18	0.16
ILP	Model metrics	B_p	0.13	0.15	0.24	0.48	0.31	0.14	0.07	0.27	0.05	0.14	0.18	0.1	0.24	0.34	0.28	0.36	0.26	0.25
		B_r	0.3	0.48	0.83	0.33	0.62	0.42	0.21	0.81	0.17	0.44	0.36	0.45	0.59	0.64	0.4	0.69	0.75	0.49
		S_p	0.23	0.23	0.32	0.26	0.26	0.17	0.16	0.33	0.23	0.21	0.13	0.09	0.16	0.16	0.23	0.3	0.27	0.13
		S_r	1.0	1.0	0.8	0.95	0.93	1.0	1.0	1.0	1.0	0.92	0.92	0.94	1.0	0.78	1.0	1.0	0.91	0.96
	Log metrics	P	0.32	0.41	0.42	0.31	0.0	0.27	0.18	0.45	0.58	0.3	0.39	0.0	0.42	0.42	0.48	0.33	0.31	0.36
		C	1.0	1.0	1.0	1.0	0.0	1.0	0.63	1.0	0.57	1.0	0.48	0.0	1.0	0.0	1.0	1.0	0.0	0.5
		S	0.1	0.18	0.27	0.12	0.17	0.16	0.09	0.24	0.16	0.14	0.14	0.11	0.13	0.14	0.15	0.21	0.19	0.11

Logs with 10% of noise

Method	Metrics		Caminate	A8	D2	M11Skip	MaS	M21	MDriverLicensed	allLoops	Ma7	l2la	MexampleLog	MherbstFig6p37	Ma6nfc	MParallel5	NC	L21P	NCB	DWS
ProDiGen	Model metrics	B_p	0.57	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		B_r	0.99	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S_p	0.73	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S_r	0.78	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
	Log metrics	P	0.86	1.0	1.0	0.82	0.91	0.86	0.9	0.86	1.0	0.96	0.96	0.96	1.0	1.0	1.0	0.56	0.87	0.78
		C	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
		S	0.25	0.32	0.3	0.32	0.31	0.34	0.29	0.3	0.29	0.3	0.3	0.34	0.32	0.31	0.3	0.31	0.31	0.3
GM	Model metrics	B_p	0.92	0.83	0.8	1.0	0.87	1.0	0.92	0.99	0.77	0.87	0.99	0.92	0.88	0.86	0.89	1.0	0.82	0.84
		B_r	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	1.0	0.99	0.99	0.99	0.99	0.99	0.99	0.83
		S_p	0.84	0.6	0.75	0.92	0.9	0.88	0.9	0.98	0.57	0.75	0.92	0.95	0.83	0.66	0.91	0.9	0.72	0.79
		S_r	0.78	0.75	0.75	1.0	0.83	1.0	0.85	0.96	0.66	0.81	1.0	0.95	0.83	0.66	0.91	1.0	0.88	0.95
	Log metrics	P	0.79	0.56	0.46	0.68	0.51	0.44	0.63	0.51	0.52	0.44	0.67	0.75	0.43	0.58	0.76	0.58	0.57	0.54
		C	0.23	0.5	1.0	1.0	1.0	1.0	1.0	0.61	0.59	1.0	0.53	1.0	1.0	1.0	0.17	1.0	1.0	0.57
		S	0.28	0.28	0.29	0.29	0.29	0.3	0.28	0.28	0.27	0.29	0.28	0.32	0.29	0.28	0.28	0.29	0.3	0.28
HM	Model metrics	B_p	0.96	1.0	0.91	1.0	0.92	0.86	0.92	0.93	1.0	0.94	1.0	1.0	0.9	1.0	0.87	0.96	0.88	0.86
		B_r	1.0	1.0	1.0	0.89	0.79	0.99	0.95	0.91	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	0.81
		S_p	1.0	1.0	1.0	1.0	1.0	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	0.83
		S_r	0.93	1.0	0.75	0.84	1.0	1.0	0.9	0.83	1.0	0.9	1.0	1.0	0.91	1.0	0.91	0.9	0.88	0.95
	Log metrics	P	1.0	1.0	0.92	0.0	0.0	0.57	0.89	0.82	1.0	0.89	0.96	0.96	0.91	1.0	0.88	0.0	0.0	0.0
		C	1.0	1.0	1.0	0.9	0.0	1.0	0.62	0.34	1.0	0.3	1.0	1.0	0.69	1.0	0.84	0.0	0.0	0.0
		S	0.25	0.32	0.33	0.31	0.31	0.32	0.3	0.3	0.29	0.32	0.3	0.34	0.33	0.31	0.32	0.3	0.32	0.29
α^{++}	Model metrics	B_p	0.85	0.84	0.84	0.86	0.64	0.61	0.35	0.74	0.84	0.65	0.63	0.33	0.54	0.63	0.62	0.68	0.32	0.43
		B_r	0.95	0.99	0.99	0.66	0.64	0.65	0.45	0.77	0.99	0.68	0.69	0.41	0.6	0.75	0.97	0.52	0.36	0.56
		S_p	0.61	0.46	0.71	0.46	0.3	0.35	0.35	0.41	0.44	0.58	0.37	0.36	0.31	0.29	0.25	0.45	0.66	0.31
		S_r	0.75	0.75	0.62	0.53	0.5	0.62	0.61	0.41	0.66	0.63	0.75	0.45	0.58	0.66	0.33	0.5	0.44	0.42
	Log metrics	P	0.0	0.85	0.85	0.0	0.0	0.0	0.0	0.0	0.77	0.71	0.0	0.0	0.0	0.0	0.76	0.7	0.0	0.0
		C	0.0	1.0	1.0	0.0	0.0	1.0	0.0	0.0	0.57	0.82	0.0	0.0	0.0	0.0	0.84	1.0	0.0	0.0
		S	0.21	0.27	0.32	0.19	0.19	0.22	0.13	0.19	0.25	0.21	0.2	0.34	0.21	0.21	0.25	0.29	0.31	0.15
ILP	Model metrics	B_p	0.16	0.25	0.2	0.33	0.32	0.2	0.12	0.28	0.21	0.12	0.15	0.05	0.13	0.17	0.1	0.41	0.24	0.15
		B_r	0.25	0.72	0.58	0.47	0.64	0.5	0.19	0.64	0.38	0.42	0.29	0.15	0.43	0.29	0.32	0.72	0.47	0.24
		S_p	0.24	0.19	0.34	0.27	0.18	0.18	0.15	0.36	0.19	0.25	0.16	0.06	0.16	0.13	0.18	0.28	0.28	0.12
		S_r	1.0	0.83	0.8	1.0	0.87	1.0	1.0	0.95	1.0	1.0	1.0	1.0	1.0	1.0	0.93	1.0	1.0	0.92
	Log metrics	P	0.36	0.41	0.44	0.34	0.0	0.26	0.17	0.45	0.43	0.37	0.37	0.0	0.37	0.3	0.32	0.36	0.28	0.28
		C	0.13	1.0	1.0	1.0	0.0	1.0	0.62	1.0	1.0	1.0	0.5	0.0	0.69	1.0	1.0	1.0	1.0	0.26
		S	0.12	0.19	0.22	0.15	0.18	0.18	0.09	0.21	0.15	0.16	0.15	0.07	0.15	0.1	0.14	0.2	0.18	0.11

CiTIUS

ProDiGen

Experimentation - Results on balanced logs (iii)

			Logs with 20% of noise																	
			Caminatas	A8	D2	M11Skip	Ma5	M12l	MDriverLL	aLLoops	Ma7	l2la	MexampleL	Herbst6p37	Ma6nfc	MParallel5	NC	L2LP	NCB	DWS-20
ProDiGen	Model metrics	B_p	0.57	1.0	0.91	1.0	1.0	1.0	0.62	1.0	0.78	0.8	0.75	1.0	0.75	0.92	0.72	1.0	1.0	0.5
		B_r	0.99	1.0	1.0	1.0	1.0	1.0	0.99	1.0	0.99	0.99	0.99	1.0	0.99	0.99	0.99	1.0	1.0	0.99
		S_p	0.75	1.0	0.87	1.0	1.0	1.0	0.77	1.0	0.79	0.81	0.76	1.0	0.76	0.78	0.76	1.0	1.0	0.47
		S_r	0.85	1.0	0.87	1.0	1.0	1.0	0.89	1.0	1.0	1.0	0.83	1.0	0.83	0.84	0.75	1.0	1.0	0.95
	Log metrics	P	0.62	1.0	0.95	0.81	0.91	0.85	0.89	0.35	1.0	0.95	0.9	0.95	0.68	0.79	0.7	0.55	0.86	0.75
		C	0.34	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.52	1.0	0.69	0.72	0.62	1.0	1.0	0.37
		S	0.27	0.31	0.31	0.31	0.3	0.33	0.28	0.29	0.28	0.29	0.3	0.33	0.3	0.29	0.29	0.3	0.3	0.28
GM	Model metrics	B_p	0.56	0.67	0.66	0.85	0.82	1.0	0.44	0.9	0.67	0.77	0.73	0.59	0.65	0.83	0.57	0.95	0.88	0.78
		B_r	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.99	0.99	0.99	0.97	0.99	0.99	0.99	0.99	0.99	0.81
		S_p	0.56	0.46	0.6	0.69	0.72	0.88	0.52	0.88	0.47	0.69	0.64	0.44	0.61	0.52	0.5	0.9	0.8	0.6
		S_r	0.6	0.75	0.75	0.69	0.66	1.0	0.52	0.95	0.66	0.81	0.91	0.65	0.66	0.66	0.66	0.9	0.88	0.95
	Log metrics	P	0.39	0.42	0.45	0.65	0.36	0.65	0.0	0.46	0.42	0.42	0.38	0.0	0.38	0.59	0.37	0.31	0.51	0.53
		C	0.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	1.0	0.0	0.0	0.3	1.0	0.16	1.0	1.0	0.54
		S	0.26	0.26	0.28	0.29	0.27	0.31	0.28	0.24	0.26	0.27	0.25	0.27	0.27	0.29	0.27	0.26	0.29	0.27
HM	Model metrics	B_p	0.97	0.7	0.95	0.9	0.9	0.67	0.9	0.93	1.0	0.94	1.0	1.0	0.89	1.0	0.75	0.77	0.88	0.7
		B_r	1.0	0.85	1.0	0.8	0.89	0.91	0.92	0.89	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.95	1.0	0.6
		S_p	1.0	0.77	1.0	1.0	1.0	0.6	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.8	1.0	0.76
		S_r	0.95	0.87	0.87	0.76	0.83	0.75	0.9	0.83	1.0	0.9	1.0	1.0	0.91	1.0	0.83	0.8	0.88	0.95
	Log metrics	P	0.89	0.0	0.95	0.0	0.85	0.0	0.0	0.89	1.0	0.88	0.95	0.95	0.9	0.99	0.0	0.0	0.0	0.0
		C	0.34	0.0	1.0	0.0	0.65	0.0	0.63	0.58	1.0	0.3	1.0	1.0	0.69	1.0	0.0	0.0	0.0	0.0
		S	0.28	0.3	0.31	0.28	0.31	0.3	0.3	0.28	0.28	0.31	0.29	0.33	0.32	0.3	0.32	0.3	0.32	0.28
α^{++}	Model metrics	B_p	0.83	0.88	0.84	0.39	0.61	0.4	0.29	0.73	0.76	0.42	0.4	0.29	0.44	0.75	0.53	0.62	0.29	0.26
		B_r	0.91	0.99	0.99	0.55	0.56	0.49	0.4	0.82	0.87	0.57	0.75	0.42	0.76	0.99	0.89	0.54	0.4	0.24
		S_p	0.59	0.46	0.62	0.23	0.26	0.26	0.18	0.2	0.27	0.25	0.17	0.21	0.23	0.22	0.26	0.15	0.27	0.29
		S_r	0.67	0.75	0.62	0.23	0.5	0.5	0.28	0.16	0.41	0.27	0.33	0.3	0.41	0.46	0.33	0.2	0.33	0.42
	Log metrics	P	0.0	0.72	0.84	0.0	0.0	0.54	0.0	0.6	0.65	0.0	0.59	0.95	0.0	0.75	0.0	0.0	0.0	0.0
		C	0.0	0.49	1.0	0.0	0.0	1.0	0.0	0.24	1.0	0.0	0.52	1.0	0.0	1.0	0.83	0.0	0.0	0.0
		S	0.2	0.25	0.29	0.15	0.17	0.31	0.08	0.28	0.24	0.15	0.2	0.33	0.21	0.21	0.24	0.16	0.19	0.14
ILP	Model metrics	B_p	0.1	0.09	0.27	0.35	0.31	0.11	0.24	0.29	0.17	0.1	0.2	0.05	0.07	0.2	0.17	0.35	0.41	0.14
		B_r	0.19	0.23	0.83	0.46	0.57	0.27	0.19	0.63	0.34	0.33	0.31	0.06	0.31	0.35	0.49	0.75	0.84	0.21
		S_p	0.23	0.23	0.45	0.28	0.21	0.17	0.19	0.26	0.2	0.22	0.14	0.05	0.14	0.1	0.19	0.29	0.25	0.12
		S_r	1.0	0.64	0.8	1.0	1.0	1.0	0.93	0.7	0.97	1.0	1.0	1.0	0.77	0.76	1.0	1.0	1.0	0.96
	Log metrics	P	0.31	0.4	0.0	0.3	0.28	0.2	0.16	0.44	0.38	0.27	0.3	0.13	0.28	0.26	0.31	0.26	0.22	0.22
		C	0.24	1.0	0.0	1.0	0.0	1.0	1.0	0.39	1.0	1.0	0.52	0.0	0.0	0.0	1.0	1.0	1.0	0.25
		S	0.08	0.13	0.26	0.14	0.12	0.12	0.08	0.22	0.14	0.12	0.11	0.04	0.12	0.1	0.16	0.18	0.16	0.07

CiTIUS

ProDiGen

Results on unbalanced logs

Logs

			g2	g3	g4	g5	g6	g7	g8	g9	g10	g12	g13	g14	g15	g19	g20	g21	g22	g23	g24	g25	EMT
ProDiGen	Model metrics	B_p	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.96	1.0
		B_r	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.99	1.0
		S_p	1.0	1.0	0.96	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.91	1.0
		S_r	1.0	1.0	0.97	1.0	1.0	1.0	0.94	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.98	0.91	1.0
	Log metrics	P	0.9	0.82	0.98	0.98	0.95	0.88	0.86	0.92	0.89	0.97	0.93	0.93	0.86	0.92	0.78	0.91	0.9	0.58	0.89	0.74	0.87
		C	1.0	1.0	1.0	1.0	1.0	1.0	0.52	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.98	1.0
		S	0.3	0.3	0.3	0.31	0.31	0.32	0.28	0.31	0.3	0.31	0.3	0.31	0.25	0.3	0.29	0.31	0.3	0.3	0.29	0.31	0.27
GM	Model metrics	B_p	1.0	0.61	0.78	1.0	1.0	1.0	0.84	0.96	0.99	1.0	0.98	0.61	0.8	0.98	1.0	1.0	0.97	0.57	0.83	0.81	1.0
		B_r	1.0	0.97	0.97	1.0	1.0	1.0	1.0	1.0	0.97	1.0	0.99	1.0	0.97	0.9	1.0	1.0	1.0	0.88	0.88	0.96	0.83
		S_p	1.0	0.81	0.81	1.0	1.0	1.0	1.0	0.97	0.9	1.0	0.95	0.95	0.88	0.95	1.0	1.0	0.85	0.76	0.75	0.76	0.85
		S_r	1.0	0.81	0.81	1.0	1.0	1.0	0.94	0.98	0.92	1.0	0.94	0.94	0.87	0.89	1.0	1.0	0.85	0.74	0.75	0.74	0.85
	Log metrics	P	0.9	0.42	0.98	0.98	0.95	0.88	0.0	0.94	0.91	0.97	0.96	0.74	0.0	0.0	0.78	0.91	0.86	0.0	0.88	0.49	0.81
		C	1.0	0.31	0.59	1.0	1.0	1.0	0.26	0.48	0.48	1.0	0.75	1.0	0.15	0.2	1.0	1.0	0.43	0.2	0.72	0.41	0.3
		S	0.3	0.31	0.3	0.31	0.31	0.32	0.26	0.3	0.29	0.31	0.3	0.31	0.24	0.29	0.29	0.31	0.3	0.28	0.3	0.28	0.3
HM	Model metrics	B_p	1.0	1.0	0.94	1.0	0.9	0.97	0.87	1.0	0.96	1.0	1.0	0.97	0.96	0.97	1.0	1.0	0.99	0.6	0.92	0.76	0.81
		B_r	1.0	0.98	0.92	1.0	0.98	0.97	0.99	0.98	0.95	1.0	1.0	0.97	0.98	1.0	1.0	1.0	0.99	1.0	0.88	0.94	0.96
		S_p	1.0	0.97	0.96	1.0	0.93	0.97	0.95	1.0	0.96	1.0	1.0	0.96	1.0	1.0	1.0	1.0	0.97	0.91	0.89	0.85	0.76
		S_r	1.0	0.97	0.86	1.0	0.97	1.0	0.86	1.0	0.96	1.0	1.0	0.92	0.86	0.9	1.0	1.0	0.91	0.94	0.81	0.85	0.74
	Log metrics	P	0.9	0.83	0.0	0.98	0.93	0.9	0.86	0.93	0.9	0.97	0.93	0.92	0.87	0.93	0.78	0.91	0.9	0.0	0.86	0.71	0.85
		C	1.0	1.0	1.0	1.0	0.66	1.0	0.52	0.74	0.78	1.0	1.0	0.91	0.87	0.85	1.0	1.0	0.9	0.0	0.93	0.23	0.37
		S	0.3	0.3	0.31	0.31	0.31	0.31	0.28	0.31	0.3	0.31	0.3	0.32	0.26	0.3	0.29	0.31	0.3	0.29	0.29	0.3	0.29
α^{++}	Model metrics	B_p	0.8	0.81	0.95	1.0	0.88	0.92	0.87	0.96	0.94	0.95	0.93	0.58	0.85	0.9	0.64	0.84	0.93	0.71	0.81	0.94	0.99
		B_r	0.85	0.91	0.94	1.0	0.99	0.9	0.94	0.94	0.9	1.0	0.94	0.92	0.85	0.94	0.94	0.89	0.85	0.67	0.7	0.92	0.93
		S_p	0.89	0.94	0.98	1.0	0.79	0.98	0.83	0.98	0.81	0.96	0.88	0.86	0.9	0.9	0.92	0.93	0.82	0.66	0.85	0.91	1.0
		S_r	1.0	0.98	0.97	1.0	1.0	1.0	0.94	1.0	1.0	1.0	0.97	1.0	0.97	1.0	0.87	1.0	0.97	1.0	1.0	0.98	0.91
	Log metrics	P	0.86	0.0	0.95	0.98	0.94	0.0	0.85	0.94	0.91	0.94	0.95	0.75	0.0	0.9	0.64	0.0	1.0	0.0	0.0	0.0	0.99
		C	0.33	0.0	1.0	1.0	0.45	0.0	0.35	0.48	0.563	1.0	0.48	0.0	0.05	0.25	0.46	0.68	0.43	0.0	0.0	0.97	0.89
		S	0.27	0.28	0.31	0.31	0.28	0.31	0.26	0.28	0.28	0.3	0.28	0.29	0.25	0.27	0.3	0.29	0.18	0.18	0.2	0.29	0.28
ILP	Model metrics	B_p	0.9	0.89	1.0	1.0	0.96	0.96	0.9	0.94	0.92	0.95	0.93	0.85	0.87	0.91	0.72	0.87	0.9	0.58	0.78	0.66	0.98
		B_r	1.0	0.98	0.99	1.0	0.99	0.99	0.99	0.99	0.97	1.0	0.99	0.99	1.0	0.99	0.95	1.0	0.99	0.88	0.94	0.96	1.0
		S_p	0.83	0.85	0.98	1.0	0.78	0.94	0.76	0.89	0.73	0.96	0.78	0.78	0.67	0.85	0.85	0.92	0.72	0.5	0.77	0.64	0.91
		S_r	1.0	0.98	0.97	1.0	1.0	1.0	0.96	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.97	1.0	1.0	0.98	1.0
	Log metrics	P	0.87	0.79	0.99	0.98	0.93	0.89	0.83	0.9	0.84	0.94	0.9	0.88	0.82	0.9	0.83	0.87	0.88	0.41	0.78	0.54	0.87
		C	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.52	1.0	0.98	1.0	1.0	1.0	1.0
		S	0.29	0.29	0.29	0.31	0.28	0.31	0.25	0.3	0.26	0.3	0.3	0.27	0.23	0.28	0.27	0.31	0.23	0.17	0.26	0.25	0.27

CiTIUS

ProDiGen

Experimentation – Non-parametric tests

- Friedman test and Holm post hoc test:
  - Balanced Logs

Algorithm	Ranking
ProDiGen	1.52
HM	2.74
GM	2.89
ILP	3.74
α^{++}	4.10
Friedman p-value: 5.34E-11	

i	Comp.	z	p	α/i	Hypothesis
4	α^{++}	10.9	7.70E-28	0.012	Rejected
3	ILP	9.43	4.17E-21	0.016	Rejected
2	GM	5.80	6.70E-9	0.025	Rejected
1	HM	5.19	2.15E-7	0.05	Rejected

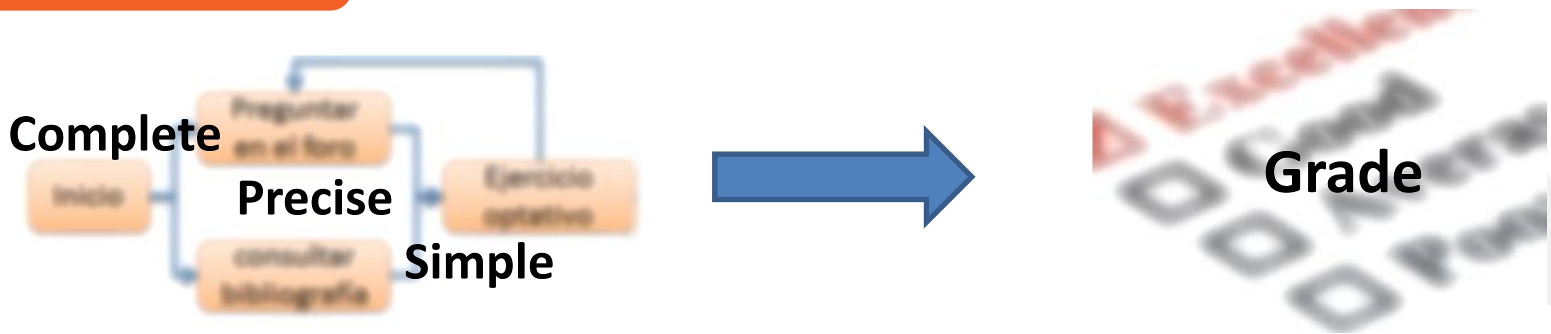
Unbalance Logs

Algorithm	Ranking
ProDiGen	1.55
HM	2.62
ILP	2.95
GM	3.62
α^{++}	4.26
Friedman p-value: 3.58E-7	

i	Comp.	z	p	α/i	Hypothesis
4	α^{++}	5.56	2.65E-8	0.012	Rejected
3	GM	4.25	2.18E-5	0.016	Rejected
2	ILP	2.88	0.003	0.025	Rejected
1	HM	2.20	0.03	0.05	Rejected

SoftLearn

Education

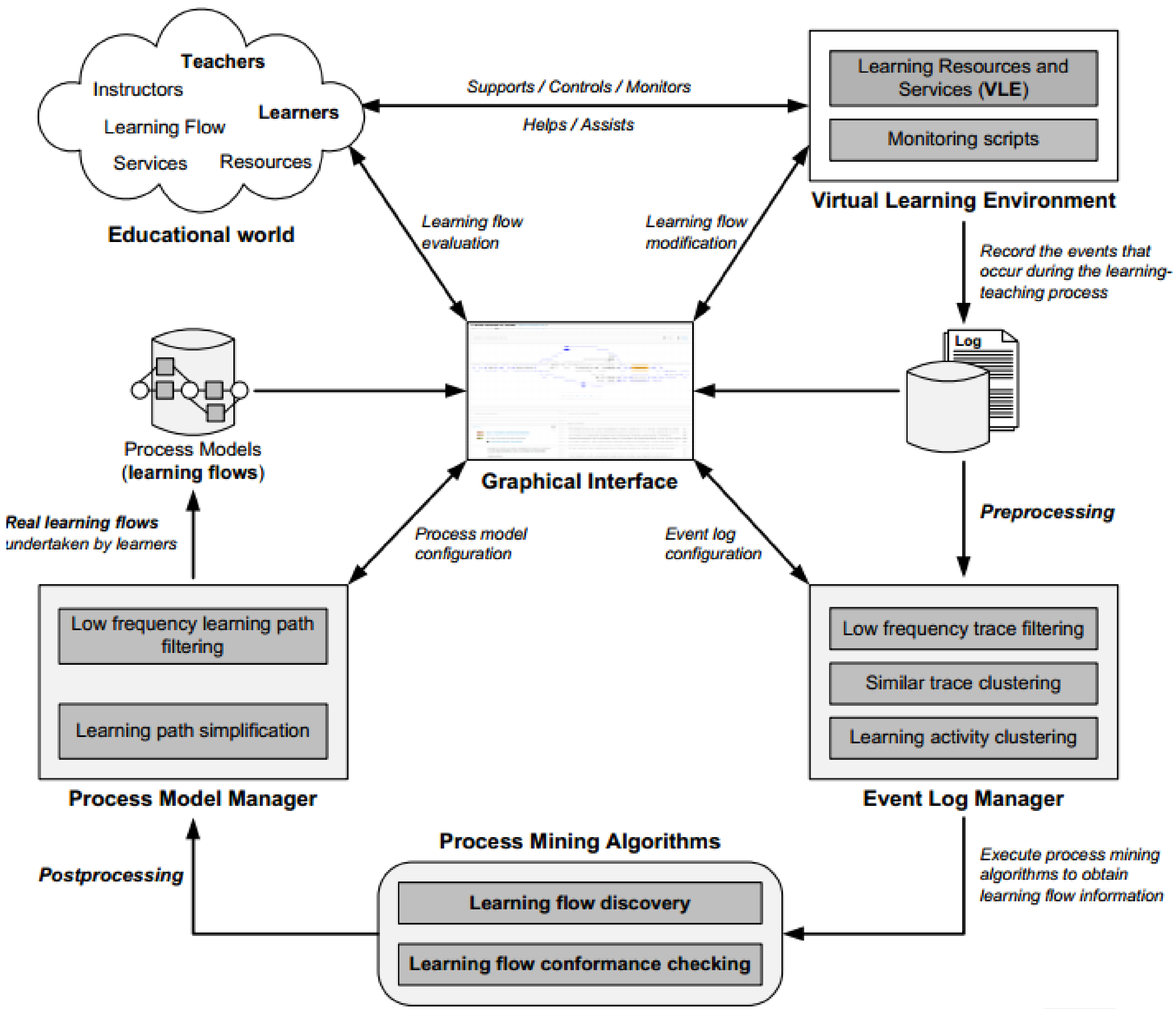


- Evaluate learning paths from a virtual learning enviroment
- Intuitive GUI to visualize:
  - The real learning path of the course
  - The learning content generated in the VLE

SoftLearn

SoftLearn

Framework



CiTIUS

SoftLearn

- Integrated in the virtual learning enviroment Elgg.
- Used as evaluation software in the present course 13/14.
- **Tecnología Educativa**, *Departamento de Didáctica y Organización Escolar – USC*
  - **72** enrolled students.

CiTIUS

Future work

SO, WHAT'S NEXT?

1. Multi-objective algorithm*.*
2. Generalization
3. Conformance
4. Other domains: Medicine (**QUIRAV**)

Publications

JOURNALS

1. Borja Vazquez-Barreiros, Manuel Mucientes, Manuel Lama: ***ProDiGen: Mining complete, precise, and minimal structure process models with a genetic algorithm***. Information Sciences. (**Under review**)
   - JCR 3.64; Ranking 6/131 (Q1) in Computer Science, Information Systems

CONFERENCES

1. Borja Vazquez-Barreiros, Manuel Lama, Manuel Mucientes, Juan C. Vidal*:* ***SoftLearn, a process mining platform for the discovery of learning paths****. IEEE International Conference on Advanced Learning Technologies* (ICALT 2014)*. (***Accepted***)*
   - Ranking 8/58 in Computer Education

2. A. Rodriguez, A. Gewerc, M. Lama, B. Vazquez-Barreiros, M. Mucientes: ***Using a learning analytics tool for evaluation in self-regulated learning.*** *Frontiers in Education* (FIE 2014) (**Abstract accepted**)
   - Ranking 7/58 in Computer Education

CiTIUS

ProDiGen: a genetic algorithm for process discovery guided by completeness, precision and simplicity

Borja Vázquez Barreiros

Supervisors : Manuel Mucientes, Manuel Lama

Centro Singular de Investigación en Tecnoloxías da Información

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA