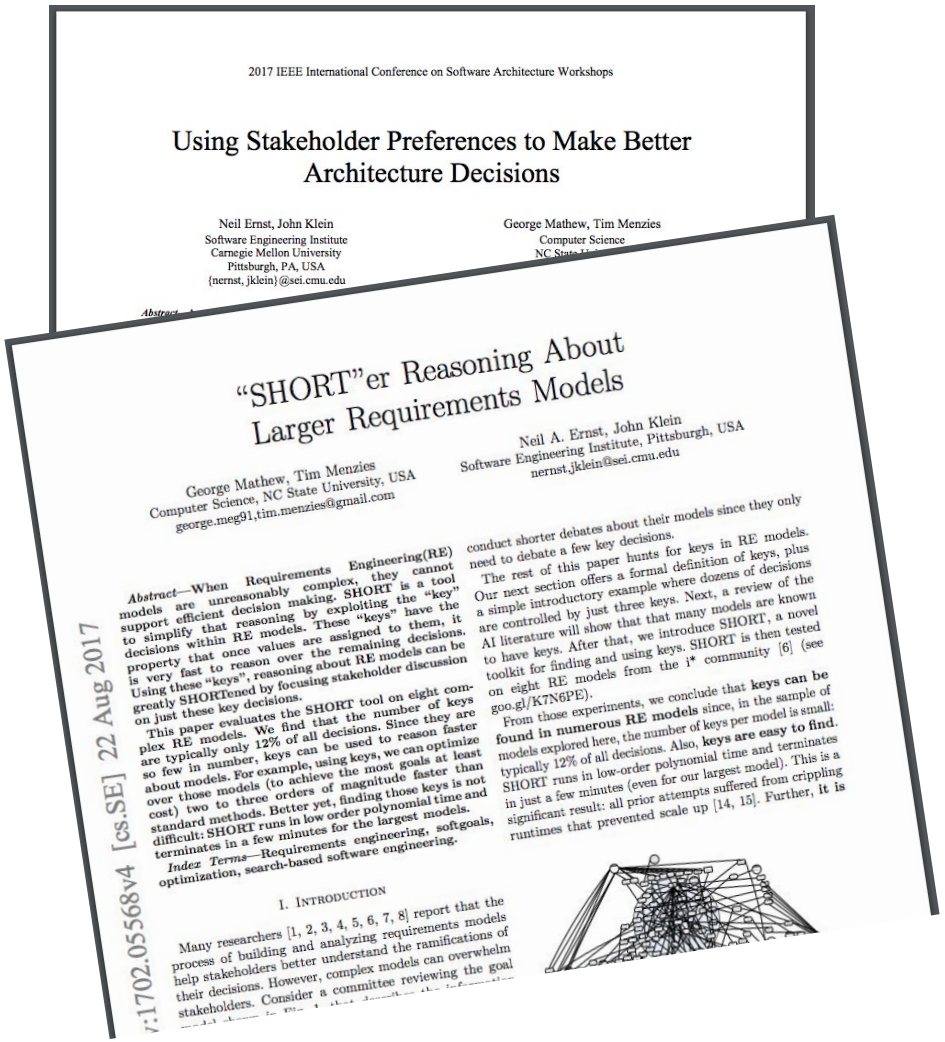
"SHORT"er Reasoning About Larger Requirement Models

George Mathew
bigfatnoob.us
Nov 15, 2017

Adviser: Dr. Tim Menzies

This Exam



2017 IEEE International Conference on Software Architecture Workshops

Using Stakeholder Preferences to Make Better Architecture Decisions

Neil Ernst, John Klein
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA, USA
{nernst, jklein}@sei.cmu.edu

George Mathew, Tim Menzies
Computer Science
NC State

"SHORT"er Reasoning About Larger Requirements Models

George Mathew, Tim Menzies
Computer Science, NC State University, USA
george.meg91,tim.menzies@gmail.com

Neil A. Ernst, John Klein
Software Engineering Institute, Pittsburgh, USA
nernst,jklein@sei.cmu.edu

Abstract—When Requirements Engineering(RE) models are unreasonably complex, they cannot support efficient decision making. SHORT is a tool to simplify that reasoning by exploiting the "key" decisions within RE models. These "keys" have the property that once values are assigned to them, it is very fast to reason over the remaining decisions. Using these "keys", reasoning about RE models can be greatly SHORTened by focusing stakeholder discussion on just these key decisions.

This paper evaluates the SHORT tool on eight complex RE models. We find that the number of keys are typically only 12% of all decisions. Since they are so few in number, keys can be used to reason faster about models. For example, using keys, we can optimize over those models (to achieve the most goals at least cost) two to three orders of magnitude faster than standard methods. Better yet, finding those keys is not difficult: SHORT runs in low order polynomial time and terminates in a few minutes for the largest models.

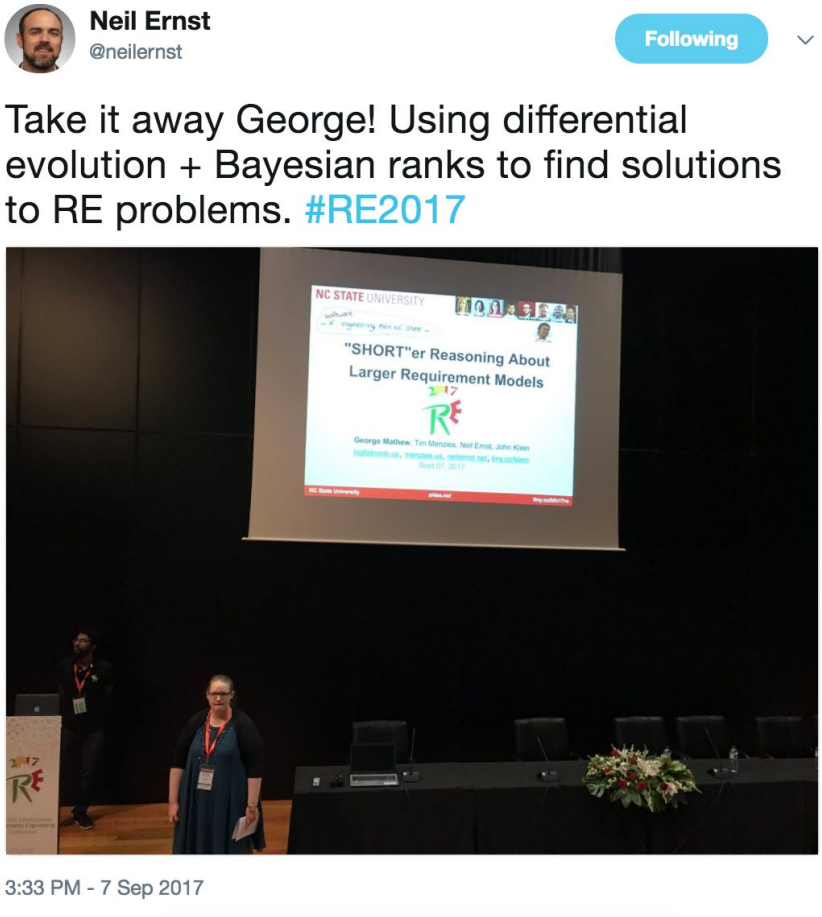
Index Terms—Requirements engineering, softgoals, optimization, search-based software engineering.

I. Introduction

Many researchers [1, 2, 3, 4, 5, 6, 7, 8] report that the process of building and analyzing requirements models help stakeholders better understand the ramifications of their decisions. However, complex models can overwhelm stakeholders. Consider a committee reviewing the goal conduct shorter debates about their models since they only need to debate a few key decisions.

The rest of this paper hunts for keys in RE models. Our next section offers a formal definition of keys, plus a simple introductory example where dozens of decisions are controlled by just three keys. Next, a review of the AI literature will show that that many models are known to have keys. After that, we introduce SHORT, a novel toolkit for finding and using keys. SHORT is then tested on eight RE models from the i* community [6] (see goo.gl/K7N6PE).

From those experiments, we conclude that **keys can be found in numerous RE models** since, in the sample of models explored here, the number of keys per model is small: typically 12% of all decisions. Also, **keys are easy to find.** SHORT runs in low-order polynomial time and terminates in just a few minutes (even for our largest model). This is a significant result: all prior attempts suffered from crippling runtimes that prevented scale up [14, 15]. Further, **it is**



Neil Ernst
@neilernst

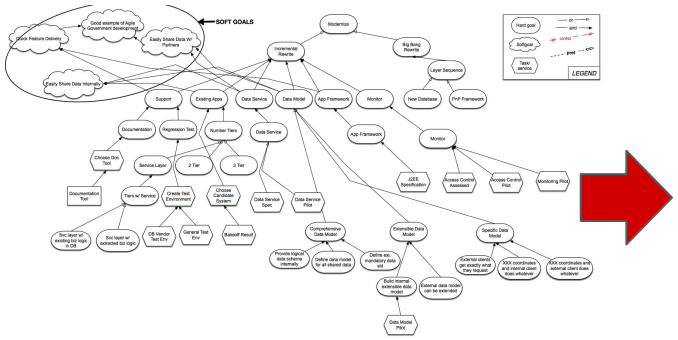
Following

Take it away George! Using differential evolution + Bayesian ranks to find solutions to RE problems. #RE2017

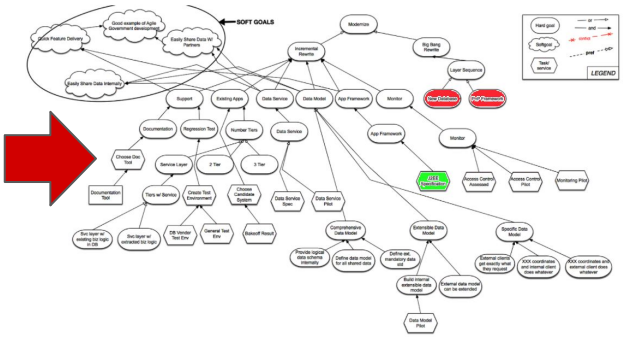
3:33 PM - 7 Sep 2017

Research Statement

"Given an interacting requirements modelling problem with a large search space and conflicts, identify key decisions in the models such that the subsequent search space can be drastically reduced."

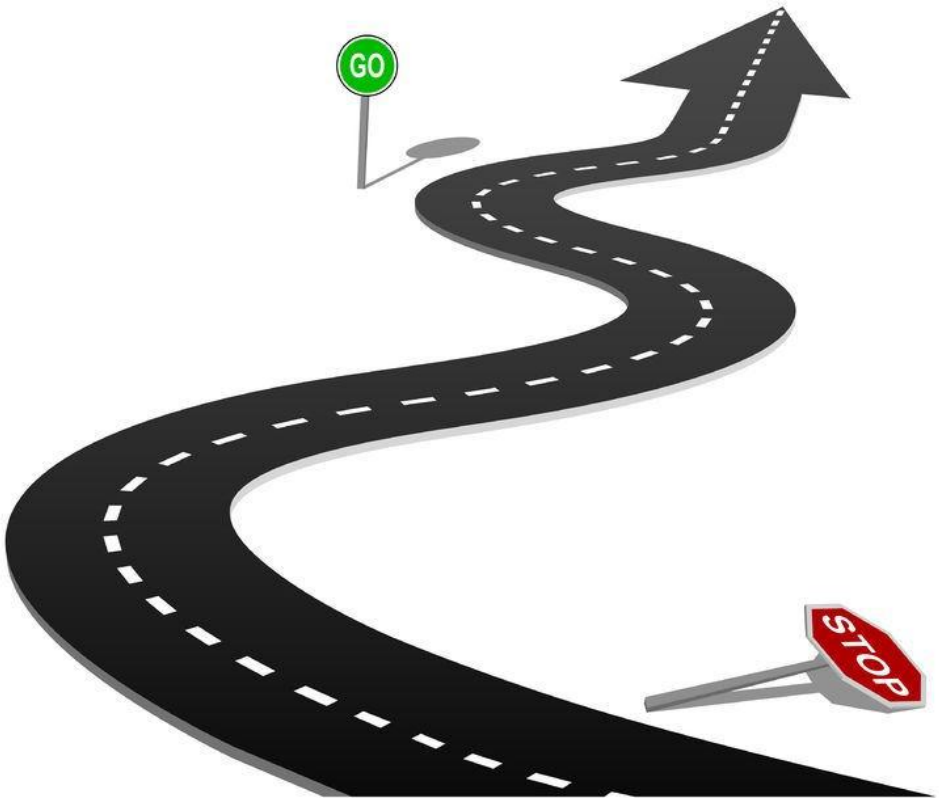


Rank	Node	Status	Support
1	J2EE Specification	ON	0.129
2	Pnp Framework	OFF	0.124
3	New Database	OFF	0.115
4	Documentation Tool	ON	0.114
5	Access Control Assessed	ON	0.113
6	Monitoring Pilot	ON	0.112
7	General Test Env	ON	0.110
8	Bakeoff Result	ON	0.110
9	Access Control Pilot	ON	0.108
10	DB Vendor Test Env	ON	0.105
11	Data Service Spec	ON	0.099
12	External clients get their request	ON	0.098
13	XXX coordinates & internal client	ON	0.098
14	XXX coordinates & external client	ON	0.097
15	Data Model Pilot	ON	0.095
16	Data Service Pilot	ON	0.095
17	2 Tier	ON	0.094
18	3 Tier	ON	0.090
19	Define data model for shared data	ON	0.085
20	Svc layer w/ extracted biz logic	OFF	0.080
21	Define ext mandatory data std	ON	0.079
22	Svc layer w/ extracted biz logic in DB	ON	0.066
23	External data model can be extended	ON	0.062
24	Provide logical data scheme internally	ON	0.052



Roadmap

- **Background**
- Related Work
- Data
- SHORT
- Research Questions
- Conclusions
- Future Work



Example - CS Department Services

- Decisions:
  - Tech Portal
  - Counselling Section
  - Help Module

Example - CS Department Services

- Decisions:
  - Tech Portal
  - Counselling Section
  - Help Module
- Hard Goals
  - Provide Phone Services?
  - Does tech allow dialogue?
  - Is content public?

Example - CS Department Services

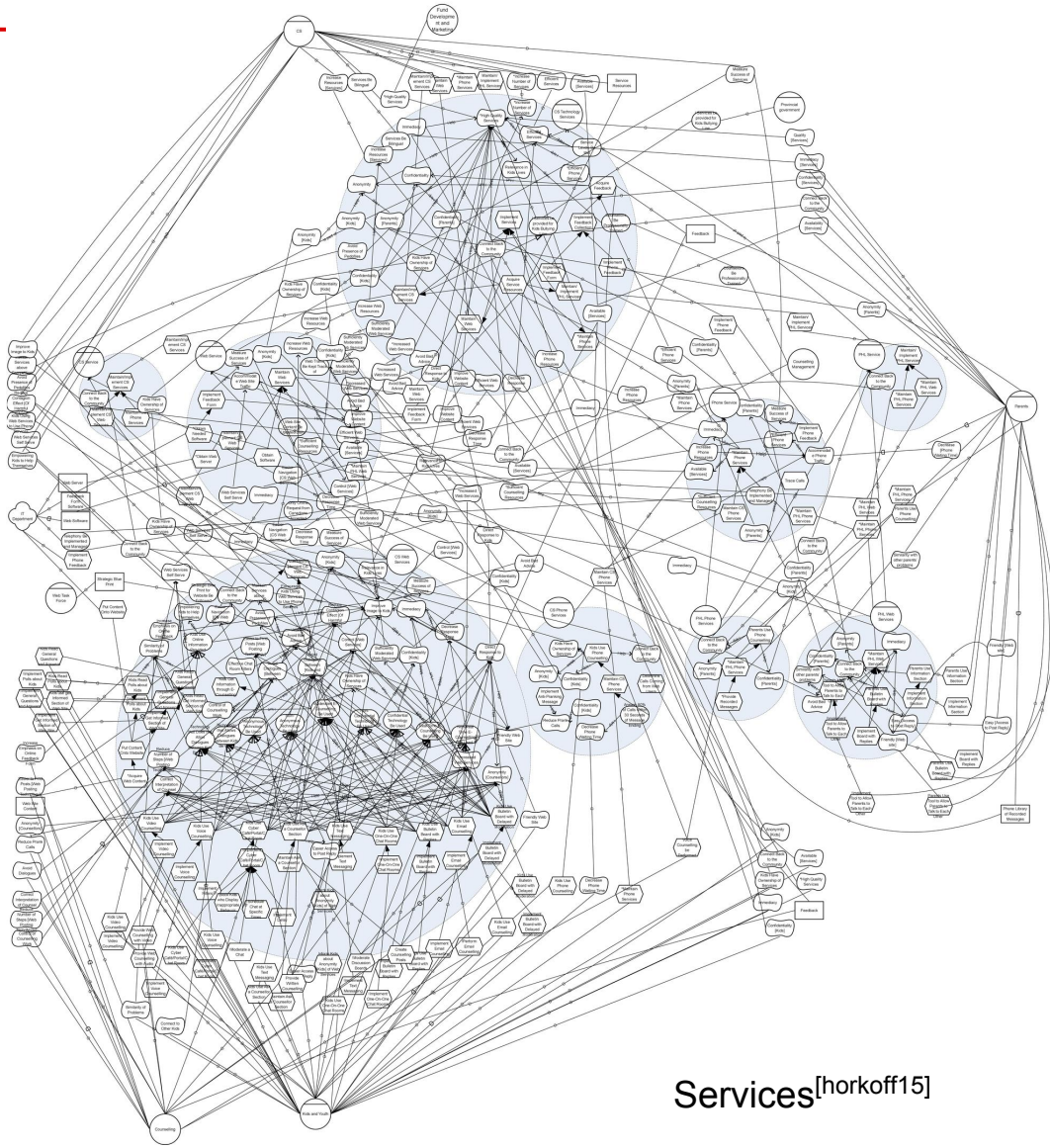
- Decisions:
  - Tech Portal
  - Counselling Section
  - Help Module
- Hard Goals
  - Provide Phone Services?
  - Does tech allow dialogue?
  - Is Content Public?
- Soft Goals:
  - Improve student feedback.
  - Maintain anonymity.
  - Decrease Response Time

Example - CS Department Services

- Decisions:
  - Tech Portal
  - Counselling Section
  - Help Module
- Hard Goals
  - Provide Phone Services?
  - Does tech allow dialogue?
  - Is Content Public?
- Soft Goals:
  - Improve student feedback.
  - Maintain anonymity.
  - Decrease Response Time
- Conflicts:
  - Encryption but content is public.
  - Tech portal is phone but communication protocol is SMTP

Example - CS Department Services

- Decisions
- Hard Goals
- Soft Goals
- Conflicts
- Nodes : 351
- Edges : 510
- Conflicts: ??



Services[horkoff15]

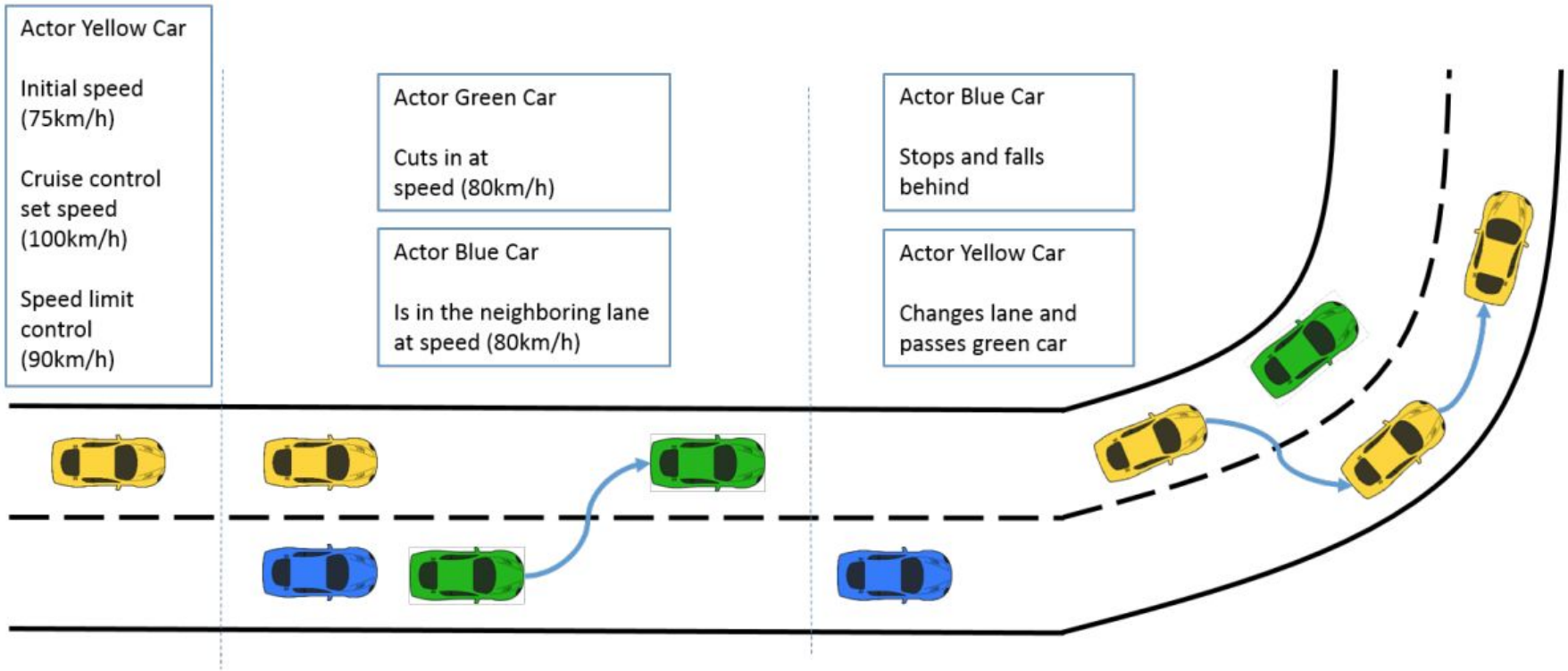
Example - CS Department Services

- Decisions
- Hard Goals
- Soft Goals
- Conflicts
- Nodes : 351
- Edges : 510
- Conflicts: ??

Services[horkoff15]

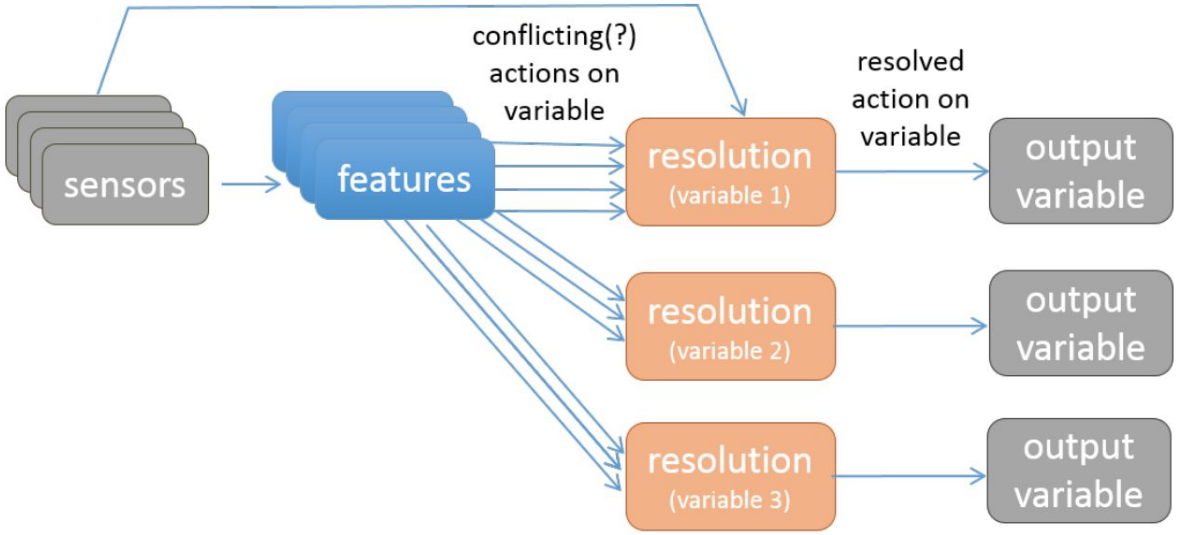
Other Real World Examples

Self Driving Cars[Zibaeenejad17]



Other Real World Examples

Self Driving Cars



- Appropriate resolution for interaction within a context.
- Contexts to consider exponentially with features

Other Real World Examples

Software for Robots

> ***….. Designing this software involves many choices, which can manifest as constants or variables in the source code and can range in granularity, from selecting between different types of algorithms for a task to selecting values for some algorithm's parameters …..***
>
> **[Kawthekar16]**

Other Real World Examples

Software for Robots

> ***….. Designing this software involves many choices, which can manifest as constants or variables in the source code and can range in granularity, from selecting between different types of algorithms for a task to selecting values for some algorithm's parameters …..***
>
> **[Kawthekar16]**

Robot Configuration[Jamshidi17]:

- Thousands to Million config settings
- Config interactions leads to exponential search space

Shortening Search across CS

Shortening Search across CS

- **Machine Learning**:
  - Feature subset selection[Kohavi97]
  - Instance Selection[Wilson00]
  - Principal Component Analysis[Pearson01]

Shortening Search across CS

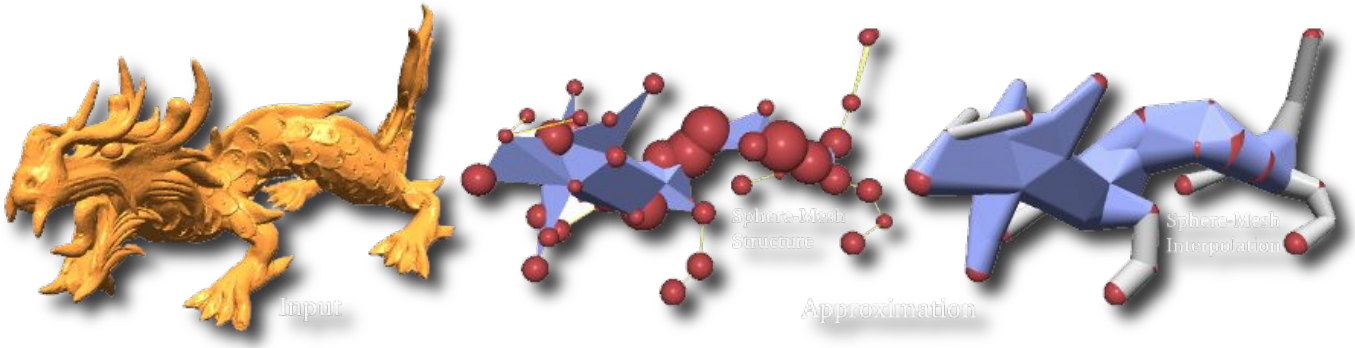
- Machine Learning
- **Theorem Proving**:
  - Narrows[Amarel66]
  - Master Variables[Crawford94]
  - Backdoors[Williams03]

Shortening Search across CS

- Machine Learning
- Theorem Proving
- **Software Engineering**:
  - Mutation Testing[Budd81]
  - Software Effort Estimation[Boehm81][Menizes16]

In the same world

- Machine Learning
- Theorem Proving
- Software Engineering
- **Computer Vision:**



[Thiery13]

Shortening Search across CS

- Machine Learning
- Theorem Proving
- Software Engineering
- Computer Vision

"Keys"

What are “Keys”?

Decisions that most reduce variance

What are “Keys”?

Decisions that most reduce variance

Stable model @ optimal value

What are “Keys”?

Decisions that most reduce variance

Stable model @ optimal value

Once set, other decisions irrelavant

What are “Keys”?

Decisions that most reduce variance

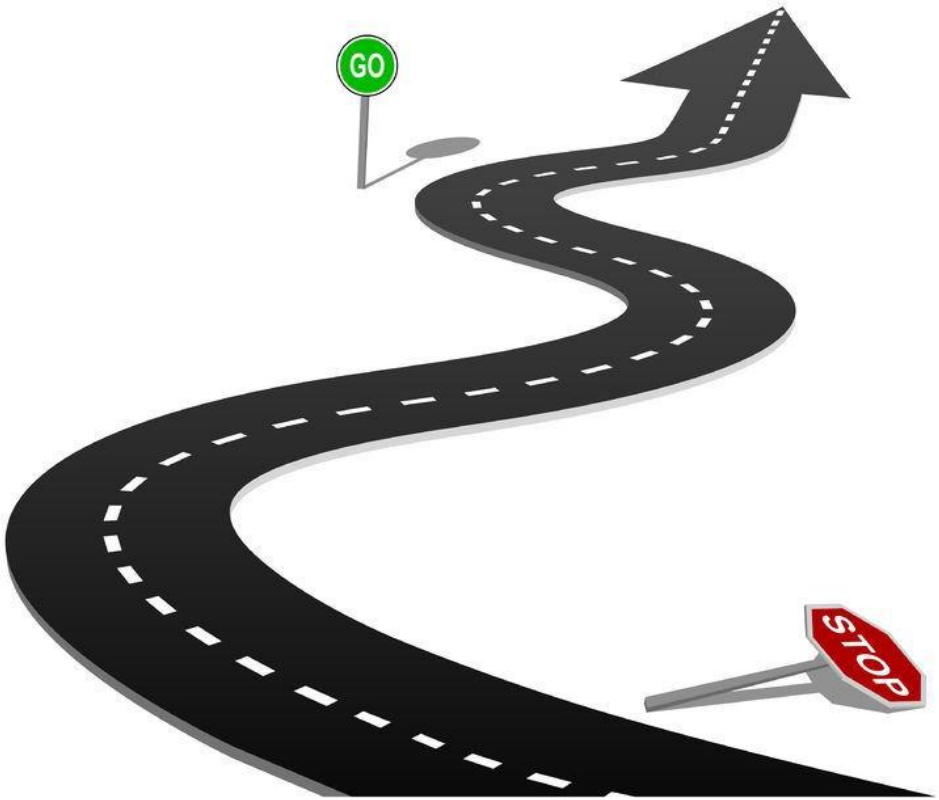
Stable model @ optimal value

Once set, other decisions irrelavant

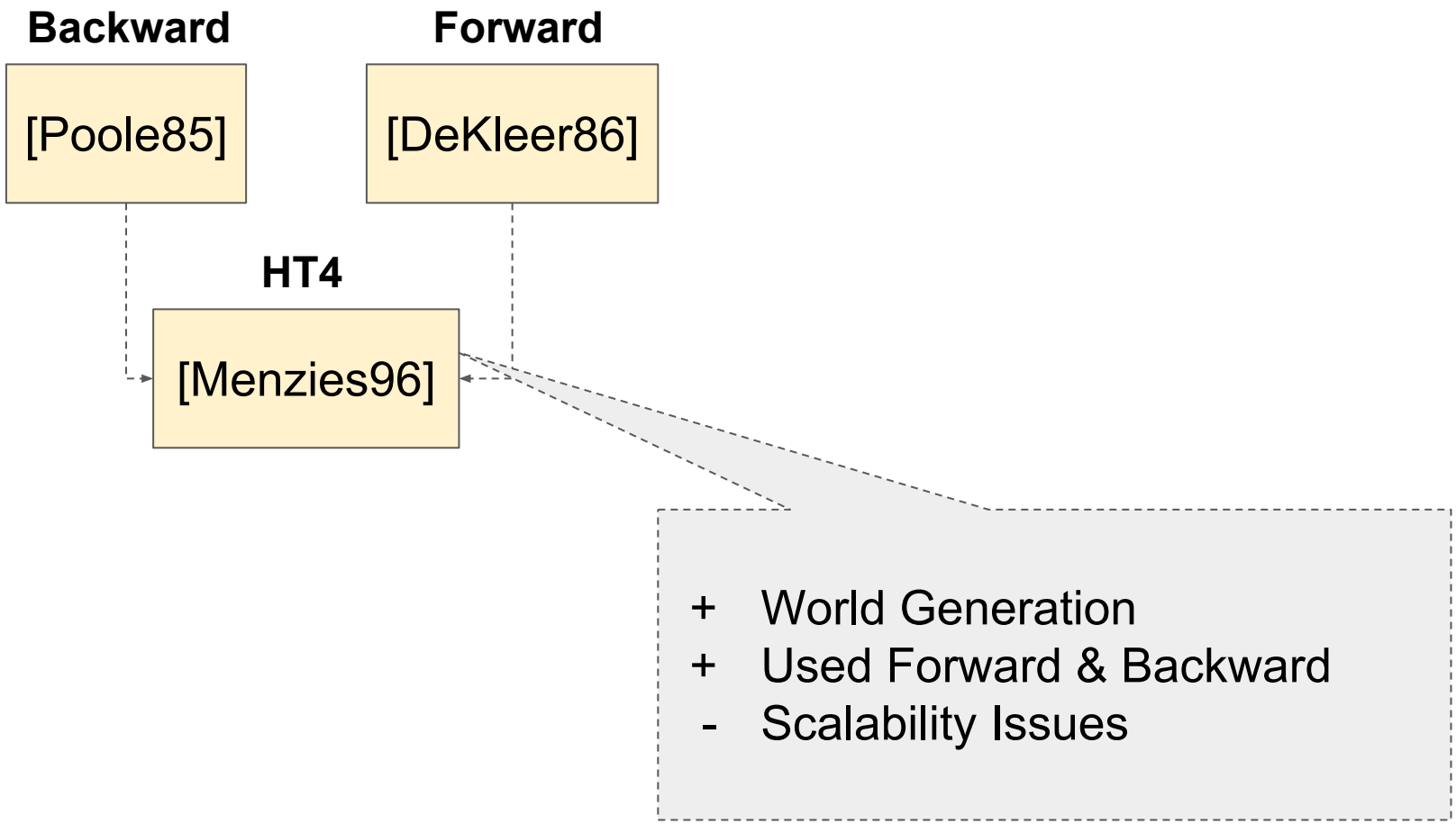
Decisions with large effects

Roadmap

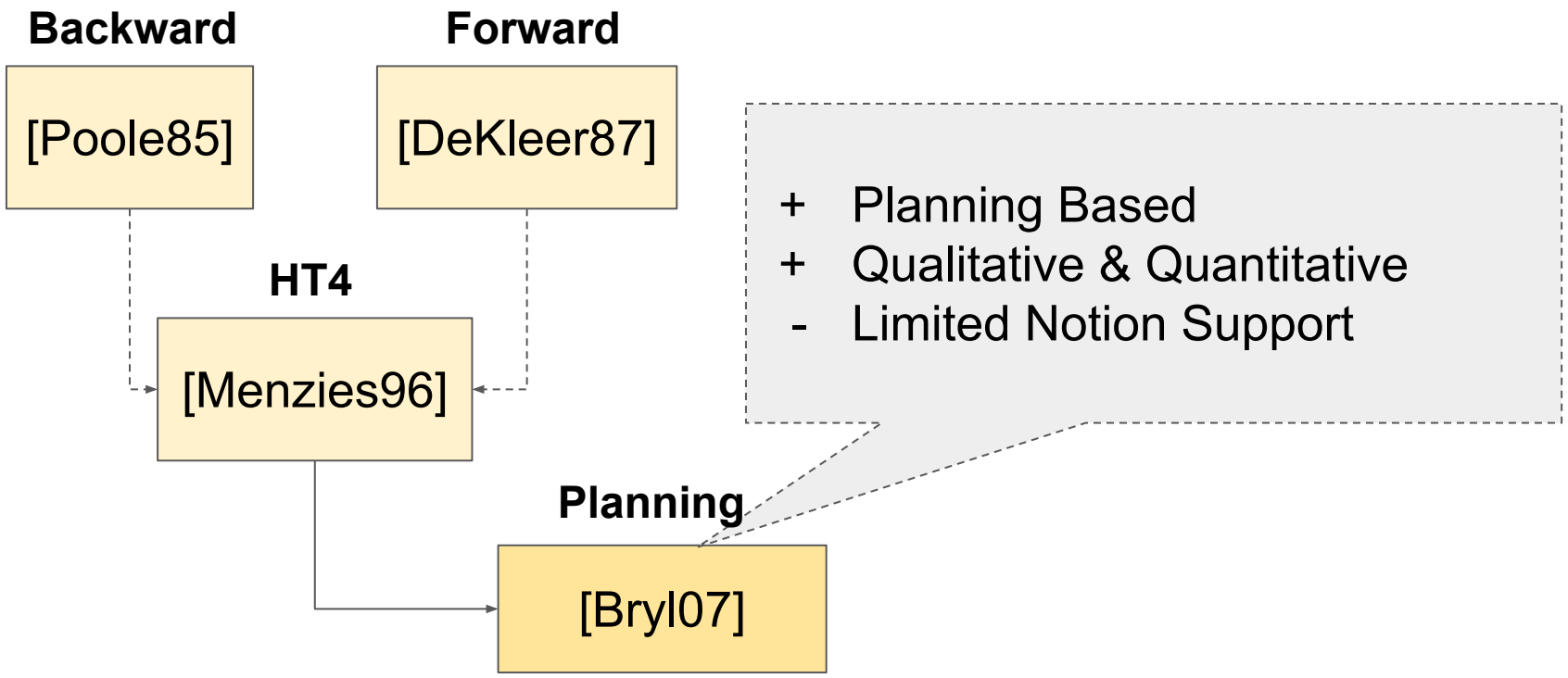
- **Background**
- **Related Work**
- Data
- SHORT
- Research Questions
- Conclusions
- Future Work



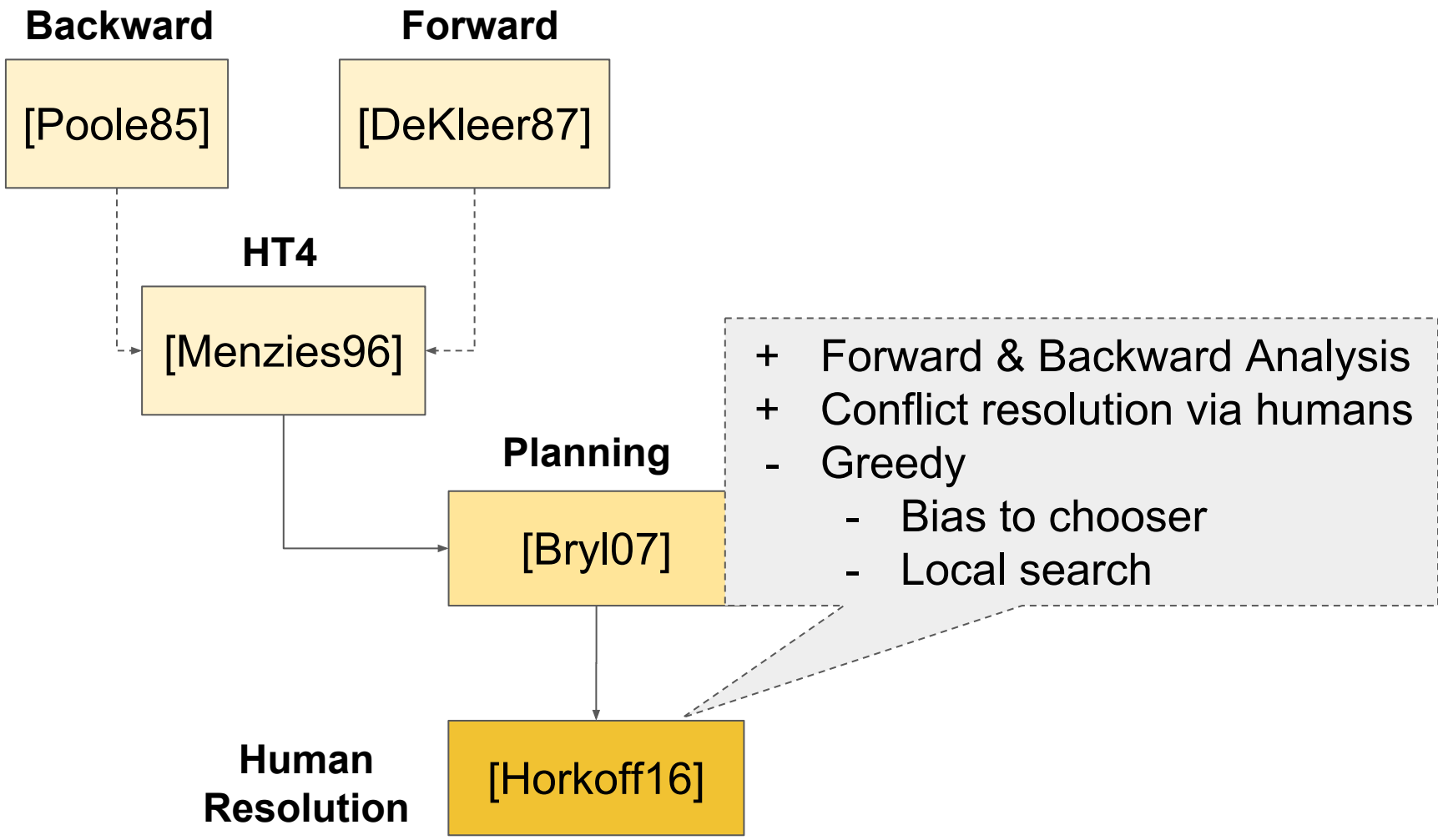
Related Work



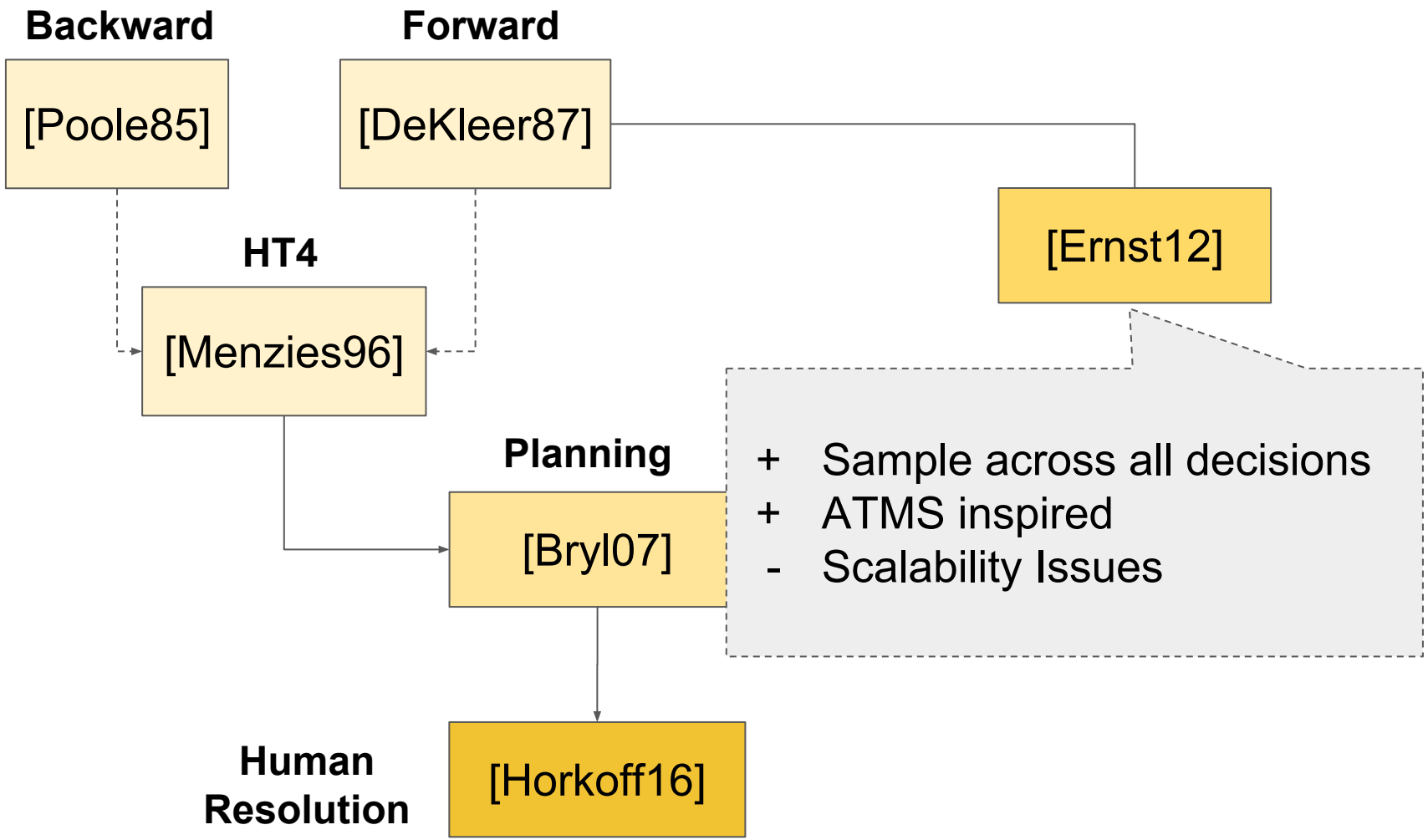
Related Work



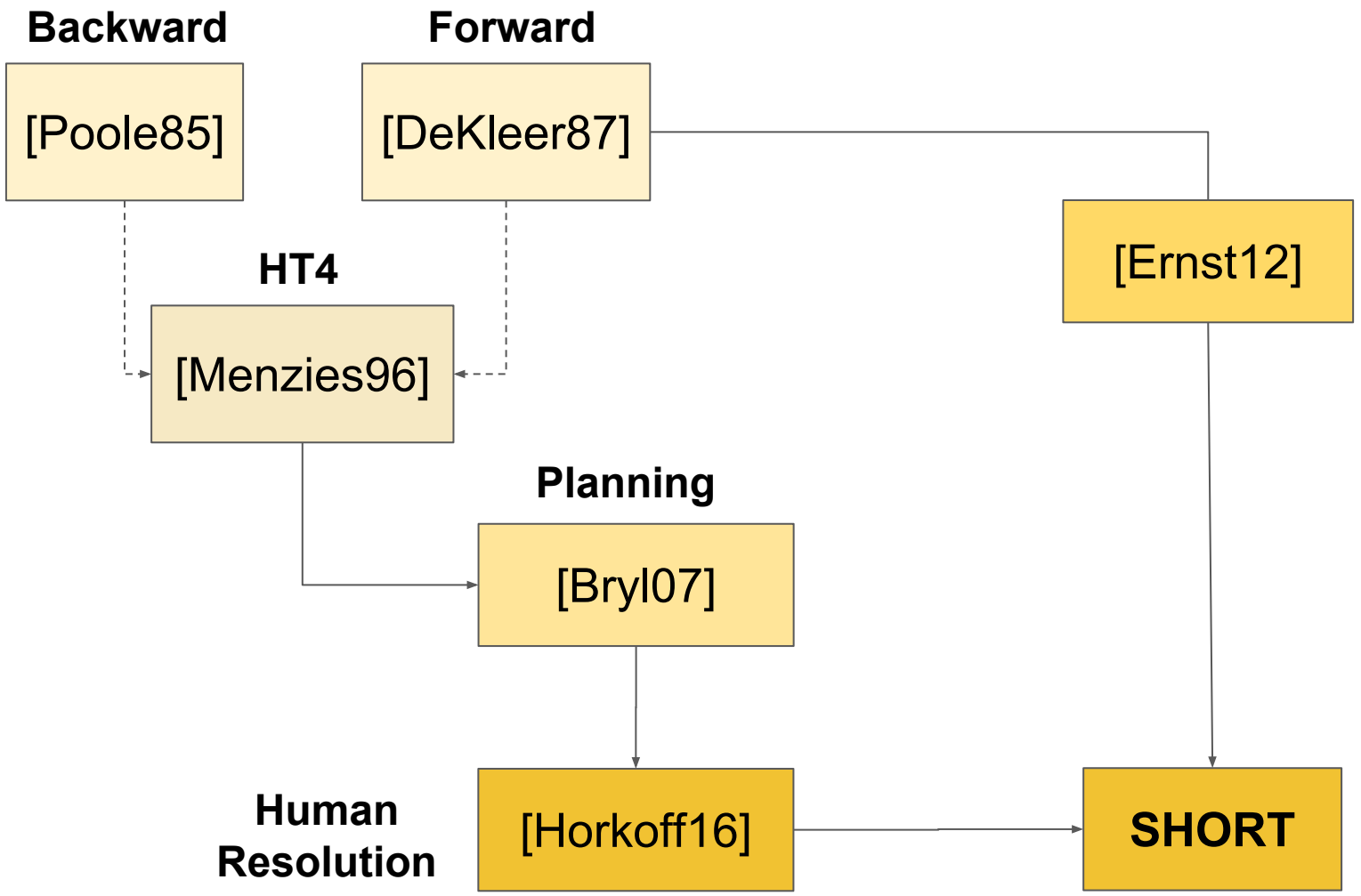
Related Work



Related Work

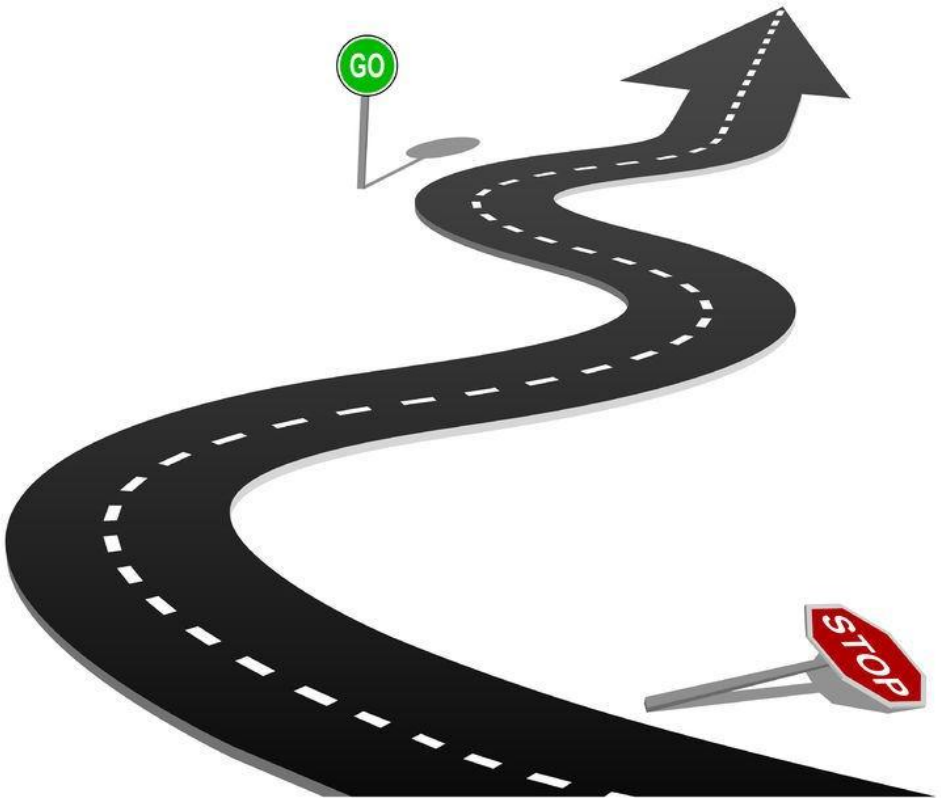


Related Work

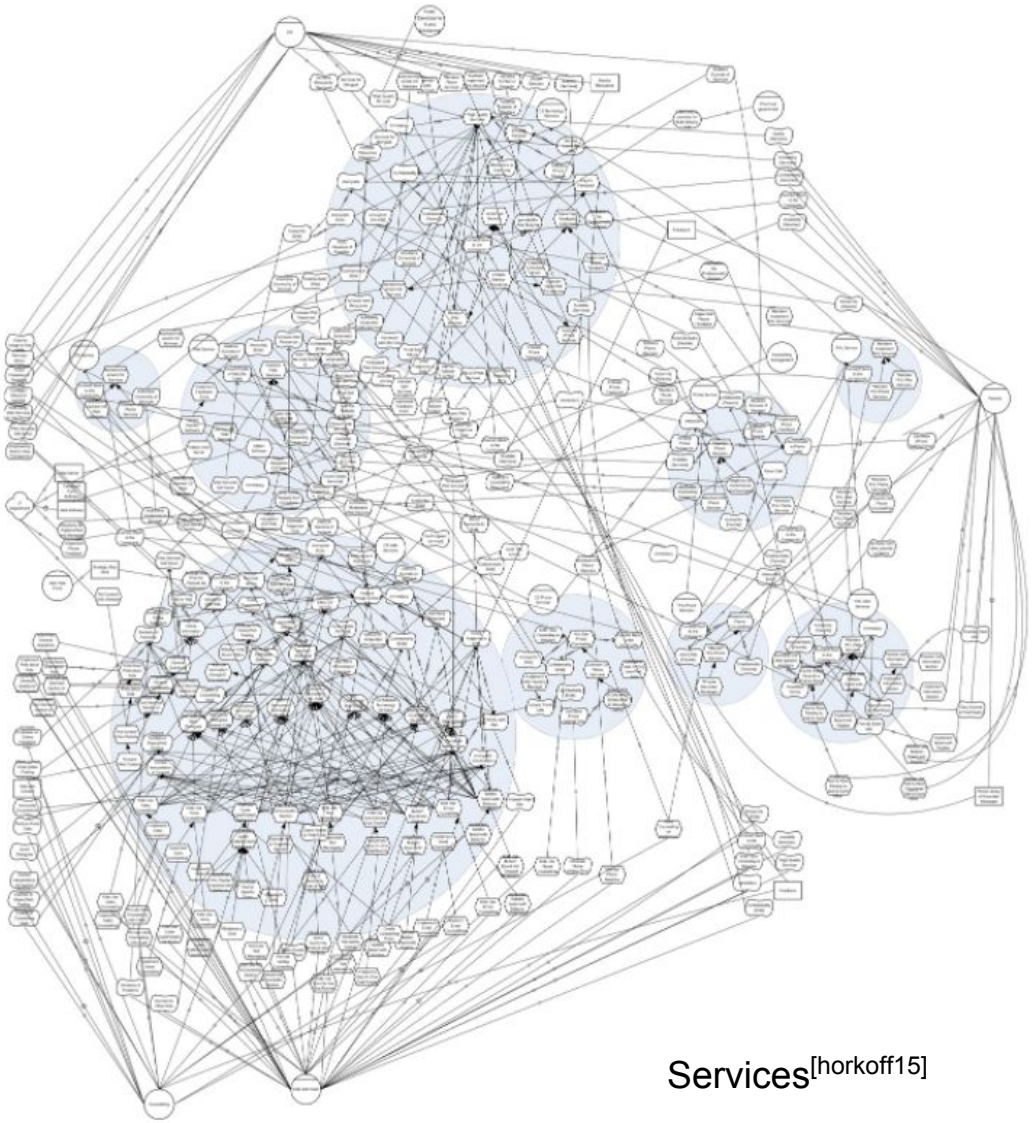


Roadmap

- **Background**
- **Related Work**
- **Data**
- SHORT
- Research Questions
- Conclusions
- Future Work



Data

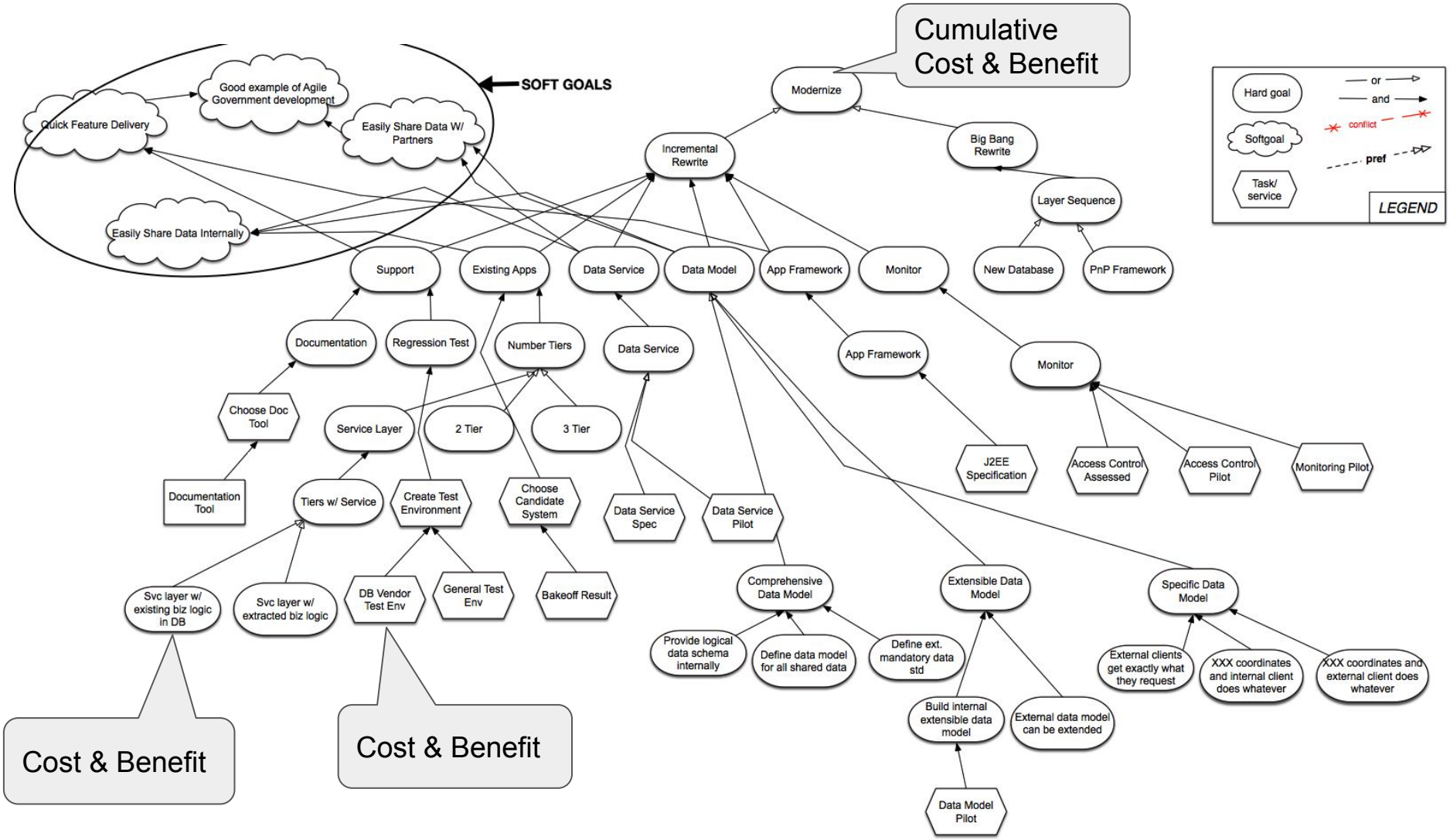


Services[horkoff15]

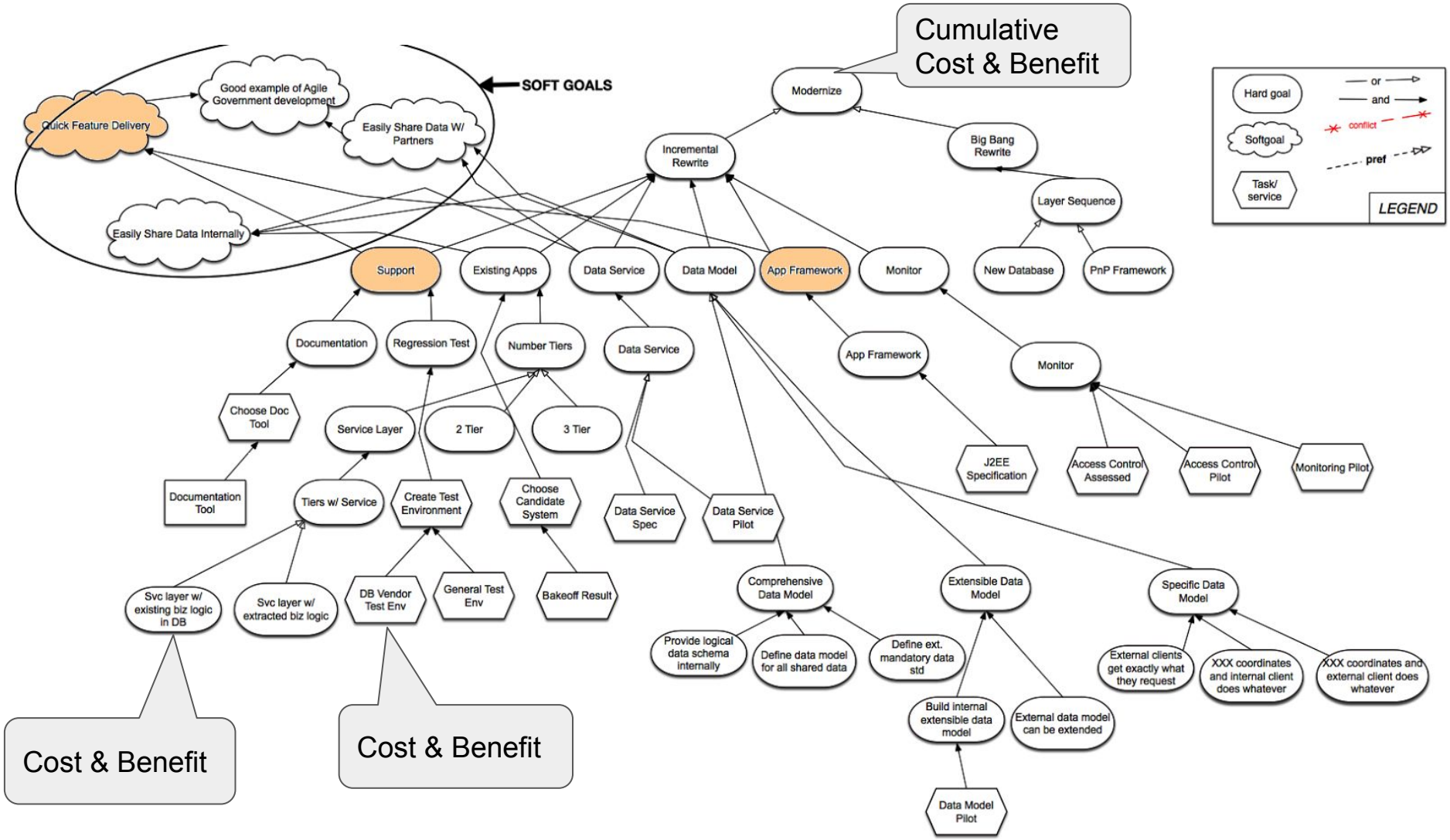
Data

Model	Nodes	Edges	# solutions
Services	351	510	> 10M
Marketing	326	422	> 10M
Counselling	350	470	> 10M
Management	206	239	> 4M
ITDepartment	126	162	> 2M
SAProgram	114	168	> 2M
Kids&Youth	81	81	~ 200K
Modernize	54	92	~ 1M

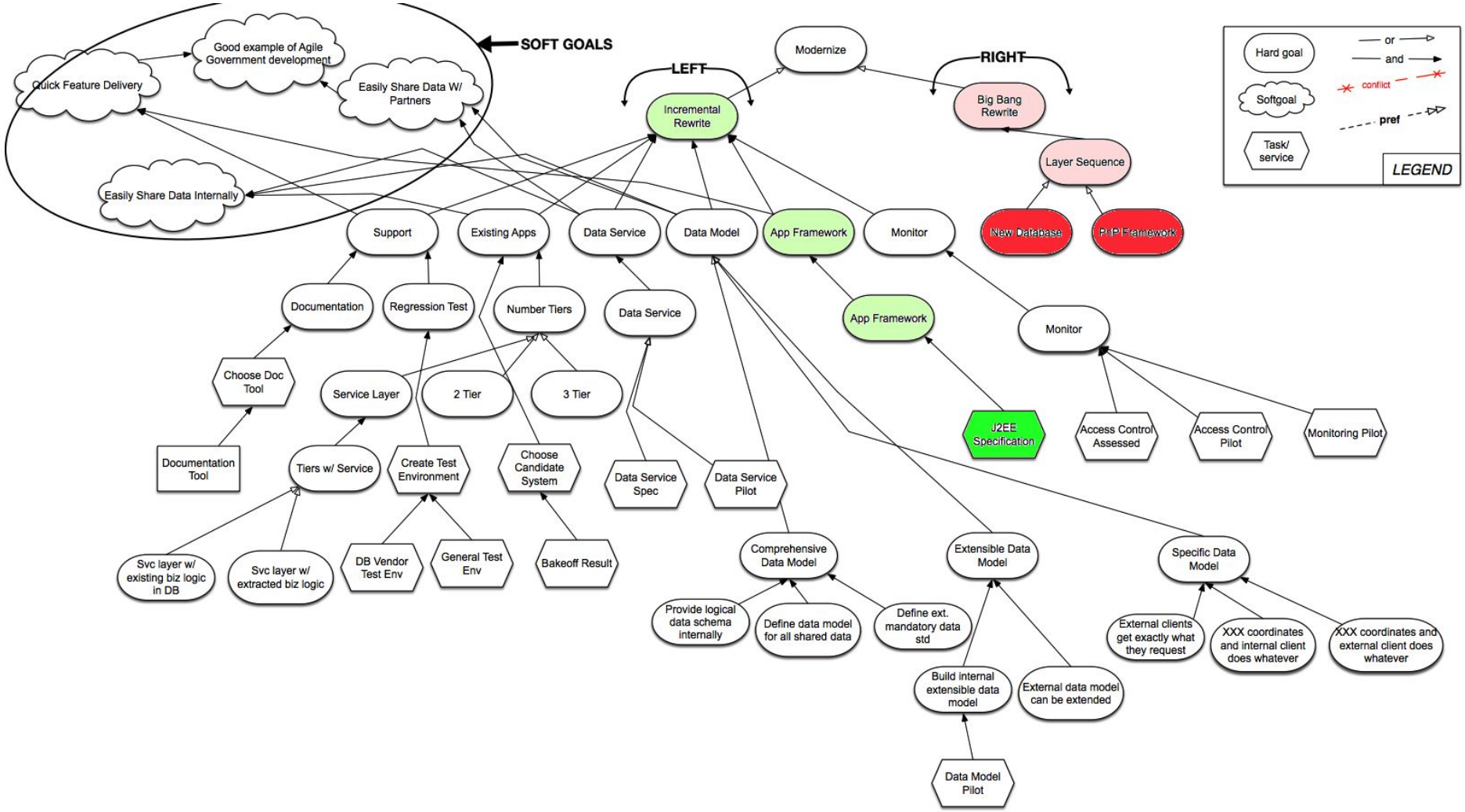
Simpler Example - Modernize



Simpler Example - Modernize

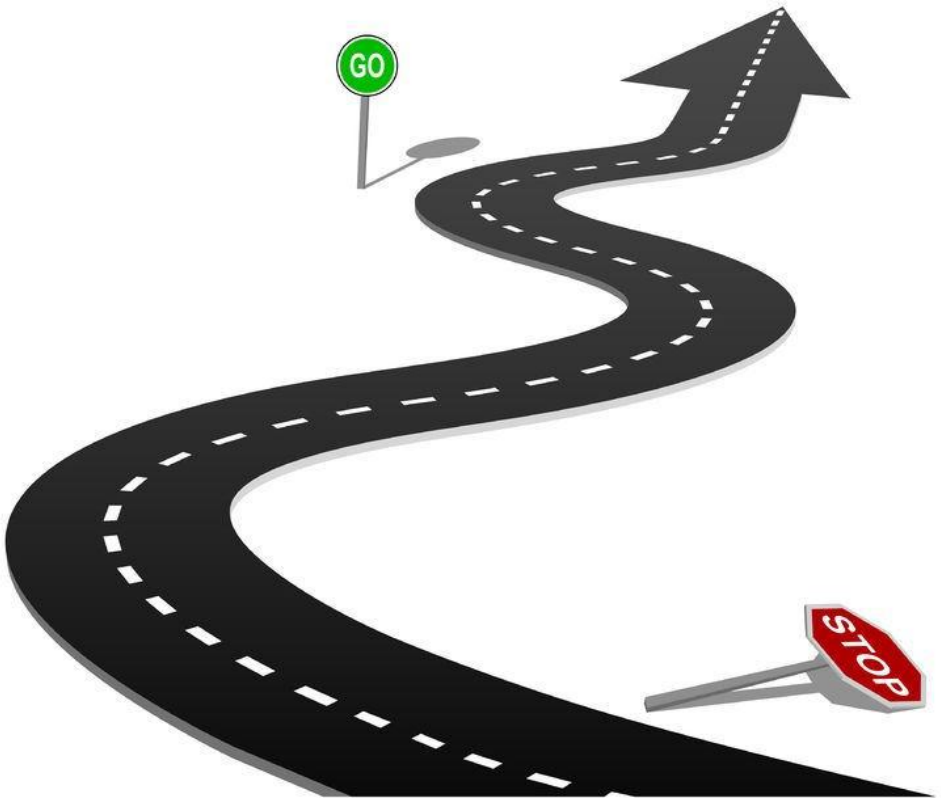


Simpler Example - Modernize

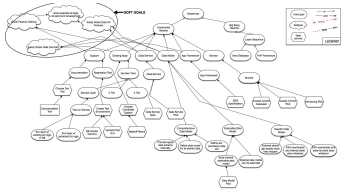


Roadmap

- **Background**
- **Related Work**
- **Data**
- **SHORT**
- Research Questions
- Conclusions
- Future Work



Framework



Sample Heuristically

Rest

Initial Population P_0

Optimizer (DE)

Pareto Front P_n

Best

Bayesian Ranker

Most

Rank

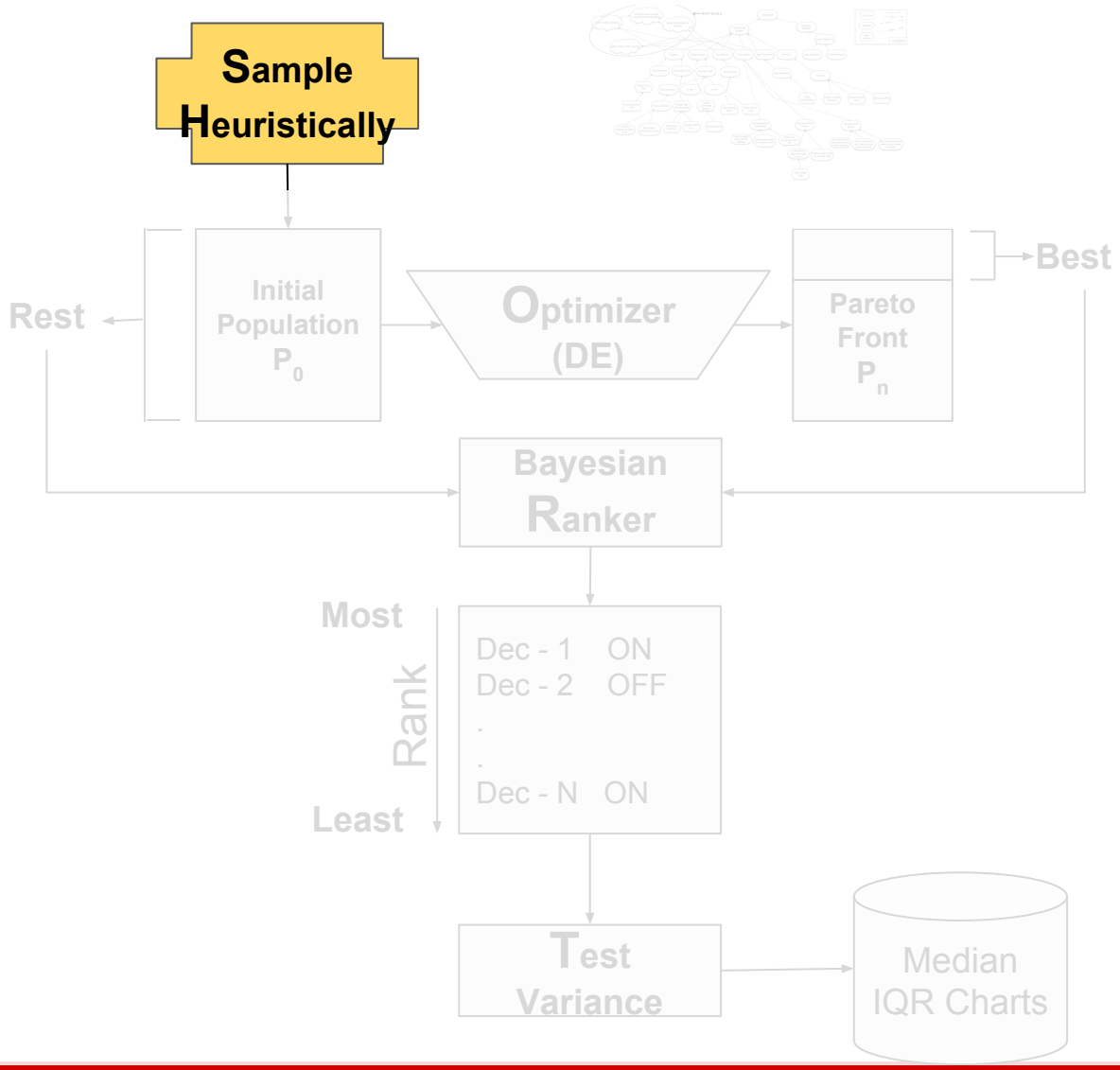
Least

Dec - 1 ON
Dec - 2 OFF
.
.
Dec - N ON

Test Variance

Median IQR Charts

Sample Heuristically



Sampling Techniques

- **Random Sampling**:
  - ○ Set decisions to arbitrary values
  - ○ Evaluate the goals(hard and soft)
  - \+ Fast Strawman approach
  - \- Fails to generate valid samples in conflicting spaces

Sampling Techniques

- Random Sampling
- **Forward Sampling**:
  - Set decisions to arbitrary value.
  - Traverse forward
  - Incase of conflicts
    - restart all over again
    - (or) resolve using an expert
  + Guarantees valid samples
  - Lack of adequate samples in constrained models
  - Biased towards experts

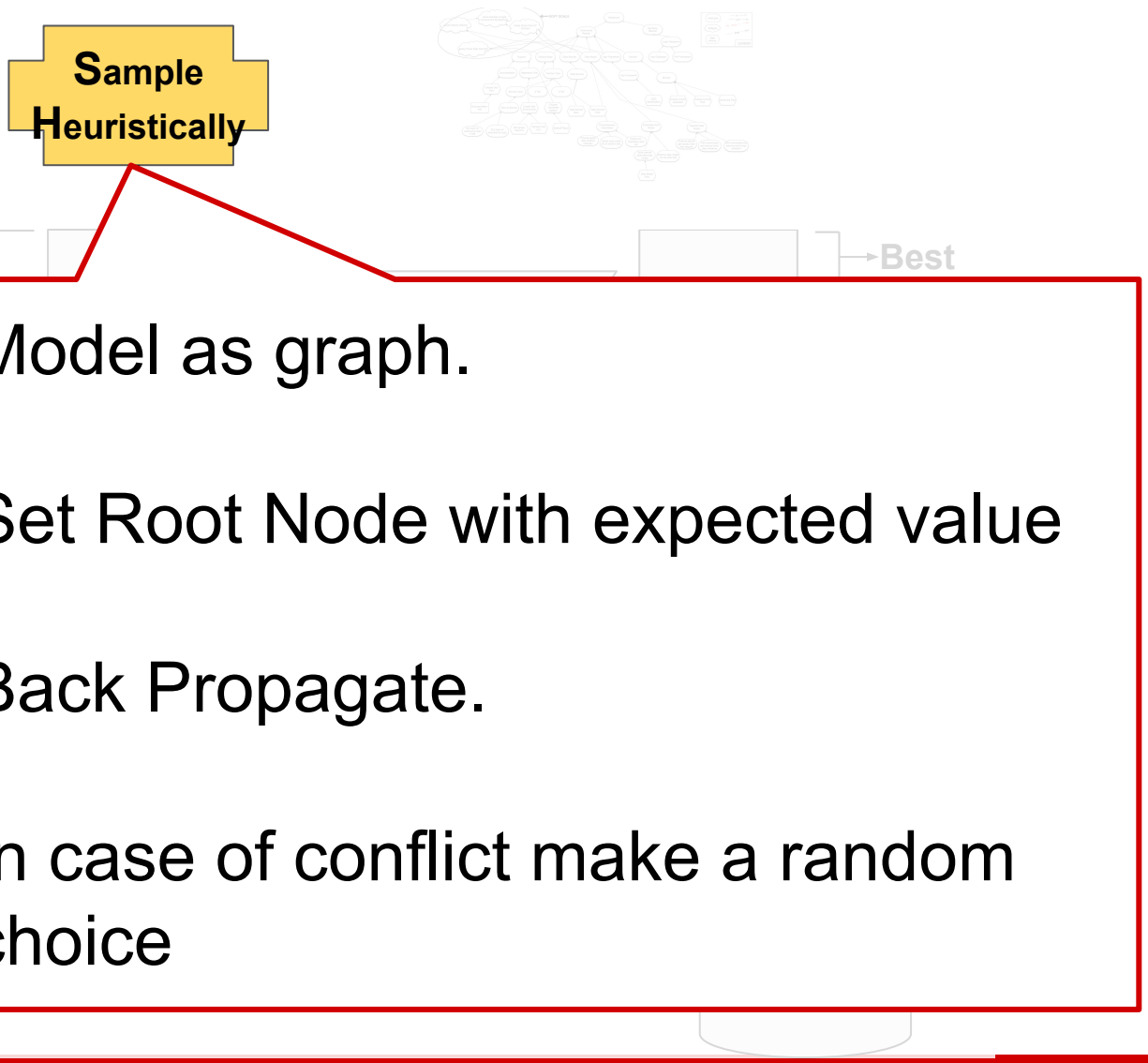
Sampling Techniques

- Random Sampling
- Forward Sampling
- **Backward Sampling**:
  - Set root nodes to desired value.
  - Traverse via BFS/DFS
  - Incase of conflicts
    - restart all over again
    - (or) resolve using an expert
  + Guarantees valid samples
  - Lack of adequate samples in constrained models
  - Biased towards experts

Sampling Techniques

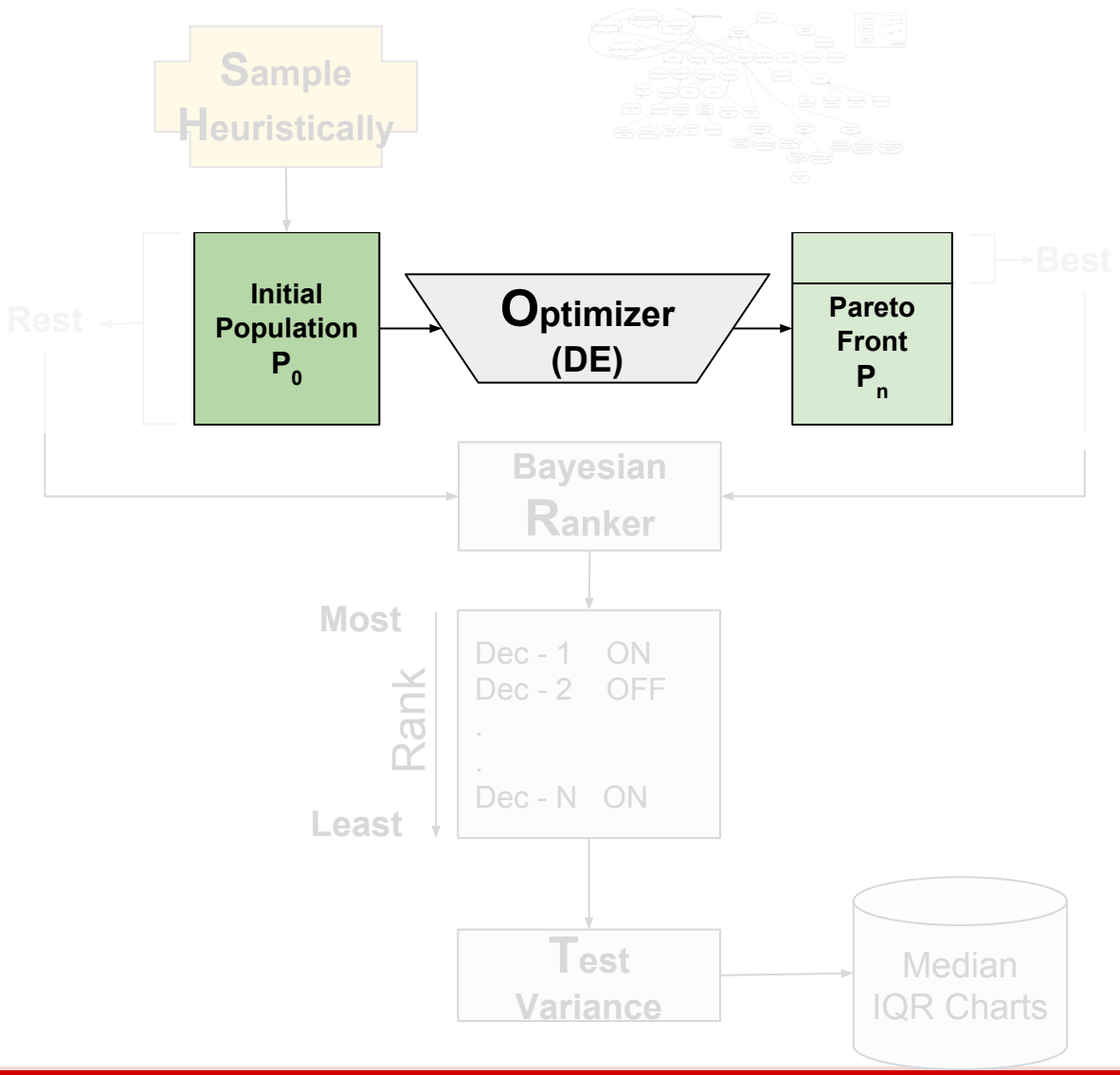
- Random Sampling
- Forward Sampling
- Backward Sampling
- **Heuristic Sampling**

Sample Heuristically

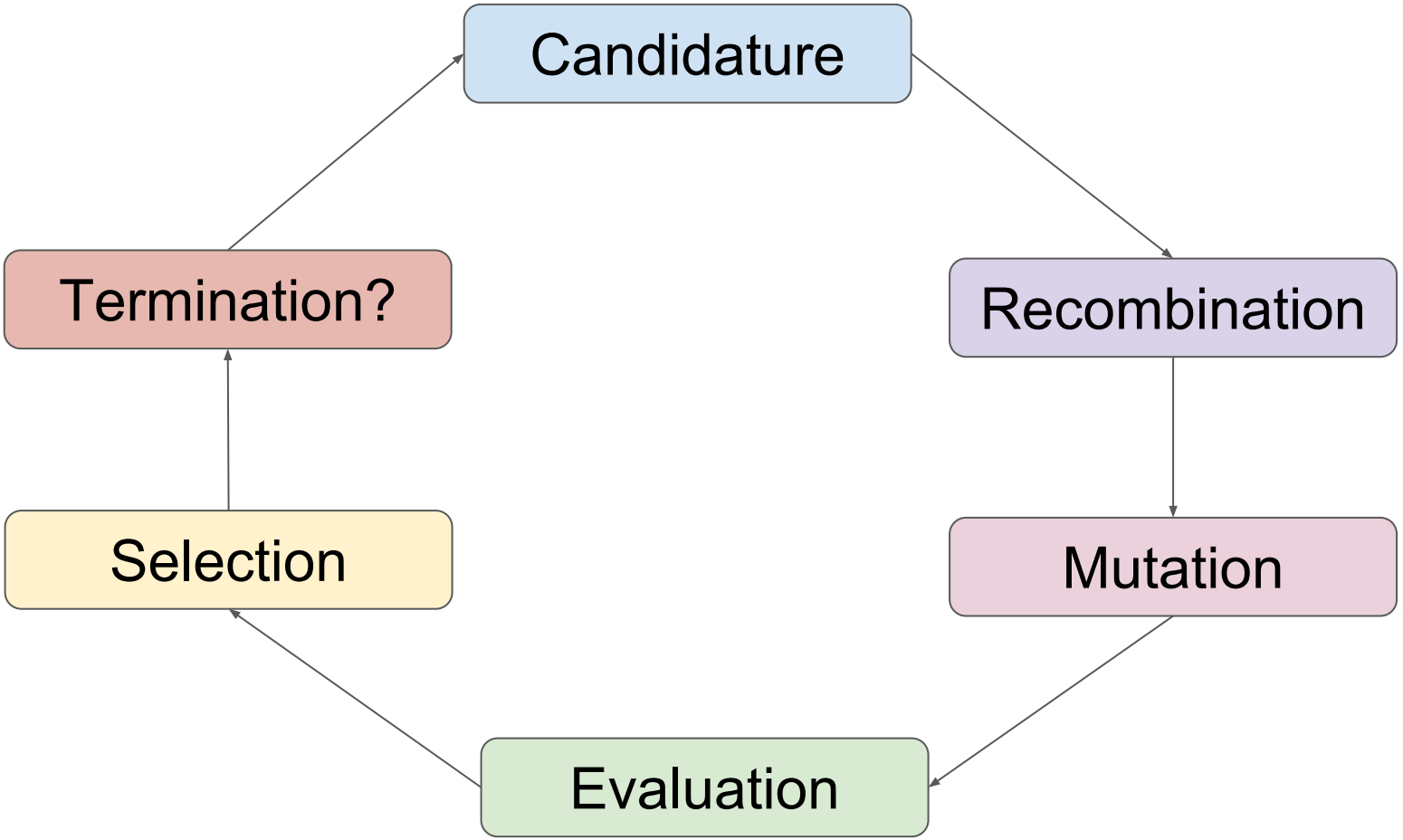


1. Model as graph.
2. Set Root Node with expected value
3. Back Propagate.
4. In case of conflict make a random choice

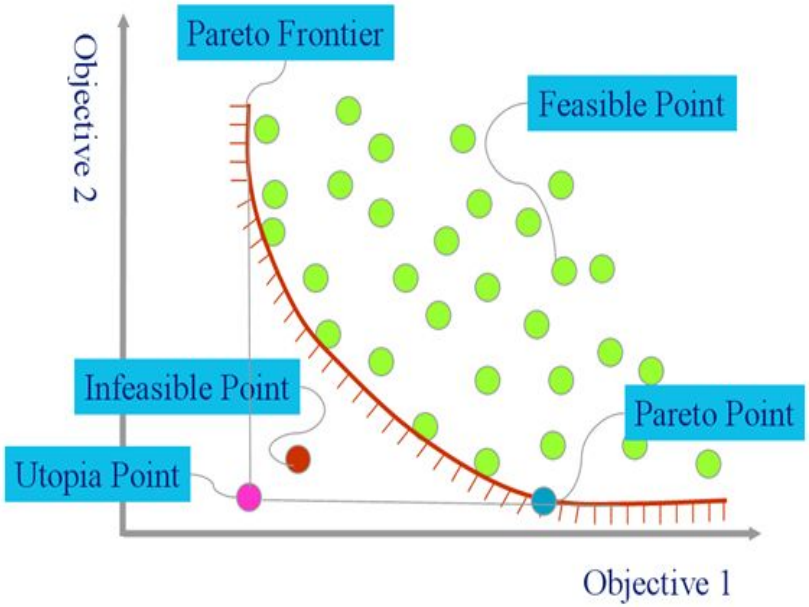
Optimization



Optimization



Optimization - Measures



- **Convergence**:
  - Distance between obtained and theoretical pareto front.
  - ***Smaller the better.***
- **Diversity**:
  - How well the actual pareto front is spread out across the theoretical front
  - ***Larger the better***

Optimization - Meta Heuristic Techniques

Algorithms	Factors			
	Family	Convergence	Diversity	Runtime
DE[Storn97]	Evolutionary			~O(n) 
NSGA-II[Deb02]	Evolutionary			$O(n^2)$ 
MOEA/D[Zhang07]	Evolutionary			$O(n^3)$ 

Optimization - Meta Heuristic Techniques

Algorithms	Factors			
	Family	Convergence	Diversity	Runtime
DE[Storn97]	Evolutionary	○	✓	~O(n) ✓
NSGA-II[Deb02]	Evolutionary	✓	○	$O(n^2)$ ✗
MOEA/D[Zhang07]	Evolutionary	✓	✓	$O(n^3)$ ✗
SWAY[Chen16]	Sampling	✓	✓	~O(k)* ✓
FLASH[Nair17]	Surrogate	✓	✓	~O(kl)* ✓

* k & l are size of sample and surrogate respectively

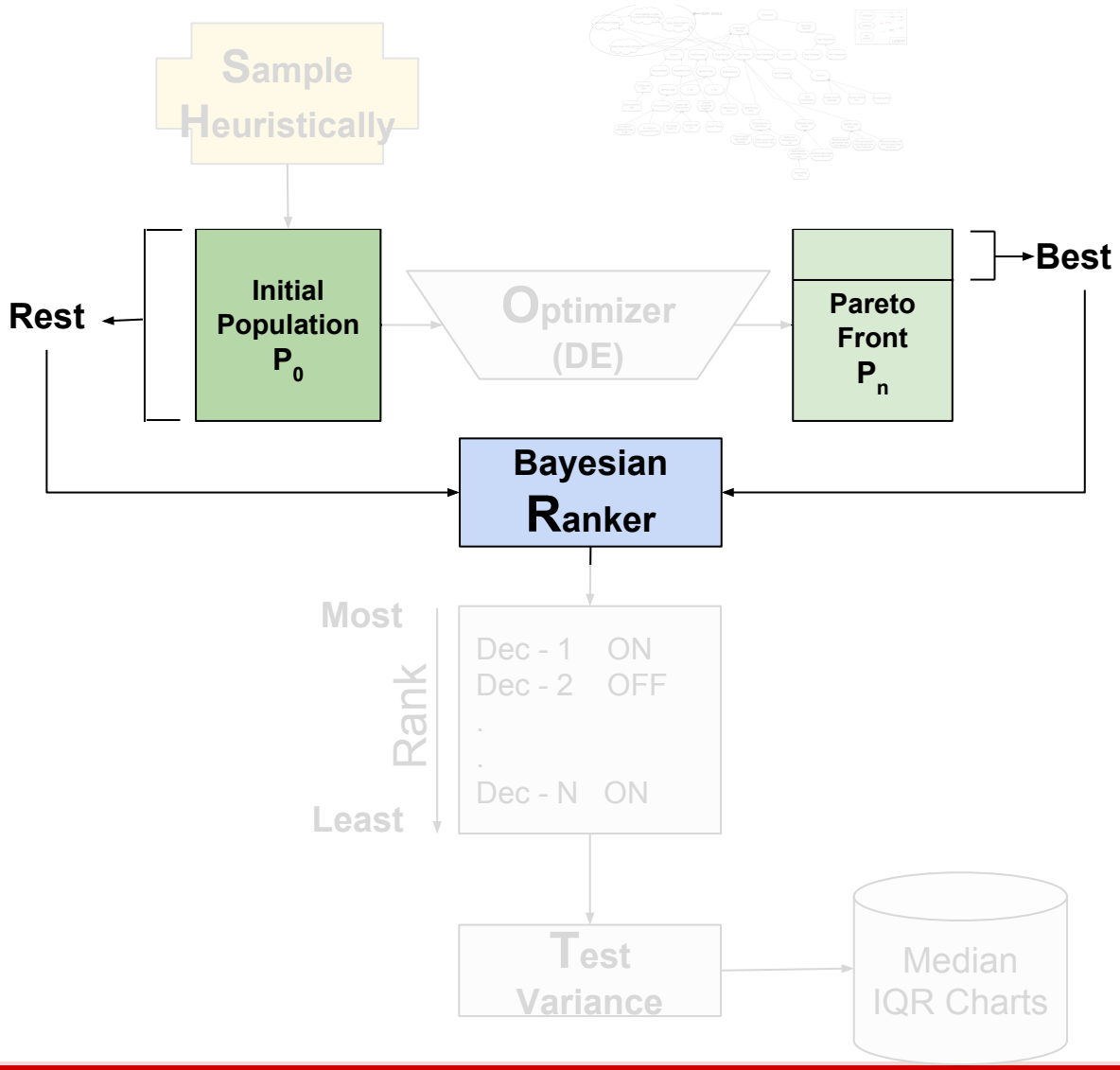
Optimization - Meta Heuristic Techniques

Algorithms	Factors			
	Family	Convergence	Diversity	Runtime
DE[Storn97]	Evolutionary	○	✓	~O(n) ✓
NSGA-II[Deb02]	Evolutionary	✓	○	$O(n^2)$ ✗
MOEA/D[Zhang07]	Evolutionary	✓	✓	$O(n^3)$ ✗
SWAY[Chen16]	Sampl	✓	✓	~O(k)* ✓
FLASH[Nair17]	Surrogate	✓	✓	~O(kl)* ✓

FUTURE WORK

* k & l are size of sample and surrogate respectively

Bayesian Ranking



Bayesian **R**anking[Menzies07]

$$P(Best) = \frac{Best}{Best + Rest}$$

$$P(Rest) = \frac{Rest}{Best + Rest}$$

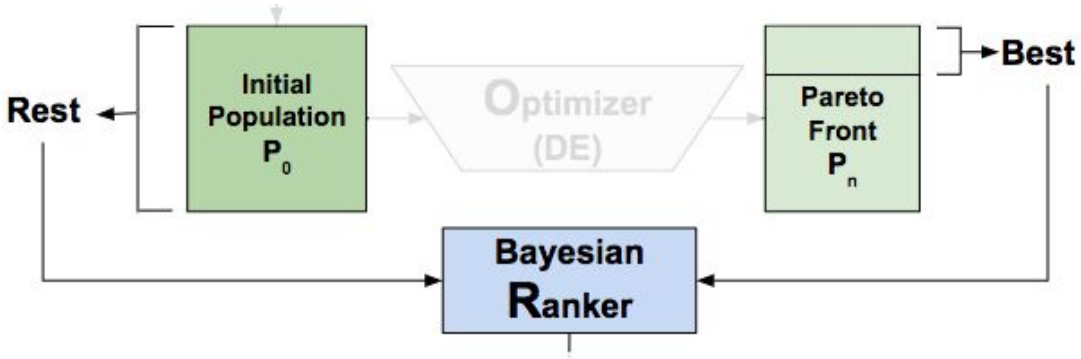
$$P(E|Best) = \frac{Best_E}{Best}$$

$$P(E|Rest) = \frac{Rest_E}{Rest}$$

$$like(E|Best) = P(E|Best) * P(Best)$$

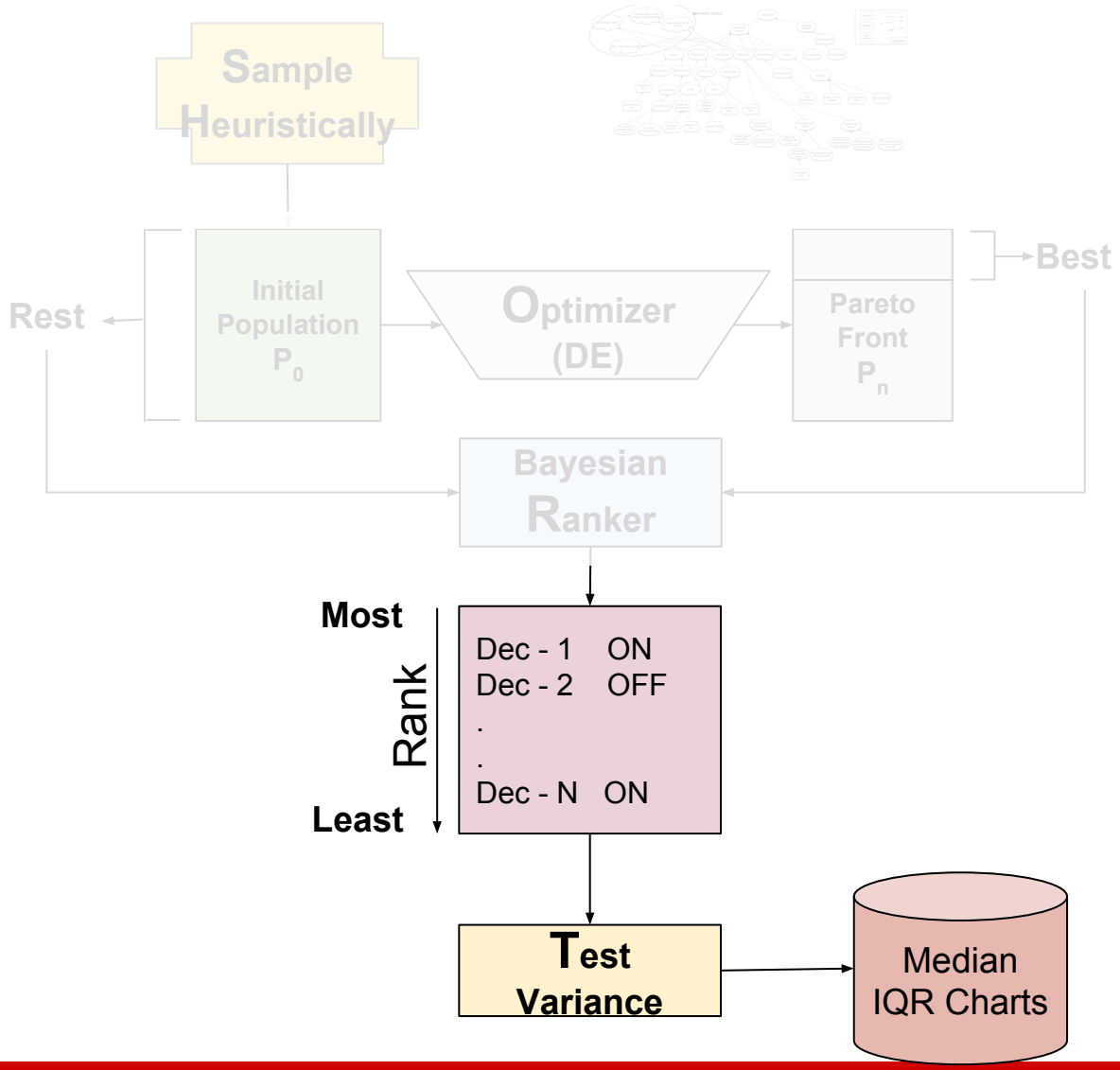
$$like(E|Rest) = P(E|Rest) * P(Rest)$$

$$support(E) = \frac{like(E|Best)}{like(E|Best) + like(E|Rest)}$$



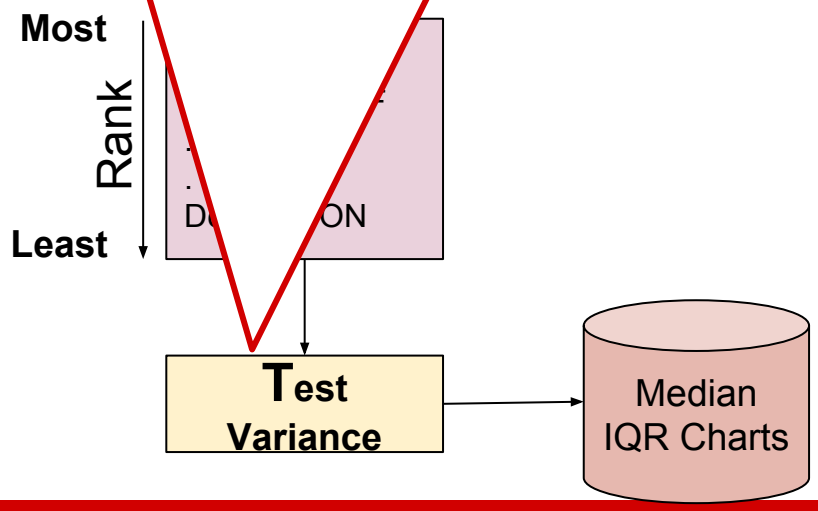
- Rank decisions in decreasing order of support

Variance Testing



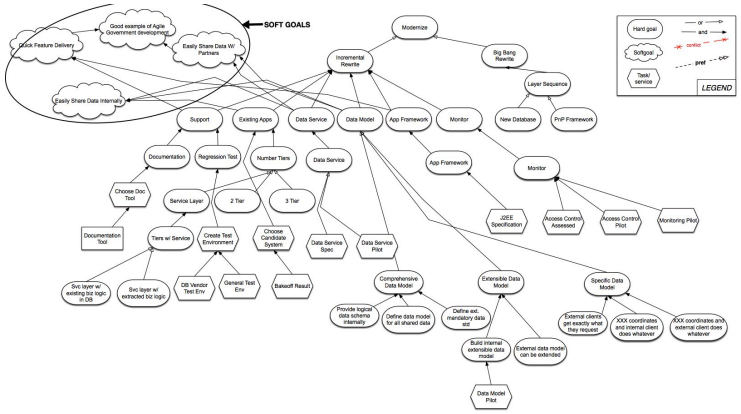
Variance Testing

1. Random Population(**P**)
2. Iterate **x** b/w **1** … **N**
   a. Decisions 1 … x to opt. value
   b. Decisions **x**+1 … **N** to random
   c. Plot median and IQR of each goal in the population
3. Examine the point of greatest change.



Example Workflow - Modernize

Ranking Decisions

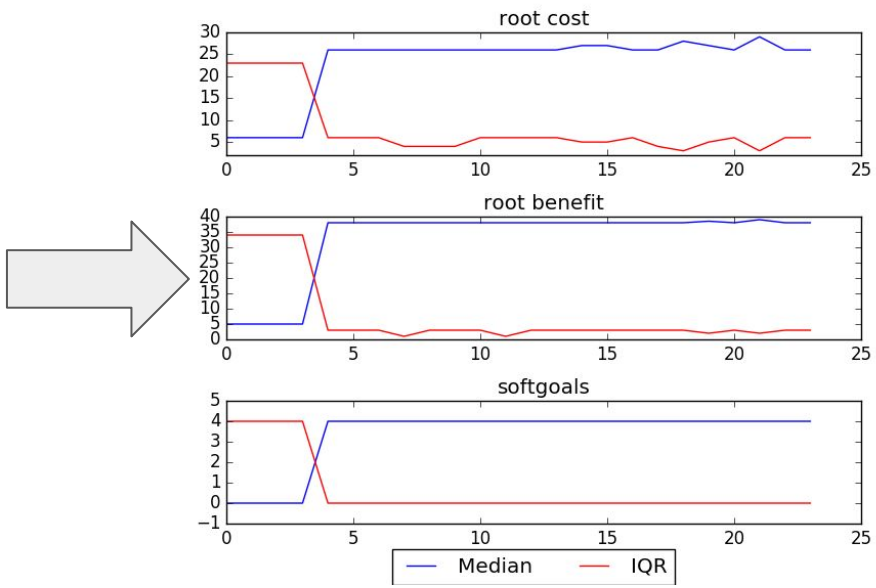


Rank	Node	Status	Support
1	J2EE Specification	ON	0.129
2	Pnp Framework	OFF	0.124
3	New Database	OFF	0.115
4	Documentation Tool	ON	0.114
5	Access Control Assessed	ON	0.113
6	Monitoring Pilot	ON	0.112
7	General Test Env	ON	0.110
8	Bakeoff Result	ON	0.110
9	Access Control Pilot	ON	0.108
10	DB Vendor Test Env	ON	0.105
11	Data Service Spec	ON	0.099
12	External Clients get their request	ON	0.098
13	XXX coordinates & internal client	ON	0.098
14	XXX coordinates & external client	ON	0.097
15	Data Model Pilot	ON	0.095
16	Data Service Pilot	ON	0.095
17	2 Tier	ON	0.094
18	3 Tier	ON	0.090
19	Define data model for shared data	ON	0.085
20	Svc layer w/ extracted biz logic	OFF	0.080
21	Define ext mandatory data std	ON	0.079
22	Svc layer w/ extracted biz logic in DB	ON	0.066
23	External data model can be extended	ON	0.062
24	Provide logical data scheme internally	ON	0.062

Example Workflow - Modernize

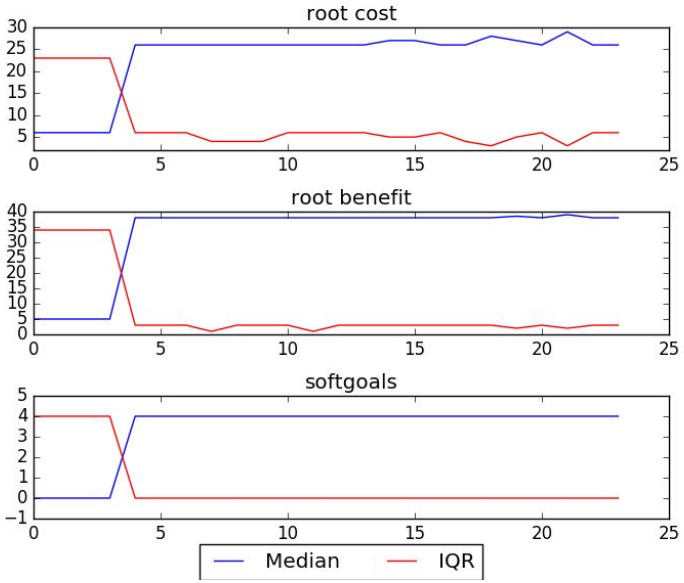
Variance Testing

Rank	Node	Status	Support
1	J2EE Specification	ON	0.129
2	Pnp Framework	OFF	0.124
3	New Database	OFF	0.115
4	Documentation Tool	ON	0.114
5	Access Control Assessed	ON	0.113
6	Monitoring Pilot	ON	0.112
7	General Test Env	ON	0.110
8	Bakeoff Result	ON	0.110
9	Access Control Pilot	ON	0.108
10	DB Vendor Test Env	ON	0.105
11	Data Service Spec	ON	0.099
12	External Clients get their request	ON	0.098
13	XXX coordinates & internal client	ON	0.098
14	XXX coordinates & external client	ON	0.097
15	Data Model Pilot	ON	0.095
16	Data Service Pilot	ON	0.095
17	2 Tier	ON	0.094
18	3 Tier	ON	0.090
19	Define data model for shared data	ON	0.085
20	Svc layer w/ extracted biz logic	OFF	0.080
21	Define ext mandatory data std	ON	0.079
22	Svc layer w/ extracted biz logic in DB	ON	0.066
23	External data model can be extended	ON	0.062
24	Provide logical data scheme internally	ON	0.062



Example Workflow - Modernize

Identifying Keys

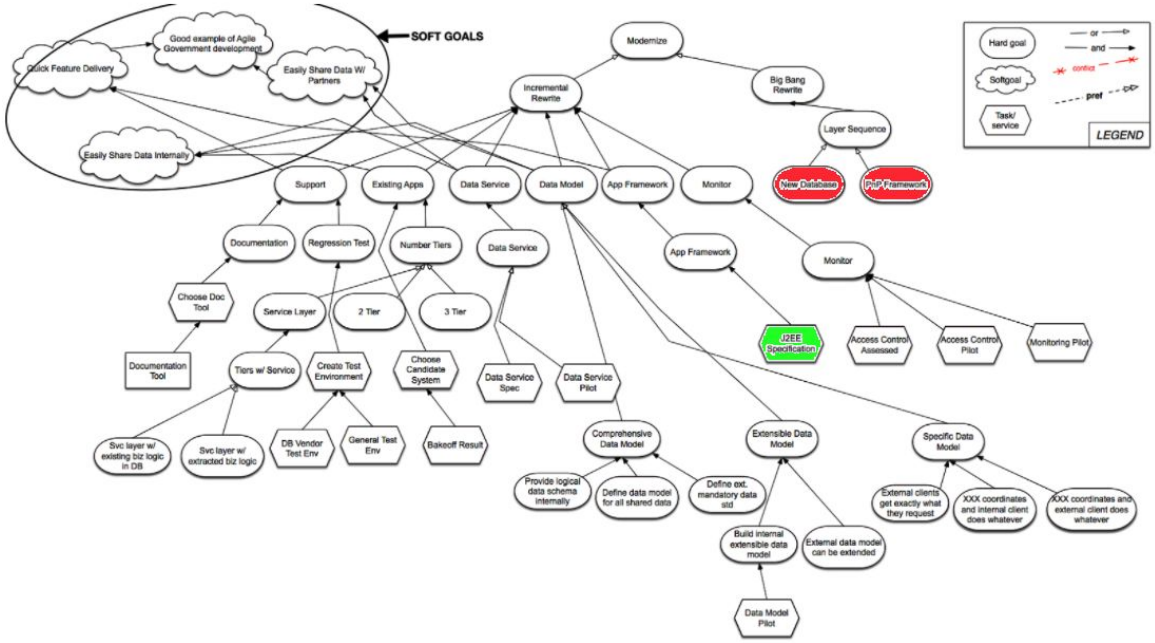


Rank	Node	Status	Support
1	J2EE Specification	ON	0.129
2	Pnp Framework	OFF	0.124
3	New Database	OFF	0.115
4	Documentation Tool	ON	0.114
5	Access Control Assessed	ON	0.113
6	Monitoring Pilot	ON	0.112
7	General Test Env	ON	0.110
8	Bakeoff Result	ON	0.110
9	Access Control Pilot	ON	0.108
10	DB Vendor Test Env	ON	0.105
11	Data Service Spec	ON	0.099
12	External clients get their request	ON	0.098
13	XXX coordinates & internal client	ON	0.098
14	XXX coordinates & external client	ON	0.097
15	Data Model Pilot	ON	0.095
16	Data Service Pilot	ON	0.095
17	2 Tier	ON	0.094
18	3 Tier	ON	0.090
19	Define data model for shared data	ON	0.085
20	Svc layer w/ extracted biz logic	OFF	0.080
21	Define ext mandatory data std	ON	0.079
22	Svc layer w/ extracted biz logic in DB	ON	0.066
23	External data model can be extended	ON	0.062
24	Provide logical data scheme internally	ON	0.052

Example Workflow - Modernize

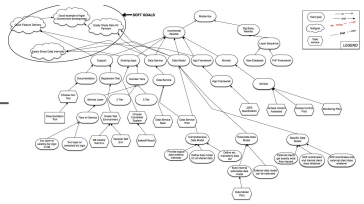
Identifying Keys

Rank	Node	Status	Support
1	J2EE Specification	ON	0.129
2	Pnp Framework	OFF	0.124
3	New Database	OFF	0.115
4	Documentation Tool	ON	0.114
5	Access Control Assessed	ON	0.113
6	Monitoring Pilot	ON	0.112
7	General Test Env	ON	0.110
8	Bakeoff Result	ON	0.110
9	Access Control Pilot	ON	0.108
10	DB Vendor Test Env	ON	0.105
11	Data Service Spec	ON	0.099
12	External clients get their request	ON	0.098
13	XXX coordinates & internal client	ON	0.098
14	XXX coordinates & external client	ON	0.097
15	Data Model Pilot	ON	0.095
16	Data Service Pilot	ON	0.095
17	2 Tier	ON	0.094
18	3 Tier	ON	0.090
19	Define data model for shared data	ON	0.085
20	Svc layer w/ extracted biz logic	OFF	0.080
21	Define ext mandatory data std	ON	0.079
22	Svc layer w/ extracted biz logic in DB	ON	0.066
23	External data model can be extended	ON	0.062
24	Provide logical data scheme internally	ON	0.052



SHORT - Retrospect

1. Model as graph.
2. Set Root Node with expected value
3. Back Propagate.
4. In case of conflict make a random choice



Sample Heuristically

Rest

Initial Population P_0

Optimizer (DE)

Pareto Front P_n

Best

Bayesian Ranker

Rank a decision 'E' using the below measure

$$P(best|E) * support(best|E) = \frac{like(best|x)^2}{like(best|x) + like(rest|x)}$$

$$P(best|E) = \frac{like(best|E)}{like(best|E) + like(rest|E)}$$

$$like(best|E) = freq(E|best) \cdot P(best)$$
$$like(rest|E) = freq(E|rest) \cdot P(rest)$$

Dec - 1 ON
Dec - 2 OFF
.
.
Dec - N ON

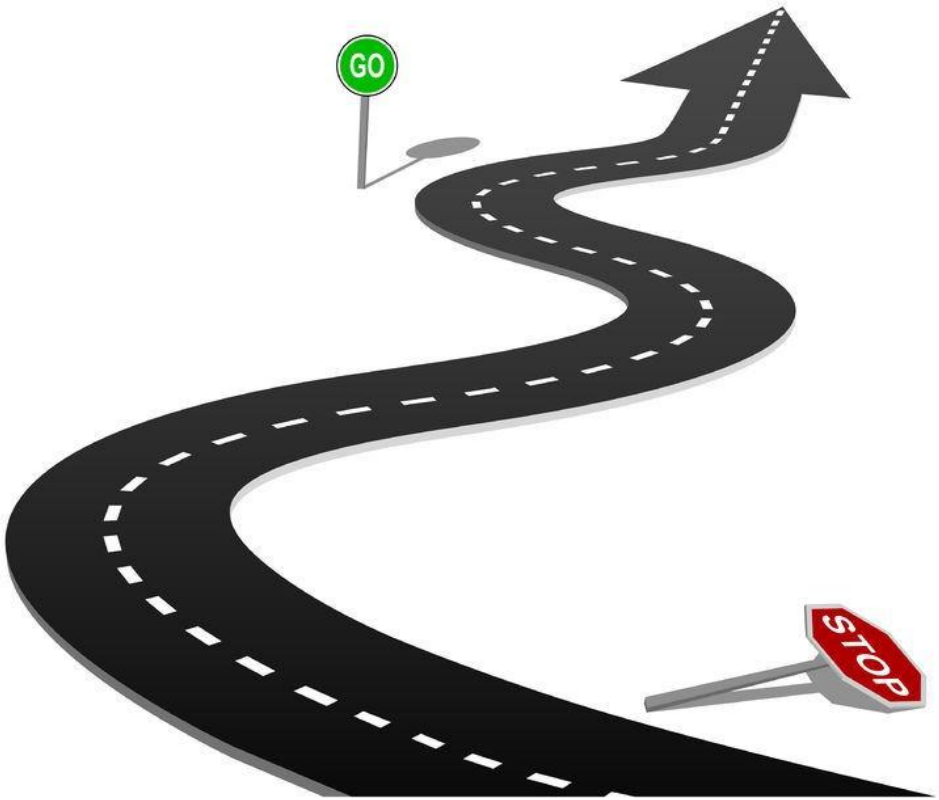
1. Random Population(**P**)
2. Iterate **x** b/w **1** … **N**
   a. Decisions 1 … x to opt. value
   b. Decisions x+1 … N to random
   c. Plot median and IQR of each goal in the population
3. Examine the point of greatest change.

Test Variance

Median IQR Charts

Roadmap

- **Background**
- **Related Work**
- **Data**
- **SHORT**
- **Research Questions**
- Conclusions
- Future Work



RQ1: Can heuristic sampling be used in RE models?

Model	Runtime in seconds per 100 samples			
	Random	**Forward**	**Backward**	**Hueristic**
Services	0.21	121	106	0.9
Marketing	0.2	101	101	0.8
Counselling	0.22	104	112	0.9
Management	0.18	52	48	0.5
ITDepartment	0.15	64	65	0.6
SAProgram	0.14	71	67	0.5
Kids&Youth	0.12	42	38	0.3
Modernize	0.1	40	40	0.25

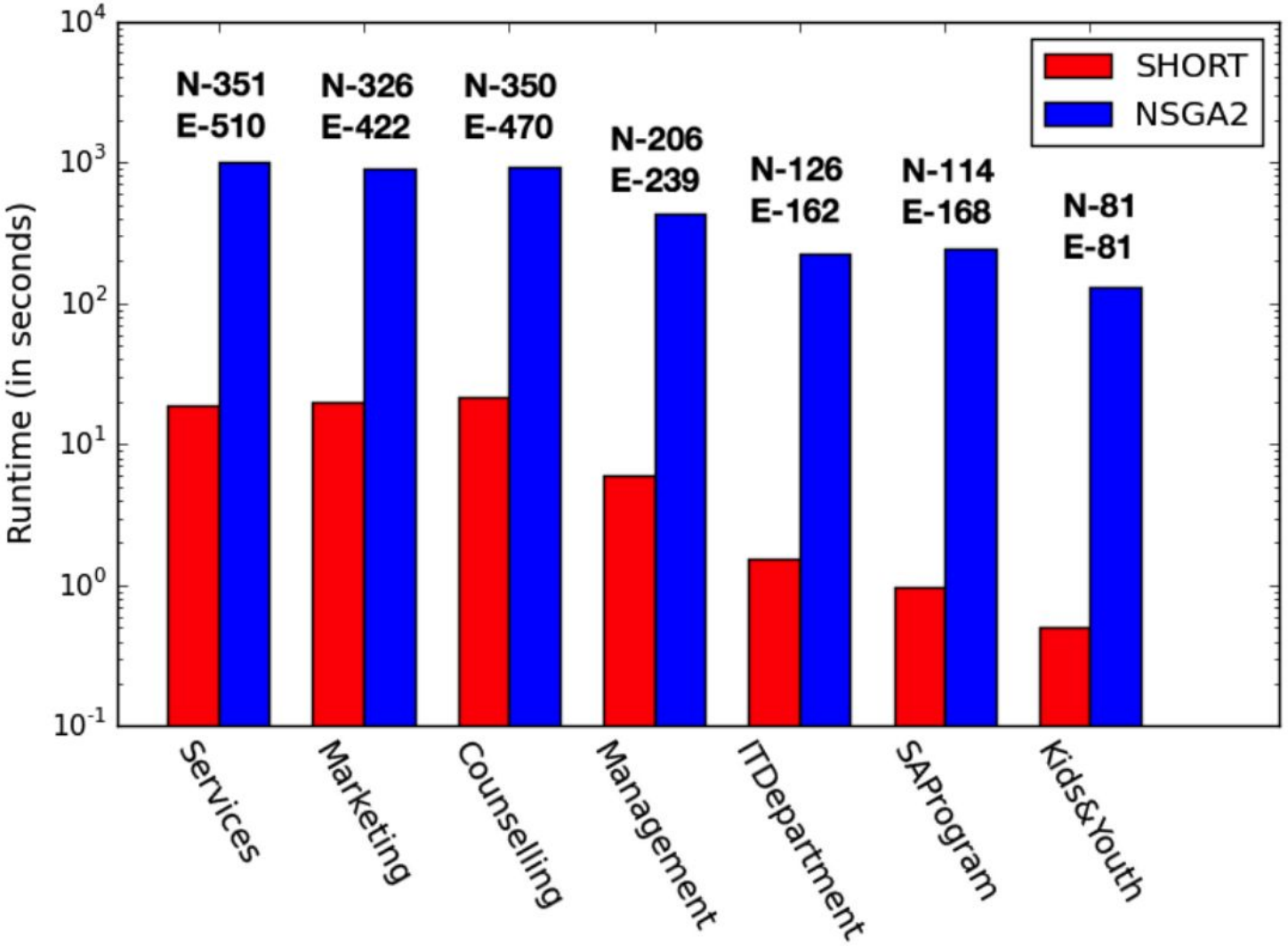
RQ1: Can heuristic sampling be used in RE models?

Model	# of valid samples in 100 samples			
	Random	**Forward**	**Backward**	**Hueristic**
Services	3	100	100	84
Marketing	7	100	100	89
Counselling	1	100	100	65
Management	21	100	100	92
ITDepartment	8	100	100	79
SAProgram	14	100	100	75
Kids&Youth	80	100	100	100
Modernize	30	100	100	100

RQ1: Can heuristic sampling be used in RE models?

- As fast as random sampling.
- > 75% valid solutions in ⅞ models

RQ2: Can DE optimize RE models?

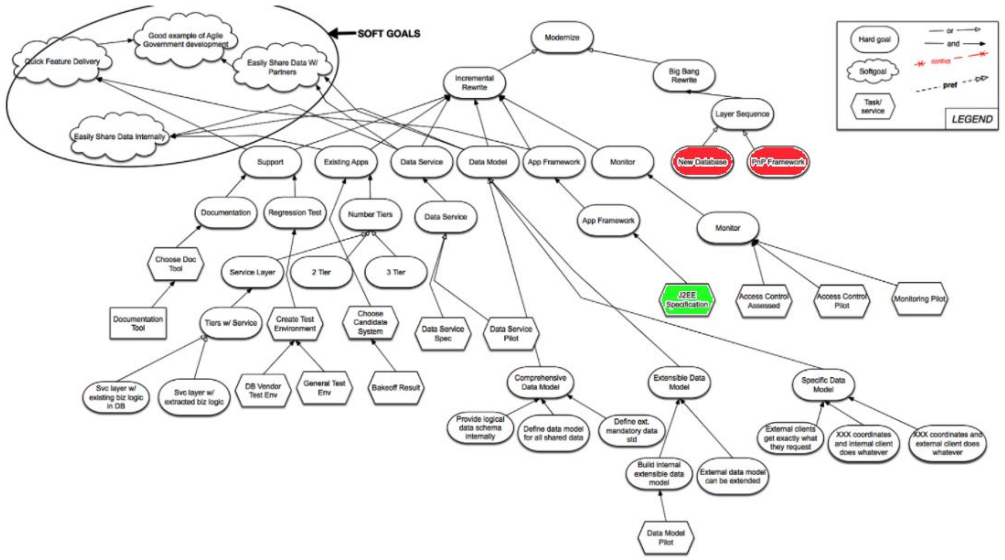


RQ2: Can DE optimize RE models?

Model	Convergence				Diversity			
	softgoals %		hard goals %		softgoals %		hard goals %	
	NSGA-II	DE	NSGA-II	DE	NSGA-II	DE	NSGA-II	DE
Services	46.77	46.24	65.22	65.22	0	0	0	0
Marketing	32.47	32.47	34.38	34.38	0	0	0	0
Counselling	54.29	53.29	44.83	44.83	1.41	2.76	3.45	3.45
Management	44.63	42.98	61.11	61.11	0.82	1.65	2.78	2.78
ITDepartment	68.42	68.42	69.57	73.91	2.63	2.63	0	8.69
SAProgram	78.69	78.69	66.67	66.67	1.64	1.64	0	0
Kids&Youth	22.86	22.86	100	100	0	0	0	0

RQ3: Do “Keys” exist in RE models

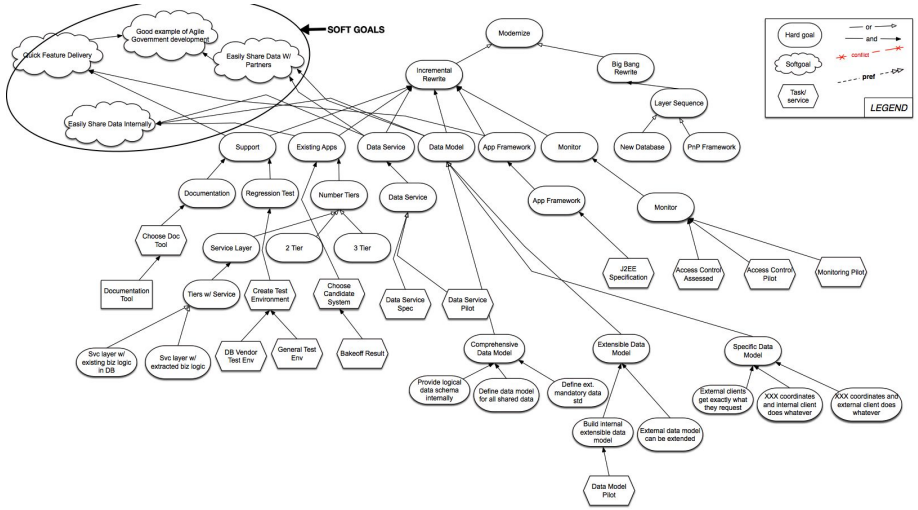
For Modernize



Rank	Node	Status	Support
1	J2EE Specification	ON	0.129
2	Pnp Framework	OFF	0.124
3	New Database	OFF	0.115
4	Documentation Tool	ON	0.114
5	Access Control Assessed	ON	0.113
6	Monitoring Pilot	ON	0.112
7	General Test Env	ON	0.110
8	Bakeoff Result	ON	0.110
9	Access Control Pilot	ON	0.108
10	DB Vendor Test Env	ON	0.105
11	Data Service Spec	ON	0.099
12	External clients get their request	ON	0.098
13	XXX coordinates & internal client	ON	0.098
14	XXX coordinates & external client	ON	0.097
15	Data Model Pilot	ON	0.095
16	Data Service Pilot	ON	0.095
17	2 Tier	ON	0.094
18	3 Tier	ON	0.090
19	Define data model for shared data	ON	0.085
20	Svc layer w/ extracted biz logic	OFF	0.080
21	Define ext mandatory data std	ON	0.079
22	Svc layer w/ extracted biz logic in DB	ON	0.066
23	External data model can be extended	ON	0.062
24	Provide logical data scheme internally	ON	0.052

RQ3: Do “Keys” exist in RE models

What about larger Models?

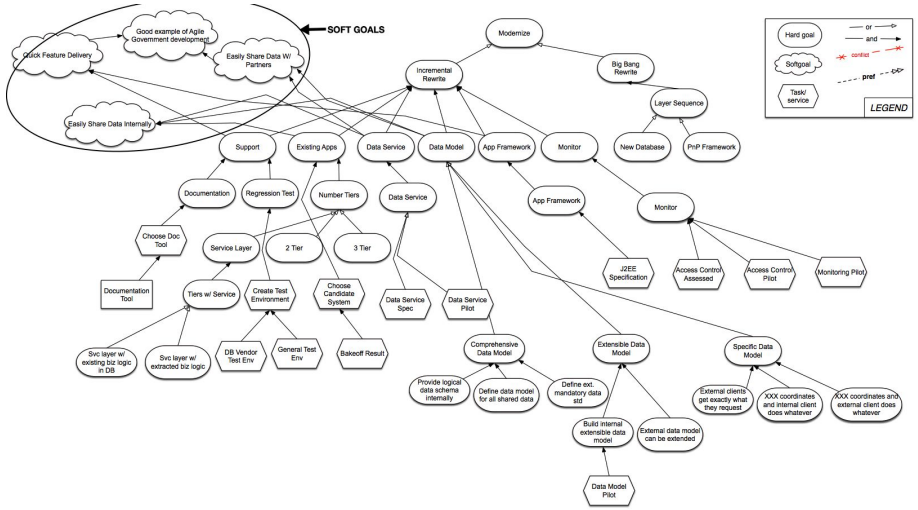


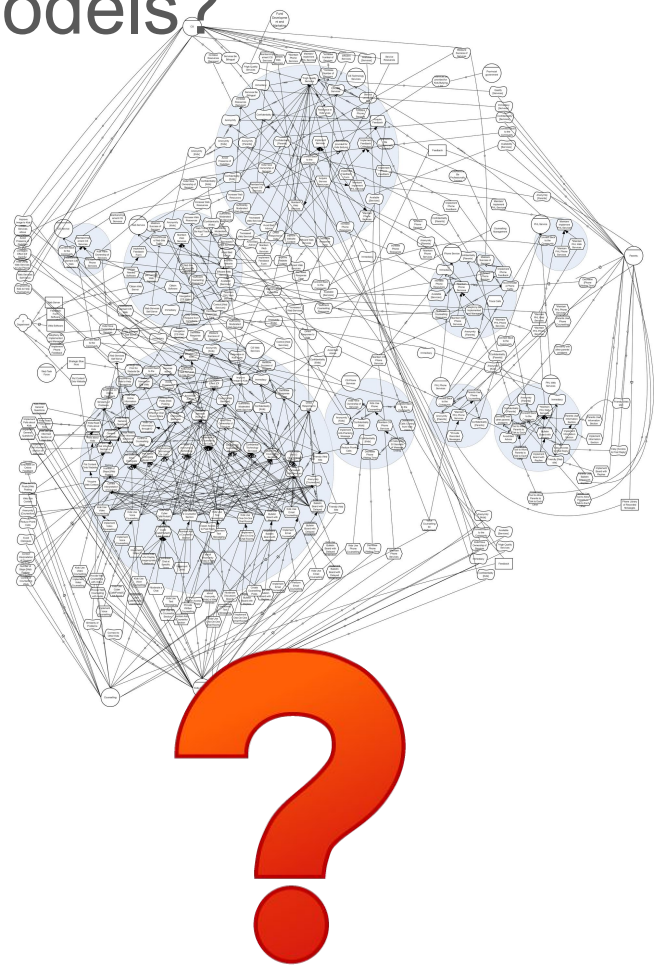
Objectives:

- Min Root Cost
- Max Root Benefit
- Max Softgoals satisfied

RQ3: Do “Keys” exist in RE models

What about larger Models?



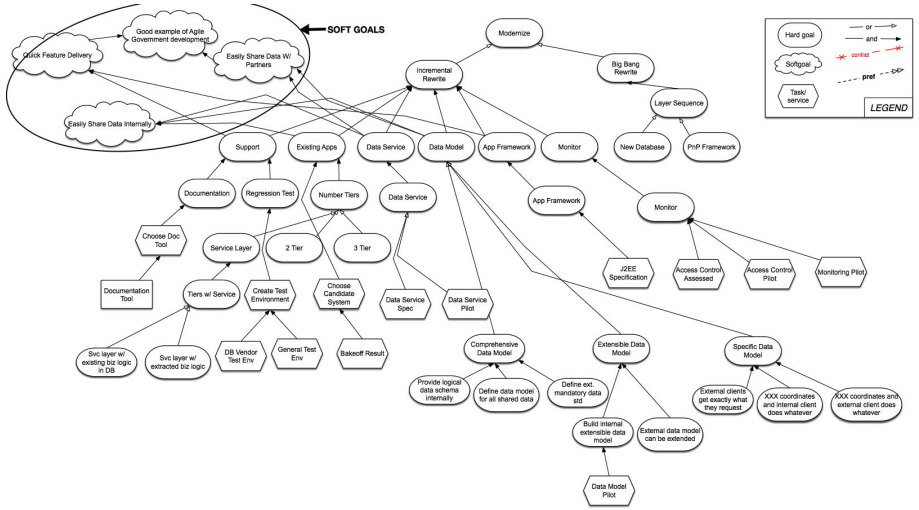


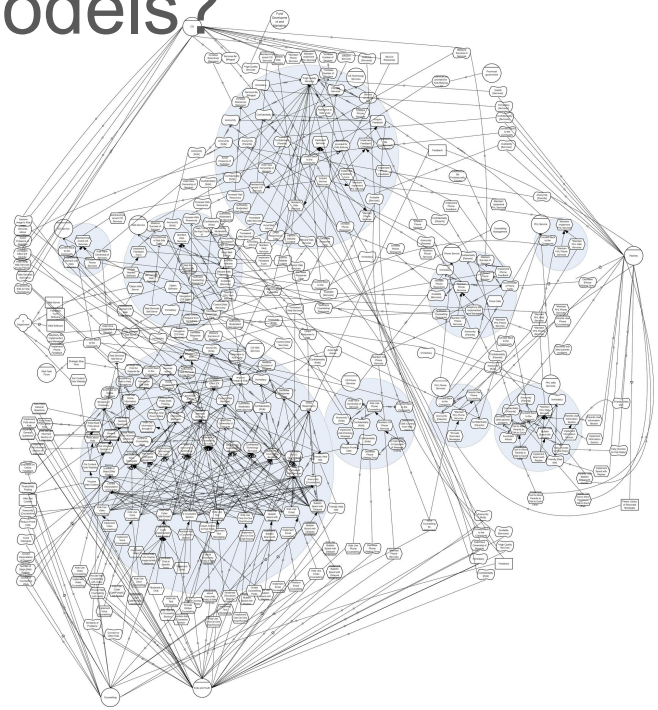
Objectives:

- Min Root Cost
- Max Root Benefit
- Max Softgoals satisfied

RQ3: Do “Keys” exist in RE models

What about larger Models?





Objectives:

- Min Root Cost
- Max Root Benefit
- Max Softgoals satisfied

Objectives:

- Max Hard Goals Satisfied
- Max SoftGoals Satisfied
- Min Decisions Set

RQ3: Do “Keys” exist in RE models

Model	Result	Percentage of Decisions				
		6	12	25	50	100
ITDepartment	SGs	70	60	70	70	60
	Goals	50	50	80	80	70
	Costs	2	4	9	17	18
Services	SGs	50	50	50	50	50
	Goals	50	70	70	60	50
	Costs	4	7	13	19	24

RQ3: Do "Keys" exist in RE models

Model	Result	Percentage of Decisions				
		6	12	25	50	100
ITDepartment	SGs	70	60	70	70	60
	Goals	50	50	80	80	70
	Costs	2	4	9	17	18
Services	SGs	50	50	50	50	50
	Goals	50	70	70	60	50
	Costs	4	7	13	19	24

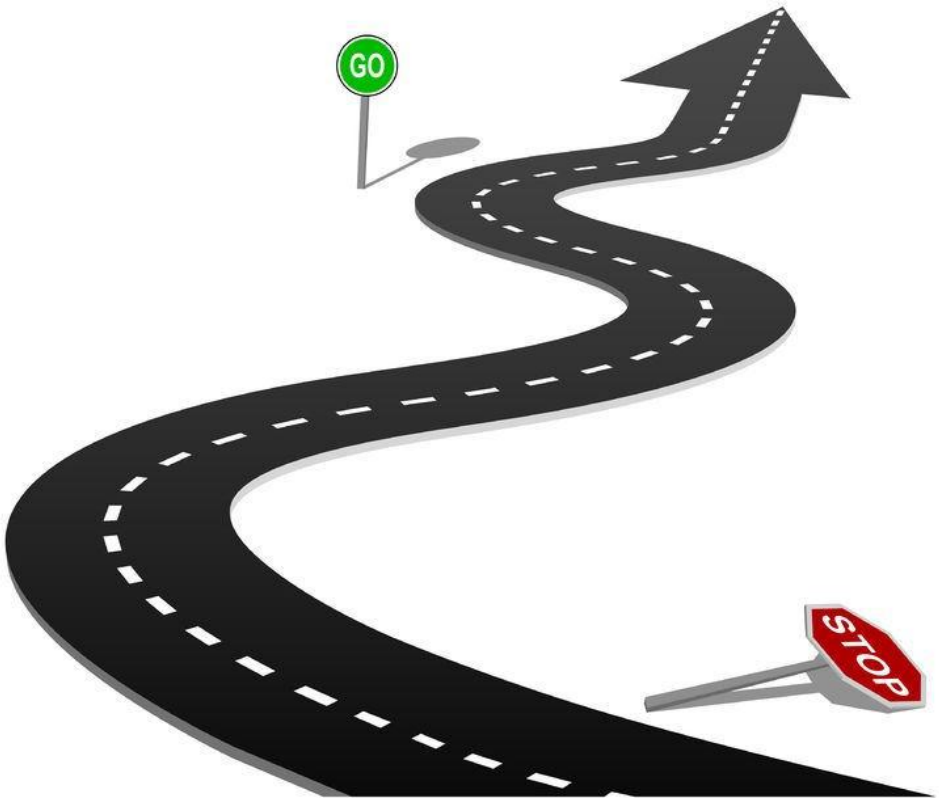
RQ3: Do "Keys" exist in RE models

Model	Result	Percentage of Decisions				
		6	12	25	50	100
ITDepartment	SGs	70	60	70	70	60
	Goals	50	50	80	80	70
	Costs	2	4	9	17	18
Services	SGs	50	50	50	50	50
	Goals	50	70	70	60	50
	Costs	4	7	13	19	24

Model	Counselling	Management	Marketing	SAProgram	Kids&Youth
Keys@	12%	25%	12%	12%	12%

Roadmap

- Background
- Related Work
- Data
- SHORT
- Research Questions
- **Conclusions**
- Future Work



Conclusion

- Heuristic sampling can be used to sample solutions from RE models.
  Open Problems:
  - Can we get 100% valid samples?
  - How does it scale to much larger models?

Conclusion

- Heuristic sampling can be used to sample solutions from RE models.

- Fast and diverse optimization of RE models can be performed using DE.
  Open Problems:
  - Can we get better quality solutions?
  - Is evolutionary the only way forward?

Conclusion

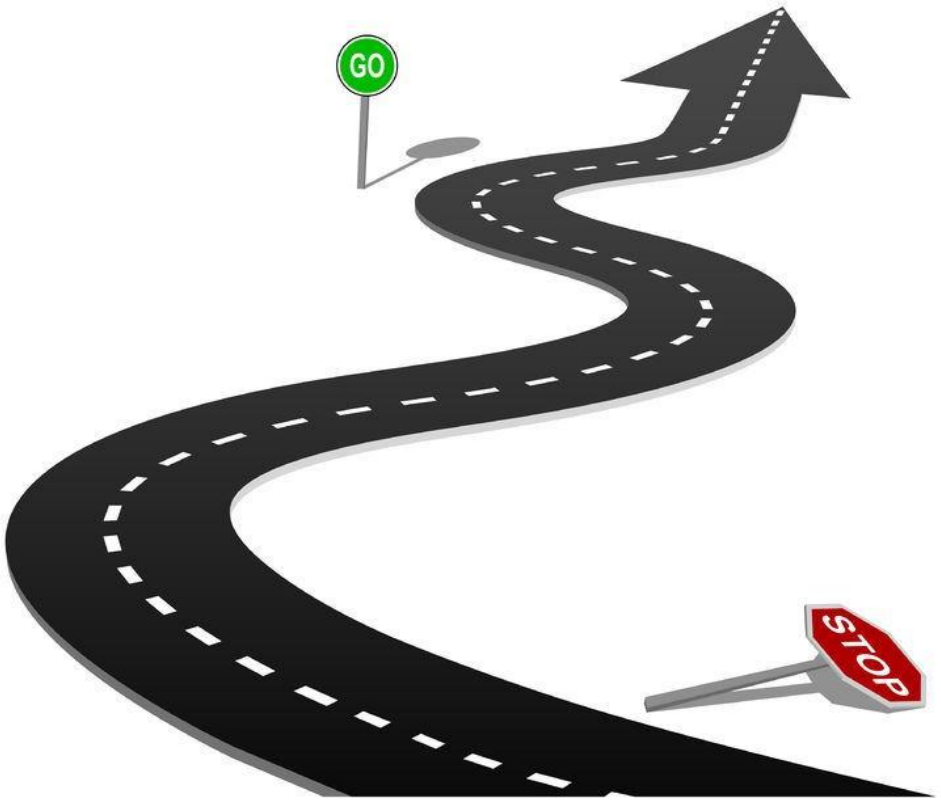
- Heuristic sampling can be used to sample solutions from RE models.
- Fast and diverse optimization of RE models can be performed using DE.
- “Keys” can be found in RE models which
  - are easy to find
  - help reason about the model

Open Problems:

- But what about non RE models?
- What if the top decisions contradict each other?

Roadmap

- Background
- Related Work
- Data
- SHORT
- Research Questions
- Conclusions
- **Future Work**



Future Work 1

Other RE Models

- Largest model in this study
  - Services : 351 Nodes & 510 Edges
  - Synthetic models demonstrating scale
- Currently on ***i**** based softgoal models. Extend to
  - NFR
  - TROPOS
  - KAOS
  - UML *e.t.c*
- GUI based tool.

Future Work 2

Other Real World Models

- **Autonomous Vehicles**[Zibaeenejad17]:
  - Appropriate resolution for interaction within a context.
  - Contexts to consider exponentially with features

- **Robot Configuration**[Jamshidi17]:
  - Thousands to Million config settings
  - Config interactions leads to exponential search space

Future Work 3

Sophisticated optimization algorithms

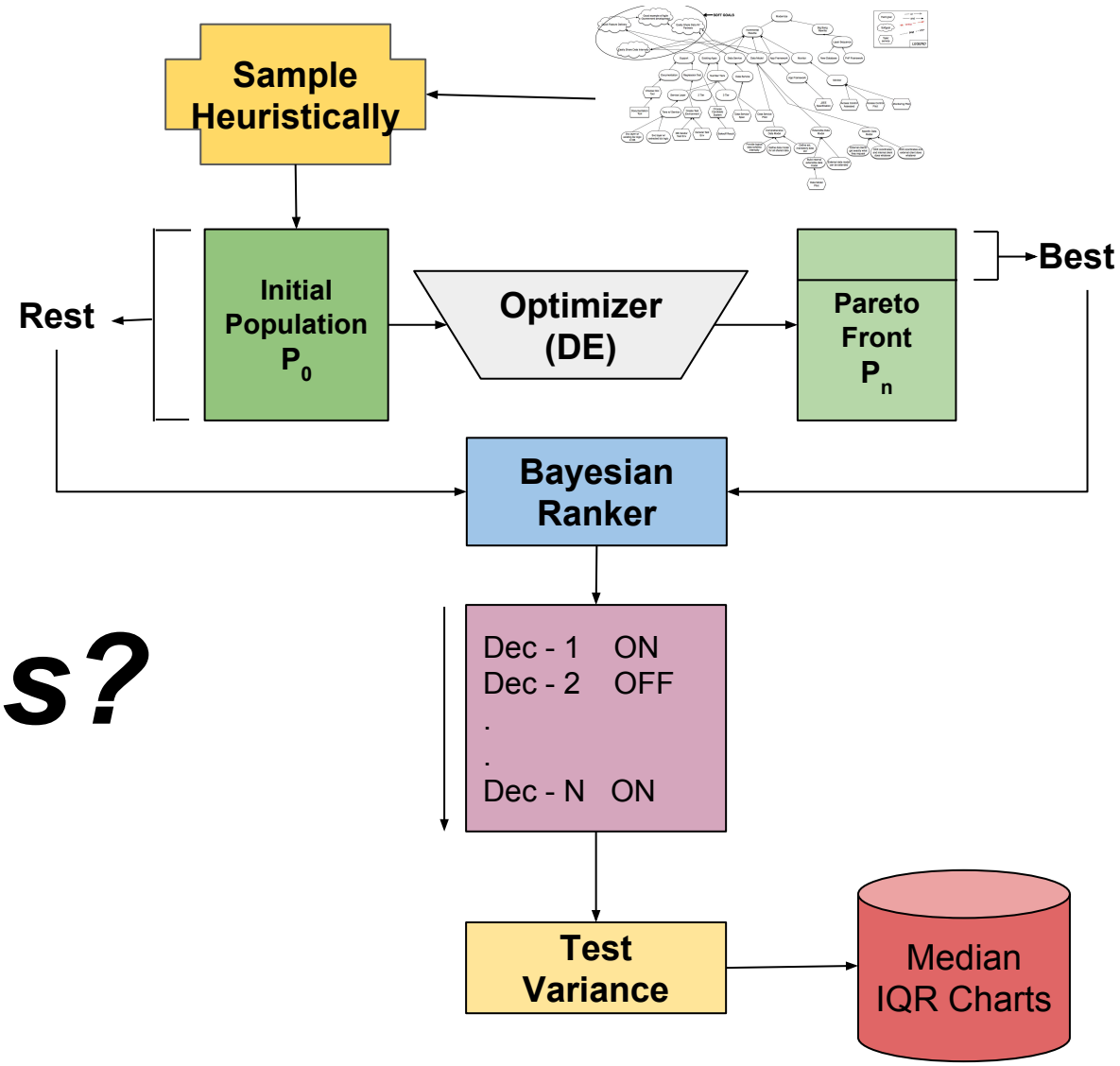
- **MOEA/D**[Zhang07]
- **SWAY**[Chen16]
- **FLASH**[Nair17]

Future Work 4

Storypoint Estimation[Choetkiertikul16]

- Estimation of Effort in modern agile SE
- Text summarization part of prior work[Mathew17ICSE].
- Featurization via Vectorizing
- Identifying prominent features via SHORT

FIN.


Sample Heuristically
Initial Population P_0
Rest
Optimizer (DE)
Pareto Front P_n
Best
Bayesian Ranker
Dec - 1 ON
Dec - 2 OFF
.
.
Dec - N ON
Test Variance
Median IQR Charts

Questions?

References

[Storn97] - Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization, 11(4), 341-359.

[Horkoff16] - J. Horkoff and E. Yu, “Interactive goal model analysis for early requirements engineering,” RE, vol. 21, 2016

[Feather00]-M.S. Feather et.al. “Scalable mechanisms for requirements interaction management.” Requirements Engineering, 2000.

[Park98] S. Park and J. Nang. "Requirements management in large software system development." Systems, Man, and Cybernetics, 1998.

[Regnell06] S. Regnell, “Assessing Requirements Compliance Scenarios in System Platform Subcontracting” Product Focused Software Process Improvement, 2006.

[Menzies96] T. Menzies, “Applications of abduction: knowledge-level modelling,” Intl Journal of Human-Computer Studies, vol. 45, 1996

[Bryl07] V. Bryl et.al., “Supporting Requirements Analysis in Tropos: a Planning-Based Approach”. PRIMA’07. Springer, 2007

[Wilson00] R.Wilson et.al., "Reduction techniques for instance-based learning algorithms." Machine learning 38.3 (2000): 257-286.

References

[Ernst12] N. Ernst et.al, “Agile Requirements Evolution via Paraconsistent Reasoning,” AISE’12, 2012

[Amarel66] S. Amarel, “Program synthesis as a theory formation task: problem representations & solution methods,” ML: An AI Approach’86.

[Crawford94] J.Crawford et.al, “Experimental results on the application of satisfiability algorithms to scheduling problems,” NCAI, 1994

[Williams03] R. Williams et.al, “Backdoors to typical case complexity,” International Joint Conference on Artificial Intelligence, 2003

[Menzies07] T. Menzies et.al, “The business case for automated software engineering,” International Conference on Automated Software Engineering, 2007

[Poole85] D .Poole, “On the comparison of theories : preferring the most specific explanation”, IJCAI, 1985

[DeKleer86] J. DeKleer, “An assumption-based TMS”, Artificial Intelligence, 1986.

[DeKleer87] J. de Kleer et.al, “Foundations for assumption-based truth maintenance systems: Preliminary report”, In Proc. American Assoc. for Artificial Intelligence Nat. Conf (pp. 183-188), July, 1987

[Pearson01] K.Pearson,. "On lines and planes of closest fit to systems of points in space." The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 2.11 (1901): 559-572.

References

[Deb01] K. Deb et. al, "A fast and elitist multiobjective genetic algorithm: NSGA-II." IEEE transactions on evolutionary computation 6.2 (2002): 182-197.
[Zhang07] Q. Zhang et. al, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition." IEEE Transactions on evolutionary computation 11.6 (2007): 712-731.
[Chen16] J. Chen et. al, “Is "sampling" better than "evolution" for search-based software engineering?” CoRR, vol. abs/1608.07617, 2016.
[Nair17] V. Nair et. al, "FLASH: A Faster Optimizer for SBSE Tasks." arXiv preprint arXiv:1705.05018 (2017).
[Menzies16] T. Menzies, et. al, "Negative results for software effort estimation." Empirical Software Engineering 22.5 (2016): 2658-2683.
[Mathew17ICSE] G. Mathew et. al, "Trends in topics at SE conferences (1993--2013)." Proceedings of the 39th International Conference on Software Engineering Companion. IEEE Press, 2017.
[Mathew17RE] G.Mathew et. al, "Shorter Reasoning About Larger Requirements Models." arXiv preprint arXiv:1702.05568 (2017).
[Boehm81] B. Boehm, Software engineering economics. Vol. 197. Englewood Cliffs (NJ): Prentice-hall, 1981.

References

[Ernst17] N. Ernst et. al, "Using Stakeholder Preferences to Make Better Architecture Decisions." Software Architecture Workshops (ICSAW), 2017 IEEE International Conference on. IEEE, 2017.

[Jamshidi17] P. Jamshidi et. al, "Transfer learning for performance modeling of configurable systems: An exploratory analysis." arXiv preprint arXiv:1709.02280 (2017).

[Zibaeenejad17] M. H. Zibaeenejad et. al, "Continuous variable-specific resolutions of feature interactions," in Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering. ACM, 2017, pp. 408–418.

[Choetkiertikul16] M. Choetkiertikul et. al, "A deep learning model for estimating storypoints,"arXiv preprint arXiv:1609.00489, 2016

[Kohavi97] R.Kohavi et. al, "Wrappers for feature subset selection." Artificial intelligence 97.1-2 (1997): 273-324.

[Thiery13] J. Thiery et. al, "Sphere-Meshes: shape approximation using spherical quadric error metrics." ACM Transactions on Graphics (TOG) 32.6 (2013): 178.

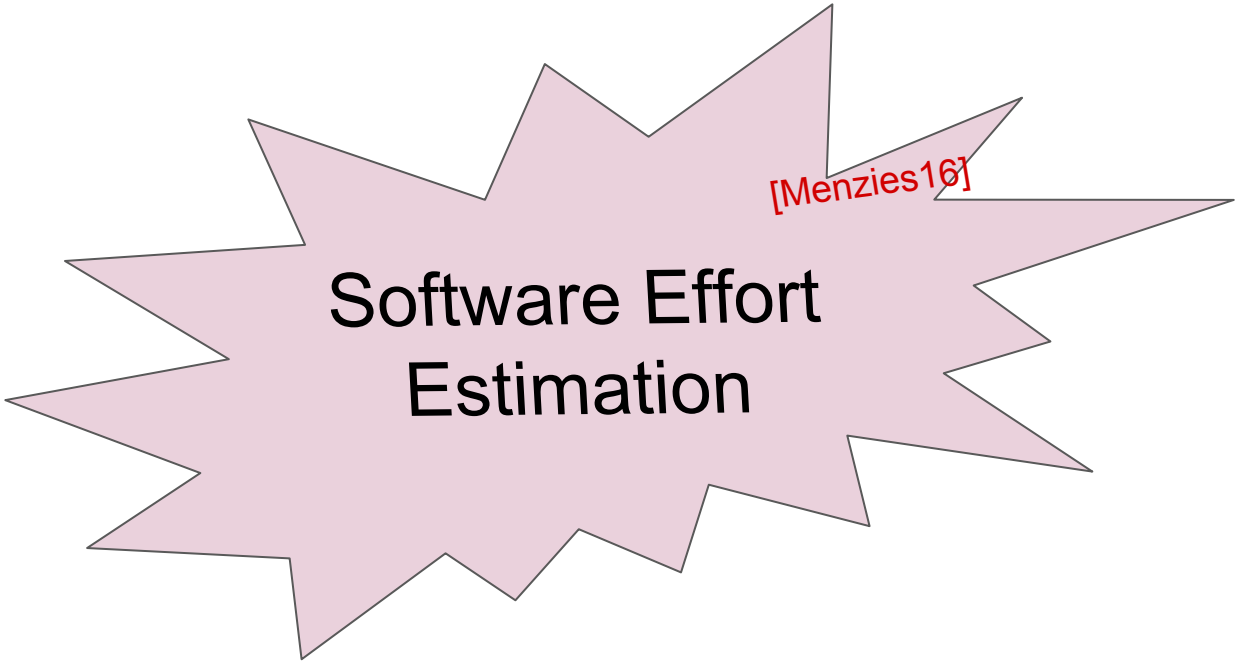
[Budd81] T. Budd, "Mutation Analysis of Program Test Data" (1981): 1826-1826.

[Kawthekar16] P. Kawthekar et.al, "Sensitivity analysis for building evolving & adaptive robotic software." IJCAI Workshop on Autonomous Mobile Service Robots. 2016.

Backups

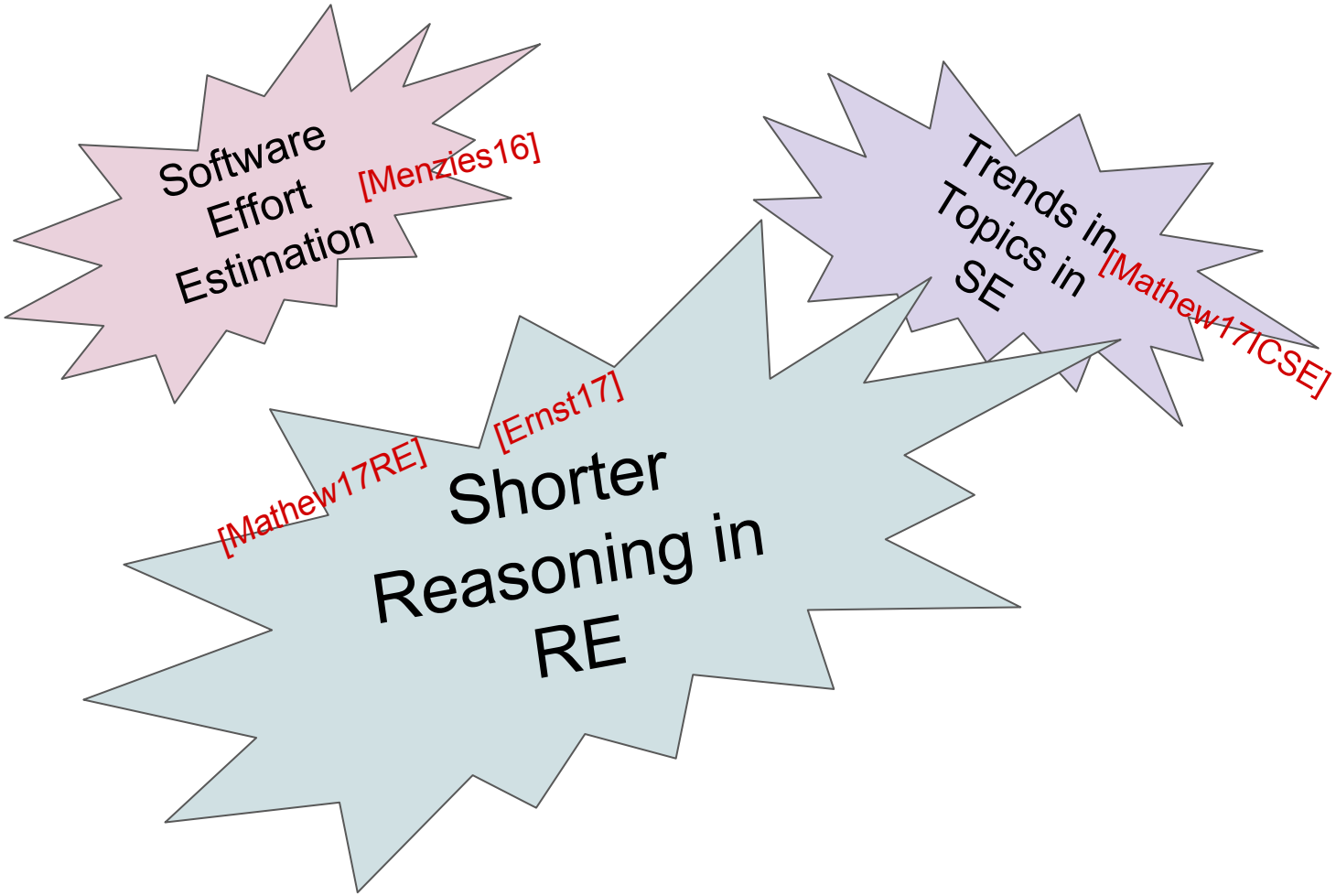
[Slide intentionally left blank]

“SHORT” Encounters



"SHORT" Encounters

"SHORT" Encounters



How these models are created

How these models are created

- Techniques for interacting requirements
- Consolidated into a tool (Defect Detection & Prevention)
- Used by NASA - JPL

[Feather00]

We're not alone

- (~ 100 Nodes)[Feather00]:
  - Techniques for interacting requirements
  - Consolidated into a tool (Defect Detection & Prevention)
  - Used by NASA - JPL

We're not alone

- (~ 100 Nodes)[Feather00]:
  - Techniques for interacting requirements
  - Tool (Defect Detection & Prevention)
  - Used by NASA - JPL
- (~ 500 Nodes) "Engineering Baselines"[Park98]:
  - Machine processing of requirements statements.
  - Correlation b/w requirements and system elements
  - Toggle elements and control.

We're not alone

- (~ 100 Nodes)[Feather00]:
  - Techniques for interacting requirements
  - Tool (Defect Detection & Prevention)
  - Used by NASA - JPL
- (~ 500 Nodes) "Engineering Baselines"[Park98]:
  - Machine processing of requirements statements.
  - Correlation b/w requirements and system elements
  - Toggle elements and control.
- (~ 1000 Nodes)[Regnell06]:
  - Asses RE collaboration of SOC negotiation in mobile platform subcontracting.
  - Classification of platform compliance scenarios

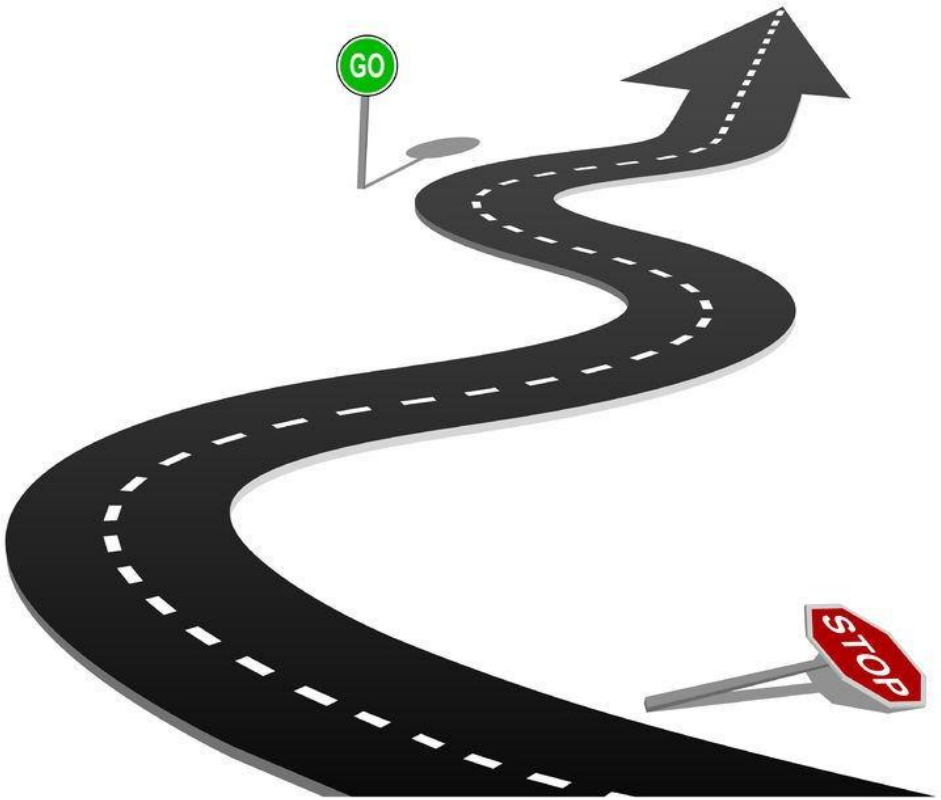
We're not alone

Large Effort / Time

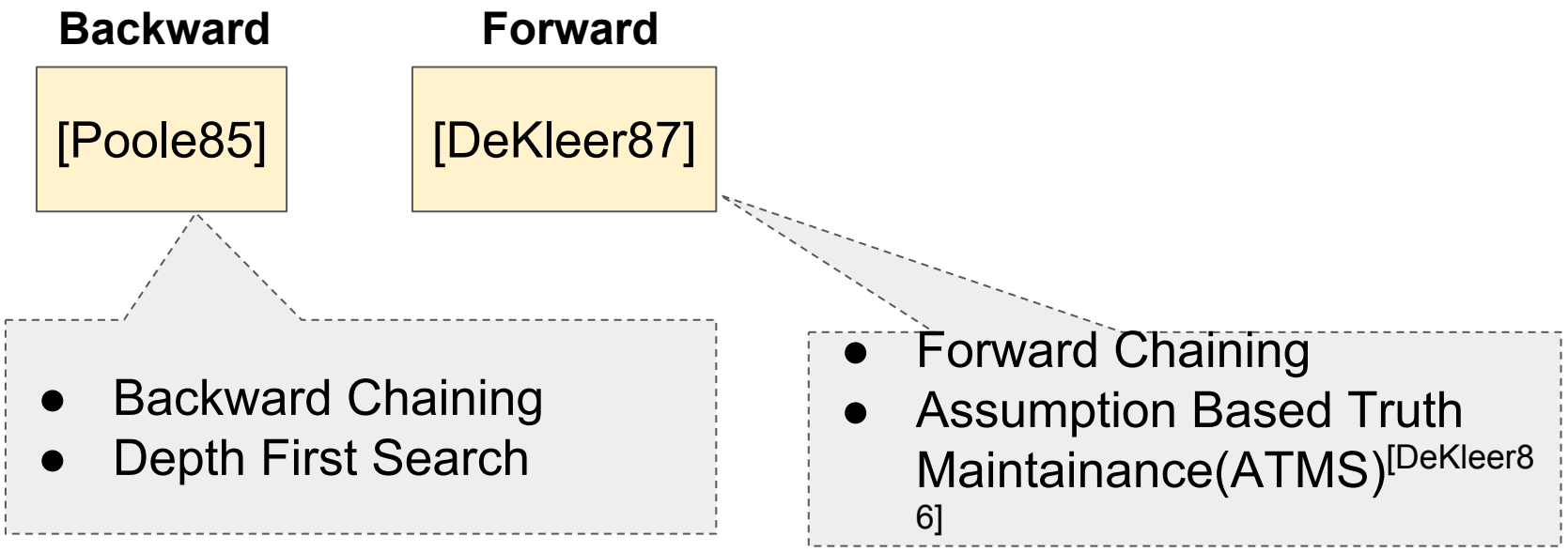
Ignore / Fail Conflicts

Roadmap

- **Background**
- **Related Work**
  - **Reasoning in RE**
  - **“Keys” in AI**
- Data
- SHORT
- Research Questions
- Conclusions
- Future Work



Related Work



“Keys” in Artificial Intelligence

Narrows

[Amarel86]

- Tiny sets of variable settings
- Macros = path b/w narrows

\+ Aids jumping b/w solutions

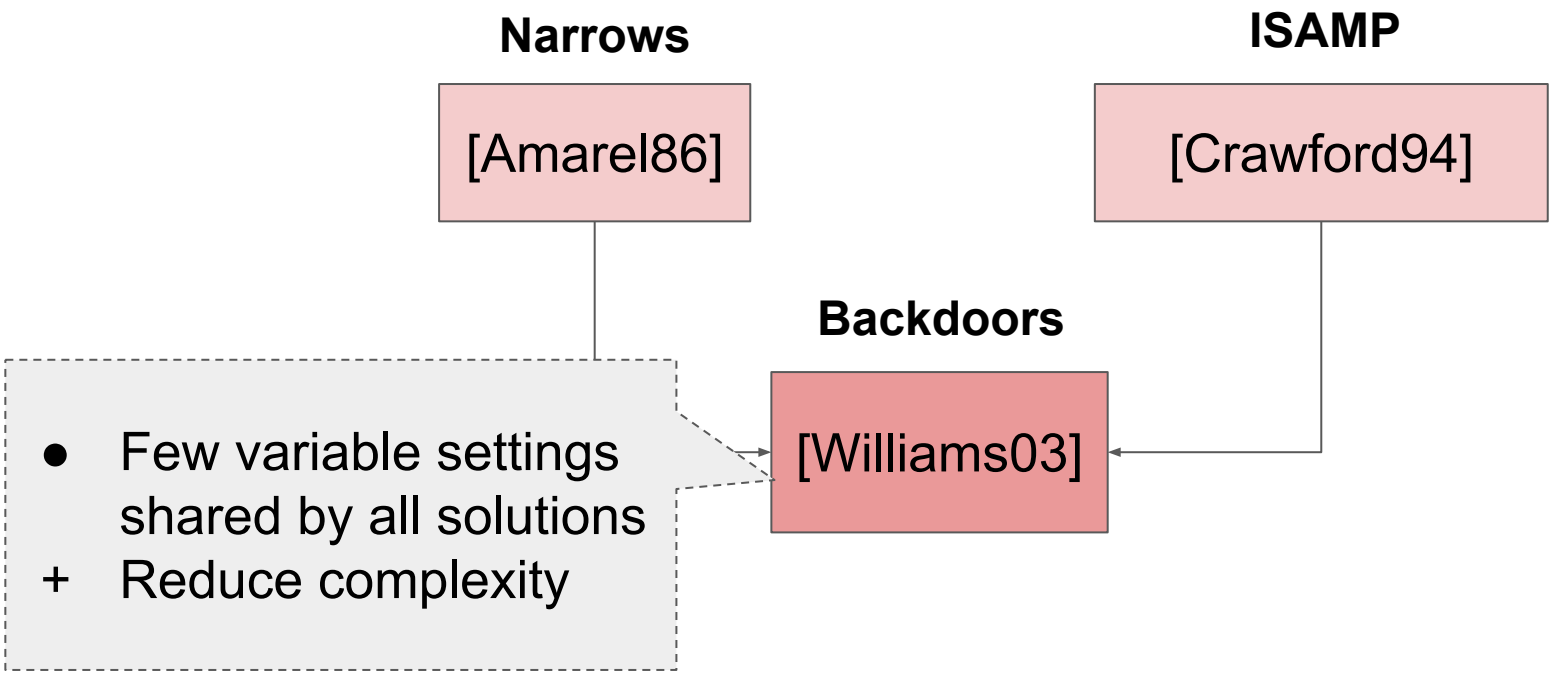
ISAMP

[Crawford94]

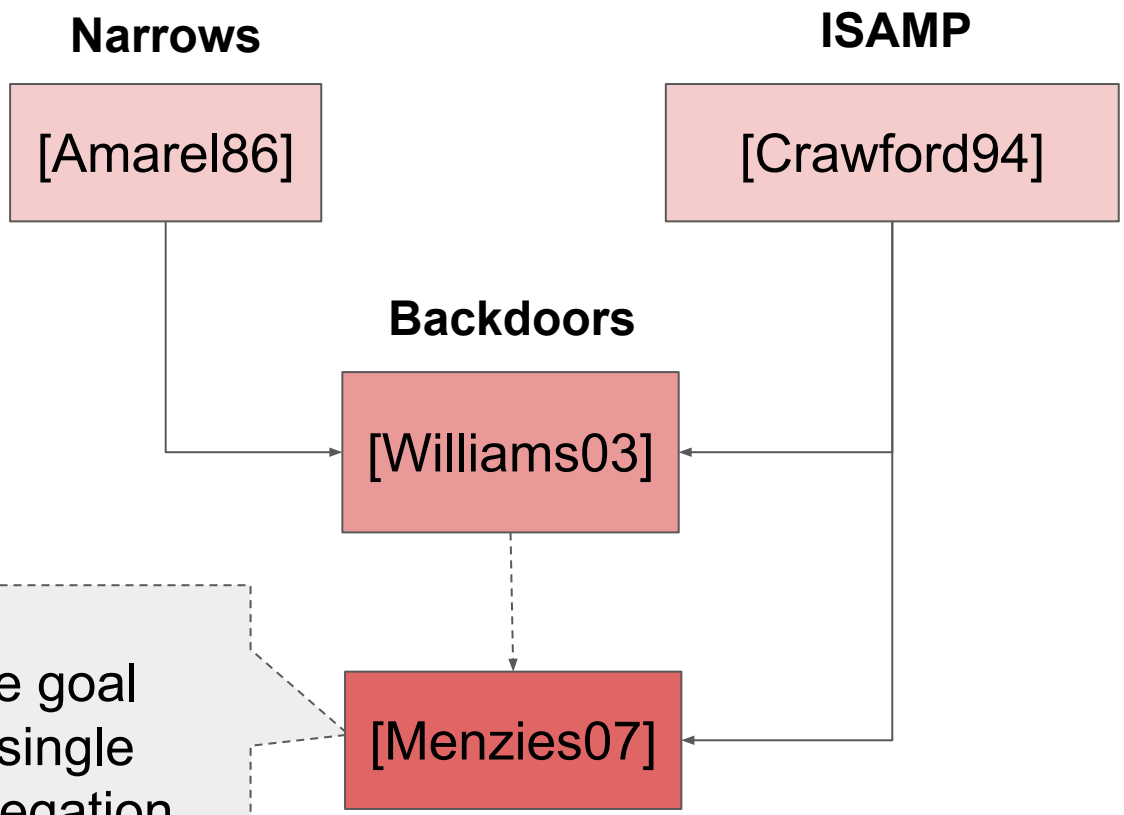
- Iterative Sampler
- Random search with retries when stuck

\+ Faster than full search

“Keys” in Artificial Intelligence

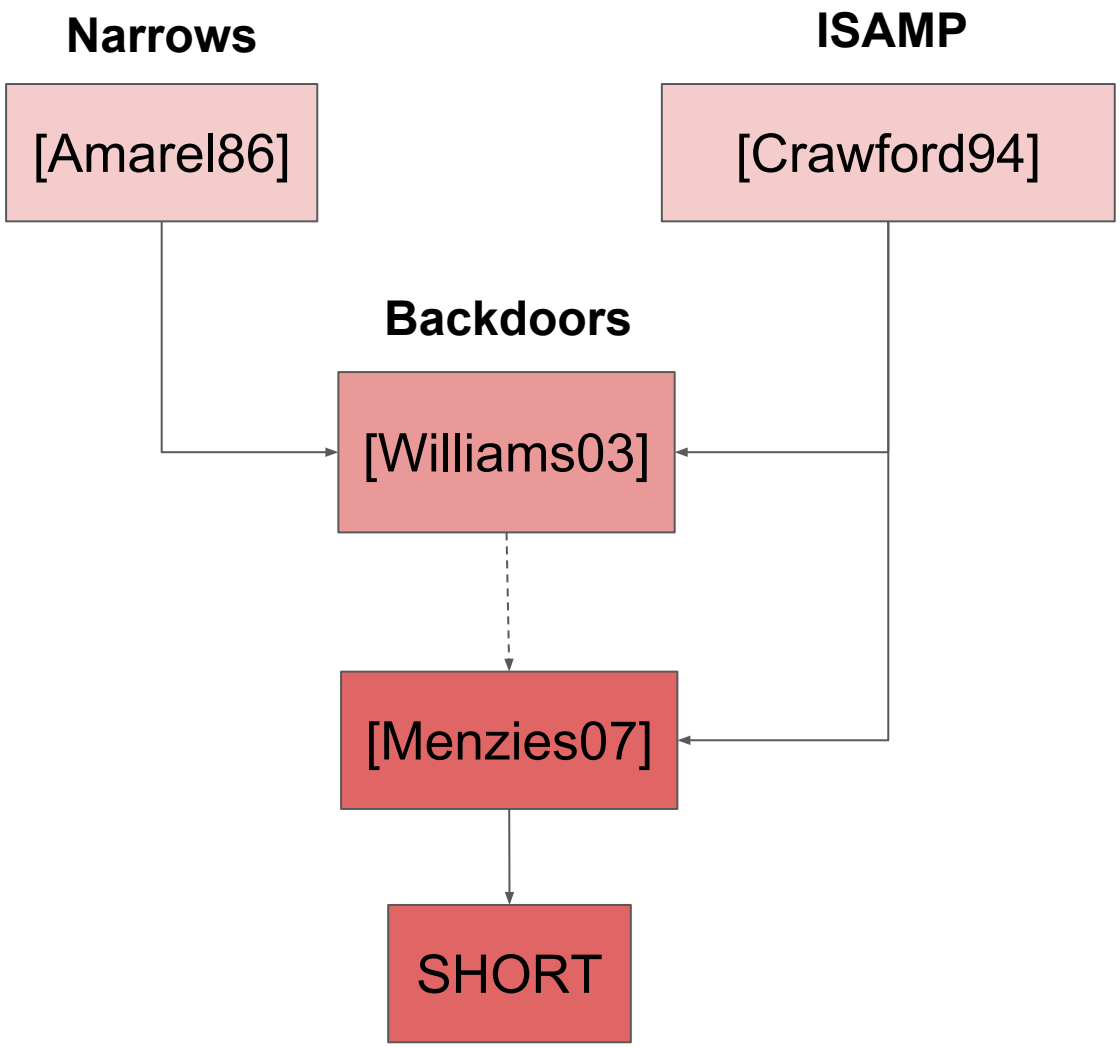


“Keys” in Artificial Intelligence

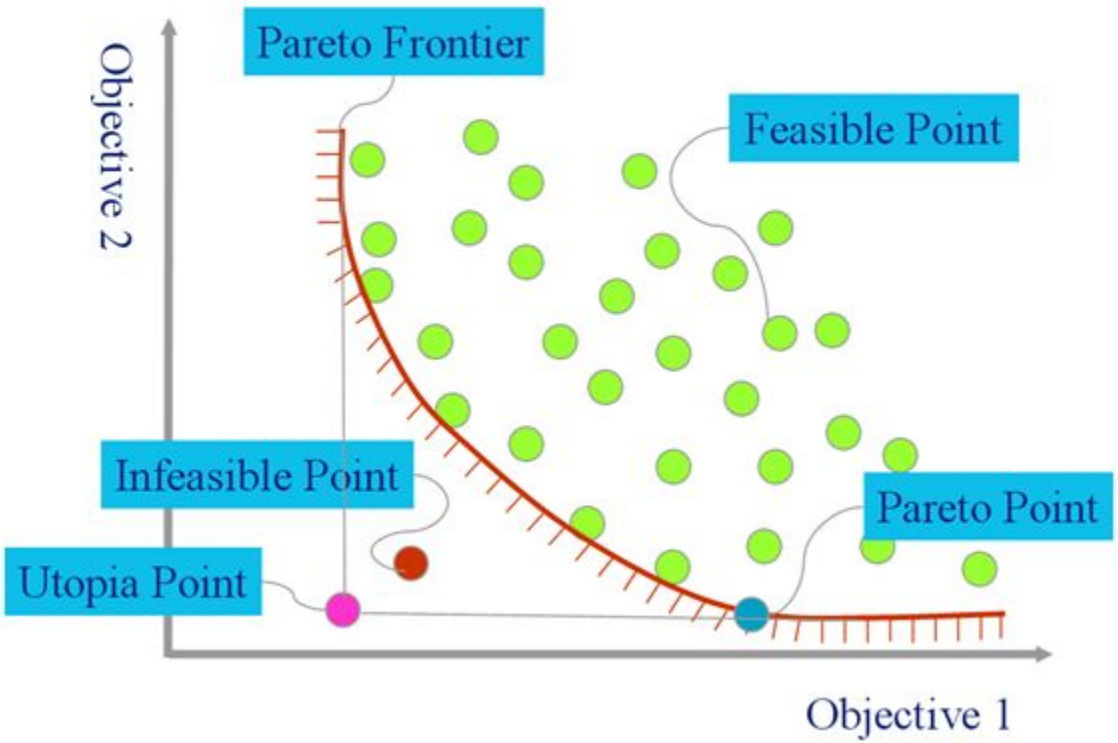


- Keys in single goal
  - Multi goal to single goal via aggregation
  - Weak optimizer

“Keys” in Artificial Intelligence



Optimization



Optimization - Meta Heuristic Techniques

- **NSGA-II**[2Deb01]:
  - Non-dominated Sorting Genetic Algorithm
  - Most used Evolutionary Meta Heuristic
  + Relatively fast(~ $O(n^2)$)
  - Poor diversity of solutions

Optimization - Meta Heuristic Techniques

- NSGA-II[Deb01]
- **MOEA/D**[Zhang05]
  - Multi-Objective Evolutionary Algorithm via Decomposition
  - Clusters population and evolves them independently
  - \+ Large diversity
  - \- Extremely slow(~ $O(n^3)$)

Optimization - Meta Heuristic Techniques

- NSGA-II[Deb01]
- MOEA/D[Zhang07]
- **Modern Techniques**:
  - SWAY[Chen16]: Sampling based
  - FLASH[Nair17]: Surrogate based
  - Part of future work.

Optimization - Meta Heuristic Techniques

- NSGA-II[Deb01]
- MOEA/D[Zhang07]
- Modern Techniques
- **DE**[Storn97]
  - Differential Evolution
  - Extrapolation based evolution.
  - Near Linear time(~ *O(n)*)

Optimization - Differential Evolution(DE)

- **Candidature**: Randomly select 4 candidates **x**, **a**, **b**, **c**

Optimization - Differential Evolution

- **Candidature**: Randomly select 4 candidates **x**, **a**, **b**, **c**
- **Reproduction + Mutation:**
  y = a + F*(b - c)

F = Mutation Rate ∈ [0, 1]

Optimization - Differential Evolution

- **Candidature**: Randomly select 4 candidates **x**, **a**, **b**, **c**
- **Reproduction + Mutation:**
  y = **a** ∪ **F** ∩ (**b** / **c**)

Optimization - Differential Evolution

- **Candidature**: Randomly select 4 candidates **x**, **a**, **b**, **c**
- **Reproduction + Mutation:**
  y = **a** ∪ **F** ∩ (**b** / **c**)
- **Evaluation**

Optimization - Differential Evolution

- **Candidature**: Randomly select 4 candidates **x**, **a**, **b**, **c**
- **Reproduction + Mutation:**
  y = **a** ∪ **F** ∩ (**b** / **c**)
- **Evaluation**
- **Selection:**

if **y** "dominates" **x** **then**
replace **x** with **y**

Domination[Storn97]

$$\forall i \; f_i(y) \geq f_i(x)$$
$$\exists j \; f_j(y) > f_j(x)$$

Optimization - Differential Evolution

- **Candidature**: Randomly select 4 candidates **x**, **a**, **b**, **c**
- **Reproduction + Mutation:**
  y = **a** ∪ **F** ∩ (**b** / **c**)
- **Evaluation**
- **Selection:**

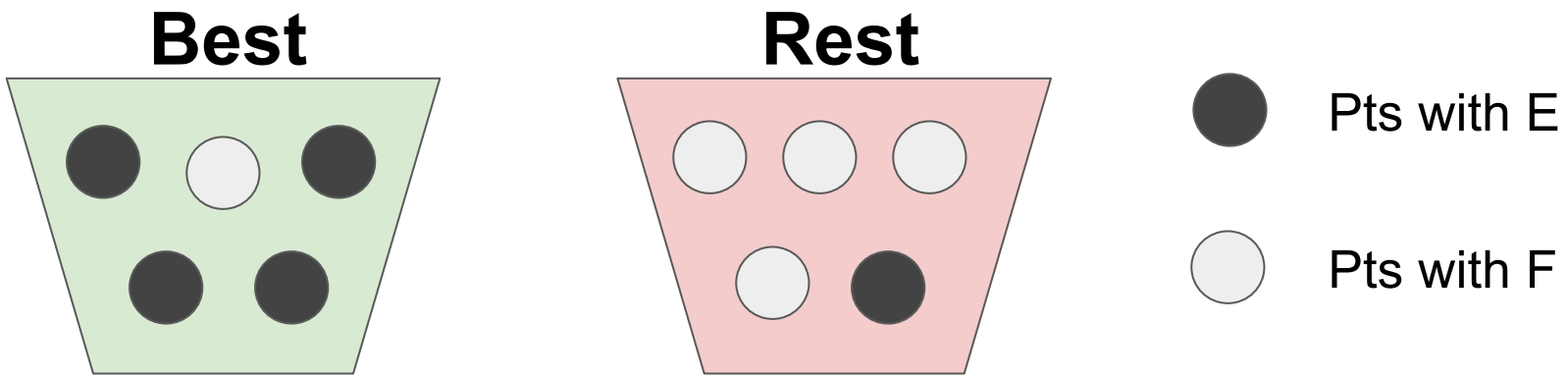
if **y** "dominates" **x** **then**
replace **x** with **y**

Domination

$$\forall i \; f_i(y) \geq f_i(x)$$
$$\exists i \; f_i(y) > f_j(x)$$

- **Termination**

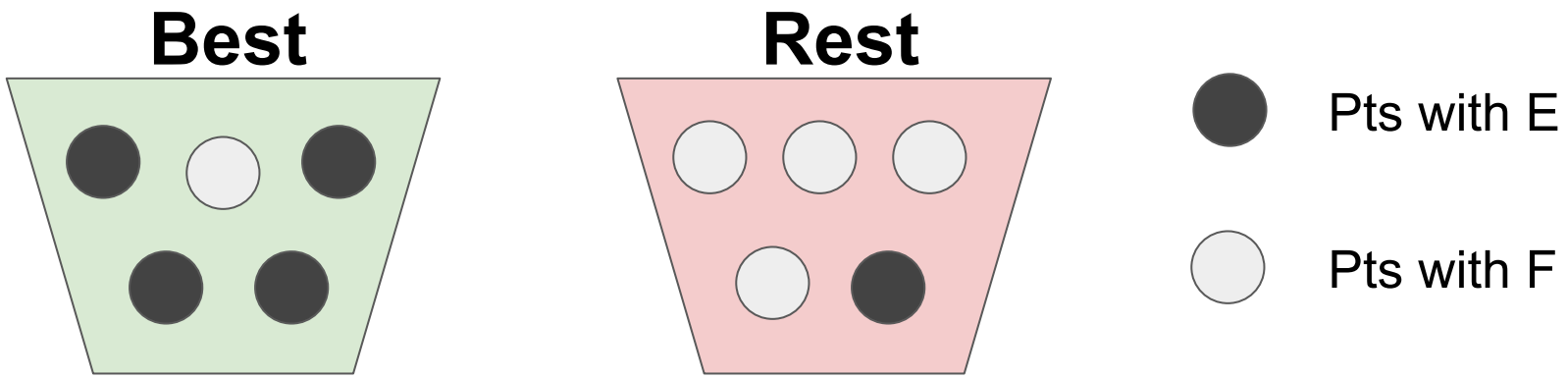
Bayesian Ranking



$$P(Best) = \frac{Best}{Best + Rest} = \frac{5}{5+5} = \frac{1}{2}$$

$$P(Rest) = \frac{Rest}{Best + Rest} = \frac{5}{5+5} = \frac{1}{2}$$

Bayesian Ranking

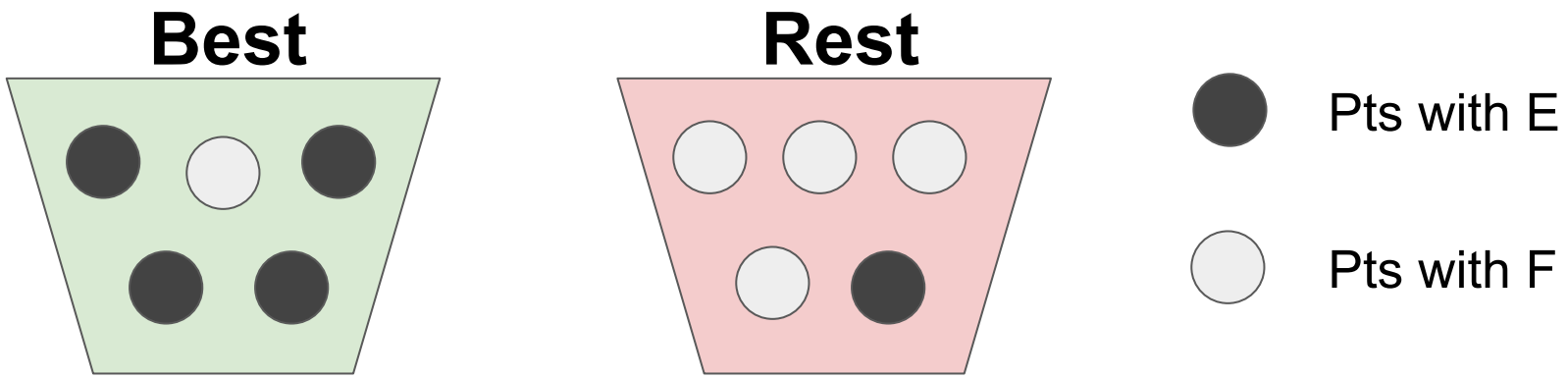


$$P(Best) = P(Rest) = \frac{1}{2}$$

$$P(E|Best) = \frac{Best_E}{Best} = \frac{4}{5}$$

$$P(E|Rest) = \frac{Rest_E}{Rest} = \frac{1}{5}$$

Bayesian Ranking



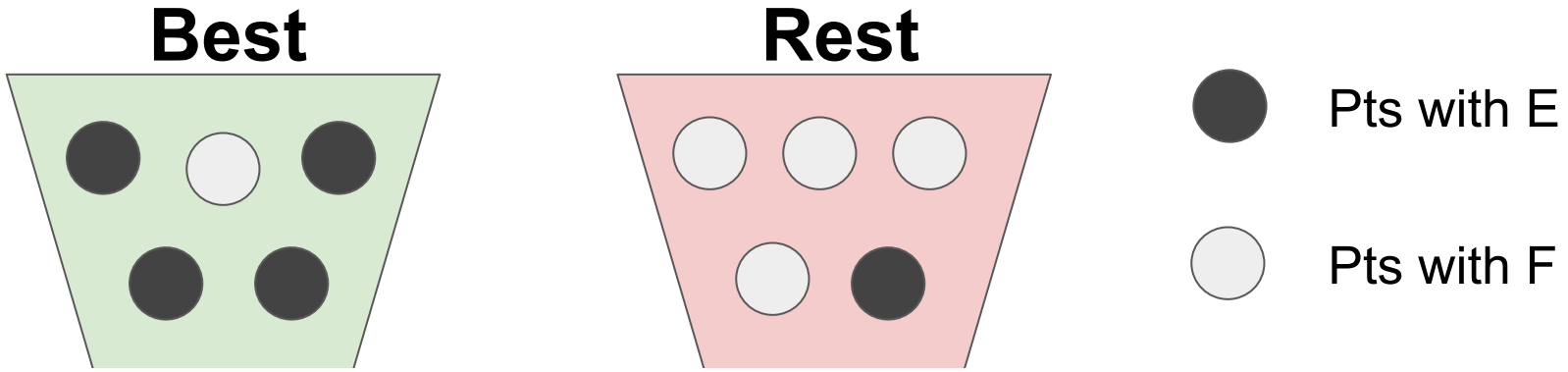
$$P(Best) = P(Rest) = \frac{1}{2}$$

$$P(E|Best) = \frac{4}{5};\ P(E|Rest) = \frac{1}{5}$$

$$like(E|Best) = P(E|Best) * P(Best) = \frac{4}{10}$$

$$like(E|Rest) = P(E|Rest) * P(Rest) = \frac{1}{10}$$

Bayesian Ranking



$$like(E|Best) = \frac{4}{10}; like(E|Rest) = \frac{1}{10}$$

$$support(E) = \frac{like(E|Best)}{like(E|Best) + like(E|Rest)} = \frac{4}{5}$$

Similarly,

$$support(F) = \frac{1}{5}$$

- Rank decisions in decreasing order of support

RQ3: Do "Keys" exist in RE models

Model	Result	Percentage of Decisions				
		6	12	25	50	100
Counselling	SGs	60	70	70	70	60
	Goals	60	70	70	70	40
	Costs	5	10	20	28	35
Management	SGs	50	60	60	50	40
	Goals	50	50	60	60	60
	Costs	4	8	20	24	32
Marketing	SGs	70	70	60	60	30
	Goals	70	70	70	60	40
	Costs	6	8	11	12	20
ITDepartment	SGs	70	60	70	70	60
	Goals	50	50	80	80	70
	Costs	2	4	9	17	18
SAProgram	SGs	80	80	80	80	80
	Goals	60	70	60	55	60
	Costs	1	2	3	3	5
Services	SGs	50	50	50	50	50
	Goals	50	60	60	60	70
	Costs	4	7	13	19	24
Kids&Youth	SGs	40	50	50	40	20
	Goals	50	70	60	60	70
	Costs	1	2	4	7	10