Search Based Requirements Selection and Optimisation

Yuanyuan Zhang
CREST Centre
University College London

Agenda

Background

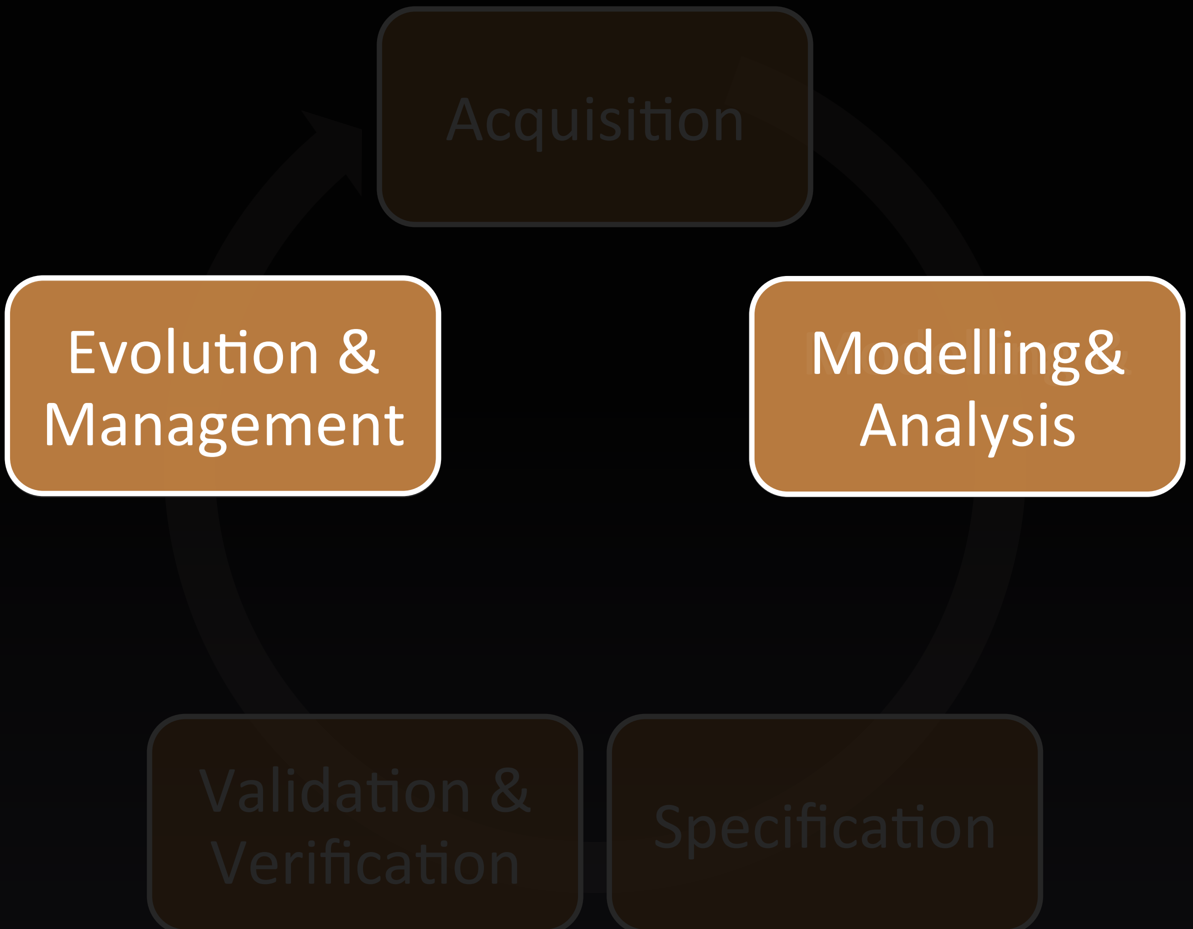
Problem

Solution

Empirical Study

Conclusion

Requirements Engineering Process



University College London yuanyuan.zhang@cs.ucl.ac.uk

Requirements Selection & Optimisation

Task

Using prioritisation, visualisation, and optimisation techniques helps decision maker to select the optimal or near optimal subset from all possible requirements to be implemented.

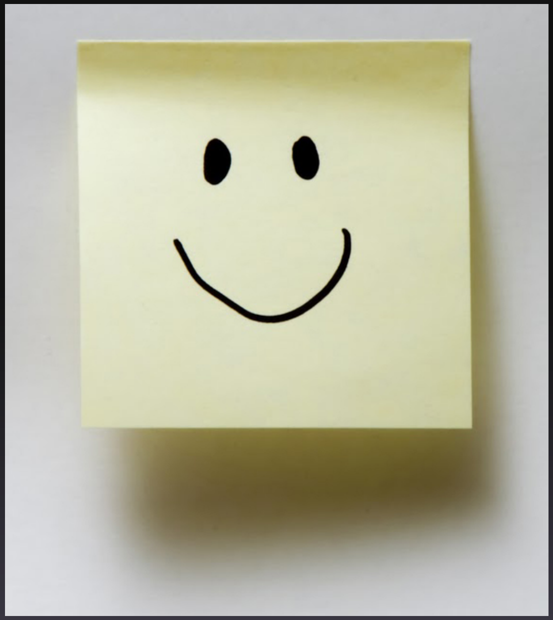
University College London yuanyuan.zhang@cs.ucl.ac.uk

Requirements Interaction Management

"the set of activities directed towards the discovery, management, and disposition of critical relationships among sets of requirements. "

Requirements Change

Unpredictable change



Predictable change

Problem

Requirements Selection

University College London yuanyuan.zhang@cs.ucl.ac.uk

Problem

Requirements Selection

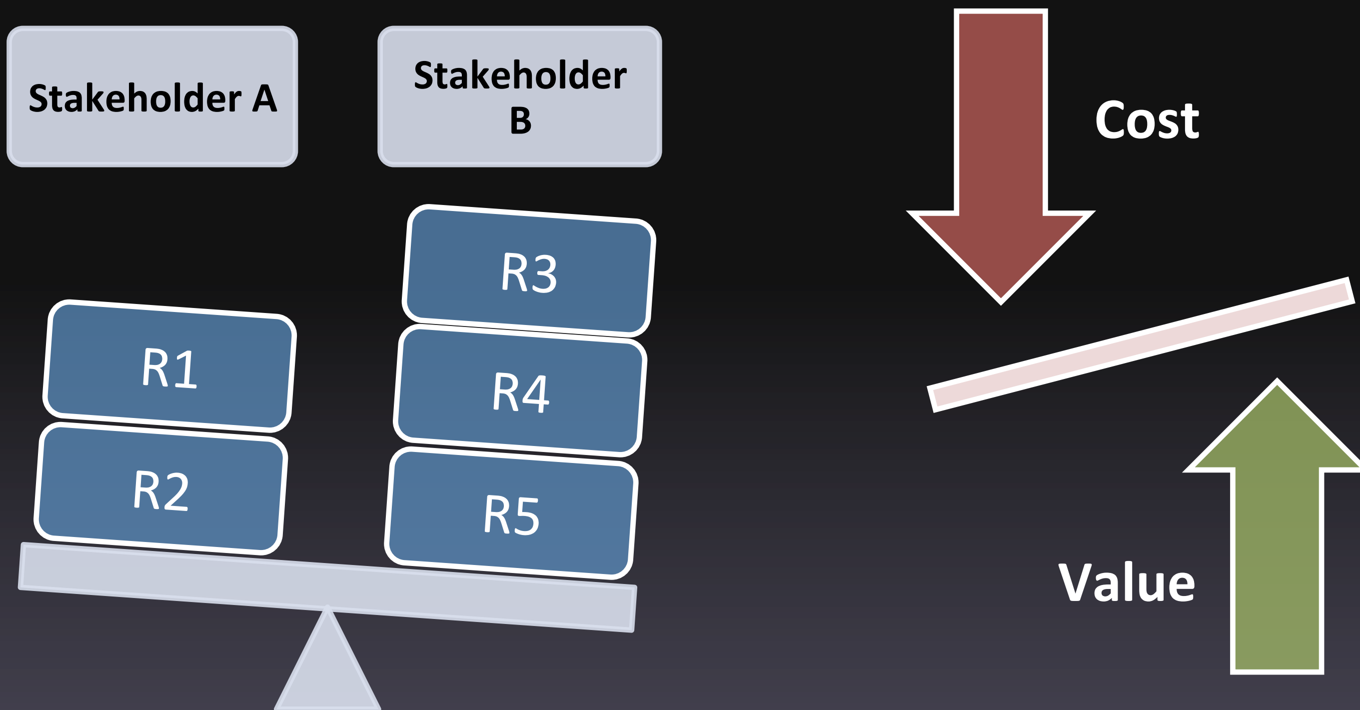
University College London yuanyuan.zhang@cs.ucl.ac.uk

Problem

Requirements Selection

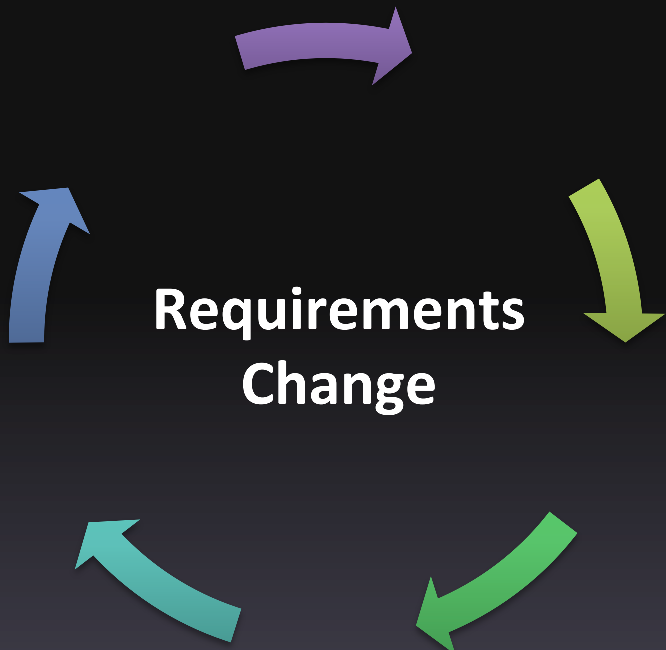
University College London yuanyuan.zhang@cs.ucl.ac.uk

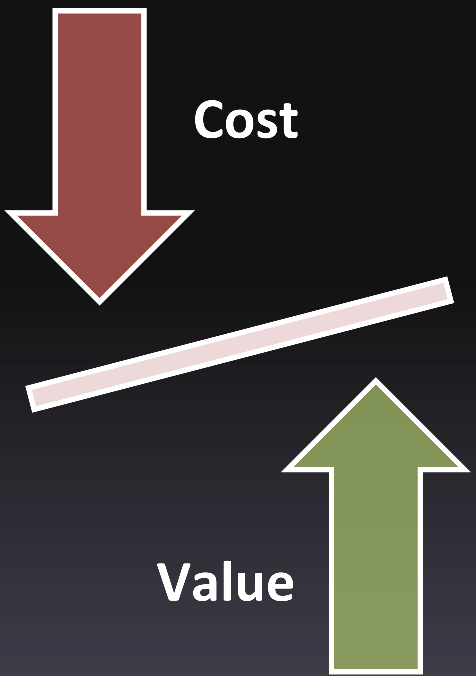
Goals



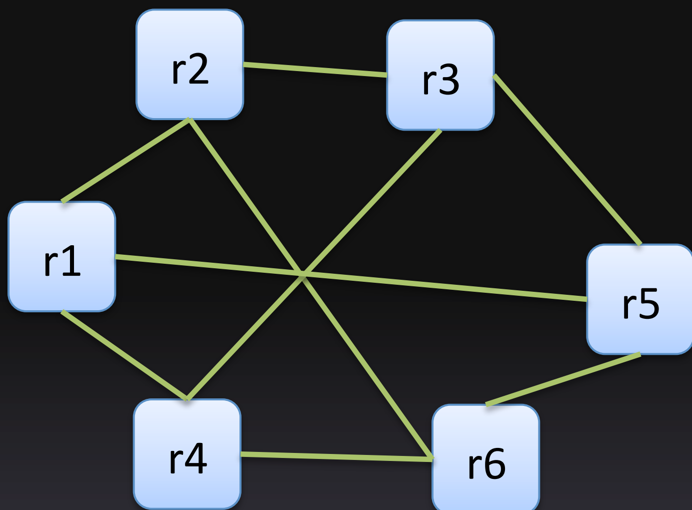
University College London yuanyuan.zhang@cs.ucl.ac.uk

Goals

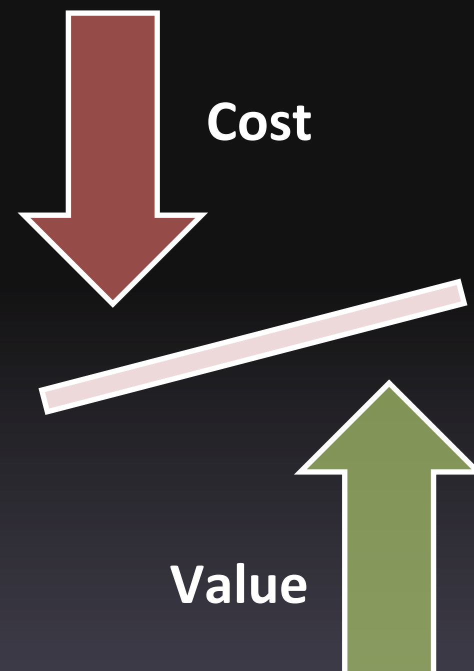




Goals



Requirements Dependency



University College London yuanyuan.zhang@cs.ucl.ac.uk

Search based Requirements Optimisation

Use of meta-heuristic algorithms to automate and optimise requirements selection process

- Choose appropriate representation of problem
- Define problem specific fitness function (to evaluate potential solutions)
- Use search based techniques to lead the search towards optimal points in the solution space

University College London yuanyuan.zhang@cs.ucl.ac.uk

Why Search Based Approach?

Robustness

Scalability

Sensitivity analysis

Insight

Feedback & Explanation of results

…

University College London yuanyuan.zhang@cs.ucl.ac.uk

Model

Stakeholder:

$$C = \{c_1, \ldots, c_j, \ldots, c_m\}$$

Weight:

$$Weight = \{w_1, \ldots, w_j, \ldots, w_m\}$$

Requirements:

$$R = \{r_1, \ldots, r_i, \ldots, r_n\}$$

Cost:

$$Cost = \{cost_1, \ldots, cost_n\}$$

University College London yuanyuan.zhang@cs.ucl.ac.uk

Model

- Each stakeholder c_j assigns a *value* to requirements r_i :

$$value\left(r_i, c_j\right)$$

- Each stakeholder c_j has a subset of requirements that expect to be fulfilled denoted by R_j

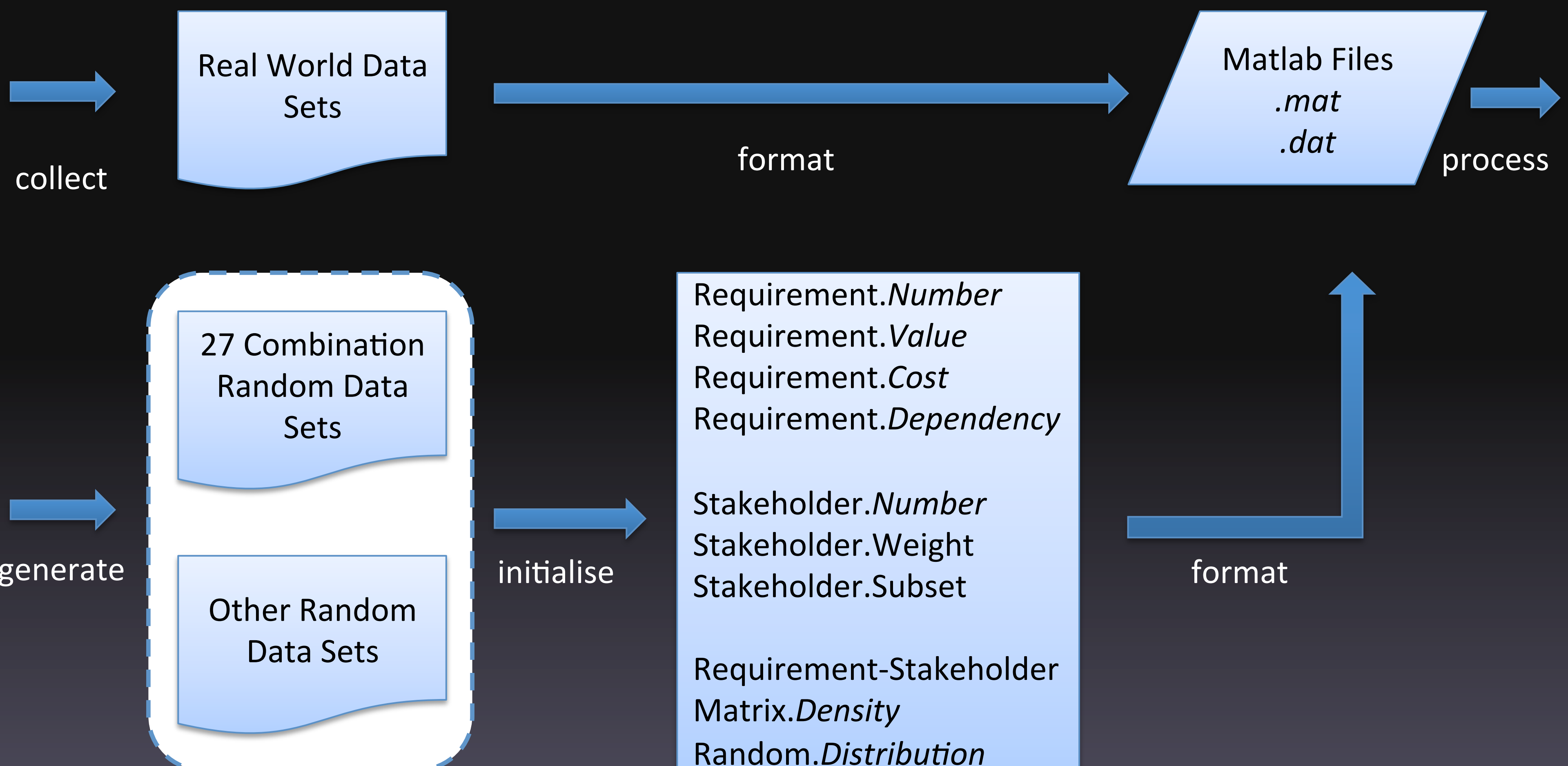
$$R_j \subseteq R, \qquad \forall r \in R_j \qquad value\left(r, c_j\right) > 0$$

- The overall *score* of a given requirement r_i can be calculated by:

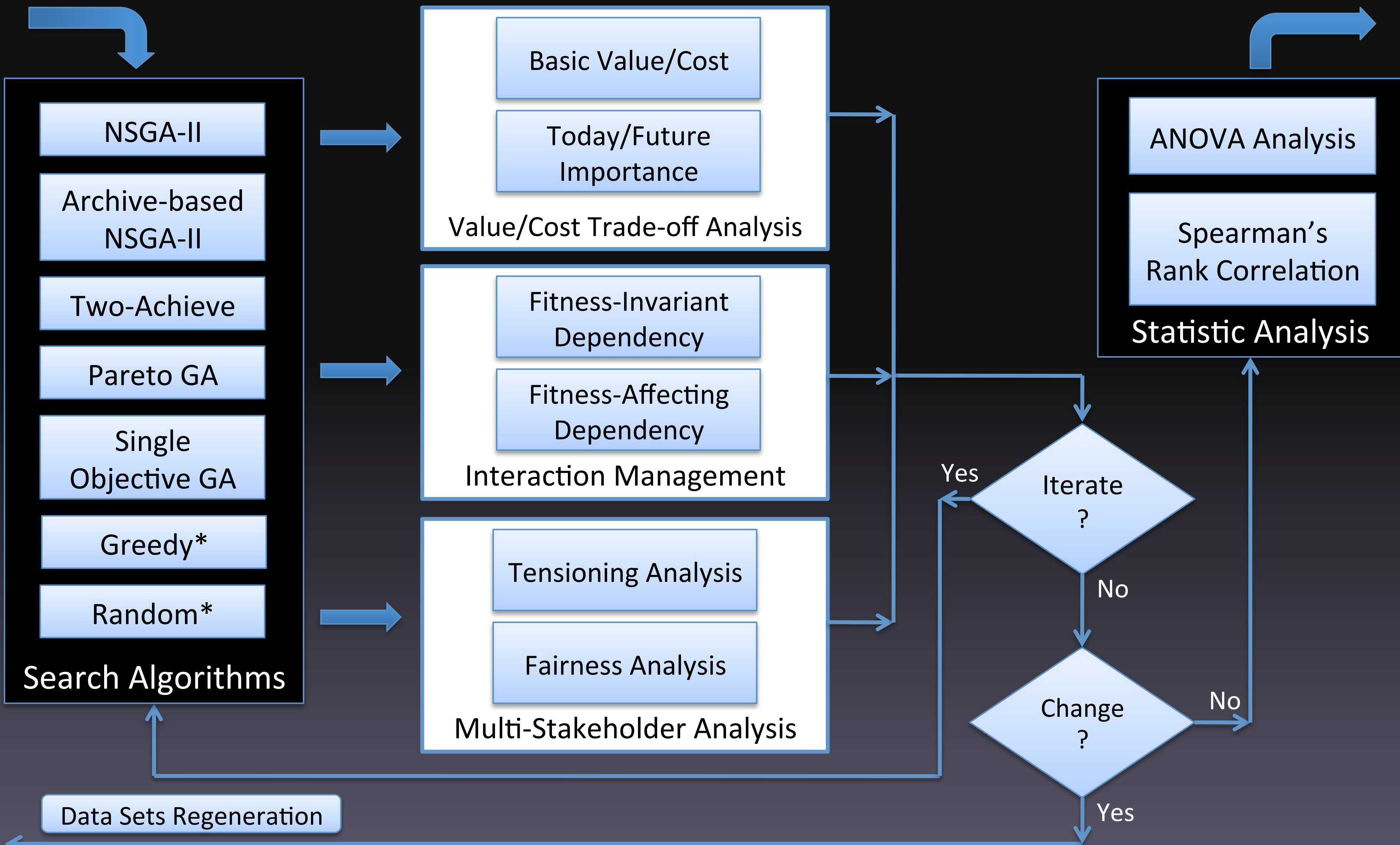
$$score_i = \sum_{j=1}^{m} w_j \cdot value\left(r_i, c_j\right)$$

University College London yuanyuan.zhang@cs.ucl.ac.uk

Data Set Collection & Initialisation



Requirements Selection Process



* Strictly speaking, these are not search algorithms.

Result Representation and Visualisation

Results

- Requirements Subsets for Release Planning
- Insight Characteristic of Data Sets
- Performance of the Algorithms

→ represent & visualise →

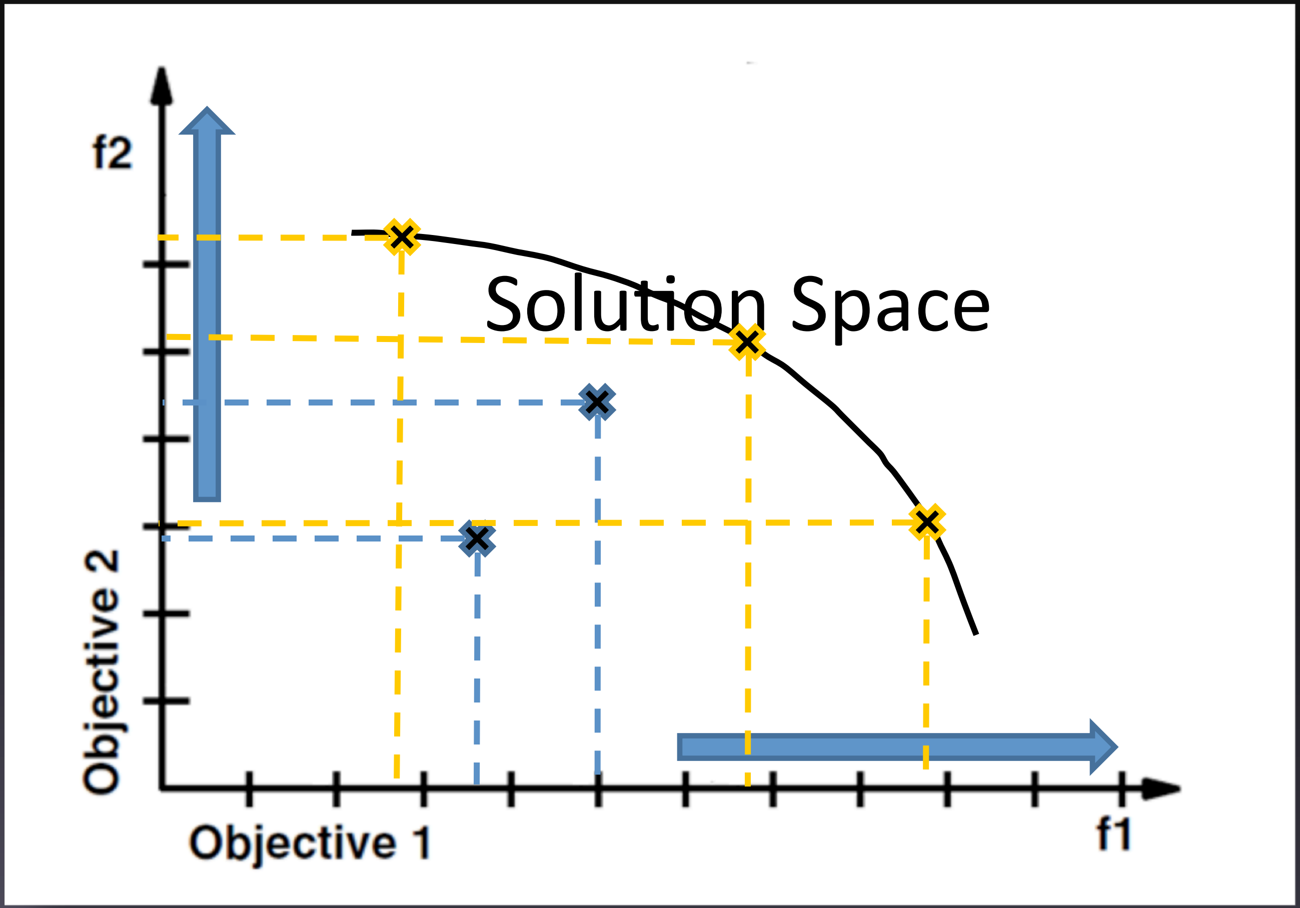
- 2D and 3D Pareto Fronts
- Kiviat Diagrams
- Marked Line Charts
- Convergence
- Diversity

↔ communicate & feedback

University College London yuanyuan.zhang@kcl.ac.uk

Visualisation

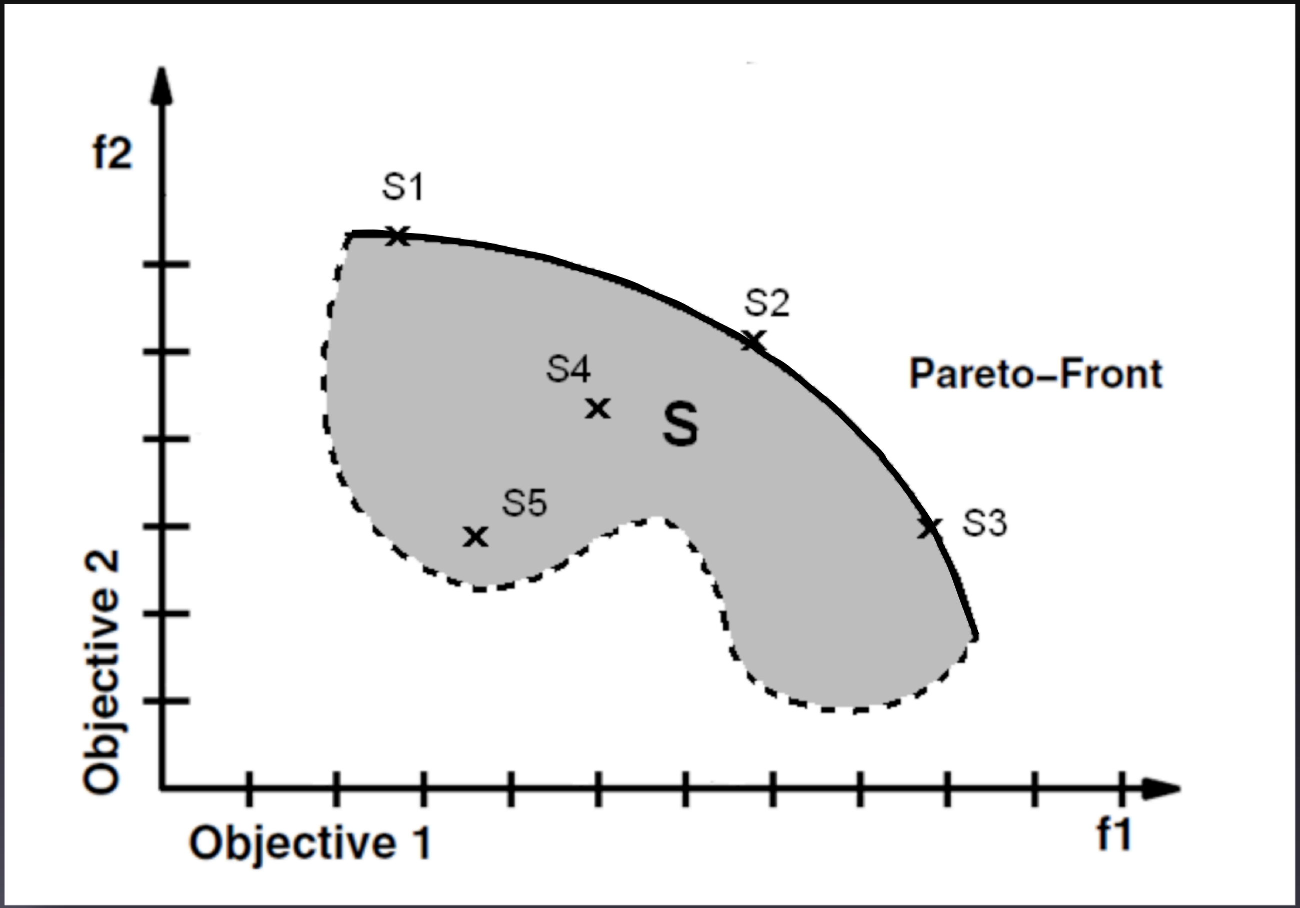
Pareto Optimal Front



University College London yuanyuan.zhang@cs.ucl.ac.uk

Visualisation

Pareto Optimal Front



1. Basic Value/Cost Trade-off

The problem is to select a set of requirements that maximise customers' satisfaction (total value) and minimise required cost.

The model of fitness functions represented as:

Maximise $$f_1\left(\vec{x}\right) = \sum_{i=1}^{n} score_i \cdot x_i$$

Minimise $$f_2\left(\vec{x}\right) = \sum_{i=1}^{n} cost_i \cdot x_i$$

University College London yuanyuan.zhang@cs.ucl.ac.uk

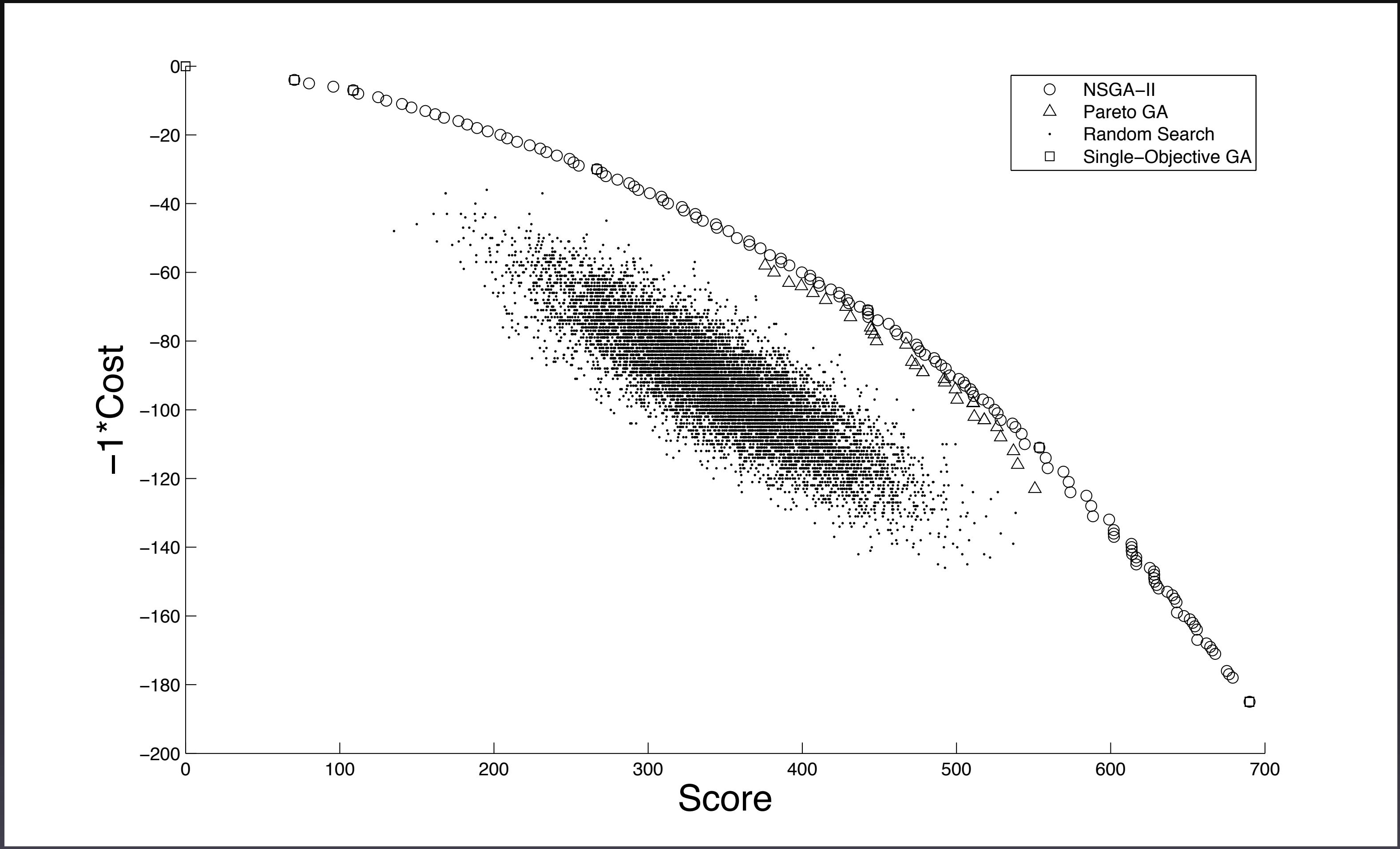
1. Basic Value/Cost Trade-off

Scale Problem

- consider three typical 'scales' cases of problem, with the number of customers ranging from 15 to 100 and the number of requirements ranging from 40 to 140.
- Investigate the relative performance of the approaches for cases.

University College London yuanyuan.zhang@cs.ucl.ac.uk

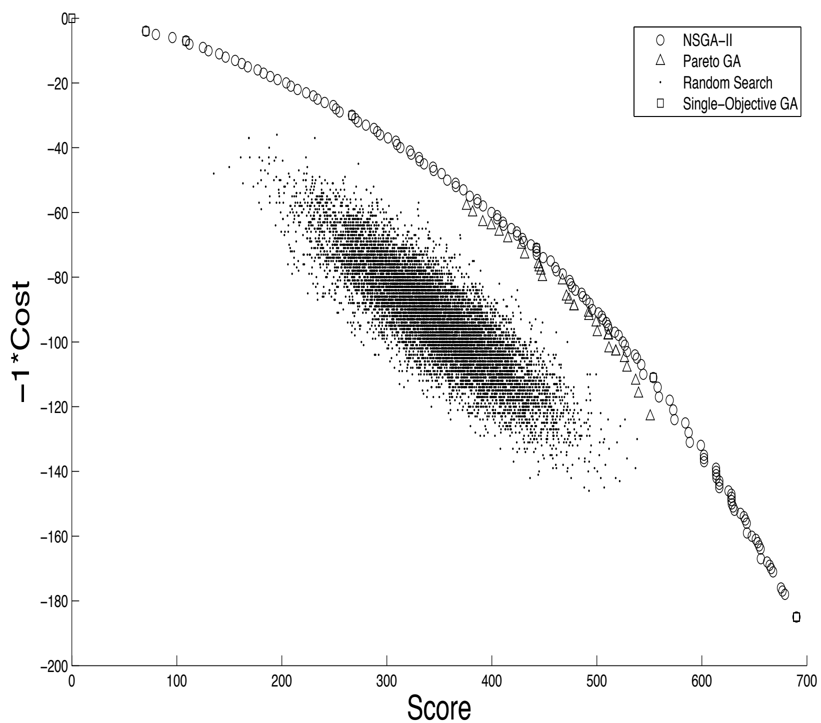
1. Basic Value/Cost Trade-off

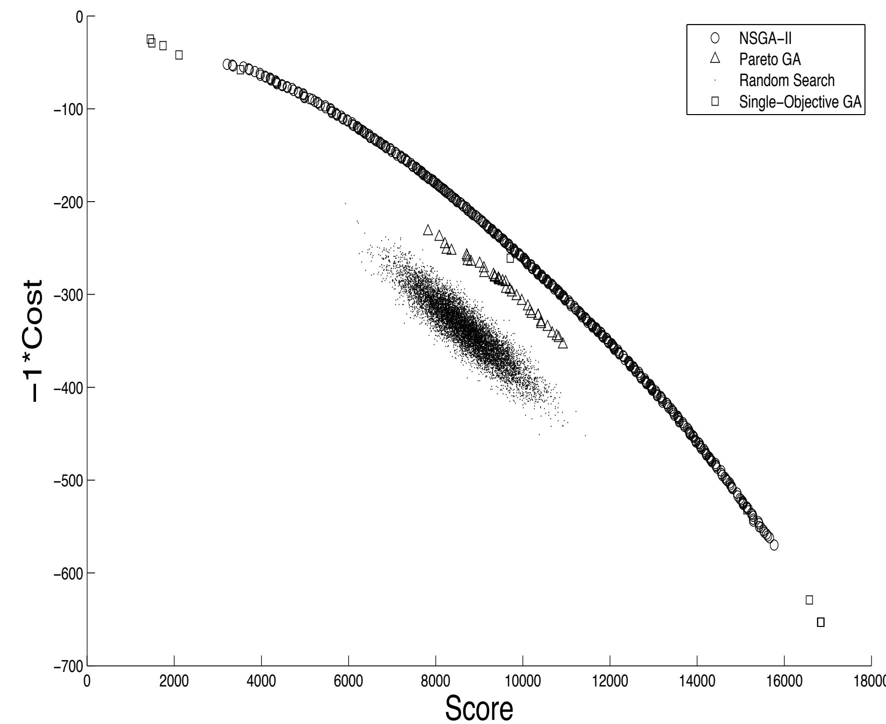


Synthetic data set: 15 stakeholders; 40 requirements

University College London yuanyuan.zhang@cs.ucl.ac.uk

1. Basic Value/Cost Trade-off





15 Stakeholders
40 Requirements

100 Stakeholders
140 Requirements

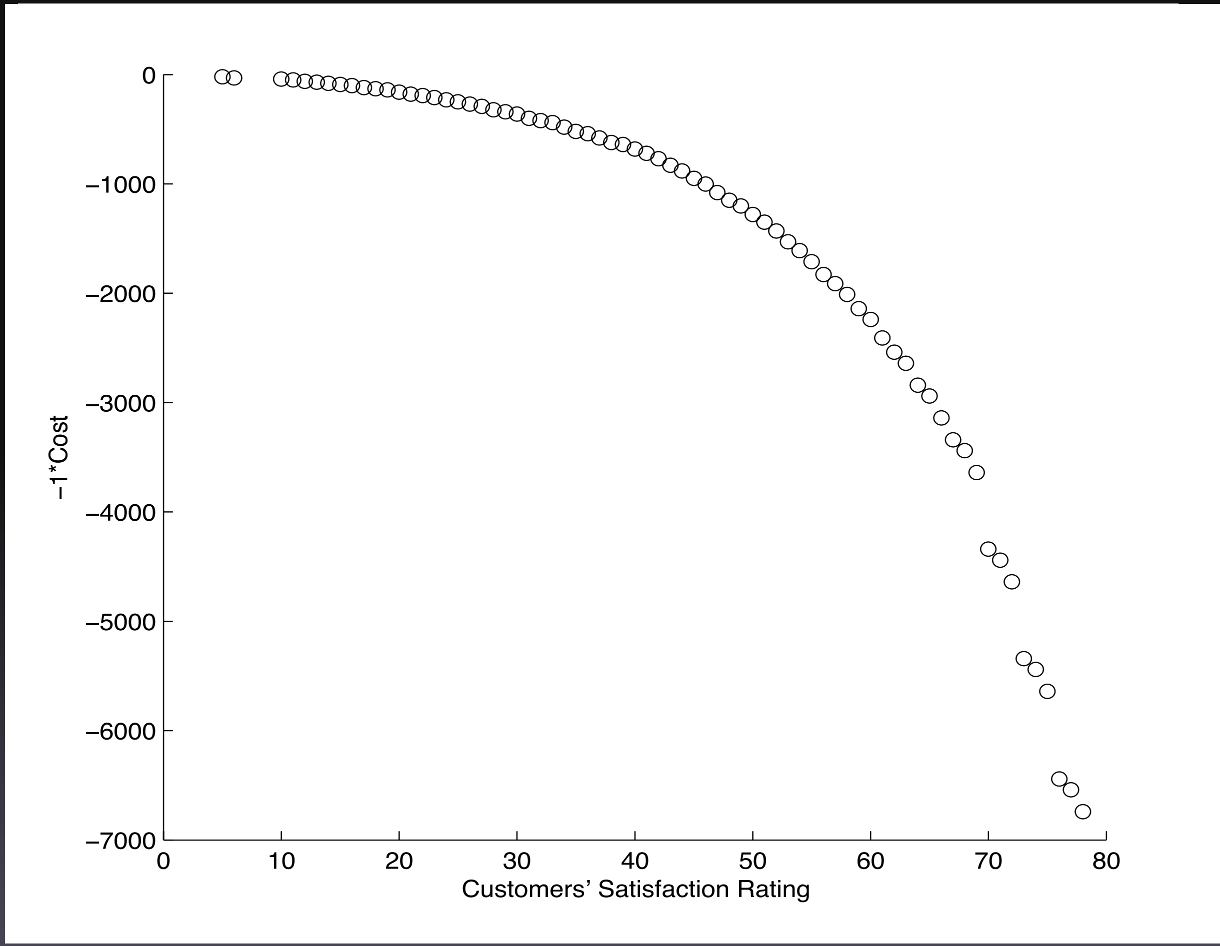
University College London yuanyuan.zhang@cs.ucl.ac.uk

1. Basic Value/Cost Trade-off

Motorola Data set

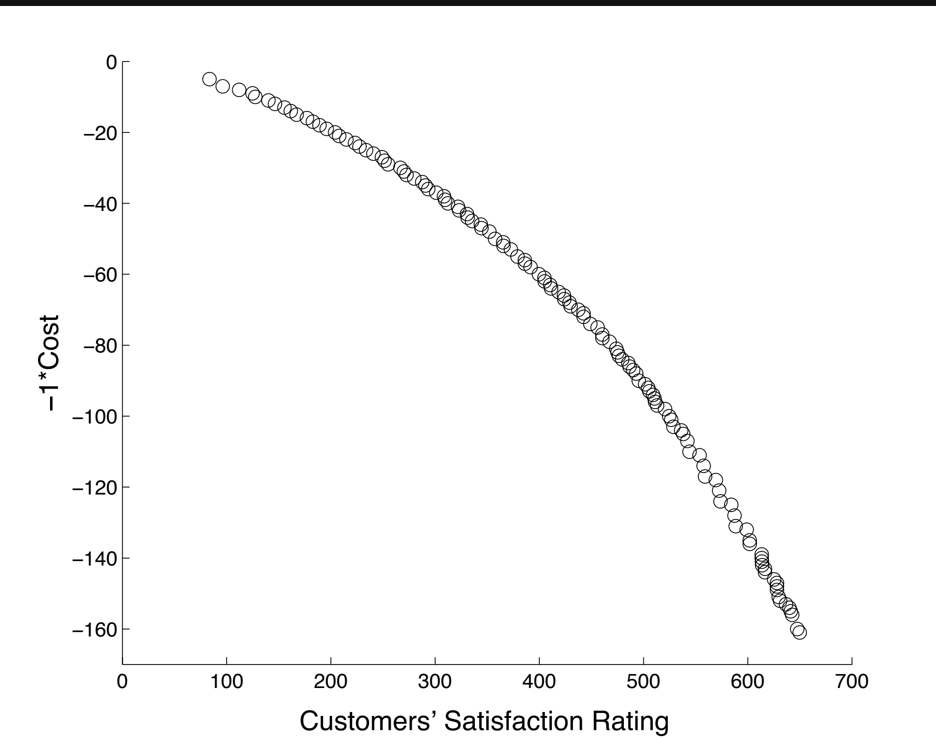
4 Stakeholders

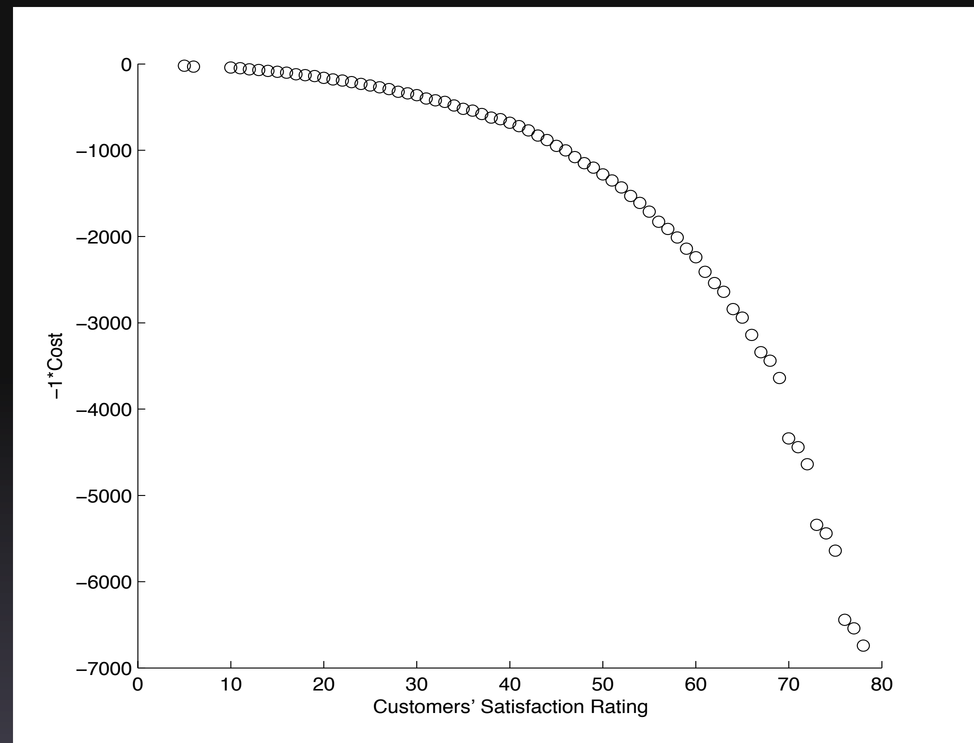
35 Requirements



University College London yuanyuan.zhang@cs.ucl.ac.uk

Results comparison



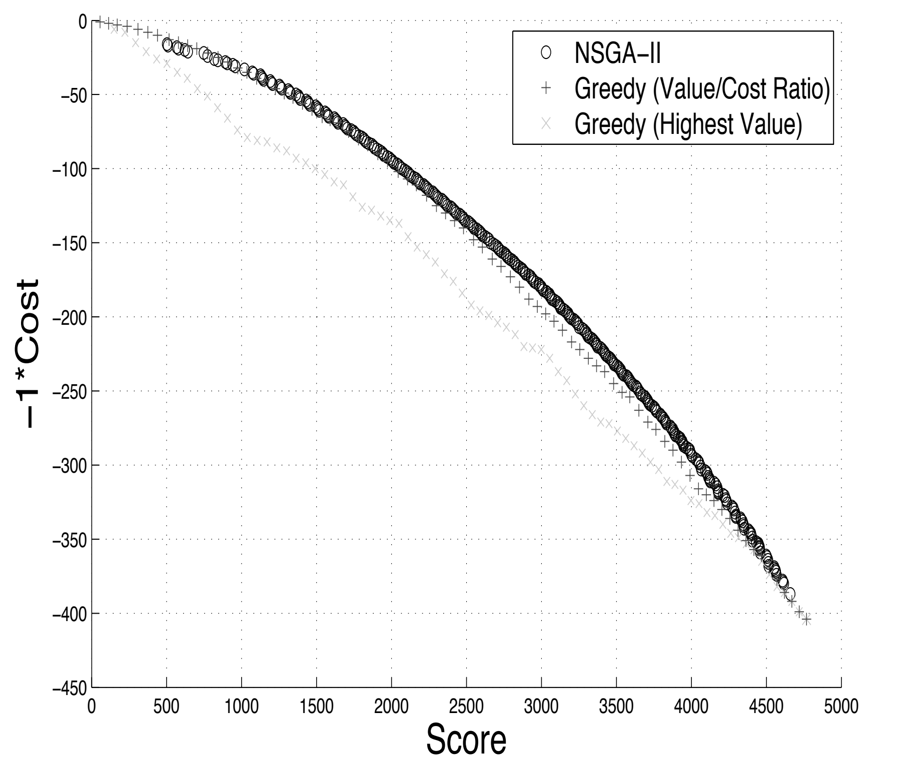


Synthetic

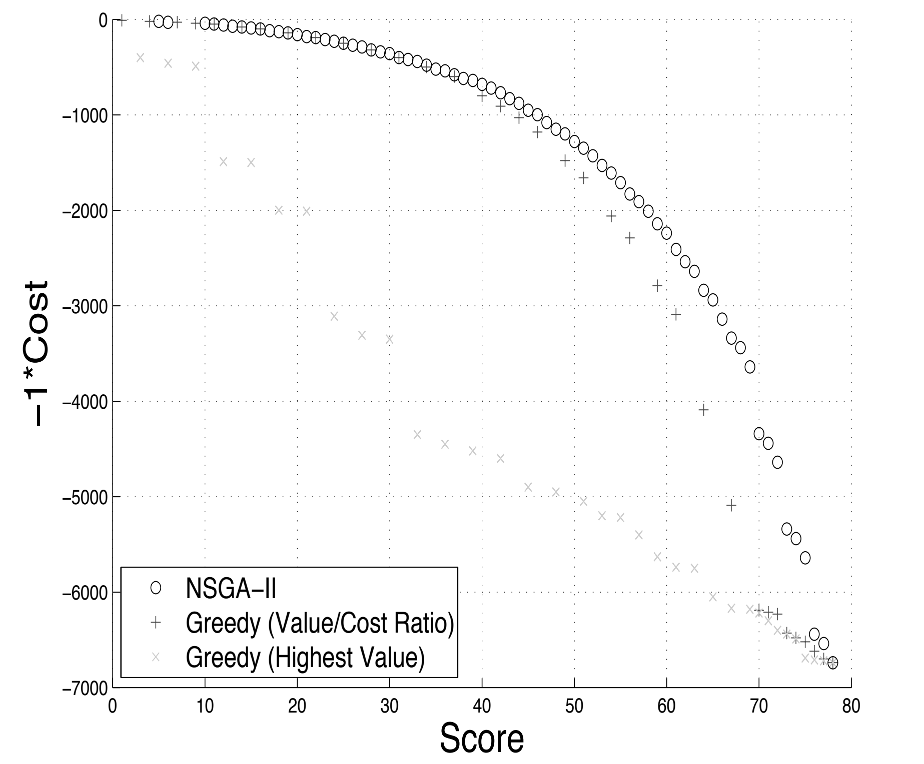
Motorola

University College London yuanyuan.zhang@cs.ucl.ac.uk

Search vs. Greedy



Synthetic



Motorola

2. Today/Future Importance Analysis

To provide robust solutions not only in the context of present conditions but also in response to those future changes that can be anticipated

Maximise

$$f_1\left(\vec{x}\right) = \sum_{i=1}^{n} score_{i,today} \cdot x_i$$

Maximise

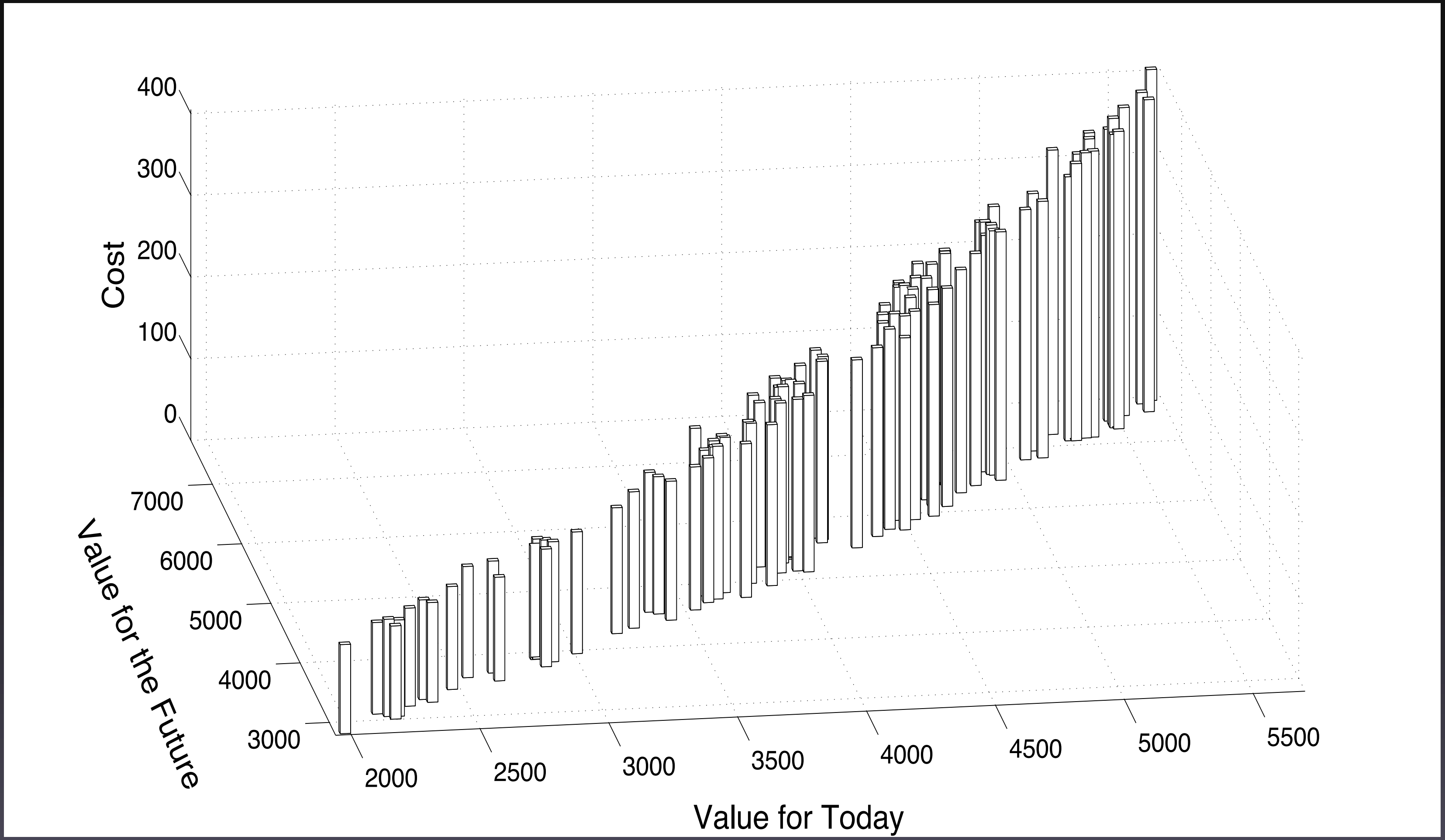
$$f_1\left(\vec{x}\right) = \sum_{i=1}^{n} score_{i,future} \cdot x_i$$

Minimise

$$f_3\left(\vec{x}\right) = \sum_{i=1}^{n} cost_i \cdot x_i$$

University College London yuanyuan.zhang@cs.ucl.ac.uk

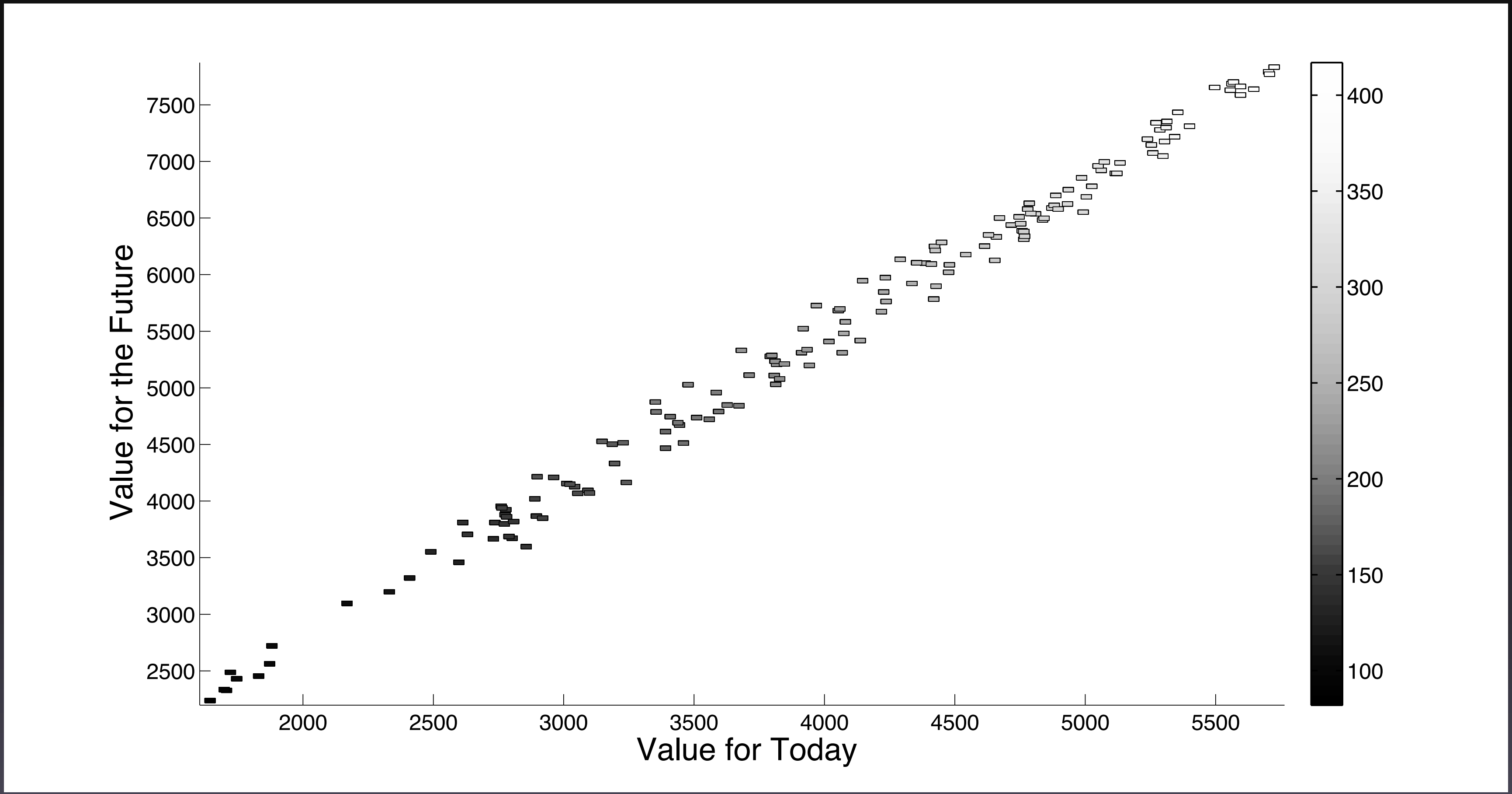
2. Today/Future Importance Analysis



Results from Ericsson Data Sets: 124 Requirements, 14 Customers

University College London yuanyuan.zhang@cs.ucl.ac.uk

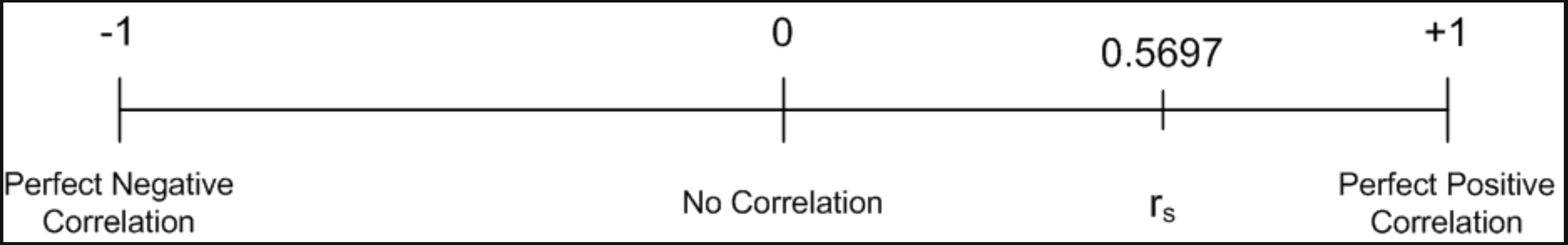
2. Today/Future Importance Analysis



Projection onto the X-Y Plane

University College London yuanyuan.zhang@cs.ucl.ac.uk

2. Today/Future Importance Analysis



Spearman's Rank Correlation Coefficient

This indicates a positive correlation between the value for today and value for the future.

University College London yuanyuan.zhang@cs.ucl.ac.uk

3. Multi-Stakeholder Tensioning Anslysis

The problem is to select a set of requirements that maximise the total value to each stakeholder, which is expressed as a percentage.

The model of fitness functions represented as:

Maximise
$$\frac{\sum_{i=1}^{n} value(r_i, c_j) \cdot x_i}{\sum_{r \in R_j} value(r, c_j)}$$

subject to
$$\sum_{i=1}^{n} cost_i \leq B, \qquad B > 0$$

Data Sets Used

1. Motorola Data Set:

   35 Requirements and 4 Stakeholders

2. Greer and Ruhe Data Set:

   20 Requirements and 5 Stakeholders

University College London yuanyuan.zhang@ucl.ac.uk

Data Sets Used

3. 27 Combination Levels of Random Data Sets:

the No. of requirements

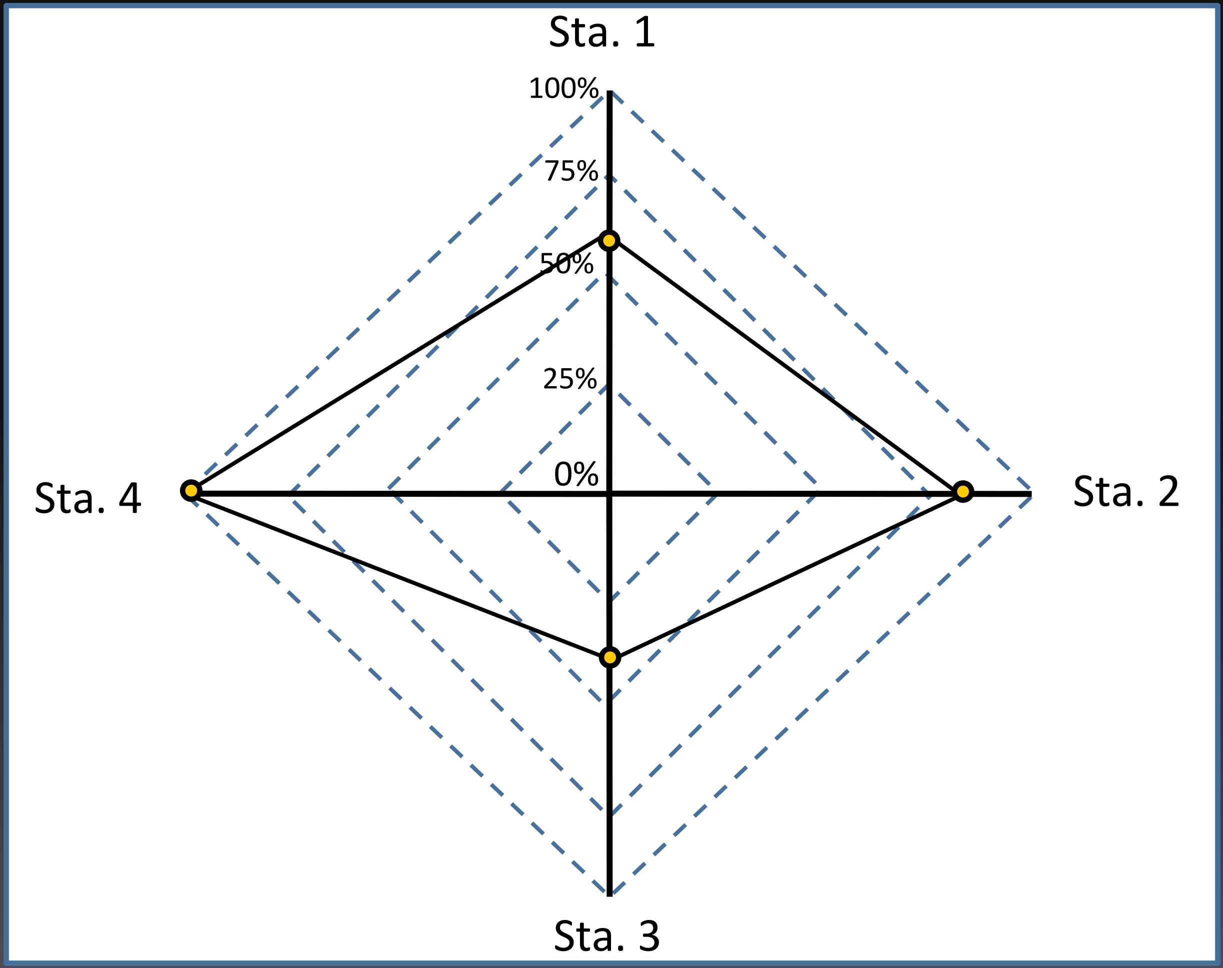
the No. of stakeholders

the density of the stakeholder-requirement matrix

	R_{small}	R_{medium}	R_{large}
C_{small}	$\boldsymbol{C_s\ R_s\ D_{low}}$ $C_s\ R_s\ D_m$ $C_s\ R_s\ D_h$	$C_s\ R_m\ D_{low}$ $C_s\ R_m\ D_m$ $C_s\ R_m\ D_h$	$C_s\ R_l\ D_{low}$ $C_s\ R_l\ D_m$ $C_s\ R_l\ D_h$
C_{medium}	$C_m\ R_s\ D_{low}$ $C_m\ R_s\ D_m$ $C_m\ R_s\ D_h$	$C_m\ R_m\ D_{low}$ $\boldsymbol{C_m\ R_m\ D_m}$ $C_m\ R_m\ D_h$	$C_m\ R_l\ D_{low}$ $C_m\ R_l\ D_m$ $C_m\ R_l\ D_h$
C_{large}	$C_l\ R_s\ D_{low}$ $C_l\ R_s\ D_m$ $C_l\ R_s\ D_h$	$C_l\ R_m\ D_{low}$ $C_l\ R_m\ D_m$ $C_l\ R_m\ D_h$	$C_l\ R_l\ D_{low}$ $C_l\ R_l\ D_m$ $\boldsymbol{C_l\ R_l\ D_h}$

King's College London yuanyuan.zhang@kcl.ac.uk

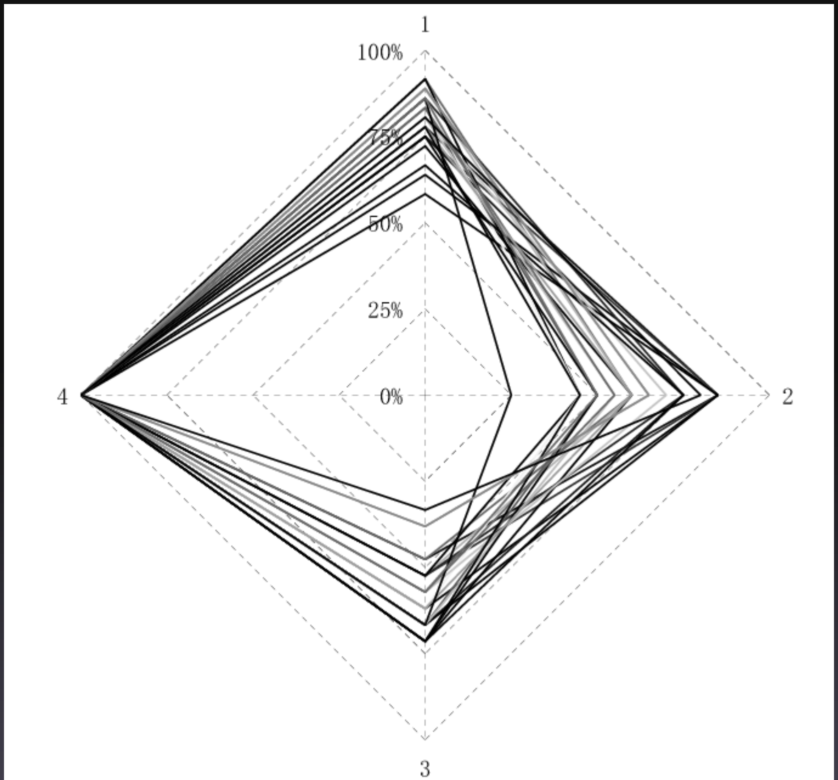
Multi-Stakeholder Kiviat Diagram

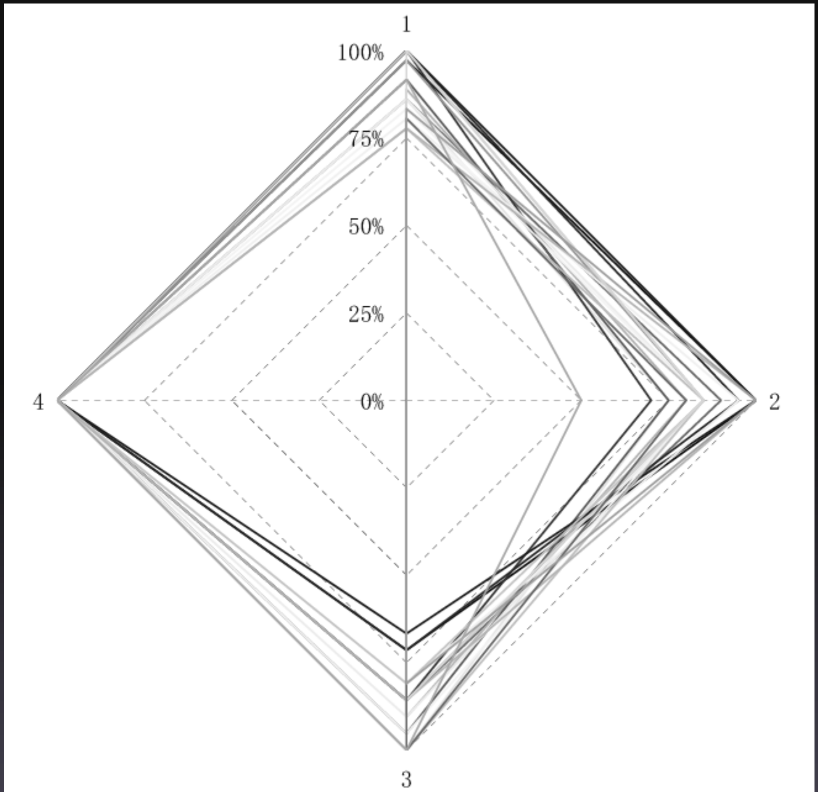


University College London yuanyuan.zhang@cs.ucl.ac.uk

Multi-Stakeholder Tensioning Analysis

Motorola data set

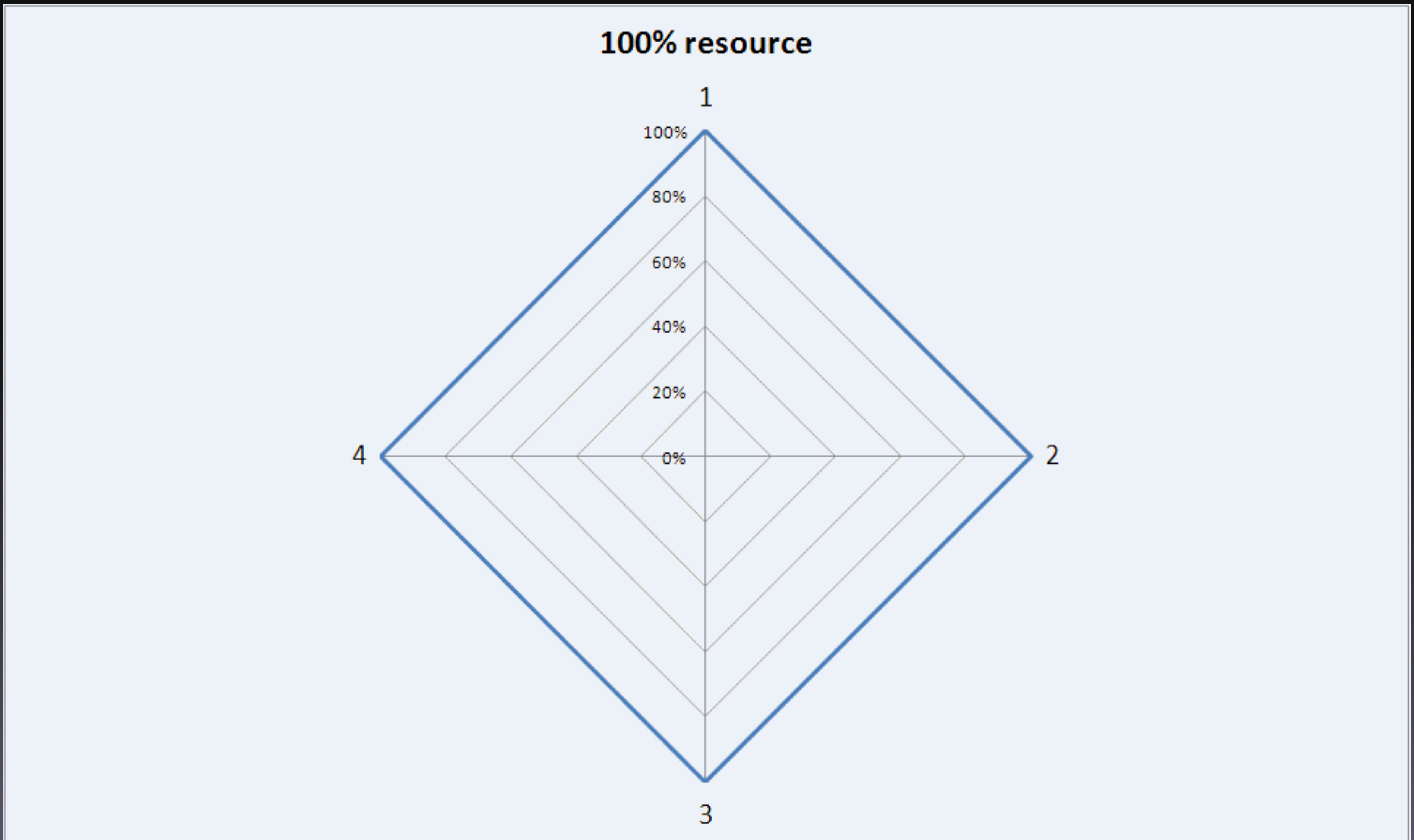




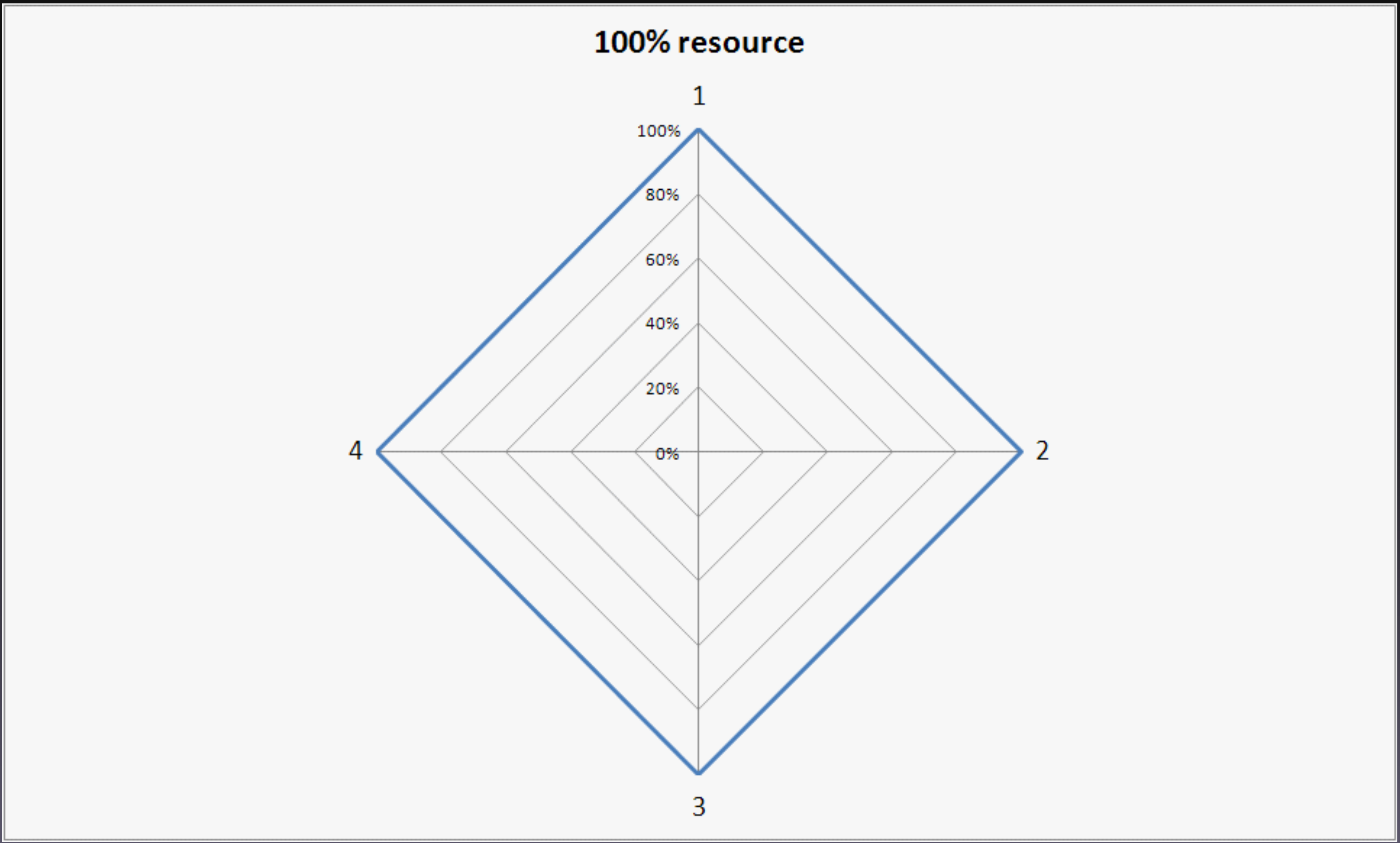
30% Budgetary Resource Constraint

70%

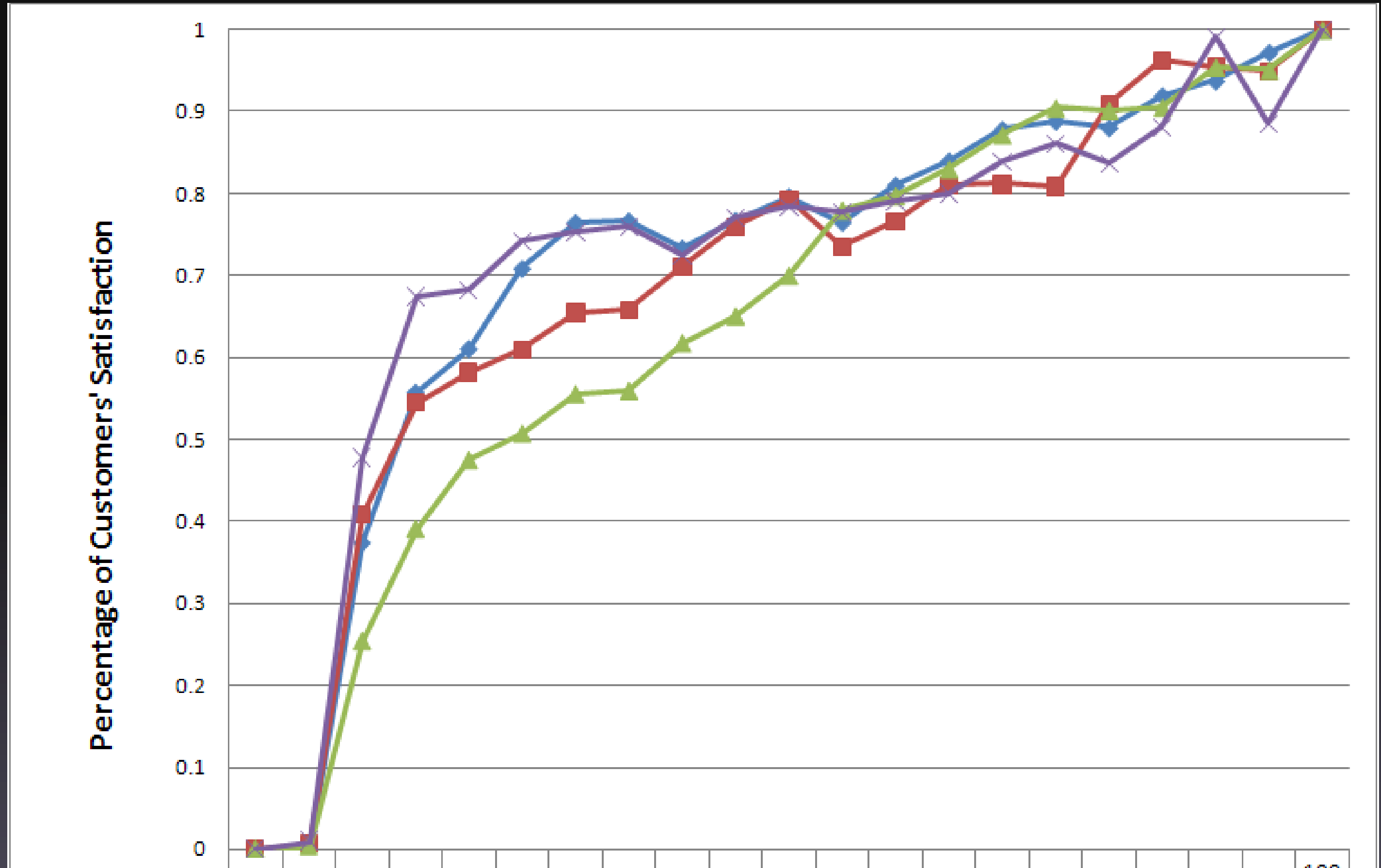
Solutions on the Pareto Front



Average Solutions



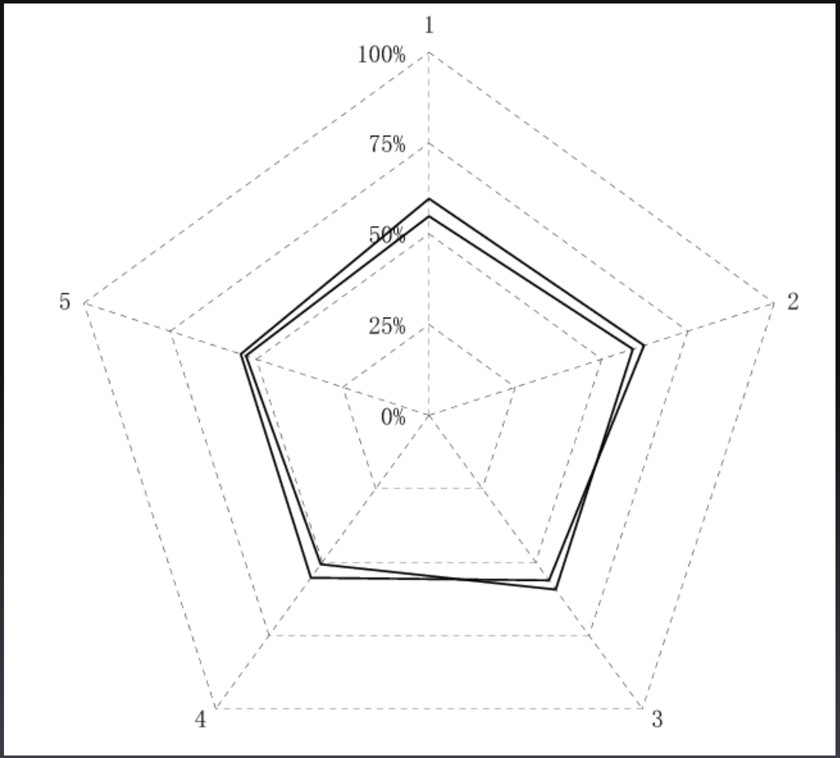
Tensions between the Stakeholders' Satisfaction for Different Budgetary Resource Constraints

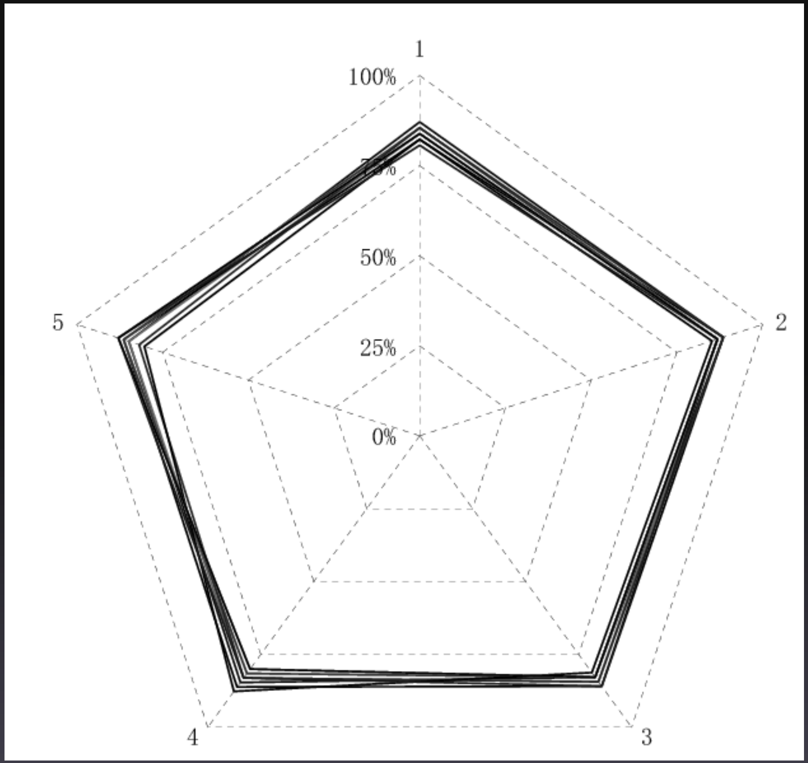


	0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%	75%	80%	85%	90%	95%	100 %
Customer 1	0	0	0.37	0.56	0.61	0.71	0.76	0.77	0.73	0.77	0.8	0.76	0.81	0.84	0.88	0.89	0.88	0.92	0.94	0.97	1
Customer 2	0	0.01	0.41	0.54	0.58	0.61	0.65	0.66	0.71	0.76	0.79	0.74	0.77	0.81	0.81	0.81	0.91	0.96	0.95	0.95	1
Customer 3	0	0	0.25	0.39	0.47	0.51	0.56	0.56	0.62	0.65	0.7	0.78	0.8	0.83	0.87	0.9	0.9	0.91	0.95	0.95	1
Customer 4	0	0.01	0.48	0.67	0.68	0.74	0.75	0.76	0.73	0.77	0.78	0.78	0.79	0.8	0.84	0.86	0.84	0.88	0.99	0.89	1

Multi-Stakeholder Tensioning Analysis

Greer and Ruhe data set





30% Budgetary Resource Constraint 70%

University College London yuanyuan.zhang@cs.ucl.ac.uk

Algorithms' Performance

$$C = \frac{\sum_{i=1}^{N} d_i}{N}$$

Rank Order for Convergence

	winner	runner up	loser
Random Search	0%	0%	100%
Two-Archive	95.19%	4.81%	0%
NSGA-II	7.04%	92.96%	0%

$$P = \frac{num}{NUM}$$

Solutions on the Reference front

Random Search	Two-Archive	NSGA-II
2.68%	94.57%	38.25%

University College London yuanyuan.zhang@cs.ucl.ac.uk

Algorithms' Performance

1 . The diversity of the Two-archive algorithm is significant in most cases

2. The Two-archive and NSGA-II algorithms always have a better convergence than the Random Search

3. The Two-Archive algorithm outperforms NSGA-II and Random Search in terms of convergence in some case

University College London yuanyuan.zhang@cs.ucl.ac.uk

4. Multi-Stakeholder Fairness Analysis

Fairness on Absolute *number* of fulfilled requirements:

Maximise $\overline{NA}$

Minimise $\sigma\left(NA\right)$

Fairness on absolute *value* of fulfilled requirements:

Maximise $\overline{VA}$

Minimise $\sigma\left(VA\right)$

where $$VA_j = \sum_{i=1}^{n} value\left(r_i, c_j\right) \cdot x_i$$

4. Multi-Stakeholder Fairness Analysis

Fairness on the percentage of value and cost of fulfiled requirements:

Minimise $\sigma(Cost\_C)$

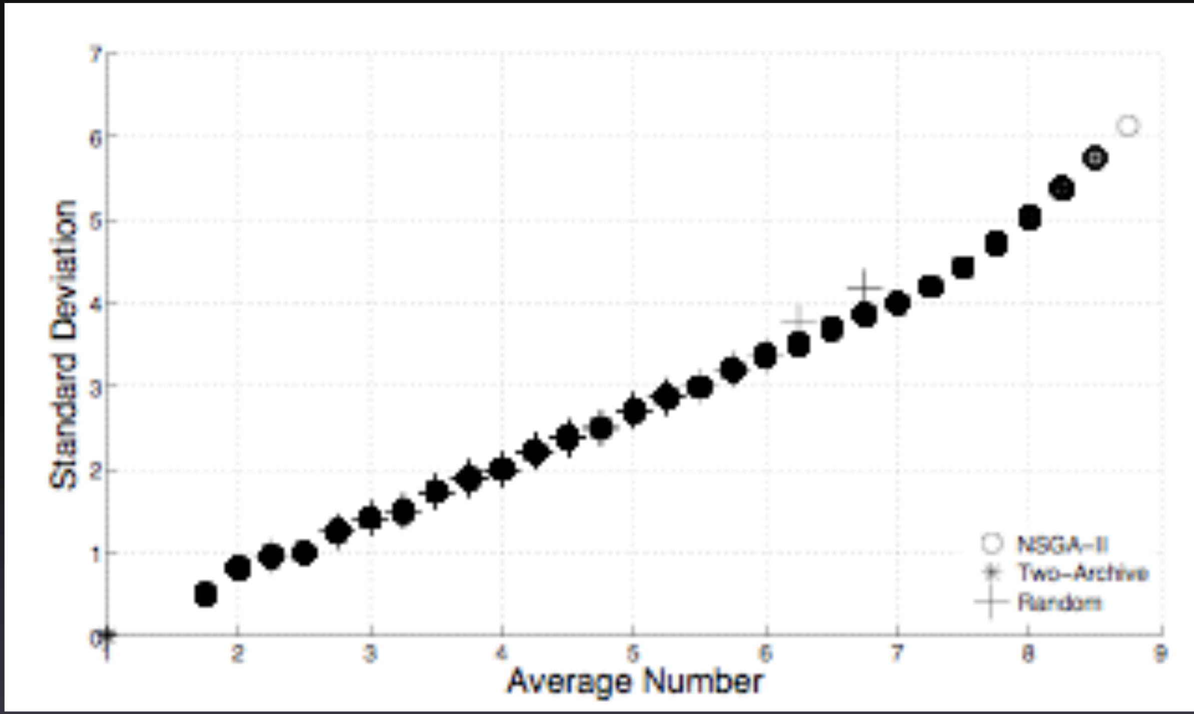
Maximise $\overline{VP}$

Minimise $\sigma(VP)$

Minimise $\sum_{i=1}^{n} cost_i \cdot x_i$

University College London yuanyuan.zhang@cs.ucl.ac.uk

4. Multi-Stakeholder Fairness Analysis

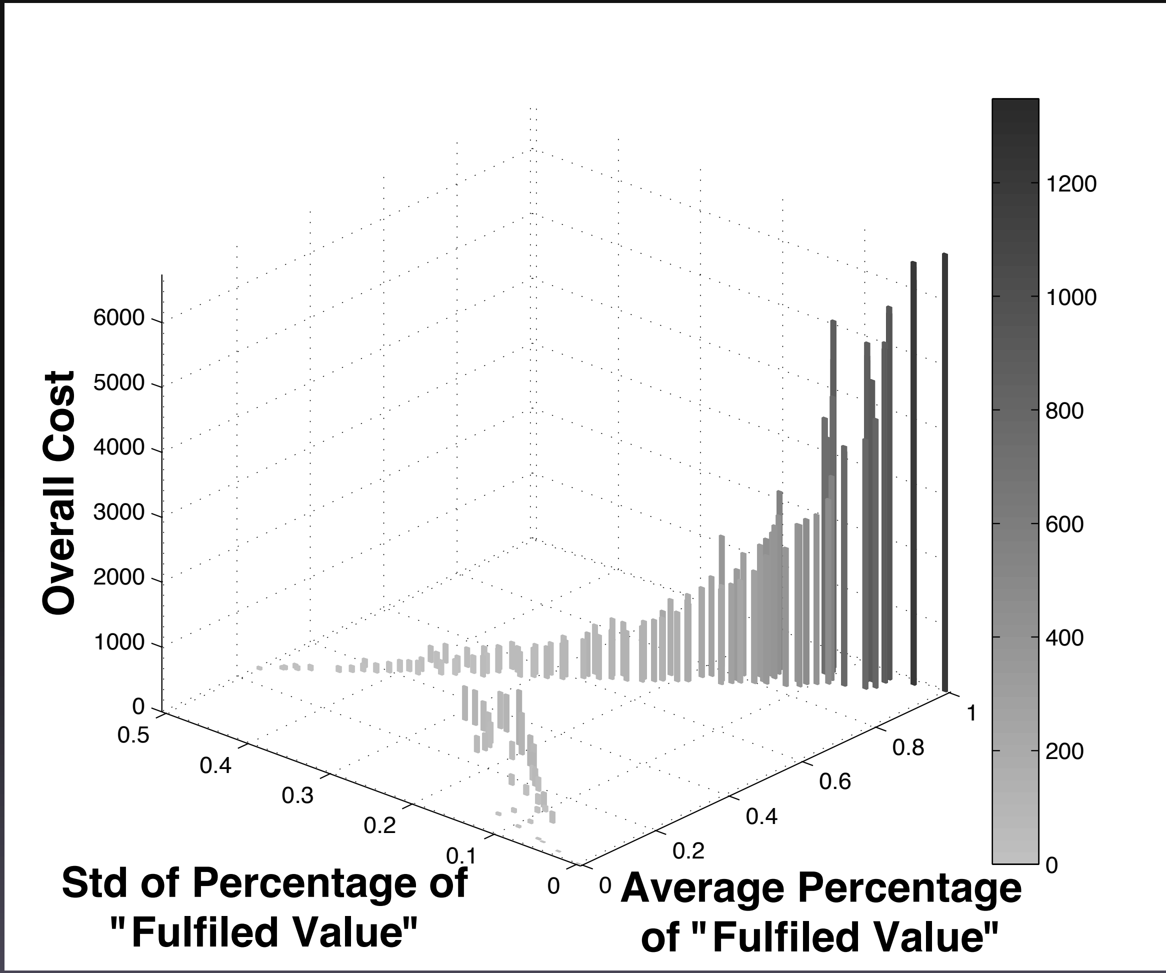


University College London yuanyuan.zhang@cs.ucl.ac.uk

4. Multi-Stakeholder Fairness Analysis

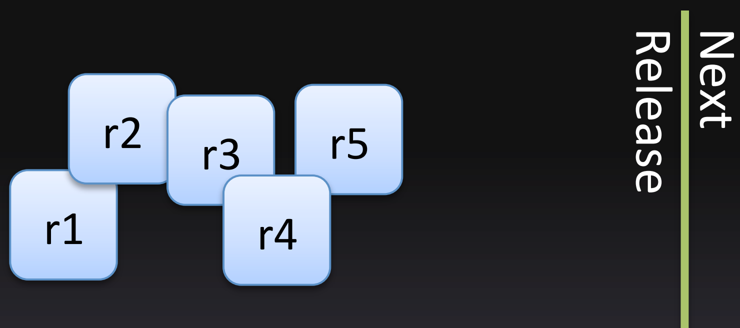
Motorola Data Set

Fairness on Percentage of Fulfilled Value



University College London yuanyuan.zhang@cs.ucl.ac.uk

5. Requirements Interaction Management (RIM)



Requirements

And

Or

Precedence

Value-related

Cost-related

University College London yuanyuan.zhang@cs.ucl.ac.uk

And

Given requirement r_i is selected,
then requirement r_j has to be chosen.

Define an equivalence relation ξ on the requirements array R such that $r(i,j) \in \xi$

University College London yuanyuan.zhang@cs.ucl.ac.uk

Or

Requirements r_i and r_j are conflicting to each other, only one of r_i , r_j can be selected.

Define an equivalence relation φ on the requirements array R such that $r(i, j) \in \varphi$

University College London yuanyuan.zhang@cs.ucl.ac.uk

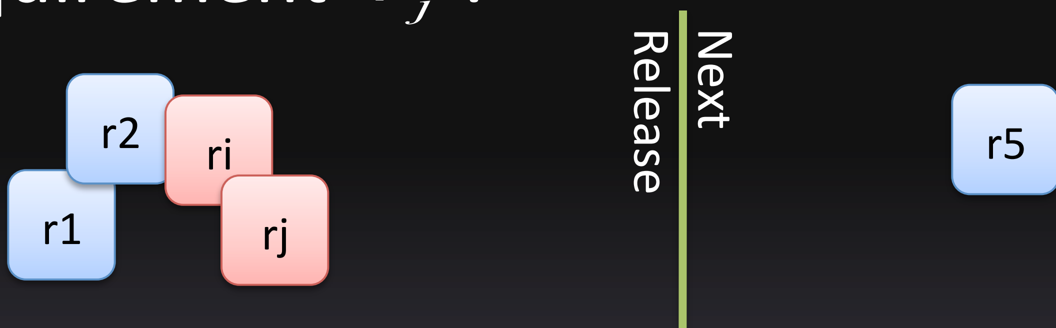
Precedence

Given requirement r_i has to be implemented before requirement r_j

Define an partial order χ on the requirements array R such that $r(i, j) \in \chi$

Cost-related

Given requirement r_i is selected, then this selection affects the cost of implementing requirement r_j .



Define an partial order ω on the requirements array R such that $r(i, j) \in \omega$

University College London yuanyuan.zhang@cs.ucl.ac.uk

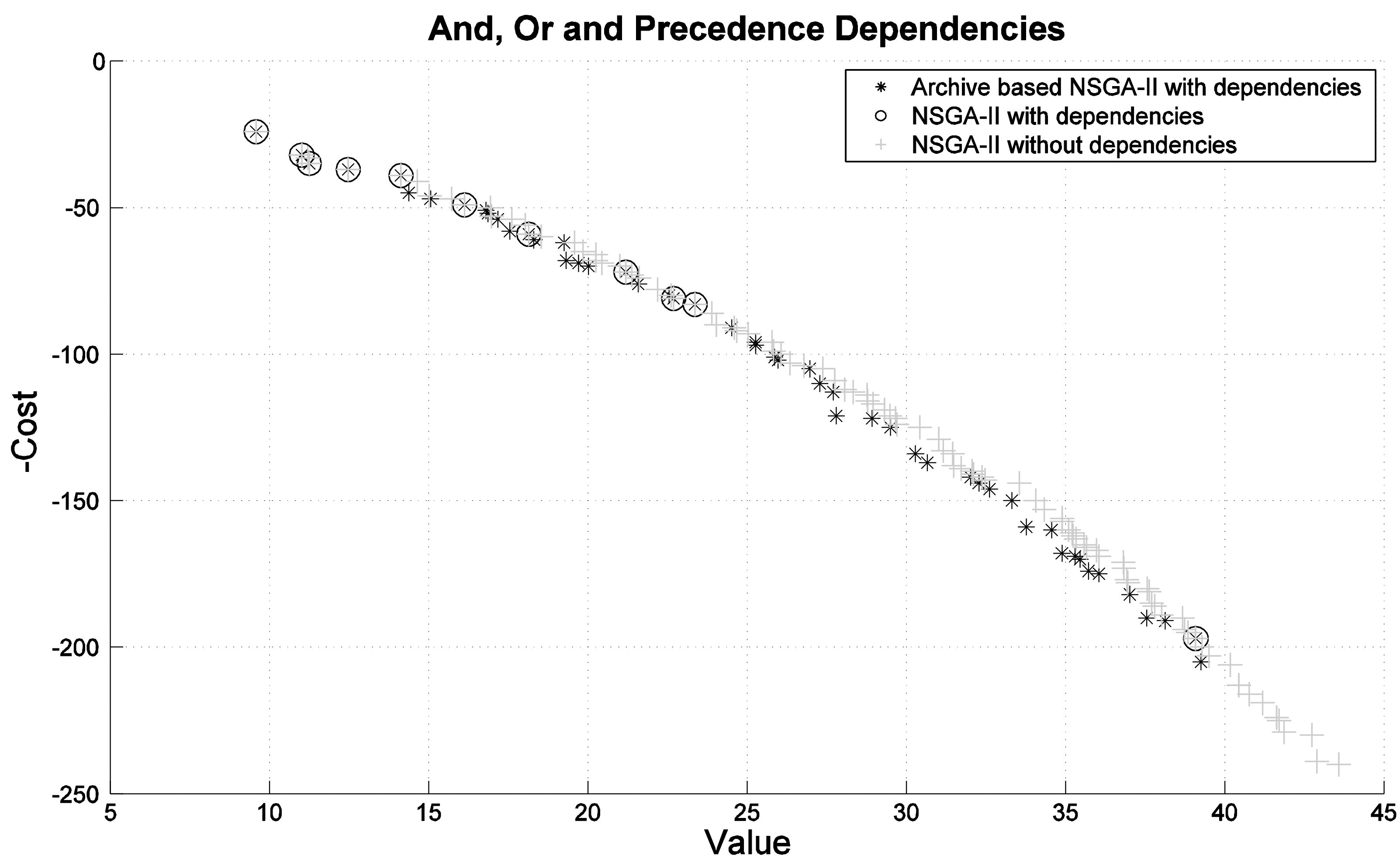
Value-related

Given requirement r_i is selected, then this selection affects the value of requirement r_j for the stakeholder.

Define an partial order ψ on the requirements array R such that $r(i, j) \in \psi$

University College London yuanyuan.zhang@cs.ucl.ac.uk

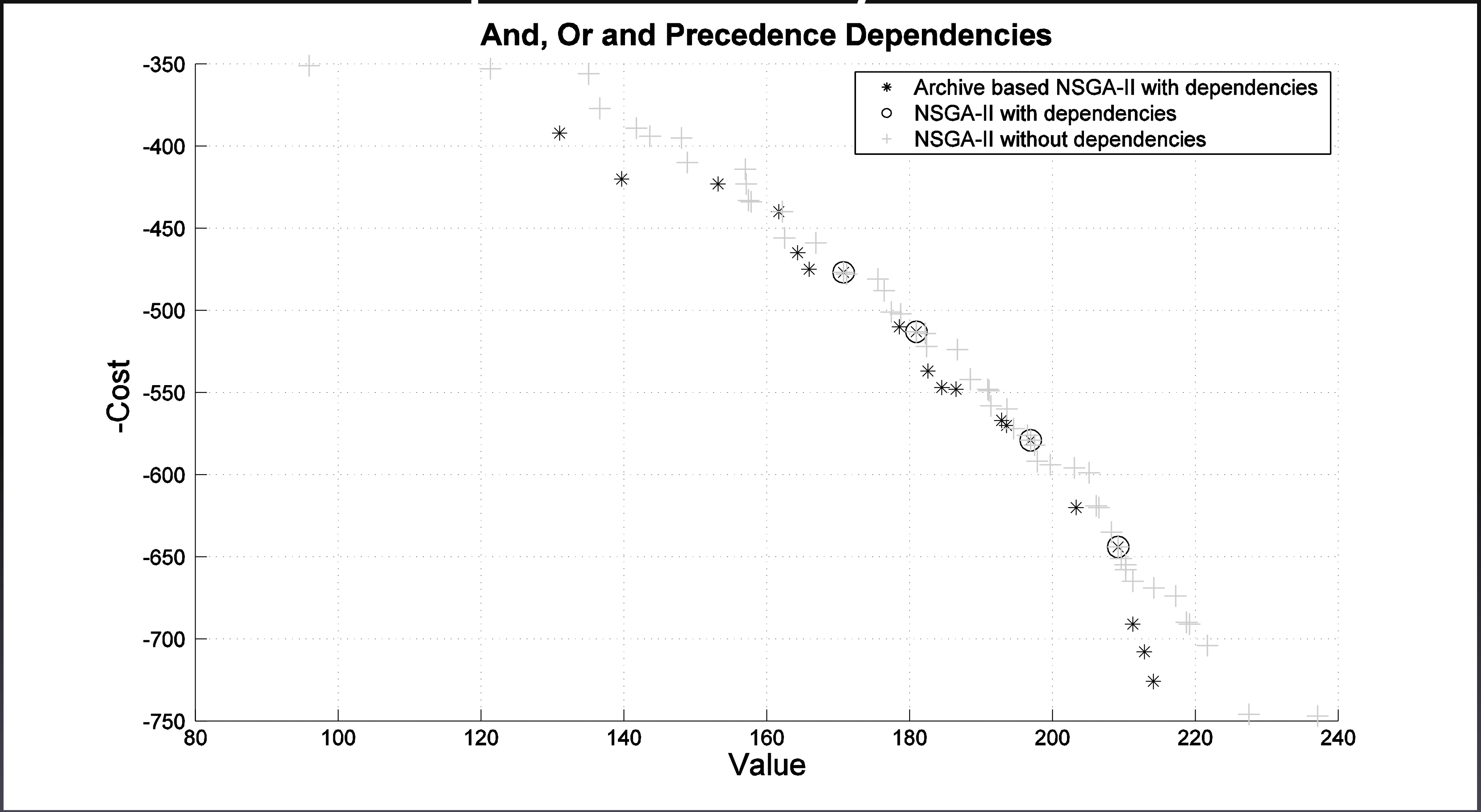
Empirical Study 5: RIM



34 Customers, 50 Requirements

University College London yuanyuan.zhang@cs.ucl.ac.uk

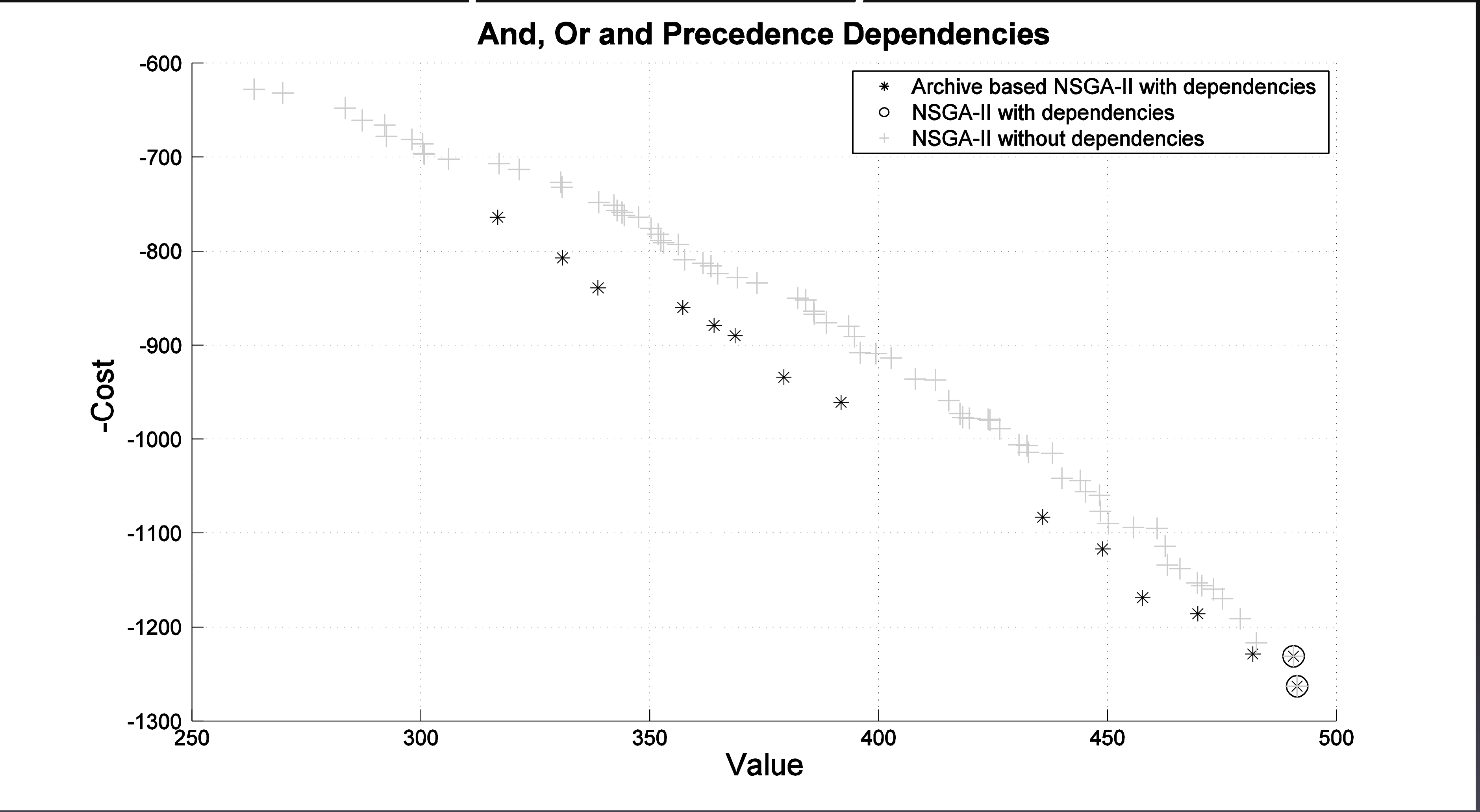
Empirical Study 5: RIM



4 Customers, 258 Requirements

University College London yuanyuan.zhang@cs.ucl.ac.uk

Empirical Study 5: RIM



21 Customers, 412 Requirements

University College London yuanyuan.zhang@cs.ucl.ac.uk

Conclusion

- Basic Value/Cost Trade-off analysis
- Today/Future Importance Analysis
- Multi-Stakeholder tension and fairness analysis
- Requirements Interaction Management

SBSE http://crestweb.cs.ucl.ac.uk/resources/sbse_repository/

UCL CREST

University College London   yuanyuan.zhang@cs.ucl.ac.uk