Using Unsupervised Learning for Data-Driven Procurement Demand Aggregation

Eran Shaham, Adam Westerski, Rajaraman Kanagasabai, Amudha Narayanan, Samuel Ong, Jiayu Wong, Manjeet Singh

Agency for Science, Technology and Research
SINGAPORE

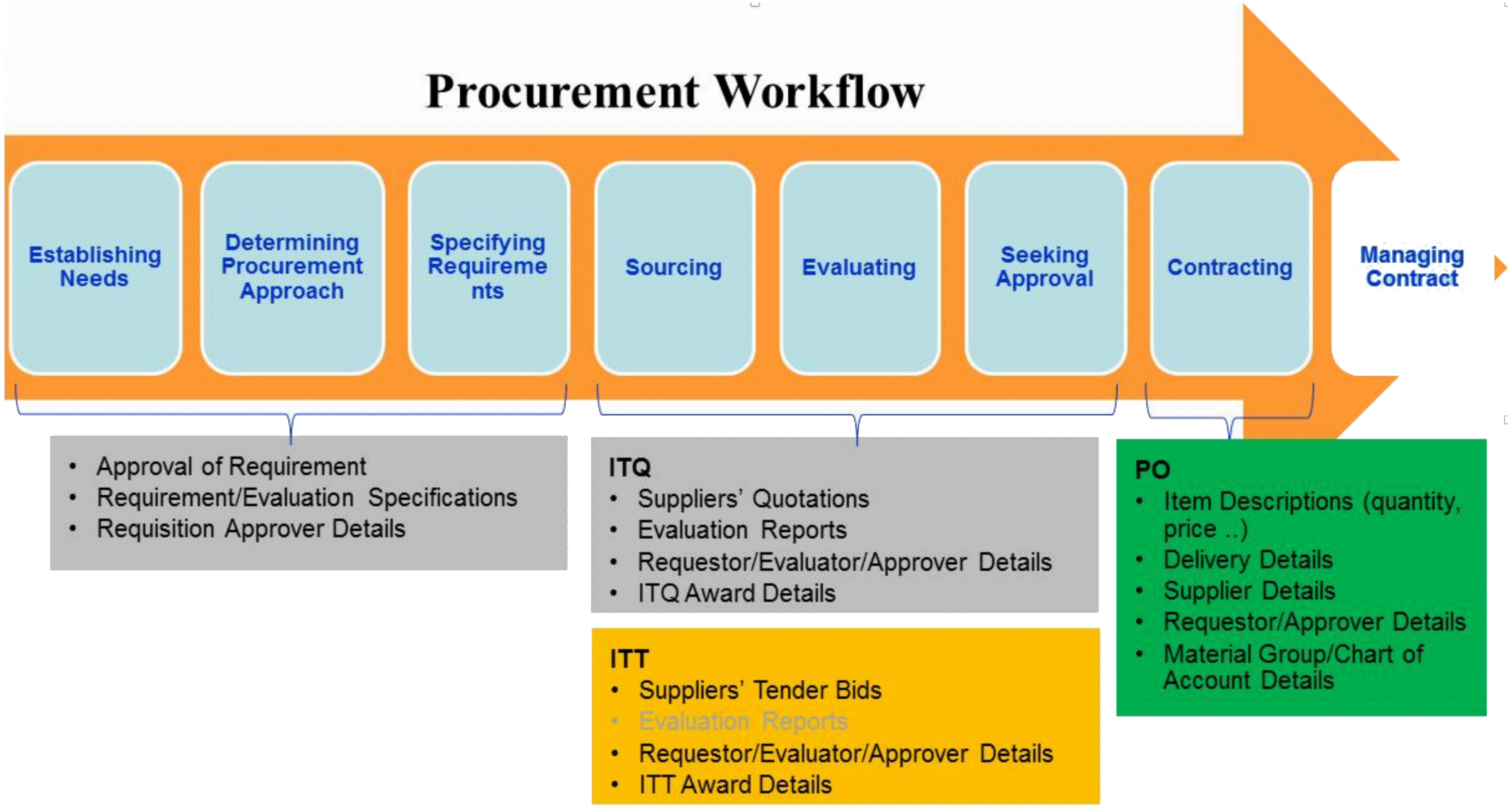
Collocated with AAAI-21
Conference on Artificial Intelligence

IAAI-21
Thirty-Third Annual Conference on Innovative Applications of Artificial Intelligence

background what is procurement ?

PROCUREMENT

- **Acquisition** of variety of goods or services from an outside external sources
- **Form:** orders, transactions, tenders, quotations
- **Data:** dates, values, quantities, requesters, vendors …



background A*STAR Procurement Office - the client (1)

A*STAR

- Large governmental agency in Singapore (>5k employees)
- Dealing with research in multiple areas, divided into Research Institutes
- 10k-magnitude volume of purchases monthly / 100k yearly

A*STAR Procurement Office

- Oversight of procurement of all Research Institutes
- Optimise purchasing processes
- Detect lapses, inconsistencies, potential fraud

background A*STAR Procurement Office - the client (2)

A*PO - I2R collaboration on Procurement Analytics | 2014 - onwards

- Technologies
- Deployed systems
- Patents
- Research publications

COMPASS

Comprehensive **P**rocurement **A**nalytics **S**ystem

Procurement
Fraud Detection

PACE

Westerski et al. 2017

Procurement
Demand Forecasting

PACE

Westerski et al. 2015

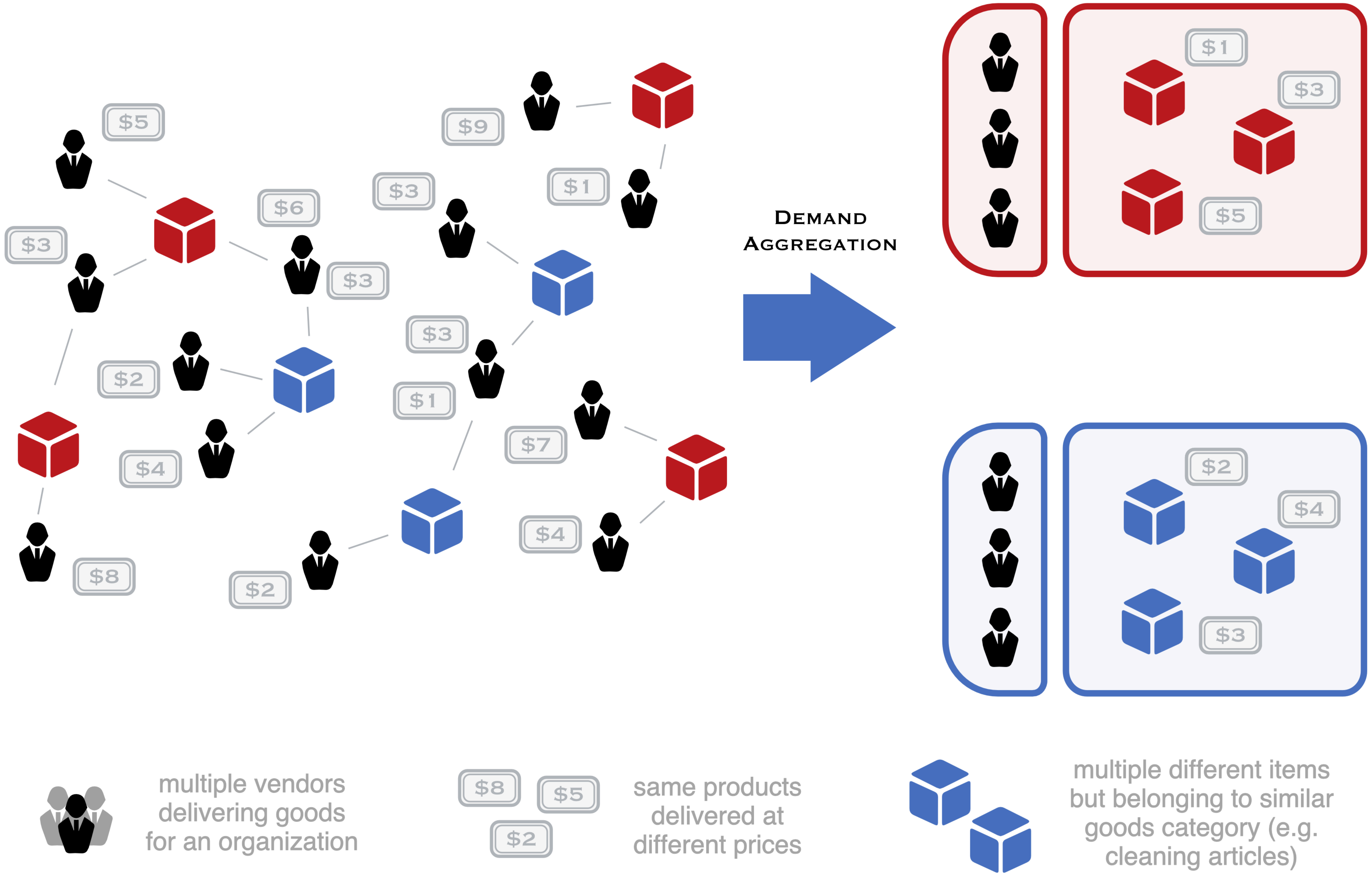
Procurement
Demand Aggregation

DART

THIS TALK

background demand aggregation - the problem

How does Demand Aggregation work in an organization?



Understand the process

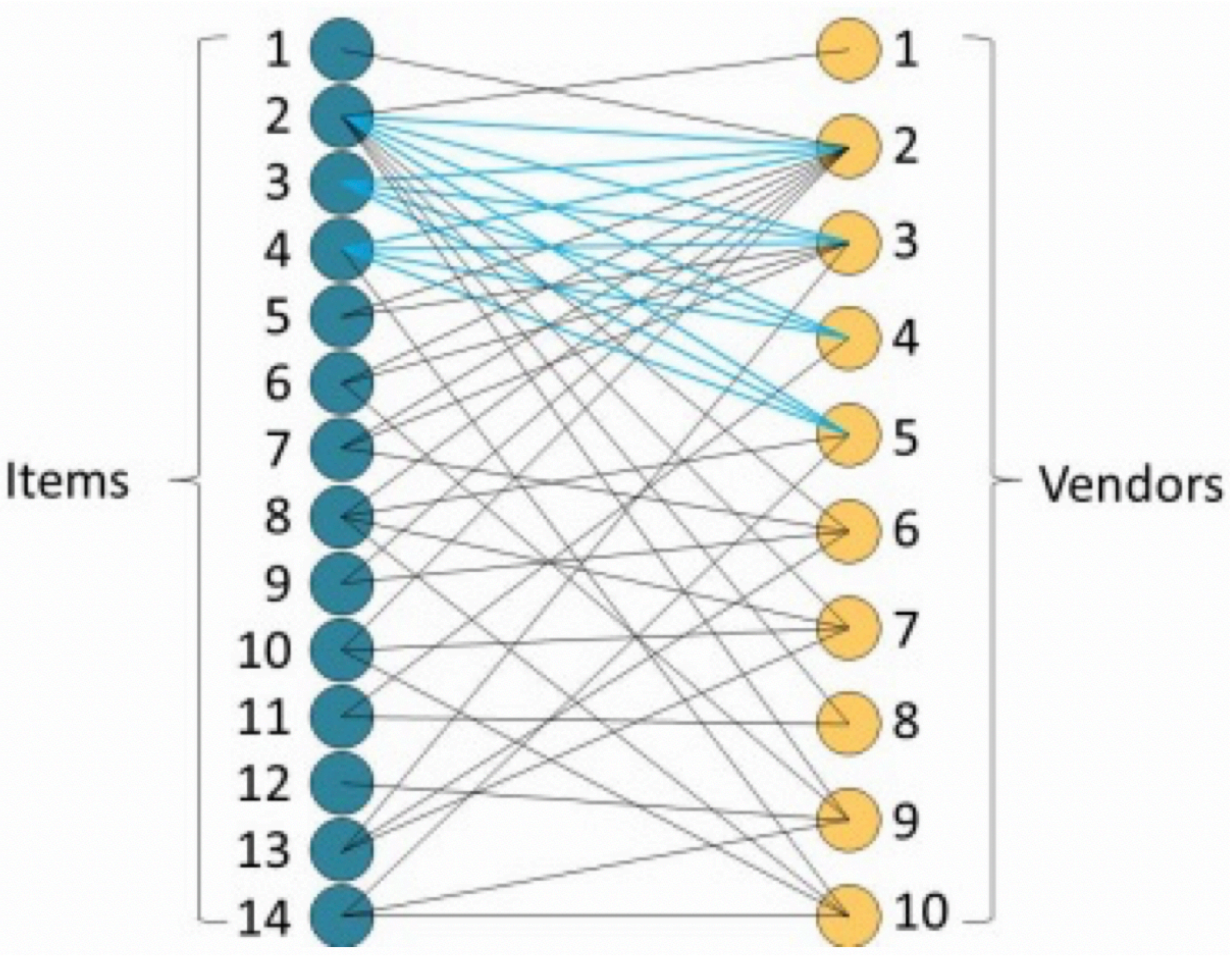
- **Collate and combine** requirements of multiple buys
- **Contract** suppliers based on combined demand
- **Standardise and establish** best-buy strategy

Know the problem

- **Fractured purchasing process** done individually by departments and units results in similar buys for different prices from multiple suppliers
- **Big amount** of data makes it difficult to understand best options for cost savings
- **Complex network** of suppliers, items they provide, prices and demand over time makes it difficult to solve this multi-variate problem with standard approaches

solution bi-clique clustering

- **<u>frame demand aggregation problem as bi-clique clustering problem</u>**
  - Map purchase orders into a bipartite graph - items vs. vendors
  - Bi-clique - every vendor connected to every item
  - Find max bi-cliques = biggest potential Demand Aggregation (DA) patterns

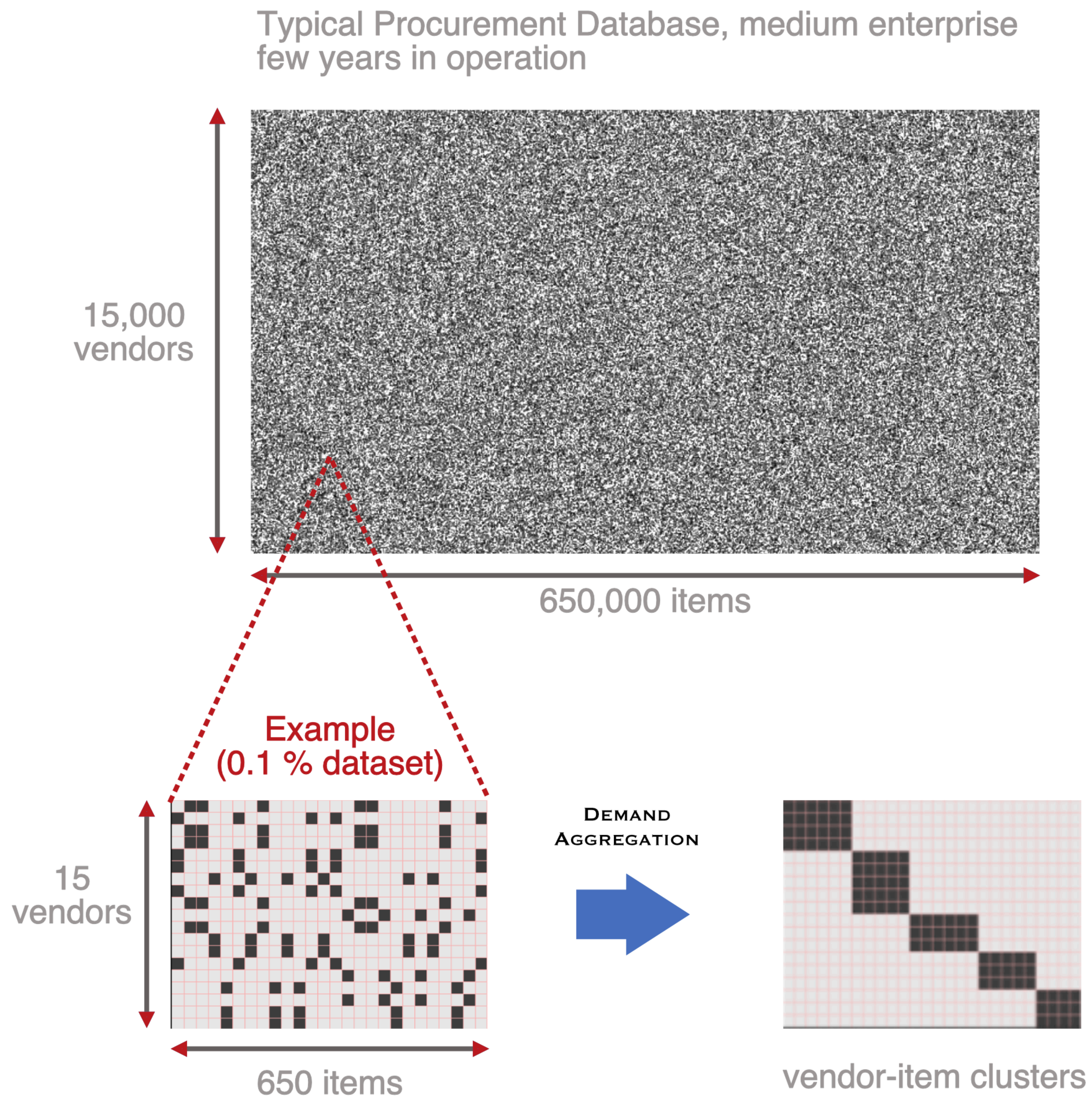


Sample procurement bipartite graph
(blue coloured edges denote 3x4 bi-clique)

solution BSC algorithm

- **Practical problem**
  - Finding maximum edge bi-clique is an NP-complete problem
  - Multiple publications with different algorithmic approaches
    - iMBEA [Zhang et al., 2014]
    - LCM-MBC [Li et al., 2007]
  - On small dataset not a big issue but for our "medium" sized data -> infeasible
  - Our approach: Monte Carlo algorithm (polynomial): doesn't give full list but optimal solution is good enough

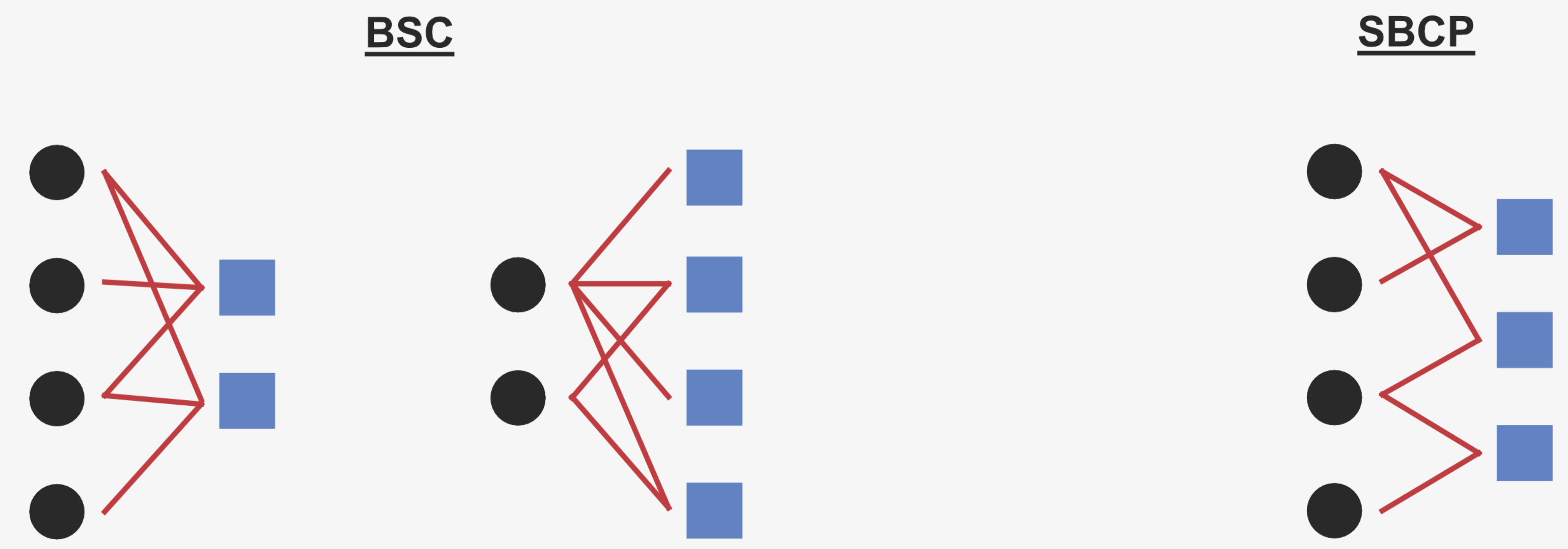
Optimality proof → Shaham et al. 2016



solution SBCP algorithm

Subspace Bi-clique Clustering for Procurement (SBCP) - procurement domain specific adjustments

- (1) **Practical BSC algorithm modifications**
  - Different bi-clique expansion strategy to give more “interesting” maximal bi-cliques (ie. more even amount of vendors vs items)

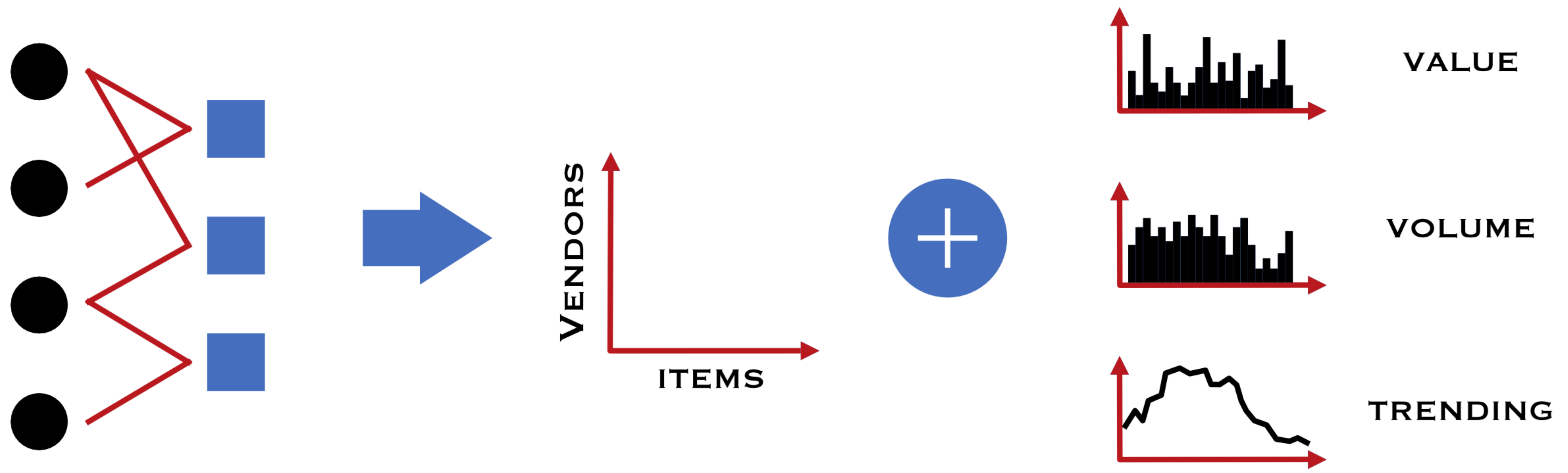


vendors >> items

vendors << items

vendors ~ items

- (2) **Post-processing filters**
  - Value constraints - check purchasing patterns with significant value only
  - Volume constraints - remove patterns with few purchase orders
  - Purchasing trend constraints - look for purchases that have potential to keep demand over years



multiple vendors deliver multiple items creating a graph of relationships

finding best DA is a max bi-clique clustering problem

multiple procurement specific conditions super-imposed

evaluation setup

- **Running SBCP algorithm over the A*PO procurement database**
  - **Input:** Using 3 most recent years (2014 - 2016); **271,219** items x **7319** vendors [**1,032,275** POs]
  - **Output:** Demand Aggregation patterns along with associated Purchase Orders
- **A*PO officers to assess quality of detected aggregation patterns:**
  - In comparison to past bulk tenders
  - Assessment for creating new bulk tenders
- **3 rounds of evaluation** (earlier mentioned SBCP improvements in between)

evaluation results

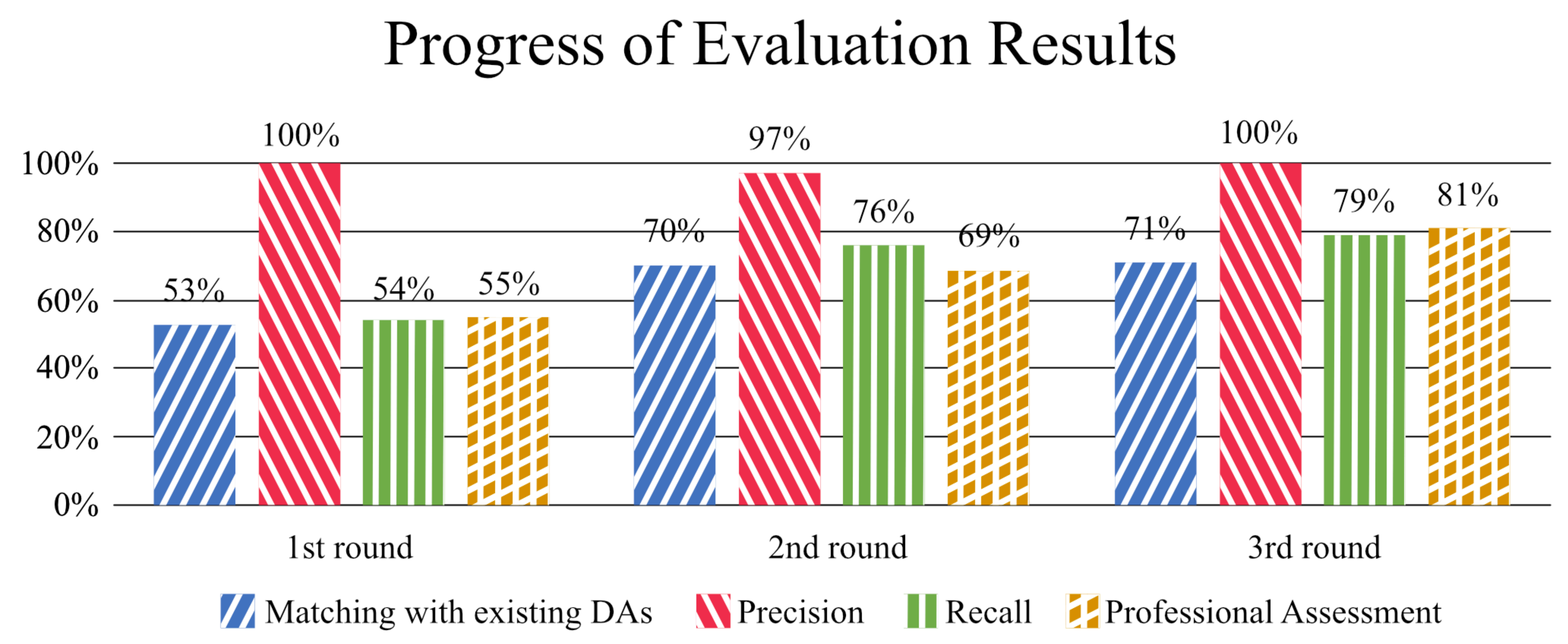
- Overall quality of DA patterns

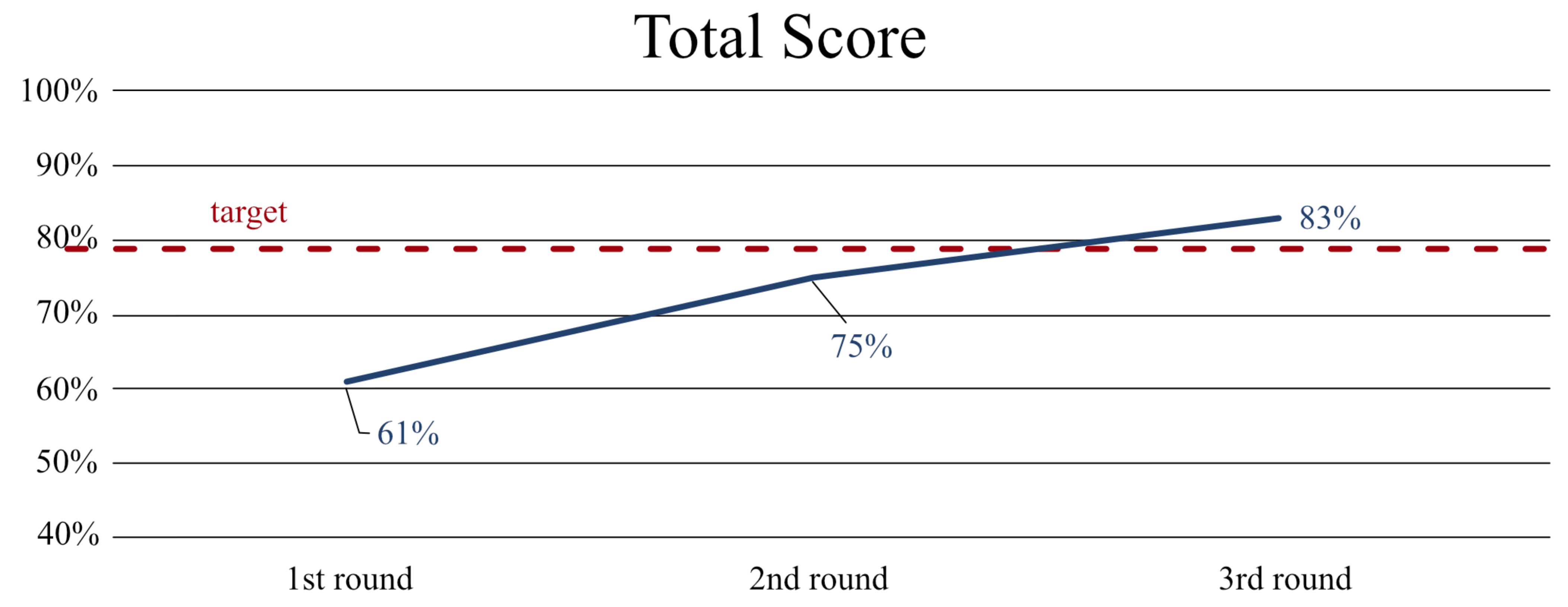
$$proifessional\ assessment = \frac{new\ valid\ DAs}{all\ new\ mined\ DAs}$$

$$matching\ existing\ DAs = \frac{detected\ past\ DAs}{total\ past\ DAs}$$

- Assessment of POs inside patterns

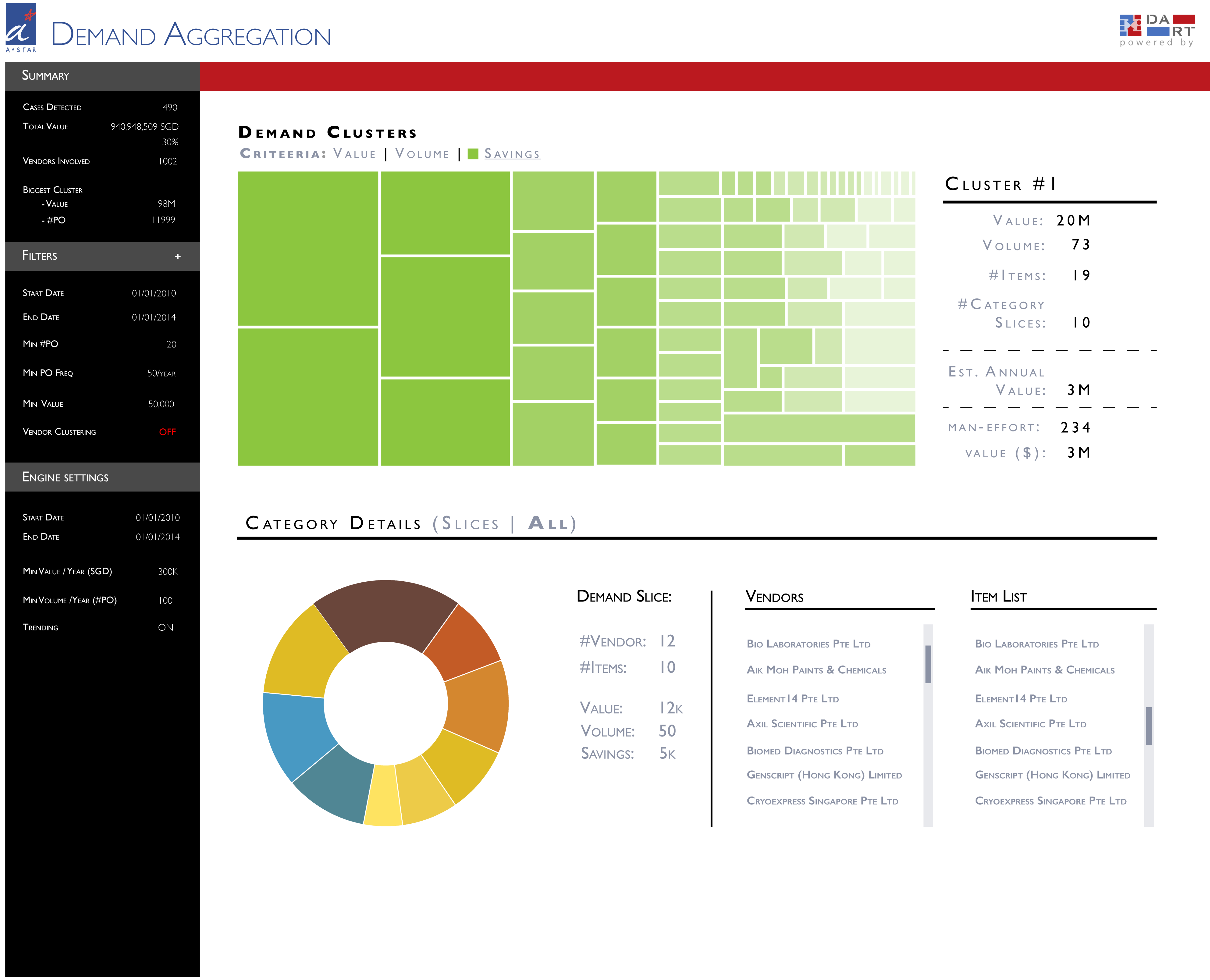
$$precision = \frac{valid\ POs}{all\ mined\ POs} \qquad recall = \frac{valid\ POs}{relevant\ POs}$$





deployment setup in production

- **Integration with A*PO workflow**
  - SBCP executed periodically as decision-support system for annual reports to management
    - Suggest new tenders
    - Update old tenders with new items/vendor



- **Input/output**
  - Taking a dump of procurement database
  - Analytics dashboard over the DA pattern list output

conclusions lessons learned

- **Complete list of maximal bi-cliques not necessary (and unwanted)**
  - Too many aggregation patterns counterproductive for end users
  - Mine and prioritise “useful” patterns (as per business requirements)
- **Data Science** metrics vs. **Business** metics
  - People who don’t deal with computer science rarely understand what is “precision” / “recall”
  - Listen to end user what they want and assess quality from their perspective (and stick to established metrics behind the scenes only)
  - “Interesting” , “non-obvious” patterns vs. “good quality”
- **Clients often doesn’t know what they want** (multiple times during the project)
  - Redefining meaning of valid demand aggregation patterns
  - Be prepared to adjust the algorithm many times
  - Early engagement with client helps
  - Long way from reaching (1) metric goals to (2) deployment to (3) active use by client

future work

- **Algorithm improvements**
  - Finding Quasi bi-cliques (ie. not every vendor selling every item, allow some freedom)
  - Assessment of DA (maximal bi-clique) quality
- **Going outside A*STAR and commercialising the technology**
  - Method and Apparatus for Procurement Demand Aggregation. Patent [Shaham et al. 2019]
  - Startup to bring our DA technology to the market - **Semantist**