On Skyline Groups

Chengkai Li[1], Nan Zhang[2], Naeemul Hassan[1], Sundaresan Rajasekaran[2], Gautam Das[1,3]

[1]University of Texas at Arlington, [2]George Washington University, [3]Qatar Computing Research Institute

UNIVERSITY OF TEXAS ARLINGTON

GW UNIVERSITY WASHINGTON DC

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute

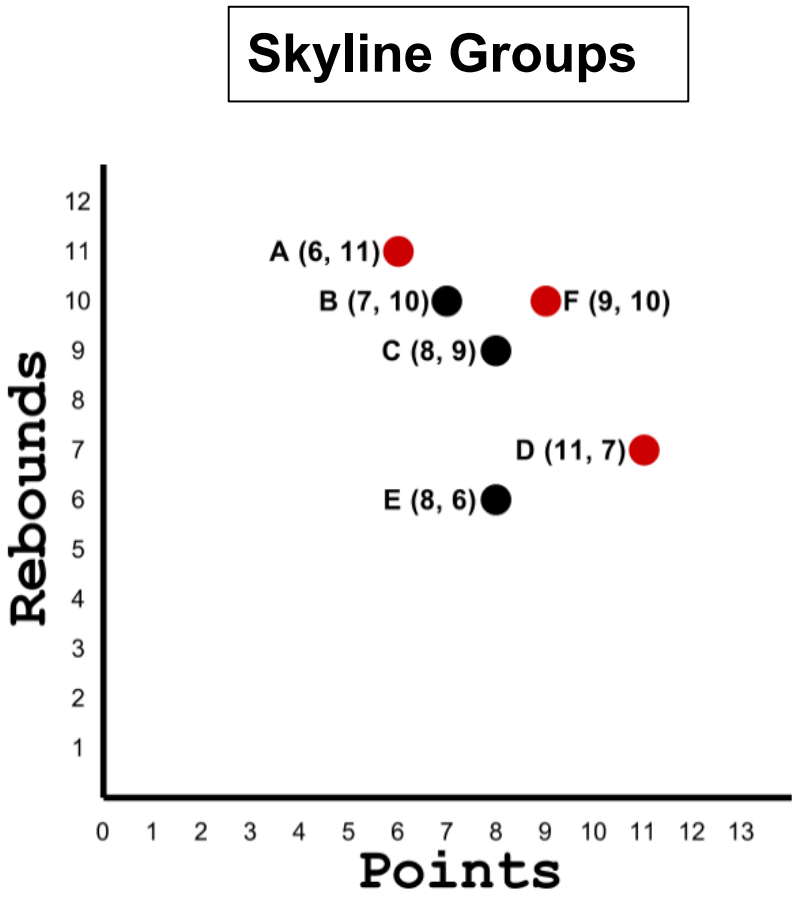
Motivating Example

Dream Team

	Points	Rebounds	Blocks
Michael Jordan	3	4	5
Lebron James	4	2	3
Kobe Bryant	4	5	3
SUM	11	11	11
MIN	3	2	3
MAX	4	5	5

Another Team

SUM	12	11	11



Applications

- Find a group of experts
  - Software Development

	Testing	Coding	Design
Applicant_1	10	20	15
Applicant_2	8	15	16
Applicant_3	11	18	15

  - Review a Paper

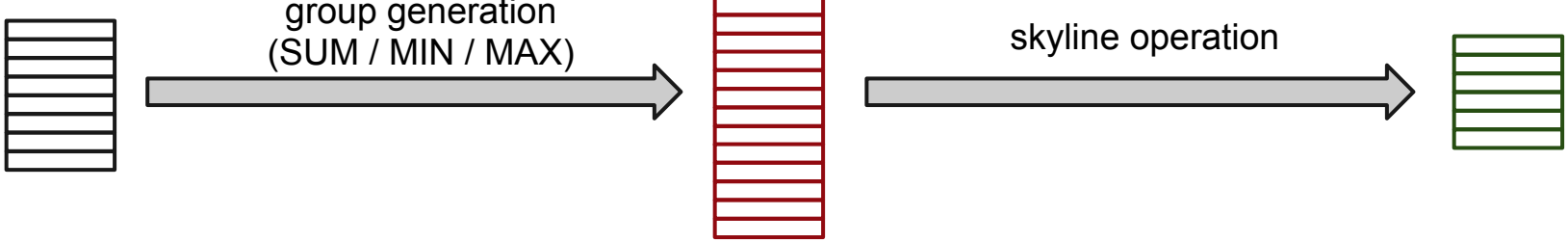
	Database	Security	Algorithm
Reviewer_1	41	35	23
Reviewer_2	45	31	34

Problem & Challenges

n tuples

group size k

Baseline Framework



n = 1 Million
k = 6

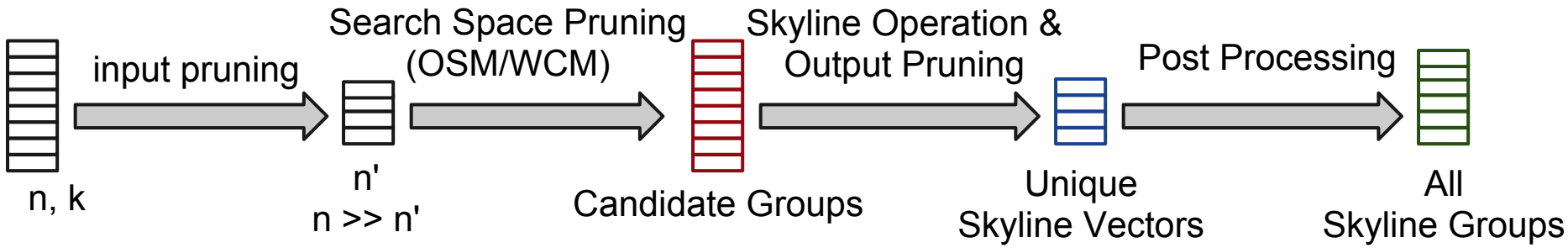
$\binom{n}{k}$ = 1 X 10^{33}

all skyline groups

12816

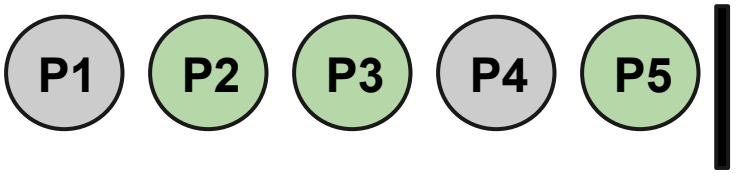
- n choose k is very large, we may not afford to compute or store that.
- Number of skyline groups can also be large.

Our Framework

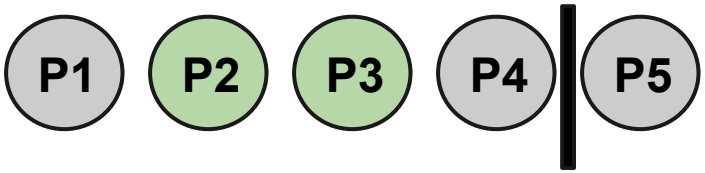


- These Skyline Groups can be input of further post-processing algorithms.
  - Representative Skyline Groups
  - Rank the Skyline Groups

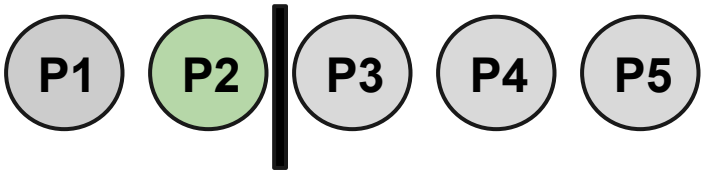
Search Space Pruning:OSM



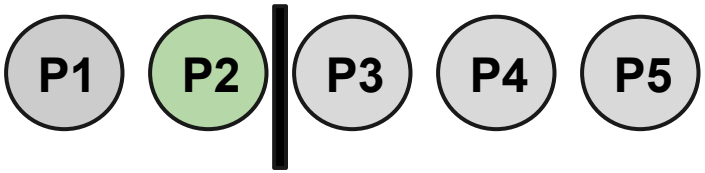
Search Space Pruning:OSM



Search Space Pruning:OSM



Search Space Pruning:OSM



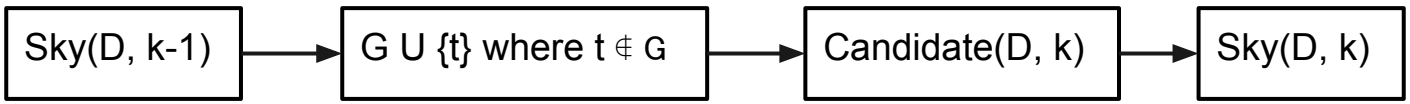
Order the tuples arbitrarily as
D_n = {P1, P2, ..., Pn}

Sky(Dn,k)	
A; Pn is present	← Sky(D_{n-1}, k-1) U {Pn}
B; Pn is absent	← Sky(D_{n-1}, k)

- An order based Anti-Monotonic property can be formed.
- SUM satisfies this property and it is extended for MIN and MAX by handling corner cases.

Search Space Pruning:WCM

- If a k-tuple group is in skyline then at least one (k-1)-tuple subset of it will also be in skyline.
- It is applicable in distinct value assumption. We extend this to general cases.
- We develop an iterative algorithm based on this property.
- WCM is satisfied by MIN and MAX. SUM does not satisfy this property.



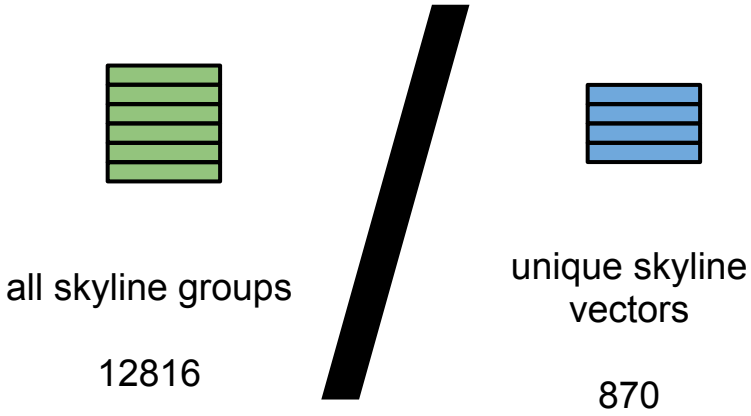
Input Pruning

- If a tuple is dominated by k or more than k tuples, it can be discarded.
- Example:
  - P4 is dominated by 4 players.
  - All **unique skyline vectors** can be found without requiring P4.
  - So, we can exclude P4 from input tuples.
- For MAX, it is sufficient to consider only skyline tuples.

	Points	Rebounds	Blocks
P1	3	4	5
P2	4	2	3
P3	4	5	3
P4	2	1	2
P5	4	1	2

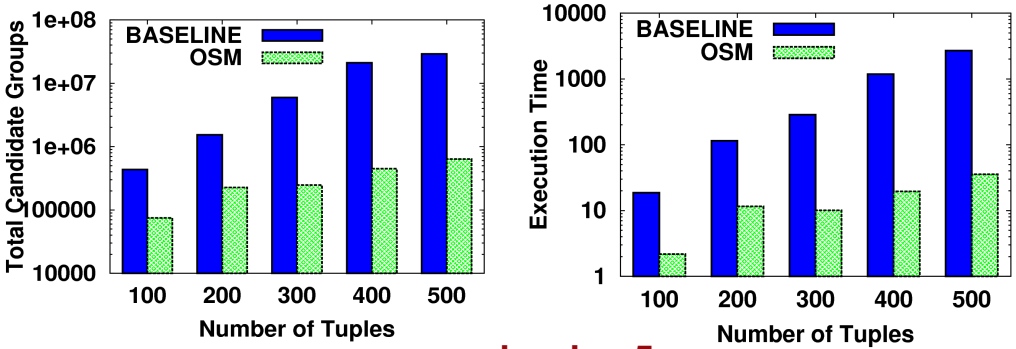
Output Pruning

- Multiple groups share the same aggregate score.
- Instead of all skyline groups, find **unique vectors**.
- All groups can be found by post-processing.
- MIN: It is sufficient to find all input tuples which are equal to or dominate a skyline vector and then find k-tuple combination of these; time complexity O(n).
- MAX: The problem is NP-hard. But simple brute-force is practically efficient because of small input size.

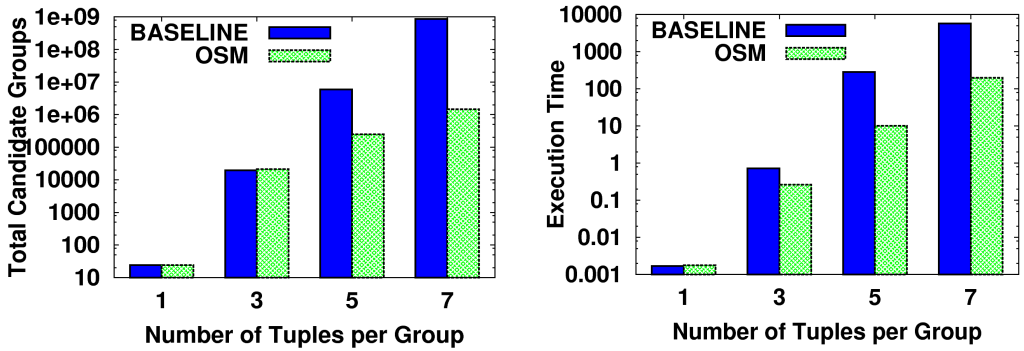


	Points	Rebounds	Blocks
Michael Jordan Lebron James Kobe Bryant	4	5	5
Michael Jordan Lebron James Carmelo Anthony	4	5	5

Experiment



group size, k = 5



Total tuples, n = 300

- NBA Dataset
- Synthetic Dataset
- Details in our CIKM paper.

Sample Skyline Groups

						PPG	RBG	APG	SPG	BPG
G1	Carmelo Anthony	Kobe Bryant	Kevin Durant	LeBron James	Dwyane Wade	283.2	63.4	52.2	15.2	7.6
G2	Andrew Bogut	Marcus Camby	Monta Ellis	Dwight Howard	Josh Smith	166.2	96.4	32.2	13.4	19.4
G3	Trevor Ariza	Monta Ellis	Dwyane Wade	Dwight Howard	Josh Smith	202	72.6	43.2	16.6	14
G4	Carlos Boozer	Baron Davis	LeBron James	Rajon Rondo	Chris Paul	193.8	61.2	80.6	17.6	4.8
G5	Andrew Bogut	LeBron James	Chris Paul	Dwight Howard	Jason Kidd	185.8	81	64	14	13.8

Table 4: Sample skyline groups from 512 players, 5 players per group

Future Work

- Generalize group aggregate function.
- Consume skyline groups.

Journal Link: http://ranger.uta.edu/~cli/

Acknowledgement
Travel Support

Mahalo :-)

feel free to drop any questions/suggestions...

naeemulhassan@gmail.com

Question