On Skyline Groups

Chengkai Li
Naeemul Hassan
Gautam Das
University of Texas at Arlington

Nan Zhang
Sundaresan Rajasekaran

George Washington University

Motivation
Question-Answer Platforms

Question

What is the difference between an int and an Integer in Java and C#?

92

I was just sitting at my local Borders sipping coffee and reading *More Joel on Software* (for free) when I came across Joel Spolsky saying something about a particular type of programmer knowing the difference between an int and an Integer in Java/C# (Object Oriented Programming Languages).

After a quick 'brain check,' I realized, to my dismay, that I didn't know the answer.

Skills

| c# | java | integer | int |

31

Expert's Name	c#	java	int	integer	Total Weights
Eric Lippert	68	3	43	0	114
Bozho	0	51	52	0	103
Marc Gravell	100	2	0	0	102
BalusC	0	100	0	0	100
Ben Hoffstein	0	0	0	100	100
DannyT	0	0	100	0	100
Matthieu M.	0	0	87	0	87
Rohit Jain	0	18	66	0	84
dasblinkenlight	11	15	46	0	72
arshajii	0	10	56	0	66

| Expert's Name | | | | | |

Showing 1 to 10 of 4,654 entries   Previous Next

Goal: Find a group of experts who can answer this question

1

Motivation
Journal/Paper Review

Keywords (7)
- Data Exploration
- Experimental Evaluation
- Faceted Search
- Information Discovery
- Internal Structure
- Semantic Information
- User Study

Skills

Academic › Publications › Facetedpedia: dynamic generation of query-dependent faceted interfaces for wikipedia

Subscribe

Facetedpedia: dynamic generation of query-dependent faceted interfaces for wikipedia (Citations: 3) Export

Task   Edit

Chengkai Li, Ning Yan, Senjuti Basu Roy, Lekhendro Lisham, Gautam Das

This paper proposes Facetedpedia, a faceted retrieval system for **information discovery** and exploration in Wikipedia. Given the set of Wikipedia articles resulting from a keyword query, Facetedpedia generates a faceted interface for navigating the result articles. Compared with other faceted retrieval systems, Facetedpedia is fully automatic and dynamic in both facet generation and hierarchy construction, and the facets are based on th... upon the collaborative vocabulary in Wikipe... (category system). Given the sheer size an...

Expert's Name	Experimental Evaluation	Semantic Information	User Study	Total Weights
Wolfram Burgard	4	100	0	104
Patrick Baudisch	2	0	100	102
Wei-ying Ma	11	33	58	102
Hans-Peter Kriegel	100	0	0	100
Martha Palmer	0	100	0	100
Juan-zi Li	0	100	0	100
Mary Czerwinski	0	0	88	88
Yong Yu	4	67	12	83
Wolfgang Nejdl	13	50	12	75
Pearl Pu	0	17	58	75

Expert's Name				

Showing 1 to 10 of 14,522 entries   Previous Next ▶

Goal: Find a group of experts who can review this paper

Motivation
Fantasy Games
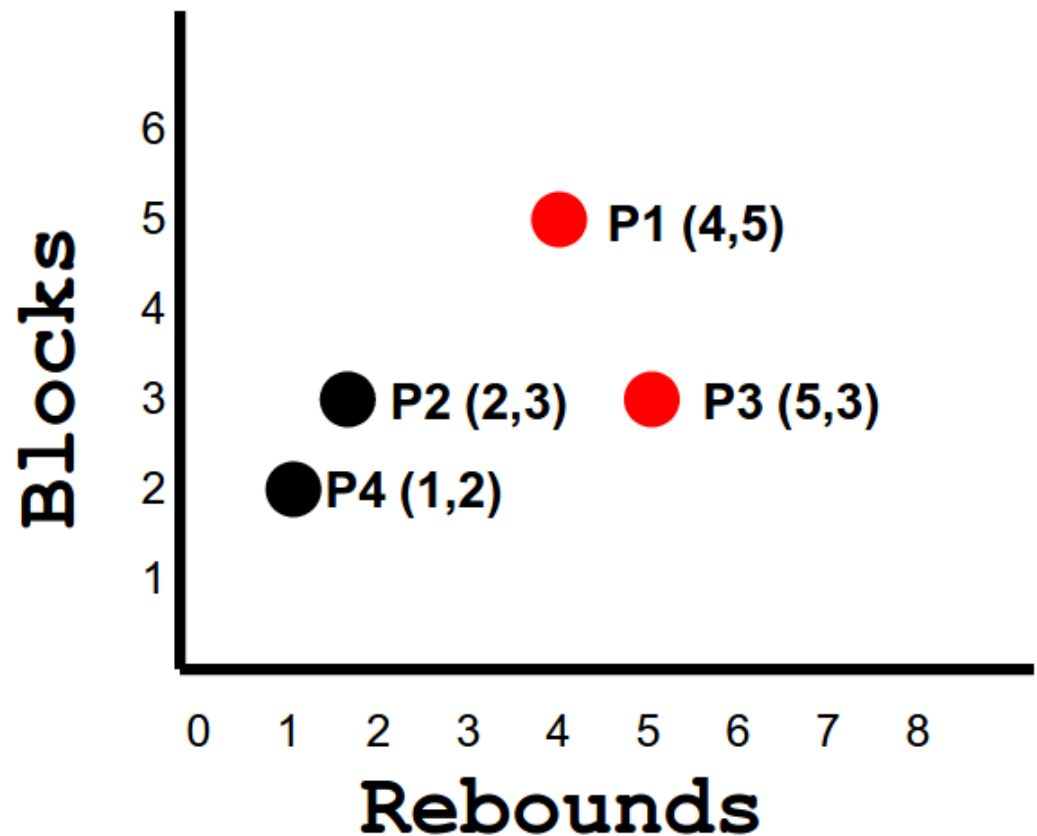

Goal: Find a group of players for Fantasy Basketball

Problem Definition
What is Skyline?

NBA Players Score

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

Skyline tuple: A tuple which is not dominated by any other tuple.

4

Problem Definition
What is Skyline Group?

NBA Players Score

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

Skyline Players

Skyline Groups

Find a group of 3 players

5 Choose 3 = 10 possible groups

	SUM			MIN			MAX		
	P	R	B	P	R	B	P	R	B
P1, P2, P3	11	11	11	3	2	3	4	5	5
P1, P2, P4	9	7	10	2	1	2	4	4	5
P1, P2, P5	11	7	10	3	1	2	4	4	5
P1, P3, P4	9	10	10	2	1	2	4	5	5
P1, P3, P5	11	10	10	3	1	2	4	5	5
P1, P4, P5	9	6	9	2	1	2	4	4	5
P2, P3, P4	10	8	8	2	1	2	4	5	3
P2, P3, P5	12	8	8	4	1	2	4	5	3
P2, P4, P5	10	4	7	2	1	2	4	2	3
P3, P4, P5	10	7	7	2	1	2	4	5	3

Problem Definition
Why Skyline Group?

NBA Players Score

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

What's wrong with taking most expert in each field?

Any other group is dominated by a Skyline

	SUM			MIN			MAX		
	P	R	B	P	R	B	P	R	B
P1, P2, P3	11	11	11	3	2	3	4	5	5
P1, P2, P4	9	7	10	2	1	2	4	4	5
P1, P2, P5	11	7	10	3	1	2	4	4	5
P1, P3, P4	9	10	10	2	1	2	4	5	5
P1, P3, P5	11	10	10	3	1	2	4	5	5
P1, P4, P5	9	6	9	2	1	2	4	4	5
P2, P3, P4	10	8	8	2	1	2	4	5	3
P2, P3, P5	12	8	8	4	1	2	4	5	3
P2, P4, P5	10	4	7	2	1	2	4	2	3
P3, P4, P5	10	7	7	2	1	2	4	5	3

6

Solution Framework
Baseline Method

Input

- n players/tuples
- group size k
- aggregate function
  (sum/min/max)

group generation
(SUM / MIN / MAX)

skyline operation

n, k

$\binom{n}{k}$

all skyline groups
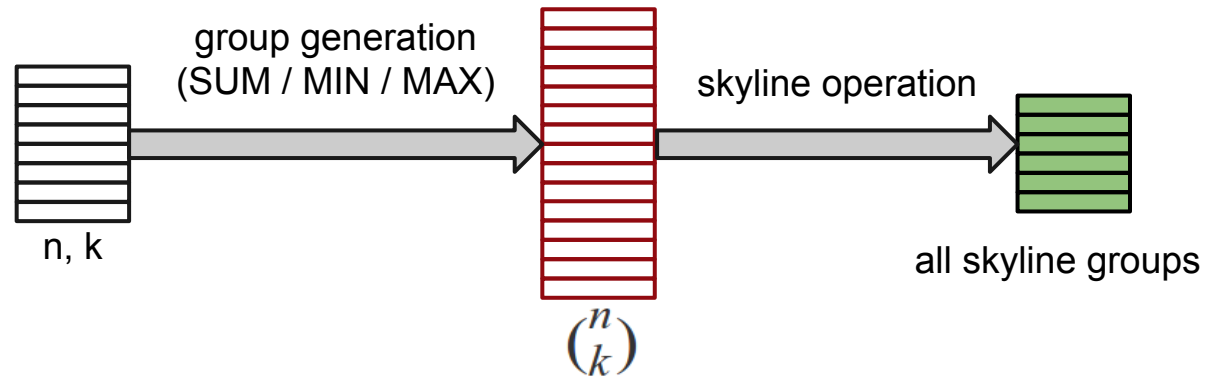
Problems
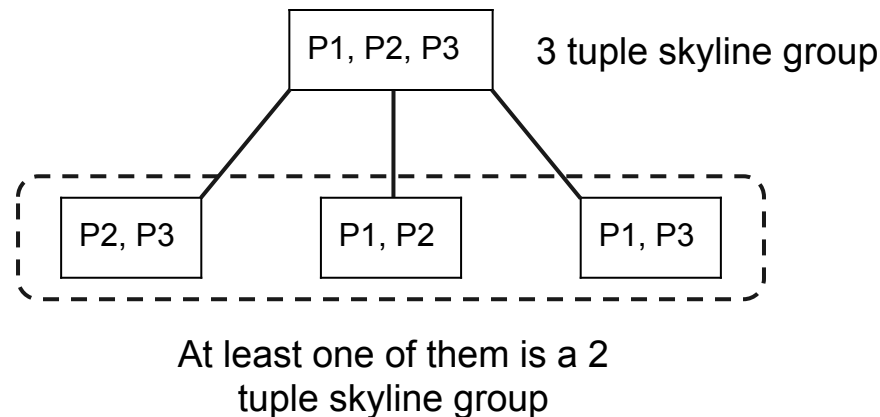
- Exponential group generation. We may not afford to compute or store them.
  - Example: For n = 2000, k = 3.
    - 1331334000 groups
    - 30 GB space [assuming 24B for each group]
    - 15 days time [assuming 1 millisecond for each group]

Solution Framework Advanced Method: WCM

<u>Weak Candidate Generation Property</u>: If G is a k tuple skyline group, then there is **at least one (k-1) tuple subset of G** such that it is a (k-1) tuple skyline group.
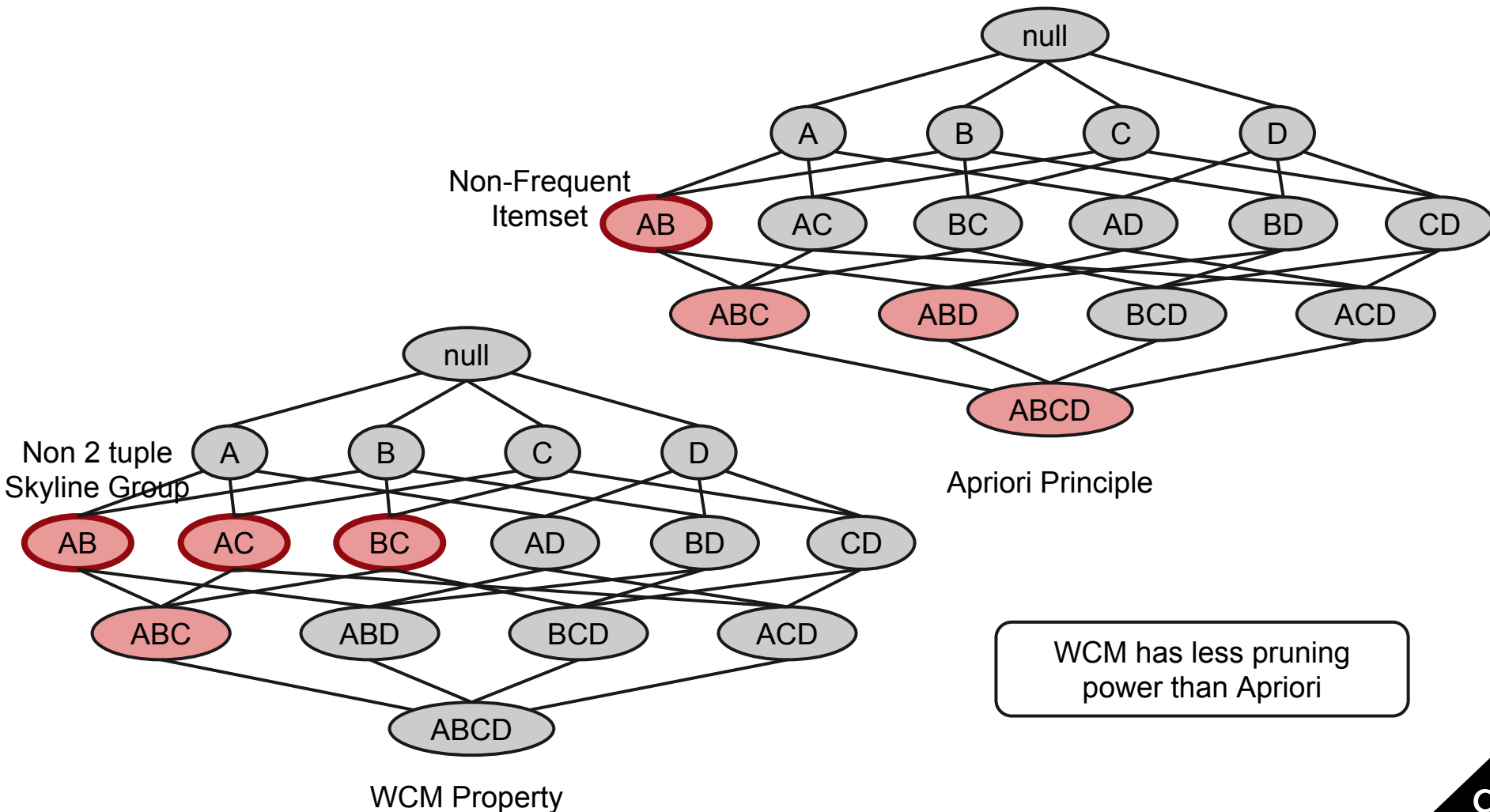
<u>Example</u>:

```
                    ┌─────────────┐
                    │  P1, P2, P3 │   3 tuple skyline group
                    └─────────────┘
                   ╱       │       ╲
         ┌ ─ ─ ─ ─╱─ ─ ─ ─ │ ─ ─ ─ ─╲─ ─ ─ ─ ┐
           ┌──────────┐ ┌──────────┐ ┌──────────┐
         │ │  P2, P3  │ │  P1, P2  │ │  P1, P3  │ │
           └──────────┘ └──────────┘ └──────────┘
         └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
```

At least one of them is a 2
tuple skyline group

<u>Does this property sound familiar?</u>

<u>Apriori Principle</u>: If an itemset is frequent, then **all of its subsets** must also be frequent
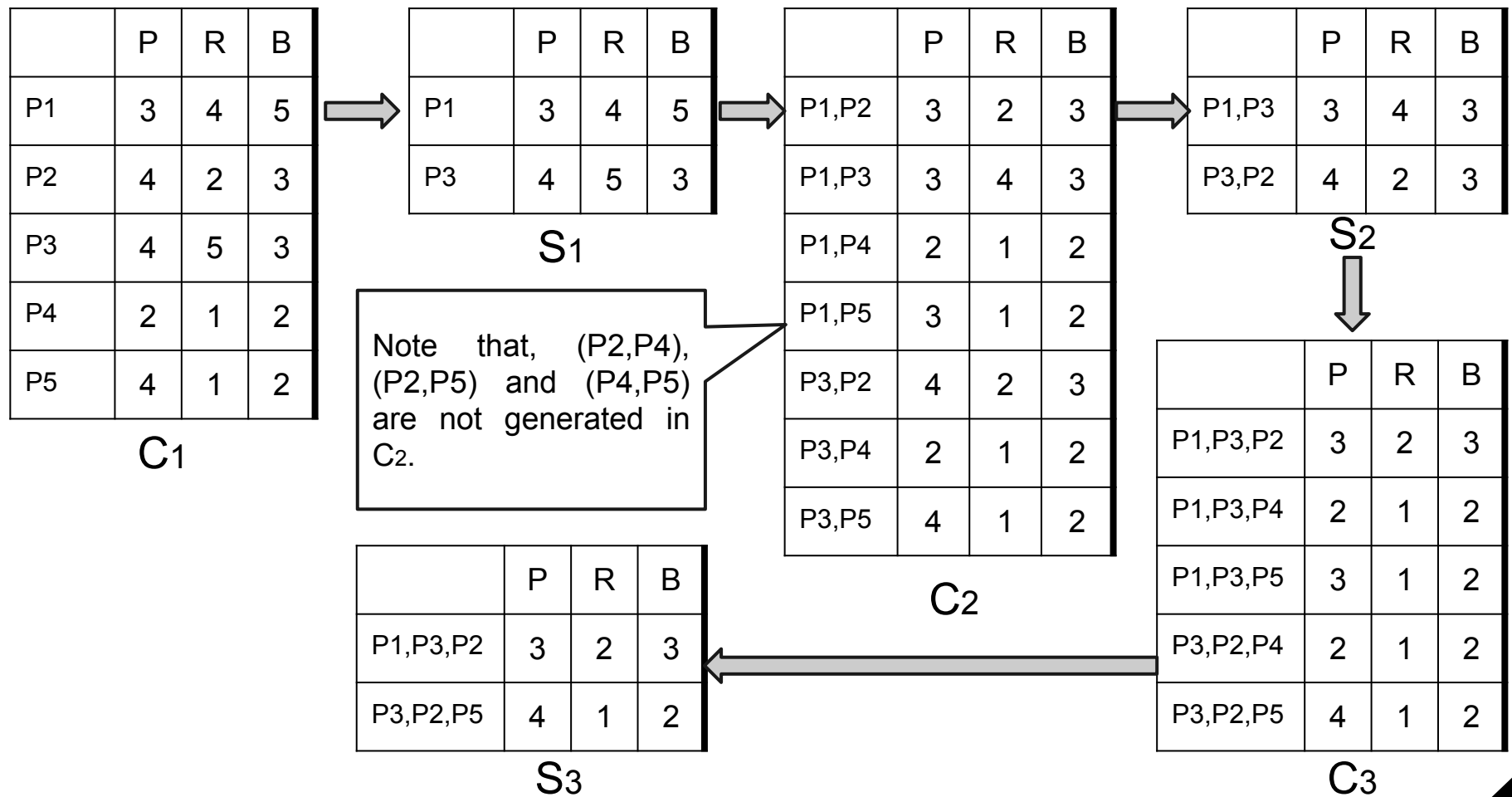
Comparison Between Apriori & WCM Property

Non-Frequent Itemset

Apriori Principle

Non 2 tuple Skyline Group

WCM Property

WCM has less pruning power than Apriori

WCM Algorithm

Input: n tuples, group size k, aggregate function = **min/max** (**not sum**)

1. Let, i = 1
2. Generate 1 tuple Candidate groups, C_1 = all n tuples
3. Generate 1 tuple Skyline groups, S_1 = skyline_operation(C_1)
4. for i = 2 to k
   a. Generate i tuple Candidate groups, C_i from S_{i-1}
   b. Generate i tuple Skyline groups, S_i = skyline_operation(C_i)
5. Return S_k

WCM Algorithm Explained with Example

Input: n tuple {P1, P2, P3, P4, P5}, group size k = 3, aggregate function = min

	P	R	B
P1	3	4	5
P2	4	2	3
P3	4	5	3
P4	2	1	2
P5	4	1	2

C_1

	P	R	B
P1	3	4	5
P3	4	5	3

S_1

	P	R	B
P1,P2	3	2	3
P1,P3	3	4	3
P1,P4	2	1	2
P1,P5	3	1	2
P3,P2	4	2	3
P3,P4	2	1	2
P3,P5	4	1	2

C_2

Note that, (P2,P4), (P2,P5) and (P4,P5) are not generated in C_2.

	P	R	B
P1,P3	3	4	3
P3,P2	4	2	3

S_2

	P	R	B
P1,P3,P2	3	2	3
P1,P3,P4	2	1	2
P1,P3,P5	3	1	2
P3,P2,P4	2	1	2
P3,P2,P5	4	1	2

C_3

	P	R	B
P1,P3,P2	3	2	3
P3,P2,P5	4	1	2

S_3

Question

?

CrewScout System

http://idir.uta.edu/crewscout

Thank You!