



# The skyline operator

## recent research directions and applications

Matteo Magnani

*matteo.magnani@it.uu.se*

*Computing Science Division  
Uppsala University, Sweden*



# A very basic operation in data management

Extract the **best** elements  
from a set of options

According to **multiple criteria**.

# Motivation (hotel booking)

The screenshot shows a search interface for hotels in Paris. At the top, there's a result for "Lemon Hotel Longperrier Roissy" located near Charles de Gaulle Airport, with a price of \$43 per night. Below this, several filters are listed: "Cheap", "Close to Ile de la Cité", "Cheap and close to Ile de la Cité", "Cheap and high rating", and "Cheap, high rating and many stars". A detailed view of the "Albe Hotel" is shown, which is located in the Latin Quarter - Pantheon area, has a rating of 4.5 out of 5 based on 21+ reviews, and costs \$174 per night. The background map of Paris highlights the location of the hotel.

**Lemon Hotel Longperrier Roissy**  
Longperrier (Roissy - Charles de Gaulle Airport (CDG)) - Map

1-800-264-5744

Cheap

Close to Ile de la Cité

Cheap and close to Ile de la Cité

Cheap and high rating

Cheap, high rating and many stars

...

**Albe Hotel**  
Paris,   
\$174 avg/night

**Area: Latin Quarter - Pantheon (5)**  
4.5 out of 5 | 21+ reviews  
Located in the heart of Paris, this hotel is within walking distance of Pont Saint-Michel, Notre Dame Cathedral, and Sainte Chapelle. Also nearby are Pont Neuf and Sorbonne University.

# Motivation (hotel booking)



## Lemon Hotel Longperrier Roissy

Longperrier (Roissy - Charles de Gaulle Airport (CDG)) - [Map](#)

1-866-264-5744

\$43

rate per night

The map highlights several hotel locations in the Latin Quarter of Paris, including the Albe Hotel, L'orangerie, and Lemmon Hotel.

**Albe Hotel**  
Paris,  
 \$174 avg/night

**Area: Latin Quarter - Pantheon (5)**  
4.5 out of 5 | 21+ reviews  
Located in the heart of Paris, this hotel is within walking distance of Pont Saint-Michel, Notre Dame Cathedral, and Sainte Chapelle. Also nearby are Pont Neuf and Sorbonne University.

Every Km is worth 10\$



- 1) Albe Hotel
- 2) L'orangerie
- 3) Lemmon Hotel
- 4) ...

# Outline

- The skyline (Pareto front) operator.
- Practical issues.
- Current research directions.

# **THE SKYLINE OPERATOR**

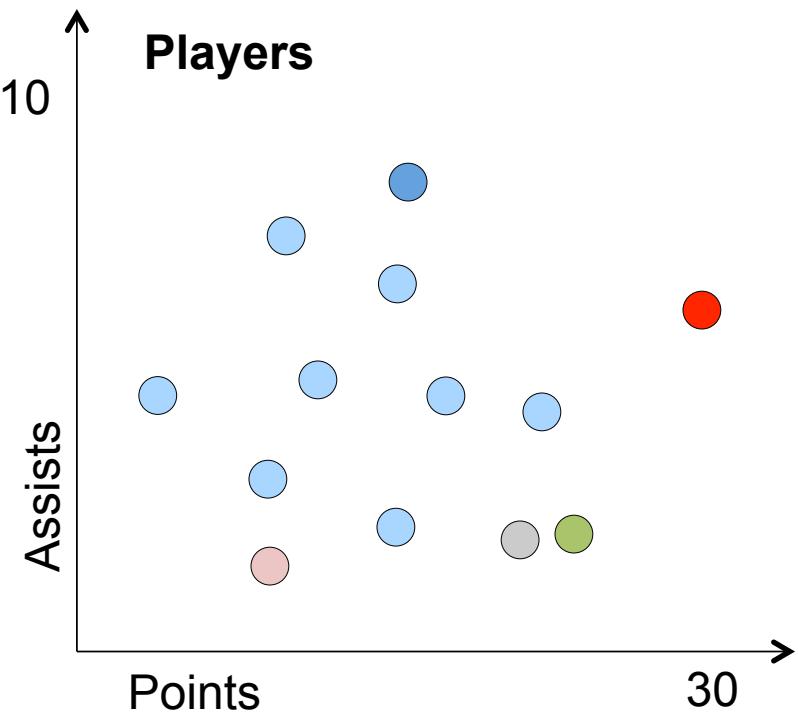
```

SELECT *
FROM PLAYERS
SKYLINE OF Assists MAX, Points MAX

```

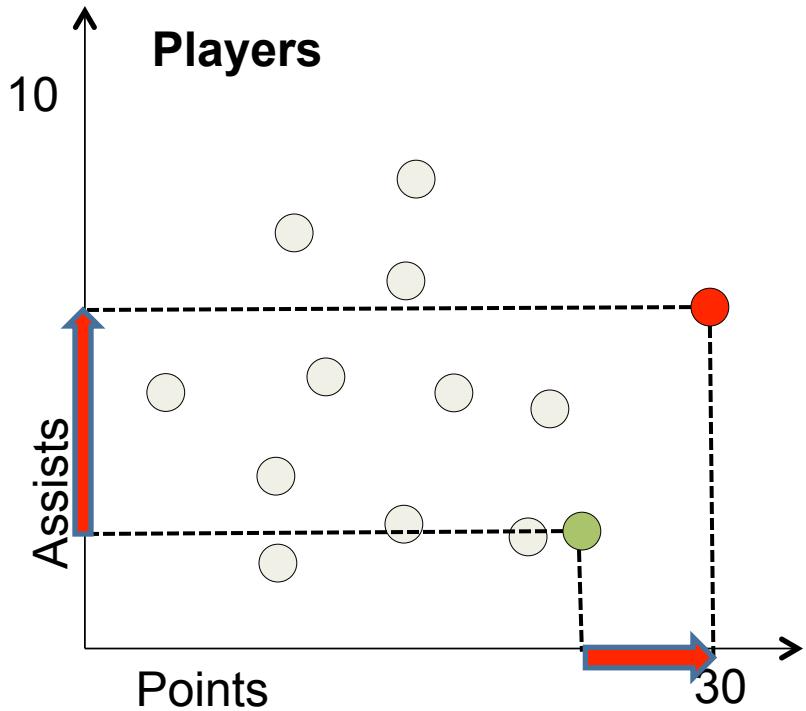
Player	Assists	Points
Yao Ming	1.6	19
Steve Nash	8.5	14.6
Michael Jordan	5.3	30
Dirk Novitzki	1.8	22
Marco Belinelli	1.3	8

# Skyline



# Skyline

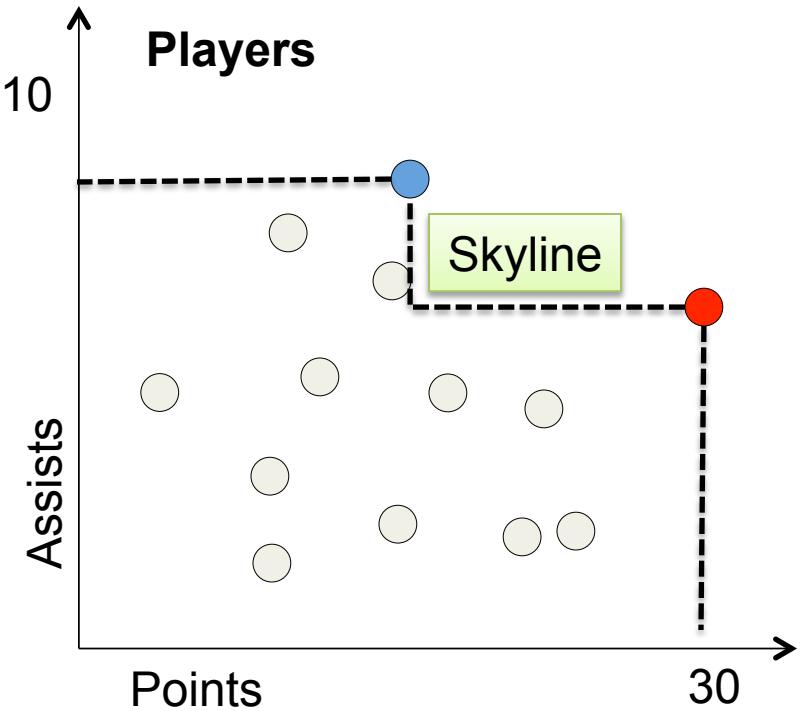
Player	Assists	Points
Yao Ming	1.6	19
Steve Nash	8.5	14.6
Michael Jordan	5.3	30
Dirk Novitzki	1.8	22
Marco Belinelli	1.3	8



DEFINITION 1 (DOMINANCE). Let  $r$  and  $s$  be two records in  $d$  dimensions.  $r$  dominates  $s$  iff  $\forall i \in [1, d] r_i \geq s_i \wedge \exists i r_i > s_i$

# Skyline

Player	Assists	Points
Yao Ming	1.6	19
Steve Nash	8.5	14.6
Michael Jordan	5.3	30
Dirk Novitzki	1.8	22
Marco Belinelli	1.3	8



# Properties of the skyline

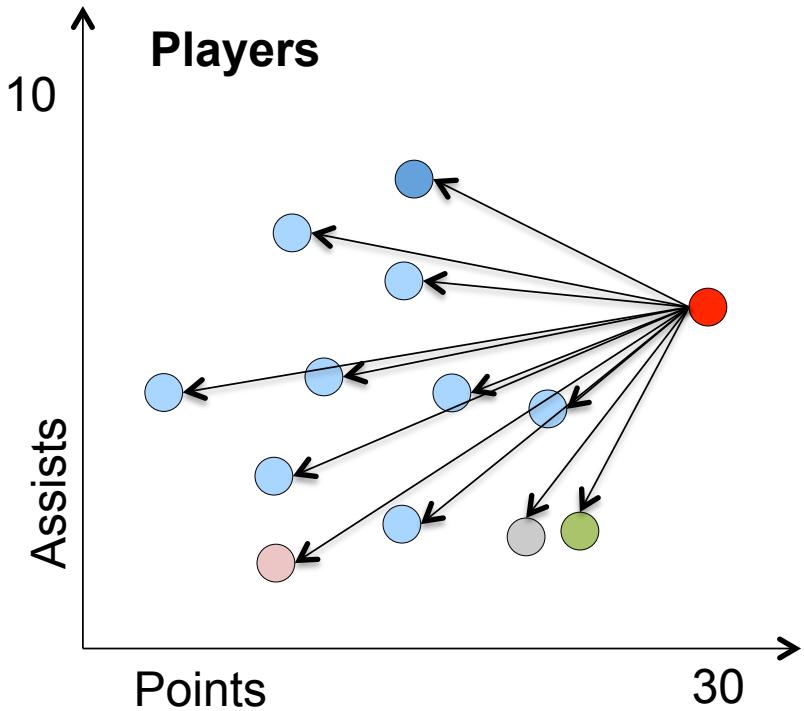
- Selects the best option for all reasonable users.
- It is the tightest result with this property.
- Requires nearly no parameters.
  - Differently from ranking.

# **PRACTICAL ISSUES**

# Practical issues

Player	Assists	Points
Yao Ming	1.6	19
Steve Nash	8.5	14.6
Michael Jordan	5.3	30
Dirk Novitzki	1.8	22
Marco Belinelli	1.3	8

To decide if a record is in the skyline, it may be necessary to **compare it with all other records.**

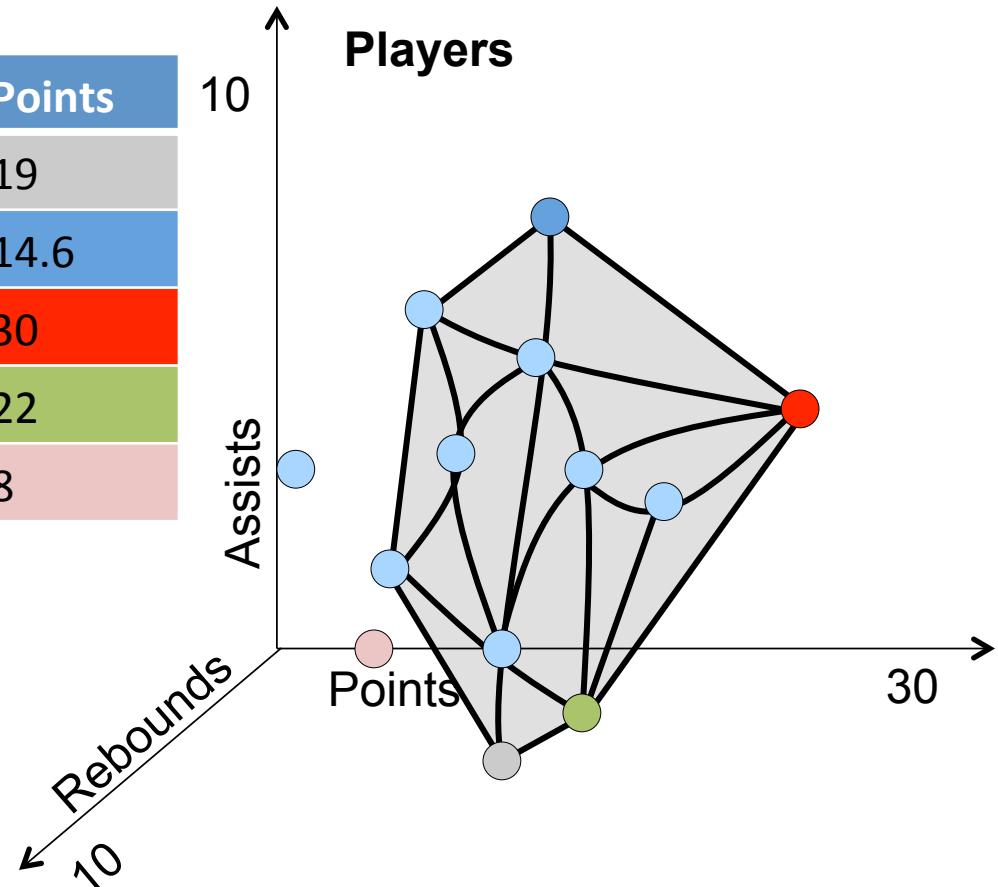


# Practical issues

Player	Assists	Rebounds	Points
Yao Ming	1.6	9.2	19
Steve Nash	8.5	3	14.6
Michael Jordan	5.3	6.2	30
Dirk Novitzki	1.8	7.1	22
Marco Belinelli	1.3	1.5	8

## Computational complexity.

The **size** of a skyline can be very large depending on the number of attributes and the data distribution



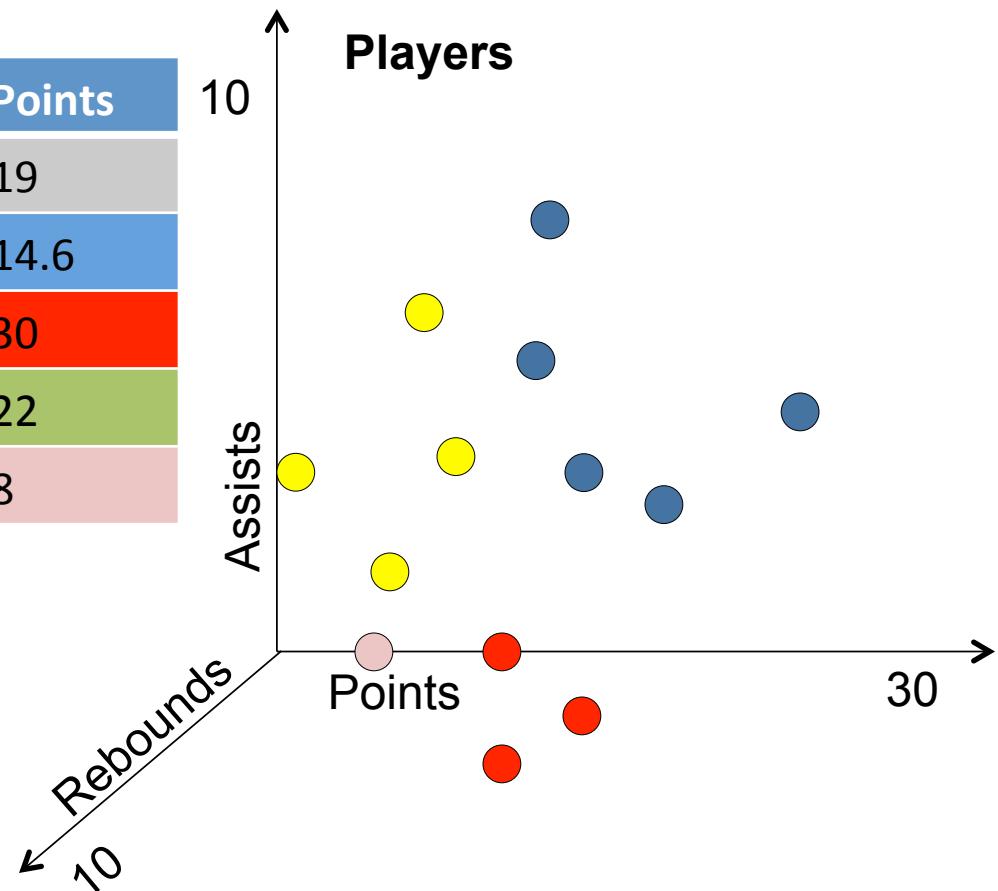
# Practical issues

Player	Assists	Rebounds	Points
Yao Ming	1.6	9.2	19
Steve Nash	8.5	3	14.6
Michael Jordan	5.3	6.2	30
Dirk Novitzki	1.8	7.1	22
Marco Belinelli	1.3	1.5	8

**Computational complexity.**

**Size** of the result.

Not used in **isolation**.



# Practical issues

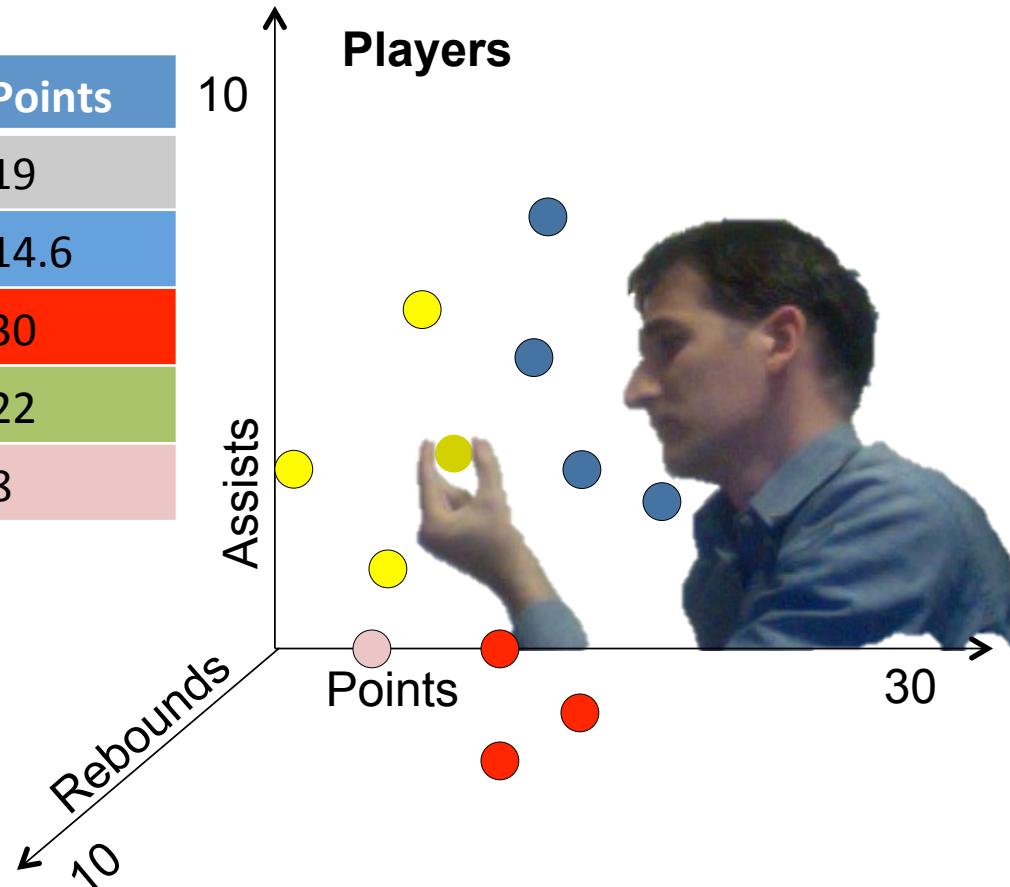
Player	Assists	Rebounds	Points
Yao Ming	1.6	9.2	19
Steve Nash	8.5	3	14.6
Michael Jordan	5.3	6.2	30
Dirk Novitzki	1.8	7.1	22
Marco Belinelli	1.3	1.5	8

**Computational complexity.**

**Size** of the result.

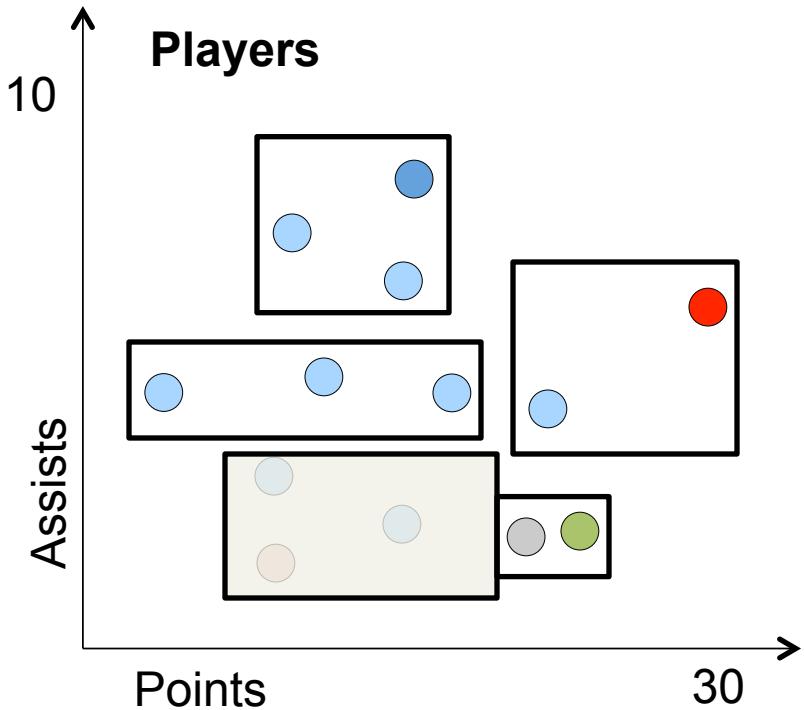
Not used in **isolation**.

Which interface to the **users**?



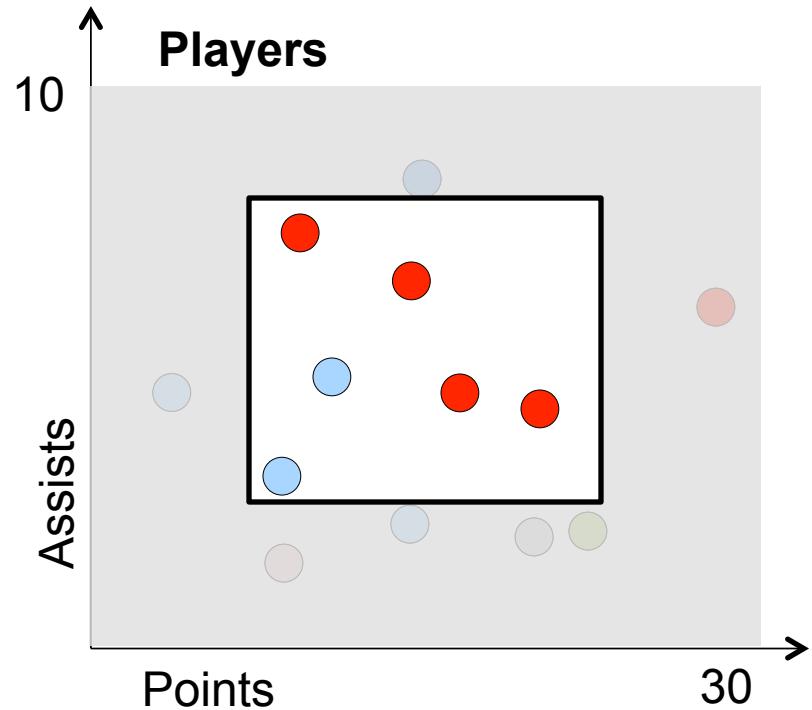
# Issue 1: Computational Complexity

- Traditional approach: multidimensional index.
- Very efficient with limited number of parameters.
- May not be fast enough for interactive systems.



# Instant skyline computation

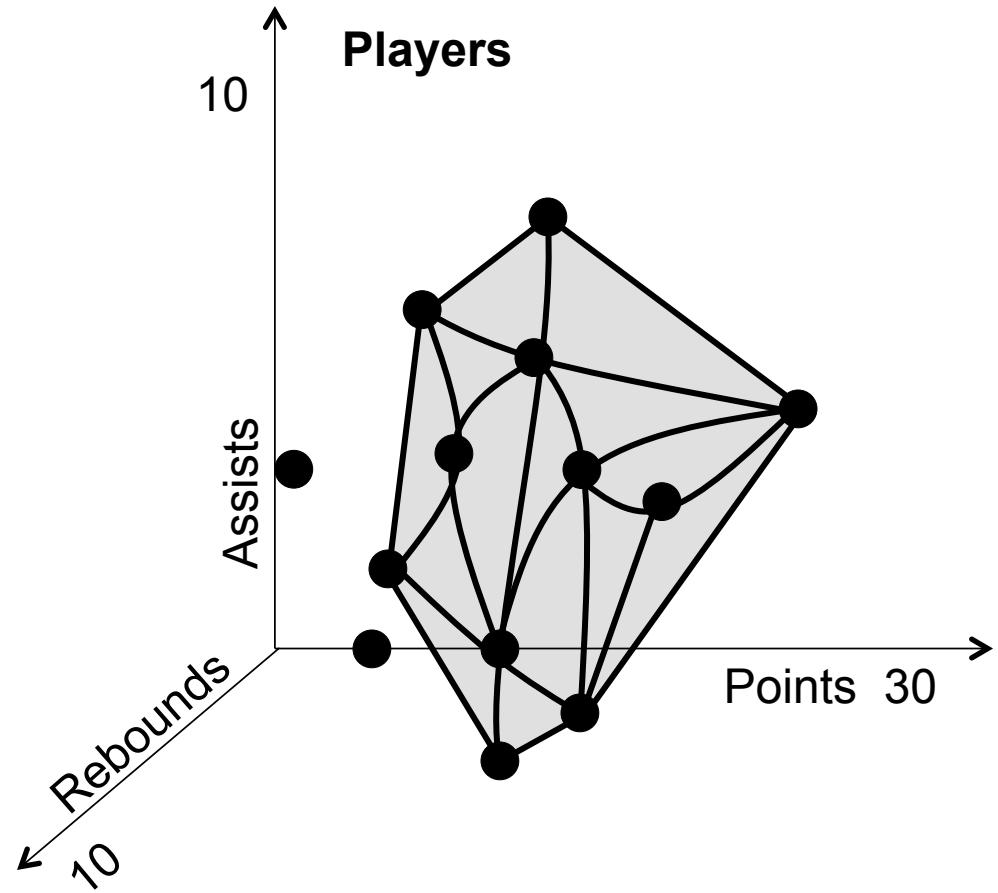
- GP-GPU computing.
- Anytime skylines.



Kenneth S. Bøgh, Ira Assent, Matteo Magnani. **Efficient GPU-based skyline computation.** SIGMOD Workshops (DaMoN), 2013

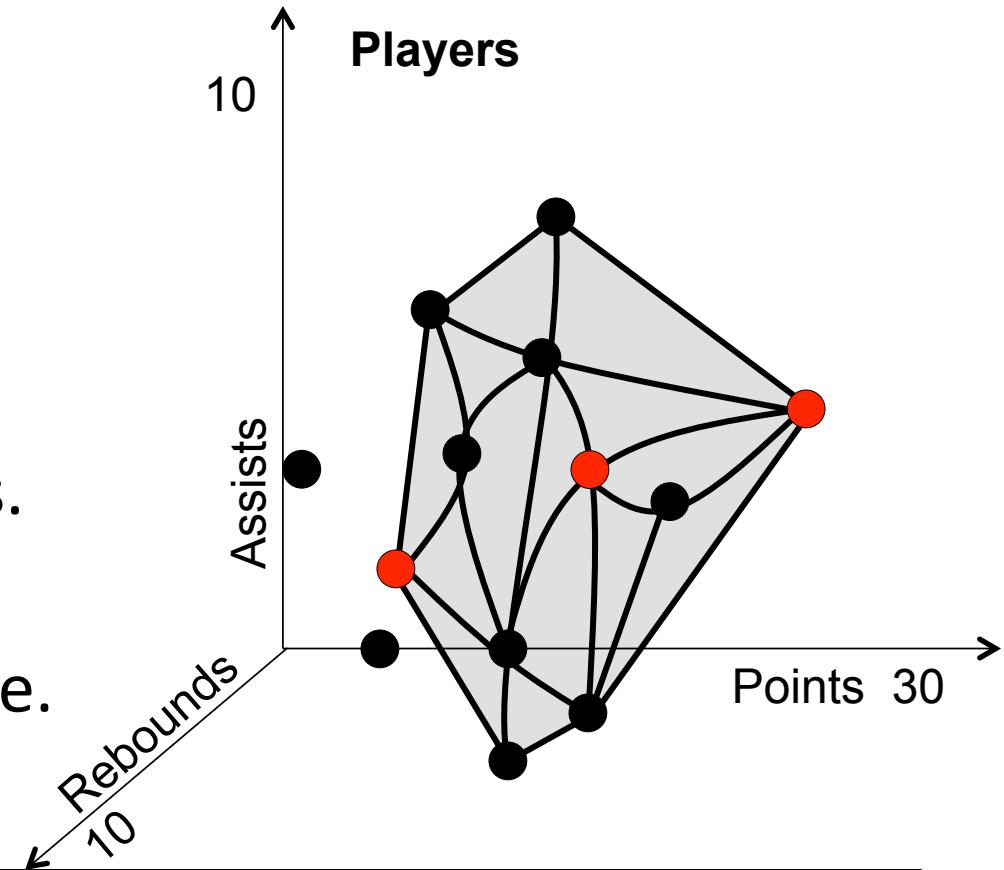
Matteo Magnani, Ira Assent, Michael L. Mortensen. **Anytime skyline query processing for interactive systems.** VLDB Workshops (DBRank), 2012.

# Issue 2: skyline size



# Representative skyline

- Computation of the skyline.
- Followed by the selection of a set of representative options.
- Based on notions of coverage and relevance.



M. Magnani, I. Assent, M. L. Mortensen.

**Taking the Big Picture: representative skylines based on relevance and diversity.**  
Under submission at VLDB Journal.

# Issue 3: combination with other operators

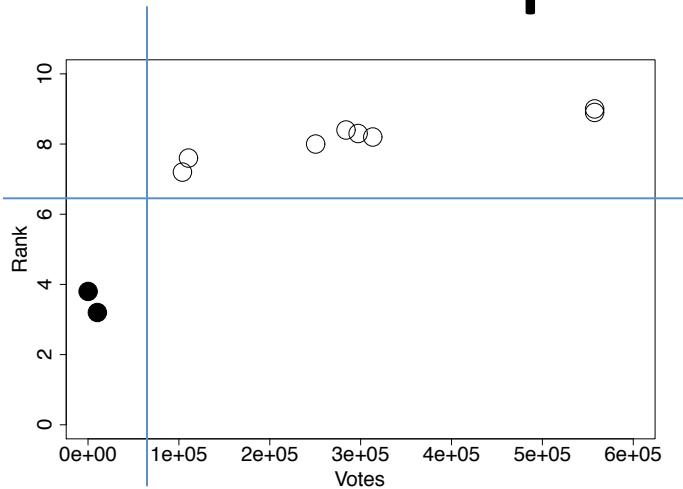
Title	Year	Director	Pop	Qual
Avatar	2009	Cameron	404	8.0
Batman Begins	2005	Nolan	371	8.3
Kill Bill	2003	Tarantino	313	8.2
Pulp Fiction	1994	Tarantino	557	9.0
Star Wars (V)	1980	Kershner	362	8.8
Terminator (II)	1991	Cameron	326	8.6
The Godfather	1972	Coppola	531	9.2
The Lord of the Rings	2001	Jackson	518	8.7
The Room	2003	Wiseau	10	3.2
Dracula	1992	Coppola	76	7.3

# *Find the best directors according to the movies they have directed*

Title	Year	Director	Pop	Qual
Avatar	2009	Cameron	404	8.0
Batman Begins	2005	Nolan	371	8.3
Kill Bill	2003	Tarantino	313	8.2
Pulp Fiction	1994	Tarantino	557	9.0
Star Wars (V)	1980	Kershner	362	8.8
Terminator (II)	1991	Cameron	326	8.6
The Godfather	1972	Coppola	531	9.2
The Lord of the Rings	2001	Jackson	518	8.7
The Room	2003	Wiseau	10	3.2
Dracula	1992	Coppola	76	7.3

```
SELECT director  
FROM movies  
GROUP BY Director  
SKYLINE OF Pop MAX, Qual MAX
```

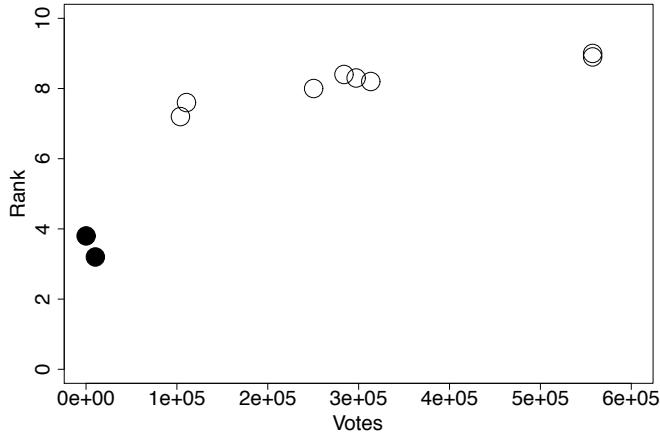
# Group dominance



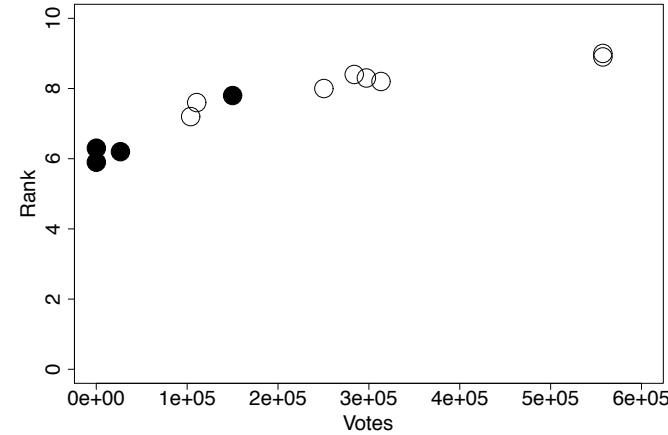
(a) Tarantino vs Wiseau



# Group dominance



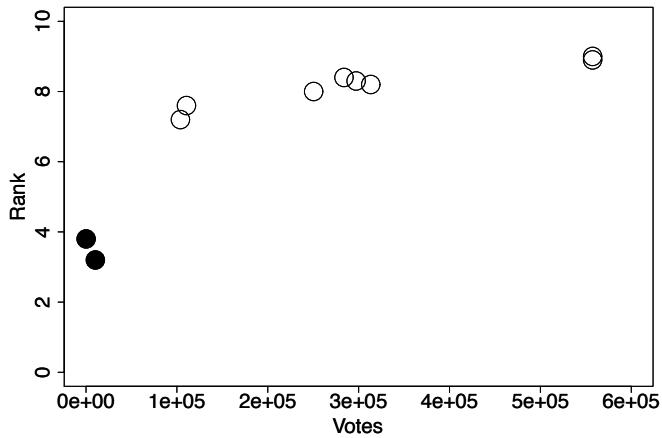
(a) Tarantino vs Wiseau



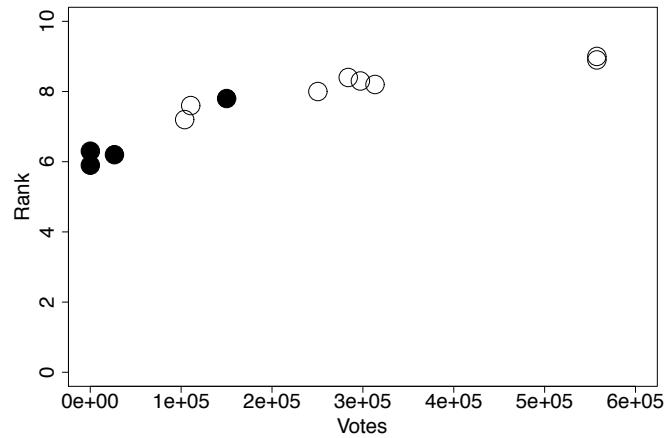
(b) Tarantino vs Fleischer



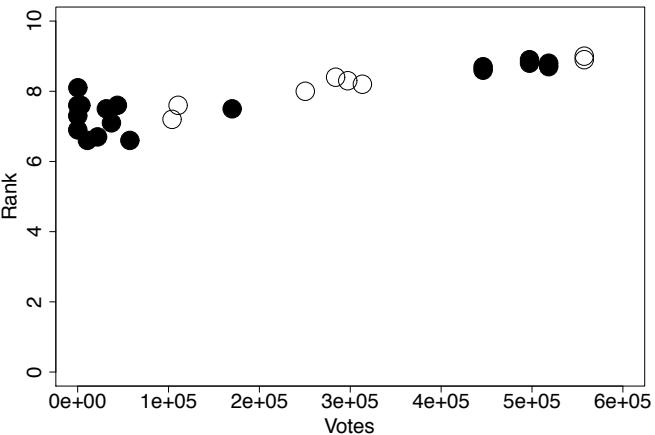
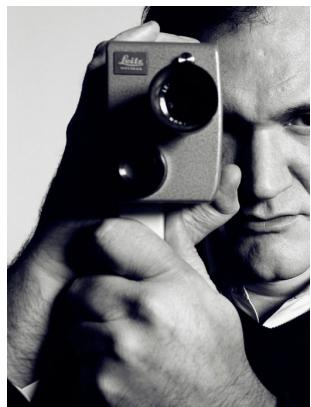
# Group dominance



(a) Tarantino vs Wiseau



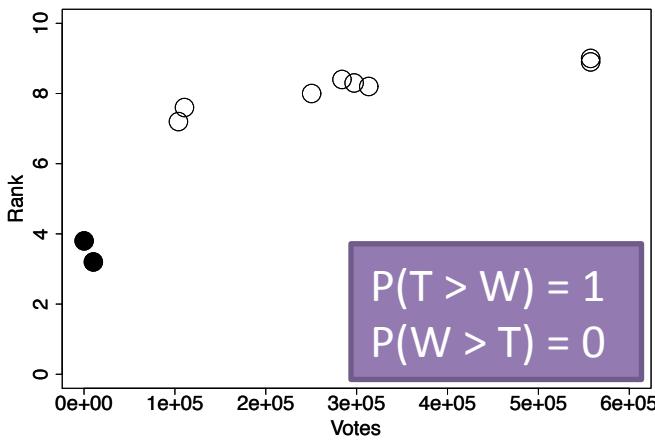
(b) Tarantino vs Fleischer



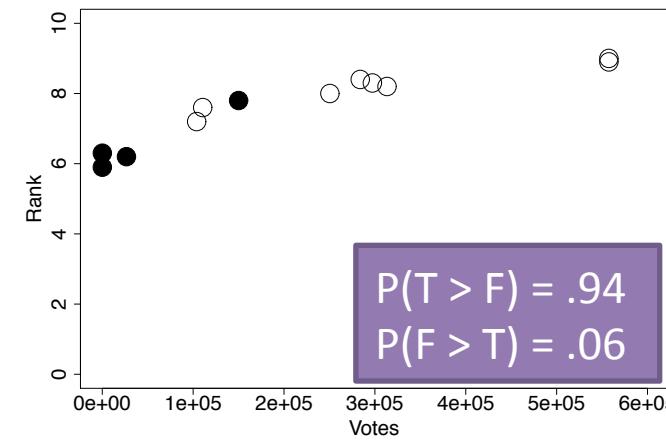
(c) Tarantino vs Jackson



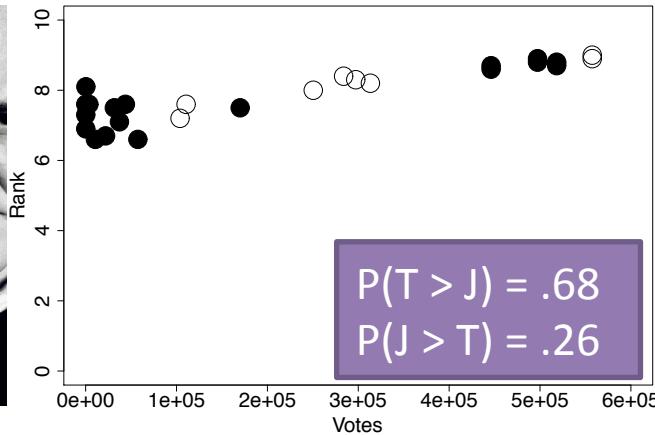
**DEFINITION 3** ( $\gamma$ -DOMINATION). Let  $p(S \succ R)$  be the probability that, given a random pair of records  $(r, s) \in R \times S$ ,  $s \succ r$ . We say that  $S \succ_\gamma R$  iff  $p(S \succ R) = 1 \vee p(S \succ R) > \gamma$ .



(a) Tarantino vs Wiseau



(b) Tarantino vs Fleischer

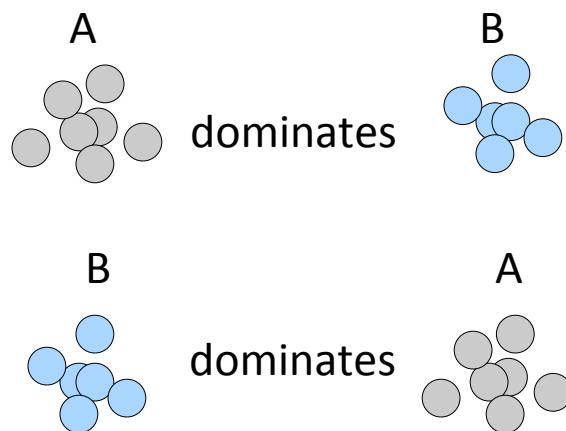


(c) Tarantino vs Jackson



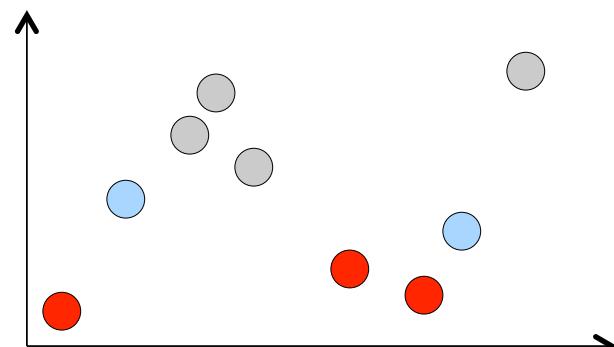
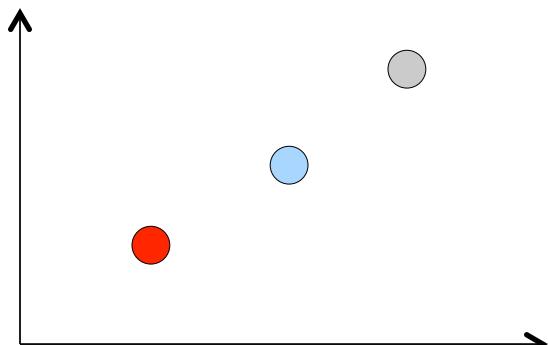
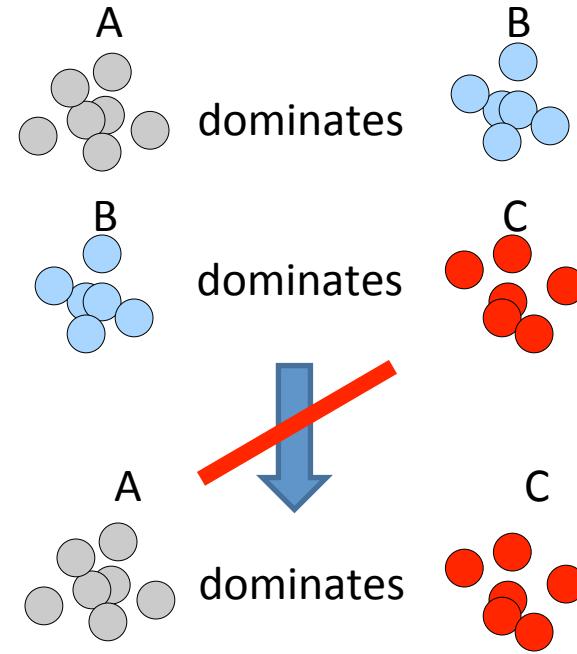
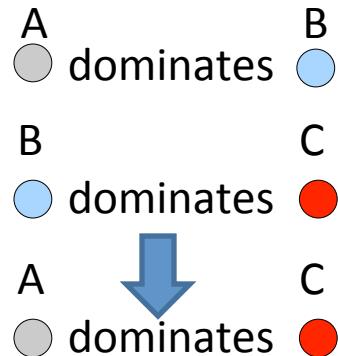
# Thresholds and asymmetry

- Above which probability should we filter out a group of records?
- Relevant threshold: probability  $> .5$
- Otherwise, no asymmetry.



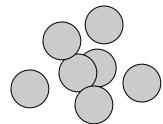
# Transitivity

- In general transitivity does not hold.

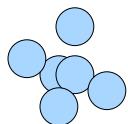
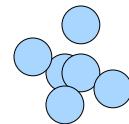


# Weak transitivity

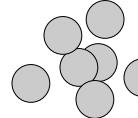
PROPOSITION 5 (WEAK TRANSITIVITY). If  $R \succ_{\bar{\gamma}} S$  and  $S \succ_{\bar{\gamma}} T$  then  $R \succ_{\gamma} T$ , where  $\bar{\gamma} = 1 - \frac{\sqrt{1-\gamma}}{2}$ .



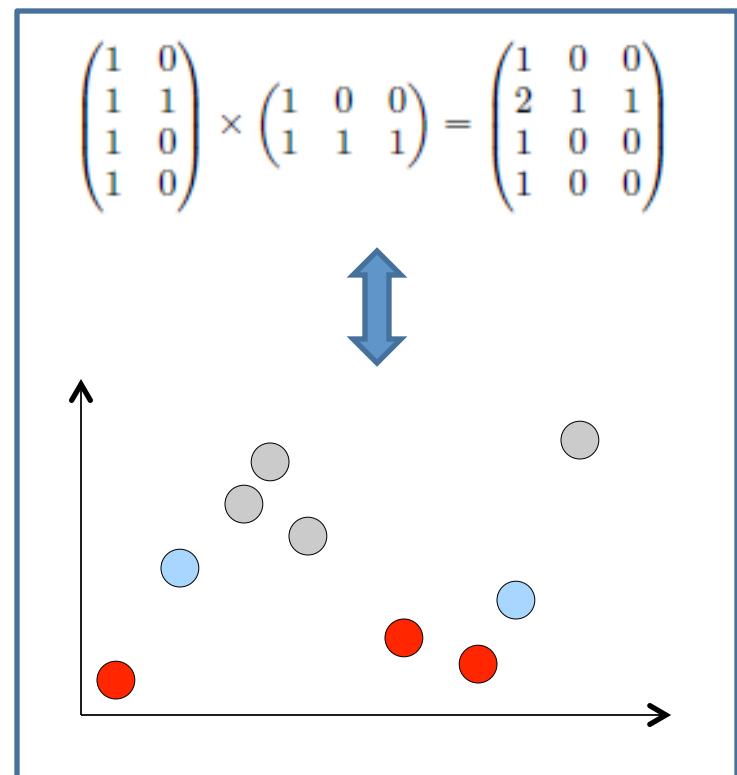
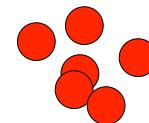
Strongly dominates



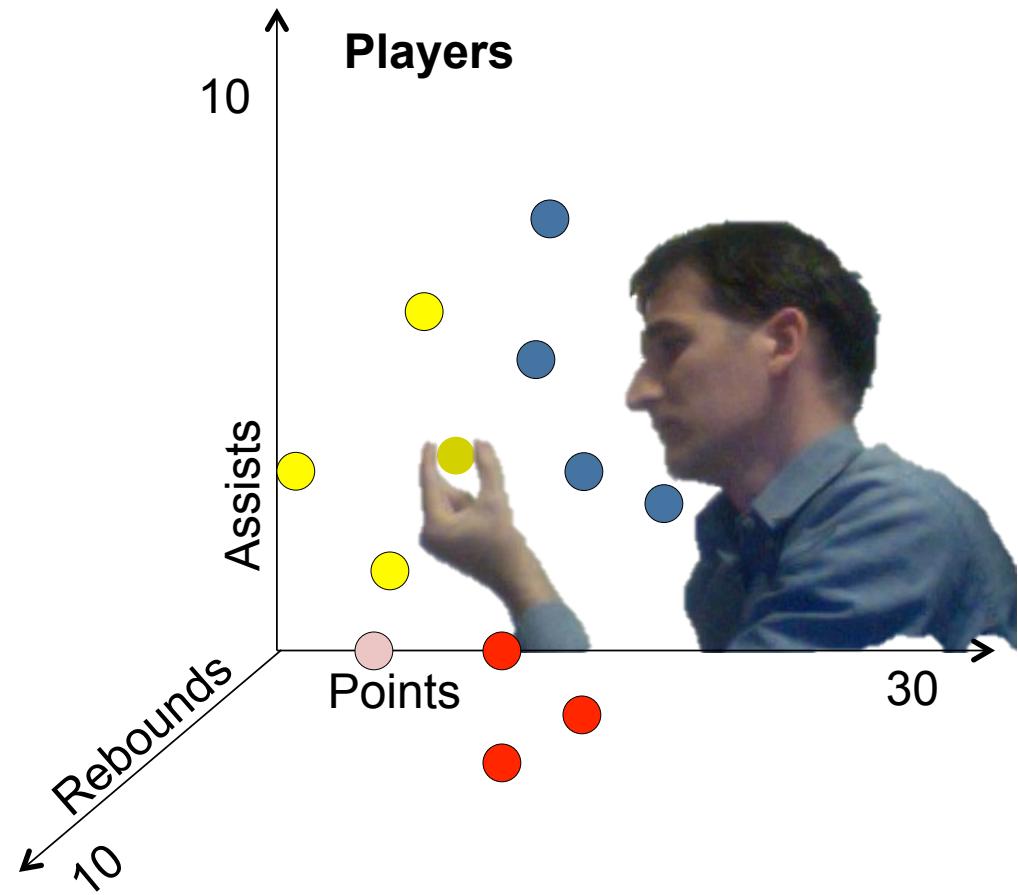
Strongly dominates



dominates



# Issue 4: user interface?



# User evaluation of a skyline-enhanced hotel booking interface

## Price and Location

Price per Night:

Distance from Colosseum:

Distance from Trevi Fountain:

Distance to the Vatican:

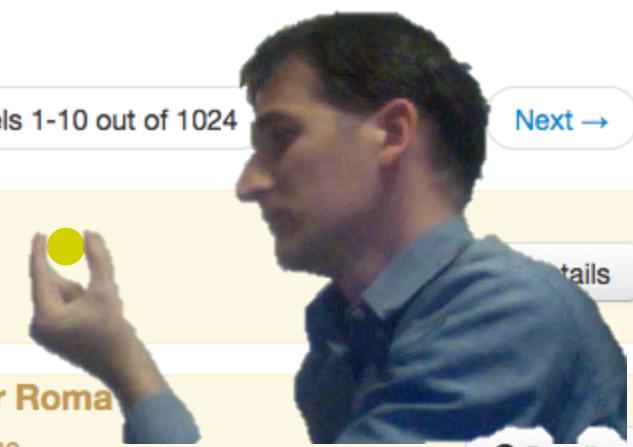
## Facilities

Internet

Pool

Order by: Price per Night (lowest first)

← Previous      Showing hotels 1-10 out of 1024      Next →



**Seven Hills Village**  
Via Vittorio Trucchi 10, 00189 Rome  
Price per Night: 23€. 0 stars

**Quality Hotel Rouge et Noir Roma**  
Via Cosimo de Giorgi, 8, 00158 Rome  
Price per Night: 44€. 4 stars

**Camping Village Fabulous**  
Via Di Malafede 225, 00125 Rome  
Price per Night: 45€. 0 stars

**Quality Hotel Excel Ciampino Airport**

# Multiple options, think-aloud study

Distance from Colosseum:

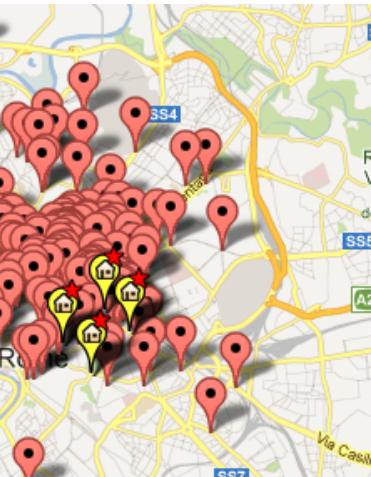
(a)

Distance from Colosseum: 0 - 30 km

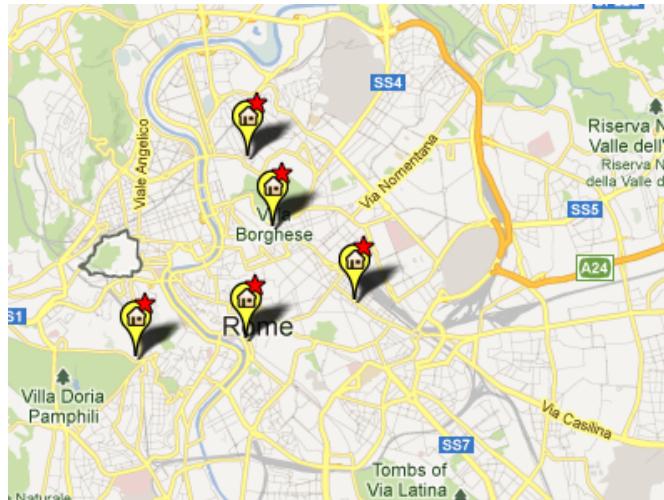
(b)

← Previous Showing hotels 1-10 out of 1019 Next →

	<b>Cesare Balbo Inn</b> Via Cesare Balbo 43, 00184 Rome Price per Night: 10€. 0 stars	<input type="button" value="Details"/>
	<b>Seven Hills Village</b> Via Vittorio Trucchi 10, 00189 Rome Price per Night: 23€. 0 stars	<input type="button" value="Details"/>
	<b>Quality Hotel Rouge et Noir Roma</b> Via Cosimo de Giorgi, 8, 00158 Rome Price per Night: 44€. 4 stars	<input type="button" value="Details"/>
	<b>Santi Quattro al Colosseo</b> Via Santi Quattro, 64, 00184 Rome Price per Night: 195€. 0 stars	<input type="button" value="Details"/>
	<b>Sleeping Beauty</b> Viale Carlo Felice 89 interno 1, 00185 Rome Price per Night: 49€. 0 stars	<input type="button" value="Details"/>
	<b>Sheraton Roma Hotel &amp; Conference Center</b> Viale Del Pattiaggio 100, 00144 Rome Price per Night: 151€. 4 stars	<input type="button" value="Details"/>



(c)



(d)

# Result highlights (1)

- No complete understanding of the result.
  - “Probably based on prices, and distance to the Colosseum, and what do I know” (after having set only the number of stars).
  - Setting max price while skylining on lower prices.
- Some algorithmic features confusing.
  - E.g., non deterministic behavior (toggle/untoggle).
  - E.g., non-monotone result generation.

# Result highlights (2)

- Confusing interaction of skyline and ranking.
  - Representative skyline on stars, sorting on price.
  - After a selection of 5-star hotels, 1-star options coming before other 5-star hotels.
- Difficult to operate explicit skyline selections.
  - Implicit selection works better, but not possible to decouple skyline and filtering.
- Highlighting hotels in the map-view works well.
  - Not necessarily in the list-view.

# Summary

- Skyline operator.
  - Low-parameter selection of best options.
- Practical issues.
  - Complexity, size, embedding and user interfaces.
- Recent advances.
  - Anytime skyline queries.
  - Representative skylines.
  - Aggregate skyline queries.
- Need to test data management operators *in vivo*.

# References

- S. Börzsöny, D. Kossmann, and K. Stocker. **The Skyline operator**, ICDE Conference, 2001.
- M. Magnani, I. Assent, M. L. Mortensen. **Taking the Big Picture: representative skylines based on relevance and diversity**. Under submission, VLDB Journal.
- M. Magnani, I. Assent, M.L. Mortensen. **Anytime skyline processing for interactive systems**. VLDB Workshops (DBRank), 2012.
- Kenneth S. Bøgh, Ira Assent, Matteo Magnani. **Efficient GPU-based skyline computation**. SIGMOD Workshops (DaMoN), 2013.
- M. Magnani, I. Assent. **From stars to galaxies: skyline queries on aggregate data**. EDBT, 2013.
- Magnani, M., Assent, I., Hornbæk, K., Jakobsen, M., Larsen, K. **SkyView: a user evaluation of the skyline operator**. CIKM, 2013.



# The skyline operator

## recent research directions and applications

Matteo Magnani

*matteo.magnani@it.uu.se*

*Computing Science Division  
Uppsala University, Sweden*