

Project Group 2: Visualizing Ranked Theta-Join Results

Zixuan Chen, Aamod Khatiwada

CS 7250 Spring 2021 - Prof. Cody Dunne, Northeastern University

Main Objective

Visually explaining the ranking results of Theta Joins

Background

Background

Theta-join

Car

CarModel	CarPrice
CarA	20,000
CarB	30,000
CarC	50,000

Boat

BoatModel	BoatPrice
Boat1	10,000
Boat2	40,000
Boat3	60,000

Car ⋈ *Boat*
CarPrice ≥ *BoatPrice* → θ

CarModel	CarPrice	BoatModel	BoatPrice
CarA	20,000	Boat1	10,000
CarB	30,000	Boat1	10,000
CarC	50,000	Boat1	10,000
CarC	50,000	Boat2	40,000

Background

Ranked theta-join results between 2 tables

Car

CarModel	CarPrice
CarA	20,000
CarB	30,000
CarC	50,000

Boat

BoatModel	BoatPrice
Boat1	10,000
Boat2	40,000
Boat3	60,000

Ranked by carprice ascending

Ranked by (carprice - boatprice) ascending

This is usually defined to be a link strength of this theta-join
between these two tables

Car ⋈ *Boat*

CarPrice ≥ *BoatPrice*

CarModel	CarPrice	BoatModel	BoatPrice
CarA	20,000	Boat1	10,000
CarB	30,000	Boat1	10,000
CarC	50,000	Boat1	10,000
CarC	50,000	Boat2	40,000

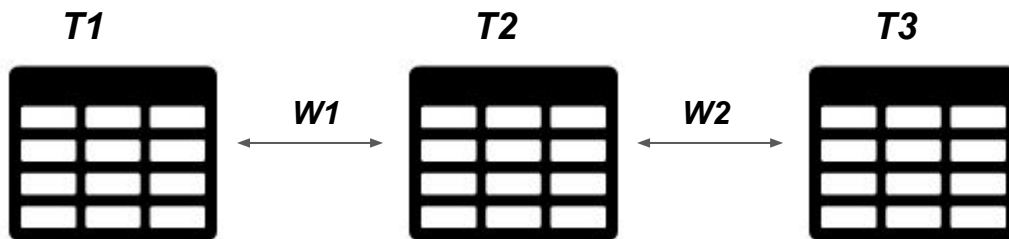
Car ⋈ *Boat*

CarPrice ≥ *BoatPrice*

CarModel	CarPrice	BoatModel	BoatPrice
CarA	20,000	Boat1	10,000
CarC	50,000	Boat2	40,000
CarB	30,000	Boat1	10,000
CarC	50,000	Boat1	10,000

Background

Ranked theta-join results
between multiple tables

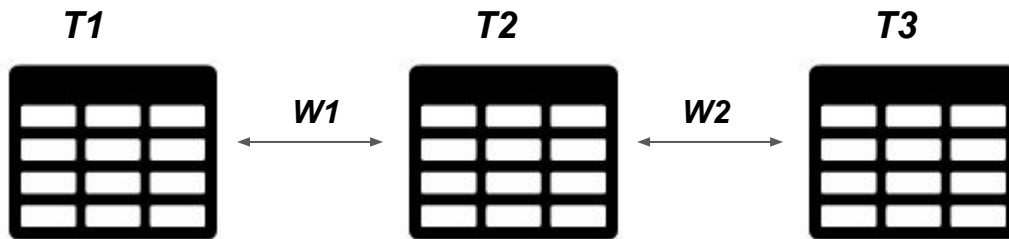


How to rank the join results?

By **a ranking function**. E.g., $f(x) = x.W1 + x.W2$. x is a result.

Background

Ranked theta-join results
between multiple tables



How to rank the join results?

By a **ranking function**. E.g., $f(x) = x.W1 + x.W2$. x is a result.

Application

This ranking can be used for route plan. Let's say you're hungry and want to go outside for dinner. After that you plan to go to a supermarket and then go back home.

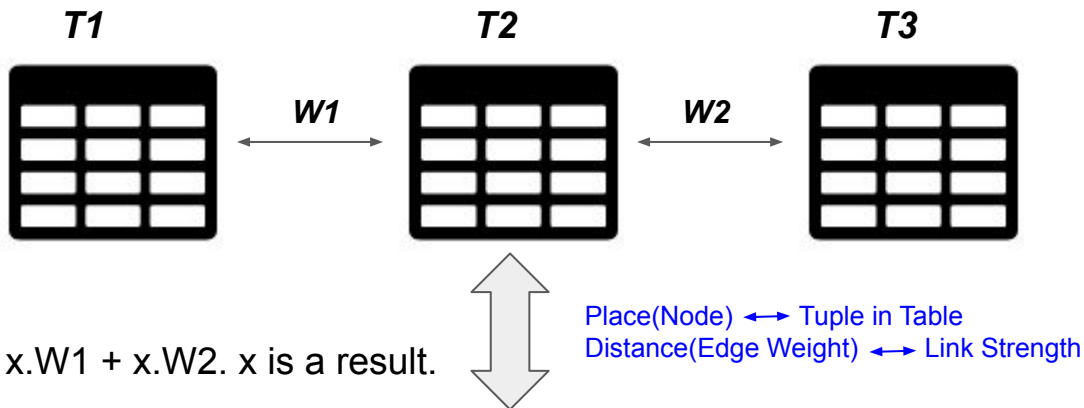
Since you're quite hungry, you need the distance from your home to the restaurant to be as short as possible but for the distance 2 and 3, you don't really care much.

Ranking function $f(x) = 0.8 * x.Distance1 + 0.1 * x.Distance2 + 0.1 * x.Distance3$



Background

Ranked theta-join results
between multiple tables



How to rank the join results?

By a **ranking function**. E.g., $f(x) = x.W1 + x.W2$. x is a result.

Application

This ranking can be used for route plan. Let's say you're hungry and want to go outside for dinner. After that you plan to go to a supermarket and then go back home.

Since you're quite hungry, you need the distance from your home to the restaurant to be as short as possible but for the distance 2 and 3, you don't really care much.

Ranking function $f(x) = 0.8 * x.Distance1 + 0.1 * x.Distance2 + 0.1 * x.Distance3$



Partner

Partner

Prof. Mirek Riedewald
Data Lab @ Northeastern

- Author of several papers about ranked enumeration over joins
- Thought it would be good to use visualizations to help him better understand the ranking
- Several questions he proposed:
 - Can you use visualizations to help me understand **why certain results are at the top of a ranking?**
 - Can you use visualizations to show the **influence of different ranking functions on the ranked join results?**

Tasks

Tasks

- Present
 - Present the ranked results with link strengths and the parameter coefficients
- Lookup
 - Look up for the appearances of one result in different rankings
- Compare
 - Compare the ranked results of multiple functions

Data

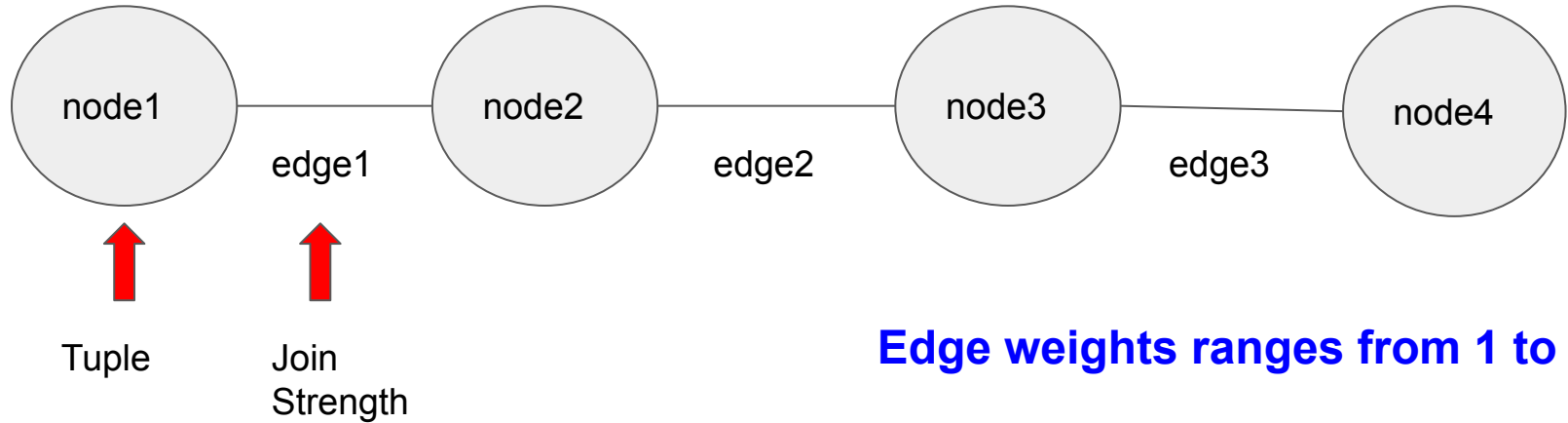
Original Data: Bitcoin OTC Trust Dataset

- **Semantics:** Trust Score in Bitcoin Transaction
- **Nodes:** Bitcoin User IDs
- **Edges:** Trust Score ranging from -10 to +10
- **Sample Tuple:**
 - 2,21,5,1289370557



Preprocessed Dataset:

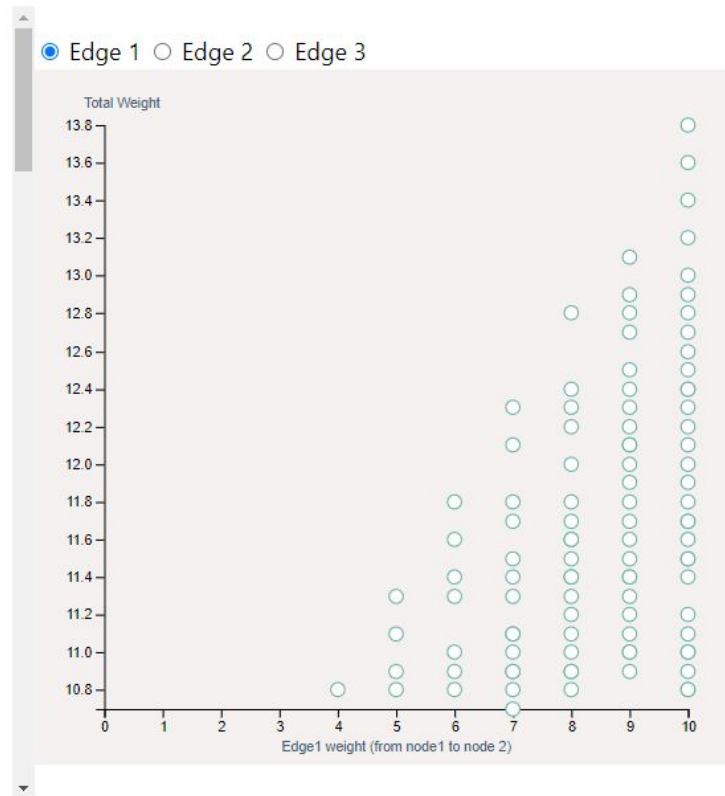
- 3-hop paths
- Generated Schema: node1, node2, node3, node4, edge1, edge2, edge3



We use graph dataset to simulate the Theta-Join Condition

Visualizations

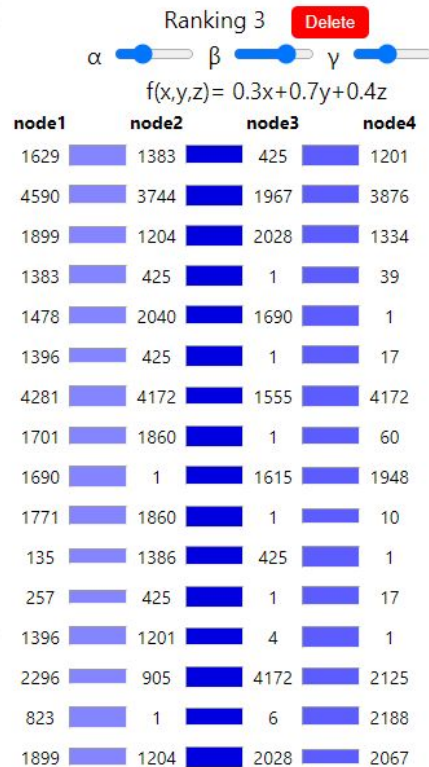
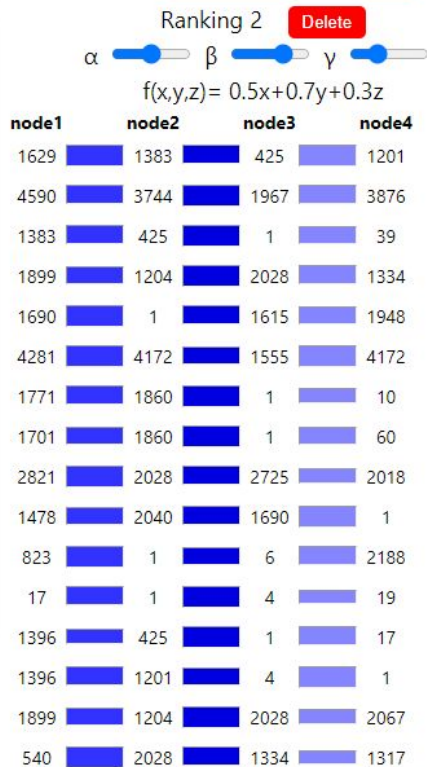
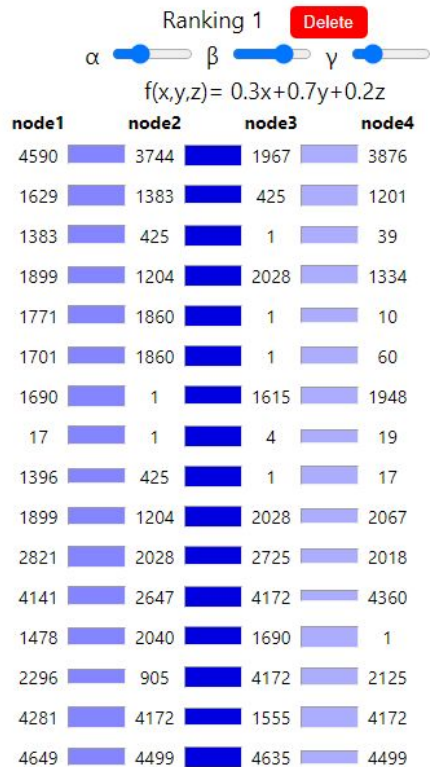
Visualization 1: Single Ranking



Visualization 2: Multiple Ranking

K:

All three functions used



Encodings

K: 100

Order: Descending

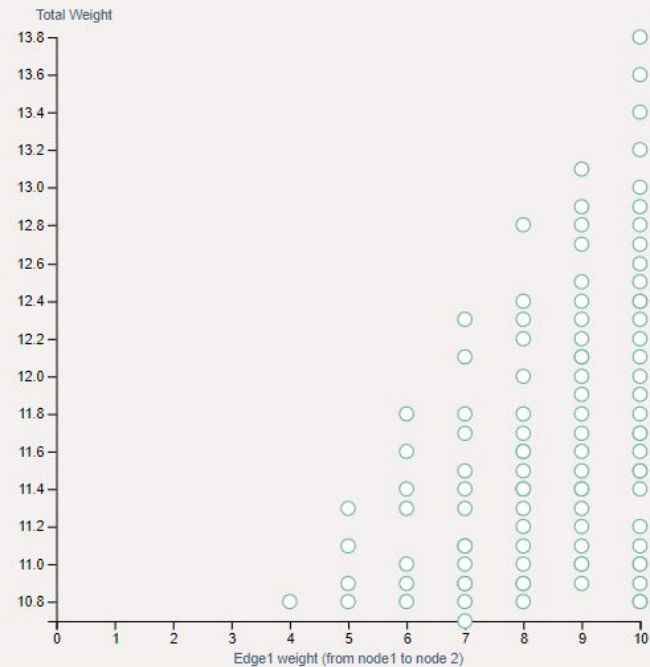
α β γ

$$f(x,y,z) = 0.5x + 0.2y + 0.7z$$

node1	node2	node3	node4
1629	1383	425	1201
4281	4172	1555	4172
5411	4649	3429	3649
4609	3837	4365	192
1396	1201	4	1
486	1	61	109
3219	1018	2763	2127
823	1	6	2188
353	1	1620	2200
1478	2040	1690	1
4590	3744	1967	3876
2214	2028	1	17
4698	4635	4499	4400
1383	1566	1620	1386
425	1317	905	1316
5636	1	21	8
3759	2934	4291	3897

Odd Columns: User Ids

Edge 1 Edge 2 Edge 3



K:

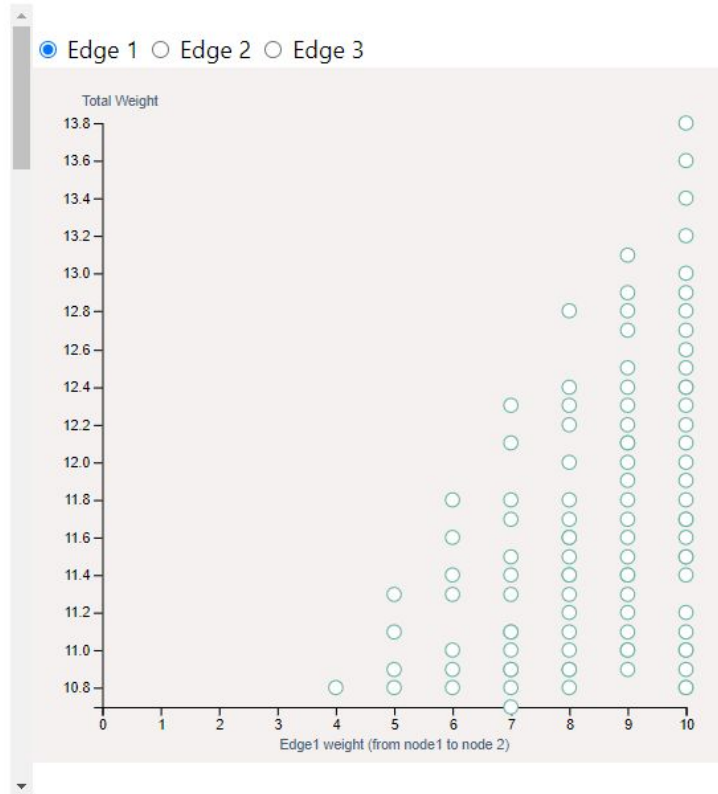
Order:

α β γ

$$f(x,y,z) = 0.5x + 0.2y + 0.7z$$

node1	node2	node3	node4
1629	1383	425	1201
4281	4172	1555	4172
5411	4649	3429	3649
4609	3837	4365	192
1396	1201	4	1
486	1	61	109
3219	1018	2763	2127
823	1	6	2188
353	1	1620	2200
1478	2040	1690	1
4590	3744	1967	3876
2214	2028	1	17
4698	4635	4499	4400
1383	1566	1620	1386
425	1317	905	1316
5636	1	21	8
3759	2934	4291	3897

↑ ↑ ↑
Even Columns: Edges



K:

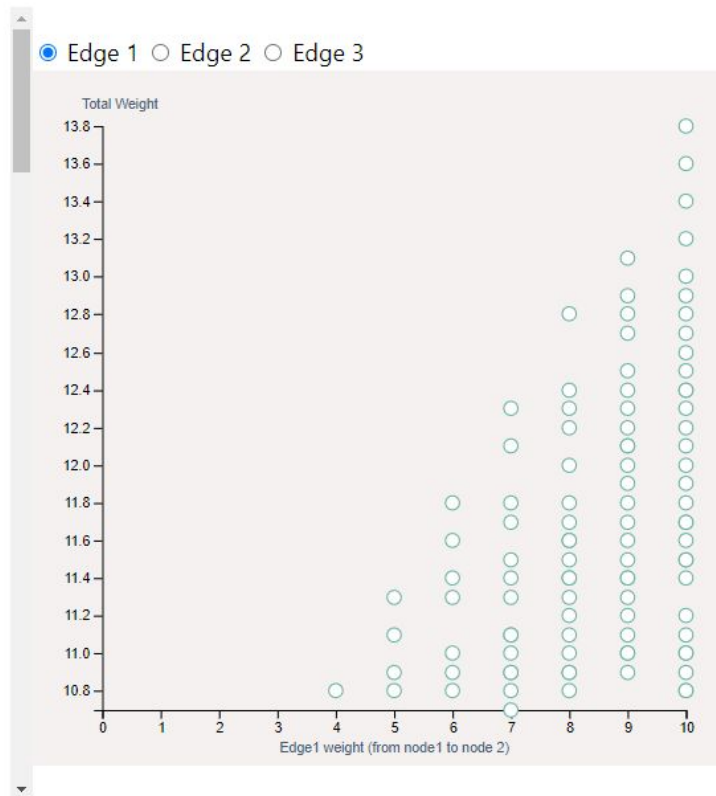
Order:

α β γ

$$f(x,y,z) = 0.5x + 0.2y + 0.7z$$

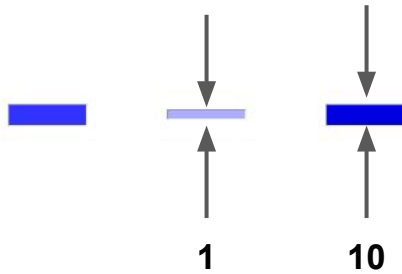
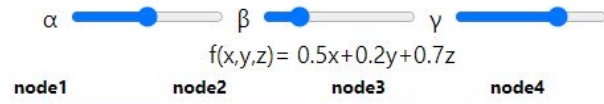
node1	node2	node3	node4
1629	1383	425	1201
4281	4172	1555	4172
5411	4649	3429	3649
4609	3837	4365	192
1396	1201	4	1
486	1	61	109
3219	1018	2763	2127
823	1	6	2188
353	1	1620	2200
1478	2040	1690	1
4590	3744	1967	3876
2214	2028	1	17
4698	4635	4499	4400
1383	1566	1620	1386
425	1317	905	1316
5636	1	21	8
3759	2934	4291	3897

Size: Edge Weight

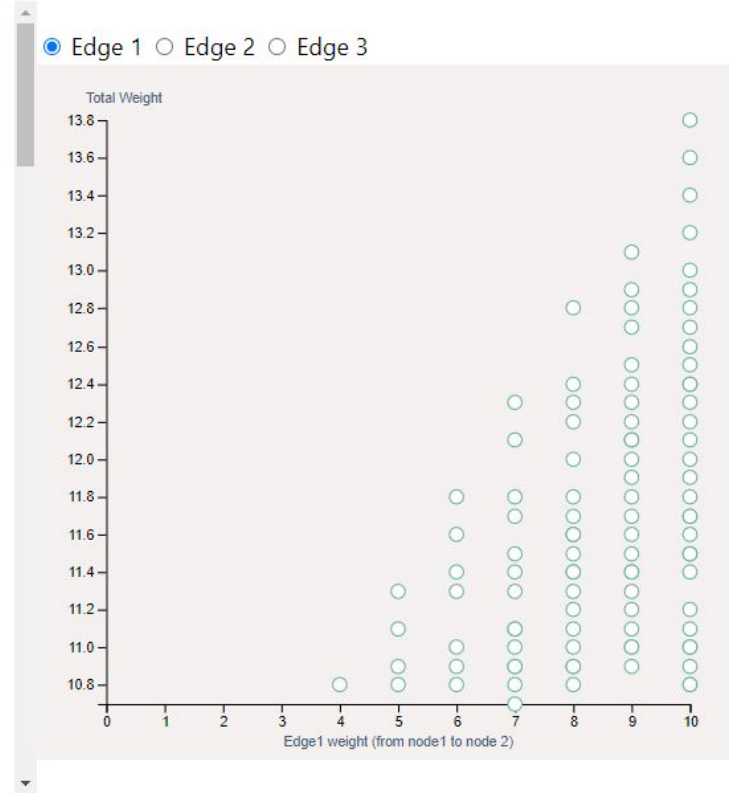


K: 100

Order: Descending



Size: Edge Weight



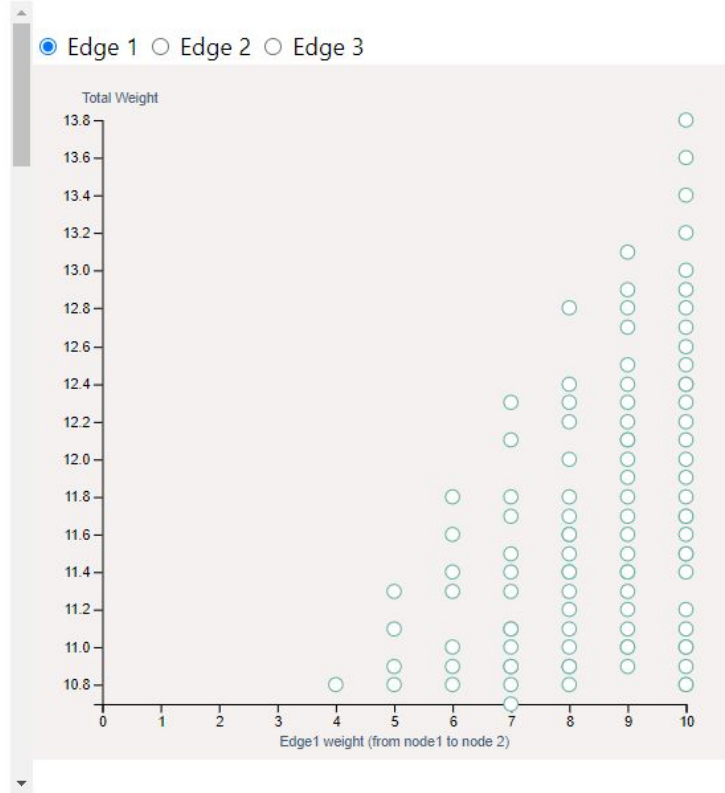
K: 100

Order: Descending

α β γ

$$f(x,y,z) = 0.5x + 0.2y + 0.7z$$

node1	node2	node3	node4
1629	1383	425	1201
4281	4172	1555	4172
5411	4649	3429	3649
4609	3837	4365	192
1396	1201	4	1
486	1	61	109
3219	1018	2763	2127
823	1	6	2188
353	1	1620	2200
1478	2040	1690	1
4590	3744	1967	3876
2214	2028	1	17
4698	4635	4499	4400
1383	1566	1620	1386
425	1317	905	1316
5636	1	21	8
3759	2934	4291	3897



Luminance: Importance of a Column

K:

Order:

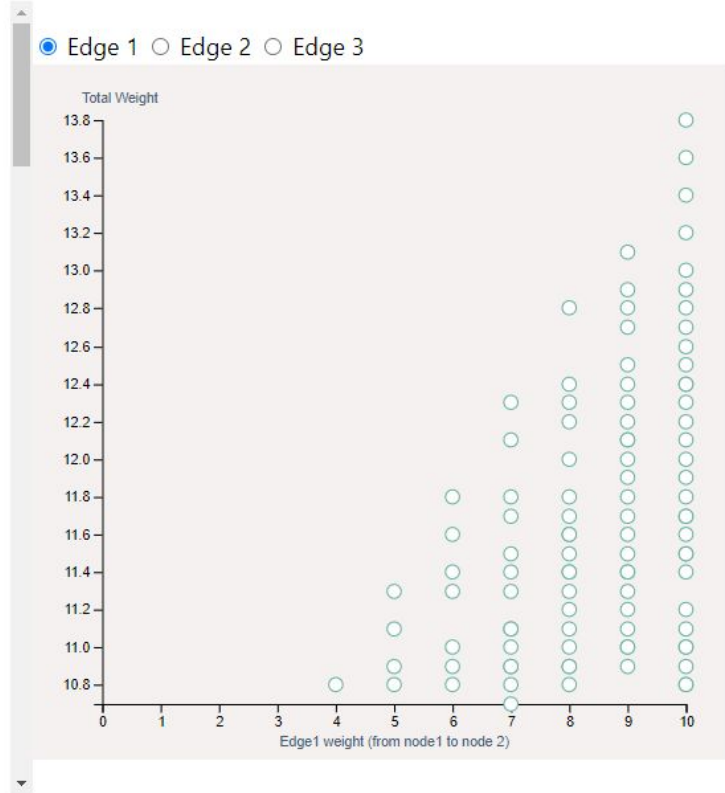
Least
Important



Most
Important



1396		1201		4		1
486		1		61		109
3219		1018		2763		2127
823		1		6		2188
353		1		1620		2200
1478		2040		1690		1
4590		3744		1967		3876
2214		2028		1		17
4698		4635		4499		4400
1383		1566		1620		1386
425		1317		905		1316
5636		1		21		8
3759		2934		4291		3897



Luminance: Importance of a Column

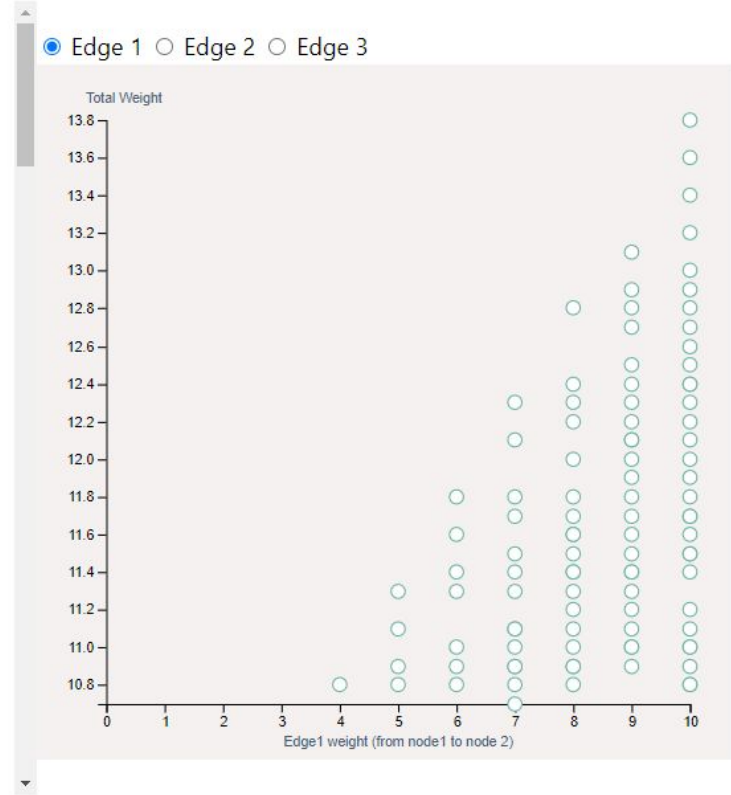
K:

Order:

α β γ

$$f(x,y,z) = 0.5x + 0.2y + 0.7z$$

node1	node2	node3	node4
1629	1383	425	1201
4281	4172	1555	4172
5411	4649	3429	3649
4609	3837	4365	192
1396	1201	4	1
486	1	61	109
3219	1018	2763	2127
823	1	6	2188
353	1	1620	2200
1478	2040	1690	1
4590	3744	1967	3876
2214	2028	1	17
4698	4635	4499	4400
1383	1566	1620	1386
425	1317	905	1316
5636	1	21	8
3759	2934	4291	3897



X-Axis: Edge Weight
Y-Axis: Total Weight = $f(x,y,z)$

Interactions

Visualization 1: Single Ranking

K : 100

Order : Descending

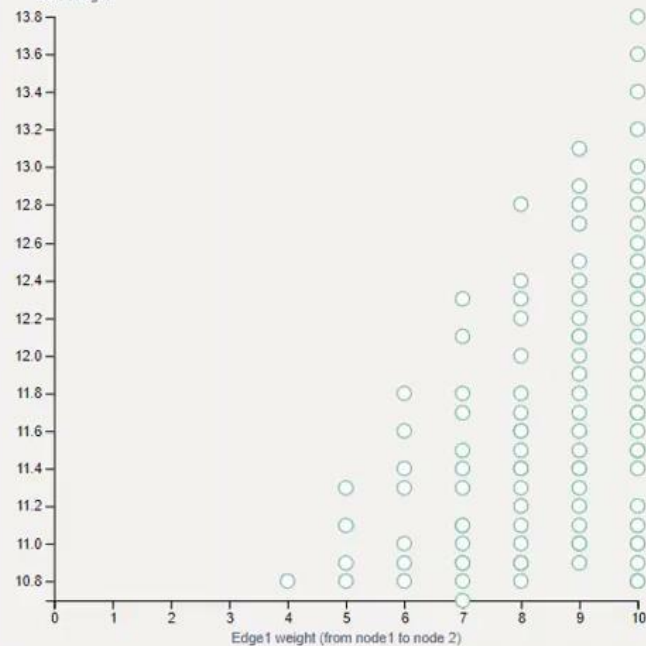
α β γ

$$f(x,y,z) = 0.5x + 0.2y + 0.7z$$

node1	node2	node3	node4
1629	1383	425	1201
4281	4172	1555	4172
5411	4649	3429	3649
4609	3837	4365	192
1396	1201	4	1
486	1	61	109
3219	1018	2763	2127
823	1	6	2188
353	1	1620	2200
1478	2040	1690	1
4590	3744	1967	3876
2214	2028	1	17
4698	4635	4499	4400
1383	1566	1620	1386
425	1317	905	1316
5636	1	21	8
3759	2934	4291	3897

Edge 1 Edge 2 Edge 3

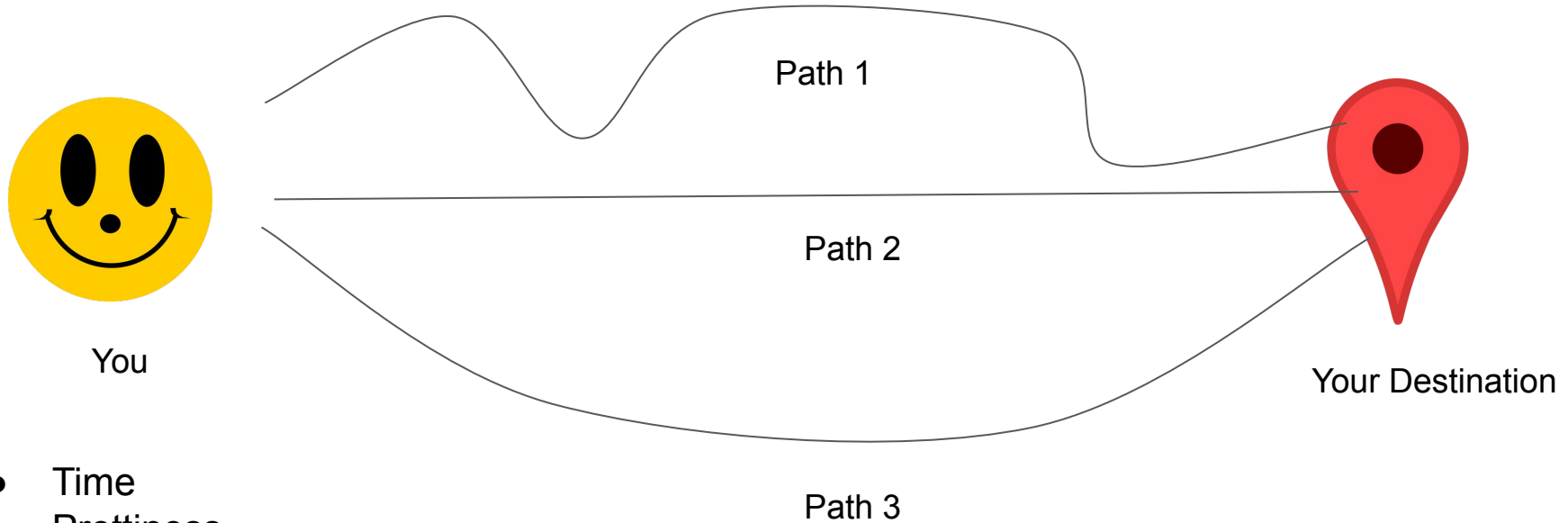
Total Weight



Other Applications

Other Applications

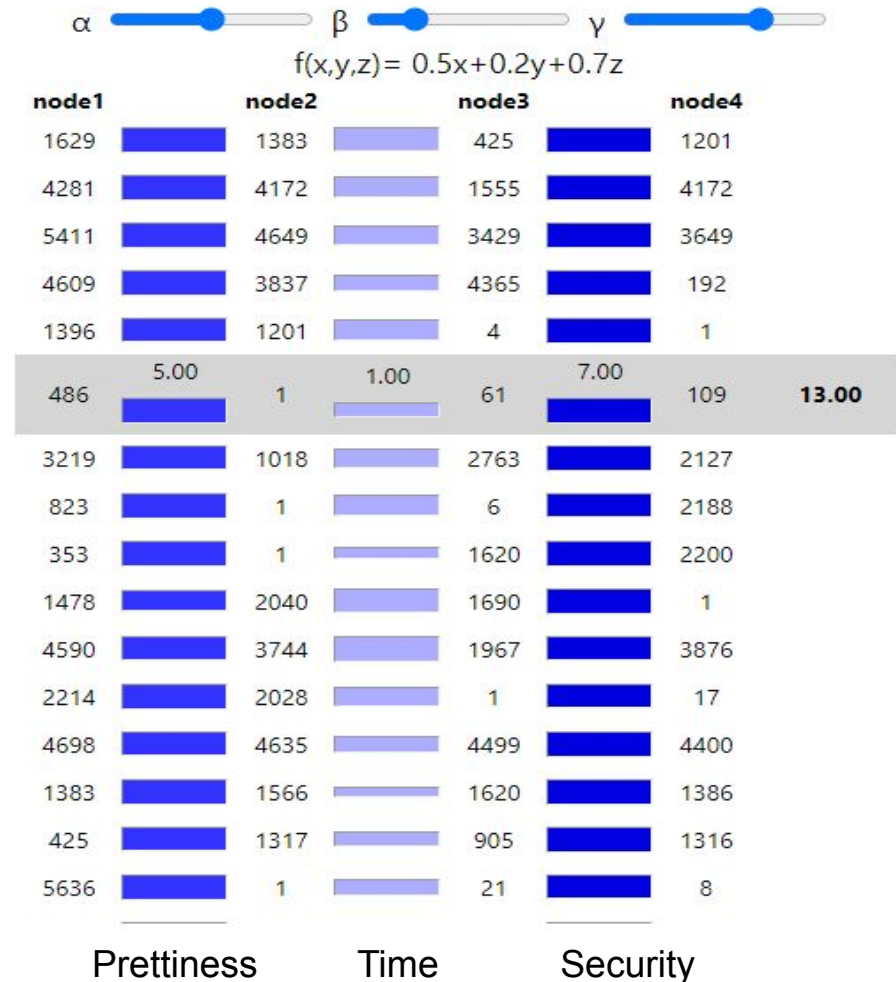
- Making decision on the basis of multiple parameters



- Time
- Prettiness
- Security

Other Applications:

Selecting the best path from source to destination on the basis of prettiness, Time and Security.



Conclusion:

Conclusion

- Helpful in understanding the outliers
- Effective in testing the ranking functions
- Performs well in comparing the multiple functions

Thank you