Alexander Lex
@alexander_lex
http://alexander-lex.net

Enabling Discovery through Interactive Visual Data Analysis

THE UNIVERSITY OF UTAH

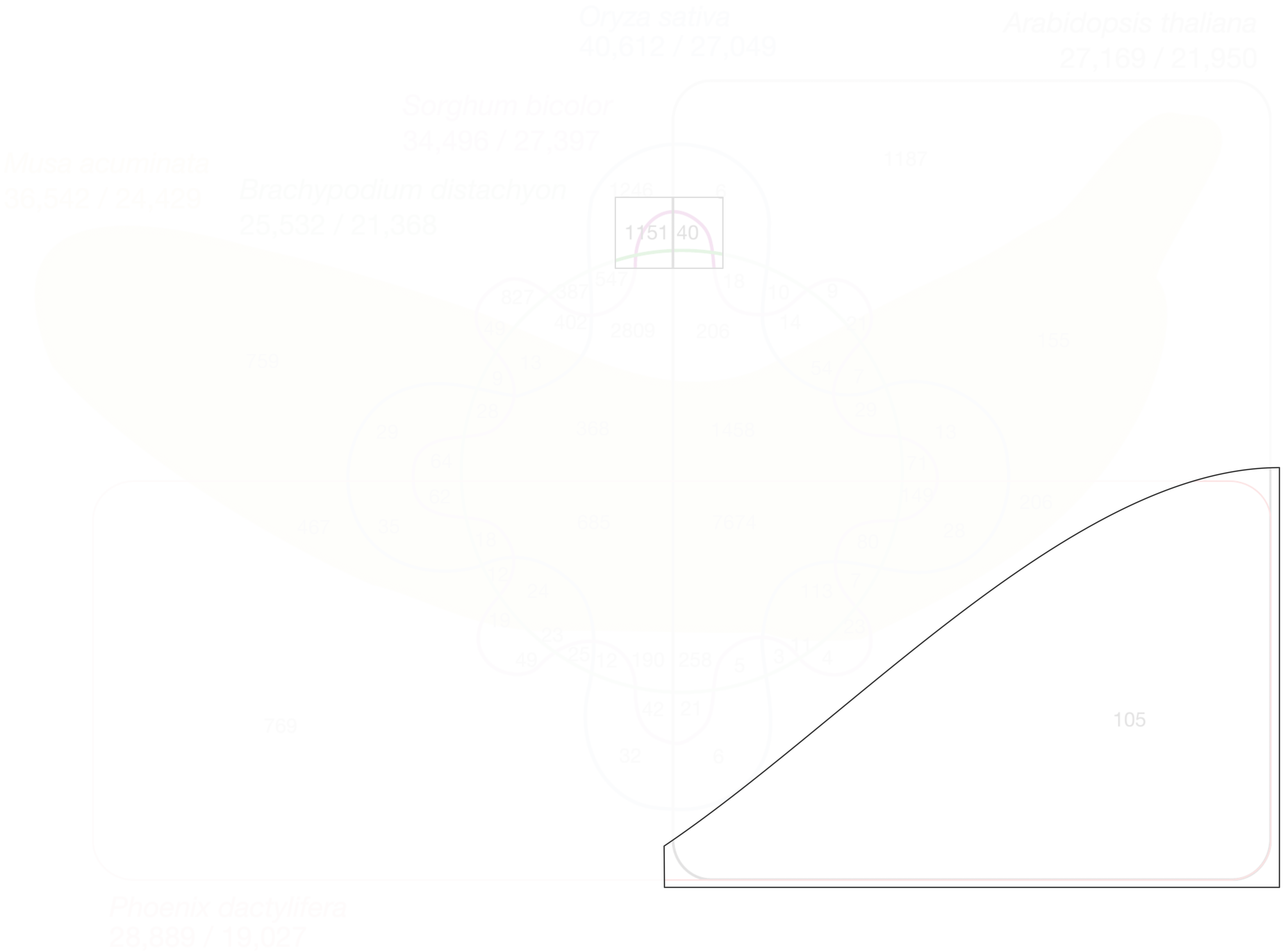
visualization *pictures*

The purpose of computing is insight, not numbers.

- Richard Wesley Hamming

- Card, Mackinlay, Shneiderman

Banana	***M. acuminata***
Date	***P. dactylifera***
Cress	***Arabidopsis thaliana***
Rice	***Oryza sativa***
Sorghum	***Sorghum bicolor***
Brome	***Brachypodium distachyon***



[D'Hont et al., Nature, 2012]

Good Data Visualization

... makes data accessible

... combines strengths of humans and computers

... enables insight

... communicates

Why Visualize?

To inform humans: Communication

How did the unemployment and labor force develop over the last years?

When questions are not well defined: Exploration

Which combination of gene mutations is a good cancer marker?

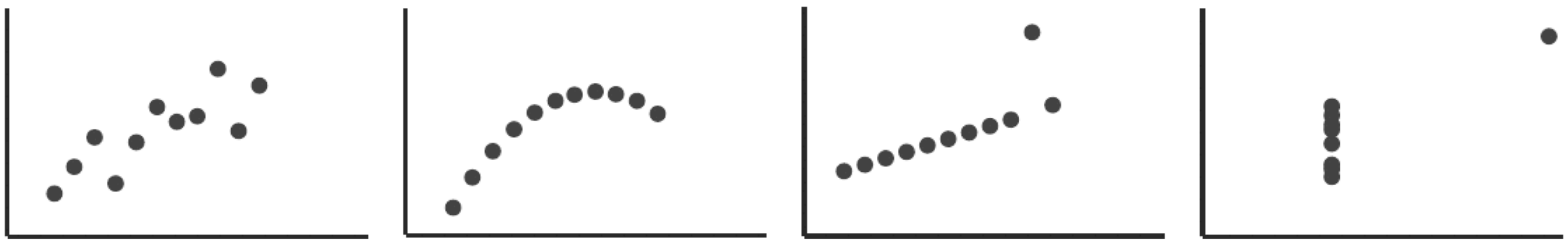
Which drug can help patient X?

Can We Trust Statistics?

I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	8	6.89

Mean x: 9 y: 7.50
Variance x: 11 y: 4.122
Correlation x – y: 0.816
Linear regression: y = 3.00 + 0.500x

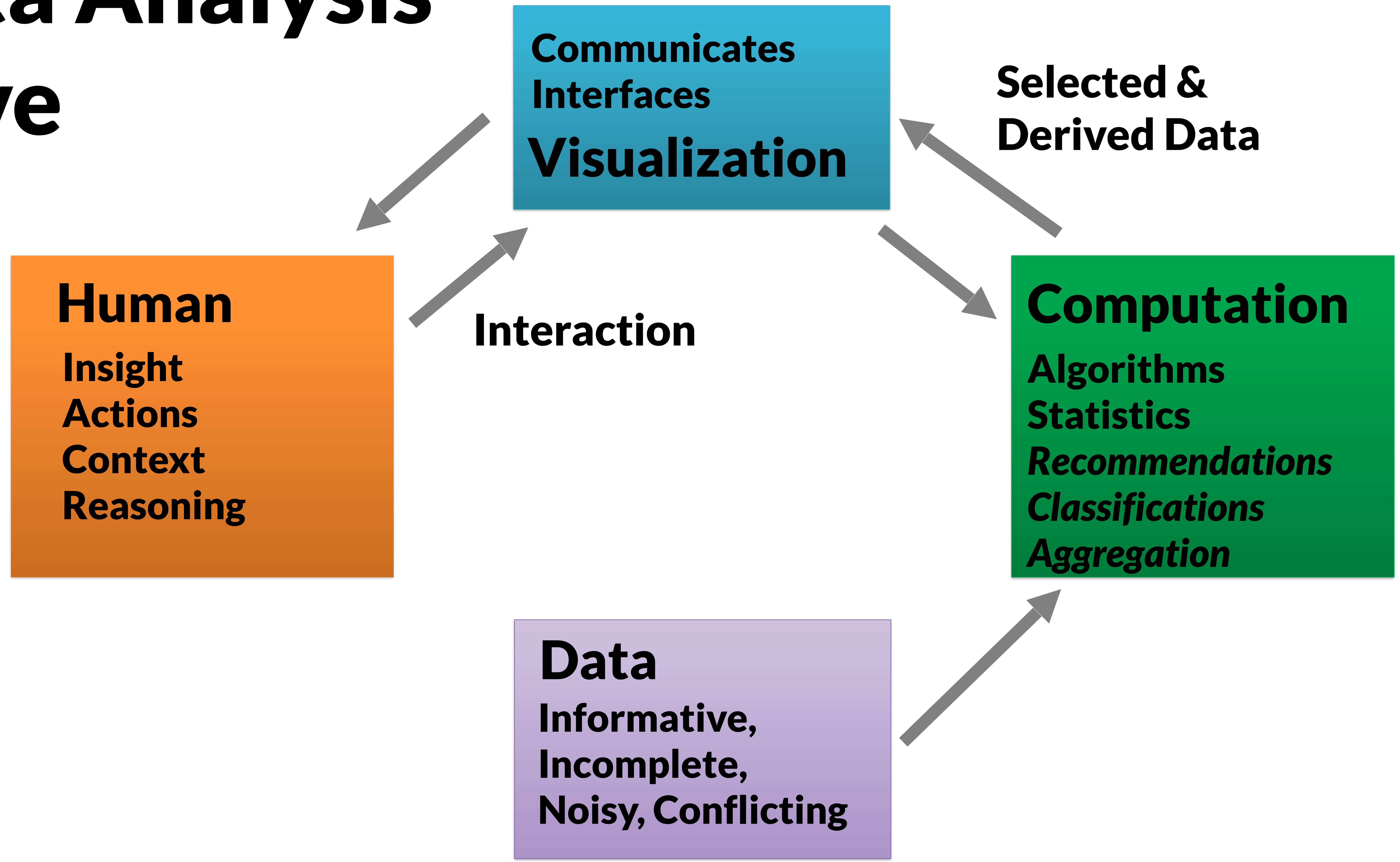
Anscombe’s Quartett



Mean x: 9 y: 7.50
Variance x: 11 y: 4.122
Correlation x – y: 0.816
Linear regression: y = 3.00 + 0.500x

Interacting with Data

The Future of Data Analysis is (also) Interactive



Visualization =

Human Data Interaction

visualization design lab

www.sci.utah.edu


Miriah Meyer


Alexander Lex

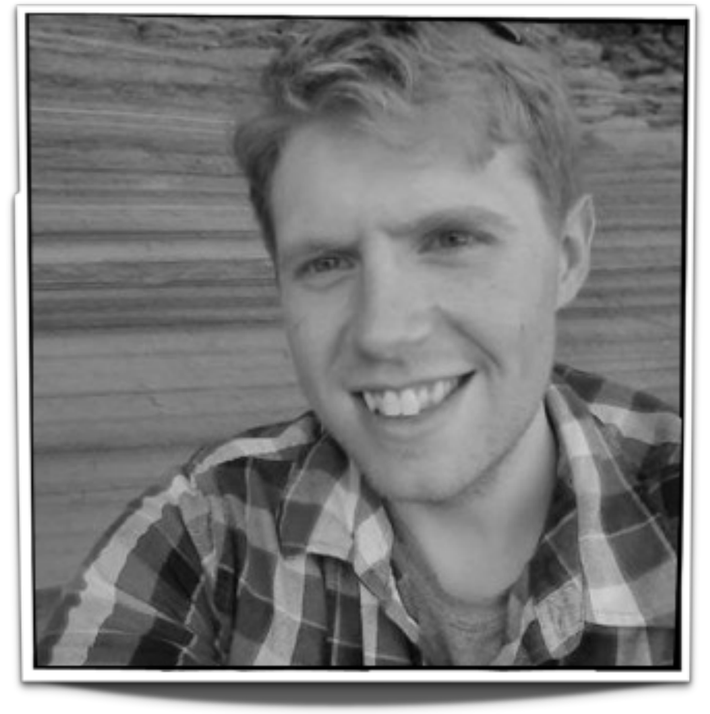

Ethan Kerzner


Alex Bigelow


Sean McKenna


Sam Quinan


Nina McCurdy


Jimmy Moore


Asmaa Aljuhani


Carolina Nobre


Cameron Waller


Magdalena Schwarzl

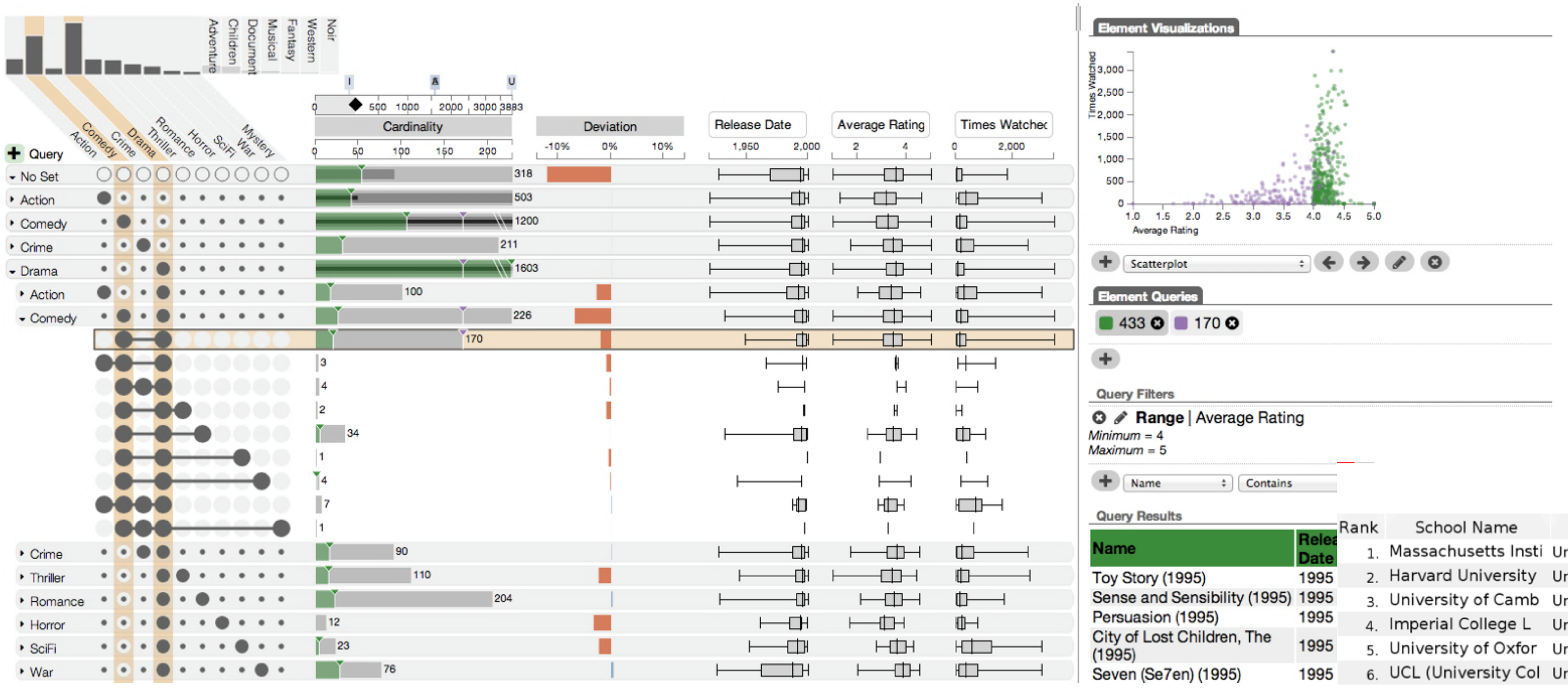

Sunny Hardasani

http://vdl.sci.utah.edu/

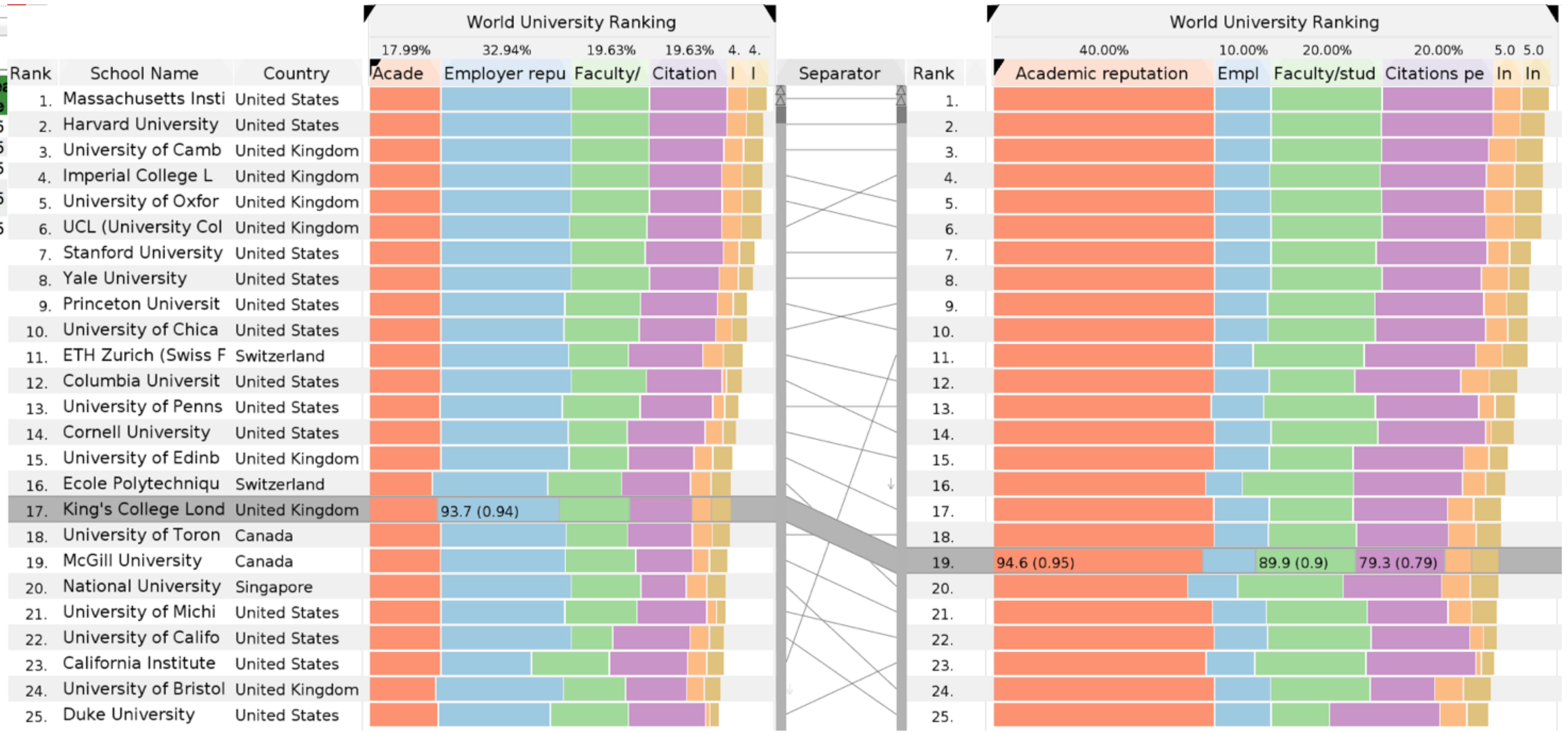
Research Areas

Large, Multivariate (Biological) Networks

Multidimensional Data

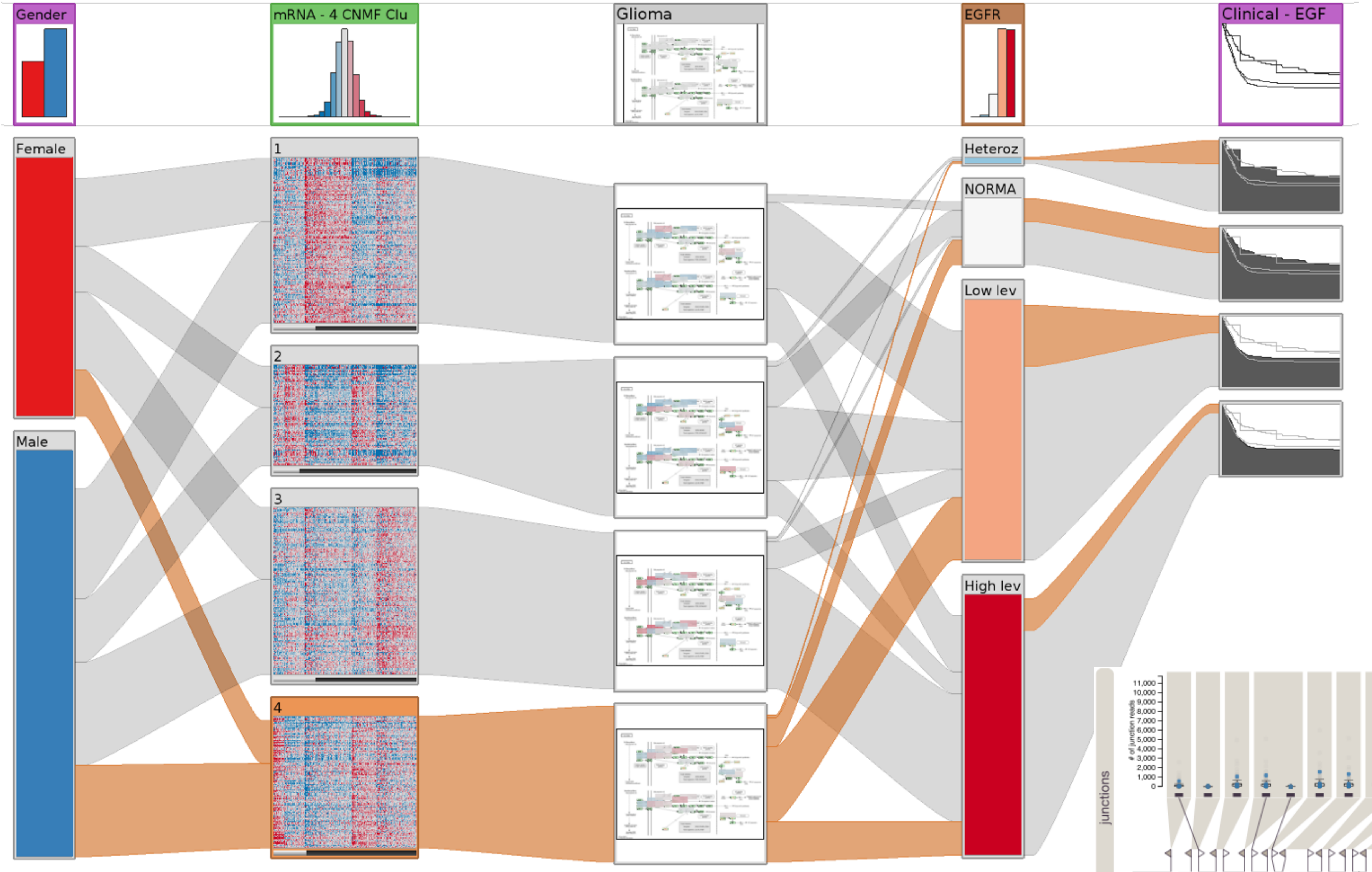


Multivariate Rankings

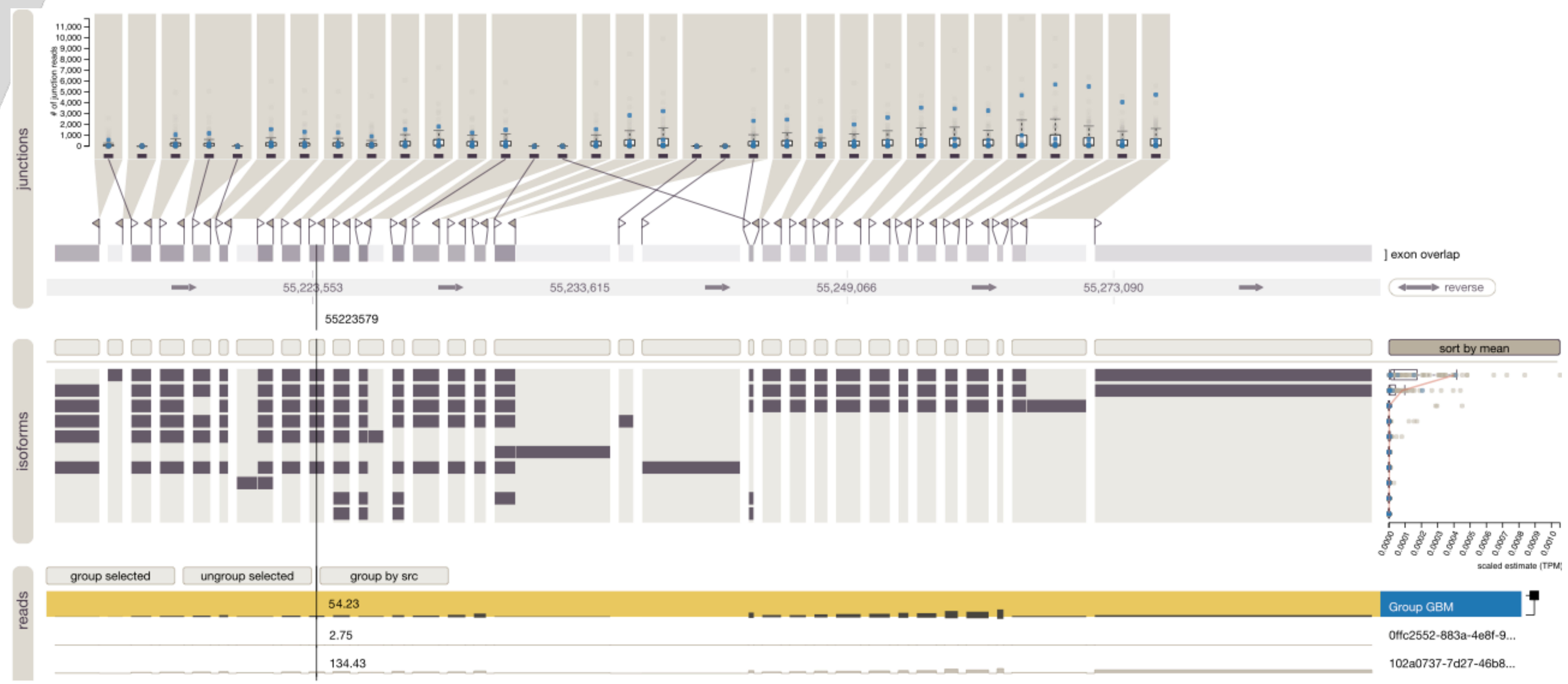


Set Visualization

Genomic Data



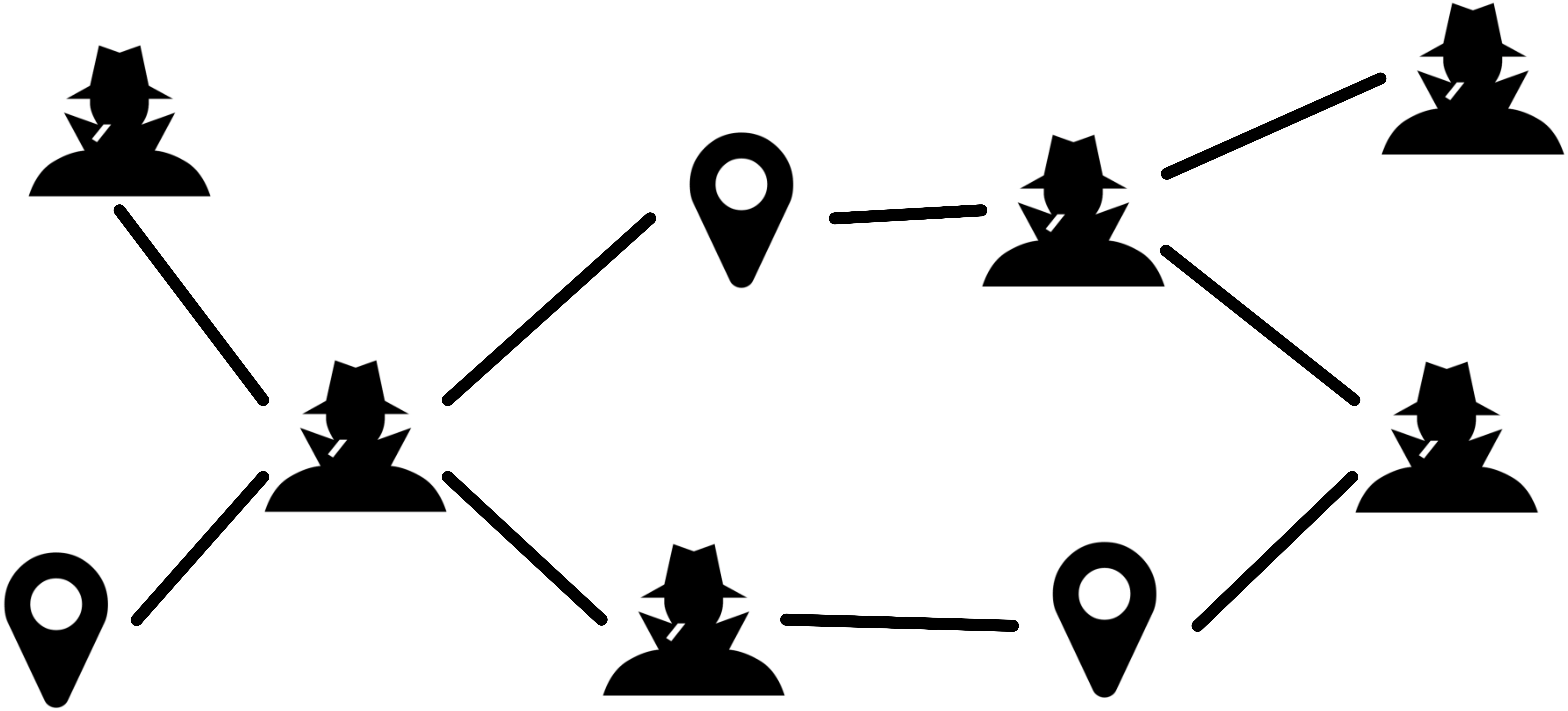
Alternative Splicing / mRNA-seq



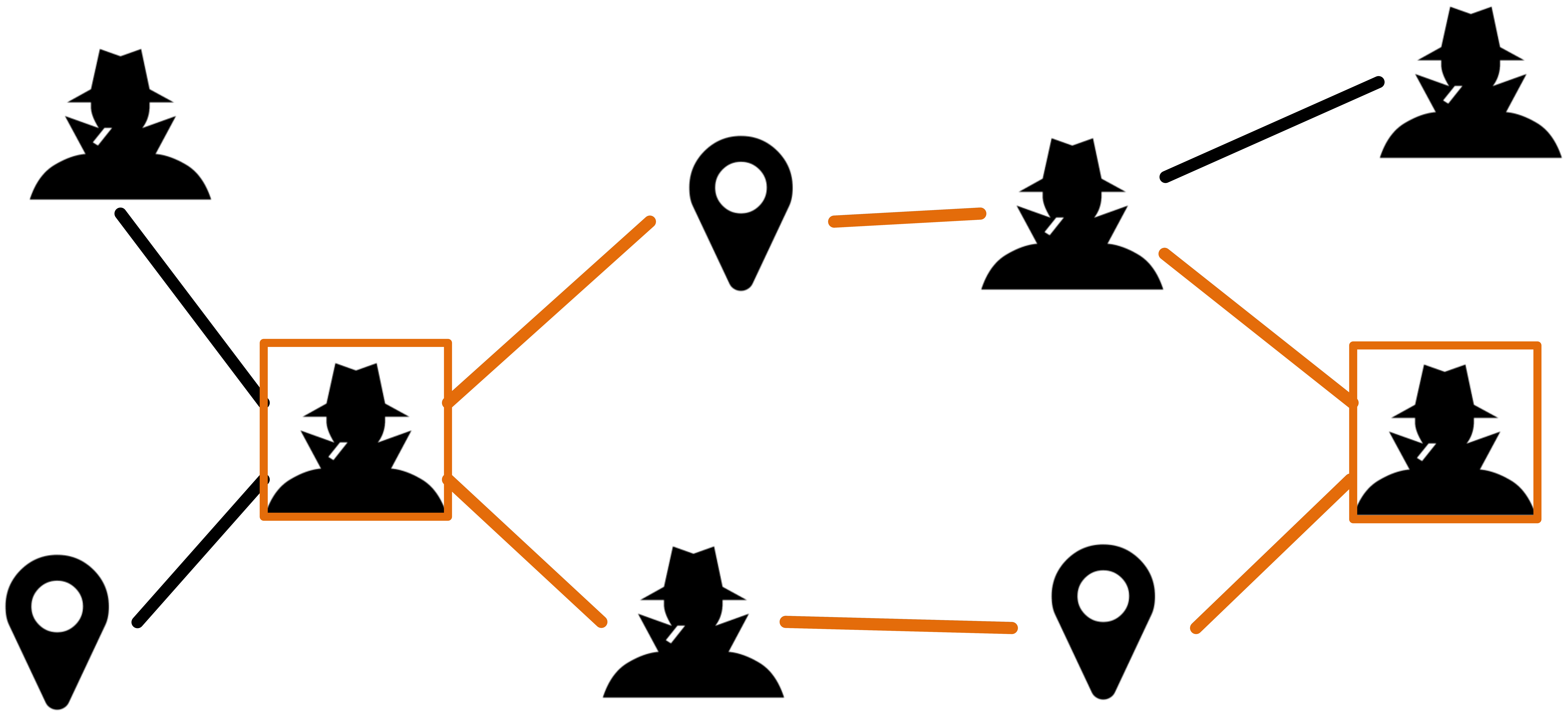
Cancer Subtypes / Omics Clustering and Stratification

[EuroVis '16]
Honorable Mention Award

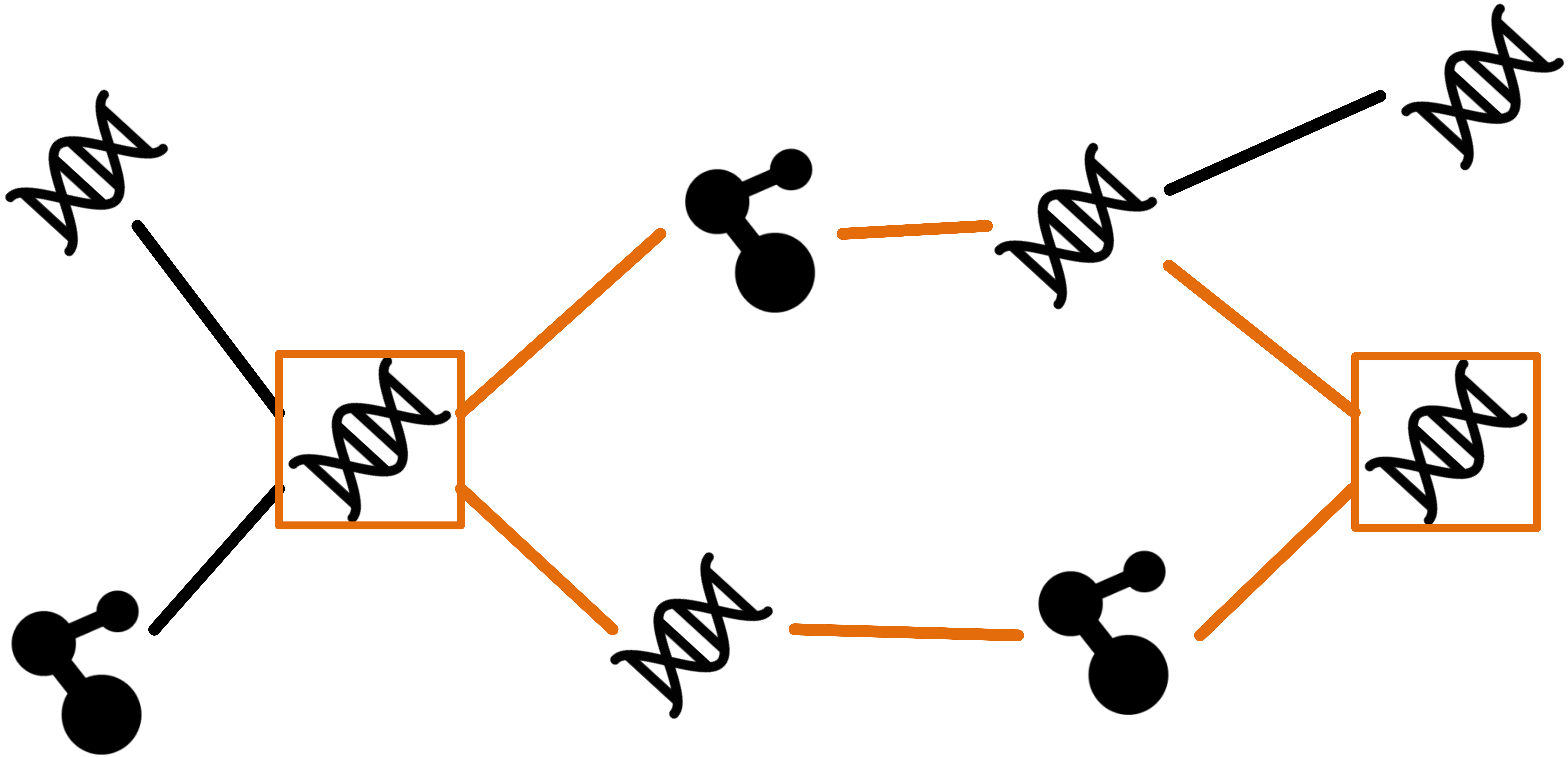
Pathfinder:
Visual Analysis of Paths in Graphs



Intelligence Data: How are two suspects connected?



Intelligence Data: How are two suspects connected?



Biological Network: How do two genes interact?

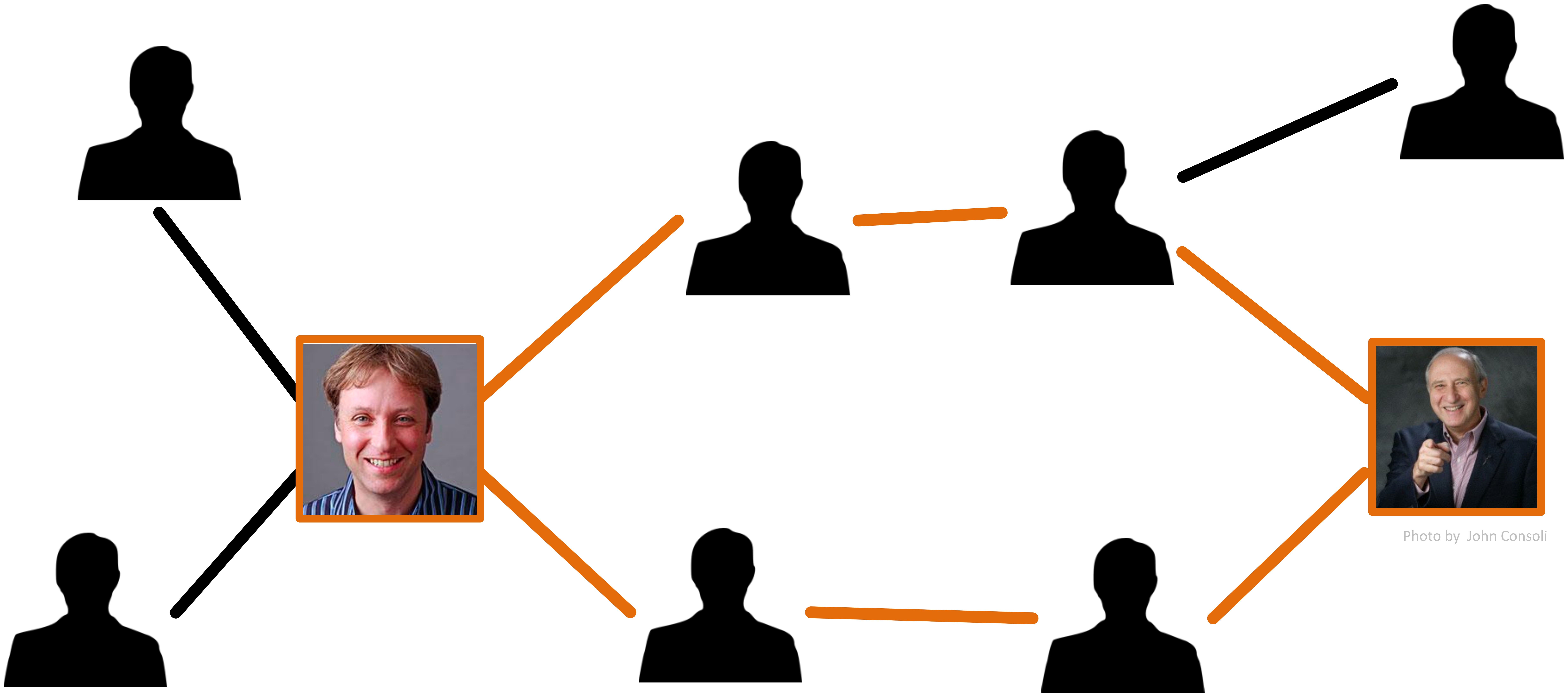
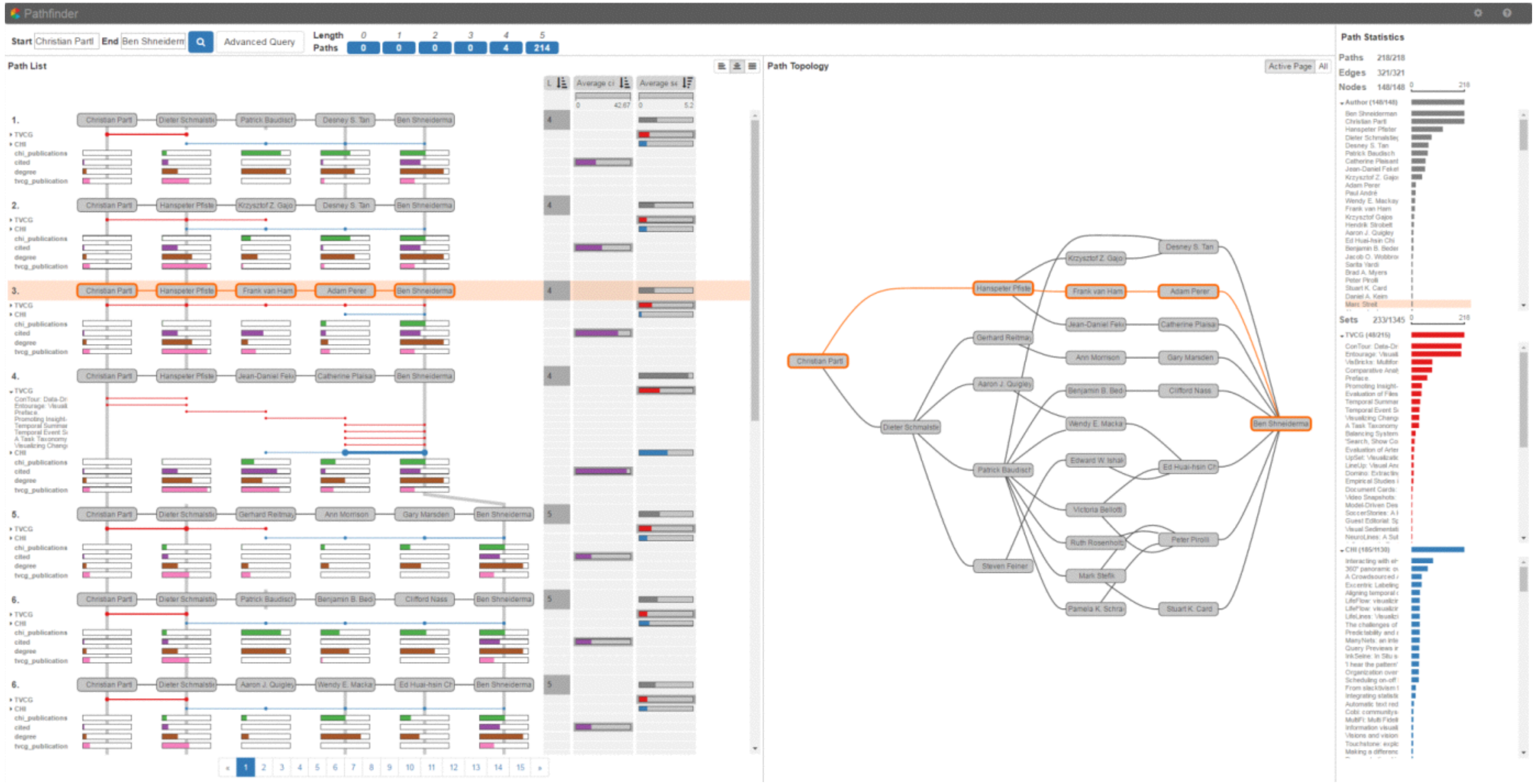


Photo by John Consoli

Coauthor Network: How is HP Pfister connected to Ben Shneiderman?

Pathfinder



Visual Analysis of Paths in Large Multivariate Graphs

Challenge 1: Graph Size

How can we deal with graphs too large to sensibly render at once?

Approach: **(Path) Queries**

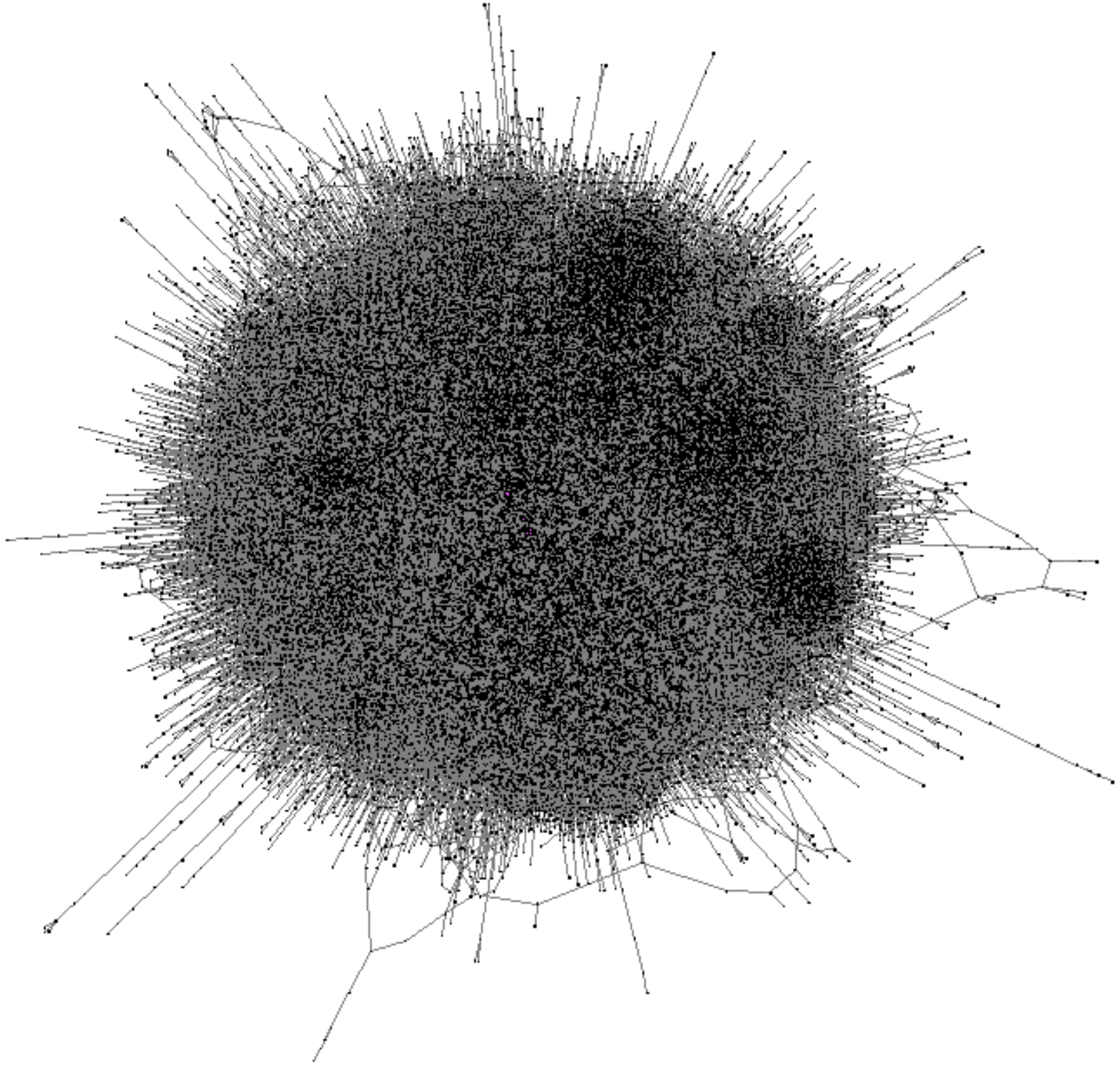
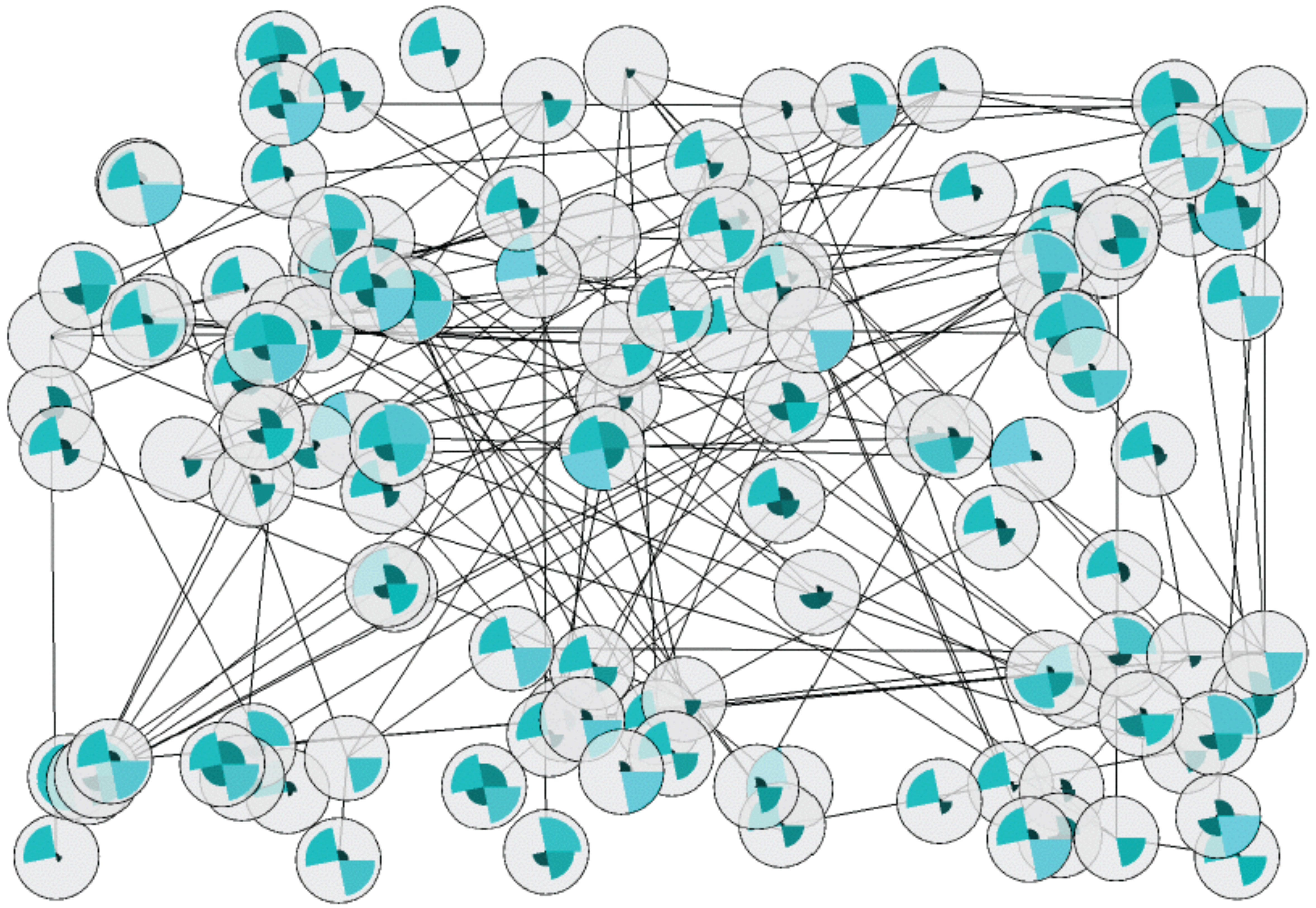


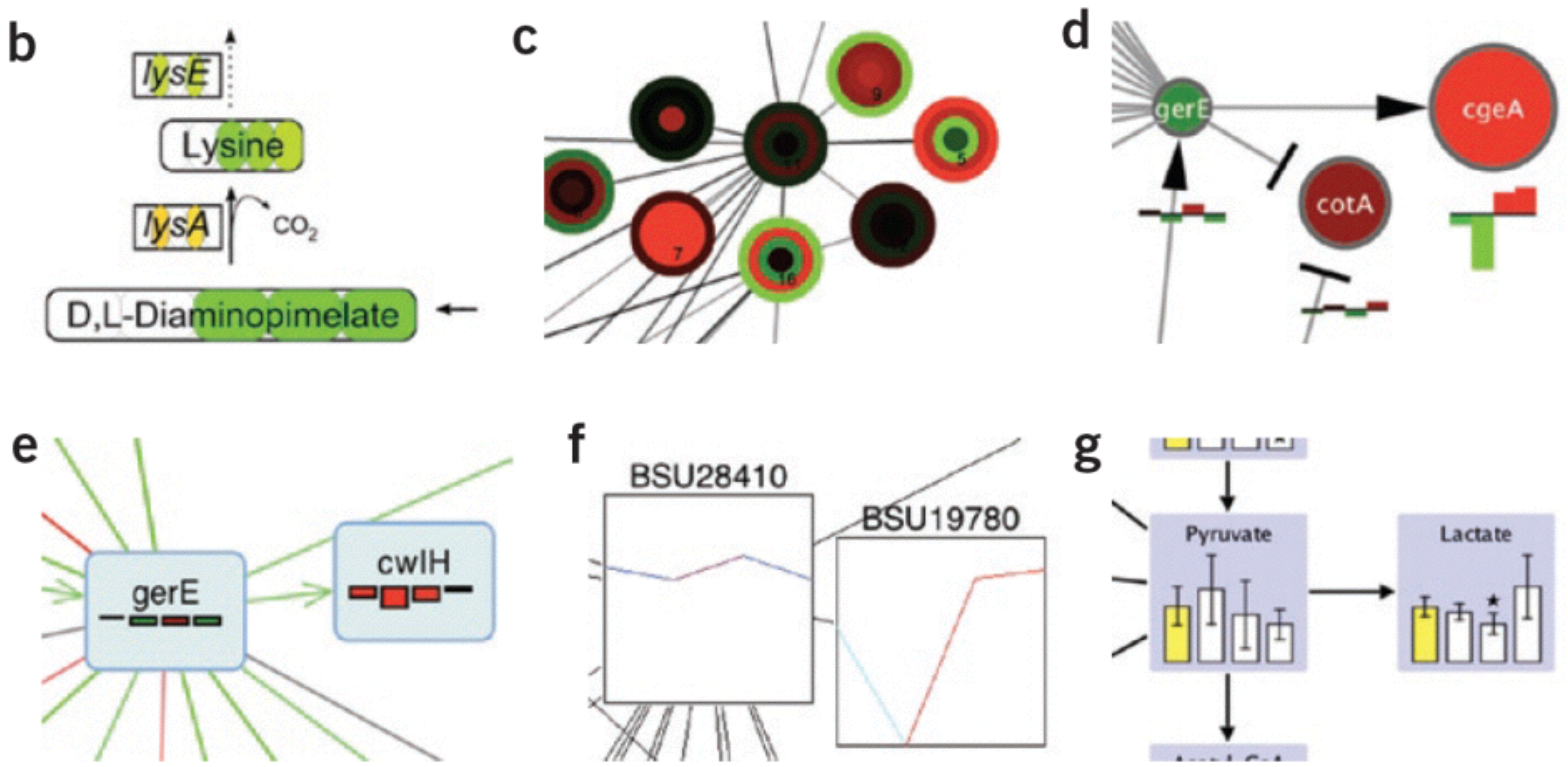
Image Source:
http://www.thenetworkthinkers.com/2013/03/big-data.html

Challenge 2: Multivariate Attributes

How can we deal with graphs that contain **rich attribute data**?



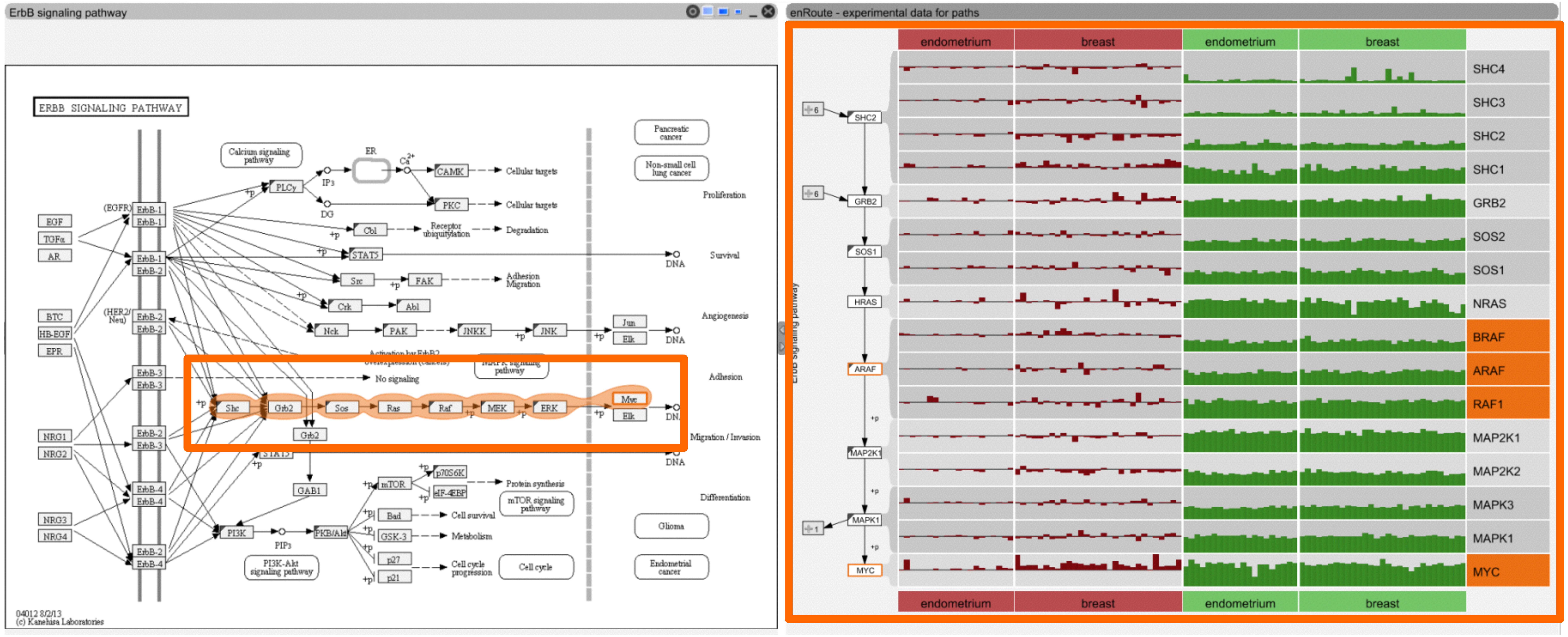
[McDonnel2009]



[Gehlenborg2010]

Approach: Attribute Visualization for Paths

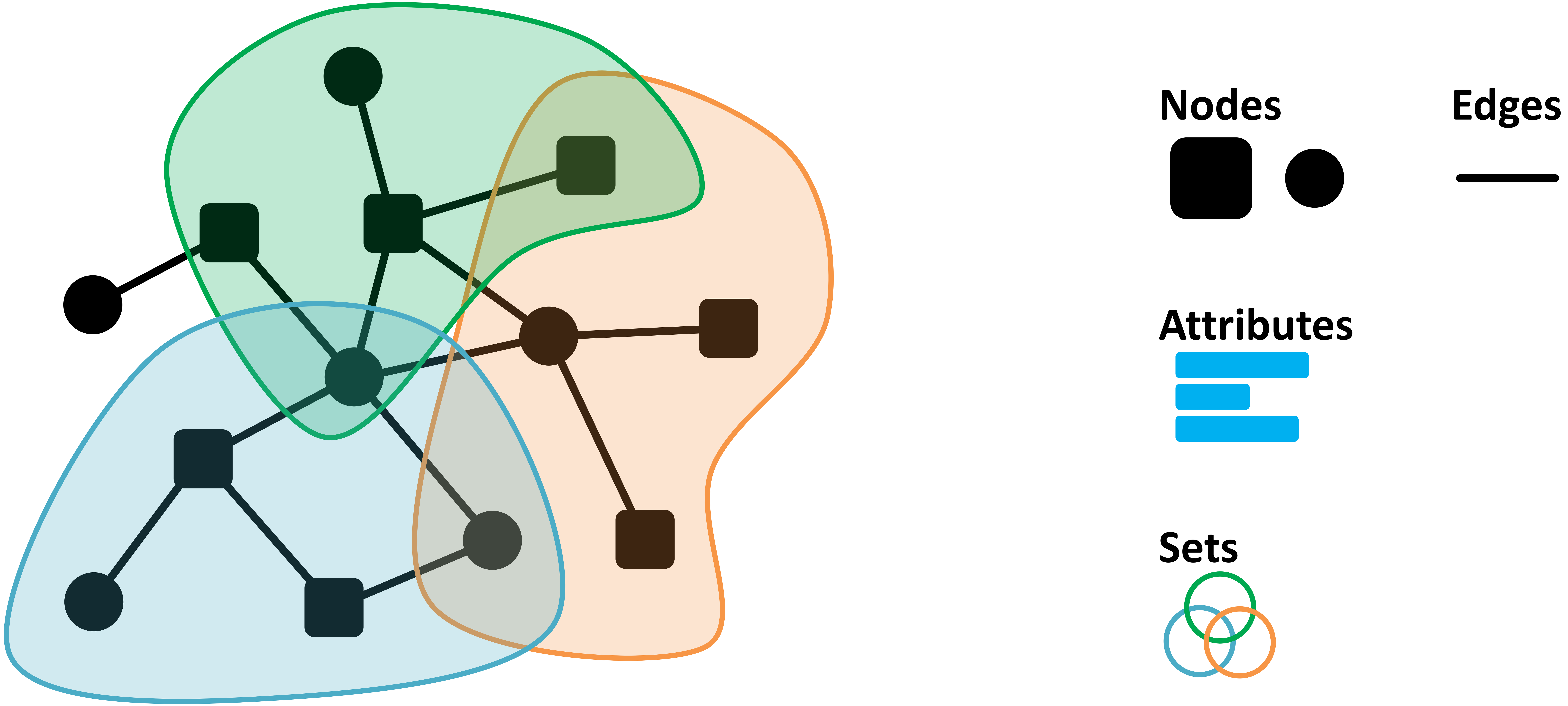
EnRoute http://enroute.caleydo.org



[Partl2012]

The Data

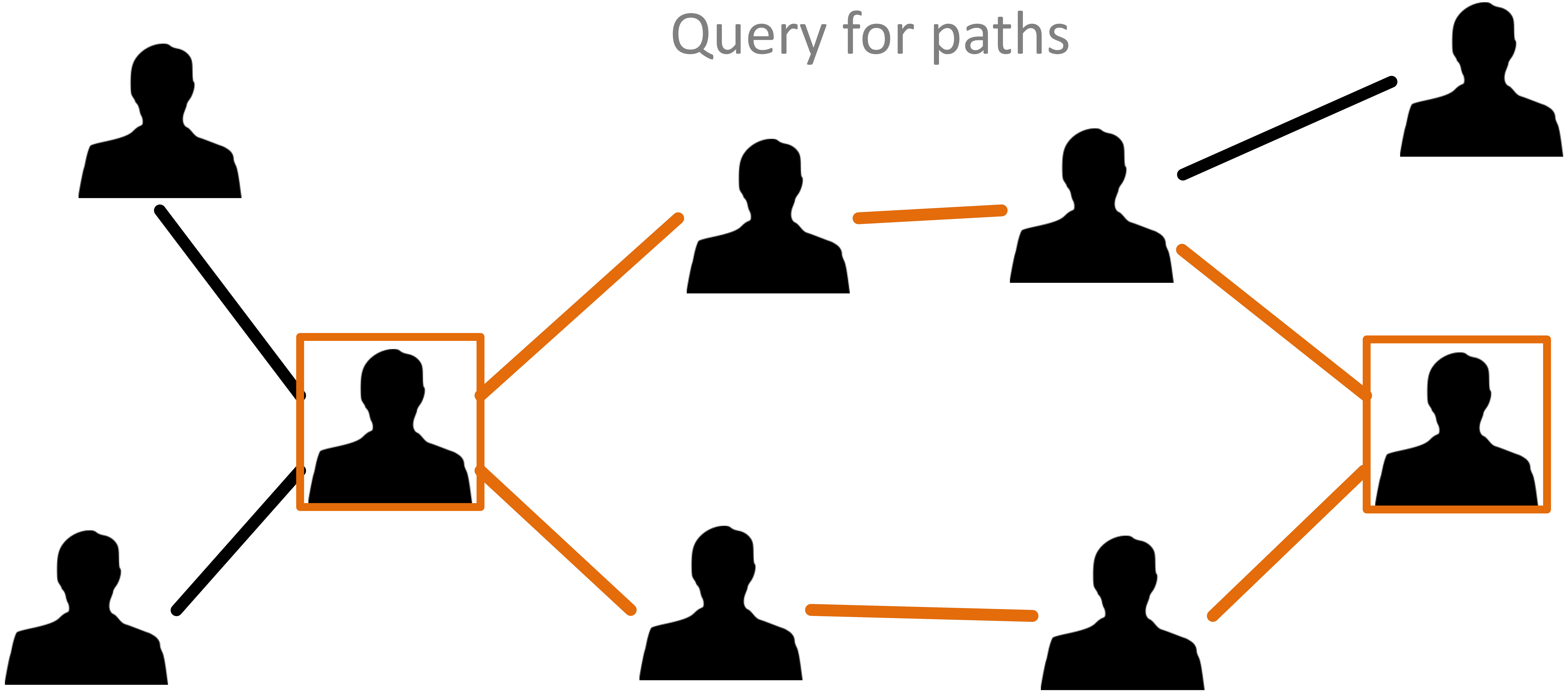
Network Data



The Approach

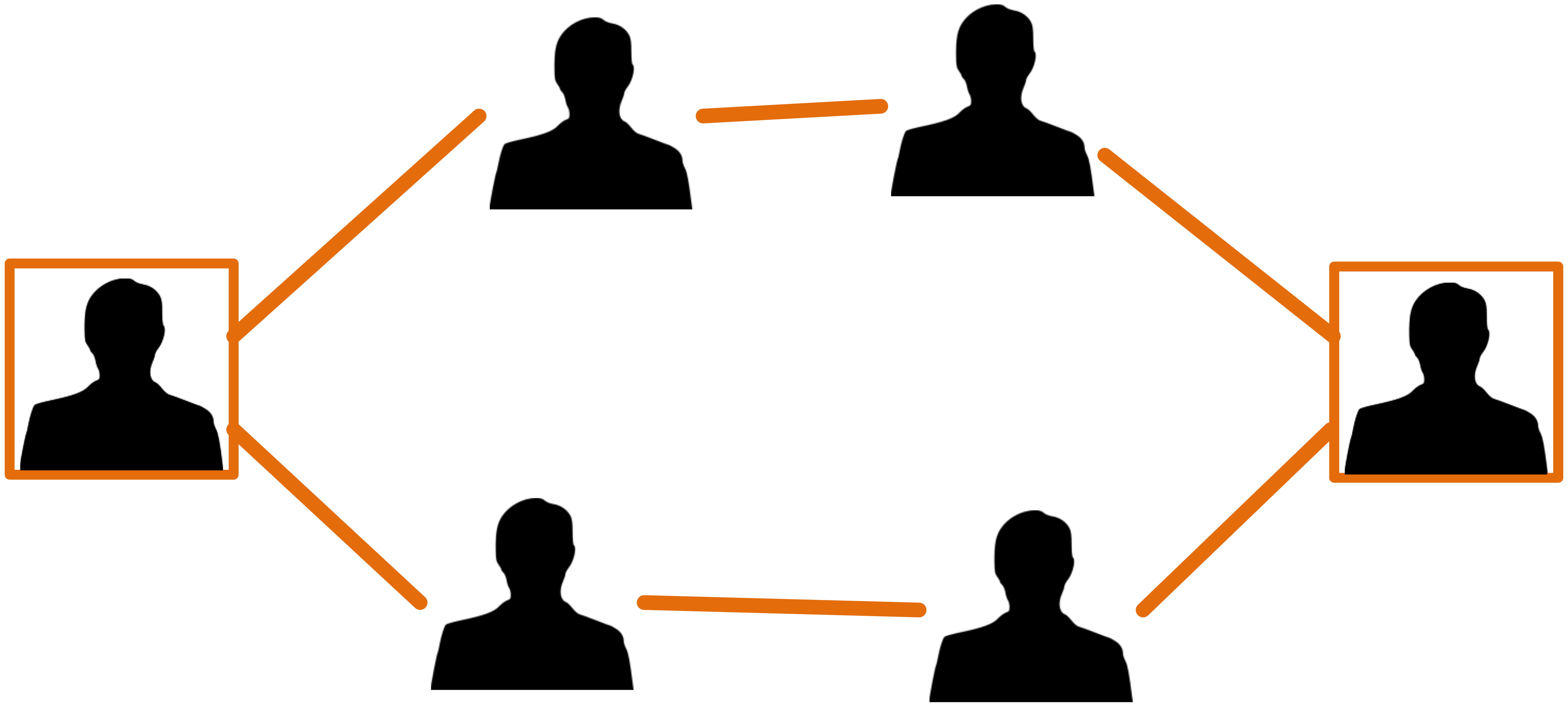
Pathfinder Approach

Query for paths



Pathfinder Approach

Show query results...
...as node-link diagram...



Pathfinder Approach

Update ranking and create list

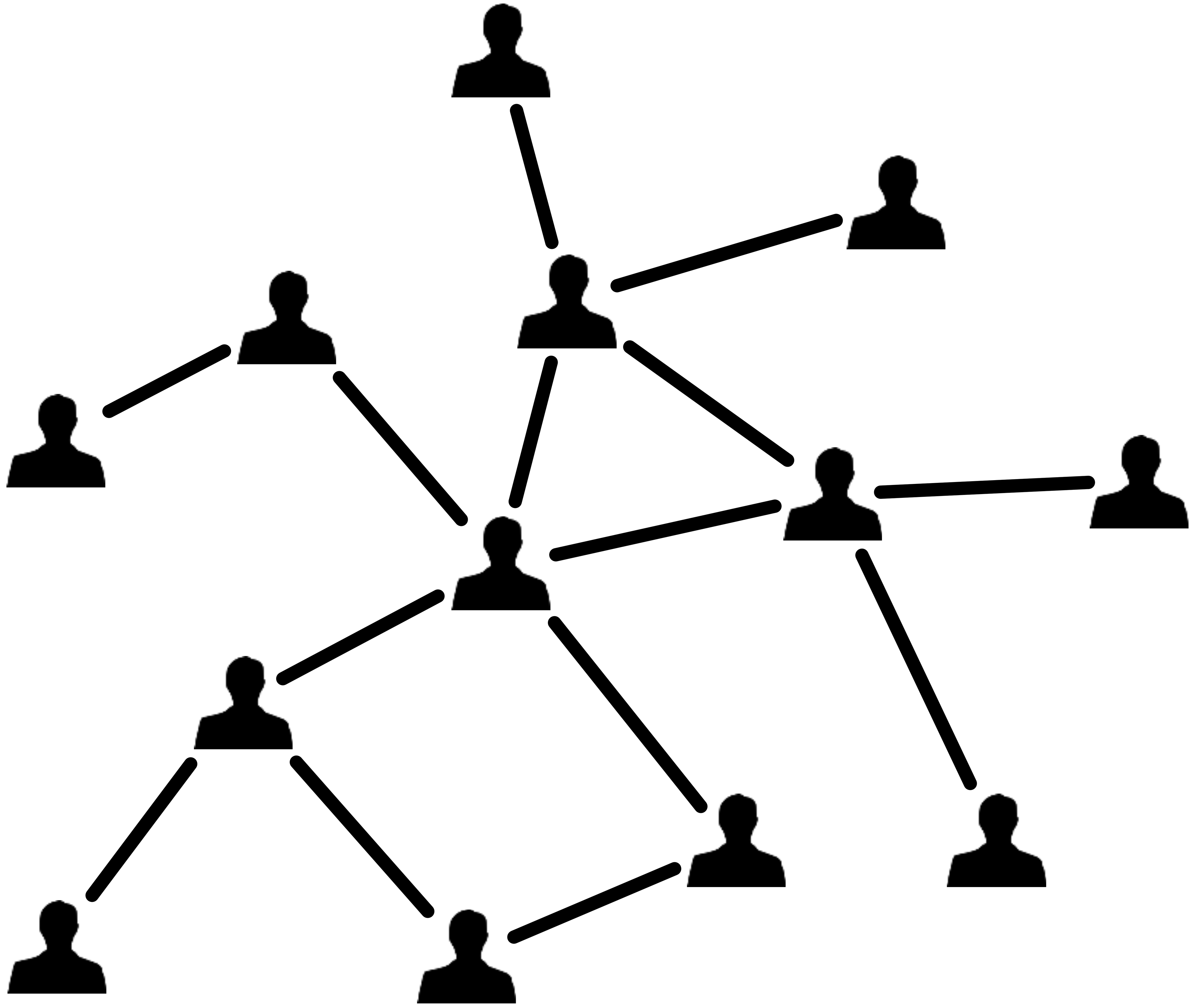
Update ranking to identify important paths

Path Score

Pathfinder Approach

Update ranking to identify important paths

Coauthor Network



Generated from TVCG and CHI publication data in DBLP

Nodes: Author

Edges: Co-author-ship

34k Nodes, 45k Edges

Attributes

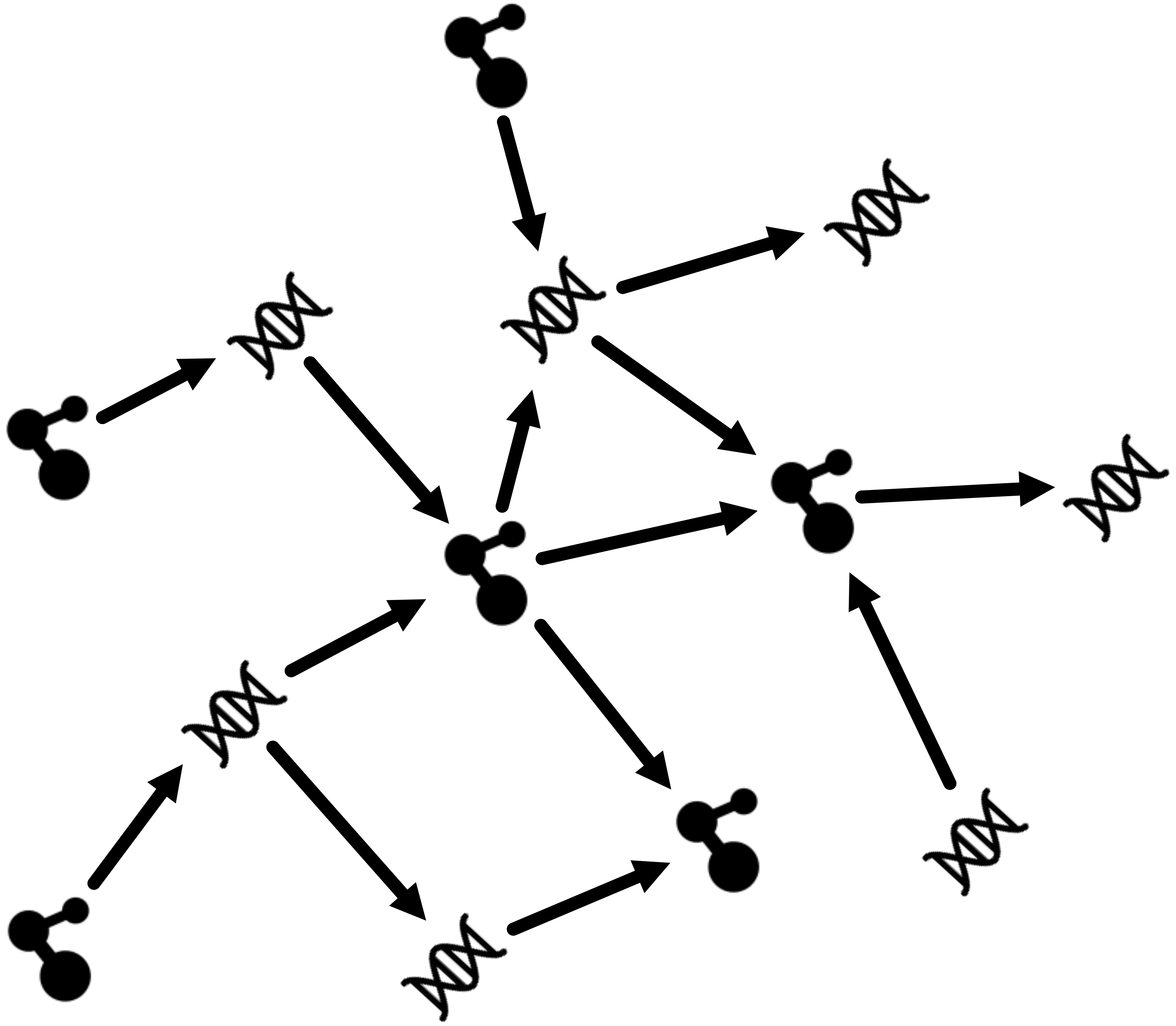
Publications

Citations

...

Sets

CHI/TVCG Publications



Biological Network

Generated from KEGG Pathway Da

11k Nodes, 71k Edges

Gene

Attributes

Gene Expression Data

Copy Number Data

Sets

Biological Pathways

Compound

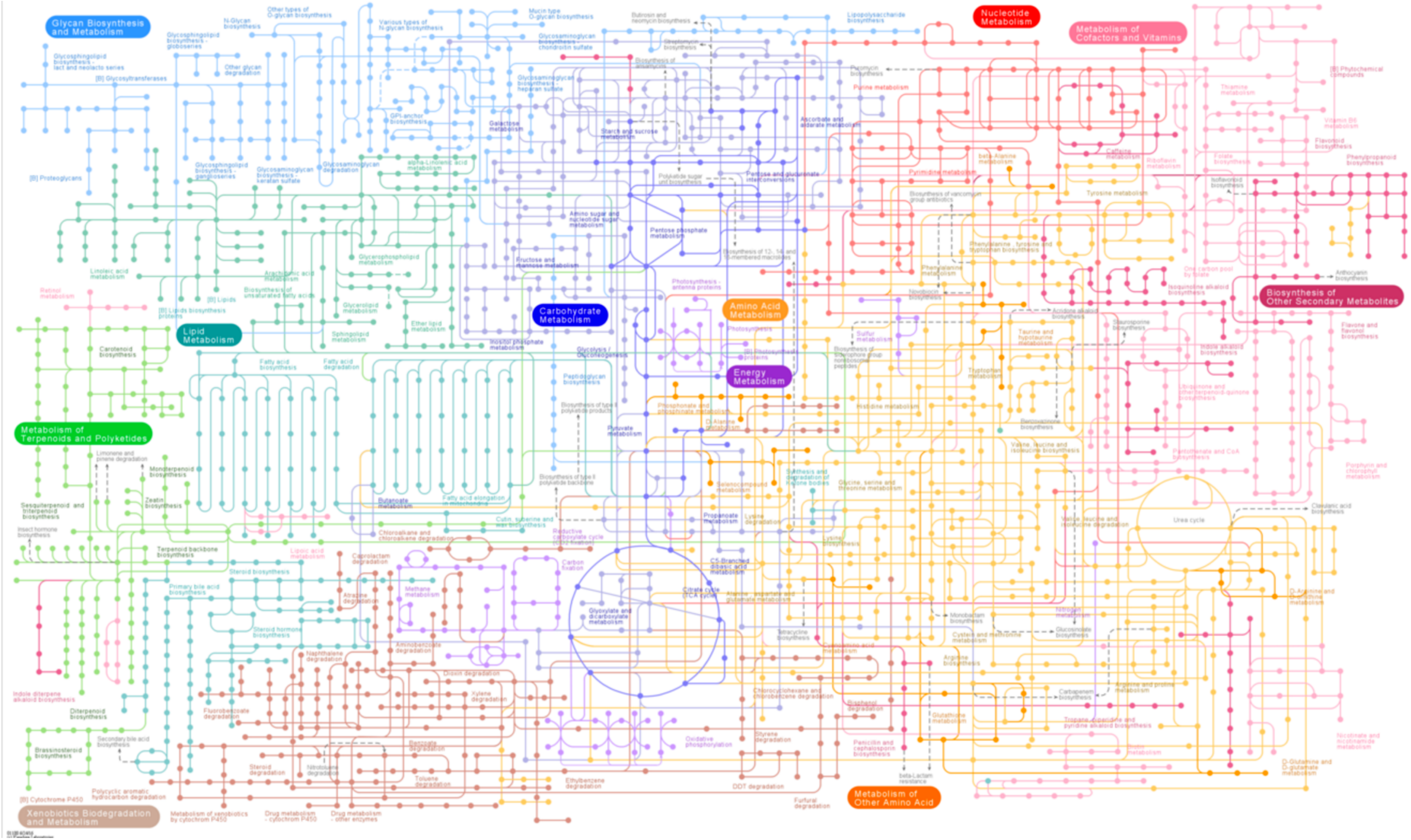
Attributes

-

Sets

Biological Pathways

Example: KEGG Metabolism Overview



Analysis Questions

Two genes are co-expressed. What is their underlying connection?

What are alternative routes connecting A to B?

How is gene A connected to Pathway B?

Is the connection from A to B the same in many pathways?

Is the route connecting A and B active?

Pathfinder Views

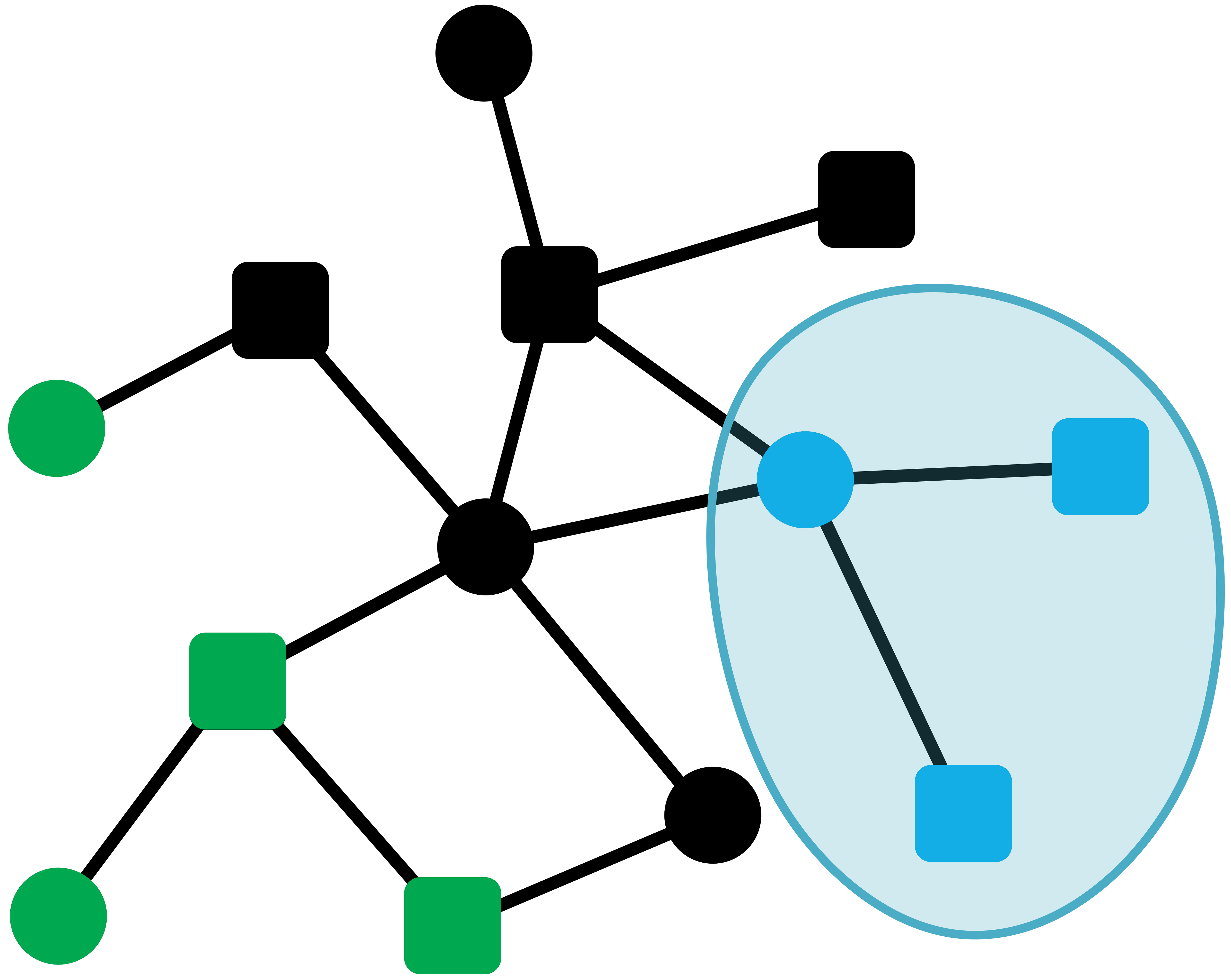

Pathfinder
Start Hanspeter Pfis End Ben Shneiderm
Advanced Query
Length 0 1 2 3 4
Paths 0 0 0 3 105
Path List
Path List View
Path Topology
Path Statistics View
Path Statistics
Paths 108/108
Edges 196/196
Nodes 95/95
Author (95/95)
Ben Shneiderman
Hanspeter Pfister
Jean-Daniel Feke
Catherine Plaisar
Krzysztof Z. Gajo
Desney S. Tan
Krzysztof Gajos
Paul André
Adam Perer
Jacob O. Wobbro
Frank van Ham
Wendy E. Mackay
Jeffrey Nichols
Juho Kim
Brad A. Myers
Sarita Yardi
Ed Huai-hsin Chi
Nam Wook Kim
Torsten Möller
m. c. schraefel
Erin Treacy Solov
Alan Borning
Sets 137/843
CHI (107/667)
TVCG (30/176)
CHI
TVCG
chi_publications
cited
degree
tvcg_publication
Hanspeter Pfiste
Krzysztof Z. Gajc
Desney S. Tan
Ben Shneiderman
Jean-Daniel Feke
Catherine Plaisan
Jennifer Golbeck
Wendy E. Macka
Ed Huai-hsin Ch
Jeffrey Heer
Stuart K. Card
Frank van Ham
Adam Perer
m. c. schraefel
Tim Berners-Lee
Robert Jacobs
Isaac S. Kohane
Peter Szolovits
Daniel J. Weitzne
Krist Wongsupha
Seth Widoff
« 1 2 3 4 5 6 7 8 »

Path Query and Search

Query and Search

Specify **start** and **end**

- **Start/end** can be node lists
- **Start/end** can be defined through set membership

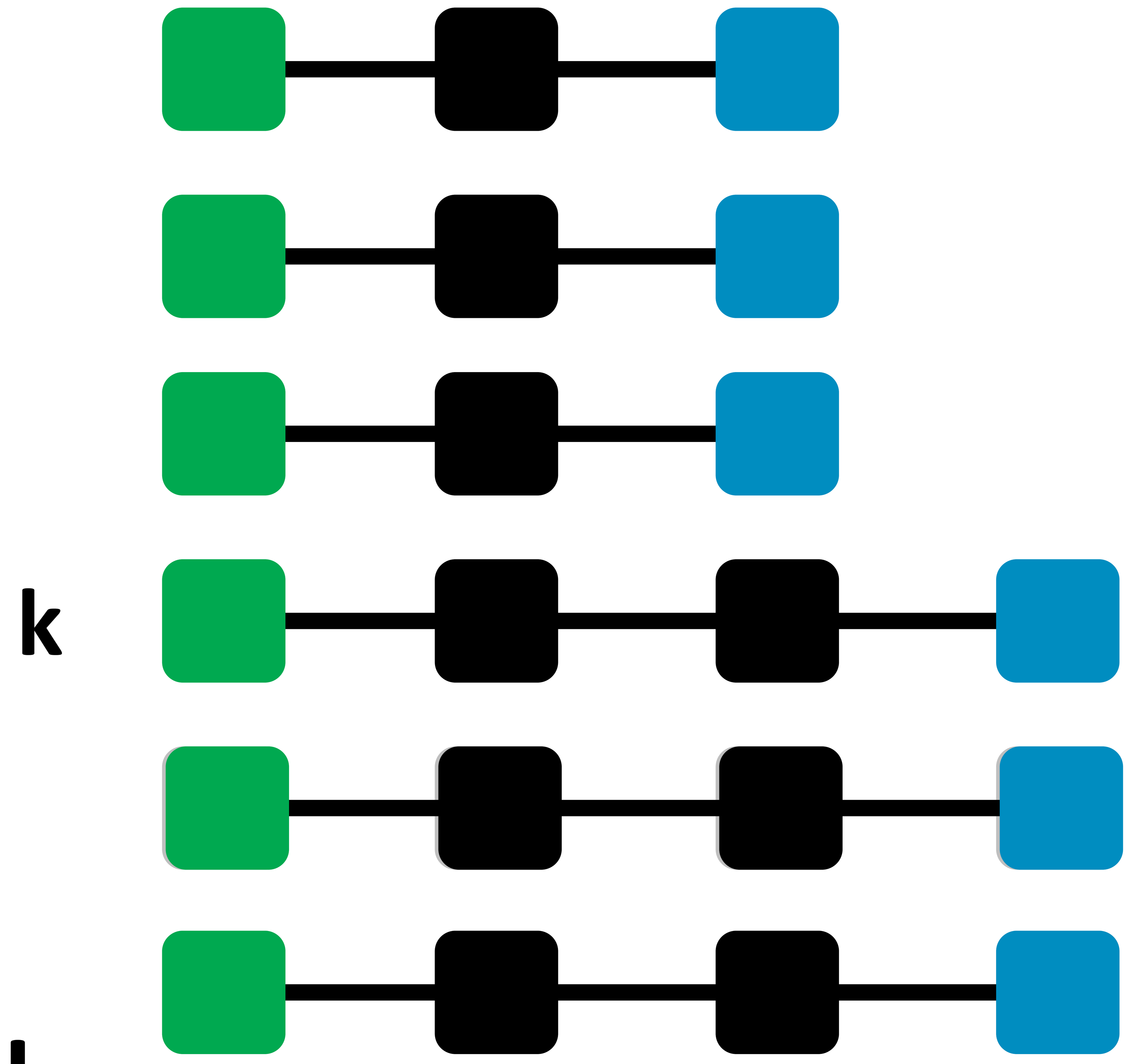


Query and Search

Specify start and end

K-shortest path search

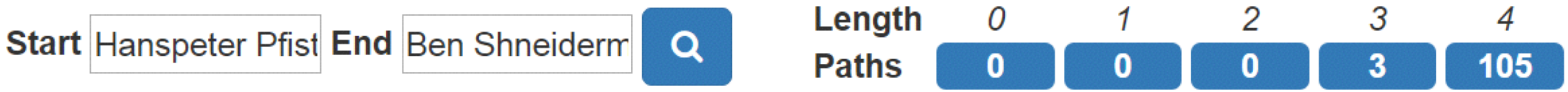
Continue until all path of length of k-th path are found




Pathfinder
Start
End
Advanced Query
Length
Paths
Path Statistics
Path List
Query Interface
Path Topology
Active Page
All

Query Interface

Simple



Advanced

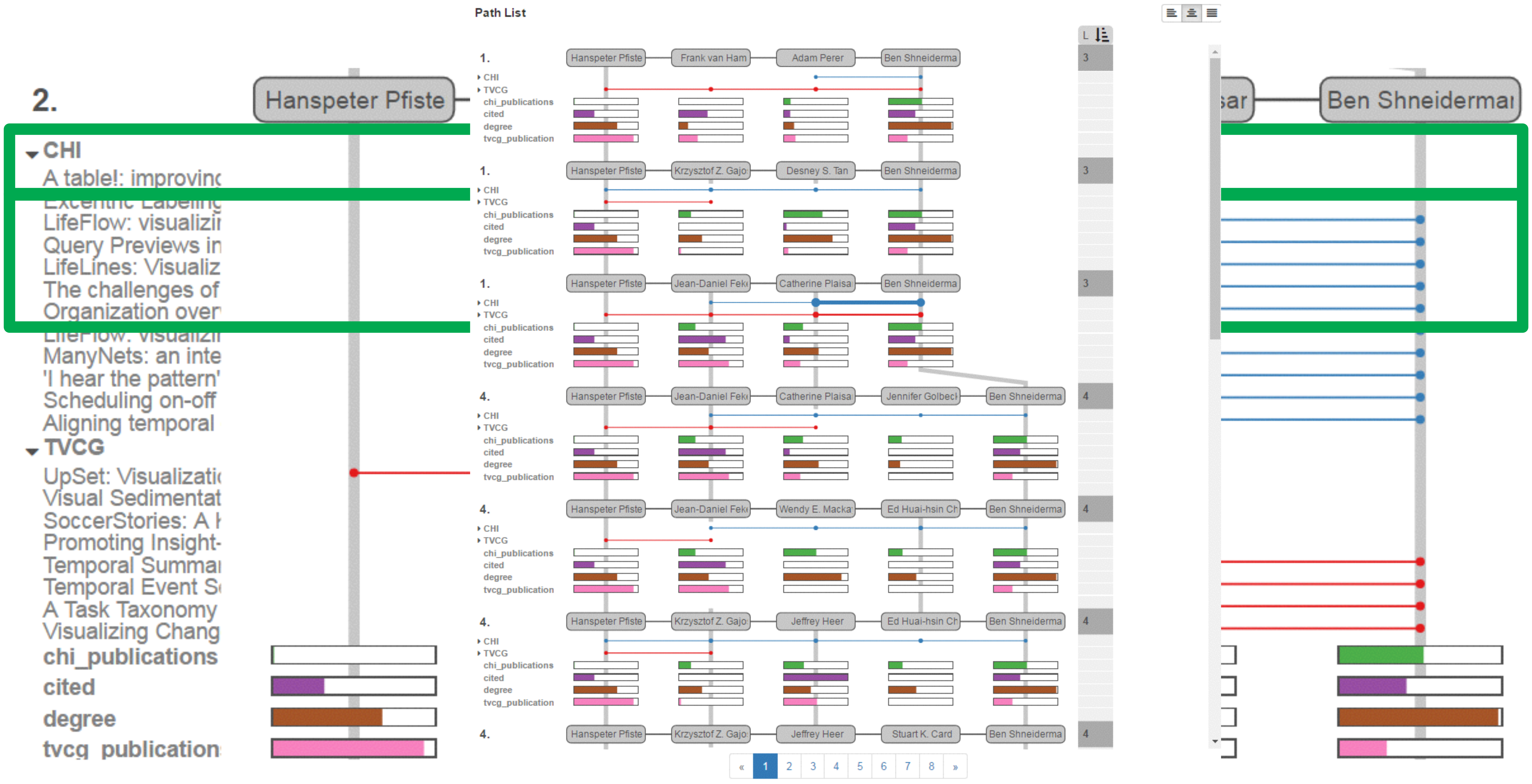
Pathfinder
Start
Hanspeter Pfis
End
Ben Shneiderm
Advanced Query
Length
0
1
2
3
4
Paths
0
0
0
3
105
Path List
1.
Hanspeter Pfiste
Frank van Ham
Adam Perer
Ben Shneidermar
CHI
TVCG
chi_publications
cited
degree
tvcg_publication
Krzysztof Z. Gajc
Desney S. Tan
Jean-Daniel Feke
Catherine Plaisan
4.
Jennifer Golbeck
Wendy E. Macka
Ed Huai-hsin Ch
Jeffrey Heer
Stuart K. Card
1
2
3
4
5
6
7
8
Path Topology
Active Page
All
m. c. schraefel
Tim Berners-Lee
Robert Jacobs
Isaac S. Kohane
Peter Szolovits
Daniel J. Weitzne
Ed Huai-hsin Chi
Krist Wongsupha
Seth Widoff
Path Statistics
Paths 108/108
Edges 196/196
Nodes 95/95
Author (95/95)
Ben Shneidermai
Hanspeter Pfister
Jean-Daniel Feke
Catherine Plaisar
Krzysztof Z. Gajo
Desney S. Tan
Krzysztof Gajos
Paul André
Adam Perer
Jacob O. Wobbro
Frank van Ham
Wendy E. Macka
Jeffrey Nichols
Juho Kim
Brad A. Myers
Sarita Yardi
Ed Huai-hsin Chi
Nam Wook Kim
Torsten Möller
m. c. schraefel
Erin Treacy Solov
Alan Borning
Sets 137/843
CHI (107/667)
Interacting with e
A Crowdsourced
Excentric Labelin
Predictability and
LifeFlow: visualizi
LifeFlow: visualizi
Aligning temporal
The challenges o
LifeLines: Visuali
ManyNets: an int
Query Previews ir
'I hear the pattern
Organization ove
Scheduling on-of
Integrating statist
Touchstone: expl
Cobi: communitys
From slacktivism t
Demonstrational i
Improving the per
Frenzy: collabora
The many faces o
TVCG (30/176)
Preface.
Promoting Insight
Evaluation of File
Temporal Summa
Temporal Event S
Visualizing Chang
A Task Taxonomy
Evaluation of Arte
Balancing Systen
'Search, Show Co
Empirical Studies
Video Snapshots
MizBee: A Multisc
MulteeSum: A To
Overview Use in
Melange: Space
A Principled Way
What Makes a Vi
FacetMap: A Sca
SoccerStories: A
ConTour: Data-Dr
Entourage: Visua

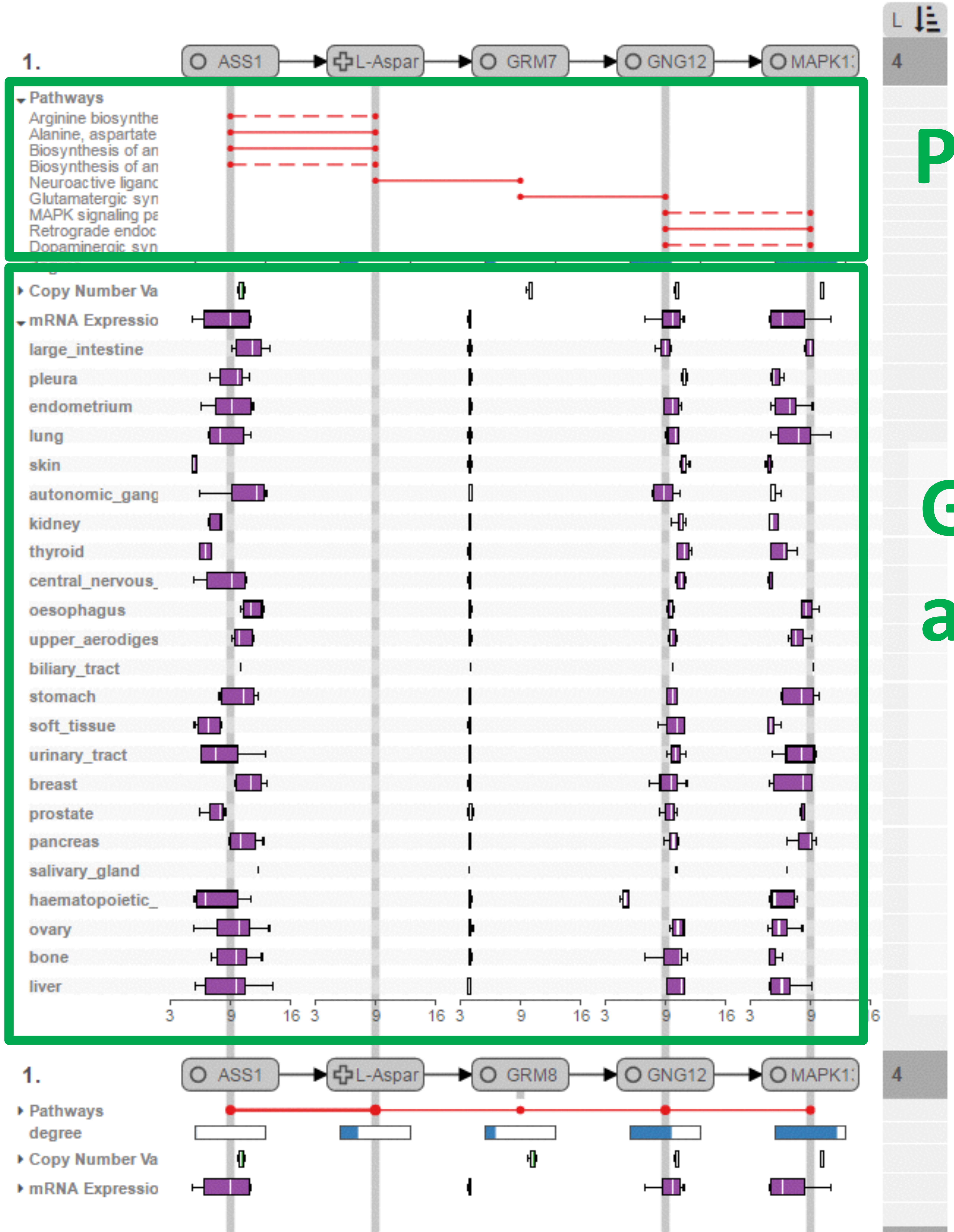
Path Topology View:
Getting an Overview

Pathfinder
Start
End
Advanced Query
Length
Paths
0
1
2
3
4
0
0
0
3
105
Path List
Path Topology
Active Page
All
Path Topology View
Path Statistics
Paths 108/108
Edges 196/196
Nodes 95/95
Author (95/95)
Sets 137/843
CHI (107/667)
TVCG (30/176)
CHI
TVCG
chi_publications
cited
degree
tvcg_publication
Hanspeter Pfister
Frank van Ham
Adam Perer
Ben Shneiderman
Desney S. Tan
Catherine Plaisant
Jennifer Golbeck
Ed Huai-hsin Chi
Jeffrey Heer
Stuart K. Card
m. c. schraefel
Tim Berners-Lee
Robert Jacobs
Isaac S. Kohane
Peter Szolovits
Krist Wongsupha
Seth Widoff
1
2
3
4
5
6
7
8

Path List View:
Investigating Paths in Detail

Path Representation





Pathways

Grouped Copy Number and Gene Expression Data

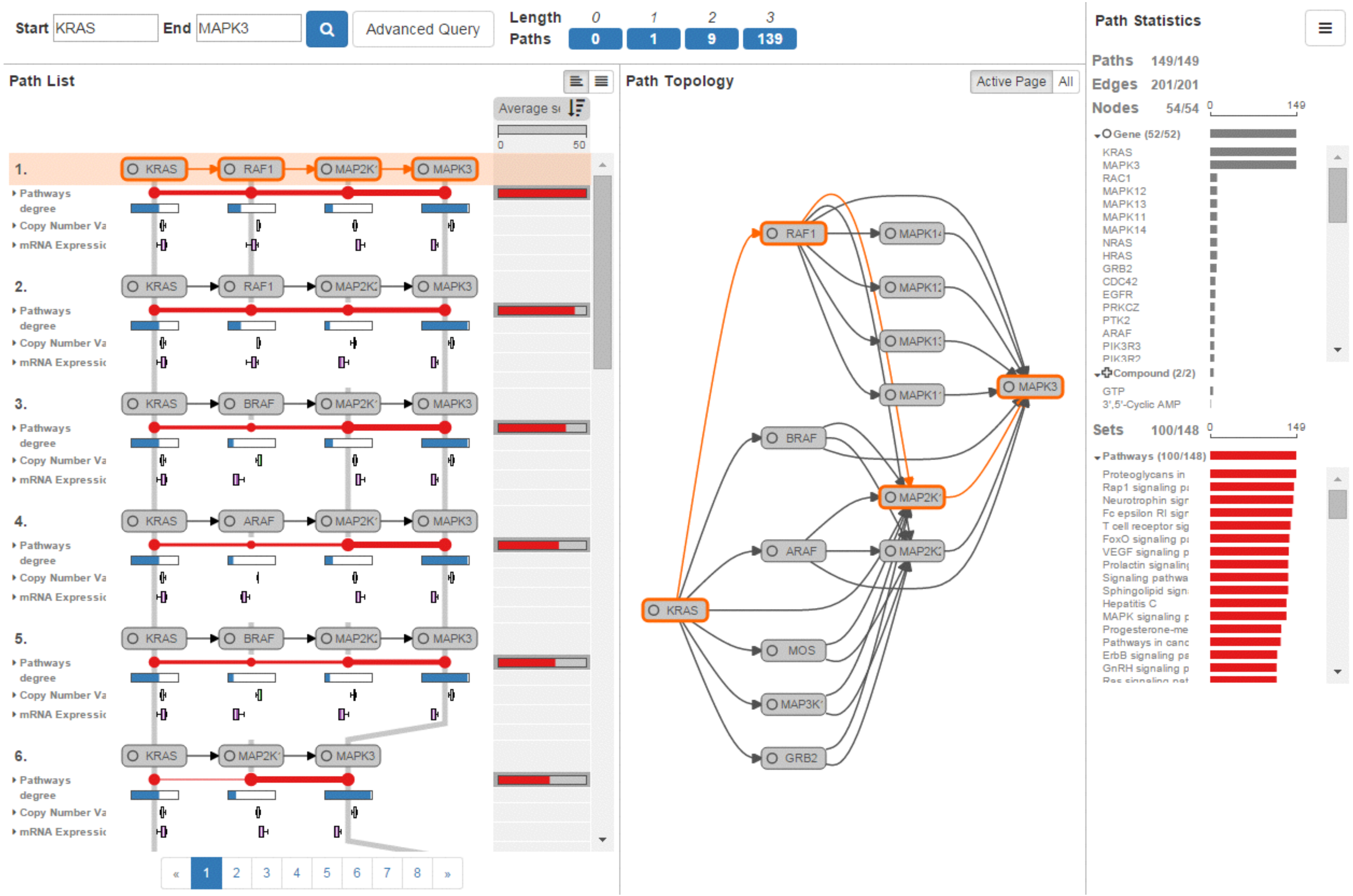
Path Ranking:
Identifying Relevant Paths


Pathfinder
Start Hanspeter Pfis End Ben Shneidern
Advanced Query
Length 0 1 2 3 4
Paths 0 0 0 3 105
Path List
Path Topology
Active Page All
Path Statistics
Paths 108/108
Edges 196/196
Nodes 95/95
Sets 137/843
Author (95/95)
CHI (107/667)
TVCG (30/176)
Rank Column
chi_publications
cited
degree
tvcg_publication
Hanspeter Pfiste
Frank van Ham
Adam Perer
Ben Shneidermar
Krzysztof Z. Gajc
Desney S. Tan
Jean-Daniel Feke
Catherine Plaisan
Jennifer Golbeck
Wendy E. Macka
Ed Huai-hsin Ch
Jeffrey Heer
Stuart K. Card
m. c. schraefel
Krist Wongsupha
Seth Widoff

Pathfinder
Start
Hanspeter Pfis
End
Ben Shneiderm
Advanced Query
Length
Paths
0
1
2
3
4
0
0
0
3
105
Path List
Average s
Rank by Average Set
Connection Strength
Rank by Journal Set
Connection Strength
Path Topology
Active Page
All
Path Statistics
Paths 108/108
Edges 196/196
Nodes 95/95
Sets 137/843
Author (95/95)
CHI (107/667)
TVCG (30/176)
CHI
TVCG
chi_publications
cited
degree
tvcg_publication
Hanspeter Pfiste
Jean-Daniel Feke
Catherine Plaisan
Ben Shneiderman
Romain Vuillemot
Petra Isenberg
Krzysztof Z. Gajc
Mary Czerwinsk
Desney S. Tan
Enrico Bertini
Tamara Munzner
Heidi Lam
Torsten Möller
Frank van Ham
Raghu Machiraju
Nathalie Henry R

Case Study:
Biological Network

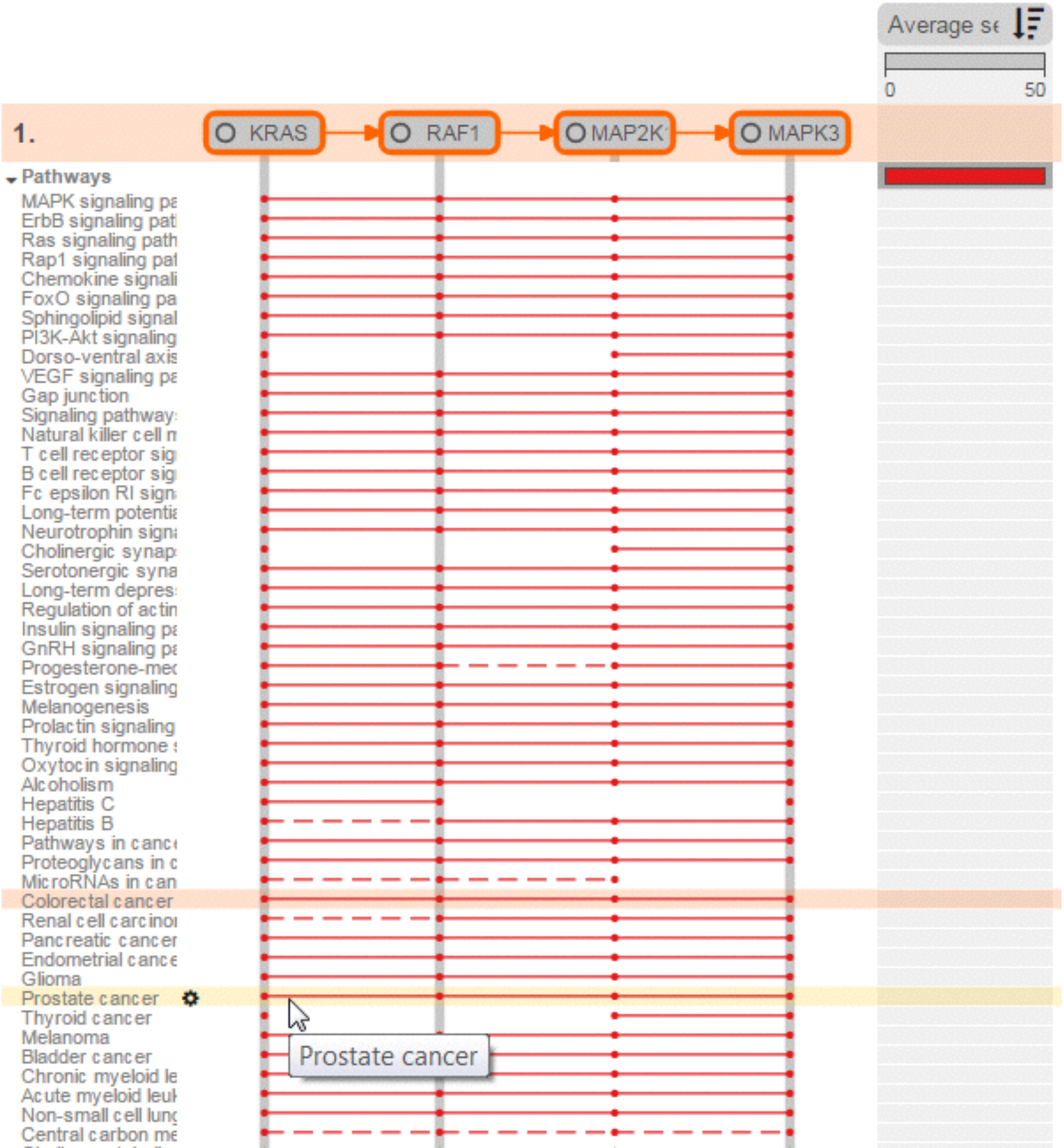
ERK-MAPK signaling cascade



Search Paths from KRAS to MAPK3

Ranking by set connection reveals ERK-MAPK signaling cascade

ERK-MAPK signaling cascade



This cascade is important in many pathways.

pathfinder.caleydo.org

[InfoVis'15]

Vials

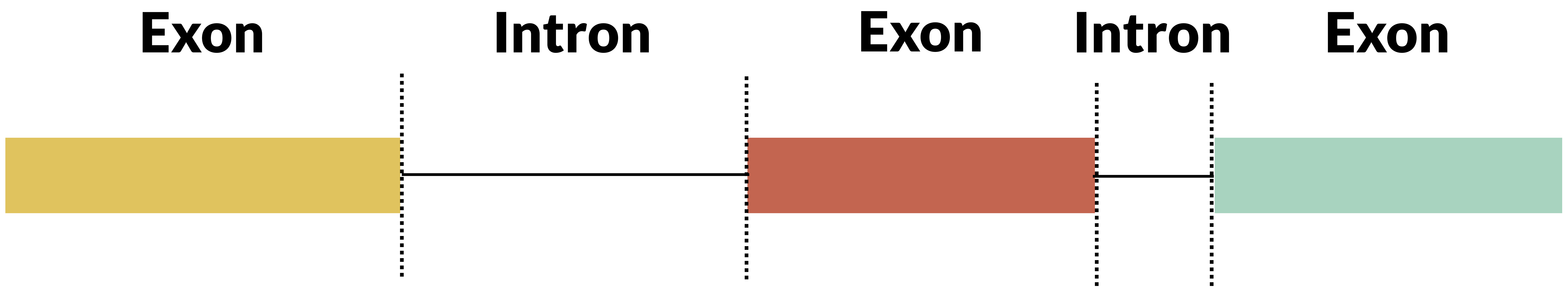
Visualizing Alternative Splicing

Scalable solution for isoform / alternative splicing analysis

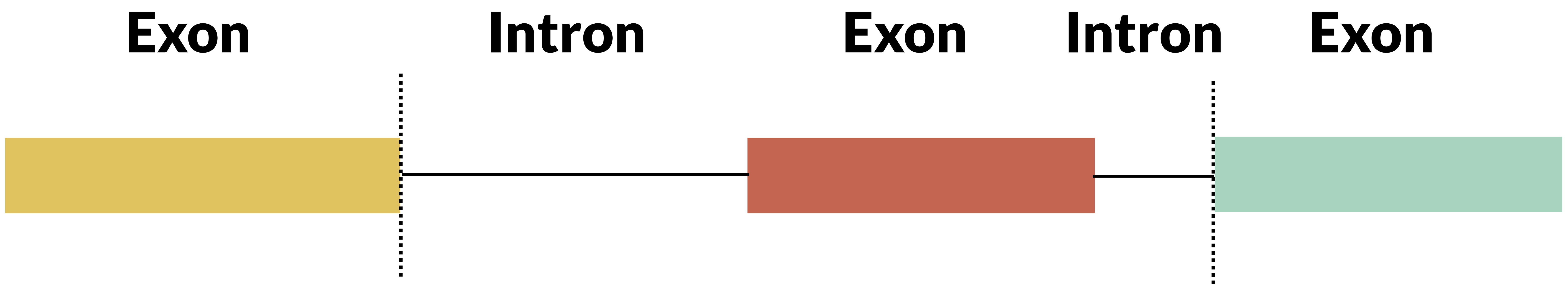
Visualize measures for
- isoform abundance,
- base-pair/exon expression,
- junction reads

for a large set of samples

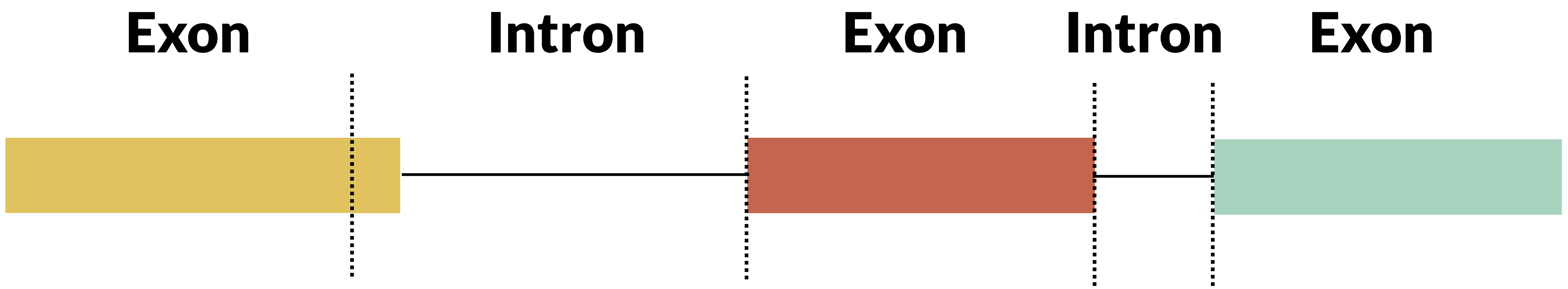
Splicing



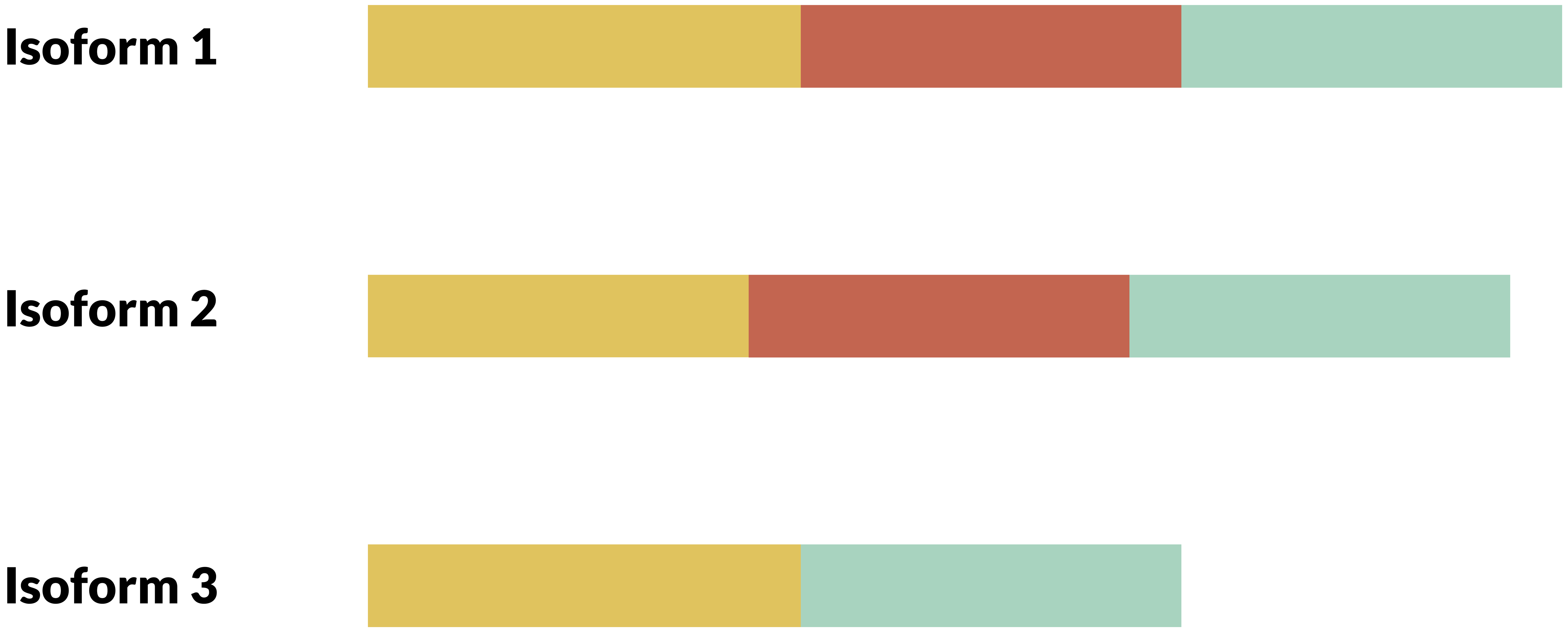
Omitted Exons



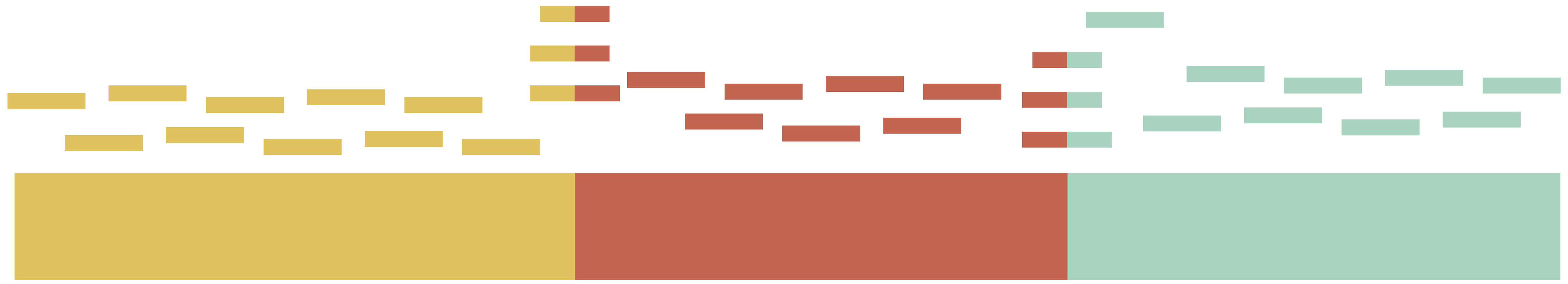
Truncated Exons



Isoforms



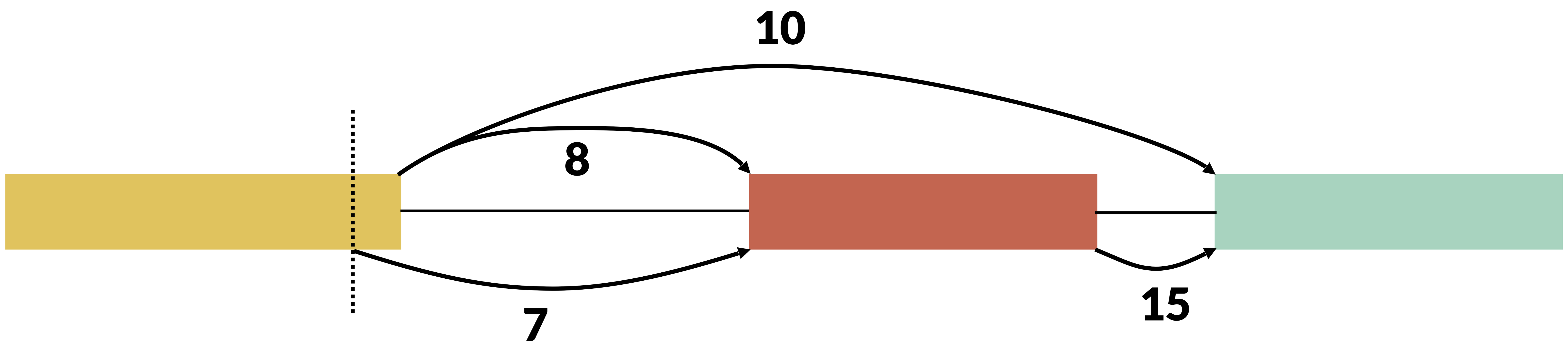
Reads



Reads from mRNA-seq data tell us about expression of exons

Junction reads tell us about which (parts of) exons are spliced out

Junction Information



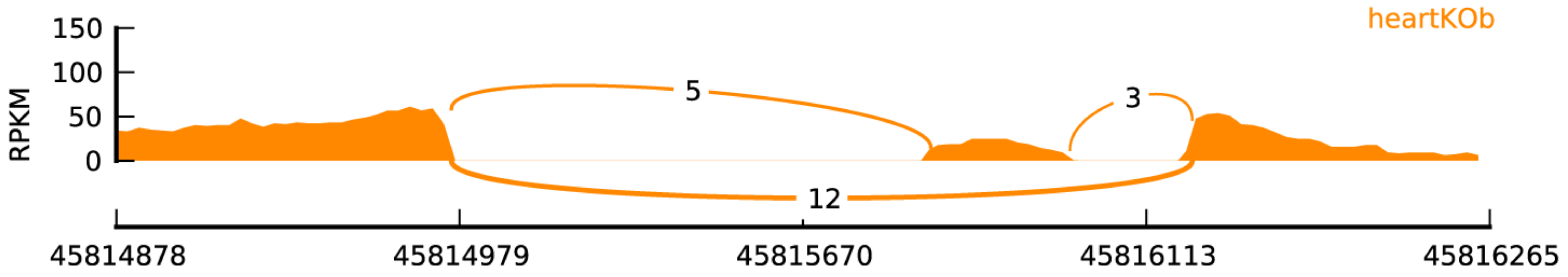
Goals

Explore differences between samples and groups of samples

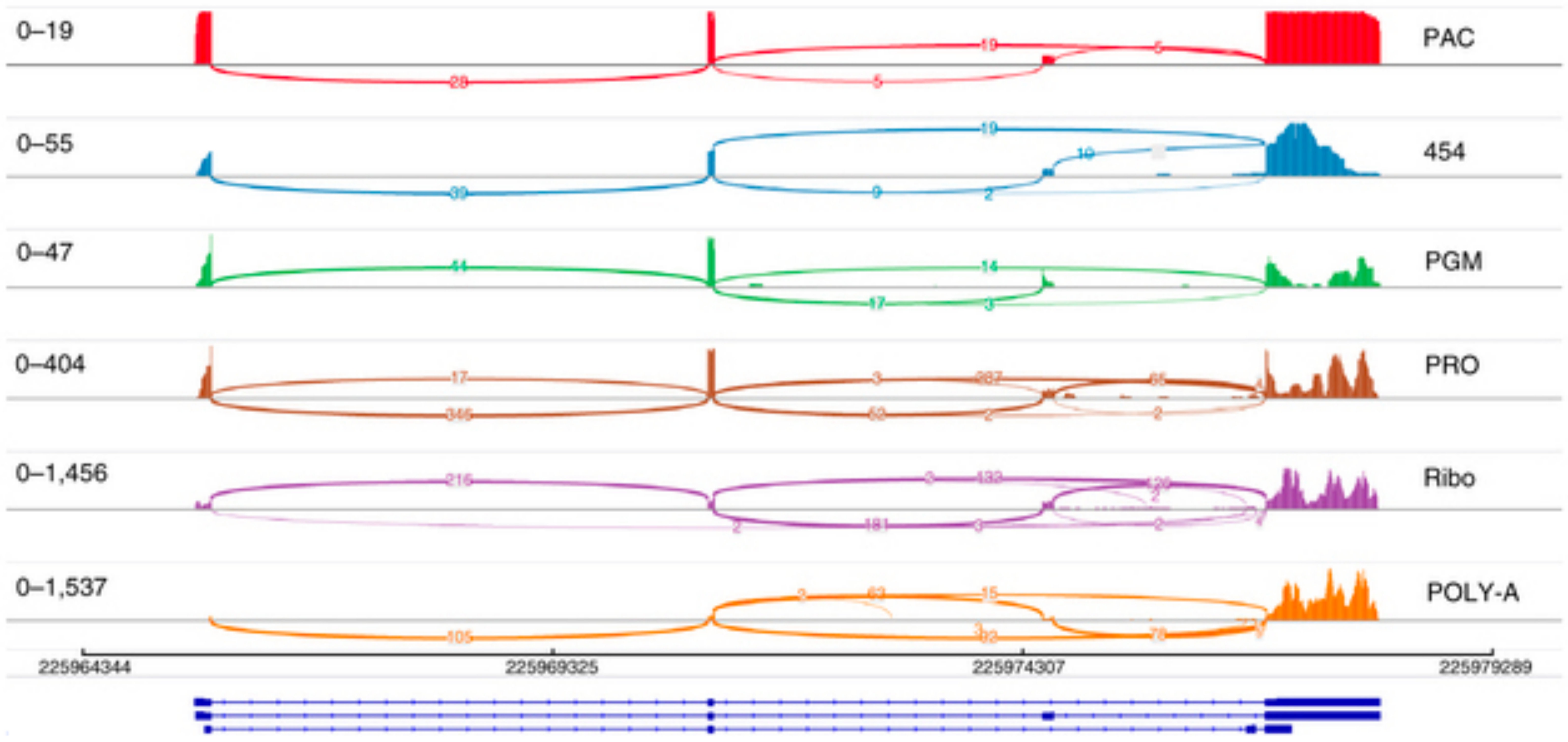
Quality control

Discover novel isoforms

The Competition: Sashimi Plots



[Katz et al, 2010]



[Sheng et al, 2014]

Data & Visualization

Isoform abundance

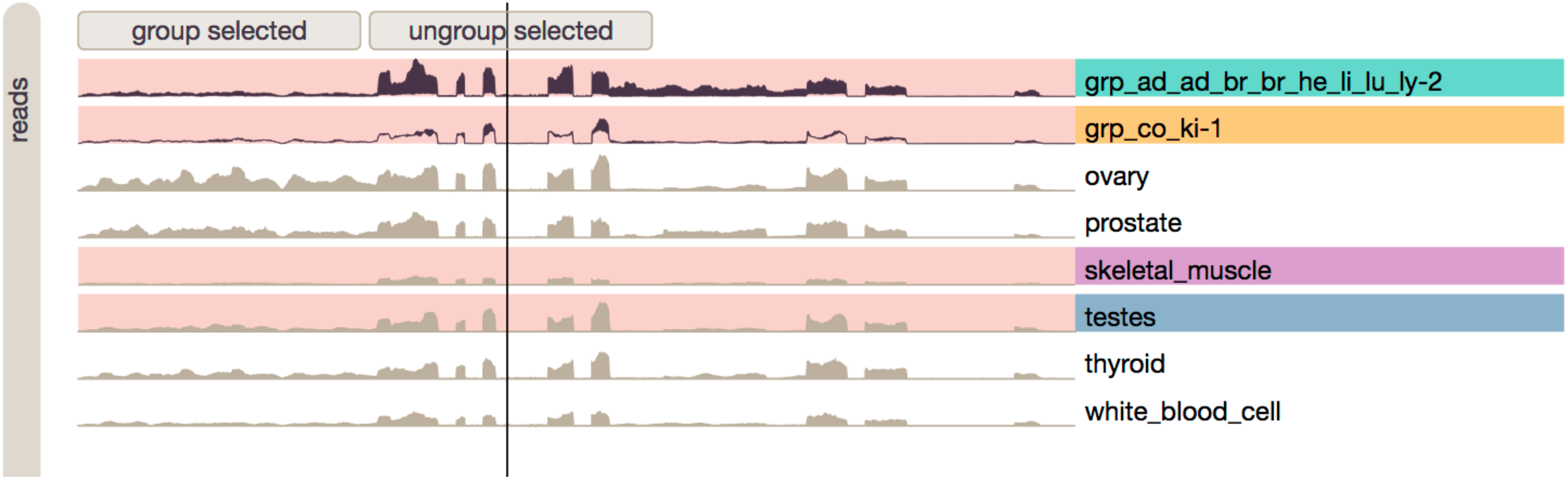
Exon expression

Junction reads

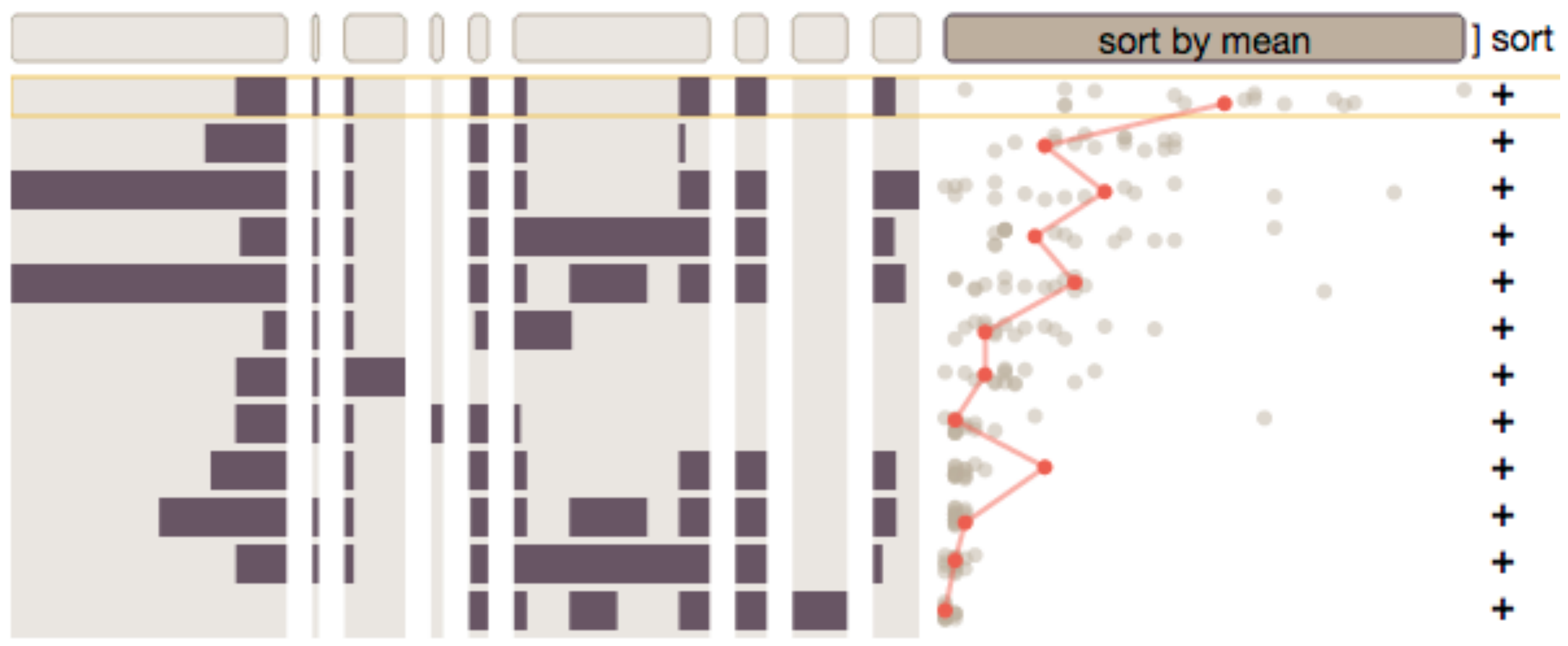
...for hundreds of samples

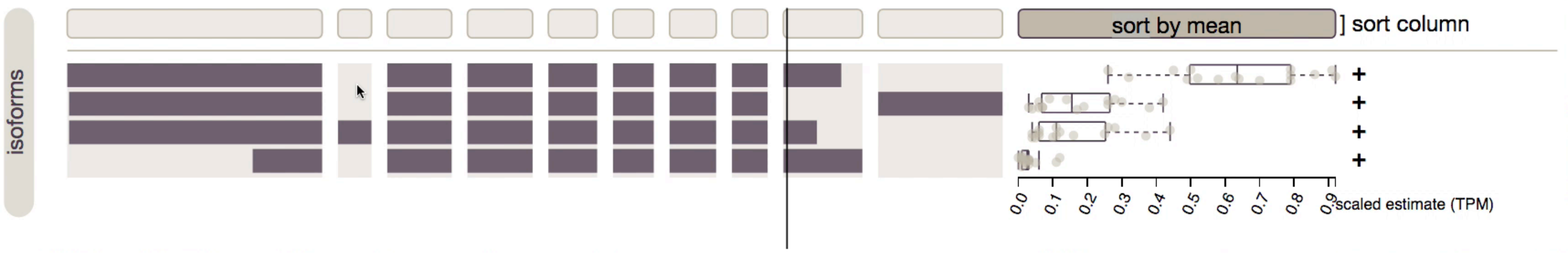
...for multiple groups

Expression View

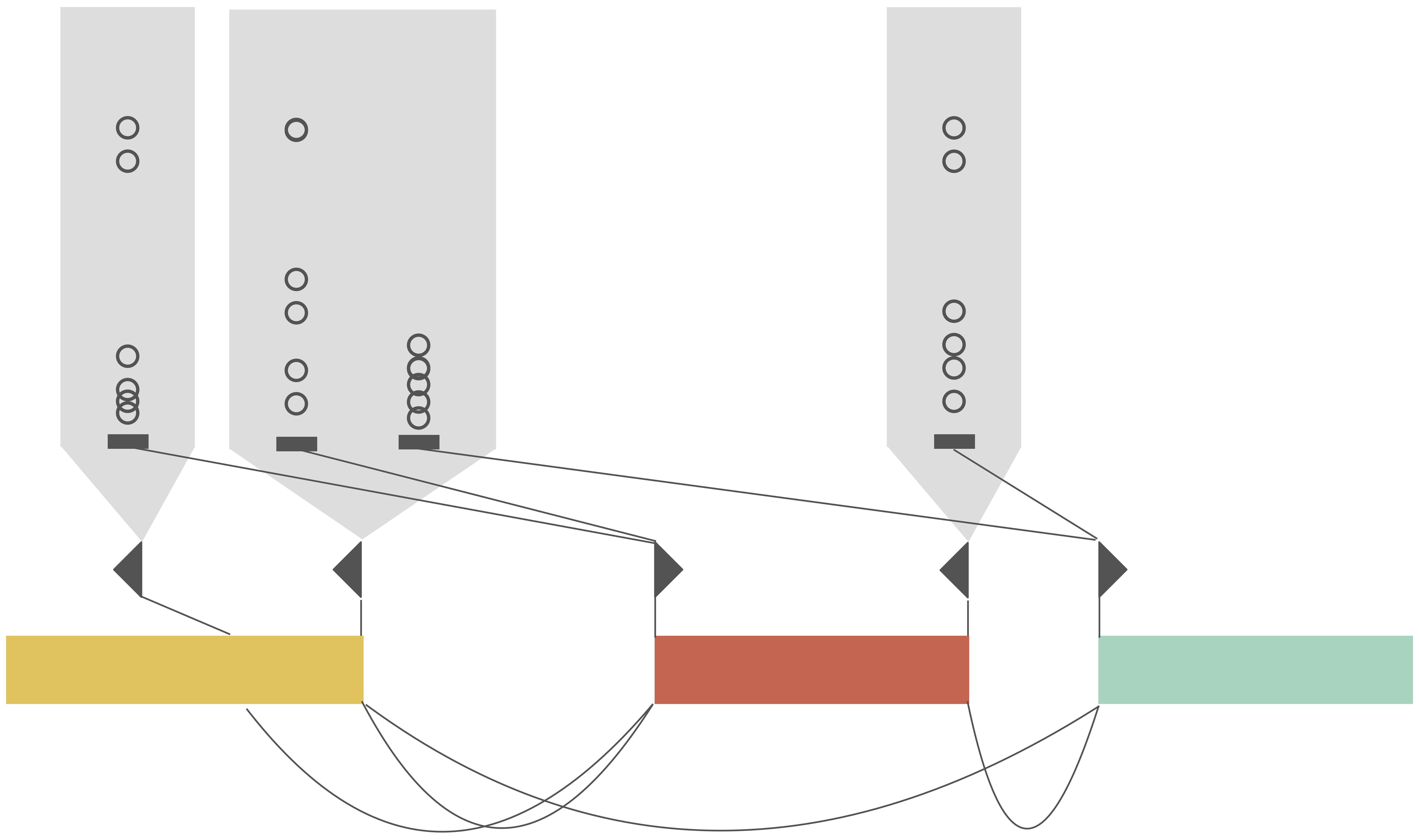


Isoform View

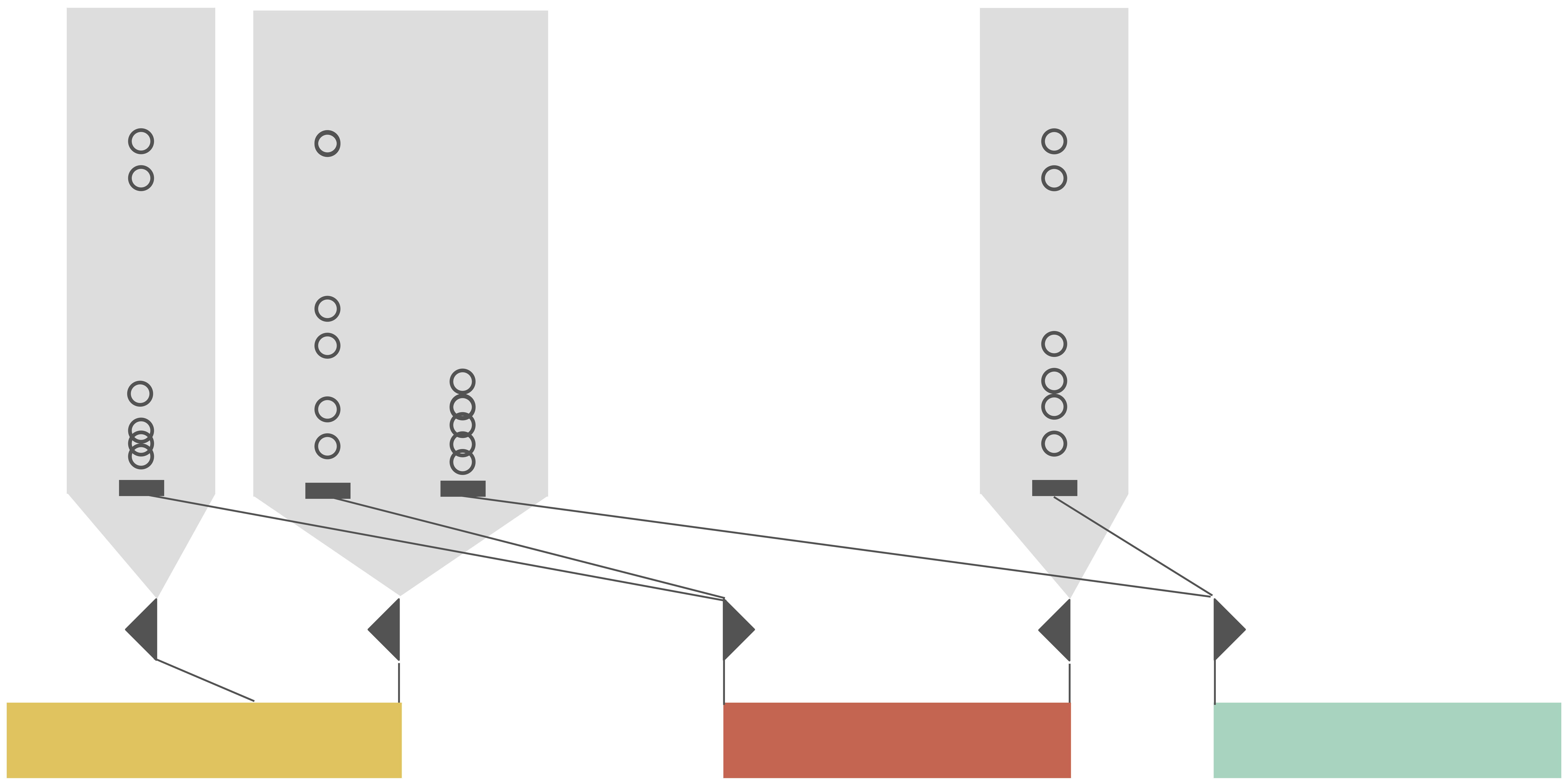




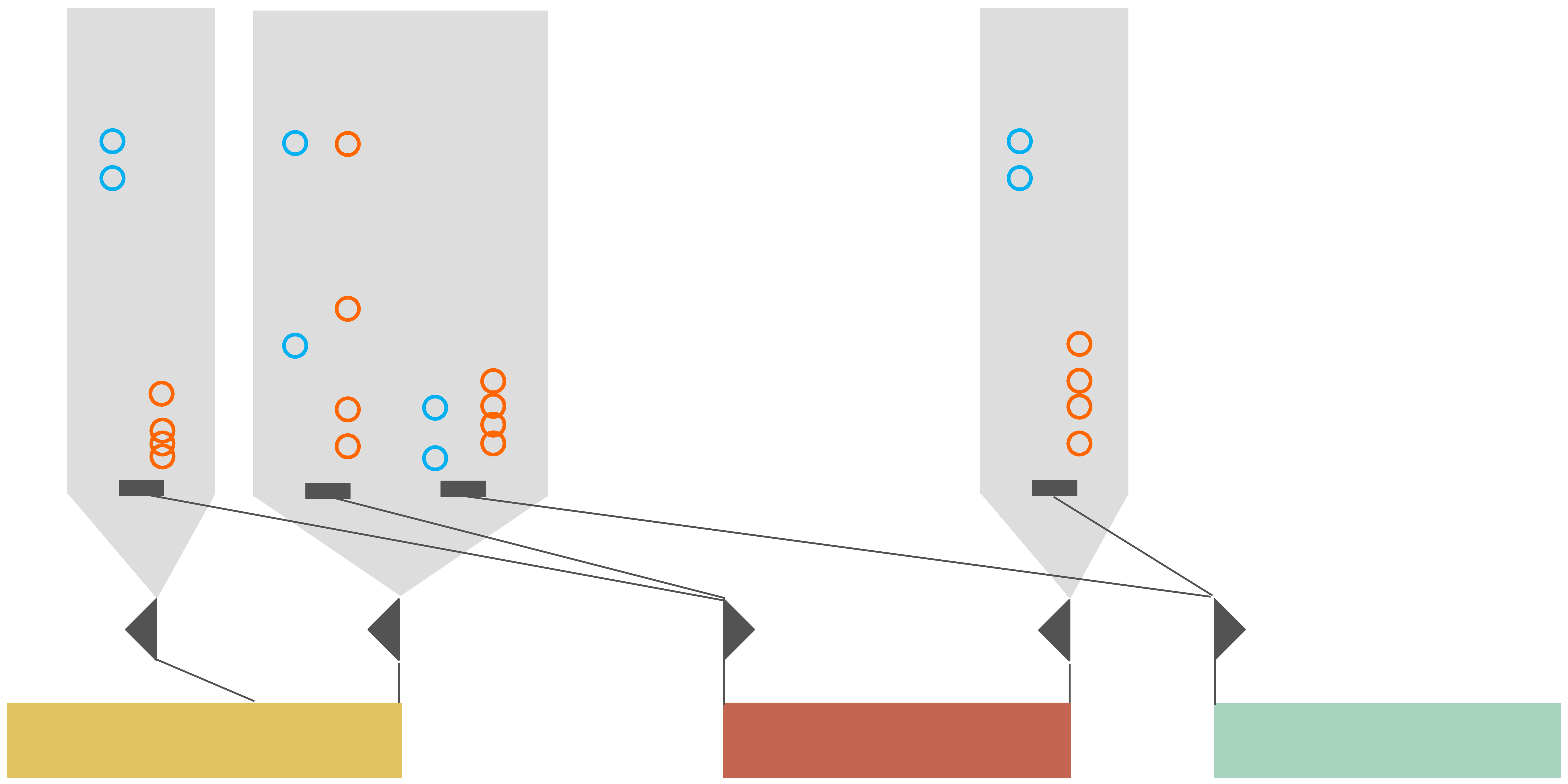
Junction View



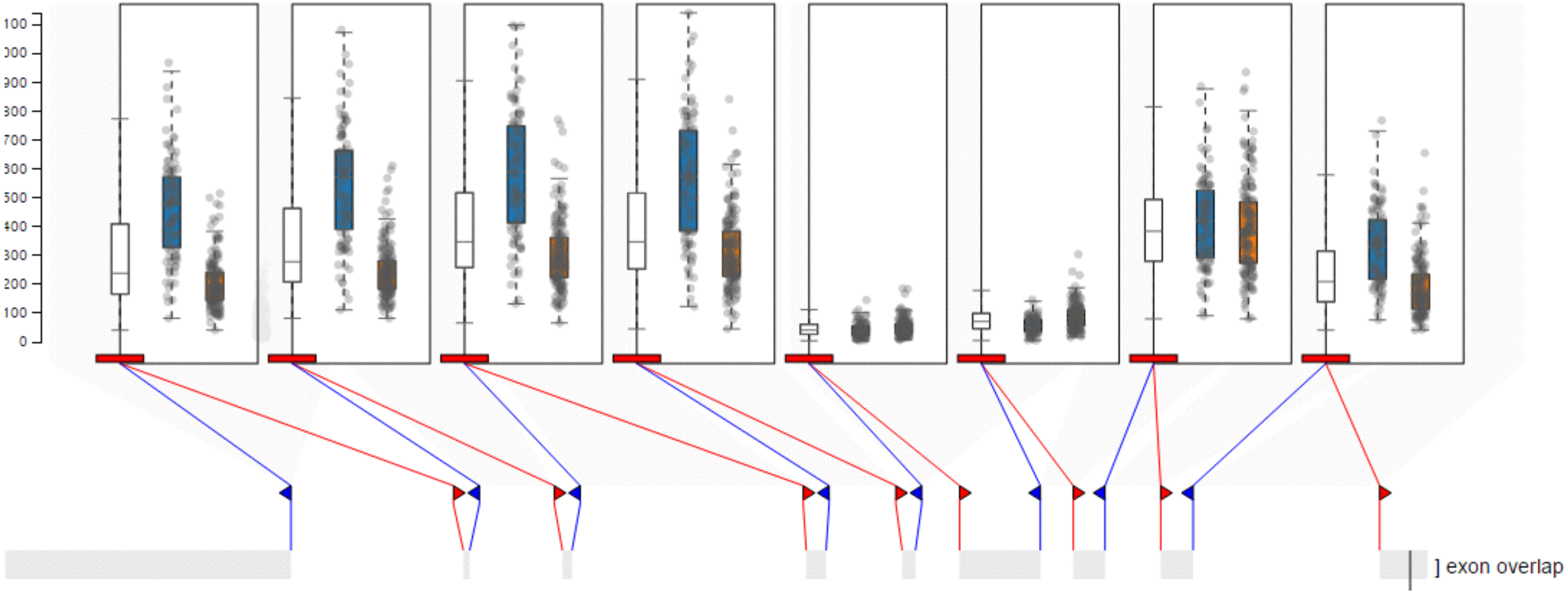
Junction View - Group Comparison



Junction View - Group Comparison



Junction View - Group Comparison

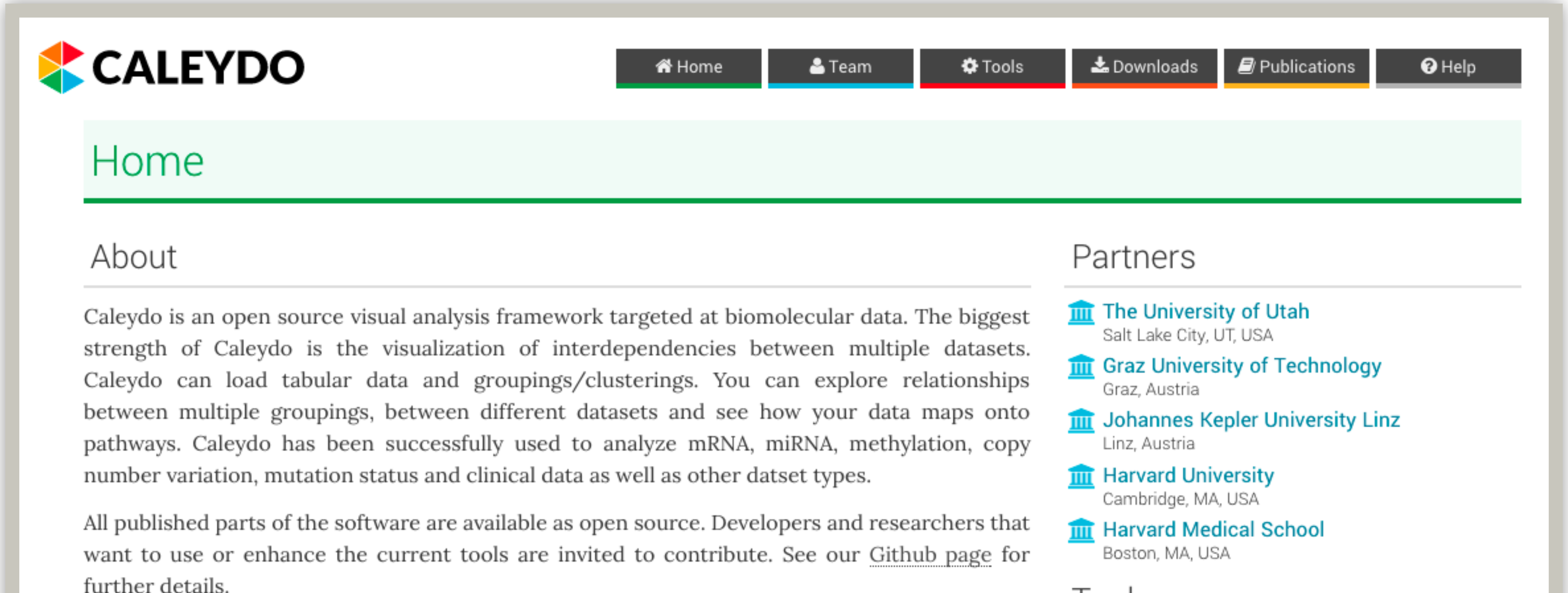


Case Study: Leukemia vs Glioblastoma

http://vials.io

CALEYDO

http://caleydo.org


CALEYDO

Home | Team | Tools | Downloads | Publications | Help

Home

About

Caleydo is an open source visual analysis framework targeted at biomolecular data. The biggest strength of Caleydo is the visualization of interdependencies between multiple datasets. Caleydo can load tabular data and groupings/clusterings. You can explore relationships between multiple groupings, between different datasets and see how your data maps onto pathways. Caleydo has been successfully used to analyze mRNA, miRNA, methylation, copy number variation, mutation status and clinical data as well as other datset types.

All published parts of the software are available as open source. Developers and researchers that want to use or enhance the current tools are invited to contribute. See our Github page for further details.

Partners

- The University of Utah
  Salt Lake City, UT, USA
- Graz University of Technology
  Graz, Austria
- Johannes Kepler University Linz
  Linz, Austria
- Harvard University
  Cambridge, MA, USA
- Harvard Medical School
  Boston, MA, USA

The Theme

Use the best visual channel for the key data

Use color sparingly, primarily for highlighting

Deal with big data through:

- definable queries
- dynamic, interactive ranking

Alexander Lex
@alexander_lex
http://alexander-lex.com

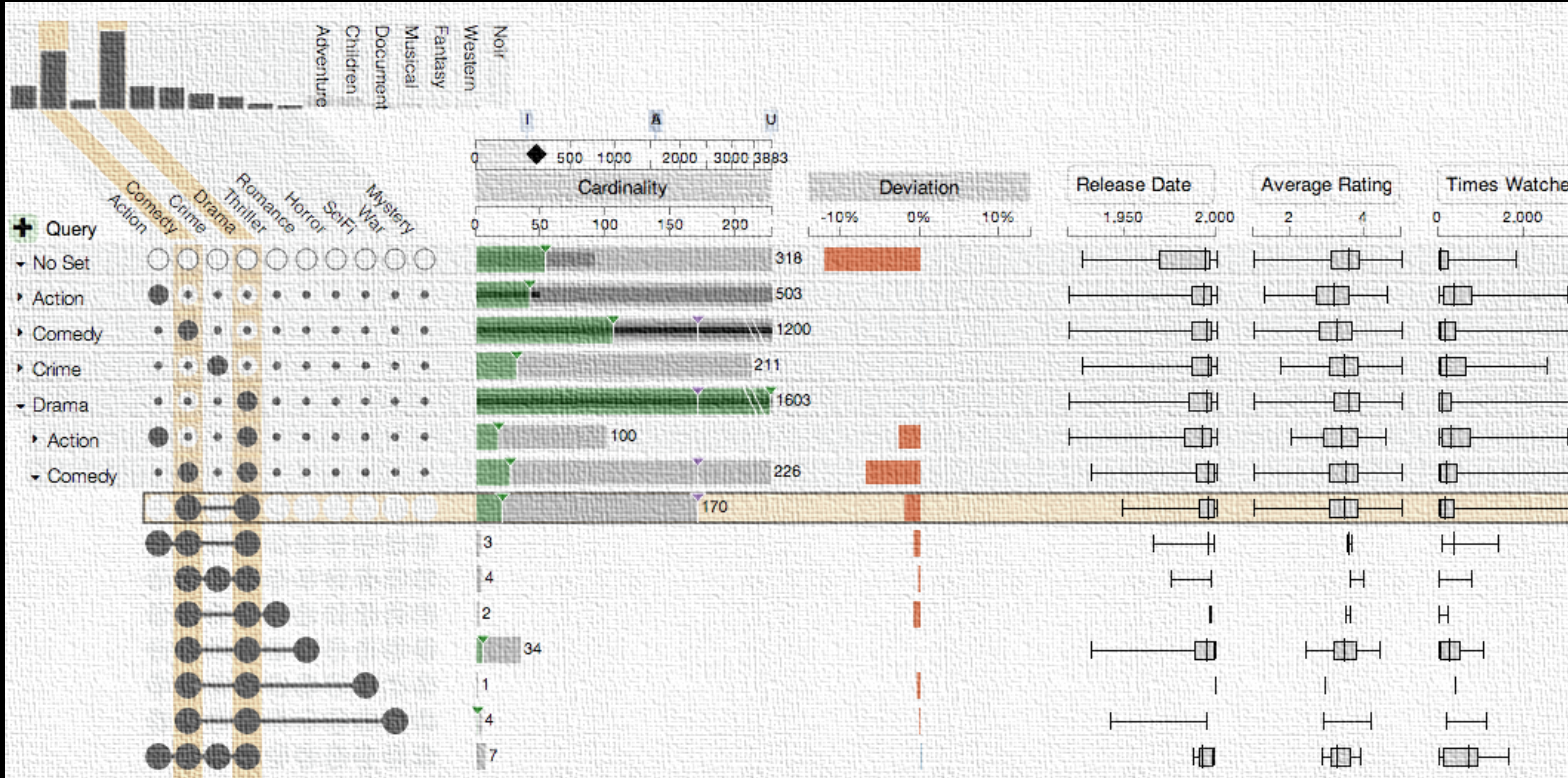
THE UNIVERSITY OF UTAH

Enabling Discovery through Interactive Visual Data Analysis

?

Credits:

Marc Streit, Nils Gehlenborg, Christian Partl, Samuel Gratzl, Hendrik Strobelt, Romain Vuillemot, Dieter Schmalstieg, Denis Kalkofen, Mark Borowsky, Anne Mai Wasserman, Hanspeter Pfister



We're hiring PostDocs and accept PhD Students!

visualization design lab

Miriah Meyer

Alexander Lex

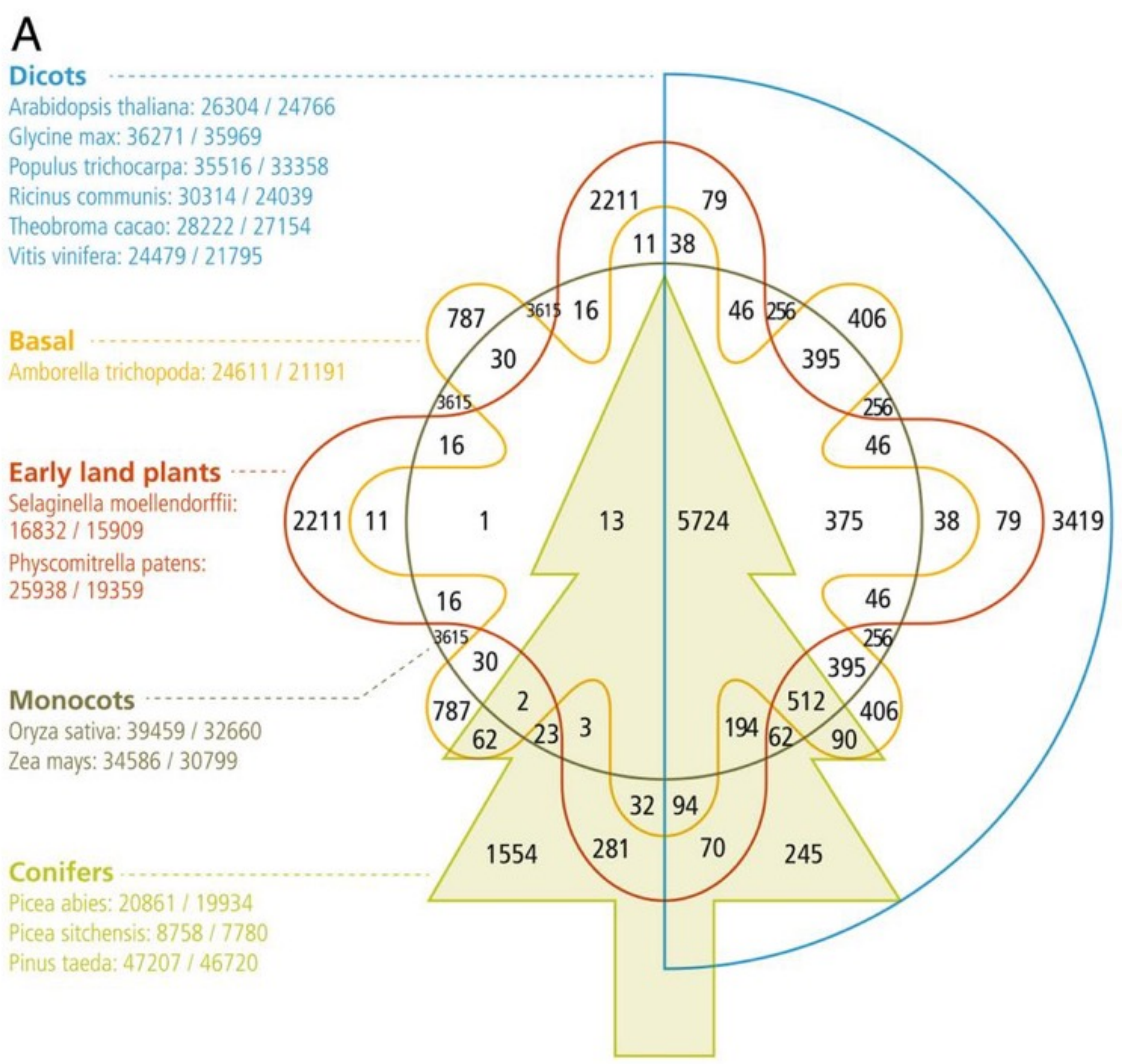
http://goo.gl/Zmm4L8

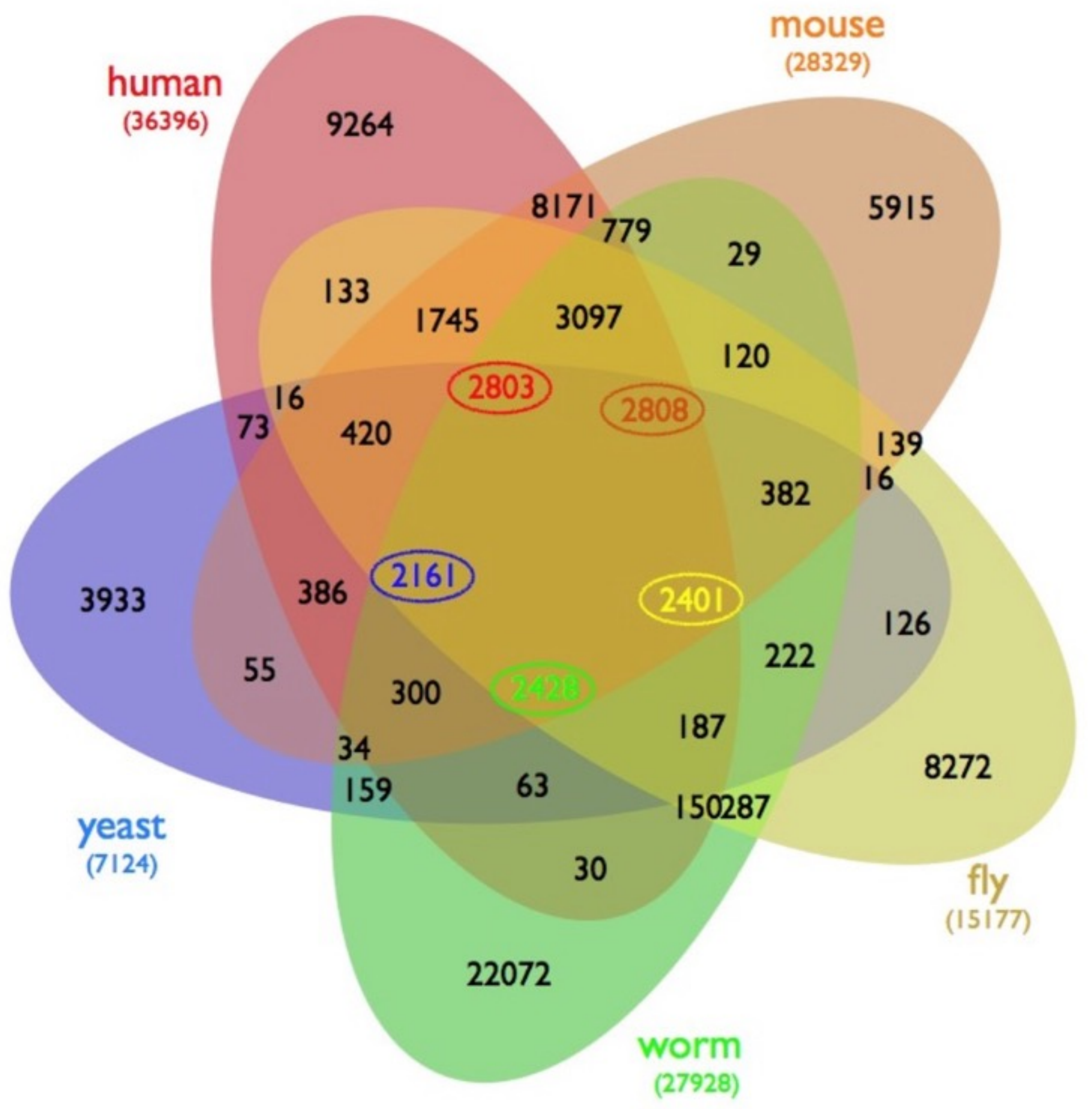
UpSet

[InfoVis'14]

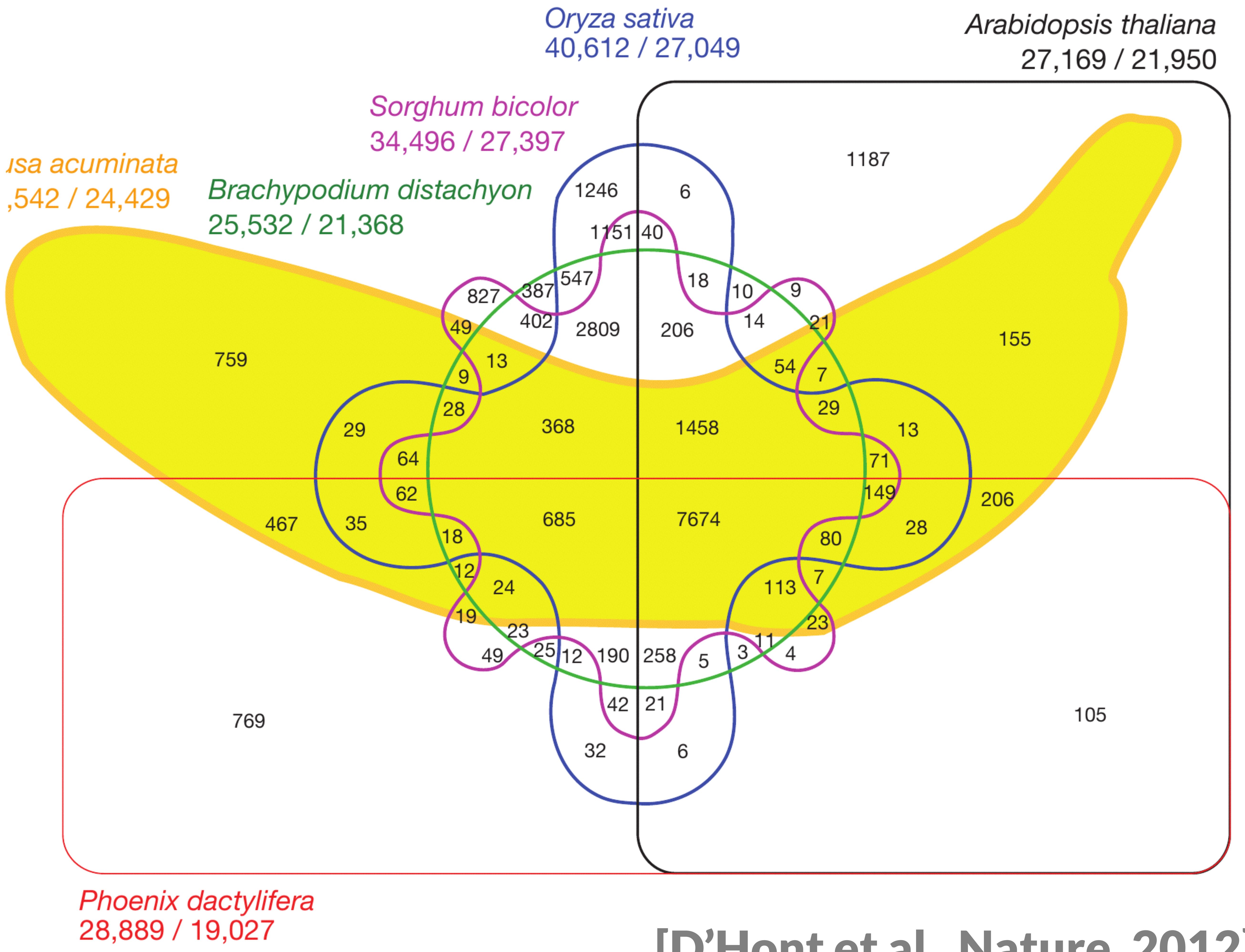
Visualizing Intersecting Sets

[Neale et al., BMC Genome Biology, 2014]

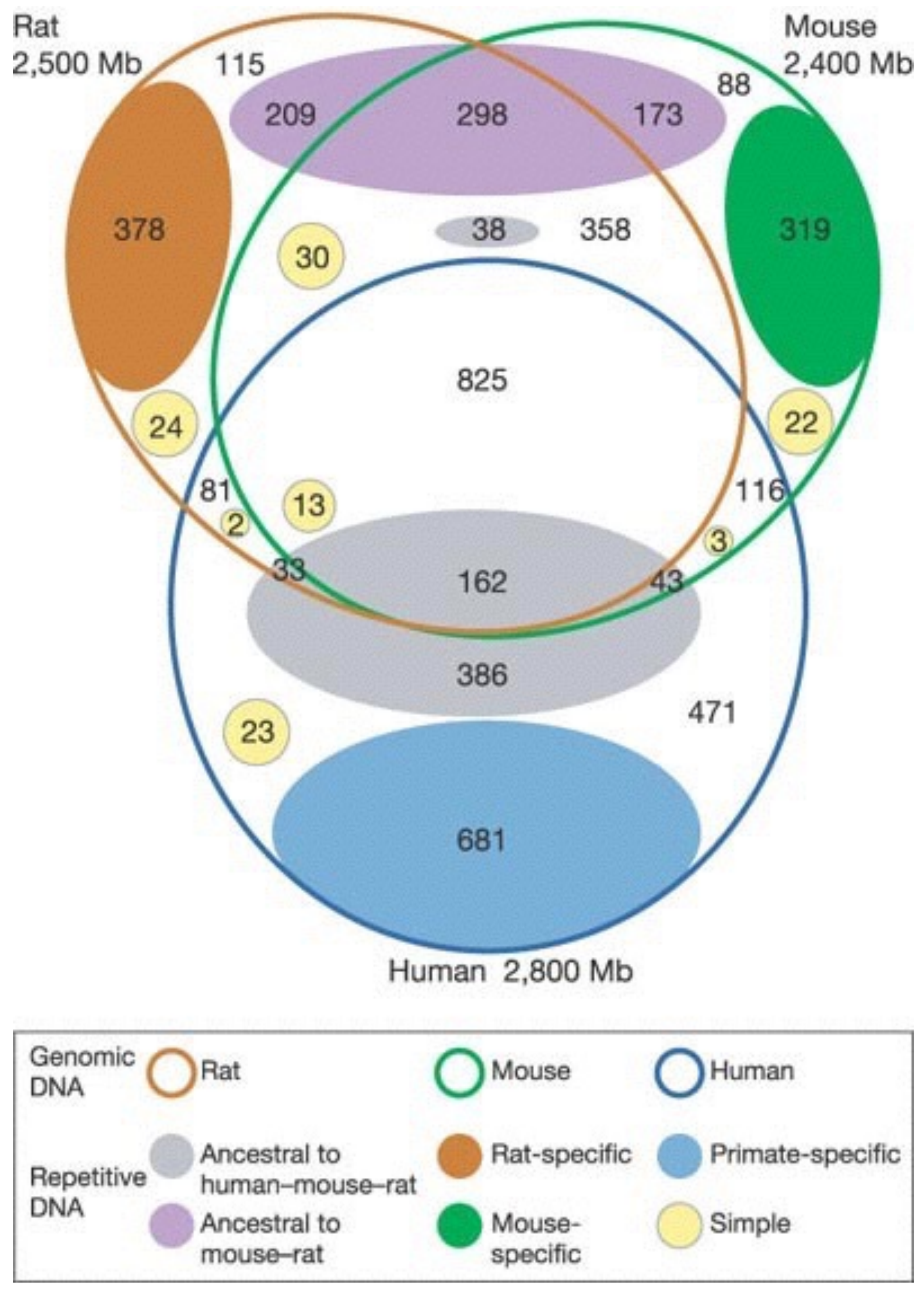




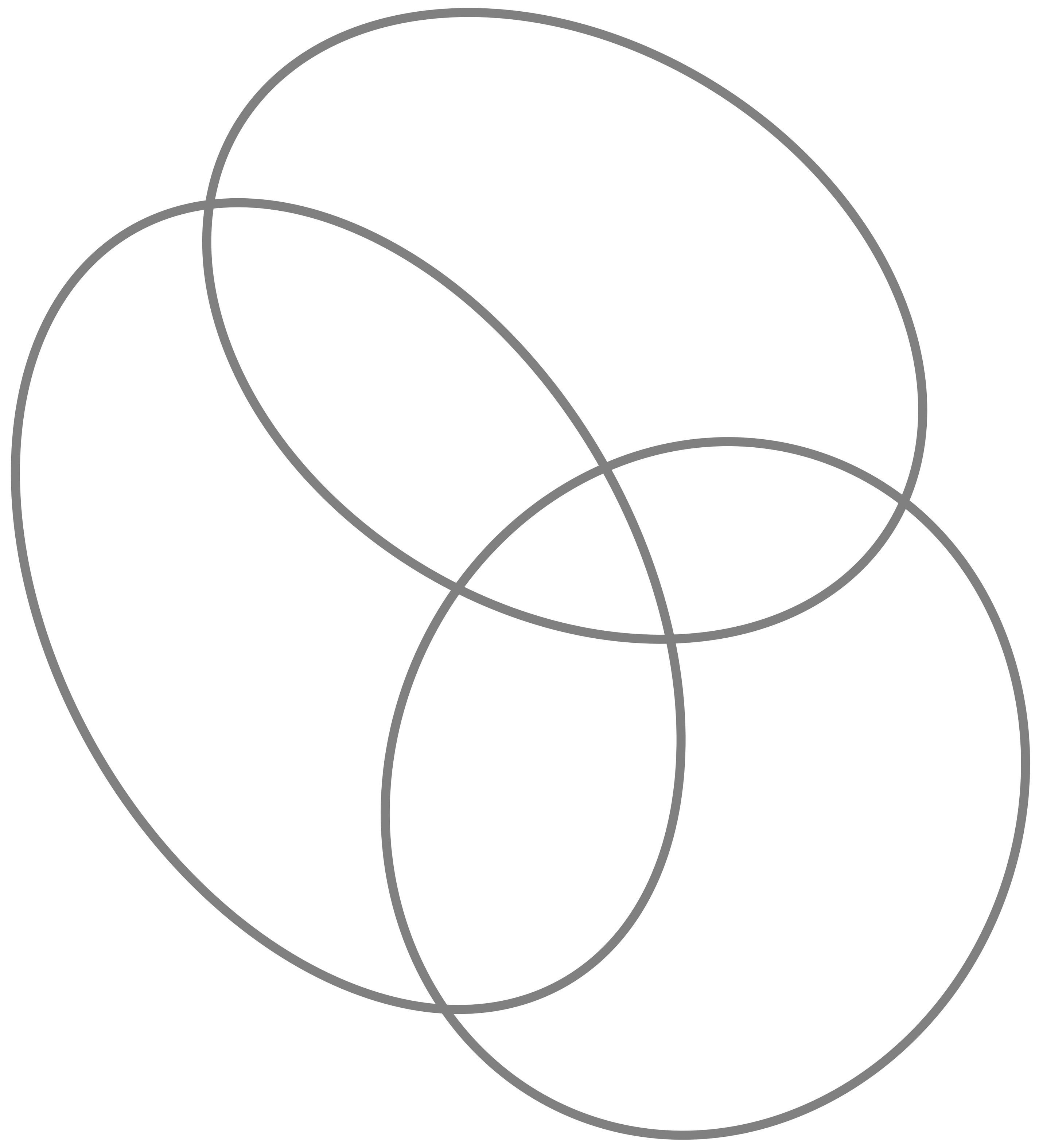
[Wiles et al., BMC Systems Biology]



[D'Hont et al., Nature, 2012]



[Gibbs et al., Nature, 2004]



[created with EulerAPE]

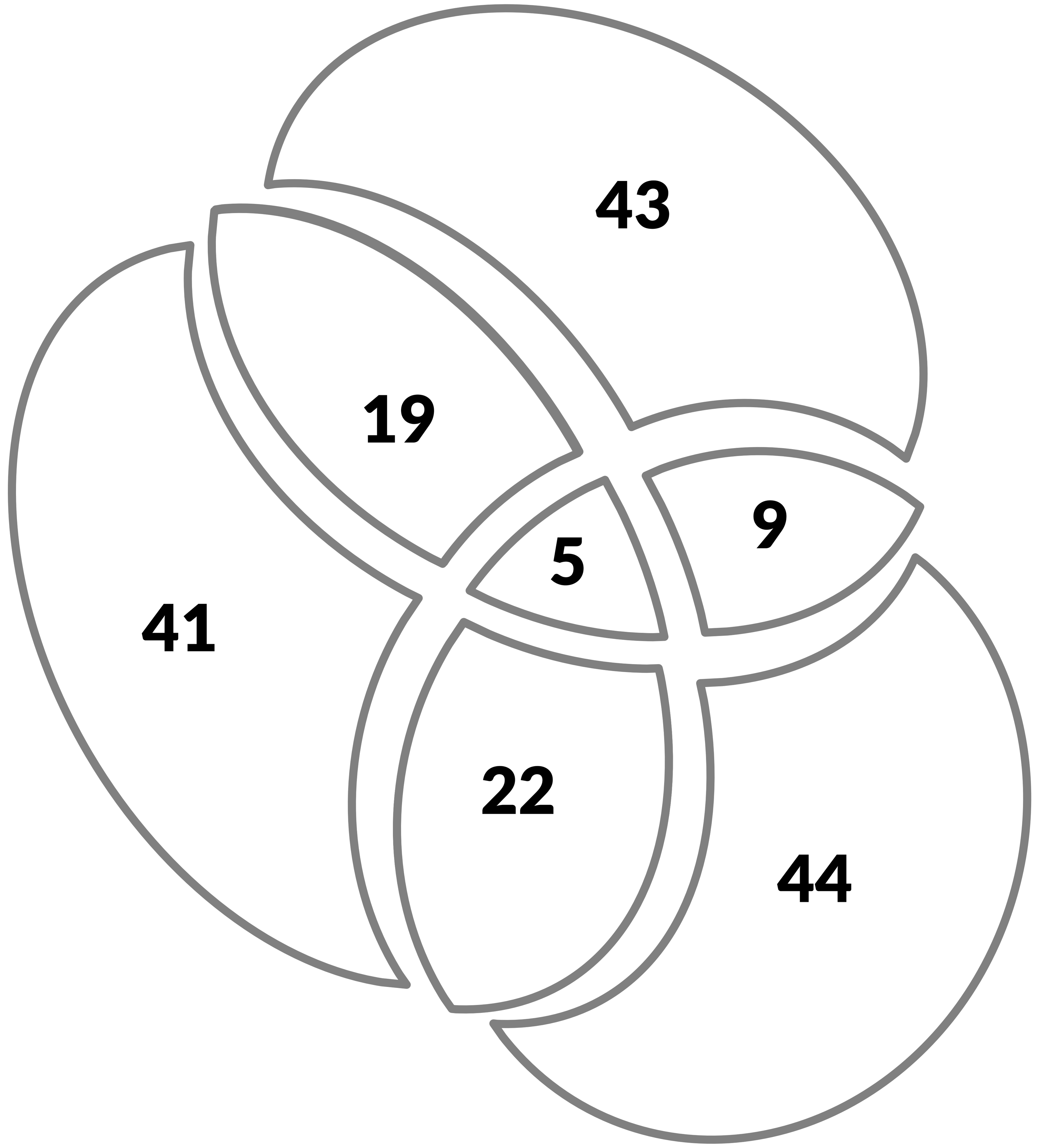
22

?

>

<

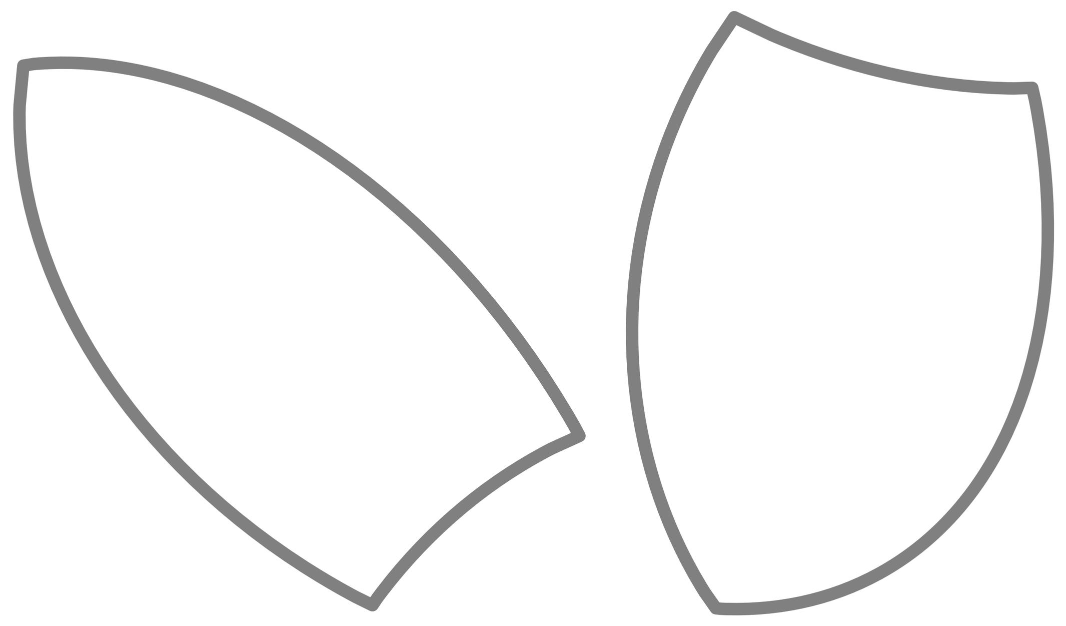
19



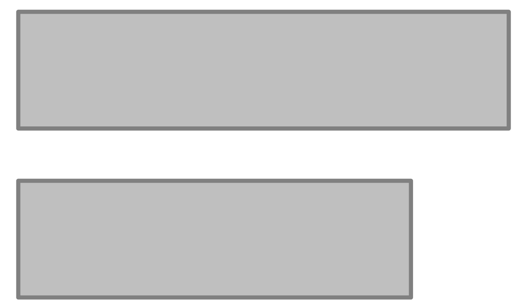
[created with EulerAPE]

Set Vis Goals

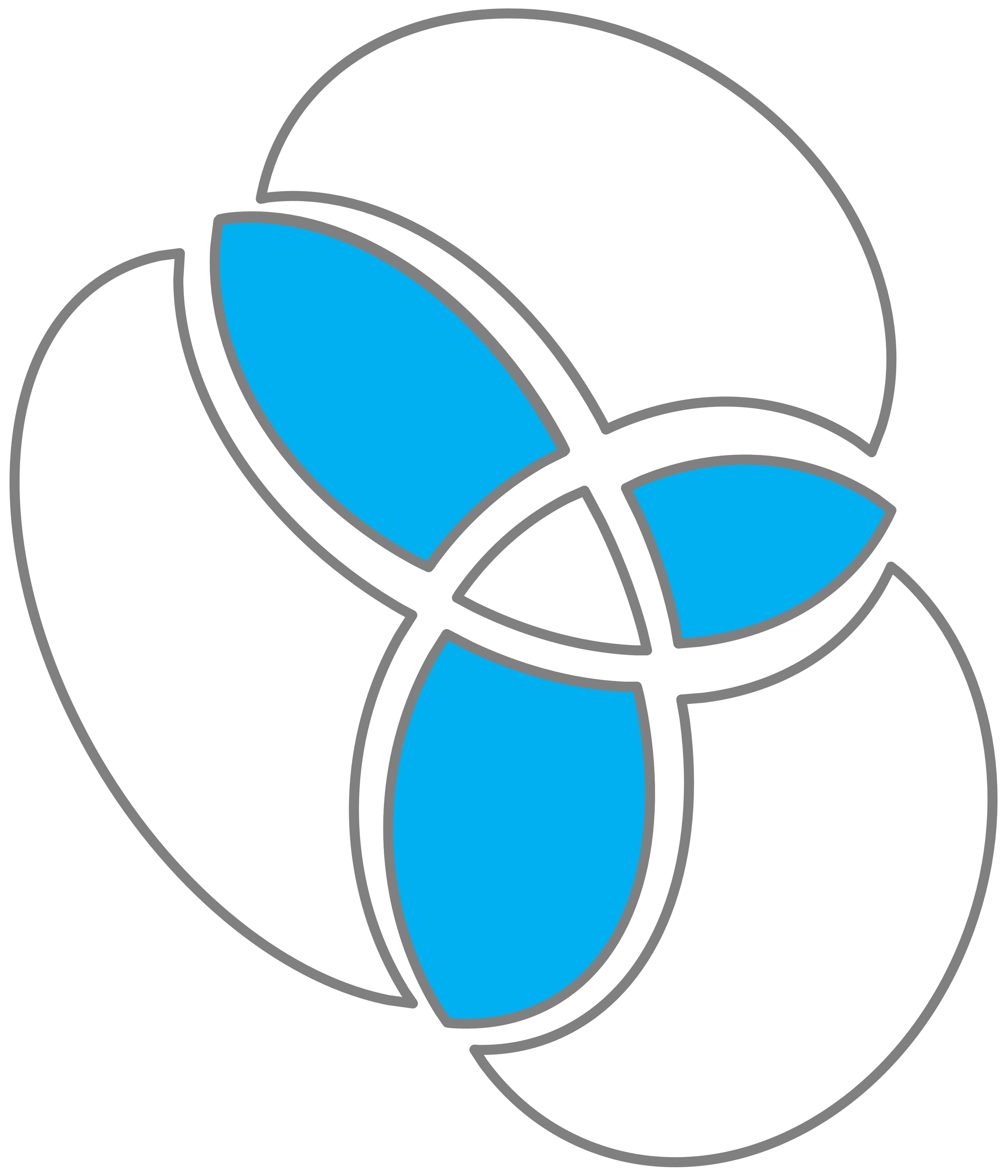
1. Efficient visual encoding



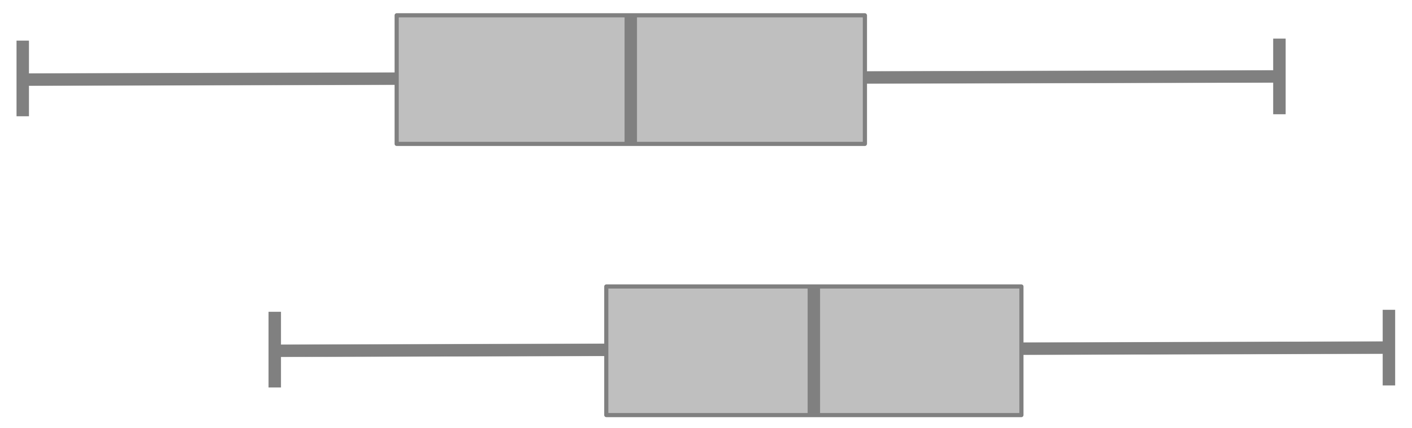
vs.



2. Creating complex slices of a dataset



3. Visualize attributes



[Movie Lens Dataset]

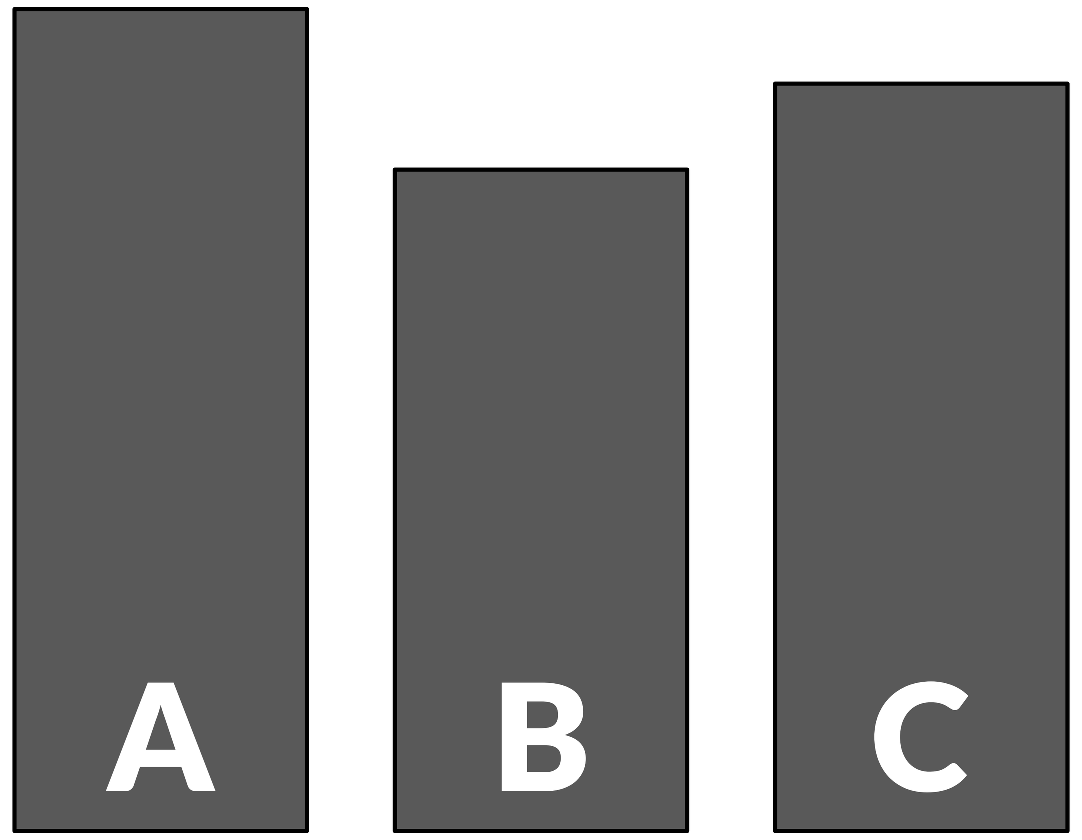
Attribute Details

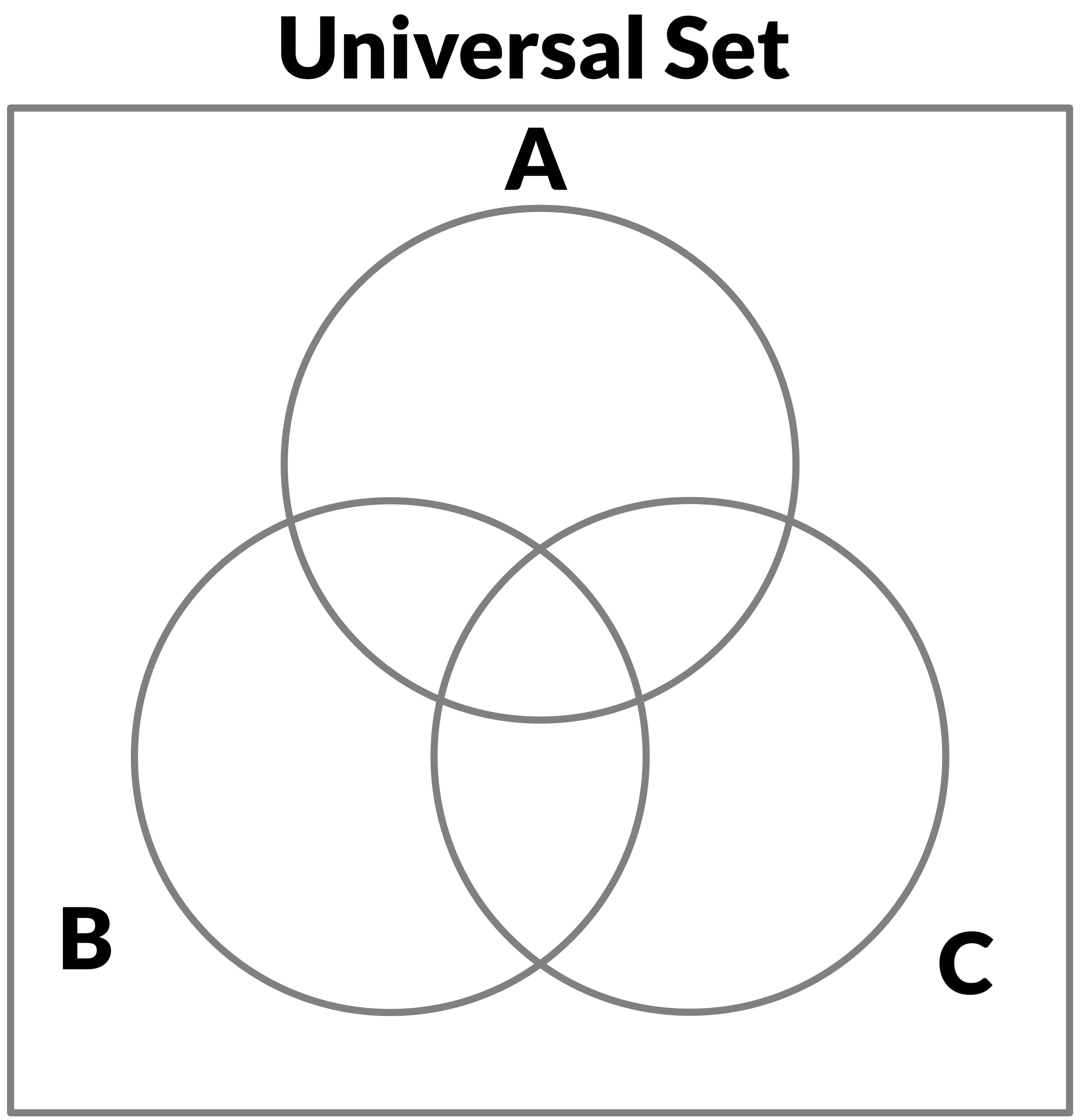
Visualizing Intersections

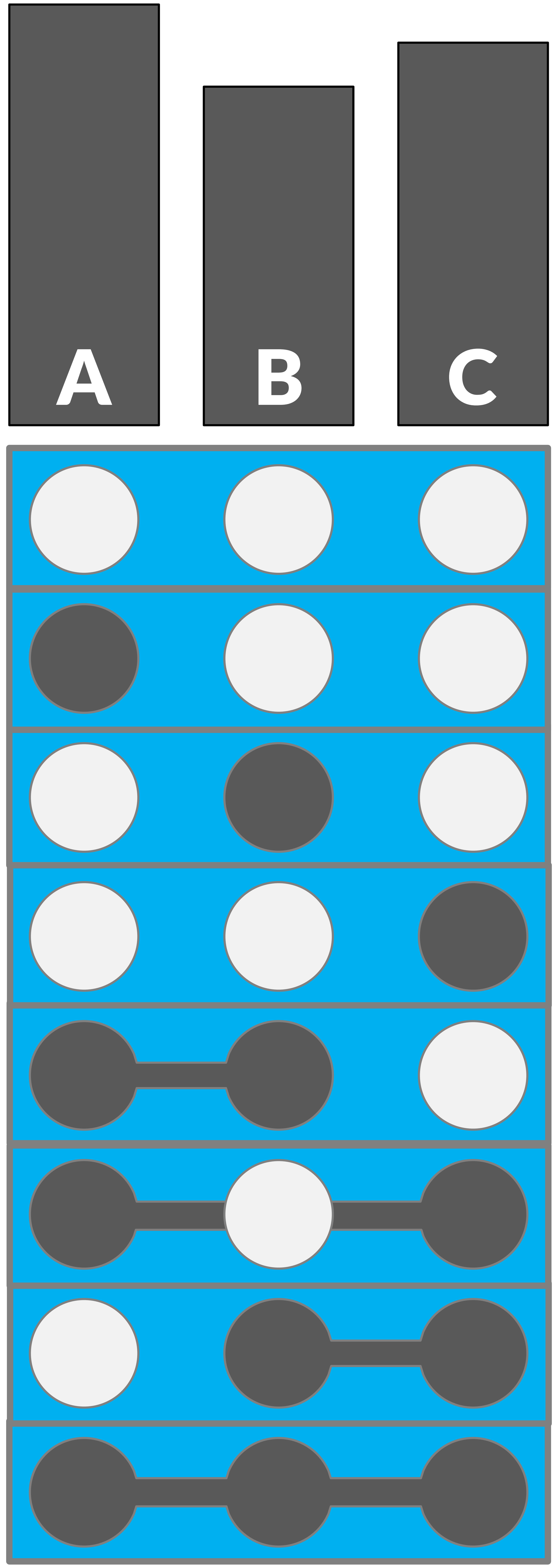
Visualizing Properties

Element List & Queries

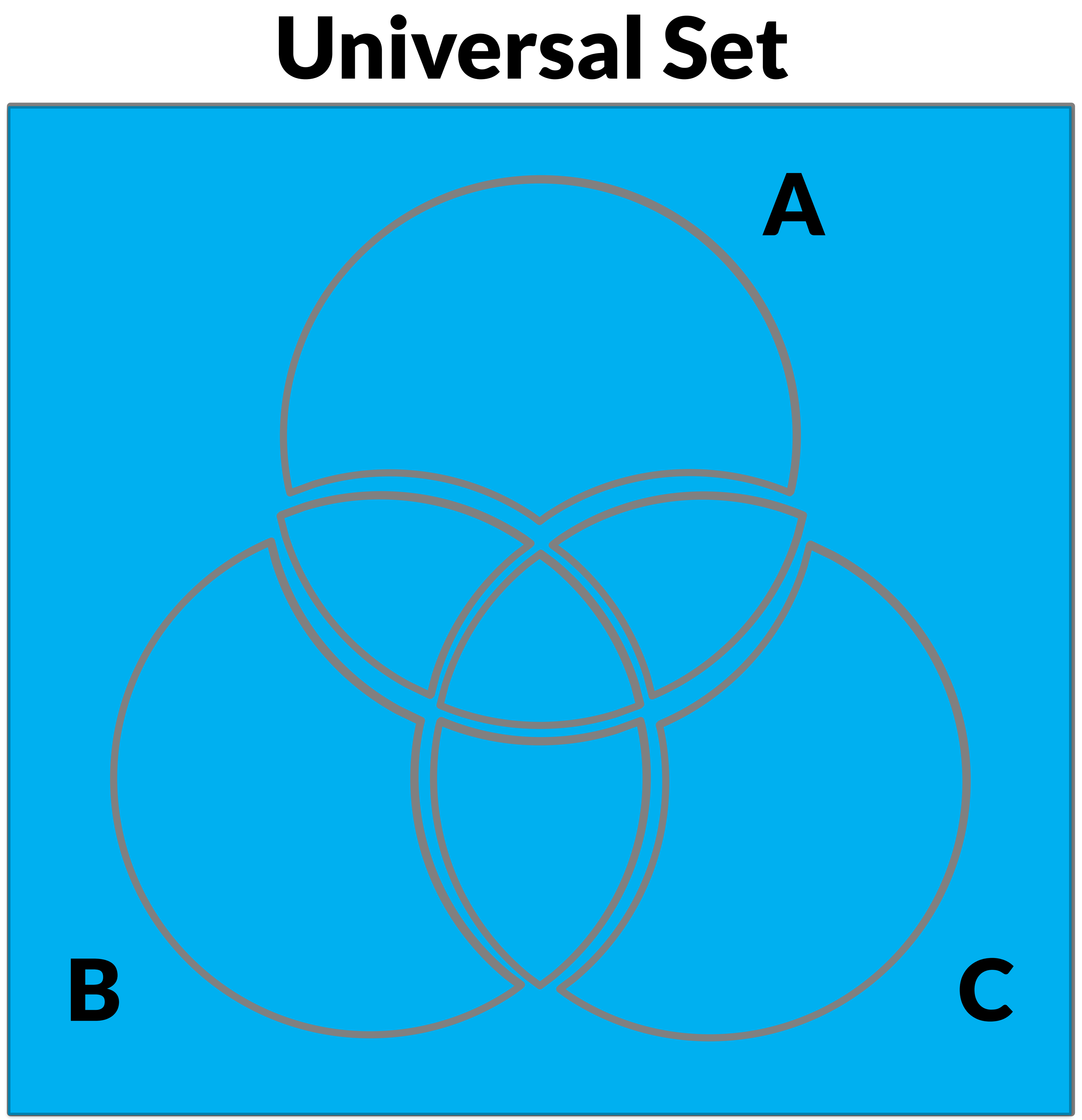
Visualizing Intersections

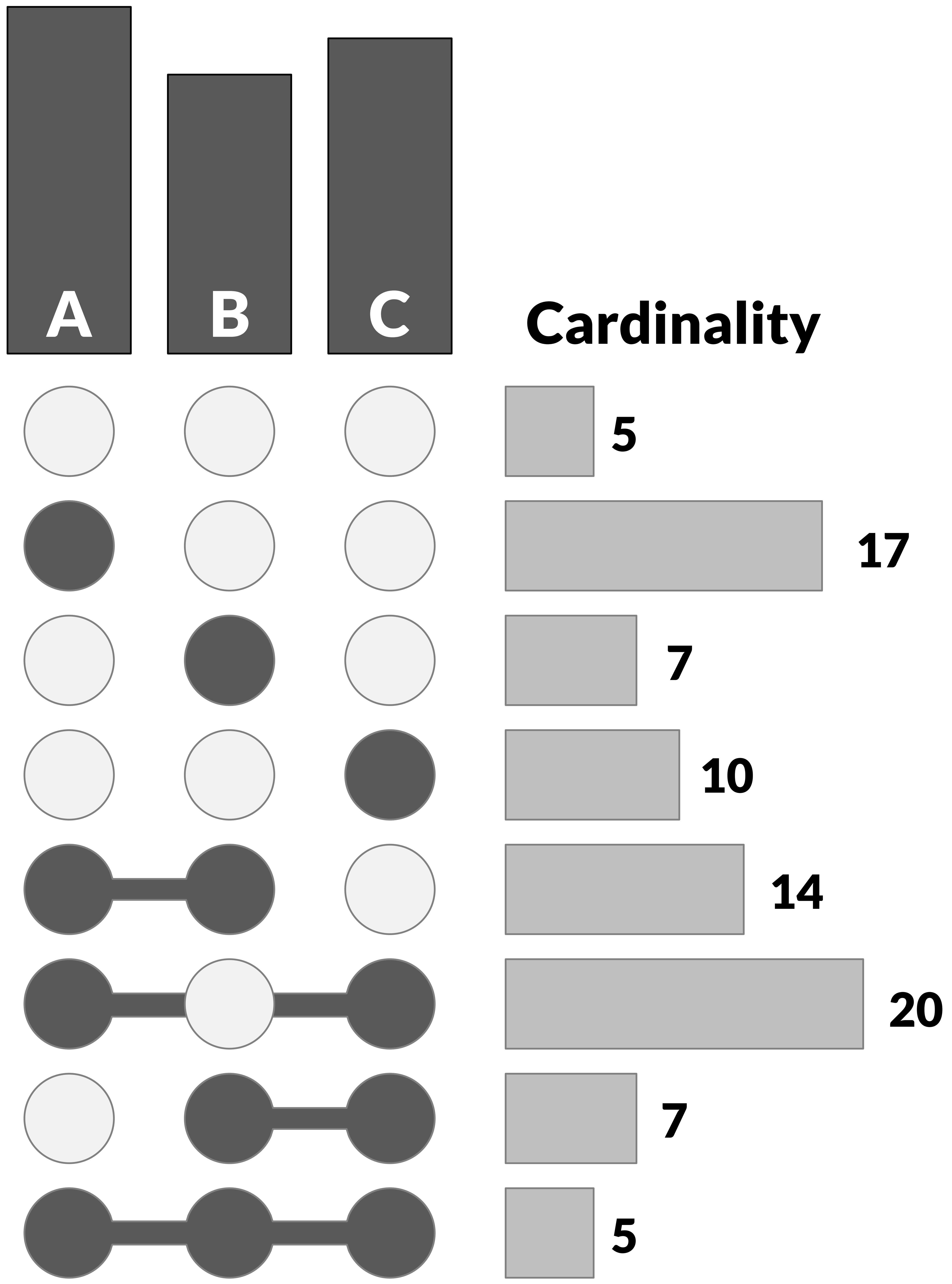

A
B
C

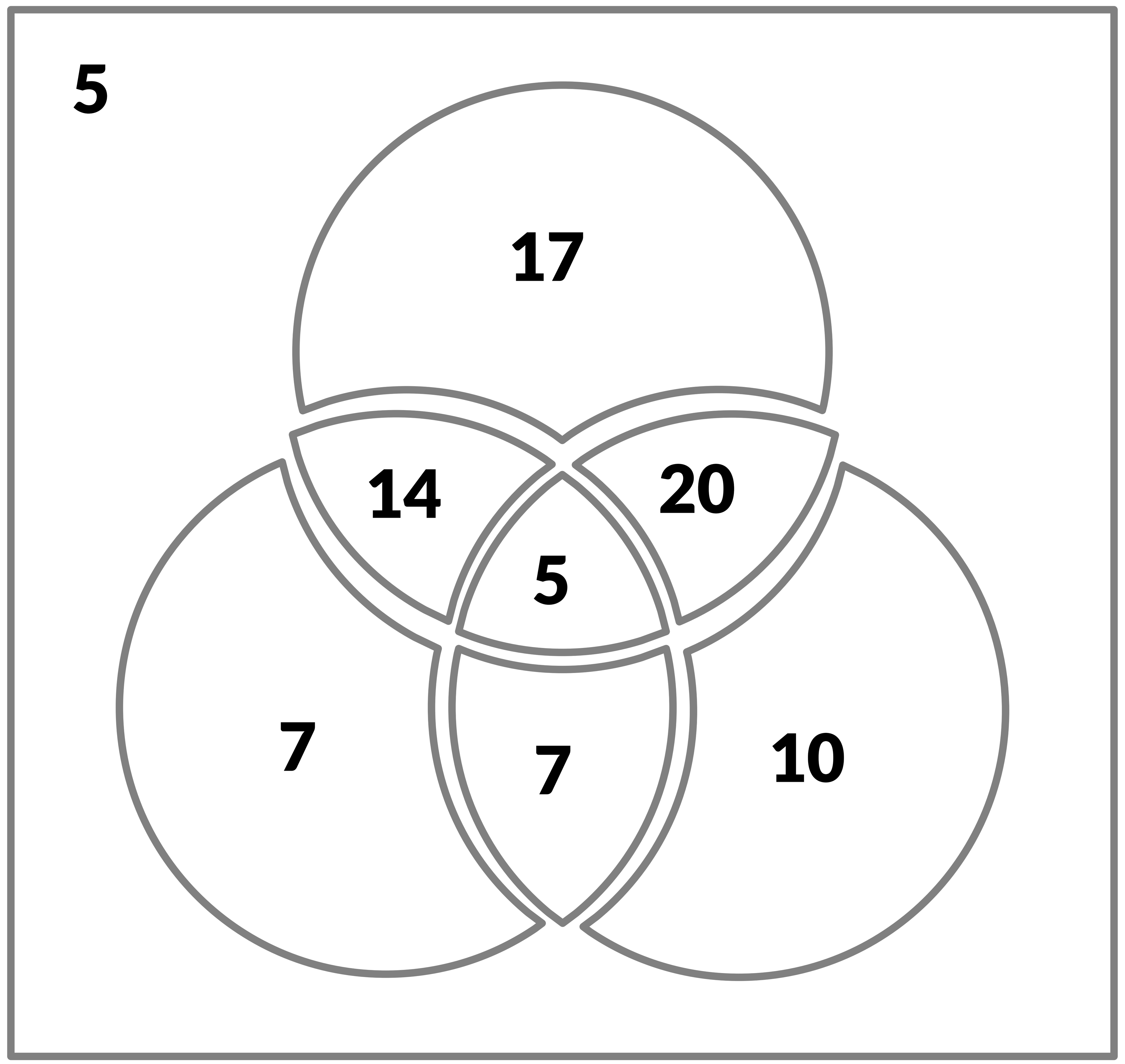

Universal Set
A
B
C



Must
Must Not




A
B
C
Cardinality
5
17
7
10
14
20
7
5


5
17
14
20
5
7
7
10

Plotting Attributes

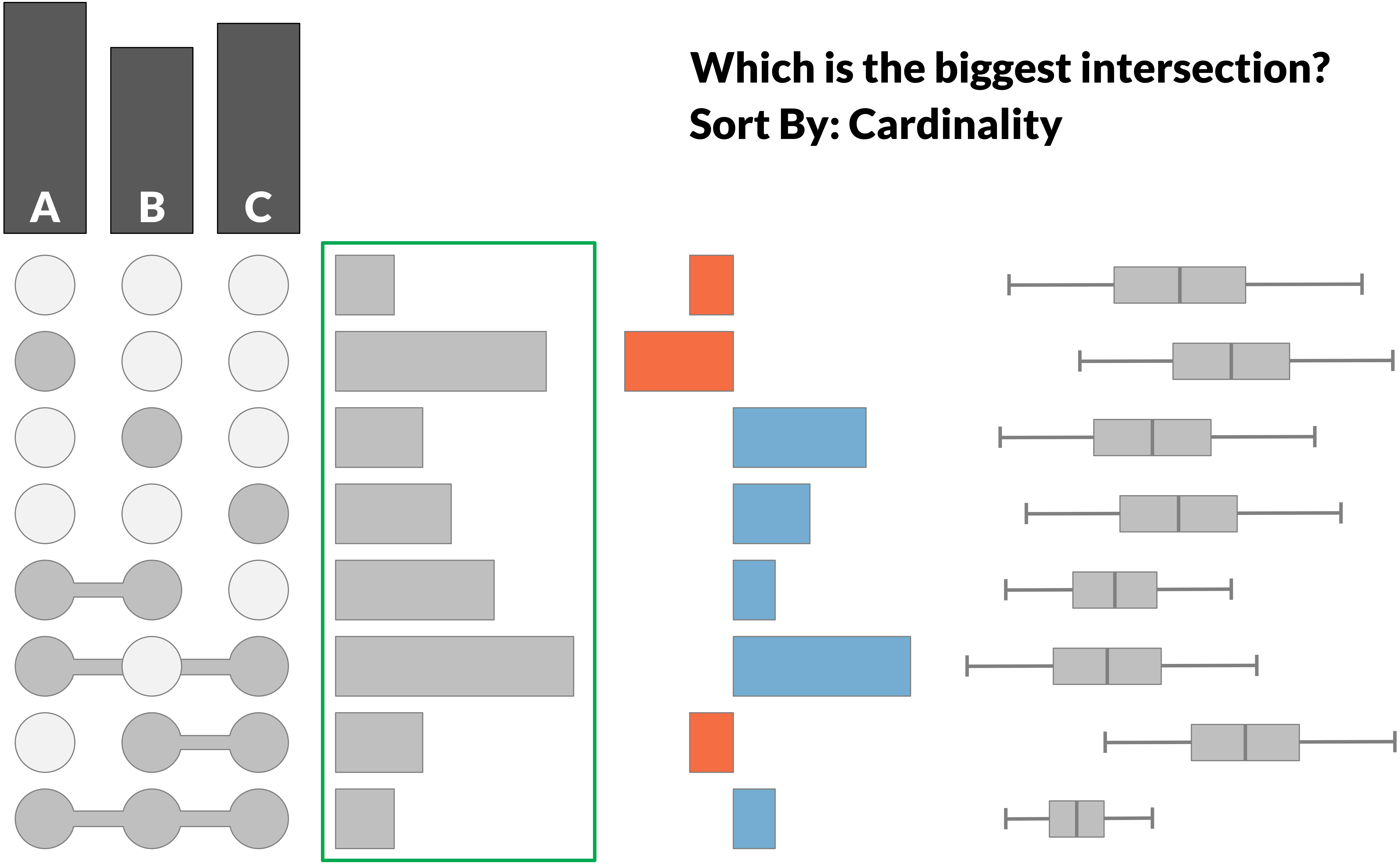

How surprising is the size of an intersection?
What's the distribution of an attribute in an intersection?
Additional Plots
A
B
C
Deviation
Attributes

First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
Description: MovieLens ratings dataset, curated and filtered by Alsallakh.
Source: http://grouplens.org/d..
Set Selection
0 - 9
Batch Add Sets
Sort Sets
Thriller
Romance
Children
Crime
Musical
Mystery
Fantasy
Western
Noir
Query
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Cardinality
Deviation
Release Date
Average Ratir
Times Watche
Drama-Comedy
Action-Comedy

Sorting

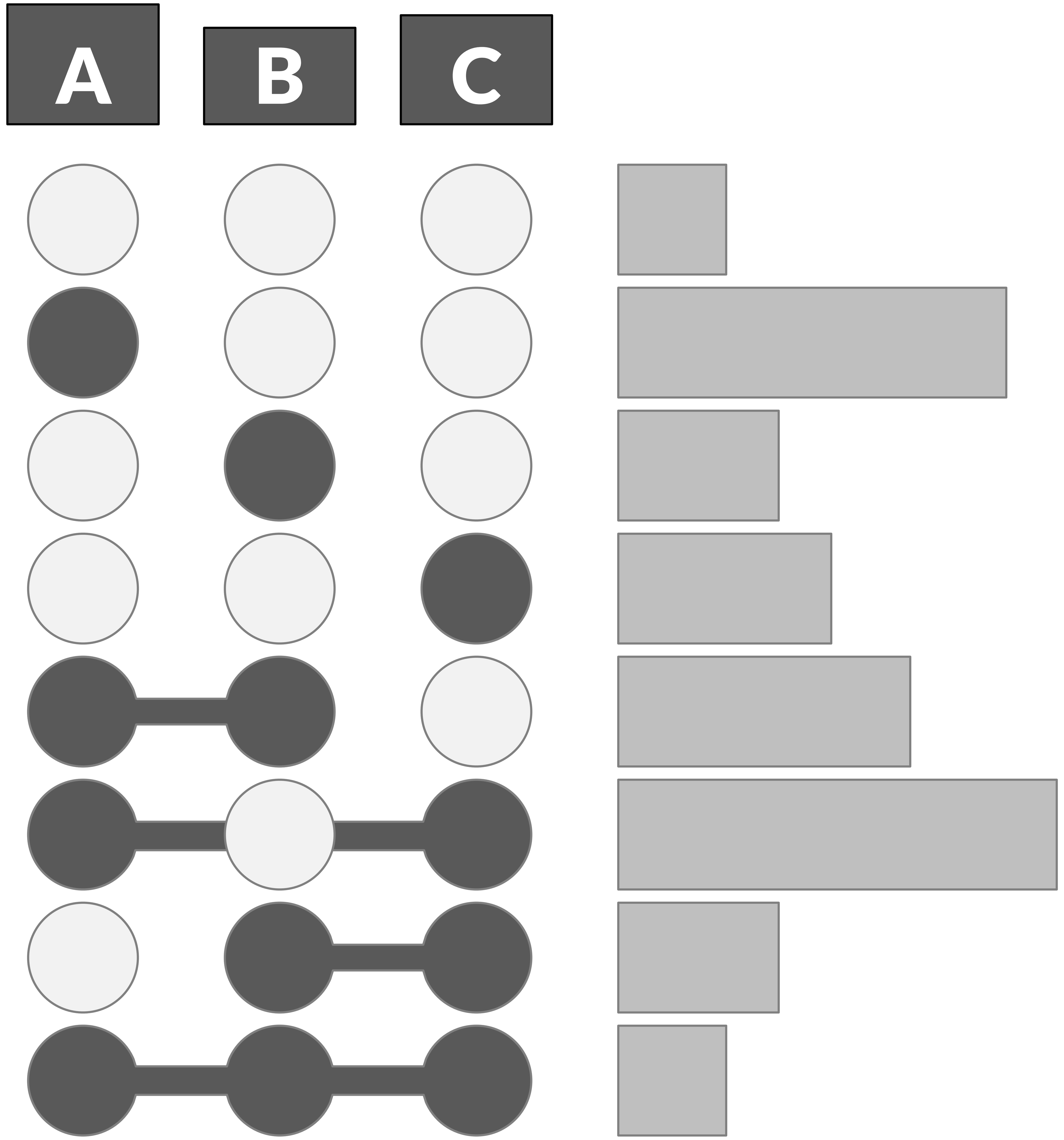
Which is the biggest intersection?

Sort By: Cardinality

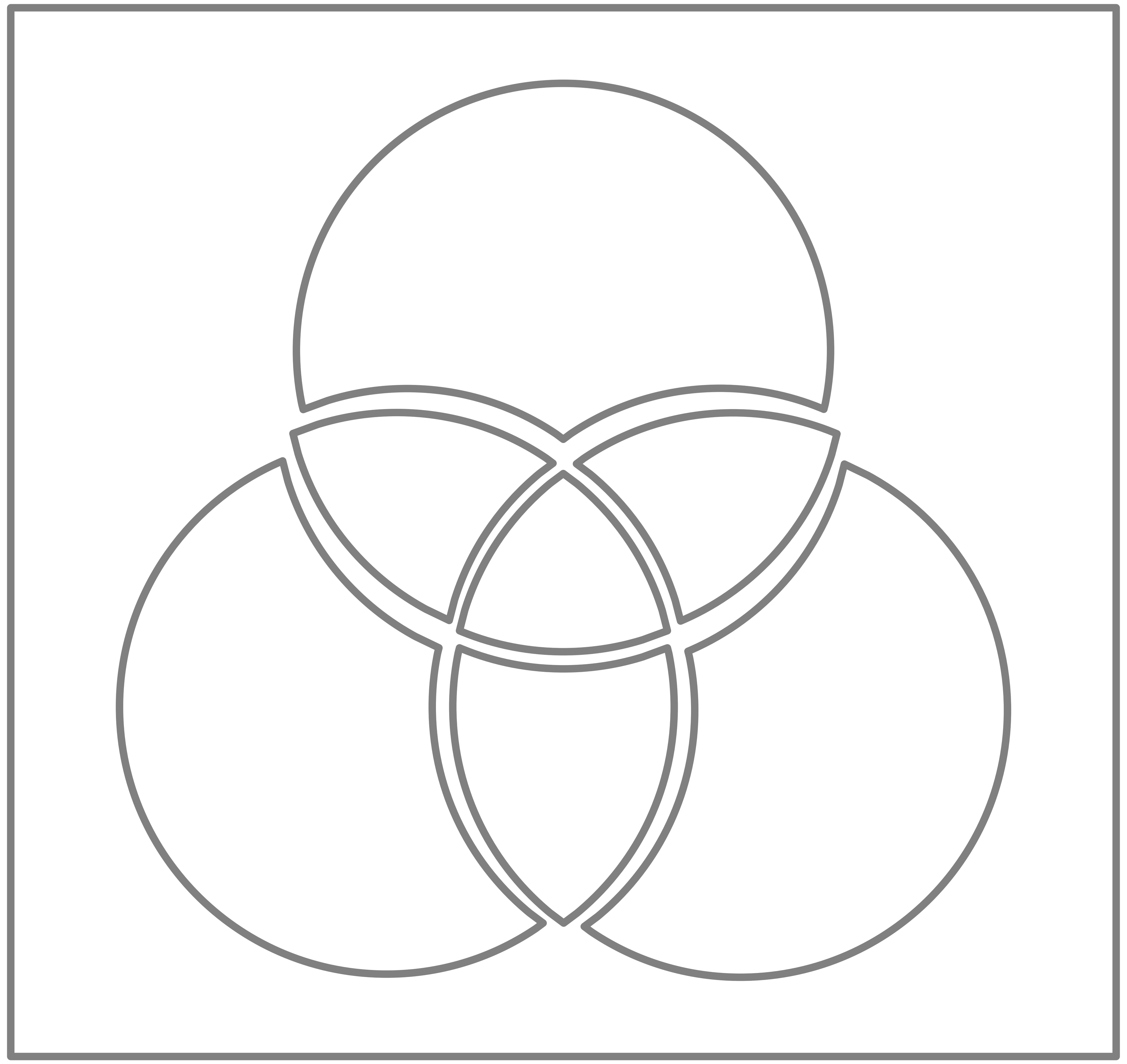



First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
Description: MovieLens ratings dataset, curated and filtered by Alsallakh.
Source: http://grouplens.org/d...
Query
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Set Selection
0 - 9
Batch Add Sets
Sort Sets
Thriller
Romance
Children
Crime
Musical
Mystery
Fantasy
Western
Noir
I
S
U
500
2000
3000
3883
Cardinality
0
200
400
600
800
1,000
Deviation
-20%
-10%
0%
10%
20%
Release Date
1,950
2,000
Average Ratin
2
4
Times Watche
0
2,000
389
823
1172
161
80
118
224
55
19
210
31
21
4
36
19
10
66
22
4
12
19

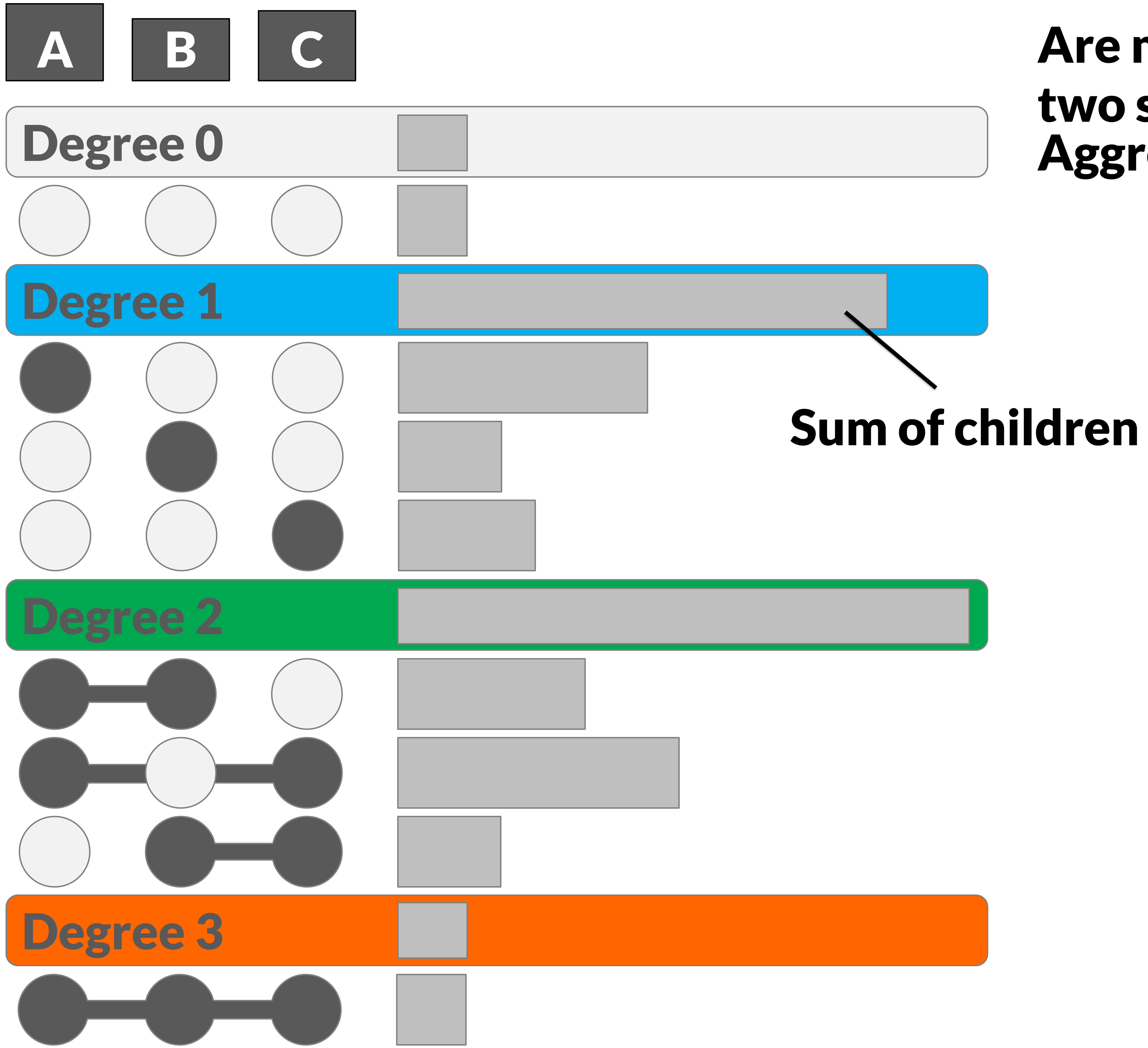
Aggregation

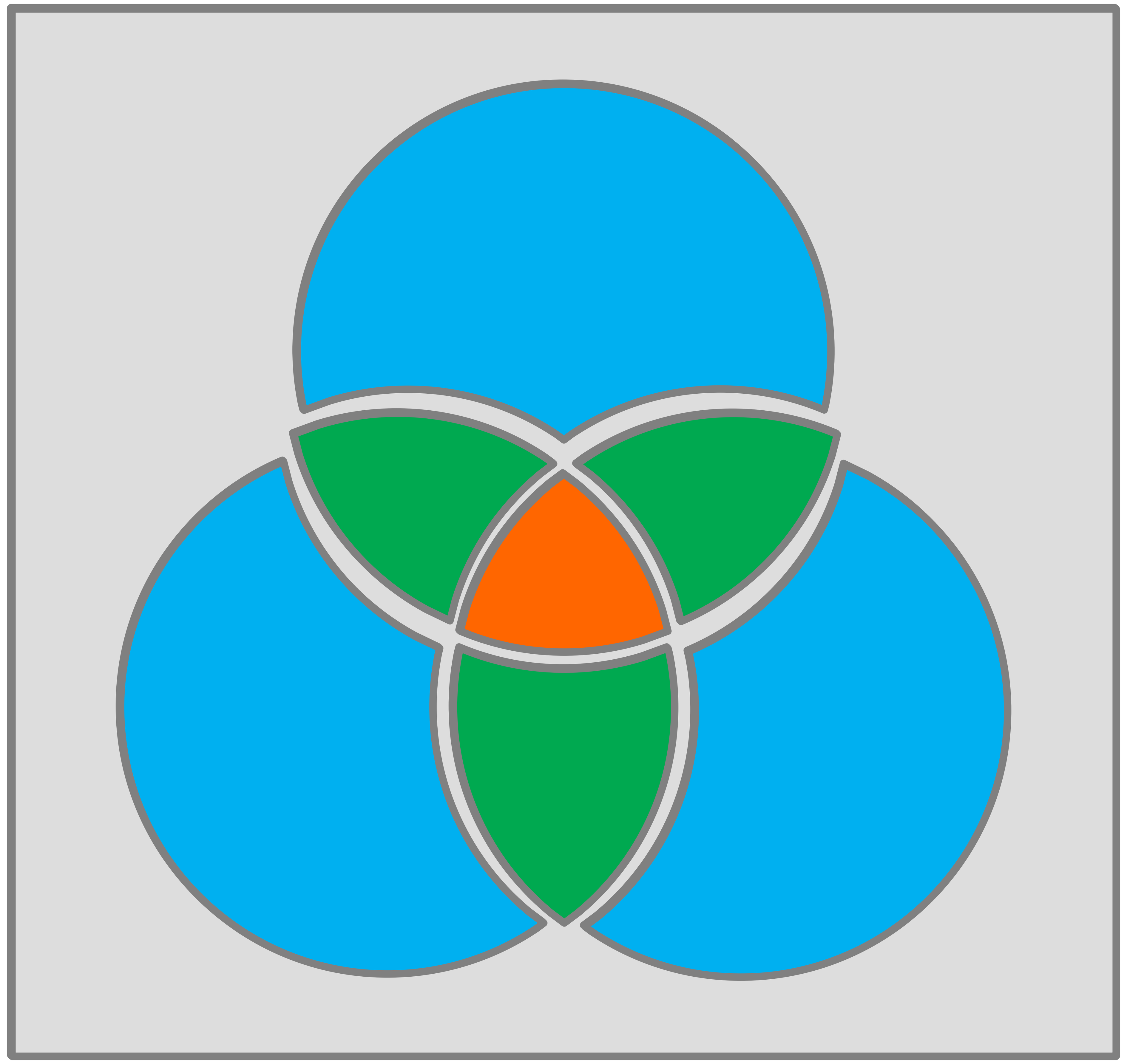


Are many items shared between two sets?
Aggregate By: Degree

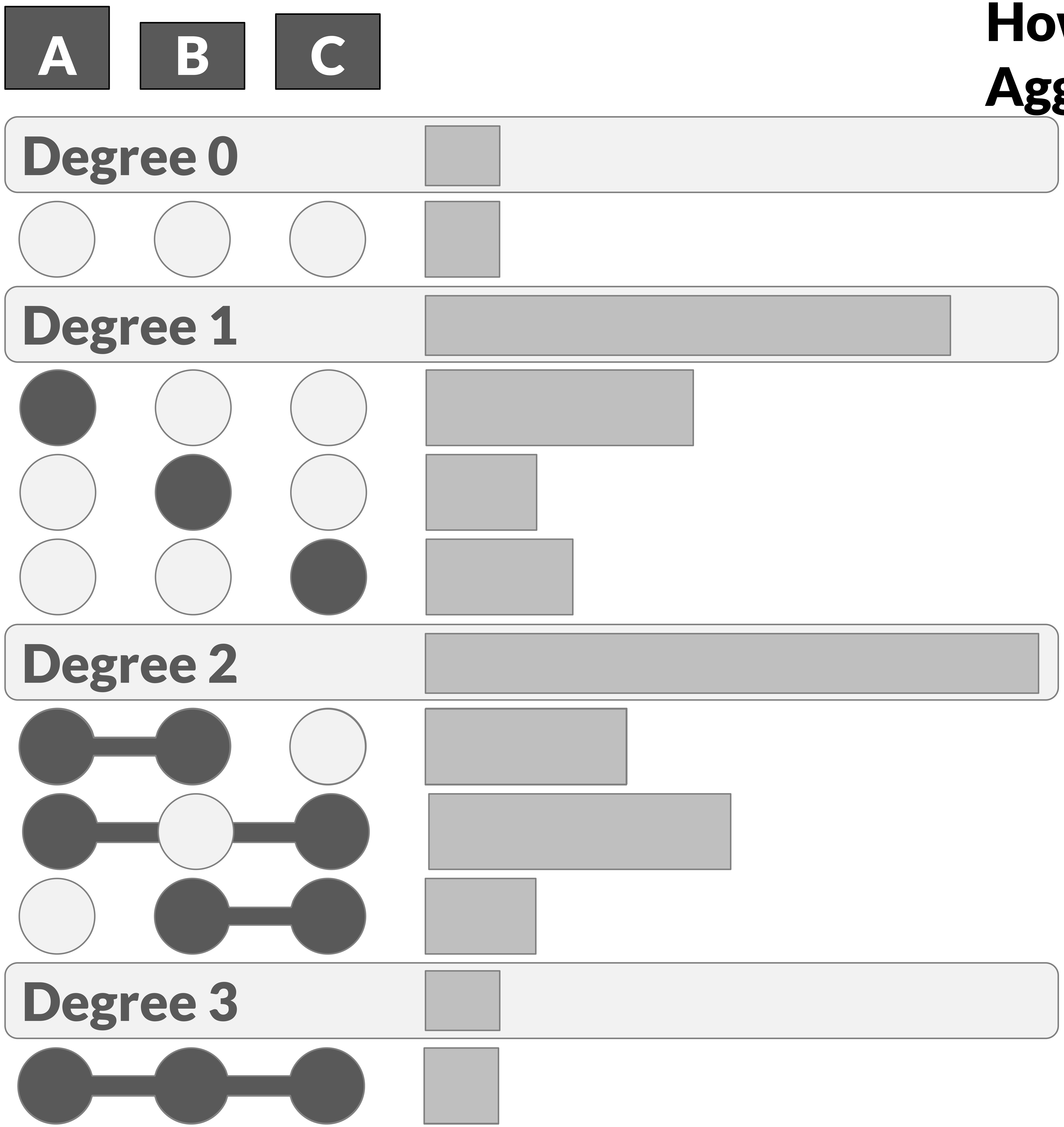


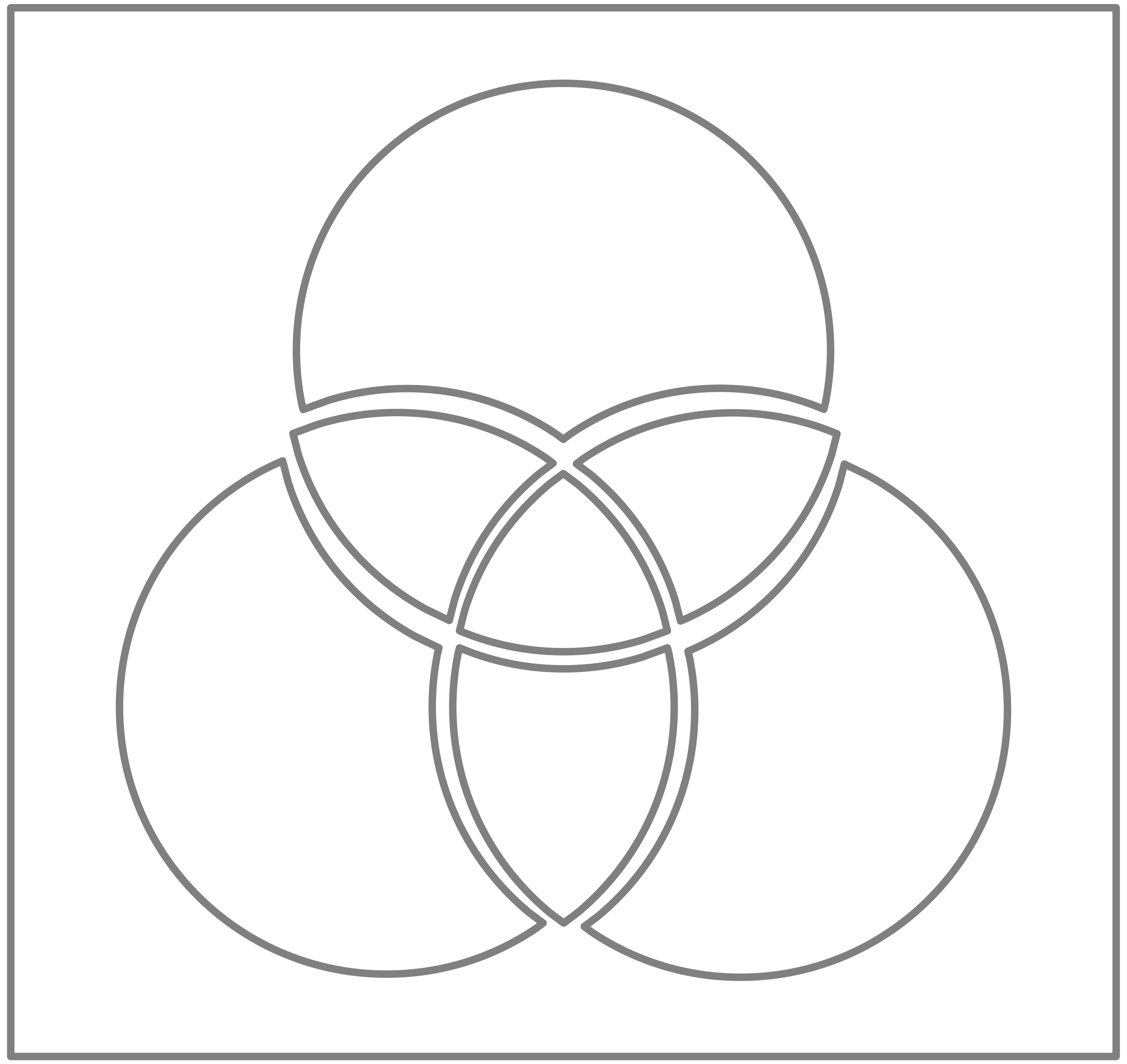
Are many items shared between two sets?
Aggregate By: Degree





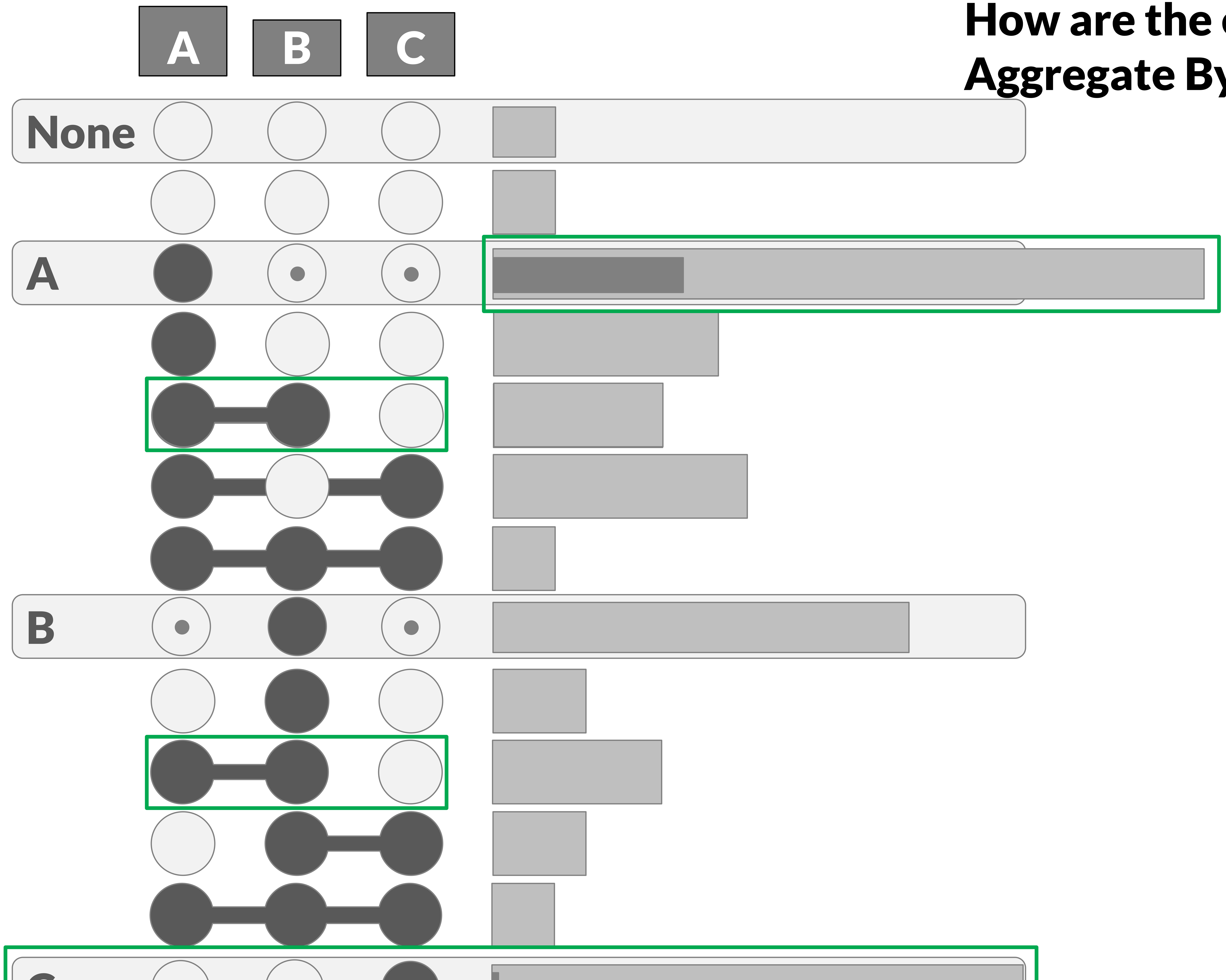
How are the elements of ‘B’ distributed?
Aggregate By: Set





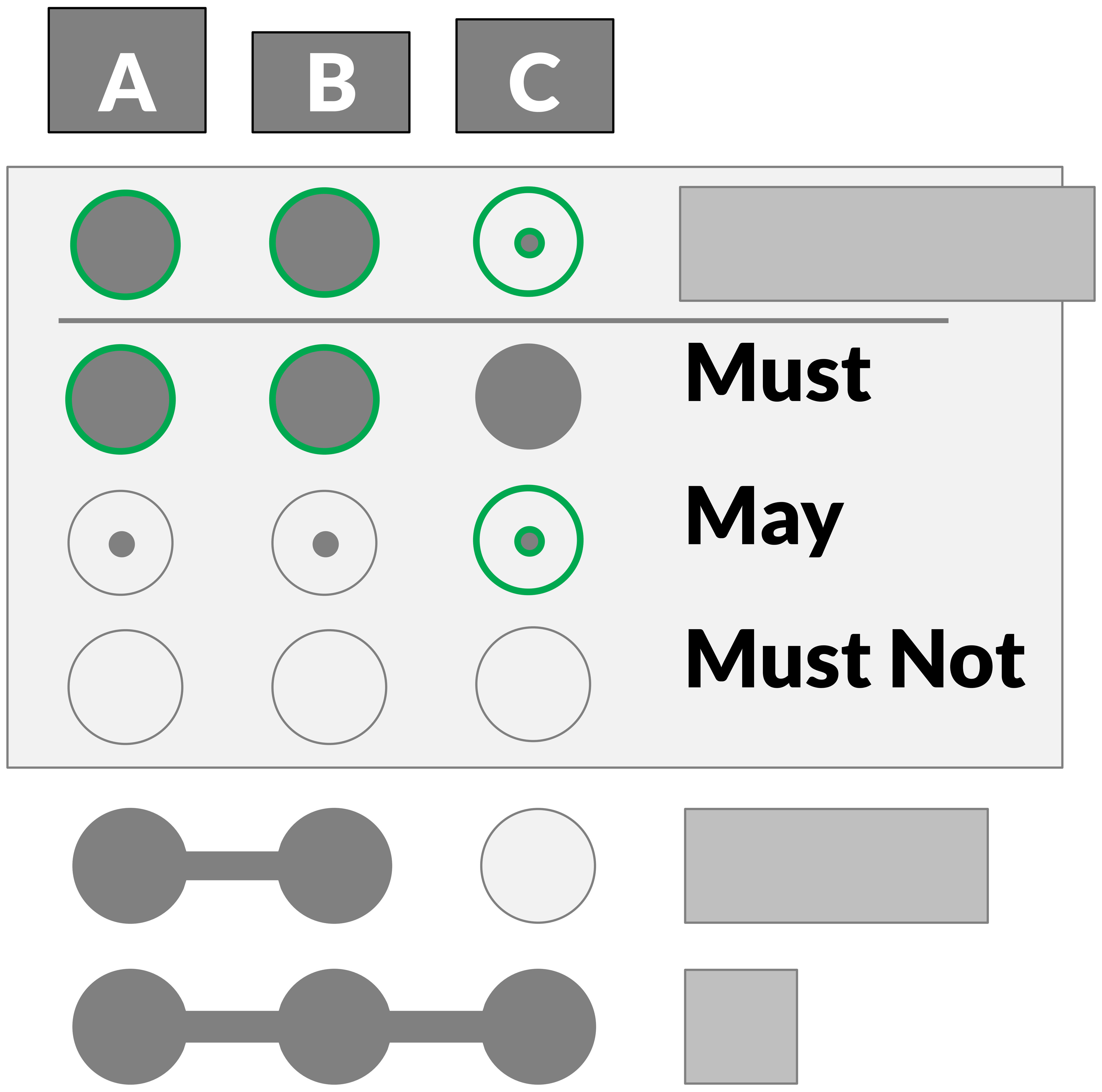
How are the elements of ‘B’ distributed?
Aggregate By: Set

How are the elements of ‘B’ distributed?
Aggregate By: Set



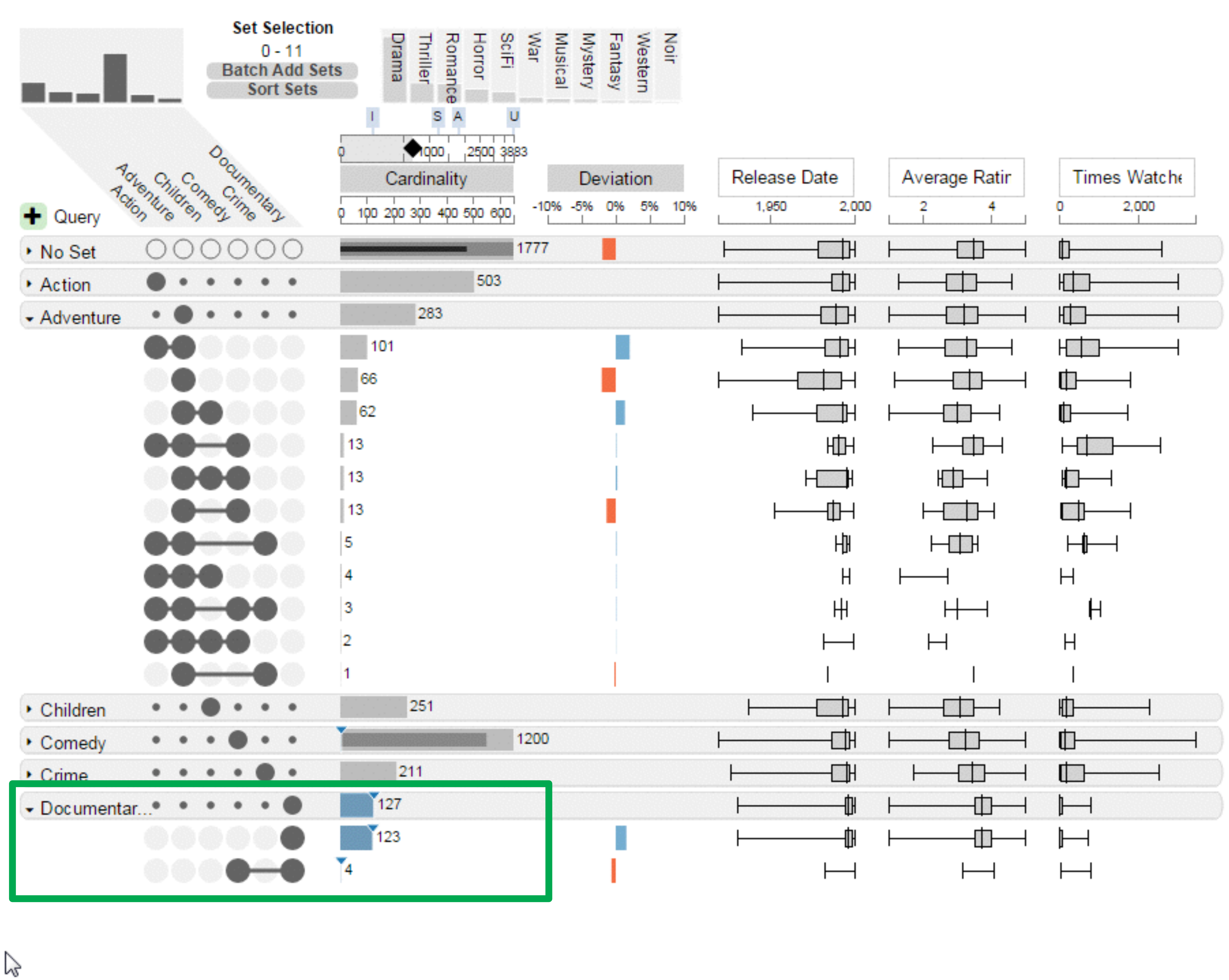
First, aggregate by
Degree
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
Description: MovieLens ratings dataset, curated and filtered by Alsallakh.
Source: http://grouplens.org/d..
Set Selection
0 - 9
Batch Add Sets
Sort Sets
Thriller
Romance
Children
Crime
Musical
Mystery
Fantasy
Western
Noir
Comedy
Drama
Action
Adventure
Documentary
Horror
SciFi
War
Query
Cardinality
Deviation
Release Date
Average Ratir
Times Watche
Degree 0 (in no set)
389
389
Degree 1 (1 set intersect.)
2652
823
1172
161
80
118
224
55
19
Degree 2 (2 set intersect.)
712
210
31
21
4
36
19
10
66
22

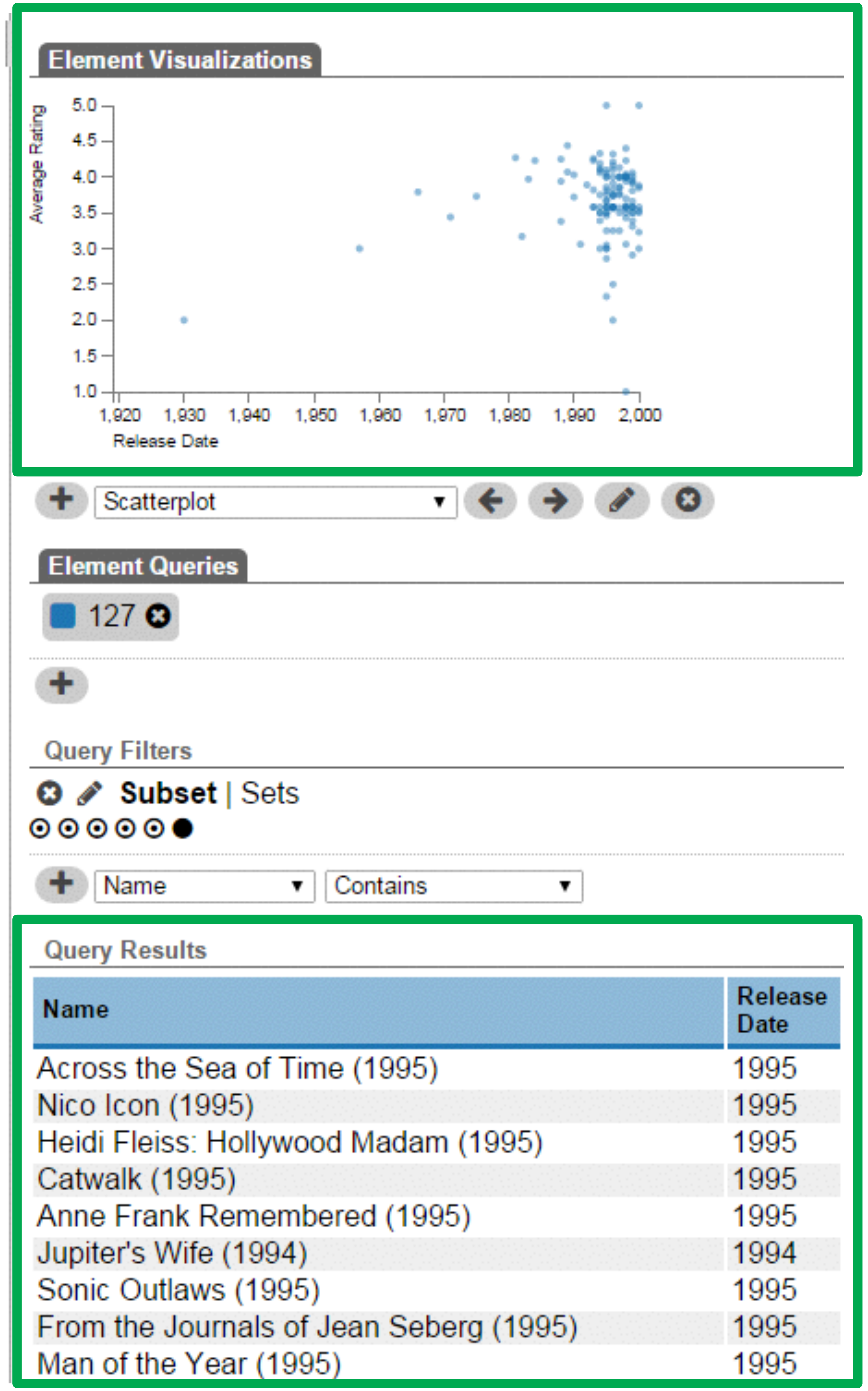
Queries


A
B
C
Must
May
Must Not

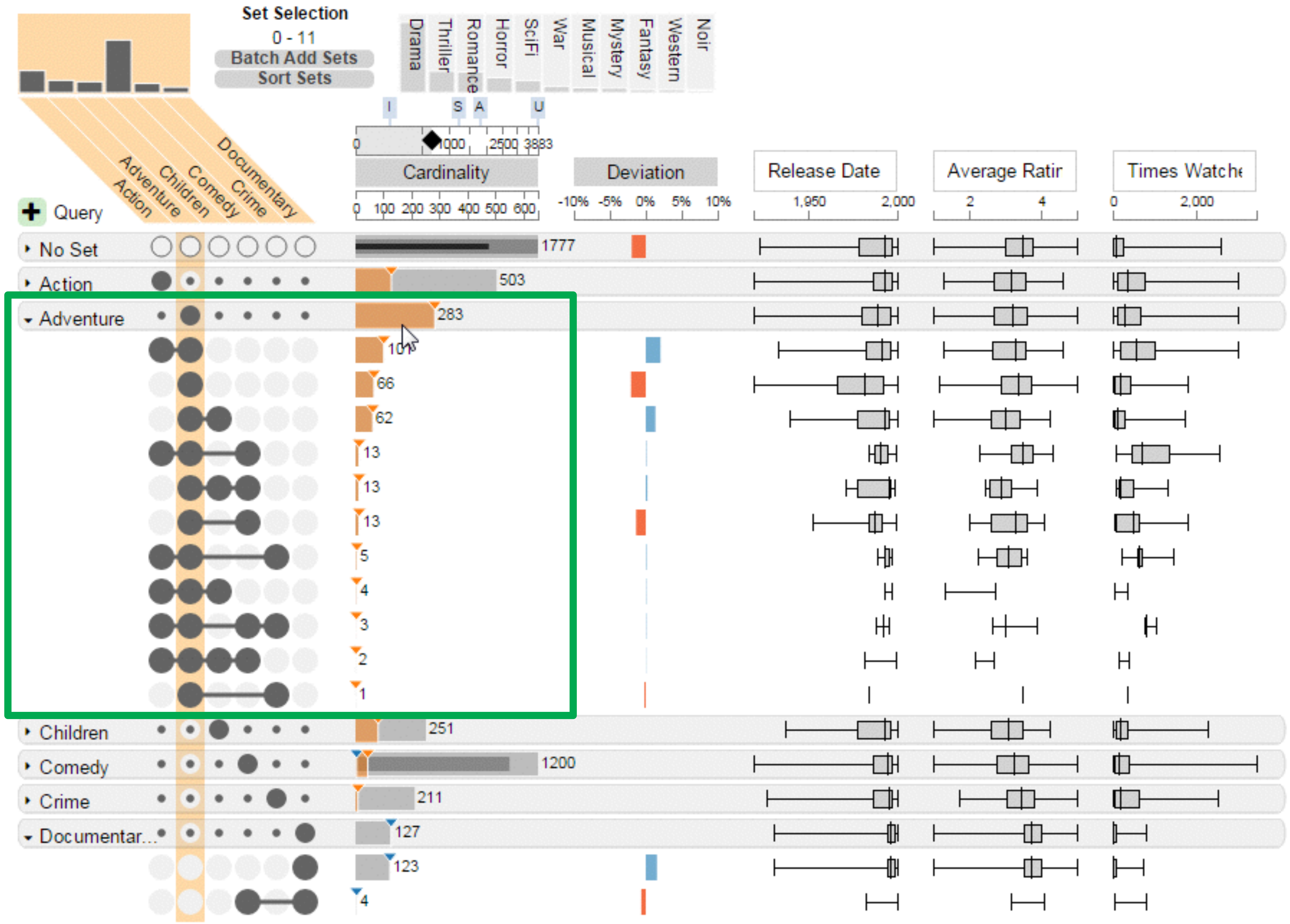
Set Selection
0 - 7
Batch Add Sets
Sort Sets
SciFi
War
Musical
Mystery
Fantasy
Western
Noir
I
S
U
2000
3883
Cardinality
Deviation
Query
Action
Adventure
Children
Comedy
Crime
Documentary
Drama
Thriller
Horror
Romance
0
50
100
150
200
250
-20%
-10%
0%
10%
128
113
37
67
585
32
119
938
150
213
51
62
6
33
18
Element Visualizations
No visualizations configured. Click + button to add a new visua
Scatterplot
Element Queries
No queries. Click + button to add a new query.
Query Filters
No active query.
Query Results
No active query.

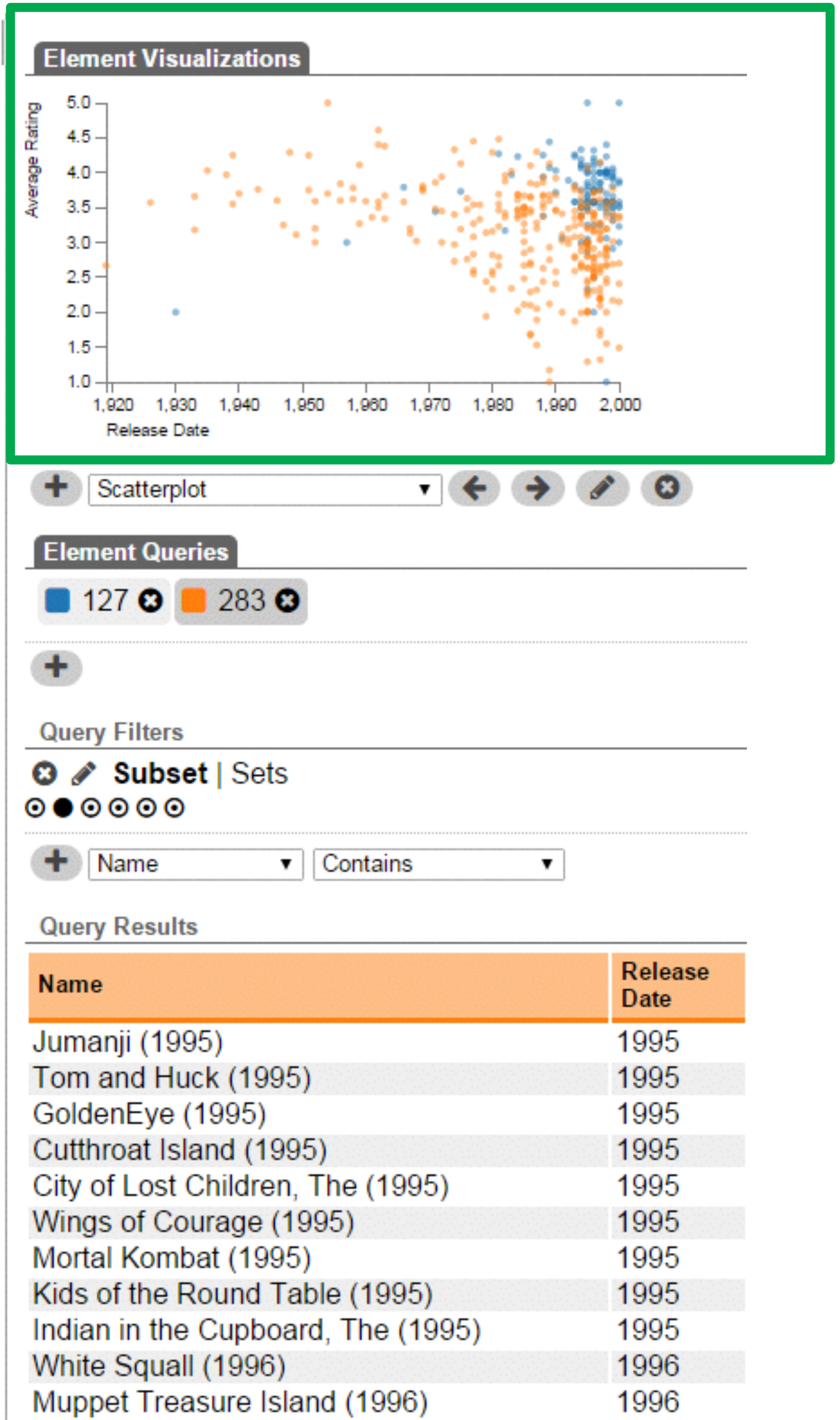
Elements & Attributes





How do documentaries compare to adventure movies?



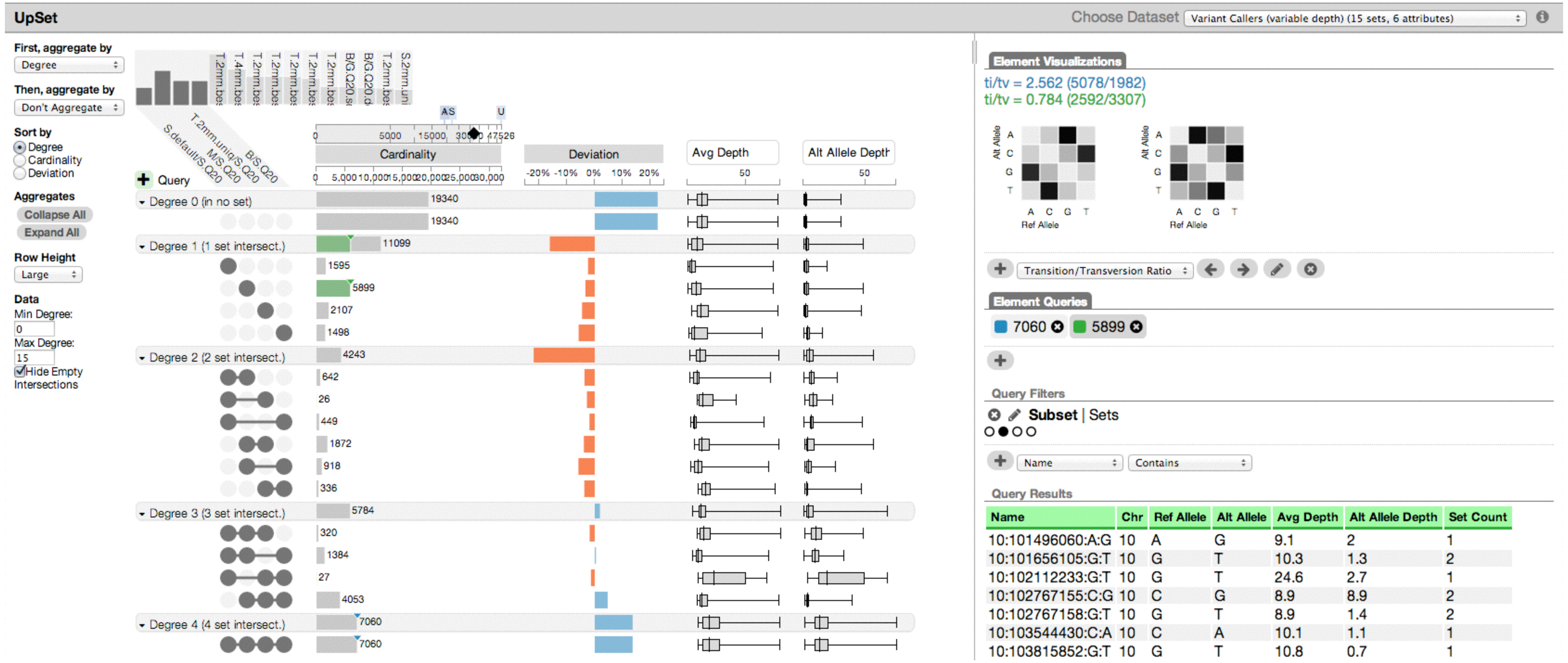


How do documentaries compare to adventure movies?

Applications

Applications

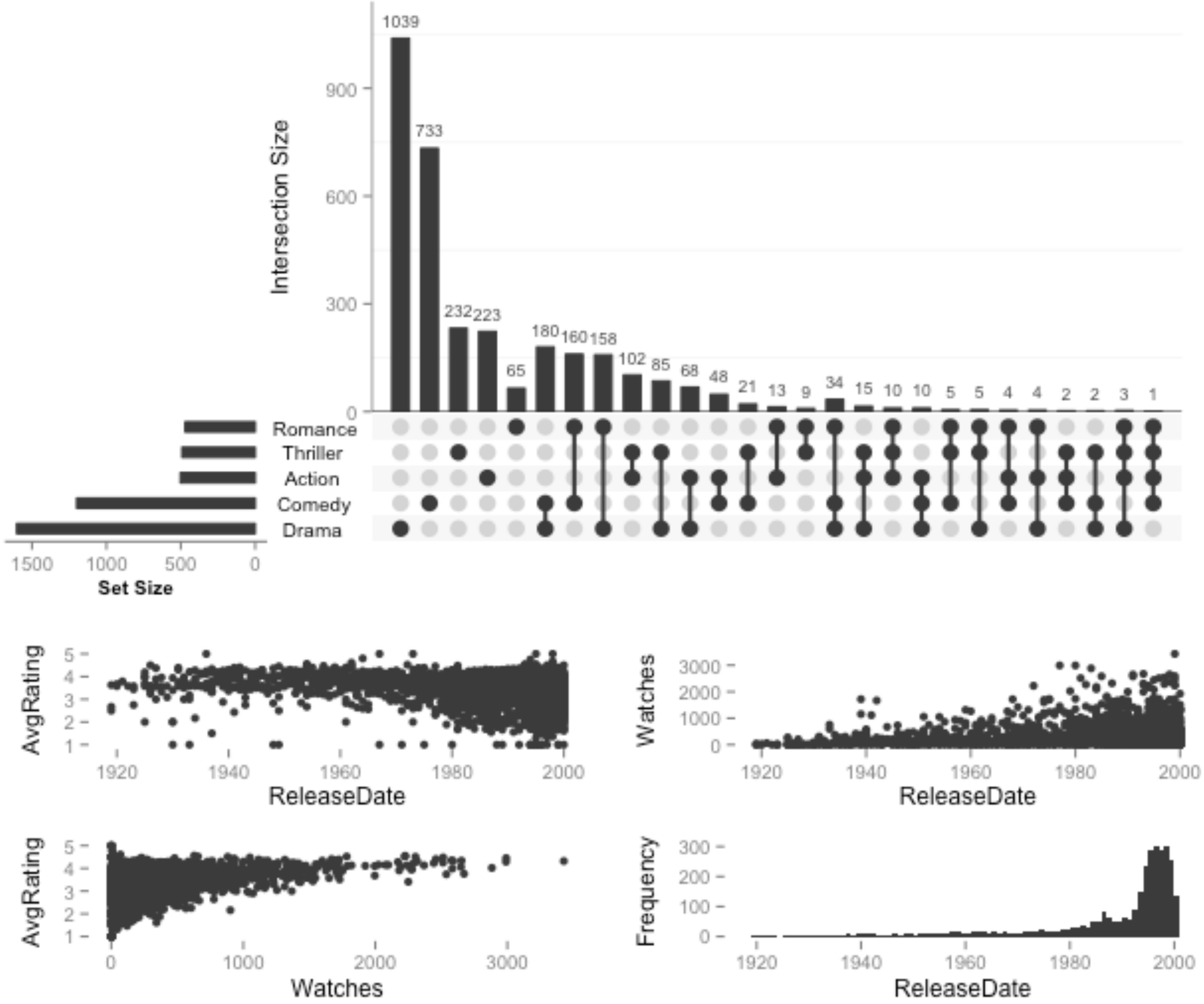
- Genetics
- Economics
- Pharmacology
- Social Networks



R-Version: UpSetR

Developed at HMS

Some design adaptions



The Banana Chart Redesigned

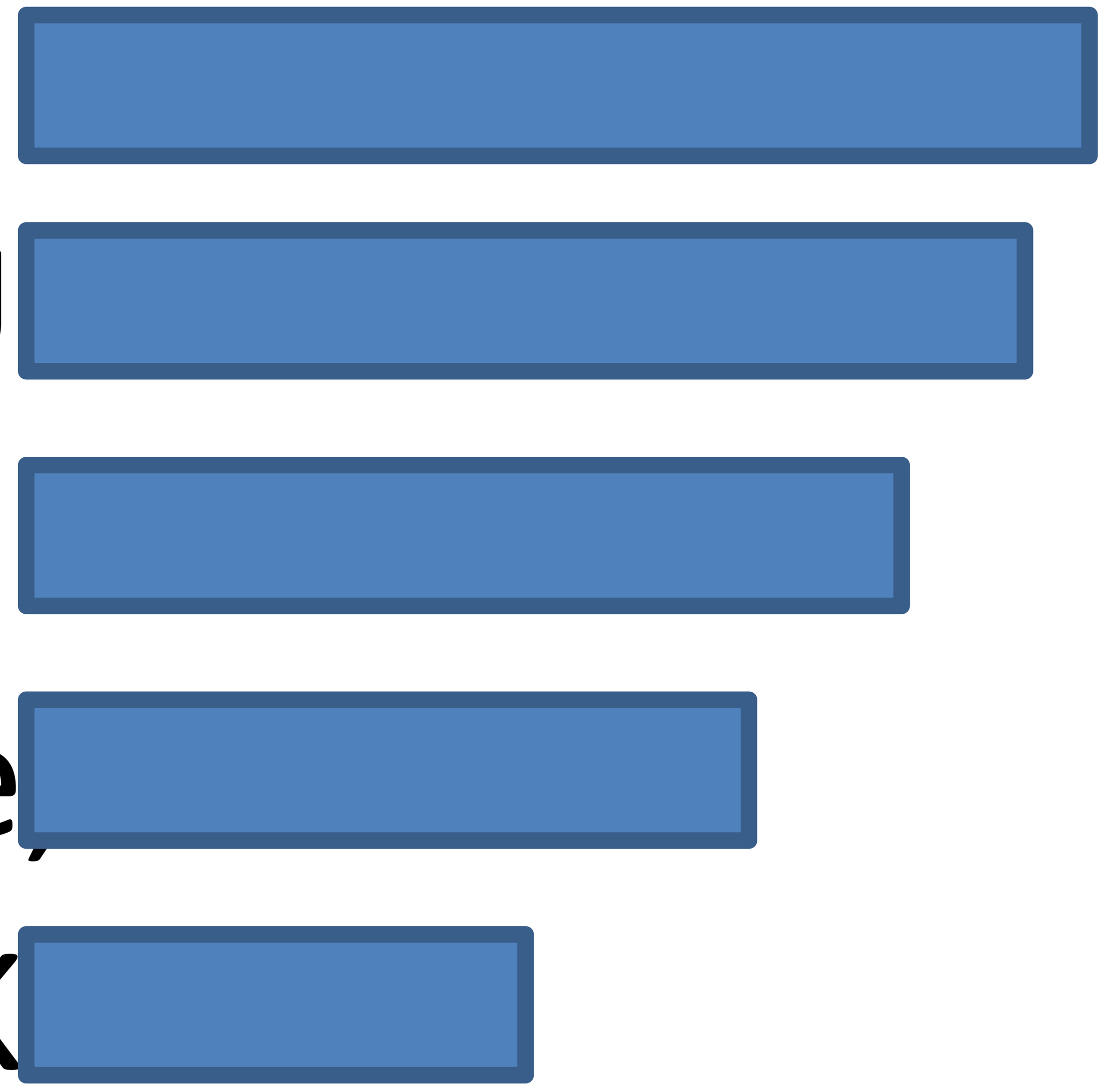
caleydo.org/tools/upset

[InfoVis '13]
Best Paper Award

Lineup: Visualizing Multivariate Rankings

Rankings are Popular

Rank	University	Score
1.	MIT, USA	
2.	Harvard, U	
3.	Princeton,	
4.	Cambridge,	
5.	Oxford, UK	



Support Multiple Attributes

Combiner functions: f(A,B,C)

(Weighted) sum

$Score = w_a A + w_b B + w_c C$

→Serial

Maximum

$Score = max(A, B, C)$

→Parallel

Product

Nesting

…

→Complex Combiners

Serial Combiner (as Stacked Bar)

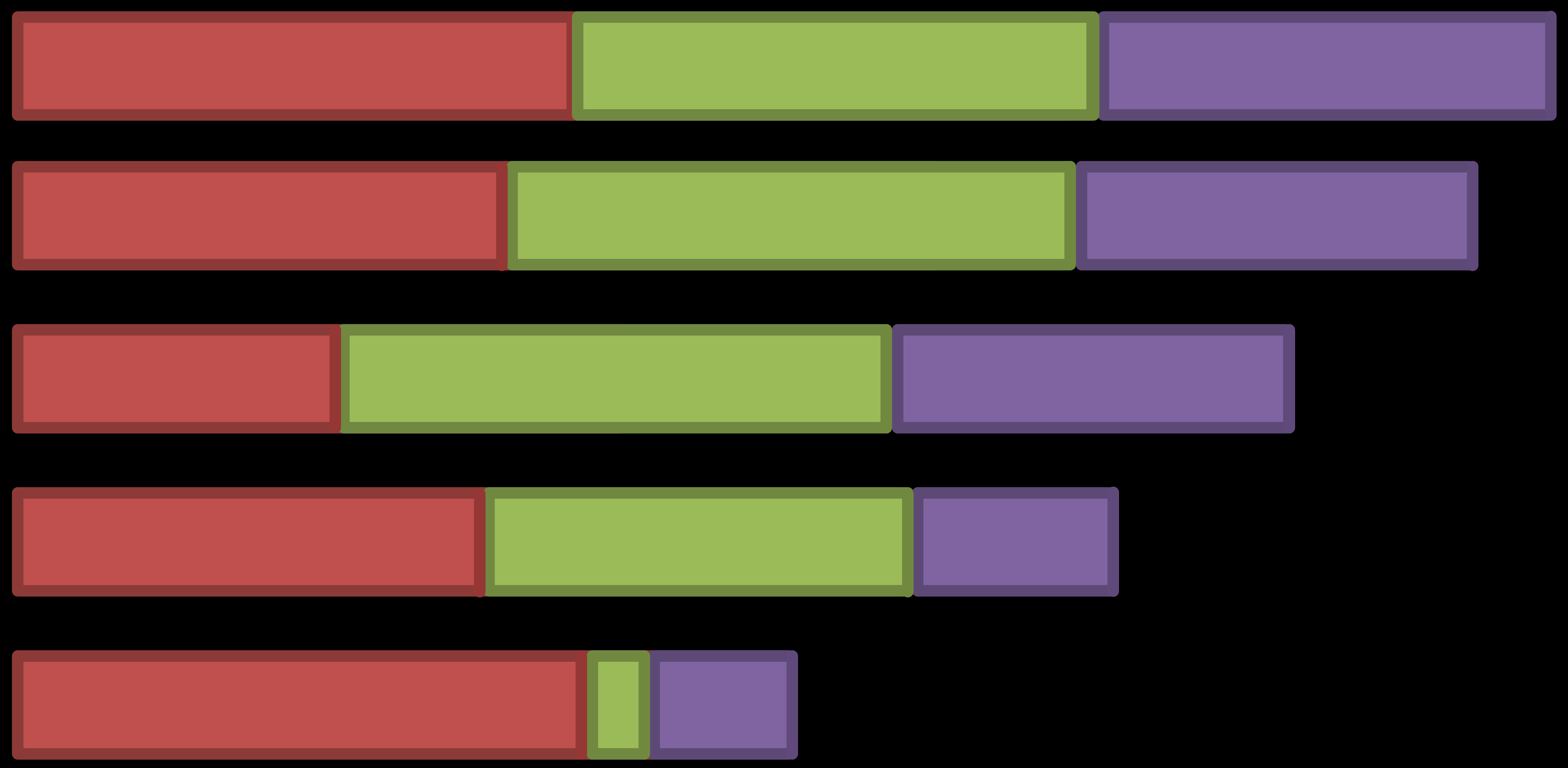
$$w_a A \quad + \quad w_b B \quad + \quad w_c C$$

Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, USA			
4.	Cambridge, UK			
5.	Oxford, UK			

Serial Combiner (as Stacked Bar)

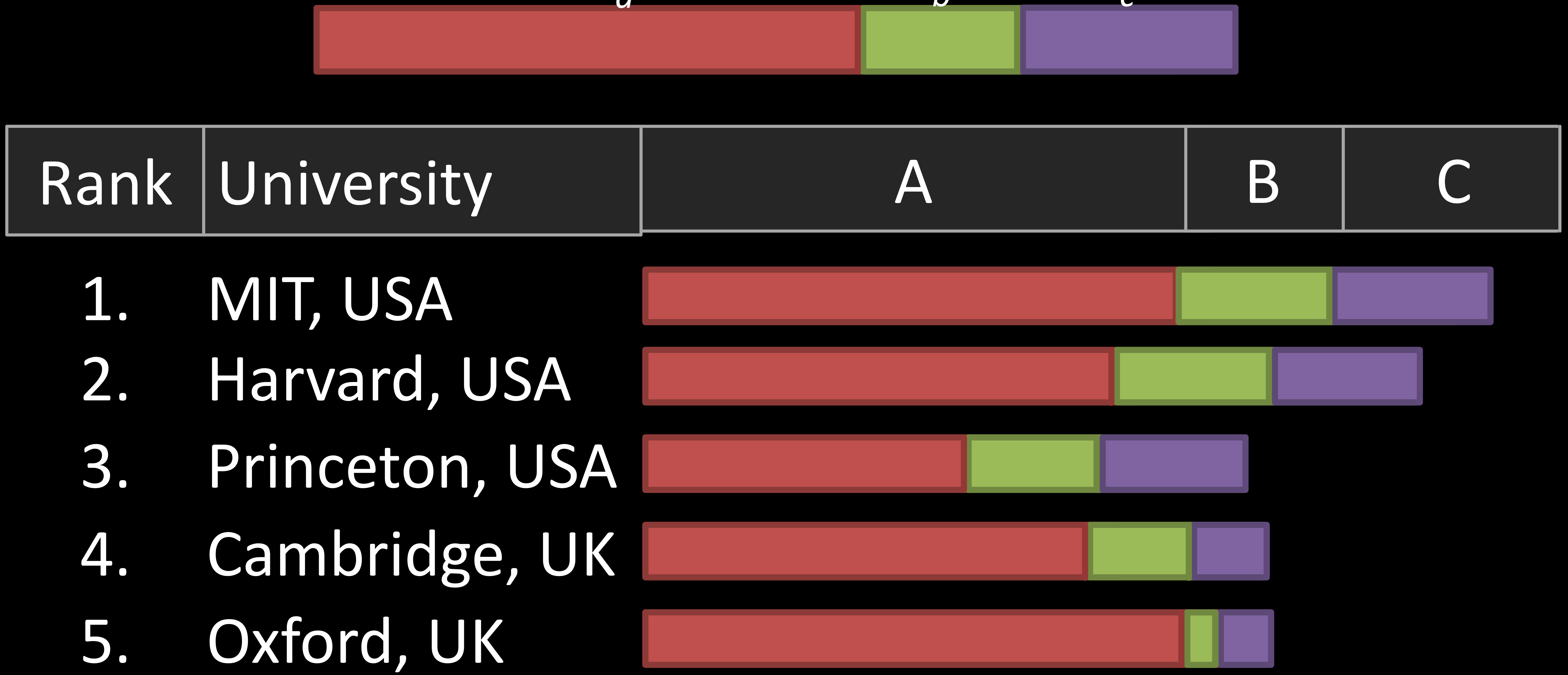
$$w_a A \quad + \quad w_b B \quad + \quad w_c C$$

Rank	University	A	B	C
1.	MIT, USA			
2.	Harvard, USA			
3.	Princeton, USA			
4.	Cambridge, UK			
5.	Oxford, UK			



Serial Combiner (as Stacked Bar)

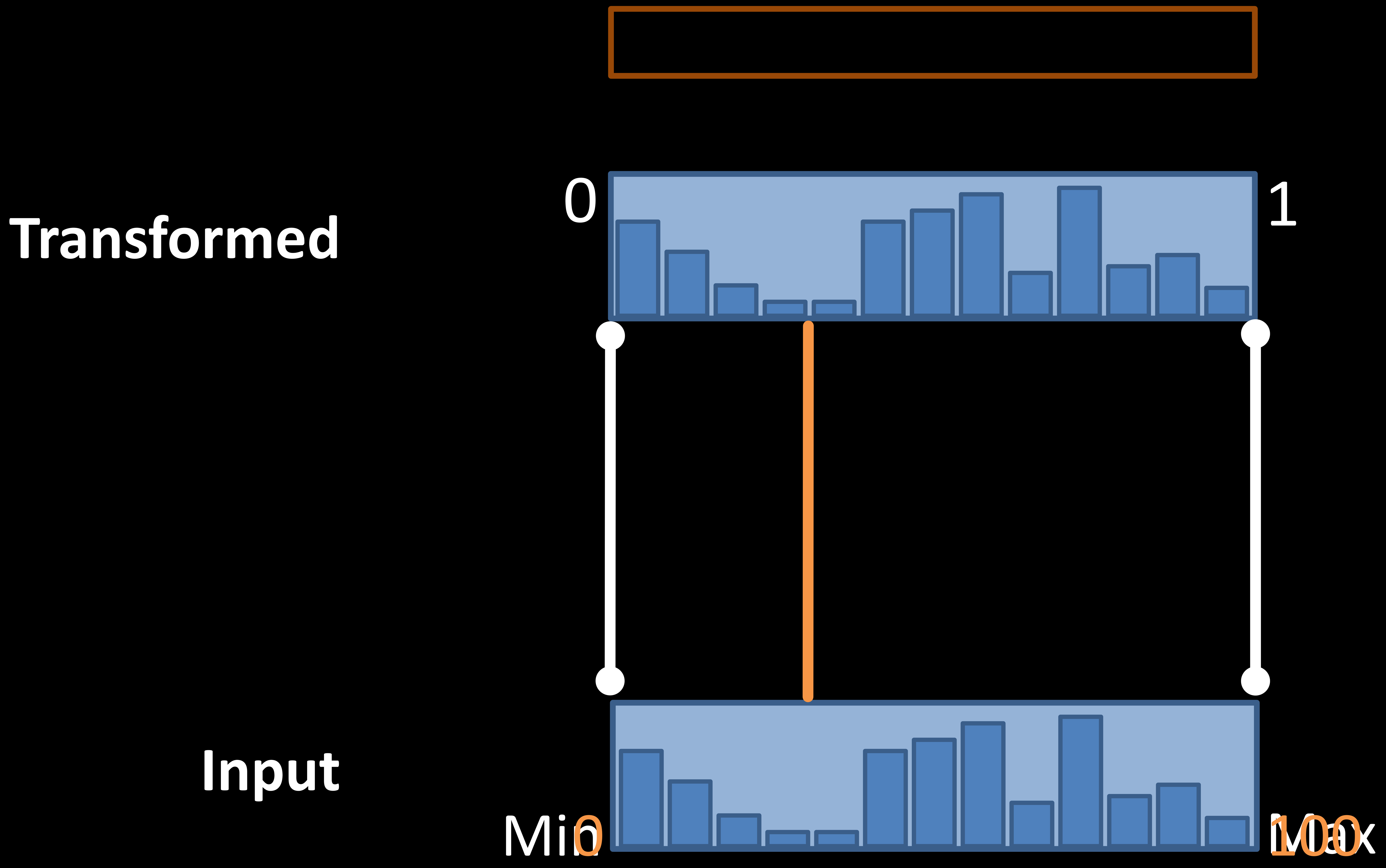
$$w_a A + w_b B + w_c C$$

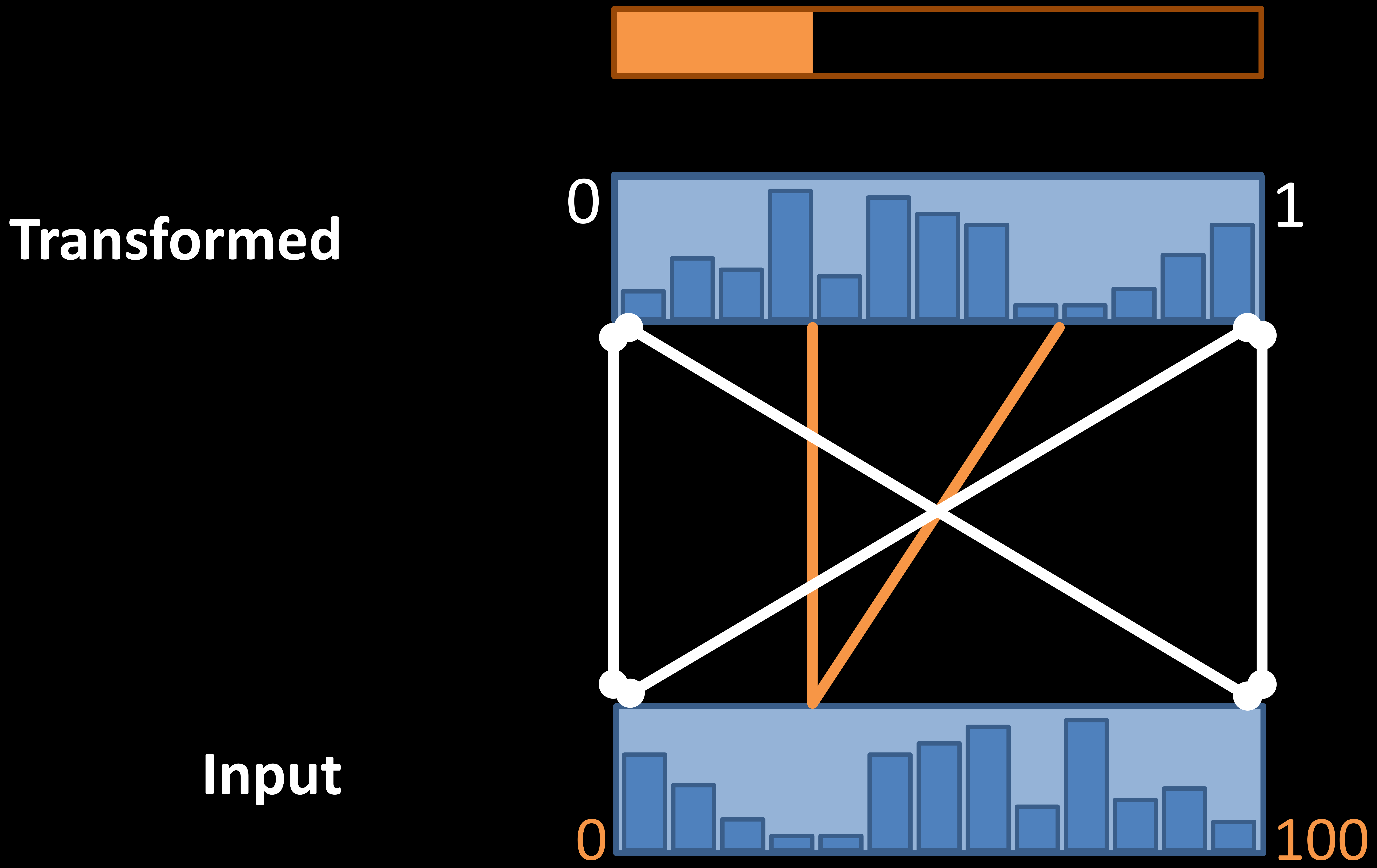


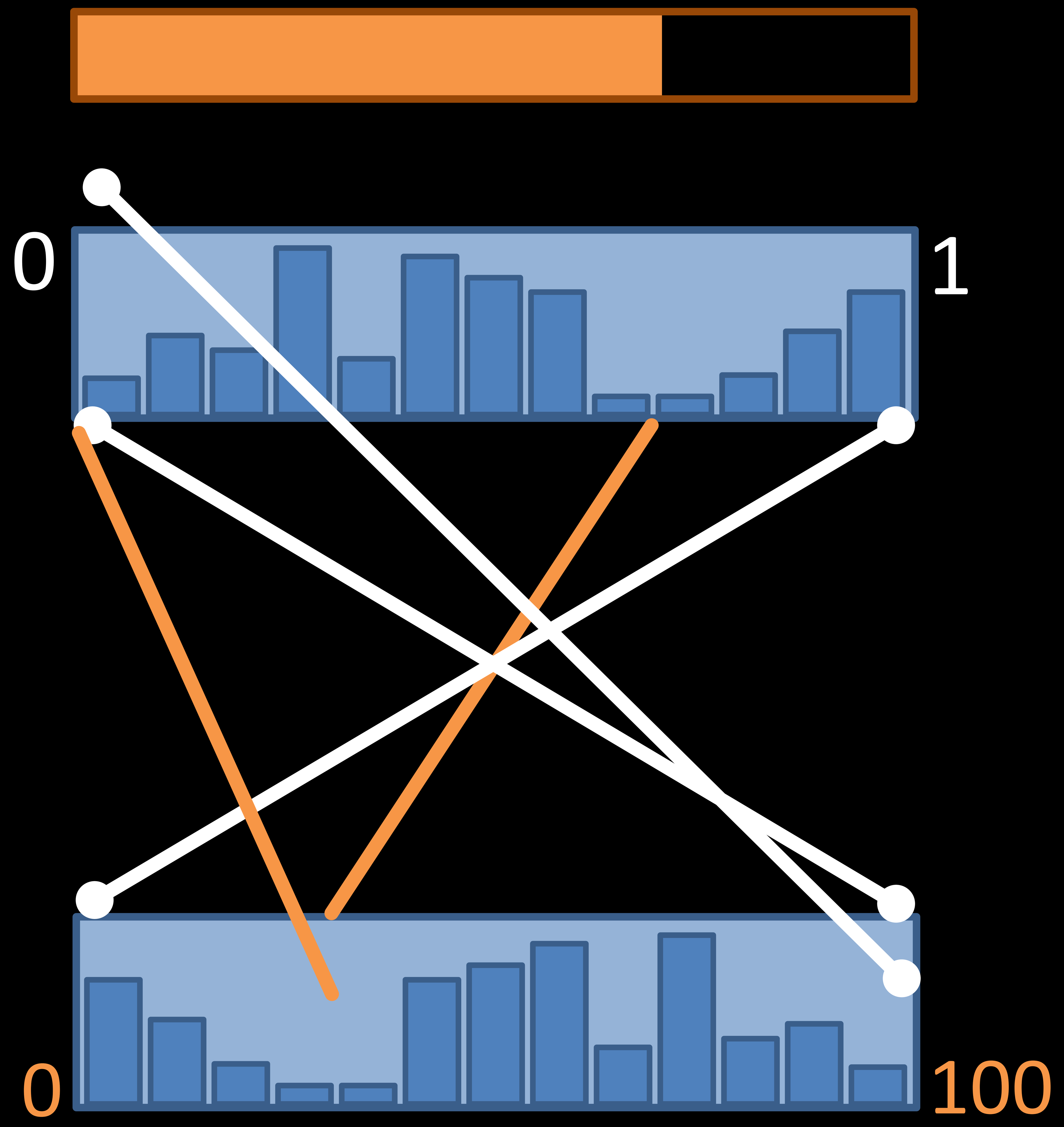
128 visible of 906 (14.13%)

Ran	School Name	Country	Faculty/student ratio	Employer reputation	Citations per faculty
	Filter: <None>	Filter: 2 out of 72			
1.	American University	United States			
2.	Arizona State University	United States			
3.	Aston University	United Kingdom			
4.	Birkbeck College, University of L	United Kingdom			
5.	Boston College	United States			
6.	Boston University	United States			
7.	Brandeis University	United States			
8.	Brown University	United States			
9.	Brunel University	United Kingdom			
10.	California Institute of Technology	United States			
11.	Cardiff University	United Kingdom			
12.	Case Western Reserve University	United States			
13.	City University London	United Kingdom			
14.	College of William & Mary	United States			
15.	Colorado State University	United States			
16.	Columbia University	United States			
17.	Cornell University	United States			
18.	Cranfield University	United Kingdom			
19.	Dartmouth College	United States			
20.	Drexel University	United States			
21.	Duke University	United States			
22.	Durham University	United Kingdom			

Flexible Mapping of Attributes to Scores


Transformed
0
1
Input
Min
0
Max
100

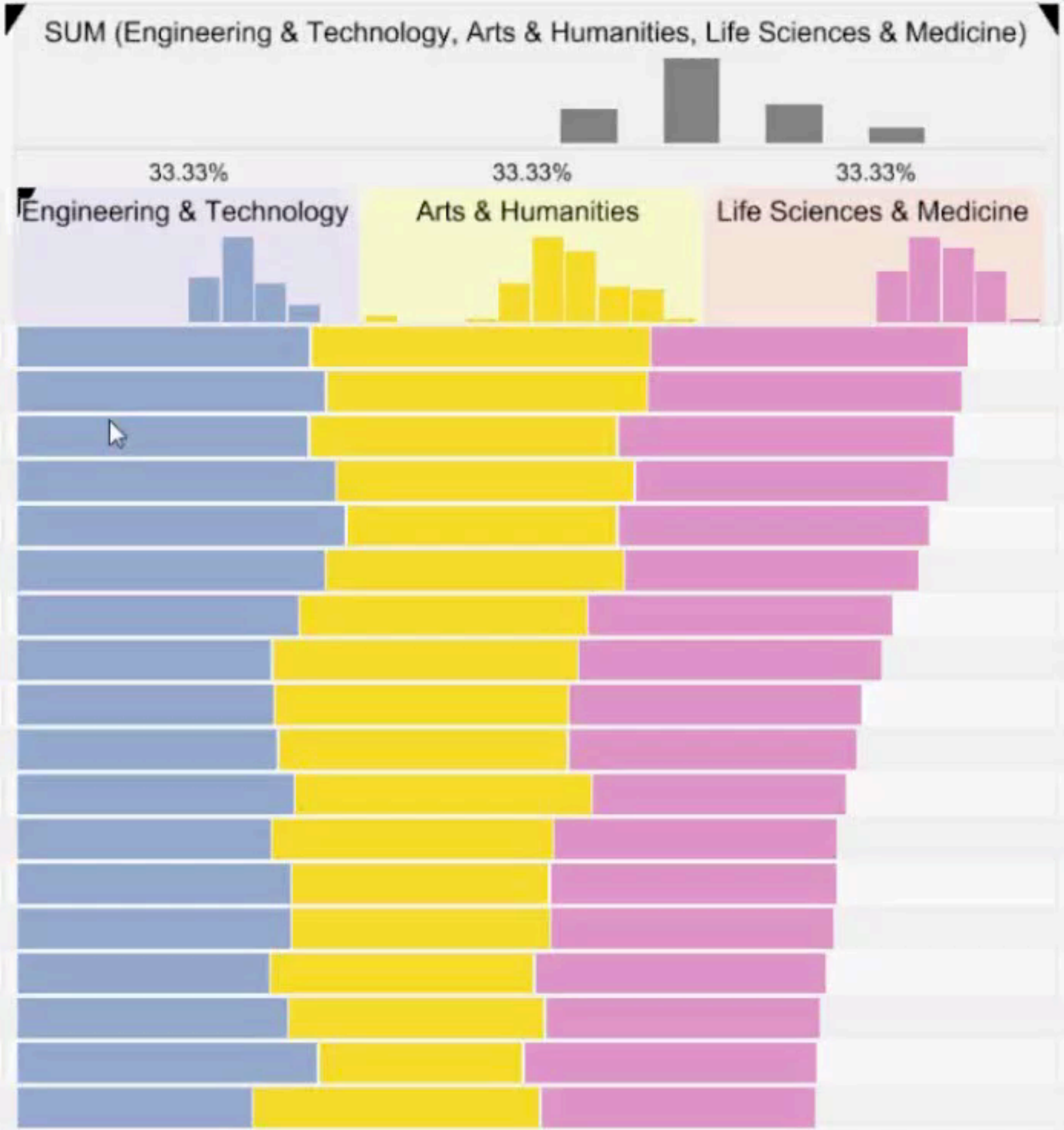

Transformed
0
1
Input
0
100



0
1
0
100

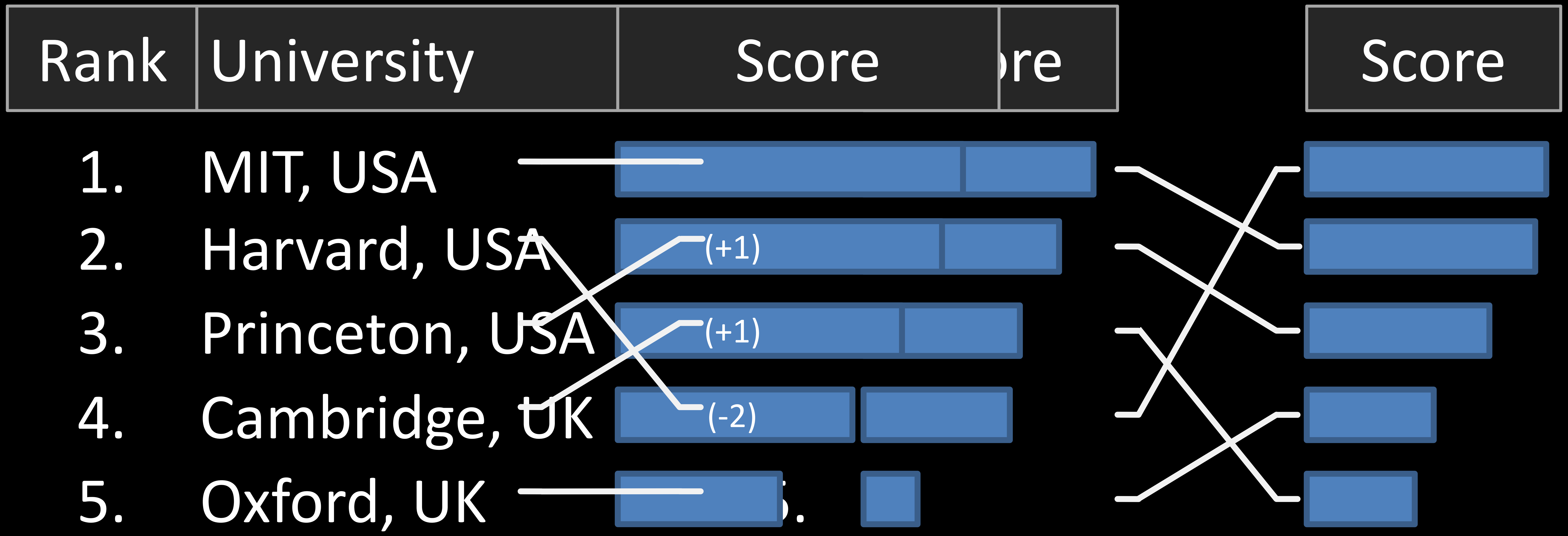
93 visible of 399 (23.31%)

Ran	School Name	Country
	Filter: <None>	Filter: 2 out of 43
1.	University of Oxford	United Kingdom
2.	University of Cambridge	United Kingdom
3.	Harvard University	United States
4.	Stanford University	United States
5.	Massachusetts Institute of Technology (MIT	United States
6.	University of California, Berkeley (UCB)	United States
7.	University of California, Los Angeles (UCL	United States
8.	Yale University	United States
9.	UCL (University College London)	United Kingdom
10.	Columbia University	United States
11.	Princeton University	United States
12.	University of Edinburgh	United Kingdom
13.	University of Michigan	United States
14.	Cornell University	United States
15.	University of Pennsylvania	United States
16.	The University of Manchester	United Kingdom
17.	Imperial College London	United Kingdom
18.	University of Chicago	United States



Compare Rankings

Bump Charts


Rank
University
Score
Score
1. MIT, USA
2. Harvard, USA
3. Princeton, USA
4. Cambridge, UK
5. Oxford, UK
(+1)
(+1)
(-2)

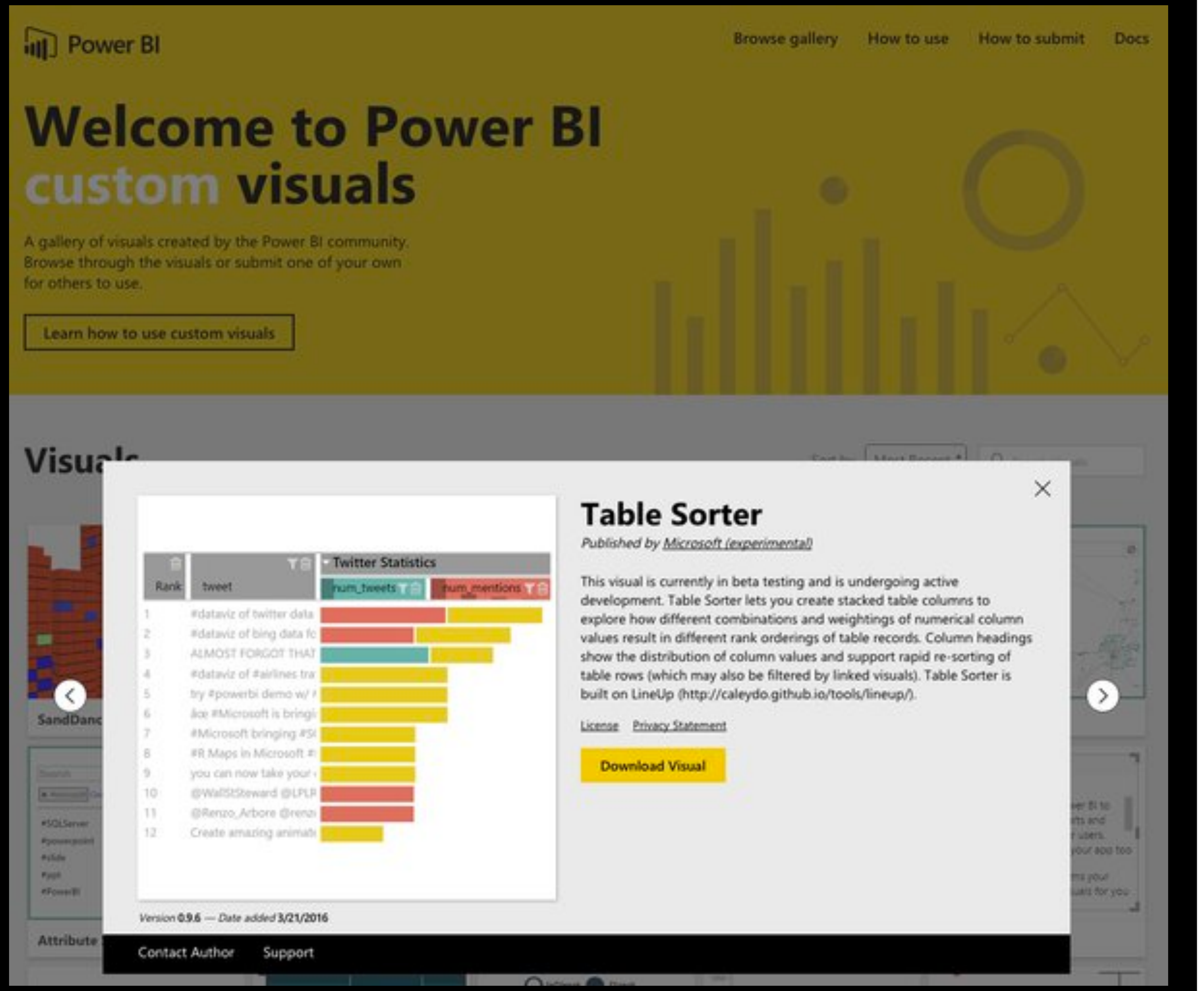
Bump Charts


Rank
University
Score
1. MIT, USA
2. Harvard, USA
3. Princeton, USA
4. Cambridge, UK
5. Oxford, UK
Rank
Score
1.
(+1)2.
(+1)3.
(-2)4.
5.
Score


453 visible of 453 (100.00%)
World University Ranking
46.55%
32.72%
20.72%
Rank
School Name
Filter:
<None>
MAX (Academic re
Faculty/stud
Citatio
1. Massachusetts Institute of Te
2. California Institute of Technol
3. Harvard University
4. University of Cambridge
5. UCL (University College Lond
6. University of Oxford
7. Princeton University
8. Imperial College London
9. University of Chicago
10. Stanford University
11. Columbia University
12. Duke University
13. University of Pennsylvania
14. Johns Hopkins University
15. Yale University
16. University of Michigan
17. Ecole normale supérieure, Pa
18. Northwestern University

Applications

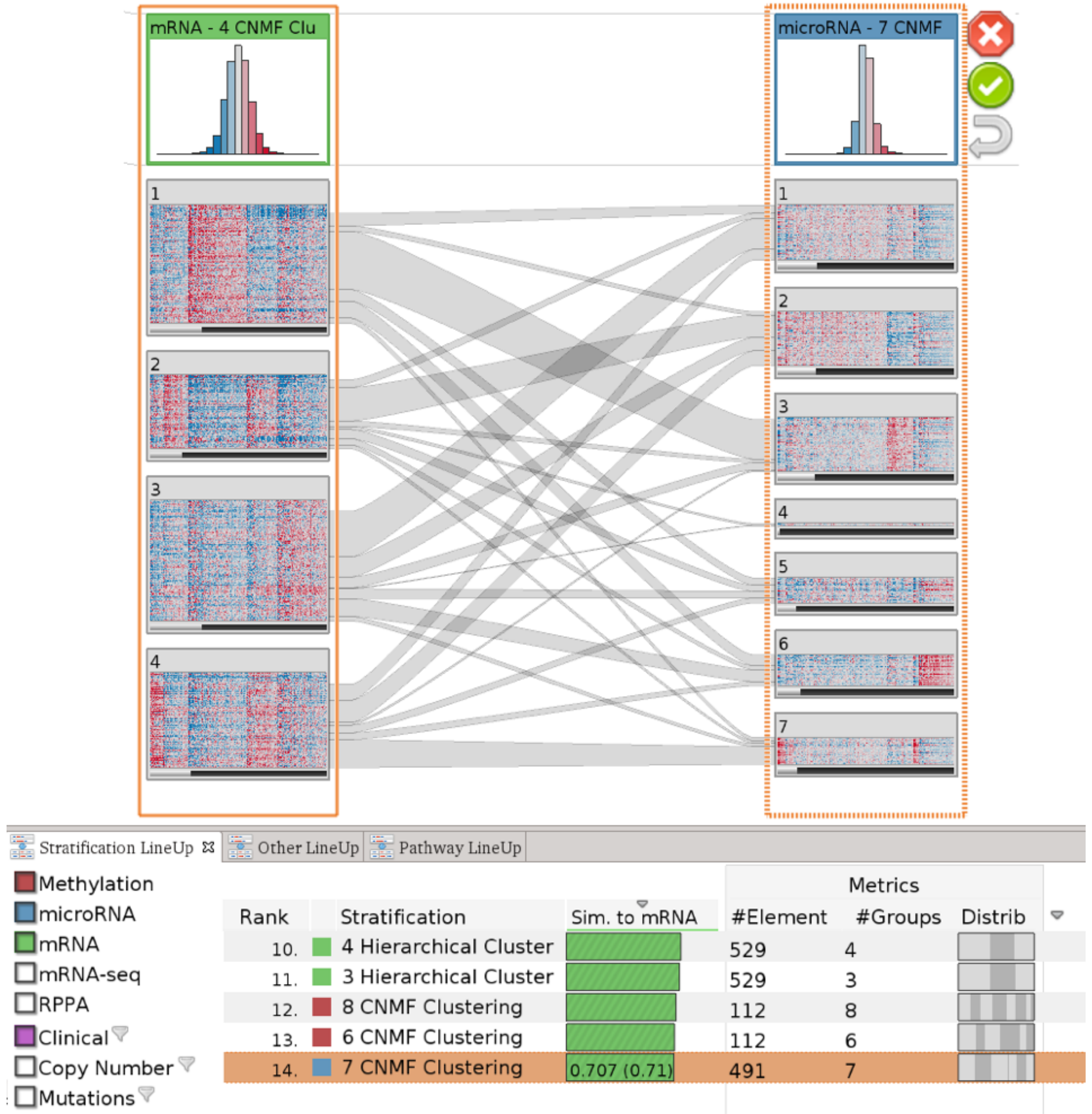
Applications



General Purpose Web Tool

Integrated by Microsoft into PowerBI

Part of Genomics Data Analysis Platform



http://lineup.caleydo.org

[Partl, BioVis '12]
Best Paper Award
[Partl, BMC Bioinf. '13]
[Lex, InfoVis '13]

Pathways

Experi-mental Data and Pathways

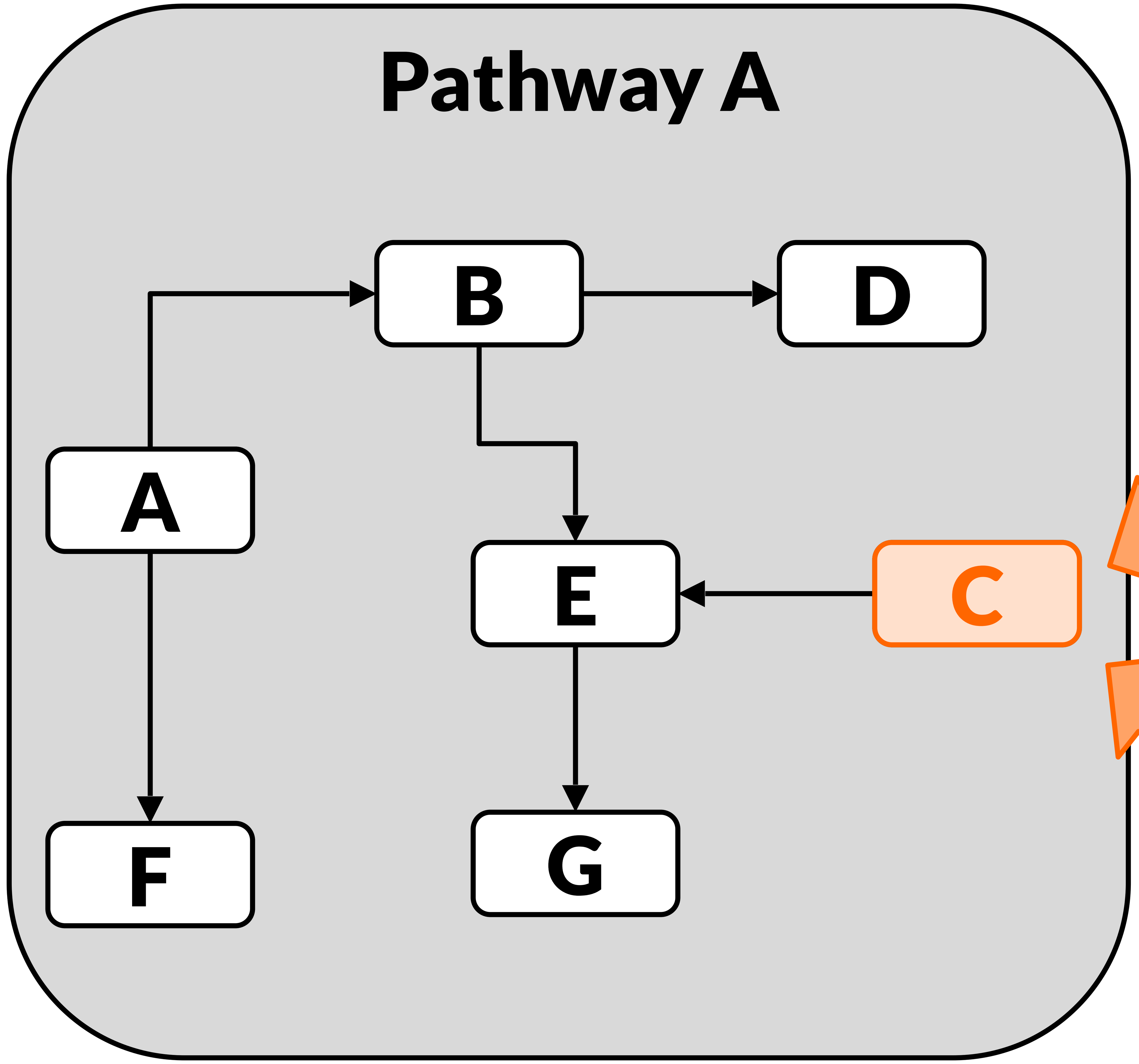
[Partl, BioVis '12]

Cannot account for variation found in real-world data

Branches can be (in)activated due to

- mutation,
- changed gene expression,
- modulation due to drug treatment,
- etc.

Many Node Attributes

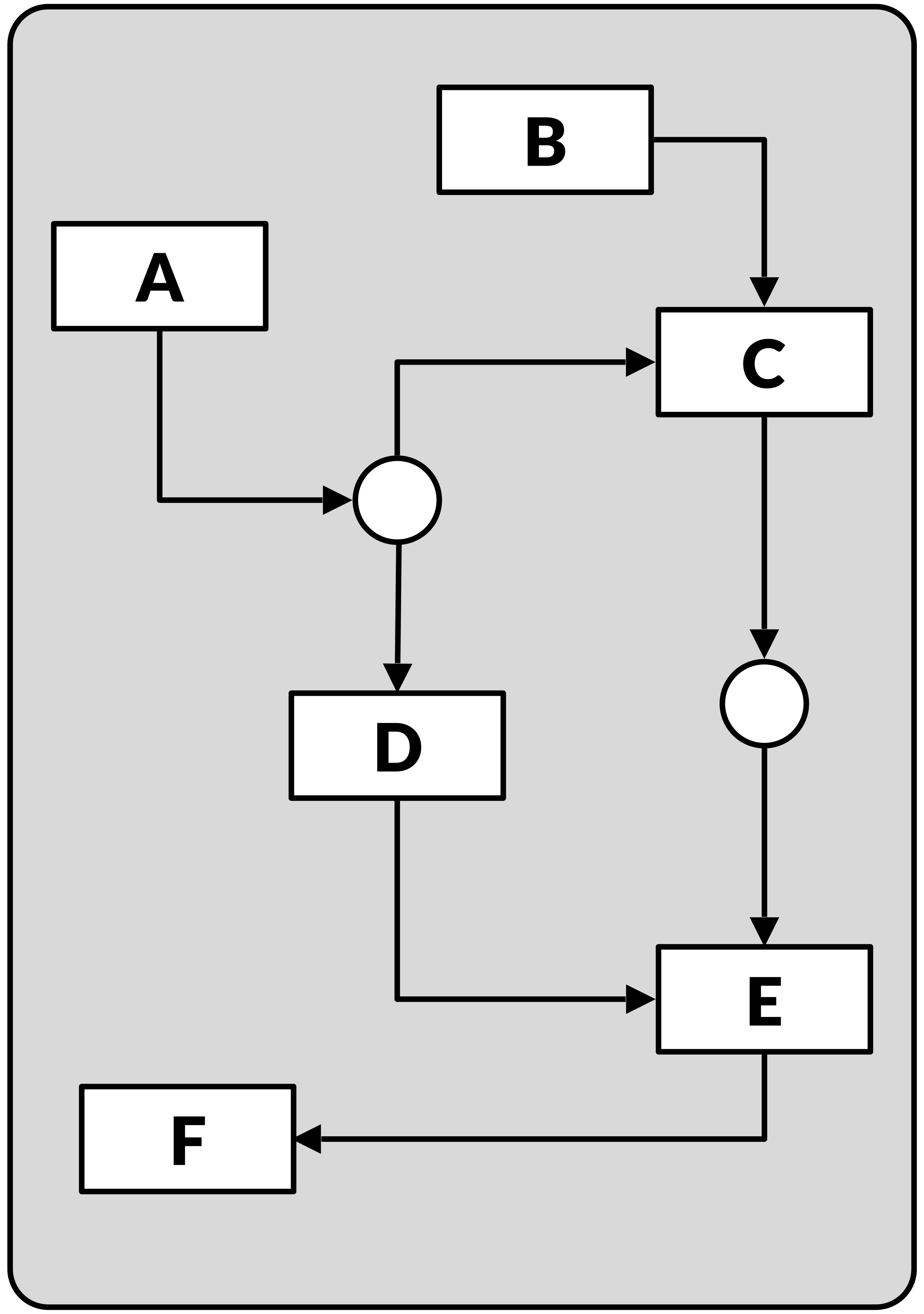


Node	Sample 1	Sample 2	Sample 3	...
A	0.55	0.95	0.83	...
B	0.12	0.42	0.16	...
C	0.33	0.65	0.38	...
...	...	...	...	

Node	Sample 1	Sample 2	Sample 3	...
A	low	low	very high	...
B	normal	low	high	...
C	high	very low	normal	...
...	...	...	...	

How to visualize experimental data on pathways?

Good Old Color Coding



A	-3.4	4.2	5.1	4.2
B	2.8	1.8	1.3	1.1
C	3.1	-2.2	2.4	2.2
D	-3	-2.8	1.6	1.0
E	0.5	0.3	-1.1	1.3
F	0.3	0.3	1.8	-0.3

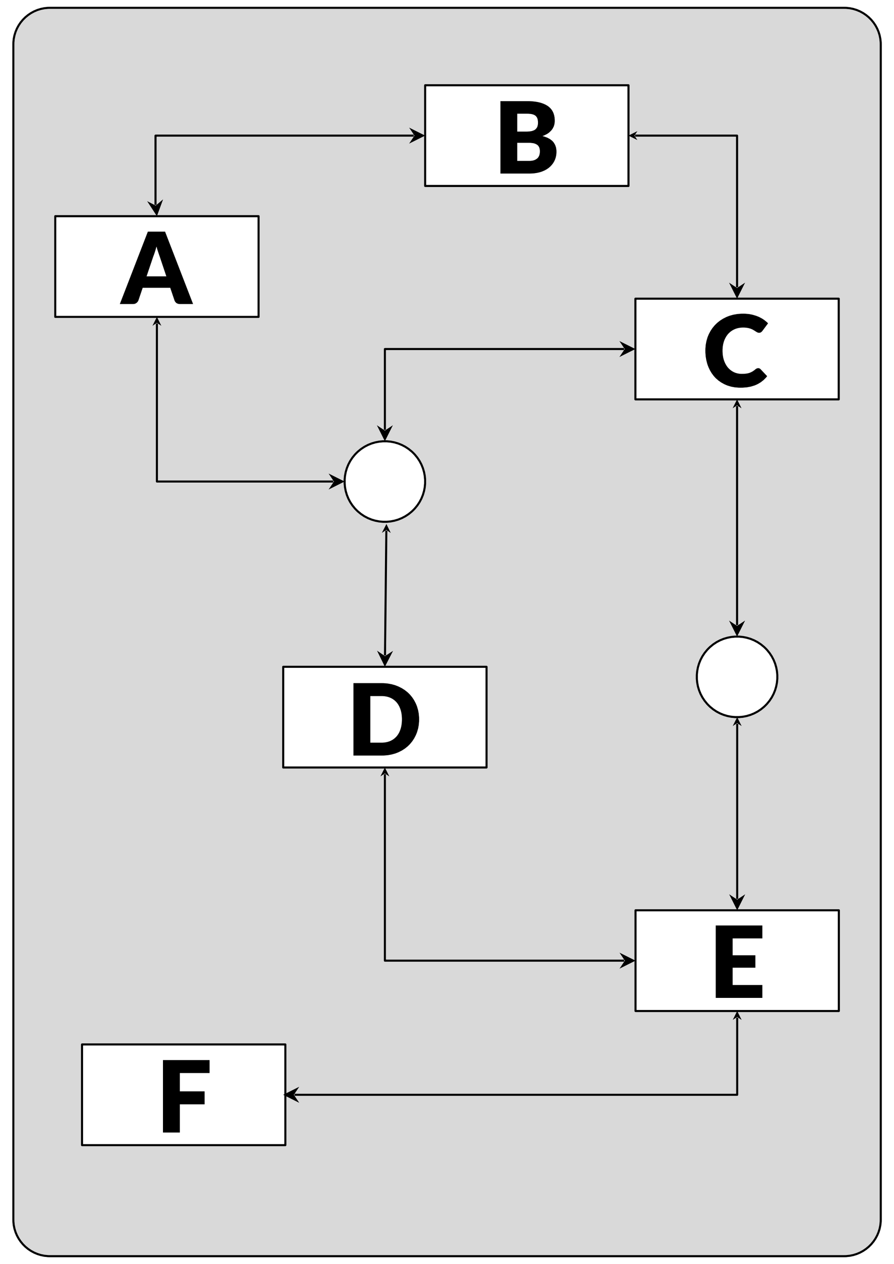
[Lindroos2002]

Challenge: Supporting Multiple Tasks

Two central tasks:

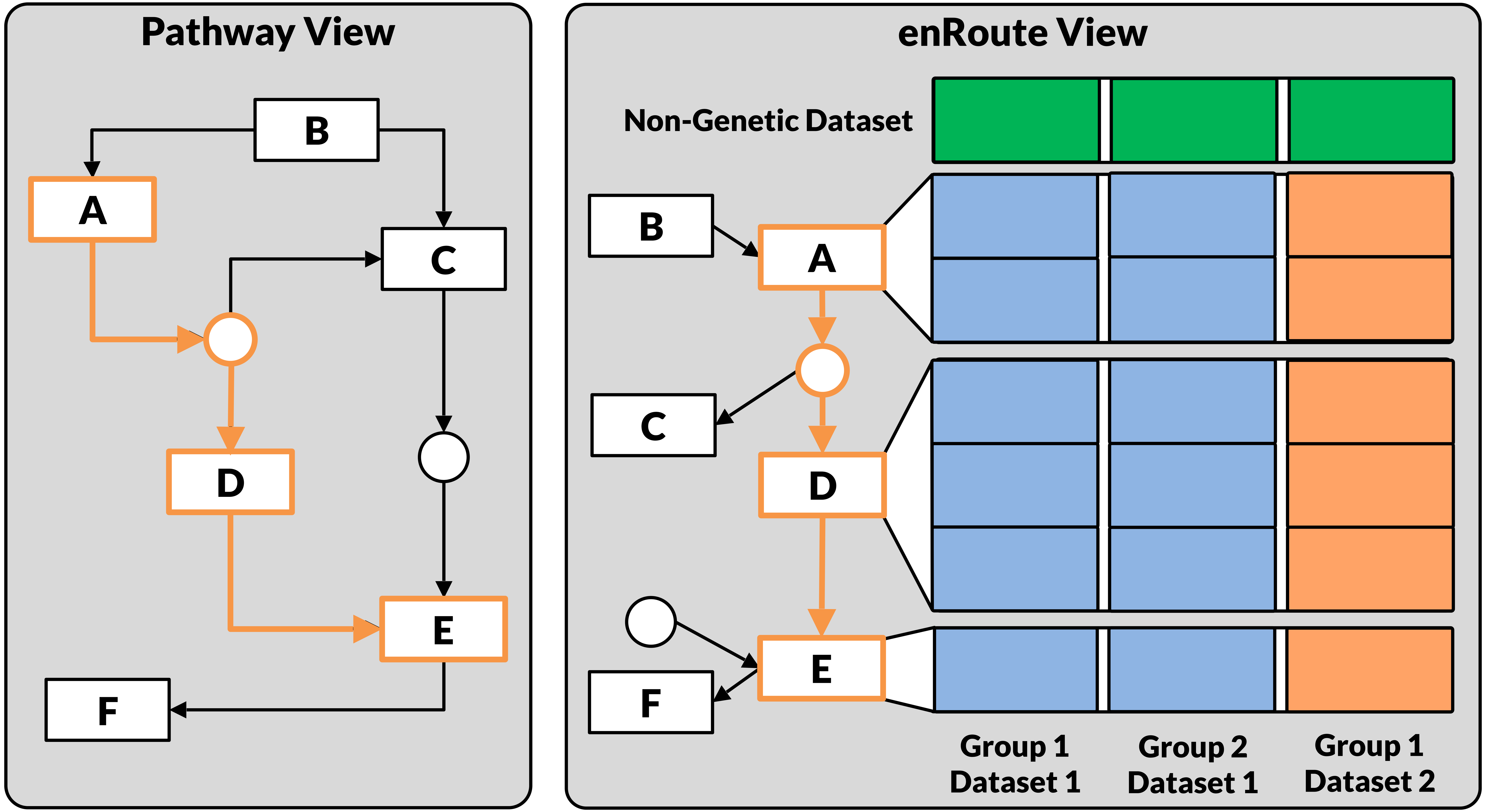
- Explore topology of pathway
- Explore the attributes of the nodes (experimental data)

Need to support both!



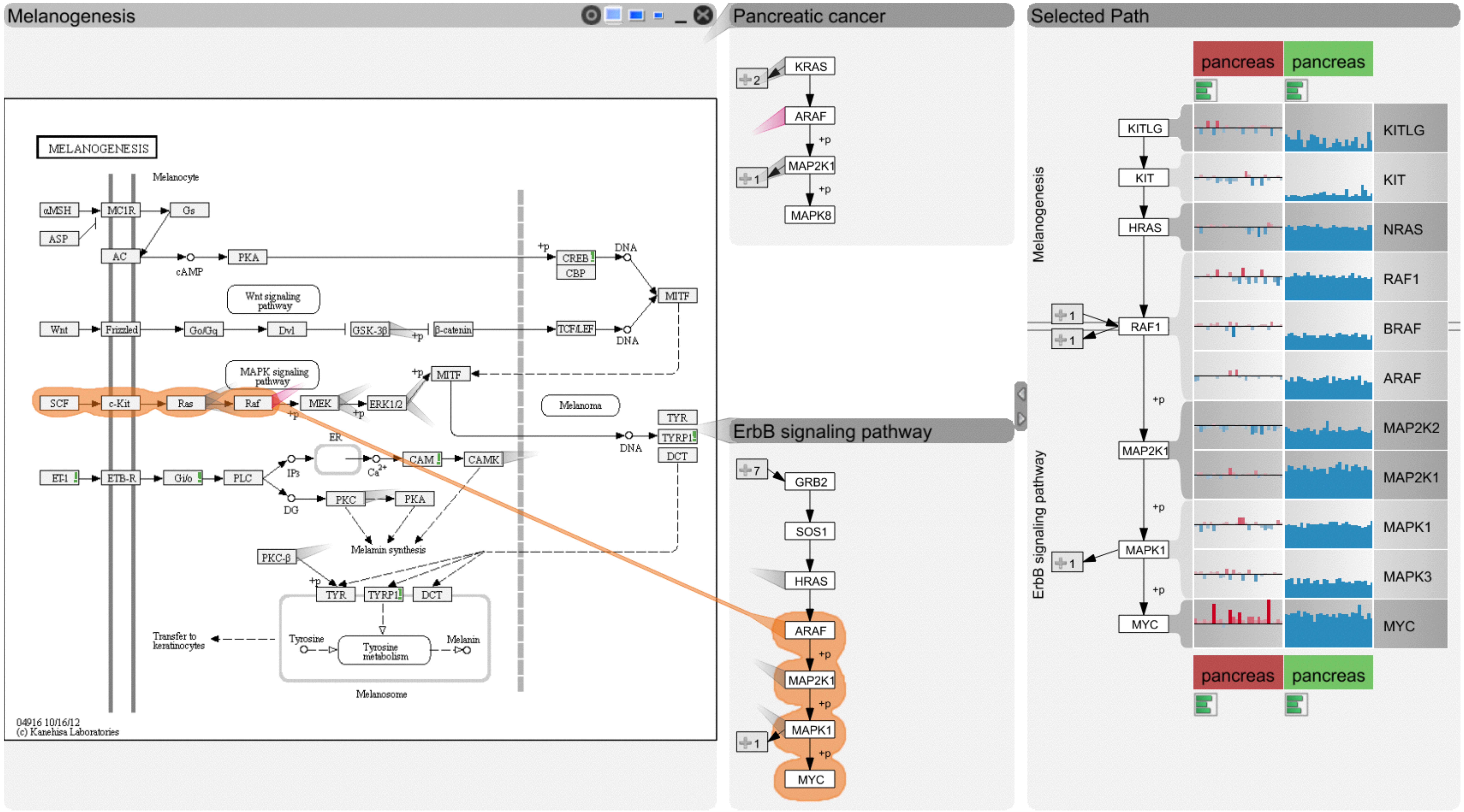
	Sample 1	Sample 2	Sample 3
Gene 1	1	1.1	0.4
Gene 2	2	0.5	1.2
Gene 3	1.4	0.2	0.5
Gene 4	0.3	0.5	0.7

Concept



Pathway View
B
A
C
D
E
F
enRoute View
Non-Genetic Dataset
B
A
C
D
E
F
Group 1
Dataset 1
Group 2
Dataset 1
Group 1
Dataset 2

enRoute



Melanogenesis
Pancreatic cancer
Selected Path
ErbB signaling pathway
MELANOGENESIS
Melanocyte
KRAS
ARAF
MAP2K1
MAPK8
GRB2
SOS1
HRAS
MAPK1
MYC
pancreas
KITLG
KIT
NRAS
RAF1
BRAF
MAP2K2
MAPK3
ErbB signaling pathway
04916 10/16/12
(c) Kanehisa Laboratories

File Data Window View Help

Entourage

Pathways

Pathway

Filter:
<None>

1 C donor
2-Oxocarboxylic acid
ABC transporters
ABC-family proteins
ACE Inhibitor Pathwa
Acetylcholine Synthes
Acute myeloid leukem
Adherens junction
Adipocyte TarBase
Adipocytokine signali
Adipogenesis
Advanced glycosylatio
Aflatoxin B1 metaboli
African trypanosomias
AGE/RAGE pathway
AhR pathway
Alanine and aspartate
Alanine, aspartate an
Alcoholism
Aldosterone-regulated
Allograft rejection
Allograft rejection
Alpha 6 Beta 4 signal
alpha-Linolenic acid
Alzheimer's disease
Alzheimers Disease
amino acid conjugatio
amino acid conjugatio
Amino sugar and nucl
Aminoacyl-tRNA bios
Amoebiasis
Amphetamine addicti
AMPK signaling
Amyotrophic lateral sc
Androgen receptor si
Angiogenesis
Angiogenesis
angiogenesis overvie
Antigen processing an
APC/C-mediated degra
Apoptosis
Apoptosis
Apoptosis Meta Path
Apoptosis Modulation
Apoptosis Modulation
Apoptosis, anoikis an

Selected Path


Toggle summary button
Summary of all dimensions
All dimensions shown separately
Not Mutated
Mutated
Not Mutated
Mutated
Race
Days to Deat
Gender
Context (non-genetic) datasets
EGF
EGFR
PLCG1
PLCG2
PRKCB1
PRKCA
PRKCG
NRAS
Dimensions sorted by this row
Multi-mapping of gene family
Collapsed incoming node
Expanded outgoing nodes
Glioma
ARAF
PIK3R5
Groups of same dataset
Different Datsets

Case Study: CCLE Data

ERBB2
GRB2
SOS1
HRAS
ARAF
MAP2K1
MAPK1
CCND1
PIK3R5
AKT3
FOXO3
Pancreatic cancer
Non-small cell lung cancer
Proliferation
Survival
Angiogenesis
Adhesion
Migration / Invasion
Differentiation
DNA
Jun
Elk
Myc
Glioma
CALML6
CAMK2A
ErbB signaling pathway
SHC2
Erlotinib
Lapatinib
SHC4
SHC3
SHC1
SOS2
NRAS
breast

http://entourage.caleydo.org

[EuroVis '12]
3rd Best Paper Award
[Nature Methods '14]

Heterogeneous Data
Cancer Subtypes

Cancer Subtypes

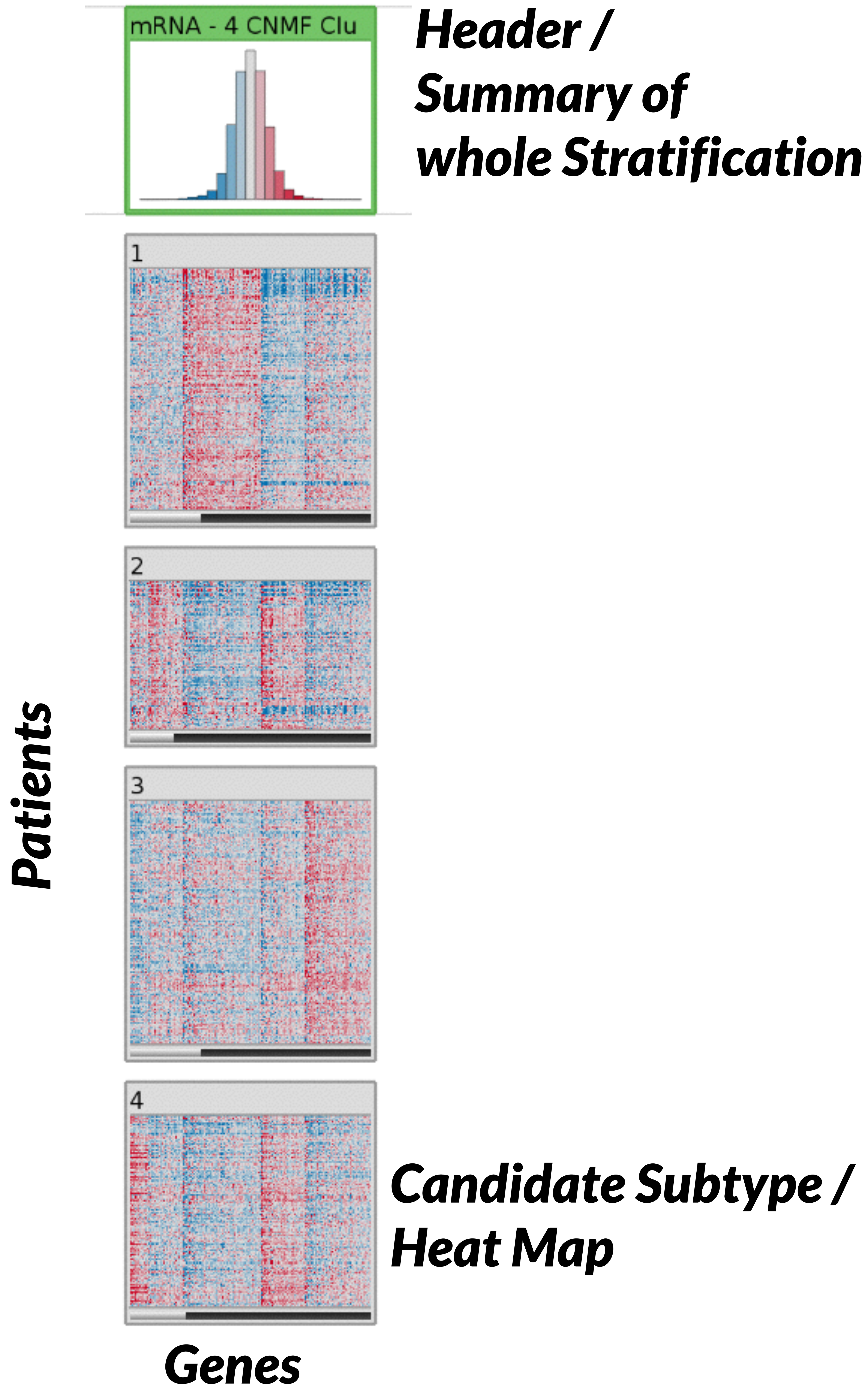
Cancer is not homogeneous

- different histology
- different molecular alterations

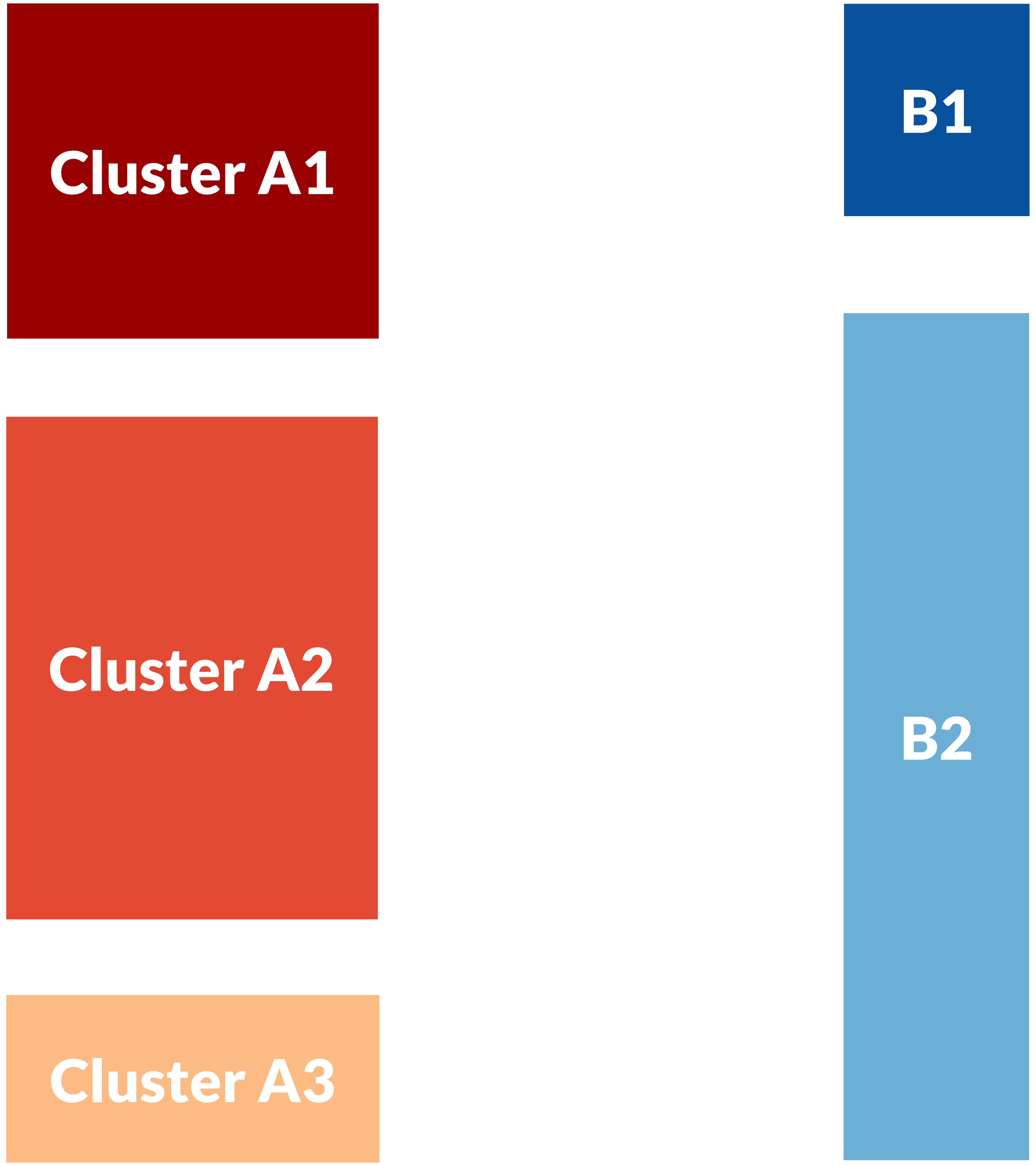
Subtypes have serious implications

- different treatment for subtypes
- prognosis varies between subtypes

Comparing & Evaluating Patient Groupings (Stratifications)


mRNA - 4 CNMF Clu
Header /
Summary of
whole Stratification
1
2
3
4
Patients
Candidate Subtype /
Heat Map
Genes

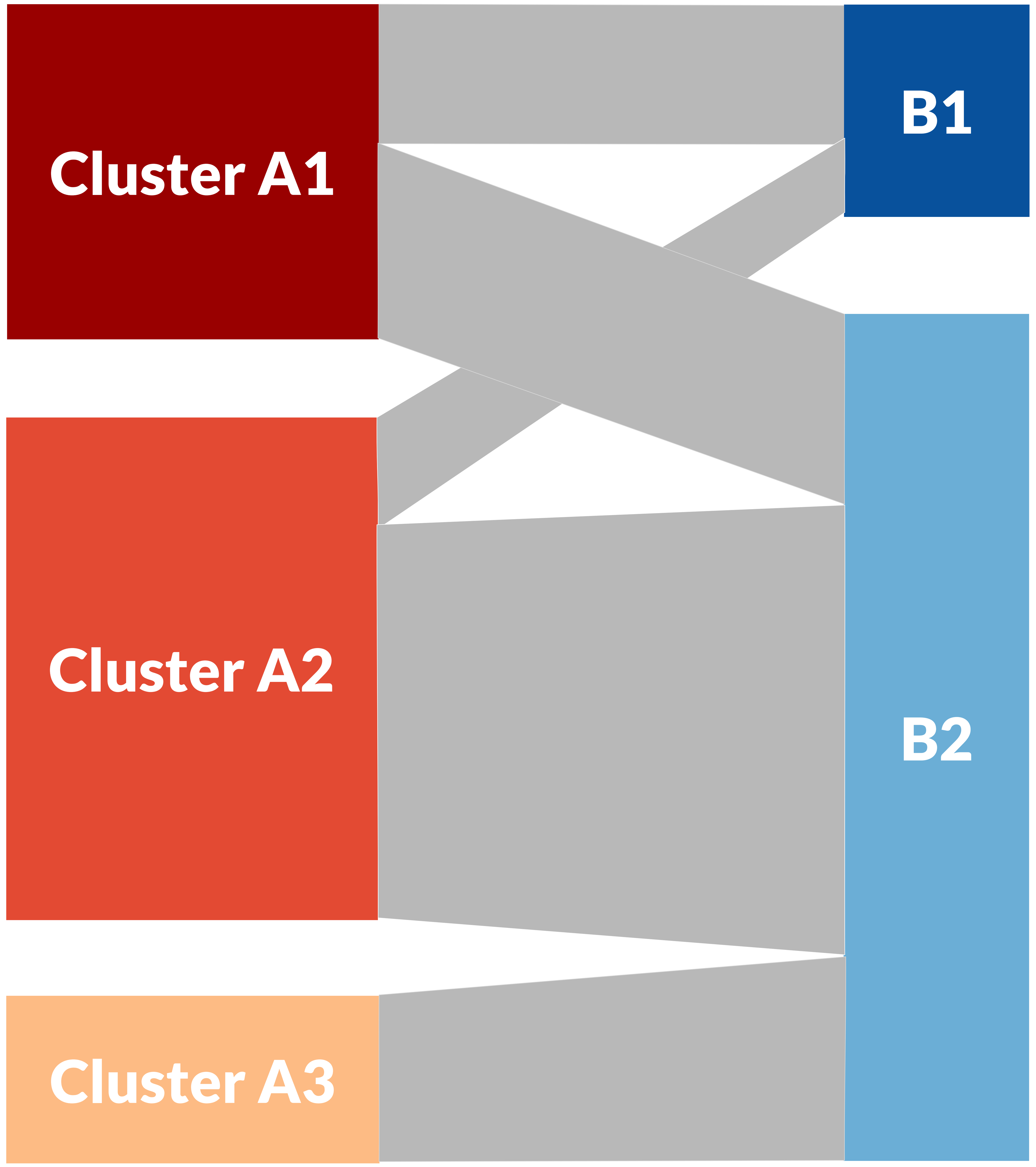
Comparing Stratifications



Tabular
e.g., mRNA

Categorical,
e.g., mutation status

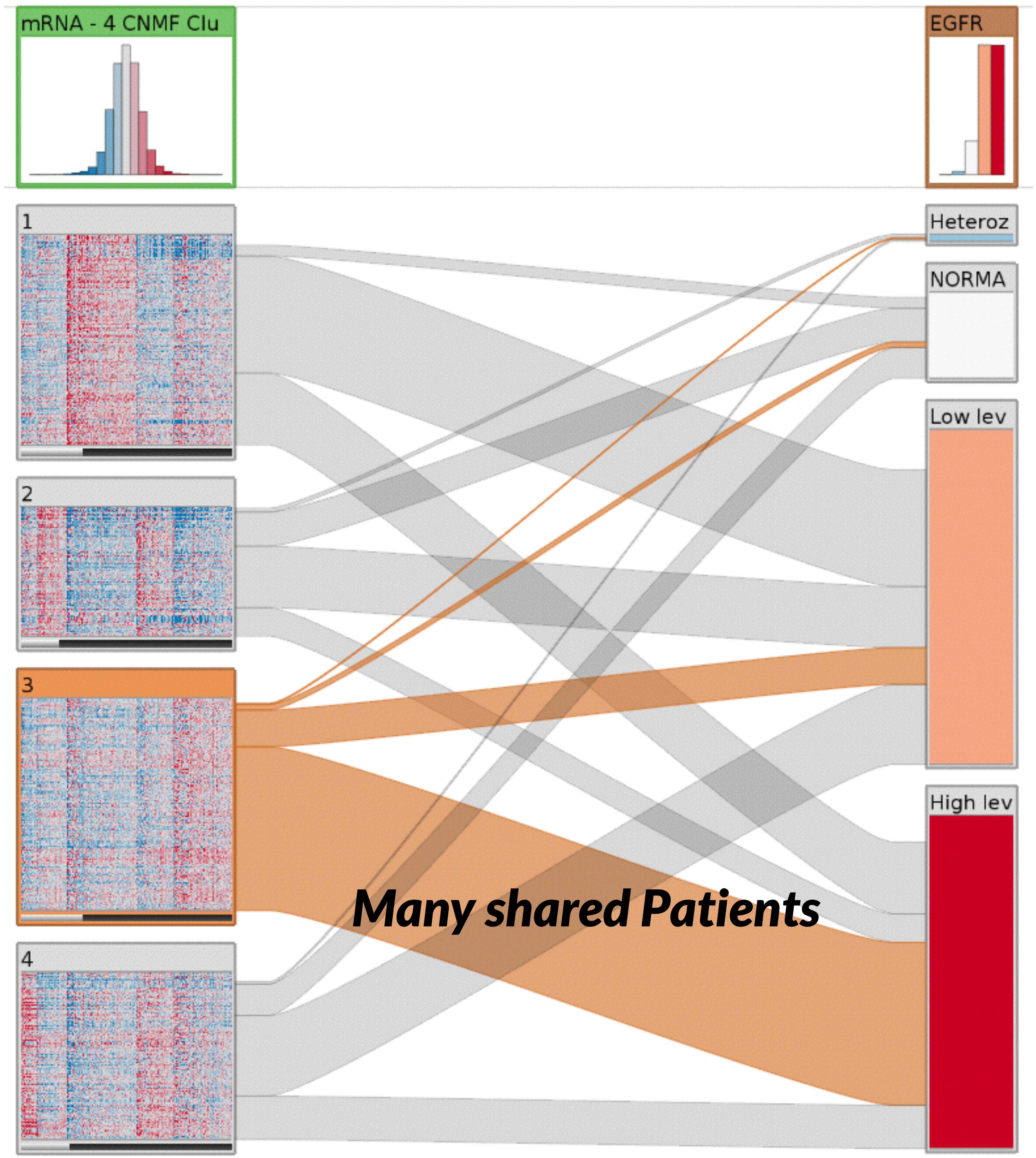
Comparing Stratifications



Tabular
e.g., mRNA

Categorical,
e.g., mutation status

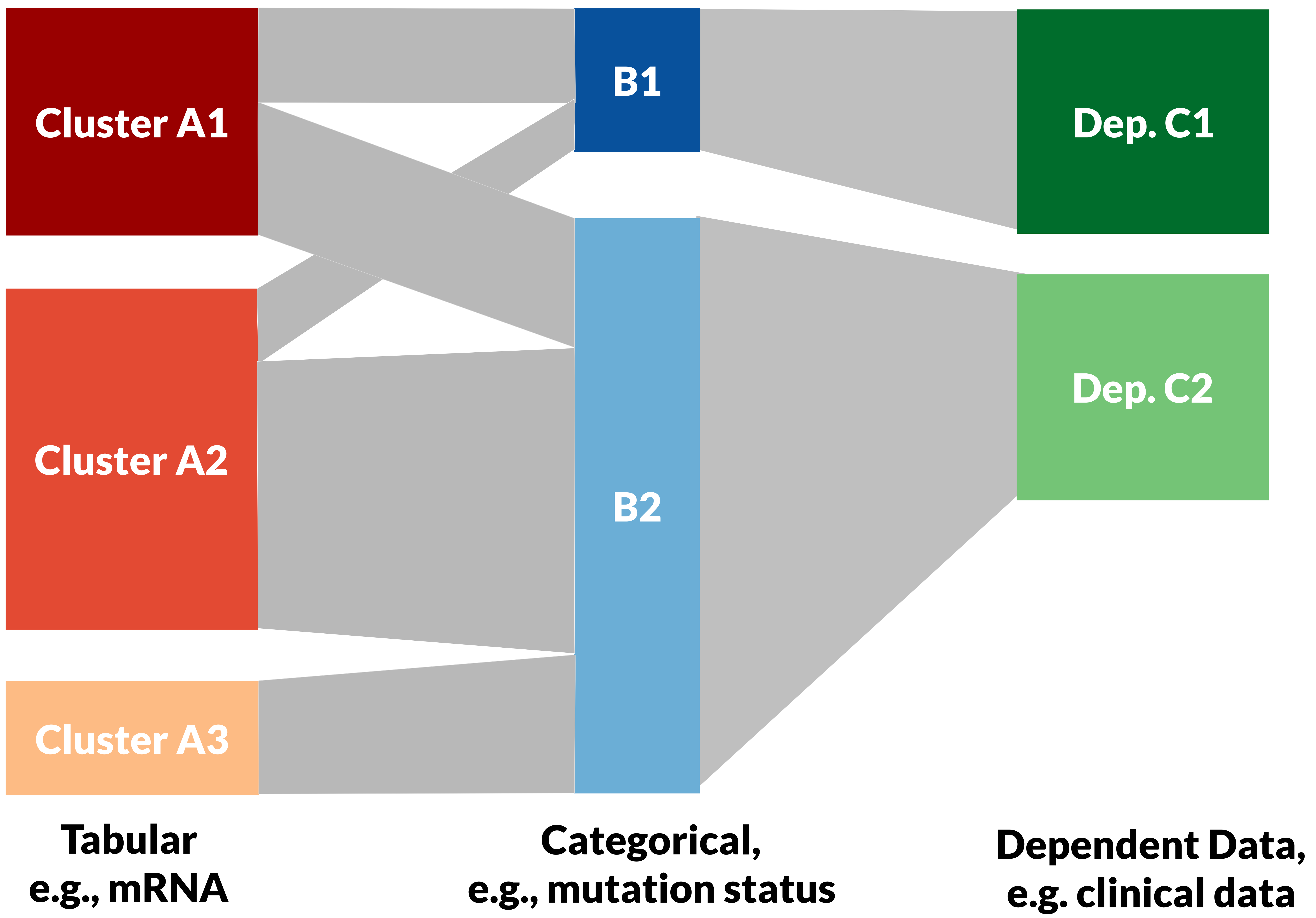
[Parallel Sets, Kosara]

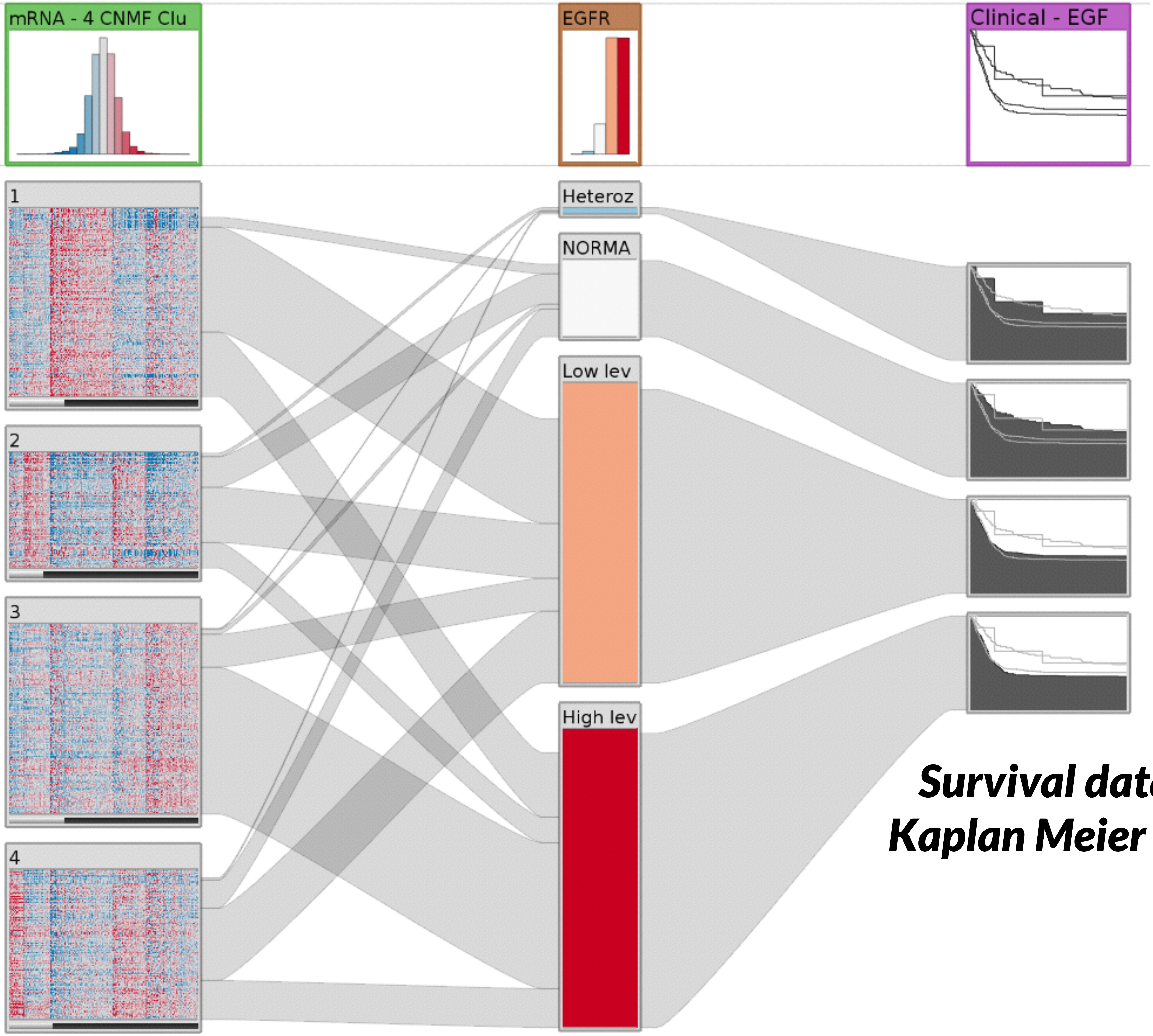


Clustering of mRNA Data

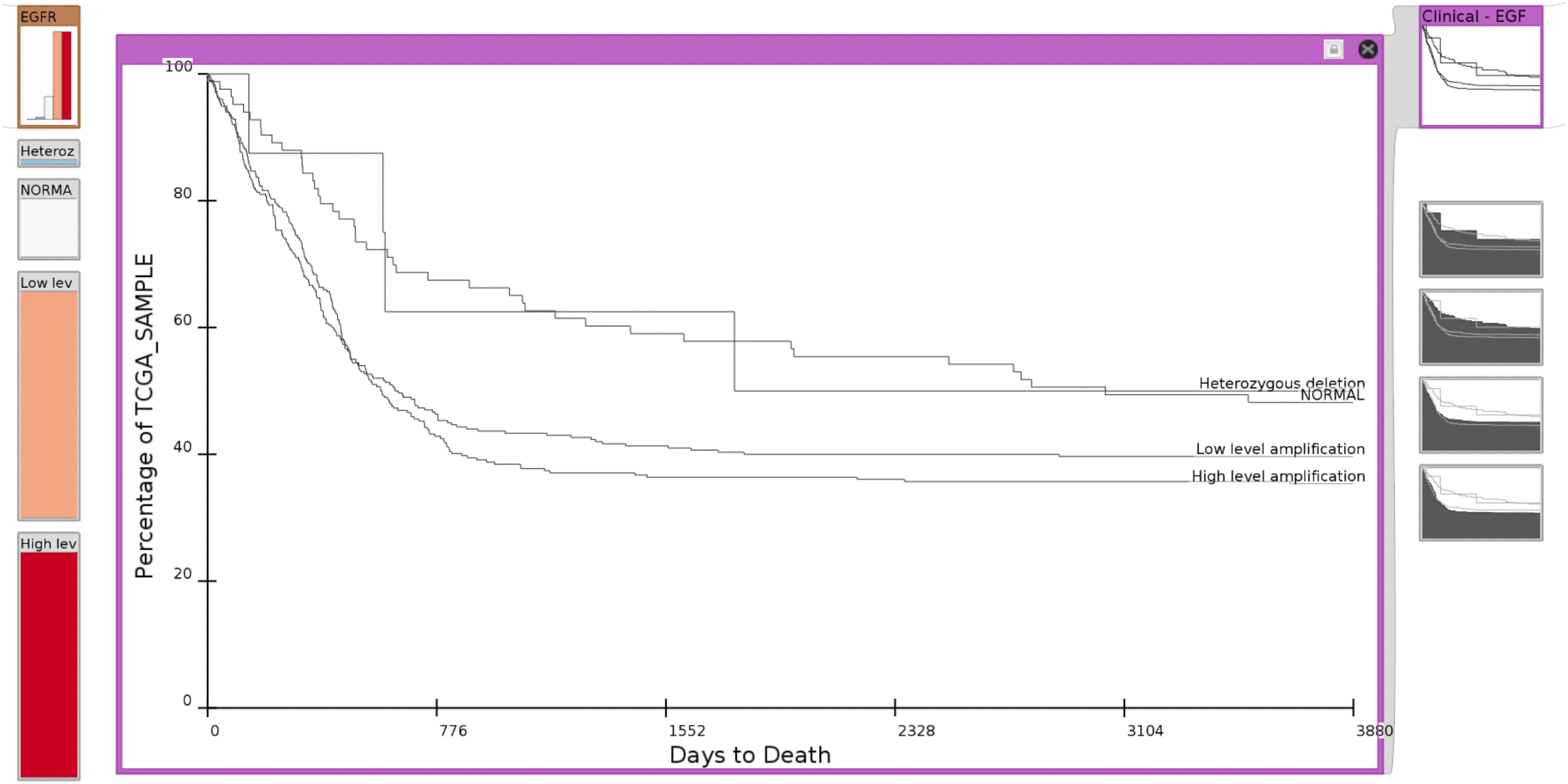
Stratification on Copy Number Status

Other Data – Same Stratification





Survival data in Kaplan Meier plots



EGFR
Heteroz
NORMA
Low lev
High lev
Clinical - EGF
100
80
60
40
20
0
Percentage of TCGA_SAMPLE
0
776
1552
2328
3104
3880
Days to Death
Heterozygous deletion
NORMAL
Low level amplification
High level amplification

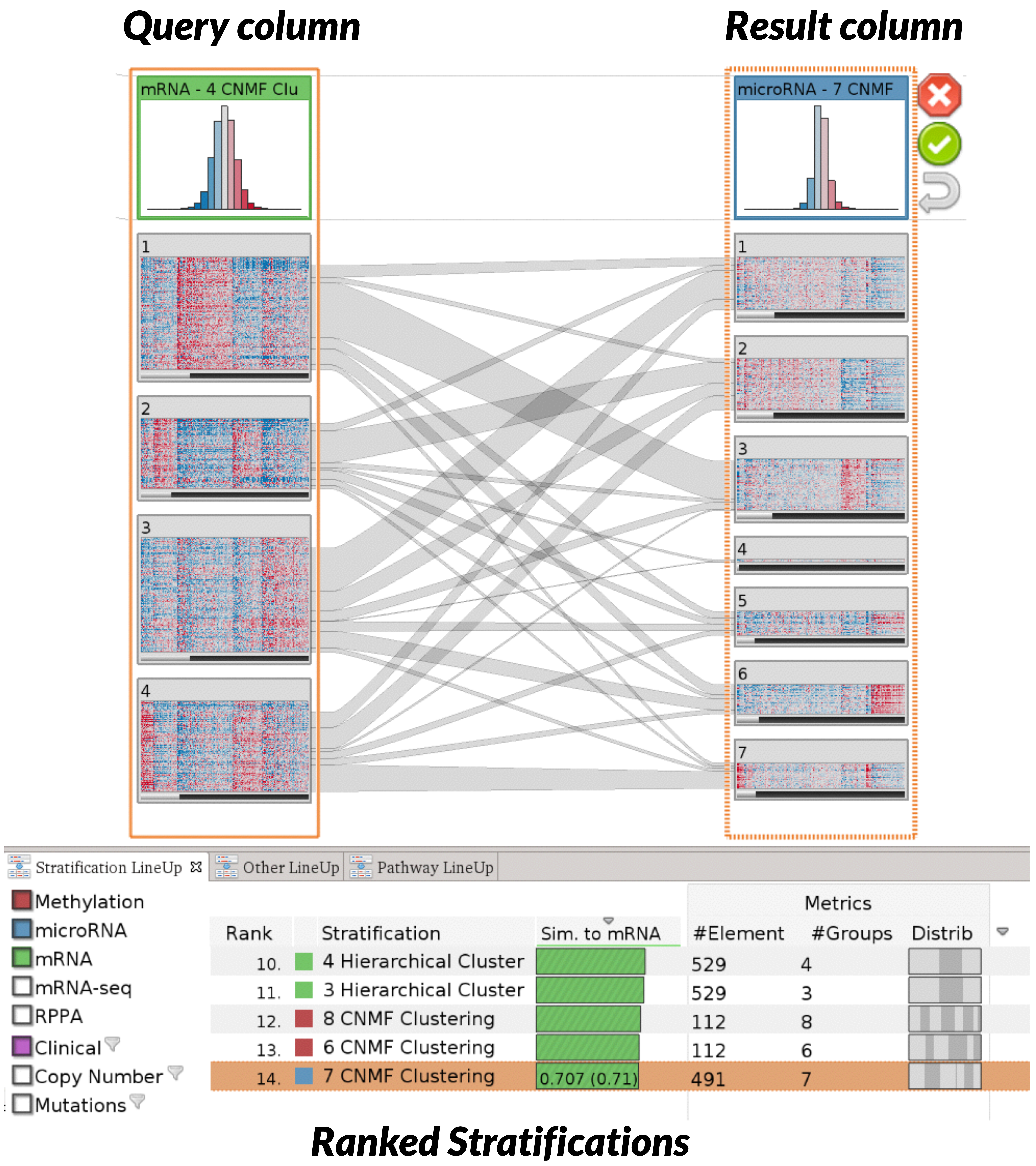
Detail View

Integrating Computation

Data Visualization

Algorithms

Algorithm Output Visualization



More Information:

http://stratomex.caleydo.org