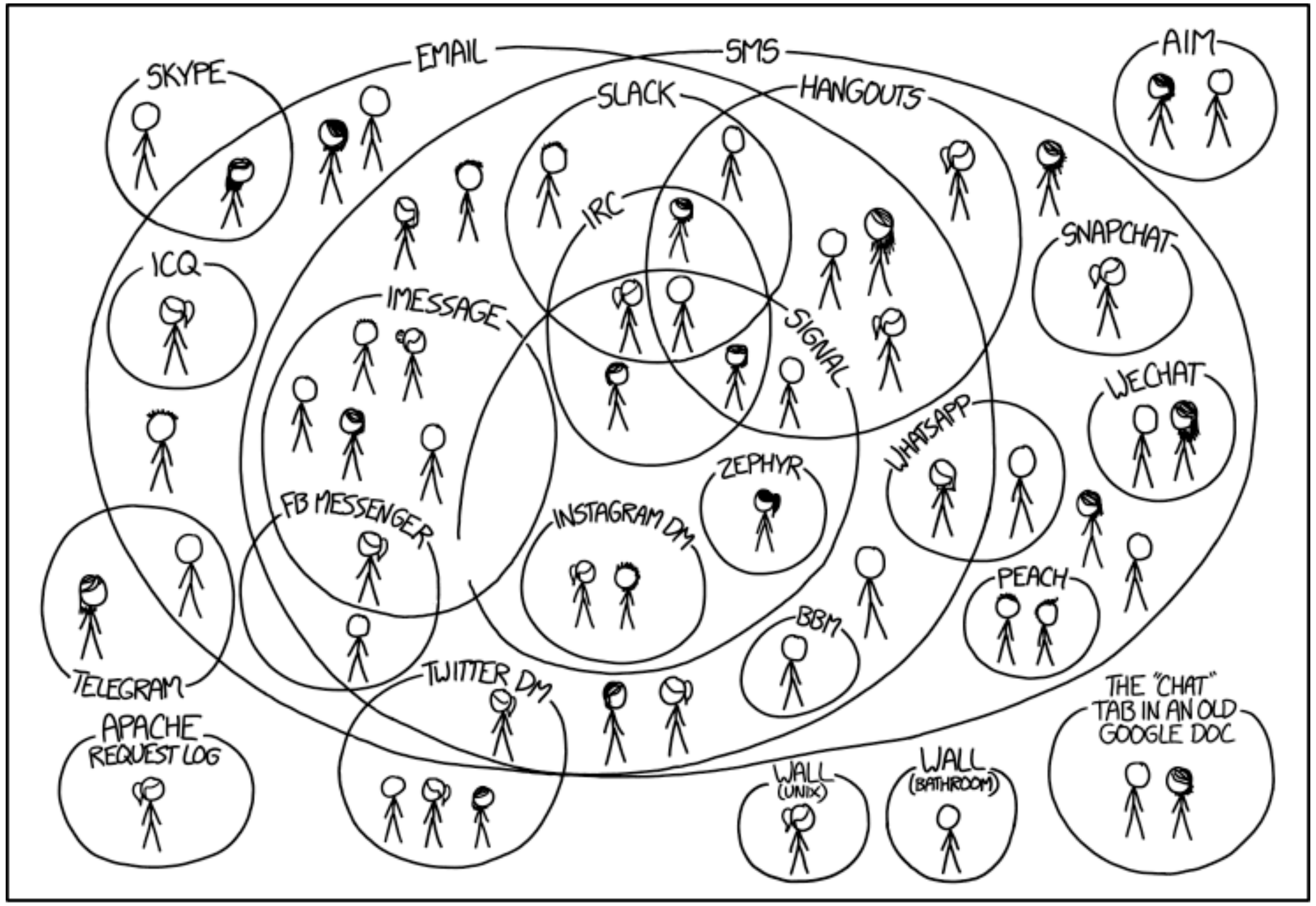
CS-5630 / CS-6630 Visualization for Data Science
Sets and Text

Alexander Lex
alex@sci.utah.edu

THE UNIVERSITY OF UTAH



I HAVE A HARD TIME KEEPING TRACK OF WHICH CONTACTS USE WHICH CHAT SYSTEMS.

[xkcd]

Text and Document Visualization

Slides adapted from Hendrik Strobelt

Text / Language

Abstract, general

Extremely expressive

Different across population groups (countries, accents, religions,…)

Linear perception

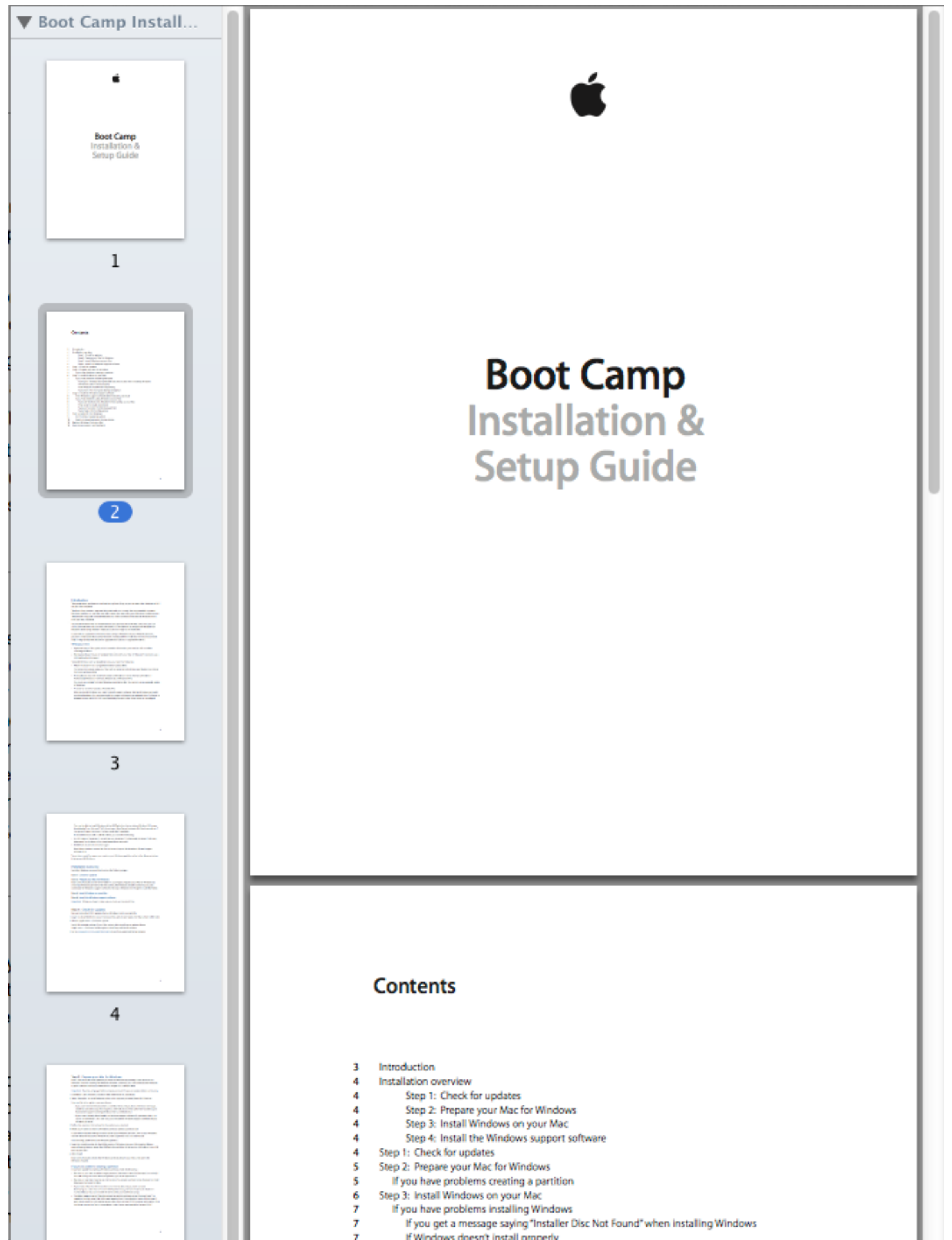
Semi-structured (headings, paragraphs, sentences, grammar, words)

Visualization for “Raw” Text

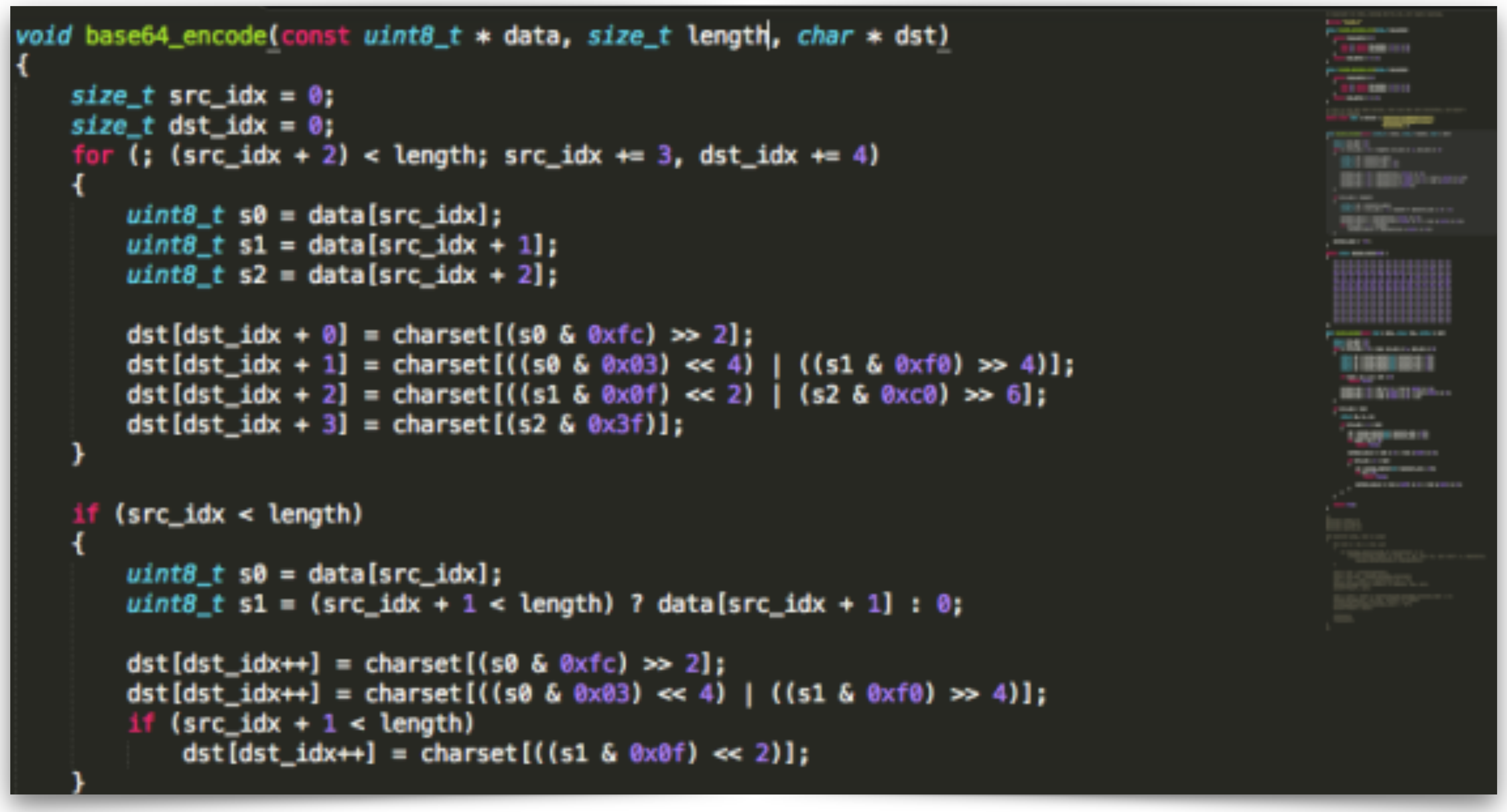
in daily use..

enriched text - hypertext linking (graph navigation)

overview & detail



highlighting semantics



Visualization for "Raw" Text

Document Lens

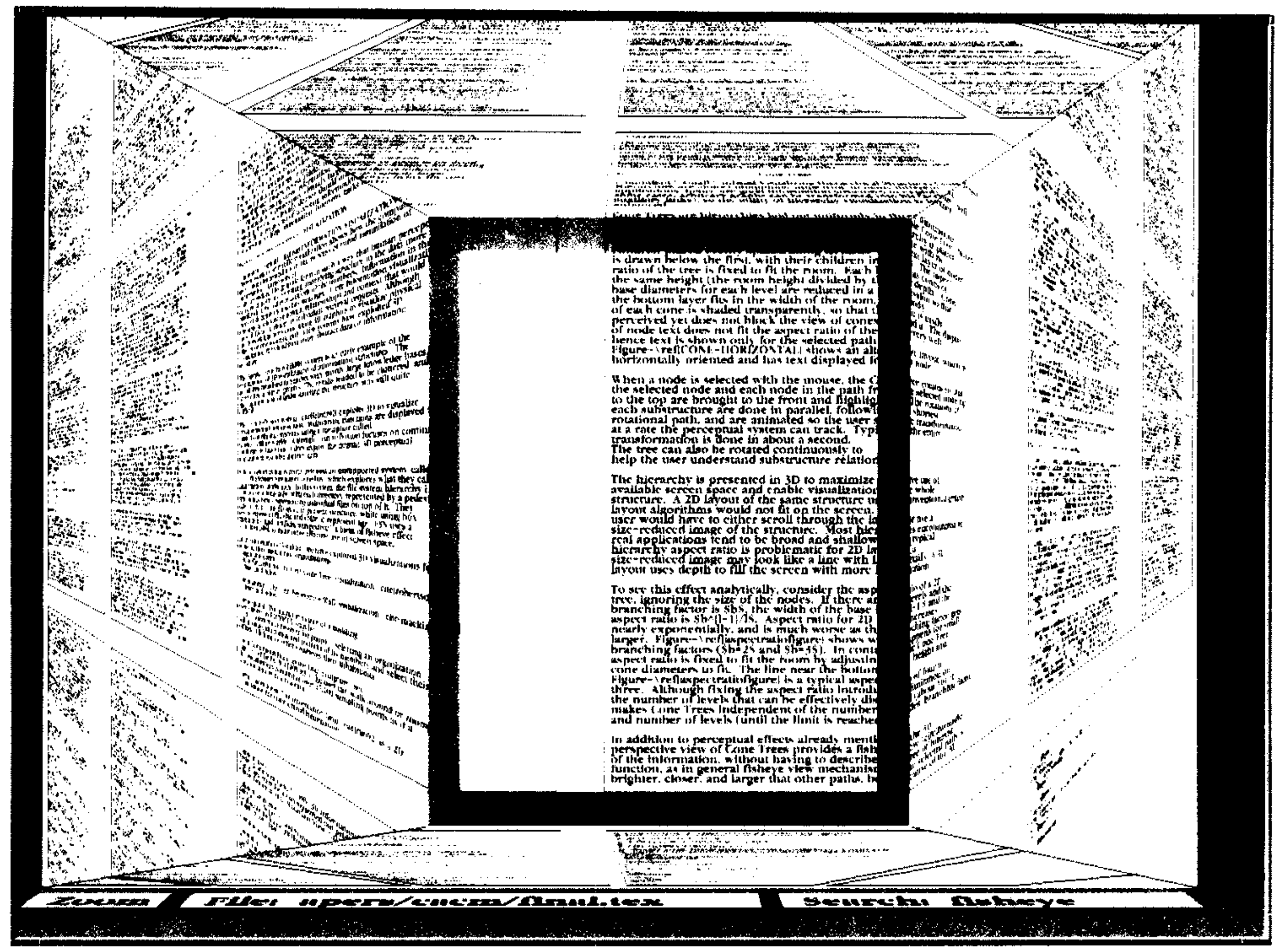
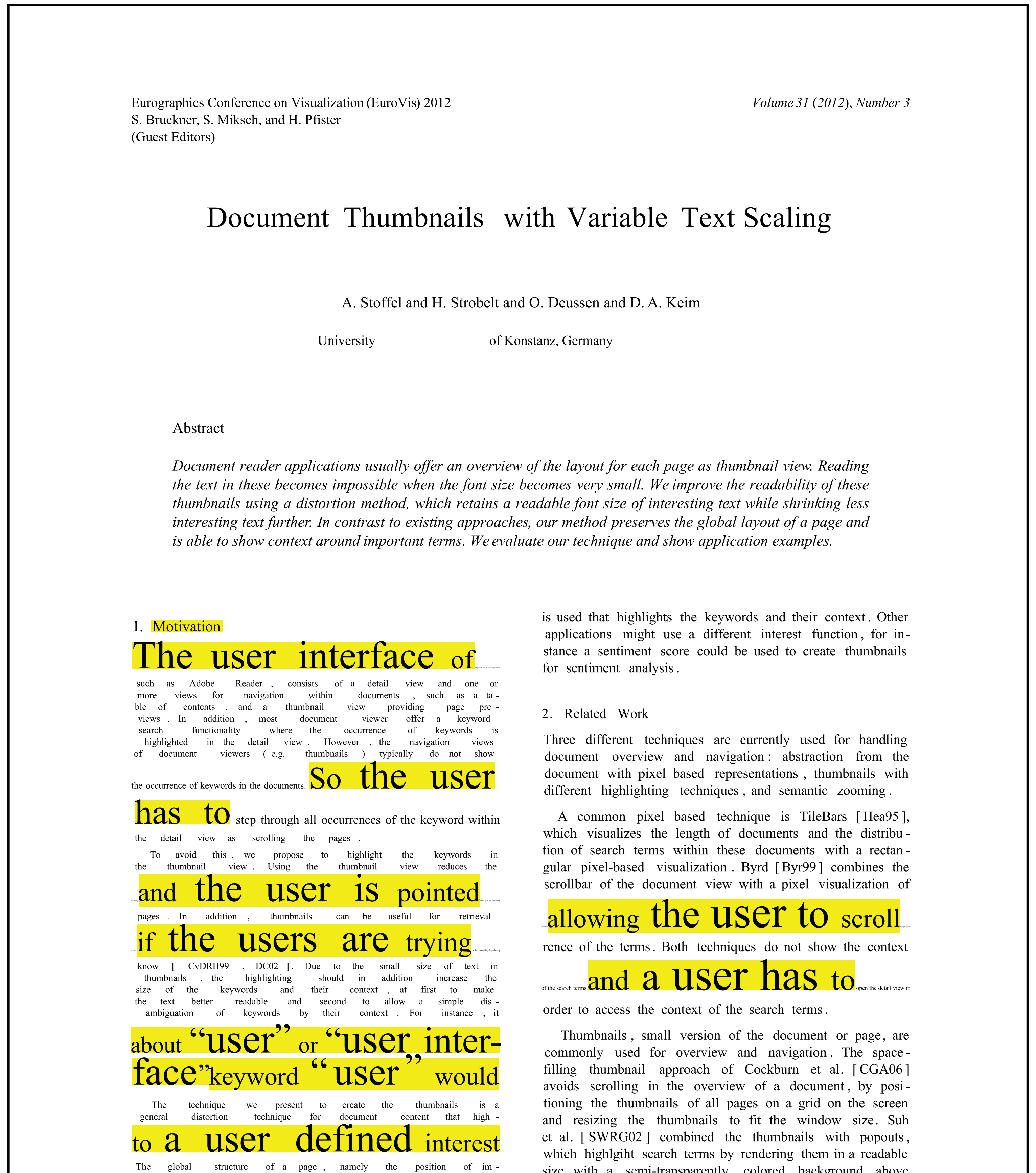


Figure 3: Document Lens with lens pulled toward the user. The resulting truncated pyramid makes text near the lens' edges readable.

Robertson, George G., and Jock D. Mackinlay
The document lens
Proceedings of the 6th annual ACM symposium on User interface software and technology. ACM, 1993.

Visualizing Search Results



Eurographics Conference on Visualization (EuroVis) 2012
S. Bruckner, S. Miksch, and H. Pfister
(Guest Editors)

Volume 31 (2012), Number 3

Document Thumbnails with Variable Text Scaling

A. Stoffel and H. Strobelt and O. Deussen and D. A. Keim

University of Konstanz, Germany

Abstract

Document reader applications usually offer an overview of the layout for each page as thumbnail view. Reading the text in these becomes impossible when the font size becomes very small. We improve the readability of these thumbnails using a distortion method, which retains a readable font size of interesting text while shrinking less interesting text further. In contrast to existing approaches, our method preserves the global layout of a page and is able to show context around important terms. We evaluate our technique and show application examples.

1. Motivation

The user interface of such as Adobe Reader, consists of a detail view and one or more views for navigation within documents, such as a table of contents, and a thumbnail view providing page previews. In addition, most document viewer offer a keyword search functionality where the occurrence of keywords is highlighted in the detail view. However, the navigation views of document viewers (e.g. thumbnails) typically do not show the occurrence of keywords in the documents. So the user has to step through all occurrences of the keyword within the detail view as scrolling the pages.

To avoid this, we propose to highlight the keywords in the thumbnail view. Using the thumbnail view reduces the and the user is pointed pages. In addition, thumbnails can be useful for retrieval if the users are trying know [CvDRH99, DC02]. Due to the small size of text in thumbnails, the highlighting should in addition increase the size of the keywords and their context, at first to make the text better readable and second to allow a simple disambiguation of keywords by their context. For instance, it about "user" or "user interface" keyword "user" would

The technique we present to create the thumbnails is a general distortion technique for document content that high- to a user defined interest

The global structure of a page, namely the position of im-

is used that highlights the keywords and their context. Other applications might use a different interest function, for instance a sentiment score could be used to create thumbnails for sentiment analysis.

2. Related Work

Three different techniques are currently used for handling document overview and navigation: abstraction from the document with pixel based representations, thumbnails with different highlighting techniques, and semantic zooming.

A common pixel based technique is TileBars [Hea95], which visualizes the length of documents and the distribution of search terms within these documents with a rectangular pixel-based visualization. Byrd [Byr99] combines the scrollbar of the document view with a pixel visualization of allowing the user to scroll rence of the terms. Both techniques do not show the context of the search terms and a user has to open the detail view in order to access the context of the search terms.

Thumbnails, small version of the document or page, are commonly used for overview and navigation. The space-filling thumbnail approach of Cockburn et al. [CGA06] avoids scrolling in the overview of a document, by positioning the thumbnails of all pages on a grid on the screen and resizing the thumbnails to fit the window size. Suh et al. [SWRG02] combined the thumbnails with popouts, which highlgiht search terms by rendering them in a readable size with a semi-transparently colored background above

Document Thumbnails with Variable Text Scaling
A. Stoffel, H. Strobelt, O. Deussen, D. A. Keim
Computer Graphics Forum, volume 31 issue 3 pp.

Working with Text

unstructured text

4 x ’t'
3 x ‘u’
2 x ‘r’
2 x ‘e’
…

structured data

Structured Text Features

simple counts (bag of words)
used for similarity measures

	princess	dragon	castle
doc1	1	1	1
doc2	0	0	1

Typical Steps of Processing to derive Text Features

Large collections require pre-processing of text to extract information and align text. Typical steps are:

- cleaning (regular expressions)
- sentence splitting
- change to lower case
- stopword removal (most frequent words in a language)
- stemming - demo porter stemmer
- POS tagging (part of speech) - demo
- noun chunking
- NER (name entity recognition) - demo opencalais
- deep parsing - try to “understand” text.

Text features are complicated

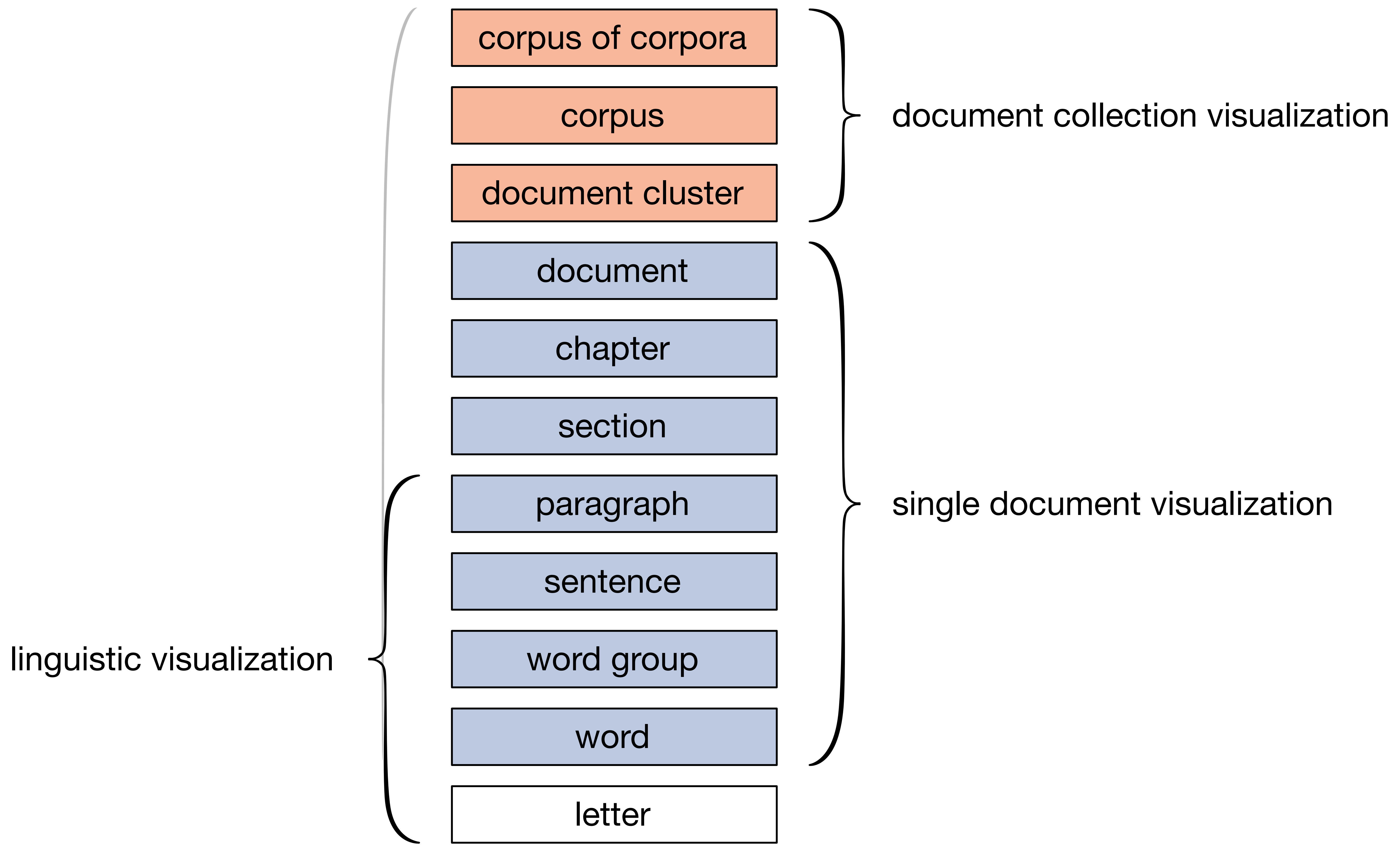
Toilet out of order. Please use floor below.

One morning I shot an elephant in my pajamas. How he got in my pajamas, I don't know.

Did you ever hear the story about the blind carpenter who picked up his hammer and saw?

http://en.wikipedia.org/wiki/List_of_linguistic_example_sentences

Text Units Hierarchy



Wordle

Frequency-based

words that occur often are large

Can vary font type,
size, color, etc.

[Viegas 2009]

http://www.wordle.net

Wordle vs Tag Cloud

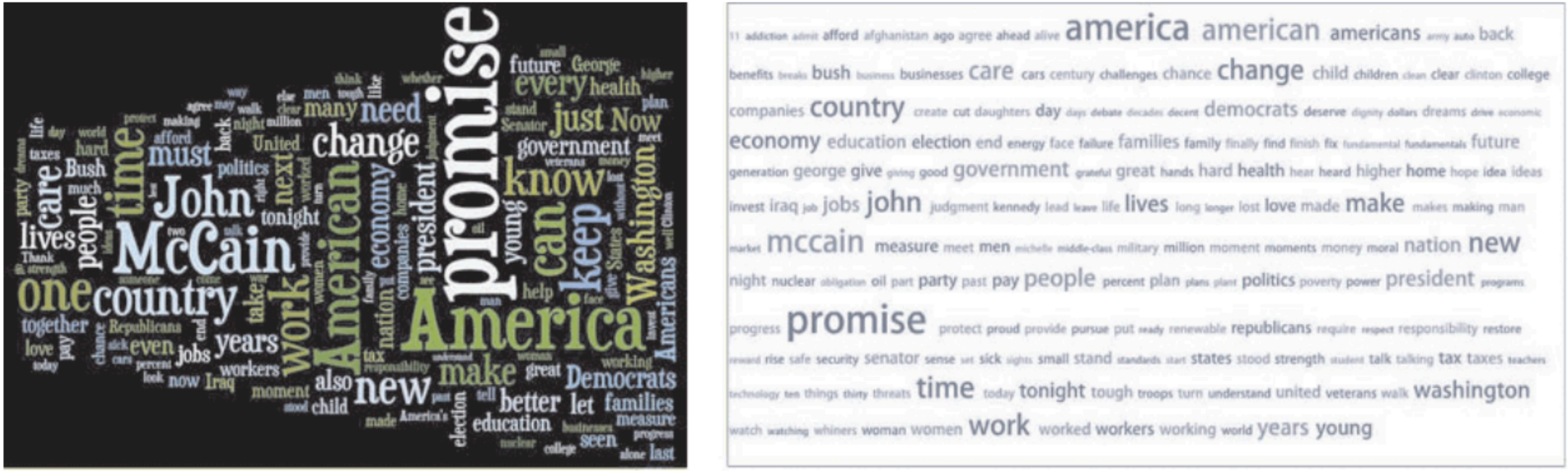


Fig 2: Wordle vs. Tag Cloud of Barack Obama’s speech at the Democratic Convention in 2008.


PARTY
CRAFT
BIRTHDAY
BAGS
KIDS
BOX
ACTIVITY
TWEEN
COLORED
BAG
CRAFTING
GIFT
SALT
PAPER
PLASTIC
FUN
CHALK
MAKING
CRAFTY
POST
LIKED
MAKE
PACKED
CELEBRATE
EASY
HAPPY
PERSONAL
NIECE
WEEKEND
SUBSCRIBE
JUST
USED
MAILING
WELL
WANT
CHOC
FRIENDS
BOXES
LOCAL
GROCERY
ONE
USE
STORE
BALLOONS
SAND
LOOKING
MOM

ideas
idealings
great
idea
imagination
friends
environment
Napoleon
creative
like
Steven
Johnson
time
good
mind
bouncing
person
new
Hill
come
word
flash
one
play
appears
born
individuals
Champions
example
little
just

malicetowardsnone
God
War
shall
nation
Union
years
right
offenses
Woe
wills
gives
come
came
cause
hope
men
needs
public
conflicts
laves
judge
extend
drawn
ago

Word Tree

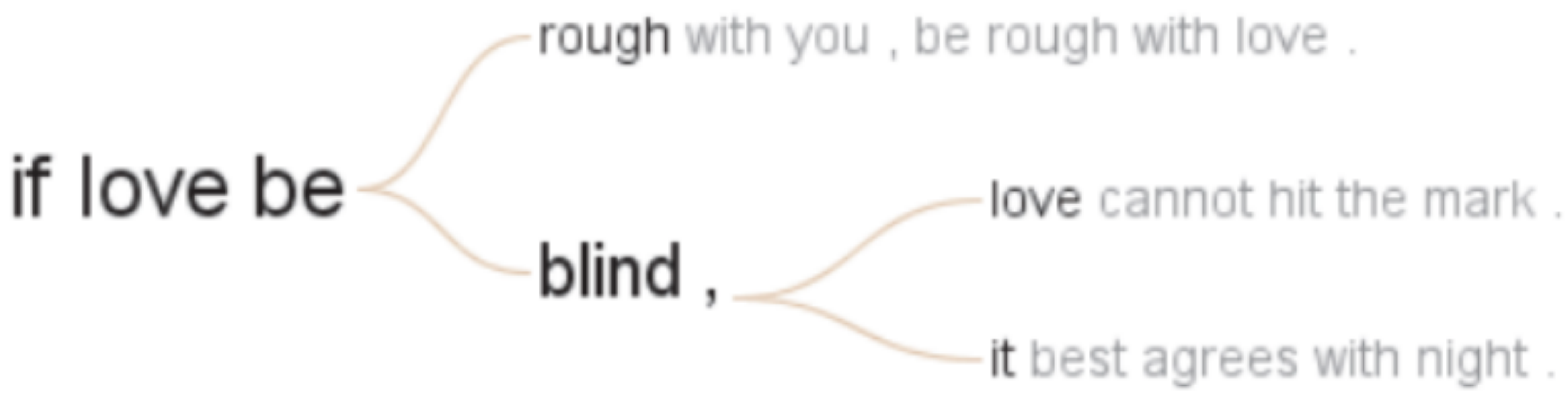
Text

if love be rough with you , be rough with love .

if love be blind , love cannot hit the mark .

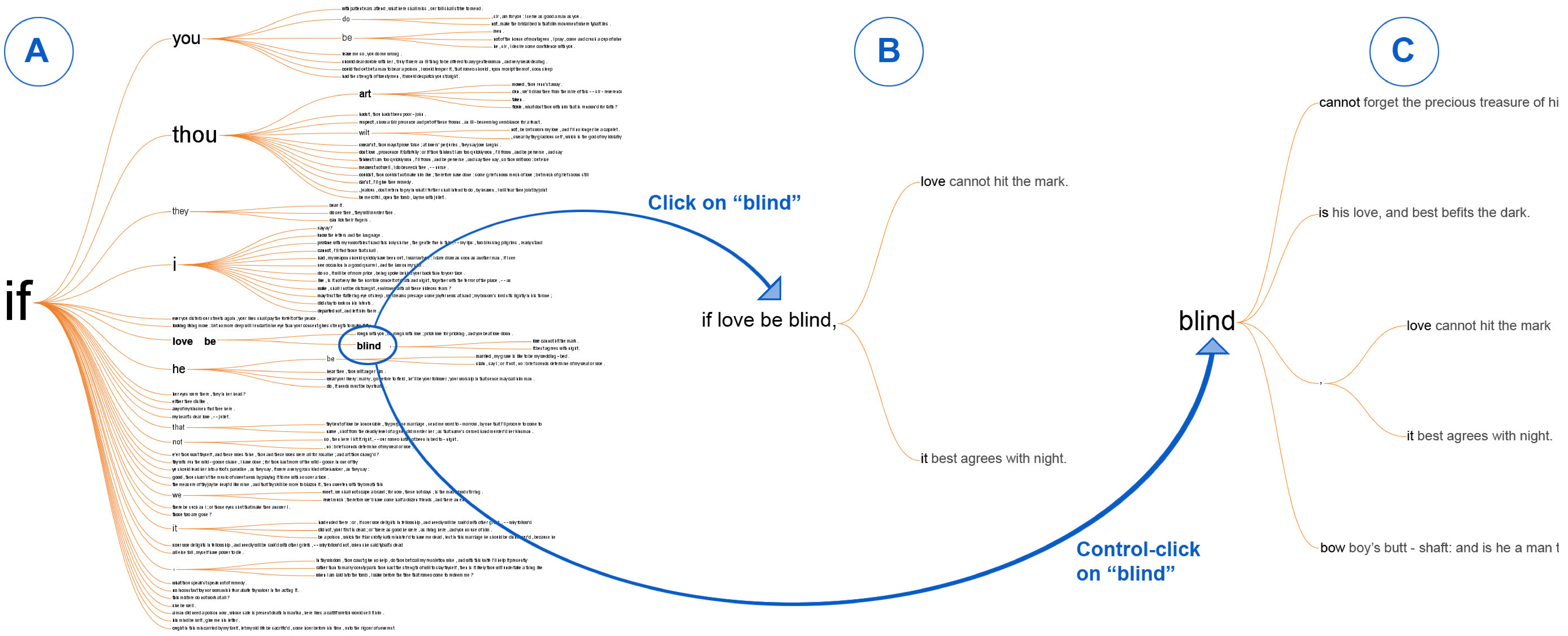
if love be blind , it best agrees with night .

WordTree



[Wattenberg 2008]

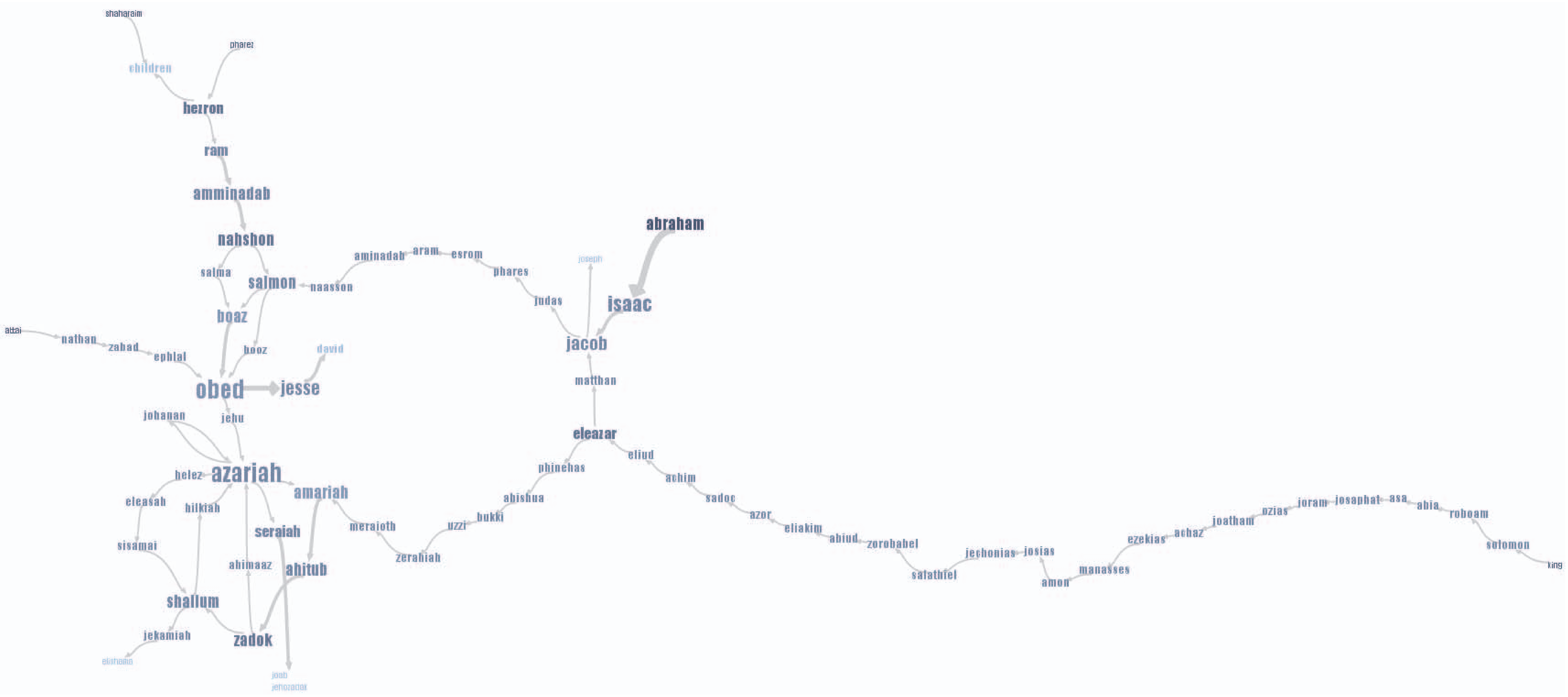
Search for “if” in romeo & Juliet



The word tree, an interactive visual concordance
M Wattenberg, FB Viégas
Visualization and Computer Graphics, IEEE Transactions on 14 (6), 1221-1228

PhraseNets

Frank van Ham, Martin Wattenberg, and Fernanda B. Viegas.
Mapping Text with Phrase Nets.
IEEE Transactions on Visualization and Computer Graphics 15, 6 (November 2009)



shaharaim
pharez
children
hezron
ram
amminadab
nahshon
salma
salmon
naasson
aminadab
aram
esrom
phares
judas
joseph
abraham
isaac
jacob
boaz
booz
david
attai
nathan
zabad
ephlal
obed
jesse
johanan
jehu
matthan
azariah
helez
amariah
eleazar
eliud
phinehas
achim
eleasah
hilkiah
seraiah
meraioth
abishua
sadoc
azor
bukki
uzzi
eliakim
sisamai
abiud
zorobabel
ahimaaz
ahitub
zerahiah
jechonias
josias
salathiel
amon
manasses
ezekias
achaz
joatham
ozias
joram
josaphat
asa
abia
roboam
solomon
king
shallum
jekamiah
zadok
elishama
joab
jehozadak

Corpora: MDS Approaches

use bag-of-word to project documents w.r.t. text similarity into a landscape

(only) one example

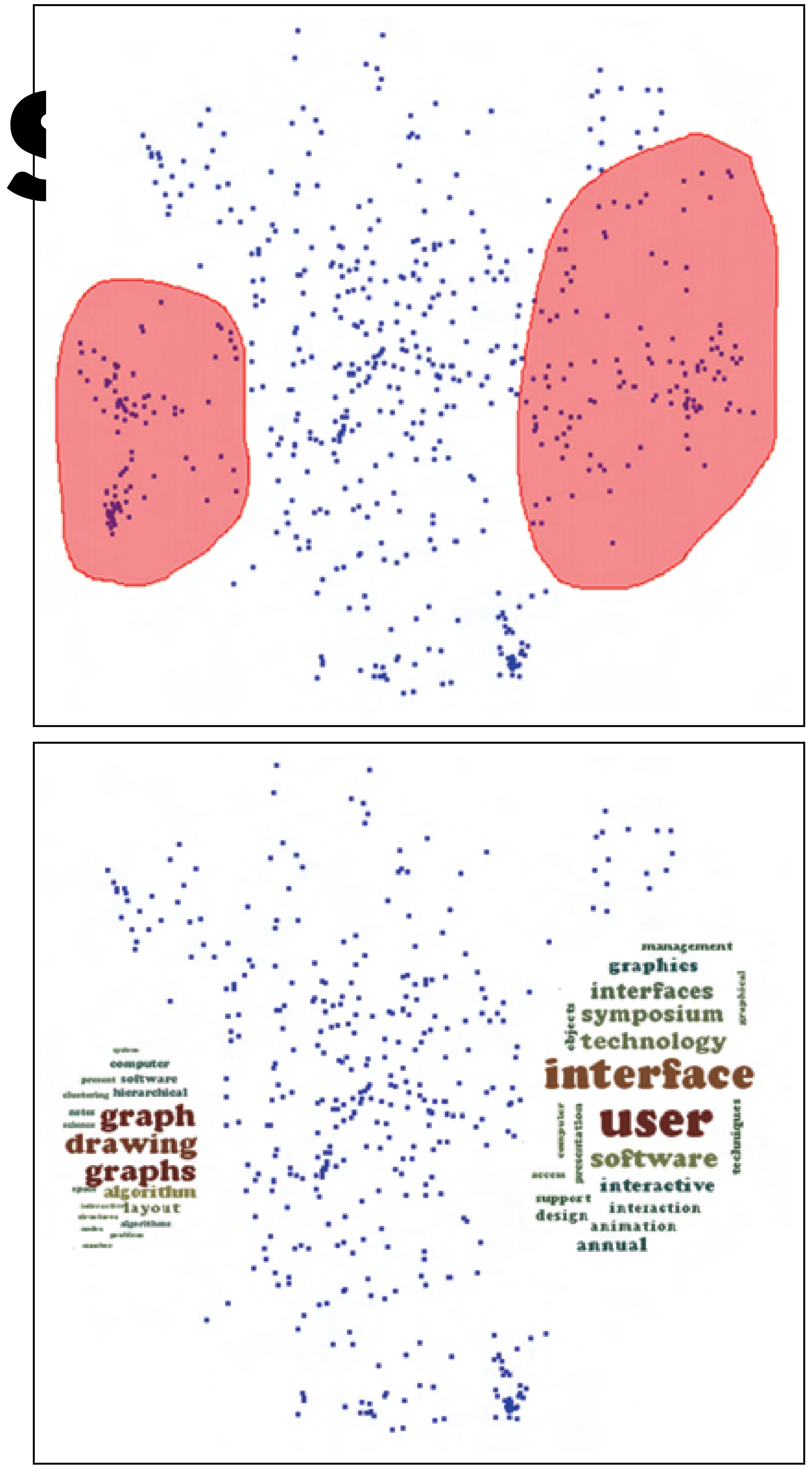


Figure 5: *A user can interactively draw a region (polygon) containing a subset of documents of interest (top figure). Keywords are extracted from the selected document and their corresponding word could is built inside the user-defined region (bottom figure).*

Fernando V. Paulovich, Franklina M. B. Toledo, Guilherme P. Telles, Rosane Minghim, and Luis Gustavo Nonato.
Semantic Wordification of Document Collections.
Comp. Graph. Forum 31, 3pt3 (June 2012)

JigSaw

Document Cluster View
Edit View Bookmarks Export Options
Highlight Viewed Documents
Filters
All Filters
Group by Filters
Undo Filters
Hide Unfiltered
Clusters
Text (Default)
Freq Words
Unique Words
All Documents
center,faa,air: 54
staff,professional,member: 3
•,national,director: 24
ksm,interrogation,2003: 38
muslim,islamic,ladin: 5
muslims,saudi,governments: 21
terrorist,unit,state: 28
studied,pause,reflect: 2
ladin,qaeda,attacks: 90
/,deputy,staff: 1
center,faa,air
staff,professional,member
director,departments,agency
ksm,interrogation,2003
muslim,islamic,fight
muslims,saudi,governments
terrorist,unit,state
studied,pause,reflect
attacks,operators,taliban
binalshibh,atta,flight
interview,general,agents
committee,senate,department
flight,11,centers
deputy,staff
interview,2004,fdny
towers,north,south
9/11,saudi,deceased
agency,officers,department
offices,director,moussaoui
islamic,v,laden
List View
OR
AND
Person
Add all
Clear
Bill Clinton
George W. Bush
Usama Bin Ladin
Richard Clarke
George Tenet
Condoleezza Rice
Dick Cheney
Nawaf al Hazmi
Donald Rumsfeld
Ramzi Binalshibh
Dale Watson
Khalid al Mihdhar
Samuel Berger
Janet Reno
Hugh Shelton
Waleed al Shehri
James Pavitt
Khalid Sheikh Mohammed
Ramzi Yousef
Salem al Hazmi
Stephen Hadley
Zacarias Moussaoui
Ahmed al Ghamdi
John McLaughlin
Abu Zubaydah
Ayman al Zawahiri
Saddam Hussein
Abdullah
Charles Allen
Mohamed Atta
Roger Cressey
Ahmed al Nami
Ali Abdul Aziz Ali
Andrew Card
Bob Woodward
Colin Powell
Gary Schroen
Hani Hanjour
Marwan al Shehhi
Saeed al Ghamdi
Satam al Suqami

Compare Corpora

Compare topics between text collections

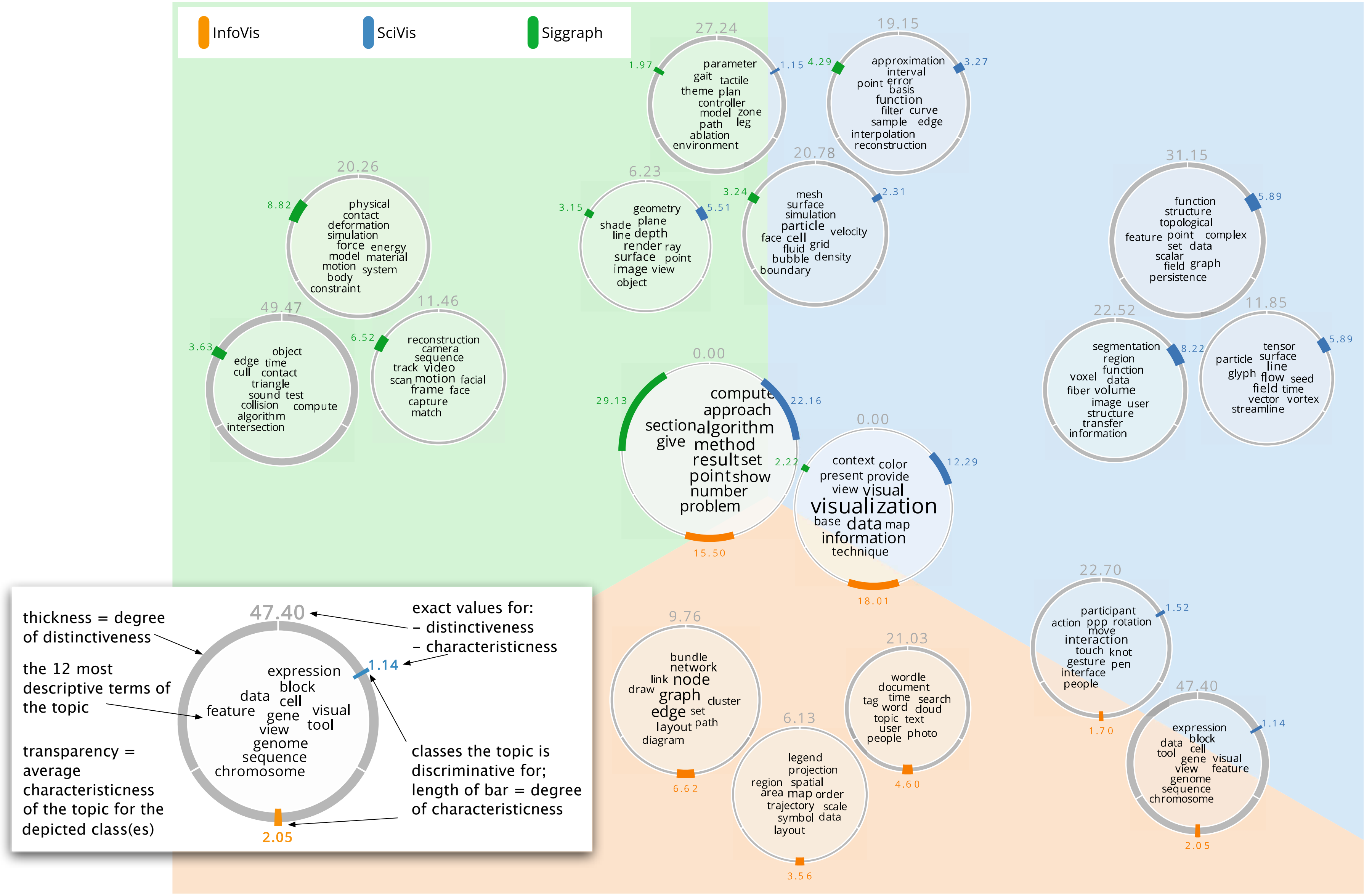
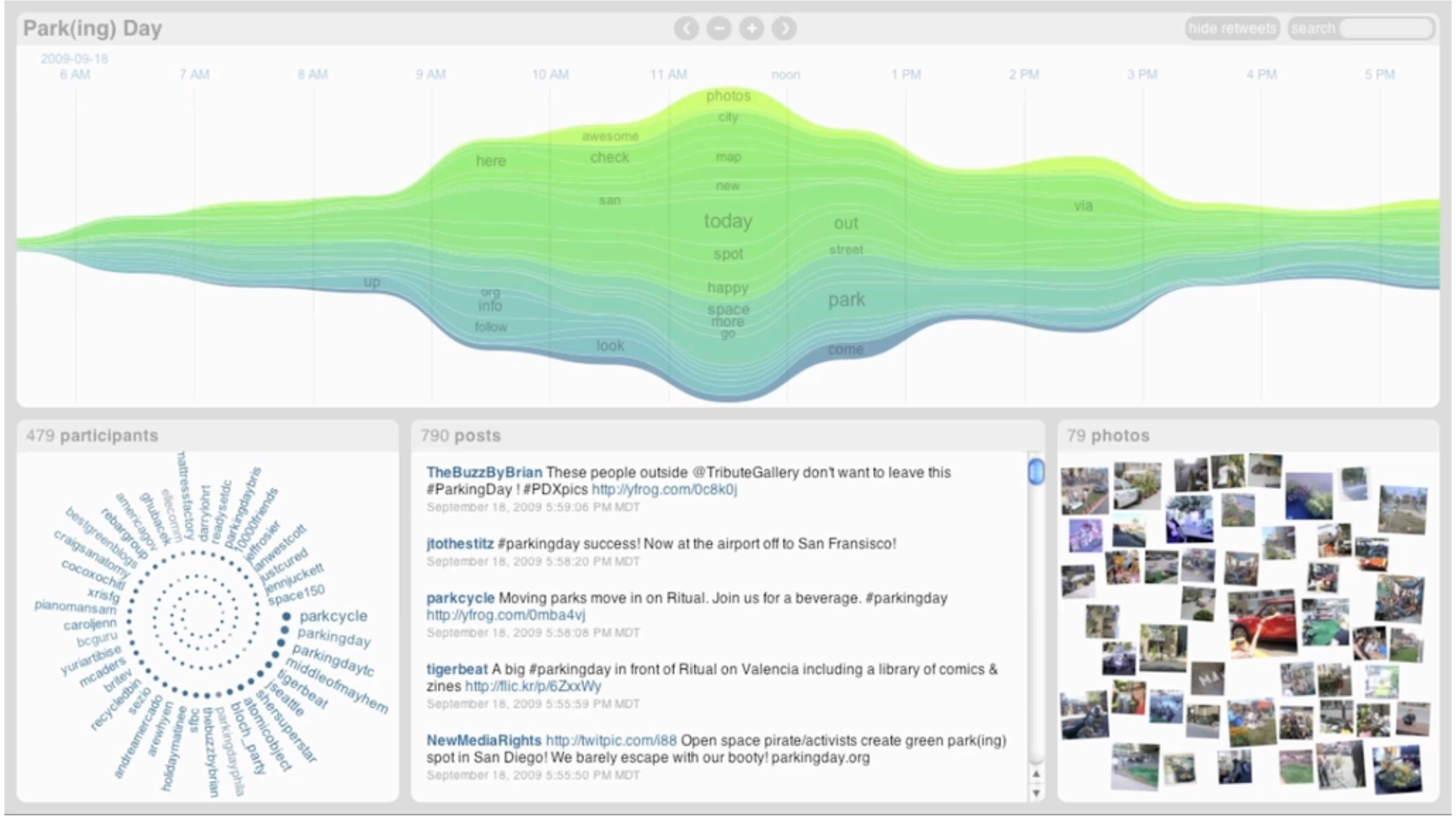


Figure 1: Comparison of 495 papers of InfoVis, SciVis, and Siggraph (discrimination threshold = 6, number of topics = 30)

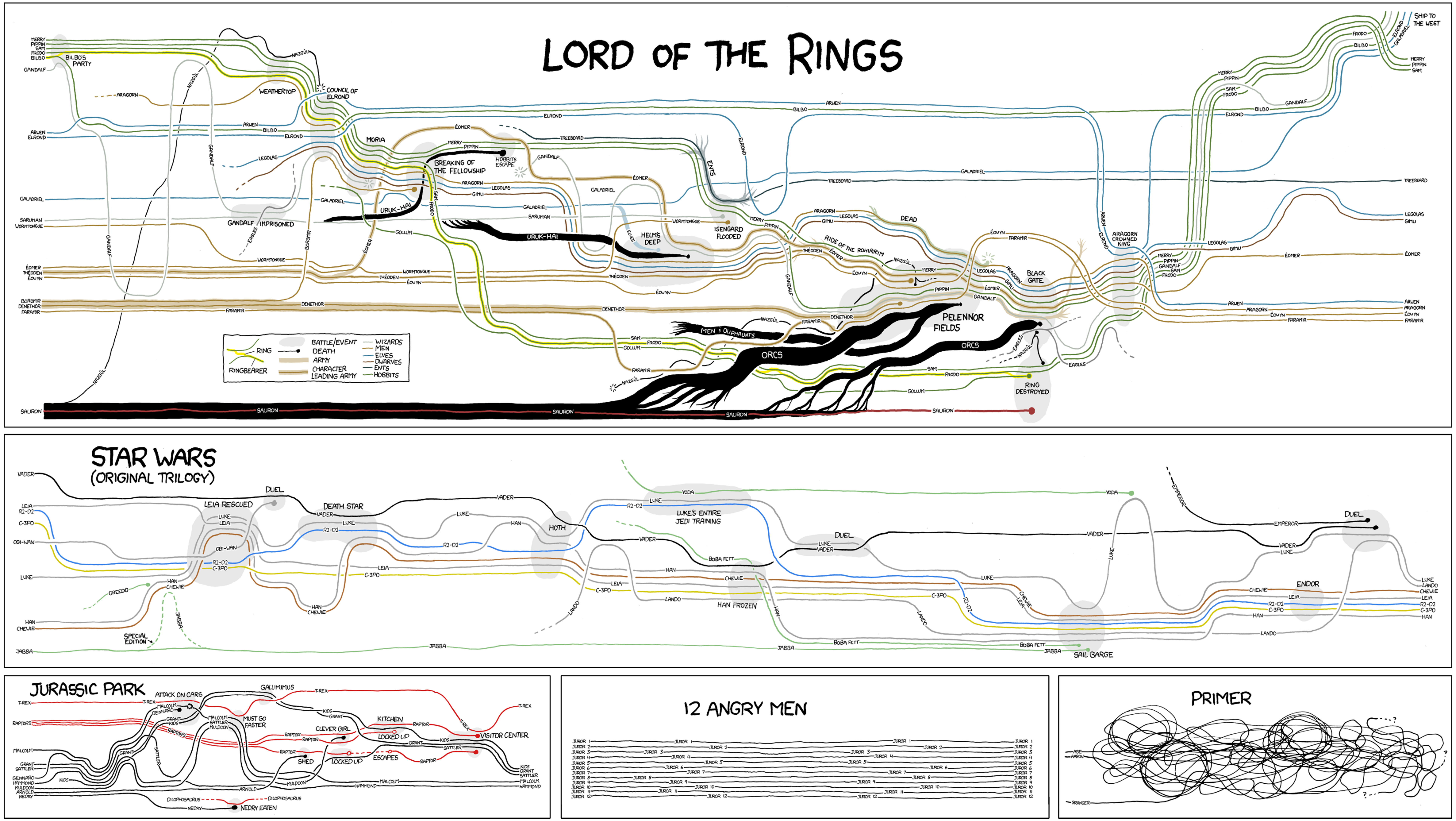
Comparative Exploration of Document Collections: a Visual Analytics Approach (http://ditop.hs8.de)
D. Oelke, H. Strobelt, C. Rohrdantz, I. Gurevych, and O. Deussen

Vis for Time-Evolving Document Collections



Marian Dörk, Daniel Gruen, Carey Williamson, and Sheelagh Carpendale. **A Visual Backchannel for Large-Scale Events**.
TVCG: Transactions on Visualization and Computer Graphics (Proceedings Information Visualization 2010

THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.



https://xkcd.com/657/

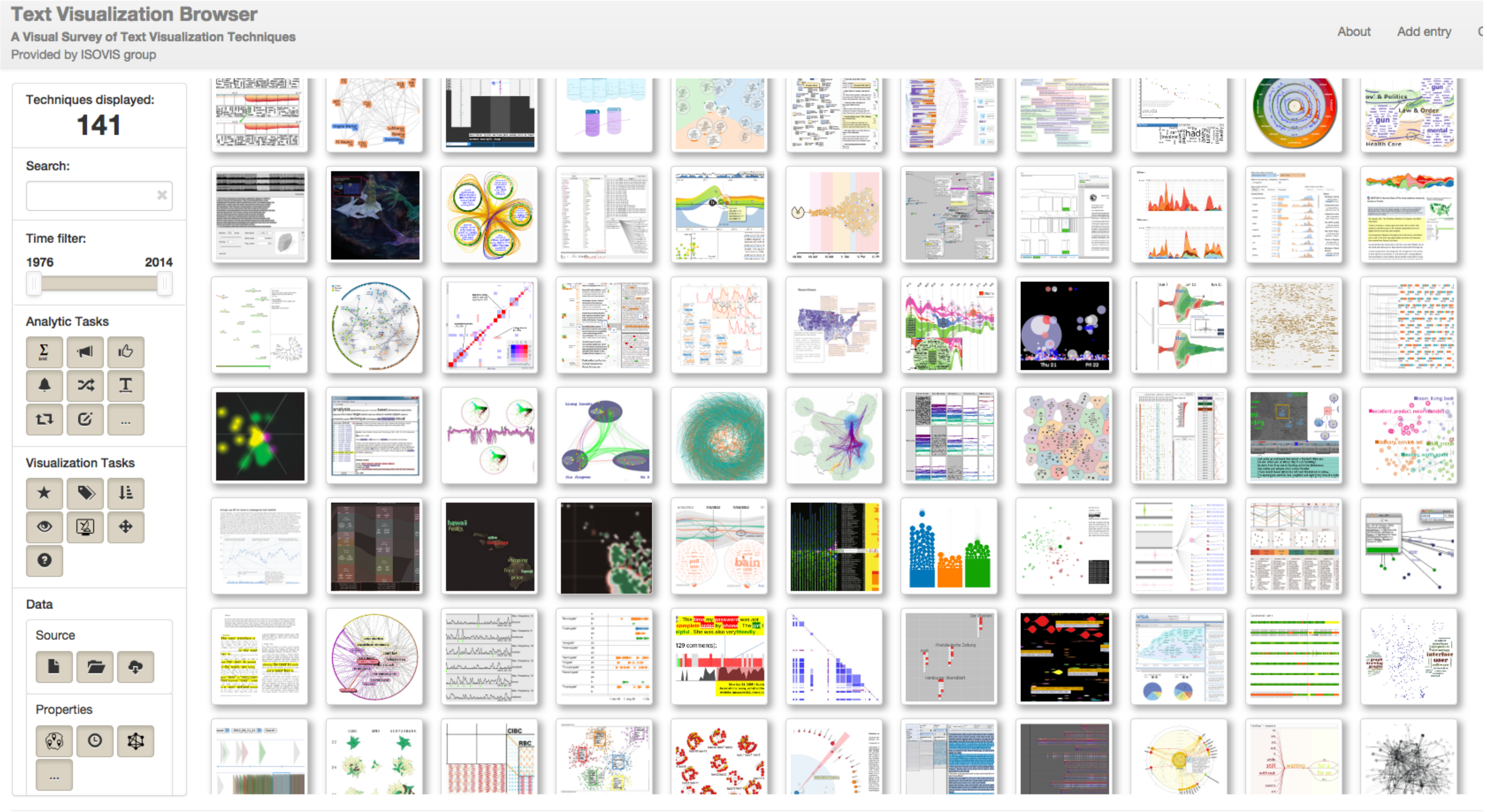
StoryFlow: Tracking the Evolution of Stories

[Liu 2013]

Visualization and NLP

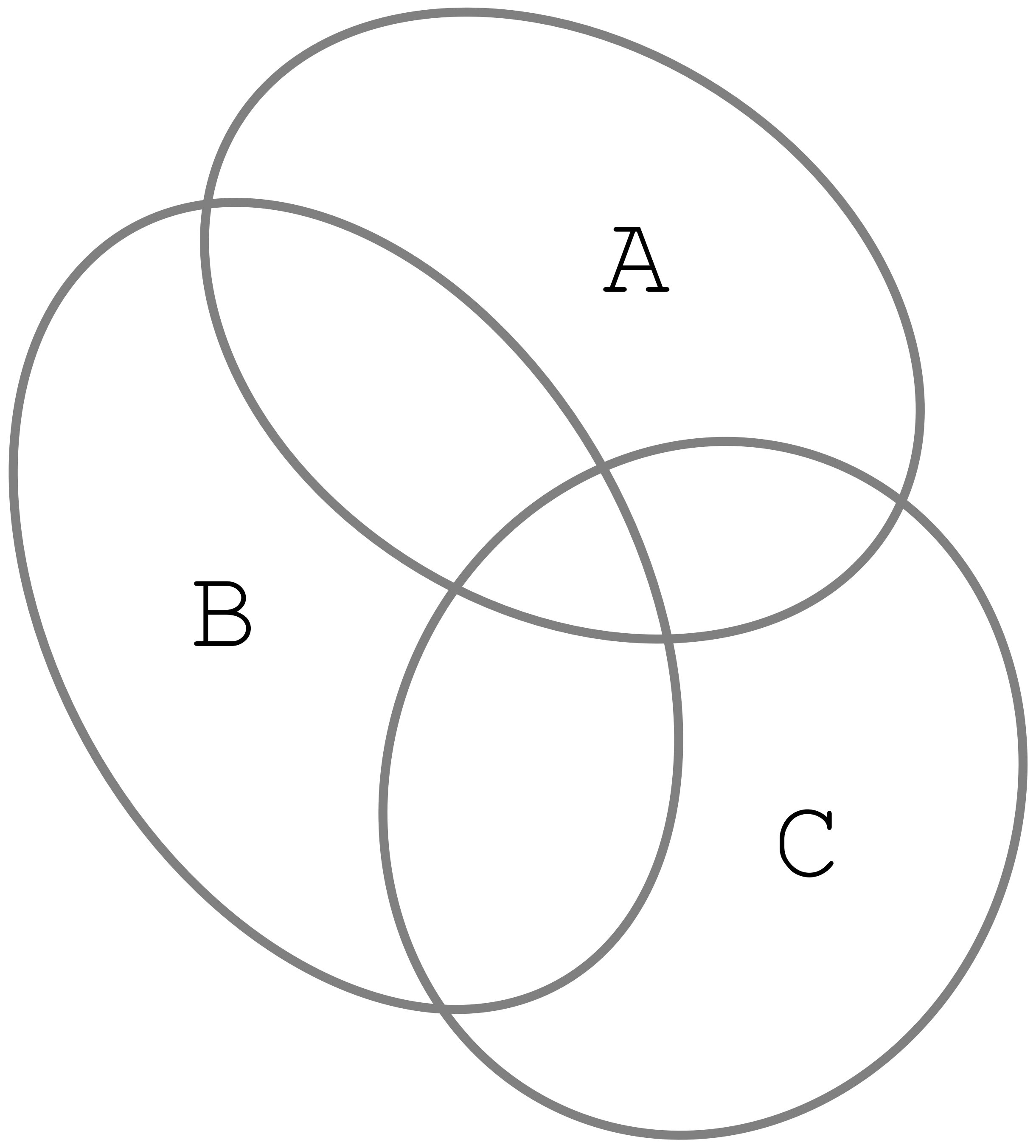
http://gltr.io/dist/index.html

http://textvis.lnu.se/



Design Workshop

```
item1 : A
item2 : A
item3 : A, B
item4 : A, C
item5 : A, B, C
item6 : B
item7 : B, C
item8 : C
…
```



Venn diagram

LETTER

doi:10.1038/nature11241

The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants

Angélique D'Hont[1]*, France Denoeud[2,3,4]*, Jean-Marc Aury[2], Franc-Christophe Baurens[1], Françoise Carreel[1,5], Olivier Garsmeur[1], Benjamin Noel[2], Stéphanie Bocs[1], Gaëtan Droc[1], Mathieu Rouard[6], Corinne Da Silva[2], Kamel Jabbari[2,3,4], Céline Cardi[1], Julie Poulain[2], Marlène Souquet[1], Karine Labadie[2], Cyril Jourda[1], Juliette Lengellé[1], Marguerite Rodier-Goud[1], Adriana Alberti[2], Maria Bernard[2], Margot Correa[2], Saravanaraj Ayyampalayam[7], Michael R. Mckain[7], Jim Leebens-Mack[7], Diane Burgess[8], Mike Freeling[8], Didier Mbéguié-A-Mbéguié[9], Matthieu Chabannes[5], Thomas Wicker[10], Olivier Panaud[11], Jose Barbosa[11], Eva Hribova[12], Pat Heslop-Harrison[13], Rémy Habas[5], Ronan Rivallan[1], Philippe Francois[1], Claire Poiron[1], Andrzej Kilian[14], Dheema Burthia[1], Christophe Jenny[1], Frédéric Bakry[1], Spencer Brown[15], Valentin Guignon[1,6], Gert Kema[16], Miguel Dita[19], Cees Waalwijk[16], Steeve Joseph[1], Anne Dievart[1], Olivier Jaillon[2,3,4], Julie Leclercq[1], Xavier Argout[1], Eric Lyons[17], Ana Almeida[8], Mouna Jeridi[1], Jaroslav Dolezel[12], Nicolas Roux[6], Ange-Marie Risterucci[1], Jean Weissenbach[2,3,4], Manuel Ruiz[1], Jean-Christophe Glaszmann[1], Francis Quétier[18], Nabila Yahiaoui[1] & Patrick Wincker[2,3,4]

Bananas (*Musa* spp.), including dessert and cooking types, are giant perennial monocotyledonous herbs of the order Zingiberales, a sister group to the well-studied Poales, which include cereals. Bananas are vital for food security in many tropical and subtropical countries and the most popular fruit in industrialized countries[1]. The *Musa* domestication process started some 7,000 years ago in Southeast Asia. It involved hybridizations between diverse species and subspecies, fostered by human migrations[2], and selection of diploid and triploid seedless, parthenocarpic hybrids thereafter widely dispersed by vegetative propagation. Half of the current production relies on somaclones derived from a single triploid genotype (Cavendish)[1]. Pests and diseases have gradually become adapted, representing an imminent danger for global banana production[3,4]. Here we describe the draft sequence of the 523-megabase genome of a *Musa acuminata* doubled-haploid genotype, providing a crucial stepping-stone for genetic improvement of banana. We detected three rounds of whole-genome duplications in the *Musa* lineage, independently of those previously described in the Poales lineage and the one we detected in the Arecales lineage. This first monocotyledon high-continuity whole-genome sequence reported outside Poales represents an essential bridge for comparative genome analysis in plants. As such, it clarifies commelinid-

sequence errors. The assembly consisted of 24,425 contigs and 7,513 scaffolds with a total length of 472.2 Mb, which represented 90% of the estimated DH-Pahang genome size. Ninety per cent of the assembly was in 647 scaffolds, and the N50 (the scaffold size above which 50% of the total length of the sequence assembly can be found) was 1.3 Mb (Supplementary Text and Supplementary Tables 1–3). We anchored 70% of the assembly (332 Mb) along the 11 *Musa* linkage groups of the Pahang genetic map. This corresponded to 258 scaffolds and included 98.0% of the scaffolds larger than 1 Mb and 92% of the annotated genes (Supplementary Text, Supplementary Table 4 and Supplementary Fig. 1).

We identified 36,542 protein-coding gene models in the *Musa* genome (Supplementary Tables 1 and 5). A total of 235 microRNAs from 37 families were identified, including only one of the eight microRNA gene (*MIR*) families found so far solely in Poaceae[8] (Supplementary Tables 6 and 7).

Viral sequences related to the banana streak virus (BSV) dsDNA plant pararetrovirus were found to be integrated in the Pahang genome, with 24 loci spanning 10 chromosomes (Supplementary Text and Supplementary Fig. 2). They belonged to a badnavirus phylogenetic group that differed from the endogenous BSV species (eBSV) found in *M. balbisiana*[9] and most of them formed a new

Nature 2012

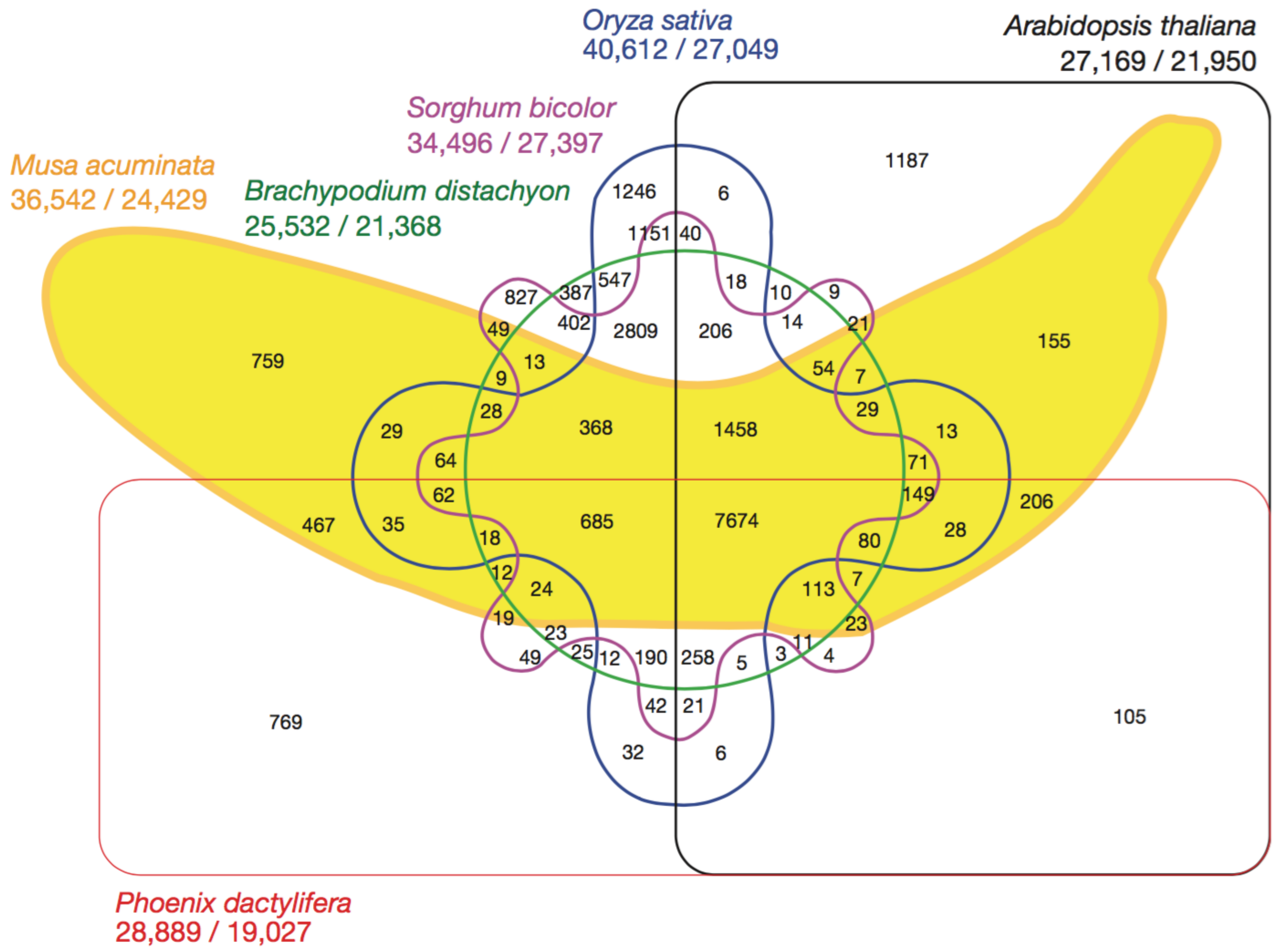
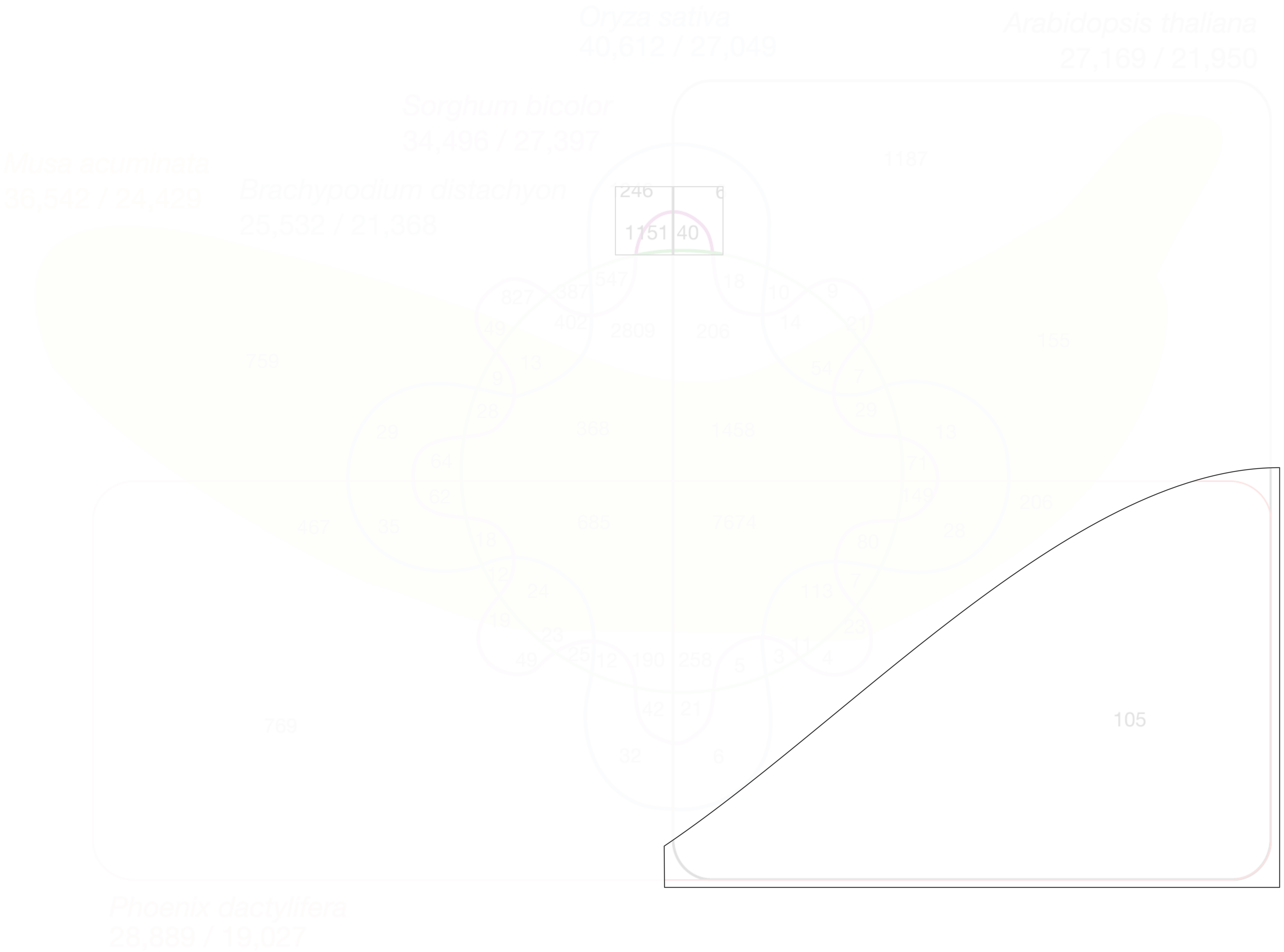
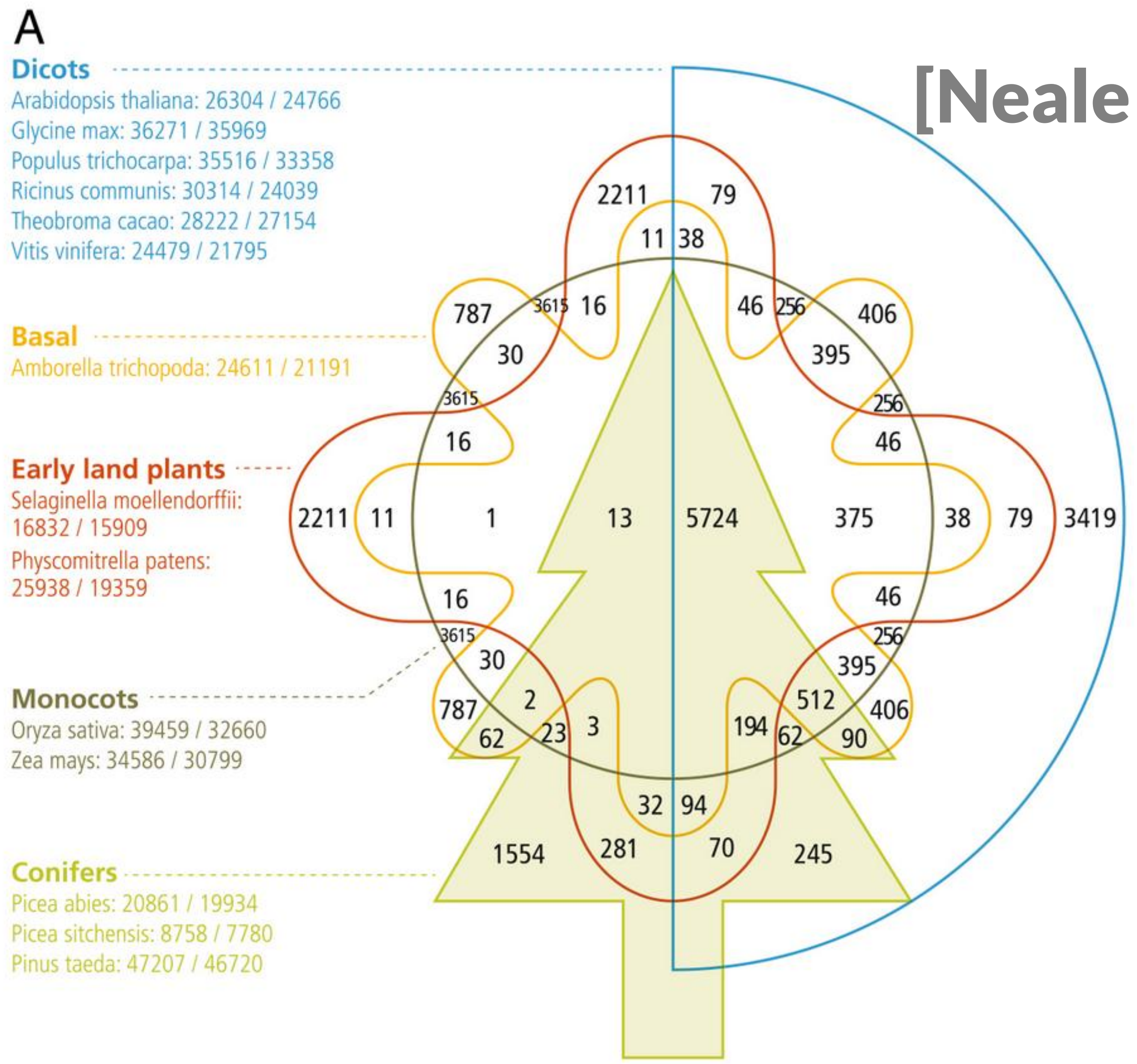


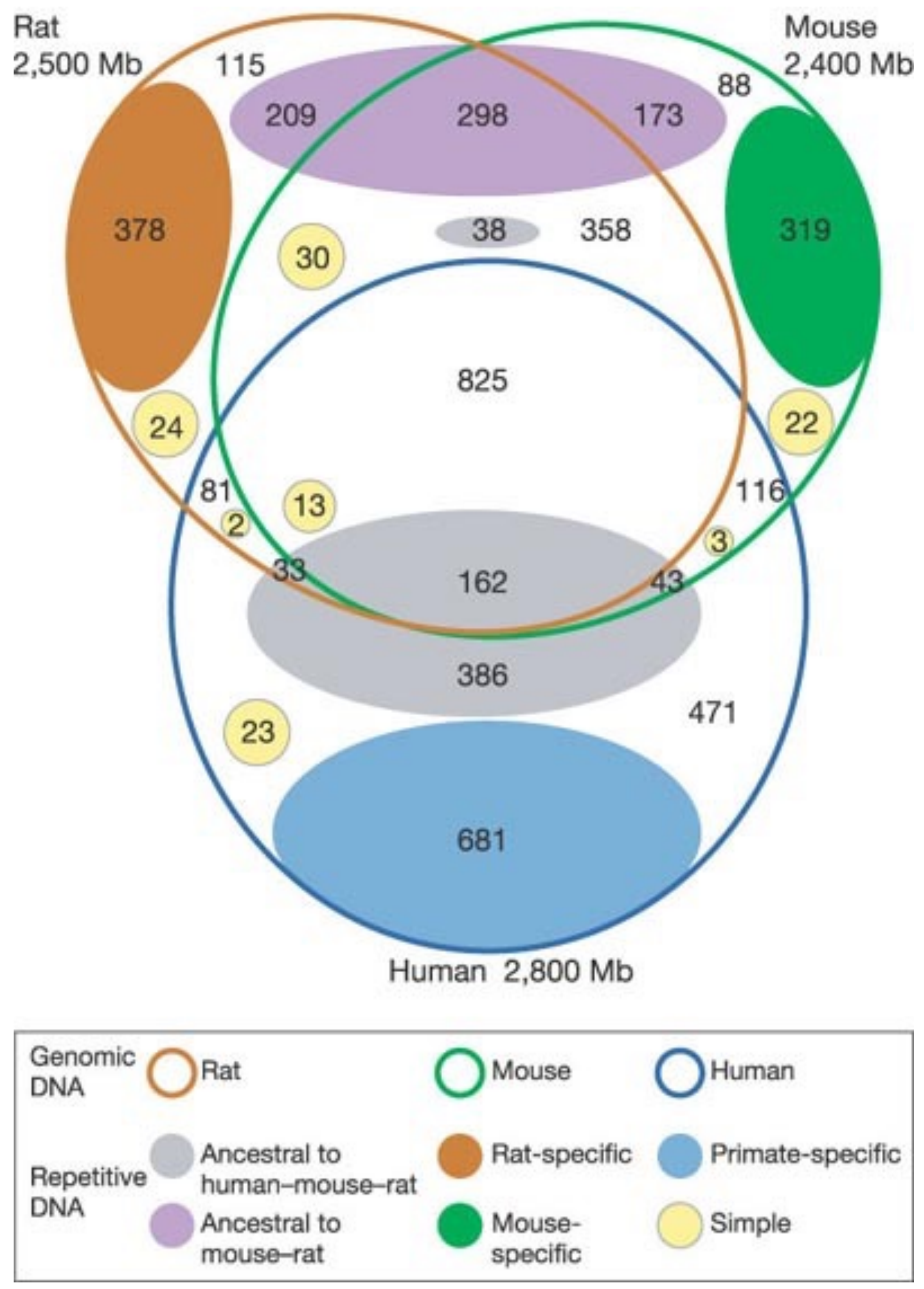
Figure 4 | Six-way Venn diagram showing the distribution of shared gene families (sequence clusters) among *M. acuminata, P. dactylifera, Arabidopsis thaliana, Oryza sativa, Sorghum bicolor* and *Brachypodium distachyon* genomes. Numbers of clusters are provided in the intersections. The total number of sequences for each species is provided under the species name (total number of sequences/total number of clustered sequences).



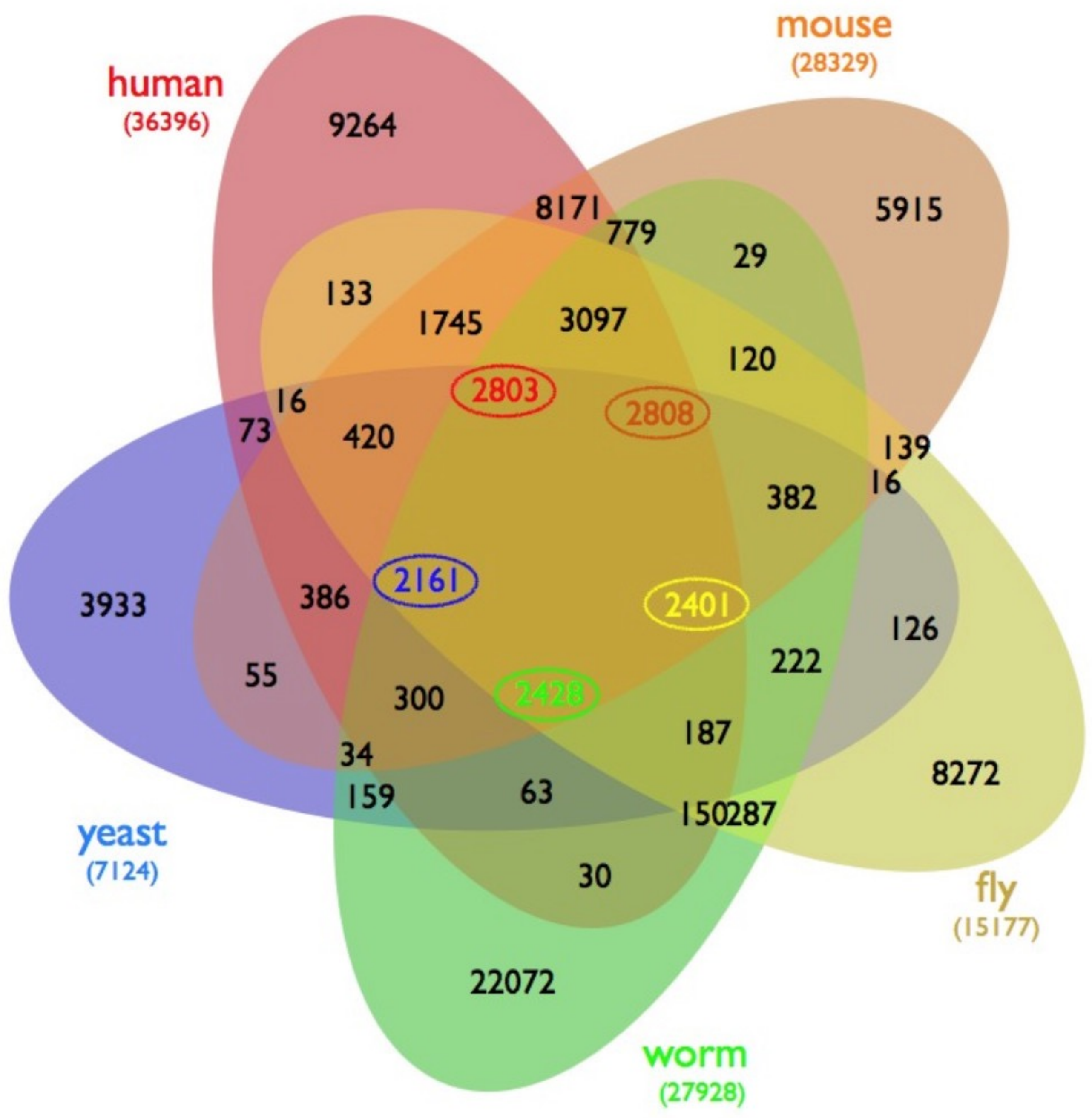
Oryza sativa
40,612 / 27,049
Arabidopsis thaliana
27,169 / 21,950
Sorghum bicolor
34,496 / 27,397
Musa acuminata
36,542 / 24,420
Brachypodium distachyon
25,532 / 21,368
Phoenix dactylifera
28,889 / 19,027
105



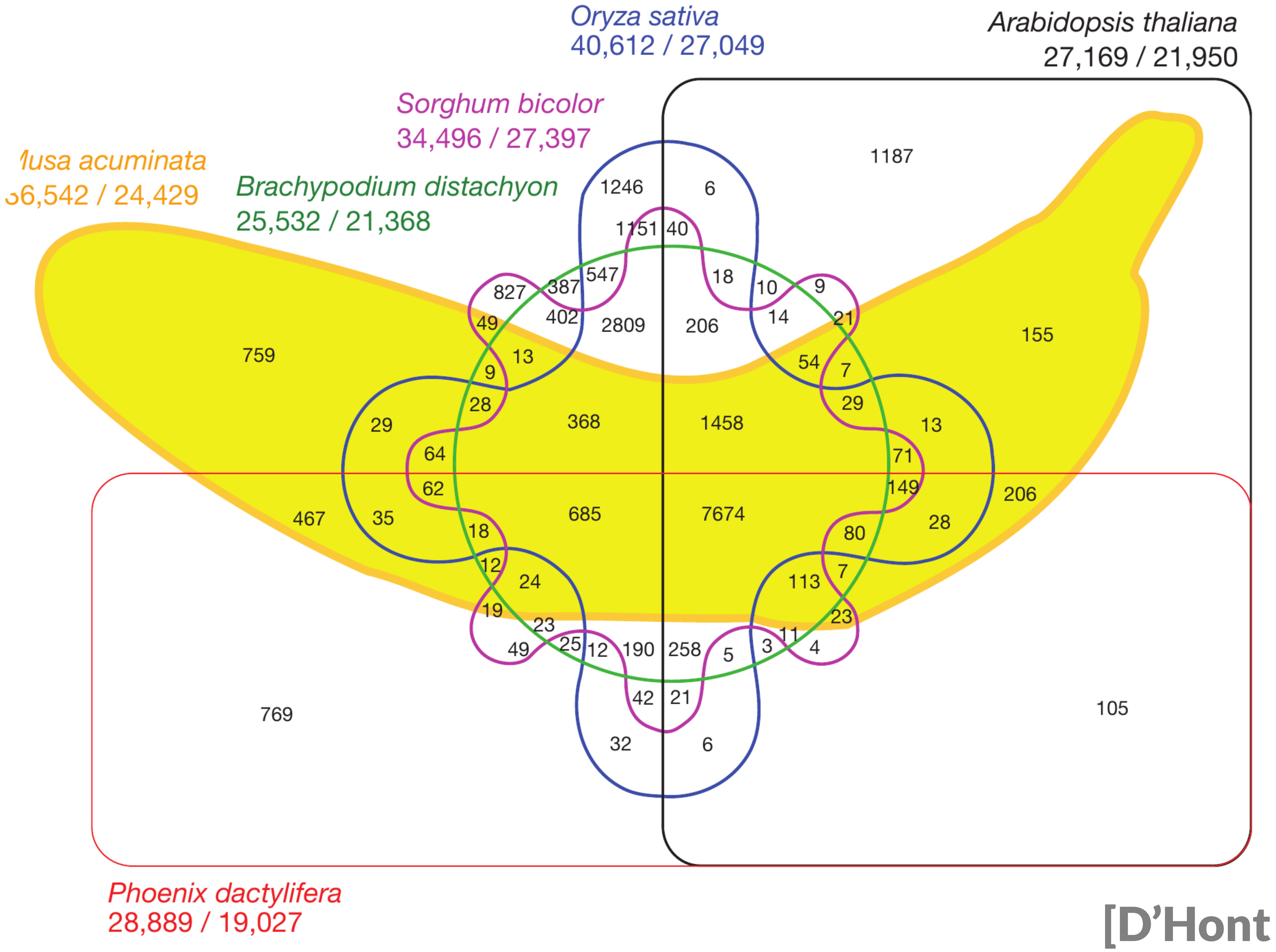
[Neale et al., BMC Genome Biology, 2014]



[Gibbs et al., Nature, 2004]



[Wiles et al., BMC Systems Biology]



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

Element ID	Sets	Attribute(s)
Name	**Characteristics**	**Age**
Lisa	School, Female	8
Bart	School, Male	10
Homer	Power Plant, Male	40
Mr. Burns	Evil, Power Plant, Male	90

What are some questions we'd like to ask?

Element ID	Sets	Attribute(s)
Name	**Characteristics**	**Age**
Lisa	School, Female	8
Bart	School, Male	10
Homer	Power Plant, Male	40
Mr. Burns	Evil, Power Plant, Male	90

1. Don’t always try to show all individuals
2. What is the biggest intersection?
3. Which sets make up an intersection?
4. How big is an intersection?
5. Does it work for more than four sets?

Design Workshop

work in groups

get to know the data (5 mins)

create three (rapid!) prototypes (3x10 mins)

Write up your two favorites (15 mins) in google docs

Upload to “Bonus” Canvas Dropbox by 4pm

Set Visualization

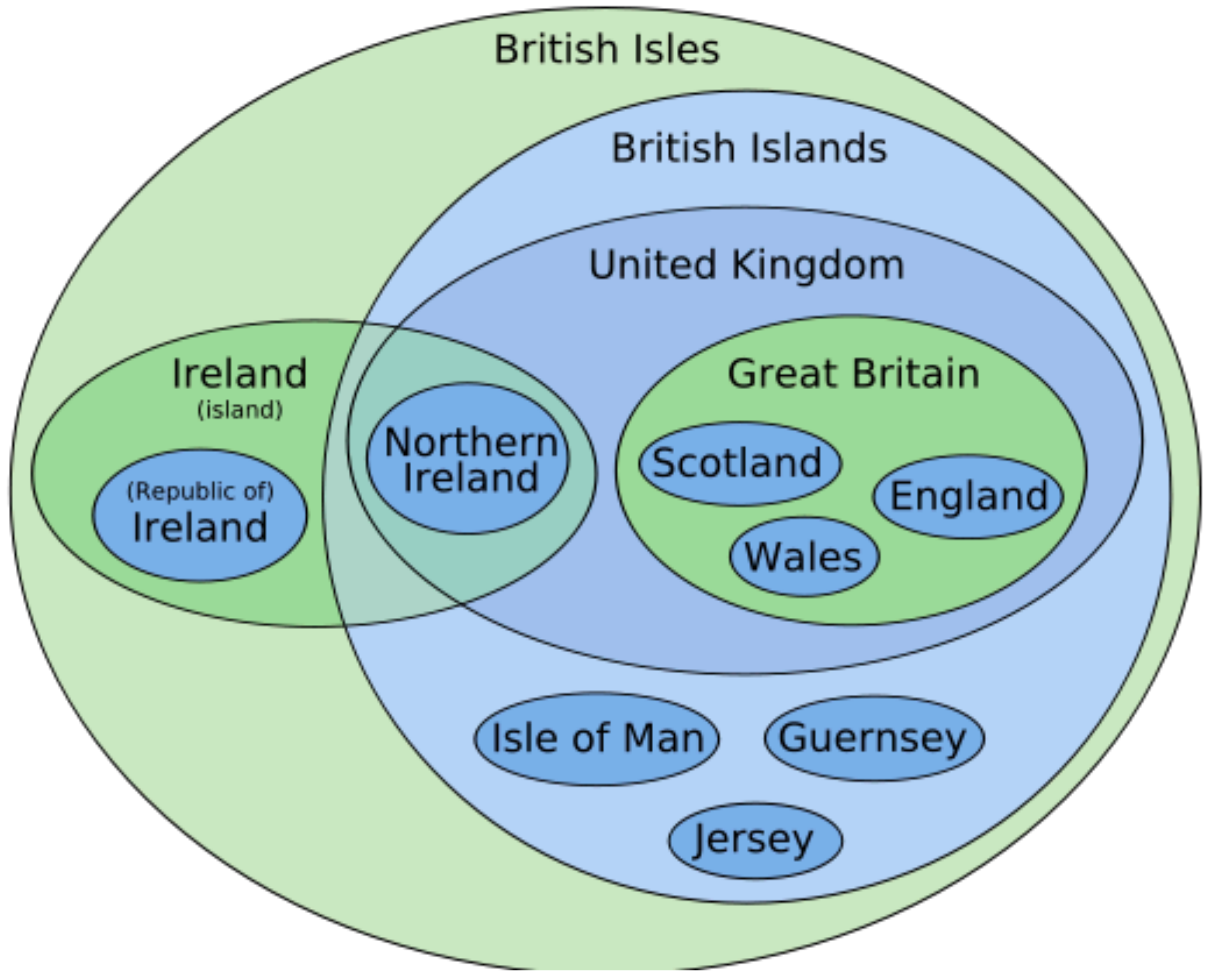
Venn and Euler Diagrams

Venn vs Euler

Euler Diagram

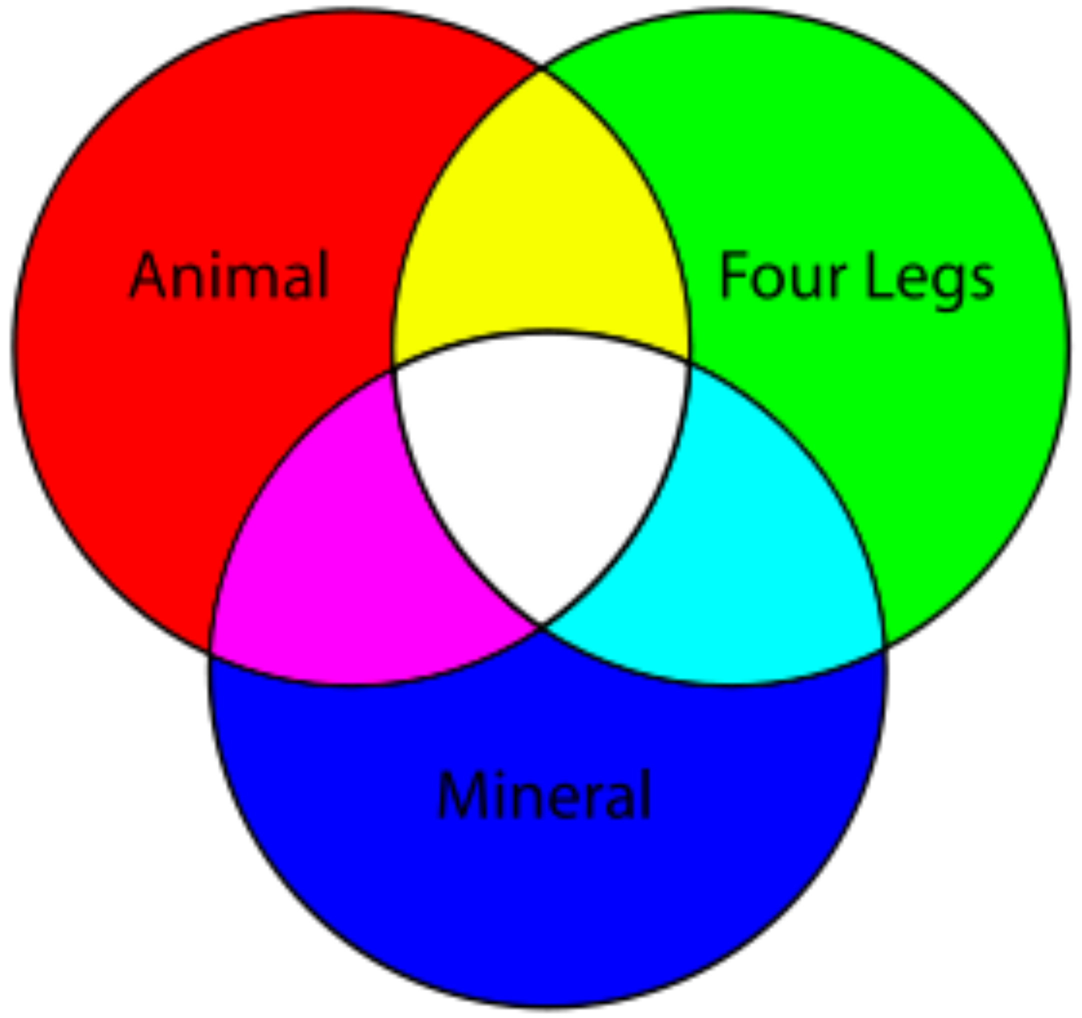
Shows logical relations

May omit empty intersections



Venn Diagram

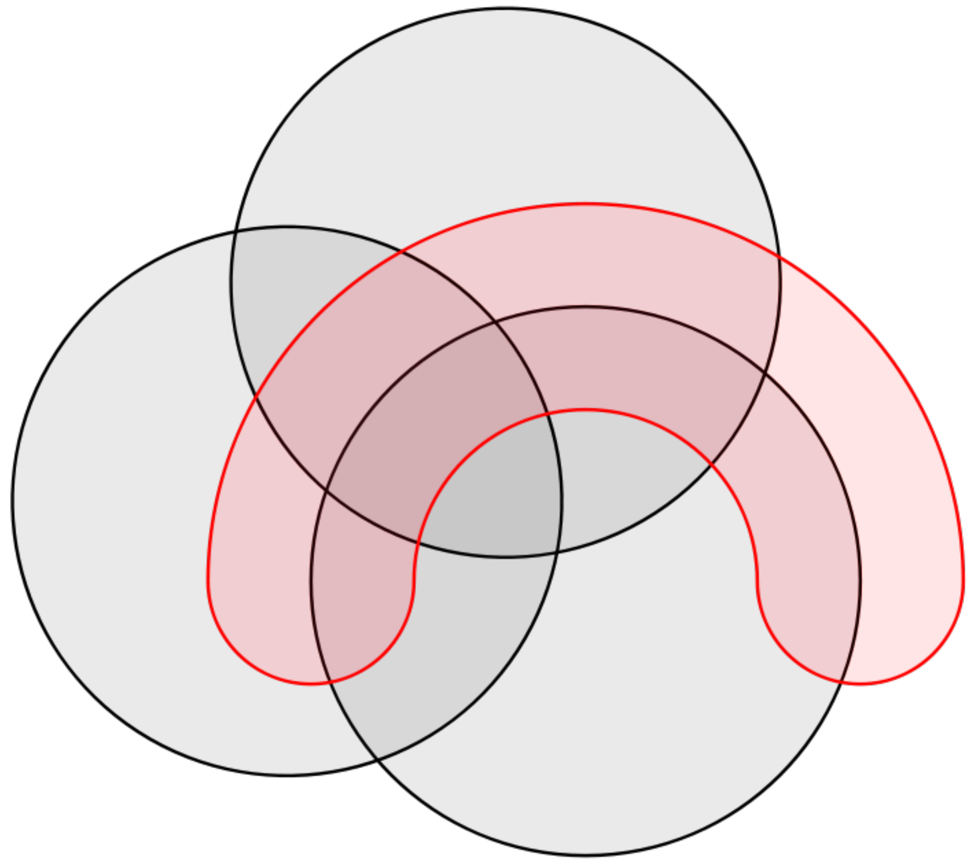
Shows all possible logical relations between sets (even if empty)

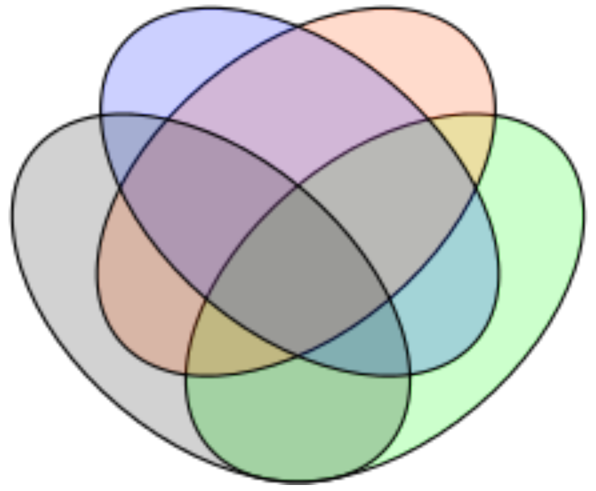


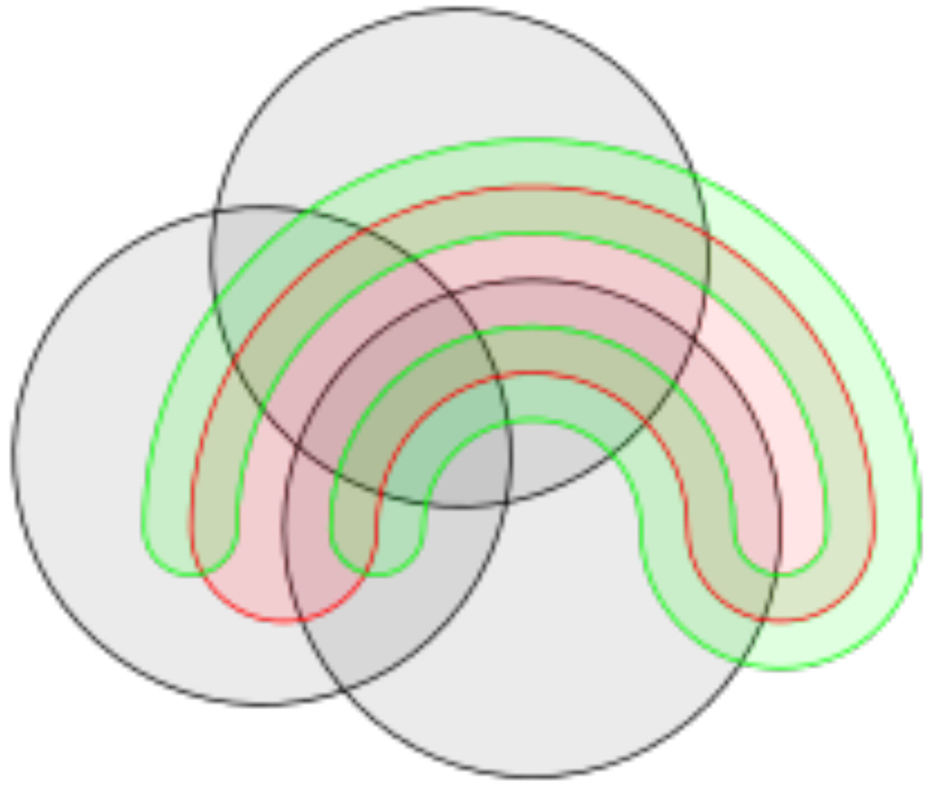
Venn Diagrams

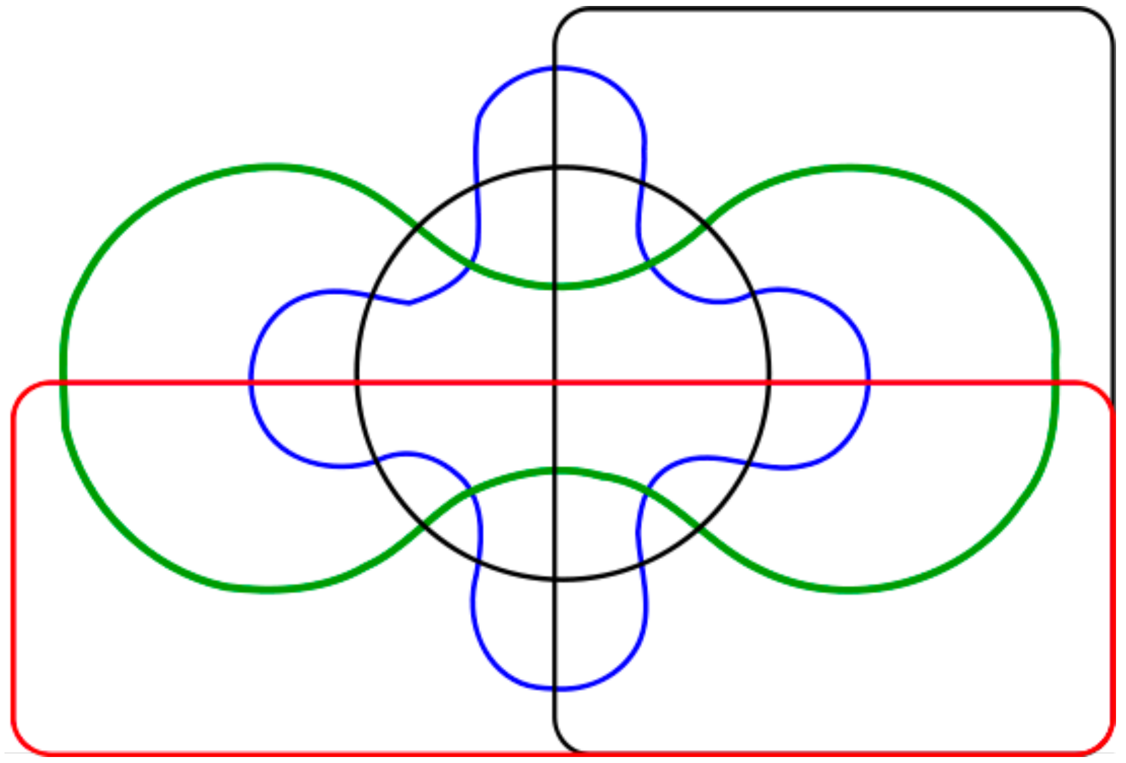
Venn diagrams for many sets are hard

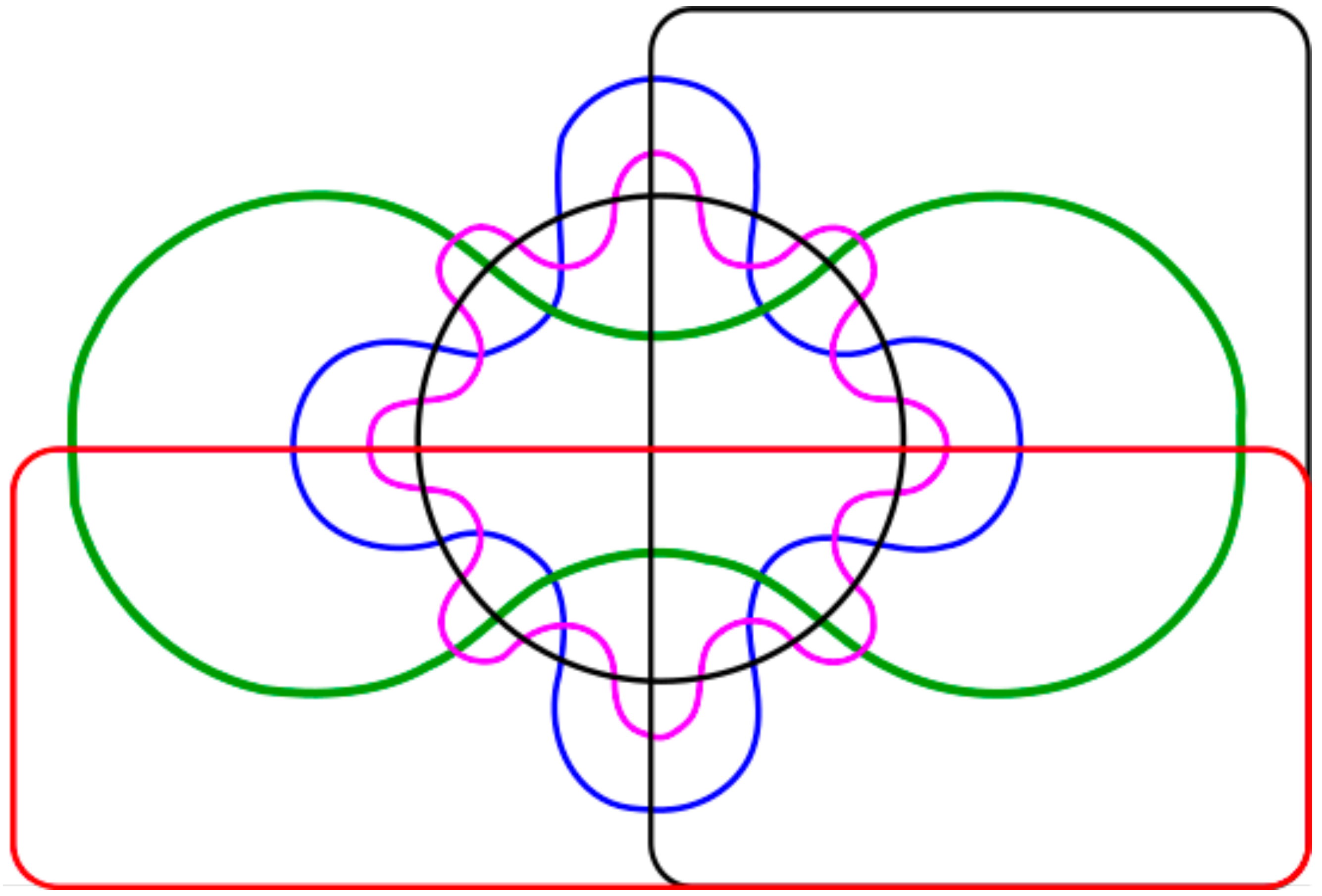
# of intersections is 2^n

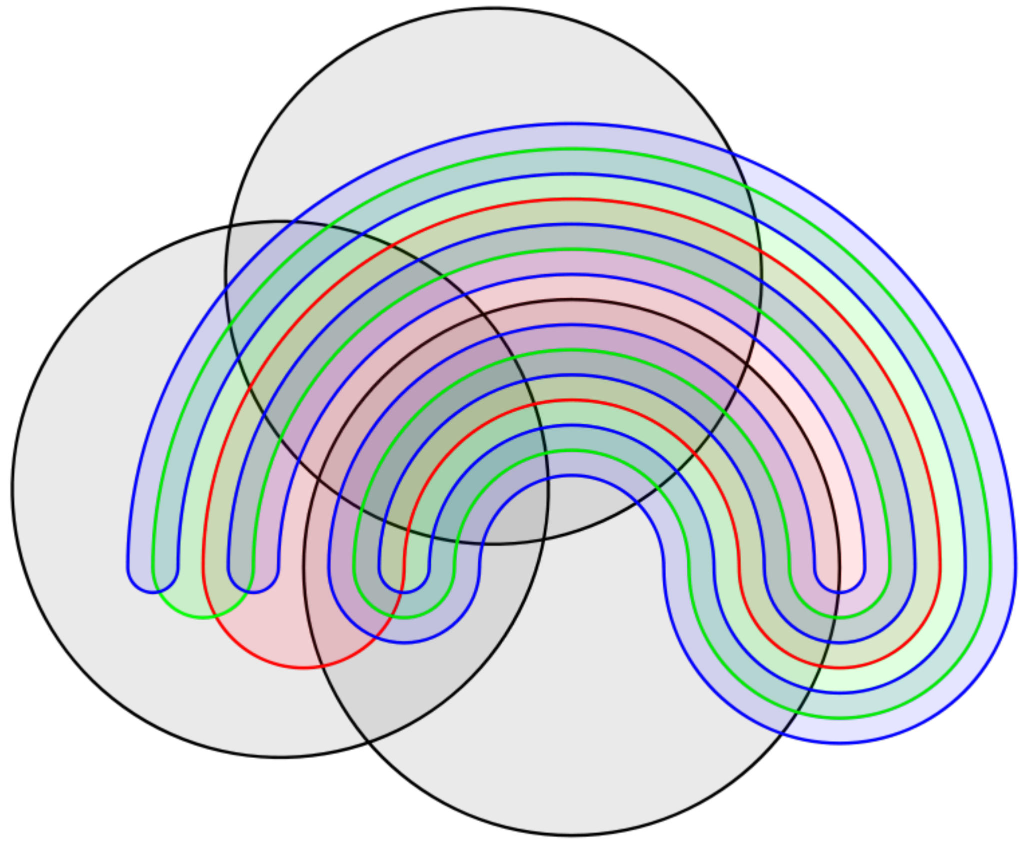


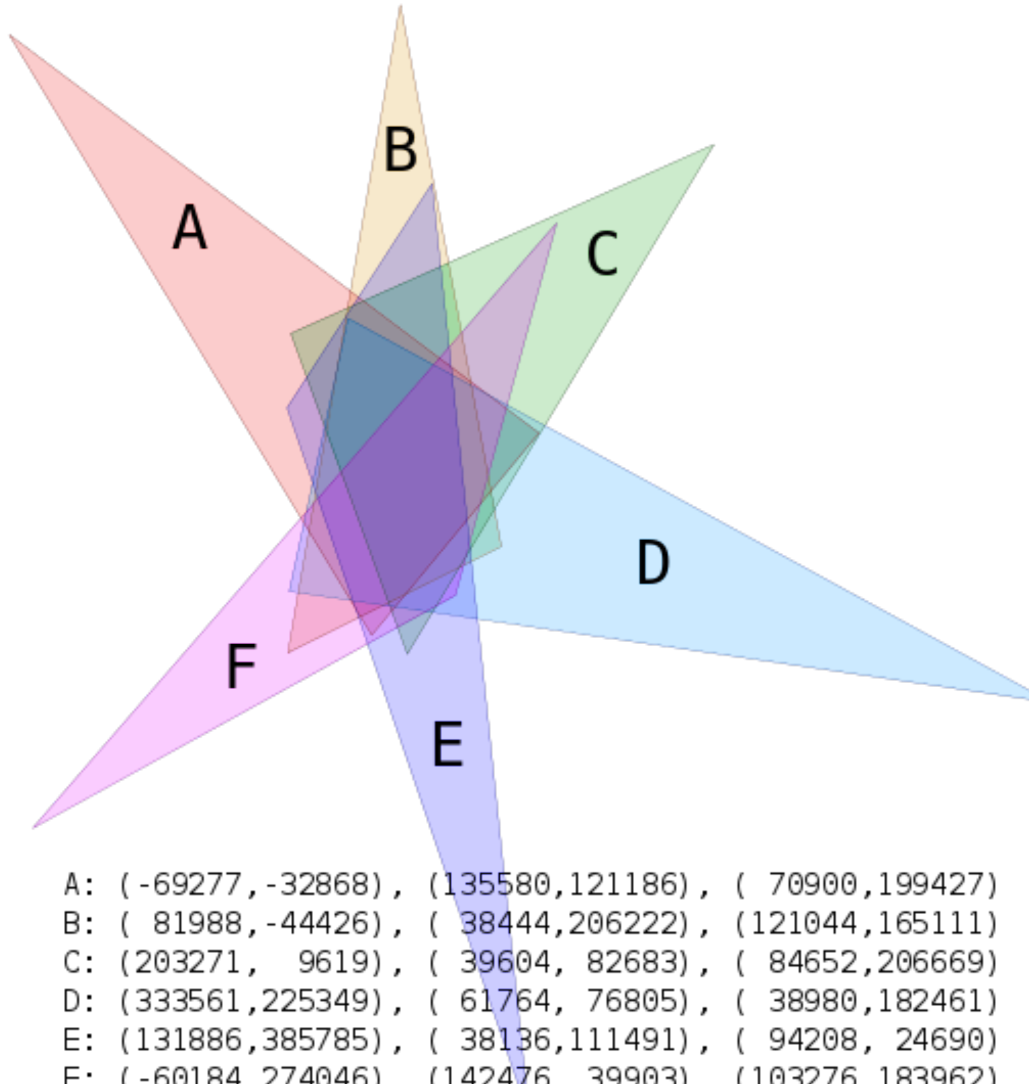












https://en.wikipedia.org/wiki/Venn_diagram

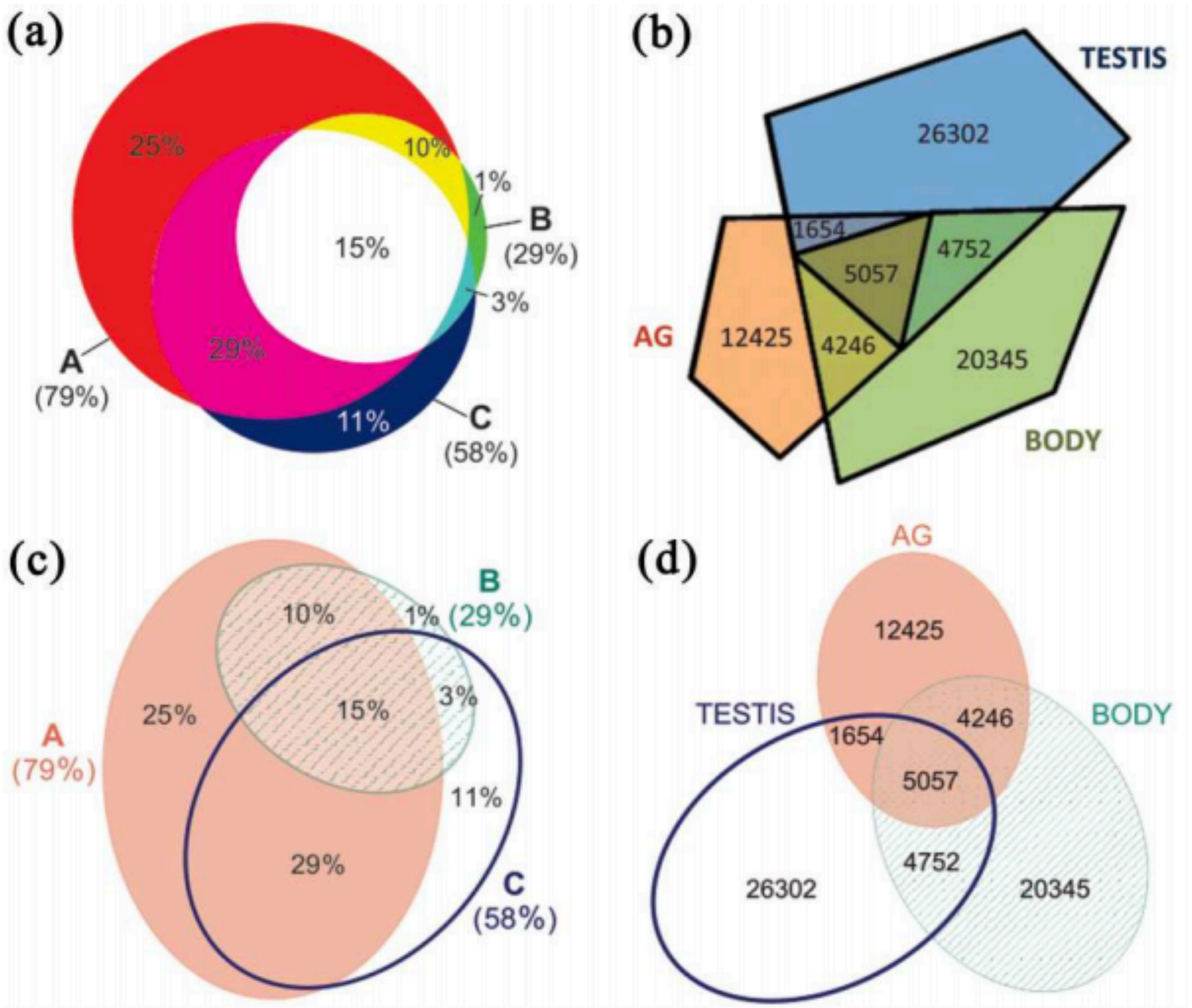
Area-Proportional Euler Diagrams

Problem with Venn: size doesn’t correspond to the data.

Creating area-proportional Euler diagrams is hard.

Layout criteria:

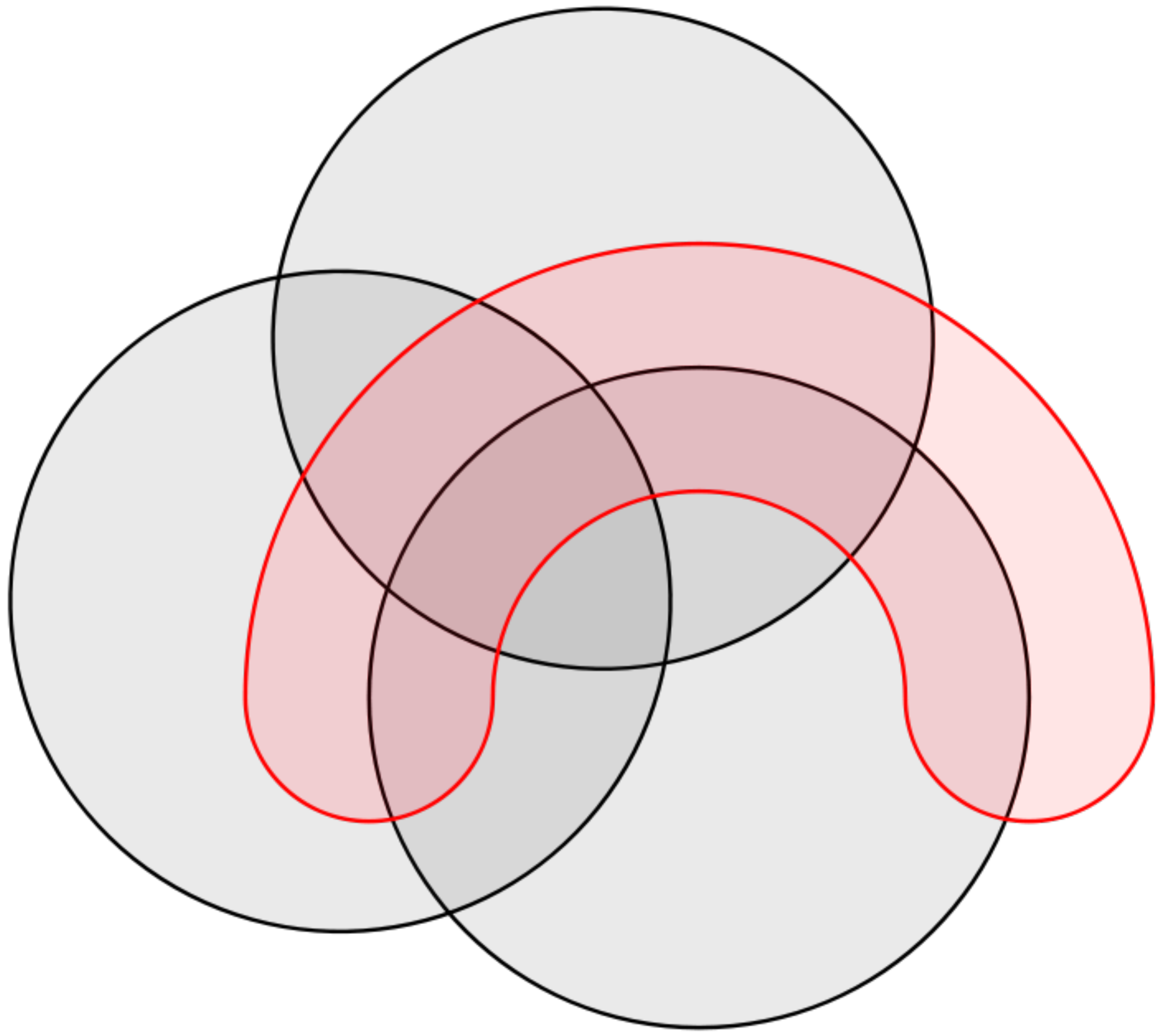
- simple curves (circles are best)
  - makes it easy to identify which sets are participating in intersection
  - Gestalt-principle: good continuation
- area proportional



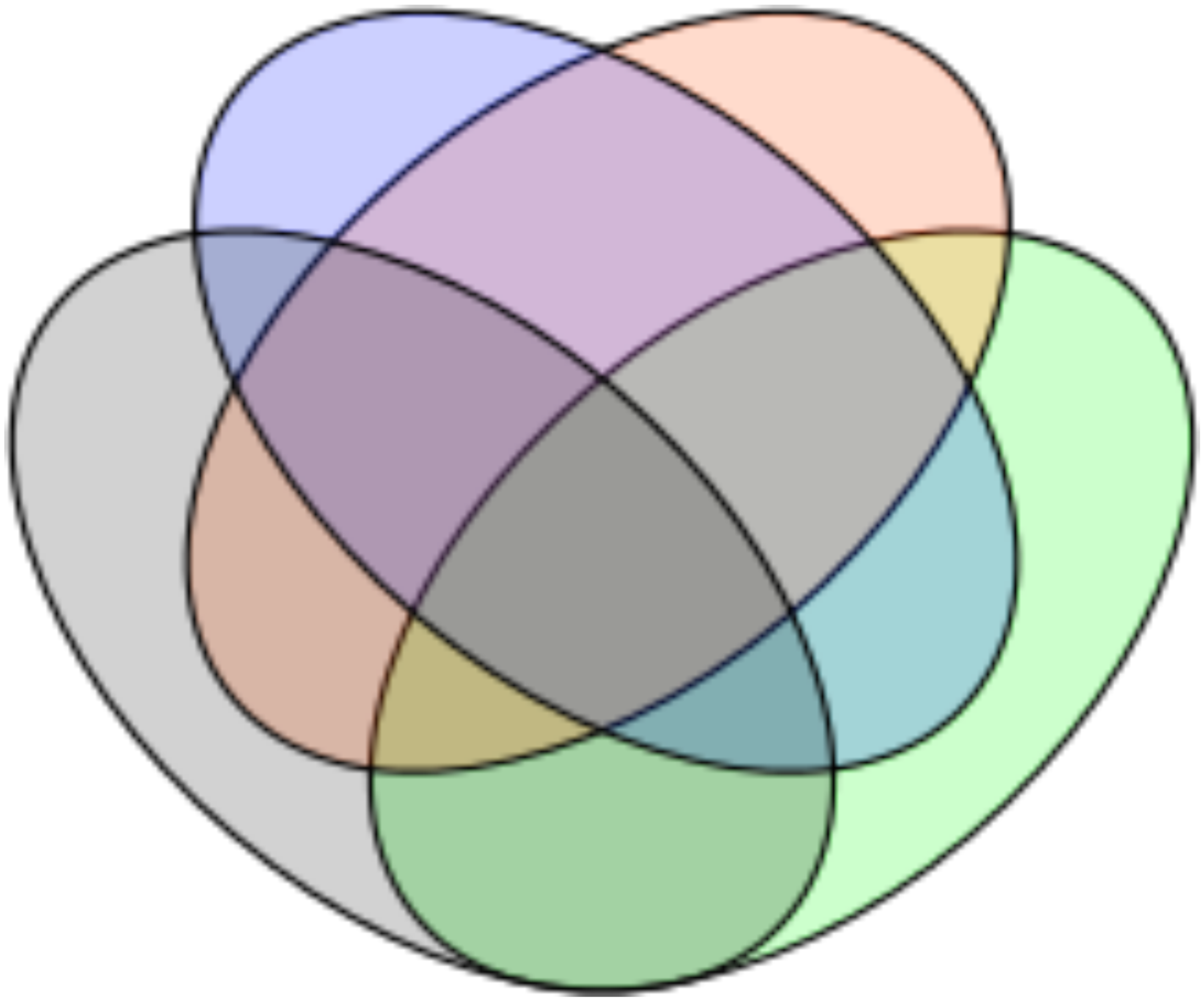
[Alsallakh 2015]

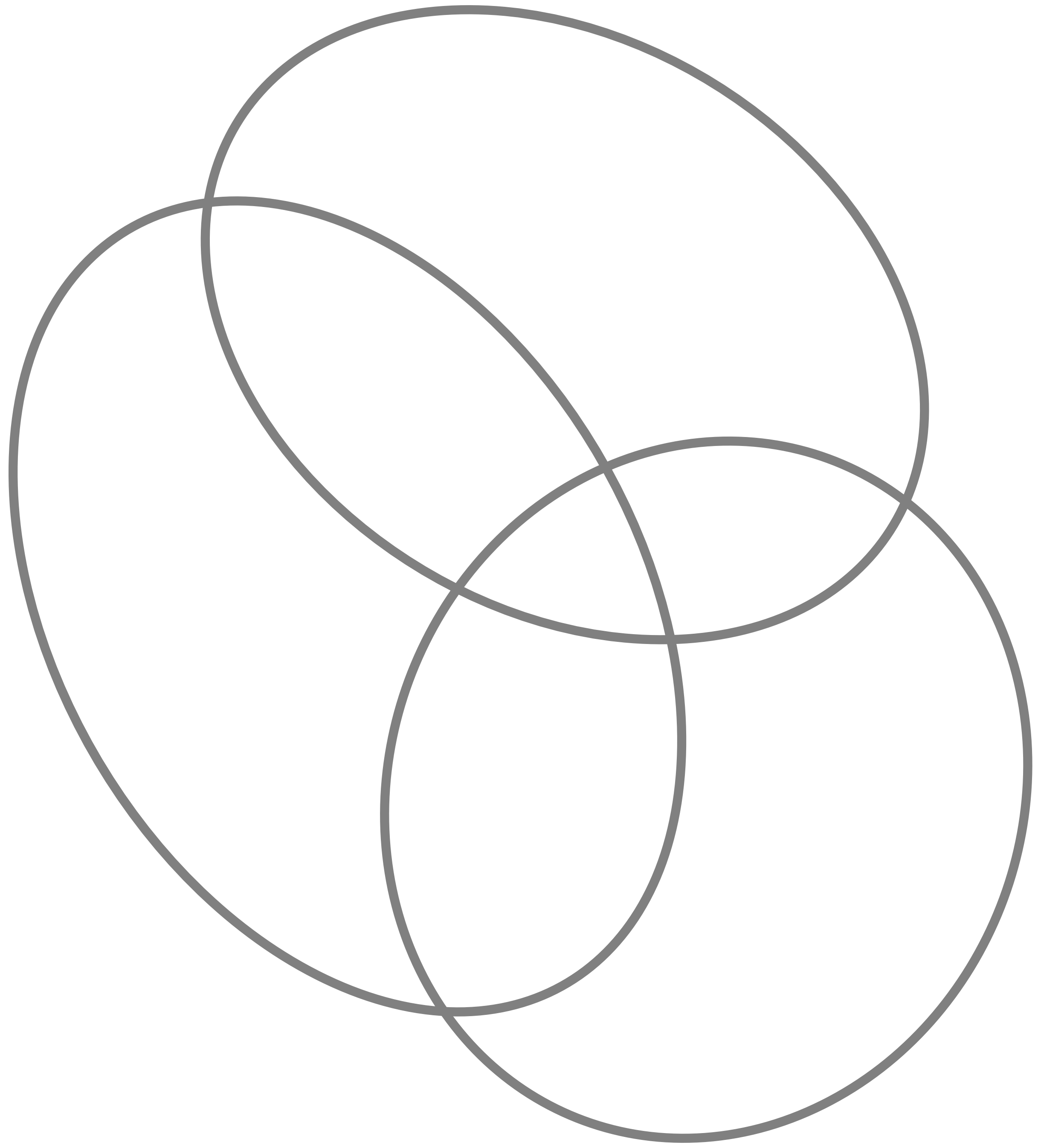
Compare Simple vs Complex Shape

Complex



Simple





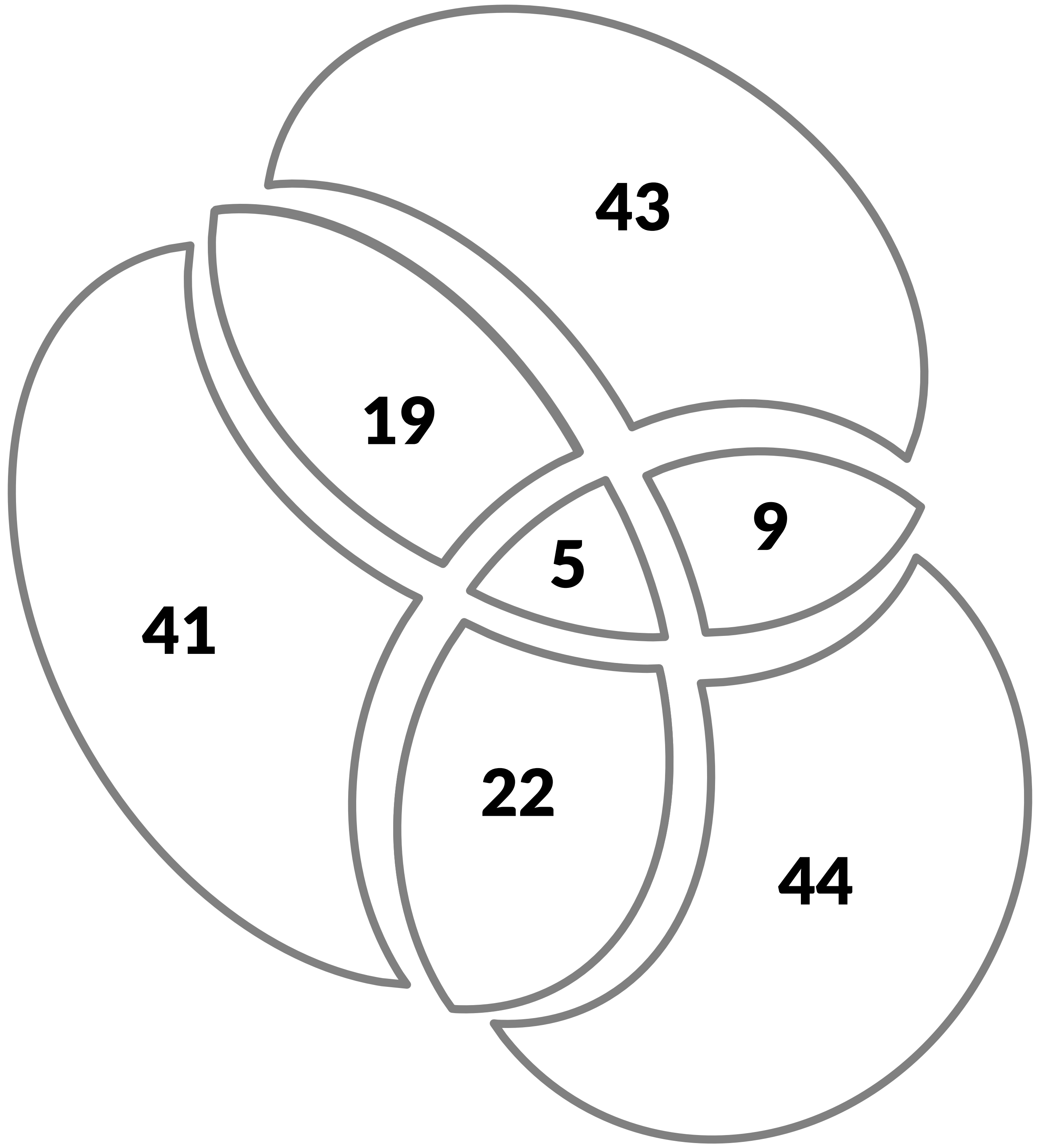
[created with EulerAPE]

?

>
<

22

19



[created with EulerAPE]

Venn-Euler Pros/Cons

Pros

Familiar

Intuitive

Work well for 2-4 sets

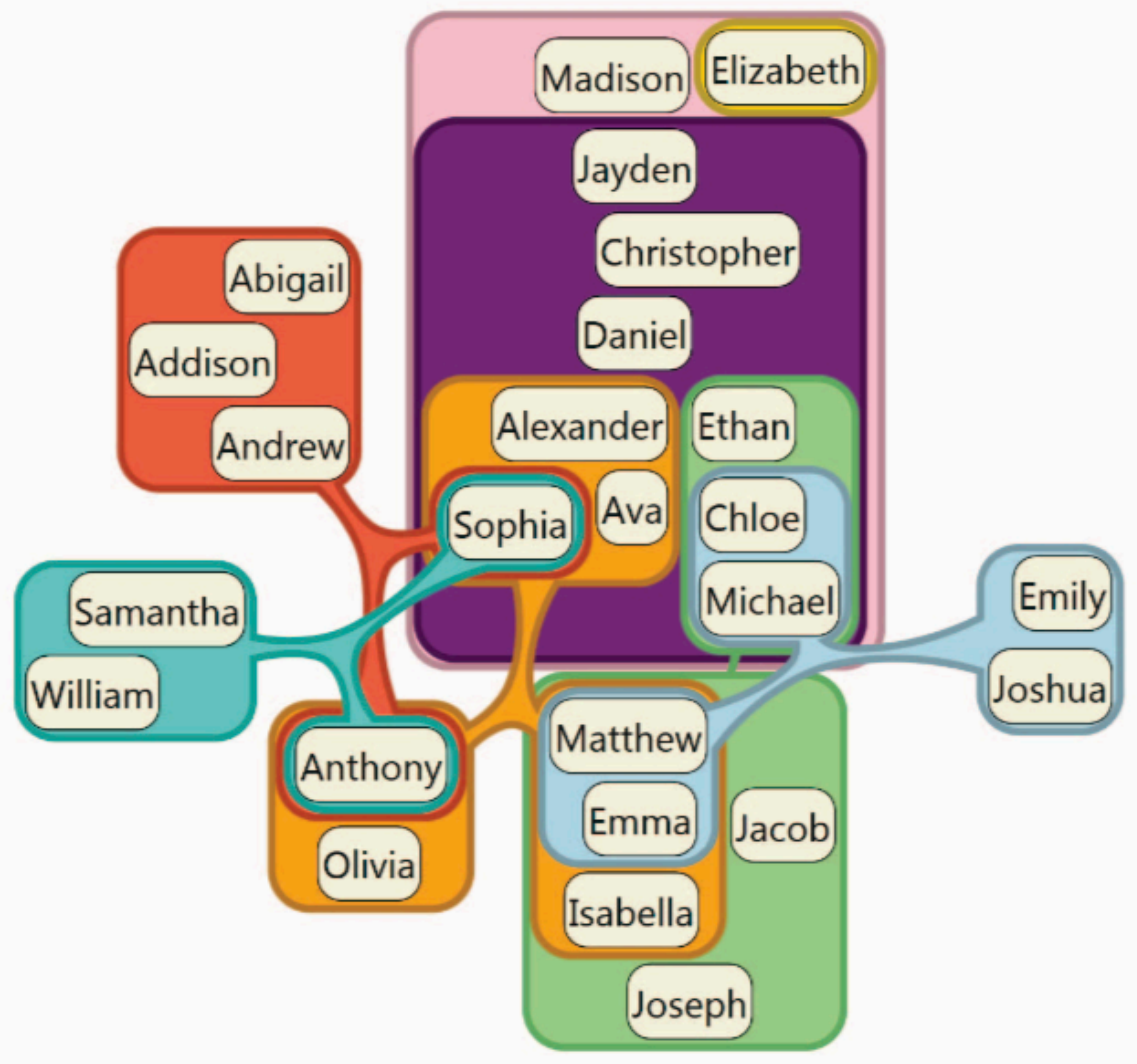
Cons

Don’t work well for more than 4 sets

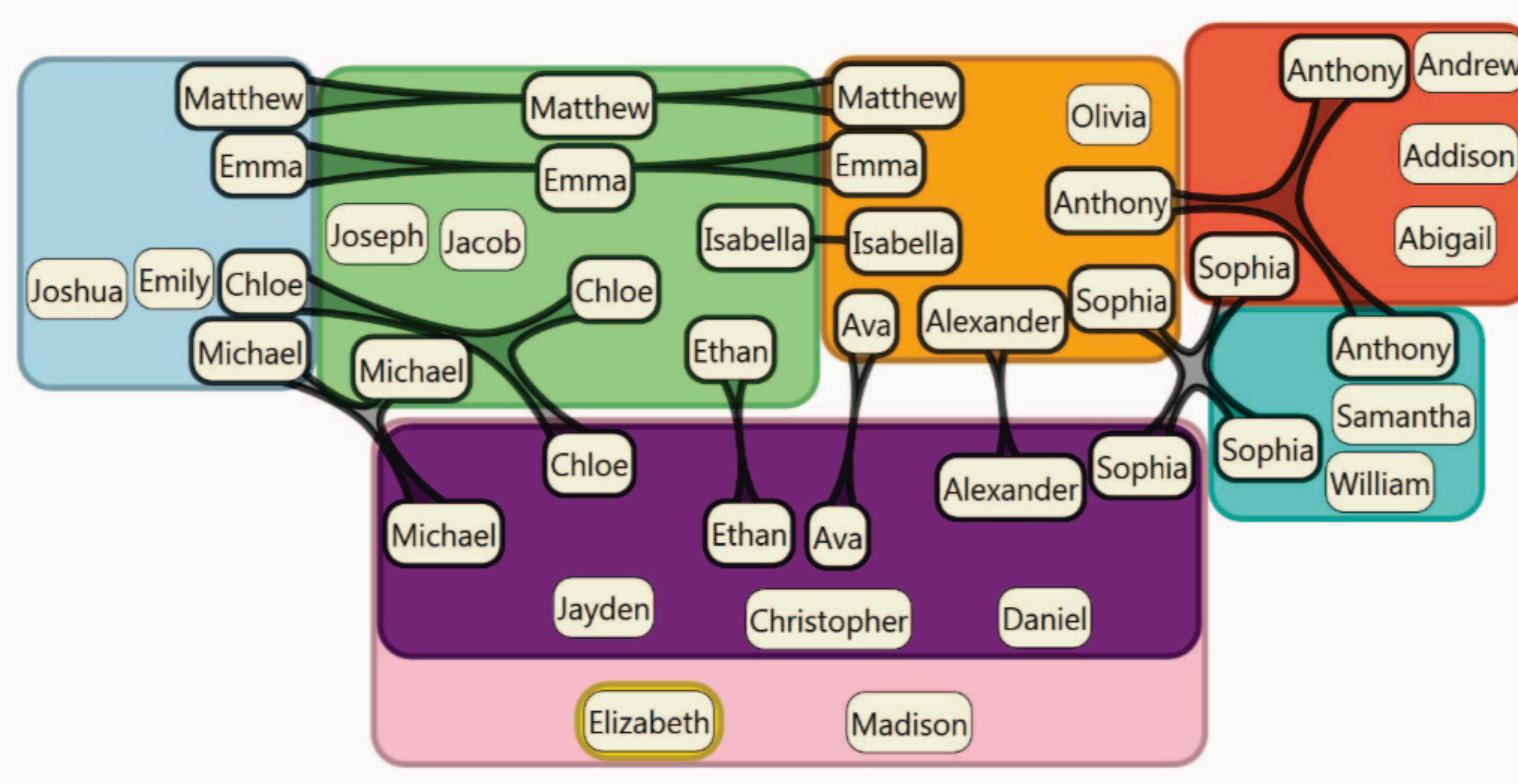
Area proportional hard to do

Not well suited to show attributes

Relationships for specific Items

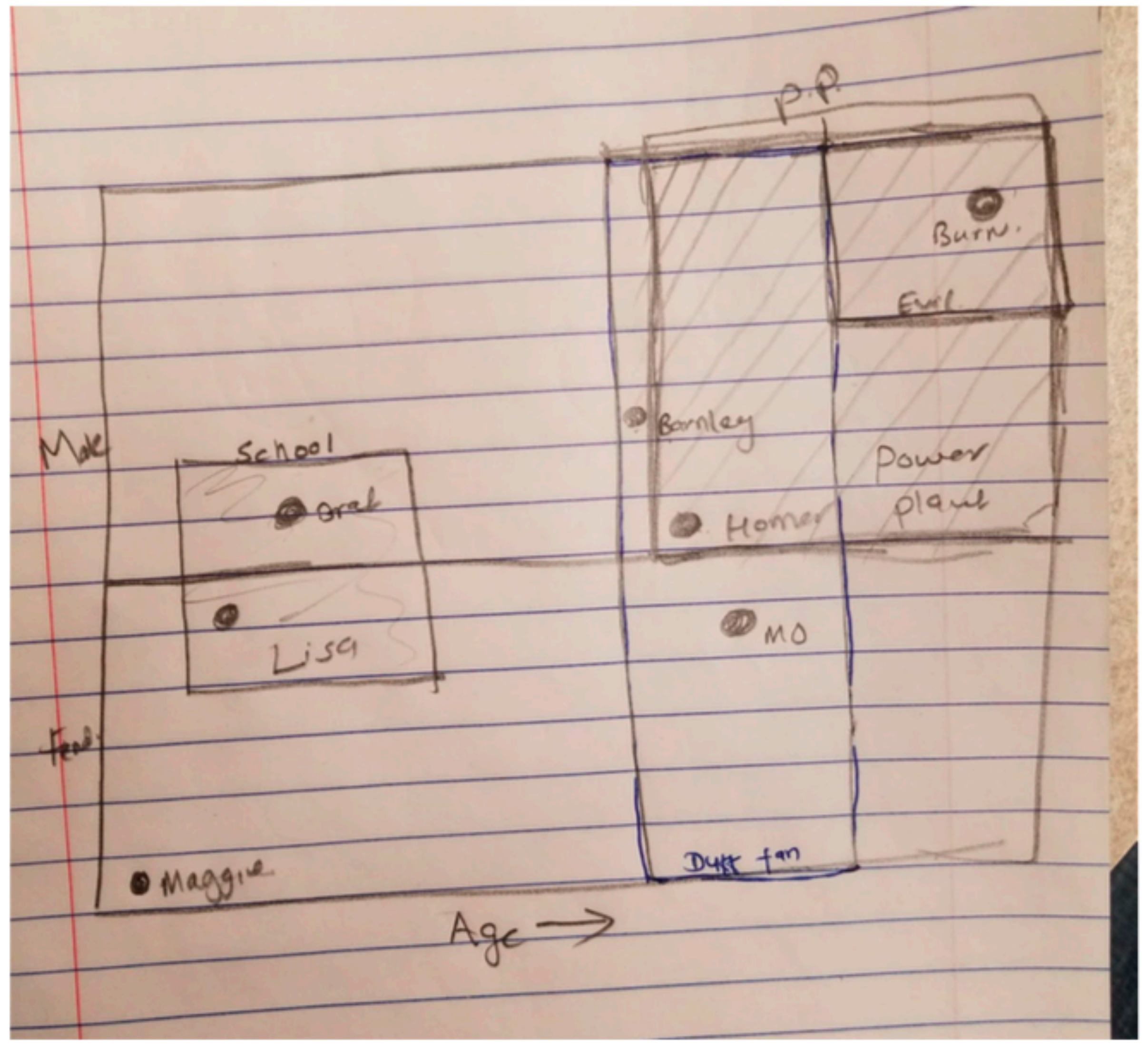


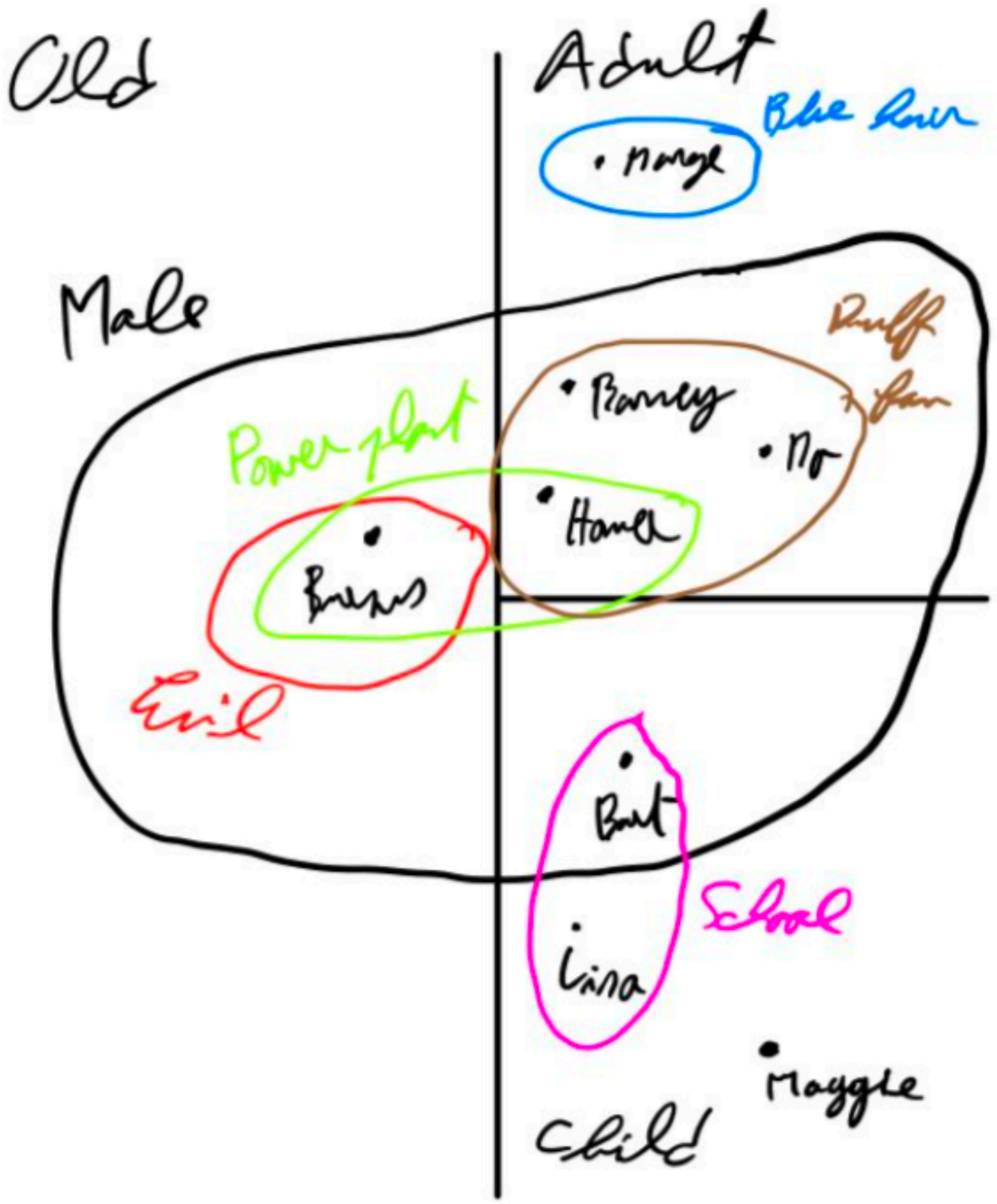
No Duplicate Nodes
Complex Shapes
Notice the Nesting

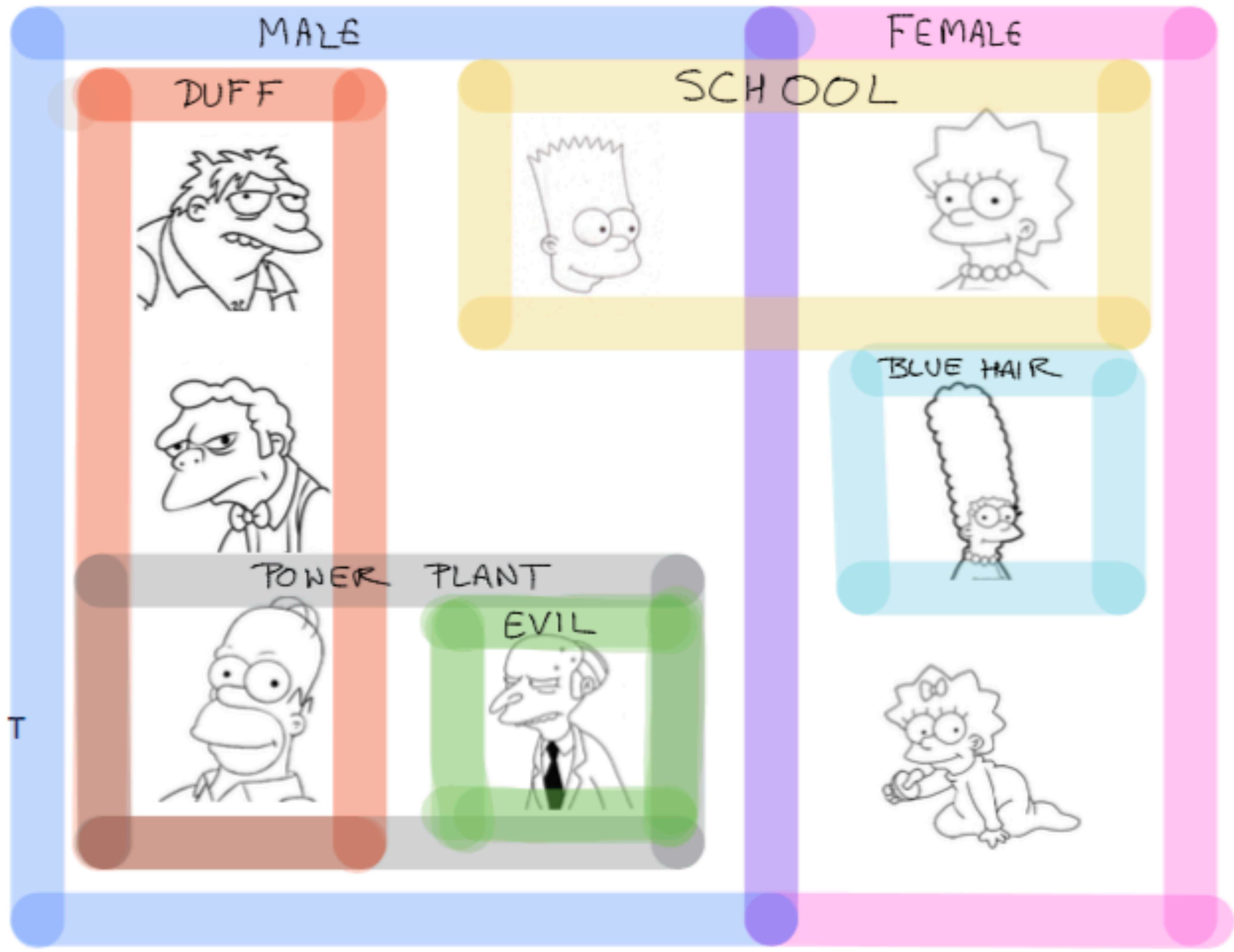


Duplicate Nodes
Simple Shapes

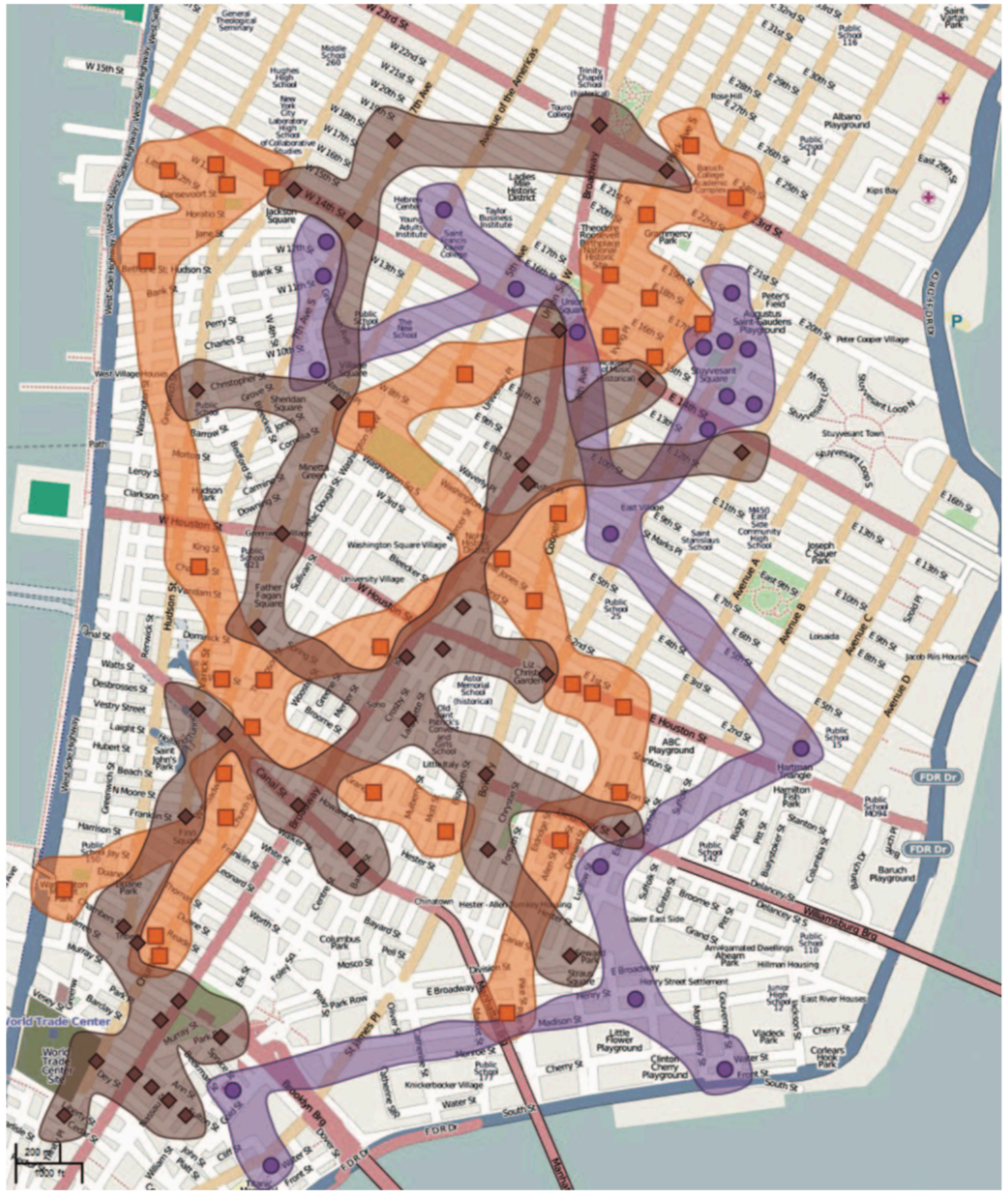
[Riche 2010]

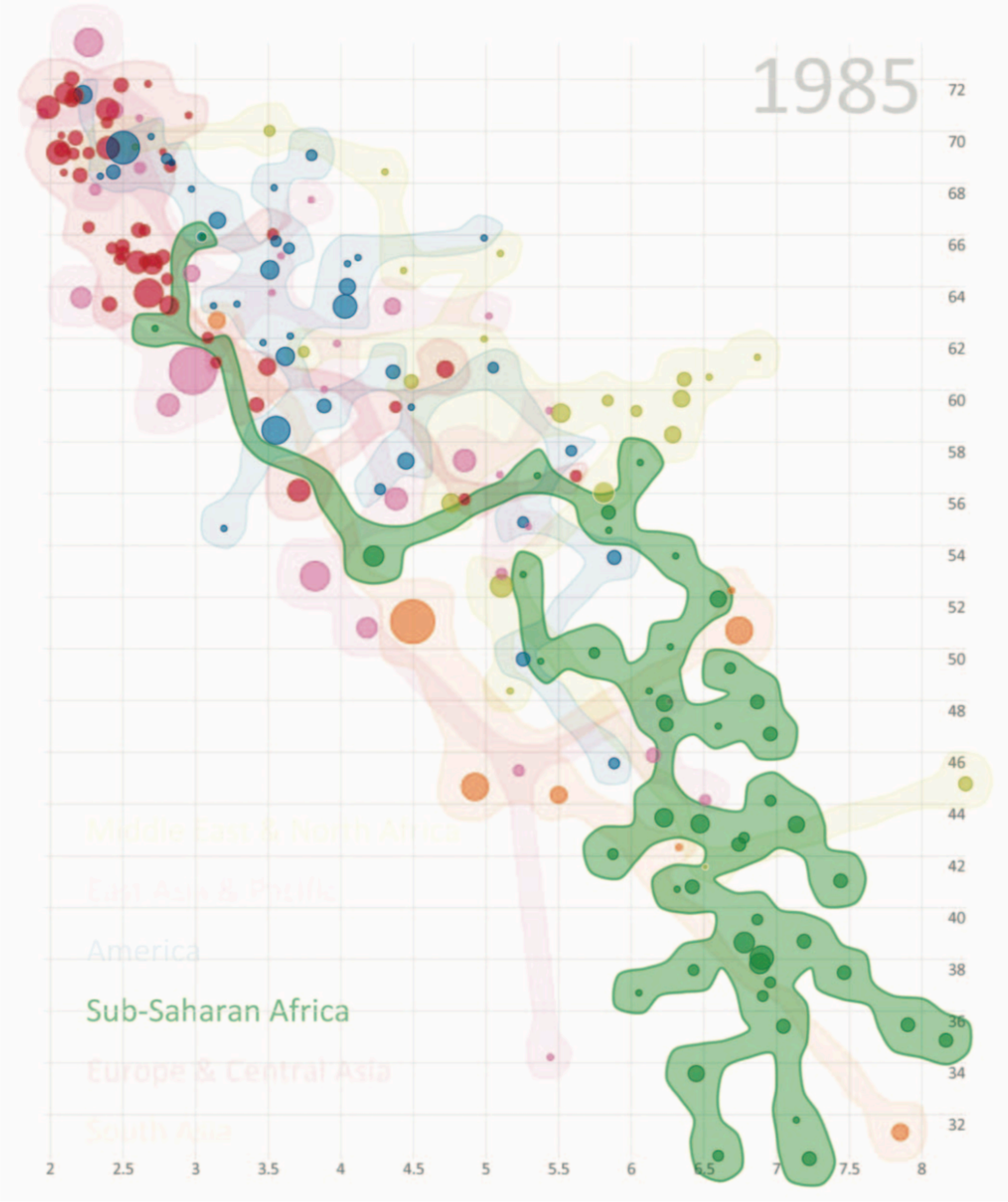

PP
Burn
Evil
Barnley
Power plant
Homer
MO
School
Lisa
Male
Maggie
Duff fan
Age →


Old
Adult
Blue hair
Marge
Male
Duff fan
Barney
Mo
Power plant
Homer
Burns
Evil
Bart
School
Lisa
Maggie
Child


MALE
FEMALE
DUFF
SCHOOL
BLUE HAIR
POWER PLANT
EVIL

Sets on top of a fixed layout



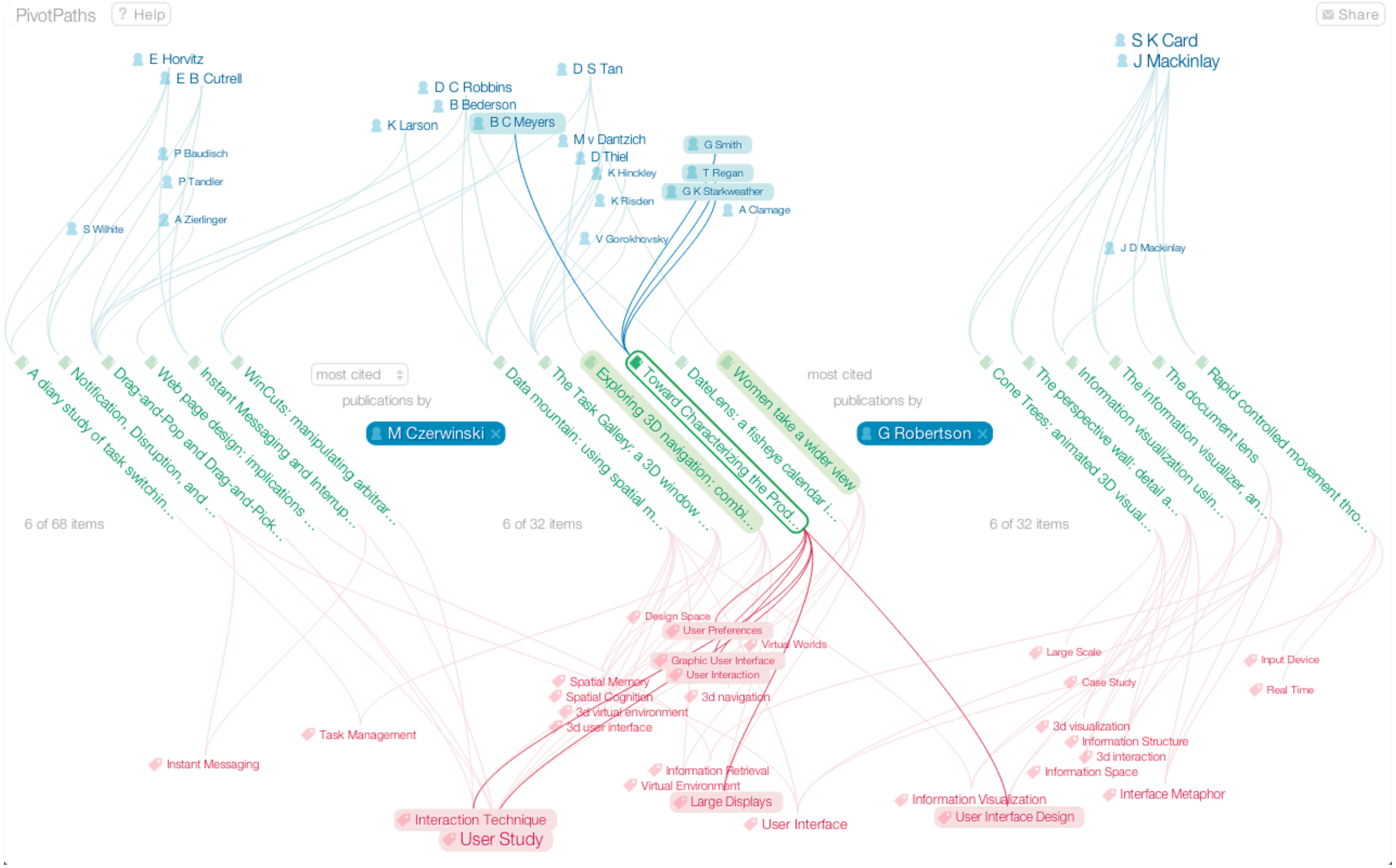


https://www.youtube.com/watch?v=Ju2hSThmPWA

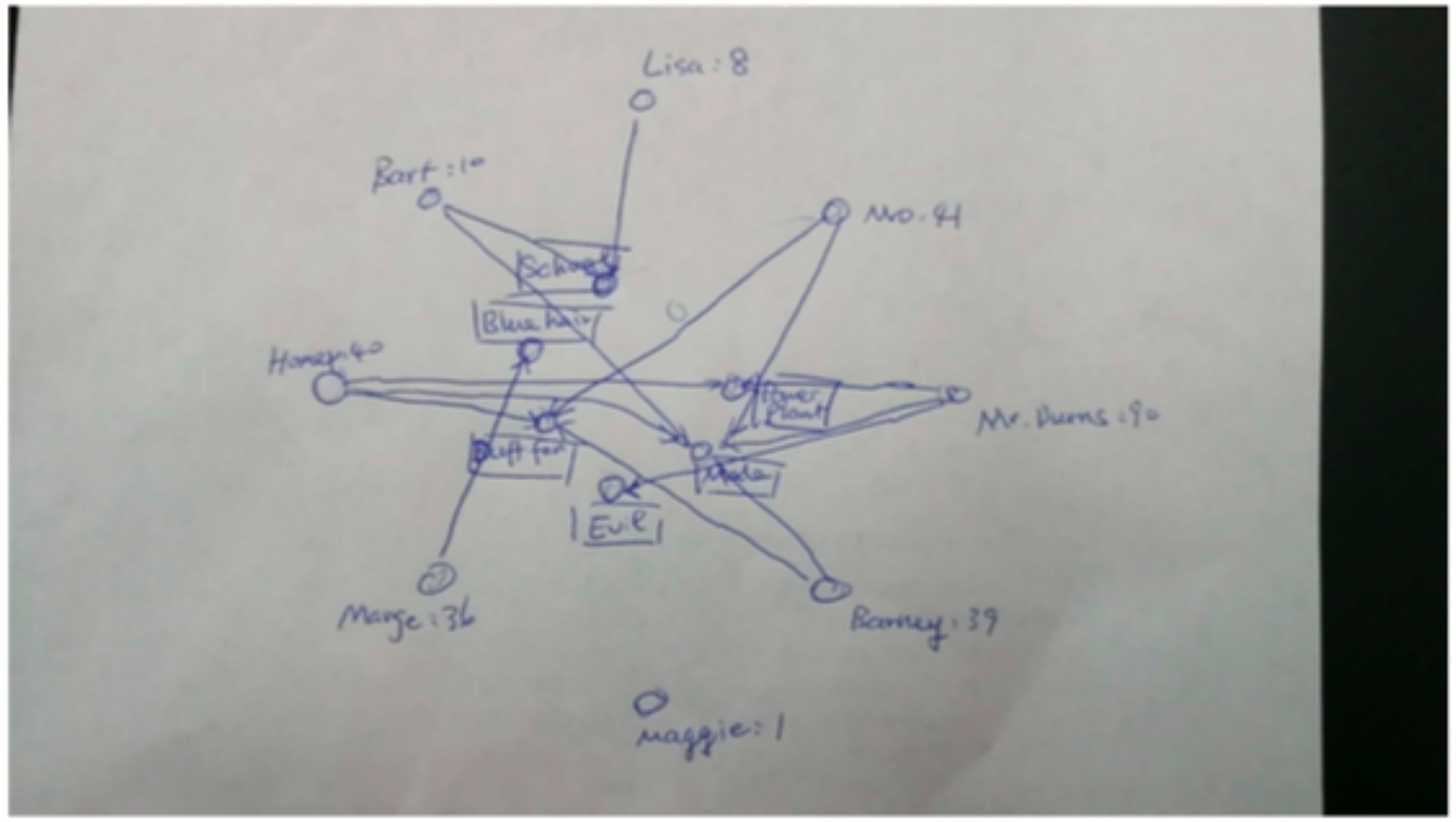
Node-Link Techniques

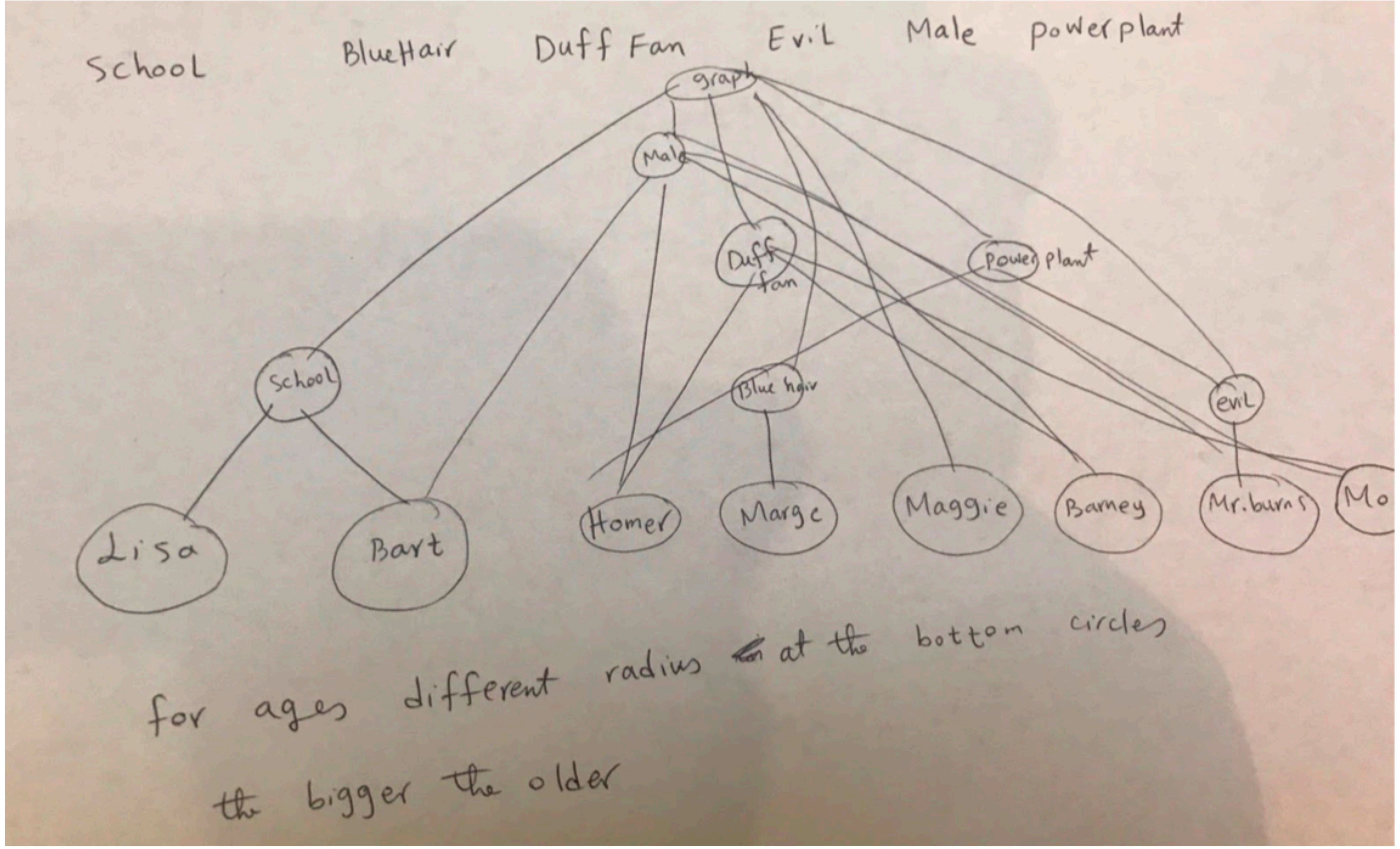
Treat sets as nodes

Connect to elements that are in set



http://mariandoerk.de/pivotpaths/demo/#/1:0_497686


Lisa: 8
Bart: 10
Homer: 40
Mr. Burns: 90
Marge: 36
Barney: 39
maggie: 1
School
Blue hair
Duff fan
Evil


School
BlueHair
Duff Fan
Evil
Male
powerplant
graph
Male
Duff fan
Power plant
School
Blue hair
evil
Lisa
Bart
Homer
Marge
Maggie
Barney
Mr. burns
Mo
for ages different radius at the bottom circles
the bigger the older

Showing Pairwise Overlap

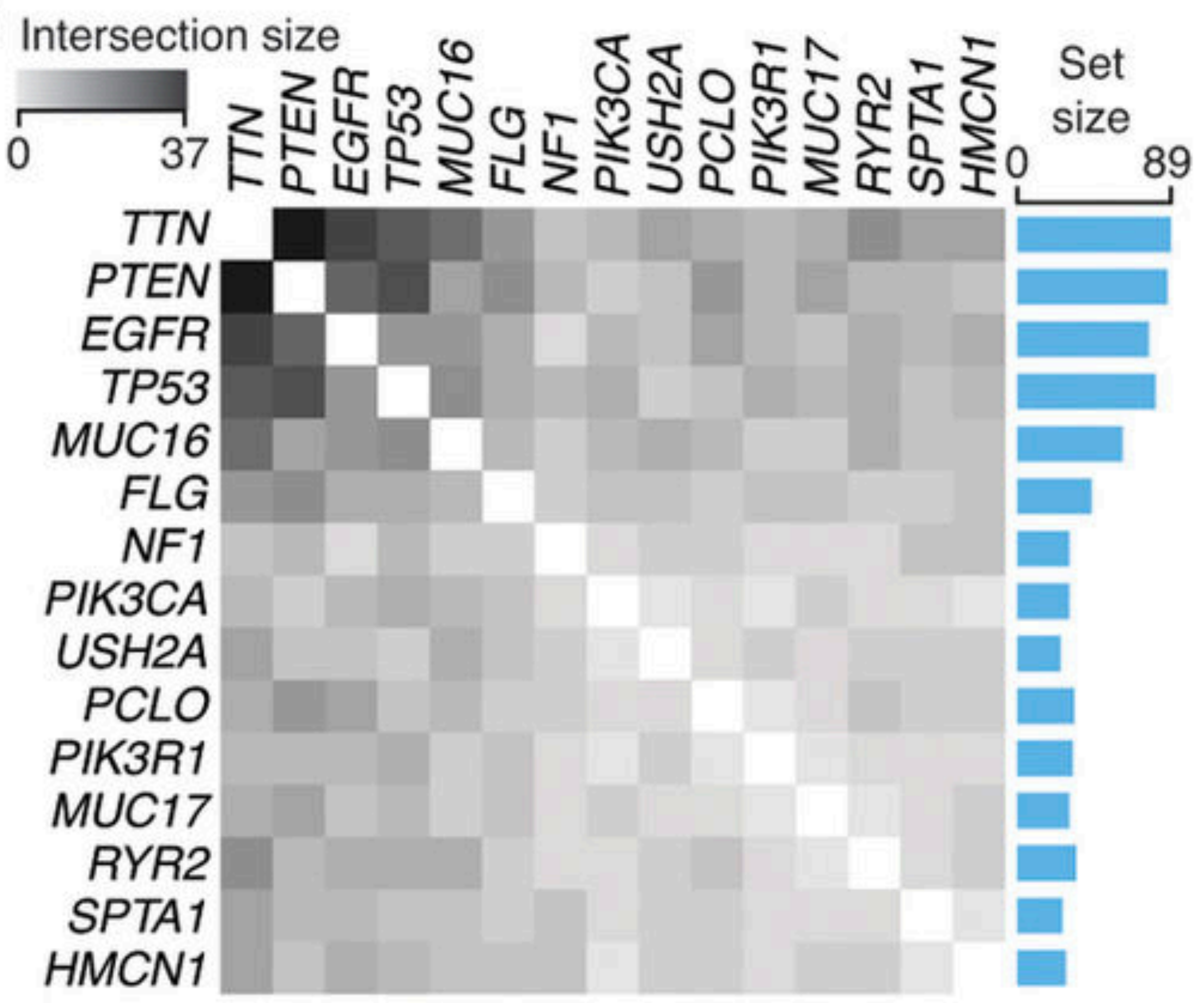
Shows pairwise overlap of sets

Doesn't show higher-order overlaps

Very scalable

Can't show attributes

Co-Mutations of genes

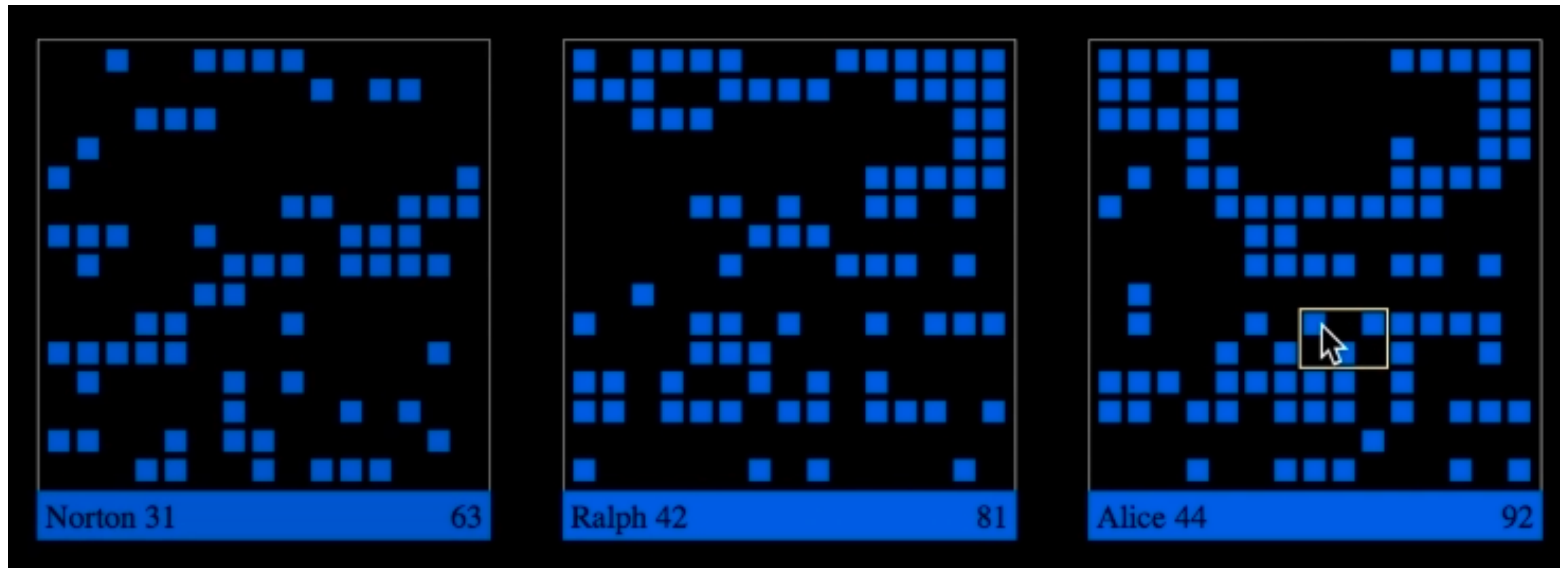


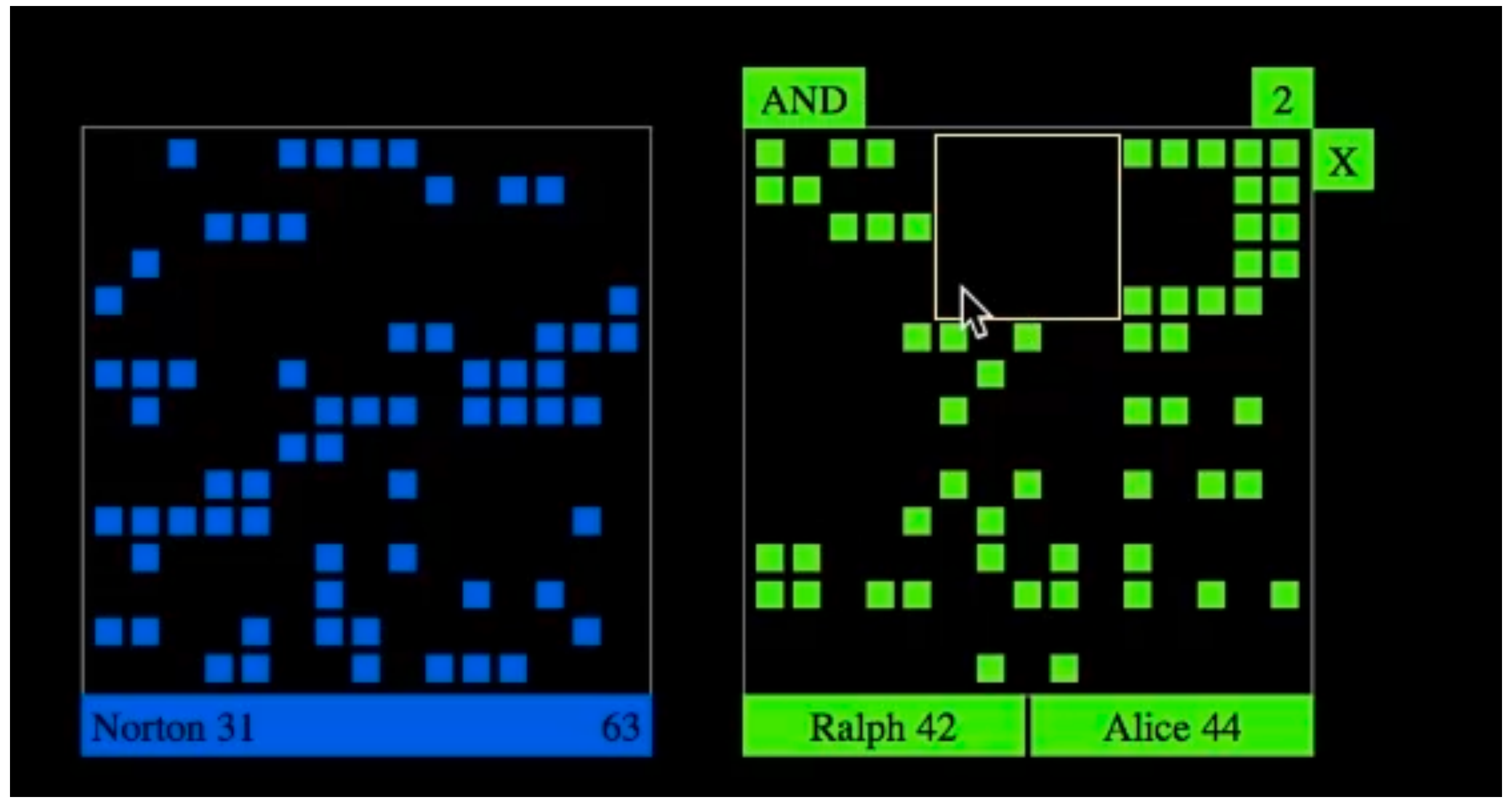
Set Matrices: OnSet

Set membership for each item shown in matrix

Comparisons can be made using AND or OR operations

Good for many sets and few items





https://vimeo.com/213029678#at=0

[Sadana 14]

Linear Diagrams

Fig. 1. Visualizing sets: linear diagrams.

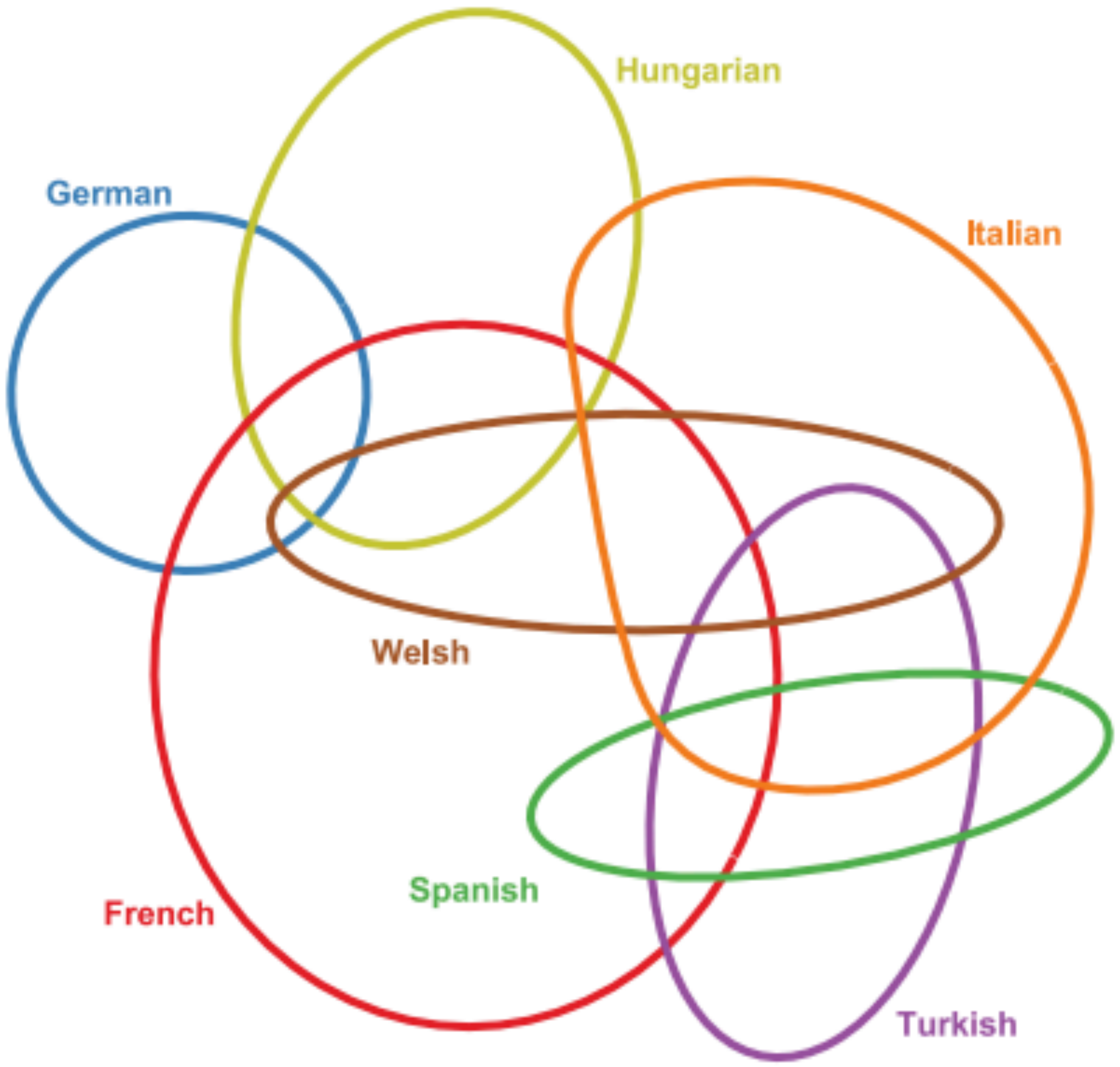
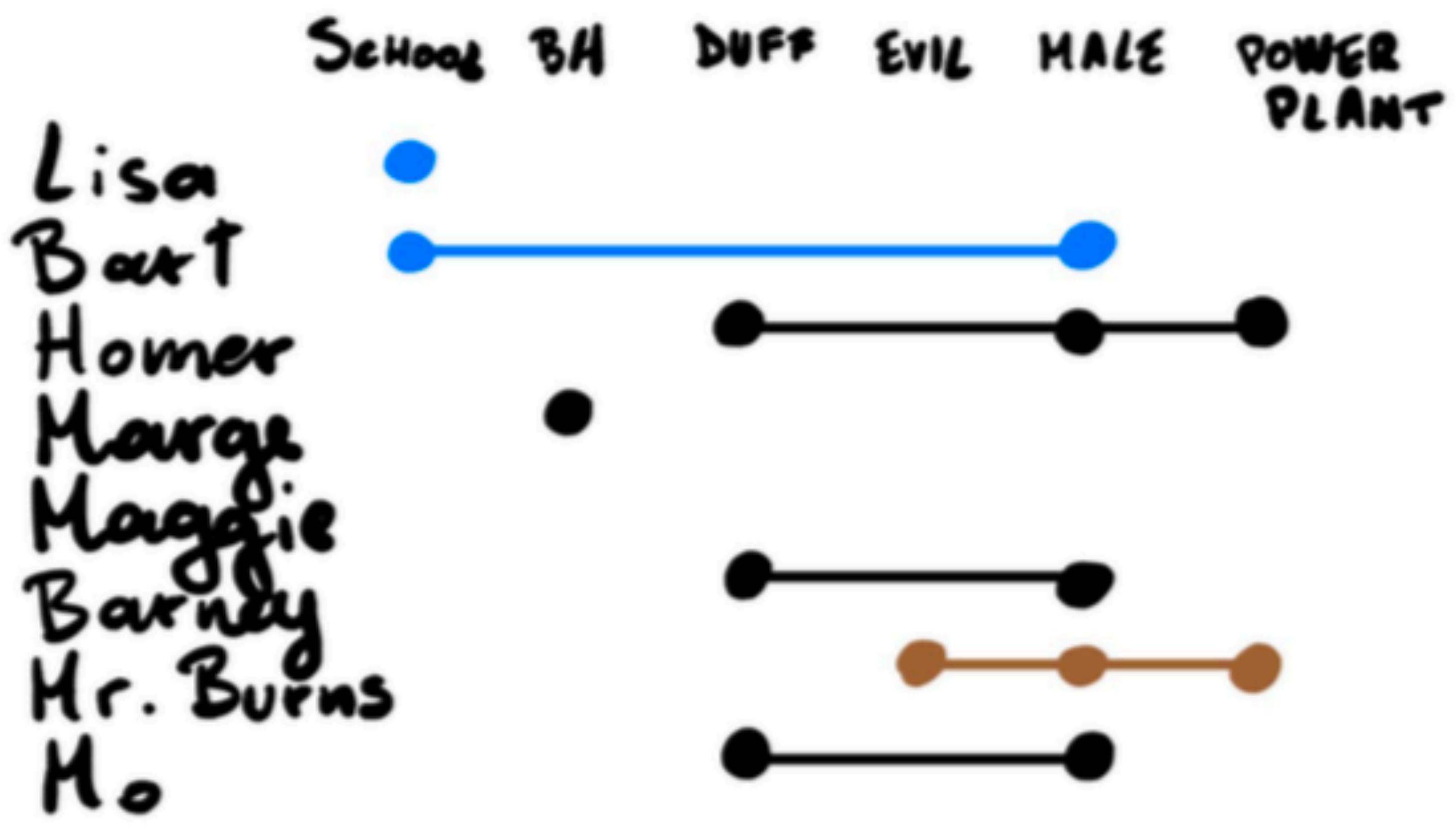
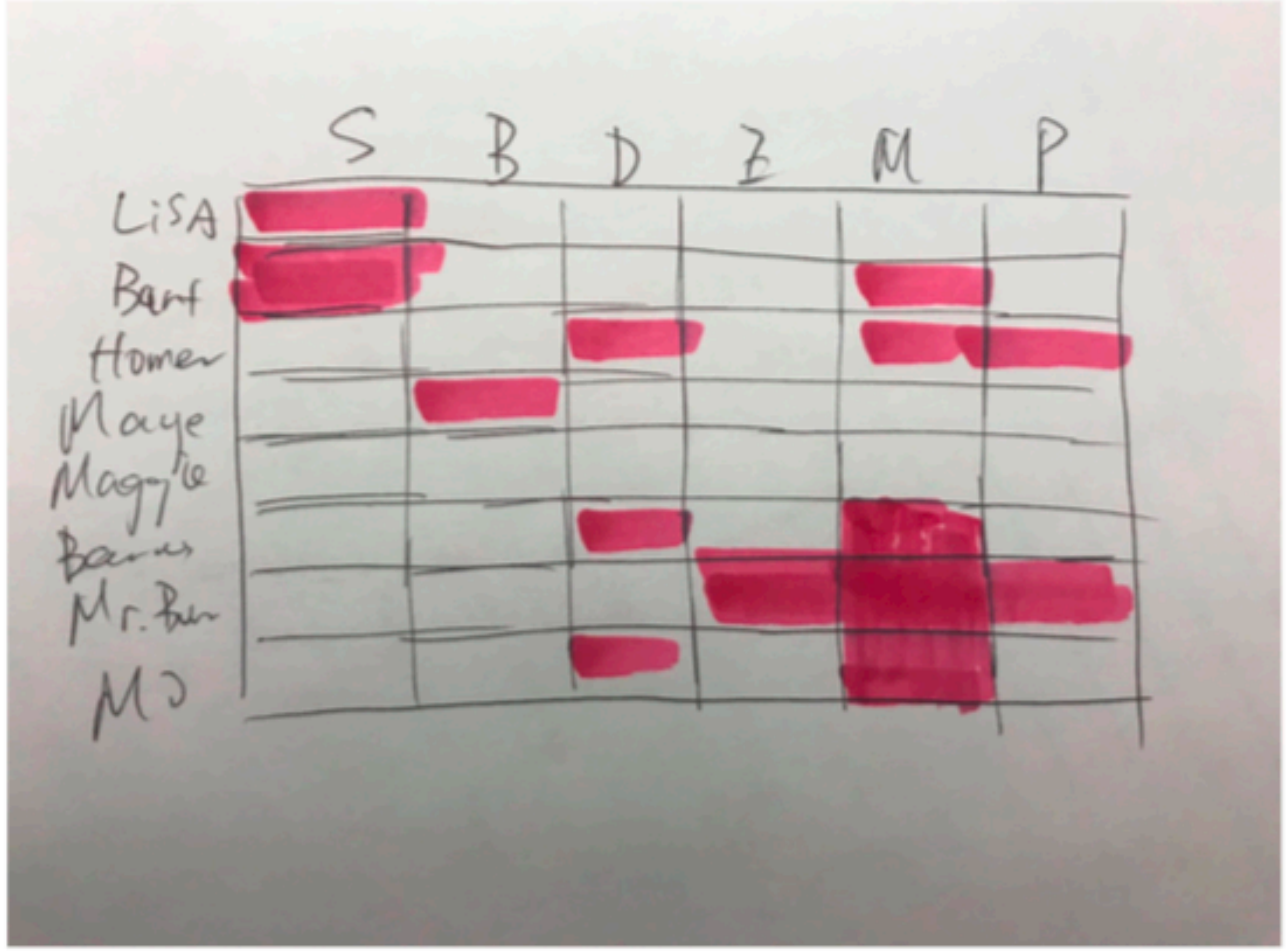
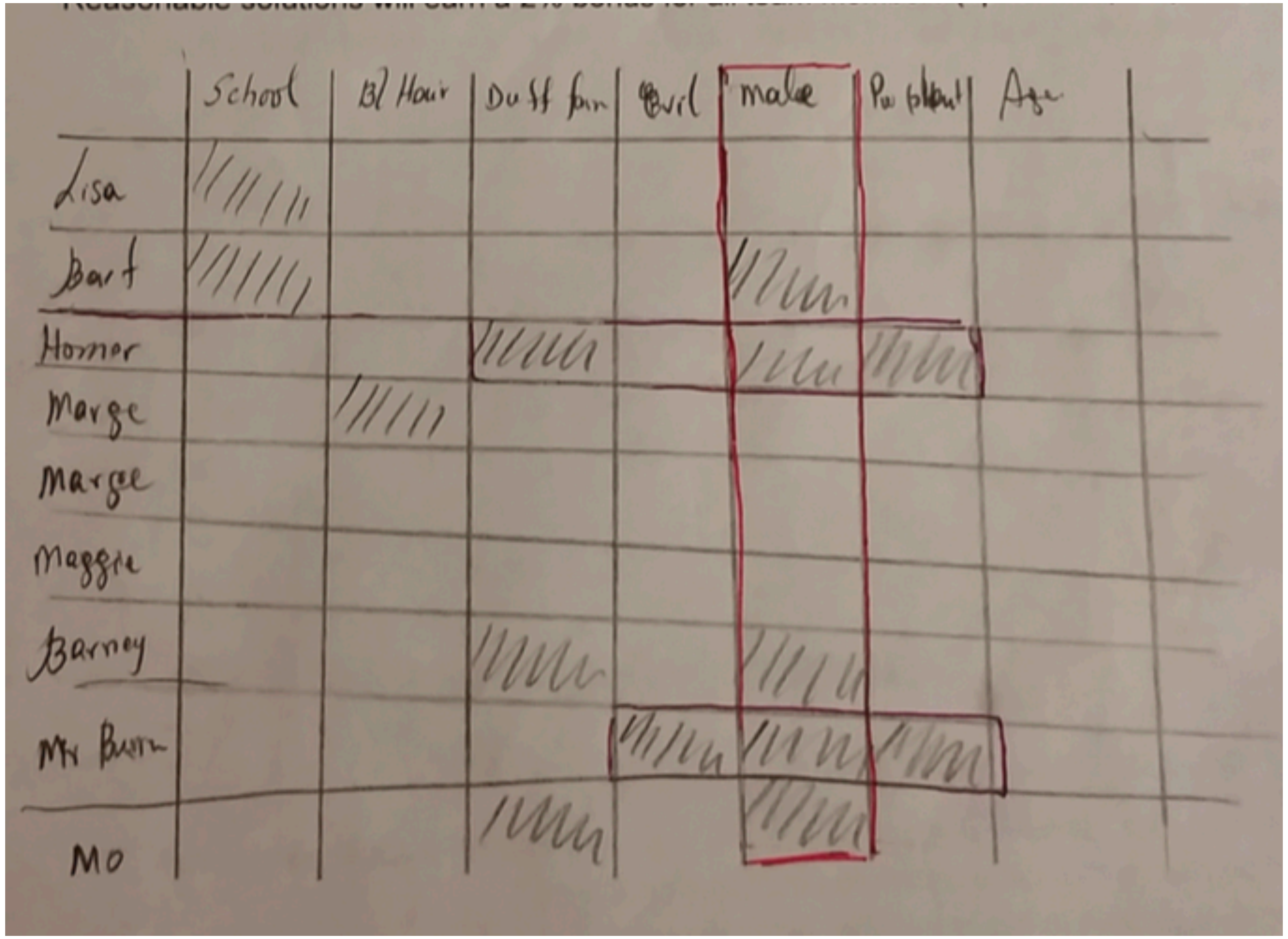


Fig. 2. Visualizing sets: Euler diagrams.

[RODGERS 2015]


School
BH
DUFF
EVIL
MALE
POWER PLANT
Lisa
Bart
Homer
Marge
Maggie
Barney
Mr. Burns
Mo


S
B
D
Z
M
P
LiSA
Bart
Homer
Maggie
MO


School
Bl Hair
Duff fan
Evil
male
Pw plant
Age
Lisa
Bart
Homer
Marge
Marge
Maggie
Barney
Mr Burns
Mo

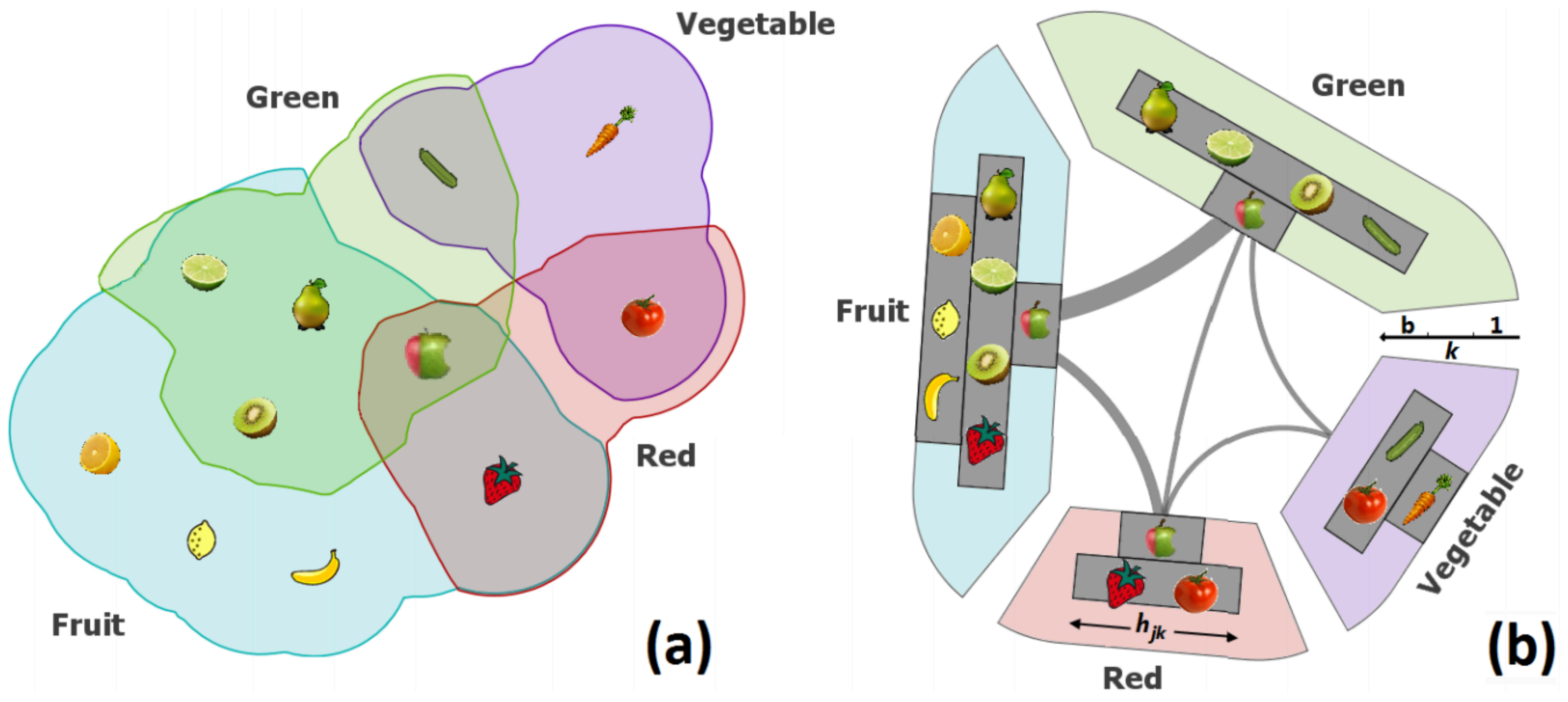
Radial Sets

Sets are segments on a “circle”

Relationships are encoded as ribbons

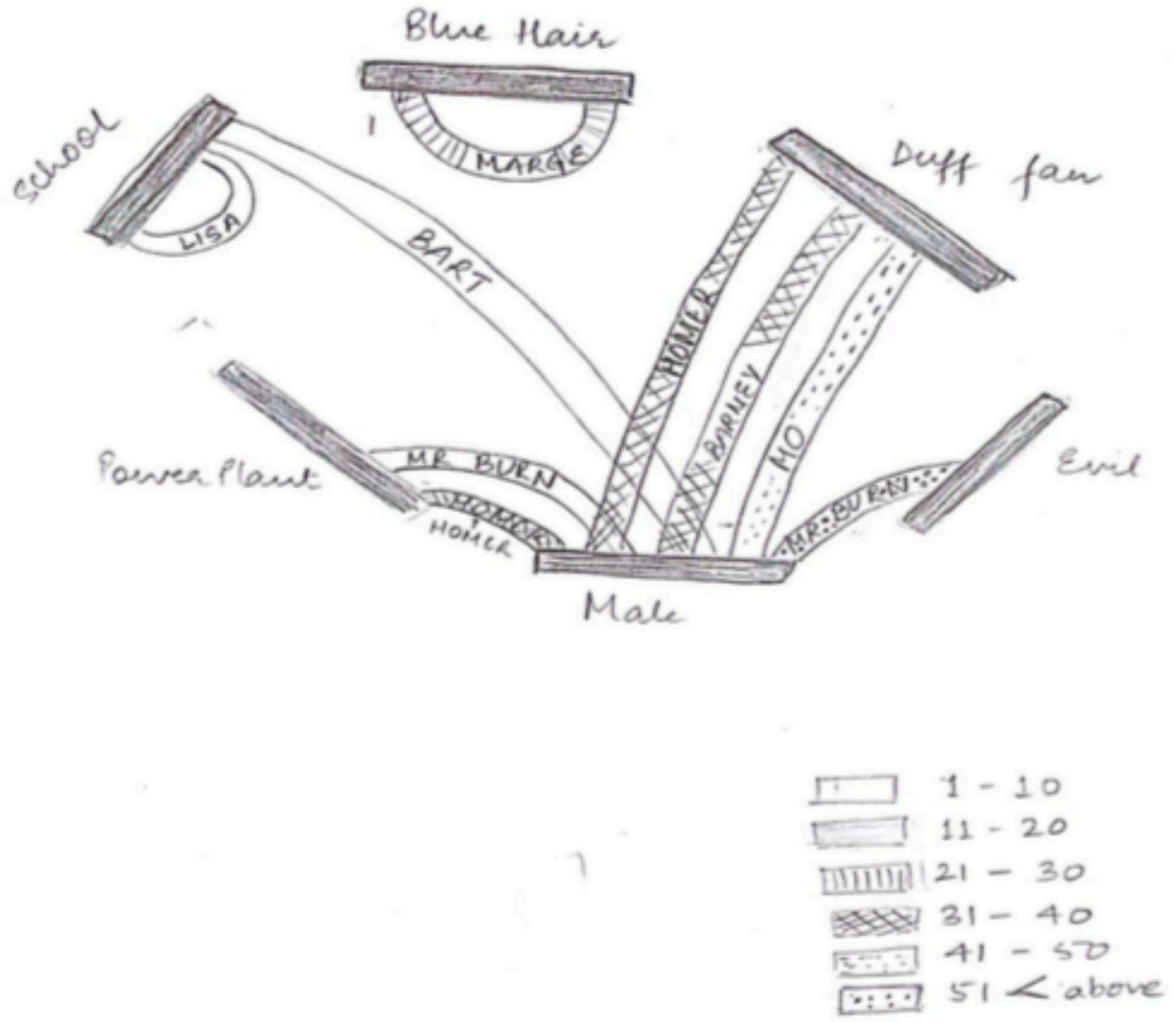
Size of segments encodes size of sets

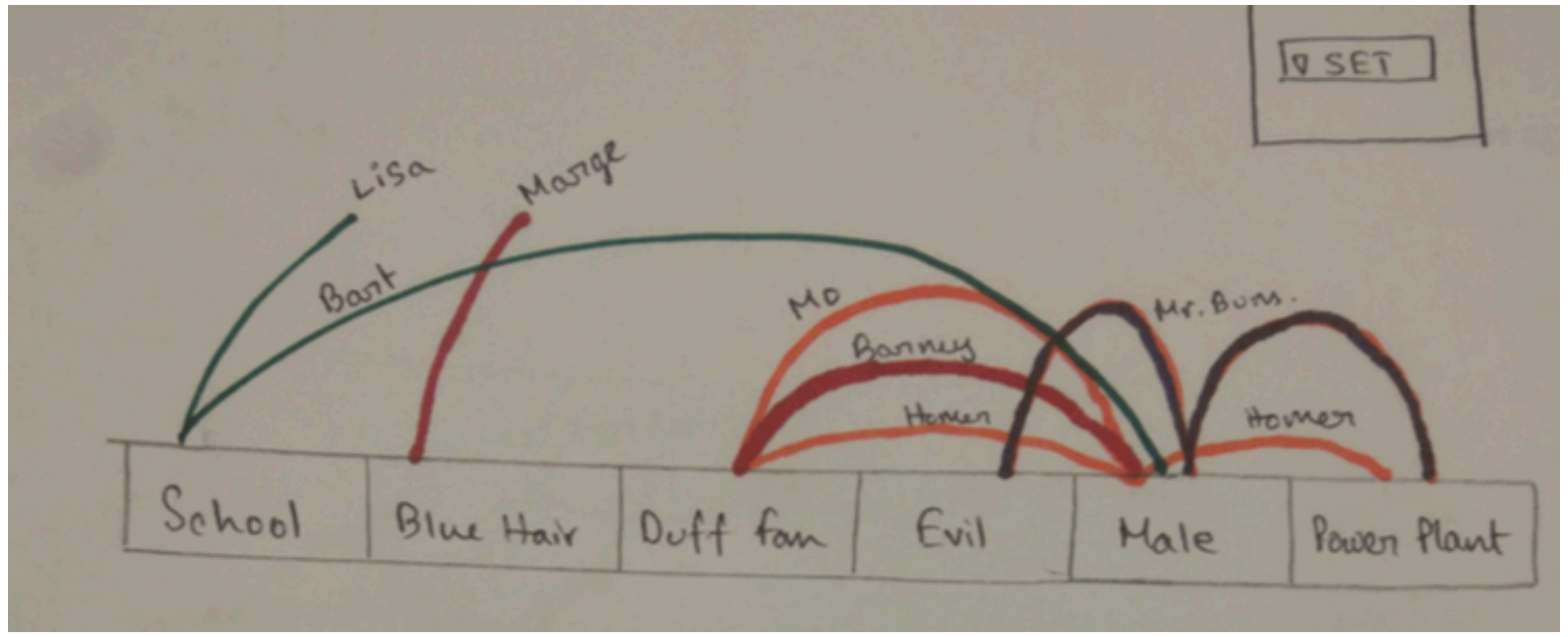
Histograms in segments show degrees

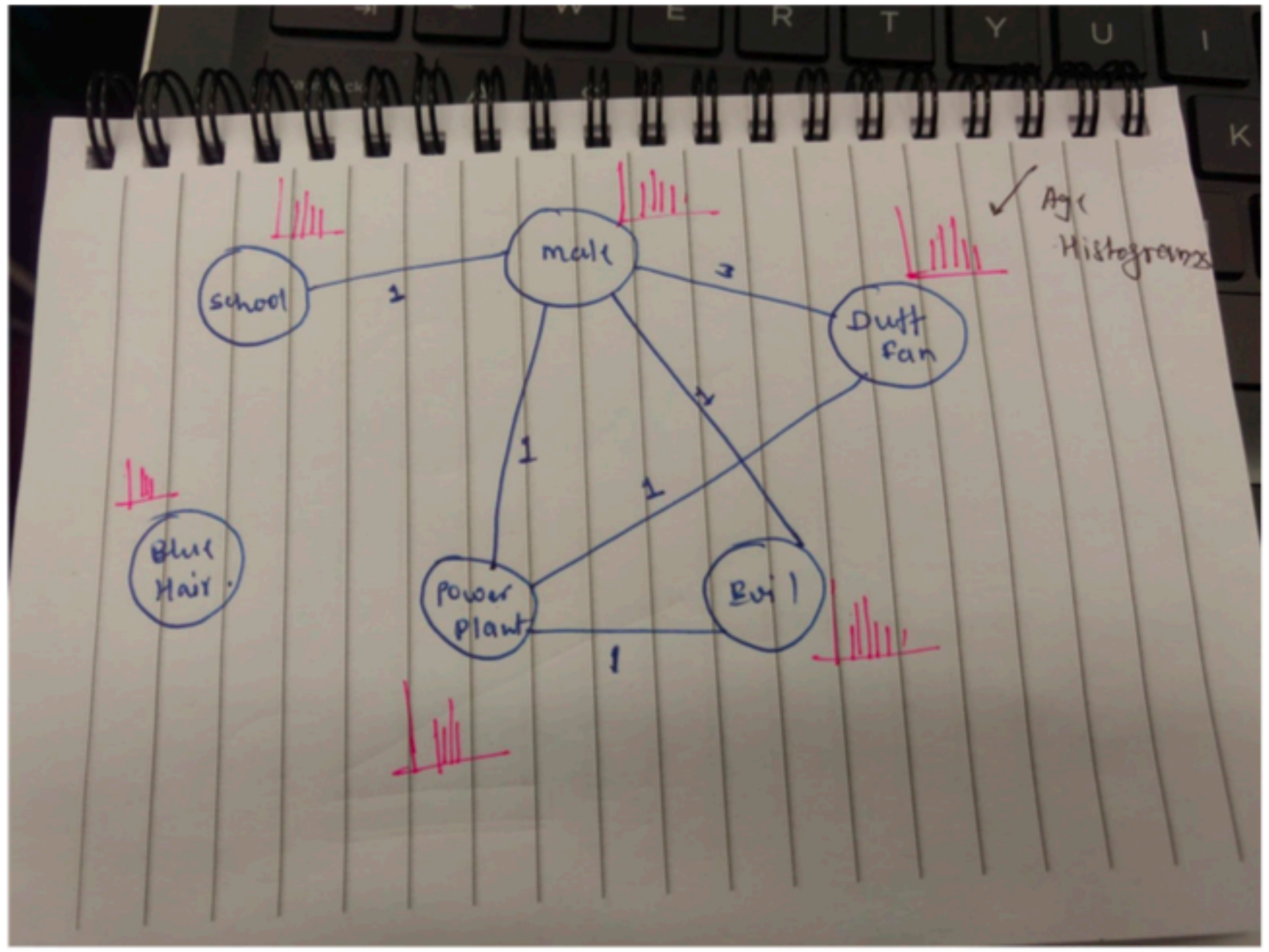


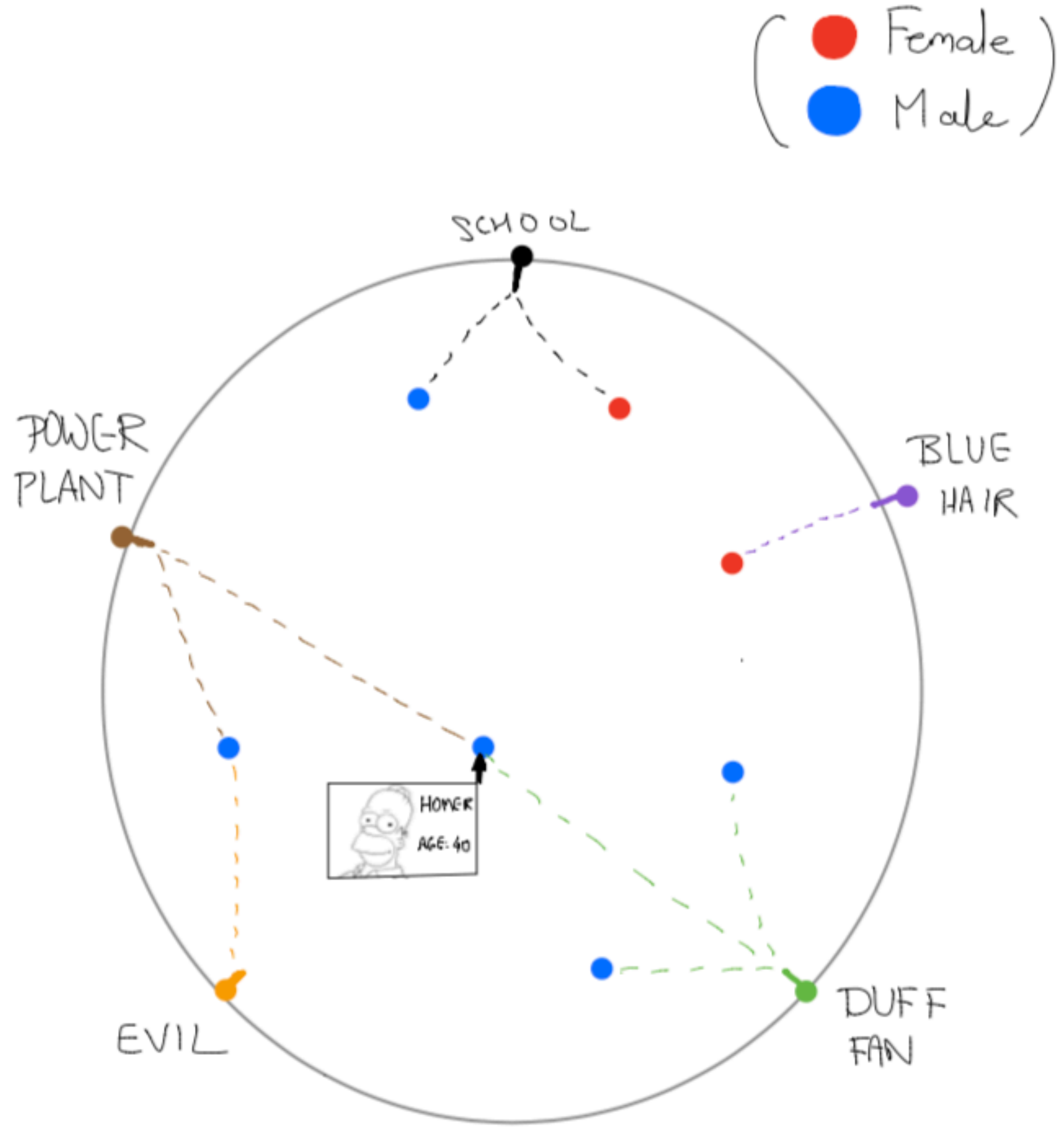
https://www.youtube.com/watch?v=UcYRrPqC5A8

[Alsallakh 2013]


Blue Hair
School
Duff fan
MARGE
LISA
BART
HOMER
BARNEY
MO
Power Plant
MR BURN
HOMER
MR. BURNS
Evil
Male
1 - 10
11 - 20
21 - 30
31 - 40
41 - 50
51 < above


SET
Lisa
Marge
Bart
Mo
Barney
Homer
Mr. Burns
Homer
School
Blue Hair
Duff fan
Evil
Male
Power Plant


Age Histograms
School
male
Duff fan
Blue Hair
Power Plant
Evil

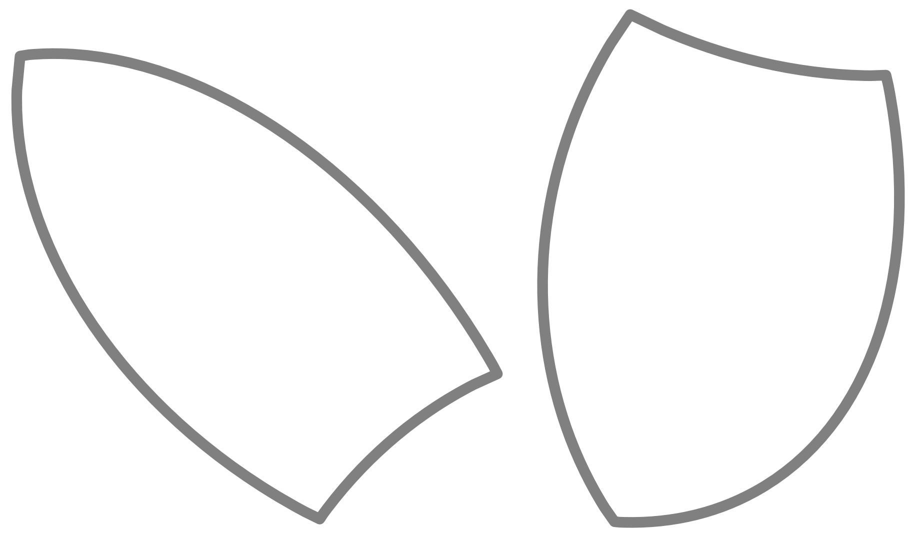

Female
Male
SCHOOL
POWER PLANT
BLUE HAIR
HOMER
AGE: 40
EVIL
DUFF FAN

[InfoVis'14]

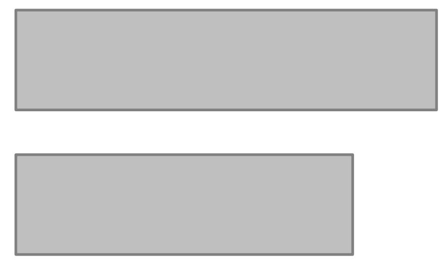
UpSet
Visualizing Intersecting Sets

Set Vis Goals

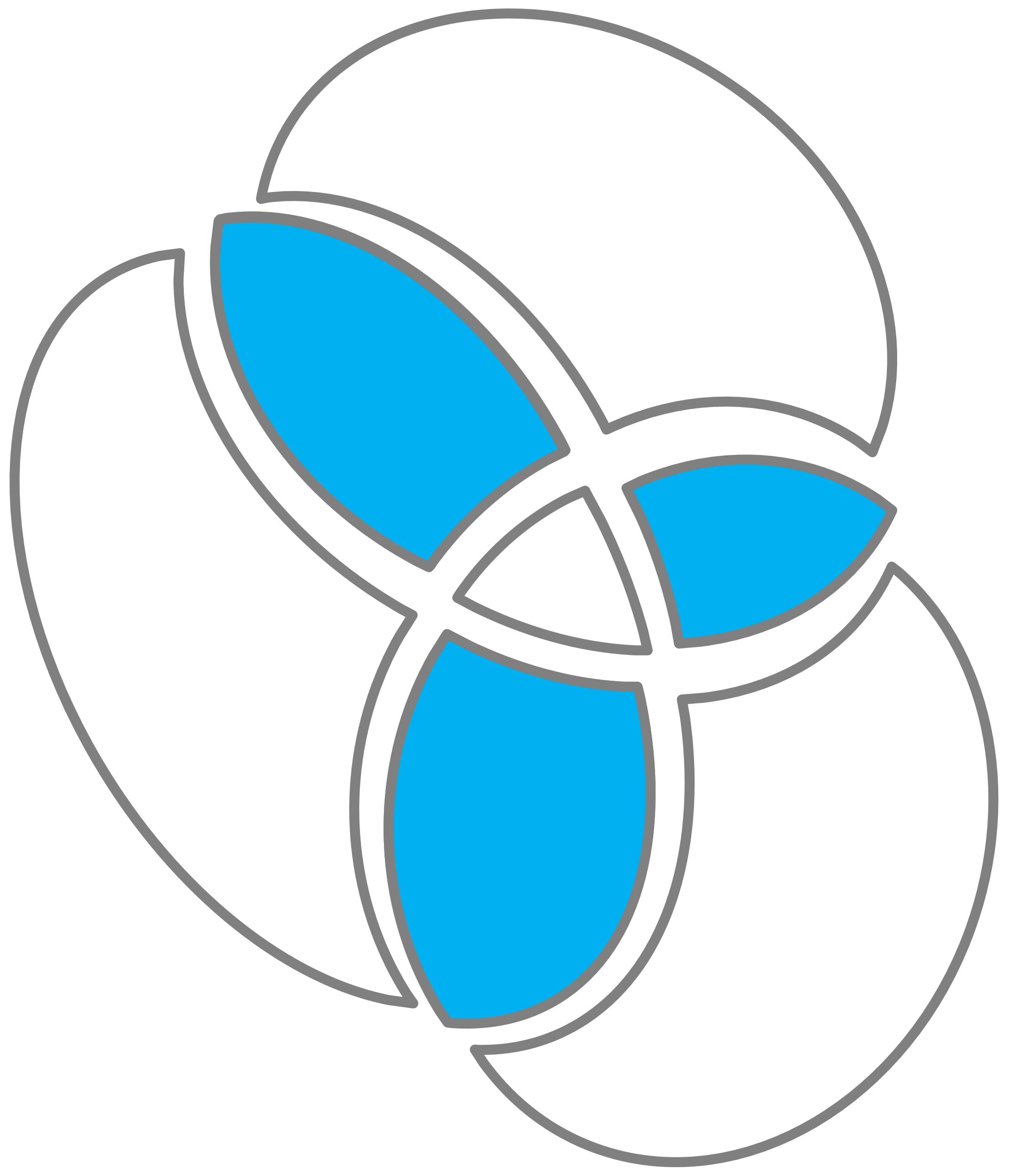
1. Efficient visual encoding



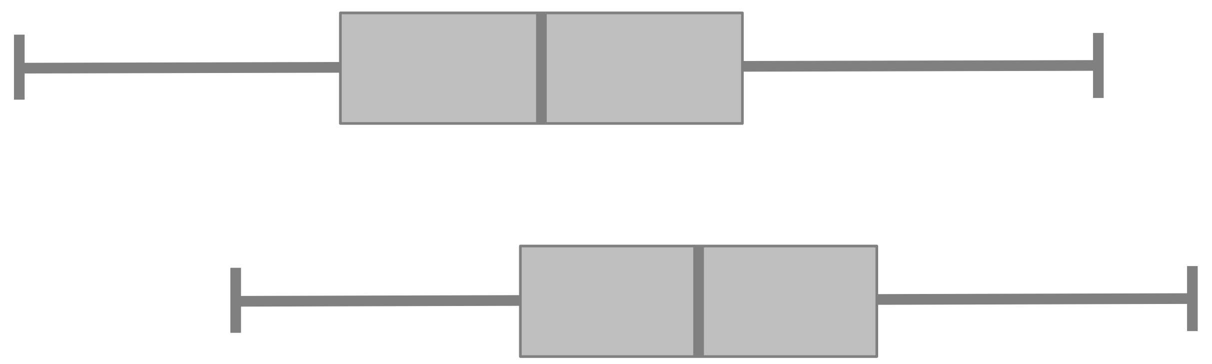
vs.



2. Creating complex slices of a dataset



3. Visualize attributes

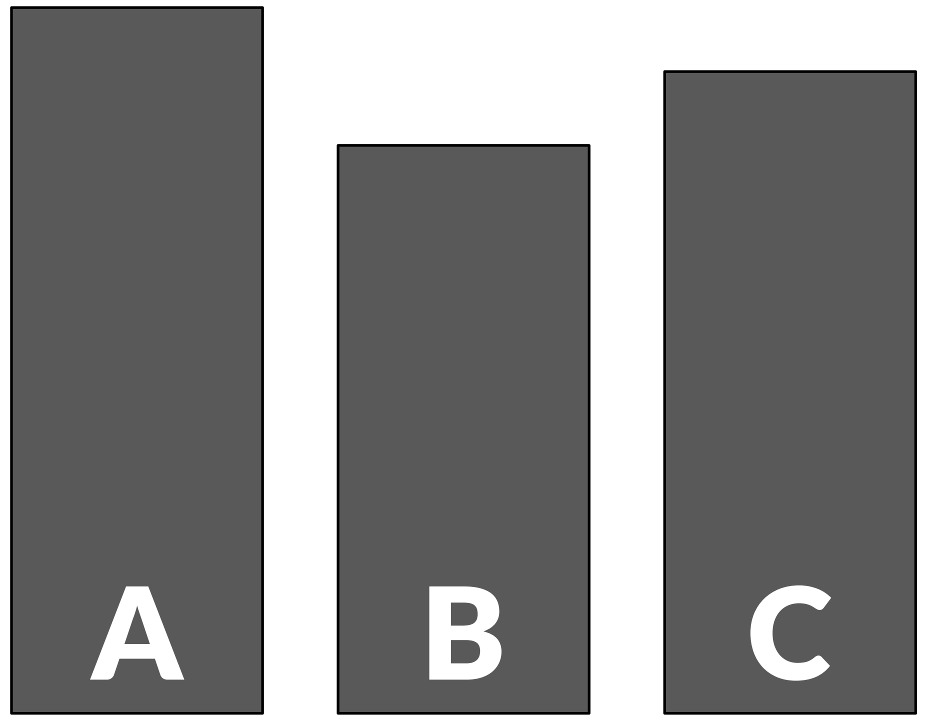


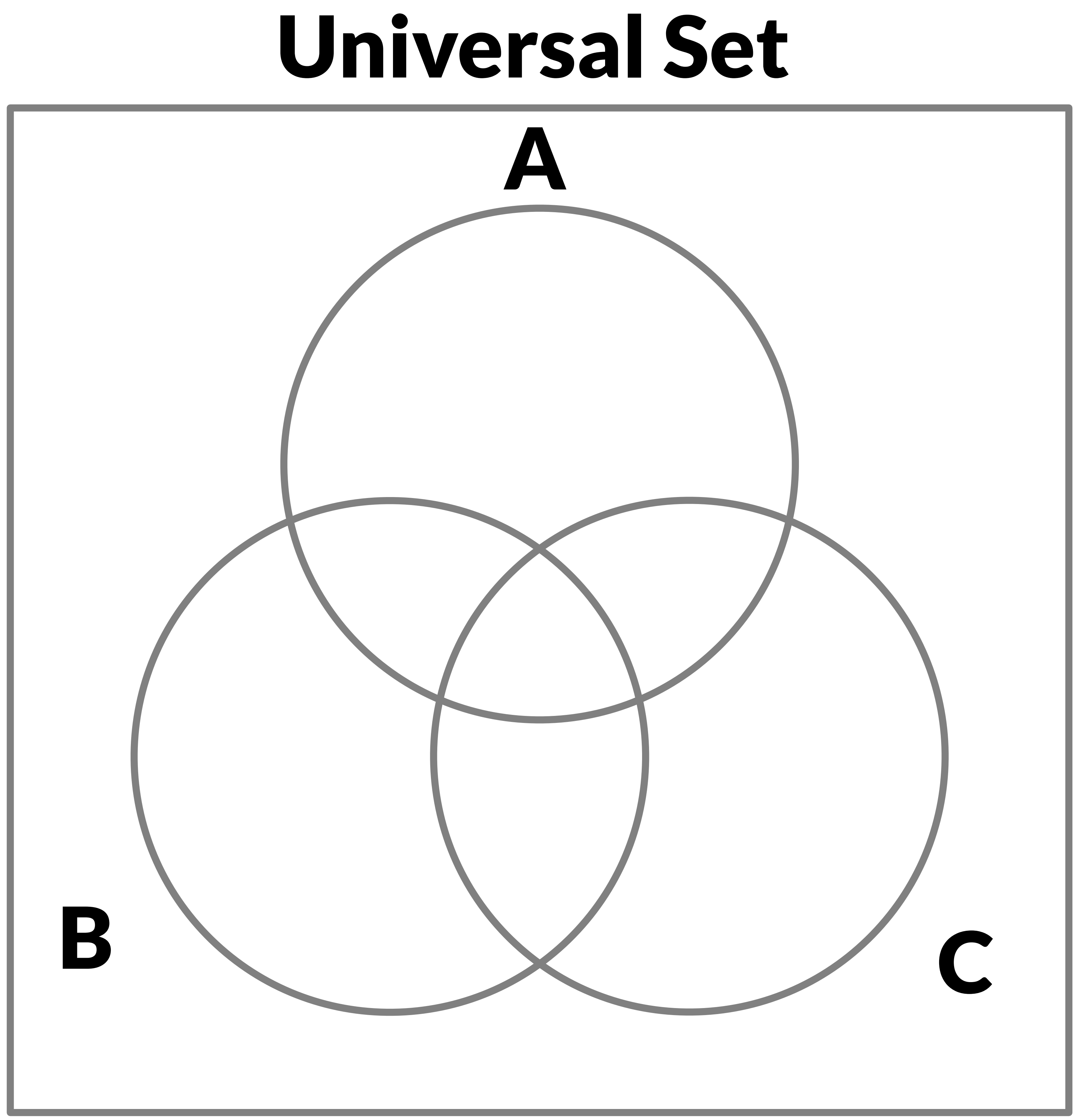
[Movie Lens Dataset]

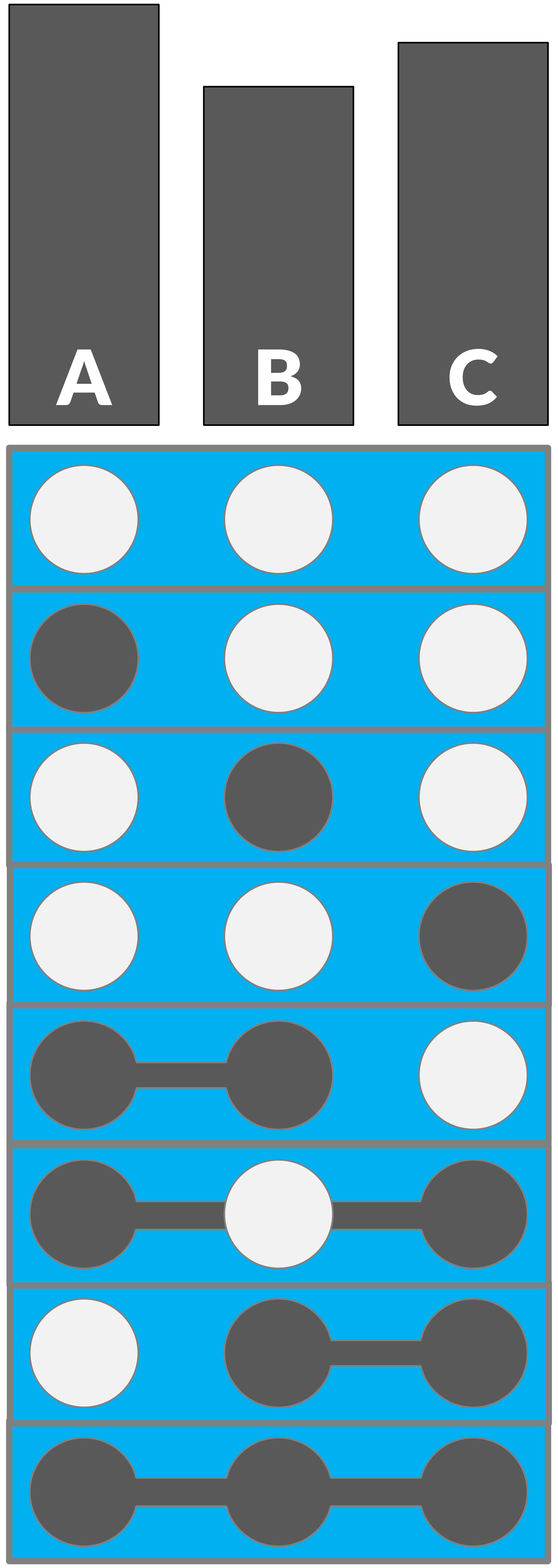
Attribute Details

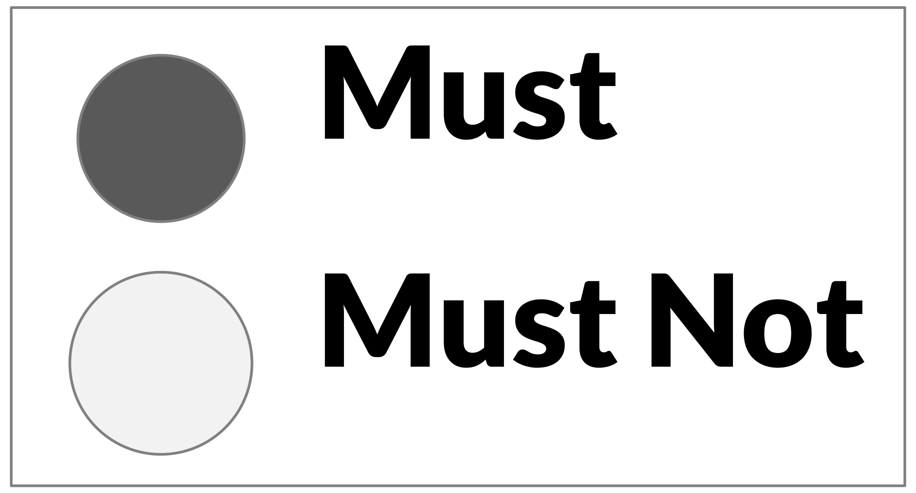
Visualizing Intersections **Visualizing Properties** **Element List & Queries**

Visualizing Intersections

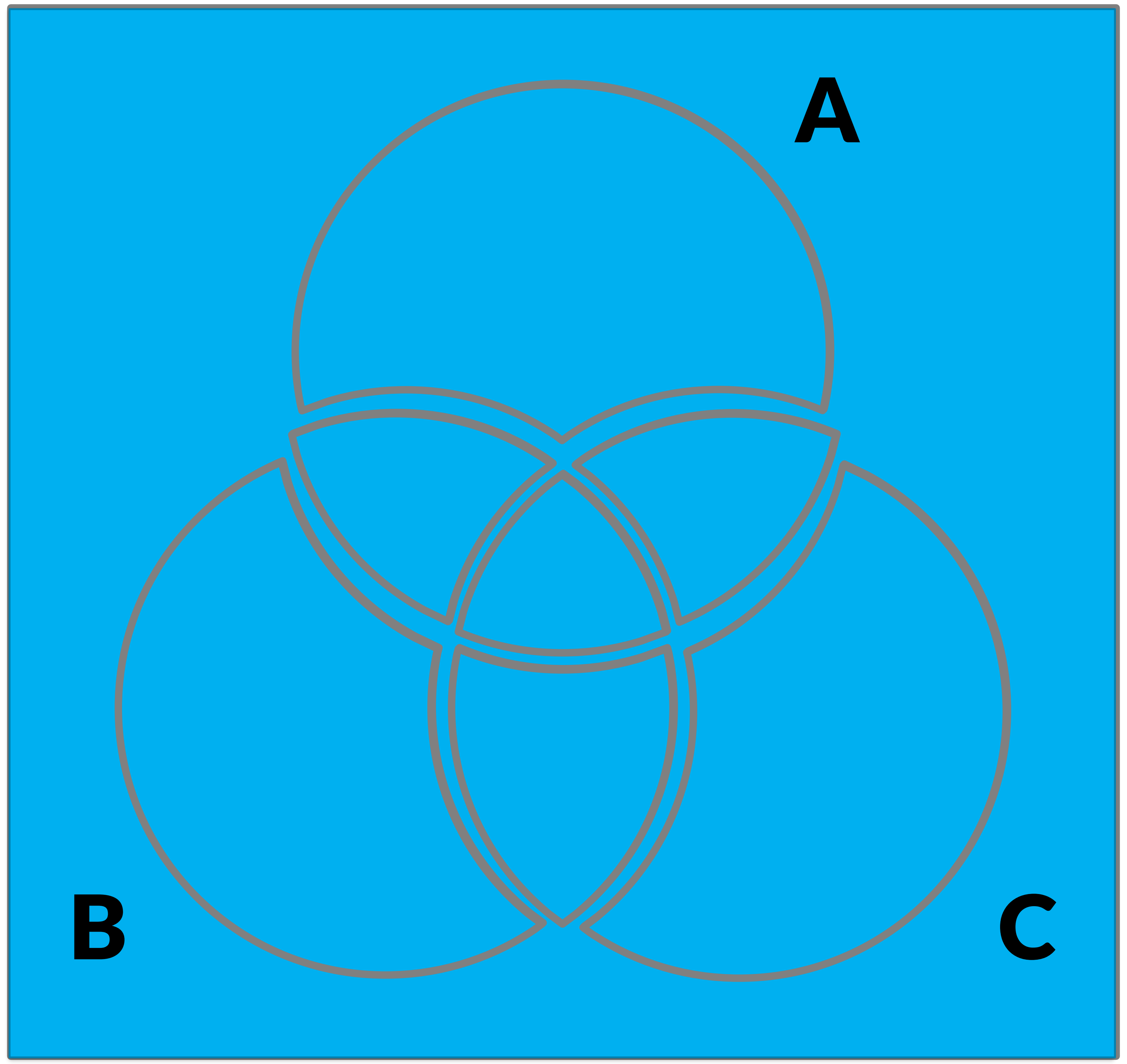

A
B
C

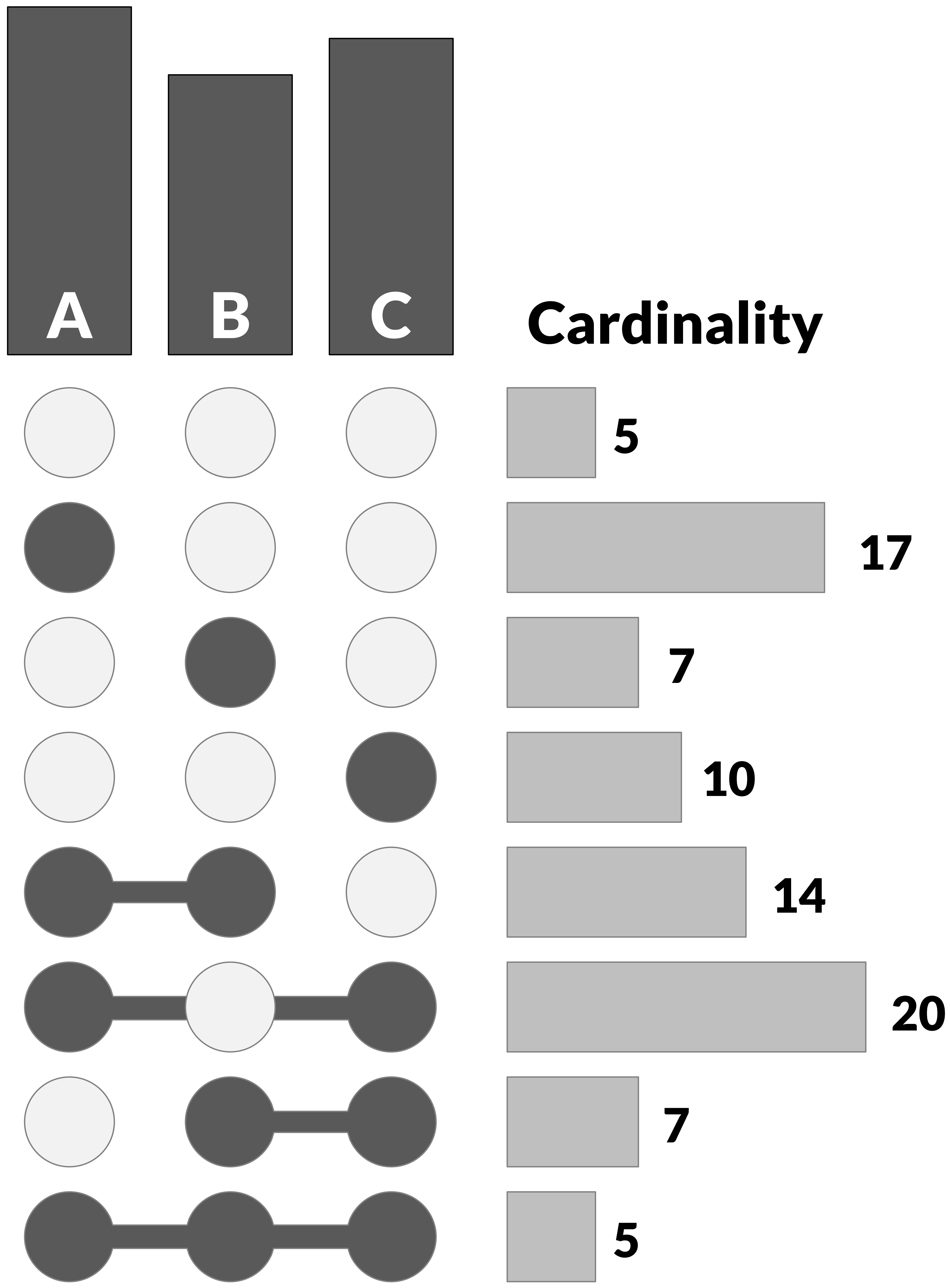

Universal Set
A
B
C

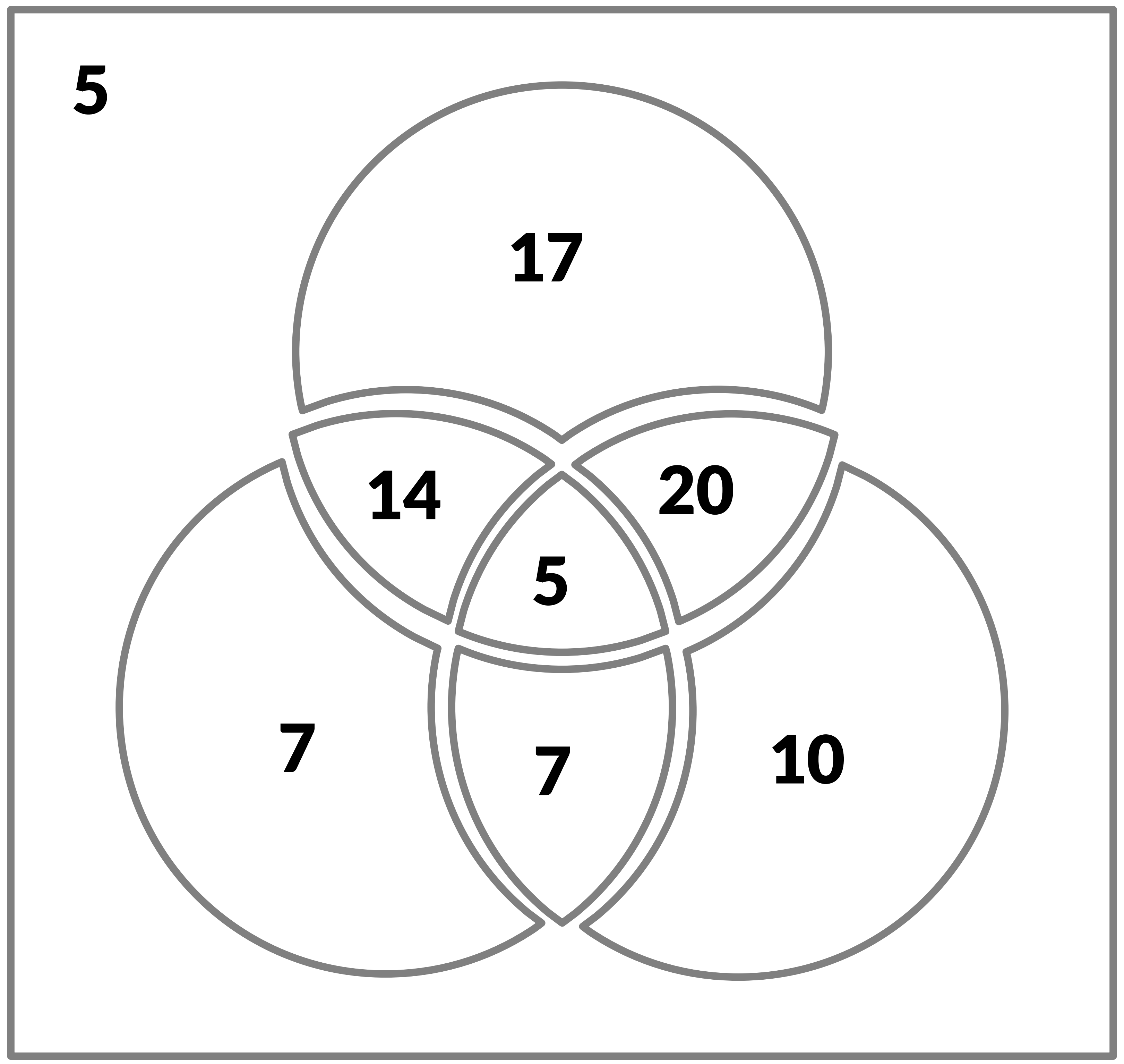




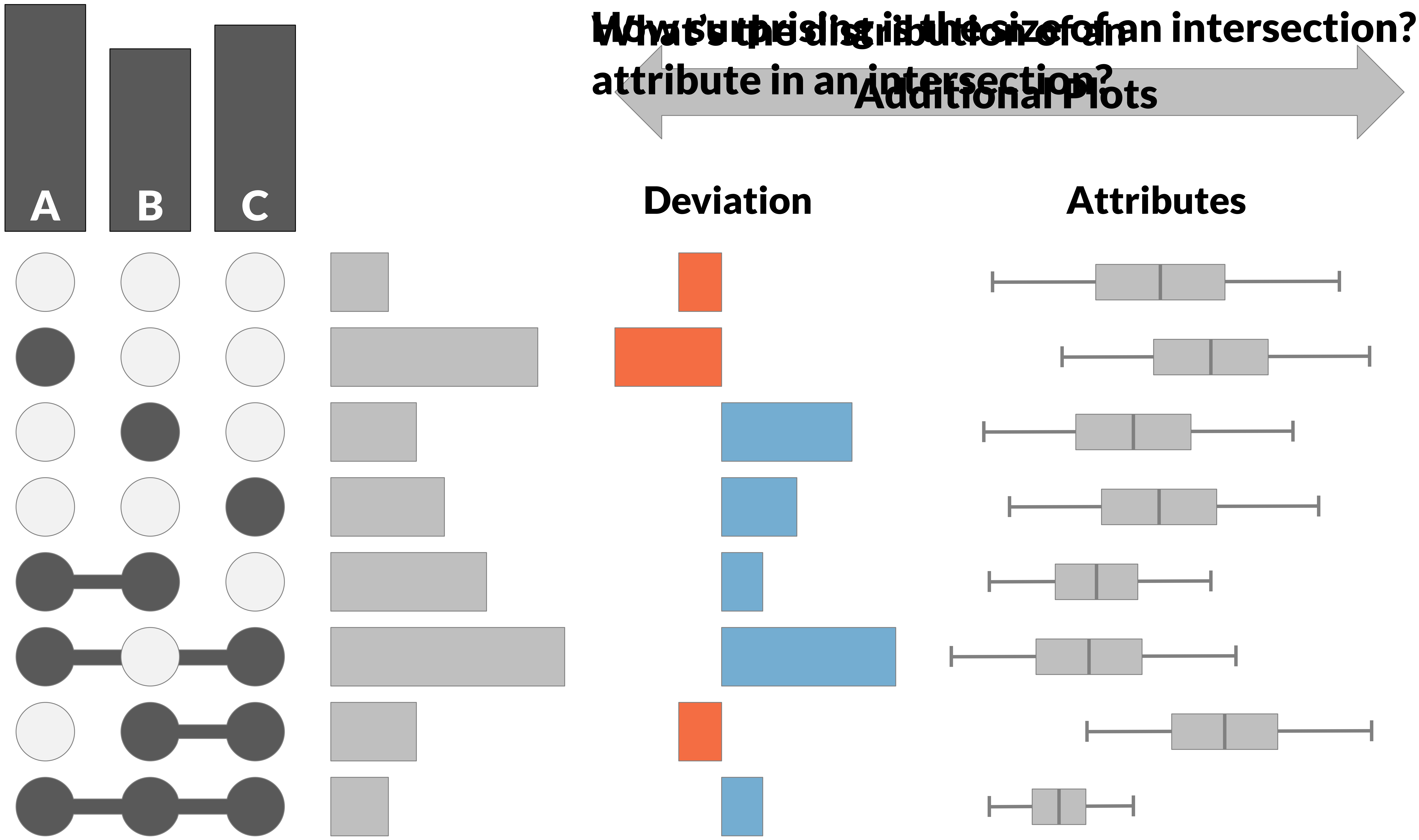
Universal Set

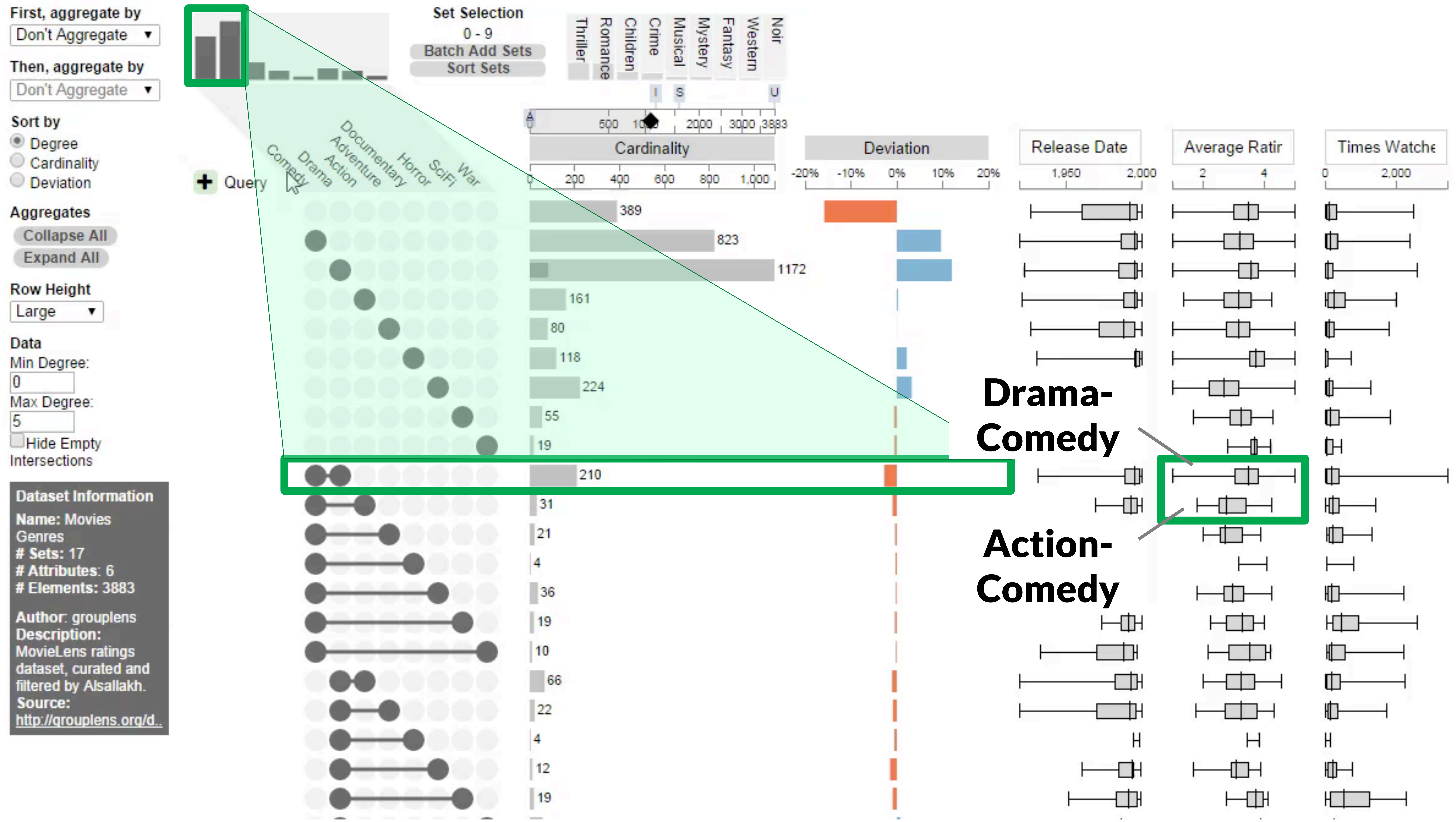



A
B
C
Cardinality
5
17
7
10
14
20
7
5


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

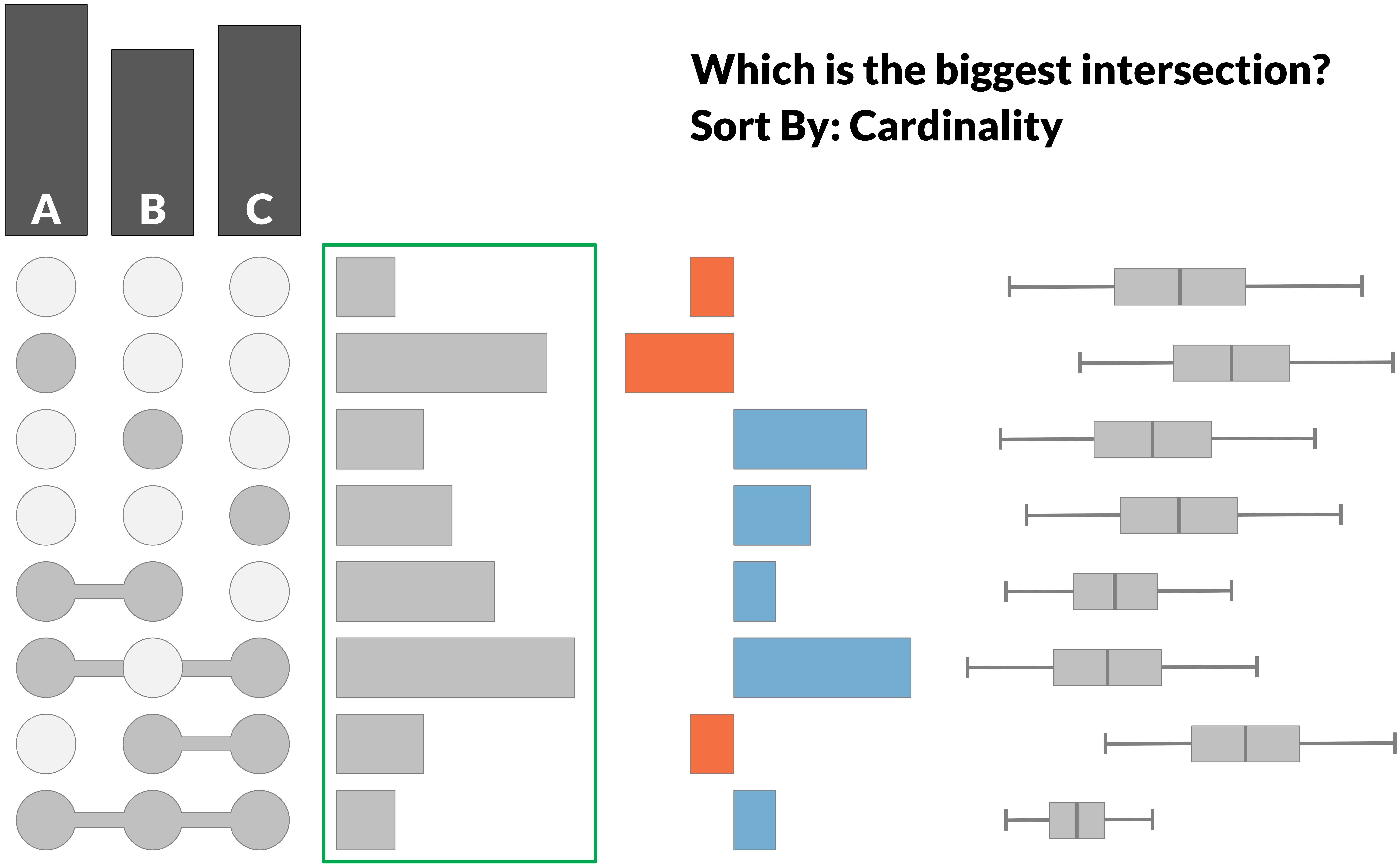


First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: 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

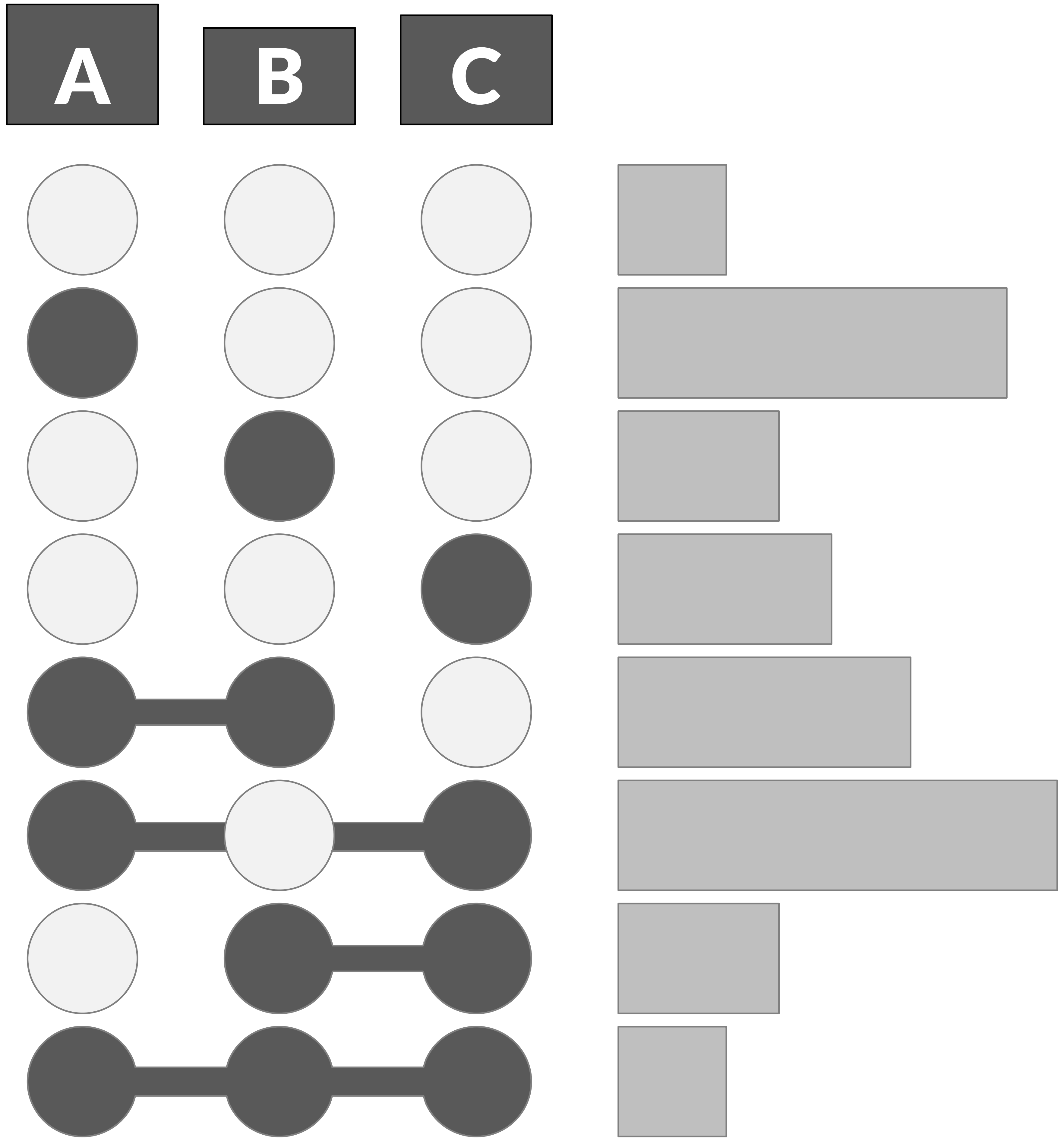
Which is the biggest intersection?

Sort By: Cardinality

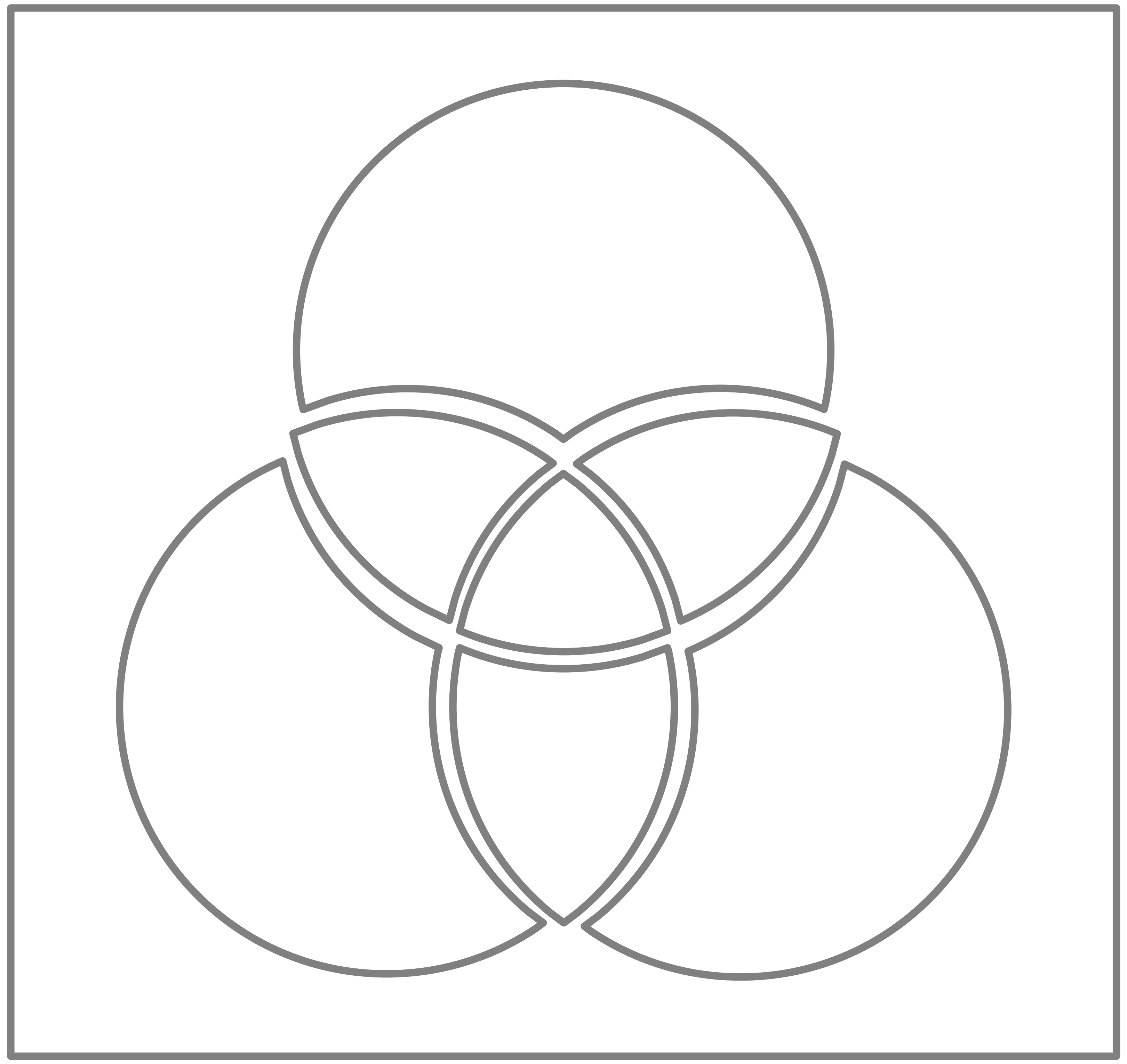



First, aggregate by
Don't Aggregate
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
Description: MovieLens ratings dataset, curated and filtered by Alsallakh.
Source: http://grouplens.org/d...
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
389
823
1172
161
80
118
224
55
19
210
31
21
4
36
19
10
66
22
4
12
19

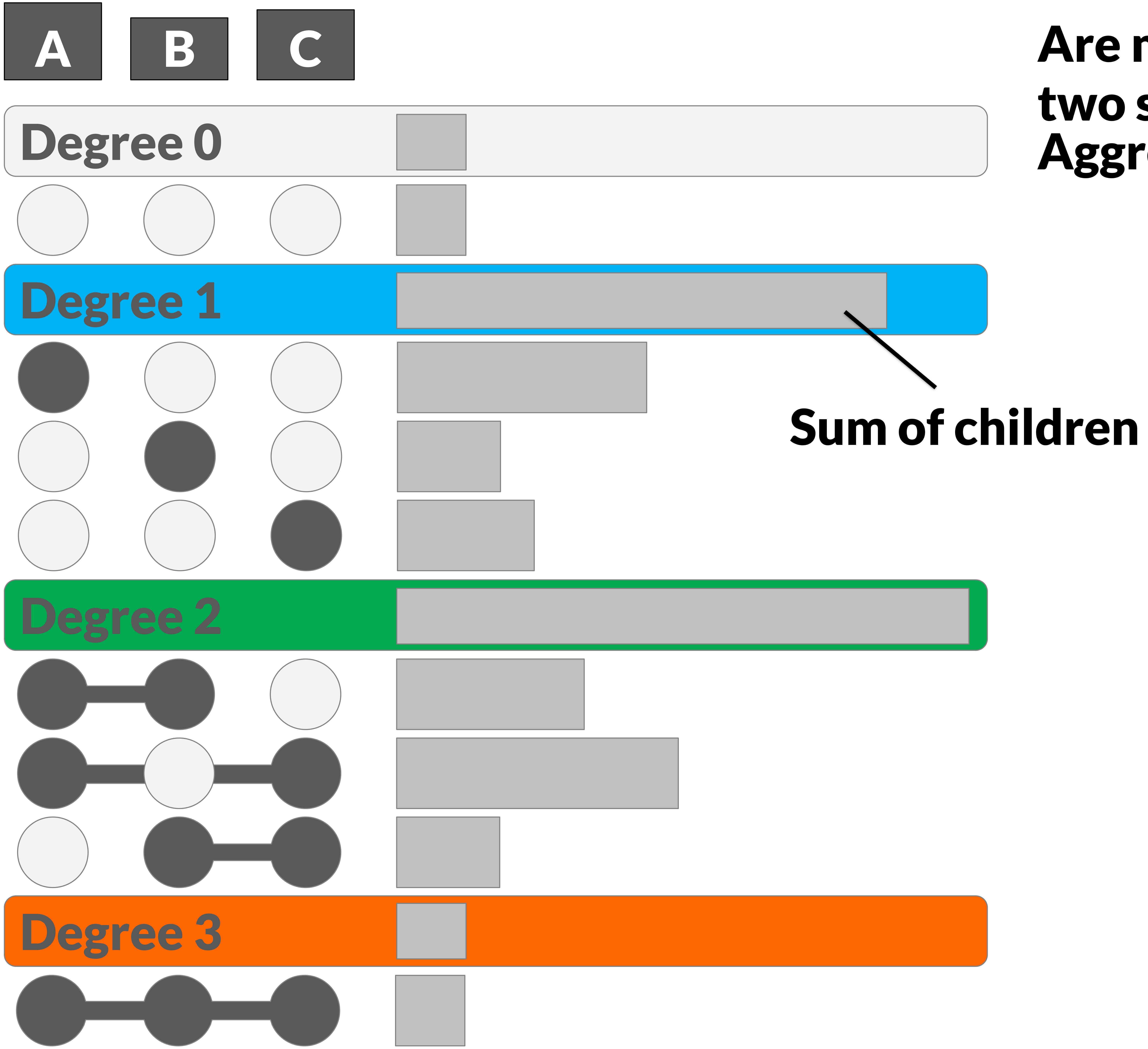
Aggregation

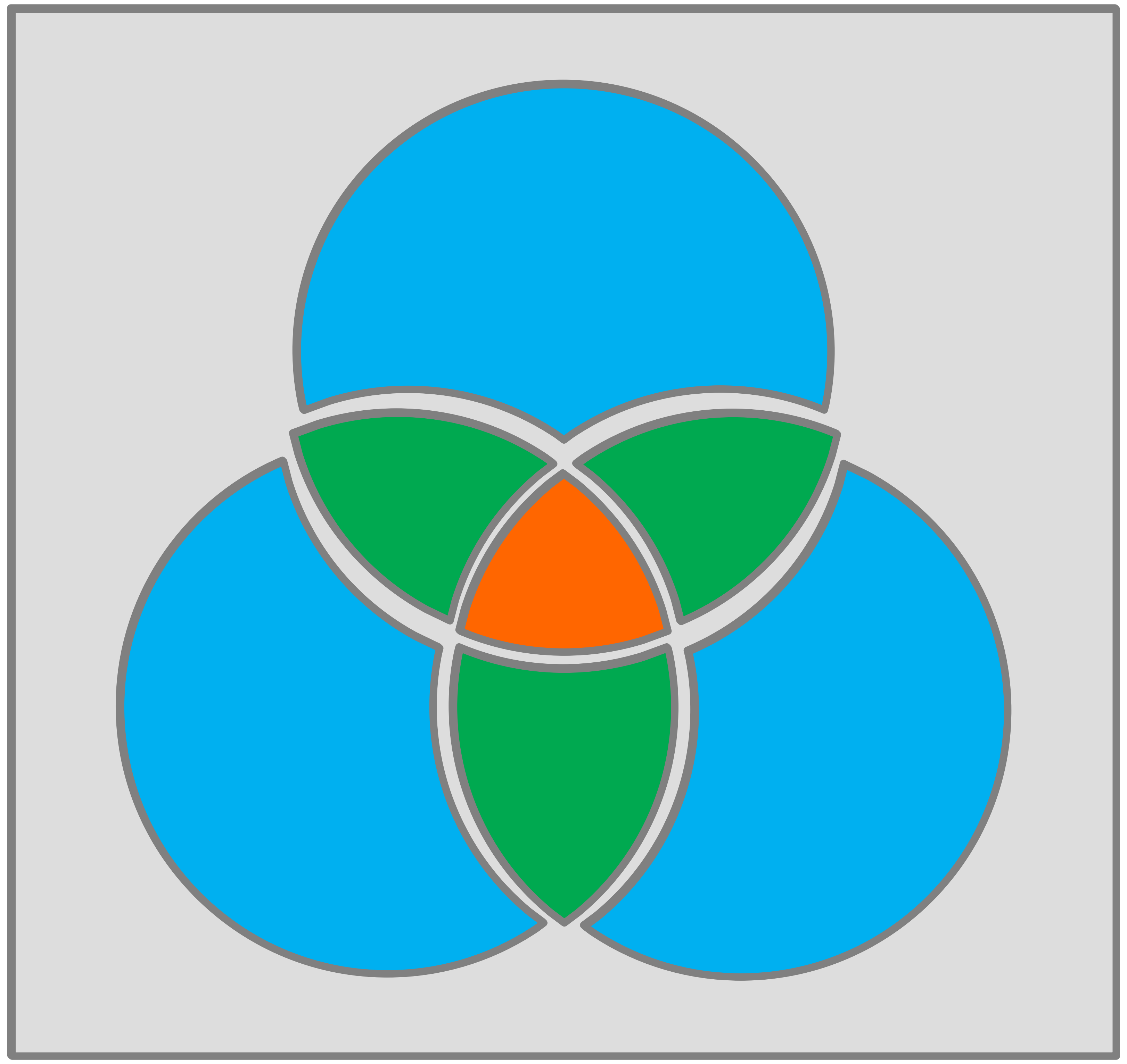


Are many items shared between two sets?
Aggregate By: Degree

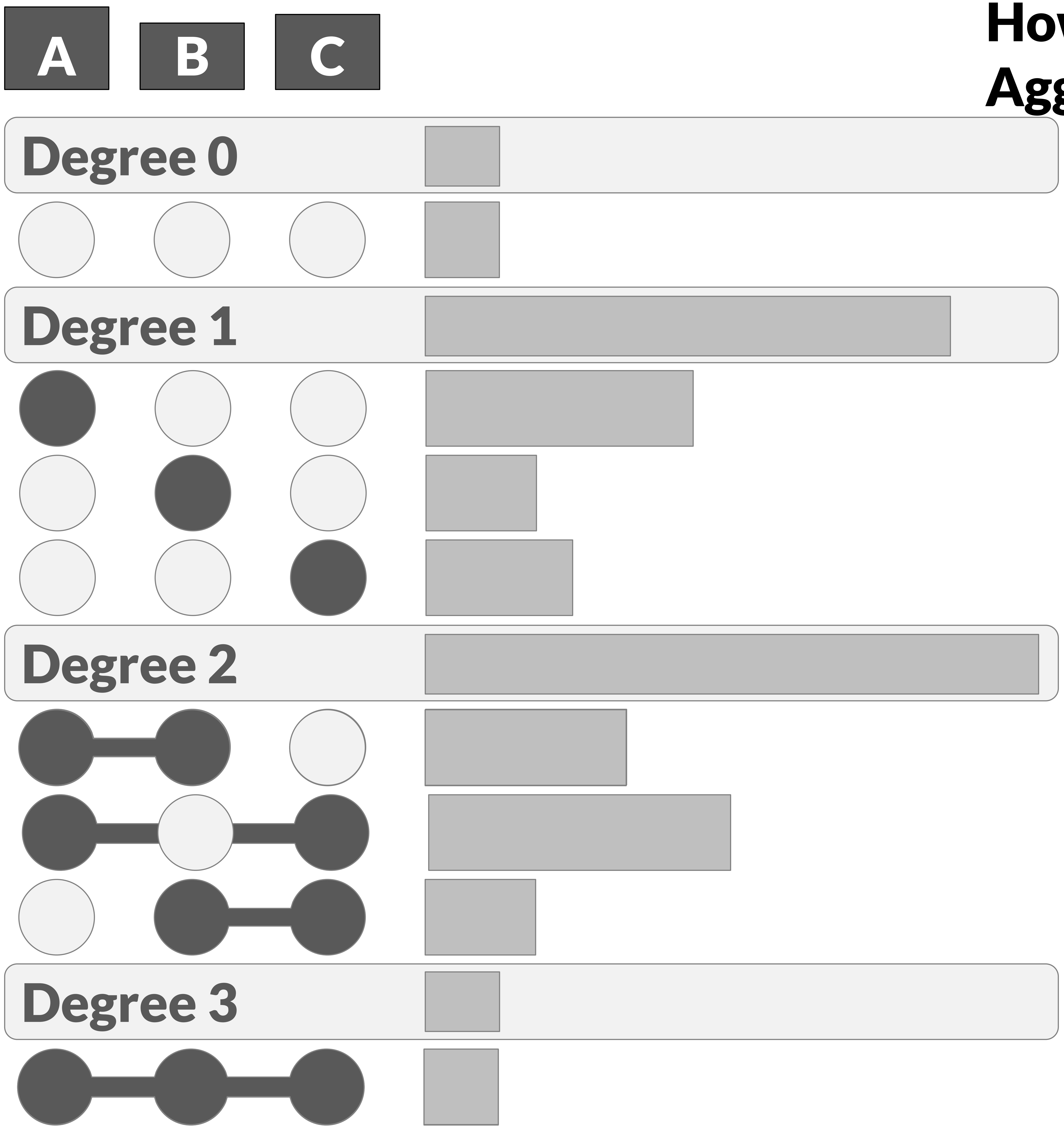


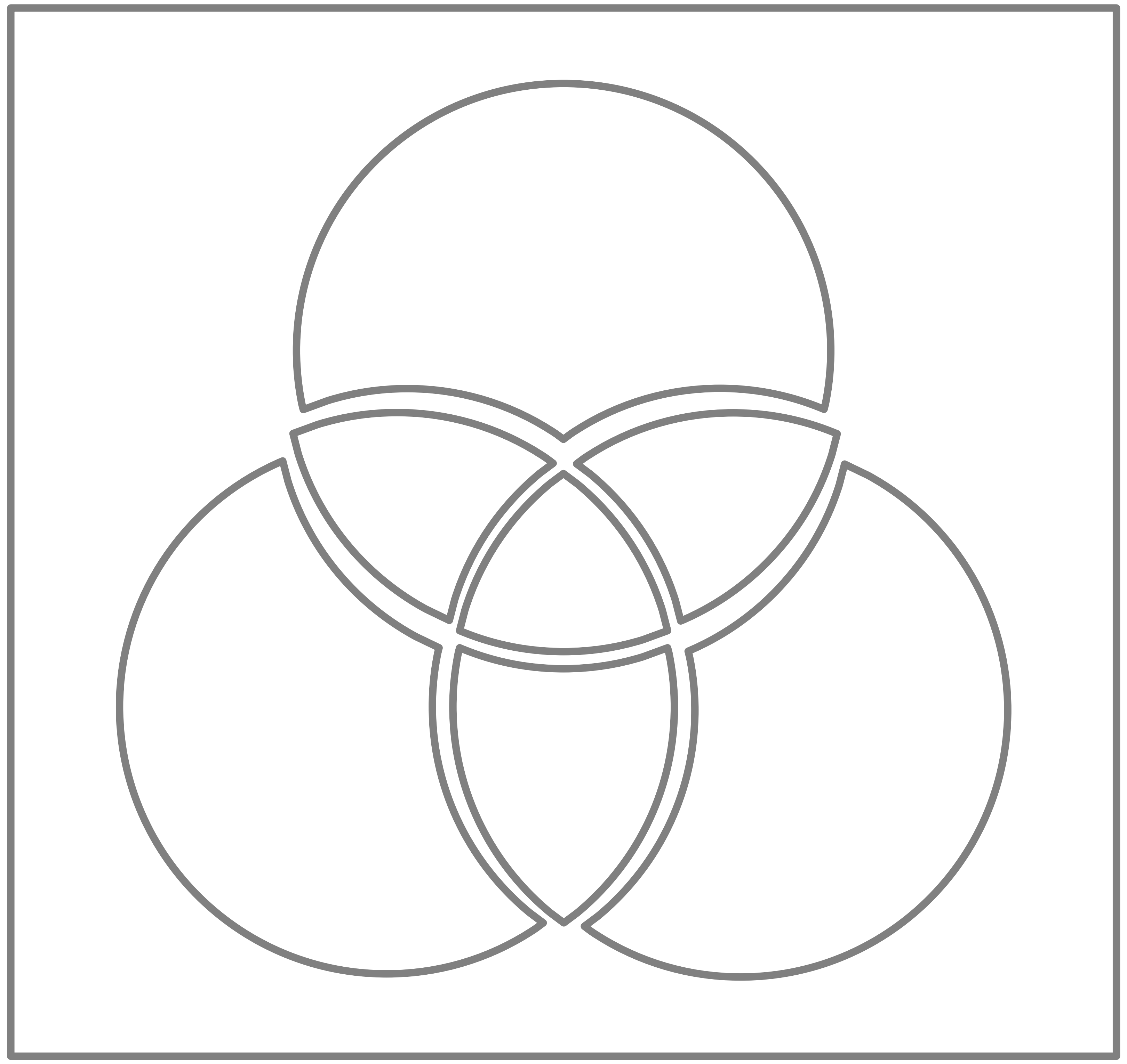
Are many items shared between two sets?
Aggregate By: Degree





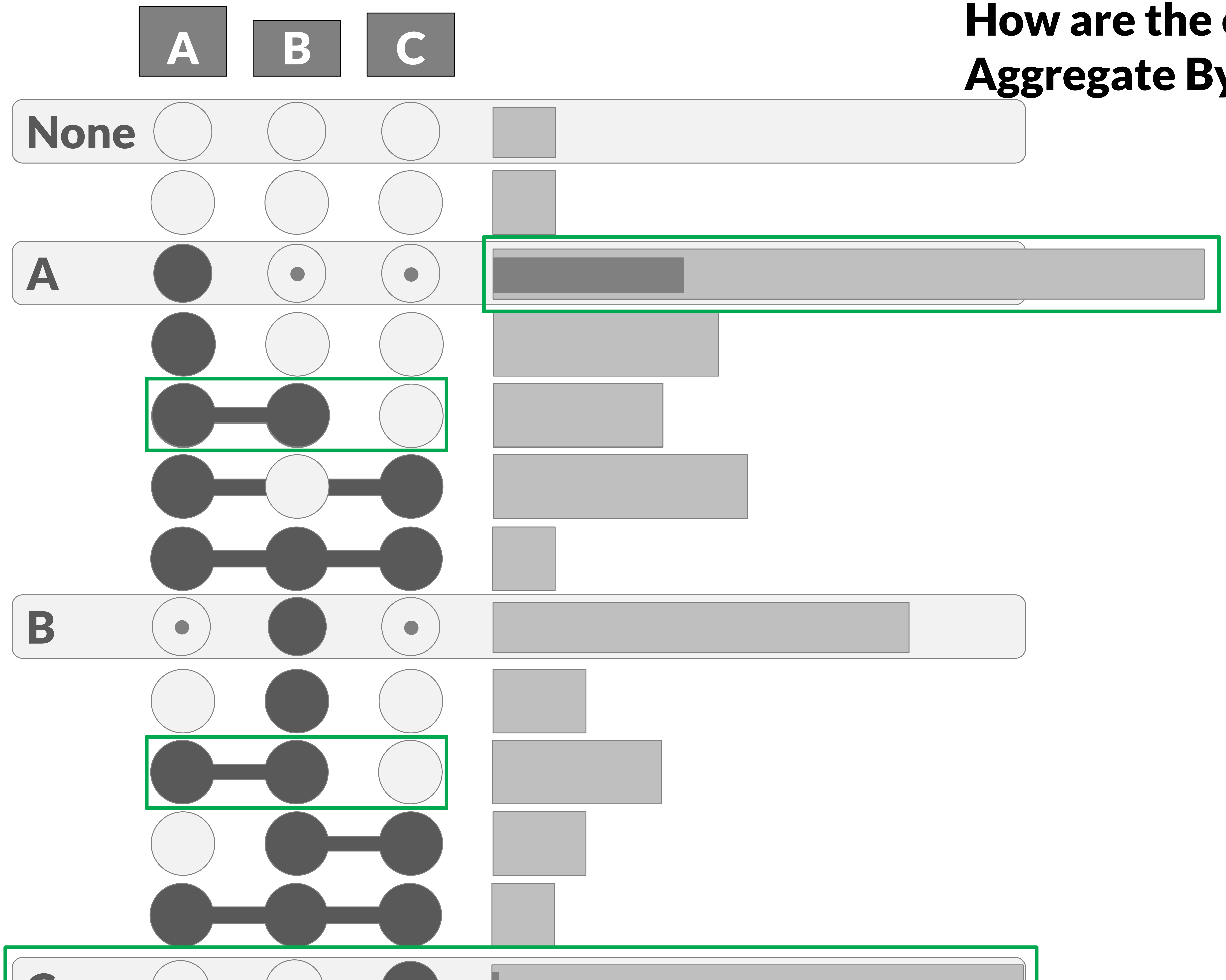
How are the elements of ‘B’ distributed?
Aggregate By: Set





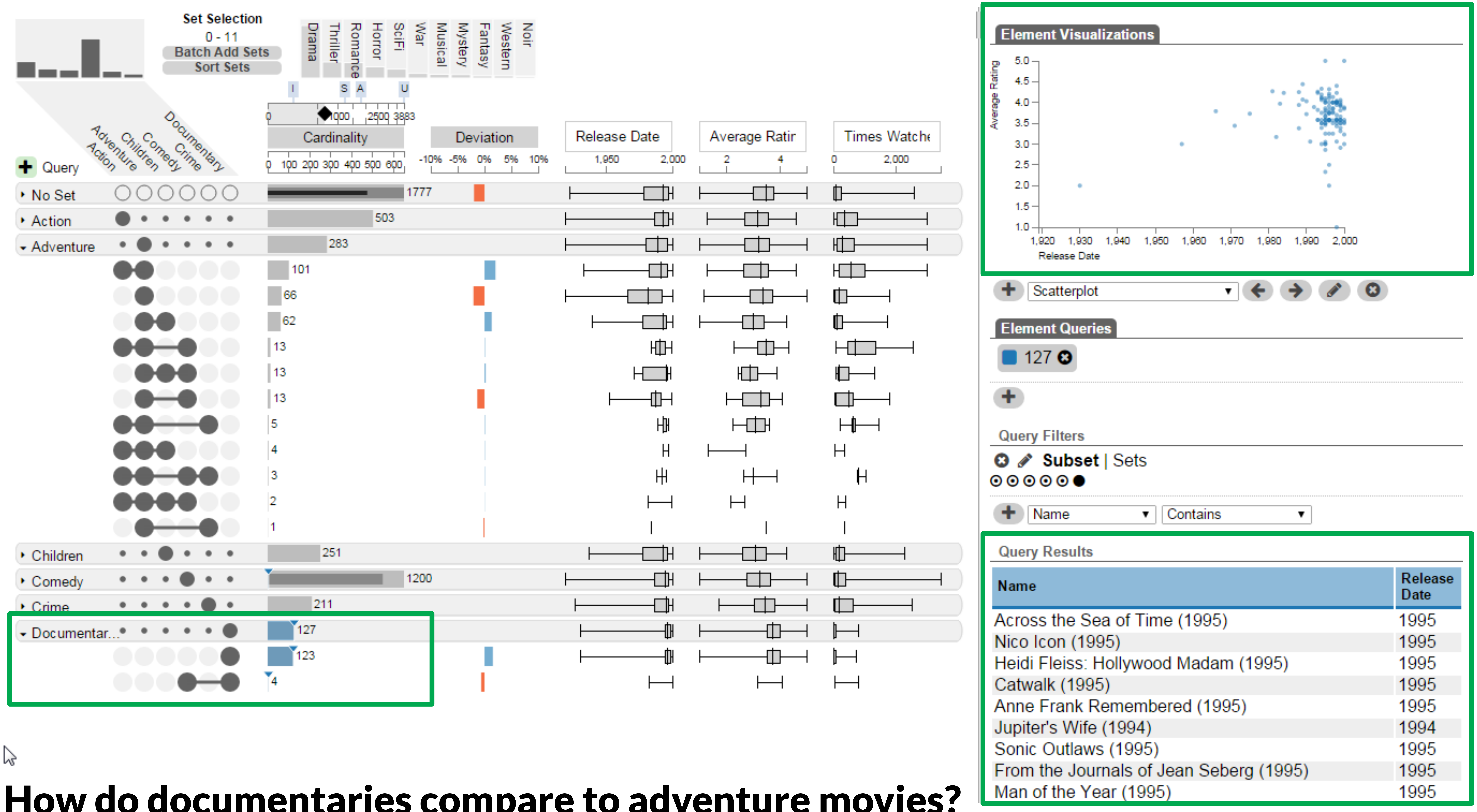
How are the elements of ‘B’ distributed?
Aggregate By: Set

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

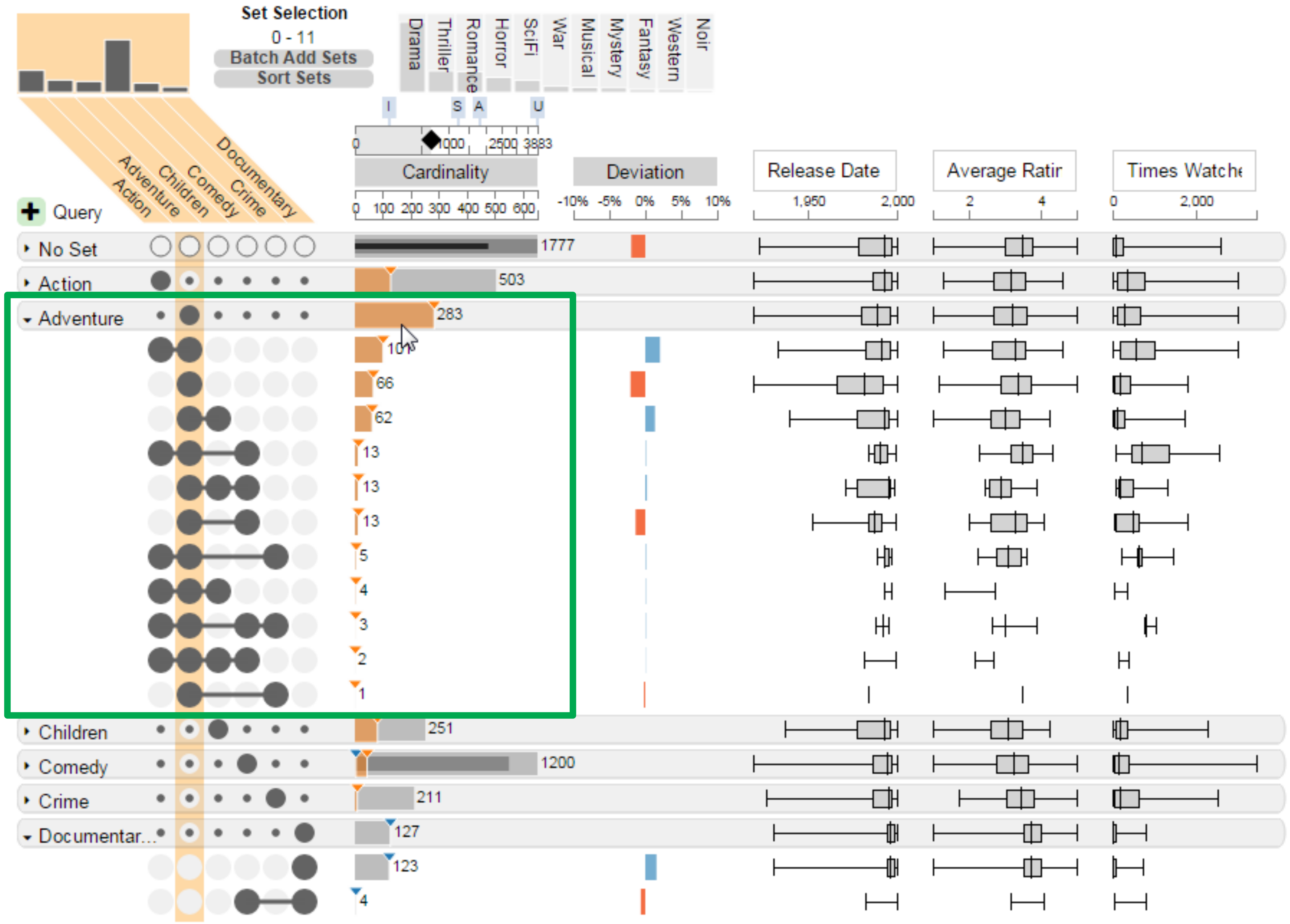


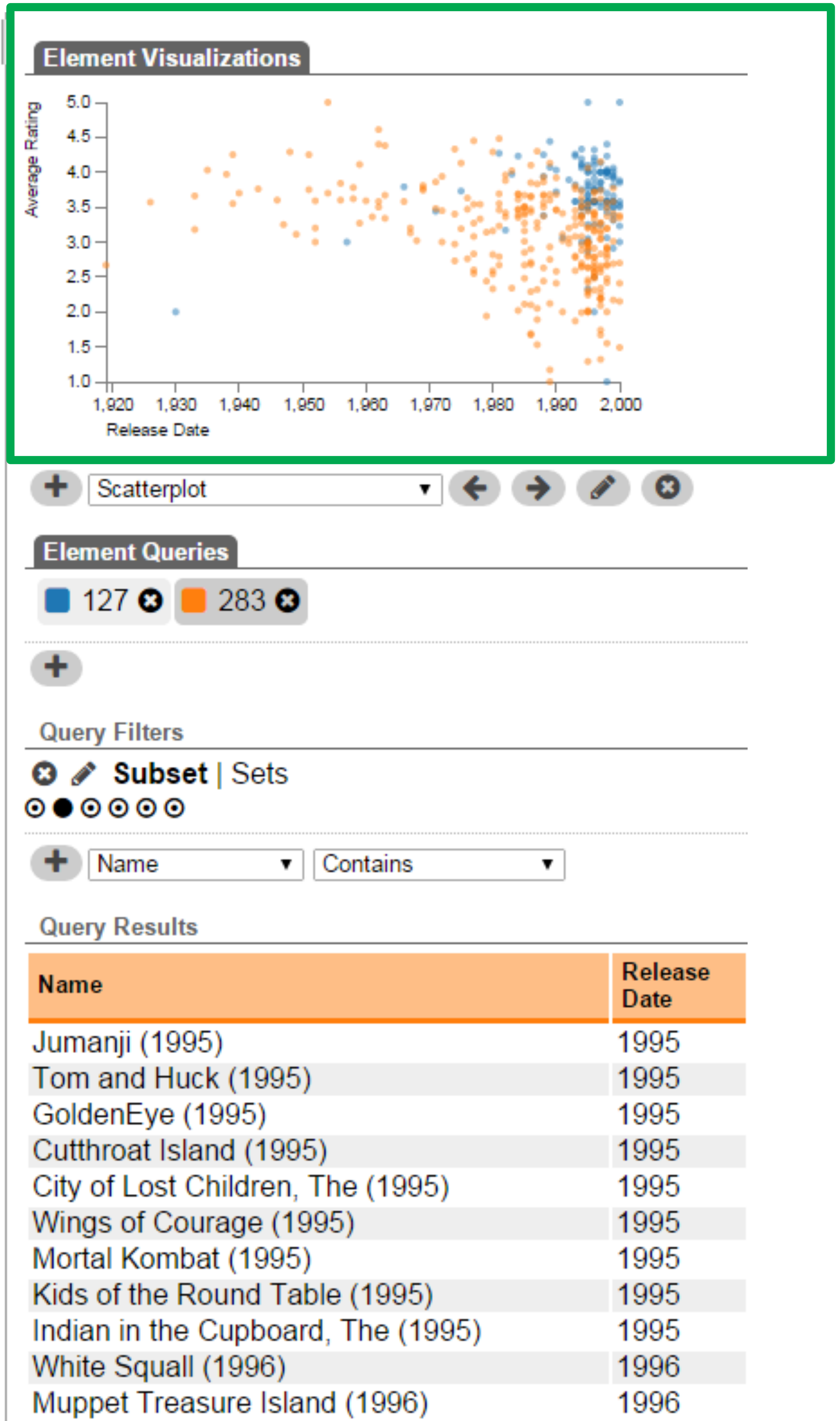

First, aggregate by
Degree
Then, aggregate by
Don't Aggregate
Sort by
Degree
Cardinality
Deviation
Aggregates
Collapse All
Expand All
Row Height
Large
Data
Min Degree:
0
Max Degree:
5
Hide Empty Intersections
Dataset Information
Name: Movies Genres
Sets: 17
Attributes: 6
Elements: 3883
Author: grouplens
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

Elements & Attributes



How do documentaries compare to adventure movies?





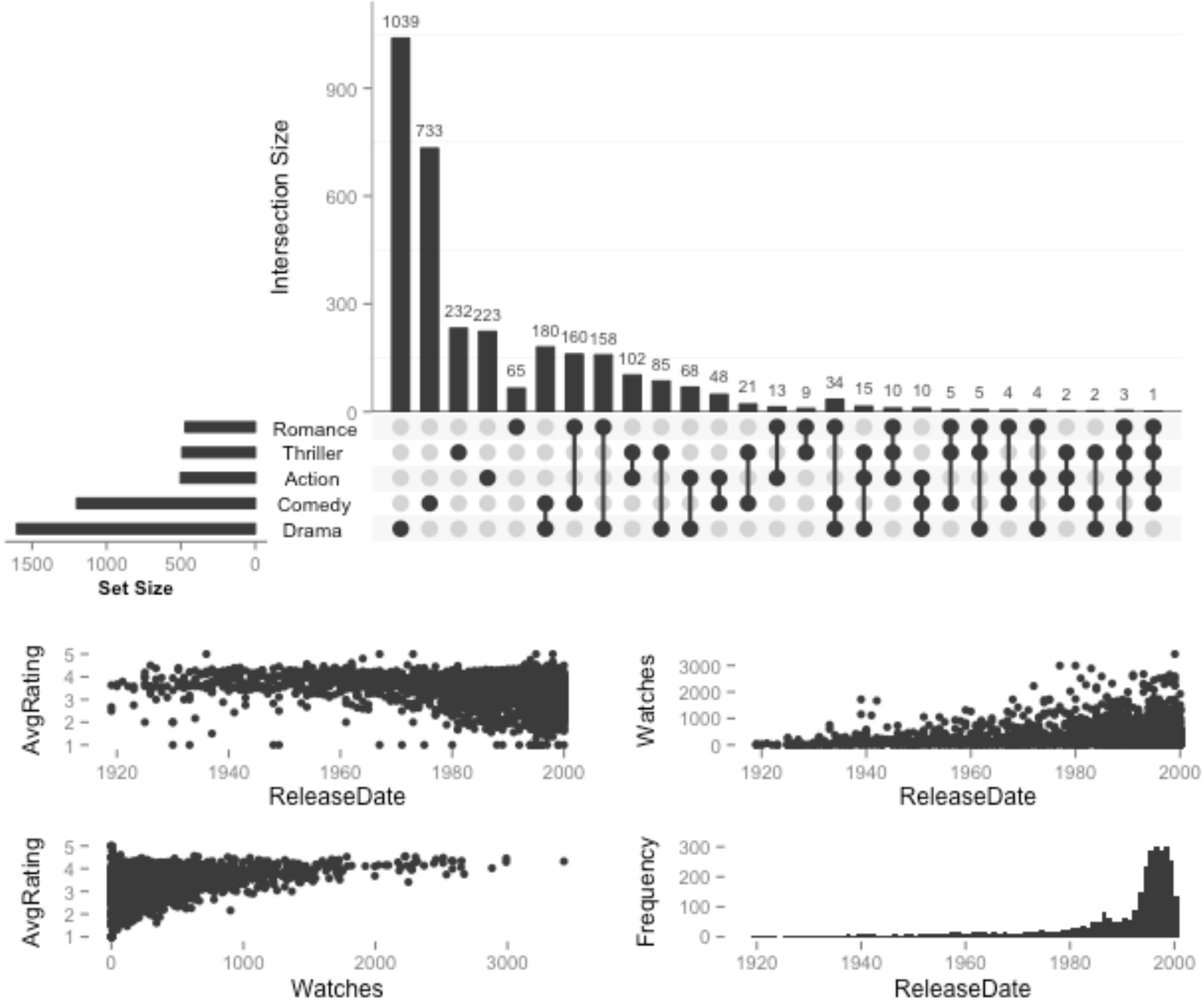
How do documentaries compare to adventure movies?

Applications

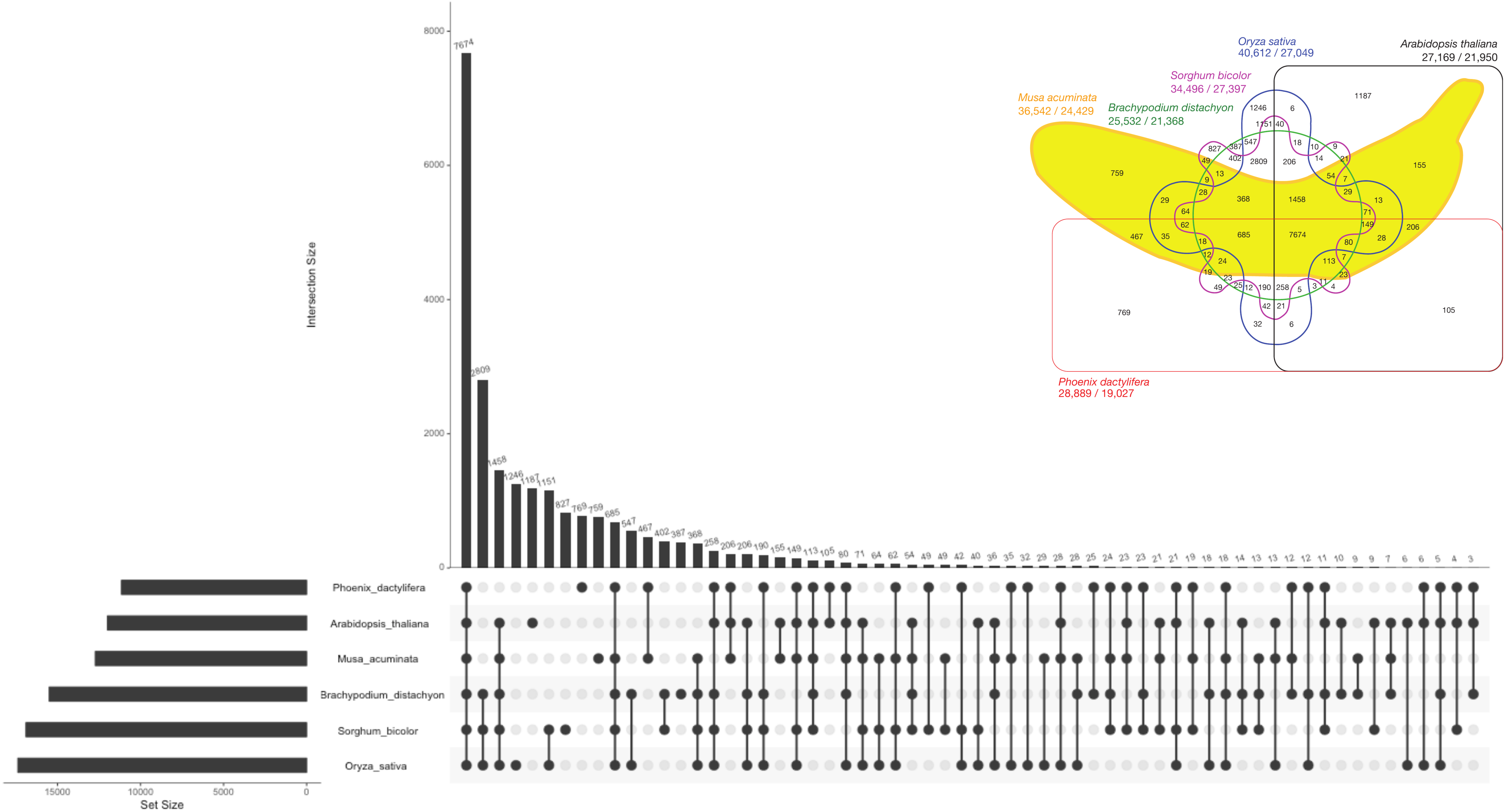
R-Version: UpSetR

Developed at HMS

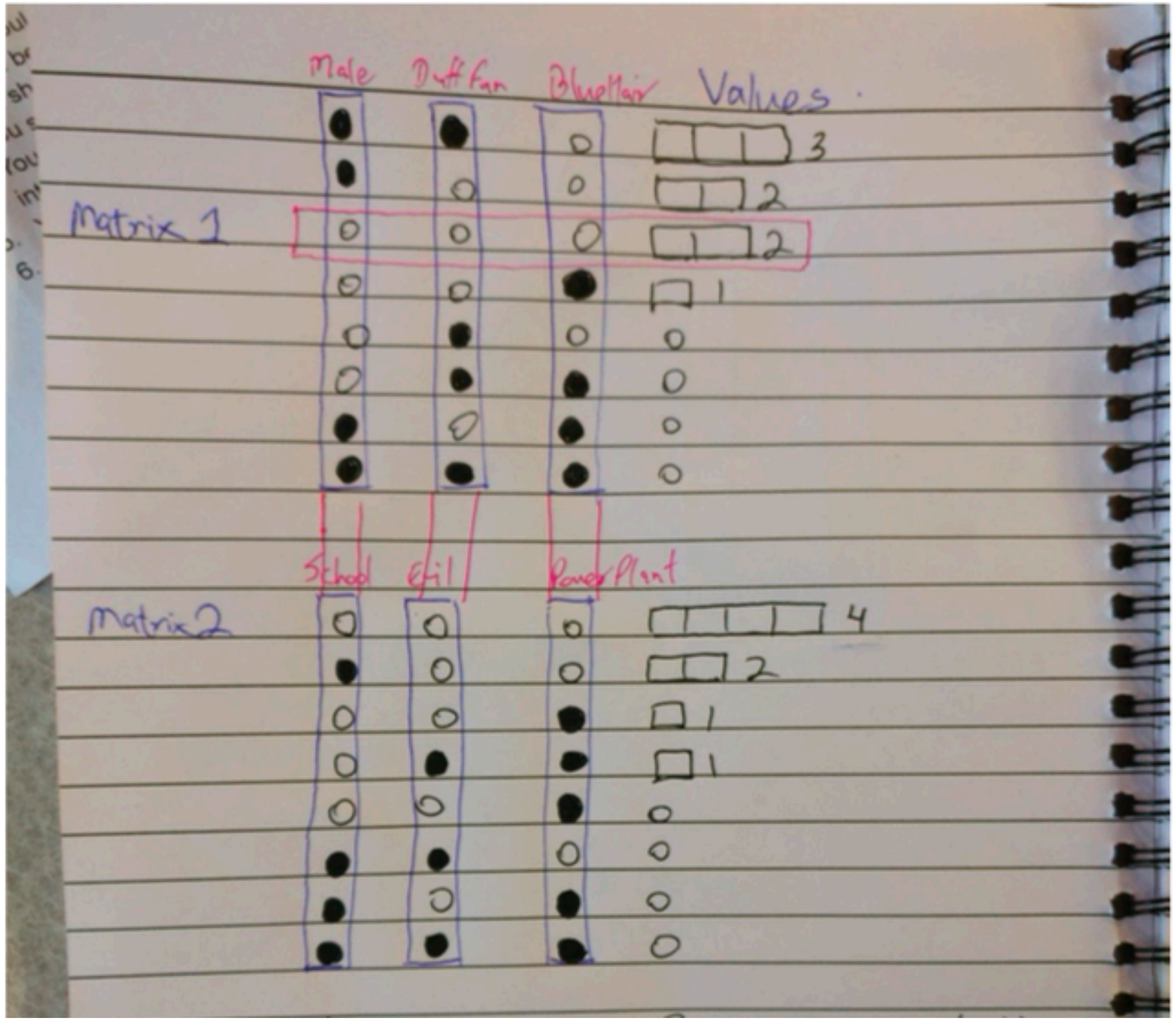
Some design adaptions



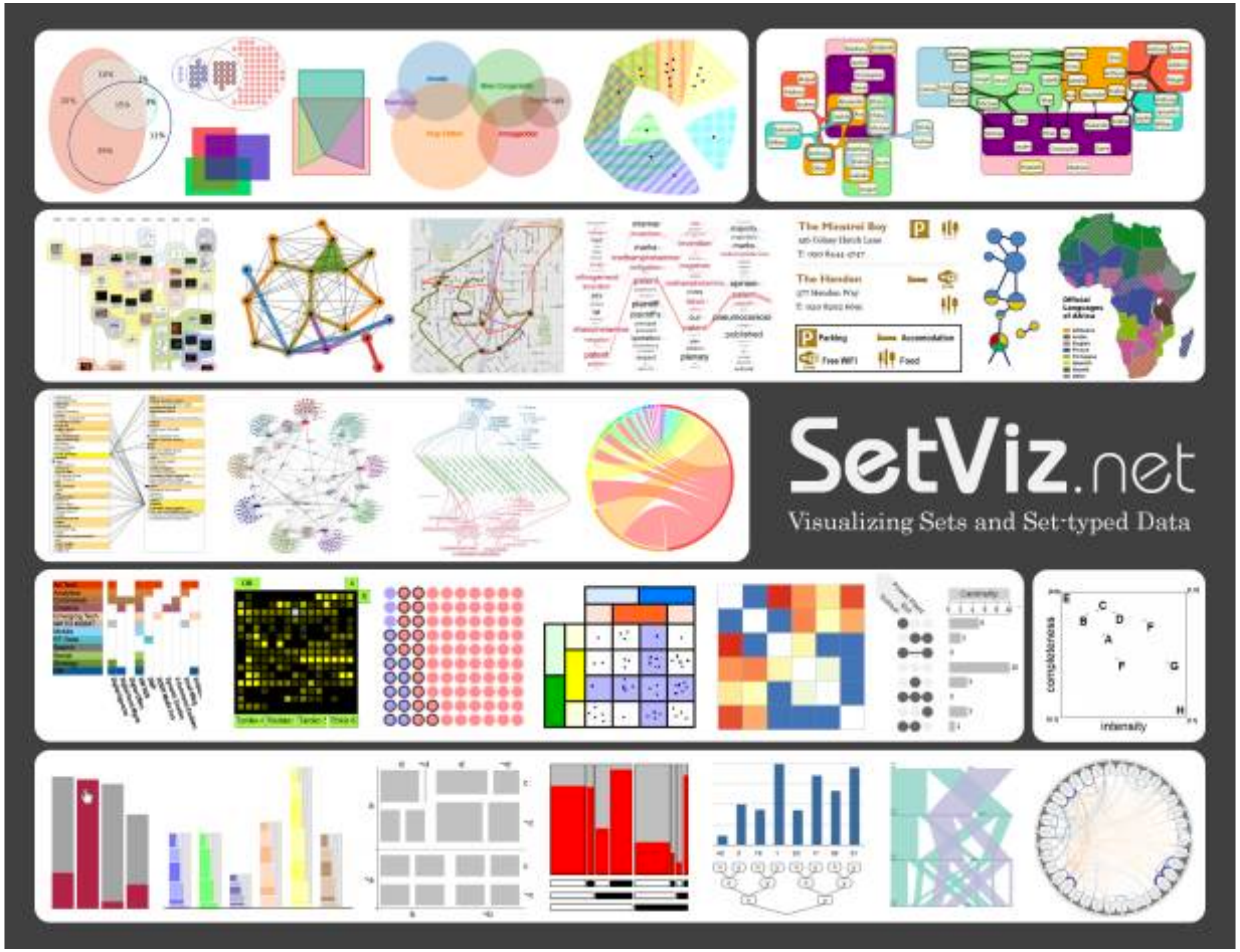
The Banana Chart Redesigned



DESIGN 2


Male
Duff Fan
BlueHair
Values
Matrix 1
3
2
2
1
School
Evil
PowerPlant
Matrix2
4
2
1
1

Other Options



http://setviz.net