

ARViz: Interactive Visualization of Association Rules for RDF Data Exploration

Aline MENIN, Lucie CADOREL, Andrea TETTAMANZI,

Alain GIBOIN, Fabien GANDON et Marco WINCKLER

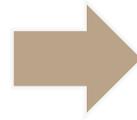
The CovidOnTheWeb Project



Domain Questions

Does coronavirus can cause cancer?

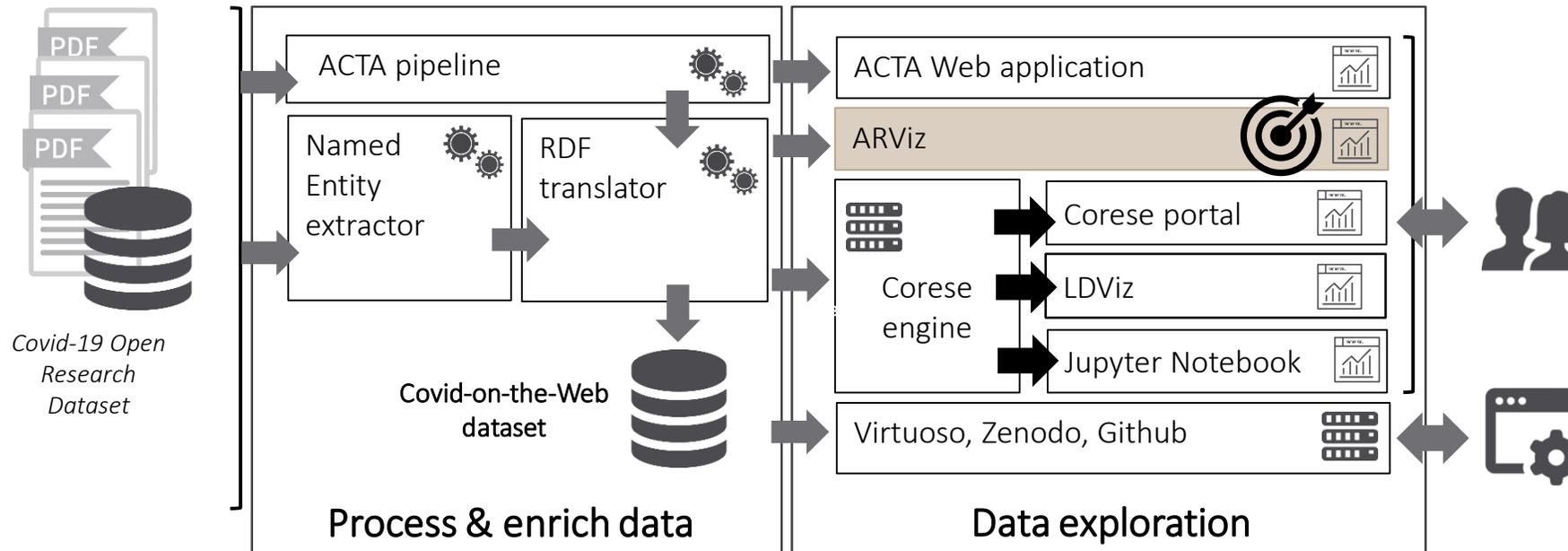
What are the side effects of SARS Cov 1/2 and MERS infections?



Research Questions

What methods to extract knowledge from the literature?

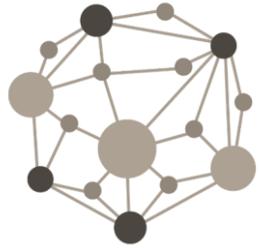
What visual methods to support the exploration of knowledge?



Covid-on-the-Web dataset

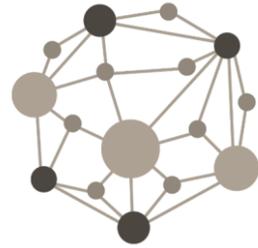


CORD-19 v7
50,000 full-text scholarly articles



CORD-19
Named Entities
Knowledge Graph
(NEKG)

Linked to Dbpedia,
Wikidata and
BioPortal

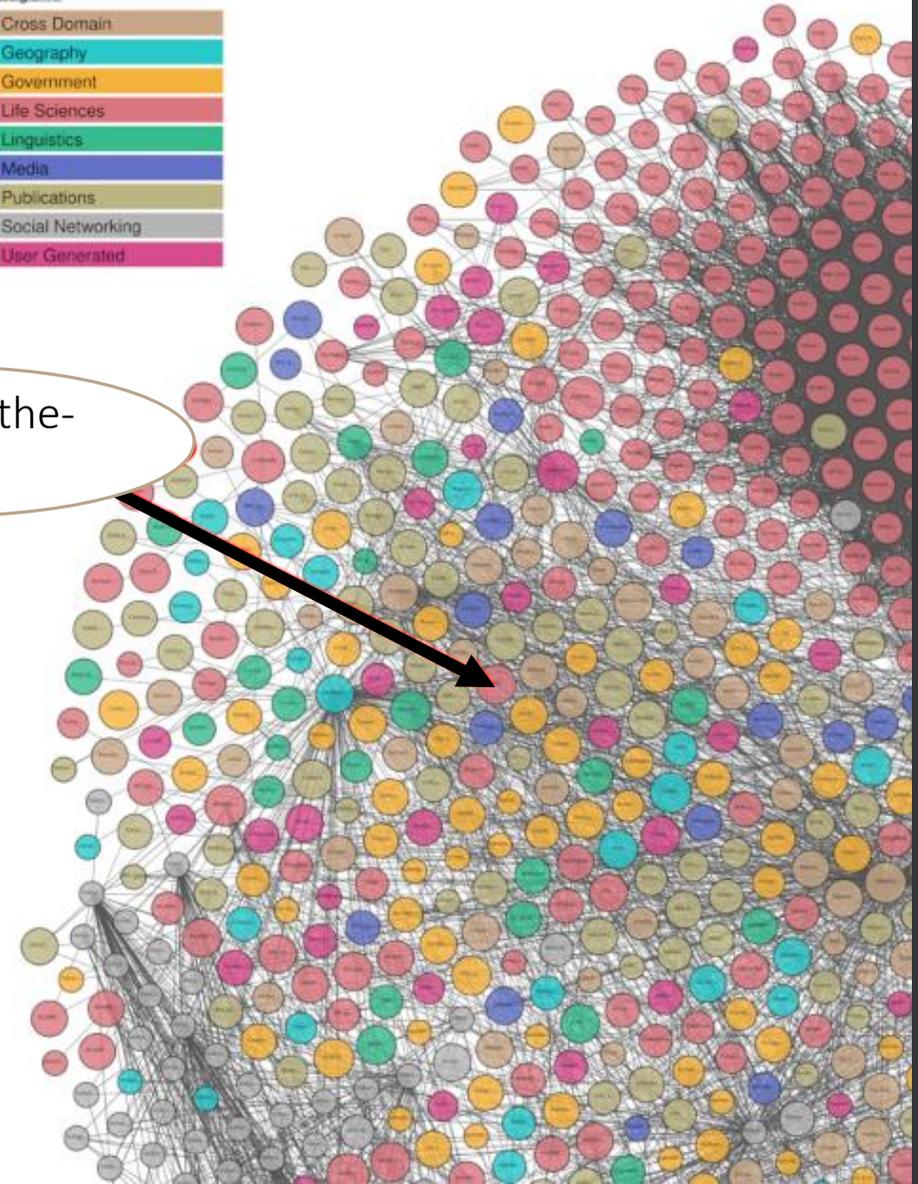


CORD-19
Argumentative
Knowledge Graph
(AKG)

Graph of arguments
linking claims and
evidences



Covid-on-the-Web



Association Rules Mining

Discover interesting correlations, frequent patterns, associations or casual structures among transactions

$$X \rightarrow Y$$

Transactions are publications and items are named entities

Measures of interest

- **Support:** the probability of finding items X and Y in a transaction
- **Confidence:** the probability of finding Y in a transaction, knowing that X is in the same transaction
- **Interestingness:** the serendipity of the rule
- **Symmetry:** whether the rule works inversely

About the Data

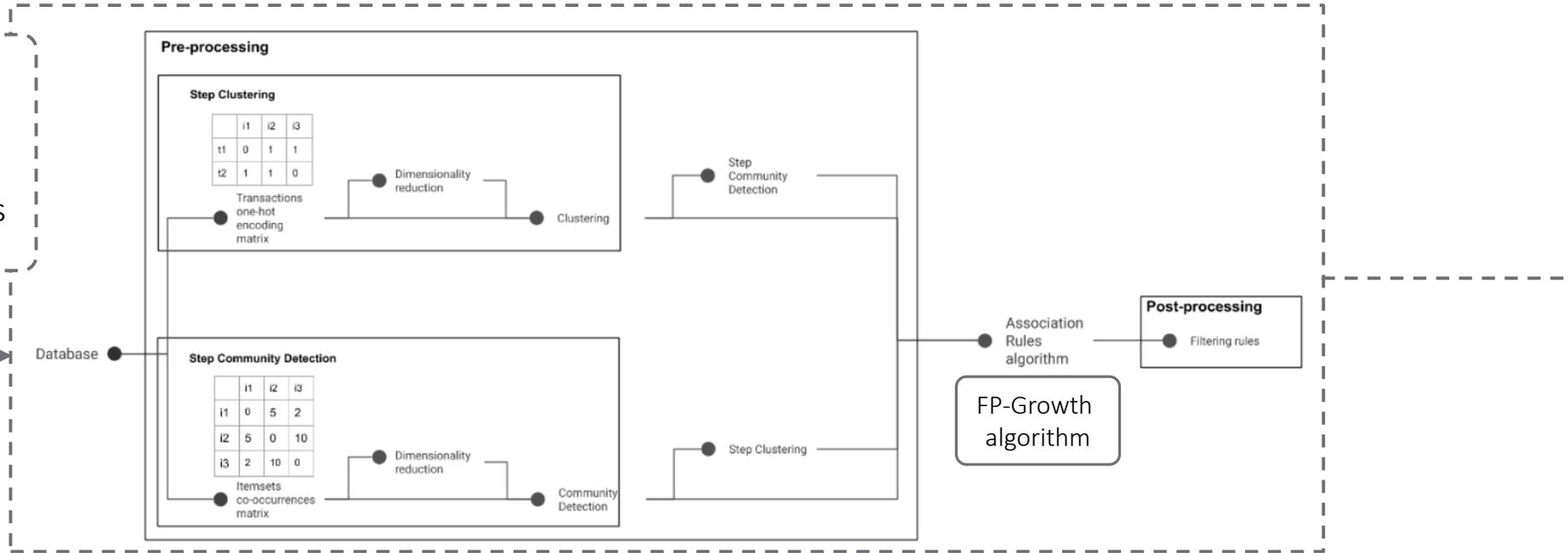
(Cadorel and Tettamanzi, 2020)

Input Dataset

CORD19-NEKG



- Wikidata
- 50,000 publications



ARViz Input Dataset

- 1772 rules
- Confidence (min. 0.7)
- Support
- Interestingness (min. 0.3)
- Symmetry



antecedent	consequent	confidence	interestingness	support	isSymmetric	cluster
ritonavir	lopinavir	0.89	0.59	0.0003414	FALSE	label_cluster2
insect	baculovirus	0.71	0.36	0.0024765	FALSE	label_cluster2 article_cluster1
steroid, osteoblast	saon	0.83	0.83	0.0002134	FALSE	no_clustering
duchenne muscular dystrophy	muscular dystrophy	1	1	0.0052854	TRUE	article_cluster1

ARViz Overview

a) Legend



The rule is
■ Symmetric
■ Not symmetric

b) Data Filtering

Data Filtering

Clusters

- No Clustering
- of papers
- of terms
- of papers and terms

Terms

select from the list

Measures of Interest

Confidence: 0.7 - 1

Interestingness: 0.3 - 1

- Symmetric Rules
- Non-Symmetric Rules

c) Data Sorting

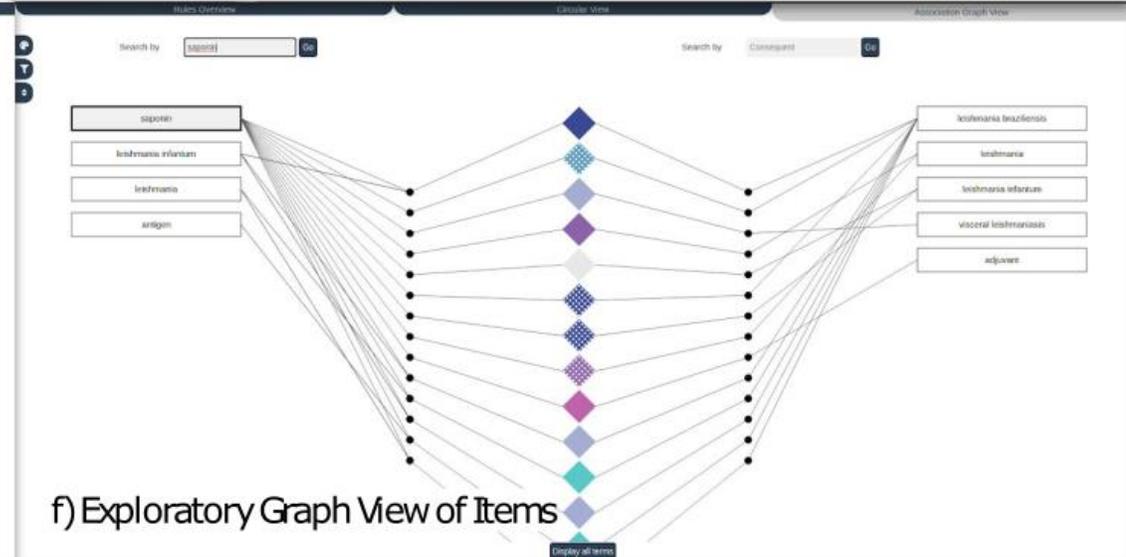
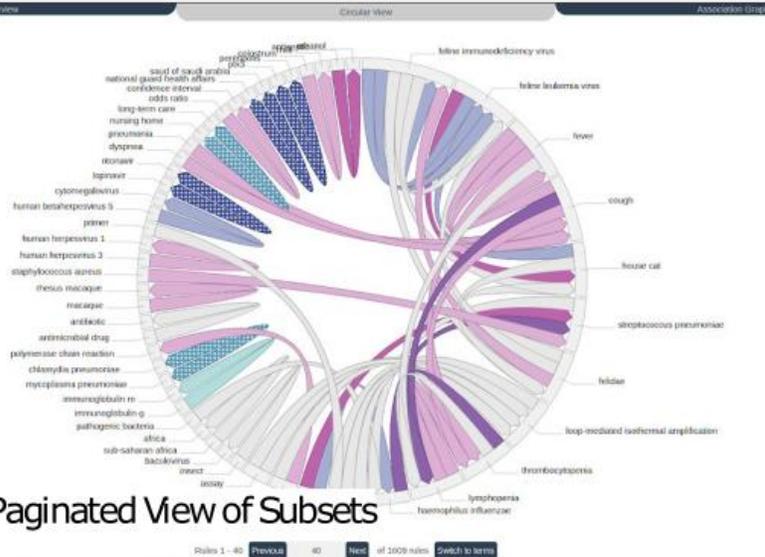
Data Sorting

Sort terms by

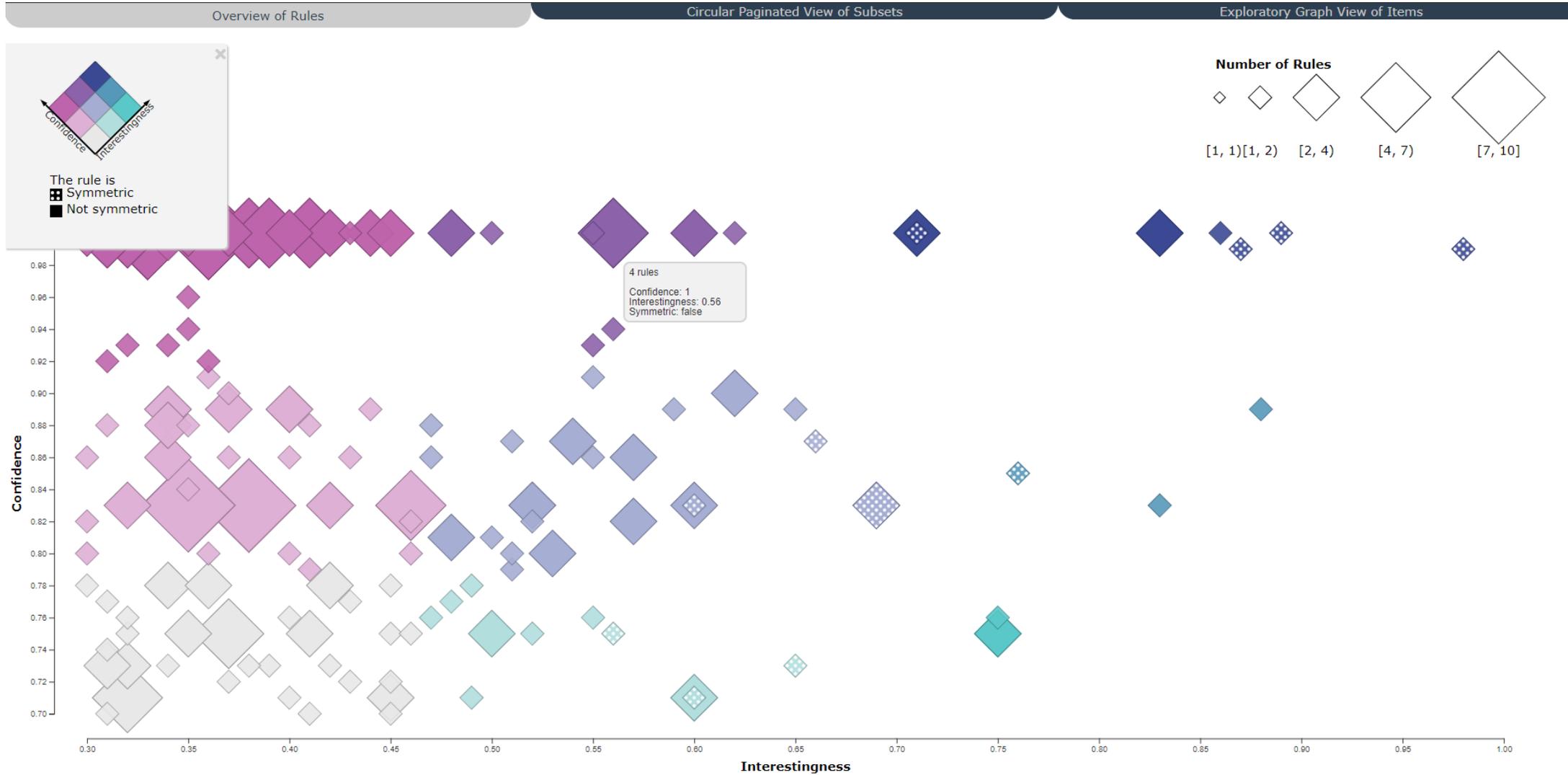
Number of Rules

Sort rules by

None



Overview of Rules

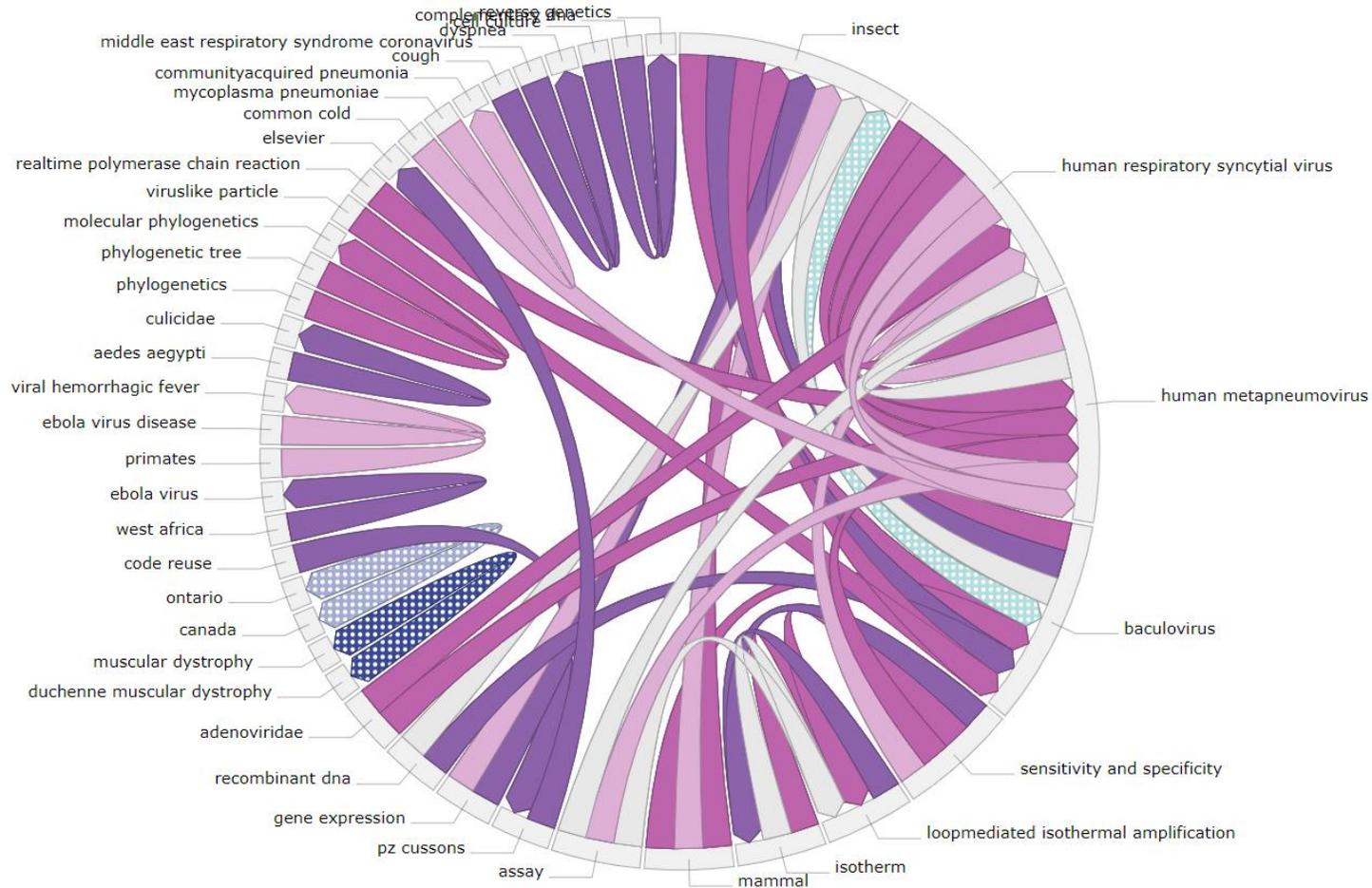


Circular Paginated View of Subsets

Overview of Rules

Circular Paginated View of Subsets

Exploratory Graph View of Items



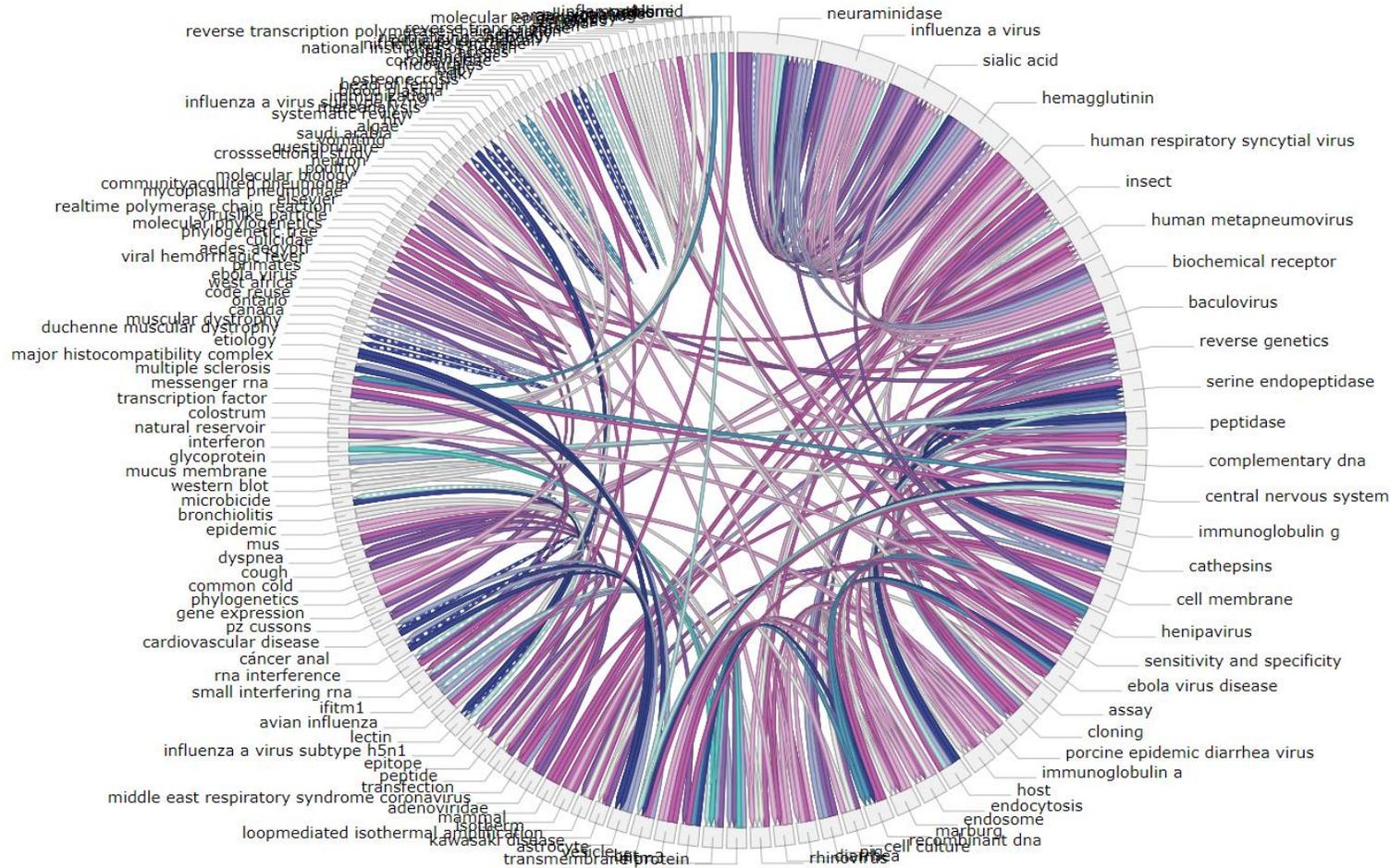
Rules 1 - 30 [Previous](#) 30 [Next](#) of 882 rules [Switch to terms](#)

Circular Paginated View of Subsets

Overview of Rules

Circular Paginated View of Subsets

Exploratory Graph View of Items



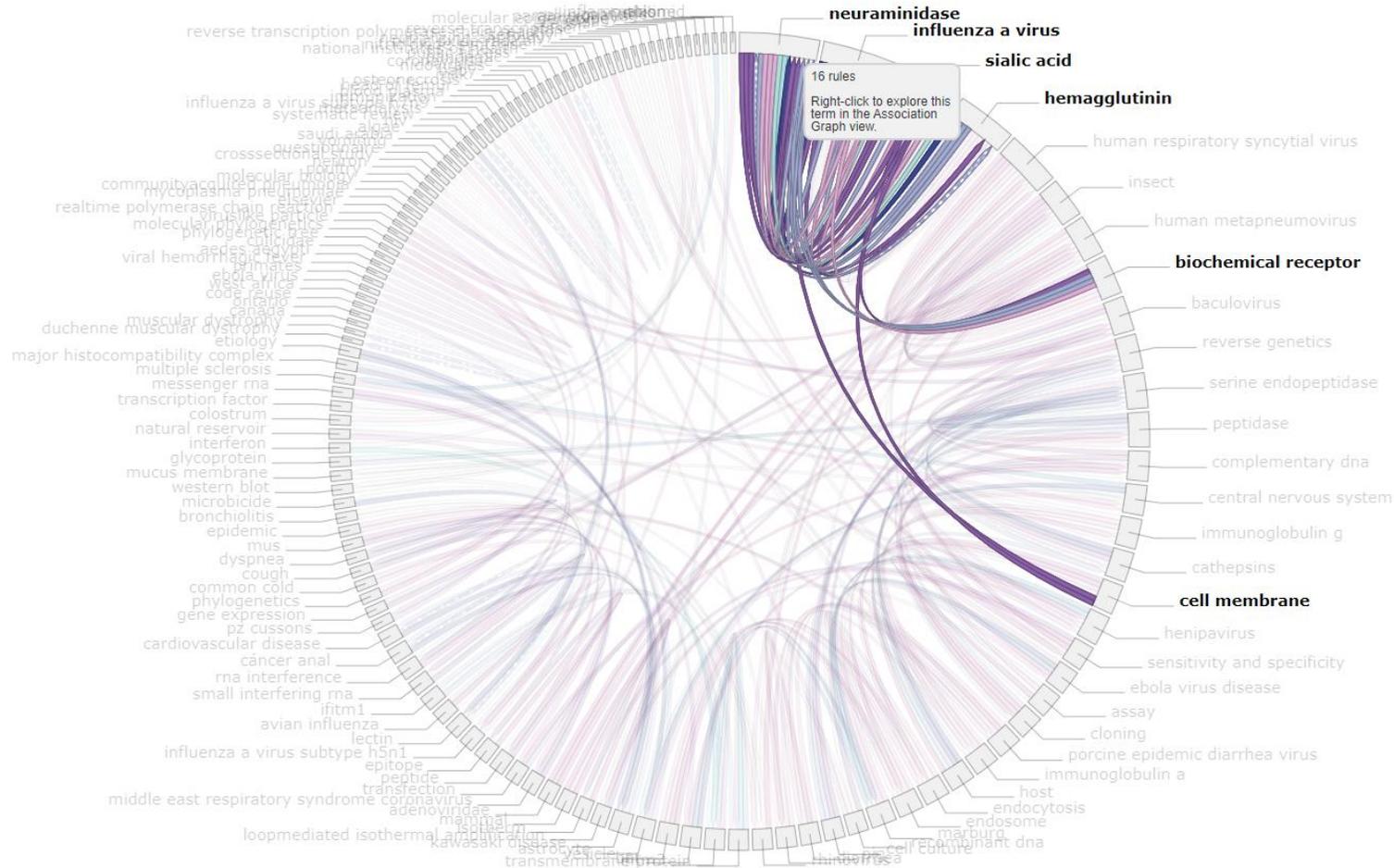
Rules 1 - 150 [Previous](#) 150 [Next](#) of 882 rules [Switch to terms](#)

Circular Paginated View of Subsets

Overview of Rules

Circular Paginated View of Subsets

Exploratory Graph View of Items



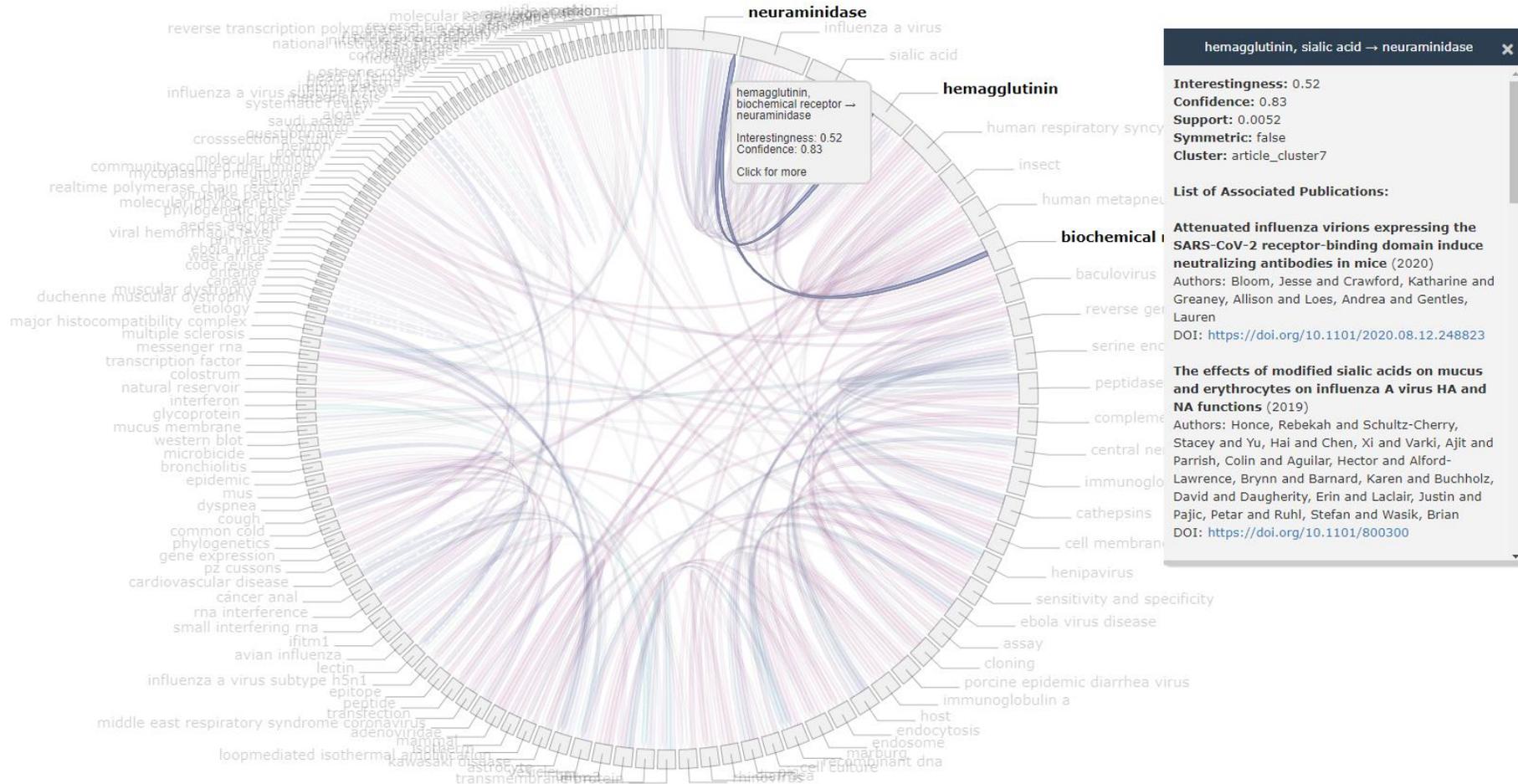
Rules 1 - 150 [Previous](#) 150 [Next](#) of 882 rules [Switch to terms](#)

Circular Paginated View of Subsets

Overview of Rules

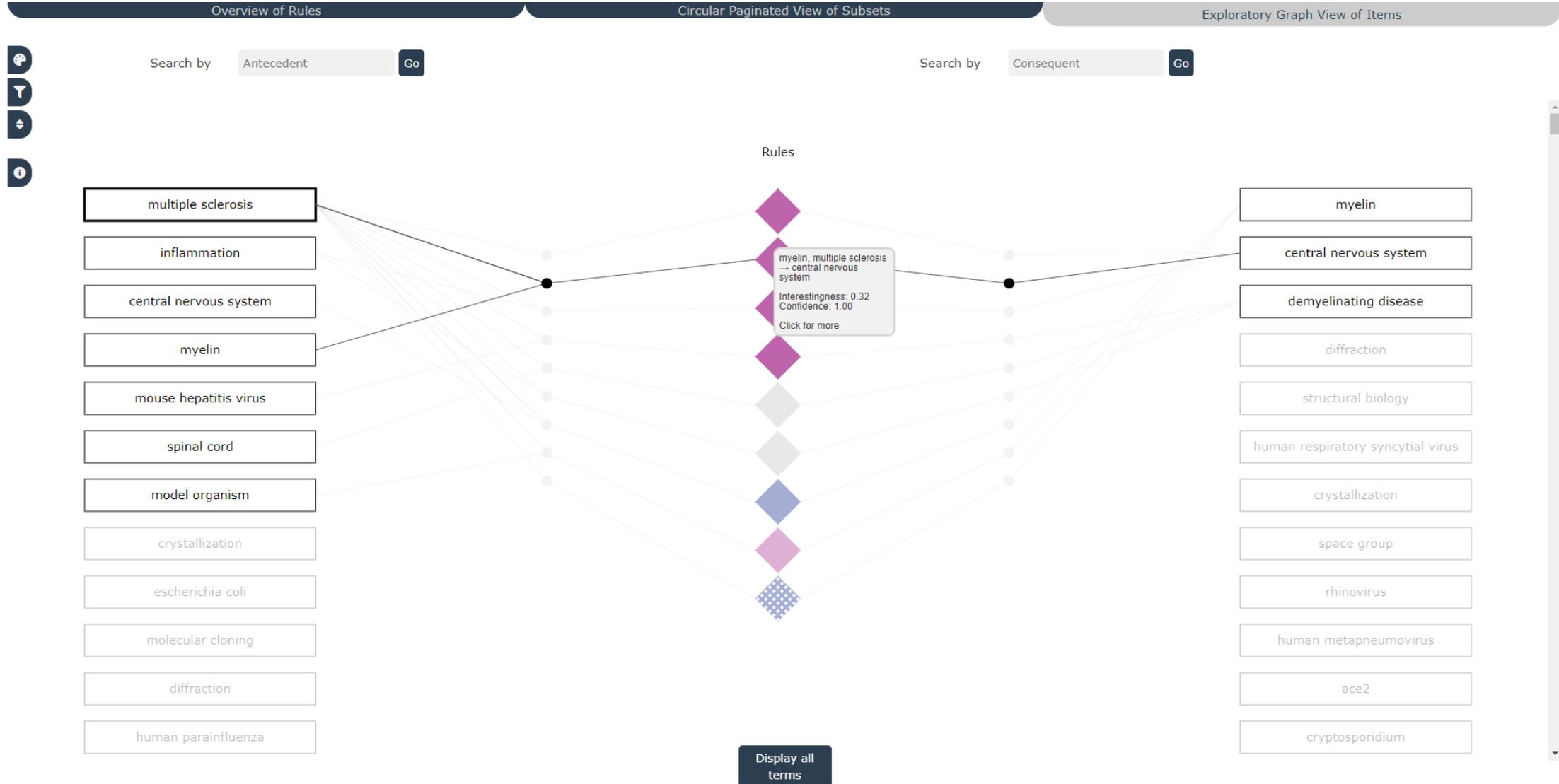
Circular Paginated View of Subsets

Exploratory Graph View of Items



Rules 1 - 150 [Previous](#) 150 [Next](#) of 882 rules [Switch to terms](#)

Exploratory Graph View of Items



Exploratory Graph View of Items

Overview of Rules Circular Paginated View of Subsets Exploratory Graph View of Items

Search by Antecedent Search by Consequent

Rules

multiple sclerosis
inflammation
central nervous system
myelin
mouse hepatitis virus
spinal cord
model organism
crystallization
escherichia coli
molecular cloning
diffraction
human parainfluenza

myelin
central nervous system
demyelinating disease
diffraction
structural biology
human respiratory syncytial virus
crystallization
space group
rhinovirus
human metapneumovirus
ace2
cryptosporidium

myelin, multiple sclerosis → central nervous system

Interestingness: 0.32
Confidence: 1.00
Support: 0.0083
Symmetric: false
Cluster: article_cluster12

List of Associated Publications:

Azadirachta indica A. Juss Ameliorates Mouse Hepatitis Virus-Induced Neuroinflammatory Demyelination by Modulating Cell-to-Cell Fusion in an Experimental Animal Model of Multiple Sclerosis (2020)
Authors: Das Sarma, Jayasri and Sato, Fumitaka and Safriyu, Abass and Sarkar, Lucky and Kiran Putchala, Ravi and Kujawska, Malgorzata
DOI: <https://doi.org/10.3389/fncel.2020.00116>

Neuroinflammation in the normal appearing white matter of multiple sclerosis brain causes abnormalities at the node of Ranvier (2020)
Authors: Reynolds, Richard and James, Rachel and Browne, Eleanor and Delgado, Patricia and Faisal, A and Howell, Owain and Mazarakis, Nicholas and Meng, Joanna and Picon, Carmen and Umashankar, Swetha
DOI: <https://doi.org/10.1101/2020.06.10.142281>

Display all terms

Exploratory Graph View of Items

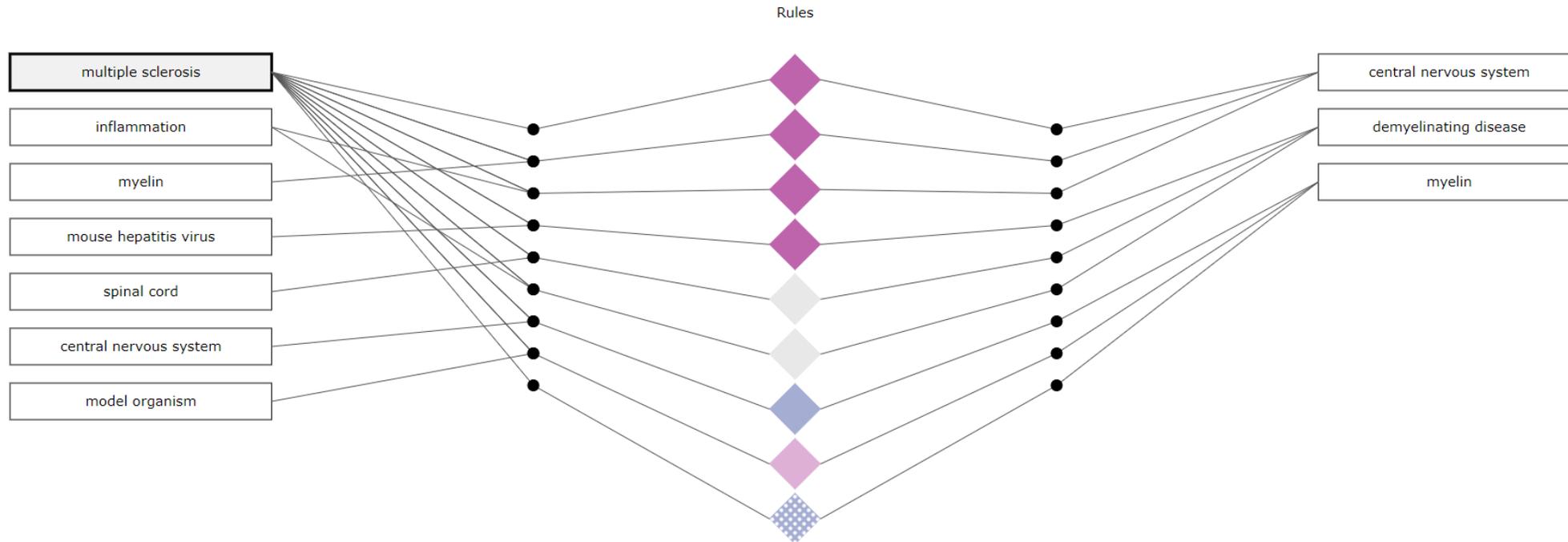
Overview of Rules

Circular Paginated View of Subsets

Exploratory Graph View of Items

Search by

Search by



Display all terms

Evaluation



- Semi-structured interview
- Videoconference



- Two participants
- +15 year of experience
- Researchers on data mining and biomedicine



- Usability
 - SUS questionnaire
 - NASA-TLX
- Usefulness

Exploratory Tasks

Given item i , identify which other terms are related to it.

Identify the items(s) with the most association rules

Recover the list of publications that mention a particular list of terms (e.g., cough, dyspnea and fever)

Identify the rules with the highest confidence

Results



Workload 🙄
55 and 46.7 points over 100

Usability 😊
85 and 87.5 points over 100



- It was easy to solve the tasks using Arviz
- It is important to display
 - the relationship between items
 - the symmetry of rules
 - the source of data
- The link between views ease the exploration
- The interactive prompting helps to extend and enrich the analysis



- Most people would not learn to use the system quickly
- The concepts of data mining could hinder the usage for people in different domains (e.g. biomedicine)
- Difficult to understand the usage of clusters

Conclusion and Future Work



Three complementary visualization techniques

Representation of

- measures of interest (confidence and interestingness)
- symmetry
- relationship between items

Interactive prompting

Access to data source



Resolution of item- and rule-based tasks



Effective tool to assist the exploration of association rules
Feasible to solve domain-related tasks



Studies with a larger sample of expert users



Generalizable to any dataset of association rules



Redesign of visualization pipeline to support different RDF datasets

ARViz: Interactive Visualization of Association Rules for RDF Data Exploration

Aline MENIN, Lucie CADOREL, Andrea TETTAMANZI,
Alain GIBOIN, Fabien GANDON et Marco WINCKLER

Available at <http://covid19.i3s.unice.fr:8080/arviz/>