

# PROPOSING RICH VIEWS OF LINKED OPEN DATA SETS THE S-PATHS PROTOTYPE AND THE VISUALIZATION OF FRBR-IZED DATA IN DATA.BNF

27/11/2019

Raphaëlle Lapôtre, Marie Destandau & Emmanuel Pietriga

**{BnF}** | Bibliothèque  
nationale de France

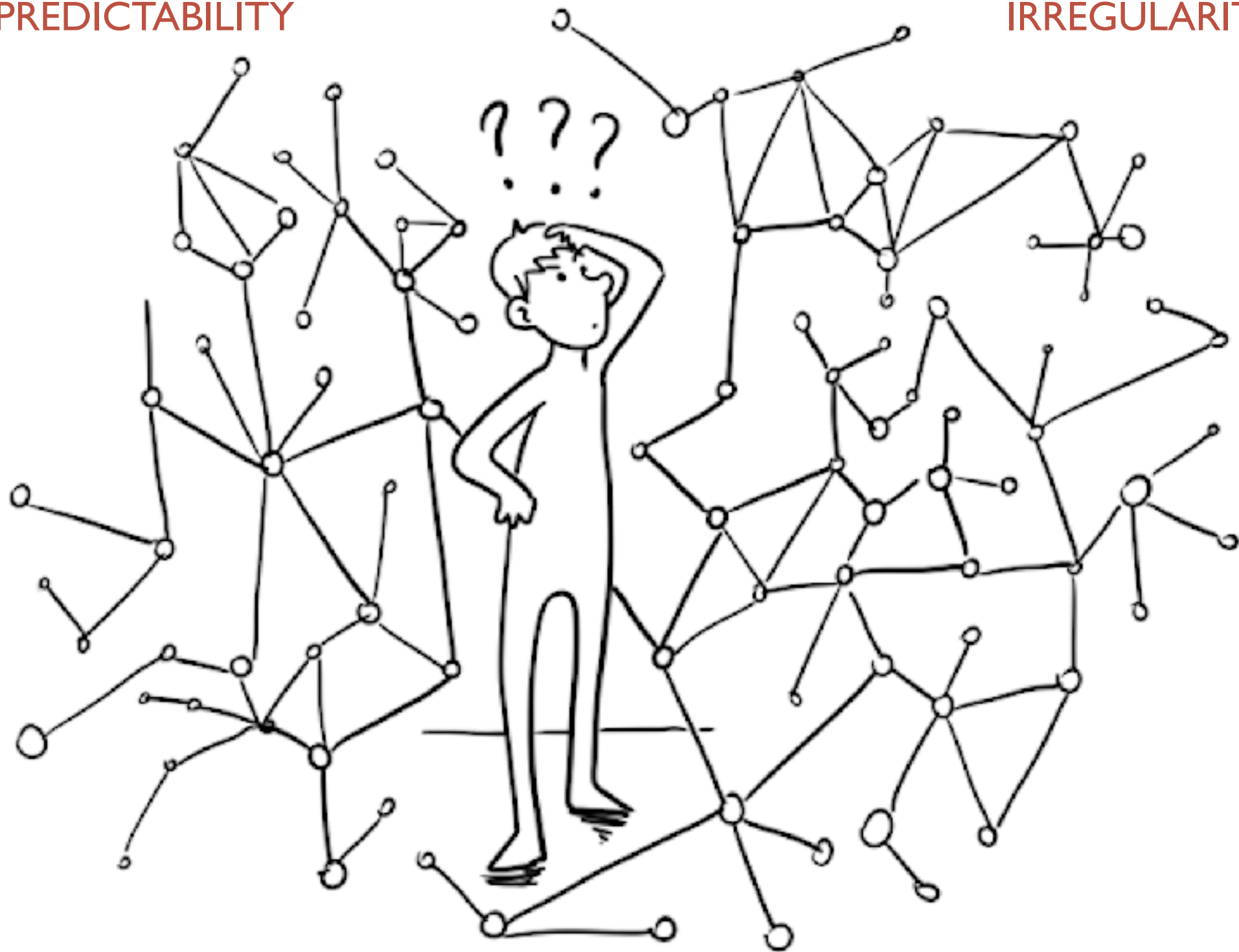
*data*.bnf.fr



*inria*  
informatiques mathématiques

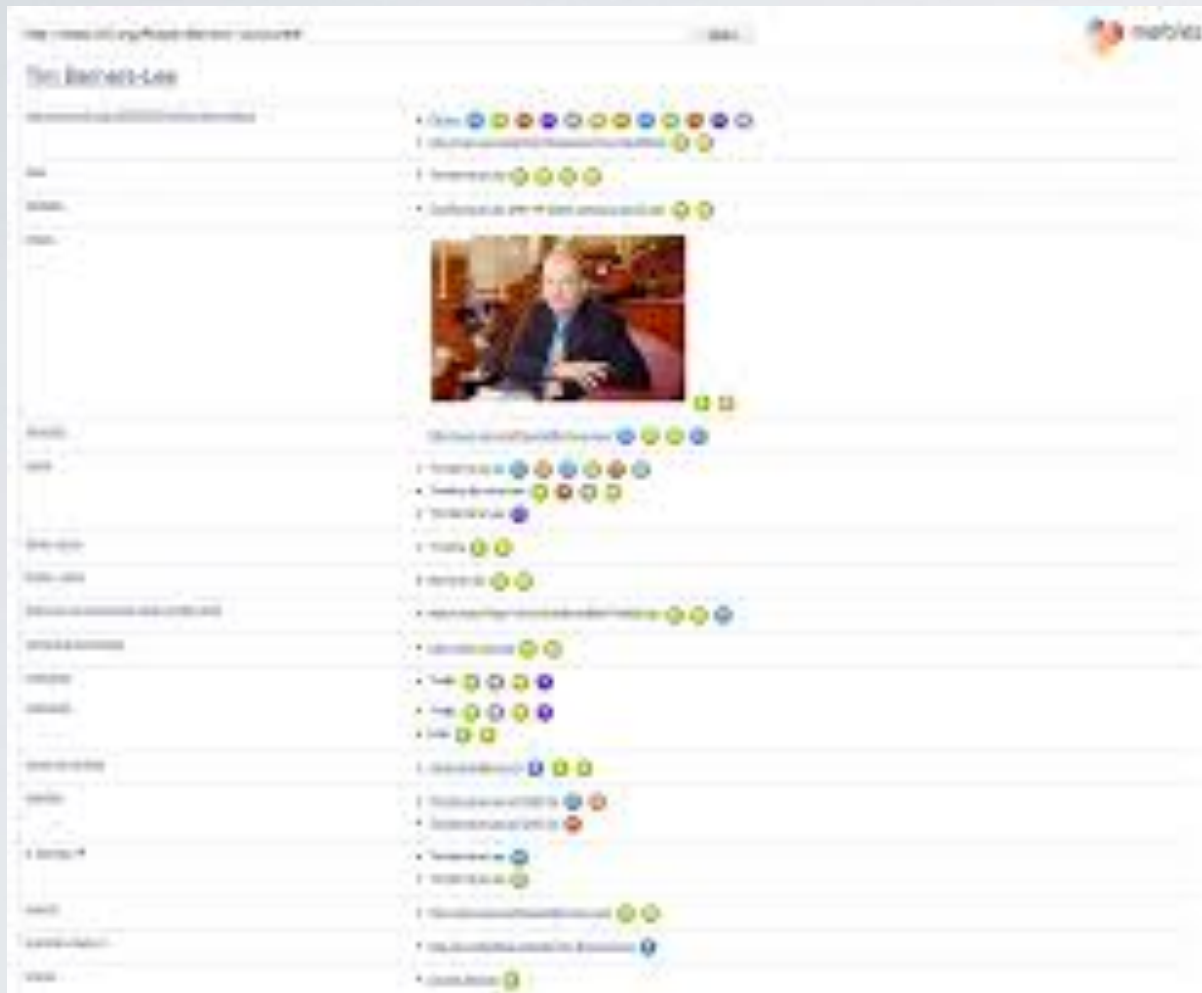
UNPREDICTABILITY

IRREGULARITY



VOLUME

# FOLLOW-YOUR-NOSE BROWSERS



## Marble browser



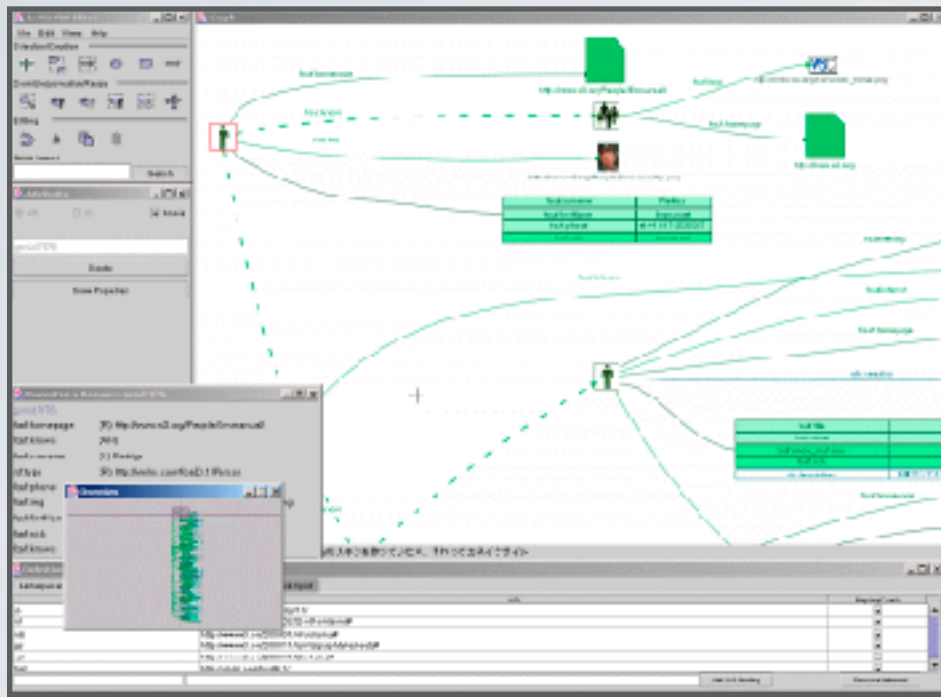
Brownsauce browser



Figure 1: Screenshots of LENA showing different lenses.

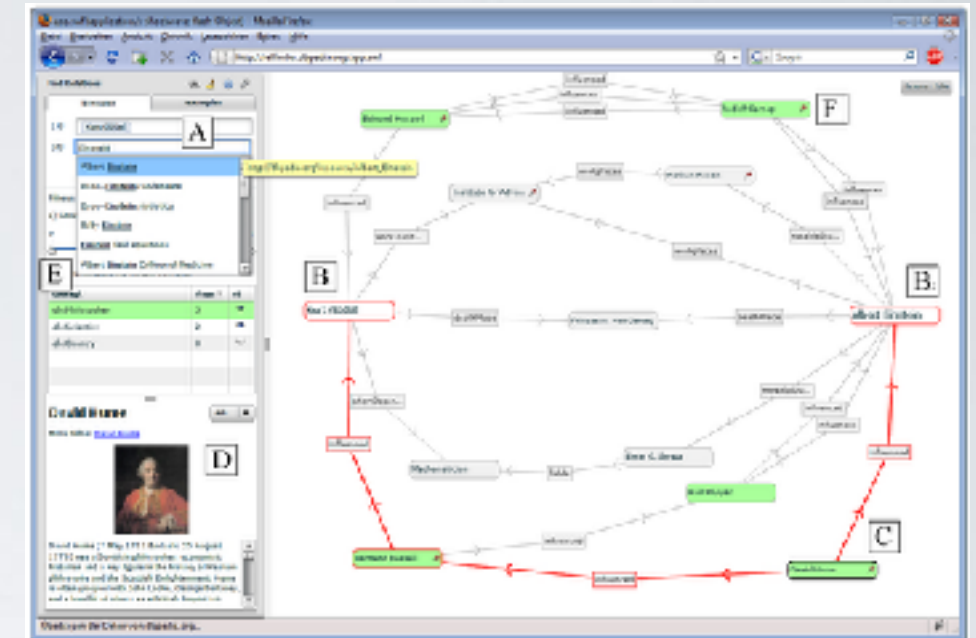
Koch, J., Franz, T., & Staab, S. (2008, October). LENA-Browsing RDF Data More Complex Than Foaf. In *International Semantic Web Conference (Posters & Demos)*.

# NODE-LINK DIAGRAMS

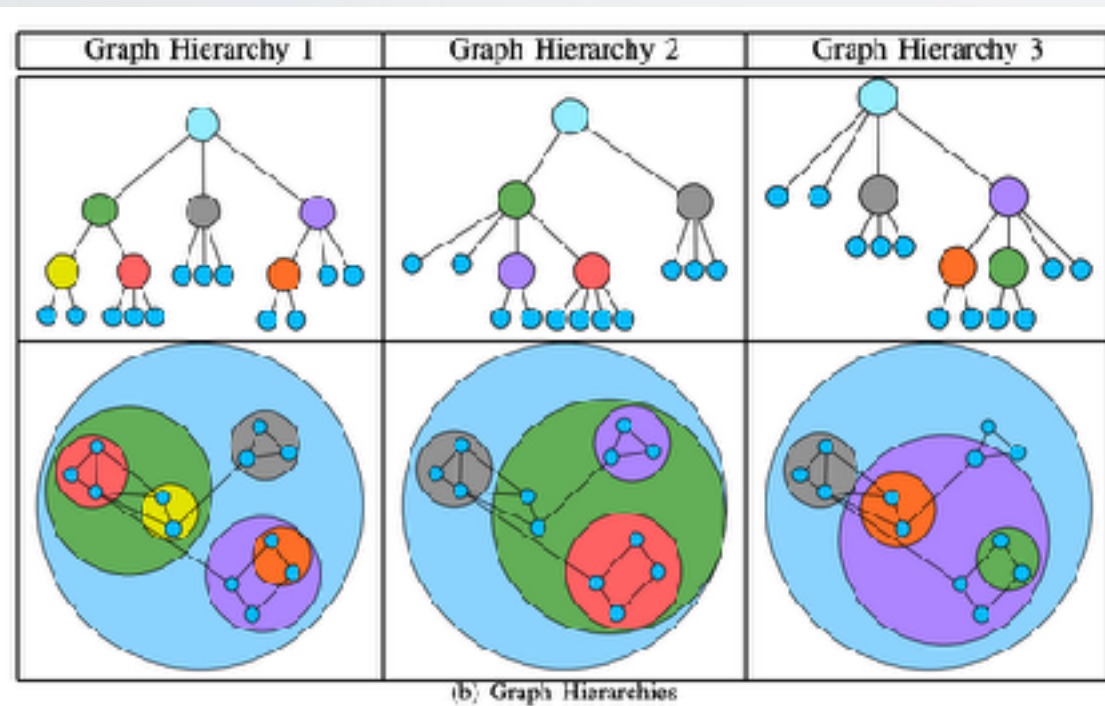


Pietriga, Emmanuel. "Isaviz, a visual environment for browsing and authoring rdf models." In *Eleventh International World Wide Web Conference Developers Day, 2002*. 2002.

Pietriga, Emmanuel. "Semantic web data visualization with graph style sheets." In *Proceedings of the 2006 ACM symposium on Software visualization*, pp. 177-178. ACM, 2006.



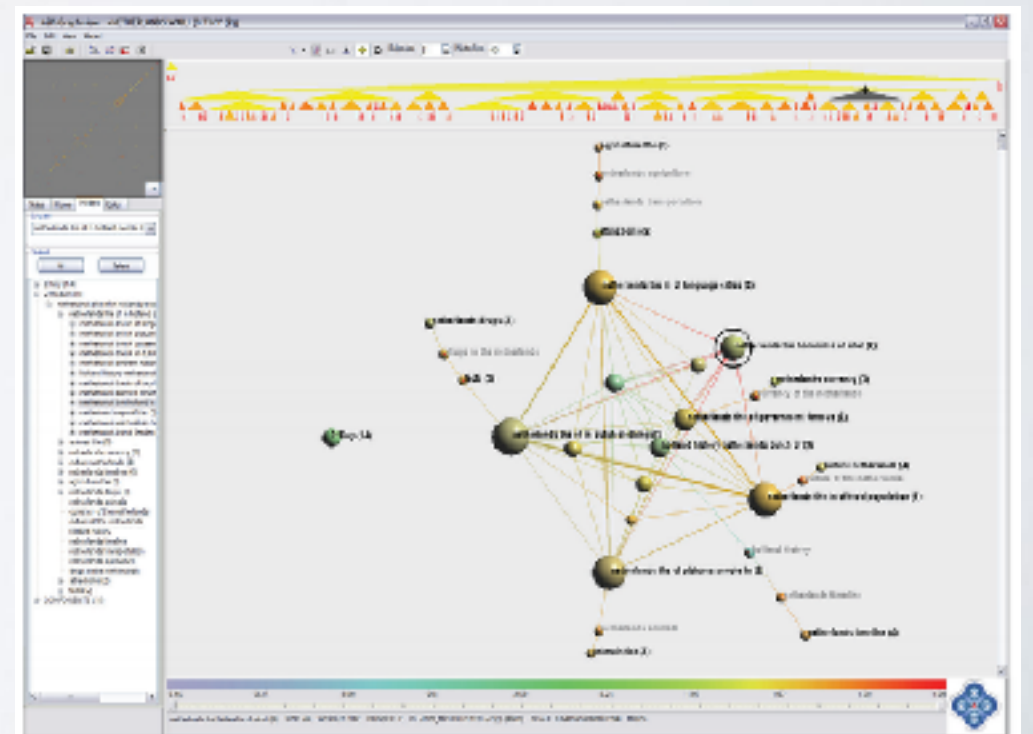
Heim, Philipp, Sebastian Hellmann, Jens Lehmann, Steffen Lohmann, and Timo Stegemann. "RelFinder: Revealing relationships in RDF knowledge bases." In *International Conference on Semantic and Digital Media Technologies*, pp. 182-187. Springer, Berlin, Heidelberg, 2009.



(b) Graph Hierarchies

Archambault, D., Munzner, T., & Auber, D. (2008). GrouseFlocks: Steerable exploration of graph hierarchy space. *IEEE transactions on visualization and computer graphics*, 14(4), 900-913.

Abello, J., Van Ham, F., & Krishnan, N. (2006). Ask-graphview: A large scale graph visualization system. *IEEE transactions on visualization and computer graphics*, 12(5), 669-676.



# FACETED BROWSERS



Fig. 3. Facet search on type `vra:Work`, but with a still active constraint on `mainPerson` (birthplace Provence-Alpes-Côte d'Azur). Also note the timeline in the bottom, visualizing multiple time-related facets. Images courtesy of Mark Harden, used with permission.

Heim, P., Ziegler, J., & Lohmann, S. (2008, December). gFacet: A Browser for the Web of Data. In Proceedings of the International Workshop on Interacting with Multimedia Content in the Social Semantic Web (IMC-SSW'08) (Vol. 417, pp. 49-58).

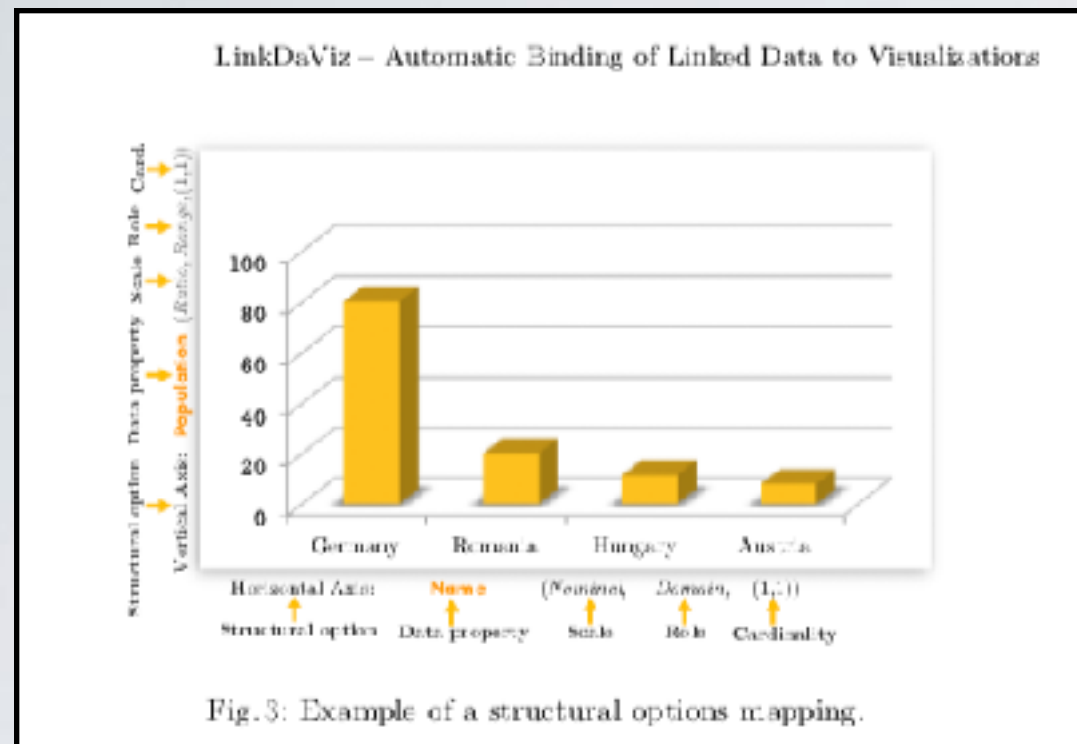
Sven Buschbeck, Anthony Jameson, Adrian Spirescu, Tanja Schneeberger, Raphaël Troncy, Houda Khrouf, Osmo Suominen, and Eero Hyvönen. 2013. Parallel faceted browsing. In CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13). ACM, New York, NY, USA, 3023-3026.

Schraefel, M. C., Alex, D., Smith, E., Russel, A., Owens, A., Harris, C., & Wilson, M. (2005). The mSpace classical music explorer: improving access to classical music for real people. In *MusicNetwork Open Workshop, Integration of Music in Multimedia Applications*.

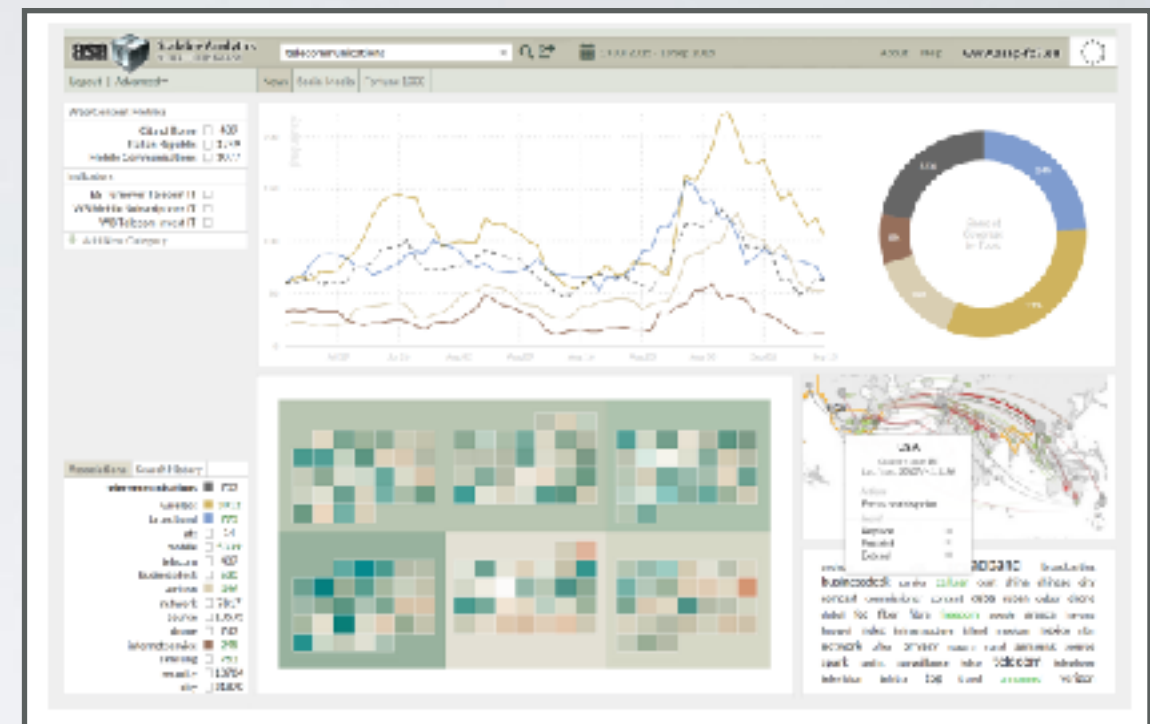


Figure 1: Screenshot of the parallel faceted browsing demonstrator.

# SET-BASED VISUALISATIONS



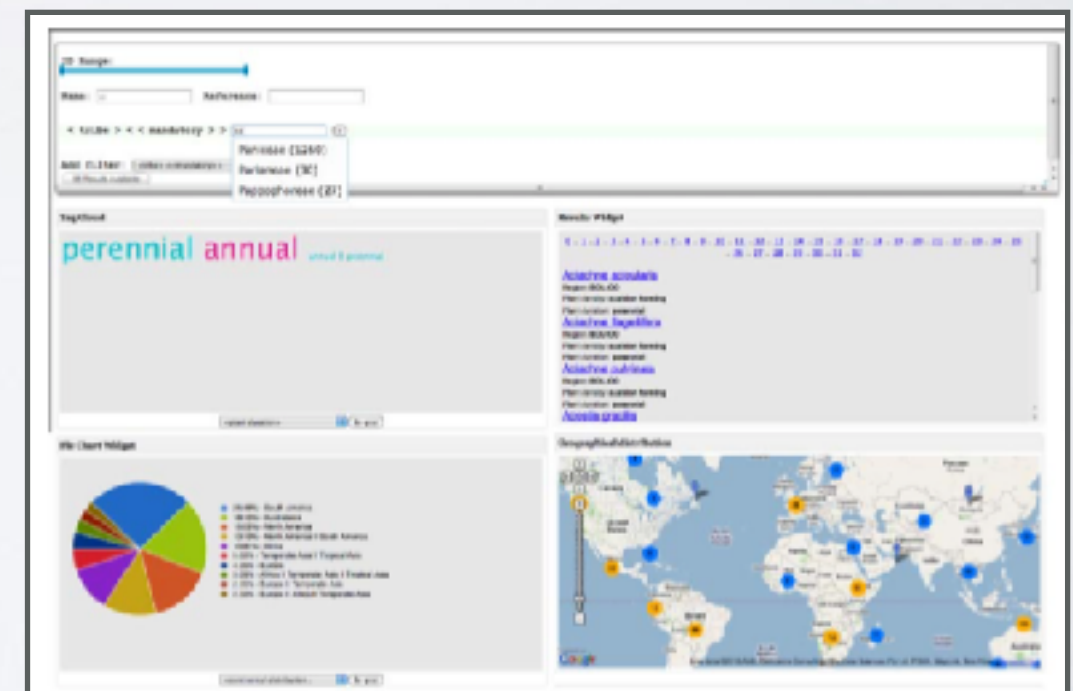
Theilmann, K., Galkin, M., Orlandi, F., & Auer, S. (2015, October). LinkDaViz—automatic binding of linked data to visualizations. In *International Semantic Web Conference* (pp. 147-162). Springer, Cham.



Braşoveanu, Adrian MP, et al. "Visualizing statistical linked knowledge for decision support." *Semantic Web* 8.1 (2017): 113-137.

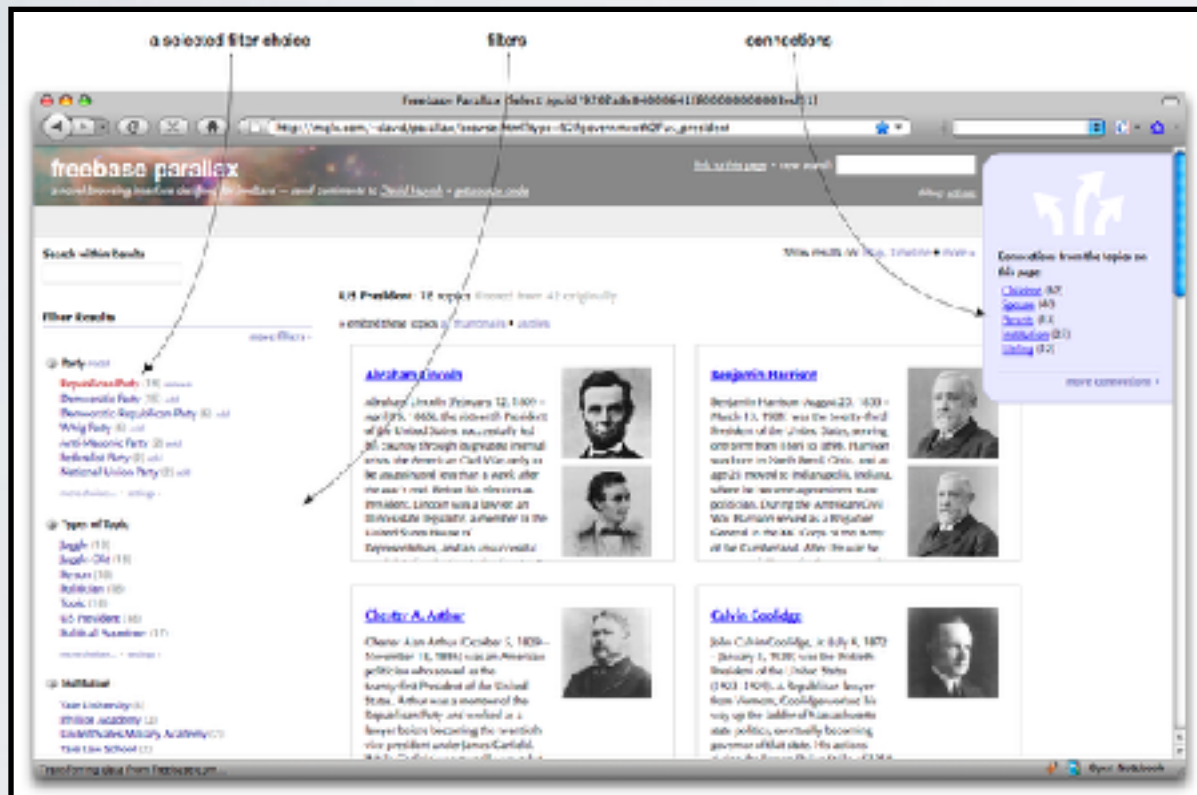


Berners-Lee, Tim, Yuhsin Chen, Lydia Chilton, Dan Connolly, Ruth Dhanaraj, James Hollenbach, Adam Lerer, and David Sheets. "Tabulator: Exploring and analyzing linked data on the semantic web." In *Proceedings of the 3rd international semantic web user interaction workshop*, vol. 2006, p. 159. 2006.

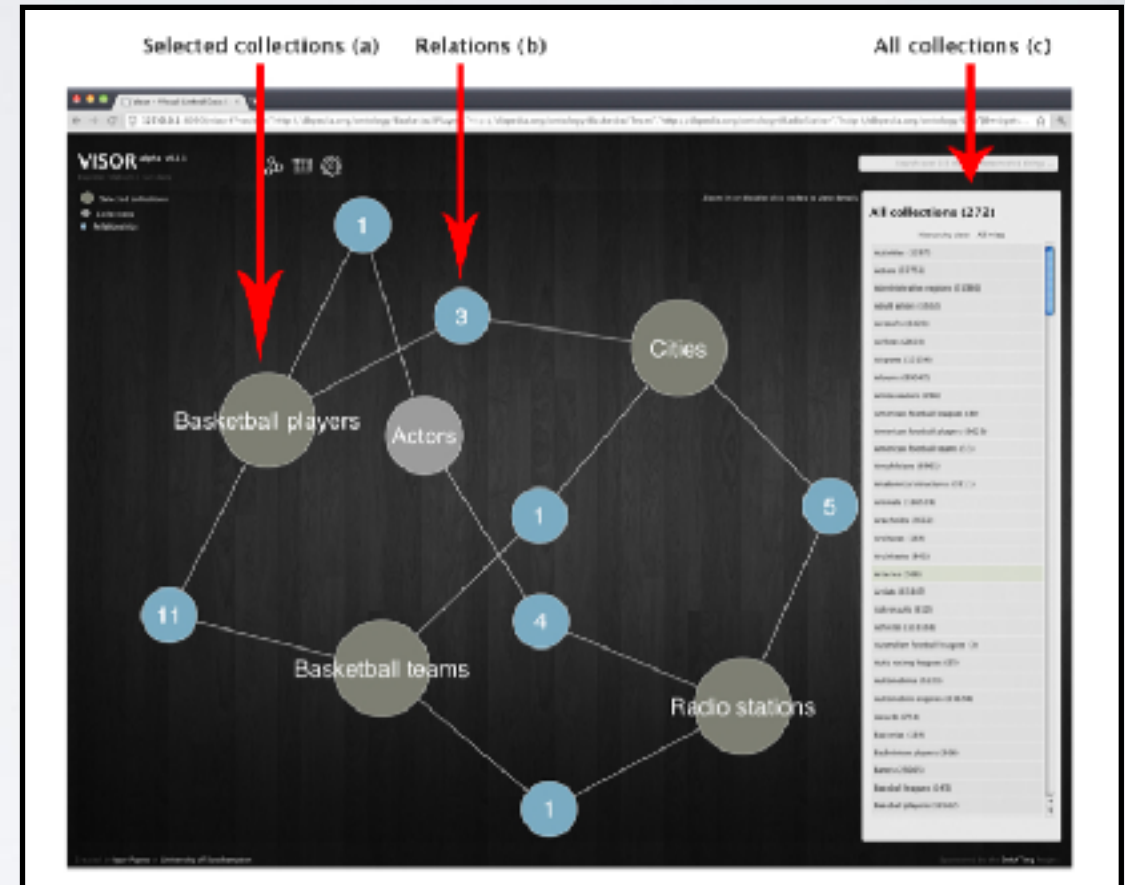


Mazumdar, Suvodeep, Daniela Petrelli, and Fabio Ciravegna. "Exploring user and system requirements of linked data visualization through a visual dashboard approach." *Semantic Web* 5.3 (2014): 203-220.

# SET-BASED PIVOT



Huynh, David F., and David Karger. "Parallax and companion: Set-based browsing for the data web." In *WWW Conference*. ACM, p. 6. 2009.



Popov, I. O., Schraefel, M. C., Hall, W., & Shadbolt, N. (2011, October). Connecting the dots: a multi-pivot approach to data exploration. In *International semantic web conference* (pp. 553-568). Springer, Berlin, Heidelberg.

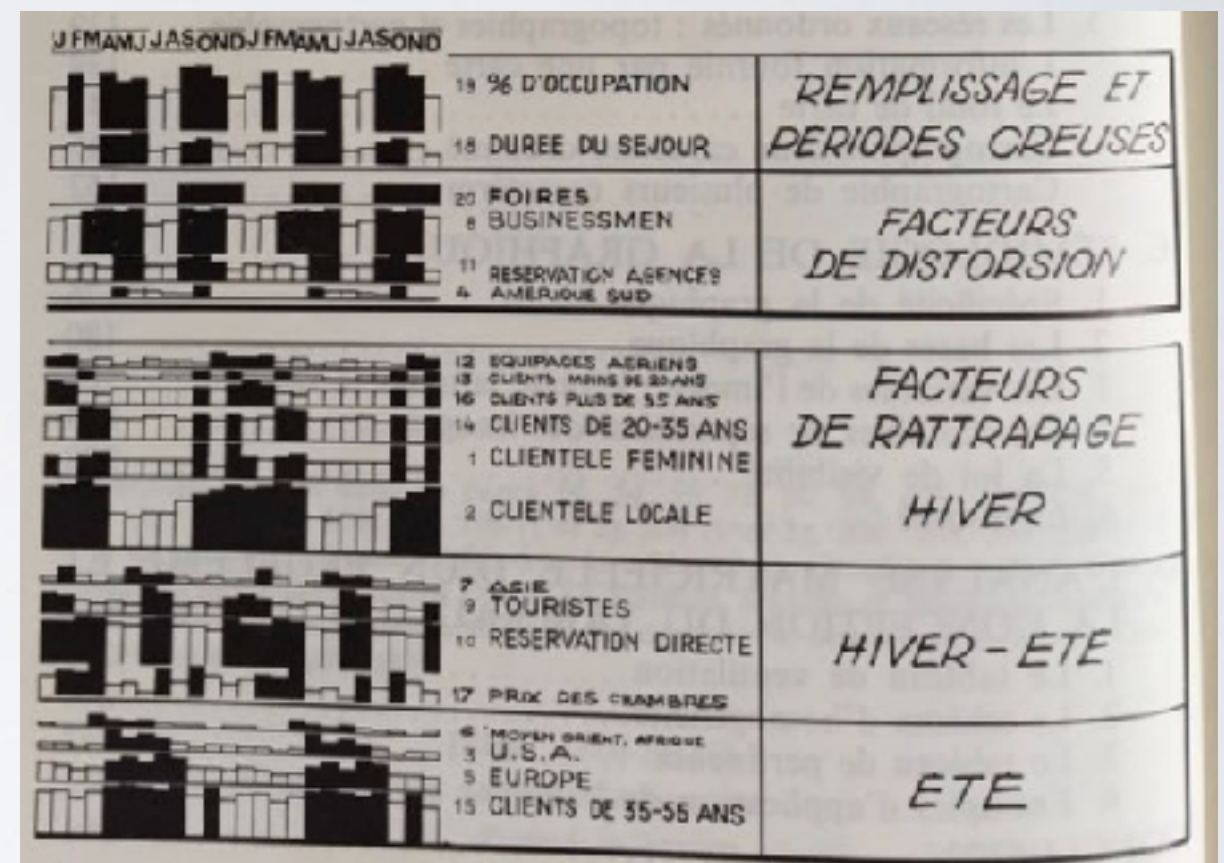
# S-PATHS

- A generic approach to continuously explore (from overview to detail) sets of RDF data with no a-priori knowledge of the model
- Offer a readable default view at any stage, and let the user explore other configurations at will
- Enable advanced selection

# READABLE DIMENSIONS ?

quantitative / categorical

| J    | F    | M    | A    | M    | J   | J    | A    | S    | O    | N    | D    |    |                           |
|------|------|------|------|------|-----|------|------|------|------|------|------|----|---------------------------|
| 26   | 21   | 26   | 28   | 20   | 20  | 20   | 20   | 40   | 15   | 40   |      | 1  | % CLIENTELE FEMININE      |
| 69   | 70   | 77   | 71   | 37   | 36  | 39   | 39   | 55   | 60   | 68   | 72   | 2  | % " " LOCALE              |
| 7    | 6    | 3    | 6    | 23   | 14  | 19   | 14   | 9    | 6    | 8    | 8    | 3  | % " " U.S.A.              |
| 0    | 0    | 0    | 0    | 8    | 6   | 6    | 4    | 2    | 12   | 0    | 0    | 4  | % " " AMERIQUE SUD        |
| 20   | 15   | 14   | 15   | 23   | 27  | 22   | 30   | 27   | 19   | 19   | 17   | 5  | % " " EUROPE              |
| 1    | 0    | 0    | 8    | 6    | 4   | 6    | 4    | 2    | 1    | 0    | 1    | 6  | % " " M.ORIENT.AFRIQUE    |
| 3    | 10   | 6    | 0    | 3    | 13  | 8    | 9    | 5    | 2    | 5    | 2    | 7  | % " " ASIE                |
| 78   | 80   | 85   | 86   | 85   | 87  | 70   | 76   | 87   | 85   | 87   | 80   | 8  | % BUSINESSMEN             |
| 22   | 20   | 15   | 14   | 15   | 13  | 30   | 24   | 13   | 15   | 13   | 20   | 9  | % TOURISTES               |
| 70   | 70   | 75   | 74   | 69   | 68  | 74   | 75   | 68   | 68   | 64   | 75   | 10 | % RESERVATION DIRECTE     |
| 20   | 18   | 19   | 17   | 27   | 27  | 19   | 19   | 26   | 27   | 21   | 15   | 11 | % " " AGENCES             |
| 10   | 12   | 6    | 9    | 4    | 5   | 7    | 6    | 6    | 5    | 15   | 10   | 12 | % EQUIPAGES AERIENS       |
| 2    | 2    | 4    | 2    | 2    | 1   | 1    | 2    | 2    | 4    | 2    | 5    | 13 | % CLIENTS MOINS DE 20 ANS |
| 25   | 27   | 37   | 35   | 25   | 25  | 27   | 28   | 24   | 30   | 24   | 30   | 14 | % " " DE 20-35 ANS        |
| 48   | 49   | 42   | 48   | 54   | 55  | 53   | 57   | 55   | 46   | 55   | 43   | 15 | % " " DE 35-55 ANS        |
| 25   | 22   | 17   | 15   | 19   | 19  | 19   | 19   | 20   | 19   | 22   | 16   | 16 | % " " PLUS DE 55 ANS      |
| 153  | 167  | 166  | 174  | 152  | 155 | 145  | 170  | 157  | 174  | 165  | 156  | 17 | PRIX DES CHAMBRES         |
| 1.65 | 1.71 | 1.65 | 1.91 | 1.90 | 2.  | 1.54 | 1.60 | 1.73 | 1.82 | 1.66 | 1.44 | 18 | DUREE DU SEJOUR           |
| 67   | 82   | 70   | 83   | 74   | 77  | 56   | 62   | 90   | 92   | 78   | 55   | 19 | % D'OCCUPATION            |
|      |      |      | X    | X    | X   |      |      | X    | X    | X    | X    | 20 | FOIRES                    |



Jacques Bertin "Useful information is a cluster"

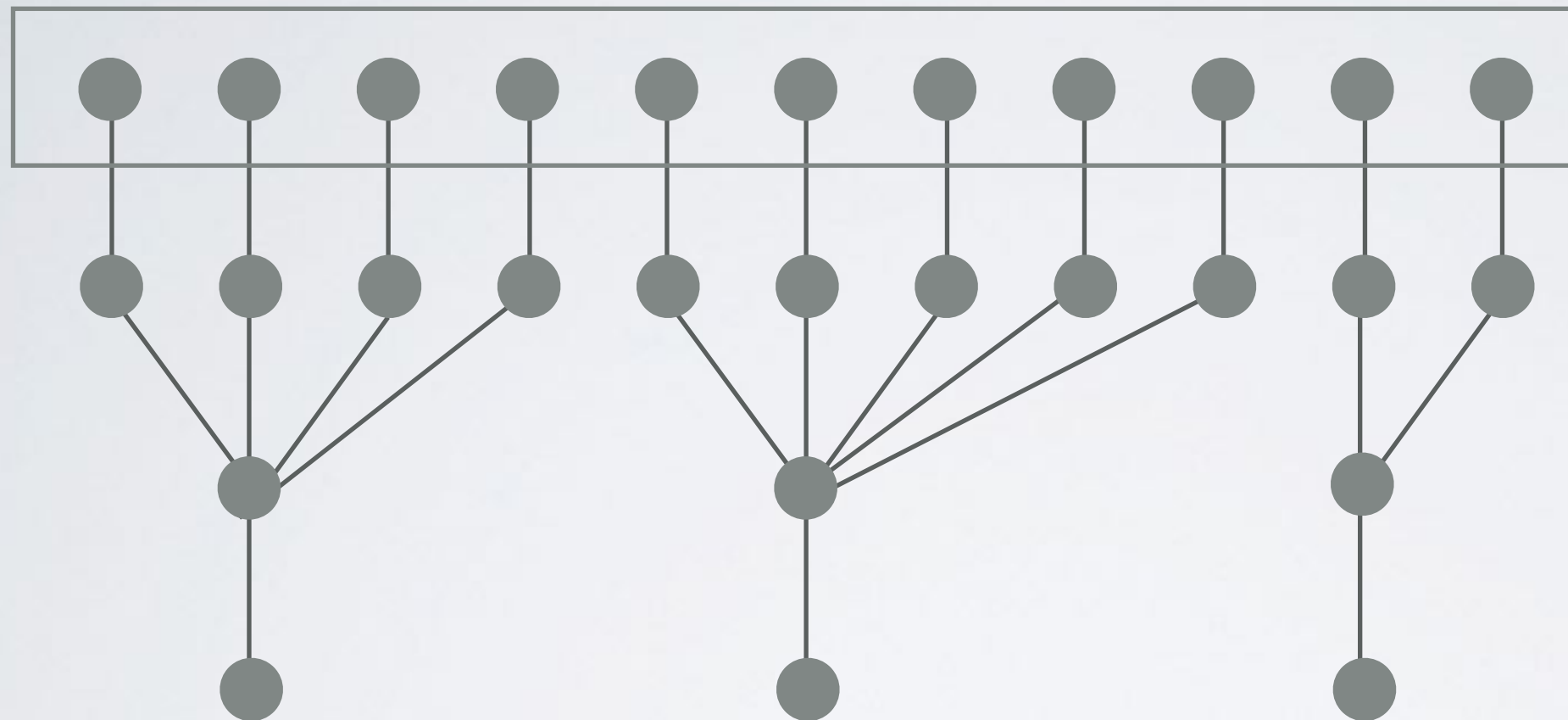
La Graphique et le traitement de l'information

# READABLE DIMENSIONS ?

|             |    |                                                          |
|-------------|----|----------------------------------------------------------|
| Datetime    | => | minutes, hours, days, months, years, decades, century... |
| Number      | => | thousands, millions, billiards...                        |
| String, URI | => | number of unique values                                  |

Jacques Bertin “Useful information is a cluster”  
La Graphique et le traitement de l'information

# FOLLOW THE PATHS



nobel:Laureate

<http://data.nobelprize.org/terms/laureateAward>

<http://data.nobelprize.org/terms/category>

<http://www.w3.org/2000/01/rdf-schema#label>

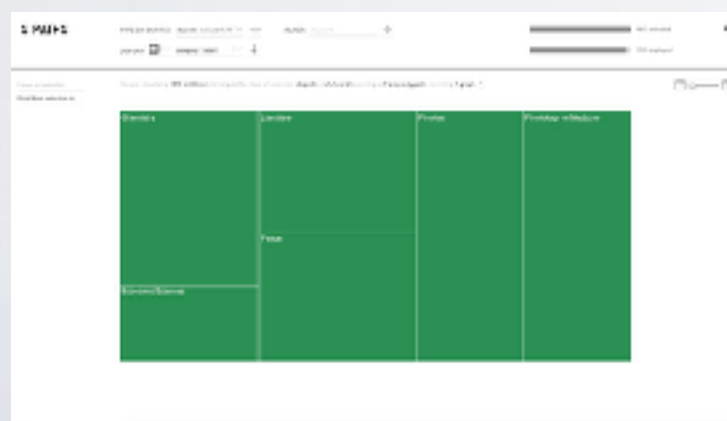
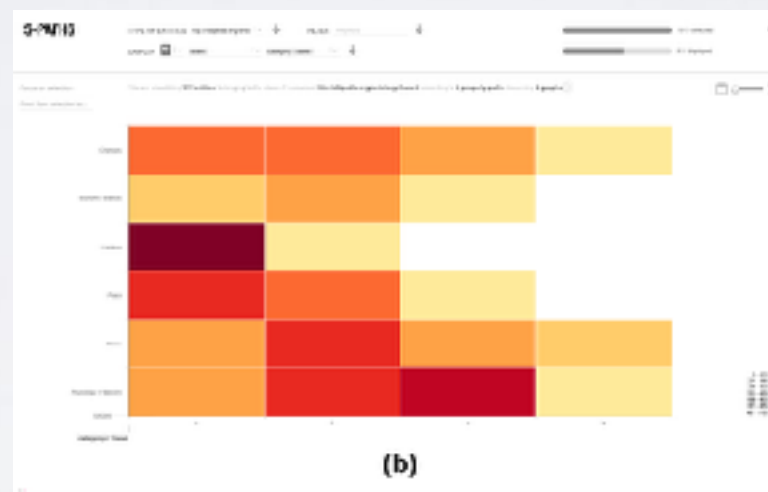
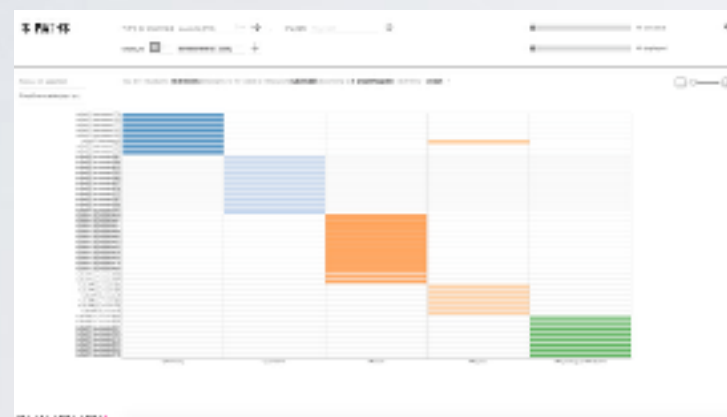
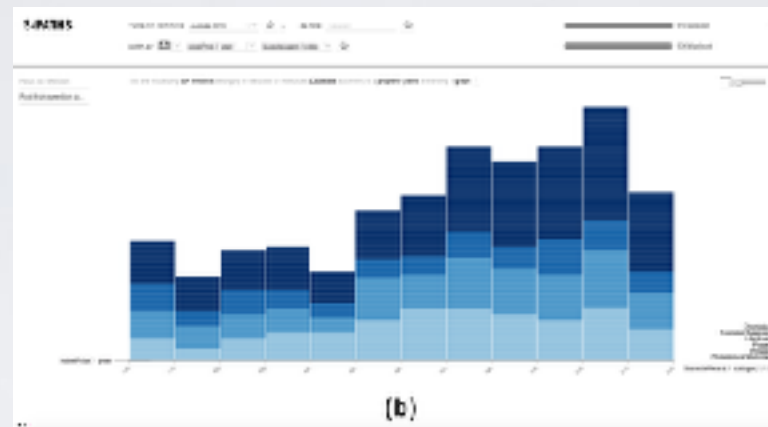
# ANALYSIS OF THE PATHS

| Characteristic      | Description                                                                         |
|---------------------|-------------------------------------------------------------------------------------|
| <i>category</i>     | one of: datetime, geographical coordinate, image, number, text or URI               |
| <i>depth</i>        | number of hops (statements) from the subject to the final property                  |
| <i>coverage</i>     | percentage of resources in the set for which this path actually exists              |
| <i>count</i>        | total number of values for the property at the end of the path, over all resources  |
| <i>unique count</i> | number of unique values for the property at the end of the path, over all resources |

```
1 SELECT DISTINCT
2   ?p1 ?p2 ?pn path of depth n
3   ?datatype datatype used to determine the category
4   (COUNT(?totalEntities) as ?nbEntities)
5   (COUNT(?values) as ?uniqueValues)
6   (COUNT(DISTINCT ?entities) as ?nbCoveredEntities)
7   (COUNT(?values) as ?totalValues)
8 WHERE {
9   ?totalEntities rdf:type <TYPE_URI> .
10  ?coveredEntities rdf:type <TYPE_URI> .
11  ?coveredEntities ?p1 ?o1 . ?o1 ?p2 ?o2 . ?o2 ?pn ?values .
12  BIND(datatype(VALUE) as ?datatype)
13 }
```

Fig. 1. Theoretical query to retrieve paths characteristics. S-Paths splits this query into multiple queries performed recursively.

# SET OF VIEWS

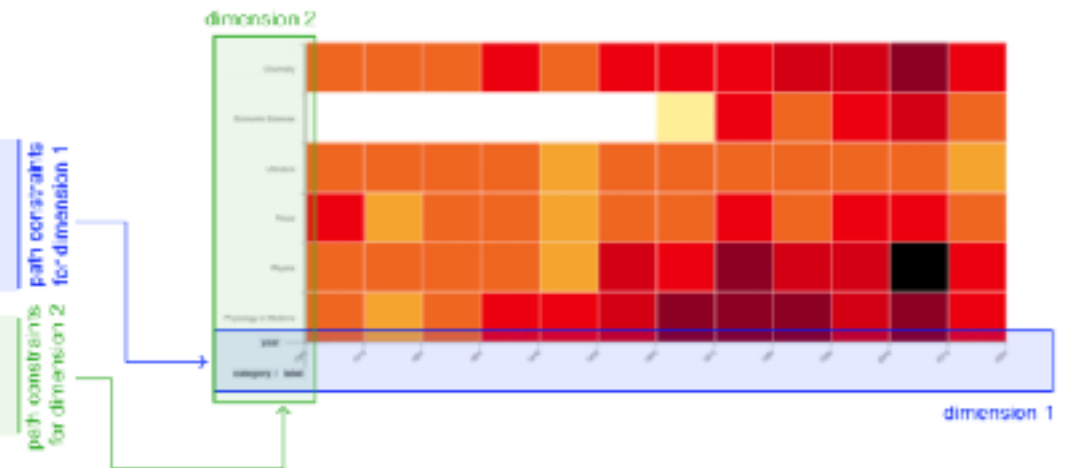


# SET OF VIEWS

```

1| let views = [
2|   {
3|     id: '2DDensityPlot',
4|     name: '2ddensityplot',
5|     thumb: '/images/2ddensityplot.svg',
6|     constraints: [
7|       [
8|         { category: 'datetime', unique: { min: 2 } },
9|         { category: 'text', avg: { max: 70, optimal: [10, 40] }, unique: { min: 2, max: 150 } },
10|        { category: 'uri', unique: { min: 2, max: 150 } }
11|      ],
12|      [
13|        { category: 'text', unique: { min: 2, max: 70, optimal: [10, 40] } },
14|        { category: 'uri', unique: { min: 2, max: 70, optimal: [10, 40] } }
15|      ]
16|    ]
17|  }
18| ]

```



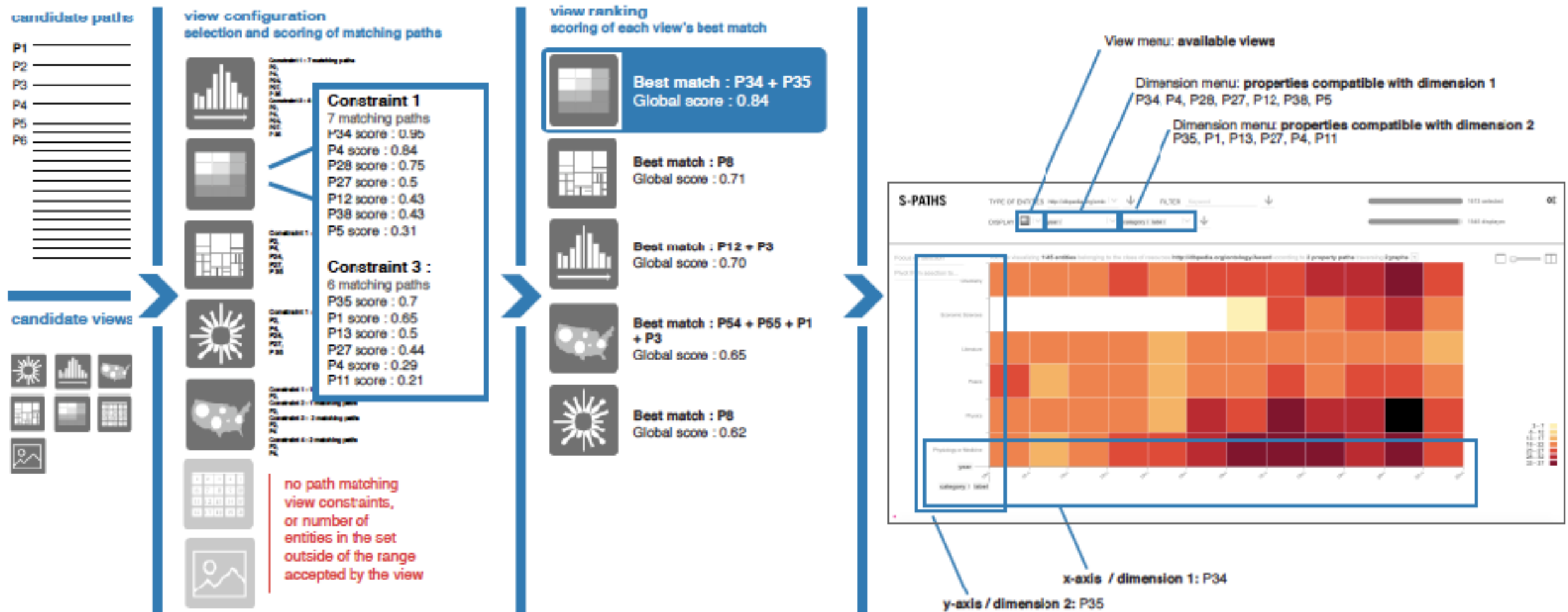
# SET OF VIEWS

| View                           | Type              | Weight | Number of resources                     | dimension 1                           | dimension 2                 | dimension 3 |
|--------------------------------|-------------------|--------|-----------------------------------------|---------------------------------------|-----------------------------|-------------|
| <i>density plot</i>            | aggregate         | 0.5    |                                         | datetime, text or uri                 | text or uri                 |             |
| <i>treemap</i>                 | aggregate         | 0.3    |                                         | single path: text or uri              |                             |             |
| <i>stacked chart</i>           | multiple distinct | 0.9    | min/max: [2, 1000]<br>optimal: [4, 200] | single path:<br>datetime, text or uri | single path:<br>text or uri |             |
| <i>timeline</i>                | multiple distinct | 0.85   | min/max: [2, 50]<br>optimal: [10, 20]   | all datetime paths                    | single path:<br>text or uri |             |
| <i>URI wheel</i>               | multiple distinct | 0.4    | min: 2                                  | uri                                   |                             |             |
| <i>map</i>                     | multiple distinct | 0.85   | min/max: [2, 1000]                      | geo                                   | geo                         | text        |
| <i>breakdown<br/>by values</i> | multiple distinct | 0.7    | max: 50, optimal: [1, 30]               | single path:<br>any category          |                             |             |
| <i>images</i>                  | multiple distinct | 0.8    | min/max: [2, 1000]                      | single path: image                    | single path: text           |             |
| <i>info card</i>               | single entity     | 1      | min/max: [1, 1]                         | all paths                             |                             |             |
| <i>node link<br/>diagram</i>   | single entity     | 0.5    | min/max: [1, 1]                         | all paths                             |                             |             |

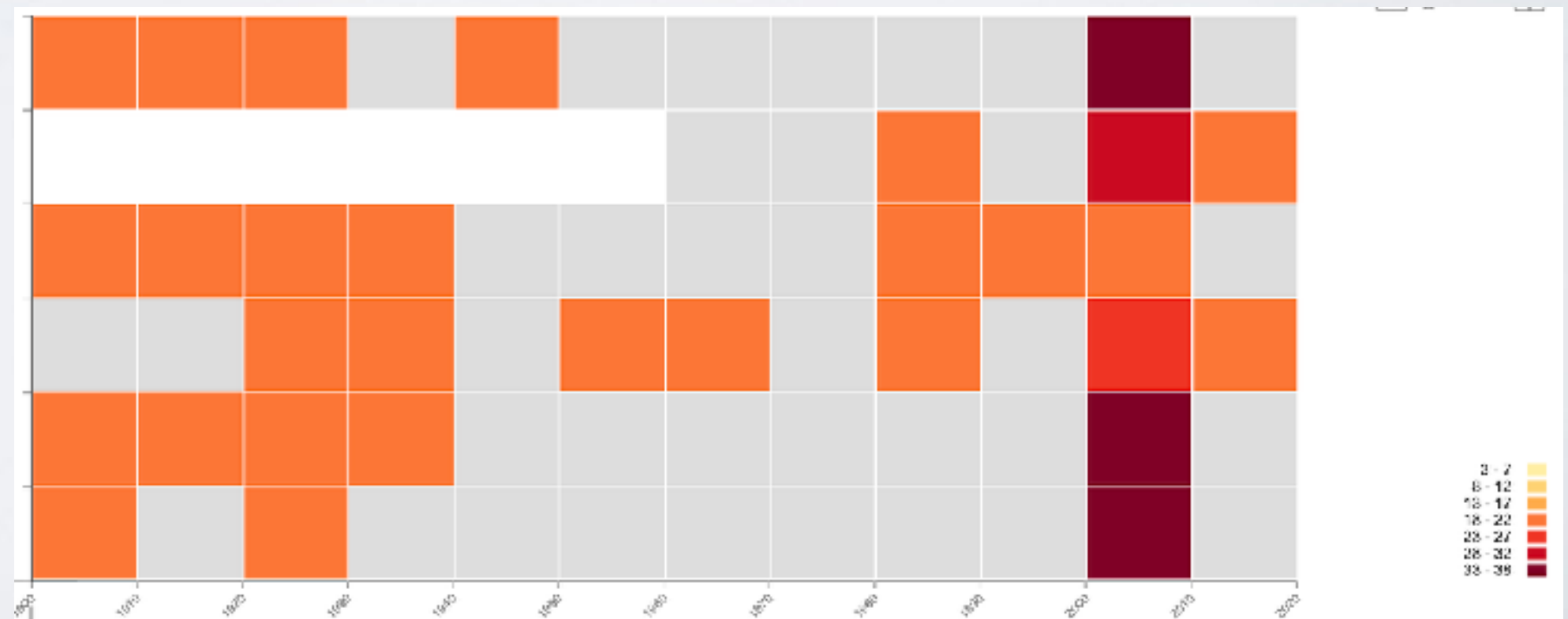
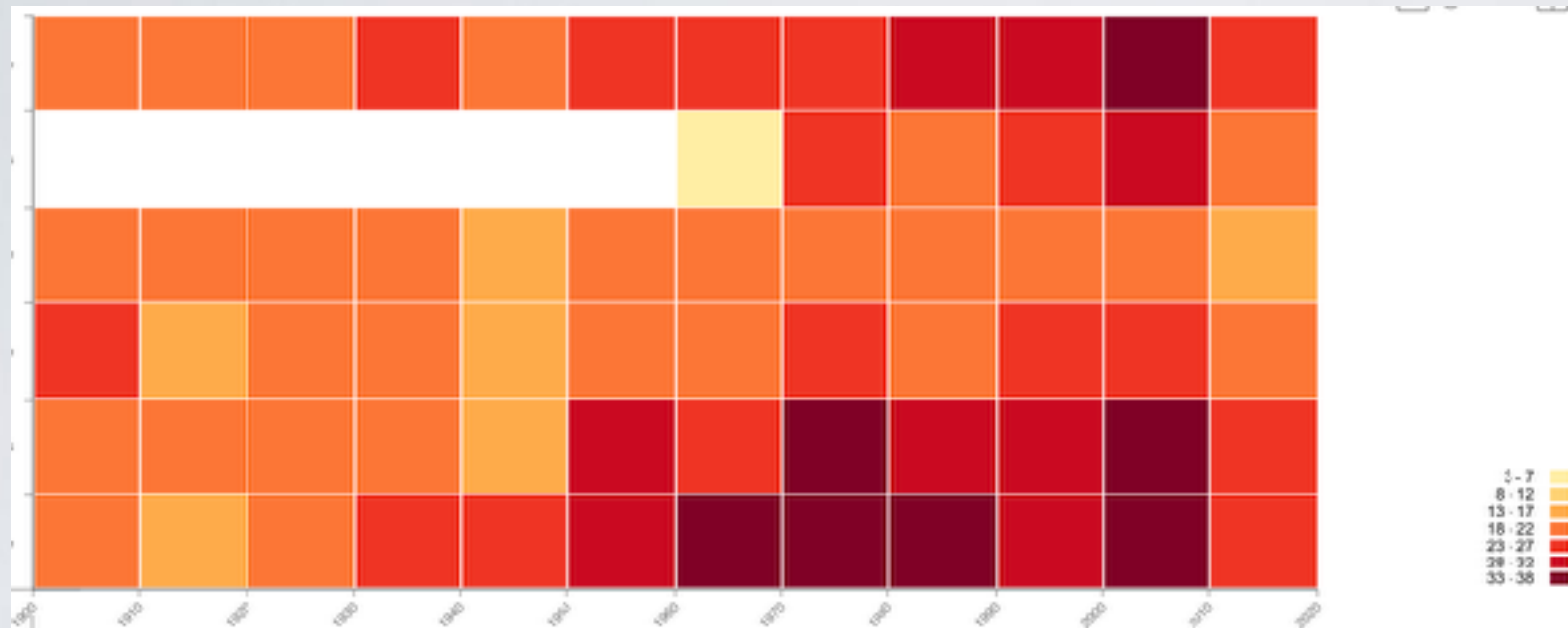
Table 2

Default configuration of view templates as configured for the Nobel dataset.

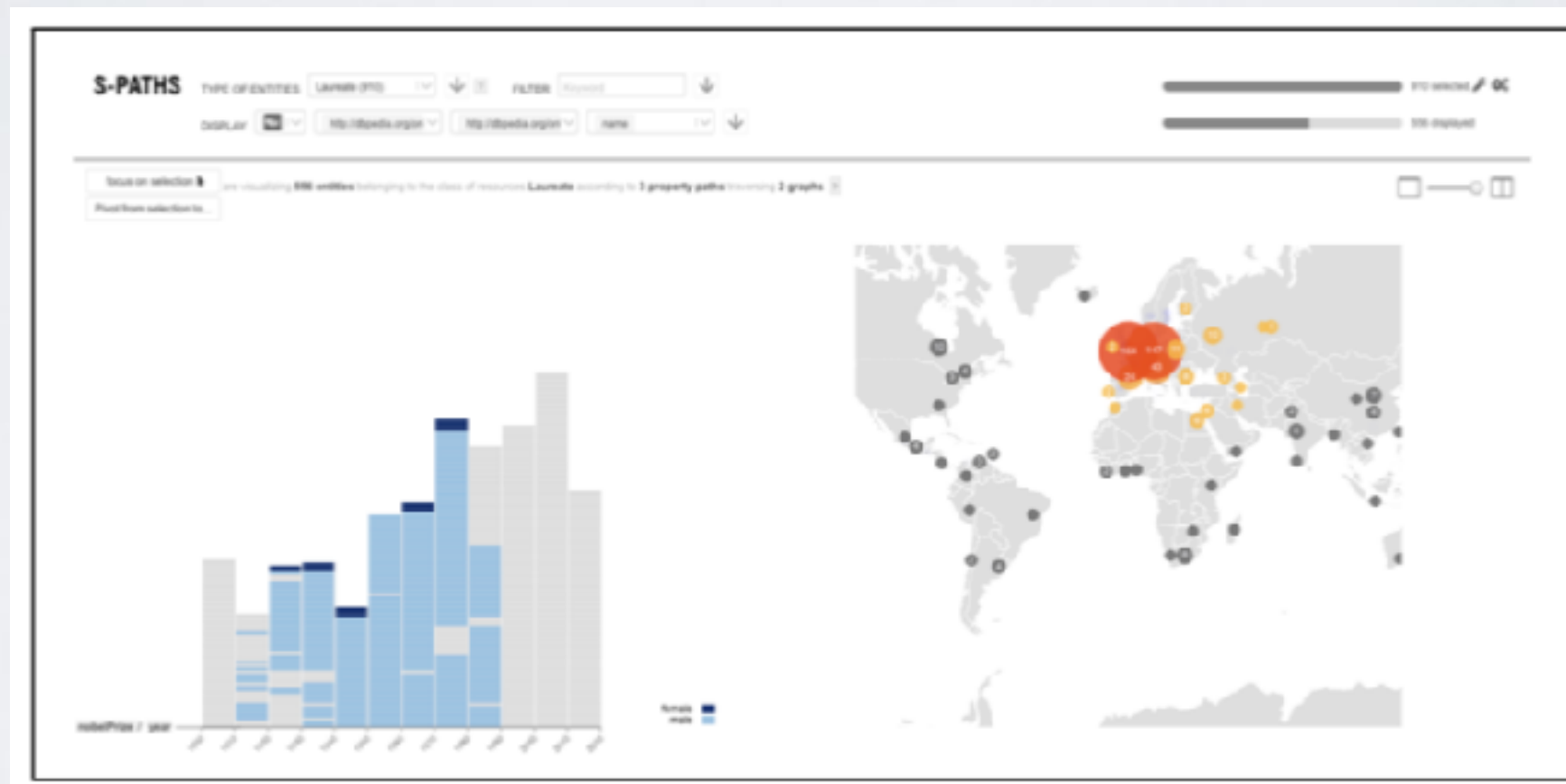
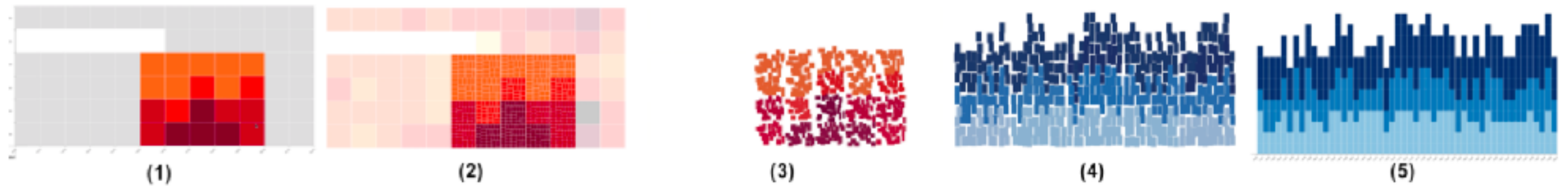
# A MATCHING ALGORITHM



# SELECTIONS



# TRANSITIONS, BRUSHING & LINKING



# A SAMPLE OF DATA.BNF

Nobel :  $\approx$  87 000 triplets

---

Data: representative sample = 33 118 586 triplets  
 $\approx$  10 %

Main entities:

**skos:Concept (395297)**

---

frbr-rda:Work (80991)

---

bnffr:expositionVirtuelle (3818)

---

Person (165722)

---

Organization (38450)

---

Document (4218)

---

S-PATHS

TYPE OF ENTITIES Person (165722) ↓

FILTER Keyword ↓

DISPLAY list ↓

firstYear ↓

gender ↓

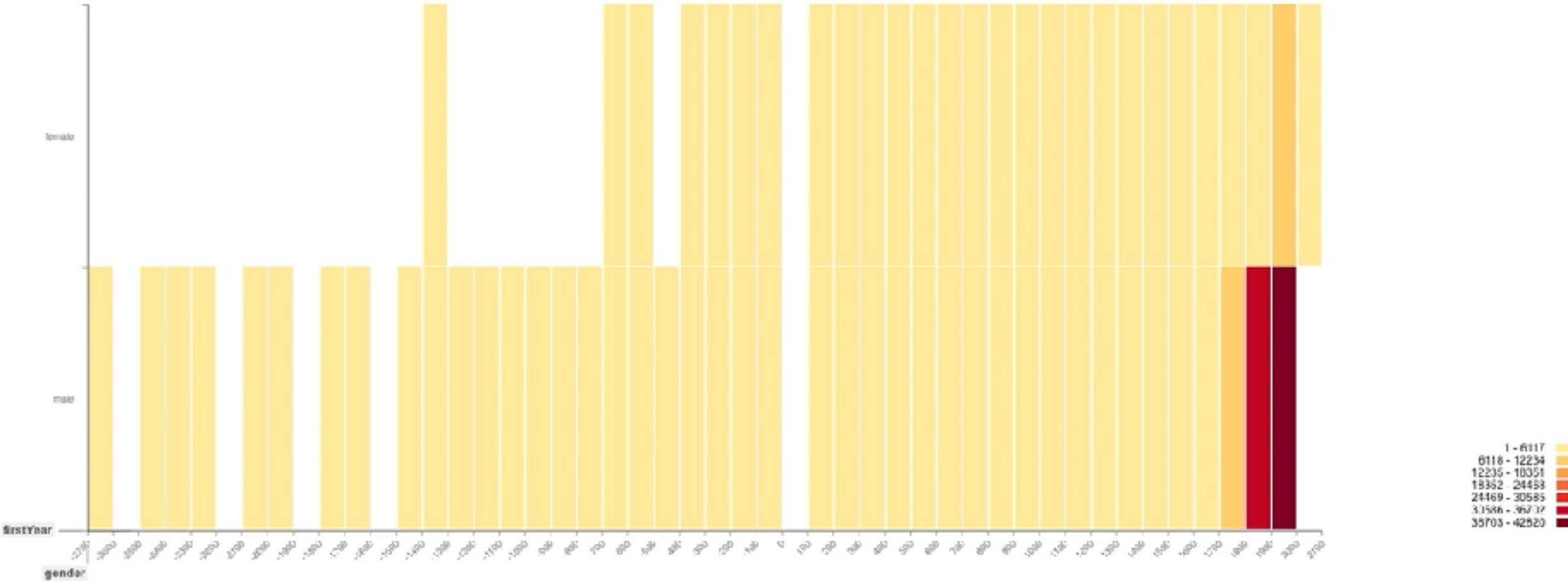
165722 selected

105538 displayed

Focus on collection

You are visualizing 105538 entities belonging to the class of resources **Person** according to 2 property paths traversing 1 graph

Phot from selection list...



S-PATHS

TYPE OF ENTITIES Person (165722) ↓

FILTER Keyword ↓

DISPLAY list ↓ firstYear ↓ gender ↓

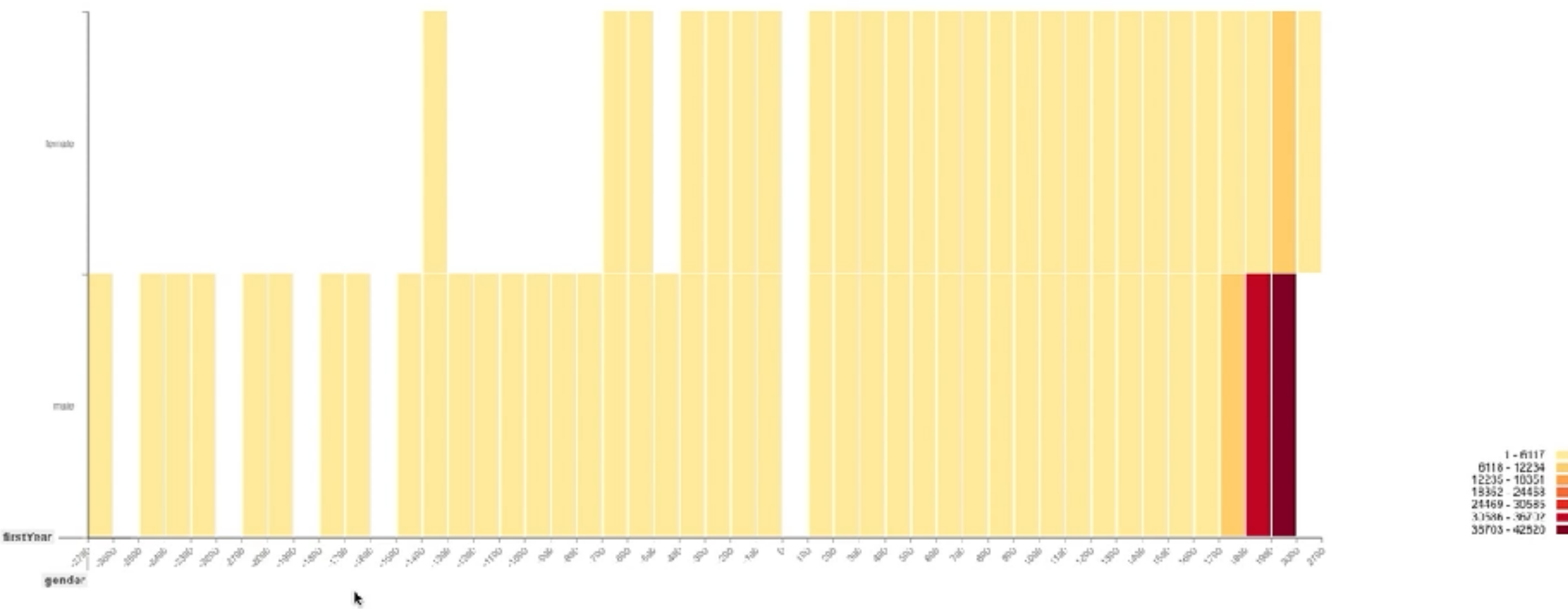
165722 selected

105538 displayed

Focus on collection

You are visualizing 105538 entities belonging to the class of resources **Person** according to 2 property paths traversing 1 graph

Print from selection list...



## S-PATHS

TYPE OF ENTITIES **for:rd:work(2036)** |  

**FILTER** Keyword 

DISPLAY  1stYear?  <http://data.bnf.fr/en/role>  

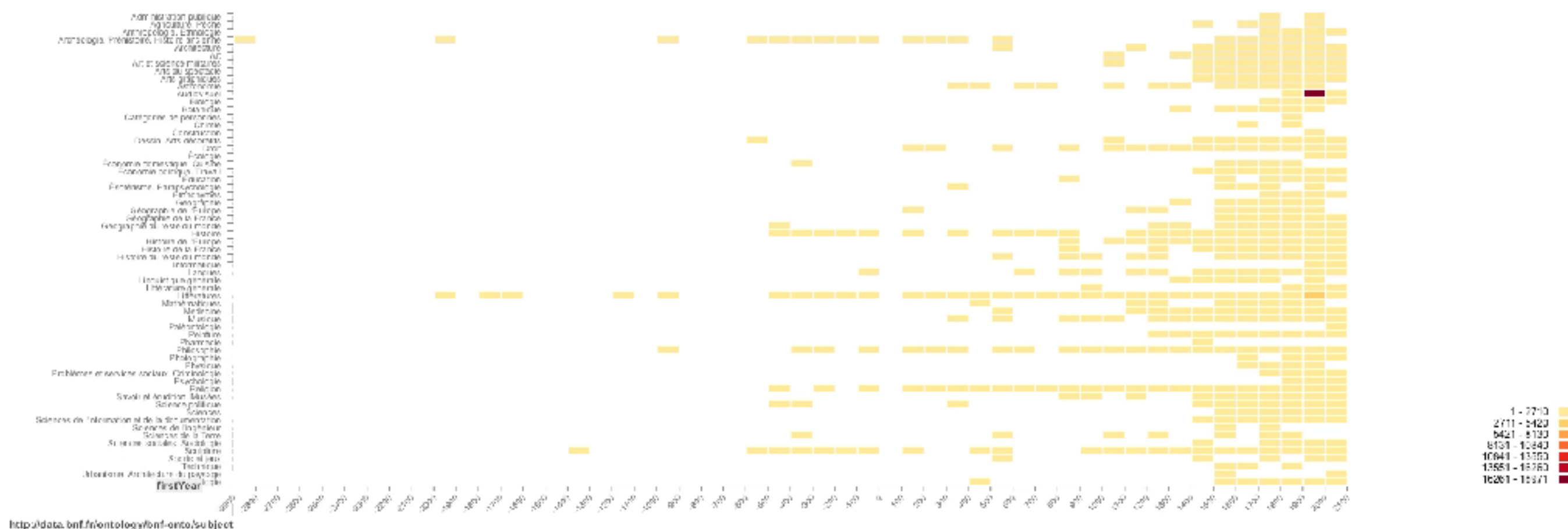
● MUST be noted

93582 cf-played

### Focus on selection

You are visualizing **33682** entities belonging to the class of resources **frbr-rda:Work** according to **2** property paths traversing **1** graph.

Print from selection to...



## S-PATHS

TYPE OF ENTITIES **for:da:Work** (83961) |  

**FILTER** Keyword 

DISPLAY  firstYear?  <http://data.bnf.fr/en/role>  

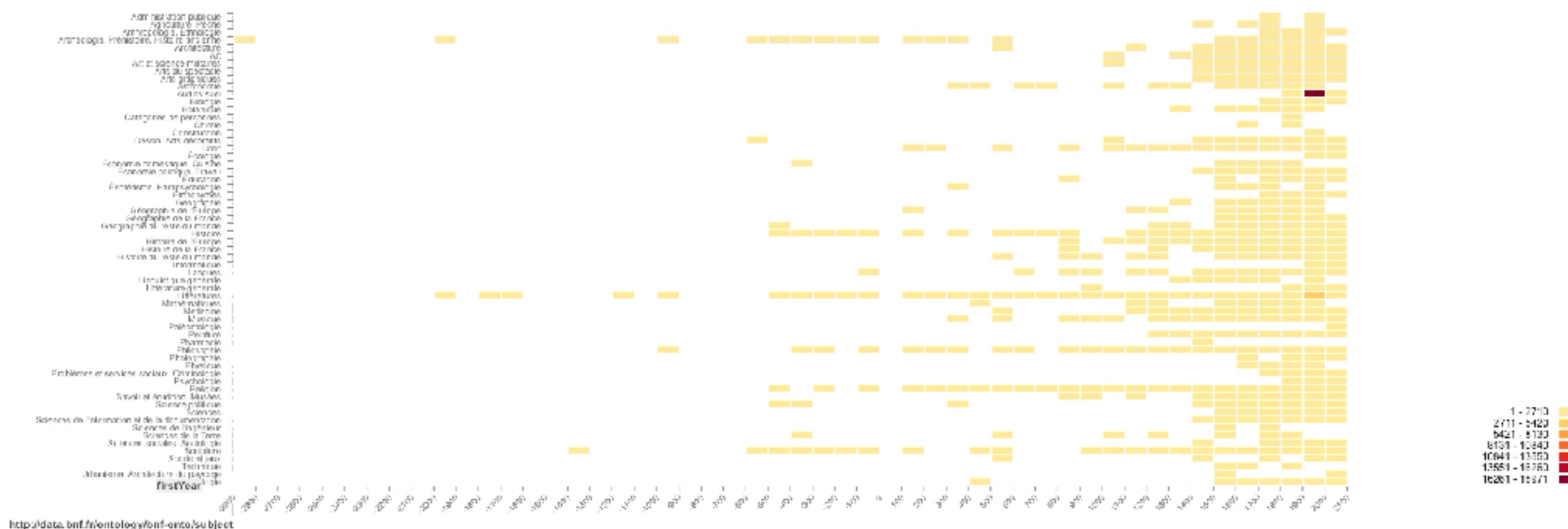
● MUST be noted

95582 ref-pubmed

### Focus on selection

You are visualizing **33682** entities belonging to the class of resources **frbr-rda:Work** according to **2** property paths traversing **1** graph.

Print from selection to...



S-PATHS

TYPE OF ENTITIES skos:Concept (385287) ↓

DISPLAY 📊 ↓ Date Created ↓ focus / gendarm ↓

FILTER Keyword ↓

385287 selected

157485 displayed

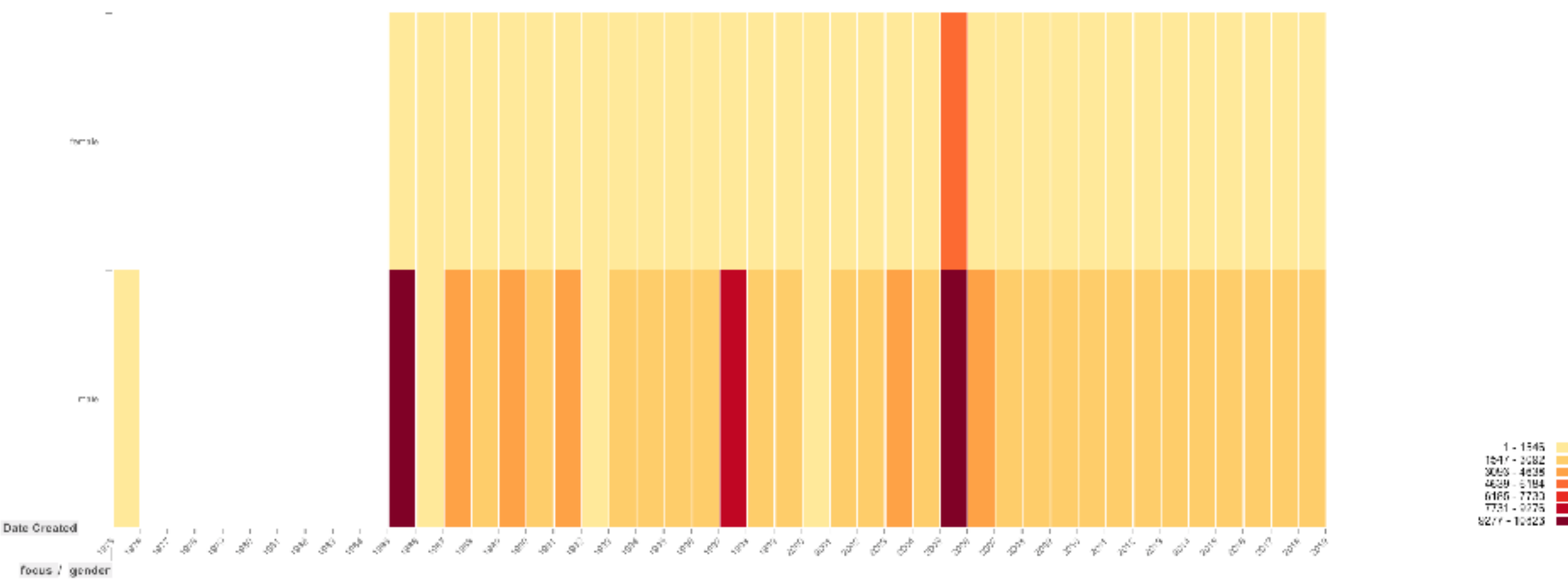
🔍 🗨

Focus on selection

You are visualizing 137480 entities belonging to the class of resources **skos:Concept** according to 2 property paths traversing 1 graph ?

🗨 🔍 🗨

Pivot from selection to...



## S-PATHS

TYPE OF ENTITIES Document (4218)   ? FILTER Keyword 

4218 selected



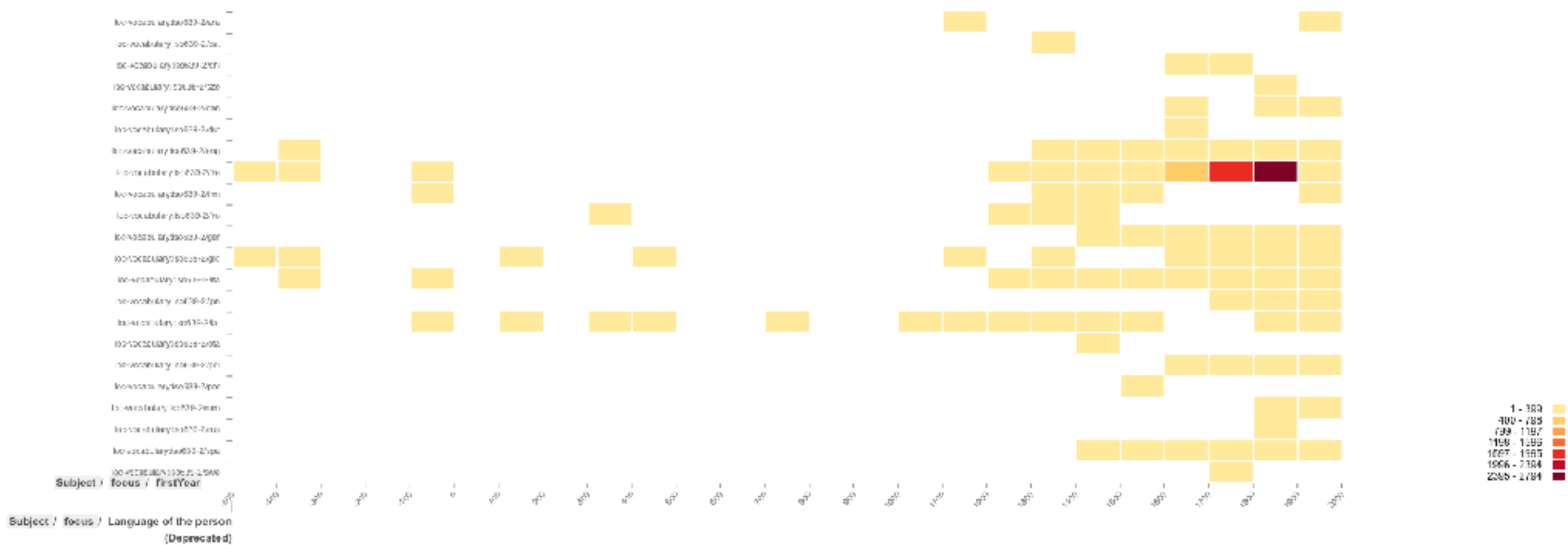
DISPLAY  Subject / focus / first |  Subject / focus / Long | 

2037 display: inline-block; width: 100%; height: 10px; background: linear-gradient(to right, #ccc, #000 49%, #000 49%, #000 51%, #ccc 51%, #ccc 52%); background-size: 100%; background-repeat: no-repeat; margin-bottom: 5px;">

### Focus on selection

You are visualizing **2937** entities belonging to the class of resources **Document** according to **2** property paths traversing **1** graph ?

Print from selection to...



S-PATHS

TYPE OF ENTITIES Document (4218) FILTER Keyword

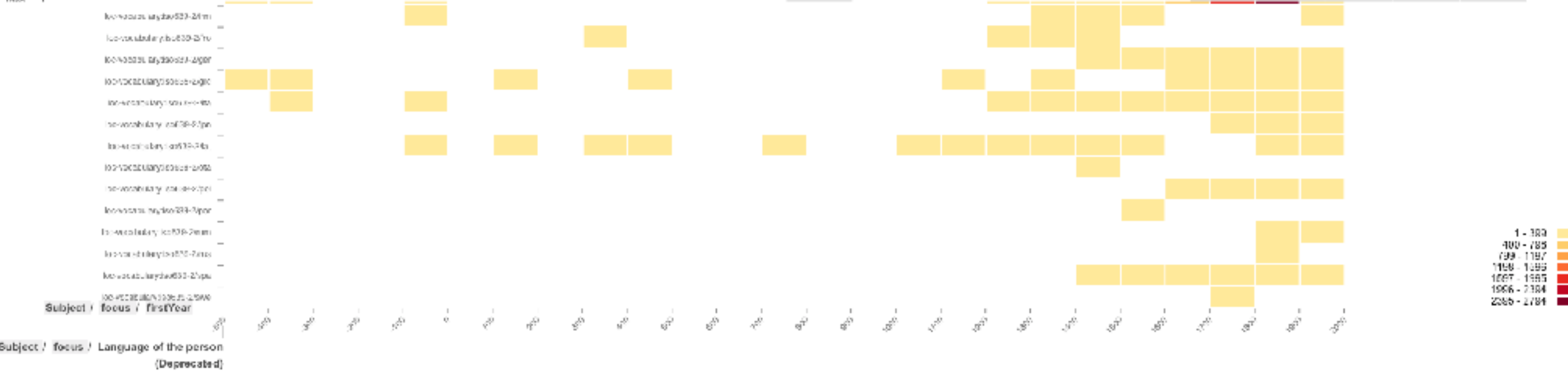
4218 selected 2617 discarded

ara  
cat  
ichi  
ces  
dan  
den  
eng  
fin  
frn  
ita

Sample of entities in the current selection

Path 1:  
<http://ark.bnf.fr/ark:/12148/mm106200310r> / Subject / <http://data.bnf.fr/ark:/12148/cb145562923> / Date Modified / 2013-08-03

Path 2:  
<http://ark.bnf.fr/ark:/12148/mm106200310r> / Subject / <http://data.bnf.fr/ark:/12148/cb145562923> / focus / <http://data.bnf.fr/ark:/12148/cb145562923#about> / Language of the person (Deprecated) / <http://vocabulary.iso639-2/ara>



# 'Abd Allāh ibn 'Umar ibn Muḥammad ibn 'Alī al- Bayḍāwī (12..?-1286?)



Langue : **arabe**

Sexe : masculin

Naissance : Al-Bayḍā' (Iran) **12..**

Mort : **1286**

Note : Savant spécialiste du droit islamique

Autres formes du nom : 1286-?-12 (عبد الله بن عمر بن محمد بن علي البغدادي) (arabe)  
Abdalla Baidawi (12..?-1286?)  
Nāṣir al Dīn 'Abd Allāh ibn 'Umar al- Bayḍāwī (12..?-1286?)  
**Voir plus**

ISNI : ISNI 0000 0001 1897 9016

Ses activités  
(16 documents)

Documents à propos  
de cet auteur

Pages dans data.bnf.fr  
(2 pages)

Sources  
et références

## Ses activités

☒ Voir tous les documents (16)

☐ Voir les documents numérisés (1)

► Auteur du texte (14)

► Auteur du commentaire (1)

► Autre (1)

## Auteur du texte

14 documents

### → تفسير القرآن

Description : Note : Tafsīr al-Qor'ān. Commentaire du Coran en marge du texte sacré  
Édition : Le Caire, impr. de 'Abd al-Raḥmān Moḥammad, 1305 (1888). Gr. in-4°, 25-816-2 p. [Acq. 21230]  
[\[catalogue\]](#)

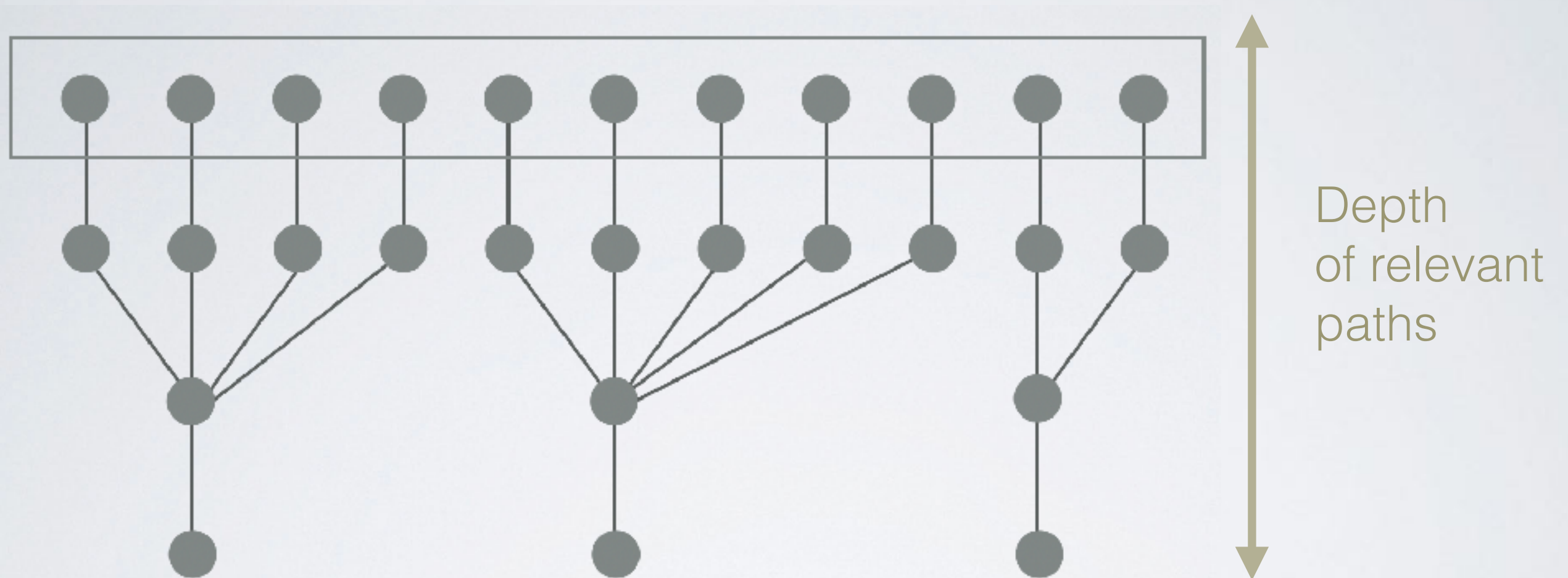
### → Contenu dans : Nature, man and God in medieval Islam

'Abd Allah Baydawi's text "Tawālī 'al-Anwar min Matalī 'al-Anzar" along with Mahmud Isfahani's commentary "Matalī 'al-Anzar, Sharḥ Tawālī 'al-Anwar". - [1]  
Description matérielle : 2 vol. ([LIX]-1183 p.)  
Description : Note : Trad. anglaise seule. - Index  
Édition : Leiden ; Boston : Brill, 2002  
[\[catalogue\]](#)

# DIFFICULTIES

- Recursive analysis of paths
- Binning in the query
- Number of paths => select relevant branches to explore for subselections

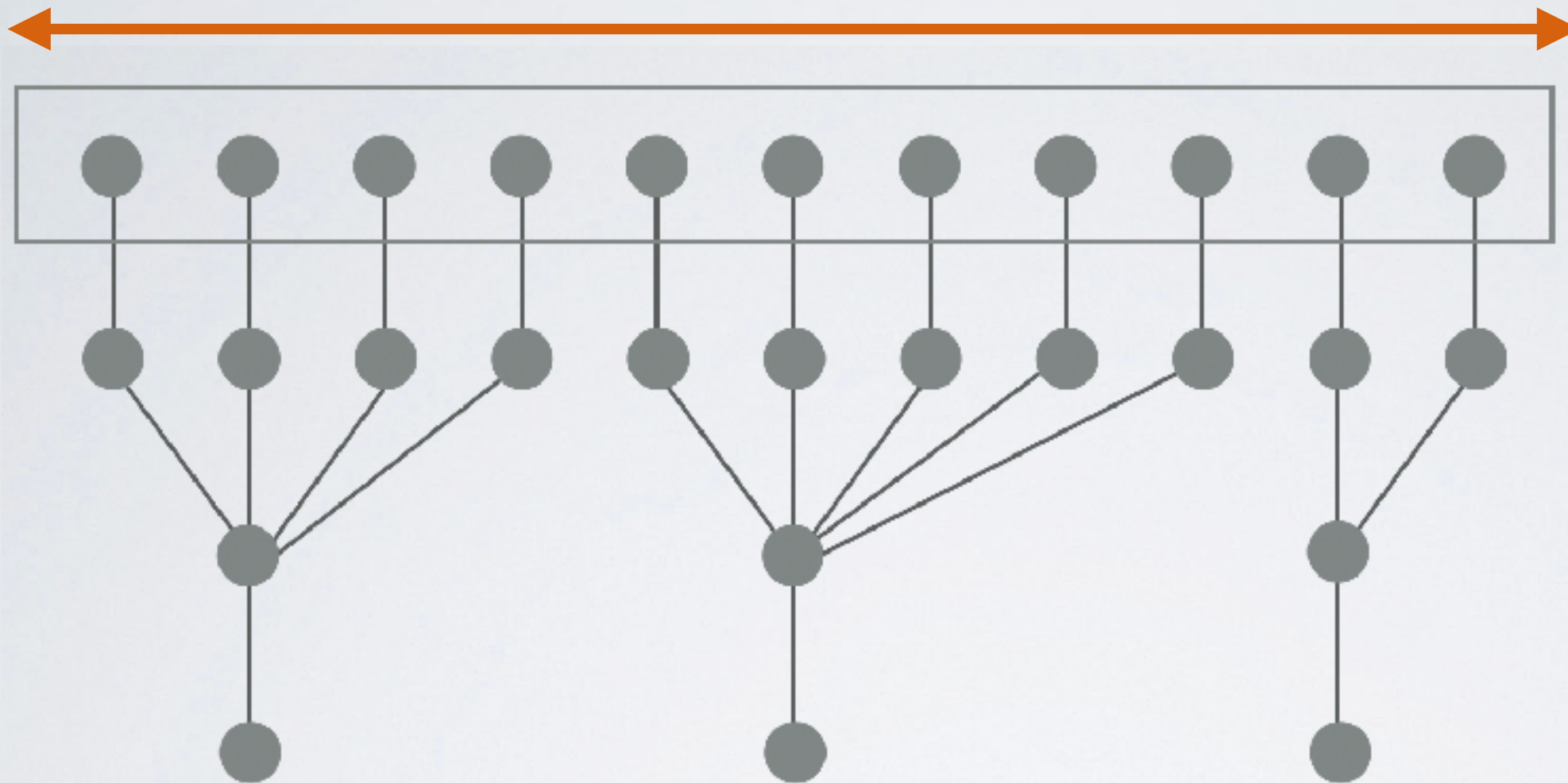
# IN PRINCIPLE



S-Paths can handle a deep model

# IN REALITY

number of entities of a type



Depth  
of relevant  
paths

Cost of query

# APPLICATIONS

- reveal defects in data sources
- visualize modeling specificities
- show trends in the data that can be used for communication towards end users.

# EVALUATION

- On the applications
- About readability / understanding

<http://s-paths.lri.fr>

marie.destandau@inria.fr  
@marie\_ototoi