

Radar : View of complex related data

Patrice Godard (Thomson Reuters), Jonathan van Eyll and Frédéric Vanclef

UCB, Chemin du Foriest, Braine-l'Alleud, Belgium.

UCB, 216 Bath Road, Slough, Berkshire, UK.



Introduction: Complex Data

Main Question

How to compare targets and browse the different elements linked to these targets in an easy way.

What is complex data?

In our case it means:

- Many different data types.
- Direct and indirect relationships.
- All instances are not equivalent. A score indicates that relative information.
- Some extra status/metadata have to be presented.
 - Positive controls.
 - Negative controls.
 - Candidates
 - Highest phase for a compound related to the target.

The figure 1 is representing the data model of the current data set we would like to show to the scientist.

Targets: 900
Compounds: 17330
Diseases: 991
Pathways: 1203

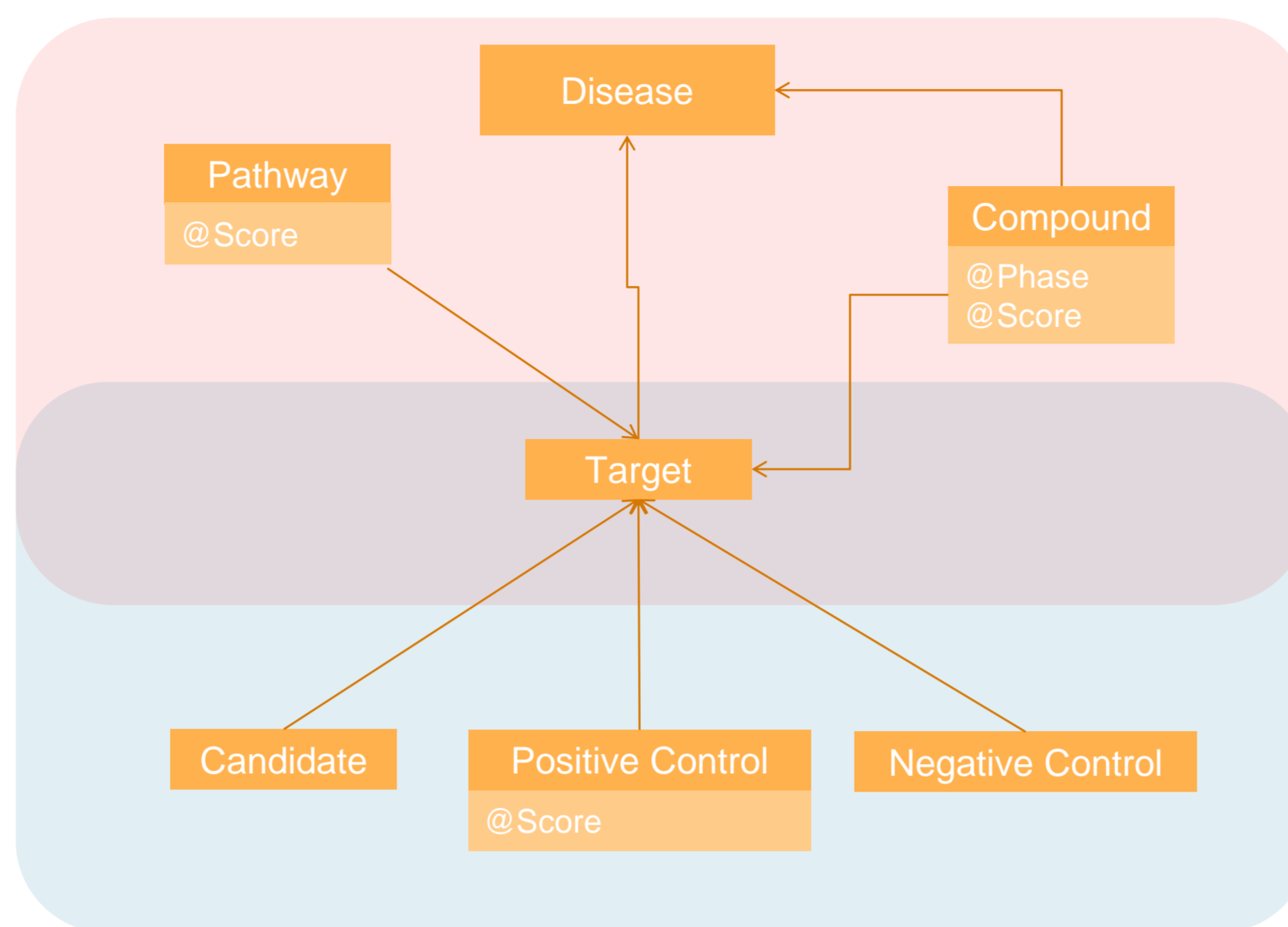


Fig. 1 Data Model (all scores are calculated for a specific disease)

Obvious Options

1. Master and detail tables

With this representation it is difficult for the scientist to focus on interesting records. He has to filter on many attributes and to use the sort function to discover the most interesting elements. The main disadvantage is that records are made invisible by filtering. Scientist has to remove the filter and start again. When the number of records is greater than 50 you need also to implement scrolling.

2. Network

Relationships are well represented but it is really difficult to represent the score. The relationships drawing is making the representation confuse and difficult to read. How to highlight interesting elements when there is about 1000 nodes and more relationships.

The Designed Solution

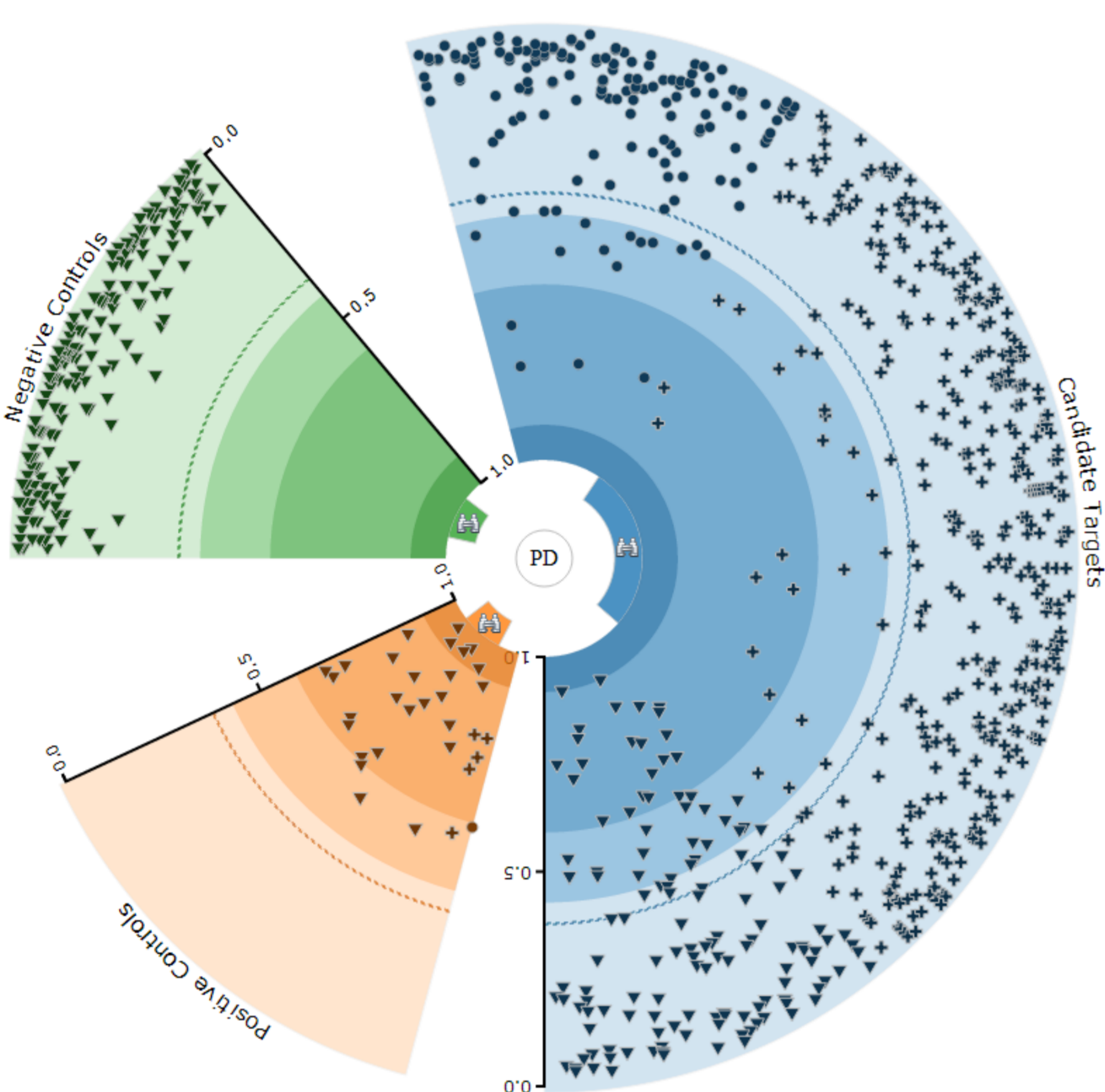


Fig. 2. Targets positioning

- The first choice was to position targets into a **disk** with the disease of interest placed at the center. The usage of a disk was helping to set the different points relative to the center. This provide a useful information to the scientist. If the point is closer to the center then it means that the target is strongly related to the disease of interest (This score calculation is out of the scope of this presentation).

- The second choice was to cut the disk into three arcs with a marge between them. This approach provides a quick overview of the status

- ✓ Blue Arc → Candidate Targets
- ✓ Orange Arc → Positive Controls
- ✓ Green Arc → Negative Controls

The size of each arc is adapted to the number of targets.

Positive controls are all pretty close to the center. At the opposite, the negative controls are all far from the center. If we apply some basic statistics to the positive controls we can then trace more sub arcs with different colors to show the distribution of the different targets. The idea was to apply the same values also for candidates and negatives.

- The third choice was to use 3 different shapes for target. The target inherits, the highest phase of all compounds in relation with the target and we regroup targets in three major groups based on that information.

- ✓ Circle → Preclinical
- ✓ Cross → Clinical
- ✓ Triangle → Launched

Advantages

This Radar view presents at a glance all the information without scrolling and sorting. You can find easily targets that are close to the center and select them.

We have implemented a filter and a lookup function in order to point quickly a target. A tooltip is also provided in order to see important properties

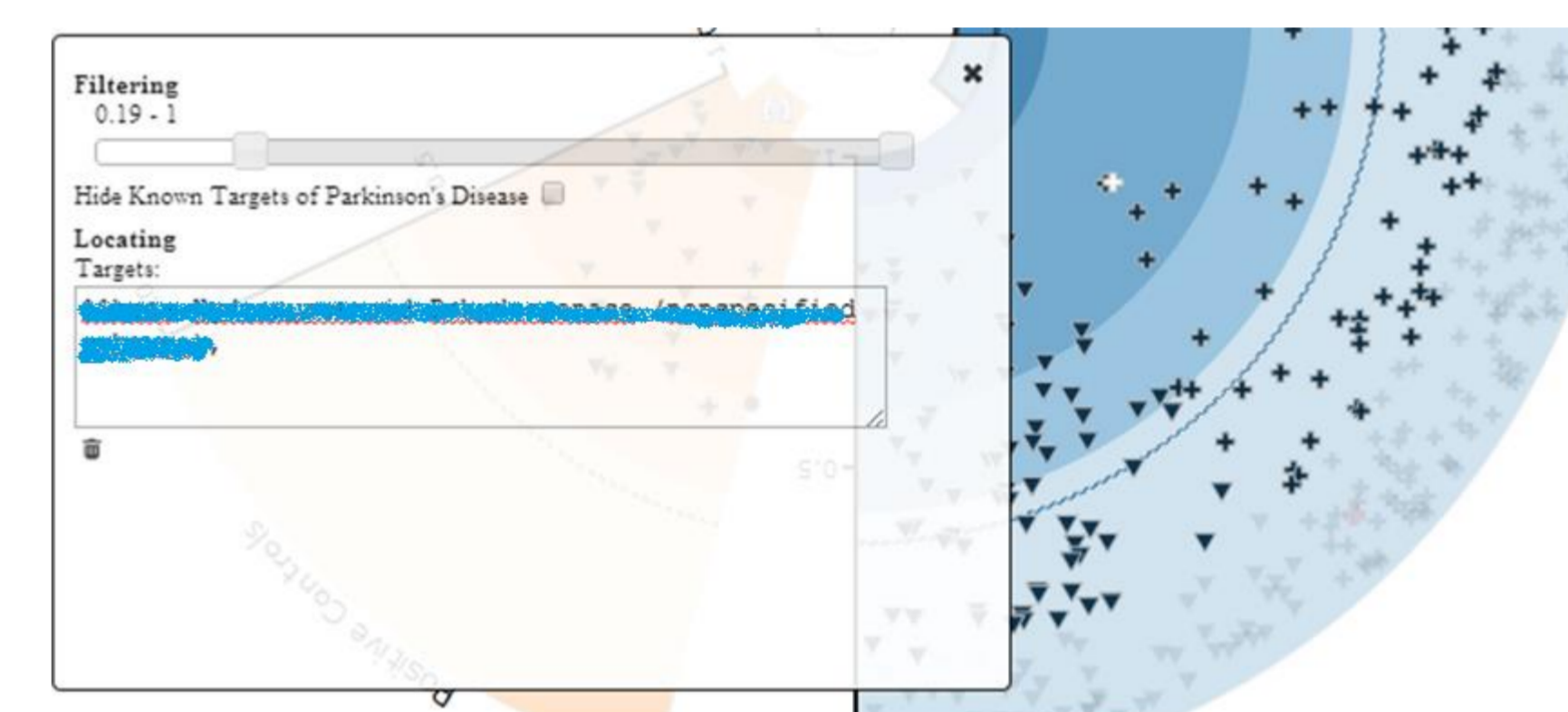


Fig. 3.1. Filtering and lookup

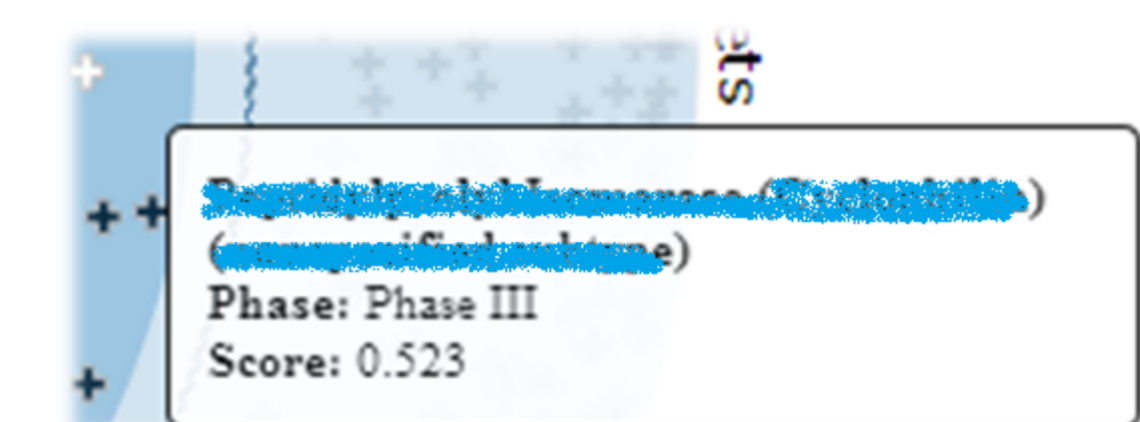


Fig. 3.2. Show Detail Tooltip

Showing Additional Information

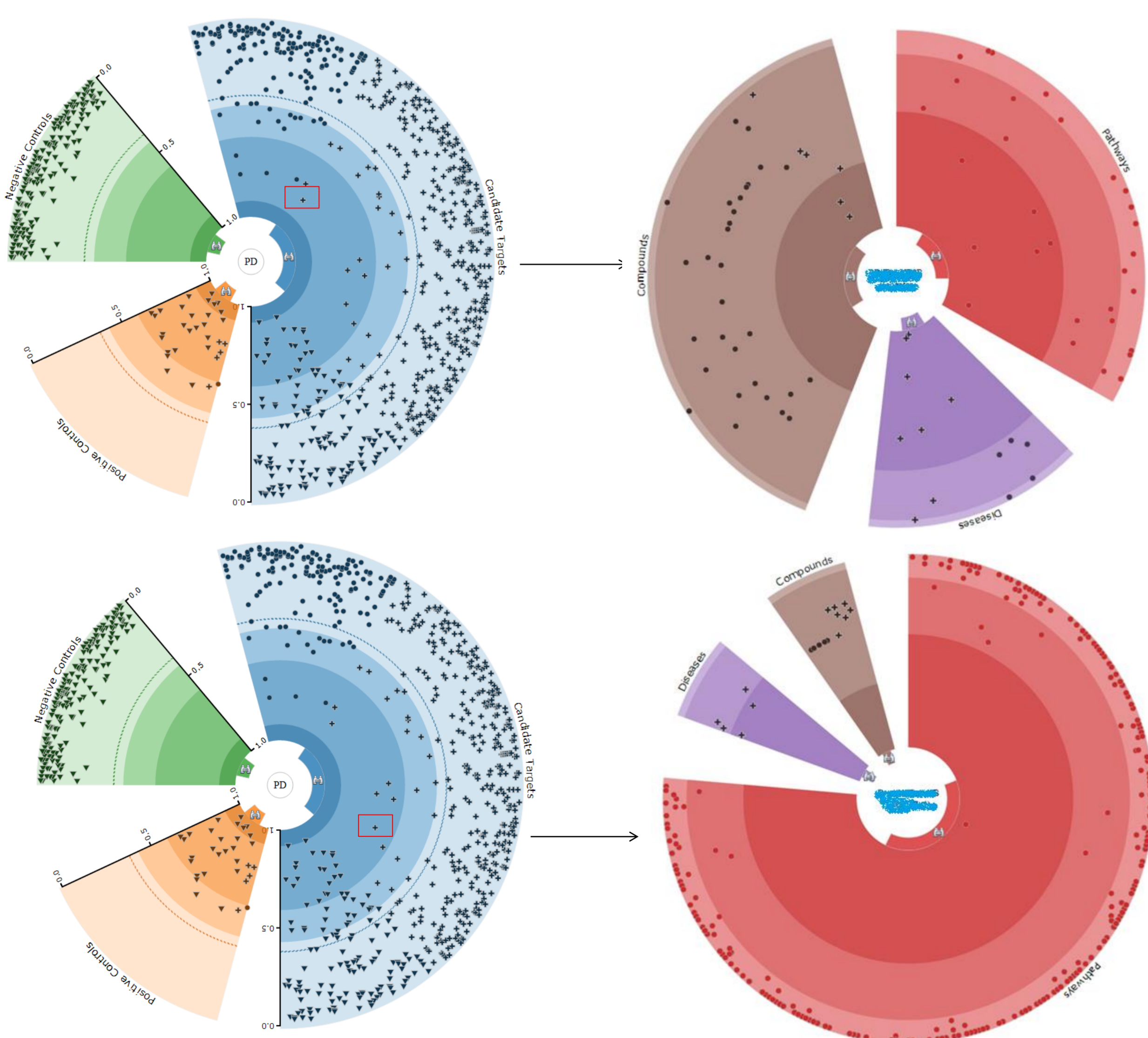


Fig. 4. Target selection and second disk rearrangement

It is important to apply some consistency in data visualization.

Also We have used the same disk representation for compounds, related diseases and pathways.

- Each time a target is selected the second disk is adapting. The size of each disk arc is depending on the number of items (figure 4).

- ✓ Brown Arc → Compounds
- ✓ Red Arc → Pathways
- ✓ Purple Arc → Diseases

- We have also implemented retro filter. It means when the user click on a disease or a pathway, targets that are not concerned are made fade.

- We are also showing relationships between diseases and compounds. As soon as a disease or a compound is selected the corresponding compounds or diseases are highlighted (figures 5.1 and 5.2).

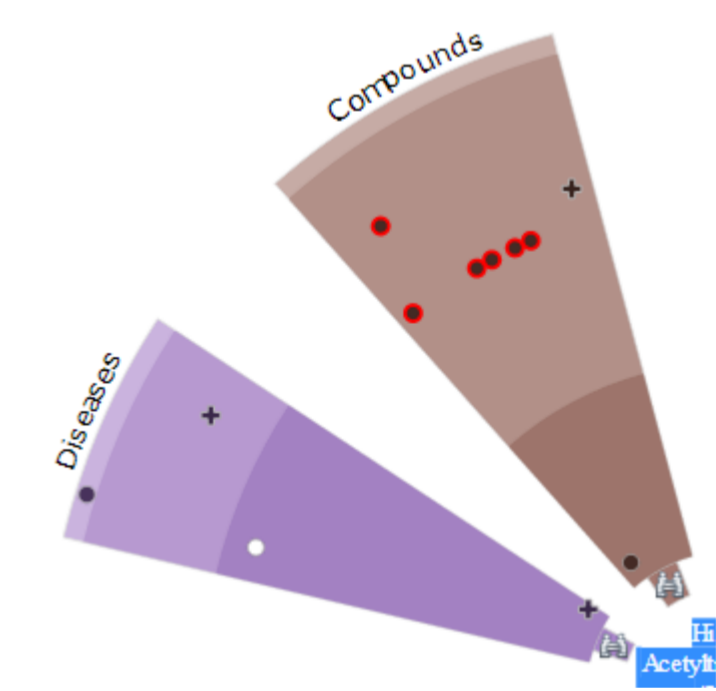


Fig. 5.1. Disease to compounds relation

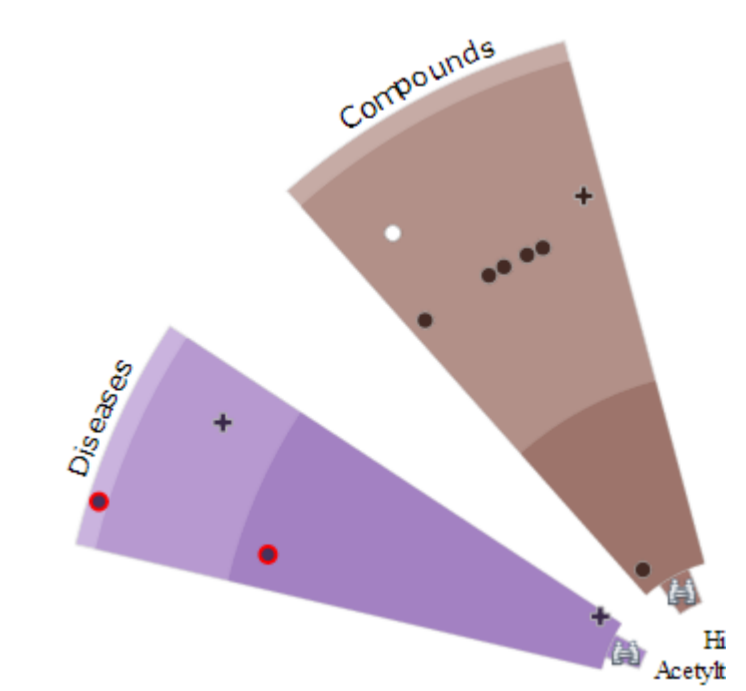


Fig. 5.2. Compound to diseases relation

Key Success Factors

- Clear question implies clear answer.
- The data preparation has been considered the most important step. Because of that approach that data structure was really stable and so the visualization was developed straightforward.
- The D3 library for generating the graph is a wonderful toolbox for generating SVG inside Web page. Implementation duration was about 7 days. (<http://d3js.org>)

Acknowledgments

UCB: Daniel Smaragd.

Thomson Reuters: Integrity (Compounds, Diseases, Targets); Metabase (Pathways)

Mike Bostock for his great D3 library.