

A few thoughts about similarity and sorting

Markus Christen

The problem

We want to measure people's appreciation of the «similarity» between terms, concepts, ideas, arguments (phrased in single statements) for doing two things:

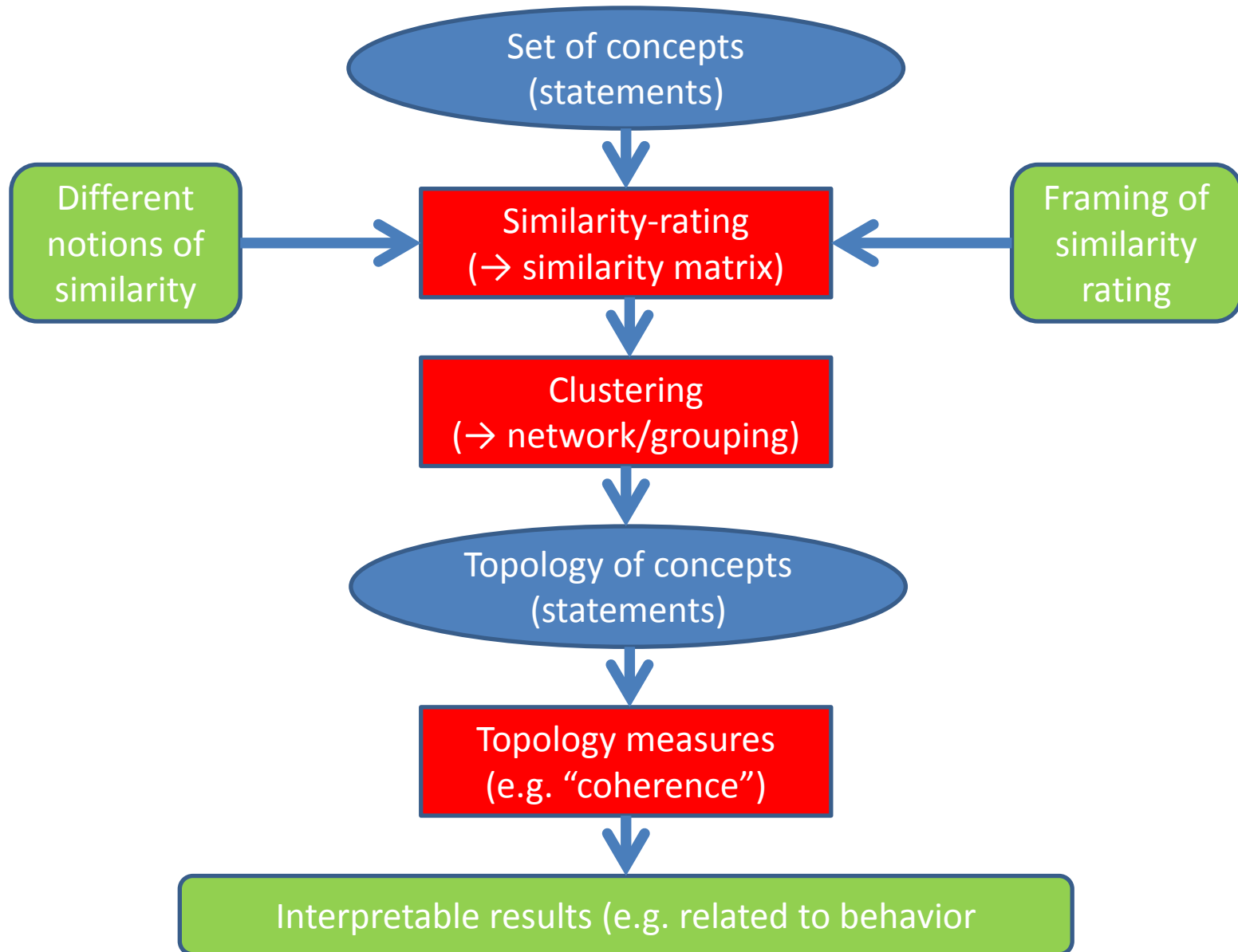
- To get an idea of the «topology» of the conceptual space
- To see how this topology changes based on different notions of «similarity» or on ways to influence similarity due to e.g. framing.

Similarity

With respect to methodology, we want to get similarity ratings (or distance ratings) in order to be able to apply them in a clustering framework.

Clustering should provide us a categorization in terms of a network (either hierarchy, or projection of a 2-D-space) and the network-topology is then investigated further.

Basic methodological setup

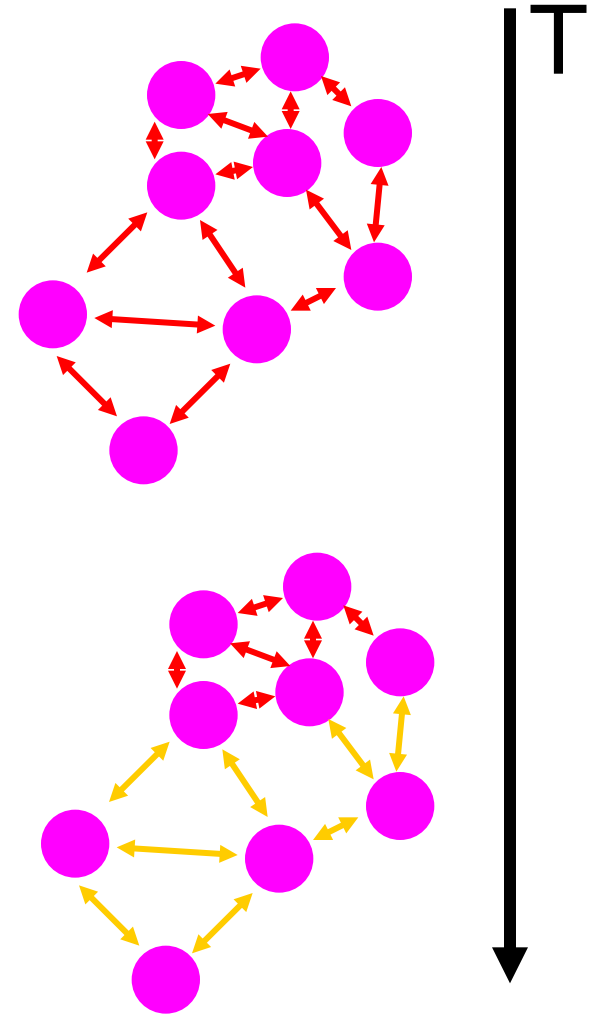


Superparamagnetic Clustering

Superparamagnetic clustering SPC (Blatt et al. 1996) is inspired by a self-organization phenomenon in magnetic spin-system.

It has been extended by Ott et al. (2005) to a powerful classification tool (sequential SPC) that has several advantages:

- No pre-definition of number and size of clusters required
- Temperature as “stability parameter”
- Natural hierarchy of sub-groups
- Choice of distance function allows adaptation to specific problem



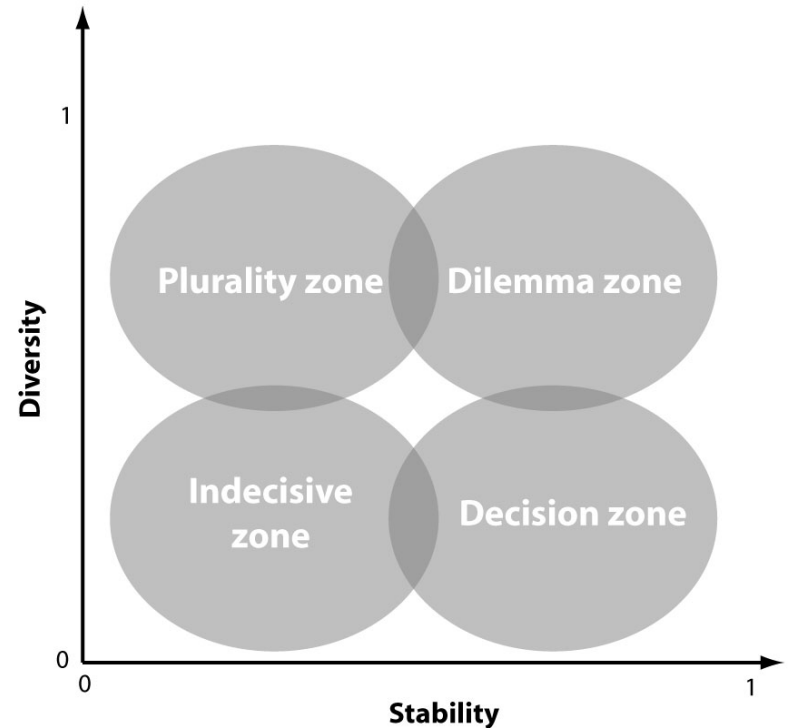
Topology-measure: Coherence

If we understand a belief system as a network of beliefs, this network probably displays sub-structures that can be understood as clusters of beliefs with stronger mutual interrelations compared to beliefs from other clusters.

Furthermore, these structures may display some property of stability that depends on the strength of the mutual interrelations of beliefs.

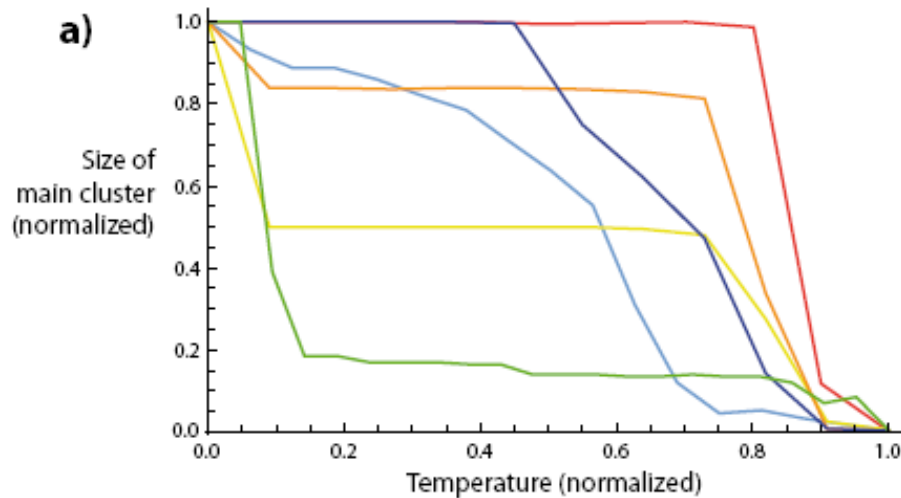
Therefore, we distinguish between

- **Diversity**
- **Stability**



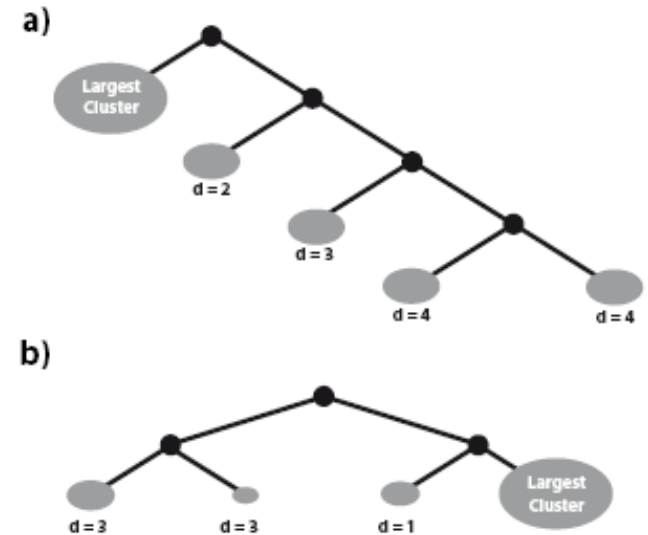
Quantified Coherence

Stability



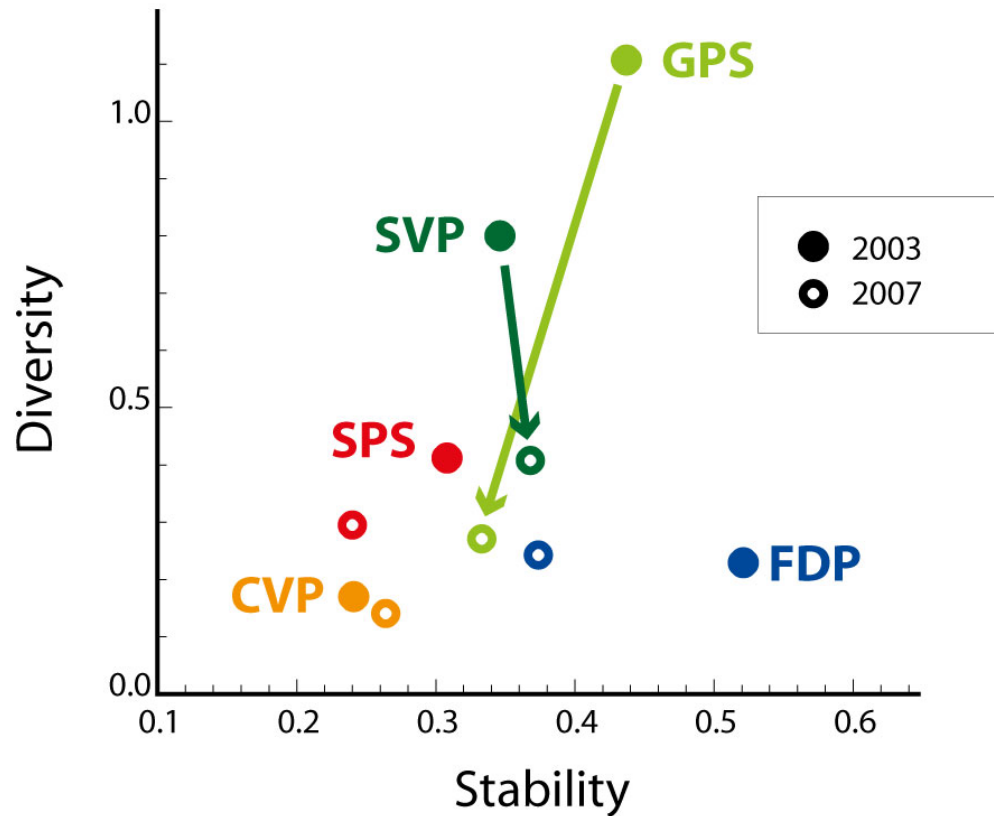
$$C_{dynamic} = \sum_{i=0}^{l-1} \frac{CS(i\Delta T) + CS((i+1)\Delta T)}{2nl} \quad (3)$$

Diversity



$$C_{static} = \sum_{i=1}^k \frac{d_i}{d_{max}} \cdot \frac{|c_i|}{|\bar{c}|}$$

Coherence and Party Stability



Similarity-Assessment (1)

We want to have a similarity assessment task that is:

- Suitable for web-interface
- “Bottom-up” (i.e. not “holistic” like a task that involves, e.g., having all concepts printed on paper and let people sort on tables”
- Robust for order-effects.

Similarity-Assessment (2)

Possible methods:

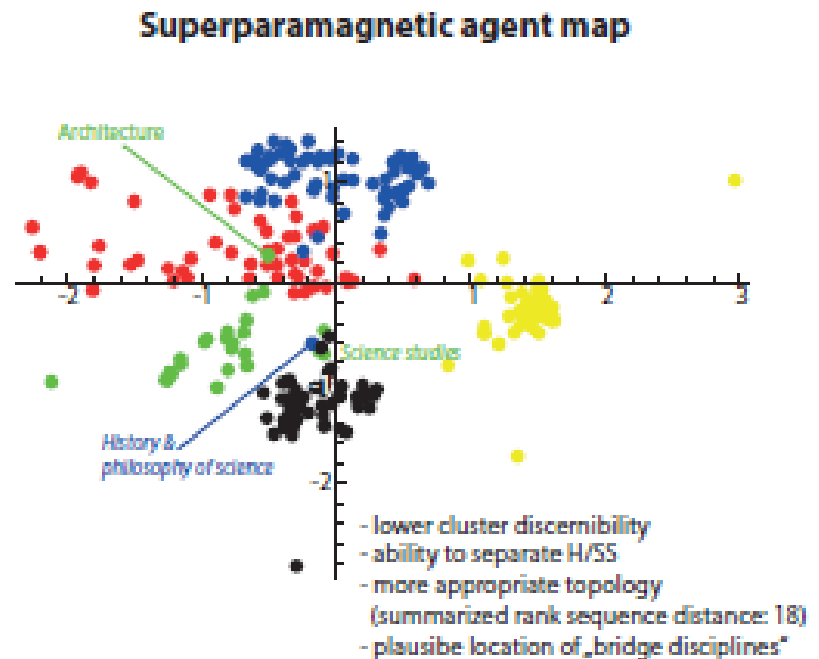
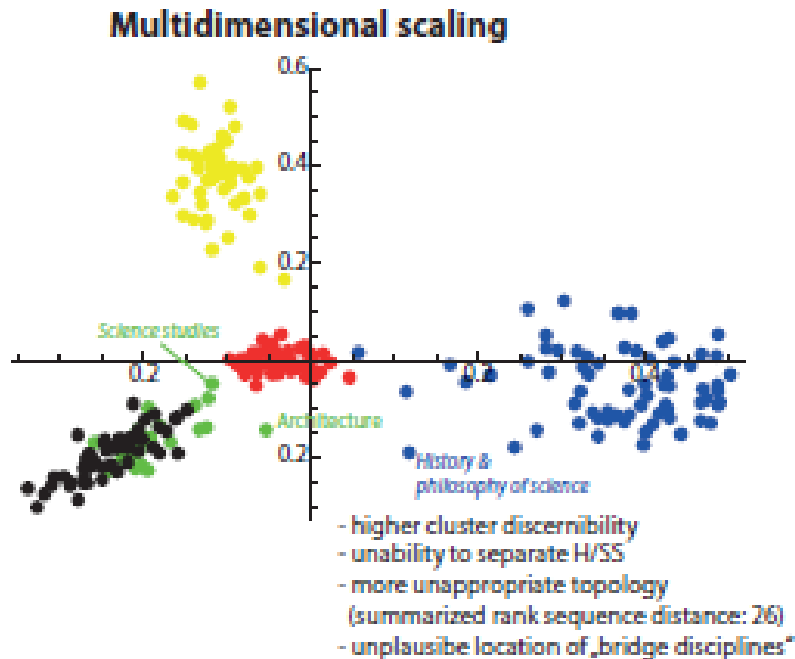
- Pairwise-comparison, using a Lickert-scale for assessing similarity
- Triplet-comparison:



We believe that latter may be more robust for order effects (but requires more comparisons)

Current projects (1)

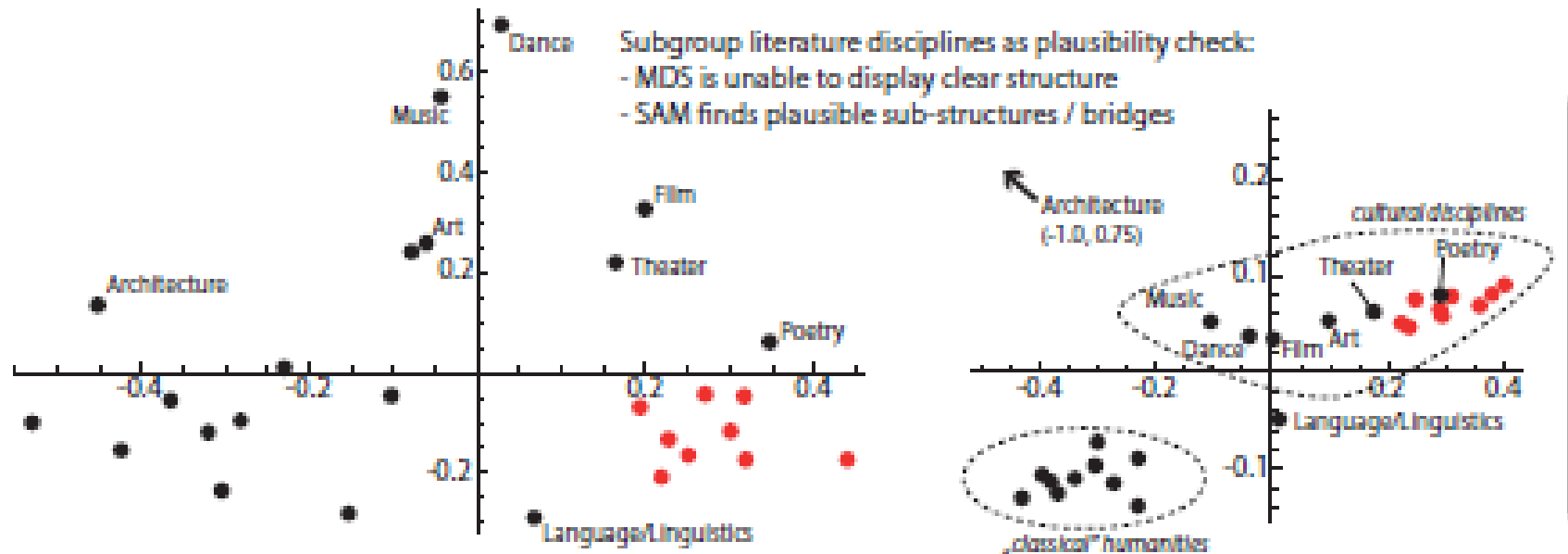
Creating a new “map of science”:



Current projects (2)

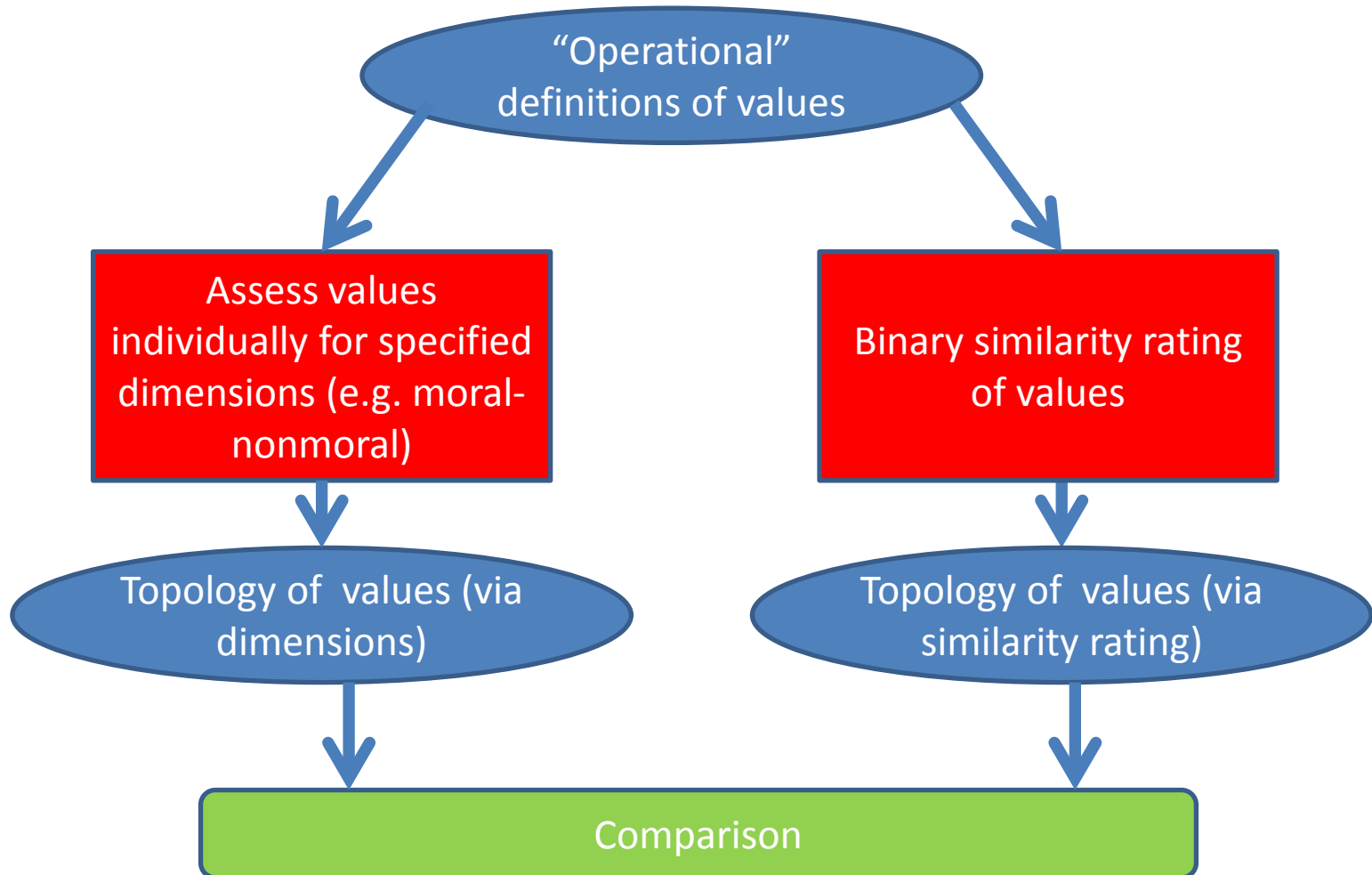
Creating a new “map of science”:

Comparison MDS and SAM for Humanities disciplines



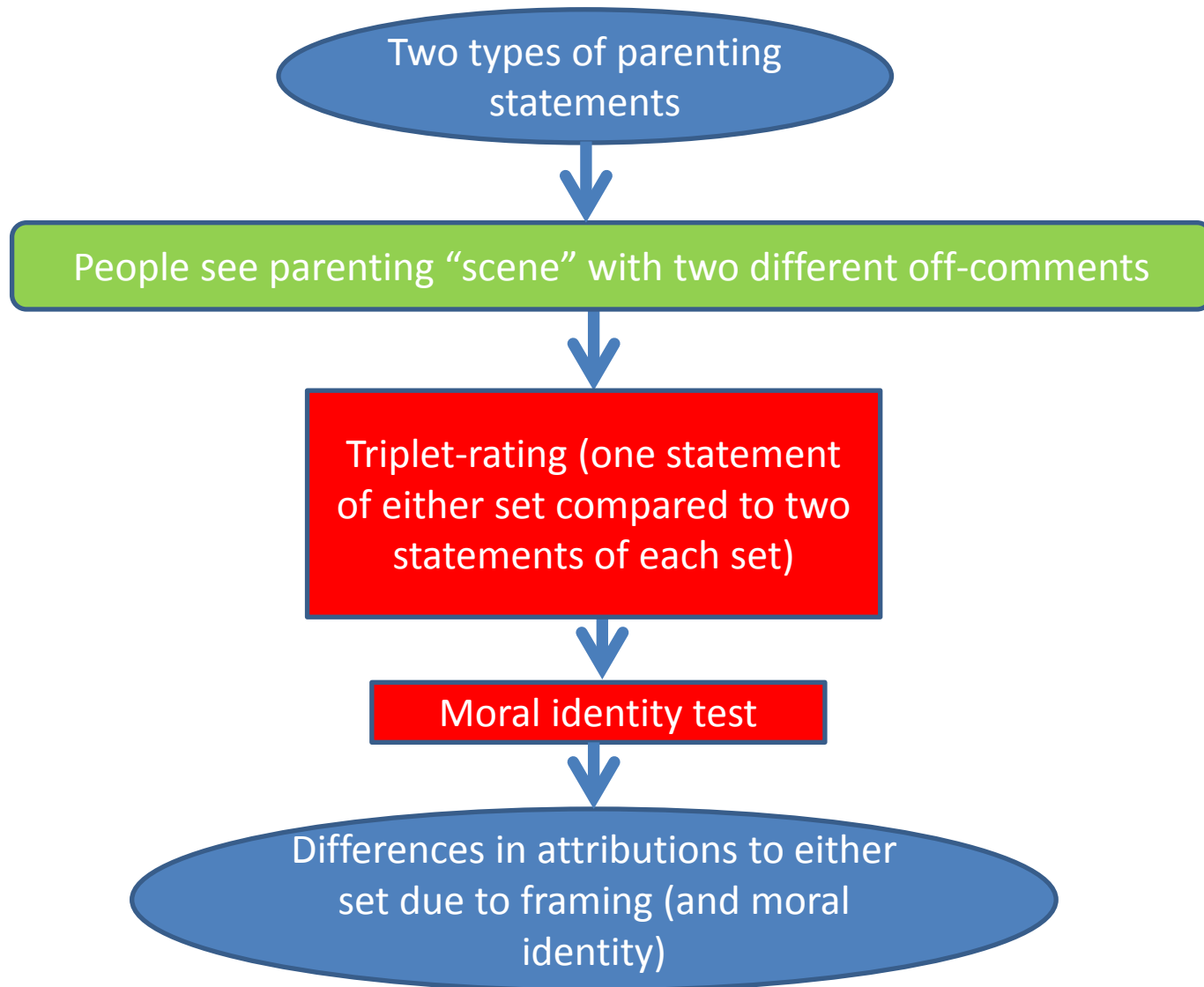
Current projects (3)

Similarity of Values:



Current projects (4)

Changing similarity rating through framing:



Questions

- Similarity: How to address different types of similarity (and to be sure that subjects actually used this type)?
- Sorting paradigms? Do they make sense? Are there better ones?
- Is there maybe a “holistic” sorting approach for a web-based instrument?