Exploring Co-Occurrence on a Meso and Global Level Using Network Analysis and Rule Mining

Abstract

Starting from a bipartite classification network of objects and classification criteria, we use a natured algorithm to explore the ecology of classification co-occurrences. Enabling meso-level exploration, we construct and sketch a weighted network of classification co-occurrence with a useful size-sigificance measure based on learned association rules. Existing global level explorations, we use a hierarchical link clustering (HCL) to connect term-ranking communities to co-occurrence networks. Finally, we use a co-occurrence network to connect term-ranking communities to co-occurrence networks, resulting in a community overlap network. Finally, exposing clear stories from the results including evolution in time, we offer important insights regarding the structure of classical archaeology as an discipline, while making interesting and applying our techniques to similar datasets covering other disciplines.

Walk through the figures

The figure sequence makes clear how we enable meso- and global-level exploration of subject classification beyond the traditional.

A. A data model for a source dataset Archeological Bibliography, including (left) a visualisation of the inherent tree of subject headings with general subject themes, topics, periods, and events; (below right) the full distribution for classification co-occurrences in publications from 1980 to 2020.

B. Our data analysis and visualisation pipelines includes (a) a network projection of the publication-classification or author-classification network to create the creation and visualisation of sub-driven classification by significance weight in addition to the regular co-occurrence link weight; and (b) the creation and visualisation of an overlapping community network, using VantagePoint (backdrop) and hierarchical link clustering (HCL).

C. A meso level classification maps act as cheat sheets. The meso level classification is a simple network of two identical networks, thresholded in different ways. Histo co-occurring topics are taken into account in a simple network, while additional topics are included if their significance is at least 0.50.

D. Another example, observed in figure 6, where several topics normally fall themselves in between clusters from the network. A classification network is a result of the process.

E. An example of classification evolution in publications with topic significance for the binary topic Tree and Trunk, i.e., a sub- classification in the network of subject headings of Archeological Bibliography. The picture, which can be seen in an instant click for any user, shows how themes are interconnected in the network, while additional topics are only included if their significance is at least 0.50.

F. Details of the global data community overlap network for co-occurrences, visualised as an activity diagram and edges split by topic, indicating the inherent frequency of classification topics, i.e., object themes, locations, periods, persons and objects.

G. Isolating the key community overlap network for each period, corresponding to subject themes, locations, persons, and persons, reveals that key themes are distinct in different ways.

H. Global level classification evolutions in publications as well as in networks evolve over time. Following such structure that emerges early in the process. Histo co-occurrences of various temporal snapshots, can be connected using a diagonal algorithm that results in interesting merging and splitting over time, indicating both diversification and generalisation.

Conclusion

In this paper, we explore a novel data-driven theme in classical archaeology, as measured in Archeological Bibliography, that granular components of co-occurrence, which we can exploit to offer a meso level where themes are connected by co-occurrence as well as on a global level where communities are connected by these overlaps. Figure F and G points to another, even more global level, where we may detect changes in the same evolution. In order to detect these conceptual contents of the academic discipline, which might be more meaningful in other domains, it seems not to use our pipeline for another time.

Acknowledgements

Maximilian Schich and Gerhard Schreiber (Delft) have, in an earlier version of this work, developed a method to measure the significance of classification co-occurrences. Further, the authors would like to thank the reviewers for their valuable feedback. This work has been supported in part by the Netherlands Organization for Scientific Research (NWO) under the VEN. The authors received a grant from the Netherlands Organization for Scientific Research (NWO) under the VEN. The authors also thank the Netherlands Organization for Scientific Research (NWO) for financial support.

Reference summary


Full paper: http://go.to/Osrl5