Concepts co-occurrence for the identification of higher order concepts in Mathematics articles.

Vsevolod Salnikov[1], Renaud Lambiotte[1,2], and Daniele Cassese[1]

[1] University of Namur, Namur 5000, Belgium,
[2] Mathematical Institute, University of Oxford, Oxford OX2 6GG, UK,
vsevolod.salnikov@unamur.be
renaud.lambiotte@maths.ox.ac.uk
daniele.cassese@unamur.be

We analyse co-occurrence of mathematical concepts in a simplicial complex framework, going beyond the network description that reduces all the structural properties to pairwise interaction and their combinations where we are interested in higher-order relations and we have a rich data structure.

The dataset includes 54K arXiv articles up to 2008, from each we extract its mathematical content, taken from a 1618 concepts list obtained from Wikipedia. The resulting network (1618 nodes, 25277 edges) is very dense, hence exploratory analysis of communities with usual methods is not very informative as we prefer not to threshold edges not to rule out that low-weight connections reveal important bridges between conceptually different areas.

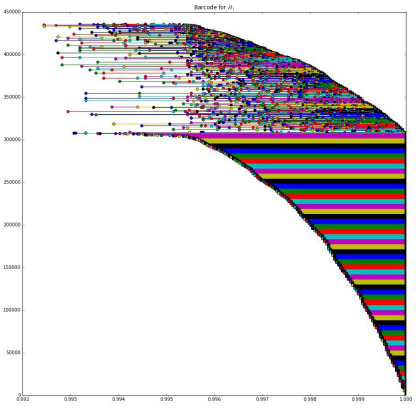

Fig. 1. Barcodes for H1

In the co-occurrence simplicial complex each article containing k concepts is a $(k-1)$-simplex, and we use articles dates to build a temporal filtration $\mathscr{F} = \{\mathscr{F}_0, \mathscr{F}_1, \ldots, \mathscr{F}_T\}$ where $(0,T)$ are first and last date in our dataset and $\mathscr{F}_i \subseteq \mathscr{F}_j$ for $i < j$ and each $\mathscr{F}_i$ co-occurrence network contains concepts from articles up to date i. We then compute Persistent Homology (PH) to study homological cycles, and to reduce the computational burden we find cliques up to dimension $k_{\max} = h+1$, where h is the dimension of the homological holes we are interested in. We focus on homological cycles of dimension 1 and 2 (2 and 3 dimensional holes bounded by edges and triangles respectively) that do not persist till the end of the filtration (at the top of the barcode in Fig. [1]), and we find that longer cycles have longer persistence.

Out of 10 most frequent concepts in H1 and H2, around 6 of them are also in the cores of S^3 stars, providing insights on their structural importance, and motivation to perform further analysis, including classification of authors according to their expertise and their role in the creation of homological holes and stars.