

# Magnitude Homology - Report

November 21, 2022

## 1 Goals

1. Develop a library to efficiently compute magnitude homology. To do this we construct a matrix representing the MH boundary operator.
2. Use MH for network analysis. Our hope is that MH can give us information about the structure of a network and how a particular structure influences information flow.

## 2 Magnitude Homology framework

**Definition 1.** Let  $G = (V, E)$  be a simple graph. We define the  $(k, \ell)$ -magnitude chain,  $MC_{k, \ell}(G)$ , as the free abelian group generated by the  $(k + 1)$ -tuples of vertices of  $G$  such that the path  $(x_0, \dots, x_k)$  has length  $\ell$ . That is,

$$MC_{k, \ell} = \langle (x_0, \dots, x_k) : (x_0, \dots, x_k) \in V^{k+1}, x_i \neq x_{i+1}, l(x_0, \dots, x_k) = \ell \rangle.$$

The  $\ell$ -magnitude complex  $MC_{*, \ell}(G)$  is the direct sum over  $k$  of all  $(k, \ell)$ -magnitude chains,

$$MC_{*, \ell}(G) = \bigoplus_{k \geq 0} MC_{k, \ell}(G).$$

**Definition 2.** The boundary operator  $\partial_{k, \ell} : MC_{k, \ell}(G) \rightarrow MC_{k-1, \ell}(G)$  is defined as follows:

$$\partial_{k, \ell}(x_0, \dots, x_k) = \sum_{i=0}^k a_i(x_0, \dots, \hat{x}_i, \dots, x_k), \text{ where}$$

$$a_i = \begin{cases} (-1)^i, & \text{if } l(x_0, \dots, \hat{x}_i, \dots, x_k) = \ell \\ 0, & \text{otherwise.} \end{cases}$$

**Definition 3.** The  $(k, \ell)$ -magnitude homology group,  $MH_{k, \ell}(G)$ , is defined as

$$MH_{k, \ell}(G) = \frac{\ker \partial_{k, \ell}}{\text{im } \partial_{k+1, \ell}}$$

## 2.1 Operative definition

Notice that the information provided by  $MC_{k,l}$  is extremely noisy. This is because the definition only asks that each vertex  $x_i$  of the  $k$ -tuple is different from  $x_{i+1}$ , and this does not prevent going back and forth between the same two vertices. That is, a generator of  $MC_{4,4}$  is for example  $(x_0, x_1, x_0, x_1, x_0)$ .

We attempt to overcome this issue with a slight modification of the definition, in which we erase from  $MC_{k,l}$  all paths revisiting a vertex.

**Definition 4.** We define the  $(k, \ell)$ -reduced magnitude chain,  $MC_{k,\ell}^{red}(G)$ , as the free abelian group generated by the  $(k+1)$ -tuples of different vertices of  $G$  such that the path  $(x_0, \dots, x_k)$  has length  $\ell$ . That is,

$$MC_{k,\ell}^{red} = \langle (x_0, \dots, x_k) : (x_0, \dots, x_k) \in V^{k+1}, x_i \neq x_j, l(x_0, \dots, x_k) = \ell \rangle.$$

The  $\ell$ -reduced magnitude complex  $MC_{*,\ell}^{red}(G)$  is the direct sum over  $k$  of all  $(k, \ell)$ -reduced magnitude chains,

$$MC_{*,\ell}^{red}(G) = \bigoplus_{k \geq 0} MC_{k,\ell}^{red}(G).$$

Since from now on we will only rely on  $MC_{k,\ell}^{red}(G)$ , we will indicate (with an abuse of notation) the  $(k, \ell)$ -reduced magnitude chain as  $MC_{k,\ell}(G)$ .

## 3 Matrix representing $\partial_{k,\ell}$

### MODIFY HERE AND DESCRIBE SPARSE MATRIX

The matrix  $\Delta_{k,\ell}$  representing  $\partial_{k,\ell}$  is constructed using the following algorithm:

1. find the tuples generating  $MC_{k,\ell}(G)$  and  $MC_{k-1,\ell}(G)$
2. initialize an all-zeros matrix of dimension  $MC_{k-1,\ell}(G) \times MC_{k,\ell}(G)$
3. for  $t \in MC_{k,\ell}(G)$ , if  $\partial_{k,\ell}(t) = t' \in MC_{k-1,\ell}(G)$  change the entry  $(t', t)$  to  $-1$ .

### 3.1 Complexity

### 3.2 Code

The code can be found in this repository.

### 3.3 Experiments

This matrix was proven to be effective in the computation of magnitude Betti numbers  $\beta_{k,\ell}$  using examples taken from the paper “Categorifying the magnitude of a graph”, Hepworth and Willerton (arXiv: 1505.04125v2).

## 4 Interpretation of the rank of $MH_{k,l}$ .

### 4.1 Diagonal $MH_{k,k}$

As for the original definition given by Hepworth and Willerton, the ranks of the  $(0,0)$  and  $(1,1)$  MH groups of a graph  $G = (V, E)$  represent the cardinality of  $V$  and the cardinality of  $E$ , respectively.

$MH_{k,k}$  provides us with the precise number of 3-cycles and 4-cycles contained in a graph (modulo automorphisms). Indeed, consider the following facts:

- the dimension of the kernel of our matrix  $\Delta_{k,k}$  is equal to the number of all-zero columns plus the number of columns that are “copies” of a previously written column.
- the number of all-zero columns is (modulo automorphisms) equal to the number of triangles contained in the graph. This is because if a  $(k,k)$ -tuple  $(x_0, x_1, \dots, x_k)$  is sent to zero after removing a vertex  $x_i$ , it means that the shortest path between  $x_{i-1}$  and  $x_{i+1}$  has length smaller than 2. So there exists an edge  $(x_{i-1}, x_{i+1})$  and equivalently a triangle  $(x_{i-1}, x_i, x_{i+1})$ .
- the number of “repeated” columns indicates how many  $(k,k)$ -tuples  $(x_0, \dots, x_i, \dots, x_k)$  are sent to the same  $(k-1, k)$ -tuple  $(x_0, \dots, \hat{x}_i, \dots, x_k)$ , and this provides us with upper and lower bounds for the number (modulo automorphisms) of 4-cycles (EXPAND).

Therefore, to obtain the number of 3-cycles contained in our graph we need to divide the number of all-zero columns by 6, i.e. by the cardinality of the automorphisms group of the triangle  $D_3$ .

**Remark 5.** *We can link this value to the global and local clustering coefficients, and to the cycle ratio defined in [PUT REFERENCE HERE]. In case of a directed graph (i.e. network) this value accounts for transitivity.*

To obtain the number of 4-cycles we first divide the number of “repeated” columns by two, in order to disregard the orientation, and then we again divide by two, so that two  $(k-1, k)$ -paths are glued into the same 4-cycle.

**Example 6.** *PUT HERE EXAMPLE OF ICOSAHEDRAL GRAPH.*

Although the  $(k,k)$ -magnitude homology groups  $MH_{k,k}$  all provide the same information, we point out that the rank of  $MK_{k,k}$  might be different from the rank of  $MK_{k',k'}$  when  $k \neq k'$ , and in particular  $\text{rank}(MH_{k,k}) < \text{rank}(MH_{k',k'})$  for  $k < k'$ . Consider for example the case of the tetrahedral graph presented above. Here we have  $\text{rank}(MH_{2,2}) = 180$  and  $\text{rank}(MH_{3,3}) = 252$ , and this is because some cycles are counted more than once in  $MH_{3,3}$ . For example, the tuples  $(11, 7, 9, 10), (4, 10, 9, 7) \in MC_{3,3}$  are both sent to 0 because of the triangle  $(7, 9, 10)$  (PUT DIFFERENT EXAMPLE).

Given this and considering the (obvious) fact that  $MH_{2,2}$  is much faster to compute, in the future analysis we will just make use of  $MH_{2,2}$ .

## 4.2 Second diagonal $MH_{k-1,k}$

We are not yet able to provide a general interpretation for all magnitude homology groups  $MH_{k-1,k}$  on the second diagonal, but we believe that  $MH_{2,3}$  contains information regarding the number of 4-cliques, 5-cliques and 6-cliques in the graph.

Consider the following chain

$$\begin{array}{ccccccc} \dots & \rightarrow & MC_{3,3} & \rightarrow & MC_{2,3} & \rightarrow & MC_{1,3} & \rightarrow & 0 \\ & & (x_0, x_1, x_2, x_3) & \rightarrow & (x_0, \hat{x}_1, x_2, x_3) & \rightarrow & (x_0, \hat{x}_1, \hat{x}_2, x_3) & & \end{array}$$

and assume  $x_3 \neq \hat{x}_1$ .

If  $(x_0, \hat{x}_1, x_2, x_3) \in \ker(\partial_{2,3})$  then one of the following is true:

- Two different tuples in  $MC_{2,3}$  are sent to the same tuple in  $MC_{1,3}$ , which means there is either a 4-cycle or a 6-cycle.
- The considered tuple in  $MC_{2,3}$  is sent to zero, which means there exists either a 4-cycle or a 5-cycle.

Now, when we quotient by the image of  $\partial_{3,3}$  we are in fact disregarding the elements  $(x_0, \hat{x}_1, x_2, x_3) \in MC_{2,3}$  such that  $\ell(x_0, \hat{x}_1, x_2, x_3) = \ell(x_0, x_1, x_2, x_3)$ . That is, we are disregarding the tuples that do now contain the triangle  $(x_0, x_1, x_2)$ , and that therefore cannot be part of a clique.

Summarizing,  $MH_{2,3}$  contains information about 4, 5, 6-cycles that contain all triangles, and this means counting 4-cliques and candidates 5, 6-cliques.

**Remark 7.** *We point out that the hypothesis " $x_3 \neq \hat{x}_1$ " is crucial to obtain this interpretation. Indeed, without this assumption it could happen that  $x_3 = \hat{x}_1$ , which would mean "revisiting an edge" and adding a lot of noise to  $MH_{2,3}$ .*

We recall the definition of *diagonal graph* introduced by Hepworth and Willerton in [ADD REF].

**Definition 8.** *A graph  $G$  is called diagonal if  $MH_{k,l}(G) = 0$  whenever  $k \neq l$ .*

What we noticed until now suggest the following fact

**Proposition 9.** *If a graph  $G$  is diagonal, then it is clique-free.*

*Proof.* Suppose  $G$  is diagonal, then  $MH_{k,l}(G) = 0$  whenever  $k \neq l$ . In particular  $MH_{2,3} = 0$ , meaning the graph contains no 4-clique, and therefore no bigger clique.  $\square$

## 5 Problems to solve

1. The information in the magnitude homology groups  $MH_{2,2}$  and  $MH_{2,3}$  is "not divided", meaning we are just given the dimension of the kernel without the distinction between all-zero columns and repeated columns.
2. Add in the software the hypothesis " $x_3 \neq \hat{x}_1$ ". This is not trivial because the software doesn't really "see"  $\hat{x}_1$ , it just computes the length of the path supported by the tuple  $(x_0, \hat{x}_1, x_2, x_3)$ .

## 6 Ideas for the future

### 1. Network analysis:

- Construct a time series using active nodes of a network
- Detect small cycles and cliques using MH
- Detect persistent structures using PH

### 2. Prove that (see if.. but I think so) MH is a stable tool:

- Define an "interaction index" which should be a measure the tendency of a general vertex  $v$  to interact with other vertices in the graph (maybe taking inspiration from connectivity index [ADD RED] and cycle index [ADD RED]). Maybe the global clustering coefficient can be used?
- Define a distance between two graphs  $G$  and  $G'$  using the interaction index
- See if a small variation in the interaction implies a small variation in MH
- Maybe define a "MH diagram" following the idea of persistence diagrams and do the above point for this MH diagram