Magnitude Homology - Report

March 9, 2023

Contents

1 Goals

1. Study magnitude homology for Erdos-Renyi graphs.

2. Develop an algorithm tool to efficiently compute magnitude homology. To do this we start by constructing a matrix representing the MH boundary operator.

3. 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

An undirected graph is a pair $G = (V, E)$ where V is a set of vertices and E is a set of edges (unordered pairs of vertices). A *trail* in G is an ordered sequence of vertices $x_0, x_1, \ldots, x_n$ such that there is an edge $\{x_i, x_{i+1}\}$ for all $0 \leq i < k$; a *path* is a trail with no repeated vertices. For the purposes of defining magnitude homology, we assume all graphs to have no self-loops and no multiedges [6]. We may view such a graph G as an extended metric space (i.e. a metric space with infinity allowed as a distance) whose points are the vertices of G by setting each edge to be of length one and defining an extended metric $d : V \times V \to [0, \infty]$ by declaring $d(u, v)$ to be equal to the length of a shortest path in G from u to v. By definition we let $d(u, v) = \infty$ if u and v lie in different components of G. We recall Hepworth and Willerton's construction [3] of the magnitude homology groups of a graph.

A k-trail in a graph G is a $(k+1)$-tuple $(x_0, \ldots, x_k)$ of vertices with $x_i \neq x_{i+1}$ and $d(x_i, x_{i+1}) < \infty$ for every $i \leq k-1$. The length of a k-trail $(x_0, \ldots, x_k)$ in G is defined as the minimum length of a trail that visits $x_0, x_1, \ldots, x_k$ in this order, namely, $\ell(x_0, \ldots, x_k) = d(x_0, x_1) + \cdots + d(x_{k-1}, x_k)$. We define the (k, ℓ)-magnitude chain $MC_{k,\ell}(G)$ to be the free abelian group generated by the k-trails of length ℓ.

Definition 1. Let $(x_0, \ldots, \hat{x_i}, \ldots, x_k)$ denote the k-tuple obtained by removing the i-th vertex from the $(k+1)$-tuple $(x_0, \ldots, x_k)$. We define the differential

$$\partial_k : MC_{k,\ell}(G) \to MC_{k-1,\ell}(G)$$

as the sum $\partial_k = \sum_{i=1}^{k-1} \partial_{k,i}$ of the maps defined by

$$\partial_{k,i}(x_0, \ldots, x_k) = a_i \cdot (x_0, \ldots, \hat{x_i}, \ldots, x_k),$$

where

$$a_i = \begin{cases} (-1)^{i+1}, & \text{if } \ell(x_0, \ldots, \hat{x_i}, \ldots, x_k) = \ell, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 2 (Magnitude chain complex)**.** We indicate as $MC_{*,\ell}(G)$ the following sequence of free abelian groups connected by differentials

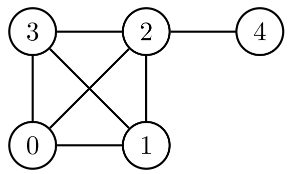
$$\cdots \to MC_{k+2,\ell}(G) \xrightarrow{\partial_{k+2}} MC_{k+1,\ell}(G) \xrightarrow{\partial_{k+1}} MC_{k,\ell}(G) \xrightarrow{\partial_k} MC_{k-1,\ell}(G) \to \cdots$$

It is shown in [3, Lemma 11] that the composition of two consecutive differentials $\partial_{k+1} \circ \partial_k$ vanishes, so that each chain $MC_{*,\ell}(G)$ is indeed a chain complex and it is thus possible to define its k-th homology group.

Definition 3. The k-magnitude homology group of the graph G at level ℓ is the abelian group defined by

$$MH_{k,l}(G) = H_k(MC_{*,l}(G)) = \frac{\ker(\partial_k)}{\operatorname{imm}(\partial_{k+1})}.$$

Example 4. Consider the following graph G



We want to compute $MH_{2,2}(G)$. From Definition 3 we need to evaluate the quotient between the kernel of ∂_2 and the image of ∂_3. So consider the map $\partial_2 : MC_{2,2}(G) \to MC_{1,2}(G)$. $MC_{2,2}(G)$ is generated by the 2-trails (i.e., triplets) in G of length 2. Therefore, $MC_{2,2}(G)$ is generated by the following 44 elements: (0, 1, 0), (0, 1, 2), (0, 1, 3), (0, 2, 0), (0, 2, 1), (0, 2, 3), (0, 2, 4), (0, 3, 0), (0, 3, 1), (0, 3, 2), (1, 0, 1), (1, 0, 2), (1, 0, 3), (1, 2, 0), (1, 2, 1), (1, 2, 3), (1, 2, 4), (1, 3, 0), (1, 3, 1), (1, 3, 2), (2, 0, 1), (2, 0, 2), (2, 0, 3), (2, 1, 0), (2, 1, 2), (2, 1, 3), (2, 3, 0), (2, 3, 1), (2, 3, 2), (2, 4, 2), (3, 0, 1), (3, 0, 2), (3, 0, 3), (3, 1, 0), (3, 1, 2), (3, 1, 3), (3, 2, 0), (3, 2, 1), (3, 2, 3), (3, 2, 4), (4, 2, 0), (4, 2, 1), (4, 2, 3), (4, 2, 4).

Similarly, $MC_{1,2}(G)$ is generated by the pairs representing 1-trails in G of length 2: (0, 4), (1, 4), (3, 4), (4, 0), (4, 1), (4, 3).

We thus have that the kernel of ∂_2 is generated by the 38 elements whose length diminishes when the middle vertex is removed. That is all elements in $MC_{2,2}(G)$ except (0, 2, 4), (1, 2, 4), (3, 2, 4), (4, 2, 0), (4, 2, 1), (4, 2, 3).

Being ∂_2 a linear operator we can represent it in the following way: we construct the associated matrix indexing the rows and the columns with the elements of $MC_{1,2}(G)$ and $MC_{2,2}(G)$ respectively. We then evaluate the differential as stated in Definition 1 and fill the matrix with the coefficients appearing in the evaluation. So for example, $\partial_2(0,1,2) = 0 \cdot (0,\hat{1},2) = 0 \cdot (0,2)$, since $\ell(0,1) < 2$, and thus the entry in position $((0,2),(0,1,2))$ will be zero. Similarly, $\partial_2(0,2,4) = 1 \cdot (0,\hat{2},4) = 1 \cdot (0,4)$, since $\ell(0,4) = 2$, and thus the entry in position $((0,4),(0,2,4))$ will be one.

For what concerns the image of ∂_3, we need to see which elements of $MC_{3,2}$ are sent to $MC_{2,2}$. But since $MC_{3,2}(G)$ is generated by the 4-tuples representing 3-trails in G of length 2, and since any path connecting 4 vertices must have length at least 3, $MC_{3,2}(G)$ is the trivial group $\langle 0 \rangle$. Therefore the image of ∂_3 is $\langle 0 \rangle$. In conclusion, we have that

$$MH_{2,2}(G) = \frac{\ker(\partial_2)}{\operatorname{imm}(\partial_3)} = \ker(\partial_2),$$

and so $|MH_{2,2}(G)| = 38$.

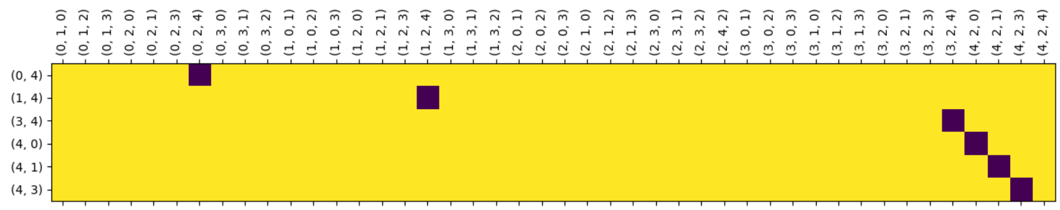
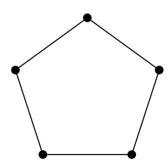


Figure 1: Matrix representation of ∂_2. The $+1$ entries are represented in purple, while the 0 entries are filled with yellow.

Remark 5. We point out that if we represent the ranks of the magnitude homology groups of a graph in a (k,ℓ)-table as in Table 1, we will always have that the table is lower triangular. In other words, $MH_{k,\ell}(G) \neq 0$ implies that $k \leq \ell$. This is because if $MH_{k,\ell}(G) \neq 0$ then $MC_{k,\ell}(G) \neq 0$, and so there is at least a tuple $(x_0, \dots, x_k)$ satisfying $\ell(x_0, \dots, x_k) = d(x_0, x_1) + \cdots + d(x_{k-1}, x_k) = \ell$. Now, since consecutive vertices are distinct by construction, $d(x_i, x_{i+1})$ is at least 1 for every $0 \leq i \leq k-1$, which means k can be at most ℓ.



	$\ell \backslash k$	0	1	2	3	4	5
	0	5					
	1		10				
	2			10			
l	3			10	10		
	4				30	10	
	5					50	10

Table 1: Ranks of $MH_{k,\ell}(C_5)$ computed using the Python code available at [8].

It is proven in [3, Proposition 9] that given a graph G, if we indicate by V the set of vertices and by $2E$ the set of oriented edges, then $MH_{0,0}(G)$ is the free abelian group on V and $MH_{1,1}(G)$ is the free abelian group on $2E$.

This provides us with an interpretation of the first two groups on the diagonal, but leaves the question about the meaning of other magnitude homology groups open.

3 Normalized and blurry magnitude chains

A major problem in producing an interpretation for the magnitude homology groups $MH_{k,\ell}(G)$ for any choice of k and ℓ comes from the fact that the definition of $MC_{k,\ell}(G)$ only asks for *consecutive* vertices to be different. That is, if x_0 and x_1 are two adjacent vertices in G an acceptable tuple in $MC_{5,4}(G)$ is $(x_0, x_1, x_0, x_1, x_0)$.

Tuples of this kind inducing a path that just revisits again and again the same edge (an more in general, tuples inducing non-eulerian trails) do not provide any insight about the meaning of magnitude homology. With this motivation, we revise the definition of magnitude chain considering the *normalized* subgroup of $MC_{k,l}(G)$ where a vertex is never required to be revisited.

Definition 6. (Normalized magnitude chain) Let G be a graph. We define the normalized (k, ℓ)-magnitude chain $NMC_{k,\ell}(G)$ to be the free abelian group generated by tuples $(x_0, \ldots, x_k)$ of vertices of G such that $(x_0, \ldots, x_k)$ is a k-path of length ℓ (i.e. a trail with $x_i \neq x_j$ for every $0 \leq i, j \leq k$).

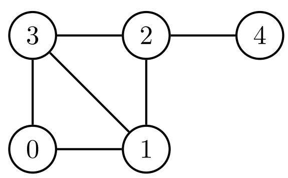
Taking as differential the one induced by $MC_{*,\ell}(G)$ we can construct the *normalized magnitude chain complex* $NMC_{*,\ell}(G)$

$$\cdots \to NMC_{k+1,\ell}(G) \xrightarrow{\partial_{k+1}} NMC_{k,\ell}(G) \xrightarrow{\partial_k} NMC_{k-1,\ell}(G) \to \cdots$$

and subsequently define the *normalized* (k, ℓ)*-magnitude homology group*

$$NMH_{k,\ell}(G) = H_k(NMC_{*,\ell}(G)) = \frac{\ker(\partial_k)}{\mathrm{imm}(\partial_{k+1})}.$$

Example 7. Consider the following graph G



We want to compute $NMH_{2,2}(G)$. The plain magnitude chain $NMC_{2,2}(G)$ is generated by (0, 1, 2), (0, 1, 3), (0, 3, 1), (0, 3, 2), (1, 0, 3), (1, 2, 3), (1, 2, 4), (1, 3, 0), (1, 3, 2), (2, 1, 0), (2, 1, 3), (2, 3, 0), (2, 3, 1), (3, 0, 1), (3, 1, 0), (3, 1, 2), (3, 2, 1), (3, 2, 4), (4, 2, 1), (4, 2, 3), while $PMC_{1,2}(G)$ is generated by (0, 2), (1, 4), (2, 0), (3, 4), (4, 1), (4, 3), and the matrix representing our differential is the one displayed in Figure 2.

We see that there are twelve all-zero columns, counting the triangles (0, 1, 3, 0) and (1, 2, 3, 1). We also have four non-columns, so modulo orientation they are just two and therefore they identify one 4-cycle. In this example they represent the square (0, 1, 2, 3, 0).

Remark 8. We point out that all definitions and properties regarding magnitude homology proved in [3] and [7] continue to be valid for normalized magnitude homology. In particular, with this new definition, $NMH_{0,0}(G)$ and $NMH_{1,1}(G)$ are still counting the number of vertices and edges in a graph respectively, since the generators of the groups $MC_{0,0}(G)$ and $MC_{1,1}(G)$ already satisfy the condition of not revisiting vertices.

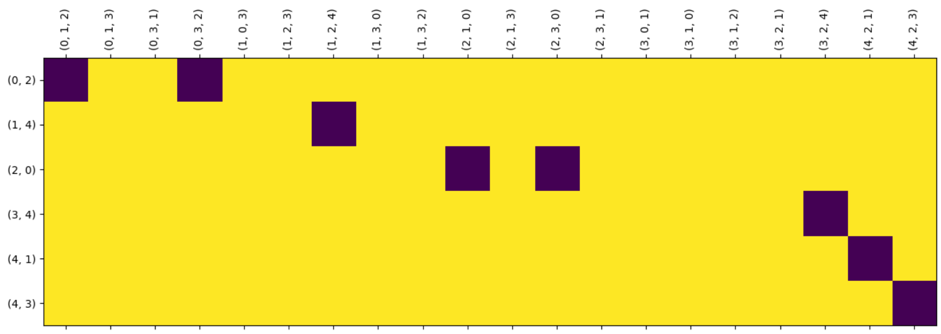


Figure 2: Matrix representation of $\partial_2 : NMC_{2,2}(G) \to NMC_{1,2}(G)$. The $+1$ entries are represented in purple, while the 0 entries are filled with yellow.

Now, in order to account for the elements in $MC_{k,\ell}(G)$ containing the repetition of at least a vertex we define the *blurry* magnitude chain as the quotient between the standard magnitude chain and the normalized one.

Definition 9. (Blurry magnitude chain) Let G be a graph. We define the blurry (k,ℓ)-magnitude chain $BMC_{k,\ell}(G)$ as

$$BMC_{k,\ell}(G) = \frac{MC_{k,\ell}(G)}{NMC_{k,\ell}(G)}.$$

Denoting by $[\cdot]_N$ the equivalence classes in $BMC_{k,\ell}(G)$, we define the differential map $\tilde{\partial}_k$ as

$$\tilde{\partial}_k([x_0,\dots,x_k]_N) = [\partial_k(x_0,\dots,x_k)]_N.$$

3.1 Long exact sequence

By definition of $NMC_{k,\ell}(G)$ and $BMC_{k,\ell}(G)$ we can build the short exact sequence

$$0 \to NMC_{k,\ell}(G) \xrightarrow{\iota} MC_{k,\ell}(G) \xrightarrow{\pi} BMC_{k,\ell}(G) \to 0.$$

Therefore we can construct the long exact sequence in homology

$$\cdots \to BMH_{k+1,\ell}(G) \xrightarrow{\delta_{k+1}} NMH_{k,\ell}(G) \xrightarrow{\iota_*} MH_{k,\ell}(G) \xrightarrow{\pi_*} BMH_{k,\ell}(G) \xrightarrow{\delta_k} \cdots.$$

Lemma 10. The map δ_k is the zero map for every value of k.

Proof. Using the Snake Lemma we obtain that

$$\begin{array}{rccc} \delta_{k+1}: & BMH_{k+1,\ell}(G) & \to & NMH_{k,\ell}(G) \\ & [[x_0,\dots,x_{k+1}]_N] & \mapsto & [\partial_{k+1}(x_0,\dots,x_{k+1})], \end{array}$$

where

$$[\partial_{k+1}(x_0,\dots,x_{k+1})] = \begin{cases} [y_0,\dots,y_k], \text{ if } y_i \neq y_j \text{ for all } i,j, \\ [0], \text{ otherwise.} \end{cases}$$

Now, since $[x_0,\dots,x_{k+1}]_N$ is a cycle and from Definition 1, we have $\partial_{k+1}(x_0,\dots,x_{k+1}) = \sum_{i=1}^{k} a_i(x_0,\dots\hat{x}_i,\dots,x_{k+1}) = 0$. So in particular $a_i = 0$ for the repeated vertex x_i, meaning there cannot exists a subtuple of $k+1$ all different vertices $(y_0,\dots,y_k)$ such that $\ell(y_0,\dots,y_k) = \ell$. □

It follows that $MH_{k,\ell}(G) \cong NMH_{k,\ell}(G) \oplus BMH_{k,\ell}(G)$, and from now on we carry on our analysis by considering the normalized and blurry magnitude groups separately.

3.2 Normalized magnitude homology

3.2.1 First diagonal

In this section we focus on the groups $NMH_{k,k}$, which we recall coincides with $\ker(\partial_k)$.

Let us start by considering the normalized homology group $NMH_{2,2} = \ker(\partial_2)$ of an undirected graph G, and consider the matrix Δ_2 associated with the differential. Recall that the columns of Δ_2 correspond to the generators of $NMC_{2,2}$, i.e. the 3-paths of length 2 in G with no repeated required vertices. We have the following simple lemma.

Lemma 11. Let Z be the number of null columns of Δ_2. The number of triangles occurring in G is $\frac{Z}{6}$.

Proof. Consider a representative 3-path $\overline{x} = (x_0,x_1,x_2)$ of length 2. Since $\overline{x}$ has length 2, $\{x_0,x_1\}$ and $\{x_1,x_2\}$ are edges of G. Also, $\overline{x}$ corresponds to a zero column of Δ_2 if and only if the shortest path between x_0 and x_2 has length smaller than 2, that is, if removing the required vertex x_1 we obtain a 2-path of length 1. This implies that there exists and edge $\{x_0,x_2\}$ and thus a triangle with vertices x_0,x_1,x_2.

It is immediate to see that the above holds for all permutations of $\overline{x}$, which implies that the number of triangles occurring in G is given by the number of null columns of Δ_2 divided by the cardinality of the automorphisms group of the triangle D_3. □

Let us now focus on the linearly dependent nonzero columns of Δ_2. We start by observing the following.

Remark 12. The linearly dependent nonzero columns of Δ_2 are all and only the columns that are repeated at least twice.

Proof. By construction of Δ_2, an entry of the matrix is equal to 1 if and only if $\ell(x_0,x_1,x_2) = \ell(x_0,x_2) = 2$. Thus in the column corresponding to (x_0,x_1,x_2)

the only entry that can possibly be 1 is the one at the row corresponding to (x_0, x_2). Therefore, two nonzero columns can only be linearly dependent if they are equal, and this can happen only if they correspond to two triplets $(x_0, x_1, x_2), (x_0, x_1', x_2)$ that share the same first and last vertex. This implies that the subset of linearly dependent columns of Δ_2 coincides with the columns that are repeated at least twice. □

The following lemma provides a first (rough) upper bound on the number of chordless squares in G.

Lemma 13. Let N be the number of linearly dependent nonzero columns of Δ_2. The number of chordless squares in G is upper bounded by $\frac{(N-1)N}{2}$.

Proof. Notice that a necessary (yet not sufficient) condition for the presence of a chordless square of vertices w, x, y, z in G is that there exist two generators of $NMC_{2,2}(G)$, (x, y, z) and (x, w, z), such that there is no edge linking x and z. In particular, the absence of the edge $\{x, z\}$ implies that the length of (x, y, z) and (x, w, z) does not decrease removing the respective intermediate required vertices, and thus the corresponding columns of Δ_2 both have a 1 at the row corresponding to (x, z). Thus two equal nonzero columns imply the presence of a square in G which is not a clique (yet there could be the edge $\{w, y\}$ that makes the square not chordless).

Suppose now there is a third generator with the same endpoints (x, v, z). Since $\{x, z\} \notin E$, the column of Δ_2 corresponding to (x, v, z) has a 1 at the row corresponding to (x, z) too. It is easy to see that these three generators imply the presence of 3 squares that are not cliques: one having vertices x, w, z, y, another with vertices x, w, v, y and a last one with vertices v, y, z, w. Three linearly dependent columns thus imply $\frac{(3-1)3}{2} = 3$ squares occurring in G.

Let us now generalize this formula to M equal columns and prove by induction that they imply the presence of $\frac{(M-1)M}{2}$ squares that are not cliques. We already verified that the formula holds for the base case $M = 2$ and for $M = 3$. Let us assume it holds for $M - 1$ equal columns, which thus identify $\frac{(M-2)(M-1)}{2}$ non-cliques squares, and suppose there is an M-th equal column. The 3-path corresponding to such column gives rise to a new distinct square with each one of the other $M - 1$ 3-paths with the same endpoints, thus giving a total of $\frac{(M-2)(M-1)}{2} + (M - 1) = \frac{(M-1)M}{2}$ non-cliques squares.

It is immediate to see that if the generators corresponding to all the N linearly dependent nonzero columns do not have all the same endpoints, the number of non-cliques squares in G is strictly less than $\frac{(N-1)N}{2}$; and since the chordless squares are included in the set of non-cliques squares, this quantity trivially bounds their occurrences in G. □

Notice that the bound provided by Lemma 13 is rather coarse. In fact, it actually bounds the total number of 4-cycles that are not 4-cliques in the graph, which also include the 4-cycles with one crossing edge. A closer look at the linearly dependent columns allows us to provide a stricter bound, given in

Lemma 14, which again bounds the total number of 4-cycles that are no cliques in the graph.

In order to compute the exact number of *chordless* squares, not only do we need to consider the linearly dependent columns themselves but also the generators they correspond to: we give this result in Lemma 15.

Lemma 14. Let D be the collection of linearly dependent nonzero columns of Δ_2, let $c_1, \dots, c_u$ be the largest subset of columns in D s.t. $c_i \neq c_j \; \forall i, j \in [1, u]$ and let n_i be the number of times a column c_i occurs in D. The number of chordless squares in G is upper bounded by

$$\sum_{i=1}^{u} \frac{(n_i - 1)n_i}{2}.$$

Proof. From the proof of Lemma 13 it follows that the number of non-clique squares identified by the generators corresponding to the columns equal to c_i is $\frac{(n_i-1)n_i}{2}$, for every $1 \leq i \leq u$. Since the total number of chordless squares is bounded by the number of these squares, and since

$$\sum_{i=1}^{u} \frac{(n_i - 1)n_i}{2} \leq \frac{(\sum_{i=1}^{u}(n_i) - 1)\sum_{i=1}^{u} n_i}{2} = \frac{(N-1)N}{2},$$

the statement follows. □

Lemma 15. In the same hypotheses of Lemma 14, let (x, w, z) and (x, y, z) be two generators of $NMC_{2,2}(G)$ corresponding to nonzero columns equal to c_i for some i. If there exists $1 \leq j \leq u$ such that $(w, x, y), (w, z, y)$ are generators whose columns are equal to c_j then there is a chordless square of vertices w, x, y, z.

Proof. From Lemma 13, the fact that $(x, x_1, y), (x, x_2, y) \in c_i$ means the presence of the square (x, x_1, y, x_2) where the diagonal (x, y) is missing. Similarly, from $(x_1, x, x_2), (x_1, y, x_2) \in c_j$ we deduce that the square (x, x_1, y, x_2) is missing the diagonal (x_1, x_2). Putting these facts together we obtain the thesis. □

Let us now consider the other magnitude homology groups $NMH_{k,k}(G)$ for $k \geq 3$ and the corresponding matrices Δ_k. The null columns of Δ_k correspond to k-paths $(x_0, \dots, x_k)$ s.t. if any required vertex x_i is removed, then $\ell(x_0, \dots, \hat{x_i}, \dots, x_k) < k-1$. In particular, a null column corresponding to a generator $(x_0, \dots, x_k)$ implies that there is an edge $\{x_{i-1}, x_{i+1}\}$ for all $1 \leq i \leq k-1$.

Intuitively, the presence of such generators is related to the presence of k-cliques in G. The following lemma formalizes this intuition and provides an upper bound on the number of k-cliques in G.

Lemma 16. Let Z be the number of null columns of Δ_k. The number of k-cliques in G is upper bounded by

$$\left\lfloor \frac{Z}{k!} \right\rfloor.$$

Proof. We claim that a set of k vertices $x_1, \ldots, x_k$ is a k-clique if and only if all their permutations are generators of $NMC_{k,k}$ and they correspond to null columns of Δ_k. Let us prove the two implications.

(⇒) Since $x_0, \ldots, x_k$ constitute a clique, there is an edge $\{x_i, x_j\}$ for all $i, j \in [0, k]$, implying that any k-path with $x_0, \ldots, x_k$ as required vertices (in any order) is of length k, and thus it is a generator of $NMC_{k,k}$. Again because there is an edge between any pair of such vertices, removing any required vertex from one of these generators leads to a path of length $k-1$, thus they all correspond to null columns of Δ_k.

(⇐)

□

3.2.2 Second diagonal

We are concerned in this section with the analysis of the information contained in the groups $NMH_{k-1,k}$.

While we were not able to provide a general interpretation for all magnitude homology groups $NMH_{k-1,k}$ on the second diagonal, we were able to understand part of the information contained in $NMH_{2,3}$: specifically, in the null columns of Δ_2. We believe the third normalized magnitude homology group on the second diagonal provides us with information regarding the number of chordless 4-cycles and 5-cycles in the graph.

Consider the following chain

$$\begin{array}{ccccccccc} \ldots & \to & NMC_{3,3} & \to & NMC_{2,3} & \to & NMC_{1,3} & \to & 0. \\ & & \Cup & & \Cup & & \Cup & & \\ & & (x_0, x_1, x_2, x_3) & & (x_0, \hat{x}, x_2, x_3) & & (x_0, \hat{x}, \hat{y}, x_3). & & \end{array}$$

We focus on the null columns of the matrix Δ_2, that is on the basis elements $(x_0, \hat{x}, x_2, x_3)$ such that $\Delta_2(x_0, \hat{x}, x_2, x_3) = 0$. Notice that if $(x_0, \hat{x}, x_2, x_3) \in \ker(\Delta_2)$ then one of the following is true: either $\ell(x_0, \hat{x}, \hat{x_2}, x_3) = 1$ or $\ell(x_0, \hat{x}, \hat{x_2}, x_3) = 2$.

Remark 17. Call L_1 and L_2 the subsets of the basis of $NMC_{2,3}$ such that $\ell(x_0, \hat{x_1}, \hat{x_2}, x_3) = 1$ and $\ell(x_0, \hat{x_1}, \hat{x_2}, x_3) = 2$, respectively. Then L_1 and L_2 for a partition of the null columns of $\ker(\Delta_2)$ and, consequently, $NMH_{2,3}$.

Lemma 18. The number of basis elements $(x_0, \hat{x_1}, x_2, x_3)$ of $NMC_{2,3}$ such that the induced path (x_0, x_2, x_3) is eulerian, (x_0, x_2, x_3) is the image of Δ_3 and $(x_0, \hat{x_1}, x_2, x_3) \in L_1$ is exactly the number of chordless 4-cycles in the graph.

Proof. First notice that assuming (x_0, x_2, x_3) eulerian while $\ell(x_0, \hat{x_1}, x_2, x_3) = 3$ implies the absence of the edge (x_0, x_2). Further, it is clear the $\ell(x_0, \hat{x_1}, \hat{x_2}, x_3) = \ell(x_0, x_3) = 1$ implies the existence of the edge (x_0, x_3), and thus (x_0, x_1, x_2, x_3) is an induced 4-cycle.

Now, the image of Δ_3 contains the elements $(x_0, \hat{x_1}, x_2, x_3) \in NMC_{2,3}$ such that $\ell(x_0, \hat{x_1}, x_2, x_3) = \ell(x_0, x_1, x_2, x_3)$. In other words, paths that do not contain the triangle (x_1, x_2, x_3), and in particular the edge (x_1, x_3). □

Lemma 19. The number of basis elements $(x_0, \hat{x_1}, x_2, x_3)$ of $NMC_{2,3}$ such that the induced path (x_0, x_2, x_3) is eulerian, (x_0, x_2, x_3) is the image of Δ_3 and $(x_0, \hat{x_1}, x_2, x_3) \in L_2$ bounds from above the number of chordless 5-cycles in the graph.

Proof. Proceeding similarly as in the proof of Lemma 18, we point out that the hypotheses of $()x_0, x_2, x_3)$ being eulerian and $\ell(x_0, \hat{x_1}, x_2, x_3) = 3$ mean the absence of the edge (x_0, x_2). Now, $\ell(x_0, \hat{x_1}, \hat{x_2}, x_3) = \ell(x_0, x_3) = 2$ tells us that either the edge (x_1, x_3) exists, or there is a different path of length 2 from x_0 to x_3 involving a new vertex x_4.

In addition, being $(x_0, \hat{x_1}, x_2, x_3)$ in the image of Δ_3 we know that the path does not contain the triangle (x_1, x_2, x_3), and in particular the edge (x_1, x_3), so we are able to exclude the first case.

In conclusion, we are left with a 5-cycle $(x_0, x_1, x_2, x_3, x_4)$ which is surely missing diagonals (x_0, x_2), (x_0, x_3) and (x_1, x_3). □

3.3 Blurry magnitude homology

3.3.1 First diagonal

To do 1. Consider a filtration given by the number of repeated vertices (something like "$BMC^i_{k,k}(G)$ repeats at most i times each vertex"), build a spectral sequence and check how ∂_k behaves.

4 $MH_{k,k}(G)$ of Erdos-Renyi graphs

Let $G = G = (n, p) = G(n, n^{-\alpha})$ be an Erdos-Renyi graph. In order to produce an expected value for the ranks of magnitude homology groups on the diagonal we proceed as follows:

- work separately with $NMH_{k,k}(G)$ and $BMH_{k,k}(G)$
- identify what kind of subgraph H is induced by a cycle in $NMH_{k,k}(G)$ and $BMH_{k,k}(G)$ respectively
- give an estimate for the occurrences of H in G.

4.1 $NMH_{k,k}(G)$

Take $[x_0, \ldots, x_k] \in NMH_{k,k}(G)$. Then for every $i = 1, \ldots, k-1$ the edge (x_{i-1}, x_{i+1}) exists and the induced subgraph H is as shown in Figure 3.

Remark 20. Notice that, by construction of $(x_0, \ldots, x_k) \in NMH_{k,k}$ each edge (x_i, x_{i+1}) will be different and so the induced path $(x_0, \ldots, x_k)$ will be eulerian.

Figure 3: Subgraph corresponding to a normalized magnitude cycle. The edges (x_{i-1}, x_{i+1}) are shown in blue.

Now, the number of edges contained in such graph is $k+(k-1)$ (black edges plus blue edges). Hence, calling a_H the number of automorphisms of H, the number of copies of H expected in G is

$$\begin{aligned} N_H &= \binom{n}{k+1} a_H p^{2k-1} \\ &\sim n^{k+1} \frac{a_H}{(k+1)!} n^{\alpha(1-2k)} \\ &= \frac{a_H}{(k+1)!} n^{\alpha(1-2k)+(k+1)} \xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } \alpha > \frac{k+1}{2k-1} \\ \infty, \text{ if } 0 < \alpha < \frac{k+1}{2k-1} \end{cases} \end{aligned}$$

The computation above implies the following.

Lemma 21. Let G be an Erdos-Renyi graph on n vertices and set $p = n^{-\alpha}$. The normalized magnitude homology $NMH_{k,k}(G)$ vanishes for $\alpha > \frac{k+1}{2k-1}$.

4.2 $BMH_{k,k}(G)$

Take $[[x_0, \ldots, x_k]_N] \in BMH_{k,k}(G)$ and suppose there are at most $r \leq \lceil \frac{k}{2} \rceil$ copies of each vertex. Further, suppose the induced path $(x_0, \ldots, x_k)$ is eulerian. Then, since we are not revisiting any edges, we are a situation similar to normalized groups case, in which we are looking for a subgraph with $k+1$ nodes and $k+(k-1)$ edges, as shown in Figure 4.

Then, calling H the subgraph in Figure 4 and denoting by a_H the cardinality of its automorphisms group, the number of copies of this subgraph will be

$$N_H = \binom{n}{k+1} a_H p^{2k-1} \xrightarrow{n\to\infty} \begin{cases} 0, \text{ if } \alpha > \frac{k+1}{2k-1} \\ \infty, \text{ if } 0 < \alpha < \frac{k+1}{2k-1} \end{cases}$$

Therefore we have the following.

Lemma 22. Let G be an Erdos-Renyi graph on n vertices and set $p = n^{-\alpha}$. Suppose all cycles $[[x_0, \ldots, x_k]_N] \in BMH_{k,k}(G)$ induce an eulerian path in G. Then the blurry magnitude homology $BMH_{k,k}(G)$ vanishes for $\alpha > \frac{k+1}{2k-1}$.

To do 2. Non-eulerian case.

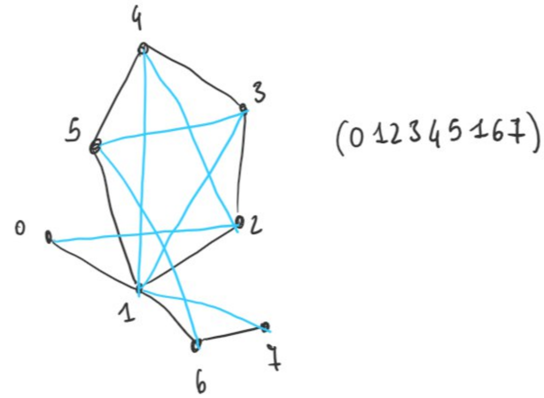


Figure 4: Subgraph corresponding to an eulerian blurry magnitude cycle. The edges (x_{i-1}, x_{i+1}) are shown in blue.

5 Geometric realization of $NMC_{k,\ell}$

In the spirit of Matthew Khale's papers [4] and [5], we look for (non)vanishing results by exploring the topology of the normalized magnitude complex.

To do this we take advantage of the Asao-Izumihara simplicial complex introduced in [1].

5.1 Asao-Izumihara simplicial complex

We recall in this section the construction of the Asao-Izumihara complex.

Let G be a graph. From [1, Proposition 2.9] we have the following.

Corollary 23. For $\ell \geq 0$ we have a direct sum decomposition of a normalized magnitude chain complex

$$NMC_{*,\ell}(G) = \bigoplus_{a,b \in V(G)} NMC_{*,\ell}(a,b),$$

where $NMC_{*,\ell}(a,b)$ is the subcomplex of $NMC_{*,\ell}(G)$ generated by sequences starting at a and ending at b.

Therefore the computation of NMH reduces to the computation of each (a,b)-component. Recall the [1, Definition 2.11].

Definition 24. Let $\bar{x} = (x_0, \ldots, x_k)$ and $\bar{y} = (y_0, \ldots, y_{k'})$ be two tuples. We call $\bar{y}$ a subtuple of $\bar{x}$ if there exists integers $0 = i_0 < \cdots < i_{k'} = k$ such that $x_{ij} = y_j$ for each $0 \leq j \leq k'$. When $\bar{y}$ is a subtuple of $\bar{x}$ we write $\bar{y} \prec \bar{x}$.

Now fix an integer $\ell \geq 3$ and denote the set of paths for a to b of length at most ℓ with $P_{\leq \ell}(a,b)$.

Definition 25. We call a *sequence* a path such that $d(x_i, x_{i+1}) = 1$.

Definition 26. Let $K_\ell(a, b)$ be the set whose elements are subsets

$$\{(x_{i_1}, i_1), \ldots, (x_{i_k}, i_k)\} \subset V(G) \times \{1, 2, \ldots, \ell - 1\}$$

such that there exists a sequence $(a, x_1, \cdots, x_{\ell-1}, b) \in P_{\leq \ell}(a, b)$ with $(a, x_{i_1}, \ldots, x_{i_k}, b) \prec (a, x_1, \cdots, x_{\ell-1}, b)$.

Now, the set $K_\ell(a, b)$ is a simplicial complex and clearly $K_{\ell-1}(a, b)$ is a subcomplex. We give the definition of another subcomplex $K'_\ell(a, b) \subset K_\ell(a, b)$ by

$$K'_\ell(a, b) := \{(x_{i_1}, \ldots, x_{i_k}) \in K_\ell(a, b) | len(a, x_{i_1}, \ldots, x_{i_k}, b) \leq \ell - 1\}.$$

With the above definitions it is possible to construct an isomorphism between $C_*(K_\ell(a, b), K'_\ell(a, b))$ and $NMC_{*+2,\ell}(a, b)$, as shown in [1, Theorem 4.3].

5.2 Geometry of $NMC_{k,\ell}(a, b)$

To do 3. Formalize the following: $\Delta^n \subseteq C_n(K_\ell(a, b), K'_\ell(a, b))$, i.e. $\Delta^n \subseteq ||NMC_{n+2,\ell}(a, b)||$ indicates the presence of an n+1-path from a to b, $(a, x_1, \cdots, x_n, b)$, such that no edge (x_i, x_{i+1}) exists. Therefore, if we allow endpoints to be the same, $\Delta^n \subseteq ||NMC_{n+2,\ell}(a, a)||$ indicates the presence of an $n + 1$ chordless cycle C_{n+1}.

6 Relation with clustering coefficients

In Graph Theory, a clustering coefficient is a structural feature that measures the degree to which nodes in a graph tend to cluster together. In other words, it tells how connected a vertex's neighbors are to one another. There are two existing versions of this measure. The *global*, which was designed by Wasserman and Faust in [9] to give an overall indication of the clustering in the network, and the *local*, first defined by Watts and Strogatz in [10] to give an indication about the tendency to cluster near a specific node.

In this section we provide a way to compute both clustering coefficients of a graph $G = (V, E)$ via $NMH_{2,2}(G)$, determining thus a close relation between these tools.

6.1 Local clustering coefficient

The local clustering coefficient C_i of a node x_i describes the likelihood that the neighbors of x_i are also connected. To compute C_i we consider the neighborhood N_i of x_i, where $N_i = \{x_j : (x_i, x_j) = e_{ij} \in E\}$ and compute the fraction of the number of links between the vertices within N_i divided by the number of links that could possibly exist between them. That is, we set

$$C_i = \frac{2\{e_{jk} : x_j, x_k \in N_i \text{ and } e_{jk} \in E\}}{d_i(d_i - 1)},$$

where $d_i = |N_i|$ is the degree of the vertex x_i.

In other words, we are dividing the number of triangles x_i is part of by the number of 2-paths of length 2 containing x_i.

Therefore, call $NMC^i_{2,2}(G)$ the subgroup of $NMC_{2,2}(G)$ such that x_i is the middle vertex of any 2-path, so $PMC^i_{2,2}(G) = \{(x, x_i, y) : \ell(x, x_i, y) = 2\}$. Then the number of triangles containing x_i is precisely the number of all-zero columns of $\ker(\partial_2(NMC^i_{2,2}(G)))$. Calling this number Z_i, we can write the local clustering coefficient as

$$C_i = \frac{2Z_i}{d_i(d_i - 1)}.$$

Remark 27. Given the connection just established between the local clustering coefficient and plain magnitude homology, one could think of using $NMH_{2,2}$ in a network analysis context as a *centrality measure*: if for a given vertex v_i the number Z_i defined above takes low values it means there are few connections between neighbors of x_i, meaning x_i has a lot of power over information flow.

6.2 Global clustering coefficient

The global clustering coefficient C is based on 3-tuples, i.e. on elements of $NMC_{2,2}$, and is computed as the number of closed 3-tuples (or $3\times$ triangles) over the total number of 3-tuples (both open and closed). That is, calling Z the number of all-zero columns in $\ker(\partial_2(NMC_{2,2}(G)))$

$$C = \frac{Z}{|NMC_{2,2}(G)|}.$$

7 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 and see how they influence the information flow.

References

[1] Asao, Y., Izumihara, K.: Geometric approach to graph magnitude homology. arXiv preprint arXiv:2003.08058 (2020)

[2] Hatcher, A.: Algebraic topology (2005)

[3] Hepworth, R., Willerton, S.: Categorifying the magnitude of a graph. arXiv preprint arXiv:1505.04125 (2015)

[4] Kahle, M.: The neighborhood complex of a random graph. Journal of Combinatorial Theory, Series A **114**(2), 380–387 (2007)

[5] Kahle, M.: Topology of random clique complexes. Discrete mathematics **309**(6), 1658–1671 (2009)

[6] Leinster, T.: The magnitude of a graph. In: Mathematical Proceedings of the Cambridge Philosophical Society. vol. 166, pp. 247–264. Cambridge University Press (2019)

[7] Leinster, T., Shulman, M.: Magnitude homology of enriched categories and metric spaces. Algebraic & Geometric Topology **21**(5), 2175–2221 (2021)

[8] Menara, G.: magnitude-homology. `https://github.com/GMenara/magnitude-homology` (2022)

[9] Wasserman, S., Faust, K., et al.: Social network analysis: Methods and applications (1994)

[10] Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world'networks. nature **393**(6684), 440–442 (1998)