On the Relations between Graph Magnitude Homology and Subgraph Counting*

February 6, 2023

Abstract

The abstract should briefly summarize the contents of the paper in 150–250 words.

1 Introduction (DA SISTEMARE)

Graphs are used in a variety of sciences to model and to analyze complex relationships. In this framework, the search for interesting and relevant substructures is a standard procedure, and the detection of cliques and clique-like subgraphs is a fundamental tool in graph analysis. Such substructures have been applied in many different situations including community detection in social networks [?], [?], identification of real-time stories in the news [?] and graph visualization [?], [?]. In practice, due to noise in data, one is also interested in large "near-cliques". While this is not a standard term, applications involve cliques that are missing a small sparse subgraph. For example, incomplete cliques have been used to predict missing pairwise interactions [?] and for identifying functional groups [?] in a protein interaction network. Also, they were exploited for community detection [?] and for detecting test collusion [?]. Recent works have used the fraction of near-cliques to k-cliques to define higher order variants of clustering coefficients [?].

In the present work, in order to quantitatively characterize these structures, we employ magnitude homology, a tool that comes from the field of algebraic topology. Magnitude is an isometric invariant of metric spaces, so-named for its web of connections to "size-like" quantities of significance in various corners of mathematics. Defined and first studied by Leinster [?], it is a special case of a general theory of magnitude of an enriched category, and has found applications in areas like biodiversity (e.g., Leinster and Cobbold [?]). As a finite graph naturally gives rise to a finite metric space, it is possible to associate magnitude with it. Magnitude homology has been invented by Hepworth and Willerton [?] as an enrichment of the magnitude of a graph which is equipped with a graph

*Titolo provvisorio! Altre proposte sono benvenute.

metric. The magnitude homology of graphs has been well studied in recent years and has proven to be a rich invariant [?, ?, ?, ?, ?]. Also, several theoretical tools for computing the magnitude homology of a graph have been studied so far. For example, Hepworth and Willerton proved in [?] a Mayer-Vietoris type exact sequence and a Kunneth type formula, and Gu [?] uses algebraic Morse theory for computation for some graphs. Although, in general, computation of magnitude homology remains a difficult problem.

TBC

In this work,...

The paper is organized as follows,...

2 Background

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 a sequence of vertices $x_1, x_2, \ldots, x_n$ such that there is an edge $\{x_i, x_{i+1}\}$ for all $0 \leq i < n$; a *path* is a trail with no repeated vertices; a *cycle* is a closed path. For the purposes of defining magnitude homology, we assume all graphs to have no self-loops and no multi-edges [?]. We define the distance $d : V \times V \to [0, \infty]$ between two vertices $u, v \in V$ as the length of the shortest path from u to v, with $d(u, v) = \infty$ if u and v lie in different connected components.

A k-trail in a graph G is a k-tuple $\overline{x}^k = (x_1, \ldots, x_k)$ of *required vertices* with $x_i \neq x_{i+1}$ and $d(x_i, x_{i+1}) < \infty$ for every $1 \leq i \leq k-1$. The length of a k-trail $\overline{x}^k$ is defined as the minimum length of a trail that visits $x_1, \ldots, x_k$ in this order (i.e., whose visited vertices form a supersequence of the required vertices): namely, $\mathsf{len}(\overline{x}^k) = d(x_1, x_2) + \cdots + d(x_{k-1}, x_k)$.

A triangle in G is a cycle on 3 distinct vertices; a *chordless* or *induced* square is a cycle connecting four distinct vertices x_1, x_2, x_3, x_4 such that there are no edges between them except the edges of the cycle: in other words, $\{x_1, x_3\} \notin E$ and $\{x_2, x_4\} \notin E$. A k-clique is a subset of k distinct vertices s.t. every two distinct vertices n the clique are connected by an edge.

A *free abelian group* is an abelian group (a set of elements with a sum operation) that has a basis, that is, a finite subset of its elements s.t. all the other elements can be expressed as an integer combination of the elements of the basis. We will denote by $\langle B \rangle$ the free abelian group *generated* by the basis B. The *rank* of a group is the number of elements of a basis.

The (k, ℓ)-*magnitude group* of a graph, denoted by M_k^ℓ, is the free abelian group generated by all the k-trails of length ℓ in the graph[1]. Assuming that the vertices of the graph are in a one-to-one correspondence with the integers in $[1, |V|]$, we can lexicographically order the k-tuples: we denote by $M_k^\ell[j]$ the j-th lexicographically smallest element of M_k^ℓ. We denote by $\overline{x}^k_{\hat{i}}$ the $(k-1)$-tuple $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k)$ obtained removing the i-th required vertex from $\overline{x}^k$.

[1]Formally, M_k^ℓ is called the (k, ℓ)-magnitude chain [?]. See definition 4 in Appendix A.

Definition 1. *Consider* M_k^ℓ *and* M_{k-1}^ℓ. *The* differential matrix $\Delta_k^\ell = \{\delta_{r,c}\}$ *is the matrix whose* c*-th column corresponds to* $M_k^\ell[c]$, *whose* r*-th row corresponds to* $M_{k-1}^\ell[r]$, *and with*

$$\delta_{r,c} = \begin{cases} (-1)^i & \text{if } M_k^\ell[c] = \overline{x}^k \wedge M_{k-1}^\ell[r] = \overline{x}_{\hat{i}}^k, \\ 0 & \text{otherwise.} \end{cases}$$

Note that each nonzero entry of Δ_k^ℓ corresponds to some k-trail of length ℓ whose length does not decrease when some required vertex x_i is removed[2], implying that there is no shortcut that connects x_{i-1} and x_{i+1} without using x_i. This provides some information on the structure of G, in particular on the occurrences of specific subgraphs, as we show in Section 3.

Clearly, we can define a sequence of differential matrices $\Delta_k^\ell, \Delta_{k-1}^\ell, \Delta_{k-2}^\ell \dots$ that relate the elements of the magnitude groups for a fixed ℓ and decreasing k. This sequence has an analogous algebraic definition, given in Appendix A, that motivates the following definition.

Definition 2. *The* (k, ℓ)*-magnitude homology group of a graph* G *is the group*

@GiuliaM, con questa definizione però il ker è generato solo dalle colonne nulle, è giusto?

No così dovremmo beccare anche le combinazioni, perché M_k^ℓ *è un gruppo libero e quindi contiene anche le somme formali. Quindi (prendendo l'esempio sotto) abbiamo che* $(0,1,2) - (0,3,2) \in M_k^\ell$ *e* $\mathsf{len}((0,\hat{1},2) - (0,\hat{3},2)) = 0$, *per cui* $(0,1,2) - (0,3,2)$ *sta nel ker.*

Ma $\overline{x}^k$ *è definita come tupla che determina un* k*-trail, non come generico elemento del gruppo...e la lunghezza di una somma formale non è definita (non avrebbe neanche senso, non corrispondendo a un percorso). secondo me quindi dobbiamo formularlo diversamente, ma possiamo pensarci dopo.*

Forse possiamo definire il kernel direttamente come kernel della matrice, visto che il nostro operatore è proprio costruito così. Perché quello che si fa di solito è definire l'operatore differenziale e poi dire "siccome è lineare ha una matrice associata e calcoliamo rango e dimensione del nucleo della matrice associata". Qua non dovremmo neanche fare questo passaggio

concordo pienamente. Poi sotto mettiamo l'osservazione sulle colonne nulle e bon

$$H_k^\ell = \ker(\Delta_k^\ell) \setminus \operatorname{im}(\Delta_{k+1}^\ell), \textit{where } \ker(\Delta_k^\ell) = \langle \overline{x}^k \in M_k^\ell \mid \mathsf{len}(\overline{x}_{\hat{i}}^k) < \ell \ \forall\ i \rangle \textit{ and}$$

$$\operatorname{im}(\Delta_k^\ell) = \langle \overline{x}^{k-1} \in M_{k-1}^\ell \mid \exists i \in [1, k-1] \textit{and } \overline{x}^k \in M_k^\ell \textit{ s.t. } \overline{x}^{k-1} = \overline{x}_{\hat{i}}^k \rangle.$$

[2] Formally, Δ_k^ℓ is the matrix representation of a differential map between free abelian groups. See Definition 5 in Appendix A.

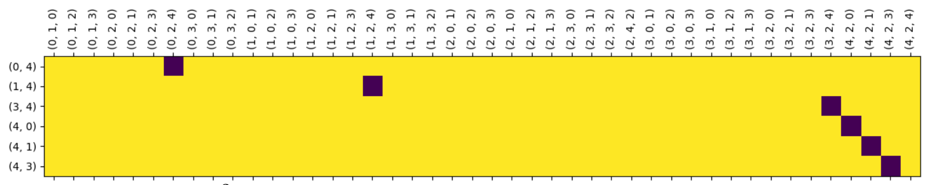
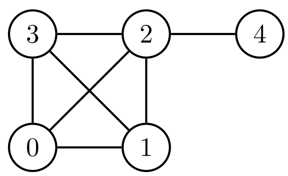


Figure 1: Matrix Δ_3^2. The $+1$ entries are in purple, while the 0 entries are in yellow.

In other words, the (k, ℓ)-magnitude homology group is generated by the k-trails of length ℓ that cannot be obtained by removing a required vertex from some $(k+1)$-trail of length ℓ, and whose length decreases if any of their required vertices are removed.

Example 1. *Consider the following graph G, for which we want to compute H_3^2.*



From Definition 2 we need to compute the difference set between the kernel of Δ_3^2 and the image of Δ_4^2, so we are concerned with the three magnitude sets M_2^2, M_3^2, M_4^2. By definition, M_2^2 consists of the 2-trails of length 2, that is, $M_2^2 = \{(0,4),(1,4),(3,4),(4,0),(4,1),(4,3)\}$. Similarly, M_3^2 consists of the 3-trails of length 2, i.e., it is the following set of 44 elements: $\{(h,i,j) \mid h,i,j \in [0,3], h \neq i, i \neq j\} \cup \{(i,2,4) \mid i \neq 2\} \cup \{(4,2,i) \mid i \neq 2\}$. Finally, M_4^2 consists of the 4-trails of length 2, and since any trail connecting 4 vertices has length at least 3, we have $M_4^2 = \emptyset$.

We see from Figure 1 that Δ_3^2 has no linear dependent columns, thus $\ker(\Delta_3^2)$ is generated by the elements of M_3^2 whose length diminishes when the middle vertex is removed (corresponding to the null coulmns):

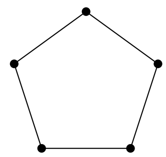
$$\ker(\Delta_3^2) = M_3^2 \setminus \{(0,2,4),(1,2,4),(3,2,4),(4,2,0),(4,2,1),(4,2,3)\}.$$

Since $M_4^2 = \emptyset$, we have that $\mathrm{im}(\Delta_4^2) = \emptyset$ and thus the 3-magnitude homology group of G coincides with the 38 elements in the kernel of Δ_3^2.

In the following we will focus on the matrix $R = \{r_{\ell,k}\}$ of the ranks of the homology groups H_k^ℓ, that is, $r_{\ell,k} = |H_k^\ell|$. We call R the *rank matrix* of G.

Questo paragrafo può essere migliorato.

Observation 1. *Note that R is such that $r_{k,\ell} = 0$ for all k, ℓ such that $\ell > k-1$. This is because if $H_k^\ell \neq \langle 0 \rangle$ then $M_k^\ell \neq \langle 0 \rangle$, and so there is at least a tuple $(x_1, \ldots, x_k)$ satisfying $\mathsf{len}(x_1, \ldots, x_k) = d(x_1, x_2) + \cdots + d(x_{k-1}, x_k) = \ell$. Now, since consecutive vertices are distinct by construction, $d(x_i, x_{i+1})$ is at least 1 for every $1 \leq i \leq k-1$, which means k can be at most $\ell + 1$.*



				k		
		1	2	3	4	5
	0	5				
	1		10			
	2			10		
l	3			10	10	
	4				30	10
	5					50

Table 1: Ranks of $H_k^\ell(C_5)$ computed using Python [?]

In what follows we refer to the set of the ranks of the H_k^{k-1} homology groups (that is, the top non-zero diagonal in the rank matrix) as *first diagonal*. Accordingly, the *n-th diagonal* will identify the ranks of the H_k^{k-n} homology groups.

It is proven in [?] (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 $H_1^0(G)$ is the free abelian group on V and $H_2^1(G)$ is the free abelian group on $2E$.

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

3 Interpretation of magnitude homology

In this section, we provide an interpretation for some of the magnitude homology groups corresponding to elements on the first and second diagonals of the rank matrix. To this aim, we introduce a slightly different notion of magnitude group, which avoids considering as generators the k-trails that carry redundant information on the structure of the graph.

First, notice that the k-tuples $(x_1, \ldots, x_k)$ and $(x_k, \ldots, x_1)$ represent exactly the same subset of vertices and edges of G, and thus one of the two is redundant. Without loss of generality, we define the lexicographically smallest of the two as the *representative* of the two possible orientations of the k-trail they represent, and we will disregard the other.

The trails with some repeated required vertices are redundant too. An example can be seen on the magnitude group M_3^2 for the graph of Example 1: the 3-trail $(0, 1, 0)$ consists in traversing the edge $(0, 1)$ twice, thus it does not provide any additional information about the structure of the graph with respect to the 2-trail $(0, 1)$.

We thus define the *normalized* magnitude group as the subgroup of M_k^ℓ generated by the representative k-trails of length ℓ with no repeated required vertices.

Definition 3. *The normalized (k, ℓ)-magnitude group $\widehat{M}_k^\ell$ of a graph G is the free abelian group generated by the representative k-trails $\overline{x}^k = (x_1, \ldots, x_k)$ of length ℓ and such that $x_i \neq x_j$ for every $i \neq j$.*

Observation 2. *If $\overline{x}^k$ is a generator of a normalized magnitude group, then $\overline{x}^k_{\hat{i}}$ is a representative $(k-1)$-trail for all $1 < i < k$.*

Proof. It follows immediately from the fact that all the vertices in a generator $\overline{x}^k = (x_1, \ldots, x_k)$ of a normalized magnitude group are distinct, and since by definition $\overline{x}^k$ is a representative k-trail it holds that $x_1 < x_k$, which remains true removing any intermediate required vertex. □ □

The differential matrix $\widehat{\Delta}^\ell_k$ and the homology group $\widehat{H}^\ell_k$ for the normalized magnitude groups are defined entirely analogously to the magnitude groups.

Example 2. *Consider the same graph G as in Example 1. By Definition 3, the normalized magnitude group $\widehat{M}^2_3$ is generated by the subset of generators of M^2_3 consisting of the following 15 elements (see also Figure 2):*

$$\{(h, i, j) \mid i, j \in [0, 3], h \in [0, j), h \neq i \neq j\} \cup \{(i, 2, 4) \mid i \in [0, 3]\}.$$

Remark 1. *We point out that all definitions and properties regarding magnitude homology proved in [**?**] and [**?**] are still valid for the normalized magnitude homology groups. This is because, by construction, $\widehat{M}^\ell_k$ is a subgroup of M^ℓ_k (as it is generated by a subset of its generators), and since the definition of the differential matrix is unchanged it follows that the normalized magnitude homology group $\widehat{H}^\ell_k$ is a subgroup of H^ℓ_k for every $k, \ell \geq 0$.*

In particular, with this new definition, $\widehat{H}^0_1$ and $\widehat{H}^1_2$ are still counting the number of vertices and edges in a graph respectively, because the generators of the groups M^0_1 and M^1_2 already satisfy the condition of not revisiting vertices.

3.1 First diagonal: counting triangles and squares

In this section, we focus on the groups $\widehat{H}^{k-1}_k$. As already noticed in Observation 1, the image of $\widehat{\Delta}^{k-1}_{k+1}$ is the zero group $\langle 0 \rangle$ for all values of k, and thus the normalized homology group $\widehat{H}^{k-1}_k$ coincides, by Definition 2, with $\ker(\widehat{\Delta}^{k-1}_k)$.

Let us start by considering the normalized homology group $\widehat{H}^2_3 = \ker(\widehat{\Delta}^2_3)$ of an undirected graph G, and consider the differential matrix $\widehat{\Delta}^2_3$. Recall that the columns of $\widehat{\Delta}^2_3$ correspond to the generators of $\widehat{H}^2_3$, i.e., the representative 3-trails of length 2 in G with no repeated required vertices. We have the following simple lemma.

Lemma 1. *Let Z be the number of null columns of $\widehat{\Delta}^2_3$. The number of triangles occurring in G is $\frac{Z}{3}$.*

Proof. Consider a representative 3-trail $\overline{x} = (x_1, x_2, x_3)$ of length 2 and suppose w.l.o.g. that $x_1 < x_2 < x_3$. Since $\overline{x}$ has length 2, $\{x_1, x_2\}$ and $\{x_2, x_3\}$ are edges of G. By Definition 1, $\overline{x}$ corresponds to a zero column of $\widehat{\Delta}^2_3$ if and only if the shortest path between x_1 and x_3 has length smaller than 2, that is, if removing the required vertex x_2 we obtain a 2-trail of length 1. This implies that there exists and edge $\{x_1, x_3\}$ and thus a triangle with vertices x_1, x_2, x_3.

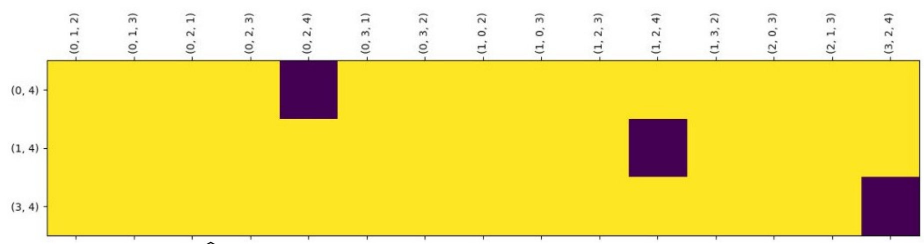


Figure 2: Matrix $\widehat{\Delta}_3^2$. The $+1$ entries are in purple, while the 0 entries are in yellow.

It is immediate to see that $\overline{x}' = (x_1, x_3, x_2)$ and $\overline{x}'' = (x_2, x_1, x_3)$ are the only other representative 3-trails whose vertices are a permutation of x_1, x_2, x_3, that $\overline{x}'$ and $\overline{x}''$ correspond to null columns and thus that they identify the same triangle in G as $\overline{x}$. $\square$

Example 3. *Consider the same graph G as in Example 1. We want to compute $\widehat{H}_3^2$. The generators of $\widehat{M}_3^2$ are shown in Example 2; further, $\widehat{M}_2^2$ is generated by the 2-trails $(0,4)$, $(1,4)$ and $(3,4)$ and the differential matrix $\widehat{\Delta}_3^2$ (shown in Figure 2) is the submatrix of the one displayed in Figure 1 whose columns and rows correspond to the generators of $\widehat{M}_3^2$ and $\widehat{M}_2^2$, respectively. We see that $\widehat{\Delta}_3^2$ has 12 null columns, corresponding to the four triangles of vertices $\{0,1,3\}$, $\{0,1,2\}$, $\{0,2,3\}$ and $\{1,2,3\}$, respectively.*

Let us now focus on the linearly dependent nonzero columns of $\widehat{\Delta}_3^2$. We start by observing the following.

Observation 3. *The linearly dependent nonzero columns of $\widehat{\Delta}_3^2$ are all and only the columns that are repeated at least twice.*

Proof. By Definition 1, an entry of $\widehat{\Delta}_3^2$ is equal to 1 if and only if $\mathsf{len}(x_1, x_2, x_3) = \mathsf{len}(x_1, x_3) = 2$. Thus in the column corresponding to (x_1, x_2, x_3) the only entry that can possibly be 1 is the one at the row corresponding to (x_1, x_3). Therefore, two nonzero columns can only be linearly dependent if they are equal, and this can happen only if they correspond to two triplets $(x_1, x_2, x_3), (x_1, x_2', x_3)$ that share the same first and last vertex. This implies that the subset of linearly dependent columns of $\widehat{\Delta}_3^2$ coincides with the columns that are repeated at least twice. $\square$ $\square$

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

Lemma 2. *Let N be the number of linearly dependent nonzero columns of $\widehat{\Delta}_3^2$. 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 $\widehat{M}_3^2$, (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 $\widehat{\Delta}_3^2$ 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 $\widehat{\Delta}_3^2$ 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-trail corresponding to such column gives rise to a new distinct square with each one of the other $M - 1$ 3-trails 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 2 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 (see Figure...........) A closer look at the linearly dependent columns allows us to provide a stricter bound, given in Lemma 3, which again bounds the total number of 4-cycles that are no cliques in the graph.

Possiamo mettere un piccolo esempio solo se abbiamo spazio alla fine: @GiuliaM, se ti va comunque potresti fare una figurina in cui ci sono un tot di quadrati inscatolati e quello più all'interno ha un lato orizzontale, poi se non c'è posto la piazziamo in appendice.

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 4.

Lemma 3. *Let D be the collection of linearly dependent nonzero columns of $\widehat{\Delta}_3^2$, let $c_1, \ldots, c_u$ 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 2 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 bouunded by the number of these squares, and since

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

the statement follows. □ □

Lemma 4. *In the same hypotheses of Lemma 3, let (x, w, z) and (x, y, z) two generators of $\widehat{M}_3^2$ 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.*

in caso può andare scritto così? perchè non sto riuscendo a trovare una formula per il numero preciso di chordless

Proof. From Lemma 2, 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 $\widehat{H}_k^{k-1}$ for $k > 3$ and the corresponding differential matrices $\widehat{\Delta}_k^{k-1}$. By Definition 1, the null columns $\widehat{\Delta}_k^{k-1}$ correspond to k-trails $\overline{x}^k$ s.t. for any required vertex x_i to be removed, $\mathsf{len}(\overline{x}_{\hat{i}}^k) < k-1$. In particular, a null column corresponding to a generator $\overline{x}^k$ implies that there is an edge $\{x_{i-1}, x_{i+1}\}$ $\forall 1 < i < k$.

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 5. *Let Z be the number of null columns of $\widehat{\Delta}_k^{k-1}$. The number of k-cliques in G is upper bounded by*

$$\left\lfloor \frac{2Z}{k!} \right\rfloor.$$

Proof. We claim that a set of k vertices $x_1, \ldots, x_k$ is a k-clique if and only if all their representative permutations are generators of $\widehat{M}_k^{k-1}$ and they correspond to null columns of $\widehat{\Delta}_k^{k-1}$. Let us prove the two implications.

($\Rightarrow$) Since $x_1, \ldots, x_k$ constitute a clique, there is an edge $\{x_i, x_j\}$ $\forall i, j \in [1, k]$, implying that any k-trail with $x_1, \ldots, x_k$ as required vertices (in any order) is of length $k-1$, and thus if it is a representative k-trail it is a generator of $\widehat{M}_k^{k-1}$. Again because there is an edge between any pair of such vertices, removing any required vertex from one of these generators leads to a $(k-1)$-trail of length $k-2$, thus by Definition 1 they all correspond to null columns of $\widehat{\Delta}_k^{k-1}$.

($\Leftarrow$)

□

3.2 Second diagonal

We are concerned in this section with the analysis of the information contained in the groups $\widehat{H}_k^k$.

While the question about a general interpretation for all magnitude homology groups $\widehat{H}_k^k$ on the second diagonal remains open, we were able to understand part of the information contained in $\widehat{H}_3^3$: specifically, in the null columns of Δ_3^3. We believe the third normalized magnitude homology group on the second diagonal provides us with information regarding the number of chordless 4-cliques and 5-cycles in the graph.

Consider the following chain

$$
\begin{array}{ccccccccc}
\dots & \to & \widehat{M}_4^3 & \to & \widehat{M}_3^3 & \to & \widehat{M}_2^3 & \to & 0. \\
 & & \cup & & \cup & & \cup & & \\
 & & (x_1, x_2, x_3, x_4) & & (x_1, \hat{x}, x_3, x_4) & & (x_1, \hat{x}, \hat{y}, x_4). & &
\end{array}
$$

We focus on the null columns of the differential matrix Δ_3^3, that is on the basis elements $(x_1, \hat{x_2}, x_3, x_4)$ such that $\Delta_3^3(x_1, \hat{x_2}, x_3, x_4) = 0$. Notice that if $(x_1, \hat{x_2}, x_3, x_4) \in \ker(\Delta_3^3)$ then one of the following is true: either $\mathsf{len}(x_1, \hat{x_2}, \hat{x_3}, x_4) = 1$ or $\mathsf{len}(x_1, \hat{x_2}, \hat{x_3}, x_4) = 2$.

Observation 4. *Call L_1 and L_2 the subsets of the basis of $\widehat{M}_3^3$ such that $\mathsf{len}(x_1, \hat{x_2}, \hat{x_3}, x_4) = 1$ and $\mathsf{len}(x_1, \hat{x_2}, \hat{x_3}, x_4) = 2$, respectively. Then L_1 and L_2 for a partition of the null columns of $\ker(\Delta_3^3)$ and, consequently, of $\widehat{H}_3^3$. Indeed,...*

Lemma 6. *The number of basis elements $(x_1, \hat{x_2}, x_3, x_4)$ of $\widehat{M}_3^3$ such that the induced path (x_1, x_3, x_4) is eulerian, (x_1, x_3, x_4) is the image of Δ_4^3 and $\mathsf{len}(x_1, \hat{x_2}, \hat{x_3}, x_4) = 1$ is exactly the number of chordless 4-cycles in the graph.*

Proof. First notice that assuming (x_1, x_3, x_4) eulerian while $\mathsf{len}(x_1, \hat{x_2}, x_3, x_4) = 3$, imply the absence of the edge (x_1, x_3). Further, it is clear the $\mathsf{len}(x_1, \hat{x_2}, \hat{x_3}, x_4) = \mathsf{len}(x_1, x_4) = 1$ implies the existence of the edge (x_1, x_4), and thus (x_2, x_2, x_3, x_4) is an induced 4-cycle.

Now, the image of Δ_4^3 contains the elements $(x_1, \hat{x_2}, x_3, x_4) \in \widehat{M}_3^3$ such that $\mathsf{len}(x_1, \hat{x_2}, x_3, x_4) = \mathsf{len}(x_1, x_2, x_3, x_4)$. In other words, trails that do not contain the triangle (x_2, x_3, x_4), and in particular the edge (x_2, x_4). $\square$

dire che bisogna dividere per 8 perchè per ogni quadrato (1,2,3,4) abbiamo 4 4-trails ((1,2,3,4), (2,3,4,1), (3,4,2,1), (4,3,2,1)) e ad ognuno togliano i due vertici intermedi uno alla volta per avere un elemento di $\widehat{M}_3^3$.

Lemma 7. *The number of basis elements $(x_1, \hat{x_2}, x_3, x_4)$ of $\widehat{M}_3^3$ such that the induced path (x_1, x_3, x_4) is eulerian, (x_1, x_3, x_4) is the image of Δ_4^3 and $\mathsf{len}(x_1, \hat{x_2}, \hat{x_3}, x_4) = 2$ is bounds from above the number of chordless 5-cycles in the graph.*

Proof. Proceeding similarly as in the proof of Lemma 6, we point out that the hypotheses of (x_1, x_3, x_4) being eulerian and $\mathsf{len}(x_1, \hat{x_2}, x_3, x_4) = 3$, mean the absence of the edge (x_1, x_3). Now, $\mathsf{len}(x_1, \hat{x_2}, \hat{x_3}, x_4) = \mathsf{len}(x_1, x_4) = 2$ tells us that either the edge (x_2, x_4) exists, or there is a different path of length 2 from x_1 to x_4 involving a new vertex x_5.

credo che andando a guardare i generatori si possa anche trovare il

In addition, being $(x_1, \hat{x_2}, x_3, x_4)$ in the image of Δ_4^3 we know that the trail does not contain the triangle (x_2, x_3, x_4), and in particular the edge (x_2, x_4), so we are able to exclude the first case.

In conclusion, we are left with a 5-cycle $(x_1, x_2, x_3, x_4, x_5)$ which is surely missing diagonals (x_1, x_3), (x_1, x_4) and (x_2, x_4). $\square$

4 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 [?] to give an overall indication of the clustering in the network, and the *local*, first defined by Watts and Strogatz in [?] 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 $\widehat{H}_3^2$, determining thus a close relation between these tools.

4.1 Local clustering coefficient

The local clustering coefficient C_i of a node x_i describes the likelihood that the neighbours 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-trails of length 2 containing x_i.

Therefore, call $\widehat{M}_{3,i}^2$ the subgroup of $\widehat{M}_3^2$ such that x_i is the middle vertex of any 2-trail, so $\widehat{M}_{3,i}^2 = \{(x, x_i, y) : \mathsf{len}(x, x_i, y) = 2\}$. Then, by Section 3.1, the number of triangles containing x_i is precisely the number of null columns of $\ker(\Delta_3^2(\widehat{M}_{3,i}^2))$. Calling this number Z_i, we can write the local clustering coefficient as

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

Remark 2. *Given the connection just established between the local clustering coefficient and normalized magnitude homology, one could think of using $\widehat{H}_3^2$ 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.*

4.2 Global clustering coefficient

The global clustering coefficient C is based on 3-trails, i.e. on elements of $\widehat{M}_3^2$, and is computed as the number of closed 3-trails (or $3\times$ triangles) over the total number of 3-trails (both open and closed). That is, calling Z the number of null columns in $\ker(\Delta_3^2(\widehat{M}_3^2))$

$$C = \frac{Z}{|\widehat{M}_3^2|}.$$

5 Algorithm complexity

6 Conclusions

A Formal Algebraic Definitions

Definition 4 ([?]). *Consider the graph $G = (V, E)$. The (k, ℓ)-magnitude chain $MC_{k,\ell}(G)$ is the free abelian group generated by the k-trails of length ℓ in G.*

Definition 5 ([?]). *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

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

Definition 6 (Magnitude chain complex). *We indicate as $M_{*,\ell}(G)$ the following sequence of free abelian groups connected by differentials*

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

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

Definition 7. *The k-magnitude homology group of the graph G at level ℓ is the abelian group defined by*

$$MH_{k,l}(G) = H_k(M_{*,l}(G)) = \frac{\ker(\partial_k)}{\mathrm{im}(\partial_{k+1})}.$$

Contare colonne nulle è in P-space; elencarle no, perchè sono un numero esponenziale. Macchina di touring non det. può contare colonne nulle: il numero di combo accettate sono il numero di colonne nulle. Oracolo per problemi in $P^{\{}\#P\} \subseteq P$ space (problemi di conteggio su NDTM). Scriverli tutti richiede spazio esponenziale.