

# SIMPLIFYING THE HOMOLOGY OF NETWORKS VIA STRONG COLLAPSES

Adam C. Wilkerson\*

Terrence J. Moore†

Ananthram Swami†

Hamid Krim\*

\* North Carolina State University

† U.S. Army Research Laboratory

## ABSTRACT

There has recently been increased interest in applications of topology to areas ranging from control and sensing, to social network analysis, to high-dimensional point cloud data analysis. Here we use simplicial complexes to represent the group relationship structure in a network. We detail a novel algorithm for simplifying homology and “hole location” computations on a complex by reducing it to its core using a strong collapse. We show that the homology and hole locations are preserved and provide motivation for interest in this reduction technique with applications in sensor and social networks. Since the complexity of finding “holes” is quintic in the number of simplices, the proposed reduction leads to significant savings in complexity.

**Index Terms**— Simplicial complex, homology, simplicial collapse, sensor network, social network

## 1. INTRODUCTION

There is a growing interest in the practical applications of topology in many areas, including control and sensing [4, 5, 11], social network analysis [10], and the analysis of point cloud data [3]. In particular, simplicial complexes have been useful for representing relationships between objects beyond the pairwise connections in a graph. However, the combinatorial nature of these structures often results in significant computational challenges. For example, the computational complexity of finding topological “holes” in a simplicial complex can be quintic in the number of simplices [5]. Hence, there is a need for methods that help mitigate these challenges.

One such approach for simplifying homology computation is utilizing simplicial collapses, which eliminate redundant parts of the simplicial complex that do not affect the topological structure (i.e., the homology). Certain trial-and-error-type collapses have already been exploited in sensor networks [12]. In this work, we independently develop a method of collapsing nearly equivalent to that of [2], called a strong

collapse; however, unlike [2], we use and keep track of labels for certain objects in the simplicial complex. We further show that this strong collapse preserves the “locations” of the topological holes in the complex. We also discuss some characteristics of the reduction in complexity by using a strong collapse prior to computing homology. And finally, we address particular applications of strong collapses relevant to homology computations and hole location in simplicial complex representations of sensor and social networks.

## 2. SIMPLICIAL COMPLEXES AND HOMOLOGY

A *simplicial complex*, i.e., a collection of sets closed under the subset operation, is a generalization of a graph useful in representing higher-than-pairwise connectivity relationships. The elements of any set are called *vertices* and the set itself is called a *simplex* (or  $k$ -simplex to denote it is  $k$ -dimensional and has exactly  $k + 1$  vertices). Any proper subset  $\Delta$  of a simplex  $\Gamma$  is called a *face* of  $\Gamma$ . A maximal simplex, i.e., a simplex that is not a face of any other simplex, is called a *facet* of the complex. The dimension of the simplicial complex is the largest dimension of its facets. One example of a simplicial complex is the representation of the coverage regions in a sensor network, where the vertices correspond to the sensors and the simplices represent that the sensors share a coverage region. Another example is a network of collaborations of members on a project, where the members and collaborations correspond to vertices and simplices, respectively.

Note, a  $k$ -simplex can be geometrically represented as the convex hull of  $k + 1$  affinely independent points in  $k$ -dimensional space. Hence, in this geometric realization, a simplicial complex is a collection of simplices (in a space of sufficient dimension) that are closed under the subset operation and that can only intersect other simplices along a face.

*Simplicial homology* is the study of the homology spaces on a simplicial complex, i.e., the sequence of vector spaces  $\{H_0(X), H_1(X), \dots, H_d(X)\}$  on a  $d$ -dimensional simplicial complex  $X$ . The rank of each vector space  $H_k(X)$  is called the  $k^{\text{th}}$  *Betti number* of  $X$  and is the count of the distinct  $(k + 1)$ -dimensional “holes” in  $X$ . A  $k$ -dimensional hole, in the geometric realization, is the empty space bounded by a collection of  $(k - 1)$ -simplices. More accurately, this hole means there is a  $(k - 1)$ -cycle that does not bound. (See

Adam Wilkerson was supported as a summer student by the U.S. Army Research Laboratory under OARU contract 1120-1120-99. Hamid Krim was partially supported by AFOSR grant FA9550-10-1-0240 and by DTRA grant HDTRA1-08-0024.

Figure 1 for an example. A more precise definition can be found in [7].)

## 2.1. Labeled Simplicial Complexes

A relation  $\nabla$  between elements of two sets  $A$  and  $B$  naturally induces two simplicial complexes with labeled simplices. For instance, for every element  $a \in A$ , one can associate a *labelled simplex* of points in  $B$  that are related to the label  $a$ ,  $\sigma_a = \{b \in B : a \nabla b\}$ . This complex can be denoted as  $K_A(B, \nabla)$ . Similarly, there exists the complex  $K_B(A, \nabla^{-1})$ . These labeled simplicial complexes are called *conjugate complexes* to one another and are defined by this set relation. If the relation is understood, we can denote these complexes as  $K_A(B)$  and  $K_B(A)$ . A classical result by Dowker [6] states that  $K_A(B)$  and  $K_B(A)$  have identical homologies. One example of a labeled simplicial complex is the set  $\mathcal{A}$  of actors and the set  $\mathcal{M}$  of movies with the relation  $m \nabla a$  between  $a \in \mathcal{A}$  and  $m \in \mathcal{M}$  if the actor  $a$  appeared in movie  $m$ . More examples and other applications can be found in the work of Atkin [1] and Johnson [8].

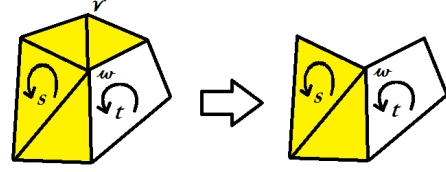
For our purposes, we shall refer to one of the two sets as the vertex set  $V = \{v_1, v_2, \dots, v_n\}$  and the other set as the label set  $L = \{l_1, l_2, \dots, l_m\}$ . Hence, the complex of interest is  $X = K_L(V)$ , which we shall often denote simply by  $X$ , when no confusion can arise. Every label is unique, i.e., no label represents two unique simplices; however, any simplex might be multi-labeled. For example, the same actors might appear in a movie sequel. By definition, all facets are labeled. We shall denote the conjugate of  $X$  as  $X^c = K_V(L)$ . It should be noted that in this context, conjugation creates a complex  $X^c$  with vertices corresponding to the labels in  $X$ , and simplex labels corresponding to the vertices in  $X$ .

## 3. STRONG COLLAPSES

We need a few more definitions before proceeding to the main results of this paper. Two labeled simplices  $l_i$  and  $l_j$  in  $L$  are said to be *q-near* to one another if they share a  $q$ -dimensional face. Note, a labeled simplex must be at least  $q$ -dimensional to be  $q$ -near another labeled simplex. The (normalized) *eccentricity* of a labeled simplex  $\sigma \in L$  is [9]

$$\text{ecc}(\sigma) = \frac{\hat{q} - \tilde{q}}{\hat{q} + 1}, \quad (1)$$

where  $\hat{q}$  is the dimension of  $\sigma$  and  $\tilde{q}$  is the largest  $q$  for which  $\sigma$  is  $q$ -near another labeled simplex. (For an isolated labeled simplex, we define  $\tilde{q} = -1$ .) Since  $\hat{q} \geq \tilde{q}$  and  $\hat{q} \geq 0$ , then  $0 \leq \text{ecc}(\sigma) \leq 1$ . Note that the eccentricity of a labeled simplex  $\sigma$  is zero if and only if  $\sigma$  is a face of a larger simplex or if the simplex labeled by  $\sigma$  has another label in  $L$ . Define the labeled subcomplex  $\tilde{X} = K_{\tilde{L}}(\tilde{V})$  as that obtained by first reducing the label set so that no simplex is multi-labeled and then removing all labels in  $L$  with 0 eccentricity. This leaves



**Fig. 1.** Vertex  $v$  is dominated by vertex  $w$ . Cycle  $s$  bounds in the topology, whereas cycle  $t$  does not bound.

us with a labeled complex where only the facets are labeled (and only labeled once), such that  $\tilde{X} = X$ ,  $\tilde{V} = V$ , and  $\tilde{L} \subset L$ . We then define a *reduction* on a labeled simplicial complex  $X$  as  $X^r = (\tilde{X}^c)^c$ .

**Theorem 1.**  $H_i(X^r)$  and  $H_i(X)$  are isomorphic as vector spaces for every index  $i$ .

The proof of this theorem is a consequence of the previously stated result from Dowker [6] that conjugate complexes have isomorphic homology. Interpreted geometrically, this states that the number of holes in the complex is preserved by a reduction of a labeled simplicial complex. Recall that since the labeled simplices  $L$  in the original complex are the vertices in the conjugate complex, then this process of eliminating simplices using  $(\cdot)^c$  reduces the set of vertices in the conjugate to  $\tilde{L}$  and, consequently, reduces the number of (unlabeled) simplices there as well. Then in the conjugate, the set of labeled simplices is again reduced from  $\tilde{V} = V$  to  $\tilde{V}$ . The reduction leaves us with a subcomplex of the original complex that preserves the holes. The successive reduction of a labeled simplicial complex must converge to a stable complex since every bounded monotone sequence converges.

This approach was developed independently by Barmak and Minian [2]. In that work, the authors developed a homotopy theory of *strong collapses*. These collapses occur by way of the sequential deletion of vertices if the vertex  $v \in V$  is *dominated* by another vertex  $w \in V$ , i.e., if every facet containing  $v$  also contains  $w$  (e.g., see Figure 1). This collapse was streamlined by successively taking the *nerve* of a simplicial complex, where the nerve happens to be the same as the conjugate complex if only facets are labeled (although labels are not tracked).

This collapse is equivalent to the reduction executed by our approach because every simplex label that is eliminated in the conjugate complex at each step has eccentricity 0. It is evident that a vertex  $v \in X$  is dominated if and only if its corresponding simplex in the conjugate  $v^c \in X^c$  has eccentricity 0. Then, when iterated until it converges, our strong collapsing algorithm, found in Section 3.1, collapses the complex  $X$  to its core since the stable limit of the algorithm can have no 0-eccentricity simplices in the conjugate, and is therefore minimal.

A result not found in [2] is that not only are the holes preserved, but their “locations” are also preserved, where a hole location in a simplicial complex is defined to be any of the shortest-length cycles that does not bound (in the homology coset or generator class corresponding to that hole in  $H_k(X)$ ).

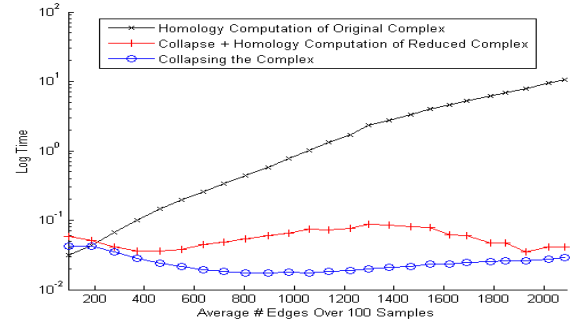
**Theorem 2.** *For each hole, at least one shortest-length cycle corresponding to each hole’s generator class will remain after a strong collapse.*

*Sketch of Proof.* The collapse results in the deletion of dominated vertices [2]: If a vertex  $v$  is in a shortest-cycle corresponding to a  $k$ -hole’s generator class, it can be shown that it can only be dominated by a vertex  $w$  with which it shares no  $k$ -simplices in the cycle. Thus, collapsing  $v$  into  $w$  collapses the  $k$ -simplices incident to  $v$  into  $k$ -simplices incident to  $w$  in a one-to-one fashion along  $(k + 1)$ -simplices shared by  $v$  and  $w$ . Hence, any shortest-length generator of an element of  $H_k$  can be replaced by another generator of the same length which contains no dominated vertices.  $\square$

This theorem shows that collapsing a network via our algorithm preserves not only the homology of the complex, but also the location and size of the features which determine its topology. This is relevant when the topological location of a hole, as it is defined here, has significance in the network. In the sensor networks setting, the algorithm preserves hole-location, providing a reduced representation of the network that retains the topological location of uncovered areas so that they may be found and repaired. This scenario is covered in more detail in Section 4.1. In the social network setting, this corresponds to preserving a shortest path between any two individuals remaining in the complex after the collapse. Since the collapse maintains those individuals most vital to the network’s topology, this is an indicator of the core entities binding the network together, as detailed in the example in Section 4.2.

Before presenting our algorithm to execute this approach, we comment on the reduction in complexity due to a strong collapse. Since the strong collapse process reduces the number of simplices in a given complex, the computation of a combinatorial invariant like homology (or hole location) benefits greatly from it. For a labeled  $d$ -dimensional simplicial complex with  $n$  vertices and  $m$  labels, the algorithm in Section 3.1 takes on the order of  $d(n^2 + m^2)$  operations for each iteration. This low cost potentially leads to a significant reduction in complexity as the homology computation can be quintic in the number of simplices. [5]

To show the dramatic effect on time complexity of a strong collapse on homology calculations in this scenario, we generate 100 random (Erdős-Rényi) graphs at each of 25 different average node degrees and then use the *flag complexes* of the graphs (i.e., the complex generated by a list of every clique of every size in the graph) for geometric realizations of the simplicial complexes. In Figure 2, we show the log average time to compute the homology of the original simplicial



**Fig. 2.** Time complexity of homology computations and time complexity of a strong collapse and homology computation of the core

complex realizations compared with the time taken to compute the strong collapse and the homology of the (reduced) core of the complex.

### 3.1. Our strong collapsing algorithm

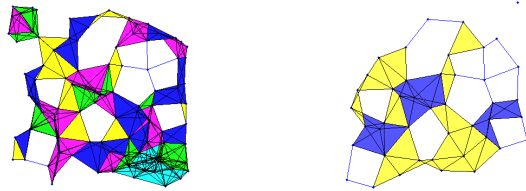
To execute this reduction, we first construct an  $m \times n$  incidence matrix  $M_{L,V}(X) = [m_{ij}]$  for a labeled simplicial complex  $X = K_L(V)$  with  $L = \{l_1, l_2, \dots, l_m\}$  and  $V = \{v_1, v_2, \dots, v_n\}$ , where

$$m_{ij} = \begin{cases} 1 & : l_i \nabla v_j \\ 0 & : \text{else} \end{cases} \quad (2)$$

Note that the incidence matrix for the conjugate complex  $X^c$  is simply the transpose, i.e.,  $M_{V,L}(X^c) = M_{L,V}^T(X)$ . The algorithm is as follows:

1. Sort labels  $L$  so that the dimension of simplices is monotone increasing to obtain  $L_s = \{l_{s_1}, l_{s_2}, \dots, l_{s_m}\}$ .
2. For each row  $i$  in  $M_{L_s,V}(X)$ , determine  $\text{ecc}(l_{s_i})$  using (1) (this is done with the dot product with rows  $> i$ ).
  - If  $\text{ecc}(l_{s_i})$  is zero, mark label  $s_i$  for removal.
3.  $\tilde{L}_s$  is the new label set with all zero-eccentricity labels removed,  $\tilde{V} = V$  is the new vertex set,  $M_{\tilde{L}_s,\tilde{V}}(X)$  is the new incidence matrix.
4. Repeat steps 1-3, interchanging the notation of vertices and labels above and using the transpose of the current incidence matrix.
5. Both the label and vertex set have been reduced to  $\tilde{\tilde{L}}$  and  $\tilde{\tilde{V}}$  and the complex reduced to  $X_1 \triangleq X^r \subset X$  from Theorem 1. Iterate steps 1-4, until  $X_j = X_{j-1}$ .

This final simplicial complex  $X_n$  is the core after strong collapsing.



(a) Original Sensor Network (b) Reduced Sensor Network

**Fig. 3.** Strong collapse of the simplicial complex representation of sensor coverage

## 4. APPLICATIONS

### 4.1. Sensor Networks

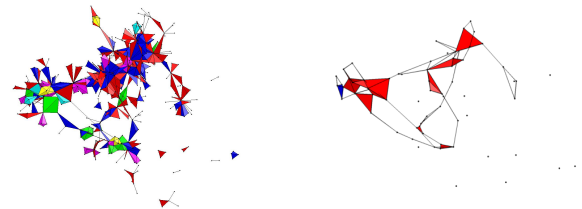
Topological approaches have proven useful in distributively detecting and autonomously repairing gaps in sensor coverage for sensors that lack location information [4, 11]. In this scenario each vertex corresponds to a sensor and each simplex represents the coverage overlap between the sensors associated with the vertices of the simplex. Hence a two-dimensional hole in this simplicial complex representation corresponds to a gap in coverage inside the sensor field (see Figure 3(a)).

The holes in the network need to be found, localized, and corrected. The overall computational complexity of this process depends on the number of 1-simplices (edges) and 2-simplices (triangles) in the complex. As we have shown in Theorem 2, a strong collapse of this simplicial complex will reduce the number of vertices (and therefore the number of 1- and 2-simplices) while preserving the hole locations relevant to the sensor coverage gaps, thereby leading to a significant reduction in complexity. Figure 3(b) shows the core of the original complex after the strong collapse and shows that the existing holes and the shortest-length cycles locating the holes are preserved. Clearly, the sensors corresponding to the removed vertices are still being used in the coverage, but they are not needed in locating the coverage holes.

### 4.2. Social Networks

Simplicial complexes have long been used to represent group structure in social networks [1]. In particular, it was recently shown that the vertices (corresponding to authors) incident to the hole locations in a simplicial complex representing the co-authorship collaborations have interesting properties such as high centrality statistics [10]. In this work, a simplicial complex was constructed from the publications of academia, industry, and government scientists in the Communications & Networks Collaborative Technology Alliance (C&N CTA) program.<sup>1</sup> The data set consists of 960 publications by 518

<sup>1</sup><http://www.arl.army.mil/www/default.cfm?page=390>



(a) Original Network (b) Reduced Network

**Fig. 4.** The C&N CTA coauthorship network and its core; in each visualization there are 16 components with 24 2-dimensional holes.

authors, which provides a natural relation between the set of authors  $\mathcal{A}$  and the set of papers  $\mathcal{P}$ , inducing a simplicial complex  $K_{\mathcal{P}}(\mathcal{A})$  where the vertices correspond to authors and the labeled simplices correspond to papers with the vertices of each labeled simplex identifying the authors of each paper (see Figure 4(a)).

A strong collapse on the C&N CTA collaboration network, shown in Figure 4(b), sequentially eliminates papers and authors that are not critical to the underlying topology. The vertices retained by this strong collapse are predominantly principal investigators in the program, i.e., the most qualitatively important vertices in the complex. For example, the papers written by a graduate student would almost always be coauthored with the student's advisor. Hence, the vertex corresponding to that student, being dominated by the vertex corresponding to his or her advisor, would not remain in the core. Similarly, a paper published by a subgroup of a larger publishing group also does not survive as a labeled simplex. The core, then, is left with the 67 relevant coauthors and the fundamental collaborations (in this case, 78 papers) that determine the network structure.

## 5. SUMMARY AND CONCLUSION

Simplicial complexes are valuable tools for analyzing network data, but extracting topological information from them can be expensive. We develop an algorithm for strong collapsing that reduces a complex through vertex deletion down to a core complex that maintains its topological structure and describe an algorithm executing this reduction. Moreover, this reduction maintains the location of holes within the network. For sensor networks, detecting and locating the sensors nearby the hole can assist in coverage repair. For collaboration networks, a strong collapse identifies the authors who are most central to the connections in the network. Furthermore, we demonstrate that in practice, this algorithm reduces the complexity of computing the homology of networks dramatically, thus making it a valuable tool in the field of computational topology.

## 6. REFERENCES

- [1] R. H. Atkin, "From cohomology in physics to q-connectivity in social sciences," *International Journal of Man-Machine Studies*, vol. 4, pp. 341–362, 1972.
- [2] J. A. Barmak and E. G. Minian, "Strong homotopy types, nerves, and collapses," *Discrete & Computational Geometry*, vol. 47, pp. 301–328, 2012.
- [3] G. Carlsson, "Topology and data," *Bulletin of the American Mathematical Society*, vol. 46, pp. 255–308, 2009.
- [4] H. Chintakunta and H. Krim, "Divide and conquer: Localizing coverage holes in sensor networks," in *Proceeding of the 2010 7th Annual IEEE Communications Society Conference Sensor Mesh and Ad Hoc Communications and Networks (SECON)*, Boston, MA, 2010, pp. 1–8.
- [5] V. de Silva and R. Ghrist "Coordinate-free coverage in sensor networks with controlled boundaries via homology," *International Journal of Robotics Research*, vol. 25, pp. 1205–1222, 2006.
- [6] C. H. Dowker, "Homology groups of relations," *The Annals of Mathematics*, vol. 56, pp. 84–95, 1952.
- [7] A. Hatcher, *Algebraic Topology*. Cambridge University Press: Cambridge, 2001.
- [8] J. J. Johnson, "Some structures and notation in q-analysis," *Environment and Planning B*, vol. 8, pp. 77–86, 1981.
- [9] S. Maletić, M. Rajković, and D. Vasiljević, "Simplicial complexes of networks and their statistical properties," In: M. Bubak, G. D. van Albada, J. Dongarra, P. M. A. Sloot (eds.) *ICCS 2008, Part II*. LNCS, vol. 5102, pp. 568–575. Springer: Heidelberg, 2008.
- [10] T. J. Moore, R. J. Drost, P. Basu, R. Ramanathan, and A. Swami, "Analyzing collaboration networks using simplicial complexes: A case study," in *Proceedings of the IEEE INFOCOM 2012 Workshop (NetSciCom)*, March 2012, pp. 238–243.
- [11] A. Tahbaz-Salehi and A. Jadbabaie, "Distributed coverage verification in sensor networks without location information," *Proceedings of the 47th IEEE Conference on Decision and Control*, Cancun, Mexico, 2008, pp. 4170–4176.
- [12] A. Vergne, L. Decreusefond, and P. Martins, "Reduction algorithm for simplicial complexes," preprint: <http://hal.archives-ouvertes.fr/hal-00688919>, pp. 1–8, 2012.