Chowdhury *et al. Applied Network Science* (2022) 7:4
https://doi.org/10.1007/s41109-021-00441-z

Applied Network Science

RESEARCH Open Access

Path homologies of motifs and temporal network representations

Samir Chowdhury[1], Steve Huntsman[2*] and Matvey Yutin[3]

*Correspondence: sch213@nyu.edu
[2] Alexandria, Virginia, USA
Full list of author information is available at the end of the article

Abstract

Path homology is a powerful method for attaching algebraic invariants to digraphs. While there have been growing theoretical developments on the algebro-topological framework surrounding path homology, *bona fide* applications to the study of complex networks have remained stagnant. We address this gap by presenting an algorithm for path homology that combines efficient pruning and indexing techniques and using it to topologically analyze a variety of real-world complex temporal networks. A crucial step in our analysis is the complete characterization of path homologies of certain families of small digraphs that appear as subgraphs in these complex networks. These families include all digraphs, directed acyclic graphs, and undirected graphs up to certain numbers of vertices, as well as some specially constructed cases. Using information from this analysis, we identify small digraphs contributing to path homology in dimension two for three temporal networks in an aggregated representation and relate these digraphs to network behavior. We then investigate alternative temporal network representations and identify complementary subgraphs as well as behavior that is preserved across representations. We conclude that path homology provides insight into temporal network structure, and in turn, emergent structures in temporal networks provide us with new subgraphs having interesting path homology.

Keywords: Path homology, Topological data analysis, Temporal networks

Introduction

Understanding network structure through the lens of higher-order organization has gained considerable attention in recent years (Battiston et al. 2020). In one sense, higher-order structures correspond to small subgraphs in contrast to lower-order structures such as nodes or edges; such subgraphs have been shown to be crucial in driving the function of complex networks (Benson et al. 2016). In a different sense, higher-order organization refers to the mesoscale structure of networks reflected in the hierarchical organization of its clusters and communities (Ravasz and Barabási 2003; Lancichinetti et al. 2009). The mathematical discipline of algebraic topology provides a convenient, well-established framework for studying the latter form of higher-order organization (Sizemore et al. 2019). Specifically within algebraic topology, a particular linear-algebraic tool known as *homology* has achieved great success in data applications due to its tradeoff between computational ease and ability to represent complex, higher-order

structures (Carlsson 2009). Moreover, such methods have been applied on network datasets to discover global structures that could not be easily obtained through standard, statistical mechanics-driven approaches to the study of complex networks (Petri et al. 2013). Crucial to these discoveries was the early development of computational libraries (Adams et al. 2014) that enabled network scientists to rapidly test new mathematical tools and discover new questions that may be answered using such methods.

The particular flavor of homology that is currently best known across disciplines is *simplicial homology* (Giusti et al. 2016). Here the basic objects are *simplices*—points, edges, triangles, tetrahedra, and so on—that encode relationships beyond the dyadic relationships captured by the edges of a graph. These simplices assemble together to form a structure called a *simplicial complex*, which in turn provides a principled method for representing complex shapes (Edelsbrunner and Harer 2010). Simplicial homology then searches for the presence of *holes* or *cavities* in the network that may signify regions where small subgroups of nodes participate in correlated activity, but without overall consensus in the region of interest (Sizemore et al. 2019).

A caveat, however, is that simplicial homology is not immediately compatible with *directed* graphs (digraphs). In the linear-algebraic representation central to simplicial homology, directed edges of the form $a \to b$ and $b \to a$ are assigned to the same vector subspace, thus leading to a potential loss of information (Chowdhury and Mémoli 2018) (e.g. an email from an employee to a superior is semantically different from an email from a superior to the employee). A remedy to this situation is obtained via the notion of *path homology*, a version of homology defined on directed graphs that was developed in Grigor'yan et al. (2014a) and associated works (Grigor'yan et al. 2012, 2014b, 2015, 2017, 2018a, b). Path homology resolves the $(a \to b)$-vs-$(b \to a)$ situation by assigning different vector subspaces to each direction. Moreover, path homology generalizes simplicial homology as follows. Given a simplicial complex S representing a shape, there is a natural ordering from lower to higher order simplices, given by the subset relation $\sigma \subseteq \tau$ for simplices σ, τ (e.g. a point belonging to an edge, or an edge belonging to a triangle). A digraph G_S can then be constructed by taking the simplices as nodes, and directed edges $\sigma \to \tau$ representing inclusion relations. Then the path homology of G_S is naturally isomorphic to the simplicial homology of the original complex S (Grigor'yan et al. 2014a). In this sense, path homology generalizes simplicial homology. Additionally, path homology serves as a general framework for computing global structures in arbitrary digraphs via linear algebra. The tradeoff for this added generality is that the intuition for path homology can be more involved than finding non-local loops and cavities in a network. However, a useful (if imperfect) interpretation is that path homology measures the consistency and robustness of directional flow in a digraph. Specifically, in a digraph with the architecture of a multilayer perceptron (MLP)—i.e. layers of nodes with unidirectional edges across consecutive layers, also referred to as a deep feedforward neural network—path homology is positively related to both the width of the layers (robustness) and the agreement of edge directions (consistency) (Chowdhury et al. 2020).

Towards grounding the preceding discussion in a concrete application, consider the problem of examining *control flow* in computer programs. Control flow of code refers to the order in which information is passed among variables to carry out operations, and it can naturally be represented by a graphical structure that can in turn by quantified by

various metrics to predict defects and points of failure. One popular metric is *cyclomatic complexity* (McCabe 1976; Ebert et al. 2016), which measures the number of linearly independent paths in the control flow graph. More explicitly, cyclomatic complexity is calculated via the simplicial homology of a graph. It has been shown that path homology is a stronger analogue of cyclomatic complexity that benefits by capturing the natural directionality of control flow graphs that is ignored by cyclomatic complexity (Huntsman 2020).

For further applications, multiscale versions of path homology have been developed in Chowdhury and Mémoli (2018), Dey et al. (2020), Lin et al. (2019). In sum, however, the empirical application of path homology to large-scale data structured as digraphs remains largely unexplored. The bottleneck for such exploration is the lack of methodological developments at different stages of the pipeline, including algorithmic development, implementation, application to real-world networks, and posthoc analysis of the insights contributed by path homology.

To help bridge this perceptual gap between algebraic and combinatorial structure in service of analyzing real-world networks, we elaborate on the conference paper (Chowdhury et al. 2020). After reviewing the basics of path homology in §2, we present an algorithm and implementation for computing path homology in arbitrary dimension in §3. In § 4, we then use this algorithm to compute the path homologies of (1) all digraphs on ≤ 4 vertices, (2) all directed acyclic graphs on ≤ 6 vertices, (3) all undirected graphs on ≤ 6 vertices, (4) Erdős-Rényi random graphs, and (5) small digraphs that exhibit torsion. These examples[1] help develop intuition about path homology and yield digraph families whose path homology has surprising behavior. We then use this algorithm in §5-7 to compute path homologies of digraphs that represent the same three real-world temporal networks in complementary ways. Here we identify salient subgraphs with exemplars that appear in § 4 and relate them to broader network behavior within a given representation. In §8 we perform additional analysis on these real-world temporal networks using popular existing network measures such as density and clustering coefficient to highlight differences from path homology. Finally, we discuss gross network behaviors that are preserved across representations and make concluding remarks in § 9.

Related literature

Parallel to path homology and its intermediate constructions, there have been recent developments in extending notions related to standard (i.e. simplicial) homology to account for directionality in the graph setting. Such advances started with the use of *directed flag complexes* in Reimann et al. (2017) and continued with theoretical, algorithmic, and empirical developments in Turner (2019), Lütgehetmann et al. (2020), Gebhart and Funk (2020). Further developments of related ideas have appeared in Chowdhury and Mémoli (2018), Méndez and Sánchez-García (2020), Bergomi et al. (2020).

[1] While our analyses have revealed recurring motifs in the sense of small subgraphs that are statistically more frequent than in a null model (Smoly et al. 2017), our techniques are designed to identify *individual* subgraphs that exhibit robust and consistent directionality.

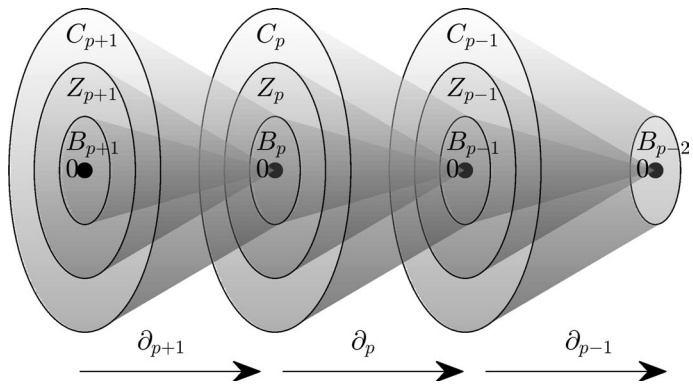

Fig. 1 Schematic of a chain complex. Here C_p, B_{p-1}, and Z_p are respectively the domain, codomain, and kernel of ∂_p, which conspire to make the homology $H_p := Z_p/B_p$ well defined

Path homology

Here we sketch the basic elements of *path homology* as treated in Grigor'yan et al. (2012), Chowdhury and Mémoli (2018). Although our development is nominally self-contained, a reader who wants background in topology (e.g. the foundational theory of simplicial homology that shares many similarities with path homology) is commended to Ghrist (2014), Hatcher (2001).

Let X be a finite set, and let $\mathbb{F}$ be a field. A standard algebraic construction is the free $\mathbb{F}$-vector space on X, i.e. a vector space with standard basis $\{e_x : x \in X\}$. We denote this space by $\mathbb{F}^X \cong \mathbb{F}^{|X|}$, and additionally set $\mathbb{F}^{\varnothing} := \{0\}$. Next let $D = (V, A)$ be a loopless digraph, and consider the sets $V^p = \{(v_0, \ldots, v_{p-1}) : v_i \in V,\ 0 \le i \le p-1\}$ for $p \ge 0$. The *non-regular boundary operator* $\partial_{[p]} : \mathbb{F}^{V^{p+1}} \to \mathbb{F}^{V^p}$ is the linear map whose action on the standard basis is

$$\partial_{[p]} e_{(v_0,\ldots,v_p)} = \sum_{j=0}^{p} (-1)^j e_{\nabla_j(v_0,\ldots,v_p)}, \tag{1}$$

where $\nabla_j(v_0, \ldots, v_p) := (v_0, \ldots, v_{j-1}, v_{j+1}, \ldots, v_p)$ is the result of deleting v_j from $(v_0, \ldots, v_p)$. A few lines of algebra focused on index bookkeeping shows that $\partial_{[p-1]} \circ \partial_{[p]} \equiv 0$, i.e. $(\mathbb{F}^{V^{p+1}}, \partial_{[p]})$ is a *chain complex* (Ghrist 2014) as depicted in Fig. 1. The "boundary of a boundary is zero" condition admits a topological interpretation: as an example, consider that a 2-dimensional disk has boundary given by a circle, but the circle itself has no boundary.

Any chain complex (C_p, ∂_p) gives rise to an algebraic invariant called *homology*. The invariance property of homology is that it behaves nicely with respect to maps on the chain complex that are induced by an underlying transformation of a common structure (here, a digraph). Writing $Z_p := \ker \partial_p$ and $B_p := \operatorname{im} \partial_{p+1}$, i.e. the kernel and image of the boundary maps, the dimension p homology of the chain complex (C_p, ∂_p) is the quotient

$$H_p := Z_p/B_p. \tag{2}$$

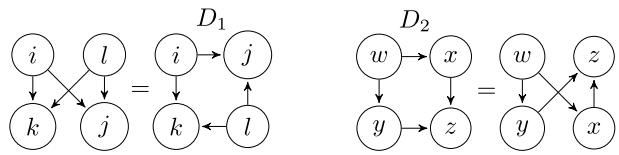

Fig. 2 (Center) The digraph D_2 has trivial path homology but the digraph D_1 does not. (Left) The nontrivial path homology of D_1 can be interpreted as a consistent and robust unidirectional flow. (Right) The trivial path homology of D_2 can be attributed to w and z appearing as "bottlenecks" that prevent robustness of the flow. This intuition is formalized by results in Chowdhury et al. (2019)

H_p is a finitely generated abelian group, and therefore has the form $\mathbb{Z}^{\beta_p} \oplus T_p$, where the *torsion* T_p is a finite abelian group. The topological interpretation of this construction is that the $\mathbb{Z}^{\beta_p}$ and T_p terms correspond to "voids" and "twistedness" in a topological space, respectively. When computed over a field $\mathbb{F}$, the C_p are vector spaces and the torsion is zero. In this setting the *Betti numbers* $\beta_p := \dim H_p = \dim Z_p - \dim B_p$ completely characterize homology up to isomorphism.

Returning to the loopless digraph D, consider the set $\mathcal{A}_p(D)$ of *allowed p-paths*:

$$\{(v_0, \ldots, v_p) \in V^{p+1} : (v_{j-1}, v_j) \in A,\ 1 \le j \le p\}. \tag{3}$$

By convention, we set $\mathcal{A}_0 := V$, $V^0 \equiv \mathcal{A}_{-1} := \{0\}$ and $V^{-1} \equiv \mathcal{A}_{-2} := \varnothing$. Path homology arises indirectly from a chain complex derived from $(\mathbb{F}^{V^{p+1}}, \partial_{[p]})$[2]. Write

$$\Omega_p := \left\{\omega \in \mathbb{F}^{\mathcal{A}_p} : \partial_{[p]}\omega \in \mathbb{F}^{\mathcal{A}_{p-1}}\right\}, \tag{4}$$

$\Omega_{-1} := \mathbb{F}^{\{0\}} \cong \mathbb{F}$, and $\Omega_{-2} := \mathbb{F}^{\varnothing} = \{0\}$. Now $\partial_{[p]}\Omega_p \subseteq \mathbb{F}^{\mathcal{A}_{p-1}}$, so $\partial_{[p-1]}\partial_{[p]}\Omega_p = 0 \in \mathbb{F}^{\mathcal{A}_{p-2}}$ and $\partial_{[p]}\Omega_p \subseteq \Omega_{p-1}$. The *(non-regular) path complex* of D is accordingly defined to be the chain complex (Ω_p, ∂_p), where $\partial_p := \partial_{[p]}|_{\Omega_p}$.[3] The *(non-regular) path homology* of D is just the homology of the path complex (Ω_p, ∂_p).

A basic example illustrating the mechanics of path homology is provided by considering the digraphs D_1 and D_2 in Fig. 2. $\mathcal{A}_1(D_1)$ and $\mathcal{A}_1(D_2)$ are given by the digraph arcs, $\mathcal{A}_2(D_1) = \varnothing$, and $\mathcal{A}_2(D_2) = \{(w,x,z),(w,y,z)\}$. We have $\partial_{[1]}(e_{(i,j)} - e_{(i,k)} - e_{(l,j)} + e_{(l,k)}) = 0$, and similarly $\partial_{[1]}(e_{(x,z)} + e_{(w,x)} - e_{(y,z)} - e_{(w,y)}) = 0$. Now $\partial_{[2]}e_{(w,x,z)} = e_{(x,z)} - e_{(w,z)} + e_{(w,x)} \notin \mathbb{F}^{\mathcal{A}_1(D_2)}$ and $\partial_{[2]}e_{(w,y,z)} = e_{(y,z)} - e_{(w,z)} + e_{(w,y)} \notin \mathbb{F}^{\mathcal{A}_1(D_2)}$ (because the arc $w \to z$ is not present), but

$$\begin{aligned}\partial_{[2]}(e_{(w,x,z)} - e_{(w,y,z)}) &= e_{(x,z)} - e_{(w,z)} + e_{(w,x)} - e_{(y,z)} + e_{(w,z)} - e_{(w,y)} \\ &= e_{(x,z)} + e_{(w,x)} - e_{(y,z)} - e_{(w,y)} \in \mathbb{F}^{\mathcal{A}_1(D_2)}.\end{aligned}$$

It follows that $H_1(D_1) = \mathbb{F}$ and $H_1(D_2) = \{0\}$, i.e., the Betti numbers are different: $\beta_1(D_1) = 1$ and $\beta_1(D_2) = 0$.

[2] Path homology can also be defined over rings such as $\mathbb{Z}$ without any modifications besides a change of notation. This definition gives additional power: M. Yutin has exhibited digraphs on as few as six vertices that have torsion: see Fig. 5.

[3] The implied *regular path complex* is constructed so as to prevent a directed 2-cycle from having nontrivial 1-homology. While (Grigor'yan et al. 2012) evinces a preference for regular path homology, in our view non-regular path homology is simpler, richer, and more likely useful in applications. As a practical matter, our rationale amounts to the convention that a directed 2-cycle should count as a "hole."

For convenience, henceforth we generally replace the path complex (Ω_p, ∂_p) with its *reduction*

$$\ldots \Omega_{p+1} \xrightarrow{\partial_{p+1}} \Omega_p \xrightarrow{\partial_p} \Omega_{p-1} \xrightarrow{\partial_{p-1}} \ldots \xrightarrow{\partial_1} \Omega_0 \xrightarrow{\tilde{\partial}_0} \mathbb{F} \longrightarrow 0. \tag{5}$$

Assuming the original complex is nondegenerate and using an obvious notational device, this has the minor effect $\tilde{H}_0 \oplus \mathbb{F} \cong H_0$, while $\tilde{H}_p \cong H_p$ for $p > 0$. Similarly, $\tilde{\beta}_p = \beta_p - \delta_{p0}$, where the Kronecker delta $\delta_{jk} := 1$ iff $j = k$ and $\delta_{jk} := 0$ otherwise.

Path homologies of the mutual dyad subgraphs

As another example with practical relevance that will be exhibited in § 5, we use path homology to characterize a family of network motifs that we call the *n-uplinked mutual dyads*—or dually, the n-downlinked mutual dyads—in reference to the original terminology from Milo et al. (2002). Given an integer $n \geq 1$, the n-uplinked mutual dyad W_n is a digraph with vertex set $\{a, b, 1, 2, \ldots, n\}$ and edge set $\{(a, b), (b, a)\} \cup \{(a, i) : 1 \leq i \leq n\} \cup \{(b, i) : 1 \leq i \leq n\}$, as illustrated in Fig. 8. The n-downlinked mutual dyad is obtained by reversing all the arcs (cf. Fig. 7). From the lens of path homology, the number of uplinks (resp. downlinks) contributes robustness to an upward (resp. downward) flow. This intuition is formalized as:

Proposition 1 *Let $n \in \mathbb{Z}_{>0}$. Then $\tilde{\beta}_2(W_n) = n - 1$, and $\tilde{\beta}_p(W_n) = 0$ for all $p \geq 0, p \neq 2$. The n-downlinked mutual dyad has the same Betti numbers.*

Proof From Grigor'yan et al. (2012) we know that β_0 counts the connected components of the underlying undirected graph, so $\beta_0 = 1$ and $\tilde{\beta}_0 = 0$.
Now suppose $p = 1$. We have $\partial_{[1]}(e_{(a,b)} + e_{(b,a)}) = 0$, but also $\partial_{[2]}(e_{(a,b,1)} + e_{(b,a,1)}) = e_{(a,b)} + e_{(b,a)}$, and so $e_{(a,b)} + e_{(b,a)}$ cannot contribute to $\tilde{\beta}_1$. Similarly terms of the form $e_{(a,b)} + e_{(b,i)} - e_{(a,i)} \in Z_1$, $1 \leq i \leq n$, cannot contribute to $\tilde{\beta}_1$ as they belong to B_1, being the images of $e_{(a,b,i)}$ for $1 \leq i \leq n$.

Next suppose $p \geq 3$: then $\Omega_p = \{0\}$, since all the 3-paths have boundaries with non-allowed paths, and taking linear combinations does not eliminate these non-allowed paths.

Finally we deal with the case $p = 2$. Let $1 \leq i \neq j \leq n$. Then $\partial_{[2]}(e_{(a,b,i)} + e_{(b,a,i)} - e_{(a,b,j)} - e_{(b,a,j)}) = e_{(a,b)} + e_{(b,a)} - e_{(a,b)} - e_{(b,a)} = 0$.
Therefore all 2-paths of the form $e_{(a,b,i)} + e_{(b,a,i)} - e_{(a,b,j)} - e_{(b,a,j)}$ belong to Z_2, but not to B_2 as Ω_3 is trivial. Some linear algebra shows that the collection $\{e_{(a,b,1)} + e_{(b,a,1)} - e_{(a,b,j)} - e_{(b,a,j)} : 2 \leq j \leq n\}$ is a basis for Z_2. It follows that $\tilde{\beta}_2 = n - 1$.

To conclude the proof, we note that the above arguments hold for the downlinked mutual dyad by replacing terms of the form $e_{(a,b,i)}$ with $e_{(i,a,b)}$. □

Algorithm

To compute non-regular path homology, one needs to produce the requisite paths and perform the necessary linear algebra. However, performing this computation efficiently is a challenge that is addressed by our implementation, available at Yutin (2020). Our method introduces some nuances and is apparently among the first for dimension > 1,[4] so we outline our approach here.

First, for efficiency we remove nonbranching limbs (i.e., chains of vertices of total degree 2 that terminate in leaves of degree 1), since these do not affect homology by Theorem 5.1 of Grigor'yan et al. (2012). We also exploit Proposition 3.25 of Grigor'yan et al. (2012), by decomposing the graph into weak components before computing homology componentwise. For each component D, we extend an order on vertices $V(D)$ to order paths lexicographically. We construct $\mathcal{A}_p(D)$ for $0 \leq p \leq p_{\max}$ inductively. Starting from $\mathcal{A}_0(D) = V(D)$, we construct $\mathcal{A}_p(D)$ by appending every vertex that has an arc from the terminal vertex of a path in $\mathcal{A}_{p-1}(D)$. The paths are produced in lexicographical order for each p.

Next, using a radix-$|V(D)|$ expansion, we compute the indices that specify the inclusion $\mathcal{A}_p \hookrightarrow V^{p+1}$ under lexicographical ordering. We then construct (in the standard basis) the matrix representation $\partial_{[p,\mathcal{A}]}$ of the restriction of $\partial_{[p]}$ to $\mathbb{F}^{\mathcal{A}_p}$. Let $\nabla_{[p,\mathcal{A}]}$ denote the projection of $\partial_{[p,\mathcal{A}]}$ onto $\mathbb{F}^{V^p \setminus \mathcal{A}_{p-1}}$ (i.e., the matrix formed by removing rows of $\partial_{[p,\mathcal{A}]}$ that correspond to elements of $\mathcal{A}_{p-1}$), and let $\Delta_{[p,\mathcal{A}]}$ denote the projection onto $\mathbb{F}^{\mathcal{A}_{p-1}}$. The kernel of $\nabla_{[p,\mathcal{A}]}$ is Ω_p. To produce Ω_p as efficiently as possible, we remove rows of $\nabla_{[p,\mathcal{A}]}$ that are identically zero before computing this kernel. Once we have produced a matrix representation $\Omega_{[p,\mathcal{A}]}$ for the kernel above, we get the chain boundary operator $\partial_p = \Omega^{-1}_{[p-1,\mathcal{A}]}\Delta_{[p,\mathcal{A}]}\Omega_{[p,\mathcal{A}]}$ (i.e., the projection of $\partial_{[p,\mathcal{A}]}$ onto $\mathbb{F}^{\mathcal{A}_{p-1}}$, projected onto and restricted to the invariant space).

Finally, we compute the homology of this chain complex, using the rank-nullity theorem and a singular value decomposition to compute the matrix ranks and in turn obtain the Betti numbers. We compute representatives for homology groups (without assurances that the representatives are the best possible) as cokernels of $[\ker \partial_p]^T \partial_{p+1}$. To find torsion over $\mathbb{Z}$, we computed the Smith normal form of boundary matrices as in usual practice (cf. Fig. 5).[5]

In the worst case, our algorithm requires $O(|V|^{2p+1})$ memory (to store the boundary matrices) and $O(|V|^{(p+1)\omega})$ runtime (to compute matrix ranks), where the matrix multiplication complexity exponent ω is 3 in practice. In practice, these estimates are overly pessimistic (most digraphs analyzed in practice are very different from complete digraphs), though the memory requirements are still exponential in p unless the digraph being analyzed is acyclic, in which case the exponential relationship only holds up to the

[4] Though (Shajii 2013; Slawinski 2013) significantly predate our implementation, we were unaware of these until after we had produced our implementation, and we could not find any published work drawing on them.

[5] To optimize further, we could preprocess the digraph in accord with Theorem 5.7 of Grigor'yan et al. (2012), though this would raise its own issues in low dimension. Instead of performing a singular value decomposition on rather large boundary matrices to compute their ranks, we could recursively construct invariant spaces — each sub-path of an invariant path is itself an invariant path (since we only consider loopless digraphs). A simple approach in this vein might be to check every pair of paths in dimension p against every vertex to see where we can append 'triangles' and 'squares' in the sense of Grigor'yan et al. (2012). While promising, this approach generates too many paths, and reducing to a basis is computationally nontrivial. For low dimensions, we could also directly compute Betti numbers from the digraph itself (cf. Proposition 3.24 of Grigor'yan et al. (2012)).

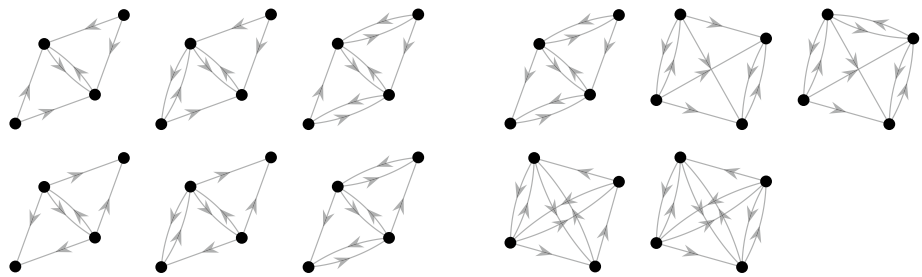

Fig. 3 (L) $\tilde{\beta}_2 > 0$ for these six (of 218 total) digraphs on four vertices. In each case it turns out that $\tilde{\beta}_p = \delta_{p,2}$. As a "suspension" of the 2-cycle, the digraph in the upper left can be thought of as an analogue of a homology 2-sphere obtained by gluing two cones along a common equator. (R) $\tilde{\beta}_3 > 0$ for these five digraphs on four vertices. In each case it turns out that $\tilde{\beta}_p = \delta_{p,3}$. Continuing the geometrical-topological analogy from before, the upper middle digraph in the right panel is akin to that in the upper left of the left panel with its "poles" glued together by a circular path in another dimension, thus giving rise to homology in dimension 3

length of the longest path. Nevertheless, it is still essential to limit computations to fairly small p and use natural filtrations (e.g., time, weight, etc.) to isolate portions of ambient digraphs. While it would be ideal in this vein to compute *persistent* path homology (Chowdhury and Mémoli 2018; Lin et al. 2019; Dey et al. 2020), no practical algorithms for this are known in dimension > 1.

Phenomenology of path homology for small digraphs

Small digraphs.

We show digraphs on four vertices with $\tilde{\beta}_p > 0$ for $p > 1$ in Fig. 3. Surprisingly, four vertices are enough for nontrivial homology to occur even in dimension three. Meanwhile, in Fig. 4, the left panel shows directed acyclic graphs (DAGs) on six vertices with $\tilde{\beta}_2 > 0$. Examining these DAGs led us to formulate and prove a conjecture about the path homology of DAGs that model the connectivity of deep feedforward neural networks (Chowdhury et al. 2019) and characterize temporal networks in the representation of Pósfai and Hövel (2014), as detailed in § 6. Finally, the right panel of Fig. 4 shows undirected graphs (which we treat as digraphs with arcs in both directions) on six vertices with $\tilde{\beta}_2 > 0$. This highlights that path homology is also relevant for the analysis of undirected graphs.

Torsion.

Though nominally defined over fields, path homology makes sense over rings, e.g. $\mathbb{Z}$, with scarcely any modifications required. This is a more powerful invariant, as it gives rise to torsion. Yutin (2019) was able to identify the digraphs in Fig. 5 by sampling Erdős-Rényi digraphs and carefully decomposing an instance with nonzero torsion.

These digraphs are the smallest members of a family that we conjecture always exhibits torsion: larger members can be formed by taking a longer central unidirected closed path and linking each of its vertices to one of two external "polar" vertices (in an alternating fashion). Specifically, we conjecture that digraphs in this family with central paths of length $2n$ have torsion subgroups $\mathbb{Z}/n\mathbb{Z}$ in $\tilde{H}_1$. (We have computationally verified this conjecture for $n \leq 8$.) These digraphs seem to be analogues of so-called *lens spaces* that are formed by gluing two tori together with a twist, and which

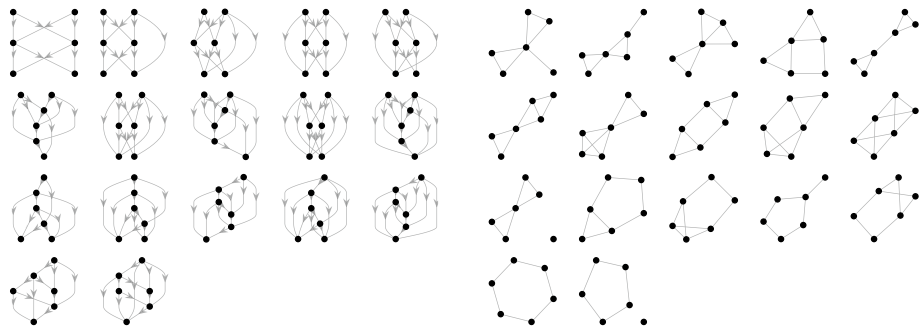

Fig. 4 (L) $\tilde{\beta}_2 > 0$ for these 17 (of 5984 total) DAGs on six vertices. Note that the DAG in the upper left is a ubiquitous subgraph of the others; the path homology of this graph and others like it is analyzed in Chowdhury et al. (2019). (R) $\tilde{\beta}_2 > 0$ for these 17 (of 156 total) undirected graphs on six vertices. The graphs in the first two rows all have $\tilde{\beta}_\bullet = (0, 0, 1, 0, \ldots)$, and the remaining seven graphs (from left to right, top to bottom) respectively have $\tilde{\beta}_\bullet = (1, 0, 1, 0, \ldots)$; $\tilde{\beta}_\bullet = (0, 1, 1, 0, \ldots)$; $\tilde{\beta}_\bullet = (0, 1, 1, 0, \ldots)$; $\tilde{\beta}_\bullet = (0, 2, 1, 0, \ldots)$; $\tilde{\beta}_\bullet = (0, 2, 1, 0, \ldots)$; $\tilde{\beta}_\bullet = (0, 2, 1, 0, \ldots)$, and $\tilde{\beta}_\bullet = (1, 2, 1, 0, \ldots)$. The common "bow tie" motif here appears to be the cause for emergence of 2-homology in transportation networks as capacities are filtered (this will be elaborated on in future work); meanwhile, polygons with ≥ 5 sides have too many sides for paths in opposing directions to "destructively interfere." That is, although these graphs are undirected, the directed paths of length 4 through them exhibit more coherence than in other graphs of the same size

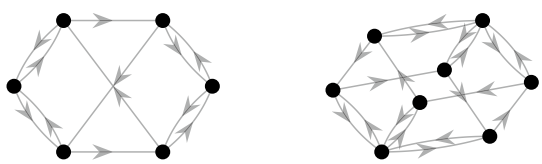

Fig. 5 These two digraphs have torsion subgroups of $\mathbb{Z}/2\mathbb{Z}$ (left) and $\mathbb{Z}/3\mathbb{Z}$ (right), i.e., their homology over $\mathbb{Z}$ contains these finite groups as a direct summand as described in §2

are archetypal examples of spaces with torsion. While we do not pursue this analogy further in the current work, we note that such lens spaces have recently been used for nonlinear topological dimension reduction (Polanco and Perea 2019).

Erdős-Rényi random graphs.

In Fig. 6 we show empirical distributions for Betti numbers of Erdős-Rényi random graphs (Frieze and Karoński 2016) on four nodes. Further knowledge of these distributions for different numbers of nodes would provide a useful method for testing if a stochastic digraph generating process could be described via an Erdős-Rényi model.

Applications to window-aggregated temporal networks

We analyze three *temporal networks* (Holme 2015; Masuda and Lambiotte 2020): MathOverflow, an email network, and activity on a Facebook group. These are all *directed contact networks* (DCNs) in the sense of Cybenko and Huntsman (2019): i.e., each network event ("contact") is specified by a source, a target, and a time. Formally, a DCN with (spatial) vertex set $V \equiv [n] \equiv \{1, \ldots, n\}$ is a finite nonempty set $\mathcal{C}$ for which each *contact* $c \in \mathcal{C}$ corresponds to a unique triple encoding a source s, a target t, and a time τ in the form $(s(c), t(c), \tau(c)) \in [n] \times [n] \times \mathbb{R}$ with $s(c) \neq t(c)$.

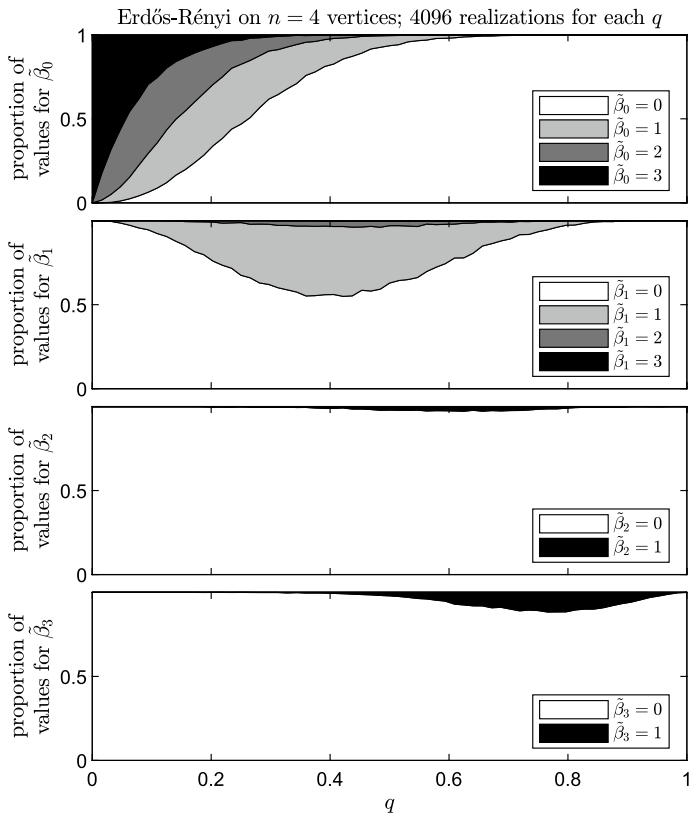

Fig. 6 Empirical distributions of $\tilde{\beta}_p(D_{4,q})$, where $D_{n,q}$ is the Erdős-Rényi random digraph on n vertices with probability q of a given arc occuring

In this section, we represent the three DCNs above by aggregating the activity over a time window into a single digraph which then varies as the window moves. In later sections, we will examine the same three networks, but via different network representations. In both this and later sections, we exhibit high-order interactions identified using path homology that are respectively indicators of dilution, recurring motifs, and concentration within network behavior that is preserved across the various representations.

MathOverflow

The answer-to-question portion of the `sx-mathoverflow` DCN available at Leskovec and Krevl (2014) provides our first illustration of the ability of path homology to identify structurally relevant subgraphs. This network has 21688 vertices and 107581 directed temporal contacts, spanning 2350 days from 29 Sep 2009 to 6 March 2016; it was previously analyzed in Montoya et al. (2013), and Tausczik et al. (2014) contains a discussion of question/answer phenomenology on MathOverflow.

In our analysis, we considered contacts within a time window of 24 hours, moving every eight hours. We aggregated contacts in a given window into a static digraph and computed the first three Betti numbers.[6]

[6] NB. Our path homology code removes any loops from digraphs.

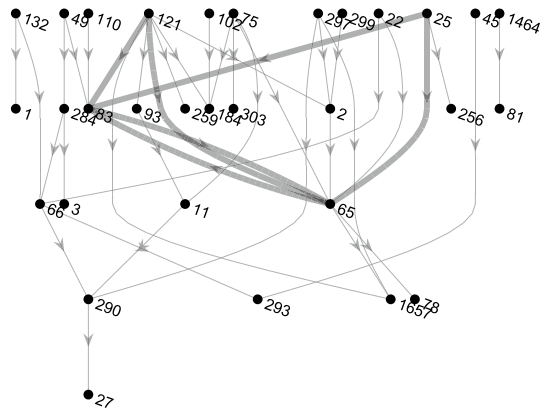

Fig. 7 Digraph of MathOverflow activity over a 24-hour period during 13-14 Oct 2009. User IDs serve as vertex labels, and arcs are directed from answerer to questioner (any parallel arcs are merged). We highlight arcs participating in a 2-homology representative. For completeness, the participating (arc, question) pairs are ((25, 65), 437), ((25, 83), 451), ((65, 83), 446), ((83, 65), 437), ((121, 65), 433), and ((121, 83), 446) as well as ((121, 83), 451). Three of the four users {25, 65, 83, 121} share the same first subject tag (i.e., area of interest/expertise) at time of writing; all four share the same second subject tag

Only two (immediately adjacent and overlapping) windows, over 13-14 Oct 2009, had $\beta_2 > 0$. Inspecting the homology representatives revealed an underlying motif, viz., the 2-downlinked mutual dyad (cf. Sec. 2.1). The particular questions and answers involved are shown in Fig. 7, which also highlights the dyad.

This (effectively) single occurrence of 2-homology happened very early in the history of MathOverflow–in fact, just two weeks after the website launched. As MathOverflow evolved, interactions on it also diluted. For example, most of the first 200 users asked and answered many fewer questions over time, while the overall size of and activity on MathOverflow grew much larger (not shown here). One consequence of this dilution of activity is that opportunities for tightly coupled patterns of questions and answers to occur diminished, with 2-homology serving as an indicator of this phenomenon.

An email network

The presence of linked dyads is actually ubiquitous and generalized in email networks, because of well-known behavior common to the medium (e.g., multiple people sending to a common mailing list). We isolated this behavior in our analysis of the `email-Eu-core-temporal` network available at Leskovec and Krevl (2014). This network has 986 vertices and 332334 directed temporal contacts; it spans 804 days of activity.

We considered contacts within a window of the most recent 100 contacts (emails), moving every 50 contacts. As before, we aggregated contacts in a given window into a static digraph and computed the first three Betti numbers. Many windows exhibited high values of β_2 that we traced to occurrences of the n-uplinked mutual dyad (cf. Sec. 2.1) motif shown in Fig. 8. The underlying dynamics is common: two people ("Alice" and "Bob") both send email to the same wide distribution and to each other.

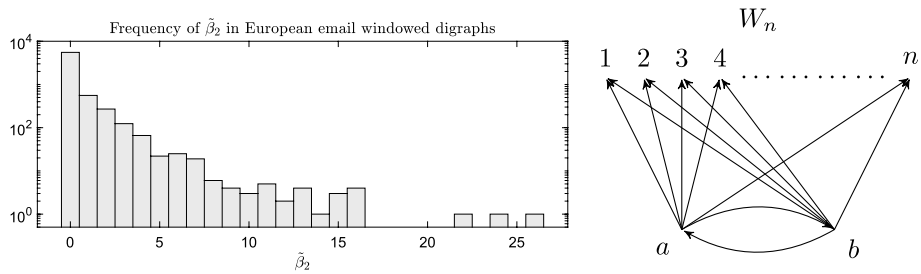

Fig. 8 (L) The distribution of $\tilde{\beta}_2$ for windowed digraphs produced from the email network. (R) The digraph W_n depicted here has $\tilde{\beta}_2(W_n) = n - 1$; it causes high values of $\tilde{\beta}_2$ in windowed digraphs produced from the email network

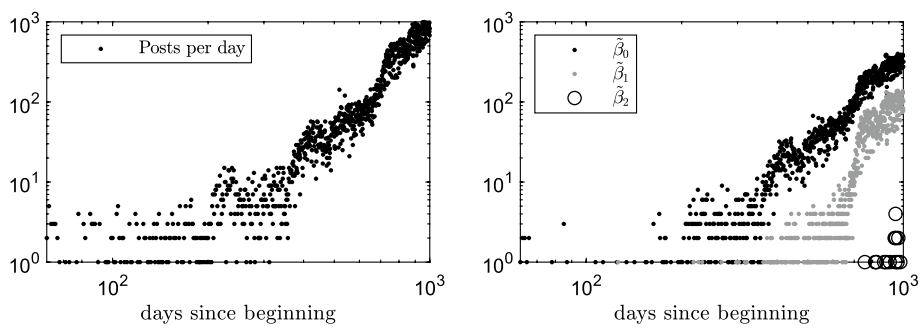

Fig. 9 (L) Daily Facebook group posts. (R) Betti numbers of daily digraphs. As activity increases, so do topological features in dimensions 0 through 2

A Facebook group

As a final example using the window-aggregated DCN representation, we considered the first 1000 days of activity on a Facebook group (Viswanath et al. 2009; Kunegis 2013) starting from 14 September 2004 and ending on 11 June 2007. We aggregated this DCN (13295 vertices; 187750 contacts) into daily digraphs with no sliding windows because it has a daily lull with virtually no activity. Fig. 9 shows the number of posts per day and the first three Betti numbers. Besides an obvious correlation between network activity and $\tilde{\beta}_0$, it is also evident that progressively more and higher-dimensional homology classes appear over time. This emergent higher-order network structure indicates concentration of activity. Fig. 10 shows the first daily digraph with $\tilde{\beta}_2 > 0$. Because this sort of concentration of activity is tied to just a few specific loci (i.e., appropriate homology representatives in the corresponding time windows), statistical or other straightforwardly quantitative network measures are unlikely to capture it *de novo*. Though it is probably the case that such measures can be reverse-engineered, this misses the point of a qualitative measure: i.e., robustness in the face of network perturbations or even (up to a point) differing representations.

Layered representations

A "layered" representation of DCNs that naturally leads to rich path homology characterizations is that of Pósfai and Hövel (2014). The essential idea is to represent a contact of the form (s, t, τ) as an arc from (s, j) to $(t, j + 1)$ where time is discretized into bins

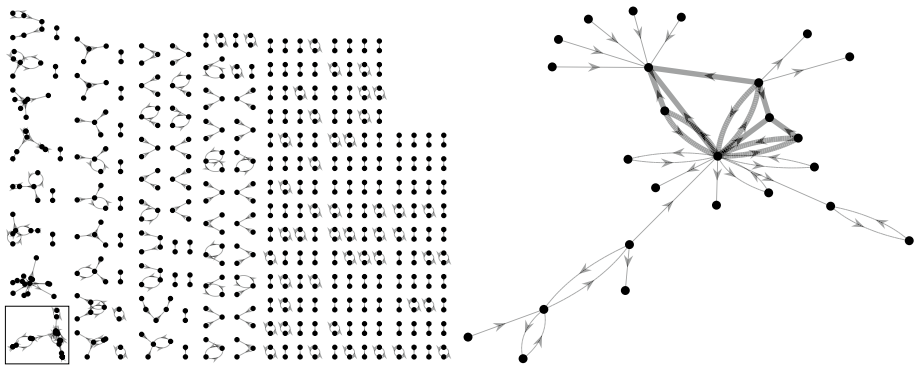

Fig. 10 (L) First daily digraph with $\tilde{\beta}_2 > 0$: day 756. The only weak component with $\tilde{\beta}_2 > 0$ is indicated with a box. (R) Detail with (different graph layout and) arcs representing $\tilde{H}_2$ highlighted. This homology representative is highly symmetrical

such that τ is in the jth bin. That is, contacts are modeled as going *forward* in time from the current time bin to the next one.

This representation casts DCNs in a structural light virtually identical to that of weight-filtered multilayer perceptrons (MLPs) as discussed in Chowdhury et al. (2019), where it was shown that homology generically occurs in dimension up to $L - 1$, where L indicates the number of time bins or layers being considered at a given time. However, while trained weight-filtered MLPs exhibit high-dimensional path homology generically because neural activations propagate across layers by design, the equivalent phenomenology in DCNs signals potential propagation of information (or whatever the DCN models) that is highly correlated across time windows in a way that there is no *a priori* reason to expect.

Tellingly, the path homology of DCNs in this layered representation again indicates gross dynamics of activity dilution (for the MathOverflow DCN); time-invariance (for the email DCN); or concentration (for the Facebook DCN). High-dimensional homology occurs (if at all) only in rare episodes where network participants engage in highly correlated activity that has structural significance, e.g. co-clustering of question/answer behavior for MathOverflow; (presumably) organizational cliques for email; and joint self- and cross-posting for Facebook, respectively.

In detail, for each DCN in the previous section, we employ a time discretization into bins of duration δt, and a sliding window of duration $d \cdot \delta t$, where d indicates the top dimension for which we compute $\tilde{\beta}_d$ and the window slides by δt. This alignment of dimension and temporal parameters ensures that we compute precisely the homologies which might be nontrivial. Here we use the preceding discussion on path homologies of MLPs, which suggests that for d layers, as in the sliding window of duration $d \cdot \delta t$, we expect only trivial path homology in dimensions beyond d.

For the MathOverflow network (see Figs. 11 and 12), there is no evidence of 2-homology for reasonable choices of time discretization and window duration, so we set $\delta t = 24$ hr and $d = 1$. Here $\tilde{\beta}_1$ is nonzero only transiently, shortly after the network began, in line with the general thrust of activity "dilution" mentioned previously. The absence of 2-homology can be attributed to an asymmetry between questioners and

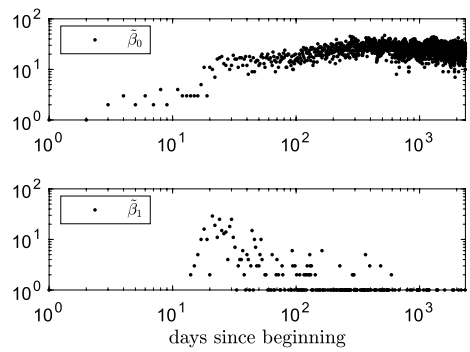

Fig. 11 Betti numbers of the layered representation of the MathOverflow DCN

answerers: in any given window, these two sets do not overlap enough for the layers of the digraph representation to produce MLP-like subgraphs.

For the email network, we set $\delta t = 1$ hr and $d = 2$. Here the network behavior is roughly time invariant (apart from a prolonged lull towards the end of the data) and several windows give rise to 2-homology (Figs. 13 and 14). This is apparently due to shared to/cc lists in emails that are related in subject matter and localized in time.

For the Facebook network, we set $\delta t = 24$ hr and $d = 2$. Here $\tilde{\beta}_2$ is nonzero only once, when two users self-post and interact with each other. Meanwhile, the increase of $\tilde{\beta}_0$ and $\tilde{\beta}_1$ (cf. Fig. 15) over time is consistent with the general thrust of activity "concentration" mentioned previously.

Temporal digraph representations

Yet another representation of DCNs is that of Cybenko and Huntsman (2019). Here, the entire network is losslessly encoded into a *temporal digraph* in which arcs are all either "spatial" (i.e., between vertices in the aggregated digraph) or "temporal" (i.e., connecting a vertex at one time to itself at a later time). A temporal digraph and the associated notion of a temporally coherent path are indicated in Fig. 16.

Our investigations suggest that temporal digraphs are uninteresting from the point of view of higher homology.

Conjecture 1 The temporal digraph of a DCN has $\tilde{\beta}_p = 0$ for $p > 1$.

The idea of the conjecture (which is based more on intuition and extensive experimentation than exhaustive computation *per se*) is that there are no "diagonal" connections from a vertex at one time to a different vertex at a different time, which in turn constrains the algebra tightly enough that there are no linear combinations of p-paths of the kind required for nontrivial phenomenology with $p \geq 2$. For example, none of the digraphs in Fig. 3 can be subgraphs of the temporal digraph of a DCN, so any production of homology for $p \geq 2$ must involve a subgraph on > 4 vertices. On the other hand, for $p = 2$, the participating paths must be of length 3, and thus cannot involve more than four vertices each. In other words, higher homology would have to arise in an intrinsically distributed way.

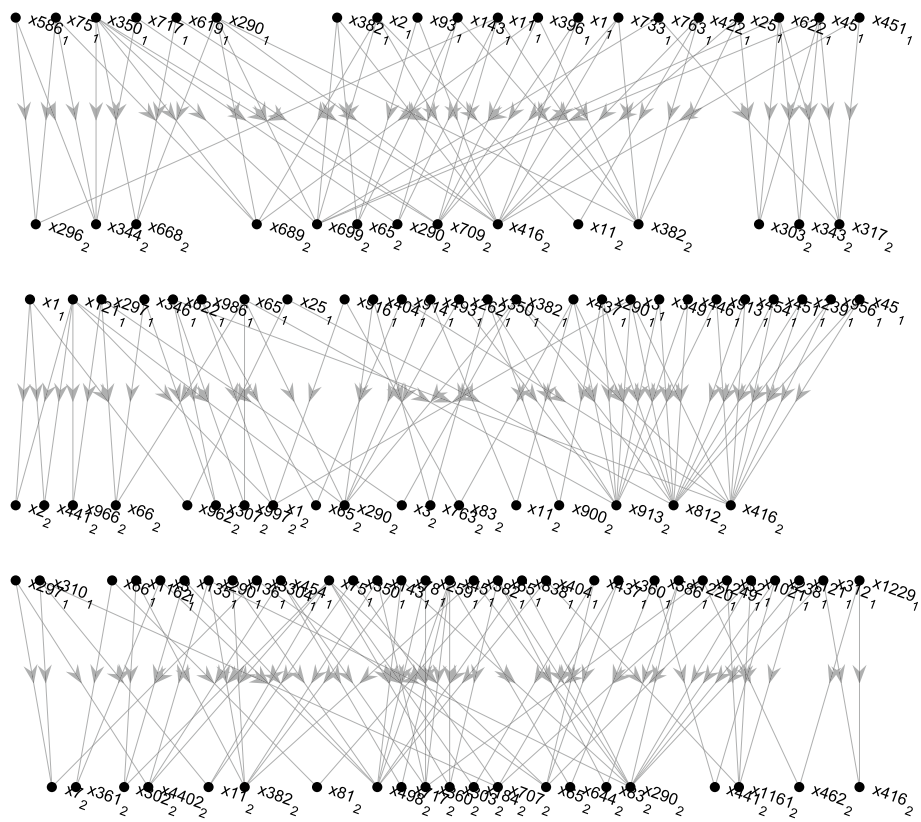

Fig. 12 In the layered representation of the MathOverflow DCN, there are three windows with $\tilde{\beta}_1 > 20$. Here we depict the weak components contributing nontrivial homology in dimension one for each of these windows. Node labels indicate user IDs; subscripts indicate bins within the window. Essentially, the dynamics here are that groups of users answer multiple questions within a window in a way that overlaps. As the network matured and diluted, this tightly coupled behavior disappeared almost entirely

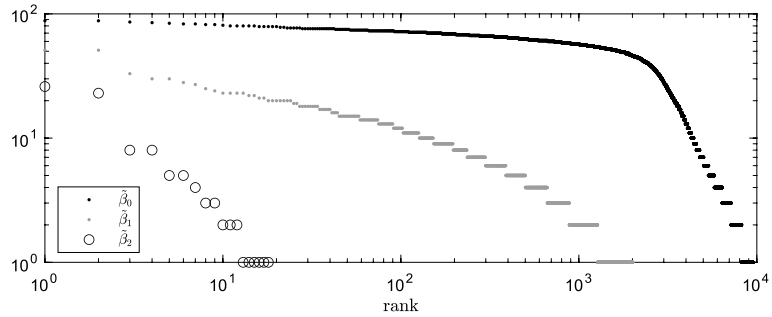

Fig. 13 Sorted Betti numbers of the layered representation of the email DCN

Although Conjecture 1 was formulated based on small, synthetically generated temporal digraphs, it was borne out in analyses of the same data sets–and using the same time windows–as in earlier sections. While at some points the resulting temporal digraphs were large enough to preclude computing β_2 due to the size of the boundary matrices involved, we always observed $\beta_2 = 0$ for temporal digraphs of DCNs. These results are illustrated in Figs. 17 and 18.

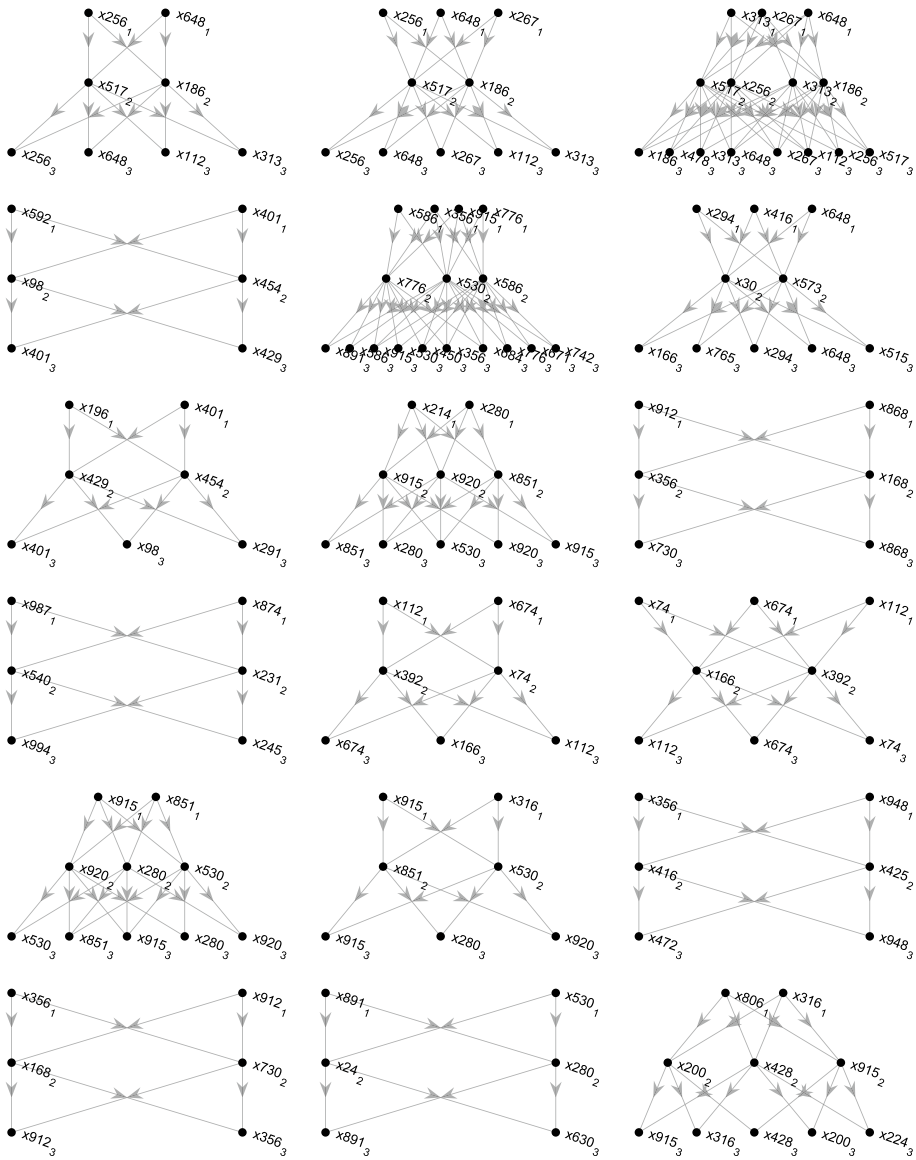

Fig. 14 As Fig. 13 indicates, there are 18 windows in the layered representation of the email DCN with $\tilde{\beta}_2 > 0$. Here, we show for each instance the corresponding subgraphs on vertices that participate in 2-homology representatives (it turns out in these particular instances that these subgraphs all have only a single weak component). These subgraphs are all very similar to the fully-connected deep feedforward networks (the difference here being that some of the layers are not fully connected, e.g. the graph in row 2, column 2) whose homology was analytically characterized in Chowdhury et al. (2019), and more generally are of the sort that would be encountered in weight-induced filtrations of multilayer perceptrons

We complement Conjecture 1 with a computationally supported conjecture about 1-homology in temporal digraphs with two (spatial) vertices.

Conjecture 2 Let $\mathcal{C}$ be a DCN with two vertices and contacts at times in $[T]$ for $T \in \{3, 4, ...\}$. The maximum possible value of $\tilde{\beta}_1$ for the corresponding temporal digraph is $T - 1$, achieved for the eight DCNs with contacts in alternating directions and possibly both directions at times 1 and T. If there are no contact pairs of the form $(1, 2, t)$ and

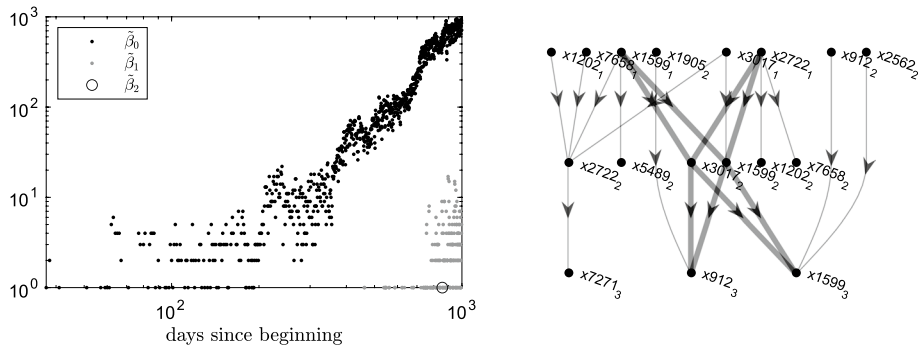

Fig. 15 (L) Betti numbers of the layered representation of the Facebook DCN. (R) The weak component in the window contributing to 2-homology, with the homology representative highlighted. Note that two users are posting to themselves and each other

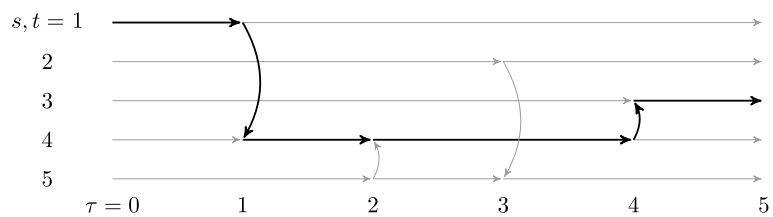

Fig. 16 Temporal digraph of the DCN $\mathcal{C} := \{(1,4,\tau_1),(5,4,\tau_2),(2,5,\tau_3),(4,3,\tau_4)\}$ with $\tau_1 < \tau_2 < \tau_3 < \tau_4$. We indicate a temporally coherent path from $\mathcal{C}$-vertices 1 to 3 (with bold versus gray arrows). By comparison, there is no temporally coherent path from 2 to 3. Temporal (resp. spatial) arcs are horizontal (resp. vertical); temporal fibers are vertices along horizontal paths

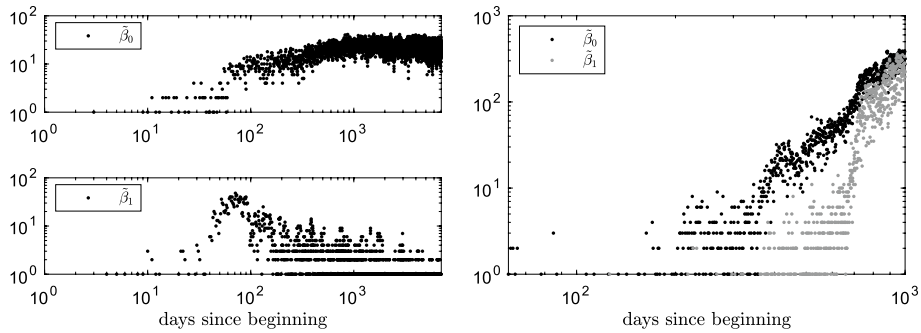

Fig. 17 (L) Betti numbers of the temporal digraph representation of the (L) MathOverflow and (R) Facebook DCNs. In the former case, we used a window of eight hours (this is different than in other representations, and was done for the sake of scaling), whereas in the latter case we used a window of 24 hours, as in other representations. Note that in the latter case β_0 is also the same as for the MLP representation. For some later times with both networks, β_2 could not be computed due to memory constraints, but was always zero when computed, including near the beginning of the MathOverflow DCN. Recomputing the Betti numbers for $p < 2$ required just a few minutes

$(2, 1, t)$ for $t \in [T]$, then $\tilde{\beta}_1$ (for the temporal digraph representation) counts the number of times that contacts change direction.

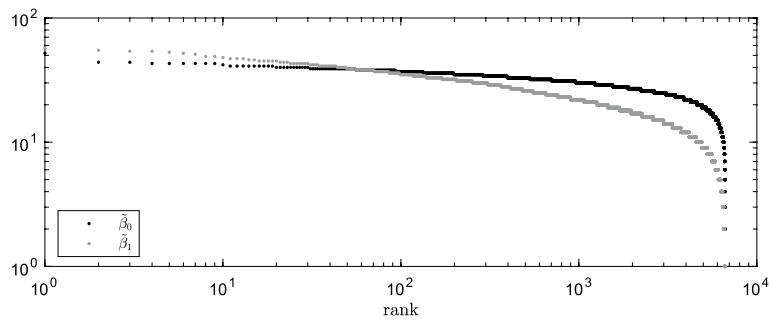

Fig. 18 Sorted Betti numbers of the temporal digraph representation of the email DCN. β_2 was computed for all time windows but always equaled zero

The idea behind this conjecture is that "flips break squares" (where "squares" are in the sense of Grigor'yan et al. (2012)) and contribute to $\tilde{\beta}_1$, whereas instantaneous back-and-forth contacts other than in the first and last instances prevent such contributions by adding enough terms to cause algebraic cancellations.

It might be tempting to consider a slightly different temporal network representation that allows "diagonal" arcs of the sort that seem necessary to support nontrivial path homology for $p > 1$. For example, we might replace a "temporal" arc followed by a "spatial" arc with a single "diagonal" arc. Such a representation indeed gives rise to nontrivial path homology for $p > 1$. However, this representation is not robust and/or meaningful unless the timestamps can actually differentiate between back-and-forth contacts that are merely close together in time versus exactly simultaneous, since these two cases can give different homologies. Meanwhile, if exact simultaneity is possible, the temporal network must actually inhabit discrete (if granular) time, and in this event the temporal digraph representation is essentially the same as the layered representation of $\mathbb{S}$, albeit with self-transitions automatically included. Taking the other observations of this section into account, we therefore see that the temporal digraph representation and its siblings either do not yield substantive insight, or require still further nuance to apply fruitfully.

Comparison with other activity measures

Thus far, we have performed extensive analysis on the dependency of path homology on the chosen representation of a temporal network. For completeness, we now compare path homology against two popular network measures—density and clustering coefficient—to show that path homology captures essentially different features of network data from these classical measures.

Recall that the density of a network describes the ratio of the number of true connections to the number of possible connections. For a digraph $D = (V, A)$, the density is given by $\frac{|A|}{|V|\cdot(|V|-1|)}$. Also recall that the clustering coefficient measures how often the neighbors of an edge are themselves neighbors. For the directed setting, we appeal to the notion of clustering coefficient developed in [?]. Following [?], we temporarily write A to denote the adjacency matrix (instead of the set of arcs), d_j to denote total (i.e., in- plus out-) degree of vertex j, and $d_j^{\leftrightarrow}$ to denote the number of vertices that have arcs to and from j. Then the (directed) clustering coefficient is given by:

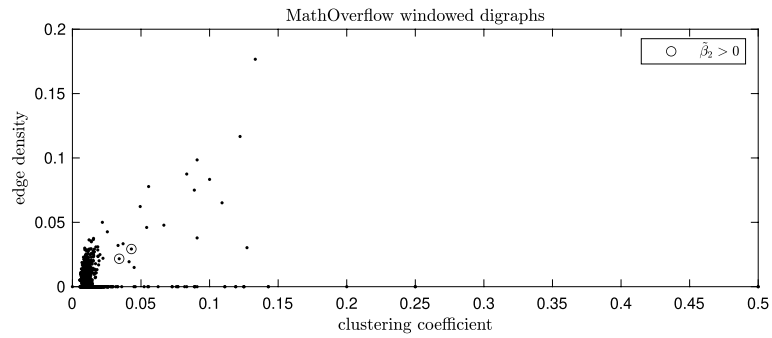

Fig. 19 Clustering coefficient (6) and density for the windowed representation of the MathOverflow temporal network, with instances of $\tilde{\beta}_2 > 0$ circled

$$C := \frac{1}{2|V|} \sum_{j \in [|V|]} \frac{((A + A^T)^3)_{jj}}{d_j(d_j - 1) - 2d_j^{\leftrightarrow}} \tag{6}$$

Overall, we find that density and clustering coefficient are generally uncorrelated with path homology Betti numbers (cf. Figs. 19, 21, and 22). There is one key exception to this rule (cf. Fig. 20): C and $\tilde{\beta}_2$ are fairly strongly correlated when mutual dyad motifs W_n are present either in large numbers or for n large. The reason for this is simple: C essentially measures the number of triangles, and in the event above, there are many triangles to be found.

However, C and $\tilde{\beta}_2$ are totally uncorrelated in layered representations of temporal networks for the equally simple reason that, by construction, the digraphs involved never have triangles. Meanwhile, in each temporal network considered here, the density and Betti numbers are also essentially uncorrelated (not shown here except for the MathOverflow network, which is broadly phenomenologically representative in this respect).

While a comprehensive comparison to network measures beyond density and clustering coefficient is out of the scope of the current paper, we make a few additional remarks about the general landscape of such statistical network measures before concluding this section. A reader more comfortable with, e.g., network centralities or motif analysis may wonder why we have not sought to compare them to the techniques of this paper. The reason is that centralities and motif analysis (along with density and clustering coefficient) are intrinsically quantitative measures focused on the statistical properties of nodes, edges, and subgraphs *in toto*, while path homology is an intrinsically qualitative measure focused on generic characteristics of individual subgraphs that may or may not be present in data. Meanwhile, clustering or community detection in digraphs also answers a fundamentally different question (viz., when/where are interconnected regions?) than we address (viz., when/where are certain directionally coordinated activities?).

Remarks

As a generalization of simplicial homology, path homology holds the promise to advance the state-of-the-art in applications of algebraic topology to network science. In this work, we have addressed two critical bottlenecks in the application of path

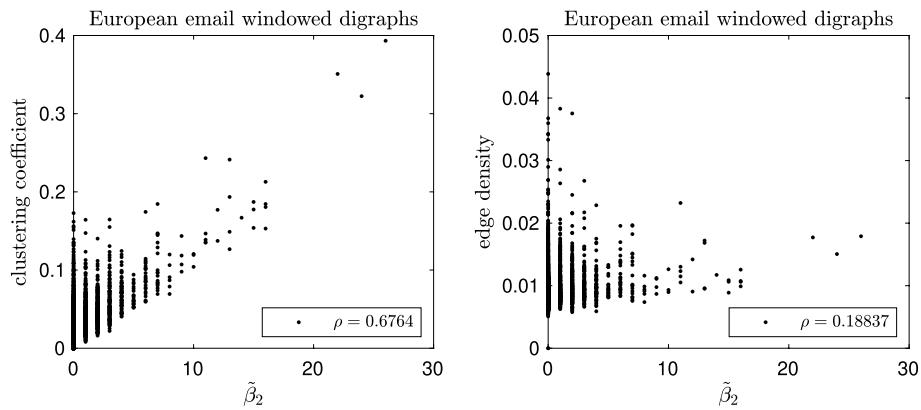

Fig. 20 $\tilde{\beta}_2$ versus (left) clustering coefficient (6) and (right) density versus for the windowed representation of the email temporal network. The strong correlation between $\tilde{\beta}_2$ and C is entirely due to the particular structure of the mutual dyad motifs that contribute to each activity measure. In other temporal network representations, or when these motifs are not dominant, these activity measures will not remain correlated (cf. Figs. 19 and 21)

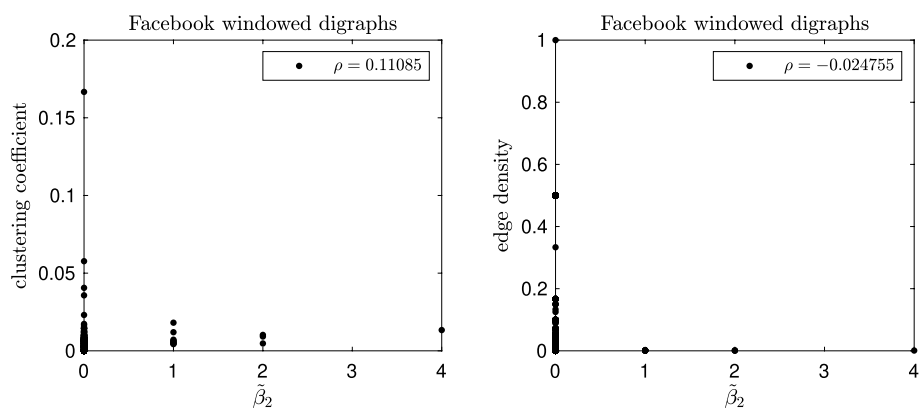

Fig. 21 $\tilde{\beta}_2$ versus (left) clustering coefficient (6) and (right) density for the windowed representation of the Facebook temporal network. The Betti number is essentially uncorrelated to either traditional activity measure

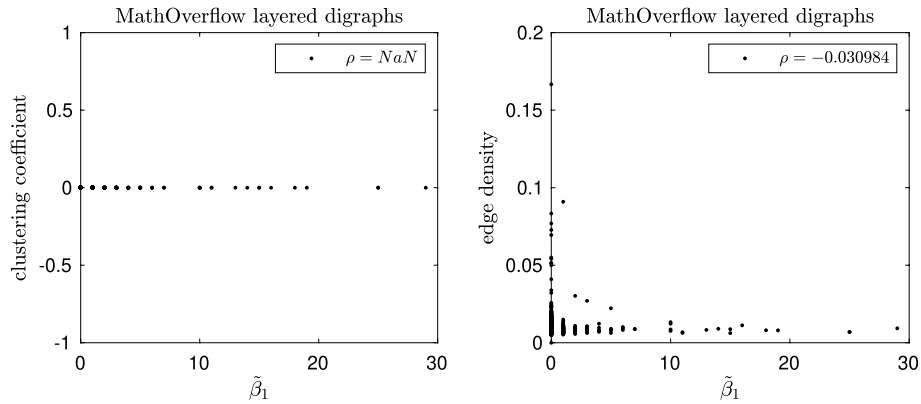

Fig. 22 $\tilde{\beta}_2$ versus (left) clustering coefficient (6) and (right) density for the layered representation of the MathOverflow temporal network. The clustering coefficient is always trivial for such representations and the density is essentially uncorrelated to $\tilde{\beta}_1$

homology: (1) fast computation, and (2) intuition for the phenomena that it captures. Specifically, we have developed an efficient implementation (Yutin 2020) and reported on an exhaustive library of computational examples that we have in turn used to explain the subgraphs in real-world networks that drive path homology activity. Looking to the future, our work could be a precursor to follow-up works that combine network dictionary learning (Lyu et al. 2020; Xu 2020; Vincent-Cuaz et al. 2021) with topological signatures (Gómez and Mémoli 2021).

Toward a comprehensive analysis of real-world directed contact networks (DCNs), we employed three different views of the data: window-aggregated, layered, and temporal digraph representations. From our experiments it is evident that the structure and even existence of path homology-carrying subgraphs can vary greatly depending on the network representation. However, one key takeaway is that the bulk dynamics indicated by the time series of Betti numbers is more robust to the particulars of the representation, particularly in the case of window-aggregated and layered representations. We also note that although the computational requirements for path homology scale exponentially with homological dimension, even the two-dimensional case can highlight salient network structure and behavior. By using the window-aggregated or layered representations, path homology can be successfully brought to bear in this regard, illuminating both subgraphs with nontrivial path homology as well as the temporal networks themselves.

We also observed that although temporal digraph representations have an obvious advantage of efficiently representing temporally coherent paths, currently they do not appear capable of yielding nontrivial path homology in higher dimensions. The possibility of developing a similar but robust and meaningful temporal network representation that can encode temporal coherence while also admitting nontrivial path homology in higher dimensions is an interesting avenue for future work. Overall, our findings suggest that different network representations of DCNs yield different perspectives of the data (see Singh et al. (2007); Lum et al. (2013) for related approaches), and considering families of path homology profiles built on top of a variety of network representations (cf. Chowdhury et al. (2020)) may provide the most comprehensive understanding of the data.

Acknowledgements
The authors thank Michael Robinson for many helpful discussions, and a reviewer for particularly careful reading and advice that improved the paper.

Authors' contributions
SC contributed to all sections of the manuscript; advised on experimental findings; and proved the result on the homology of the mutual dyads. SH designed and implemented an early version of the path homology algorithm, performed the analysis of path homology for small digraphs except for torsion digraphs; performed the data analyses; and organized the manuscript. MY designed and implemented the path homology algorithm presented here; identified torsion digraphs; and advised on the manuscript. All authors read and approved the final manuscript.

Funding
Portions of this material are based upon work partially supported by the Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) while SH and MY were at BAE Systems. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of DARPA or AFRL.

Availability of data and materials
Data, MATLAB scripts, and ancillary code sufficient to reproduce our results are available at https://github.com/SteveHuntsman/PathHomologyDataAndScripts and the path homology MATLAB code itself is available at https://github.com/SteveHuntsmanBAESystems/PerformantPathHomology.

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Stanford University, Stanford, USA. [2]Alexandria, Virginia, USA. [3]Fairfax, Virginia, USA.

Received: 31 March 2021 Accepted: 19 November 2021
Published online: 06 January 2022

References
Adams H, Tausz A, Vejdemo-Johansson M (2014) Javaplex: a research software package for persistent (co) homology. In: International congress on mathematical software. Springer, pp 129–136
Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G, Petri G (2020) Networks beyond pairwise interactions: structure and dynamics. Phys Rep
Benson AR, Gleich DF, Leskovec J (2016) Higher-order organization of complex networks. Science 353(6295):163–166
Bergomi MG, Ferri M, Tavaglione A (2020) Steady and ranging sets in graph persistence. arXiv preprint arXiv:2009.06897
Carlsson G (2009) Topology and data. Bull Am Math Soc 46(2):255–308
Chowdhury S, Clause N, Mémoli F, Sánchez JÁ, Wellner Z (2020) New families of stable simplicial filtration functors. Topol Appl 279:107254n
Chowdhury S, Mémoli F (2018) A functorial Dowker theorem and persistent homology of asymmetric networks. J Appl Comput Topol 2(1):115–175
Chowdhury S, Mémoli F (2018) Persistent path homology of directed networks. In: Proceedings of the 29th annual ACM-SIAM symposium on discrete algorithms. SIAM, pp 1152–1169
Chowdhury S, Huntsman S, Yutin M (2020) Path homology and temporal networks. In: International conference on complex networks and their applications. Springer, pp 639–650
Chowdhury S, Gebhart T, Huntsman S, Yutin M (2019) Path homologies of deep feedforward networks. In: 2019 18th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp 1077–1082
Cybenko G, Huntsman S (2019) Analytics for directed contact networks. Appl Netw Sci 4(1):1–21
Dey TK, Li T, Wang Y (2020) An efficient algorithm for 1-dimensional (persistent) path homology. In: 36th international symposium on computational geometry (SoCG 2020). Schloss Dagstuhl-Leibniz-Zentrum für Informatik
Ebert C, Cain J, Antoniol G, Counsell S, Laplante P (2016) Cyclomatic complexity. IEEE Softw 33(6):27–29
Edelsbrunner H, Harer J (2010) Computational topology: an introduction. American Mathematical Soc, Providence
Frieze A, Karoński M (2016) Introduction to random graphs. Cambridge University Press, Cambridge
Gebhart T, Funk RJ (2020) The emergence of higher-order structure in scientific and technological knowledge networks. arXiv preprint arXiv:2009.13620
Ghrist RW (2014) Elementary applied topology, vol 1. Createspace Seattle, Scotts Valley
Giusti C, Ghrist R, Bassett DS (2016) Two's company, three (or more) is a simplex. J Comput Neurosci 41(1):1–14
Grigor'yan A, Jimenez R, Muranov Y, Yau S-T (2018) On the path homology theory of digraphs and Eilenberg–Steenrod axioms. Homol Homot Appl 20(2):179–205
Grigor'yan A, Lin Y, Muranov Y, Yau S-T (2014) Homotopy theory for digraphs. Pure Appl Math Q 10(4):619–674
Grigor'yan A, Lin Y, Muranov Y, Yau S-T (2015) Cohomology of digraphs and (undirected) graphs. Asian J Math 19(5):887–932
Grigor'yan A, Muranov YV, Yau S-T et al (2014) Graphs associated with simplicial complexes. Homol Homot Appl 16(1):295–311
Grigor'yan A, Muranov Y, Yau S-T (2017) Homologies of digraphs and künneth formulas. Commun Anal Geom 25(5):969–1018
Grigor'yan A, Lin Y, Muranov Y, Yau S-T (2012) Homologies of path complexes and digraphs. arXiv preprint arXiv:1207.2834
Grigor'yan A, Muranov Y, Vershinin V, Yau S-T (2018) Path homology theory of multigraphs and quivers. In: Forum mathematicum, vol 30. De Gruyter, pp 1319–1337
Gómez M, Mémoli F (2021) Curvature sets over persistence diagrams. arXiv preprint arXiv:2103.04470
Hatcher A (2001) Algebraic topology
Holme P (2015) Modern temporal network theory: a colloquium. Eur Phys J B 88(9):1–30
Huntsman S (2020) Path homology as a stronger analogue of cyclomatic complexity. arXiv preprint arXiv:2003.00944
Kunegis J (2013) Konect: the koblenz network collection. In: Proceedings of the 22nd international conference on World Wide Web, pp 1343–1350
Lancichinetti A, Fortunato S, Kertész J (2009) Detecting the overlapping and hierarchical community structure in complex networks. New J Phys 11(3):033015
Leskovec J, Krevl A (2014) SNAP datasets: Stanford large network dataset collection
Lin Y, Ren S, Wang C, Wu J (2019) Weighted path homology of weighted digraphs and persistence. arXiv preprint arXiv:1910.09891
Lum PY, Singh G, Lehman A, Ishkanov T, Vejdemo-Johansson M, Alagappan M, Carlsson J, Carlsson G (2013) Extracting insights from the shape of complex data using topology. Sci Rep 3(1):1–8
Lütgehetmann D, Govc D, Smith JP, Levi R (2020) Computing persistent homology of directed flag complexes. Algorithms 13(1):19
Lyu H, Needell D, Balzano L (2020) Online matrix factorization for Markovian data and applications to network dictionary learning. J Mach Learn Res 21(251):1–49

Masuda N, Lambiotte R (2020) Guide to temporal networks, vol 6. A World Scientific, Singapore
McCabe TJ (1976) A complexity measure. IEEE Trans Softw Eng 4:308–320
Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U (2002) Network motifs: simple building blocks of complex networks. Science 298(5594):824–827
Montoya LV, Ma A, Mondragón RJ (2013) Social achievement and centrality in mathoverflow. In: Complex networks IV. Springer, pp 27–38
Méndez D, Sánchez-García RJ (2020) A directed persistent homology theory for dissimilarity functions. arXiv preprint arXiv:2008.00711
Petri G, Scolamiero M, Donato I, Vaccarino F (2013) Topological strata of weighted complex networks. PLoS ONE 8(6):66506
Polanco L, Perea J (2019) Coordinatizing data with lens spaces and persistent cohomology. In: Proceedings of the 31st Canadian conference on computational geometry (CCCG), pp 49–57
Pósfai M, Hövel P (2014) Structural controllability of temporal networks. New J Phys 16(12):123055
Ravasz E, Barabási A-L (2003) Hierarchical organization in complex networks. Phys Rev E 67(2):026112
Reimann MW, Nolte M, Scolamiero M, Turner K, Perin R, Chindemi G, Dłotko P, Levi R, Hess K, Markram H (2017) Cliques of neurons bound into cavities provide a missing link between structure and function. Front Comput Neurosci 11:48
Shajii AR (2013) Digraph homology. https://github.com/arshajii/digraph-homology
Singh G, Mémoli F, Carlsson G (2007) Topological methods for the analysis of high dimensional data sets and 3d object recognition. In: Eurographics symposium on point-based graphics, vol 2
Sizemore AE, Phillips-Cremins JE, Ghrist R, Bassett DS (2019) The importance of the whole: topological data analysis for the network neuroscientist. Netw Neurosci 3(3):656–673
Slawinski M (2013) Digraph homology. https://github.com/gtownrocks/digraph_homology
Smoly IY, Lerman E, Ziv-Ukelson M, Yeger-Lotem E (2017) Motifnet: a web-server for network motif analysis. Bioinformatics 33(12):1907–1909
Tausczik YR, Kittur A, Kraut RE (2014) Collaborative problem solving: a study of mathoverflow. In: Proceedings of the 17th ACM conference on computer supported cooperative work & social computing, pp 355–367
Turner K (2019) Rips filtrations for quasimetric spaces and asymmetric functions with stability results. Algebraic Geometric Topol 19(3):1135–1170
Vincent-Cuaz C, Vayer T, Flamary R, Corneli M, Courty N (2021) Online graph dictionary learning. In: Meila M, Zhang T (eds) Proceedings of the 38th international conference on machine learning. Proceedings of machine learning research, vol 139. PMLR, pp 10564–10574. https://proceedings.mlr.press/v139/vincent-cuaz21a.html
Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in facebook. In: Proceedings of the 2nd ACM workshop on online social networks, pp 37–42
Xu H (2020) Gromov-Wasserstein factorization models for graph clustering. In: Proceedings of the AAAI conference on artificial intelligence, vol 34, pp 6478–6485
Yutin M (2020) Performant path homology. https://github.com/SteveHuntsmanBAESystems/PerformantPathHomology
Yutin M (2019) Personal communication

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.