



Consistent tie-strength labeling for multilayer strong triadic closure

Lutz Oettershagen¹ · Athanasios L. Konstantinidis² · Fariba Ranjbar³ · Giuseppe F. Italiano³

Received: 29 October 2025 / Accepted: 21 April 2026
© The Author(s) 2026

Abstract

Inferring tie strengths (*strong* vs. *weak*) is a core task in network analysis, often guided by the Strong Triadic Closure (STC) principle. In multilayer networks, such as social platforms or biological systems, applying STC independently to each layer can lead to inconsistent tie labels, undermining interpretations that rely on coherent relationship semantics across layers. We propose new formulations, multilayer STC and its extension STC+, which are axiomatically grounded and enforce cross-layer consistency. These problems are NP-hard; we present efficient 2- and 6-approximation algorithms alongside exact solutions. Experiments on real-world networks demonstrate that our methods produce consistent tie strength labelings with a transparent structural justification, significantly improving over the baselines.

Keywords Strong triadic closure · Multilayer networks · Tie strength

✉ Lutz Oettershagen
lutz.oettershagen@liverpool.ac.uk

Athanasios L. Konstantinidis
a.konstantinidis@uoi.gr

Fariba Ranjbar
fariba.ranjbar@luiss.it

Giuseppe F. Italiano
gitaliano@luiss.it

- ¹ University of Liverpool, Liverpool, UK
- ² University of Ioannina, Ioannina, Greece
- ³ Luiss University, Rome, Italy

1 Introduction

Inferring the strength of relationships between entities is a fundamental task in network analysis, with applications in social network mining and related interaction networks (Adriaens et al. 2020; Gilbert et al. 2009; Kahanda et al. 2009; Lü and Zhou 2010; Oettershagen et al. 2025; Rozenshtein et al. 2017; Stolz and Schlereth 2021). In many real-world networks, edges are not equally meaningful: some represent *strong* (close, persistent) interactions, while others capture incidental or *weak* relationships (Granovetter 1973). Distinguishing between *strong* and *weak* ties is therefore essential for understanding network structure and for supporting downstream tasks such as recommendation, community detection, or robustness analysis.

A prominent topology-based principle for tie strength inference is the *Strong Triadic Closure* (STC) property, introduced by Sintos et al. (2014). Informally, STC states that if a node has strong ties to two others, then those two nodes should be at least weakly connected. That is, if person *A* is strongly connected to *B*, and *B* is strongly connected to *C*, then *A* and *C* should know each other. This captures the intuition that strong relationships tend to close triangles. Figure 1 shows an example. A key advantage of the STC is that we are able to infer tie strength using only network topology, without requiring node or edge attributes, which are often unavailable due to data collection limitations or privacy concerns.

Therefore, Sintos et al. (2014) introduced optimization formulations of the STC problem that seek to label edges as *strong* or *weak*, using only the network structure, so as to satisfy the strong triadic closure constraints. In addition to the basic STC formulation, they proposed an additional variant, denoted STC+, in which new edges may be inserted to resolve violations of the STC, allowing a more flexible trade-off between edge insertions and weak tie assignments.

In this work, we introduce new definitions and algorithms to compute the STC and STC+ variants in multilayer networks. In the following we motivate this new direction and discuss the arising challenges as well as our contributions in solving them.

Multilayer networks. Many complex systems are more naturally modeled not as a single graph, but as a collection of graphs over the same set of nodes, each representing a different type of interaction. Such systems are commonly referred to as *multi-*

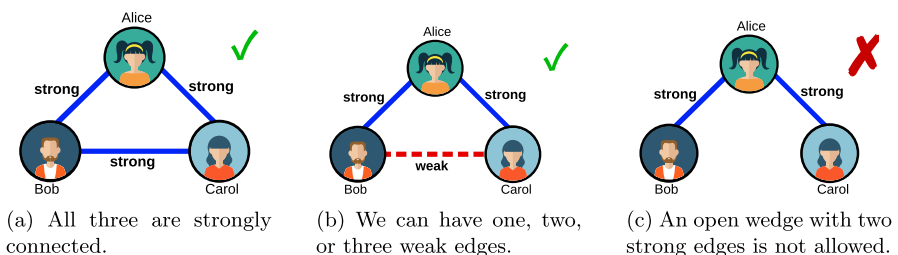


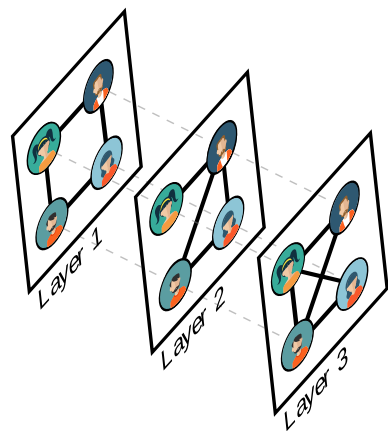
Fig. 1 Illustration of the *Strong Triadic Closure* (STC) principle. Alice is strongly connected to Bob and Carol. In (a), all three edges are strong and the STC is satisfied. In (b), Bob and Carol are connected by a weak edge, which still satisfies the STC. In (c), Bob and Carol are not connected, resulting in an open wedge with two strong edges and thus a violation of the STC

*layer networks*¹ (De Domenico 2023; De Domenico et al. 2013; Kivelä et al. 2014). Typical examples include social networks combining online and offline interactions, biological networks integrating multiple types of molecular interactions, or collaboration networks where each layer corresponds to a different venue or context. Unlike temporal snapshots, where layers represent the same relation observed at different times, multilayer networks capture distinct but coexisting types of relationships. Figure 2 shows an example of a 3-layer network.

We strictly focus on multilayer networks with *semantically aligned layers*, where different layers capture interactions in distinct contexts or modalities but are intended to support a *single* notion of tie strength. We model this by assuming that each dyad that is observed in at least one layer is associated with one latent binary tie strength (*strong* vs. *weak*), whose semantics are shared across layers. Individual layers provide complementary, potentially noisy and incomplete evidence about this latent tie strength: in particular, the absence of an edge in a layer is not interpreted as “weak”, but merely as missing or unobserved interaction evidence in that layer. Equivalently, we view the layers as partial observations that provide *hints* about an underlying binary tie strength, rather than as carrying layer-specific strength semantics.

The challenge of consistency. A straightforward way of applying STC to multilayer networks is to solve the STC problem independently in each layer. This produces a *strong/weak* labeling for every layer, but can easily lead to *inconsistent tie strength semantics*: the same edge may be inferred as strong in one layer and weak in another. Although each layer-wise labeling may satisfy STC in isolation, such inconsistencies undermine interpretability and complicate cross-layer analysis. Figure 3 illustrates this phenomenon in a two-layer network. While both labelings in Fig. 3a and 3b satisfy STC within each layer, only the labeling in Fig. 3b preserves consistent edge semantics across layers. Moreover, enforcing the STC on an aggregated graph (e.g., merging all layers into a single weighted network in which the weights correspond to the multiplicities of the edges in the multilayer graph) is generally insufficient for guaranteeing STC validity in each layer, since a wedge that is open in a

Fig. 2 A multilayer network with three layers over a common node set. Each layer represents a different interaction type and may have a different edge set. We assume a *single latent tie strength for each node pair*, whose semantics are shared across all layers, independently of edge existence. This motivates consistent tie-strength inference across layers



¹Here, the term “layer” refers to an interaction layer and should not be confused with layers in neural-network models.

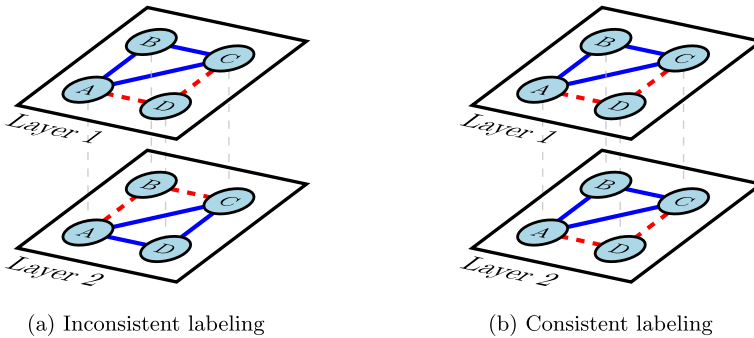


Fig. 3 Illustration of inconsistent versus consistent edge labelings in a two-layer network. Strong edges (solid blue) and weak edges (dashed red) represent inferred tie strengths. While both (a) and (b) satisfy the STC property in each layer, only (b) maintains consistent edge semantics across layers

particular layer may appear closed in the aggregated graph due to an edge that exists only in a different layer.

Therefore, our goal is to infer a single, coherent assignment of tie strengths that simultaneously satisfies the STC constraints induced by all layers and preserves a consistent semantic interpretation across layers.

When is consistency appropriate? We emphasize that cross-layer consistency is a *modeling assumption*, not a universal requirement. It is well suited for multilayer networks in which layers provide complementary views of the same underlying dyadic relationships, and where the meaning of tie strength (e.g., close vs. distant) is expected to remain stable across interaction types. Typical examples include social networks combining digital and physical interactions, as well as biological networks in which multiple experimental layers capture complementary aspects of the same underlying molecular interactions. In such settings, inconsistent tie strength labels across layers can distort the interpretation of relationships and hinder meaningful cross-layer analysis. Accordingly, we model tie strength as a binary latent variable (*strong* or *weak*) whose semantics are preserved across layers.

In contrast, in highly heterogeneous multilayer systems, such as networks combining family ties and professional collaborations, it may be entirely reasonable for a relationship to be strong in one layer and weak in another. Our framework is designed for the former class of networks, where consistent semantics are meaningful.

Our approach. To avoid ad hoc design choices, we adopt an axiomatic approach to formulating STC in multilayer networks. We identify five natural axioms that any reasonable multilayer STC objective should satisfy: (1) penalization of weak edges, (2) layer permutation invariance, (3) edge-wise decomposability, (4) non-decreasing disagreement penalties with more strong labels, and (5) invariance to duplicating weak-only layers. These axioms characterize a family of admissible objective functions parameterized by a non-decreasing weighting function and lead to principled optimization problems for multilayer STC and its extension STC+, which allows edge insertions.

Both problems are NP-hard. We develop efficient approximation algorithms with constant-factor guarantees and provide exact integer and linear programming formu-

lations. Our methods infer a single, coherent labeling of tie strength that simultaneously respects STC constraints in all layers.

Contributions. Our main contributions are:

- We introduce the first *axiomatically grounded* formulations of the STC and STC+ problems for multilayer networks, ensuring consistent tie strength semantics across all layers.
- We design efficient approximation algorithms with provable guarantees: a 2-approximation for multilayer STC and a 6-approximation for multilayer STC+, along with exact integer and linear programming formulations.
- We conduct extensive experiments on real-world multilayer networks from diverse domains, demonstrating that layer-wise baselines produce inconsistent and often misleading labelings, while our approach yields fully consistent and interpretable results.

2 Related work

There are several recent surveys on multilayer networks, e.g., Boccaletti et al. (2014); Crainic et al. (2022); De Domenico et al. (2013); Dickison et al. (2016); Hammoud and Kramer (2020). Pappalardo et al. (2013) measure tie strength in multilayer networks, where a greater number of connections across different layers corresponds to a higher (continuous) tie strength between two users. However, their work does not consider the STC property. Other studies have predicted tie strength based on node and edge features in conventional networks, e.g., Gilbert et al. (2009); Jones et al. (2013); Pham et al. (2016); Xiang et al. (2010), but these approaches also do not classify edges with respect to the STC. In contrast, our work is grounded in the STC framework, and we infer binary tie strengths in each layer purely from the network topology. Sintos et al. (2014) introduced the STC optimization problems, characterizing edges as strong or weak based solely on structural information. They proved that maximizing the number of strong edges is NP-hard and provided two approximation algorithms. An extensive analysis of STC can be found in Easley and Kleinberg (2010). Subsequent work has explored restricted network classes to further analyze the complexity of STC maximization (Grüttemeier and Komusiewicz 2020; Konstantinidis et al. 2018; Konstantinidis and Papadopoulos 2020). Rozenshtein et al. (2017) incorporated community connectivity constraints, while Adriaens et al. (2020) proposed integer linear programming formulations and relaxations. More recently, Matakos and Gionis (2022) studied the problem of reinforcing connectivity by adding new edges guided by strong ties. The STC framework has also been extended to dynamic and temporal settings. Oettershagen et al. (2022) considered temporal networks with evolving topology using a sliding time window, and later extended this approach to the STC+ variant (Oettershagen et al. 2025). Arachchi et al. (2024) combined the STC with the densest subgraph problem, while Veldt et al. (2022) examined connections between STC+ and cluster editing.

To the best of our knowledge, our work is the first to address the STC in multilayer networks. We ground our formulation in a set of explicit axioms that capture essential principles any consistent tie strength inference method should satisfy.

3 Preliminaries

Table 6 in the appendix gives an overview of the notation and symbols. We use $[k]$ with $k \in \mathbb{N}$ to denote the set $\{1, \dots, k\}$.

Multilayer Graphs, and (Hyper-)graphs: We define a *multilayer graph* $G = (V, E_1, \dots, E_k)$ consisting of a finite set of nodes V and k finite sets E_i with $i \in [k]$ of undirected edges $e = \{u, v\}$ with u and v in V , and $u \neq v$. We define $n = |V|$, $m_i = |E_i|$, and $m = \sum_{i=1}^k m_i$. We call a 1-layer graph just a *graph*; e.g., each individual layer $G_i = (V, E_i)$, with $i \in [k]$, of a multilayer graph $G = (V, E_1, \dots, E_k)$ is itself a graph. Given a graph $G = (V, E)$, we define a *wedge* as a triplet of nodes $u, v, w \in V$ such that $\{\{u, v\}, \{v, w\}\} \subseteq E$ and $\{u, w\} \notin E$. We denote such a wedge by $(v, \{u, w\})$, and with $\mathcal{W}(E)$, the set of wedges in the set of edges E . Moreover, $\mathcal{W}(E, \{u, w\}) = \{(v, \{u, w\}) \mid (v, \{u, w\}) \in \mathcal{W}(E)\}$ denotes the set of wedges w.r.t. $\{u, w\}$. Note that $\mathcal{W}(E, \{u, w\}) = \emptyset$ whenever $\{u, w\} \in E$, since wedges are defined only with respect to missing edges. A *hypergraph* $H = (V, E)$ consists of a finite set of nodes V and a finite set of *hyperedges* $E \subseteq 2^V \setminus \emptyset$, i.e., each hyperedge connects a non-empty subset of V . We consider the special case of *3-uniform* hypergraphs in which each hyperedge consists of an unordered triple. We define the union $H = H_1 \cup H_2$ of two (hyper-)graphs $H_1 = (V_1, E_1)$ and $H_2 = (V_2, E_2)$ as $H = (V_1 \cup V_2, E_1 \cup E_2)$.

Given a graph $G = (V, E)$, we assign each edge $e \in E$ one of the labels *weak* or *strong*. Such a labeling is called a *strong-weak labeling* and is represented by a subset $S \subseteq E$, where each edge $e \in S$ is labeled *strong*, and each $e \in E \setminus S$ is labeled *weak*.

The *strong triadic closure (STC)* property requires that for any two strong edges $\{u, v\} \in S$ and $\{v, w\} \in S$, the edge $\{u, w\}$ must also exist in E (either weak or strong). We say that a labeling $S \subseteq E$ *fulfills* the STC if this condition holds. Equivalently, the STC is fulfilled if and only if no wedge in $\mathcal{W}(E)$ contains more than one strong edge; that is, there is no triple $(v, \{u, w\})$ such that both $\{u, v\}$ and $\{v, w\}$ are strong and $\{u, w\} \notin E$; cf. Sintos et al. (2014). Two natural optimization variants have been defined and studied:

- **MAXSTC:** maximize $|S|$ subject to $S \subseteq E$ fulfilling STC.
- **MINSTC:** minimize $|E \setminus S|$ subject to $S \subseteq E$ fulfilling STC.

These two formulations are complementary: they optimize over the same feasible set of STC-valid labelings, but with opposing objectives, i.e., maximizing the number of strong edges versus minimizing the number of weak edges. As a result, an

optimal solution to one formulation can be obtained from an optimal solution to the other. However, the two problems differ substantially from an algorithmic perspective: approximation guarantees do not carry over, since they are defined with respect to different objective values. MINSTC admits a 2-approximation, while MAXSTC is hard to approximate within $n^{1-\epsilon}$ for any constant $\epsilon > 0$ unless $P = NP$ (Sintos et al. 2014).

Approximation in Single-Layer Graphs: To approximate MINSTC in a single-layer graph $G = (V, E)$, we first construct the wedge graph $W(G) = (V_W, E_W)$. The wedge graph $W(G)$ contains for each edge $\{i, j\} \in E$ one vertex $n_{ij} \in V_W$. Two vertices n_{uv}, n_{uw} in $W(G)$ are adjacent if and only if there exists a wedge $(u, \{v, w\}) \in \mathcal{W}(E)$. Sintos et al. (2014) showed that solving the *vertex cover* (VC) problem on $W(G)$ leads then to a solution for the STC of G , where VC is defined as follows: Given a graph $G = (V, E)$, a minimum vertex cover is a subset $C \subseteq V$ such that each edge $e \in E$ is incident to at least one vertex $v \in C$. Following the definition of $W(G)$, choosing the set of *weak* edges to be the edges $\{u, v\} \in E$ such that $n_{uv} \in C$, each wedge in $\mathcal{W}(E)$ has at least one *weak* edge, i.e., the labeling fulfills the STC.

Recently, the approximation was generalized for edge weighted graphs (Oettershagen et al. 2022) by mapping the edge weights of the input graph to node weights in the wedge graph, and solving the *Minimum Weighted Vertex Cover* (MWVC) problem. The running time of this approximations is in $\mathcal{O}(nm)$ due to the size of the wedge graph. We adopt and extend this wedge-graph-based approach to approximate the multilayer STC problem, incorporating additional structural constraints across layers.

Strong Triadic Closure with Edge Additions (STC+): In the STC+ variant, apart from labeling the edges of the graph as *strong* or *weak*, new *weak* edges between non-adjacent nodes can be added. The MINSTC+ problem is defined as follows (Sintos et al. 2014): Given a graph $G = (V, E)$, the goal is to find a set $F \subseteq \binom{V}{2} \setminus E$ and a set $E' \subseteq E$ such that the strong edge set $E \setminus E'$ fulfills the strong triadic closure with respect to the augmented graph $(V, E \cup F)$, while minimizing $|E' \cup F|$.

STC+ is useful, for example, when a single missing edge can force many others to be labeled weak under STC (Sintos et al. 2014). Adding such edges can significantly reduce the number of weak labels required.

In previous work, we extended the approximation of the STC to the STC+ problem (Oettershagen et al. 2025). Given a graph G , we construct a 3-uniform wedge hypergraph in which nodes represent edges of the input graph or edges that can be newly inserted. Hyperedges correspond to the wedges in G . We obtain an approximation solving a MWVC analogously to the approximation of the weighted MINSTC problem; the running time is in $\mathcal{O}(nm)$. Our algorithm for the multilayer STC+ builds on this idea and generalizes it to multilayer networks.

4 Strong triadic closure in multilayer graphs

Our goal is to find an STC labeling for each layer that is consistent across the multilayer network. Formally defining a cost function that properly balances the penalization of *weak* edges with disagreements across layers is a non-trivial task. To ensure our formulation is robust, interpretable, and free from arbitrary design choices, we adopt an axiomatic approach. We define the following five axioms that any reasonable multilayer STC objective should fulfill:

1. *Weakness Minimization*: The cost function must include a cost for every edge instance labeled *weak*.
2. *Layer Permutation Invariance*: The final cost of a labeling must be independent of the order in which the layers are presented.
3. *Edge-wise Decomposability*: The total cost of a labeling for the entire multilayer network can be expressed as the sum of costs calculated for each individual edge.
4. *Strong Persistence Penalty*: If and only if an edge has conflicting labels (is *strong* in at least one layer and *weak* in at least one), it must incur a disagreement penalty. This penalty must not decrease as the number of layers labeling it *strong* increases.
5. *Weak Duplication Invariance*: Duplicating a layer where an edge is already known to be *weak* (or is absent) must not alter the optimal resolution for that edge.

Axiom (1) anchors the objective in the classical MinSTC tradition: strong ties are the scarce resource and every weak instance carries budgetary cost. Axiom (2) rules out artifacts from data representation: permuting, batching, or reordering layers must not change the value of any labeling. Axiom (3) requires separability across edges. It lets us write the global objective as a sum of per-edge terms plus feasibility via STC constraints. Axiom (4) captures the semantics of *strong*: if an edge is called *strong* somewhere and *weak* elsewhere, that disagreement must be penalized, and the penalty should not decrease as the number of *strong* occurrences grows. Intuitively, each *strong* claim creates closure obligations in its layer; disagreeing against more such claims should never be cheaper. Axiom (5) prevents gaming by data replication. Duplicating a layer in which an edge is *weak* (or absent) must not alter which labelings are optimal for that edge. In particular, disagreement penalties must not grow with the number of *weak* occurrences, otherwise duplicating layers that contain only weak edges would unduly bias the optimizer's outcome.

We show that these axioms naturally lead to a specific cost objective. In the following, we use $E = \bigcup_{i \in [k]} E_i$ and for each $e \in E$ we define $s(e)$ to be the number of layers in which e is labeled *strong* and $w(e)$ the number of layers in which e is labeled *weak*.

Theorem 1 *Any cost function that assigns constant costs (without loss of generality, unit costs) satisfies Axioms (1)–(5) if and only if it has the form*

$$Cost(S_1, \dots, S_k) = \sum_{e \in E} f(s(e), w(e))$$

where $f : \mathbb{N}_0 \times \mathbb{N}_0 \rightarrow \mathbb{R}_{\geq 0}$ satisfies:

- (i) $f(s(e), w(e)) \geq w(e)$ for all $s(e), w(e)$,
- (ii) $f(s(e), 0) = 0$ for all $s(e)$ (no cost if edge is consistently strong),
- (iii) $f(s(e), w(e)) = w(e) + g(s(e))$ for some non-decreasing function $g : \mathbb{N}_0 \rightarrow \mathbb{R}_{\geq 0}$ with $g(0) = 0$ when $w(e) > 0$.

Guided by Axioms (1)-(5) and Theorem 1, we now formalize the multilayer STC objective. Theorem 1 characterizes the class of valid cost functions under the axioms but leaves the disagreement penalty function g undetermined. We first define *disagreements* between per-layer labelings. We say that labelings S_i and S_j , with $i \neq j \in [k]$, *disagree* on edge e if exactly one of them labels e strong, i.e., $e \in S_i \cup S_j$ and $e \notin S_i \cap S_j$. We then define the *disagreement cost* of the labelings S_1, \dots, S_k as

$$d_k(S_1, \dots, S_k) = \sum_{\substack{e \in E \\ w(e) > 0}} s(e).$$

A labeling is *consistent* if and only if $d_k(S_1, \dots, S_k) = 0$. We can now define the STC problem in multilayer graphs as follows.

Problem 1 Given a multilayer graph $G = (V, E_1, \dots, E_k)$, find labelings $S_i \subseteq E_i$ for $i \in [k]$ such that

1. for each $i \in [k]$, S_i satisfies the STC for E_i , and
2. $Cost(S_1, \dots, S_k) = \sum_{i \in [k]} |E_i \setminus S_i| + d_k(S_1, \dots, S_k)$ is minimized.

This objective directly instantiates Theorem 1 with the linear disagreement penalty $g(s(e)) = s(e)$ when $w(e) > 0$, which is the minimal non-decreasing function satisfying Axiom (4). Each *strong* occurrence of an edge that is not consistently labeled across layers contributes one unit to the disagreement penalty.

The MINMULTILAYERSTC problem is a generalization of the single-layer MINSTC. Moreover, by *layer anonymity* and *edge-wise decomposability* (Axioms (2)-(3)), any disagreement surcharge for an edge e can depend only on its occurrence counts $(s(e), w(e))$, not on the order of layers. *Weak Duplication Invariance* (Axiom (5)) rules out any dependence on $w(e)$ beyond the event $w(e) > 0$ to prevent that duplicating a weak-only layer would spuriously change the objective. *Strong Persistence Penalty* (Axiom (4)) then requires the surcharge to be nondecreasing in $s(e)$. Consequently, we penalize *each strong occurrence* of e whenever e is not uniformly labeled across layers. Operationally, each strong claim creates STC closure obligations in its layer; disagreeing with more such claims should never be cheaper.

Note that solving the single-layer STC optimally in each layer does not lead to an optimal solution for the multilayer variant. For example, in Fig. 3a each layer is optimally solved separately; however, the optimal solution for the multilayer STC is shown in Fig. 3b. Furthermore, for MINMULTILAYERSTC, the STC constraints are evaluated *within each layer*: if $\{u, v\}$ and $\{v, w\}$ are labeled strong in layer i , then the closing edge $\{u, w\}$ must be present in E_i (weak or strong) to avoid an open wedge in that layer. Edges that exist only in other layers do not close wedges in layer i . Hence, enforcing STC on an aggregated union graph is generally insufficient for guaranteeing STC validity in each layer.

Finally, a maximization variant generalizing MAXSTC can be defined, but it inherits the hardness of approximation. Thus, here we focus on the minimization variant, which admits constant-factor approximations.

In the remainder of this section, we give a 2-approximation for MINMULTILAYERSTC. The following lemma provides the basis for our approximation algorithm and the ILP formulation. It establishes that we can focus on finding consistent solutions without loss of optimality. Intuitively, it states that if there is an optimal solution with disagreements, then there exists also an optimal solution without disagreements.

Lemma 1 *Let S_1^*, \dots, S_k^* be labelings of strong edges for a multilayer graph $G = (V, E_1, \dots, E_k)$, where $S_i^* \subseteq E_i$ for each $i \in [k]$. Let C^* be their objective value for the MINMULTILAYERSTC problem, i.e., $C^* = \sum_{i \in [k]} |E_i \setminus S_i^*| + d_k(S_1^*, \dots, S_k^*)$. There exists a set of labelings of strong edges $H^* = \{H_1^*, \dots, H_k^*\}$ (where $H_i^* \subseteq E_i$) s.t.:*

1. *The labelings in H^* are consistent, i.e., $d_k(H_1^*, \dots, H_k^*) = 0$.*
2. *The objective value of H^* for MINMULTILAYERSTC is equal to C^* , i.e., $\sum_{i \in [k]} |E_i \setminus H_i^*| + d_k(H_1^*, \dots, H_k^*) = C^*$.*

Lemma 1 is crucial because it allows us to simplify the search space for MINMULTILAYERSTC. Instead of needing to find a labeling that might have disagreements and then account for the d_k penalty, we can seek an optimal solution that is inherently consistent (i.e., $d_k(S_1, \dots, S_k) = 0$). In such a consistent labeling, an edge e is either *strong* in all layers it appears in, or *weak* in all layers it appears in.

Our approximation algorithm for MINMULTILAYERSTC, presented as Algorithm 1, is designed based on this principle. It first constructs individual wedge graphs W_i for each layer $G_i = (V, E_i)$. These are then combined into a single, node-weighted wedge graph $W = (V_W, E_W)$. The nodes V_W in W correspond to the unique edges $e \in E$ present in the input multilayer graph. The weight $\omega(n_e)$ of a node $n_e \in V_W$ (representing edge e) is set to the multiplicity of e . An edge (n_{e_1}, n_{e_2}) exists in E_W if e_1 and e_2 form two sides of a wedge in some layer G_j . The algorithm then finds an

approximate Minimum Weighted Vertex Cover (MWVC) in W . Edges corresponding to nodes in this MWVC are labeled *weak* across all their layers, yielding a consistent STC labeling.

Theorem 2 *Given an α -approximation for the MWVC with running time T_{VC} , Algorithm 1 computes an α -approximation for the MINMULTILAYERSTC in $\mathcal{O}(k \cdot nm + T_{VC})$ running time.*

Algorithm 1 A 2-approximation algorithm for MINMULTILAYERSTC

```

Input: Multilayer graph  $G = (V, E_1, \dots, E_k)$ 
Output: Consistent labelings  $S_1, \dots, S_k$ 
1  $V_W \leftarrow \emptyset; E_W \leftarrow \emptyset$  // Initialize combined wedge graph  $W$ 
2 for  $i \in \{1, \dots, k\}$  do
3   Compute wedge graph  $W_i = (V_i^{wg}, E_i^{wg})$  of  $G_i = (V, E_i)$ 
4    $V_W \leftarrow V_W \cup V_i^{wg}$  // Nodes in  $V_i^{wg}$  are  $n_e$  for  $e \in E_i$ 
5    $E_W \leftarrow E_W \cup E_i^{wg}$ 
   // For each unique edge  $e \in E = \bigcup E_j$  represented by  $n_e$ :
6 for  $n_e \in V_W$  do
7    $\omega(n_e) \leftarrow |\{j \mid e \in E_j\}|$  // Weight is multiplicity of edge  $e$ 
8   Approx. MWVC  $C \subseteq V_W$  on  $W$  with node weights  $\omega$ 
9    $E_{weak} \leftarrow \{e \mid n_e \in C\}$  // Set of unique edges labeled weak
10 for  $i \in [k]$  do
11    $S_i \leftarrow E_i \setminus E_{weak}$  // Strong labeled edges in layer  $i$ 
12 return strong edge sets  $S_i$  for each layer  $i \in [k]$ 

```

The MWVC solver used in Line 8 of Algorithm 1 can be chosen flexibly. For example, employing the *pricing algorithm* yields a 2-approximation for MWVC with running time linear in the number of edges of the wedge graph (Bar-Yehuda and Even 1981). By Theorem 2, this implies an overall 2-approximation for MINMULTILAYERSTC with running time $\mathcal{O}(k \cdot nm)$, where $n = |V|$, $m = \sum_{i=1}^k |E_i|$ denotes the total number of edges across all layers, and k is the number of layers.

Regarding practical applicability, the computational bottleneck of our approach lies in enumerating wedges across layers, making scalability primarily dependent on local density and clustering. In our experimental evaluation (Sect. 6), we demonstrate scalability to large empirical multilayer networks with hundreds of thousands of nodes and over a million edges, while extremely dense or web-scale networks would require additional sparsification or parallelization.

Moreover, the space complexity of Algorithm 1 is dominated by storing the wedge graph W , which requires $\mathcal{O}(knm)$ space in the worst case. Finally, we provide a linear programming formulation for the exact computation in Appendix B Fig. 4.

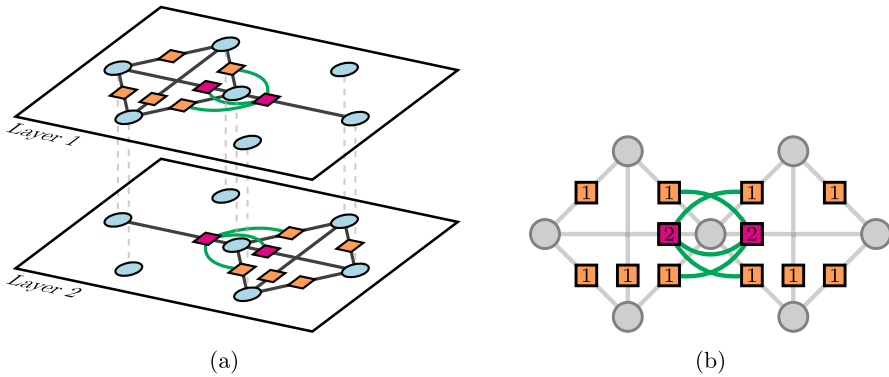


Fig. 4 Construction of the combined wedge graph for Algorithm 1. **(a)** A multilayer graph with two layers. In each layer, the wedge graph is overlaid with its nodes shown as squares and its edges in green. **(b)** The combined node-weighted wedge graph, with the aggregated multilayer graph shown in gray in the background

5 Multilayer STC with edge insertions

We now consider the MINMULTILAYERSTC+ problem (Problem 2), where new *weak* edges $E_i^+ \subseteq \binom{V}{2} \setminus E_i$ can be added to layer $i \in [k]$ to potentially achieve a better overall STC labeling. The problem formulation follows the same axiomatic principles introduced in Sect. 4 (Axioms 1-5) to govern the cost function. The goal is to minimize the total number of *weak* edges (original ones in $E_i \setminus S_i$ plus newly added ones in E_i^+) plus the disagreement cost $d_k(S_1, \dots, S_k)$. The disagreement cost d_k is now defined over edges in $(E_i \cup E_i^+) \cap (E_j \cup E_j^+)$.

Problem 2 Given a multilayer graph $G = (V, E_1, \dots, E_k)$, find k subsets of new edges $E_i^+ \subseteq \binom{V}{2} \setminus E_i$ and labelings $S_i \subseteq E_i$ for $i \in [k]$ such that

1. for each $i \in [k]$, S_i satisfies the STC for $E_i \cup E_i^+$, and
2. $\sum_{i \in [k]} |E_i^+ \cup E_i \setminus S_i| + d_k(S_1, \dots, S_k)$ is minimized.

Problem 2 generalizes the single-layer variant of the problem defined by Sintos et al. (2014). Similar to Lemma 1 for MINMULTILAYERSTC, the following lemma establishes that an optimal solution to MINMULTILAYERSTC+ can be found that is consistent across layers, even when considering newly added edges.

Lemma 2 Let S_1^*, \dots, S_k^* be labelings of strong edges for a multilayer graph $G = (V, E_1, \dots, E_k)$ with potential new edges $E_i^+ \subseteq \binom{V}{2} \setminus E_i$, where $S_i^* \subseteq E_i$ for

each $i \in [k]$. Let C_+^* be their objective value for the MINMULTILAYERSTC+ problem, i.e., $C_+^* = \sum_{i \in [k]} |(E_i \cup E_i^+) \setminus S_i^*| + d_k(S_1^*, \dots, S_k^*)$. There exists another set of labelings $H^* = \{H_1^*, \dots, H_k^*\}$ (with $H_i^* \subseteq E_i$) s.t.:

1. The labeling H^* is consistent, i.e., $d_k(H_1^*, \dots, H_k^*) = 0$.
2. The objective value of H^* for MINMULTILAYERSTC+ is equal to C_+^* , i.e., $\sum_{i \in [k]} |(E_i \cup E_i^+) \setminus H_i^*| + d_k(H_1^*, \dots, H_k^*) = C_+^*$.

Lemma 2 implies that the cost of an optimal consistent solution involves summing, for each edge $e \in E \cup \bigcup E_i^+$, its contribution if it is weak. If $e \in E_i$ (original), it contributes to $|E_i \setminus S_i|$. If $e \in E_i^+$ (new), it contributes to $|E_i^+|$. Due to consistency, if e is weak, it is weak in all layers it's part of (original or added). To develop our approximation algorithm, we first introduce an intermediate problem variant.

Problem 3 As MINMULTILAYERSTC+ (Problem 2), with the additional constraint (3): for any $i \in [k]$, if a new edge $e \in E_i^+$ is added to close a wedge in $G_i = (V, E_i)$, and if for some other layer $j \neq i$, e also closes a wedge in G_j (i.e., $(v, e) \in \mathcal{W}(E_j)$ for some v), then $e \in E_j^+$ must also hold.

The new constraint (3) means that if a new edge e is chosen to be added as weak to resolve STC violations, its addition is propagated to all layers where it could serve a similar purpose. This aids in constructing a unified hypergraph for approximation. The relationship between this variant and our target problem is given by:

Lemma 3 An α -approximation for CONSTRAINTMMLSTC+ (Problem 3) provides a 2α -approximation for MINMULTILAYERSTC+ (Problem 2).

Our approximation algorithm, Algorithm 2, addresses MINMULTILAYERSTC+ by first targeting CONSTRAINTMMLSTC+. It constructs a combined 3-uniform wedge hypergraph $W = (V_W, E_W)$. The vertices V_W correspond to nodes n_e for all existing edges $e \in E$ (the aggregated set of unique edges) and all potential new edges $e' \in \binom{V}{2} \setminus E$. Hyperedges in E_W represent open wedges from any layer $G_i = (V, E_i)$

. The weight $\omega(n_e)$ for a node $n_e \in V_W$ is:

- If $e \in E$: $\omega(n_e)$ is its multiplicity, $|\{j \mid e \in E_j\}|$.
- If $e \notin E$ (a potential new edge): $\omega(n_e)$ is the number of layers j where adding e as weak would close an open wedge in G_j .

The algorithm then finds an approximate Minimum Weighted Vertex Cover (MWVC) $C \subseteq V_W$ on this hypergraph W (e.g., using a 3-approximation algorithm like pricing for 3-uniform hypergraphs). Edges $e \in E$ with $n_e \in C$ are designated as weak. Potential new edges $e' \notin E$ with $n_{e'} \in C$ are provisionally added to the sets E_i^+ for all layers i where their addition was counted in .

Algorithm 2 A 6-approximation algorithm for MINMULTILAYERSTC+

```

Input: Multilayer graph  $G = (V, E_1, \dots, E_k)$ 
Output: Labelings  $S_1, \dots, S_k$  and new edge sets  $E_1^+, \dots, E_k^+$ 
// Construct 3-uniform wedge hypergraph  $W = (V_W, E_W)$ 
1  $V_W \leftarrow \{n_e \mid e \in E \text{ or } e \notin E \text{ and closes at least one open wedge}\}$ 
2  $E_W \leftarrow \emptyset$ 
3 for  $i \in \{1, \dots, k\}$  do
4   for each open wedge  $(v, \{u, x\})$  in  $G_i = (V, E_i)$  do
5      $E_W \leftarrow E_W \cup \{\{n_{vu}, n_{vx}, n_{ux}\}\}$ 
6 for  $n_e \in V_W$  do
7   if  $e \in E = \bigcup E_j$  then
8      $\omega(n_e) \leftarrow |\{j \mid e \in E_j\}|$ 
9   else //  $e$  is potential new edge
10     $\omega(n_e) \leftarrow |\{j \mid \exists v : (v, \{u, w\}) \text{ is wedge in } G_j \text{ and } e = \{u, w\}\}|$ 
11 Approximate MWVC  $C \subseteq V_W$  on  $W$  with weights  $\omega$ 
12  $E_{\text{weak}} \leftarrow \{e \mid n_e \in C \text{ and } e \in E\}$ 
13 Initialize  $E_i^+ \leftarrow \emptyset$  for all  $i \in [k]$ 
14 for  $n_e \in C$  where  $e \notin E$  (new edge selected by MWVC) do
15   for  $i \in [k]$  such that adding  $e$  to  $G_i$  closes an open wedge in  $G_i$  do
16      $E_i^+ \leftarrow E_i^+ \cup \{e\}$ 
17 for  $i \in [k]$  do
18    $S_i \leftarrow E_i \setminus E_{\text{weak}}$  // Strong labeled edges
19 return strong edges  $S_i$  and new edges  $E_i^+$  for each  $i \in [k]$ 

```

$$\omega(n_{e'})$$

Theorem 3 Given an α -approximation for the MWVC in 3-uniform hypergraphs with running time T_{VC_H} , Algorithm 2 computes a 2α -approximation for MINMULTILAYERSTC+ (Problem 2) in $\mathcal{O}(k \cdot nm + T_{VC_H})$ running time.

Our algorithm (Algorithm 2) achieves a 6-approximation for the MINMULTILAYERSTC+ problem by employing the pricing method, a linear-time 3-approximation algorithm for MWVC in 3-uniform hypergraphs (Bar-Yehuda and Even 1981). The overall time and space complexity is $\mathcal{O}(k \cdot nm)$. For completeness, we also provide an ILP formulation for the exact computation in Appendix B.

6 Experiments

We conduct experimental evaluations to answer the following research questions:

- Q1. Consistency:** How much does the multilayer STC improve the consistency compared to the standard STC?
- Q2. Multilayer STC+:** How does the MINMULTILAYERSTC+ improve the labeling compared to MINMULTILAYERSTC?
- Q3. Approximation quality:** How is the empirical approximation quality of our

algorithms?

Q4. Efficiency: How efficient are our algorithms in terms of running time and memory usage?

Additionally, we present a qualitative case study in Sect. 6.3.6.1 Algorithms, datasets, and metrics

We implemented and compared the following three groups of algorithms:

- **Ours:** `Cosmic` and `Cosmic+` are our approximation algorithms for the multilayer STC and STC+ problems (Algorithm 1 and Algorithm 2). `ExactML` and `ExactML+` are the exact computations of the STC and STC+ using the ILPs introduced in Appendix B.
- **Layer-wise baselines:** `BLExact` and `BLExact+` are exact baselines that apply the single-layer ILP formulations for STC (Adriaens et al. 2020) and STC+ (Veldt et al. 2022), respectively, independently to each layer. Similarly, `BLstc` and `BLstc+` are baseline heuristics that approximate the solutions of STC and STC+, respectively, in each layer independently.
- **Aggregation baselines:** `BLAgg` and `BLAgg+` are baselines that aggregate layers via edge multiplicity and then solve a single weighted STC or STC+ instance on the resulting graph. The resulting labels are mapped back to all layers in which an edge appears. In addition, we report exact variants of the aggregated baselines (`BLAggExact`, `BLAggExact+`), obtained by solving the corresponding weighted STC/STC+ instance via ILP.

For the approximation-based methods, we compute the required vertex covers in the wedge graphs using the pricing algorithm (with unit weights in the unweighted case) (Bar-Yehuda and Even 1981) and a greedy algorithm that iteratively adds the vertex covering the largest number of uncovered edges (see Vazirani (2001)). We append (P) for pricing and (G) for greedy to the algorithm names, e.g., `Cosmic(P)` for our STC approximation using the pricing method.

The aggregation-based baselines capture a common heuristic in multilayer analysis: reducing the network to a single weighted graph before applying a single-layer method. This enforces identical labels across layers for shared edges and thus eliminates cross-layer conflicts. However, since STC constraints are evaluated independently in each layer, aggregation can mask layer-specific open wedges and may produce solutions that violate STC in individual layers. We therefore additionally report the number of STC violations per layer for these baselines.

Finally, we do not compare against learning-based tie-strength prediction methods. These are typically studied in supervised, single-layer settings under application-specific definitions of tie strength (see Cheng et al. (2025)), whereas our work addresses an unsupervised structural optimization problem with explicit multilayer STC constraints. We therefore focus on optimization-based baselines targeting the same problem.

We implemented our algorithms in C++ using GNU CC Compiler 11.4.0 with the flag `-O3`. We used Gurobi 11 for solving ILPs. All experiments ran on a computer

cluster. Each experiment had an exclusive node with an Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz and 96 GB of RAM. We used a time limit of 12 h. The source code and all datasets are available at <https://gitlab.com/multilayergraphs/multilayerstc>.

Datasets. We use eight real-world datasets from different domains. Table 1 provides an overview and Appendix C further details.

Metrics. To assess cross-layer coherence, we evaluate two complementary measures:

1. **Edge disagreement rate.** We define

$$\kappa_e(S) = 1 - \frac{1}{|F|} \sum_{e \in F} \frac{s(e)}{s(e) + w(e)}, \quad F = \{e \in E : s(e) > 0\}.$$

This measures the average cross-layer inconsistency of edges that are labeled strong in at least one layer. It equals 0 when all such edges are consistently strong and increases up to 1 as strong labels become fragmented across layers.

2. **Weak-edge disagreement rate.** We define

$$\kappa_w(S) = \frac{|\{e \in E : s(e) > 0 \wedge w(e) > 0\}|}{|\{e \in E : w(e) > 0\}|},$$

which measures the fraction of weak-labeled edges that are also labeled strong in at least one other layer, i.e., how frequently conflicts occur among edges that are classified as weak.

In addition, we report the fraction of weak edges, running times, and where applicable, the number of STC violations for aggregation baselines that do not enforce layer-wise feasibility. An overview of the reported measures per method type is given in Table 2.

6.2 Results

We demonstrate that our methods achieve perfect consistency while maintaining competitive solution quality, strong approximation guarantees, and practical efficiency.

Table 1 Statistics of the real-world datasets

Dataset	$ V(G) $	$ \bigcup_i E_i(G) $	$\sum_i E_i(G) $	#Layers	Domain
<i>AUCS</i>	61	353	620	5	Social network
<i>Hospital</i>	75	1 139	1 885	5	Face-to-face
<i>Airports</i>	417	2 953	3 588	37	Transportation
<i>Rattus</i>	2 634	3 677	3 905	6	Biological
<i>FtWYt</i>	6 401	60 583	74 810	3	Social network
<i>Knowledge</i>	14 505	210 946	225 898	30	Knowledge graph
<i>HomoSap</i>	18 190	137 659	153 922	7	Biological
<i>DBLP</i>	344 814	1 528 399	1 837 562	168	Collaboration

Table 2 Reported measures. A ✓ indicates that the metric is computed and reported; otherwise, the metric is 0 by construction

Method type	Weak frac	$\kappa_e(S)_w$	STC viol	Time	
Layer-wise baselines (BLExact, BLExact+, BLstc, BLstc+)	✓	✓	✓	0	✓
Aggregation baselines (BLAggExact, BLAggExact+, BLAgg, BLAgg+)	✓	0	0	✓	✓
Ours (ExactML, Cosmic, Cosmic+)	✓	0	0	0	✓

Q1. Consistency. Table 3a reveals fundamental limitations of layer-wise approaches. The exact baseline (BLExact) exhibits $\kappa_e(S) \in [0.05, 0.25]$ across all datasets, confirming that independent layer optimization systematically produces conflicting tie semantics. Approximation baselines amplify this issue dramatically: BLstc(P) and BLstc(G) yield $\kappa_e(S) > 0.8$ on six of eight datasets, with values reaching 0.90 (*DBLP*, *Airports*, *Knowledge*), indicating that 90% of strong edges have inconsistent labels across layers. Moreover, $\kappa_w(S)$ values up to 0.32 show that conflicts contaminate substantial fractions of weak edges, undermining interpretability.

Aggregation baselines fare worse: while producing lower weak fractions by exploiting cross-layer evidence, they violate layer-specific STC constraints extensively (up to 26 975 violations on *DBLP*, 10 550 on *Hospital* for STC+), rendering solutions invalid.

In contrast, our methods enforce $\kappa_e(S) = \kappa_w(S) = 0$ by construction with zero STC violations, while achieving weak fractions competitive with or superior to layer-wise baselines. Critically, Cosmic(G) consistently outperforms BLstc(G) in weak fraction (e.g., 0.47 vs. 0.55 on *DBLP*), demonstrating that consistency does not require sacrificing solution quality.

Q2. Multilayer STC+. Table 3b confirms that edge insertion systematically improves solution quality while preserving perfect consistency. Comparing Tables 3a and 3b, STC+ reduces weak fractions substantially: ExactML+ achieves 0.58 vs. 0.65 (*AUCS*), 0.76 vs. 0.84 (*Hospital*), and 0.84 vs. 0.87 (*Airports*) compared to ExactML. Our approximations maintain this advantage across all datasets and Cosmic+(G) reduces weak fractions by 3-8 percentage points versus Cosmic(G).

Layer-wise baselines again exhibit severe consistency violations ($\kappa_e(S)$ up to 0.90), while aggregation baselines produce massive STC violations (up to 42 271 on *DBLP*). The pattern is clear: allowing edge insertions provides flexibility to reduce weak assignments, but only our formulation ensures this flexibility is exercised consistently across layers while maintaining STC validity everywhere.

Q3. Approximation Quality. Fig. 5a shows the empirical approximation ratios, i.e., the approximated objective value divided by the optimal objective value computed by the ILPs for the multilayer STC and the four smaller datasets. Even though Algorithm 1 using the pricing method (Cosmic(P)) is theoretically a 2-approximation, the empirical approximation quality is much better with values between 1.24 for *AUCS* and 1.08 for *Airports*. The greedy variant Cosmic(G), even though it is an $\mathcal{O}(\log n)$ approximation, has even better empirical approximation ratios between 1.005 for *Rattus* and 1.03 for *AUCS*. This is expected, as the greedy vertex cover algorithm is known to often outperform the pricing method (Gomes et al. 2006).

Table 3 Results for multilayer STC and STC+ (lower is better for all metrics). Our algorithms enforce consistency by construction and satisfy all STC constraints

Method	Metric	AUCS	Hospital	Air-ports	Rattus	Fft-wYt	Knowl-edge	Ho-moSap	DBLP		
<i>(a) Multilayer STC results.</i>											
Layer-wise bl.	BLExact	Weak	0.55	0.75	0.85	0.78	–	–	–	–	
		$\kappa_e(S)$	0.16	0.25	0.08	0.06	–	–	–	–	
		$\kappa_w(S)$	0.23	0.16	0.03	0.03	–	–	–	–	
	BLstc (P)	Weak	0.94	0.98	0.95	0.90	0.99	0.97	0.97	0.64	
		$\kappa_e(S)$	0.59	0.65	0.90	0.86	0.68	0.89	0.83	0.90	
		$\kappa_w(S)$	0.25	0.12	0.02	0.03	0.01	0.01	0.01	0.11	
	BLstc (G)	Weak	0.91	0.95	0.91	0.84	0.96	0.93	0.94	0.55	
		$\kappa_e(S)$	0.44	0.58	0.87	0.88	0.69	0.90	0.84	0.89	
		$\kappa_w(S)$	0.32	0.19	0.04	0.03	0.03	0.01	0.02	0.15	
	Aggregated bl.	BLAggExact	Weak	0.52	0.72	0.82	0.81	–	–	–	–
		STC viol	140	824	2 107	31	–	–	–	–	
BLAgg (P)		Weak	0.73	0.89	0.96	0.90	0.98	0.98	0.98	0.77	
		STC viol	110	332	118	7	122	569	621	4 732	
BLAgg (G)		Weak	0.64	0.83	0.88	0.84	0.94	0.94	0.94	0.66	
		STC viol	161	794	2 031	19	1 555	4 043	4 337	26 975	
Ours		ExactML	Weak	0.65	0.84	0.87	0.79	–	–	–	
		Cosmic (P)	Weak	0.81	0.96	0.94	0.88	0.97	0.96	0.97	0.58
		Cosmic (G)	Weak	0.67	0.85	0.88	0.79	0.93	0.92	0.92	0.47
<i>(b) Multilayer STC+ results.</i>											
Layer-wise bl.	BLExact+	Weak	0.45	0.54	0.80	0.75*	–	–	–	–	
		$\kappa_e(S)$	0.12	0.19	0.14	0.05*	–	–	–	–	
		$\kappa_w(S)$	0.23	0.23	0.04	0.03	–	–	–	–	
	BLstc+ (P)	Weak	0.95	0.96	0.96	0.90	0.99	0.97	0.98	0.64	
		$\kappa_e(S)$	0.54	0.61	0.89	0.86	0.69	0.90	0.83	0.90	
		$\kappa_w(S)$	0.21	0.08	0.02	0.03	0.01	0.01	0.01	0.10	
	BLstc+ (G)	Weak	0.88	0.95	0.83	0.79	0.90	0.89	0.92	0.54	
		$\kappa_e(S)$	0.39	0.58	0.84	0.86	0.67	0.90	0.84	0.89	
		$\kappa_w(S)$	0.36	0.28	0.07	0.03	0.04	0.02	0.03	0.15	
	Aggregated bl.	BLAggEx-act+	Weak	0.34	0.24	0.68	0.77	–	–	–	–
		STC viol	583	10 550	7 807	224	–	–	–	–	
BLAgg+ (P)		Weak	0.72	0.91	0.97	0.90	0.98	0.98	0.98	0.77	

Table 3 (continued)

Method	Metric	<i>AUCS</i>	<i>Hospital</i>	<i>Airports</i>	<i>Rattus</i>	<i>FfT-wYt</i>	<i>Knowledge</i>	<i>HomoSap</i>	<i>DBLP</i>	
BLAgg+ (G)	STC viol	102	278	35	18	142	556	537	5 293	
	Weak	0.46	0.49	0.78	0.78	0.91	0.91	0.90	0.64	
Ours	STC viol	466	6 205	5 629	226	6 137	22 831	37 966	42 271	
	ExactML+	Weak	0.58	0.76*	0.84*	0.77*	–	–	–	
	Cosmic+ (P)	Weak	0.80	0.92	0.91	0.87	0.97	0.95	0.96	0.58
	Cosmic+ (G)	Weak	0.64	0.83	0.81	0.74	0.92	0.90	0.90	0.46

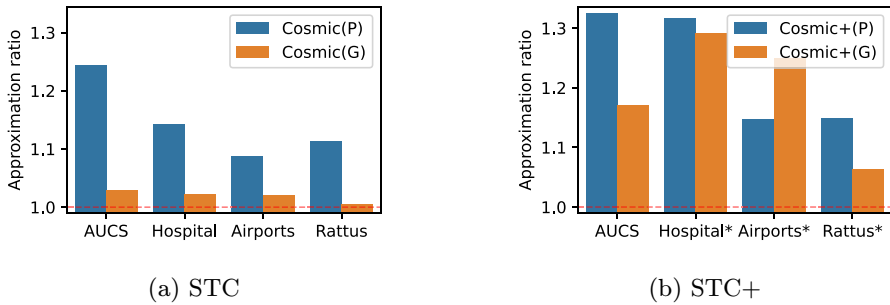


Fig. 5 The empirical approximation ratios. For datasets with (*), the reported ratio is an upper bound on the true empirical approximation ratio, computed using the best-known lower bound on the optimal solution found within the 12 h time limit

Figure 5b shows the results for the STC+. Here, we use the best lower bounds for the optimal solutions for the datasets for which no optimal solution could be obtained within the time limit, and hence, the corresponding approximation ratios reported are upper bounds. We see low empirical approximation ratios for both Cosmic+ (P) and Cosmic+ (G) with values between 1.14 for *Airports* and 1.32 for *AUCS* in the case of Cosmic+ (P) and 1.06 for *Rattus* and 1.29 for *Hospital* in the case of Cosmic+ (G). Here, Cosmic+ (P) beats Cosmic+ (G) for the *Airports* dataset.

Q4. Efficiency. Table 4 shows that running times scale with wedge (hyper-)graph sizes (cf. Table 5). On smaller networks (*AUCS*, *Hospital*), all methods complete in < 0.1s. On larger networks, our algorithms remain competitive: Cosmic(P) processes *DBLP* (168 layers, 1.8M edges) in 13 s versus 70 s for BLstc (P), despite solving a harder problem. The advantage is structural: we solve a single multilayer wedge graph, while layer-wise baselines invoke vertex cover once per layer (168 times for *DBLP*), which becomes costly as the number of layers grows. Aggregation baselines are often competitive in runtime since they operate on a single aggregated graph, but this simplification ignores layer-specific wedge constraints, explaining their substantial STC violations. For STC+, similar patterns hold. In particular, Cosmic+(G) outperforms BLstc+ (G) on *HomoSap*, as the baseline processes each layer separately while our method works on one multilayer wedge hypergraph. Memory usage is dominated by wedge (hyper-)graph size, peaking on *HomoSap* at 9.81 GB (STC) and 18.64 GB (STC+).

Table 4 Running times in seconds (s)

	Method	AUCS	Hospital	Airports	Rattus	FfTwYt	Knowledge	HomoSap	DBLP
STC	BLstc (P)	0.005	0.01	0.01	0.12	7.94	22.99	41.20	69.59
	BLstc (G)	0.005	0.01	0.01	0.11	9.74	26.56	45.78	94.01
	BLAgg (P)	0.001	0.01	0.03	0.10	4.70	26.40	27.01	18.79
	BLAgg (G)	0.001	0.01	0.04	0.15	11.37	70.34	66.72	71.00
	Cosmic (P)	0.005	0.01	0.03	0.21	9.35	32.82	54.75	12.85
	Cosmic (G)	0.005	0.01	0.04	0.31	16.83	58.91	100.04	25.87
STC+	BLstc+ (P)	0.003	0.01	0.04	0.59	29.45	90.87	231.46	59.03
	BLstc+ (G)	0.005	0.03	0.10	1.51	103.20	340.73	905.07	127.38
	BLAgg+ (P)	0.003	0.02	0.10	0.70	31.87	199.61	242.05	184.75
	BLAgg+ (G)	0.005	0.04	0.27	1.68	106.12	673.57	775.10	512.68
	Cosmic+ (P)	0.003	0.01	0.07	0.67	42.52	102.82	272.38	50.21
	Cosmic+ (G)	0.005	0.03	0.16	1.57	118.27	366.63	731.67	115.28

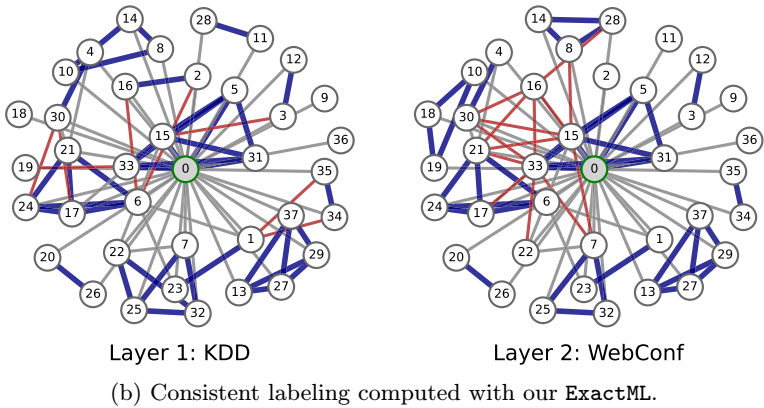
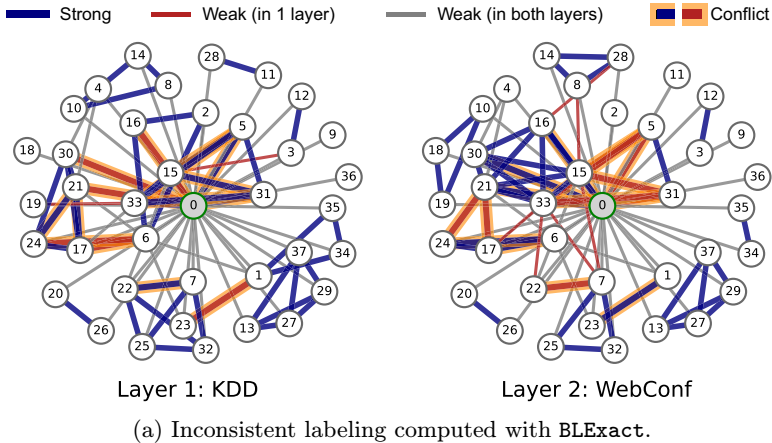
Table 5 Sizes of the wedge (hyper-)graphs (the number of (hyper-)edges is equal for STC and STC+)

Dataset	(STC) $ V(W) $	(STC+) $ V(W) $	$ E(W) $
AUCS	353	845	2 152
Hospital	1 139	2 333	15 196
Airports	2 953	17 053	50 226
Rattus	3 677	380 818	386 877
FfTwYt	60 583	6 410 083	12 037 201
Knowledge	210 946	19 795 949	42 820 276
HomoSap	137 659	50 630 405	62 634 802
DBLP	1 528 399	7 925 839	12 356 011

6.3 Case study: consistent tie strength in DBLP

We extracted a 2-layer ego network (of illustrative size) of *Christos Faloutsos* from the DBLP co-authorship graph, using publications from KDD and WebConf to define the two layers. Figure 6 shows the results of computing STC separately for each layer and using our multilayer model. *Strong* ties are shown in thick blue; *weak* ties that appear in both layers in gray and otherwise in thinner red. Edges with disagreeing labels in the two layers are highlighted with an orange halo.

The case study highlights a concrete issue caused by conflicting tie-strength labels. As shown in Fig. 6a, solving STC independently in each layer can assign different strengths to the same coauthorship edge, labeling it *strong* in one venue layer and *weak* in another. In this ego network of Christos Faloutsos, 30 out of 175 total edges ($\approx 17\%$) are in disagreement. This inconsistency has immediate practical consequences. Consider a collaborator of Faloutsos whose coauthorship edge appears in both layers but is labeled *strong* in KDD and *weak* in WebConf. A strong label carries a concrete structural obligation under STC: if Faloutsos is strongly connected to this collaborator, then any other node to whom Faloutsos is also strongly connected must be at least weakly linked to that collaborator within the same layer. Labeling the same edge *weak* in WebConf removes this obligation entirely in that layer, even if the underlying relationship has not changed. The edge therefore plays two struc-



0: C. Faloutsos	8: D. Koutra	16: K. Shin	24: S. Yang	32: X. Song
1: A. Beutel	9: D. Eswaran	17: M. Jiang	25: S. Adeshina	33: X. Dong
2: B. Prakash	10: D. Chau	18: M. Lee	26: S. Günnemann	34: Y. Matsubara
3: B. Hooi	11: E. Papalexakis	19: N. Park	27: T. Januschowski	35: Y. Sakurai
4: C. Liu	12: F. Guo	20: N. Günnemann	28: U. Kang	36: Y. Yamaguchi
5: C. Zang	13: J. Gasthaus	21: P. Cui	29: V. Flunkert	37: Y. Wang
6: C. Faloutsos	14: J. Lee	22: Q. Zhu	30: W. Zhu	
7: D. Zheng	15: J. Han	23: R. Rosenfeld	31: X. Li	

Fig. 6 Labelings in an ego-network of *Christos Faloutsos* (node 0 in the center) in a coauthorship multilayer network. The layers correspond to *KDD* and *WebConf*

turally incompatible roles simultaneously: it enforces triangle closure in KDD but imposes no such constraint in WebConf. Consequently, whether a given tie acts as a closure-enforcing strong relationship or a peripheral weak one is no longer a property of the relationship itself, but an artifact of which layer is consulted. This means that tasks relying on a single, coherent notion of tie strength, such as identifying Faloutsos’s core collaborators, recommending new collaborations, or explaining which ties enforce closure constraints in the co-authorship network, yield different and incompatible answers depending on which layer is consulted. This issue is distinct from

edges being present in one layer and absent in another, which merely reflects differences in publication activity across venues. Instead, the ambiguity arises when an edge is present in multiple layers but receives incompatible labels. The multilayer STC model (Fig. 6b) eliminates these conflicts by enforcing a uniform label for each shared edge, producing a single tie-strength assignment that is valid across all layers in which the edge appears.

7 Conclusion and future work

We introduced the first principled framework for inferring consistent tie strengths in multilayer networks. Our approach is grounded in a set of five natural axioms that lead to our objective function, ensuring that the resulting tie strength labels are both consistent and structurally well-justified. Our contributions include efficient approximation algorithms with constant-factor guarantees (2-approx. for STC, 6-approx. for STC+) and ILP formulations for the exact solution. Experiments on real-world networks confirm our methods yield consistent, interpretable labelings, outperforming baselines.

Our framework is designed for semantically aligned multilayer networks where layers provide complementary views of the same underlying relationships. The explicit wedge enumeration ($O(knm)$ complexity) presents scalability challenges for web-scale graphs. Therefore, future work includes parallelization, sparsification, and sampling-based approximations to handle massive networks, and relaxed or learned consistency models for heterogeneous layers with different tie semantics.

Appendix

See Table 6

Table 6 Commonly used notations

Symbol	Definition
$G = (V, E_1, \dots, E_k)$	multilayer graph
V	finite set of nodes
k	number of layers
E_1, \dots, E_k	finite sets E_i with $i \in [k]$ of undirected edges
$E = \bigcup_{i \in [k]} E_i$	aggregated set of edges in G
$n = V $	number of nodes
$m_i = E_i $	number of edges in layer i
$m = \sum_{i \in [k]} m_i$	total number of edges in layers
$(v, \{u, w\})$	wedge or open triangle with edge $\{u, w\}$ missing
$\mathcal{W}(E)$	set of wedges in the set of edges E
$\mathcal{W}(E, \{u, w\})$	set of wedges in E w.r.t. $\{u, w\}$
$W(G)$	wedge graph of G
$\omega(n_e)$	weight of node n_e in wedge (hyper-)graph
$S_i \subseteq E_i$	subset of <i>strong</i> labeled edges in E_i (labeling)

Table 6 Commonly used notations

Symbol	Definition
$s(e)$	number of layers in which e is labeled <i>strong</i>
$w(e)$	number of layers in which e is labeled <i>weak</i>
$d_k(S_1, \dots, S_k)$	number of disagreements between labelings S_1, \dots, S_k
$\kappa_e(S)$	Edge disagreement rate under labeling S
$\kappa_w(S)$	Weak-edge disagreement rate under labeling S

Omitted proofs

Proof of Theorem 1 \Rightarrow : By Axioms (2)-(3), the total cost is a sum over edges and each edge’s contribution depends only on its counts of strong and weak labels; hence $\text{Cost} = \sum_{e \in E} f(s(e), w(e))$ for some f .

By Axiom (1) with unit weak cost, each weak instance contributes 1, so write $f(s(e), w(e)) = w(e) + h(s(e), w(e))$ with $h(s(e), w(e)) \geq 0$. If $w(e) = 0$, there is no strong/weak conflict, so by Axiom (4) there is no disagreement penalty; hence, (ii) follows with $h(s(e), 0) = 0$, i.e., $f(s(e), 0) = 0$.

By Axiom (5), duplicating a weak-only layer must not change the *disagreement* part; thus $h(s(e), w(e))$ cannot vary with the exact value of $w(e)$ once $w(e) > 0$. Hence there exists g with $h(s(e), w(e)) = g(s(e))$ for all $w(e) > 0$, giving $f(s(e), w(e)) = w(e) + g(s(e))$ when $w(e) > 0$ showing (iii). Axiom (4) further requires that disagreement does not get cheaper as more layers label the edge strong, so g is non-decreasing; also $g(0) = 0$ (no conflict when $s(e) = 0$). Finally, since $h \geq 0$, (i) follows as we have $f(s(e), w(e)) \geq w(e)$.

\Leftarrow : Assume there exists a non-decreasing function $g : \mathbb{N}_0 \rightarrow \mathbb{R}_{\geq 0}$ with $g(0) = 0$ and define

$$f(s(e), w(e)) = \begin{cases} 0, & w(e) = 0, \\ w(e) + g(s(e)), & w(e) > 0. \end{cases}$$

Then $\text{Cost} = \sum_{e \in E} f(s(e), w(e))$ satisfies Axioms (1)–(5).

Axioms (2)–(3). Layer permutation invariance and edge-wise decomposability hold because f depends only on the counts $(s(e), w(e))$ and the total cost is a sum over edges.

Axiom (1). With unit weak cost, each weak instance contributes 1: when $w(e) > 0$ the term $w(e)$ adds exactly one per weak label, and when $w(e) = 0$ the cost is 0.

Axiom (4). If $w(e) = 0$ there is no strong/weak conflict, so $f(s, 0) = 0$. If $w(e) > 0$, the disagreement surcharge is $g(s(e))$, which is non-decreasing in $s(e)$ and satisfies $g(0) = 0$.

Axiom (5). Define the disagreement surcharge at counts $(s(e), w(e))$ by $\delta(s(e), w(e)) := f(s(e), w(e)) - w(e)$ (the part beyond the unit weak-instance cost). Duplicating a layer l where $e \in E_l$ is labeled weak changes $(s(e), w(e))$ to $(s(e), w(e) + 1)$, and Axiom (5) requires $\delta(s(e), w(e) + 1) = \delta(s(e), w(e))$ for all $w(e) \geq 1$. For our f , $\delta(s(e), 0) = 0$, while for $w(e) > 0$,

$$\delta(s(e), w(e)) = f(s(e), w(e)) - w(e) = (w(e) + g(s(e))) - w(e) = g(s(e)),$$

which is independent of $w(e)$. Hence duplication leaves the disagreement part unchanged:

$$f(s(e), w(e) + 1) - (w(e) + 1) = g(s(e)) = f(s(e), w(e)) - w(e).$$

□

Proof of Lemma 1 Let S_1^*, \dots, S_k^* be the initial labelings of strong edges. Let F be the multiset of all strong edge instances in S^* that participate in a disagreement. That is, $e \in F_i = F \cap E_i$ if $e \in S_i^*$ (edge e is strong in layer i) and there exists some layer $j \neq i$ such that $e \in E_j \setminus S_j^*$ (edge e is weak in layer j). The size $|F| = \sum_{i \in [k]} |F_i|$ represents the total count of such strong edge instances involved in disagreements. By definition, the disagreement score $d_k(S_1^*, \dots, S_k^*) = \sum_{e \in E} 1_{\{e\}} \cdot s(e)$, where $s(e)$ is the number of layers in which edge e is strong, and $1_{\{e\}} = 1$ if e is weak in at least one layer. This sum precisely counts each strong instance of an edge that is part of a disagreement. Therefore, $d_k(S_1^*, \dots, S_k^*) = |F|$.

Let $W_i^* = E_i \setminus S_i^*$ be the set of weak edges in layer i for the initial labeling S^* . The objective value for S^* is $C^* = \sum_{i \in [k]} |W_i^*| + d_k(S_1^*, \dots, S_k^*) = \sum_{i \in [k]} |W_i^*| + |F|$.

We construct a new set of strong edge labelings H_1^*, \dots, H_k^* by setting $H_i^* = S_i^* \setminus F_i$ for all $i \in [k]$. Consider an edge e . If e was not involved in any disagreement in S^* , its labeling remains unchanged in H^* . If e was involved in a disagreement in S^* , then all its strong instances (i.e., $e \in F_i$ for relevant i) are made weak in H^* . Thus, in H^* , such an edge e is now weak in all layers where it was previously strong and involved in F , and it remains weak in layers where it was already weak. Consequently, the labeling H^* is consistent, which means $d_k(H_1^*, \dots, H_k^*) = 0$.

Let $W'_i = E_i \setminus H_i^*$ be the set of weak edges in layer i for the new labeling H^* . Then $W'_i = E_i \setminus (S_i^* \setminus F_i)$. Since $F_i \subseteq S_i^*$, this can be written as $W'_i = (E_i \setminus S_i^*) \cup F_i = W_i^* \cup F_i$. As F_i contains edges that were strong in S_i^* (and thus not in W_i^*), the sets W_i^* and F_i are disjoint for each layer i . Therefore, the total number of weak edges in the labeling H^* is $\sum_{i \in [k]} |W'_i| = \sum_{i \in [k]} |W_i^* \cup F_i| = \sum_{i \in [k]} (|W_i^*| + |F_i|) = \left(\sum_{i \in [k]} |W_i^*| \right) + |F|$. Thus, the objective value for H^* for MINMULTILAYERSTC is

$$\sum_{i \in [k]} |W'_i| + d_k(H_1^*, \dots, H_k^*) = \left(\sum_{i \in [k]} |W_i^*| \right) + |F| + 0 = C^*.$$

□

Proof of Theorem 2 By Lemma 1, it suffices to consider consistent labelings when solving MINMULTILAYERSTC. Let $G = (V, E_1, \dots, E_k)$ be the multilayer graph and

$W = (V_W, E_W)$ the combined wedge graph constructed in Algorithm 1, where each node $n_e \in V_W$ corresponds to a unique edge $e \in \bigcup_j E_j$, and $(n_{e_1}, n_{e_2}) \in E_W$ if e_1 and e_2 form an open wedge in some layer. Node weights $\omega(n_e)$ are set to the number of layers in which e appears. We reduce MINMULTILAYERSTC to a weighted vertex cover (MWVC) problem on W :

- Given a vertex cover $C \subseteq V_W$, label each e as *weak* in all layers if $n_e \in C$, and *strong* otherwise. Since C covers all edges in W , every open wedge is blocked by a weak edge, ensuring the STC property in all layers. The cost of the labeling is $\sum_{n_e \in C} \omega(n_e)$.
- Conversely, any consistent STC solution (S_1^*, \dots, S_k^*) defines a vertex cover $C^* = \{n_e \mid e \text{ is weak in all layers}\}$. If C^* were not a cover, some wedge would remain with both edges labeled *strong*, violating STC. Thus, solving MWVC on W is equivalent to solving MINMULTILAYERSTC under the consistency constraint. An α -approximation for MWVC yields an α -approximation for MINMULTILAYERSTC.

Constructing W takes $\mathcal{O}(knm)$ time, as each layer contributes $\mathcal{O}(nm)$ wedges. Assigning weights and extracting the final labeling is asymptotically dominated. Thus, the total running time is in $\mathcal{O}(knm + T_{MWVC})$. □

Proof of Lemma 2 The proof follows the same line of reasoning as the proof of Lemma 1. If an optimal solution S^* (with new edges E^N) has disagreements $d_k(S_1^*, \dots, S_k^*) > 0$, we can construct a consistent solution H^* by changing all *strong* instances of disagreeing edges in S^* to *weak*. This increases the count of *weak* edges by exactly $d_k(S_1^*, \dots, S_k^*)$ while reducing the disagreement penalty to zero, resulting in the same objective value for H^* while the STC properties are preserved. □

Proof of Lemma 3 Due to Lemma 2, we know that there is a (optimal) solution for the problem that does not contain any disagreement. Assume that there exists an α -approximation for CONSTRAINTMMLSTC+. Namely, we can find a solution for CONSTRAINTMMLSTC+ that is at most α times worse than the optimal solution. Now, for solving Problem 3, if we insert a new *weak* edge in one layer for its corresponding wedge then we have to insert the same *weak* edge in all the layers $\mathcal{L} \subseteq [k]$ in which a corresponding wedge exists (due to (3) in Problem 3). However, it might not be required that the corresponding wedge gets closed by the new *weak* edge in each such layer. To be specific, adding a new *weak* edge in all layers \mathcal{L} , might not improve the labeling for those layers. Therefore, we might already have a *weak* edge for the corresponding wedge, thus by adding an additional *weak* edge, we have two *weak* edges, i.e., in total the number of *weak* edges is at most doubled. Therefore, the solution we get for MINMULTILAYERSTC+ is at most 2α times worse than the optimal solution. □

Proof of Theorem 3 Let $G = (V, E_1, \dots, E_k)$ be the input multilayer graph. Algorithm 2 constructs a combined 3-uniform wedge hypergraph $W = (V_W, E_W)$. Nodes

$n_e \in V_W$ represent all possible edges $e \in \binom{V}{2}$. Hyperedges $\{n_{vu}, n_{vx}, n_{ux}\} \in E_W$ correspond to open wedges $(v, \{u, x\})$ in any layer G_i . The weights $\omega(n_e)$ are defined such that an optimal MWVC on W corresponds to an optimal solution to CONSTRAINTMMLSTC+ (Problem 3), assuming consistent labelings as per Lemma 2. Specifically, $\omega(n_e)$ for an existing edge $e \in E = \bigcup E_j$ is its multiplicity. For a potential new edge $e' \notin E$, $\omega(n_{e'})$ is the number of layers where adding e' would close an open wedge, implicitly reflecting the cost under constraint (3) of the variant (Problem 3).

The algorithm uses an α -approximation solver to find a MWVC $C \subseteq V_W$ on the hypergraph W . The solution derived from C thus constitutes an α -approximate solution for CONSTRAINTMMLSTC+.

By Lemma 3, an α -approximation for CONSTRAINTMMLSTC+ provides a 2α -approximation for the original MINMULTILAYERSTC+ problem (Problem 2). Therefore, the solution obtained from C is a 2α -approximation for MINMULTILAYERSTC+.

For the running time analysis, the number of vertices in W is $\mathcal{O}(n^2)$, and the number of hyperedges is $\mathcal{O}(k \cdot nm)$ (as each layer contributes $\mathcal{O}(nm)$ potential wedges). Constructing W takes $\mathcal{O}(k \cdot nm)$. If the MWVC algorithm (e.g., pricing for hypergraphs) runs in time linear to the hypergraph size, $T_{MWVC_H} = \mathcal{O}(k \cdot nm)$. Thus, the total running time is dominated by $\mathcal{O}(k \cdot nm)$. □

Exact ILP solutions

In order to solve our multilayer STC and STC+ variants exactly, we develop corresponding ILP formulations. First, we define the minimization ILP for MINMULTILAYERSTC, where we use binary variables y_{ij} which encode whether edge $\{i, j\} \in E$ is *strong* ($y_{ij} = 0$) or *weak* ($y_{ij} = 1$), and m_{ij} which denotes the multiplicity of edge $\{i, j\}$, i.e., the number of layers E_ℓ such that $\{i, j\} \in E_\ell$. We can define the ILP as follows because Lemma 1 ensures the existence of an optimal solution without disagreements:

$$\min \sum_{\{i,j\} \in E} y_{ij} \cdot m_{ij} \tag{1}$$

$$\text{s.t. } y_{ij} + y_{ih} \geq 1 \quad \forall (i, \{j, h\}) \in \mathcal{W}(E^\ell) \text{ and } \ell \in [k] \tag{2}$$

$$y_{ij} \in \{0, 1\} \quad \forall \{i, j\} \in E. \tag{3}$$

For the case of the MINMULTILAYERSTC+ problem, we first introduce a binary quadratic program which then can be linearized. We use binary variables y_{ij}^ℓ which encode whether a (possibly new) edge $\{i, j\} \in E_\ell \cup E_\ell^+$ is *strong* ($y_{ij}^\ell = 0$) or *weak* ($y_{ij}^\ell = 1$). Additionally, we use variables u_{ij}^ℓ that encode whether edge $\{i, j\}$ exists in layer $\ell \in [k]$ with $u_{ij}^\ell = 1$ if edge $\{i, j\} \in E_\ell \cup E_\ell^+$, and $u_{ij}^\ell = 0$, otherwise. Clearly,

the product $u_{ij}^\ell y_{ij}^\ell$ is one if and only if edge $\{i, j\} \in E_\ell \cup E_\ell^+$ will be a *weak* edge in the final result, leading to the quadratic program:

$$\min \sum_{\ell \in [k]} \sum_{ij \in \binom{V}{2}} u_{ij}^\ell \cdot y_{ij}^\ell \tag{4}$$

$$\text{s.t. } u_{ij}^\ell y_{ij}^\ell + u_{ih}^\ell y_{ih}^\ell + u_{jh}^\ell y_{jh}^\ell \geq 1 \tag{5}$$

$$\begin{aligned} & \forall (i, \{j, h\}) \in \mathcal{W}(E_\ell) \text{ and } \ell \in [k] \\ y_{ij}^\ell - y_{ij}^h = 0 & \quad \forall \{i, j\} \in \binom{V}{2} \text{ and } 1 \leq \ell \leq h \leq k \end{aligned} \tag{6}$$

$$u_{ij}^\ell = 1 \quad \forall \{i, j\} \in E_\ell \text{ and } \ell \in [k] \tag{7}$$

$$y_{ij}^\ell \geq u_{ij}^\ell \quad \forall \{i, j\} \in \binom{V}{2} \setminus E_\ell \text{ and } \ell \in [k] \tag{8}$$

$$y_{ij}^\ell \in \{0, 1\} \quad \forall \{i, j\} \in \binom{V}{2} \text{ and } \ell \in [k] \tag{9}$$

$$u_{ij}^\ell \in \{0, 1\} \quad \forall \{i, j\} \in \binom{V}{2} \text{ and } \ell \in [k]. \tag{10}$$

Eq. 5 ensures that for each wedge in any of the layers either one of the wedge edges are *weak*, or a new edge is inserted. Moreover, Eq. 6 guarantees that if an edge is *weak* in one layer, it is *weak* in all layers. Equation 7 forces all edges that exist in the multilayer graph to exist in the solution. Finally, Eq. 8 ensures that all newly inserted edges are labeled *weak*. By linearization of the binary products, we obtain the ILP:

$$\min \sum_{\ell \in [k]} \sum_{ij \in \binom{V}{2}} z_{ij}^\ell \tag{11}$$

$$\begin{aligned} \text{s.t. } z_{ij}^\ell + z_{ih}^\ell + z_{jh}^\ell & \geq 1 \quad \forall (i, \{j, h\}) \in \mathcal{W}(E_\ell) \text{ and } \ell \in [k] \\ z_{ij}^\ell & \leq y_{ij}^\ell \quad \{i, j\} \in \binom{V}{2} \text{ and } \ell \in [k] \end{aligned} \tag{12}$$

$$z_{ij}^\ell \leq u_{ij}^\ell \quad \{i, j\} \in \binom{V}{2} \text{ and } \ell \in [k] \tag{13}$$

$$z_{ij}^\ell \geq y_{ij}^\ell + u_{ij}^\ell - 1 \quad \{i, j\} \in \binom{V}{2} \text{ and } \ell \in [k] \tag{14}$$

$$y_{ij}^\ell - y_{ij}^h = 0 \quad \forall \{i, j\} \in \binom{V}{2} \text{ and } 1 \leq \ell \leq h \leq k \tag{15}$$

$$u_{ij}^\ell = 1 \quad \forall \{i, j\} \in E_\ell \text{ and } \ell \in [k] \tag{16}$$

$$y_{ij}^\ell \geq u_{ij}^\ell \quad \text{for all } \{i, j\} \in \binom{V}{2} \setminus E_\ell \text{ and } \ell \in [k] \tag{17}$$

$$y_{ij}^{\ell} \in \{0, 1\} \quad \forall \{i, j\} \in \binom{V}{2} \text{ and } \ell \in [k] \quad (18)$$

$$u_{ij}^{\ell} \in \{0, 1\} \quad \forall \{i, j\} \in \binom{V}{2} \text{ and } \ell \in [k] \quad (19)$$

$$z_{ij}^{\ell} \in \{0, 1\} \quad \forall \{i, j\} \in \binom{V}{2} \text{ and } \ell \in [k]. \quad (20)$$

Details of the datasets

We provide the details of the eight dataset.

- *AUCS* (5 layers) contains five different online and offline relations (working relationship, leisure activities, regular lunch, co-authorship, and Facebook friendship) among faculty members of the CS department at Aarhus University (Magnani et al. 2013).
- *Hospital* (5 layers) is a face-to-face contact network between hospital patients and health care workers (Vanhems et al. 2013), where each layer represents one day of recorded interactions.
- *Airports* (37 layers) is an aviation transport network of European airports (Cardillo et al. 2013), where layers correspond to different airline companies.
- *Rattus* (6 layers) contains different types of genetic interactions of *Rattus norvegicus* (De Domenico et al. 2015a), with each layer representing a distinct interaction type.
- *FfTwyT* (3 layers) is a multilayer social network linking user accounts across *FriendFeed*, *Twitter*, and *YouTube* (Dickison et al. 2016), where each platform forms one layer.
- *Knowledge* (30 layers) is derived from the *FB15K-237* knowledge graph (Toutanova et al. 2015), where nodes are entities and each layer corresponds to a distinct relation type.
- *HomoSap* (7 layers) represents different types of genetic and protein interactions between human genes curated from the BioGRID database (De Domenico et al. 2015b). The layers correspond to: Physical Association, Association, Direct Interaction, Colocalization, Additive Genetic Interaction, Suppression Genetic Interaction, and Synthetic Genetic Interaction. Each layer captures a distinct biological interaction mechanism between genes or proteins.
- *DBLP* (168 layers) is a subgraph of the DBLP bibliography (Ley 2002) restricted to *A* and *A** ranked conferences (CORE ranking). Each layer corresponds to one conference, and edges represent co-authorships within that venue. Across all datasets, tie strength is defined structurally via STC. We do not impose domain-specific semantics on strong versus weak ties and we assume that tie strength is semantically comparable across layers.

Author contributions All authors contributed to the conception and design of the study, the development of theoretical results and algorithms, and the writing and revision of the manuscript. Implementation and experimentation were primarily carried out by LO. All authors reviewed the manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adriaens F, Bie TD, Gionis A et al (2020) Relaxing the strong triadic closure problem for edge strength inference. *Data Min Knowl Discov* 34(3):611–651. <https://doi.org/10.1007/S10618-020-00673-0>
- Arachchi CW, Kumpulainen I, Tatti N (2024) Dense subgraph discovery meets strong triadic closure. In: Baeza-Yates R, Bonchi F (eds) *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, pp 3334–3344. <https://doi.org/10.1145/3637528.3671697>
- Bar-Yehuda R, Even S (1981) A linear-time approximation algorithm for the weighted vertex cover problem. *J Algorithms* 2(2):198–203. [https://doi.org/10.1016/0196-6774\(81\)90020-1](https://doi.org/10.1016/0196-6774(81)90020-1)
- Boccaletti S, Bianconi G, Criado R et al (2014) The structure and dynamics of multilayer networks. *Phys Rep* 544(1):1–122. <https://doi.org/10.1016/j.physrep.2014.07.001>
- Cardillo A, Gómez-Gardenes J, Zanin M et al (2013) Emergence of network features from multiplexity. *Sci Rep* 3(1):1344. <https://doi.org/10.1038/srep01344>
- Cheng X, Yang C, Zhao Y, et al (2025) BTS: A comprehensive benchmark for tie strength prediction. In: Antonie L, Pei J, Yu X, et al (eds) *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, pp 5345–5354. <https://doi.org/10.1145/3711896.3737441>
- Crainic TG, Gendron B, Kazemzadeh MRA (2022) A taxonomy of multilayer network design and a survey of transportation and telecommunication applications. *Eur J Oper Res* 303(1):1–13. <https://doi.org/10.1016/j.ejor.2021.12.028>
- De Domenico M (2023) More is different in real-world multilayer networks. *Nat Phys* 19(9):1247–1262. <https://doi.org/10.1038/s41567-023-02132-1>
- De Domenico M, Solé-Ribalta A, Cozzo E et al (2013) Mathematical formulation of multilayer networks. *Phys Rev X* 3(4):041022. <https://doi.org/10.1103/PhysRevX.3.041022>
- De Domenico M, Nicosia V, Arenas A et al (2015) Structural reducibility of multilayer networks. *Nat Commun* 6(1):6864. <https://doi.org/10.1038/ncomms7864>
- De Domenico M, Porter MA, Arenas A (2015) Muxviz: a tool for multilayer analysis and visualization of networks. *Journal of Complex Networks* 3(2):159–176. <https://doi.org/10.1093/comnet/cnu038>
- Dickson ME, Magnani M, Rossi L (2016) *Multilayer Social Networks*. Cambridge University Press Cambridge. <https://doi.org/10.1017/CBO9781139941907>
- Easley DA, Kleinberg JM (2010) *Networks, Crowds, and Markets - Reasoning About a Highly Connected World*. Cambridge University Press Cambridge. <https://doi.org/10.1017/CBO9780511761942>
- Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Jr. DRO, Arthur RB, Hinckley K, et al (eds) *Proceedings of the 27th International Conference on Human Factors in Computing Systems*. CHI 2009, Boston, MA, USA, April 4–9, 2009. ACM, New York, NY, USA, pp 211–220. <https://doi.org/10.1145/1518701.1518736>

- Gomes FC, Meneses CN, Pardalos PM et al (2006) Experimental analysis of approximation algorithms for the vertex cover and set covering problems. *Computers & Operations Research* 33(12):3520–3534. <https://doi.org/10.1016/j.cor.2005.03.030>
- Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380. <https://doi.org/10.1086/225469>
- Grüttemeier N, Komusiewicz C (2020) On the relation of strong triadic closure and cluster deletion. *Algorithmica* 82:853–880. <https://doi.org/10.1007/s00453-019-00617-1>
- Hammoud Z, Kramer F (2020) Multilayer networks: aspects, implementations, and application in bio-medicine. *Big Data Analytics* 5(1):2. <https://doi.org/10.1186/s41044-020-00046-0>
- Jones JJ, Settle JE, Bond RM et al (2013) Inferring tie strength from online directed behavior. *PLoS ONE* 8(1):e52168. <https://doi.org/10.1371/journal.pone.0052168>
- Kahanda I, Neville J (2009) Using transactional information to predict link strength in online social networks. In: Adar E, Hurst M, Finin T, et al (eds) *Proceedings of the Third International Conference on Weblogs and Social Media, ICWSM*. The AAAI Press, Washington, DC, USA, pp 74–81, <https://doi.org/10.1609/icwsml.v3i1.13957>
- Kivelä M, Arenas A, Barthelemy M et al (2014) Multilayer networks *Journal of complex networks* 2(3):203–271. <https://doi.org/10.1093/comnet/cnu016>
- Konstantinidis AL, Papadopoulos C (2020) Maximizing the strong triadic closure in split graphs and proper interval graphs. *Discret Appl Math* 285:79–95. <https://doi.org/10.1016/j.dam.2020.05.035>
- Konstantinidis AL, Nikolopoulos SD, Papadopoulos C (2018) Strong triadic closure in cographs and graphs of low maximum degree. *Theor Comput Sci* 740:76–84. <https://doi.org/10.1016/j.tcs.2018.05.012>
- Ley M (2002) The DBLP computer science bibliography: Evolution, research issues, perspectives. In: Laender AHF, Oliveira AL (eds) *String Processing and Information Retrieval, 9th International Symposium, SPIRE*. Springer, Berlin, Heidelberg, pp 1–10, https://doi.org/10.1007/3-540-45735-6_1
- Lü L, Zhou T (2010) Link prediction in weighted networks: The role of weak ties. *Europhys Lett* 89(1):18001. <https://doi.org/10.1209/0295-5075/89/18001>
- Magnani M, Micenkova B, Rossi L (2013) Combinatorial analysis of multiple networks. *arXiv preprint arXiv:1303.4986*<https://doi.org/10.48550/arXiv.1303.4986>
- Matakos A, Gionis A (2022) Strengthening ties towards a highly-connected world. *Data Min Knowl Discov* 36(1):448–476. <https://doi.org/10.1007/s10618-021-00812-1>
- Oettershagen L, Konstantinidis AL, Italiano GF (2022) Inferring tie strength in temporal networks. In: Amini M, Canu S, Fischer A, et al (eds) *European Conference on Machine Learning and Knowledge Discovery in Databases, ECML PKDD*. Springer International Publishing, Cham, pp 69–85, https://doi.org/10.1007/978-3-031-26390-3_5
- Oettershagen L, Konstantinidis AL, Italiano GF (2025) Inferring tie strength in temporal networks. *Data Min Knowl Discov* 39(2):1–31. <https://doi.org/10.1007/s10618-025-01093-8>
- Pappalardo L, Rossetti G, Pedreschi D (2013) Measuring tie strength in multidimensional networks. In: Buccafurri F, Saccà D (eds) *21st Italian Symposium on Advanced Database Systems, SEBD*. Curran Associates, Red Hook, NY, USA, pp 223–230
- Pham H, Shahabi C, Liu Y (2016) Inferring social strength from spatiotemporal data. *ACM Trans Database Syst* 41(1):1–47. <https://doi.org/10.1145/2877200>
- Rozenshtein P, Tatti N, Gionis A (2017) Inferring the strength of social ties: A community-driven approach. In: *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, pp 1017–1025, <https://doi.org/10.1145/3097983.3098199>
- Sintos S, Tsaparas P (2014) Using strong triadic closure to characterize ties in social networks. In: Macskassy SA, Perlich C, Leskovec J, et al (eds) *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, New York, NY, USA, pp 1466–1475, <https://doi.org/10.1145/2623330.2623664>
- Stolz S, Schlereth C (2021) Predicting tie strength with ego network structures. *J Interact Mark* 54(1):40–52. <https://doi.org/10.1016/j.intmar.2020.10.001>
- Toutanova K, Chen D, Pantel P, et al (2015) Representing text for joint embedding of text and knowledge bases. In: Márquez L, Callison-Burch C, Su J (eds) *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Lisbon, Portugal, pp 1499–1509, <https://doi.org/10.18653/v1/D15-1174>
- Vanhems P, Barrat A, Cattuto C et al (2013) Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS ONE* 8(9):e73970. <https://doi.org/10.1371/journal.pone.0073970>

- Vazirani VV (2001) *Approximation algorithms*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-04565-7>
- Veldt N (2022) Correlation clustering via strong triadic closure labeling: Fast approximation algorithms and practical lower bounds. In: Chaudhuri K, Jegelka S, Song L, et al (eds) *Proceedings of the 39th International Conference on Machine Learning*, pp 22060–22083
- Xiang R, Neville J, Rogati M (2010) Modeling relationship strength in online social networks. In: *Proceedings of the 19th International Conference on World Wide Web*, pp 981–990, <https://doi.org/10.1145/1772690.1772790>

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