TRIÈST: Counting Local and Global Triangles in Fully-dynamic Streams with Fixed Memory Size

Lorenzo De Stefani*, Alessandro Epasto*, Matteo Riondato†, Eli Upfal*
{lorenzo,epasto,eli}@cs.brown.edu, matteo@twosigma.com

February 12, 2016

"Ogni lassada xe persa"[1] – Proverb from Trieste, Italy.

Abstract

We present TRIÈST, a suite of one-pass streaming algorithms to compute unbiased, low-variance, high-quality approximations of the global and local (i.e., incident to each vertex) number of triangles in a fully-dynamic graph represented as an adversarial stream of edge insertions and deletions. Our algorithms use reservoir sampling and its variants to exploit the user-specified memory space at all times. This is in contrast with previous approaches which use hard-to-choose parameters (e.g., a fixed sampling probability) and offer no guarantees on the amount of memory they will use. We show a full analysis of the variance of the estimations and novel concentration bounds for these quantities.

Our experimental results on very large graphs show that TRIÈST outperforms state-of-the-art approaches in accuracy and exhibits a small update time.

1 Introduction

Exact computation of characteristic quantities of Web-scale networks is often impractical or even infeasible due to the humongous size of these graphs. In these cases, it is natural to resort to *efficient-to-compute approximations* of these quantities that, when of sufficiently high quality, can be used as proxies for the exact values.

In addition to being huge, many interesting networks are *fully-dynamic* and can be represented as a *stream* whose elements are edges/nodes insertions and deletions which occur in an *arbitrary* (even adversarial) order. Characteristic quantities in these graphs are *intrinsically volatile*, hence there is limited added value in maintaining them exactly. Rather, one is interested in keeping track, *at all times*, of a high-quality approximation of these quantities. For efficiency, the algorithms should *aim at exploiting the available memory space as much as possible* and they should *require only one pass over the stream*.

We introduce TRIÈST, a suite of *sampling-based, one-pass algorithms for adversarial fully-dynamic streams to approximate the global number of triangles and the local number of triangles incident to each vertex*, which are primitives with many applications (e.g., community detection [4], topic mining [10], spam/anomaly detection [3, 26], and the analysis of protein interaction networks [28].)

Many previous works on triangle estimation in streams also employ sampling (see Sect. 3), but they usually require the user to specify *in advance* a *edge sampling probability* p that is fixed for the entire stream. This approach presents several significant drawbacks. First, choosing a p that allows to obtain the desired approximation quality requires to know or guess a number of properties of the input (e.g., the size of the stream). Second, a fixed p implies that the sample size grows with the size of the stream, which is problematic when the stream size is not known in advance: if the user specifies a large p, the algorithm may run out of

*Department of Computer Science – Brown University.
†Two Sigma Investments, LLC. Part of the work done while affiliated to Brown University.
[1]Any missed chance is lost forever.

memory, while for a smaller p it will provide a suboptimal estimation. Third, even assuming to be able to compute a p that ensures (in expectation) full use of the available space, the memory would be fully utilized only the end of the stream, and the estimations computed throughout the execution would be suboptimal.

Contributions. We address all the above issues by taking a significant departure from the fixed-probability, independent edge sampling approach taken even by state-of-the-art methods. Specifically:

- We introduce TRIÈST (*TRI*angle *E*stimation from *ST*reams), a suite of *one-pass streaming algorithms* to approximate, at each time instant, the global and local number of triangles in a *fully-dynamic* graph stream (i.e., a sequence of edges additions and deletions in arbitrary order) using a *fixed amount of memory*. This is the first contribution that enjoys all these properties. TRIÈST only requires the user to specify *the amount of available memory*, an interpretable parameter that is definitively known to the user.
- Our algorithms maintain a sample of edges, using the *reservoir sampling* [38] and *random pairing* [13] sampling schemes to exploit the available memory as much as possible. To the best of our knowledge, ours is the first application of these techniques to subgraph counting in fully-dynamic, arbitrarily long, adversarially ordered streams. We present an analysis of the unbiasedness and of the variance of our estimators, and establish strong concentration results for them. The use of reservoir sampling and random pairing requires additional sophistication in the analysis, as the presence of an edge in the sample is not independent from the concurrent presence of another edge. Hence, in our proofs we must consider the complex dependencies in events involving sets of edges. The gain is worth the effort: we prove that the variance of our algorithms is smaller than that of state-of-the-art methods [26], and this is confirmed by our experiments.
- We conduct an extensive experimental evaluation of TRIÈST on very large graphs, some with billions of edges, comparing the performances of our algorithms to those of existing state-of-the-art contributions [17, 26, 32]. *Our algorithms significantly and consistently reduce the average estimation error by up to* 90% w.r.t. the state of the art, both in the global and local estimation problems, while using the same amount of memory. Our algorithms are also extremely scalable, showing update times in the order of hundreds of microseconds for graphs with billions of edges.

Paper organization. We formally introduce the settings and the problem in Sect. 2. In Sect. 3 we discuss related works. We present and analyze our algorithm TRIÈST and discuss our design choices in Sect. 4. The results of our experimental evaluation are presented in Sect. 5. We draw our conclusions and outline direction for future work in Sect. 6. The proofs of our theoretical results can be found in Appendix A.

2 Preliminaries

We study the problem of counting global and local triangles in a fully-dynamic undirected graph as an arbitrary (adversarial) stream of edge insertions and deletions.

Formally, for any (discrete) time instant $t \geq 0$, let $G^{(t)} = (V^{(t)}, E^{(t)})$ be the graph observed up to and including time t. At time $t = 0$ we have $V^{(t)} = E^{(t)} = \emptyset$. For any $t > 0$, at time $t+1$ we receive an element $e_{t+1} = (\bullet, (u, v))$ from a stream, where $\bullet \in \{+, -\}$ and u, v are two distinct vertices. The graph $G^{(t+1)} = (V^{(t+1)}, E^{(t+1)})$ is obtained by *inserting a new edge or deleting an existing edge* as follows:

$$E^{(t+1)} = \begin{cases} E^{(t)} \cup \{(u, v)\} & \text{if } \bullet = \text{“}+\text{”} \\ E^{(t)} \setminus \{(u, v)\} & \text{if } \bullet = \text{“}-\text{”} \end{cases} .$$

If u or v do not belong to $V^{(t)}$, they are added to $V^{(t+1)}$. Nodes are deleted from $V^{(t)}$ when they have degree zero.

Edges can be added and deleted in the graph in an arbitrary adversarial order: i.e., as to cause the worst outcome for the algorithm, but we assume that the adversary has *no access to the random bits* used by the algorithm. We assume that *all operations have effect*: if $e \in E^{(t)}$ (resp. $e \notin E^{(t)}$), $(+, e)$ (resp. $(-, e)$) can not be on the stream at time $t + 1$.

Given a graph $G^{(t)} = (V^{(t)}, E^{(t)})$, a *triangle* in $G^{(t)}$ is a *set* of three vertices $\{u, v, w\} \subseteq V^{(t)}$ such that $\{(u, v), (v, w), (w, u)\} \subseteq E^{(t)}$. We refer to the vertices forming a triangle as its *corners*. We denote with $\Delta^{(t)}$ the set of *all* triangles in $G^{(t)}$, and, for any vertex $u \in V^{(t)}$, with $\Delta_u^{(t)}$ the subset of $\Delta^{(t)}$ containing all and only the triangles that have u as a corner.

Problem definition. We study the *Global* (resp. *Local*) Triangle Counting Problem in Fully-dynamic Streams, which requires to compute, at each time $t \geq 0$ an estimation of $|\Delta^{(t)}|$ (resp. for each $u \in V$ an estimation of $|\Delta_u^{(t)}|$).

3 Related work

The literature on exact and approximate triangle counting is extremely rich, including exact algorithms, graph sparsifiers [36, 37], complex-valued sketches [19, 27], and MapReduce algorithms [30, 31, 34]. Here we restrict the discussion to the works most related to ours, i.e., to those presenting algorithms for counting or approximating the number of triangles from data streams. We refer to the survey by Latapy [24] for an in-depth discussion of other works.

Many previous contributions presented algorithms for more restricted (i.e., less generic) settings than ours, or for which the constraints on the computation are more lax [2, 6, 18, 21]. For example, Becchetti et al. [3] and Kolountzakis et al. [20] present algorithms for approximate triangle counting from *static* graphs by performing multiple passes over the data. Pavan et al. [32] and Jha et al. [17] propose algorithms for approximating only the global number of triangles from *edge-insertion-only* streams. Kutzkov and Pagh [22] present a one-pass algorithm for fully-dynamic graphs, but the triangle count estimation is (expensively) computed only at the end of the stream and the algorithm requires, in the worst case, more memory than what is needed to store the entire graph. Ahmed et al. [1] apply the sampling-and-hold approach to insertion-only graph stream mining to obtain, only at the end of the stream and using non-constant space, an estimation of many network measures including triangles.

None of these works has *all* the features offered by TRIÈST: performs a single pass over the data, handles fully-dynamic streams, uses a fixed amount of memory space, requires a single interpretable parameter, and returns an estimation at each time instant. Furthermore, our experimental results show that we outperform the algorithms from [17, 32] on insertion-only streams.

Lim and Kang [26] present an algorithm for insertion-only streams that is based on independent edge sampling with a fixed probability. Since the memory is not fully utilized during most of the stream, the variance of the estimate is large. Our approach handles fully-dynamic streams and makes better use of the available memory space at each time instant, resulting in a better estimation, as shown by our analytical and experimental results.

Vitter [38] presents a detailed analysis of the reservoir sampling scheme and discusses methods to speed up the algorithm by reducing the number of calls to the random number generator. Random Pairing [13] is an extension of reservoir sampling to handle fully-dynamic streams with insertions and deletions. Cohen et al. [8] generalize and extend the Random Pairing approach to the case where the elements on the stream are key-value pairs, where the value may be negative (and less than -1). In our settings, where the value is not less than -1 (for an edge deletion), these generalizations do not apply and the algorithm presented by Cohen et al. [8] reduces essentially to Random Pairing.

4 Algorithms

We present three novel algorithms for approximate global and local triangle counting from edge streams. The first two work on insertion-only streams, while the third can handle fully-dynamic streams where edge deletions are allowed.

Parameters. Our algorithms keep an edge sample $\mathcal{S}$ containing up to M edges from the stream, where M is a positive integer parameter. For ease of presentation we realistically assume $M \geq 6$. In Sect. 1 we motivated the design choice of only requiring M as a parameter and remarked its advantages over using a fixed sampling probability p. Our algorithms are designed to use the available space as much as possible.

Counters. Our algorithms keep *counters* which are used for computing the estimations of the global and local number of triangles. They *always* keep one global counter τ for the estimation of the global number of triangles. Only the global counter is needed to estimate the total triangle count. To estimate the local triangle counts, we keep a set of local counters τ_u for a subset of the nodes $u \in V^{(t)}$. The local counters are created on the fly as needed, and *always* destroyed as soon as they have a value of 0. Hence our algorithms use $O(M)$ space (with one exception, see Sect. 4.2).

Notation. For any $t \geq 0$, let $G^{\mathcal{S}} = (V^{\mathcal{S}}, E^{\mathcal{S}})$ be the subgraph of $G^{(t)}$ containing all and only the edges in the current sample $\mathcal{S}$. We denote with $\mathcal{N}_u^{\mathcal{S}}$ the *neighborhood* of u in $G^{\mathcal{S}}$: $\mathcal{N}_u^{\mathcal{S}} = \{v \in V^{(t)} \ : \ (u,v) \in \mathcal{S}\}$ and with $\mathcal{N}_{u,v}^{\mathcal{S}} = \mathcal{N}_u^{\mathcal{S}} \cap \mathcal{N}_v^{\mathcal{S}}$ the *shared neighborhood* of u and v in $G^{\mathcal{S}}$.

4.1 A first algorithm – TRIÈST-BASE

We first present TRIÈST-BASE, which works on insertion-only streams and uses standard reservoir sampling [38] to maintain the edge sample $\mathcal{S}$:

- If $t \leq M$, then the edge $e_t = (u,v)$ on the stream at time t is deterministically inserted in $\mathcal{S}$.
- If $t > M$, we flip a biased coin with heads probability M/t. If the outcome is tails, $\mathcal{S}$ is not modified. Otherwise, we choose an edge $(w,z) \in \mathcal{S}$ uniformly at random, remove (w,z) from $\mathcal{S}$, and insert (u,v) in $\mathcal{S}$.

After each insertion (resp. removal) of an edge (u,v) from $\mathcal{S}$, TRIÈST-BASE calls the procedure UPDATE-COUNTERS that increments (resp. decrements) τ, τ_u and τ_v by $|\mathcal{N}_{u,v}^{\mathcal{S}}|$, and τ_c by one, for each $c \in \mathcal{N}_{u,v}^{\mathcal{S}}$.

The pseudocode for TRIÈST-BASE is presented in Alg. 1.

Estimation. For any $t \geq 0$, let $\xi^{(t)} = \max\left\{1, \frac{t(t-1)(t-2)}{M(M-1)(M-2)}\right\}$. Let $\tau^{(t)}$ (resp. $\tau_u^{(t)}$) be the value of the counter τ at the end of time step t (i.e., after the edge on the stream at time t has been processed by TRIÈST-BASE) (resp. the value of the counter τ_u at the end of time step t if there is such a counter, 0 otherwise). When queried at the end of time t, TRIÈST-BASE returns $\xi^{(t)}\tau^{(t)}$ (resp. $\xi^{(t)}\tau_u^{(t)}$) as the estimation for the global (resp. local for $u \in V^{(t)}$) triangle count.

Analysis.

Theorem 1. *We have*

$$\xi^{(t)}\tau^{(t)} = \tau^{(t)} = |\Delta^{(t)}| \text{ if } t \leq M$$
$$\mathbb{E}\left[\xi^{(t)}\tau^{(t)}\right] = |\Delta^{(t)}| \text{ if } t > M \ .$$

Algorithm 1 TRIÈST-BASE

```
Input: Graph stream Σ, integer M ≥ 6
S ← ∅, t ← 0, τ ← 0
for each edge (u, v) from Σ do
    t ← t + 1
    if SampleEdge((u, v), t) then
        S ← S ∪ {(u, v)}
        UpdateCounters(+, (u, v))

function SampleEdge((u, v), t)
    if t ≤ M then
        return True
    else if FlipBiasedCoin(M/t) = heads then
        (u′, v′) ← random edge from S
        S ← S \ {(u′, v′)}
        UpdateCounters(−, (u′, v′))
        return True
    return False

function UpdateCounters((•, (u, v)))
    N^S_{u,v} ← N^S_u ∩ N^S_v
    for all c ∈ N^S_{u,v} do
        τ ← τ • 1
        τ_c ← τ_c • 1
        τ_u ← τ_u • 1
        τ_v ← τ_v • 1
```

The TRIÈST-BASE estimations are not only *unbiased in all cases*, but actually *exact for* $t \leq M$, i.e., for $t \leq M$, they are the true global/local number of triangles in $G^{(t)}$. We now analyze the variance of the estimation returned by TRIÈST-BASE for $t > M$ (the variance is 0 for $t \leq M$.)

Let $t \geq 0$. For any $u \in V^{(t)}$, let $r_u^{(t)}$ be the number of *unordered* pairs of distinct triangles from $\Delta_u^{(t)}$ that share an edge.[2] Similarly, let $r^{(t)} = \frac{1}{3}\sum_{u \in V^{(t)}} r_u^{(t)}$ be the *total* number of unordered pairs of distinct triangles from $\Delta^{(t)}$ that share an edge. We also define $w^{(t)} = \binom{|\Delta^{(t)}|}{2} - r^{(t)}$ as the number of unordered pairs of distinct triangles that do not share any edge, and analogously for $w_u^{(t)}$.

Theorem 2. *For any* $t > M$*, let* $f(t) = \xi^{(t)} - 1$*,*

$$g(t) = \xi^{(t)} \frac{(M-3)(M-4)}{(t-3)(t-4)} - 1$$

and

$$h(t) = \xi^{(t)} \frac{(M-3)(M-4)(M-5)}{(t-3)(t-4)(t-5)} - 1 \ \ (\leq 0).$$

We have:

$$\mathrm{Var}\left[\xi(t)\tau^{(t)}\right] = |\Delta^{(t)}| f(t) + r^{(t)} g(t) + w^{(t)} h(t). \tag{1}$$

In our proofs, we carefully account for the fact that, as we use reservoir sampling [38], the presence of an edge a in $\mathcal{S}$ is *not independent* from the concurrent presence of another edge b in $\mathcal{S}$. This is not the case for samples built using fixed-probability independent edge sampling. When computing the variance, we must consider not only pairs of triangles that share an edge (as for independent edge sampling approaches), but also pairs of triangles that share no edge, as their respective presences in the sample are not independent events. The gain is worth the additional sophistication needed in the analysis, because the contribution of triangles that do not share edges to the variance is *non-positive* ($h(t) \leq 0$), i.e., it reduces the variance. A comparison of the variance of our estimator with that obtained with a fixed-probability independent edge sampling approach, is discussed below.

Let $h^{(t)}$ denote the maximum number of triangles sharing a single edge in $G^{(t)}$. The following concentration theorem relies on *1.* a result by Hajnal and Szemerédi [14] on graph coloring, *2.* a novel concentration

[2] Two distinct triangles can share at most one edge.

result for fixed-probability independent edge sampling, and *3.* a Poisson-approximation-like result on probabilities of general events under reservoir sampling w.r.t. their probabilities under independent edge sampling. These ingredients are then combined to obtain the following result. The details can be found in our extended online version [9].

Theorem 3. *Let $t \geq 0$ and assume $|\Delta^{(t)}| > 0$.*[3] *For any $\varepsilon, \delta \in (0,1)$, let*

$$\Phi = \sqrt[3]{8\varepsilon^{-2}\frac{3h^{(t)}+1}{|\Delta^{(t)}|}\ln\left(\frac{(3h^{(t)}+1)e}{\delta}\right)}\ .$$

If

$$M \geq \max\left\{t\Phi\left(1+\frac{1}{2}\ln^{2/3}(t\Phi)\right), 12\varepsilon^{-1}+e^2, 25\right\},$$

then $|\xi^{(t)}\tau^{(t)} - |\Delta^{(t)}|| < \varepsilon|\Delta^{(t)}|$ with probability $> 1-\delta$.

Results analogous to those in Thms. 1, 2, and 3 hold for the local triangle count for any $u \in V^{(t)}$, replacing the global quantities with the corresponding local ones.

Comparison with fixed-probability approaches. We now compare the variance of TRIÈST-BASE to the variance of a fixed probability sampling approach (MASCOT-C [26]), which samples edges *independently* with a fixed probability p and uses the $p^{-3}|\Delta_{\mathcal{S}}|$ as the estimate for the global number of triangles at time t. As shown by Lim and Kang [26, Lemma 2], the variance of this estimator is

$$\mathrm{Var}[p^{-3}|\Delta_{\mathcal{S}}|] = |\Delta^{(t)}|\bar{f}(p) + r^{(t)}\bar{g}(p),$$

where $\bar{f}(p) = p^{-3}-1$ and $\bar{g}(p) = p^{-1}-1$.

Assume that we give MASCOT-C the additional information that the stream has finite length T, and assume we run MASCOT-C with $p = M/T$ so that the expected sample size at the end of the stream is M.[4] Let $\mathbb{V}^{(t)}_{\text{fix}}$ be the resulting variance of the MASCOT-C estimator at time t, and let $\mathbb{V}^{(t)}$ be the variance of our estimator at time t (see (1)). For $t \leq M$, $\mathbb{V}^{(t)} = 0$, hence $\mathbb{V}^{(t)} \leq \mathbb{V}^{(t)}_{\text{fix}}$.

For $M < t < T$, we can show the following result.

Lemma 1. *Let $0 < \alpha < 1$ be a constant. For any constant $M > \max(\frac{8\alpha}{1-\alpha}, 42)$ and any $t \leq \alpha T$ we have $\mathbb{V}^{(t)} < \mathbb{V}^{(t)}_{\text{fix}}$.*

For example, if we set $\alpha = 0.99$ and run TRIÈST-BASE with $M \geq 400$ and MASCOT-C with $p = M/T$, we have that TRIÈST-BASE has strictly smaller variance than MASCOT-C for 99% of the stream.

What about $t = T$? The difference between the definitions of $\mathbb{V}^{(t)}_{\text{fix}}$ and $\mathbb{V}^{(t)}$ is in the presence of $\bar{f}(M/T)$ instead of $f(t)$ (resp. $\bar{g}(M/T)$ instead of $g(t)$) as well as the additional term $w^{(t)}h(M,t) \leq 0$ in our $\mathbb{V}^{(t)}$. Let $M(T)$ be an arbitrary slowly increasing function of T. For $T \to \infty$ we can show that $\lim_{T\to\infty}\frac{\bar{f}(M(T)/T)}{f(T)} = \lim_{T\to\infty}\frac{\bar{g}(M(T)/T)}{g(T)} = 1$, hence, *informally*, $\mathbb{V}^{(T)} \to \mathbb{V}^{(T)}_{\text{fix}}$, for $T \to \infty$.

A similar discussion also holds for the method we present in Sect. 4.2, and explains the results of our experimental evaluations, which shows that our algorithms have strictly lower (empirical) variance than fixed probability approaches for most of the stream.

Update time. The time to process an element of the stream is dominated by the computation of the shared neighborhood $\mathcal{N}_{u,v}$ in UPDATECOUNTERS. A MERGESORT-based algorithm for the intersection requires $O(\deg(u) + \deg(v))$ time, where the degrees are w.r.t. the graph $G_{\mathcal{S}}$. By storing the neighborhood of each vertex in a Hash Table (resp. an AVL tree), the update time can be reduced to $O(\min\{\deg(v), \deg(u)\})$ (resp. amortized time $O(\min\{\deg(v), \deg(u)\} + \log\max\{\deg(v), \deg(u)\})$).

[3] If $|\Delta^{(t)}| = 0$, our algorithms correctly estimate 0 triangles.

[4] We are giving MASCOT-C a relevant advantage: if only space M were available, we should run MASCOT-C with a sufficiently smaller $p' < p$, otherwise there would be a constant probability that MASCOT-C would run out of memory.

4.2 Improved insertion algorithm - TRIÈST-IMPR

TRIÈST-IMPR is a variant of TRIÈST-BASE with small modifications that result in higher-quality (i.e., lower variance) estimations. The changes are:

1. UPDATECOUNTERS is called *unconditionally for each element on the stream*, before the algorithm decides whether or not to insert the edge into $\mathcal{S}$. W.r.t. the pseudocode in Alg. 1, this change corresponds to moving the call to UPDATECOUNTERS on line 6 to *before* the **if** block. MASCOT [26] uses a similar idea, but TRIÈST-IMPR is significantly different as we allow edges to be removed from the sample, since we use reservoir sampling.

2. TRIÈST-IMPR *never* decrements the counters when an edge is removed from $\mathcal{S}$. W.r.t. the pseudocode in Alg. 1, we remove the call to UPDATECOUNTERS on line 13.

3. UPDATECOUNTERS performs a *weighted* increase of the counters using $\eta^{(t)} = \max\{1, (t-1)(t-2)/(M(M-1))\}$ as weight. W.r.t. the pseudocode in Alg. 1, we replace "1" with $\eta^{(t)}$ on lines 19–22 (given change 2 above, all the calls to UPDATECOUNTERS have $\bullet = +$).

Counters. If we are interested only in estimating the global number of triangles in $G^{(t)}$, TRIÈST-IMPR needs to maintain only the counter τ and the edge sample $\mathcal{S}$ of size M, so it still requires space $O(M)$. If instead we are interested in estimating the local triangle counts, at any time t TRIÈST maintains (non-zero) local counters *only* for the nodes u such that at least one triangle with a corner u has been detected by the algorithm up until time t. The number of such nodes may be greater than $O(M)$, but this is the price to pay to obtain estimations with lower variance (Thm. 5).

Estimation. When queried for an estimation, TRIÈST-IMPR returns the value of the corresponding counter, unmodified.

Analysis.

Theorem 4. *We have $\tau^{(t)} = |\Delta^{(t)}|$ if $t \leq M$ and $\mathbb{E}\left[\tau^{(t)}\right] = |\Delta^{(t)}|$ if $t > M$.*

As in TRIÈST-BASE, the estimations by TRIÈST-IMPR are *exact* at time $t \leq M$ and unbiased for $t > M$. We now show an *upper bound* to the variance of the TRIÈST-IMPR estimations for $t > M$. The proof relies on a very careful analysis of the covariance of two triangles which depends on the order of arrival of the edges in the stream (which we assume to be adversarial).

Let $z^{(t)}$ be the number of unordered pairs (λ, γ) of distinct triangles in $G^{(t)}$ that share an edge g and are such that g is neither the last edge of λ on the stream nor the last edge of γ on the stream. For any node $u \in V^{(t)}$, let $z_u^{(t)}$ be similarly defined, but considering only the triangles incident to u.

Theorem 5. *Then, for any time $t > M$, we have*

$$\mathrm{Var}\left[\tau^{(t)}\right] \leq |\Delta^{(t)}|(\eta^{(t)} - 1) + z^{(t)} \frac{t-2-M}{M} \quad .$$

For the sake of clarity, in Thm. 5, we chose not to present a stricter but more complex bound involving triangles that do not share any edge, which, as in Thm. 2, would add a *non-positive* term to the variance.

The following result relies on Chebyshev's inequality, leveraging the above result on the variance.

Theorem 6. *Let $t \geq 0$ and assume $|\Delta^{(t)}| > 0$. For any $\varepsilon, \delta \in (0,1)$, if*

$$M > \max\left\{\sqrt{\frac{2(t-1)(t-2)}{\delta\varepsilon^2|\Delta^{(t)}| + 2} + \frac{1}{4}} + \frac{1}{2}, \frac{2z^{(t)}(t-2)}{\delta\varepsilon^2|\Delta^{(t)}|^2 + 2z^{(t)}}\right\}$$

then $|\tau^{(t)} - |\Delta^{(t)}|| < \varepsilon|\Delta^{(t)}|$ with probability $> 1 - \delta$.

Algorithm 2 TRIÈST-FD

```
Input: Fully Dynamic edge stream Σ, integer M ≥ 6
S ← ∅, d_i ← 0, d_o ← 0, t ← 0, s ← 0
for each element (•, (u, v)) from Σ do
    t ← t + 1
    s ← s • 1
    if • = + then
        if SampleEdge (u, v) then
            UpdateCounters(+, (u, v))
    else if (u, v) ∈ S then
        UpdateCounters(−, (u, v))
        S ← S \ {(u, v)}
        d_i ← d_i + 1
    else   d_o ← d_o + 1

function SampleEdge(u, v)
    if d_o + d_i = 0 then
        if |S| < M then
            S ← S ∪ {(u, v)}
            return True
        else if FlipBiasedCoin(M/t) = heads then
            Select (z, w) uniformly at random from S
            UpdateCounters(−, (z, w))
            S ← (S \ {(z, w)}) ∪ {(u, v)}
            return True
    else if FlipBiasedCoin(d_i/(d_i+d_o)) = heads then
        S ← S ∪ {(u, v)}
        d_i ← d_i − 1
        return True
    else
        d_o ← d_o − 1
        return False
```

In Thms. 5 and 6, it is possible to replace the value $z^{(t)}$ with the more interpretable $r^{(t)}$, which is agnostic to the order of the edges on the stream but gives a looser upper bound to the variance.

Results analogous to those in Thms. 4, 5, and 6 hold for the local triangle count for any $u \in V^{(t)}$, replacing the global quantities with the corresponding local ones.

4.3 Fully-dynamic algorithm - TRIÈST-FD

TRIÈST-FD computes unbiased estimates of the global and local triangle counts in a *fully-dynamic stream where edges are inserted/deleted in any arbitrary, adversarial order.* It is based on *random pairing* (RP) [13], a sampling scheme that extends reservoir sampling and can handle deletions. The idea behind the RP scheme is that edge deletions seen on the stream will be "compensated" by future edge insertions. Following RP, TRIÈST-FD keeps a counter d_i (resp. d_o) to keep track of the number of uncompensated edge deletions involving an edge e that was (resp. was *not*) in $\mathcal{S}$ at the time the deletion for e was on the stream.

When an edge deletion for an edge $e \in E^{(t-1)}$ is on the stream at the beginning of time step t, then, if $e \in \mathcal{S}$ at this time, TRIÈST-FD removes e from $\mathcal{S}$ (effectively decreasing the number of edges stored in the sample by one) and increases d_i by one. Otherwise, it just increases d_o by one. When an edge insertion for an edge $e \notin E^{(t-1)}$ is on the stream at the beginning of time step t, if $d_\mathrm{i} + d_\mathrm{o} = 0$, then TRIÈST-FD follows the standard reservoir sampling scheme. If $|\mathcal{S}| < M$, then e is deterministically inserted in $\mathcal{S}$ without removing any edge from $\mathcal{S}$ already in $\mathcal{S}$, otherwise it is inserted in $\mathcal{S}$ with probability M/t, replacing an uniformly-chosen edge already in $\mathcal{S}$. If instead $d_\mathrm{i}+d_\mathrm{o} > 0$, then e is inserted in $\mathcal{S}$ with probability $d_\mathrm{i}/(d_\mathrm{i}+d_\mathrm{o})$; since it must be $d_\mathrm{i} > 0$, then it must be $|\mathcal{S}| < M$ and no edge already in $\mathcal{S}$ needs to be removed. In any case, after having handled the eventual insertion of e into $\mathcal{S}$, the algorithm decreases d_i by 1 if e was inserted in $\mathcal{S}$, otherwise it decreases d_o by 1. TRIÈST-FD also keeps track of $s^{(t)} = |E^{(t)}|$ by appropriately incrementing or decrementing a counter by 1 depending on whether the element on the stream is an edge insertion or deletion. The pseudocode for TRIÈST-FD is presented in Alg. 2 where the UpdateCounters procedure is the same as Alg. 1.

Estimation. We denote as $M^{(t)}$ the size of $\mathcal{S}$ at the end of time t (we always have $M^{(t)} \leq M$). For any time t, let $d_{\mathrm{i}}^{(t)}$ and $d_{\mathrm{o}}^{(t)}$ be the value of the counters d_{i} and d_{o} at the end of time t respectively, and let $\omega^{(t)} = \min\{M, s^{(t)} + d_{\mathrm{i}}^{(t)} + d_{\mathrm{o}}^{(t)}\}$. Define

$$\kappa^{(t)} = 1 - \sum_{j=0}^{2} \binom{s^{(t)}}{j}\binom{d_{\mathrm{i}}^{(t)} + d_{\mathrm{o}}^{(t)}}{\omega^{(t)} - j} \Big/ \binom{s^{(t)} + d_{\mathrm{i}}^{(t)} + d_{\mathrm{o}}^{(t)}}{\omega^{(t)}} \ . \tag{2}$$

For any three positive integers a, b, c s.t. $a \leq b \leq c$, define

$$\psi_{a,b,c} = \prod_{i=0}^{a-1} \frac{c-i}{b-i} \ .$$

When queried at the end of time t, for an estimation of the global number of triangles, TRIÈST-FD returns

$$\rho^{(t)} = \begin{cases} 0 \text{ if } M^{(t)} < 3 \\ \frac{\tau^{(t)}}{\kappa^{(t)}}\psi_{3,M^{(t)},s^{(t)}} = \frac{\tau^{(t)}}{\kappa^{(t)}}\frac{s^{(t)}(s^{(t)}-1)(s^{(t)}-2)}{M^{(t)}(M^{(t)}-1)(M^{(t)}-2)} \text{ othw.} \end{cases}$$

When estimating $|\Delta_u^{(t)}|$ for $u \in V^{(t)}$, the definition for $\rho_u^{(t)}$ uses $\tau_u^{(t)}$ and has the additional condition that $\rho_u^{(t)} = 0$ if there is no counter τ_u. TRIÈST-FD can keep track of $\kappa^{(t)}$ during the execution, each update of $\kappa^{(t)}$ taking time $O(1)$. Hence the time to return the estimations is still $O(1)$.

Analysis. Let t^* be the first $t \geq M+1$ such that $|E^{(t)}| = M+1$, if such a time step exists (otherwise $t^* = +\infty$).

Theorem 7. *We have $\rho^{(t)} = |\Delta^{(t)}|$ for all $t < t^*$, and $\mathbb{E}\left[\rho^{(t)}\right] = |\Delta^{(t)}|$ for $t \geq t^*$.*

The proof relies on properties of RP and on the definitions of $\kappa^{(t)}$ and $\rho^{(t)}$.

Theorem 8. *Let $t > t^*$ s.t. $|\Delta^{(t)}| > 0$ and $s^{(t)} \geq M$. Suppose we have $d^{(t)} = d_o^{(t)} + d_i^{(t)} \leq \alpha s^{(t)}$ total unpaired deletions at time t, with $0 \leq \alpha < 1$. If $M \geq \frac{1}{2\sqrt{\alpha' - \alpha}} 7 \ln s^{(t)}$ for some $\alpha < \alpha' < 1$, we have:*

$$\begin{aligned}\mathrm{Var}\left[\rho^{(t)}\right] \leq (\kappa^{(t)})^{-2}|\Delta^{(t)}|\left(\psi_{3,M(1-\alpha'),s^{(t)}} - 1\right) + (\kappa^{(t)})^{-2}2 \\ + (\kappa^{(t)})^{-2} r^{(t)}\left(\psi^2_{3,M(1-\alpha'),s^{(t)}}\psi^{-1}_{5,M(1-\alpha'),s^{(t)}} - 1\right)\end{aligned}$$

The following result relies on Chebyshev's inequality, relying on the above result on the variance.

Theorem 9. *Let $t \geq t^*$ s.t. $|\Delta^{(t)}| > 0$ and $s^{(t)} \geq M$. Let $d^{(t)} = d_o^{(t)} + d_i^{(t)} \leq \alpha s^{(t)}$ for some $0 \leq \alpha < 1$. For any $\varepsilon, \delta \in (0,1)$, if for some $\alpha < \alpha' < 1$*

$$\begin{aligned} M > \max\Bigg\{ &\frac{1}{\sqrt{a' - \alpha}} 7 \ln s^{(t)}, \\ &(1-\alpha')^{-1}\left(\sqrt[3]{\frac{2s^{(t)}(s^{(t)}-1)(s^{(t)}-2)}{\delta\varepsilon^2|\Delta^{(t)}|(\kappa^{(t)})^2 + 2\frac{|\Delta^{(t)}|-2}{|\Delta^{(t)}|}}} + 2\right), \\ &\frac{(1-\alpha')^{-1}}{3}\left(\frac{r^{(t)}s^{(t)}}{\delta\varepsilon^2|\Delta^{(t)}|^2(\kappa^{(t)})^{-2} + 2r^{(t)}}\right)\Bigg\}\end{aligned}$$

then $|\rho^{(t)} - |\Delta^{(t)}|| < \varepsilon|\Delta^{(t)}|$ with probability $> 1 - \delta$.

Results analogous to those in Thms. 7, 8, and 9 hold for the local triangle count for any $u \in V^{(t)}$, replacing the global quantities with the corresponding local ones.

Graph	$\|V\|$	$\|E\|$	$\|E_u\|$	$\|\Delta\|$
Patent (Co-Aut.)	1,162,227	3,660,945	2,724,036	3.53×10^6
Patent (Cit.)	2,745,762	13,965,410	13,965,132	6.91×10^6
LastFm	681,387	43,518,693	30,311,117	1.13×10^9
Yahoo! Answers	2,432,573	1.21×10^9	1.08×10^9	7.86×10^{10}
Twitter	41,652,230	1.47×10^9	1.20×10^9	3.46×10^{10}

Table 1: Properties of the dynamic graph streams analyzed. $|V|$, $|E|$, $|E_u|$, $|\Delta|$ refer respectively to the number of nodes appearing in the graph, the number of edge addition events, the number of distinct edges additions, and the maximum number of triangles in the graph (for Yahoo! Answers and Twitter estimated with TRIÈST-IMPR $M = 1000000$, otherwise computed exactly with the naïve algorithm).

5 Experimental evaluation

We evaluated TRIÈST on several real-world graphs with up to a billion edges. The algorithms were implemented in C++, and ran on the Brown University CS department cluster.[5] Each run employed a single core and used at most 4 GB of RAM. We will make our code available upon publication.

Datasets We created the streams from the following publicly available graphs (properties in Table 1).

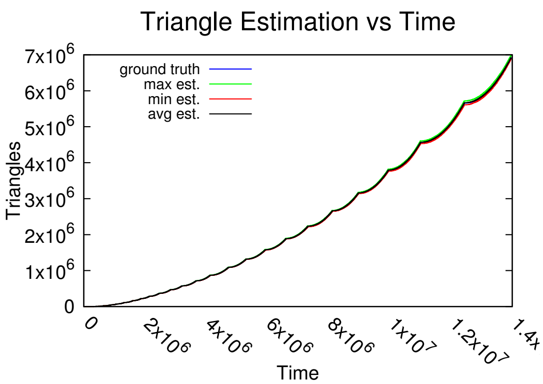
Patent (Co-Aut.) and Patent (Cit.). The *Patent (Co-Aut.)* and *Patent (Cit.)* graphs are obtained from a dataset of ≈ 2 million U.S. patents granted between '75 and '99 [15]. In *Patent (Co-Aut.)*, the nodes represent inventors and there is an edge with timestamp t between two co-inventors of a patent if the patent was granted in year t. In *Patent (Cit.)*, nodes are patents and there is an edge (a, b) with timestamp t if patent a cites b and a was granted in year t.

LastFm. The LastFm graph is based on a dataset [7, 35] of ≈ 20 million `last.fm` song listenings, ≈ 1 million songs and ≈ 1000 users. There is a node for each song and an edge between two songs if ≥ 3 users listened to both on day t.

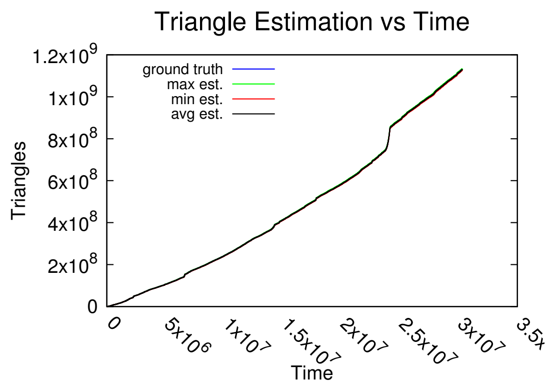
Yahoo! Answers. The Yahoo! Answers graph is obtained from a sample of ≈ 160 million answers to ≈ 25 millions questions posted on Yahoo! Answers [39]. An edge connects two users at time $max(t_1, t_2)$ if they both answered the same question at times t_1, t_2 respectively. We removed 6 outliers questions with more than 5000 answers.

Twitter. This is a snapshot [5, 23] of the Twitter followers/following network with ≈ 41 million nodes and ≈ 1.5 billions edges. We do not have time information for the edges, hence we assign a random timestamp to the edges (of which we ignore the direction).

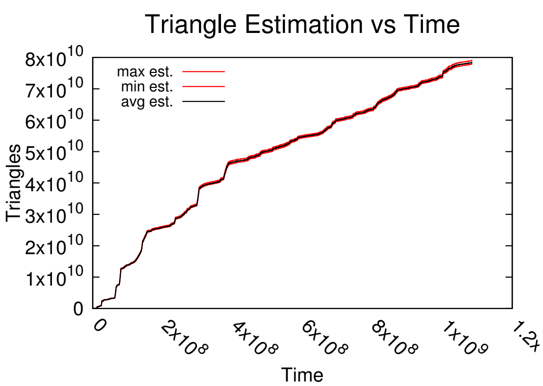
Ground truth. To evaluate the accuracy of our algorithms, we computed the *ground truth* for our smaller graphs (i.e., the exact number of global and local triangles for each time step), using an exact algorithm. The entire current graph is stored in memory and when an edge u, v is inserted (or deleted) we update the current count of local and global triangles by checking how many triangles are completed (or broken). As exact algorithms are not scalable, computing the exact triangle count is feasible only for small graphs such as Patent (Co-Aut.), Patent (Cit.) and LastFm. Table 1 reports the exact total number of triangles at the end of the stream for those graphs (and an estimate for the larger ones using TRIÈST-IMPR with $M = 1000000$).



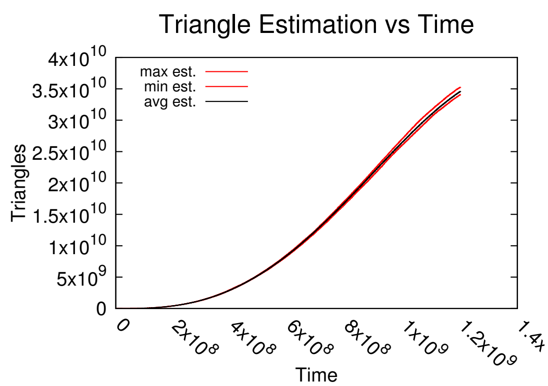
(a) Patent (Cit.)



(b) LastFm



(c) Yahoo! Answers



(d) Twitter

Figure 1: Estimation by TRIÈST-IMPR of the global number of triangles over time. Our estimations have very small error and variance: the ground truth is indistinguishable from max/min estimation.

5.1 Insertion-only case

We now evaluate TRIÈST on insertion-only streams and compare its performances with those of state-of-the-art approaches [17, 26, 32], showing that TRIÈST has an average estimation error significantly smaller than these methods both for the global and local estimation problems, while using the same amount of memory.

Estimation of the global number of triangles. Starting from an empty graph we add one edge at a time, in timestamp order. Figure 1 illustrates the evolution, over time, of the estimation computed by TRIÈST-IMPR with $M = 1{,}000{,}000$. For smaller graphs for which the ground truth can be computed exactly, the curve of the exact count is practically indistinguishable from our estimation showing the precision of the method. Our estimators have very small variance even on the very large Yahoo! Answers graph (point-wise max/min estimation over ten runs is almost coincident with the average estimation). These results show that our unbiased estimator is very accurate even when storing less than a 0.001 fraction of the total edges of the graph.

Comparison with the state of the art. We compare quantitatively with three state-of-the-art methods: MASCOT [26], JHA ET AL. [17] and PAVAN ET AL. [32]. MASCOT is a suite of local triangle counting methods (but provides also a global estimation). The other two are global triangle counting approaches. None of

[5] `https://cs.brown.edu/about/system/services/hpc/grid/`

Graph	Impr.	p	Max. MAPE Mascot	Max. MAPE trièst	Avg. MAPE Mascot	Avg. MAPE trièst	Change
Patent (Cit.)	N	0.01	0.9231	0.2583	0.6517	0.1811	-72.2%
	Y	0.01	0.1907	0.0363	0.1149	0.0213	-81.4%
	N	0.1	0.0839	0.0124	0.0605	0.0070	-88.5%
	Y	0.1	0.0317	0.0037	0.0245	0.0022	-91.1%
Patent (Co-A.)	N	0.01	2.3017	0.3029	0.8055	0.1820	-77.4%
	Y	0.01	0.1741	0.0261	0.1063	0.0177	-83.4%
	N	0.1	0.0648	0.0175	0.0390	0.0079	-79.8%
	Y	0.1	0.0225	0.0034	0.0174	0.0022	-87.2%
LastFm	N	0.01	0.1525	0.0185	0.0627	0.0118	-81.2%
	Y	0.01	0.0273	0.0046	0.0141	0.0034	-76.2%
	N	0.1	0.0075	0.0028	0.0047	0.0015	-68.1%
	Y	0.1	0.0048	0.0013	0.0031	0.0009	-72.1%

Table 2: Global triangle estimation MAPE for triѐst and Mascot. The rightmost column shows the reduction in terms of the avg. MAPE obtained by using triѐst. Rows with Y in column "Impr." refer to improved algorithms (triѐst-impr and mascot-i) while those with N to basic algorithms (triѐst-base and mascot-c).

these can handle fully-dynamic streams, in contrast with triѐst-fd. We first compare the three methods to triѐst for the global triangle counting estimation. Mascot comes in two memory efficient variants: the basic Mascot-C variant and an improved Mascot-I variant.[6] Both variants sample edges with fixed probability p, so there is no guarantee on the amount of memory used during the execution. To ensure fairness of comparison, we devised the following experiment. First, we run both Mascot-C and Mascot-I for $\ell = 10$ times with a fixed p using the same random bits for the two algorithms run-by-run (i.e. the same coin tosses used to select the edges) measuring each time the number of edges M'_i stored in the sample at the end of the stream (by construction this the is same for the two variants run-by-run). Then, we run our algorithm using $M = M'_i$ (for $i \in [\ell]$). We do the same to fix the size of the edge memory for Jha et al. [17] and Pavan et al. [32].[7] This way, all algorithms use the same amount of memory for storing edges (run-by-run).

We use the *MAPE* (Mean Average Percentage Error) to assess the accuracy of the global triangle estimators over time. The MAPE assesses the average percentage of the prediction error with respect to the ground truth, and is widely used in the prediction literature [16]. For $t = 1, \ldots, T$, let $\overline{\Delta}^{(t)}$ be the estimator of the number of triangles at time t, the MAPE is defined as $\frac{1}{T}\sum_{t=1}^{T}\left|\frac{|\Delta^{(t)}|-\overline{\Delta}^{(t)}}{|\Delta^{(t)}|}\right|$.[8]

In Fig. 2(a), we compare the average MAPE of triѐst-base and triѐst-impr as well as the two Mascot variants and the other two streaming algorithms for the Patent (Co-Aut.) graph, fixing $p = 0.01$. triѐst-impr has the smallest error of all the algorithms compared.

We now turn our attention to the efficiency of the methods. Figure 2(b) shows the average update time per operation in Patent (Co-Aut.) graph, fixing $p = 0.01$. Both Jha et al. [17] and Pavan et al. [32] are up to ≈ 3 orders of magnitude slower than the Mascot variants and triѐst. This is expected as both algorithms have an update complexity of $\Omega(M)$ (they have to go through the entire reservoir graph at each step), while both Mascot algorithms and triѐst need only to access the neighborhood of the nodes involved in the edge addition.[9] This allows both algorithms to efficiently exploit larger memory sizes. We can use efficiently M up to 1 million edges in our experiments, which only requires few megabytes of RAM.[10]

[6] In the original work [26], this variant had no suffix and was simply called Mascot. We add the -I suffix to avoid confusion. Another variant Mascot-A can be forced to store the entire graph with probability 1 so we do not consider it here.

[7] More precisely, we use $M'_i/2$ estimators in Pavan et al. as each estimator stores two edges. For Jha et al. we set the two reservoirs in the algorithm to have each size $M'_i/2$. This way, all algorithms use M'_i cells for storing (w)edges.

[8] The MAPE is not defined for t s.t. $\Delta^{(t)} = 0$ so we compute it only for t s.t. $|\Delta^{(t)}| > 0$. All algorithms we consider are guaranteed to output the correct answer for t s.t. $|\Delta^{(t)}| = 0$.

[9] We observe that Pavan et al. [32] would be more efficient with batch updates. However, we want to estimate the triangles continuously at each update. In their experiments they use batch sizes of million of updates for efficiency.

[10] The experiments in [17] use M in the order of 10^3, and in [32], large M values require large batches for efficiency.

Graph	Impr.	p	Avg. Pearson			Avg. ε Err		
			MASCOT	TRIÈST	Change	MASCOT	TRIÈST	Change
LastFm	Y	0.1	0.99	1.00	+1.18%	0.79	0.30	-62.02%
		0.05	0.97	1.00	+2.48%	0.99	0.47	-52.79%
		0.01	0.85	0.98	+14.28%	1.35	0.89	-34.24%
	N	0.1	0.97	0.99	+2.04%	1.08	0.70	-35.65%
		0.05	0.92	0.98	+6.61%	1.32	0.97	-26.53%
		0.01	0.32	0.70	+117.74%	1.48	1.34	-9.16%
Patent (Cit.)	Y	0.1	0.41	0.82	+99.09%	0.62	0.37	-39.15%
		0.05	0.24	0.61	+156.30%	0.65	0.51	-20.78%
		0.01	0.05	0.18	+233.05%	0.65	0.64	-1.68%
	N	0.1	0.16	0.48	+191.85%	0.66	0.60	-8.22%
		0.05	0.06	0.24	+300.46%	0.67	0.65	-3.21%
		0.01	0.00	0.003	+922.02%	0.86	0.68	-21.02%
Patent (Co-aut.)	Y	0.1	0.55	0.87	+58.40%	0.86	0.45	-47.91%
		0.05	0.34	0.71	+108.80%	0.91	0.63	-31.12%
		0.01	0.08	0.26	+222.84%	0.96	0.88	-8.31%
	N	0.1	0.25	0.52	+112.40%	0.92	0.83	-10.18%
		0.05	0.09	0.28	+204.98%	0.92	0.92	0.10%
		0.01	0.01	0.03	+191.46%	0.70	0.84	20.06%

Table 3: Comparison of the quality of the local triangle estimations between our algorithms and the state-of-the-art approach in [26]. Rows with Y in column "Impr." refer to improved algorithms (TRIÈST-IMPR and MASCOT-I) while those with N to basic algorithms (TRIÈST-BASE and MASCOT-C). In virtually all cases we significantly outperform MASCOT using the same amount of memory.

MASCOT is one order of magnitude faster than TRIÈST (which runs in ≈ 28 micros/op), because it does not have to handle edge removal from the sample, as it offers no guarantees on the used memory. As we will show, TRIÈST has much higher precision and scales well on billion-edges graphs.

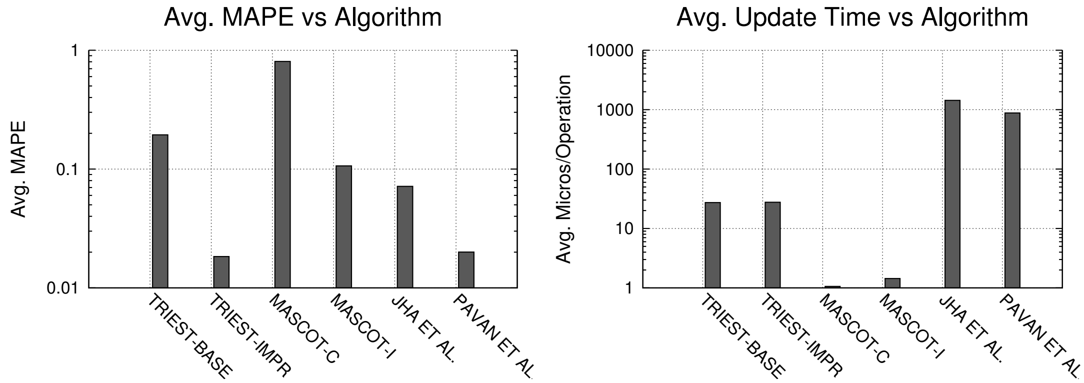
Given the slow execution of the other algorithms on the larger datasets we compare in details TRIÈST only with MASCOT.[11] Table 2 shows the average MAPE of the two approaches. The results confirm the pattern observed in Figure 2(a): TRIÈST-BASE and TRIÈST-IMPR both have an average error significantly smaller than that of the basic MASCOT-C and improved MASCOT variant respectively. We achieve up to a 91% (i.e., 9-fold) reduction in the MAPE while using the same amount of memory. This experiment confirms the theory: reservoir sampling has overall lower or equal variance in all steps for the same expected total number of sampled edges. To further validate this observation we run TRIÈST-IMPR and of the improved MASCOT variant using the same (expected memory) $M = 10000$. Figure 3 shows the max-min estimation over 10 runs. TRIÈST-IMPR shows significantly lower variance over the evolution: the maxmin estimation lines are closer to the ground truth virtually all time. This confirms our theoretical observations in the previous sections. Even with very low M (about 2/10000 of the size of the graph) TRIÈST gives a good estimation.

Local triangle counting. We compare the precision in local triangle count estimation of TRIÈST with that of MASCOT [26] using the same approach of the previous experiment. We can not compare with JHA ET AL. and PAVAN ET AL. algorithms as they provide only global estimation. As in [26], we measure the Pearson coefficient and the average ε error (see [26] for definitions). In Table 3 we report the Pearson coefficient and average ε error over all timestamps for the smaller graphs.[12] TRIÈST (significantly) improves (higher correlation and lower error) over the state-of-the-art MASCOT, using the same amount of memory.

Memory vs accuracy trade-offs. We study the tradeoff between the sample size M vs the running time and accuracy of the estimators. Figure 4(a) shows the tradeoffs between the accuracy of the estimation and the size M for the smaller graphs for which the ground truth number of triangles can be computed

[11] We attempted to run the other two algorithms but they did not complete after 12 hours for the larger datasets in Table 2 with the prescribed p parameter setting.

[12] For efficiency, in this test we evaluate the local number of triangles every 1000 edge updates.



(a) MAPE (b) Update Time

Figure 2: Average MAPE and average update time of the various methods on the Patent (Co-Aut.) graph with $p = 0.01$ – insertion only. TRIÈST-IMPR has the lowest error. Both PAVAN ET AL. and JHA ET AL. have very high update times compared to our method and the two MASCOT variants.

exactly using the naive algorithm. Even with small M TRIÈST-IMPR achieves very low MAPE value. As expected, larger M corresponds to higher accuracy and for the same M TRIÈST-IMPR outperforms TRIÈST-BASE. Figure 4(b) shows the average time per update in microseconds (μs) for TRIÈST-IMPR as function of M. Larger M requires longer update times (a larger sample implies larger graph on which to count triangles). On average a few hundreds of microseconds are sufficient for handling any update even in very large graphs with billions of edges. Our algorithms can handle hundreds of thousands of edge updates per second with very small error (Fig. 4(a)), and therefore can be used efficiently and effectively in high-velocity contexts.

Alternative edge orders. In all previous experiments the edges are added in their natural order (i.e., in order of their appearance).[13] While the natural order is the most important use case, we have assessed the impact of other ordering on the accuracy of the algorithms. We experiment with both the uniform-at-random (u.a.r.) order of the edges and the random BFS order: until all the graph is explored a BFS is started from a u.a.r. unvisited node and edges are added in order of their visit (neighbors are explored in u.a.r. order). The results for the random BFS order (Fig. 5) and for the u.a.r. (omitted for lack of space) confirm that TRIÈST has the lowest error and is very scalable in every tested ordering.

5.2 Fully-dynamic case

We evaluate TRIÈST-FD on fully-dynamic streams. We cannot compare TRIÈST-FD with the algorithms previously used [17, 26, 32] as they only handle insertion-only streams.

In the first set of experiments we model deletions using the widely used *sliding window model*, where a sliding window of the most recent edges defines the current graph. The sliding window model is of practical interest as it allow to observe recent trends in the stream. For Patent (Co-Aut.) & (Cit.) we keep in the sliding window the edges generated in the last 5 years, while for LastFm we keep the edges generated in the last 30 days. For Yahoo! Answers we keep the last 100 millions edges in the window[14].

Figure 6 shows the evolution of the global number of triangles in the sliding window model using TRIÈST-FD using $M = 200{,}000$ ($M = 1{,}000{,}000$ for Yahoo! Answers). The sliding window scenario is significantly more challenging than the addition-only case (very often the entire sample of edges is flushed away) but

[13]Excluding twitter for which we used the random order, given the lack of timestamps.

[14]The sliding window model is not interesting for the Twitter dataset, whose edges have a random timestamp, as it would create uniform samples of the graph. We do not report the results on this graph, but TRIÈST-FD is fast and has low variance also in this case.

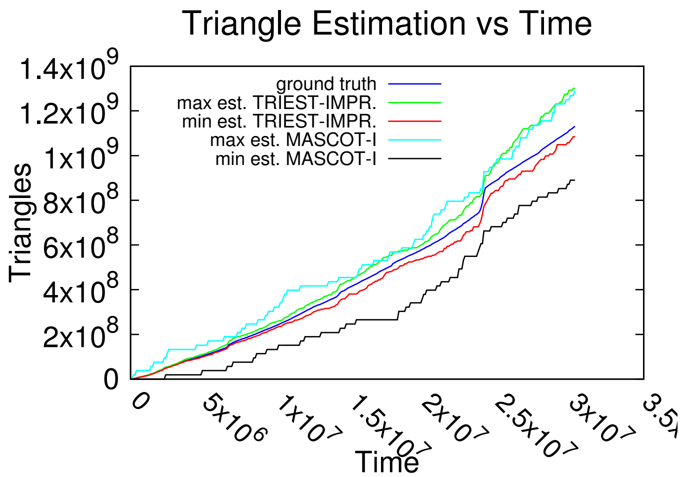


Figure 3: Variance of TRIÈST-IMPR with $M = 10000$ and of MASCOT with same expected memory, on LastFM. TRIÈST-IMPR has a smaller variance: the max/min estimation lines are closer to the ground truth. (Average estimations are qualitatively similar and not shown).

		Avg. Global	Avg. Local	
Graph	M	MAPE	Pearson	ε Err.
LastFM	200000	0.005	0.98	0.02
	1000000	0.002	0.999	0.001
Pat. (Co-Aut.)	200000	0.01	0.66	0.30
	1000000	0.001	0.99	0.006
Pat. (Cit.)	200000	0.17	0.09	0.16
	1000000	0.04	0.60	0.13

Table 4: Estimation errors for TRIÈST-FD.

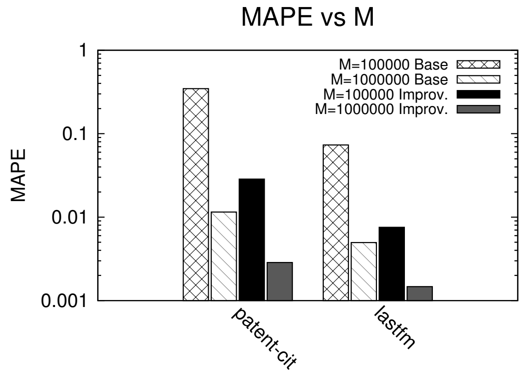
TRIÈST-FD maintains good variance and scalability even when, as for LastFm and Yahoo! Answers, the global number of triangles varies quickly.

Continuous monitoring of triangle counts with TRIÈST-FD allows to detect patterns that would otherwise be difficult to notice. For LastFm (Fig. 6(c)) we observe a sudden spike of several order of magnitudes. The dataset is anonymized so we cannot establish which songs are responsible for this spike. In Yahoo! Answers (Fig. 6(d)) a popular topic can create a sudden (and shortly lived) increase in the number of triangles, while the evolution of the Patent co-authorship and co-citation networks is slower, as the creation of an edge requires filing a patent (Fig. 6(a) and (b)). The almost constant increase over time[15] of the number of triangles in Patent graphs is consistent with previous observations of *densification* in collaboration networks as in the case of nodes' degrees [25] and the observations on the density of the densest subgraph [12].

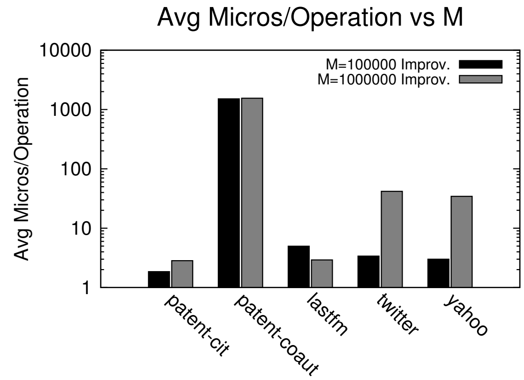
Table 4 shows the results for both the local and global triangle counting estimation provided by TRIÈST-FD. In this case we can not compare with previous works, as they only handle insertions. It is evident that precision improves with M values, and even relatively small M values result in a low MAPE (global estimation), high Pearson correlation and low ε error (local estimation). Figure 7 shows the tradeoffs between memory (i.e., accuracy) and time. In all cases our algorithm is very fast and it presents update times in the order of hundreds of microseconds for datasets with billions of updates (Yahoo).

Alternative models for deletion. We evaluate TRIÈST-FD using other models for deletions than the sliding window model. To assess the resilience of the algorithm to massive deletions we run the following experiments: we added edges in their natural order. Each edge addition is followed with probability q by a

[15]The decline at the end is due to the removal of the last edges from the sliding window after there are no more edge additions.



(a) M vs MAPE



(b) M vs Update Time

Figure 4: Trade-offs between M and MAPE or avg. update time (μs) – edge insertion only. Higher M implies lower errors but higher update times.

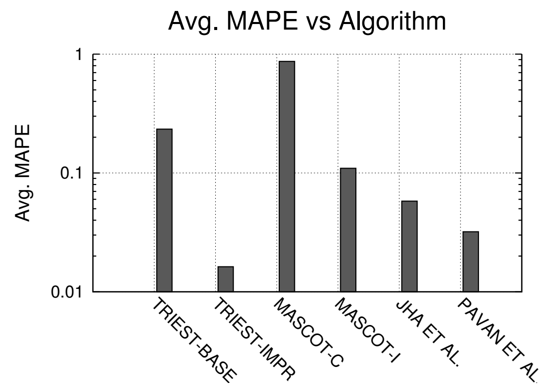
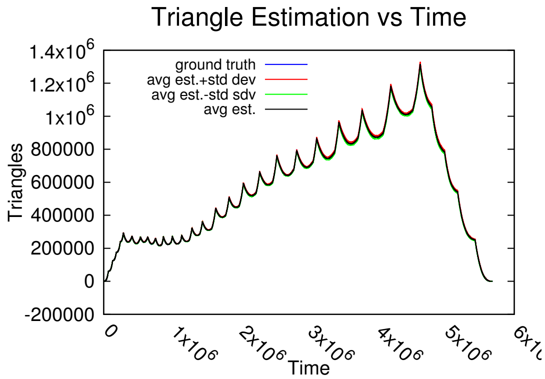


Figure 5: Average MAPE on Patent (Co-Aut.), with $p = 0.01$ – insertion only in Random BFS order. TRIÈST-IMPR has the lowest error.

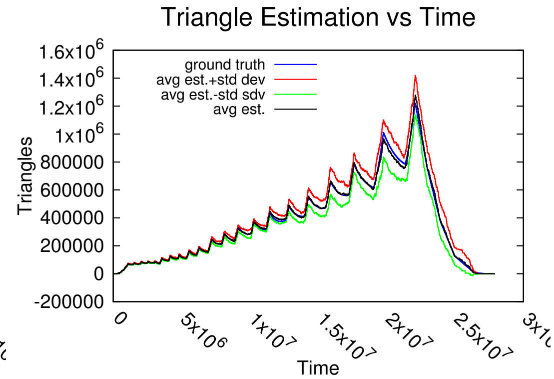
mass deletion event where each edge currently in the the graph is deleted with probability d independently. We run experiments with $q = 3{,}000{,}000^{-1}$ (i.e., a mass deletion expected every 3 millions edges) and $d = 0.80$ (in expectation 80% of edges are deleted). The results are shown in Table 5.

6 Conclusions

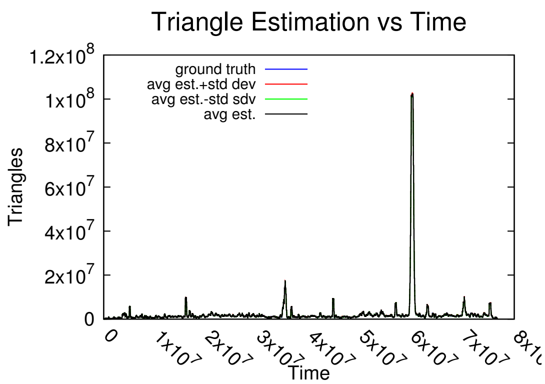
We presented TRIÈST, the first suite of algorithms that use reservoir sampling and its variants to continuously maintain unbiased, low-variance estimates of the local and global number of triangles in fully-dynamic graphs streams of arbitrary edge/vertex insertions and deletions using a fixed, user-specified amount of space. Our experimental evaluation shows that TRIÈST outperforms state-of-the-art approaches and achieves high accuracy on real-world datasets with more than one billion of edges, with update times of hundreds of microseconds. Interesting directions for future work include the use of color-coding techniques [30], and the extension to 3-profiles and complex graph motifs [11].



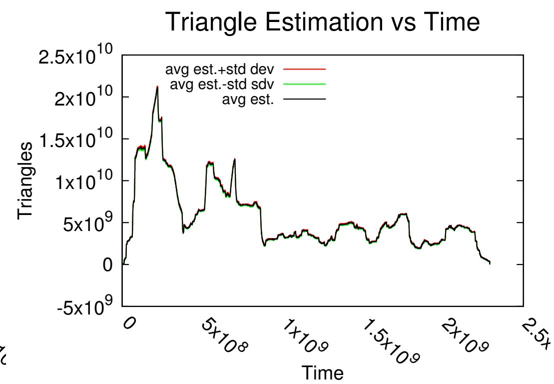
(a) Patent (Co-Aut.)



(b) Patent (Cit.)



(c) LastFm



(d) Yahoo! Answers

Figure 6: Evolution of the global number of triangles – fully dynamic case.

References

[1] N. K. Ahmed, N. Duffield, J. Neville, and R. Kompella. Graph Sample and Hold: A framework for big-graph analytics. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1446–1455. ACM, 2014.

[2] Z. Bar-Yossef, R. Kumar, and D. Sivakumar. Reductions in streaming algorithms, with an application to counting triangles in graphs. In *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '02, pages 623–632, Philadelphia, PA, USA, 2002. Society for Industrial and Applied Mathematics. ISBN 0-89871-513-X. URL `http://dl.acm.org/citation.cfm?id=545381.545464`.

[3] L. Becchetti, P. Boldi, C. Castillo, and A. Gionis. Efficient algorithms for large-scale local triangle counting. *ACM Trans. Knowl. Discov. Data*, 4(3):13:1–13:28, Oct. 2010. ISSN 1556-4681. doi: 10.1145/1839490.1839494. URL `http://doi.acm.org/10.1145/1839490.1839494`.

[4] J. W. Berry, B. Hendrickson, R. A. LaViolette, and C. A. Phillips. Tolerating the community detection resolution limit with edge weighting. *Physical Review E*, 83(5):056119, 2011.

[5] P. Boldi, M. Rosa, M. Santini, and S. Vigna. Layered label propagation: A multiresolution coordinate-free ordering for compressing social networks. In *WWW*. ACM Press, 2011.

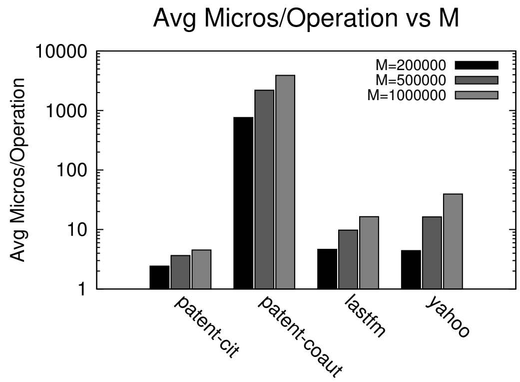


Figure 7: Trade-offs between the avg. update time (μs) and M for TRIÈST-FD.

Graph	M	Avg. Global MAPE	Avg. Local Pearson	Avg. Local ε Err.
LastFM	200000	0.04	0.62	0.53
	1000000	0.006	0.95	0.33
Pat. (Co-Aut.)	200000	0.06	0.278	0.50
	1000000	0.006	0.79	0.21
Pat. (Cit.)	200000	0.28	0.068	0.06
	1000000	0.026	0.51	0.04

Table 5: Estimation errors for TRIÈST-FD. – mass deletion experiment $q = 3{,}000{,}000^{-1}$ and $d = 0.80$.

[6] L. S. Buriol, G. Frahling, S. Leonardi, A. Marchetti-Spaccamela, and C. Sohler. Counting triangles in data streams. In *Proceedings of the Twenty-fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '06, pages 253–262, New York, NY, USA, 2006. ACM. ISBN 1-59593-318-2. doi: 10.1145/1142351.1142388. URL `http://doi.acm.org/10.1145/1142351.1142388`.

[7] Ò. Celma Herrada. Music recommendation and discovery in the long tail. Technical report, Universitat Pompeu Fabra, 2009.

[8] E. Cohen, G. Cormode, and N. Duffield. Don't let the negatives bring you down: sampling from streams of signed updates. *ACM SIGMETRICS Performance Evaluation Review*, 40(1):343–354, 2012.

[9] L. De Stefani, A. Epasto, M. Riondato, and E. Upfal. TRIÈST: Counting local and global triangles in fully-dynamic streams with fixed memory size. extended version, available from `http://www.epasto.org/papers/triangles.pdf`.

[10] J.-P. Eckmann and E. Moses. Curvature of co-links uncovers hidden thematic layers in the world wide web. *Proceedings of the National Academy of Sciences*, 99(9):5825–5829, 2002.

[11] E. R. Elenberg, K. Shanmugam, M. Borokhovich, and A. G. Dimakis. Beyond triangles: A distributed framework for estimating 3-profiles of large graphs. In *KDD'15*, 2015.

[12] A. Epasto, S. Lattanzi, and M. Sozio. Efficient densest subgraph computation in evolving graph. In *WWW*, 2015.

[13] R. Gemulla, W. Lehner, and P. J. Haas. Maintaining bounded-size sample synopses of evolving datasets. *The VLDB Journal*, 17(2):173–201, 2008.

[14] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. In *Combinatorial theory and its applications, II (Proc. Colloq., Balatonfüred, 1969)*, pages 601–623, 1970.

[15] B. H. Hall, A. B. Jaffe, and M. Trajtenberg. The NBER patent citation data file: Lessons, insights and methodological tools. Technical report, National Bureau of Economic Research, 2001.

[16] R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *Int. J. Forecasting*, 22(4):679–688, 2006. URL http://www.sciencedirect.com/science/article/pii/S0169207006000239.

[17] M. Jha, C. Seshadhri, and A. Pinar. A space-efficient streaming algorithm for estimating transitivity and triangle counts using the birthday paradox. *ACM Trans. Knowl. Discov. Data*, 9(3):15:1–15:21, Feb. 2015. ISSN 1556-4681. doi: 10.1145/2700395. URL http://doi.acm.org/10.1145/2700395.

[18] H. Jowhari and M. Ghodsi. New streaming algorithms for counting triangles in graphs. In L. Wang, editor, *Computing and Combinatorics*, volume 3595 of *Lecture Notes in Computer Science*, pages 710–716. Springer Berlin Heidelberg, 2005. ISBN 978-3-540-28061-3. doi: 10.1007/11533719_72. URL http://dx.doi.org/10.1007/11533719_72.

[19] D. M. Kane, K. Mehlhorn, T. Sauerwald, and H. Sun. Counting arbitrary subgraphs in data streams. In A. Czumaj, K. Mehlhorn, A. Pitts, and R. Wattenhofer, editors, *Automata, Languages, and Programming*, volume 7392 of *Lecture Notes in Computer Science*, pages 598–609. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-31584-8. doi: 10.1007/978-3-642-31585-5_53. URL http://dx.doi.org/10.1007/978-3-642-31585-5_53.

[20] M. N. Kolountzakis, G. L. Miller, R. Peng, and C. E. Tsourakakis. Efficient triangle counting in large graphs via degree-based vertex partitioning. *Internet Mathematics*, 8(1–2):161–185, 2012. doi: 10.1080/15427951.2012.625260. URL http://dx.doi.org/10.1080/15427951.2012.625260.

[21] K. Kutzkov and R. Pagh. On the streaming complexity of computing local clustering coefficients. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM '13, pages 677–686, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-1869-3. doi: 10.1145/2433396.2433480. URL http://doi.acm.org/10.1145/2433396.2433480.

[22] K. Kutzkov and R. Pagh. Triangle counting in dynamic graph streams. In *Algorithm Theory–SWAT 2014*, pages 306–318. Springer, 2014.

[23] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *WWW*, pages 591–600. ACM, 2010.

[24] M. Latapy. Main-memory triangle computations for very large (sparse (power-law)) graphs. *Theoretical Computer Science*, 407(1):458–473, 2008.

[25] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM Trans. KDD*, 1(1):2, 2007.

[26] Y. Lim and U. Kang. MASCOT: Memory-efficient and accurate sampling for counting local triangles in graph streams. In *KDD*, 2015.

[27] M. Manjunath, K. Mehlhorn, K. Panagiotou, and H. Sun. Approximate counting of cycles in streams. In *Algorithms–ESA 2011*, pages 677–688. Springer, 2011.

[28] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[29] M. Mitzenmacher and E. Upfal. *Probability and computing: Randomized algorithms and probabilistic analysis*. Cambridge University Press, 2005.

[30] R. Pagh and C. E. Tsourakakis. Colorful triangle counting and a MapReduce implementation. *Inf. Process. Lett.*, 112(7):277–281, Mar. 2012. ISSN 0020-0190. doi: 10.1016/j.ipl.2011.12.007. URL http://dx.doi.org/10.1016/j.ipl.2011.12.007.

[31] H.-M. Park and C.-W. Chung. An efficient MapReduce algorithm for counting triangles in a very large graph. In *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management*, CIKM '13, pages 539–548, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2263-8. doi: 10.1145/2505515.2505563. URL http://doi.acm.org/10.1145/2505515.2505563.

[32] A. Pavan, K. Tangwongsan, S. Tirthapura, and K.-L. Wu. Counting and sampling triangles from a graph stream. *Proceedings of the VLDB Endowment*, 6(14):1870–1881, 2013.

[33] M. Skala. Hypergeometric tail inequalities: ending the insanity. *arXiv preprint arXiv:1311.5939*, 2013.

[34] S. Suri and S. Vassilvitskii. Counting triangles and the curse of the last reducer. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 607–614, New York, NY, USA, 2011. ACM. ISBN 978-1-4503-0632-4. doi: 10.1145/1963405.1963491. URL http://doi.acm.org/10.1145/1963405.1963491.

[35] The Koblenz Network Collection (KONECT). Last.fm song network dataset. http://konect.uni-koblenz.de/networks/lastfm_song.

[36] C. E. Tsourakakis, U. Kang, G. L. Miller, and C. Faloutsos. Doulion: counting triangles in massive graphs with a coin. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 837–846. ACM, 2009.

[37] C. E. Tsourakakis, M. N. Kolountzakis, and G. L. Miller. Triangle sparsifiers. *J. Graph Algorithms Appl.*, 15(6):703–726, 2011.

[38] J. S. Vitter. Random sampling with a reservoir. *ACM Trans. Math. Softw.*, 11(1):37–57, Mar. 1985. ISSN 0098-3500. doi: 10.1145/3147.3165. URL http://doi.acm.org/10.1145/3147.3165.

[39] Yahoo! Research Webscope Datasets. Yahoo! Answers browsing behavior version 1.0. http://webscope.sandbox.yahoo.com.

A Additional theoretical results

In this section we present the theoretical results (statements and proofs) not included in the main text.

A.1 Theoretical results for TRIÈST-BASE

Before proving Thm. 1, we need to introduce three technical lemmas. We denote with $\Delta^{\mathcal{S}}$ the set of triangles in $G^{\mathcal{S}}$.

Lemma 2. *After each call to* UPDATECOUNTERS*, we have* $\tau = |\Delta^{\mathcal{S}}|$ *and* $\tau_v = |\Delta_v^{\mathcal{S}}|$ *for any* $v \in V_{\mathcal{S}}$ *s.t.* $|\Delta_v^{\mathcal{S}}| \geq 1$.

Proof. We only show the proof for τ, as the proof for the local counters follows the same steps.

The proof proceeds by induction. The thesis is true after the first call to UPDATECOUNTERS at time $t = 1$. Since only one edge is in $\mathcal{S}$ at this point, we have $\Delta^{\mathcal{S}} = 0$, and $\mathcal{N}_{u,v}^{\mathcal{S}} = \emptyset$, so UPDATECOUNTERS does not modify τ, which was initialized to 0. Hence $\tau = 0 = \Delta^{\mathcal{S}}$.

Assume now that the thesis is true for any subsequent call to UPDATECOUNTERS up to some point in the execution of the algorithm where an edge (u, v) is inserted or removed from $\mathcal{S}$. We now show that the thesis is still true after the call to UPDATECOUNTERS that follows this change in $\mathcal{S}$. Assume that (u, v) was *inserted* in $\mathcal{S}$, as the proof for the removal follows the same steps. Let $\mathcal{S}^{\rm b} = \mathcal{S} \setminus \{(u, v)\}$ and $\tau^{\rm b}$ be the value of τ before the call to UPDATECOUNTERS and, for any $w \in V_{\mathcal{S}^{\rm b}}$, let $\tau_w^{\rm b}$ be the value of τ_w before the call to

UpdateCounters. Let $\Delta^{\mathcal{S}}_{u,v}$ be the set of triangles in $G_{\mathcal{S}}$ that have u and v as corners. We need to show that, after the call, $\tau = |\Delta^{\mathcal{S}}|$. Clearly we have $\Delta^{\mathcal{S}} = \Delta^{\mathcal{S}^{\mathrm{b}}} \cup \Delta^{\mathcal{S}}_{u,v}$ and $\Delta^{\mathcal{S}^{\mathrm{b}}} \cap \Delta^{\mathcal{S}}_{u,v} = \emptyset$, so

$$|\Delta^{\mathcal{S}}| = |\Delta^{\mathcal{S}^{\mathrm{b}}}| + |\Delta^{\mathcal{S}}_{u,v}|$$

We have $|\Delta^{\mathcal{S}}_{u,v}| = |\mathcal{N}^{\mathcal{S}}_{u,v,}|$ and, by the inductive hypothesis, we have that $\tau^{\mathrm{b}} = |\Delta^{\mathcal{S}^{\mathrm{b}}}|$. Since UpdateCounters increments τ by $|\mathcal{N}^{\mathcal{S}}_{u,v,}|$, the value of τ after UpdateCounters has completed is exactly $|\Delta^{\mathcal{S}}|$. □

The following lemma states a well known property of the reservoir sampling scheme.

Lemma 3 (Sect. 2 [38])**.** *For any $t > M$, let A be any subset of $E^{(t)}$ of size $|A| = M$. Then, at the end of time step t,*

$$\Pr(\mathcal{S} = A) = \frac{1}{\binom{|E^{(t)}|}{M}} = \frac{1}{\binom{t}{M}},$$

i.e., the set of edges in $\mathcal{S}$ at the end of time t is a subset of $E^{(t)}$ of size M chosen uniformly *at random from all subsets of $E^{(t)}$ of the same size.*

The following lemma deals with the probability that multiple edges of $G^{(t)}$ are simultaneously in $\mathcal{S}$, i.e., with the higher-order inclusion probability of the reservoir sampling scheme.

For any pair of positive integers a and b such that $a \leq \min\{M, b\}$ let

$$\xi_{a,b} = \max\left\{1, \prod_{i=0}^{a-1} \frac{b-i}{M-i}\right\} \ .$$

Lemma 4. *For any time step t and any positive integer $k \leq t$, let B be any subset of $E^{(t)}$ of size $|B| = k \leq t$. Then, at the end of time step t,*

$$\Pr(B \subseteq \mathcal{S}) = \begin{cases} 0 & \text{if } k > M \\ \frac{1}{\xi_{k,t}} & \text{otherwise} \end{cases} \ .$$

Proof. If $k > \min\{M, t\}$, we have $\Pr(B \subseteq \mathcal{S}) = 0$ because it is impossible for B to be equal to $\mathcal{S}$ in these cases. From now on we then assume $k \leq \min\{M, t\}$.

If $t \leq M$, then $E^{(t)} \subseteq \mathcal{S}$ and $\Pr(B \subseteq \mathcal{S}) = 1 = \xi^{-1}_{k,t}$.

Assume instead that $t > M$, and let $\mathcal{B}$ be the family of subsets of $E^{(t)}$ that *1.* have size M, and *2.* contain B:

$$\mathcal{B} = \{C \subset E^{(t)} \ : \ |C| = M, B \subseteq C\} \ .$$

We have

$$|\mathcal{B}| = \binom{|E^{(t)}| - k}{M-k} = \binom{t-k}{M-k} \ . \tag{3}$$

From this, (3), and Lemma 3 we then have

$$\begin{aligned} \Pr(B \subseteq \mathcal{S}) &= \Pr(\mathcal{S} \in \mathcal{B}) = \sum_{C \in \mathcal{B}} \Pr(\mathcal{S} = C) \\ &= \frac{\binom{t-k}{M-k}}{\binom{t}{M}} = \frac{\binom{t-k}{M-k}}{\binom{t-k}{M-k} \prod_{i=0}^{k-1} \frac{t-i}{M-i}} = \prod_{i=0}^{k-1} \frac{M-i}{t-i} = \xi^{-1}_{k,t} \ . \end{aligned}$$

□

We can now prove Thm. 1 on the unbiasedness of the estimation computed by trièst-base (and on their exactness for $t \leq M$).

Proof of Thm. 1. We prove the statement for the estimation of global triangle count. The proof for the local triangle counts follows the same steps.

If $t \leq M$, we have $G_{\mathcal{S}} = G^{(t)}$ and from Lemma 2 we have $\tau^{(t)} = |\Delta^{\mathcal{S}}| = |\Delta^{(t)}|$, hence the thesis holds.

Assume now that $t > M$, and assume that $|\Delta^{(t)}| > 0$, otherwise, from Lemma 2, we have $\tau^{(t)} = |\Delta^{\mathcal{S}}| = 0$ and our estimation is deterministically correct. Let $\lambda = (a, b, c) \in \Delta^{(t)}$, (where a, b, c are edges in $E^{(t)}$) and let $\delta_\lambda^{(t)}$ be a random variable that takes value $\xi^{(t)}$ if $\lambda \in \Delta_{\mathcal{S}}$ (i.e.,$\{a, b, c\} \subseteq \mathcal{S}$) at the end of the step instant t, and 0 otherwise. From Lemma 4, we have that

$$\mathbb{E}\left[\delta_\lambda^{(t)}\right] = \xi^{(t)} \Pr(\{a, b, c\} \subseteq \mathcal{S}) = \xi^{(t)} \frac{1}{\xi_{3,t}} = \xi^{(t)} \frac{1}{\xi^{(t)}} = 1 \ . \tag{4}$$

We can write

$$\xi^{(t)} \tau^{(t)} = \sum_{\lambda \in \Delta^{(t)}} \delta_\lambda^{(t)}$$

and from this, (4), and linearity of expectation, we have

$$\mathbb{E}\left[\xi^{(t)} \tau^{(t)}\right] = \sum_{\lambda \in \Delta^{(t)}} \mathbb{E}\left[\delta_\lambda^{(t)}\right] = |\Delta^{(t)}| \ . \qquad \square$$

We now prove Thm. 2 on the variance of the estimation computed by TRIÈST-BASE for $t > M$ such that $|\Delta^{(t)}| > 0$ (we recall that the variance is 0 for $t < M$ and for all t such that $|\Delta^{(t)}| = 0$) because the estimation is deterministically exact).

Proof of Thm. 2. We prove the statement for the estimation of global triangle count. The proof for the local triangle counts follows the same steps.

Assume $|\Delta^{(t)}| > 0$, otherwise our estimation is deterministically correct and has variance 0 and the thesis holds. Let $\lambda \in \Delta^{(t)}$ and $\delta_\lambda^{(t)}$ be as in the proof of Thm. 1. We have $\mathrm{Var}[\delta_\lambda^{(t)}] = \xi^{(t)} - 1$ and from this and the definition of variance and covariance we obtain

$$\begin{aligned}
\mathrm{Var}\left[\xi^{(t)} \tau^{(t)}\right] &= \mathrm{Var}\left[\sum_{\lambda \in \Delta^{(t)}} \delta_\lambda^{(t)}\right] = \sum_{\lambda \in \Delta^{(t)}} \sum_{\gamma \in \Delta^{(t)}} \mathrm{Cov}\left[\delta_\lambda^{(t)}, \delta_\gamma^{(t)}\right] \\
&= \sum_{\lambda \in \Delta^{(t)}} \mathrm{Var}\left[\delta_\lambda^{(t)}\right] + \sum_{\substack{\lambda, \gamma \in \Delta^{(t)} \\ \lambda \neq \gamma}} \mathrm{Cov}\left[\delta_\lambda^{(t)}, \delta_\gamma^{(t)}\right] \\
&= |\Delta^{(t)}|(\xi^{(t)} - 1) + \sum_{\substack{\lambda, \gamma \in \Delta^{(t)} \\ \lambda \neq \gamma}} \mathrm{Cov}\left[\delta_\lambda^{(t)}, \delta_\gamma^{(t)}\right] \\
&= |\Delta^{(t)}|(\xi^{(t)} - 1) + \sum_{\substack{\lambda, \gamma \in \Delta^{(t)} \\ \lambda \neq \gamma}} \left(\mathbb{E}\left[\delta_\lambda^{(t)} \delta_\gamma^{(t)}\right] - \mathbb{E}\left[\delta_\lambda^{(t)}\right] \mathbb{E}\left[\delta_\gamma^{(t)}\right]\right) \\
&= |\Delta^{(t)}|(\xi^{(t)} - 1) + \sum_{\substack{\lambda, \gamma \in \Delta^{(t)} \\ \lambda \neq \gamma}} \left(\mathbb{E}\left[\delta_\lambda^{(t)} \delta_\gamma^{(t)}\right] - 1\right) \ . \qquad (5)
\end{aligned}$$

Assume now $|\Delta^{(t)}| \geq 2$, otherwise we have $r^{(t)} = w^{(t)} = 0$ and the thesis holds as the second term on the r.h.s. of (5) is 0. Let λ and γ be two distinct triangles in $\Delta^{(t)}$. If λ and γ do not share an edge, we have $\delta_\lambda^{(t)} \delta_\gamma^{(t)} = \xi^{(t)} \xi^{(t)} = \xi_{3,t}^2$ if all *six* edges composing λ and γ are in $\mathcal{S}$ at the end of time step t, and $\delta_\lambda^{(t)} \delta_\gamma^{(t)} = 0$

otherwise. From Lemma 4 we then have that

$$\mathbb{E}\left[\delta_\lambda^{(t)}\delta_\gamma^{(t)}\right] = \xi_{3,t}^2 \Pr\left(\delta_\lambda^{(t)}\delta_\gamma^{(t)} = \xi_{3,t}^2\right) = \xi_{3,t}^2 \frac{1}{\xi_{6,t}} = \xi_{3,t} \prod_{j=3}^{5} \frac{M-j}{t-j}$$
$$= \xi^{(t)} \frac{(M-3)(M-4)(M-5)}{(t-3)(t-4)(t-5)} \quad . \tag{6}$$

If instead λ and γ share exactly an edge we have $\delta_\lambda^{(t)}\delta_\gamma^{(t)} = \xi_{3,t}^2$ if all *five* edges composing λ and γ are in $\mathcal{S}$ at the end of time step t, and $\delta_\lambda^{(t)}\delta_\gamma^{(t)} = 0$ otherwise. From Lemma 4 we then have that

$$\mathbb{E}\left[\delta_\lambda^{(t)}\delta_\gamma^{(t)}\right] = \xi_{3,t}^2 \Pr\left(\delta_\lambda^{(t)}\delta_\gamma^{(t)} = \xi_{3,t}^2\right) = \xi_{3,t}^2 \frac{1}{\xi_{5,t}} = \xi_{3,t} \prod_{j=3}^{4} \frac{M-j}{t-j}$$
$$= \xi^{(t)} \frac{(M-3)(M-4)}{(t-3)(t-4)} \quad . \tag{7}$$

The thesis follows by combining (5), (6), (7), recalling the definitions of $r^{(t)}$ and $w^{(t)}$, and slightly reorganizing the terms. □

We now prove Lemma 1, about the fact that the variance of the estimations computed by TRIÈST-BASE is smaller, for most of the stream, than the variance of the estimations computed by MASCOT-C [26]. We first need the following technical facts.

Fact 1. *For any $x > 2$, we have*

$$\frac{x^2}{(x-1)(x-2)} \le 1 + \frac{4}{x-2} \quad .$$

Fact 2. *For any $x > 42$, we have*

$$\frac{x^2}{(x-3)(x-4)} \le 1 + \frac{8}{x} \quad .$$

Proof of Lemma 1. We focus on $t > M > 42$ otherwise the theorem is immediate. We show that for such conditions $f(M,t) < \bar{f}(M/T)$ and $g(M,t) < \bar{g}(M/T)$. Using the fact that $t \le \alpha T$ and Fact 1, we have

$$\begin{aligned} f(M,t) - \bar{f}(M/T) &= \frac{t(t-1)(t-2)}{M(M-1)(M-2)} - \frac{T^3}{M^3} \\ &< \frac{\alpha^3 T^3}{M^3} \frac{M^2}{(M-1)(M-2)} - \frac{T^3}{M^3} \\ &\le \frac{\alpha^3 T^3}{M^3}\left(1 + \frac{4}{M-2}\right) - \frac{T^3}{M^3} \\ &\le \frac{T^3}{M^3}\left(\alpha^3 + \frac{4\alpha^3}{M-2} - 1\right) \quad . \end{aligned} \tag{8}$$

Given that T and M are ≥ 42, the r.h.s. of (8) is non-positive iff

$$\alpha^3 + \frac{4\alpha^3}{M-2} - 1 \le 0 \quad .$$

Solving for M we have that the above is verified when $M \ge \frac{4\alpha^3}{1-\alpha^3} + 2$. This is always true given our assumption that $M > \max(\frac{8\alpha}{1-\alpha}, 42)$: for any $0 < \alpha < 0.6$, we have $\frac{4\alpha^3}{1-\alpha^3} + 2 < 42 \le M$ and for any $0.6 \le \alpha < 1$ we have $\frac{4\alpha^3}{1-\alpha^3} + 2 < \frac{8\alpha}{1-\alpha} \le M$. Hence the r.h.s. of (8) is ≤ 0 and $f(M,t) < \bar{f}(M/T)$.

We also have:

$$\begin{aligned}
g(M,t) - \bar{g}(M/T) &= \frac{t(t-1)(t-2)(M-3)(M-4)}{(t-3)(t-4)M(M-1)(M-2)} - \frac{T}{M} \\
&< \frac{t}{M}\frac{t^2}{(t-3)(t-4)} - \frac{T}{M} \\
&\le \frac{t}{M}\left(1+\frac{8}{t}\right) - \frac{T}{M}, \qquad (9)
\end{aligned}$$

where the last inequality follow from Fact 2, since $t > M > 42$. Now, from (9) since $t \le \alpha T$ and $t > M$, we can write:

$$g(M,t) - \bar{g}(M/T) < \frac{T}{M}\left(\alpha + \frac{8\alpha}{M} - 1\right) \ .$$

The r.h.s. of this equation is non-positive given the assumption $M > \frac{8\alpha}{1-\alpha}$, hence $g(M,t) < \bar{g}(M/T)$. □

The roadmap to proving Thm. 3 is the following:

1. we first define two simpler algorithms, named INDEP and MIX. The algorithms use, respectively, fixed-probability independent sampling of edges and reservoir sampling;
2. we then prove concentration results on the estimators of INDEP and MIX. Specifically the concentration result for MIX will depend on the concentration result for INDEP and on a technical result stating that probabilities of events when using reservoir sampling are close to the probabilities of those events when using fixed-probability independent sampling;
3. we then show that the estimates returned by TRIÈST-BASE are close to the estimates returned by MIX;
4. finally, we combine the above results and show that, if M is large enough, then the estimation provided by MIX is likely to be close to $|\Delta^{(t)}|$ and since the estimation computed by TRIÈST-BASE is close to that of MIX, then it must also be close to $|\Delta^{(t)}|$.

Note: for ease of presentation, in the following we use $\phi^{(t)}$ to denote the estimation returned by TRIÈST-BASE, i.e., $\phi^{(t)} = \xi^{(t)}\tau^{(t)}$.

The INDEP algorithm. The INDEP algorithm works as follows: it creates a sample $\mathcal{S}_{\text{IN}}$ by sampling edges in $E^{(t)}$ independently with a fixed probability p. It estimates the global number of triangles in $G^{(t)}$ as

$$\phi^{(t)}_{\text{IN}} = \frac{\tau^{(t)}_{\text{IN}}}{p^3},$$

where $\tau^{(t)}_{\text{IN}}$ is the number of triangles in $\mathcal{S}_{\text{IN}}$. This is for example the approach taken by MASCOT-C [26].

The MIX algorithm. The MIX algorithm works as follows: it uses reservoir sampling (like TRIÈST-BASE) to create a sample $\mathcal{S}_{\text{MIX}}$ of M edges from $E^{(t)}$. MIX uses

$$\phi^{(t)}_{\text{MIX}} = \left(\frac{t}{M}\right)^3 \tau^{(t)}$$

as an estimator for $|\Delta^{(t)}|$, where $\tau^{(t)}$ is, as in TRIÈST-BASE, the number of triangles in $G^{\mathcal{S}}$ (TRIÈST-BASE uses $\phi^{(t)} = \frac{t(t-1)(t-2)}{M(M-1)(M-2)}\tau^{(t)}$ as an estimator.)

We call this algorithm MIX because it uses reservoir sampling to create the sample, but computes the estimate as if it used fixed-probability independent sampling, hence in some sense it "mixes" the two approaches.

Concentration results for INDEP and MIX. We now show a concentration result for INDEP. Then we show a technical lemma relating the probabilities of events when using reservoir sampling to the probabilities of those events when using fixed-probability independent sampling. Finally, we use these results to show that the estimator used by MIX is also concentrated (Lemma 7).

Lemma 5. *Let $t \geq 0$ and assume $|\Delta^{(t)}| > 0$.*[16] *For any $\varepsilon, \delta \in (0,1)$, if*

$$p \geq \sqrt[3]{2\varepsilon^{-2} \ln\left(\frac{3h^{(t)}+1}{\delta}\right) \frac{3h^{(t)}+1}{|\Delta^{(t)}|}} \tag{10}$$

then

$$\Pr\left(|\phi_{\text{IN}}^{(t)} - \Delta^{(t)}|| < \varepsilon|\Delta^{(t)}|\right) > 1 - \delta \ .$$

Proof. Let H be a graph built as follows: H has one node for each triangle in $G^{(t)}$ and there is an edge between two nodes in H if the corresponding triangles in $G^{(t)}$ share an edge. By this construction, the maximum degree in H is $3h^{(t)}$. Hence by the Hajanal-Szemérédi's theorem [14] there is a proper coloring of H with at most $3h^{(t)}+1$ colors such that for each color there are at least $L = \frac{|\Delta^{(t)}|}{3h^{(t)}+1}$ nodes with that color.

Assign an arbitrary numbering to the triangles of $G^{(t)}$ (and, therefore, to the nodes of H) and let X_i be a Bernoulli random variable, indicating whether the triangle i in $G^{(t)}$ is in the sample at time t. From the properties of independent sampling of edges we have $\Pr(X_i = 1) = p^3$ for any triangle i. For any color c of the coloring of H, let $\mathcal{X}_c$ be the set of r.v.'s X_i such that the node i in H has color c. Since the coloring of H which we are considering is proper, the r.v.'s in $\mathcal{X}_c$ are independent, as they correspond to triangles which do not share any edge and edges are sampled independent of each other. Let Y_c be the sum of the r.v.'s in $\mathcal{X}_c$. The r.v. Y_c has a binomial distribution with parameters $|\mathcal{X}_c|$ and p_t^3. By the Chernoff bound for binomial r.v.'s, we have that

$$\begin{aligned}\Pr\left(|p^{-3}Y_c - |\mathcal{X}_c|| > \varepsilon|\mathcal{X}_c|\right) &< 2\exp\left(-\varepsilon^2 p^3 |\mathcal{X}_c|/2\right) \\ &< 2\exp\left(-\varepsilon^2 p^3 L/2\right) \\ &\leq \frac{\delta}{3h^{(t)}+1},\end{aligned}$$

where the last step comes from the requirement in (10).Then by applying the union bound over all the (at most) $3h^{(t)}+1$ colors we get

$$\Pr(\exists \text{ color } c \text{ s.t. } |p^{-3}Y_c - |\mathcal{X}_c|| > \varepsilon|\mathcal{X}_c|) < \delta \ .$$

Since $\phi_{\text{IN}}(t) = p^{-3} \sum_{\text{color } c} Y_c$, from the above equation we have that, with probability at least $1-\delta$,

$$\begin{aligned}|\phi_{\text{IN}}^{(t)} - |\Delta^{(t)}|| &\leq \left|\sum_{\text{color } c} p^{-3}Y_c - \sum_{\text{color } c} |\mathcal{X}_c|\right| \\ &\leq \sum_{\text{color } c} |p^{-3}Y_c - |\mathcal{X}_c|| \leq \sum_{\text{color } c} \varepsilon|\mathcal{X}_c| \leq \varepsilon|\Delta^{(t)}| \ .\end{aligned}$$

□

We remark that we can not use the same approach to show a concentration result for TRIÈST-BASE because it uses reservoir sampling, hence the event of having a triangle a in $\mathcal{S}$ and the event of having another triangle b in $\mathcal{S}$ are not independent.

We can however show the following general result, similar in spirit to the well-know Poisson approximation of balls-and-bins processes [29].

Fix the parameter M and a time $t > M$. Let $\mathcal{S}_{\text{MIX}}$ be a sample of M edges from $E^{(t)}$ obtained through reservoir sampling (as MIX would do), and let $\mathcal{S}_{\text{IN}}$ be a sample of the edges in $E^{(t)}$ obtained by sampling edges independently with probability M/t (as INDEP would do). We remark that the size of $\mathcal{S}_{\text{IN}}$ is in $[0, t]$ but not necessarily M.

[16] For $|\Delta^{(t)}| = 0$, INDEP correctly and deterministically returns 0 as the estimation.

Lemma 6. *Let $f : 2^{E^{(t)}} \to \{0,1\}$ be an arbitrary binary function from the powerset of $E^{(t)}$ to $\{0,1\}$. We have*

$$\Pr\left(f(\mathcal{S}_{\text{MIX}})=1\right) \leq e\sqrt{M}\,\Pr\left(f(\mathcal{S}_{\text{IN}})=1\right) \ .$$

Proof. Using the law of total probability, we have

$$\begin{aligned}\Pr\left(f(\mathcal{S}_{\text{IN}})=1\right) &= \sum_{k=0}^{t} \Pr\left(f(\mathcal{S}_{\text{IN}})=1 \mid |\mathcal{S}_{\text{IN}}|=k\right)\Pr(|\mathcal{S}_{\text{IN}}|=k)\\ &\geq \Pr\left(f(\mathcal{S}_{\text{IN}})=1 \mid |\mathcal{S}_{\text{IN}}|=M\right)\Pr(|\mathcal{S}_{\text{IN}}|=M)\\ &\geq \Pr\left(f(\mathcal{S}_{\text{MIX}})=1\right)\Pr\left(|\mathcal{S}_{\text{IN}}|=M\right), \qquad (11)\end{aligned}$$

where the last inequality comes from Lemma 3: the set of edges included in $\mathcal{S}_{\text{MIX}}$ is a uniformly-at-random subset of M edges from $E^{(t)}$, and the same holds for $\mathcal{S}_{\text{IN}}$ when conditioning its size being M.

Using the Stirling approximation $\sqrt{2\pi n}(\frac{n}{e})^n \leq n! \leq e\sqrt{n}(\frac{n}{e})^n$ for any positive integer n, we have

$$\begin{aligned}\Pr\left(|\mathcal{S}_{\text{IN}}|=M\right) &= \binom{t}{M}\left(\frac{M}{t}\right)^M\left(\frac{t-M}{t}\right)^{t-M}\\ &\geq \frac{t^t\sqrt{t}\sqrt{2\pi}e^{-t}}{e^2\sqrt{M}\sqrt{t-M}e^{-t}M^M(t-M)^{t-M}}\frac{M^M(t-M)^{t-M}}{t^t}\\ &\geq \frac{1}{e\sqrt{M}} \ .\end{aligned}$$

Plugging this into (11) concludes the proof. ☐

We now use the above two lemmas to show that the estimator $\phi^{(t)}_{\text{MIX}}$ computed by MIX is concentrated. We will first need the following technical fact.

Fact 3. *For any $x \geq 5$, we have*

$$\ln\left(x(1+\ln^{2/3}x)\right) \leq \ln^2 x \ .$$

Lemma 7. *Let $t \geq 0$ and assume $|\Delta^{(t)}| < 0$. For any $\varepsilon, \delta \in (0,1)$, let*

$$\Psi = 2\varepsilon^{-2}\frac{3h^{(t)}+1}{|\Delta^{(t)}|}\ln\left(e\frac{3h^{(t)}+1}{\delta}\right) \ .$$

If

$$M \geq \max\left\{t\sqrt[3]{\Psi}\left(1+\frac{1}{2}\ln^{2/3}\left(t\sqrt[3]{\Psi}\right)\right), 25\right\}$$

then

$$\Pr\left(|\phi^{(t)}_{\text{MIX}} - |\Delta^{(t)}|| < \varepsilon|\Delta^{(t)}|\right) \geq 1-\delta \ .$$

Proof. For any $S \subseteq E^{(t)}$ let $\tau(S)$ be the number of triangles in S, i.e., the number of triplets of edges in S that compose a triangle in $G^{(t)}$. Define the function $g : 2^{E^{(t)}} \to \mathbb{R}$ as

$$g(S) = \left(\frac{t}{M}\right)^3\tau(S) \ .$$

Assume that we run INDEP with $p = M/t$, and let $\mathcal{S}_{\text{IN}} \subseteq E^{(t)}$ be the sample built by INDEP (through independent sampling with fixed probability p). Assume also that we run MIX with parameter M, and let

$\mathcal{S}_{\text{MIX}}$ be the sample built by MIX (through reservoir sampling with a reservoir of size M). We have that $\phi^{(t)}_{\text{IN}} = g(\mathcal{S}_{\text{IN}})$ and $\phi^{(t)}_{\text{MIX}} = g(\mathcal{S}_{\text{MIX}})$. Define now the binary function $f : 2^{E^{(t)}} \to \{0,1\}$ as

$$f(S) = \begin{cases} 1 & \text{if } |g(S) - |\Delta^{(t)}|| > \varepsilon|\Delta^{(t)}| \\ 0 & \text{otherwise} \end{cases} \quad .$$

We now show that, for M as in the hypothesis, we have

$$p \geq \sqrt[3]{2\varepsilon^{-2}\frac{3h^{(t)}+1}{|\Delta^{(t)}|}\ln\left(e\sqrt{M}\frac{3h^{(t)}+1}{\delta}\right)} \quad . \tag{12}$$

Assume for now that the above is true. From this, using Lemma 5 and the above fact about g we get that

$$\Pr\left(|\phi^{(t)}_{\text{IN}} - |\Delta^{(t)}|| > \varepsilon|\Delta^{(t)}|\right) = \Pr\left(f(\mathcal{S}_{\text{IN}}) = 1\right) < \frac{\delta}{e\sqrt{M}} \quad .$$

From this and Lemma 6, we get that

$$\Pr\left(f(\mathcal{S}_{\text{MIX}}) = 1\right) \leq \delta$$

which, from the definition of f and the properties of g, is equivalent to

$$\Pr\left(|\phi^{(t)}_{\text{MIX}} - |\Delta^{(t)}|| > \varepsilon|\Delta^{(t)}|\right) \leq \delta$$

and the proof is complete. All that is left is to show that (12) holds for M as in the hypothesis.

Since $p = M/t$, we have that (12) holds for

$$\begin{aligned} M^3 &\geq t^3 2\varepsilon^{-2}\frac{3h^{(t)}+1}{|\Delta^{(t)}|}\ln\left(\sqrt{M}e\frac{3h^{(t)}+1}{\delta}\right) \\ &= t^3 2\varepsilon^{-2}\frac{3h^{(t)}+1}{|\Delta^{(t)}|}\left(\ln\left(e\frac{3h^{(t)}+1}{\delta}\right) + \frac{1}{2}\ln M\right) \quad . \end{aligned} \tag{13}$$

We now show that (13) holds.

Let $A = t\sqrt[3]{\Psi}$ and let $B = t\sqrt[3]{\Psi}\ln^{2/3}\left(t\sqrt[3]{\Psi}\right)$. We now show that $A^3 + B^3$ is greater or equal to the r.h.s. of (13), hence $M^3 = (A+B)^3 > A^3 + B^3$ must also be greater or equal to the r.h.s. of (13), i.e., (13) holds. This really reduces to show that

$$B^3 \geq t^3 2\varepsilon^{-2}\frac{3h^{(t)}+1}{|\Delta^{(t)}|}\frac{1}{2}\ln M \tag{14}$$

as the r.h.s.of (13) can be written as

$$A^3 + t^3 2\varepsilon^{-2}\frac{3h^{(t)}+1}{|\Delta^{(t)}|}\frac{1}{2}\ln M \quad .$$

We actually show that

$$B^3 \geq t^3\Psi\frac{1}{2}\ln M \tag{15}$$

which implies (14) which, as discussed, in turn implies (13). Consider the ratio

$$\frac{B^3}{t^3\Psi\frac{1}{2}\ln M} = \frac{\frac{1}{2}t^3\Psi\ln^2(t\sqrt[3]{\Psi})}{t^3\Psi\frac{1}{2}\ln M} = \frac{\ln^2(t\sqrt[3]{\Psi})}{\ln M} \geq \frac{\ln^2(t\sqrt[3]{\Psi})}{\ln\left(t\sqrt[3]{\Psi}\left(1+\ln^{2/3}\left(t\sqrt[3]{\Psi}\right)\right)\right)} \quad . \tag{16}$$

We now show that $t\sqrt[3]{\Psi} \geq 5$. By the assumptions $t > M \geq 25$ and by

$$t\sqrt[3]{\Psi} \geq \frac{t}{\sqrt[3]{|\Delta^{(t)}|}} \geq \sqrt{t}$$

which holds because $|\Delta^{(t)}| \leq t^{3/2}$ (in a graph with t edges there can not be more than $t^{3/2}$ triangles) we have that $t\sqrt[3]{\Psi} \geq 5$. Hence Fact 3 holds and we can write, from (16):

$$\frac{\ln^2(t\sqrt[3]{\Psi})}{\ln\left(t\sqrt[3]{\Psi}\left(1+\ln^{2/3}\left(t\sqrt[3]{\Psi}\right)\right)\right)} \geq \frac{\ln^2(t\sqrt[3]{\Psi})}{\ln^2\left(t\sqrt[3]{\Psi}\right)} \geq 1,$$

which proves (15), and in cascade (14), (13), (12), and the thesis. □

Relationship between TRIÈST-BASE and MIX. We now discuss the relationship between the $\phi^{(t)}$ and $\phi^{(t)}_{\text{MIX}}$, when both TRIÈST-BASE and MIX use a reservoir of size M.

Lemma 8. *For any $t > M$ we have*

$$\left|\phi^{(t)} - \phi^{(t)}_{\text{MIX}}\right| \leq \phi^{(t)}_{\text{MIX}} \frac{4}{M-2} \ .$$

Proof. We start by looking at the ratio between $\frac{t(t-1)(t-2)}{M(M-1)(M-2)}$ and $(t/M)^3$. We have:

$$\begin{aligned}
1 \leq \frac{t(t-1)(t-2)}{M(M-1)(M-2)}\left(\frac{M}{t}\right)^3 &= \frac{M^2}{(M-1)(M-2)}\frac{(t-1)(t-2)}{t^2}\\
&\leq \frac{M^2}{(M-1)(M-2)}\\
&\leq 1+\frac{4}{M-2}
\end{aligned}$$

where the last step follows from Fact 1. Using this, we obtain

$$\begin{aligned}
\left|\phi^{(t)} - \phi^{(t)}_{\text{MIX}}\right| &= \left|\tau^{(t)}\frac{t(t-1)(t-2)}{M(M-1)(M-2)} - \tau^{(t)}\left(\frac{t}{M}\right)^3\right|\\
&= \left|\tau^{(t)}\left(\frac{t}{M}\right)^3\left(\frac{t(t-1)(t-2)}{M(M-1)(M-2)}\left(\frac{M}{t}\right)^3 - 1\right)\right|\\
&\leq \tau^{(t)}\left(\frac{t}{M}\right)^3\frac{4}{M-2}\\
&= \phi^{(t)}_{\text{MIX}}\frac{4}{M-2} \ .
\end{aligned}$$

□

Tying everything together. Finally we can use the previous lemmas to prove our concentration result for TRIÈST-BASE.

Proof of Thm. 3. For M as in the hypothesis we have, from Lemma 7, that

$$\Pr\left(\phi^{(t)}_{\text{MIX}} \leq (1+\varepsilon/2)|\Delta^{(t)}|\right) \geq 1-\delta \ .$$

Suppose the event $\phi^{(t)}_{\text{MIX}} \leq (1+\varepsilon/2)|\Delta^{(t)}|$ (i.e., $|\phi^{(t)}_{\text{MIX}} - |\Delta^{(t)}|| \leq \varepsilon|\Delta^{(t)}|/2$) is indeed verified. From this and Lemma 8 we have

$$|\phi^{(t)} - \phi^{(t)}_{\text{MIX}}| \leq \left(1+\frac{\varepsilon}{2}\right)|\Delta^{(t)}|\frac{4}{M-2} \leq |\Delta^{(t)}|\frac{6}{M-2},$$

where the last inequality follows from the fact that $\varepsilon < 1$. Hence, given that $M \geq 12\varepsilon^{-1} + e^2 \geq 12\varepsilon^{-1} + 2$, we have

$$|\phi^{(t)} - \phi^{(t)}_{\text{MIX}}| \leq |\Delta^{(t)}|\frac{\varepsilon}{2} \ .$$

Using the above, we can then write:

$$\begin{aligned}
|\phi^{(t)} - |\Delta^{(t)}|| &= |\phi^{(t)} - \phi^{(t)}_{\text{MIX}} + \phi^{(t)}_{\text{MIX}} - |\Delta^{(t)}|| \\
&\leq |\phi^{(t)} - \phi^{(t)}_{\text{MIX}}| + |\phi^{(t)}_{\text{MIX}} - |\Delta^{(t)}|| \\
&\leq \frac{\varepsilon}{2}|\Delta^{(t)}| + \frac{\varepsilon}{2}|\Delta^{(t)}| = \varepsilon|\Delta^{(t)}|
\end{aligned}$$

which completes the proof. □

A.2 Theoretical results for TRIÈST-IMPR

Proof of Thm. 4. We prove the statement for the estimation of global triangle count. The proof for the local triangle counts follows the same steps.

If $t \leq M$ TRIÈST-IMPR behaves exactly like TRIÈST-BASE, and the statement follows from Lemma 1.

Assume now $t > M$ and assume that $|\Delta^{(t)}| > 0$, otherwise, we deterministically return 0 as an estimation and the thesis follows. Let $\lambda \in \Delta^{(t)}$ and denote with a, b, and c the edges of λ and assume, w.l.o.g., that they appear in this order (not necessarily consecutively) on the stream. Let t_λ be the time step at which c is on the stream. Let δ_λ be a random variable that takes value $\xi_{2,t_\lambda-1}$ if a and b are in $\mathcal{S}$ at the end of time step $t_\lambda - 1$, and 0 otherwise. Since it must be $t_\lambda - 1 \geq 2$, from Lemma 4 we have that

$$\Pr\left(\delta_\lambda = \xi_{2,t_\lambda-1}\right) = \frac{1}{\xi_{2,t_\lambda-1}} \ . \tag{17}$$

When $c = (u, v)$ is on the stream, i.e., at time t_λ, TRIÈST-IMPR calls UPDATECOUNTERS and increments the counter τ by $|\mathcal{N}^{\mathcal{S}}_{u,v}|\xi_{2,t_\lambda-1}$, where $|\mathcal{N}^{\mathcal{S}}_{u,v}|$ is the number of triangles with (u, v) as an edge in $\Delta^{\mathcal{S}\cup\{c\}}$. All these triangles have the corresponding random variables taking the same value $\xi_{2,t_\lambda-1}$. This means that the random variable $\tau^{(t)}$ can be expressed as

$$\tau^{(t)} = \sum_{\lambda\in\Delta^{(t)}} \delta_\lambda \ .$$

From this, linearity of expectation, and (17), we get

$$\mathbb{E}\left[\tau^{(t)}\right] = \sum_{\lambda\in\Delta^{(t)}} \mathbb{E}[\delta_\lambda] = \sum_{\lambda\in\Delta^{(t)}} \xi_{2,t_\lambda-1} \Pr\left(\delta_\lambda = \xi_{2,t_\lambda-1}\right) = \sum_{\lambda\in\Delta^{(t)}} \xi_{2,t_\lambda-1} \frac{1}{\xi_{2,t_\lambda-1}} = |\Delta^{(t)}| \ .$$ □

Before proving Thm. 5, we need the following technical lemma.

For any time step t and any edge $e \in E^{(t)}$, we denote with t_e the time step at which e is on the stream. For any $\lambda \in \Delta^{(t)}$ we denote as t_λ the time at which the last edge of λ is on the stream. W.l.o.g., let $\lambda = (\ell_1, \ell_2, \ell_3)$, where the edges are numbered in order of appearance on the stream. We define the event D_λ as the event that ℓ_1 and ℓ_2 are both in the edge sample $\mathcal{S}$ at the end of time step $t_\lambda - 1$.

Lemma 9. *Let $\lambda = (\ell_1, \ell_2, \ell_3)$ and $\gamma = (g_1, g_2, g_3)$ be two* disjoint *triangles, where the edges are numbered in order of appearance on the stream, and assume, w.l.o.g., that the last edge of λ is on the stream* before *the last edge of γ. Then*

$$\Pr(D_\gamma \mid D_\lambda) \leq \Pr(D_\gamma) \ .$$

Proof. Consider first the case where *all* edges of λ appear on the stream before *any* edge of γ, i.e.,

$$t_{\ell_1} < t_{\ell_2} < t_{\ell_3} < t_{g_1} < t_{g_2} < t_{g_3} \ .$$

The presence or absence of either or both ℓ_1 and ℓ_2 in $\mathcal{S}$ at the beginning of time step t_{ℓ_3} (i.e., whether D_λ happens or not) has no effect whatsoever on the probability that g_1 and g_2 are in the sample $\mathcal{S}$ at the beginning of time step t_{g_3}. Hence in this case,

$$\Pr(D_\gamma \mid D_\lambda) = \Pr(D_\gamma) \ .$$

Consider now the case where, for any $i \in \{1, 2\}$, the edges $g_1, \ldots, g_i$ appear on the stream before ℓ_3 does. Define now the events

A_i**:** "the edges $g_1, \ldots, g_i$ are in the sample $\mathcal{S}$ at the beginning of time step t_{ℓ_3}."

B_i**:** if $i = 1$, this is the event "the edge g_2 is inserted in the sample $\mathcal{S}$ during time step t_{g_2}." If $i = 2$, this event is the whole event space, i.e., the event that happens with probability 1.

C**:** "neither g_1 nor g_2 were among the edges removed from $\mathcal{S}$ between the beginning of time step t_{ℓ_3} and the beginning of time step t_{g_3}."

We can rewrite D_γ as

$$D_\gamma = A_i \cap B_i \cap C \ .$$

Hence

$$\begin{aligned} \Pr(D_\gamma \mid D_\lambda) &= \Pr\left(A_i \cap B_i \cap C \mid D_\lambda\right) \\ &= \Pr\left(A_i \mid D_\lambda\right) \Pr\left(B_i \cap C \mid A_i \cap D_\lambda\right) \ . \end{aligned} \tag{18}$$

We now show that

$$\Pr\left(A_i \mid D_\lambda\right) \leq \Pr\left(A_i\right) \ .$$

If we assume that $t_{\ell_3} \leq M + 1$, then all the edges that appeared on the stream up until the beginning of t_{ℓ_3} are in $\mathcal{S}$. Therefore,

$$\Pr\left(A_i \mid D_\lambda\right) = \Pr(A_i) = 1 \ .$$

Assume instead that $t_{\ell_3} > M + 1$. Among the $\binom{t_{\ell_3}-1}{M}$ subsets of $E^{(t_{\ell_3}-1)}$ of size M, there are $\binom{t_{\ell_3}-3}{M-2}$ that contain ℓ_1 and ℓ_2. From Lemma 3, we have that at the beginning of time t_{ℓ_3}, $\mathcal{S}$ is a subset of size M of $E^{(t_{\ell_3}-1)}$ chosen uniformly at random. Hence, if we condition on the fact that $\{\ell_1, \ell_2\} \subset \mathcal{S}$, we have that $\mathcal{S}$ is chosen uniformly at random from the $\binom{t_{\ell_3}-3}{M-2}$ subsets of $E^{(t_{\ell_3}-1)}$ of size M that contain ℓ_1 and ℓ_2. Among these, there are $\binom{t_{\ell_3}-3-i}{M-2-i}$ that also contain $g_1, \ldots, g_i$. Therefore,

$$\Pr(A_i \mid D_\lambda) = \frac{\binom{t_{\ell_3}-3-i}{M-2-i}}{\binom{t_{\ell_3}-3}{M-2}} = \prod_{j=0}^{i-1} \frac{M-2-j}{t_{\ell_3}-3-j} \ .$$

From Lemma 4 we have

$$\Pr(A_i) = \frac{1}{\xi_{i,t_{\ell_3}-1}} = \prod_{j=0}^{i-1} \frac{M-j}{t_{\ell_3}-1-j},$$

where the last equality comes from the assumption $t_{\ell_3} > M + 1$. From the same assumption and from the fact that for any $j \geq 0$ and any $y \geq x > j$ it holds $\frac{x-j}{y-j} \leq \frac{x}{y}$, then we have

$$\Pr(A_i \mid D_\lambda) \leq \Pr(A_i) \ .$$

This implies, from (18), that

$$\Pr(D_\gamma \mid D_\lambda) \leq \Pr(A_i) \Pr(B_i \cap C \mid A_i \cap D_\lambda) \ . \tag{19}$$

Consider now the events B_i and C. When conditioned on A_i, these event are *both independent* from D_λ: if the edges $g_1, \ldots, g_i$ are in $\mathcal{S}$ at the beginning of time t_{ℓ_3}, the fact that the edges ℓ_1 and ℓ_2 were also in

$\mathcal{S}$ at the beginning of time t_{ℓ_3} has no influence whatsoever on the actions of the algorithm (i.e., whether an edge is inserted in or removed from $\mathcal{S}$). Thus,

$$\Pr(A_i)\Pr(B_i \cap C \mid A_i \cap D_\lambda) = \Pr(A_i)\Pr(B_i \cap C \mid A_i) \ .$$

Putting this together with (19), we obtain

$$\Pr(D_\gamma \mid D_\lambda) \leq \Pr(A_i)\Pr(B_i \cap C \mid A_i) \leq \Pr(D_\gamma) \ .$$

□

We can now prove Thm. 5.

Proof of Thm. 5. We prove the statement for the estimation of global triangle count. The proof for the local triangle counts follows the same steps.

Assume $|\Delta^{(t)}| > 0$, otherwise our estimation is deterministically correct and has variance 0 and the thesis holds. Let $\lambda \in \Delta^{(t)}$ and let δ_λ be as in the proof of Thm. 4. Since

$$\mathrm{Var}\left[\delta_\lambda\right] = \xi_{2,t_\lambda-1} - 1 \leq \xi_{2,t-1},$$

we have:

$$\begin{aligned}
\mathrm{Var}\left[\tau^{(t)}\right] &= \mathrm{Var}\left[\sum_{\lambda\in\Delta^{(t)}} \delta_\lambda\right] = \sum_{\lambda\in\Delta^{(t)}}\sum_{\gamma\in\Delta^{(t)}} \mathrm{Cov}\left[\delta_\lambda, \delta_\gamma\right] \\
&= \sum_{\lambda\in\Delta^{(t)}} \mathrm{Var}\left[\delta_\lambda\right] + \sum_{\substack{\lambda,\gamma\in\Delta^{(t)}\\ \lambda\neq\gamma}} \mathrm{Cov}\left[\delta_\lambda, \delta_\gamma\right] \\
&\leq |\Delta^{(t)}|(\xi_{2,t-1} - 1) + \sum_{\substack{\lambda,\gamma\in\Delta^{(t)}\\ \lambda\neq\gamma}} \left(\mathbb{E}[\delta_\lambda\delta_\gamma] - \mathbb{E}[\delta_\lambda]\mathbb{E}[\delta_\gamma]\right) \\
&\leq |\Delta^{(t)}|(\xi_{2,t-1} - 1) + \sum_{\substack{\lambda,\gamma\in\Delta^{(t)}\\ \lambda\neq\gamma}} \left(\mathbb{E}[\delta_\lambda\delta_\gamma] - 1\right) \ . \qquad (20)
\end{aligned}$$

For any $\lambda \in \Delta^{(t)}$ define $q_\lambda = \xi_{2,t_\lambda-1}$. Assume now $|\Delta^{(t)}| \geq 2$, otherwise we have $r^{(t)} = w^{(t)} = 0$ and the thesis holds as the second term on the r.h.s. of (20) is 0. Let now λ and γ be two distinct triangles in $\Delta^{(t)}$ (hence $t \geq 5$). We have

$$\mathbb{E}\left[\delta_\lambda\delta_\gamma\right] = q_\lambda q_\gamma \Pr\left(\delta_\lambda\delta_\gamma = q_\lambda q_\gamma\right)$$

The event "$\delta_\lambda\delta_\gamma = q_\lambda q_\gamma$" is the intersection of events $D_\lambda \cap D_\gamma$, where D_λ is the event that the first two edges of λ are in $\mathcal{S}$ at the end of time step $t_\lambda - 1$, and similarly for D_γ. We now look at $\Pr(D_\lambda \cap D_\gamma)$ in the various possible cases.

Assume that λ and γ do not share any edge, and, w.l.o.g., that the third (and last) edge of λ appears on the stream before the third (and last) edge of γ, i.e., $t_\lambda < t_\gamma$. From Lemma 9 and Lemma 4 we then have

$$\Pr(D_\lambda \cap D_\gamma) = \Pr(D_\gamma|D_\lambda)\Pr(D_\lambda) \leq \Pr(D_\gamma)\Pr(D_\lambda) \leq \frac{1}{q_\lambda q_\gamma} \ .$$

Consider now the case where λ and γ share an edge g. W.l.o.g., let us assume that $t_\lambda \leq t_\gamma$ (since the shared edge may be the last on the stream both for λ and for γ, we may have $t_\lambda = t_\gamma$). There are the following possible sub-cases :

g **is the last on the stream among all the edges of λ and γ:** In this case we have $t_\lambda = t_\gamma$. The event "$D_\lambda \cap D_\gamma$" happens if and only if the *four* edges that, together with g, compose λ and γ are all in $\mathcal{S}$ at the end of time step $t_\lambda - 1$. Then, from Lemma 4 we have

$$\Pr(D_\lambda \cap D_\gamma) = \frac{1}{\xi_{4,t_\lambda-1}} \le \frac{1}{q_\lambda}\frac{(M-2)(M-3)}{(t_\lambda-3)(t_\lambda-4)} \le \frac{1}{q_\lambda}\frac{M(M-1)}{(t_\lambda-1)(t_\lambda-2)} \le \frac{1}{q_\lambda q_\gamma} \ .$$

g **is the last on the stream among all the edges of λ and the first among all the edges of γ:** In this case, we have that D_λ and D_γ are independent. Indeed the fact that the first two edges of λ (neither of which is g) are in $\mathcal{S}$ when g arrives on the stream has no influence on the probability that g and the second edge of γ are inserted in $\mathcal{S}$ and are not evicted until the third edge of γ is on the stream. Hence we have

$$\Pr(D_\lambda \cap D_\gamma) = \Pr(D_\gamma)\Pr(D_\lambda) = \frac{1}{q_\lambda q_\gamma} \ .$$

g **is the last on the stream among all the edges of λ and the second among all the edges of γ:** In this case we can follow an approach similar to the one in the proof for Lemma 9 and have that

$$\Pr(D_\lambda \cap D_\gamma) \le \Pr(D_\gamma)\Pr(D_\lambda) \le \frac{1}{q_\lambda q_\gamma} \ .$$

The intuition behind this is that if the first two edges of λ are in $\mathcal{S}$ when g is on the stream, their presence lowers the probability that the first edge of γ is in $\mathcal{S}$ at the same time, and hence that the first edge of γ and g are in $\mathcal{S}$ when the last edge of γ is on the stream.

g **is not the last on the stream among all the edges of λ:** There are two situations to consider, or better, one situation and all other possibilities. The situation we consider is that

1. g is the first edge of γ on the stream; and
2. the second edge of γ to be on the stream is on the stream at time $t_2 > t_\lambda$.

Suppose this is the case. Recall that if D_λ is verified, than we know that g is in $\mathcal{S}$ at the beginning of time step t_λ. Define the following events:

E_1**:** "the set of edges evicted from $\mathcal{S}$ between the beginning of time step t_λ and the beginning of time step t_2 does not contain g."

E_2**:** "the second edge of γ, which is on the stream at time t_2, is inserted in $\mathcal{S}$ and the edge that is evicted is not g."

E_3**:** "the set of edges evicted from $\mathcal{S}$ between the beginning of time step t_2+1 and the beginning of time step t_γ does not contain either g or the second edge of γ."

We can then write

$$\Pr(D_\gamma \mid D_\lambda) = \Pr(E_1 \mid D_\lambda)\Pr(E_2 \mid E_1 \cap D_\lambda)\Pr(E_3 \mid E_2 \cap E_1 \cap D_\lambda) \ .$$

We now compute the probabilities on the r.h.s., where we denote with $1_{t_2>M}(1)$ the function that has value 1 if $t_2 > M$, and value 0 otherwise:

$$\Pr(E_1 \mid D_\lambda) = \prod_{j=\max\{t_\lambda, M+1\}}^{t_2-1} \left(\left(1-\frac{M}{j}\right) + \frac{M}{j}\left(\frac{M-1}{M}\right)\right)$$

$$= \prod_{j=\max\{t_\lambda, M+1\}}^{t_2-1} \frac{j-1}{j} = \frac{\max\{t_\lambda - 1, M\}}{\max\{M, t_2-1\}} \ ;$$

$$\Pr(E_2 \mid E_1 \cap D_\lambda) = \frac{M}{\max\{t_2, M\}} \frac{M - 1_{t_2>M}(1)}{M} = \frac{M - 1_{t_2>M}(1)}{\max\{t_2, M\}} \ ;$$

$$\Pr(E_3 \mid E_2 \cap E_1 \cap D_\lambda) = \prod_{j=\max\{t_2+1, M+1\}}^{t_\gamma-1} \left(\left(1-\frac{M}{j}\right) + \frac{M}{j}\left(\frac{M-2}{M}\right)\right)$$

$$= \prod_{j=\max\{t_2+1, M+1\}}^{t_\gamma-1} \frac{j-2}{j} = \frac{\max\{t_2, M\}\max\{t_2-1, M-1\}}{\max\{t_\gamma-2, M-1\}\max\{t_\gamma-1, M\}} \ .$$

Hence, we have

$$\Pr(D_\gamma \mid D_\lambda) = \frac{\max\{t_\lambda-1, M\}(M - 1_{t_2>M}(1))\max\{t_2-1, M-1\}}{\max\{M, t_2-1\}\max\{(t_\gamma-2)(t_\gamma-1), M(M-1)\}} \ .$$

With a (somewhat tedious) case analysis we can verify that

$$\Pr(D_\gamma \mid D_\lambda) \le \frac{1}{q_\gamma}\frac{\max\{M, t_\lambda-1\}}{M} \ .$$

Consider now the complement of the situation we just analyzed. In this case we have that two edges of γ, that is, g and another edge h are on the stream before time t_λ, in some non-relevant order (i.e., g could be the first or the second edge of γ on the stream). Define now the following events:

E_1: "h and g are both in $\mathcal{S}$ at the beginning of time step t_λ."

E_2: "the set of edges evicted from $\mathcal{S}$ between the beginning of time step t_λ and the beginning of time step t_γ does not contain either g or h."

We can then write

$$\Pr(D_\gamma \mid D_\lambda) = \Pr(E_1 \mid D_\lambda)\Pr(E_2 \mid E_1 \cap D_\lambda) \ .$$

If $t_\lambda \le M+1$, we have that $\Pr(E_1 \mid D_\lambda) = 1$. Consider instead the case $t_\lambda > M+1$. If D_λ is verified, then both g and the other edge of λ are in $\mathcal{S}$ at the beginning of time step t_λ. At this time, all subsets of $E^{(t_\lambda-1)}$ of size M and containing both g and the other edge of λ have an equal probability of being $\mathcal{S}$, from Lemma 3. There are $\binom{t_\lambda-3}{M-2}$ such sets. Among these, there are $\binom{t_\lambda-4}{M-3}$ sets that also contain h. Therefore, if $t_\lambda > M+1$, we have

$$\Pr(E_1 \mid D_\lambda) = \frac{\binom{t_\lambda-4}{M-3}}{\binom{t_\lambda-3}{M-2}} = \frac{M-2}{t_\lambda-3} \ .$$

Considering what we said before for the case $t_\lambda \le M+1$, we then have

$$\Pr(E_1 \mid D_\lambda) = \min\left\{1, \frac{M-2}{t_\lambda-3}\right\} \ .$$

We also have

$$\Pr(E_2 \mid E_1 \cap D_\lambda) = \prod_{j=\max\{t_\lambda, M+1\}}^{t_\gamma-1} \frac{j-2}{j} = \frac{\max\{(t_\lambda-2)(t_\lambda-1), M(M-1)\}}{\max\{(t_\gamma-2)(t_\gamma-1), M(M-1)\}} \ .$$

Therefore,

$$\Pr(D_\gamma \mid D_\lambda) = \min\left\{1, \frac{M-2}{t_\lambda - 3}\right\} \frac{\max\{(t_\lambda-2)(t_\lambda-1), M(M-1)\}}{\max\{(t_\gamma-2)(t_\gamma-1), M(M-1)\}} \enspace .$$

With a case analysis, one can show that

$$\Pr(D_\gamma \mid D_\lambda) \le \frac{1}{q_\gamma}\frac{\max\{M, t_\lambda - 1\}}{M} \enspace .$$

To recap we have the following two scenarios when considering two distinct triangles γ and λ:

1. if λ and γ share an edge and, assuming w.l.o.g. that the third edge of λ is on the stream no later than the third edge of γ, and the shared edge is neither the last among all edges of λ to appear on the stream nor the last among all edges of γ to appear on the stream, then we have

$$\begin{aligned}\mathrm{Cov}[\delta_\lambda, \delta_\gamma] &\le \mathbb{E}[\delta_\lambda\delta_\gamma] - 1 = q_\lambda q_\gamma \Pr(\delta_\lambda\delta_\gamma = q_\lambda q_\gamma) - 1 \\ &\le q_\lambda q_\gamma \frac{1}{q_\lambda q_\gamma}\frac{\max\{M, t_\lambda - 1\}}{M} - 1 \le \frac{\max\{M, t_\lambda - 1\}}{M} - 1 \le \frac{t-2-M}{M};\end{aligned}$$

   where the last inequality follows from the fact that $t_\lambda \le t-1$ and $t-1 \ge M$.

2. in all other cases, including when γ and λ do not share an edge, we have $\mathbb{E}[\delta_\lambda\delta_\gamma] \le 1$, and since $\mathbb{E}[\delta_\lambda]\mathbb{E}[\delta_\gamma] = 1$, we have

$$\mathrm{Cov}[\delta_\lambda, \delta_\gamma] \le 0 \enspace .$$

Hence, we can bound

$$\sum_{\substack{\lambda,\gamma\in\Delta^{(t)}\\ \lambda\neq\gamma}} \mathrm{Cov}[\delta_\lambda, \delta_\gamma] \le z^{(t)}\frac{t-1-M}{M}$$

and the thesis follows by combining this into (20). □

We now prove Thm. 6 about TRIÈST-IMPR.

Proof of Thm. 6. By Chebyshev's inequality it is sufficient to prove that

$$\frac{\mathrm{Var}[\tau^{(t)}]}{\varepsilon^2|\Delta^{(t)}|^2} < \delta \enspace .$$

We can write

$$\frac{\mathrm{Var}[\tau^{(t)}]}{\varepsilon^2|\Delta^{(t)}|^2} \le \frac{1}{\varepsilon^2|\Delta^{(t)}|}\left((\eta(t)-1) + z^{(t)}\frac{t-2-M}{M|\Delta^{(t)}|}\right) \enspace .$$

Hence it is sufficient to impose the following two conditions:

1.

$$\begin{aligned}\frac{\delta}{2} &> \frac{\eta(t)-1}{\varepsilon^2|\Delta^{(t)}|} \\ &> \frac{1}{\varepsilon^2|\Delta^{(t)}|}\frac{(t-1)(t-2)-M(M-1)}{M(M-1)},\end{aligned} \tag{21}$$

   which is verified for:

$$M(M-1) > \frac{2(t-1)(t-2)}{\delta\varepsilon^2|\Delta^{(t)}|+2}.$$

   As $t > M$, we have $\frac{2(t-1)(t-2)}{\delta\varepsilon^2|\Delta^{(t)}|+2} > 0$. The condition in (21) is thus verified for:

$$M > \frac{1}{2}\left(\sqrt{4\frac{2(t-1)(t-2)}{\delta\varepsilon^2|\Delta^{(t)}|+2}+1}+1\right)$$

2.

$$\frac{\delta}{2} > z^{(t)}\frac{t-2-M}{M\varepsilon^2|\Delta^{(t)}|^2},$$

which is verified for:

$$M > \frac{2z^{(t)}(t-2)}{\delta\varepsilon^2|\Delta^{(t)}|^2+2z^{(t)}}.$$

The theorem follows. □

A.3 Theoretical results for TRIÈST-FD

Before proving Thm. 7 we need the following technical lemmas.

The following is a corollary of [13, Thm. 1].

Lemma 10. *For any $t > 0$, and any j, $0 \le j \le s^{(t)}$, let $\mathcal{B}^{(t)}$ be the collection of subsets of $E^{(t)}$ of size j. For any $B \in \mathcal{B}^{(t)}$ it holds*

$$\Pr\left(\mathcal{S} = B \mid M^{(t)} = j\right) = \frac{1}{\binom{|E^{(t)}|}{j}} \enspace .$$

That is, conditioned on its size at the end of time step t, $\mathcal{S}$ is equally likely to be, at the end of time step t, any of the subsets of $E^{(t)}$ of that size.

The next lemma is an immediate corollary of [13, Thm. 2]

Lemma 11. *Recall the definition of $\kappa^{(t)}$ from (2). We have*

$$\kappa^{(t)} = \Pr(M^{(t)} \ge 3) \enspace .$$

The next lemma follows from Lemma 10 in the same way as Lemma 4 follows from Lemma 3.

Lemma 12. *For any time step t and any j, $0 \le j \le s^{(t)}$, let B be any subset of $E^{(t)}$ of size $|B| = k \le s^{(t)}$. Then, at the end of time step t,*

$$\Pr\left(B \subseteq \mathcal{S} \mid M^{(t)} = j\right) = \begin{cases} 0 & \text{if } k > j \\ \frac{1}{\psi_{k,j,s^{(t)}}} & \text{otherwise} \end{cases} \enspace .$$

The next two lemmas discuss properties of TRIÈST-FD for $t < t^*$, where t^* is the first time that $|E^{(t)}|$ had size $M+1$ ($t^* \ge M+1$).

Lemma 13. *For all $t < t^*$, we have:*

1. *$d_{\mathrm{o}}^{(t)} = 0$; and*
2. *$\mathcal{S} = E^{(t)}$; and*
3. *$M^{(t)} = s^{(t)}$.*

Proof. Since the third point in the thesis follows immediately from the second, we focus on the first two points.

The proof is by induction on t. In the base base for $t = 1$: the element on the stream must be an insertion, and we deterministically insert the edge in $\mathcal{S}$. Assume now that it is true for all time steps up to (but excluding) some $t \le t^* - 1$. We now show that it is also true for t.

Assume the element on the stream at time t is a deletion. The corresponding edge must be in $\mathcal{S}$, from the inductive hypothesis. Hence TRIÈST-FD removes it from $\mathcal{S}$ and increments the counter d_{i} by 1. Thus it is still true that $\mathcal{S} = E^{(t)}$ and $d_{\mathrm{o}}^{(t)} = 0$, and the thesis holds.

Assume now that the element on the stream at time t is an insertion. From the inductive hypothesis we have that the current value of the counter d_o is 0.

If the counter d_i has currently value 0 as well, then, because of the hypothesis that $t < t^*$, it must be that $|\mathcal{S}| = M^{(t-1} = s^{(t-1)} < M$. Therefore TRIÈST-FD always inserts the edge in $\mathcal{S}$. Thus it is still true that $\mathcal{S} = E^{(t)}$ and $d_\mathrm{o}^{(t)} = 0$, and the thesis holds.

If otherwise $d_\mathrm{i} > 0$, then TRIÈST-FD flips a biased coin with probability of heads equal to

$$\frac{d_\mathrm{i}}{d_\mathrm{i} + d_\mathrm{o}} = \frac{d_\mathrm{i}}{d_\mathrm{i}} = 1,$$

therefore TRIÈST-FD always inserts the edge in $\mathcal{S}$ and decrements d_i by one. Thus it is still true that $\mathcal{S} = E^{(t)}$ and $d_\mathrm{o}^{(t)} = 0$, and the thesis holds. □

The following result is an immediate consequence of Lemma 11 and Lemma 13.

Lemma 14. *For all $t < t^*$ such that $s^{(t)} \geq 3$, we have $\kappa^{(t)} = 1$.*

We can now prove Thm. 7.

Proof of Thm. 7. We prove the statement for the estimation of global triangle count. The proof for the local triangle counts follows the same steps.

Assume for now that $t < t^*$. From Lemma 13, we have that $s^{(t)} = M^{(t)}$. If $M^{(t)} < 3$, then it must be $s^{(t)} < 3$, hence $|\Delta^{(t)}| = 0$ and indeed we return $\rho^{(t)} = 0$ in this case. If instead $M^{(t)} = s^{(t)} \geq 3$, then we have

$$\rho^{(t)} = \frac{\tau^{(t)}}{\kappa^{(t)}} \quad .$$

From Lemma 14 we have that $\kappa^{(t)} = 1$ for all $t < t^*$, hence $\rho^{(t)} = \tau^{(t)}$ in these cases. Since (an identical version of) Lemma 2 also holds for TRIÈST-FD, we have $\tau^{(t)} = |\Delta^{\mathcal{S}}| = |\Delta^{(t)}|$, where the last equality comes from the fact that $\mathcal{S} = E^{(t)}$ (Lemma 13). Hence $\rho^{(t)} = |\Delta^{(t)}|$ for any $t \leq t^*$, as in the thesis.

Assume now that $t \geq t^*$. Using the law of total expectation, we can write

$$\mathbb{E}\left[\rho^{(t)}\right] = \sum_{j=0}^{\min\{s^{(t)},M\}} \mathbb{E}\left[\rho^{(t)} \mid M^{(t)} = j\right] \Pr\left(M^{(t)} = j\right) \quad . \tag{22}$$

Assume that $|\Delta^{(t)}| > 0$, otherwise, we deterministically return 0 as an estimation and the thesis follows. Let λ be a triangle in $\Delta^{(t)}$, and let $\delta_\lambda^{(t)}$ be a random variable that takes value

$$\frac{\psi_{3,M^{(t)},s^{(t)}}}{\kappa^{(t)}} = \frac{s^{(t)}(s^{(t)}-2)(s^{(t)}-2)}{M^{(t)}(M^{(t)}-1)(M^{(t)}-2)} \frac{1}{\kappa^{(t)}}$$

if all edges of λ are in $\mathcal{S}$ at the end of the time instant t, and 0 otherwise. Since (an identical version of) Lemma 2 also holds for TRIÈST-FD, we can write

$$\rho^{(t)} = \sum_{\lambda \in \Delta^{(t)}} \delta_\lambda^{(t)} \quad .$$

Then, using Lemma 11 and Lemma 12, we have, for $3 \leq j \leq \min\{M, s^{(t)}\}$,

$$\begin{aligned}
\mathbb{E}\left[\rho^{(t)} \mid M^{(t)} = j\right] &= \sum_{\lambda \in \Delta^{(t)}} \frac{\psi_{3,j,s^{(t)}}}{\kappa^{(t)}} \Pr\left(\delta_\lambda^{(t)} = \frac{\psi_{3,j,s^{(t)}}}{\kappa^{(t)}} \mid M^{(t)} = j\right) \\
&= |\Delta^{(t)}| \frac{\psi_{3,j,s^{(t)}}}{\kappa^{(t)}} \frac{1}{\psi_{3,j,s^{(t)}}} = \frac{1}{\kappa^{(t)}} |\Delta^{(t)}|,
\end{aligned} \tag{23}$$

and

$$\mathbb{E}\left[\rho^{(t)} \mid M^{(t)} = j\right] = 0, \text{if } 0 \leq j \leq 2. \tag{24}$$

Plugging this into (22), and using Lemma 11, we have

$$\mathbb{E}\left[\rho^{(t)}\right] = |\Delta^{(t)}| \frac{1}{\kappa^{(t)}} \sum_{j=3}^{\min\{s^{(t)},M\}} \Pr(M^{(t)} = j) = |\Delta^{(t)}| \ . \qquad \square$$

We now move to prove Thm. 8 about the variance of the TRIÈST-FD estimator. We first need some technical lemmas.

Lemma 15. *For any time $t \geq t^*$, and any j, $3 \leq j \leq \min\{s^{(t)}, M\}$, we have:*

$$\mathrm{Var}\left[\rho^{(t)} | M^{(t)} = j\right] = (\kappa^{(t)})^{-2} \left(|\Delta^{(t)}| \left(\psi_{3,j,s^{(t)}} - 1\right) + r^{(t)} \left(\psi^2_{3,j,s^{(t)}} \psi^{-1}_{5,j,s^{(t)}} - 1\right) + w^{(t)} \left(\psi^2_{3,j,s^{(t)}} \psi^{-1}_{6,j,s^{(t)}} - 1\right)\right) \tag{25}$$

An analogous result holds for any $u \in V^{(t)}$, replacing the global quantities with the corresponding local ones.

Proof. The proof is analogous to that of Theorem 2, using j in place of M, $s^{(t)}$ in place of t, $\psi_{a,j,s^{(t)}}$ in place of $\xi_{a,t}$, and using Lemma 12 instead of Lemma 4. The additional $(k^{(t)})^{-2}$ multiplicative term comes from the $(k^{(t)})^{-1}$ term used in the definition of $\rho^{(t)}$. $\square$

The term $w^{(t)} \left(\psi^2_{3,j,s^{(t)}} \psi^{-1}_{6,j,s^{(t)}} - 1\right)$ is non-positive.

Lemma 16. *For any time $t \geq t^*$, and any j, $6 < j \leq \min\{s^{(t)}, M\}$, if $s^{(t)} \geq M$ we have:*

$$\mathrm{Var}\left[\rho^{(t)} | M^{(t)} = i\right] \leq (\kappa^{(t)})^{-2} \left(|\Delta^{(t)}| \left(\psi_{3,j,s^{(t)}} - 1\right) + r^{(t)} \left(\psi^2_{3,j,s^{(t)}} \psi^{-1}_{5,j,s^{(t)}} - 1\right)\right), \ \textit{for } i \geq j$$
$$\mathrm{Var}\left[\rho^{(t)} | M^{(t)} = i\right] \leq (\kappa^{(t)})^{-2} \left(|\Delta^{(t)}| \left(\psi_{3,3,s^{(t)}} - 1\right) + r^{(t)} \left(\psi^2_{3,5,s^{(t)}} \psi^{-1}_{5,5,s^{(t)}} - 1\right)\right), \ \textit{for } i < j$$

An analogous result holds for any $u \in V^{(t)}$, replacing the global quantities with the corresponding local ones.

Proof. The proof follows by observing that the term $w^{(t)} \left(\psi^2_{3,j,s^{(t)}} \psi^{-1}_{6,j,s^{(t)}} - 1\right)$ is non-positive, and that (25) is a non-increasing function of the sample size. $\square$

The following lemma deals with properties of the r.v. $M^{(t)}$.

Lemma 17. *Let $t > t^*$, with $s^{(t)} \geq M$. Let $d^{(t)} = d_o^{(t)} + d_i^{(t)}$ denote the total number of unpaired deletions at time t.*[17] *The sample size $M^{(t)}$ follows the hypergeometric distribution:*[18]

$$\Pr\left(M^{(t)} = j\right) = \begin{cases} \binom{s^{(t)}}{j}\binom{d^{(t)}}{M-j} / \binom{s^{(t)}+d^{(t)}}{M} & \textit{for } \max\{M - d^{(t)}, 0\} \leq j \leq M \\ 0 & \textit{otherwise} \end{cases} . \tag{26}$$

We have

$$\mathbb{E}\left[M^{(t)}\right] = M \frac{s^{(t)}}{s^{(t)} + d^{(t)}}, \tag{27}$$

and for any $0 < c < 1$

$$\Pr\left(M^{(t)} > \mathbb{E}\left[M^{(t)}\right] - cM\right) \geq 1 - \frac{1}{e^{2c^2 M}} \ . \tag{28}$$

[17] While both $d_o^{(t)}$ and $d_i^{(t)}$ are r.v.s, their sum is not.

[18] We use here the convention that $\binom{0}{0} = 1$, and $\binom{k}{0} = 1$.

Proof. Since $t > t*$, from the definition of t^* we have that the $M^{(t)}$ has reached size M at least once (at t^*). From this and the definition of $d^{(t)}$ (number of uncompensated deletion), we have that $M^{(t)}$ can not be less than $M - d^{(t)}$. The rest of the proof for (26) and for (27) follows from [13, Thm. 2].

The concentration bound in (28) follows from the properties of the hypergeometric distribution discussed in [33]. □

The following is an immediate corollary from the above.

Corollary 1. *Consider the execution of* TRIÈST-FD *at time $t > t^*$. Suppose we have $d^{(t)} \leq \alpha s^{(t)}$, with $0 \leq \alpha < 1$ and $s^{(t)} \geq M$. If $M \geq \frac{1}{2\sqrt{\alpha' - \alpha}} c' \ln s^{(t)}$ for $\alpha < \alpha' < 1$, we have:*

$$\Pr\left(M^{(t)} \geq M(1-\alpha')\right) > 1 - \frac{1}{\left(s^{(t)}\right)^{c'}}.$$

We can now prove Thm. 8.

Proof of Thm. 8. From the law of total variance we have:

$$\begin{aligned}
\mathrm{Var}\left[\rho^{(t)}\right] &= \sum_{j=0}^{M} \mathrm{Var}\left[\rho^{(t)} | M^{(t)} = j\right] \Pr\left(M^{(t)} = j\right) \\
&+ \sum_{j=0}^{M} \mathbb{E}\left[\rho^{(t)} | M^{(t)} = j\right]^2 (1 - \Pr\left(M^{(t)} = j\right)) \Pr\left(M^{(t)} = j\right) \\
&- 2\sum_{j=1}^{M}\sum_{i=0}^{j-1} \mathbb{E}\left[\rho^{(t)} | M^{(t)} = j\right] \Pr\left(M^{(t)} = j\right) \mathbb{E}\left[\rho^{(t)} | M^{(t)} = i\right] \Pr\left(M^{(t)} = i\right).
\end{aligned}$$

As shown in (23) and (24), for any $j = 0, 1, \ldots, M$ we have $\mathbb{E}\left[\rho^{(t)} | M^{(t)} = j\right] \geq 0$. This in turn implies:

$$\begin{aligned}
\mathrm{Var}\left[\rho^{(t)}\right] &\leq \sum_{j=0}^{M} \mathrm{Var}\left[\rho^{(t)} | M^{(t)} = j\right] \Pr\left(M^{(t)} = j\right) \\
&+ \sum_{j=0}^{M} \mathbb{E}\left[\rho^{(t)} | M^{(t)} = j\right]^2 (1 - \Pr\left(M^{(t)} = j\right)) \Pr\left(M^{(t)} = j\right). \qquad (29)
\end{aligned}$$

Let us consider separately the two main components of (29), from Lemma 16 we have:

$$\begin{aligned}
\sum_{j=0}^{M} \mathrm{Var}\left[\rho^{(t)} | M^{(t)} = j\right] \Pr\left(M^{(t)} = j\right) &= \sum_{j \geq M(1-\alpha')}^{M} \mathrm{Var}\left[\rho^{(t)} | M^{(t)} = j\right] \Pr\left(M^{(t)} = j)\right) \\
&+ \sum_{j=0}^{M(1-\alpha')} \mathrm{Var}\left[\rho^{(t)} | M^{(t)} = j\right] \Pr\left(M^{(t)} = j\right) \\
&\leq (\kappa^{(t)})^{-2} \left(|\Delta^{(t)}| \left(\psi_{3,j,s^{(t)}} - 1\right) + r^{(t)} \left(\psi^2_{3,j,s^{(t)}} \psi^{-1}_{5,j,s^{(t)}} - 1\right)\right) \Pr\left(M^{(t)} > M(1-\alpha')\right) \\
&+ \leq (\kappa^{(t)})^{-2} \left(|\Delta^{(t)}| \frac{\left(s^{(t)}\right)^3}{6} + r^{(t)} \frac{s^{(t)}}{6}\right) \Pr\left(M^{(t)} \leq M(1-\alpha')\right) \qquad (30)
\end{aligned}$$

According to our hypothesis $M \geq \frac{1}{2\sqrt{\alpha' - \alpha}} 7 \ln s^{(t)}$, from Corollary 1 we thus have:

$$\Pr\left(M^{(t)} \leq M(1-\alpha'))\right) \leq \frac{1}{(s^{(t)})^7}.$$

As $r^{(t)} < |\Delta^{(t)}|^2$ and $|\Delta^{(t)}| \leq (s^{(t)})^3$ we have:

$$(\kappa^{(t)})^{-2}\left(|\Delta^{(t)}|\frac{\left(s^{(t)}\right)^3}{6} + r^{(t)}\frac{s^{(t)}}{6}\right)\Pr\left(M^{(t)} \leq M(1-\alpha')\right) \leq (\kappa^{(t)})^{-2}$$

We can therefore rewrite (30) as:

$$\sum_{j=0}^{M}\mathrm{Var}\left[\rho^{(t)}|M^{(t)}=j\right]\Pr\left(M^{(t)}=j\right) \leq (\kappa^{(t)})^{-2}\left(|\Delta^{(t)}|\left(\psi_{3,M(1-\alpha'),s^{(t)}}-1\right)\right)$$
$$+(\kappa^{(t)})^{-2}\left(r^{(t)}\left(\psi^2_{3,M(1-\alpha'),s^{(t)}}\psi^{-1}_{5,M(1-\alpha'),s^{(t)}}-1\right)+1\right). \tag{31}$$

Let us now consider the term $\sum_{j=0}^{M}\mathbb{E}\left[\rho^{(t)}|M^{(t)}=j\right]^2\left(1-\Pr\left(M^{(t)}=j\right)\right)\Pr\left(M^{(t)}=j\right)$. Recall that, from (23) and (24), we have $\mathbb{E}\left[\rho^{(t)}|M^{(t)}=j\right]=|\Delta^{(t)}|(\kappa^{(t)})^{-1}$ for $j=3,\ldots,M$, and $\mathbb{E}\left[\rho^{(t)}|M^{(t)}=j\right]=0$ for $j=0,1,2$. From Corollary 1 we have that for $j\leq(1-\alpha')M$ and $M\geq\frac{1}{2\sqrt{\alpha'-\alpha}}7\ln s^{(t)}$

$$\Pr\left(M^{(t)}=j\right)\leq\Pr\left(M^{(t)}\leq(1-\alpha')M\right)\leq\frac{1}{\left(s^{(t)}\right)^7},$$

and thus:

$$\sum_{j=0}^{(1-\alpha')M}\mathbb{E}\left[\rho^{(t)}|M^{(t)}=j\right]^2\left(1-\Pr\left(M^{(t)}=j\right)\right)\Pr\left(M^{(t)}=j\right)\leq\frac{(1-\alpha')M|\Delta^{(t)}|^2(\kappa^{(t)})^{-2}}{\left(s^{(t)}\right)^7}$$
$$\leq(1-\alpha')(\kappa^{(t)})^{-2}, \tag{32}$$

where the last passage follows since, by hypothesis, $M\leq s^{(t)}$.

Let us now consider the values $j>(1-\alpha')M$, we have:

$$\sum_{j>(1-\alpha')M}^{M}\mathbb{E}\left[\rho^{(t)}|M^{(t)}=j\right]^2\left(1-\Pr\left(M^{(t)}=j\right)\right)\Pr\left(M^{(t)}=j\right)$$
$$\leq\alpha'M|\Delta^{(t)}|^2(\kappa^{(t)})^{-2}\left(1-\sum_{j>(1-\alpha')M}^{M}\Pr\left(M^{(t)}=j\right)\right)$$
$$\leq\alpha'M|\Delta^{(t)}|^2(\kappa^{(t)})^{-2}\left(1-\Pr\left(M^{(t)}>(1-\alpha')M\right)\right)$$
$$\leq\frac{\alpha'M|\Delta^{(t)}|^2(\kappa^{(t)})^{-2}}{\left(s^{(t)}\right)^7}$$
$$\leq\alpha'(\kappa^{(t)})^{-2}, \tag{33}$$

where the last passage follows since, by hypothesis, $M\leq s^{(t)}$.

The theorem follows from composing the upper bounds obtained in (31), (32) and (33) according to (29). □

We now prove Thm. 9 about TRIÈST-FD.

Proof of Thm. 9. By Chebyshev's inequality it is sufficient to prove that

$$\frac{\mathrm{Var}[\rho^{(t)}]}{\varepsilon^2|\Delta^{(t)}|^2}<\delta\ .$$

From Lemma 8, for $M \geq \frac{1}{\sqrt{a'-\alpha}} 7 \ln s^{(t)}$ we have:

$$\begin{aligned}\mathrm{Var}\left[\rho^{(t)}\right] \leq & (\kappa^{(t)})^{-2} |\Delta^{(t)}| \left(\psi_{3,M(1-\alpha'),s^{(t)}} - 1\right) \\ & + (\kappa^{(t)})^{-2} r^{(t)} \left(\psi^2_{3,M(1-\alpha'),s^{(t)}} \psi^{-1}_{5,M(1-\alpha'),s^{(t)}} - 1\right) \\ & + (\kappa^{(t)})^{-2} 2\end{aligned}$$

Let $M' = (1-\alpha')M$. In order to verify that the lemma holds, it is sufficient to impose the following two conditions:

1. $$\frac{\delta}{2} > \frac{(\kappa^{(t)})^{-2} \left(|\Delta^{(t)}| \left(\psi_{3,M(1-\alpha'),s^{(t)}} - 1\right) + 2\right)}{\varepsilon^2 |\Delta^{(t)}|^2}.$$

   As by hypothesis $|\Delta^{(t)}| > 0$, we can rewrite this condition as:

   $$\frac{\delta}{2} > \frac{(\kappa^{(t)})^{-2} \left(\psi_{3,M(1-\alpha'),s^{(t)}} - (\frac{|\Delta^{(t)}|-2}{|\Delta^{(t)}|}\right)}{\varepsilon^2 |\Delta^{(t)}|}$$

   which is verified for:

   $$\begin{aligned}M'(M'-1)(M'-2) &> \frac{2s^{(t)}(s^{(t)}-1)(s^{(t)}-2)}{\delta\varepsilon^2 |\Delta^{(t)}| (\kappa^{(t)})^2 + 2\frac{|\Delta^{(t)}|-2}{|\Delta^{(t)}|}}, \\ M' &> \sqrt[3]{\frac{2s^{(t)}(s^{(t)}-1)(s^{(t)}-2)}{\delta\varepsilon^2 |\Delta^{(t)}| (\kappa^{(t)})^2 + 2\frac{|\Delta^{(t)}|-2}{|\Delta^{(t)}|}}} + 2, \\ M &> (1-\alpha')^{-1} \left(\sqrt[3]{\frac{2s^{(t)}(s^{(t)}-1)(s^{(t)}-2)}{\delta\varepsilon^2 |\Delta^{(t)}| (\kappa^{(t)})^2 + 2\frac{|\Delta^{(t)}|-2}{|\Delta^{(t)}|}}} + 2\right).\end{aligned}$$

2. $$\frac{\delta}{2} > \frac{(\kappa^{(t)})^{-2}}{\varepsilon^2 |\Delta^{(t)}|^2} r^{(t)} \left(\psi^2_{3,M(1-\alpha'),s^{(t)}} \psi^{-1}_{5,M(1-\alpha'),s^{(t)}} - 1\right). \tag{34}$$

   As we have:

   $$(\kappa^{(t)})^{-2} r^{(t)} \left(\psi^2_{3,M(1-\alpha'),s^{(t)}} \psi^{-1}_{5,M(1-\alpha'),s^{(t)}} - 1\right) \leq (\kappa^{(t)})^{-2} r^{(t)} \left(\frac{s^{(t)}}{6M(1-\alpha')} - 1\right)$$

   The condition (34) is verified for:

   $$M > \frac{(1-\alpha')^{-1}}{3} \left(\frac{r^{(t)} s^{(t)}}{\delta\varepsilon^2 |\Delta^{(t)}|^2 (\kappa^{(t)})^{-2} + 2r^{(t)}}\right).$$

The theorem follows. □