

Original citation:

Chitnis, Rajesh, Cormode, Graham, Hajiaghayi, Mohammad Taghi and Monezimedeh, Morteza (2015) Parameterized streaming : maximal matching and vertex cover. In: ACM SIAM Symposium on Discrete Algorithms (SODA), San Diego, USA, 4-6 Jan 2015. Published in: Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms pp. 1234-1251.

Permanent WRAP url:

<http://wrap.warwick.ac.uk/63453>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Copyright © 2015 by the Society for Industrial and Applied Mathematics

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription. For more information, please contact the WRAP Team at: publications@warwick.ac.uk

warwick**publications**wrap

highlight your research

<http://wrap.warwick.ac.uk>

Parameterized Streaming: Maximal Matching and Vertex Cover*

Rajesh Chitnis[†]

Graham Cormode[‡]

MohammadTaghi Hajiaghayi[§]

Morteza Monemizadeh[¶]

Abstract

As graphs continue to grow in size, we seek ways to effectively process such data at scale. The model of streaming graph processing, in which a compact summary is maintained as each edge insertion/deletion is observed, is an attractive one. However, few results are known for optimization problems over such dynamic graph streams.

In this paper, we introduce a new approach to handling graph streams, by instead seeking solutions for the parameterized versions of these problems. Here, we are given a parameter k and the objective is to decide whether there is a solution bounded by k . By combining kernelization techniques with randomized sketch structures, we obtain the first streaming algorithms for the parameterized versions of Maximal Matching and Vertex Cover. We consider various models for a graph stream on n nodes: the insertion-only model where the edges can only be added, and the dynamic model where edges can be both inserted and deleted. More formally, we show the following results:

- In the insertion only model, there is a one-pass deterministic algorithm for the parameterized Vertex Cover problem which computes a sketch using

$\tilde{O}(k^2)$ space¹ such that at each timestamp in time $\tilde{O}(2^k)$ it can either extract a solution of size at most k for the current instance, or report that no such solution exists. We also show a tight lower bound of $\Omega(k^2)$ for the space complexity of any (randomized) streaming algorithms for the parameterized Vertex Cover, even in the insertion-only model.

- In the dynamic model, and under the *promise* that at each timestamp there is a maximal matching of size at most k , there is a one-pass $\tilde{O}(k^2)$ -space (sketch-based) dynamic algorithm that maintains a maximal matching with worst-case update time² $\tilde{O}(k^2)$. This algorithm partially solves Open Problem 64 from [1]. An application of this dynamic matching algorithm is a one-pass $\tilde{O}(k^2)$ -space streaming algorithm for the parameterized Vertex Cover problem that in time $\tilde{O}(2^k)$ extracts a solution for the final instance with probability $1 - \delta/n^{O(1)}$, where $\delta < 1$. To the best of our knowledge, this is the first graph streaming algorithm that combines linear sketching with sequential operations that depend on the graph at the current time.
- In the dynamic model without any promise, there is a one-pass randomized algorithm for the parameterized Vertex Cover problem which computes a sketch using $\tilde{O}(nk)$ space such that in time $\tilde{O}(nk + 2^k)$ it can either extract a solution of size at most k for the final instance, or report that no such solution exists.

*An earlier draft of this paper was made available online as <http://arxiv.org/abs/1405.0093>

[†]Department of Computer Science, University of Maryland at College Park, USA. rchitnis@cs.umd.edu. Supported in part by NSF CAREER award 1053605, NSF grant CCF-1161626, ONR YIP award N000141110662, DARPA/AFOSR grant FA9550-12-1-0423 and a Simons Award for Graduate Students in Theoretical Computer Science.

[‡]Department of Computer Science, University of Warwick, UK. g.cormode@warwick.ac.uk. Supported in part by the Yahoo Faculty Research and Engagement Program and a Royal Society Wolfson Research Merit Award.

[§]Department of Computer Science, University of Maryland, USA. hajiagha@cs.umd.edu. Supported in part by NSF CAREER award 1053605, NSF grant CCF-1161626, ONR YIP award N000141110662, and DARPA/AFOSR grant FA9550-12-1-0423.

[¶]Goethe-Universität Frankfurt, Germany and Department of Computer Science, University of Maryland at College Park, USA. monemizadeh@em.uni-frankfurt.de, Supported in part by MO 2200/1-1.

1 Introduction

Many large graphs are presented in the form of a sequence of edges. This stream of edges may be a simple stream of edge arrivals, where each edge adds to the graph seen so far, or may include a mixture of arrivals and departures of edges. In either case, we want to

¹ $\tilde{O}(f(k)) = O(f(k) \cdot \log^{O(1)} m)$, where m is the number of edges.

²The time to update the current maximal matching upon an insertion or deletion.

be able to quickly answer basic optimization questions over the current state of the graph, such as finding a (maximal) matching over the current graph edges, or finding a (minimum) vertex cover, while storing only a limited amount of information, sublinear in the size of the current graph.

The semi-streaming model introduced by Feigenbaum, Kannan, McGregor, Suri and Zhang [15] is a classical streaming model in which maximal matching and vertex cover are studied. In the semi-streaming model we are interested to solve (mostly approximately) graph problems using one pass over the graph and using $O(n \text{ polylog } n)$ space. Numerous problems have been studied in this setting, such as maintaining random walks and page rank over large graphs [32].

However, in many real world applications, we often observe instances of graph problems whose solutions are small comparing to the size of input. Consider for example the problem of finding the minimum number of fire stations to cover an entire city, or other cases where we expect a small number of facilities will serve a large number of locations. In these scenarios, assuming that the number of fire stations or facilities is a small number k is very practical. So, it is meaningful to solve instances of graph problems whose solutions are small (say, sublinear in the input size) in a streaming fashion using space which is bounded with respect to the size of their solutions, not the input size.

In order to make progress on this objective, we *parameterize* problems with a parameter k , and look for a solution whose size is bounded by k . We therefore seek *parameterized streaming algorithms* whose space and time complexities are bounded with respect to k , i.e., sublinear in the size of the input.

There are several ways to formalize this question, and we give results for the most natural formalizations. The basic case is when the input consists of a sequence of edge arrivals only, for which we seek a *parameterized streaming algorithm* (PSA). More challenging problems arise when the input stream is more dynamic, and can contain both deletions and insertions of edges. In this case we seek a *dynamic parameterized streaming algorithm* (DPSA). The challenge here is that when an edge in the matching is deleted, we sometimes need substantial work to repair the solution, and have to ensure that the algorithm has enough information to do so, while keeping only a bounded amount of working space. If we are *promised* that at every timestamp there is a solution of cost k , then we seek a *promised dynamic parameterized streaming algorithm* (PDPSA). We give examples of PSAs, DPSAs and PDPSAs for the problems of Maximal Matching and Vertex Cover.

1.1 Parameterized Complexity Most interesting optimization problems on graphs are NP-hard, implying that, unless $P=NP$, there is no polynomial time algorithm that solves all the instances of an NP-hard problem exactly. However as noted by Garey and Johnson [19], hardness results such as NP-hardness should merely constitute the beginning of research. The traditional way to combat intractability is to design approximation algorithms or randomized algorithms which run in polynomial time. These methods have their own shortcomings: we either get an approximate solution or lose the guarantee that the output is always correct.

Parameterized complexity is essentially a two-dimensional analogue of “P vs NP”. The running time is analyzed in finer detail: instead of expressing it as a function of only the input size n , one or more parameters of the input instance are defined, and we investigate the effects of these parameters on the running time. The goal is to design algorithms that work efficiently if the parameters of the input instance are small, even if the size of the input is large. We refer the reader to [12, 17] for more background.

A *parameterization* of a decision problem P is a function that assigns an integer parameter k to each instance I of P . We assume that instance I of problem P has the corresponding input $X = \{x_1, \dots, x_i, \dots, x_m\}$ consisting of elements x_i (e.g. edges defining a graph). We denote the input size of instance I by $|I| = m$. In what follows, we assume that $f(k)$ and $g(k)$ are functions of an integer parameter k .

DEFINITION 1.1. (FIXED-PARAMETER TRACTABILITY (FPT)) *A parameterized problem P is fixed-parameter tractable (FPT) if there is an algorithm that in time $f(k) \cdot m^{O(1)}$ returns a solution for each instance I whose size fulfills a given condition corresponding to k (say, at most k or at least k) or reports that such a solution does not exist.*

To illustrate this concept, we define the parameterized version of Vertex Cover as follows. A *vertex cover* of an undirected graph $G = (V, E)$ is a subset S of vertices such that for every edge $e \in E$ at least one of the endpoints (or vertices) of e is in S .

DEFINITION 1.2. (PARAMETERIZED VERTEX COVER (VC(k))) *Given an instance (I, k) where I is an undirected graph $G = (V, E)$ (with input size $|I| = |E| = m$ and $|V| = n$) and parameter $k \in \mathbb{N}$, the goal in the parameterized Vertex Cover problem (VC(k) for short) is to develop an algorithm that in time $f(k) \cdot m^{O(1)}$ either returns a vertex cover of size at most k for G , or reports that G does not have any vertex cover of size at most k .*

A simple branching method gives a $2^k \cdot m^{O(1)}$

algorithm for $VC(k)$: choose any edge and branch on choosing either end-point of the edge into the solution. The current fastest FPT algorithm for $VC(k)$ is due to Chen et al. [10] and runs in time $1.2738^k + k \cdot n$.

We also study the problem of maintaining a maximal matching, which becomes challenging in streaming models where edges are inserted and deleted.

DEFINITION 1.3. (PARAMETERIZED MAXIMAL MATCHING ($MM(k)$)) *Given an instance (I, k) where I is an undirected graph $G = (V, E)$ (with input size $|I| = |E| = m$ and $|V| = n$) and parameter $k \in \mathbb{N}$, the goal in the parameterized Maximal Matching problem ($MM(k)$ for short) is to develop an algorithm that in time $f(k) \cdot m^{O(1)}$ either returns a maximal matching of size at most k for G , or reports that G has a maximal matching of size more than k .*

One of the techniques used to obtain FPT algorithms is *kernelization*. In fact, it is known that a problem is FPT if and only if it has a kernel [17]. Kernelization has been used to design efficient algorithms by using polynomial-time preprocessing to replace the input by another equivalent input of smaller size. More formally, we have:

DEFINITION 1.4. (KERNELIZATION) *For a parameterized problem P , its kernelization is a polynomial-time transformation that maps an instance (I, k) of P to an instance (I', k') such that*

- (I, k) is a yes-instance if and only if (I', k') is a yes-instance;
- $k' \leq g(k)$ for some computable function g ;
- the size of I' is bounded by some computable function f of k , i.e., $|I'| \leq f(k)$.

The output (I', k') of a kernelization algorithm is called a kernel.

In Section 3.1 we review the kernelization algorithm of Buss and Goldsmith [7] for the parameterized Vertex Cover problem which relies on finding a maximal matching of a graph $G = (V, E)$. This kernel gives a graph with $O(k^2)$ vertices and $O(k^2)$ edges. Another kernelization algorithm given in [17] exploits the half-integrality property of LP-relaxation for vertex cover due to Nemhauser and Trotter, and produces a graph with at most $2k$ vertices.

1.2 Parameterized Streaming Algorithms: Our Results In order to state our results for *parameterized streaming* we first define the notion of a *sketch* in a very general form.

DEFINITION 1.5. (SKETCH [4, 16, 20]) *A sketch is a sublinear-space data structure that supports a fixed set of queries and updates.*

Insertion-Only Streaming. Let P be a problem parameterized by $k \in \mathbb{N}$. Let I be an instance of P that has the input $X = \{x_1, \dots, x_i, \dots, x_m\}$. Let S be a stream of $\text{INSERT}(x_i)$ (i.e., the insertion of an element x_i) operations of underlying instance (I, k) . In particular, stream S is a permutation $X' = \{x'_1, \dots, x'_i, \dots, x'_m\}$ for $x'_i \in X$ of an input X . Here we denote the time when an input x'_i is inserted by *time* i . At time i , the input which corresponds to instance I is $X'_i = \{x'_1, \dots, x'_i\}$.

DEFINITION 1.6. (PARAMETERIZED STREAMING ALGORITHM (PSA)) *Given stream S , let \mathcal{A} be an algorithm that computes a sketch for problem P using $\tilde{O}(f(k))$ -space and with one pass over stream S . Suppose at a time i , algorithm \mathcal{A} in time $\tilde{O}(g(k))$ extracts, from the sketch, a solution for input X'_i (of instance I) whose size fulfills the condition corresponding to k or reports that such a solution does not exist. Then we say \mathcal{A} is a $(f(k), g(k))$ -PSA.*

For many problems, whether or not there is a solution of size at most k is monotonic under edge additions, and so if at time i , algorithm \mathcal{A} reports that a solution for input X'_i does not exist, then there is also no solution for any input X'_t of instance I at all times $t > i$. Consequently, we can terminate the algorithm \mathcal{A} . For example, there is a trivial (k, k) -PSA for Maximal Matching in the insertion-only model: simply greedily maintain a maximal matching on the prefix of the stream so far. If the maintained matching exceeds size k , then we have evidence that there exists a matching in excess of this size. We state a simple result on the parameterized streaming algorithm for Vertex Cover and prove in Section 3.

THEOREM 1.1. *Let S be a stream of insertions of edges of an underlying graph G . Then there exists a deterministic $(k^2, 2^{O(k)})$ -PSA for $VC(k)$ problem.*

The best known kernel size for the $VC(k)$ problem is $O(k^2)$ edges [7]. In fact, Dell and van Melkebeek [11] showed that it is not possible to get a kernel for the $VC(k)$ problem with $O(k^{2-\epsilon})$ edges for any $\epsilon > 0$, under some assumptions from classical complexity. Interestingly, the space complexity of our PSA of Theorem 3.1 matches this best known kernel size. In Section 4 we show that the space complexity of above PSA is optimal even if we use randomization. More precisely, we prove the following result.

THEOREM 1.2. *Any (randomized) PSA for the $VC(k)$ problem requires $\Omega(k^2)$ space.*

Dynamic Streaming. We define *dynamic parameterized stream* as a generalization of *dynamic graph stream* introduced by Ahn, Guha and McGregor [2].

DEFINITION 1.7. (DYNAMIC PARAMETERIZED STREAM) *Let P be a problem parameterized by $k \in \mathbb{N}$. Let I be an instance of P that has an input $X = \{x_1, \dots, x_i, \dots, x_m\}$ with input size $|I| = m$. We say stream S is a dynamic parameterized stream if S is a stream of $\text{INSERT}(x_i)$ (i.e., the insertion of an element x_i) and $\text{DELETE}(x_i)$ (i.e., the deletion of an element x_i) operations applying to the underlying instance (I, k) of P .*

Now stream S is not simply a permutation $X' = \{x'_1, \dots, x'_i, \dots, x'_m\}$ for $x'_i \in X$ of an input X , but rather a sequence of transactions that collectively define a graph. We assume the size of stream S is $|S| \leq m^c$ for a constant c which means $\log |S| \leq c \log m$ or asymptotically, $O(\log |S|) = O(\log m)$. We denote the time which corresponds to the i -th update operation of S by *time i* . The i -th update operation can be $\text{INSERT}(x'_i)$ or $\text{DELETE}(x'_i)$ for $x'_i \in X$ (note that we can perform $\text{DELETE}(x'_i)$ only if x'_i is present at time $i - 1$). At time i , the input of instance I is a subset $X'_i \subseteq X$ of inputs which are, up to time i , inserted but not deleted.

We next define a *promised* streaming model as follows. Suppose we know for sure that at every time i of a dynamic parameterized stream S , the size of the vertex cover of underlying graph $G(V, E)$ (where E is the set of edges that are inserted up to time i but not deleted) is at most k . We show that within the framework of the promised streaming model we are able to develop a dynamic parameterized streaming algorithm whose space usage matches the lower bound of Theorem 4.1 up to $\tilde{O}(1)$ factor.

We formulate a dynamic parameterized streaming algorithm within the framework of the promised streaming model as follows.

DEFINITION 1.8. (PROMISED DYNAMIC PARAMETERIZED STREAMING ALGORITHM (PDPSA)) *Let S be a promised dynamic parameterized stream, i.e., we are promised that at every time i , there is a solution for input X'_i whose size fulfills the condition corresponding to k . Let \mathcal{A} be an algorithm that computes a sketch for problem P using $\tilde{O}(f(k))$ -space in one pass over stream S . Suppose at the end of stream S , i.e., time $|S|$, algorithm \mathcal{A} in time $\tilde{O}(g(k))$ extracts, from the sketch, a solution for input $X'_{|S|}$ (of instance I) whose size ful-*

fills the condition corresponding to k . We say \mathcal{A} is an $(f(k), g(k))$ -PDPSA.

In this model, maintaining a maximal matching turns out to be the more challenging problem. We summarize this main result in the following theorem and we develop it in Section 5.

THEOREM 1.3. *Suppose at every timestep the size of the vertex cover of underlying graph $G(V, E)$ is at most k . There exists a (k^2, k) -PDPSA for $MM(k)$ and $(k^2, 2^{O(k)})$ -PDPSA for $VC(k)$ with probability $\geq 1 - \delta/n^c$, where $\delta < 1$ and c is a constant.*

Our algorithm takes the novel approach of combining linear sketching with sequential operations that depend on the current state of the graph. Prior work in sketching has instead only performed updates of sketches for each stream update, and postponed inspecting them until the end of the stream. As an example, Ahn, Guha, and McGregor [2] proposed a multi-pass streaming algorithm for $MM(k)$. Their algorithm repeatedly samples an edge set of size $\tilde{O}(n^{1+1/p})$ in each pass, and finds the maximal matching for the sampled edges, for p rounds, and remove the vertices in the matching. Indeed, our algorithm partially solves Open Problem 64 from [1] as posed by McGregor at ‘‘Bertinoro Workshop on Sublinear Algorithms 2014’’. The problem as stated is ‘‘Consider an unweighted graph on n nodes defined by a stream of edge insertions and deletions. Is it possible to approximate the size of the maximum cardinality matching up to constant factor given a single pass and $o(n^2)$ space?’’ Even stronger for $k = o(n)$, our algorithm maintains a maximal matching of size $o(n)$ using $o(n^2)$ space. As an example for $k = \tilde{O}(\sqrt{n})$, this gives a dynamic algorithm for maximal matching whose space, worst-case update and query times are $\tilde{O}(n)$, $\tilde{O}(\sqrt{n})$ and $\tilde{O}(\sqrt{n})$, respectively.

Finally, we formulate a dynamic parameterized streaming algorithm without any promise as follows.

DEFINITION 1.9. (DYNAMIC PARAMETERIZED STREAMING ALGORITHM (DPSA)) *Let S be a dynamic parameterized stream S . Let \mathcal{A} be an algorithm that computes a sketch for problem P using $\tilde{o}(m) \cdot f(k)$ -space and with one pass over stream S . Suppose at the end of stream S , i.e., time $|S|$, algorithm \mathcal{A} in time $\tilde{o}(m) \cdot g(k)$ extracts, from the sketch, a solution for input $X'_{|S|}$ whose size fulfills the condition corresponding to k or reports that such a solution does not exist. We say \mathcal{A} is an $(\tilde{o}(m) \cdot f(k), \tilde{o}(m) \cdot g(k))$ -DPSA.*

We state our result on the DPSA (without any promise) for Maximal Matching and Vertex Cover and prove it in Section 6.

THEOREM 1.4. *Let S be a dynamic parameterized stream of insertions and deletions of edges of an underlying graph G . There exists a randomized $(\min(m, nk), \min(m, nk) + 2^{O(k)})$ -DPSA for $VC(k)$ problem and a $(\min(m, nk), \min(m, nk))$ -DPSA for $MM(k)$.*

For graphs which are not sparse (i.e., $m > O(nk)$) the algorithm of Theorem 1.4 gives $(\tilde{o}(m) \cdot f(k), \tilde{o}(m) \cdot g(k))$ -DPSA for $VC(k)$. The space usage of PDPSA of Theorem 1.3 matches the lower bound of Theorem 4.1. On the other hand, there is a gap between space bound $\tilde{O}(nk)$ of DPSA of Theorem 1.4 and lower bound $\Omega(k^2)$ of Theorem 4.1.

1.3 Related Work The question of finding maximal and maximum cardinality matchings has been heavily studied in the model of (insert-only) graph streams. The greedy algorithm to find a maximal matching (simply store every edge that links two currently unmatched nodes) can also be shown to be a 0.5-approximation to the maximum cardinality matching [15]. By taking multiple passes over the input streams, this can be improved to a $1 - \epsilon$ approximation, by finding augmenting paths with successive passes [25, 26].

Subsequent work has extended to the case of weighted edges (when a maximum weight matching is sought), and reducing the number of passes to provide a guaranteed approximation [14, 13]. While approximating the size of the vertex cover has been studied in other sublinear models, such as sampling [31, 30], we are not aware of prior work that has addressed the question of finding a vertex cover over a graph stream. Likewise, parameterized complexity has not been previously studied in dynamic graph streams with both insertions and deletions.

The model of dynamic graph streams has recently received much attention, due to breakthroughs by Ahn, Guha and McGregor [2, 3]. Over two papers, they showed the first results for a number of graph problems over dynamic streams, including determining connected components, testing bipartiteness, minimum spanning tree weight and building a sparsifier. They also gave multipass algorithms for maximum weight matchings and spanner constructions. This has provoked much interest into what can be computed over dynamic graph streams.

Outline. Section 2 provides background on techniques for kernelization of graph problems, and on streaming algorithms for building a sketch to recover a compact set. Our results on PSA and DPSA for matching and vertex cover are stated in Section 3 and Section 6, respectively, while Section 4 provides lower bounds for

these problems. Section 5 is the most involved, as it addresses the most difficult dynamic case in the promised model. Some observations on the (parameterized) feedback vertex set problem are presented in Section 7.

2 Preliminaries

In this section, we present the definitions of streaming model and the graph sketching that we use.

Streaming Model. Let S be a stream of insertions (or similarly, insertions and deletions) of edges of an underlying graph $G(V, E)$. We assume that vertex set V is fixed and given, and the size of V is $|V| = n$. We assume that the size of stream S is $|S| \leq n^c$ for some large enough constant c so that we may assume that $O(\log |S|) = O(\log n)$. Here $[x] = \{1, 2, 3, \dots, x\}$ when $x \in \mathbb{N}$. Throughout the paper we denote failure probabilities by δ , and approximation parameters by ϵ .

We assume that there is a unique numbering for the vertices in V so that we can treat $v \in V$ as a unique number v for $1 \leq v \leq n = |V|$. We denote an undirected edge in E with two endpoints $u, v \in V$ by (u, v) . The graph G can have at most $\binom{n}{2} = n(n-1)/2$ edges. Thus, each edge can also be thought of as referring to a unique number between 1 and $\binom{n}{2}$.

At the start of stream S , edge set E is an empty set. We assume in the course of stream S , the maximum size of E is a number m , i.e., $m' = |E| \leq m$. Counter m' stores the current number of edges of stream S , i.e., after every insertion we increment m' by one and after every deletion we decrement m' by one.

Let M be a maximal matching that we maintain for stream S . Edges in M are called *matched* edges; the other edges are *free*. If uv is a matched edge, then u is the *mate* of v and v is the *mate* of u . Let V_M be the vertices of M and $\bar{V}_M = V \setminus V_M$. A vertex v which is in V_M is called a *matched* vertex, otherwise, i.e., if $v \in \bar{V}_M$, v is called an *exposed* vertex.

The *neighborhood* of a vertex $u \in V$ is defined as $\mathcal{N}_u = \{v \in V : uv \in E\}$. Hence the degree of a vertex $u \in V$ is $d_u = |\{uv \in E\}| = |\mathcal{N}_u|$. We split the neighborhood of u into the set of matched neighbors of u , $\mathcal{N}_u \cap V_M$, and the set of exposed neighbors of u , i.e., $\mathcal{N}_u \setminus V_M$.

Oblivious Adversarial Model. We work in the *oblivious adversarial model* as is common for analysis of randomized data structures such as universal hashing [9]. This model has been used in a series of papers on dynamic maximal matching and dynamic connectivity problems: see for example [29, 6, 22, 28]. The model allows the adversary to know all the edges in the graph $G(V, E)$ and their arrival order, as well as the algorithm to be used. However, the adversary is not aware of the random bits used by the algorithm, and so cannot

choose updates adaptively in response to the randomly guided choices of the algorithm. This effectively means that we can assume that the adversary prepares the full input (inserts and deletes) before the algorithm runs.

k -Sparse Recovery Sketch and Graph Sketching. We first define an ℓ_0 -Sampler as follows.

DEFINITION 2.1. (ℓ_0 -SAMPLER [18, 27]) *Let $0 < \delta < 1$ be a parameter. Let $S = (a_1, t_1), \dots, (a_i, t_i), \dots$ be a stream of updates of an underlying vector $x \in \mathbb{R}^n$ where $a_i \in [n]$ and $t_i \in \mathbb{R}$. The i -th update (a_i, t_i) updates the a_i -th element of x using $x[a_i] = x[a_i] + t_i$. A ℓ_0 -sampler algorithm for $x \neq 0$ returns FAIL with probability at most δ . Else, with probability $1 - \delta$, it returns an element $j \in [n]$ such that the probability that j -th element is returned is $\Pr[j] = \frac{|x_j|_0}{\ell_0(x)}$.*

Here, $\ell_0(x) = (\sum_{i \in [n]} |x_i|_0)$ is the (so-called) “0-norm” of x that counts the number of non-zero entries.

LEMMA 2.1. ([21]) *Let $0 < \delta < 1$ be a parameter. There exists a linear sketch-based algorithm for ℓ_0 -sampling using $O(\log^2 n \log \delta^{-1})$ bits of space.*

The concepts behind sketches for ℓ_0 -sampling can be generalized to draw k distinct elements from the support set of x :

DEFINITION 2.2. (k -SAMPLE RECOVERY) *A k -sample recovery algorithm recovers $\min(k, \|x\|_0)$ elements from x such that sampled index i has $x_i \neq 0$ and is sampled uniformly.*

Constructions of k -sample recovery mechanisms are known which require space $\tilde{O}(k)$ and fail only with probability polynomially small in n [5]. We apply this algorithm to the neighborhood of vertices: for each node v , we can maintain an instance of the k -sample recovery sketch (or algorithm) to the vector corresponding to the row of the adjacency matrix for v . Note that as edges are inserted or deleted, we can propagate these to the appropriate k -sample recovery algorithms, without needing knowledge of the full neighborhood of nodes.

Specifically, let $a_1, \dots, a_v, \dots, a_n$ be the rows of the adjacency matrix of G , \mathcal{A}_G , where a_v encodes the neighborhood of a vertex $v \in V$. We define the sketch of \mathcal{A}_G as follows. Let S be a stream of insertions and deletions of edges to an underlying graph $G = (V, E)$. We sketch each row a_u of \mathcal{A}_G using the sketching matrix of Lemma 2.1. Let us denote this sketch by S_u . Since sketch S is linear, the following operations can be done in the sketch space.

- **QUERY(S_u):** This operation queries sketch S_u to find a uniformly random neighbor of vertex u .

Since S_u is a k -sample recovery sketch, we can query up to k uniformly random neighbors of vertex u .

- **UPDATE($S_u, \pm(u, v)$):** This operation updates the sketch of a vertex u . In particular, operation **UPDATE($S_u, (u, v)$)** means that edge (u, v) is added to sketch S_u . And, operation **UPDATE($S_u, -(u, v)$)** means that edge (u, v) is deleted from sketch S_u .

3 Parameterized Streaming Algorithm (PSA) for $VC(k)$

To build intuition, we give a simple $(k^2, 2^{O(k)})$ -PSA for $VC(k)$. In Section 4 we show that the space complexity of this PSA is optimal even if we use randomization. First, we review the kernelization algorithm of Buss and Goldsmith [7] since we use it in our PSA for $VC(k)$.

3.1 Kernel for $VC(k)$ Let (G, k) be the original instance of the problem which is initialized by graph $G = (V, E)$ and parameter k . Let d_v denote the degree of v in G . While one of the following rules can be applied, we follow it.

- (1) **There exists a vertex $v \in G$ with $d_v > k$:** Observe that if we do not include v in the vertex cover, then we must include all of \mathcal{N}_v . Since $|\mathcal{N}_v| = d_v > k$, we must include v in our vertex cover for now. Update $G \leftarrow G \setminus \{v\}$ and $k \leftarrow k - 1$.
- (2) **There is an isolated vertex $v \in G$:** Remove v from G , since v cannot cover any edge.

If neither of above rules can be applied, then we look at the number of edges of G . Note that the maximum degree of G is now $\leq k$. Hence, if G has a vertex cover of size $\leq k$, then the maximum number of edges in G is k^2 . If $|E| > k^2$, then we can safely answer NO. Otherwise we now have a kernel graph $G = (V, E)$ such that $|E| \leq k^2$. Since G does not have any isolated vertex, we have $|V| \leq 2|E| \leq 2k^2$. Observe that we obtain the kernel graph G in polynomial time.

Remark: The FPT algorithm of Chen et al. runs in time $1.2378^k + k \cdot n$, where n is the number of vertices. In the above kernel graph, we have $|V| \leq 2k^2$ and hence the Chen et al. algorithm runs in time $1.2378^k + k \cdot 2k^2 = 2^{O(k)}$.

3.2 $(k^2, 2^{O(k)})$ -PSA for $VC(k)$ We now prove Theorem 3.1, which is restated below:

THEOREM 3.1. *Let S be a stream of insertions of edges of an underlying graph G . Then there exists a deterministic $(k^2, 2^{O(k)})$ -PSA for $VC(k)$ problem.*

The proof is divided into three parts: first we describe the algorithm, analyze its complexity and then show its correctness.

Algorithm. Let S be a stream of insertions of edges of an underlying graph $G(V, E)$. We maintain a maximal matching M of stream S in a greedy fashion. Let V_M be the vertices of matching M . For every matched vertex v , we also store up to k edges incident on v in a set E_M . If at the i th update of stream S we observe that $|M| > k$, we report that the size of any vertex cover of $G = (V, E)$ is more than k and quit. At the end of stream S , we run the kernelization algorithm of Section 3.1 on instance $(G_M = (V_M, E_M), k)$.

Complexity of the Algorithm. We observe that the space complexity of the algorithm is $O(k^2)$. In fact, for each vertex $v \in V_M$ assuming $|M| \leq k$ we keep at most k incident edges, thus we need space of at most $2k \cdot k = 2k^2$. If $|M| > k$, as soon as the size of the matching M goes beyond k we quit the algorithm and so in this case we also use space of at most $2k \cdot k = 2k^2$. The query time of this algorithm is dominated by the time to extract the vertex cover of G_M (and hence also of G) using the FPT algorithm of Chen et al. [10] which runs in time $1.2378^k + k \cdot |V_M| = 2^{O(k)}$, since $|V_M| = O(k^2)$.

Correctness proof. We argue that

- (1) if the kernelization algorithm succeeds on instance $(G_M = (V_M, E_M), k)$ and finds a vertex cover of size at most k for G_M , then that vertex cover is also a vertex cover of size at most k for G .
- (2) On the other hand, if the kernelization algorithm reports that instance $(G_M = (V_M, E_M), k)$ does not have a vertex cover of size at most k , then instance $(G = (V, E), k)$ does not have a vertex cover of size at most k .

First, note that trivially, any matching provides a lower bound on the size of the vertex cover, and hence we are correct to reject if $|M| > k$.

Otherwise, i.e., if $|M| \leq k$, we write d_v and d'_v for the degree of v in G and G_M , respectively. We follow rules of the kernelization algorithm on G and G_M in lockstep. Observe that since every edge $e \in E$ is incident on at least one matched vertex $v \in V_M$, when an edge $(u, v) \in E$ is not stored in E_M it is in one of the following cases.

- (1) $u \in V_M$ and $v \in V_M$: Then, we must have $d'_u > k$ and $d'_v > k$ which means that $d_u > k$ and $d_v > k$.
- (2) Only $u \in V_M$: Then, we must have $d'_u > k$ which means that $d_u > k$.
- (3) Only $v \in V_M$: Then, we must have $d'_v > k$ which means that $d_v > k$.

Now, let us consider a set $X = \{v_k, v_{k-1}, \dots, v_r\}$ (for $r \geq 0$) of vertices that Rule (1) of the kernelization algorithm for G_M removes. According to Rule (1), for a vertex $v_{k'} \in X$ (for $k \geq k' \geq r$) with $d'_{v_{k'}} > k'$, we remove $v_{k'}$ and all edges incident on $v_{k'}$ from G_M and decrease k' by one. Note that $d'_{v_{k'}} > k'$ if and only if $d_{v_{k'}} > k'$. This is due to the fact that, before we remove vertex $v_{k'}$ from G_M , we have removed only those neighbors of $v_{k'}$ that are matched and the number of such vertices is less than $k - k'$. Thus, Rule (1) of the kernelization algorithm can be applied on G and we remove $v_{k'}$ and all edges incident on $v_{k'}$ from G and decrease k' by one.

Next we consider Rule (2). Assume in one step of the kernelization algorithm for G_M , we have an isolated vertex $v \in G_M$. Observe that those neighbors of v that we have removed using Rule (1) (before vertex v becomes isolated) are all matched vertices and the number of such vertices is less than k . Moreover, v never had any neighbor in $V \setminus V_M$ otherwise, v is not isolated. Thus, if v has a neighbor u in the remaining vertices of V_M , edge (u, v) must be in E_M as we store up to k edges incident on v in set E_M which means v is not isolated in G_M and that is in contradiction to our assumption that v is isolated in G_M . Since we run the kernelization algorithm on G_M and on G for the vertices in set X , the same thing happens for G , i.e., v in G is also isolated. So, using Rule (2), v is removed from G_M if and only if v is removed from G .

Now assume neither Rule (1) nor Rule (2) can be applied for G_M , but the number of edges in E_M is more than k^2 . The same thing must happen for E . Therefore, G_M and G do not have a vertex cover of size at most k .

If none of the above rules can be applied for G_M , we have a kernel (G_M, k') such that $|V_M| \leq 2k$ and $|E_M| \leq k'^2 \leq k^2$. Now observe that after removal of all vertices of X and their incident edges from G , for every remaining vertex v in G_M , $d_v \leq k'$; otherwise $d_v > k'$ and $d'_v > k'$; so we can apply Rule (1) which is in contradiction to our assumption that none of the above rules can be applied for G_M . Therefore, kernel (G_M, k') is also a kernel for (G, k') and this proves the correctness of our algorithm.

4 $\Omega(k^2)$ Lower Bound for $VC(k)$

We prove Theorem 4.1 which is restated below:

THEOREM 4.1. *Any (randomized) PSA for the $VC(k)$ problem requires $\Omega(k^2)$ space.*

Proof. We will reduce from the INDEX problem in communication complexity:

INDEX

Input: Alice has a string $X \in \{0,1\}^n$ given by $x_1x_2 \dots x_n$. Bob has an index $\iota \in [n]$

Question: Bob wants to find x_ι , i.e., the ι^{th} bit of X .

It is well-known that there is a lower bound of $\Omega(n)$ bits in the one-way randomized communication model for Bob to compute x_i [24]. We assume an instance of the index problem where n is a perfect square, and let $k = \sqrt{n}$. Fix a canonical mapping from $[n] \rightarrow [k] \times [k]$. Consequently we can interpret the bit string as an adjacency matrix for a bipartite graph with k vertices on each side.

From the instance of INDEX, we construct an instance G_X of Vertex Cover. Assume that Alice has an algorithm which solves the $VC(k)$ problem using $f(k)$ bits. For each $i \in [k]$, we have vertices, v_i, v'_i, v''_i , and w_i, w'_i, w''_i . First, we insert the edges corresponding to the edge interpretation of X between nodes v_i and w_j : for each $i, j \in [k]$, Alice adds the edge (v_i, w_j) if the corresponding entry in X is 1. Alice then sends the memory contents of her algorithm to Bob, using $f(k)$ bits.

Bob has the index $\iota \in [n]$, which he interprets as (I, J) under the same canonical remapping to $[k] \times [k]$. He receives the memory contents of the algorithm, and proceeds to add edges to the instance of vertex cover. For each $i \in [k], i \neq I$, Bob adds the edges (v_i, v'_i) and (v_i, v''_i) . Similarly, for each $j \in [k], j \neq J$, Bob adds the edges (w_j, w'_j) and (w_j, w''_j) .

The next lemma shows that finding the minimum vertex cover of G_X allows us to solve the corresponding instance of INDEX.

LEMMA 4.1. *The minimum size of a vertex cover of G_X is $2k - 1$ if and only if $x_\iota = 1$.*

Proof. Suppose $x_\iota = 0$. Then it is easy to check that the set $\{v_i : i \in [k], i \neq I\} \cup \{w_j : j \in [k], j \neq J\}$ forms a vertex cover of size $2k - 2$ for G_X .

Now suppose $x_\iota = 1$, and let Y be a minimum vertex cover for G_X . For any $i \in [k], i \neq I$ the vertices v'_i and v''_i have degree one in G_X . Hence, without loss of generality, we can assume that $v_i \in Y$. Similarly, $w_j \in Y$ for each $j \in [k], j \neq J$. This covers all edges except (v_I, w_J) . To cover this we need to pick one of v_I or w_J , which shows that $|Y| = 2k - 1$.

Thus, by checking whether the output of \mathcal{A} on the instance G_X of $VC(k)$ is $2k - 1$ or $2k - 2$, Bob can determine the index x_ι . The total communication between Alice and Bob was $O(f(k))$ bits, and hence we can solve the INDEX problem in $f(k)$ bits. Recall that the lower bound for the INDEX problem is $\Omega(n) = \Omega(k^2)$, and hence we have $f(k) = \Omega(k^2)$.

COROLLARY 4.1. *Let $1 > \epsilon > 0$. Any (randomized) PSA that approximates $VC(k)$ within a relative error of ϵ requires $\Omega(\frac{1}{\epsilon^2})$ space.*

Proof. Choose $\epsilon = \frac{1}{2k}$. Theorem 4.1 shows that the relative error is at most $\frac{1}{2k-1}$, which is less than ϵ . Hence finding an approximation within ϵ relative error amounts to finding the exact value of the vertex cover. The lower bound of $\Omega(k^2)$ from Theorem 4.1 translates to $\Omega(\frac{1}{\epsilon^2})$ here.

5 Promised Dynamic Parameterized Streaming Algorithm (PDPSA) for $VC(k)$

In this section we prove our main theorem, i.e., Theorem 1.3. We first explain the outline of our algorithm. We then give the detailed description of the algorithm and the proof of Theorem 1.3.

It is natural to first think of solutions which keep some summary (sketch) information for various vertices. However, many natural such attempts end up in keeping a large number of sketches. Our aim is to provide a solution whose cost is bounded by a polynomial of k , which means we cannot allow such solutions. Instead, we must only materialize a small number of sketches of vertices, and add/remove these so as to bound the total quantity of sketches. This distinguishes this work from prior algorithms for problems in graph streaming which maintain a sketch for each vertex [2, 3, 22]

5.1 Outline We develop a streaming algorithm that maintains a maximal matching of underlying graph $G(V, E)$ in a streaming fashion. At the end of stream S we run the kernelization algorithm of Section 3.1 on the maintained maximal matching. Our data structure to maintain a maximal matching M of stream S consists of two parts.

First, for each matched vertex u , we maintain an x -sample recovery sketch S_u of its incident edges, where x is chosen to be $\tilde{O}(k)$. Insertions of new edges are relatively easy to handle: we update the matching with the edge if we can, and update the sketches if the new edge is incident on matched nodes. The difficulty arises with deletions of edges: we must try to “patch up” the matching, so that it remains maximal, using only the stored information, which is constrained to be $O(k^2)$. The intuition behind our algorithm is that, given the promise, there cannot be more than k matched nodes at any time. Therefore, keeping $\tilde{O}(k)$ information about the neighborhood of each matched node can be sufficient to identify any adjacent unmatched nodes with which it can be paired if it becomes unmatched. However, this intuition requires significant care and case-analysis to put into practice. The reason is we need

some extra book-keeping to record where information is stored, since nodes are entering and leaving the matching, and we do not necessarily have access to the full neighborhood of a node when it is admitted to the matching. Nevertheless, we show that additional book-keeping information of size $O(k^2)$ is sufficient for our purposes, allowing us to meet the $O(k^2)$ space bound.

This book-keeping comes in the form of another data structure \mathcal{T} , that stores a set of edges (u, v) such that both endpoints are matched (not necessarily to each other), and (u, v) has been inserted into sketches S_u and S_v , but not deleted from them. The size of \mathcal{T} is clearly $O(k^2)$. To implement \mathcal{T} , we can adopt any fast dictionary data structure (AVL-tree, red-black tree, or hash-tables).

The update at a time t is either the insertion or the deletion of an edge (u, v) for $1 \leq t \leq |S|$ where $|S| \leq n^c$ is the length of stream S . We continue our outline of the algorithm by describing the behavior in each case informally, with the formal details given in subsequent sections.

Insertion of an Edge (u, v) at Time t . When the update at time t is insertion of an edge (u, v) two cases can occur. The first case is if at least one of u and v is matched, we insert edge (u, v) to the sketches of those vertices (from u and v) which are matched. If both u and v are matched, we also insert (u, v) to \mathcal{T} .

The second case occurs if both vertices u and v are exposed. We add edge (u, v) to the current matching and to \mathcal{T} , and initialize sketches S_u and S_v by insertion of edge (u, v) to S_u and S_v . However, we also need to perform some additional book-keeping updates to ensure that the information is up to date. Fix vertex u . There can be matched vertices, say $w \in V_M$, which are neighbors of u . If previously an edge (w, u) arrived while u was not in the matching, then we inserted (w, u) to sketch S_w , but (w, u) was not inserted to sketch S_u as u was an exposed vertex at that time. If at some subsequent point w becomes an exposed vertex and matching edge (u, v) is deleted then vertex u must have the option of choosing an unexposed vertex w to be rematched. For that, we need to ensure that some information about (w, u) is accessible to the algorithm.

A first attempt to address this is to try interrogating each sketch S_w for all edges incident on u , say when u is first added to the matching. However, this may not work while respecting the space bounds: w may have a large number of neighbors, much larger than the limit x . In this case, we can only use S_w to recover a sample of the neighbors of w , and u may not be among them.

To solve this problem we must wait until w has low enough degree that we can retrieve its complete neighborhood from S_w . At this point, we can use

these recovered edges to update the sketches of other matched nodes. We use the structure \mathcal{T} to track information about edges on matched vertices that are already represented in sketches, to avoid duplicate representations of an edge. This is handled during the deletion of an edge, since this is the only event that can cause the degree of a node w to drop.

Deletion of an Edge (u, v) at Time t . When the update at time t is deletion of an edge (u, v) , we have three cases to consider. The first case is if only one of vertices u and v is matched, we delete edge (u, v) from the sketch of that matched vertex.

The second case is if both u and v are matched vertices, but $(u, v) \notin M$. We want to delete edge (u, v) from sketches S_u and S_v , but (u, v) might not be represented in both these sketches. We need to find out if (u, v) has been inserted to S_u and S_v , or only to one of them. This can be found from \mathcal{T} . If $(u, v) \in \mathcal{T}$, edge (u, v) has been inserted to both S_u and S_v . So, we delete (u, v) from both sketches safely. Otherwise, i.e., if $(u, v) \notin \mathcal{T}$, (u, v) has been inserted to the sketch of only one of u and v . Assume that this is u . To discover this we define *timestamps* for matched vertices. The timestamp t_i of a matched vertex u is the (most recent) time when u was matched. We show that edge (u, v) is only in sketch S_u (not S_v) if and only if $(u, v) \notin \mathcal{T}$ and $t_u < t_v$. Therefore, if $t_u < t_v$, we delete (u, v) from sketch S_u . Otherwise, i.e., if $t_v < t_u$, we delete (u, v) from sketch S_v . Observe that if $t_u = t_v$, we have inserted (u, v) to S_u and S_v as well as \mathcal{T} .

The third case is when $(u, v) \in M$. We delete edge (u, v) from sketches S_u and S_v as well as matching M and \mathcal{T} . To maintain the maximality of matching M we need to see whether we can rematch u and v . Let us consider u (the case for v is identical). If u has high degree, we sample edges (u, z) from sketch S_u . Given the size of the sketch, we argue that there is high probability of finding an edge to rematch u . Meanwhile, if u is low degree, then we can recover its full neighborhood, and test whether any of these can match u . Otherwise, u is an exposed vertex, and its sketch is deleted. We also remove all edges incident on u from \mathcal{T} .

We now describe and prove the properties of PDPSA for $VC(k)$ in full. We begin with notations, data structures and invariants.

5.2 Notations, Data Structures and Invariants

Timestamp of a Vertex and an Edge. We define *time* t corresponding to the t -th update operation (insert or delete of an edge) in stream S . We define the *timestamp of a matched vertex* as follows. Let u be a matched vertex at time t . Let $t' \leq t$ be the greatest time such that u was unmatched before time

t' and is matched in the interval $[t', t]$. Then we say the timestamp t_u of vertex u is t' and we set $t_u = t'$. If at time t , vertex u is exposed we define $t_u = \infty$, i.e. a value larger than any timestamp.

We define the *timestamp of an edge* as follows. Let E_t denote the set of edges present at time t , i.e. which have been inserted without a corresponding deletion. Let $t' \leq t$ be the last time in which the edge $(u, v) \in E_t$ is inserted but not deleted in the interval $[t', t]$. Then we say the timestamp $t_{(u,v)}$ of edge (u, v) is t' and we set $t_{(u,v)} = t'$. If at timestamp t , edge (u, v) is deleted we define $t_{(u,v)} = \infty$.

Low and High Degree Vertices. Let $x = 8ck \cdot \log(n/\delta)$, for constant c (where, we assume that $|S| = O(n^c)$). At time t we say a vertex u is a *high-degree* vertex if $d_u > x$; otherwise, if $d_u \leq x$, we say u is a *low-degree* vertex.

Data Structures: For every matched vertex u , i.e., $u \in V_M$, we maintain an x -sample recovery sketch S_u of edges incident on u . We also maintain a dictionary data structure \mathcal{T} of size $O(k^2)$. We assume the insertion, deletion and query times of \mathcal{T} are all worst-case $O(\log k)$. At every time t , \mathcal{T} stores edges (u, v) for which vertices u and v are matched at time t (not necessarily to each other); and also edge (u, v) is represented in both sketches S_u and S_v , i.e. there is a time $t' \leq t$ at which we invoked $\text{UPDATE}(S_u, (u, v))$, but there is no time in interval $[t', t]$ in which we have invoked $\text{UPDATE}(S_u, -(u, v))$, and symmetrically for S_v .

Sketched Neighbors of a Vertex: Let u be a matched vertex at some time t , i.e., $u \in V_M$. Recall that $\mathcal{N}_u = \{v \in V : (u, v) \in E_t\}$ is the full neighborhood of u at time t . Let $\mathcal{N}'_u \subseteq \mathcal{N}_u$ be the set of neighbors of u that up to time t are inserted to S_u but not deleted from S_u , that is for every vertex $v \in \mathcal{N}'_u$ we have invoked $\text{UPDATE}(S_u, (u, v))$ at a time $t' \leq t$ but have not invoked $\text{UPDATE}(S_u, -(u, v))$ in time interval $[t', t]$. We call the vertices in \mathcal{N}'_u the sketched neighbors of vertex u . Note that we can recover \mathcal{N}'_u exactly when $|\mathcal{N}'_u| < x$.

Invariants. Recall that at every time t of stream S , set E_t is the set of edges which are inserted but not deleted up to time t . We maintain three invariants.

- **Invariant 1:** For every edge $(u, v) \in E_t$ at time t we have at least one of $v \in \mathcal{N}'_u$ or $u \in \mathcal{N}'_v$.
- Let $(u, v) \in E_t$ be an edge at time t such that $u, v \in V_M$. At time t ,
 - **Invariant 2:** $u \notin \mathcal{N}'_v$ iff $t_u < t_v$ and $(u, v) \notin \mathcal{T}$.
 - **Invariant 3:** $v \in \mathcal{N}'_u$ and $u \in \mathcal{N}'_v$ iff $(u, v) \in \mathcal{T}$.

Observe that these invariants imply that at any time $|\mathcal{T}| < 2k^2$. That is, since \mathcal{T} only holds edges such that both ends are matched, and we assume that the matching has at most $2k$ nodes, then the number of edges can be at most $\binom{2k}{2} < 2k^2$.

5.3 Adding an Edge to Matching M The first primitive that we develop is Procedure $\text{AddEdgeToMatching}((u, v), t)$. This procedure first adds edge (u, v) to matching M and data structure \mathcal{T} . Then it inserts vertex u to V_M , sets timestamp t_u to the current time t , and initializes sketch S_u by inserting edge (u, v) to sketch S_u . It also repeats these steps for v . We invoke this procedure in Procedures $\text{Rematch}((u, v), t)$ and $\text{Insertion}((u, v), t)$.

Insertion $((u, v), t)$

- (1) If $u \notin V_M$ and $v \notin V_M$, then $\text{AddEdgeToMatching}((u, v), t)$.
- (2) Else $\text{InsertToDS}((u, v))$.

AddEdgeToMatching $((u, v), t)$

- (1) Add edge (u, v) to M and \mathcal{T} .
- (2) For $z \in \{u, v\}$
 - (a) $V_M \leftarrow V_M \cup \{z\}$
 - (b) $t_z \leftarrow t$
 - (c) Initialize sketch S_z with $\text{UPDATE}(S_z, (u, v))$.

LEMMA 5.1. *Let t be a time when we invoke Procedure $\text{AddEdgeToMatching}((u, v), t)$. Suppose before time t , Invariants 1, 2 and 3 hold. Then, Invariants 1, 2 and 3 hold after time t .*

Proof. Recall that t_u is the last time $t' \leq t$ such that u before time t' was unmatched and is matched in the interval $[t', t]$. Similarly, t_v is the last time $t' \leq t$ such that v before time t' was unmatched and is matched in the interval $[t', t]$.

In Procedure $\text{AddEdgeToMatching}((u, v), t)$ we insert (u, v) to sketches S_u and/or S_v if the edge has not been inserted to these sketches. So, at time t , Invariant 1 for edge (u, v) holds. Since $(u, v) \in M$, nothing changes for Invariants 2 and 3. Therefore, if Invariants 2 and 3 hold at time $t - 1$, they also hold at time t .

5.4 Maintenance of Data Structure \mathcal{T} To maintain data structure \mathcal{T} at every time t of stream S , we develop two procedures to handle insertions and deletions

to the structure. If u and v are matched vertices, Procedure $\text{InsertToDS}((u, v))$ inserts edge (u, v) to sketches S_u and S_v as well as to data structure \mathcal{T} . If only one of u and v is matched, we insert (u, v) to the sketch of the matched vertex. We invoke this procedure upon insertion of an arbitrary edge (u, v) inside Procedure $\text{Insertion}((u, v), t)$.

InsertToDS $((u, v))$

- (1) If $u \in V_M$ and $v \in V_M$ then insert edge (u, v) into \mathcal{T} .
- (2) If $u \in V_M$ then $\text{UPDATE}(S_u, (u, v))$.
- (3) If $v \in V_M$ then $\text{UPDATE}(S_v, (u, v))$.

LEMMA 5.2. *Let t be a time of stream S when we invoke Procedure $\text{InsertToDS}((u, v))$. Suppose before time t , Invariants 1, 2 and 3 hold. Then, Invariants 1, 2 and 3 hold after time t .*

Proof. First assume at time t when we invoke Procedure $\text{InsertToDS}((u, v))$, vertices u and v are already matched. In Procedure $\text{InsertToDS}((u, v))$ we insert (u, v) to sketches S_u and S_v using $\text{UPDATE}(S_u, (u, v))$ and $\text{UPDATE}(S_v, (u, v))$. So, $u \in \mathcal{N}'_v$ and $v \in \mathcal{N}'_u$ and Invariant 1 holds. Moreover, we insert (u, v) to \mathcal{T} . Therefore, Invariant 3 holds. Invariant 2 also holds as neither condition is true ($v \notin \mathcal{N}'_u$ and $(u, v) \notin \mathcal{T}$).

Next assume only vertex u is matched. We insert (u, v) to sketch S_u , but not to S_v and \mathcal{T} . Since $v \in \mathcal{N}'_u$, Invariant 1 is correct. Invariant 2 and 3 are correct as v is not matched at time t . The case when only vertex v is matched is symmetric.

The second procedure is $\text{DeleteFromDS}((u, v))$ which is invoked in Procedure $\text{Deletion}((u, v), t)$ when $(u, v) \notin M$. There are three main cases to consider. If $(u, v) \in \mathcal{T}$, we delete (u, v) from sketches S_u and S_v as well as data structure \mathcal{T} . If not, we know that (u, v) is only in one of S_u and S_v .

If $t_u < t_v$ and both u and v are matched, we delete the edge from S_u , otherwise, if $t_v < t_u$ and u and v are matched from S_v , we delete the edge from S_v . If none of these cases occur, then only one of u and v is matched. If the matched vertex is u , we delete (u, v) from S_u . Otherwise, we delete (u, v) from S_v .

Deletion $((u, v), t)$

- (1) If $(u, v) \in M$ then invoke $\text{Rematch}((u, v), t)$
- (2) Else invoke $\text{DeleteFromDS}((u, v))$.
- (3) Invoke $\text{AnnounceNeighborhood}(u)$ and $\text{AnnounceNeighborhood}(v)$

DeleteFromDS $((u, v))$

- (1) If $(u, v) \in \mathcal{T}$ then
 - (a) $\text{UPDATE}(S_u, -(u, v))$ and $\text{UPDATE}(S_v, -(u, v))$.
 - (b) Remove (u, v) from \mathcal{T} .
- (2) Else if $t_u < t_v$ and $u, v \in V_M$ then $\text{UPDATE}(S_u, -(u, v))$.
- (3) Else if $t_v < t_u$ and $u, v \in V_M$ then $\text{UPDATE}(S_v, -(u, v))$.
- (4) Else if $u \in V_M$ and $v \notin V_M$ then $\text{UPDATE}(S_u, -(u, v))$.
- (5) Else if $v \in V_M$ and $u \notin V_M$ then $\text{UPDATE}(S_v, -(u, v))$.

LEMMA 5.3. *Assume Invariants 1, 2 and 3 hold at time t when Procedure $\text{DeleteFromDS}((u, v))$ is invoked. Then, Procedure $\text{DeleteFromDS}((u, v))$ chooses the correct case.*

Proof. First, we consider the case that both u and v are matched vertices. Since Invariant 3 holds, we know that edge (u, v) at time t is in \mathcal{T} if and only if $v \in \mathcal{N}'_u$ and $u \in \mathcal{N}'_v$. Procedure $\text{DeleteFromDS}((u, v))$ searches for (u, v) in \mathcal{T} . If this finds (u, v) in \mathcal{T} , we then know that $v \in \mathcal{N}'_u$ and $u \in \mathcal{N}'_v$. So, we can safely delete the edge from sketches S_u and S_v and data structure \mathcal{T} .

On the other hand, if $(u, v) \notin \mathcal{T}$, we ensure that the edge is in only one of S_u and S_v . Now, we can use the claim of Invariant 2 which says $u \notin \mathcal{N}'_v$ if and only if $t_u < t_v$ and $(u, v) \notin \mathcal{T}$. We compare t_u and t_v . If $t_u < t_v$, then $u \notin \mathcal{N}'_v$. Recall that since Invariant 1 holds, we know that at least one of $v \in \mathcal{N}'_u$ and $u \in \mathcal{N}'_v$ is correct. Because $u \notin \mathcal{N}'_v$, we must have $v \in \mathcal{N}'_u$. So deleting edge (u, v) from sketch S_u is the correct operation. On the other hand, if $t_v < t_u$, then $v \notin \mathcal{N}'_u$ and so edge (u, v) is only in sketch S_v . Thus, deleting edge (u, v) from sketch S_v is the correct operation.

Next we consider the case that only one of u and v is matched. Let us assume u is the matched vertex. Since Invariant 1 holds, we know that at least one of $v \in \mathcal{N}'_u$ and $u \in \mathcal{N}'_v$ is correct. Because u is the matched vertex and we maintain the sketch of matched vertices, (u, v) has been inserted to sketch S_u that is $v \in \mathcal{N}'_u$. Therefore, deleting edge (u, v) from sketch S_u is the correct operation. The case when v is the matched vertex is symmetric.

5.5 Announcement and Deletion of Neighborhood of a Vertex In this section we develop ba-

sic primitives for the announcement and deletion of the neighborhood of a vertex. Announcement is performed by Procedure **AnnounceNeighborhood**(u) which is invoked in **Deletion**((u, v), t). Suppose that node u has low degree. For every matched vertex $v \in \mathcal{N}'_u$, we search for edge (u, v) in \mathcal{T} . If $(u, v) \in \mathcal{T}$, (u, v) is in both S_u and S_v and no action is needed. If not, we insert edge (u, v) into tree \mathcal{T} as well as sketch S_v .

AnnounceNeighborhood(u)

- (1) If $u \in V_M$ and $d_u \leq x$, then
 - (a) For every edge (u, v) in sketch S_u
 - i. Add v to set \mathcal{N}'_u .
 - (b) For every $v \in \mathcal{N}'_u \cap V_M$
 - i. If edge $(u, v) \notin \mathcal{T}$, then insert (u, v) to \mathcal{T} ; **UPDATE**($S_v, (u, v)$).

We also introduce a deletion primitive in the form of Procedure **DeleteNeighborhood**(u). This is invoked in **Rematch**((u, v), t) when the matched edge (u, v) is removed. The **DeleteNeighborhood**(u) procedure is called on a node u when all the following three conditions hold.

- (1) The matched edge of matched vertex u is deleted.
- (2) Vertex u is a low-degree vertex.
- (3) Vertex u does not have any exposed neighbor.

In this case, we need to delete u from V_M and delete incident edges on u from data structure \mathcal{T} as Invariant 3 for u is not valid anymore. More precisely, for a low-degree matched vertex whose neighborhood are all matched we do as follows.

We recover all edges from the sketch S_u (i.e. \mathcal{N}'_u). For every edge $(u, v) \in \mathcal{N}'_u$, we check to see if $(u, v) \in \mathcal{T}$. If so, we know that (u, v) is represented in both sketches S_u and S_v . We also delete (u, v) from \mathcal{T} as u is not matched and Invariant 3 does not hold. But if $(u, v) \notin \mathcal{T}$, since Invariant 1 holds we know that (u, v) is inserted only in S_u not in S_v . Observe that since u does not have any exposed neighbor, vertex v must be a matched vertex, and so vertex v has an associated sketch S_v . Therefore, in order to fulfill Invariant 1, we first insert (u, v) to sketch S_v . Finally, we delete the whole sketch S_u , and remove u from V_M .

DeleteNeighborhood(u)

- (1) For every edge (u, v) in sketch S_u
 - (a) If edge $(u, v) \in \mathcal{T}$, then Remove (u, v) from \mathcal{T} .
 - (b) Else **UPDATE**($S_v, (u, v)$).
- (2) Delete sketch S_u and remove u from V_M .

LEMMA 5.4. *Let t be a time when we invoke Procedure **AnnounceNeighborhood**(u). Suppose u is a low-degree matched vertex at time t . Suppose before time t , Invariants 1, 2 and 3 hold. Then after time t , Invariants 1, 2 and 3 hold.*

Proof. Let \mathcal{N}'_u be the set of neighbors of u that up to time t are inserted into sketch S_u but not deleted from S_u . Since u at time t is a low-degree vertex we can use Definition 2.2 to recover \mathcal{N}'_u in its entirety. We assume Invariants 1, 2 and 3 hold before time t . We prove that all three invariants continue to hold after invocation of **AnnounceNeighborhood**(u).

Fix a matched neighbor v of u in \mathcal{N}'_u that is $v \in V_M \cap \mathcal{N}'_u$. In Procedure **AnnounceNeighborhood**(u) for v we do the following. If edge (u, v) has not been already inserted in \mathcal{T} , we insert edge (u, v) to \mathcal{T} and S_v . So, now $v \in \mathcal{N}'_u$ and $u \in \mathcal{N}'_v$, and $(u, v) \in \mathcal{T}$. Invariants 1, 2 and 3 hold for (u, v) , and continue to hold for all other edges.

After processing this deletion, edge (u, v) is no longer in E_t , and so the invariants trivially hold in regard of this edge. Meanwhile, for any other edge, if the invariants held before, then they continue to hold, since the changes only affected edge (u, v) .

LEMMA 5.5. *Suppose before time t , Invariants 1, 2 and 3 hold and we invoke **DeleteNeighborhood**(u) at time t . Here we assume u is a matched vertex whose neighbors are all matched, i.e., $\mathcal{N}_u \cap \bar{V}_M = \emptyset$. Then after time t , Invariants 1, 2 and 3 hold.*

Proof. Let (u, v) be an edge in sketch S_u . Since we assume Invariant 1 holds before time t , (u, v) must be inserted into at least one of S_u and S_v . We know edge (u, v) is in S_u . Since Invariants 2 and 3 hold, we have one of the two following cases.

(i) If edge (u, v) is also inserted to S_v , this means this edge must be in \mathcal{T} . In **DeleteNeighborhood**(u) if edge (u, v) is in \mathcal{T} , we delete the edge from \mathcal{T} as well as sketch S_u . As (u, v) is still in S_v , Invariant 1 after time t holds.

(ii) Else, edge (u, v) is not in S_v . Using Invariant 2 this happens if and only if $t_u < t_v$ and $(u, v) \notin \mathcal{T}$. We

want to delete all edges which are inserted to S_u and delete sketch S_u . Observe that since u does not have any exposed neighbor, vertex v must be a matched vertex and so has an associated sketch S_v . We insert (u, v) to sketch S_v , and subsequently S_u is deleted. Therefore, Invariant 1 still holds.

Finally, Invariants 2 and 3 hold after time t since u is not a matched vertex anymore.

5.6 Rematching Matched Vertices In this section we develop the last (and most involved) primitive, $\text{Rematch}((u, v), t)$. We invoke this procedure in Procedure $\text{Deletion}((u, v), t)$ when the matched edge (u, v) is deleted. We first delete edge (u, v) from sketches S_u and S_v as well as data structure \mathcal{T} . We also delete (u, v) from current set M of matched edges. To see if we can rematch u and v to one of their exposed neighbors, we perform the subsequent steps for u (and then repeat for v).

If u is a low degree vertex, by querying S_u we recover \mathcal{N}'_u , i.e., the set of neighbors of u that up to time t are inserted into sketch S_u but not deleted from S_u . We then check whether there is an exposed vertex $z \in \mathcal{N}'_u$. If so, we rematch u to z .

Rematch $((u, v), t)$

- (1) $\text{DeleteFromDS}((u, v))$, remove (u, v) from M , remove u, v from V_M
- (2) For $w \in \{u, v\}$
 - (a) If $d_w \leq x$ then
 - i. For every edge (w, z) in sketch S_w , add z to set \mathcal{N}'_w .
 - ii. If there is an exposed $z \in \mathcal{N}'_w$ then invoke $\text{AddEdgeToMatching}((w, z), t)$.
 - iii. Else invoke $\text{DeleteNeighborhood}(\text{vertex } w)$.
 - (b) If $d_w > x$ then
 - i. Query edges $(w, z_1), \dots, (w, z_y)$ from sketch S_w for $y = 8c \log(n/\delta)$.
 - ii. If there is an exposed $z \in \{z_1, \dots, z_y\}$ then invoke $\text{AddEdgeToMatching}((w, z), t)$.

But if there is no exposed vertex in \mathcal{N}'_u , we announce u as an exposed vertex. We also remove sketch S_u as u is not a matched vertex anymore. Moreover, we remove all incident edges on u from \mathcal{T} as our third invariant does not hold anymore. Lemma 5.6 shows that in both cases, the matching after invoking Procedure

$\text{Rematch}((u, v), t)$ is maximal if the matching before this invocation was maximal.

If u is a high degree vertex, it samples an edge (u, z) from sketch S_u . In Lemma 5.7 we show that with high probability z is an exposed vertex, so we rematch u to z . Therefore, if the matching before the invocation of Procedure $\text{Rematch}((u, v), t)$ is maximal, the matching after this invocation would be maximal as well.

5.6.1 Analyzing Rematching of a Low-Degree Vertex.

LEMMA 5.6. *Let u be a low-degree matched vertex at time t . Assuming the matching M before time t is maximal, then, after the invocation of Procedure $\text{Rematch}((u, v), t)$, the matching M is maximal. The running time of Procedure $\text{Rematch}((u, v), t)$ when u is a low-degree vertex is $O(k \log^4(n/\delta))$.*

Proof. Let \mathcal{N}'_u be the set of neighbors of u up to time t that are inserted into sketch S_u but not deleted from S_u . From Definition 2.2, by querying S_u and with probability at least $1 - \frac{\delta}{2n^c}$, we can recover \mathcal{N}'_u . Observe that assuming Invariants 1, 2 and 3 hold, we must have $\mathcal{N}_u \setminus \mathcal{N}'_u \subseteq V_M$, that is, those neighbors of u that are not in \mathcal{N}'_u at time t must be matched. Therefore, all exposed neighbors of u must be in \mathcal{N}'_u .

Two cases can occur. The first is if there is an exposed vertex z in \mathcal{N}'_u . Then, Procedure $\text{Rematch}((u, v), t)$ will rematch u using exposed vertex z . The second is when all neighbors of u are already matched. Since all neighbors of u are matched, vertex u cannot be matched to one of its neighbors and so we announce u as an exposed vertex and release its sketch S_u . Therefore, assuming M before time t is maximal, M after time t would be maximal as well.

We next discuss the running time of Procedure $\text{Rematch}((u, v), t)$ when u is a low-degree vertex. By properties of the sketch data structures, the time to query x sampled edges from sketch S_u and construct set \mathcal{N}'_u is $O(x \log^2 n \log(n/\delta))$. If the second case happens, since we assume at every time of stream S , $|M| \leq k$, we then have $d_u = |\mathcal{N}'_u| \leq 2k$.

Recall that \mathcal{T} is a data structure with at most k^2 edges whose space is $O(k^2)$. The insertion, deletion and search times of \mathcal{T} are all worst-case $O(\log k)$. In the second case, the main cost is to remove incident edges on u from \mathcal{T} . For every neighbor $z \in \mathcal{N}'_u$ we search, in time $O(\log k)$, if edge (u, z) has been inserted into \mathcal{T} ; so overall the deletion of incident edges on u from \mathcal{T} is done in time $O(k \log k) = O(x \log k)$ as $|\mathcal{N}'_u| \leq 2k$. So, the running time of Procedure $\text{Rematch}((u, v), t)$ when u is a low-degree vertex is $O(x \log^2 n \log(n/\delta)) = O(k \log^4(n/\delta))$, as we set $x = O(k \log(n/\delta))$.

5.6.2 Analyzing Rematching of a High-Degree Vertex.

LEMMA 5.7. *Let $x = 8ck \log(n/\delta)$ and $y = 8c \log(n/\delta)$. Let u be a high degree vertex, i.e., $d_u > x$. Suppose we query edges $(u, z_1), \dots, (u, z_i), \dots, (u, z_y)$ from sketch S_u . The probability that there exists an exposed vertex $z \in \{z_1, \dots, z_y\}$ is at least $1 - \delta/n^c$. Further, the running time of Procedure **Rematch** $((u, v), t)$ when u is a high-degree vertex is $O(\log^4(n/\delta))$.*

Proof. From Definition 2.1, a ℓ_0 -sampler returns an element $i \in [n]$ with probability $\Pr[i] = \frac{|x_i|}{\ell_0(x)}$ and returns FAIL with probability at most δ . Using Lemma 2.1, there exists a linear sketch-based algorithm for ℓ_0 -sampling using $O(\log^2 n \log \delta^{-1})$ bits of space.

Sketch S_u is a x -sample recovery sketch which means we can recover $\min(x, d_u)$ items (here, edges) that are inserted into sketch S_u . We can think of S_u as x instances of a ℓ_0 -sampler. Note that in this way the space to implement S_u would be x times the space to implement a ℓ_0 -sampler which is $O(x \log^2 n \log \delta^{-1})$ bits of space. Each one of these x ℓ_0 -samplers returns FAIL with probability at most δ . Using a union bound the probability that S_u returns FAIL is $x\delta$. We rescale the failure probability δ to $\frac{\delta}{2xn^c}$ for a constant c . Therefore, the probability that sketch S_u returns FAIL is $\frac{\delta}{2n^c}$, and hence the overall space of S_u is $O(x \log^2 n \log(xn^c/\delta)) = O(cx \log^2 n (\log(n/\delta) + \log \log(n/\delta))) = O(cx \log^2 n \log(n/\delta))$ as $x = 8ck \log(n/\delta)$ and $k \leq n$.

Let $(u, z_1), \dots, (u, z_i), \dots, (u, z_y)$ be the edges queried from sketch S_u for $y = 8c \log(n/\delta)$. Note that the time to query y edges from sketch S_u is $O(y \log^2 n \log(n/\delta)) = O(\log^4(n/\delta))$. Let us define event NoFAIL if S_u does not return FAIL. Let us condition on event NoFAIL which happens with probability $\Pr[\text{NoFAIL}] \geq 1 - \frac{\delta}{2n^c}$.

Fix a returned edge (u, z_i) . Recall that \mathcal{N}_u is the neighborhood of u that is, $\mathcal{N}_u = \{v \in V : (u, v) \in E_t\}$. The number of matched vertices is at most $2k$, i.e., $|V_M| \leq 2k$. Thus, $|\mathcal{N}_u \cap V_M| \leq 2k$ and $|\mathcal{N}_u \setminus \mathcal{N}'_u| = |\mathcal{N}_u| - |\mathcal{N}'_u| \leq 2k$. The probability that (u, z_i) is a fixed edge (u, z) is $\Pr[(u, z_i) = (u, z)] = \Pr[z_i = z] = \frac{1}{|\mathcal{N}'_u|} \leq \frac{1}{|\mathcal{N}_u| - 2k} = \frac{1}{d_u - 2k}$. Using a union bound and since $d_u > x = 8ck \log(n/\delta)$ we obtain

$$\begin{aligned} \Pr[z_i \in V_M] &\leq \sum_{y \in \mathcal{N}'_u \cap V_M} \Pr[z_i = y] \leq \sum_{y \in \mathcal{N}'_u \cap V_M} \frac{1}{d_u - 2k} \\ &\leq \frac{2k}{d_u - 2k} \leq \frac{1}{2c \log(n/\delta)} \leq \frac{1}{2c}. \end{aligned}$$

Therefore the probability that z_i is an exposed vertex, i.e., $z_i \notin V_M$ is $\Pr[z_i \notin V_M] \geq 1 - \frac{1}{2c}$.

We define an indicator variable I_i for queried edge (u, z_i) for $i \in [y]$ which is one if $z_i \notin V_M$ and zero otherwise. Note that $\Pr[I_i = 1] \geq 1 - \frac{1}{2c}$. Let $I = \sum_{i=1}^y I_i$. Then, since y ℓ_0 -samplers of S_u use independent hash functions we obtain

$$\begin{aligned} \Pr[I = 0] &= \Pr[z_1 \in V_M \wedge \dots \wedge z_i \in V_M \wedge \dots \wedge z_y \in V_M] \\ &= \prod_{i=1}^y \Pr[z_i \in V_M] \leq \left(\frac{1}{2c}\right)^y = \left(\frac{1}{2c}\right)^{8c \log(n/\delta)} \leq \frac{\delta}{2n^c}. \end{aligned}$$

Therefore, the probability that there exists an exposed vertex $z \in \{z_1, \dots, z_y\}$ is $1 - \frac{\delta}{2n^c}$. Overall, the probability that sketch S_u does not return FAIL and there exists an exposed vertex $z \in \{z_1, \dots, z_y\}$ is

$$\Pr[\text{NoFAIL} \wedge \{z_1, \dots, z_y\} \setminus V_M \neq \emptyset] \geq 1 - \delta/n^c.$$

LEMMA 5.8. *Suppose that we invoke **Rematch** $((u, v), t)$, and before time t , Invariants 1, 2 and 3 hold, and matching M is maximal. Then after time t , Invariants 1, 2 and 3 hold and matching M is maximal. The running time of **Rematch** $((u, v), t)$ is $O(k \log^4(n/\delta))$.*

Proof. First of all, we invoke **AddEdgeToMatching** $((u, v), t)$ to add edge (u, v) to matching M . In Procedure **AddEdgeToMatching** $((u, v), t')$, we insert the edge to M as well as \mathcal{T} for some $t' \leq t$. We also insert (u, v) to the sketch of whichever vertex (u or v) was exposed before time t' . So at the end of **AddEdgeToMatching** $((u, v), t')$ edge (u, v) is in S_u, S_v and \mathcal{T} .

Once we invoke, Procedure **DeleteFromDS** $((u, v))$, it deletes edge (u, v) from S_u, S_v and \mathcal{T} . We also delete the edge from M . So after invocation of **DeleteFromDS** $((u, v))$, Invariants 1, 2 and 3 hold. Let us fix vertex u . The following proof is the same for vertex v . We consider two cases for u .

(i) First, u is a low-degree vertex, i.e., $d_u \leq x$ assuming Invariants 1, 2 and 3 hold. Observe that using Lemma 5.6, after the invocation of Procedure **Rematch** $((u, v), t)$, matching M is maximal. Moreover, the running time of **Rematch** $((u, v), t)$ when u is a low-degree vertex is $O(x \log^2 n \log(n/\delta)) = O(x \log^3(n/\delta))$. Let \mathcal{N}'_u be the set of neighbors of u that up to time t are inserted into sketch S_u but not deleted from S_u . By Definition 2.2, by querying S_u and with probability at least $1 - \frac{\delta}{2n^c}$, we can recover \mathcal{N}'_u . Observe that assuming Invariants 1, 2 and 3 hold, we must have $(\mathcal{N}_u \setminus \mathcal{N}'_u) \subseteq V_M$. That is, those neighbors of u that are not in \mathcal{N}'_u at time t must be matched. Therefore, all exposed neighbors of u must be in \mathcal{N}'_u . We have two sub-cases. First, if there is an exposed $z \in \mathcal{N}'_u$ then we invoke **AddEdgeToMatching** $((u, z), t)$. Lemma 5.1 shows that Invariants 1, 2 and 3 hold after

invocation of `AddEdgeToMatching` $((w, z), t)$. The second subcase is if there is no exposed node in \mathcal{N}'_w , we then invoke `DeleteNeighborhood`(vertex w). Lemma 5.5 shows that Invariants 1, 2 and 3 hold after invocation of `DeleteNeighborhood`(vertex w).

(ii) Second, u is a high-degree vertex assuming Invariants 1, 2 and 3 hold. Observe that using Lemma 5.7, after the invocation of Procedure `Rematch` $((u, v), t)$, matching M with probability at least $1 - \delta/n^c$ is maximal and the running time of Procedure `Rematch` $((u, v), t)$ when u is a high-degree vertex is $O(\log^4(n/\delta))$. Since with probability at least $1 - \delta/n^c$ there exists an exposed vertex $z \in \{z_1, \dots, z_y\}$, with this probability we invoke `AddEdgeToMatching` $((w, z), t)$. Lemma 5.1 then shows that Invariants 1, 2 and 3 hold after invocation of `AddEdgeToMatching` $((w, z), t)$.

5.7 Completing the Proof of Theorem 1.3 First we prove the claim for the space complexity of our algorithm. We maintain at most $2k$ sketches (for matched vertices), each one is an x -sample recovery sketch for $x = 8ck \cdot \log(n/\delta)$. From Definition 2.2 and the proof of Lemma 5.8, the space to maintain an x -sample recovery sketch is $O(k \log^4(n/\delta))$. So, we need $O(k^2 \log^4(n/\delta))$ bits of space to maintain the sketches of matched vertices. The size of data structure \mathcal{T} , i.e., the number of edges stored in \mathcal{T} is $|\mathcal{T}| \leq (2k)^2$. Thus, overall the space complexity of our algorithm is $O(k^2 \log^4(n/\delta))$ bits.

Next we prove the update time and query time of our dynamic algorithm for maximal matching is $\tilde{O}(k)$. In fact, the deletion or the insertion time of an edge (u, v) is dominated by the running time of most expensive procedures which are `AnnounceNeighborhood` (u) , `DeleteNeighborhood` (u) , and `Rematch` $((u, v), t)$. The running times of these procedures are dominated by the time to query at most x edges from sketches S_u and S_v plus the time to search for x edges in data structure \mathcal{T} . The time to query at most x edges from sketches S_u and S_v using Lemma 5.8 is $O(k \log^4(n/\delta))$. The time to search for x edges in data structure \mathcal{T} is $O(x \log k) = O(k \log^2(n/\delta))$ as we assume the insertion, deletion and query times of \mathcal{T} are all worst-case $O(\log k)$. Therefore, the update time and query time of our dynamic algorithm for maximal matching is $O(k \log^4(n/\delta))$.

Finally, we give the correctness proof of Theorem 1.3. Observe that since after every time t of stream S , Invariants 1, 2 and 3 hold, the matching M is maximal. In fact, since Invariant 1 holds, for every edge $(u, v) \in E_t$ we have at least one of $v \in \mathcal{N}'_u$ or $u \in \mathcal{N}'_v$ which means M is maximal. Recall that V_M is the set of vertices of matched edges in M . Note that for every matched vertex u , we maintain an x -sample recovery sketch S_u .

Next, similar to the algorithm of Theorem 3.1 (Section 3) we construct a graph $(G_M = (V_M, E_M), k)$. For every matched vertex v , we extract up to k edges incident on v from sketch S_v and store them in set E_M . At the end, we run the kernelization algorithm of Section 3.1 on instance $(G_M = (V_M, E_M), k)$. The rest of proof of correctness of Theorem 1.3 requires showing that maintaining a maximal matching is sufficient to obtain a kernel for vertex cover, which is what was exactly argued in proof of Theorem 3.1.

6 Dynamic Parameterized Streaming Algorithm (DPSA for $VC(k)$)

In this section we prove Theorem 1.4 restated below:

THEOREM 6.1. *Let S be a dynamic parameterized stream of insertions and deletions of edges of an underlying graph G . There exists a randomized $(nk, nk + 2^{O(k)})$ -DPSA for $VC(k)$ problem.*

Proof. Let S be a stream of insertions and deletions of edges to an underlying graph $G(V, E)$. We maintain a kn -sample recovery algorithm (Definition 2.2), which processes all the edges seen in the stream; we also keep a counter to record the degree of the vertex. At the end of the stream S , we recover a graph G' by extracting the at most kn edges from the recovery algorithm data structure, or outputting “NO” if there are more than kn edges currently in the graph. We then run the kernelization algorithm of Section 3.1 on instance (G', k) .

Observe that if a graph has a vertex cover of size at most k , then there can be at most nk edges. Each node in the cover has degree at most n , and every node must either be in the cover, or be adjacent to a node in the cover. Therefore, if the graph has more than nk edges, it cannot have a vertex cover of size k . We take advantage of this fact to bound the overall cost of the algorithm in the dynamic case. We maintain a structure which allows us to recover at most nk edges from the input graph, along with a counter for the current number of “live” edges. This can be implemented using a k -sample recovery algorithm (Definition 2.2), or indeed by a deterministic algorithm (e.g. Reed-Solomon syndromes).

The algorithm now proceeds as follows. To test for a vertex cover of size k , we first test whether the number of edges is above nk : if so, there can be no such cover, and we can immediately reject. Otherwise, we can recover the full graph, get the kernel and then run the algorithm of Chen et al. [10] (see Section 3.1). The total time for this algorithm is then $O(nk + 2^{O(k)})$, and the space used is that to store the k -sample recovery algorithm, which is $\tilde{O}(nk)$.

This assumes that each edge is inserted at most once, i.e. the same edge is not inserted multiple times without intervening deletion. This assumption can be removed, if we replace the edge counter with a data structure which counts the (approximate) number of distinct edges currently in the data structure. This can provide a constant factor approximation with polylogarithmic space. This is sufficient to determine if the number of edges is greater than nk , and if not, to recover the at most (say) $1.01nk$ edges in the graph from the data structure storing the edges, and apply the kernelization algorithm of Section 3.1.

7 Feedback Vertex Set

In the Feedback Vertex Set ($FVS(k)$) problem we are given a graph $G = (V, E)$ and an integer k . The question is whether there exists a set $V' \subseteq V$ such that $G \setminus V'$ has no cycles. We can show the following results for $FVS(k)$.

THEOREM 7.1. *There is a deterministic PSA for $FVS(k)$ which uses $O(nk)$ space.*

THEOREM 7.2. *Any (randomized) PSA for $FVS(k)$ requires $\Omega(n)$ space.*

7.1 Parameterized Streaming Algorithm (PSA) for $FVS(k)$

Proof. To prove Theorem 7.1, we use the following lemma to bound the number of edges of a graph with small feedback vertex set.

LEMMA 7.1. *Any graph with a feedback vertex set of size at most k can have at most $n(k+1)$ edges, where n is the number of vertices of the graph.*

Proof. Let the graph be $V = (G, E)$ and $S \subseteq V$ be the feedback vertex set of size at most k . Then the graph $G \setminus S$ is a forest, and hence has at most $n - |S| - 1$ edges. Now each of the vertices in S is adjacent to at most $n - 1$ vertices in G . Hence the total number of edges of G is at most $(n - |S| - 1) + (n - 1)|S| = n + (n - 2)|S| - 1 \leq n + nk$ since $|S| \leq k$.

The PSA algorithm for $FVS(k)$ runs as follows:

- Store all the edges that appear in the stream.
- If the number of edges exceeds $n(k+1)$, output NO.
- Otherwise the total number of edges (and hence the space complexity) is $n + nk$. Now that we have stored the entire graph, use any one of the various known FPT algorithms [8, 23] to solve the $FVS(k)$ problem.

This concludes the proof of Theorem 7.1.

7.2 $\Omega(n)$ Lower Bound for $FVS(k)$

Here, we prove Theorem 7.2.

Proof. We show the proof by reduction to the Disjointness problem in communication complexity.

DISJOINTNESS

Input: Alice has a string $x \in \{0, 1\}^n$ given by $x_1 x_2 \dots x_n$.

Bob has a string $y \in \{0, 1\}^n$.

Question: Bob wants to check if $\exists i : x_i = y_i = 1$.

There is a lower bound of $\Omega(n)$ bits of communication between Alice and Bob, even allowing randomization [24].

Given an instance of Disjointness, we create a graph on $8n$ nodes as follows. We create nodes a_i, b_i, \dots, h_i , and insert edges $(b_i, g_i), (c_i, e_i), (d_i, f_i)$ for all i . We also create edges (h_i, a_{i+1}) for $1 \leq i < n$. This is illustrated in the first graph in Figure 1.

For each i , we add 2 edges corresponding to x_i , and two according to y_i . If $x_i = 0$, we add (a_i, c_i) and (b_i, d_i) ; else we add (a_i, b_i) and (c_i, d_i) . If $y_i = 0$, we add (f_i, h_i) and (e_i, g_i) ; else we add (f_i, e_i) and (g_i, h_i) .

We now observe that the resulting graph is a tree (in fact it is a path) if the two strings are disjoint, but it has at least one cycle if there is any i such that $x_i = y_i = 1$. This can be seen by inspecting Figure 1, which shows the configuration for each possibility for x_i and y_i . Thus, any streaming algorithm that can determine whether a graph stream is cycle-free or has one (or more) cycles implies a communication protocol for DISJOINTNESS, and hence requires $\Omega(n)$ space.

Since $FVS(k)$ must, in the extreme case $k = 0$, determine whether G is acyclic, then $\Omega(n)$ space is required for this problem also. This generalizes to any constant k by simply adding k triangles on $3k$ new nodes to the graph: one node from each must be removed, leaving the question whether the original graph is acyclic.

8 Concluding Remarks

By combining techniques of kernelization with randomized sketch structures, we have initiated the study of *parameterized streaming algorithms*. We considered the widely-studied Vertex Cover problem, and obtained results in three models: insertion only streams, dynamic streams and promised dynamic streams. There are several natural directions for further study. We mention some of the below.

Dynamic Algorithms. Recent work has uncovered connections between streaming algorithms and dynamic algorithms [22]. It is natural to ask whether we can make the algorithms provided dynamic: that is, ensure

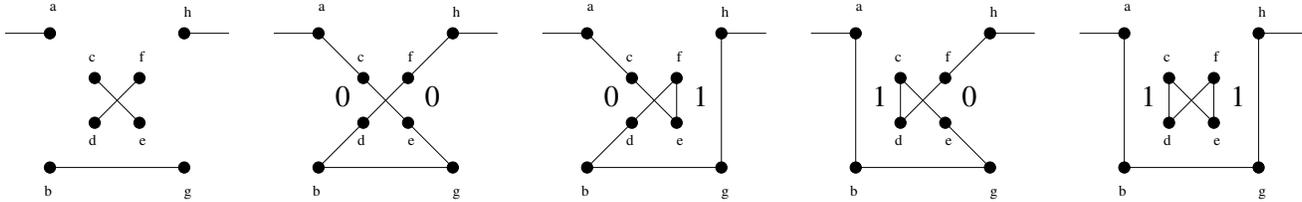


Figure 1: Gadget for reduction to Disjointness

that after each step they provide a current answer to the desired problem. The current algorithm for maximal matching sometimes takes time polynomial in k to process an update: can this be made sublinear in k ?

Our main algorithm in Section 5 applies in the case where there is a promise on the size of the maximal matching. Can this requirement be relaxed? That is, the main open question is whether there exists a dynamic algorithm that will succeed in finding a maximal matching of size k at time t , even if some intermediate maximal matching has exceeded this bound? Or can the cost be made proportional to the largest maximal matching encountered, i.e. remove the requirement for k to be specified at the time, and allow the algorithm to adapt to the input instance.

Other Problems. In this paper, we primarily studied the related problems of Maximal Matching and Vertex Cover. It follows to consider other NP-hard problems in the framework of parameterized streaming, where kernelization algorithms can also be helpful. In some cases, one might be able to obtain parameterized streaming algorithms with small modifications of the existing kernelization methods. This is the case for the Feedback Vertex Set ($FVS(k)$) problem for which we obtain parameterized streaming as discussed in Section 7.

Acknowledgments: The third author would like to thank Marek Cygan for fruitful discussion on early stages of this project in a Dagstuhl workshop. We thank Catalin Stefan Tiseanu for some useful discussions regarding the Feedback Vertex Set problem.

References

- [1] List of open problems in sublinear algorithms. <http://sublinear.info/>.
- [2] K. J. Ahn, S. Guha, and A. McGregor. Analyzing graph structure via linear measurements. In *Proceedings of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 459–467, 2012.
- [3] Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Graph sketches: sparsification, spanners, and subgraphs. In *ACM Principles of Database Systems*, 2012.
- [4] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.
- [5] Neta Barkay, Ely Porat, and Bar Shalem. Efficient sampling of non-strict turnstile data streams. In *Fundamentals of Computation Theory*, pages 48–59, 2013.
- [6] S. Baswana, M. Gupta, and S. Sen. Fully dynamic maximal matching in $O(\log n)$ update time. In *Proceedings of the 52nd IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 383–392, 2011.
- [7] J. F. Buss and J. Goldsmith. Nondeterminism within P. *SIAM Journal on Computing*, 22(3):560–572, 1993.
- [8] Yixin Cao, Jianer Chen, and Yang Liu. On feedback vertex set new measure and new structures. In *SWAT*, pages 93–104, 2010.
- [9] L. Carter and M. N. Wegman. Universal classes of hash functions (extended abstract). In *Proceedings of the 9th Annual ACM Symposium on Theory of Computing (STOC)*, pages 106–112, 1977.
- [10] Jianer Chen, Iyad A. Kanj, and Ge Xia. Improved upper bounds for vertex cover. *Theor. Comput. Sci.*, 411(40-42):3736–3756, 2010.
- [11] Holger Dell and Dieter van Melkebeek. Satisfiability allows no nontrivial sparsification unless the polynomial-time hierarchy collapses. In *STOC*, pages 251–260, 2010.
- [12] R. G. Downey and M. R. Fellows. *Parameterized Complexity*. Springer, New York, 1999.
- [13] Sebastian Eggert, Lasse Kliemann, Peter Munstermann, and Anand Srivastav. Bipartite matching in the semi-streaming model. *Algorithmica*, 63(1-2):490–508, 2012.
- [14] Sebastian Eggert, Lasse Kliemann, and Anand Srivastav. Bipartite graph matchings in the semi-streaming model. In *ESA*, pages 492–503, 2009.
- [15] J. Feigenbaum, S. Kannan, McGregor, S. Suri, and J. Zhang. On graph problems in a semi-streaming model. *Theoretical Computer Science*, 348(2-3):207–216, 2005.
- [16] J. Feigenbaum, S. Kannan, M. Strauss, and M. Viswanathan. An approximate l_1 -difference algorithm for massive data streams. *SIAM Journal on Computing*, 32(1):131–151, 2002.
- [17] J. Flum and M. Grohe. *Parameterized Complexity Theory*. Springer, 2006.
- [18] G. Frahling, P. Indyk, and C. Sohler. Sampling in dynamic data streams and applications. In *Proceedings of the 21st Annual Symposium on Computational Ge-*

- ometry (*SoCG*), pages 142–149, 2005.
- [19] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Macmillan Higher Education, 1979.
 - [20] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM*, 53(3):307–323, 2006.
 - [21] H. Jowhari, M. Saglam, and G Tardos. Tight bounds for L_p samplers, finding duplicates in streams, and related problems. In *Proceedings of the 17th ACM SIGMOD Symposium on Principles of Database Systems (PODS)*, pages 49–58, 2011.
 - [22] B. Kapron, V. King, and Mountjoy. Dynamic graph connectivity in polylogarithmic worst case time. In *Proceedings of the 24th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1131–1142, 2013.
 - [23] Tomasz Kociumaka and Marcin Pilipczuk. Faster deterministic feedback vertex set. *Inf. Process. Lett.*, 114(10):556–560, 2014.
 - [24] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.
 - [25] Andrew McGregor. Finding graph matchings in data streams. In *APPROX-RANDOM*, pages 170–181, 2005.
 - [26] Andrew McGregor. Graph mining on streams. In *Encyclopedia of Database Systems*, pages 1271–1275. Springer, 2009.
 - [27] M. Monemizadeh and D. Woodruff. 1-Pass Relative-Error L_p -Sampling with Applications. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1143–1160, 2010.
 - [28] O. Neiman and S. Solomon. Simple deterministic algorithms for fully dynamic maximal matching. *Proceedings of the 45th Annual ACM Symposium on Theory of Computing (STOC)*, 2013.
 - [29] K. Onak and R. Rubinfeld. Maintaining a large matching and a small vertex cover. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 457–464, 2010.
 - [30] Krzysztof Onak, Dana Ron, Michal Rosen, and Ronitt Rubinfeld. A near-optimal sublinear-time algorithm for approximating the minimum vertex cover size. In *SODA*, pages 1123–1131, 2012.
 - [31] Michal Parnas and Dana Ron. Approximating the minimum vertex cover in sublinear time and a connection to distributed algorithms. *Theor. Comput. Sci.*, 381(1-3):183–196, 2007.
 - [32] A. Das Sarma, S. Gollapudi, and Rina Panigrahy. Estimating pagerank on graph streams. In *Proceedings of the 14th ACM SIGMOD Symposium on Principles of Database Systems (PODS)*, pages 69–78, 2008.