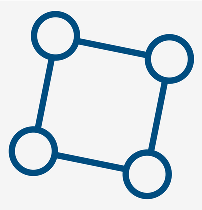
Near-Optimal Four-Cycle Counting in Graph Streams

SODA'26

Sebastian Lüderssen[1] Stefan Neumann[1] Pan Peng[2]

[1] TU Wien [2] USTC

Four-Cycle Counting



Let T be the number of subgraphs H of G with $H \cong C_4$.

Goal: Compute a $(1 \pm \varepsilon)$-approximate estimate $\hat{T}$ of T with prob. $1 - \delta$.

Graph Streaming Model

Graph G is given as a stream of edges in arbitrary order.

We get $O(1)$ passes over the stream.

Goal: Minimize space usage.

Our Result

Theorem: Four-Cycle Counting in arbitrary order edge streams can be solved in $\tilde{O}\left(m/\sqrt{T}\right)$ space using 3 passes.

$\Rightarrow$ Tight space complexity up to $\mathrm{poly}(\log n, \varepsilon^{-1})$ factors.

Previous Works

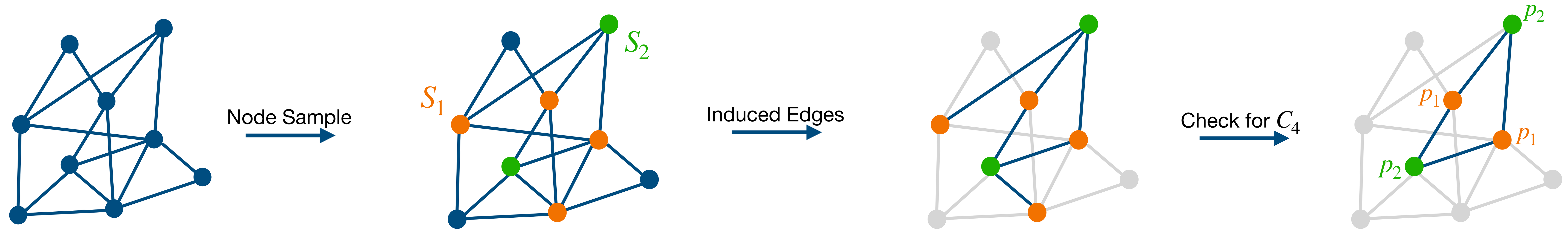
$O(m^2/T)$ in 2 passes [Bera, Chakrabarti. STACS 2017]

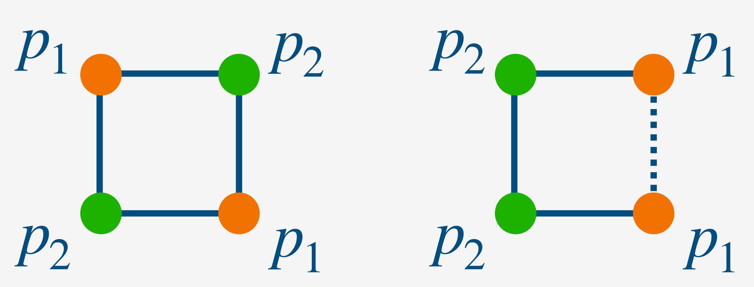
$O(m/T^{1/4})$ in 3 passes [McGregor, Vorotnikova. PODS 2020]

$O(m/T^{1/3})$ in 3 passes [Vorotnikova. arXiv 2020]

$\Omega(m/\sqrt{T})$ for $O(1)$ passes [McGregor, Vorotnikova. PODS 2020]

Our Algorithm — Sampling Induced Subgraphs





We can find each four-cycle in $6 \log T$ *configurations*.

For $\kappa = 1, 2, 4, \ldots, T^{1/4}$:

Sample node sets S_1, S_2 with

$$p_1 = \Theta\left(\frac{\kappa}{T^{1/4}}\right) \text{ and } p_2 = \Theta\left(\frac{1}{\kappa T^{1/4}}\right).$$

Pass 1: Collect induced edges $E[S_1, S_2]$ and $E[S_2, S_2]$.

Pass 2: Collect edges in $E[S_1, S_1]$ that close a C_4.

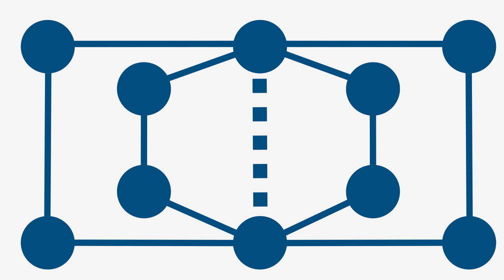
Let A be the number of C_4 sampled in ***LLM*** configurations.

Return $\hat{T} = A/(p_1^2 p_2^2)$.

We expect to find $\Omega(Tp_1^2 p_2^2) = \Omega(1)$ four-cycles. ✓

We expect to collect $O(mp_1p_2) = O(m/\sqrt{T})$ edges. ✓

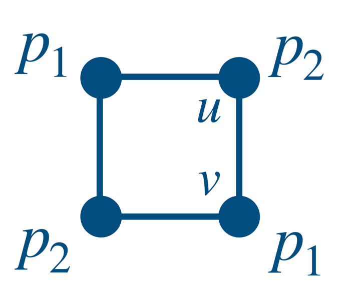
To find *heavy edges.*



Light Configurations

The **heaviness** $t(v)$ of a node v is the number of four-cycles containing v.

A node sampled with probability p is **light** if $t(v) < Tp$.



u is light :	✓	✓	✓	…	✓	✓	✗	…	✗
v is light :	✗	✗	✗	…	✗	✓	✓	…	✓
κ :	2^0	2^1	2^2	…	2^{i-1}	2^i	2^{i+1}	…	$T^{1/4}$

A configuration is **light** if all its nodes, edges and wedges are light.

- Only counting light configurations implies low variance overall.
- We prove that $(1 - \varepsilon)T$ four-cycles admit a light configuration.
- We need to check whether a configuration is light during the algorithm.
- But checking node heaviness is infeasible within space bound.

Refined Heaviness and LLM Configurations

The **refined heaviness** $t(v, p)$ is the number of four-cycles containing v that do not contain heavy edges or wedges w.r.t. p.

This can be checked by a node-sampling oracle using a 3rd pass.

But a node being light is not monotone in p anymore.

A configuration is **light and locally minimal (LLM)**, i.e. light for κ and heavy for all $\kappa' \in [\varepsilon\kappa, \kappa/2]$ with respect to refined heaviness.

		LLM					not LLM					LLM		
light :	✗	✓	…	✓	✗	✗	✓	…	✗	✗	✗	✓	…	✗
κ :	2^0	2^1	…	2^i	2^{i+1}	2^{i+2}	2^{i+3}	…	2^i	2^{i+1}	2^{i+2}	2^{i+3}	…	$T^{1/4}$
				$\varepsilon\kappa$					$\varepsilon\kappa$					

One four-cycle can have multiple LLM configurations.

Counting all LLM configurations double counts at most εT four-cycles.