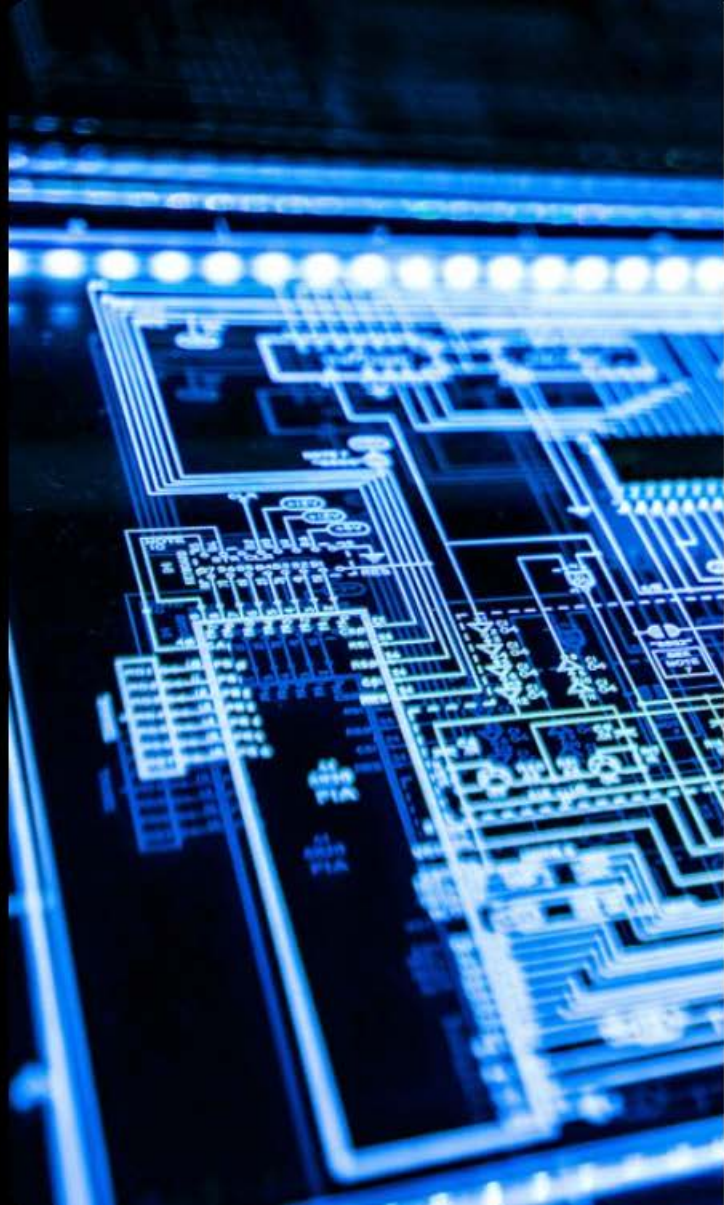


Triangle Counting Optimization in Graph Streams Leveraging Degeneracy and Memory Constraints



1 Importance of Triangle Counting in Network Science and Applications

Triangles in Network Science



Importance of Triangles in Networks

Triangles serve as foundational structures for clustering, anomaly detection, and accurate database join size estimations in networks.



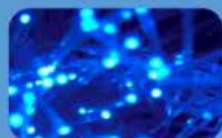
Triangle Counting Challenges

In graph streams, edge arrival is sequential, requiring efficient triangle counting methods due to limited memory resources.



Worst-Case Space Bounds

For arbitrary-order constant-pass streams, the worst-case space requirement is established at $\min\{m^3/2T, m\sqrt{T}\}$, highlighting significant computational challenges.



Real-World Graph Applications

Practical applications of triangle counting span social networks and transportation systems, where understanding connectivity is crucial.

Graph Degeneracy and Estimation

1 Understanding Graph Degeneracy

Graph degeneracy, characterized by the core number, plays a crucial role in optimizing triangle counting by reducing complexity.



2 Input Constraints for Triangle Counting

The input consists of an edge stream across n vertices, aiming for a $1 \pm \epsilon$ approximation of the true triangle count T while adhering to strict memory limits.



Key Findings and Future Directions

1

Leveraging Low Degeneracy

Utilizing low degeneracy allows for circumventing worst-case space limitations, advancing triangle counting efficiency.

2

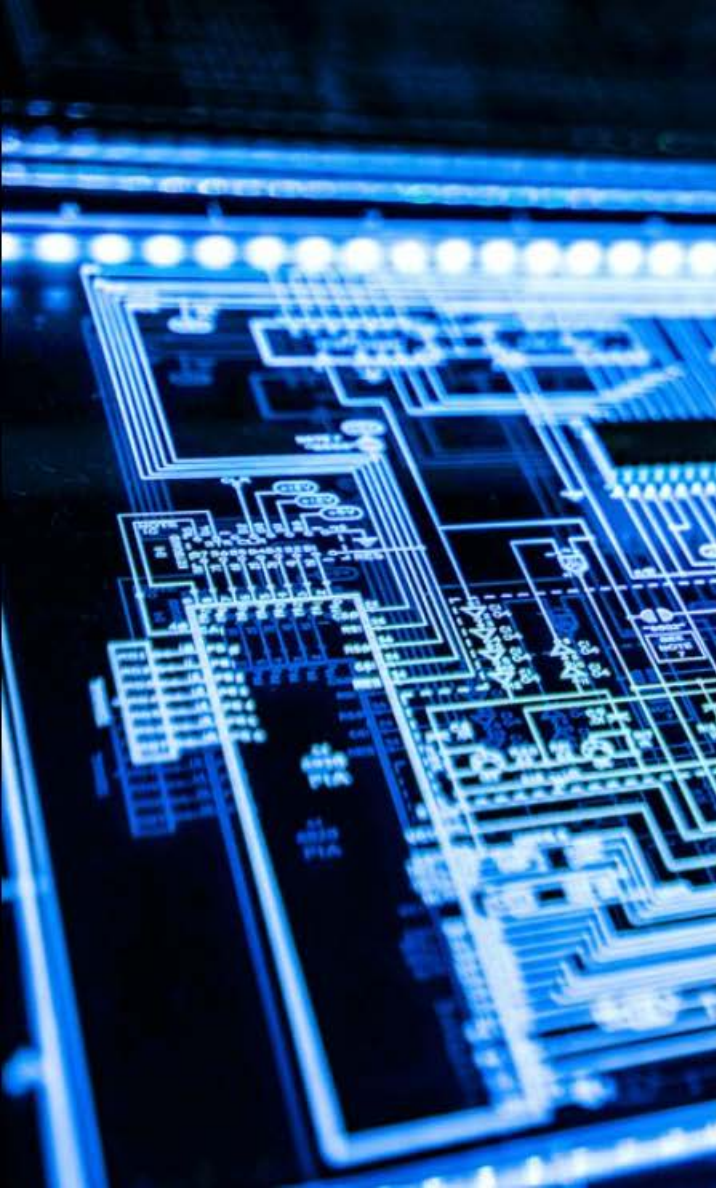
Potential Extension to Other Problems

The methodology shows promise for addressing other subgraph counting challenges under similar bounded degeneracy conditions.

3

Need for Empirical Validation

Future work should focus on empirical evaluations to ascertain practical performance across large-scale graph streams.



2 Challenges in Streaming Algorithms for Triangle Counting

Streaming algorithms face significant limitations due to memory constraints and edge arrival order.

1

Memory Limitations in Algorithms

Efficient triangle counting in graph streams is hindered by memory constraints, necessitating innovative approaches to data processing.

2

Edge Arrival Order Impact

The order of edge arrivals in streams greatly affects the performance and accuracy of triangle counting algorithms in graph processing.

3

Worst-case Scenarios

Algorithms face challenges in worst-case scenarios, emphasizing the need for resilient and adaptable solutions in triangle counting.

4

Efficient Resource Utilization

Maximizing resource utilization in streaming algorithms requires a thorough understanding of memory constraints and edge processing order.

Worst-case scenarios pose challenges, requiring innovative solutions for efficient triangle counting.

1 Triangle Counting Challenges

Accurate counting of triangles in graphs under worst-case conditions necessitates advanced algorithmic strategies to ensure efficiency.



2 Innovation in Algorithms

Developing innovative solutions is critical for overcoming the inherent challenges presented by worst-case scenarios in graph processing.



Graph degeneracy impacts stream processing efficiency, demanding sublinear memory usage strategies.

1 Understanding Graph Degeneracy

Graph degeneracy plays a key role in stream processing efficiency, influencing the design of memory-efficient algorithms.

2 Core Number Significance

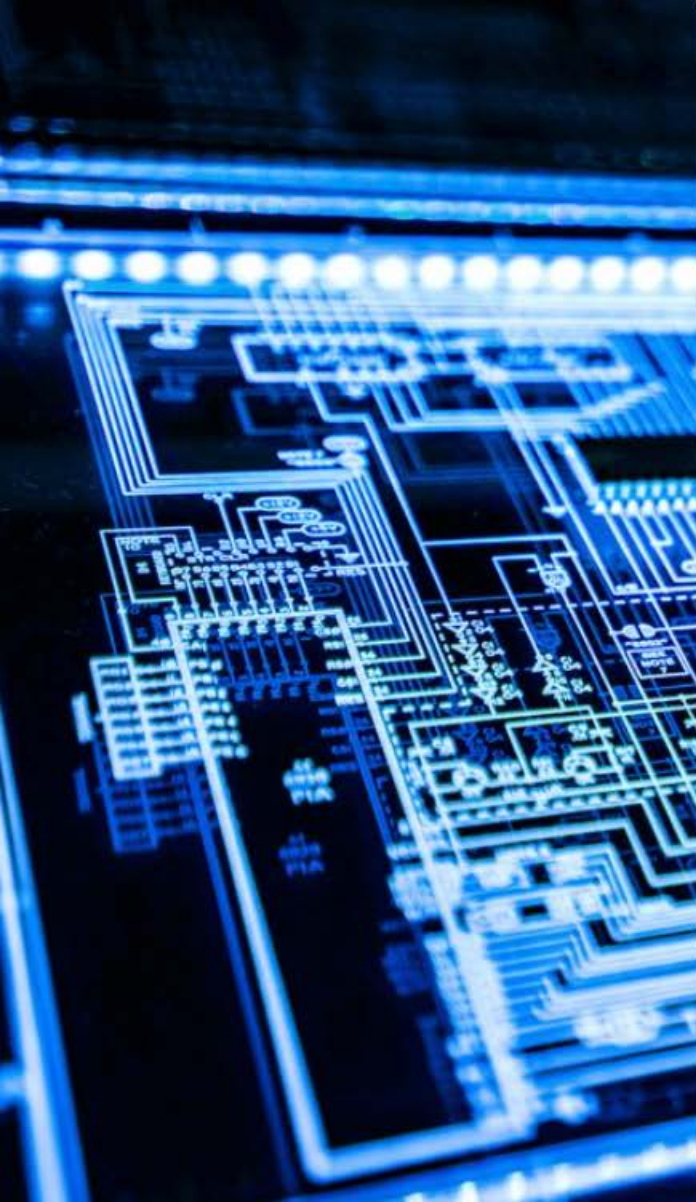
The core number offers insights into graph structure, impacting the effectiveness of triangle counting in streaming contexts.

3 Real-world Graph Applications

Many real-world graphs exhibit properties suitable for triangle counting, highlighting the importance of degeneracy in practical scenarios.

4 Advanced Estimation Techniques

Utilizing advanced estimation techniques under low degeneracy can significantly improve the efficiency of triangle counting algorithms.



3 Algorithm Design and Space Bound Analysis

Context Motivation

◆ Importance of Triangles in Networks

Triangles are crucial in network science for applications like clustering, anomaly detection, and database join size estimation.

◆ Streaming Data Challenges

In streaming settings, edges arrive sequentially, necessitating algorithms that function under stringent memory constraints.

◆ Space Bounds in Triangle Counting

The worst-case space requirements for constant-pass streams with arbitrary edge order are proportional to $\min\{m^3/2T, m\sqrt{T}\}$, indicating significant limitations.



High-Level Algorithm



Three-Pass Estimation Procedure

The proposed algorithm follows a three-pass approach, sampling edges and checking for triangle closure without needing a degree oracle.



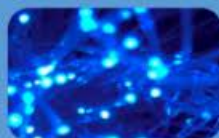
Sampling and Neighbor Selection

During the first pass, edges are sampled based on their degrees, and in the second pass, a neighbor is selected uniformly.



Triangle Closure Check

In the third pass, the algorithm checks if the sampled edge and neighbor form a triangle, assigning values accordingly.



Parallel Execution for Estimation

The final estimator runs multiple independent copies in parallel, utilizing a median-of-means strategy to enhance accuracy.

Key Takeaways

1

Exploiting Low Degeneracy

By leveraging graphs with low degeneracy, the algorithm effectively circumvents the worst-case space lower bounds.

2

Improved Space Complexity

For low degeneracy graphs, the space complexity is reduced to $\tilde{O}(mT)$, outperforming traditional bounds like $m^3/2T$ and $m\sqrt{T}$.

3

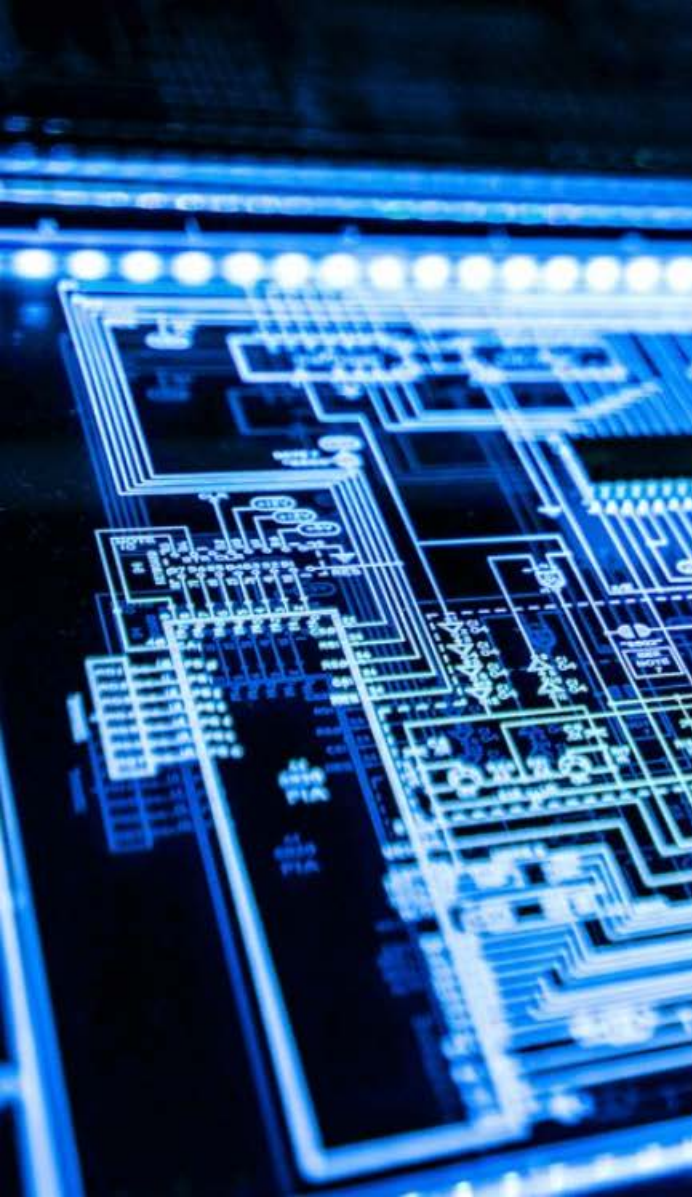
Broader Applications

The methodology may be extended to other subgraph counting problems, demonstrating versatility in bounded degeneracy contexts.

4

Future Research Directions

Future work may include refinements for dynamic scenarios, expanding to larger motifs, and validating performance on substantial graph streams.



4 Future Research Directions and Practical Implementations

Context Motivation

1 Importance of Triangles in Networks

Triangles play a crucial role in network analysis, aiding in functions such as clustering, anomaly detection, and database join size estimation.

2 Challenges in Streaming Settings

In streaming contexts, edges arrive sequentially, necessitating efficient algorithms due to limited memory availability.

3 Worst-case Space Bounds

The worst-case scenario for arbitrary-order constant-pass streams is constrained to $\min(m^3/2T, m\sqrt{T})$ as shown in previous research.

4 Significance of Degeneracy

Understanding graph degeneracy is essential, as it influences algorithm efficiency and memory usage in triangle counting.

Graph Degeneracy Core Number



Definition of Core Number

The core number (k_G) is defined as the maximum minimum degree across all subgraphs of G , indicating the degeneracy of the graph.



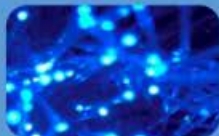
Vertex Ordering

Vertices can be arranged so that each vertex has at most k later neighbors, impacting the triangle counting efficiency.



Characteristics of Real-world Graphs

Many real-world graphs exhibit properties such as being planar or having constant degeneracy, affecting algorithm design.



Importance of Degeneracy in Algorithms

Graphs with low degeneracy allow for improved algorithmic strategies, overcoming worst-case space constraints.

Key Takeaways and Future Directions

1

Leverage Low Degeneracy

Utilizing low degeneracy can help circumvent the worst-case space lower bounds, enhancing algorithm performance.

2

Space Efficiency Achievements

For graphs with degeneracy k , algorithms can achieve space complexity of $\tilde{O}(mT)$, surpassing previous bounds of $m^3/2T$ and $m\sqrt{T}$.

3

Extension to Other Problems

The proposed techniques can be adapted to address other subgraph counting challenges under bounded degeneracy.

4

Future Research Directions

Potential future work includes multipass refinements, exploring dynamic weighted streams, and validating performance on large-scale graphs.

Thank You

