**Paper Reading for Triangle Count**

# P1

New Streaming Algorithms for Counting Triangles in Graphs

**Abstract**

我们提供了三种近似计算triangle count的streaming algorithm。和先前的算法相似，算法的空间复杂度和三角形的数目成正比关系，对于某些类型的图数据，这个空间复杂度会降低。我们也提供了一个基于triangle数量的下界，这意味着我们的第一个算法在常量度的情况下有优化。

**1. Introduction**

在这篇文章中，我们展示了如何在大量的图数据中，流式计算图中三角形的数目。换句话说，设图，是一个有n个顶点，m条边，t个三角形的无向图。我们的着眼点是希望能够找到一个亚线性的空间复杂度的算法来近似计算图中t的数目，而图中的边是作为流来引入的。所谓的亚线性空间利用率，是指我们可以使用的空间，所谓的边流，是指图的边可以是以任何顺序组成的边的序列。为了达到空间利用率，我们严格要求算法只需要

在流上扫描一遍，而且每条边在O(m)的处理时间。

Bar-Yossef et al在文章中说明了所有的列举出triangle的算法，如果要达到置信度为99/100，则至少需要的空间。这说明了针对近似算法有一个的下界，但是Bar-Yossef et al也说明了针对带有大量triangle的图，空间复杂度逼近亚线性是有可能的。换句话说，我们假设表示带有i条边组成的子图的数目，显然有.

**Our Contribution.** 这里作者简单介绍了自己三种算法的复杂度。

**Related Work. 文献2和文献7**给出了在流式模型下的某些图计算的复杂度的下界。现在也有很多人在诸如最短路径、权重最大化等不同图算法上进行研究。

**[文献2]** N. Alon, Y. Matias, M. Szegedy. The space complexity of approximating the frequency moments. STOC 96.

**[文献3]** Bar-Yossef, R. Kumar, S. Sivakumar. Reduction in streaming algorithms with an application of counting triangles in graphs. SODA 2002.

**[文献7]** M. R. Henzinger, P. Raghavan, and S. Rajagopalan, Computing on data streams, Technical Report 1998-001, DEC Systems Research Center .1998.

**2. Algorithms**

首先，我们定义一系列的符号。对于一个无向图，假设n为顶点数目，m为边的数目，d为节点最大的度。是G的有若干个顶点构成的元组的集合。我们将分成四个部分：,其中是包含i条边的元组构成的子图，设。对于,假设是长度为j的图G中的环。

**2.1 One-pass algorithms**

首先我们将展示由文献2提到的两种算法。我们定义一个随机变量X使得E(X) = 。通过利用适量数目的随机变量X的实例，利用切比雪夫不等式，我们就能够有效的减少差异。我们能够在误差至多为的下估计的值。因此，算法的空间使用情况依赖于Var(X)和计算计算随机变量X的空间。

**First Estimator.** 这随机变量X的计算过程如下。随机均匀的从流中选取一条边,统计流中顶点和的公共近邻的数量，假设其数量为c。我们定义。

现在我们来计算变量X的期望，假设我们要求的triangle有一个顺序，针对第i个triangle，我们定义如下指示器：



根据上面的定义，