Abstract

We study the problem of estimating the number of triangles in a graph stream. No streaming algorithm can get sublinear space on all graphs, so methods in this area bound the space in terms of parameters of the input graph such as the maximum number of triangles sharing a single edge. We give a sampling algorithm that is additionally parameterized by the maximum number of triangles sharing a single vertex. Our bound matches the best known turnstile results in all graphs, and gets better performance on simple graphs like $G(n, p)$ or a set of independent triangles. We complement the upper bound with a lower bound showing that no sampling algorithm can do better on those graphs by more than a log factor. In particular, any insertion stream algorithm must use $\sqrt{T}$ space when all the triangles share a common vertex, and any sampling algorithm must take $T^{1/3}$ samples when all the triangles are independent. We add another lower bound, also matching our algorithm's performance, which applies to *all* graph classes. This lower bound covers "triangle-dependent" sampling algorithms, a subclass that includes our algorithm and all previous sampling algorithms for the problem.