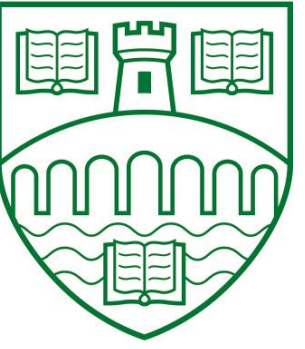


Counting Triangles In Large Graphs

Ioannis Panagiotas

MSc in Big Data

UNIVERSITY of
STIRLING



Problem

Given a graph G , we wish to find the number of triangles, $\tau(G)$, in it.

This problem has some interesting applications in Social Networks for example community or spam detection but it can also be applicable in the field of bioinformatics or CAD design.

Best known single machine exact algorithms are inefficient for large graphs due to memory requirements and are computationally expensive.

Node-Iterator++ specifically needs $O(E)$ space and runs in $O(E \cdot \sqrt{E})$ time.

There are two approaches to deal with the problem:

Streaming Algorithms

Instead of finding the exact number of triangles, we settle for an approximation of this metric.

With minimal space, it is possible to find an approximation with a relative error below 5% in almost linear to E time.

Distributed Algorithms

We can parallelize this problem and take advantage of the opportunities offered by Hadoop.

Using Hadoop, we can find exact answers for graphs with billion of edges efficiently.

Triangular Graphs

