

We approach the problem of computing geometric centralities, such as closeness and harmonic centrality, on very large graphs; traditionally this task requires an all-pairs shortest-path computation in the exact case, or a number of breadth-first traversals for approximated computations, but these techniques yield very weak statistical guarantees on highly disconnected graphs. We rather assume that the graph is accessed in a semi-streaming fashion, that is, that adjacency lists are scanned almost sequentially, and that a very small amount of memory (in the order of a dozen bytes) per node is available in core memory. We leverage the newly discovered algorithms based on HyperLogLog counters, making it possible to approximate a number of geometric centralities at a very high speed and with high accuracy. While the application of similar algorithms for the approximation of closeness was attempted in the MapReduce framework, our exploitation of HyperLogLog counters reduces exponentially the memory footprint, paving the way for in-core processing of networks with a hundred billion nodes using just 2TiB of RAM. Moreover, the computations we describe are inherently parallelizable, and scale linearly with the number of available cores.