Abstract

LFR is a popular benchmark graph generator used to evaluate community detection algorithms. We present EM-LFR, the first external memory algorithm able to generate massive complex networks following the LFR benchmark. Its most expensive component is the generation of random graphs with prescribed degree sequences which can be divided into two steps: the graphs are first materialized deterministically using the Havel-Hakimi algorithm, and then randomized. Our main contributions are EM-HH and EM-ES, two I/O-efficient external memory algorithms for these two steps. In an experimental evaluation we demonstrate their performance: our implementation is able to handle graphs with more than 37 billion edges on a single machine, is competitive with a massive parallel distributed algorithm, and is faster than a state-of-the-art internal memory implementation even on instances fitting in main memory. EM-LFR's implementation is capable of generating large graph instances orders of magnitude faster than the original implementation. We give evidence that both implementations yield graphs with matching properties by applying clustering algorithms to generated instances.