

BDA Assignment-1 Nishant Kumar (2022326)

Summary Table:

Metric	Ground Truth	My Computed	Notes on Difference
Nodes	7,115	7,115	Exact match. Parser drops comments/self-loops/dupes; all vertices kept.
Edges	103,689	103,689	Exact match. Unique directed (<code>src, dst</code>) after de-dup; nothing lost.
Largest WCC (nodes)	7,066 (0.933)	7,066	Exact match. <code>connectedComponents()</code> (treats edges as undirected) finds main weak component correctly.
Largest WCC (edges)	103,663 (1.000)	103,663	Exact match. Edges counted only when both ends are in that WCC.
Largest SCC (nodes)	1,300 (0.183)	1,300	Exact match. <code>stronglyConnectedComponents(maxIter=50)</code> on the directed graph matches GT.
Largest SCC (edges)	39,456 (0.381)	39,456	Exact match. Count restricted to edges internal to the largest SCC.
Avg. clustering coefficient	0.1409	0.1409	Exact match. I count triangles on the

			undirected graph, set nodes with degree < 2 to 0, and took average over all nodes.
Number of triangles	608,389	608,389	Exact match. <code>triangleCount()</code> per-vertex, then <code>sum(t_i)/3</code> avoids triple-counting.
Fraction of closed triangles	0.04564	0.04183	Close match (slightly lower). I divide unique triangles by wedges from undirected degrees; the reference likely uses directed wedges, so mine is a bit smaller.
Diameter	7	7	Match. Double-sweep (farthest-from-farthest) on the largest WCC returns 7.
Effective diameter (90%)	3.8	4.0	Close match. 90th-percentile via sampled multi-source shortest paths + <code>approxQuantile</code> ; minor deviation expected from sampling/quantile approximation.