Graph Similarity Metrics

Jonathan H. Morgan

8 October 2025

**Core Node-Level Features (definitely scalable)**

1. **Degree measures**
   * In-degree, out-degree (unweighted)
   * In-strength, out-strength (weighted)
   * In/out degree ratio
2. **Local structure**
   * Local clustering coefficient (directed)
   * Local reciprocity (fraction of reciprocated edges)
   * Weighted clustering coefficient
3. **Iterative centralities** (good convergence)
   * PageRank
   * Hub score (HITS)
   * Authority score (HITS)
4. **Core decomposition**
   * In-core number
   * Out-core number
   * Combined core number
5. **Local reach**
   * 2-hop in-reach (how many can reach this node in 2 steps)
   * 2-hop out-reach (how many this node can reach in 2 steps)

**Graph-Level Features**

* **Sampled triad census** (e.g., sample 10,000 random triads)
  + Get approximate frequencies of the 16 triad types
  + Focus on key patterns: reciprocal (030T), transitive (030C), cycles (030C)
* **Component structure**
  + SCC size distribution (largest, second largest)
  + Bow-tie fractions (IN, OUT, SCC, etc.)
* **Global measures**
  + Global reciprocity
  + Density (or edge/node ratio for large graphs)
  + Degree assortativity
  + Modularity (with fast community detection)