CSE 5311 – Programming Project

Disjoint sets

Union-Find refers to a class of algorithms used to maintain a data structure for disjoint sets, many of which have been analyzed by Tarjan (For references, see [Tarjan75], [wiki]) These algorithms are often used in connected components algorithms, which motivates the following problem:

Consider a graph that initially contains n vertices and no edges. Randomly include one edge at a time until the graph is connected; that is, until there is only one connected component. <u>Plot</u> the expected number of edges that must be included before the graph is connected against n. Can you find a function *f(n)* such that *f(n)* is close to the expected number of edges that must be included? For a few values of n, use your samples to estimate $g_n(m)$ = **Pr***{after m edges, the n vertex graph is connected}*. Plot your estimates of g_n; do you notice anything surprising? Can you find any threshold beyond which the graph gets connected with high probability? Use graph plots to show your results.

[Tarjan75] Tarjan, Robert Endre (1975). "Efficiency of a Good But Not Linear Set Union Algorithm".
    Journal of the ACM 22 (2): 215–225. doi:10.1145/321879.321884
[wiki] http://en.wikipedia.org/wiki/Disjoint-set_data_structure.