Linear Time Algorithms for Pairwise Statistical Problems

**Single Query Case:** Effect of a set on a query can be computed in O(N) time
- Can be solved efficiently with spatial partitioning
- Trees used – kd-tree (Freidman, et al., '77) for NN, Barnes-Hut tree (Barnes & Hut, '86) for potential summation
- Have non-rigorous O(log N) bound
- Analysis of NN in low intrinsic dimensional (c) data
- Rigorous O(log N) bounds for random walk (Karger & Ruhl, '02), navigating nets (Krauthgamer, et al., '04), cover trees (Beygelzimer, et al., '06)
- Answering O(N) queries is only improved to O(N log N)

**Dual-tree NN analysis with cover tree:**
- Need to bound degree of bichromaticityk:(number of query descends between any two reference descends)
- kquantifies difference in scale between the query and reference set
- Unbounded k => runtime no better than O(N log N)
- Query tree completely traversed down to the leaves
- O(N) single query computations
- => O(N log N) runtime bound
- k=1 in monochromatic case

**Cover tree approximate KDE:**
- Approximation with absolute error ε
- *prune when every point in the subtree can be approximately computed within ε error
- Single query KDE – first ever O(k log N) runtime bound where k is a function of the kernel function, the bandwidth, and the error ε
- Multi (O(N)) query KDE – first ever O(N) runtime bound for the dual-tree KDE algorithm
- Similar results for approx. KDE with relative error

**Future Directions:** Apply cover tree analysis to
- Boruvka’s algorithm for Euclidean Minimum Spanning Tree
- Range Search
- Multi-tree algorithms for n-point correlation problems