

CSE 5800 Mining/Learning and the Internet—HW4

Due 6:30pm, Mon, Nov 9

Submit Server: `course=cse5800 , project=hw4`

1. Implement BridgeCut with four versions:
 - (a) edge with the highest Bridging Centrality ($C_{Br}(e)$ in the paper)
 - (b) vertex with the highest Bridging Centrality ($C_{Br}(v)$)
 - (c) edge with the highest Betweenness ($\Phi(e)$)
 - (d) vertex with the highest Betweenness ($\Phi(v)$)
2. Allow this parameter:
 - (a) density threshold (densityThreshold in the paper)
3. Measure performance using:
 - (a) Davies-Bouldin index
 - (b) Silhouette Coefficient (handout)
4. Use three data sets:
 - (a) toy data set on the course web site
 - (b) real data set on the course web site
 - (c) your own data set
5. Discuss in a report (in pdf):
 - (a) Sensitivity analysis of parameters for the second data set:
 - i. vary each of the parameters (keeping the rest constant),
 - ii. calculate each performance measurement,
 - iii. plot performance vs. value of a parameter,
 - iv. discuss the value for each parameter that seems to achieve the highest performance.
 - (b) Compare the algorithms for the second data set:
 - i. plot performance vs. density threshold for different algorithms
 - ii. discuss the relative performance of different algorithms
6. Implementation:
 - (a) use one of these programming languages: C, C++, Java, or LISP.
 - (b) input file: a file for vertices and edges
 - (c) two modules:
 - i. BridgeCut: input graph; output:
 - top edge/vertex when it is removed
 - for each cluster, output vertices in the cluster
 - ii. Evaluate: input vertices and cluster membership; output performance