DATA MINING
Spatial Clustering

Margaret H. Dunham
Department of Computer Science and Engineering
Southern Methodist University

Companion slides for the text by Dr. M.H.Dunham, *Data Mining, Introductory and Advanced Topics*, Prentice Hall, 2002.

Nearest Neighbor

- Items are iteratively merged into the existing clusters that are closest.
- Incremental
- Threshold, t, used to determine if items are added to existing clusters or a new cluster is created.

Nearest Neighbor Algorithm

Input:

$D = \{t_1, t_2, ..., t_n\}$ // Set of elements

A // Adjacency matrix showing distance between elements.

Output:

K // Set of clusters.

Nearest Neighbor Algorithm:

$K_1 = \{t_1\}$;

$K = \{K_1\}$;

$k = 1$;

for $i = 1$ *to* n **do**

  find the t_m **in some cluster** K_m **in** K **such that** $dis(t_i, t_m)$ **is the smallest;**

  if $dis(t_i, t_m) \leq t$ **then**

    $K_m = K_m \cup t_i$

  else

    $k = k + 1$;

    $K_k = \{t_i\}$;