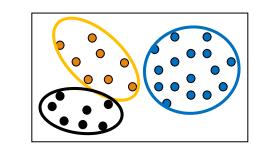
Internal Measures (I): BetaCV Measure

- □ A trade-off in maximizing intra-cluster compactness and inter-cluster separation
- \square Given a clustering $C = \{C_1, \ldots, C_k\}$ with k clusters, cluster C_i containing $n_i = |C_i|$ points
 - Let W(S, R) be sum of weights on all edges with one vertex in S and the other in R

 - The sum of all the intra-cluster weights over all clusters: $W_{in} = \frac{1}{2} \sum_{i=1}^{k} W(C_i, C_i)$ The sum of all the inter-cluster weights: $W_{out} = \frac{1}{2} \sum_{i=1}^{k} W(C_i, \overline{C_i}) = \sum_{i=1}^{k-1} \sum_{i>i} W(C_i, C_i)$
 - The number of distinct intra-cluster edges: $N_{in} = \sum_{i=1}^{k} {n_i \choose 2}$
 - The number of distinct inter-cluster edges: $N_{out} = \sum_{i=1}^{k-1} \sum_{i=1}^{k} n_i n_i$



- Beta-CV measure: $BetaCV = \frac{W_{in} / N_{in}}{W_{out} / N_{out}}$
 - The ratio of the mean intra-cluster distance to the mean inter-cluster distance
 - The smaller, the better the clustering