

CSE 40647/60647 Data Science (Spring 2018)

Lecture 19: Cluster Analysis: Evaluation

Cluster Quality

External

Matching-based

Purity

Maximum matching

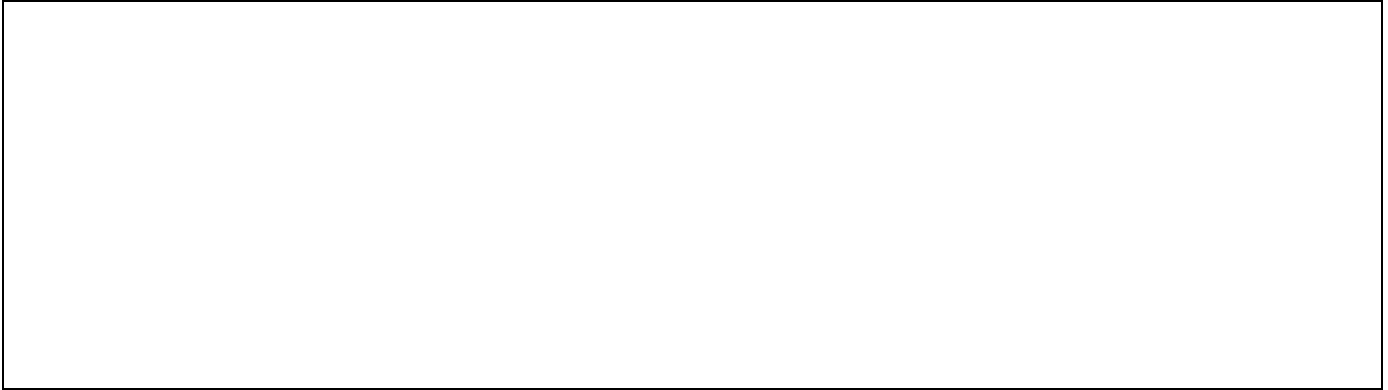
Precision, Recall, F-measure

Pairwise

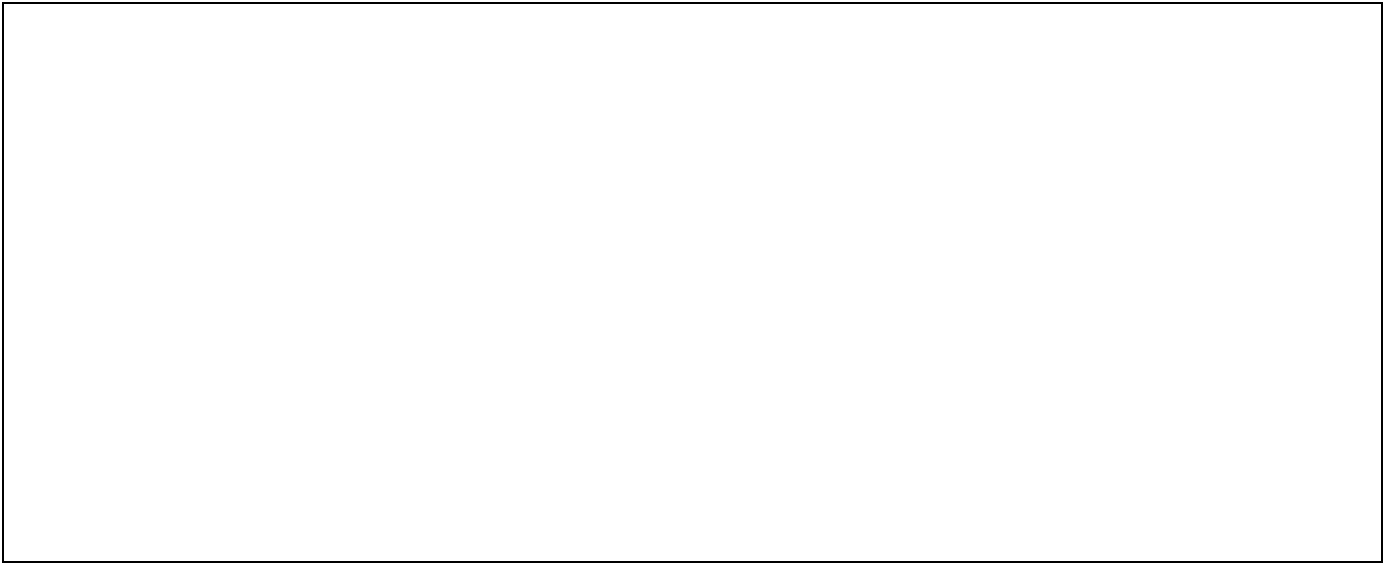
TP, FN, FP, TN

Jaccard coefficient, Rand statistic, FM-measure (in exercise)

Entropy-based
Conditional entropy
NMI



Internal
BetaCV



Normalized cut
Modularity
Silhouette coefficient

Relative
Silhouette coefficient

Cluster Stability

Elbow method

Cluster Tendency

Hypothesis tests: Hopkins Statistic



Name (NetID):

Data: (C: clusters; T: Truth partitions)

$C \backslash T$	T_1	T_2	T_3	Sum
C_1	0	20	30	50
C_2	0	20	5	25
C_3	25	0	0	25
m_j	25	40	35	100

Pair-wise external (clustering evaluation) measures:

1. TP

2. FN

3. FP

4. TN

5. Jaccard coefficient

6. Rand Statistic

7. FM Measure