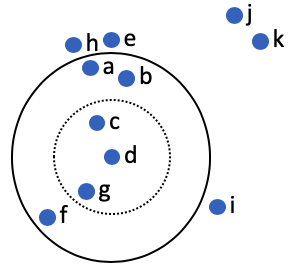
­­­

DSCI 558: Building Knowledge Graphs

Quiz 17 (5 minutes)

**Question 1 (3pts):**

Recall the Canopy Clustering algorithm (McCallum et al, KDD00). The following illustrates the outcome of the 1st iteration:



1. Which of the following represents the **valid group of elements** from which we choose a random element to generate the next canopy (iteration 2):
2. {a, b, c, e, f, g, h, i, j, k}
3. {a, b, e, f, h, i, j, k}
4. {a, b, f}
5. {e, h, i, j, k}
6. Assuming that the chosen element (iteration 2) is **f,** which of the following represents the valid group of elements for iteration 3:
7. {j, k}
8. {a, b, e, h, i, j, k}
9. {h, e, j, k}
10. {h, e, i, j, k}

**Question 2 (4pts):**

Let F be a (0.3, 0.8, 0.7, 0.2)-sensitive family of minHash functions. Let F1 be the family constructed using a “band of r=2 minHashes” (AND construction on F) and F2 be the family constructed using a “band of b=2 minHashes” (OR construction on F). Let F1 is a (d1, d2, d3, d4)-sensitive family and F2 is a (e1, e2, e3, e4)-sensitive family of minHash functions. Calculate (d1, d2, d3, d4) and (e1, e2, e3, e4).

**Question 3: (3 pts)**

Match the following PSL rules to their correct explanations:

!Same(E1, E2). Nothing is the same without evidence

CandSame(E1, E2) => Same(E1, E2): Candidate pairs can refers to the same entity

CandSame(E1, E2) ^ Sim(E1, E2) => Same(E1, E2): Similar candidate pairs can refers to the same entity

Same(E1, E2) => Same(E2, E1): Symmetry rule

CandSame(A, B) ^ CandSame(A,C) ^ Same(A,B) => !Same(A,C): Sparsity rule

CandSame(E1, E2) ^ NewEntity(E1) => Same(E1, E2): Linking to new entities rule