**Problem 1:**

1. T = {1, 2, 3, 4, 6, 7, 8, 9}

**TD1 = <1, 0, 0, 1, 1, 1, 1, 0>**

**TD2 = <0, 1, 1, 1, 0, 1, 0, 1>**

1. Jac(D1, D2) = |SD1 and SD2|/| SD1 or SD2|

|SD1 and SD2| = |{4, 7}| = 2

| SD1 or SD2| = |{1, 2, 3, 4, 6, 7, 8, 9}| = 8

**Jac(D1, D2) = 2/8 = 1/4**

1. CosSim(D1, D2) = TD1 dot TD2 / (L2(TD1) \* L2(TD2))

TD1 dot TD2 = |SD1 and SD2| = 2

L2(TD1) = sqrt(5), L2(TD2) = sqrt(5)

**CosSim(D1, D2) = 2 / 5**

**Problem 2:**

1. C = TD1 dot TD2 / (L2(TD1) \* L2(TD2))

C2 = (TD1 dot TD2)2 / (L2(TD1) \* L2(TD2))2

J = TD1 dot TD2 / (L2(TD1)2 + L2(TD2)2 – (TD1 dot TD2))

Let (TD1 dot TD2) = P, L2(TD1) = A, and L2(TD2) = B

C2= P2 / (A \* B)2

= P2 / A2B2

J = P / (A2 + B2 - P)

Divide both sides by P gives: P / A2B2 <= 1 / (A2 + B2 – P)

Multiply both sides by (A2 + B2 – P) gives: P(A2 + B2 – P) / A2B2 <= 1

P(A2 + B2 – P) / A2B2 = (PA2 / A2B2) + (B2/ A2B2) – (P2/ A2B2)

= (P/A2) + (P/B2) – (P2/ A2B2) <= 1

We know P <= A2 and P <= B

So, (P/A2) + (P/B2) – (P2/ A2B2) <= (A2/A2) + (B2/B2) – (A2B2/ A2B2)

And (A2/A2) + (B2/B2) – (A2B2/ A2B2) = 1 + 1 – 1 = 1

So, (P/A2) + (P/B2) – (P2/ A2B2) <= 1

**Thus, C2 <= J**

1. Let (TD1 dot TD2) = P, L2(TD1) = A, and L2(TD2) = B

J = P / (A2 + B2 - P)

C = P / (A\*B)

C / (2 – C) = (P / (A\*B)) / (2 – (P / (A\*B))) = P / (-P + 2AB)

Divide both sides by P results in (1 / (A2 + B2 - P)) <= (1 / (-P + 2AB))

Multiply both sides by both denominators gives us: (-P + 2AB) < = (A2 + B2 - P)

Add P to both sides gives us : 2AB <= A2 + B2

Since A and B are real numbers, so is A – B, and we know 0 <= (A – B)2, and (A – B)2 = A2 + B2 – 2AB, so 0 <= A2 + B2 – 2AB

Add 2AB to both sides gives us 2AB <= A2 + B2

**Since 2AB <= A2 + B2, we can conclude that J <= (C / (2 – C))**

**Problem 3:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | D1 | D2 | D3 | D4 |
| Perm1 (2x+1)%5 | 0 | 1 | 2 | 0 |
| Perm2 (3x+4)%5 | 0 | 2 | 1 | 0 |
| Perm3 (x+3)%5 | 0 | 3 | 1 | 0 |

**Problem 4:**

Note that for claim 1, Pr[P(i) = j] = 1/(m+1)

Observation 1: For every permutation P, at least one of min[P(Da)] or min[P(Db)] equals 1 or 2. Since P is a permutation, there exists x in {1, … , n+1} such that P(x) = 1 or P(x) = 2. Since Da and Db = {1, … , n+1}, x is an element of Da or Db. Thus, x must belong to at least one of Da or Db.

Observation 2: let x be an element of Da such that min[P(Da)] = P(x), and let y be an element of Db such that min[P(Db)] = P(y). if min[P(Da)] = min[P(Db)], then x = y, and thus x belongs to Da and Db.

If min[P(Da)] = min[P(Db)], then P(x) = P(y). since P is still one-one, it must still be the case that x = y, and thus x appears in both Da and Db.

Thus:

Pr[min[P(Da)] = min[P(Db)]] = Pr[min[P(Da)] = min[P(Db)] = 1 or 2]

= Pr[for x, an element of Da and Db, P(x) = 1 or 2]

= |Da and Db| / |m+1| by claim 1

= |Da and D­b­| / |Da or Db|

= Jac(Da, Db)

**Problem 5:**

Any permutation hash function would preserve the given similarity measure. For example, suppose the vectors are 4 dimensional, and you have two vectors with 1 shared item at the same index (IE. U = {1, 2, 3, 4}, V = {1, 6, 7, 9}). The similarity of those two items is obviously ¼. We know that for a random permutation of U and V, there are 4 possible positions for the matching number, the chance of the number being put into any of those positions is ¼, and the chance of the other vectors number being in the same position is ¼. So, the chance of the numbers matching up is 4 \* ¼ \* ¼ = 1/4.

In general, the chance of of the elements matching position is m \* 1/m \* 1/m \* |{i| ui = vi}|

= |{i| ui = vi}| / m