Tanimoto coefficient

$$S_{AB} = \frac{c}{a + b - c}$$

a: A count of „1“ in a fingerprint of a molecule A
b: A count of „1“ in a fingerprint of a molecule B
c: A count of „1“, which have both fingerprints in the same positions

A	1	0	1	1	1	0	1	1	0	0	1	1
B	0	0	1	1	0	0	1	0	1	0	1	1

$a=8$

$b=6$

$c=5$

$$S_{AB} = \frac{5}{8+6-5} = 0.56$$

Next binary similarity coefficients

Name	Formula for binary (dichotomous) variables
Tanimoto (Jaccard) coefficient)	$S_{AB} = \frac{c}{a+b-c}$ Range: 0 to 1
Dice coefficient (Hodgkin index)	$S_{AB} = \frac{2c}{a+b}$ Range: 0 to 1
Cosine similarity (Carbó index)	$S_{AB} = \frac{c}{\sqrt{ab}}$ Range: 0 to 1
Euclidean distance	$D_{AB} = \sqrt{a+b-2c}$ Range: 0 to N
Hamming (Manhattan or City-block) distance	$D_{AB} = a+b-2c$ Range: 0 to N
Soergel distance	$D_{AB} = \frac{a+b-2c}{a+b-c}$ Range: 0 to 1

Similarity coefficients – binary and real numbers

Name	Formula for binary (dichotomous) variables	Formula for continuous variables
Tanimoto (Jaccard) coefficient)	$S_{AB} = \frac{c}{a+b-c}$ Range: 0 to 1	$S_{AB} = \frac{\sum_{i=1}^{N} x_{iA}x_{iB}}{\sum_{i=1}^{N}(x_{iA})^2+\sum_{i=1}^{N}(x_{iB})^2-\sum_{i=1}^{N} x_{iA}x_{iB}}$ Range: -0.333 to $+1$
Dice coefficient (Hodgkin index)	$S_{AB} = \frac{2c}{a+b}$ Range: 0 to 1	$S_{AB} = \frac{2\sum_{i=1}^{N} x_{iA}x_{iB}}{\sum_{i=1}^{N}(x_{iA})^2+\sum_{i=1}^{N}(x_{iB})^2}$ Range: -1 to $+1$
Cosine similarity (Carbó index)	$S_{AB} = \frac{c}{\sqrt{ab}}$ Range: 0 to 1	$S_{AB} = \frac{\sum_{i=1}^{N} x_{iA}x_{iB}}{\left[\sum_{i=1}^{N}(x_{iA})^2 \sum_{i=1}^{N}(x_{iB})^2\right]^{1/2}}$ Range: -1 to $+1$
Euclidean distance	$D_{AB} = \sqrt{a+b-2c}$ Range: 0 to N	$D_{AB} = \left[\sum_{i=1}^{N}(x_{iA}-x_{iB})^2\right]^{1/2}$ Range: 0 to ∞
Hamming (Manhattan or City-block) distance	$D_{AB} = a+b-2c$ Range: 0 to N	$D_{AB} = \sum_{i=1}^{N}\lvert x_{iA}-x_{iB}\rvert$ Range: 0 to ∞
Soergel distance	$D_{AB} = \frac{a+b-2c}{a+b-c}$ Range: 0 to 1	$D_{AB} = \frac{\sum_{i=1}^{N}\lvert x_{iA}-x_{iB}\rvert}{\sum_{i=1}^{N}\max(x_{iA},x_{iB})}$ Range: 0 to 1