

# Intelligent Data Management - Exercise 3

Christoph Prinz

November 14, 2017

## Assignment 1

Element	$S_1$	$S_2$	$S_3$	$S_4$	$h_1$	$h_2$	$h_3$
0	0	1	0	1	1	2	2
1	0	1	0	0	3	5	1
2	1	0	0	1	5	2	0
3	0	0	1	0	1	5	5
4	0	0	1	1	3	2	4
5	1	0	0	0	5	5	3

Minhash-Signature Step 1:

	$S_1$	$S_2$	$S_3$	$S_4$
$h_1$	$\infty$	$\infty$	$\infty$	$\infty$
$h_2$	$\infty$	$\infty$	$\infty$	$\infty$
$h_3$	$\infty$	$\infty$	$\infty$	$\infty$

Minhash-Signature Step 2:

	$S_1$	$S_2$	$S_3$	$S_4$
$h_1$	$\infty$	1	$\infty$	1
$h_2$	$\infty$	2	$\infty$	2
$h_3$	$\infty$	2	$\infty$	2

Minhash-Signature Step 3:

	$S_1$	$S_2$	$S_3$	$S_4$
$h_1$	$\infty$	1	$\infty$	1
$h_2$	$\infty$	2	$\infty$	2
$h_3$	$\infty$	1	$\infty$	2

Minhash-Signature Step 4:

	$S_1$	$S_2$	$S_3$	$S_4$
$h_1$	5	1	$\infty$	1
$h_2$	2	2	$\infty$	2
$h_3$	0	1	$\infty$	0

Minhash-Signature Step 5:

	$S_1$	$S_2$	$S_3$	$S_4$
$h_1$	5	1	3	1
$h_2$	2	2	2	2
$h_3$	0	1	4	0

Final Minhash-Signature (After Step 6):

	$S_1$	$S_2$	$S_3$	$S_4$
$h_1$	5	1	3	1
$h_2$	2	2	2	2
$h_3$	0	1	4	0

Jaccard Similarities formula =  $|S \cap T| \div |S \cup T|$

$$SIM(S_1, S_2) = SIM(\{2, 5\}, \{0, 1\}) = 0$$

$$SIM(S_1, S_3) = SIM(\{2, 5\}, \{3, 4\}) = 0$$

$$SIM(S_1, S_4) = SIM(\{2, 5\}, \{0, 2, 4\}) = \frac{1}{4}$$

$$SIM(S_2, S_3) = SIM(\{0, 1\}, \{3, 4\}) = 0$$

$$SIM(S_2, S_4) = SIM(\{0, 1\}, \{0, 2, 4\}) = \frac{1}{4}$$

$$SIM(S_3, S_4) = SIM(\{3, 4\}, \{0, 2, 4\}) = \frac{1}{4}$$

## Assignment 2

Plots for s-curve:

