

③ $V_1 = [0, 1, 0, 1, 1, 0, 0]$
 $V_2 = [1, 0, 1, 0, 1, 0, 1]$
 $V_3 = [1, 1, 1, 0, 0, 1, 0]$
 $V_4 = [0, 1, 0, 0, 1, 0, 1]$
 $V_5 = [1, 1, 1, 1, 1, 1, 0]$
 $V_6 = [1, 1, 1, 1, 0, 0, 0]$
 $V_7 = [0, 0, 0, 0, 0, 0, 1]$
 $V_8 = [0, 0, 0, 1, 1, 1, 0]$

- DIST DE HAMMING P/ BITS 0, 1 Y 5 $\rightarrow r=3, b=1 \Rightarrow 3 \text{ MH}$
- P/USH, TABLA DE 6 REGISTROS CON BUCKETS DE TAMAÑO 4.

DIST DE HAMMING = NORMALIZADO $\Rightarrow P = 1 - H(X)/\text{LEN}(X)$

- VAMOS A TENER UNA TABLA CON 3 VALORES FUNCIONES DE HASHING. (1 POR CADA BIT DADO)

	MH1	MH2	MH3
V1	0	1	0
V2	1	0	0
V3	1	1	1
V4	0	1	0
V5	1	1	1
V6	1	1	0
V7	0	0	0
V8	0	0	1

\rightarrow CADA VALOR ES EL BIT 0, 1 Y 5 DE CADA VECTOR.

• V1 VA A LA POS 2, V2 A LA POS 4, V3 A LA 7
 V4 A LA 2, V5 A LA 7, V6 A LA 6, V7 A LA 0
 Y V8 A LA 1

- LA CAMBIO DE LOS 3 BITS (O VALORES DE LOS MH) NOS INDICAN LA POSICIÓN EN LA TABLA DE HASH EN BUENAS. COMO MIS VECTORES DIFERENTES SON DE 7 DIMENSIONES, Y LA TABLA DE 6, DEBERÁN APARECER UNA FUNCIÓN DE HASHING UNIVERSAL A LA POSICIÓN QUE ME INDICAN ESOS 3 BITS PARA REDIRECCIONAR LA POSICIÓN A UNA TABLA DE 6 POSICIONES.

• VAY A PLANTEAR $H(X) = [(ax + b) \% P] \% M \rightarrow M = 6, P: \text{NUMERO PRIMO} \geq M \Rightarrow P = 117$
 $\rightarrow A \in [1, P-1] \Rightarrow A = 1$

$\therefore H(X) = (x + 7) \% 6 \rightarrow$ ESTA FUNCIÓN SE

LA AFILIA AL VALOR DECIMAL QUE ME DAN LOS 3 MH DE UNA VECTOR

V3, V5 V6, V7	V8	V1, V2		V2	
0	1	2	3	4	5

$H(V_1) = 2$
 $H(V_2) = 4$
 $H(V_3) = 0$
 $H(V_4) = 2$
 $H(V_5) = 0$
 $H(V_6) = 0$
 $H(V_7) = 0$
 $H(V_8) = 1$

B) VECTOR MÁS CERCANO A $Q = [1, 0, 1, 0, 0, 0, 1]$

A ESTE VECTOR, LE BUSCO EL VALOR DE SUS 3 MH SEGÚN EL VALOR DE LOS BITS 0, 1 Y 5.

$MH_1(Q) = 1$, $MH_2(Q) = 0$, $MH_3(Q) = 0$. \Rightarrow ESTO ME DA LA POSICIÓN ⁴ DE LA TABLA DE HASH, PERO DEBO APLICAR LA FUNCIÓN DE HASH.

~~$MH_1(Q) = 1$~~ $H(Q) = H(4) = (4 + 7) \cdot 1 \cdot 6 = 4$

SI MIRO LA TABLA, EN LA POSICIÓN 4 ~~ESTÁ~~ ESTÁ EL VECTOR V_2

\therefore VECTOR MÁS CERCANO A Q : $V_2 = [1, 0, 1, 0, 1, 0, 1]$

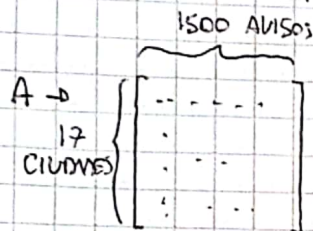
C) PARA REDUCIR LA CANTIDAD DE ESPACIO USADA, PODRÍAN HABER UTILIZADO UNA FUNCIÓN DE HASHING PERFECTA Y MÍNIMA, QUE ASEGURA $O(M)$ EN ESPACIO, LO CUAL ES POSIBLE SÓLO SI CONOCEMOS CUÁLES SON TODAS LAS CLAVES QUE QUEREMOS. SI ESTO SUCEDE, COMO EN NUESTRO CASO, ES POSIBLE DESARROLLAR UNA SOLUCIÓN QUE MAPEE CADA CLAVE A UNA ÚNICA POSICIÓN.

④ TENGO UNA MATRIZ DE 17 FILAS X 1500 COLUMNAS. \rightarrow MATRIZ A

A) SI DESCOMPONGO LA MATRIZ "A" EN SU SVD, OBTENGO:

$$A = U \Sigma V^T$$

\rightarrow EN CADA CELDA ESTÁ LA CANTIDAD DE VECES QUE EL AVISO FUE CLICKEADO



• SI QUIERO QUEDARME CON GRUPOS DE AVISOS DIVIDIDOS EN 4 CATEGORÍAS, VOY A TENER 4 COLUMNAS EN LA MATRIZ.

• PARA ESO, VOY A HACER UNA APROXIMACIÓN DE RANGO $K=4$, UTILIZANDO LOS 4 VALORES SINGULARES DE MAYOR IMPORTANCIA DE LA MATRIZ Σ (LAS PRIMERAS 4 FILAS Y COLUMNAS).

- ADEMÁS, TENDRÍA QUE QUEDARME CON LAS PRIMERAS 4 COLUMNAS DE LA MATRIZ U Y LAS PRIMERAS 4 COLUMNAS DE V .
- HACIENDO ESTA APROXIMACIÓN, OBTENDRÍA UNA MATRIZ QUE APROXIMA BASTANTE BIEN A LA MATRIZ A, YA QUE ESTÁ DEMOSTRANDO QUE LA SVD DA LA MEJOR APROXIMACIÓN DE RANGO K POSIBLE A LA MATRIZ ORIGINAL (SE DEMUESTRA CON EL TEOREMA DE ECKERT-YOUNG)
- ENTONCES, NUESTRA MATRIZ U_r "CONECTA" CIUDADES CON CATEGORÍAS DE AVISOS; NUESTRA MATRIZ V_r RELACIONA AVISOS CON LAS CATEGORÍAS; Y NUESTRA MATRIZ Σ_r NOS DA UNA IDEA DE LA IMPORTANCIA DE CADA CATEGORÍA.

$U_r (17 \times 4) \quad \Sigma_r (4 \times 4) \quad V_r (4 \times 1500)$

B) PARA SABER CUANTO INVERTIR EN CADA CIUDAD, ME FICARÍA QUE CIUDAD TIENE LA MAYOR CANTIDAD DE CLICKS EN LOS AVISOS, Y A ESA CIUDAD LE DARÍA MÁS DIVERSO. PARA ESTO TENDRÍA QUE MIRAR LOS VALORES DE LAS CELDAS DE LA MATRIZ U_r , Y VER CUAL DE ESAS CELDAS TIENE UN VALOR MAYOR.

• PARA TENER UNA NOCIÓN DE LA ESCALIDAD DE LA SOLUCIÓN PLANTEADA, TENDRÍAN QUE VER QUE PORCENTAJE DE LA ENERGÍA DE LA MATRIZ ORIGINAL SE CAPTURA CON ESTA APROXIMACIÓN DE RANGO $K=4$

ESE PORCENTAJE SE PUEDE CALCULAR COMO

$$\alpha = \frac{\sum_{i=1}^K \sigma_i^2}{\sum_{i=1}^N \sigma_i^2}$$

\rightarrow K AUTOVALORES DE LA APROXIMACIÓN
 \rightarrow N AUTOVALORES TOTALES