

# Introduction to Topological Data Analysis

	H	M	R	C	W
H	0	11	10	14	22
M	11	0	3	13	21
R	10	3	0	12	20
C	14	13	12	0	16
W	22	21	20	16	0

- Human, Mouse, Rat, Cat, Whale
- Halfpenny, MBS, River Island, Currys, whitbread

$$\text{dist}(H, M) = \text{dist}(M, H)$$

$$\text{dist}(H, H) = 0$$

Choose some  $\Sigma > 0$ , called a threshold

and consider the graph  $G_\Sigma$   
with vertices

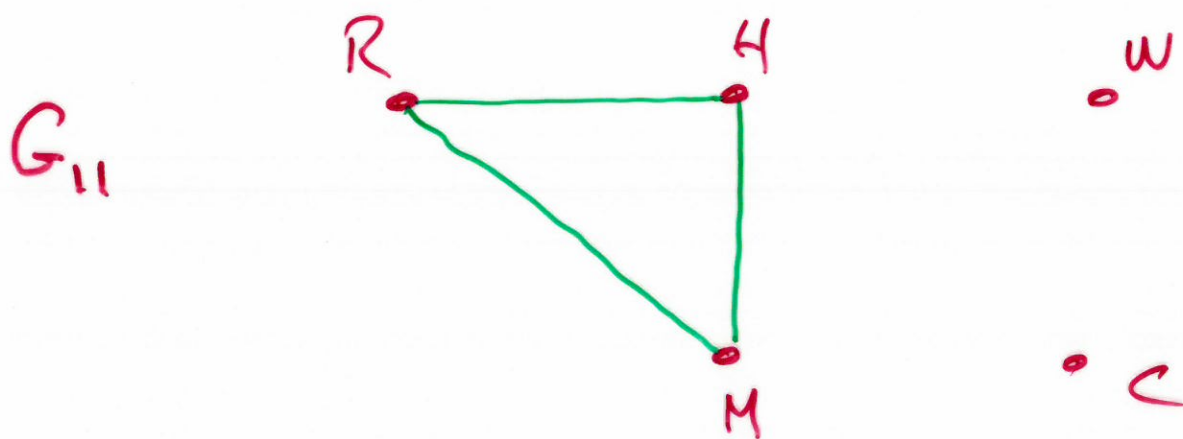
$H, M, R, C, W$

and with an edge



if  $\text{dist}(X, Y) \leq \Sigma$ .

for  $\Sigma = 11$  we have



We regard this graph as a  
subspace of  $\mathbb{R}^5$  by  
identifying

$$H = (1, 0, 0, 0, 0) = e_1$$

$$H = (0, 1, 0, 0, 0) = e_2$$

$$R = (0, 0, 1, 0, 0) = e_3$$

$$C = (0, 0, 0, 1, 0) = e_4$$

$$W = (0, 0, 0, 0, 1) = e_5$$

The graph  $G_{11}$  can be thought of as a subspace of  $\mathbb{R}^5$  with points  $e_1, e_2, e_3, e_4, e_5$  and line segments

$$e_1 e_3, e_2 e_3, e_1 e_2.$$

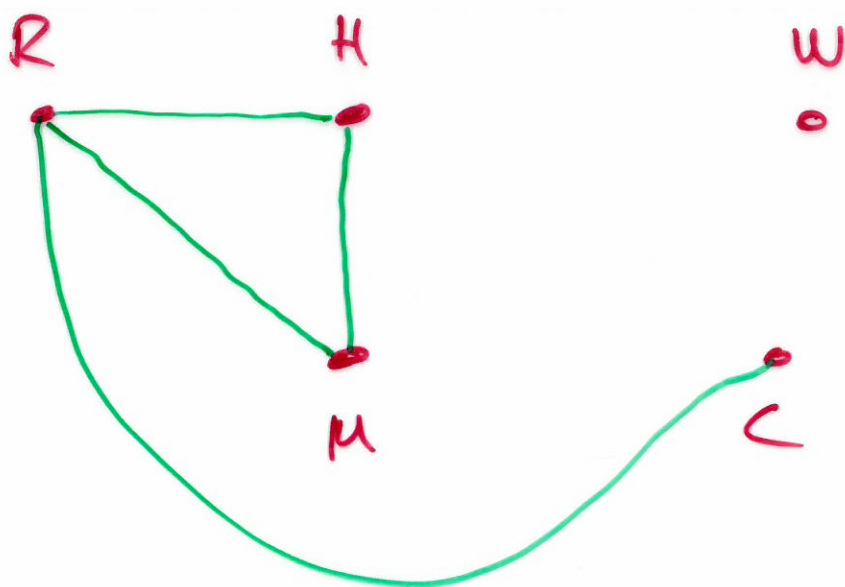
The space  $G_{11}$  has **three**

connected components,  $X_W, X_C,$

$X_{RHM}.$

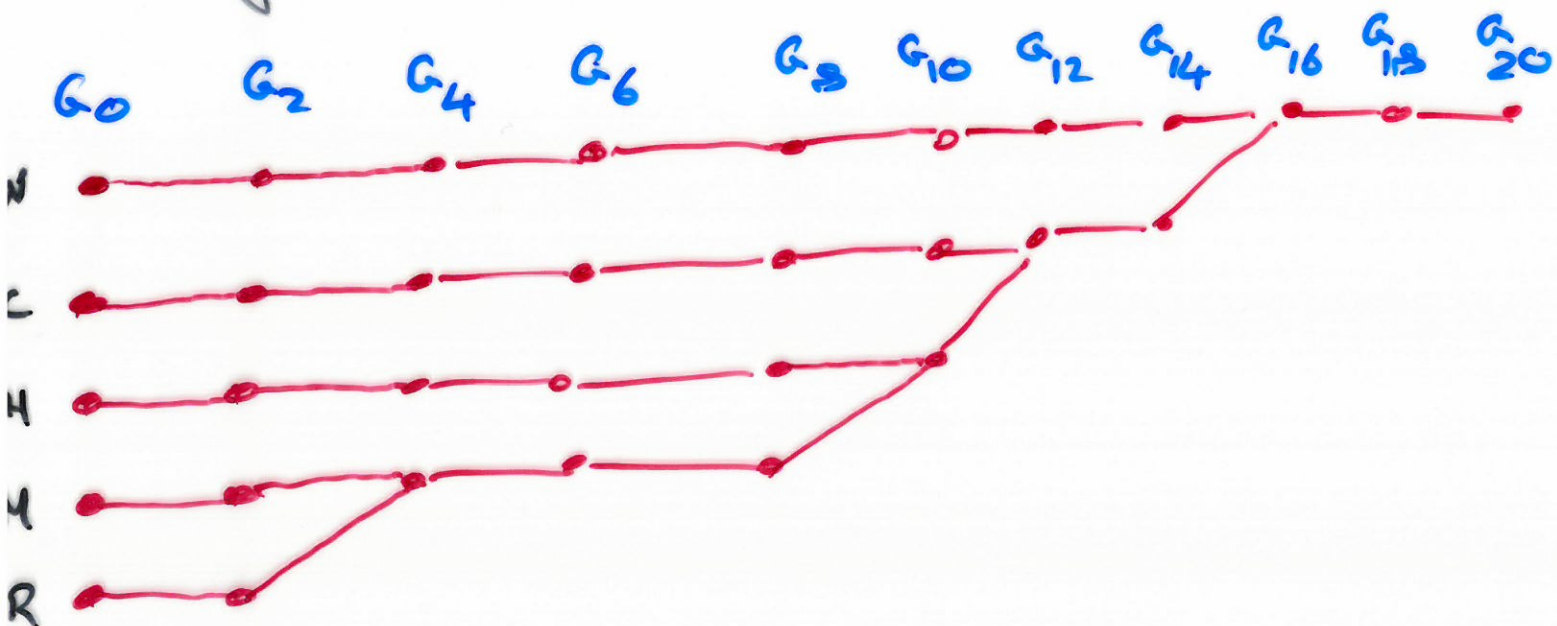


Now let's look at  $G_{12}$ .



So  $G_{12}$  has two connected components  $X_W$ ,  $X_{RHMC}$

A dendrogram summarizes the inclusions of connected components as follows:





Then between A & B. various  
kinds of relation. C & B. The  
first predation, B & D  
rather greater interaction  
Then some could be  
formed. - binary relation

# Continuity

Defn Let  $X, Y$  be topological spaces. A function  $f: X \rightarrow Y$  is continuous if the inverse image of every open set in  $Y$  is an open set in  $X$ .

In other words, if  $U \subseteq Y$  is an open set in  $Y$  then

$$f^{-1}U = \{x \in X : f(x) \in U\}$$

is open in  $X$ .

We often use the term map to mean a continuous function.