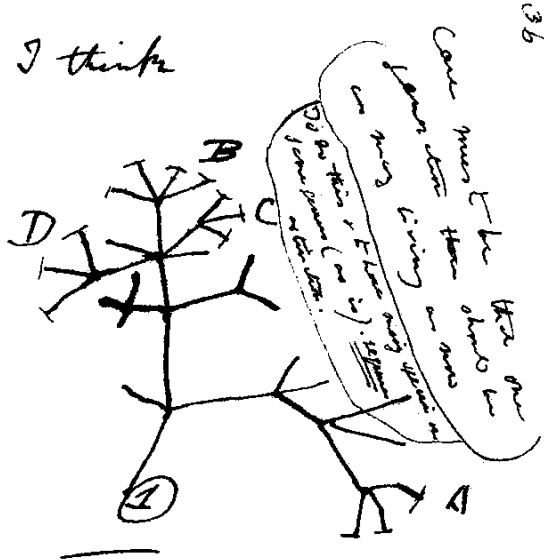


I think



Thus between A & B. various
 sort of relation. C & B. The
 finest gradation, B & D
 rather greater distinction
 Thus genera would be
 formed. - bearing relation

I think case must be that one
 generation should have as many
 living as now. To do this and
 to have as many species in the
 same genus requires extinction. Thus
 between A & B the immense gap of
 relation, C & B the finest gradation,
 B & D rather greater distinction. Thus
 genera would be formed. Bearing
 relation to ancient types with several
 extinct forms. Charles Darwin 1837

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
- Honda, Microsoft, Ripple, Colsoft, Walmart

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

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

Choose $\Sigma > 0$, called a threshold, and consider the graph G_Σ

with vertices

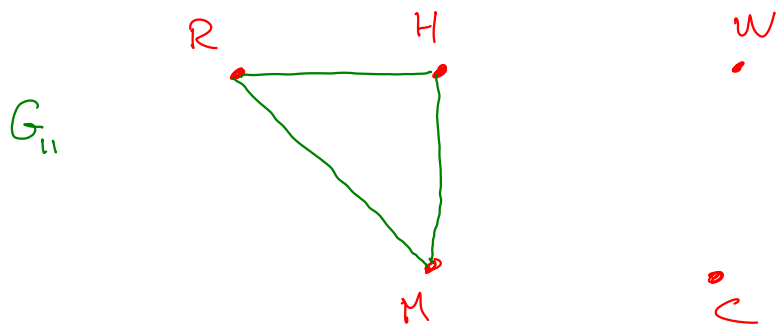
H, M, R, C, W

and with edge



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

e.g. $\Sigma = 11$



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

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

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

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

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

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

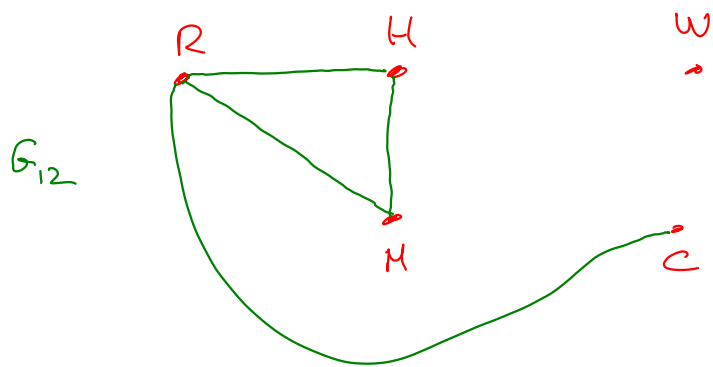
The graph G_{11} can be thought of as a subspace of \mathbb{E}^5 with points e_1, e_2, \dots, e_5 and the line segments

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

The graph G_{11} has **three** connected components:

$$X_W, X_C, X_{RHM}$$

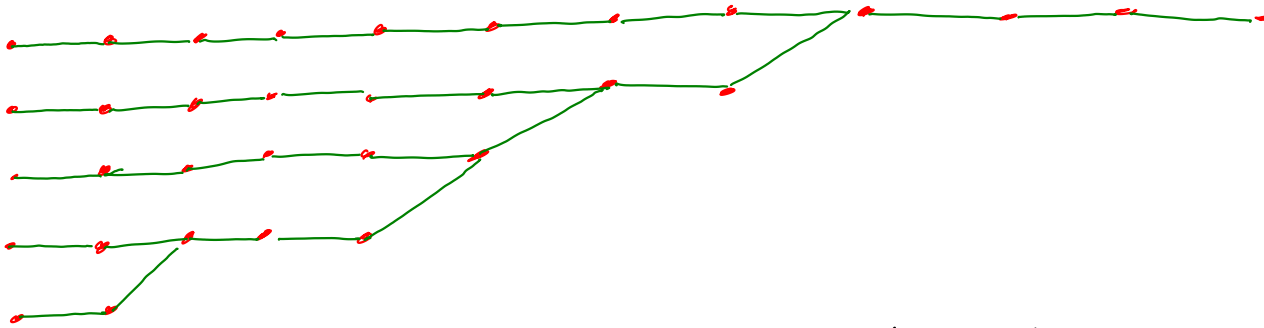
Now let's look at G_{12} :



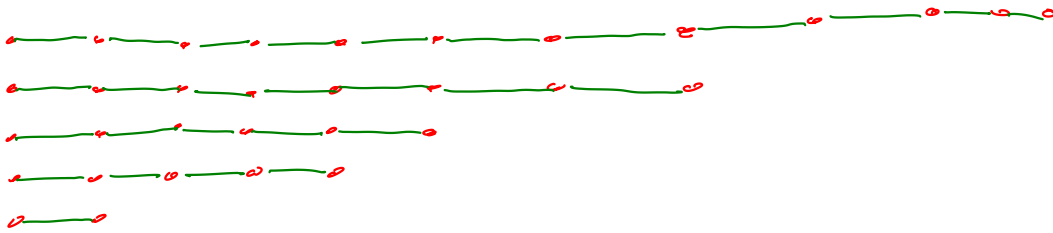
So G_{12} has **two** connected components.

A dendrogram or phylogenetic tree summarizes the inclusion of connected components.

G_0 G_2 G_4 G_6 G_8 G_{10} G_{12} G_{14} G_{16} G_{18} G_{20}



We can also view this as a bar code



This is just standard cluster analysis but using the language of topology.

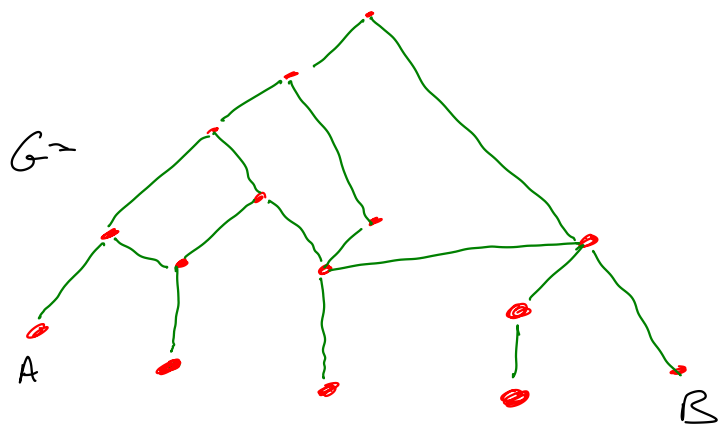
Phylogenetic trees are used throughout biology.

However :

Reassortment

If a single host (e.g. human) is infected by two different strains of an influenza virus, it is possible that new assembled viral particles will be created from segments, some segments from one strain and some from the other strain.

A phylogenetic network needs to be used when studying evolution of viruses.



Exercise For any connected graph

we have

$$\chi(G) = 1 - \text{number of holes}$$

Above example $\chi(G) = 1 - 3 = -2$

check $V = 15, E = 17, V - E = -2.$

Gunnar Carlsson, J. Chen, R. Rabadan
use a version of the ~~E~~ Euler characteristic
(known as "persistent homology") to study
horizontal evolution of influenza pandemics.