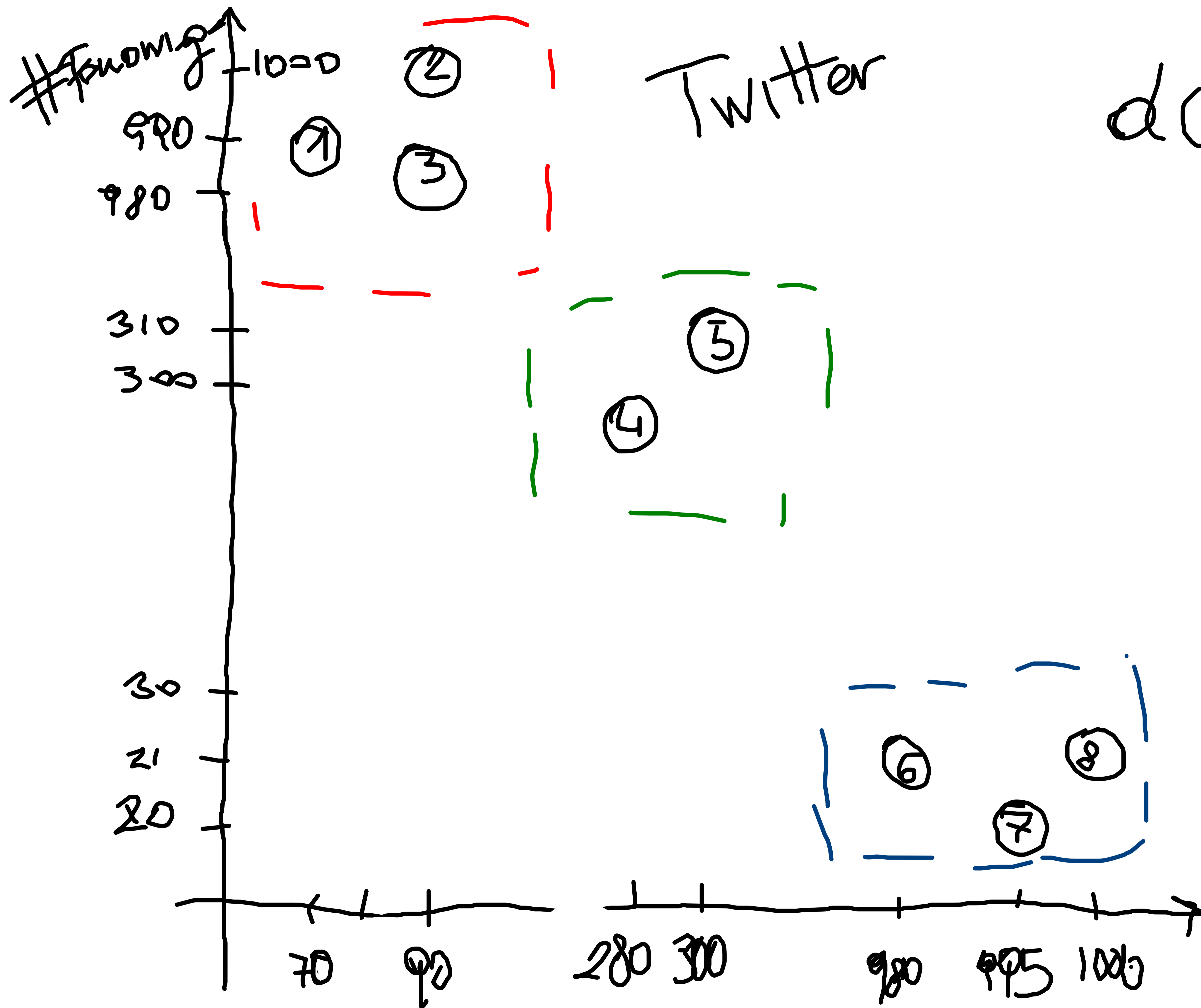


Come separamo i punti in gruppi*?

Questo set di punti può essere modellato come un grafo pesato, non diretto calcolando la distanza* tra tutte le coppie di nodi.
 ↳ Avremo un grafo completo

* Euclidea ma ci sono alternative

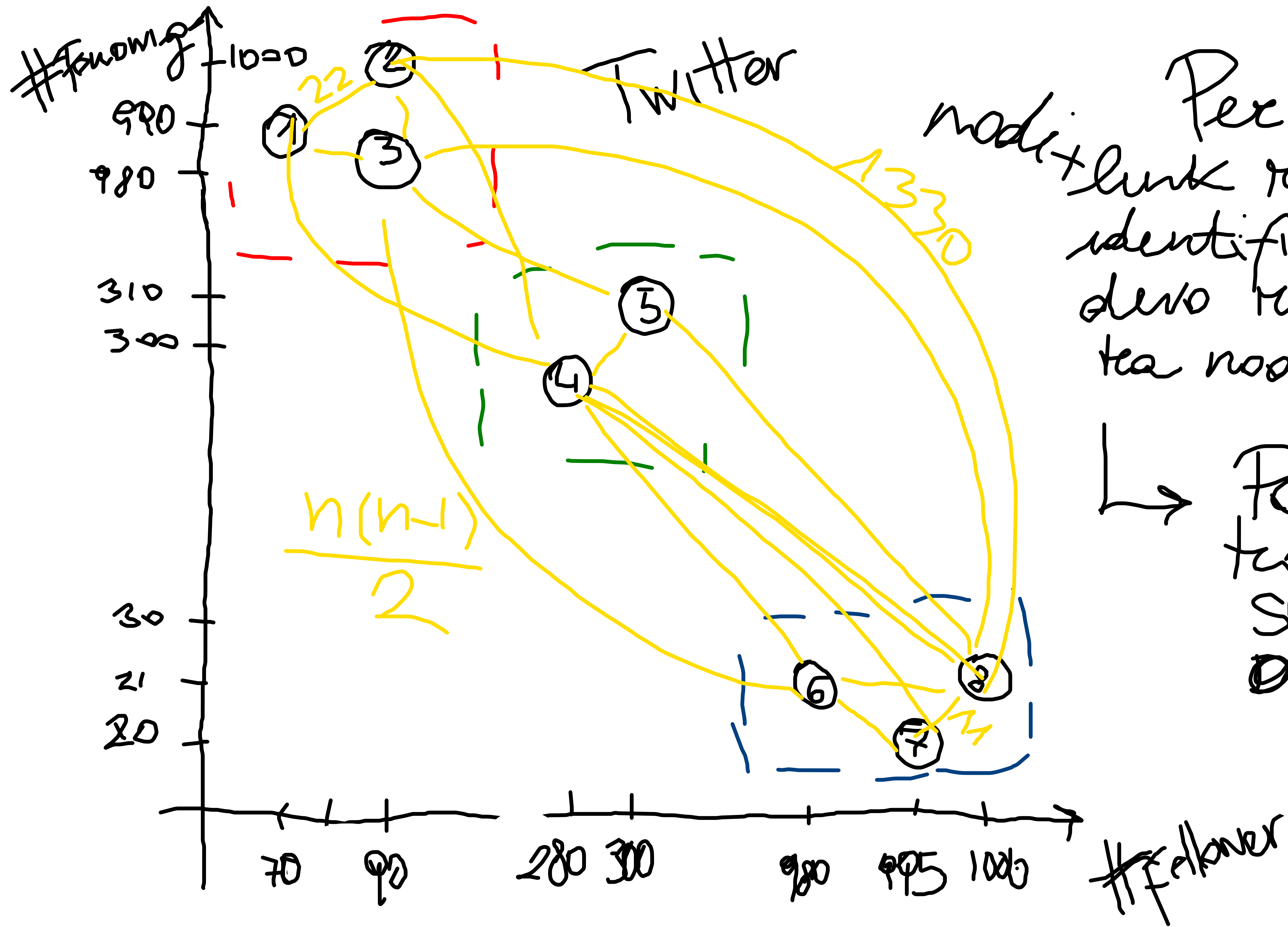
* I punti di un gruppo sono vicini tra loro e lontani dai punti fuori dal gruppo



$$d(p, q) = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2}$$

	1	2	3	4	5	6	7	8
1	\	22	22	721	718	1329	1340	1336
2		\	20	725	721	1323	1334	1320
3			\	706	702	1308	1319	1316
4				\	22	753	768	769
5					\	739	753	754
6						\	15	21
7							\	11
8								\

#Followers



Per mantenere solo i link rilevanti al fine della identificazione dei gruppi devo rimuovere le connessioni tra nodi distanti.

↳ Posso, per esempio, trovare il MINIMUM SPANNING TREE del graph.

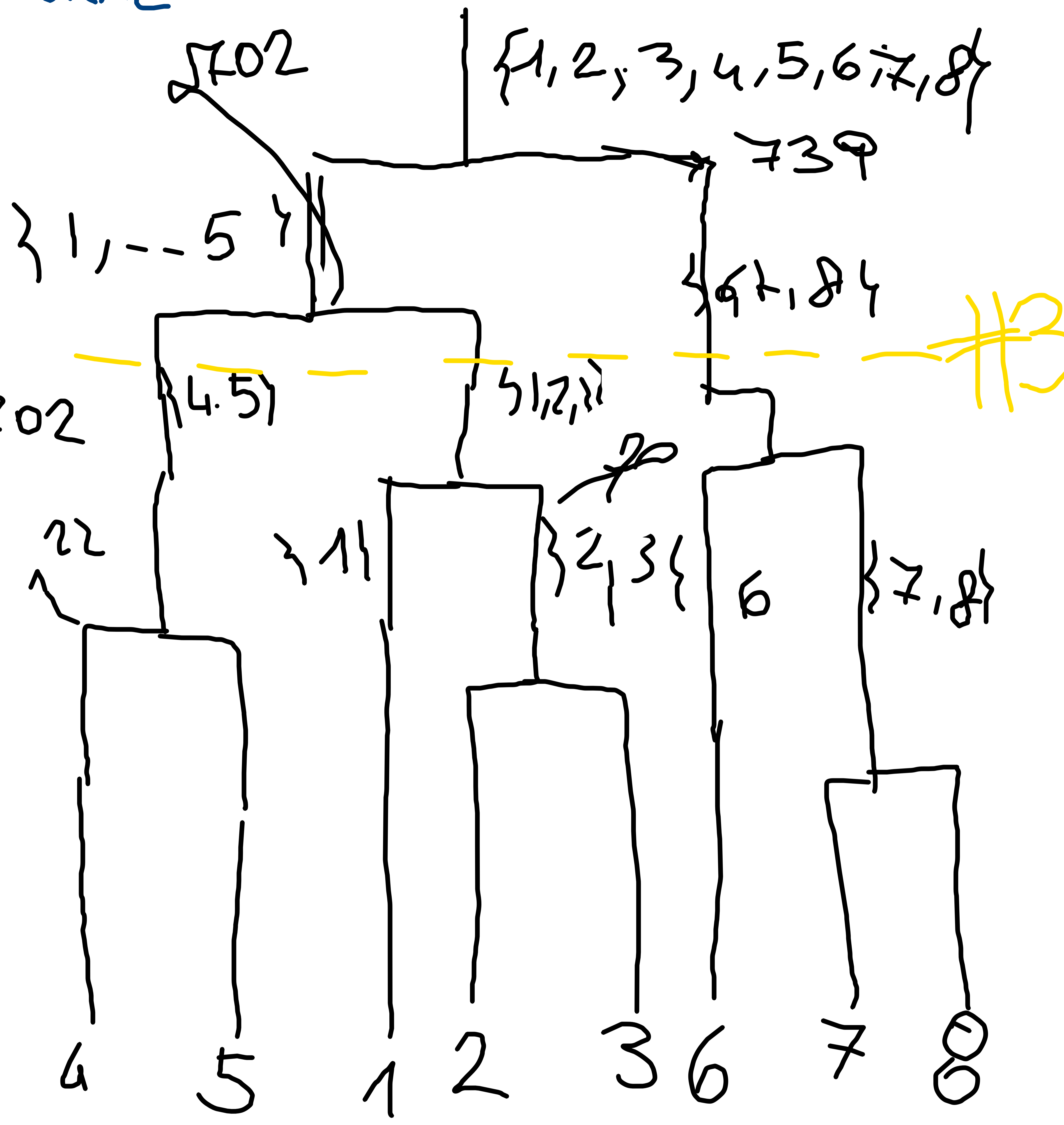
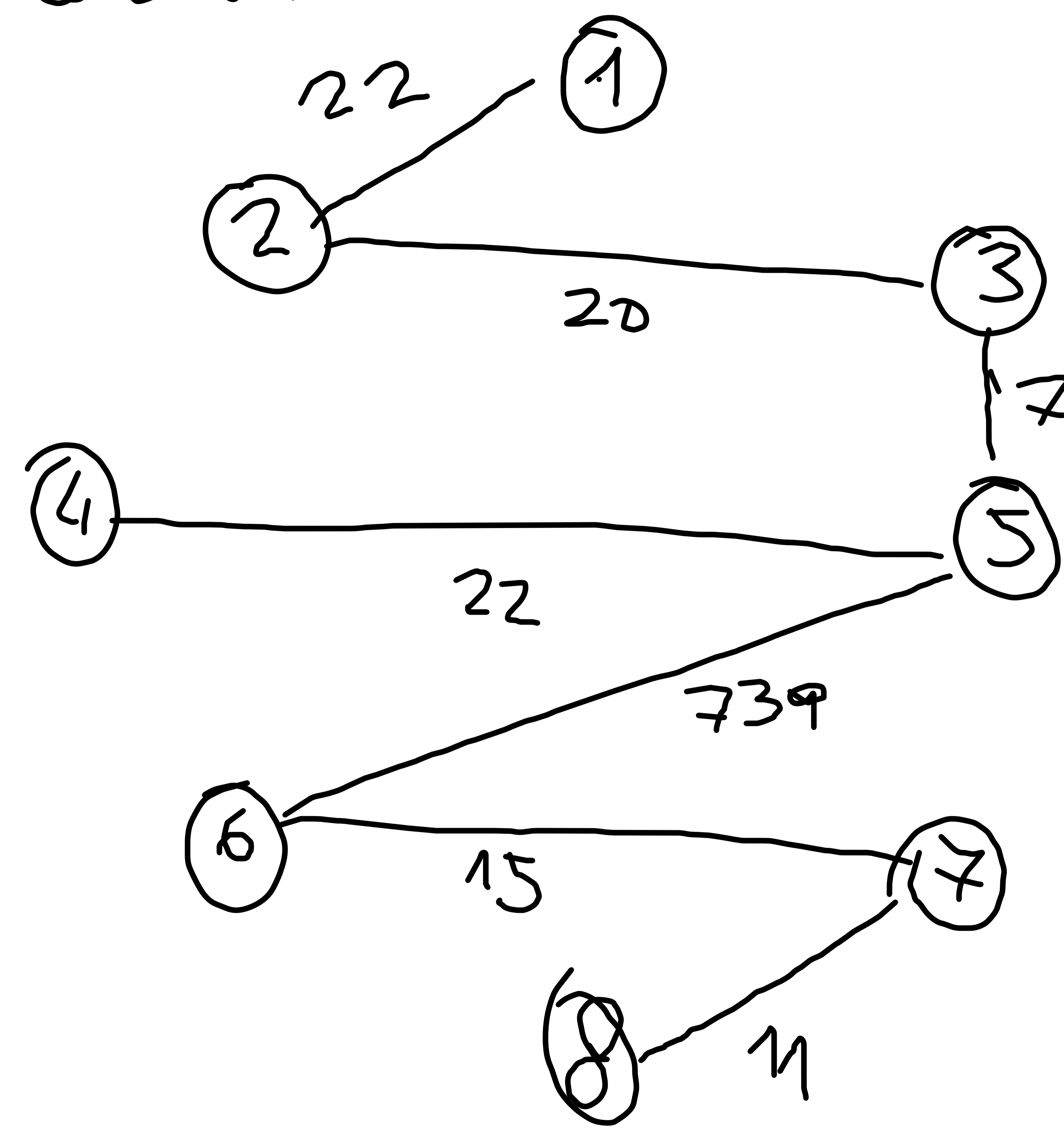
ALGORITMO DI KRUSKAL - Trova un Minimum Spanning Tree

- Inizializza A come un insieme vuoto di archi
- Crea $|V|$ alberi
- Ordina i pesi degli archi in ordine non decrescente e scorri la lista
- Se u e v \in ad alberi diversi aggiungi (u, v) ad A
- Fondi i due alberi che contengono u e v .

APPLICO ALGORITMO DI KRUSKAL

Ordinamento archi:

u	v	w	
7	8	41	✓
6	7	15	✓
2	3	20	✓
6	8	22	✗
1	2	22	✓
1	3	22	✗
4	5	22	✓
3	5	702	✓
3	4	706	✗
1	5	710	✗
1	4	721	✗
2	5	721	✗
2	4	725	✗
5	6	739	✓
...	

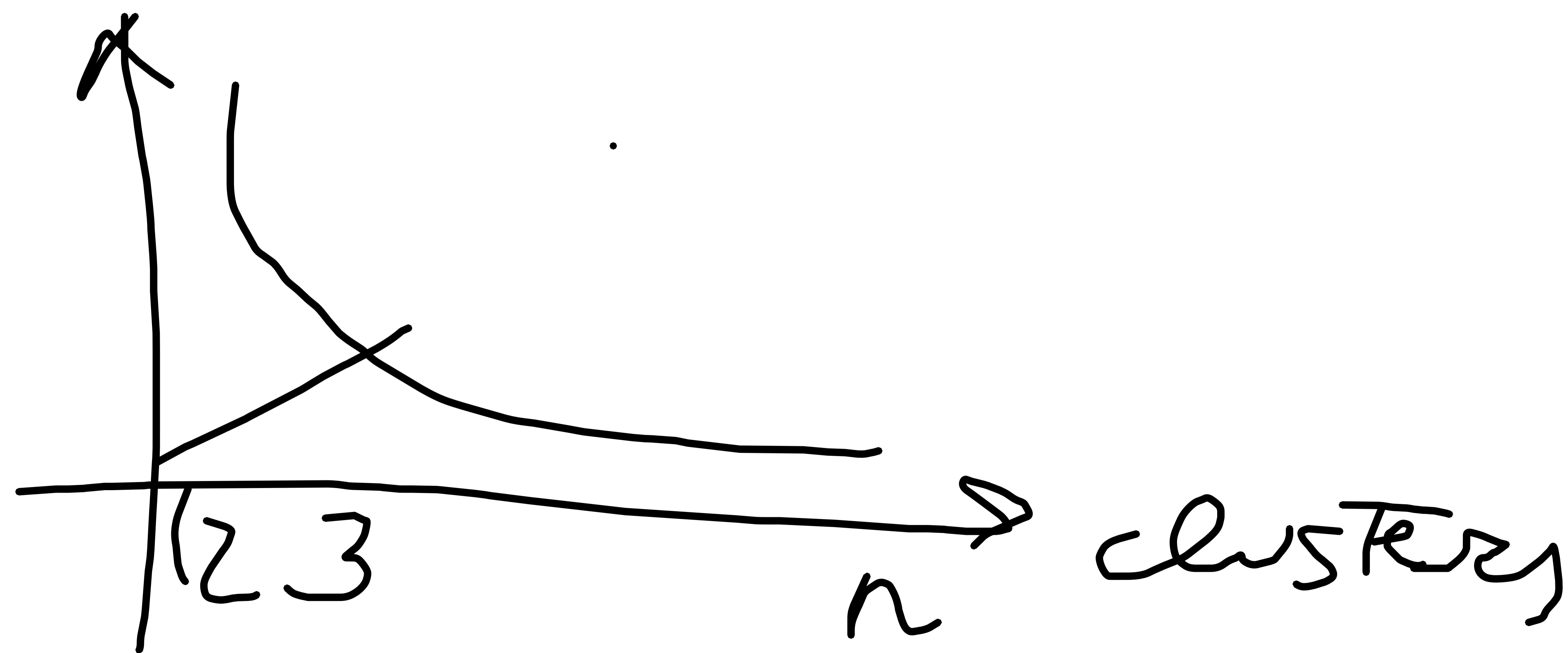


Dendrogramma.

Per trovare il numero di cluster dobbiamo dividere il grafo di prima in componenti connesse. Possiamo:

1) Andare "ad occhio" disegnando un istogramma dei per

2) Utilizzare il metodo detto Elbow Method per cui proviamo varie divisioni in cluster e calcoliamo la somma delle distanze da ogni punto del cluster al loro baricentro



QUESTO APPROCCIO
AL CLUSTERING SI USA
AD ESEMPIO NELL'ALG. K-MEANS