

Taller 1

Árbol generador mínimo
Camino mínimo

Objetivo

- Ver más aplicaciones de AGM en otros contextos.
- Aprender un poco haciendo.
- Charlar un poco sobre el análisis cualitativo.
- Primera aproximación a competencias de programación con caminos mínimos.
- Llevarse código para el TP.

¿Por qué un taller?

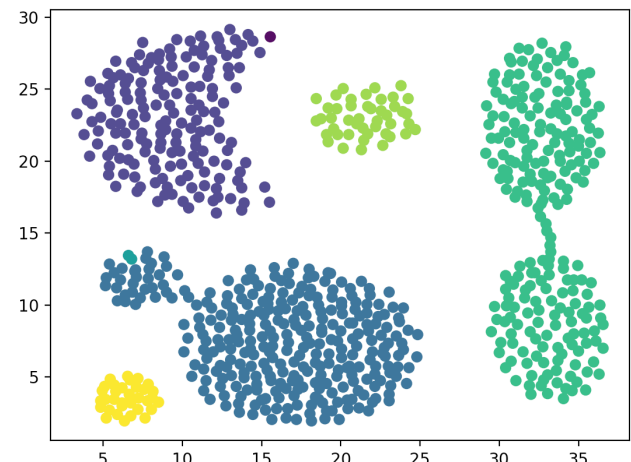
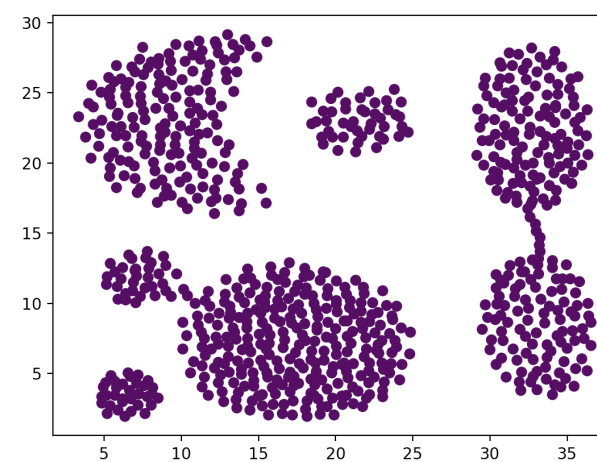
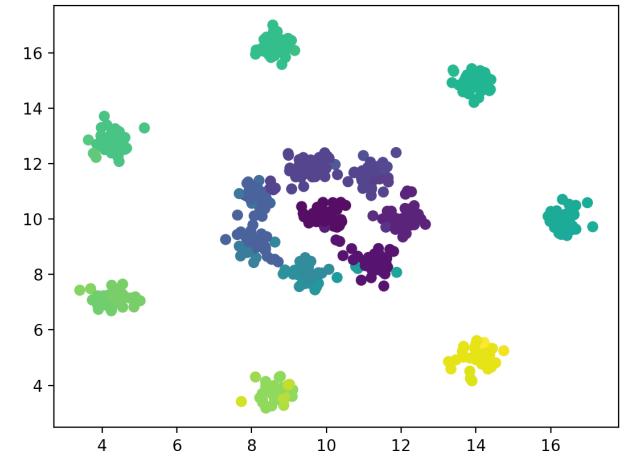
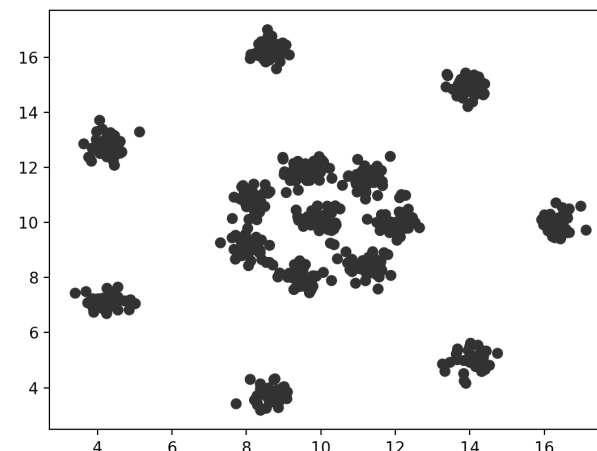
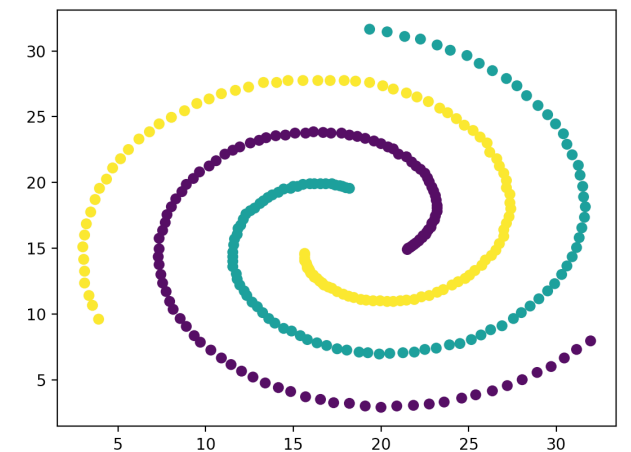
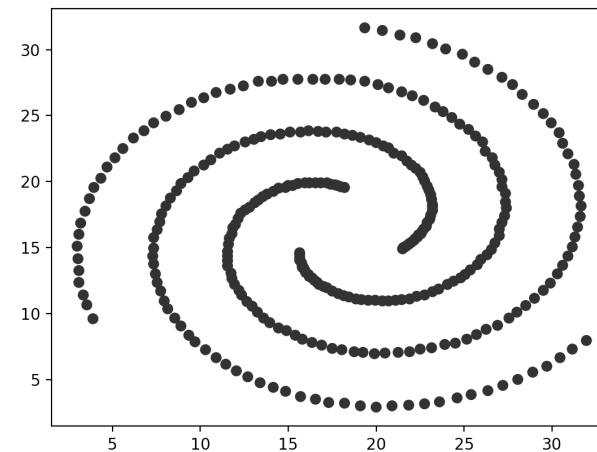


“Lo que vemos cambia lo que sabemos. Lo que conocemos cambia lo que vemos.”

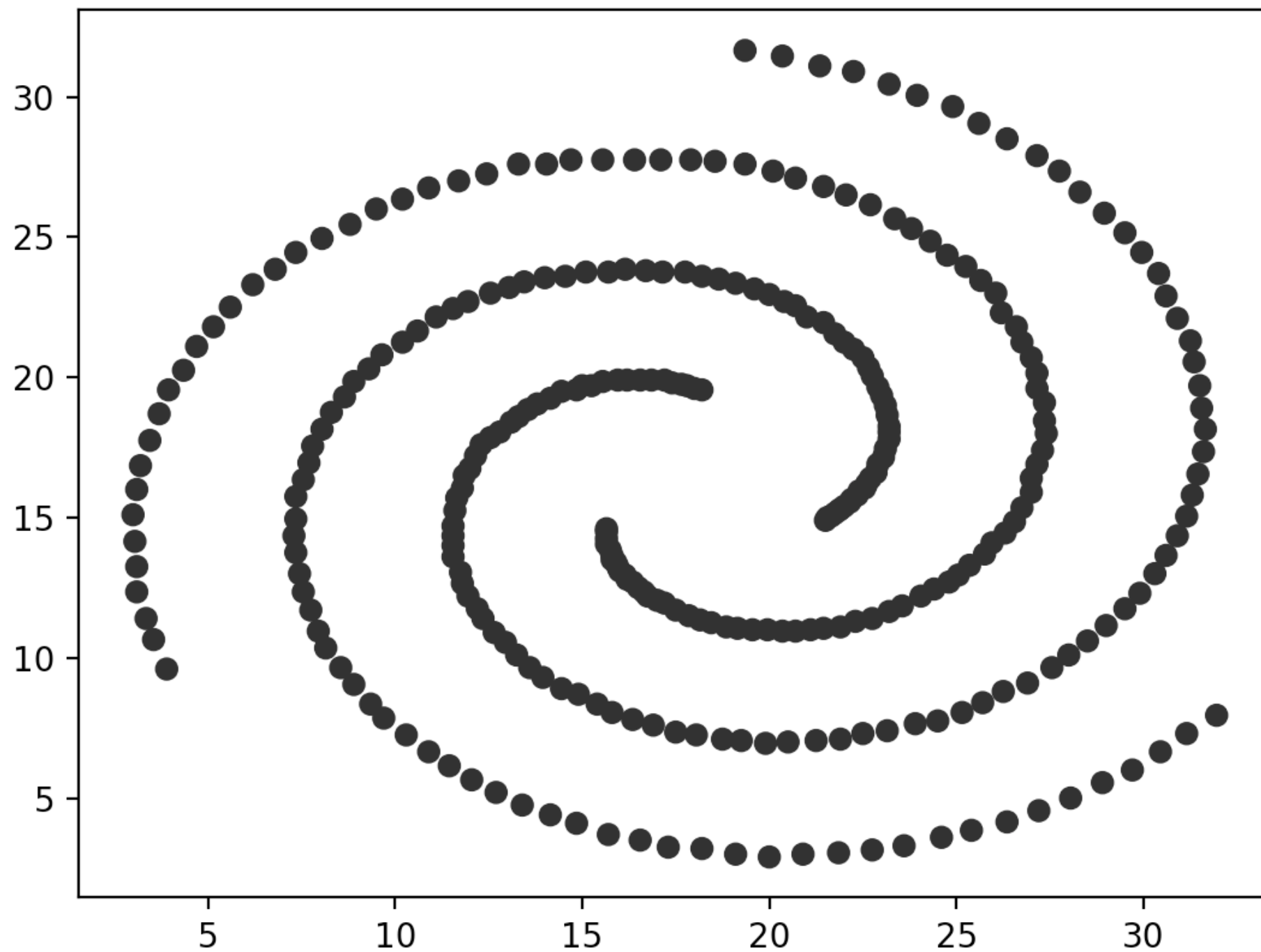
–Jean Piaget

Problema de Clustering

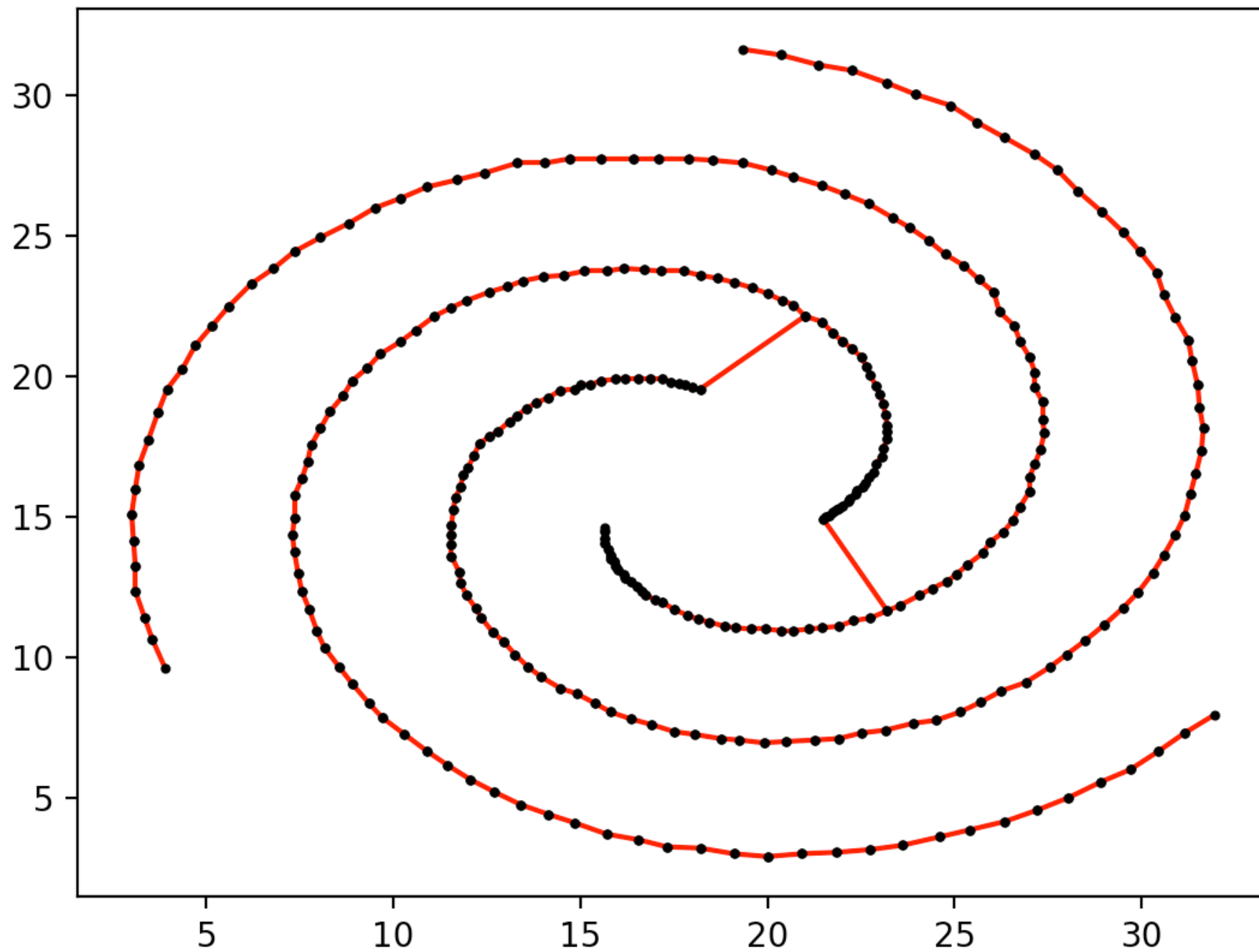
En *aprendizaje por computadora* uno de los problemas fundamentales es el de *clusterizar*. Clusterizar significa agrupar un conjunto de objetos en diferentes subconjuntos por alguna noción de similitud intraconjunto (o disimilitud interconjunto).



¿Ideas?

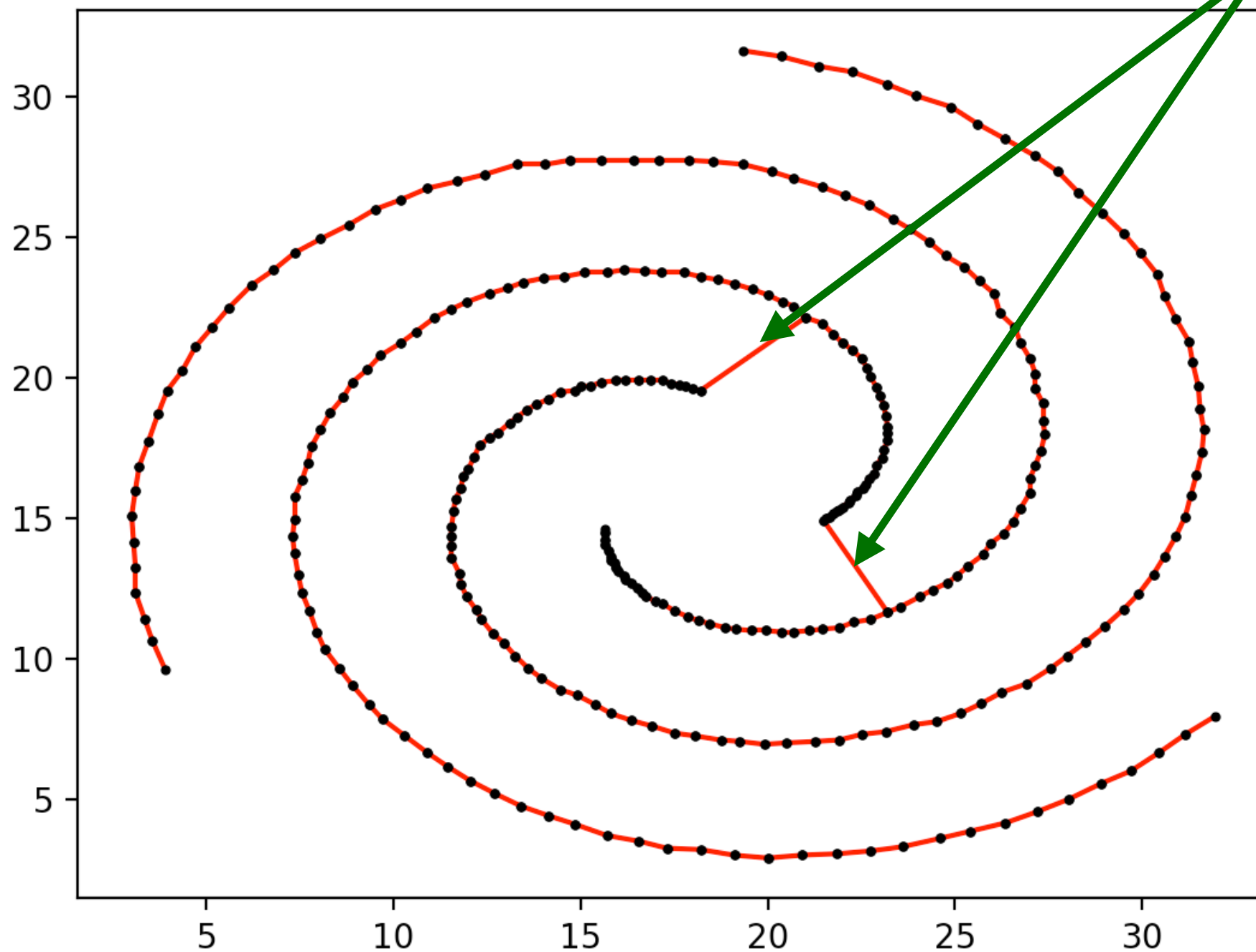


¿Ideas? (2)



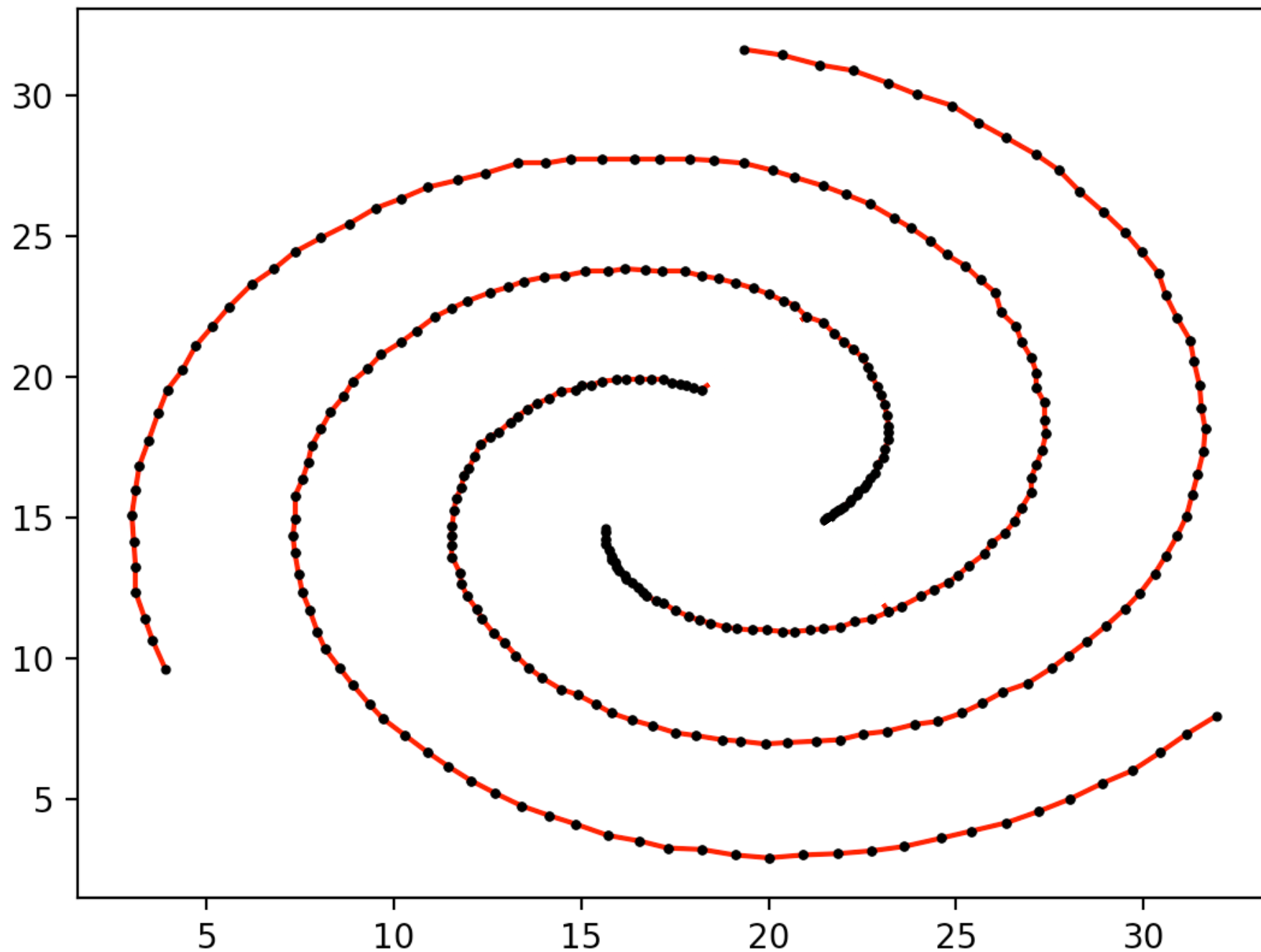
¿Ideas? (3)

¿Son muy largos?



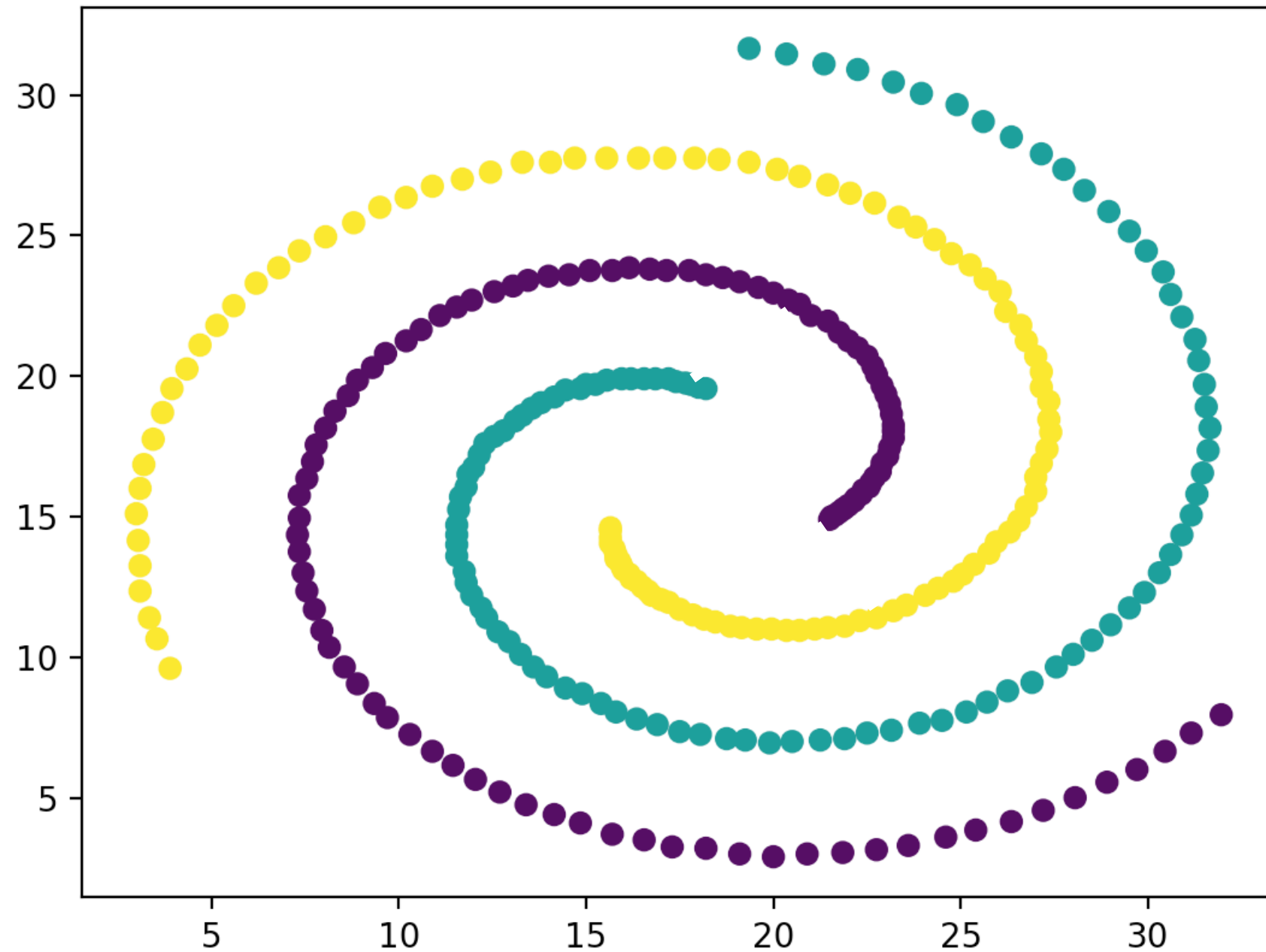
¿Ideas? (4)

¡Los sacamos!



Ahora cada componente conexa es un cluster.

¿Ideas? (5)



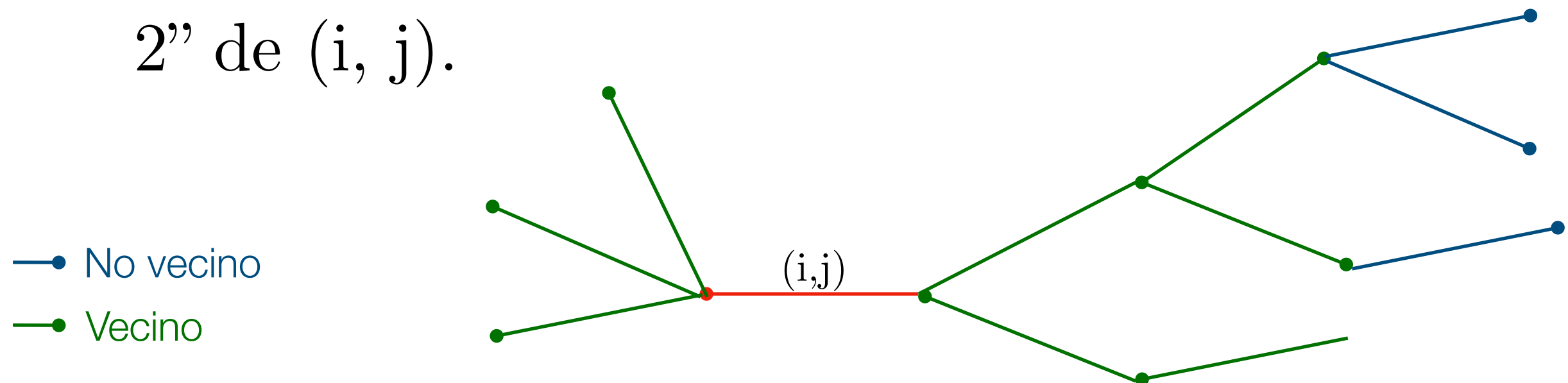
¿Cuándo sacamos una arista?

Decimos que una arista (i, j) de un AGM T es *inconsistente* si:

$$w_{ij} > K \text{ PesoPromedio}(\text{Vecinos}(i, j))$$

Donde:

- K es un parámetro de implementación,
- $\text{Vecinos}(i, j)$ son las aristas “a distancia a lo sumo 2” de (i, j) .



Formalizamos

- Input: Conjunto P de n puntos (x_i, y_i) en 2D.
- Output: Un número de cluster c_i para cada punto i .

Algorithm 1 Algoritmo de clustering basado en AGM (Charles T. Zahn)

```
function CLUSTERING( $P, n$ )  
   $G \leftarrow \langle \{1, \dots, n\}, \{(i, j) \mid i, j \in 1 \dots n, i \neq j\} \rangle$ .  
   $w_{ij} \leftarrow \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$   
   $T \leftarrow AGM(G)$   
  for  $(i, j) \in T$  do  
    if EsInconsistente( $(i, j)$ ) then  $T \leftarrow T \setminus \{(i, j)\}$   
  return ComponentesConexas( $T$ ).
```

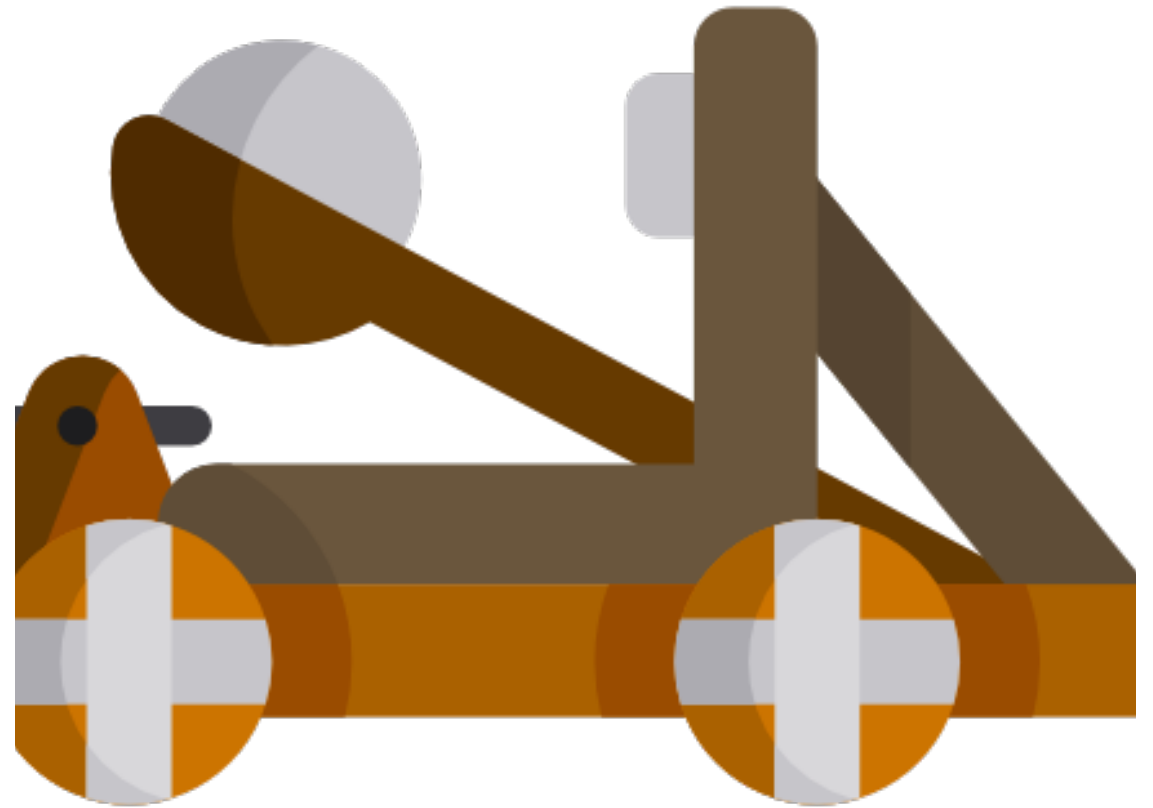
Su misión

- Completar los algoritmos de Prim y Kruskal.
- Tener en cuenta que G está implementado con listas de adyacencia.
- El diccionario de distancias w está implementado como matriz.
- Variar el parámetro K y ver qué pasa con la calidad del clustering en las distintas instancias.
- Sacar conclusiones de la calidad de las soluciones.

Puesta en común

HEISLAZY - He is Lazy

- Humberto tiene un campo que es una grilla.
- Puede caminar en linea recta desde cualquier punto a cualquier otro.
- Pero... es vago. Entonces instaló catapultas que permiten saltar F celdas.
- Humberto quiere ir desde (P_x, P_y) hasta (L_x, L_y) caminando lo menos posible.
- ¿Lo ayudás?



A

B

C

D

E

F

1



2



3



3

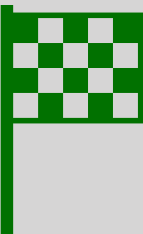
4

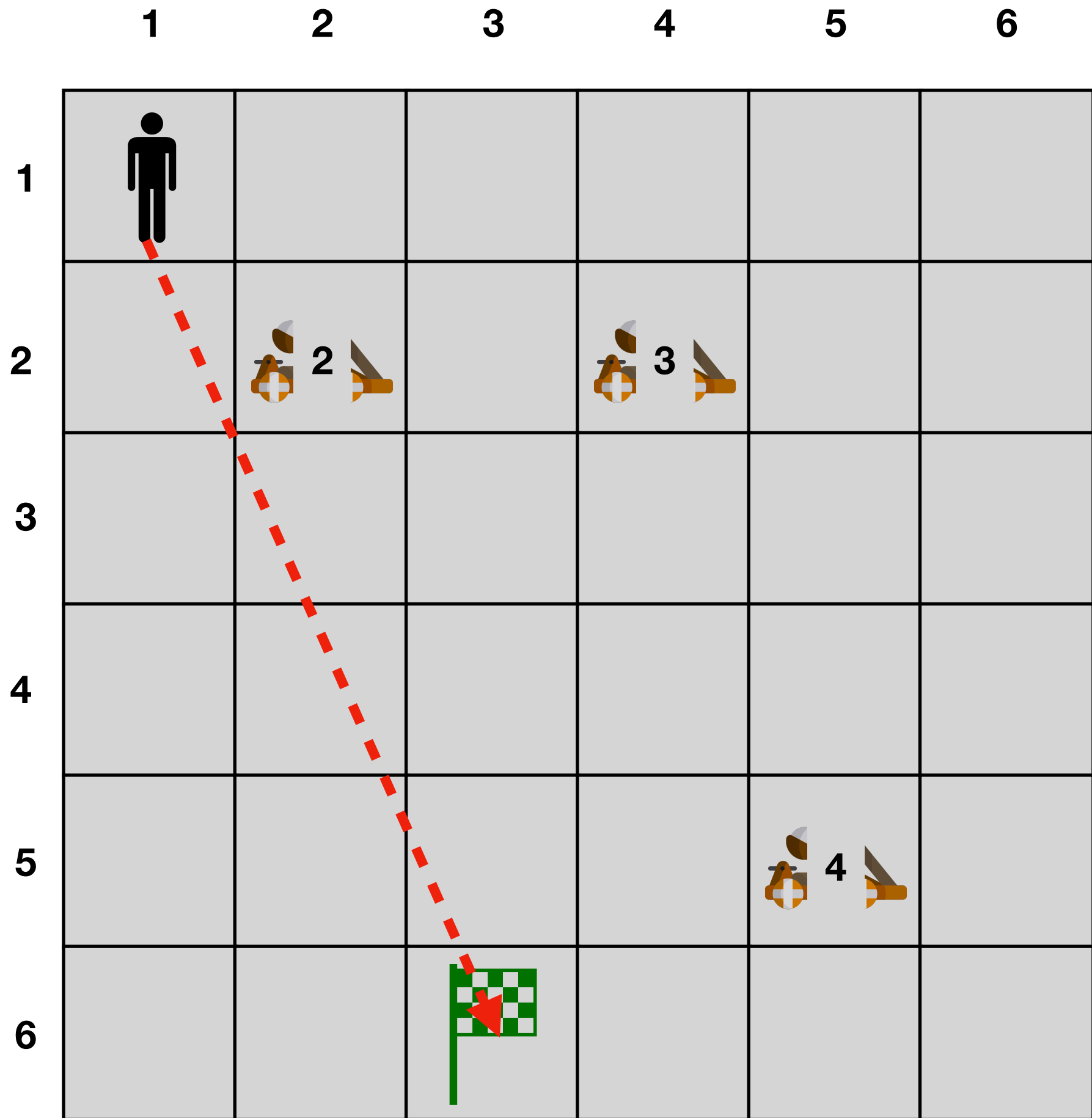
5

4

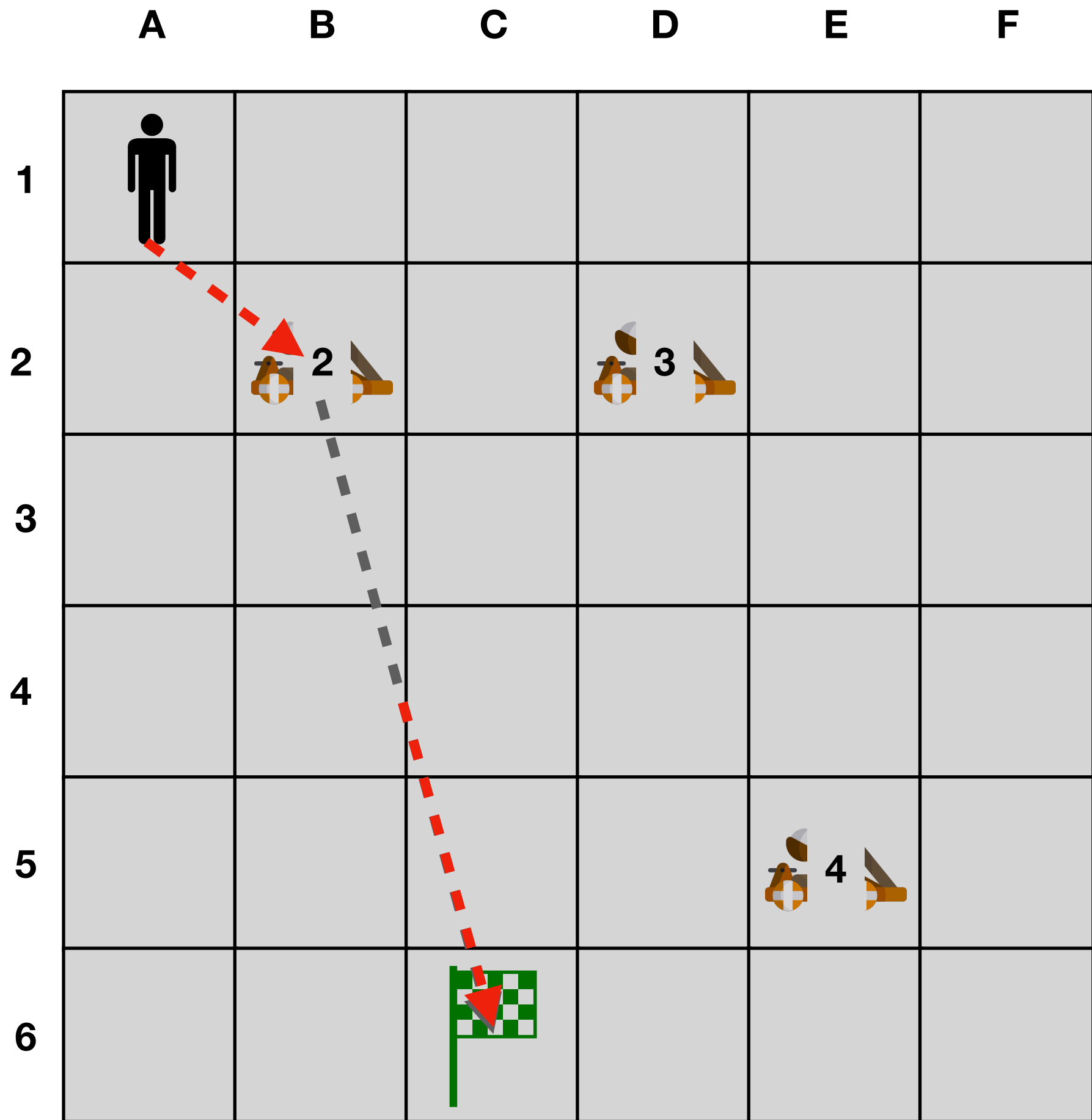


6





Costo = $\text{dist}((1,1), (3,6)) = 5.38$



Costo = 1.41 + 2.112 = 3.53

¿Cómo probar el la solución?

- El problema está en el juez online SPOJ.
 - <https://www.spoj.com/problems/HEISLAZY/>
- En la carpeta del taller ya hay una cáscara que lee los casos de entrada e implementa una solución, pero no sabíamos algoritmos de camino mínimo, así que falta implementar la solución.
- Una vez que se crea tener una solución correcta, enviar a SPOJ.