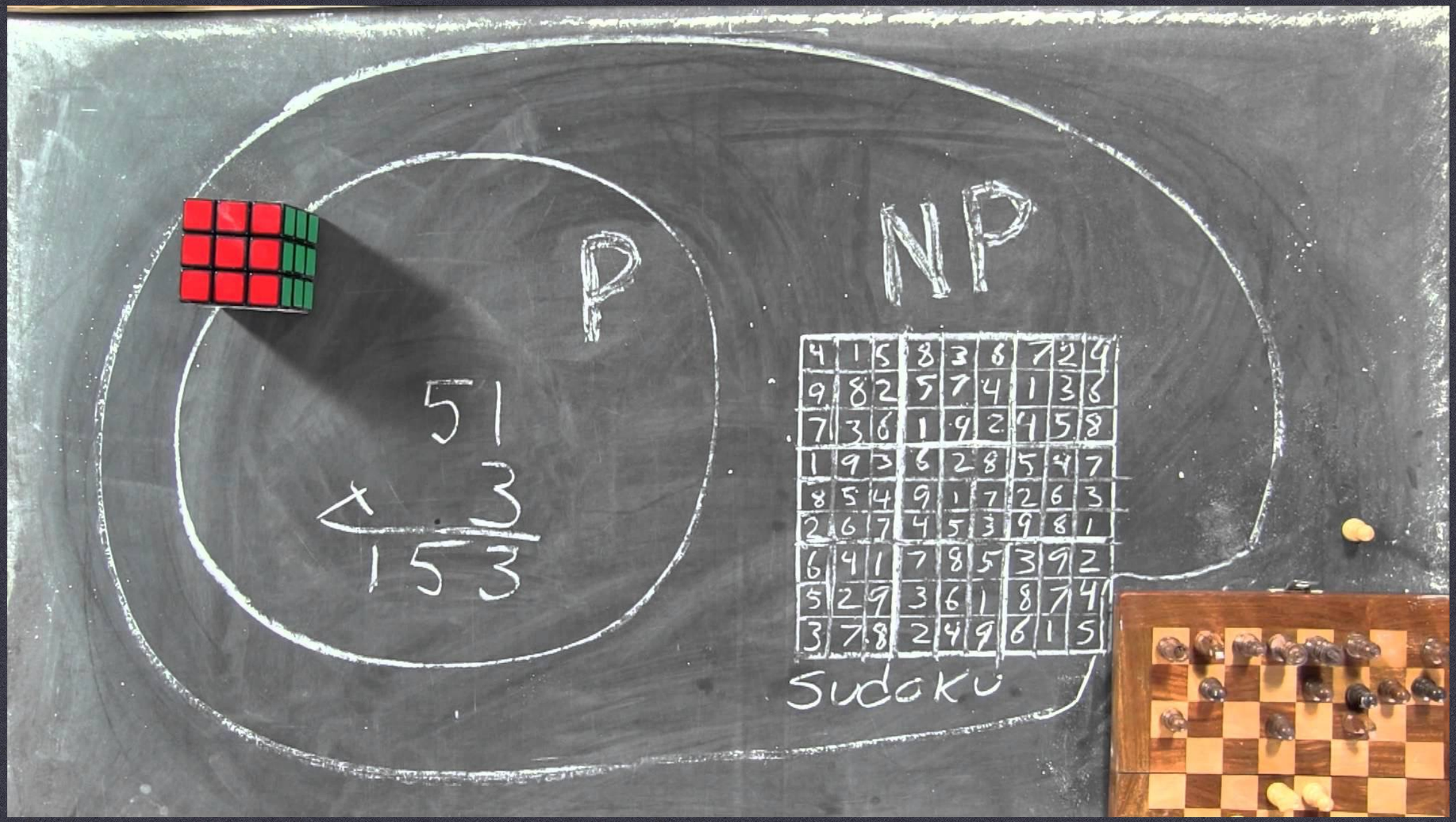


P VS NP

ALESSANDRO PANCONESI
COMPUTER SCIENCE - SAPIENZA UNIVERSITY OF ROME

PARIS, MARCH 24, 2000

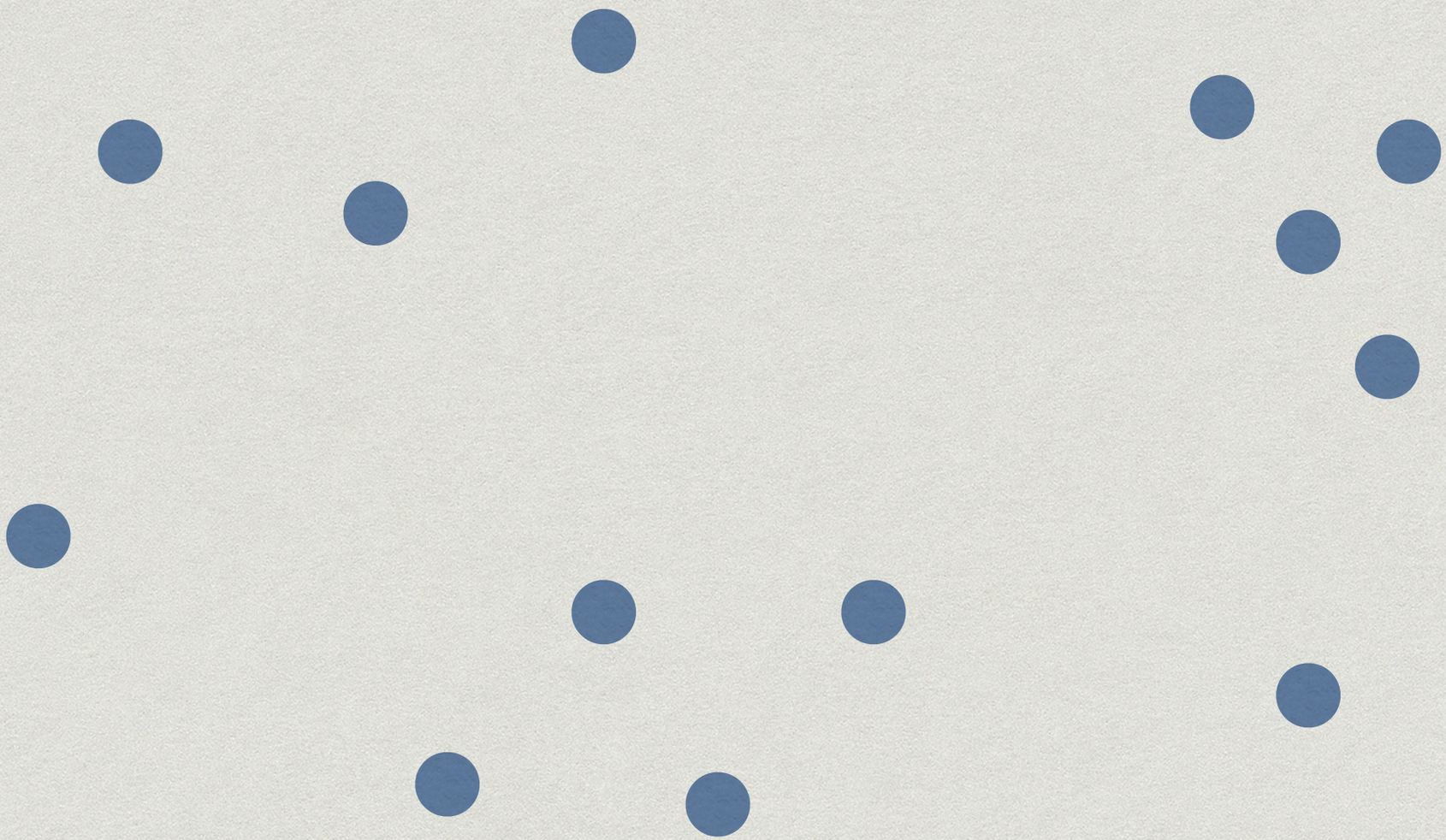
COLLEGE DE FRANCE

CLAY MATHEMATICS INSTITUTE'S ANNOUNCEMENT: THE MILLENNIUM PRIZE PROBLEMS

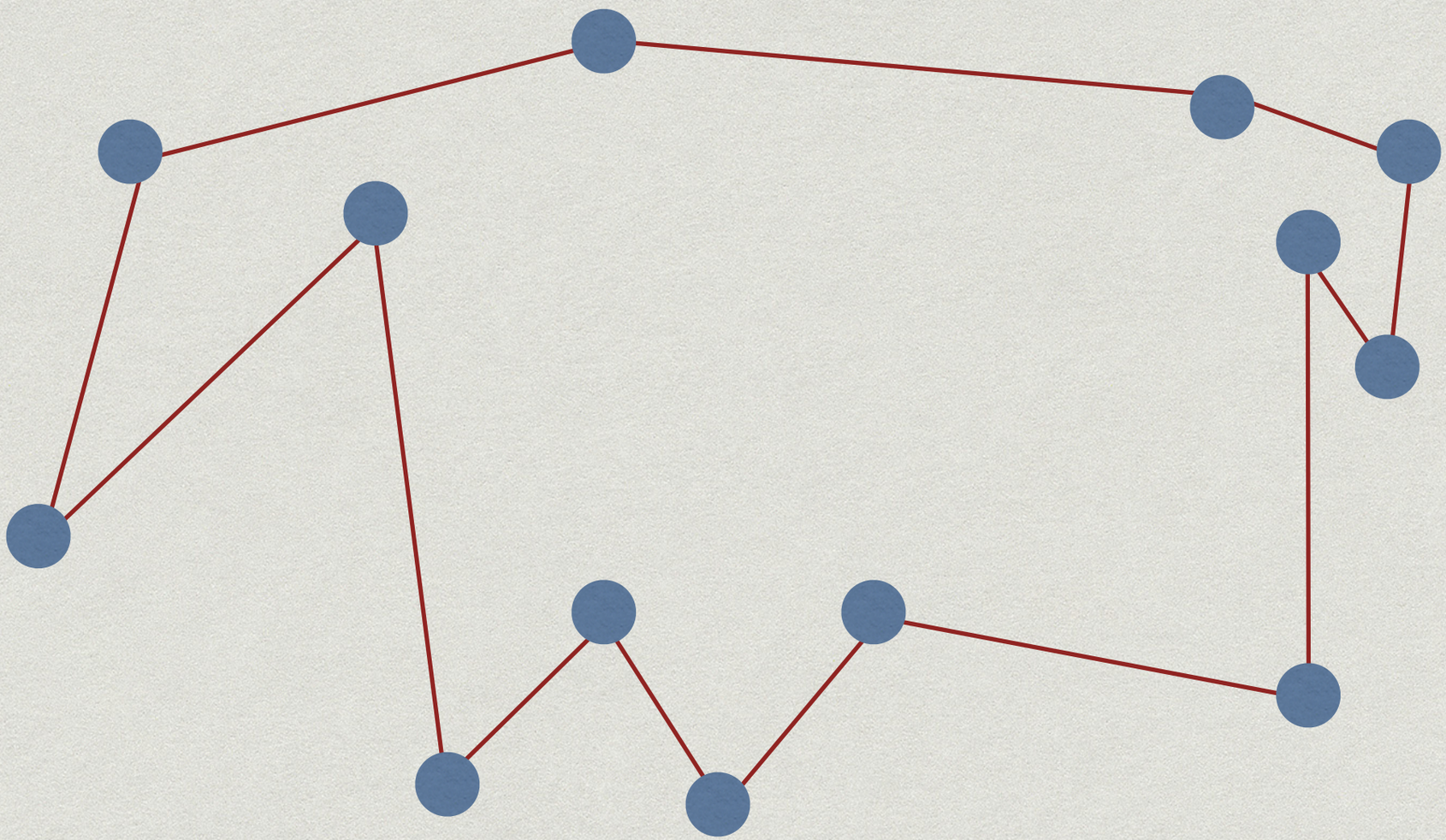
Paris, August 1900

- One hundred years earlier David Hilbert presented his famous list of 23 problems

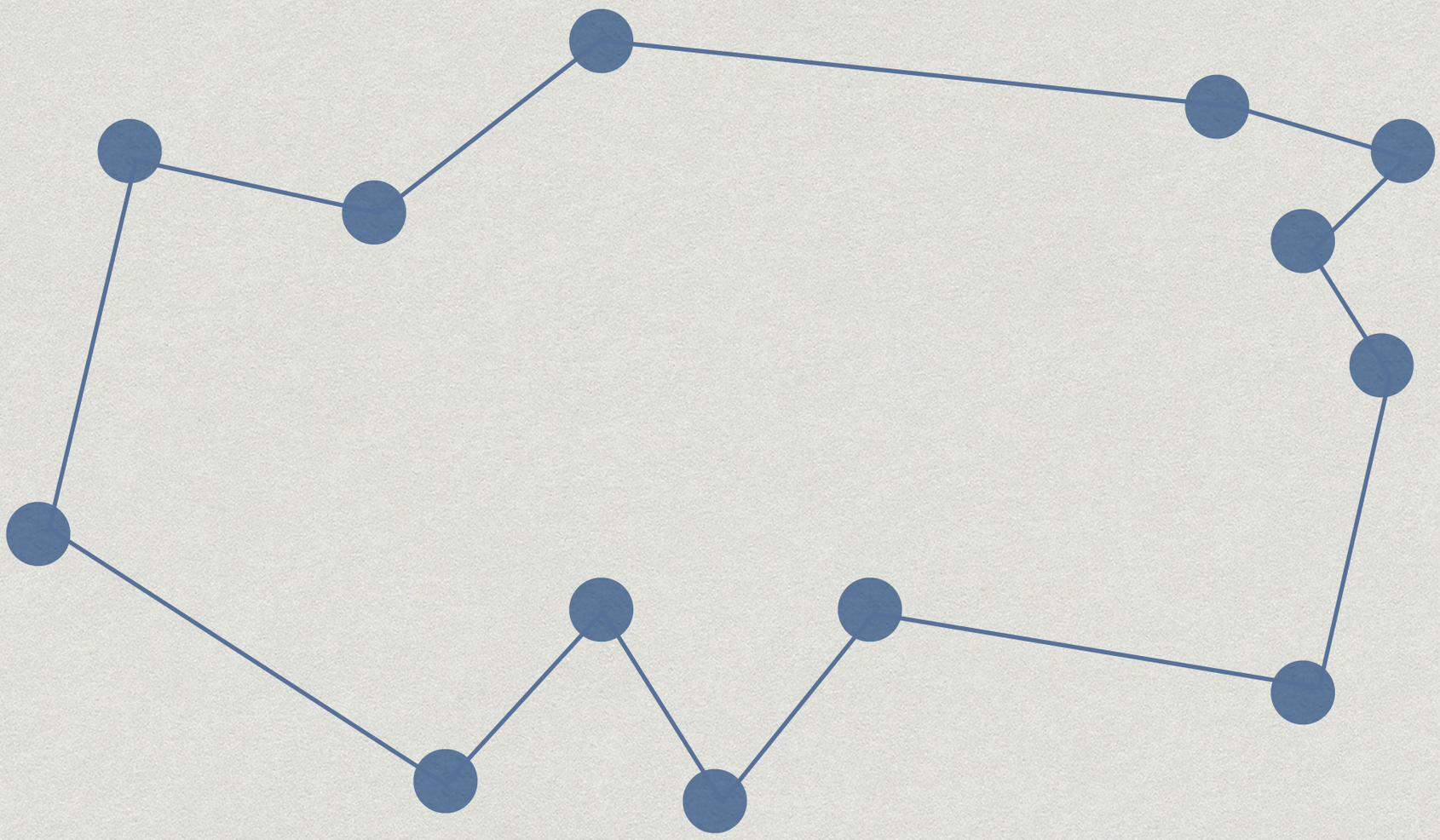
The Travelling Salesman Problem



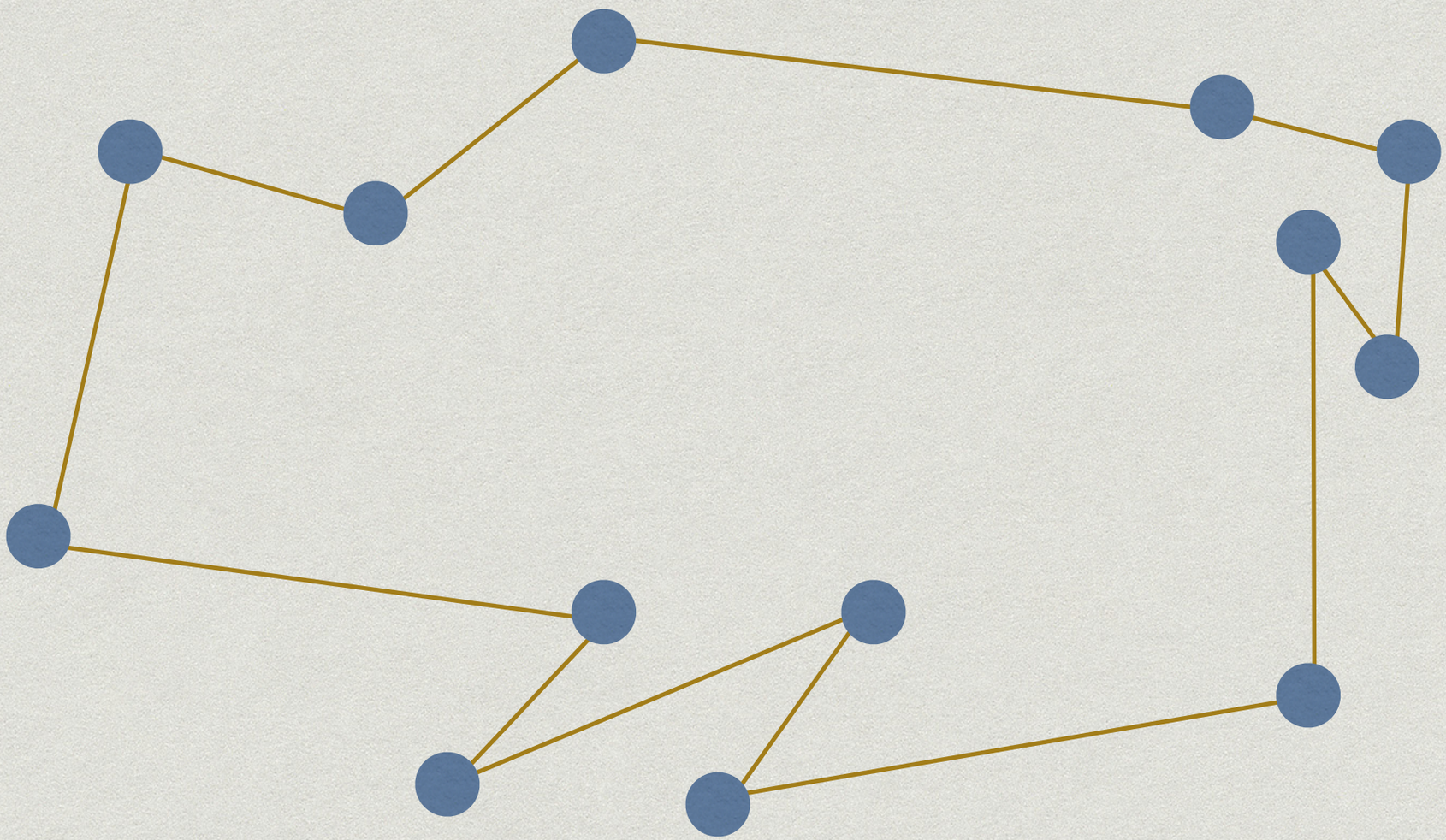
The Travelling Salesman Problem



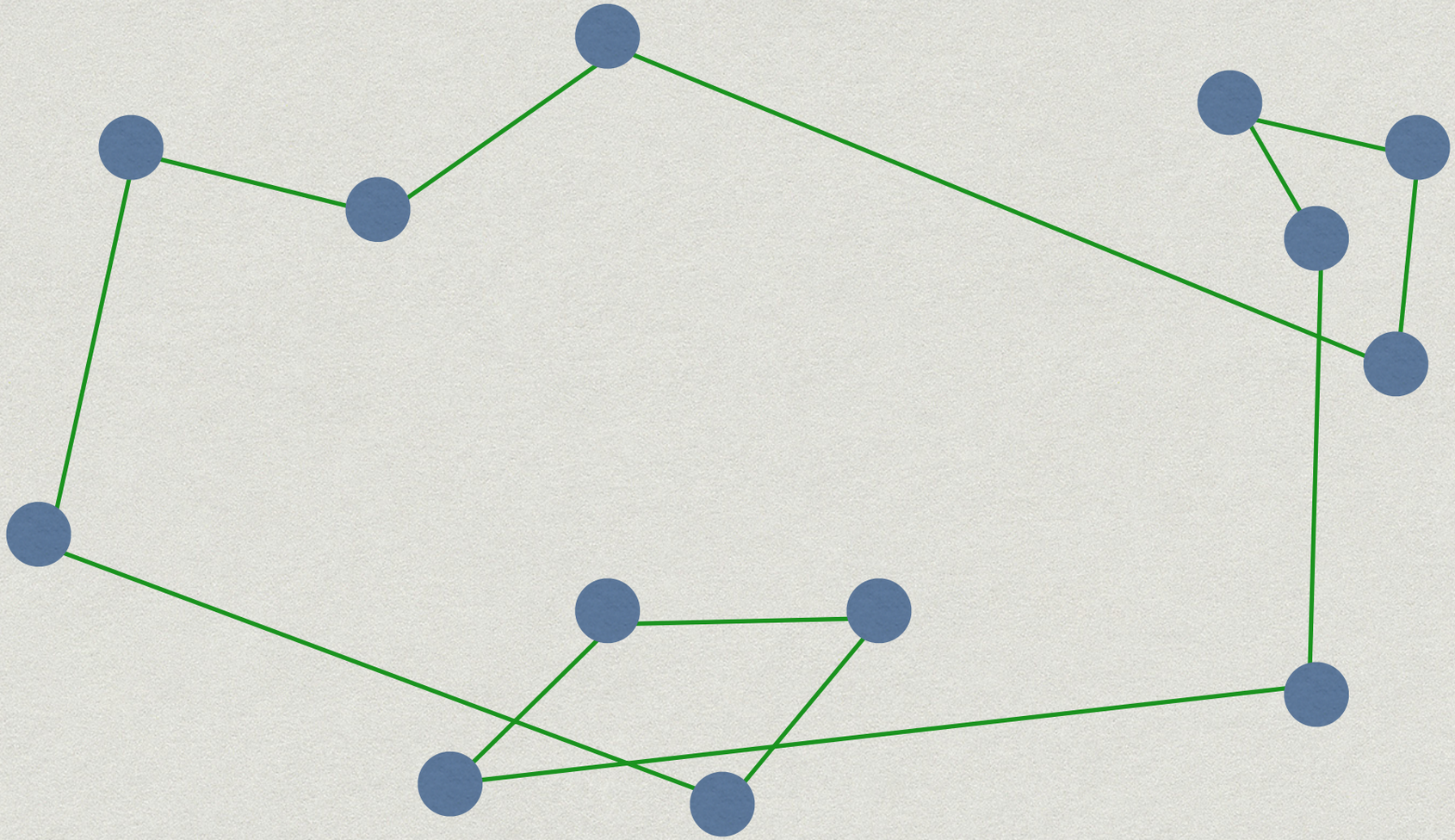
The Travelling Salesman Problem



The Travelling Salesman Problem



The Travelling Salesman Problem



P vs NP, Take 1

- Find an **efficient** (i.e. polynomial time) algorithm for TSP or prove that none exists

Time is of the essence

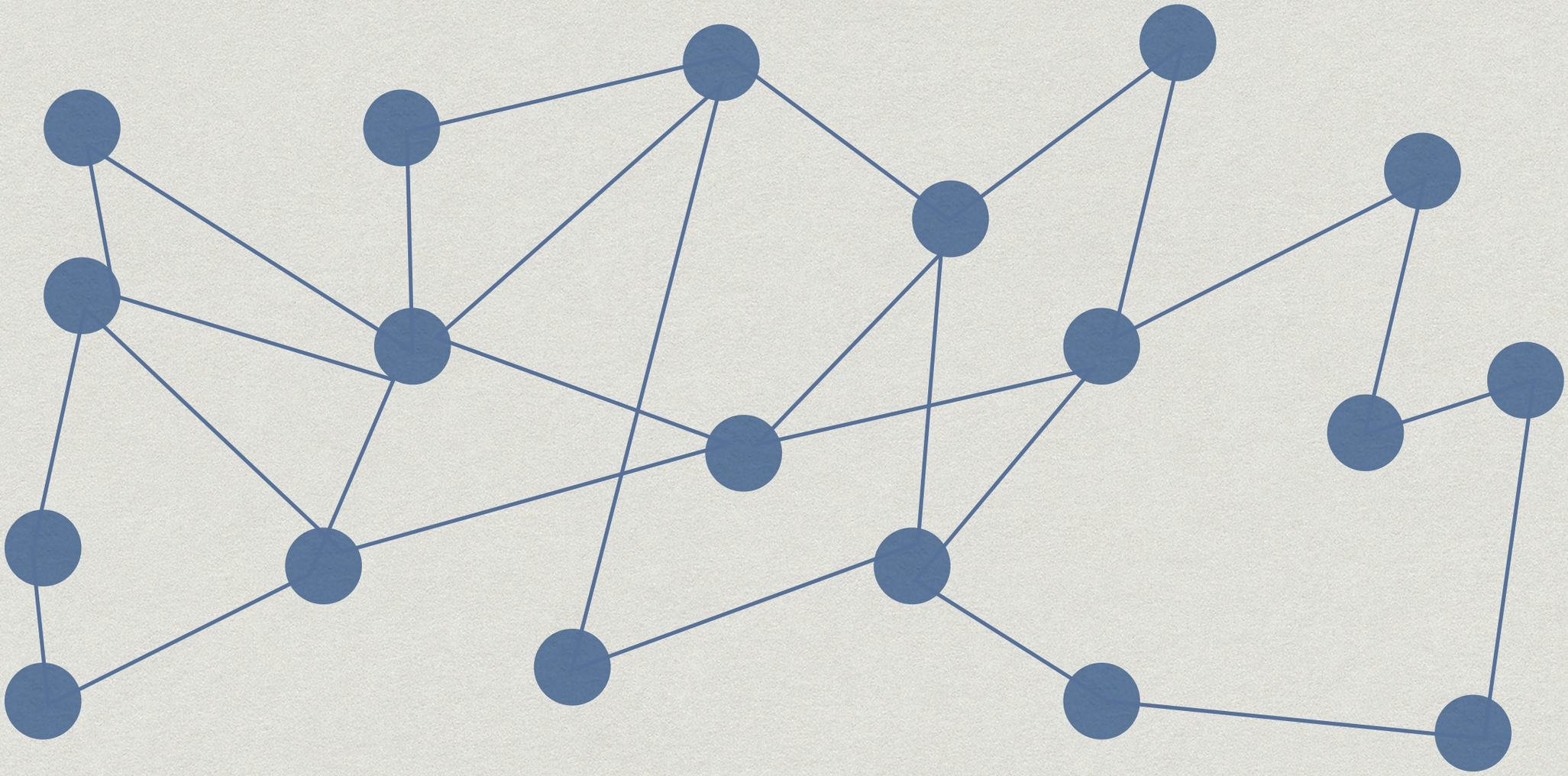
	N=50	N=100	N=1000	$N = 10^6$	$N = 10^9$
N	~0	~0	~0	0.001 sec	1 sec
N^2	~0	~0	0.001 sec	16 min	31 years
2^N	13 days	2871 U	$>10^{283}$ U	!	!
3^N	227,643 centuries	$> 10^{21}$ U	!	!	!
N!	$> 10^{46}$ U	$> 10^{140}$ U	!	!	!

computer speed = 1 billion operations per second
U is the Age of the Universe

Moral of the story

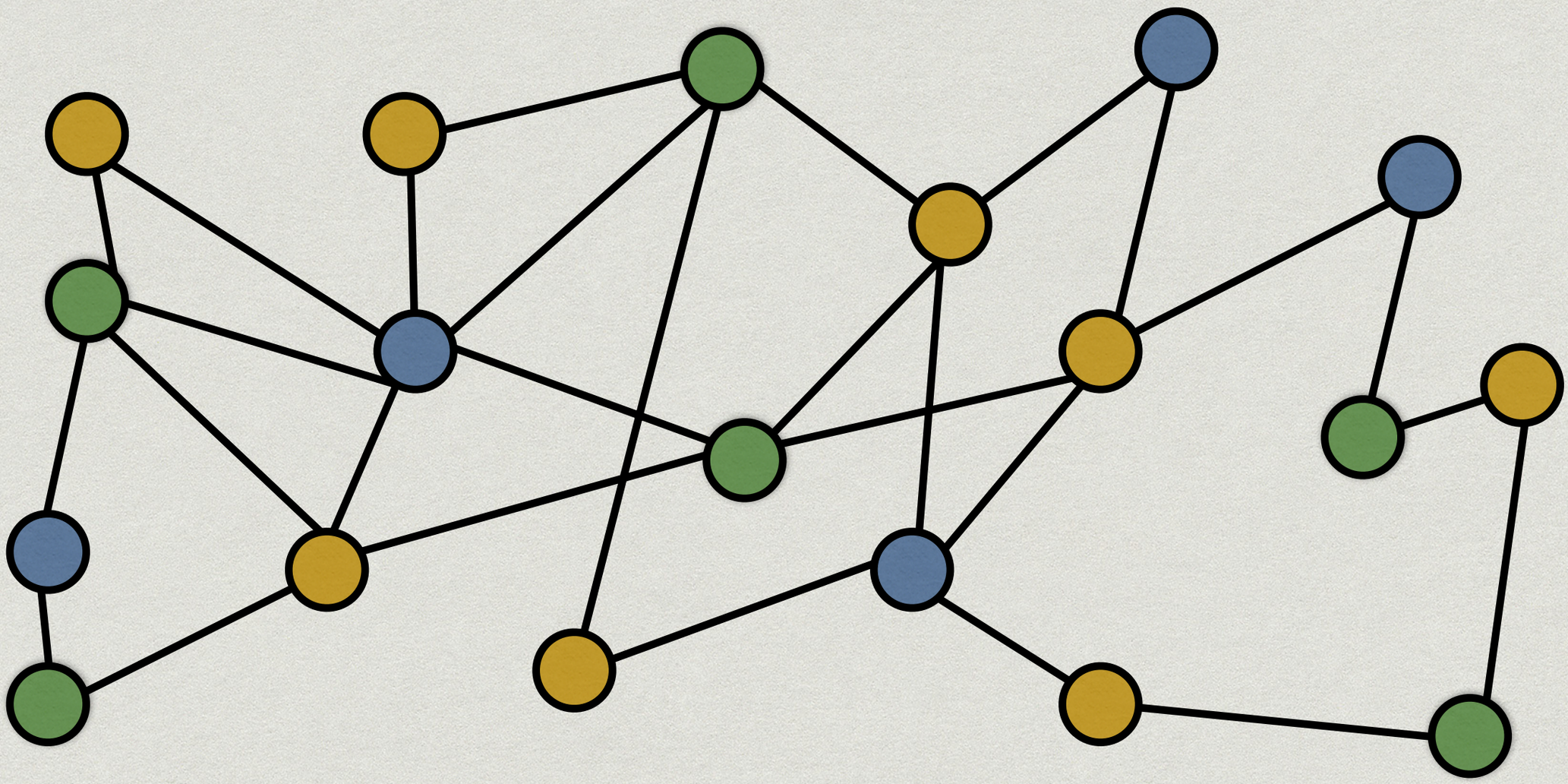
- No technological improvement will ever make an exponential-time algorithm practical

Colouring Graphs

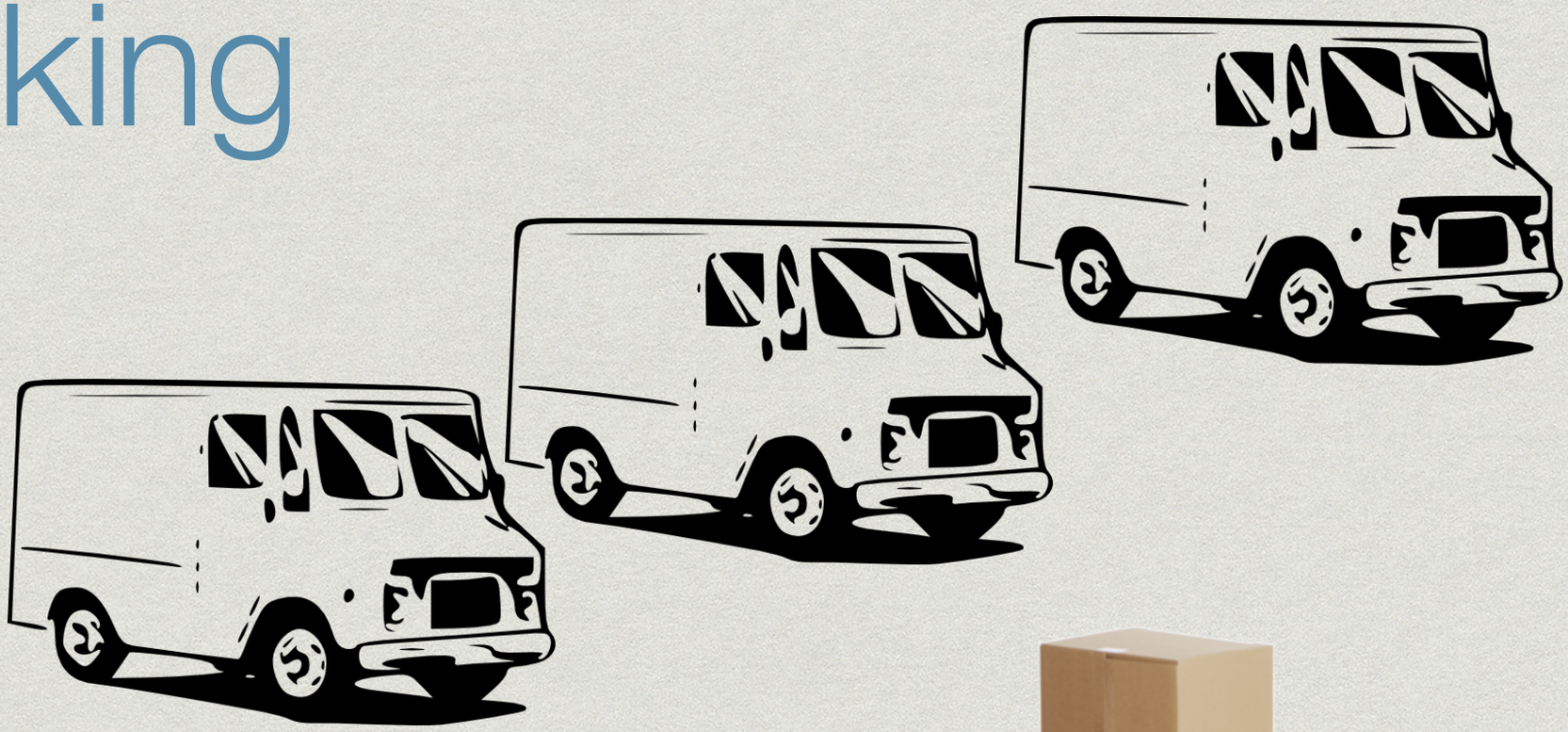


Can we colour this graph with three colours?

Colouring Graphs



Bin Packing



Partition

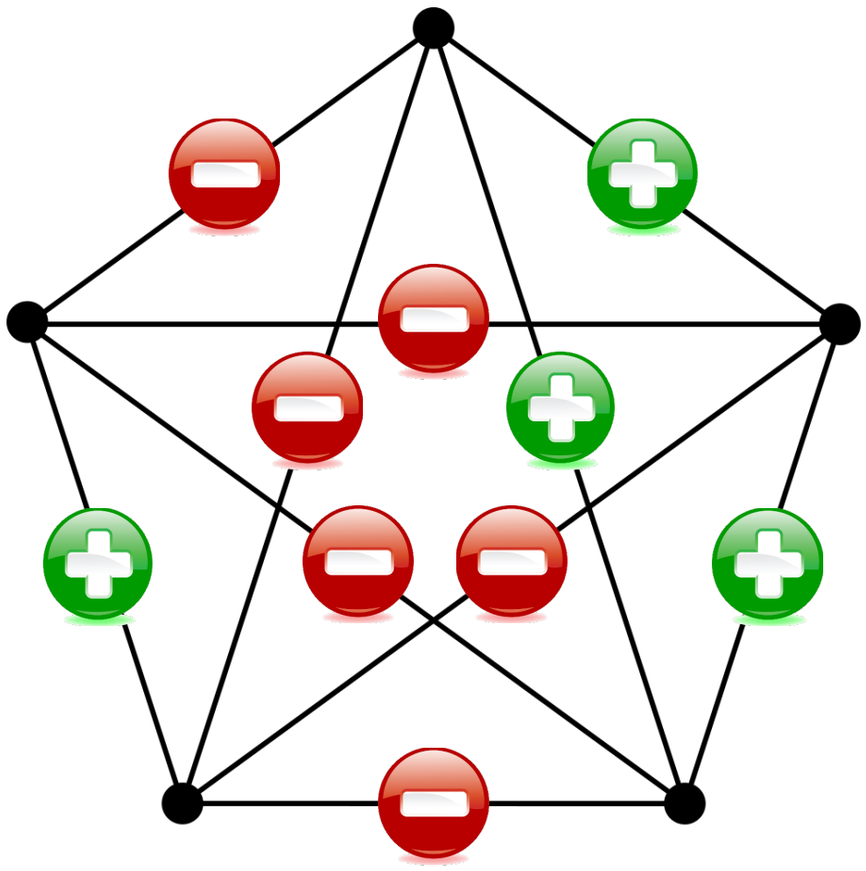
5 3 5 11 8 3 7 8 2 3 4 1 3 2 3 2

Partition

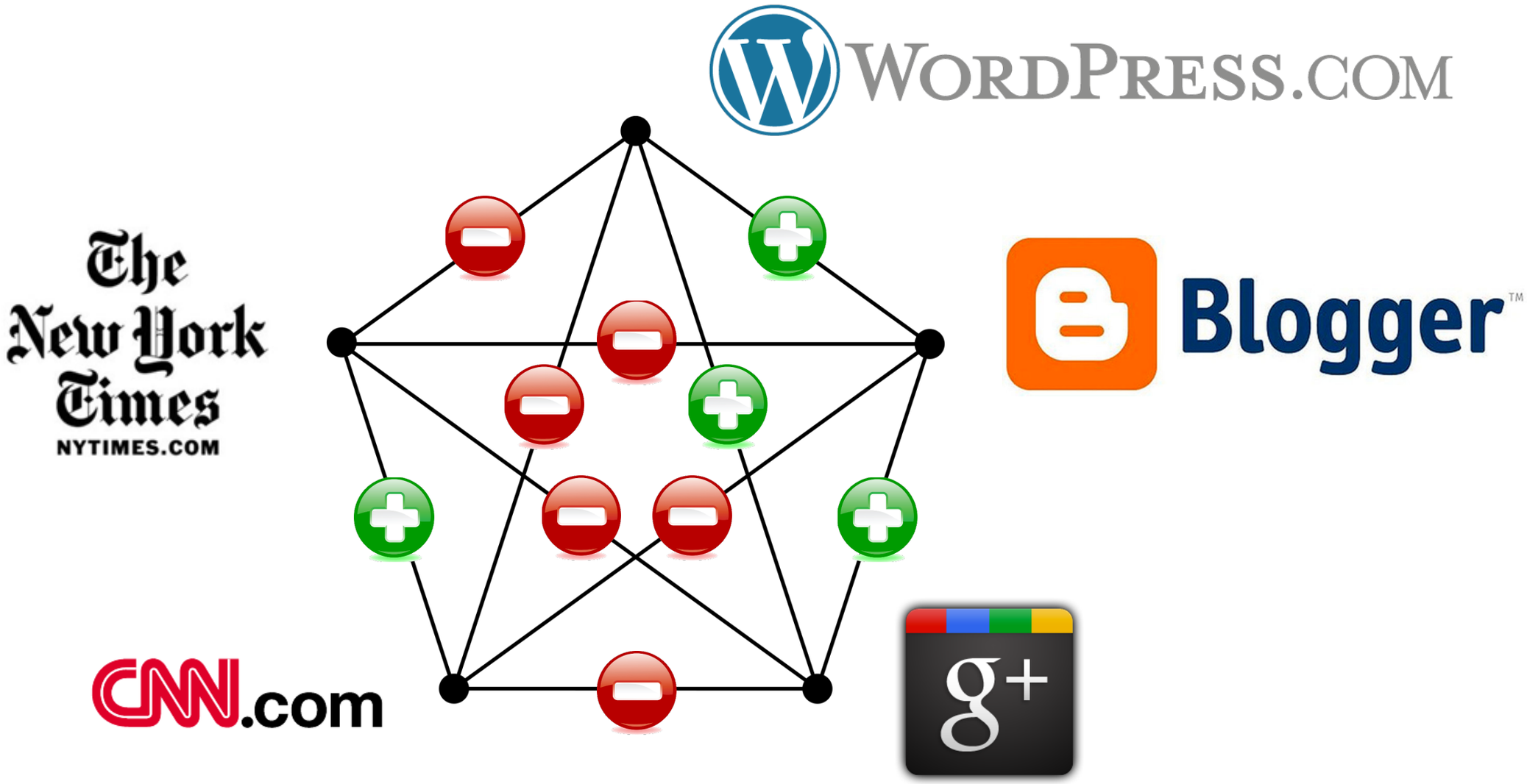
5 3 5 11 8 3 7 8 2 3 4 1 3 2 3 2

THE CLUSTERING ZOO

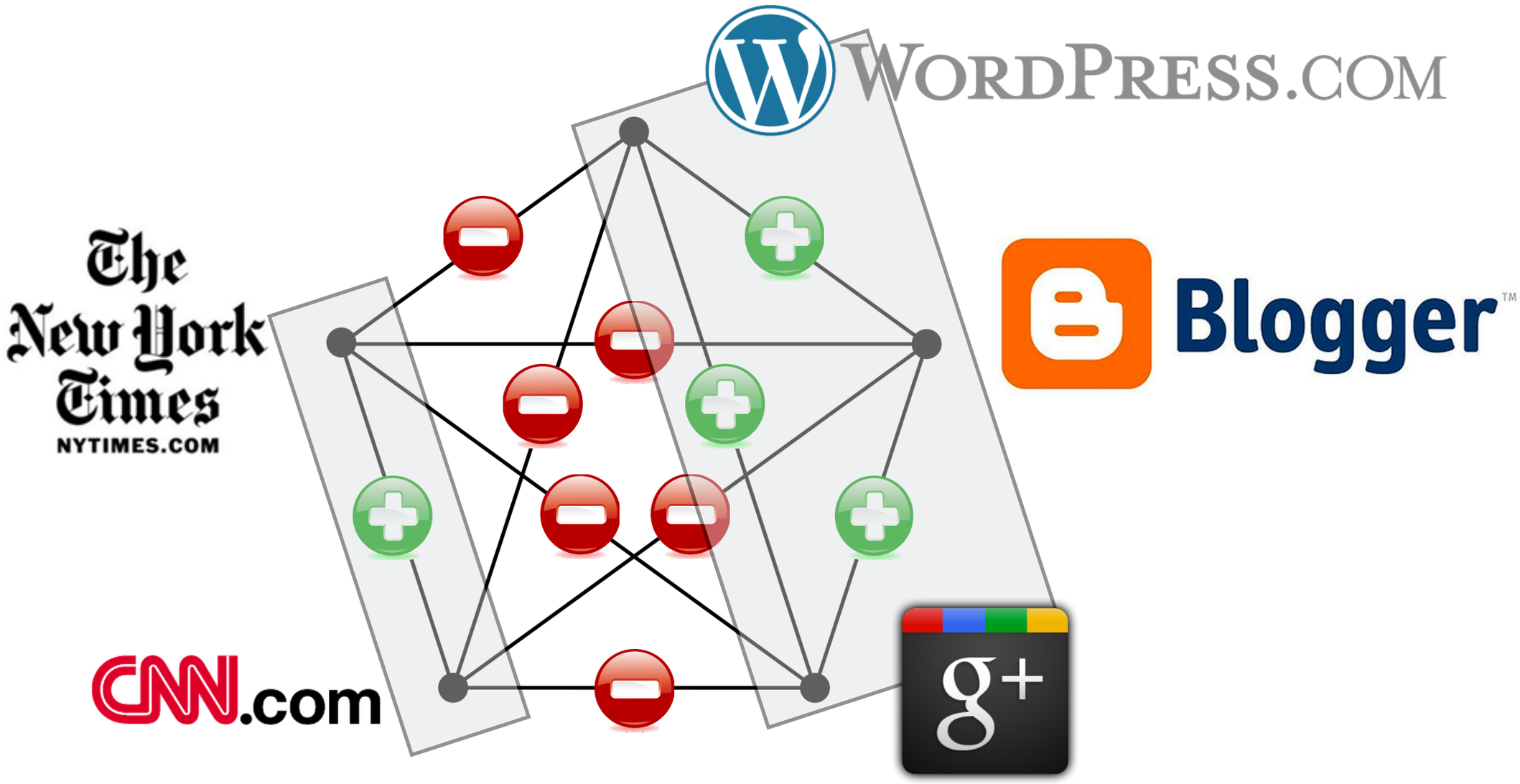
Correlation Clustering



Correlation Clustering



Correlation Clustering



Correlation Clustering

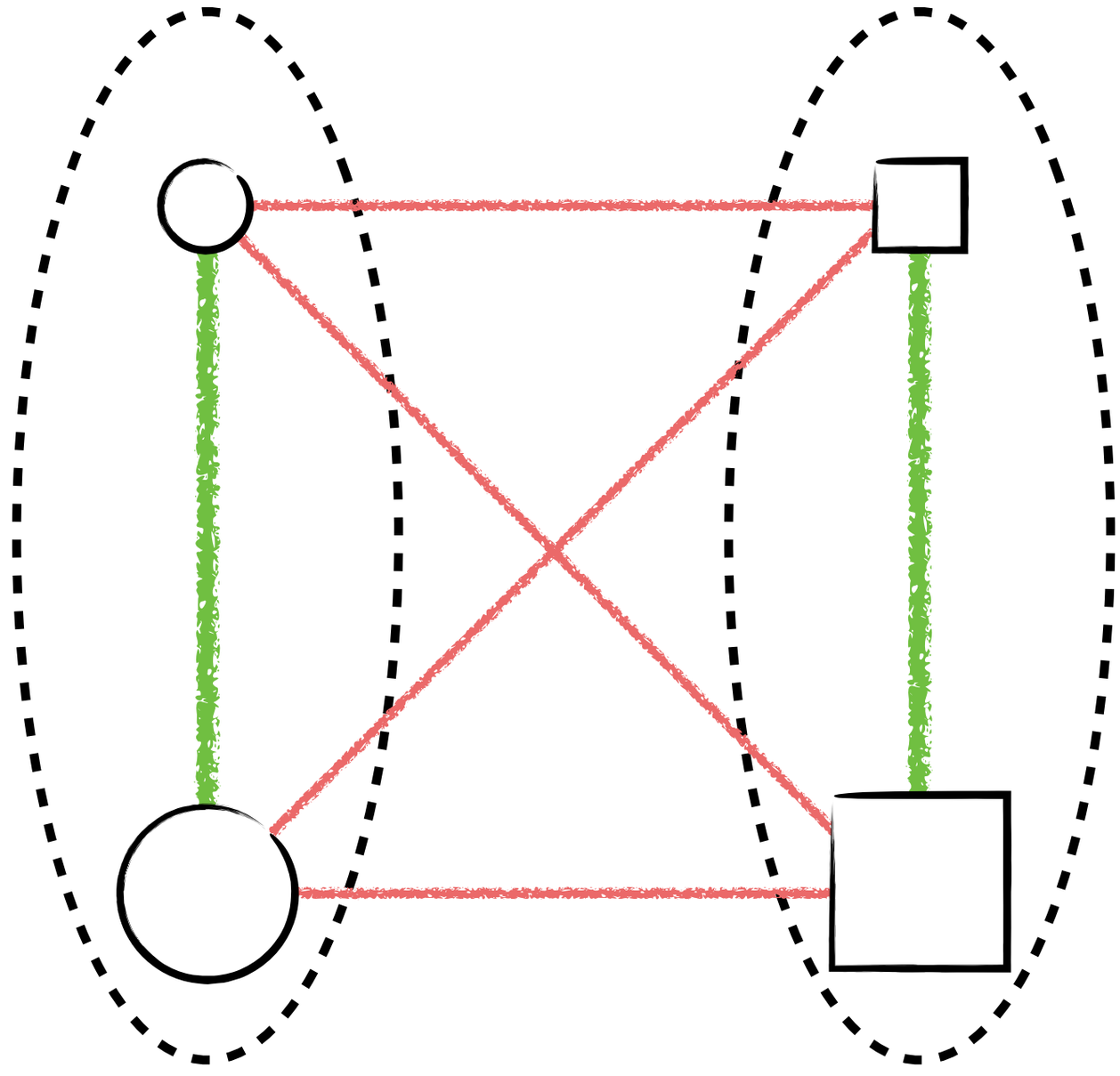
The goal is to partition the elements in
disjoint clusters
so as to **minimize** the number of **mistakes**

Correlation Clustering

The goal is to partition the elements in
disjoint clusters
so as to **minimize** the number of **mistakes**

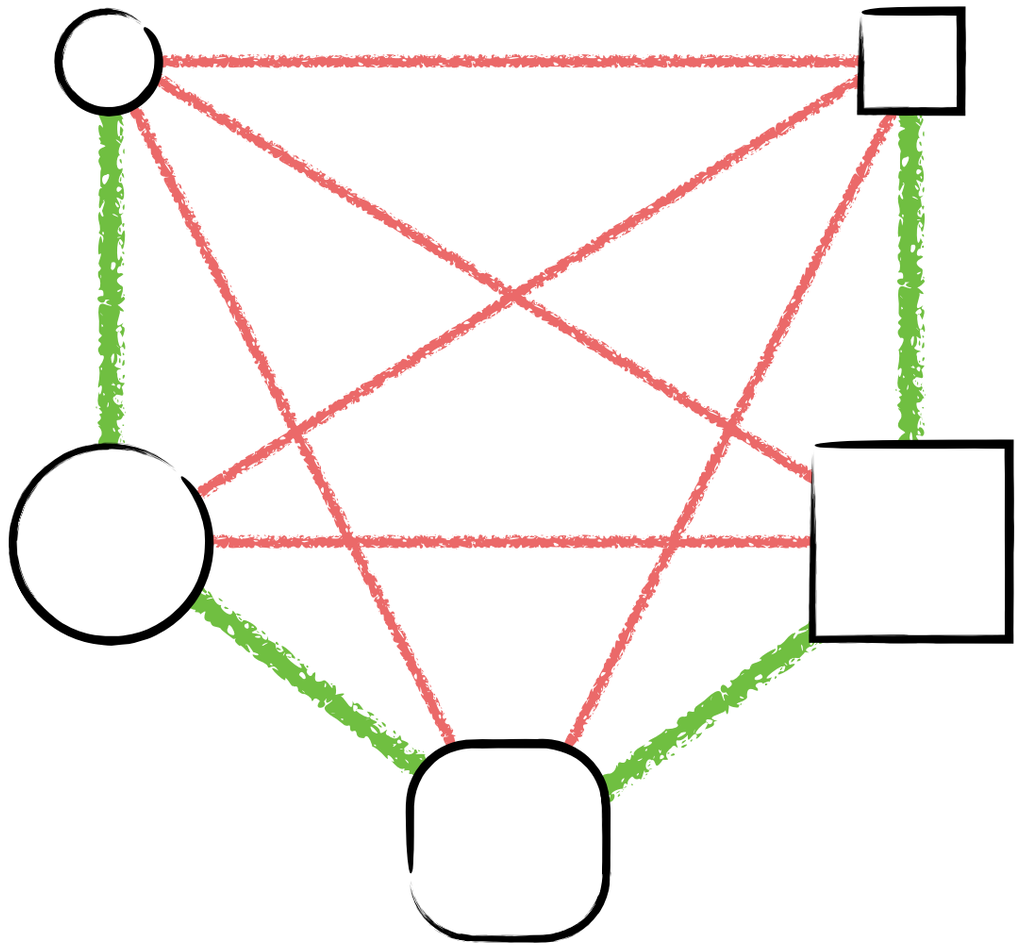
The **number of clusters**
is not decided a priori

Correlation Clustering



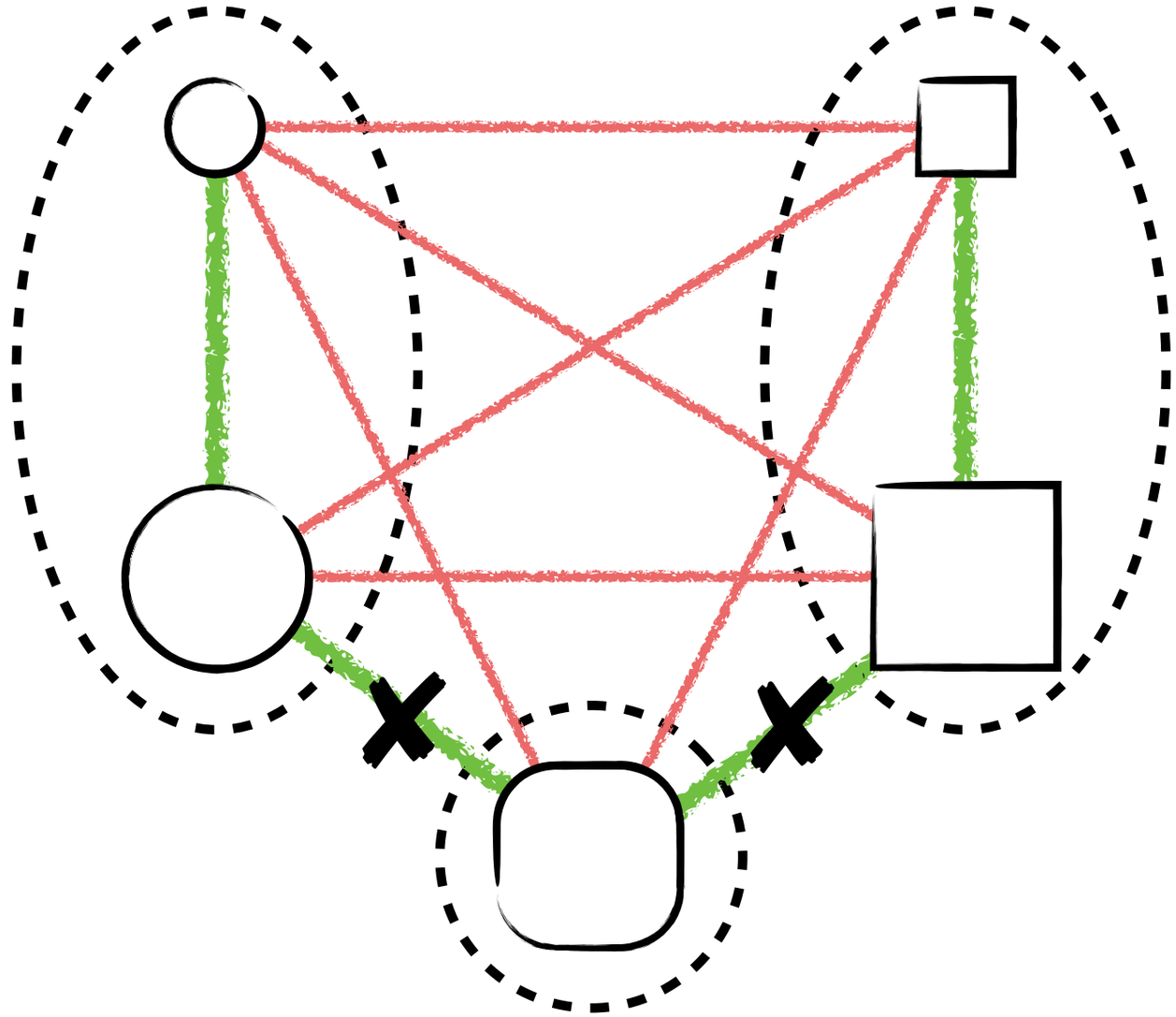
"Perfect clustering"
(0 mistakes)

Correlation Clustering



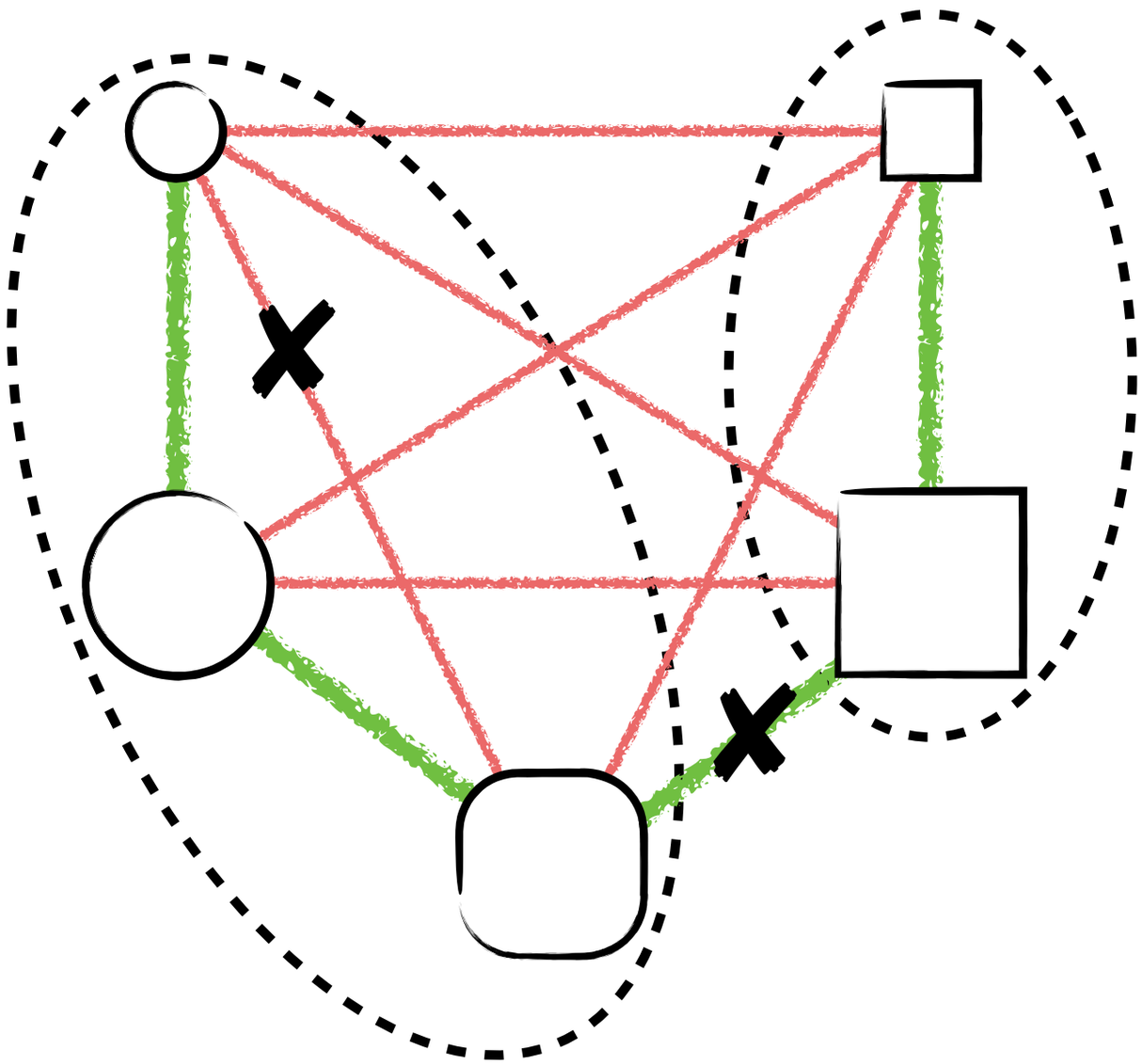
In this instance you cannot have zero mistakes

Correlation Clustering



In this instance you cannot have zero mistakes

Correlation Clustering



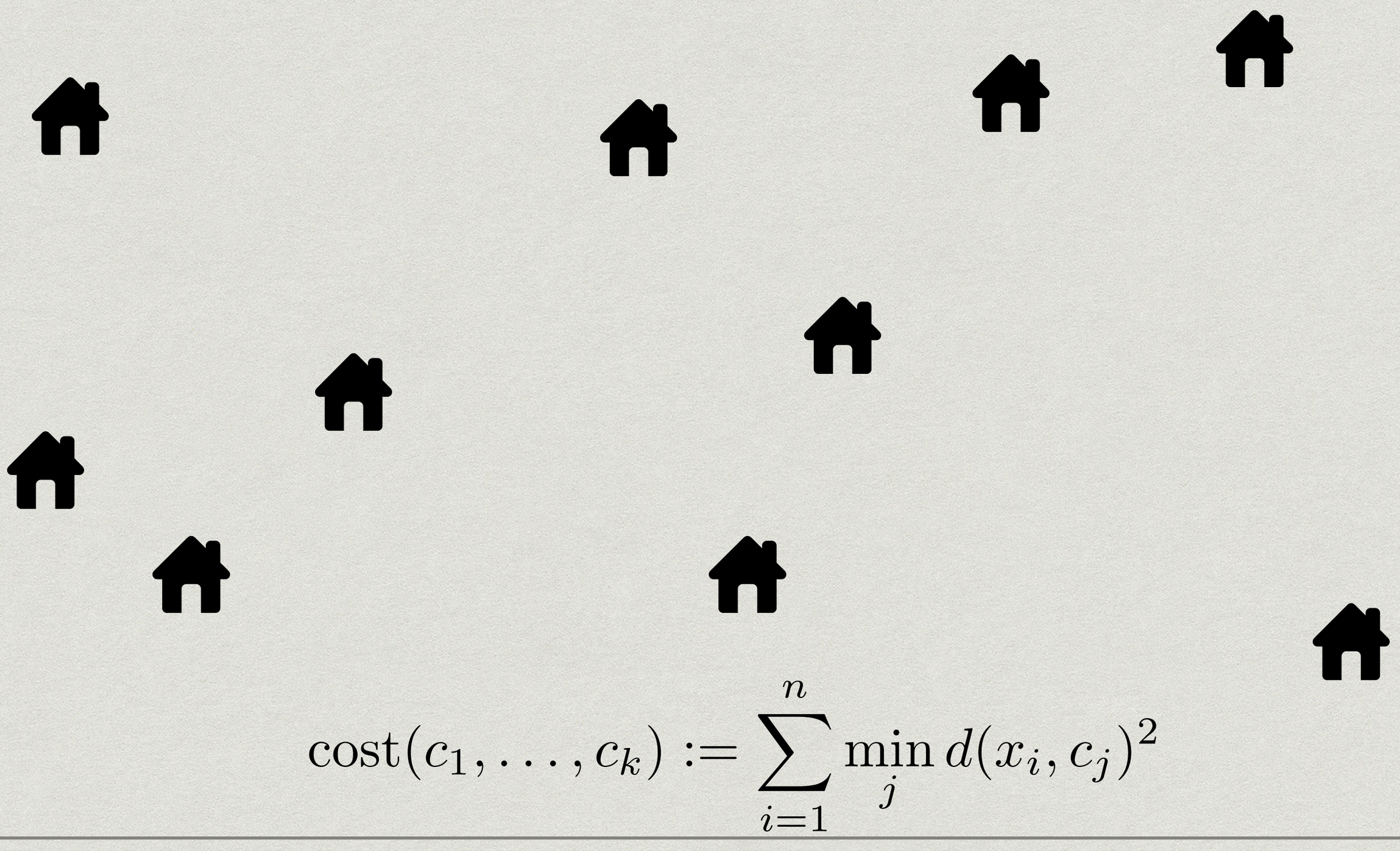
In this instance you cannot have zero mistakes

k-means Clustering

- We are given a set S of n points in R^d
- Assume k is also given
- Find k points $c_1,..,c_k$, called centers (or means) such that the following cost is minimised

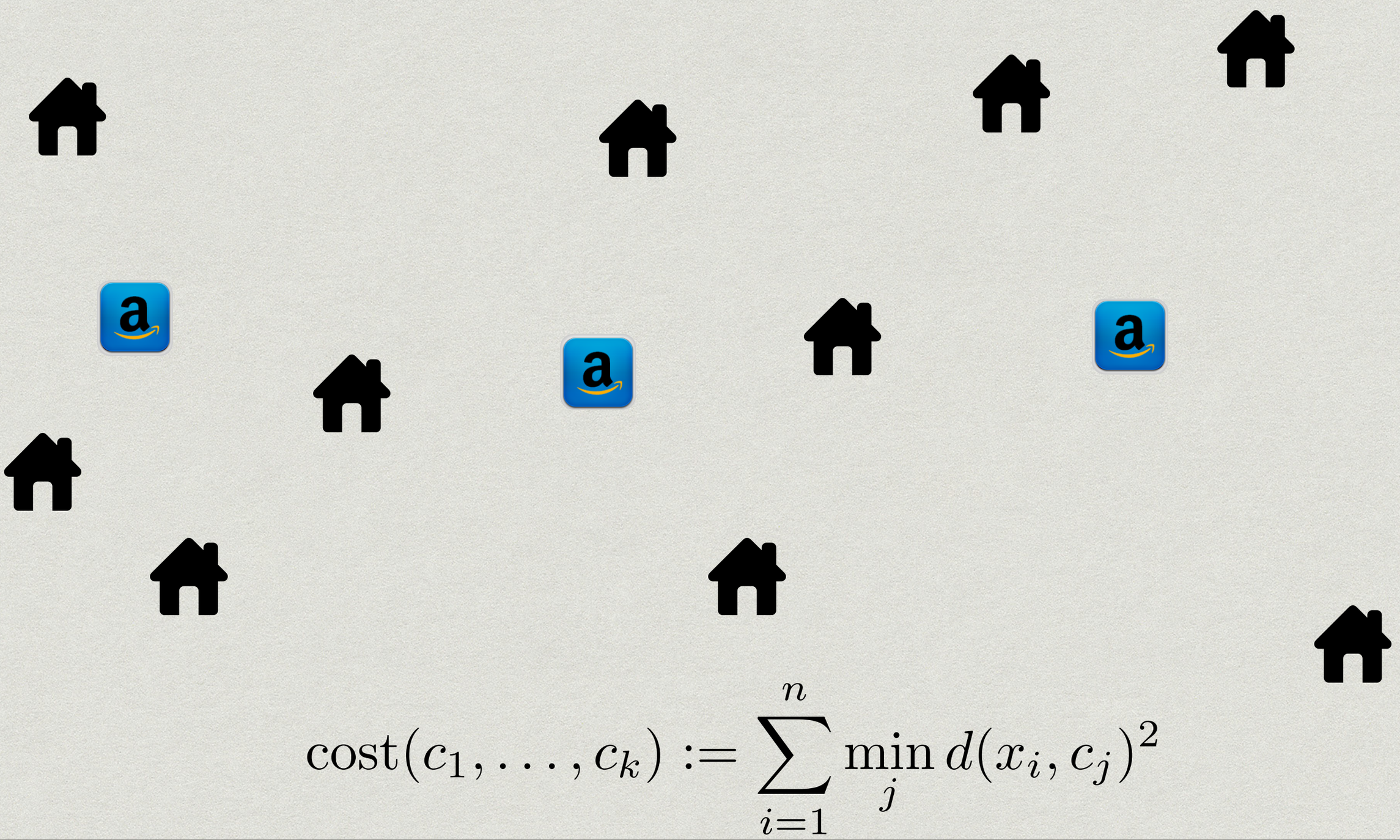
$$\text{cost}(c_1, \ldots, c_k) := \sum_{i=1}^{n} \min_j d(x_i, c_j)^2$$

k-means Clustering



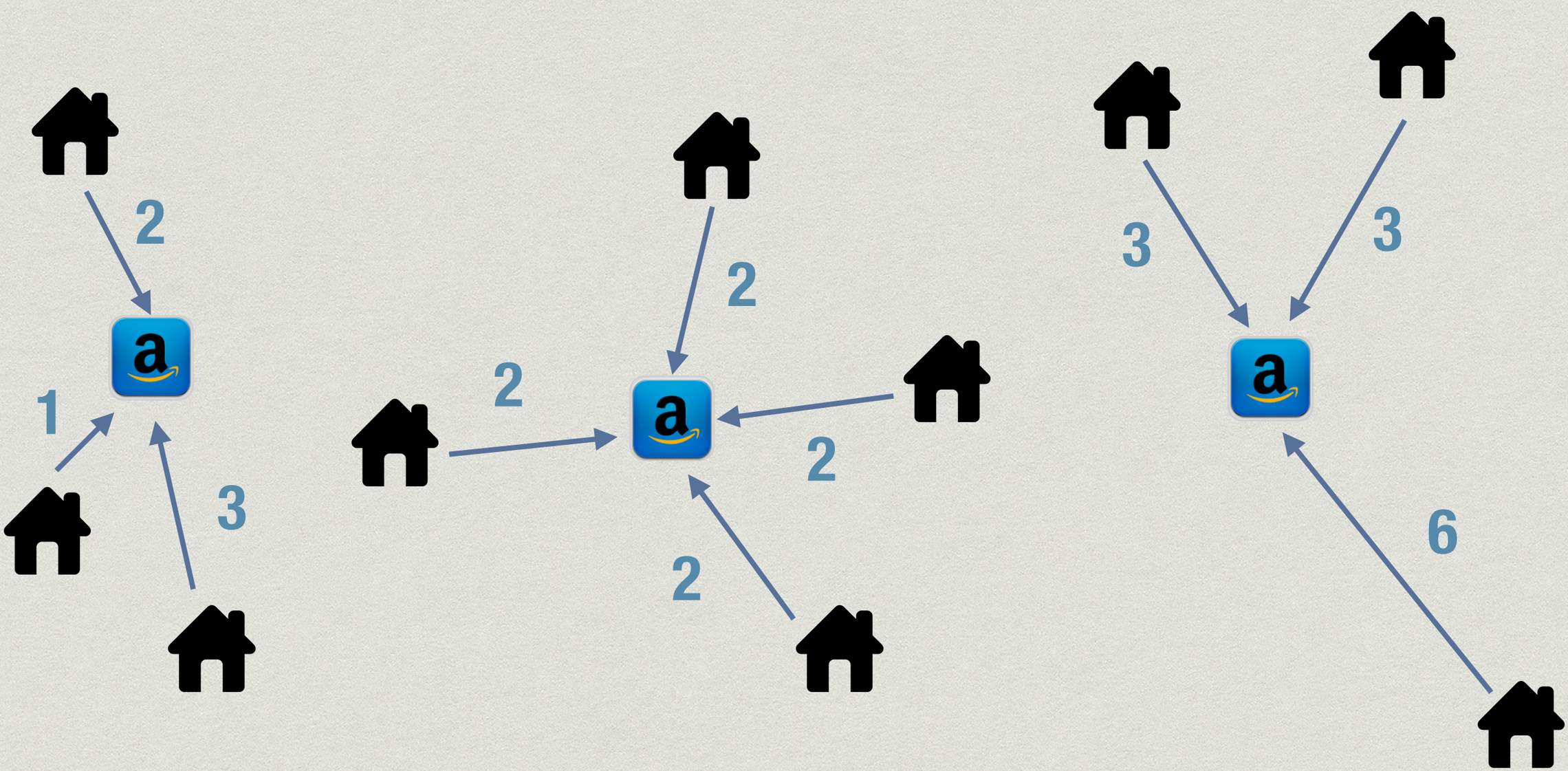
$$\text{cost}(c_1, \ldots, c_k) := \sum_{i=1}^{n} \min_j d(x_i, c_j)^2$$

k-means Clustering



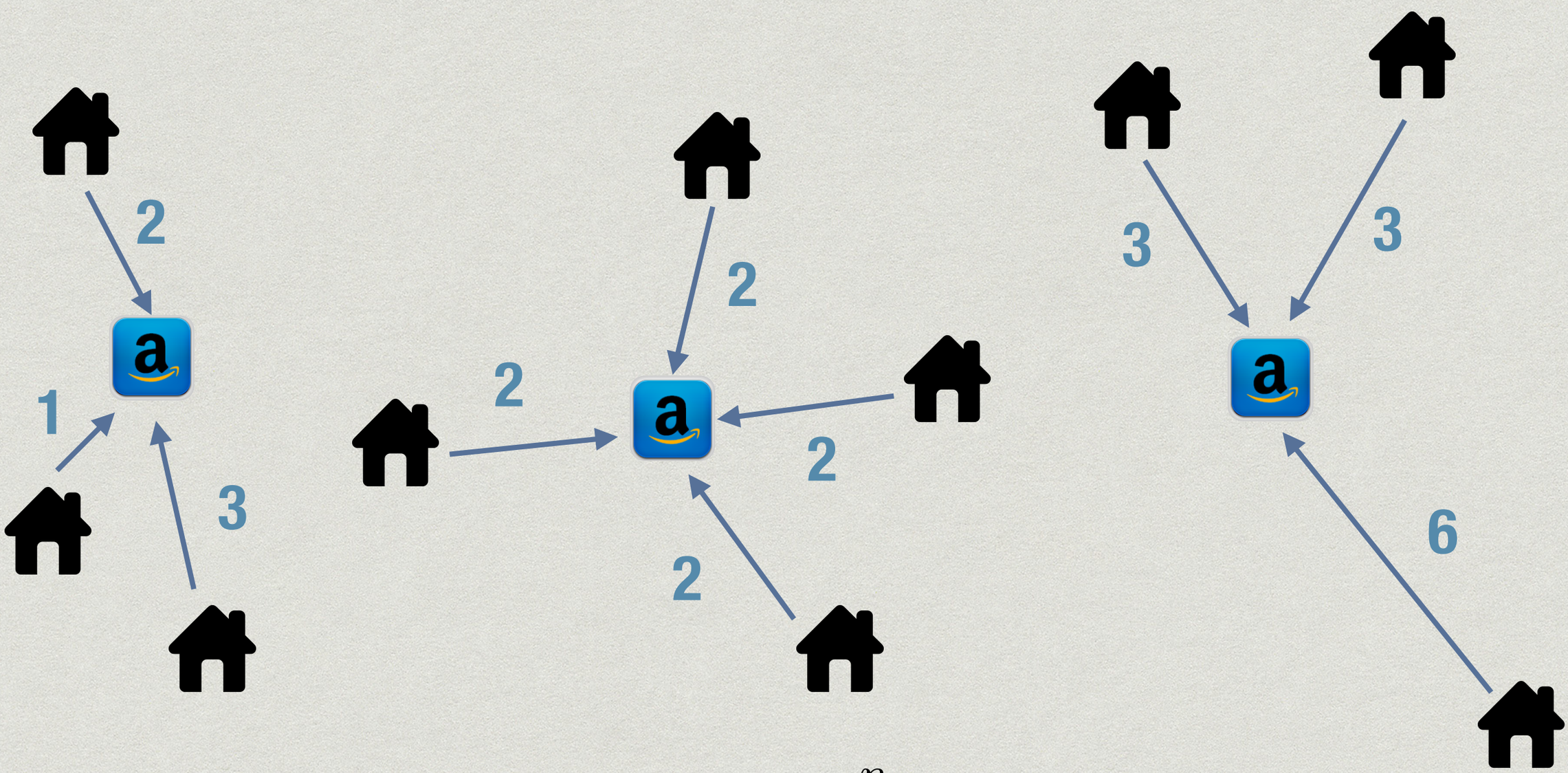
$$\text{cost}(c_1, \ldots, c_k) := \sum_{i=1}^{n} \min_j d(x_i, c_j)^2$$

k-means Clustering



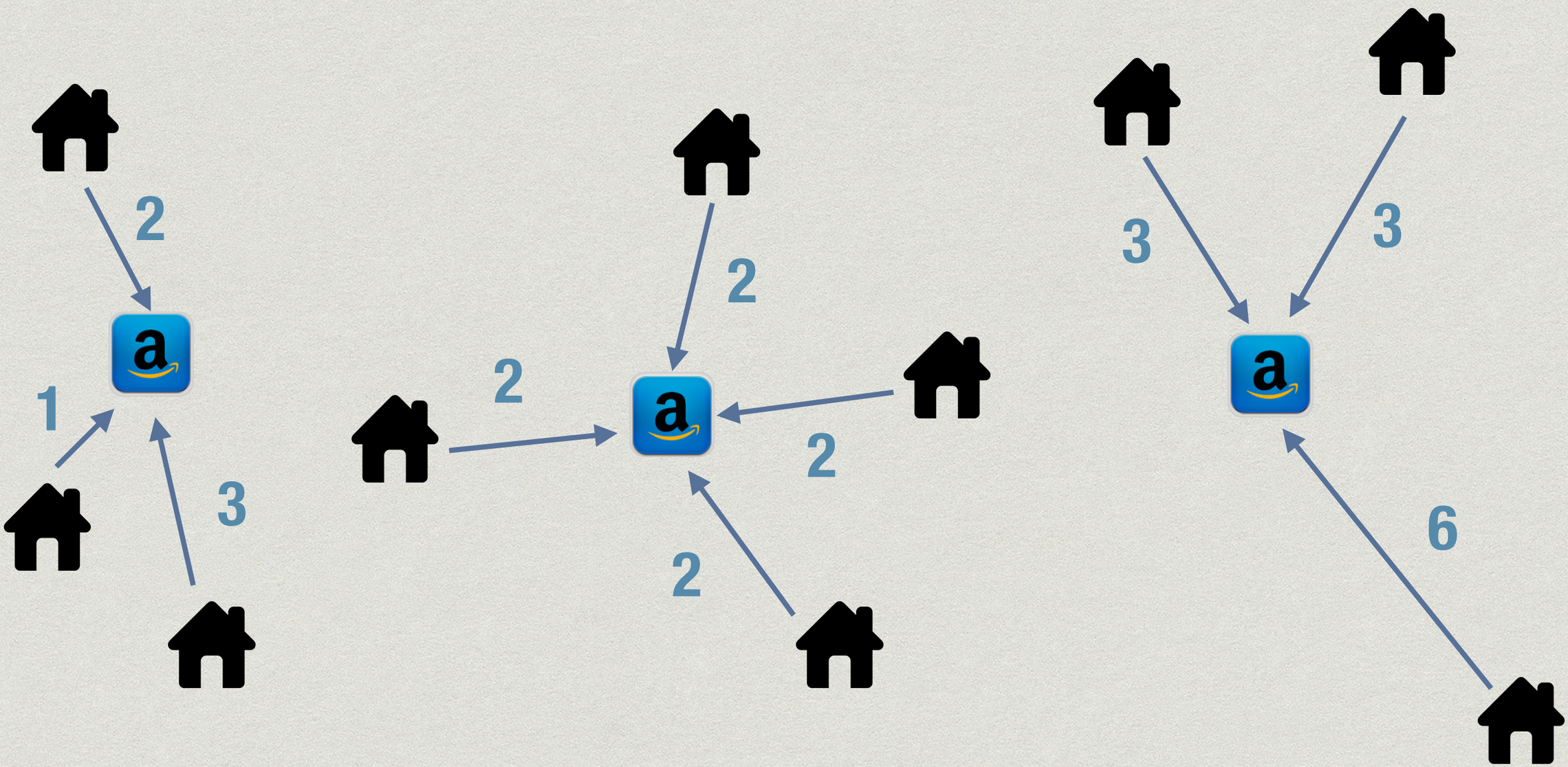
$$cost(c_1, c_2, c_3) = 84$$

k-median Clustering



$$\text{cost}(c_1, \ldots, c_k) := \sum_{i=1}^{n} \min_j d(x_i, c_j)$$

k-median Clustering



$$cost(c_1, c_2, c_3) = 26$$

k-median Clustering

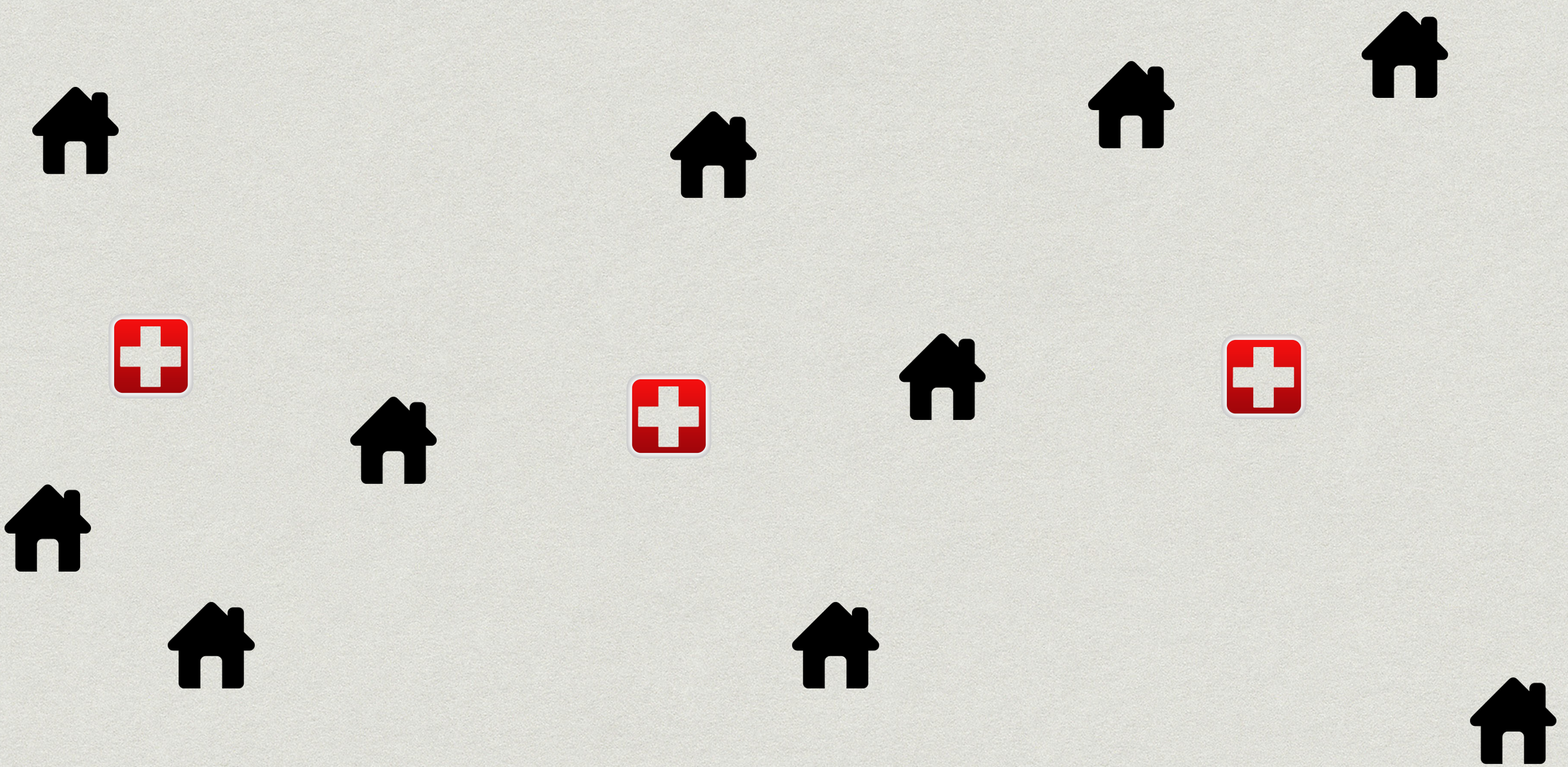
- Same input as in k-means clustering, but objective function becomes

$$\text{cost}(c_1, \ldots, c_k) := \sum_{i=1}^{n} \min_j d(x_i, c_j)$$

k-center Clustering

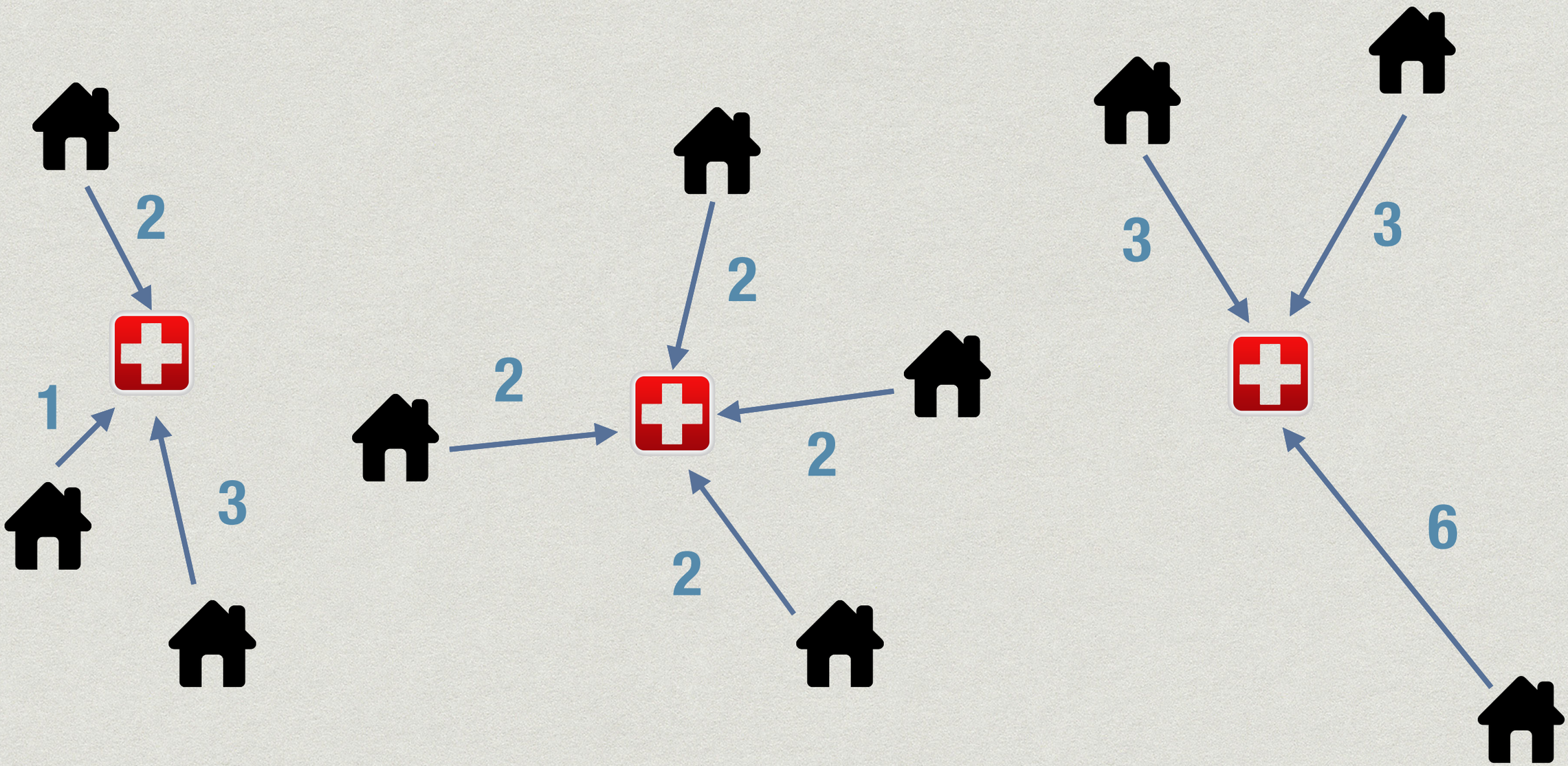
$$\text{cost}(c_1, \ldots, c_k) := \max_{x \in S} \min_{j} d(x, c_i)$$

k-center Clustering



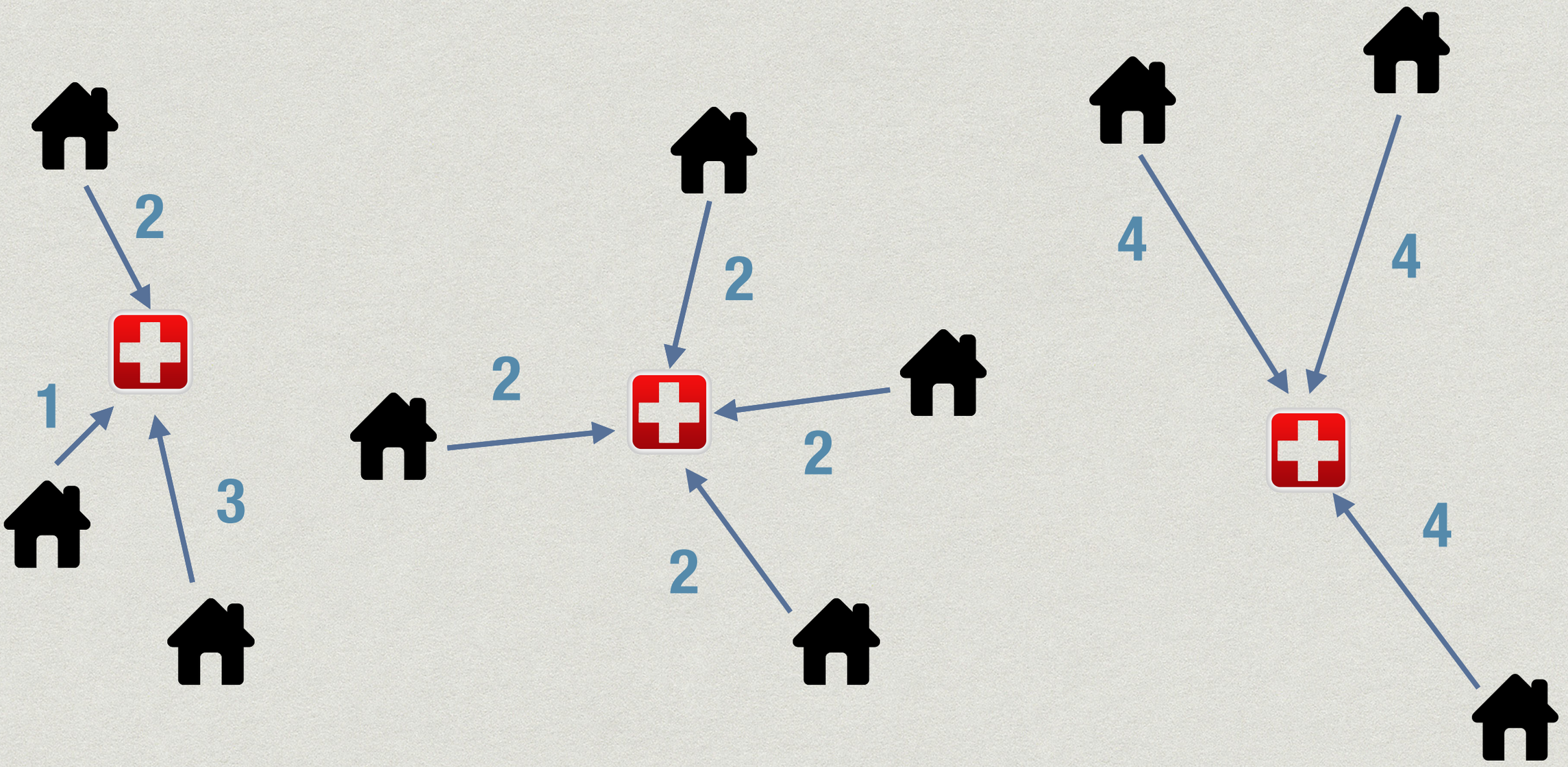
$$\text{cost}(c_1, \ldots, c_k) := \max_{x \in S} \min_{j} d(x, c_i)$$

k-center Clustering



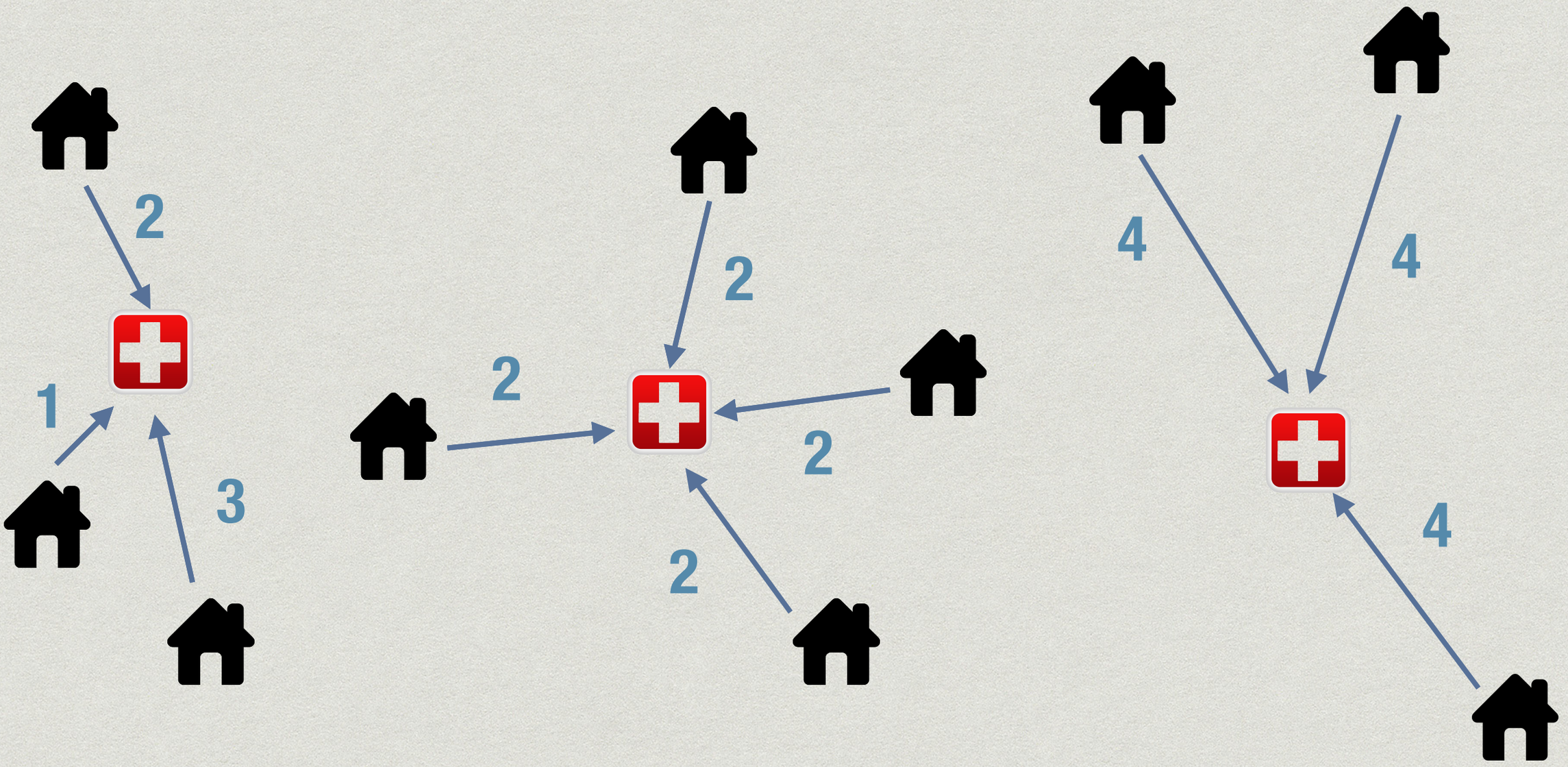
$$\text{cost}(c_1, \ldots, c_k) := \max_{x \in S} \min_{j} d(x, c_i)$$

k-center Clustering



$$\text{cost}(c_1, \ldots, c_k) := \max_{x \in S} \min_{j} d(x, c_i)$$

k-center Clustering



$$\text{cost}(c_1, \ldots, c_k) := \max_{x \in S} \min_{j} d(x, c_i)$$

k-center Clustering

- We want to build k facilities (say hospitals) in such a way that no point in S is too penalised, i.e. we want to minimise the maximum distance between every point and its closest center

$$\text{cost}(c_1, \ldots, c_k) := \max_{x \in S} \min_{j} d(x, c_i)$$

P vs NP

- All of these problems, and many, many more, are **polynomial-time equivalent**
- For all of them, no efficient algorithms is known..
- …and it is conjectured that none exists

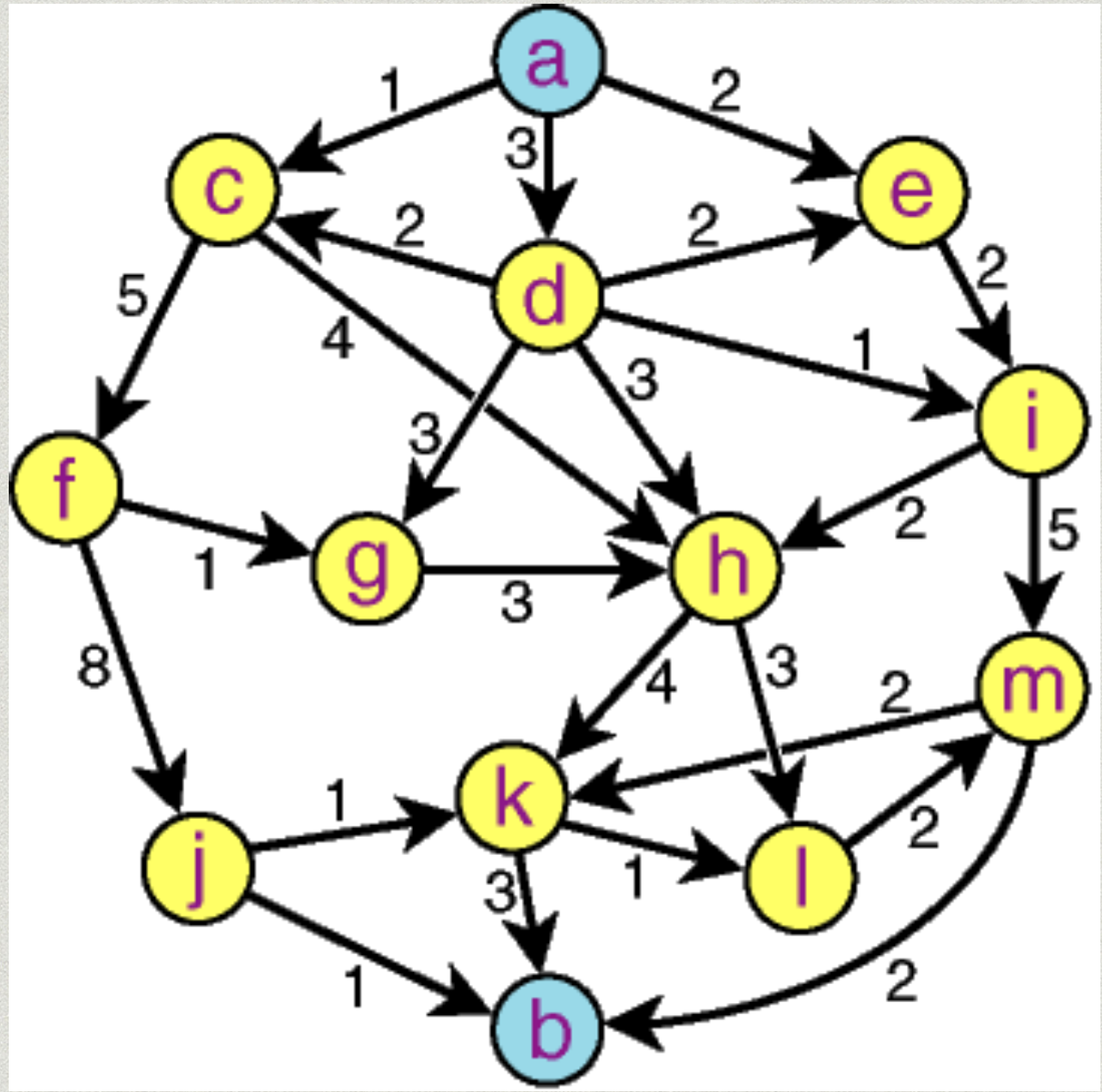
NP-hardness (informal)

- An optimisation problem is **NP-hard** if it is polynomial-time equivalent to the ones we have seen

PROBLEMS IN P

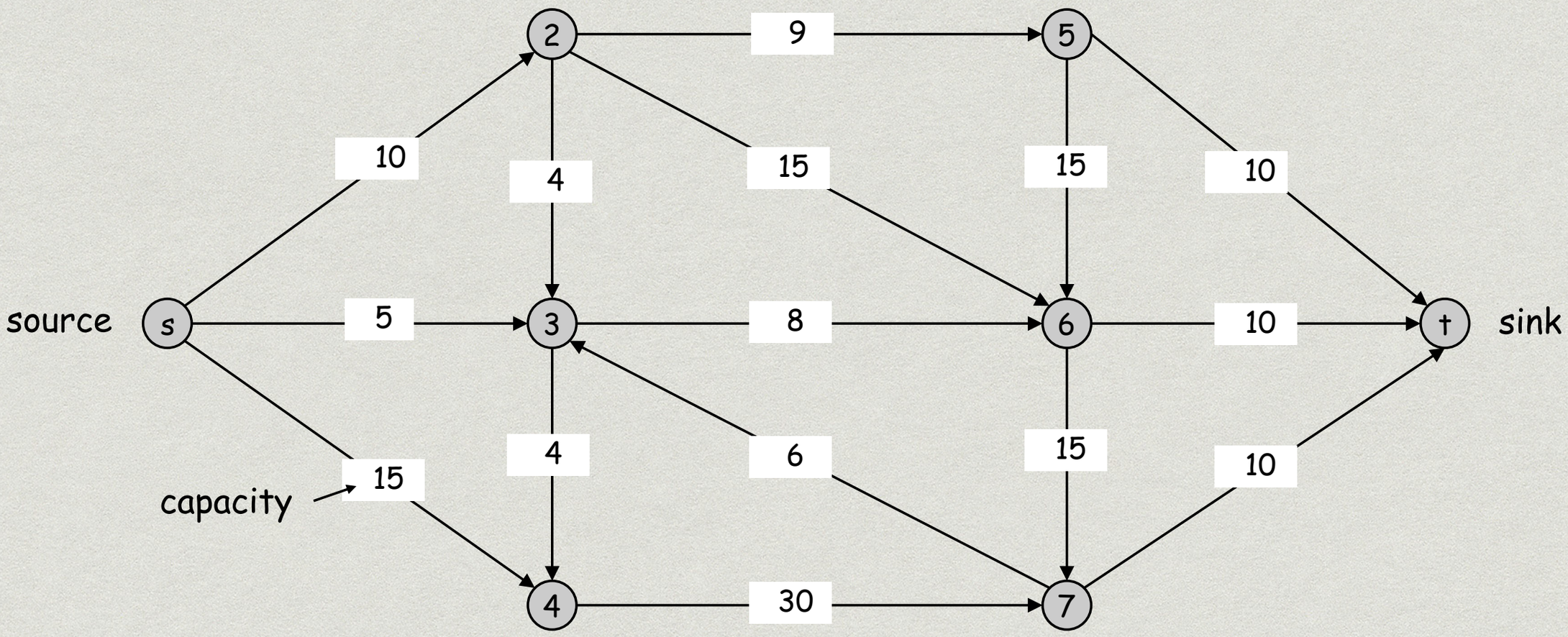
THE IMPORTANCE OF BEING POLYNOMIAL

Shortest Path

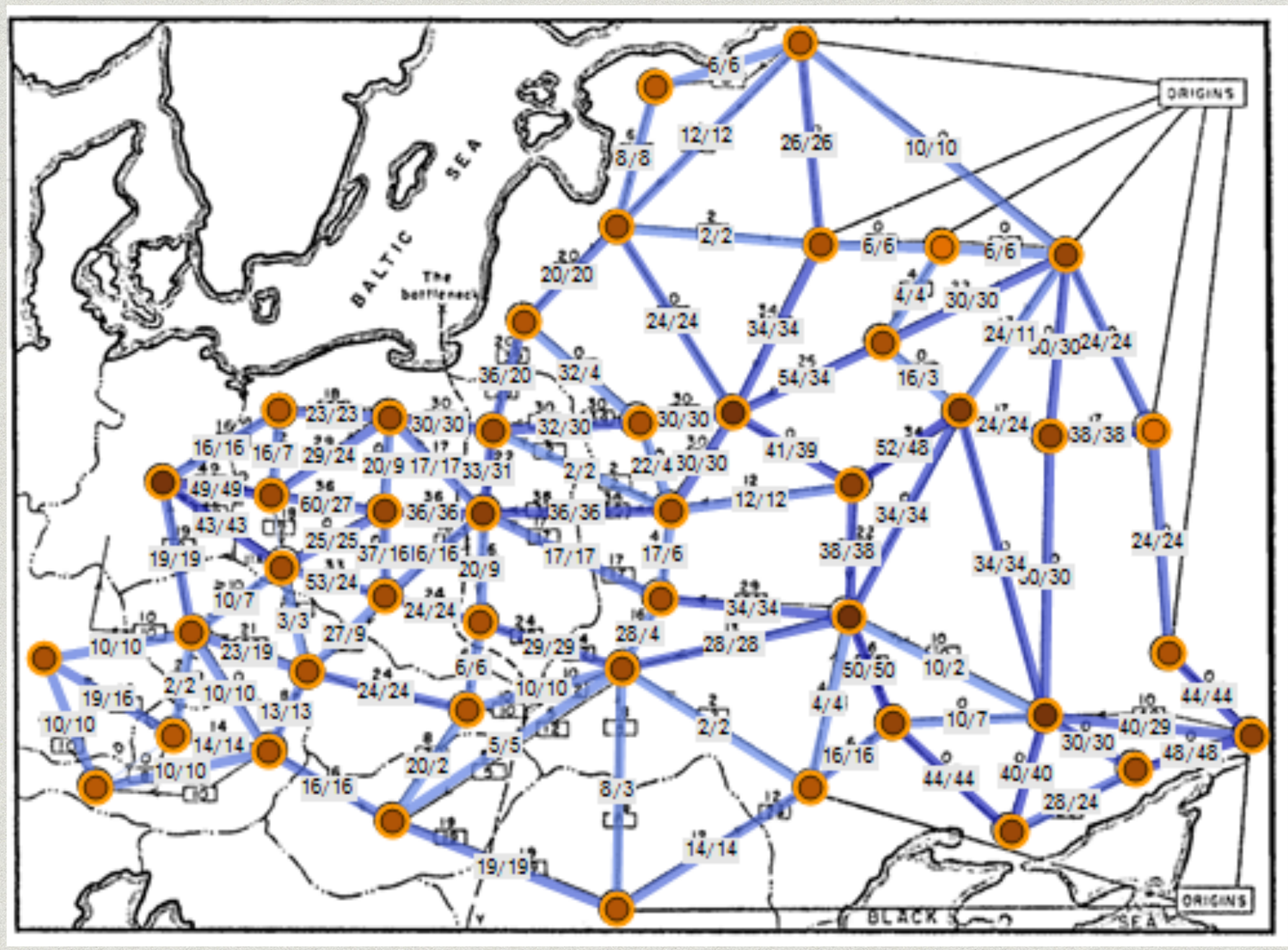


Shortest Path

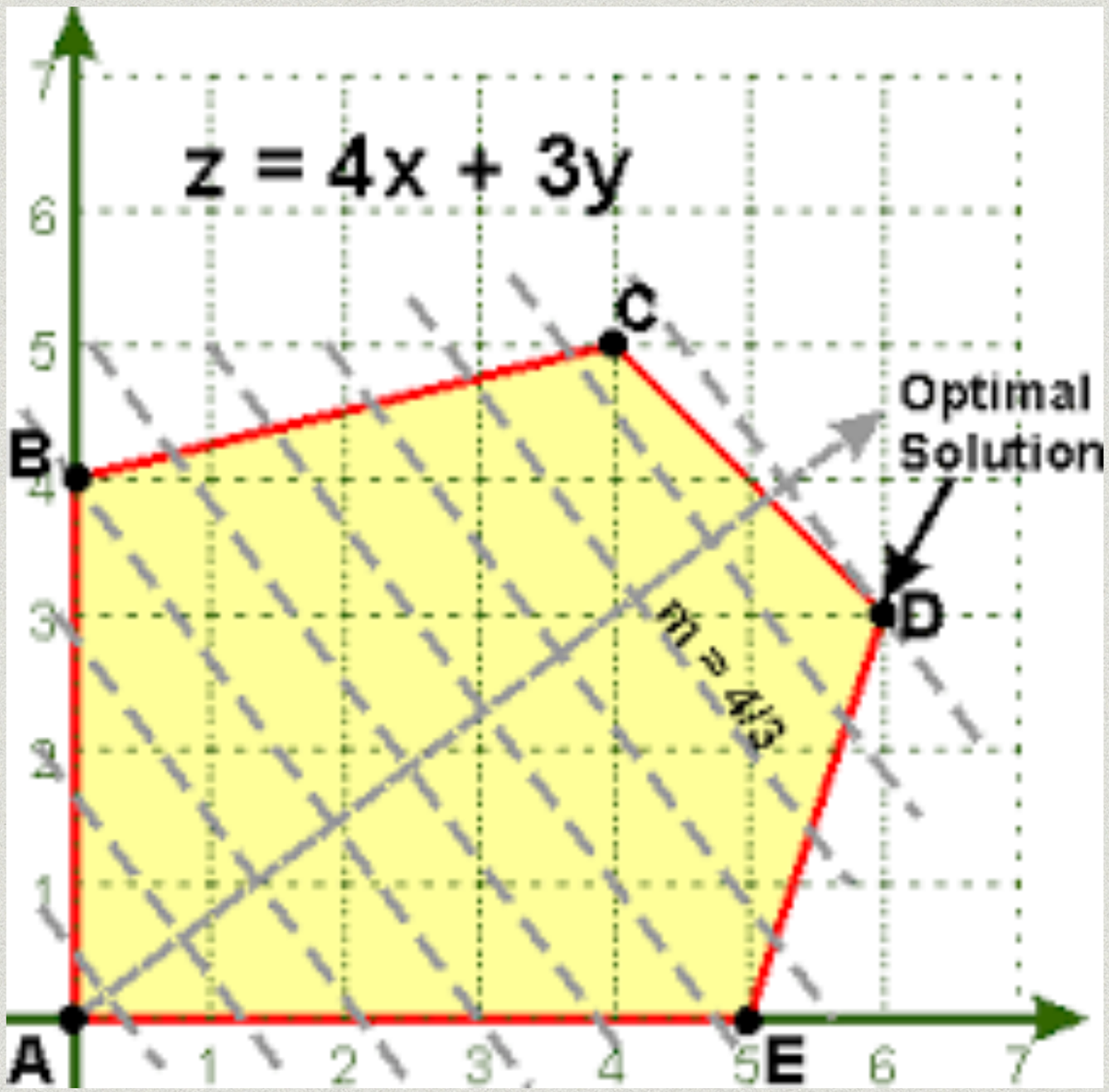
Max Flow


2
9
5
10
4
15
15
10
source
s
5
3
8
6
10
t
sink
4
6
15
10
capacity
15
4
30
7

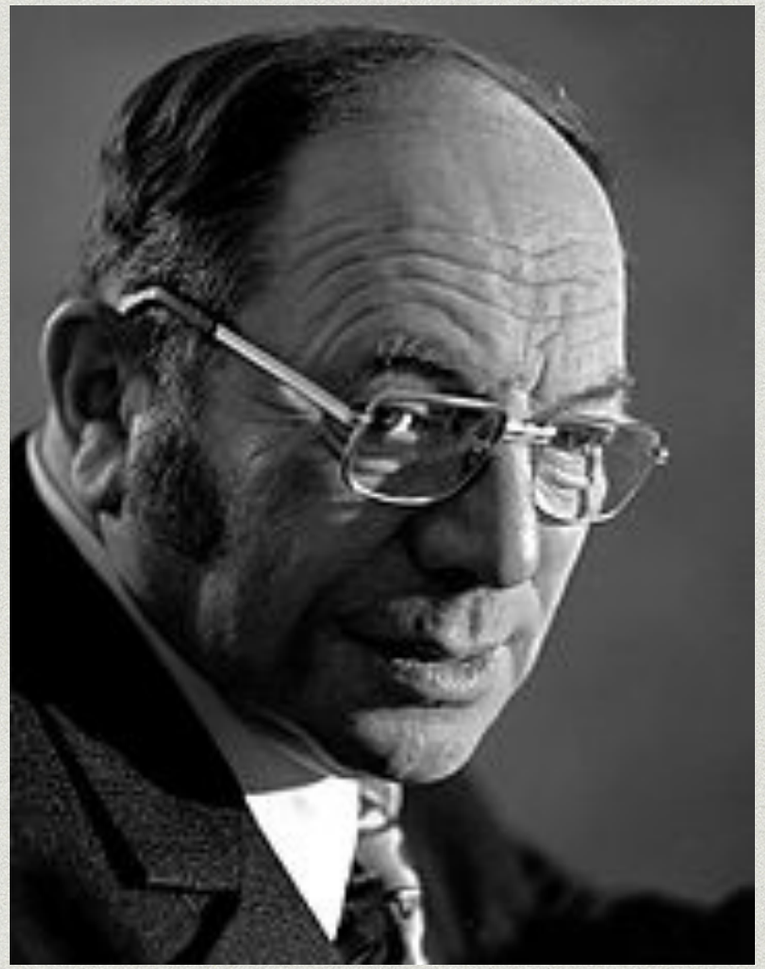
максимальный поток



Linear Programming



Nobel Prize



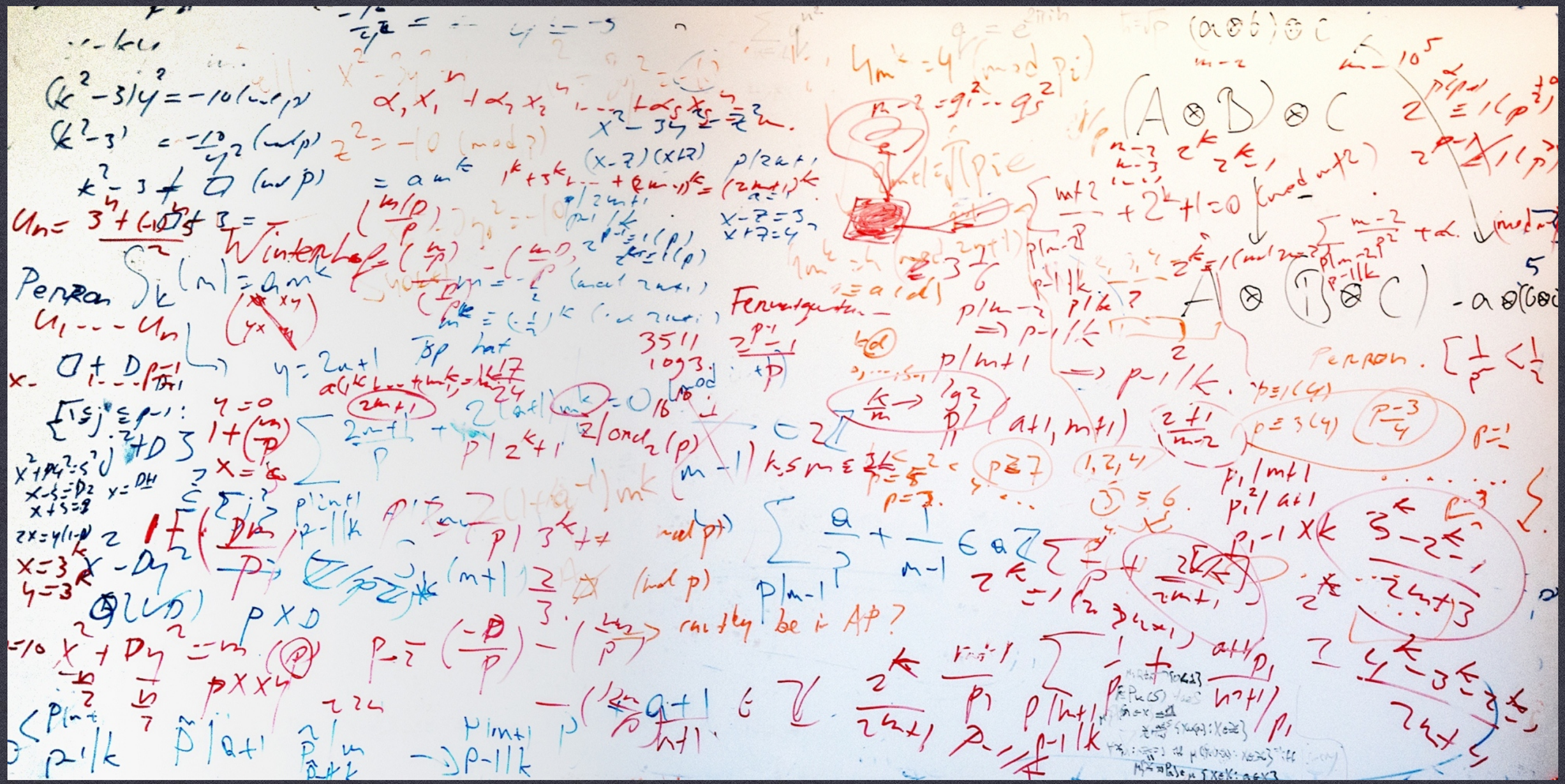
Leonid V. Kantorovich

Tjalling Koopmans

SIEGE OF LENINGRAD SEPTEMBER 1941- JANUARY 1944

THE ROAD OF LIFE

A FAMOUS SYMPHONY



BACK TO THE BOARD