Exponential Clocks, VC dimension, and TU matrices

Algorithmic Toolbox

Ola Svensson

Exponential Clocks

Exponential Distribution

- A random variable X distributed according to the exponential distribution with rate λ, denoted by $X \sim exp(\lambda)$ has
  - pdf $f_X(x) = \lambda e^{-\lambda x}$
  - Cdf $\Pr[X \leq x] = F_X(x) = 1 - e^{-\lambda x}$

NICE PROPERTIES:

- The exponential distribution is memoryless: $\Pr[X \geq s + t \mid X \geq s] = \Pr[X \geq t]$
- Let $X_1, \ldots, X_k$ be independent random variables with $X_i \sim exp(\lambda_i)$
  - $\min\{X_1, \ldots, X_k\} \sim exp(\lambda_1 + \ldots + \lambda_k)$
  - $\Pr[X_i \leq \min_{j \neq i} X_j] = \dfrac{\lambda_i}{\lambda_1 + \ldots + \lambda_k}$

The algorithm

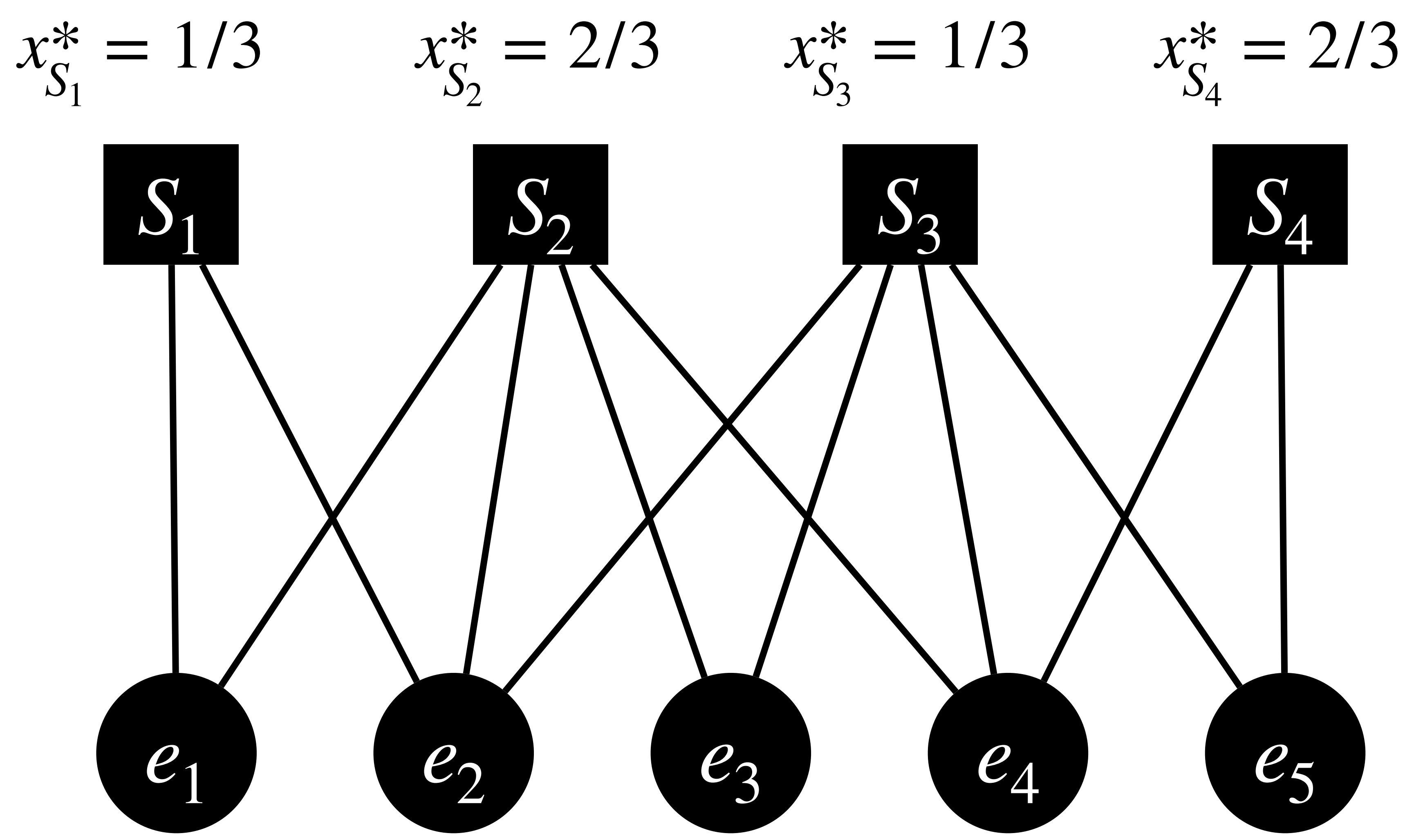
- Solve Set Cover LP:

$$\text{minimize} \quad \sum_i c(S_i) \cdot x_i$$

$$\text{Subject to} \quad \sum_{i:e \in S_i} x_i \geq 1 \text{ for every } e \in U$$

$$x_i \geq 0 \text{ for every } S_i \in T$$

- Let x^* be an optimal solution
- For each set S_i sample $Z_{S_i} \sim \exp(x_i)$
- Output $\bigcup_{e \in U} \arg\min\{Z_{S_i} \mid e \in S_i\}$



The algorithm

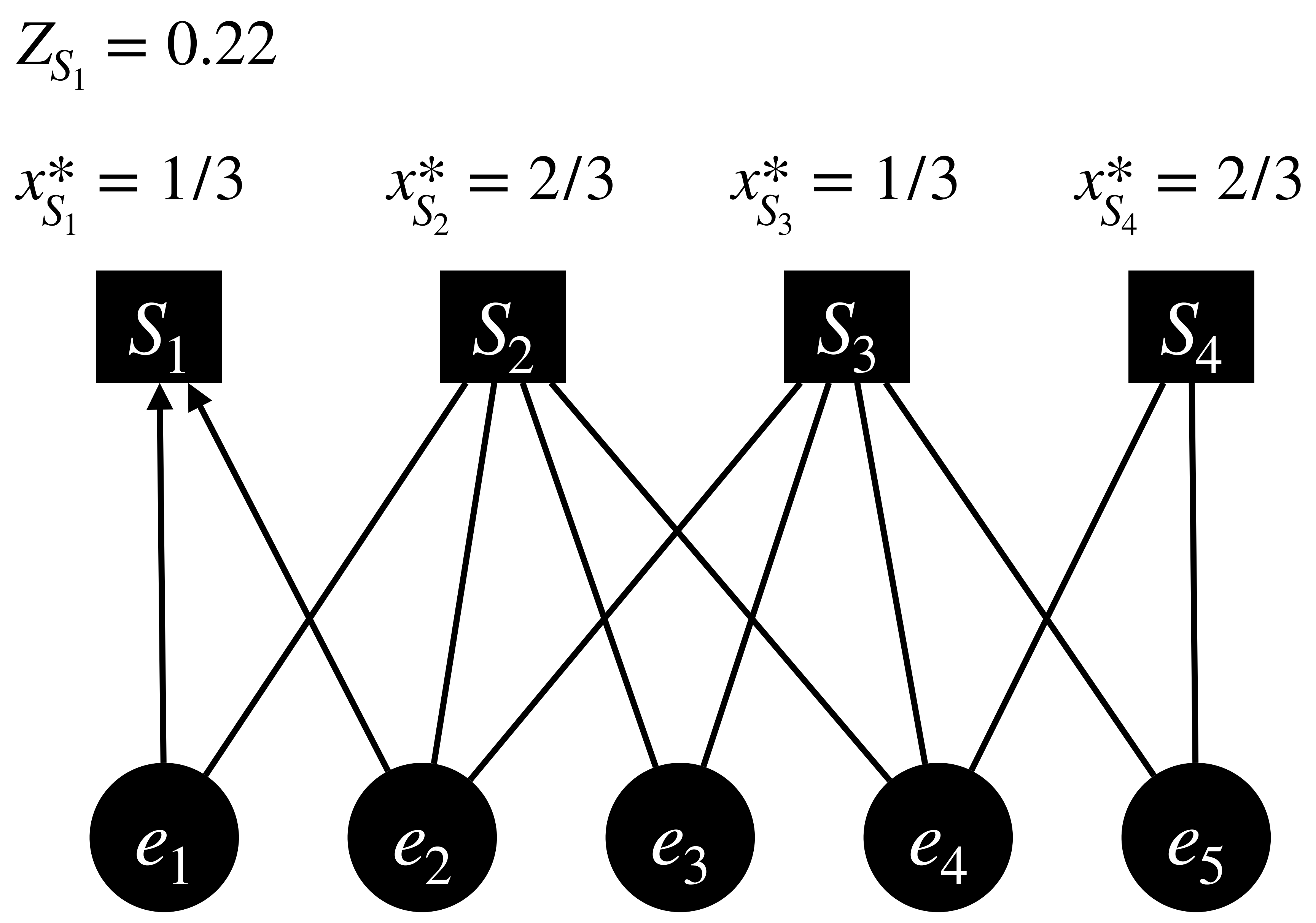
- Solve Set Cover LP:

$$\text{minimize} \quad \sum_i c(S_i) \cdot x_i$$

$$\text{Subject to} \quad \sum_{i: e \in S_i} x_i \geq 1 \text{ for every } e \in U$$

$$x_i \geq 0 \text{ for every } S_i \in T$$

- Let x^* be an optimal solution
- For each set S_i sample $Z_{S_i} \sim \exp(x_i)$
- Output $\bigcup_{e \in U} \arg\min\{Z_{S_i} \mid e \in S_i\}$



The algorithm

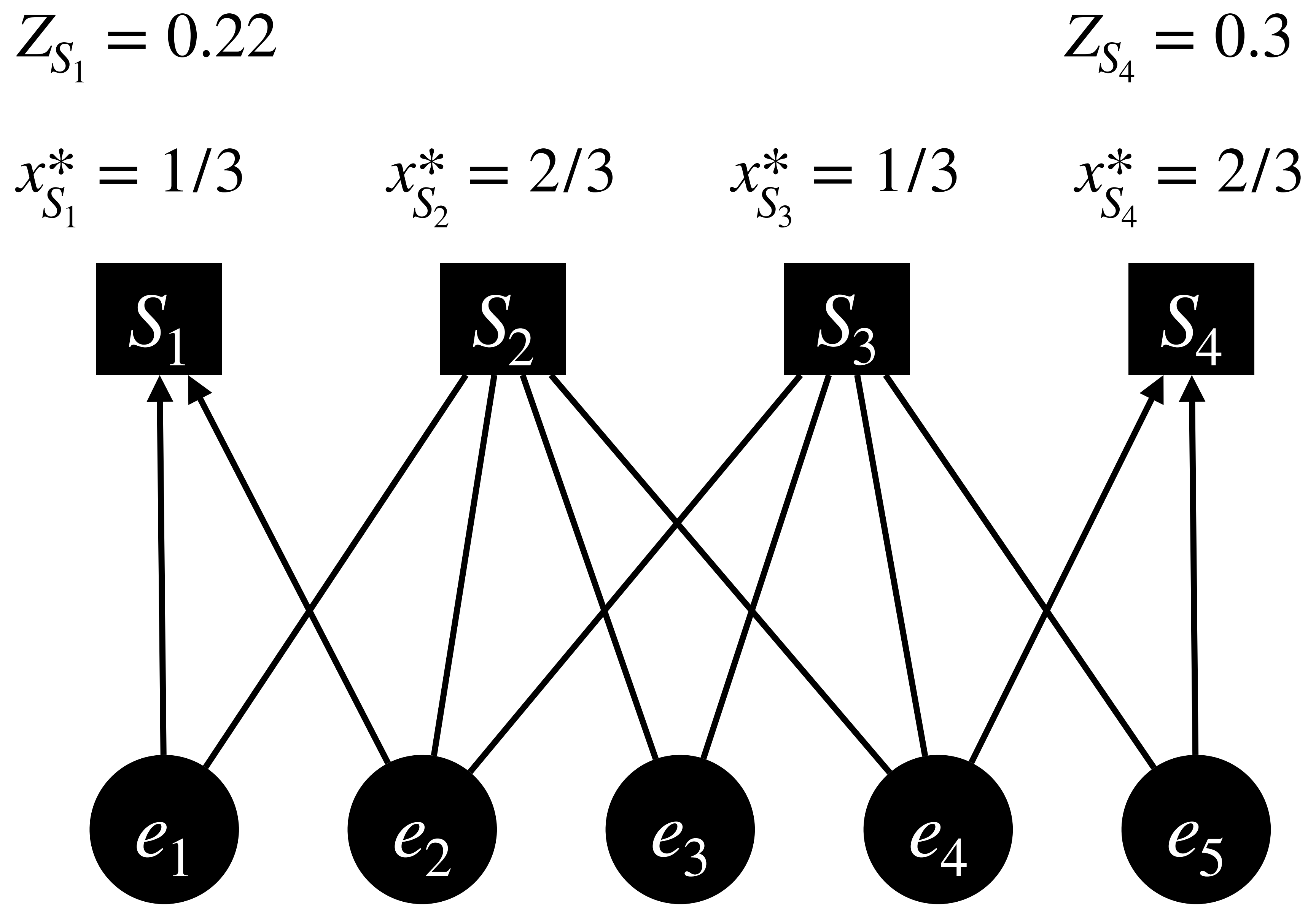
- Solve Set Cover LP:

$$\text{minimize} \quad \sum_i c(S_i) \cdot x_i$$

$$\text{Subject to} \quad \sum_{i: e \in S_i} x_i \geq 1 \text{ for every } e \in U$$

$$x_i \geq 0 \text{ for every } S_i \in T$$

- Let x^* be an optimal solution
- For each set S_i sample $Z_{S_i} \sim \exp(x_i^*)$
- Output $\bigcup_{e \in U} \arg\min\{Z_{S_i} \mid e \in S_i\}$



The algorithm

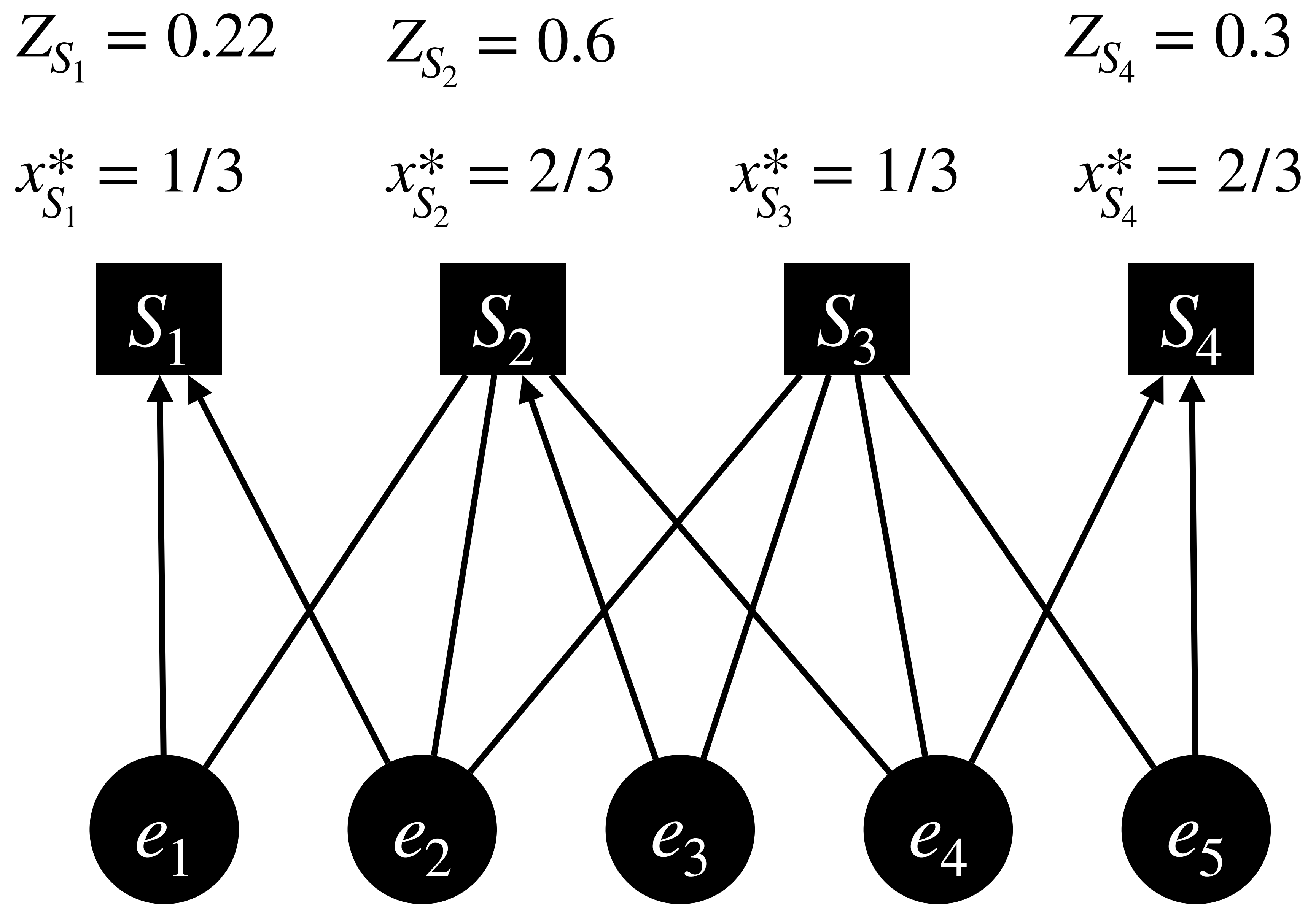
- Solve Set Cover LP:

$$\text{minimize} \quad \sum_i c(S_i) \cdot x_i$$

$$\text{Subject to} \quad \sum_{i:e \in S_i} x_i \geq 1 \text{ for every } e \in U$$

$$x_i \geq 0 \text{ for every } S_i \in T$$

- Let x^* be an optimal solution
- For each set S_i sample $Z_{S_i} \sim \exp(x_i)$
- Output $\bigcup_{e \in U} \arg\min\{Z_{S_i} \mid e \in S_i\}$



The algorithm

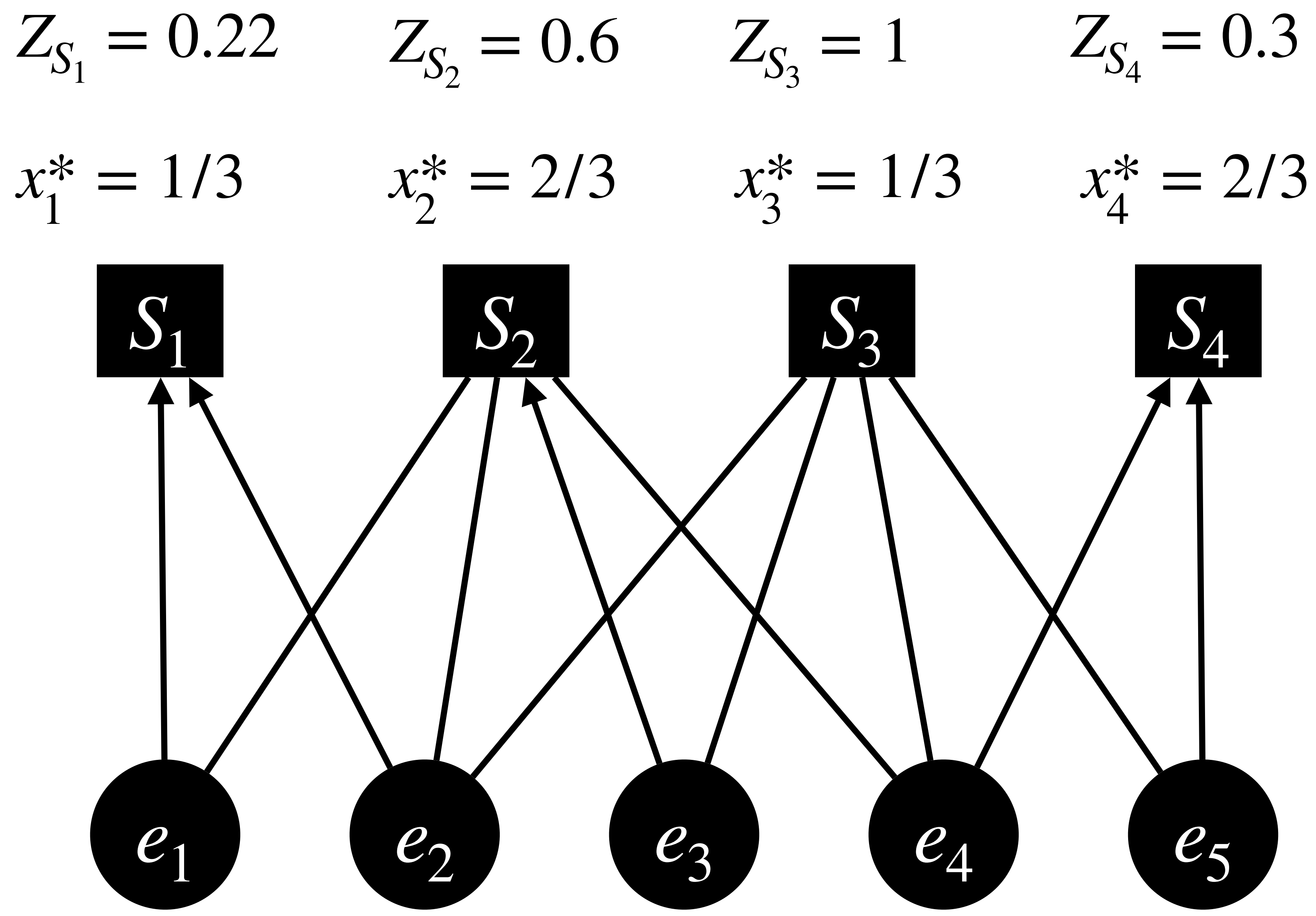
- Solve Set Cover LP:

$$\text{minimize} \quad \sum_i c(S_i) \cdot x_i$$

$$\text{Subject to} \quad \sum_{i: e \in S_i} x_i \geq 1 \text{ for every } e \in U$$

$$x_i \geq 0 \text{ for every } S_i \in T$$

- Let x^* be an optimal solution
- For each set S_i sample $Z_{S_i} \sim \exp(x_i)$
- Output $\bigcup_{e \in U} \arg\min\{Z_{S_i} \mid e \in S_i\}$



Output $\{S_1, S_2, S_4\}$

Analysis (1/2)

The probability that we output S_i is at most $(1 + \ln|S_i|)x_i^*$

- For element $e \in S_i$, let A_e be the event that e chooses S_i, i.e., that $Z_{S_i} = \min\{Z_{S_j} \mid e \in S_j\}$
- With this notation, $\Pr[S_i \text{ output}] = \Pr[\vee_{e \in S_i} A_e]$
- Now $\Pr[\vee_{e \in S_i} A_e] = \Pr[\vee_{e \in S_i} A_e \mid Z_{S_i} \leq \alpha] \Pr[Z_{S_i} \leq \alpha] + \Pr[\vee_{e \in S_i} A_e \mid Z_{S_i} \geq \alpha] \Pr[Z_{S_i} \geq \alpha]$

$$\Pr[\vee_{e \in S_i} A_e \mid Z_{S_i} \leq \alpha] \Pr[Z_{S_i} \leq \alpha]$$

$$\leq \Pr[Z_{S_i} \leq \alpha] \leq 1 - e^{-x_i^* \alpha} \leq x_i^* \alpha$$

$$\Pr[\vee_{e \in S_i} A_e \mid Z_{S_i} \geq \alpha] \Pr[Z_{S_i} \geq \alpha]$$

$$\leq \sum_{e \in S_i} \Pr[A_e \mid Z_{S_i} \geq \alpha] \cdot e^{-x_i^* \alpha}$$

So the probability that we output S_i is at most $\alpha x_i^* + \sum_{e \in S_i} \Pr[A_e \mid Z_{S_i} \geq \alpha] \cdot e^{-x_i^* \alpha}$

Analysis (2/2)

The probability that we output S_i is at most $(1 + \ln|S_i|)x_i^*$

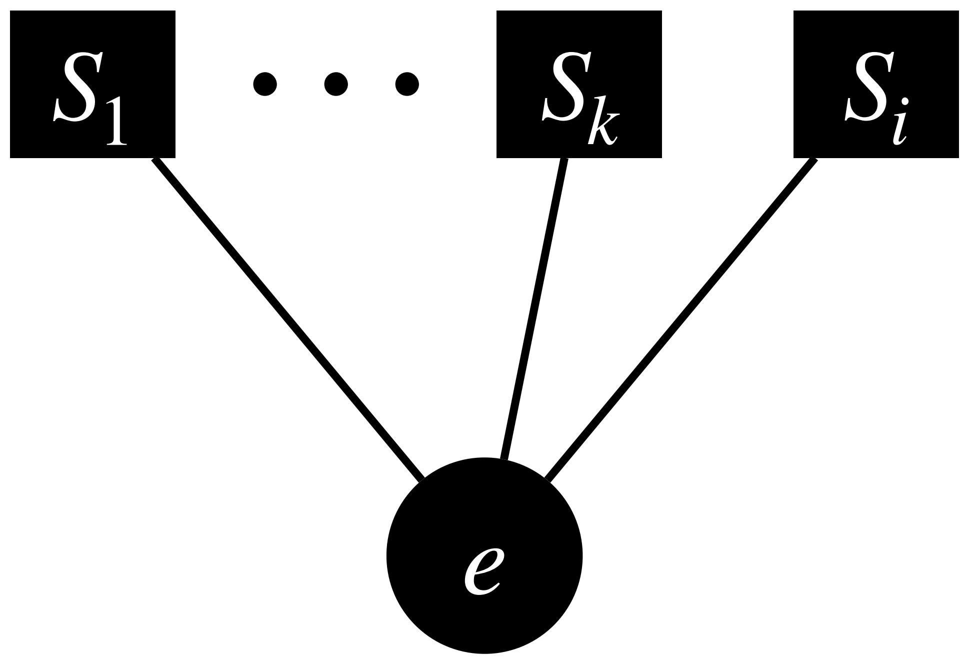
The probability that we output S_i is at most $\alpha x_i^* + \sum_{e \in S_i} \Pr[A_e \mid Z_{S_i} \geq \alpha] \cdot e^{-x_i^* \alpha}$

To analyze $\Pr[A_e \mid Z_{S_i} \geq \alpha]$, let $S_1, \ldots, S_k, S_i$ be the sets that cover e and let $Y = \min\{Z_{S_1}, \ldots, Z_{S_k}\}$

Note that $Y \sim \exp(x_1^* + \ldots + x_k^*)$, and that A_e does not happen if $Y < \alpha$

Hence, $e^{-x_i^* \alpha} \Pr[A_e \mid Z_{S_i} \geq \alpha] = e^{-x_i^* \alpha} \Pr[Y \geq \alpha] \Pr[A_e \mid Z_{S_i} \geq \alpha, Y \geq \alpha]$

$$= e^{-\alpha(x_1^* + \ldots x_k^* + x_i^*)} \cdot \frac{x_i^*}{x_1^* + \ldots + x_k^* + x_i^*} \geq e^{-\alpha} \cdot x_i^*$$



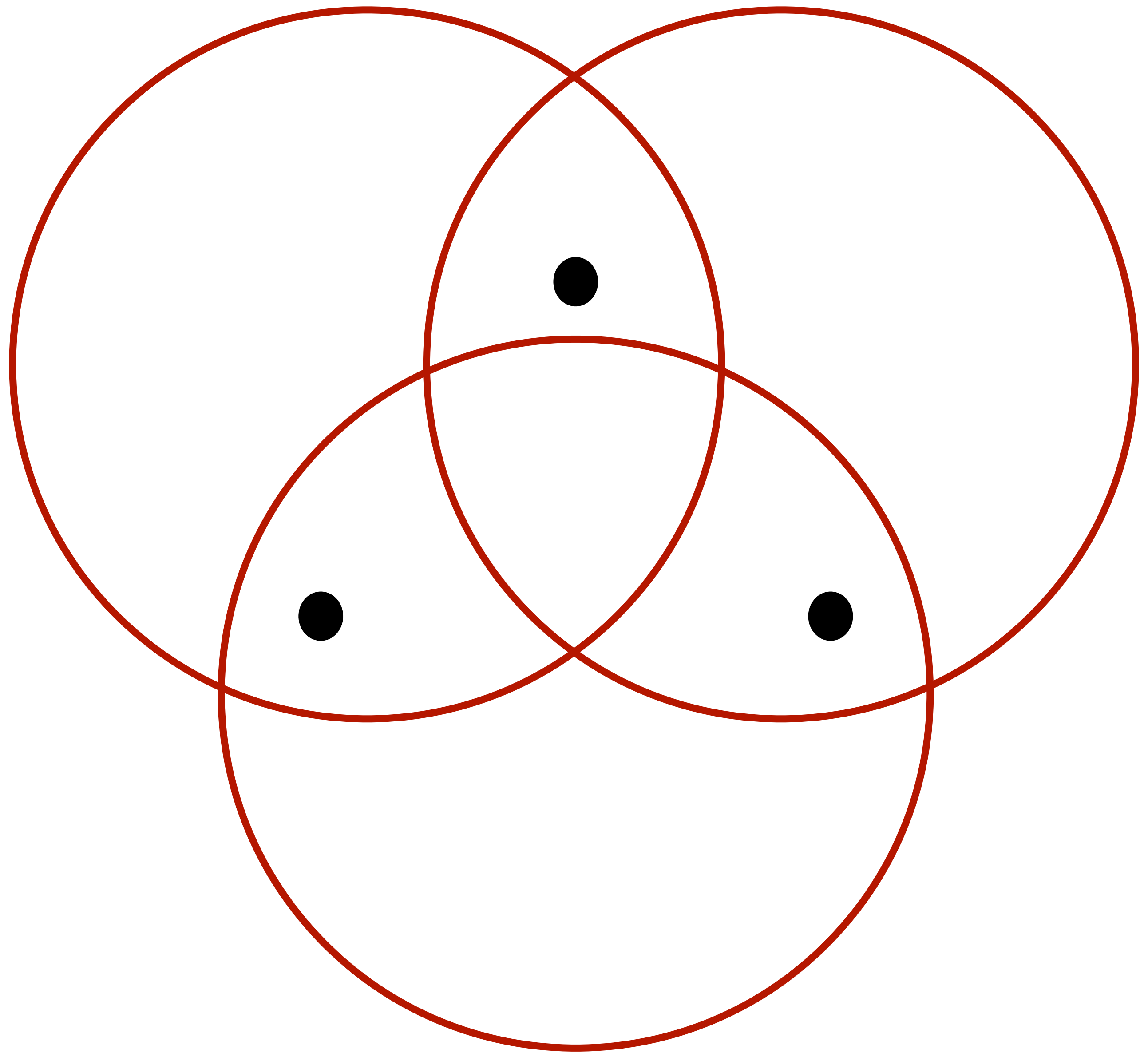
It follows that the probability that we output S_i is at most $\alpha x_i^* + \sum_{e \in S_i} e^{-\alpha} x_i^*$

Selecting $\alpha = \ln|S_i|$ now gives the result

Vapnik-Chervonenkis (VC) dimension

Definition of VC dimension

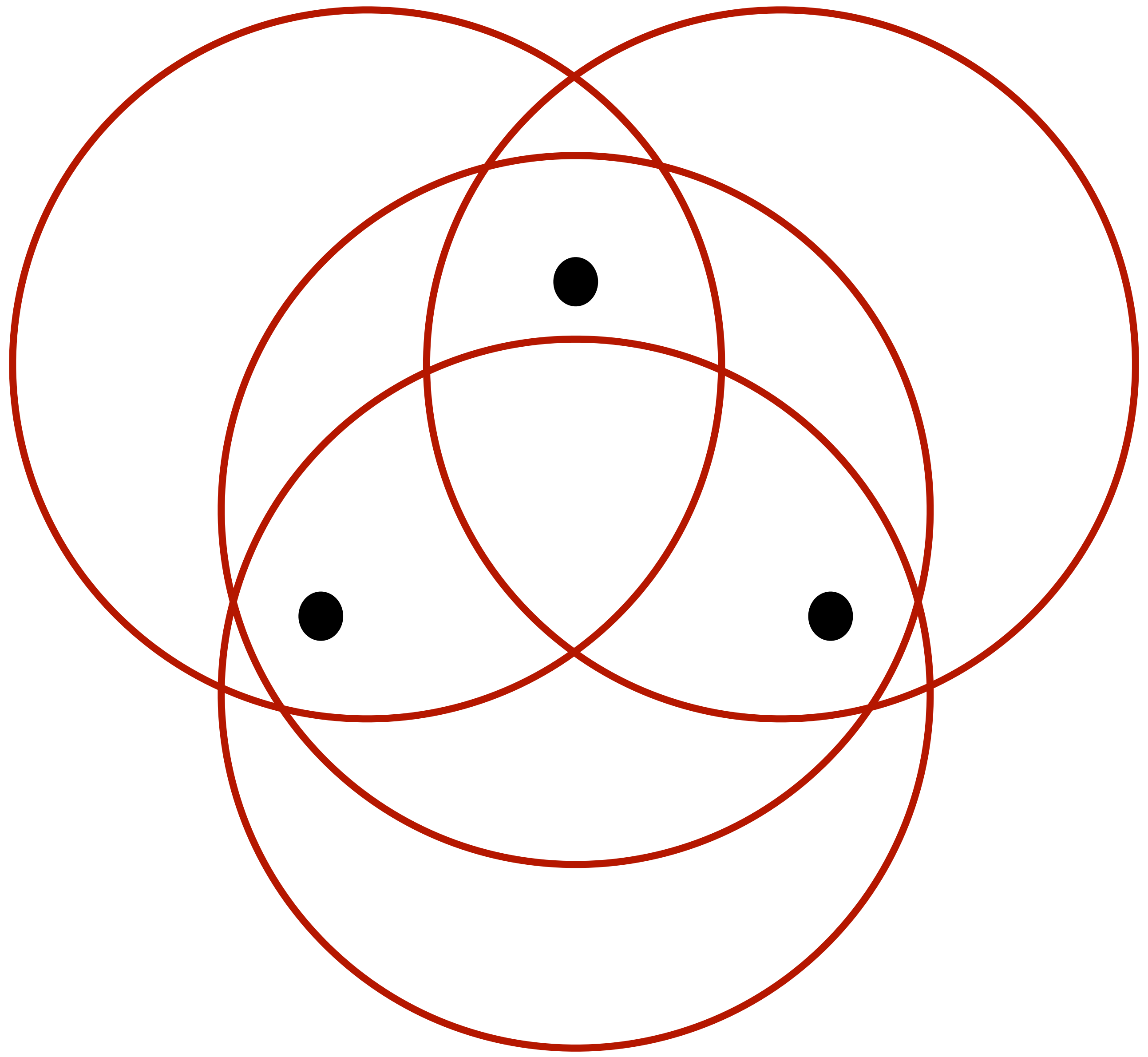
- Consider a ground set U and a family of subsets $T = \{S_1, S_2, \ldots, S_m\}$
- We say that T *shatters* a subset $U' \subseteq U$ if $\{U' \cap S \mid S \in T\}$ contains all subsets of U'
- The *VC dimension* of (U, T) is the maximum d so that T shatters a subset $U' \subseteq U$ of cardinality d



+ singleton sets have VC dimension 2

Definition of VC dimension

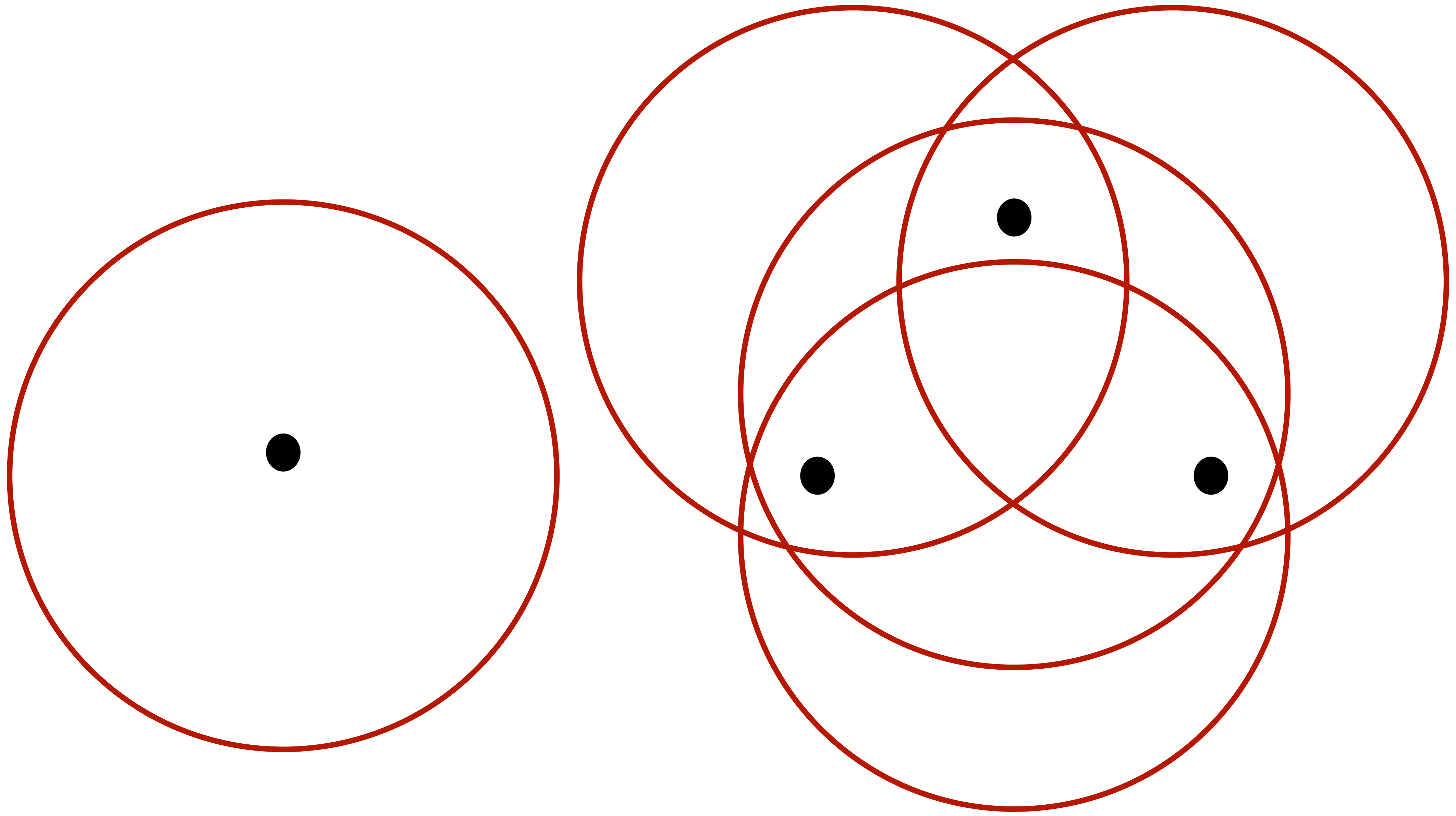
- Consider a ground set U and a family of subsets $T = \{S_1, S_2, \ldots, S_m\}$
- We say that T *shatters* a subset $U' \subseteq U$ if $\{U' \cap S \mid S \in T\}$ contains all subsets of U'
- The *VC dimension* of (U, T) is the maximum d so that T shatters a subset $U' \subseteq U$ of cardinality d



+ singleton sets have VC dimension 2

Definition of VC dimension

- Consider a ground set U and a family of subsets $T = \{S_1, S_2, \ldots, S_m\}$
- We say that T *shatters* a subset $U' \subseteq U$ if $\{U' \cap S \mid S \in T\}$ contains all subsets of U'
- The *VC dimension* of (U, T) is the maximum d so that T shatters a subset $U' \subseteq U$ of cardinality d

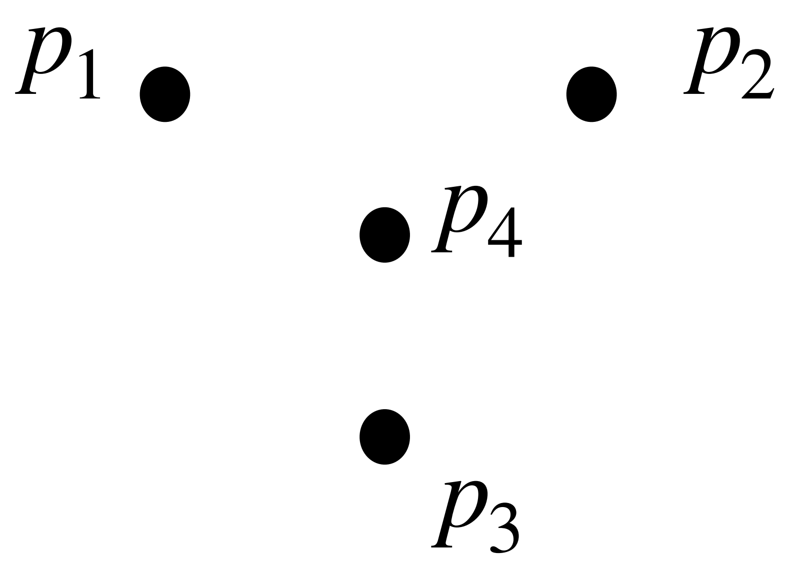


+ singleton sets have VC dimension 3

VC Dimension of disks in plane ≤ 3

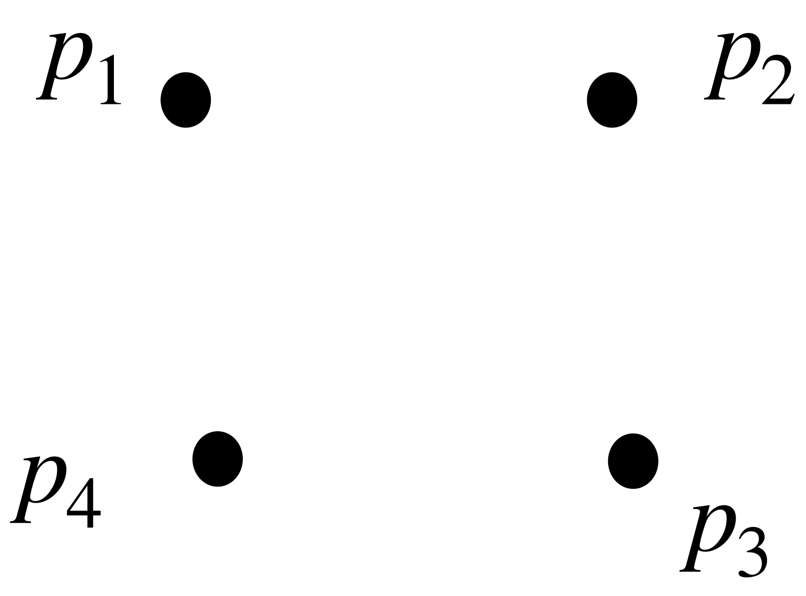
Consider 4 points

- Case 1: one point is in the interior of the convex hull of the others:



No disk can realize the set $\{p_1, p_2, p_3\}$

- Case 2: All four points are at the boundary of the convex hull



Can't exist two disks where one contains $\{p_1, p_3\}$ and the other $\{p_2, p_4\}$

Because any two such disks would intersect at 4 points. But any disks intersect at most 2 points

ϵ-Net Theorem

Fix $\epsilon, \delta > 0$

- Suppose (U, T) has VC-dimension d.
- If we select $m \geq \max\{\frac{4}{\epsilon} \log \frac{2}{\delta}, \frac{8d}{\epsilon} \log \frac{8d}{\epsilon}\}$ many samples from U independently at random
- Then, with probability $\geq 1 - \delta$, we sample at least one element from every set $S \in T$ of cardinality at least $\epsilon |U|$

What is cool here is that the number of samples is independent of $|U|$

Sauer's Lemma

If (U, T) with $|U| = n$ has VC-dimension d then $|T| \leq g(n, d)$ where $g(n, d) = \sum_{i=0}^{d} \binom{n}{i}$

- We will prove the following stronger claim by Pajor 1985:
  - the number of different subsets that are shattered by T is at least $|T|$

- This implies Sauer's lemma since the if $|T| > g(n, d)$ there must be a subset shattered of size at least $d + 1$ by counting

- The proof is credited to Noga Alon or to Ron Aharoni and Ron Holzman.

The number of different subsets that are shattered by T is at least $|T|$

- Base: Every family of only one set shatters the empty set.
- Inductive step: Let T be a family of two or more sets and let x be an element that appear in some but not all sets.
- Split T into two subfamilies, those that contain x and those that don't.
- By IH, these two subfamilies shatter two collections of sets whose sizes add to at least $|T|$
- None of these shattered sets contain x.
- If a shattered set appears in one subfamily then it contributes one unit to the subfamilies and one unit to #shattered sets of T
- If a shattered set appears twice then it counts twice for the subfamilies and then also S and S+x are shattered by $|T|$
- Therefore the number of shattered sets by subfamilies and T is the same and so T shatters iat least $|T|$ sets.

Proof of ϵ-Net Theorem

- Let E_1 be the event that our sampled points N fail to be an ϵ-net, i.e., $E_1 = \{\exists S \in T \mid |S| \geq \epsilon |U|, S \cap N = \emptyset\}$
- We wish to prove $\Pr[E_1] \leq \delta$. This turns out to be hard as there can be $g(n, d)$ sets in total and our sample size doesn't depend on n.
- Instead, consider what happens if we first sample $N = \{x_1, \ldots, x_m\}$ and then $Y = \{y_1, \ldots, y_m\}$ from the same distribution.
- Let $E_2 = \{\exists S \in T \mid |S| \geq \epsilon |U|, S \cap N = \emptyset, |S \cap Y| \geq \epsilon m/2\}$
- Note that $\Pr[E_2] \leq \Pr[E_1]$ but we have also $\Pr[E_1] \leq 2\Pr[E_2]$ since each large set has in expectation $|S \cap Y| = \epsilon m$ and our samples are independent, so we can apply standard concentration bounds.
- It is thus sufficient to upper bound $\Pr[E_2]$.
- To do this, we will upper bound $E_2' = \{\exists S \in T \mid S \cap N = \emptyset, |S \cap Y| \geq \epsilon m/2\}$.
- Clearly $\Pr[E_2] \leq \Pr[E_2']$ and note crucially that U doesn't appear in the definition of the event anymore.

Upper bounding $E_2' = \{ \exists S \in T \mid S \cap N = \emptyset, |S \cap Y| \geq \epsilon m/2\}$

$\Pr[E_2] \leq \Pr[E_2'] \leq g(d,2m) \cdot 2^{-\epsilon m/2}$

- We imagine that we sample $Z = N \cup Y$ together and *then* randomly decide which elements belong to N and which belong to Y.
- We have $\Pr[E_2'] = \sum_Z \Pr[E_2' \mid Z] \Pr[Z]$. We now fix a set Z and bound $\Pr[E_2' \mid Z]$.
- To do this, it is enough to consider the set system $T_Z = \{S \cap Z \mid S \in T\}$, i.e., the projection onto Z.
- By Sauer's lemma, T_Z contains at most $g(d,2m)$ sets.
- Let us now fix any set $S \in T_Z$ and consider the event $E_S = \{S \cap N = \emptyset, S \cap Y \geq \epsilon m/2\}$.
- For $k = |S \cap Z|$, we have $\Pr[N \cap S = \emptyset \mid N \cap Z \geq \epsilon m/2] = \frac{\binom{2m-k}{m}}{\binom{2m}{m}} \leq \ldots \leq 2^{-\epsilon m/2}$
- Thus by union bound $\Pr[E_2' \mid Z] \leq g(d,2m) \cdot 2^{-\epsilon m/2}$

ϵ-Net Theorem

Fix $\epsilon, \delta > 0$

- Suppose (U, T) has VC-dimension d.
- If we select $m \geq \max\{\frac{4}{\epsilon}\log\frac{2}{\delta}, \frac{8d}{\epsilon}\log\frac{8d}{\epsilon}\}$ many samples from U independently at random
- Then, with probability $\geq 1 - \delta$, we sample at least one element from every set $S \in T$ of cardinality at least $\epsilon|U|$

By the previous argument, we have that the success probability is at least $g(d,2m) \cdot 2^{\epsilon m/2}$

The statement follows by the selection of m

What does this have to do with Set Cover???

Hitting Set

- Input: A universe U, and a family of sets T.
- Output: The smallest subset $U' \subseteq U$ that hits every set in T, i.e., $U' \cap S \neq \emptyset$ for every $S \in T$
- LP relaxation

Minimize $\sum_{e \in U} x_e$

Subject to $\sum_{e \in S} x_e \geq 1$ for every set $S \in T$

$x_e \geq 0$ for every element $e \in U$

Same as set cover we just swapped the meaning of sets and elements

Suppose T has VC-dimension d

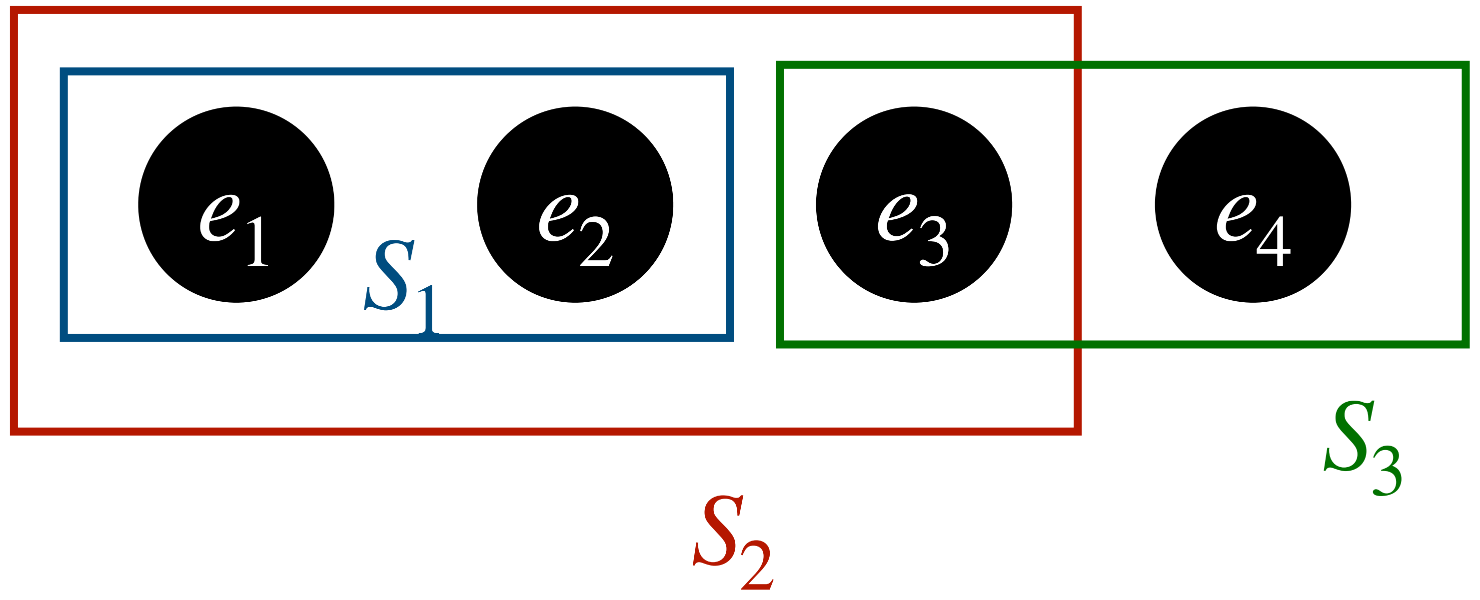
Then we have an $O(d \log(d \cdot OPT))$-approximation algorithm

- Solve LP to obtain optimal solution x^*, let $x' = x^*/|x^*|$ and so
  - $\sum_{e \in U} x'_e = 1$ and $\sum_{e \in S} x_e \geq 1/|x^*|$ for every $S \in T$
- Now find an ϵ-net $U' \subseteq U$ of size $O(\frac{1}{\epsilon} d \log(d/\epsilon))$ where $\epsilon = 1/|x^*|$
- This is a hitting set of size $O(|x^*| d \log(|x^*| d))$ and since $|x^*| \leq OPT$ this gives the guarantee.

Totally unimodularity

Hitting set with consecutive ones

- Suppose elements of U can be ordered so that all sets in T are consecutive subsets in this order.
- Example:



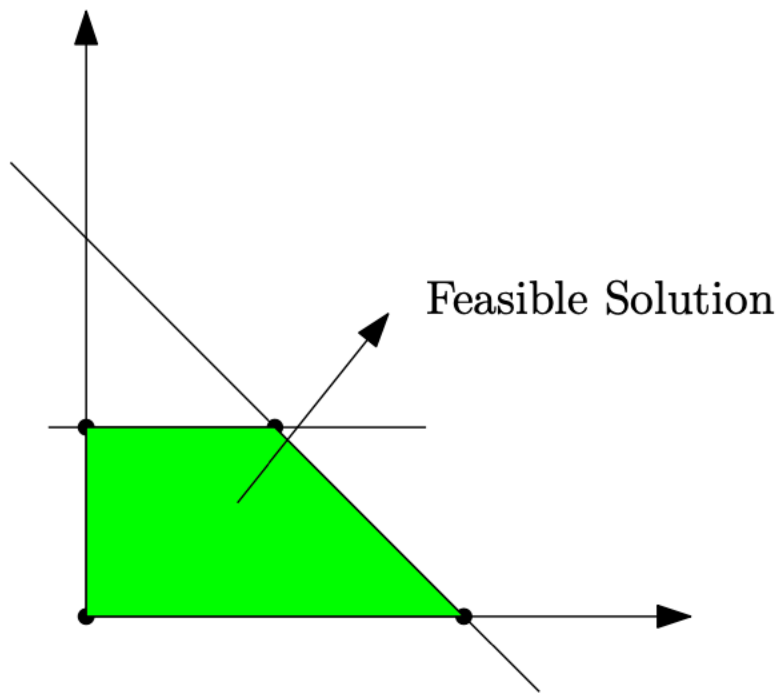
In this case the linear program is integral, i.e., solves the problem exactly! WHY???

Totally unimodularity

- A matrix A is totally unimodular if every square submatrix has determinant $0, +1$, or -1. In particular, this implies that all entries are 0 or ± 1.

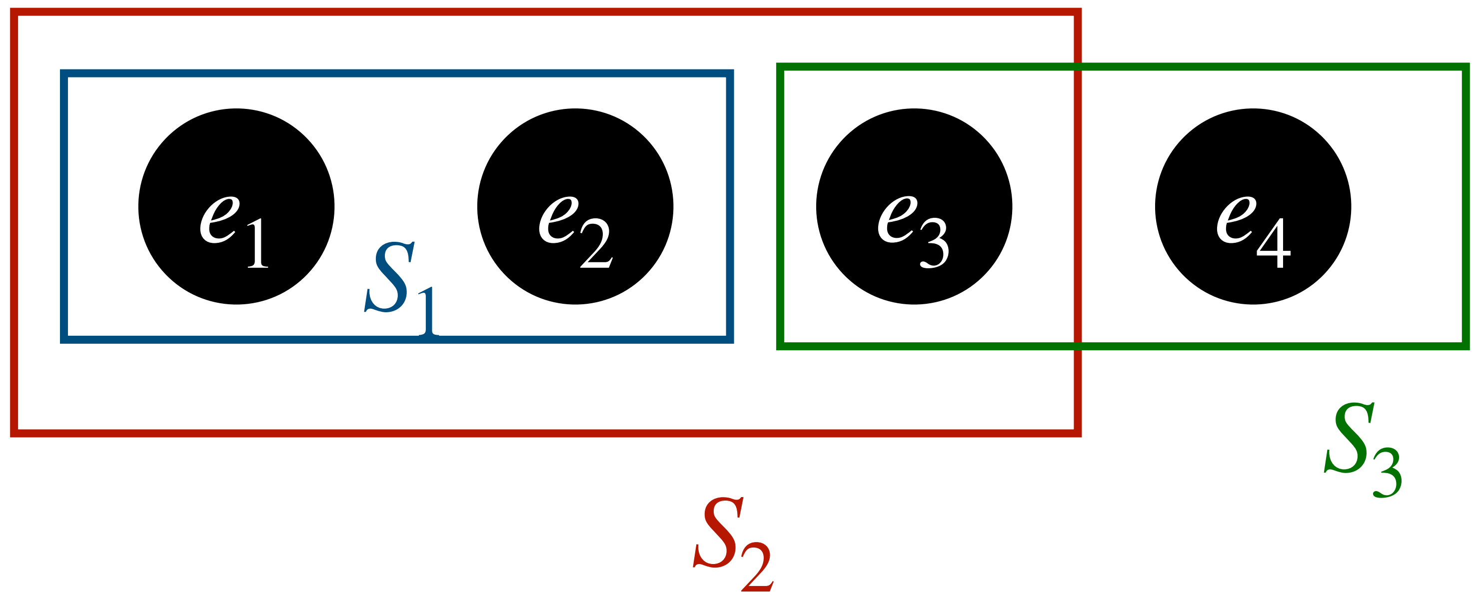
- Theorem: If A is totally unimodular and b is an integer vector, then $P = \{x \mid Ax \geq b\}$ has integer vertices.

- Proof: Let v be a vertex of P. There exists a non-singular square sub-matrix A' of A such that $A'v = b$. We have $\det(A') = \pm 1$ by totally unimodularity. By Cramer's rule, we have $v_i = \dfrac{\det(A'_i \mid b)}{\det(A')}$ where $A'_i \mid b$ is A' with the i:th column replaced by b. Therefore, v_i is an integer.



$$
\begin{aligned}
\text{Maximize} \quad & x + y \\
\text{Subject to} \quad & x + y \leq 2 \\
& y \leq 1 \\
& x, y \geq 0
\end{aligned}
$$

Linear programming relaxation



$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Every square submatrix of A satisfies the consecutive ones property!

A matrix with consecutive ones are totally unimodular

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

- Define matrix C by $C_{r,c} = \begin{cases} B_{r,c} - B_{r,c+1} & \text{for c< \#columns} \\ B_{r,c} & \text{otherwise} \end{cases}$ We have that $\det C = \det B$
- Each row of C has at most two entries in ± 1
  - If some row has no non-zero entries, the determinant is 0
  - If some row has one non-zero entry then do Laplace expansion and consider the only minor that has a non-zero coefficient
  - After all expansions, each row has exactly one $+1$ and one -1. Call this matrix C' and observe $C'\mathbf{1} = \mathbf{0}$ and hence $\det C' = 0$
  - Hence $\det B = \det C \in \{-1,0,1\}$

Other prominent examples of TU matrices

- Incidence matrices of bipartite graphs
- Incidence matrices of directed graphs
- Network flow matrices
- Seymour’80 gave a complete characterisation of TU matrices.