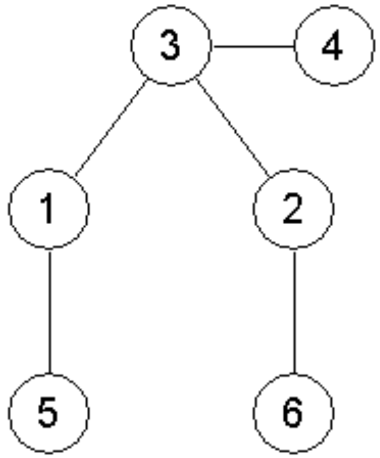
A fast Monte Carlo algorithm for the homogeneous set sandwich problem

Vinícius Gusmão Pereira de Sá
Guilherme Dias da Fonseca
Celina Miraglia Herrera de Figueiredo

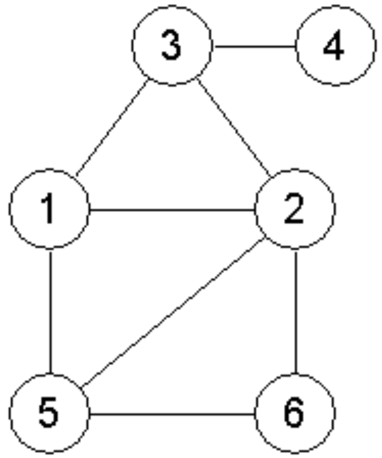
Universidade Federal do Rio de Janeiro

Sandwich Graphs

A graph $G_2(V_2,E_2)$ is said to be a *supergraph* of a graph $G_1(V_1,E_1)$ if and only if $V_2 = V_1$ and $E_2 \subseteq E_1$.

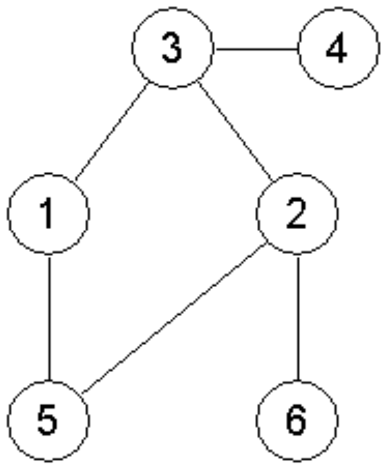


G_1



G_2
(supergraph of G_1)

A graph $G_S(V_S,E_S)$ is said to be a *sandwich graph* of a pair $G_1(V,E_1)$, $G_2(V,E_2)$ if and only if $V_S = V$ and $E_1 \subseteq E_S \subseteq E_2$.

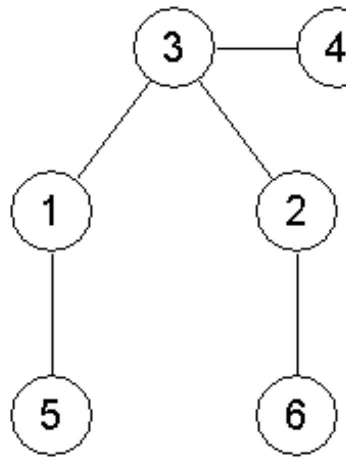


G_S
(sandwich graph of G_1,G_2)

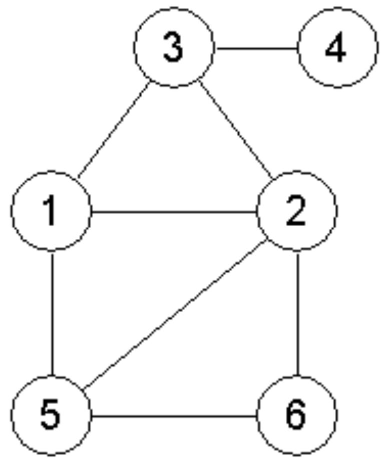
Sandwich Problems

Input: two graphs $G_1(V,E_1)$ and $G_2(V,E_2)$ such that G_2 is a *supergraph* of G_1.

Question: is there any sandwich graph G_S of pair (G_1,G_2) that has property $\prod$?



G_1



G_2
(supergraph of G_1)

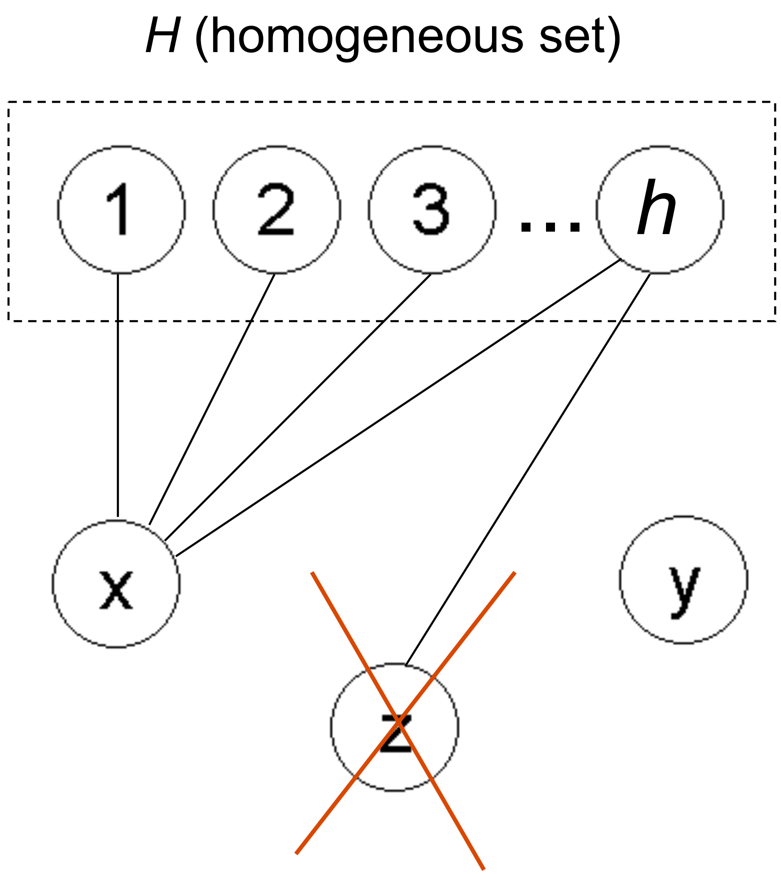
An edge (x,y) is said to be

- *mandatory*, if $(x,y) \in E_1$
- *forbidden*, if $(x,y) \notin E_2$
- *optional*, if $(x,y) \in E_2 \setminus E_1$

Homogeneous Sets

Let G (V, E) be a graph.

A set $H \subset V$ is a *homogeneous set* of G if and only if all vertices in H have exactly the same neighborhood outside H and $1 < |H| < |V|$.



The Homogeneous Set Sandwich Problem (HSSP): is there a sandwich graph of (G_1,G_2) which admits a homogeneous set?

History

$m = \min \{m_M, m_F\}$
$M = \max \{m_M, m_F\}$

AUTHORS	PUBLICATION	ALGORITHM		TIME	DET./RAND.
Cerioli, Everett, Figueiredo, Klein	IPL, 1998	1	Ehaustive Envelopment	$\cancel{O(n^4)}$ $\boxed{O(M\ n^2)}$	det.
~~Tang, Wang, Yeh~~	~~IPL, 2001~~	~~2~~	~~Strongly Connected Sinks~~	$\cancel{O(\Delta_2\ n^2)}$ $\boxed{O(m\ n)}$	~~det.~~
Figueiredo, P. de Sá	M.Sc., 2003 **+ IPL, 2004**	3	Two-Phase	$O(m\ M)$	det.
Figueiredo, Fonseca, P. de Sá, Spinrad	**LNCS, 2004 (WEA)**	4	Balanced Subsets	$O(n^{3,5})$	det.
		5	Monte Carlo	$O(n^3)$	rand.
Figueiredo, Fonseca, P. de Sá	**Algorithmica, 2005**	6	Harmonic Series	$O(n^3 \log n)$	det.
		7	Growing Cliques	$O(n^3 \log n)$	det.
		8	Las Vegas	$O(n^3)$	rand.
		9	Quick Fill	$O(n^3 \log m/n)$	det.
Bornstein, P. de Sá, Figueiredo	**IPL, 2006**	10	Pair Completion	$O(m\ n \log n)$	det.

QUALIFYING

Randomized Monte Carlo algorithms

- Give the correct answer with known probability *p*
- Their time complexity is computed deterministically

One-sided error Monte Carlo algorithms

<u>YES-biased Monte Carlo</u>

- Whenever the answer is YES, the answer is correct for certain (a certificate is given)

<u>NO-biased Monte Carlo</u>

Defined analogously

Several independent runs ⇔ any desired error ratio

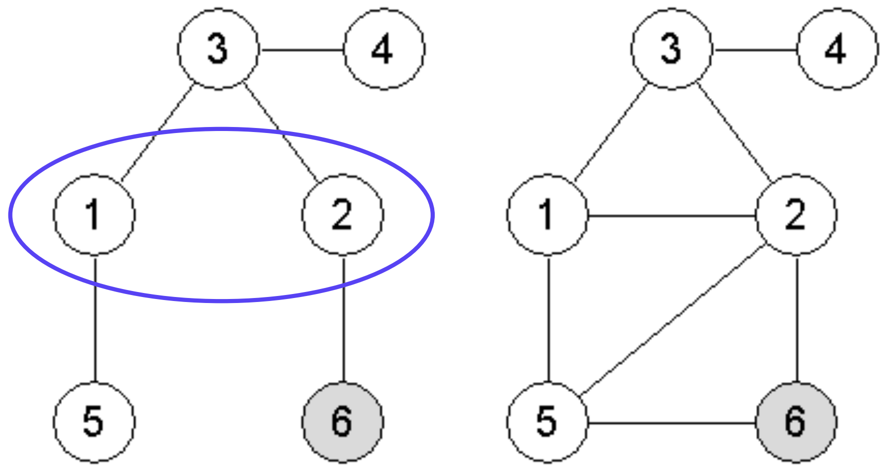
Bias vertices

Let $G_1(V,E_1)$, $G_2(V,E_2)$ be an input instance for the HSSP.

A vertex $b \in V$ is said to be a *bias vertex* of set $S \subseteq V \setminus \{ b \}$ iff there exists at least one mandatory edge $[b, x] \in E_1$ between b and some vertex $x \in S$ and, also, at least one forbidden edge $[b, y] \notin E_2$ between b and some vertex $y \in S$.

The set $B(S)$ containing all bias vertices of S is called its *bias set*.

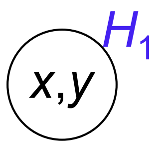
Sandwich instance (HSSP input)



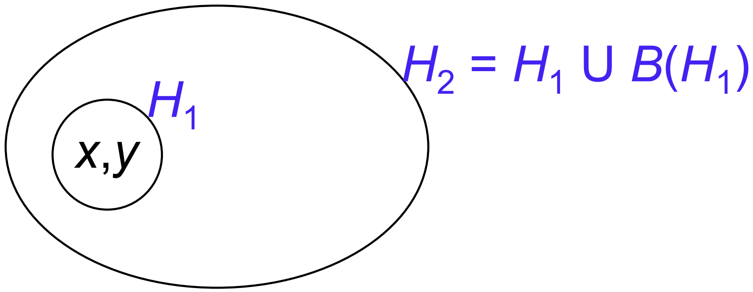
Sandwich Homogeneous Sets Characterization

A set $H \subset V$ is a sandwich homogeneous set of pair $G_1(V,E_1)$, $G_2(V,E_2)$ if and only if its bias set is the empty set.

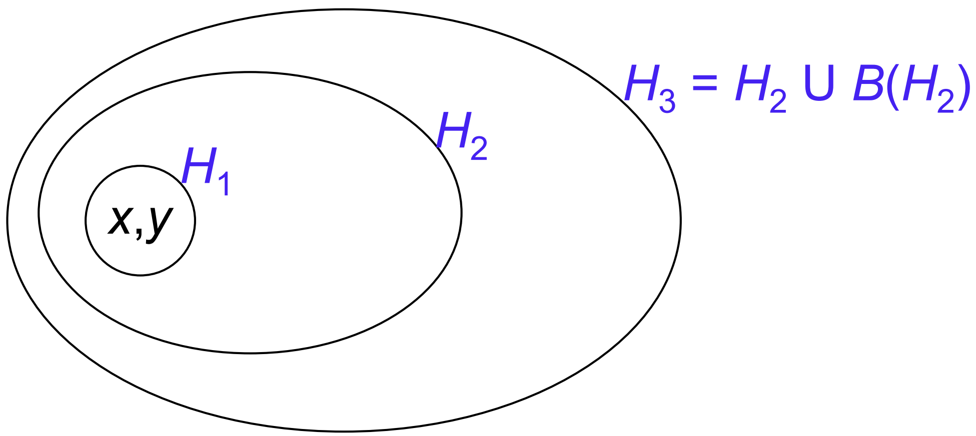
Bias Envelopment



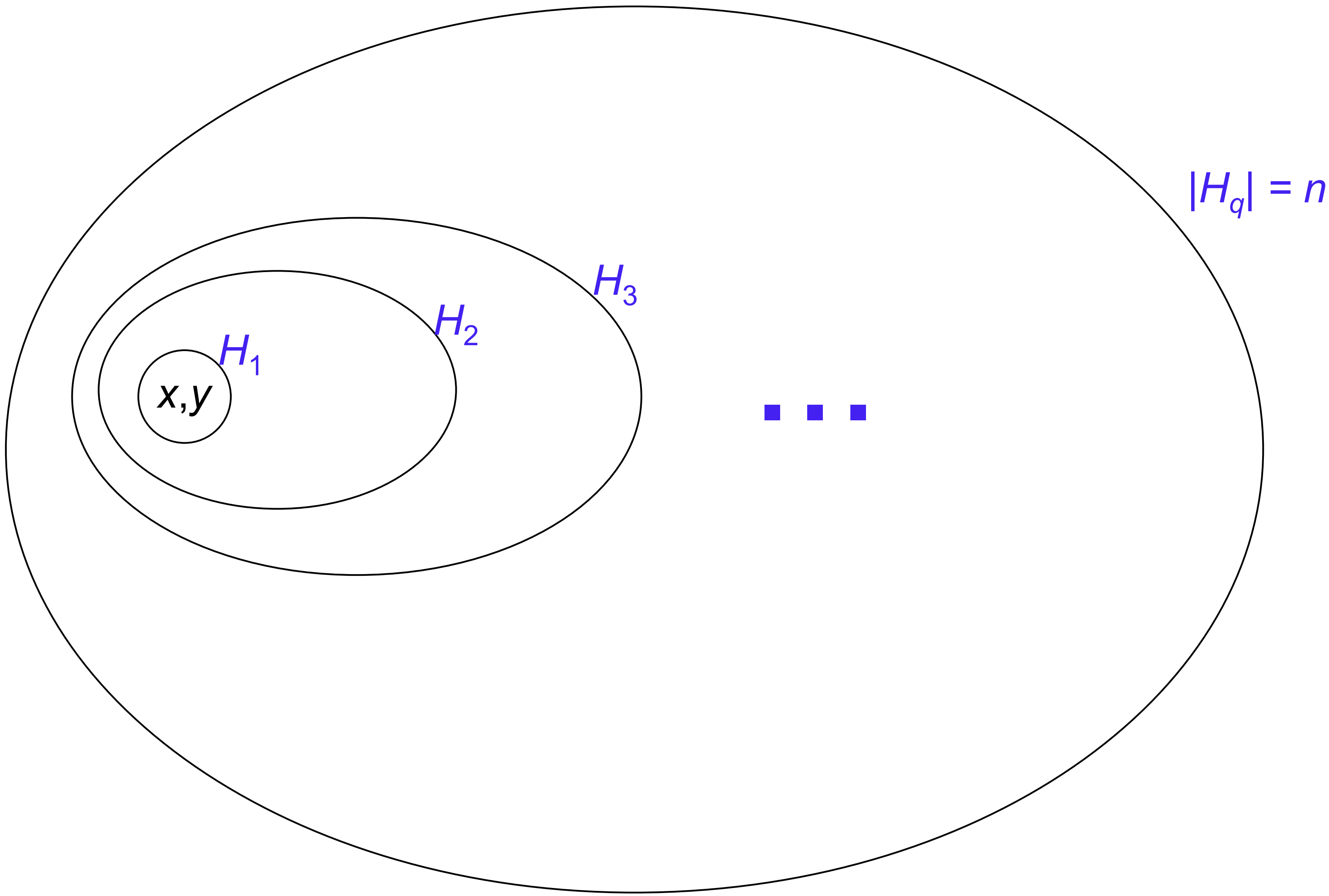
Bias Envelopment



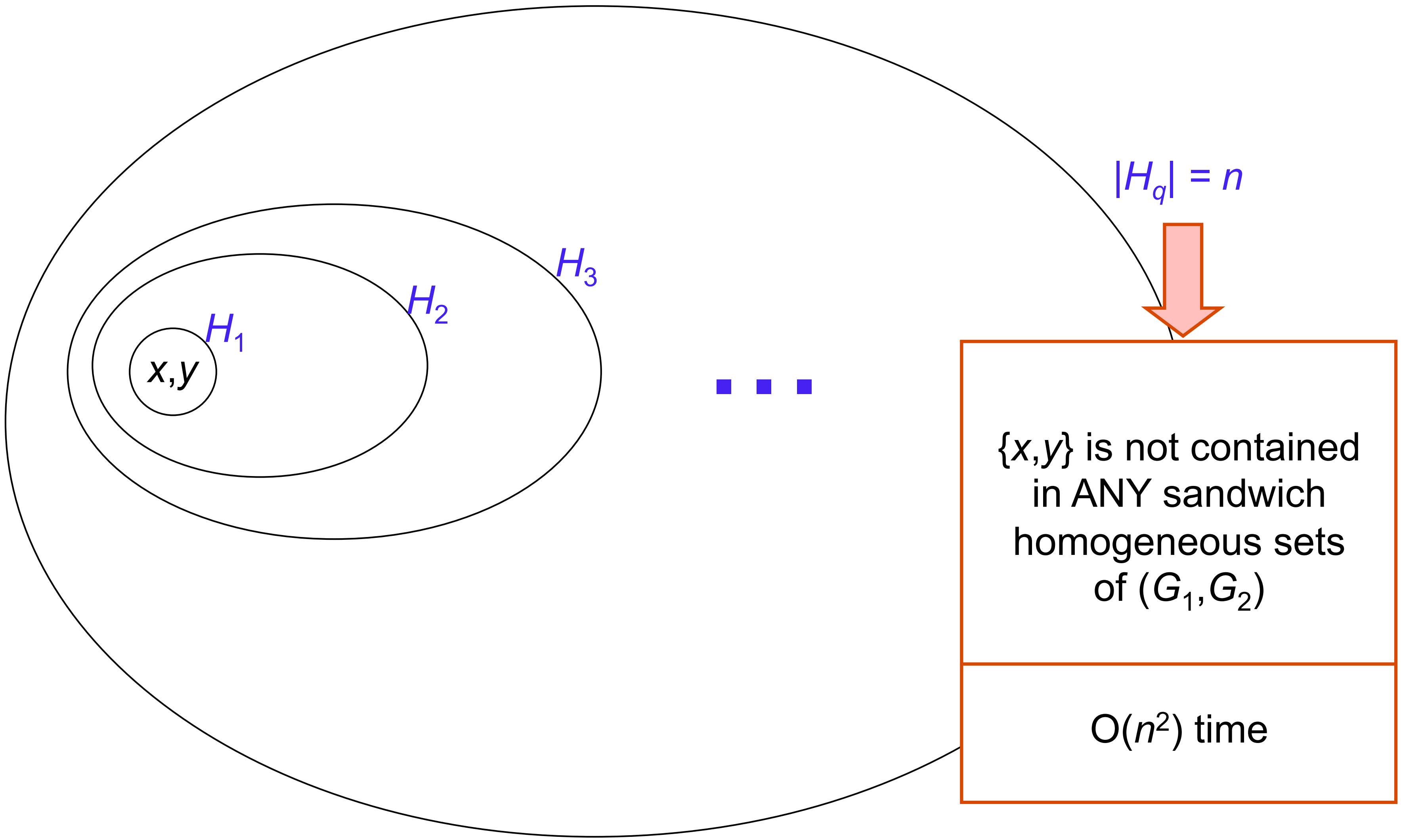
Bias Envelopment



Bias Envelopment



Bias Envelopment



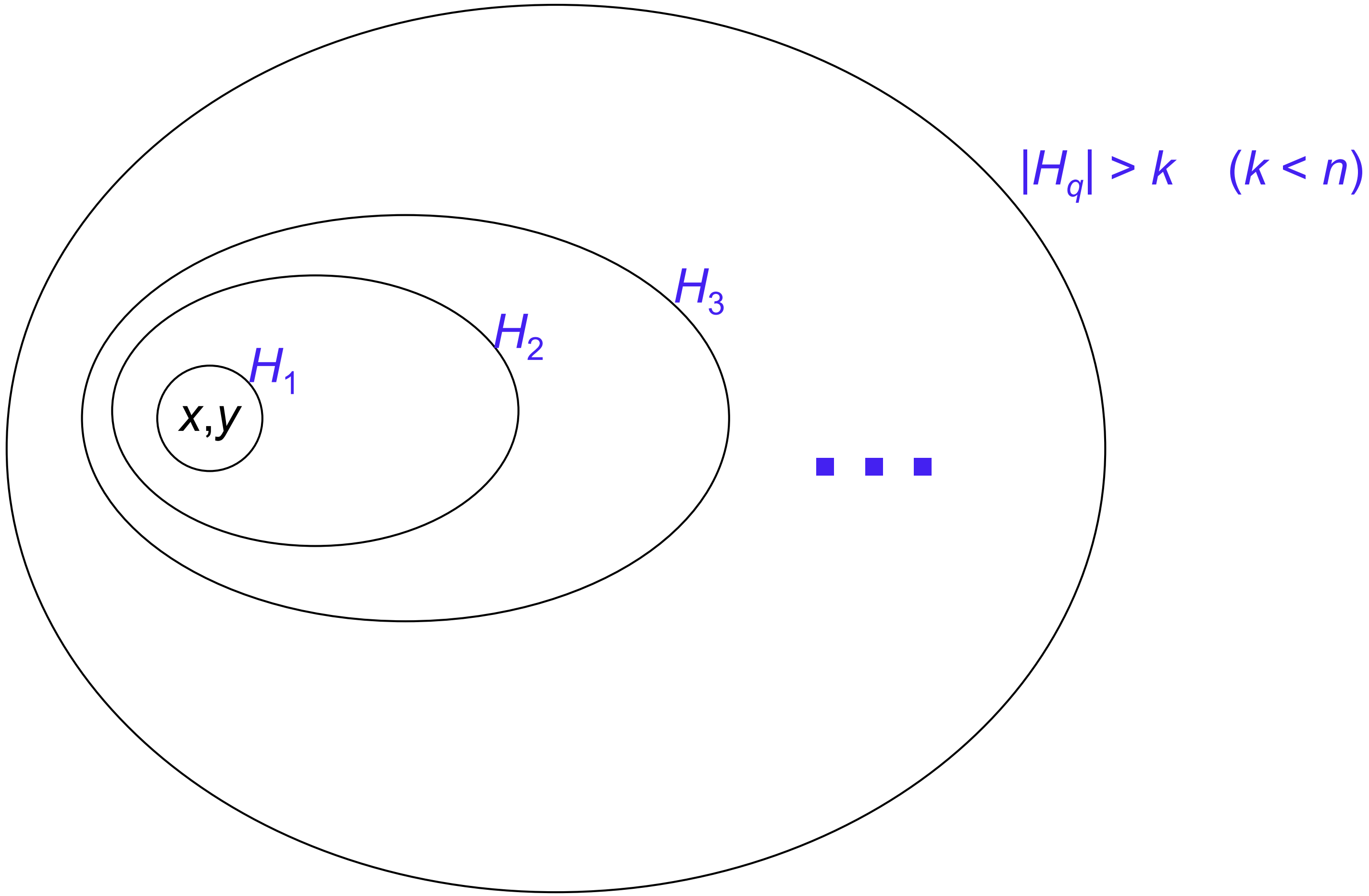
The Exhaustive Bias Envelopment algorithm

(CERIOLI, EVERETT, FIGUEIREDO, KLEIN, 1998)

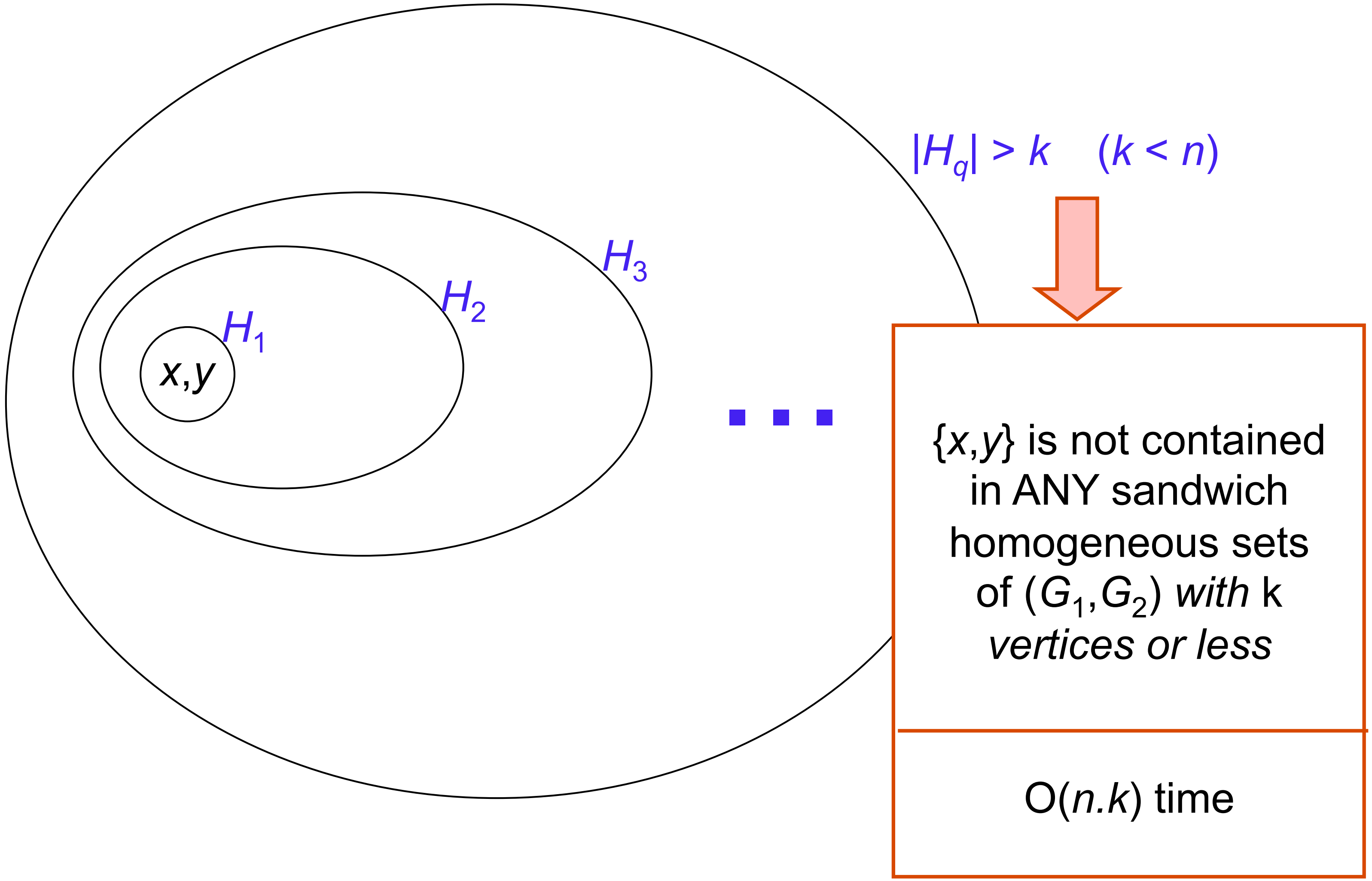
1. For each pair of vertices $x, y \in V$ do
   1.1. $H \leftarrow \{x, y\}$
   1.2. While $|H| < n$ do
      1.2.1. Find the bias set T of H
      1.2.2. If $T = \emptyset$ then
         1.2.1.1 Return YES.
      1.2.3. $H \leftarrow H \cup T$
2. Return NO.

Bias Envelopment

Incomplete Bias Envelopment



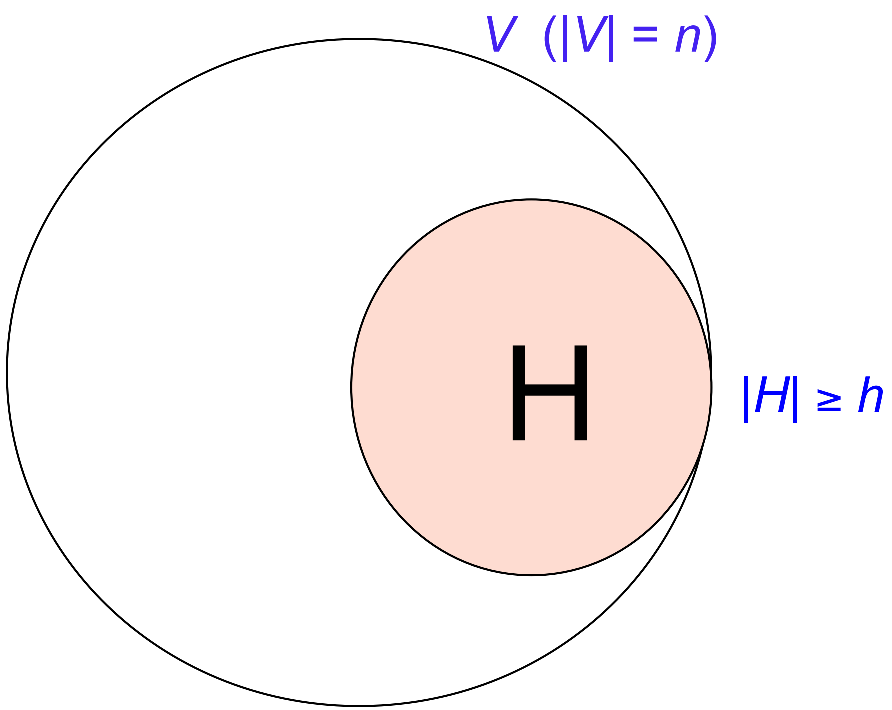
Incomplete Bias Envelopment



The Monte Carlo HSSP algorithm

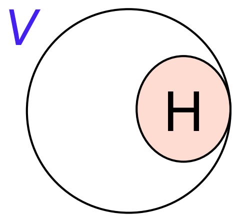
Let $G_1(V,E_1)$, $G_2(V,E_2)$ be an input instance for the HSSP.

Suppose there is a sandwich homogeneous set $H \subset V$ with h vertices or more.



The Monte Carlo HSSP algorithm

Hypothesis:

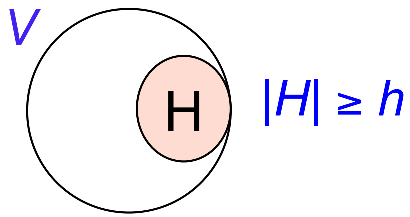


$|H| \geq h$

What is the probability $\overline{p_1}$ that a random pair of vertices $\{x,y\} \subset V$ is NOT contained in H ?

The Monte Carlo HSSP algorithm

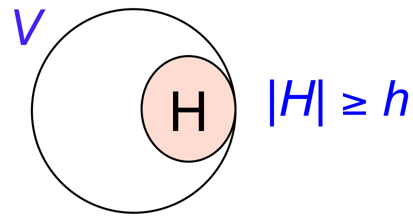
Hypothesis:



What is the probability $\overline{p_1}$ that a random pair of vertices $\{x,y\} \subset V$ is NOT contained in H ?

$$\overline{p_1} \leq 1 - \frac{h(h-1)}{n(n-1)}$$

The Monte Carlo HSSP algorithm

Hypothesis: 

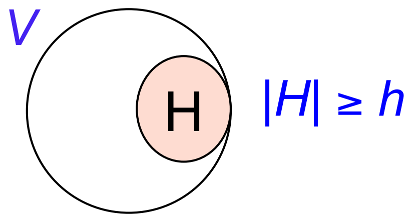
What is the probability $\overline{p_1}$ that a random pair of vertices $\{x,y\} \subset V$ is NOT contained in H ?

$$\overline{p_1} \leq 1 - \frac{h(h-1)}{n(n-1)}$$

What is the probability $\overline{p_t}$ that t random pairs of vertices fail to be contained in H ?

The Monte Carlo HSSP algorithm

Hypothesis:



What is the probability $\overline{p_1}$ that a random pair of vertices $\{x,y\} \subset V$ is NOT contained in H ?

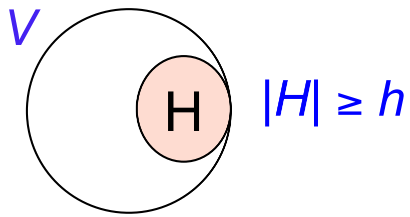
$$\overline{p_1} \leq 1 - \frac{h(h-1)}{n(n-1)}$$

What is the probability $\overline{p_t}$ that t random pairs of vertices fail to be contained in H ?

$$\overline{p_t} \leq \left(1 - \frac{h(h-1)}{n(n-1)}\right)^t$$

The Monte Carlo HSSP algorithm

Hypothesis:



What is the probability $\overline{p_1}$ that a random pair of vertices $\{x,y\} \subset V$ is NOT contained in H ?

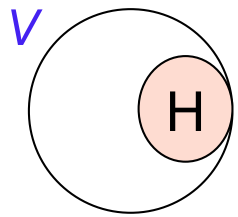
$$\overline{p_1} \leq 1 - \frac{h(h-1)}{n(n-1)}$$

What is the probability $\overline{p_t}$ that t random pairs of vertices fail to be contained in H ?

$$\overline{p_t} \leq \left(1 - \frac{h(h-1)}{n(n-1)}\right)^t$$

Now, what is the probability p_t that, running t Bias Envelopment procedures (starting from t random pairs of vertices), a sandwich homogeneous set is successfully found?

The Monte Carlo HSSP algorithm

Hypothesis: 

$|H| \geq h$

What is the probability $\overline{p_1}$ that a random pair of vertices $\{x,y\} \subset V$ is NOT contained in H ?

$$\overline{p_1} \leq 1 - \frac{h(h-1)}{n(n-1)}$$

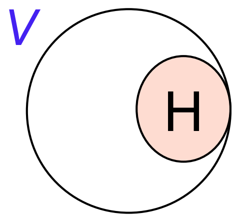
What is the probability $\overline{p_t}$ that t random pairs of vertices fail to be contained in H ?

$$\overline{p_t} \leq \left(1 - \frac{h(h-1)}{n(n-1)}\right)^t$$

Now, what is the probability p_t that, running t Bias Envelopment procedures (starting from t random pairs of vertices), a sandwich homogeneous set is successfully found?

$$p_t \geq 1 - \left(1 - \frac{h(h-1)}{n(n-1)}\right)^t$$

The Monte Carlo HSSP algorithm

Hypothesis: 

$|H| \geq h$

$$p_t \geq 1 - \left(1 - \frac{h(h-1)}{n(n-1)}\right)^t$$

Fix $p_t \geq \boldsymbol{p} = 1 - \varepsilon$

$$h(t) = \left\lfloor \frac{1 + \sqrt{1 + 4(n^2 - n)(1 - (1-p)^{1/t})}}{2} \right\rfloor$$

Running the Bias Envelopment on *t* random pairs suffices to find a sandwich homogeneous set with probability at least ***p***, *in case there exists any with* h(t) *vertices or more*.

The Monte Carlo HSSP algorithm

But the algorithm is meant to find one, if there exists any, *no matter its size*.

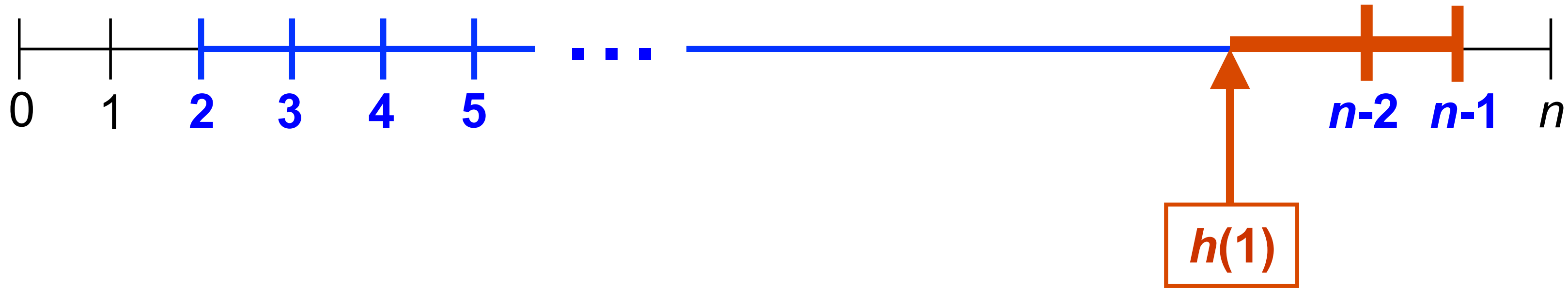
What is the number *t'* of Bias Envelopment procedures (on random pairs) that grants this?

$$h(t') = 2$$

The Monte Carlo HSSP algorithm

Number *t* of Bias Envelopment procedures undertaken on random pairs of vertices:	Minimum integer *h*(*t*) such that *t* Bias Envelopment executions (on random pairs) suffice to find some sandwich homogeneous set, *in case there exists any with* h(t) *vertices or more*:
0	we don't know anything

The Monte Carlo HSSP algorithm



Number t of Bias Envelopment procedures undertaken on random pairs of vertices:	Minimum integer $h(t)$ such that t Bias Envelopment executions (on random pairs) suffice to find some sandwich homogeneous set, *in case there exists any with* h(t) *vertices or more*:
0	we don't know anything
1	$h(1)$

The Monte Carlo HSSP algorithm

Number t of Bias Envelopment procedures undertaken on random pairs of vertices:	Minimum integer $h(t)$ such that t Bias Envelopment executions (on random pairs) suffice to find some sandwich homogeneous set, *in case there exists any with* h(t) *vertices or more*:
0	we don't know anything
1	$h(1)$
2	$h(2)$

The Monte Carlo HSSP algorithm

Number t of Bias Envelopment procedures undertaken on random pairs of vertices:	Minimum integer $h(t)$ such that t Bias Envelopment executions (on random pairs) suffice to find some sandwich homogeneous set, *in case there exists any with* h(t) *vertices or more*:
0	we don't know anything
1	$h(1)$
2	$h(2)$
...	...

The Monte Carlo HSSP algorithm

Number *t* of Bias Envelopment procedures undertaken on random pairs of vertices:	Minimum integer *h*(*t*) such that *t* Bias Envelopment executions (on random pairs) suffice to find some sandwich homogeneous set, *in case there exists any with* h(t) *vertices or more*:
0	we don't know anything
1	*h*(1)
2	*h*(2)
...	...
t'	*h*(*t'*) = 2

The Monte Carlo HSSP algorithm

Determining *t'* ...

$$h(t) = \left\lfloor \frac{1 + \sqrt{1 + 4(n^2 - n)(1 - (1-p)^{1/t})}}{2} \right\rfloor$$, given a fixed $\boldsymbol{p} = 1 - \varepsilon$

$$t' = \frac{\ln(1-p)}{\ln\left(1 - \frac{2}{n(n-1)}\right)} = \left(\ln \frac{1}{\boldsymbol{\varepsilon}}\right) \Theta(n^2) \;=\; \Theta(n^2)$$

The Monte Carlo HSSP algorithm

Determining *t'* ...

$$h(t) = \left\lfloor \frac{1 + \sqrt{1 + 4(n^2 - n)(1 - (1-p)^{1/t})}}{2} \right\rfloor$$, given a fixed ***p*** = 1 – ε

$$t' = \frac{\ln(1-p)}{\ln\left(1 - \frac{2}{n(n-1)}\right)} = \left(\ln \frac{1}{\boldsymbol{\varepsilon}}\right) \Theta(n^2) \;=\; \Theta(n^2)$$

But this leads to an O(n^4) algorithm!!!!!!

The Monte Carlo HSSP algorithm

Determining *t'* ...

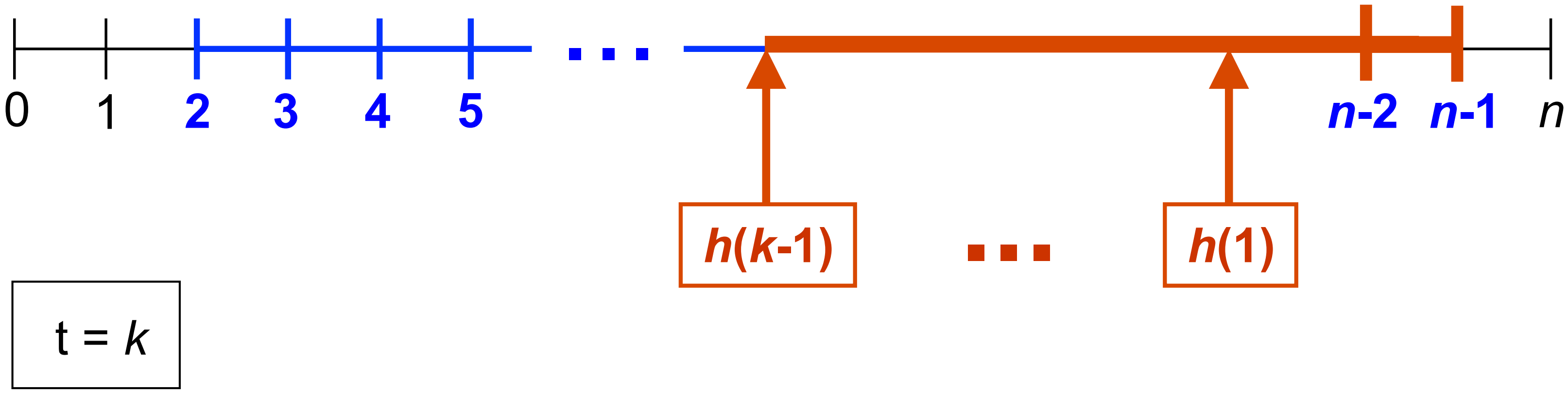
$$h(t) = \left\lfloor \frac{1 + \sqrt{1 + 4(n^2 - n)(1 - (1-p)^{1/t})}}{2} \right\rfloor$$, given a fixed $\boldsymbol{p} = 1 - \varepsilon$

$$t' = \frac{\ln(1-p)}{\ln\left(1 - \frac{2}{n(n-1)}\right)} = \left(\ln \frac{1}{\boldsymbol{\varepsilon}}\right) \Theta(n^2) \;=\; \Theta(n^2)$$

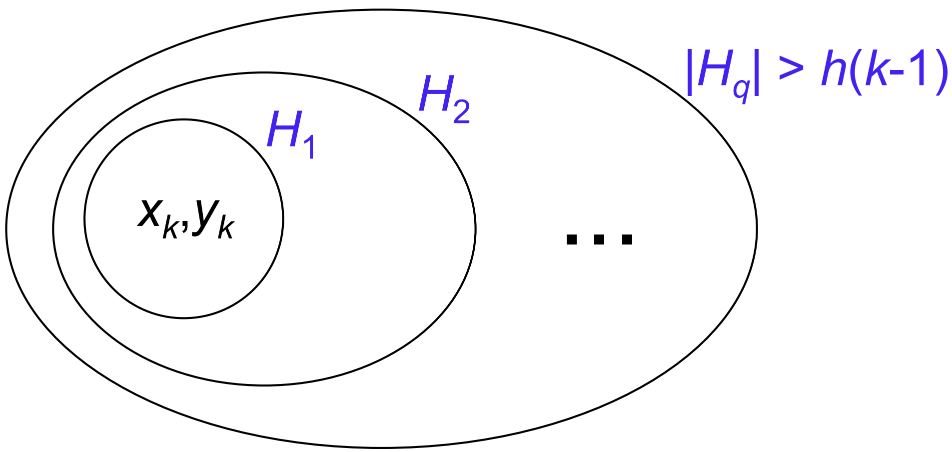
~~But this leads to an $O(n^4)$ algorithm!!!!!!~~

NO, IT DOESN'T.

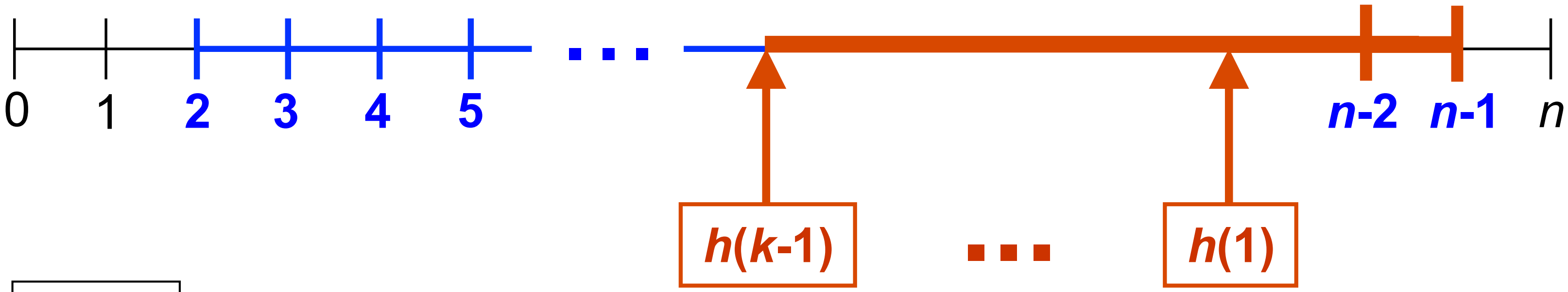
The Monte Carlo HSSP algorithm



- pick a random pair $\{x_k, y_k\}$
- Bias Envelopment



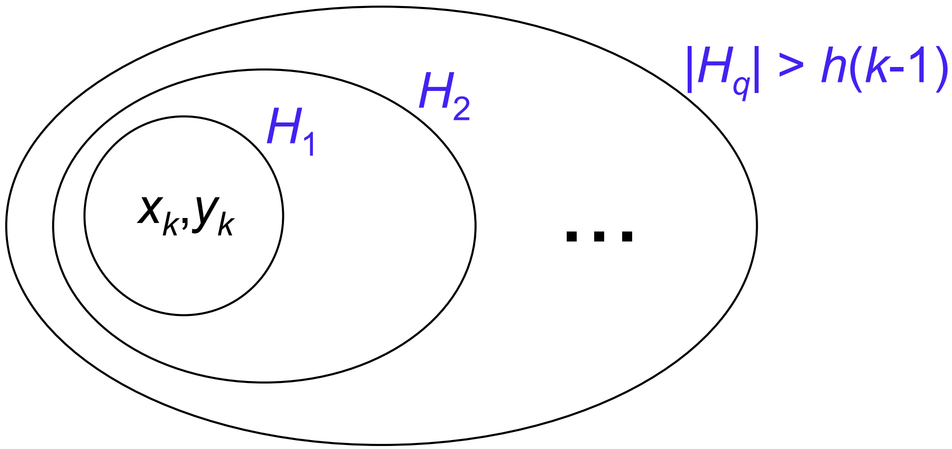
The Monte Carlo HSSP algorithm



t = k

Two possibilities:

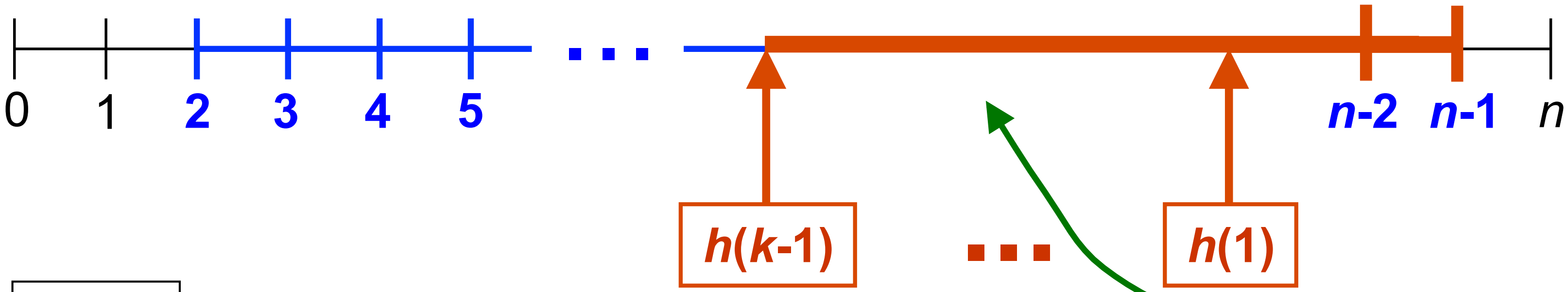
- pick a random pair $\{x_k, y_k\}$
- Bias Envelopment



(1) THERE IS a sandwich homogeneous set with more than $h(k\text{-}1)$ vertices

(2) THERE IS NO sandwich homogeneous set with more than $h(k\text{-}1)$ vertices

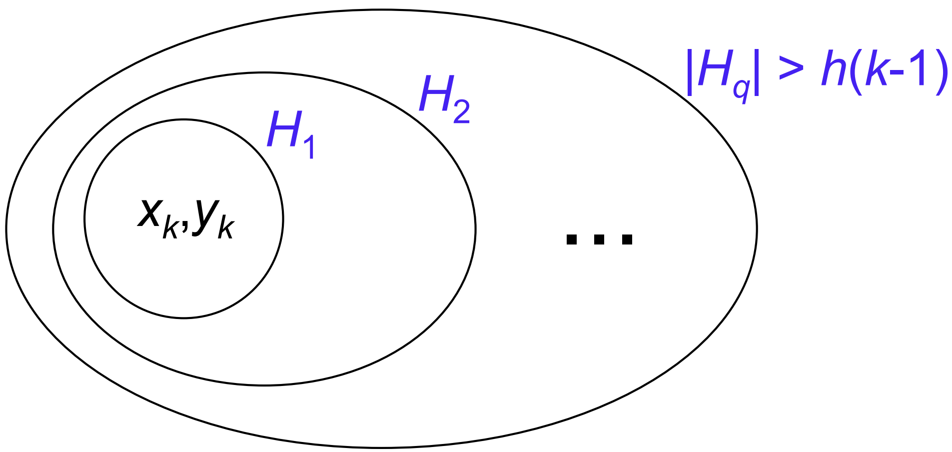
The Monte Carlo HSSP algorithm



t = k

Two possibilities:

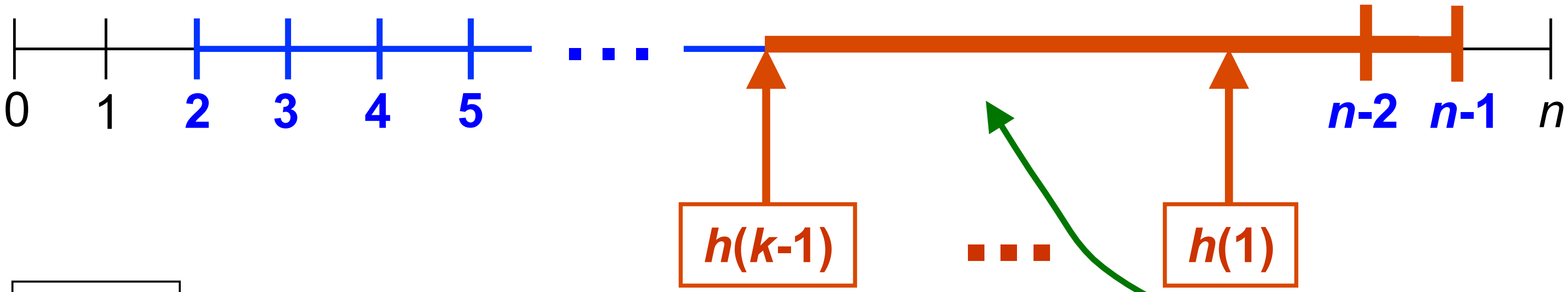
- pick a random pair $\{x_k, y_k\}$
- Bias Envelopment



(1) THERE IS a sandwich homogeneous set with more than $h(k\text{-}1)$ vertices **OK!**

(2) THERE IS NO sandwich homogeneous set with more than $h(k\text{-}1)$ vertices

The Monte Carlo HSSP algorithm



t = k

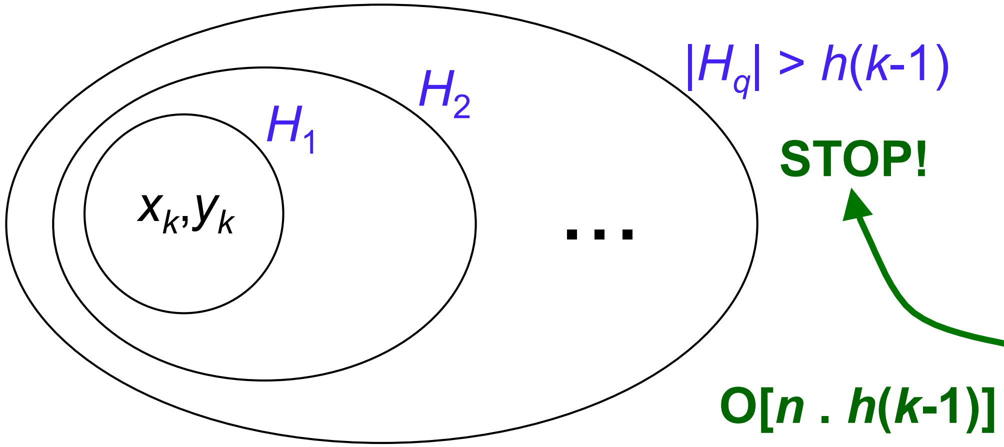
- pick a random pair $\{x_k, y_k\}$
- Bias Envelopment

Two possibilities:

(1) THERE IS a sandwich homogeneous set with more than $h(k\text{-}1)$ vertices

OK!

(2) THERE IS NO sandwich homogeneous set with more than $h(k\text{-}1)$ vertices



The Monte Carlo HSSP algorithm

1. $h \leftarrow n$
3. $t \leftarrow 0$
4. While $h > 2$ do
3.1. $t \leftarrow t + 1$
3.2. $(v_1, v_2) \leftarrow$ random pair of vertices
3.3. If *Incomplete Bias Envelopment* (v_1, v_2, h) = YES
3.3.1. Return YES.
3.4. $$h = \left\lfloor \frac{1 + \sqrt{1 + 4(n^2 - n)(1 - (1-p)^{1/t})}}{2} \right\rfloor$$
4. Return NO.

The Monte Carlo HSSP algorithm

Analysis:

$$\sum_{t=1}^{t'} O\left[nh(t)\right]$$

The Monte Carlo HSSP algorithm

Analysis:

$$\sum_{t=1}^{t'} O\left[nh(t)\right]$$

$$h(t) = \left\lfloor \frac{1+\sqrt{1+4(n^2-n)(1-(1-p)^{1/t})}}{2} \right\rfloor$$

$$h(t) \leq \Theta(n)\sqrt{\frac{1}{\Theta(t)}} = \frac{\Theta(n)}{\Theta(\sqrt{t})}$$

The Monte Carlo HSSP algorithm

Analysis:

$$\sum_{t=1}^{t'} O\big[nh(t)\big] = \sum_{t=1}^{\Theta(n^2)} \frac{O(n^2)}{O(\sqrt{t})}$$

$$h(t) = \left\lfloor \frac{1 + \sqrt{1 + 4(n^2 - n)(1 - (1-p)^{1/t})}}{2} \right\rfloor$$

$$h(t) \leq \Theta(n)\sqrt{\frac{1}{\Theta(t)}} = \frac{\Theta(n)}{\Theta(\sqrt{t})}$$

The Monte Carlo HSSP algorithm

Analysis:

$$\sum_{t=1}^{t'} O\left[nh(t)\right] = \sum_{t=1}^{\Theta(n^2)} \frac{O(n^2)}{O(\sqrt{t})} \quad = O(n^3)$$

$$h(t) = \left\lfloor \frac{1+\sqrt{1+4(n^2-n)(1-(1-p)^{1/t})}}{2} \right\rfloor$$

$$h(t) \le \Theta(n)\sqrt{\frac{1}{\Theta(t)}} = \frac{\Theta(n)}{\Theta(\sqrt{t})}$$

MUITO OBRIGADO.

THANKS.