Recall

- $H(X) = -\sum P(X=x) \log_2 P(X=x)$
- $H(X|Y) = H(X,Y) - H(Y)$
  $= \mathbb{E}_y H(X|Y=y)$
- $0 \le H(X) \le \log_2 |supp(X)|$ "=" iff X unif
- $H(X,Y) \le H(X) + H(Y)$
- $H(X|Y) \le H(X)$
- $H(X, f(X,Y) \,|\, Y) = H(X|Y) = H(X \,|\, Y, f(Y))$ for deterministic f
- $H(f(X) \,|\, Y) \le H(X|Y) \le H(X \,|\, f(Y))$
- Bit-wise entropy calculation

Graph hom

homomorphism

$$Hom(H,G) = \{ f : V(H) \to V(G) \mid f(u)f(v) \in E(G) \ \forall uv \in E(H) \}$$

$$hom(H,G) = |Hom(H,G)|$$

hom density $$t(H,G) = \frac{t(H,G)}{|V(G)|^{|V(H)|}}$$

Sidorenko

Def A graph H is Sidorenko if

$$t(H,G) \ge t(K_2,G)^{|E(H)|} \quad \forall G$$

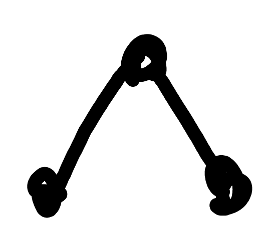
Remark • If $G = G(n,p)$, then $\mathbb{E}\, t(H,G) = p^{|E(H)|}$

$$= \left(\mathbb{E}\, t(e,G)\right)^{|E(H)|}$$

Roughly speaking, Sidorenko property says random graph minimizes the density of H

• If H NOT bipartite, then H NOT sidorenko

pf take $G = e$

Thm P_2 :  sido.

pf $\hom(P_2,G) = \sum_v \deg(v)^2 \geq \frac{\left(\sum \deg(v)\right)^2}{n}$, $n = |V(G)|$

$$= \frac{(2m)^2}{n} \qquad m = |E(G)|$$

$$t(K_2,G) \leq \frac{(2m)^2}{n} \Big/ n^3 = \left(\frac{2m}{n^2}\right)^2 = t(e,G)^2$$

Alt pf Take $XY \in E(G)$ u.a.r.,

Condition on Y, resample X indep. and get Z

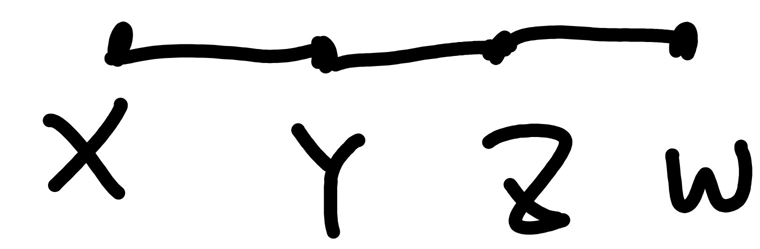
$\left(\text{i.e. } P(Z=z, X=x \mid Y=y) = P(X=z \mid Y=y) \cdot P(X=x \mid Y=y) \right)$

$$\Rightarrow H(X,Z \mid Y) = H(X \mid Y) + H(Z \mid Y) = 2H(X \mid Y)$$

$$\underbrace{H(XYZ)}_{\leq \log \hom(P_2,G)} = 2H(X \mid Y) + H(Y) = \underbrace{2H(XY)}_{= \log \hom(K_2,G)} \underbrace{-\, H(Y)}_{\geq -\log_2 n} \qquad (n = |V(G)|)$$

Thm P_3 sido

pf Take $XY \in E(G)$ u.a.r.



Condition on Y, resample X indeply. get Z

Condition on Z, resample Y indeply, get W

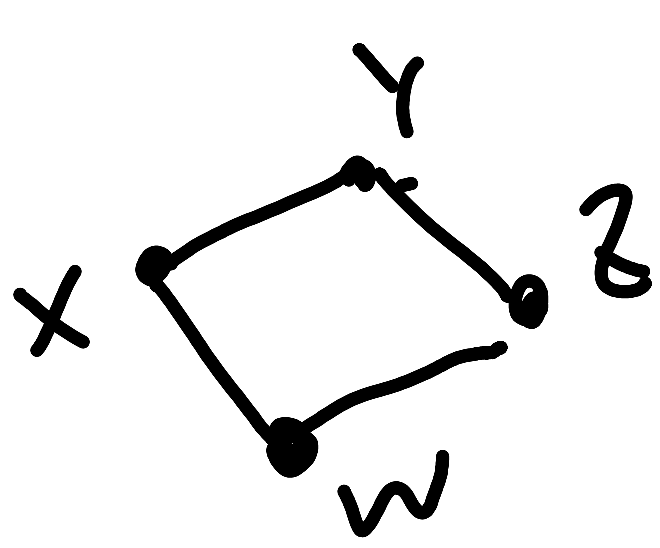
$$H(X,Y,Z,W) \le \log_2(\hom(P_3,G))$$

$$= \underbrace{H(X,Y,Z)}_{=H(X,Y)+H(Z|Y)} + \underbrace{H(W|X,Y,Z)}_{=H(W|Z)} = \underbrace{H(X,Y)+H(Y,Z)+H(Z,W)}_{=3\log_2(2m)} \underbrace{-H(Y)-H(Z)}_{\ge -2\log_2 n}$$

$$\Rightarrow \frac{\hom(P_3,G)}{n^4} \ge \frac{(2m)^3}{n^6} \quad \square$$

Thm C_4 sido.

pf



Condition on X, Z, resample Y indeply. get W

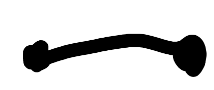
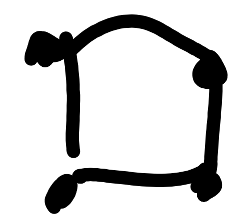
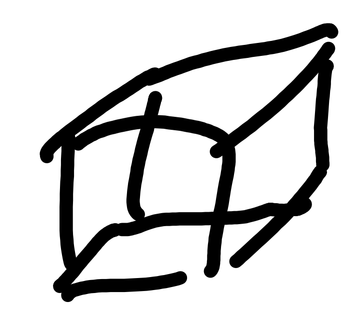
$$H(X,Y,Z,W) \le \log_2(\hom(P_4,G))$$

$$= H(X,Y,Z) + H(W|X,Z) = 2H(X,Y,Z) - H(X,Z)$$

$$= 4H(X,Y) - 2H(Y) - H(X,Z)$$

$H(X,Z) \le H(X) + H(Z) \le 2\log_2 n$

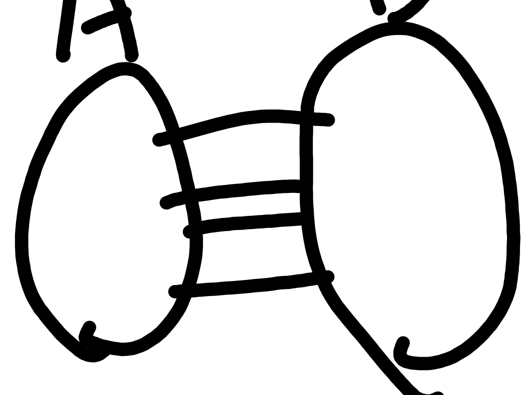
$$\Rightarrow \frac{\hom(C_4, G)}{n^4} \ge \frac{(2m)^4}{n^8}$$

Thm $Q_0 =$  , $Q_1 =$  , $Q_2 =$  , $Q_3 =$ 

Q_n sido.

pf Assume we sampled Q_k so that the marginal distribution of $A \sim B$. Sample Q_{k+1} by

A B

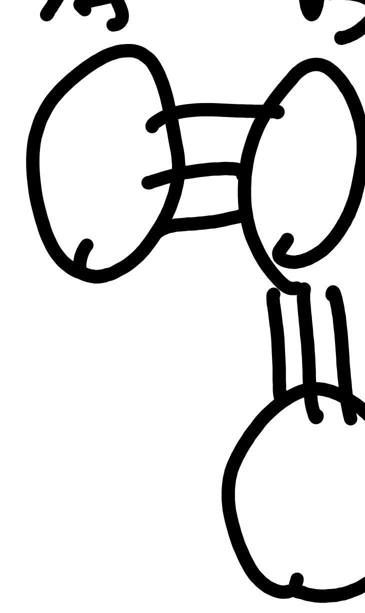


Q_{k-1} Q_{k-1}

Condition on B resample A indeply

$\rightsquigarrow$

A B

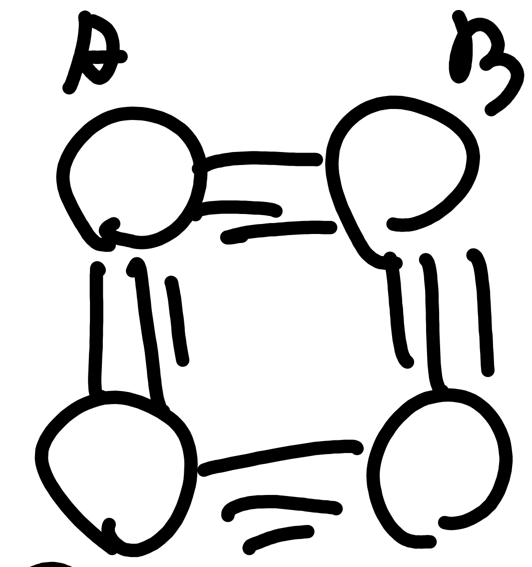


C

Condition on A, C, resample B indeply

$\rightsquigarrow$

A B



D C

Q_{k+1}

$$\Rightarrow H(Q_{k+1}) = H(A,B,C) + H(D \mid A,C)$$

$$\ge 4H(A,B) - 4H(A) = 4H(Q_k) - 4H(Q_{k-1})$$

Let $D_k = H(Q_k) - 2^k \log_2 n \le \log_2 \hom(Q_k, G)$

$$\Rightarrow D_{k+1} \ge 4D_k - 4D_{k-1}$$

$$D_n \geq 4D_{n-1} - 4D_{n-2}$$

$$4D_{n-1} \geq 16 D_{n-2} - 16 D_{n-3}$$

$$\vdots$$

$$(k+1)\cdot 2^k D_{n-k} \geq (k+1)2^{k+2} D_{n-k-1} - (k+1)2^{k+2} D_{n-k-2}$$

$$\vdots$$

+)

$$D_n \geq (n-1)2^n D_1 - (n-1)2^n D_0$$

$$= (n-1)2^n \log_2\left(\frac{2m}{n}\right)$$

$D_1 = \log_2\left(\frac{2m}{n}\right)$, $D_0 = 0$ $\square$

---

hom density of trees

Thm (Sidorenko)

Let T be a tree w/ k vtxs,

$$t(T, G) \leq t(K_{1,k-1}, G) \quad \forall G$$

$$\hom(T, G) \leq \hom(K_{1,k-1}, G)$$

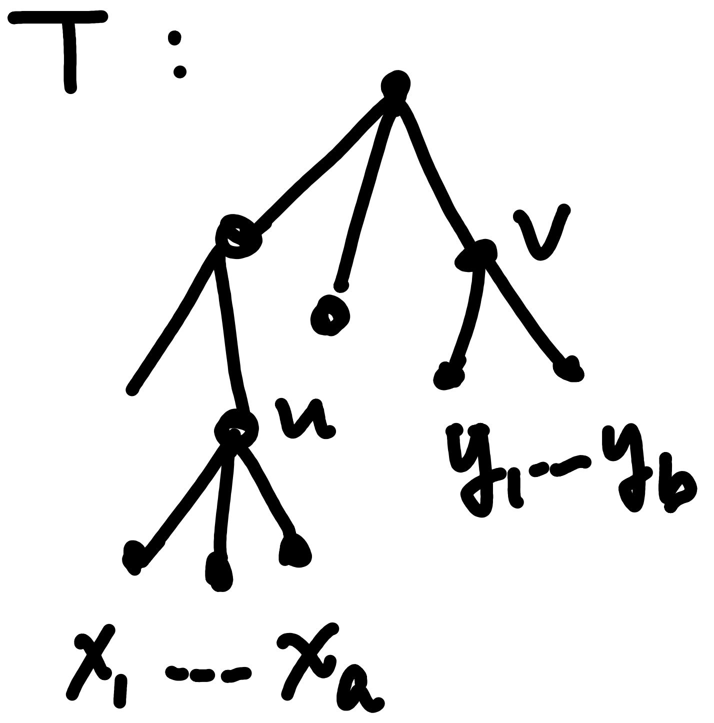
Pf **Claim** Fix T, G, there exists T': tree w/ k vtxs and has more leaves than T s.t.

$$\hom(T, G) \leq \hom(T', G)$$

Pf Pick $X_T = (X_v)_{v \in V(T)} \in \mathrm{Hom}(T, G)$ unif at random.

Pick two vtxs u, v that are distance one from some leaves in T

T :



$$H(X_T) = H(X_{\underbrace{T\setminus\{x_1\cdots x_a\, y_1\cdots y_b\}}_{\tilde T}}) + H(X_{x_1}\cdots X_{x_a}X_{y_1}\cdots X_{y_b} | X_{\tilde T})$$

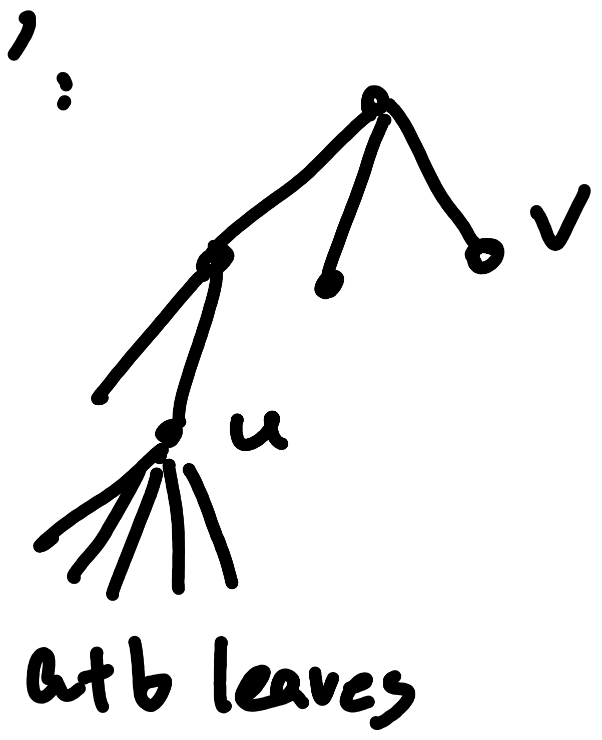
$$\le H(X_{\tilde T}) + aH(X_x | X_{\tilde T}) + bH(X_y | X_{\tilde T})$$

(actually $=$)

$$\le H(X_{\tilde T}) + aH(X_x | X_u) + bH(X_y | X_v)$$

<u>Case 1</u> $H(X_x | X_u) \ge H(X_y | X_v)$

T':



$$H(X_T) \le H(X_{\tilde T}) + (a+b) H(X_x | X_u)$$

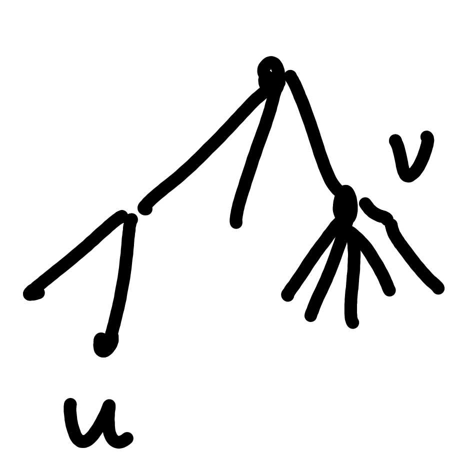
resample $X_{x_{a+i}} | X_u \underset{\text{C.I.}}{\sim} X_{x_1} | X_u$, $i = 1, \cdots, b$

$$H(\underbrace{X_{\tilde T}, X_{x_1}, \cdots X_{x_{a+b}}}_{\text{supported on } Hom(T', G)}) = H(X_{\tilde T}) + (a+b) H(X_x | X_u)$$

$\Rightarrow \log_2 hom(T, G) = H(X_T) \le H(\,\text{〃}\,) \le \log_2 Hom(T', G)$ □

<u>Case 2</u> $H(X_x | X_u) \le H(X_y | X_v)$

T' :



□

Shearer's inequality

Thm $H(X,Y,Z) \le H(X,Y) + H(Y,Z) + H(X,Z)$

pf $H(X,Y,Z) = H(X,Y) + \underline{H(Z|X,Y)}$

$\quad \text{"} \quad = H(Y,Z) + H(X|Y,Z)$

$\quad \text{"} \quad = H(X,Z) + \underline{H(Y|X,Z)}$

$$H(Z|XY) + H(X|YZ) + H(Y|XZ) \le H(Z) + H(X|Z) + H(Y|XZ)$$
$$= H(XYZ)$$

Thm Suppose $A_1, \dots, A_m \subseteq [n]$ covers i at least k times $\forall i \in [n]$

then $k H(X_1, \dots, X_n) \le \sum_{i=1}^{m} H(X_{A_i})$

where $H(X_A) : (X_{a_1}, \dots, X_{a_t})$ if $A = \{a_1, \dots, a_t\}$

pf Submodularity $\Rightarrow H(X_{A_i}) + H(X_{A_j}) \ge H(X_{A_i \cup A_j}) + H(X_{A_i \cap A_j})$

Replace A_i, A_j by $A_i \cup A_j$, $A_i \cap A_j$, RHS decreases

the family still covers i $\ge k$ times $\forall i \in [n]$

Keep applying this move until we can't (stop in finite time)
why?

WLOG we get $A_1 \supseteq \cdots \supseteq A_k \supseteq \cdots \supseteq A_m$

Since it covers i $\ge k$ times $\forall i$, $A_1 = \cdots = A_k = [n]$

$\Rightarrow$ RHS $\ge k H(X_{[n]})$