Q. How good is belief propagation for $|\mathcal{X}| < \infty$?

- If G is tree $\Rightarrow$ it converges and exact
- If $\exists$ only 1 loop $\Rightarrow$ Converges but could be wrong [Weiss 2000]
- Maximum Weight Matching Problem $\Rightarrow$ Converges & Exact [Bayati, Shah, Sharma 2005]
- In the limit of large graph, <u>Density Evolution</u> provides an asymptotic performance estimate.

  - Analysis of compressed sensing

$$\min_{x \in \mathbb{R}^d} \|\underset{\text{fat = underdetermined system}}{A}x - b\|^2 + \lambda \cdot \|x\|_{L1}$$

$$\|x\|_{L1} = \sum_{i=1}^{d} |x_i|$$

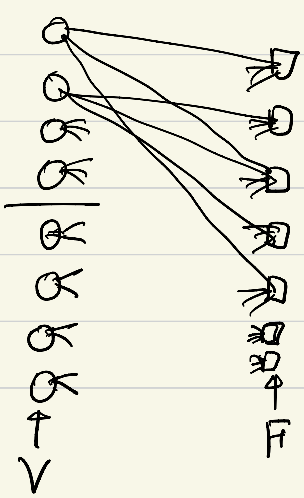
$$\boxed{A} \; \boxed{x} \leftarrow \text{sparse.}$$

  - Analysis of LDPC decoders
    $\Rightarrow$ Design of the first provably capacity achieving codes [Luby 1998]
  - Analysis of near-optimal crowdsourcing algorithm [KargerOhShah 2011].
  - Community detection for stochastic block models.

* Assumption (for density evolution): Factor Graph model on Random Graphs.

Def. Random Graph $(n, m, l = [l_1, l_2, l_3, \dots, l_{d_l}], r = [r_1, r_2, r_3, r_4, r_{d_r}])$

$n = |V|$, $m = |F|$, $l = [0, .5, .5, \dots, 0]$, $r = [0, 0, 0, 1]$.



s.t. $n \cdot \bar{l} = m \cdot \bar{r} = |E|$.

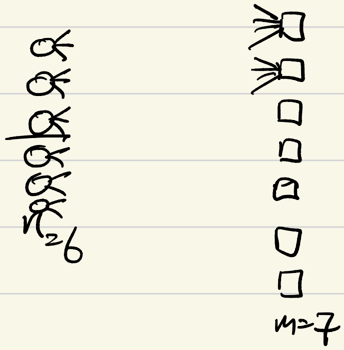
$\bar{l} = \sum_i l_i \cdot i$, $\quad \bar{r} = \sum_j r_j \cdot j$

G is drawn from Random Graph (n, m, l, r) if it is chosen uniformly over all graphs satisfying the (l, r) degree distribution.

• given (l, r). how do you generate such graphs?

Def. Configuration model. $RG(n, m, l, r)$

$$\begin{array}{cc} l_1 \\ l_2 \\ l_3 \\ l_4 \end{array}\begin{bmatrix} 0 \\ 0 \\ .5 \\ .5 \end{bmatrix} \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}\begin{array}{c} r_1 \\ r_2 \\ r_3 \\ r_4 \end{array}$$



$\bar{l} = 3.5 \quad \bar{r} = 4$

$n = 6 \quad m = 7 \qquad |E| = 21$

① draw a Random Permutation π from $[|E|] \to [|E|]$
$\{1, 2, \cdots 21\} \quad \{1, 2, \cdots, 21\}$

② and match left half edges to the right w.r.t π.

③ locally fix double-edges.

* When do we see random graphs? when we have the control to design a system.

Density Evolution for Random Graphs.

ex) LDPC codes.

Factor Graph Model: $X_i \in \{0, 1\}$, $Y_i \in \{0, 1, *\}$ ← erasure.

$\mathbb{P}(Y_i = * \mid X_i) = \varepsilon$

$\mathbb{P}(Y_i = X_i \mid X_i) = 1 - \varepsilon$.

$$P_Y(x) = \frac{1}{Z} \cdot \underbrace{\prod_{i=1}^{n} \mathbb{P}(Y_i \mid X_i)}_{\text{channel}} \cdot \prod_{a \in F} \underbrace{\mathbb{I}(\oplus X_a = 0)}_{\text{LDPC Parity}}$$

$\oplus$: XOR of all binary variables.

Q. what is the error achieved by BP to get $\hat{P}_Y(X_i)$'s.

* Belief Propagation on LDPC codes.

$$m_{i\to a}(X_i) = \mathbb{P}(Y_i, X_i)\cdot \prod_{b\in\partial i\setminus\{a\}} \tilde{m}_{b\to i}(X_i)$$

$$\mathbb{P}(Y_i,X_i) = \begin{cases} \varepsilon & Y_i = * \\ 1-\varepsilon & Y_i = X_i \\ 0 & Y_i = \text{not } X_i \end{cases}$$

binary operation

$$\tilde{m}_{a\to i}(X_i) = \sum_{X_{\partial a\setminus\{i\}}} \left\{\prod_{j\in\partial a\setminus\{i\}} m_{j\to a}(X_j)\right\} \mathbb{I}(\oplus X_{\partial a}=0)$$

claim: $m_{i\to a}(X_i) \in \left\{ \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}\frac12\\\frac12\end{bmatrix} \right\}$ ← The messages are Discrete.

Def. $X_{i\to a}=0$ $\quad X_{i\to a}=1$ $\quad X_{i\to a}=*$

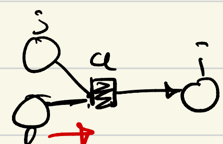
proof: by induction

Init: $\tilde{m}^{(0)}_{b\to i}(X_i)$ is initialized as $\begin{bmatrix}\frac12\\\frac12\end{bmatrix}$,

$m^{(1)}_{i\to a}$ tell you what you received from channel. $\{0, 1, *\}$

$$m^{(1)}_{i\to a}(X_i) = \begin{cases} \begin{bmatrix}\varepsilon\\\varepsilon\end{bmatrix} \propto \begin{bmatrix}\frac12\\\frac12\end{bmatrix} & \text{if } Y_i = * \\ \begin{bmatrix}1-\varepsilon\\0\end{bmatrix} \propto \begin{bmatrix}1\\0\end{bmatrix} & Y_i = 0 \\ \begin{bmatrix}0\\1\end{bmatrix} & Y_i = 1 \end{cases}$$

Induction: ① $\tilde{m}^{(t)}_{a\to i}$:



Case 1: all incoming messages are $m_{l\to a}(X_l)=\begin{bmatrix}1\\0\end{bmatrix}$, $m_{j\to a}(X_j)=\begin{bmatrix}1\\0\end{bmatrix}$ $\quad X_{l\to a}=0$ OR $X_{l\to a}=1$

then
$$\tilde{m}^{(t)}_{a\to i}(X_i) = \begin{bmatrix} m(X_j=0)\,m(X_l=0) + \cancel{m(X_j=1)\,m(X_l=1)} \\ m(X_j=0)\,\cancel{m(X_l=1)} + \cancel{m(X_j=1)\,m(X_l=0)} \end{bmatrix} = \begin{bmatrix}1\\0\end{bmatrix}$$

$\{X_l=0$ or $X_l=1$ or $X_l=*$

$\tilde{m}^{(t)}_{a\to i} = \mathbb{I}(X_i \oplus \underbrace{X_{\partial a\setminus\{i\}}}_{\text{either 0 or 1.}})$

* If incoming messages are all $\begin{bmatrix}0\\1\end{bmatrix}$ or $\begin{bmatrix}1\\0\end{bmatrix}$

then $X^{(t)}_{a\to i} = \bigoplus_{j\in\partial a\setminus\{i\}} X_{j\to a}$

Case 2: at least one incoming message is $m_{j\to a}(X_i)=\begin{bmatrix}\frac12\\\frac12\end{bmatrix}$, $X_{j\to a}=*$

then
$$\tilde{m}^{(t)}_{a\to i}(X_i) = \begin{bmatrix} \underbrace{m(X_j=0)}_{\frac12}\,\underbrace{m(X_l=0)}_{1} + m(X_j=1)\,\cancel{m(X_l=1)} \\ m(X_j=0)\,\cancel{m(X_l=1)} + \underbrace{m(X_j=1)}_{\frac12}\,\underbrace{m(X_l=0)}_{1} \end{bmatrix} = \begin{bmatrix}\frac12\\\frac12\end{bmatrix}$$

$$\tilde{m}^{(t)}_{a\to i} = \sum_{X_{j\to a}\in\{0,1\}} \frac12\, \mathbb{I}(X_i \oplus \underbrace{X_{j\to a}}_{*} \oplus \underbrace{X_{\partial a\setminus\{i,j\}}}_{\{0,1\}}) = \frac12$$

* If incoming message has at least one $\begin{bmatrix}\frac12\\\frac12\end{bmatrix}$, then $\tilde{m}^{(t)}_{a\to i} = \begin{bmatrix}\frac12\\\frac12\end{bmatrix}$

$* \quad \tilde{m}^{(t)}_{a\to i}(x_i) \in \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1/2\\1/2\end{bmatrix}$

$\tilde{x}^{(t)}_{a\to i} = 0 \ , \ 1 \ , \ *$

② $m^{(t)}_{i\to a}(x_i) = \prod_{b\in\partial i\setminus\{a\}} \hat{m}^{(t)}_{b\to i}(x_i)$

Case 1: if all incoming is $\tilde{m}_{b\to i} = \begin{bmatrix}1/2\\1/2\end{bmatrix}$, $\tilde{x}_{b\to i} = *$

then $m_{a\to i} = \begin{bmatrix}1/2\\1/2\end{bmatrix}$, $x_{a\to i} = *$

Case 2: if at least one incoming is $\tilde{m}_{b\to i} = \begin{bmatrix}0\\1\end{bmatrix}$ or $\begin{bmatrix}1\\0\end{bmatrix}$

$\tilde{x}_{b\to i} = 1 \quad 0$

then $m_{a\to i} = \tilde{m}_{b\to i} \in \begin{bmatrix}0\\1\end{bmatrix}$ or $\begin{bmatrix}1\\0\end{bmatrix}$.

With this claim, we can simplify the BP update.

$x^{(t)}_{i\to a} \in \{0, 1, *\}$, $\tilde{x}^{(t)}_{a\to i} \in \{0, 1, *\}$

initialize $\quad x_{i\to a} = y_i$

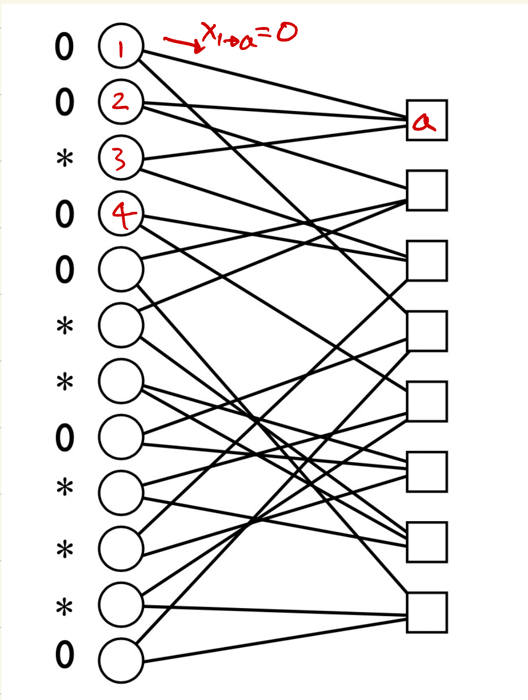
update $\quad \tilde{x}_{a\to i} = \begin{cases} * & \text{, at least one incoming is } *. \\ \oplus x_{\partial a\setminus i\to i}, & \end{cases}$

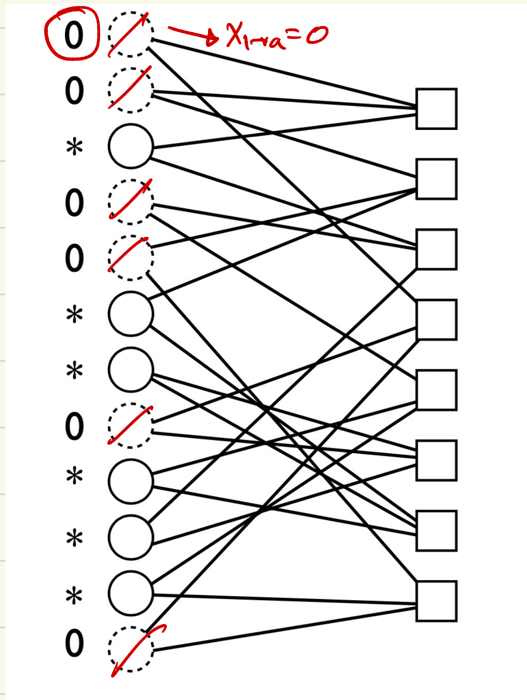
$x_{i\to a} = \begin{cases} * & \text{, all incoming is } * \\ \tilde{x}_{b\to i} & , \end{cases}$

Claim: this is equivalent to the following <u>peeling algorithm</u>.

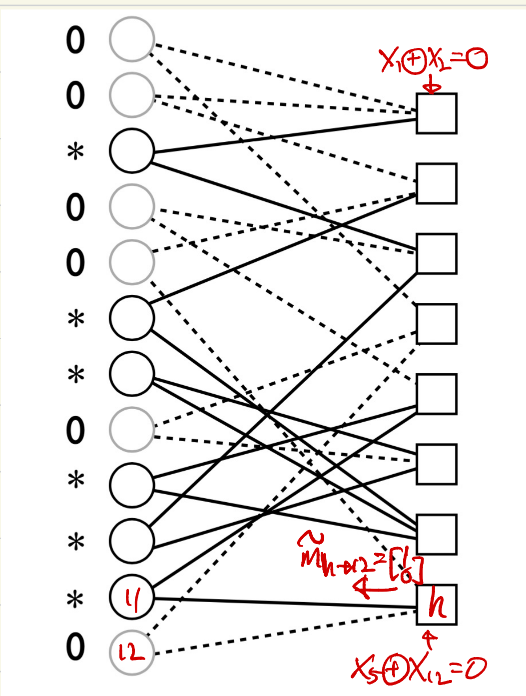
Without loss of generality, suppose all 0's were sent,

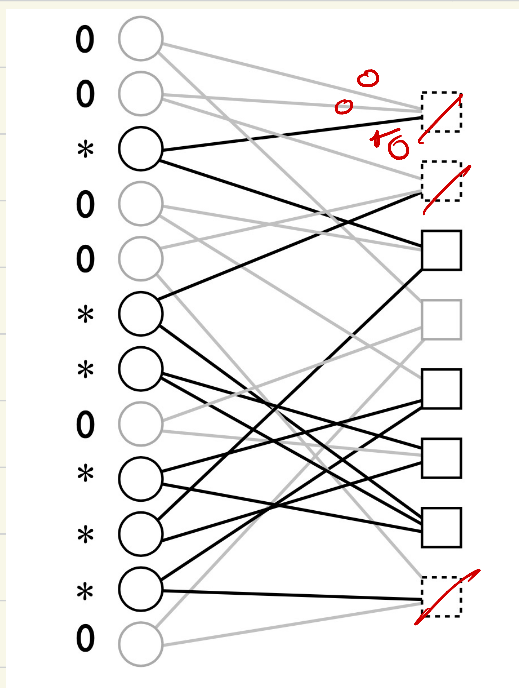
Peel nodes/edges whose messages are $\begin{bmatrix}0\\1\end{bmatrix}$ or $\begin{bmatrix}1\\0\end{bmatrix}$.



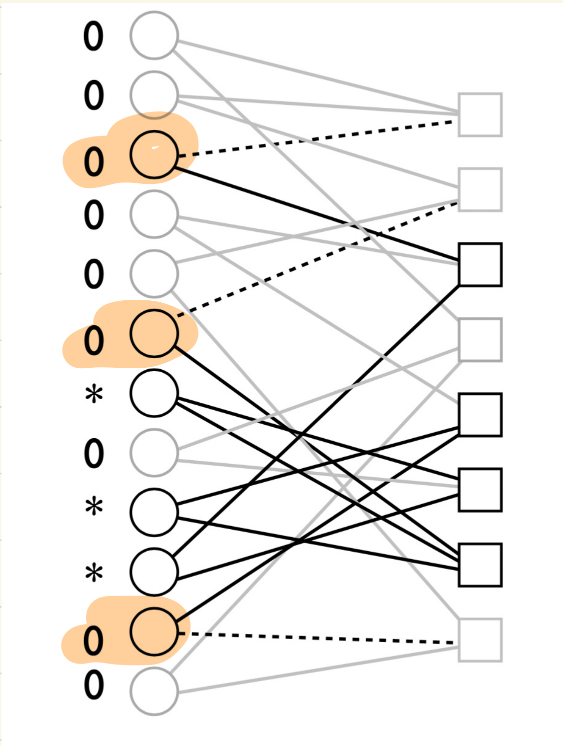


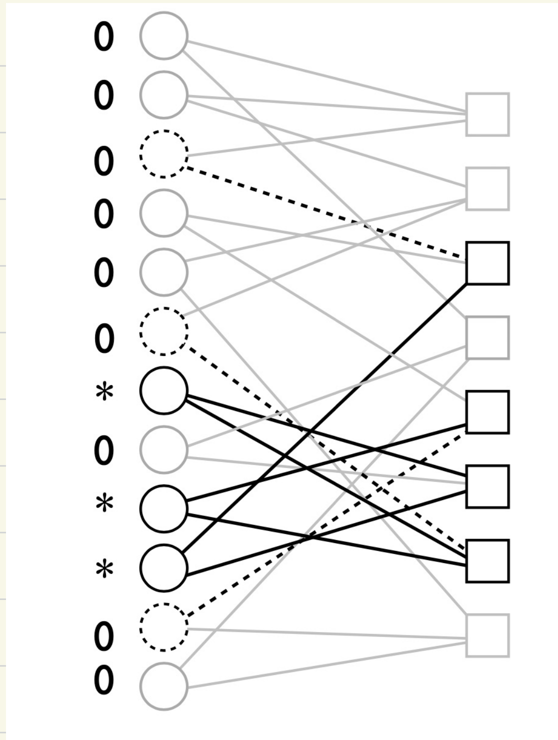
$x_1=0$

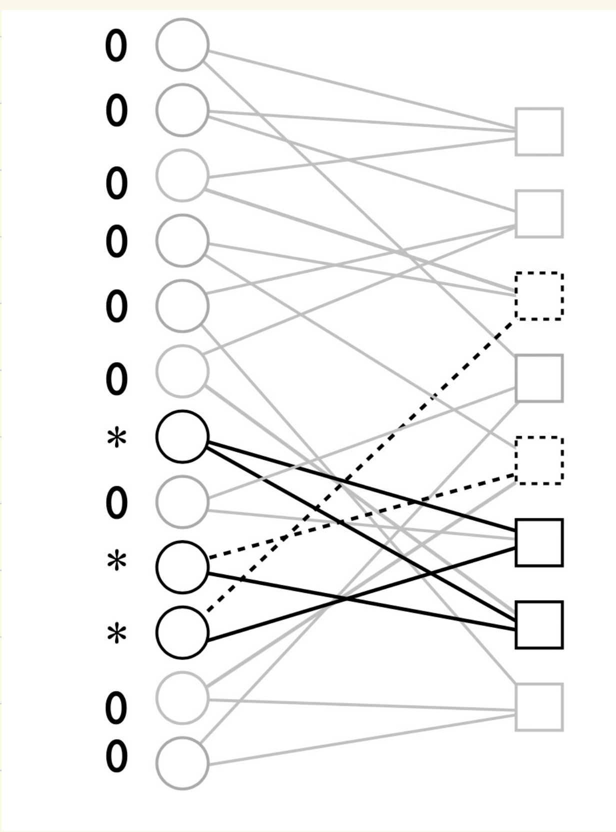


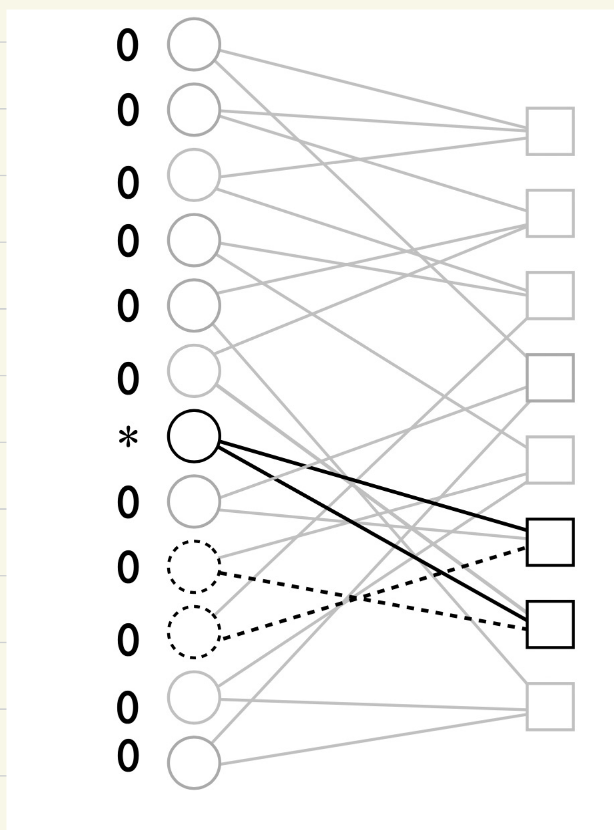


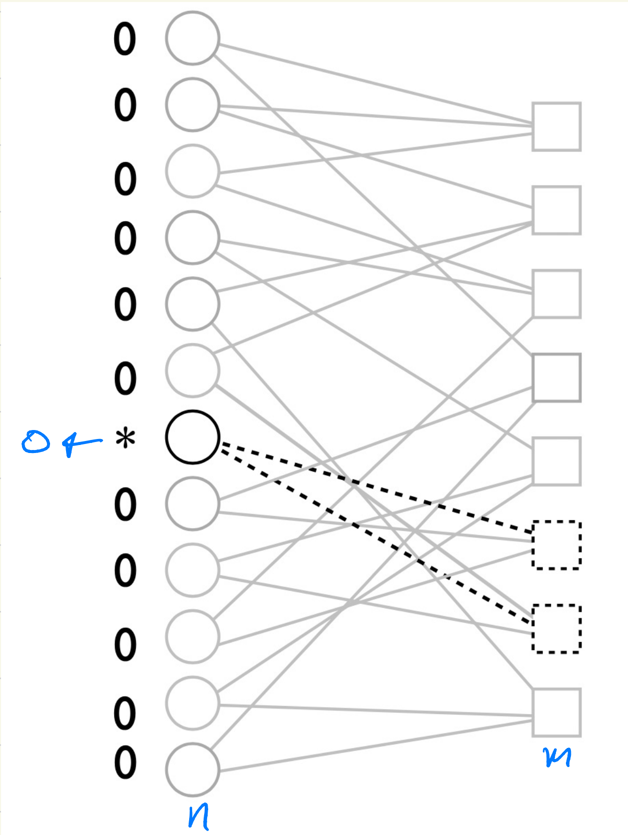
* factor nodes with 1 remaining edge can be decoded.


0
0
0
0
0
0
*
0
*
*
0
0


0
0
0
0
0
0
*
0
*
*
0
0

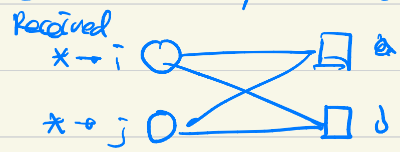

0
0
0
0
0
0
*
0
*
*
0
0


0
0
0
0
0
0
*
0
0
0
0
0



* If all nodes are peeled, then Decoding success.

* If a subset of nodes are left whose uncertainty is not resolves ef



then those bits cannot be decoded.

* Design Goal: for a given n & m, find a graph that minimizes the error probability.

↑ Density Evolution is critical in achieving this breakthrough

---

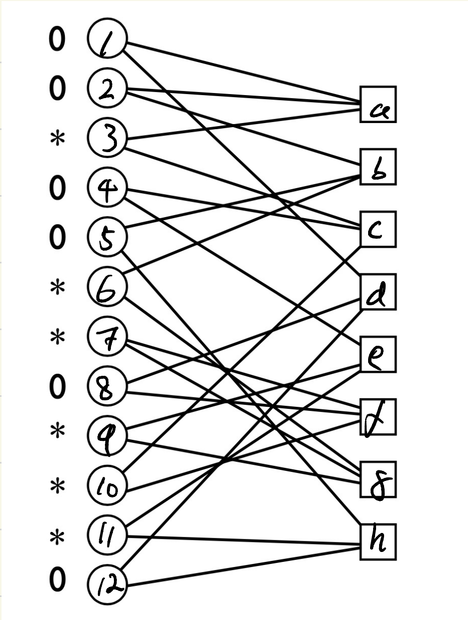
Strategy under Random Factor Graph with $m, n \to \infty$.

① suppose we update BP as follows.

repeat
- Update $\{m^{(t)}_{i\to a}\}$
- Draw fresh random graph
- Update $\{\tilde{m}^{(t)}_{a\to i}\}$
- Draw fresh random graph
- $t = t+1$

and analyze this process

② Justify it is accurate if FG is random & $m, n \uparrow \infty$. for fixed $t \le \log n$



* Computation tree for $m^{(t)}_{i\to a}(x_i)$

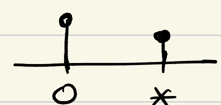
message from $i \xrightarrow{\text{to}} a$, after t-iterations of BP.

for $t=1$

$$\left(\begin{array}{c} m^{(1)}_{1\to a}(x_1) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ \vdots \\ m^{(1)}_{3\to c}(x_3) = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \\ \vdots \end{array} \right) \Rightarrow \left(\begin{array}{c} X^{(1)}_{1\to a} = 0 \\ \vdots \\ X^{(1)}_{3\to c} = * \\ \vdots \end{array} \right)$$

Set. Set $\{X^{(1)}_{i\to a}\}$

Histogram of this $\{X^{(1)}_{i\to a}\}$



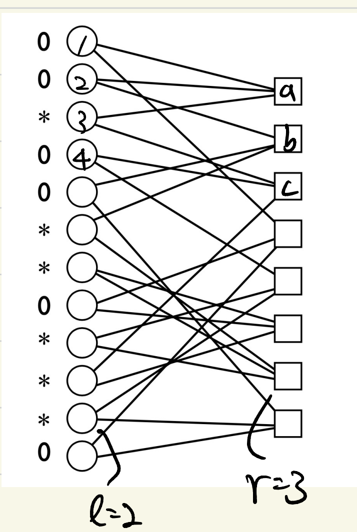
Density : $\hat{\mathbb{P}}^{(1)}_{G_{m,n}\,BEC(\varepsilon)}(X=*) \begin{array}{c} 0 \\ * \end{array}\begin{bmatrix} \\ \bigcirc \end{bmatrix} = \xi^{(1)}$ keep

for a randomly chosen edge

Q. $\mathbb{E}\left[\hat{\mathbb{P}}^{(1)}_{G_{m,n},BEC(\varepsilon)} \right] =$

Q. $\lim_{m,n\to\infty} \hat{\mathbb{P}}^{(1)}_{G_{m,n},BEC(\varepsilon)} =$

for $t=1$, $\tilde{m}^{(1)}_{a\to i}(x_i)$



$\tilde{X}_{a\to 1} \in \{0,1,*\}$

local computation tree. at depth $t=1$

$X^{(1)}_{2\to a}=0$, $X^{(1)}_{3\to a}=*$, $r-1$

$$\tilde{X}^{(1)}_{a\to 1} = \begin{cases} * & \text{if at least 1 star} \\ \bigoplus X_{\partial a \setminus i\to a} \end{cases} = *$$

Histogram $\{\tilde{X}^{(1)}_{a\to i}\} = \tilde{\mathbb{P}}^{(1)}_{G_{m,n},BEC(\varepsilon),\,l,r}(X=*) = \tilde{\xi}^{(1)}$

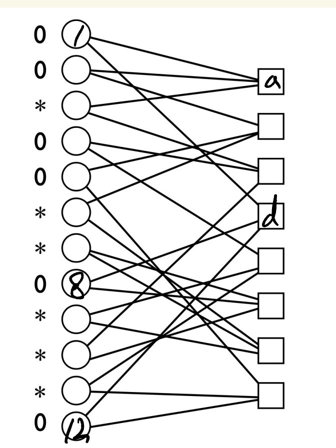
$$\mathbb{P}(\tilde{X}^{(1)}_{a\to 1}=*) = \mathbb{P}(\underbrace{X^{(1)}_{2\to a}=*}_{\tilde{q}^{(1)}} \text{ or } X^{(1)}_{3\to a}=*)$$

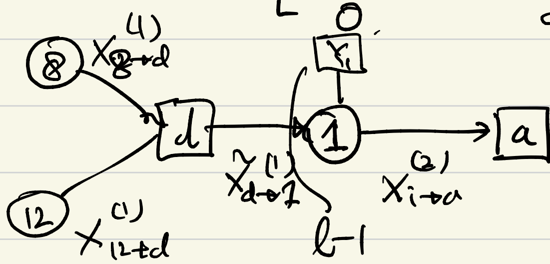
$$\tilde{q}^{(1)} = 1-(1-q^{(1)})^{r-1=2}$$

Completely justified if
$\{m^{(1)}_{i\to a}\}$ was computed
and then we draw random edges again,
and compute $\{\tilde{m}^{(1)}_{a\to i}\}$ and repeat.

$$q^{(2)} = \mathbb{P}(X^{(2)}_{i\to a}=*) = (\tilde{q}^{(1)})^{\ell-1}\cdot\varepsilon$$

$$\begin{cases} * & \text{if incoming is } * \ \& \ Y_i=* \\ 0 & \text{otherwise} \end{cases}$$





Evolution of messages		Evolution of density (Histogram)
$\{m^{(1)}_{i\to a}\}$	$\longleftrightarrow$	$q^{(1)}\in[0,1]$
$\{\tilde{m}^{(1)}_{a\to i}\}$		$\tilde{q}^{(1)}$
$\{m^{(2)}_{i\to a}\}$		$q^{(2)}$
$\{\tilde{m}^{(2)}_{a\to i}\}$		$\tilde{q}^{(2)}$
$\vdots$		

$q^{(1)}=\varepsilon$

$\tilde{q}^{(t)}=1-(1-q^{(t)})^{r-1}$

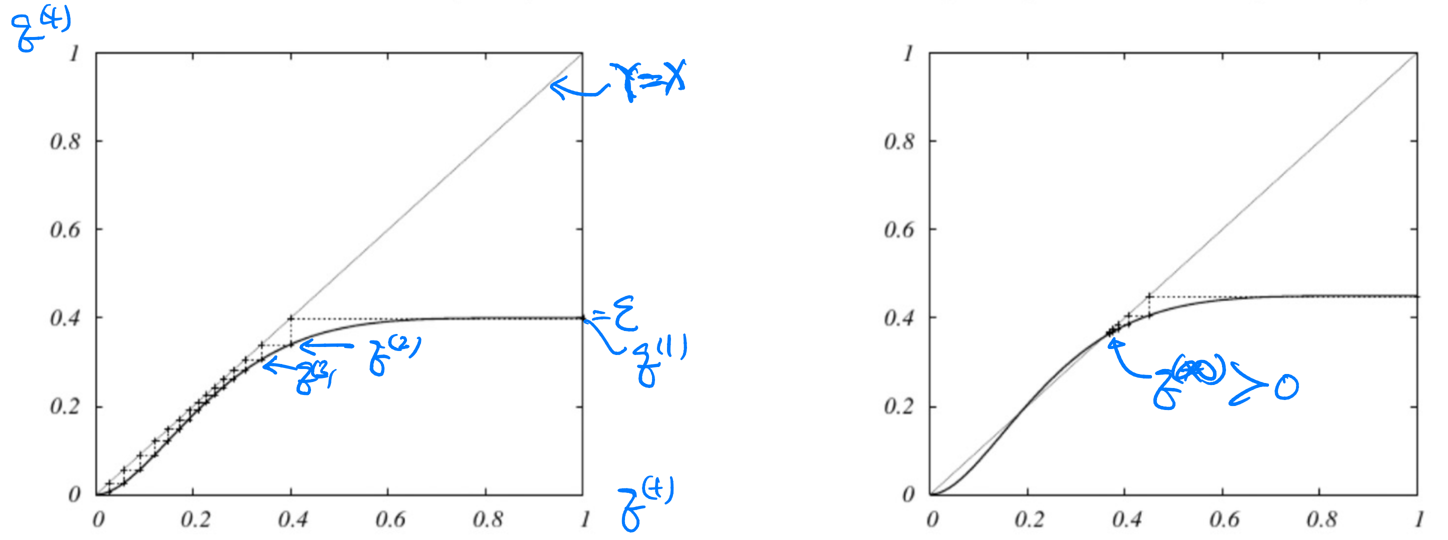
$q^{(t)}=\varepsilon\cdot(\tilde{q}^{(t-1)})^{\ell-1}$

Density Evolution Update

$$q^{(1)}=\varepsilon,\quad q^{(t)}=\varepsilon\cdot\left(1-(1-q^{(t-1)})^{r-1}\right)^{\ell-1}$$

$$z^{(t+1)} = \varepsilon \cdot \left(1-\left(1-z^{(t)}\right)^{r-1}\right)^{\ell-1}$$

density evolution for $(\underset{\ell}{3},\underset{r}{6})$ code with $\epsilon = 0.4$(left) and 0.45(right)



rate of this code = 0.5, threshold $\epsilon^* \simeq 0.4$xxx,

eventually $z^{(t)} \to 0$.

$P(X_{i\to a} = *) \to 0$

Perfect decoding.

if $\varepsilon = 0.4$

$z^{(\infty)} > 0$, 0.38xxx

$\lim_{t\to\infty} \lim_{n\to\infty} P(X_{i\to a} = *) = 0.38$

if $\varepsilon = 0.45$

bit-error-rate.

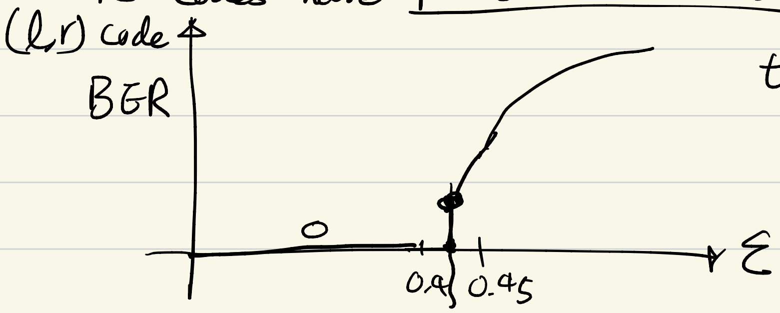
$$P(X_i = *) = \varepsilon \cdot \left(1-(1-z)^{r-1}\right)^{\ell}$$

$$= \prod_{a\in\partial i} \hat{X}_{a\to i} =$$

---

* Important Consequence ①.

LDPC Codes have phase transition



the asymptotic BER curve has a discontinuity

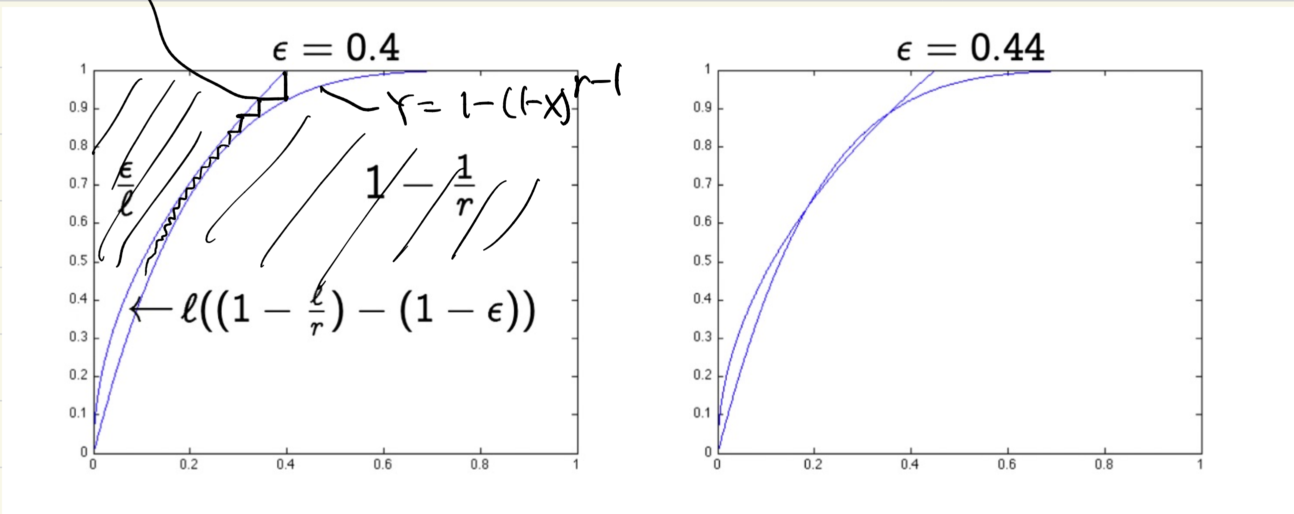
ε^* ← can compute with density evolution

* Important Consequence (2)

we had $\quad z = \varepsilon(1-(1-z)^{r-1})^{\ell-1}$ ← the solution is the final error rate.

let's change it to $\left(\frac{z}{\varepsilon}\right)^{\frac{1}{\ell-1}} = 1-(1-z)^{r-1}$

$Y = \left(\frac{x}{\varepsilon}\right)^{\frac{1}{\ell-1}}$ plot $\quad = Y \qquad = Y$



If we can design degree distribution $\ell = [\ell_1 \ \ell_2 \ \ell_3 \ \dots \ \ell_{max}]$ with $\ell_1 = 0$, $\ell_2 = 0.3$, $\ell_3 = 0.4$

$r = [\qquad]$

such that the Gap closes ⟶ then we have a capacity achieving code.

Claim: for fixed t, the ℓ, r random graph is locally-treelike with probability one, as $m, n \to \infty$.

$\lim_{n\to\infty} \mathbb{P}(\text{local tree of depth } t \text{ is a tree}) \to 0$

⇒ t-iterations of BP is performed on a random tree, so the 2 processes are statistically equivalent.