Chapter 7 PageRank

Angsheng Li

Institute of Software
Chinese Academy of Sciences

Advanced Algorithms
U CAS
1st, April, 2017

Outline

1. Backgrounds
2. Web graph
3. Google's matrix
4. Teleportation
5. Personalised vector
6. Sensitivity
7. Proofs
8. Local algorithms
9. Exrcises

The new phenomena

Brin and Page, 1995 - 1998

1. The current-generation search engine
2. Billions of queries everyday
3. What is the principle behind?
4. How good is the current-generation search engine?

The graph

- Massive directed graph
- Nodes: webpages
- Directed edges, hyperlines, including inlinks and outlinks
- The question: Rank the web pages by importance.

The PageRank thesis

A page is important, if it is pointed to by many important pages.

Brin and Page, 1998

Established the equation of the PageRank thesis.
The PageRank of a page P_i, written $r(P_i)$, is the sum of the PageRanks of all the pages pointing to P_i, that is,

$$r(P_i) = \sum_{P_j \in B_i} \frac{r(P_j)}{|P_j|}, \qquad (1)$$

- B_i: the set of pages pointing to P_i,
- $|P_j|$: the number of outlinks from page P_j.

Recurrence of the PageRank

$$\begin{cases} r_{k+1}(P_i) = \sum_{P_j \in B_i} \frac{r_k(P_j)}{|P_j|} \\ r_0(P_i) = \frac{1}{n} \end{cases} \tag{2}$$

The stationary solution of the recursive equation in Equation (2) gives rise to the PageRank of a graph G.

Matrix representation

$$H_{ij} = \begin{cases} \frac{1}{|P_i|} & \text{if there is an edge from node } i \text{ to node } j, \\ 0 & \text{o.w.} \end{cases} \tag{3}$$

$|P_i|$: The number of outlinks from node i.
$H = (H_{ij})$ is the PageRank matrix of G.

PageRank solution

Let π^{T} be a $1 \times n$ vector.
Set

$$\begin{cases} \pi^{(k+1)\mathrm{T}} = \pi^{(k)\mathrm{T}} H, \\ \pi^{(0)\mathrm{T}} = \frac{1}{n} e^{\mathrm{T}}, \end{cases} \tag{4}$$

where $e^{\mathrm{T}} = (1, 1, \cdots, 1)$.
For the equation (4), we require:

- convergence and the interpretation of the solution
- Uniqueness of the solution
- Invariance of $\pi^{(0)}$
- The number of iterations of the convergent solution

Rank sinks

$$\begin{tikzcd} 1 \arrow[rr, shift left] \arrow[dr] & & 2 \arrow[ll, shift left] \arrow[dl] \\ & 3 & \end{tikzcd}$$

All the PageRanks go to node 3.

Matrix S

To solve the sink problem, define a vector **a**,

$$a_i = \begin{cases} 1 & \text{if node } i \text{ has no outgoing links,} \\ 0 & \text{o.w.} \end{cases} \tag{5}$$

Definition

Define

$$S = H + \frac{1}{n}\mathbf{a}\mathbf{e}^{\mathrm{T}},$$

where $\mathbf{e}^{\mathrm{T}} = (1, 1, \cdots, 1)$.

Intuition: If node i has no outgoing link, then from node i, the randomly walks to any other nodes uniformly.

S is the transition probability matrix of a Markov chain.

Google's matrix G

Definition

Define the Google's matrix by

$$G = \alpha S + (1-\alpha)J,$$

where $J_{ij} = \frac{1}{n}$.

- J is called *teleportation matrix*
- $1-\alpha$ is called the *teleportation parameter*.

Expander

Recall: If G is a graph with $\lambda = \lambda(G) < 1$, then for $A = A_G$,

$$A = (1-\lambda)J + \lambda C,$$

for some C with $\|C\| \leq 1$.
We thus know that Google's matrix is an expander. However, the parameter α is chosen arbitrarily. Of course, α determines the spectral gap of the graph.

Properties of G - I

(1) G is stochastic 随机游走
It is a convex combination of two stochastic matrices S and J.

(2) G is irreducible. 因为有J，可以直接连其他点
Every page is directly connected to every other page.

(3) G is aperiodic.
$G_{ii} > 0$. Every node has a self-loop.

(4) G is primitive.
There exists a k such that $G^k > 0$
Because: G is an expander. There is a unique π^{T} such that

$$\|pG^l - \pi^{\mathrm{T}}\| \approx 0$$

for a small l. – Power method works

Properties of G - II

(5) G is rank-one updated

$$\begin{aligned} G &= \alpha S + (1-\alpha)\frac{1}{n}ee^{\mathrm{T}} \\ &= \alpha(H + \frac{1}{n}ae^{\mathrm{T}}) + (1-\alpha)\frac{1}{n}ee^{\mathrm{T}} \\ &= \alpha H + (\alpha\frac{1}{n}a + (1-\alpha)\frac{1}{n}e)e^{\mathrm{T}}. \end{aligned} \tag{6}$$

- H is sparse
- $\alpha\frac{1}{n}a + (1-\alpha)\frac{1}{n}e$ is dense, but only one-dimensional vector.

(6) G is artificial due to the choice of α.
G may not well reflect the real world H.

Computation of π^{T}

Power method

$$\begin{aligned} \pi^{(k+1)\mathrm{T}} &= \pi^{(k)\mathrm{T}} G \\ &= \alpha \pi^{(k)\mathrm{T}} S + \frac{1-\alpha}{n} \pi^{(k)\mathrm{T}} e e^{\mathrm{T}} \\ &= \alpha \pi^{(k)\mathrm{T}} H + (\alpha \pi^{(k)\mathrm{T}} a + (1-\alpha) e) e^{\mathrm{T}} / n. \end{aligned} \tag{7}$$

Suppose that $1, \lambda_2, \cdots, \lambda_n$ are the eigenvalues of G with $1 > |\lambda_2| \geq \cdots \geq |\lambda_n|$.
Then:

$$G = G_1 + \lambda_2 G_2 + \cdots + \lambda_n G_n,$$

– $G_i^2 = G_i$,
– For $i \neq j$, $G_i G_j = 0$.
Then

$$G^l = G_1 + \lambda_2^l G_2 + \cdots \lambda_n^l G_n$$

Since $\lambda_2 < 1$, G^l quickly converges to G_1.

$\lambda(G)$

Lemma

For the Google matrix $G = \alpha S + (1-\alpha)J$*,*

$$|\lambda_2(G)| \leq \alpha.$$

$\lambda(G)$ again

Lemma

If the spectrum of the stochastic matrix S is $\{1, \lambda_2, \cdots, \lambda_n\}$, then the spectrum of the Google matrix $G = \alpha S + (1-\alpha) e v^{\mathrm{T}}$ is

$$\{1, \alpha\lambda_2, \cdots, \alpha\lambda_n\},$$

where v^{T} is the personalised vector.

Proofs - I

Since S is stochastic, $(1, \mathbf{e})$ is an eigenpair of S. Let $Q = (eX)$ be a nonsingular matrix that has the eigenvector $\mathbf{e}$ as its first column.

Set

$$Q^{-1} = \begin{pmatrix} y^{\mathrm{T}} \\ Y^{\mathrm{T}} \end{pmatrix} \tag{8}$$

Then:

$$Q^{-1}Q = \begin{pmatrix} y^{\mathrm{T}}\mathbf{e} & y^{\mathrm{T}}X \\ Y^{\mathrm{T}}\mathbf{e} & Y^{\mathrm{T}}X \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & I \end{pmatrix} \tag{9}$$

Proofs - II

Similarly,

$$Q^{-1}SQ = \begin{pmatrix} y^{\mathrm{T}}e & Y^{\mathrm{T}}SX \\ Y^{\mathrm{T}}\mathbf{e} & Y^{\mathrm{T}}SX \end{pmatrix} = \begin{pmatrix} 1 & y^{\mathrm{T}}SX \\ 0 & Y^{\mathrm{T}}SX \end{pmatrix} \tag{10}$$

This implies that $Y^{\mathrm{T}}SX$ contains the remaining eigenvalues of S, i.e., $\lambda_2, \cdots, \lambda_n$.
In addition,

$$Q^{-1}GQ = \begin{pmatrix} 1 & \alpha y^{\mathrm{T}}SX + (1-\alpha)v^{\mathrm{T}}X \\ 0 & \alpha Y^{\mathrm{T}}SX \end{pmatrix} \tag{11}$$

The eigenvalues of G are

$$\{1, \alpha\lambda_2, \cdots, \alpha\lambda_n\}.$$

Since $\lambda_2 \leq 1$, $\alpha\lambda_2 \leq \alpha$.

The role α

$$G = (1-\alpha)J + \alpha S.$$

If α is small, then $1-\alpha$ is large, G is basically an artificial random graph, failing to reflect the real world matrix S.
If α is large, then

- there is no unique stationary distribution
- even if there is a stationary distribution, it is hard to compute
- the power method fails

Google's choice: $\alpha = 0.85$.

Personalised PageRank

For a personalised probability vector v^{T},

$$G = \alpha S + (1-\alpha) e v^{\mathrm{T}}.$$

The power method works as before.
The stationary distribution is a personalised PageRank.
Significance: Real applications.

The stationary distribution

Theorem

The Pagerank $\pi^{\mathrm{T}}(\alpha)$ of G_α is

$$\pi^{\mathrm{T}}(\alpha) = \frac{1}{\sum_{i=1}^{n} D_i(\alpha)} (D_1(\alpha), D_2(\alpha), \cdots, D_n(\alpha))$$

where $D_i(\alpha)$ is the i-th principal minor determinant of order $n-1$ in $I - G_\alpha$.
Furthermore, every $D_i(\alpha)$ is differentiable for α.

Proof.

By definition. $\square$

Differential

Theorem

If $\pi^{\mathrm{T}}(\alpha) = (\pi_1(\alpha), \pi_2(\alpha), \cdots, \pi_n(\alpha))$, *then*

1. *For each* j,
$$|\frac{d\pi_j(\alpha)}{d\alpha}| \leq \frac{1}{1-\alpha}.$$
2.
$$\|\frac{d\pi^{\mathrm{T}}(\alpha)}{d\alpha}\|_1 \leq \frac{2}{1-\alpha}.$$

- If α is small, then the PageRank $\pi^{\mathrm{T}}(\alpha)$ is not sensitive.
- If α is large, then the upper bounds $\frac{1}{1-\alpha}$ and $\frac{2}{1-\alpha}$ are both approaching to infinity.

Representation

Theorem

$$\frac{d\pi^{\mathrm{T}}(\alpha)}{d\alpha} = -v^{\mathrm{T}}(I-S)(I-\alpha S)^{-2}.$$

Sensitive to H

1. $$\frac{d\pi^{\mathrm{T}}(h_{ij})}{dh_{ij}} = \alpha\pi_i(e_j^T - v^T)(I-\alpha S)^{-1}$$
2. $$(I-\alpha S)^{-1} \to \infty,$$

   as α goes to 1.

π^{T} is sensitive to perturbations in H is $\alpha \approx 1$.
Therefore, if $\alpha \approx 1$, then π^{T} is sensitive to small changes of the matrix H.

Sensitive to v^T

$$\frac{d\pi^{\mathrm{T}}(v^T)}{dv^T} = (1-\alpha+\alpha\sum_{i\in D}\pi_i)(I-\alpha S)^{-1},$$

D is the set of nodes that have no outgoing links.
The same as before, as α goes to 1, $(I-\alpha S)^{-1}$ goes to ∞.

Summary of sensitivity

If $\alpha \approx 1$, then

1. Computing $\pi^{\mathrm{T}}(\alpha)$ is hard, since the power method fails
2. $\pi^{\mathrm{T}}(\alpha)$ is sensitive to the perturbation of H
3. $\pi^{\mathrm{T}}(\alpha)$ is sensitive to the personalised vector v^{T}

Google's tradeoff:

$$\alpha = 0.85$$

Proof of upper bounds - I

Theorem

If $\pi^{\mathrm{T}}(\alpha) = (\pi_1(\alpha), \pi_2(\alpha), \cdots, \pi_n(\alpha))$, *then*

1. *For each j,*
$$\left|\frac{d\pi_j(\alpha)}{d\alpha}\right| \leq \frac{1}{1-\alpha}.$$
2. $$\left\|\frac{d\pi^{\mathrm{T}}(\alpha)}{d\alpha}\right\|_1 \leq \frac{2}{1-\alpha}.$$

$\pi^T(\alpha)$ is a probability vector, so

$$\sum_{i=1}^{n} \pi_i(\alpha) = 1$$

giving
$\pi^{\mathrm{T}}(\alpha)e = 1$, $e^T = (1, 1, \cdots, 1)$.

Proof of upper bounds- II

By definition,
$\pi^T(\alpha) = \pi^T(\alpha)G(\alpha) = \pi^T(\alpha)(\alpha S + (1-\alpha)ev^T)$.
By differential,

$$\frac{d\pi^T(\alpha)}{d\alpha} = \pi^T(\alpha)(S - ev^T)(I - \alpha S)^{-1}. \tag{12}$$

For (1). For every real x, $xT \perp e$, i.e., $\sum x_i = 0$, and for all real vector y, column vector,

$$\begin{aligned} |x^{\mathrm{T}}y| &= |\sum_{i=1}^{n} x_i y_i| \\ &\leq \|x^{\mathrm{T}}\|_1 \cdot \frac{y_{\max} - y_{\min}}{2}. \end{aligned} \tag{13}$$

By Equation (12),

$$\frac{d\pi_j(\alpha)}{d\alpha} = \pi^{\mathrm{T}}(\alpha)(S - ev^{\mathrm{T}})(I - \alpha S)^{-1}e_j.$$

Prof of upper bounds - III

Since $\pi^{\mathrm{T}}(\alpha)(S - ev^{\mathrm{T}})e = 0$, set $x^{\mathrm{T}} = \pi^{\mathrm{T}}(\alpha)(S - ev^{\mathrm{T}})$ and $y = (I - \alpha S)^{-1} e_j$.
By Inequality (13),

$$\left|\frac{d\pi_j(\alpha)}{d\alpha}\right| \leq \|\pi^{\mathrm{T}}(\alpha)(S - ev^{\mathrm{T}})\|_1 \cdot \frac{y_{\max} - y_{\min}}{2}.$$

Since $\|\pi^{\mathrm{T}}(\alpha)(S - ev^{\mathrm{T}})\|_1 \leq 2$,
$\left|\frac{d\pi_j(\alpha)}{d\alpha}\right| \leq y_{\max} - y_{\min}$.
Since $(I - \alpha S)^{-1} \geq 0$ and $(I - \alpha S)e = (1 - \alpha)e$, and hence $(I - \alpha S)^{-1} = (1 - \alpha)^{-1} e$.
This shows that $y_{\min} \geq 0$.
For $y_{\max}$, we have
$y_{\max} \leq \max_{i,j}[(I - \alpha S)^{-1}]_{ij} \leq \frac{1}{1-\alpha}$.
(1) follows.

Proof of upper bounds - IV

For (2).

$$\begin{aligned}\|\frac{d\pi^{\mathrm{T}}(\alpha)}{d\alpha}\|_1 &= \|\pi^{\mathrm{T}}(\alpha)(S-ev^{\mathrm{T}})(I-\alpha S)^{-1}\|_1 \\ &\le \|\pi^{\mathrm{T}}(\alpha)(S-ev^{\mathrm{T}})\|_1 \cdot \|(I-\alpha S)^{-1}\|_\infty \\ &\le 2\frac{1}{1-\alpha} = \frac{2}{1-\alpha}. \end{aligned} \tag{14}$$

Conductance

Given a graph $G = (V, E)$ and $S \subset V$, the conductance of S in G is:

$$\Phi(S) = \frac{|E(S, \bar{S})|}{\min\{\mathrm{vol}(S), \mathrm{vol}(\bar{S})\}}.$$

The conductance of G is
$\Phi = \min\{\Phi(S) \mid |S| \leq \frac{n}{2}\}$.

$\mathrm{Push}(u)$

Andersen, Chung and Lang, FOCS, 2006.
Define an operator
$\mathrm{Push}(u)$:

1. $p(u) \leftarrow p(u) + \alpha r(u)$
2. $r(u) \leftarrow (1-\alpha) r(u)/2$
3. For each v with $v \sim u$, set

$$r(v) \leftarrow r(v) + (1-\alpha) r(u)/(2d(u)).$$

Approximate PageRank

Given a node v,

1. set $p = 0$, $r(v) = 1$, and $r(u) = 0$ for all $u \neq v$.
2. For every u, if $r(u) \geq \epsilon d(u)$, then:
   – Apply $\mathrm{push}(u)$.
3. Otherwise, Then output p and r.

ACL local algorithm

1. To find the RageRank from a given input vertex v,
2. To rank the pages by decreasing of the normalised PageRank, i.e., $\frac{p_v}{d(v)}$. Suppose that $v_1, v_2, \cdots, v_l$ is listed such that

$$\frac{p_{v_1}}{d(v_1)} \geq \frac{p_{v_2}}{d(v_2)} \geq \cdots \geq \frac{p_{v_l}}{d(v_l)}.$$

3. (Pruning) To take an initial segment of the list as a community associated with the given input v.
   Let j be such that

$$\Phi(X_j) = \min\{\Phi(X_i) \mid 1 \leq i \leq l\},$$

   where $\phi(X)$ is the conductance of X in G, and
   $X_k = \{v_1, \cdots, v_k\}$. conductance小，图出去的少
   Output X_j.

Question for local algorithm

For every query Q, we rank the set of answers for the query by PageRank, however, the list is a too long list.
The question is to determine a short list of ranks as the output of the query.
Still open.

The great idea

- The PageRank thesis
- The teleportation parameter $1-\alpha$.
  This is a great idea, which may be used in many other areas, such as learning, data processing.
  The essence of the idea here is to make sure that the Ranking matrix is a well-defined stochastic procedure so that PageRank exists and can be computed.
  We may also regard the introduction of $1-\alpha$ as amplifying noises, playing a role similar to that in the error correcting codes.
- Google's success: Making big money by randomness

A grand challenge

- What is the principle for determining α? Is there a metric of networks which determines the optimum α?
- What are principles for structuring the unstructured and noisy data?
- Making money by connection and interaction???

Reference

1. Amy N. Langville and Carl D. Meyer, Google's PageRank and Beyond: The Science of Search Engine Ranking, Princeton University Press, 2006.
2. Andersen, Chung and Lang, Local graph partitioning using PageRank vectors, FOCS, 2006.

Natural rank

The natural rank based on the structural information theory is the answer.

Exercise 1

Let $X_1, \cdots, X_n$ be independent random variables such that X_i is equal to 1 with probability $1-\delta$ and equal to 0 with probability δ. Let $X = \sum_{i=1}^{n} X_i (\mod 2)$. Prove that

$$\Pr[X = 1] = \begin{cases} \frac{1}{2} + (1-2\delta)^n/2, & n \text{ is odd,} \\ \frac{1}{2} - (1-2\delta)^n/2, & n \text{ is even.} \end{cases}$$

Significance?

Exercise 1 - proof 1

Let $Y_i = (-1)^{X_i}$, and $Y = \prod_{i=1}^{n} Y_i$.
Assume n odd.
Let $\Pr[X = 1] = \alpha$.
Then

$$E[Y] = 1 - 2\alpha.$$

Since X_i and then Y_i are independent,

$$E[Y] = (-1 + 2\delta)^n$$

Therefore

$$(-1 + 2\delta)^n = 1 - 2\alpha$$

$$\alpha = \frac{1}{2} + \frac{(1 - 2\delta)^n}{2}.$$

Exercise 1 - proof 2

Proof.

$$\begin{aligned}(1-2\delta)^n &= ((1-\delta)-\delta)^n \\ &= \sum_{i=0}^{n}\binom{n}{i}(1-\delta)^i(-\delta)^{n-i}.\end{aligned}$$

Case 1. n odd
$(1-2\delta)^n = \Pr[X=1]-\Pr[X=0]$,

$$\Pr[X=1] = \frac{1}{2}+\frac{1}{2}(1-2\delta)^n$$

Case 2. n even $(1-2\delta)^n = \Pr[X=0]-\Pr[X=1]$,

$$\Pr[X=1] = \frac{1}{2}-\frac{1}{2}(1-2\delta)^n$$

Exercise 2

Prove that if there exists a δ-density distribution H such that $\Pr_{x \in_R H}[C(x) = f(x)] \leq \frac{1}{2} + \epsilon$ for every circuit C of size at most s with $s \leq \sqrt{\epsilon^2 \delta 2^n / 100}$, then there exists a subset $I \subseteq \{0,1\}^n$ of size at least $\frac{\delta}{2} 2^n$ such that

$$\Pr_{x \in_R I}[C(x) = f(x)] \leq \frac{1}{2} + 2\epsilon$$

for every circuit C of size at most s.

Exercise 2 - Proof

Some problems?
Leave this to Mingji

Exercise 3

1. Let $f : \mathbb{F} \to \mathbb{F}$ be any function. Suppose integer $d \geq 0$ and number $\epsilon > 2\sqrt{\frac{d}{|\mathbb{F}|}}$. Prove that there are at most $2/\epsilon$ degree d polynomials that agree with f on at least an ϵ fraction of its coordinates.
   Significance?
2. Prove that if $Q(x, y)$ is a bivariate polynomial over some field $\mathbb{F}$ and $P(x)$ is a univariate polynomial over $\mathbb{F}$ such that $Q(x, P(x))$ is the zero polynomial, then $Q(x, y) = (y - P(x))A(x, y)$ for some polynomial $A(x, y)$.

Exercise 3 - proof 1

Suppose that

$$P_1, P_2, \cdots, P_l$$

are the all degree d polynomials that agree with f in at least ϵ fraction of coordinates.
For each i, define a vector v_i by

$$v_i(j) = \begin{cases} 1, & \text{if } P_i(j) = f(j), \\ 0, & \text{otherwise} \end{cases}$$

for every $j \in \mathbb{F}$.
Then for every i,

$$\|v_i\|_1 \geq \epsilon \cdot m,$$

where $m = |\mathbb{F}|$.

$$\epsilon m \leq \langle v_i, v_i \rangle \leq m$$

Exercise 3 - proof 2

Set

$$v = \sum_{i=1}^{l} v_i$$

Then

$$\begin{aligned} \langle v, v \rangle &= \sum_{i=1}^{l} \langle v_i \rangle + \sum_{i \neq j} \langle v_i v_j \rangle \\ &\leq l \cdot m + (l^2 - l)d. \end{aligned}$$

Exercise 3 - proof 3

And

$$
\begin{aligned}
\langle v, v\rangle &= \sum_{k\in\mathbb{F}} (v(k))^2 \\
&= \sum_k \Big(\sum_{i=1}^{l} v_i(k)\Big)^2 \\
&\geq \frac{(\sum_k \sum_i v_i(k))^2}{m} \\
&= \frac{(\sum_i \sum_k v_i(k))^2}{m} \\
&\geq \frac{(l\epsilon m)^2}{m}.
\end{aligned}
$$

Exercise 3 - proof 4

This gives

$$I \leq \frac{1-\frac{d}{m}}{\epsilon^2-\frac{d}{m}}$$

for $\epsilon > \sqrt{\frac{d}{m}}$.

Exercise 3 - proof 5

For

$$\epsilon > 2\sqrt{\frac{d}{m}}$$

and

$$l \leq \frac{2}{\epsilon}.$$

$$\epsilon + (\epsilon - \frac{d}{m}) + \cdots (\epsilon - \frac{(l-1)d}{m}) \geq 1$$

with

$$\epsilon - \frac{(l-1)d}{m} \geq 0$$

Solving this, we have

$$l \leq \frac{2}{\epsilon}.$$

Exercise 3 - proof 4

Take $Q(x,y)$ as a polynomial of y with coefficients being polynomials of x.
Divide $Q(x,y)$ by the linear function $y-P(x)$, linear in variable y, giving

$$Q(x,y)=(y-P(x))A(x,y)+R(x)$$

By the assumption,

$$Q(x,P(x))=R(x)\equiv 0.$$

Exercise 4

Linear codes We say that an ECC $E : \{0,1\}^n \to \{0,1\}^m$ is *linear*, if for every $x, x' \in \{0,1\}^n$, $E(x + x') = E(x) + E(x')$ (componentwise addition modulo 2). A linear ECC can be seen an $m \times n$ matrix A such that $E(x) = Ax$, thinking of x as a column vector.

1. Prove that the distance of a linear ECC is equal to the minimum over all nonzero $x \in \{0,1\}^n$ of the fraction of 1's in $E(x)$.
2. Prove that for every $\delta > 0$, there exists a linear ECC $E : \{0,1\}^n \to \{0,1\}^m$ for $m = \Omega(n)/(1 - H(\delta))$ with distance δ.
3. Prove that for some $\delta > 0$, there is an ECC $E : \{0,1\}^n \to \{0,1\}^{\mathrm{poly}(n)}$ of distance δ with poly time encoding, and decoding algorithms.

Exercise 4 - proof 1

Let A be an $m \times n$ $0,1$ matrix which defines a linear ECC.
The distance of A is:

$$\delta = \min_{x \neq x'} \frac{1}{m} \cdot |\{i \mid y_i \neq y_i'\}|$$

where

$$y_i = a_{i,1}x_1 + a_{i,2}x_2 + \cdots a_{i,n}x_n$$

and

$$y_i' = a_{i,1}x_1' + a_{i,2}x_2' + \cdots a_{i,n}x_n'$$

This is

$$\delta = \min_{x \neq 0} \frac{1}{m} \cdot |\{i \ |y_i = 1\}|.$$

Exercise 4 - proof 2

Remove the condition that E is linear.
Given a vector $y \in \{0,1\}^m$, define the δ-ball of y to be the set of the vectors $z \in \{0,1\}^m$ such that the distance between y and z is less than δ.
Denoted by B_y^δ. Then

$$|B_y^\delta| \leq \binom{m}{\delta \cdot m} = o(1) \cdot 2^{H(\delta)\cdot m}$$

In increasing order, for each $x \in \{0,1\}^n$, we define $E(x)$ to be a $y \in \{0,1\}^m$ such that B_y^δ disjoins all the δ-balls associated with the codewords of $x' < x$.
Suppose that $m \geq \frac{n}{1-H(\delta)}$. Then the definition above never stops, since there are at least 2^n many disjoint δ-balls in $\{0,1\}^m$.

Exercise 4 - proof 3

Consider now the linear ECC.
Each linear ECC is given by an $m \times n$ matrix A.
Two approaches:
Case 1. Consider the random matrix A.
With nonzero probability that A is such an ECC.
Case 2. Counting the number of linear ECC that have distance $< \delta$.

Exercise 4 - proof 4

Consider the first approach.
Let A be a random $m \times n$ matrix.
We say that $x = (x_1, x_2, \cdots, x_n)$ is a witness showing that A has distance $< \delta$, if $x \neq 0$ and there are $< \delta m$ many j such that $y_j = 1$, where

$$y_j = a_{j1}x_1 + \cdots + a_{jn}x_n.$$

For each j, define

$$Y_j = \begin{cases} 1, & \text{if } y_j = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$Y = \sum_{j=1}^{m} Y_j.$$

Exercise 4 - proof 5

Then for each j,

$$E[Y_j] = \frac{1}{2}$$

$$E[Y] = \mu = \frac{m}{2}$$

Clearly, all Y_j's are independent.
By the Chernoff bound, for $\epsilon = 1 - 2\delta$,

$$\begin{aligned}
\Pr[Y < \delta m] &= \Pr[Y < (1-\epsilon)\mu] \\
&\leq \left[\frac{e^{-\epsilon}}{(1-\epsilon)^{(1-\epsilon)}}\right]^{\frac{m}{2}} \\
&\leq \frac{1}{2^{c \cdot m}},
\end{aligned}$$

for some constant c.

Exercise 4 - proof 6

By the union bound,
the probability that A has a witness for distance $< \delta$ is

$$\frac{1}{2^{c \cdot m - n}}$$

which is ≈ 0 if

$$m = \Omega(n).$$

Exercise 4 - proof 7

Consider the Reed-Solomon code

$$RS : \mathbb{F}^n \to \mathbb{F}^m$$

It is an ECC with distance $\delta_1 = 1 - \frac{n}{m}$.
For every $x = (a_0, a_1, \cdots, a_{n-1}) \in \mathbb{F}^n$,

$$RS(x) = (z_0, z_1, \cdots, z_{m-1})$$

where

$$z_j = \sum_{i=0}^{n-1} a_i j^i$$

$j \in \mathbb{F}$.

Exercise 4 - proof 8

Let $|\mathbb{F}| = 2^k$.
Then each element $f \in \mathcal{F}$ is interpreted as an element in $GF(2^k)$.
For $x \in \mathbb{F}^n$, we interpret it as an element in $\{0,1\}^{k \cdot n}$. We encode $RS(x)$ by

$$WH(z_0), WH(z_1), \cdots, WH(z_{m-1})$$

This is an ECC from $\{0,1\}^{k \cdot n}$ to $\{0,1\}^{m \cdot 2^k}$.
Choosing k such that $m \cdot 2^k$ is a polynomial of $k \cdot n$.

Exercise 5

1. Recall the spectral norm of a matrix A, written $\|A\|$ to be the maximum $\|Av\|_2$ for unit v. Let A be symmetric stochastic, i.e., $A = A^{\mathrm{T}}$, and every row and column of A has nonnegative entries summing up to 1. Prove that $\|A\| \leq 1$.
2. Let A, B be symmetric stochastic matrices. Prove that $\lambda(A+B) \leq \lambda(A) + \lambda(B)$.
3. Let A, B be two $n \times n$ matrices.
   (a) Prove that $\|A+B\| \leq \|A\| + \|B\|$.
   (b) Prove that $\|AB\| \leq \|A\| \cdot \|B\|$

Exercise 5 - proof

For 1.
First,

$$\|A\| \leq n^2.$$

Second, for every such A,

- A^2 is symmetric stochastic matrix
- $$\|A^2\| \geq \|A\|^2.$$
- If there is an A such that $\|A\| = 1 + \alpha$ for $\alpha > 0$. Then there is such a B with $\|B\|$ unbounded.

Exercise 6

Let G be an (n, d, λ)-expander graph, and $\mathcal{B}$ be a set of vertices of size at most βn for $0 < \beta < 1$. Let $X_1, X_2, \cdots, X_k$ be a random walk of k steps in G from X_1 that is randomly and uniformly chosen.

1. Prove that for every subset $I \subseteq [k]$,

$$\Pr[(\forall i \in I)[X_i \in \mathcal{B}]] \leq \left((1-\lambda)\sqrt{\beta} + \lambda\right)^{|I|-1}.$$

2. Conclude that if $\mathcal{B} < n/100$ and $\lambda < 1/100$, then the probability that there exists a subset $I \subseteq [k]$ such that $|I| > k/10$ and $\forall_{i \leq |I|} X_i \in \mathcal{B}$ is at most $2^{-k/100}$.
3. To show that every BPP algorithm that uses m coins and decides a language L with probability 0.99 into an algorithm B that uses $m + O(k)$ coins and decides the language L with probability $1 - 2^{-k}$.

Exercise 6: Proof - I

For each i, $1 \leq i \leq k$, let
B_i: the event $X_i \in \mathcal{B}$. For $I \subseteq [k]$, let $I = \{j_1 < j_2 < \cdots j_i\}$. Then:

$$\begin{aligned} & \Pr[\wedge_{i \in I} B_i] \\ = \; & \Pr[B_{j_1}] \cdot \Pr[B_{j_2} | B_{j_1}] \cdot \cdots \cdot \Pr[B_{j_i} | B_{j_1}, \cdots, B_{j_{i-1}}]. \end{aligned} \quad (15)$$

Define B to be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ that keeps the values indexed in $\mathcal{B}$. That is, for $(u_1, u_2, \cdots, u_n)$, define

$$(Bu)_i = \begin{cases} u_i, & \text{if } i \in \mathcal{B}, \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 6: Proof - II

For every probability vector p,
(i) Bp is the vector whose coordinates sum to the probability that a vertex i is chosen according to p, is in $\mathcal{B}$.
(ii) The normalised Bp is the distribution of p conditioned to the event that the vertex is in $\mathcal{B}$.

Exercise 6: Proof - III

Let p^i be the distribution of X_j conditioned on the events $B_{j_1}, \cdots, B_{j_i}$. Then:

$$p^1 = \frac{1}{\Pr[B_{j_1}]} \cdot B1$$

$$p^2 = \frac{1}{\Pr[B_{j_2}|B_{j_1}]\Pr[B_{j_1}]} BAB1$$

$$p^i = \frac{1}{\Pr[B_{j_i}\ |B_{j_{i-1}}, \cdots B_{j_1}] \cdots \Pr[B_{j_1}]} (BA)^{i-1} B1.$$

Hence,

$$\Pr[B_{j_1}] \cdots \Pr[B_{j_i}|B_{j_{i-1}} \cdots B_{j_1}] p^i = (BA)^{i-1} B1.$$

Exercise 6: Proof - IV

$$\Pr[\wedge_{j\in I} B_j] = \Pr[B_1]\cdots\Pr[B_{j_i}|B_{j_{i-1}}\cdots B_{j_1}] = \|(BA)^{i-1}B1\|_1.$$

Let $A=(1-\lambda)J+\lambda C$ for some C with $\|C\|\le 1$.
Then $BA=(1-\lambda)BJ+\lambda BC$.
Noting:
(i) $\|B1\|_2 \le \sqrt{\beta}\|1\|_2$
(ii) $\|BJ\|\le\sqrt{\beta}$, $\|B\|\le 1$, $\|BC\|\le 1$.
(iii) $\|BA\|\le(1-\lambda)\sqrt{\beta}+\lambda$
Therefore,

$$\begin{aligned} |(BA)^{i-1}B1|_1 &\le \|(BA)^{i-1}B1\|_2\cdot\sqrt{n} \\ &\le ((1-\lambda)\sqrt{\beta}+\lambda)^{i-1}. \end{aligned} \tag{16}$$

Exercise 7

(1) Give a probabilistic polynomial time algorithm that given a 3CNF formula ϕ with exactly three distinct variables in each clause, outputs an assignment satisfying at least a $\frac{7}{8}$ fraction of ϕ's clauses.

(2) Give a deterministic polynomial time algorithm with the same approximation guarantee as Exercise 1 above.

(3) Show a polynomial time algorithm that given a satisfiable 2CSP instance ϕ over binary alphabet with m clauses outputs a satisfying assignment for ϕ.

(4) Show a deterministic poly $(n, 2^q)$-time algorithm that given a qCSP-instance ϕ over binary alphabet with m clauses outputs an assignment satisfying $m/2^q$ of the constraints of ϕ.

Exercise 7 - proof

Easy

Exercise 8

(5) Suppose that $G = (V, E)$ is an (n, d, λ)-expander. Show that for any $S \subset V$ of size $\leq \frac{n}{2}$, the following holds:

$$\Pr_{(u,v) \in_{\mathrm{R}} E}[u \in S \wedge v \in S] \leq \frac{|S|}{n}\left(\frac{1}{2} + \frac{\lambda}{2}\right).$$

Exercise 8 - proof- 1

$$\begin{aligned}&\Pr_{e=(u,v)\in_{\mathrm{R}}E}[u\in S \,\&\, v\in S]\\ =\;&\Pr[u\in S]\cdot\Pr[v\in S\mid u\in S].\end{aligned}$$

Clearly,

$$\Pr[u\in S]=\frac{s}{n},$$

where $s=|S|$.
Recall the expander mixing lemma, for any X, and Y,

$$|e(X,Y)-\frac{\mathrm{vol}\,X\cdot\mathrm{vol}\,Y}{\mathrm{vol}\,G}|\leq\lambda\sqrt{\mathrm{vol}\,X\cdot\mathrm{vol}\,Y}.$$

Exercise 8 - proof -2

For $X = Y = S$, using the lemma,

$$\Pr[v \in S \mid u \in S] \leq \frac{1}{2}(1 + \lambda).$$