Sublinear Algorithms

LECTURE 17

Last time

- Canonical testers for the dense graph model
- Approximating the average degree

Today

- Finish approximating the average degree
- Testing linearity of Boolean functions

[Blum Luby Rubinfeld]

Average Degree Estimation [Eden Ron Seshadhri]

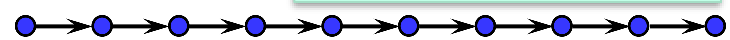
Intuition: To reduce variance,
we will ``count'' each edge towards its endpoint with smaller degree.

- Define ordering on V: for $u, v \in V$, we say $u \prec v$ if $d(u) < d(v)$ or if $d(u) = d(v)$ *and* $id(u) < id(v)$.

to break ties

- ``Orient'' the edges towards higher-degree nodes

vertices with higher degree



Algorithm (Input: ε, n; degree and neighbor query access to $G=(V,E)$)

1. Set $k = \frac{12}{\varepsilon^2} \cdot \sqrt{n}$ and initialize $X_i = 0$ for all $i \in [k]$
2. For $i = 1$ to k **do**

   Define $N(v)$ to be the set of neighbors of v.

   a. Sample a vertex $u \in V$ u.i.r. and query its degree $d(u)$
   b. Sample a vertex $v \in N(u)$ u.i.r. by making a neighbor query to v.
   c. If $u \prec v$, set $X_i = 2d(u)$
3. Return $\hat{d} = \frac{1}{k} \cdot \sum_{i \in [k]} X_i$

Outdegree Lemma

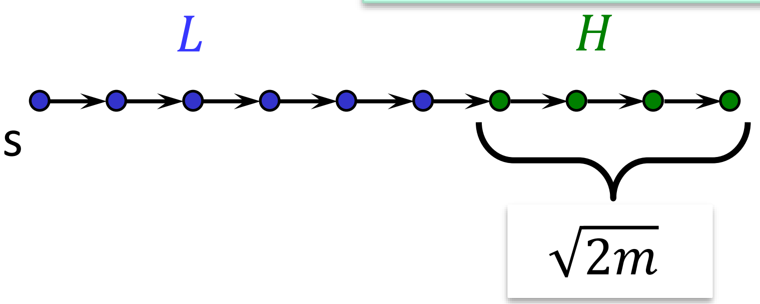
Let $d^+(u)$ denote the number of neighbors v of u with $u \prec v$.

Outdegree Lemma

For all vertices v, the outdegree $d^+(v) < \sqrt{2m}$.

vertices with the highest degree

Proof:



- Let $H \subseteq V$ be the set of the $\sqrt{2m}$ vertices with highest rank according to $\prec$.
- Let $L = V \backslash \mathrm{H}$.

1. Consider $v \in H$.
   $d^+(v)$ is the number of neighbors of v of rank higher than v.
   v is among the $\sqrt{2m}$ vertices of the highest rank, so $d^+(v) < \sqrt{2m}$
2. Consider $v \in L$. All $u \in H$, by definition, have degree at least $d(v)$.

   Then the sum of all degrees, $2m$, is greater than $\sqrt{2m} \cdot d(v)$.

$$d^{+(v)} \leq d(v) < \frac{2m}{\sqrt{2m}} = \sqrt{2m}$$

Analysis: Expectation

Algorithm (**Input:** ε, n; vertex and neighbor query access to $G=(V,E)$)

1. Set $k = \frac{12}{\varepsilon^2} \cdot \sqrt{n}$ and initialize $X_i = 0$ for all $i \in [k]$
2. For $i = 1$ to k **do**
   a. Sample a vertex $u \in V$ u.i.r. and query its degree $d(u)$
   b. Sample a vertex $v \in N(u)$ u.i.r. by making a neighbor query to v.
   c. If $u \prec v$, set $X_i = 2d(u)$
3. Return $\hat{d} = \frac{1}{k} \cdot \sum_{i \in [k]} X_i$

- Let $d^+(u)$ denote the number of neighbors v of u with $u \prec v$.
- Let X denote one of the variables X_i. (They all have the same distribution.)
- Let U denote the random variable equal to the node u sampled in Step 2a.

$$\mathbb{E}[X] = \mathbb{E}\big[\mathbb{E}[X|U]\big]$$

By the compact form of the Law of Total Expectation

$$\mathbb{E}[X|U] = \frac{d^+(U)}{d(U)} \cdot 2d(U) = 2d^+(U).$$

$d^+(U)$ is # of neighbors v of U for which $X = 2d(U)$

$$\mathbb{E}[X] = \mathbb{E}[2d^+(U)] = 2\sum_{u \in V} \frac{1}{n} \cdot d^+(u) = \frac{2m}{n} = \bar{d}$$

Analysis: Variance

Reminders:
$d^+(u) =$ the # of neighbors v of u with $u \prec v$.
RV X denotes X_i.
RV $U =$ the node u sampled in Step 2a.

- $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 < \mathbb{E}[X^2]$
- $\mathbb{E}[X^2] = \left[\mathbb{E}[X^2 | U]\right]$

By the compact form of the Law of Total Expectation

- $\mathbb{E}[X^2 | U] = \frac{d^+(U)}{d(U)} \cdot \left(2d(U)\right)^2 = 4d^+(U) \cdot d(U).$

Outdegree Lemma

$\forall v, \quad d^+(v) < \sqrt{2m}.$

- $\mathbb{E}[X^2] = \mathbb{E}[4d^+(U) \cdot d(U)]$

$$< \mathbb{E}\left[4 \cdot \sqrt{2m} \cdot d(U)\right]$$
$$= 4\sqrt{2m} \cdot \mathbb{E}[d(U)]$$
$$= 4\sqrt{2m} \cdot \bar{d}.$$

By linearity of expectation

By definition of expectation

We get that $\quad \text{Var}[X] < 4\sqrt{2m} \cdot \bar{d}.$

Analysis: Putting It All Together

Algorithm (**Input:** ε, n; vertex and neighbor query access to $G=(V,E)$)

1. Set $k = \frac{12}{\varepsilon^2} \cdot \sqrt{n}$ and initialize $X_i = 0$ for all $i \in [k]$
2. For $i = 1$ to k **do**
   a. Sample a vertex $u \in V$ u.i.r. and query its degree $d(u)$
   b. Sample a vertex $v \in N(u)$ u.i.r. by making a neighbor query to v.
   c. If $u \prec v$, set $X_i = 2d(u)$
3. Return $\hat{d} = \frac{1}{k} \cdot \sum_{i \in [k]} X_i$

$d^+(u) =$ the # of neighbors v of u with $u \prec v$.
RV X denotes X_i.
RV $U =$ the node u sampled in Step 2a.

- $\mathbb{E}[\hat{d}] = \mathbb{E}[X] = \bar{d}$
- $\mathrm{Var}[\hat{d}] = \frac{\mathrm{Var}[X]}{k} \le \frac{4\sqrt{2m} \cdot \bar{d}}{k}$
- $\Pr[|\hat{d} - \bar{d}| \ge \varepsilon \cdot \bar{d}] = \Pr[|\hat{d} - \mathbb{E}[\hat{d}]| \ge \varepsilon \cdot \bar{d}] \overset{\text{By Chebyshev}}{\le} \frac{\mathrm{Var}[\hat{d}]}{(\varepsilon \cdot \bar{d})^2}$

$$\le \frac{4\sqrt{2m} \cdot \bar{d}}{k \cdot \varepsilon^2 \cdot \bar{d}^2} = \frac{4\sqrt{2m} \cdot n}{k \cdot \varepsilon^2 \cdot 2m} = \frac{4n}{k \cdot \varepsilon^2 \cdot \sqrt{2m}} = \frac{4\sqrt{n}}{k \cdot \varepsilon^2 \cdot \sqrt{\bar{d}}} \overset{\text{Our choice of } k}{=} \frac{1}{3\sqrt{\bar{d}}} \overset{\text{Assumption } \bar{d} \ge 1}{\le} \frac{1}{3}$$

Approximating the Average Degree: Run Time

Algorithm (Input: ε, n; vertex and neighbor query access to $G=(V,E)$)

1. Set $k = \frac{12}{\varepsilon^2} \cdot \sqrt{n}$ and initialize $X_i = 0$ for all $i \in [k]$
2. For $i = 1$ to k **do**
   a. Sample a vertex $u \in V$ u.i.r. and query its degree $d(u)$
   b. Sample a vertex $v \in N(u)$ u.i.r. by making a neighbor query to v.
   c. If $u \prec v$, set $X_i = 2d(u)$
3. Return $\hat{d} = \frac{1}{k} \cdot \sum_{i\in[k]} X_i$

Running time:

$$O\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$$

to get $\Pr\left[\left|\hat{d} - \bar{d}\right| \geq \varepsilon \cdot \bar{d}\right] \leq \frac{1}{3}$

Technical Writing Tips: Citations

- Be generous in acknowledging the source of your ideas.

Don't be a citation scrooge. If someone inspired you, give them credit.

- Use dblp to get citations in the bibtex format.

Because manually typing BibTeX is a rite of passage only once.
After that, it's just masochism.

- Fix issues with capitalization in BibTeX items by using {curly braces}: e.g., ``{LCAs} for {Lipschitz} functions.'' (You should also add dollar signs around math expressions).

Dblp likes to lowercase Everything Like It's Stuck in the '90s.

BibTeX deserves LaTeX, too.

- If there are multiple versions of the paper, cite the most recently published one. (Archival versions are not considered published).

  -- journal > conference > archival preprint

  > that PDF on someone's website called "final_final_revised3_REAL.pdf"

  -- Beware of paper mergers.

- When you cite multiple papers, give multiple arguments to the same \cite command. The result will look like, for example: [BLR93,GGR98].

Technical Writing Questions

For each of the sentences,[1] specify what needs fixing and how you'd fix it.

1. A found p_{square} f-violated squares and $q_{triangle}$ g-violated triangles.

2. We proved interesting results in our project. We will describe them below. We focus on the lower bounds.

3. Given a n-node graph G, design an efficient algorithm that estimates the average degree of G up to a factor of 2[2]

4. If this works for any function, our algorithm will succeed.

5. We note that we plan to spend some of the future time devoting it to coming up with the right notion of the model.

6. Let's eat grandma.

---

[1] Any resemblance to sentences in course projects is purely coincidental.
[2] The reason we cannot beat the factor of 2 is...

Linear Functions Over Finite Field $\mathbb{F}_2$

A Boolean function $f: \{0,1\}^n \to \{0,1\}$ is *linear* if

$$f(x_1, \dots, x_n) = a_1 x_1 + \cdots + a_n x_n \text{ for some } a_1, \dots, a_n \in \{0,1\}$$

no free term

- Work in finite field $\mathbb{F}_2$
  - Other accepted notation for $\mathbb{F}_2$: GF_2 and $\mathbb{Z}_2$
  - Addition and multiplication is mod 2
  - $\boldsymbol{x}=(x_1, \dots, x_n)$, $\boldsymbol{y}=(y_1, \dots, y_n)$, that is, $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$
    $\boldsymbol{x} + \boldsymbol{y}=(x_1 + y_1, \dots, x_n + y_n)$

example

$$\begin{array}{r} 001001 \\ +\ 011001 \\ \hline 010000 \end{array}$$

Testing If a Boolean Function Is Linear

Input: Boolean function $f: \{0,1\}^n \to \{0,1\}$

Question:

Is the function linear or ε-far from linear

($\geq \varepsilon 2^n$ values need to be changed to make it linear)?

Today: can answer in $O\left(\frac{1}{\varepsilon}\right)$ time

Motivation

- Linearity test is one of the most celebrated testing algorithms
  - A special case of many important property tests
  - Computations over finite fields are used in
    - Cryptography
    - Coding Theory
  - Originally designed for program checkers and self-correctors
  - Low-degree testing is needed in constructions of Probabilistically Checkable Proofs (PCPs)
    - Used for proving inapproximability
- Main tool in the correctness proof: Fourier analysis of Boolean functions
  - Powerful and widely used technique in understanding the structure of Boolean functions

Equivalent Definitions of Linear Functions

Definition. f is *linear* if $f(x_1, \dots, x_n) = a_1 x_1 + \cdots + a_n x_n$ for some $a_1, \dots, a_n \in \mathbb{F}_2$

$$\Updownarrow$$

$[n]$ is a shorthand for $\{1, \dots n\}$

$f(x_1, \dots, x_n) = \sum_{i \in S} x_i$ for some $S \subseteq [n]$.

Definition′. f is *linear* if $f(\boldsymbol{x} + \boldsymbol{y}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$ for all $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$.

- Definition ⇒ Definition′

$$f(\boldsymbol{x} + \boldsymbol{y}) = \sum_{i \in S} (\boldsymbol{x} + \boldsymbol{y})_i = \sum_{i \in S} (x_i + y_i) = \sum_{i \in S} x_i + \sum_{i \in S} y_i = f(\boldsymbol{x}) + f(\boldsymbol{y}).$$

- Definition′ ⇒ Definition

Let $\alpha_i = f((\overbrace{0, \dots, 0, 1, 0, \dots, 0}^{e_i}))$

Repeatedly apply Definition′:

$$f\big((x_1, \dots, x_n)\big) = f\left(\sum x_i e_i\right) = \sum x_i f(e_i) = \sum \alpha_i x_i.$$

Based on Ryan O'Donell's lecture notes: http://www.cs.cmu.edu/~odonnell/boolean-analysis/

Linearity Test [Blum Luby Rubinfeld 90]

BLR Test (ε, query access to f)

1. Pick $\boldsymbol{x}$ and $\boldsymbol{y}$ independently and uniformly at random from $\{0,1\}^n$.
2. Set $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$ and query f on $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{z}$. **Accept** iff $f(\boldsymbol{z}) = f(\boldsymbol{x}) + f(\boldsymbol{y})$.

Analysis

If f is linear, BLR always accepts.

Correctness Theorem [Bellare Coppersmith Hastad Kiwi Sudan 95]

If f is ε-far from linear then $> \varepsilon$ fraction of pairs $\boldsymbol{x}$ and $\boldsymbol{y}$ fail BLR test.

- Then, by Witness Lemma (Lecture 1), $2/\varepsilon$ iterations suffice.

Analysis Technique: Fourier Expansion

Representing Functions as Vectors

Stack the 2^n values of $f(\boldsymbol{x})$ and treat it as a vector in $\{0,1\}^{2^n}$.

$$f = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ 0 \\ 0 \end{bmatrix} \qquad \begin{bmatrix} f(0000) \\ f(0001) \\ f(0010) \\ f(0011) \\ f(0100) \\ \cdot \\ \cdot \\ \cdot \\ f(1101) \\ f(1110) \\ f(1111) \end{bmatrix}$$

Linear functions

There are 2^n linear functions: one for each subset $S \subseteq [n]$.

$$\chi_\emptyset = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \chi_{\{1\}} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ 0 \\ 1 \end{bmatrix}, \qquad \cdots \cdots, \qquad \chi_{[n]} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

Parity on the positions indexed by set S is $\chi_S(x_1, \ldots, x_n) = \sum_{i \in S} x_i$

Great Notational Switch

Idea: Change notation, so that we work over reals instead of a finite field.

- Vectors in $\{0,1\}^{2^n}$ $\longrightarrow$ Vectors in $\mathbb{R}^{2^n}$.
- 0/False $\longrightarrow$ 1 1/True $\longrightarrow$ -1.
- Addition (mod 2) $\longrightarrow$ Multiplication in $\mathbb{R}$.
- Boolean function: $f : \{-1,1\}^n \to \{-1,1\}$.
- Linear function χ_S: $\{-1,1\}^n \to \{-1,1\}$ is given by $\chi_S(\boldsymbol{x}) = \prod_{i \in S} x_i$.

Benefit 1 of New Notation

- The dot product of f and g as vectors in $\{-1,1\}^{2^n}$:

$$(\#\ \boldsymbol{x}\text{'s such that } f(\boldsymbol{x}) = g(\boldsymbol{x})) - (\#\ \boldsymbol{x}\text{'s such that } f(\boldsymbol{x}) \neq g(\boldsymbol{x}))$$

$$= 2^n - 2 \cdot \underbrace{(\#\ \boldsymbol{x}\text{'s such that } f(\boldsymbol{x}) \neq g(\boldsymbol{x}))}_{\textit{disagreements between } f \textit{ and } g}$$

Inner product of functions $f, g : \{-1,1\}^n \to \{-1,1\}$

$$\langle f, g \rangle = \frac{1}{2^n}(\text{dot product of } f \text{ and } g \text{ as vectors})$$

$$= \operatorname*{avg}_{\boldsymbol{x} \in \{-1,1\}^n}[f(\boldsymbol{x})g(\boldsymbol{x})] = \mathop{\mathbb{E}}_{\boldsymbol{x} \in \{-1,1\}^n}[f(\boldsymbol{x})g(\boldsymbol{x})].$$

$\langle f, g \rangle = 1 - 2 \cdot$ (fraction of *disagreements* between f and g)

Benefit 2 of New Notation

Claim. The functions $(\chi_S)_{S\subseteq[n]}$ form an orthonormal basis for $\mathbb{R}^{2^n}$.

- If $S \neq T$ then χ_S and χ_T are orthogonal: $\langle \chi_S, \chi_T \rangle = 0$.
  - Let i be an element on which S and T differ (w.l.o.g. $i \in S \setminus T$)
  - Pair up all n-bit strings: $(\boldsymbol{x}, \boldsymbol{x}^{(i)})$ where $\boldsymbol{x}^{(i)}$ is $\boldsymbol{x}$ with the i^{th} bit flipped.
  - Each such pair contributes $ab - ab = 0$ to $\langle \chi_S, \chi_T \rangle$.
  - Since all $\boldsymbol{x}$'s are paired up, $\langle \chi_S, \chi_T \rangle = 0$.
- Recall that there are 2^n linear functions χ_S .
- $\langle \chi_S, \chi_S \rangle = 1$
  - In fact, $\langle f, f \rangle = 1$ for all $f : \{-1,1\}^n \to \{-1,1\}$.
  - (The norm of f, denoted $|f|$, is $\sqrt{\langle f, f \rangle}$)

$$\begin{array}{c|c|c|}
 & +1 & -1 \\
 & -1 & +1 \\
 & +1 & +1 \\
\boldsymbol{x} & +a & b \\
 & +1 & +1 \\
 & \vdots & \vdots \\
\boldsymbol{x}^{(i)} & -a & b \\
 & +1 & -1 \\
 & -1 & +1 \\
 & -1 & +1 \\
 & \chi_S & \chi_T
\end{array}$$

Fourier Expansion Theorem

Idea: Work in the basis $(\chi_S)_{S\subseteq[n]}$, so it is easy to see how close a specific function f is to each of the linear functions.

Fourier Expansion Theorem

Every function $f : \{-1,1\}^n \to \mathbb{R}$ is uniquely expressible as a linear combination (over $\mathbb{R}$) of the 2^n linear functions:

$$f = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S,$$

where $\hat{f}(S) = \langle f, \chi_S\rangle$ is the Fourier Coefficient of f on set S.

Proof: f can be written uniquely as a linear combination of basis vectors:

$$f = \sum_{S\subseteq[n]} c_S \cdot \chi_S$$

It remains to prove that c_S=$\hat{f}(S)$ for all S.

$$\hat{f}(S) = \langle f, \chi_S\rangle = \left\langle \sum_{T\subseteq[n]} c_T \cdot \chi_T, \chi_S \right\rangle = \sum_{T\subseteq[n]} c_T \cdot \langle \chi_T, \chi_S\rangle = c_S$$

Definition of Fourier coefficients

Linearity of $\langle\cdot,\cdot\rangle$

$$\langle \chi_T, \chi_S\rangle = \begin{cases} 1 & \text{if } T = S \\ 0 & \text{otherwise} \end{cases}$$

Examples: Fourier Expansion

$\boldsymbol{f}$	**Fourier transform**
$f(\boldsymbol{x}) = 1$	1
$f(\boldsymbol{x}) = x_i$	x_i
$\mathrm{AND}(x_1, x_2)$	$\frac{1}{2} + \frac{1}{2}x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_1x_2$
$\mathrm{MAJORITY}(x_1, x_2, x_3)$	$\frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_3 - \frac{1}{2}x_1x_2x_3$