Randomized Computation
Eugene Santos looked at computability for Probabilistic TM.
John Gill studied complexity classes defined by Probabilistic TM.

Synopsis

1. Tail Distribution
2. Probabilistic Turing Machine
3. \( PP \)
4. \( BPP \)
5. \( ZPP \)
6. Random Walk and \( RL \)
Tail Distribution
Markov’s Inequality

For all \( k > 0 \),

\[
\Pr[X \geq kE[X]] \leq \frac{1}{k},
\]

or equivalently

\[
\Pr[X \geq \nu] \leq \frac{E[X]}{\nu}.
\]

- Observe that \( d \cdot \Pr[X \geq d] \leq E[X] \).
- We are done by letting \( d = kE[X] \).
Moment and Variance

Information about a random variable is often expressed in terms of moments.

- The $k$-th moment of a random variable $X$ is $E[X^k]$.

The variance of a random variable $X$ is

$$
$$

The standard deviation of $X$ is

$$
\sigma(X) = \sqrt{\text{Var}(X)}.
$$

**Fact.** If $X_1, \ldots, X_n$ are pairwise independent, then

$$
\text{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \text{Var}(X_i).
$$
Chebyshev Inequality

For all $k > 0$,

$$\Pr[|X - E[X]| \geq k\sigma] \leq \frac{1}{k^2},$$

or equivalently

$$\Pr[|X - E[X]| \geq k] \leq \frac{\sigma^2}{k^2}.$$

Apply Markov’s Inequality to the random variable $(X - E[X])^2$. 
The moment generating function of a random variable $X$ is $M_X(t) = \mathbb{E}[e^{tX}]$.

- If $X$ and $Y$ are independent, then $M_{X+Y}(t) = M_X(t)M_Y(t)$.
- If differentiation commutes with expectation then the $n$-th moment $\mathbb{E}[X^n] = M_X^{(n)}(0)$.

1. If $t > 0$ then $\Pr[X \geq a] = \Pr[e^{tX} \geq e^{ta}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$. Hence $\Pr[X \geq a] \leq \min_{t > 0} \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$.

2. If $t < 0$ then $\Pr[X \leq a] = \Pr[e^{tX} \geq e^{ta}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$. Hence $\Pr[X \leq a] \leq \min_{t < 0} \frac{\mathbb{E}[e^{tX}]}{e^{ta}}$.

For a specific distribution one chooses some $t$ to get a convenient bound. Bounds derived by this approach are collectively called Chernoff bounds.
Chernoff Bounds for Poisson Trials

Let $X_1, \ldots, X_n$ be independent Poisson trials with $\Pr[X_i = 1] = p_i$. Let $X = \sum_{i=1}^{n} X_i$.

- $M_{X_i}(t) = \mathbb{E}[e^{tX_i}] = p_i e^t + (1 - p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t - 1)}$. [$1 + x \leq e^x$]

- Let $\mu = \mathbb{E}[X] = \sum_{i=1}^{n} p_i$. Then

$$M_X(t) \leq e^{(e^t - 1)\mu}.$$

For Bernoulli trials

$$M_X(t) \leq e^{(e^t - 1)np}.$$
Chernoff Bounds for Poisson Trials

**Theorem.** Suppose $0 < \delta < 1$. Then

\[
\Pr [X \geq (1 + \delta)\mu] \leq \left[\frac{e^\delta}{(1 + \delta)^{(1+\delta)}}\right]^\mu \leq e^{-\mu\delta^2/3}, \\
\Pr [X \leq (1 - \delta)\mu] \leq \left[\frac{e^{-\delta}}{(1 - \delta)^{(1-\delta)}}\right]^\mu \leq e^{-\mu\delta^2/2}.
\]

**Corollary.** Suppose $0 < \delta < 1$. Then

\[
\Pr [|X - \mu| \geq \delta\mu] \leq 2e^{-\mu\delta^2/3}.
\]

If $t > 0$ then $\Pr[X \geq (1 + \delta)\mu] = \Pr[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu}} \leq \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}}$. We get the first inequality by setting $t = \ln(1 + \delta)$. For $t < 0$ we set $t = \ln(1 - \delta)$. 


When using pairwise independent samples, the error probability decreases \textit{linearly} with the number of samples.

When using totally independent samples, the error probability decreases \textit{exponentially} with the number of samples.


Probabilistic Turing Machine
A Probabilistic Turing Machine (PTM) $\mathbb{P}$ is a Turing Machine with two transition functions $\delta_0, \delta_1$.

- To execute $\mathbb{P}$ on an input $x$, we choose in each step with probability $1/2$ to apply transition function $\delta_0$ and with probability $1/2$ to apply transition function $\delta_1$.
- All choices are independent.

We denote by $\mathbb{P}(x)$ the random variable corresponding to the value $\mathbb{P}$ produces on input $x$.

$\Pr[\mathbb{P}(x) = y]$ is the probability of $\mathbb{P}$ outputting $y$ on the input $x$. 

Probabilistic TM vs Nondeterministic TM:

1. What does it mean for a PTM to compute a function?
2. How about time complexity?
A function $\phi$ is computable by a PTM $\mathbb{P}$ in the following sense:

$$\phi(x) = \begin{cases} y, & \text{if } \Pr[\mathbb{P}(x) = y] > 1/2, \\ \uparrow, & \text{if no such } y \text{ exists.} \end{cases}$$
Probabilistically Decidable Problem

A language $L$ is decided by a PTM $\mathbb{P}$ if the following holds:

$$\Pr[\mathbb{P}(x) = L(x)] > 1/2.$$
Fact. The functions computable by PTM’s are precisely the computable functions.

Proof.
By fixing a Gödel encoding, it is routine to prove S-m-n Theorem, Enumeration Theorem and Recursion Theorem.

PTM’s are equivalent to TM’s from the point of view of computability.
Blum Time Complexity for Probabilistic Turing Machine

**Definition** (Trakhtenbrot, 1975; Gill, 1977). The Blum time complexity $T_i$ of PTM $P_i$ is defined by

$$T_i(x) = \begin{cases} \mu n. \Pr[P_i(x) = \phi_i(x) \text{ in } n \text{ steps}] > 1/2, & \text{if } \phi_i(x) \downarrow, \\ \uparrow, & \text{if } \phi_i(x) \uparrow. \end{cases}$$

Neither the average time complexity nor the worst case time complexity is a Blum complexity measure.
Average Case Time Complexity

It turns out that average time complexity is a pathological complexity measure.

Lemma (Gill, 1977). Every recursive set is decided by some PTM with constant average run time.

Proof.
Suppose recursive set $W$ is decided by TM $M$. Define PTM $P$ by

- repeat
  - simulate one step of $M(x)$;
    - if $M(x)$ accepts then accept; if $M(x)$ rejects then reject;
  - until head;
    - if head then accept else reject.

The average run time is bounded by a small constant.
Worst Case Time Complexity

A PTM $\mathbb{P}$ runs in $T(n)$-time if for any input $x$, $\mathbb{P}$ halts on $x$ within $T(|x|)$ steps regardless of the random choices it makes.

The worst case time complexity is subtle since the execution tree of a PTM upon receiving an input is normally unbounded.

- The problem is due to the fact that the error probability $\rho(x)$ could tend to $1/2$ fast, for example $\rho(x) = 1/2 - 2^{-2|x|}$. 

Computation with Bounded Error

A function $\phi$ is computable by a PTM $\mathbb{P}$ with bounded error probability if there is some positive $\epsilon < 1/2$ such that for all $x, y$

$$\phi(x) = \begin{cases} y, & \text{if } \Pr[\mathbb{P}(x) = y] \geq 1/2 + \epsilon, \\ \uparrow, & \text{if no such } y \text{ exists.} \end{cases}$$

Both average time complexity and worst case time complexity are good for bounded error computability.
Biased Random Source

In practice our coin is pseudorandom. It has a face-up probability $\rho \neq 1/2$.

PTM's with biased random choices $\equiv$ PTM's with fair random choices?
Biased Random Source

**Fact.** A coin with $P_r[Heads] = 0.p_1p_2p_3\ldots$ can be simulated by a PTM in expected $O(1)$ time if $p_i$ is computable in $\text{poly}(i)$ time.

Our PTM $\mathbb{P}$ generates a sequence of random bits $b_1, b_2, \ldots$ one by one.

- If $b_i < p_i$, the machine outputs ‘Head’ and stops;
- If $b_i > p_i$, the machine outputs ‘Tail’ and stops;
- If $b_i = p_i$, the machine goes to step $i + 1$.

$\mathbb{P}$ outputs ‘Head’ at step $i$ if $b_i < p_i \land \forall j < i.b_j = p_j$, which happens with probability $1/2^i$.

Thus the probability of ‘Heads’ is $\sum_i p_i \frac{1}{2^i} = 0.p_1p_2p_3\ldots$.

The expected number of coin flipping is $\sum_i i \frac{1}{2^i} = 2$. 
Fact. (von Neumann, 1951) A coin with $\Pr[\text{Heads}] = 1/2$ can be simulated by a PTM with access to a $\rho$-biased coin in expected time $O(1)$.

The machine tosses pairs of coin until it gets ‘Head-Tail’ or ‘Tail-Head’. In the former case it outputs ‘Head’, and in the latter case it outputs ‘Tail’.

The probability of ‘Head-Tail’/‘Tail-Head’ is $\rho(1 - \rho)$.

The expected running time is $1/2\rho(1 - \rho)$. 
Finding the $k$-th Element

$$\text{FINDKthELEMENT}(k, \{a_1, \ldots, a_n\})$$

1. Pick a random $i \in [n]$ and let $x = a_i$.
2. Count the number $m$ of $a_j$'s such that $a_j \leq x$.
3. Split $a_1, \ldots, a_n$ to two lists $L \leq x < H$ by the pivotal element $x$.
4. If $m = k$ then output $x$.
5. If $m > k$ then $\text{FINDKthELEMENT}(k, L)$.
6. If $m < k$ then $\text{FINDKthELEMENT}(k - m, H)$.
Finding the $k$-th Element

Let $T(n)$ be the expected worst case running time of the algorithm.

Suppose the running time of the nonrecursive part is $cn$.

We prove by induction that $T(n) \leq 10cn$.

\[
T(n) \leq cn + \frac{1}{n} \left( \sum_{j > k} T(j) + \sum_{j < k} T(n - j) \right)
\]
\[
\leq cn + \frac{10c}{n} \left( \sum_{j > k} j + \sum_{j < k} (n - j) \right)
\]
\[
\leq 10cn.
\]

This is a ZPP algorithm.
Polynomial Identity Testing

An algebraic circuit has gates implementing $+, -, \times$ operators.

$\text{ZERO}$ is the set of algebraic circuits calculating the zero polynomial.

- Given polynomials $p(x)$ and $q(x)$, is $p(x) = q(x)$?
Polynomial Identity Testing

Let $C$ be an algebraic circuit. The polynomial computed by $C$ has degree at most $2^{|C|}$.

Our algorithm does the following:

1. Randomly choose $x_1, \ldots, x_n$ from $[10 \cdot 2^{|C|}]$;
2. Accept if $C(x_1, \ldots, x_n) = 0$ and reject otherwise.

By Schwartz-Zippel Lemma, the error probability is at most $1/10$. However the intermediate values could be as large as $(10 \cdot 2^{|C|})^{2^{|C|}}$.

Schwartz-Zippel Lemma. If a polynomial $p(x_1, x_2, \ldots, x_n)$ over $GF(q)$ is nonzero and has total degree at most $d$, then $\Pr_{a_1, \ldots, a_n \in GF(q)}[p(a_1, \ldots, a_n) \neq 0] \geq 1 - d/q.$
Polynomial Identity Testing

A solution is to use the so-called *fingerprinting* technique. Let $m = |C|$.

- Evaluation is carried out modulo a number $k \in \mathbb{R}[2^m]$.
- With probability at least $1/4m$, $k$ does not divide $y$ if $y \neq 0$.
  - There are at least $\frac{2^m}{2m}$ prime numbers in $[2^m]$.
  - $y$ can have at most $\log y = O(m2^m)$ prime factors.
  - When $m$ is large enough, the number of primes in $[2^m]$ not dividing $y$ is at least $\frac{2^m}{4m}$.
- Repeat the above $4m$ times. Accept if all results are zero.

This is a **coRP** algorithm.
Lovász (1979) reduced the matching problem to the problem of zero testing of the determinant of the following matrix.

- A bipartite graph of size $2n$ is represented as an $n \times n$ matrix whose entry at $(i, j)$ is a variable $x_{i,j}$ if there is an edge from $i$ to $j$ and is 0 otherwise.

Pick a random assignment from $[2n]$ and calculate the determinant.
PP
If P-time probabilistic decidable problems are defined using worst case complexity measure without any bound on error probability, we get a complexity class that appears much bigger than $\mathbf{P}$. 
Suppose $T : \mathbb{N} \to \mathbb{N}$ and $L \subseteq \{0, 1\}^*$.

A PTM $P$ decides $L$ in time $T(n)$ if, for every $x \in \{0, 1\}^*$, $\Pr[P(x) = L(x)] > 1/2$ and $P$ halts in $T(|x|)$ steps regardless of its random choices.
We write PP for the class of problems decided by P-time PTM’s.

Alternatively $L$ is in PP if there exist a polynomial $p : \mathbb{N} \to \mathbb{N}$ and a P-time TM $M$ such that for every $x \in \{0, 1\}^*$,

$$\Pr_{r \in \{0, 1\}^{p(|x|)}}[M(x, r) = L(x)] > 1/2.$$
Another Characterization of \( \text{PP} \)

\( L \) is in \( \text{PP} \) if there exist a polynomial \( p : \mathbb{N} \rightarrow \mathbb{N} \) and a P-time TM \( M \) such that for every \( x \in \{0, 1\}^\ast \),

\[
\Pr_{r \in \{0, 1\}^{|x|}}[M(x, r) = 1] \geq 1/2, \quad \text{if } x \in L,
\]
\[
\Pr_{r \in \{0, 1\}^{|x|}}[M(x, r) = 0] > 1/2, \quad \text{if } x \notin L.
\]

1. If a computation that uses some \( \delta_1 \) transition ends up with a ‘yes’/’no’ answer, toss the coin three more times and produce seven ‘yes’s/’no’s and one ‘no’/’yes’.
2. If the computation using only \( \delta_0 \) transitions ends up with a ‘no’ answer, toss the coin and announces the result.
3. If the computation using only \( \delta_0 \) transitions ends up with a ‘yes’ answer, answers ‘yes’.
Lemma (Gill, 1977). \( \text{NP, coNP} \subseteq \text{PP} \subseteq \text{PSPACE} \).

Suppose \( L \) is accepted by some NDTM \( N \) running in P-time. Design \( \mathbb{P} \) that upon receiving \( x \) executes the following:

1. Simulate \( N(x) \) probabilistically.

2. If a computation terminates with a ‘yes’ answer, then accept; otherwise toss a coin and decide accordingly.

3. If the computation using only \( \delta_0 \) transitions ends up with a ‘no’ answer, then toss the coin two more times and produce three ‘no’s and one ‘yes’.

Clearly \( \mathbb{P} \) decides \( L \).
PP-Completeness

Probabilistic version of SAT:

1. \((\varphi, i) \in \mathcal{HSAT}\) if more than \(i\) assignments make \(\varphi\) true.
2. \(\varphi \in \text{MajSAT}\) if more than half assignments make \(\varphi\) true.

**PP-Completeness**

**Theorem** (Simon, 1975). \( \sharp \text{SAT} \) is \( \text{PP} \)-complete.

**Theorem** (Gill, 1977). \( \text{MajSAT} \leq_K \sharp \text{SAT} \leq_K \text{MajSAT} \).

1. Probabilistically produce an assignment. Then evaluate the formula under the assignment. This shows that \( \text{MajSAT} \in \text{PP} \). Completeness by Cook-Levin reduction.

2. The reduction \( \text{MajSAT} \leq_K \sharp \text{SAT} \) is clear. Conversely given \( \langle \varphi, i \rangle \), where \( \varphi \) contains \( n \) variables, construct a formula \( \psi \) with \( 2^n - 2^i - \ldots - 2^i \) true assignments, where \( i = \sum_{h=1}^{j} 2^h \).

   - For example \( (x_{k+1} \lor \ldots \lor x_n) \) has \( 2^n - 2^k \) true assignments.

Let \( x \) be a fresh variable. Then \( \langle \varphi, i \rangle \in \sharp \text{SAT} \) if and only if \( x \land \varphi \lor \overline{x} \land \psi \in \text{MajSAT} \).


Theorem. \textbf{PP} is closed under union and intersection.

\begin{itemize}
  \item 1. R. Beigel, N. Reingold and D. Spielman. PP is Closed under Intersection, STOC, 1-9, 1991.
\end{itemize}
BPP
If P-time probabilistic decidable problems are defined using worst case complexity measure with bound on error probability, we get a complexity class that is believed to be very close to $\mathbf{P}$. 
**Problem Decided by PTM with Bounded-Error**

Suppose $T : \mathbb{N} \rightarrow \mathbb{N}$ and $L \subseteq \{0, 1\}^*$. A PTM $\mathbb{P}$ with **bounded error** decides $L$ in time $T(n)$ if for every $x \in \{0, 1\}^*$, $\mathbb{P}$ halts in $T(|x|)$ steps, and $\Pr[\mathbb{P}(x) = L(x)] \geq 2/3$.

$L \in \text{BPTIME}(T(n))$ if there is some $c$ such that $L$ is decided by a PTM in $cT(n)$ time.
We write $\text{BPP}$ for $\bigcup_c B\text{PTIME}(n^c)$.

Alternatively $L \in \text{BPP}$ if there exist a polynomial $p : \mathbb{N} \rightarrow \mathbb{N}$ and a P-time TM $M$ such that for every $x \in \{0, 1\}^*$,

$$\Pr_{r \in R\{0,1\}^{|x|}}[M(x, r) = L(x)] \geq 2/3.$$
1. $P \subseteq BPP \subseteq PP$.

2. $BPP = \text{coBPP}$.
How robust is our definition of \textbf{BPP}?
**Fact.** In the definition of $\text{BPP}$, we could use the expected running time instead of the worst case running time.

Let $L$ be decided by a bounded error PTM $P$ in average $T(n)$ time. Design a PTM that simulates $P$ for $9T(n)$ steps. It outputs ‘yes’ if $P$ does not stop in $9T(n)$ steps. By Markov’s inequality the probability that $P$ does not stop in $9T(n)$ steps is at most $1/9$. 
Error Reduction Theorem

Let $\text{BPP}(\rho)$ denote the $\text{BPP}$ defined with error probability $\rho$.

Theorem. $\text{BPP}(1/2 - 1/n^c) = \text{BPP}(2^{-n^d})$ for all $c, d > 1$. 
Error Reduction Theorem

Let $L$ be decided by a bounded error PTM $P$ in $\text{BPP}(1/2 - 1/n^c)$. Design a PTM $P'$ as follows:

1. $P'$ simulates $P$ on $x$ for $k = 12|x|^{2c+d} + 1$ times, obtaining $k$ results $y_1, \ldots, y_k \in \{0, 1\}$.

2. If the majority of $y_1, \ldots, y_k$ are 1, $P'$ accepts $x$; otherwise $P'$ rejects $x$.

For each $i \in [k]$ let $X_i$ be the random variable that equals to 1 if $y_i = 1$ and is 0 if $y_i = 0$.

Let $X = \sum_{i=1}^{k} X_i$. Let $\delta = |x|^{-c}$. Let $p = 1/2 + \delta$ and $\bar{p} = 1/2 - \delta$.

- By linearity $E[X] \geq kp$ if $x \in L$, and $E[X] \leq k\bar{p}$ if $x \notin L$.

- If $x \in L$ then $\Pr[X < \frac{k}{2}] < \Pr[X < (1-\delta)kp] \leq \Pr[X < (1-\delta)E[X]] < e^{-\frac{\delta^2}{2}kp} < \frac{1}{2|x|^d}$.

- If $x \notin L$ then $\Pr[X > \frac{k}{2}] < \Pr[X > (1+\delta)k\bar{p}] \leq \Pr[X > (1+\delta)E[X]] < e^{-\frac{\delta^2}{2}k\bar{p}} < \frac{1}{2|x|^d}$.

The inequality $<$ is due to Chernoff Bound. Conclude that the error probability of $P'$ is $\leq \frac{1}{2^n^d}$.
Conclusion: In the definition of $\text{BPP}$,

- we can replace $2/3$ by a constant arbitrarily close to $1/2$;
- we can even replace $2/3$ by $\frac{1}{2} + \frac{1}{n^c}$ for any constant $c$.

Error Reduction Theorem offers a powerful tool to study BPP.
“Nonuniformity is more powerful than randomness.”

**Adleman Theorem.** \( BPP \subseteq \mathbf{P}/\text{poly}. \)

Proof of Adleman Theorem

Suppose \( L \in \text{BPP} \). There exist a polynomial \( p(x) \) and a P-time TM \( M \) such that

\[
\Pr_{r \in \{0,1\}^{p(n)}}[M(x, r) \neq L(x)] \leq 1/2^{n+1}
\]

for every \( x \in \{0,1\}^n \).

Say \( r \in \{0,1\}^{p(n)} \) is bad for \( x \in \{0,1\}^n \) if \( M(x, r) \neq L(x) \); otherwise \( r \) is good for \( x \).

- For each \( x \) of size \( n \), the number of \( r \)’s bad for \( x \) is at most \( 2^{p(n)}/2^{n+1} \).
- The number of \( r \)’s bad for some \( x \) of size \( n \) is at most \( 2^n 2^{p(n)}/2^{n+1} = 2^{p(n)}/2 \).
- There must be some \( r_n \) that is good for every \( x \) of size \( n \).

We may construct a P-time TM \( M \) with advice \( \{r_n\}_{n \in \mathbb{N}} \).
Theorem. \( \text{BPP} \subseteq \Sigma_2^p \cap \Pi_2^p \).

Sipser proved \( \text{BPP} \subseteq \Sigma_4^p \cap \Pi_4^p \). Gács pointed out that \( \text{BPP} \subseteq \Sigma_2^p \cap \Pi_2^p \). This is reported in Sipser’s paper. Lautemann provided a simplified proof using probabilistic method.

Lautemann’s Proof

Suppose $L \in \text{BPP}$. There is a polynomial $p$ and a P-time TM $\mathbb{M}$ such that for all $x \in \{0, 1\}^n$,

\[
\Pr_{r \in \{0, 1\}^p(n)}[\mathbb{M}(x, r) = 1] \geq 1 - 2^{-n}, \text{ if } x \in L,
\]

\[
\Pr_{r \in \{0, 1\}^p(n)}[\mathbb{M}(x, r) = 1] \leq 2^{-n}, \text{ if } x \notin L.
\]

Let $S_x$ be the set of $r$’s such that $\mathbb{M}(x, r) = 1$. Then

\[
|S_x| \geq (1 - 2^{-n})2^p(n), \text{ if } x \in L,
\]

\[
|S_x| \leq 2^{-n}2^p(n), \text{ if } x \notin L.
\]

For a set $S \subseteq \{0, 1\}^p(n)$ and string $u \in \{0, 1\}^p(n)$, let $S + u$ be $\{r + u \mid r \in S\}$, where $+$ is the bitwise exclusive $\lor$. 

Computational Complexity, by Fu Yuxi Randomized Computation 53 / 109
Lautemann’s Proof

Let \( k = \lceil \frac{p(n)}{n} \rceil + 1 \).

---

**Claim 1.** For every set \( S \subseteq \{0, 1\}^{p(n)} \) such that \( |S| \leq 2^{-n}2^{p(n)} \) and every \( k \) vectors \( u_1, \ldots, u_k \), one has \( \bigcup_{i=1}^{k} (S + u_i) \neq \{0, 1\}^{p(n)} \).

---

**Claim 2.** For every set \( S \subseteq \{0, 1\}^{p(n)} \) such that \( |S| \geq (1 - 2^{-n})2^{p(n)} \) there exist \( u_1, \ldots, u_k \) rendering \( \bigcup_{i=1}^{k} (S + u_i) = \{0, 1\}^{p(n)} \).

---

**Proof.**

Fix \( r \in \{0, 1\}^{p(n)} \). Now \( \Pr_{u_i \in \{0, 1\}^{p(n)}} [u_i \in S + r] \geq 1 - 2^{-n} \).

So \( \Pr_{u_1, \ldots, u_k \in \{0, 1\}^{p(n)}} \left[ \land_{i=1}^{k} u_i \notin S + r \right] \leq 2^{-kn} < 2^{-p(n)} \).

Notice that \( u_i \notin S + r \) if and only if \( r \notin S + u_i \), we get by union bound that

\[
\Pr_{u_1, \ldots, u_k \in \{0, 1\}^{p(n)}} \left[ \exists r \in \{0, 1\}^{p(n)}. r \notin \bigcup_{i=1}^{k} (S + u_i) \right] < 1. \]
Lautemann’s Proof

Now Claim 1 and Claim 2 imply that $x \in L$ if and only if

$$\exists u_1, \ldots, u_k \in \{0, 1\}^{p(n)}. \forall r \in \{0, 1\}^{p(n)}. r \in \bigcup_{i=1}^{k} (S_x + u_i),$$

or equivalently

$$\exists u_1, \ldots, u_k \in \{0, 1\}^{p(n)}. \forall r \in \{0, 1\}^{p(n)}. \bigvee_{i=1}^{k} M(x, r + u_i) = 1.$$  

Since $k$ is polynomial in $n$, we may conclude that $L \in \sum_p^2$. 

BPP is Low for Itself

Lemma. $BPP^{BPP} = BPP$. 
Complete Problem for $\text{BPP}$?

$\text{PP}$ is a syntactical class in the sense that every P-time PTM decides a language in $\text{PP}$.

$\text{BPP}$ is a semantic class. It is undecidable to check if a PTM both accepts and rejects with probability $2/3$.

1. We are unable to prove that PTMSAT is $\text{BPP}$-complete.
2. We are unable to construct universal machines. Consequently we are unable to prove any hierarchy theorem.

But if $\text{BPP} = \text{P}$, there should be complete problems for $\text{BPP}$.
ZPP
If P-time probabilistic decidable problems are defined using average complexity measure with bound on error probability, we get a complexity class that is even closer to \( \mathbf{P} \).
Suppose $T : \mathbb{N} \to \mathbb{N}$ and $L \subseteq \{0, 1\}^*$.

A PTM $\mathbb{P}$ with zero-sided error decides $L$ in time $T(n)$ if for every $x \in \{0, 1\}^*$, the expected running time of $\mathbb{P}(x)$ is at most $T(|x|)$, and it outputs $L(x)$ if $\mathbb{P}(x)$ halts.

$L \in \text{ZTIME}(T(n))$ if there is some $c$ such that $L$ is decided by some zero-sided error PTM in $cT(n)$ average time.
\[ ZPP = \bigcup_{c \in \mathbb{N}} ZTIME(n^c). \]
**Lemma.** $L \in \text{ZPP}$ if and only if there exists a P-time PTM $\mathcal{P}$ with outputs in \{0, 1, ?\} such that, for every $x \in \{0, 1\}^*$ and for all choices, $\mathcal{P}(x)$ outputs either $L(x)$ or ?, and $\Pr[\mathcal{P}(x) =?] \leq 1/3$.

If a PTM $\mathcal{P}$ answers in $O(n^c)$ time ‘dont-know’ with probability at most 1/3, then we can design a zero sided error PTM that simply runs $\mathcal{P}$ repetitively until it gets a proper answer. The expected running time of the new PTM is also $O(n^c)$.

Given a zero sided error PTM $\mathcal{P}$ with expected running time $T(n)$, we can design a PTM that answers ‘?’ if a sequence of $3T(n)$ choices have not led to a proper answer. By Markov’s inequality, this machines answers ‘?’ with a probability no more than 1/3.
PTM with One Sided Error

Suppose $T : \mathbb{N} \rightarrow \mathbb{N}$ and $L \subseteq \{0, 1\}^*$. A PTM $\mathbb{P}$ with one-sided error decides $L$ in time $T(n)$ if for every $x \in \{0, 1\}^*$, $\mathbb{P}$ halts in $T(|x|)$ steps, and

\[
\Pr[\mathbb{P}(x) = 1] \geq \frac{2}{3}, \text{ if } x \in L,
\]

\[
\Pr[\mathbb{P}(x) = 1] = 0, \text{ if } x \notin L.
\]

$L \in \text{RTIME}(T(n))$ if there is some $c$ such that $L$ is decided in $cT(n)$ time by some PTM with one-sided error.
\[ \text{RP} = \bigcup_{c \in \mathbb{N}} \text{RTIME}(n^c). \]
Theorem. \( \text{ZPP} = \text{RP} \cap \text{coRP}. \)

A ‘?’ answer can be replaced by a yes/no answer consistently.
Theorem. \( \text{ZPP}(1 - 1/n^c) = \text{ZPP}(2^{-n^d}) \) for all \( c, d > 1 \).

Suppose \( L \in \text{ZPP}(1 - 1/n^c) \) is decided by a PTM \( \mathbb{P} \) with a “don’t know” probability \( 1 - 1/n^c \) in expected running time \( T(n) \).

Let \( \mathbb{P}' \) be the PTM that on input \( x \) of size \( n \), repeat \( \mathbb{P} \) a total of \( \ln(2)n^{c+d} \) times. The “don’t know” probability of \( \mathbb{P}' \) is

\[
(1 - 1/n^c)\ln(2)n^{c+d} < e^{-\ln(2)n^d} = 2^{-n^d}.
\]

The running time of \( \mathbb{P}' \) on \( x \) is bounded by \( \ln(2)n^{c+d}T(n) \).
Error Reduction for $\text{RP}$

**Theorem.** $\text{RP}(1 - 1/n^c) = \text{RP}(2^{-n^d})$ for all $c, d > 1$. 
Random Walk and **RL**
Randomized Logspace Complexity

$L \in \mathbf{BPL}$ if there is a logspace PTM $\mathbf{P}$ such that $\Pr[\mathbf{P}(x) = L(x)] \geq \frac{2}{3}$.

**Fact.** $\mathbf{BPL} \subseteq \mathbf{P}$.

**Proof.**
Upon receiving an input the algorithm produces the adjacent matrix $\mathcal{A}$ of the configuration graph, in which $a_{ij} \in \{0, \frac{1}{2}, 1\}$ indicates the probability $C_i$ reaches $C_j$ in $\leq$ one step. It then computes $\mathcal{A}^{n-1}$. 

\[\square\]
Randomized Logspace Complexity

\[ L \in \text{RL} \text{ if } x \in L \text{ implies } \Pr[\mathbb{P}(x)=1] \geq \frac{2}{3} \text{ and } x \notin L \text{ implies } \Pr[\mathbb{P}(x)=1] = 0 \text{ for some logspace PTM } \mathbb{P}. \]

**Fact.** \( \text{RL} \subseteq \text{NL}. \)
Undirected Path Problem

Let UPATH be the reachability problem of undirected graph. Is UPATH in L?
Theorem. \( \text{UPATH} \in \text{RL} \).

To prove the theorem we need preliminary properties about Markov chains.

Markov chains were introduced by Andreĭ Andreevich Markov (1856-1922).
A stochastic process \( X = \{X_t \mid t \in T\} \) is a set of random variables taking values in a single state space \( \Omega \).

- If \( T \) is countably infinite, \( X \) is a **discrete time** process.
- If \( \Omega \) is countably infinite, \( X \) is a **discrete space** process.
- If \( \Omega \) is finite, \( X \) is a **finite** process.

A discrete space is often identified to \( \{0, 1, 2, \ldots\} \) and a finite space to \( \{0, 1, 2, \ldots, n\} \).
In the discrete time case a stochastic process starts with a state distribution $X_0$. It becomes another distribution $X_1$ on the states in the next step, and so on. In the $t$-th step $X_t$ may depend on all the histories $X_0, \ldots, X_{t-1}$. 
Markov Chain

A discrete time, discrete space stochastic process $X_0, X_1, X_2, \ldots$, is a Markov chain if

$$\Pr[X_t = a_t \mid X_{t-1} = a_{t-1}] = \Pr[X_t = a_t \mid X_{t-1} = a_{t-1}, \ldots, X_0 = a_0].$$

The dependency on the past is captured by the value of $X_{t-1}$. This is the Markov property.

A Markov chain is time homogeneous if for all $t \geq 1$,

$$\Pr[X_{t+1} = j \mid X_t = i] = \Pr[X_t = j \mid X_{t-1} = i].$$

These are the Markov chains we are interested in. We write $M_{j,i}$ for $\Pr[X_{t+1} = j \mid X_t = i]$. 
Transition Matrix

The transition matrix $M$ is $(M_{j,i})_{j,i}$ such that $\sum_j M_{j,i} = 1$ for all $i$. For example

$$M = \begin{pmatrix}
0 & 1/2 & 1/2 & 0 & \ldots \\
1/4 & 0 & 1/3 & 1/2 & \ldots \\
0 & 1/3 & 1/9 & 1/4 & \ldots \\
1/2 & 1/6 & 0 & 1/8 & \ldots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}$$
Transition Graph
Let $m_t$ denote a probability distribution on the state space at time $t$. Then

$$m_{t+1} = M \cdot m_t.$$  

The $t$ step transition matrix is clearly given by

$$M^t.$$
Irreducibility

A state $j$ is accessible from state $i$ if $(M^n)_{j,i} > 0$ for some $n \geq 0$. If $i$ and $j$ are accessible from each other, they communicate.

A Markov chain is irreducible if all states belong to one communication class.
A period of a state $i$ is the greatest common divisor of $T_i = \{ t \geq 1 \mid (M^t)_{i,i} > 0 \}$.

A state $i$ is aperiodic if $\gcd T_i = 1$.

---

**Lemma.** If $M$ is irreducible, then $\gcd T_i = \gcd T_j$ for all states $i, j$.

**Proof.**

By irreducibility $(M^s)_{j,i} > 0$ and $(M^t)_{i,j} > 0$ for some $s, t > 0$. Clearly $T_i + (s + t) \subseteq T_j$. It follows that $\gcd T_i \geq \gcd T_j$. Symmetrically $\gcd T_j \geq \gcd T_i$. □

---

The period of an irreducible Markov chain is the period of the states.
Classification of State

Let \( r_{j,i}^t \) denote the probability that, starting at \( i \), the first transition to \( j \) occurs at time \( t \); that is

\[
 r_{j,i}^t = \Pr[X_t = j \land \forall s \in [t-1].X_s \neq j \mid X_0 = i].
\]

A state \( i \) is recurrent if

\[
 \sum_{t \geq 1} r_{i,i}^t = 1.
\]

A state \( i \) is transient if

\[
 \sum_{t \geq 1} r_{i,i}^t < 1.
\]

A recurrent state \( i \) is absorbing if

\[
 M_{i,i} = 1.
\]
If one state in an irreducible Markov chain is recurrent, respectively transient, all states in the chain are recurrent, respectively transient.
Ergodic State

The expected hitting time to $j$ from $i$ is

$$h_{j,i} = \sum_{t \geq 1} t \cdot r_{j,i}^t.$$

A recurrent state $i$ is positive recurrent if the expected first return time $h_{i,i} < \infty$.

A recurrent state $i$ is null recurrent if $h_{i,i} = \infty$.

An aperiodic, positive recurrent state is ergodic.
For the presence of null recursive state, the number of states must be infinite.
A Markov chain $M$ is **recurrent** if every state in $M$ is recurrent.

A Markov chain $M$ is **aperiodic** if the period of $M$ is 1.

A Markov chain $M$ is **ergodic** if all states in $M$ are ergodic.

A Markov chain $M$ is **regular** if $\exists r > 0. \forall i, j. M^r_{j,i} > 0$.

A Markov chain $M$ is **absorbing** if there is at least one absorbing state and from every state it is possible to go to an absorbing state.
The Gambler’s Ruin

A fair gambling game between Player I and Player II.

- In each round a player wins/loses with probability 1/2.
- The state at time $t$ is the number of dollars won by Player I. Initially the state is 0.
- Player I can afford to lose $\ell_1$ dollars, Player II $\ell_2$ dollars.
- The states $-\ell_1$ and $\ell_2$ are absorbing. The state $i$ is transient if $-\ell_1 < i < \ell_2$.
- Let $M^t_i$ be the probability that the chain is in state $i$ after $t$ steps.
- Clearly $\lim_{t \to \infty} M^t_i = 0$ if $-\ell_1 < i < \ell_2$.
- Let $q$ be the probability the game ends in state $\ell_2$. By definition $\lim_{t \to \infty} M^t_{\ell_2} = q$.
- Let $W^t$ be the gain of Player I at step $t$. Then $E[W^t] = 0$ since the game is fair.

Now $E[W^t] = \sum_{i=-\ell_1}^{\ell_2} i M^t_i = 0$ and $\lim_{t \to \infty} E[W^t] = \ell_2 q - \ell_1 (1 - q) = 0$.

Conclude that $q = \frac{\ell_1}{\ell_1 + \ell_2}$. 
In the rest of the lecture we confine our attention to finite Markov chains.
Lemma. In a finite Markov chain, at least one state is recurrent; and all recurrent states are positive recurrent.

In a finite Markov chain $\mathbf{M}$ there must be a communication class without any outgoing edges. Starting from any state $k$ in the class the probability that the chain will return to $k$ in $d$ steps is at least $p$ for some $p > 0$, where $d$ is the diameter of the class. The probability that the chain never returns to $k$ is $\lim_{t \to \infty} (1 - p)^{dt} = 0$. Hence $\sum_{t \geq 1} M_{k,k}^t = 1$.

Starting from a recurrent state $i$, the probability that the chain returns to $i$ in $dt$ steps is at most $q$ for some $q \in (0, 1)$. Thus $\sum_{t \geq 1} t r_{i,i}^t$ is bounded by $\sum_{t \geq 1} dt q^{dt} < \infty$.

Corollary. In a finite irreducible Markov chain, all states are positive recurrent.
Proposition. Suppose $M$ is a finite irreducible Markov chain. The following are equivalent:

(i) $M$ is aperiodic.  
(ii) $M$ is ergodic.  
(iii) $M$ is regular.

(i$\iff$ii) This is a consequence of the previous corollary.

(i$\Rightarrow$iii) Assume $\forall i. \gcd T_i = 1$. Since $T_i$ is closed under addition, Fact implies that some $t_i$ exists such that $t \in T_i$ whenever $t \geq t_i$. By irreducibility for every $j$, $(M^{t_j,i})_{j,i} > 0$ for some $t_j,i$.

Set $t = \prod_i t_i \cdot \prod_{i \neq j} t_{j,i}$. Then $(M^t)_{i,j} > 0$ for all $i,j$.

(iii$\Rightarrow$i) If $M$ has period $t > 1$, for any $k > 1$ some entries in the diagonal of $M^{kt-1}$ are 0.

Fact. If a set of natural number is closed under addition and has greatest common divisor 1, then it contains all but finitely many natural numbers.
The graph of a finite Markov chain contains two types of maximal strongly connected components (MSCC).

- Recurrent MSCC’s that have no outgoing edges. There is at least one such MSCC.
- Transient MSCC’s that have at least one outgoing edge.

If we think of an MSCC as a big node, the graph is a dag.

How fast does the chain leave the transient states? What is the limit behaviour of the chain on the recurrent states?
Let $Q$ be the matrix for the transient states, $E$ for the recurrent states, assuming that the graph has only one recurrent MSCC. We shall assume that $E$ is ergodic.

\[
\begin{pmatrix}
Q & 0 \\
L & E
\end{pmatrix}
\]

It is clear that

\[
\begin{pmatrix}
Q & 0 \\
L & E
\end{pmatrix}^n = \begin{pmatrix} Q^n & 0 \\ L' & E^n \end{pmatrix}.
\]

Limit Theorem for Transient Chain. $\lim_{n \to \infty} Q^n = 0$. 
**Theorem.** \( N = \sum_{n \geq 0} Q^n \) is the inverse of \( I - Q \). The entry \( N_{j,i} \) is the expected number of visits to \( j \) starting from \( i \).

\( I - Q \) is nonsingular because \( x(I - Q) = 0 \) implies \( x = 0 \). Then \( N(I - Q^{n+1}) = \sum_{i=0}^{n} Q^i \) follows from \( N(I - Q) = I \). Thus \( N = \sum_{i=0}^{\infty} Q^n \).

Let \( X_k \) be the Poisson trial with \( \Pr[X_k = 1] = (Q^k)_{j,i} \), the probability that starting from \( i \) the chain visits \( j \) at the \( k \)-th step. Let \( X = \sum_{k=1}^{\infty} X_k \). Clearly \( \mathbb{E}[X] = N_{j,i} \). Notice that \( N_{i,i} \) counts the visit at the 0-th step.
Theorem. \( \sum_j N_{j,i} \) is the expected number of steps to stay in transient states after starting from \( i \).

\[ \sum_j N_{j,i} \] is the expected number of visits to any transient states after starting from \( i \). This is precisely the expected number of steps.
A stationary distribution of a Markov chain $M$ is a distribution $\pi$ such that

$$\pi = M\pi.$$ 

If the Markov chain is finite, then $\pi = \begin{pmatrix} \pi_0 \\ \pi_1 \\ \vdots \\ \pi_n \end{pmatrix}$ satisfies

$$\sum_{j=0}^{n} M_{i,j} \pi_j = \pi_i = \sum_{j=0}^{n} M_{j,i} \pi_i.$$ 

[probability entering $i =$ probability leaving $i]$
Theorem. The power $E^n$ approaches to a limit as $n \to \infty$. Suppose $W = \lim_{n \to \infty} E^n$. Then $W = (\pi, \pi, \ldots, \pi)$ for some positive $\pi$. Moreover $\pi$ is a stationary distribution of $E$.

We may assume that $E > 0$. Let $r$ be a row of $E$, and let $\Delta(r) = \max r - \min r$.

- It is easily seen that $\Delta(rE) < (1 - 2p)\Delta(r)$, where $p$ is the minimal entry in $E$.
- It follows that $\lim_{n \to \infty} E^n = W = (\pi, \pi, \ldots, \pi)$ for some distribution $\pi$.
- $\pi$ is positive since $rE$ is already positive.

Moreover $W = \lim_{n \to \infty} E^n = E \lim_{n \to \infty} E^n = EW$. That is $\pi = E\pi$. 

Computational Complexity, by Fu Yuxi Randomized Computation 96 / 109
**Lemma.** \( \mathbf{E} \) has a unique stationary distribution. [\( \pi \) can be calculated by solving linear equations.]

Suppose \( \pi, \pi' \) are stationary distributions. Let

\[
\frac{\pi_i}{\pi'_i} = \min_{0 \leq k \leq n} \left\{ \frac{\pi_k}{\pi'_k} \right\}.
\]

It follows from the regularity property that \( \pi_i / \pi'_i = \pi_j / \pi'_j \) for all \( j \in \{0, \ldots, n\} \).
Limit Theorem for Ergodic Chains

**Theorem.** \( \pi = \lim_{n \to \infty} E^n v \) for every distribution \( v \).

Suppose \( E = (m_0, \ldots, m_k) \). Then \( E^{n+1} = (E^n m_0, \ldots, E^n m_k) \). It follows from

\[
\left( \lim_{n \to \infty} E^n m_0, \ldots, \lim_{n \to \infty} E^n m_k \right) = \lim_{n \to \infty} E^{n+1} = (\pi, \ldots, \pi)
\]

that \( \lim_{n \to \infty} E^n m_0 = \ldots = \lim_{n \to \infty} E^n m_k = \pi \). Now

\[
\lim_{n \to \infty} E^n v = \lim_{n \to \infty} E^n (v_0 m_0 + \ldots + v_k m_k) = v_0 \pi + \ldots + v_k \pi = \pi.
\]
**Limit Theorem for Ergodic Chains**

**H** is the hitting time matrix whose entries at \((j, i)\) is \(h_{j,i}\).

**D** is the diagonal matrix whose entry at \((i, i)\) is \(h_{i,i}\).

**J** is the matrix whose entries are all 1.

---

**Lemma.** \(H = J + (H - D)E\).

**Proof.**

For \(i \neq j\), the hitting time is \(h_{j,i} = E_{j,i} + \sum_{k \neq j} E_{k,i}(h_{j,k} + 1) = 1 + \sum_{k \neq j} E_{k,i}h_{j,k}\), and the first recurrence time is \(h_{i,i} = E_{i,i} + \sum_{k \neq i} E_{k,i}(h_{i,k} + 1) = 1 + \sum_{k \neq i} E_{k,i}h_{i,k}\). \(\square\)

---

**Theorem.** \(h_{i,i} = 1/\pi_i\) for all \(i\).

**Proof.**

\[1 = J\pi = H\pi - (H - D)E\pi = H\pi - (H - D)\pi = D\pi.\] \(\square\)
Queue

Let $X_t$ be the number of customers in the queue at time $t$. At each time step exactly one of the following happens.

- If $|\text{queue}| < n$, with probability $\lambda$ a new customer joins the queue.
- If $|\text{queue}| > 0$, with probability $\mu$ the head leaves the queue after service.
- The queue is unchanged with probability $1 - \lambda - \mu$.

The finite Markov chain is ergodic. Therefore it has a unique stationary distribution.

\[
\begin{pmatrix}
1 - \lambda & \mu & 0 & \ldots & 0 & 0 & 0 \\
\lambda & 1 - \lambda - \mu & \mu & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 1 - \lambda - \mu & \mu \\
0 & 0 & 0 & \ldots & 0 & 1 - \mu & \mu
\end{pmatrix}
\]
Time Reversibility

A distribution $\pi$ for a finite Markov chain $M$ is **time reversible** if $M_{j,i}\pi_i = M_{i,j}\pi_j$.

**Lemma.** A time reversible distribution is stationary.

**Proof.**
\[
\sum_i M_{j,i}\pi_i = \sum_i M_{i,j}\pi_j = \pi_j.
\]

Suppose $\pi$ is a stationary distribution of a finite Markov chain $M$.
Consider $X_0, \ldots, X_n$, a finite run of the chain. We see the reverse sequence $X_n, \ldots, X_0$ as a Markov chain with transition matrix $R$ defined by $R_{i,j} = \frac{1}{\pi_j} M_{j,i}\pi_i$.

- If $M$ is time reversible, then $R = M$, hence the terminology.
Using the equality $EW = W$ and $W^k = W$, one proves $\lim_{n \to \infty} (E - W)^n = 0$ using

$$
(E - W)^n = \sum_{i=0}^{n} (-1)^i \binom{n}{i} E^{n-i} W^i = E^n + \sum_{i=1}^{n} (-1)^i \binom{n}{i} W = E^n - W.
$$

It follows from the above result that $x(I - E + W) = 0$ implies $x = 0$. So $(I - E + W)^{-1}$ exists.

Let $Z = (I - E + W)^{-1}$. This is the fundamental matrix of $E$. 
Lemma. (i) $1Z = 1$. (ii) $Z\pi = \pi$. (iii) $(I - E)Z = I - W$.

Proof. (i) is a consequence of $1E = 1$ and $1W = 1$.
(ii) is a consequence of $E\pi = \pi$ and $W\pi = \pi$.

Theorem. $h_{j,i} = (z_{j,j} - z_{j,i})/\pi_j$. [This equality can be used to calculate $h_{j,i}$.


$$H - D = J - DZ + (H - D)W.$$ 

For $i \neq j$ one has $h_{j,i} = 1 - z_{j,i}h_{j,j} + ((H - D)\pi)_j$. Also $0 = 1 - z_{j,j}h_{j,j} + ((H - D)\pi)_j$. Hence $h_{j,i} = (z_{j,j} - z_{j,i})h_{j,j} = (z_{j,j} - z_{j,i})/\pi_j$. 

□
Theorem. A finite irreducible Markov chain has a unique stationary distribution.

Proof. 
$(I + M)/2$ is regular because it is aperiodic. If $\pi$ is a stationary distribution of $(I + M)/2$, it is a stationary distribution of $M$, and vice versa. Hence the uniqueness.

The stationary distribution $\pi$ is no longer a stable distribution. But $\pi_i$ can still be interpreted as the frequency of the occurrence of state $i$. 

A random walk on an undirected graph $G$ is the Markov chain whose transition matrix $A$ is the normalized adjacent matrix of $G$.

**Lemma.** A random walk on an undirected connected graph $G$ is aperiodic if and only if $G$ is not bipartite.

**Proof.**

$(\Rightarrow)$ If $G$ is bipartite, the period of $G$ is 2.

$(\Leftarrow)$ If one node has a cycle of odd length, every node has a cycle of length $2k + 1$ for all large $k$. So the gcd must be 1. [In an undirected graph every node has a cycle of length 2.]

**Fact.** A graph is bipartite if and only if it has only cycles of even length.
Random Walk on Undirected Graph

**Theorem.** A random walk on $G = (V, E)$ converges to the stationary distribution

$$
\pi = \begin{pmatrix}
d_0 \\
2|E| \\
\vdots \\
d_n \\
2|E|
\end{pmatrix}.
$$

**Proof.**
The degree of vertex $i$ is $d_i$. Clearly $\sum_v \frac{d_v}{2|E|} = 1$ and $A\pi = \pi$.

**Lemma.** If $(u, v) \in E$ then $h_{u,v} < 2|E|$.

**Proof.**
Omitting possible self-loops, $2|E|/d_u = h_{u,u} \geq \sum_{v \neq u} (1 + h_{u,v})/d_u$. Hence $h_{u,v} < 2|E|$.
Random Walk on Undirected Graph

The cover time of $G = (V, E)$ is the maximum over all vertices $v$ of the expected time to visit all nodes in the graph $G$ by a random walk from $v$.

Lemma. The cover time of $G = (V, E)$ is bounded by $4|V||E|$.

Proof.
Fix a spanning tree of the graph. A depth first walk along the edges of the tree is a cycle of length $2(|V| - 1)$. The cover time is bounded by

$$
\sum_{i=1}^{2|V|-2} h_{v_i, v_{i+1}} < (2|V| - 2)(2|E|) < 4|V||E|.
$$

\Box
An **RL** algorithm for **UPATH** can now be designed. Let \(((V, E), s, t)\) be the input.

1. Starting from \(s\), walk randomly for \(12|V||E|\) steps;

2. If \(t\) has been hit, answer ‘yes’, otherwise answer ‘no’.

Add self loops if \(G\) is bipartite.

By Markov inequality the error probability is less than \(\frac{1}{3}\).
BPP $\equiv$ P