Journal of Algorithms and Computation

journal homepage: http://jac.ut.ac.ir

On some main parameters of stochastic processes on directed graphs

Mehdi Shams *[1] and Gholamreza Hesamian†[2]

[1]Department of Mathematical Sciences, University of Kashan, Kashan, Iran
[2]Department of Statistics, Payame Noor University, Tehran, 19395-3697, Iran

ABSTRACT

A random walk is a special kind of stochastic process of the Markov chain type. Some stochastic processes can be represented as a random walk on a graph. In this paper, the main parameters for a random walk on graph are examined.

ARTICLE INFO

Article history:
Research paper
Received 03, January 2023
Received in revised form 11, April 2023
Accepted 15 May 2023
Available online 01, June 2023

Keyword: Markov chain, Martingale, connected graph, stationary distribution, first hit time, cover time

AMS subject Classification: 05C81, 94C15, 60Gxx

*Corresponding author: M. Shams . Email: mehdishams@kashanu.ac.ir
†gh.hesamian@pnu.ac.ir

1 Introduction

Graph theory is an important branch of mathematics that is also used in probability and stochastic processes. In a graph we start moving from a point and select a neighbor from that point at random and go to that neighborhood from that point and then randomly select a neighbor from that point.

One of the most popular stochastic process is random walk. A graph in this context is made up of vertices which are connected by edges. A random sequence of selected points is a random walk on a graph. If the edges are weighted, all Markov chains can be shown as a random walk on the directed graph. Similarly, reversible Markov chains can be represented as a random walk on a undirected graph and symmetric Markov chains can be represented as a random walk on a regular symmetric graph. Random walk is included in many mathematical and physics models. Brownian motion of dust in a room is also a random walk. In some cases, statistical mechanics models also involve random walk.

Polya [12] proves that in a random walk on d-dimensional grid, if $d = 2$, it returns to its starting point for infinity, and for a finite number of motions. See [5, 15] for more results on random walk on an infinite graph. Recently, random walk has become more common, but finite graphs have received more attention and their qualitative aspects have been further studied. For example, how long should we walk before returning to the starting point? Before we see a given node? Before we see all nodes? How fast does the random walk distribution tend to its final distribution? Here are some general references to random walk and finite Markov chains: [1, 4, 5].

Some stochastic processes can be represented as a random walk on a graph. In this paper, the main parameters for a random walk on graph are examined.

2 Basic notions

A graph is a set of objects called vertices (denoted V and also called nodes) along with a set of unordered pairs of vertices called edges (denoted E and also called links or lines).

Let $G = (V, E)$ be a connected graph with $|V| = n$ nodes and $|E| = m$ edges. A graph is said to be connected, if for any two points $i, j \in V$ there is always a path joining them. A directed graph (oriented graph or digraph) is a graph in which edges have orientations [10]. In the following we will consider only directed graphs. The degree of a vertex i in a graph is the number $d(i)$ of edges [1].

On an graph $G = (V, E)$ a graph automorphism is a one-to-one (injective) map $f : V \to V$ s.t.

$$(i, j) \in E \iff (f(i), f((j)) \in E.$$

The graph G is vertex-transitive if its group of automorphisms i.e. $Aut(G)$ (or equivalently the graph automorphism) acts transitively on its vertex set V. In other words, given any vertices $i, j \in V$, there is an $f \in Aut(G)$ s.t. $f(i) = j$. This simply means that all vertices look the same locally, i.e. we cannot uniquely identify any vertex based on the edges and vertices around it. Clearly, random walk on a (unweighted) graph is a symmetric reversible chain iff the graph is vertex-transitive[1]. For example a Cayley graph is vertex-transitive [11]. The distance $d(i, j)$ between any two vertices i, j of G is the length of a shortest path between i and j in G. A distance-transitive graph is a graph for which if we are given two pairs of vertices i_1, j_1, i_2, j_2 with $d(i_1, j_1) = d(i_2, j_2)$, then we can always find an automorphism f of the graph s.t. $f(i_1) = i_2$ and $f(j_1) = j_2$. Distance-regular graphs are natural combinatorial generalizations of distance-transitive graphs [13].

An graph G is bipartite if we can divide V into two disjoint sets U and W s.t. no edge in E connects two vertices from U or two vertices from W. Equivalently, all edges connect a vertex in U to a vertex in W [10].

A random walk, in probability theory, is the stochastic process by which randomly-moving objects wander away from where they started. Random walks are an example of Markov processes, in which future behaviour is independent of past history. Stock market fluctuations, at least over the short run, are random walks [2].

Consider a random walk on G: we start at a node v_0 if at the t^{th} step we are at a node v_t, we move neighbor of v_t with probability $1/d(v_t)$. Clearly, the sequence of random nodes $\{v_t\}_{t\in\mathbb{N}\cup\{0\}}$ or equivalently the random sequence $\{X_t\}_{t\in\mathbb{N}\cup\{0\}}$ is a discrete-time stochastic process defined on the state space V where X_t denote the location of the random walk at time t.

The node v_0 may be fixed, but may itself be drawn from some initial distribution π_0. We denote by π_t the distribution of v_t s.t. $\pi_t(i) = \mathrm{P}(X_t = i)$. We denote by $P = (p_{ij})_{i,j\in V}$ the matrix of transition probabilities of this Markov chain, so

$$p_{ij} = P(X_{k+1} = j|X_k = i) = \begin{cases} \frac{1}{d(i)} & i, j \in V \\ 0 & o.w. \end{cases}$$

These transition probabilities do not depend on "time" k.

The Markov property holds: conditional on the present, the future is independent of the past:

$$P(X_{k+1} = j|X_k = i, X_{k-1} = i_{k-1}, ...X_0 = i_0) = P(X_{k+1} = j|X_k = i) = p_{ij}.$$

Therefore the random sequence of vertices visited by the walk, $\{X_t\}_{t\in\mathbb{N}\cup\{0\}}$ is a Markov chain with state space V and matrix of transition probabilities $P = (p_{ij})_{i,j\in V}$. Note that P is a stochastic matrix, i.e. $\sum_{j\in V} p_{ij} = 1$ [2].

Similarly for lazy random walk, transition probabilities are as follows:

$$p_{ij} = P(X_{k+1} = j|X_k = i) = \begin{cases} \frac{1}{2d(i)} & (i, j) \in E \\ \frac{1}{2} & i = j \end{cases}$$

(At each step, the lazy random walk will do the following: with probability $\frac{1}{2}$ stay at the current vertex and with probability $\frac{1}{2}$ take a usual random step). The graph topology can be algebraically represented introducing its adjacency matrix $A_G = (A_{ij})$ given by:

$$A_{ij} = \begin{cases} 1 & (i, j) \in E \\ 0 & o.w. \end{cases}$$

When an undirected edge connects i and j, we say that i and j are adjacent, or that i and j are neighbors. We denote adjacency as $i \sim j$. In directed graph, an edge connecting nodes $i \in V$ and $j \in V$ will be denoted as (i, j).
Assuming the time (t) to be discrete, we define at each time step t the jumping probability p_{ij} between nearest neighbour sites i and j:

$$p_{ij} = \frac{A_{ij}}{\sum_j A_{ij}} = \begin{cases} \dfrac{1}{\sum_j A_{ij}} = \dfrac{1}{d(i)} & (i,j) \in E \\ 0 & o.w. \end{cases}$$

This is the simplest case we can consider: the jumping probabilities are isotropic at each point and they do not depend on time; in addition the walker is forced to jump at every time step.
Let $D = (D_{ij})$ denote the diagonal matrix with $D_{ii} = 1/d(i)$ and $D_{ij} = 0$, for all $i \neq j$, then $P = DA_G$. If G is d-regular, then $P = \frac{1}{d}A_G$ [9]. Also we have $\pi_{t+1} = \pi_t P$ and thus $\pi_t = \pi_0 P^t$. It follows that the probability p_{ij}^t that, starting at i, we reach j in t steps is given by the ij-entry of the matrix P^t. If G is regular, then this Markov chain is symmetric, the probability of moving to i given that we are at node j is the same as the probability of moving to node j given that we are at node i. For a non-regular graph G, this property is replaced by time-reversibility (see relation (1)).
The probability distributions $\pi_0, \pi_1, ...$ are of course different in general. We say that the distribution π_0 is stationary (steady-state or invariant) for the graph G, if $\pi_1 = \pi_0$. In this case, $\pi_t = \pi_0$ for all $t \geq 0$. We call this random walk the stationary random walk [2]. It is easy to show that for every graph G, the distribution $\pi(i) = \frac{d(i)}{2m}$ is stationary. Hence a connected non-bipartite undirected graph has a stationary distribution proportional to the degree distribution. In particular, the uniform distribution on V is stationary if the graph is regular (see (2)). It is not difficult to show that the stationary distribution on connected graph is unique. In terms of the stationary distribution, it is easy to formulate the property of time-reversibility, it is equivalent to saying that for every pair $i, j \in V$,

$$\pi(i)p_{ij} = \pi(j)p_{ji}. \tag{1}$$

This means that in a stationary random walk, we step as often from i to j as from j to i.
Recall P denotes the transition probability matrix of a finite irreducible (there is only one equivalence class) discrete-time stochastic process $\{X_t\}_{t \in \mathbb{N} \cup \{0\}}$ and a vector $\boldsymbol{\pi} = (\pi_j)_{j \in \mathbb{N} \cup \{0\}}$ denotes the stationary distribution where $\pi_j(i) = \mathrm{P}(X_j = i)$. The discrete-time stochastic process is said to be reversible, if the equation (1) is satisfied [16]. Equivalently, suppose (for given irreducible P) that $\boldsymbol{\pi}$ is a probability distribution satisfying (1). Then $\boldsymbol{\pi}$ is the unique stationary distribution and the discrete-time stochastic process is reversible. Certainly, this is true because (1), sometimes called the detailed balance equations, implies for all $i \in V$:

$$\pi(i) = \pi(i) \sum_{j \in V} p_{ij} = \sum_{j \in V} \pi(i)\, p_{ij} = \sum_{j \in V} \pi(j)\, p_{ji}$$

that is, $\boldsymbol{\pi} = \boldsymbol{\pi}P$ and so $\boldsymbol{\pi}$ is the unique stationary distribution and the discrete-time stochastic process is reversible. The name reversible comes from the following fact. If $\{X_t\}_{t\in\mathbb{N}\cup\{0\}}$ is the stationary discrete-time stochastic process, that is, if X_0 has distribution π, then $(X_0, X_1, ..., X_t) \stackrel{d}{=} (X_t, X_{t-1}, ..., X_0)$. It is elementary that the same symmetry property (1) holds for the t-step transition matrix P^t: $\pi(i)p_{ij}^t = \pi(j)p_{ji}^t$. For a random walk on G, it is straightforward to find a probability vector satisfying the detailed balance condition, for every pair $i, j \in V$, $\frac{\pi(i)}{d(i)} = \frac{\pi(j)}{d(j)} = k$. But $1 = \sum_{i\in V}\pi(i) = k\sum_{i\in V} d(i) = 2mk$ and hence, $k = \frac{1}{2m}$. In particular, if G is d-regular, since the number of edges (i.e. m) of a d-regular graph with n vertices is equal to $\frac{n\times d}{2}$, so for all $i \in V$:

$$\pi(i) = \frac{d}{2m} = \frac{1}{n} \tag{2}$$

and $\boldsymbol{\pi}$ is the uniform distribution [16].

From (1) we have $\pi(i)p_{ij} = \frac{1}{2m}$, for all $i, j \in V$, so we move along every edge, in every given direction, with the same frequency. If we are sitting on an edge and the random walk just passed through it, then the expected number of steps before it passes through it in the same direction again is $2m$. There is a similar fact for nodes, if we are sitting at a node i and the random walk just visited this node, then the expected number of steps before it returns is $\frac{1}{\pi(i)} = \frac{2m}{d(i)}$. If G is regular, then this return time is just $\frac{1}{\pi(i)} = \frac{2m}{d(i)} = n$ the number of nodes.

If the graph G is non-bipartite, then for any $i, j \in V$:

$$\lim_{k\to\infty} P(X_k = j|X_0 = i) = \lim_{k\to\infty} p_{ij}^k = \pi(j). \tag{3}$$

The convergence to $\pi(j)$ does not depend on the initial vertex i. Hence, by (3) we have:

$$\begin{aligned}
\lim_{k\to\infty} \pi_k(j) &= \lim_{k\to\infty} P(X_k = j) \\
&= \lim_{k\to\infty} \sum_{i\in V} P(X_k = j|X_0 = i)P(X_0 = i) \\
&= \sum_{i\in V} \lim_{k\to\infty} P(X_k = j|X_0 = i)P(X_0 = i) \\
&= \pi(j)\sum_{i\in V} \pi_0(i) = \pi(j)
\end{aligned}$$

that is, $\lim_{k\to\infty} \boldsymbol{\pi_k} = \boldsymbol{\pi}$ independently of the initial distribution. This is equivalent to $\lim_{k\to\infty} P^k = \Pi$ where Π is a stochastic matrix with all its rows equal to $\boldsymbol{\pi}$.

3 Main parameters

We now introduce the parameters of random walk discussion, which play the most important role in quantitative random walk.

In the study of stochastic processes in mathematics, a first hit time (access time or hitting time or discovery time) is the first time at which a given process "hits" a given subset of the state space. Exit times and return times (time a random walk takes to leave and return to one node) are also examples of hitting times. For a discrete-time stochastic process $\{X_t\}_{t\in\mathbb{N}\cup\{0\}}$, write $T_i = min\{t \geqslant 0 | X_t = i\}$ for the first hitting time on state i [1]. More generally, a subset A of states has first hitting time $T_A = min\{t \geqslant 0 | X_t \in A\}$ and last exit time $L_A = max\{t \geqslant 0 | X_t \in A\}$. The maximal mean hitting time $max_{i,j} E_i(T_j)$ arises in many contexts. The cover time for a n-state Markov chain is the random time C taken for the entire state-space to be visited. Formally, $C = max_j T_j$. It is sometimes mathematically nicer to work with the "cover-and-return" time $C^+ = min\{t \geqslant C | X_t = X_0\}$.

Definition 1 The first hit time $H_i(j)$ is the expected number of steps before node j is visited, starting from node i. That is, for two states $i, j \in V$, the hitting time of j from i is defined as $H_i(j) = E_i[T_j] = E[T_j | X_0 = i]$. Notice that $H_i(i) = 0$ for all $i \in V$.

The sum $\kappa_i(j) = H_i(j) + H_j(i)$ is called the commute time, this is the expected number of steps in a random walk starting at i, before node j is visited and then node i is reached again.

If vertices i and j are connected by an edge, i.e. $(i, j) \in E$, then

$$\kappa_i(j) = H_i(j) + H_j(i) \leq 2m. \tag{4}$$

To prove Equation (4), we do the following. The probability of traversing any edge is equally likely, i.e. $\frac{1}{2m}$. Therefore, the frequency of seeing a given edge in a random walk is $2m$. Based on the condition that the walk traversed edge (i, j), the expected time until the next traversal of edge (i, j) is $2m$. But since random walks are memoryless, we can drop the condition. The path taken by the random walk could have touched i and gone back to j multiple times before actually taking edge (i, j). Therefore, since the total path length is expected to be $2m$, the path may go to i and come back to j in less than $2m$ steps. Hence $H_i(j) + H_j(i) \leq 2m$. For example in circular graph, by symmetry, (4) concludes that $H_i(j) + H_j(i) = 2H_i(j) \leq 2m$ and so $H_i(j) \leq m$.
In a graph G on n vertices, $n \geq 13$, it is shown in [3] that for any three distinct vertices i, j, k we have $\kappa_i(j)+\kappa_j(k)+\kappa_k(i) \leq \frac{8}{27}n^3+\frac{8}{3}n^2+\frac{4}{9}n-\frac{592}{27}$. There is also a way to express first hit times in terms of commute times, due to [14]: $H_i(j) = \frac{1}{2}\big(\kappa_i(j) + \sum\limits_u \pi(u)[\kappa_u(j) - \kappa_u(i)]\big)$.
It is easy to show that

$$H_i(j) = \begin{cases} 1 + \sum_k p_{ik} H_k(j) & i \neq j \\ 0 & o.w. \end{cases} = \begin{cases} 1 + \dfrac{1}{d(i)} \sum_{k\in\Gamma(i)} H_k(j) & i \neq j \\ 0 & o.w. \end{cases}$$

where $\Gamma(i)$ is the set of neighbors of node i. This equation can be put together in matrix notation. Let $H = (H_{ij}) = (H_i(j))$ be a square matrix s.t. $H_{ii} = 0$. Then, $H + 2mD =$

$J + PH$. That is, $(I - P)H = J - 2mD$. We can not solve for H, because $I - P$ is singular. Let $Z = (I - P + \Pi)^{-1}$. It is easily checked that $H = J - 2mZD + \Pi H$. Hence,

$$H_i(j) = \begin{cases} 1 - \dfrac{2m}{d(j)} Z_{ij} + (\boldsymbol{\pi} H)_j & i \neq j \\ 1 - \dfrac{2m}{d(j)} Z_{jj} + (\boldsymbol{\pi} H)_j = 0 & o.w. \end{cases}$$

Thus, $1 + (\boldsymbol{\pi} H)_j = \frac{2m}{d(j)} Z_{jj}$ and we can compute the access times from the fundamental matrix Z: $H_i(j) = 2m \frac{Z_{jj} - Z_{ij}}{d(j)}$. Diagonalizing

$$N = D^{-1/2} P D^{1/2} = D^{-1/2} D A_G D^{1/2} = D^{1/2} A_G D^{1/2}$$

which has the same eigenvalues as P, $\lambda_1 \geq ... \geq \lambda_n$. Write N in spectral form $N = \sum_{r=1}^{n} \lambda_r \boldsymbol{v}_r^T \boldsymbol{v}_r$ where the row eigenvectors $\boldsymbol{v}_r$ are unitary and orthogonal. It is easily checked that $\boldsymbol{w} = (\sqrt{d(1)}, ..., \sqrt{d(n)})$ is a positive eigenvector of N with eigenvalue 1. So we get:

$$H_i(j) = 2m \sum_{k=2}^{n} \frac{1}{1 - \lambda_k} \Big(\frac{v_{kj}^2}{d(j)} - \frac{v_{ki} v_{kj}}{\sqrt{d(i) d(j)}} \Big). \tag{5}$$

Thus for stationary $\pi(v) = \frac{d(v)}{2m}$, by using (5) we have:

$$\begin{aligned} \sum_{j \in V} \pi(j) H_i(j) &= \sum_{j \in V} \frac{d(j)}{2m} \cdot 2m \sum_{k=2}^{n} \frac{1}{1 - \lambda_k} \cdot \left(\frac{v_{kj}^2}{d(j)} - \frac{v_{ki} v_{k,j}}{\sqrt{d(i) d(j)}} \right) \\ &= \sum_{j \in V} \sum_{k=2}^{n} \frac{1}{1 - \lambda_k} \cdot \left(v_{kj}^2 - \frac{v_{ki} v_{kj} \sqrt{d(j)}}{\sqrt{d(i)}} \right) \\ &= \sum_{k=2}^{n} \frac{1}{1 - \lambda_k} \left(\sum_{j \in V} v_{kj}^2 - v_{ki} \frac{1}{\sqrt{d(i)}} \cdot \sum_{j \in V} v_{kj} \sqrt{d(j)} \right) \\ &= \sum_{k=2}^{n} \frac{1}{1 - \lambda_k} \left(1 - v_{ki} \frac{1}{\sqrt{d(i)}} \cdot 0 \right) = \sum_{k=2}^{n} \frac{1}{1 - \lambda_k} \end{aligned}$$

where in the last line we used the fact that the eigenvectors $v_1, v_2, ..., v_n$ are orthonormal. Also it is easy to show that: $\kappa_i(j) = 2m \sum_{k=2}^{n} \frac{1}{1-\lambda_k} \left(\frac{v_{kj}}{\sqrt{d(j)}} - \frac{v_{ki}}{\sqrt{d(i)}} \right)^2$. Using that $\frac{1}{2} \leq \frac{1}{1-\lambda_k} \leq \frac{1}{1-\lambda_2}$ along with the orthogonality of the matrix (v_{ks}) we get $m \left(\frac{1}{d(i)} + \frac{1}{d(j)} \right) \leq \kappa_i(j) \leq \frac{2m}{1-\lambda_2} \left(\frac{1}{d(i)} + \frac{1}{d(j)} \right)$.

The following theorem is the result of [1, 6, 7, 8].

Theorem 3.1. *a) The first hit time between any two nodes of a graph on n nodes is at most*

$$\left(\frac{4}{27}\right)n^3-\left(\frac{1}{9}\right)n^2+\left(\frac{2}{3}\right)n-1,\ \ if\ \ n\overset{3}{\equiv}0$$
$$\left(\frac{4}{27}\right)n^3-\left(\frac{1}{9}\right)n^2+\left(\frac{2}{3}\right)n-\left(\frac{29}{27}\right),\ \ if\ \ n\overset{3}{\equiv}1$$
$$\left(\frac{4}{27}\right)n^3-\left(\frac{1}{9}\right)n^2+\left(\frac{4}{9}\right)n-\left(\frac{13}{27}\right),\ \ if\ \ n\overset{3}{\equiv}2.$$

b) The cover time from any starting node in a graph with n nodes is at least $(1-o(1))\,n\log n$ and at most $(4/27\ +o(1))\,n^3$.

c) The cover time of a regular graph on n nodes is at most $2n^2$.

Proposition 1 The probability that a random walk starting at i visits j before returning to i is $\frac{1}{\kappa_i(j)\pi(i)}$.

Definition 2 The cover time (with initial distribution) is defined as the average number of steps to reach each vertex. The worst case scenario is when no vertex is specified for the start (initial distribution is not defined). In this case, we should start from the vertex that maximizes the cover time. Hence the cover time, $cover_i(G)$, is the expected time of a random walk starting at vertex i in the graph G to reach each vertex at least once. We write $cover_i$ when G is understood. The cover time of an undirected graph G, denoted $cover_i(G)$, is $cover(G) = max_i cover_i(G)$.

Theorem 3.2. *Let G be a connected graph with n vertices and m edges. The time for a random walk to cover all vertices of the graph G is bounded above by $4m(n-1)$.*

Proof. Consider a depth first search of the graph G starting from some vertex k and let T be the resulting depth first search spanning tree of G. (Depth first search is an algorithm for traversing or searching tree or graph data structures. The algorithm starts at the root node and explores as far as possible along each branch before backtracking.) The depth first search covers every vertex. Consider the expected time to cover every vertex in the order visited by the depth first search. Clearly this bounds the cover time of G starting from vertex k. Note that each edge in T is traversed twice, once in each direction: $cover_k(G) \leq \sum_{(i,j),(j,i)\in T} h_i(j)$. If (i,j) is an edge in T, then i and j are adjacent, i.e. $(i,j) \in E$ and thus (4) implies $h_i(j) \leq 2m$. Since there are $n-1$ edges in the depth first search spanning tree and each edge is traversed twice, once in each direction, $cover_k(G) \leq 4m(n-1)$. This holds for all starting vertices k. Thus $cover_i(G) = max_k cover_k(G) \leq 4m(n-1)$. □

Example 1 Determine the first hit time for two points of a path on nodes $0, 1, ..., n-1$. The first hit time $H_{k-1}(k)$ is one less than the expected return time of a random walk on a path with $k+1$ nodes, starting at the last node. This return time is $2k$ so, $H_{k-1}(k) = 2k-1$. In other words,

$$\begin{aligned}
H_{k-1}(k) &= 1 + \frac{1}{d(k-1)} \sum_{j\in\Gamma(k-1)} H_j(k) \\
&= 1 + \frac{1}{2}[H_k(k) + H_{k-2}(k)] \\
&= 1 + \frac{1}{2}[H_{k-2}(k-1) + H_{k-1}(k)].
\end{aligned}$$

Let $a_k = H_{k-1}(k)$, thus solving for a_k yields the recurrence $a_k - a_{k-1} - 2 = 0$. Solving the recurrence equation yields $a_k = H_{k-1}(k) = 2k-1$.
Now consider the first hit times $H_i(k)$ where $0 \leq i < k \leq n$. In order to reach k we have to reach node $k-1$, this takes, on the average $H_i(k-1)$ steps. We have to get to k which takes, on the average $2k-1$ steps (the nodes beyond the k^{th} play no role). This yields the recurrence

$$H_i(k) = H_i(k-1) + H_{k-1}(k) = H_i(k-1) + 2k - 1$$

and then

$$\begin{aligned}
H_i(k) &= H_i(k-1) + 2k - 1 \\
&= H_i(k-2) + (2k-3) + (2k-1) \\
&= ... \\
&= H_i(i+1) + (2i+3) + ... + (2k-1) \\
&= (2i+1) + (2i+3) + ... + (2k-1) = k^2 - i^2
\end{aligned}$$

The second method is as follows:

$$\begin{aligned}
H_i(k) &= \sum_{j=i+1}^{k} H_j - 1(j) \\
&= \sum_{j=1}^{k-i} H_{j+i-1}(j+i) \\
&= \sum_{j=1}^{k-i} \big(2(j+i) - 1\big) \\
&= 2\frac{(k-i)(k-i+1)}{2} + (k-i)(2i-1) \\
&= (k-i)(k+i) = k^2 - i^2.
\end{aligned}$$

In particular, $H_0(k) = k^2$. Assuming that we start from 0, the cover time of the path on n nodes will also be $(n-1)^2$ since it suffices to reach the other endnode.

Example 2 Here of course we may assume that we start from 0 and to find the first hit times, it suffices to determine $H_0(1)$. The probability that we first reach node 1 in the t^{th} step is clearly $\left(\frac{n-2}{n-1}\right)^{t-1}\frac{1}{n-1}$, and so the expected time this happens is

$$H_0(1) = \sum_{t=1}^{\infty} t\left(\frac{n-2}{n-1}\right)^{t-1}\frac{1}{n-1} = n-1.$$

Let T_i denote the first time when i vertices have been visited. Hence

$$T_1 = 0 < T_2 = 1 < T_3 < ... < T_n.$$

Now $T_{i+1} - T_i$ is the number of steps while we wait for a new vertex to occur an event with probability $\frac{n-i}{n-1}$ independently of the previous steps. Hence $E(T_{i-1} - T_i) = \frac{n-1}{n-i}$, and so the cover time is

$$E(T_n) = \sum_{i=1}^{n-1} E(T_{i+1} - T_i) = \sum_{i=1}^{n-1} \frac{n-1}{n-i} \approx n \log n.$$

The following theorem is the result of [1, 6, 7, 8].

Theorem 3.3. *a) The first hit time between any two nodes of a graph on n nodes is at most*

$$\begin{array}{l}
\left(\frac{4}{27}\right)n^3 - \left(\frac{1}{9}\right)n^2 + \left(\frac{2}{3}\right)n - 1, \;\; if \quad n \overset{3}{\equiv} 0 \\
\left(\frac{4}{27}\right)n^3 - \left(\frac{1}{9}\right)n^2 + \left(\frac{2}{3}\right)n - \left(\frac{29}{27}\right), \;\; if \quad n \overset{3}{\equiv} 1 \\
\left(\frac{4}{27}\right)n^3 - \left(\frac{1}{9}\right)n^2 + \left(\frac{4}{9}\right)n - \left(\frac{13}{27}\right), \;\; if \quad n \overset{3}{\equiv} 2.
\end{array}$$

b) The cover time from any starting node in a graph with n nodes is at least $(1 - o(1))\, n \log n$ and at most $(4/27\ + o(1))\, n^3$.

c) The cover time of a regular graph on n nodes is at most $2n^2$.

Poposition 2 The probability that a random walk starting at i visits j before returning to i is $\frac{1}{\kappa_i(j)\pi(i)}$.

Proof. Let T_i be the first time when a random walk starting at i returns to i and T_{ij} the first time when it returns to i after visiting j. Observe that $T_i \leq T_{ij}$.
Let $p = P(T_i = T_{ij})$ be the probability that a random walk starting at i visits j before returning to i. Therefore since $T_i \leq T_{ij}$, we can say that

$$1 - p = P(T_i \neq T_{ij}) = P(T_i < T_{ij}) + P(T_i > T_{ij}) = P(T_i < T_{ij}).$$

Notice that $E(T_i) = \frac{1}{\pi(i)} = \frac{2m}{d(i)}$, $E(T_{ij}) = \kappa_i(j)$.

Since if $T_i < T_{ij}$, then after the first T_i steps we have to walk from i until we reach j and then return to i, we have

$$\begin{aligned} E(T_{ij}) - E(T_i) &= E(T_{ij} - T_i) \\ &= pE(T_{ij} - T_i | T_i = T_{ij}) + (1-p)E(T_{ij} - T_i | T_i < T_{ij}) \\ &= (1-p)E(T_{ij}) = E(T_{ij}) - pE(T_{ij}). \end{aligned}$$

Hence, $p = \frac{E(T_i)}{E(T_{ij})} = \frac{2m}{d(i)\kappa_i(j)} = \frac{1}{\kappa_i(j)\pi(i)}$. □

Theorem 3.4. *For any three nodes i, j and k of a connected, undirected graph G: $H_i(j) + H_j(k) + H_k(i) = H_i(k) + H_k(j) + H_j(i)$.*

Proof. Essentially, this equality is a consequence of the reversibility of the Markov chain for random walks on an undirected graph. Note that the left-hand side of the equation in the theorem is the expected time for a random walk to go from i to j, then to k and back to i, and similarly for the right.
Now fix a number r and begin a random walk at i, ending when i is reached again for the r^{th} time. Let $i, i_1, i_2, ...i_r, i$ be the outcome of the walk; its probability is $\frac{1}{d(i)} \prod_{x=1}^{r} \frac{1}{d(i_x)}$. However, this value is, of course, the same as the probability of the reverse walk $i, i_r, i_{r-1}, ...i_1, i$, i.e. $\frac{1}{d(i)} \prod_{x=r}^{1} \frac{1}{d(i_x)}$. Now, we claim that the number of $i \to j \to k \to i$ tours in one of these walks is the same as the number of $i \to k \to j \to i$ tours in its reverse. To see this, note that the greedy algorithm for finding such tours starting from the left is optimal and thus yields at least as many such tours as we can find by listing $i \to k \to j \to i$ tours from the right; the symmetric argument establishes equality. It follows that the expected lengths of the two types of tours from i to i are the same, proving the theorem. □

Note that the nodes of any graph can be ordered so that if i precedes j then $H_i(j) \leq H_j(i)$. Such an ordering can be obtained by fixing any node t and order the nodes according to the value of $H_i(t) - H_t(i)$. Also we define the potential function as $H_i(t) - H_t(i)$. However, from the next proposition, we can deduce the fact that there is a pure hitting time strategy whose tournament is transitive.

Proposition 3 On any graph G, the vertex-relation given by

$$i \leq j \Leftrightarrow H_i(j) \leq H_j(i)$$

is transitive, i.e., constitutes a preorder on the vertices of G.

Proof. The proof is immediate from the equation of Theorem 3.4. Assume that i precedes j in the ordering described. Then $H_i(t) - H_t(i) \leq H_j(t) - H_t(j)$ and hence $H_i(t) + H_t(j) \leq H_j(t) + H_t(i)$. By Theorem 3.4, this is equivalent to saying that $H_i(j) \leq H_j(i)$. □

For us, the important consequence of Proposition 3 is that there is always a vertex that is minimal in this preorder and thus satisfies $H_j(t) \leq H_t(j)$ for every other vertex j of G. Such a vertex will be called hidden. (As an example, the reader may verify that a vertex of a tree is hidden just if its average distance to other vertices of the tree is maximum.)

This ordering is not unique because of the ties. But if we partition the nodes by putting i and j in the same class if $H_i(j) = H_j(i)$ (this is an equivalence relation by Proposition 3), then there is a well-defined ordering of the equivalence classes independent of the reference node t. The nodes in the lowest class are "difficult to reach but easy to get out of", the nodes in the highest class are "easy to reach but difficult to get out of". One of the important results regarding this symmetric property is as follows:

Proposition 4 If a graph has a vertex-transitive automorphism group, then $H_i(j) = H_j(i)$ for all nodes i and j.
Let $G = (V, E)$ be a connected graph and $S \subset V$. A function $\varphi : V \longrightarrow \mathbb{R}$ is a harmonic function with boundary S if $\frac{1}{d(i)} \sum_{j\in\Gamma(i)} \varphi(j) = \varphi(i)$ holds for every $i \in V \setminus S$.
For a random walk, a harmonic function has the following interpretation. Suppose at a given time k the random walk is visiting vertex i. Then $E(\varphi(X_{k+1})|X_k = i) = \sum_{j\in V} \varphi(j)p_{ij} = \frac{1}{d(i)} \sum_{j\in\Gamma(i)} \varphi(j) = \varphi(i)$. Thus, the stochastic process $\{\varphi(X_t)\}_{t\in\mathbb{N}\cup\{0\}}$ is a martingale with respect to X and so we are playing a fair game.

Example 3 Let $S = \{s, t\}$. Let $\varphi(i)$ denote the probability that a random walk starting at i hits s before it hits t. By conditioning on the first step we have $\varphi(i) = \sum_{j\in V} \varphi(j)p_{ij} = \frac{1}{d(i)} \sum_{j\in\Gamma(i)} \varphi(j)$, for every $i \in V \setminus S$. Also, $\varphi(s) = 1$, $\varphi(t) = 0$. That is, φ is harmonic with boundary $\{s, t\}$.
This example is applicable in the Gambler's Ruin problem on a graph. A gambler enters a casino with a plan to play the following game. At each turn he will bet 1 dollar to win 2 dollars with probability $\frac{1}{2}$ and lose his money with probability $\frac{1}{2}$. He is determined to leave either when he is ruined (i.e. he has no money left) or as soon as he collects N dollars. Now, Consider a generalization. Suppose we have a random walk in a graph. We are placed on some starting vertex and may walk through the maze until we reach either vertex t or vertex s. If we hit vertex s we win 1 dollar and we get nothing if we reach vertex t. What is the probability φ_i we win 1 dollar starting at vertex i? We can write the system of equations: $\varphi(t) = 0$, $\varphi(s) = 1$, $\varphi(i) = \frac{1}{d(i)} \sum_{j\in\Gamma(i)} \varphi(j)$.
More generally, in Example 3, let $S \subset V$ and suppose we have a function $\varphi_0 : S \longrightarrow \mathbb{R}$. Let $\varphi(i)$ be the expected valued of $\varphi_0(s)$, where s is the random vertex where the random walk started at i first hits S. Again

$$\varphi(i) = \frac{1}{d(i)} \sum_{j\in\Gamma(i)} \varphi(j)$$

and φ is harmonic with boundary S.

References

[1] Aldous, D. J., and Fill, J. A., *Reversible Markov Chains and Random Walks on Graphs.* University of California, Berkeley, (2002).

[2] Brémaud, P., *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues.* Texts in Applied Mathematics, 31, Ed 2, Springer Nature Switzerland AG, (2020).

[3] Coppersmith, D., Tetali, P. and Winkler P., Collisions among random walks on a graph. *SIAM Journal on Discrete Mathematics*, 6 (1993) 363-374.

[4] Diaconis, P., Group Representations in Probability and Statistics. *Institute of Mathematical Statistics*, Hayward, Californis, (1988).

[5] Doyle, P. G., and Snell, J. L., *Random walks and Electric Networks.* Mathematical Association of America, (1984).

[6] Feige, U. A., tight upper bound on the cover time for random walks on graphs. *Random Structures and Algorithms*, 6 (1995) 51-54.

[7] Feige, U. A., tight lower bound on the cover time for random walks on graphs. *Random Structures and Algorithms*, 6 (1995) 433-438.

[8] Feige, U., *Collecting coupons on trees, and the analysis of random walks.* Technical report CS93-20 of the Weizmann Institute, (1393).

[9] Grimmett, G., *Probability on graphs.* Cambridge University Press, (2010).

[10] Gross, J. L., Yellen, J. and Anderson, M., *Graph Theory and its Applications.* Chapman. Hall/CRC, Third Edition, (2018).

[11] Kelarev, A. V., On undirected Cayley graphs. *The Australasian Journal of Combinatorics*, 25 (2002) 73-78.

[12] Polya, G., Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz. *Mathematische Annalen*, (1921) 149-160.

[13] Schmidt, S., On the quantum symmetry of distance-transitive graphs. *Advances in Mathematics*, 368 (2020), Article ID 107150, 1-50.

[14] Tetali, P., Random walks and effective resistance of networks. *Journal of Theoretical Probability*, (1991) 101-109.

[15] Thomassen, C., Resistences and currents in infinite electrical networks. *Journal of Combinatorial Theory*, (1990) 87-102.

[16] Woess, W., *Random walks on infinite graphs and groups.* Cambridge university press, (2000).