# Estimating Transition Matrices and their associated quantities from Random Walks

Kessem Clein

January 24, 2021

## 1 Introduction

1. Covering time in classic random walks
2. RWs with arbitrary transition probabilities
3. RWs with unknown transition probabilities
4. RWs with unknown transition probabilities and unknown vertices

## 2 Preliminaries

Let $T \in \mathbb{R}^{n \times n}$ be a row-stochastic transition probability matrix, describing the probabilities of hopping amongst a discrete set of states. The elements $T_{ij}$ give the probability of an $i \to j$ transition in a single step. We denote this probability as:

$$P_T(i \to j)$$

In the following we refer to a *random walk* as a random walk on the fully connected graph with $n$ vertices $K_n$, where the transition probabilities are given by $T$.
The covering time $C_G$ of a graph $G$ is the expected number of steps until a random walk on $G$ visits all the vertices of $G$. In the case of a highly regular, or very connected graph, it has been shown that the expected number of steps it takes to cover **any** two sets of vertices of equal size should be very similar [ref].
*REPHRASE THIS:* We are interested in investigating the timescales of expected covering times, in cases where there are nearly disconnected sets in the graph, such that the covering time of a set of vertices can vary drastically with the starting point.

**Definition 1.** *first passage time (FPT)*
*For any $i, j \in V$ we define the random variable $FPT_{ij}$ to be the first time a random walk that originates from $i$ reaches $j$.*

It can be useful to understand the distribution of the FPT. In particular, it has been shown [ref 10.1103/PhysRevA.42.2047] that for $t$ larger than a fraction of $T = \mathbb{E}[FPT_{ij}]$, the value $P(FPT_{ij} = t)$ is approximated to a high degree of accuracy by:

$$P(FPT_{ij} = t) \approx \frac{1}{T} \exp\left(\frac{-t}{T}\right) \tag{1}$$

This means that for any $\epsilon > 0$, there is a $t > T$ for which we can determine that $P(FPT_{ij} > t) < \epsilon$.

**Definition 2.** *$\epsilon$-certain passage time*
*For $\epsilon > 0$ we define the $\epsilon$-certain passage time from $i$ to $j$ as:*

$$\epsilon\text{-}CPT_{ij} = \min_k \{k \in \mathbb{N} | P(FPT_{ij} \geq k) < \epsilon\}$$

In words, if $\epsilon$-CPT$_{ij} = t$, this means that it is nearly certain (with probability at least $1 - \epsilon$) that a random walk originating from $i$ will visit $j$ in at most $t$ steps.
**Note:** if $\epsilon$ is chosen to be a relatively large fraction of 1, although the definitions stay the same, a more appropriate semantic interpretation of the $\epsilon$-certain passage time is that "if $\epsilon$-CPT$_{ij} = t$, this means that it is **plausible** (with probability at least $1 - \epsilon$) that a random walk originating from $i$ will visit $j$ in at most $t$ steps. "

**Definition 3.** *$\epsilon$-covering time NOT SURE WE NEED THIS*
*For any set $B \subset \{1, \ldots, n\}$, and vertex $i$ we define the $\epsilon$-covering time of $B$ from $i$ to be:*

$$C_\epsilon(i, B) = \max\{\epsilon\text{-}CPT_{ij} | j \in B\}$$

**Definition 4.** *$\epsilon$-covering ball*
*For any vertex $i$ and radius $r$ we define the $\epsilon$-covering ball of radius $r$ around $i$ to be:*

$$B_\epsilon(i, r) = \{j \in [n] | \epsilon\text{-}CPT_{ij} \leq r\}$$

*Equivalently, $B_\epsilon(i, r)$ is the maximal set $B \subset \{1, \ldots, n\}$ such that*

$$C_\epsilon(i, B) \leq r.$$

**Note:** From here on we will assume some fixed $\epsilon$, and emit the $\epsilon$ from our notations, *e.g. we will write $CPT_{ij}$ instead of $\epsilon$-$CPT_{ij}$*
¡SOME DISCUSSION EXPLAINING THESE DEFINITIONS ¿

# 3 Estimation of First Passage Times

In cases where precise transition probabilities are unknown apriori, they can be estimated if there is a way to produce samples of random walks.
Denote by $C$ a transition count matrix, describing the number of transitions observed between vertices in a set of sampled random walks. The elements $C_{ij}$

count the number of $i \to j$ transitions observed.

By the Markovian assumption on random walks, each time the current state is a vertex $i$, the next state is sampled from the categorical distribution parameterized by the vector $T_i$ (denoting transition probabilities from vertex $i$) Therefore, each row $C_i$, which counts all the transitions observed *from* the vertex $i$, is a random variable sampled from the multinomial distribution parameterized by the vector $T_i$:

$$C_i \sim Mul(T_i, k_i) \tag{2}$$

where $k_i = \sum_{j=1}^{n} C_{ij}$.

However if we don't know the precise value of $T$, we can still estimate it via it's posterior distribution using the observed count matrix $C$. Since each $C_i$ has a multinomial distribution parameterized by $T_i$, if we assume a dirichlet prior distribution on $T_i$:

$$T_i \sim Dir(\alpha_i) \tag{3}$$

then the posterior distribution $P(T_i|C_i)$ is also a dirichlet distribution:

$$T_i|C_i \sim Dir(C_i + \alpha_i) \tag{4}$$

This observation will allow us to compute and estimate some some useful values. Namely, we will be able to estimate the distribution $P(CPT_{ij})$. Lets define a few values which will be of particular use:

**Definition 5.** *$\delta$-probable certain passage time*
*We define the $\delta$-probable certain passage time from $i$ to $j$ to be:*

$$d_\delta(i, j|C) = \min_n \{k \in \mathbb{N} | P(CPT_{ij} \leq k) > 1 - \delta\}.$$

The $\delta$-probable certain passage time from $i$ to $j$ is essentially a rough upper bound on $CPT_{ij}$, that expresses, given our observation $C$, the first time we are almost sure that that $j$ is in "$CPT - range$" of $i$.
It is useful to think of it as a sort of *non-symmetric-distance* from $i$ to $j$. Since we have defined a sort of distance, the next natural question arises: *What does a ball look like under this distance?*. For example, say we want to define a ball of radius $r$ using this distance function around the vertex $i$. If for a vertex $j$ it holds that

$$D_\delta(i, j|C) < r$$

we could assign $j$ to the ball, and be confident that $j$ will stay inside this ball even after we get more data and update our observations $C$. However, if

$$D_\delta(i, j|C) > r$$

we are not necessarily confident that $j$ will remain *outside* of the ball if we obtain more data, so we can not define such a ball with confidence.

To address this, we define the complementary matching lower bound:

**Definition 6.** *$\delta$-probable uncertain passage time (I AM $\delta$-CERTAIN WE NEED A BETTER NAME FOR THIS)*
*We define the $\delta$-probable uncertain passage time from $i$ to $j$ to be:*

$$D_\delta(i, j|C) = \max_n \{k \in \mathbb{N}|P(CPT_{ij} > k) > 1 - \delta\}.$$

Similarly to $d_\delta(i, j|C)$, the value $D_\delta(i, j|C)$ represents a rough lower bound to $CPT_{ij}$.
With these two bounds in hand, we can determine exactly for which $r$'s we can confidently define a ball around a vertex.

**Definition 7.** *CPT confidence time.*
*We define the $\delta-CPT$ confidence time of a vertex $i$ to be:*

$$Conf_\delta(i|C) = \{r|\forall j \in [n] : r \notin (d_\delta(i, j|C), D_\delta(i, j|C))\}.$$

For a vertex $i$, if $r \in Conf_\delta(i|C)$, then for each vertex $j$ we can determine with confidence if $CFP_{ij} < r$ or $CFP_{ij} > r$.
Now we have all we need to have a well defined $\epsilon-$covering ball in cases where there is uncertainty about $T$.

**Definition 8.** *$\epsilon, \delta$-covering ball around a vertex*
*For any vertex $i$ and radius $r \in Conf_\delta(i|C)$ we define the $\epsilon, \delta$-covering ball of radius $r$ around $i$ to be:*

$$B_{\epsilon,\delta}(i, r) = \{j \in [n]|d_\delta(i, j|C) \leq r\}.$$

Note that the complement of the ball $B_{\epsilon,\delta}(i, r)$ is exactly

$$B^c_{\epsilon,\delta}(i, r) = \{j \in [n]|D_\delta(i, j|C) > r\}$$

meaning that each vertex is assigned with probability $> 1 - \delta$ to it's correct set.
**Note:** Once again, to avoid excessive notation, from here on we will assume some fixed $\delta$ and $\epsilon$, and omit writing them in our notation.

### 3.1 Calculating first passage time probabilities

On a random walk with transition probabilities given by a transition matrix $T$, where $T_{ij}$ gives the probability of a $i \to j$ transition in one step, we can see that:

$$T^2_{ij} = \sum_{k=1}^{n} T_{ik} \cdot T_{kj} = \sum_{k=1}^{n} P(i \to k)P(k \to j) = \sum_{k=1}^{n} P(i \to k \to j) \tag{5}$$

Which by the law of total probability, is exactly the probability of an $i \to j$ transition in two steps. Likewise it can be seen that for any number of steps $k$, exponentiating the matrix $T$ to the $k$'th power gives the probabilities of transitions in exactly $k$ steps.

To find the probability that a transition $i \to j$ occurred at least once in $k$ steps, we can modify the transition probabilities to make $j$ an absorbing vertex (a sink), and then calculate the probability of an $i \to j$ occurring in $k$ steps with the modified transition probability matrix.
In the modified setting, any random walk that reaches the vertex $j$ will remain there for the rest of the random walk, therefore if $\tilde{T}(j)$ is the modified transition matrix (with $\tilde{T}_{jk}(j) = \mathbb{1}_{j=k}$) then

$$\tilde{T}^k_{ij}(j) = P(FPT_{ij} \leq k) \tag{6}$$

So that in the case where the transition probability matrix $T$ is known, we have found a closed form representation of the distribution $P(FPT_{ij})$, and therefore of the value:

$$CPT_{ij} = \min_k \{k \in \mathbb{N} | \tilde{T}^k_{ij}(j) > 1 - \epsilon\} \tag{7}$$

Using this representation we can calculate the exact value of $CPT_{ij}$ by simply exponentiating $\tilde{T}(j)$ until the first time $\tilde{T}^k_{ij}(j) > 1 - \epsilon$.
The complexity of this calculation is $O(n^3 \log k)$.
*TODO:* I'm pretty sure we can find $k$ by solving a linear system of the form $Tx = \epsilon b$ (which is also $O(n^3)$).

### 3.2 Estimating first passage time probabilities from data

We now return to the setting where the precise transition probabilities are unknown to us, but we have a transition count matrix $C$ calculated from a set of sampled random walks. Using (4), we can generate samples from the posterior distribution of the rows of $T$ given $C$, and therefore generate samples from the posterior distribution of $T$ given $C$.
By modifying these samples and setting one of the vertices as an absorbing vertex, and exponentiating the samples to the $k$'th power, using (**??**) and (7) we can directly sample from the posterior distribution $P(FPT_{ij} \leq k|C)$ and $P(CPT_{ij}|C)$ to estimate them, and determine the values of $d_\delta(i, j|C)$ and $D_\delta(i, j|C)$.

For example, a naive way to determine if $d_\delta(i, j|C) < k$ is to sample a large set of transition probability matrices $S$ *s.t.* $T \in S \sim T|C$, and estimate:

$$P(CPT_{ij} \leq k) \approx \frac{|\{T \in S : CPT_{ij}(T) \leq k\}|}{|S|} \tag{8}$$

Where $CPT_{ij}(T)$ is the value of $CPT_{ij}$ calculated as in (7) using the matrix $T$.

Other ways of estimating the values $d_\delta(i, j|C)$, $D_\delta(i, j|C)$ more efficiently may include estimating the distribution of $P(CPT_{ij}|C)$, for example by fitting a Gaussian distribution to it.

### 3.3 Spectral properties of T

1. classifying balls and transition rates with eigenvectors
2. estimation and uncertainty of eigenvalues

# 4 Random walks with unknown vertices

1. unseen vertex trick
2. $Conf_\delta(i|C) = \{r < D_\delta(i,n|C) | \forall j \in [n] : r < D_\delta(i,j|C) \text{ or } r > d_\delta(i,j|C)\}$

# 5 Transition Matrices of Hierarchical Graphs

## 5.1 Hierarchical representation of a graph

#some sentences about when the graph has a nested graph structure, and how representing it with a coarse grained graph may be useful

**Definition 9.** *Graph Partition*
*Let $G$ be a graph, associated with a transition probability matrix $T$, and let $W = \{V_1, \ldots, V_k\}$ be a partition of $V$, the vertices of $G$.*
*Let the graph $G_i$ be the graph with the vertices $V_i$, and an additional vertex $w_j$ for each $V_j \in W$ s.t. $j \neq i$.*
*Further, let the transition probability matrix $T_i$ be the transition probability matrix of $G_i$, such that*

$$\forall v, w \in V_i : P_{T_i}(v \to w) = P_T(v \to w)$$

*And for any $j \neq i$:*

$$\forall v \in V_i : P_{T_i}(v \to w_j) = \sum_{w \in V_j} P_T(v \to w)$$
$$P_{T_i}(w_j \to v) = 0$$
$$P_{T_i}(w_j \to w_j) = 1$$

*Then we define the set of graphs with their associated transition probability matrices:*

$$G_W = \{(G_1, T_1), \ldots, (G_k, T_k)\}$$

*to be a graph partition of $G$.*

For an arbitrary partition $W$, a graph partition $G_W$ is unlikely to be useful in any meaningful way.
However, if $W$ partitions $G$ in such a way that a random walk on $G$ sampled via the matrix $T^k$ for $k > 1$, can be estimated to a high degree of accuracy as a random walk between on $G_W$ (how this graph is defined exactly will be

defined soon), then such a partition can provide a useful model to understand the behavior of random walks on G - both in terms of human understanding and computation.

To formalize this requirement, we must introduce the following standard definition:

**Definition 10.** *Mixing time*
*Let $\pi$ be the stationary distribution of a random walk on a graph $G$. For a vertex $i$, denote by $Q_i^k$ the probability distribution of the random walk on $G$ after $k$ steps, starting at $i$.*
*We say that the random walk mixes after $k$ steps if:*

$$\forall i \in [n] : \|Q_i^k - \pi\| < \frac{1}{4}.$$

*The mixing time $mix(G)$ of a random walk on $G$ is the smallest such $k$.*

With this definition in hand, we can now introduce a criterion for a graph partition to give a good estimation of long term transitions on a graph.
For ease of notation, for a subset $V_i \subset V$ we denote:

$$P_T(v \to V_i) = \sum_{w \in V_i} P_T(v \to w).$$

**Definition 11.** *$\epsilon$-Graph partition*
*Let $\epsilon > 0$, and $W = \{V_1, \dots, V_k\}$ be a partition of $V$, the vertices of $G$.*
*Let $mix(G|_{V_i})$ be the mixing time of the subgraph of $G$ restricted to the vertices $V_i$.*
*If for all $i \neq j$:*

$$\forall v \in V_i : P_{T^{mix(G|_{V_i})}}(v \to V_j) < \epsilon.$$

*Then $G_W$ is an $\epsilon$-graph partition of $G$.*

In essence, if $G_W$ is an $\epsilon$-graph partition, then for any $t \geq \max_{i \in [k]} mix(G|_{V_i})$, a random walk sampled at intervals of $t$ steps, 'forgets' (up to a factor of $\epsilon$) which vertex it visited last, and only remembers which subgraph $G|_{V_i}$ it was in.
We will give a formal proof for this once we have defined how to construct a 'Metagraph' out of a graph partition, which describes random walks in timesteps of $t$ steps.

# example: two cliques with very small probability of traversing between them. Two graphs with a 'narrow bridge' between them

**Definition 12.** *Metagraph / integrative graph?*
*Let $G_W = \{(G_1, T_1), \dots, (G_k, T_k)\}$ be an $\epsilon$-graph partition of $G$, and $\tau > \max_{i \in [k]} mix(G|_{V_i})$. We define the metagraph with timescale $\tau$ of $G$ w.r.t. $G_W$*

*to be the graph with $k$ vertices, and associated transition probability matrix $T_W$, such that for $i \neq j$:*

$$P_{T_W}(i \to j) = \mathbb{E}_{v \sim \pi_i}[P_{T_i^\tau}(v \to w_j)].$$

*Where $\pi_i$ is the stationary distribution on $G|_{V_i}$, and $T_i^\tau$ is the transition probability matrix $T_i$ exponentiated $\tau$ times, which gives the probabilities of transitions occurring in $\tau$ steps.*

**Note:** From here on we will refer to $G_W$ as the Metagraph induced by the graph partition $G_W$, and we will assume it is an $\epsilon$-graph partition, and omit the $\epsilon$ notation.

**Claim 1.** *Let $G_W$ be a Metagraph of $G$ with timescale $\tau$. Then for all $i, j \in [k]$, and for all $v \in V_i$:*

$$|P_{T^\tau}(v \to V_j) - P_{T_W}(i \to j)| < 3\epsilon.$$

*Proof.* (Okay theres a bit of a hole in my proof...)

□

# 6 Adaptive Sampling Strategies