ON ONLINE AND APPROXIMATE COVER TIME PROBLEMS

by
Mehmet Akif Yıldız
B.S., Industrial Engineering, Boğaziçi University, 2017
B.S., Mathematics, Boğaziçi University, 2017

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Mathematics
Boğaziçi University
2020

ACKNOWLEDGEMENTS

First of all, I am extremely grateful to my thesis advisor Ümit Işlak and I feel very lucky to have such a supervisor due to his passion about mathematics and particularly his research. He always pushed me toward scientific studies and helped me to improve my vision. Moreover, he has been more than just a professor for me, we were like fellows during all my master years. I will always remember those days well and try to be a mathematician like him in the future.

Secondly, I would like to thank Tınaz Ekim due to her guidance since she had been very influential on me while I was deciding to be a graduate student on mathematics. Throughout the last two years, I have been a part of her project and I learned the processes of a collaborative work and being a researcher via this project.

Next, I am very happy to meet Mohan Ravichandran due to his intellectual and mathematical thinking capability, and he always improved my knowledge about different fields of mathematics. Through his study group, I had chance to discuss many interesting studies with other students and him. Also, I am very grateful to Özgür Martin for being a part of my thesis committee.

I have difficulties while choosing words to describe how I am thankful to Süheyla, my wife, for her support and understanding in this long journey. She was always there when I need her in good and bad days. Also, I am appreciated to my parents Reyhan and Servet, my brother Emin, and especially my sister Ayşe since she was also a great housemate for years.

Finally, I would like to thank my dear genius friend Yunus Emre Demirci. Since fifteen years, we always had long discussions about various mathematical problems and he has a significant impact on me due to his approach style towards to the problems.

ABSTRACT

ON ONLINE AND APPROXIMATE COVER TIME PROBLEMS

As a generalization of the classical coupon collector problem in the probability theory, the cover time in random walks on Markov chains has been investigated in numerous studies in the literature. Especially, there are several results for the cover time of a simple random walk on connected and undirected graphs. In this thesis, we study two new problems about the cover time of graphs.

Firstly, we build an on-line model where there is a walker moving with random time intervals on a graph growing in time. We initiate this study by examining the number of vertices covered up to a fixed time for a simple model, and we discuss further research directions.

Secondly, we generalize the classical cover time definition in order to understand the differences between the partial covering and the full covering. We initiate this study with the investigation of the approximate covering time on specific graph families such as path graphs and complete graphs, and our main motivation is to explore the structure of the graphs allowing easy partial covering in terms of the order of magnitude.

For the sake of completeness, we also give some preliminary results about the classical cover time problem and several variations of the problem from the literature such as edge covering and dynamic versions in the thesis.

ÖZET

ÇEVRİMİÇİ VE YAKLAŞIK KAPLAMA ZAMANI PROBLEMLERİ ÜZERİNE

Olasılık teorisindeki klasikleşmiş kupon toplama probleminin bir genellemesi olarak, Markov zincirleri üzerindeki rastgele dolaşmanın kaplama zamanı literatürdeki birçok çalışmada incelenmiştir. Özellikle, bağlantılı ve yönsüz çizgeler üzerindeki basit rastgele dolaşmanın kaplama zamanı üzerine birçok sonuç bulunmaktadır. Bu tezde, çizgelerin kaplama zamanı üzerine iki yeni problem üzerine çalışılmıştır.

İlk olarak, zamanla büyüyen bir çizge üzerinde hareketlerinin zaman aralıkları rassal olan bir gezicinin olduğu bir çevrimiçi model inşa edilmiştir. Belirli bir zamana kadar kaplanan köşelerin sayısı basit bir model üzerinde incelenerek bu çalışma başlatılmış ve çeşitli araştırma alanları tartışılmıştır.

İkincisi, klasik kaplama zamanı tanımı genelleştirilerek, kısmi kaplama ve tam kaplama arasındaki farklar anlaşılmaya çalışılmıştır. Yolak çizgesi ve tam çizge gibi özel çizge ailelerinde yaklaşık kaplama zamanı incelenerek bu çalışma başlatılmış, büyüklük derecesi olarak daha kolay yaklaşık zamanına izin veren çizge yapıları keşfedilmesi temel motivasyon olmuştur.

Bütünlük açısından, klasik kaplama zamanı üzerine temel sonuçlar ile kenar kaplama ve dinamik versiyonlar gibi problemin literatürde yer alan çeşitli varyasyonları da tezde verilmiştir.

TABLE OF CONTENTS

LIST OF SYMBOLS

$\mathbb{1}_A$	The indicator function of the set A
$\|A\|$	The cardinality of a set A
$Bin(n,p)$	The binomial distribution with parameters n and p
C	The first time that all vertices of a graph have been visited by the random walk
$C_{1-\alpha}$	The first time that at least $(1-\alpha)n$ vertices of a graph of order n have been visited by the random walk
$com(G)$	Commute time of a graph G
$cov(G)$	Cover time of a graph G
$cov(G_{1-\alpha})$	Approximate covering time of a graph with parameter α
$E(G)$	The edge set of a graph G
$\mathbb{E}[X]$	The expectation of the random variable X
$\mathbb{E}[X\|Y]$	The expectation of the random variable X when the event Y is given
$\mathbb{E}_u[X]$	The expectation of the random variable X in a simple random walk on a graph starting at the vertex u
$\mathcal{H}(n)$	n^{th} harmonic sum of natural numbers: $\mathcal{H}(n) = \sum_{k=1}^{n} \frac{1}{k}$
$hit(G)$	Hitting time of a graph G
K_n	Complete graph on n vertices
$K_{n,m}$	Complete bipartite graph with partitions size n and m
$\mathbb{N}$	The set of natural numbers
$\mathbb{N}_0$	The set of extended natural numbers including 0
N_T	The number of vertices visited at least once by the random walk up to time T
$o(f(n))$	A function whose growth rate is strictly smaller than $f(n)$
$O(f(n))$	A function whose growth rate is less than or equal $f(n)$
P_n	The path graph on n vertices
$Poi(\lambda)$	The Poisson distribution with expectation λ

$\mathbb{P}[E]$	The probability of the event E
$\mathbb{P}[E\|Y]$	The probability of the event E when the event Y is given
$\mathbb{P}_u[E]$	The probability of the event E in a simple random walk on a graph starting at the vertex u
$\mathbb{R}$	The set of real numbers
S_n	The cycle graph on n vertices
T_j	Hitting time of the state j in a Markov chain
T_j^+	Expected return time of the state j in a Markov chain
$U(a,b)$	The uniform distribution on the interval (a,b)
$V(G)$	The vertex set of a graph G
$\mathrm{Var}[X]$	Variance of a random variable X
$\Delta(u,v)$	The length of the shortest path between the vertices u and v in a connected graph
$\Theta(f(n))$	A function whose growth rate is equal to $f(n)$
$\Omega(f(n))$	A function whose growth rate greater than or equal to $f(n)$

LIST OF ACRONYMS/ABBREVIATIONS

i.e.	in other words
i.i.d.	independent and identically distributed
w.h.p.	with high probability

1. INTRODUCTION

Remember the times that cards having footballer's photos with their skills were given as a gift by gums during world cup tournaments. Suppose you are interested in collecting all Brazil squad of 23 people, and each gum has exactly one gift card. How many gums are necessary for completing the collection? Of course, it is dependent on your luck, but can we say anything for the average number of gums?

Let us investigate this phenomena on a very simple example. Assume each gum has one card belonging *Ronaldo*, *Rivaldo*, or *Ronaldinho*, and we are trying to complete this great triple. On each day, we buy a gum, and obtain a random footballer's card. After some days, we have the following cards in the order:

$$Ronaldo, Rivaldo, Ronaldo, Ronaldo, Ronaldinho, Rivaldo, Ronaldinho, ...$$

In this example, we completed the collection in 5 days. Then, can we say that we are lucky? We can answer this question by calculating the expected number of days required.

This is a very simple example of the famous problem, known as *coupon collector problem*, in the probability theory (see [1]). In the general setting, mathematicians are interested in the time required for all elements of a certain set at least once, which is called as *cover time* by a walker who traverses randomly on those elements with respect to some probabilistic rules.

In this thesis, we are interested in the cover time of a simple random walk on graphs. Starting from some of the preliminaries about probability and graph theory in Chapter 2, first we explore some nice relations between statistics related with cover time in Chapter 3. Then, in Chapter 4, we give exact formulas for the cover time in specific graphs such as complete graphs or cycles.

It is known that the calculation of the cover time in general graphs is not so easy. Therefore, we may need to bound the cover time to understand its magnitude. Especially, it turns out that the decision on whether it can be exponential is very important. In Chapter 5, we discuss how a polynomial bound can be obtained through a spanning tree argument in any connected graph.

The cover time problem has numerous extensions and variations in the literature. In Chapter 6, we give several of them where we also discuss the proofs in some cases for the sake of completeness. We will start with an edge covering version of the classical problem, and then continue with the variation that there are more than one walker who are simultaneously covering the graph. Moreover, we will state two models about the cover time which are dynamic in terms of the walker and graph, and then we give a result about the case that our graph is random at the beginning. At the end of the Chapter 6, we discuss some theorems that allow us to improve the general bound when we have an additional properties on the graph studied.

In Chapter 7 and 8, we give two new investigations of the cover time problem, of course up to our the best knowledge, namely *on-line covering* and *approximate covering*. In the first new variation, we are interested in the number of covered vertices up to a certain time by the walker while our initial graph is growing with time for which the time intervals between walker's moves are randomly determined. In the latter, we extend the definition of the cover time to reflect the partial covering of a graph.

As well as the classical cover time problem has numerous open questions such as the minimum value of the cover time that can be achieved on a graph of fixed size, our new models are also very open to generalize for further studies. In the Chapter 9, we discuss some of them.

2. PRELIMINARIES

In this chapter, we discuss some preliminaries that are prerequisites of the topics to be investigated. This includes some basic probabilistic and graph theoretical definitions and theorems as well as an introduction to Markov chains. At the end of the chapter, we define simple random walks on graphs.

2.1. Basic Probability

In this section, we give the necessary basics for probability theory used throughout the thesis. In general, following concepts are widely common in the literature, and for sake of the completeness we follow the books [2] and [3]. The reader can look at those books for further information as well as other references such as [4], [5], and [6] can be used to introduce the ideas in probability theory.

In simplest words, a *random variable* can be thought as a result of some experiment such as tossing a coin or rolling a die. The set of all possible outcomes is called *sample space*. For instance, let X be a random variable taking the values “blue”, “red”, or “green”. Here, our sample space becomes the set $\{$blue,red,green$\}$ of size three. Suppose that, in the long-run, X takes the value ”blue” with frequency $\frac{1}{2}$, in other words, X is assigned with “blue” in the half of the all cases. Similarly, assume X takes the value “red” with frequency $\frac{1}{3}$, then we can say X is assigned with “green” with frequency $\frac{1}{6}$.

The *distribution function* is a map that assigns the frequencies to possible outcomes. In this example, if we write λ for the distribution function, we get $\lambda(\text{blue}) = \frac{1}{2}$, $\lambda(\text{red}) = \frac{1}{3}$, and $\lambda(\text{green}) = \frac{1}{6}$. Thus, we say X has the distribution λ, and we write $\mathbb{P}(X = \text{blue}) = \frac{1}{2}$, $\mathbb{P}(X = \text{red}) = \frac{1}{3}$, and $\mathbb{P}(X = \text{green}) = \frac{1}{6}$.

Let X be a random variable associated with the sample space S and the distribution function ω. It is clear from the definition that $\sum_{x \in S} \omega(x) = 1$, and $\omega(x) \geq 0$ for all $x \in S$. In formal language, any function satisfying these two properties represents a distribution function. Any subset $E \subseteq S$ is called as an *event.* For an event E, we write $\mathbb{P}(E)$ for $\sum_{x \in E} \omega(x)$, and call it as the *probability* of the event E.

Our examples at the beginning were *discrete* random variables, in other words, the sample space was countable. On the other hand, for example, we can examine a random variable Y that can take any real number between 0 and 1. Clearly, for any particular point $y \in (0, 1)$, we get $\mathbb{P}(Y = y) = 0$. However, we need to distinguish elements in the sample space. Therefore, we are required to express the distribution in another way. In simplest words, a random variable is called as *continuous* if the sample space uncountable.

Accordingly, the distribution function ω of a continuous random variable associated with the sample space S satisfies $\int_S \omega(x)dx = 1$ and $\omega(x) \geq 0$ for all $x \in S$. Similarly, the probability of an event E is defined as $\int_E \omega(x)dx$. For a real-valued random variable X, a *density function* $f(x)$ is a real-valued function satisfying

$$\mathbb{P}(u \leq X \leq v) = \int_u^v f(x)dx \text{ for all } u, v \in \mathbb{R}.$$

Let X be a real-valued random variable associated with the sample space S. If X is discrete, let us write ω for its distribution function, and if it is continuous say $f(x)$ is its density function. The *expectation* of X is defined as:

i. $\mathbb{E}[X] := \sum_{x \in S} x \cdot \omega(x)$ if X is discrete, and
ii. $\mathbb{E}[X] := \int_{x \in S} x \cdot f(x)$ if X is continuous.

The *variance* of X is defined as $\mathrm{Var}[X] := \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ in both cases.

Now, assume we somehow learned that an event E has occurred. Of course, this affect the probability and the expectation of X. In such cases, we say the random variable X is *conditioned on* E. For another event E_1, the probability that is changed with respect to E is said to be *conditional probability of* E_1 *given* E, and denoted by $\mathbb{P}(E_1|E)$. Accordingly, the *conditional expectation* can be defined and denoted by $\mathbb{E}[X|E]$.

For instance, let X be a random variable that is determined by the consequence of a fair die. It is clear that $\mathbb{P}(X = j) = \frac{1}{6}$ for all $j \in \{1, 2, 3, 4, 5, 6\}$, and it can be calculated that $\mathbb{E}[X] = \frac{7}{2}$. However, assume it is given that X is even. Thus, we get $\mathbb{P}(X = j|X \text{ is even}) = \begin{cases} \frac{1}{3}, & \text{if } j \in \{2, 4, 6\} \\ 0, & \text{if } j \in \{1, 3, 5\} \end{cases}$ and then $\mathbb{E}[X|X \text{ is even}] = 4$.

Formally, for two events E_1 and E_2, the conditional probability is given by

$$\mathbb{P}(E_1|E_2) = \frac{\mathbb{P}(E_1 \cap E_2)}{\mathbb{P}(E_2)}.$$

Accordingly, the conditional expectation can be given as:

$$\mathbb{E}[X|E] := \sum_{x \in S} x \cdot \mathbb{P}(X = x|E).$$

Two random variables X and Y are called *independent* if

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B),$$

for all sets A and B. On the other hand, they are called *identically distributed* if $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$ for all sets A. They are said to be *i.i.d.* shortly, if they are independent and identically distributed.

For a set A, the *indicator* function of A is defined as

$$\mathbb{1}_A := \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

For an event E, $\mathbb{1}_E$ becomes a random variable whose expectation is equal to $\mathbb{P}(E)$.

Finally, we discuss two properties of the expectation. Let X and Y be random variables with finite expectation. Then, we have

i. $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, and
ii. $\mathbb{E}[cX] = c \cdot \mathbb{E}[X]$ for any constant c.

These two properties are called as the *linearity of the expectation*, and we frequently use this term throughout the thesis. On the other hand, if X is a non-negative and integer valued random variable, then we can calculate the expectation of X from the formula

$$\mathbb{E}[X] = \sum_{j=0}^{\infty} \mathbb{P}(X > j).$$

This formula is known as *tail sum formula.*

2.1.1. Special Distributions

Here, we give the definitions of some well-known and widely-used special distribution. Firstly, let X be a discrete random variable with the distribution function ω and sample space S. We say X has *uniform distribution on S* if $\omega(x) = \omega(y)$ for all $x, y \in S$.

Let Y be a real-valued continuous random variable with density function f. We say Y has uniform distribution on $[a,b]$ if f is given by

$$f(y) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \leq y \leq b, \\ 0, & \text{otherwise,} \end{cases}$$

and denoted by $Y \sim U(a,b)$.

We say a random variable B has *binomial distribution with parameters n and p*, and denoted by $B \sim Bin(n,p)$, if B is the number of successes in a sequence of n trials where each trial results in the success with probability p.

A continuous random variable Z whose density function is f is said to be

i. *exponentially distributed with parameter λ* if f is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0, \end{cases}$$

ii. *normally distributed with parameters μ and σ* if f is given by

$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}.$$

Finally, we say a discrete random variable L taking only non-negative integers has *Poisson distribution with parameter λ*, and denoted by $L \sim Poi(\lambda)$, if

$$\mathbb{P}(L = k) = \frac{\lambda^k e^{-k}}{k!} \text{ for all } k \in \mathbb{N}_0.$$

Poisson distribution is generally used for modeling the number of events for time intervals such as incoming customers in a restaurant. It is well-known that the interarrival times between consecutive events are independent and exponentially distributed.

2.1.2. Inequalities

In this section, we give two well-known inequalities which are used in the thesis.

Markov Inequality: Let X be a non-negative random variable with finite expectation. Then, we have

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a} \text{ for any } a > 0.$$

Union Bound Inequality: For a countable set of events E_1, E_2, ..., we have

$$\mathbb{P}\big(\bigcup_i E_i\big) \leq \sum_i \mathbb{P}(E_i).$$

2.1.3. Order of Magnitude

Throughout the thesis, we sometimes need to compare two functions. Let f and g are two functions on natural numbers. Even though these notations may be expressed differently in the general case, we used the following ones for simplicity.

We say $f(n) = o(g(n))$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

We say $f(n) = O(g(n))$ if there exist constants c and K such that

$$f(n) \leq c \cdot g(n) \text{ for all } x \geq K.$$

Finally, we say $f(n) = \Theta(g(n))$ if there exist constants c_1 and c_2 so that

$$c_1 g(n) \leq f(n) \leq c_2 g(n),$$

when n is sufficiently large.

2.2. Basic Graph Theory

In this section, we give the basic definitions and properties about graph theory which are used throughout the thesis. All concepts in this section are quite well-known and they can be found in several graph theory books such as [7] and [8].

2.2.1. Definitions

Let V be a finite set, and E be a subset of $V \times V$. We say G is a *graph* on the vertex set V with the edge set E by representing the elements of V with dots and connecting two dots $i, j \in V$ whenever $(i, j) \in E$. In general, we denote a graph G as $G = (V, E)$. If the vertex and edge sets are not specified at the beginning, we write $V(G)$ and $E(G)$ for corresponding sets. The *size* of the graph G, denoted by $|G|$, is defined as the number of elements in the set $V(G)$.

Two vertices $u, v \in V(G)$ are called as *adjacent* or *neighbor* if $(u, v) \in E(G)$. We sometimes denote the edge (u, v) by uv shortly, and an edge $uv \in E(G)$ is said to be *incident* to the vertices u and v. The *degree* of a vertex $v \in V(G)$ is defined as the number of neighbors in G. A vertex of degree zero is called as an *isolated* vertex. We say the edge $(i, i) \in E(G)$ for some $i \in V(G)$ is a *self-loop*, and we call G as a *simple graph* if it has no self-loops. If $E(G)$ is a subset containing unordered (resp. ordered) pairs from $V(G)$, we say G is an *undirected* (resp. *directed*) graph. Unless otherwise stated, we will consider only simple and undirected graphs throughout the thesis. For simplicity in the notation, we generally write uv instead of (u, v) for the edges in undirected graphs.

A graph H is called a *subgraph* of G if $V(H) \subseteq V(G)$ and $uv \in E(H)$ for $u, v \in V(H)$ whenever $uv \in E(G)$. If H is a graph that can be obtained by deleting some vertices and edges from G, then we say H is a *partial subgraph* of G. For a vertex u, we write $G - u$ for the subgraph of G whose vertices are $V(G) - \{u\}$.

A tuple $(v_1, v_2, ..., v_k)$ is called a *path* between v_1 and v_k if $v_j v_{j+1} \in E(G)$ for all $j \in \{1, 2, ..., k-1\}$. Similarly, a tuple $(v_1, v_2, ..., v_k)$ is called a *cycle* if $v_j v_{j+1} \in E(G)$ for all $j \in \{1, 2, ..., k-1\}$ and $v_k v_1 \in E(G)$. A graph G is said to be *connected* if there is path between u and v for all $u, v \in V(G)$. If G is not connected, it is called a *disconnected* graph. A connected subgraph of G is said to be a *connected component* if any additional vertex from G into this subgraph leads to have a disconnected subgraph.

For given two graphs G_1 and G_2 on disjoint vertex sets, their *disjoint union* is defined as the graph on the vertex set $V(G_1) \cup V(G_2)$ with the edge set $E(G_1) \cup E(G_2)$. Conversely, their *join* is defined as the graph on the vertex set $V(G_1) \cup V(G_2)$ with the edge set $E(G_1) \cup E(G_2) \cup E$ where $E = \{(u, v) : u \in V(G_1), v \in V(G_2)\}$.

2.2.2. Special Graph Families

A graph G is called as *complete* if $uv \in E(G)$ for all $u, v \in V(G)$. Conversely, we say G is an *empty* graph if $uv \in E(G)$ for all $u, v \in V(G)$. We denote the complete graph and empty graph on n vertices by K_n and I_n, respectively. A graph admitting all vertices have the same degree is called as *regular*. A subgraph H of G called as a *clique* or an *independent set* if H itself forms a complete graph or an empty graph, respectively.

We call a graph as a *path graph*, and denote by P_k, if P_k itself forms a path on k vertices and there are no edges other than the path itself. Similarly, we call a graph as a *cycle graph*, and denote by S_k, if S_k itself forms a cycle on k vertices and there are no edges other than the cycle itself.

We say G is a *bipartite* graph with *bipartition* $G = (A, B)$ where $A \cup B = V(G)$ and each of A and B forms an independent set. The *complete bipartite graph* $K_{a,b}$ is defined as the join of the independent sets I_a and I_b.

A graph G is called as a *tree* if it is connected and it has no cycles. In a tree G, we say $u \in V(G)$ is a *leaf* if its degree is exactly one. For a directed edge (u, v) in the tree, we call u as an *ancestor* of v.

For a given graph G, a partial subgraph T is called as a *spanning tree* if it forms a tree and $|T| = |G|$. It is clear that every connected graph should have at least one spanning tree. A graph whose all connected components are trees is called as a *forest.* Similarly, a partial subgraph F is called as a *spanning forest* if it forms a forest and $|F| = |G|$. Again, it is clear that every graph has at least one spanning forest.

2.2.3. Basic Properties of Graphs

In this subsection, we give some facts about graphs without proofs since all of them are either trivial or they can be seen directly. We state them as propositions because they are frequently used throughout the thesis.

Proposition 2.1. *Let G be a graph with m edges. Then, we have $\sum\limits_{u \in V(G)} d_u = 2m$ where d_u denotes the degree of the vertex u. In particular, $\sum\limits_{u \in V(G)} d_u$ is always even.*

Proposition 2.2. *If G is a tree on n vertices, it has exactly $n - 1$ edges.*

Proposition 2.3. *If G is a tree on n vertices and $v \in V(G)$, then there is a sequence of vertices $(v_0, v_1, ..., v_{2n-2})$ of G such that each directed edge $xy \in E(G)$ appears exactly once in this sequence, i.e. there is a bijective map from the set of directed edges to $\{1, 2, ..., 2n - 2\}$ that maps $xy \mapsto t_{xy}$ so that $x = v_{t_{xy}-1}$ and $y = v_{t_{xy}}$.*

2.3. Markov Chains

A sequence $\{X_n\}_{n\in\mathbb{N}_0}$ is called as a *stochastic process* if each of X_n is a random variable. If this stochastic process has some additional properties, then it is said to be *Markov chain.* Throughout the thesis, our main investigation, namely cover times, is a problem studied on Markov chains. Therefore, we need to discuss some basic facts about Markov chains. Even though there exists also a continuous-time chains in the literature (see [9]), we only pay attention to the discrete-time models. The reader can look at the book [10] for further information about Markov chains.

Definition 2.4. *Let* S *be a finite set,* $P = (p_{ij} \geq 0 : i, j \in S)$ *be a matrix with* $\sum_{j\in S} p_{ij} = 1$ *for all* $i \in S$ *and* $\lambda = (\lambda_i \geq 0 : i \in S)$ *be a row vector with* $\sum_{i\in S} \lambda_i = 1$. *The stochastic process* $\{X_n\}_{n\in\mathbb{N}_0}$ *is called as a Markov chain with state space* S*, initial distribution* λ *and transition matrix* P *if the following hold:*

i. $\mathbb{P}(X_0 = i) = \lambda_i$ *for all* $i \in S$.

ii. $\mathbb{P}(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = \mathbb{P}(X_{n+1} = j|X_n = i) = p_{ij}$ *for all* $i_0, i_1, ..., i_{n-1}, i, j \in S$ *and* $n \in \mathbb{N}_0$.

For a meaningful examination, the cover time problem requires some technical assumptions on the Markov chain. As an abuse of the notation, we sometimes say the transition matrix has some property instead of saying corresponding Markov chain has that property, and vice versa. Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain with state space S and transition matrix P.

We say P (or the Markov chain) is *irreducible* if

$$\mathbb{P}(X_n = i \text{ for some } n \in \mathbb{N}|X_0 = j) > 0 \text{ for all } i, j \in S.$$

We say P (or the Markov chain) is *aperiodic* if

$$\mathbb{P}(X_n = i | X_0 = i) > 0 \text{ for all } i \in S \text{ and for all sufficiently large } n \in \mathbb{N}.$$

Let λ be a probability distribution on S. We say λ is *invariant* for P (or the Markov chain) if $\lambda P = \lambda$.

The following theorem summarizes the technical condition that guarantees to have a *nice* background for further investigations.

Theorem 2.5. *Let P be an irreducible and aperiodic transition matrix, and suppose that P has an invariant distribution π. Let $\{X_n\}_{n \in \mathbb{N}_0}$ be a Markov chain with state space S, transition matrix P and any initial distribution. Then π is unique and*

$$\mathbb{P}(X_n = j) \longrightarrow \pi_j \text{ as } n \longrightarrow \infty \text{ for all } j \in S.$$

In particular, π is called the stationary distribution of the Markov chain.

2.4. Simple Random Walks on Graphs

Consider a finite Markov chain with transition matrix P such that the next move is purely arbitrary among all possible choices for each state. More formally, we assume that all nonzero entries are equal in each row of P. Moreover, for all states $i, j \in S$, we assume that if the state j is an option for the state i then the state i is an option for the state j, too.

It is clear that, under these assumptions, the Markov chain can be represented by a random walk that traverses on vertices of an undirected graph by using adjacency relations. Further, we study on simple graphs in general, which means diagonal entries of the transition matrix are assumed to be zero.

Moreover, the irreducibility of the Markov chain corresponds to the connectivity of the graph. Therefore, we will study on connected graphs throughout the thesis unless otherwise is stated.

Let G be a simple undirected graph on n vertices. We define the random walk $\{X_n\}_{n\in\mathbb{N}_0}$ on the vertices of G where the transition probabilities are determined as follows for all vertices $u, v \in V(G)$:

$$\mathbb{P}(X_{n+1} = v|X_n = u) = \begin{cases} \dfrac{1}{d_u}, & \text{if } uv \in E(G), \\ 0, & \text{otherwise,} \end{cases}$$

where d_u denotes the degree of the vertex u.

This stochastic process is said to be *simple random walk* on the graph G. For notational convenience, for a random variable Y on the Markov chain, we will write $\mathbb{E}_u Y$ instead of $\mathbb{E}[Y|X_0 = u]$. Similarly, for an event Z, we write $\mathbb{P}_u Z$ instead of $\mathbb{P}(Z|X_0 = u)$. Moreover, we will write $\mathbb{E}_\pi Y$ and $\mathbb{P}_\pi Z$ if X_0 is determined with respect to the stationary distribution π.

3. CHAIN STATISTICS

In this chapter, first we give the formal definition of some chain statistics used throughout the thesis, and then discuss the some relations between them.

3.1. Basic Definitions

Let $\{X_n\}_{n\in\mathbb{N}_0}$ be a Markov chain with a state space S and transition matrix P where S is finite and P is irreducible and aperiodic. Also, let π be the unique stationary distribution of the chain.

Hitting time for the state j, denoted by T_j, is defined as

$$T_j := inf\{n \geq 0 : X_n = j\}.$$

Similarly, *expected return time for the state* j, denoted by m_j, is defined as

$$m_j := \mathbb{E}[T_j^+ | X_0 = j],$$

where $T_j^+ = inf\{n > 0 : X_n = j\}$. Not surprisingly, there is a strong relation between the expected return time and the stationary distribution.

In the long-run, π_j measures how often the chain visits the vertex j where m_j is the number of steps required between two consecutive visits to that vertex. Therefore, these two quantities become the reciprocal of each other. We state this fact as a theorem since it is frequently used throughout the thesis.

Theorem 3.1. *For a Markov chain with irreducible transition matrix P, finite state space S and invariant distribution π, the expected return time for a state is the reciprocal of the corresponding entry of the invariant distribution, that is*

$$m_j = \frac{1}{\pi_j} \text{ for all } j \in S.$$

Since we study on simple, undirected and connected graphs in general, let us manage our terminology into the graph theoretical language. Let G be a graph, and consider the simple random walk on its vertices. Recall $\mathbb{E}_u T_v$ means the expected time needed for the first hit to the vertex v if we start at the vertex u. Accordingly, *hitting time for the graph G*, denoted by $hit(G)$, is defined as

$$hit(G) := \max_{u,v \in V(G)} \mathbb{E}_u T_v.$$

Similarly, *commute time for the graph G*, denoted by $com(G)$, is defined as

$$com(G) := \max_{u,v \in V(G)} (\mathbb{E}_u T_v + \mathbb{E}_v T_u).$$

Finally, let C be the first time that all vertices of G have been visited by the random walk at least once. Then, the *cover time of the graph G*, denoted by $cov(G)$, is defined as

$$cov(G) := \max_{u \in V} \mathbb{E}_u C.$$

3.2. Random Target Lemma

Consider a Markov chain $\{X_n\}_{n\in\mathbb{N}_0}$, let us take a state u. If we start a random walk at u, then the following lemma asserts that the weighted sum of hitting times of other states does not depend on u. This can be interpreted as the time needed to reach a *random* state is the same for each starting point. In general, we use the ideas stated in the book [11] in this section.

Lemma 3.2. *Consider a Markov chain $\{X_n\}_{n\in\mathbb{N}_0}$ with state space V, irreducible transition matrix P and stationary distribution π, and let $u \in V$. Then,*

$$\sum_{w\in V} \mathbb{E}_u[T_w]\cdot \pi_w$$

does not depend on u.

Proof. Let $P = \{p_{ij} : i, j \in V\}$, and write $h_u = \sum\limits_{w\in V} \mathbb{E}_u[T_w]\cdot\pi_w$. We need to show that h is a constant vector. In the first two steps, we use the conditional expectation on the first move of the random walk. The main approach in the proof is to show $Ph = h$, which implies h being constant due to the irreducibility of the transition matrix.

Step 1: $\mathbb{E}_u[T_w] = 1 + \sum\limits_{v\in V} p_{uv}\cdot\mathbb{E}_v[T_w]$ for $w \neq u$.

For $w \neq u$, we consider the conditional expectation on X_1.

$$\begin{aligned}
\mathbb{E}_u[T_w] &= \sum_{v\in V} \mathbb{E}_u[T_w | X_1 = v]\cdot p_{uv} \\
&= \sum_{v\in V} (1 + \mathbb{E}_v[T_w])\cdot p_{uv} \\
&= \sum_{v\in V} p_{uv} + \sum_{v\in V} \mathbb{E}_v[T_w]\cdot p_{uv} \\
&= 1 + \sum_{v\in V} \mathbb{E}_v[T_w]\cdot p_{uv}.
\end{aligned}$$

Step 2: $\dfrac{1}{\pi_u} = 1 + \sum_{v \in V} p_{uv} \cdot \mathbb{E}_v[T_u]$.

From the Theorem 3.1, the expected return time for the state u is equal to $\dfrac{1}{\pi_u}$. On the other hand, the expected return time can be found by conditional expectation on X_1.

$$\begin{aligned} m_u &= \sum_{v \in V} p_{uv} \cdot \mathbb{E}_u[T_u^+ | X_1 = v] \\ &= \sum_{v \in V} p_{uv} \cdot (1 + \mathbb{E}_v[T_u]) \\ &= \sum_{v \in V} p_{uv} + \sum_{v \in V} \mathbb{E}_v[T_w] \cdot p_{uv} \\ &= 1 + \sum_{v \in V} \mathbb{E}_v[T_w] \cdot p_{uv}. \end{aligned}$$

Step 3: $Ph = h$.

We will show that $(Ph)_u = h_u$ for all $u \in V$. Note that

$$\begin{aligned} (Ph)_u = \sum_{v \in V} p_{uv} \cdot h_v &= \sum_{v \in V} p_{uv} \cdot \sum_{w \in V} \mathbb{E}_v[T_w]\pi_w \\ &= \sum_{v \in V} \sum_{w \in V} p_{uv} \cdot \mathbb{E}_v[T_w]\pi_w \\ &= \sum_{w \in V} \sum_{v \in V} p_{uv} \cdot \mathbb{E}_v[T_w]\pi_w. \end{aligned}$$

Therefore, we need to prove that

$$\sum_{w \in V} \sum_{v \in V} p_{uv} \cdot \mathbb{E}_v[T_w]\pi_w = \sum_{w \in V} \mathbb{E}_u[T_w] \cdot \pi_w.$$

Firstly, it is clear that we can write

$$\sum_{w\in V}\sum_{v\in V} p_{uv}\cdot \mathbb{E}_v[T_w]\pi_w = \sum_{w\neq u}\sum_{v\in V} p_{uv}\cdot \mathbb{E}_u[T_w]\pi_w + \sum_{v\in V} p_{uv}\cdot \mathbb{E}_v[T_u]\pi_u.$$

Thus, by using Step 1, we get $\sum_{w\neq u}\sum_{v\in V} p_{uv}\cdot \mathbb{E}_u[T_w]\pi_w = \sum_{w\neq u}(\mathbb{E}_u[T_w]-1)\pi_w$. Since $\mathbb{E}_u[T_u]=0$ from the definition and $\sum_{w\in V}\pi_w = 1$, we have

$$\begin{aligned}
\sum_{w\neq u}(\mathbb{E}_u[T_w]-1)\pi_w &= \sum_{w\in V}(\mathbb{E}_u[T_w]-1)\pi_w - (\mathbb{E}_u[T_u]-1)\pi_u \\
&= \sum_{w\in V}\mathbb{E}_u[T_w]\pi_w - \sum_{w\in V}\pi_w + \pi_u \\
&= \sum_{w\in V}\mathbb{E}_u[T_w]\pi_w - 1 + \pi_u.
\end{aligned}$$

On the other hand, by using Step 2, we have $\sum_{v\in V} p_{uv}\cdot \mathbb{E}_v[T_u] = \frac{1}{\pi_u} - 1$. Therefore, we obtain

$$\sum_{v\in V} p_{uv}\cdot \mathbb{E}_v[T_u]\pi_u = \left(\frac{1}{\pi_u}-1\right)\cdot \pi_u = 1-\pi_u.$$

As a result, we can write

$$\begin{aligned}
\sum_{w\in V}\sum_{v\in V} p_{uv}\cdot \mathbb{E}_v[T_w]\pi_w &= \sum_{w\neq u}\sum_{v\in V} p_{uv}\cdot \mathbb{E}_u[T_w]\pi_w + \sum_{v\in V} p_{uv}\cdot \mathbb{E}_v[T_u]\pi_u \\
&= \Big[\sum_{w\in V}\mathbb{E}_u[T_w]\pi_w - 1 + \pi_u\Big] + \Big[1-\pi_u\Big] \\
&= \sum_{w\in V}\mathbb{E}_u[T_w]\pi_w.
\end{aligned}$$

Step 4: If $Ps = s$ for some nonnegative vector s, then s is a constant vector.

If $s = 0$ then it is trivial and so assume at least one of the components of s is nonzero. Since V is finite, there exists $M > 0$ such that $s_u = M$ is maximal for some $u \in V$. If $p_{uv} > 0$ and $s_v < M$ for some $v \in V$, then we would have

$$s_u = p_{uv}s_v + \sum_{w \neq v} p_{uw}s_w < M,$$

which is a contradiction. Therefore, $p_{uv} > 0$ implies $s_v = M$. On the other hand irreducibility implies that there exists a sequence $u = z_0, z_1, z_2, ..., z_l = w$ for which $p_{z_i z_{i+1}} > 0$ for any $w \in V$. Repeating the same argument, we can conclude that $s_w = s_{z_{l-1}} = ... = s_{z_1} = s_{z_0} = s_u = M$, which completes the proof. $\square$

This lemma is used in the proof of the following theorem that gives an upper bound the hitting time of the graph in terms the hitting times of the vertices.

Theorem 3.3. *Let G be a connected graph, and π be the stationary distribution of the random walk on its vertices. Then, we have*

$$hit(G) \leq 2 \cdot \max_{w \in V(G)} \mathbb{E}_\pi[T_w].$$

Proof. Let $u, v \in V(G)$. Suppose the random walk starts at the vertex u, it goes to a vertex with respect to the stationary distribution, and then it traverses until reaching to the vertex v. Then we can write $\mathbb{E}_u[T_v] \leq \mathbb{E}_u[T_\pi] + \mathbb{E}_\pi[T_v]$. From the Lemma 3.2, we have $\mathbb{E}_u[T_\pi] = \mathbb{E}_\pi[T_\pi]$. Therefore, we can write

$$\mathbb{E}_u[T_v] \leq \mathbb{E}_\pi[T_\pi] + \mathbb{E}_\pi[T_v] \leq 2 \cdot \max_{w \in V(G)} \mathbb{E}_\pi[T_w].$$

Since it holds for any $u, v \in V(G)$, the result follows:

$$hit(G) = \max_{u,v \in V(G)} \mathbb{E}_u[T_v] \leq 2 \cdot \max_{w \in V(G)} \mathbb{E}_\pi[T_w]. \qquad \square$$

3.3. Matthew's Theorem

Let G be a connected graph and consider the random walk on its vertices. There is a strong relation between $hit(G)$ and $cov(G)$ since the cover time means all vertices are to be hitted by the random walk. The one direction is obvious since

$$\mathbb{E}_u T_v \leq \mathbb{E}_u C \leq cov(G),$$

for any $u, v \in V(G)$, so we get

$$hit(G) = \max_{u,v \in V(G)} \mathbb{E}_u T_v \leq cov(G).$$

For the reverse direction, Matthew constructed an inequality in [12].

Let us write $\mathcal{H}(n)$ for the n^{th} harmonic sum in the rest of the thesis, i.e. let $\mathcal{H}(n) = \sum_{j=1}^{n} \frac{1}{j}$ for any natural number n. We will give the proof in our words for the sake of completeness.

Theorem 3.4. *Let G be a graph on n vertices, and consider the random walk on its vertices. Then, we have $cov(G) \leq hit(G) \cdot \mathcal{H}(n)$.*

Proof. For the simplicity, let $V(G) = \{1, 2, ..., n\}$. Let us choose an arbitrary initial vertex $u \in V(G)$ and let σ be a uniform random permutation of $\{1, 2, ..., n\}$, that is chosen independently from the random walk.

Let t_k be the first time that the vertices $\{\sigma(1), \sigma(2), ..., \sigma(k)\}$ have all been visited, and write $t_0 = 0$. Moreover, let $l_k = X_{t_k}$ be the last vertex among $\{\sigma(1), \sigma(2), ..., \sigma(k)\}$ to be visited.

We claim $\mathbb{E}_u[t_k - t_{k-1}] \leq \dfrac{hit(G)}{k}$ for all $1 \leq k \leq n$. Firstly, let us show this claim completes the proof. Now, by using the linearity of the expectation, we get $\mathbb{E}_u[t_n] = \sum_{k=1}^{n} \mathbb{E}_u[t_k - t_{k-1}]$. Therefore, if the claim holds, then we get

$$\mathbb{E}_u[t_n] = \sum_{k=1}^{n} \mathbb{E}_u[t_k - t_{k-1}] \leq \sum_{k=1}^{n} \frac{hit(G)}{k} = hit(G) \cdot \sum_{k=1}^{n} \frac{1}{k}.$$

Thus, the result follows by taking the maximum over all $u \in V(G)$ in the following inequality since we know $cov(G) = \max_{u \in V(G)} \mathbb{E}_u C$:

$$\mathbb{E}_u C = \mathbb{E}_u[t_n] \leq hit(G) \cdot \sum_{k=1}^{n} \frac{1}{k} = hit(G) \cdot \mathcal{H}(n).$$

Let us prove the claim. First, let us examine the case $k = 1$. We can use the conditional expectation on whether $\sigma(1) = u$ or not:

$$\begin{aligned}
\mathbb{E}_u[t_1 - t_0] &= \mathbb{E}_u[T_{\sigma(1)}] \\
&= \mathbb{E}_u[T_{\sigma(1)} | \sigma(1) \neq u] \cdot \mathbb{P}(\sigma(1) \neq u) + \mathbb{E}_u[T_{\sigma(1)} | \sigma(1) = u] \cdot \mathbb{P}(\sigma(1) = u) \\
&= \sum_{v \neq u} \mathbb{E}_u[T_v] \cdot \mathbb{P}(\sigma(1) = v) + 0 \\
&\leq \sum_{v \neq u} hit(G) \cdot \mathbb{P}(\sigma(1) = v) \leq hit(G).
\end{aligned}$$

Then, the claim holds for $k = 1$. Let us consider the case $k = 2$. Now, we condition on the vertex l_2:

$$\begin{aligned}
\mathbb{E}_u[t_2 - t_1] &= \mathbb{E}_u[t_2 - t_1 | l_2 = \sigma(1)] \cdot \mathbb{P}(l_2 = \sigma(1)) + \mathbb{E}_u[t_2 - t_1 | l_2 = \sigma(2)] \cdot \mathbb{P}(l_2 = \sigma(2)) \\
&= \mathbb{E}_u[t_2 - t_1 | l_2 = \sigma(2)] \cdot \mathbb{P}(l_2 = \sigma(2)) \text{ since } l_2 = \sigma(1) \text{ implies } t_2 = t_1.
\end{aligned}$$

Moreover $\mathbb{P}(l_2 = \sigma(1)) = \mathbb{P}(l_2 = \sigma(2)) = \dfrac{1}{2}$ since σ was chosen independently and uniformly from the chain. Let q_{rs} be the probability of the event $\{\sigma(1) = r, \sigma(2) = s\}$ for $r \neq s$ when $l_2 = \sigma(2)$ is given. Then, we can write

$$\begin{aligned}
\mathbb{E}_u[t_2 - t_1 | l_2 = \sigma(2)] &= \sum_{r \neq s} \mathbb{E}_u[t_2 - t_1 | l_2 = \sigma(2), \sigma(1) = r, \sigma(2) = s] \cdot q_{rs} \\
&= \sum_{r \neq s} \mathbb{E}_u[t_2 - t_1 | l_2 = \sigma(2), X_{t_1} = r, X_{t_2} = s] \cdot q_{rs} \\
&= \sum_{r \neq s} \mathbb{E}_r[T_s] \cdot q_{rs} \\
&\leq \sum_{r \neq s} hit(G) \cdot q_{rs} \leq hit(G).
\end{aligned}$$

Finally, let us assume $3 \leq k \leq n$. We calculate $\mathbb{E}_u[t_k - t_{k-1}]$ by conditioning whether $l_k = \sigma(k)$ or not. Note that $\mathbb{P}(l_k = \sigma(k)) = \dfrac{1}{k}$ since σ was chosen independently and uniformly from the chain. Therefore, we have

$$\begin{aligned}
\mathbb{E}_u[t_k - t_{k-1}] &= \frac{\mathbb{E}_u[t_k - t_{k-1} | l_k = \sigma(k)]}{k} + \frac{(k-1)\mathbb{E}_u[t_k - t_{k-1} | l_k \neq \sigma(k)]}{k} \\
&= \frac{\mathbb{E}_u[t_k - t_{k-1} | l_k = \sigma(k)]}{k} \text{ since } l_k \neq \sigma(k) \text{ implies } t_k = t_{k-1}.
\end{aligned}$$

Similarly, let q_{rs} be the probability of the event $\{\sigma(k-1) = r, \sigma(k) = s\}$ for $r \neq s$ when $l_k = \sigma(k)$ is given. Then, we have

$$\begin{aligned}
\mathbb{E}_u[t_k - t_{k-1} | l_k = \sigma(k)] &= \sum_{r \neq s} \mathbb{E}_u[t_k - t_{k-1} | l_k = \sigma(k), \sigma(k-1) = r, \sigma(k) = s] \cdot q_{rs} \\
&= \sum_{r \neq s} \mathbb{E}_u[t_k - t_{k-1} | l_k = \sigma(k), X_{t_{k-1}} = r, X_{t_k} = s] \cdot q_{rs} \\
&= \sum_{s \neq r} \sum_{r=1}^{n} \mathbb{P}(X_{t_{k-1}} = r) \mathbb{E}_r[T_s] \cdot q_{rs} \\
&\leq \sum_{s \neq r} \sum_{r=1}^{n} \mathbb{P}(X_{t_{k-1}} = r) \cdot hit(G) \cdot q_{rs} \\
&\leq \sum_{s \neq r} hit(G) \cdot q_{rs} \leq hit(G),
\end{aligned}$$

which completes the proof. $\square$

In the proof of Theorem 3.4, *hit*(G) is used as an upper bound for $\mathbb{E}_u[T_v]$ for any $u, v \in V(G), u \neq v$. On the other hand, consider a subset of $V(G)$ so that elements in this subset are far away from each other in terms of hitting time.

It is clear that we need to cover this subset in order to cover all vertices, therefore the cover time for the subset can give a useful lower bound. By using almost the same procedure in Theorem 3.4 can be applied to obtain a lower bound for *cov*(G) in terms of a *good* subset.

Theorem 3.5. *Let G be a graph on n vertices, and consider the random walk on its vertices. Take an arbitrary subset $A \subset V(G)$ with $|A| = k$. Then, we have*

$$cov(G) \geq \mathcal{H}(k-1) \cdot \min_{u,v \in A, u \neq v} \mathbb{E}_u[T_v].$$

Proof. Let $a \in A$ and write $t_{min} = \min_{u,v \in A, u \neq v} \mathbb{E}_u[T_v]$. Note that *cov*$(G) \geq \mathbb{E}_a C_A$ where C_A is the time that all states of A have been visited by the chain at least once.

As similar to the proof of Theorem 3.4, let σ be a uniform random permutation of $\{1, 2, ..., k\}$, that is chosen independently from the chain. Moreover, let t_j be the first time that the states $\{\sigma(1), \sigma(2), ..., \sigma(j)\}$ have all been visited, and write $t_0 = 0$. Hence, by using the linearity of the expectation, we can obtain

$$\mathbb{E}_a C_A = \mathbb{E}_a[t_k] = \sum_{j=1}^{k} \mathbb{E}_a[t_j - t_{j-1}].$$

Note that $\mathbb{E}_a[t_1 - t_0] = \mathbb{E}_a[T_{\sigma(1)}]$. Moreover, observe that

$$\begin{aligned}
\mathbb{E}_a[t_1 - t_0] &= \mathbb{E}_a[T_{\sigma(1)}] \\
&= \mathbb{E}_a[T_{\sigma(1)}|\sigma(1) \neq a] \cdot \mathbb{P}(\sigma(1) \neq a) + \mathbb{E}_a[T_{\sigma(1)}|\sigma(1) = a] \cdot \mathbb{P}(\sigma(1) = a) \\
&= \mathbb{E}_a[T_{\sigma(1)}|\sigma(1) \neq a] \cdot \Big(1 - \frac{1}{k}\Big) + 0 \cdot \frac{1}{k} \\
&\geq t_{min} \cdot \Big(1 - \frac{1}{k}\Big).
\end{aligned}$$

Again, as similar to Theorem 3.4, we can conclude that $\mathbb{E}_a[t_j - t_{j-1}] \geq \dfrac{t_{min}}{j}$ for all $2 \leq j \leq k$. Hence, we get

$$\begin{aligned}
cov(G) &\geq \mathbb{E}_a C_A \\
&\geq t_{min} \cdot \Big(1 - \frac{1}{k}\Big) + \sum_{j=2}^{k} \frac{t_{min}}{j} \\
&= t_{min} \cdot \Big(1 - \frac{1}{k} + \sum_{j=2}^{k} \frac{1}{j}\Big) \\
&= \mathcal{H}(k-1) \cdot t_{min},
\end{aligned}$$

so the result follows. □

4. EXACT CALCULATIONS OF COVER TIME

It is unknown whether the cover time can be calculated deterministically in a polynomial time (see [11]). In this chapter, we discuss the special cases so that the exact value of the cover time can be found by simple probabilistic and combinatorial arguments. For the sake of completeness, we will give the detailed proofs of those cases in our words.

We will start with the calculation of the complete graph.

Theorem 4.1. *Let G be a complete graph of n vertices. Then, we have*

$$cov(G) = (n-1) \cdot \mathcal{H}(n-1).$$

Proof. Let C_r be the time when the random walk has first visited r distinct vertices. Clearly, $C_1 = 0$, $C_2 = 1$ and we can write $C_n = \sum_{r=1}^{n-1} \big(C_{r+1} - C_r\big)$. Then by using the linearity of the expectation, we have $cov(G) = \mathbb{E}[C_n] = \sum_{r=1}^{n-1} \mathbb{E}[C_{r+1} - C_r]$.

At the time the random walk has first visited r distinct vertices, the probability of reaching $(r+1)^{th}$ new vertex in the next move is $\frac{n-r}{n-1}$, so we get

$$\mathbb{P}(C_{r+1} - C_r = 1) = \frac{n-r}{n-1}.$$

Similarly, we reach the $(r+1)^{th}$ new vertex at the step k if the random walk traverses already visited r vertices in the first $k-1$ steps, therefore we can write

$$\mathbb{P}(C_{r+1} - C_r = k) = \frac{n-r}{n-1} \cdot \left(\frac{r-1}{n-1}\right)^{k-1},$$

for any $k \in \mathbb{N}$ and $2 \leq r \leq n-1$. Hence we get

$$\begin{aligned}
\mathbb{E}[C_{r+1} - C_r] &= \sum_{k=1}^{\infty} k \cdot \mathbb{P}(C_{r+1} - C_r = k) \\
&= \sum_{k=1}^{\infty} k \frac{n-r}{n-1} \Big(\frac{r-1}{n-1}\Big)^{k-1} \\
&= \frac{n-r}{n-1} \sum_{k=1}^{\infty} k \Big(\frac{r-1}{n-1}\Big)^{k-1}.
\end{aligned}$$

Now, for simplicity, let us write $\rho = \dfrac{r-1}{n-1}$. Then, it is clear that $\dfrac{n-r}{n-1} = 1 - \rho$ and we get $\mathbb{E}[C_{r+1} - C_r] = (1-\rho) \sum_{k=1}^{\infty} k\rho^{k-1}$.

Note that $\sum_{k=1}^{\infty} k\rho^{k-1}$ is equal to the derivative of $\sum_{k=0}^{\infty} \rho^k = \dfrac{1}{1-\rho}$ with respect to ρ, which equals $\dfrac{1}{(1-\rho)^2}$, so we have

$$\mathbb{E}[C_{r+1} - C_r] = (1-\rho)\Big(\frac{1}{1-\rho}\Big)' = (1-\rho)\frac{1}{(1-\rho)^2} = \frac{1}{1-\rho} = \frac{1}{1-\frac{r-1}{n-1}} = \frac{n-1}{n-r}.$$

As a result, we get

$$\begin{aligned}
cov(G) = \mathbb{E}[C_n] &= \mathbb{E}[C_2 - C_1] + \sum_{r=2}^{n-1} \frac{n-1}{n-r} \\
&= 1 + (n-1) \sum_{r=2}^{n-1} \frac{1}{n-r} \\
&= 1 + (n-1) \cdot \mathcal{H}(n-2) \\
&= (n-1) \cdot \mathcal{H}(n-1),
\end{aligned}$$

as claimed. $\square$

Next, consider a game in which we pick randomly choose one of available n different coupons in each turn and we are waiting for the completion of the collection. This problem is well-studied in the literature and known as *coupon collector problem* (see [1]). The expected number of turns needed for reaching the goal is pretty related with the cover time concept. The game is equivalent to the covering all vertices of a complete graph where each vertex has an additional self-loop.

Theorem 4.2. *Let G be a complete graph on n vertices with addition a self-loop at each vertex. Then, we have*

$$cov(G) = n \cdot \mathcal{H}(n-1).$$

Proof. The proof can be completed by applying the same procedure as in the Theorem 4.1. Similarly, we define C_r, but in this case we have $C_1 = 0$ and

$$\mathbb{P}(C_{r+1} - C_r = k) = \frac{n-r}{n}\Big(\frac{r}{n}\Big)^{k-1},$$

for any $k \in \mathbb{N}$ and $1 \leq r \leq n-1$. Again, by using the linearity of the expectation and the formula for geometric series, we can write

$$\mathbb{E}[C_{r+1} - C_r] = \frac{n-r}{n}\sum_{k=1}^{\infty} k\Big(\frac{r}{n}\Big)^{k-1} = \frac{n-r}{n}\frac{1}{(1-\frac{r}{n})^2} = \frac{n}{n-r}.$$

Hence, the result easily follows:

$$\mathbb{E}[C_n] = \sum_{r=1}^{n-1}\mathbb{E}[C_{r+1} - C_r] = \sum_{r=1}^{n-1}\frac{n}{n-r} = n\sum_{r=1}^{n-1}\frac{1}{n-r} = n \cdot \mathcal{H}(n-1). \qquad \square$$

Now, we need to state a famous probability problem, known as *gambler's ruin problem* in the literature (see [13]), since we will use it frequently in the rest of the thesis. Here, we give the proof of the most classical version as a proposition.

Proposition 4.3. *Consider a gamble that ends when either the gambler has* 0 *dollar (he loses the game) or* N *dollars (he wins the game). Starting with* n *dollars where* $1 \leq n \leq N-1$*, in each turn, the gambler wins or loses* 1 *dollar with equal probability. Then, the expected number of turns needed for the end of the gamble is* $n(N-n)$*. Moreover, the probability of winning is* $\frac{n}{N}$*.*

Proof. Let E_n be the expected number of turns needed for the end of the gamble, and P_n be the probability of winning when having n dollars at the beginning for $0 \leq n \leq N$. It is clear from the definitions that $E_0 = E_N = 0$, $P_0 = 0$, $P_N = 1$. Moreover, the following recursions can be obtained by conditioning on the result of the first turn:

$$\begin{aligned} E_n &= 1 + \frac{1}{2}E_{n-1} + \frac{1}{2}E_{n+1}, \text{ and} \\ P_n &= \frac{1}{2}P_{n-1} + \frac{1}{2}P_{n+1}, \end{aligned}$$

for $1 \leq n \leq N-1$. Now, by adding up the first recursion from $n = 1$ to $n = N-1$, we have

$$\sum_{n=1}^{N-1} E_n = (N-1) + \frac{E_0 + E_1 + E_{N-1} + E_N}{2} + \sum_{n=2}^{N-2} E_n.$$

After the cancellations and by using $E_0 = E_N = 0$, we get $E_1 + E_{N-1} = 2N - 2$. Moreover, by the symmetry, $E_n = E_{N-n}$ for each n, which leads $E_1 = E_{N-1} = N-1$.

Now, we have $E_0 = 0$, $E_1 = N-1$ and $E_{n+1} = 2E_n - E_{n-1} - 2$ for $1 \leq n \leq N-1$. Hence, we can prove our claim by strong induction on n. The claim holds for $n = 0$ and $n = 1$, and assume $E_n = n(N-n)$ for $n \leq k$ for some $k \geq 1$. Then, we get

$$
\begin{aligned}
E_{k+1} &= 2E_k - E_{k-1} - 2 \\
&= 2k(N-k) - (k-1)(N-k+1) - 2 \\
&= 2kN - 2k^2 - kN + N + k^2 - 2k + 1 - 2 \\
&= kN - k^2 - 2k + N - 1 \\
&= (k+1)N - (k+1)^2 \\
&= (k+1)(N-k-1),
\end{aligned}
$$

and so the result follows.

On the other hand, we have $P_2 = 2P_1$ for $n = 1$ from the second recursion since $P_0 = 0$. If we write the same recursion for $n = 2$, we get $P_3 = 2P_2 - P_1 = 3P_1$ by using $P_2 = 2P_1$. Similarly, it can be easily seen that we have $P_k = kP_1$ for $k \geq 1$ inductively, which implies $P_N = NP_1$ and so $P_1 = \dfrac{1}{N}$. Therefore, we get $P_n = \dfrac{n}{N}$ for all $1 \leq n \leq N$, and the result follows. $\square$

Hence, we give the result for star graphs, again which is quite similar to the complete graph case.

Theorem 4.4. *Let G be a star with n vertices where $u \in V$ is the center and $v \in V$ is a leaf. Then, we have $\mathbb{E}_u C = 2(n-1) \cdot \mathcal{H}(n-2)$ and $\mathbb{E}_v C = 2(n-1) \cdot \mathcal{H}(n-2) - 1$, therefore we get*

$$cov(G) = 2(n-1) \cdot \mathcal{H}(n-2).$$

Proof. If we start the random walk at the center u, then visits to the leaves (in every second step) have exactly same distribution as in the coupon collector problem with $n-1$ distinct coupons since we pick one of the leaves uniformly in each step. From Theorem 4.2, we have $\mathbb{E}_u C = 2(n-1) \cdot \mathcal{H}(n-2)$.

If we start at a leaf v, then $\mathbb{E}_u C = 2(n-1)\cdot\mathcal{H}(n-2)-1$ since we do not need one more visit to the center u in the last step. Hence, we get $cov(G) = 2(n-1)\cdot\mathcal{H}(n-2)$ and the result follows. □

Now, we discuss the cover time of another tree, namely the path graph.

Theorem 4.5. *Let G be a path with n vertices where $V(G) = \{1, 2, ..., n\}$ from left to the right. Then, we have $\mathbb{E}_i C = (i-1)(n-i) + (n-1)^2$ for $i \in V$, therefore we get*

$$cov(G) = \left\lfloor \frac{5(n-1)^2}{4} \right\rfloor.$$

Proof. Suppose we start the random walk at the vertex 1 and let us write $h_i = \mathbb{E}_i T_n$ for $1 \le i \le n$. Since hitting the vertex n is equivalent to the covering all vertices, we have $\mathbb{E}_1 C = h_1$. Moreover, clearly $h_n = 0$ from the definition, and $h_1 = 1 + h_2$ since we will be at the vertex 2 after the first step when we start at the vertex 1.

On the other hand, for any $2 \le k \le n-1$, if we are at the vertex k, then by conditioning on the next step, we have the following recursion:

$$h_k = 1 + \frac{1}{2}h_{k-1} + \frac{1}{2}h_{k+1},$$

for $2 \le k \le n-1$. We claim $h_i = h_1 - (i-1)^2$ for $2 \le i \le n$ and we will prove this by strong induction. We know this equality holds for $i = 2$. At the induction step, assume $h_k = h_1 - (k-1)^2$ for some $2 \le k \le n-1$. From the recurrence relation, we have

$$h_1 - (k-1)^2 = 1 + \frac{h_1 - (k-2)^2}{2} + \frac{h_{k+1}}{2}.$$

Then, we get

$$\begin{aligned} h_{k+1} &= 2h_1 - 2(k-1)^2 - 2 - h_1 + (k-2)^2 \\ &= h_1 - 2k^2 + 4k - 2 - 2 + k^2 - 4k + 4 \\ &= h_1 - k^2, \end{aligned}$$

and the claim follows.

Thus, since $h_n = 0$, we get $h_1 - (n-1)^2 = 0$ from the claim, and therefore we get $\mathbb{E}_1 C = h_1 = (n-1)^2$. By the symmetry, we can also conclude that $\mathbb{E}_n C = (n-1)^2$.

On the other hand, suppose we start at the vertex i for some $2 \leq i \leq n-1$. Observe that we must first either hit to the vertex 1 or vertex n to cover all vertices. Therefore, $E_i C$ equals to the summation of the expected time needed for the first hit to one of the endpoints and $E_1 C$ (or $E_n C$). However, the first hit to the one of those endpoints is equivalent to the gambler's ruin problem: we have i dollars at initial and the game ends when we reach either 1 dollar or n dollars. Therefore, the expected time needed for the first hit is $(i-1)(n-1-i+1) = (i-1)(n-i)$ from Proposition 4.3, and so we get $\mathbb{E}_i C = (i-1)(n-i) + (n-1)^2$ for $2 \leq i \leq n-1$.

Finally, we need to choose $i = \left\lfloor \frac{n+1}{2} \right\rfloor$ to maximize the quantity $(i-1)(n-i)$, therefore we get $cov(G) = \left\lfloor \frac{5(n-1)^2}{4} \right\rfloor$. □

Remark 4.6. *Let n be a natural number. We remark that a star on n vertices has the minimum cover time over all trees of size n whereas a path on n vertices has the maximum cover time over all trees of size n. A detailed explanation will be given in Chapter 6.*

We close this chapter by giving the cover time of cycles. In the proof, again the gambler's ruin problem will be used.

Theorem 4.7. *Let G be a cycle on n vertices. Then, we have $cov(G) = \dfrac{n(n-1)}{2}$.*

Proof. Again, let us write C_r as in the Theorem 4.1. At the time we have exactly r distinct states, already visited vertices form a segment and reaching the $(r+1)^{th}$ new vertex is again equivalent to the gambler's ruin problem: the current position of the random walk corresponds to having 1 dollar (or r dollars) and we are waiting for him to reach either 0 or $r+1$ dollars.

From the Proposition 4.3, the expected time is $1 \cdot r = r$ and so we get $C_{r+1} = C_r + r$ for $r \geq 1$. Since $C_1 = 0$, it can be easily shown by induction that $C_n = \dfrac{n(n-1)}{2}$. $\square$

5. POLYNOMIAL BOUND FOR GENERAL GRAPHS

For a given connected graph G, it turns out that the calculation of $cov(G)$ is not an easy task with some exceptional cases which are discussed in Chapter 4. To our best knowledge, there are no further nontrivial graph families whose cover time can be exactly determined. When it is not possible to calculate a parameter exactly, the first thing to do would be to understand its order of magnitude. Therefore, we need to decide whether $cov(G)$ can be exponential or not.

In this chapter, we will discuss how a polynomial bound on the cover time for general graphs can be obtained. We use the ideas in [14] and give the proofs in detail. Firstly, we need a simple proposition for expected return time.

Proposition 5.1. *Let G be a connected graph with m edges. For any $u \in V(G)$, we have $m_u = \dfrac{2m}{d_u}$ where d_u denotes the degree of u.*

Proof. For the convenience, let us write $V(G) = \{1, 2, ..., n\}$. From the Theorem 3.1, we have $m_j = \dfrac{1}{\pi_j}$ for the invariant distribution π if the transition matrix is irreducible. Since G is connected, irreducibility is clear and we must show that $\pi_j = \dfrac{d_j}{2m}$ for any vertex $j \in V$. Moreover, aperiodicity of the transition matrix is trivial and hence an invariant distribution is the unique stationary distribution from the Theorem 2.5. Therefore we only need to check the equality $\pi P = \pi$ where $\pi_j = \dfrac{d_j}{2m}$. To satisfy the equality, we need $\displaystyle\sum_{i=1}^{n} \pi_i p_{ij} = \pi_j$. Indeed, we have

$$\sum_{i=1}^{n} \pi_i p_{ij} = \sum_{i:ij \in E} \frac{d_i}{2m}\frac{1}{d_i} + \sum_{i:ij \notin E} \frac{d_i}{2m} 0 = \frac{|i : ij \in E|}{2m} = \frac{d_j}{2m} = \pi_j,$$

and we are done. $\square$

We can interpret the Proposition 5.1 as follows: We will visit the vertex i with the frequency $\frac{d_i}{2m}$ in the long-run and by the symmetry we can say that each transition from i to j for any $ij \in E$ has the frequency $\frac{\frac{d_i}{2m}}{d_i} = \frac{1}{2m}$, i.e. each directed edge has the same long-run frequency $\frac{1}{2m}$. We state this fact as a corollary.

Corollary 5.2. *Let G be a graph with m edges. For a simple random walk on the vertices of G, any directed edge (u, v) has the same long-run frequency $\frac{1}{2m}$.*

Hence, by using Corollary 5.2, we get an upper bound for the quantity that is also known as *commute time.*

Proposition 5.3. *Let G be a graph with m edges. For any edge $uv \in E$, we have*

$$\mathbb{E}_u T_v + \mathbb{E}_v T_u \leq 2m.$$

Proof. Consider a long journey and observe the period that starts at u, visits to v at least once, and returns to u. We must show that the expected time needed for this observation period is at most $2m$. Since each directed edge has the same long-run frequency from Corollary 5.2, the expected time of the observation period is exactly equal to the number of directed edges times the expected number of occurrences of any particular directed edge.

On the other hand, the transition vu leads to the end of the observation period, so the expected number of occurrences of the edge vu is less than or equal to 1 with the equality if and only if $uv \in E$ is an cut edge of G, i.e. the removal of the edge uv makes G disconnected. Hence, we get $\mathbb{E}_u T_v + \mathbb{E}_v T_u \leq 2m \cdot 1$ and the result follows. □

Now, we are ready to obtain a polynomial upper bound for the cover time in a general graph by using a classical spanning tree argument.

Theorem 5.4. *Let G be a connected graph on n vertices with m edges. Then, we have*

$$cov(G) \leq 2m(n-1).$$

Proof. For any $u \in V$, we will show $\mathbb{E}_u C \leq 2m(n-1)$. Let H be a spanning tree of G. Then, there is a walk beginning and ending at u which traverses each edge of H exactly once in each direction from Proposition 2.3. The expected time for this walk is clearly larger than $\mathbb{E}_u C$ since we restrict our route while covering all vertices. Therefore, we can write

$$\begin{aligned}
\mathbb{E}_u C &\leq \sum_{xy \in H} (\mathbb{E}_x T_y + \mathbb{E}_y T_x) \\
&\leq \max_{xy \in H} (\mathbb{E}_x T_y + \mathbb{E}_y T_x) \cdot |E(H)| \\
&\leq \max_{xy \in H} (\mathbb{E}_x T_y + \mathbb{E}_y T_x) \cdot (n-1) \text{ from the Proposition 2.2} \\
&\leq 2m(n-1) \text{ from the Proposition 5.3}
\end{aligned}$$

and we are done. □

After finding a polynomial upper bound for the cover time of a general graph, we must question whether this bound is optimal up to order of magnitude.

For dense graphs, we know $m = \Theta(n^2)$ and so $2m(n-1) = \Theta(n^3)$. To conclude that this order is optimal in the general case, we need an example.

Example 5.5. *Let G be a graph with $3n$ vertices where $V(G) = \{1, 2, ..., 3n\}$. If we have only these adjacency relations, then $cov(G)$ has the order $\Theta(n^3)$.*

i. $\{1, 2, ..., n\}$ *forms a clique.*
ii. $\{2n+1, 2n+2, ..., 3n\}$ *forms a clique.*
iii. $\{1, n+1, n+2, ..., 2n-1, 2n, 3n\}$ *forms a path from left to the right.*

Proof. We will show that $\mathbb{E}_1 C$ has the order $\Theta(n^3)$. Firstly, let us only consider the left clique. Starting with the vertex 1, the expected return time is exactly n from the Theorem 3.1 since we have a uniform stationary distribution, i.e. $\pi_j = \dfrac{1}{n}$ for any $j \in \{1, 2, ..., n\}$. In each turn at the vertex 1, the probability of leaving the left clique, or explicitly probability of hitting the vertex $(n+1)$ is $\dfrac{1}{n}$. Hence, we will hit the vertex $(n+1)$ after the n^{th} visit to the vertex 1 in average, i.e.

$$\mathbb{E}_1 T_{n+1} = n(n-1).$$

On the other hand, the only way we can reach to the vertex $3n$ is to use the path $\{1, n+1, n+2, ..., 2n-1, 2n, 3n\}$. Hence, let us only consider this path and start at the vertex $(n+1)$. Again, we have the same situation as in the gambler's ruin problem: we have 1 dollar and the game will end either we have 0 dollar or $(n+1)$ dollars. The probability of winning the game is $\dfrac{1}{n+1}$ from the Proposition 4.3, and in the case of losing, we need $n(n-1)$ steps to return to the back the vertex $n+1$ from the vertex 1 in average. Therefore,

$$\begin{aligned} n(n-1) \cdot \sum_{k=1}^{\infty} k \Big(\frac{n}{n+1} \Big)^{k-1} \frac{1}{n+1} &= \frac{n(n-1)}{n+1} \Big(\frac{1}{1 - \frac{n}{n+1}} \Big)^2 \\ &= n(n-1)(n+1) \\ &= \Theta(n^3) \end{aligned}$$

steps are necessary to reach the vertex $3n$ in average. Obviously, we have

$$cov(G) \geq \mathbb{E}_1 C \geq \mathbb{E}_1 T_{3n},$$

and the result follows. □

6. VARIATIONS OF COVER TIMES FROM THE LITERATURE

In this chapter, we discuss several variations and extensions of the classical cover time problem from the literature.

6.1. Edge Covering

Firstly, the edge covering version was discussed from Zuckerman in 1991. Consider the simple random walk on a connected graph G on n vertices with m edges. Let C_{edge} be the first time that all edges of G have been traversed in both direction at least once.

In [15], the author proved that the order of the magnitude of the upper bound for the cover time in general graphs, $O(mn)$, would be the same if we wait up to time C_{edge}. Then, Aldous made the proof cleaner in his book [11] with Fill, and we will follow the same idea in that proof here.

Theorem 6.1. *Let G be a connected graph on n vertices with m edges. Then, we have*

$$\max_{u \in V(G)} \mathbb{E}_u C_{edge} \leq \frac{64mn}{3}.$$

Proof. Let us write $T = \lceil \sqrt{32m^3} \rceil$ and take a vertex $u \in V(G)$.

Let us define a *tour* as follows:

- The tour starts at the vertex u, and continue until all vertices have been visited.
- Then, wait for additional T steps.
- The tour ends at the time the random walk hits the vertex u.

Let Y_i be the time that i^{th} tour ends, and let N_{edge} be the number of tours required for all edges have been visited in each direction at least once. Note that Y_i's are i.i.d. random variables and we have $Y_{N_{edge}} = \min\{Y_i : Y_i \geq C_{edge}\}$. Thus, from Wald's identity, we get

$$\mathbb{E}_u C_{edge} \leq \mathbb{E}_u Y_{N_{edge}} = \mathbb{E}_u N_{edge} \cdot \mathbb{E}_u Y_1.$$

Firstly, clearly we can bound $\mathbb{E}_u Y_1$ in terms of T and $cov(G)$ as follows:

$$\mathbb{E}_u Y_1 \leq cov(G) + T + \max_{w \in V(G)} \mathbb{E}_w T_u \leq T + 2 \cdot cov(G).$$

To bound $\mathbb{E}_u N_{edge}$, we will use the tail sum formula. Note that we can write

$$\mathbb{E}_u N_{edge} = \sum_{n=0}^{\infty} \mathbb{P}_u(N_{edge} > n) \leq 2 \cdot \sum_{j=0}^{\infty} \mathbb{P}_u(N_{edge} > 2j),$$

since it is clear that $\mathbb{P}_u(N_{edge} > 2j) + \mathbb{P}_u(N_{edge} > 2j+1) \leq 2 \cdot \mathbb{P}_u(N_{edge} > 2j)$ for all $j \geq 0$. Now, we will show that $\mathbb{P}_u(N_{edge} > 2j) \leq \left(\frac{8m^3}{T^2}\right)^j$ for $j \geq 1$.

Take a fix directed edge (y, z). Note that the long-run frequency of this directed edge is $\frac{1}{2m}$ from Corollary 5.2, i.e. the expected time until (y, z) is traversed starting from z is equal to $2m$. Therefore, (y, z) is not traversed in T steps starting from z with probability at most $\frac{2m}{T}$ by Markov's inequality. Then, we can conclude the directed edge (y, z) is not traversed in a tour with probability at most $\frac{2m}{T}$. Hence, by using the union bound, we have

$$\begin{aligned}
\mathbb{P}_u(N_{edge} > l) &= \mathbb{P}_u(\text{At least one directed edge is not visited in first } l \text{ tours}) \\
&\leq \sum_{\text{all directed edges}} \mathbb{P}_u((y, z) \text{ is not visited in the first } l \text{ tours}) \\
&= \sum_{\text{all directed edges}} \mathbb{P}_u((y, z) \text{ is not visited in a tour})^l \\
&\leq 2m \cdot \left(\frac{2m}{T}\right)^l = \frac{(2m)^{l+1}}{T^l}.
\end{aligned}$$

Then, we get $\mathbb{P}_u(N_{edge} > 2j) \leq \dfrac{(2m)^{2j+1}}{T^{2j}} \leq \dfrac{(2m)^{3j}}{T^{2j}} \leq \left(\dfrac{8m^3}{T^2}\right)^j$ for $j \geq 1$. Thus, by using the summation in the infinite geometric series since $T^2 > 8m^3$, we have

$$\begin{aligned}
\mathbb{E}_u N_{edge} &\leq 2 \cdot \sum_{j=0}^{\infty} \mathbb{P}_u(N_{edge} > 2j) \\
&\leq 1 + \left(\frac{8m^3}{T^2}\right)^j \\
&= \frac{2}{1 - \dfrac{8m^3}{T^2}}.
\end{aligned}$$

Since $T \geq \sqrt{32m^3}$ from its definition, we get $\mathbb{E}_u N_{edge} \leq \dfrac{2}{1 - \dfrac{1}{4}} = \dfrac{8}{3}$. Therefore, we get $\mathbb{E}_u C_{edge} \leq \dfrac{8}{3} \cdot (\lceil \sqrt{32m^3} \rceil + 2 \cdot cov(G)) \leq \dfrac{8}{3} \cdot (\sqrt{32m^3} + 1 + 2 \cdot cov(G))$. Recall we know $cov(G) \leq 2m(n-1)$ from Theorem 5.4, which implies $1 + 2 \cdot cov(G) \leq 4mn$. As a result, by using $m \leq \dfrac{n}{\sqrt{2}}$, we have

$$\begin{aligned}
\mathbb{E}_u C_{edge} &\leq \frac{8}{3} \cdot (\sqrt{32m^3} + 4mn) \\
&= \frac{8}{3} \cdot (4m\sqrt{2}\sqrt{m} + 4mn) \\
&\leq \frac{8}{3} \cdot (4m\sqrt{2}\frac{n}{\sqrt{2}} + 4mn) \\
&= \frac{64mn}{3},
\end{aligned}$$

as desired. □

6.2. Multiple Walkers

As a second variation, cover time problem can be considered when there are more than one walker. Formally, consider l many independent random walk on a connected graph G. We may interest the expectation of the first time that all vertices have been visited at least once by at least one of those l random walks.

In [16], this problem appeared through a different motivation. Given an undirected graph G with its two vertices s and t, we want to decide whether s and t are in the same connected component. This decision can be made through a deterministic graph search algorithms, or alternatively it can be simulated via random walks.

Authors examined l many independent random walks of length $O(m^2/l^2 \log^3 n)$ starting with respect to stationary distribution (which is $\pi_j = \dfrac{d_j}{2m}$ from Proposition 5.1) to check whether s and t are in the same component through the covered vertices from each of those random walks. Here, we just state the theorem.

Theorem 6.2. *Let G be a connected graph on n vertices with m edges. Define C_l as the time needed for covering all vertices by l independent random walks where each of them starts at a vertex with respect to stationary distribution. Then, we have*

$$\mathbb{E}[C_l] = O\Big(\frac{m^2 \log^3 n}{l^2}\Big).$$

On the other hand, Aldous and Fill examined a similar variation of this problem on regular graphs in [11]. As contrary to the Theorem 6.2, they started each of those random walks at a vertex chosen uniformly.

Theorem 6.3. *Let G be a regular graph on n vertices. Consider a k independent random walks, each started at a uniform random vertex. Define $C^{[k]}$ as the time needed for covering all vertices by those random walks. Then, we have*

$$\mathbb{E}[C^{[k]}] \leq \frac{(25 + o(1))n^2 \log^2 n}{k^2} \quad \textit{as } n \to \infty \textit{ provided that } k \geq 6 \log n.$$

6.3. Dynamic Models

In 2003, a dynamic version of cover time problem have been studied in [17]. Authors built two similar scenarios for growing graph model by inspiration from web graphs.

In both model, there is a sequence of connected graphs $\mathcal{G} = \{G(t) : t = 1, 2, ...\}$ for which $G(t)$ is constructed from $G(t-1)$ by adding a new vertex and a fixed number of edges between the new vertex and the vertices of $G(t-1)$ where the initial graph $G(1)$ consists of a single vertex with some self-loops. Their two models only differ by how we connect the new vertex with the previous graph. While the neighbors of the new vertex is chosen independently and uniformly with possibility of multiple connection in the first model, the new vertex is linked with previous vertices with probability proportional to their degrees in the second model. For both models, there is a walker traversing on the vertices where he walks for a fixed length at each step.

Authors examined the expectation of the number of vertices which have not been visited by the walker up to step t. Intuitively, it is clear that a portion of the vertices will not be visited by the walker and they showed that this portion is asymptotically equal to a constant times the reciprocal of the fixed length of the walker's movements at each step and the number of edges between the new vertex and the previous graph is sufficiently large as $t \to \infty$.

Formally, let $G(1)$ be a graph with a single vertex 1 and m self-loops where m is fixed. For $t \geq 2$, let $G(t)$ be a graph constructed from $G(t-1)$ by adding the vertex t and m randomly chosen edges (t, u_i), $i = 1, 2, ..., m$ where:

i. *First Model:* The vertices $u_1, u_2, ..., u_m$ are chosen from $\{1, 2, ..., t-1\}$ independently and uniformly.

ii. *Second Model:* $\mathbb{P}(u_i = u) = \dfrac{d_u}{2m(t-1)}$, $u \in \{1, 2, ..., t-1\}$ where d_u denotes the degree of the vertex u in the graph $G(t-1)$.

As parallel to the grow of the graph, let us say there is a walker at the vertex 1 at the beginning. For $t \geq 2$, assume the walker is at a vertex of $G(t-1)$, and he makes a simple random walk on length l after the addition of the vertex t where l is a fixed natural number independent from the time.

Authors interest the random variable $T_{l,m}(t)$ that is defined as the expectation of the number of vertices not visited by the walker up to the end of step t. We state the main theorem here:

Theorem 6.4. *For both models, if m is sufficiently large, then as $t \to \infty$,*

$$T_{l,m}(t) \sim \mathbb{E}\left[\sum_{s=1}^{t}\prod_{q=s}^{t}\Big(1 - \frac{d(s,q)}{2mq}\big(1 + O(1/m)\big)\Big)^{l}\right],$$

where $d(s,q)$ denotes the degree of the vertex s in the graph $G(q)$.

Moreover, authors examined the limits of $T_{l,m}(t)$. Write $T_{l,m} = \lim_{t\to\infty} \frac{T_{l,m}(t)}{t}$ and $T_l = \lim_{m\to\infty} T_{l,m}$. They give the formula of T_l for both models with an integration.

Theorem 6.5. *With the preceding notation,*

i. *For the first model, we have*

$$T_l = \sqrt{\frac{2}{l}}e^{(l+2)^2/4l}\int_{(l+2)/\sqrt{2l}}^{\infty} e^{-y^2/2}dy.$$

In particular, we have $T_1 = 0.57\cdots$ and $T_l \sim \frac{2}{l}$ as $l \to \infty$.

ii. *For the second model, we have*

$$T_l = e^{l}2l^2\int_{l}^{\infty} y^{-3}e^{-y}dy.$$

In particular, we have $T_1 = 0.59\cdots$ and $T_l \sim \frac{2}{l}$ as $l \to \infty$.

On the other hand, in 2008, another dynamic model for cover time has examined in [18]. As contrary to the previous study, authors worked on a fixed set of vertices. Formally, let V be a finite set, and $\mathcal{G} = \{G_i : i = 1, 2, ...\}$ be a set of graphs where $V(G_i) = V$ for all $i \in \mathbb{N}$.

They call this graph sequence $\mathcal{G}$ as an *evolving* graph, and it is said to be *explorable* if each of G_i is connected and every vertex of G_i has a self-loop to handle the ergodicity issue. Moreover, they say G is *evolving with rate* $\frac{1}{\tau}$ if for all $i \geq 1$, $G_i \neq G_{i+1} \implies G_{i+1} = G_{i+1+j}$ for all $j \in \{0, 1, 2, ..., \tau - 1\}$.

From Theorem 5.4, we know the covering of static graphs always requires a polynomial time. Frist of all, authors proved that this is not the case for evolving graph models. They constructed a sequence of stars in a way that the hitting time between two specific vertices is exponential in terms of the size of stars, which implies the cover time is not polynomial.

Theorem 6.6. *Let* $\mathcal{G} = \{G_i : i = 1, 2, ...\}$ *be an evolving graph where* G_1 *is a star with the addition of a self-loop at each vertex where the center is labelled as* $n - 1$*, and* $1, 2, ..., n - 2$ *and* n *are the leaves. Assume the graph* G_{j+1} *is obtained from* G_j *for* $j \geq 1$ *by just renaming the vertices:*

i. *The vertex* i *is renamed as* $i + 1$ *for* $1 \leq i \leq n - 2$.
ii. *The vertex* $n - 1$ *is renamed as* 1.
iii. *The vertex* n *does not change its name.*

Then, starting from the vertex labelled as 1 *in* G_1*, reaching the vertex* n *requires* $\Omega(2^n)$ *steps in expectation. This implies, there exists an explorable evolving graph that has an exponential cover time.*

Secondly, authors turned their attention to the evolving graphs which have a polynomial cover time like in the case static graphs. Indeed, they showed that if the graphs that form the evolving graph are i.i.d. from a distribution, then we can expect to cover all vertices in polynomial time.

Theorem 6.7. *Let* $\mathcal{G}$ *be a set of graphs on the same vertex set* V*, and let* P *be probability distribution over* $\mathcal{G}$*. Consider the evolving graph* $\mathcal{B} = \{G_1, G_2, ...\}$ *so that each of* G_i *is chosen according to the distribution* P *for* $i \geq 1$*. Thus, the cover time is* $O(n^3 \log n)$ *and the maximum hitting time is* $O(n^3)$ *for* $\mathcal{B}$.

Moreover, authors remark that graphs in the set $\mathcal{G}$ of Theorem 6.7 are not necessarily connected to have a finite cover time unlike in the case of static graphs. To illustrate this fact, authors give the following special case of Theorem 6.7.

Theorem 6.8. *Consider the setting in the Theorem 6.7 where $\mathcal{G}$ denotes the set of all maximum matchings of complete graphs on n vertices and P denotes the uniform distribution of $\mathcal{G}$. Then, cover time is the same as with the cover time of the complete graph, which is $n \log n(1 + o(1))$.*

Finally, authors claimed the real world applications include graphs that usually evolve slower than rate 1. Intuitively, it can be expected that the cover time might be still polynomial for slowly evolving graphs since they behave like static graphs locally. However, authors showed that this is not the case: for any $\epsilon \in (0, 1)$ and a large integer n, it is possible to construct a graph on $O(n)$ vertices that evolves at rate $\dfrac{1}{n^{1-\epsilon}}$ such that covering all vertices requires $2^{\Omega(n^{\epsilon})}$ steps in expectation.

6.4. Random Graph Covering

In 2011, Mohammed Abdullah studied the asymptotic behavior of the cover time of a random graph with given degree sequence in his PhD thesis. Let G be a graph with $V(G) = \{1, 2, ..., n\}$ where $n \to \infty$, and let us write the *degree sequence* of G as $\mathcal{D} = (d_1, d_2, ..., d_n)$ where d_i denotes the degree of the vertex i. The author examined the cover time of a graph that is chosen from uniformly from the set of all connected and simple graphs having the degree sequence $\mathcal{D}$.

First of all, it is trivial that there are some sequences which do not allow a simple graph. For instance, the sum of the entries in the sequence should be even. However, this condition may be insufficient to give a simple graph, too. From this intuition, some technical assumptions are required for $\mathcal{D}$, and the author built further conditions on $\mathcal{D}$ to study the cover time in a proper way, which can be found in Section 7.4 of his thesis [19].

He called the degree sequence as *nice* if those assumptions hold, and the *effective minimum degree* is defined by him as the first entry in the sorted degree sequence which occurs $\Omega(n)$ times. We state his main theorem here:

Theorem 6.9. *Let $\mathcal{D}$ be nice degree sequence of length n where $n \to \infty$, and let G be graph chosen uniformly from the set of all simple and connected graphs having the degree sequence $\mathcal{D}$. Let d be the effective minimum degree of $\mathcal{D}$, and m be the number of edges in G. Then, w.h.p.*

$$cov(G) \sim \frac{2(d-1)m}{(d-2)} \log n.$$

6.5. Bound with Minimum Degree

For a general graph G with n vertices and m edges, we have a bound for $cov(G)$ as $2m(n-1)$ from Theorem 5.4, and we know this bound is optimal up to order of magnitude due to the Example 5.5. In this example, there were some vertices whose degrees are relatively small than the average degree. On the other hand, recall the cover time of the complete graph has order $\Theta(n \log n)$ from Theorem 4.1.

One of the reasons behind the difference between the order of magnitudes can be thought as the balance of the degrees. Therefore, it is reasonable to expect to have a smaller cover time when the minimum degree is large enough in a graph. In [20], authors proved it is the case. In this section, we discuss their proof in detail for the sake of completeness. Firstly, we need a lemma related with hitting time, which comes from as a direct consequence of Proposition 5.1.

Lemma 6.10. *Let G be a graph with m edges and $v \in V$. Then, we have*

$$\sum_{u: uv \in E(G)} \mathbb{E}_u[T_v] = 2m - d_v,$$

where d_v denotes the degree of the vertex v.

Proof. From the Proposition 5.1, we have $m_v = \dfrac{2m}{d_v}$. On the other hand, we can write

$$m_v = 1 + \sum_{u:uv \in E(G)} \mathbb{E}_u[T_v] \cdot \frac{1}{d_v},$$

by conditioning the next step from the vertex v. Therefore, we get

$$\sum_{u:uv \in E(G)} \mathbb{E}_u[T_v] = d_v \cdot \Big(\frac{2m}{d_v} - 1\Big),$$

and so the result follows. $\square$

Secondly, authors use another combinatorial lemma about spanning forests.

Lemma 6.11. *Let G be graph with n vertices and let d_{min} be the minimum degree. Then there exists a collection of $\lceil \frac{d_{min}}{2} \rceil$ spanning forests such that*

i. Each edge in E appears in at most 2 forests.

ii. Each forest has at most $\dfrac{2n}{d_{min}}$ components.

Proof. Construct a directed graph G_d by using the vertices of G where G_d only contains both of directed edges $u \to v$ and $v \to u$ for each $uv \in E$. Let $q = \left\lceil \dfrac{d_{min}}{2} \right\rceil$ and we will construct the collection of spanning forests $F_1, F_2, ..., F_q$ on G as follows:

Step 0: Define $F_0 = \emptyset$ and $s = 0$.

Step 1: Let H be the directed graph obtained by deleting the edges of $F_0, F_1, ..., F_s$ from G_d. Start with an arbitrary $u \in V$, let $X_0 = u$. Construct a path $\{X_0, X_1, ..., X_r\}$ on distinct vertices of F_{s+1} such that H contains the directed edge $X_j \to X_{j+1}$ for each $0 \leq j \leq r-1$ until there is no remaining directed edge whose ancestor is X_r in H. Mark the vertices $\{X_0, X_1, ..., X_r\}$.

Step 2: Update the graph H by deleting marked vertices. Similarly, choose a vertex Y_0 in H and construct a path $\{Y_0, Y_1, ..., Y_t\}$ on distinct vertices of F_{s+1} such that H contains the directed edge $Y_j \to Y_{j+1}$ for each $0 \leq j \leq t-1$ until there is no remaining directed edge whose ancestor is Y_t in H. In the last step, if there are some directed edges from Y_t to marked vertices, choose one of those edges and add into F_{s+1}. Mark the vertices $\{Y_0, Y_1, ..., Y_t\}$.

Step 3: If there are some unmarked vertices, then return Step 2. If all vertices are marked, then we have a spanning forest F_{s+1} on G by making all directed edges undirected. Observe that if there is no remaining directed edge whose ancestor is either X_r or Y_t, then all neighbors of them must be in the constructing path and hence each connected component of F_{s+1} has at least $d_{min} - s + 1$ vertices since at most s edges incident to the stopping vertex X_r or Y_t are deleted from G_d before the construction of F_{s+1}.

Step 4: If $s = q-1$, then stop. If $s < q-1$, then increase s by 1 and return Step 1.

Finally, we have q spanning forests and each edge in E clearly appears in at most two forests due to the construction. Moreover, each connected component of each forest has at least $d_{min} - q + 2$ vertices. Therefore, each forest has at most

$$\frac{n}{d_{min} - q + 2} \leq \frac{n}{d_{min} - \frac{d_{min}}{2} + 1} \leq \frac{2n}{d_{min}} \text{ components,}$$

and we are done. □

Now, we are ready to prove their main theorem.

Theorem 6.12. *Let G be a graph with n vertices and m edges. Then, we have*

$$cov(G) \leq \frac{12mn}{d_{min}},$$

where d_{min} denotes the minimum degree.

Proof. Let us define of the weight of an edge as $W(e) = \mathbb{E}_u[T_v] + \mathbb{E}_v[T_u]$ for any $e = uv \in E(G)$, and define the total weight of a graph $W(S) = \sum_{s\in E(S)} W(s)$ for any graph S. Write $q = \left\lceil \frac{d_{min}}{2} \right\rceil$ and construct the spanning forests $F_1, F_2, ..., F_q$ according to the Lemma 6.11.

Then, we have

$$\sum_{i=1}^{q} W(F_i) \leq 2 \cdot \sum_{e\in E(G)} W(e) = 2 \cdot \sum_{v\in V(G)} \sum_{u:uv\in E(G)} \mathbb{E}_u[T_v],$$

since each edge appears in at most two forests. Thus, by using the Lemma 6.10 and the Proposition 2.1, we get

$$\begin{aligned}
\sum_{i=1}^{q} W(F_i) &\leq 2 \cdot \sum_{v\in V(G)} \sum_{u:uv\in E(G)} \mathbb{E}_u[T_v] \\
&\leq 2 \cdot \sum_{v\in V(G)} (2m - d_v) \\
&= 2 \cdot (2mn - \sum_{v\in V(G)} d_v) \\
&= 2 \cdot (2mn - 2m) = 4m(n-1).
\end{aligned}$$

Now, we have $\sum_{i=1}^{q} W(F_i) \leq 4m(n-1)$, so there exists a spanning forest F_t with

$$W(F_t) \leq \frac{4m(n-1)}{q} \leq \frac{4mn}{\frac{d_{min}}{2}} = \frac{8mn}{d_{min}} \text{ by pigeonhole principle.}$$

On the other hand, the spanning forest F_t has at most $\frac{2n}{d_{min}}$ connected components and so there can be added at most $\frac{2n}{d_{min}} - 1$ edges to obtain a spanning tree. Moreover, we have $W(e) \leq 2m$ for each edge $e \in E$ from the Proposition 5.3.

Therefore, when we obtain a spanning tree T, we would have

$$W(T) \leq \frac{8mn}{d_{min}} + 2m \cdot \Big(\frac{2n}{d_{min}} - 1\Big) \leq \frac{12mn}{d_{min}}.$$

Finally, as in the proof of the Theorem 5.4, we can conclude that $W(T)$ is an upper bound for $cov(G)$ and the result follows. ☐

From the Theorem 6.12, we have a nice corollary for regular graphs.

Corollary 6.13. *Let G be a regular graph with n vertices. Then, $cov(G) \leq 6n^2$.*

Proof. Let m be the number of edges in G. Since G is regular, we get $d_{min} = \dfrac{2m}{n}$. Thus, by using the Theorem 6.12, we get

$$cov(G) \leq \frac{12mn}{d_{min}} = \frac{12mn}{\frac{2m}{n}} = 6n^2,$$

and the result follows. ☐

Moreover, this $O(n^2)$ upper bound is optimal up to order of magnitude, in other words one can construct a graph on n vertices whose cover time has the order $\Theta(n^2)$. Let us examine the following example.

Example 6.14. *Let G be a graph on $4n^2$ vertices where $V(G) = \{1, 2, ..., 4n^2\}$. If we have only these adjacency relations, then $cov(G)$ has the order $\Theta(n^4)$.*

i. *$\{2jn+1, 2jn+2, ..., 2jn+2n\}$ forms a clique with the deletion of the edge between the vertices $jn+1$ and $jn+2$ for each $0 \leq j \leq 2n-1$.*

ii. *For each $0 \leq j \leq 2n-2$, there is an edge between the vertices $2jn+2$ and $2jn+2n+1$.*

iii. *There is an edge between the vertices 1 and $4n^2-2n+2$.*

Proof. Firstly note that G is a regular graph since each vertex has the degree $2n-1$. We will show that $\mathbb{E}_1C$ has the order $\Theta(n^4)$. For simplicity, let $Y_j = 2jn+1$ for $0 \leq j \leq 2n$ and write $Y_{2n} = Y_0$. Observe that, for any $0 \leq i < j \leq 2n-1$, one can reach to the vertex Y_j from the vertex Y_i only

- by passing through $Y_{i+1}, Y_{i+2}, ..., Y_{j-1}$ or
- by passing through $Y_{i-1}, Y_{i-2}, ..., Y_1, Y_0, Y_{2n-1}, Y_{2n-2}, ..., Y_{j+1}$

due to the adjacency relations. Let us consider the cycle formed by the vertices $\{Y_0, Y_1, Y_2, ..., Y_{2n-1}, Y_0\}$.

Starting from Y_0, time needed for covering the vertices $\{Y_1, Y_2, ..., Y_{2n-1}\}$ is less than or equal to $\mathbb{E}_1C$. Let us write $u_i = \mathbb{E}_{Y_i}[T_{Y_{i+1}}]$ for $0 \leq i \leq 2n-1$. It can be clearly observed that u is a constant vector. Since cover time of a cycle of $2n$ vertices is $n(2n-1)$ from the Theorem 4.7, $\mathbb{E}_1C$ is greater than or equal to $n(2n-1)u_0$.

Hence, it is enough to show that $u_0 = \mathbb{E}_1[T_{2n+1}]$ has the order $\Theta(n^2)$. Due to the graph structure, we have

$$\mathbb{P}(2 \text{ is visited before } 2n+1 | X_0 = 1) > \mathbb{P}(2 \text{ is visited after } 2n+1 | X_0 = 1).$$

On the other hand, if we are at the vertex 2, the probability that the next step is $2n+1$ equals $\dfrac{1}{2n-1}$, so visiting the vertex $2n+1$ just after the vertex 2 requires $2n-1$ visits to the vertex 2 in average.

Moreover, since we start the random walk at the vertex 1, the probability that the last visited vertex just before the first hit to $2n+1$ is the vertex 2 is greater than $\dfrac{1}{2}$ due to the graph structure. Therefore, we get

$$u_0 \geq \frac{\mathbb{E}_1[T_2] \cdot (2n-1)}{4}.$$

Now, let $\rho = \dfrac{1}{2n-1}$, $a = \mathbb{E}_1[T_2]$ and $b_j = \mathbb{E}_j[T_2]$ for $3 \leq j \leq 2n$. Since b_j values are the same due to the symmetry, we can write $b = b_j$. Hence, by conditioning on the first step, we have

(i) $a = \rho \cdot (1 + \mathbb{E}_{4n^2-2n+1}[T_2]) + (1-\rho) \cdot (1+b) \geq (1-\rho) \cdot (1+b)$, and

(ii) $b = \rho \cdot 1 + \rho \cdot (1+a) + (1-2\rho) \cdot (1+b)$.

From the second equation we have $2b = a + 2n - 1$. Then, by putting this equality into the first inequality, we get the following:

$$a \geq \frac{2n-2}{2n-1} \cdot \frac{a+2n-1}{2} = \frac{(n-1) \cdot (a+2n+1)}{2n-1},$$

and then $(2n-1) \cdot a \geq (n-1)a + (2n+1)(n-1)$. Therefore, we have

$$a \geq \frac{(2n+1)(n-1)}{n} \geq 2(n-1).$$

Finally, we get $u_0 = \mathbb{E}_1[T_{2n+1}] \geq \dfrac{2(n-1)(2n-1)}{4}$, and therefore we have

$$\mathbb{E}_1 C \geq \frac{n(n-1)(2n-1)^2}{2},$$

which implies it has the order $\Theta(n^4)$ and the result follows. □

6.6. Trees and Bound with Time Difference

Feige gives an improvement for bounding the cover time by combining the classical spanning tree approach used in the Theorem 5.4 with a coupon collector argument in [21]. Let us define *time difference* of u and v, denoted by D_{uv}, as $\mathbb{E}_u[T_v] - \mathbb{E}_v[T_u]$ for any $u, v \in V(G)$. We state the author's main theorem here:

Theorem 6.15. *Let G be a connected graph and $u \in V(G)$. Then, we have*

$$E_u C \leq \frac{1}{2} \cdot \big(\min_{S \in \mathcal{S}} W(S) + \max_{v \in V(G)} D_{uv} \big).$$

where $\mathcal{S}$ denotes the set of all spanning trees of G and $W(S)$ denotes the total weight of the graph S as in the Theorem 6.12.

Theorem 6.15 has nice corollaries. The author improves the upper bound for regular graphs found in Corollary 6.13 by using this theorem with some helpful lemmas (see [21] for details). We just state the result here:

Theorem 6.16. *Let n be a natural number. Then, there exists a constant $\epsilon > 0$ such that $cov(G) < (2-\epsilon)n^2$ for all connected and regular graphs G on n vertices.*

On the other hand, recall the Proposition 5.1, that is $\mathbb{E}_u[T_v] + \mathbb{E}_v[T_u] \leq 2m$ if $uv \in E(G)$. This inequality can be easily improved as $\mathbb{E}_u[T_v] + \mathbb{E}_v[T_u] \leq 2m \cdot \Delta(u,v)$ for all $u, v \in V(G)$ where $\Delta(u,v)$ denotes the length of the shortest path between u and v by using the induction on $\Delta(u,v)$ (see [14] for the details).

Then, for a tree T on n vertices, take two vertices u and v which maximize the quantity D_{uv}. Let l be the number of edges of the path connecting u and v. It is clear that $\mathbb{E}_v T_u \geq l^2$ from the idea used in the proof of Theorem 4.5. Therefore, we get $D_{uv} \leq 2(n-1)l - 2l^2 \leq \left\lfloor \frac{(n-1)^2}{4} \right\rfloor$ for this tree. Thus, by using the Theorem 6.15, we get $cov(T) \leq \left\lfloor \frac{5(n-1)^2}{4} \right\rfloor$, which is the cover time of a path. As a result, we can conclude a path on n vertices has the maximum cover time over all trees of size n.

Next, we can question which tree of order n has the minimum cover time over all trees on n vertices. First of all, we can ask for a meaningful lower bound. In [20], authors constructed a nice recursion that eventually leads the correct asymptotic for the minima on trees. We give the full proof for the sake of completeness.

Theorem 6.17. *Let $\mathcal{T}_n$ be the set of all trees on n vertices. Then, we have*

$$\min_{T\in\mathcal{T}_n} cov(T) \geq 1 + \frac{n-1}{n-2}\min_{T\in\mathcal{T}_{n-1}} cov(T) \text{ for } n \geq 2.$$

Proof. Let T be a tree on n vertices and $u \in V(T)$ where u is a leaf. If we can show $\mathbb{E}_u C \geq 1 + \frac{n-1}{n-2}\min_{S\in\mathcal{T}_{n-1}} cov(S)$, then the result follows. Let v be a unique neighbor of u in T, and d denotes the degree of v in $T-u$. Starting from the vertex u, our first move is (u,v), and we need to wait until the graph $T-u$ is covered. Note that $cov(T-u) \geq \min_{S\in\mathcal{T}_{n-1}} cov(S)$ from the definition. However, the random walk spends time on the edge uv during its journey, and we can add the expected time spending on this edge while covering $T-u$ to obtain a more accurate lower bound.

From Wald's identity, this additional time can be calculated as the product of the expected number of visits to v while covering $T-u$ and the expected number of time spending on the edge uv in each visit. Recall the Corollary 5.2, that states each edge has the same long-run frequency $\frac{1}{2m}$ where m denotes the number of edges in the graph. Since v has degree d in $T-u$, it appears d many edges and so v is visited $cov(T-u)\cdot\frac{d}{2|E(T-u)|}$ times in average while the random walk is covering $T-u$. On the other hand, in each visit to v, the next move of the random walk is u with probability $\frac{1}{d+1}$. Moreover, we have to return back from u to v, which implies each visit to u takes 2 steps. Therefore, the random walk spends $\sum_{j=1}^{\infty} 2j\left(\frac{1}{d+1}\right)^j = \frac{2}{d}$ time on the edge uv in average. As a result, we get

$$\begin{aligned}
\mathbb{E}_u C &\geq 1 + \min_{S\in\mathcal{T}_{n-1}} cov(S) + cov(T-u)\cdot\frac{d}{2|E(T-u)|}\cdot\frac{2}{d} \\
&= 1 + \min_{S\in\mathcal{T}_{n-1}} cov(S) + \frac{cov(T-u)}{n-2} \text{ (since } T-u \text{ has } n-2 \text{ edges)} \\
&\geq 1 + \min_{S\in\mathcal{T}_{n-1}} cov(S) + \frac{\min_{S\in\mathcal{T}_{n-1}} cov(S)}{n-2} \\
&= 1 + \frac{n-1}{n-2}\min_{S\in\mathcal{T}_{n-1}} cov(S),
\end{aligned}$$

which completes the proof. □

Now, we have recursive lower bound on the minimum cover time for trees. The author states that by solving this recursion, it can be obtained that

$$\min_{S \in \mathcal{T}_n} cov(S) \geq n \log n - O(n),$$

which gives the correct asymptotics since we already know that a star on n vertices has the cover time in this order.

We close this section with a remark. Even though this lower bound is optimal up to order of magnitude, it is not exact. In [22], the authors proved that the star graph indeed has minimum cover time over all trees by using an inductive argument.

7. ON-LINE COVERING

In this chapter, we give a new variation of the cover time problem in our best knowledge. As contrary to the classical cover time problem, we are interested in a dynamic version. However, our model has significant differences from the models discussed in [17] and [18].

First of all, our graph is growing in time and we add a single vertex in each time $t \in \mathbb{N}$, as similar to the variation in [17]. In that model, authors proposed two possibilities to connect the new vertex to the previous ones, and here we link all vertices to each other as a first difference. Moreover, we will give additional scenarios for connections as a future work.

On the other hand, the main difference between our model and the previous ones is the moves of the walker. In our investigation, the random walk moves such that the time interval between two consecutive steps is a random variable. We believe this enhances the study of the cover time problem in a dynamic setting. As an initial point, we study the case where those time intervals are exponentially distributed.

Let us formally define our *on-line covering* process:

i. Assume there are k_0 (call v_1, v_2, ..., v_{k_0}) vertices at the beginning (at time 0) for some $k_0 \geq 3$, and there is a walker at the vertex 1.
ii. For each $t \in \mathbb{N}$, create a new vertex v_{k_0+t} at time t.
iii. The walker jumps from his vertex to one of the remaining current vertices at random where time intervals between moves are exponentially distributed with parameter $\dfrac{1}{\lambda}$.

Fix T. We interest the random variable that counts the number of vertices visited at least once by the walker up to time T, which is denoted by N_T.

Let u_j be the probability that the vertex j is visited at least once from the walker up to time T. Clearly, $u_1 = 1$ and $u_s = 0$ for $s \geq k_0 + T$. Firstly, we investigate the expectation of N_T. Observe that

$$\begin{aligned} \mathbb{E}[N_T] &= \mathbb{E}[\sum_{j=1}^{\infty} \mathbb{1}_{\{\text{the vertex } j \text{ is visited up to time } T\}}] \\ &= \sum_{j=1}^{\infty} \mathbb{E}[\mathbb{1}_{\{\text{the vertex } j \text{ is visited up to time } T\}}] \\ &= \sum_{j=1}^{\infty} \mathbb{P}\big(\text{the vertex } j \text{ is visited up to time } T\big) \\ &= 1 + \sum_{j=2}^{k_0+T-1} u_j. \end{aligned}$$

Therefore, we need to calculate the probability u_j for all $1 \leq j \leq T$. On the other hand, let M_t be the number of moves on $[t-1, t)$ for $t \geq 1$. We know M_t has Poisson distribution with parameter λ. Also, it is well-known that M_t and M_s are independent for $t \neq s$. Therefore, for any $1 \leq j \leq T$ and $m_j, m_{j+1}, ..., m_T \in \mathbb{N}_{\geq 0}$, we get

$$\mathbb{P}(M_j = m_j, M_{j+1} = m_{j+1}, ..., M_T = m_t) = \prod_{s=j}^{T} \frac{e^{-\lambda} \lambda^{m_s}}{m_s!}.$$

Now, suppose that $M_1 = m_1$, $M_2 = m_2$, $\ldots$, $M_T = m_T$ are given where $m_1, m_2, ..., m_T \in \mathbb{N}_{\geq 0}$. If $2 \leq j \leq k_0$, then the walker does not visit the vertex v_j up to time T with probability

$$\Big(\frac{k_0 - 2}{k_0 - 1}\Big)^{m_1} \cdot \Big(\frac{k_0 - 1}{k_0}\Big)^{m_2} \cdot \Big(\frac{k_0}{k_0 + 1}\Big)^{m_3} \cdots \Big(\frac{k_0 + T - 3}{k_0 + T - 2}\Big)^{m_T}.$$

Let us call this quantity as $R^j_{m_1, m_2, ..., m_T}$. Similarly, if $k_0 + 1 \leq j \leq k_0 + T - 1$, then the walker does not visit the vertex v_j up to time T with probability

$$\Big(\frac{j-2}{j-1}\Big)^{m_{j-k_0+1}}\cdot\Big(\frac{j-1}{j}\Big)^{m_{j-k_0+2}}\cdot\Big(\frac{j}{j+1}\Big)^{m_{j-k_0+3}}\cdots\Big(\frac{k_0+T-3}{k_0+T-2}\Big)^{m_T}.$$

Similarly, write $S^j_{m_1,m_2,\dots,m_T}$ for this quantity. Thus, for $2\le j\le k_0$, by using conditional probability on M_1, M_2, ... , M_T, we have

$$\begin{aligned}1-u_j &= \mathbb{P}\big(\text{the vertex } j \text{ is not visited up to time } T\big)\\ &= \sum_{m_1,m_2,\dots,m_T\in\mathbb{N}_{\ge 0}} R^j_{m_1,m_2,\dots,m_T}\cdot\mathbb{P}\big(M_i=m_i \text{ for } 1\le i\le T\big).\end{aligned}$$

On the other hand, observe that we can express $R^j_{m_1,m_2,\dots,m_T}$ as

$$\prod_{s=1}^{T}\big(1-\frac{1}{k_0+s-2}\big)^{m_s}.$$

Therefore, we can write

$$\begin{aligned}1-u_j &= \sum_{m_1,m_2,\dots,m_T\in\mathbb{N}_{\ge 0}}\left(\prod_{s=1}^{T}\big(1-\frac{1}{k_0+s-2}\big)^{m_s}\right)\cdot\frac{e^{-\lambda}\lambda^{m_s}}{m_s!}\\ &= \sum_{m_1,m_2,\dots,m_T\in\mathbb{N}_{\ge 0}}\prod_{s=1}^{T}e^{-\lambda}\frac{\big(\lambda-\frac{\lambda}{k_0+s-2}\big)^{m_s}}{m_s!}\\ &= \sum_{m_1,m_2,\dots,m_T\in\mathbb{N}_{\ge 0}} e^{-\lambda T}\cdot\prod_{s=1}^{T}\frac{\big(\lambda-\frac{\lambda}{k_0+s-2}\big)^{m_s}}{m_s!}\\ &= e^{-\lambda T}\cdot\sum_{m_1,m_2,\dots,m_T\in\mathbb{N}_{\ge 0}}\prod_{s=1}^{T}\frac{\big(\lambda-\frac{\lambda}{k_0+s-2}\big)^{m_s}}{m_s!}\\ &= e^{-\lambda T}\cdot\prod_{s=1}^{T}\sum_{l=0}^{\infty}\frac{\big(\lambda-\frac{\lambda}{k_0+s-2}\big)^{l}}{l!}\\ &= e^{-\lambda T}\cdot\prod_{s=1}^{T}e^{\lambda-\frac{\lambda}{k_0+s-2}} = e^{-\lambda\cdot\big(\frac{1}{k_0-1}+\frac{1}{k_0}+\dots+\frac{1}{k_0+T-2}\big)}.\end{aligned}$$

Similarly, for $k_0 + 1 \leq j \leq k_0 + T - 1$, again by using conditional probability on M_{k-j_0+1}, M_{k-j_0+2}, ..., M_T, we have

$$\begin{aligned} 1 - u_j &= \mathbb{P}\big(\text{the vertex } j \text{ is not visited up to time } T\big) \\ &= \sum_{m_{j-k_0+1}, m_{j-k_0+2}, \ldots, m_T \in \mathbb{N}_{\geq 0}} S^j_{m_1, m_2, \ldots, m_T} \cdot \mathbb{P}\big(M_i = m_i \text{ for } j - k_0 + 1 \leq i \leq T\big). \end{aligned}$$

On the other hand, observe that we can express $S^j_{m_1, m_2, \ldots, m_T}$ as

$$\prod_{s=j-k_0+1}^{T} (1 - \frac{1}{k_0 + s - 2})^{m_s}.$$

Therefore, we have

$$\begin{aligned} 1 - u_j &= \sum_{m_{j-k_0+1}, m_{j-k_0+2}, \ldots, m_T \in \mathbb{N}_{\geq 0}} \Big(\prod_{s=j-k_0+1}^{T} (1 - \frac{1}{k_0 + s - 2})^{m_s} \Big) \cdot \frac{e^{-\lambda} \lambda^{m_s}}{m_s!} \\ &= \sum_{m_{j-k_0+1}, m_{j-k_0+2}, \ldots, m_T \in \mathbb{N}_{\geq 0}} \prod_{s=j-k_0+1}^{T} e^{-\lambda} \frac{\big(\lambda - \frac{\lambda}{k_0+s-2}\big)^{m_s}}{m_s!} \\ &= \sum_{m_{j-k_0+1}, m_{j-k_0+2}, \ldots, m_T \in \mathbb{N}_{\geq 0}} e^{-\lambda T} \cdot \prod_{s=j-k_0+1}^{T} \frac{\big(\lambda - \frac{\lambda}{k_0+s-2}\big)^{m_s}}{m_s!} \\ &= e^{-\lambda T} \cdot \sum_{m_{j-k_0+1}, m_{j-k_0+2}, \ldots, m_T \in \mathbb{N}_{\geq 0}} \prod_{s=j-k_0+1}^{T} \frac{\big(\lambda - \frac{\lambda}{k_0+s-2}\big)^{m_s}}{m_s!} \\ &= e^{-\lambda T} \cdot \prod_{s=j-k_0+1}^{T} \sum_{l=0}^{\infty} \frac{\big(\lambda - \frac{\lambda}{k_0+s-2}\big)^{l}}{l!} \\ &= e^{-\lambda T} \cdot \prod_{s=j-k_0+1}^{T} e^{\lambda - \frac{\lambda}{k_0+s-2}} = e^{-\lambda \cdot \big(\frac{1}{j-1} + \frac{1}{j} + \ldots + \frac{1}{k_0+T-2}\big)}. \end{aligned}$$

As a result, we calculated the probabilities u_j as follows:

$$u_j = \begin{cases} 1 - e^{-\lambda \cdot \left(\mathcal{H}(k_0+T-2) - \mathcal{H}(k_0-2)\right)}, & \text{if } 2 \leq j \leq k_0, \\ 1 - e^{-\lambda \cdot \left(\mathcal{H}(k_0+T-2) - \mathcal{H}(j-2)\right)}, & \text{if } k_0 + 1 \leq j \leq k_0 + T - 1. \end{cases}$$

Therefore, by using $\mathbb{E}[N_T] = 1 + \sum_{j=2}^{k_0+T-1} u_j$, we get the following result:

$$\mathbb{E}[N_T] = (k_0 + T - 1) - e^{-\lambda \cdot \mathcal{H}(k_0+T-2)} \Big((k_0 - 1) e^{\lambda \cdot \mathcal{H}(k_0-2)} + \sum_{j=k_0+1}^{k_0+T-1} e^{\lambda \cdot \mathcal{H}(j-2)} \Big).$$

Secondly, we investigate the variance of the random variable N_T. Let A_j be the event that the vertex j is not visited by the walker up to time T. Suppose $M_1 = m_1$, $M_2 = m_2$, ..., $M_T = m_T$ are given where $m_1, m_2, ..., m_T \in \mathbb{N}_{\geq 0}$.

Let us assume that $k_0 + 1 \leq r < s \leq k_0 + T - 1$. Then, the event $A_r \cap A_s$ occurs with probability

$$\Big(\frac{r-2}{r-1}\Big)^{m_{r+1}} \cdot \Big(\frac{r-1}{r}\Big)^{m_{r+2}} \cdots \Big(\frac{s-3}{s-2}\Big)^{m_s} \cdot \Big(\frac{s-3}{s-1}\Big)^{m_{s+1}} \cdot \Big(\frac{s-2}{s}\Big)^{m_{s+2}} \cdots \Big(\frac{k_0+T-4}{k_0+T-2}\Big)^{m_T}.$$

Thus, with the same conditioning argument, we can write

$$\begin{aligned} \mathbb{P}(A_r \cap A_s) &= e^{-\lambda \cdot \left(\frac{1}{r-1} + \frac{1}{r} + \cdots \frac{1}{s-2} + \frac{2}{s-1} + \frac{2}{s} + \cdots + \frac{2}{k_0+T-2}\right)} \\ &= e^{-\lambda \cdot \left(\frac{1}{r-1} + \frac{1}{r} + \cdots \frac{1}{k_0+T-2}\right)} \cdot e^{-\lambda \cdot \left(\frac{1}{s-1} + \frac{1}{s} + \cdots \frac{1}{k_0+T-2}\right)}. \end{aligned}$$

As a result, we have

$$\mathbb{P}(A_r \cap A_s) = (1 - u_r)(1 - u_s) = \mathbb{P}(A_r)\mathbb{P}(A_s),$$

for $k_0 + 1 \leq r < s \leq k_0 + T - 1$. By the same way, it can be seen that the equality $\mathbb{P}(A_r \cap A_s) = \mathbb{P}(A_r)\mathbb{P}(A_s)$ holds for all $2 \leq r < s \leq k_0 + T - 1$.

Define U_T as the number of unvisited vertices among $\{v_1, v_2, ..., v_{k_0+T-1}\}$ by the walker up to time T. Since $N_T + U_T = k_0 + T - 1$, we have $\mathrm{Var}[N_T] = \mathrm{Var}[U_T]$. Now, by using the equalities $U_T = \sum_{j=2}^{k_0+T-1} \mathbb{1}_{A_j}$, $\mathbb{P}(A_j) = 1 - u_j$, and the fact that equality $\mathbb{P}(A_r \cap A_s) = \mathbb{P}(A_r)\mathbb{P}(A_s)$ holds for all $2 \leq r < s \leq k_0 + T - 1$, we get the following:

$$
\begin{aligned}
\mathrm{Var}[U_T] &= \mathbb{E}[U_T^2] - \mathbb{E}[U_T]^2 \\
&= \mathbb{E}\left[\Big(\sum_{j=2}^{k_0+T-1} \mathbb{1}_{A_j}\Big)^2\right] - \mathbb{E}\left[\sum_{j=2}^{k_0+T-1} \mathbb{1}_{A_j}\right]^2 \\
&= \left(\sum_{j=2}^{k_0+T-1} \mathbb{E}[\mathbb{1}_{A_j}^2] - \mathbb{E}[\mathbb{1}_{A_j}]^2\right) + 2 \cdot \left(\sum_{2 \leq r < s \leq k_0+T-1} \mathbb{E}[\mathbb{1}_{A_r}\mathbb{1}_{A_s}] - \mathbb{E}[\mathbb{1}_{A_r}]\mathbb{E}[\mathbb{1}_{A_s}]\right) \\
&= \left(\sum_{j=2}^{k_0+T-1} \mathrm{Var}[\mathbb{1}_{A_j}]\right) + 2 \cdot \left(\sum_{2 \leq r < s \leq k_0+T-1} \mathbb{P}(A_r \cap A_s) - \mathbb{P}(A_r)\mathbb{P}(A_s)\right) \\
&= \sum_{j=2}^{k_0+T-1} (1 - u_j)u_j.
\end{aligned}
$$

As a result, we have the variance for N_T as $\mathrm{Var}[N_T] = \sum_{j=2}^{k_0+T-1} (1 - u_j) \cdot u_j$.

8. APPROXIMATE COVERING

In this chapter, we discuss a similar problem related with the classical cover time. Our main motivation is to explore the graphs with the following property:

The time required for covering all vertices is significantly larger than the time needed to cover a large portion of the vertices.

In other words, we are trying to understand what are the conditions on the structure of the graph G which allow the proportion between the time required to cover $(1-\alpha)|G|$ vertices and $cov(G)$ converges to zero. We believe this is a meaningful research question since it may be valuable to cover an important portion of the states in real world applications such as web sites.

Formally, let $0 \leq \alpha < 1$ be a constant. As a generalization of the classical cover time problem, we define the *approximate covering time with parameter* α of the graph G as the expectation of the first time that at least $(1-\alpha)|G|$ vertices have been visited. Note that the case $\alpha = 0$ corresponds to the original problem. We denote it by $cov(G_{1-\alpha})$, and write $C_{1-\alpha}$ for corresponding random time.

Here, we initiate this study by examining this problem in the specific graph whose the classical cover time can be exactly determined.

We start with the complete graph, and we see it leads to an *easy* covering.

Theorem 8.1. *Let G be a complete graph on n vertices and $0 \leq \alpha < 1$ be a constant. Then, we have*

$$cov(G_{1-\alpha}) = (n-1) \cdot \big(\mathcal{H}(n-1) - \mathcal{H}(\lfloor \alpha \cdot n \rfloor)\big).$$

Proof. We basically use the same idea in the proof of Theorem 4.1. If we write C_r for the first time that exactly r distinct vertices have been covered, we know

$$\mathbb{E}[C_{r+1} - C_r] = \frac{n-1}{n-r}.$$

Thus, we can write

$$cov(G_{1-\alpha}) = \sum_{r=1}^{\lceil (1-\alpha)n \rceil - 1} \mathbb{E}[C_{r+1} - C_r],$$

by using the linearity of the expectation. Hence we get

$$cov(G_{1-\alpha}) = (n-1) \cdot \sum_{r=1}^{\lceil (1-\alpha)n \rceil - 1} \frac{1}{n-r} = (n-1) \cdot (\mathcal{H}(n-1) - \mathcal{H}(\lfloor \alpha \cdot n \rfloor)). \qquad \square$$

Now, by using the fact that $\mathcal{H}(n) \sim \log n$, it is clear that

$$\lim_{n\to\infty} \frac{cov(G_{1-\alpha})}{cov(G)} = \lim_{n\to\infty} \frac{(n-1)(\mathcal{H}(n-1) - \mathcal{H}(\lfloor \alpha \cdot n \rfloor))}{(n-1)\mathcal{H}(n-1)} = \lim_{n\to\infty} \frac{\ln n - \ln \alpha n}{\ln n} = 0,$$

which is independent from α.

Therefore, we can conclude that approximate covering in the complete graph is linear as independent from the parameter α where the classical cover time has the order $\Theta(n \log n)$.

Secondly, we discuss the approximate cover time problem for the cycle.

Theorem 8.2. *Let G be a cycle on n vertices and $0 \leq \alpha < 1$ be a constant. Write $q = \lceil (1-\alpha)n \rceil$. Then, we have*

$$cov(G_{1-\alpha}) = \frac{q(q-1)}{2}.$$

Proof. Again, the proof is almost the same as the proof of Theorem 4.7. In that theorem, we proved that $\mathbb{E}[C_r] = \dfrac{r(r-1)}{2}$ where it denotes the first time that exactly r distinct vertices have been covered. Therefore, we get $cov(G_{1-\alpha}) = \mathbb{E}[C_q] = \dfrac{q(q-1)}{2}$ as desired. ☐

As contrary to the complete graph case, the cycle graph does not allow easy covering. Indeed, it can be observed that

$$\lim_{n\to\infty} \frac{cov(G_{1-\alpha})}{cov(G)} = \lim_{n\to\infty} \frac{q(q-1)}{n(n-1)} = (1-\alpha)^2,$$

which depends on α and does not converge to zero.

Thirdly, we state the approximate covering time for star graphs. Since the proof can be directly follows from Theorem 4.4 and Theorem 8.1, we give the following without proof.

Theorem 8.3. *Let G be a star with n vertices. Then, we have*

$$\mathbb{E}_u[C_{1-\alpha}] = \begin{cases} 2(n-1)(\mathcal{H}(n-2) - \mathcal{H}(\lfloor \alpha \cdot n \rfloor - 1)), & \textit{if } u \textit{ is the center,} \\ 2(n-1)(\mathcal{H}(n-2) - \mathcal{H}(\lfloor \alpha \cdot n \rfloor - 1)) - 1, & \textit{if } u \textit{ is a leaf.} \end{cases}$$

As a result, we get $cov(G_{1-\alpha}) = 2(n-1)(\mathcal{H}(n-2) - \mathcal{H}(\lfloor \alpha \cdot n \rfloor - 1))$.

We remark that, as similar with the complete graph case, asymptotically we get

$$\lim_{n\to\infty} \frac{cov(G_{1-\alpha})}{cov(G)} = 0,$$

as independent from α in star graphs. Hence, it can be concluded that a star graph also allows easy covering.

Finally, we investigate the approximate covering problem on path graphs. Unlike the previous cases, its proof is quite different from the classical cover time case.

Theorem 8.4. *Let G be a path with n vertices where $V = \{1, 2, ..., n\}$ from left to the right, and $0 \leq \alpha < \frac{1}{2}$. Write $q = \lceil (1-\alpha)n \rceil$ and $r = n - q + 1$. Then, we have*

$$\mathbb{E}_s[C_{1-\alpha}] = \begin{cases} (q-1)^2 - \dfrac{(s-1)(s-2)}{2}, & \textit{if } s \leq r, \\ (s-r)(q-s) + (q-1)^2 - \dfrac{(r-1)(r-2)}{2}, & \textit{if } r \leq s \leq q-1, \\ (q-1)^2 - \dfrac{(n-s)(n-s-1)}{2}, & \textit{if } s \geq q. \end{cases}$$

Proof. Let us write $G^s_{1-\alpha}$ for $\mathbb{E}_s[C_{1-\alpha}]$. Firstly, $G^1_{1-\alpha}$ is equal to the usual hitting time for the vertex q where the starting point is the vertex 1. Hence, we can write $G^1_{1-\alpha} = (q-1)^2$ from Theorem 4.5.

Suppose we start at the vertex 2. Now, if we hit the vertex 1 before the vertex q, the situation is the same as covering at least q vertices starting at the vertex 1. However, if we hit the vertex q before the vertex 1, then we must wait until the first hit to the vertex 1 or the vertex $q+1$ by considering the starting point as the vertex q. Therefore, by using the conditional expectation, $G^2_{1-\alpha}$ equals to the following:

$$\mathbb{P}_2(T_1 < T_q)\cdot(G^1_{1-\alpha} + \mathbb{E}_2[T_1|T_1 < T_q]) + \mathbb{P}_2(T_q < T_1)\cdot(\mathbb{E}_q[\min\{T_1, T_{q+1}\}] + \mathbb{E}_2[T_q|T_q < T_1]).$$

Hence we can rearrange these terms. Since

$$\mathbb{E}_2[\min\{T_1, T_q\}] = \mathbb{P}_2(T_1 < T_q)\cdot \mathbb{E}_2[T_1|T_1 < T_q] + \mathbb{P}_2(T_q < T_1)\cdot \mathbb{E}_2[T_q|T_q < T_1],$$

we can write

$$G^2_{1-\alpha} = \mathbb{E}_2[\min\{T_1, T_q\}] + \mathbb{P}_2(T_1 < T_q)\cdot G^1_{1-\alpha} + \mathbb{P}(T_q < T_1)\cdot \mathbb{E}_q[\min\{T_1, T_{q+1}\}].$$

On the other hand, recall the Proposition 4.3. We know $\mathbb{P}_2(T_1 < T_q) = \frac{q-2}{q-1}$, $\mathbb{P}_2(T_q < T_1) = \frac{1}{q-1}$, $\mathbb{E}_2[\min\{T_1, T_q\}] = (q-2)$, and $\mathbb{E}_q[\min\{T_1, T_{q+1}\}] = (q-1)$. Therefore, we get

$$G^2_{1-\alpha} = (q-2) + \frac{q-2}{q-1} \cdot (q-1)^2 + \frac{1}{q-1} \cdot (q-1) = (q-1)^2.$$

Now, let us examine a little bit more general case. Suppose we start at the vertex s for some $2 \leq s \leq r$. As similar to the previous case, if we hit the vertex $s-1$ before the vertex q, the situation is the same as covering at least q vertices starting at the vertex $s-1$.

However, if we hit the vertex q before the vertex $s-1$, then we have a subsegment of length $(q-s+1)$ and it is possible to move exactly $(s-1)$ steps in both directions because $s \leq n-q+1$.

Therefore, the situation in this case is the same as traveling on a circle. As in the proof of Theorem 4.7, we get the expectation of the time needed for visiting $(s-1)$ new vertices is $\sum_{l=1}^{s-1}(q-l) = \frac{(2q-s)(s-1)}{2}$.

Again, with the same conditioning in the calculating $G^2_{1-\alpha}$, for $s \leq r$, we get

$$G^s_{1-\alpha} = (q-s) + \frac{q-s}{q-s+1} \cdot G^{s-1}_{1-\alpha} + \frac{1}{q-s} \frac{(2q-s)(s-1)}{2}.$$

We claim we can get $G^s_{1-\alpha} = (q-1)^2 - \frac{(s-1)(s-2)}{2}$ inductively from this recursion. Let us prove our claim. Firstly, we showed that it holds for $s = 1$ and $s = 2$.

Now, assume $G_{1-\alpha}^{t} = (q-1)^2 - \dfrac{(t-1)(t-2)}{2}$ for some $2 \leq t \leq r-1$. Then, we need to show that $G_{1-\alpha}^{t+1} = (q-1)^2 - \dfrac{t(t-1)}{2}$. From the induction hypothesis and the formula obtained above, we have

$$G_{1-\alpha}^{t+1} = (q-t-1) + \frac{q-t-1}{q-t} \cdot \Big((q-1)^2 - \frac{(t-1)(t-2)}{2}\Big) + \frac{1}{q-t}\frac{(2q-t-1)t}{2}.$$

Hence, it can be observed that

$$\begin{aligned}
G_{1-\alpha}^{t+1} &= \frac{(q-t-1)\Big(q-t+q^2-2q+1-\dfrac{(t-1)(t-2)}{2}\Big)+\dfrac{(2q-t-1)t}{2}}{q-t} \\
&= q^2-q-\frac{t(t-1)}{2}+\frac{-q^2+q+\dfrac{t(t-1)}{2}+qt-\dfrac{t(t+1)}{2}}{q-t} \\
&= q^2-q-\frac{t(t-1)}{2}-\frac{q^2-(t+1)q+t}{q-t} \\
&= q^2-q-\frac{t(t-1)}{2}-\frac{(q-1)(q-t)}{q-t} \\
&= q^2-q-\frac{t(t-1)}{2}-(q-1) = (q-1)^2-\frac{t(t-1)}{2},
\end{aligned}$$

and the claim follows.

Now, suppose we start at the vertex s for some $r+1 \leq s \leq q-1$. Again, if we hit the vertex $s-1$ before the vertex q, the situation is the same as covering at least q vertices starting at the vertex $s-1$. However, if we hit the vertex q before the vertex $s-1$, then we have the same situation with covering at least q vertices starting at the vertex q because we have to visit all vertices from $r+1$ to $q-1$ in the case that starting point is q. Moreover, due to the symmetry, we get $G_{1-\alpha}^{q} = G_{1-\alpha}^{r}$. Therefore, as similar to the previous conditionings, for $r+1 \leq s \leq q-1$, we get

$$G_{1-\alpha}^{s} = (q-s) + \frac{q-s}{q-s+1} \cdot G_{1-\alpha}^{s-1} + \frac{1}{q-s+1} \cdot G_{1-\alpha}^{r}.$$

Firstly, observe that $G_{1-\alpha}^{r+1} = (q-r-1) + (q-1)^2 - \dfrac{(r-1)(r-2)}{2}$. Assume $s \geq r+2$ and let us write the equality above for $s-1$, we get

$$G_{1-\alpha}^{s-1} = (q-s+1) + \frac{q-s+1}{q-s+2} \cdot G_{1-\alpha}^{s-2} + \frac{1}{q-s+2} \cdot G_{1-\alpha}^{r}.$$

By subtracting $(q-s+2)$ times the second equality from $(q-s+1)$ times the first one, we can write

$$(q-s+1)G_{1-\alpha}^{s} - (q-s+2)G_{1-\alpha}^{s-1} = -2(q-s+1) + (q-s)G_{1-\alpha}^{s-1} - (q-s+1)G_{1-\alpha}^{s-2}.$$

Hence, we have

$$(q-s+1)G_{1-\alpha}^{s} = -2(q-s+1) + (2q-2s+2)G_{1-\alpha}^{s-1} - (q-s+1)G_{1-\alpha}^{s-2}.$$

Note that $q+1 > s$ and the factor $(q-s+1)$ can be canceled, which leads

$$G_{1-\alpha}^{s} = 2 \cdot G_{1-\alpha}^{s-1} - G_{1-\alpha}^{s-2} - 2 \text{ for } r+2 \leq s \leq q-1.$$

Moreover, we have two initial values as $G_{1-\alpha}^{r} = (q-1)^2 - \dfrac{(r-1)(r-2)}{2}$ and $G_{1-\alpha}^{r+1} = (q-r-1) + (q-1)^2 - \dfrac{(r-1)(r-2)}{2}$. Thus, we can inductively get

$$G_{1-\alpha}^{s} = (s-r)(q-s) + (q-1)^2 - \frac{(r-1)(r-2)}{2} \text{ for } r \leq s \leq q-1,$$

by using these recurrence relation and initial values as follows:

First of all, the equality holds for $s = r$ and $s = r + 1$. Assume it holds for $s = t$ and $s = t + 1$ for some $r \leq t \leq q - 3$. To be short, let us write

$$A = (q-1)^2 - \frac{(r-1)(r-2)}{2}.$$

Then, we can observe

$$\begin{aligned}
G_{1-\alpha}^{t+2} &= 2 \cdot G_{1-\alpha}^{t+1} - G_{1-\alpha}^{t} - 2 \\
&= 2 \cdot \Big[(t+1-r)(q-t-1) + A\Big] - \Big[(t-r)(q-t) + A\Big] - 2 \\
&= 2(t-r)(q-t) + 2(q-t) - 2(t-r) - 2 + (t-r)(q-t) - 2 + A \\
&= (t-r)(q-t) + 2(q-t) - 2(t-r) - 4 + A \\
&= (t-r+2)(q-t-2) + A \\
&= (t+2-r)(q-(t+2)) + (q-1)^2 - \frac{(r-1)(r-2)}{2},
\end{aligned}$$

and we are done.

Therefore, we proved that

$$G_{1-\alpha}^{s} = (s-r)(q-s) + (q-1)^2 - \frac{(r-1)(r-2)}{2} \text{ for } r+1 \leq s \leq q-1.$$

Finally, due to the symmetry, we know $G_{1-\alpha}^{t} = G_{1-\alpha}^{n+1-t}$ for any t. Therefore, the formula found for case of $s \leq r$ will be the same with the cases for which $s \geq q$. Then, we get

$$\begin{aligned}
G_{1-\alpha}^{s} &= (q-1)^2 - \frac{(n+1-s-1)(n+1-s-2)}{2} \\
&= (q-1)^2 - \frac{(n-s)(n-s-1)}{2} \text{ for } s \geq q,
\end{aligned}$$

which completes the proof. $\square$

9. CONCLUSION

In this chapter, we summarize the results from the thesis, state some open questions and discuss some research directions by inspiration from our work done.

First of all, the cover time problem is pretty natural and simple to define. Moreover, it is very appropriate to investigate by just using combinatorial ideas at the initial. In many real world applications, it may be the case that the problem studied is equivalent or at least has similarities with the cover time concept. The coupon collector problem can be considered as the simplest example, and also there are some problems in other combinatorial problems having a strong connection between the cover time such as the problem discussed in the Section 6.2.

Even though it is easy to express, the cover time problem still has numerous open questions. The most surprising one is that it is still unknown whether there is a graph on n vertices whose cover time is smaller than the cover time of K_n. In [11], authors give several open questions about hitting time and cover time, and a very few of them have been answered in our best knowledge.

Moreover, there are only few graph families whose cover time can be exactly determined. For instance, the cover time of $K_{n,n}$ is still open, which is basically equivalent to the expectation of the time that at least one of two independent collections of size n is completed. On the other hand, for a given graph G of order n, it is unknown that whether $cov(G)$ can be calculated in polynomial number of steps.

In our first contribution for the cover time, namely on-line covering, there are several research directions. First of all, the creation of the new vertices is made deterministically, in other words, we create a new vertex in each time step. Naturally, it can be randomized by assigning a probability distribution for the time intervals between two consecutive vertex creation such as exponential distribution.

On the other hand, the distribution of the time intervals between walker's moves can be changed into another distribution. For instance, it can be questioned that what will be the case when the walker traverses according to the uniform distribution. We remark that we use the memoryless property of the exponential distribution in our solution for the model, therefore it may be challenging when the new distribution has not this property.

The most important extension can be considered as the graph structure in on-line covering. We have studied on the complete graph, in other words, every new vertex is automatically connected with the previous ones. However, this may be struggling when we deal with a real world application, and therefore studying the same problem with a *more realistic* graph structure would be better. Of course, the definition of "realistic" is very unclear, and there must be more investigation on this topic. Just as a candidate, the second model studied in [17] can be seen as an appropriate to model the real world, however it becomes very difficult to analyze.

Finally, our second contribution has some interesting research questions. We have examined a very few example to calculate the approximate cover time, however it seems that we are limited to extend graph families examined since those are already the only graphs whose classical cover time is known. However, the main research question in this study can be giving some structural information about the graphs whose partial covering requires strictly less time than the full covering in terms of magnitude.

We know a complete graph is an example for such graphs, but it is very unclear that the secret property of it allowing easy covering. Therefore, it may be valuable if one can reach a result about those graphs at least intuitively.

REFERENCES

1. Flajolet, P., D. Gardy and L. Thimonier, “Birthday paradox, coupon collectors, caching algorithms and self-organizing search”, *Discrete Applied Mathematics*, Vol. 39, No. 3, pp. 207 – 229, 1992.

2. Grinstead, C. and J. Snell, *Grinstead and Snell’s Introduction to Probability*, Titolo collana, University Press of Florida, 2009.

3. Ross, S., *A First Course in Probability*, Pearson Prentice Hall, 2010.

4. Kolmogorov, A., *Foundations of the Theory of Probability*, Martino Fine Books, 2013.

5. Durrett, R., *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2010.

6. Renyi, A., *Probability Theory*, Dover Books on Mathematics Series, Dover Publications, Incorporated, 2012.

7. West, D., *Introduction to Graph Theory*, Featured Titles for Graph Theory, Prentice Hall, 2001.

8. Bollobas, B., *Graph Theory: An Introductory Course*, Graduate Texts in Mathematics, Springer New York, 2012.

9. Norris, J. R., *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1997.

10. Gagniuc, P. A., *Markov Chains: From Theory to Implementation and Experimentation*, USA, NJ: John Wiley & Sons., 2017.

11. Aldous, D. and J. A. Fill, *Reversible Markov Chains and Random Walks on Graphs*, 2002, unfinished monograph, recompiled 2014.

12. Matthews, P., "Covering Problems for Brownian Motion on Spheres", *The Annals of Probability*, Vol. 16, No. 1, 1988.

13. Song, S. and J. Song, "A Note on the History of the Gambler's Ruin Problem", *Communications for Statistical Applications and Methods*, Vol. 20, pp. 1–12, 2013.

14. Aleliunas, R., R. M. Karp, R. J. Lipton, L. Lovász and C. Rackoff, "Random Walks, Universal Traversal Sequences, and the Complexity of Maze Problems", *20th Annual Symposium on Foundations of Computer Science (San Juan, Puerto Rico, 1979)*, pp. 218–223, 1979.

15. Zuckerman, D., "On the time to traverse all edges in a graph", *Information Process Letters*, Vol. 38, pp. 335–337, 1991.

16. Broder, A. Z., P. Raghavan, A. R. Karlin and E. Upfal, "Trading space for the time in undirected $s-t$ connectivity", *SIAM Journal of Computing*, Vol. 23, pp. 324–334, 1994.

17. Cooper, C. and A. Frieze, "Crawling on simple models of web graphs.", *Internet Mathematics*, Vol. 1, pp. 57–90, 2003.

18. Avin, C., M. Koucký and Z. Lotker, "How to Explore a Fast-Changing World (Cover Time of a Simple Random Walk on Evolving Graphs)", *ICALP 2008: Automata, Languages and Programming*, pp. 121–132, 2008.

19. Abdullah, M., *The Cover Time of Random Walks on Graphs*, Ph.D. Thesis, King's College London, 2011.

20. Kahn, J., N. Linial, N. Nisan and M. Saks, "On the cover time of random walks on graphs", *Journal of Theoretical Probability*, Vol. 2, pp. 121–128, 1989.

21. Feige, U., “Collecting Coupons on Trees, and the Analysis of Random Walks”, *Computational Complexity*, Vol. 6, 2000.

22. Brightwell, G. and P. Winkler, “Extremal cover time for random walks on trees”, *Journal of Graph Theory*, Vol. 14, pp. 547 – 554, 1990.