# 國 立 交 通 大 學

## 應 用 數 學 系

## 博士論文

以圖為基礎之存取結構上的秘密分享機制的
平均訊息比率之研究

**The Average Information Ratio of Secret-Sharing Schemes for Graph-Based Access Structures**


研 究 生：呂惠娟

指 導 教 授：傅恆霖 教授


中 華 民 國 一 百 零 二 年 六 月

# The Average Information Ratio of Secret-Sharing Schemes for Graph-Based Access Structures

# 以圖為基礎之存取結構上的秘密分享機制的平均訊息比率之研究


研 究 生： 呂惠娟　　Student: Hui-Chuan Lu

指導教授： 傅恆霖 教授　Advisor: Hung-Lin Fu

國 立 交 通 大 學

應 用 數 學 系


博 士 論 文

A Dissertation
Submitted to Department of Applied Mathematics
College of Science
National Chiao Tung University
in Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy
in
Applied Mathematics
June 2013
Hsinchu, Taiwan, Republic of China

中 華 民 國 一 百 零 二 年 六 月

# Abstract

A perfect secret-sharing scheme is a method of distributing a secret among a set of $n$ participants in such a way that only qualified subsets of participants can recover the secret and the joint share of the participants in any unqualified subset is statistically independent of the secret. The collection of all qualified subsets is called the access structure of the scheme. In a graph-based access structure, each vertex of a graph $G$ represents a participant and each edge of $G$ represents a minimal qualified subset. The information ratio of a perfect secret-sharing scheme realizing a given access structure is the ratio of the maximum length of the share given to a participant to the length of the secret, while the average information ratio is the ratio of the average length of the shares given to the participants to the length of the secret. The infimum of the (average) information ratio of all possible perfect secret-sharing schemes realizing an access structure is called the optimal (average) information ratio of that access structure. In this thesis, we focus on the average information ratio of graph-based access structures.

In a weighted threshold scheme, each participant has his or her own weight. A subset is qualified if and only if the sum of the weights of participants in the subset is not less than the given threshold. Morillo et al. considered the scheme for a weighted threshold access structure that can be represented by a graph which is referred to as a $k$-weighted graph. They characterized this kind of access structures and derived a bound on the optimal information ratio. In Chapter 2, we deal with the average information ratio of the secret-sharing schemes for these access structures. Two sophisti-

cated constructions are presented. Bounds on the average information ratio of them are derived. Each of our constructions has its own advantages and both of them perform very well when $n/k$ is large.

Due to the difficulty of finding the exact values of the optimal information ratio and the optimal average information ratio, most results give bounds on them. Before 2007, apart from one specially defined class of graphs, the paths and cycles are the only infinite classes of graph-based access structures whose optimal information ratio and optimal average information ratio are known. Csirmaz and Tardos found the exact values of the optimal information ratio of all tree-based access structures in 2007. In 2009, Csirmaz and Ligeti determined the exact values of the optimal information ratio of broader classes of graph-based access structures.

Following in their footsteps, we devote our efforts to the discussion the optimal average information ratio of tree-based access structures in Chapter 3. We successfully determine the exact values of the optimal average information ratio of all tree-based access structures. Our idea also formulates a complicated problem in secret-sharing into a problem in Graph Theory with easy description.

Extending our work in Chapter 3, we are dedicated to the study the optimal average information ratio of the access structures based on bipartite graphs in Chapter 4. We determine the optimal average information ratio of some classes of bipartite graphs. In addition, we also give a bound on the optimal average information ratio of the rest classes of bipartite graphs. This bound is the best for some classes of bipartite graphs using our approach.

In the final chapter, we summarize our work in this thesis and introduce possible directions of future research.

# 摘要

所謂祕密分享機制(secret-sharing scheme)的概念是一個將秘密(secret)分成許多 shares 給所有參與者(participants)，使得只有授權子集(qualified subset)中的參與者將所分配到的 shares 組合起來才能重建出這個秘密，而非授權子集(nonqualified subset) 中的參與者則無法從分配到的shares得到任何有關這個秘密的任何資訊的機制。其中，所有授權子集所成的集合稱為該機制的存取結構(access structure)。一個存取結構中所有最小授權子集所成的集合則稱為該存取結構的基底(basis)。

所謂以圖$G$為基礎的存取結構是指將圖$G$中的每個點視為一個參與者而且以圖$G$的邊集合為基底的存取結構。在秘密分享的問題中被廣為討論的訊息比率(information ratio)與平均訊息比率(average information ratio)則分別定義為參與者所分到的share 的最大長度與秘密的長度的比值，以及所有參與者所分到的share 的平均長度與秘密的長度的比值。一個存取結構上所能構造出的所有秘密分享機制的(平均)訊息比率的infimum則稱為該存取結構的最佳(平均)訊息比率(optimal (average) information ratio)。在此論文中我們要探討的是以圖為基礎的存取結構的最佳平均訊息比率的問題。

首先我們討論權重門檻型的秘密分享機制。給定一個門檻(threshold) $t > 0$ 與一個定義在參與者集合上的權重函數，若一子集中所有參與者的權重和不小於給定的門檻 $t$，則該子集即為一個授權子集。這種授權子集所成的存取結構稱為權重門檻型的存取結構。Morillo等人研究了可以用圖表示的權重門檻型的存取結構並將此種圖稱為一個 $k-$權重圖($k-$weighted graph)。他們清楚刻劃了這種圖的結構並推導了這種存取結構的最佳訊息比率的一個上限。在本論文的第二章中，我們將探討 $k-$權重圖的最佳平

均訊息比率的問題。我們提出兩種秘密分享機制的構造方法並推導它們的平均訊息比率的範圍。兩種構造方式的平均訊息比率都很低，且各有各的優點。當參與者的個數趨近無窮時，我們構造的秘密分享機制的平均訊息比率會趨近於最佳值 1。

由於推導最佳訊息比率與最佳平均訊息比率是相當困難的問題，所以大部分的結果都是提供上下限。在2007年之前被求出最佳訊息比率與最佳平均訊息比率的無窮圖類只有paths和cycles，以及Blundo等人定義出的一種圖類。Csirmaz 與 Tardos 在2007年求出了所有樹圖的最佳訊息比率的正確值。而 Csirmaz 與 Ligeti 則在2009年求得了更廣的圖類的最佳訊息比率正確值。

在本論文的第三章與第四章中，我們則是致力於最佳平均訊息比率的正確值的研究。我們將在論文的第三章提出我們求出以圖為基礎的存取結構的最佳平均訊息比率的正確值的做法。我們的方法將這個秘密分享方面的複雜問題數學模式化為圖論上用簡單語言便能描述的max-min的問題。我們利用這個方法求出所有樹圖的最佳平均訊息比率的正確值，並提供一個有系統的方法求出該值。

而在第四章中，我們更進一步討論二部圖(bipartite graph) 的最佳平均訊息比率的問題。我們求出一個簡單圖的任意 even-subdivision 與一些二部圖類的最佳平均訊息比率的正確值。同時，對於尚未被求出最佳平均訊息比率的正確值的二部圖，我們也推導了一個最佳平均訊息比率的上下限。對一些圖類而言，我們的給上下限是用我們的做法可以得到的最佳上下限。

最後在第五章，我們作了簡短的總整理並介紹未來可以繼續努力的研究方向。

# 謝誌

身兼學生、老師、媽媽與家庭主婦多重身份，在這四年的博士班求學過程中，時間永遠不夠用，壓力是如影隨形地未曾間斷過。若不是我身邊許許多多貴人的大力相助，我無法獨立完成博士學位的追求。所以，我要藉由這篇謝誌衷心感謝這些貴人的相助。

首先要感謝的是我的指導教授傅恆霖教授。與傅老師結緣，是讀博士班之前的事。當時只是透過同事吳順良老師的引介拜訪了傅老師，老師就在許多方面給予很多指導與協助，讓我得以在幾年後順利的憑藉著作升等副教授。甚至在我母親罹癌之後，也是透過傅老師的介紹轉到大林慈濟醫院的胸腔科權威賴醫師的門診，使我的母親可以在她最後幾年的生命中獲得良好的生活品質，安詳而有尊嚴的走完人生最後一程。傅老師對我的這份恩情已是無以為報。進入博士班就讀之後，老師對我有更多的指導、協助與照顧。在老師門下，我體認到老師不只是學術上的巨擘，也是最佳的人生導師。老師不只在專業知識上給予很多教導與建議，老師無形的身教更深深的影響著我們家族中的每個學生。我們學習到的不只是做研究的視野，還有人生的態度與面對事物的謙卑。對老師一直以來的提攜與照顧的恩情，心中的感謝實非言語所能形容。

接下來要感謝翁志文教授，早在我博士入學推甄口試答不出問題時，翁老師就出言相救了。過去四年修了多門翁老師的課，每一門課都獲益良多。翁老師做事親力親為、實事求是的務實態度讓我非常佩服，是我們學習的最佳榜樣。另外，我也要感謝系主任陳秋媛教授。雖然沒機會修到陳老師的課，但老師關愛學生與做事認真、細心、堅持的態度讓我非常敬佩。感謝陳主任如此用心地把應數系經營得這麼好，讓我有個愉快的博士求學歷程。還要感謝的是同事吳順良教授，當初要不是他的催促與幫忙並

引薦我認識傅老師，讓碩士畢業好幾年的我重新走上研究之路，我無法在現在完成我攻讀博士的理想。還有系辦公室優雅的助理們盈吟、麗君與慧珊，她們總是在我面對討厭的行政程序時大力幫忙，讓我節省很多麻煩，真是感謝。

此外，也要感謝家族中許多學弟妹們，惠蘭、貓頭、敏筠、智懷、軒軒...等人。由於我對學校很多事情都不甚清楚，學弟妹們不僅跟我討論研究上的問題，更是我學校事務的最佳諮詢者，他們總是在我最需要幫忙時給予最大的協助，讓我省掉很多摸索的時間。除了家族學弟妹，還要特別感謝在博士班入學口試時認識的光祥。在博一博二一起修課的時候，光祥真的給我很多及時的協助與建議，讓碩士時不是離散領域的我得以在初入這領域研讀時能順利跟上他們，並且得到最佳的學習效果。跟光祥一起經歷了入學口試、修課與準備兩門資格考，總覺得跟光祥有特殊的”革命情感”，他幫忙我學到很多。除了光祥之外，還要感謝當初也一起準備資格考的鈺傑、忠逵與智懷，跟你們的討論給我很多啟發，讓我對所學知識有更深的了解，很懷念當初一起討論的時光。

最後，當然要感謝我親愛的家人。先要感謝我兩個可愛的孩子志群與季晴，你們的乖巧懂事，讓媽媽不用為你們煩心，可以安心的追求理想。你們也是我努力的最大動力來源。而貼心的女兒季晴，永遠都是我最佳的小秘書，論文口試的投影片多虧有她的幫忙才得以及時完成，而且她總是跟媽媽一樣”龜毛”的細心地做到最好。此外，最感謝的還是我的先生心漢，他無條件的包容與支持，是我向前努力的堅強後盾。在我蠟燭兩頭燒時，他為我分擔家務與工作；在我壓力太大情緒低落時，他給我最佳的安慰與倚靠的肩膀，讓我有了再重新振作的力量。我不知道上輩子種了什麼善因，讓老天賜給我這個無可挑剔的人生伴侶！

僅將這本博士論文獻給我的家人，還有我在天上的父母，謝謝您們賦予我充分的智慧與堅強的意志—女兒博士畢業了！

# Contents

# List of Figures

# Chapter 1

# Introduction

Originally motivated by the problem of secure information storage, secret-sharing schemes have found numerous applications in cryptography and distributed computing such as access control, attribute-based encryption and secure multiparty computations. A *secret-sharing scheme* involves a dealer who has a secret, a finite set $\mathcal{P}$ of participants and a collection $\Gamma$ of subsets of $\mathcal{P}$ called the *access structure*. Each subset in $\Gamma$ is a qualified subset. A secret-sharing scheme is a method by which the dealer distributes a secret among the participants in $\mathcal{P}$ such that only the participants in a qualified subset can recover the secret from the shares they received. If, in addition, the joint share of the participants in any unqualified subset is statistically independent of the secret, then the secret-sharing scheme is called *perfect*. We will use "secret-sharing scheme" for "perfect secret-sharing scheme" since only perfect ones are considered in the thesis. An access structure is naturally required to be *monotone*, that is, any subset of $\mathcal{P}$ containing a qualified subset must also be qualified. Therefore, an access structure is completely determined by the family of its minimal subsets. This family of the minimal subsets in $\Gamma$ is called the *basis* of $\Gamma$.

Shamir [31] and Blakley [3] independently introduced the first kind of secret-sharing schemes called the $(t, n)$-*threshold schemes* in 1979. In such a scheme, the basis of the access structure consists of all $t$-subsets of the participant set of size $n$. Their work has raised a great deal of interest in the research of many aspects of secret-sharing problems. Related problems have received considerable attention since then. Secret-sharing schemes for various access structures as well as many modified versions with additional capacities were widely studied [11, 12, 19, 21, 22, 24, 29]. The *information ratio* and the *average information ratio* of secret-sharing schemes have long been the main subjects of discussion. The information ratio of a secret-sharing scheme is the ratio of the maximum length (in bits) of the share given to a participant to the length of the secret, while the average information ratio of a secret-sharing scheme specifies the ratio of the average length of the shares given to the participants to the length of the secret. These ratios respectively represent the maximum and the average number of bits a participant has to remember for each bit of the secret. As opposed to them, some literature uses information rate and average information rate which are exactly the reciprocal of the information ratio and the average information ratio respectively. For lower storage and communication complexity, these ratios are expected to be as low as possible. The question of constructing secret-sharing schemes with the lowest ratios arose naturally. Given an access structure $\Gamma$, the infimum of the (average) information ratio of all possible secret-sharing schemes realizing this access structure $\Gamma$ is referred to as the *optimal (average) information ratio* of $\Gamma$. It has been shown that, for general access structures, the infimum is not always a minimum [2]. The reader is referred to [1] and its references for a comprehensive survey and recent developments in secret-sharing. Secret sharing has been an interesting branch of modern cryptography.

## 1.1 Preliminaries

Let $\mathcal{P}$ be the set of all participants and $\Gamma \subseteq 2^{\mathcal{P}}$ be the access structure. We use $\Gamma_0$ to denote the basis of $\Gamma$. Then $\Gamma$ is called the *closure* of $\Gamma_0$, written $\Gamma = Cl(\Gamma_0)$. Let $\mathcal{K}$ be the set of all secrets and $S$ be the set of all possible shares. Given a secret $d \in \mathcal{K}$, a dealer $D$ gives to participant $p$ a share $s \in S_p$ where $S_p$ is the set of all shares participant $p$ receives from the dealer corresponding to all secrets in $\mathcal{K}$. A *distribution rule* is a function $f : \{D\} \cup \mathcal{P} \to \mathcal{K} \cup S$ with $f(D) \in \mathcal{K}$ and $f(p) \in S$ for all $p \in \mathcal{P}$. $f(D)$ is the secret to be distributed and $f(p)$ is the share participant $p$ receives from the dealer for secret $f(D)$. Let $\mathcal{F}$ be a collection of distribution rules and $\mathcal{F}_d = \{f \in \mathcal{F} : f(D) = d\}$. We call $\mathcal{F}$ a perfect secret-sharing scheme if the following two conditions are satisfied:

i) Given any $B \in \Gamma$ and $f, g \in \mathcal{F}$, if $f(p) = g(p)$ for all $p \in B$, then $f(D) = g(D)$.

ii) Given any $B \notin \Gamma$ and any function $g : B \to S$, there exists a nonnegative integer $\lambda(g, B)$ such that, for each $d \in \mathcal{K}$,

$$|\{f \in \mathcal{F}_d | f(p) = g(p), \forall p \in B\}| = \lambda(g, B).$$

The first condition guarantees that the shares given to a qualified subset uniquely determine the secret. The second ensures that the shares given to an unqualified subset reveal no information about the secret. When these two conditions are made, we say that this secret-sharing scheme $\mathcal{F}$ realizes the access structure $\Gamma$. Since all schemes mentioned in this thesis are perfect, we will simply use “secret-sharing scheme” for “perfect secret-sharing scheme” throughout. The information ratio of the secret-sharing scheme $\mathcal{F}$, denoted as $R_{\mathcal{F}}$, is defined as

$$R_{\mathcal{F}} = \frac{\max\{\log_2 |S_p| : p \in \mathcal{P}\}}{\log_2 |\mathcal{K}|}$$

and the average information ratio of $\mathcal{F}$, written as $AR_{\mathcal{F}}$, is

$$AR_{\mathcal{F}} = \frac{\sum_{p\in\mathcal{P}} \log_2 |S_p|}{|\mathcal{P}| \log_2 |\mathcal{K}|}.$$

The optimal information ratio and the optimal average information ratio of the access structure $\Gamma$ are denoted as $R(\Gamma)$ and $AR(\Gamma)$, respectively. It is well known that $R(\Gamma) \geq AR(\Gamma) \geq 1$ and that $R(\Gamma) = 1$ if and only if

$AR(\Gamma) = 1$. A secret-sharing scheme with information ratio equal to one is then called an *ideal* secret-sharing scheme. An access structure is said to be ideal if there exist an ideal secret-sharing scheme for it.

**Example 1.1.1.** Consider the case where the set of participants $\mathcal{P} = \{a, b, c\}$, the basis of the access structure $\Gamma_0 = \{\{a, b\}, \{b, c\}\}$ and the set of secret $\mathcal{K} = GF(3)$. Define the set of distribution rules as $\mathcal{F} = \{f_{r,d} | r, d \in GF(3)\}$ where $f_{r,d}(D) = d$, $f_{r,d}(a) = f_{r,d}(c) = r$ and $f_{r,d}(b) = r + d$, then this scheme can be represented by the following table:

| $D$ | $a$ | $b$ | $c$ |
|---|---|---|---|
| 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 1 |
| 0 | 2 | 2 | 2 |
| 1 | 0 | 1 | 0 |
| 1 | 1 | 2 | 1 |
| 1 | 2 | 0 | 2 |
| 2 | 0 | 2 | 0 |
| 2 | 1 | 0 | 1 |
| 2 | 2 | 1 | 2 |

Note that each row in the table represents a distribution rule. One can easily check that this scheme is a secret-sharing scheme and $R_{\mathcal{F}} = AR_{\mathcal{F}} = 1$ since $\mathcal{K} = S_a = S_b = S_c = GF(3)$. This scheme is in fact an ideal one. Therefore, $Cl(\Gamma_0)$ is an ideal access structure.

Reseachers have characterized many kinds of ideal access structures by taking advantage of the theory of matroid and linear algebra [8, 25, 26, 27]. In this thesis, we only consider graph-based access structures.

## 1.2 Graph-Based Access Structures

These structures have been widely studied during the past decades. In such an access structure, each vertex of a graph $G$ represents a participant and each edge represents a minimal qualified subset, that is, $\mathcal{P} = V(G)$ and $\Gamma =$

$Cl(E(G))$. We shall introduce another definition of secret-sharing scheme next. The equivalence of this definition and the previous one has been shown in [1]. The information ratio and the average information ratio of a secret-sharing scheme can then be defined alternativly in a way that is especially convient for deriving lower bounds on $R(G)$ and $AR(G)$.

A secret-sharing scheme $\Sigma$ for the access structure based on $G$ is a collection of random variables $\zeta_S$ and $\zeta_v$ for $v \in V(G)$ with a joint distribution such that

(i) $\zeta_S$ is the secret and $\zeta_v$ is the share of $v$;

(ii) if $uv \in E(G)$, then $\zeta_u$ and $\zeta_v$ together determine the value of $\zeta_S$;

(iii) if $A \subseteq V(G)$ is an independent set in $G$, then $\zeta_S$ and the collection $\{\zeta_v | v \in A\}$ are statistically independent.

Before introducing the alternative definition of the (average) information ratio, we recall some basic property of the Shannon entropy function. Given a discrete random variable $X$ with possible values $\{x_1, x_2, \ldots, x_n\}$ and a probability distribution $\{p(x_i)\}_{i=1}^{n}$, the Shannon entropy of $X$ is defined as $H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i)$ which is a measure of the average uncertainty associated with $X$. It holds that $0 \leq H(X) \leq \log |X|$. Note that $H(X)$ takes its minimum value 0 if there is a value $x_i$ of $X$ with $p(x_i) = 1$ and it attains its maximum value $\log |X|$ if $p$ is a uniform distribution [17]. Let us assume the probability distributions involved are uniform. Then the information ratio of the scheme $\Sigma$ can be defined as $R_\Sigma = \max_{v \in V(G)}\{H(\zeta_v)/H(\zeta_S)\}$ and the average information ratio of $\Sigma$ is $AR_\Sigma = (\sum_{v \in V(G)} H(\zeta_v))/(|V(G)|H(\zeta_S))$. For simplicity, with the same symbol $G$, we will denote both the graph as well as the access structure based on it. For example, "a secret-sharing scheme on $G$" refers to "a secret-sharing scheme for the access structure based on $G$". Furthermore, the optimal information ratio, $R(G)$, of $G$ and the optimal average information ratio, $AR(G)$, of $G$ are the infimum of the information ratio $R_\Sigma$ and the average information ratio $AR_\Sigma$ over all possible secret-sharing schemes $\Sigma$ on $G$ respectively. Then one has that $R(G) \geq AR(G) \geq 1$

[13] and that $R(G) = 1$ if and only if $AR(G) = 1$. A secret-sharing scheme $\Sigma$ on $G$ with the optimal ratio $R_\Sigma = 1$ or $AR_\Sigma = 1$ is then called *ideal*. An access structure $G$ is ideal if there exists an ideal secret-sharing scheme on it.

The ideal graph-based access structures have been completely characterized in terms of matroid by Brickell and Devenport .

**Theorem 1.2.1** ([8]). *Suppose that $G$ is a connected graph, then $R(G) = AR(G) = 1$ if and only if $G$ is a complete multipartite graph.*

The basis of the access structure in Example 1.1.1 is in fact the complete multipartite graph $K_{1,2}$. This also shows that $R(K_{1,2}) = 1$.

## 1.3 Approaches to the Derivation of Bounds on the Ratios

In this section, we introduce the main tools for deriving upper bounds and lower bounds on $R(G)$ and $AR(G)$ for non-ideal graph-based access structures.

### 1.3.1 The Derivation of Upper Bounds

By constructing a secret-sharing scheme $\Sigma$ on a graph $G$, we naturally have an upper bound $R_\Sigma$ ($AR_\Sigma$) on the optimal (average) information ratio of $G$. Stinson [34] has proposed a very useful method for constructing secret-sharing schemes for a graph from its *complete multipartite covering*. A complete multipartite covering of a graph $G$ is a collection (multiset) $\Pi = \{G_1, G_2, \ldots, G_l\}$ of complete multipartite subgraphs of $G$ such that each edge of $G$ belongs to at least one subgraph in this collection. Since ideal secret-sharing schemes on all $G_i$'s are known, each vertex (participant) receives a share from the secret-sharing scheme constructed on each $G_i$ containing this vertex. Stinson's ideal is to obtain the share of a vertex in the secret-sharing scheme for the whole graph by joining together the shares the vertex receives from all secret-sharing schemes on the complete multipartite subgraphs containing it in the covering. This method has been a major tool for the derivatin of upper bounds on the optimal (average) information ratio of a graph. Let us introduce some important parameters of a complete multipartite covering of a graph before stating Stinson's method. The *occurrence* $t_e$ of an edge $e$ in the covering $\Pi$ is defined as $t_e = |\{j|e \in E(G_j)\}|$ and the occurrence $r_v$ of a vertex $v$ is $r_v = |\{j|v \in V(G_j)\}|$. The *minimum edge occurrence* of a covering $\Pi$ is the minimum occurrence of an edge in $\Pi$, denoted as $t_\Pi$, and the *maximum vertex occurrence* of a covering $\Pi$ is the maximum occurrence of a vertex in $\Pi$, denoted as $r_\Pi$. In dealing with the average information ratio, the most important concern is the total occurrences of all vertices in $\Pi$. This number also represents the total of the vertex numbers of all subgraphs in this covering. We call it the *vertex-number sum* of the covering $\Pi$, written as $m_\Pi = \sum_{i=1}^{l} |V(G_i)|$.

**Theorem 1.3.1** ([34]). *Suppose that $\Pi = \{G_1, G_2, \ldots, G_l\}$ is a complete multipartite covering of a graph $G$ with $|V(G)| = n$. Then there exists a secret-sharing scheme $\Sigma$ on $G$ with information ratio $R_\Sigma$ and average information ratio $AR_\Sigma$ where*

$$R_\Sigma = r_\Pi / t_\Pi \text{ and } AR_\Sigma = \frac{1}{t_\Pi n} \sum_{v \in V(G)} r_v = \frac{m_\Pi}{t_\Pi n}.$$

This theorem suggests that in order to construct a secret-sharing scheme with lower information ratio, we need a complete multipartite covering with less maximum vertex occurrence and larger minimum edge occurrence. However, the problem of how many copies of each complete multipartite subgraph of $G$ should we use to compose a covering(multiset) in order to reach to the optimal value of the ratio $r_\Pi/t_\Pi$ is a crucial issue to handle. Linear programming technique plays an important role in solving this problem. We introduce the approach by Stinson [34] which is a modification of the version by Blundo et.al [7].

Let $\mathcal{L} = \{G_1, G_2, \ldots, G_h\}$ be the collecction of all complete multipartite subgraphs of $G$. For $v \in V(G)$, $e \in E(G)$ and $i = 1, 2, \ldots, h$, define

$$c_{vi} = \begin{cases} 1, & \text{if } v \in V(G_i); \\ 0, & \text{if } v \notin V(G_i) \end{cases}$$

and

$$b_{ei} = \begin{cases} 1, & \text{if } e \in E(G_i); \\ 0, & \text{if } e \notin E(G_i). \end{cases}$$

Suppose we construct a covering using $\alpha_i$ copies of $G_i$, for $i = 1, 2, \ldots, h$. Then we have $t_\Pi = \min_{e \in E(G)}\{\Sigma_{i=1}^h \alpha_i b_{ei}\}$ and $r_\Pi = \max_{v \in V(G)}\{\Sigma_{i=1}^h \alpha_i c_{vi}\}$. The secret-sharing scheme $\Sigma$ constructed via the covering has information ratio $R_\Sigma = r_\Pi / t_\Pi$. Since taking a scalar multiple of all the $\alpha_i$'s does not affect the value of the ratio, we may allow the $\alpha_i$'s to be nonnegative rationals and "normalize" them by stipulating that

$$\max_{v \in V(G)} \{\Sigma_{i=1}^h \alpha_i c_{vi}\} = 1.$$

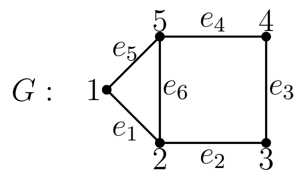
Then our objective is to maximize $t_\Pi$. The linear programming problem can describe as follows.
(*)Maximize $R$ subject to

$$\begin{array}{ll} \alpha_i \geq 0, & 1 \leq i \leq h \\ \Sigma_{i=1}^h \alpha_i c_{vi} \leq 1, & v \in V(G) \\ \Sigma_{i=1}^h \alpha_i b_{ei} \geq R, & e \in E(G) \end{array}$$

By solving this linear programming problem, the optimal solution will involve rational values of $\alpha_i$'s. We can make all the $\alpha_i$'s integral by multiplying an appropriate integer. Then take the resulting integral combination of the $G_i$'s as the covering. We demonstrate this process in the following example.

**Example 1.3.2.** Consider the access structure based on the graph $G$ depicted below.

The list $\mathcal{L}$ of complete multipartite subgraphs consists of the subgraphs $G_i$'s induced by the following sets of edges, respectively.




$E(G_i) = \{e_i\}, i = 1, 2, \ldots, 6$
$E(G_{6+i}) = \{e_i, e_{i+1}\}, i = 1, 2, \ldots, 5$
$E(G_{12}) = \{e_1, e_5\}$
$E(G_{13}) = \{e_1, e_6\}$
$E(G_{14}) = \{e_2, e_6\}$
$E(G_{15}) = \{e_4, e_6\}$
$E(G_{16}) = \{e_1, e_2, e_6\}$
$E(G_{17}) = \{e_1, e_5, e_6\}$
$E(G_{18}) = \{e_4, e_5, e_6\}$
$E(G_{19}) = \{e_2, e_3, e_4, e_6\}$

The optimal solution to the linear programming problem(*) is

$$\alpha_i = \begin{cases} 1/3, & \text{if } i \in \{3, 7, 10, 17, 19\}; \\ 0, & \text{otherwise} \end{cases}$$

and $R = 3/2$. In this case, we have the desired covering $\Pi$ consisting of one copy of $G_3, G_7, G_{10}, G_{17}$ and $G_{19}$. One can easily check the fact that $t_\Pi = 2$ and $r_\Pi = 3$.

Besides these major approaches, there are other results that may sometimes serve as good tools in deriving upper bounds on $R(G)$.

**Lemma 1.3.3** ([9]). *Suppose that $u$ and $v$ are two vertices of a graph $G$ who have the same neighbors, then $R(G) = R(G - v)$.*

Complete multipartite coverings with $t_\Pi > 1$ are easpecially helpful when dealing with information ratio, whereas they do not necessarily lead to good results for average information ratio. In our approach, we use covering with $t_\Pi = 1$. In this case, complete multipartie coverings with less vertex-number sum are what we are aiming for in finding a good upper bound on $AR(G)$.

In the case when $G$ is of girth not less than five, the stars are the only possible subgraphs to use in a complete multipartite covering. A complete multipartite covering in which each subgraph is a star is called a *star covering*. A star covering is indeed most useful for graphs of larger girth. It in general does not result in the least vertex-number sum for a graph of girth less than five. In Chapter 3 and 4, we are dealing with graphs with larger girth. A suitable star covering is our main tool to establish upper bounds on $AR(G)$.

### 1.3.2 The Derivation of Lower Bounds

Finding lower bounds on the opitaml (average) information ratio is generally much more challenging. The only known tool to do this job is the information theoretic approach [4, 13]. Lower bounds are obtained by manipulating information equalities and inequalities. Adopting the result in [10], Blundo et al.[7] showed the following result.

**Theorem 1.3.4** ([7])**.** *Let $G$ be a graph with $V(G) = \{v_i | i = 1, 2, \ldots, 4\}$. If $v_1v_2, v_2v_3, v_3v_4 \in E(G)$ and $v_1v_4, v_1v_3 \notin E(G)$. Then $R(G) \geq 3/2$.*

van Dijk also used the this approach to characterize graphs of order six whose information ratio is not less than 5/3.

**Theorem 1.3.5** ([35])**.** *Let $G$ be a graph with $V(G) = \{v_i | i = 1, 2, \ldots, 6\}$. If $G$ satisfies both*

*(i)* $v_1v_2, v_3v_4, v_5v_6 \in E(G)$ *and*

*(ii)* $v_1v_5, v_1v_6, v_2v_5, v_2v_6, v_3v_5, v_3v_6 \notin E(G)$

*and at least one of the following conditions.*

- $v_2v_4, v_4v_6 \in E(G)$,

- $v_2v_3, v_3v_4 \in E(G)$,
- $v_2v_3, v_2v_4 \in E(G)$, *or*
- $v_3v_4, v_2v_4 \in E(G)$.

*Then* $R(G) \geq 5/3$.

When dealing with information ratio, the following lemma is especially helpful.

**Lemma 1.3.6** ([7]). *If* $G'$ *is an induced subgraph of a graph* $G$*, then* $R(G) \geq R(G')$.

Theorem 1.2.1 guarantees that the ideal graph-based access structures are exactly the complete multipartite graphs. By Theorem 1.3.4 and Lemma 1.3.6, the result for graphs which are not complte multipartite follows.

**Theorem 1.3.7** ([7]). *Suppose that* $G$ *is a connected graph which is not complete multipartite, then* $R(G) \geq \frac{3}{2}$ *and* $AR(G) \geq \frac{n+1}{n}$ *where* $n = |V(G)|$.

It shows that there is a gap in the inforamtion ratio between the ideal and non-ideal graph-based access structures.

In addition to these results, Blundo et al.[7] defined a so-called "*fundation*" of a graph to cope with the optimal average information ratio of graphs. The fundation of a graph $G$ is a subgraph $G_0$ of $G$ which satisfies (i) $xy \in E(G_0)$ if and only if there exist vertices $w, z \in V(G)$ such that the subgraph induced by $\{w, x, y, z\}$ has edge set $\{wx, xy, yz\}$ or $\{wx, xy, yz, xz\}$ and (ii) the edge set of $G_0$ consist of all vertices in $V(G)$ which are incident with at least one edge in $E(G)$. Then, they considered the linear programming problem.

(**)Minimize $C = \Sigma_{v \in V(G)} a_v$ subject to

$$\begin{array}{ll} a_v \geq 0, & v \in V(G) \\ a_v + a_w \geq 1, & vw \in V(G_0) \end{array}$$

They obtain a lower bound with the optimal solution $C^*$ to this linear programming problem.

**Theorem 1.3.8** ([7]). *Let $G_0$ be the fundation of a graph $G$ and $C^*$ be the optimal solution to the linear programming problem (**). Then*

$$AR(G) \geq \frac{C^* + |V(G)|}{|V(G)|}$$

.

Csirmaz [13] put the information theoretic approach in a neater way which is what we place much reliance on in Chpater 3.

Let $\Sigma$ be a secret-sharing scheme in which $\zeta_S$ is the random variable of the secret and each $\zeta_v$ is the random variable of the share of $v$, $v \in V(G)$. Define a real-valued function $f$ as $f(A) = H(\{\zeta_v : v \in A\})/H(\zeta_S)$ for each subset $A \subseteq V(G)$, where $H$ is the Shannon entropy. Then, $R_\Sigma = \max_{v \in V(G)} f(v)$ and $AR_\Sigma = \frac{1}{n}\sum_{v \in V(G)} f(v)$, where $n = |V(G)|$. Using properties of the entropy function and the definition of a secret-sharing scheme, one can show that $f$ satisfies the following inequalities [13]:

(a) $f(\emptyset) = 0$, and $f(A) \geq 0$;

(b) if $A \subseteq B \subseteq V(G)$, then $f(A) \leq f(B)$;

(c) $f(A) + f(B) \geq f(A \cap B) + f(A \cup B)$;

(d) if $A \subseteq B \subseteq V(G)$, $A$ is an unqualified set and $B$ is not, then $f(A)+1 \leq f(B)$;

(e) if neither $A$ nor $B$ is unqualified but $A \cap B$ is, then $f(A) + f(B) \geq 1 + f(A \cap B) + f(A \cup B)$.

A subset $V_0$ of $V(G)$ is called *connected* if it induces a connected subgraph of $G$. Csirmaz and Tardos [16] defined a *core* $V_0$ of a graph $G$ as a connected subset $V_0$ of $V(G)$ satisfies that (i) each $v \in V_0$ has a neighbor $\bar{v}$ outside $V_0$ and is not adjacent to any other vertices in $V_0$ and (ii) $\{\bar{v}|v \in V_0\}$ is an independent set in $G$. The neighbor $\bar{v}$ in the definition is referred to as the *designated outside neighbor* of $v$ throughout this thesis. By employing inequalities (a) to (e), they showed the following result .

**Theorem 1.3.9** ([16])**.** *Let $V_0$ be a core of a graph $G$. If $f$ is defined as above, then $\sum_{v \in V_0} f(v) \geq 2|V_0| - 1$.*

Based on this fact, we will derive a lower bound on $AR(G)$ and rewrite Theorem 1.3.1 as an upper bound on $AR(G)$ of particular form in Chapter 3. Our approach to determining the exact value of $AR(G)$ will then be introduced.

## 1.4 Known Reults on $R(G)$ and $AR(G)$

For non-ideal graphs, Stinson's [34] bound has been shown to be the best for general graphs among known upper bounds on $R(G)$. The complete multipartite covering he used was a star covering. For a general graph $G$, let $S_v$ be the star on vertex set $\{v\} \bigcup N_G(v)$ having center $v$. Then $\Pi = \{S_v | v \in V(G)\}$ form a star covering with minimum edge occurrence 2 and maximum vertex occurrence $d+1$. By Theorem 1.3.1, Stinson [34] improved previous results and showed that $R(G) \leq \frac{d+1}{2}$ where $d$ is the maximum degree of $G$ and $AR(G) \leq \frac{2m+n}{2n}$ where $n = |V(G)|$ and $m = |E(G)|$. Blundo et al [4] defined an infinite class of graphs $H_n$ and use the information theoretic approach to show that $R(H_n) \geq \frac{d+1}{2}$. This result shows that Stinson's result on $A(G)$ is tight. In addition, Stinson's upper bound on $AR(G)$ is also the best for general graph so far.

Due to the difficulty of the derivation of good results on general graphs, most efforts have been focused on small graphs [7, 23, 32, 33, 34, 35] and graphs with better structures [4, 7, 15, 17, 34]. Stinson [32, 33, 34], van Dijk[35] and Blundo et al. [7] used various combinations of the methods described in Section 1.3 to derive the exact velues or bounds on $R(G)$ for all graphs of order not less than six. Stinson [32, 33, 34] and Blundo et al. [7]

have also found the exact velues or bounds on $AR(G)$ for all graphs of order not less than five.

Let $C_n$ and $P_n$ be the cycle and the path of length $n$, respectively. Stinson [34] showed that $R(C_n) = 3/2$ for $n \geq 5$ and $R(P_n) = 3/2$ for $n \geq 3$, which are direct results from the bound $R(G) \leq \frac{d+1}{2}$ and Theorem 1.3.7. The values of $AR(C_n) = 3/2$ for $n \geq 5$ and $AR(P_n) = \frac{3n+\delta}{2(n+1)}$ for $n \geq 3$ [7], where $\delta = 0$ when $n$ is even and $\delta = 1$ when $n$ is odd, come from constructing suitable star covering (Theorem 1.3.1) and the fundation of the graphs (Therem 1.3.8).

Morillo et al.[28] considered the weighted threshold secret-sharing schemes. This is the case when every participant is given a weight depending on his or her position in an organization. A set of participants is in the access structure if and only if the sum of the weights of all participants in the set is not less than the given threshold. They characterized the wieghted threshold access structure that can be represented by a graph $G_k$ which is called $k$-weighted graphs, and constructed a complet multipartite covering $\Pi_{G_k}$ for $k = 2^q - 1$ with the maximum vertex occurrence $r_{\Pi_{G_k}} = q$. By Lemma 1.3.6, they obtained an upper bound $\lceil \log_2(k+1) \rceil$ on $R(G_k)$ for each value of $k$.

Before 2007, apart from the aforementioned class of graphs $H_n$ defined by Blundo et al.[4], the paths and cycles are the only infinite classes of graphs which have known exact values of the optimal information ratio and the optimal average information ratio. Csirmaz and Tardos's [17] excellent work appeared in 2007. They determined the exact values of the optimal information ratio of all trees as $R(G) = 2 - \frac{1}{c(T)}$, where $c(T)$ is the maximum size of a core in the tree $T$. They showed $R(G) \geq 2 - \frac{1}{c(T)}$ from Theorem 1.3.9 and obtained that $R(G) \leq 2 - \frac{1}{c(T)}$ by constructing a star covering $\Pi$ with minimum edge occurrence $t_\Pi = c(T)$ and maximum vertex occurrence $r_\Pi = 2c(T) - 1$.

By generalizing this approach, Csirmaz and Ligeti [16] made an even greater achievement in 2009. They showed that $R(G) = 2 - 1/d$, where $d$ is the maximum degree of $G$, for any graph $G$ satisfying the following properties: (i) every vertex has at most one neighbor of degree one, (ii) vertices of degree at least three are not connected by an edge, and (iii) the girth of $G$ is at least six. This has been the greatest accomplishment regarding exact values of the information ratio of non-ideal graph-based access structures. During the past decades, the information ratio has apparently attracted a lot more attention than the average information ratio has. This is partly due to the complicated essence of treating the average inforamtion ratio. Despite the complexity, we devote our effort to the discussion of the average information ratio of graphs. Hope to make a contribution to the study of efficiency of secret-sharing schemes.

## 1.5 Overview of the Thesis

As mentioned above, Morillo et al. [28] characterized weighted threshold access structures based on graphs and studied their optimal information ratio. Since these access structures are more applicable in real-life situation, we are motivated to construct better secret-sharing schemes for them and have a more detailed analysis of the average information ratio of our schemes in Chapter 2. We start this chapter with Morillo's characterization of the graphs that represent weighted threshold access structures and the upper bound on $R(G)$ they have derived. We then present an observation on the structure of this kind of graphs. Subsequently, two sophisticated constructions of secret-sharing schemes are proposed and bounds on the average information ratio of these schemes are calculated. A comparison of the efficiency of them will be given in the final section of this chapter.

Next, we engage in the pursuit of the exact values of the optimal average information ratio of graphs in Chapter 3 and 4. We begin with completing the work of Csirmaz and Tardos's [17] on the study of tree-based access structure by determinig the exact values of the optimal average information ratio of all trees in Chapter 3. Extending this result, we deal with bipartite graphs in Chapter 4. We obtain the exact values of the optimal average information ratio of some classes of bipartite graphs. For the rest classes of bipartite graphs, a bound on the optimal average information ratio is provided subsequently. Our bound is the first one regarding the optimal average information ratio of bipartite graphs. This bound is the best possible for some classess of bipartite graphs using our approach. In the final chapter, we summarize our work in this thesis and introduce possible directions of future research.

# Chapter 2

# Average Information Ratio of Weighted Threshold Secret-Sharing Schemes

In this thesis, we only take care of graph-based access structures. The graphs considered in Chapter 2 and 3 are connected. Chapter 4 deals with bipartite graphs which may not be connected. In all chapters, each graph considered contains no isolated vertices.

## 2.1 Weighted Threshold Access Structures

Given a set of $n$ participants $\mathcal{P}$, a threshold $t > 0$ and a weight function $w : \mathcal{P} \to \mathbb{R}$ with $w(p) \geq 0$ for all $p \in \mathcal{P}$, the $(t, n, w)$-weighted threshold access structure consists of all subset $A \subseteq \mathcal{P}$ such that $w(A) = \sum_{p \in A} w(p) \geq t$. Morillo et al. [28] showed that any weighted access structure determined by a non-integer-valued weight function and a non-integer threshold can also be determined by an integer-valued weight function and an integer threshold. Therefore, considering integer-valued weight functions is sufficient in our problem. In the remainder of the chapter, we assume that a weight function $w$ is given. An access structure $\Gamma = Cl(\Gamma_0)$ is called $r$-homogeneous if each subset in $\Gamma_0$ is of size $r$. Throughout this chapter, we consider 2-homogeneous weighted threshold access structure and exclude the case where any participant has zero-weight. This kind of access structure can be represented by a graph $G$. In this graph, there is a set $C$ of vertices, each of which is adjacent to all other vertices in $G$. The weight of each vertex in $C$ is higher than the weight of any vertex not in $C$. If $C \neq V(G)$, removing $C$ from the graph $G$ produces a nonempty set $A$ of isolated vertices, each of which has lower weight than any other vertex not in $A$. If $C \cup A \neq V(G)$, the subgraph $G'$ induced by $V(G)\backslash(C \cup A)$ represents a 2-homogeneous weighted threshold access structure $\Gamma' = \{B \subseteq \mathcal{P}\backslash(C \cup A) | w(B) \geq t\}$. By repeating this process, Morillo et al. has a clear characteriztion of the structure of $G$ in the following theorem.

**Theorem 2.1.1** ([28]). *Let $G$ be a graph that represents the 2-homogeneous weighted threshold access structure $\Gamma$. Then, there exists a unique partition of the vertices of $G$,*

$$P = C_1 \cup A_1 \cup C_2 \cup A_2 \cup \cdots \cup C_k \cup A_k,$$

*where $C_i \neq \emptyset$ for $i = 1, \ldots, k$, $A_i \neq \emptyset$ if $i = 1, \ldots, k-1$ and either $A_k = \emptyset$ and $|C_k| \geq 2$ or $|A_k| \geq 2$, such that the set of edges of $G$ is*

$$\Gamma_0 = \left\{ \{u, v\} \middle| u, v \in \bigcup_{i=1}^{k} C_i, u \neq v \right\} \cup \{\{v, p\} | v \in C_i, p \in A_j, 1 \leq i \leq j \leq k\}.$$

They also showed that any graph with a partition described in Theorem 2.1.1 represents a 2-homogeneous weighted threshold access structure. Such a graph is then called *$k$-weighted* where $k$ is the parameter used in Theorem 2.1.1. Since the structure of a $k$-weighted graph is completely determined by the values $|A_i|$'s and $|C_i|$'s, $i = 1, 2, \ldots, k$, we denote the $k$-weighted graph by $W(|A_1|, \ldots, |A_k|, |C_1|, \ldots, |C_k|)$. Observe that the subgraph induced by $\bigcup_{i=1}^{l}(A_{j_i} \cup C_{j_i})$ where $1 \leq j_1 < j_2 < \cdots < j_l \leq k$ is an $l$-weighted graph $W(|A_{j_1}|, \ldots, |A_{j_l}|, |C_{j_1}|, \ldots, |C_{j_l}|)$. Morillo et al. gave a complete multipartite decomposition for $(2^q - 1)$-weighted graph of which the minimum edge occurrence is one and the maximum vertex occurrence is not greater than $q$.

Then, by Lemma 1.3.6, a lower bound on the optimal information ratio for $k$-weighted graph, for all $k$, follows.

**Theorem 2.1.2** ([28]). *Let $\Gamma = \{A \subseteq \mathcal{P} | w(A) \geq t\}$ be an access structure that is represented by a $k$-weighted graph $G$. Then $R(G) \leq \lceil \log_2(k+1) \rceil$.*

While dealing with information ratio, one can obtain upper bound of a graph from its subgraph using Lemma 1.3.6. However, for the average information ratio, we do not have the advantage to take. The complete multipartite covering must be constructed for each value of $k$. For convenience, we make a slight modification to the notation given in Theorem 2.1.1. In the case where $A_k = \emptyset$ and $|C_k| \geq 2$, we move one (arbitrarily chosen) vertex from $C_k$ to $A_k$. Thus, none of $A_i$'s and $C_i$'s are empty in our model. Next, we will present an observation on the construction of $k$-weighted graphs before introducing our constructions in the following sections.

## 2.2 An Observation

We observe that any $k$-weighted graph can be obtained by alternately applying two graph operations starting with a single vertex. Let us introduce these operations first. By "*splitting* vertex $v$ of a graph $G$ into $m$ vertices $v_1, \ldots, v_m$", denoted $Spt(v; \{v_1, \ldots, v_m\})$, we obtain a graph $G^{Spt(v;\{v_1,\ldots,v_m\})}$ whose vertex set is $V(G^{Spt(v;\{v_1,\ldots,v_m\})}) = (V(G) - \{v\}) \cup \{v_1, v_2, \ldots, v_m\}$ and the edge set is $E(G^{Spt(v;\{v_1,\ldots,v_m\})}) = E(G - v) \cup \{v_i u | vu \in E(G)$ and $i = 1, 2, \ldots, m\}$. If we further add all edges in $\{v_i v_j | 1 \leq i < j \leq m\}$ to $E(G^{Spt(v;\{v_1,\ldots,v_m\})})$, then we obtain a graph $G^{Exp(v;\{v_1,\ldots,v_m\})}$. This resulting graph is said to be obtained by "*expanding* vertex $v$ into $m$ vertices $v_1, \ldots, v_m$ from the original graph $G$ and this operation is denoted by $Exp(v; \{v_1, \ldots, v_m\})$. In what follows, we use $\langle V_1, V_2 \rangle_G$ to denote the set of edges $\{uv | u \in V_1, v \in V_2$ and $uv \in E(G)\}$ for any two disjoint subsets of vertices $V_1$ and $V_2$ in $G$.

Given a $k$-weighted graph $G = W(a_1, a_2, \ldots, a_k, c_1, c_2, \ldots, c_k)$, where $a_i = |A_i|$ and $c_i = |C_i|$, we let $A_i = \{u_1^i, u_2^i, \ldots, u_{a_i}^i\}$ and $C_i = \{v_1^i, v_2^i, \ldots, v_{c_i}^i\}$,

$i = 1, 2, \ldots, k$. We explain how the given graph can be constructed start with a single vertex by splitting and expandingan in the following algorithm.

***Algorithm;***

$G_0 \leftarrow \{u_0\}$.

For $i \leftarrow 1$ to $k$ do

$$G_i \leftarrow G_{i-1}^{Exp(u_0; C_i \cup \{u_0\})}$$

$$G_i \leftarrow G_i^{Spt(u_0; A_i^*)} \quad \text{where} \quad A_i^* = \begin{cases} A_i \cup \{u_0\}, & \text{if } 1 \le i < k; \\ A_k, & \text{if } i = k. \end{cases}$$

Output the $k$-weighted graph $G_k$.

**Theorem 2.2.1.** *The proposed algorithm produces the given $k$-weighted graph $G$ from a single vertex.*

**Proof.** Observe that the edges in $\langle A_i, C_j \rangle$, $j \le i$, are produced by the operation $Spt(u_0; A_i^*)$ and edges in $\langle C_i, C_j \rangle$, $j < i$, and within the part $C_i$ are all produced by $Exp(u_0, C_i^*)$. So, $G$ is a subgraph of $G_k$. Next, the number of edges produced in this algorithm is

$$\sum_{i=1}^{k-1} \left( \binom{c_i + 1}{2} + c_i \sum_{j=1}^{i-1} c_j + a_i \sum_{j=1}^{i} c_j \right) + \binom{c_k + 1}{2} + c_k \sum_{j=1}^{k-1} c_j + (a_k - 1) \sum_{j=1}^{k} c_j$$
$$= \sum_{i=1}^{k} \left( \binom{c_i + 1}{2} + c_i \sum_{j=1}^{i-1} c_j + a_i \sum_{j=1}^{i} c_j \right) - \sum_{j=1}^{k} c_j$$
$$= \sum_{j=1}^{k} \left( \binom{c_i}{2} + c_i \sum_{j=1}^{i-1} c_j + a_i \sum_{j=1}^{i} c_j \right)$$

which is exactly the size of the given graph $G$. The proof is completed. ∎

## 2.3 Construction (I)

Before we can literally describe our first construction, there are some more notations needed to be introduced. For any $l$ disjoint sets of vertices $V_1, V_2, \ldots, V_l$, we use $K(V_1, V_2, \ldots, V_l)$ to denote the complete multipartite graph with partite sets $V_1, V_2, \ldots$ and $V_l$. Let $G_l = W(|A_1|, \ldots, |A_l|, |C_1|, \ldots, |C_l|)$ be the $l$-weighted graph with vertex set $(\bigcup_{i=1}^{l} A_i) \cup (\bigcup_{i=1}^{l} C_i)$, $l \leq k$. Define $B_l$, $l \leq k$, to be the graph obtained from $G_l$ by removing all edges connecting vertices in $\bigcup_{i=1}^{l} C_i$. Then $B_l$ is a bipartite graph with partite sets $\bigcup_{i=1}^{l} A_i$ and $\bigcup_{i=1}^{l} C_i$. Next, we use $M_{l_1,l_2}$ to denote the complete multipartite graph $K(C_1, C_2, \ldots, C_{l_1-1}, \{v_1^{l_1}\}, \{v_2^{l_1}\}, \ldots, \{v_{c_{l_1}}^{l_1}\}, (\bigcup_{j=l_1+1}^{l_2} C_j) \cup (\bigcup_{j=l_1}^{l_2} A_j))$, $1 \leq l_1 \leq l_2 \leq k$. In what follows, the complete multipartite graph $K(C_1, C_2, \ldots, C_{j-1}, A_{j-1}, A_j)$ is written as $H_j$, $2 \leq j \leq k$.

**Lemma 2.3.1.** $\Pi_l^B$ *is a complete multipartite covering of* $B_l$ *where*

$$\Pi_l^B = \begin{cases} \{H_{2i}, K(A_{2i}, C_{2i}) | i = 1, 2, \ldots, \frac{l}{2}\}, & \text{if } l \text{ is even;} \\ \{K(A_1, C_1), H_{2i+1}, K(A_{2i+1}, C_{2i+1}) | i = 1, 2, \ldots, \frac{l-1}{2}\}, & \text{if } l \text{ is odd.} \end{cases}$$

**Proof.** When $l$ is even, the edges in $\langle A_{2i}, C_j \rangle_{B_l}$ with $j < 2i$ and the edges in $\langle A_{2i-1}, C_j \rangle_{B_l}$ with $j \leq 2i-1$ appear in the subgraph $H_{2i}$, for $i = 1, 2, \ldots, \frac{l}{2}$, while the edges in $\langle A_{2i}, C_{2i} \rangle_{B_l}$ appear in the subgraph $K(A_{2i}, C_{2i})$. The edges of $B_l$ are then all used up. For odd $l$, the argument is similar. ∎

With these notations in mind, we are able to give our complete multipartite covering $\Pi_k$ of $G_k$. Let $\Pi_k$ be obtained recursively by letting $\Pi_1 = \{G_1\}$, $\Pi_2 = \{K(\{v_1^1\}, \{v_2^1\}, \ldots, \{v_{c_1}^1\}, A_1), M_{2,2}\}$, $\Pi_3 = \{K(\{v_1^1\}, \{v_2^1\}, \ldots, \{v_{c_1}^1\}, A_1),$ $K(\{v_1^3\}, \ldots, \{v_{c_3}^3\}, A_3), M_{2,3}\}$ and, for $k \geq 4$, $\Pi_k = \Pi^B_{\lfloor \frac{k+1}{2} \rfloor} \cup \left\{ M_{\lfloor \frac{k+1}{2} \rfloor + 1, k} \right\} \cup \Pi_{\lfloor \frac{k}{2} \rfloor - 1}$ where $\Pi_{\lfloor \frac{k}{2} \rfloor - 1}$ is the complete multipartite covering of the $(\lfloor \frac{k}{2} \rfloor - 1)$-weighted subgraph $W\left(a_{\lfloor \frac{k+1}{2} \rfloor + 2}, a_{\lfloor \frac{k+1}{2} \rfloor + 3}, \ldots, a_k, c_{\lfloor \frac{k+1}{2} \rfloor + 2}, c_{\lfloor \frac{k+1}{2} \rfloor + 3}, \ldots, c_k\right)$.

It can be easily checked that the edges of $G_k$ which are not in $B_{\lfloor \frac{k+1}{2} \rfloor}$ and $W\left(a_{\lfloor \frac{k+1}{2} \rfloor + 2}, \ldots, a_k, c_{\lfloor \frac{k+1}{2} \rfloor + 2}, \ldots, c_k\right)$ all lie in $M_{\lfloor \frac{k+1}{2} \rfloor + 1, k}$. These three subgraphs virtually make up the $k$-weighted graph $G_k$. We have the following lemma.

**Lemma 2.3.2.** *The collection* $\Pi_k$ *stated above is a complete multipartite covering of* $G_k$ *with mimimum edge occurence one.*

Our next goal is to evaluate the vertex-number sum $m_k$ of $\Pi_k$. Due to the complexity of the enumeration, we consider the reduced forms first. We call $G_k^0 = W(1,\ldots,1,1,\ldots,1)$ the *reduced form* of a general $k$-weighted graph $W(a_1,\ldots,a_k,c_1,\ldots,c_k)$. We also let $B_l^0$, $M_{l_1,l_2}^0$ and $H_j^0$ be the graphs defined in the same ways as $B_l$, $M_{l_1,l_2}$ and $H_j$ respectively, except that $a_i$'s and $c_j$'s involved are all set to be one. Then $G_k^0$ and $B_k^0$ have the complete multipartite covering $\Pi_k^0$ and $\Pi_k^{B^0}$ reduced from $\Pi_k$ and $\Pi_k^B$ respectively. Note here that $G_k^0$ has $2k$ vertices. By applying suitable splitting and expanding operations mentioned in Section 2.2 accordingly to the reduced form $G_k^0$, one can recover the general $k$-weighted graph $W(a_1,\ldots,a_k,c_1,\ldots,c_k)$. For the description of the evaluation of the vertex-number sum $m_k^0$ of $\Pi_k^0$, we introduce a specially designed binary tree.

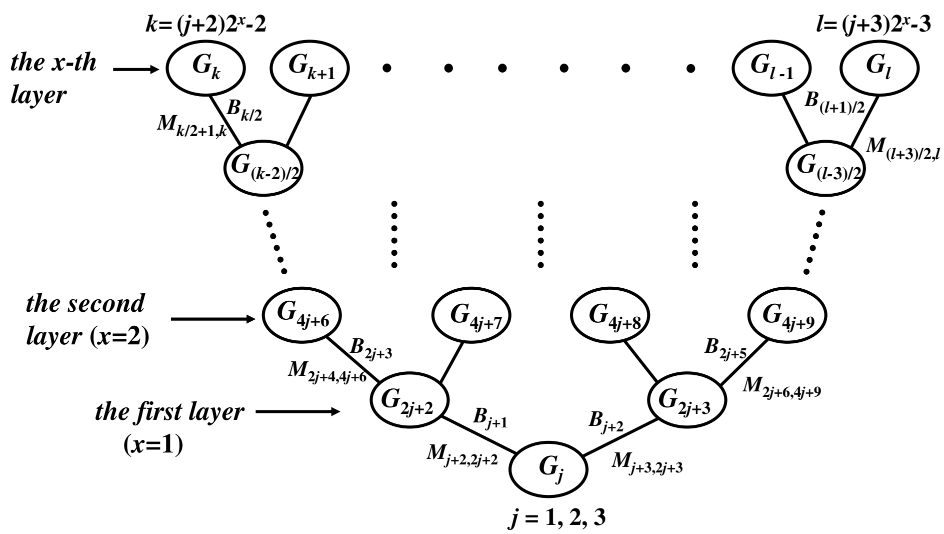



Figure 2.1: The binary tree for Construction (I)

Note that we have decomposed $G_k^0$ into $B^0_{\lfloor\frac{k+1}{2}\rfloor}$, $M^0_{\lfloor\frac{k+1}{2}\rfloor+1,k}$ and $G^0_{\lfloor\frac{k}{2}\rfloor-1}$. Since $\lfloor\frac{k+1}{2}\rfloor$ equals $(\lfloor\frac{k}{2}\rfloor-1)+1$ or $(\lfloor\frac{k}{2}\rfloor-1)+2$, $G_j^0$ can either go with $B_{j+1}^0$ and $M_{j+2,2j+2}^0$ to compose $G_{2j+2}^0$ or go with $B_{j+2}^0$ and $M_{j+3,2j+3}^0$ to compose

$G^0_{2j+3}$. By recursively repeating this process, we observe that all $G^0_k$'s can be built up from some $B^0_l$'s, $M^0_{l_1,k}$'s and just $G_1$, $G_2$ and $G_3$. We illustrate this relation by means of a binary tree in Figure 2.1. In this tree, each path from the root represents the conformation of a $k$-weighted graph of the reduced form in our covering. For example, the leftmost path from the root $G_j$ to $G_{4j+6}$ represents that $G^0_{2j+2}$ is composed of $G^0_j$, $B^0_{j+1}$ and $M^0_{j+2,2j+2}$ and then $G^0_{4j+6}$ is composed of $G^0_{2j+2}$, $B^0_{2j+3}$ and $M^0_{2j+4,4j+6}$. Hence the path shows how $G^0_{4j+6}$ is built up. The $2^x$ paths of length $x$ from the root give the conformations of the $2^x$ $k$-weighted graphs where $k$ ranges from $(j+2)2^x-2$ to $(j+3)2^x-3$, $j=1,2,3$.

**Theorem 2.3.3.** *Let $\Gamma=\{A\subseteq\mathcal{P}|w(A)\geq t\}$ be an access structure represented by a $k$-weighted graph $G^0_k$ of reduced form, $k_1=(j+2)2^x-2$ and $k_2=(j+3)2^x-3$, $x\geq 1$, $j=1,2,3$. If $k_1\leq k\leq k_2$, then there exists a secret-sharing scheme $\Sigma$ for the access structure $\Gamma$ whose average information ratio $AR_\Sigma$ satisfies*

$$\frac{k_1^2+58k_1-60\log_2(\frac{k_1+2}{j+2})-32-\delta_1^{(j)}}{24k_1}\leq AR_\Sigma \leq \frac{k_2^2+60k_2-84\log_2(\frac{k_2+2}{j+3})-37-\delta_2^{(j)}}{24k_2}$$

*where* $(\delta_1^{(j)},\delta_2^{(j)})=\begin{cases}(0,0), & \text{if } j=1;\\ (28,24), & \text{if } j=2;\\ (40,44), & \text{if } j=3.\end{cases}$

**Proof.** Let $m^0_k$ and $m^{B^0}_l$ be the vertex-number sum of $\Pi^0_k$ and $\Pi^{B^0}_l$ respectively and $m^{M^0}_{l_1,l_2}$ be the order of $M^0_{l_1,l_2}$, then $m^{M^0}_{l_1,l_2}=2l_2-l_1+1$. In $\Pi^{B^0}_l$, $|V(K(C_i,A_i))|=|V(K_2)|=2$ and $|V(H^0_i)|=i+1$ for each $i$. So $m^{B^0}_l$ can be evaluated as follows.

$$\begin{aligned} m^{B^0}_l &= \begin{cases}\sum_{i=1}^{\frac{l}{2}}|V(H^0_{2i})|+|V(K(C_{2i},A_{2i})|, & \text{if } l \text{ is even};\\ \sum_{i=1}^{\frac{l-1}{2}}|V(H^0_{2i+1})|+\sum_{i=0}^{\frac{l-1}{2}}|V(K(C_{2i+1},A_{2i+1}))|, & \text{if } l \text{ is odd};\end{cases}\\ &= \begin{cases}\sum_{i=1}^{\frac{l}{2}}((2i+1)+2), & \text{if } l \text{ is even};\\ \sum_{i=1}^{\frac{l-1}{2}}(2i+2)+\sum_{i=0}^{\frac{l-1}{2}}2, & \text{if } l \text{ is odd};\end{cases}\end{aligned}$$

$$= \begin{cases} \frac{1}{4}(l^2+8l), & \text{if } l \text{ is even;} \\ \frac{1}{4}(l^2+8l-1), & \text{if } l \text{ is odd;} \end{cases}$$

(1) First, we consider $G^0_{k_1}$ whose composition process is shown by the leftmost path of length $x$ from the root. Adding up the orders of all subgraphs involved, we have

$$\begin{aligned} m^0_{k_1} &= m^0_j + \sum_{i=1}^{x} m^{B^0}_{(j+2)2^{i-1}-1} + \sum_{i=1}^{x} m^{M^0}_{(j+2)2^{i-1},(j+2)2^i-2} \\ &= \begin{cases} m^0_j + \frac{1}{4}\left[(j+1)^2+8(j+1)\right] \\ \quad + \sum\limits_{i=2}^{x} \frac{1}{4}\left[((j+2)2^{i-1}-1)^2+8((j+2)2^{i-1}-1)-1\right] \\ \quad + \sum\limits_{i=1}^{x}\left[2((j+2)2^i-2)-(j+2)2^{i-1}+1\right], & \text{if } j=1,3; \\ m^0_j + \sum\limits_{i=1}^{x} \frac{1}{4}\left[((j+2)2^{i-1}-1)^2+8((j+2)2^{i-1}-1)-1\right] \\ \quad + \sum\limits_{i=1}^{x}\left[2((j+2)2^i-2)-(j+2)2^{i-1}+1\right], & \text{if } j=2. \end{cases} \\ &= m^0_j + \frac{1}{12}((j+2)2^x)^2 + \frac{9}{2}(j+2)2^x - 5x - \varepsilon^{(j)}_1 \\ &= \frac{1}{12}(k_1+2)^2 + \frac{9}{2}(k_1+2) - 5\log_2\left(\frac{k_1+2}{j+2}\right) - \tilde{\varepsilon}^{(j)}_1 \\ &= \frac{1}{12}\left[k_1^2 + 58k_1 - 60\log_2\left(\frac{k_1+2}{j+2}\right) - 32 - \delta^{(j)}_1\right], \end{aligned}$$

where $\varepsilon^{(j)}_1 = \begin{cases} \frac{j^2+58j+109}{12}, & \text{if } j=1,3; \\ \frac{j^2+58j+112}{12}, & \text{if } j=2. \end{cases}$ and $(\tilde{\varepsilon}^{(1)}_1, \tilde{\varepsilon}^{(2)}_1, \tilde{\varepsilon}^{(3)}_1) = (12, \frac{43}{3}, \frac{46}{3})$.

In the second last step, we combine the value of $\varepsilon^{(j)}_1$ with $m^0_1 = 2$, $m^0_2 = 5$ and $m^0_3 = 9$ to calculate the value of $\tilde{\varepsilon}^{(j)}_1$. With this covering of $G^0_{k_1}$, we are able to construct a secret-sharing scheme with average information ratio $AR_{\Sigma_1} = \frac{m^0_{k_1}}{2k_1}$.

(2) We consider $G^0_{k_2}$ whose composition process is shown by the rightmost path of length $x$ from the root. Similar to (1), we have

$$m^0_{k_2} = m^0_j + \sum_{i=1}^{x} m^{B^0}_{(j+3)2^{i-1}-1} + \sum_{i=1}^{x} m^{M^0}_{(j+3)2^{i-1},(j+3)2^i-3}$$

$$= \begin{cases} m^0_j + \sum\limits_{i=1}^{x} \frac{1}{4}\left[((j+3)2^{i-1}-1)^2 + 8((j+3)2^{i-1}-1) - 1\right] \\ \quad + \sum\limits_{i=1}^{x} \left[2((j+3)2^i - 3) - (j+3)2^{i-1} + 1\right], & \text{if } j = 1, 3; \\ m^0_j + \frac{1}{4}\left[(j+2)^2 + 8(j+2)\right] \\ \quad + \sum\limits_{i=2}^{x} \frac{1}{4}\left[((j+3)2^{i-1}-1)^2 + 8((j+3)2^{i-1}-1) - 1\right] \\ \quad + \sum\limits_{i=1}^{x} \left[2((j+3)2^i - 3) - (j+3)2^{i-1} + 1\right], & \text{if } j = 2. \end{cases}$$

$$= m^0_j + \frac{1}{12}((j+3)2^x)^2 + \frac{9}{2}(j+3)2^x - 7x - \varepsilon^{(j)}_2$$

$$= \frac{1}{12}\left(k_2^2 + 60k_2 - 84\log_2\left(\frac{k_2+3}{j+3}\right) - 37 - \delta^{(j)}_2\right),$$

where $\varepsilon^{(j)}_2 = \begin{cases} \frac{j^2+60j+171}{12}, & j = 1, 3; \\ \frac{j^2+60j+168}{12}, & j = 2. \end{cases}$

With this covering of $G^0_{k_2}$, we have constructed a secret-sharing scheme with average information ratio $AR_{\Sigma_2} = \frac{m^0_{k_2}}{2k_0}$. The result then follows. ∎

As a matter of fact, the vertex-number sum $m^0_k$ of each $G^0_k$ can be evaluated in a similar way. The resulting expression only slightly differs from the ones for $m^0_{k_1}$ and $m^0_{k_2}$ at some nonleading coefficients.

After dealing with the reduced forms we shall turn back to the general forms. Let us introduce some more notations to simplify our description. Let $\vec{\mathbf{z}}_l = (1\ 1\ 2\ 1\ 2\ 1\ 2\ 1\ \cdots\ 2\ 1)$, $\vec{\mathbf{y}}_l = \left((\frac{l}{2}+1)\ \frac{l}{2}\ \frac{l}{2}\ (\frac{l}{2}-1)\ (\frac{l}{2}-1)\ \cdots\ 2\ 2\ 1\right)$ and $\vec{\mathbf{1}}_l = (1\ 1\ \cdots\ 1)$ be three $l$-dimensional vectors. For $l_1 \le l_2$, let $\vec{\mathbf{a}}(l_1, l_2) = (a_{l_1}\ a_{l_1+1}\ a_{l_1+2}\ \cdots\ a_{l_2})$ and $\vec{\mathbf{c}}(l_1, l_2) = (c_{l_1}\ c_{l_1+1}\ c_{l_1+2}\ \cdots\ c_{l_2})$ where $a_i = |A_i|$ and $c_i = |C_i|$, $i = l_1, l_1+1, \ldots, l_2$.

**Lemma 2.3.4.** *For $k = 3 \cdot 2^x - 2$ and $x \ge 1$, the vertex-number sum $m_k$ of the covering $\Pi_k$ is given as follows.*

$$
\begin{aligned}
m_k = &\sum_{i=1}^{x-1}\left(\vec{\mathbf{z}}_{\frac{k+2}{2^i}}+(i-1)\vec{\mathbf{1}}_{\frac{k+2}{2^i}}\right)\cdot\vec{\mathbf{a}}\left(\frac{(k+2)(2^{i-1}-1)}{2^{i-1}}+1,\frac{(k+2)(2^i-1)}{2^i}\right)\\
&+xa_{k-3}+(x+1)a_{k-2}+xa_{k-1}+(x+1)a_k\\
&+\sum_{i=1}^{x-1}\left(\vec{\mathbf{y}}_{\frac{k+2}{2^i}}+(i-1)\vec{\mathbf{1}}_{\frac{k+2}{2^i}}\right)\cdot\vec{\mathbf{c}}\left(\frac{(k+2)(2^{i-1}-1)}{2^{i-1}}+1,\frac{(k+2)(2^i-1)}{2^i}\right)\\
&+(x+1)c_{k-3}+(x+1)c_{k-2}+xc_{k-1}+(x+1)c_k.
\end{aligned}
$$

**Proof.** Note that the expression for $m_k$ depends on all $a_i$'s and $c_i$'s, each of whose coefficients represents the occurrence of the vertices of that part in the covering $\Pi_k$.

(1) First, let us examine the occurrence of vertices of $B_l$, whose partite sets are $\bigcup_{i=1}^{l} A_i$ and $\bigcup_{i=1}^{l} C_i$, in its covering $\Pi_l^B$. For odd $l$, by Lemma 2.3.1, one can easily see that the vertices in $A_1$ have occurrence 1 (only in $K(A_1, C_1)$), the vertices in $A_{2j}$, $j = 1, \ldots, \frac{l-1}{2}$, also have occurrence 1 (only in $H_{2j+1}$) and the vertices in $A_{2j+1}$, $j = 1, \ldots, \frac{l-1}{2}$, have occurrence 2 (in $H_{2j+1}$ and $K(A_{2j+1}, C_{2j+1})$). Hence, the occurrences of the vertices in $A_1, A_2, \ldots, A_l$ are exactly the first $l$ coordinates in $\vec{\mathbf{z}}_{l+1}$. Similarly, the vertices in $C_1$ have occurrence $\frac{l+1}{2}$ (in $K(A_1, C_1)$ and $H_{2i+1}$'s, $i = 1, \ldots, \frac{l-1}{2}$), the vertices in $C_{2j}$, $j = 1, \ldots, \frac{l-1}{2}$, have occurrence $\frac{l-1}{2} - j + 1$ (in $H_{2i+1}$'s, $i \geq j$) and the vertices in $C_{2j+1}$, $j = 1, \ldots, \frac{l-1}{2}$, have occurrence $\frac{l-1}{2} - j + 1$ (in $H_{2i+1}$'s, $i \geq j+1$ and $K(A_{2j+1}, C_{2j+1})$). Hence, the occurrences of the vertices in $C_1, C_2, \ldots, C_l$ are exactly the first $l$ coordinates in $\vec{\mathbf{y}}_{l+1} - \vec{\mathbf{1}}_{l+1}$.

(2) Let us consider the value of $m_k$ now. We prove the result by induction on $x$. When $x = 1$, $m_4 = a_1 + 2a_2 + a_3 + 2a_4 + 2c_1 + 2c_2 + c_3 + 2c_4$ by direct counting the occurrences of vertices in $\Pi_4$. So, the result holds when $x = 1$. Next, for $k = 3 \cdot 2^{x+1} - 2$, $G_k = W(a_1, \ldots, a_k, c_1, \ldots, c_k)$ is composed of $B_{3\cdot 2^x - 1}$, $M_{3\cdot 2^x, 3\cdot 2^{x+1}-2}$ and $G_{3\cdot 2^x - 2}$. For convenience, denote $M_{3\cdot 2^x, 3\cdot 2^{x+1}-2}$ by $M$ for now. Observe that the vertices in $A_i$, $1 \leq i \leq 3 \cdot 2^x - 1$, have the same occurrences in $\Pi_k$ as they do in the covering $\Pi^B_{3\cdot 2^x - 1}$ because they do not lie in $M$ and $G_{3\cdot 2^x - 2}$, while the vertices in $C_i$, $1 \leq i \leq 3 \cdot 2^x - 1$, gain

one more occurrences in $\Pi_k$ than they do in $\Pi^B_{3\cdot 2^x-1}$ because they also occur in $M$. Notice that the vertices in $A_{3\cdot 2^x}$ and $C_{3\cdot 2^x}$ only occur once in $\Pi_k$. Besides, the vertices in $A_i$'s and $C_i$'s, $i = 3\cdot 2^x+1,\ldots,k$, also gain one more occurrence in $\Pi_k$ than they do in the covering $\Pi_{3\cdot 2^x-2}$ of $G_{3\cdot 2^x-2}$. Therefore, by (1) and the induction hypothesis, we have

$$
\begin{aligned}
&m_{3\cdot 2^{x+1}-2}\\
&= \vec{\mathbf{z}}_{3\cdot 2^x}\cdot\vec{\mathbf{a}}(1,3\cdot 2^x) + (\vec{\mathbf{y}}_{3\cdot 2^x}-\vec{\mathbf{1}}_{3\cdot 2^x})\cdot\vec{\mathbf{c}}(1,3\cdot 2^x) + \vec{\mathbf{1}}_{3\cdot 2^x}\cdot\vec{\mathbf{c}}(1,3\cdot 2^x)\\
&+\sum_{i=1}^{x-1}\left(\vec{\mathbf{z}}_{\frac{3\cdot 2^x}{2^i}}+(i-1)\vec{\mathbf{1}}_{\frac{3\cdot 2^x}{2^i}}+\vec{\mathbf{1}}_{\frac{3\cdot 2^x}{2^i}}\right)\cdot\vec{\mathbf{a}}\left(\frac{3\cdot 2^x(2^{i-1}-1)}{2^{i-1}}+1+3\cdot 2^x,\frac{3\cdot 2^x(2^i-1)}{2^i}+3\cdot 2^x\right)\\
&+(x+1)a_{3\cdot 2^x-5+3\cdot 2^x}+(x+2)a_{3\cdot 2^x-4+3\cdot 2^x}+(x+1)a_{3\cdot 2^x-3+3\cdot 2^x}+(x+2)a_{3\cdot 2^x-2+3\cdot 2^x}\\
&+\sum_{i=1}^{x-1}\left(\vec{\mathbf{y}}_{\frac{3\cdot 2^x}{2^i}}+(i-1)\vec{\mathbf{1}}_{\frac{3\cdot 2^x}{2^i}}+\vec{\mathbf{1}}_{\frac{3\cdot 2^x}{2^i}}\right)\cdot\vec{\mathbf{c}}\left(\frac{3\cdot 2^x(2^{i-1}-1)}{2^{i-1}}+1+3\cdot 2^x,\frac{3\cdot 2^x(2^i-1)}{2^i}+3\cdot 2^x\right)\\
&+(x+2)c_{3\cdot 2^x-5+3\cdot 2^x}+(x+2)c_{3\cdot 2^x-4+3\cdot 2^x}+(x+1)c_{3\cdot 2^x-3+3\cdot 2^x}+(x+2)c_{3\cdot 2^x-2+3\cdot 2^x}\\
&=\vec{\mathbf{z}}_{\frac{3\cdot 2^{x+1}}{2}}\cdot\vec{\mathbf{a}}\left(1,\frac{3\cdot 2^{x+1}}{2}\right)+\vec{\mathbf{y}}_{\frac{3\cdot 2^{x+1}}{2}}\cdot\vec{\mathbf{c}}\left(1,\frac{3\cdot 2^{x+1}}{2}\right)\\
&+\sum_{i=1}^{x-1}\left(\vec{\mathbf{z}}_{\frac{3\cdot 2^{x+1}}{2^{i+1}}}+((i+1)-1)\vec{\mathbf{1}}_{\frac{3\cdot 2^{x+1}}{2^{i+1}}}\right)\cdot\vec{\mathbf{a}}\left(\frac{3\cdot 2^{x+1}(2^i-1)}{2^i}+1,\frac{3\cdot 2^{x+1}(2^{i+1}-1)}{2^{i+1}}\right)\\
&+(x+1)a_{(3\cdot 2^{x+1}-2)-3}+(x+2)a_{(3\cdot 2^{x+1}-2)-2}+(x+1)a_{(3\cdot 2^{x+1}-2)1}+(x+2)a_{(3\cdot 2^{x+1}-2)}\\
&+\sum_{i=1}^{x-1}\left(\vec{\mathbf{y}}_{\frac{3\cdot 2^{x+1}}{2^{i+1}}}+((i+1)-1)\vec{\mathbf{1}}_{\frac{3\cdot 2^{x+1}}{2^{i+1}}}\right)\cdot\vec{\mathbf{c}}\left(\frac{3\cdot 2^{x+1}(2^i-1)}{2^i}+1,\frac{3\cdot 2^{x+1}(2^{i+1}-1)}{2^{i+1}}\right)\\
&+(x+2)c_{(3\cdot 2^{x+1}-2)-3}+(x+2)c_{(3\cdot 2^{x+1}-2)-2}+(x+1)c_{(3\cdot 2^{x+1}-2)1}+(x+2)c_{(3\cdot 2^{x+1}-2)}\\
&=\sum_{i=1}^{x}\left(\vec{\mathbf{z}}_{\frac{k+2}{2^i}}+(i-1)\vec{\mathbf{1}}_{\frac{k+2}{2^i}}\right)\cdot\vec{\mathbf{a}}\left(\frac{(k+2)(2^{i-1}-1)}{2^{i-1}}+1,\frac{(k+2)(2^i-1)}{2^i}\right)\\
&+(x+1)a_{k-3}+(x+2)a_{k-2}+(x+1)a_{k-1}+(x+2)a_k\\
&+\sum_{i=1}^{x}\left(\vec{\mathbf{y}}_{\frac{k+2}{2^i}}+(i-1)\vec{\mathbf{1}}_{\frac{k+2}{2^i}}\right)\cdot\vec{\mathbf{c}}\left(\frac{(k+2)(2^{i-1}-1)}{2^{i-1}}+1,\frac{(k+2)(2^i-1)}{2^i}\right)\\
&+(x+2)c_{k-3}+(x+2)c_{k-2}+(x+1)c_{k-1}+(x+2)c_k.
\end{aligned}
$$

∎

This lemma presents a sophisticated expression for $m_k$ in terms of $a_i$'s and $c_i$'s. In what follows, we give the conditions on the values of $a_i$'s and $c_i$'s under which $m_k$ attains its minimum value when $n = \sum_{i=1}^{k}(a_i + c_i)$ is fixed. Thereby, the lowest possible average information ratio of the secret-sharing scheme constructed via this covering is obtained.

**Theorem 2.3.5.** *Let $\Gamma$ be a weighted threshold access structure represented by a $k$-weighted graph $G = W(a_1, \ldots, a_k, c_1, \ldots, c_k)$ of order $n$ and $k = 3\cdot 2^x - 2$. If $c_i = 1$ for all $i \neq \frac{k}{2} + 1$ and $a_i = 1$ for all $i \notin T = \{1, 2, 4, 6, \ldots, \frac{k}{2} + 1\}$. Then*

$$AR(G) \leq \frac{12n + k^2 + 34k - 60\log_2(\frac{k+2}{3}) - 32}{12n}.$$

**Proof.** Observe that only $c_{\frac{k}{2}+1}$ and $a_i$, $i \in T$, have coefficient equal to one in the expression for $m_k$ in Lemma 2.3.4. So $m_k$ is minimized if $c_i = 1$ for all $i \neq \frac{k}{2}+1$ and $a_i = 1$ for all $i \notin T$ since this expression for $m_k$ is linear. This case is similar to the reduced form. So, we make an adjustment in the expression for $m_{k_1}^0$ (with $j = 1$) in the proof of Theorem 2.3.3 to derive what we need here. The vertex-number sum $m_k$ of this covering is $m_{k_1}^0 + \sum_{i\in T} a_i + c_{\frac{k}{2}+1} - (|T|+1)$. Note that $n = \sum_{i=1}^{k}(a_i + c_i) = \sum_{i\in T} a_i + c_{\frac{k}{2}+1} + \sum_{i\notin T} a_i + \sum_{i\neq\frac{k}{2}+1} c_i = \sum_{i\in T} a_i + c_{\frac{k}{2}+1} + (k - |T|) + (k - 1) = \sum_{i\in T} a_i + c_{\frac{k}{2}+1} + 2k - (|T| + 1)$. Therefore, in this case $m_k = \frac{1}{12}[k^2 + 58k - 60\log_2(\frac{k+2}{3}) - 32] + n - 2k = \frac{1}{12}[12n + k^2 + 34k - 60\log_2(\frac{k+2}{3}) - 32]$. The average information ratio of the secret-sharing scheme constructed with this covering attains its minimum value $\frac{m_k}{n}$ and the proof is completed. ∎

Our result appears to be quite good if $k$ is relatively small compared with $n$. In fact, as $k$ fixed, the ratio given in Theorem 2.3.5 asymptotically approaches "1" which is the optimal value for this ratio.

## 2.4 Construction (II)

Our second construction is similar to the first, while it performs better than Construction I when $k \geq 31$. The major difference is that $B_l$ is replaced with $G_l$ in the covering. With the notations used before, we define our second covering $\widetilde{\Pi}_k$ of $G_k = W(a_1, \ldots, a_k, c_1, \ldots, c_k)$ recursively as follows. $\widetilde{\Pi}_i = \Pi_i$, $i = 1, 2, 3$. For $k \geq 4$, $\widetilde{\Pi}_k = \widetilde{\Pi}_{\lfloor \frac{k-1}{2} \rfloor} \cup \left\{ M_{\lfloor \frac{k-1}{2} \rfloor + 1, k} \right\} \cup \widetilde{\Pi}_{\lfloor \frac{k}{2} \rfloor}$ where the $\widetilde{\Pi}_{\lfloor \frac{k}{2} \rfloor}$ is the complete multipartite covering of the $\lfloor \frac{k}{2} \rfloor$-weighted subgraph $\underset{\sim}{W} = W\left( a_{\lfloor \frac{k-1}{2} \rfloor + 2}, a_{\lfloor \frac{k-1}{2} \rfloor + 3}, \ldots, a_k, c_{\lfloor \frac{k-1}{2} \rfloor + 2},\ c_{\lfloor \frac{k-1}{2} \rfloor + 3}, \ldots, c_k \right)$. It is obvious that the edges not in the subgraphs $W\left( a_1, \ldots, a_{\lfloor \frac{k-1}{2} \rfloor},\ c_1, \ldots, c_{\lfloor \frac{k-1}{2} \rfloor} \right)$ and $\underset{\sim}{W}$ all lie in $M_{\lfloor \frac{k-1}{2} \rfloor + 1, k}$. So, $\widetilde{\Pi}_k$ is a complete multipartite covering of $G_k$.

**Lemma 2.4.1.** *The collection $\widetilde{\Pi}_k$ is a complete multipartite covering of $G_k$ with minimum edge occurrence one.*

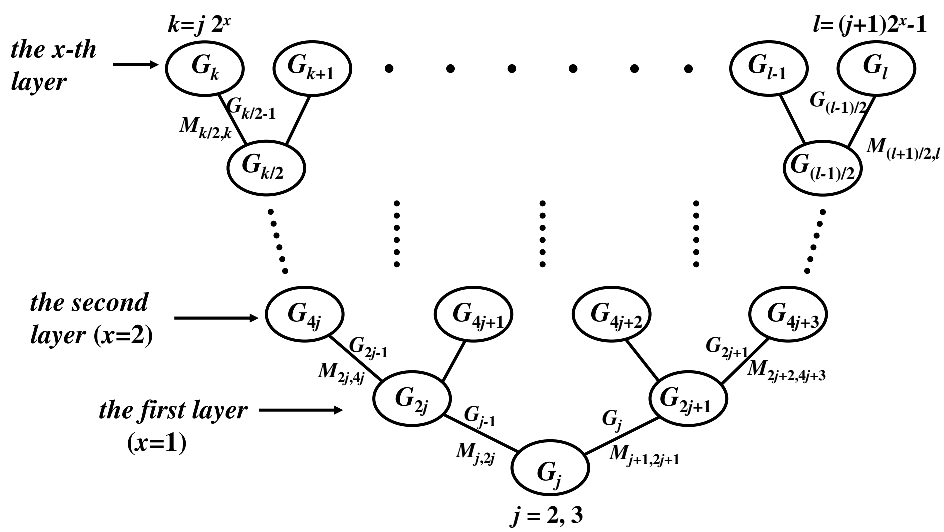



Figure 2.2: The binary tree for Construction (II)

In order to evaluate the vertex-number sum $\widetilde{m}_k$ of $\widetilde{\Pi}_k$, we consider the reduced form first. Let $\widetilde{\Pi}_k^0$ and $\widetilde{m}_k^0$ be the reduced version of $\widetilde{\Pi}_k$ and $\widetilde{m}_k$ respectively. In the covering $\widetilde{\Pi}_k^0$, we decompose $G_k^0$ into $G^0_{\lfloor \frac{k-1}{2} \rfloor}$, $M^0_{\lfloor \frac{k-1}{2} \rfloor + 1, k}$ and $G^0_{\lfloor \frac{k}{2} \rfloor}$. Since $\lfloor \frac{k-1}{2} \rfloor$ equals $\lfloor \frac{k}{2} \rfloor - 1$ or $\lfloor \frac{k}{2} \rfloor$, $G_j^0$ can either go with $G_{j-1}^0$

and $M^0_{j,2j}$ to compose $G^0_{2j}$ or go with $G^0_j$ and $M^0_{j+1,2j+1}$ to compose $G^0_{2j+1}$. Recursively, all $G^0_k$'s can be obtained by using this process repeatly from $G_1$, $G_2$, $G_3$ and some $M^0_{i,k}$'s. As we have done in Section 2.3, this relation is depicted by a binary tree in Figure 2.2. The $2^x$ paths of length $x$ from the root give the conformations of the $2^x$ $k$-weight graphs where $2^{x+1} \le k \le 3 \cdot 2^x - 1$ or $3 \cdot 2^x \le k \le 2^{x+2} - 1$.

**Theorem 2.4.2.** *Let $\Gamma$ be an weighted threshold access structure represented by a $k$-weighted graph $G^0_k$ of reduced form, $k_1 = j \cdot 2^x$ and $k_2 = (j+1) \cdot 2^x - 1$, $x \ge 0$, $j = 2, 3$. If $k_1 \le k \le k_2$, then there exists a secret-sharing scheme $\Sigma$ for the access structure $\Gamma$ whose average information ratio $AR_\Sigma$ satifies*

$$\begin{aligned}\frac{(\frac{3}{2}k_1 + 2)\log_2 k_1 + \delta_1^{(j)} k_1 + \delta_0^{(j)}}{2k_1} &\le AR_\Sigma \\ &\le \frac{\frac{3}{2}(k_2+1)\log_2(k_2+1) + \delta^{(j)}(k_2+1) + 1}{2k_2}\end{aligned}$$

*where* $(\delta^{(j)}, \delta_1^{(j)}, \delta_0^{(j)}) = \begin{cases} (\frac{4}{3} - \frac{3}{2}\log_2 3, -1, 2), & \text{if } j = 2; \\ (-1, \frac{4}{3} - \frac{3}{2}\log_2 3, 5 - 2\log_2 3), & \text{if } j = 3. \end{cases}$

**Proof.** Recall that $M^0_{l_1,l_2}$ has order $m^{M^0}_{l_1,l_2} = 2l_2 - l_1 + 1$, $\widetilde{m}^0_i = m^0_i$, $i = 1, 2, 3$. $m^0_1 = 2$, $m^0_2 = 5$, and $m^0_3 = 9$.

(1) First, we consider $G^0_{k_2}$. For each $l = 2^i(j+1) - 1$, $G_l$ is composed of two $G_{\frac{l-1}{2}}$'s and one $M_{\frac{l+1}{2},l}$. So $\widetilde{m}^0_k$ can be evaluated recursively as follows.

$$\begin{aligned}\widetilde{m}^0_{k_2} &= 2\widetilde{m}^0_{2^{x-1}(j+1)-1} + 3 \cdot 2^{x-1}(j+1) - 1 \\ &= 2^x m^0_j + \sum_{i=1}^{x}(2^{i-1}(3 \cdot 2^{x-i}(j+1) - 1)) \\ &= 2^x \cdot m^0_j + 3x \cdot 2^{x-1}(j+1) - (2^x - 1) \\ &= 3 \cdot \frac{k_2+1}{2}\log_2\left(\frac{k_2+1}{j+1}\right) + \frac{m^0_j - 1}{j+1} \cdot (k_2+1) + 1 \\ &= \frac{3}{2}(k_2+1)\log_2(k_2+1) + \left(\frac{m^0_j - 1}{j+1} - \frac{3}{2}\log_2(j+1)\right)(k_2+1) + 1 \\ &= \frac{3}{2}(k_2+1)\log_2(k_2+1) + \delta^{(j)}(k_2+1) + 1.\end{aligned}$$

Hence, the secret-sharing scheme constructed with $\widetilde{\Pi}^0_{k_2}$ has average information ratio $AR_{\Sigma_2} = \frac{\widetilde{m}^0_{k_2}}{2k_2}$.

(2) The composition process of $G^0_{k_1}$ is shown on the leftmost path of length $x$ from the root. Adding up the orders of all subgraphs involved, we have $\widetilde{m}^0_{k_1} = \widetilde{m}^0_j + \widetilde{m}^0_{j-1} + \sum_{i=1}^{x-1} \widetilde{m}^0_{2^i \cdot j-1} + \sum_{i=1}^{x} m^{M^0}_{2^{i-1}j, 2^i j}$. Making use of the equation $\widetilde{m}^0_{2^x(j+1)-1} = 2^x \cdot m^0_j + 3x \cdot 2^{x-1}(j+1) - (2^x - 1)$ from the derivation in (1), we can continue to evaluate $\widetilde{m}^0_{k_1}$ according to the value of $j$ as follows.

(i) If $j = 3$,

$$
\begin{aligned}
&\widetilde{m}^0_{3\cdot 2^x} \\
&= m^0_j + m^0_{j-1} + \sum_{i=1}^{x-1}[2^i \cdot m^0_{j-1} + 3 \cdot i \cdot 2^{i-1} \cdot j - (2^i - 1)] + \sum_{i=1}^{x}(3 \cdot 2^{i-1} \cdot j + 1) \\
&= m^0_3 + m^0_2 + m^0_2(2^x - 2) + 9((x-2)2^{x-1} + 1) - (2^x - 1 - x) + 9(2^x - 1) + x \\
&= 9x2^{x-1} + 4 \cdot 2^x + 2x + 5 \\
&= \frac{3k}{2}\log_2 k_1 + \left(\frac{4}{3} - \frac{3}{2}\log_2 3\right) k_1 + 2\log_2 k_1 + (5 - 2\log_2 3).
\end{aligned}
$$

(ii) If $j = 2$,

$$
\begin{aligned}
&\widetilde{m}^0_{2^{x+1}} \\
&= m^0_j + m^0_{j-1} + \sum_{i=1}^{x-1}[2^{i-1}m^0_3 + 3(i-1)2^{i-2} \cdot 4 - (2^{i-1} - 1)] + \sum_{i=1}^{x}(3 \cdot 2^{i-1} \cdot j + 1) \\
&= 3x \cdot 2^x + 2^x + 2x + 4 \\
&= \frac{3}{2}k_1 \log_2 k_1 - k_1 + 2\log_2 k_1 + 2.
\end{aligned}
$$

Hence $\widetilde{m}^0_{k_1} = (\frac{3}{2}k_1 + 2)\log_2 k_1 + \delta^{(j)}_1 k_1 + \delta^{(j)}_0$ and we have a secret-sharing scheme with average information ratio $AR_{\Sigma_1} = \frac{\widetilde{m}^0_{k_1}}{2k_1}$. The result follows immediately. ∎

Next, we give the expression for $\widetilde{m}_k$ for a $k$-weighted graph of general form.

**Lemma 2.4.3.** *Let $k = 2^x \cdot (j+1) - 1$, $x \geq 0$, $j = 2, 3$. If $\widetilde{m}_k = \sum_{i=1}^{k} \alpha_{j,i}^x a_i + \sum_{i=1}^{k} \beta_{j,i}^x c_i$ is the vertex-number sum of the covering $\widetilde{\Pi}_k$ of the k-weighted graph $G_k = W(a_1, \ldots, a_k, c_1, \ldots, c_k)$. Then the values of $\alpha_{j,i}^x$'s and $\beta_{j,i}^x$'s can be obtained by the recursive relations $\alpha_{j,i}^x = \alpha_{j,\frac{k+1}{2}+i}^x - 1 = \alpha_{j,i}^{x-1}$, $\beta_{j,i}^x = \beta_{j,\frac{k+1}{2}+i}^x = \beta_{j,i}^{x-1} + 1$ and $\alpha_{j,\frac{k+1}{2}}^x = \beta_{j,\frac{k+1}{2}}^x = 1$, $1 \leq i \leq \frac{k-1}{2}$, with initial values $\alpha_{j,1}^0 = \alpha_{j,2}^0 = \beta_{j,2}^0 = 1$ and $\beta_{j,1}^0 = \alpha_{3,3}^0 = \beta_{3,3}^0 = 2$.*

**Proof.** We prove this result by induction on $x$. When $x = 0$, $k = j$, the occurrences of the vertices in $A_i$'s and $C_i$'s in $\widetilde{\Pi}_j$ are exactly the initial values $\alpha_{j,i}^0$'s and $\beta_{j,i}^0$'s respectively. For $x > 0$, recall that $G_k$ is composed of $W_1 = W(a_1, \ldots, a_{2^{x-1}(j+1)-1}, c_1, \ldots, c_{2^{x-1}(j+1)-1})$, $W_2 = W(a_{2^{x-1}(j+1)+1}, \ldots, a_k, c_{2^{x-1}(j+1)+1}, \ldots, c_k)$ and $M = M_{2^{x-1}(j+1), 2^x(j+1)-1}$. Each vertex in $A_i$, $1 \leq i \leq \frac{k-1}{2} = 2^{x-1}(j+1) - 1$, has the same occurrence in $\widetilde{\Pi}_k$ as it does in the covering of $W_1$ since it does not occur in either $W_2$ or $M$. So, $\alpha_{j,i}^x = \alpha_{j,i}^{x-1}$. However, each vertex in $C_i$, $1 \leq i \leq \frac{k-1}{2}$, gains one more occurrence in $\widetilde{\Pi}_k$ than it does in the covering of $W_1$ because it also occurs in $M$. This is also true for vertices in $A_i$ and $C_i$, $\frac{k+1}{2} = 2^{x-1}(j+1) + 1 \leq i \leq k$, because all of them occur in graph $M$ as well. Hence, we also have $\beta_{j,i}^x = \beta_{j,i}^{x-1} + 1$, $\alpha_{j,\frac{k+1}{2}+i}^x = \alpha_{j,i}^{x-1} + 1$ and $\beta_{j,\frac{k+1}{2}+i}^x = \beta_{j,i}^{x-1} + 1$ for $1 \leq i \leq \frac{k-1}{2}$. Besides, the vertices in $A_{\frac{k+1}{2}}$ and $C_{\frac{k+1}{2}}$ have occurrence one because they only appear in $M$. Hence, $\alpha_{j,\frac{k+1}{2}}^x = \beta_{j,\frac{k+1}{2}}^x = 1$. This proves that the coefficients $\alpha_{j,i}^x$'s and $\beta_{j,i}^x$'s satisfy the given recursive relations. ∎

Now, we consider the case when $n = \sum_{i=1}^{k}(a_i + c_i)$ is fixed. By evaluating the minimum value of $\widetilde{m}_k$, we obtain the lowest possible average information ratio of a secret-sharing scheme constructed with this covering.

**Theorem 2.4.4.** *Let $\Gamma$ be a weighted threshold access structure represented by a k-weighted graph $G = W(a_1, \ldots, a_k, c_1, \ldots, c_k)$ of order $n$ and $k = (j+1)2^x - 1$. If $c_i = 1$ for all $i \neq \frac{k+1}{2}$ and $a_i = 1$ for all $i \notin T = \{1, 2\} \cup \{(j+1)2^i | i = 0, 1, \ldots, x-1\}$. Then*

$$AR(G) \leq \frac{n + \frac{3}{2}(k+1)\log_2(k+1) + (\delta^{(j)} - 2)k + (\delta^{(j)} + 1)}{n}$$

*where $\delta^{(j)}$ is given in Theorem 2.4.2.*

**Proof.** The argument is similar to the proof of Theorem 2.3.5. From the relations given in Lemma 2.4.3, among all the coefficients of $a_i$'s and $c_i$'s, only $\alpha^x_{j,i}$, $i \in T$, and $\beta^x_{j,\frac{k+1}{2}}$ are equal to one. So $\widetilde{m}_k$ is minimized if $a_i = 1$ for all $i \notin T$ and $c_i = 1$ for all $i \neq \frac{k+1}{2}$. We modify the expression for $\widetilde{m}^0_{k_2}$ in the proof of Theorem 2.4.2 to meet what we need here. In this case, $\widetilde{m}_k = \widetilde{m}^0_{k_2} + \sum_{i \in T} a_i + c_{\frac{k+1}{2}} - (|T|+1) = \widetilde{m}^0_k + n - 2k = n + \frac{3}{2}(k+1)\log_2(k+1) + (\delta^{(j)}-2)k + (\delta^{(j)}+1)$. The secret-sharing scheme for this access structure has average information ratio $\frac{\widetilde{m}_k}{n}$. ∎

This result is also very good when $k$ is relatively small compared with $n$. The ratio also approaches "1" asymptotically as $k$ fixed. After analyzing the average information ratio produced from each of our constructions separately, we shall give a comparison of them in Section 2.5. For a fair comparison, we consider the same class of $k$-weighted graphs where $k = 3 \cdot 2^x - 2$. We present the lowest possible average information rate for this class as follows.

**Theorem 2.4.5.** *Let $\Gamma$ be a weighted threshold access structure represented by a $k$-weighted graph $G_k = W(a_1, \ldots, a_k, c_1, \ldots, c_k)$ of order $n$ and $k = 3 \cdot 2^x - 2$. If $c_i = 1$ for all $i \neq \frac{k}{2}$ and $a_i = 1$ for all $i \notin T = \{1\} \cup \{3 \cdot 2^i - 1 | i = 0, 1, \ldots, x-1\}$. Then*

$$AR(G_k) \leq \frac{n + (\frac{3}{2}k+2)\log_2(k+2) - (\frac{2}{3} + \frac{3}{2}\log_2 3)k + \frac{2}{3} - 2\log_2 3}{n}.$$

**Proof.** Suppose $(\bigcup_{i=1}^k A_i) \cup (\bigcup_{i=1}^k C_i)$ is the vertex set of $G_k$ where $|A_i| = a_i$ and $|C_i| = c_i$, $i = 1, 2, \ldots, k$. Denote $\{u\}$ by $A_0$ and $\{v\}$ by $C_0$. Let $(\bigcup_{i=0}^k A_i) \cup (\bigcup_{i=0}^k C_i)$ be the vertex set of the $(k+1)$-weighted graph $G_{k+1} = W(|A_0|, a_1, \ldots, a_k, |C_0|, c_1, \ldots, c_k)$ of order $n+2$ where $k+1 = 3 \cdot 2^x - 1$. Then $G_{k+1}$ satisfies the criteria in Theorem 2.4.4, and the vertex-number sum $\widetilde{m}_{k+1}$ of its covering $\widetilde{\Pi}_{k+1}$ is $n + 2 + \frac{3}{2}(k+2)\log_2(k+2) + (\delta^{(2)}-2)(k+1) + \delta^{(2)} + 1$. Now, observe that $G_k = G_{k+1} - (A_0 \cup C_0)$ and the collection of subgraphs obtained from $\widetilde{\Pi}_{k+1}$ by deleting $u$ and $v$ from each subgraphs in $\widetilde{\Pi}_{k+1}$ is exactly the complete multipartite covering $\widetilde{\Pi}_k$ of $G_k$ since $G_{k+1}$ is composed of $W(|A_0|, a_1, \ldots, a_{\frac{k}{2}-1}, |C_0|, c_1, \ldots, c_{\frac{k}{2}-1})$, $M_{\frac{k}{2}+1,k+1}$ (in $G_{k+1}$) and $W(a_{\frac{k}{2}+1}, \ldots, a_k, c_{\frac{k}{2}+1}, \ldots, c_k)$ and $G_k$ is composed of $W(a_1, \ldots, a_{\frac{k}{2}-1}, c_1, \ldots, c_{\frac{k}{2}-1})$, $M_{\frac{k}{2},k}$ (in $G_k$) and $W(a_{\frac{k}{2}+1}, \ldots, a_k, c_{\frac{k}{2}+1}, \ldots, c_k)$. From the relations in Lemma 2.4.3, one can see that the occurrence of $u$ in $\widetilde{\Pi}_{k+1}$ is one and the occurrence of $v$ in $\widetilde{\Pi}_{k+1}$ is $\beta^x_{2,1} = x + 2 = \log_2(\frac{k+2}{3}) + 2$. Hence, the vertex-number sum $\widetilde{m}_k$ of $\widetilde{\Pi}_k$ is $\widetilde{m}_{k+1} - 1 - (\log_2(\frac{k+2}{3}) + 2) = n + (\frac{3}{2}k + 2)\log_2(k + 2) - (\frac{2}{3} + \frac{3}{2}\log_2 3)k + \frac{2}{3} - 2\log_2 3$. The result is then obtained. ∎

## 2.5 Concluding Remark

The weighted threshold access structure is a more applicable structure of secret-sharing schemes in reality. In the implementation of such a scheme, the value of $k$ can be thought of as the number of departments or divisions in an organization. In order to have a comparison of the efficiency of our constructions of secret-sharing scchmes, we let $AR_1 = \frac{12n+k^2+34k-60\log_2(\frac{k+2}{3})-32}{12n}$ and $AR_2 = \frac{n+(\frac{3}{2}k+2)\log_2(k+2)-(\frac{2}{3}+\frac{3}{2}\log_2 3)k+\frac{2}{3}-2\log_2 3}{n}$ which are the lowest possible average information ratio derived from our two constructions in Theorem 2.3.5 and Theorem 2.4.5, respectively. Both ratios perform very well when $n/k$ is large. If $k$ is constant, both of them approaches "1" asymptotically. Let $n = \mu k$ where $\mu$ can be thought of as the average size of departments in the organization. When $\mu$ is larger, both $AR_1$ and $AR_2$ become lower for each fixed value of $k$. Figures 2.3 and 2.4 show the behavior of Morillo's ratio [28], $AR_1$ and $AR_2$ in the case when $\mu = 20$. As indicated in the figure, $AR_1$ performs better than $AR_2$ when $k \leq 30$, whereas $AR_2$ becomes superior to $AR_1$ for all $k \geq 31$. Actually, this fact remains true for all values of $\mu$. Therefore, Construction I is more suitable for organizations with fewer departments, whereas Construction II performs especially well for organizations with more departments.

The results in this chpater have been included in the following paper.

"H.-C. Lu and H.-L. Fu, New bounds on the average information rate of secret-sharing schemes for graph-based weighted threshold access structures, *Information Sciences*, **240** (2013), 83-94."
(http://dx.doi.org/10.1016/j.ins.2013.03.047)

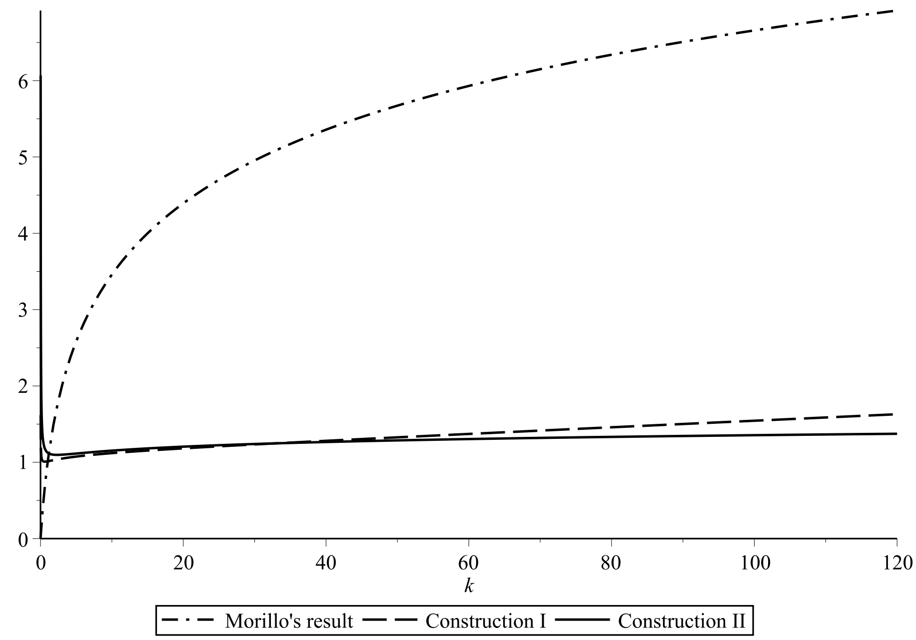



Figure 2.3: A comparison of the results in the case when $\mu = 20$.

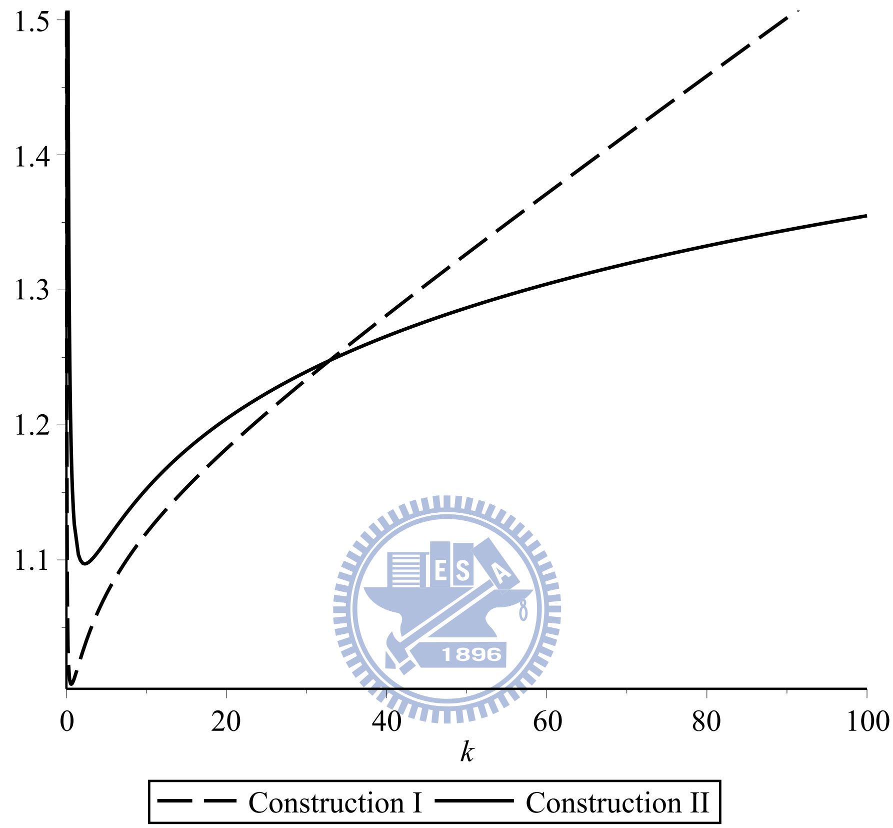



Figure 2.4: A comparison of $AR_1$ and $AR_2$ in the case when $\mu = 20$.

# Chapter 3

# Optimal Average Information Ratio for Trees

Before taking care of trees, we start this chapter with the introduction of our approach to the determination of the exact values of the optimal average information ratio of graphs of larger girth.

## 3.1 Our Approach to the Determination of the Exact Values of $AR(G)$

Let $IN(G) = \{v \in V(G) | \deg_G(v) \geq 2\}$ and $in(G) = |IN(G)|$. Given a star covering $\Pi$ of $G$ with vertex-number sum $m_\Pi$, the *deduction* of $\Pi$ is defined as $d_\Pi = |V(G)| + in(G) - m_\Pi$. A star covering with the least vertex-number sum gives the largest deduction. We also denote the largest deduction over all star coverings of $G$ as $d^*(G)$, called the deduction of $G$. A star covering $\Pi$ with $d_\Pi = d^*(G)$ is referred to as an *optimal star covering* of $G$. The following upper bound on $AR(G)$ is simply a rephrasemant of Theorem 1.3.1 in terms of the deduction of $G$.

**Corollary 3.1.1** ([34]). *If* $\Pi$ *is a star covering of a graph* $G$ *with deduction* $d_\Pi$, *then* $AR(G) \leq \frac{|V(G)|+in(G)-d_\Pi}{|V(G)|}$.

For the derivation of lower bounds on $AR(G)$, we follow Csirmaz's approach stated in Section 1.3.2. Recall that a core of $G$ is a connected subset $V_0 \subseteq V(G)$ such that each vertex $v \in V_0$ has a *designated outside neighbor* $\bar{v}$, which refers to a neighbor of $v$ that is outside $V_0$ and is not adjacent to any other vertex in $V_0$, and $\{\bar{v}|v \in V_0\}$ is an independent set. In the case of trees, all neighbors of the vertices in a connected set naturally form an independnet set. Therefore a core of a tree can be simplified as a connected subset $V_0 \subseteq V(G)$ such that each vertex $v \in V_0$ has a designated outside neighbor. In order to cope with the average information ratio, we extend the idea of a core of $G$. For $G \neq K_{1,1}$, we define a *core cluster of $G$ of size $k$* as a partition $\mathcal{C} = \{V_1, V_2, \ldots, V_k\}$ of $IN(G)$ such that each $V_i$, $i \in \{1, 2, \ldots, k\}$, is a core of $G$. The size of a core cluster $\mathcal{C}$ is written as $c_{\mathcal{C}}$. We also denote the minimum size of all core clusters of $G$ as $c^*(G)$, called the *core number* of $G$. Note that $\bigcup_{i=1}^{k} V_i$ may not be a core of $G$, if so, then $c^*(G) = 1$ for $G \neq K_{1,1}$. The core number of $K_{1,1}$ is naturally defined as $c^*(K_{1,1}) = 0$. A core cluster of size $c^*(G)$ is then called an *optimal core cluster* of $G$. The idea of a core cluster helps us establish a lower bound on $AR(G)$.

**Theorem 3.1.2.** *If $\mathcal{C}$ is a core cluster of a graph $G$, then*

$$AR(G) \geq \frac{|V(G)| + in(G) - c_{\mathcal{C}}}{|V(G)|}.$$

**Proof.** Let $\mathcal{C} = \{V_1, V_2, \ldots, V_k\}$ and $\Sigma$ be a secret-sharing scheme on $G$. Then the function $f$ defined in Section 1.3.2 by the random variables from $\Sigma$ satisfies inequalities (a) to (e) and Theorem 1.3.9. Since $G$ has no isolated vertices, $f(v) \geq 1$ for all $v \in V(G)$ [13]. We have $\sum_{v \in V(G)} f(v) = \sum_{v \in IN(G)} f(v) + \sum_{v:\deg_G(v)=1} f(v) \geq \sum_{i=1}^{k} \sum_{v \in V_i} f(v) + |\{v|\deg_G(v) = 1\}| \geq \sum_{i=1}^{k}(2|V_i| - 1) + |\{v|\deg_G(v) = 1\}| = |V(G)| + in(G) - k$. Hence, $AR_\Sigma \geq \frac{1}{|V(G)|}(|V(G)| + in(G) - k)$ for any secret-sharing scheme $\Sigma$ on $G$. The result follows. ∎

Combining Corollary 3.1.1 and Theorem 3.1.2, we have the following results.

**Theorem 3.1.3.** *The inequality $c_{\mathcal{C}} \geq d_{\Pi}$ holds for any star covering $\Pi$ and core cluster $\mathcal{C}$ of a graph $G$. In particular, $c^*(G) \geq d^*(G)$.*

**Corollary 3.1.4.** *If there exists a star covering $\Pi$ and a core cluster $\mathcal{C}$ of a graph $G$ such that $c_{\mathcal{C}} = d_{\Pi}$, then $c^*(G) = c_{\mathcal{C}} = d_{\Pi} = d^*(G)$ and $AR(G) = \frac{|V(G)|+in(G)-c^*(G)}{|V(G)|}$.*

As indicated in this result, the equality $c^*(G) = d^*(G)$ makes a criterion for examining whether the lower bound and the upper bound on $AR(G)$ will match. We call $G$ *realizable* if $c^*(G) = d^*(G)$ holds. In the next section, we shall show that all trees are realizable.

## 3.2 The Exact Values of the Optimal Information Ratio of All Trees

Given a tree $T$, we let $IN(T)$ and $LF(T)$ be the sets of all internal vertices and leaves of $T$ respectively. Denote $|IN(T)|$ as $in(T)$ and $|LF(T)|$ as $lf(T)$. Blundo et al.[7] gave an algorithm for producing a star covering of a tree $T$. We make a slight modification to it and restate it for completeness. Let $N_T(v)$ be the set of all neighbors of $v$ in $T$ and $S_v$ be the star centered at $v$ with $N_T(v)$ as its leaf set.

***Algorithm;***

| **Covering($T$)** | **Cover($v$)** |
|---|---|
| Let $v \in IN(T)$ | $A(v) \leftarrow N_T(v) \cap IN(T)$ |
| $\Pi \leftarrow \phi$ | $\Pi \leftarrow \Pi \cup \{S_v\}$ |
| Cover($v$) | $E(T) \leftarrow E(T) \backslash E(S_v)$ |
| Output the star covering $\Pi$ | $V(T) \leftarrow V(T) \backslash ((N_T(v) \cap LF(T)) \cup \{v\})$ |
| | for all $v' \in A(v)$ do Cover($v'$) |

**Lemma 3.2.1.** *Let $T$ be a tree. The star covering $\Pi$ of $T$ produced by* Covering($T$) *has deduction $d_{\Pi} = 1$ if $T \neq K_{1,1}$ and $d_{\Pi} = 0$ if $T = K_{1,1}$.*

**Proof.** For $T \neq K_{1,1}$, the initial vertex $v$ and all leaves of $T$ appear in exactly one star in $\Pi$. All internal vertices but the initial one appear twice in the covering. So the vertex-number sum $m_\Pi = lf(T)+1+2(in(T)-1) = |V(T)|+in(T)-1$, and we have $d_\Pi = 1$. ∎

We shall refine this process and obtain star coverings with higher deductions next.

A vertex $v \in IN(T)$ is called a *critical vertex* of $T$ if $N_T(v) \cap LF(T) = \emptyset$. In the structure of a tree $T$, critical vertices play an important role in our discussion. We use $\mathbb{X}_T$ to denote the set of all critical vertices of $T$. Consider the subgraph $H_T$ of $T$ induced by $\mathbb{X}_T$ and let $\Lambda_T$ (resp. $\mathbb{Y}_T$) be the set of all nontrivial (resp. trivial) components in $H_T$. Then the set $\mathbb{Y}_T$ is in fact the set of all isolated vertices in $H_T$. So, $\mathbb{Y}_T$ can been seen as a subset of $\mathbb{X}_T$. In addition, for any $V' \subseteq V(T)$ and $E' \subseteq E(T)$, the graph $T - V'$ is obtained by removing from $T$ all vertices in $V'$ as well as the edges incident to them. $T - E'$ is resulted from removing all edges in $E'$ from $T$. Both $T - V'$ and $T - E'$ may contain isolated vertices.

**Proposition 3.2.2.** *Let $T \neq K_{1,1}$ be a tree. If $\Lambda_T = \emptyset$ and $|\mathbb{Y}_T| = y \geq 0$, then there exists a star covering $\Pi$ of $T$ with deduction $d_\Pi = y + 1$.*

**Proof.** Let $G$ be an arbitrary component in $T - \mathbb{Y}_T$. If $w_1, \ldots, w_l$ are all of the vertices in $\mathbb{Y}_T$ that are adjacent to some vertices in $G$, then we define $\tilde{G}$ as the subgraph of $T$ induced by $V(G) \cup \{w_1, \ldots, w_l\}$. Let $\mathcal{H} = \{\tilde{G} | G$ is a component in $T - \mathbb{Y}_T\}$ and $\Pi_{\tilde{G}}$ be the star covering produced by algorithm Covering($\tilde{G}$). By the definition of $\mathbb{Y}_T$, no $\tilde{G}$ is isomorphic to $K_{1,1}$, so $d_{\Pi_{\tilde{G}}} = 1$ by Lemma 3.2.1. Since $\bigcup_{\tilde{G} \in \mathcal{H}} E(\tilde{G}) = E(T)$, the covering $\Pi = \bigcup_{\tilde{G} \in \mathcal{H}} \Pi_{\tilde{G}}$ is a star covering of $T$ with vertex-number sum

$$\begin{aligned} m_\Pi &= \sum_{\tilde{G} \in \mathcal{H}} (|V(\tilde{G})| + in(\tilde{G}) - 1) \\ &= \left( V(T) + \sum_{v \in \mathbb{Y}_T} (\deg_T(v) - 1) \right) + (in(T) - y) \\ &\quad - \left( \sum_{v \in \mathbb{Y}_T} \deg_T(v) - (y-1) \right) \end{aligned}$$

$$= V(T) + in(T) - (y + 1).$$

∎

Next, we consider the core number of $T$. For a tree $T$ with $\mathbb{X}_T = \emptyset$, $\{IN(T)\}$ is obviously a core cluster of minimum size. The following lemma is straight forward.

**Lemma 3.2.3.** *Let $T \neq K_{1,1}$ be a tree. If $\mathbb{X}_T = \emptyset$, then $c^*(T) = 1$.*

Now, we introduce the way we decompose a tree in order to define a core cluster we need. Let $V' \subseteq V(T)$. Given a vertex $\tilde{v} \in N_T(v) \cap IN(T)$ for each $v \in V'$, we set $E' = \{v\tilde{v} | v \in V'\}$. For each component $G$ in $T - E'$, let $G^+$ be the subtree of $T$ obtained by attaching to $G$ all edges of the form $v\tilde{v}$ if $\tilde{v} \in V(G)$, then $G^+ = G$ if $G$ does not contain any $\tilde{v}$. We also denote the collection of all $G^+$'s, where $G$ is a component in $T - E'$, as $\mathcal{H}^+(T, V', E')$. Observe that, if $v \in V'$ and $\deg_T(v) = 2$, then $v \in LF(G^+)$ for exactly two $G^+$'s in the collection $\mathcal{H}^+(T, V', E')$.

**Proposition 3.2.4.** *Let $T \neq K_{1,1}$ be a tree. If $\Lambda_T = \emptyset$ and $|\mathbb{Y}_T| = y \geq 0$, then $c^*(T) = d^*(T) = y + 1$.*

**Proof.** It suffices to show that there is a core cluster of $T$ of size $y + 1$. For each $v \in \mathbb{Y}_T$, choose an arbitrary neighbor of $v$ as $\tilde{v}$, then $\tilde{v} \in IN(T)$. Let $E' = \{v\tilde{v} | v \in \mathbb{Y}_T\}$. There are $y + 1$ subgraphs in $\mathcal{H}^+(T, \mathbb{Y}_T, E')$. Let $\mathcal{H}^+(T, \mathbb{Y}_T, E') = \{G_0^+, G_1^+, \ldots, G_y^+\}$ where $G_i$'s, $i = 0, 1, \ldots, y$ are the components in $T - E'$. Note that any two vertices in $\mathbb{Y}_T$ have distance at least two, so $IN(G_i^+) \neq \emptyset$. Let $V_i = IN(G_i^+) \cup \{v | v \in V(G_i) \cap \mathbb{Y}_T$ and $\deg_T(v) = 2\}$. We claim that $\{V_0, V_1, \ldots, V_y\}$ is a core cluster of $T$. First, each vertex $u \in IN(T) \backslash \mathbb{Y}_T$ belongs to exactly one $IN(G_i^+)$ and also exactly one $V_i$. Each $v \in \mathbb{Y}_T$ belongs to exactly two $G_i^+$'s. If $\deg_T(v) \geq 3$, then $v$ is an internal vertex of one $G_i^+$ and a leaf of the other. It belongs to exactly one $IN(G_i^+)$ and hence exactly one $V_i$. If $\deg_T(v) = 2$, then $v$ is a leaf of exactly one component $G_i$ in $T - E'$ and is a leaf of two subgraphs in $\mathcal{H}^+(T, \mathbb{Y}_T, E')$. Hence it belongs to exactly one $V_i$ and none of the $IN(G_j^+)$'s, $j = 0, 1, \ldots, y$.

This shows that $\{V_0, V_1, \ldots, V_y\}$ is a partition of $IN(T)$. Next, each $V_i$ certainly induces a connected subgraph of $T$. In addition, each $v \in V_i \cap \mathbb{Y}_T$ has a neighbor $\tilde{v}$ not in $V_i$. Each $u \in V_i \backslash \mathbb{Y}_T$ has a leaf neighbor in $T$ which does not belongs to $V_i$. Hence, $V_i$ is a core of $T$. Since we have a core cluster of size $y+1$, the result then follows immediately by Proposition 3.2.2 and Corollary 3.1.4. ∎

Before literally proving our main theorem, we examine the relation between the deductions of star coverings of the subtrees in $\mathcal{H}^+(T, V', E')$ and the deduction of a star covering of $T$ more closely.

**Lemma 3.2.5.** *Let $V'$ be an independent subset of $IN(T)$ and $z = |\{v \in V' | \deg_T(v) \geq 3\}|$. For each $v \in V'$, let $\tilde{v}$ be a nonleaf neighbor of $v$ in $T$ and $E' = \{v\tilde{v} | v \in V'\}$. If there is a star covering $\Pi_{T'}$ of each $T' \in \mathcal{H}^+(T, V', E')$ with deduction $d_{\Pi_{T'}}$, then $\Pi = \bigcup_{T' \in \mathcal{H}^+(T,V',E')} \Pi_{T'}$ is a star covering of $T$ with deduction $d_\Pi = \sum_{T' \in \mathcal{H}^+(T,V',E')} d_{\Pi_{T'}} - z$.*

**Proof.** Denote $\mathcal{H}^+(T, V', E')$ as $\mathcal{H}^+$ for now. Since $\bigcup_{T' \in \mathcal{H}^+} E(T') = E(T)$, $\Pi$ is a star covering of $T$. The vertex-number sum $m_\Pi$ of $\Pi$ is

$$\begin{aligned} m_\Pi &= \sum_{T' \in \mathcal{H}^+} (|V(T')| + in(T') - d_{\Pi_{T'}}) \\ &= |V(T)| + |V'| + in(T) - (|V'| - z) - \sum_{T' \in \mathcal{H}^+} d_{\Pi_{T'}} \\ &= |V(T)| + in(T) - \left( \sum_{T' \in \mathcal{H}^+} d_{\Pi_{T'}} - z \right). \end{aligned}$$

∎

Now, we are in a position to present our main theorem in this chapter.

**Theorem 3.2.6.** *Any tree $T$ is realizable and*

$$AR(T) = \frac{n + in(T) - c^*(T)}{n}.$$

**Proof.** We prove this result by induction on $|\mathbb{X}_T|$.

(1) If $|\mathbb{X}_T| = 0$ or $1$, then $\Lambda_T = \emptyset$. The result holds by Proposition 3.2.4.

(2) Suppose that $|\mathbb{X}_T| \geq 2$. By Proposition 3.2.4, we may assume that $\Lambda_T \neq \emptyset$. Choose a vertex $v \in LF(T')$ for some $T' \in \Lambda_T$ and let $\tilde{v}$ be the neighbor of $v$ in $T'$. There are two subtrees $G_0^+$ and $G_1^+$ in $\mathcal{H}^+(T, \{v\}, \{v\tilde{v}\})$, each of which is not a $K_{1,1}$. Let $G_0^+$ be the one not containing $\tilde{v}$, then $|\mathbb{X}_{G_0^+}| < |\mathbb{X}_T|$ is obviously true. Since $v \in LF(G_1^+)$, it is no longer a critical vertex of $G_1^+$, we also have $|\mathbb{X}_{G_1^+}| < |\mathbb{X}_T|$. By induction hypothesis, there exist a star covering $\Pi_i$ of $G_i^+$ and a core cluster $\mathcal{C}_i = \{V_{i1}, V_{i2}, \ldots, V_{ik_i}\}$ with $d_{\Pi_i} = c_{\mathcal{C}_i} = k_i > 0$, $i = 0, 1$. Then $\Pi = \Pi_0 \cup \Pi_1$ is a star covering of $T$. We construct a core cluster of size $d_\Pi$ next.

(i) If $\deg_T(v) \geq 3$, then $d_\Pi = k_0 + k_1 - 1$ by Lemma 3.2.5. Suppose that $v \in V_{01}$. Since $V_{01}$ is a core of $G_0^+$, there is a designated outside neighbor $v'$ of $v$ in $G_0^+$ and outside $V_{01}$. Now, $v'$ is an internal vertex of $G_0^+$ because $v$ is critical both in $T$ and in $G_0^+$. We may assume that $v' \in V_{02}$. Now, let $\mathcal{C} = \{V_{01} \cup V_{02}, V_{03}, \ldots, V_{0k_0}, V_{11}, \ldots, V_{1k_1}\}$, then $|\mathcal{C}| = k_0 + k_1 - 1$. We claim that $\mathcal{C}$ is a core cluster of $T$. First note that $IN(G_0^+) \cup IN(G_1^+) = IN(T)$ and any two sets in $\mathcal{C}$ are disjoint. Each set in $\mathcal{C} \backslash \{V_{01} \cup V_{02}\}$ is a core of $G_0^+$ or $G_1^+$, hence a core of $T$. For $V_{01} \cup V_{02}$, $\tilde{v}$ is a neighbor of $v$ in $T$ not in $V_{01} \cup V_{02}$. Since $v \in LF(T')$, $v'$ is not critical and then has a leaf neighbor $v'' \neq v$ in $G_0^+$ (and in $T$) not in $V_{02}$, so $v'' \notin V_{01} \cup V_{02}$ is the designated outside neighbor of $v'$ with respect to $V_{01} \cup V_{02}$, and $V_{01} \cup V_{02}$ is qualified as a core of $T$. Therefore, $\mathcal{C}$ is a core cluster of $T$ of size $d_\Pi$.

(ii) If $\deg_T(v) = 2$, then $d_\Pi = k_0 + k_1$ by Lemma 3.2.5. Since $v$ is a critical vertex of $T$, the neighbor $v' \neq \tilde{v}$ in $T$ is an internal vertex of $G_0^+$. We may assume that $v' \in V_{01}$. Let $\mathcal{C} = \{V_{01} \cup \{v\}, V_{02}, \ldots, V_{0k_0}, V_{11}, \ldots, V_{1k_1}\}$, then $|\mathcal{C}| = k_0 + k_1$. To show that $\mathcal{C}$ is a core cluster of $T$, it suffices to show that $V_{01} \cup \{v\}$ is a core of $T$. Note that $v'$ is not critical in both $G_0^+$ and $T$. It has a leaf neighbor $v'' \neq v$ not in $V_{01} \cup \{v\}$ which serves as a qualified designated outside neighbor of $v'$ with respect to $V_{01} \cup \{v\}$. Besides, $\tilde{v}$ is also a qualified designated outside neighbor of

$v$ with respect to $V_{01} \cup \{v\}$. The set $V_{01} \cup \{v\}$ is indeed a core of $T$. Therefore, $T$ also has a core cluster of size $d_{\Pi}$ in this case.

In both cases, we have $c^*(T) = d^*(T)$, which implies that the lower bound and the upper bound on $AR(T)$ coincide. Hence, $AR(T) = \frac{n+in(T)-c^*(T)}{n}$. ∎

## 3.3 The Evaluation of $AR(T)$ for Some Classes of Trees Using Our Approach

In this section, we evaluate the optimal average information ratio systematically for two infinite classes of trees using our approach.

The only infinite class of trees which has known optimal average information ratio is the paths. By evaluating the core number, we can easily obtain the known result.

**Proposition 3.3.1** ([34]). *Let $P_n$ be a path of length $n$. Then*

$$AR(P_n) = \begin{cases} \frac{3n}{2(n+1)}, & \text{if } n \text{ is even;} \\ \frac{3n+1}{2(n+1)}, & \text{if } n \text{ is odd.} \end{cases}$$

**Proof.** By Proposition 3.2.4, we have $c^*(P_1) = 0$, $c^*(P_2) = c^*(P_3) = 1$ and $c^*(P_4) = 2$. Observe that $\Lambda_{P_n} = \{P_{n-4}\}$ for all $n \geq 5$. Since any leaf of the $P_{n-4}$ in $\Lambda_{P_n}$ has degree two in $P_n$, from the proof of Theorem 3.2.6, we have $c^*(P_n) = c^*(P_{n-4}) + 2$. Recursively, we have

$$\begin{aligned} c^*(P_n) &= \begin{cases} c^*(P_i) + 2k, & \text{if } n = 4k+i,\ i = 1,2,3; \\ c^*(P_4) + 2(k-1), & \text{if } n = 4k. \end{cases} \\ &= \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even;} \\ \frac{n-1}{2}, & \text{if } n \text{ is odd.} \end{cases} \end{aligned}$$

Hence,

$$AR(P_n) = \frac{(n+1)+(n-1)-c^*(P_n)}{n+1} = \begin{cases} \frac{3n}{2(n+1)}, & \text{if } n \text{ is even;} \\ \frac{3n+1}{2(n+1)}, & \text{if } n \text{ is odd.} \end{cases}$$

∎

Next, we evaluate the average information ratio of complete $q$-ary trees. A complete $q$-ary tree with $k$ levels is a rooted tree such that each nonleaf vertex has $q$ children and the distance from the root to each leaf is $k$.

**Theorem 3.3.2.** *Let $T_k$ be a complete $q$-ary tree with $k$ levels, $q \geq 2$. Then*

$$AR(T_k) = \begin{cases} \frac{q^{k+2}+2q^{k+1}-q^2-2q}{(q+1)(q^{k+1}-1)}, & \text{if } k \text{ is even;} \\ \frac{q^{k+2}+2q^{k+1}-q^2-q-1}{(q+1)(q^{k+1}-1)}, & \text{if } k \text{ is odd.} \end{cases}$$

**Proof.** By Proposition 3.2.4, $c^*(T_1) = 1$ and $c^*(T_2) = 2$. Observe that $\Lambda_{T_k} = \{T_{k-2}\}$, $k \geq 3$, and the $T_{k-2}$ has $q^{k-2}$ leaves, each of which has degree $q+1 \geq 3$ in $T_k$. Since each leaf of the $T_{k-2}$ and its descendants in $T_k$ compose a $T_2$, from the proof of Theorem 3.2.6, we get $c^*(T_k) = c^*(T_{k-2}) + q^{k-2}(c^*(T_2) - 1) = c^*(T_{k-2}) + q^{k-2}$. Recursively, the core number of $T_k$ can be evaluated as follows.

$$\begin{aligned} c^*(T_k) &= \begin{cases} q^{k-2} + q^{k-4} + \cdots + q^2 + c^*(T_2), & \text{if } k \text{ is even;} \\ q^{k-2} + q^{k-4} + \cdots + q + c^*(T_1), & \text{if } k \text{ is odd.} \end{cases} \\ &= \begin{cases} \frac{q^k+q^2-2}{q^2-1}, & \text{if } k \text{ is even;} \\ \frac{q^k+q^2-q-1}{q^2-1}, & \text{if } k \text{ is odd.} \end{cases} \end{aligned}$$

Therefore,

$$\begin{aligned} AR(T_k) &= \frac{\frac{q^{k+1}-1}{q-1} + \frac{q^k-1}{q-1} - c^*(T_k)}{\frac{q^{k+1}-1}{q-1}} \\ &= \begin{cases} \frac{q^{k+2}+2q^{k+1}-q^2-2q}{(q+1)(q^{k+1}-1)}, & \text{if } k \text{ is even;} \\ \frac{q^{k+2}+2q^{k+1}-q^2-q-1}{(q+1)(q^{k+1}-1)}, & \text{if } k \text{ is odd.} \end{cases} \end{aligned}$$

∎

## 3.4 Concluding Remark

We have proposed the idea of the deduction $d^*(G)$ and the core number $c^*(G)$ of a graph $G$ and showed that these values are the same for any tree

$T$, thereby proving the upper bound and the lower bound on the optimal average information ratio of a tree coincide. By doing so, we also present a systematic way of evaluating the core number of a tree.

In addition, the condition $d^*(G) = c^*(G)$ makes a criterion for examining whether the upper bound and the lower bound on $AR(G)$ will match. The idea formulates a complicated problem of secret-sharing schemes into a problem in graph theory with easy description. "For what kind of graphs will the identity be true?" is indeed an interesting question to investigate. One obvious restriction to set on $G$ is that $G$ must be of larger girth. A star covering generally does not serve as a complete multipartite covering with the least vertex-number sum for a graph of small girth. In the next chapter, we study the optimal average information ratio of bipartite graphs of larger girth. Finding a star covering whose deduction matches the size of a core cluster is in general very difficult. However, there have not been any bounds or asymptotic results on the complexity of the problem yet.

The results in this chapter have been included in the following paper.
"H.-C. Lu and H.-L. Fu, The exact values of the optimal average information ratio of perfect secret-sharing schemes for tree-based access structures, *Designs, Codes and Cryptography* (2013), http://dx.doi.org/10.1007/s10623-012-9792-1"

# Chapter 4

# The Average Information Ratio of Bipartite Graphs

## 4.1 Some Classes of Realizable Graphs

In this chapter, we need more definitions and notations to facilitate the whole discussion process for bipartite graphs. The girth of $G$ is written as $\mathrm{girth}(G)$. $N_G(v)$ denotes the set of all neighbors of $v$ in $G$ and $N(S) = \bigcup_{v \in S} N_G(v)$ for any $S \subseteq V(G)$. A vertex $v$ is called a *$k$-vertex* of $G$ if $\deg_G(v) = k$. Let $G = (X, Y)$ be a bipartite graph with bipartitions $X$ and $Y$. If $H$ is a subgraph of $G$, we use $X_H$ and $Y_H$ to denote $X \cap V(H)$ and $Y \cap V(H)$ respectively and then $H = (X_H, Y_H)$. In addition, let $X_H^{(k)} = \{x \in X_H | \deg_H(x) = k\}$ and $X_H^{k^+} = \{x \in X_H | \deg_H(x) \geq k\}$. The sets $Y_H^{(k)}$ and $Y_H^{k^+}$ are defined correspondingly. In the case when $H = G$, we use $X^{(k)}$ and $X^{k^+}$ for $X_G^{(k)}$ and $X_G^{k^+}$ respectively and also use $Y^{(k)}$ and $Y^{k^+}$ for $Y_G^{(k)}$ and $Y_G^{k^+}$ respectively for simplicity. In order to have a better description of our approach to the problem regarding bipartite graphs, we give an alternative definition of a core cluster of $G$. A *core cluster* $g$ of $G$ is defined as a vertex labeling $g : IN(G) \to \mathbb{N} \cup \{0\}$ such that each $g^{-1}(i)$, $i \in g(IN(G))$, is a core of $G$. The size $|g(IN(G))|$ of the clore cluster is denoted as $c_g$ in this chapter. The core number of $G$ is still written as $c^*(G)$. As a reminder, for any $V' \subseteq V(G)$ and any $E' \subseteq E(G)$, we do not remove resulting isolated vertices from the subgraphs $G - V'$ and $G - E'$. Each isolated vertex is considered as a trivial component in both subgraphs.

As we define an orientation on a specified trail $v_0 - v_1 - \cdots - v_l$ (the $v_i$'s may repeat) in the proof of Theorem 4.1.1, "orienting the trail from $v_0$ to $v_l$" means choosing the orientation $v_i \to v_{i+1}$ for each edge $v_iv_{i+1}$, $i = 0, 1, \ldots, l-1$, of this trail. For any subgraph $H$ of $G$, we denote as $S_v^H$ the star centered at $v$ and having all neighbors of $v$ in $H$ as its leaves. In what follows, we let $\Pi_X(H) = \{S_x^H | x \in X_H\}$ and $\Pi_Y(H) = \{S_y^H | y \in Y_H\}$. Both of them are star coverings of $H$. Unless otherwise specified, a graph $G = (X, Y)$ always represents a bipartite graph which contains no isolated vertices.

**Theorem 4.1.1.** *Let $G = (X, Y)$ with $|X| \geq |Y|$ and* $\text{girth}(G) \geq 6$. *If $\deg_G(x) \leq 2$ for all $x \in X$, then $G$ is realizable and $c^*(G) = |Y^{2^+}|$.*

**Proof.** Before constructing the desired core cluster, we define an orientation on $G$ first. (i) If $G$ contains a cycle $C$, then we start with an orientation on $C$ so that $C$ becomes a directed cycle. Next, we repeat the following process until all edges of $G$ are oriented. We take a $uv$-trail passing through unoriented edges where $u$ is a vertex to which at least two oriented edges are incident and $v$ is a 1-vertex or a repeated vertex on this trail or also a vertex to which at least two oriented edges are incident, and then we orient the trail from $u$ to $v$. Since $G$ is connected, we will eventually arrive at an orientation of $G$ by repeatedly doing this process. (ii) In the case when $G$ is a tree, counting the number of edges of $G$ gives $|X^{(1)}| + 2(|X| - |X^{(1)}|) = |X| + |Y| - 1 \leq 2|X| - 1$ which implies $|X^{(1)}| \geq 1$. Let $x_0 \in X^{(1)}$ be the root of $G$ and orient all edges toward the leaves. Now, we have the orientation we need. Observe that in both cases, each vertex $v \in IN(G)$ has at least one in-neighbor and one out-neighbor. Let us construct a core cluster of $G$ by virtue of this orientation. Initially, we label the vertices in $Y^{2^+}$ differently, that is, let $g : Y^{2^+} \to \{1, 2, \ldots, |Y^{2^+}|\}$ be a bijection. Next, we will extend the domain of $g$ to $IN(G)$ and keep the image of $g$ unchanged at the same time. For each $x \in X^{2^+}$, define $g(x) = g(y)$ if $(y, x)$ is an arc in the orientation.

Being a 2-vertex of $G$, $x$ has exactly one in-neighbor $y$, and then the extended labeling $g: IN(G) \to \{1, 2, \ldots, |Y^{2^+}|\}$ is well-defined.

We claim that $g$ is a core cluster of $G$. Note that each $y \in Y^{2^+}$ has at least one in-neighbor which is either a 1-vertex or a vertex $x \in X$ who receives the label from its in-neighbor $y' \neq y$. Hence each $y \in Y^{2^+}$ has a neighbor not in $g^{-1}(g(y))$. Similarly, each $x \in X^{2^+}$ receives the label from its in-neighbor $y \in Y^{2^+}$ and also has at least one out-neighbor $y' \neq y$ which is a 1-vertex or has initially gotten a label different from $y$'s. So each $x \in X^{2^+}$ also has at least one neighbor not in $g^{-1}(g(x))$. Now, each vertex in $g^{-1}(i)$ does have a neighbor outside $g^{-1}(i)$ and these outside neighbors of vertices in $g^{-1}(i)$ certainly form an independent set in $G$ because $g^{-1}(i)$ induces a connected subgraph of diameter at most two and $G$ has girth not less then six. This shows that $g$ is indeed a core cluster of size $|Y^{2^+}|$. On the other hand, the star covering $\Pi = \Pi_Y(G) = \{S_y^G | y \in Y\}$ has the vertex-number sum $m_\Pi = |V(G)| + |X^{2^+}|$ which gives the deduction $d_\Pi = |V(G)| + in(G) - m_\Pi = |Y^{2^+}|$. The proof is then completed. ∎

In a graph $G$, *k-subdividing* an edge is the operation of replacing the edge with a path of length $k$. A graph $G'$ is called an *even-subdivision* of $G$ if it is obtained by $2k_e$-subdividing each edge $e \in E(G)$, where $k_e \geq 1$.

**Corollary 4.1.2.** *If $G$ is a simple graph, then any even-subdivision $G'$ of $G$ is realizable. In addition, if $G'$ is obtained by $2k_e$-subdividing each edge $e$ of $G$ and $G$ is not a tree, then $AR(G') = \frac{|V(G)|-|E(G)|+3\sum_{e\in E(G)} k_e}{|V(G)|-|E(G)|+2\sum_{e\in E(G)} k_e}$.*

**Proof.** We may assume that $G$ is not a tree. Let $v_1^e, v_2^e, \ldots, v_{2k_e-1}^e$ be the consecutive internal vertices of the path in $G'$ that replaces the edge $e$ in $G$. Then $G'$ is a bipartite graph with bipartition $X = \{v_{2i+1}^e | e \in E(G),\ i = 0, 1, \ldots, k_e - 1\}$ and $Y = \{v_{2i}^e | e \in E(G),\ i = 1, \ldots, k_e - 1\} \cup V(G)$. So, $|X| = \sum_{e\in E(G)} k_e$ and $|Y| = \sum_{e\in E(G)}(k_e - 1) + |V(G)| = \sum_{e\in E(G)} k_e - |E(G)| + |V(G)| \leq |X|$. Since the girth of $G'$ is not less than six and $\deg_{G'}(x) = 2$ for all $x \in X$, we know that $G'$ is realizable by Theorem 4.1.1 and $c^*(G') = |Y^{2^+}| = \sum_{e\in E(G)}(k_e - 1) + in(G)$. Since $|V(G')| = \sum_{e\in E(G)}(2k_e - 1) + |V(G)|$

and $in(G') = \sum_{e\in E(G)}(2k_e - 1) + in(G)$, the optimal average information ratio of $G'$ can be easily evaluated as follows.

$$\begin{aligned} AR(G') &= \frac{|V(G')| + in(G') - c^*(G')}{|V(G')|} \\ &= \frac{|V(G)| - |E(G)| + 3\sum_{e\in E(G)} k_e}{|V(G)| - |E(G)| + 2\sum_{e\in E(G)} k_e}. \end{aligned}$$

∎

This proof actually also works when $G$ is not simple and $G'$ has girth not less than six.

**Corollary 4.1.3.** *If $G$ is a graph with loops and multiple edges, then any even-subdivision $G'$ of $G$ is realizable provided that $G'$ is of girth not less than six.*

**Theorem 4.1.4.** *Let $G = (X, Y)$ and $|X| \geq |Y|$. Suppose that* $\text{girth}(G) \geq 8$ *and $N_G(u) \cap N_G(v) \cap Y^{3^+} = \emptyset$ for all distinct $u, v \in X^{3^+}$. If for each $v \in X^{3^+}$, there exists a set $N^-(v) = \{v_i | i = 1, \ldots, \deg_G(v) - 1\} \subseteq IN(G) \cap N_G(v)$ such that each component $\widetilde{G}$ in $G - E'$, where $E' = \{vv_i | v_i \in N^-(v), v \in X^{3^+}\}$, satisfies $|X_{\widetilde{G}}| \geq |Y_{\widetilde{G}}|$, then $G$ is realizable and $c^*(G) = |Y^{2^+}| - \sum_{v\in X^{3^+}}(\deg_G(v) - 2)$.*

**Proof.** First note that for all distinct $u, v \in X^{3^+}$, $N^-(u)$ and $N^-(v)$ are disjoint because a vertex in $N^-(u) \cap N^-(v)$ would otherwise turn out to be a trivial component in $G - E'$ which violates the assumption. Now let us initially define $g : Y^{2^+} \to \{1, 2, \ldots, |Y^{2^+}|\}$ to be a bijection and then, for each $v \in X^{3^+}$, we further define $g(v) = g(v_1)$ and alter the labels of $v_i$'s, $i \geq 2$, by redefining $g(v_i) = g(v_1)$ for $i = 2, 3, \ldots, \deg_G(v) - 1$. After this alteration, $|g(Y^{2^+} \cup X^{3^+})| = |Y^{2^+}| - \sum_{v\in X^{3^+}}(\deg_G(v) - 2)$.

Let $\{\widetilde{G}_i = (X_{\widetilde{G}_i}, Y_{\widetilde{G}_i}) | i = 1, 2, \ldots, s\}$ be the collection of all components in $G - E'$. Applying the construction of a core cluster used in the proof of Theorem 4.1.1 to each $\widetilde{G}_i$ if $\widetilde{G}_i \neq K_{1,1}$, we extend the domain of $g|_{Y^{2^+} \cap IN(\widetilde{G}_i)}$ to $IN(\widetilde{G}_i)$ and keep its image unchanged. As a consequence, we have jointly extended the domain of $g$ to $IN(G)$ and keep its image unchanged.

Next, it will be verified that $g^{-1}(g(u))$ is a core for each $u \in IN(G)$. If $u = v$ for some $v \in X^{3^+}$, there exists $y' \in N_G(v)\backslash N^-(v)$ which is either a 1-vertex or has a different label from $v$'s because $y'$ was initially given a label different from $v_1$'s and has never been altered. If $u = v_i \in N^-(v)$ for some $v \in X^{3^+}$ and $\deg_G(v_i) = 2$, then $v_i$ is a 1-vertex of some $\widetilde{G}_j$. According to the manner we extend $g|_{Y^{2+} \cap IN(\widetilde{G}_j)}$, the neighbor of $v_i$ in $\widetilde{G}_j$ has a label different from $v_i$'s. Finally, if $u \in IN(G)\backslash X^{3^+}\backslash\{v_i \in N^-(v)|v \in X^{3^+}, \deg_G(v_i) = 2\}$, then $u \in IN(\widetilde{G}_j)$ for some $j$. It has been shown that $u$ has a neighbor in $\widetilde{G}_j$ which is either a 1-vertex or has a label different from $u$'s in the proof of Theorem 4.1.1. Hence, each vertex $u \in IN(G)$ has a neighbor not in $g^{-1}(g(u))$. These outside neighbors of the vertices in $g^{-1}(g(u))$ certainly form an independent set in $G$ because $g^{-1}(g(u))$ induces a connected subgraph of diameter at most four and the girth of $G$ is at least eight. We conclude that $g$ is a core cluster of $G$ of size $|g(IN(G))| = |Y^{2^+}| - \sum_{v \in X^{3+}}(\deg_G(v) - 2)$. On the other hand, the star covering $\Pi = \Pi_Y(G)$ has the vertex-number sum

$$\begin{aligned} m_\Pi &= |V(G)| + \sum_{v \in X^{2^+}} (\deg_G(v) - 1) \\ &= |V(G)| + \sum_{v \in X_{3^+}} (\deg_G(v) - 2) + |X^{2^+}|. \end{aligned}$$

Therefore, it has deduction $d_\Pi = |V(G)| + |X^{2^+}| + |Y^{2^+}| - m_\Pi = |g(IN(G))|$ as desired and the proof is completed. ∎

A component $H$ in $G - X^{3^+}$ with $|X_H| \geq |Y_H|$ will give rise to a component $H^*$ in $G - E'$ with $|X_{H^*}| \geq |Y_{H^*}|$. We have a complete characterization of this kind of components. In the next lemma, we consider a more general case for later use.

**Lemma 4.1.5.** *Let $G = (X, Y)$ with $|X| \geq |Y|$ and $H = (X_H, Y_H)$ be a component in $G - S$ for some $S$ satisfying $X^{3^+} \subseteq S \subseteq X$. Then $|X_H| \geq |Y_H|$ if and only if $H$ contains a cycle or $H$ is a tree with at least one leaf in $X_H$.*

**Proof.** For each $x \in X_H$, $\deg_H(x) = \deg_G(x) \leq 2$. Since $H$ is connected, counting the edges of $H$ gives $|X_H^{(1)}| + 2|X_H^{(2)}| \geq |X_H| + |Y_H| - 1$ which implies that $|X_H^{(2)}| \geq |Y_H|-1$. In addition, $H$ is a tree if and only if $|X_H^{(2)}| = |Y_H|-1$. Now, it is clear that $|X_H| < |Y_H|$ and $|X_H^{(2)}| \geq |Y_H|-1$ if and only if $|X_H^{(1)}| = 0$ and $|X_H^{(2)}| = |Y_H| - 1$. The result follows immediately. ∎

Let $G = (X, Y)$ with $|X| \geq |Y|$ and $S$ satisfy $X^{3^+} \subseteq S \subseteq X$. Then, a component $H$ in $G - S$ with $|X_H| \geq |Y_H|$ is called a *proper component* in $G - S$. A component in $G - S$ is *improper* if it is not proper. In other words, an improper component $H$ in $G - S$, where $X^{3^+} \subseteq S \subseteq X$, is a tree component with all its leaves in $Y_H$.

Theorem 4.1.4 suggests that proper components in $G - X^{3^+}$ do not hinder $G$ from being realizable. However, improper components may cause trouble while constructing core clusters of $G$. To deal with improper components in $G - X^{3^+}$, it will be convenient to define an *improper-component-adjacency graph* $A_G$ as follows. Let $\mathbb{U}_0 = \{T_i | i \in I_0\}$ be the collection of improper components in $G - X^{3^+}$ and let $\widetilde{X}^{3^+} = \{v \in X^{3^+} | v$ is adjacent to some $T_i \in \mathbb{U}_0$ in $G\}$. The improper-component-adjacency graph is a bipartite graph $A_G = (\mathbb{U}_0, \widetilde{X}^{3^+})$ such that for all $T_i \in \mathbb{U}_0$ and $v \in \widetilde{X}^{3^+}$, $(T_i, v)$ is an edge in $A_G$ if and only if $v$ is adjacent to some vertex of $T_i$ in $G$. Suppose that $M_0 = \{(T_j, v_j) | j \in J_0\}$ ($J_0 \subseteq I_0$) is a maximum matching in $A_G$. Each $T_i$, $i \in I_0 \backslash J_0$, is called an *excess improper component* of $G$. The number of excess improper components of $G$ is independent of the choices of the maximum matchings. We denote the number $|I_0 \backslash J_0|$ as $\mathrm{exc}(G)$. This parameter plays an important role in finding $c^*(G)$ and $d^*(G)$.

We take care of star coverings first. In what follows, we shall identify a subgraph $G'$ of $G$ and show that $\Pi = \Pi_X(G') \cup \Pi_Y(G - G')$ is an optimal star covering of $G$. Note that the graph $G - G'$ is obtained by removing all edges of $G'$ as well as the resulting isolated vertices from $G$.

**Lemma 4.1.6.** *Suppose that $G'$ is a subgraph of $G = (X, Y)$ with $|X| \geq |Y|$ and $V(G') \cap V(G - G') \subseteq X$. Then, the deduction of the star covering $\Pi = \Pi_X(G') \cup \Pi_Y(G - G')$ is given as*

$$d_\Pi = |Y^{2^+}| - \sum_{v \in X^{3^+}} (\deg_G(v) - 2) + |Y_{G'}| - |X_{G'}|.$$

**Proof.** Denote $G - G'$ as $G_0$ for now. Let $S = V(G') \cap V(G_0)$ and $|S| = s$. The vertex-number sum $m_\Pi$ of $\Pi$ can be evaluated as follows.

$$\begin{aligned}
m_\Pi &= |V(G')| + \sum_{y \in Y_{G'}} (\deg_{G'}(y) - 1) + |V(G_0)| + \sum_{x \in X_{G_0}} (\deg_{G_0}(x) - 1) \\
&= |V(G)| + s + \sum_{x \in X_{G'}} \deg_{G'}(x) - |Y_{G'}| + \sum_{x \in X_{G_0}} \deg_{G_0}(x) - |X_{G_0}| \\
&= |V(G)| + s + \sum_{x \in X} \deg_G(x) - |X_{G_0}| - |Y_{G'}| \\
&= |V(G)| + s + \left( \sum_{x \in X^{3+}} (\deg_G(x) - 2) + |X| + |X^{2^+}| \right) - |X_{G_0}| - |Y_{G'}| \\
&= |V(G)| + in(G) - \Bigg[ |X^{2^+}| + |Y^{2^+}| - s \\
&\qquad - \left( \sum_{x \in X^{3+}} (\deg_G(x) - 2) + |X| + |X^{2^+}| \right) + |X_{G_0}| + |Y_{G'}| \Bigg] \\
&= |V(G)| + in(G) - \left[ |Y^{2^+}| - \sum_{x \in X^{3+}} (\deg_G(x) - 2) - |X_{G'}| + |Y_{G'}| \right]
\end{aligned}$$

In the last step, we use the fact that $|X| + s = |X_{G_0}| + |X_{G'}|$. Therefore, we have the deduction as desired. ∎

**Lemma 4.1.7.** *Suppose that $G = (X, Y)$ with $|X| \geq |Y|$ and $X^{3^+} \subseteq S \subseteq X$. Let $\mathbb{U}$ be the collection of all components in $G - S$. If every component $H$ in $G - S$ is improper, namely, $|X_H| < |Y_H|$, then $|\mathbb{U}| - |S| = |Y| - |X|$.*

**Proof.** Since every component $H \in \mathbb{U}$ is improper, $H$ is a tree with all leaves in $Y_H$ and then $\deg_G(x) = 2$ for all $x \in X_H$. Counting the edges of $H$ gives $2|X_H| = |X_H| + |Y_H| - 1$ which implies $|X_H| = |Y_H| - 1$. As a consequence, we have $|Y| - |X| = \sum_{H \in \mathbb{U}} |Y_H| - (\sum_{H \in \mathbb{U}} |X_H| + |S|) = |\mathbb{U}| - |S|$. ∎

The notion of a maximum matching in a bipartite graph and a cut in a network is at the core of our process of identifying the subgraph $G'$ in $G$. We recall some basic properties before further discussion. We follow the terms and notations used in [36] in the following review.

Given a matching $M$ in $G$, an *$M$-augmenting path* is a path that alternates between edges in $M$ and edges not in $M$ and the endpoints of the path are unsaturated by $M$. It is well known that a matching $M$ in a graph $G$ is a maximum matching in $G$ if and only if $G$ has no $M$-augmenting path.

A *network* $N$ is a digraph with a nonnegative *capacity* $c(e)$ on each edge $e$ and a distinguished *source vertex* $s$ and *sink vertex* $t$. A *flow* $f$ assigns a value $f(e)$ to each edge satisfying $0 \le f(e) \le c(e)$ and the in-flow $f^-(v)$ and the out-flow $f^+(v)$ of each vertex $v \notin \{s,t\}$ are the same. Given a flow $f$ in a network, an *$f$-augmenting path* is a source-sink path $P$ in the underlying graph such that, for each $e \in E(P)$, (i) if $P$ follows $e$ in the forward direction, then $f(e) < c(e)$, and (ii) if $P$ follows $e$ in the backward direction, then $f(e) > 0$. In a network $N$, a *source/sink cut* $[S,T]$ consists of the edges from the *source set* $S$ to the *sink set* $T$, where $S$ and $T$ partition $V(N)$ with $s \in S$ and $t \in T$. The capacity of the cut is the total of the capacity on the edges of $[S,T]$. The well-known Ford-Fulkerson algorithm [20] produces an $f$-augmenting path or a cut with capacity $f^-(t) - f^+(t)$ in a network. We will take advantage of it in our approach later.

We are now in a position to introduce our star coverings.

**Theorem 4.1.8.** *If $G = (X,Y)$ and $|X| \ge |Y|$, then there exists a star covering $\Pi$ of $G$ with $d_\Pi = |Y^{2^+}| - \sum_{v \in X^{3^+}} (\deg_G(v) - 2) + \mathrm{exc}(G)$.*

**Proof.** Let $\mathbb{H}_0$ and $\mathbb{U}_0 = \{T_i | i \in I_0\}$ be the collection of all proper and improper components in $G - X^{3^+}$, respectively. Suppose that the improper-component-adjacency graph $A_G = (\mathbb{U}_0, \widetilde{X}^{3^+})$ has a maximum matching $M_0 = \{(T_j, v_j) | j \in J_0\}$, $J_0 \subseteq I_0$, and let $X^{(M)} = \{v_j \in \widetilde{X}^{3^+} | j \in J_0\}$.

**Case 1.** If $J_0 = I_0$, that is, $\mathrm{exc}(G) = 0$, we have shown that $\Pi = \Pi_Y(G)$ has the given deduction in the proof of Theorem 4.1.4.

**Case 2.** If $J_0 \subsetneq I_0$, then $\mathrm{exc}(G) = |I_0 \backslash J_0| > 0$. Consider the subgraph $G_1$ defined as the union of nontrivial components in $G - (\bigcup_{H \in \mathbb{H}_0} H)$ containing some excess improper component $T_i$, where $i \in I_0 \backslash J_0$. Let $\mathbb{U}_1 = \{T_i | i \in I_1\}$, $I_1 \subseteq I_0$, be the subset of $\mathbb{U}_0$ consisting of the $T_i$'s, $i \in I_0$, which are contained in $G_1$.

Denote $A_1 = A_G|_{G_1} = (\mathbb{U}_1, X_{G_1} \cap X^{3^+})$ in which for all $T \in \mathbb{U}_1$ and $v \in X_{G_1} \cap X^{3^+}$, $(T, v) \in E(A_1)$ if $(T, v) \in E(A_G)$. Then $A_1$ is an induced subgraph of $A_G$. Note that $A_1$ may differ from $A_{G_1}$ because in general $X_{G_1} \cap X^{3^+} \neq X^{3^+}_{G_1} = \{x \in X_{G_1} | \deg_{G_1}(x) \geq 3\}$. Let us examine the matching $M_1 = M_0|_{A_1} = \{(T_j, v_j) | (T_j, v_j) \in E(A_1) \cap M_0\}$ in $G_1$ more closely. Let $M_1 = \{(T_j, v_j) | j \in J_1\}$, $J_1 \subseteq J_0$. Observe that, by the definition of $G_1$, for each $(T_j, v_j) \in M_0$, we have $T_j \in \mathbb{U}_1$ if and only if $v_j \in X_{G_1} \cap X^{3^+}$. So, each edge in $M_0 \backslash M_1$ is not incident to any vertex in the subgraph $A_1$ of $A_G$. This fact guarantees that $M_1$ is a maximum matching in $A_1$ because any maximum matching $M'$ in $A_1$ would otherwise result in a matching $M = M' \cup (M_0 \backslash M_1)$ in $A_G$ with $|M| > |M_0|$, giving a contradiction. Since each $T_i$, $i \in I_0 \backslash J_0$, belongs to $\mathbb{U}_1$, we have $|I_1 \backslash J_1| = |I_0 \backslash J_0| = \mathrm{exc}(G)$.

(i) If $X_{G_1} \cap X^{3^+} \subseteq X^{(M)}$, then $M_1$ saturates $X_{G_1} \cap X^{3^+}$ and thus $|X_{G_1} \cap X^{3^+}| = |M_1| = |J_1|$. Now, $G_1$ is a bipartite graph in which every component in $G_1 - (X_{G_1} \cap X^{3^+})$ is improper and $X^{3^+}_{G_1} \subseteq X_{G_1} \cap X^{3^+} \subseteq X_{G_1}$. By Lemma 4.1.7, we have $|Y_{G_1}| - |X_{G_1}| = |\mathbb{U}_1| - |X_{G_1} \cap X^{3^+}| = |I_1| - |J_1|$. With the aid of Lemma 4.1.6, the deduction of the star covering $\Pi = \Pi_X(G_1) \cup \Pi_Y(G - G_1)$ can be easily calculated as $d_\Pi = |Y^{2^+}| - \sum_{v \in X^{3^+}} (\deg_G(v) - 2) + \mathrm{exc}(G)$.

(ii) If $(X_{G_1} \cap X^{3^+}) \backslash X^{(M)} \neq \emptyset$, then the vertices in $(X_{G_1} \cap X^{3^+}) \backslash X^{(M)}$ are not incident to any edge in $M_1$. In this case, we transform the graph $A_1$ into a network $A'$ through the following process. First, we define an orientation on $A_1$ by choosing $T_j \to v_j$ for $(T_j, v_j) \in M_1$ and $v_i \to T_l$ if $(T_l, v_i) \in E(A_1) \backslash M_1$. Second, let $A'$ be the graph obtained from the oriented $A_1$ by identifying all vertices in $(X_{G_1} \cap X^{3^+}) \backslash X^{(M)}$ and then renaming the resulting new vertex as the source vertex $s$, and also by identifying all $T_i$'s for $i \in I_1 \backslash J_1$ and then renaming the resulting vertex as the sink vertex $t$. Additionally, we assign the capacity $c(e) = 1$ to each $e \in E(A')$ and let $f$ be a zero flow in $A'$. Now, applying the Ford-Fulkerson algorithm to the network $A'$, we claim that the result from carrying out this algorithm must be a cut $[S, \overline{S}]$ as opposed to an $f$-augmenting path. For simplicity, let us call the edges in $M_1$ red and the other edges in $A'$ black. Observe that at each $T_j$, $j \in J_1$, the only leaving edge is red and each entering edge (if there is any) is black, whereas at each $v_j$, $j \in J_1$, the only entering edge is red and each leaving edge (if there is any) is black. If this algorithm results in an $f$-augmenting path $s - T_{j_1} - v_{j_1} - T_{j_2} - v_{j_2} - \cdots - T_{j_k} - v_{j_k} - t$, then each $(T_{j_i}, v_{j_i})$ must be red and the remaining edges must be black. This path naturally corresponds to an $M_1$-augmenting path in $A_1$ which contradicts to the fact that $M_1$ is a maximum matching in $A_1$.

Now we have a cut $[S, \overline{S}]$ from this algorithm. Define $G_2$ to be the subgraph of $G_1$ induced by $\{v|v \in V(T_i) \text{ where } T_i \in \overline{S}\backslash\{t\} \text{ or } i \in I_1\backslash J_1\} \cup (X_{G_1} \cap X^{3^+} \cap \overline{S})$. In order to have a better understanding of $G_2$, we need to point out some features of the cut $[S, \overline{S}]$. When this algorithm is running, some $T_j$, $j \in J_1$, must be reached. Searching from such $T_j$ reaches exactly one vertex $v_j$, where $(T_j, v_j) \in M_1$. It can not go any further only when the reached $v_j$ has no leaving edges. Hence, for each $(T_j, v_j) \in M_1$, $T_j \in S$ if and only if $v_j \in S$, and each edge in $[S, \overline{S}]$ is a black one of the form $T_j \leftarrow v_l$ where $T_j \in S$, $v_l \in \overline{S}$ and $(T_j, v_l) \notin M_1$. Now, it is clear that if $T_j \in \overline{S}$, then all its neighbors in $A_1$ lie in $\overline{S}$ as well. This accounts for the fact $V(G_2) \cap V(G - G_2) \subseteq X$. By Lemma 4.1.6, the deduction of the star covering $\Pi = \Pi_X(G_2) \cup \Pi_Y(G - G_2)$ has deduction $d_\Pi = |Y^{2^+}| - \sum_{v \in X^{3^+}} (\deg_G(v) - 2) + |Y_{G_2}| - |X_{G_2}|$.

This proof will be completed after the equality $|Y_{G_2}| - |X_{G_2}| = \text{exc}(G)$ is assured. Let $\mathbb{U}_2 = \{T_i | T_i \in \overline{S}\backslash\{t\} \text{ or } i \in I_1\backslash J_1\}$, $A_2 = A_1|_{G_2}$ and $M_2 = M_1|_{G_2}$, then $A_2 = (\mathbb{U}_2, X_{G_1} \cap X^{3^+} \cap \overline{S})$. Now, each vertex $v \in (X_{G_1} \cap X^{3^+})\backslash X^{(M)}$ unsaturated by $M_1$ has been excluded from $G_2$ and $A_2$. The vertices in $(X_{G_1} \cap X^{3^+} \cap \overline{S})$ are saturated by $M_2$ and thus $|M_2| = |X_{G_1} \cap X^{3^+} \cap \overline{S}|$. Since each $T_i$, $i \in I_1\backslash J_1$, belongs to $\mathbb{U}_2$, $|I_1\backslash J_1| = |\mathbb{U}_2| - |M_2|$ holds obviously. We finally reach to a bipartite graph $G_2$ in which each component in $G_2 - (X_{G_1} \cap X^{3^+} \cap \overline{S})$ is improper and $X_{G_2}^{3^+} \subseteq (X_{G_1} \cap X^{3^+} \cap \overline{S}) \subseteq X_{G_2}$.

By Lemma 4.1.7, we conclude that $|Y_{G_2}| - |X_{G_2}| = |\mathbb{U}_2| - |X_{G_1} \cap X^{3^+} \cap \overline{S}| = |\mathbb{U}_2| - |M_2| = |I_1 \backslash J_1| = \mathrm{exc}(G)$. ∎

Let us denote this crucial value $|Y^{2^+}| - \sum_{v \in X^{3+}} (\deg_G(v) - 2) + \mathrm{exc}(G)$ as $\beta(G)$ in the remainder of this chapter. Note that our star covering is in fact a star decomposition (which requires that each edge of $G$ appears in exactly one star) of the bipartite graph $G$. It will be shown that $\beta(G)$ meets the size of some core clusters for certain classes of bipartite graphs, thereby proving the star covering (decomposition) we propose is optimal for each of those graphs. Although it would not be of the least vertex-number sum among all complete multipartite coverings (decompositions), we strongly believe that it is an optimal star covering (decomposition) for all bipartite graphs.

Next, we turn to the construction of our core clusters.

**Lemma 4.1.9.** *Let $G = (X, Y)$ and $X' \subseteq X^{3^+}$. Given a neighbor $v^*$ of each $v \in X'$ and let $N^-(v) = N_G(v) - \{v^*\}$ for all $v \in X'$. If $N^-(u) \cap N^-(v) \cap N^-(w) = \emptyset$, for all distinct $u, v, w \in X'$, then there exists $v^+ \in N^-(v)$ for each $v \in X'$ such that all $v^+$'s are distinct.*

**Proof.** Let us consider the bipartite graph $B = (X', \bigcup_{v \in X'} N^-(v))$ in which, for all $v \in X'$ and $y \in \bigcup_{v \in X'} N^-(v)$, $(v, y) \in E(B)$ if and only if $y \in N^-(v)$. Since $\deg_B(v) = \deg_G(v) - 1 \geq 2$ for all $v \in X'$ and $\deg_B(y) \leq 2$ for all $y \in \bigcup_{v \in X'} N^-(v)$, we have $|N_B(S)| \geq \frac{2|S|}{2} = |S|$ for all $S \subseteq X'$. By Hall's Theorem, there is a matching $M_B = \{(v, v^+) | v \in X', v^+ \in \bigcup_{v \in X'} N^-(v)\}$ which saturates $X'$. ∎

In the remainder of this chapter, $l(C)$ denotes the length of the cycle $C$ in $G$. We give another description of the criteria for examining whether a lableing of $IN(G)$ is a core cluster.

**Lemma 4.1.10.** *Let $G$ be a simple graph. Then a labeling $g : IN(G) \to \mathbb{N} \cup \{0\}$ is a core cluster of $G$ if the following conditions are satisfied.*

*(i) $g^{-1}(i)$ induces a connected subgraph of $G$ for all $i \in g(IN(G))$;*

*(ii)* *any vertex $v \in IN(G)$ has a neighbor $w \in N_G(v)$ such that $w \notin IN(G)$ or $g(w) \neq g(v)$;*

*(iii)* *each cycle $C$ in $G$ contains at most $l(C)-4$ consecutive edges in every subgraph induced by $g^{-1}(i)$, $i \in g(IN(G))$.*

**Proof.** Conditions (ii) ensures that each vertex in $g^{-1}(i)$, $i \in g(IN(G))$, has a neighbor outside $g^{-1}(i)$. Condition (iii) in turn guarantees that any two of these outside neighbors are of distance at least two and each of them is adjacent to only one vertex in $g^{-1}(i)$. Hence, $g$ is a core cluster of $G$. ∎

Now, we present the construction of our core clusters.

**Lemma 4.1.11.** *Let $G=(X,Y)$ with $|X| \geq |Y|$. Then there exists a labeling $g : IN(G) \to \mathbb{N}$ satisfying criterion (i) in Lemma 4.1.10. Moreover, if $g$ satisfies criterion (iii) in Lemma 4.1.10, then $g$ is a core cluster of $G$ and $|g(IN(G))| = \beta(G)$.*

**Proof.** (a) First, let us consider the case where each vertex in $X^{3^+}$ has at most one 1-vertex neighbor. Let $\mathbb{H}_0$ and $\mathbb{U}_0 = \{T_i | i \in I_0\}$, $I_0 \subseteq \mathbb{N}$, be the collection of proper and improper components in $G - X^{3^+}$ respectively. Suppose $M_0 = \{(T_j, v_j) | j \in J_0\}$, $J_0 \subseteq I_0$, is a maximum matching in the improper-component-adjacency graph $A_G = (\mathbb{U}_0, \widetilde{X}^{3^+})$. If $v \in X^{3^+}$ has a 1-vertex neighbor $y$, then $\{y\}$ is a trivial component in $\mathbb{U}_0$ and $v = v_j$ for some $j \in J_0$. In this case, we may assume that $T_j = \{y\}$. Now choose $v_j^* \in V(T_j) \cap N_G(v_j)$ for each $j \in J_0$ and choose $v^*$ arbitrarily from $N_G(v)$ for each $v \in X^{3^+} \backslash \{v_j | j \in J_0\}$. Let $N^-(v) = N_G(v) \backslash \{v^*\}$ for each $v \in X^{3^+}$ and $Y^\bullet = \{y | N_G(y) \subseteq X^{3^+}$ and $y \neq v^*$ for all $v \in X^{3^+}\}$. For each $H \in \mathbb{H}_0 \cup \mathbb{U}_0$, let $H^*$ be the graph obtained by attaching to $H$ each edge $vv^*$ with $v \in X^{3^+}$ and $v^* \in V(H)$, and $H^* = H$ if $H$ does not contain any vertex in $\{v^* | v \in X^{3^+}\}$. Observe that the collection of components in $G - E'$, where $E' = \{vw | w \in N^-(v), v \in X^{3^+}\}$, is exactly $\{H^* | H \in \mathbb{H}_0 \cup \mathbb{U}_0\}$, among which the improper ones are $\{T_i | i \in I_0 \backslash J_0\}$. Now, for each $y \in Y^\bullet$, let $N'(y)$ be a subset of $N_G(y) \subseteq X^{3^+}$ consisting of $\deg_G(y) - 2$ arbitrary neighbors of $y$ and let

$X' = X^{3^+} \backslash (\bigcup_{y \in Y^\bullet} N'(y))$. Then $N^-(u) \cap N^-(v) \cap N^-(w) = \emptyset$ for all distinct $u, v, w \in X'$. With the aid of Lemma 4.1.9, we have distinct $v^+$'s for $v \in X'$, where $v^+ \in N_G(v)$. For each $y \in Y^\bullet$, let $y^*$ be the vertex in $N_G(y) \backslash (N'(y) \cup \{v\})$ if $y = v^+$ for some $v \in X'$, and be any vertex in $N_G(y) \backslash N'(y)$ otherwise. Now, consider the subgraph $K$ induced by $X^{3^+} \cup (\bigcup_{v \in X^{3^+}} N^-(v))$ and all the components $O_1, O_2, \ldots, O_s$ in $K - \{uu^* | u \in X^{3^+} \cup Y^\bullet\}$. It is worth nothing that each $O_i$ contains at least one $v \in X'$ and its neighbor $v^+ \in Y^{2^+}$, or at least one $v \in X^{3^+}$ and its neighbor $y \in Y^\bullet$ where $v \in N'(y)$. In addition, if $v \in X^{3^+} \backslash V(O_i)$ and $v$ is adjacent (in $G$) to a vertex $y$ of $O_i$, then $y = v^*$ for some $v \in X^{3^+}$ or $v = y^*$ for some $y \in Y^\bullet$. With these facts in mind, we now start to define the desired labeling $g$. Initially, we define $g$ to be a bijection from $Y^{2^+}$ to $\{1, 2, \ldots, |Y^{2^+}|\}$. Next, for each $i \in \{1, \ldots, s\}$, we choose a vertex $y_i \in V(O_i) \cap Y^{2^+}$ and then extend the domain of $g$ to $Y^{2^+} \cup X^{3^+}$ and alter some labels in $V(K) \cap Y^{2^+}$ by redefining $g(w) = g(y_i)$ for all $w \in V(O_i)$. To evaluate the cardinality of the image of the extended $g$, we define $Y^\bullet(v) = \{y \in Y^\bullet | v = y^*\}$ for each $v \in X^{3^+}$, then these $Y^\bullet(v)$'s are disjoint and $\sum_{v \in X^{3^+}} |Y^\bullet(v)| = |Y^\bullet|$. If each $V(O_i)$ does not induce cycles in $G$, then $|g(Y^{2^+} \cup X^{3^+})| = |Y^{2^+}| - \sum_{v \in X^{3^+}} (\deg_G(v) - 2 - |Y^\bullet(v)|) = |Y^{2^+}| - \sum_{v \in X^{3^+}} (\deg_G(v) - 2) + |Y^\bullet|$.

For each $H^* \in \{H^* | H \in \mathbb{H}_0 \cup \{T_j | j \in J_0\}\}$ and $H^* \neq K_{1,1}$, since $|X_{H^*}| \geq |Y_{H^*}|$, the labeling $g|_{Y^{2+}_{H^*}}$ can be extended to a core cluster of $H^*$ with its image kept unchanged as what we have done in the proof of Theorem 4.1.1. Next, for each improper component $T_i^* = T_i$ in $G - E'$, if $T_i$ is trivial, then $T_i = \{y\}$ for some $y \in Y^\bullet$ because we assume that each $v \in X^{3^+}$ has at most one 1-vertex neighbor $v^*$ and the trivial component $\{v^*\}$ is saturated by $M_0$ in $A_G$. Since each vertex in $Y^\bullet$ has been labeled, it remains to label nontrivial improper components. If $T_i$, $i \in I_0 \backslash J_0$, is nontrivial, then all its leaves are in $Y_{T_i}$ and $|X_{T_i}| = |X^{(2)}_{T_i}| > 0$. Let us choose a vertex $x_0 \in X_{T_i}$ which has a leaf neighbor in $T_i$ and then root $T_i$ at $x_0$. Now, we define $g(x_0) = |Y^{2^+}| + i \cdot |V(G)|$ and $g(x) = g(y)$ if $x \in X^{(2)}_{T_i} \backslash \{x_0\}$ is a child of $y \in Y^{2^+}_{T_i}$ in $T_i$. This extension process is almost the same as the one used in Theorem 4.1.1 except only that we give an extra value $|Y^{2^+}| + i \cdot |V(G)|$ to the labels of each improper component $T_i$. Since $x_0$ has a leaf neighbor $y$ in $T_i$ with $y \notin IN(G)$ or $g(y) \neq g(x_0)$, $\{x_0\}$ is indeed a core of $T_i$. The extended labeling $g|_{IN(T_i)}$ from $g|_{Y^{2^+}_{T_i}}$ is a core cluster of $T_i$.

Now, we have a labeling $g : IN(G) \to \mathbb{N}$ obviously satisfying criterion (i) in Lemma 4.1.10. Let us further assume that $g$ satisfies criterion (iii),then $u$ and $u^*$ are in different components in $K - \{uu^* | u \in X^{3^+} \cup Y^\bullet\}$. This implies that $g(u) \neq g(u^*)$ for all $u \in X^{3^+} \cup Y^\bullet$. If $u \in V(K) \cap IN(G) \backslash (X^{3^+} \cup Y^\bullet)$, then $\deg_{H^*}(u) = 1$ or $u \in IN(H^*)$ for some component $H^*$ in $G - E'$. From the construction of $g|_{IN(H^*)}$, we know that each vertex $u \in IN(G)$ with $\deg_{H^*}(u) = 1$ satisfies $g(u) \neq g(w)$ where $w$ is the unique neighbor of $u$ in $H^*$. Also, the proof of Theorem 4.1.1 guarantees that each vertex $u \in IN(H^*)$ has a neighbor $w$ such that $w \notin IN(G)$ or $g(w) \neq g(u)$. Criterion (ii) in Lemma 4.1.10 is satisfied as well. We therefore conclude that $g$ is a core cluster of $G$. Finally, since there are $\mathrm{exc}(G) - |Y^\bullet|$ nontrivial improper components in $G - E'$, $|g(IN(G))| = |Y^{2^+}| - \sum_{v \in X^{3^+}} (\deg_G(v) - 2) + |Y^\bullet| + (\mathrm{exc}(G) - |Y^\bullet|) = \beta(G)$.

(b) For the case where $G$ has some vertices in $X^{3^+}$ which have more than one 1-vertex neighbors, we let $t_v$ be the number of 1-vertex neighbors of $v \in X^{3^+}$ and $X' = \{v \in X^{3^+} | t_v \geq 2\}$. Denote as $G'$ the subgraph obtained by removing $(t_v - 1)$ 1-vertex neighbors of each $v \in X'$ from $G$, then $\mathrm{exc}(G') = \mathrm{exc}(G) - \sum_{v \in X'} (t_v - 1)$ and $\beta(G') = \beta(G)$. The core cluster $g$ of $G'$ obtained from part (a) is also a core cluster of $G$ with $|g(IN(G))| = |g(IN(G'))| = \beta(G') = \beta(G)$. The proof is completed. ∎

Let us call any labeling of $IN(G)$ defined in the way stated in this proof a *candidate labeling* of $G$ in the remainder of this section. If a candidate labeling $g$ of $G$ satisfies criterion (iii) in Lemma 4.1.10, then $g$ is a core cluster of $G$.

In the case where $N_G(u) \cap N_G(v) \cap Y^{3^+} = \emptyset$ for all distinct $u, v \in X^{3^+}$ and $\mathrm{girth}(G) \geq 8$, a candidate labeling obviously satisfies criterion (iii). The following consequence extends Theorem 4.1.4.

**Corollary 4.1.12.** *Let $G = (X, Y)$ with $|X| \geq |Y|$ and* $\text{girth}(G) \geq 8$*. If $N_G(u) \cap N_G(v) \cap Y^{3^+} = \emptyset$ for all distinct $u, v \in X^{3^+}$, then $G$ is realizable and $c^*(G) = \beta(G)$.*

In what follows, we call a cycle *feasible* if it contains two 2-vertices of distance at least four. A feasible cycle is of length at least eight. If every cycle in a graph $G$ is feasible, then $G$ is called feasible as well.

**Theorem 4.1.13.** *Let $G = (X, Y)$ and $|X| \geq |Y|$. If $G$ is feasible, then $G$ is realizable and $c^*(G) = \beta(G)$.*

**Proof.** Consider a candidate labeling $g$ of $G$. It suffices to show that criterion (iii) in Lemma 4.1.10 is made in this situation. We adopt the notations used in the proof of Lemma 4.1.11. Let $w$ be a 2-vertex on a cycle of $G$. Then $w \in V(H)$ for some component $H$ in $G - X^{3^+}$ or $w \in Y^\bullet$. By the construction of $g$, if $w \in V(H)$, then $w$ certainly has a neighbor $w'$ in $G$ with $g(w) \neq g(w')$. If $w \in Y^\bullet$, since each cycle containing $w$ must contain another 2-vertex, $w$ and $w^*$ must be in different components in $K - \{uu^* | u \in X^{3^+} \cup Y^\bullet\}$. We therefore conclude that each 2-vertex $w$ on a cycle has at least one neighbor which has a label different from $w$'s. Next, let $C = (w_0, w_1, \ldots, w_{l-1})$ be a cycle of $G$ in which $\deg_G(w_0) = \deg_G(w_d) = 2$ and $4 \leq d \leq \frac{1}{2}l$, then $g(w_{l-1}) \neq g(w_1)$ and $g(w_{d-1}) \neq g(w_{d+1})$. This implies that this cycle contains at most $l(C) - d$ consecutive edges in the subgraph induced by $g^{-1}(i)$ for all $i \in g(IN(G))$ and the result follows. ∎

An unfeasible cycle can be made feasible by subdividing an edge on it. We have the following observations regarding the effect of subdividing an edge of $G$ on the size of a core cluster and the deduction of a star covering of $G$ which is not necessarily bipartite.

**Proposition 4.1.14.** *Let $G$ be realizable and* $\text{girth}(G) \geq 4$*. If $g$ is an optimal core cluster of $G$, then every cycle of $G$ contains at most $l(C) - 3$ consecutive edges in every subgraph induced by $g^{-1}(i)$, $i \in g(IN(G))$.*

**Proof.** Let $g$ and $\Pi$ be a core cluster and a star covering of $G$, respectively, with $|g(IN(G))| = d_\Pi$. Suppose, on the contrary, that $C = (u_0, u_1, \ldots, u_{k-1})$ is a cycle containing $t$ consecutive edges in the subgraph induced by $g^{-1}(0)$ with $t \geq k-2$. We may assume that $g(u_i) = 0$ for $i = 1, 2, \ldots, k-1$ and $g(u_0) = i_0$, then $i_0 = 0$ if $t = k$ and $i_0 \neq 0$ if $t = k-2$. In the latter case, $u_0$ can not be the designated outside neighbor of $u_1$ or $u_{k-1}$ because $\{u_1, u_{k-1}\} \subseteq g^{-1}(0)$ and we may further assume that $u_1$ is not the designated outside neighbor of $u_0$. Now, we subdivide the edge $u_0u_1$ by replacing it with a path which has consecutive vertices $u_0 = w_0, w_1, \ldots, w_{2l+1} = u_1$, $l \geq 3$, and let the resulting graph be $G'$. We then define a labeling $g'$ on $IN(G')$ as $g'|_{IN(G)} = g$, $g'(w_1) = i_0$, $g'(w_{2l}) = 0$ and $g'(w_{2i}) = g'(w_{2i+1}) = \max(g(IN(G))) + i$ for all $i = 1, 2, \ldots, l-1$. Since in both cases $u_0$ and $u_1$ are not the designated outside neighbors of one another, $g'$ is a core cluster of $G'$ of size $|g'(IN(G'))| = |g(IN(G))| + l - 1$. On the other hand, a star covering of $G'$ can be constructed in a natural way. Let us denote the star with only two edges $w_{i-1}w_i$ and $w_iw_{i+1}$ as $S_i$. Since we may assume that $u_0u_1$ belong to a star $S_{u_0}$ centered at $u_0$ in $\Pi$, $\Pi' = (\Pi \backslash \{S_{u_0}\}) \cup \{(S_{u_0} - u_0u_1) + w_0w_1, S_{w_2}, S_{w_4}, \ldots, S_{w_{2l}}\}$ is a star covering of $G'$ with vertex-number sum $m_{\Pi'} = m_\Pi + 3l$. The deduction of $\Pi'$ will then be $d_{\Pi'} = (|V(G)| + 2l) + (in(G) + 2l) - (m_\Pi + 3l) = d_\Pi + l = |g'(IN(G'))| + 1$ which contradicts to Theorem 3.1.3 and we have the proof. ∎

**Proposition 4.1.15.** *Let $G'$ be a graph obtained by $(2l+1)$-subdividing an edge $e$ of $G$ where $e$ is not pendant and $l \geq 3$. If $G$ is realizable, then $G'$ is realizable.*

**Proof.** Suppose that $G'$ is obtained by replacing the edge $u_0u_1$ with a path which has consecutive vertices $u_0 = w_0, w_1, w_2, \ldots, w_{2l+1} = u_1$. Let $g$ and $\Pi$ be a core cluster and a star covering of $G$, respectively, with $|g(IN(G))| = d_\Pi$. We give $G'$ the same star covering $\Pi'$ defined in the previous proof. Then $d_{\Pi'} = d_\Pi + l$. Now, we need a core cluster $g'$ of $G'$ with $|g'(IN(G'))| = d_\Pi + l$ as well. If $g(u_0) \neq g(u_1)$, then we define $g'$ as $g'|_{IN(G)} = g$ and $g'(w_{2i-1}) =$

$g'(w_{2i}) = \max(g(IN(G))) + i$, for $i = 1, 2, \ldots, l$. $g'$ is clearly a core cluster of $G'$ as desired. For the case where $g(u_0) = g(u_1)$, the subgraph induced by $V_0 = g^{-1}(g(u_0))$ in $G$ is no longer connected after removing the edge $u_0u_1$ from it by Proposition 4.1.14 and this removal results in two components, say $U_0$ and $U_1$. Assume that $u_0 \in V(U_0)$, then $u_1 \in V(U_1)$. Let us define $g'$ as $g'|_{IN(G)\setminus V(U_1)} = g|_{IN(G)\setminus V(U_1)}$, $g'(u) = \max(g(IN(G)))+1$ for all $u \in V(U_1) \cup \{w_{2l}\}$, $g'(w_1) = g(u_0)$ and $g'(w_{2i}) = g'(w_{2i+1}) = \max(g(IN(G))) + i + 1$, for all $i = 1, \ldots, l-1$. One can easily verify that $g'$ is a core cluster of $G'$ with the desired size. ∎

## 4.2 A Bound on the Optimal Average Information Ratio of Bipartite Graphs

Proposition 4.1.15 states that $(2l + 1)$-subdivision ($l \geq 3$) of a nonpendant edge preserves realizability. As for graphs which have not been determined to be realizable or not, suitable 7-subdividing some selected edges can transform them into feasible ones. This suggests a possibility to derive bounds on the optimal average information ratio of them. In the discussion of the following results, we assume that $G'$ is obtained by replacing an edge $u_0u_1$ of $G$ with a path which has consecutive vertices $u_0 = w_0, w_1, \ldots, w_{2l+1} = u_1$.

**Theorem 4.2.1.** *If $G'$ is a graph obtained by $(2l+1)$-subdividing a nonpendant edge of $G$ where $l \geq 3$, then $d^*(G) = d^*(G') - l$.*

**Proof.** In the proof of Proposition 4.1.14, we have given a construction of a star covering $\Pi'$ of $G'$ from an optimal star covering $\Pi$ of $G$ and obtained that $d_{\Pi'} = d_\Pi + l$. Therefore, we have $d^*(G') \geq d^*(G) + l$. On the other hand, if $\Pi'$ is an optimal star covering of $G'$, then a star covering of $G$ can be constructed from $\Pi'$ as follows. First, if none of $w_0$ and $w_{2l+1}$ is the center of any star in $\Pi'$ which has some leaves in $V(G)$, then we let $S$ be the star with a unique edge $u_0u_1$. For the rest case, since the $w_0w_{2l+1}$-path which replaces $u_0u_1$ is of odd length, we may assume that only $w_0$ is the center of a star $S'_{w_0}$ in $\Pi'$

which has leaves in both $V(G)$ and $\{w_i|i = 1, \ldots, 2l\}$, and that $w_{2l+1}$ is not the center of such kind of stars. In this case, we let $S = (S'_{w_0} - \{w_1\}) + u_0u_1$. Now, discarding all stars containing vertices in $\{w_1, w_2, \ldots, w_{2l}\}$ from $\Pi'$ and adding the star $S$ to it, we have a star covering $\Pi$ of $G$ which has vertex-number sum $m_\Pi = m_{\Pi'} - 3l$ where $m_{\Pi'}$ is the vertex-number sum of $\Pi'$ and the deduction $d_\Pi = (|V(G')| - 2l) + (in(G') - 2l) - (m_{\Pi'} - 3l) = d_{\Pi'} - l$. This gives $d^*(G) \geq d^*(G') - l$ and the result follows. ∎

The gap between $c^*(G)$ and $c^*(G')$ depends largely on the edge that is being subdivided. We classify the edges of $G$ as follows. An edge $u_0u_1$ is said to be of *type 1* if either one of the following two conditions is true: (1) $u_0u_1$ does not belong to any cycle in $G$, or (2) it belongs to some cycle $(u_0u_1 \cdots u_l)$ and there is no path in $G$ which connects $u_0$ and some $u_i$, $i \in \{1, 2, \ldots, l\}$, without traversing any edge of this cycle. In case (1), any vertex in $N_G(u_0)\backslash\{u_1\}$ is called a *friendly neighbor* of the edge $u_0u_1$. In case (2), the vertex $u_l$ of $u_0$ is assigned to be the friendly neighbor of $u_0u_1$. An edge not of type 1 is said to be of *type* $r+1$, $r \in \mathbb{N}$, if it is the unique common edge of exactly $r$ cycles and any two of these $r$ cycles have no common vertices other than $u_0$ and $u_1$. In the proof of the next two lemmas, the construction of desired core cluster involves fiddly description. We make use of the following notations and an operation to facilitate the discussion. If $g$ is a core cluster of $G$ and $u \in IN(G)$, then we denote the designated outside neighbor of $u$ as $(u)^*_g$ and let $(\widetilde{V})^*_g = \{(u)^*_g|u \in \widetilde{V}\}$. Besides, if $\widetilde{V}$ is a connected subset of $V(G)$ which induces a connected subgraph $K$ of $G$, and $A_0$ and $A_1$ are disjoint connected subsets of $\widetilde{V}$, then we define a *splitting operation* on $\widetilde{V}$ as follows. Suppose that $\mathbb{U} = \{O_i|i \in I\}$ is the collection of all components in $K - A_0$ and $O_1 \in \mathbb{U}$ is the component containing $A_1$. Let $\widetilde{V}^{[1]} = V(O_1)$ and $\widetilde{V}^{[0]} = \widetilde{V}\backslash\widetilde{V}^{[1]}$, then both $\widetilde{V}^{[0]}$ and $\widetilde{V}^{[1]}$ are connected. By applying the splitting operation to $\widetilde{V}$ w.r.t. $A_0$ and $A_1$, we have two disjoint subsets $\widetilde{V}^{[0]}$ and $\widetilde{V}^{[1]}$ with $A_i \subseteq \widetilde{V}^{[i]}$, $i = 0, 1$, such that $\widetilde{V}^{[0]} \cup \widetilde{V}^{[1]} = \widetilde{V}$. We denote this process as $\mathrm{Split}(\widetilde{V}; A_0, A_1) = (\widetilde{V}^{[0]}, \widetilde{V}^{[1]})$.

Let $g'$ be an optimal core cluster of $G'$. In the proof of Lemma 4.2.2

and 4.2.3, we initially define a labeling $g$ on $IN(G)$ as $g = g'|_{IN(G)}$ and let $(u)^*_g = (u)^*_{g'}$ for all $u \in IN(G)$ when there is no specification. The labeling $g$ may require some modification accordingly in order to reach to a core cluster of $G$. There are many cases to discuss. Let $(g')^{-1}(i) \cap V(G) = V_i$. One situation that worsens our problem the most is when $\{u_0, u_1\} \subseteq (V_a)^*_{g'}$ for some $a \in g'(IN(G'))$ where $u_0u_1$ is the edge been subdivided. This situation is referred to as Situation ($S^*$). In what follows, we assume that $u_0 = (y^i_0)^*_{g'}$ and $u_1 = (y^i_1)^*_{g'}$ where $\{y^i_0, y^i_1\} \subseteq V_{a_i}$ for all $i = 1, 2, \ldots, t$, and $\{u_0, u_1\} \nsubseteq (V_i)^*_{g'}$ for all $i \in g'(IN(G')) \backslash \{a_i | i = 1, \ldots, t\}$. Naturally, $t > 0$ when Situation ($S^*$) occurs and $t = 0$ otherwise. When $t > 0$, we use $V^{[0]}_{a_i}$ and $V^{[1]}_{a_i}$ to denote the resulting subsets from applying the splitting operation to $V_{a_i}$ w.r.t. $\{y^i_0\}$ and $\{y^i_1\}$, i.e. $\mathrm{Split}(V_{a_i}; \{y^i_0\}, \{y^i_1\}) = (V^{[0]}_{a_i}, V^{[1]}_{a_i})$, for all $i = 1, \ldots, t$. Moreover, the numbers $c_0, c_1, \ldots, c_t$, $d_0$ and $d_1$ that will be used in the proof always represent distinct integers in $\mathbb{N} \backslash g'(IN(G'))$. With the aid of these notations, we can present our construction of core clusters of $G$ in a more systematic way.

**Lemma 4.2.2.** *Let $G'$ be a graph obtained by $(2l+1)$-subdividing a nonpendant edge $u_0u_1$ of a simple graph $G$ with $\mathrm{girth}(G) \geq 6$, where $l \geq 3$. If $g'$ is an optimal core sequence of $G'$ and $g'(u_0) = g'(u_1)$, then $c^*(G) \leq c^*(G') - l + r$ provided that $u_0u_1$ is an edge of type $r$.*

**Proof.** If $\{(u_0)^*_{g'}, (u_1)^*_{g'}\} \subseteq V(G)$, then $|\{g'(w_i)|i = 1, \ldots, 2l\} \backslash \{g'(u_0)\}| \geq l - 1$ and the labeling $g = g'|_{IN(G)}$ is a core cluster of $G$ with $|g(IN(G))| \leq |g'(IN(G'))| - (l - 1)$. Now, we assume that $g'(u_0) = g'(u_1) = 0$ and $\{(u_0)^*_{g'}, (u_1)^*_{g'}\} \nsubseteq V(G)$, then $|\{g'(w_i)|i = 1, \ldots, 2l\} \backslash \{0\}| \geq l$ and $g$ may no longer be qualified as a core cluster of $G$. We shall make some local modifications of $g$ and assign $(u_0)^*_g = u_1$ and $(u_1)^*_g = u_0$ to reach our goal. Set $A_0 = \{u_0\} \cup ((N_G(u_0) \backslash \{u_1\}) \cap V_0)$ and $A_1 = \{u_1\} \cup ((N_G(u_1) \backslash \{u_0\}) \cap V_0)$. Since $u_0$ and $u_1$ have no common neighbors, $A_0$ and $A_1$ are disjoint connected subsets of the connected set $V_0$. Applying the splitting operation $\mathrm{Split}(V_0; A_0, A_1) = (V^{[0]}_0, V^{[1]}_0)$, we have two disjoint connected subsets $V^{[0]}_0$ and $V^{[1]}_0$ with $V^{[0]}_0 \cup V^{[1]}_0 = V_0$.

(1) Suppose first that $t = 0$, that is, Situation $(S^*)$ does not occur. By redefining $g(V_0^{[0]}) = \{c_0\}$, we claim that the resulting labeling $g$ is a core cluster of $G$. Note that now $g(u_0) = c_0 \neq g(u_1)$, and $u_0$ is adjacent to $u_1 \in V_0^{[1]}$ and no other vertices in $V_0^{[1]}$. Besides, $\{u_0\} \cup (V_0^{[1]})^*_{g'}$ is independent because $(g')^{-1}(0)$ is a core in $G'$ containing $\{u_0\} \cup V_0^{[1]}$ and each $(w)^*_{g'} \in (V_0^{[1]})^*_{g'}$ is adjacent to the unique vertex $w$ in $(g')^{-1}(0)$. Hence, $(u_1)^*_g = u_0$ and $(w)^*_g = (w)^*_{g'}$, for all $w \in V_0^{[1]} \backslash \{u_1\}$, are qualified designated outside neighbors of vertices in $V_0^{[1]}$ and then $V_0^{[1]} = g^{-1}(0)$ is a core of $G$. The fact $g^{-1}(c_0) = V_0^{[0]}$ is also a core of $G$ can be shown by similar reasoning. We then conclude that $g$ is a core cluster of $G$ and $|g(IN(G))| \leq |g'(IN(G'))| - l + 1$.

(2) Suppose that $t > 0$, then $r \geq t + 1$. Besides making $g(V_0^{[0]}) = \{c_0\}$, we further redefine $g(V_{a_i}^{[0]}) = \{c_i\}$ for all $i = 1, \ldots, t$. Since $g(y_0^i) = c_i \neq g(y_1^i) = a_i$, $V_{a_i}^{[0]}$ and $V_{a_i}^{[1]}$ are cores of $G$. $g$ is then a core cluster of $G$ with $|g(IN(G))| \leq |g'(IN(G'))| - l + (t + 1)$. ∎

**Lemma 4.2.3.** *Let $G'$ be a graph obtained by $(2l+1)$-subdividing a nonpendant edge $u_0u_1$ of a simple graph $G$ with* $\mathrm{girth}(G) \geq 6$*, where $l \geq 3$. If $g'$ is an optimal core cluster of $G'$ and $g'(u_0) \neq g'(u_1)$, then $c^*(G) \leq c^*(G') - l + r$ provided that $u_0u_1$ is an edge of type $r$.*

**Proof.** We split the discussion into two cases.

**Case 1.** Assume that $g'(u_0) = 0 \neq g'(u_1) = 1$ and $\{(u_0)^*_{g'}, (u_1)^*_{g'}\} \subseteq V(G)$, then $|\{g'(w_i) | i = 1, \ldots, 2l\} \backslash \{0, 1\}| \geq l - 1$ and $g = g'|_{IN(G)}$ is not a core cluster of $G$ only when any of the following three situations occurs. Situation $(S1)$ : $u_1 = (x_1)^*_{g'}$ for some $x_1 \in V_0$; Situation $(S2)$ : $u_0 = (x_0)^*_{g'}$ for some $x_0 \in V_1$; and the stated Situation $(S^*)$. We shall fix the problem by shifting some vertices between $V_0$ and $V_1$ or adding some extra values to $g(IN(G))$ as follows.

**Subcase 1-1.** Suppose that both Situation $(S1)$ and $(S2)$ do not occur, then $t > 0$. If $r = t = 1$, let us assume that $y_0^1$ is the friendly neighbor of $u_0u_1$. We redefine $g(V_{a_1}^{[0]}) = \{0\}$ and then assign $(u_0)^*_g = u_1$ and choose a neighbor of $y_0^1$ in $V_{a_1}^{[1]}$ to be $(y_0^1)_g^*$. Since $u_0u_1$ is of type 1, each vertex in $V_{a_1}^{[0]}$ is not adjacent to any vertex in $V_0\backslash\{u_0\}$ and $\{(y_0^1)_g^*\} \cup (V_{a_1}^{[0]}\backslash\{y_0^1\})_{g'}^* \cup (V_0)_{g'}^*$ is independent. This guarantees that $g^{-1}(0) = V_0 \cup V_{a_1}^{[0]}$ is a core of $G$. Besides, $g(y_0^1) \neq g(y_1^1)$ implies that $V_{a_1}^{[1]}$ is also a core. Hence, $g$ is a core cluster of $G$ with $|g(IN(G))| \leq |g'(IN(G'))| - (l-1) = c^*(G') - l + r$. If $r > 1$, then $r \geq t+1$. By redefining $g(V_{a_i}^{[0]}) = \{c_i\}$ for all $i = 1, \ldots, t$, and letting $(u)_g^* = (u)_{g'}^*$ for all $u \in IN(G)$, we have a core cluster $g$ of $G$ with $|g(IN(G))| \leq |g'(IN(G'))| - (l-1) + t \leq c^*(G') - l + r$.

**Subcase 1-2.** Suppose that Situation $(S1)$ occurs and $(S2)$ does not, then either $t = 0$ and $r \geq 1$ or $t > 0$ and $r \geq t+2$. Let $\mathrm{Split}(V_0; \{u_0\}, \{x_1\}) = (V_0^{[0]}, V_0^{[1]})$. When $r \in \{1, 2\}$ $(t = 0)$, we redefine $g(V_0^{[0]}) = \{1\}$. One can easily verify that $g^{-1}(1) = V_0^{[0]} \cup V_1$ is a core of $G$ and therefore $g$ is a core cluster of $G$ with $|g(IN(G))| \leq |g'(IN(G'))| - (l-1)$. When $r \geq 3$, redefining $g(V_0^{[0]}) = \{c_0\}$ is sufficient if $t = 0$. After assigning $u_1 = (x_1)_g^*$, $g$ is a core cluster of $G$ with $|g(IN(G))| \leq |g'(IN(G'))| - (l-1) + 1 \leq c^*(G') - l + 2$. If $t > 0$, we further redefine $g(V_{a_i}^{[0]}) = \{c_i\}$ for all $i = 1, \ldots, t$. The resulting labeling $g$ is a core cluster of $G$ with $|g(IN(G))| \leq |g'(IN(G'))| - (l-1) + t + 1 \leq c^*(G') - l + r$.

**Subcase 1-3.** Suppose that Situation $(S1)$ and $(S2)$ occur simultaneously, then $r \geq t+3$. When $t = 0$, we redefine $g(V_0^{[0]} \cup V_1^{[0]}) = \{d_0\}$ if $r = 3$, and redefine $g(V_0^{[0]}) = \{d_0\}$ and $g(V_1^{[0]}) = \{d_1\}$ if $r \geq 4$. In both cases, $g$ is a core cluster of $G$ with $|g(IN(G))| \leq c^*(G') - l + 3$. When $t > 0$, besides making $g(V_0^{[0]}) = \{d_0\}$ and $g(V_1^{[0]}) = \{d_1\}$, we further redefine $g(V_{a_i}^{[0]}) = \{c_i\}$ for all $i = 1, \ldots, t$. This results in a core cluster $g$ of $G$ that meets our requirement where $|g(IN(G))| \leq c^*(G) - l + r$.

**Case 2.** Assume that $g(u_0) = 0 \neq g(u_1) = 1$ and $\{(u_0)_{g'}^*, (u_1)_{g'}^*\} \not\subseteq V(G)$, then $|\{g'(w_i)|i = 1, \ldots, 2l\}\backslash\{0, 1\}| \geq l$. Note that if we assign $(u_0)_g^* = u_1$ and $(u_1)_g^* = u_0$, then the labeling $g = g'|_{IN(G)}$ will not be a core cluster of $G$ only when any of the following three situations occurs. Situation $(T1)$ : $N_G(u_1) \cap V_0 \neq \emptyset$ or $N_G(u_1) \cap (V_0)_{g'}^* \neq \emptyset$; Situation $(T2)$ : $N_G(u_0) \cap V_1 \neq \emptyset$ or $N_G(u_0) \cap (V_1)_{g'}^* \neq \emptyset$; and the Situation $(S^*)$.

**Subcase 2-1.** Suppose that both Situation $(T1)$ and $(T2)$ do not occur and $t > 0$, then either $r = t = 1$ or $r > 1$ and $r \geq t+1$. We redefine $g(V_{a_i}^{[0]}) = c_i$, for all $i = 1, \ldots, t$, and assign $(u_0)_g^* = u_1$ and $(u_1)_g^* = u_0$. The resulting labeling $g$ is obviously a core cluster with $|g(IN(G))| \leq |g'(IN(G'))| - l + t$.
**Subcase 2-2.** Suppose that Situation $(T1)$ occurs and $(T2)$ does not, then either $t = 0$ and $r \geq 1$ or $t > 0$ and $r \geq t + 2$. Now, let $x_1$ be a vertex in $N_G(u_1) \cap V_0$ if $N_G(u_1) \cap V_0 \neq \emptyset$, and $x_1$ be a vertex in $V_0$ such that $(x_1)_{g'}^* \in N_G(u_1)$ otherwise. Choose a vertex $z_0 \in N_G(u_0)$ which is on a $u_0x_1$-path whose vertices are in $V_0$, and then consider $\text{Split}(V_0; \{u_0\}, \{z_0\}) = (V_0^{[0]}, V_0^{[1]})$. After redefining $g(V_0^{[0]}) = \{c_0\}$ and assigning $(u_0)_g^* = z_0$ and $(u_1)_g^* = u_0$, one can easily verify that $V_0^{[0]} = g^{-1}(c_0)$ is a core. If $t > 0$, we further redefine $g(V_{a_i}^{[0]}) = \{c_i\}$ for all $i = 1, 2, \ldots, t$. Then the labeling $g$ is a core cluster of $G$ with $|g(IN(G))| \leq |g'(IN(G'))| - l + t + 1$.
**Subcase 2-3.** Suppose that both Situation $(T1)$ and $(T2)$ occur, then $r \geq t + 3$. Using the manner we chose $z_0$ in the previous subcase, we select $z_1 \in N_G(u_1)$ such that $z_1$ is on a path with vertices in $V_1$ connecting $u_1$ to a vertex $x_0$ where $x_0 \in N_G(u_0) \cap V_1$ if $N_G(u_0) \cap V_1 \neq \emptyset$, and $x_0 \in V_1$ such that $(x_0)_{g'}^* \in N_G(u_0)$ if $N_G(u_0) \cap V_1 = \emptyset$. Consider $\text{Split}(V_0; \{u_0\}, \{z_0\}) = (V_0^{[0]}, V_0^{[1]})$ and $\text{Split}(V_1; \{u_1\}, \{z_1\}) = (V_1^{[0]}, V_1^{[1]})$. By redefining $g(V_0^{[0]}) = \{d_0\}$ and $g(V_1^{[0]}) = \{d_1\}$ and assigning $(u_i)_g^* = z_i$, $i = 0, 1$, $g^{-1}(d_0) = V_0^{[0]}$ and $g^{-1}(d_1) = V_1^{[0]}$ are both cores of $G$. If $t > 0$, we further redefine $g(V_{a_i}^{[0]}) = \{c_i\}$ for all $i = 1, \ldots, t$. Then the core cluster $g$ of $G$ has $|g(IN(G))| \leq |g'(IN(G'))| - l + t + 2$. ∎

Theorem 4.2.1, Lemma 4.2.2 and Lemma 4.2.3 jointly show the following lemma.

**Lemma 4.2.4.** *Let $G'$ be a graph obtained by $(2l + 1)$-subdividing a non-pendant edge $e$ of a simple graph $G$ with* $\text{girth}(G) \geq 6$*, where $l \geq 3$. If $c^*(G') - d^*(G') = k$, then $c^*(G) - d^*(G) \leq k + r$ provided that $e$ is an edge of type $r$.*

This lemma gives rise to a bound on $AR(G)$. Let $E'$ be a set of edges of $G$. If 7-subdividing each edge in $E'$ results in a feasible graph, then

$E'$ is called a *feasiblizer* of $G$. The minimum cardinality of all feasiblizers of $G$ is denoted as $\phi(G)$, called the *feasiblizing number* of $G$. Let $\Delta(G)$ be the maximum degree of $G$. If an edge $u_0u_1$ of $G$ is of type $r$, then $r \le \min\{\deg_G(u_0), \deg_G(u_1)\} \le \Delta(G)$.

**Theorem 4.2.5.** *Let $G = (X, Y)$ with $|X| \ge |Y|$ and $\operatorname{girth}(G) \ge 8$. If $E'$ is a feasiblizer of $G$ in which there are $\alpha_r$ type-$r$ edges and $\alpha = \sum_{r=1}^{\Delta(G)} r\alpha_r$, then $c^*(G) - d^*(G) \le \alpha$ and*

$$\frac{|V(G)| + in(G) - (\beta(G) + \alpha)}{|V(G)|} \le AR(G) \le \frac{|V(G)| + in(G) - \beta(G)}{|V(G)|}.$$

The feasiblizing number is analogous to the decycling number of $G$. One major difference lies in that we only deal with unfeasible cycles instead of all cycles in $G$. More importantly, we choose edges as opposed to vertices to destroy unfeasible cycles. This gives a lot more freedom on the choices of edges in a feasiblizer. It should be clarified that choosing common edges of cycles does not necessarily lessen the number of edges needed to feasiblize a graph. For instance, let $G$ be a 16-cycle $(w_0w_1 \cdots w_{15})$ with a chord $w_0w_7$, then $\phi(G) = 2$ and both edges in a minimum feasiblizer can be chosen to be of type 1. Choosing the common edge $w_0w_7$ of two cycles does not result in a feasiblizer with lesser edges. For a graph which has a feasiblizer consisting of type-1 edges, the bound of Theorem 4.2.5 can be very good.

**Corollary 4.2.6.** *Let $G = (X, Y)$ with $|X| \ge |Y|$ and $\operatorname{girth}(G) \ge 8$. If $E'$ is a feasiblizer consisting of type-1 edges with $|E'| = \phi(G)$, then $c^*(G) - d^*(G) \le \phi(G)$ and*

$$\frac{|V(G)| + in(G) - (\beta(G) + \phi(G))}{|V(G)|} \le AR(G) \le \frac{|V(G)| + in(G) - \beta(G)}{|V(G)|}.$$

This bound is best possible using our $c^*(G)$-and-$d^*(G)$ approach. We show this fact by proposing an infinite class of graphs attaining this bound. Consider the class of connected graphs with the pattern given in Figure 4.1. The one with $k$ cycles is denoted as $G(k)$. For each $k \in \mathbb{N}$, $\phi(G(k)) = k$ is obviously true. By direct calculation, one can verify that the labeling giving all vertices of the $i$-th cycle the label $i$, for all $i = 1, \ldots, k$, is an optimal core cluster, hence $c^*(G(k)) = k$. On the other hand, the covering given in Theorem 4.1.8 is an optimal star covering of $G(k)$ and then $d^*(G(k)) = 0$. Therefore, the bound $c^*(G) - d^*(G) \leq \phi(G)$ is attained by each $G(k)$. For the classes of bipartite graphs described in this corollary, our bound on $AR(G)$ is not only the best possible using our approach but also the best bound so far.

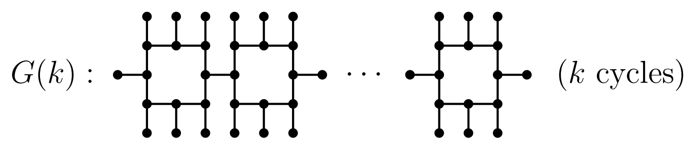



Figure 4.1: The family $G(k)$ of bipartite graphs

## 4.3 Concluding Remark

In this chapter, we have investigated the equality $c^*(G) = d^*(G)$ and have shown that it holds for any even-subdivdion of a simple graph and certain classes of bipartite graphs of larger girth. The exact values of the optimal average information ratio for those graphs can then be determined.

For bipartite graphs which have not been determined to be realizable or not, we have derived a bound on $c^*(G) - d^*(G)$, which naturally gives rise to a bound on the optimal average information ratio for them. We have also shown that our bound is the best possible using our approach for some infinite classes of graphs. To determine the exact values of the optimal average information ratio for them, new technique must be imposed.

Theorems 4.1.1 and 4.1.4 and Corollaries 4.1.2 and 4.1.3 have been presented in the 33rd International Conference on Mathematical, Computational and Statistical Sciences, and Engineering (ICMCSSE2012).

# Chapter 5

# Conclusion

## 5.1 Our Contribution

Evaluating the optimal information ratio and the optimal avergage information ratio is an important and challenging issue in secret-sharing. In this thesis, we devote our efforts to the study of the optimal average information ratio of interesting access structures.

In weighted threshold access structures, each participant has his or her own weight depending on the importance of the participant in an organization. A participant(vertex) with higher weight naturally induce more edges incident to it in the $k$-weighted graph. This makes the weighted threshold access structures more applicable in real-life situation. An in-depth investigation can have a significant contribution to the application of secret-sharing. We have examined the structure of $k$-weighted graphs and presented two constructions of secret-sharing schemes for them. Both of our constructions have low avergage information ratios and, as $k$ fixed, both ratios approach the optimal value 1 asymptotically. A comparison shows that Construction I has lower avergage information ratio when $k$ is smaller, while Constructin II gains its superiority over Construction I for larger $k$. Dealing with the average information ratio is in general very tedious. In the work of Chapter 2, we have demonstrated an approach to extracting valuable results from complicated expressions.

In Chapter 3, we propose our new approach to the determination of the exact values of the optimal average information ratio of graphs. We define the core number $c^*(G)$ and the deduction $d^*(G)$ of a graph $G$, and show that when $c^*(G) = d^*(G)$, the exact value of $AR(G)$ can be determined. This idea also formulates a complicated problem in secret-sharing into an elegant max-min problem in Graph Theory with easy description. Using our approach, we successfully determine the exact values of the optimal avergage information ratio of all trees. Along with the result by Csirmaz and Tardos [17], we complete the work of evaluating the optimal information ratio and the optimal average information ratio of all trees. In addition, our approach can also be used to recursively evaluate the core number of trees with symmetric structures. This gives a systematic way to evaluate the optimal average information ratio of them.

We then make an attempt on the average information ratio of bipartite graphs in Chapter 4. We determine the exact values of $AR(G)$ for any even-subdividion of a simple graph and some classes of bipartite graphs. It is worth noting that the value of $AR(G)$ also serves as a lower bound on the unknown optimal information ratio of those graphs. Deriving lower bounds on the optimal (average) information ratio is in general much more difficult than deriving upper bounds for any graph. Appendantly, by solving the problem of $AR(G)$, we also obtain valuable results in graph decomposition problem. We have shown that the star covering (decomposition) we constructed has the minimum vertex-number sum among all star coverings (deocmpositions) of those realizable graphs. Although we did not make an effort to show that the coverings (decompositions) given in Theorem 4.1.8 are optimal star coverings for all bipartite graphs, we conjecture that this is true.

## 5.2 Future Work

Continuing our work in this thesis, we shall explore more classes of graphs which satisfy the identity $c^*(G) = d^*(G)$. We shall also try to characterize non-realizable graphs, namely, the decuction of the graphs can never match the core number of them. By estimating the gap between the decuction and the core number of a non-realizalbe graph, one can obtain a bound on the optimal average information ratio of that graph. To find out the exact values of the optimal average information ratio of non-realizable graphs, new approach must be developed.

In our work, the deduction of $G$ is defined for a star covering of $G$. Since a star covering generally does not serve as a complete multipartite covering with the least vertex-number sum for graphs of smaller girth, our approach only works well for graphs of larger girth. However, the idea of the deduction of a star covering can be generalized. It can be defined for a complete multipartite covering in the same way. Then, the deduction of a complete multipartite covering matching the size of a core cluster still makes a criterion for examinimg whether the exact values of the optimal average information ratio of a graph can be determined. In this case, the complete multipartite covering may contain various kinds of complete multipartite subgraphs. The question of how many copies of each complete multipartite subgraph should we use in the covering in order to reach to the maximum deduction may again lead to a linear programming problem.

Under this new setting, the problem of identifying a proper complete multipartite covering with the maximum deduction which matches the core number of that graph, or estimating the gap between the maximum deduction and the core number is again worth trying. Apart from these questions, we may try to characterize the graphs of which the deduction of complete multipartite coverings can never match the core number, and develop a new strategy to determine the exact values of the optimal average information ratio of this kind of graphs. Although they may be quite challenging, these questions certainly are intriguing generalizations of our work in this thesis.

# References


[1] A. Beimel, Secret-sharing schemes: A survey, in: *Proceedings of the 3rd International Workshop on Coding and Cryptography Conference, LNCS*, **6639** (2011), 11–46.

[2] M. Ben-Or, S. Goldwasser, and A. Wigderson, Completeness for non-cryptographic fault-tolerant distributed computations, in: *Proceedings of the 20th STOC* (1988), 1–10.

[3] G. R. Blakley, Safeguarding cryptographic keys, in *American Federation of Information Processing Societies Proceedings*, **48** (1979), 313–317.

[4] C. Blundo, A. De Santis, R. De Simone and U. Vaccaro, Tight bounds on the information rate of secret sharing schemes, *Designs, Codes and Cryptography*, **11** (1997), 107–122.

[5] C. Blundo, A. De Santis, L. Gargano and U. Vaccaro, On the information rate of secret sharing schemes, *Theoretical Computer Science*, **154** (1996), 283–306.

[6] C. Blundo, A. De Santis, A. Giorgio Gaggian and U. Vaccaro, New bounds on the information rate of secret sharing schemes, *IEEE Transactions on Information Theory*, **41** (1995), 549–554.

[7] C. Blundo, A. De Santis, D. R. Stinson and U. Vaccaro, Graph decompositions and secret sharing schemes, *J. Cryptology*, **8** (1995), 39–64.

[8] E. F. Brickell and D. M. Davenport, On the classification of ideal secret sharing schemes, *J. Cryptology*, **4** (1991), 123–134.

[9] E. F. Brickell and D. R. Stinson, Some improved bounds on the information rate of perfect secret sharing schemes, *J. Cryptology*, **5** (1992), 153–166.

[10] R. M. Capocell, A. De Stantis, L. Gargano and U. Vaccaro, On the size of shares for secret sharing schemes, *J. Cryptology*, **6** (1993), 157–167.

[11] D. Chaum, C. Crépeau, and I. Damgard, Multiparty unconditionally secure protocols, in: *Proceedings of the 20th STOC* (1988), 11–19.

[12] R. Cramer, I. Damgard, and U. Maurer, General secure multi-party computation from any linear secret-sharing scheme, in *EUROCRYPT 2000, Lecture Notes in Computer Science*, **1807** (2000), 316–334.

[13] L. Csirmaz, The size of a share must be large, *J. Cryptology*, **10** (1997), 223-231.

[14] L. Csirmaz, Secret sharing schemes on graphs, *Studia Mathematica Hungarica*, **10** (1997), 297–306.

[15] L. Csirmaz, An impossibility result on graph secret sharing, *Designs, Codes and Cryptography*, **53** (2009), 195–209.

[16] L. Csirmaz and P. Ligeti, On an infinite families of graphs with information ratio $2 - \frac{1}{k}$, *Computing*, **85** (2009), 127–136.

[17] L. Csirmaz and G. Tardas, Exact bounds on tree based secret sharing schemes, *Tatracrypt* 2007, Slovakia.

[18] I. Csiszár and J. Körner, *Information Theory. Coding Theorems for Discrete Memoryless Systems*, Academic Press, New York, 1981.

[19] M. H. Dehkordi and S. Mashhadi, New efficient and practical multi-secret sharing schemes, *Information Sciences*, **178** (2008), 2262–2274.

[20] L. R. Jr. Ford and D. R. Fulkerson, “Maximal flow through a network”, *Canad. J. Math.*, vol.8, pp.399–404, 1956.

[21] V. Goyal, O. Pandey, A. Sahai, and B. Waters, Attribute based encryption for fine-grained access control of encrypted data, in: *CCS 2006* (2006).

[22] L. Harn and C. Lin, Strong $(n, t, n)$ verifiable secret sharing scheme, *Information Sciences*, **180** (2010), 3059–3064.

[23] W.-A. Jackson and K. M. Martin, Perfect secret sharing schemes on five participants, *Designs Codes and Cryptography*, **9** (1996), 267–286.

[24] K. Kaya and A. A. Selcuk, Threshold cryptography based on Asmuth-Bloom secret sharing, *Information Sciences*, **177** (2007), 4148–4160.

[25] J. Martí-Farré and C. Padró, Secret sharing schemes with three or four minimal qualified subsets, *Designs, Codes and Cryptography*, **34** (2005), 17–34.

[26] J. Martí-Farré and C. Padró, Secret sharing schemes on access structures with intersection number equal to one, *Discrete Applied Mathematics*, **154** (2006), 552–563.

[27] J. Martí-Farré and C. Padró, On secret sharing schemes, matroids and polymatroids, *J. Math. Cryptol.*, **4** (2010), 95–120.

[28] P. Morillo, C. Padro, G. Saez and J. L. Villar, Weighted threshold secret sharing schemes, *Information Processing Letters*, **704** (1999), 211–216.

[29] M. Naor and A. Wool, Access control and signatures via quorum secret sharing, *IEEE Trans. Parallel and Distributed Systems*, **9** (1998), 909–922.

[30] C. Padró and G. Sáez, Secret sharing schemes with bipartite access structure, *IEEE Transactions on Information Theory*, **46(7)** (2000), 2596–2604.

[31] A. Shamir, How to share a secret, *Communications of the ACM*, **22** (1979), 612–613.

[32] D. R. Stinson, An explication of secret sharing schemes, *Designs Codes and Cryptography*, **2** (1992), 357–390.

[33] D. R. Stinson, New general lower bounds on the information rate of perfect secret sharing schemes, in *Advances in Cryptology – CRYPTO '92, Lecture Notes in Computer Science*, **740** (1993), 168–182.

[34] D. R. Stinson, Decomposition constructions for secret sharing schemes, *IEEE Transactions on Information Theory*, **40** (1994), 118–125.

[35] M. van Dijk, On the information rate of perfect secret sharing schemes, *Designs, Codes and Cryptography*, **6** (1995), 143–169.

[36] D. B. West, *Introduction to Graph Theory*, Prentice Hall, 2001.