Winter 2019

Modeling Stochastically Intransitive Relationships in Paired Comparison Data

Ryan Patrick Alexander McShane
Southern Methodist University, rmcshane@smu.edu

MODELING STOCHASTICALLY INTRANSITIVE RELATIONSHIPS

IN PAIRED COMPARISON DATA

Approved by:

_____

Dr. Ian Richard Harris
Southern Methodist University
Department of Statistical Science

_____

Dr. Daniel Francis Heitjan
Southern Methodist University
Department of Statistical Science

_____

Dr. Eric Cooper Larson
Southern Methodist University
Department of Computer Science

_____

Dr. Cornelis Jacobus Potgieter
Texas Christian University
Department of Mathematics

_____

Dr. Sara Lynne Stokes
Southern Methodist University
Department of Statistical Science

_____

Dr. Paul Xavier Uhlig
St. Mary's University
Department of Mathematics

MODELING STOCHASTICALLY INTRANSITIVE RELATIONSHIPS IN PAIRED COMPARISON DATA

A Dissertation Presented to the Graduate Faculty of the

Dedman College

Southern Methodist University

in

Partial Fulfillment of the Requirements

for the degree of

Doctor of Philosophy

with a

Major in Statistical Science

by

Ryan Patrick Alexander McShane

B.S., Mathematics with Computer Science minor, St. Mary's University
M.S., Applied Statistics, The University of Texas at San Antonio
M.S., Statistical Science, Southern Methodist University

December 21, 2019

ACKNOWLEDGEMENTS

First, I must thank my advisor, Dr. Harris, who has worked with me in nearly every capacity over the past five years. At this point, I can recite Dr. Harris's STAT2331 lectures in my sleep! In fact, I have used whole chunks of his lectures in my first semester at Amherst College. I appreciate the way he exposits his course material in a connected way. More importantly, I must thank him for being open to the topic I picked for this thesis and for his guidance and mathematical insight. I look forward to continuing this work together, as we have opened many paths of inquiry.

Thank you to SMU's Department of Statistical Science for accepting me into their program, providing an assistantship, and providing the time and space I needed to realize a life-long goal. Thank you to my committee members (Drs. Harris, Stokes, Potgieter, Larson, Uhlig, and Heitjan) for reading my dissertation proposal, reading this thesis, and submitting dozens of recommendation letters. I must absolutely thank Dr. Heitjan for his flexibility and assistance in my time of need.

I thank Jackie Horras for hiring me to my first professional teaching position at Health Careers High School, Raul Ortega for hiring me to my first industry job at USAA, and Amy Wagaman for hiring me at my first position in academia at Amherst College — all of these experiences have formed me into the person I am today and will be in the future.

Thank you to Dr. Uhlig, who has played an important role in the last 15 years of my life — at times as a polymath role model, mathematics professor, academic advisor, supervisor, actuarial exam study buddy, and father figure. And thank you to Dr. Michael Louis Lecocke and Drs. Holder, Berndt, Weber, and Hall for providing a steadying hand in my shaky undergraduate years.

Thank you to Morgan Lewis and Jennifer Garcia who have edited my personal, teaching, and research statements and parts of my dissertation. A special thank you to Amy Nussbaum, who helped me with R's ggplot and suggested numerous edits to the dissertation.

The initial inspiration to do this work came from reading weekly NBA power rankings over the last 11 years, primarily penned or programmed by John Hollinger, Marc Stein, Nate Silver, and Neil Paine. The teams, individuals, or organizations I have been proudest to support over my lifetime include the Steve Spurrier-era Gators, the Elliott Smith-era Cowboys, the David Robinson as well as Tim Duncan-era Spurs, Jan Ulrich, and both Team USA and Team Germany at the Olympics and Soccer World Cups. Thank you to Ravensburger, Deutsche Grammophon, The Dillinger Escape Plan, Kyle Kloster and the DFW Super Smash Brothers N64 community (Matt Hamzeh, Dillon Grandy, Mateen Emad, and Sam Preminger), and Pandemic for recharging my batteries in my down time.

Thank you most of all to my mom and sister, Claudia and Janine, whom I love dearly and have supported me in all my life decisions. I thank them both for the sacrifices they have made to help me have a better life.

McShane, Ryan Patrick Alexander

B.S., Mathematics with Computer Science minor, St. Mary's University
M.S., Applied Statistics, The University of Texas at San Antonio
M.S., Statistical Science, Southern Methodist University

Modeling Stochastically Intransitive Relationships
in Paired Comparison Data

Advisor: Dr. Ian Richard Harris

Doctor of Philosophy conferred December 21, 2019

Dissertation completed August 23, 2019

If the Warriors beat the Rockets and the Rockets beat the Spurs, does that mean that the Warriors are better than the Spurs? Sophisticated fans would argue that the Warriors are better by the transitive property, but could Spurs fans make a legitimate argument that their team is better despite this chain of evidence?

We first explore the nature of intransitive (rock-scissors-paper) relationships with a graph theoretic approach to the method of paired comparisons framework popularized by Kendall and Smith (1940). Then, we focus on the setting where all pairs of items, teams, players, or objects have been compared to one another twice (i.e., home and away). We propose a novel linear model (CRSP) whose latent bilinear fixed effect allows us to estimate deviations from our transitive model (C).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF IMPORTANT ABBREVIATIONS

α-vector	Row sums of binary tournament adjacency matrix a la Kendall and Smith's (1940); score vector; alpha vector; victory counts
AHP	Analytical Hierarchy Process
BTM	Bradley-Terry Model
C	Chain [model]
CRSP	Chain Rock-Scissors-Paper [model]
d-distribution	Kendall and Smith's (1940) circular triad frequency distribution
MDS	Multi-Dimensional Scaling
NBA	National Basketball Association
NCAA	National Collegiate Athletic Association
OEIS	Online Encyclopedia of Integer Sequences
RPI	Ratings Percentage Index
SSB64	Super Smash Brothers 64
SST	Strong Stochastic Transitivity [condition]
TMM	Thurstone-Mosteller Model
WST	Weak Stochastic Transitivity [condition]

This is dedicated to Tim Duncan and the 2012-2013 San Antonio Spurs,

who taught us how to handle finishing in second place

by accruing wins,

rising from the ashes like a phoenix,

and winning the 2013-2014 NBA Championship.

CHAPTER 1:

THE NATURE OF INTRANSITIVITY

If The University of Texas (UT) beats The Ohio State University (OSU) and OSU beats Texas A&M University (A&M), does that mean that UT is better than A&M? Sophisticated Longhorn fans would argue UT is better than A&M by the transitive property, but could Aggie fans make a legitimate argument that their team is better despite this chain of evidence? We attempt to answer these questions and more!

Our initial focus is on the method of paired comparisons, used in settings where all $\binom{n}{2}$ pairs of n items, teams, players, or objects have been compared to one another by one judge. We will not be concerned with the aggregation of preferences (or ranks) from multiple judges or sources; however, extensions may be considered later.

Applications of pairwise comparisons reach far beyond sports. Examples include ranking political candidates (Borda, 1781), experimentally predicting animal behavior (Kendall & Smith, 1940), ranking children's handwriting quality (Thurstone, 1927), measuring utility (May, 1954), and exploring dominance relations within or among animal species (Rapoport, 1949). The earliest known application of paired comparisons was documented in 1283 by Llull, where he described a method for selecting the pope (1283). The most accessible examples and data, however, typically involve sporting competitions.

We first explore the nature of intransitive relationships with a graph theoretic approach to the method of paired comparisons framework popularized by M.G. Kendall and B.B. Smith (1940). Then, models which measure relationships between objects that have been compared pairwise, predominately with the objective of creating ratings and subsequently rankings, are discussed. Next, we propose a model known as a Chain Rock-Paper-Scissors, or CRSP, whose latent bilinear fixed effect allows us to estimate deviations from a transitive model. Finally, a method to extend Kendall and Smith's d-distribution by parallelized computing is proposed.

The rest of this chapter introduces concepts fundamental to the remaining chapters, including the transitive property of inequalities, graph theory, circular triads, and adjacency matrices.

1.1. Transitive Property Introduction

Let a, b, and c be real numbers. The transitive property of equality states that if $a = b$ and $b = c$, then $a = c$. The transitive property of inequality states that if $a > b$ and $b > c$, then $a > c$ or alternatively that if $a < b$ and $b < c$, then $a < c$. We will only be concerned with the transitive property of inequality. For abstract real numbers like a, b, and c, the transitive property is undeniable. Additionally, when measuring the length of physical objects — placing them on a measurement scale — it is also clear that the transitive property should hold upon comparison. What is more difficult to prove, however, is that the transitive property should hold for unobservable, or latent, quantities.

One of the earliest attempts to derive a latent measurement scale through observation was by a psychologist, L. L. Thurstone (1927). He sought to *judge*[1] the seriousness of 19 criminal offenses. His objective was not just to order the offenses by seriousness, but to place them on a "psychophysical measurement" scale of "social values". His method assumed stochastic transitivity.

The weakest condition for stochastic transitivity is (1.1.1), where A, B, and C are specific objects. We will discuss other conditions for stochastic transitivity in Section 2.1.4.

$$\text{If } P(A \to B) \geq 0.5 \text{ and } P(B \to C) \geq 0.5, \text{then } P(A \to C) \geq 0.5. \qquad (1.1.1)$$

$A \to B$ means that A is preferred to B, that A beats B, or that A is stronger than B, etc. Then $P(A \to B)$ represents the true proportion of comparisons between A and B where A is favored over B. These objects are static either in truth or by simplifying assumption. Whether the objects being compared are static or dynamic is a philosophical or practical question for the user of a paired comparisons method to justify. Methods themselves treat objects as static. What is random, then, is how the objects are judged or perceived or how the objects perform or are expressed.

Thurstone's model, then, assumed that A, B, C, and any other criminal offense in his study could be ordered from least serious to most serious — that the seriousness of all criminal offenses could lie on a single scale where individual latent ratings were scalars and would follow the transitive property of inequality.

[1] As the first known academic pioneer of the method of paired comparisons in English-speaking countries, he effectively dictated much of the language surrounding paired comparisons in the literature. We will see, however, that many applications of paired comparisons do not involve "judges," per se. Thurstone was pre-dated by a Majorcan, a Frenchman, and at least two Germans in the method of paired comparisons literature.

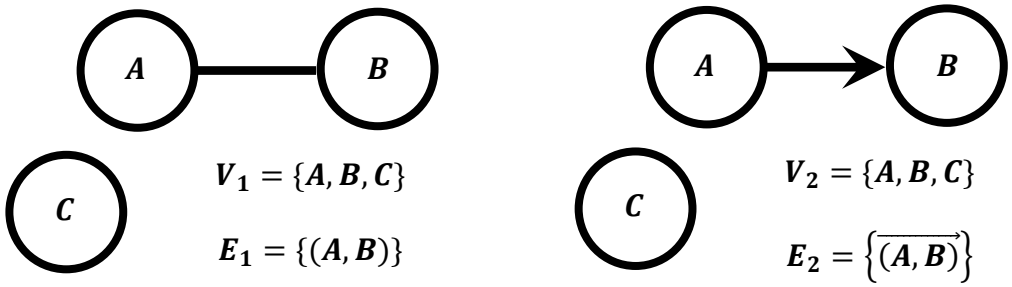
In contrast, another psychologist, Amos Tversky (1969), gave empirical evidence that single judges could have stochastically intransitive preferences (1.1.2) in various experimental settings. These experiments were designed so that stochastically intransitive preferences would be reasonable to expect, although Tversky argued that this implied that the participants were irrational. In (1.1.2), let $\hat{P}$ be an empirical probability observation.

$$\hat{P}(A \to B) \geq 0.5 \text{ and } \hat{P}(B \to C) \geq 0.5, \text{but } \hat{P}(A \to C) < 0.5 \qquad (1.1.2)$$

1.2. Review of Graph Theory for Tournaments

We give several graph theory definitions from Tucker (2012) to make our meaning clear. "A **graph** $G = (V, E)$ consists of a finite set V of **vertices** and a set E of **edges** joining different pairs of distinct vertices. […] We say that vertices a and b are **adjacent** when there is an edge (a, b)". **Directed edges** are ordered pairs of vertices, written $\overrightarrow{(a, b)}$. Undirected edges are written (a, b). In Figure 1-1, we see undirected (V_1, E_1) and directed (V_2, E_2).

Figure 1-1: An Undirected and Directed Graph

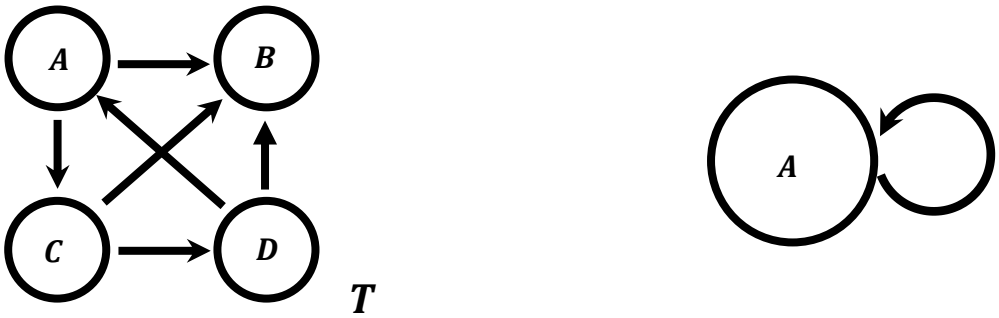


"In a **directed graph**, all edges are directed." "A **path** P is a sequence of distinct vertices in P written" $P = x_1 \to x_2 \to \cdots \to x_n$ when the edges are directed, "with each pair of

consecutive vertices in P joined by an edge. If there is an edge $\overrightarrow{(x_n, x_1)}$, the sequence is a **circuit**, written" $x_1 \to x_2 \to \cdots \to x_n \to x_1$. "A graph is **connected** if there is a path between every pair of vertices." "The number of edges incident to a vertex is called the **degree** of the vertex." "The **in-degree** [is the] number of edges pointed in toward the vertex and **out-degree** [is the] number of edges pointed out."

"A graph with n vertices in which each vertex is adjacent to all the other vertices is called a **complete graph on n vertices**, denoted K_n." "A **tournament** is a directed graph obtained from a complete (undirected) graph by giving a direction to every edge." We see an example of a tournament in Figure 1-2. The reader should not confuse tournaments, complete digraphs, with single-elimination and double-elimination tournaments in tennis and other sports. Single-elimination tournaments are trees, which are never graph theoretic tournaments ($n > 2$). However, a tournament is the same as a round robin tournament, as every pair of teams is compared once. "Two graphs G and G' are called **isomorphic** if there exists a one-to-one correspondence between the vertices in G and the vertices in G' such that a pair of vertices are adjacent in G if and only if the corresponding pair of vertices are adjacent in G'."

Figure 1-2: A Tournament and a Vertex with a Loop

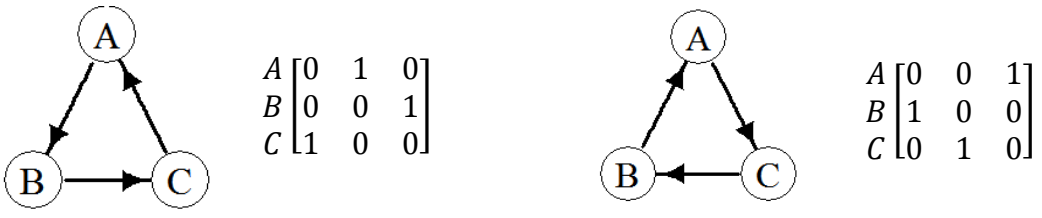


Loops are edges originating and terminating at the same vertex, (i.e. (A, A)), an example of which is in Figure 1-2. We will generally only be concerned with **simple graphs** (graphs containing no loops and no more than one edge between any two vertices), however multigraphs may be of interest in the future. A loop in our setting would only occur if an individual or team could be compared to itself.

We can represent a directed simple graph as a square matrix of zeros and ones, called an **adjacency matrix**. The i^{th} row and j^{th} column represent the i^{th} and j^{th} vertices, while the ij^{th} entry represents directed edge $\overrightarrow{(i, j)}$. Since every vertex pair can only share one edge in a simple graph, an $\overrightarrow{(i, j)}$ edge necessitates that there is no $\overrightarrow{(j, i)}$ edge. In the adjacency matrix, this means that if the ij^{th} entry is 1, then the ji^{th} entry is 0. Also, since there are no loops, the diagonal, or ii^{th}, entries of this matrix are 0.

In Figure 1-3, we see two examples of graphs with their adjacency matrices. The graph on the right is an example of a directed graph $K_3 = (V, E)$ whose vertex set is $V = \{A, B, C\}$ and edge set is $E = \{\overrightarrow{(A, C)}, \overrightarrow{(C, B)}, \overrightarrow{(B, A)}\}$. It contains the path $A \to C \to B$ and circuit $A \to C \to B \to A$. It is connected. Vertex A has a degree of 2, an in-degree of 1, and an out-degree of 1. It has no loops and every vertex pair has up to one edge — thus it is a

Figure 1-3: Two Circular Triads and Their Adjacency Matrices



simple graph. It is isomorphic with the graph on the left — this can be seen by reflecting the triangle across its altitude and swapping the labels C and B.

1.3. Kendall, Smith, and Graph Theory

Kendall and Smith (1940) describe a method of paired comparisons which requires one judge to compare every pair of n objects $A, B, C, \ldots$ directly to one another. They prescribe this method for cases when an object's property of interest is not directly measurable, or the scale of measurement is unknown. When an object A is preferred to an object B, they write that $A \rightarrow B$ or $B \leftarrow A$. Kendall and Smith demonstrate two ways to describe the $\binom{n}{2}$ preferences in their Table 1 and their Figure 1 (p. 326).

The ij^{th} entry in the table is read by row label and then column label for a right-pointing arrow preference. For example, $(A, B) = 1$, indicates $A \rightarrow B$, while $(A, D) = 0$ indicates $D \rightarrow A$. We can see the preference $A \rightarrow B$ is depicted as a directed edge between A and B, $\overrightarrow{(A, B)}$. Also note that when $(i, j) = 1$, $(j, i) = 0 \; \forall \; i \neq j$. Thus, we see in Figure 1-4 that Kendall and Smith's tabular expression of all $\binom{6}{2}$ possible pairwise comparisons between 6 objects is analogous to the adjacency matrix of a complete graph K_6 that is directed and simple.

Figure 1-4: Tabular Expression of Six Compared Objects and Equivalent K_6

	A	B	C	D	E	F
A	—	1	1	0	1	1
B	0	—	0	1	1	0
C	0	1	—	1	1	1
D	1	0	0	—	0	0
E	0	0	0	1	—	1
F	0	1	0	1	0	—

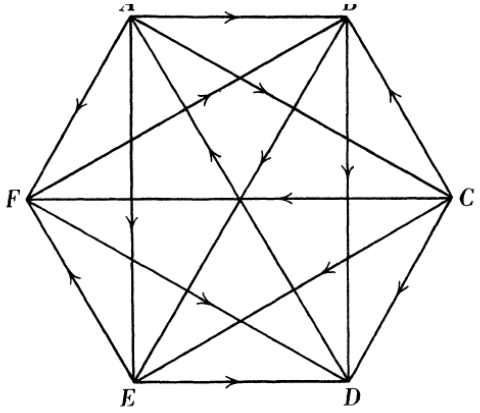


Fig. 1. Geometrical representations of the scheme of preferences of Table I.

Suppose $A \rightarrow B$ and $B \rightarrow C$. It is entirely plausible that $C \rightarrow A$, but ranking requires that A is preferred to C. Pairwise comparison permits this contradiction, while ranking denies it. Kendall and Smith identify the case $A \rightarrow B \rightarrow C \rightarrow A$ as an "inconsistent" preference. In fact, they call every circuit of any size an inconsistent preference. The smallest possible circuits in any K_n are subgraphs with three vertices (since two vertices can only have one directed edge in a simple graph). Kendall and Smith call a circuit of size three a "circular triad". Circular triads are the building blocks of larger circuits, which they call "circular polyads". Circuits of size $k \geq 4$ will always contain at least $k - 2$ circular triads. Kendall and Smith use d, the count of circular triads within K_n, as a unit of inconsistence, as circular polyads are more difficult to count[2]. The more circular triads that occur in a graph, the more inconsistent the judge is or the more difficult it is to compare the

[2] Since, 1940, however, some advances have been made. Alspach and Tabib (1982) found the maximum number of 4-circuits in a tournament and Savchenko (2016) found the maximum number of 5-circuits in a tournament. Various other developments have occurred, but we still accept circular triads as a fundamental unit of inconsistency. Other measures of inconsistency will be discussed later.

objects of interest. We leave the more technical details of d and its distribution for Chapters 2 and 6.

1.4. Bilinear Forms

A **bilinear form** B on a vector space V is a map $V \times V \to K$, where K is the field of scalars. In our case, K is usually $\mathbb{R}$. B is **symmetric** if $B(\boldsymbol{v}, \boldsymbol{w}) = B(\boldsymbol{w}, \boldsymbol{v})$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$. B is **skew-symmetric** if $B(\boldsymbol{v}, \boldsymbol{w}) = -B(\boldsymbol{w}, \boldsymbol{v})$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$. Consider, for example,

$$\boldsymbol{v} = \begin{pmatrix} r_1 \\ r_2 \end{pmatrix}, \boldsymbol{w} = \begin{pmatrix} s_1 \\ s_2 \end{pmatrix},$$

where $r_1, r_2, s_1, s_2 \in \mathbb{R}$. Then

$$B(\boldsymbol{v}, \boldsymbol{w}) = r_1 s_2 - s_1 r_2 = -(s_1 r_2 - r_1 s_2) = -B(\boldsymbol{w}, \boldsymbol{v}).$$

We introduce a model with a skew-symmetric bilinear effect in Chapter 4.

1.5. Looking Forward

Virtually every paired comparison method aims to pick an overall winner, create ratings, or create a ranking. It is common, then, to discuss the qualities of a final ranking (or ordered ratings). Since CRSP does not aim to create a final ranking, this objective is outside of scope. Similarly out of scope, paired comparisons is part of a larger rank aggregation literature — see for Li et al. (2017) for a treatment of rank aggregation. We restrict ourselves specifically to paired comparisons methods for one judge, the relationships between the compared objects, and ratings as a side effect.

In Chapter 2, we first explore how intransitivity in paired comparisons has been addressed in the literature. Then, we explore the standard paired comparison literature and provide some categorization for transitive models. Finally, we discuss the existing

intransitive models in the literature. In Chapter 3, we examine the author's previous work analyzing NBA data. In Chapter 4, our intransitive CRSP model is presented and justified. This model is capable of representing interactions between teams which are intransitive, but requires $O(n)$ parameter estimates instead of $O(n^2)$ parameter estimates, thereby minimizing the possibility of overfitting. In Chapter 5, we see how the CRSP model could be applied, primarily with the NBA. In Chapter 6, we discuss a few ways to enumerate Kendall and Smith's d-distribution and then extend it further than has been accomplished before. In Chapter 7, we discuss opportunities to improve on the methods in Chapter 4 and 6, as well as other available avenues of research into intransitivity and sports analytics.

If one is interested in understanding our work on the circular triad frequency distribution, we recommend reading Chapter 1, Section 2.1, Chapter 6, and Chapter 7.

If one is interested in understanding the CRSP method, we recommend reading Chapters 1, 4, 5, and 7.

If one is interested in sports analytics, we recommend reading Section 2.5.4, Chapter 3, and Chapter 5.

Of course, we hope you read this in its entirety!

CHAPTER 2:

LITERATURE REVIEW

Here, we study the works of our predecessors. We search the literature for evidence that our work in Chapters 4, 5, and 6 is unique and has not been achieved before. Sections 2.1.2, 2.7, 2.8, and 2.9 provide a review of the work which most closely resembles our own. We do not find any evidence of a paper which describes a model for intransitive relations like ours, but we do find a few papers that have made progress on enumerating or cataloguing Kendall and Smith's (1940) d-distribution.

2.1. More on the Nature of Intransitivity

In this section, we further describe Kendall and Smith's work, extensions thereof, competitors to d and ξ, and other work useful in studying the theory of intransitivity.

2.1.1. More on Kendall and Smith

In building the distribution of d in their Table II which we have recreated in Table 2-1, Kendall and Smith assume that for any two objects A and B, it is equally likely that $A \rightarrow B$ or that $A \leftarrow B$ (i.e., $P(A \rightarrow B) = P(A \leftarrow B) = 0.5$). Table II determines how probable it is

that the $\binom{n}{2}$ preferences were made by chance. Someone who is "incapable of making judgments" would flip a fair coin to make $\frac{n(n-1)}{2}$ choices. They would like to test

$$H_0: p_{ij} = 0.5, H_1: p_{ij} \neq 0.5 \; \forall i, j \in (1,2, \dots, n), i \neq j. \qquad (2.1.1.1)$$

Table 2-1: Table II from Kendall and Smith (1940)

n		3		4		5		6		7	
		f	P	f	P	f	P	f	P	f	P
d	0	6	1	24	1	120	1	720	1	5040	1
	1	2	.25	16	.625	120	.883	960	.978	8400	.998
	2			24	.375	240	.766	2240	.949	21840	.994
	3					240	.531	2880	.880	33600	.983
	4					280	.297	6240	.792	75600	.967
	5					24	.023	3648	.602	90384	.931
	6							8640	.491	179760	.888
	7							4800	.227	188160	.802
	8							2640	.081	277200	.713
	9									280560	.580
	10									384048	.447
	11									244160	.263
	12									233520	.147
	13									72240	.036
	14									2640	.001
Total		**8**		**64**		**1024**		**32768**		**2097152**	

$(1 - P)$ in Kendall and Smith's Table II can be interpreted as a typical p-value for this test — we reject H_0 when $1 - P$ is small (i.e., for small d). Rejecting H_0 can be interpreted as it being "improbable that the observer is completely incapable of judgment. We might then be led to suppose that his small deviation from internal consistence is due to fluctuation of attention, very close resemblance to the objects giving rise to the

inconsistencies, or both." Mosteller (1958) calls (2.1.1.1) the "nullest of null hypotheses." They also give a coefficient of consistence, which we give in (2.1.1.2). Large values of ξ indicate a capable judge. Like Kendall's τ, $\xi \in [0, 1]$ and is comparable across n.

$$\xi = 1 - \frac{d}{\max(d)} \tag{2.1.1.2}$$

Kendall and Gibbons (1990) suggest approximating the d distribution with a χ^2 distribution for hypothesis testing purposes. Given n and a particular d value, they give the test statistic as

$$X^2 = \frac{8}{n-4}\left[\frac{1}{4}\binom{n}{3} - d + \frac{1}{2}\right] + \nu, \text{where } df = \nu = \frac{n(n-1)(n-2)}{(n-4)^2}. \tag{2.1.1.3}$$

Considering the remarkable amount of time needed to enumerate the d-distribution for a given $n > 10$, even with modern computing power, the χ^2 approximation is fairly useful.

2.1.2. Counting Circular Triads

To assess how unusual it would be for a tournament to assume the null hypothesis in (2.1.1.1), Kendall and Smith (1940) provide a distribution of d in Table II (Table 2-1 above) under the null hypothesis. We discuss how to calculate the values in this table in Chapter 6, where we also extend the table to n larger than in Kendall and Smith. Since 1940, two papers have extended this table to $n \leq 10$ (Alway, 1962) and $n \leq 15$ (Knezek, et al. 1998).

Alway produced an algorithm to achieve his result. We find later that one of our independently developed algorithms matches his algorithm closely (details in Chapter 6). Knezek et al. (1998) also extended the d distribution up to $n = 15$; however, they do not

provide exact values of d for anything other than common critical values. They do show anecdotally that as n increases, the accuracy of the χ^2 approximation improves.

In describing a method to calculate d for a given tournament, Kendall and Smith (1940) 'invent' a way to summarize the tournaments — with α-vectors. An α-vector $\boldsymbol{\alpha}$ is a score vector of a tournament's adjacency matrix, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n)$. Each element of an α-vector, α_i, represents the number of times an object i is preferred over the other $n-1$ objects, or the number of wins a team i has in a round-robin tournament. So, $\alpha_i \in \{0,1, \dots, n-1\}$ and $\sum_{i=1}^{n} \alpha_i = \frac{n(n-1)}{2}$. By convention, the α_i are ordered from least to greatest or greatest to least and ignore the order of the object labels. Our convention will be to order them from least to greatest.

These α-vectors are useful for cataloging and keeping track of graphs to count d. Kendall and Smith do not appear to have advanced to using α-vectors for cataloging purposes, and instead use a more fundamental method of counting. However, the original method of counting is not described in their 1940 paper in reaching the full d-distribution for $n = 7$. They say, "The principles of [our] method are clear enough and the work may be formalized by a number of conventions which we omit to save space. In common with many similar combinatorial problems, however, troubles arise from the sheer number of possibilities and the difficulty of ensuring that nothing is overlooked."

In early animal dominance structure literature regarding 'peck right,' a more general, not necessarily transitive version of 'pecking order' describing tournaments among animals, Davis (1954) calls α-vectors 'structure sequences.' He and Rapoport (1949) and Landau (1951), (1953) appear to be unaware of Kendall and Smith (1940) —

they calculate probabilities of particular structure sequences for no longer than the $n = 6$ case.

Later, David (1959) uses a different notation for α-vectors. He orders the α_i from greatest to least and uses 'exponents' to denote repeated counts. For example, if we had $\boldsymbol{\alpha} = (1, 3, 3, 3, 3, 4, 4)$, he would compactly write this as $[4^2 3^4 1]$. Formally, he calls these partitions $[x_1^{r_1} x_2^{r_2} \ldots x_m^{r_m}]$ where the r_u are the frequency of the score x_u. In our most memory-conscious method of counting the α-vectors for the d-distribution, the r_u play a pivotal role. David (1959) is the first to list and count the α-vectors for the $n = 8$ case, the largest n for which we have seen the α-vectors listed. He gives a generating function approach to listing all possible outcomes, but this approach amounts to a brute force technique with much redundant counting.

Alway (1962), working in parallel with Slater (1961), produced a computer program to enumerate the d-distribution to the $n = 10$ case. The program is nearly identical to the one devised independently by this author in August 2016, documented in Section 6.2.4. The method involves building the list of α-vectors and their counts in sequential n. For each α-vector of length n, an additional object is included and compared to the previous n objects. Each comparison is binary — a win or a loss — so there are 2^n possible results when expanding the α-vector by one item. The resulting α-vectors of length $n + 1$ are not necessarily unique, so the redundancy is counted.

Kadane (1966) devises a method of counting the α-vector frequencies by d working largely from the counts David (1959) produces for each α-vector with the assistance of some unexplored conjectures from Kendall and Smith (1940). His proposed method appears to reduce some of the redundancy of Alway's method (1962), but it appears to

have gone largely ignored in the literature, including in Knezek et al. (1998) in which the authors claim to have extended the d-distribution to the $n = 15$ case.

Kadane (1966) has his own notation for alpha vectors. In our notation, we have $\boldsymbol{\alpha} = (1, 3, 3, 3, 3, 4, 4)$ or $[4^2 3^4 1]$ in David's notation. Kadane would write 0-0-2-4-0-1-0, where David's r_u (i.e., 2, 4, and the implicit exponent 1) are listed in a position placeholder, which may or may not include an x_u. Note that in this $n = 7$ case, the largest integer, 6, would represent the most preferences an object could have over the other objects. Then Kadane's notation refers to the number of times 6, 5, 4, 3, 2, 1, and 0, respectively, appear in an alpha vector. Henceforth, we refer to this notation as *Kadane notation*.

Kadane (1966) introduces the concept of simple α-vectors and *compound* α-vectors, but we refer to simple α-vectors as *atomic* α-vectors. The distinction between atomic and compound α-vectors and their usefulness is most easily seen in Kadane notation. A compound α-vector is one which is composed by juxtaposing two or more atomic α-vectors or by juxtaposing 1s or both, and an atomic α-vector is an α-vector which cannot be composed by juxtaposing multiple α-vectors or 1s. Atomic α-vectors ($n \geq 3$) always have a leading and ending 0 in Kadane notation, for example 0-3-0 or 0-2-2-0. If an object were added to a circular triad (0-3-0), and won every comparison, we would have the compound α-vector 1-0-3-0. Similarly, if the new object lost every comparison, we would have the compound α-vector 0-3-0-1. All three α-vectors (0-3-0, 1-0-3-0, and 0-3-0-1) have $d = 1$, and there are an infinite number of compound α-vectors with $d = 1$ which can be constructed by padding the beginning and end of the atomic 0-3-0 circular triad α-vector with 1's. We can also create compound α-vectors by pressing two or more atomic α-vectors together. For example, 0-3-0-0-3-0 is an $n = 6$ compound α-vector with $d = 2$. In

this example, the first three objects always beat the next three objects, but the first three objects are in a circular triad and the second three objects are in a circular triad. In adjacency matrix form, this is fairly easy to recognize as well:

$$\begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \Rightarrow \begin{bmatrix} 4 \\ 4 \\ 4 \\ 1 \\ 1 \\ 1 \end{bmatrix}. \qquad (2.1.2.1)$$

In block form, where $\boldsymbol{C}$ is a circular triad matrix, $\mathbf{0}$ is a matrix of 0's, and $\mathbf{1}$ is a matrix of 1's, this adjacency matrix is $\begin{bmatrix} \boldsymbol{C} & \mathbf{1} \\ \mathbf{0} & \boldsymbol{C} \end{bmatrix}$, where the top three objects always beat the bottom three objects. It is helpful to recognize that a compound α-vector can always be nontrivially represented in a block form adjacency matrix, where the blocks are $\mathbf{0}$, $\mathbf{1}$, and adjacency matrices corresponding to atomic α-vectors. Note the row sums of the adjacency matrix in (2.1.2.1) are the α-vector (1, 1, 1, 4, 4, 4), or 0-3-0-0-3-0 in Kadane notation. Finally, compound α-vectors can also be composed of one or more atomic α-vectors and one or more 1s (i.e., the $d = 2, n = 7$ α-vector 0-3-0-1-0-3-0).

Formally, Kadane (1966) notes that any tournament can be arranged into disjoint sets $S_1, S_2, \dots, S_k$ (for some k, $k \in [1, \dots, n]$) with the following properties:

a) Each object in S_h is preferred to every object in $S_{h'}$ for all $h < h'$; $h, h' \in [1, \dots, k]$;

b) For any two objects A_i, A_j in the same set S_h, either $A_i \to A_j$ or there exist other objects $A_{i_1}, A_{i_2}, \dots$, in S_h such that $A_i \to A_{i_1} \to A_{i_2} \to \cdots \to A_j$. Such a set is called a *strong subtournament*.

So, an atomic α-vectors represents a tournament where $k = 1$, while compound α-vectors are tournaments where $k > 1$ — a compound α-vector has two or more strong subtournaments and an atomic α-vector is the only strong subtournament.

Kadane's method uses the frequencies of the less common atomic α-vectors. He provides formulae for up to the $d = 5$ case on page 492. We use Kadane's method to provide the formula for $d \leq 13$, with $d \leq 20$ easily achievable with the d-distribution up to $n = 16$, and Kadane's method provides useful insights and notation for our most space-conscious d-distribution enumeration algorithm alluded to in Chapter 6.

Moran (1947), Davis (1954), and Narayana and Bent (1964) provide important results useful for contextualization and error checking. Moran proves that the first four moments of the d-distribution conjectured by Kendall and Smith are correct. Davis creates a method to count the number of isomorphic tournament graphs for any n. The On-Line Encyclopedia of Integer Sequences (2017) has stored this sequence as A000568. Narayana and Bent (1964) derive a formula to count the number of unique score vectors for any n, which is sequence A000571 in the OEIS. We list this sequence in Table 2-2. Note that one would need to use a $259{,}451{,}116 \times 20$ matrix to simply store all α-vectors for the $n = 20$ case on a computer.

Table 2-2: Number of Unique Score Vectors for $n \in (3, \ldots, 32)$

n	Number of unique α-vectors
3	2
4	4
5	9
6	22
7	59
8	167
9	490
10	1,486
11	4,649
12	14,805
13	48,107
14	158,808
15	531,469
16	1,799,659
17	6,157,068

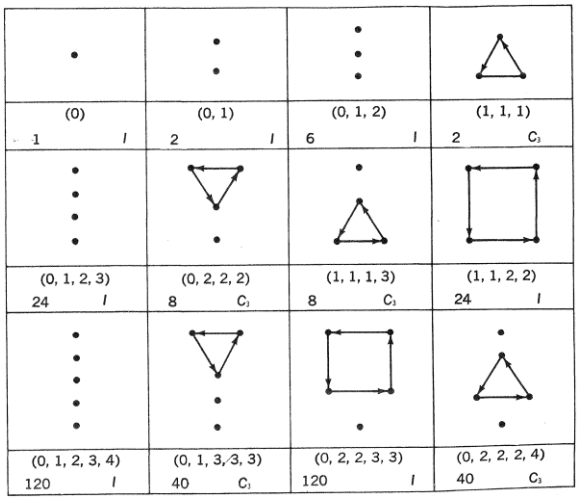
n	Number of unique α-vectors
18	21,258,104
19	73,996,100
20	259,451,116
21	951,695,102
22	3,251,073,303
23	11,605,141,649
24	41,631,194,766
25	150,021,775,417
26	542,875,459,724
27	1,972,050,156,181
28	7,189,259,574,618
29	26,295,934,251,565
30	96,478,910,768,821
31	354,998,461,378,719
32	1,309,755,903,513,480

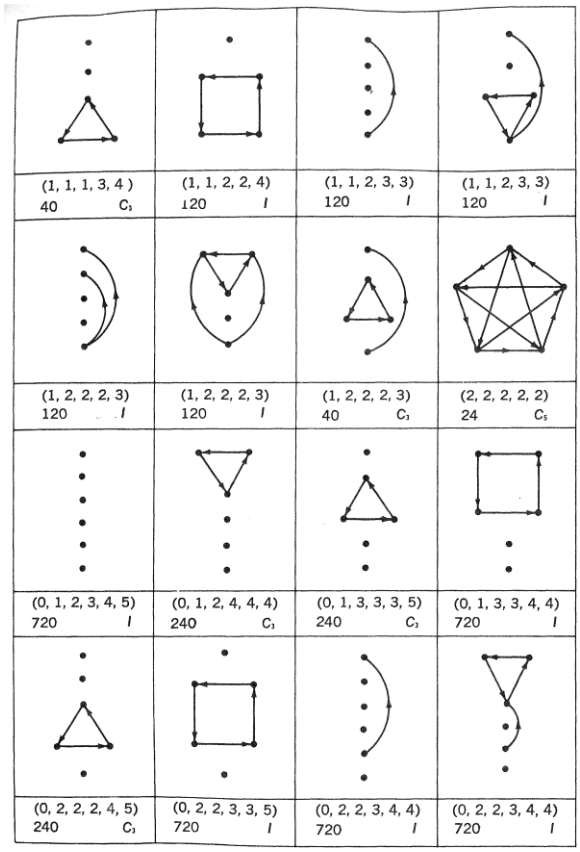
2.1.3. Moon on Tournaments and Related Graph Theory

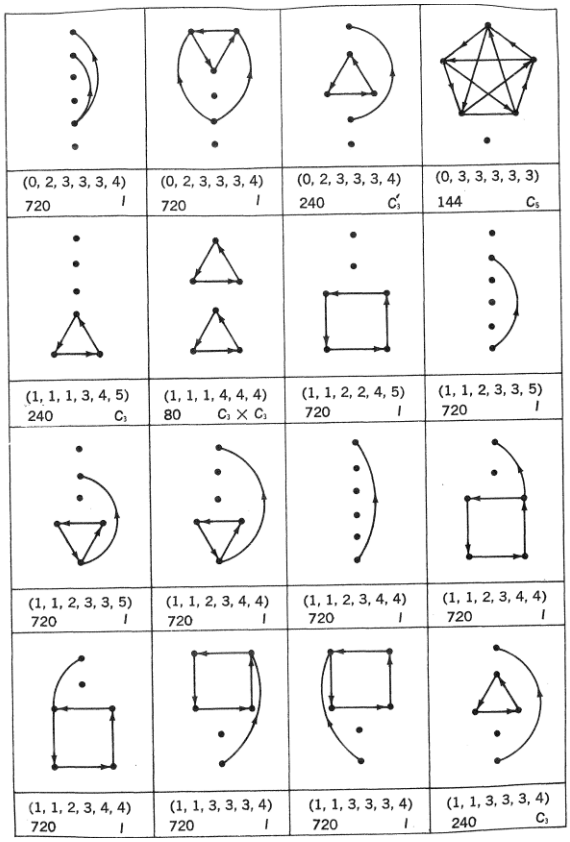
Moon (1968) provides an excellent collection of results regarding tournaments, including most of the preceding papers. Moon appears to be one of the first mathematicians that has discovered paired comparisons contributions from other fields, like animal behavior. For example, he notes contributions from Davis (1954) and Landau (1953), which provide earliest versions of some proofs regarding tournaments. Most importantly, Moon (1968) introduces an exhaustive collection of non-isomorphic tournaments for $n \in \{1, 2, \ldots, 6\}$. In Figure 2-1, unconnected vertices stacked vertically imply strictly transitive relationships. For example, the chain $(0, 1, 2)$ includes three undrawn vertices — two vertices pointing from the top object to the two lower objects, and a vertex pointing from the middle object to the bottom object. We call the initial appearance of a form a *canonical*

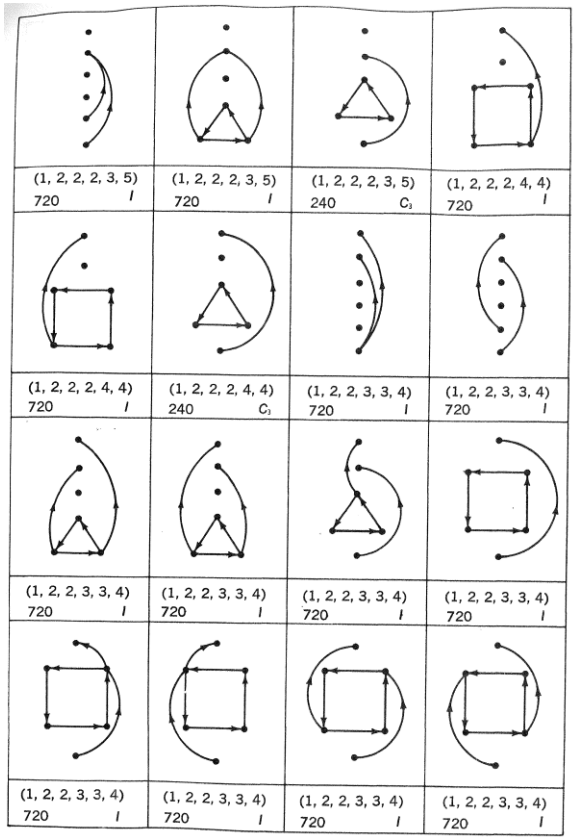
form. For example, a circular triad (1, 1, 1) is the only canonical form for $d = 1$. The tournaments below, corresponding to (0, 2, 2, 2), (1, 1, 1, 3), (0, 1, 3, 3, 3), (0, 2, 2, 2, 4), (1, 1, 1, 3, 4), and the similar graphs for $n = 6$, are all $d = 1$ graphs which feature a circular triad in an otherwise strictly ordered set of objects.

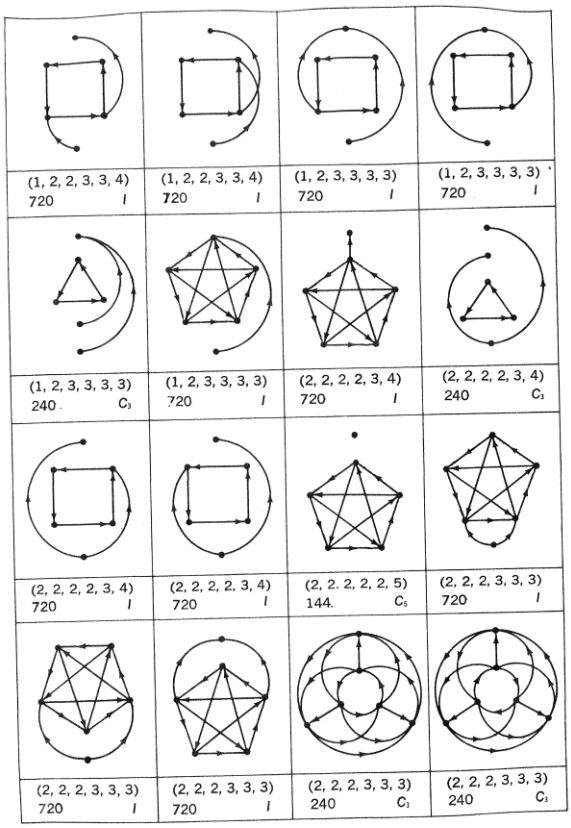
Figure 2-1: Non-Isomorphic Matrices from Moon (1968)




(1, 1, 1, 3, 4)
40 C_3
(1, 1, 2, 2, 4)
120 I
(1, 1, 2, 3, 3)
120 I
(1, 1, 2, 3, 3)
120 I
(1, 2, 2, 2, 3)
120 I
(1, 2, 2, 2, 3)
120 I
(1, 2, 2, 2, 3)
40 C_3
(2, 2, 2, 2, 2)
24 C_5
(0, 1, 2, 3, 4, 5)
720 I
(0, 1, 2, 4, 4, 4)
240 C_3
(0, 1, 3, 3, 3, 5)
240 C_3
(0, 1, 3, 3, 4, 4)
720 I
(0, 2, 2, 2, 4, 5)
240 C_3
(0, 2, 2, 3, 3, 5)
720 I
(0, 2, 2, 3, 4, 4)
720 I
(0, 2, 2, 3, 4, 4)
720 I


(0, 2, 3, 3, 3, 4)
720 I
(0, 2, 3, 3, 3, 4)
720 I
(0, 2, 3, 3, 3, 4)
240 C_3'
(0, 3, 3, 3, 3, 3)
144 C_5
(1, 1, 1, 3, 4, 5)
240 C_3
(1, 1, 1, 4, 4, 4)
80 $C_3 \times C_3$
(1, 1, 2, 2, 4, 5)
720 I
(1, 1, 2, 3, 3, 5)
720 I
(1, 1, 2, 3, 3, 5)
720 I
(1, 1, 2, 3, 4, 4)
720 I
(1, 1, 2, 3, 4, 4)
720 I
(1, 1, 2, 3, 4, 4)
720 I
(1, 1, 2, 3, 4, 4)
720 I
(1, 1, 3, 3, 3, 4)
720 I
(1, 1, 3, 3, 3, 4)
720 I
(1, 1, 3, 3, 3, 4)
240 C_3


(1, 2, 2, 2, 3, 5)
720 I
(1, 2, 2, 2, 3, 5)
720 I
(1, 2, 2, 2, 3, 5)
240 C_3
(1, 2, 2, 2, 4, 4)
720 I
(1, 2, 2, 2, 4, 4)
720 I
(1, 2, 2, 2, 4, 4)
240 C_3
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I


(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 2, 3, 3, 4)
720 I
(1, 2, 3, 3, 3, 3)
720 I
(1, 2, 3, 3, 3, 3)
720 I
(1, 2, 3, 3, 3, 3)
240 C_3
(1, 2, 3, 3, 3, 3)
720 I
(2, 2, 2, 2, 3, 4)
720 I
(2, 2, 2, 2, 3, 4)
240 C_3
(2, 2, 2, 2, 3, 4)
720 I
(2, 2, 2, 2, 3, 4)
720 I
(2, 2. 2, 2, 2, 5)
144. C_5
(2, 2, 2, 3, 3, 3)
720 I
(2, 2, 2, 3, 3, 3)
720 I
(2, 2, 2, 3, 3, 3)
720 I
(2, 2, 2, 3, 3, 3)
240 C_3
(2, 2, 2, 3, 3, 3)
240 C_3

Kernan (1967) delivers the first explicitly graph theoretic interpretation of Kendall and Smith (1940) to an audience of marketing researchers and casts doubt on the efficacy of the method of paired comparisons. As evidence, he cites another marketing paper titled "Let's Bury Paired Comparisons" by Blankenship (1966), who claims that paired comparisons do not reflect reality — consumers purchase one item of many. Additionally, paired comparison methods do not determine an actual location of item strength, just relative strengths. In their setting, that means that product A may be preferred to product B, but that does not mean that product A will sell well. He advocates additional research to address that issue but concedes that paired comparison methodology gives good additional information.

Burns (1980) produced a FORTRAN IV algorithm which performs an exhaustive search of all three-object combinations for circular triads in a set of n objects. Later, Gass (1998) gives a linear programming solution to identifying circular triads in a shipping network represented by an adjacency matrix. Brown and Peterson (2009) introduce the concept of double-sorting a tournament adjacency matrix — the matrix is sorted by both row sums (α-vectors) in descending order and column sums in ascending order. Since a strictly transitive double-sorted tournament matrix should have all 1s in the upper right triangle, it should be easy to begin to recognize which relationships are circular triads since entries below the diagonal should be 0, not 1.

In animal dominance studies, near-transitive orderings are often converted to the closest transitive ordering, but Appleby (1983) re-introduces the field to Kendall and Smith's method (1940) and cautions that this practice is presumptive and recommends testing for transitivity first. Shader (1992) gives a theorem bounding the spectral radius of

the possible adjacency matrices for a given α-vector. Füredi (2010) describes some open mathematical problems in tournament matrices.

2.1.4. Stochastic Transitivity Conditions

In Section 1.1, we introduced the weak stochastic transitivity (WST) condition (1.1.1), and repeat it here — if $P(A \to B), P(B \to C) \geq 0.5$, then $P(A \to C) \geq 0.5$, where A, B, and C are specific objects. However, there are dozens of other testable conditions for stochastic transitivity. David (1988) provides an excellent overview of the paired comparisons literature and lists three conditions — the SST, MST, and WST. However, Morrison (1962), (1963) and Fishburn (1973) list, prove, discuss, and relate over a dozen more.

The partial stochastic transitivity (PST) condition is

If $P(A \to B), P(B \to C) > 0.5$, then $P(A \to C) \geq \min\big(P(A \to B), P(B \to C)\big)$.

The moderate stochastic transitivity condition (MST) is

If $P(A \to B), P(B \to C) \geq 0.5$, then $P(A \to C) \geq \min\big(P(A \to B), P(B \to C)\big)$.

The strong stochastic transitivity condition (SST) is

If $P(A \to B), P(B \to C) \geq 0.5$, then $P(A \to C) \geq \max\big(P(A \to B), P(B \to C)\big)$.

The strict stochastic transitivity condition (SSST) is

If $P(A \to B), P(B \to C) > 0.5$, then $P(A \to C) > \max\big(P(A \to B), P(B \to C)\big)$.

Additionally, there can be a continuum between the SST and MST with $\lambda \in [0,1]$ as well as between the MST and WST. If the condition met is again that $P(A \to B), P(B \to C) \geq 0.5$, then the SST-$\lambda$-MST (2.1.4.1) and MST-$\lambda$-WST (2.1.4.2) implications follow, respectively

$$P(A \to C) \geq \left(\lambda \max\left(P(A \to B), P(B \to C)\right) + (1-\lambda) \min\left(P(A \to B), P(B \to C)\right)\right) \quad (2.1.4.1)$$

$$P(A \to C) \geq \left(0.5\lambda + (1-\lambda) \min\left(P(A \to B), P(B \to C)\right)\right). \quad (2.1.4.2)$$

The triangular stochastic transitivity condition (TST) is

$$P(A \to B) + P(B \to C) \geq P(A \to C).$$

The multiplicative stochastic transitivity condition (MuST) is

$$P(A \to B) \cdot P(B \to C) \leq P(A \to C).$$

The disjunctive hypothesis multiplicative stochastic transitivity condition (DMuSt), which requires either $P(A \to B) \geq 0.5$ or $P(B \to C) \geq 0.5$ is

$$P(A \to B) \cdot P(B \to C) \leq P(A \to C).$$

The conjunctive hypothesis multiplicative stochastic transitivity condition (CMuSt), which requires $P(A \to B), P(B \to C) \geq 0.5$ is

$$P(A \to B) \cdot P(B \to C) \leq P(A \to C).$$

The symmetric intransitive triad condition (SIT) is

$$P(A \to B) \cdot P(B \to C) \cdot P(C \to A) = P(B \to A) \cdot P(A \to C) \cdot P(C \to B).$$

The acyclic stochastic transitivity (AST) condition is

$$\text{If } P(A \to B), P(B \to C), \dots, P(Y \to Z) > 0.5, \text{then } P(A \to Z) \geq 0.5.$$

The negative stochastic transitivity (NST) condition is

$$\text{If } P(A \to C) > 0.5, \text{then } P(A \to C) \leq \max\left(P(A \to B), P(B \to C)\right).$$

The interval stochastic transitivity (IST) condition is

$$\max\left(P(A \to B), P(C \to D)\right) \geq \min\left(P(A \to D), P(C \to B)\right).$$

The just noticeable difference stochastic transitivity (JST) condition is

$$\max\left(P(A \to B), P(B \to C)\right) \geq \min\left(P(A \to D), P(D \to C)\right).$$

The K stochastic transitivity (KST) condition, so named because K follows I and J, is

$$P(A \to B) > \max(P(B \to C), 0.5) \Rightarrow \min\big(P(A \to D), P(D \to C)\big) \geq P(B \to D).$$

The substitutable stochastic transitivity (SubST) condition is

$$P(A \to C) > P(B \to C) \Leftrightarrow P(A \to B) > 0.5.$$

Fishburn (1973) lists a few more esoteric probabilistic stochastic transitivity conditions, which we will not add to the 18 conditions above (WST, PST, MST, SST, SSST, SST-λ-MST, MST-λ-WST, TST, MuST, DMuST, CMuST, SIT, AST, NST, IST, JST, KST, SubST). Additionally, Iverson and Falmagne (1985) give likelihood ratio tests for some of the stochastic transitivity conditions in this section.

2.1.5. Measures of Intransitivity

Kendall and Smith introduced d, the number of circular triads in a tournament, in 1940. They considered counting cycles of lengths greater than three as well, but there are several reasons circular triads made more sense at the time. First and foremost, every circuit with $n > 3$ nodes is guaranteed to have at least $n - 2$ circular triads. Circular triads account for larger circuits, then. Second, the mathematical theory behind circuits of size $n = 4$ or greater is an open area of research to this day. Alspach and Tabib (1982) found the maximum number of 4-circuits in a tournament of size n and Savchenko (2016) found the maximum number of 5-circuits in a tournament of size n. Wormald (2004) finds the asymptotic lower bound for the maximum number of Hamilton circuits (length n circuits) in a tournament. In short, it was entirely reasonable for circular triads to be considered the fundamental unit of inconsistency. They also favored discussing ξ, the coefficient of consistence, which is

$$\xi = 1 - \frac{24d}{n^3 - \gamma n} \qquad \begin{cases} \gamma = 1 & \text{if } n \text{ is odd} \\ \gamma = 4 & \text{if } n \text{ is even} \end{cases}. \tag{2.1.5.1}$$

ξ is standardized on [0, 1] for all $n \geq 3$.

However, there were many other attempts to make a measure of inconsistency, starting with Landau (1951), who did not appear to be aware of Kendall and Smith (1940). Following the work of Rapoport, he called α-vectors "hierarchy structures" instead of "structure sequences". He defined his measure of intransitivity, h, the hierarchy index.

$$h = \frac{12}{n^3 - n} \sum_{i=i}^{n} \left(\alpha_i - \frac{n-1}{2} \right)^2 \tag{2.1.5.2}$$

However, this is virtually identical to Kendall and Smith's ξ. We provide a proof for (2.5.1.3) in Chapter 6, but also David (1988) reports that d can be calculated as a function of $\boldsymbol{\alpha}$ and n,

$$d = \frac{n^3 - n}{24} - \frac{1}{2} \sum_{i=i}^{n} \left(\alpha_i - \frac{n-1}{2} \right)^2. \tag{2.5.1.3}$$

Let $T = \sum_{i=i}^{n} \left(\alpha_i - \frac{n-1}{2} \right)^2$. Then we can see from (2.1.5.2) and (2.5.1.3) that

$$d = \frac{n^3 - n}{24} - \frac{1}{2} T,$$

$$h = \frac{12}{n^3 - n} T,$$

$$d = \frac{n^3 - n}{24} (1 - h), \text{and}$$

$$h = 1 - \frac{24d}{n^3 - n}. \tag{2.5.1.4}$$

We can see from inspection of (2.5.1.1) and (2.5.1.4) that $\xi = h$ when n is odd, but they are not the same when n is even. Note that, when n is even, $\max(d) = \frac{n^3 - 4n}{24}$, which gives that $\min(h) = 1 - \frac{n^2 - 4}{n^2 - 1} = \frac{3}{n^2 - 1}$.

Landau knew that $\min(h)$ differed when n is odd and even. In short, Kendall and Smith's ξ and Landau's h are virtually identical and have a one-to-one mapping which are both directly related to the d-distribution.

Slater (1961) introduced a novel way of describing intransitivity. His measure is i, the minimum number of edge direction swaps required to get an entirely transitive ordering. While heuristic approaches can give a close estimate of i, to guarantee the minimum is found, all $n!$ fully transitive orders must be compared to the observed adjacency matrix. The order which minimizes the number of swaps required is called the *nearest adjoining order*. Slater's primary concern is,

> *"Triads ought not to be treated as elements. They are compound events not conceivably independent of one another, for the total number of triads in a schedule exceeds the number of responses by a factor of* $(n-2)/3$ *and each response features in* $n-2$ *triads. Moreover, there is no 1:1 relationship between* i *and* d*; schedules from the same* n *with the same* i *may differ in* d*, and vice versa."*

Consider, for example, a perfect ordering, where a single swap occurs; $i = 1$. If the formerly third-ranked object is preferred over the formerly first-ranked object, then $d = 1$. If instead the formerly lowest-ranked object is preferred over the formerly first-ranked object, then $d = n - 2$. While Slater is critical of d, he still relates i to Kendall's τ by

$$\max(\tau) = 1 - \frac{2i}{\binom{n}{2}}. \qquad (2.1.5.5)$$

Kendall and Smith (1940) note that d can be calculated by counting the number of adjacent swaps required to break all ties (elements of $\boldsymbol{\alpha}$ that are equal) in an α-vector. Similarly, Kendall's τ is calculated by counting reversals in one set of ranks to match another set of

ranks as well. We discuss the notion of reversals further in proofs in Chapter 6. David (1988) notes that since d can be obtained from an α-vector by restricted reversing, $i \leq d$. For $n \in (4, \dots, 8)$, Slater (1961) demonstrates that $\text{corr}(i, d) \approx 0.9$. He also gives enumerated distributions for i when $n \in (4, \dots, 8)$ and equations to enumerate small i for all n. We note that the distribution of i could be calculated with unique adjacency matrices, but α-vectors may not have enough information for i to be calculated. This warrants further, future examination.

Remage and Thompson (1966) give a heuristic method to enumerate only $n2^{n-1}$ tournaments instead of the $n!$ Slater proposed. Abeles (1979) notes that C.L. Dodgson (better known as Lewis Carroll) anticipated the approach Remage and Thompson give in three pamphlets on voting published between 1873 and 1876. We also note that by double-sorting an adjacency matrix of interest, it is possible to calculate i by counting the non-zero lower triangle entries in at least some cases. This may provide for a faster solution to calculating i than Remage and Thompson propose by reducing the space of possibilities. Finally, Nurmi (2014) studies Slater's nearest adjoining order, generalizes the method, and compares it to other orders popular in social choice theory.

Bezembinder (1981) is critical of both Kendall and Smith's d and Slater's i as measures of inconsistency and introduces his own measure, ρ. His measure is calculable when graphs are incomplete and "accounts" for circular polyads of any size (i.e., in $[3, 4, \dots, n]$). Bezembinder was not aware of Kadane (1966), however, his method is most easily elucidated in Kadane notation. Continuing with Kadane's notation for strong subtournaments in Section 2.1.2, where k is the number of strong subtournaments, $S = \{S_1, S_2, \dots, S_k\}$ is the set of strong subtournaments, n_j is the number of nodes in strong

subtournament S_j, and N_j is the number of edges in S_j ($N_j = \binom{n_j}{2}$ in our setting — a tournament), Bezembinder gives

$$\rho = k - n + \sum_{j=1}^{k} N_j \,. \tag{2.1.5.6}$$

Note that $\sum_{j=1}^{k} N_j$ is the number of edges in the tournament that are included in any cycle. When a node is completely transitive in relation to all strong subtournaments, it is itself a strong subtournament of size 1. If a tournament is fully transitive, then there are $k = n$ strong subtournaments and $\sum_{j=1}^{k} N_j = 0$ edges included in any cycle. Thus, $\rho = 0$ when $d = 0$. Meanwhile, if the tournament can be represented by an atomic α-vector, or rather, when a tournament is strong and $k = 1$, then every node is included in a cycle — $\sum_{j=1}^{k} N_j = \binom{n}{2}$ and $\rho = 1 - n + \binom{n}{2} = \frac{1}{2}(n-1)(n-2)$.

Consider Bezembinder's tournament example, where $\alpha = (1, 1, 1, 4, 4, 5, 5, 7)$ and $n = 8$. In Kadane notation, this is 0-3-0-0-2-2-0-1, which we may recognize a compound α-vector composed of the circular triad 0-3-0, the $d = 2$ and $n = 4$ atomic α-vector 0-2-2-0, and a node transitive in relation to the rest of the nodes, 1. There are then $k = 3$ strong subtournaments with $n_1 = 3, n_2 = 4, n_3 = 1, \sum_{j=1}^{k} N_j = 3 + 6 + 0 = 9$. Then $\rho = 3 - 8 + 9 = 4$. Meanwhile, we can see that $d = 3$ by adding the number of circular triads in each strong subtournament. Also, note that we can calculate Slater's i in the same way. In a circular triad, $i = 1$. In the $n = 4$ atomic α-vector, $i = 2$. So Slater's i in this tournament is $i = 3$.

Bezembinder does implicitly give a nice way to calculate k from the α-vector:

$$k = \sum_{i=1}^{n} \mathbb{I}\left[\binom{i}{2} = \sum_{j=1}^{i} \alpha_j\right]. \tag{2.1.5.7}$$

The calculation for $\sum_{j=1}^{k} N_j$ using the α-vector is a bit more involved, but simply requires finding the size n_j of each strong subtournament after which $\sum_{j=1}^{k} N_j = \sum_{j=1}^{k} \binom{n_j}{2}$. However, Bezembinder's ρ can be calculated with only knowledge of the α-vector and a short algorithm.

The more interesting measure Bezembinder introduces, but later neglects, is δ, the proportion of edges in the tournament that are included in a cycle:

$$\delta = \frac{\sum_{j=1}^{k} N_j}{\binom{n}{2}}. \tag{2.1.5.8}$$

This measure is in [0, 1] and is approximately comparable across all n.

Bezembinder also introduces the concepts of internal and external consistency. An external measure of consistency is one which compares a given ranking with another ranking. Examples include Kendall's τ, Pearson's r, and two measures Bezembinder introduces, and Lundh's tournament stability index (2006). An internal measure of consistency is one in which a set of comparisons are consistent with themselves. We are concerned with internal measures of intransitivity in this section, 2.1.5, including d, i, ρ, and δ.

Bezembinder claims that his internal consistency measures accounts for cycles of any size, however, no quantity in ρ or δ explicitly does that. $\sum_{j=1}^{k} N_j$ does represent the number of edges that are embedded in a cycle, but we do not know much about the size of these cycles. As we know from Kendall and Smith (1940), every cycle of size n includes $n -$

2 circular triads. Bezembinder's ρ accounts for cycles of any size no more than d does. Interestingly, Bezembinder's student Maas introduced a measure of circularity where he modified Kendall and Smith's d for tournaments where some preferences are known a priori (1993).

Monsuur and Storcken (1997) survey d, i, ρ, and δ and introduce their own measure ν which can be calculated by finding the ℓ_1 norm (or Manhattan distance) between an α-vector $\boldsymbol{\alpha}$ and the α-vector of a fully transitive tournament of size n:

$$\nu = \frac{1}{2}\|\boldsymbol{\alpha} - (0, 1, 2, \dots, n-1)\|_1. \tag{2.1.5.9}$$

They also introduce a framework to characterize measures of intransitivity. Monsuur later (2005) discusses another measure of intransitivity, although it appears to be an external measure in Bezembinder's framework (1981). Cropper (2011) explores powers of α-vectors as a way to rank α-vectors. Reid and Santana (2015) discuss a method to check if a vector is a valid α-vector. Finally, Kulakowski (2017) introduces a measure of inconsistency extending Kendall and Smith's ξ to the case where ties are allowed.

2.1.6. Returning to Order

Many authors consider intransitivities as things that need to be "solved". The first such author was Patrick Slater (1961). His method was to find the minimum number of inconsistent responses, denoted as i, and then to swap those responses in the adjacency matrix to give a fully transitive tournament which he called the "nearest adjoining order." However, this method required enumerating all $n!$ transitive tournaments. Remage and Thompson (1966) improve on the search method by enumerating only $n2^{n-1}$ tournaments. Later, Flueck and Korsh (1975) introduce an improvement on Remage and Thompson's

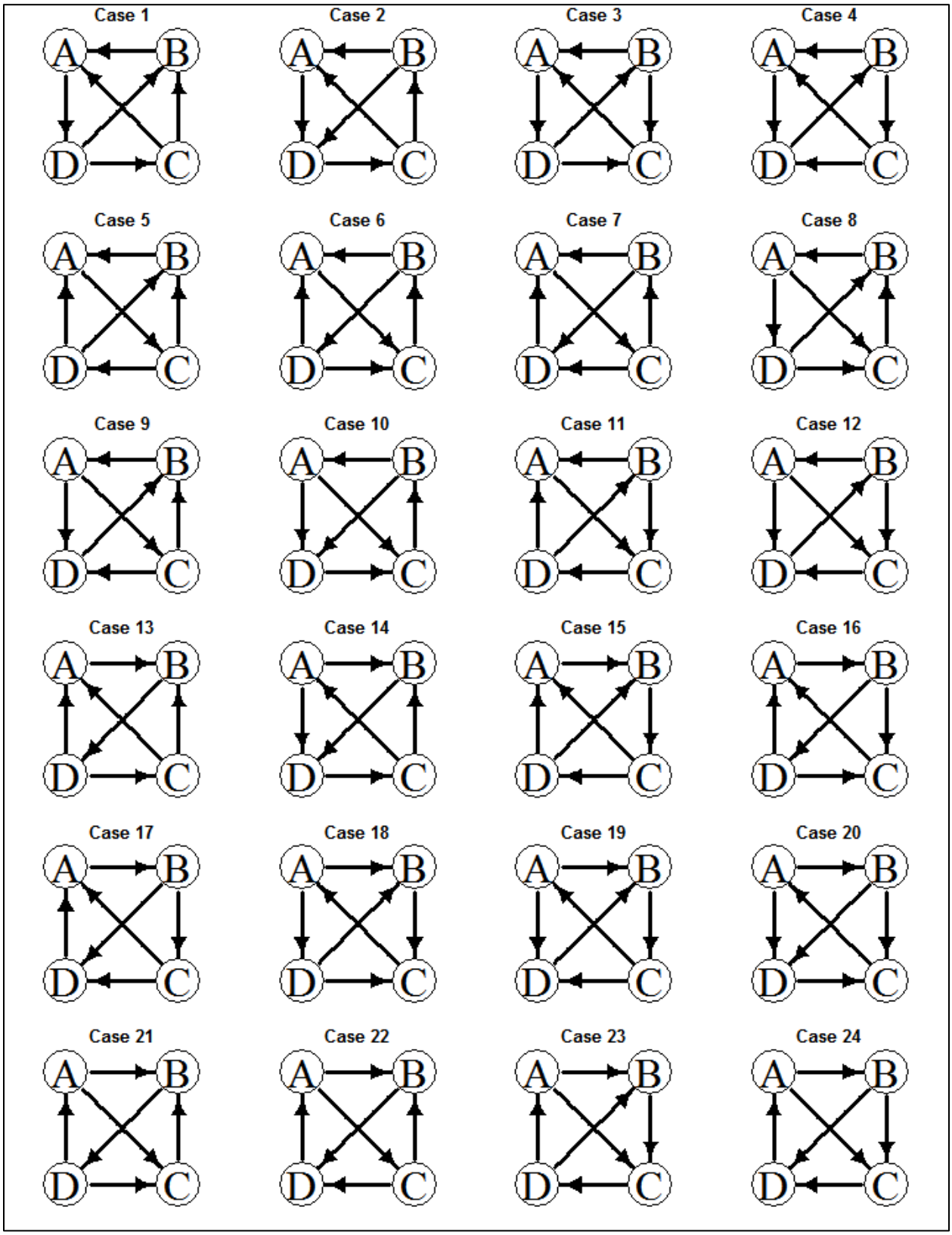
method with a branch search algorithm. They also attempt to generalize methods which summarize paired comparisons as rankings. Maas, et. al (1995) incorporate the strength of preferences to solve intransitivities. Charon and Hudry (2006) (2007) classify all these methods and introduce a branch-and-bound algorithm to solve the linear ordering problem for weighted tournaments. Koczkodaj et. al (2015) and Koczkodaj and Szybowski (2016) suggest reducing inconsistency in the pairwise comparison matrix in order to use conventional models.

Siffert (2005) disagrees with inconsistency so much that he removes all inconsistencies in his Beatpaths model. He starts by removing all pairs of inconsistent dyads, followed by all circular triads, further followed by all circular tetrads, and so on until all Hamilton circuits have been removed from the tournament. The remaining edges in the graph have no inconsistencies; however, teams are often tied.

2.1.7. Missing Values and Circular Triads

Jiang et al. (2011) give an interesting example on p. 217 where an incomplete graph has no circular triads, but does have a circular polyad of size 6: $A \to B \to C \to D \to E \to F \to A$. However, we know that if a complete graph has a circular polyad of size k, then it must contain at least $k - 2$ circular triads. In this example, the reader can see that the incomplete graph would be locally inconsistent (i.e., have at least one circular triad) if it were completed. This can be seen by construction. First, add an edge on (B, E). Now, if $B \to E$, then we have the circular triad $A \to B \to E \to A$. If $E \to B$, then we get a circular tetrad in $B \to C \to D \to E \to B$. From Figure 2-2: All Tetrads Containing Two Circular Triads ($d = 2$) on the next page, we can see that this must contain two circular triads.

Figure 2-2: All Tetrads Containing Two Circular Triads ($d = 2$)



There is a reasonable amount of literature describing how to address tournaments with missing edges. For example, Beineke and Harary (2001) study digraphs on n nodes and $m \leq \binom{n}{2}$ edges to find the maximum number of circular triads in those graphs, reaching $m < 33$. Shizuka and McDonald (2012) randomly generate the missing preferences and give measures of intransitivity. They repeat this procedure 10,000 times and explore the distributions of the measures of intransitivity. Bozóki et al. (2013) study both missing preferences and fixing inconsistent matrices to achieve a transitive order. Bauer (1978) begins work on a method to count circular triads when not every comparison has been made.

2.1.8. Related Kendall Work

Kendall introduced non-parametric correlation coefficient based on ranks commonly called Kendall's τ (1938) (1945). The method used here primes the reader for a counting mechanism we encounter in Chapter 6. Consider the toy example in Figure 2-3, where two judges X and Y have ranked items A, B, C, and D. Without loss of generality, X is in ascending order or the items are relabeled to match X's order.

Figure 2-3: Kendall's Rank Correlation Coefficient Toy Example

	A	B	C	D
X	1	2	3	4
Y	2	4	3	1

A score, s, can be calculated by comparing the rankings of two objects A and B by two rankers — X and Y. If they are in the same direction, then the pair A and B has a positive contribution. If they are different, they have a negative contribution. All contributions are weighted equally ($+1$ or -1) and they are then summed over all pairs of objects. However, since we want all scores to be on the same scale (-1 to 1) as Pearson's correlation, r, we divide this score s by the maximum score of $\binom{n}{2} = \frac{n(n-1)}{2}$, which is also the absolute value of the minimum score. This is Kendall's $\tau = \frac{2s}{n(n-1)}$.

Table 2-3: Kendall's Rank Correlation Coefficient Sign Calculation Version

i	j	$\text{sgn}(x_j - x_i) - \text{sgn}(y_j - y_i)$
A	B	$\text{sgn}(3-1) - \text{sgn}(3-2) = 1$
A	C	$\text{sgn}(4-1) - \text{sgn}(1-2) = -1$
A	D	$\text{sgn}(2-1) - \text{sgn}(4-2) = 1$
B	C	$\text{sgn}(4-3) - \text{sgn}(1-3) = -1$
B	D	$\text{sgn}(2-3) - \text{sgn}(4-1) = -1$
C	D	$\text{sgn}(2-4) - \text{sgn}(4-1) = -1$
$s = \sum \left(\text{sgn}(x_j - x_i) - \text{sgn}(y_j - y_i)\right) = -2$		

If one considers the count of the positive and negative scores, P and Q respectively, one observes that $P + Q = \binom{n}{2}$. Then, an alternative expression for τ is

$$\tau = \frac{2(P-Q)}{n(n-1)} = 1 - \frac{4Q}{n(n-1)} = \frac{4P}{n(n-1)} - 1 \qquad (1.4.1)$$

One can calculate P by first sorting the two sets of rankings X and Y by X, ascending. Then, examining only the vector of rankings $Y = \langle y_1, y_2, \ldots y_n \rangle$, one counts the number of ranks to the right of each rank y_i that are greater than y_i. This number is called P. That is,

$$P = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{I}_{[y_j > y_i]},$$

where $i \in (1, \ldots, n-1)$ since $\mathbb{I}_{[y_n > y_n]} = 0$.

$Q = \binom{n}{2} - P$ can be thought of as the minimum number of neighbor swaps that need to be made to obtain the natural order, or it can be thought of as the number of mistakes made in ranking. Consider again the second judge's preference order $Y = \langle 2, 4, 3, 1 \rangle$. To achieve a natural order, the 1 would have to swap three times to the left, giving us $Y^* = \langle 1, 2, 4, 3 \rangle$. Then, 3 and 4 would have to swap places, Y^{**} now matches X, $Q = 4$, and it follows that $P = 2$ and $s = -2$ as in Table 2-3. Kendall calls these swaps "interchanges." These same interchanges inform one version of the algorithm to enumerate the circular triad frequency distribution in Chapter 6.

Kendall and Smith (1940) also developed a pairwise comparisons procedure for two or more judges. Every one of the m judges must make all $\binom{n}{2}$ comparisons. Kendall and Smith develop a measure of agreement, u, for these judges. It goes from -1 to 1 when $m = 2$ and when m gets large, the minimum of u approaches 0 from the negative side. This is because two judges can be in perfect disagreement when making a choice between A and B, but three or more judges cannot all disagree when there are only two choices. Three or more judges, however, could be in perfect agreement, so the upper bound of u remains 1. The formula for u is given in (2.1.8.1), where γ_{ij} is the number of judges who preferred i to j:

$$u = \frac{2\sum_{i,j,i\neq j}\binom{\gamma_{ij}}{2}}{\binom{n}{2}\binom{m}{2}} - 1. \qquad (2.1.8.1)$$

When $m = 2$, u is Kendall's rank correlation coefficient, τ.

$$u = \frac{2\sum_{i,j,i\neq j}\binom{\gamma_{ij}}{2}}{\binom{n}{2}} - 1$$
$$= \frac{(\#\ of\ concordant\ pairs) - (\#\ of\ discordant\ pairs)}{\binom{n}{2}} = \tau \qquad (2.1.8.2)$$

When m is even, the lower bound of u is $-\frac{1}{m-1}$, and while when u is odd, the lower bound is $-\frac{1}{m}$. While building an exact significance test, Kendall and Smith found that calculating the variance of u, which depends on the number of possible preferences, $2^{m\binom{n}{2}}$, to be virtually impossible to enumerate for large values of m or n. An approximation could be accomplished by simulation. Kendall and Smith do, however, derive a χ^2-approximation based on the distribution of u, or rather, $\sum_{i,j,i\neq j}\binom{\gamma_{ij}}{2}$. By fitting the moments of $\sum_{i,j,i\neq j}\binom{\gamma_{ij}}{2}$ to the moments of the χ^2 distribution, they get the test statistic

$$X^2 = \left(\frac{4}{m-2}\right)\sum_{i,j,i\neq j}\binom{\gamma_{ij}}{2} - \frac{1}{2}\binom{n}{2}\binom{m}{2}\cdot\frac{m-3}{m-2} \sim \chi^2\left(\nu = \frac{\binom{n}{2}m(m-1)}{(m-2)^2}\right). \qquad (2.1.8.3)$$

They give exact critical values for $\sum_{i,j,i\neq j}\binom{\gamma_{ij}}{2}$ in their Table III, IV, V, and VI for small values of m and n.

Finally, Kendall and Smith comment that ξ can be large (a judge can be internally consistent), but the judge may not be accurate. When possible, the judge's preferences should be compared to known preferences by Kendall's τ. In their examples, they use the distribution of the m counts of d in reference to the expected value of d to assess whether

the judges were internally consistent. Kendall and Smith also recover the ranks of the n objects from this procedure by sorting the $\sum_i \binom{\gamma_i}{2}$ in descending order, although they make no comment on these ranks or that choice. They may have done this to sort the sum of the adjacency matrices, therefore minimizing the γ_{ij} in the lower triangle. Their calculation of u uses the equivalence in (2.1.8.4) which benefits from small γ_{ij} in the lower triangle —

$$\sum_{i,j,i\neq j} \binom{\gamma_{ij}}{2} = \sum_{i>j} \left(\gamma_{ij}^2 - m\gamma_{ij}\right) + \binom{n}{2}\binom{m}{2}. \qquad (2.1.8.4)$$

Basak (1993) and Raghavachari (2004) attempt to improve Kendall and Smith's m judge procedure.

2.1.9. The Favorability of Intransitivity

One of the earliest hypothetical discussions giving validity to intransitivity in preferences or strategies comes from von Neumann's version of game theory; for example, see von Neumann and Morgenstern (1944). Social choice theory, as founded in Arrow (1950), (1951) and Bergson (1954) and further developed in Kemeny (1962), Arrow (1974), Sen (1970) (1999), Young and Levenglick (1978), Young (1988), Heiser and D'Ambrosio (2013), and Nurmi (2014), involves the aggregation of voters' candidate preference orders. Since this involves rank aggregation, it is generally out of scope for our work. However, since social choice theory generally involves assuming individual voters have transitive preferences, it follows that researchers in and around the field question the validity of this assumption.

For example, May (1954) openly supports intransitive preferences. He describes an experiment in which 62 college students were asked to choose a hypothetical marriage partner A, B, or C based on three attributes alone — in intelligence they ranked $A \to B \to C$,

in appearance they ranked $B \rightarrow C \rightarrow A$, and in wealth they ranked $C \rightarrow A \rightarrow B$. Every hypothetical marriage partner was described as least acceptable in all the attributes. If these three attributes were equally weighted in importance and there were three voters who each valued a different attribute most, then this is exactly the simplest case of Condorcet's Voting Paradox (see, for example, Young and Levenglick (1978)). Participants were repeatedly faced with unlabeled alternatives A, B, and C a pair at a time until the experimenter was certain the student meant the choices that they had made. As a result, $A \rightarrow B$ $39/62$ times, $B \rightarrow C$ $57/62$ times, and $C \rightarrow A$ $33/62$ times, and so the aggregated preferences were intransitive. Additionally, $17/62$ individual preference orders were intransitive. Here, May essentially argues that unless one preference criterion dominates, intransitive relationships *may* appear.

May (1954) also gives examples of non-transitive relations in game theory, including examples of combat superiority:

Battleship → Destroyer → Submarine → Battleship,

Mongoose → Cobra → Cat → Mongoose, and

Tank → Machine Gunner → Bazookaman → Tank.

He also cites McCulloch's (1948) experiment in which rats are starved and are shown to prefer food to sex, sex to pain avoidance, and pain avoidance to food. May says, "[…] the question is no longer 'Are preferences transitive?' but rather 'Under what conditions does transitivity fail?'"

Meanwhile, Tversky (1969) lists seven studies, including May (1954), that "failed to detect any significant violation of WST" (although it seems to this author that May's aggregated preferences demonstrated intransitivity, while the individual preferences

provide weak evidence of intransitivity). Nevertheless, Tversky seeks to create situations under which the condition of transitivity fails and he succeeds, to an extent. Tversky's paper is often cited as giving validity to the notion that intransitivity is sometimes appropriate. See also Krantz et al. (1971), Fishburn (1984), Suppes et al. (1989), Luce et al. (1990), Gehrlein (1989), Laird and Schamp (2009), Linares (2009), Regenwetter et al. (2011), and Elias et al. (2012) for more examples of intransitive experiments and games.

Koczkodaj and Wajch (2015) give a logical, philosophical, and mathematical take on intransitivity in paired comparisons. To the notions of returning an inconsistent adjacency matrix to a transitive order (as discussed in Section 2.1.6) they say,

> *"Inconsistent assessments are considered as inaccurate but, after their approximation, they may reflect values that are useful for us. It must be stressed that there is no way to find the ideal values for the inconsistent input. In practice, every consistent approximation of an inconsistent [preference] matrix M differs from M and has an error. The ideal approximation, without an error, is only a product of our imagination."*

Meanwhile, there are clearly times when transitivity in measurements should absolutely be assumed or maintained as an ideal. When we guess physical measurements like weight, length, or volume, there should be no question that inconsistent errors need to be adjusted. Consider, for example, a set of medical studies comparing a pair of treatments at a time, where a treatment is found to have a significant effect, say $\hat{\delta}_{ij} = d_{ij} = x_i - x_j$, relative to the other treatment in each study. We should expect that if $d_{12} = 5$, $d_{23} = 3$, and $d_{34} = 6$ that $d_{13} \approx d_{12} + d_{23} = 8$ and $d_{14} \approx d_{12} + d_{23} + d_{34} = 14$. This is the exact kind of

"network meta-analysis" performed in Higgins and Whitehead (1996), whose methodology was later described in Schmid and Mengersen (2013), and Mengersen and Schmid (2013). This network meta-analysis methodology validly assumes that the true effects δ_{ij} should be transitive and uses the d_{ij} in the aggregate to produce a scale, not unlike Kaiser and Serlin (1978) do (as we see in Section 2.3.2).

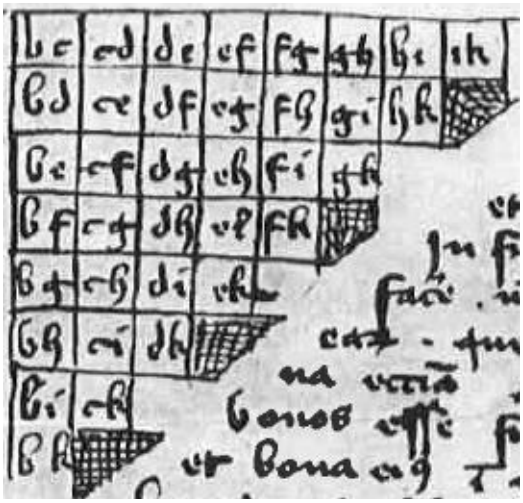
2.2. The Obscured History of Paired Comparisons

With many research topics, who first discovered the topic or made a particular contribution is often not known until later. Consider, for example, the Binomial Theorem, often attributed to Blaise Pascal. However, as Coolidge (1949) demonstrates, Pascal is not the first. Instead, the earliest documented work is from Zhu Shijie in 1303, but even he claims his triangle diagram is based on another Chinese mathematician's work in 1050.

Similarly, Thurstone (1927) was thought to be [one of] the earliest examples of a paired comparisons methodology. However, it should be noted that French mathematician Borda (1781) described a method of voting on one pair of candidates at a time, although he dismissed this notion almost immediately after proposing it in favor of the voting procedure he is now famous for (and this method pre-dates him, too). Even then, a method of voting on one pair of candidates at a time pre-dates even Borda, as noted in McLean and Lorrey (2007). Here, we learn that Majorcan and Franciscan tertiary Ramon Llull developed a method for selecting a pope for the Catholic Church in Llull (1283), (1299) . In this method with n cardinals who are candidates for pope, one pair of cardinals would leave the room at a time. The other $n - 2$ cardinals would cast votes for their cardinal of

choice outside of the room. This process was repeated for all $\binom{n}{2}$ pairs of cardinals, at which point, the cardinal with the most votes overall would be selected as the new pope. The votes were recorded in a preference matrix as in Figure 2-4.

Figure 2-4: Llull's Preference Matrix from Ars Electionis (An Electoral System) (1299)



While Bradley and Terry (1952) were believed to be the first to implement their method, Luce (1959) independently developed the method and it is often called the Bradley-Terry-Luce model out of courtesy. However, even the work of Bradley, Terry, and Luce was anticipated by German mathematician Zermelo (1929). Additionally, the Kendall-Wei method (Wei (1952) and Kendall (1955)) was anticipated by German mathematician Landau (1895) (1914) and the matrix theory on which the Kendall-Wei method is based, and indeed all spectral methods (Section 2.5.2), was proven by Perron (1907) and Frobenius (1908), (1909), (1912).

2.3. Thurstone's Model

In 1927, a psychologist named L. L. Thurstone published his *Law of Comparative Judgment* through which he wished to identify a scale for stimuli, qualitative judgments, opinions, intelligence, handwriting or drawing quality, or anything else which is physically unmeasurable but plausibly on a single continuum. A subject would be prompted by pairs of stimuli, X_i and X_j, and would then make a comparison — either $X_i > X_j$ or $X_i < X_j$. This would be repeated many times, potentially with multiple subjects, for all $\binom{n}{2}$ pairs of n stimuli so that an empirical proportion that the first stimulus was greater than the second, $\hat{p}_{i>j}$, could be calculated. From this information, he could solve for location parameter S_i of each stimulus's distribution as well as its variance σ_i^2 using what he called his law of comparative judgment, but is the statistical model in (2.3.1):

$$S_i - S_j = b_{ij} \cdot \sqrt{\sigma_i^2 + \sigma_j^2 - 2r\sigma_i\sigma_j} \tag{2.3.1}$$

Here, $X_i \sim (S_i, \sigma_i^2)$ and $r = corr(X_i, X_j)\ \forall i \neq j$. b_{ij} is then the number of standard deviations X_i and X_j are apart. He found that b_{ij} fit best when the X_i were assumed to be normally distributed as opposed to uniformly distributed. Thus, b_{ij} was a z-score corresponding to $\hat{p}_{i>j}$ as a probability in a normal area table since (2.2.1) can be re-expressed as in (2.3.2):

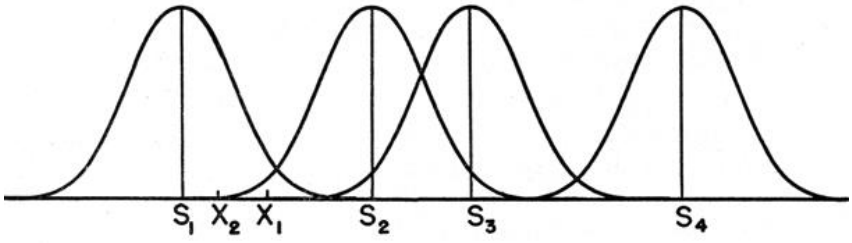
$$z = b_{ij} = \frac{(S_i - S_j) - 0}{\sqrt{\sigma_i^2 + \sigma_j^2 - 2r\sigma_i\sigma_j}} = \frac{(x_i - x_j) - E[X_i - X_j]}{SD[X_i - X_j]} \tag{2.3.2}$$

This led to a system of $\frac{1}{2}n(n-1)$ equations to solve for $2n-1$ values (he let $S_1 = 0$). In the literature, it appears that the set of simplifying assumptions he called "Case V" were the most appealing to later researchers, including Horst (1932). Thurstone let $\sigma_i = 1\ \forall i$ and set

$r = 0$. This simplified the calculations considerably as $\sqrt{\sigma_i^2 + \sigma_j^2 - 2r\sigma_i\sigma_j} = \sqrt{2}$ in (2.3.2) and thus a system of only $n - 1$ equations for the estimates of S_i are needed. His original study creates a scale of on which to compare the severity of crimes. Thurstone (1927b), (1931) explain his estimation procedure in greater depth.

2.3.1. Mosteller Improves the Thurstonian Model

Mosteller (1951) noted that it was not necessary to assume $r = 0$, just that $\sqrt{\sigma_i^2 + \sigma_j^2 - 2r\sigma_i\sigma_j} = c = 1$ by constant correlation and equal variances. Mosteller (p. 4) made the crucial point that in order to get an idea of how close S_1 and S_2 were, for example, there would have to be times where $X_2 < X_1$. That is, the method works best when there are a minority of non-transitive paired observations. He illustrated this point in the following image (p. 4):



But Mosteller's most notable contribution to the Thurstonian method of paired comparisons was his least squares solution. Letting $D_{ij} = S_i - S_j$, the $\widehat{D}_{ij}$ are then estimates of D_{ij}. As in Thurstone (1927), the $\widehat{D}_{ij} = b_{ij} \cdot c = b_{ij}$. He also makes the important distinction that while the true location differences need to be transitive in this framework,

$$D_{ij} + D_{jk} = (S_i - S_j) + (S_j - S_k) = D_{ik} \qquad (2.3.1.1)$$

the observed differences $\widehat{D}_{ij}$ do not need to maintain transitivity as it is almost never true

that $\widehat{D}_{ij} + \widehat{D}_{jk} = \widehat{D}_{ik}$. Then, we seek to minimize Q in (2.3.1.2) with respect to $\hat{S}_i$:

$$Q = \sum_{i,j}\left[\widehat{D}_{ij} - \left(\hat{S}_i - \hat{S}_j\right)\right]^2 \tag{2.3.1.2}$$

While Horst (1932) gave a solution by setting $\bar{\hat{S}} = 0$, Mosteller (1951) solved this by setting $\hat{S}_1 = 0$ as Thurstone (1927) had done before him. Thus, this gives a solution for the other -1 $\hat{S}_i$:

$$\hat{S}_i = \frac{1}{n}\sum_{\mathrm{j}=1}^{\mathrm{n}}\left(\widehat{D}_{j1} - \widehat{D}_{ji}\right) \tag{2.3.1.3}$$

Mosteller (1951) lamented that large values of $\widehat{D}_{ij}$ occur when $\hat{p}_{i>j}$ is close to 0 or 1 and recommends the ad hoc fix of bounding $\left|\widehat{D}_{ij}\right|$ by 2.

Gulliksen (1956) also gives an ad hoc correction which includes converting all $\hat{p}_{i>j} = \frac{0}{m}$ to $\frac{0.5}{m}$ and $\hat{p}_{i>j} = \frac{m}{m}$ to $\frac{m-0.5}{m}$, where m is the number of times X_i and X_j have been compared. Gulliksen also gives a solution for the incomplete case where not all $\binom{n}{2}$ comparisons are made. However, since he sets $\hat{S}_1 = 0$, his solution does not match Kaiser and Serlin's solution. His solution is equivalent to Mosteller's (1951) solution in the complete case. Finally, Schönemann (1970) showed that Gulliksen's solution for the incomplete case was unique when every row in the skew-symmetric matrix, $\boldsymbol{D} = \left(\widehat{D}_{ij}\right)$, has nonzero entries in at least half of the non-diagonal cells.

2.3.2. Kaiser and Serlin and Extensions

In the literature discussed so far, the Thurstone-Mosteller model has only been used in cases where proportions of comparisons are mapped to normal deviates to build a scale of measurement. Kaiser and Serlin (1978) examine the Thurstonian model without the assumption that the D_{ij} are normally distributed, although they still build a scale of

measurement with their D_{ij}, which is defined in a college football game as $+1$, 0, or -1 for a win, tie, or loss, respectively. The resulting scale, unadjusted, has no practical interpretation other than to rank the football teams in the group and give a sense of quality. If they had used point differential as proposed, the resulting ratings of the teams could be used to predict point differential in future games.

As in (2.3.1.2), they seek to minimize $Q = \sum_j^{n_i} \left(D_{ij} - \left(x_i - x_j\right)\right)^2$ with respect to x_i, where n_i is the number of games team i has played. Here, x_i plays the role of the location parameter, previously S_i. $D_{ij} = 0$ in games not played, and Q is not minimized over these pairs of i and j. They let $D_i = \sum_j^{n_i} D_{ij}$, which can be interpreted as team i's wins minus their losses. This is colloquially called "games over .500," a common phrase in sports reporting. They also choose their solutions for x_i such that $\sum_i x_i = 0$, as in Horst (1932). This makes the x_i interpretable as "team i is better than average" when $x_i > 0$. Then, this gives a system of linear equations which can be expressed as

$$(n_i + 1)x_i + \sum_j^{n-1-n_i} x_j = D_i \qquad (2.3.2.1)$$

In the summation in (2.3.2.1), only the teams j that did not play team i are summed over, of which there are $n - 1 - n_i$. Using team $i = 2$ in their example, Army, this is read as

$$8x_2 + (x_5 + x_7 + x_9 + x_{12}) = -1 \qquad (2.3.2.2)$$

since Army did not play teams $j = 5,7,9$, and 12. This suggests that Kaiser and Serlin's method implicitly considers strength of schedule. If $\sum_j^{n-1-n_i} x_j > 0$, then this means that the teams i hasn't played are better than average, and thus the teams i has played are worse than average. Team i is penalized in this case, while they benefit when $\sum_j^{n-1-n_i} x_j < 0$. We can re-express (2.3.2.2) as

$$0x_1 + 8x_2 + 0x_3 + 0x_4 + 1x_5 + 0x_6 + 1x_7 + 0x_8 + 1x_9 + 0x_{10} + 0x_{11} + 1x_{12} = -1 \quad (2.3.2.3)$$

From (2.3.2.3), it is easy to see that the coefficients k_{ij} on the x_{ij} are 1, 0, or $(n_i + 1)$ when i hasn't played j, i has played j, or $i = j$, respectively. The matrix of these coefficients, $\boldsymbol{K}$, multiplied by the vector of team ratings, x, are equal to the vector of $D = (D_i)$, giving $\boldsymbol{K}x = D$. The solution for the team ratings is then $x = \boldsymbol{K}^{-1}D$.

Kaiser and Serlin give a necessary and sufficient condition for a solution. Intuitively, every team can be compared directly or indirectly to every other team, even if the shortest list of comparisons goes through $n - 2$ teams. In a four-team example, if team i and j play, j and k play, and k and l play and there are no other games among them, then team i could be compared to team l through teams j and k. After all, this is the reasoning behind a single-elimination tournament (the first and third games listed above would be the first round, which would give a championship game between j and k, without loss of generality). They prove that this intuitive condition is equivalent to $\boldsymbol{K}$ being nonsingular. Alternatively, since $\boldsymbol{K} = f(\boldsymbol{D})$, an equivalent condition for the existence of a solution is that the skew-symmetric matrix $\boldsymbol{D}$ cannot be row operated into the form $\begin{bmatrix} \boldsymbol{D}_{11} & \boldsymbol{D}_{12} \\ \boldsymbol{D}_{21} & \boldsymbol{D}_{22} \end{bmatrix}$ where $\boldsymbol{D}_{12} = \boldsymbol{D}_{21} = \boldsymbol{0}$. This is comparable to equation [9] in Kaiser and Serlin, since $k_{ij} = 1 \Rightarrow D_{ij} = 0$. In their example, they have 24 of a possible 66 games, well under the 36 games Schönemann required.

Checking $\boldsymbol{K}$ or $\boldsymbol{D}$ as above would be an important condition for a schedule-maker (which is often automated to some extent). It would also be useful for an analyst (or fan) to decide at which point in a season's schedule they would be able to justify making a definitive statement about which team is the best in the league. John Hollinger, formerly of

ESPN, would not publish his team ratings until every team in the NBA had played at least 20 games. With 30 teams in the league that play each other two to four times a season, this would likely meet Kaiser and Serlin's condition.

Kaiser and Serlin develop a formula for R^2,

$$R^2 = \frac{\sum_i x_i D_i}{\sum_{i>j} D_{ij}^2} \tag{2.3.2.4}$$

but by their own admission, this does not give quite the same meaning as in other least squares settings. Finally, to make their ratings more interpretable, they rescale the x_i to closely match the score differential, p_{ij}, or point spread (points score by team i minus points scored by team j). They minimize g with respect to c in (2.3.2.5):

$$g(c) = \sum_i \sum_j \left(p_{ij} - c\left(x_i - x_j\right)\right)^2 \tag{2.3.2.5}$$

This gave $c = 22.7$. They added 100 to the cx_i to give ratings similar to an index (that also are not as *negative* to readers as the unscaled cx_i are).

Kaiser and Serlin debate whether p_{ij} should have been used in lieu of their D_{ij}. As we will see later in Barrow et al. (2013), it is almost always the case that point differential is a better predictor of team success than wins and losses.

Kaiser and Serlin (1978) extended the least squares solution for the Thurstonian model to cases where there are a small number of paired comparisons. They identify the minimum number of comparisons to be made as well as introduce a theorem which identifies under what conditions the least squares solution to the Thurstonian model can be used. This technique, they claim, suggests that Schönemann's lower bound (1970) of half of all $\binom{n}{2}$ comparisons is too conservative. They then apply their method to the games played between 12 major college football teams as a demonstration.

2.4. Bradley-Terry Model

The model we introduce here is from the classic paper by Bradley and Terry (1952). Luce independently developed the same model a few years later and so it is often referred to as the Bradley-Terry-Luce Model. The earliest work on this model, however, is believed to be that of Zermelo (1929), a German; although his work was not discovered until much later. We refer to this model as the Bradley-Terry model or BTM. Many of the methods in Chapter 2 can be related to the BTM. However, the BTM only uses wins and losses and classically gives transitive ratings and does not take the strength of victory into account.

Suppose we have n teams with true ratings $p_1, p_2, \dots, p_n$ where $p_i \geq 0 \; \forall i$ and $\sum p_i = 1$. Bradley and Terry claim that when teams i and j play each other (or as they say, "appear in the same block"), then they assume (2.4.1). They also denote r_{ijk} as the ranking of the i^{th} team in the k^{th} game against team j, and it follows that $r_{ijk} = 3 - r_{jik}$. They let n_{ij} be constant for all i, j, that is, every pair of teams plays the same number of times.

$$p_{ij} = \frac{p_i}{p_i + p_j} \tag{2.4.1}$$

They then give a likelihood function (2.4.2) which assumes the p_{ij} are independent $\forall i, j$.

$$L = \prod_i p_i^{2n_{ij}(n-1) - \sum_{i \neq j} \sum_k r_{ijk}} \prod_{i<j} (p_i + p_j)^{n_{ij}} \tag{2.4.2}$$

And a hypothesis test for equality of strength in (2.4.3):

$$H_0 : p_i = \frac{1}{n} \; \forall i, H_1 : p_i \neq \frac{1}{n} \; \exists i. \tag{2.4.3}$$

They give maximum likelihood estimates for the p_i on page 326. We note that the model is more commonly described as in (2.4.4), where β_i is the strength of team i and thus $\beta_i - \beta_j$ is the amount team i is better than team j.

$$\text{logit}(p_{ij}) = \log\left(\frac{p_i}{p_j}\right) = \beta_i - \beta_j \tag{2.4.4}$$

Also of note is their table, *The generation of treatment sums of ranks and probabilities for three treatments and two repetitions*. Here, if we have that $B \to A$, $C \to A$, and $C \to B$, then we have a transitive triple, $C \to B \to A$. Bradley and Terry would express this as $r_{ab1} = r_{ac1} = r_{bc1} = 2$ and $r_{ba1} = r_{ca1} = r_{bc1} = 1$. This would give rank sums for the first round of $r_{c1} = 2$, $r_{b1} = 3$, and $r_{a1} = 4$, where $r_i = \sum_j r_{ij1}$. Then $r_1 = (2, 3, 4)$ is the same as the score vector (or alpha vector) in the Kendall and Smith (1940) setting, $\alpha = (2, 1, 0)$. We will explore score vectors more in Chapter 6. Bradley and Terry's table resembles that of David (1959).

2.4.1. Extensions and Improvements on the Bradley-Terry Model

The Bradley-Terry model is incredibly popular. Many have made attempts to improve it. Dykstra (1960) introduces an unbalanced version of the BTM. Rao and Kupper (1967) introduce a way to deal with ties, which they define as indistinguishable results or comparisons, i.e., $|\beta_i - \beta_j| \leq \varepsilon$ for some small ε. Leonard (1977) introduces an early Bayesian approach to the BTM, where the priors on the β_i might depend on one another or depend on some explanatory variables. Simons and Yao (1999) prove a maximum likelihood estimator for the β_i is asymptotically normal. Hunter (2004) gives an iterative minorization-maximization algorithm to estimate the β_i which converge asymptotically to the maximum likelihood estimates for the β_i. Adams (2005) compares and contrasts

Bayesian and non-parametric approaches to the BTM. Usami (2010) gives a BTM for multiple judges where the judges may be externally inconsistent. Cattelan et al. (2013) gives a dynamic BTM with an exponentially weighted moving average process for the β_i. All of these models produce transitive ratings, β_i.

Mallows (1957) introduces a variation on the BTM based on the differences in ranks between teams i and j, labeled k. θ and ϕ are parameters which are $\theta = \phi = 1$ under the null hypothesis, and the model is

$$p_{ij} = \frac{1}{2} + \frac{1}{2}\tanh(k \log \theta + \log \phi) \tag{2.4.4.1}$$

2.4.2. Connecting Thurstone-Mosteller and Bradley-Terry Models

The Thurstone-Mosteller and Bradley-Terry model are remarkably similar. Here, Tsukida and Gupta (2011) note (2.4.2.1), or that Thurstone uses the probit link function instead of the logit link function as in the BTM.

$$\Phi^{-1}(\hat{p}_{ij}) = \widehat{\beta}_i - \widehat{\beta}_j \tag{2.4.2.1}$$

Then, the relationship between the BTM and the TMM is nicely summarized with Figure 4 in Tsukida and Gupta (2011) and below.

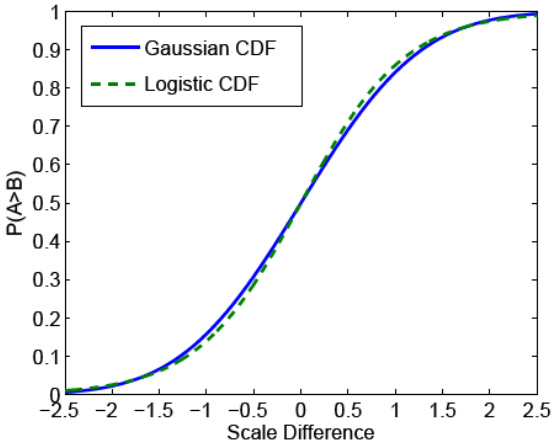


Figure 4: Gaussian vs Logistic CDF

Noether (1960) gives some early connections between the BTM and TMM. Andrews and David (1990) extend and continue Gulliksen's (1956) ideas on using incomplete or unbalanced data, but in a nonparametric fashion. Stern (1990)and Critchlow and Fligner (1991) introduce general linear model frameworks that unify the BTM and TMM.

2.5. Other Transitive Ratings Models

Here, we discuss other popular models whose goal is to estimate ratings with the assumption that a strict order exists. In the absence of any written discussion of these methods, see Barrow et al. (2013) for a comprehensive look within the application of sport rankings. All the methods they explore are fully transitive. They show empirically that average margin of victory[3] is more predictive of future success than winning percentage. Their survey builds on the survey of work in Chartier et al. (2011).

Barrow et al. (2013) discuss winning percentage, rating percentage index, least squares pairwise comparisons (i.e., Thurstone-Mosteller, Bradley-Terry, and HodgeRank), maximum posterior, Keener's direct method, PageRank, random walker, and Elo's method. These ranking methods are applied to six different sports leagues (National Basketball Association (NBA), Major League Baseball, National League, American League, NCAA Men's Basketball, and NCAA Football) on a season by season basis.

In the NBA, the least squares pairwise comparison, maximum posterior, Keener's direct, and the random walker methods give the best predictions. The Elo and least square pairwise comparison methods are noted to have a probabilistic interpretation. This is an important feature for consumers of sports rankings.

[3] Here, margin of victory is the point differential divided by the total number of points scored in the game.

Barrow et al. performed this analysis at the end of each season. However, most fans and gamblers alike want to know how their team is doing midseason. An interesting question then, might be to identify at which point in the season point differential is a better predictor of success than wins and losses. Barrow et al. inferred from the NCAA men's Basketball and NCAA Football data, which have substantially shorter seasons than the other leagues, that win/loss is more predictive during shorter seasons. Alternatively, this discrepancy could be tied to the youth of these athletes — they are collegiate athletes with few to no veteran athletes surrounding them, and the roster turnover at the collegiate level is high.

2.5.1. HodgeRank and the State of the Art

Jiang, et al., (2011) offer an improvement on the Thurstonian model seen in Kaiser and Serlin (1978) which incorporates the idea of circular triads and circular polyads in Kendall and Smith (1940). HodgeRank can be used in cases with great imbalance, including missing pairwise comparisons, solo pairwise comparisons, and many pairwise comparisons. The underlying scale values of the objects are recovered through least squares minimization. Hodge Theory is used to decompose the aggregated rank information into measures of consistency — local inconsistency and global inconsistency. HodgeRank has the drawback of producing the Borda winner, which is not necessarily the Condorcet winner when it exists. However, unlike Kemeny optimization (1959), it is not NP-hard but is instead $O(n)$ due to the least squares solution. HodgeRank draws on results from graph theory, linear algebra, vector calculus, topology, ranking, decision theory, financial economics, machine learning, and social choice theory. We will focus on the graph theory, ranking, and least squares aspects in our examination of HodgeRank.

In the diction of Saaty (1980) and other decision scientists, the purpose of HodgeRank is to rank a set of n alternatives by m voters. Each of the m voters only needs to rank, pairwise compare, or score a relatively small number of alternatives. Judgments by the α^{th} voter between the i^{th} and j^{th} objects in the α^{th} matrix Y^{α} are recorded as Y_{ij}^{α}. They let $Y_{ij}^{\alpha} = -Y_{ji}^{\alpha}$, producing a skew-symmetric matrix. If the alternatives are scored (for example, on the Likert scale), they are converted to ranks. If the alternatives are ranked, then if $i \to j$, $Y_{ij}^{\alpha} = 1$. They expect alternatives that are closely ranked to display inconsistency, but alternatives that are ranked far apart to display inconsistency infrequently.

They define the weight function:

$$w_{ij}^{\alpha} = \begin{cases} 1 & \text{if } \alpha \text{ made a pairwise comparison for } \{i,j\} \\ 0 & \text{did not make comparison} \end{cases}. \qquad (2.5.1.1)$$

However, w_{ij}^{α} may also be the number of pairwise comparisons between i and j voter α. W^{α} is symmetric. The true scale values are denoted s_i and their true differences are $X_{ij} = s_j - s_i$ in the $\mathcal{M}_G$ model class as Kaiser and Serlin (1978) had them. The s_i are used to extract a global ranking which is Borda-optimal. The X_{ij}, and in turn the s_i, are estimated by minimizing $X \in \mathcal{M}_G$ in (2.5.1.2):

$$\sum_{\alpha,i,j} w_{ij}^{\alpha}\left(X_{ij} - Y_{ij}^{\alpha}\right)^2 \qquad (2.5.1.2)$$

The $\mathcal{M}_K$ model class is an alternative to $\mathcal{M}_G$ that produces analogous results to the binary adjacency matrix in Kendall and Smith (1940), although the $Y_{ij}^{\alpha} \in \{-1,1\}$ when $w_{ij}^{\alpha} > 0$ and $Y_{ij}^{\alpha} = 0$ otherwise. In this special case, $X_{ij} = \text{sign}\left(s_j - s_i\right)$ and the resulting global ranking of the s_i gives the Kemeny-optimal solution. This has the benefit of producing the Condorcet winner if it exists. They define the pairwise comparison graph

$G = (V, E)$ as an undirected graph whose vertex set is $V = \{1, \dots, n\}$ and whose edge set is (2.5.1.3). Here, edges can have weights $w_{ij} = \sum_\alpha w_{ij}^\alpha$:

$$E = \left\{\{i, j\} \in V\text{x}V \text{ such that } \sum_\alpha w_{ij}^\alpha > 0\right\} \tag{2.5.1.3}$$

Now, they given an alternative to (2.5.1.2) to minimize $X \in \mathcal{M}_G$ in (2.5.1.4) where $\bar{Y}_{ij} = \frac{\sum_\alpha w_{ij}^\alpha Y_{ij}^\alpha}{\sum_\alpha w_{ij}^\alpha}$:

$$\sum_{\{i,j\}\in E} w_{ij}\left(X_{ij} - \bar{Y}_{ij}\right)^2 \tag{2.5.1.4}$$

If X_{ij} is a pairwise ranking (comparison), then Φ_{ijk} is a triplewise ranking (comparison). The former is the edge flow and the latter is the triangular flow. The triangular flow carries the same information as a subgraph on vertices i, j, and k and requires the subgraph be complete — K_3. Calculating the curl is analogous to counting the number of circular triads in the ordinal setting. However, the curl can also be calculated in the cardinal setting. Here, they cite a formula comparable to (6.1.1), which we derived from Kendall and Smith's work (1940) in Section 2.1.1. Formula 6.1.1 gives the fastest method to count the number of circular triads in a complete graph. (2.5.1.5) may then merit further investigation.

$$d = \frac{n}{24}(n^2 - 1) - \frac{1}{8}\sum_i\left(\sum_j X_{ij}\right)^2 \tag{2.5.1.5}$$

HodgeRank goes beyond the scope of our work as it allows for rank aggregation and multiple judges. However, it is the state of the art in paired comparisons and is a frequently cited paper. See, for example, Csató (2015a), (2015b), (2016), Csató and Rónyai (2016), and Rajkumar (2015).

2.5.2. Spectral Ranking Methods

Wei's dissertation (1952), which must be read via microfiche, introduces a method for paired comparisons based on the Perron-Frobenius Theorem. Kendall (1955) gave Wei credit for the model (see pages 92–108 in Wei (1952)), but without Kendall, the method might have languished in obscurity as Wei's dissertation was convoluted and Kendall greatly simplified the model. It is widely called the Kendall-Wei model. Their method requires an adjacency matrix (which they call P) with non-negative entries and with positive diagonal elements (the diagonal elements are usually 0.5 to indicate a tie between object A_i and A_i). Then by the Perron-Frobenius theorem, there exists a unique simple positive root of the equation $|A - \lambda I| = 0$. Then, let λ_1 be the largest root and Y_1 the corresponding [ratings] vector, where we have the relation $PY = \Lambda Y$. Then, by using a successive powering, we get $P^k Y_1 = \lambda_1^k Y_1$. We converge on a solution for Y_1 as $k \to \infty$. More intuitively, we can think of this process as weighting each object gaining the strength of the objects it beats. We demonstrate this with an example where the diagonal entries are 0, the row sums are $\boldsymbol{\alpha}$, and the relative strength vector is $\boldsymbol{u} = \frac{\boldsymbol{\alpha}}{\sum_i \alpha_i}$:

$$P = \begin{matrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{matrix} \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \boldsymbol{\alpha}^{(1)} = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \end{bmatrix} \Rightarrow \boldsymbol{u}^{(1)} = \begin{bmatrix} 0.\overline{3} \\ 0.\overline{3} \\ 0.1\overline{6} \\ 0.1\overline{6} \end{bmatrix}.$$

Then, for every A_i, we assign the number of wins from objects A_i has beaten. So, object A_1 beat A_2 and A_3, for $1(2)$, and $1(1)$ points, respectively, and a total of $\alpha_1 = 1(2) + 1(1) = 3$ points. This gives $\boldsymbol{\alpha}^{(2)} = (3, 2, 3, 3)$ and $\boldsymbol{u}^{(2)} = (0.\overline{27}, 0.\overline{18}, 0.\overline{27}, 0.\overline{27})$. Now, this process is repeated until the relative strength vector converges to $\boldsymbol{u}^{(\infty)} \approx (0.321, 0.283, 0.165, 0.230)$,

which gives the ranking (A_1, A_2, A_4, A_3). This method takes into account the strength of schedule and victories and results in a fully transitive strength vector.

David (1971) is critical of Kendall and Wei's powering matrix method — strong teams losing to weak teams are punished more heavily than weak teams. Instead, he favors other methods, including Slater's (1961) nearest adjoining order, as it is the maximum likelihood weak stochastic order.

Jech (1983) provides a similar method which Stob (1984) identifies. David (1987) provides an update on the Kendall-Wei method for unbalanced data. Keener (1993) is one of the early authors to notice that there are dozens of methods which are similar to the Kendall-Wei method. Later, Vigna (2019) produces an excellent overview of "spectral ranking methods" that arose independently over the last century, and gives Landau (1895) credit for the earliest version of this. Consider as well Redmond (2003). Meanwhile, PageRank by Google founders Brin et al. (1998) falls in this class of spectral ranking methods albeit by perturbation. They label their method a network-based stochastic method which rank webpages. Since then, the literature on PageRank has grown quickly including, for example, Langville and Meyer (2004), Callaghan et al. (2007), Govan (2008), Govan et al. (2009), Sargolzaei and Soleymani (2010), Radicchi (2011), London et al. (2014), Fogel et al. (2014), and Goodman et al. (2015). Some of these methods are time-dynamic.

Vigna (2019) also includes Saaty's Analytic Hierarchy Process (1980) among these spectral ranking methods. The difference is that entries in the paired comparison matrix are ratios, i.e., $x_{ij} = 1/x_{ji}$ instead of strength differences or win counts. Saaty and his legion of business world followers have never heard of paired comparisons. His work continues,

with Takahashi (1990), Nishizawa (1995), Kwiesielewicz and van Uden (2004), Ma (2006), Iida (2009), and Koczkodaj et al. (2016), among others. Koczkodaj et al. (2015) demystifies the connections between paired comparisons and the analytic hierarchy process. Koczkodaj introduces a measure of consistency in the AHP (1993).

2.5.3. Dynamic Ratings Models

Elo (1978) discusses the theory and implementation of his time-dynamic rating system for chess players, which had been in use by the US Chess Federation (USCF) since 1960 and was first published in 1961. The international chess federation, FIDE, began using the system in 1970 to assign international titles. FIDE uses various versions of the model in different settings. That is, there is a ratings method that is used on an annual basis, one on a continuous basis, and one for a round robin (i.e., a simultaneous method), for example. It is, however, the continuous rating formula (2.5.3.1) that is of interest. It requires that some rating scale has been established, with players rated on that scale, and who will continue to be compared to newer, untested players. The rating system has become synonymous with Elo, to the point that most assume "ELO" to be an acronym of some sort.

Elo assumes player performances are normally distributed. Here, a player i's new rating after a competition, $\beta_i{}'$, is given by

$$\beta_i' = \beta_i + K\big(W_{ij} - E\big[W_{ij}\big]\big), \tag{2.5.3.1}$$

where β_i is the pre-competition rating; K is a fixed rating point value for a single competition score, $K \in [10, 32]$; W_{ij} is the actual game score with wins being 1, draws being ½, and losses being 0; $E\big[W_{ij}\big]$ is the expected game score based on β_i and β_j; and β_i' is rounded to the nearest integer by convention. The method assumes that $\beta_i - \beta_j$ is normally

distributed like in the TMM, and so is considered an extension of the TMM. Note that $\sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2} = \sigma_i\sqrt{2}, \sigma_i = \sigma_j \; \forall \; i, j$. Additionally, $\sigma_i \coloneqq 200$, much like $\sigma \coloneqq 100$ for a single SAT exam, which gives $\sigma_{ij} = 200\sqrt{2} \approx 282.84$. Thus, $\beta_i - \beta_j \sim N(\mu = 0, \sigma_{ij} \approx 282.84)$. Then, it follows that

$$E[W_{ij}] = P\left(z < \frac{\beta_i - \beta_j}{\sigma_{ij}}\right) = \Phi\left(\frac{\beta_i - \beta_j}{\sigma_{ij}}\right).$$

Note that K represents the number of points that are at stake in a particular match. The most stable leagues usually have $K = 10$ while cohorts of unknown competitors have $K = 32$. $K(W_{ij} - E[W_{ij}])$, then, are the points that change owners. So, if i beats j, then i collects $K(W_{ij} - E[W_{ij}])$ points while j loses $K(W_{ij} - E[W_{ij}])$ points. Draws are still an increase in β' for the lower rated player and a decrease for a higher rated plater.

Elo (1978) also provides a tier list separated by individual player standard deviation $\sigma_i = 200$.

Table 2-4: Chess Player Tier List by Elo Rating

Rating β_i	**Player Classification**
$\beta_i \geq 2600$	World Championship Contenders
$2400 \leq \beta_i \leq 2599$	Most Grandmasters Most International Masters
$2200 \leq \beta_i \leq 2399$	Most National Masters
⋮	⋮
$1200 \leq \beta_i \leq 1399$	Amateurs Class D/Category 4
$\beta_i \leq 1199$	Novices

Elo (1978) also provides suggestions and solutions to maintain the model. For example, since ratings are zero-sum, there need to be methods to minimize "deflation." Elo applies his ratings to historical data to retrieve "historical Grand Masters" and studies age-related ratings, both of which Silver (2019) incorporates in his CARM-Elo model. Elo also notes that white, the first player to move, has a 57% chance of winning a game, which equates to about a 50-point rating increase. Interestingly, this is not taken into account when calculating $E[W_{ij}]$.

In Mark Glickman's dissertation (1993), he introduces an Elo-like model with $\sigma_i(t)$ called Glicko. The individual player variance decreases over time if the player is active but increases with inactivity. He continues work on time-dynamic models, including, but not limited to, Glickman and Stern (1998), (2016) and Glickman (2001). Glickman has served as the USCF ratings committee chairman since 1992.

Herbrich et al. (2007) introduce a live ratings system for online gaming under Microsoft called TrueSkill, which can take more than two players into account. Due to the scale of online gaming systems (sometimes millions of players in a single game), how to balance calculation speed and degree of approximation is a consideration. Leiva et al. (2008) discuss the role of competition observation timeframe from an animal dominance and social choice perspective a la Appleby (1983) and Tufto (1998). In short, this observation timeframe should be as brief as possible.

2.5.4. Rating and Ranking Models Used in Sports

Plackett (1975) models finishing order. He seeks to predict, for example $P(A_i \rightarrow A_j \rightarrow A_k)$, where the $A_{(\cdot)}$ are racehorses and the finishing order of three particular

horses is relevant. Each horse some probability of winning, where $\sum_{i=1}^{n} p_i = 1$, so he would say $P(A_i \to A_j \to A_k) = p_i \left(\frac{p_j}{1-p_i}\right)\left(\frac{p_k}{1-p_i-p_j}\right)$.

From 1998 to 2013, the college football championship game participants were determined by the Bowl Championship Series (BCS), an ensemble model that took into account two polls and up to six statistical ranking methods. Not all of these methods have been made public, but they include Massey (1997) and Colley (2002), a Kendall-Wei method, as well as a BTM implementation and an Elo method. See Stern et al. (2004) for more on the BCS. Mease (2003) gives a Bayesian penalized maximum likelihood model that outperforms the methods used for the BCS.

For the NCAA, West (2006) introduces an ordinal logistic regression method. The NCAA selection committee is also known for using RPI (2.5.4.1), or the Ratings Percentage Index, where $\hat{p}_i$ is team i's winning percentage, $\hat{o}_i$ is the average winning percentage of all of the teams team i has played, and $\hat{q}_i$ is the average winning percentage of all of the teams team i's opponents have faced. It takes into account the same information the first two iterations of the Kendall-Wei method, albeit in an arbitrary weighting.

$$RPI_i = 0.25\hat{p}_i + 0.5\hat{o}_i + 0.25\hat{q}_i \qquad (2.5.4.1)$$

2.6. Multi-Dimensional Scaling

Multi-Dimensional Scaling (MDS) takes observed pairwise [Euclidean] distance (or similarity) measures d_{ij} between n objects, $(i, j \in [1, 2, \dots, n])$, and uses them to create a K-dimensional mapping by minimizing a loss function. Consider the classic example where the desire is to take flying distances between n American cities and produce a flat, 2-dimensional map of the contiguous United States. There is no perfect map (in non-trivial

sets of cities) or distances δ_{ij} since we cannot represent points on a sphere with a plane. However, we can still produce estimates $\hat{\delta}_{ij}$ by, for example, by minimizing

$$SS = \sum_{ij} \left(\hat{\delta}_{ij} - d_{ij}\right)^2 . \tag{2.6.1}$$

If we were, for example, taking in a round-robin tournament's worth of score differentials Y_{ij}, then we could estimate μ_{ij} with an MDS method, and it would be less restrictive about intransitive outcomes. The larger K, the fewer restrictions there would be on intransitive outcomes.

Kruskal (1978) provides an excellent introduction to MDS. Sjöberg (1968) considers MDS to be an extension to the unidimensional latent scale of Thurstone to K dimensions. Havlena and DeSarbo (1991) and Wedel and DeSarbo (1993) gives example applications of MDS and Stigler (1994) provides interesting data for MDS purposes.

2.7. Tufto, Modeling Intransitivity, and Dominance Hierarchies

Tufto et al. (1998) describes four models which account for intransitivities in paired comparisons data. First, he alludes to Lahti et al. (1994) where each of the comparisons are simply modeled by the empirical proportions in (2.7.1), where X_{ij} is the observed number of successes $i > j$ and n_{ij} is the number of times items i and j are judged.

$$\hat{p}_{ij} = \frac{X_{ij}}{n_{ij}} \tag{2.7.1}$$

However, Lahti et al. do no theoretical examination of this model; it is only used descriptively. Lahti et al. do, however, identify three circular triads in their data.

Tufto et al. also introduce three models with an intransitive interaction term for covariates in a Bradley-Terry framework with the logit choice of link-function. For a

comparison to a transitive model, they introduce (2.7.2), where individuals i and j have trait vectors $< x_i, y_i >$ and fitted constants a, b. Here, $\beta_i = ax_i + by_i$.

$$\text{logit}(p_{ij}) = a(x_i - x_j) + b(y_i - y_j) = (ax_i + by_i) - (ax_j + by_j) = \beta_i - \beta_j \qquad (2.7.2)$$

They modify (2.7.2) with a cross-product interaction $x_i y_j - x_j y_i$ in (2.7.3), whose strength is measured with fitted constant c.

$$\text{logit}(p_{ij}) = a(x_i - x_j) + b(y_i - y_j) + c(x_i y_j - x_j y_i) \qquad (2.7.3)$$

The cross-product interaction term is remarkably similar to the term we introduce in Chapter 4, $r_i s_j - r_j s_i$, except here the traits $\boldsymbol{x}$ (age) and $\boldsymbol{y}$ (sex) are observed, while $\boldsymbol{r}$ and $\boldsymbol{s}$ are not observed. The strength of $\boldsymbol{r}$ and $\boldsymbol{s}$ is accounted for in their fitting. Finally, CRSP in Chapter 4 models observed differences in strength, not win or preference probabilities. In (2.7.3), $x_i y_j - x_j y_i$ is treated as a single observed covariate. This would not be unreasonable in a sports ranking setting if $\boldsymbol{x}$ and $\boldsymbol{y}$ were offensive efficiency and defensive efficiency measures as in Govan (2008); however, these terms would need to be centered at zero.

Tufto et al. also introduce an additional intransitive model in (2.7.4) as an extension of (2.7.2), where d is a fitted constant and i and j may be related. This is likely not applicable in a sports ranking setting as teams do not usually beget other teams in the biblical sense. A similar model is given in Sinervo and Lively (1996).

$$\text{logit}(p_{ij}) = a(x_i - x_j) + b(y_i - y_j) + dr_{ij}, \text{where } r_{ij} = \begin{cases} 1 & i \text{ is } j's \text{ offspring} \\ -1 & j \text{ is } i's \text{ offspring} \\ 0 & \text{otherwise} \end{cases} \qquad (2.7.4)$$

Finally, Tufto et al. informally allude to an extension of (2.7.3) in (2.7.5), where more than two pairs of covariates could introduce intransitivity. Tufto's student Kristiansen formally states (2.7.5) in his thesis, where there are K traits $\boldsymbol{x}_k$, the a_k and a_{uv}

are fitted constants, and $a_{uv} = 0$ if no interaction for trait pair u, v is desired (2011). Kristiansen explores a set of models where $\|a_{uv} \neq 0\| = 2$.

$$\text{logit}(p_{ij}) = \sum_{k=1}^{K} \left(a_k(x_{ik} - x_{jk}) + \sum_{u,v \in (1,K)} a_{uv}(x_{iu}x_{jv} - x_{ju}x_{iv}) \right) \tag{2.7.5}$$

Generally, animal behavior and dominance structures are rife with intransitive relationships. For example, Broom and Cannings (2002) provide an interesting study in which dominance structure is not necessarily based purely on wins and losses, but of strategies of either sharing or competing resources. This is called the Hawk-Dove game. Hawks always takes a dove's resources, doves always split their resources equally, and hawks compete with other hawks for a resource, but at some cost, C. The Hawk-Dove game allows for intransitivities to occur, as hawks are stronger than doves, but hawks may perish if C is large. A difference between the Hawk-Dove game and our setting is that preferences and the outcomes of sports contests are usually not life and death. De Vries (1998) provides a nice overview of models that attempt to return empirically intransitive adjacency matrices to their nearest order and provides a modification on Slater's method.

2.8. Random Effects Models Embracing Intransitivity

In Tsai and Böckenholt's (2006) variation on Thurstone-Mosteller framework, intransitive preferences are allowed by the model. These intransitive preferences are explained by the variability of the X_{ij} (as in (2.7.1)), which are dependent on the pair of items being compared — other models assume the variation in judgment to be the same across all pairs. Their reasoning is that some comparisons are more difficult to make than

others. However, the final model still results in transitive ordering. They do not believe that one is able to model deviations from final ratings, just that the errors are random.

Crispino et al. (2016) propose a Bayesian version of Tsai's (2006) model, where intransitive preferences occur with some probability. They are unaware of Tsai, but this model is sufficiently different by way of Bayesian Mallows framework. They propose a second kind of model where multiple users that make judgments which are clustered by their level of intransitivity. In both models, the goal is a transitive ordering. Also see Crispino et al. (2017) and Vitelli et al. (2017) for more on their Bayesian Mallows approaches. If the true model is a transitive ranking, but there is much variability in making paired comparisons, observed intransitive preferences would still occur, especially as n increases. The models of Tsai and Crispino are quite reasonable.

2.9. The Blade-Chest Family of Models

In a BTM framework, Shuo Chen and Thorsten Joachims (2016a), (2016b) introduce the Blade-Chest-Inner model that allows intransitivities with a pair of vectors. These vectors represent latent values. However, their solution is iterative and does not rely on an eigendecomposition[4] as ours does. Most models dealing with intransitivity are either identified in Chen's work or reference Chen's work. Chen and Joachims (2016) is Chapter 5 of Chen's dissertation (2016) while Chen and Joachims (2016) is Chapter 6 of Chen's dissertation (2016).

[4] The benefits of eigendecomposition chiefly include well-known and finely-tuned algorithms.

2.9.1. Chen and Joachims

As Chen (2016) notes, Tufto et al. (1998) modeled intransitivity, but required covariates to do so. In Chapter 5, Chen (2016) introduces two[5] models (2.9.2, 2.9.3, 2.9.4) which use only wins and losses or binary preferences but nevertheless allow intransitivity. The entries of Chen's matchup matrix, $M(a,b)$, are defined in (2.9.1.1), where $a = i$ and $b = j$ are items or teams as we have previously defined. It is a more general version of the logit link function (2.4.4) in the BTM as it does not enforce stochastic transitivity, only that $p_{ij} = 1 - p_{ji}$.

$$M(a,b) = \log\left(\frac{p_{ab}}{1-p_{ab}}\right) \tag{2.9.1.1}$$

In these models, each player (i.e., player a) is represented by a pair of k-dimensional vectors $\boldsymbol{a}_{chest}$ and $\boldsymbol{a}_{blade}$. The blade-chest-dist model (2.9.1.2) is then an extension of the 2-dimensional model in Causeur and Husson (2005). Notably in the blade-chest family of models, $k \neq 1$.

$$M(a,b) = \|\boldsymbol{b}_{blade} - \boldsymbol{a}_{chest}\|_2^2 - \|\boldsymbol{a}_{blade} - \boldsymbol{b}_{chest}\|_2^2 \tag{2.9.1.2}$$

They add two "bias" terms in (2.9.1.3) to modify the model in (2.9.1.2) and relate it to the BTM.

$$M(a,b) = \|\boldsymbol{b}_{blade} - \boldsymbol{a}_{chest}\|_2^2 - \|\boldsymbol{a}_{blade} - \boldsymbol{b}_{chest}\|_2^2 + \gamma_a - \gamma_b \tag{2.9.1.3}$$

Finally, the blade-chest-inner model (2.9.1.4) is similar to the k-dimensional model in Tufto et al. (1998) and Kristiansen (2011) in (2.7.5), except the blade and chest vectors are not observed.

$$M(a,b) = \boldsymbol{a}_{chest} \cdot \boldsymbol{b}_{blade} - \boldsymbol{b}_{chest} \cdot \boldsymbol{a}_{blade} + \gamma_a - \gamma_b \tag{2.9.1.4}$$

[5] Nominally two models, actually three models.

Also note that (2.9.1.4) has strength parameters which are independent of the blade and chest vectors, as Chen proves. He shows that the chest-blade-dist model (2.9.1.2) can be related to (2.9.1.4) with (2.9.1.5), where $\gamma'_a = (\|\boldsymbol{a}_{chest}\|_2^2 - \|\boldsymbol{a}_{blade}\|_2^2)/2$ and γ'_b is defined similarly.

$$M(a,b) = 2(\boldsymbol{a}_{chest} \cdot \boldsymbol{b}_{blade} - \boldsymbol{b}_{chest} \cdot \boldsymbol{a}_{blade} + \gamma'_a - \gamma'_b) \tag{2.9.1.5}$$

They then give the likelihood in (2.9.1.6) where $S(x) = (1 + e^{-x})^{-1}$ and D is the training dataset consisting of 4-tuples (a, b, x_{ab}, n_{ab}).

$$\prod_{(a,b,x_{ab},n_{ab})\in D} S\big(M(a,b)\big)^{x_{ab}} \cdot \Big(1 - S\big(M(a,b)\big)\Big)^{n_{ab}-x_{ab}} \tag{2.9.1.6}$$

To fit the models, he samples D, computes the sub-gradients of the local log-likelihood over the parameters, and repeats until the parameters converge. He uses a regularization parameter[6] to prevent over-fitting; however, this may bias the intransitive components. If the regularization parameter is too strong, the model devolves into the BTM. Their intransitive vectors are always of length 2 or more — they never consider scalars. Empirically, (2.9.1.4) performs better than (2.9.1.2), and both are usually better than the BTM. Notably, neither take score differential into account, and whether a contest is home or away is also not considered.

In Chapter 6, Chen (2016) uses context covariates, like whether a game is played at home or away. The context covariates are usually not related to a team's or player's production. In both chapters, datasets related to Street Fighter VI, Defense of the Ancients 2, StarCraft II, and ATP Tennis are used as they involve only two competitors at a time. That

[6] The orthogonality the $\boldsymbol{rs}$ pairs in the CRSP model may not be orthogonal under such a regularization parameter. Need to explore this as well as the relationship between CRSP and Blade-Chest-Inner.

is, NBA data is not examined. Player and context covariates are incorporated with NOACT and TANH activation functions in Chen and Joachims (2016) and Chen (2016) Chapter 6.

2.9.2. Works Related to Chen and Joachims

Causeur and Husson (2005) introduce a 2-dimensional extension (2.9.2.1) of the Bradley-Terry model which Chen and Joachims (2016) generalize to k-dimensional in (2.9.1.2). Causeur and Husson also call this a MDS approach to estimating team strength parameters with win and loss counts.

$$\text{logit}(p_{ij}) = \sigma_{ij}\sqrt{(\lambda_{i1} - \lambda_{j1})^2 + (\lambda_{i2} - \lambda_{j2})^2} \qquad (2.9.2.1)$$

Bilmes et al. (2002) explicitly takes cycles into account, but ultimately finds transitive ratings. Meanwhile, Pahikkala et al. (2010) introduce a kernel which permits intransitivity. They give two interesting synthetic experiments, one which extracts strategies from a $k = 2$ simplex design (i.e., an equilateral triangle whose vertices are rock, paper and scissors) and one which is not unlike the Hawk-Dove game. Volkovs and Zemel (2014) introduce the Multinomial Preference model which is a rank aggregation method that takes intransitive preferences into account.

The Blade-Chest models are wildly popular. Pelechrinis et al. (2016) introduce two models, SportsNetRank and SportsTensoRank. Like Slater (1961), SportsNetRank removes the minimum number of edges to make the graph acyclic. Meanwhile, SportsTensoRank utilizes a 3-dimensional tensor decomposition. Ragain and Ugander (2016) introduce a Pairwise Choice Markov Chain Model. Aoki et al. (2017) introduce a method to "separate skill from luck." However, this is accomplished primarily by comparing margin of victory to winning percentage. Niranjan and Rajkumar (2017) introduce a model which they claim

subsumes BTM, TMM, Blade-Chest, and other models. Chen et al. (2017) give an example of the Blade-Chest model to preferences of images — a paired comparisons application to something other than sporting competition with intransitivity.

Yang et al. (2016) introduce a model which provides live win probabilities for sports, games, and eSports. Live models at some point during a competition may produce win probabilities that are intransitive when compared to pre-match probabilities. Additionally, two line-up models are introduced by Huang et al. (2006) and Pelechrinis (2016), (2018), the latter of which appears to have at the very least inspired Nate Silver et al.'s CARM-Elo model, 2.0 and on (2019).

Gabriel (1998) generalizes linear methods and bilinear models by "criss-cross regression." Gill and Swartz (2001) introduce methods to analyze round robin interaction data in a GLM framework. Hoff (2005) introduces bilinear mixed-effects models for dyadic data on which McCormick and Zheng (2015) build. Weston et al. (2011), Aizenberg et al. (2012), and Mikolov et al. (2013) introduce models with inner product effects, i.e. symmetric bilinear forms. Duan et al. (2017) introduce a generalized model for multidimensional intransitivity which has connections to MDS, Causeur and Husson (2005), and the Blade-Chest model.

CHAPTER 3:

PREVIOUS APPLICATIONS

The author has done a fair amount of modeling with NBA data. Examining it may facilitate conversation about how to proceed in describing relationships between NBA teams, as well as how covariates might be used in the CRSP model.

The author drew inspiration from Nate Silver's PECOTA (2003), CARM-Elo (2015), and DRAYMOND (2019) player and team prediction systems, as well as from Dean Oliver — engineer, former basketball scout, former statistician for the Seattle Supersonics and Sacramento Kings, and an assistant coach for the Washington Wizards as of August 2019. In *Basketball on Paper* (2004) and Roboscout (2004), Oliver defines his four factors for success in basketball games: a high "shooting percentage from the field, getting offensive rebounds, [not] committing turnovers, and going to the foul line a lot and making the shots."

3.1. Covariates Predictive of Winning NBA Games

In 2012, the author compared season-long count statistics to winning percentages for a SAS-based graduate regression course. Every observation was an NBA team's statistical summary for a season (counts of every action performed by an NBA team,

tracked and published by the NBA). The data includes the 32 most recent seasons (1979-1980 through 2010-2011), the only seasons to incorporate the three-point shot.

Several models were produced and discussed, although the most promising was a logistic regression using stepwise variable selection:

$$\hat{p}_i = \frac{e^{-(2.1562-25.1735\cdot winloss_i+15.0924\cdot DPPA_i)}}{1+\ e^{-(2.1562-25.1735\cdot winloss_i+15.0924\cdot DPPA_i)}}. \quad (3.1.1)$$

Here, $DPPA$ is calculated by dividing the number of points a team's opponents scored in a season by the number of field goals attempted. This variable is better than opponent field goal percentage (number of field goals made divided by field goals attempted) since it accounts for all types of offense generated by a shot attempt, including foul shots and three point shooting — both considered to be more efficient than regular two point shots. This is a measure of defensive efficiency. The team's winning percentage is $winloss$.

Using the 2011-2012 season statistics, the author sought to predict the eventual champion later that summer. The top four teams by $\hat{p}_i$ are listed in Table 3-1. Note that $\sum_i \hat{p}_i > 1$ for the top four 2011-2012 teams since $\sum_i \hat{p}_i = 32$ across all 32 seasons. Some seasons had stronger teams that others. The Bulls prediction fell flat almost immediately as Derrick Rose tore his ACL in the first game of the playoffs. The Spurs lost to a team featuring three future league MVPs (Durant, Harden, and Westbrook).

Table 3-1: 2012 Logistic Regression Prediction

Team	$winloss_i$	$DPPA_i$	$\hat{p}_i$
Chicago Bulls	0.75758	1.07362	0.67066
San Antonio Spurs	0.75758	1.14532	0.40829
Oklahoma City Thunder	0.71212	1.13661	0.20039
Miami Heat	0.69697	1.15910	0.10864

3.2. Logistic Regression on PCAs to Predict NBA Championships

In 2015, the author used the same dataset as in the 2012 project with the three additional seasons (2011-2012 through 2013-2014) to complete a group project for a regression course. The work that follows is exclusively his, with the exception that Zhengyang Zhou suggested to difference offensive and defensive statistics. Principle components analysis was used to summarize the variation in the covariates. The five principal components corresponding to the four largest eigenvalues were interpreted. A logistic regression model to rate teams on their championship-caliber-ness was built selecting from the principal components and a time lag championship indicator.

There are two components to a game of basketball: offense and defense. In offense, a team tries to get good shots by being close to the basket, without defenders close by, and easy to make. Furthermore, teams should not turn the ball over before they shoot, create open space by moving, trick their opponents, and use all members of the team.
In defense, teams should stay between their opponents and the basket, eliminate space between themselves and a shooter, be aware of where the ball is, and help other members of their team. They should not let the opponents rebound.

3.2.1. NBA Subject Introduction

A team wins a game of basketball by scoring more points than their opponent. Shots known as field goals are worth 2 points; however, shots taken behind the 3-point line are worth 3 points. Free throws are worth 1 point. Over the 1980-1981 through 2013-2014 NBA seasons, teams can expect to score 0.966 points per 2 point attempt, 1.047 points per 3 point attempt, and 1.506 points per free throw.

The National Basketball Association (NBA) was founded in 1946-1947 season. The last substantial rule change in occurred in the 1979-1980 season, when league officials added the 3-point line. The NBA tracks several count statistics on the player level; however, the data used here is aggregated by team and season over 35 seasons (1980-2014). This is probably a disadvantage — it would be better to look at the data by game so that we could tie performance in each category to a win-loss outcome.

We divided by number of games team played that season, and found the difference of totals for each category to find a per game categorical difference. For example, team i had an average of 52 rebounds per game and their opponents had an average of 49.5 rebounds per game against team i. Then the resulting average per game difference was 2.5.

3.2.2. Principle Components

The variables included in the PCA are in Table 3-2.

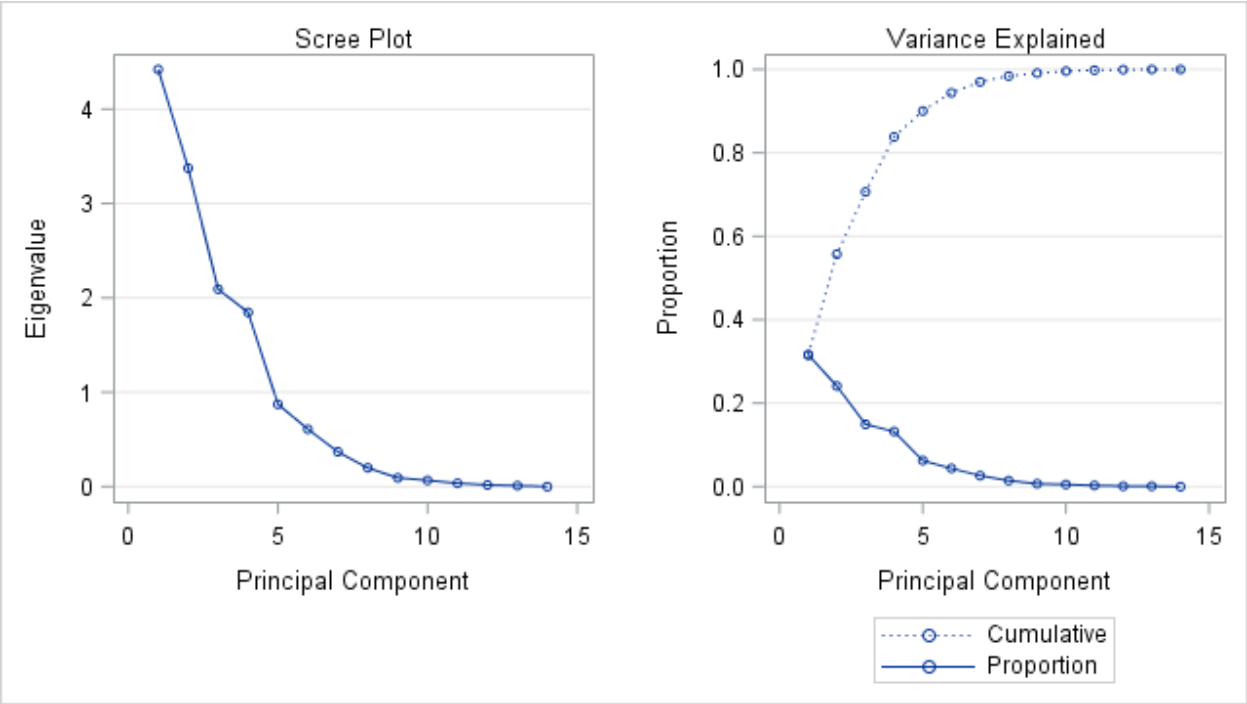
Table 3-2: PC Analysis Variables Included

Abbreviation	Variable
FGM	Field Goals Made
FGA	Field Goals Attempted
FTM	Free Throws Made
FTA	Free Throws Attempted
DR	Defensive Rebound
OR	Offensive Rebound
TR	Total Rebound

Abbreviation	Variable
A	Assists
PF	Personal Fouls
S	Steals
T	Turnovers
B	Blocks
TPM	Three Pointers Made
PT	Points

In Figure 3-1, we see that the first four principal components explain 84% of the variation in the covariates. We selected the first four components for interpretation.

Figure 3-1: PCA Scree Plot and Proportion of Variation Explained



To interpret the principle components, we found all factor loadings greater than 0.28 (an arbitrary threshold). Then, we identified the driving factor, and related all other factors to that factor. We interpreted these as either roster abilities or coaching strategies, found an exemplar of that principle component, and identified why that team was high in that principle component and what percent of games it won. We refer to the vector of described factor loadings as $\boldsymbol{\beta}_+$ and the ones we do not describe as $\boldsymbol{\beta}_-$, and similarly, the vector of observed covariates chosen and not as $\boldsymbol{X}_+$ and $\boldsymbol{X}_-$. Note that the observed covariates are centered and scaled.

The first principal component was "Foul Choices;"

$$PC1 \approx 0.357FTM + 0.355FTA + 0.396DR + 0.291TR - 0.340PF + 0.385PT + \boldsymbol{\beta}_-^T \boldsymbol{X}_-.$$

It describes teams who foul less than their opponents. This may cause them to make or attempt more free throws, have more opportunities for defensive rebounds, and score more overall points. Fouling more is a symptom of bad defense, which means a team is more likely to suffer from their poor performance. This principal component should have a positive relation with winning percentage. An exemplar of the Foul Choices principal component is the 2009-10 Cleveland Cavaliers. Their observed Foul Choice principal component is

$$PC1_{CLE10} \approx \boldsymbol{\beta}_{+}^{T}[0.6 \quad 1.1 \quad 1.8 \quad 1.4 \quad -0.8 \quad 1.4] + \boldsymbol{\beta}_{-}^{T}\boldsymbol{X}_{-} \approx 3.56.$$

This team featured 25 year-old LeBron James entering his prime. He was virtually unguardable, opposing teams gave him an average of 10.2 free throw attempts per game (an amount typically only two to five players take in a given NBA season), and he was still making a consistent effort on defense. They won 74% of their games that year.

The second principal component was "Possession Completion;"

$$PC2 \approx 0.457FGM + 0.385FGA + 0.331A + 0.340S - 0.346T + 0.283PT + \boldsymbol{\beta}_{-}^{T}\boldsymbol{X}_{-}.$$

It describes teams who have fewer turnovers or more steals, giving them more opportunities to score, which means the team makes more field goals and has more assists, therefore scoring more points overall. Turnovers are often indicative of a poor point guard or playmaker. This principal component should also have a positive relation with winning percentage. An exemplar of the Possession Completion principal component is the 2003-04 Los Angeles Clippers. Their observed Possession Completion principal component is

$$PC2_{LAC04} \approx \boldsymbol{\beta}_{+}^{T}[-1.6 \quad -0.5 \quad -1.1 \quad -1.1 \quad 1.7 \quad -1] + \boldsymbol{\beta}_{-}^{T}\boldsymbol{X}_{-} \approx -2.89.$$

This team had no true point guard and instead was run by players like Corey Maggette, a shooting guard/small forward. His assist to turnover ratio was $224/207 \approx 1.08$, but point guards usually average 2 assists per turnover or more. The team won 34% of their games.

The third principal component was "Forcing Turnovers;"

$$PC3 \approx 0.301OR + 0.324DR + 0.420TR - 0.417S + 0.485T + \boldsymbol{\beta}_{\_}^{T}\boldsymbol{X}_{\_}.$$

It describes teams who have fewer steals or more turnovers. This subsequently gives them more opportunities to rebound, since they take fewer shots. Forcing turnovers is indicative of a good wing defender. This principal component should have a negative relation with winning percentage. An exemplar of the Forcing Turnovers principal component is the 2005-06 Detroit Pistons. Their observed Forcing Turnovers principal component is

$$PC3_{DET06} \approx \boldsymbol{\beta}_{+}^{T}[-1.1 \quad 0.6 \quad -0.1 \quad 1.1 \quad -1.7] + \boldsymbol{\beta}_{\_}^{T}\boldsymbol{X}_{\_} \approx 2.66.$$

They had an excellent wing defender Tayshaun Prince and had historically great rebounders in Rasheed Wallace and Ben Wallace. The team won 78% of their games.

The fourth principal component was "Return Strategy;"

$$PC4 = 0.555OR + 0.284TR - 0.340B - 0.326TPM + \boldsymbol{\beta}_{\_}^{T}\boldsymbol{X}_{\_}.$$

Teams low in this principal component attempt 3-pointers frequently and immediately run back on defense, meaning they will have fewer opportunities to get offensive rebounds, get blocked less on their shot attempts since they are taking 3s, and get more blocks on defense. This is a common strategy for older teams. This principal component should have a negative relation with winning percentage. An exemplar of the Return Strategy principal component is the 2012-13 San Antonio Spurs. Their observed Return Strategy principal component is

$$PC4_{SAS13} \approx \boldsymbol{\beta}_{+}^{T}[-2 \quad -0.5 \quad 0.4 \quad 1.4] + \boldsymbol{\beta}_{\_}^{T}\boldsymbol{X}_{\_} \approx -2.22.$$

This Spurs team was uniquely equipped with all-time 3-point shooters like Danny Green, Marco Bellinelli, Patty Mills, and Matt Bonner and skilled big men like Tim Duncan and Tiago Splitter to earn blocks and rebounds. The Spurs won 71% of their games.

The principal components are summarized in Table 3-3.

Table 3-3: Principal Components Summary

Principal Component	Name	Sign
$PC1$	Foul Choices	$+$
$PC2$	Possession Completion	$+$
$PC3$	Forcing Turnovers	$-$
$PC4$	Return Strategy	$-$

3.2.3. Modeling Championship Probability with PCAs

To predict the probability of winning a championship, we fit a logistic regression model using the principal components as well as winning percentage Y_i and a binary variable for championship won in the previous year, W_{i-1}. Several of the PCs are collinear with these additional variables, and so do not show up in the model. The variables were selected with forward stepwise selection. The model, which had the lowest AIC and largest AUC among candidate models, is

$$\ln\left(\frac{\hat{p}_i}{1-\hat{p}_i}\right) = -19.03 + 23.47Y_i - .66W_{i-1} - 1.40PC_{i9} - 1.08PC_{i8} + .46PC_{i6} + .26PC_{i1} + \varepsilon_i.$$

And thus, the probability of winning can be formally expressed as

$$\hat{p}_i = \frac{1}{1 + e^{19.03-23.47Y_i+.66W_{i-1}+1.40X_{i9}+1.08X_{i8}-.46X_{i6}-.26X_{i1}+\varepsilon_i}}.$$

This model gave the highest rating to the eventual champion in 20 of the 35 seasons.

CHAPTER 4:

INTRODUCING THE CRSP MODEL

Here, we introduce the CRSP model (pronounced "crisp"), whose latent interaction structure is novel in an expected score differential setting. We discuss the mathematical details of the models in this chapter. Simulations and an application of the generalized CRSP model are discussed in Chapter 5, while Section 7.2 discusses ongoing and future work regarding CRSP.

4.1. A Basic Intransitive Interaction Model

We introduce a few basic models for Y_{ij}, the score of home team i minus the score of away team j. All the models in Section 4.1 are of the form $Y_{ij} = \mu_{ij} + \varepsilon_{ij}$, where errors are assumed independent with constant variance and μ_{ij} is the true point differential. The first model (4.1.1) is that all teams are equal and that outcomes are purely random.

$$\mu_{ij} = 0 \tag{4.1.1}$$

The second model (4.1.2) is that teams have equal strength, but that the home team has some advantage h, which is the same for all pairs of teams.

$$\mu_{ij} = h \tag{4.1.2}$$

The third model (4.1.3) is a purely transitive model where each team i has a

strength parameter α_i. In the complete and balanced case, the strength estimate is equivalent to average point differential, a measure displayed in most league standings summaries.

$$\mu_{ij} = h + \alpha_i - \alpha_j \tag{4.1.3}$$

We also introduce an intransitive model (4.1.4) with a fixed interaction effect β_{ij} between every pair of teams i and j in addition to the parameters in (4.1.3) to serve as a reference point later.

$$\mu_{ij} = h + \alpha_i - \alpha_j + \beta_{ij} \tag{4.1.4}$$

We assume $\beta_{ij} = -\beta_{ji}$, which is similar to the assumption in the BTM that $p_{ij} = 1 - p_{ji}$. We also note that in such a model, the α_i are estimated first to center the data, i.e., they are subject to $\sum_i \alpha_i = 0$. Asymptotically, (4.1.4) fits one parameter for every two observations, producing an overfit model. We offer a generalization of the previous two models where $\beta_{ij}\ \forall i, j$ is multiplied by a fitted constant w. When $w = 0$, (4.1.5) becomes (4.1.3) and when $w = 1$, (4.1.5) becomes (4.1.4).

$$\mu_{ij} = h + \alpha_i - \alpha_j + w\beta_{ij} \tag{4.1.5}$$

A simple approach might be to find $w \in [0,1]$ such that (4.1.5) is not overfit but still contains as much intransitivity as the data allow. In this sense, w could be considered a shrinkage parameter. This fails, however, when some interactions between team i and team j, β_{ij}, are truly nonzero and others ($\beta_{i^*j^*}$) are truly zero. We believe CRSP can strike a middle ground where systemic relationships can be identified between teams which explain why β_{ij} is nonzero in some cases but not in others.

4.2. Callback to Kendall and Smith

Remember that in Kendall and Smith (1940), a circular triad was when $A \to B \to C \to A$ and could be expressed as $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. As in HodgeRank, if we let a loss be indicated by a -1 instead of a 0, then a circular triad would be expressed as in (4.2.1)

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix}. \tag{4.2.1}$$

We could propose an alternative to $w\beta_{ij}$ in (4.1.5) where we let teams i and j interact via some function of additional parameters $\boldsymbol{\beta}_i = (r_i, s_i, p_i)$ and $\boldsymbol{\beta}_j$, say $f(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j) = \boldsymbol{\beta}_i^T \boldsymbol{A} \boldsymbol{\beta}_j$. Then we get

$$\boldsymbol{\beta}_i^T \boldsymbol{A} \boldsymbol{\beta}_j = \begin{bmatrix} r_i & s_i & p_i \end{bmatrix} \begin{bmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} r_j \\ s_j \\ p_j \end{bmatrix} = r_i s_j - s_i r_j + r_i p_j - p_i r_j + s_i p_j - p_i s_j. \tag{4.2.2}$$

We impose the restriction

$$r_i + s_i + p_i = 0, \forall i. \tag{4.2.3}$$

Substituting $p_i = -r_i - s_i$ into (4.2.2), we get

$$\boldsymbol{\beta}_i^T \boldsymbol{A} \boldsymbol{\beta}_j = 3(r_i s_j - s_i r_j). \tag{4.2.4}$$

Without loss of generality, we let our $f(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j) = r_i s_j - s_i r_j$. This could also have been accomplished by removing p_i from $\boldsymbol{\beta}_i$ and getting that

$$f(\boldsymbol{\beta}_i, \boldsymbol{\beta}_j) = \begin{bmatrix} r_i & s_i \end{bmatrix} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} r_j \\ s_j \end{bmatrix} = r_i s_j - s_i r_j$$

or similarly $\det \begin{vmatrix} r_i & s_i \\ r_j & s_j \end{vmatrix} = r_i s_j - s_i r_j$. However, the determinant analogy falls apart with the addition of p_i and p_j.

So, we propose the CRSP model in (4.2.5),

$$\mu_{ij} = h + \alpha_i - \alpha_j + r_i s_j - s_i r_j. \tag{4.2.5}$$

We also propose the generalized CRSP model in (4.2.6), where $K \in \left\{1, \ldots, \left\lfloor \frac{n-1}{2} \right\rfloor\right\}$,

$$\mu_{ij} = h + \alpha_i - \alpha_j + \sum_{k=1}^{K} \left(r_{ik} s_{jk} - s_{ik} r_{jk}\right). \tag{4.2.6}$$

4.3. Some Matrix Algebra

We now discuss some relevant facts about matrices.

4.3.1. Facts About Skew-Symmetric Matrices

1. A square matrix $\boldsymbol{A}$ is skew-symmetric if $\boldsymbol{A}^T = -\boldsymbol{A}$.
2. If a skew-symmetric matrix consists of all real numbered entries, then its eigenvalues are all purely imaginary (or zero).
   a. The non-zero eigenvalues come in pairs $\pm i\lambda_k, k = 1, 2, \ldots$, where all λ_k are real. This implies that if the matrix is $n \times n$ with n odd, at least one eigenvalue is 0. We will let d be the number of pairs of non-zero eigenvalues.
   b. Any real skew-symmetric matrix can be decomposed into diagonal form (eigenvector decomposition), as follows:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\dagger} = [\boldsymbol{u}_1\, \boldsymbol{u}_2\ \boldsymbol{u}_3\ \boldsymbol{u}_4\ \ldots] \begin{bmatrix} i\lambda_1 & 0 & 0 & 0 & \cdots \\ 0 & -i\lambda_1 & 0 & 0 & \cdots \\ 0 & 0 & i\lambda_2 & 0 & \cdots \\ 0 & 0 & 0 & -i\lambda_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1^{\dagger} \\ \boldsymbol{u}_2^{\dagger} \\ \boldsymbol{u}_3^{\dagger} \\ \boldsymbol{u}_4^{\dagger} \\ \vdots \end{bmatrix}$$

$$= \sum_{k=1}^{K} i\lambda_k \left(\boldsymbol{u}_{2k-1}\, \boldsymbol{u}_{2k-1}^{\dagger} - \boldsymbol{u}_{2k}\, \boldsymbol{u}_{2k}^{\dagger}\right).$$

In this decomposition, the $\boldsymbol{u}_k$ are complex orthogonal eigenvectors, and "†" denotes the conjugate transpose.

3. Any real skew-symmetric matrix can be decomposed into block-diagonal form as follows[7]

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^T = [\boldsymbol{q}_1\ \boldsymbol{q}_2\ \boldsymbol{q}_3\ \boldsymbol{q}_4\ \dots] \begin{bmatrix} 0 & \lambda_1 & 0 & 0 & \cdots \\ -\lambda_1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & \lambda_2 & \cdots \\ 0 & 0 & -\lambda_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^T \\ \boldsymbol{q}_2^T \\ \boldsymbol{q}_3^T \\ \boldsymbol{q}_4^T \\ \vdots \end{bmatrix}.$$

In this decomposition, the λ_k are the same as in 2.a. above, and the $\boldsymbol{q}_k$ are real, orthogonal column vectors (not the eigenvectors of A). Note without loss of generality $\boldsymbol{q}_1^T\boldsymbol{q}_1 = 1$. This reduces to

$$\begin{aligned} \boldsymbol{A} &= -\lambda_1\boldsymbol{q}_2\boldsymbol{q}_1^T + \lambda_1\boldsymbol{q}_1\boldsymbol{q}_2^T - \lambda_2\boldsymbol{q}_4\boldsymbol{q}_3^T + \lambda_2\boldsymbol{q}_3\boldsymbol{q}_4^T + \cdots \\ &= \sum_{k=1}^{K} \lambda_k \left(\boldsymbol{q}_{2k-1}\boldsymbol{q}_{2k}^T - \boldsymbol{q}_{2k}\boldsymbol{q}_{2k-1}^T\right). \end{aligned}$$

Now define $\lambda_k^{1/2}\boldsymbol{q}_{2k-1} = \boldsymbol{r}_k$ and $\lambda_k^{1/2}\boldsymbol{q}_{2k} = \boldsymbol{s}_k$ (so $\boldsymbol{r}_k^T\boldsymbol{r}_k = \lambda_k = \boldsymbol{s}_k^T\boldsymbol{s}_k$). Then,

$$\boldsymbol{A} = \sum_{k=1}^{K} \boldsymbol{r}_k\boldsymbol{s}_k^T - \boldsymbol{s}_k\boldsymbol{r}_k^T. \qquad (4.3.1.1)$$

4.3.2. Other Facts About Matrices

1. If a square matrix is of the form $\boldsymbol{B}\boldsymbol{B}^T$, then it is positive semi-definite, which means that all eigenvalues are greater than or equal to 0.

2. We have that a matrix of the form $\boldsymbol{B}\boldsymbol{B}^T$ can be decomposed as follows, where the $\boldsymbol{u}_k$ are real orthogonal eigenvectors and $\lambda_k \geq 0$ are the eigenvalues of $\boldsymbol{B}\boldsymbol{B}^T$.

[7] Standard result in matrix theory.

$$\boldsymbol{B}\boldsymbol{B}^T = [\boldsymbol{u}_1\, \boldsymbol{u}_2\ \boldsymbol{u}_3\ \boldsymbol{u}_4\ \dots] \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & \cdots \\ 0 & \lambda_2 & 0 & 0 & \cdots \\ 0 & 0 & \lambda_3 & 0 & \cdots \\ 0 & 0 & 0 & \lambda_4 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \boldsymbol{u}_1^\dagger \\ \boldsymbol{u}_2^\dagger \\ \boldsymbol{u}_3^\dagger \\ \boldsymbol{u}_4^\dagger \\ \vdots \end{bmatrix} = \sum_k \lambda_k\, \boldsymbol{u}_k \boldsymbol{u}_k^T$$

4.3.3. Relationship Between Skew-Symmetric $\boldsymbol{A}$ and $\boldsymbol{A}\boldsymbol{A}^T$

From (4.3.1.1), we can write $\boldsymbol{A}\boldsymbol{A}^T$ as

$$\boldsymbol{A}\boldsymbol{A}^T = \textstyle\sum_{k=1}^{K}(\boldsymbol{r}_k\boldsymbol{s}_k^T - \boldsymbol{s}_k\boldsymbol{r}_k^T)\,(\sum_{k=1}^{K}(\boldsymbol{r}_k\boldsymbol{s}_k^T - \boldsymbol{s}_k\boldsymbol{r}_k^T))^T.$$

Since the $\boldsymbol{r}_k$ and $\boldsymbol{s}_k$ are all orthogonal, we have

$$\boldsymbol{A}\boldsymbol{A}^T = \sum_{k=1}^{K}(\boldsymbol{r}_k\boldsymbol{s}_k^T\boldsymbol{s}_k\boldsymbol{r}_k^T + \boldsymbol{s}_k\boldsymbol{r}_k^T\boldsymbol{r}_k\boldsymbol{s}_k^T) = \sum_{k=1}^{K}\lambda_k(\boldsymbol{r}_k\boldsymbol{r}_k^T + \boldsymbol{s}_k\boldsymbol{s}_k^T)\,. \tag{4.3.3.1}$$

Careful examination of this shows us that the $\boldsymbol{r}_k$ and $\boldsymbol{s}_k$ are eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$; furthermore, these eigenvectors come in pairs. That is, each of the non-zero eigenvalues has multiplicity 2.

4.3.4. Perfect Fit for Double Round Robin

Our data consists of a square $n \times n$ matrix $\boldsymbol{Y}$ which consists of score differentials (home score $-$ away score) for all match-ups between n teams. We may or may not have zeroes on the diagonal, and for now we ignore this. Now let $\boldsymbol{Z} = \boldsymbol{Y}/2$ and define

$$\boldsymbol{D} = \frac{(\boldsymbol{Y} - \boldsymbol{Y}^T)}{2} = \boldsymbol{Z} - \boldsymbol{Z}^T.$$

Let $\widehat{\boldsymbol{\alpha}}_{n\times 1} = \frac{1}{n}\boldsymbol{D}\boldsymbol{1}$ be the estimate of the "main effects" or strength values for the n teams, and define $\boldsymbol{A} = \widehat{\boldsymbol{\alpha}}\boldsymbol{1}^T = \frac{1}{n}(\boldsymbol{Z} - \boldsymbol{Z}^T)\boldsymbol{1}\boldsymbol{1}^T$ (note that $\boldsymbol{A}$ is singular). Now define $\boldsymbol{B} = \boldsymbol{Z} - \boldsymbol{A}$; this is a square matrix, but is not symmetric or skew-symmetric.

Finally define $\boldsymbol{C} = \boldsymbol{B} - \boldsymbol{B}^T = \boldsymbol{D} - \boldsymbol{M}$, where $\boldsymbol{M} = \boldsymbol{A} - \boldsymbol{A}^T$ represents the fitting of

main effects to the data. This matrix $\boldsymbol{C}$ can be termed the "interaction matrix;" it is what is left over after fitting the main effects (overall strength). If the data are perfectly modeled by a transitive strength relationship, we will have $\boldsymbol{C} = \boldsymbol{0}$. Note that $\boldsymbol{C}^T = -\boldsymbol{C}$, so $\boldsymbol{C}$ is skew-symmetric. Thus by (4.3.1.1), where $\boldsymbol{r}_k$ and $\boldsymbol{s}_k$ are the eigenvectors of $\boldsymbol{C}\boldsymbol{C}^T$, we have

$$\boldsymbol{C} = \sum_{k=1}^{K} \boldsymbol{r}_k \boldsymbol{s}_k^T - \boldsymbol{s}_k \boldsymbol{r}_k^T . \tag{4.3.4.1}$$

We see that a perfect fit to $\boldsymbol{C}$ can be found by using all K components in (4.2.6), where $K = \left\lfloor \frac{n-1}{2} \right\rfloor$. We will prove the eigenvector pair corresponding to the largest eigenvalue pair explains the most variation in the deviations from the transitive model in Section 4.5. Subsequently, we prove that each of the remaining largest eigenvalue pairs give us the eigenvector pair that explains the most variation in the deviations from the transitive model. The $K = \left\lfloor \frac{n-1}{2} \right\rfloor$ model is equivalent to fitting the interaction term β_{ij} in (4.1.4) and thus undesirable in the (balanced) $n_{ij} = 1 \; \forall i, j$ case.

4.3.5. The Trace of $\boldsymbol{AA}^T$

Consider a square matrix $\boldsymbol{A}_{4\times4}$,

$$\boldsymbol{A}_{4\times4} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}.$$

Additionally, consider

$$\boldsymbol{AA}^T = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{31} & a_{41} \\ a_{12} & a_{22} & a_{32} & a_{42} \\ a_{13} & a_{23} & a_{33} & a_{43} \\ a_{14} & a_{24} & a_{34} & a_{44} \end{bmatrix}.$$

Then, where $[\boldsymbol{X}]_{(ij)}$ is the (i,j) entry of a matrix $\boldsymbol{X}$,

$$\begin{aligned} &\text{trace}(\boldsymbol{AA}^T) = [\boldsymbol{AA}^T]_{(11)} + [\boldsymbol{AA}^T]_{(22)} + [\boldsymbol{AA}^T]_{(33)} + [\boldsymbol{AA}^T]_{(44)} \\ &= (a_{11}^2 + a_{12}^2 + a_{13}^2 + a_{14}^2) \\ &+(a_{12}^2 + a_{22}^2 + a_{23}^2 + a_{24}^2) \\ &+(a_{13}^2 + a_{23}^2 + a_{33}^2 + a_{34}^2) \\ &+(a_{14}^2 + a_{24}^2 + a_{34}^2 + a_{44}^2) \\ &= \sum_{i=1}^{4}\sum_{j=1}^{4} a_{ij}^2. \end{aligned}$$

In general, then, for any $\boldsymbol{A}_{n\times n}$,

$$\boldsymbol{AA}^T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{bmatrix},$$

and

$$\text{trace}(\boldsymbol{AA}^T) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 . \qquad (4.3.5.1)$$

Also, note that for a vector $\boldsymbol{a}$,

$$\text{trace}(\boldsymbol{a}\boldsymbol{a}^T) = \boldsymbol{a}^T\boldsymbol{a} = \sum_i a_i^2 . \tag{4.3.5.2}$$

4.4. CRSP Model Fitting

We will fit models using minimization of SS. In non-matrix form, we wish to minimize

$$SS = \sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{n}\left(y_{ij} - \mu_{ij}\right)^2 . \tag{4.4.1}$$

The equations to be solved are of the form (4.4.2):

$$\frac{\partial SS}{\partial \theta} = (-2)\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{n}\left(y_{ij} - \mu_{ij}\right)\frac{\partial \mu_{ij}}{\partial \theta} = 0. \tag{4.4.2}$$

We can write (4.4.2) as (4.4.3):

$$\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\frac{\partial \mu_{ij}}{\partial \theta} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_{ij}\frac{\partial \mu_{ij}}{\partial \theta}. \tag{4.4.3}$$

These equations can be written in matrix form fairly easily (unlike the SS itself). For example, consider the model in (4.1.2), then we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij}\frac{\partial \mu_{ij}}{\partial h} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_{ij}\frac{\partial \mu_{ij}}{\partial h}. \tag{4.4.4}$$

Now, since $\frac{\partial \mu_{ij}}{\partial h} = 1, \forall i, j$, we have $\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} h$ or

$$\hat{h} = \bar{y}_{\cdot\cdot} . \tag{4.4.5}$$

Note that (4.4.5) could also be written as (4.4.6), (4.4.7), or (4.4.8),

$$\hat{h} = \frac{1}{2n^2}\mathbf{1}^T(\boldsymbol{Y} + \boldsymbol{Y}^T)\mathbf{1}, \text{or} \tag{4.4.6}$$

$$\hat{h}(\mathbf{1}^T(\mathbf{1}\mathbf{1}^T)\mathbf{1}) = \mathbf{1}^T\frac{(\boldsymbol{Y}+\boldsymbol{Y}^T)}{2}\mathbf{1}, \text{or} \tag{4.4.7}$$

$$\mathbf{1}^T\left[\frac{(\boldsymbol{Y}+\boldsymbol{Y}^T)}{2} - \hat{h}\mathbf{1}\mathbf{1}^T\right]\mathbf{1} = 0. \tag{4.4.8}$$

Let us revisit the equations under the CRSP model we proposed in (4.2.5),

$$\mu_{ij} = h + \alpha_i - \alpha_j + r_i s_j - s_i r_j.$$

Note we need $\frac{\partial \mu_{ij}}{\partial \theta}$ for each θ.

$$\frac{\partial \mu_{ij}}{\partial h} = 1$$

$$\frac{\partial \mu_{ij}}{\partial \alpha_1} = \begin{cases} 1 & i = 1, j \in \{2,3,\dots,n\} \\ -1 & j = 1, i \in \{2,3,\dots,n\} \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial \mu_{ij}}{\partial r_1} = \begin{cases} s_j & i = 1, j \in \{2,3,\dots,n\} \\ -s_j & j = 1, i \in \{2,3,\dots,n\} \\ 0 & \text{otherwise} \end{cases}$$

$$\frac{\partial \mu_{ij}}{\partial s_1} = \begin{cases} -r_j & i = 1, j \in \{2,3,\dots,n\} \\ r_j & j = 1, i \in \{2,3,\dots,n\} \\ 0 & \text{otherwise} \end{cases}$$

This gives us a set of fitting equations.

For h, we get

$$\sum_{i=1}^{n}\sum_{j=1}^{n} y_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} h + \alpha_j - \alpha_j + r_i s_j - s_i r_j = n^2 h. \tag{4.4.9}$$

From (4.4.9) we get (4.4.6) again;

$$\hat{h} = \frac{1}{2n^2}\mathbf{1}^T(\boldsymbol{Y}+\boldsymbol{Y}^T)\mathbf{1} = \bar{y}_{\cdot\cdot}.$$

For α_1, we have

$$\sum_{j=1}^{n} y_{1j} - \sum_{i=1}^{n} y_{i1} = \sum_{j=1}^{n} \mu_{1j} - \sum_{i=1}^{n} \mu_{i1} \ . \tag{4.4.10}$$

This gives

$$y_{1\cdot} - y_{\cdot 1} = \sum_{j=1}^{n} \big(h + \ \alpha_1 - \alpha_j + r_1 s_j - s_1 r_j\big) - \sum_{i=1}^{n} (h + \alpha_i - \alpha_1 + r_i s_1 - s_i r_1)\,. \tag{4.4.11}$$

And (4.4.11) reduces to

$$y_{1\cdot} - y_{\cdot 1} = h + \alpha_1 - \alpha_j + r_1 s_j - s_1 r_j. \tag{4.4.12}$$

Now, we know $\alpha_{\cdot} = 0$ by requirement of constraints. In addition, it is a usual constraint that any interaction parameter β_{ij} sum to zero over j for fixed i and sums to zero over i for fixed j. For us, $\beta_{1j} = r_1 s_j - \ s_1 r_j$, so $\beta_{1\cdot} = r_1 s_{\cdot} - \ s_1 r_{\cdot} = 0$ with similar algebra showing that $\beta_{\cdot 1} = s_1 r_{\cdot} - r_1 s_{\cdot} = 0$. The result is that all terms except α_1 disappear, and so we get (4.4.13):

$$\hat{\alpha}_1 = \frac{1}{2}(\bar{y}_{1\cdot} - \bar{y}_{\cdot 1}). \tag{4.4.14}$$

In general, this gives (4.4.15). As claimed earlier, the $\hat{\boldsymbol{\alpha}}$ are the average score differential in the balanced case

$$\hat{\boldsymbol{\alpha}} = \frac{1}{2n}[\boldsymbol{Y} - \boldsymbol{Y}^T]\mathbf{1}. \tag{4.4.15}$$

For r_1, we have (4.4.16):

$$\sum_{j=1}^{n} y_{1j} s_j - \sum_{i=1}^{n} y_{i1} s_i = \sum_{j=1}^{n} \mu_{1j} s_j - \sum_{i=1}^{n} \mu_{i1} s_i \,. \tag{4.4.16}$$

This gives

$$\sum_{j=1}^{n} y_{1j}s_j - \sum_{i=1}^{n} y_{i1}s_i = \sum_{j=1}^{n}\big(h + \alpha_1 - \alpha_j + r_1 s_j - s_1 r_j\big)s_j - \sum_{i=1}^{n}\big(h + \alpha_1 - \alpha_j + r_i s_1 - s_i r_1\big)s_i.$$

And the RHS of (4.4.16) is

$$\begin{aligned} & h\sum_{j=1}^{n} s_j + \alpha_1 \sum_{j=1}^{n} s_j - \sum_{j=1}^{n} \alpha_j s_j + \sum_{j=1}^{n} s_j^2 - s_1 \sum_{j=1}^{n} r_j s_j - h\sum_{i=1}^{n} s_i \\ & \qquad - \sum_{i=1}^{n} \alpha_i s_i + \alpha_1 \sum_{i=1}^{n} s_i + s_1 \sum_{i=1}^{n} r_i s_i + r_i \sum_{i=1}^{n} s_i^2 \\ & = 2\alpha_1 \sum_{i=1}^{n} s_i - 2\sum_{i=1}^{n} \alpha_i s_i + 2r_1 \sum_{i=1}^{n} s_i^2 - 2s_1 \sum_{i=1}^{n} r_i s_i. \end{aligned}$$

Now we will assume that the last term is zero, that is $\sum_{i=1}^{n} r_i s_i = 0$. This is because we will show that the solutions are orthogonal eigenvectors. Now, we have

$$\sum_{j=1}^{n} y_{1j}s_j - \sum_{i=1}^{n} y_{i1}s_i = 2\alpha_1 \sum_{i=1}^{n} s_i - 2\sum_{i=1}^{n} \alpha_i s_i + 2r_1 \sum_{i=1}^{n} s_i^2 . \tag{4.4.17}$$

Using $\lambda_{ss} = \sum_{i=1}^{n} s_i^2$, we have

$$\lambda_{ss} r_1 = \frac{1}{2}\sum_{i=1}^{n} (y_{1i} - y_{i1})s_i + \boldsymbol{\alpha}^T \boldsymbol{s} - \alpha_1(\mathbf{1}^T \boldsymbol{s}). \tag{4.4.18}$$

We get similar equations for $r_2, \dots, r_n$, so we have (4.4.19):

$$\lambda_{ss}\boldsymbol{r} = \frac{1}{2}[\boldsymbol{Y} - \boldsymbol{Y}^T]\boldsymbol{s} + \mathbf{1}\boldsymbol{\alpha}^T\boldsymbol{s} - \boldsymbol{\alpha}\mathbf{1}^T\boldsymbol{s}. \tag{4.4.19}$$

In the notation from Section 4.3, (4.4.19) is

$$\lambda_{ss}\boldsymbol{r} = \boldsymbol{D}\boldsymbol{s} - \boldsymbol{A}\boldsymbol{s} + \boldsymbol{A}^T\boldsymbol{s} = (\boldsymbol{D} - \boldsymbol{M})\boldsymbol{s}. \tag{4.4.20}$$

Similarly, we get

$$\lambda_{rr}\boldsymbol{s} = (\boldsymbol{D} - \boldsymbol{M})^T\boldsymbol{r}. \tag{4.4.21}$$

Putting these together we see

$$\lambda_{rr}\lambda_{ss}\boldsymbol{r} = (\boldsymbol{D}-\boldsymbol{M})(\boldsymbol{D}-\boldsymbol{M})^T\boldsymbol{r}. \tag{4.4.22}$$

Inspection of (4.4.22) shows us that indeed $\boldsymbol{r}$ is an eigenvector of $(\boldsymbol{D}-\boldsymbol{M})(\boldsymbol{D}-\boldsymbol{M})^T$, with eigenvalue $\lambda = \lambda_{rr}\lambda_{ss}$. The eigenvector needs to be properly scaled so that $\mathbf{1}^T\boldsymbol{r} = \lambda^{1/2}$.

4.5. Largest Eigenvalue Proof

Define Y_{ij} as home team i's score minus away team j's score, i.e., Y_{ij} is the home team's score differential, where $i, j \in \{1,2,\dots,n\}$.

Define $z_{ij} = Y_{ij} - \hat{h}$, where $\hat{h} = \bar{Y}_{\cdot\cdot} = \frac{\sum_{ij} Y_{ij}}{n^2}$ (i.e., when $n_{ij} = 1\ \forall i, j$; all teams play one game at home and one game away against all other teams, including themselves).

Define $d_{ij} = \frac{1}{2}\left(Y_{ij} - Y_{ji}\right) = \frac{1}{2}\left(z_{ij} - z_{ji}\right)$, since

$$z_{ij} - z_{ji} = \left(Y_{ij} - \hat{h}\right) - \left(Y_{ji} - \hat{h}\right) = Y_{ij} - Y_{ji}.$$

Let $\mu_{ij} = -\mu_{ji}$, which gives skew-symmetry.

Also, $\boldsymbol{D} \coloneqq \{d_{ij}\}$ and $\boldsymbol{M} \coloneqq \{\mu_{ij}\}$. $\widehat{\boldsymbol{M}} = \{\hat{\mu}_{ij}\}$, where $\widehat{\boldsymbol{M}} = \hat{\boldsymbol{\alpha}}\mathbf{1}^T - \mathbf{1}\hat{\boldsymbol{\alpha}}^T$ with $\hat{\alpha}_i = \frac{1}{2}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot i})$.

Now, our models are of the form $E[Y_{ij}] = h + \mu_{ij}$.

Consider models which have $\mu_{ij} = -\mu_{ji}$, that is, a skew-symmetric $\boldsymbol{M}$.

$$\mu_{ij} = \alpha_i - \alpha_j \tag{4.5.1}$$

$$\mu_{ij} = \alpha_i - \alpha_j + r_i s_j - s_i r_j \tag{4.5.2}$$

$$\mu_{ij} = \alpha_i - \alpha_j + \sum_{k=1}^{K} r_{ik}s_{jk} - s_{ik}r_{jk} \tag{4.5.3}$$

We want to show

1. in all models, (4.5.1–.3), $\hat{\alpha}_i = \frac{1}{2}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot i})$ is the least squares estimator,
2. the least squares estimator of $\boldsymbol{r}$, $\boldsymbol{s}$ in (4.5.2) is an eigenvector pair from $\boldsymbol{CC}^T$, and
3. once we have shown 2., this result can be extended to the 2nd, 3rd, … pair.

First,

$$SS = \sum_{ij} \left(Y_{ij} - \hat{h} - \hat{\mu}_{ij}\right)^2$$

$$= \sum_{ij} \left(z_{ij} - \hat{\mu}_{ij}\right)^2$$

$$= \sum_{ij} \left(z_{ij} - d_{ij} + d_{ij} - \hat{\mu}_{ij}\right)^2$$

$$= \sum_{ij} \left(z_{ij} - d_{ij}\right)^2 + \sum_{ij} \left(d_{ij} - \hat{\mu}_{ij}\right)^2 + 2\sum_{ij} \left(z_{ij} - d_{ij}\right)\left(d_{ij} - \hat{\mu}_{ij}\right).$$

Note that

$$\sum_{ij} \left(z_{ij} - d_{ij}\right)\left(d_{ij} - \hat{\mu}_{ij}\right) = \sum_{ij} \left(z_{ij} - d_{ij}\right) d_{ij} - \sum_{ij} \left(z_{ij} - d_{ij}\right)\hat{\mu}_{ij}$$

$$= \frac{1}{2}\sum_{ij} \left(z_{ij} - \frac{1}{2}\left(z_{ij} - z_{ji}\right)\right)\left(z_{ij} - z_{ji}\right) - \sum_{ij} \left(z_{ij} - \frac{1}{2}z_{ij} + \frac{1}{2}z_{ji}\right)\hat{\mu}_{ij}$$

$$= \frac{1}{4}\sum_{ij} \left(z_{ij} + z_{ji}\right)\left(z_{ij} - z_{ji}\right) - \frac{1}{2}\sum_{ij} \left(z_{ij} + z_{ji}\right)\hat{\mu}_{ij}$$

$$= \frac{1}{4}\left(\sum_{ij} z_{ij}^2 - \sum_{ij} z_{ji}^2\right) - \frac{1}{2}\left(\sum_{ij} z_{ij}\hat{\mu}_{ij} - \sum_{ji} z_{ji}\hat{\mu}_{ji}\right)$$

$$= \frac{1}{4}(0) - \frac{1}{2}(0) = 0.$$

So, while $\mu_{ij} = -\mu_{ji}$ in the balanced case ($n_{ij} = 1\ \forall i, j$),

$$SS = \sum_{ij} \left(z_{ij} - \hat{\mu}_{ij}\right)^2 = \sum_{ij} \left(z_{ij} - d_{ij}\right)^2 + \sum_{ij} \left(d_{ij} - \hat{\mu}_{ij}\right)^2 . \tag{4.5.4}$$

Note here that $d_{ij} = \frac{1}{2}\left(Y_{ij} - Y_{ji}\right)$ represents the average strength differential between teams i and j. So, $\sum_{ij}\left(z_{ij} - d_{ij}\right)^2$ represents the variation from individual games with respect to the mean difference between the teams after taking home court advantage into account and

$$\sum_{ij} \left(d_{ij} - \hat{\mu}_{ij}\right)^2 \tag{4.5.5}$$

represents the variation we seek to model.

If we define $\boldsymbol{D} = \left\{d_{ij}\right\}$, then $\boldsymbol{D}$ is skew-symmetric, as

$$d_{ij} = \frac{1}{2}\left(Y_{ij} - Y_{ji}\right) = -\frac{1}{2}\left(Y_{ji} - Y_{ij}\right) = -d_{ji}\ \forall i, j.$$

Additionally, we have skew-symmetric matrix $\widehat{\boldsymbol{M}} = \left\{\hat{\mu}_{ij}\right\}$ and let $\boldsymbol{C} = \boldsymbol{D} - \widehat{\boldsymbol{M}}$. We also know that the scalar multiple of a skew symmetric matrix (i.e., $-1 \times \widehat{\boldsymbol{M}}$), as well as the sum of skew-symmetric matrices (i.e., $\boldsymbol{D} + \left(-1 \times \widehat{\boldsymbol{M}}\right)$), result in a skew-symmetric matrix. So, $\boldsymbol{C} = \boldsymbol{D} - \widehat{\boldsymbol{M}}$ is skew-symmetric.

In Section 4.3.4, we prove that the least squares solution for $\hat{\alpha}_i$ is $\hat{\alpha}_i = \frac{1}{2}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot i})$ and that $\hat{\boldsymbol{r}}_k, \hat{\boldsymbol{s}}_k$ pairs must be paired eigenvectors of $\left(\boldsymbol{D} - \widehat{\boldsymbol{M}}\right)\left(\boldsymbol{D} - \widehat{\boldsymbol{M}}\right)^T = \boldsymbol{C}\boldsymbol{C}^T$ with $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{\alpha}}\mathbf{1}^T - \mathbf{1}\widehat{\boldsymbol{\alpha}}^T$. Now, consider fitting one pair of eigenvectors of $\boldsymbol{C}\boldsymbol{C}^T$ called $\boldsymbol{r}_\ell$ and $\boldsymbol{s}_\ell$. We want to show that these are the eigenvectors corresponding to the largest eigenvalue pair, λ_1, λ_1. Without loss of generality, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$ where $K = \left\lfloor \frac{n-1}{2} \right\rfloor$.

Now, notice that $\boldsymbol{D} - \widehat{\boldsymbol{M}} = \{d_{ij} - \hat{\mu}_{ij}\}$ and from (4.3.5.1) we know that

$$\sum_{ij}(d_{ij} - \hat{\mu}_{ij})^2 = \text{trace}\left((\boldsymbol{D} - \widehat{\boldsymbol{M}})(\boldsymbol{D} - \widehat{\boldsymbol{M}})^T\right) = \text{trace}(\boldsymbol{C}\boldsymbol{C}^T). \tag{4.5.6}$$

We also know that $\boldsymbol{r}_k^T\boldsymbol{r}_\ell = \boldsymbol{r}_k^T\boldsymbol{s}_\ell = \boldsymbol{s}_k^T\boldsymbol{r}_\ell = \boldsymbol{s}_k^T\boldsymbol{s}_\ell = 0\ \forall k, \ell, k \neq \ell$, and from (4.3.4.1) we have $\boldsymbol{C} = \sum_{k=1}^{K}\boldsymbol{r}_k\boldsymbol{s}_k^T - \boldsymbol{s}_k\boldsymbol{r}_k^T$. Then, letting $L = K$,

$$\text{trace}(\boldsymbol{C}\boldsymbol{C}^T) = \text{trace}\left(\left(\sum_{k=1}^{K}\boldsymbol{r}_k\boldsymbol{s}_k^T - \boldsymbol{s}_k\boldsymbol{r}_k^T\right)\left(\sum_{l=1}^{L}\boldsymbol{r}_l\boldsymbol{s}_l^T - \boldsymbol{s}_l\boldsymbol{r}_l^T\right)^T\right)$$

$$= \text{trace}\left(\left(\sum_{k=1}^{K}\boldsymbol{r}_k\boldsymbol{s}_k^T - \boldsymbol{s}_k\boldsymbol{r}_k^T\right)\left(\sum_{l=1}^{L}\boldsymbol{s}_l\boldsymbol{r}_l^T - \boldsymbol{r}_l\boldsymbol{s}_l^T\right)\right)$$

$$= \text{trace}\left(\sum_{k,l=1}^{K,L}(\boldsymbol{r}_k\boldsymbol{s}_k^T - \boldsymbol{s}_k\boldsymbol{r}_k^T)(\boldsymbol{s}_l\boldsymbol{r}_l^T - \boldsymbol{r}_l\boldsymbol{s}_l^T)\right)$$

$$= \text{trace}\left(\sum_{k,l=1}^{K,L}\boldsymbol{r}_k\boldsymbol{s}_k^T\boldsymbol{s}_l\boldsymbol{r}_l^T - \boldsymbol{r}_k\boldsymbol{s}_k^T\boldsymbol{r}_l\boldsymbol{s}_l^T - \boldsymbol{s}_k\boldsymbol{r}_k^T\boldsymbol{s}_l\boldsymbol{r}_l^T + \boldsymbol{s}_k\boldsymbol{r}_k^T\boldsymbol{r}_l\boldsymbol{s}_l^T\right)$$

$$= \text{trace}\left(\sum_{k=1}^{K}\boldsymbol{r}_k\boldsymbol{s}_k^T\boldsymbol{s}_k\boldsymbol{r}_k^T + \boldsymbol{s}_k\boldsymbol{r}_k^T\boldsymbol{r}_k\boldsymbol{s}_k^T\right)$$

$$= \text{trace}\left(\sum_{k=1}^{K}\boldsymbol{r}_k\lambda_k\boldsymbol{r}_k^T + \boldsymbol{s}_k\lambda_k\boldsymbol{s}_k^T\right)$$

$$= \sum_{k=1}^{K}\lambda_k\Big(\text{trace}(\boldsymbol{r}_k\boldsymbol{r}_k^T) + \text{trace}(\boldsymbol{s}_k\boldsymbol{s}_k^T)\Big)$$

$$= \sum_{k=1}^{K}\lambda_k(\boldsymbol{r}_k^T\boldsymbol{r}_k + \boldsymbol{s}_k^T\boldsymbol{s}_k)$$

$$= \sum_{k=1}^{K}\lambda_k(\lambda_k + \lambda_k)$$

$$= 2\sum_{k=1}^{K} \lambda_k^2$$

So, now we have

$$\sum_{ij} \left(d_{ij} - \hat{\mu}_{ij}\right)^2 = 2\sum_{k=1}^{K} \lambda_k^2 . \tag{4.5.7}$$

Since $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K$, we have that $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_K^2$. Clearly, choosing the eigenvector pair, r_1 and s_1, corresponding to the largest eigenvalue of $\boldsymbol{CC}^T$, λ_1, maximizes the amount of variation explained.

We can show this alternatively. We want to pick ℓ such that $\boldsymbol{r}_\ell$ and $\boldsymbol{s}_\ell$ maximize the amount of variation explained in $\boldsymbol{CC}^T$.

$$RSS = \sum_{ij} \left(d_{ij} - \left(\hat{\alpha}_i - \hat{\alpha}_j\right) - \left(r_{i\ell}s_{j\ell} - s_{i\ell}r_{j\ell}\right)\right)^2$$

Note that we can re-express $\boldsymbol{r}_\ell \boldsymbol{s}_\ell^T - \boldsymbol{s}_\ell \boldsymbol{r}_\ell^T$ as

$$\boldsymbol{r}_\ell \boldsymbol{s}_\ell^T - \boldsymbol{s}_\ell \boldsymbol{r}_\ell^T = \begin{bmatrix} r_{1\ell} \\ r_{2\ell} \\ \vdots \\ r_{n\ell} \end{bmatrix} \begin{bmatrix} s_{1\ell} & s_{2\ell} & \cdots & s_{n\ell} \end{bmatrix} - \begin{bmatrix} s_{1\ell} \\ s_{2\ell} \\ \vdots \\ s_{n\ell} \end{bmatrix} \begin{bmatrix} r_{1\ell} & r_{2\ell} & \cdots & r_{n\ell} \end{bmatrix}$$

$$= \begin{bmatrix} r_{1\ell}s_{1\ell} & r_{1\ell}s_{2\ell} & \cdots & r_{1\ell}s_{n\ell} \\ r_{2\ell}s_{1\ell} & r_{2\ell}s_{2\ell} & \cdots & r_{2\ell}s_{n\ell} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n\ell}s_{1\ell} & r_{n\ell}s_{2\ell} & \cdots & r_{n\ell}s_{n\ell} \end{bmatrix} - \begin{bmatrix} s_{1\ell}r_{1\ell} & s_{1\ell}r_{2\ell} & \cdots & s_{1\ell}r_{n\ell} \\ s_{2\ell}r_{1\ell} & s_{2\ell}r_{2\ell} & \cdots & s_{2\ell}r_{n\ell} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n\ell}r_{1\ell} & s_{n\ell}r_{2\ell} & \cdots & s_{n\ell}r_{n\ell} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & r_{1\ell}s_{2\ell} - s_{1\ell}r_{2\ell} & \cdots & r_{1\ell}s_{n\ell} - s_{1\ell}r_{n\ell} \\ r_{2\ell}s_{1\ell} - s_{2\ell}r_{1\ell} & 0 & \cdots & r_{2\ell}s_{n\ell} - s_{2\ell}r_{n\ell} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n\ell}s_{1\ell} - s_{n\ell}r_{1\ell} & r_{n\ell}s_{2\ell} - s_{n\ell}r_{2\ell} & \cdots & 0 \end{bmatrix}.$$

Here, we see that $\boldsymbol{r}_\ell \boldsymbol{s}_\ell^T - \boldsymbol{s}_\ell \boldsymbol{r}_\ell^T = \{r_{i\ell} s_{j\ell} - s_{i\ell} r_{j\ell}\}$ and that $\boldsymbol{r}_\ell \boldsymbol{s}_\ell^T - \boldsymbol{s}_\ell \boldsymbol{r}_\ell^T$ is a square, skew-symmetric matrix. From earlier, we have that $\boldsymbol{D} = \{d_{ij}\}$, $\widehat{\boldsymbol{M}} = \{\hat{\mu}_{ij}\}$, and $\boldsymbol{C} = \boldsymbol{D} - \widehat{\boldsymbol{M}} = \{d_{ij} - \hat{\mu}_{ij}\}$ where $\hat{\mu}_{ij} = \hat{\alpha}_i - \hat{\alpha}_j$.

So, now we have that

$$RSS = \sum_{ij} \left(d_{ij} - (\hat{\alpha}_i - \hat{\alpha}_j) - (r_{i\ell} s_{j\ell} - s_{i\ell} r_{j\ell}) \right)^2$$

$$= \text{trace}\left(\left(\boldsymbol{C} - (\boldsymbol{r}_\ell \boldsymbol{s}_\ell^T - \boldsymbol{s}_\ell \boldsymbol{r}_\ell^T) \right) \left(\boldsymbol{C} - (\boldsymbol{r}_\ell \boldsymbol{s}_\ell^T - \boldsymbol{s}_\ell \boldsymbol{r}_\ell^T) \right)^T \right)$$

$$= \text{trace}\left(\left(\sum_{\substack{k=1, \\ k \neq \ell}}^{K} \boldsymbol{r}_k \boldsymbol{s}_k^T - \boldsymbol{s}_k \boldsymbol{r}_k^T \right) \left(\sum_{\substack{l=1, \\ l \neq \ell}}^{L} \boldsymbol{r}_l \boldsymbol{s}_l^T - \boldsymbol{s}_l \boldsymbol{r}_l^T \right)^T \right)$$

$$= 2 \sum_{\substack{k=1, \\ k \neq \ell}}^{K} \lambda_k^2$$

$$= 2 \left(\left(\sum_{k=1}^{K} \lambda_k^2 \right) - \lambda_\ell^2 \right).$$

Clearly, choosing $\ell = 1$ reduces RSS by $2\lambda_1^2$, thereby maximizing the variation explained by $\boldsymbol{r}_\ell$ and $\boldsymbol{s}_\ell$.

4.6. Analysis of Variance

We can decompose all of the variation in the data, $\boldsymbol{Y}$, into parts. The first step is to identify how much variation the home advantage constant h explains.

$$SS_{TOT^*} = \sum_{ij} Y_{ij}^2$$

$$= \sum_{ij} \left(Y_{ij} - \hat{h} + \hat{h}\right)^2$$

$$= \sum_{ij} \left(Y_{ij} - \hat{h}\right)^2 - 2\hat{h} \sum_{ij} \left(Y_{ij} - \hat{h}\right) + \hat{h}^2 \sum_{ij} 1$$

$$= \sum_{ij} \left(Y_{ij} - \hat{h}\right)^2 + n^2\hat{h}^2$$

$$= SS_{TOT} + SS_{\hat{h}}$$

Specifically, we have

$$SS_{TOT} = \sum_{ij} \left(Y_{ij} - \hat{h}\right)^2 \tag{4.6.1}$$

and

$$SS_{\hat{h}} = n^2\hat{h}^2. \tag{4.6.2}$$

We leave the decomposition of $\sum_{ij}\left(Y_{ij} - \hat{h}\right)^2$ into $\sum_{ij}\left(Y_{ij} - \hat{h} - \hat{\mu}_{ij}\right)^2$ and a remainder as well as identifying the degrees of freedom for later.

CHAPTER 5:

AN APPLICATION OF THE CRSP MODEL IN THE NBA

Here, we see the CRSP model in action, performed on individual NBA seasons. The league is composed of two 15-team conferences, each of which is composed of three 5-team divisions. Special circumstances[8] excluded, every team plays 82 games in a regular season. Specifically, each team plays

- four games against all four opponents within a team's division;
- two games against every opponent in the other conference; and
- 36 games against the ten teams in the other two conference divisions (where four teams are randomly assigned to only three games).

Our current CRSP solution requires exactly two inputs per pair of teams. Ideally then, the method in its current form would be used for a double round-robin. To fit the data into the required structure, we have averaged replications of matchups in the same location. If needed, one could also use a single replication for every matchup. The effects on the variability of the parameter estimates requires further study in cases where $\exists n_{ij} \neq 1$.

[8] The 2010-2011 season was shortened to 66 games by the lockout and in 2013, a game to be played in Boston at the very end of the regular season was cancelled because of the Boston Marathon bombing and was not rescheduled.

5.1. A Simple Worked Example

We begin by looking at a single season — 2016-2017. Here, we model expected score differentials μ_{ij}, where $\mu_{ij} = \alpha_i - \alpha_j - r_i s_j - r_j s_i$; that is, the home effect $\hat{h}$ has been removed. We find that $\hat{h} = 2.887$. The remaining parameter estimates are in Table 5-1. How are we to interpret this? For one, the strength parameters should be like the point differential in a standings table one might find in a newspaper or on the internet, as these are calculated as in (5.1.1) where G_i is the number of games team i has played and X_{ig} is team i's score differential in the g^{th} game of the season,

$$\hat{a}_i = \frac{\sum_{g=1}^{G_i} X_{ig}}{G_i}. \tag{5.1.1}$$

Then, if team i were to hypothetically play the same schedule over and over, $\hat{a}_i$ would be biased by the schedule while $\hat{\alpha}_i$ would not.

Table 5-1: k=1 Parameter Estimates for each Team, 2016-2017 Season

Team (i)	Location	$\widehat{\boldsymbol{\alpha}}$	$\hat{\boldsymbol{r}}_1$	$\hat{\boldsymbol{s}}_1$
76ers	Philadelphia	−5.725	−0.014	2.503
Blazers	Portland	−0.425	0.592	−0.717
Bucks	Milwaukee	−0.458	−0.434	0.227
Bulls	Chicago	−0.558	−2.032	−3.039
Cavaliers	Cleveland	2.542	1.330	0.374
Celtics	Boston	2.417	−2.172	0.023
Clippers	Los Angeles	4.333	0.136	−3.014
Grizzlies	Memphis	0.283	0.713	1.223
Hawks	Atlanta	−1.400	1.668	2.864
Heat	Miami	0.558	1.767	0.647
Hornets	Charlotte	−0.008	−1.135	0.345
Jazz	Utah	4.858	2.352	−2.891
Kings	Sacramento	−3.708	0.164	1.025
Knicks	New York	−3.808	−0.791	1.093
Lakers	Los Angeles	−5.525	0.391	−1.368
Magic	Orlando	−6.658	0.852	−2.696
Mavericks	Dallas	−2.900	−2.606	−0.408
Nets	Brooklyn	−6.483	1.728	−0.510
Nuggets	Denver	0.975	−2.558	2.603
Pacers	Indiana	−0.600	−1.404	−2.691
Pelicans	New Orleans	−1.708	2.382	0.703
Pistons	Detroit	−0.483	3.644	0.987
Raptors	Toronto	3.083	−0.386	−0.705
Rockets	Houston	4.933	−0.138	0.792
Spurs	San Antonio	7.908	0.860	1.100
Suns	Phoenix	−5.067	−1.478	1.123
Thunder	Oklahoma City	1.492	0.736	−1.192
Timberwolves	Minnesota	−0.950	−1.027	1.737
Warriors	Golden State	11.267	−3.397	0.288
Wizards	Washington	1.817	0.252	−0.424

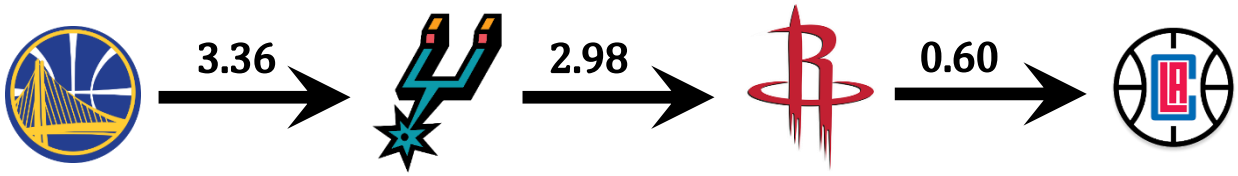
Also, consider for example the matchups between the Golden State Warriors (GSW), the San Antonio Spurs (SAS), and the Houston Rockets (HOU). Here, we see that the Chain

model (i.e., strength rating only; $k = 0$) is entirely transitive as $\hat{\alpha}_{GSW} - \hat{\alpha}_{HOU}$ could be calculated by $\hat{\alpha}_{GSW} - \hat{\alpha}_{SAS}$ and $\hat{\alpha}_{SAS} - \hat{\alpha}_{HOU}$,

$$\hat{\alpha}_{GSW} - \hat{\alpha}_{HOU} = (\hat{\alpha}_{GSW} - \hat{\alpha}_{SAS}) + (\hat{\alpha}_{SAS} - \hat{\alpha}_{HOU}) \approx 6.34.$$

As we can see in Figure 5-1, we can represent the relationships between the n teams with only $n - 1$ quantities, i.e., $\hat{\alpha}_{(2)} - \hat{\alpha}_{(1)}, \hat{\alpha}_{(3)} - \hat{\alpha}_{(2)}, \dots, \hat{\alpha}_{(n)} - \hat{\alpha}_{(n-1)}$.

Figure 5-1: Chain Example, Subset Visualization



We can also look at the intransitive contribution between GSW and SAS, for example, in the CRSP model,

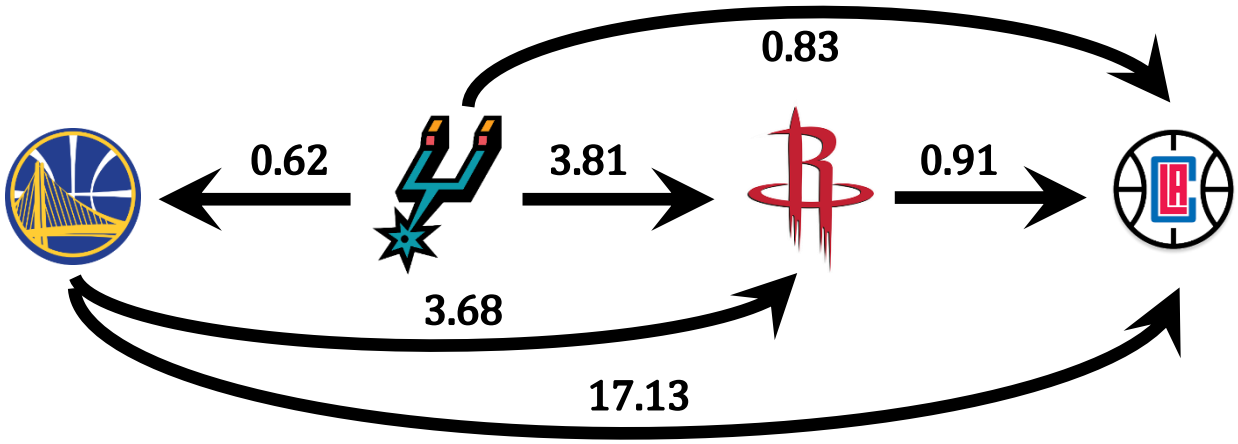
$$\hat{r}_{GSW}\hat{s}_{SAS} - \hat{r}_{SAS}\hat{s}_{GSW} \approx -3.40(1.10) - 0.86(0.29) \approx -3.98.$$

Then, the estimated strength differential between GSW and SAS is

$$\hat{\alpha}_{GSW} - \hat{\alpha}_{SAS} + \hat{r}_{GSW}\hat{s}_{SAS} - \hat{r}_{SAS}\hat{s}_{GSW} = 3.36 + (-3.98) = -0.62.$$

For the top teams from the Chain example visualization, we see that in the $k = 1$ CRSP example in Figure 5-2 that all the directional relationships remain the same, except between GSW and SAS. Here, the direction switches and now SAS is favored over GSW, albeit with a strength differential that can be overcome by home court advantage.

Figure 5-2: CRSP Example, Subset Visualization



Now, if we decide to use the generalized CRSP model and pick $k = 4$ $\boldsymbol{r}, \boldsymbol{s}$ pairs , then we get, for example, that GSW are favored over SAS by 8.02 points, where the parameter estimates can be found in Table 5-2. We do not show the $\boldsymbol{r}$ and $\boldsymbol{s}$ parameter estimates for $k \in \{5, .. 14\}$.

$$\hat{\alpha}_{GSW} - \hat{\alpha}_{SAS} + \sum_{k=1}^{4} \hat{r}_{GSW,k}\hat{s}_{SAS,k} - \hat{s}_{GSW,k}\hat{r}_{SAS,k}$$

$$= \hat{\alpha}_{GSW} - \hat{\alpha}_{SAS} + \hat{\boldsymbol{r}}^{T}_{GSW,1:4}\hat{\boldsymbol{s}}_{SAS,1:4} - \hat{\boldsymbol{s}}^{T}{}_{GSW,1:4}\hat{\boldsymbol{r}}_{SAS,1:4}$$

$$\approx 3.36 + \begin{bmatrix} -3.40 \\ 0.23 \\ -2.27 \\ 1.59 \end{bmatrix}^{T} \begin{bmatrix} 1.10 \\ 2.18 \\ -0.97 \\ 1.32 \end{bmatrix} - \begin{bmatrix} 0.29 \\ 0.87 \\ -1.27 \\ 0.69 \end{bmatrix}^{T} \begin{bmatrix} 0.86 \\ -2.10 \\ 2.07 \\ 0.90 \end{bmatrix} \approx 3.36 + 4.66 \approx 8.02.$$

Table 5-2: Generalized CRSP r,s Parameter Estimates for k=4

Team	$\hat{r}_1$	$\hat{r}_2$	$\hat{r}_3$	$\hat{r}_4$	$\hat{s}_1$	$\hat{s}_2$	$\hat{s}_3$	$\hat{s}_4$
76ers	−0.014	−1.349	0.014	−0.067	2.503	−0.153	−0.623	−2.111
Blazers	0.592	1.185	0.490	−0.046	−0.717	1.698	−0.367	0.402
Bucks	−0.434	−3.092	−2.291	−0.581	0.227	−2.386	−0.684	−0.779
Bulls	−2.032	−2.405	0.940	1.131	−3.039	1.251	−0.770	1.484
Cavaliers	1.330	2.125	−1.772	0.477	0.374	2.349	0.513	−1.468
Celtics	−2.172	2.272	−0.763	−0.891	0.023	−0.093	0.278	−1.052
Clippers	0.136	−2.149	0.568	−1.637	−3.014	1.096	0.597	−0.761
Grizzlies	0.713	−1.100	0.323	−1.297	1.223	−0.870	−0.599	1.164
Hawks	1.668	−0.967	0.375	3.318	2.864	−1.548	0.146	0.490
Heat	1.767	0.333	−0.656	0.922	0.647	−2.080	−1.194	1.116
Hornets	−1.135	0.372	0.907	0.386	0.345	1.127	1.971	0.978
Jazz	2.352	0.336	−0.393	1.210	−2.891	−1.307	−0.830	0.827
Kings	0.164	−1.837	−0.218	−0.587	1.025	2.556	1.273	−0.074
Knicks	−0.791	0.308	−0.419	−1.496	1.093	−1.940	1.438	1.086
Lakers	0.391	1.748	0.064	−1.703	−1.368	−1.828	−2.553	1.694
Magic	0.852	0.010	1.425	−0.478	−2.696	−1.138	0.189	−0.465
Mavericks	−2.606	1.405	2.416	−1.824	−0.408	−1.071	0.213	−0.876
Nets	1.728	0.331	−1.497	−1.790	−0.510	0.725	−1.937	−2.002
Nuggets	−2.558	1.279	−0.124	0.711	2.603	0.700	−2.506	−0.946
Pacers	−1.404	−0.019	−2.692	2.236	−2.691	−1.516	3.249	−0.501
Pelicans	2.382	0.622	−2.230	−1.516	0.703	0.372	0.445	0.864
Pistons	3.644	1.384	0.780	−0.196	0.987	0.979	2.265	0.182
Raptors	−0.386	2.500	1.424	1.466	−0.705	0.910	−0.073	−0.994
Rockets	−0.138	−2.121	1.275	−0.411	0.792	−1.100	1.219	−3.444
Spurs	0.860	−2.103	2.078	0.903	1.100	2.181	−0.970	1.317
Suns	−1.478	0.108	−0.962	−1.564	1.123	1.914	1.394	2.717
Thunder	0.736	1.518	1.301	1.773	−1.192	0.020	−0.693	−1.054
Timberwolves	−1.027	0.718	1.901	−0.182	1.737	−2.874	1.224	1.234
Warriors	−3.397	0.225	−2.278	1.595	0.288	0.869	−1.270	0.690
Wizards	0.252	−1.638	0.015	0.134	−0.424	1.156	−1.346	0.281

5.2. Selecting the Number of r, s Pairs

So how do we choose k in the generalized CRSP model? This question has endured some amount of discussion and will not be resolved here. Methods include using a graphical elbow, like a scree plot in PCA or using R^2 in regression, picking a percentage of variation cutoff in advance, using an information criterion like AIC or BIC, and using an ANOVA method and perform F-tests. In some way or another, we will want to use the information in Table 5-3.

Table 5-3: Variation Explained by each r, s pair

k	Intransitive Variation Explained by Pair	Cumulative Proportion of Total Intransitive Variation	Pair Signs
1	12633	0.264	1
2	9426	0.461	-1
3	5972	0.586	-1
4	5383	0.699	-1
5	3930	0.781	1
6	3035	0.844	-1
7	2722	0.901	1
8	1692	0.937	-1
9	1309	0.964	1
10	774	0.98	1
11	539	0.991	-1
12	290	0.997	1
13	76	0.999	-1
14	47	1	-1

We arbitrarily chose to represent 70% of the intransitive variation (that is, the variation from pairs of teams that disagrees with the transitive Chain model) and got $k = 4$.

We also examined the generalized CRSP model for each season from 2004-2005 to 2016-2017. Interestingly, the 2016-2017 NBA season given as an example in Section A Simple Worked Example5.1 had the most intransitive variation. This could be because the league was in a state of flux at the time — the GSW revolutionized "small ball" basketball which involved zero or one big men (generally, players 6'10" or taller) while other teams were stuck in the old NBA and played with two "big men."

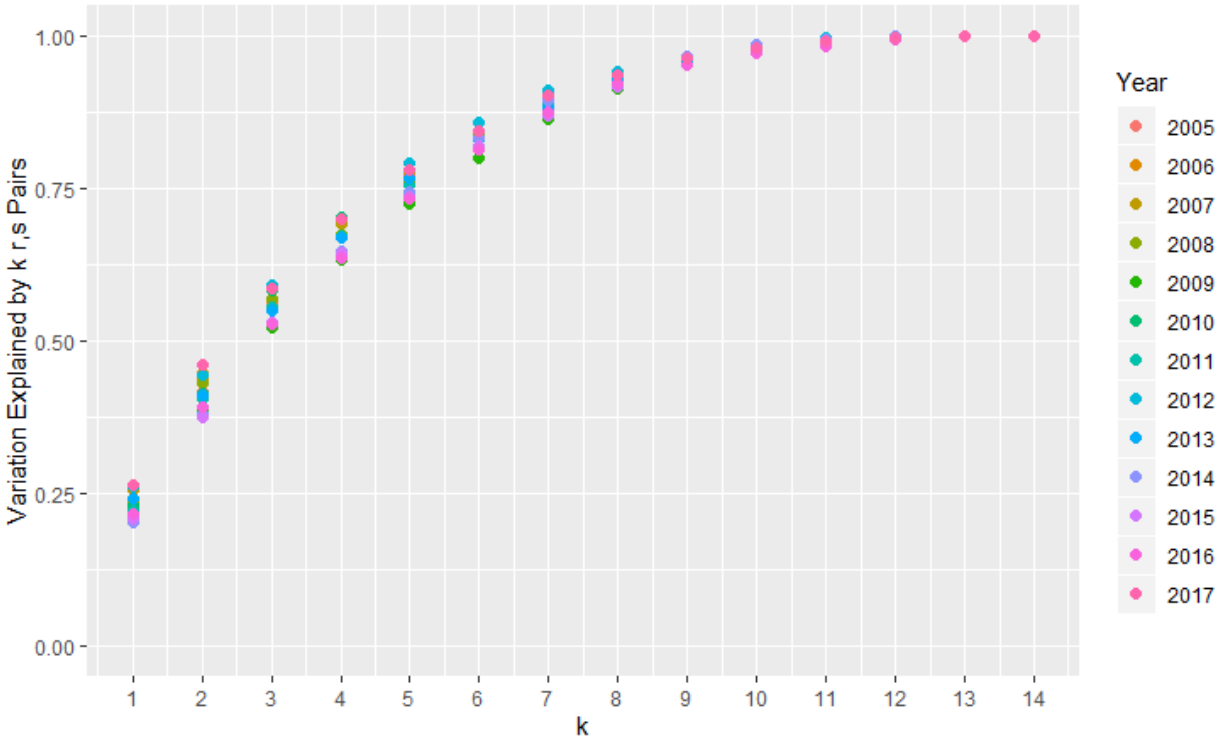
Table 5-4: Generalized CRSP by NBA Season

Year	Percent of Total Intransitive Variation Explained by up to k r, s pairs					Total Intransitive Variation
	$k \leq 1$	$k \leq 2$	$k \leq 3$	$k \leq 4$	$k \leq 5$	
2005	0.237	0.415	0.564	0.674	0.769	36241
2006	0.259	0.447	0.566	0.674	0.762	40992
2007	0.256	0.437	0.569	0.691	0.778	43142
2008	0.233	0.429	0.569	0.676	0.762	40949
2009	0.22	0.387	0.521	0.634	0.723	40300
2010	0.231	0.413	0.582	0.701	0.78	42898
2011	0.222	0.404	0.556	0.671	0.756	39759
2012	0.261	0.444	0.59	0.699	0.791	36823
2013	0.241	0.412	0.549	0.669	0.765	46119
2014	0.201	0.381	0.527	0.648	0.744	36680
2015	0.209	0.375	0.53	0.645	0.736	41558
2016	0.218	0.392	0.527	0.635	0.734	42725
2017	0.264	0.461	0.586	0.699	0.781	**47829**

We also see in Figure 5-3: Comparison of Cumulative Intransitive Variation over k and by YearFigure 5-3 that the eigendecomposition is remarkably consistent in the amount of variation explained across years. It is not clear whether that is due to the nature of the

NBA, some properties of the eigendecomposition, or both. The year used as the example, 2017, is represented using pink dots.

Figure 5-3: Comparison of Cumulative Intransitive Variation over k and by Year



5.3. Cartesian Coordinate Representation of r, s Pairs

We can represent the intransitive relationship between every team in a fairly simple plot. At least, the visual representation is simple for the basic CRSP model ($k = 1$). First, note that each team i's intransitive pair, r_i and s_i, can be plotted on a Cartesian plane with orthogonal axes R and S. Then, the magnitude of the resulting vector (r_i, s_i) is

$$\|\hat{v}_i\| = \sqrt{\hat{r}_i^2 + \hat{s}_i^2} \tag{5.3.1}$$

and the vector's counter-clockwise rotation from the positive R axis is

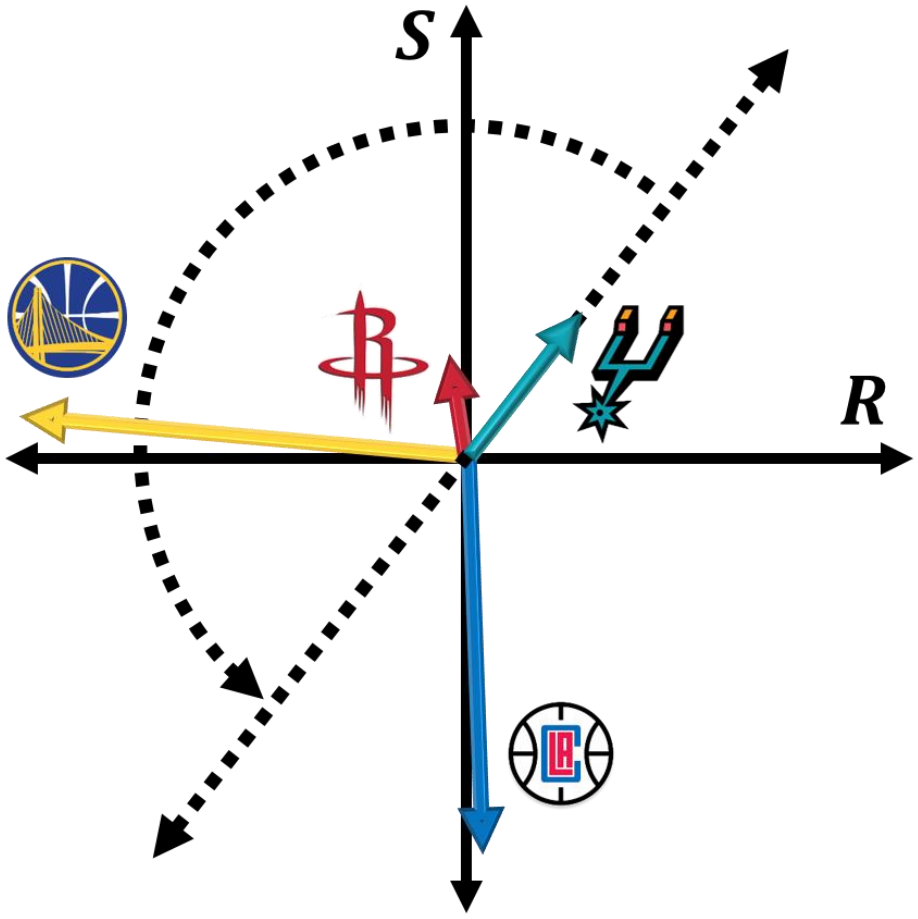
$$\hat{\theta}_{\mathrm{i}} = \tan^{-1}\left(\frac{s_{\mathrm{i}}}{r_{\mathrm{i}}}\right). \tag{5.3.2}$$

Now, continuing the example from Section 5.1, we examine the top four teams from the 2016-2017 season. The teams are described in Table 5-5 and plotted in Figure 5-4.

Table 5-5: Small Cartesian Example, 2016-2017 NBA Season

i	$\hat{r}_i$	$\hat{s}_i$	$\|\hat{v}_i\|$	$\hat{\theta}_i$
GSW	-3.40	0.29	2.10	175°
SAS	0.86	1.10	0.86	52°
HOU	-0.13	0.79	0.49	99°
LAC	0.13	-3.01	1.86	272°

Figure 5-4: Small Cartesian Example, 2016-2017 NBA Season



We notice that $r_i s_j - r_j s_i \geq 0 \ \ \forall j$ where $0° < \hat{\theta}_j - \hat{\theta}_i < 180°$. That is, team i gets a positive intransitive contribution to its strength differential estimate from any team j within 180° counter-clockwise of team i. So, for example, SAS gets a positive intransitive contribution from the Houston Rockets and GSW, but gets a negative intransitive contribution from the Los Angeles Clippers. Notice also that $r_i s_j - r_j s_i$ maximized when $\hat{\theta}_j - \hat{\theta}_i = 90°$ (when holding $\|\hat{v}_i\|$ and $\|\hat{v}_j\|$ constant). So, the largest contributions come from $\|\hat{v}_i\|$ and $\|\hat{v}_j\|$ large and $\hat{\theta}_j - \hat{\theta}_i \approx 90°$. This is why SAS got such a large contribution from GSW.

Figure 5-4 is easy to read with only four teams, however there were 30 teams in the model. We used one color per vector to distinguish the four teams in Figure 5-4, but with 30 teams (many choosing red or blue as their primary color), we need to find a way to distinguish the teams' vectors. To address this, we added a secondary (or tertiary) team color to the vector in Figure 5-5.

Observe that the original relationships and positions of the four teams from Figure 5-4 are still in Figure 5-5. Also, observe that the teams with the largest positive contribution from intransitive components in relation to the GSW (i.e., about 90° clockwise of GSW and have large $||\hat{v}_i||$) were teams that notoriously played their "big men" — Atlanta Hawks (Dwight Howard), Memphis Grizzlies (Marc Gasol and Zack Randolph), Sacramento Kings (DeMarcus Cousins), Philadelphia 76ers (Joel Embiid), and San Antonio Spurs (LaMarcus Aldridge and Pau Gasol). The Warriors had a strength rating of $\hat{\alpha}_{GSW} = 11.27$, which was substantially larger than every other team. They were incredibly fast and athletic (and relatively short) and were able to overwhelm other teams in those ways. However, against the aforementioned larger teams, this was less of an advantage and is thus reflected as a modeled decrease between those teams. It might also be helpful to display subsets of teams, like in Eastern and Western Conferences as in Figure 5-6 and Figure 5-7.

Figure 5-5: Complete Cartesian Example, 2016-2017 NBA Season

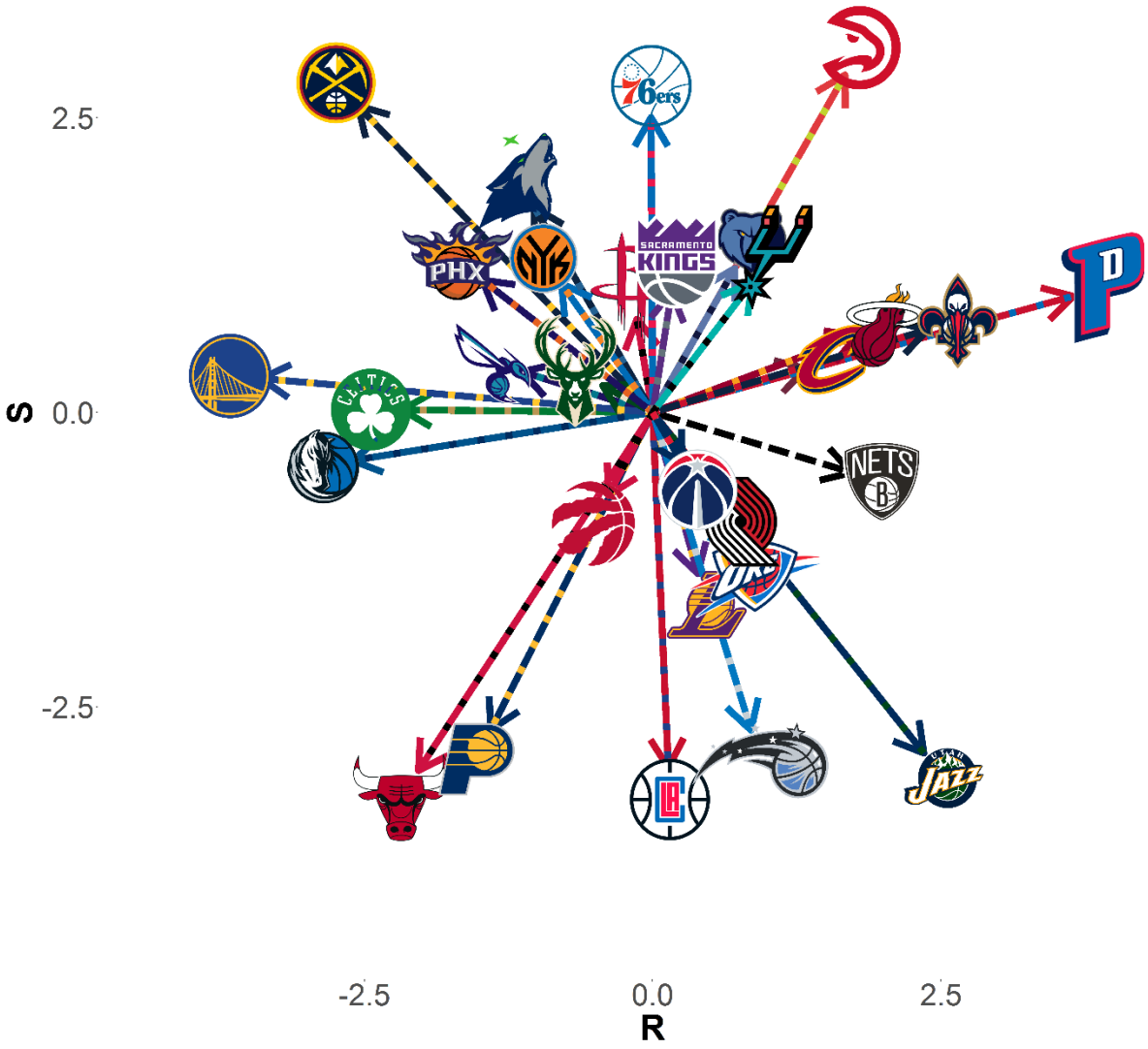


Figure 5-6: Complete Cartesian Example, Eastern Conference

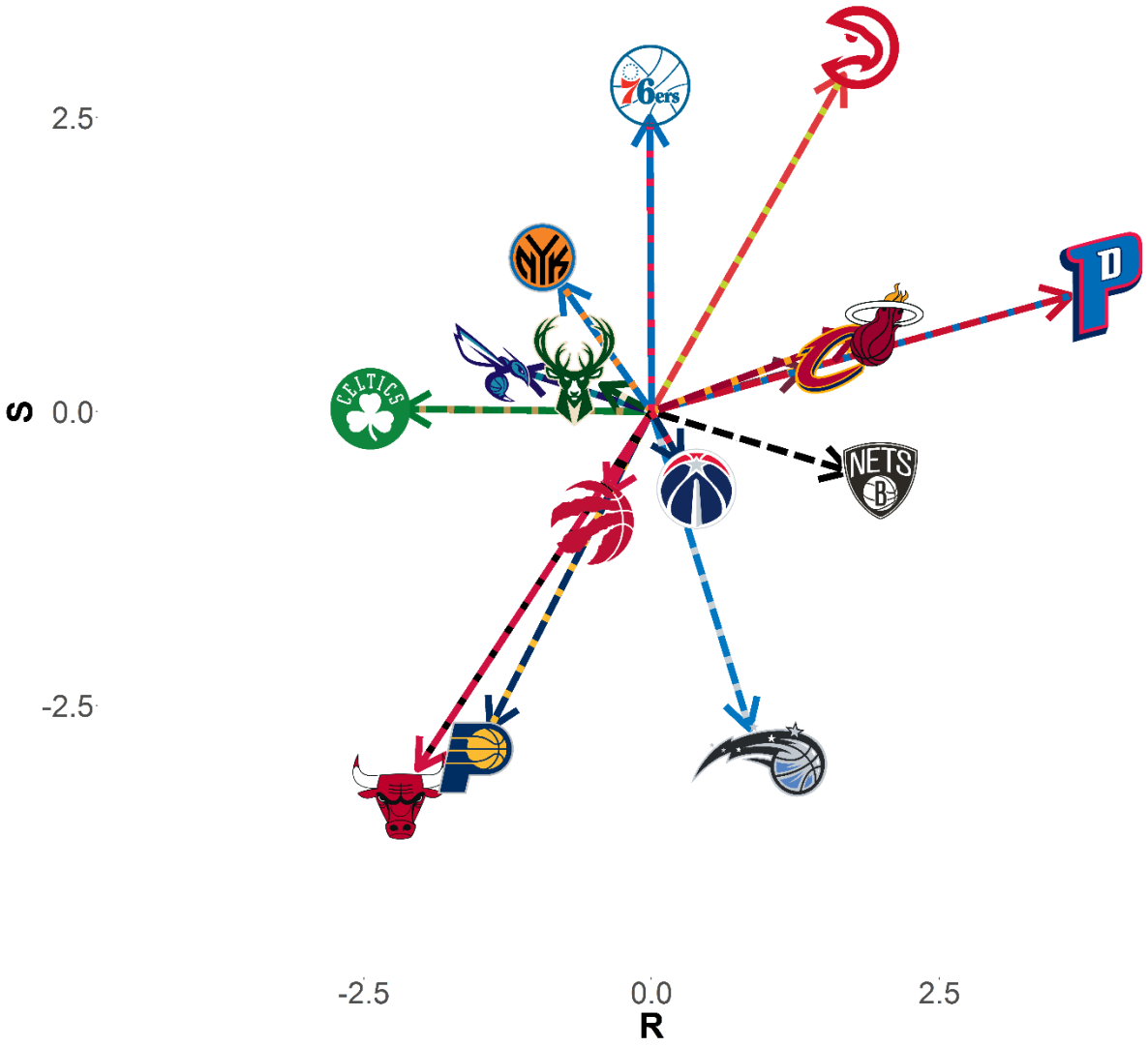
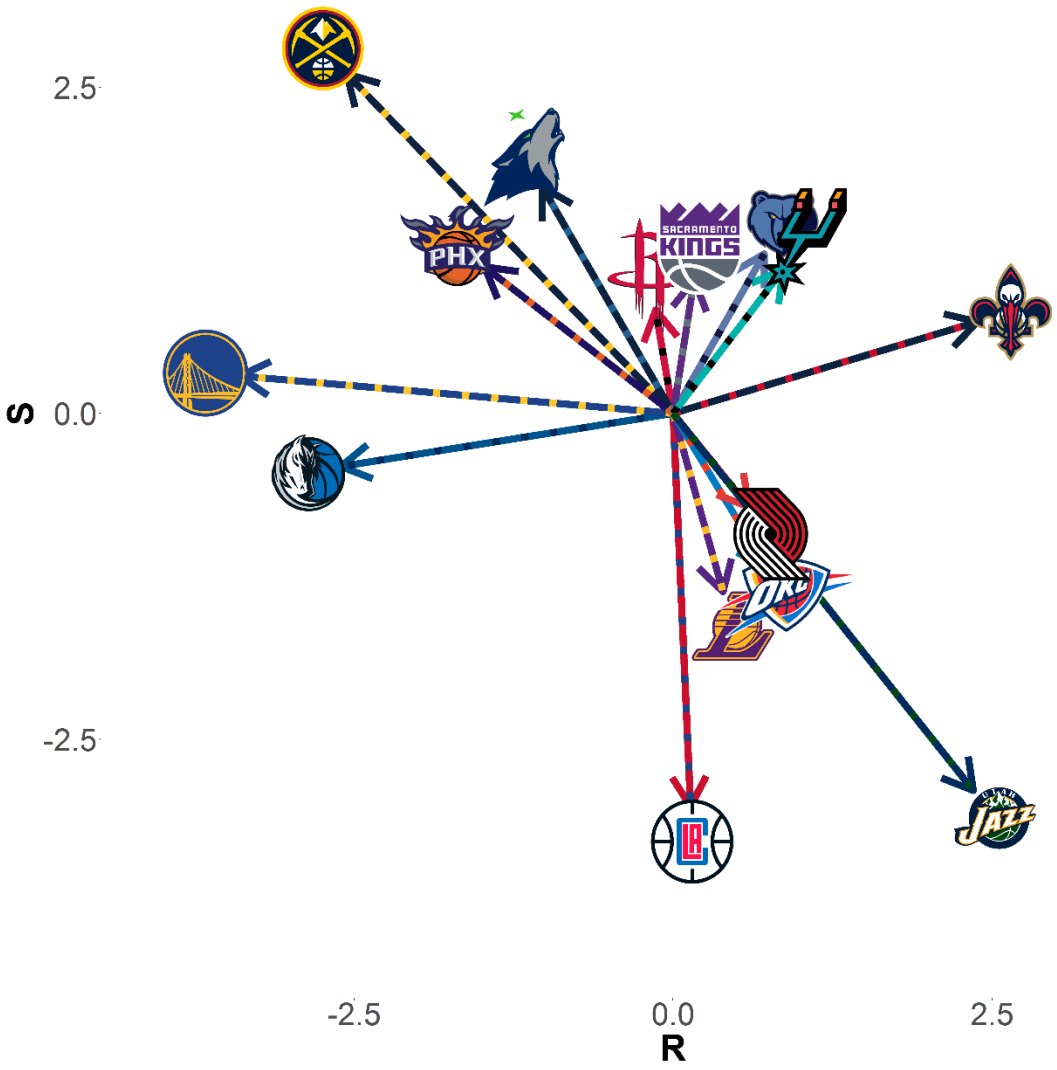


Figure 5-7: Complete Cartesian Example, Western Conference



CHAPTER 6:

CIRCULAR TRIAD COUNT DISTRIBUTION FOR LARGER n

Our initial goal was to recreate the enumerated distribution of circular triads, d, found in Table II (Kendall and Smith, 1940, p. 333, included in Chapter 2), which goes to $n = 7$. We will examine details that we skipped in Section 2.1.1, prove select statements, and apply that knowledge to write a speedy algorithm to replicate Table II to $n = 13$. We go beyond the $n = 10$ case in Kendall and Gibbons' fifth edition of their book (1990), originally published in 1948.

The algorithm independently developed here is remarkably similar to Alway (1962). Also, we run into space constraints in the $n = 14$ case. Knezek et al. (1998) explore up to the $n = 15$ case, but do not explicitly publish the d distribution in their work. To address the issues of space and to extend the d distribution to cases larger than $n = 15$, we propose modifying the algorithm with parallelization and smaller matrices of new edges.

6.1. Theoretical Examination of Kendall and Smith

Kendall and Smith define the "alpha numbers" (α-numbers) as the number of edges leaving a vertex in a graph. In graph theory, this is called the out-degree. An α-number for item r is the number of directed edges leaving item r, or the number of times item r was

preferred to other items. Then, $\alpha \in (0,1,\dots,n-1)$. Since there are $\binom{n}{2}$ pairs to compare, $\sum_{r=1}^{n}\alpha_r = \binom{n}{2} = \frac{n(n-1)}{2}$. Kendall then gives the mean and variance of the α-vectors, but calls the latter T. We derive these here:

$$E[\alpha_r] = \frac{\sum_{r=1}^{n}\alpha_r}{n} = \frac{n(n-1)/2}{n} = \frac{n-1}{2}$$

$$T = Var(\alpha_r)$$

$$= \sum_{r=1}^{n}(\alpha_r - E[\alpha_r])^2$$

$$= \sum_{r=1}^{n}\left(\alpha_r - \frac{n-1}{2}\right)^2$$

$$= \sum_{r=1}^{n}(\alpha_r)^2 - (n-1)\alpha_r + \frac{1}{4}(n-1)^2$$

$$= \sum_{r=1}^{n}(\alpha_r)^2 - (n-1)\sum_{r=1}^{n}\alpha_r + \frac{1}{4}(n-1)^2\sum_{r=1}^{n}1$$

$$= \sum_{r=1}^{n}(\alpha_r)^2 - (n-1)\cdot\frac{n-1}{2} + \frac{1}{4}(n-1)^2 n$$

$$= \sum_{r=1}^{n}(\alpha_r)^2 - \frac{n(n-1)^2}{4}$$

We also note that max (T) is achieved when the judge gives one of the $n!$ perfect rankings and the α-numbers are $0,1,\dots,n-1$. In a perfect ranking, $d = 0$ — there are no inconsistencies. Then the minimum of T is achieved when the α-numbers are most alike—that is, when max (d) is achieved. Then, as d increases, T decreases, and vice versa. So, we show:

$$\max(T) = \sum_{i=0}^{n-1} i^2 - \frac{n(n-1)^2}{4}$$

$$= 0 + \frac{(n-1)n(2n-1)}{6} - \frac{n(n-1)^2}{4}$$

$$= (n-1)\left(\frac{4n^2-2n}{12} - \frac{3n^2-3n}{12}\right)$$

$$= \frac{n(n-1)}{12}(n+1)$$

$$= \frac{n^3-n}{12}$$

By inspecting $T = \sum_{r=1}^{n}\left(\alpha_r - \frac{n-1}{2}\right)^2$, we can see that the closer the α_r are to the mean $\frac{n-1}{2}$, the smaller T becomes. Kendall and Smith show what graph configuration would minimize T, but we simply note that the α_r are subject only to the constraints $\alpha_r \in (0,1,\dots,n-1)$ and $\sum_{r=1}^{n}\alpha_r = \frac{n(n-1)}{2}$. So, if n is odd, $\alpha_r = \frac{n-1}{2}\ \forall r$, meets those conditions and $T = 0$. If n is even, then $\frac{n-1}{2}$ is not an integer, but we can have half of the α_r half a unit below the mean and have the other half of the α_r half a unit above the mean. Then the α_r meet the constraints and T will be as small as possible. When n is even, without loss of generality, this gives us:

$$\min(T) = \sum_{r=1}^{\frac{n}{2}}\left(\frac{n-2}{2} - \frac{n-1}{2}\right)^2 + \sum_{r=\frac{n}{2}+1}^{n}\left(\frac{n}{2} - \frac{n-1}{2}\right)^2$$

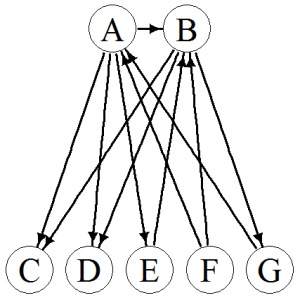
$$= \frac{1}{4}\sum_{r=1}^{\frac{n}{2}}(-1)^2 + \frac{1}{4}\sum_{r=\frac{n}{2}+1}^{n}(1)^2 = \frac{1}{4}\sum_{r=1}^{n}1 = \frac{n}{4}$$

We have shown $\max(T) = \frac{n^3-n}{12}$ and $\min(T) = \begin{cases} 0 & \text{if } n \text{ odd} \\ \frac{n}{4} & \text{if } n \text{ even} \end{cases}$. We return to Kendall and Smith's (1940) argument to find $\max(d)$, beginning by examining the example in

Figure 6-1. Here, we let $\alpha = \alpha_1$ be the number of edges leaving vertex A and $\beta = \alpha_2$ be the number of edges leaving vertex B. We are assuming that $A \rightarrow B$, and have selected an arbitrary number of other vertices with arbitrarily directed edges for demonstration purposes. We are only interested in what will happen to the number of circular triads when we flip $A \rightarrow B$ to $A \leftarrow B$. The only edges that matter, then, are the edge between A and B, and all the edges connected to A and B. Including $A \rightarrow B$, $\alpha = 3 + 1 = 4$. Including $A \rightarrow B$, $\beta = 3 + 0 = 3$. We note that there are $(n - 2)$ vertices outside of A and B for which we must eventually account.

Figure 6-1: Edge Count Example

Type	Triad Form	Figure example list
1	$A \rightarrow X \leftarrow B$	$X = C, D$
2	$A \leftarrow X \rightarrow B$	$X = F$
3	$A \rightarrow X \rightarrow B$	$X = G$
4	$A \leftarrow X \leftarrow B$	$X = E$



Let p be the number of triads of type 1 (Figure 6-1). Triads of type 3 are the same as type 1, except the (B, X) edge has been flipped. Edges of triad types 1 and 3 as well as $A \rightarrow B$ are included in α. Then the number of triads of type 3 is $\alpha - p - 1$. Similarly, triads of type 4 are the same as type 1, except the (A, X) edge has been flipped. β currently only includes edges of triad types 1 and 4. Then the number of triads of type 4 is $\beta - p$. Lastly, though not strictly necessary, the number of triads of type 2 can be counted by subtracting the number of edges we have counted so far from $n - 2$. That is, there are $(n - 2) - \big(p + (\alpha - p - 1) + (\beta - p)\big) = n + p - (\alpha + \beta + 1)$ edges of type 2. Triads containing $A \rightarrow$

B are only circular in type 4. Now, if we change $A \to B$ to $A \leftarrow B$, triads of type 4 stop being circular, and triads of type 3 start being circular. So, the number of circular triads involving the edge (A, B) increases by the new amount of inconsistencies $(\alpha - p - 1)$ minus the old amount of inconsistencies $(\beta - p)$. That is,

$$\Delta d = (\alpha - p - 1) - (\beta - p) = a - \beta - 1.$$

Also, note that $T = \sum_{r=1}^{n} (\alpha_r)^2 - \frac{n(n-1)^2}{4} = \sum_{r=1}^{2} (\alpha_r)^2 + \sum_{r=3}^{n} (\alpha_r)^2 - \frac{n(n-1)^2}{4} = T_1 + T_2$, where $T_1 = \sum_{r=1}^{2} (\alpha_r)^2$. So, the only term in T affected by changing the direction of the edge between A and B is T_1. From earlier, $T_1 = \alpha_1^2 + \alpha_2^2 = \alpha^2 + \beta^2$. So when $A \to B$ changes to $A \leftarrow B$, α decreases by one, β increases by one, and T_1 (and subsequently T) changes by

$$\Delta T = (\alpha - 1)^2 + (\beta + 1)^2 - (\alpha^2 + \beta^2) = -2\alpha + 1 + 2\beta + 1 = -2(\alpha - \beta - 1) = -2\Delta d$$

That is, $\Delta T = -2\Delta d$. This is true for any change in d or change in T. Maximizing ΔT gives us:

$$\max(T) - \min(T) = 2(\max(d) - \min(d))$$

$$\Rightarrow \max(d) = \frac{n^3 - n - 12 \cdot \min(T)}{24}.$$

We substitute what we know, solve for $\max(d)$, and reiterate that $\min(T) = \begin{cases} 0 & \text{if } n \text{ odd} \\ \frac{n}{4} & \text{if } n \text{ even} \end{cases}$. Then

$$\max(d) = \begin{cases} \dfrac{n^3 - n}{24} & \text{if } n \text{ odd} \\ \dfrac{n^3 - 4n}{24} & \text{if } n \text{ even.} \end{cases}$$

Finally, we note that for some d,

$$\max(T) - T = 2(d - \min(d)).$$

$$\Rightarrow d = \frac{1}{2}\left(\frac{n^3 - n}{12} - \left(\sum_{r=1}^{n}(\alpha_r)^2 - \frac{n(n-1)^2}{4}\right)\right)$$

$$= \frac{1}{2}\left(\frac{n(n-1)}{4}\left(\frac{n+1}{3} + \frac{3(n-1)}{3}\right) - \sum_{r=1}^{n}(\alpha_r)^2\right)$$

$$d = \frac{n(n-1)(2n-1)}{12} - \frac{1}{2}\sum_{r=1}^{n}(\alpha_r)^2 . \qquad (6.1.1)$$

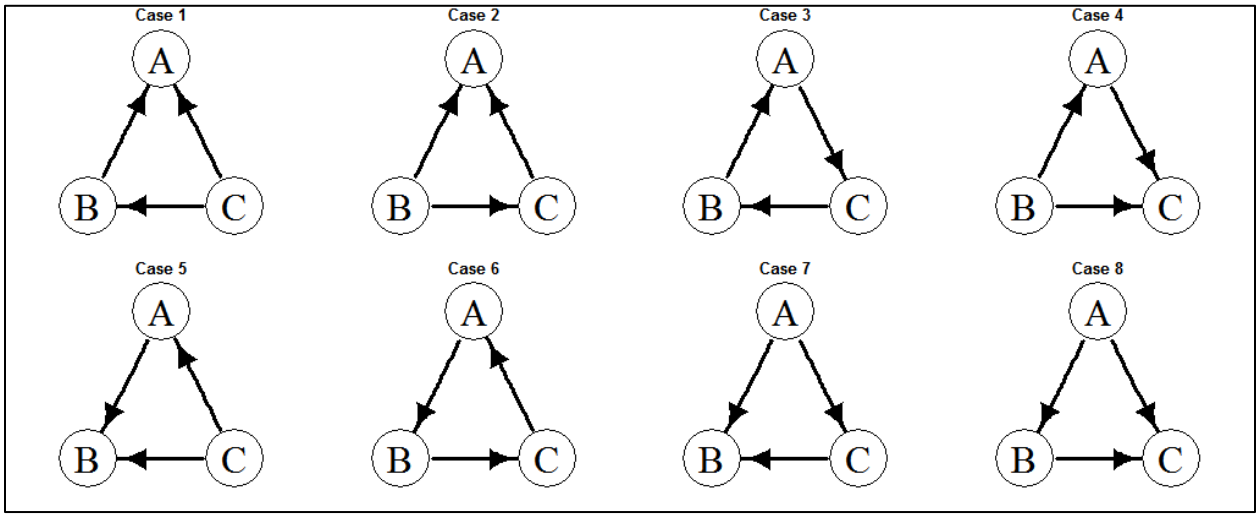
That is, we do not need to count the d directly anymore; we simply need n and an α vector to calculate d! Furthermore, the order of the α vector does not affect d, only the counts α_i, j. Thus, (6.1.1) gives the number of circular triads for the score vector of any tournament.

6.2. Algorithm to Extend to n=13

We propose various algorithms to extend the d distribution.

6.2.1. Exhaustive Tournament Search Algorithm

Figure 6-2: All Possible Triads



These triads can be represented with adjacency matrices and only Case 3 and Case 6 are circular triads. In their adjacency matrix form, they can be represented as:

$$\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix} \text{ and } \begin{bmatrix}0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}, \text{respectively.} \qquad (6.2.1.1)$$

If we check any given triad in a larger adjacency matrix, they should be equivalent to one of the two forms in (6.2.1.1). However, this is slow even in bit-vectorized form as there are $2^{\binom{n}{2}}$ checks to be made. A preliminary exhaustive search algorithm was timed and recorded in Table 6-1. Notice that the $n = 8$ case alone took 7099 minutes (just under 5 days). It would take on the order of *4 million years* under current computing conditions to obtain the result for the $n = 11$ case and about 2 years for the $n = 10$—certainly not satisfactory.

Table 6-1: Replication of Kendall and Gibbons (1990) Appendix Table II - Exhaustive Algorithm with Adjacency Matrices

n		3	4	5	6	7	8
Minutes elapsed		1.67E-05	2.00E-04	0.005933	0.3307	34.80	7099.72
Total		8	64	1024	32768	2097152	268435456
d	0	6	24	120	720	5040	40320
	1	2	16	120	960	8400	80640
	2		24	240	2240	21840	228480
	3			240	2880	33600	403200
	4			280	6240	75600	954240
	5			24	3648	90384	1304576
	6				8640	179760	3042816
	7				4800	188160	3870720
	8				2640	277200	6926080
	9					280560	8332800
	10					384048	15821568
	11					244160	14755328
	12					233520	24487680
	13					72240	24514560
	14					2640	34762240
	15						29288448
	16						37188480
	17						24487680
	18						24312960
	19						10402560
	20						3230080

6.2.2. Tournament Simulation

Since 4 million years for the $n = 11$ case is obviously too long, we attempted a simulation instead. Instead of exhaustively searching all $2^{\binom{n}{2}}$ cases, we can just examine M cases which are randomly generated. To do this, we simply need to generate M random binary vectors of length $\binom{n}{2}$ — one bit for every edge whose direction we need to simulate. In Kendall and Gibbons' 1990 book, a χ^2 approximation to the d distribution is given. For

the simulation to have utility, it would have to improve upon the χ^2 approximation. In Appendix Table 9, Kendall and Gibbons (1990) give the χ^2 approximation for the $n = 10$ case. The probabilities are within 0.005 for $P > 0.90$. So, we chose $M = 10^6$ for greater precision than the χ^2 approximation. In Table 6-2, we see that as n increases, so too does the amount of time elapsed.

Table 6-2: Simulation Performance

Table $\boldsymbol{n}$	**8**	**9**	**10**	**11**	**12**
Minutes elapsed	29.33	42.48	60.15	78.88	105.13
Number of unique adjacency matrices	998158	999996	1000000	1000000	1000000

In Table 6-3, we examine the results of the $n = 10$ case. The first column contains the frequencies of the d from the simulation, the second column contains the corresponding P values, calculated by finding the probability of d or more circular triads, and the third column contains the exact P values, obtained with the final algorithm. In the fourth column, we determine if the simulation's absolute error is smaller than the χ^2 approximation's absolute error. In all but one case, $d = 28$, the simulation was closer than the χ^2 approximation.

Table 6-3: Replication of Kendall and Gibbons (1990) Appendix Table II - Simulation with Adjacency Matrices; $n = 10$

d	Simulated Freq	Simulated P	$\sim\chi^2$ P	Exact P	Sim Closer than $\sim\chi^2$
0	0	1	0.999994	1	TRUE
1	0	1	0.999991	1	TRUE
2	0	1	0.999986	1	TRUE
3	2	1	0.999977	0.999999	TRUE
4	3	0.999998	0.999963	0.999997	TRUE
5	14	0.999995	0.999942	0.999992	TRUE
6	22	0.999981	0.999909	0.999983	TRUE
7	34	0.999959	0.999858	0.999963	TRUE
8	72	0.999925	0.999779	0.999929	TRUE
9	126	0.999853	0.999657	0.999861	TRUE
10	217	0.999727	0.999473	0.999753	TRUE
11	286	0.99951	0.999194	0.999542	TRUE
12	548	0.999224	0.998776	0.999242	TRUE
13	764	0.998676	0.998153	0.998687	TRUE
14	1299	0.997912	0.997234	0.997903	TRUE
15	1822	0.996613	0.99589	0.996564	TRUE
16	3057	0.994791	0.993941	0.994744	TRUE
17	3822	0.991734	0.991144	0.99169	TRUE
18	6270	0.987912	0.987174	0.987869	TRUE
19	7720	0.981642	0.981601	0.981646	TRUE
20	11840	0.973922	0.973875	0.973907	TRUE
21	14234	0.962082	0.963308	0.96204	TRUE
22	21292	0.947848	0.949059	0.947997	TRUE
23	23330	0.926556	0.930146	0.926872	TRUE
24	34243	0.903226	0.905467	0.903289	TRUE
25	38086	0.868983	0.873856	0.869118	TRUE
26	50832	0.830897	0.834188	0.831187	TRUE
27	53292	0.780065	0.785518	0.78077	TRUE
28	71052	0.726773	0.727275	0.727371	FALSE
29	67710	0.655721	0.659489	0.656412	TRUE
30	86154	0.588011	0.58302	0.588002	TRUE
31	80815	0.501857	0.499736	0.501827	TRUE
32	89696	0.421042	0.412592	0.421078	TRUE
33	78145	0.331346	0.3255	0.331085	TRUE
34	81771	0.253201	0.242973	0.253115	TRUE
35	60031	0.17143	0.169504	0.171323	TRUE
36	52184	0.111399	0.108776	0.111357	TRUE
37	31753	0.059215	0.062884	0.059058	TRUE
38	19488	0.027462	0.031828	0.027531	TRUE
39	6589	0.007974	0.013544	0.007928	TRUE
40	1385	0.001385	0.00456	0.001371	TRUE

6.2.3. Isomorphic Tournament Algorithm

A faster algorithm involving isomorphisms was developed. Being isomorphic means that all adjacency matrices that are isomorphic can be re-expressed (ignoring labels) as one another. This is an improvement over the simulation, but requires costly isomorphism checks. The distribution of unique adjacency matrices is documented in Table 6-6: Performance Results of α-vector Algorithm. While we do not describe this in detail, the rationale behind this approach is the same as the following score vector approach.

6.2.4. Score Vector Interchange Algorithm

We could go about collecting α vectors, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots \alpha_n)$, in a variety of ways. In the $n = 3$ case, we can calculate all the $\boldsymbol{\alpha}_i$ by summing over the rows of each adjacency matrix. There are eight of them. The six with $d = 0$ are isomorphic and the two with $d = 1$ are isomorphic. In other words, when $\boldsymbol{\alpha}$ is calculated, it will be the same for two graphs that are isomorphic. Then, we do not need to do an exhaustive search of all possible $\boldsymbol{\alpha}$. We simply need to keep track of the $\boldsymbol{\alpha}_i$ and their corresponding c_i — count of times the alpha vector occurs. Additionally, the number of unique score vectors is smaller for larger n, as we know by comparing Beineke and Harary (1974) and Davis (1954).

We examine the $n = 3$ to $n + 1 = 4$ step in Table 6-4. We start with $\boldsymbol{\alpha}_1 = (0, 1, 2)$, $\boldsymbol{\alpha}_2 = (1, 1, 1)$, $c_1 = 6$, and $c_2 = 2$. The $\boldsymbol{v}_j, j \in 1, 2, \dots, 2^3$, represent a new vertex being added. A new vertex will have n directed edges either entering or leaving. Then, there are 2^n possible combinations of edges entering ($v_{jk} = 1$) or leaving ($v_{jk} = 0$) that $(n + 1)^{th}$ vertex. These are listed in the first grand column in Table 6-4. For all $I = 2$ and $J = 2^n = 8$ pairs of $\boldsymbol{\alpha}_i$ and $\boldsymbol{v}_j$, we need to add the vectors — these are listed in the second grand column. We must also account for the edges which are leaving the $(n + 1)^{th}$ vertex. These

can be found by summing over the $\boldsymbol{v}_j$ to count the number of edges entering old vertices and subtracting that from the number of edges a new vertex must have: $\alpha_{i,(n+1)} = n - \sum_{k=1}^{n} v_{jk}$. This is in the fourth column of the second grand column. In the third grand column, we show that the prior counts of the $\boldsymbol{\alpha}_i$ are tracked. For every new $\boldsymbol{v}_j$ that is added to an $\boldsymbol{\alpha}_i$, we are tracking c_i of those $\boldsymbol{\alpha}_i$. We then combine the $\boldsymbol{\alpha}_i + \boldsymbol{v}_j$ and $a_{i,(n+1)}$ into a new vector of length $(n+1)$ and call them the $\boldsymbol{\alpha}_{ij}^*$ — these are in the fourth grand column. Lastly, we identify the I^* unique $\boldsymbol{\alpha}$ among the $\boldsymbol{\alpha}_{ij}^*$, count the number of times they occur in i, multiply that count by c_i, and add all the counts corresponding to a unique $\boldsymbol{\alpha}_{i^*}^*$together. At the point that one has the $\boldsymbol{\alpha}_{i^*}^*$ and $c_{i^*}^*$, you may want to count the number of circular triads in all $I^*$of the $\boldsymbol{\alpha}_{i^*}^*$ using (6.1.1), $d = \frac{(n+1)n(2(n+1)-1)}{6} - \frac{1}{2}\sum_{r=1}^{n+1}(\alpha_r)^2$ (where $n \to n+1$). This calculation is equivalent to counting the number of interchanges that need to be made to return an $\boldsymbol{\alpha}$ to $(0, 1, \dots, n)$. For example, in $\boldsymbol{\alpha}_{2^*}^*$, if one moves an edge from a vertex with $\alpha_r = 2$ to a different vertex with $\alpha_r = 2$, then the first vertex loses an edge, and a different vertex gains an edge. At that point, after sorting, you would have $\boldsymbol{\alpha} = (0, 1, 2, 3)$. So, with one interchange, the graph achieves the perfect rank state, and $d = 1$. The d distribution need not be passed on to the next iteration of the algorithm, but the list of unique $\boldsymbol{\alpha}_{i^*}^*$ and their counts $c_{i^*}^*$ do need to be passed on to the next step of the algorithm.

Table 6-4: Keeping Track of $\boldsymbol{\alpha}_i$ and c_i

i
1
2

$\boldsymbol{\alpha}_i$		
0	1	2
1	1	1

c_i
6
2

i	j	$\boldsymbol{v}_j$		
1	1	1	1	1
1	2	1	1	0
1	3	1	0	1
1	4	0	1	1
1	5	1	0	0
1	6	0	1	0
1	7	0	0	1
1	8	0	0	0
2	1	1	1	1
2	2	1	1	0
2	3	1	0	1
2	4	0	1	1
2	5	1	0	0
2	6	0	1	0
2	7	0	0	1
2	8	0	0	0

$\boldsymbol{\alpha}_i + \boldsymbol{v}_j$			α_{i4}
1	2	3	0
1	2	2	1
1	1	3	1
0	2	3	1
1	1	2	2
0	2	2	2
0	1	3	2
0	1	2	3
2	2	2	0
2	2	1	1
2	1	2	1
1	2	2	1
2	1	1	2
1	2	1	2
1	1	2	2
1	1	1	3

c_i
6
6
6
6
6
6
6
6
2
2
2
2
2
2
2
2

$\boldsymbol{\alpha}^*_{ij}$			
0	1	2	3
1	1	2	2
1	1	1	3
0	1	2	3
1	1	2	2
0	2	2	2
0	1	2	3
0	1	2	3
0	2	2	2
1	1	2	2
1	1	2	2
1	1	2	2
1	1	2	2
1	1	2	2
1	1	2	2
1	1	1	3

$\boldsymbol{\alpha}^*_{i^*}$			
0	1	2	3
0	2	2	2
1	1	1	3
1	1	2	2

$c^*_{i^*}$	i^*
$4 \cdot 6 = 24$	1
$1 \cdot 6 + 1 \cdot 2 = 8$	2
$1 \cdot 6 + 1 \cdot 2 = 8$	3
$2 \cdot 6 + 6 \cdot 2 = 24$	4

$\boldsymbol{\alpha}^*_{i^*}$			
0	1	2	3
0	2	2	2
1	1	1	3
1	1	2	2

i^*	$c^*_{i^*}$
1	24
2	8
3	8
4	24

d
0
1
1
2

d	Frequency
0	24
1	16
2	24

We were able to run this algorithm up to and including $n = 12$ in under 19 minutes. To get to the $n = 8$ distribution, the calculation had run for a grand total of 1.88 seconds — the same distribution that took 5 days to create with the exhaustive algorithm! In the process, we see that every pair of adjacency matrices that are isomorphic have exactly one corresponding $\boldsymbol{\alpha}$, but there are many non-isometric matrix pairs that share an $\boldsymbol{\alpha}$. For example, when $n = 8$, there are 6880 unique adjacency matrices, but only 167 unique α

vectors. This means a lot less computation when comparing the last two algorithms. The frequencies obtained for $n = 9$ to 12 can be seen in Table 6-5.

Table 6-5: Replication and Extension of Kendall and Gibbons (1990) Appendix Table II: α - Vector Algorithm

d	**9**	**10**	**11**	**12**
0	362880	3628800	39916800	479001600
1	846720	9676800	119750400	1596672000
2	2580480	31449600	412473600	5801241600
3	5093760	68275200	972787200	1.4725E+10
4	12579840	175392000	2594592000	4.0721E+10
5	19958400	311592960	5054353920	8.5709E+10
6	44698752	711728640	1.178E+10	2.0464E+11
7	70785792	1193794560	2.1574E+10	4.0036E+11
8	130032000	2393475840	4.3543E+10	8.4202E+11
9	190834560	3784596480	7.5821E+10	1.5283E+12
10	361525248	7444104192	1.4891E+11	3.0979E+12
11	443931264	1.0527E+10	2.3655E+11	5.2334E+12
12	779950080	1.9534E+10	4.3892E+11	9.9713E+12
13	1043763840	2.761E+10	6.902E+11	1.6505E+13
14	1529101440	4.7107E+10	1.1997E+12	2.9828E+13
15	1916619264	6.4016E+10	1.8061E+12	4.7204E+13
16	2912257152	1.0745E+11	3.0868E+12	8.406E+13
17	3078407808	1.3447E+11	4.3987E+12	1.2659E+14
18	4506485760	2.1894E+11	7.2564E+12	2.1709E+14
19	4946417280	2.723E+11	1.0369E+13	3.2737E+14
20	6068256768	4.1751E+11	1.6133E+13	5.3664E+14
21	6160876416	4.9408E+11	2.212E+13	7.7575E+14
22	7730384256	7.4328E+11	3.4508E+13	1.2766E+15
23	6292581120	8.2974E+11	4.5536E+13	1.7905E+15
24	6900969600	1.2023E+12	6.7968E+13	2.8301E+15
25	5479802496	1.3346E+12	8.9403E+13	3.9612E+15
26	4327787520	1.7739E+12	1.2878E+14	6.082E+15
27	2399241600	1.8788E+12	1.6343E+14	8.2653E+15
28	1197020160	2.4966E+12	2.3395E+14	1.271E+16
29	163094400	2.407E+12	2.8543E+14	1.6653E+16
30	3230080	3.032E+12	3.9366E+14	2.4896E+16
31		2.8411E+12	4.8061E+14	3.2917E+16
32		3.1664E+12	6.3705E+14	4.7515E+16
33		2.7433E+12	7.422E+14	6.0739E+16
34		2.8778E+12	9.811E+14	8.8543E+16
35		2.1099E+12	1.1054E+15	1.1008E+17
36		1.8401E+12	1.395E+15	1.5501E+17
37		1.1093E+12	1.547E+15	1.9325E+17

d	9	10	11	12
38		6.8972E+11	1.8755E+15	2.6517E+17
39		2.3068E+11	1.9892E+15	3.2173E+17
40		4.8252E+10	2.3719E+15	4.4177E+17
41			2.3625E+15	5.1745E+17
42			2.6585E+15	6.9156E+17
43			2.5956E+15	8.1838E+17
44			2.6948E+15	1.0543E+18
45			2.4315E+15	1.2012E+18
46			2.3909E+15	1.5593E+18
47			1.9246E+15	1.7239E+18
48			1.6548E+15	2.1549E+18
49			1.17E+15	2.3756E+18
50			8.2854E+14	2.8715E+18
51			4.2707E+14	3.0572E+18
52			1.981E+14	3.6686E+18
53			5.2098E+13	3.7349E+18
54			3.9226E+12	4.3032E+18
55			4.8252E+10	4.3536E+18
56				4.7516E+18
57				4.5358E+18
58				4.8678E+18
59				4.373E+18
60				4.3533E+18
61				3.7222E+18
62				3.3856E+18
63				2.614E+18
64				2.1569E+18
65				1.4112E+18
66				9.4877E+17
67				4.9169E+17
68				2.2303E+17
69				5.9991E+16
70				9.3077E+15

In R, they are stored as exact values. The $n = 13$ case was also calculated but is not shown. When calculating the $n = 14$ case, the $\boldsymbol{v}$ matrix (i.e., all possible additional edges) was too large for the memory on the author's PC. Plans to address this issue follow.

Table 6-6: Performance Results of α -vector Algorithm

n	Number of unique		Runtime in seconds	Total
	adjacency matrices	α vectors		
4	4	4	0.01	64
5	12	9	0.03	1,024
6	56	22	0.1	32,768
7	456	59	0.39	2,097,152
8	6880	167	1.35	268,435,456
9	-	490	4.4	68,719,476,736
10	-	1486	24.68	35,184,372,088,832
11	-	4639	150.45	36,028,797,018,963,968
12	-	14805	936.9	73,786,976,294,838,206,464

6.3. Reaching Larger n

First, the space constraint needs to be handled to reach larger n,. Fortunately, the planned way to address this has a side effect which is especially parallelizable. We know from Narayana and Bent (1974) the total number of unique score vectors for every n. These counts match our counts up to $n = 12$. Even with the modifications to space we suggest, we likely cannot reach the $n = 30$ case with the current algorithm, as it would require storing a minimum of 96,478,910,768,821 unique score vectors[9], each of length 30.

In our best method in 6.2, we combine all 2^{n-1} binary vectors $\boldsymbol{v}$ with each of the unique score vectors stored from the previous iteration. When adding the binary vectors, we currently store a 2^{n-1} by n binary matrix including all possible binary vectors of length n. Alternatively, we propose storing only a matrix of 2^{k_j} binary vectors, where $\sum_{j=1}^{J} k_j = n-1$, J is the number of partitions, and $k_j \in \left\{\left\lfloor \frac{n-1}{J} \right\rfloor, \left\lceil \frac{n-1}{J} \right\rceil\right\}$. For small n, for example, $J = 2$.

[9] No back of the envelope calculation has been made to estimate how much space such a matrix would require.

Then, the algorithm would loop through all combinations of the first and second partitions, concatenate the rows on the fly, and then add them to the score vectors as before. This would easily be adjusted with parallelization by sending the entire second partition and the score vectors to each of the nodes, but dividing the first partition close to evenly among the nodes. However, there would be a bottleneck at the end of each nth iteration as the results would have to be combined. Additionally, now that we know the number of unique score vectors from Narayana and Bent (1974), a row count check should be performed.

The algorithm in 6.2.4 could run to $n = 15$ in a serial computation on ManeFrame in under 3 days (assuming no space constraints) and the algorithm in 6.3 could potentially run to $n = 20$ using parallel computation on one thousand nodes on ManeFrame in about a month.

CHAPTER 7:

ONGOING AND FUTURE WORK

We briefly list some open problems and questions which we are working on and plan to work on in the future.

7.1. Further CRSP Exploration

7.1.1. Monte Carlo CRSP Simulations

Since it is commonplace to assume transitivity is guaranteed, analysts who believe in the transitivity assumption might not want to use a model that does not. To assuage these people, we could generate data from various transitive structures and show when (and when not) CRSP does a good job of recovering the transitive structure. Additionally, for the more open-minded, we would like to explore the intransitive capabilities of the CRSP and Generalized CRSP models.

In order to accomplish this, we need to designate, in advance, methods to monitor CRSP performance:

1. Under transitive structure:
    a. $d \approx 0$, and
    b. Kendall's $\tau \approx 1$.

2. Under intransitive structure:
    a. $\hat{d} \approx d$,
    b. $\widehat{\sum_i N_i} \approx \sum_i N_i$, number of edges in at least one cycle,
    c. Number of edges directionally correct under true structure, and
    d. Number of rs pairs required to represent directionality
3. $E\left[\hat{\theta} - \theta\right] \approx 0$
4. Variation accounted for by the model

7.1.1.1. Strictly Transitive Lattice

Our first transitive structure should be a transitive lattice. Of interest here would be the effect of variation on the ability to observe the transitive structure. Here,

$$X_i \sim N\left(\mu = i - \frac{n+1}{2}, \sigma = c\right), \text{where } i \in [1, 2, \dots, n] \text{ and } c \in \left\{\frac{1}{10}, 1, 10\right\},$$

$$E[\hat{\alpha}_i] = 1 - \frac{n+1}{2},$$

$$E\left[\hat{h}\right] = 0, \text{and}$$

$$d \approx 0.$$

Note that $Y_{ij} = X_i - X_j$, i plays $2(n-1)$ games, and $\hat{\alpha}_i = \frac{1}{2}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot i})$. Then,

$$\text{Var}(\hat{\alpha}_i) = \frac{1}{4}\text{Var}(\bar{Y}_{i\cdot}) + \frac{1}{4}\text{Var}(\bar{Y}_{\cdot i}) = \frac{1}{4}\left(\frac{2\sigma^2}{n-1}\right) + \frac{1}{4}\left(\frac{2\sigma^2}{n-1}\right) = \frac{\sigma^2}{n-1}.$$

It follows then that $\text{SD}(\hat{\alpha}_i - \mu_i) = \frac{\sigma}{\sqrt{n-1}}$.

Additionally, it would be interesting to see what happens if one circular triad is introduced, as well as a particularly large reversal. Would such a large reversal dominate the intransitive parameters?

7.1.1.2. Strength Ratings Randomly Distributed

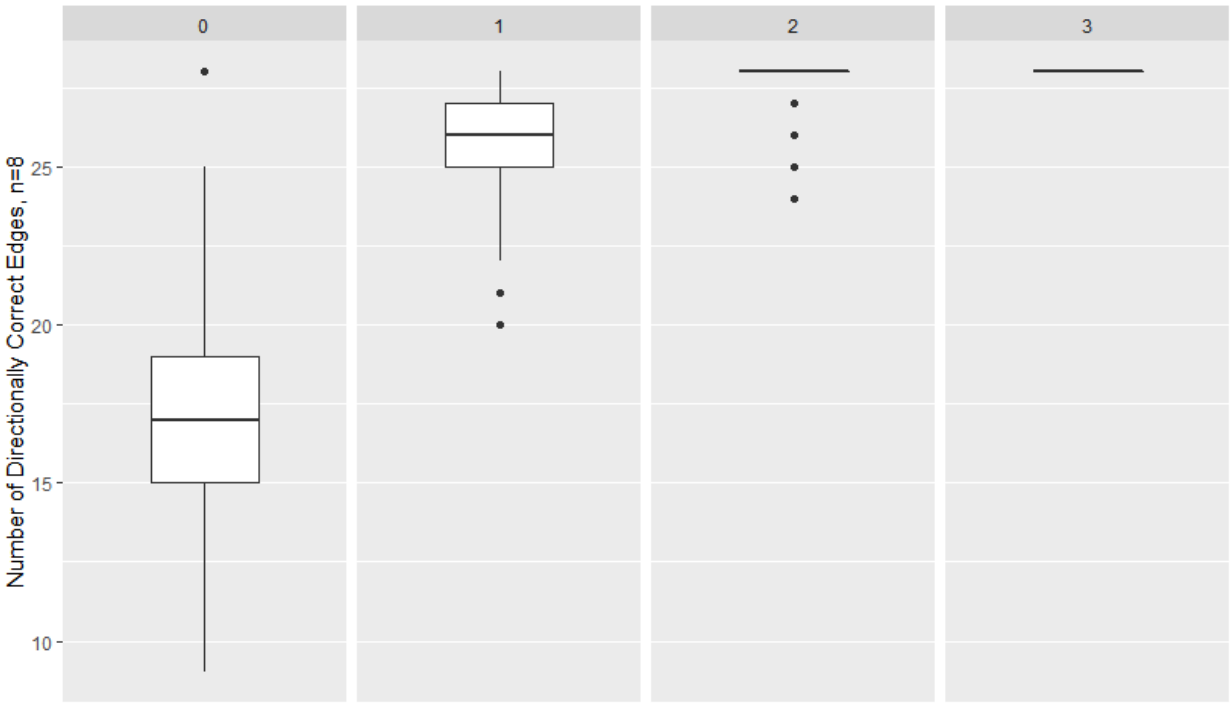
We could then examine ratings which are randomly distributed, i.e., the ratings $X_i \sim F(x)$. Then, unlike the transitive lattice, some teams could be remarkably close in truth. Example distributions could include the Uniform distribution, Normal distribution, and Gamma distribution. Again, could we find the true transitive structure? What kind of intransitivities are liable to occur when they should not?

7.1.1.3. Randomly Generated Tournament Matrices, with Replications

We could also fit CRSP to randomly generated tournament matrices with $\sigma = 0$ — just 1's and corresponding -1's in the off-diagonal entries (a la Kendall and Smith). This way, $h = 0$, and we should be able to see how many eigenvector pairs are required to fit "perfectly intransitive" data and be directionally correct. This may give us some insight into how large k should be to adequately describe certain intransitive structures, or any intransitive structure.

In Figure 7-1, we have already examined the $n = 8$ case as an example (with $M =$ 10,000 replications). Here, we see that we can typically represent about 17 out of 28 relationships in the correct direction (i.e., if $A \to B$, then $\mu_{AB} > 0$) with just the strength vector $\boldsymbol{\alpha}$. However, with just one $\boldsymbol{r}, \boldsymbol{s}$ pair, we can get 26 directionally correct relationships. With two $\boldsymbol{r}, \boldsymbol{s}$ pairs, we can represent all 28 relationships correctly 98% of the time.

Figure 7-1: $n = 8$ Tournament Simulation



7.1.2. Super Smash Brothers Application

Another application might be to rank players in the Nintendo 64 version of Super Smash Brothers. There are many factors that can introduce intransitivities and a linear ordering of players may not sufficiently describe the hierarchy of players. Such predictive factors may include character usage, counter-picking, player matchup history, and players who stream and are well known by the community versus those who are not. However, these things are often not known, and covariate data is scarce. But to answer this question, we will likely need to modify the CRSP model to handle missing observations.

Fortunately, this author helped run a 30-person round-robin tournament with players who traveled from around the world (London, California, Colorado, New York). The data needs to be cleaned but is otherwise ready. Each set was best-of-3, where the third

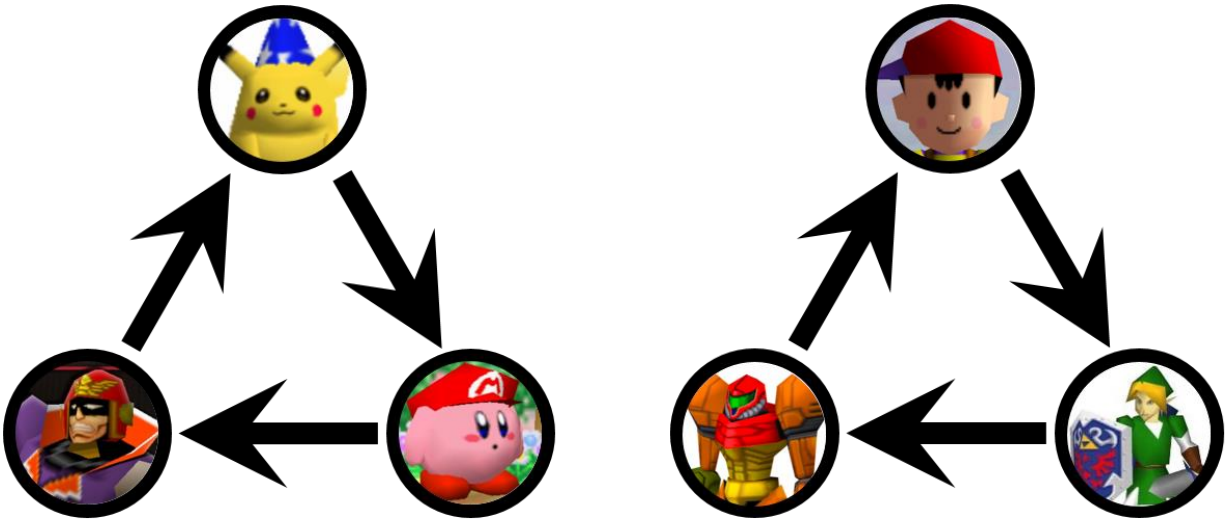
game was not played if a player won twice. Each game, the players' characters were recorded. This should enable us to model both player and character strengths, as well as character intransitivities. Then, our model might be

$$\mu_{lmij} = \gamma_l - \gamma_m + \alpha_i - \alpha_j + r_i s_j - s_i r_j, \tag{7.2.1}$$

where γ_l are player strengths, α_i are character strengths, and $r_i s_j - s_i r_j$ is character intransitivity.

It is widely believed that the top tier (left, Figure 7-2) of characters has a rock-paper-scissors relationship, while some believe the lowest tier of characters also includes a rock-paper-scissors relationship.

Figure 7-2: Intransitivities in Characters in Super Smash Brothers N64



7.1.3. NBA Applications

With a start on the NBA example, we might want to explore more specific questions. First, we could examine the claim made during a 2017 MITSSAC talk that, on average, only one team that should have made the playoffs in a strict top-16 format does not. By using the CRSP model, we identified an average of two teams per season that should have made the playoffs in a strict top-16 format. We also found that most playoff series are closer in such an optimal top-16 format. Thus, the series take more games to complete and thus draw more revenue. Finally, an optimal top-16 format features the top two teams in the NBA more often than previous formats.

Second, we could examine the gambling scandal that shook the NBA more than ten years ago which Eden (2019) documented recently. The NBA, after the scandal broke in the news, hired a private law firm to investigate. In Pedowitz's report to the NBA (2008), 17 games are identified which the legal team believed may have involved point-fixing by the referee. CRSP is well-situated to examine point differentials on which wagers are made (called point spreads).

7.2. CRSP-Related Future Work

The subsections on CRSP are ordered by importance.

7.2.1. CRSP EM Algorithm

In cases where games have not been played yet or the schedule is imbalanced, an EM Algorithm solution could be used.

7.2.2. Clustered Line-Ups

Instead of looking at entire teams, we could examine individual line-ups. Since these line-ups are usually scarce, we would have to do something like cluster them.

7.2.3. CRSP Standard Error Estimates

For which point estimates do we need standard errors? The standard errors for the home effect h and the team strength parameters α_i should not be difficult to calculate. But what about the $\boldsymbol{rs}$ pairs? How do we go about calculating the standard errors of r_i and s_j, $r_i s_j$, $r_i s_j - s_i r_j$, or $\sum_{k=1}^{K} r_{ik} s_{jk} - s_{ik} r_{jk}$?

7.2.4. CRSP Eigenvector Pair Exploration

How many $\boldsymbol{rs}$ pairs should we include in the model? Should we default to the transitive model? Should we use some kind of penalty (i.e., AIC), to determine the number of $\boldsymbol{rs}$ pairs? Would finding the elbow in a plot of the variation explained by number of $\boldsymbol{rs}$ pairs be sufficient (i.e., in LSR, R^2 and the number of covariates, or in PCA, the number of PCs)? Would a cross-validation type of procedure work better? Since sports data is broken up by seasons, would it make sense to pick the number of $\boldsymbol{rs}$ pairs by using a rule of thumb learned from modeling previous regular seasons and comparing the models for each number of $\boldsymbol{r}, \boldsymbol{s}$ pairs by their playoff prediction accuracy?

How do our eigenvector pairs relate to the blade and chest vectors in Chen and Joachim's model? Is there a simple solution like ours in a Bradley-Terry setting?

7.2.5. CRSP Model and Graph Theory

We would like the CRSP model to be able to identify any possible digraph. It can be seen that[10] if all $\left\lfloor \frac{n-1}{2} \right\rfloor \boldsymbol{rs}$ pairs are included in the model, then the β_{ij} (where $\beta_{ij} = \beta_{ji}$) can be fully recovered and thus any theoretical tournament could be recovered. We believe this would typically lead to overfitting, especially in cases of a small number of games. So, the question is what kind of tournaments could be recovered as the number of $\boldsymbol{rs}$ pairs increase. One way to do this would be to generate random tournaments, simulate games with extremely low variance, build models for all possible $\boldsymbol{rs}$ pairs, and then count the number of correctly recovered edges. If the results are satisfactory, then the same procedure could be repeated with larger variances. Additionally, it would be interesting to keep track of the number of circular triads, d, and other measures of intransitivity as the number of $\boldsymbol{rs}$ pairs increase.

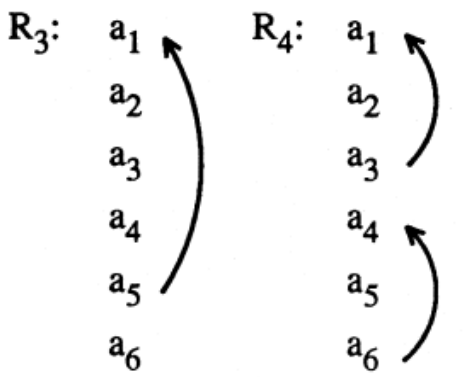
7.2.6. CRSP Model Cases and Extensions

Currently, the model only deals with a double round robin. We would like to explore cases of imbalance — where every pair is compared, but some more than twice. The analysis in 5.1 on NBA data uses a method of moments approach to estimate each edge between teams i and j. The least squares method would reduce variance in the strength estimates. However, it is unclear how this would affect the $\boldsymbol{rs}$ pairs. Next, we would examine the case of missing data — that is, where not every pair of teams have played or have played just once.

[10] Also, we have not proven this.

We have seen many interesting modifications of ranking models in the literature. For example, the composition of the team should be considered. If an NBA team plays a game without their best player, does the outcome really represent the team? Additionally, some teams go through states of change — they trade their players, younger players get in a groove and replace older players, rookies hit a wall at about half of the way through the season, etc. This could be taken into account by treating each lineup separately, as in LinNet (Pelechrinis, 2017). Alternatively, games weights could be introduced for each game. This would be a natural extension of the unbalanced and missing cases. More recent games could also be weighed more heavily. Generally, weights could be a function of player minutes (i.e., of the top seven most important players).

7.2.7. CRSP Visualization



A graphical (visual) representation of the model should also be explored. For example, Monsuur and Storcken (1997) give a couple of nice pictures. Here, vertical position implies overall strength. Team a_1 is the strongest, while team a_6 is the weakest. The arrows, then, show inconsistent edges. For example, in graph R_3, team a_5 upsets team a_1. This type of picture poses a few problems. If the number of teams is large and there are a lot of teams whose strength ratings are close, then we may expect to see many edge reversals in the middle. This may not be desirable. To address this, one might select only a

subset of edges to display. One could display the biggest upsets or produce only the reversals for one team at a time. This could be implemented in a browser by hovering over a team/node with all reversals appearing (both jumps up and jumps down). In this case, a graph with no intersecting edges would be guaranteed if all reversals to lower rated teams are drawn on the left, and all reversals to higher rated teams are drawn on the right. Also, the strength of a reversal could be illustrated by varying the edge width or the edge length (larger reversals would protrude further from the graph).

7.2.8. Dynamic CRSP Model

It would be quite useful to be able to make predictions in the middle of the season. This is related to the missing games problem. However, an additional question is whether the entire model needs to be re-calculated, or can it can be updated on a per-game basis. Elo's model, for example, can update just two teams at a time. This makes it a bit more interpretable, as well — some amount of rating points is directly transferred from one team to the other upon completion.

7.2.9. CRSP with Covariates

Additionally, covariates should eventually be considered for inclusion in the model. As seen in Chapter 3, season-long covariates (total defensive rebounds, three point percentage, opponent field goal percentage, et cetera) can also be used to predict success. A simple course of action might be to measure the overall strengths of teams first, using the covariates and point differential. Then any deviations from a transitive model could be modeled as before with $\boldsymbol{rs}$ pairs.

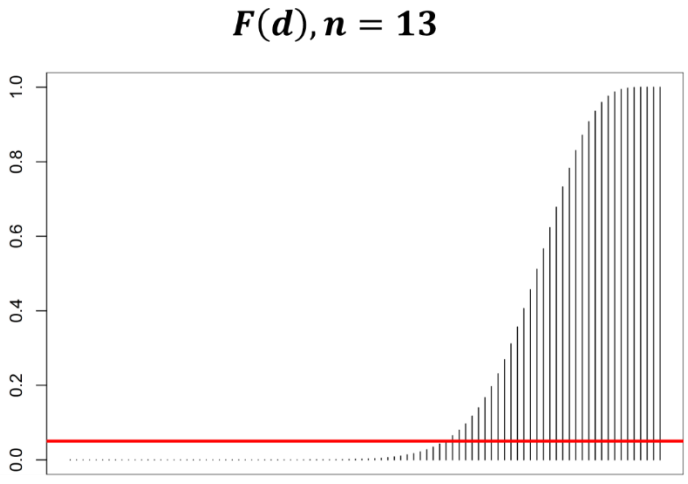
7.3. Kendall and Smith's Circular Triad Distribution

See Section 6.3. We have extended the circular triad distribution past the $n = 15$ case, but this work is unfortunately not documented here. We took advantage of Kadane's work to accomplish this, but synthesized findings from other researchers.

Table 7-1: Progress on Circular Triad Distribution

n	Number of Isomorphic Graphs, $T(n)$	Number of Unique Score Vectors, $S(n)$	Year Reached
3	2	2	1940
7	456	59	1940
8	6880	167	1961
11	9.0×10^8	4639	'98, '16
13	4.9×10^{13}	48107	'98[11], '17
16	6.4×10^{22}	1799659	2018[12]
19	2.5×10^{34}	73996100	–

Figure 7-3: Limitations of the Chi-Squared Approximation



Critical Values of d distribution

$\alpha \setminus n$	8	9	10	11	12	13
0.001	2	6	12	20	30	43
0.01	5	10	17	26	37	51
0.05	8	14	21	31	43	58
0.1	10	16	24	34	46	62

The exact count provides more accurate critical values for small α.

[11] Knezek et al. (1998) do not publish the full distribution.
[12] The $n = 16$ case is currently only partially complete.

7.4. Probabilistic Extensions of the Circular Triad Distribution

Regarding Kendall and Smith's Table II (1940), it would be interesting, instead, to see how this distribution would change under different circumstances. In this distribution, it is assumed that every preference is equally likely — $\mathrm{P}(A \to B) = P(A \leftarrow B) = .5$. For example, if it were known that dogs tend to prefer objects placed on their left, what would happen to that distribution? What if $P(A \to B) > P(A \leftarrow B)$?

Teams playing in their own stadium are known to have a higher chance of winning (sometimes this chance is slight). Across sports and leagues, this home team advantage varies. For many established leagues, a general home court advantage is known and can be expressed as a point spread or as a probability. In the NBA, for example, during the 2014-2015 season, the home team won 57.4% of the time. In the situation where there are three teams playing each other once, ignoring team strength, the probability of a circular triad increases as home team advantage increases. In general, if p is the probability of winning at home, then $P(circular\ triad) = P(3\ home\ wins) + P(3\ road\ wins) = p^3 + (1-p)^3 = 1 - 3p + 3p^2$. In the NBA, the probability of a circular triad would then increase to .267 from .25.

The problem becomes even more interesting if we assign different probabilities to each pair's outcome. To model this, perhaps we could use a graph with weighted edges where the weights are probabilities or measures of (relative) team strength. Or perhaps, the graph of interest would have a corresponding matrix of probabilities that the team in the left column would win, and so each of the $2^{\binom{n}{2}}$ possible outcomes are not equally weighted.

7.5. Alpha Vectors and Other Distributions

Can Slater's i be calculated with only α-vectors? If so, we could fully describe Slater's i-distribution using our already-calculated α-vector distribution. If not, can we find a way to reverse-engineer all adjacency matrices that relate to a particular α-vector? If so, this might be a convenient way to solve the Slater's i-distribution problem, but may also allow us to explore adjacency matrices for larger n than previously possible. One way might be to explore the $n-1$ to n step in our newest α-vector generating algorithm.

We can also enumerate any distribution for a measure that is a function of an alpha vector — we have all alpha vectors for $n \leq 15$. We can examine the relationship between d and Bezembinder's ρ as well as between k and $\sum_{j=1}^{k} N_j$. We expect the latter to be curved; as k decreases, $\sum_{j=1}^{k} N_j$ increases rapidly.

7.6. Further Graph Theory

We can also continue exploring the forms described in Moon (1968) and Goldberg (1966, 1972).

Figure 7-4: All Unique (Non-isomorphic) K_5 Graphs with Directed Edges

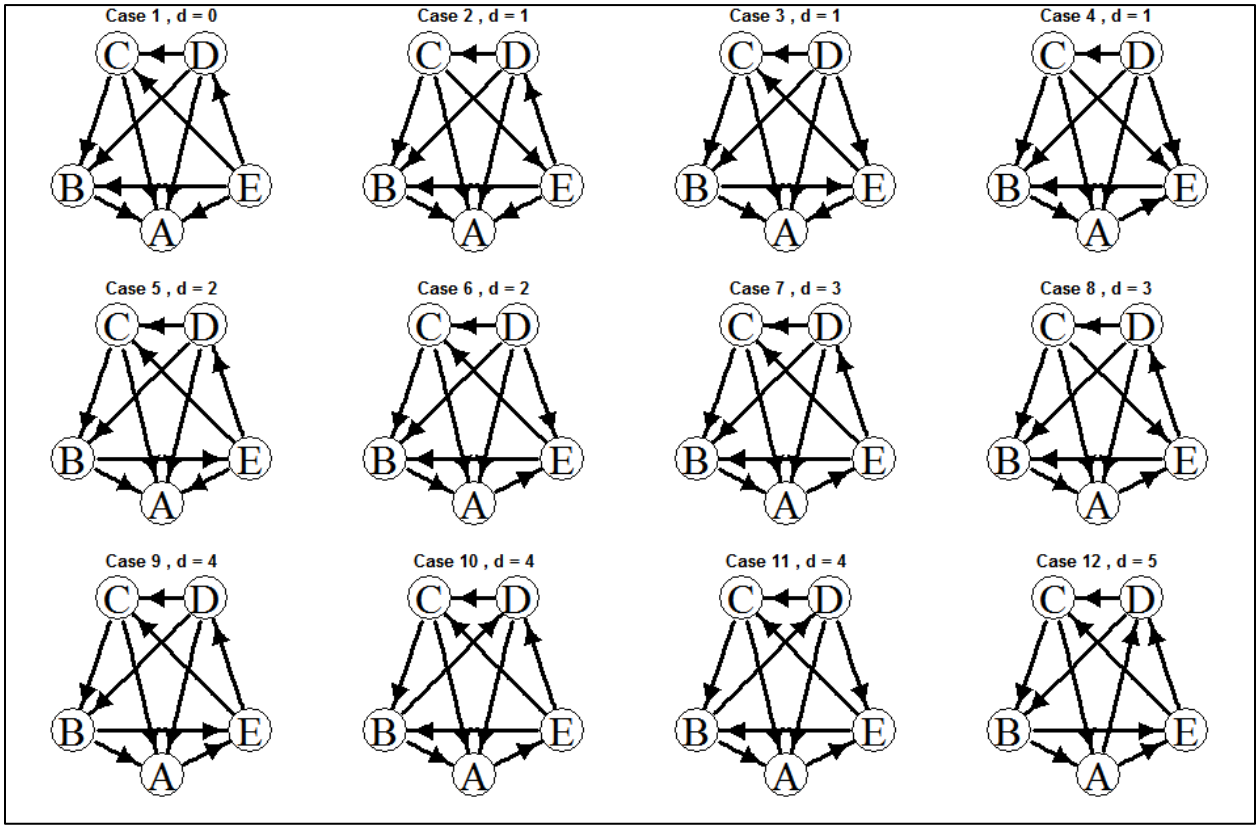
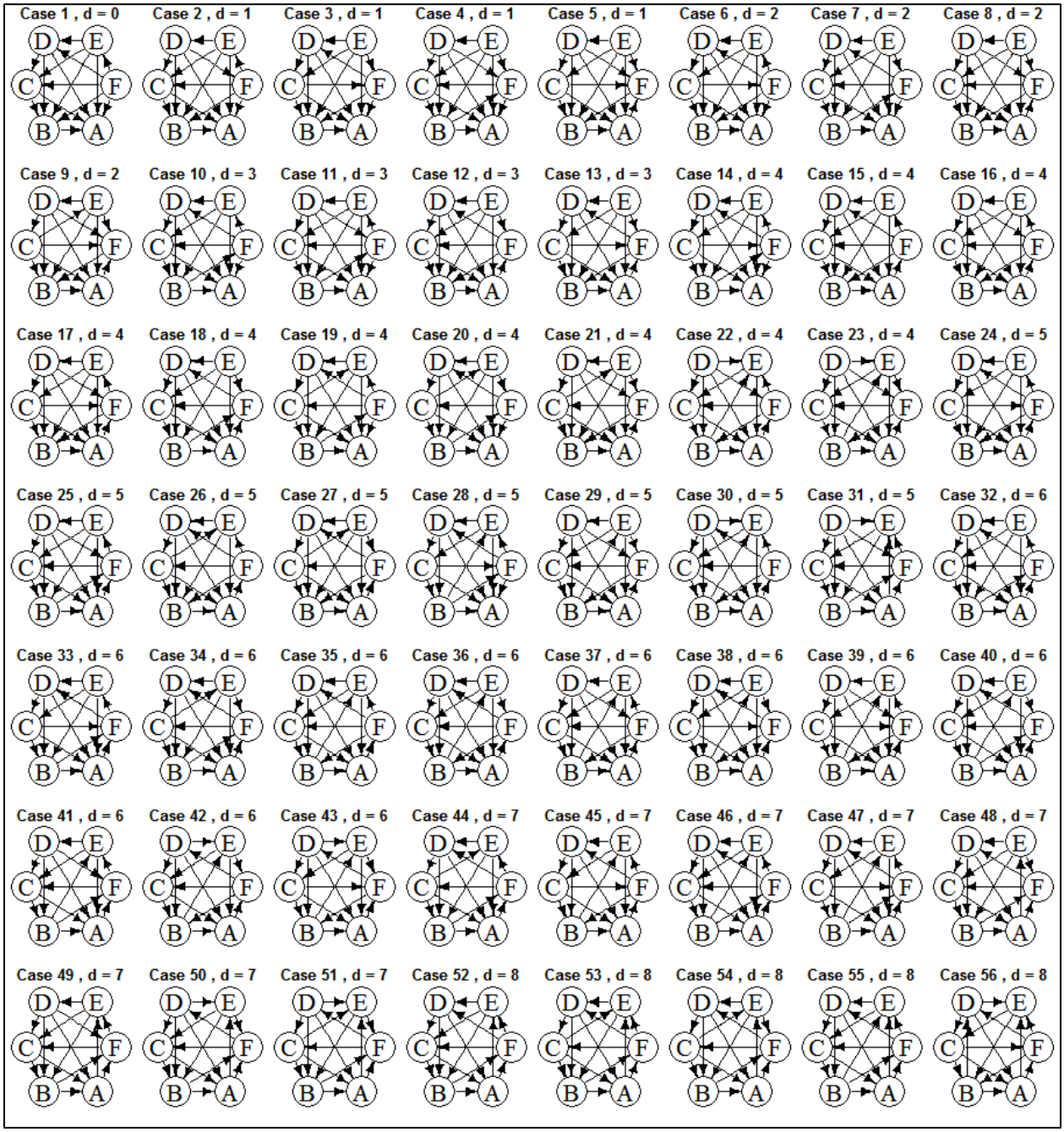


Figure 7-5: All Unique (Non-isomorphic) K_6 Graphs with Directed Edges



REFERENCES

Abeles, F. (1979). Ranking by Inversion- A Note on C.L. Dodgson. *Historia Mathematica*, 310-317. doi:10.1016/0315-0860(79)90129-0

Adams, E. S. (2005). Bayesian Analysis of Linear Dominance Hierarchies. *Animal Behaviour, 69*, 1191-1201. doi:10.1016/j.anbehav.2004.08.011

Aizenberg, N., Koren, Y., & Somekh, O. (2012). Build Your Own Music Recommender by Modeling Internet Radio Streams. *Proceedings of the 21st International Conference on World Wide Web*, 1-10. doi:10.1145/2187836.2187838

Alspach, B., & Tabib, C. (1982). A Note on the Number of 4-Circuits in a Tournament. *Annals of Discrete Mathematics, 12*, 13-19. doi:10.1016/s0304-0208(08)73485-6

Alway, G. G. (1962). The Distribution of the Number of Circular Triads in Paired Comparisons. *Biometrika, 49*, 265-269. doi:10.1093/biomet/49.1-2.265

Andrews, D. M., & David, H. A. (1990). Nonparametric Analysis of Unbalanced Paired-Comparison or Ranked Data. *Journal of the American Statistical Association, 85*, 1140-1146. doi:10.2307/2289613

Anton, H. (2009). *Elementary Linear Algebra.* John Wiley & Sons.

Aoki, R., Renato, A. M., & De Melo, P. V. (2017). Luck is Hard to Beat: The Difficulty of Sports Prediction. *arXiv preprint arXiv:1706.02447*. doi:10.1145/3097983.3098045

Appleby, M. C. (1983). The Probability of Linearity in Hierarchies. *Animal Behaviour, 31*, 600-608. doi:10.1016/s0003-3472(83)80084-0

Arrow, K. J. (1950). A Difficulty in the Concept of Social Welfare. *Journal of Political Economics, 58*, 328-346. doi:10.1086/256963

Arrow, K. J. (1951). *Social Choice and Individual Values.*

Arrow, K. J. (1974). General Economic Equilibrium: Purpose, Analytic Techniques, Collective Choice. *The American Economic Review, 64*, 253-272. Retrieved from https://www.jstor.org/stable/1808881

Barrow, D., Drayer, I., Elliott, P., Gaut, G., & Osting, B. (2013). Ranking Rankings: An Empirical Comparison of the Predictive Power of Sports Ranking Methods. *Journal of Quantitative Analysis in Sports, 9*, 187-202. doi:10.1515/jqas-2013-0013

Basak, P. (1993). Some Combinatorial Aspects of Paired Comparisons. *American Journal of Mathematical and Management Sciences, 13*, 1-30. doi:10.1080/01966324.1993.10737343

Bauer, D. F. (1978). Circular Triads When Not All Paired Comparisons Are Made. *Biometrics, 34*, 458-461. doi:10.2307/2530608

Beineke, L. W., & Haray, F. (2001). On the Maximum Number of Cyclic Triples in Oriented Graphs. *Journal of Combinatorial Theory, Series B, 81*, 165-176. doi:10.1006/jctb.2000.1995

Bergson, A. (1954). On the Concept of Social Welfare. *The Quarterly Journal of Economics, 68*, 233-252. doi:10.2307/1884448

Bezembinder, T. G. (1981). Circularity and Consistency in Paired Comparisons. *British Journal of Mathematical and Statistical Psychology, 34*, 16-37. doi:10.1111/j.2044-8317.1981.tb00615.x

Bilmes, J., Ji, G., & Meila, M. (2002). Intransitive Likelihood-Ratio Classifiers. *Advances in Neural Information Processing Systems*, 1141-1148. Retrieved from http://papers.nips.cc/paper/2084-intransitive-likelihood-ratio-classifiers.pdf

Blankenship, A. B. (1966). Lets Bury Paired Comparisons. *Journal of Advertising Research, 6*, 13-17.

Borda, J. C. (1781). Memoir on Elections by Ballot. *Isis, 44*, 42-51. Retrieved from https://www.jstor.org/stable/227640

Bozóki, S., Dezső, L., Poesz, A., & Temesi, J. (2013). Analysis of Pairwise Comparison Matrices: An Empirical Research. *Annals of Operational Research, 211*, 511-528. doi:10.1007/s10479-013-1328-1

Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika, 39*, 324-345. doi:10.1093/biomet/39.3-4.324

Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks, 56*, 3825-3833. doi:10.1016/j.comnet.2012.10.007

Broom, M., & Cannings, C. (2002). Modelling Dominance Hierarchy Formation as a Multi-Player Game. *Journal of Theoretical Biology, 219*, 397-413. doi:10.1006/jtbi.2002.3137

Brown, T. C., & Peterson, G. L. (2009). An Enquiry into the Method of Paired Comparison. 98. Retrieved from https://www.fs.fed.us/rm/pubs/rmrs_gtr216.pdf

Burns, E. (1980). A Computer Program to Evaluate Paired-Comparison Transitivity. *Behavior Research Methods |& Instrumentation, 12*, 636. doi:10.3758/bf03201860

Callaghan, T., Mucha, P. J., & Porter, M. A. (2007). Random Walker Ranking for NCAA Division IA Football. *American Mathematical Monthly, 114*, 761-777. doi:10.1080/00029890.2007.11920469

Cattelan, M., Varin, C., & Firth, D. (2013). Dynamic Bradley-Terry Modelling of Sports Tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 62*, 135-150. doi:10.1111/j.1467-9876.2012.01046.x

Causeur, D., & Husson, F. (2005). A 2-Dimensional Extension of the Bradley-Terry Model for Paired Comparisons. *Journal of Statistical Planning and Inference, 135*, 245-259. doi:10.1016/j.jspi.2004.05.008

Charon, I., & Hudry, O. (2006). A Branch-and-Bound Algorithm to Solve the Linear Ordering Problem for Weighted Tournaments. *Discrete Applied Mathematics, 154*, 2097-2116. doi:10.1016/j.dam.2005.04.020

Charon, I., & Hudry, O. (2007). A Survey on the Linear Ordering Problem for Weighted or Unweighted Tournaments. *4OR, 5*, 5-60. doi:10.1007/s10288-007-0036-6

Chartier, T., Kreutzer, E., Langville, A. N., & Pedings, K. E. (2011). Sports Ranking with Nonuniform Weighting. *Journal of Quantitative Analysis in Sports, 7*. doi:10.2202/1559-0410.1290

Chen, J., Wang, C., Wang, J., Ying, X., & Wang, X. (2017). Learning the Personalized Intransitive Preference of Images. *IEEE Transactions on Image Processing, 26*. doi:10.1109/tip.2017.2709941

Chen, S. (2016). *Representation Learning for Sequence and Comparison Data.* Ph.D. dissertation, Cornell University. Retrieved from https://hdl.handle.net/1813/43697

Chen, S., & Joachims, T. (2016, 2). Modeling Intransitivity in Matchup and Comparison Data. *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*. doi:10.1145/2835776.2835787

Chen, S., & Joachims, T. (2016, 8). Predicting Matchups and Preferences in Context. *KDD*, 775-784. doi:10.1145/2939672.2939764

Colley, W. (2002). Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained. Retrieved from https://www.colleyrankings.com/matrate.pdf

Coolidge, J. L. (1949). The Story of the Binomial Theorem. *The American Mathematical Monthly, 56*, 147-157. doi:10.2307/2305028

Crispino, M., Arjas, E., Vitelli, V., & Frigessi, A. (2016). Recommendation from Intransitive Pairwise Comparisons. Retrieved from https://pdfs.semanticscholar.org/d0c4/96192fb34c2ac48593da7a16af1b5e980961.pdf

Crispino, M., Arjas, E., Vitelli, V., Barret, N., & Frigessi, A. (2017). A Bayesian Mallows Approach to Non-Transitive Pair Comparison Data: How Human are Sounds? *arXiv preprint arXiv:1705.08805*. doi:10.1214/18-aoas1203

Critchlow, D. E., & Fligner, M. A. (1991). Paired Comparison, Triple Comparison, and Ranking Experiments as Generalized Linear Models, and Their Implementaton on GLIM. *Psychometrika, 56*, 517-533. doi:10.1007/bf02294488

Cropper, S. R. (2011). Ranking Score of Vectors of Tournaments. Retrieved from https://digitalcommons.usu.edu/gradreports/91/

Csató, L. (2015). A Graph Interpretation of the Least Squares Ranking Method. *Social Choice and Welfare, 44*, 51-69.

Csató, L. (2015). Methodological and Applicational Issues of Paired Comparison Based Ranking. doi:10.14267/phd.2015022

Csató, L. (2016). On the Ranking in Swiss System Chess Team Tournament. *arXiv preprint arXiv:1507.05045v2*. doi:10.1007/s10479-017-2440-4

Csató, L., & Rónyai, L. (2016). Incomplete Pairwise Comparison Matrices and Weighting Methods. *Fundamenta Informaticae, 144*, 309-320. doi:10.3233/fi-2016-1337

David, H. A. (1959). Tournaments and Paired Comparisons. *Biometrika, 46*, 139-149. doi:10.2307/2332816

David, H. A. (1971). Ranking the Players in a Round Robin Tournament. *Revue de l'Institut International de Statistique, 39*, 137-147. doi:10.2307/1402170

David, H. A. (1987). Ranking from Unbalanced Paired-Comparison Data. *Biometrika, 74*, 432-436. doi:10.2307/2336160

David, H. A. (1988). *The Method of Paired Comparisons.* Oxford University Press.

Davis, R. L. (1954). Structures of Dominance Relations. *Bulletin of Mathematical Biology, 16*, 131-140. doi:10.1007/bf02478368

De Vries, H. (1998). Finding a Dominance Order Most Consistent with a Linear Hierarchy: A New Procedure and Review. *Animal Behaviour, 55*, 827-843. doi:10.1006/anbe.1997.0708

Dobson, A. J. (1990). *An Introduction to Generalized Linear Models, First Edition (Chapman & Hall/CRC Texts in Statistical Science).* Chapman and Hall/CRC.

Duan, J., Jiyi, L., Yukino, B., & Hisashi, K. (2017). A Generalized Model for Multidimensional Intransitivity. Springer. doi:10.1007/978-3-319-57529-2_65

Dykstra, O. (1960). Rank Analysis of Incomplete Block Designs: A Method of Paired Comparisons Employing Unequal Repititions on Pairs. *Biometrics, 16*, 176-188. doi:10.2307/2527550

Eden, S. (2019, 2 19). How Former Ref Tim Donaghy Conspired to Fix NBA Games. Retrieved from https://www.espn.com/nba/story/_/id/25980368/how-former-ref-tim-donaghy-conspired-fix-nba-games

Elias, G. S. (2012, 8 24). *Characteristics of Games.* MIT University Press Group Ltd.

Elo, A. E. (1978). *The Rating Of Chess Players, Past & Present.* Arco Pub.

Fishburn, P. C. (1973). Binary Choice Probabilities: On the Varieties of Stochastic Transitivity. *Journal of Mathematical Psychology, 10*, 327-352. doi:10.1016/0022-2496(73)90021-7

Fishburn, P. C. (1984). SSB Utility Theory: An Economic Perspective. *Mathematical Social Sciences, 8*, 63-94. doi:10.1016/0165-4896(84)90061-1

Flueck, J. A., & Korsh, J. F. (1975). A Generalized Approach to Maximum Likelihood Paired Comparison Ranking. *The Annals of Statistics, 3*, 846-861. doi:10.1214/aos/1176343186

Fogel, F., A., & Vojnovic, M. (2014). SerialRank: Spectral Ranking Using Seriation. *Advances in Neural Information Processing Systems*. Retrieved from http://papers.nips.cc/paper/5223-serialrank-spectral-ranking-using-seriation.pdf

Frobenius, F. G. (1908). Über Matrizen aus Nicht Negativen Elementen. *Sitzungsberichte der Koeniglich Preussischen Akademie der Wissenschaften*, 471-476.

Frobenius, F. G. (1909). Uber Matrizen aus positiven Elementen II. *Sitzungsber, Berlin: Preuss Akad Wiss*, 514-518. Retrieved from https://www.worldcat.org/title/uber-matrizen-aus-positiven-elementen-ii/oclc/603330332

Frobenius, F. G. (1912). Über Matrizen aus Nicht Negativen Elementen. *Sitzungsberichte der Koeniglich Preussischen Akademie der Wissenschaften*, 456-477. Retrieved from https://upload.wikimedia.org/wikipedia/commons/4/44/Ueber_Matrizen_aus_nicht_negativen_Elementen.pdf

Füredi, Z., Kantor, I., Monti, A., & Sinameri, B. (2010). On Reverse-Free Codes and Permutations. *SIAM Journal on Discrete Mathematics, 24*, 964-978. doi:10.1016/j.endm.2011.09.062

Gabriel, K. R. (1998). Generalised Bilinear Regression. *Biometrika, 85*, 689-700. doi:10.1093/biomet/85.3.689

Gass, S. I. (1998). Tournaments, Transitivity, and Pairwise Comparison Matrices. *The Journal of the Operational Research Society, 49*, 616-624. doi:10.2307/3010670

Gehrlein, W. V. (1989). The Probability of Intransitivity of Pairwise Comparisons in Individual Preference. *Mathematical Social Sciences, 17*, 67-75. doi:10.1016/0165-4896(89)90017-6

Gill, P. S., & Swartz, T. B. (2001). Statistical Analyses for Round Robin Interaction Data. *Canadian Journal of Statistics, 29*, 321-331. doi:10.2307/3316080

Glickman, M. E. (1993). *Paired Comparison Models with Time-Varying Parameters.* Ph.D. dissertation, Harvard University, Cambridge. doi:10.21236/ada272016

Glickman, M. E. (2001). Dynamic Paired Comparison Models with Stochastic Variances. *Journal of Applied Statistics, 28*, 673-689. doi:10.1080/02664760120059219

Glickman, M. E., & Stern, H. S. (1998). A State-Space Model for National Football League Scores. *Journal of the American Statistical Association, 93*, 25-35. doi:10.1137/1.9780898718386.ch5

Glickman, M. E., & Stern, H. S. (2016). Estimating Team Strength in the NFL. In *Handbook of Statistical Methods and Analyses in Sports* (pp. 129-152).

Goodman, I., Gregory, K., & Pai, S. (2015). A Network-Based Approach to Ranking College Football Teams. Retrieved from http://snap.stanford.edu/class/cs224w-2015/projects_2015/A_Network-Based_Approach_to_Ranking_College_Football_Teams.pdf

Govan, A. Y. (2008). *Ranking Theory with Application to Popular Sports.* Ph.D. dissertation, North Carolina State University. Retrieved from https://repository.lib.ncsu.edu/bitstream/handle/1840.16/3102/etd.pdf

Govan, A. Y., & Langville, A. N. (2009). Offense-Defense Approach to Ranking Team Sports. *Journal of Quantitative Analysis in Sports, 5*. doi:10.2202/1559-0410.1151

Gulliksen, H. (1956). A Least Squares Solution for Paired Comparisons With Incomplete Data. *Psychometrika, 21*, 125-134. doi:10.1007/bf02289093

Havlena, W. J., & DeSarbo, W. S. (1991). On the Measurement of Perceived Consumer Risk. *Decision Sciences, 22*, 927-939. doi:10.1111/j.1540-5915.1991.tb00372.x

Heiser, W. J., & D'Ambrosio, A. (2013). Clustering and Prediction of Rankings within a Kemeny Distance Framework. *Algorithms from and for Nature and Life*, 19-31. doi:10.1007/978-3-319-00035-0_2

Herbrich, R., Minka, T., & Graepel, T. (2007). TrueSkill™: a Bayesian Skill Rating System. *Advances in Neural Information Processing Systems*, 569-576. Retrieved from http://papers.nips.cc/paper/3079-trueskilltm-a-bayesian-skill-rating-system.pdf

Higgins, J. P., & Whitehead, A. (1996). Borrowing Strength from External Trials in a Meta-Analysis. *Statistics in Medicine, 15*, 2733-2749. doi:10.1002/(sici)1097-0258(19961230)15:24<2733::aid-sim562>3.0.co;2-0

Hoff, P. (2005). Bilinear Mixed-Effects Models for Dyadic Data. *Journal of the American Statistical Association, 100*, 286-295. doi:10.1198/016214504000001015

Horst, P. (1932). A Method for Determining the Absolute Affective Value of a Series of Stimulus Situations. *British Journal of Educational Psychology, 23*, 418-440. doi:10.1037/h0070134

Huang, T. K., Weng, R. C., & Lin, C. J. (2006). Generalized Bradley-Terry Models and Multi-Class Probability Estimates. *Journal of Machine Learning Research, 7*, 85-115. Retrieved from http://www.jmlr.org/papers/volume7/huang06a/huang06a.pdf

Hunter, D. (2004). MM Algorithms for Generalized Bradley-Terry Models. *Annals of Statistics, 32*, 384-406. doi:10.1214/aos/1079120141

Iida, Y. (2009). The Number of Circular Triads in a Pairwise Comparison Matrix and a Consistency Test in the AHP. *Journal of the Operational Society of Japan, 52*, 174-185. Retrieved from https://www.jstage.jst.go.jp/article/jorsj/52/2/52_KJ00005622385/_pdf

Inc., O. E. (2017). The On-Line Encyclopedia of Integer Sequences. Retrieved from http://oeis.org

Iverson, G., & Falmagne, J. C. (1985). Statistical Issues in Measurement. *Mathematical Social Sciences, 10*, 131-153. doi:10.1016/0165-4896(85)90031-9

Jech, T. (1983). The Ranking of Incomplete Tournaments: A Mathematician's Guide to Popular Sports. *The American Mathematical Monthly, 90*, 246-266. doi:10.2307/2975756

Jiang, X., Lim, L. H., Yao, Y., & Ye, Y. (2011). Statistical Ranking and Combinatorial Hodge Theory. *Mathematical Programming, 127*, 203-244.

Kadane, J. B. (1966). Some Equivalence Classes in Paired Comparisons. *The Annals of Mathematical Statistics, 37*, 488-494. Retrieved from https://www.jstor.org/stable/2238622

Kaiser, H. F., & Serlin, R. C. (1978). Contributions to the Method of Paired Comparisons. *Applied Psychological Measurement, 2*, 421-430. doi:10.1177/014662167800200317

Keener, J. P. (1993). The Perron-Frobenius Theorem and the Ranking of Football Teams. *SIAM Review, 35*, 80-93. doi:10.1137/1035004

Kemeny, J. G. (1959). Mathematics without Numbers. *Daedalus, 88*, 577-591. Retrieved from https://www.jstor.org/stable/20026529

Kemeny, J. G. (1962). Mathematical Models in the Social Sciences. Cambridge, MA: MIT Press.

Kendall, M. G. (1938). A New Measure of Rank Correlation. *Biometrika, 30*, 81-93. doi:10.2307/2332226

Kendall, M. G. (1945). The Treatment of Ties in Ranking Problems. *Biometrika, 33*, 239-251. doi:10.2307/2332303

Kendall, M. G. (1955). Further Contributions to the Theory of Paired Comparisons. *Biometrics, 11*, 43-62. doi:10.2307/3001479

Kendall, M. G., & Gibbons, J. D. (1990). *Rank Correlation Methods.* Oxford University Press.

Kendall, M. G., & Smith, B. B. (1940). On the Method of Paired Comparisons. *Biometrika, 31*, 324-345. doi:10.2307/2332613

Kernan, J. (1967). Paired Comparisons and Graph Theory. *Journal of Marketing Research*, 67-72. doi:10.2307/3150167

Knezek, G., Wallace, S., & Dunn-Rankin, P. (1998). Accuracy of Kendall's Chi-Square Approximation to Circular Triad Distributions. *Psychometrika, 63*, 23-34. doi:10.1007/bf02295434

Koczkodaj, W. W. (1993). A New Definition of Consistency of Pairwise Comparisons. *Mathematical and Computer Modelling, 18*, 79-84. doi:10.1016/0895-7177(93)90059-8

Koczkodaj, W. W., & Szybowski, J. (2016). The Limit of Inconsistency Reduction in Pairwise Comparisons. *International Journal of Applied Mathematics and Computer Science, 26*, 721-729. doi:10.1515/amcs-2016-0050

Koczkodaj, W. W., & Wajch, E. (2015). Can We Compute the Degree of Truth by Pairwise Comparisons? Retrieved from https://s3.amazonaws.com/academia.edu.documents/40559044/KW2015v68.pdf

Koczkodaj, W. W., Kosiek, M., Szybowski, J., & Xu, D. (2015). Fast Convergence of Distance-based Inconsistency in Pairwise Comparisons. *Fundamenta Informaticae, 137*, 355-367. doi:10.3233/FI-2015-1184

Koczkodaj, W. W., Mikhailov, L., Redlarski, G., Soltys, M., Szybowski, J., Wajch, E., . . . Tamazian, G. (2016). Important Facts and Observations about Pairwise Comparisons. *Fundamenta Informaticae, 144*, 291-307. doi:10.3233/FI-2016-1336

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of Measurement Volume I: Additive and Polynomial Representations.* DOVER PUBN INC.

Kristiansen, T. (2011). Analysis of Dominance Hierarchies Using Generalized Mixed Models. Retrieved from https://core.ac.uk/download/pdf/30861380.pdf

Kruskal, J. B., & Wish, M. (1978). *Multidimensional Scaling*. SAGE PUBN.

Kułakowski, K. (2017). Inconsistency in the Ordinal Paired Comparisons Method With and Without Ties. *arXiv preprint arXiv:1702.01126*. doi:10.1016/j.ejor.2018.03.024

Kwiesielewicz, M., & Uden, E. (2004). Inconsistent and Contradictory Judgements in Pairwise Comparison Method in the AHP. *Computers |& Operations Research, 31*, 713-719. doi:10.1016/s0305-0548(03)00022-4

Lahti, K., Koivula, K., & Orell, M. (1994). Is the Social Hierarchy Always Linear in Tits? *Journal of Avian Biology, 25*, 347-348. doi:10.2307/3677283

Laird, R. A., & Schamp, B. S. (2009). Species Coexistence, Intransitivity, and Topological Variation in Competitive Tournaments. *Journal of Theoretical Biology, 256*, 90-95. doi:10.1016/j.jtbi.2008.09.017

Landau, E. (1895). Zur Relativen Wertbemessung der Turnierresultate. *Deusches Wochenschach, 11*, 366-369.

Landau, E. (1914). Über Preisverteilung bei Spielturnieren. *Teubner*. Retrieved from https://www.worldcat.org/title/uber-preisverteilung-bei-spielturnieren/oclc/906365725

Landau, H. G. (1951). On Dominance Relations and the Structure of Animal Societies: I. Effect of Inherent Characteristics. *Bulletin of Mathematical Biophysics, 13*, 1-19. doi:10.1007/bf02478336

Landau, H. G. (1953). On Dominance Relations and the Structure of Animal Societies: III The Condition for a Score Structure. *Bulletin of Mathematical Biophysics, 15*, 143-148. doi:10.1007/BF02476378

Langville, A. N., & Meyer, C. D. (2004). Deeper Inside PageRank. *Internet Mathematics, 1*, 335-380. doi:10.1080/15427951.2004.10129091

Leiva, D., Solanas, A., & Salafranca, L. (2008). Testing Reciprocity in Social Interactions: A Comparison Between the Directional Consistency and Skew-Symmetry Systems. *Behavior Research Methods, 40*, 626-634. doi:10.3758/BRM.40.2.626

Leonard, T. (1977). An Alternative Bayesian Approach to the Bradley-Terry Model for Paired Comparisons. *Biometrics, 33*, 121-132. doi:10.2307/2529308

Li, X., Wang, X., & Xiao, G. (2017). A Comparative Study of Rank Aggregation Methods for Partial and Top Ranked Lists in Genomic Applications. *Briefings in Bioinformatics, 20*, 178-189. doi:10.1093/bib/bbx101

Linares, P. (2009). Are Inconsistent Decisions Better? An Experiment with Pairwise Comparisons. *European Journal of Operational Research, 193*, 492-498. doi:10.1016/j.ejor.2007.11.030

Llull, R. (1283). Artifitium Electionis Personarum. *Manuscript*.

Llull, R. (1299). Ars Electionis.

London, A., Németh, J., & Németh, T. (2014). Time-Dependent Network Algorithm to Ranking in Sports. *Acta Cybernetica-Szeged, 21*, 495-506. doi:10.14232/actacyb.21.3.2014.13

Luce, D. R., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of Measurement, Volume 3: Representation, Axiomization, and Invariance.* Academic Press.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis.* J. Wiley.

Lundh, T. (2006). Which Ball is the Roundest? A Suggested Tournament Stability Index. *Journal of Quantitative Analysis in Sports, 2*. doi:10.2202/1559-0410.1005

Ma, M. (2006). A Matrix Approach to Asset Pricing in Foreign Exchange Market. *SSRN*. doi:10.2139/ssrn.921755

Maas, A. (1993). A Relativized Measure of Circularity for Tournaments. *Mathematical Social Sciences, 26*, 79-91. doi:10.1016/0165-4896(93)90013-9

Maas, A., Bezembinder, T., & Wakker, P. (1995). On Solving Intransitivities in Repeated Pairwise Choices. *Mathematical Social Sciences, 29*, 83-101. doi:10.1016/0165-4896(94)00769-5

Mallows, C. L. (1957). Non-Null Ranking Models. I. *Biometrika, 44*, 114-130. doi:10.2307/2333244

Massey, K. (1997). Statistical Models Applied to the Rating of Sports Teams. Retrieved from https://www.masseyratings.com/theory/massey97.pdf

May, K. O. (1954). Intransitivity, Utility, and the Aggregation of Preference Patterns. *Econometrica, 22*, 1-13. doi:10.2307/1909827

McCormick, T. H., & Zheng, T. (2015). Latent Surface Models for Networks Using Aggregated Relational Data. *Journal of the American Statistical Association, 110*, 1684-1695. doi:10.1080/01621459.2014.991395

McCullagh, P., & Nelder, J. A. (1989). *Generalized Linear Models.* Taylor & Francis Ltd.

McCulloch, W. S. (1948). A Recapitulation of the Theory, with a Forecast of Several Extensions. *Annals of the New York Academy of Sciences, 50*, 259-277. doi:10.1111/j.1749-6632.1948.tb39856.x

McLean, I., & Lorrey, H. (2007). Voting in the Medieval Papacy and Religious Orders. *Modeling Decisions for Artificial Intelligence*, 30-44. doi:10.1007/978-3-540-73729-2_4

Mease, D. (2003). A Penalized Maximum Likelihood Approach for the Ranking of College Football Teams Independent of Victory Margins. *The American Statistician, 57*, 241-248. doi:10.1198/0003130032396

Mengersen, K., & Schmid, C. H. (2013). Maximum Likelihood Approaches to Meta-Analysis. In *Handbook of Meta-Analysis in Ecology and Evolution* (pp. 125-144). doi:10.23943/princeton/9780691137285.003.0010

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and Their Compositionality. *Advances in Neural Information Processing Systems*.

Monsuur, H. (2005). Characterizations of the 3-Cycle Count and Backward Length of a Tournament. *European Journal of Operational Research, 164*, 778-784. doi:10.1016/j.ejor.2003.09.032

Monsuur, H., & Storcken, T. (1997). Measuring Intransitivity. *Mathematical Social Sciences, 34*, 125-152. doi:10.1016/s0165-4896(97)00010-3

Moon, J. W. (1968). *Topics on Tournaments in Graph Theory.* Dover Publications Inc.

Moran, P. A. (1947). On the Method of Paired Comparisons. *Biometrika, 34*, 363-365. doi:10.2307/2332449

Morrison, H. W. (1962). *Intransitivity of Paired Comparison Choices.* Ph.D. dissertation, University of Michigan. Retrieved from https://www.worldcat.org/title/intransitivity-of-paired-comparison-choices/oclc/43626940

Morrison, H. W. (1963). Testable Conditions for Triads of Paired Comparison Choices. *Psychometrika, 28*, 369-389. doi:10.1007/bf02289558

Mosteller, F. (1951). Remarks on the Method of Paired Comparisons: I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations. *Psychometrika, 16*, 3-9. doi:10.1007/bf02313422

Mosteller, F. (1958). The Mystery of the Missing Corpus. *Psychometrika, 23*, 279-289. doi:10.1007/bf02289778

Narayana, T. V., & Bent, D. H. (1964). Computation of the Number of Score Sequences in Round-Robin Tournaments. *Canadian Mathematical Bulletin, 7*, 133-136. doi:10.4153/cmb-1964-015-1

Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Niranjan, U. N., & Rajkumar, A. (2017). Inductive Pairwise Ranking: Going Beyond the n log (n) Barrier. Retrieved from https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14382/14427

Nishizawa, K. (1995). A Consistency Improving Method in Binary AHP. *Journal of Operations Research, Society of Japan, 38*, 21-33. doi:10.15807/jorsj.38.21

Noether, G. E. (1960). Remarks about a Paired Comparison Model. *Psychometrika, 25*, 357-367. doi:10.1007/bf02289753

Nurmi, H. (2014). Are We Done with Preference Rankings? If We are, Then What? *Operations Research and Decisions, 24*, 63-74. doi:10.5277/ord140405

Oliver, D. (2004). Roboscout and the Four Factors of Basketball Success. *Journal of Basketball Studies*. Retrieved from http://www.rawbw.com/deano/articles/20040601_roboscout.htm

Pahikkala, T., Waegman, W., Tsivtsivadze, E., Salakoski, T., & De Baets, B. (2010). Learning Instransitive Reciprocal Relations with Kernel Methods. *European Journal of Operational Research, 206*, 676-685. doi:10.1016/j.ejor.2010.03.018

Pedowitz, L. B. (2008). Report to the Board of Governors of the National Basketball Association. Retrieved from http://assets.espn.go.com/media/pdf/081002/PedowitzReport.pdf

Pelechrinis, K. (2018). LinNet: Probabilistic Lineup Evaluation Through Network Embedding. doi:10.1007/978-3-030-10997-4_2

Pelechrinis, K., Papalexakis, E., & Faloutsos, C. (2016). SportsNetRank: Network-Based Sports Team Ranking. Retrieved from https://pdfs.semanticscholar.org/23a1/b1526690675bbce9648d2d19a86847a3e0fe.pdf

Perron, O. (1907). Zur Theorie der Matrices. *Mathematische Annalen, 64*, 248-263. doi:10.1007/bf01449896

Plackett, R. L. (1975). The Analysis of Permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 24*, 193-202. doi:10.2307/2346567

Raddichi, F. (2011). Who is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis. *PLOS ONE, 6*, e17249.

Ragain, S., & Ugander, J. (2016). Pairwise Choice Markov Chains. *Advances in Neural Information Processing Systems*, 3198-3206.

Raghavachari, M. (2004). Measures of Concordances for Assessing Agreement in Ratings and Rank Order Data. (N. Balakrishnan, N. Kannan, & H. N. Nagaraja, Eds.) *Advances in Ranking and Selection, Multiple Comparisons, and Reliability*, 245-263. doi:10.1007/0-8176-4422-9_14

Rajkumar, A. (2015). Ranking from Pairwise Comparisons: The Role of the Pairwise Preference Matrix. Retrieved from http://etd.iisc.ac.in/handle/2005/3787

Rao, P. V., & Kupper, L. L. (1967). Ties in Paired-Comparison Experiments: A Generalization of the Bradley-Terry Model. *Journal of the American Statistical Association, 62*, 194-204. doi:10.2307/2282923

Rapoport, A. (1949). Outline of a Probabilistic Approach to Animal Sociology: I. *The Bulletin of Mathematical Biophysics, 11*, 183-196. doi:10.1007/bf02478364

Rapoport, A. (1950). Outline of a Probabilistic Approach to Animal Sociology: III. *The Bulletin of Mathematical Biophysics, 12*, 7-17. doi:10.1007/bf02477340

Redmond, C. (2003). A Natural Generalization of the Win-Loss Rating System. *Mathematics Magazine, 76*, 119-126. doi:10.1080/0025570x.2003.11953163

Regenwetter, M., & Dana, J. (2011). Transitivity of Preferences. *Psychological Review, 118*, 42-56. doi:10.1037/a0021150

Reid, K. B., & Santana, M. (2015). Landau's Theorem Revisited Again. *arXiv preprint arXiv:1509.04205*.

Remage, R. J., & Thompson, W. A. (1966). Maximum-Likelihood Paired Comparison Rankings. *Biometrika, 53*, 143-149. doi:10.1093/biomet/53.1-2.143

Saaty, T. L. (1980). *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation (Decision Making Series).* McGraw-Hill.

Sargolzaei, P., & Soleymani, F. (2010). PageRank Problem, Survey and Future Research Directons. *International Mathematical Forum, 5*, 937-956. Retrieved from https://pdfs.semanticscholar.org/ff34/d98113c2279d431f292276b14b29a4acd103.pdf

Savchenko, S. V. (2016). On 5-Cycles and 6-Cycles in Regular n-Tournaments. *Journal of Graph Theory, 83*, 44-77. doi:10.1002/jgt.21914

Schmid, C. H., & Mengersen, K. (2013). Bayesian Meta-Analysis. In *The Handbook of Meta-Analysis in Ecology and Evolution* (pp. 145-173). doi:10.23943/princeton/9780691137285.003.0011

Schönemann, P. H. (1970). A Note on Gulliksen's Least Squares Solution for Incomplete Data. *British Journal of Mathematical and Statistical Psychology, 23*, 69-71. doi:10.1111/j.2044-8317.1970.tb00436.x

Sen, A. (1970). The Impossibility of a Paretian Liberal. *Journal of Political Economy, 78*, 152-157. doi:10.1086/259614

Sen, A. (1999). The Possibility of Social Choice. *American Economic Review, 89*, 349-378. Retrieved from https://www.jstor.org/stable/117024

Shader, B. L. (1992). On Tournament Matrices. *Linear Algebra and Its Applications, 162*, 335-368. doi:10.1016/0024-3795(92)90384-m

Shizuka, D., & McDonald, D. B. (2012). A Social Network Perspective on Measurements of Dominance Hierarchies. *Animal Behaviour, 83*, 925-934. doi:10.1016/j.anbehav.2012.01.011

Siffert, C. (2005). Beatpaths: The Winning Ways of Winners. Retrieved from http://beatpaths.com

Silver, N. (2003). Introducing PECOTA. *Baseball Prospectus*, 507-514.

Silver, N., Boice, J., & Paine, N. (2019). How Our NBA Predictions Work. Retrieved from https://fivethirtyeight.com/methodology/how-our-nba-predictions-work/

Simons, G., & Yao, Y. (1999). Asymptotics When the Number of Parameters Tends to Infinity in the Bradley-Terry Model for Paired Comparisons. *The Annals of Statistics, 27*, 1041-1060. doi:10.1214/aos/1018031267

Sinervo, B., & Lively, C. M. (1996). The Rock-Paper-Scissors Game and the Evolution of Alternative Male Strategies. *Nature, 380*, 240-243. doi:10.1038/380240a0

Sjöberg, L. (1968). The Dimensionality Paradox in Comparative Judgment: A Resolution. *Scandinavian Journal of Psychology, 9*, 97-108. doi:10.1111/j.1467-9450.1968.tb00522.x

Slater, P. (1961). Inconsistencies in a Schedule of Paired Comparisons. *Biometrika, 48*, 303-312. doi:10.2307/2332752

Stern, H. (1990). A Continuum of Paired Comparisons Models. *Biometrika, 77*, 265-273. doi:10.2307/2336804

Stern, H. S., Massey, K., Harville, D. A., Billingsley, R., Colley, W., & Mease, D. (2004). Statistics and the College Football Championship [with Discussions, with Reply]. *The American Statistician, 58*, 179-195. Retrieved from https://www.jstor.org/stable/27643551

Stigler, S. M. (1994). Citation Patterns in the Journals of Statistics and Probability. *Statistical Science, 9*, 94-108. doi:10.1214/ss/1177010655

Stob, M. (1984). A Supplement to "A Mathematician's Guide to Popular Sports". *Mathematical Association of America, 91*, 277-282. doi:10.1080/00029890.1984.11971405

Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1989). *Foundations of Measurement, Vol. 2: Geometrical, Threshold and Probabilistic Representations.* Academic Press.

Takahashi, I. (1990). AHP Applied to Binary and Ternary Comparisons. *Journal of Operations Research, Society of Japan, 33*, 199-206. doi:10.15807/jorsj.33.199

Thurstone, L. L. (1927). A Law of Comparative Judgment. *Psychological Review, 34*, 273. doi:10.1037//0033-295x.101.2.266

Thurstone, L. L. (1927). The Method of Paired Comparisons for Social Values. *The Journal of Abnormal and Social Psychology, 21*, 1927. doi:10.1037/h0065439

Thurstone, L. L. (1931). Rank Order as a Psycho-Physical Method. *Journal of Experimental Psychology, 14*, 187. doi:10.1037/h0070025

Tsai, R., & Böckenholt, U. (2006). Modelling Intransitive Preferences: A Random-Effects Approach. *Journal of Mathematical Psychology, 50*, 1-14. doi:10.1016/j.jmp.2005.11.004

Tsukida, K., & Gupta, M. R. (2011). Rank Aggregation Via Nuclear Norm Minimization. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a543806.pdf

Tucker, A. (2012). *Applied Combinatorics.*

Tufto, J., Solberg, E. J., & Ringsby, T. H. (1998). Statistical Models of Transitive and Intransitive Dominance Structures. *Animal Behaviour, 55*, 1489-1498. doi:10.1006/anbe.1998.0755

Tversky, A. (1969). Intransitivity of Preferences. In *Preference, Belief, and Similarity.* doi:10.1037/h0026750

Usami, S. (2010). Individual Differences Multidimensional Bradley-Terry Model Using Reversible Jump Markov Chain Monte Carlo Algorithm. *Behaviormetrika, 37*, 135-155. doi:10.2333/bhmk.37.135

Vigna, S. (2019). Spectral Ranking. Retrieved from https://arxiv.org/abs/0912.0238v15

Vitelli, V., Sørensen, Ø., Crispino, M., Frigessi, A., & Arjas, E. (2017). Probabilistic Preference Learning with the Mallows Rank Model. *arXiv preprint arXiv:1405.7945*.

Volkovs, M., & Zemel, R. S. (2014). New Learning Methods for Supervised and Unsupervised Preference Aggregation. *Journal of Machine Learning Research, 15*, 1135-1176. Retrieved from http://www.jmlr.org/papers/volume15/volkovs14a/volkovs14a.pdf

Wedel, M., & DeSarbo, W. S. (1993). A Latent Class Binomial Logit Methodology for the Analysis of Paired Comparison Choice Data: An Application Reinvestigating the Determinants of Perceived Risk. *Decision Sciences, 24*, 1157-1170. doi:10.1111/j.1540-5915.1993.tb00508.x

Wei, T. H. (1952). The Algebraic Foundations of Ranking Theory. Retrieved from https://www.repository.cam.ac.uk/handle/1810/250988

West, B. T. (2006). A Simple and Flexible Rating Method for Predicting Success in the NCAA Basketball Tournament. *Journal of Quantitative Analysis in Sports, 2*. doi:10.2202/1559-0410.1039

Weston, J., Bengio, S., & Usunier, N. (2011). WSABIE: Scaling Up to Large Vocabulary Image Annotation. *JCAI, 11*, 2764-2770. doi:10.5591/978-1-57735-516-8/IJCAI11-460

Wormald, N. C. (2004). Tournaments with Many Hamilton Cycles. *preprint*. Retrieved from http://www.math.uwaterloo.ca/ nwormald/papers/hamtourn.pdf

Yang, Y., Qin, T., & Lei, Y. (2016). Real-time eSports Match Result Prediction. *arXiv preprint arXiv:1701.03162*.

Young, H. P. (1988). Condorcet's Theory of Voting. *American Political Science Review, 82*, 1231-1244. doi:10.2307/1961757

Young, H. P., & Levenglick, A. (1978). A Consistent Extension of Condorcet's Election Principle. *SIAM Journal on Applied Mathematics, 35*, 285-300. doi:10.1137/0135023

Zermelo, E. (1929). Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift, 29*, 436-460. doi:10.1007/BF01180541