

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Vergleich und Kategorisierung von Ranking- und Matchmaking-Systemen im Sport: Ein systematisches Review

verfasst von | submitted by

Paul Resch BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 826

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Sportwissenschaft

Betreut von | Supervisor:

Univ.-Prof. Dipl.-Ing. Dr. Arnold Baca

Abstract

Einleitung: Rankings sind in unserem alltäglichen Leben fest verankert. Sie geben uns einen Überblick, welche die besten Restaurants oder Hotels in unserer Nähe sind, welche Länder am unfreundlichsten sind oder wer aktuell der beste Tennisspieler oder die beste Tennisspielerin der Welt ist. Im Sport sind die Rankings ausschlaggebend, um die Entwicklung eines Teams oder einer spielenden Person ständig nachvollziehen und verfolgen zu können.

Neben den Rankings nimmt auch das Matchmaking eine Bedeutung ein. Der Prozess des Matchmaking stellt sicher, dass Wettkämpfe spannend, fair und ausgeglichen sind. Auch im E-Sport ist die Bedeutung eines ausgereiften und effizienten Matchmaking-Systems immer ein wichtiges und aktuelles Thema. Besonders die Faktoren der Fairness und der Einheitlichkeit müssen in einem guten Matchmaking-System harmonisieren.

Methode: Diese Übersichtsarbeit wird anhand der PRISMA Guidelines durchgeführt. In den Datenbanken SPORTDiscus, SURF und Web of Science wurde zwischen Oktober und Dezember 2024 eine Literatursuche durchgeführt. Inkludiert wurden Studien, die sich mit der Zusammensetzung eines Rankings- und/oder Matchmaking-Systems befassen und in englischer Sprache verfasst sind. Studien die sich mit medizinischen, trainingswissenschaftlichen oder psychologischen Themen befassen wurden ausgeschlossen. Ebenso wurden Studien, die sich hauptsächlich mit der Vorhersage von Spielergebnissen auseinandersetzten, nicht inkludiert.

Resultate: Es wurden anfänglich 1636 Studien durch die Anwendung eines Search Strings identifiziert. Nach einem Literaturscreening wurden 29 Studien für die Beantwortung der Forschungsfragen herangezogen. Es werden Ranking- und Matchmaking-Systeme für 11 verschiedene Sportarten behandelt. Von den 29 Studien, befassen sich 21 nur mit einem Rankingsystem. Nur vier Studien behandeln das Thema Matchmaking. Die restlichen Studien beschäftigen sich mit einer Kombination aus beiden Faktoren. Es ergeben sich vier Kategorien: Individual Player Performance Ranking/Rating, Team-Performance-Ranking/Rating, Matchmaking-Systeme und Kombination aus Ranking- und Matchmaking-Systemen.

Zusammenfassung: Die Ergebnisse lassen unterschiedliche Ansätze für Ranking und Matchmaking erkennen, es gibt jedoch auch viele Gemeinsamkeiten. So ergeben sich auch weitere Kategorisierungsmöglichkeiten, während sich mehrere Systeme mit einer ähnlichen mathematischen Basis wiederfinden. Das Streben nach Fairness ist in allen Ansätzen zu erkennen.

Abstract (english)

Introduction: Rankings have become an integral part of our daily lives. They give us an overview of which are the best restaurants or hotels in our neighbourhood, which countries are the least friendly, or who is currently the best tennis player in the world. In sport, rankings are crucial for understanding and tracking the development of a team or player.

In addition to the rankings, matchmaking also plays an important role. But without the matchmaking factor, something would be missing in today's sport: The process of matchmaking ensures that competitions are exciting, fair and balanced. In e-sports, the importance of a sophisticated and efficient matchmaking system is always an important and topical issue. In particular, the factors of fairness and consistency need to be balanced in a good system.

Methods: This review is based on the PRISMA guidelines. A literature search was conducted in the SPORTDiscus, SURF and Web of Science databases between October and December 2024. Studies were included if they dealt with the construction of a ranking and/or matchmaking system and were written in English. Studies dealing with medical, training or psychological topics were excluded. Studies that mainly dealt with the prediction of match results were also excluded.

Results: Initially, 1636 studies were identified using a search string. After literature screening, 29 studies were used to answer the research questions. Ranking and matchmaking systems for 11 different sports are covered. Out of 29 studies, 21 are discussing a ranking system, while only four studies deal with matchmaking. The rest deals with a combination out of both categories. There are four categories: Individual Player Performance Ranking/Rating, Team Performance Ranking, Matchmaking-Systems, Combination of Ranking and Matchmaking Systems.

Summary: The results reveal different approaches to ranking and matchmaking, but there are also similarities. This allows further categorisations to be identified, while several systems with a similar mathematical basis can be found. The pursuit of fairness is evident in all approaches.

Erklärung

„Ich erkläre, dass ich die vorliegende Arbeit *selbstständig verfasst habe* und nur die ausgewiesenen Hilfsmittel verwendet habe. Diese Arbeit wurde weder an einer anderen Stelle eingereicht (z. B. für andere Lehrveranstaltungen) noch von anderen Personen (z. B. Arbeiten von anderen Personen aus dem Internet) vorgelegt.“

Inhaltsverzeichnis

Abstract	3
Abstract (english)	5
Erklärung	7
1 Einleitung	13
1.1 Zieldefinition	13
1.2 Definition von Ranking	13
1.3 Rankings außerhalb des Sports	13
1.4 Rankings im Sport	14
1.4.1 Arten von Rankingsystemen im Sport	15
1.5 Historische Entwicklung von Rankings	16
1.6 Das ELO-Ratingsystem	17
1.7 Das ATP-Ranking	18
1.8 Matchmaking	19
1.9 Die Beziehung von Ranking und Matchmaking	20
1.10 Matchmaking im E-Sport	20
2 Methode	22
2.1 Auswahlkriterien	22
2.1.1 Population	22
2.1.2 Exposure	22
2.1.3 Outcomes	22
2.2 Ein- und Ausschlusskriterien	22
2.3 Informationsquellen	23
2.4 Studiendesign	23
2.5 Suchstrategie	24
2.6 Auswahlverfahren	25
2.7 Datenerfassung	25
2.8 Datenelemente	25
2.9 Risk of Bias	25
3 Ergebnisse	26
3.1 Ausgewählte Studien	26

3.2 Charakteristik der Studien.....	27
3.3 Ergebnisse im Detail.....	28
3.3.1 Individual Player Performance Ranking/Rating	29
3.3.2 Team-Performance-Ranking/Rating	31
3.3.3 Matchmaking-Systeme	33
3.3.4 Kombination aus Ranking- und Matchmaking-Systemen	35
4 Diskussion	37
4.1 Mathematische Hintergründe und Zusammenhänge	37
4.1.1 Colley- und Massey Methode.....	38
4.1.2 Page-Rank Algorithmus und iterative Prozesse	40
4.1.3 Matchmaking	42
4.2 Vergleich und Analyse der vier Kategorien.....	44
4.2.1 Zeitliche Verläufe.....	45
4.2.2 Die verschiedenen Sportarten und ihre Systeme	46
4.2.3 Top-down und Bottom-up-Ratings.....	48
4.2.4 Match Outcome und Match Aktion.....	49
4.2.5 Individualsport	50
4.2.6 E-Sport.....	50
4.3 Fairness und Heimvorteil	51
4.3.1 Fairness	51
4.3.2 Fairness in Onlinespielen	53
4.3.3 Einheitlichkeit in Onlinespielen.....	53
4.3.4 Heimvorteil.....	54
Literaturverzeichnis.....	59
Abbildungsverzeichnis	67
Tabellenverzeichnis	68
Abkürzungsverzeichnis	69
Anhang	71
Die Studien in der Übersicht	71
Kategorie I: Individual Player Performance Ranking/Rating.....	71
Ambrutis & Povilatis, 2024: Composite rating method: Application to European basketball leagues	71

Blackburn, 2013: Ranking the performance of tennis players: An application to women's professional tennis	72
Broadie & Rendleman, 2013: Are the official world golf rankings biased?	73
Hvattum & Gelade, 2021: Comparing bottom-up and top-down ratings for individual soccer players.....	73
Irons et al., 2014: Developing an improved tennis ranking system	74
Kim & Jeon, 2019: Proposal for implementation of a ranking model for Olympic Taekwondo competitions using PageRank.....	74
Servien, 2022: An alternative ranking system for biathlon pursuit races	75
Wolf et al., 2021: A football player rating system.....	75
Kategorie II: Team-Performance-Ranking/Rating	77
Barrow et al., 2014: Ranking rankings: an empirical comparison of the predictive power of sports ranking methods.....	77
Chartier et al., 2011: Sports Ranking with Nonuniform Weighting	77
Colley, 2002: Colley's bias free college football ranking method: The colley matrix explained...	78
Constantinou & Fenton, 2013: Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries.	78
Duncan & Parece, 2024: Population-adjusted national rankings in the Olympics.....	79
Gill & Keating, 2009: Assessing Methods for College Football Rankings	80
Govan et al., 2009: Offense-Defense Approach to Ranking Team Sports	80
Jeremic & Radojicic, 2010: A New Approach in the Evaluation of Team Chess Championships Rankings	81
Lowery et al., 2020: Functional ratings in sports.....	82
Mangan & Collins, 2016: A Rating System For Gaelic Football Teams: Factors That Influence Success.....	83
Pradhan & Abdourazakou, 2020: "Power ranking" professional circuit eSports teams using multi-criteria decision-making (MCDM).....	83
Stewart, 2013: A proposed general rating system for which the Colley Matrix Rating System is a special case	84
Wigness et al., 2010: A New Iterative Method for Ranking College Football Teams.....	85
Kategorie III: Matchmaking-Systeme.....	85
Gong et al., 2020: OptMatch: Optimized Matchmaking via Modeling the High-Order Interactions on the Arena	85

Mysłak & Deja, 2015: Developing Game-Structure Sensitive Matchmaking System for Massive-Multiplayer Online Games.....	86
Sauer et al., 2024: Improving ranking quality and fairness in Swiss-system chess tournaments	87
Wang et al., 2024: EnMatch: Matchmaking for Better Player Engagement via Neural Combinatorial Optimization.....	87
Kategorie IV: Kombination aus Ranking- und Matchmaking-Systemen.....	88
Matthews et al., 2021: Application of PageRank Algorithm to Division I NCAA men's basketball as bracket formation and outcome predictive utility	88
Miles et al., 2010: AccuV College Football Ranking Model	90
Raack et al., 2014: Standings in sports competitions using integer programming	90
Reinig & Horowitz, 2019: Analyzing the impact of the NCAA Selection committee's new quadrant system	91

1 Einleitung

1.1 Zieldefinition

Als Ziel der Arbeit kann die Klärung der Frage, welche Unterschiede es in den Ranking- und Matchmaking-Systemen zwischen den verschiedenen Sportarten gibt, bezeichnet werden. Dabei sollen verschiedene Systeme in sowohl Einzel-, als auch Teamsportarten beleuchtet und folglich kategorisiert werden. Durch eine strukturierte Literatursuche sollen Studien identifiziert werden, anhand derer sich Zusammensetzung, Anwendung und mathematischer Hintergründe erkennen lassen.

Weiters ist die Frage nach dem Einfluss von Faktoren wie Heimvorteil oder die Fairness auf sportliche Ranking- und Matchmaking-Systemen, von Relevanz.

1.2 Definition von Ranking

Ein Ranking kann als eine geordnete Liste bezeichnet werden, in welcher Personen, Objekte oder Leistungen durch bestimmte Kriterien miteinander verglichen werden. Durch die Zuordnung eines Rangs wird die relative Position auf einer Liste angegeben - so wird etwa vom schlechtesten bis zum besten sortiert (Langville & Meyer, 2012). Rankings finden in einer Vielzahl von Kontexten Anwendung, darunter im Sport, in wissenschaftlichen Disziplinen (beispielsweise Hochschulrankings), oder im Rahmen von Produkttests. Das Ziel von Rankings ist die Schaffung einer transparenten und vergleichbaren Darstellung von Leistungen oder Qualitäten (Ringel & Werron, 2020).

1.3 Rankings außerhalb des Sports

Rankings werden als soziale Operationen betrachtet, die öffentliche Wettbewerbsformen konstruieren (Ringel & Werron, 2020). Beispiele hierfür sind etwa der Human Development Index¹ oder das World University Ranking². Moderne Ranking-Systeme kombinieren vier analytisch unterschiedliche Prozesse: den Vergleich von Leistungen, die Quantifizierung, die Visualisierung und die wiederholte Veröffentlichung (Ringel & Werron, 2020). Diese vier Operationen tragen zur sozialen Konstruktion von Wettbewerbsfeldern bei. Anhand von soziologischen Vergleichen wird versucht, Einzelheiten in Kategorien einzuteilen und messbar zu machen. Dafür ist weiters aber auch ein tieferes Verständnis der sozialen Dynamik notwendig (Heintz, 2010). Während die ersten beiden Operationen die informative Dimension der Rankings darstellen, wird besonderer Fokus auf die performative Dimension gelegt, die durch die dritte und vierte Operation definiert wird. Dies bedeutet, dass Rankings

¹ Human Development Index: Beschreibt die menschliche Entwicklung in den Ländern der Vereinten Nationen. <https://hdr.undp.org/data-center/human-development-index#/indicies/HDI>

² World University Ranking: Ein Ranking der besten Universitäten von Times Higher Education. <https://www.timeshighereducation.com/world-university-rankings/latest/world-ranking>

Universitäten oder Länder dazu zwingen, ihre Leistungen zu verbessern und ihr Ansehen auf Kosten anderer zu steigern. Ein höherer Platz in einem Ranking bringt neben einem hohen Ansehen auch eine Zunahme an Studenten und Studentinnen mit sich (Pollock & D'Adderio, 2012; Ringel & Werron, 2020).

Rankings haben demnach eine große Auswirkung auf Organisationen. Deren Routinen und Strukturen werden sich je nach Einordnung in einem Ranking ändern. Ein Unternehmen, welches schlecht bewertet wurde, arbeitet wie eine Universität an seinen Strategien, um wieder mehr potenzielle Kunden und Kundinnen von sich zu überzeugen (Espeland & Sauder, 2007). Unternehmen oder Universitäten, die in einem elitären Ranking einen hohen Platz einnehmen haben einen gewissen Ruf, der es oft leichter macht an der Spitze zu bleiben (Schultz et al., 2001). Ist dieser Ruf durch ein schlechtes Ranking einmal verschlechtert, ist es schwierig diesen wieder zu ändern (Kelley & Simmons, 2015). Bei Hochschulen oder Universitäten ist der Ruf - wie auch schon weiter oben erwähnt - entscheidend in deren Fähigkeit Studierende sowie entsprechende Partner für die finanzielle Unterstützung anzuziehen (Kelley & Simmons, 2015).

Rankings werden ständig aktualisiert, erneuert und publiziert. Somit stehen etwa Unternehmen ständig im Rampenlicht und müssen um die Aufmerksamkeit des „Publikums“ konkurrieren. Auch dieser Aspekt ist demnach sehr entscheidend, weil erneut ein Wettbewerb entsteht, der durch das Bestreben der Unternehmen zu Qualitätsverbesserungen führen kann (Ringel & Werron, 2020).

1.4 Rankings im Sport

Bewertung und das Ranking sind wichtige Elemente von Sportwettbewerben und der sie umgebenden Unterhaltungslandschaft. Im Allgemeinen hat die Wertung eine informative Funktion, die Fans und profanen Beobachtern einen schnellen Einblick in die relative Stärke der Mannschaften oder der Athleten und Athletinnen gibt. Die Presse interessiert sich beispielsweise häufig für die am besten bewerteten Mannschaften oder für die Nationalmannschaft, die eine bestimmte Position in einer Rangliste erreicht hat (Szczecinski & Roatis, 2022). Noch bedeutender ist, dass die Rangliste zu Folgeentscheidungen führt, wie zum Beispiel die Setzliste, also die Bestimmung, welche Mannschaften in bestimmten Wettbewerben gegeneinander antreten müssen: Welches Land wird in welche Gruppe bei der FIFA-Weltmeisterschaft gelost. Weitere Beispiele wären etwa Auf- und Abstieg in die verschiedensten Ligen (z. B. zwischen der englischen Premier League und der EFL-Championship [2.Liga]), oder die Teilnehmer an prestigeträchtigen Wettbewerben am Ende der Saison (wie die Stanley-Cup-Serie in der National Hockey League (NHL) zu bestimmen. Die Rankings sind demnach auch im

Kontext des Matchmakings von Relevanz (Szczecinska & Roatis, 2022). Mit Blick auf den Reitsport dienen Rankings dazu, um die Leistungsfähigkeit von Pferden zu beurteilen und damit wertvolle Informationen aufbereiten zu können (Hartmann et al., 2014).

1.4.1 Arten von Rankingsystemen im Sport

Es gibt heutzutage mehrere hundert Rankingmethoden um Mannschaften und einzelne Spieler: innen bewerten zu können (Barrow et al., 2013). Einer der wichtigsten Begriffe ist der des „Power Rankings“ (Pradhan & Abdourazakou, 2020). Mit dieser Technik wird die Stärke einer Mannschaft im Vergleich zu den anderen Teams der Liga relativiert. Neben Siegen und Niederlagen fließen in diese Rankings auch die Stärke des Spielplans³, der aktuelle Leistungstrend sowie individuelle Spielerstatistiken mit ein (Barrow et al., 2013).

Wenn der Fall auftritt, dass zwei Parteien die gleiche Leistung erbracht haben, muss es eine Möglichkeit geben zu determinieren, welches Team, welcher Spieler oder welche Spielerin, höher eingestuft werden muss. Diese Systeme werden als „Tie-breaking Systeme“ bezeichnet (Pakaslahti, 2019). Darunter fallen auch sogenannte Head-to-Head-Vergleiche, durch welche der direkte Vergleich zwischen zwei Konkurrenten um einen Platz in der Rankingliste leichter gezogen werden kann (Pakaslahti, 2019).

Weiters sind die Handicap-Systeme zu nennen, die dazu dienen, gleiche Wettbewerbsbedingungen für Athleten und Athletinnen mit unterschiedlichen Fähigkeiten zu schaffen, einen fairen Wettbewerb zu ermöglichen und die Teilnahme zu fördern. Diese Systeme passen die Punktzahlen oder Leistungsanforderungen an die Fähigkeiten der einzelnen Athleten und Athletinnen an, so dass jeder die Möglichkeit hat, effektiv an Wettkämpfen teilzunehmen (Pakaslahti, 2019).

Unter den Begriff des Handicaps fällt auch der Behindertensportbereich. Anhand von Klassifizierungssystemen sollen Sportler:innen anhand ihrer Voraussetzungen und Einschränkungen in Gruppen eingeteilt werden. Dadurch sollen Leistungen vergleichbarer und der Wettkampf fairer für alle Teilnehmer:innen werden. Besonders bei Wettbewerben wie den paraolympischen Spielen, bei welchen die Differenzen in den physischen und psychischen Voraussetzungen für die Ausübung einer Sportart sehr vielfältig sind, ist eine Klassifizierung entscheidend. Bei der Klassifizierung kann zwischen drei Arten unterschieden werden: Erstens wird nach der Art der Behinderung bewertet. Das schließt etwa Blinde und Sehbehinderte, Cerebrale Parese oder Amputierte Sportler:innen mit ein. Zweitens wird nach dem Ausmaß der Funktion, trotz der Einschränkung eine Sportart ausüben zu können, klassifiziert. Drittens wird nach dem Handicap bewertet, indem der

³ Gibt die durchschnittliche Qualität der Gegner an und ist Indikator dafür, wie anspruchsvoll die kommenden Spiele eines Teams sein werden (engl. Strength of Schedule)

Funktionsverlust numerisch beziffert und somit die sportliche Leistung eingeordnet wird (Palmen, 2014).

Der Behindertensport wird im Rahmen dieser Übersichtsarbeit nicht behandelt, damit die Resultate kompakter gehalten werden. Es gilt dennoch anzumerken, dass die Rankings auch im Behindertensport eine essenzielle Rolle spielen. Durch die Klassifizierung lässt sich weiters auch das Matchmaking ableiten, weil dadurch bestimmt wird, wer gegen wen in welcher Gruppe in einem sportlichen Wettkampf antritt (Palmen, 2014).

1.5 Historische Entwicklung von Rankings

Erst in der zweiten Hälfte des 19. Jahrhunderts, als der moderne Sport durch standardisierte Regeln, neue Wettkampfmodi und regelmäßige öffentliche Leistungsvergleiche zu einem eigenständigen Bereich wurde, entstanden Sportranglisten (Werron, 2015). Durch kulturelle und soziale Faktoren wurde es wünschenswert, dass die Menschen ihre Leistungen untereinander vergleichen können. Das schulische Notensystem in Schulen gilt hierbei als ideengebende Struktur, weil auch in diesem Fall ein Wettbewerb gefördert wird und die Leistungen vergleichbar gemacht werden. Anhand der Rankings wurde folglich auch eine Vergleichbarkeit für die Öffentlichkeit garantiert. (Bührer, 2019; Werron, 2015).

Die ersten Sportranglisten entwickelten sich im Zusammenhang mit der Erfindung des Ligasystems. Dieses wurde Anfang der 1870er Jahre im US-amerikanischen Baseball eingeführt und kurz darauf von der britischen Football League übernommen. Das Ligasystem ist eine wichtige Innovation im Sport, da es eine neue Art der zeitlichen Organisation einführt. Es dauert in der Regel eine Saison, um einen Meister zu ermitteln und ermöglicht eine größere Anzahl von Teilnehmern (Ringel & Werron, 2020).

Seitdem haben sich Ligasysteme und andere serielle Wettbewerbsformen weltweit verbreitet und sind zu den wichtigsten Formen des modernen Sportwettbewerbs geworden. Um den Verlauf eines Wettbewerbs verfolgen zu können, musste es Vergleichsmöglichkeiten geben. Eine Ligatabelle dient genau diesem Zweck, sie kann aktuelle Ergebnisse widerspiegeln. Der Erfolg der Ranglisten hängt daher auch sehr viel mit der Einführung des Ligasystems zusammen (Ringel & Werron, 2020).

Das Aufkommen von Rankings führte in weiterer Folge zu einer Vielzahl an individuellen Statistiken über Teams und einzelne Spieler:innen. Diese basierten auf Kategorien, wie „batting average“ (Schlagdurchschnitt) im Baseball oder „erzielte Tore“ im Fußball. Diese individuellen Statistiken konnten nun auch als Ausdruck für besondere Leistungen über den Verlauf einer Karriere herangezogen werden. Im amerikanischen Baseball wurden solche individuellen Ranglisten bereits in den 1880er Jahren geführt (Werron, 2015).

1.6 Das ELO-Ratingsystem

Das Elo-Ratingsystem wurde in den 1950er Jahren von dem ungarisch-amerikanischen Mathematiker Arpad Elo entwickelt und ist eine Methode zur Berechnung der relativen Fähigkeiten von Spielern und Spielerinnen in Zwei-Spieler-Spielen wie Schach (Elo, 1978). Das System basiert auf der Prämisse, dass die Leistung eines Spielers oder einer Spielerin quantifiziert und mit anderen Personen verglichen werden kann, wodurch ein dynamisches Ranglistensystem ermöglicht wird, dass sich nach jedem Spiel anpassen kann (Glickman, 1999).

Das Elo-Bewertungssystem wird nicht nur im Schach, sondern auch in anderen Sportarten und Wettkampfbereichen verwendet (Powell, 2023). Es wurde angepasst, um auch Mannschaftssportarten, wie Basketball und Fußball, zu berücksichtigen, bei denen sowohl individuelle Fähigkeiten als auch Teamarbeit eine Rolle spielen. Neuerungen zielen darauf ab, das System auf Mehrspielerwettbewerbe auszuweiten, da traditionelle Paarvergleiche die Komplexität von Mehrspielerwettbewerben nicht vollständig erfassen können. Powell (2023) zeigt auf, dass man erkannt hat, wie wichtig es ist, relative Ranglisten zu haben, aber auch Faktoren wie Kontextvariablen und Spielinteraktionen in die Bewertung der Teilnehmer:innen einzubeziehen.

Im Elo-System hat jeder Spielende eine numerische Bewertung, welche die Spielstärke jener Personen darstellt. Wenn zwei Spieler:innen gegeneinander antreten, wird das erwartete Ergebnis auf der Grundlage ihrer Bewertungen berechnet. Gewinnt die spielende Person mit der höheren Wertungszahl, steigt seine Wertungszahl leicht an, während die Wertungszahl der spielenden Person mit der niedrigeren Wertungszahl sinkt. Umgekehrt gewinnt der oder diejenige mit der niedrigeren Bewertung eine beträchtliche Anzahl von Punkten, während die spielende Person mit der höheren Bewertung Punkte verliert. Dies spiegelt die Unvorhersehbarkeit des Ergebnisses wider und bietet eine genauere Darstellung der Fähigkeiten der Spieler:innen im Laufe der Zeit (Elo, 1978; Powell, 2023).

Ein entscheidender Aspekt im Elo-System ist die erwartende Gewinnwahrscheinlichkeit. Die Wahrscheinlichkeit, dass Spieler A gegen Spieler B gewinnt, kann mit folgender Formel dargestellt werden:

$$E_A = \frac{1}{1 + 10^{(R_B - R_A)/400}}$$

E_A stellt hierbei die zu erwartende Gewinnwahrscheinlichkeit von Spieler A dar. R_A steht für das Elo-Rating von Spieler A, während R_B das Elo-Rating von Spieler B bezeichnet (Wolf et al., 2021).

Wenn im Elo-System zwei Spieler:innen gegeneinander antreten, wird die erwartete Punktzahl für jeden Spieler oder jede Spielerin auf der Grundlage seiner aktuellen Bewertungen berechnet. Das tatsächliche Ergebnis (Sieg, Niederlage oder Unentschieden) wird dann mit diesem erwarteten Ergebnis verglichen, um ihre Bewertungen anzupassen. Ein entscheidender Bestandteil dieses Anpassungsprozesses ist der sogenannte k -Faktor. Dadurch kann bemessen werden, wie viele Elo-Punkte eine spielende Person in einem Spiel höchstens hinzugewinnen oder maximal verlieren kann. Somit wird die maximal mögliche Änderung der Bewertung einer Person nach einer Partie bestimmt. Ein höherer k -Faktor führt zu deutlicheren Änderungen der Bewertung eines Spielers oder einer Spielerin, sodass das System besser auf aktuelle Leistungen reagieren kann. Umgekehrt stabilisiert ein niedrigerer k -Faktor die Bewertungen, wodurch sie weniger anfällig für Schwankungen sind (Albers & Vries, 2001). In Sportarten, wie Tennis oder Fußball, in welchen häufig gespielt wird ist ein eher niedriger k -Faktor bevorzugt, um die Ratings stabil halten zu können. In wenig frequentierten Sportarten wie Gaelic Football, wird ein höherer k -Faktor präferiert. Durch diese Anpassungen können sich die Ratings schneller an die aktuellen Leistungen anpassen (Mangan & Collins, 2016).

In der Regel kann der k -Faktor je nach Spielniveau variieren. So kann er beispielsweise für neue Spieler:innen höher gesetzt werden, um schnell ihr wahres Können wiederzugeben, und für etablierte Spieler:innen niedriger, um die Stabilität aufrechtzuerhalten (Langville & Meyer, 2012; Powell, 2023).

1.7 Das ATP-Ranking

Das Ranglistensystem der ATP (Association of Tennis Professionals) ist eine Methode zur Bestimmung des relativen Leistungsniveaus männlicher Tennisprofis. Die im Jahr 1973 eingeführte ATP-Rangliste ist entscheidend für die Einstufung der Spieler bei Turnieren, für die Zulassung zu Veranstaltungen und für die Beurteilung der Leistung der Spieler im Laufe der Zeit (Dingle et al., 2013; Irons et al., 2014). Die ATP-Rangliste basiert auf der Leistung eines Spielers bei verschiedenen Turnieren über einen Zeitraum von 52 Wochen. Die Spieler erhalten Punkte auf der Grundlage ihrer Leistungen bei diesen Turnieren, wobei prestigeträchtigere Turniere mehr Punkte erhalten. Das Rankingsystem unterteilt die Turniere in verschiedene Kategorien, darunter Grand Slams, ATP-Finals, Masters 1000, ATP 500 und ATP 250. Bei Grand Slam Turnieren, wie Wimbledon und den US Open, gibt es die höchste Punktezahl. Der Sieger erhält 2000 Punkte. Die Punkte werden danach vergeben, wie weit ein Spieler im Turnier kommt. So erhält ein Spieler, der das Finale eines Grand-Slam-Turniers erreicht, deutlich mehr Punkte als ein Athlet, der in der ersten Runde ausscheidet. Außerdem erhalten die Spieler Punkte für Doppelspiele, die jedoch getrennt von der Einzelrangliste berechnet werden (atptour.com, 2024; Dingle et. al, 2012).

Die ATP-Rangliste wird wöchentlich, im Normalfall jeden Montag, aktualisiert. Die Rangliste jedes Spielers wird aus den 18 besten Ergebnissen des vergangenen Jahres ermittelt. Das bedeutet, dass sich die Platzierung eines Spielers je nach seinen Leistungen bei den letzten Turnieren ändern kann, da ältere Ergebnisse nach 52 Wochen verloren gehen. Wenn ein Spieler bei einem Turnier nicht gut abschneidet, läuft er Gefahr, Punkte zu verlieren, was sich auf seine Gesamtwertung auswirken kann. Darüber hinaus berücksichtigt die ATP-Rangliste auch die Teilnahme an Turnieren. Die Spieler müssen an einer Mindestanzahl von Turnieren teilnehmen, um ihren Rang zu halten und Strafpunkte zu vermeiden. Verletzungen und Rücktritte können sich auf die Rangliste auswirken, da die Spieler möglicherweise Chancen verpassen, Punkte zu sammeln (Irons et al., 2014).

Die ATP-Rangliste spiegelt nicht nur die aktuelle Leistung wider, sondern dient auch als historische Aufzeichnung der Karriere eines Spielers. Das Erreichen einer hohen Platzierung ist eine bedeutende Leistung und kann die Vermarktungsfähigkeit eines Spielers, seine Sponsorentätigkeit und sein gesamtes Vermächtnis in diesem Sport beeinflussen. So sind die Nachwirkungen der Verletzungen, die etwa Dominic Thiem in seiner Karriere erlitten hat, am Verlauf des ATP-Rankings nachvollziehbar (Chen et al., 2023).

1.8 Matchmaking

Matchmaking im Sport bezieht sich auf den Prozess der Zusammenstellung von Teams oder Einzelspielern und Einzelspielerinnen für Wettkämpfe, um sicherzustellen, dass die Wettkämpfe fair, spannend und konkurrenzfähig sind. Dieser Aspekt ist in einer Vielzahl von Sportarten, vom Amateur- bis hin zum Profibereich, von entscheidender Bedeutung. Dadurch wird auch das Gesamterlebnis für Sportler:innen und Zuschauer:innen verbessert. Ein effektives Matchmaking berücksichtigt verschiedene Faktoren, wie Leistungsniveau, bisherige Leistungen und Teamdynamiken, um ausgewogene Paarungen zu schaffen, die das Engagement fördern und Talente zur Geltung bringen (Powell, 2023).

In Sportarten, in denen mehrere Teams oder Einzelspieler:innen um den Titel kämpfen, ist Matchmaking besonders wichtig für Turniere und Ligen. Durch die sorgfältige Analyse von Statistiken und vergangenen Leistungen können Spielpläne erstellt werden, die das Potenzial für spannende Wettkämpfe maximieren und gleichzeitig Ungleichgewichte minimieren, die zu einseitigen Ergebnissen führen könnten. Dieser Prozess ist besonders wichtig im Jugend- und Amateursport, wo der Schwerpunkt häufig noch auf Entwicklung und Teilnahme und nicht auf dem Einfahren von Gewinn liegt (Zrike & Zrike, 2010).

Darüber hinaus hat der technologische Fortschritt den Matchmaking-Prozess revolutioniert. Datenanalyse, künstliche Intelligenz und Algorithmen spielen heute eine wichtige Rolle bei

der Bewertung der Fähigkeiten von Spielern und Spielerinnen und der Vorhersage von Spielergebnissen. Dies erhöht nicht nur die Fairness der Wettbewerbe, sondern fügt auch ein strategisches Element für Trainer:innen und Mannschaften hinzu, wenn sie sich auf ihre Gegner vorbereiten (Zrike & Zrike , 2010).

Es geht beim Matchmaking letzten Endes um mehr als nur um das Organisieren von Spielen. Es kann als Kunst angesehen werden, die Wettbewerb, Fairness und Unterhaltung miteinander in Einklang bringt (Gebser et al., 2013).

1.9 Die Beziehung von Ranking und Matchmaking

Ranking und Matchmaking sind von ihrer Definition her unterschiedliche Begriffe, doch sie sind dennoch zwei miteinander verbundene Komponenten, die aus dem Sport nicht wegzudenken sind. Die Rankings geben den Rahmen vor und helfen dabei, dass Match-ups ausgeglichen gestaltet und faire Wettbewerbsbedingungen geschaffen werden können. So zeigt sich diese Synergie zwischen Ranking und Matchmaking in den diversen Sportligen und Turnieren. Beispielsweise legt etwa die Rangliste in den deutschen Fußballprofiligen fest, welche Teams in der ersten Liga , und welche Teams in der zweiten Liga gesetzt sind (Raack et al., 2014). Ebenso bestimmen die ATP- und WTA Rankings die Setzlisten für diverse Turniere. Effektives Matchmaking basierend auf genauen Rankingsystemen erhöht in allen Sportarten die Spannung im Wettbewerb und es trägt dadurch auch zu einem Wachstum im Sport bei (Blackburn M., 2013).

1.10 Matchmaking im E-Sport

Nicht nur in der „realen“ Welt ist Matchmaking von Bedeutung: Besonders im E-Sport nimmt es eine sehr wichtige Stellung ein. Ein Onlinespiel kann noch so gut gestaltet sein, wenn kein passendes Match-up gefunden wird, kann dadurch das Spiel auf viele Spieler:innen schnell uninteressant wirken. Ein computergesteuertes und vor allem effizientes Matchmaking-System ist daher im E-Sport eminent wichtig. Für die Millionen Spieler:innen weltweit müssen schnellstmöglich passende Match-ups kreiert werden (Alman & McKay, 2017).

Die meisten Matchmaking-Systeme für Onlinespiele sind meist „skill-based“. Die spielende Person wird anhand ihrer absoluten Stärke oder Fähigkeit eingeordnet und der Algorithmus versucht somit Teams ausgeglichen zu gestalten (Gong et al.,2020). Dieser Fähigkeitswert wird meist durch probabilistische Algorithmen wie, ELO, Glicko, dem TrueSkill-Ratingsystem oder dem Bradley-Terry-Model sowie durch deren Erweiterungen ermittelt (Glickman, 1999; Herbrich et al., 2007).

Immer wieder kommt es im E-Sport zu Problemen im Matchmaking, wie beispielsweise die Warteschlange vor Beginn eines Onlinespiels oder die Rollenauswahl verschiedener

„Ingame-Charaktere“. Durch neu entwickelte Algorithmen und Systeme wird dem entgegengewirkt, indem unter anderem die Fähigkeiten der einzelnen Spieler:innen einbezogen werden, um für ausgeglichene Teams zu sorgen. Da das Feld des E-Sports immer mehr an Nutzer und Nutzerinnen gewinnt, gilt es ein genau definiertes Maß zu bestimmen, um die besten Match-ups kreieren zu können (Gong et al., 2020; Wang et al., 2024). „League of Legends“ („LOL“) und „Counter Strike“ (CS), zwei der populärsten Onlinespiel, gelten informell dann als ausgewogen, wenn folgende zwei Faktoren erfüllt sind: Fairness und Einheitlichkeit (Alman & McKay, 2017)

2 Methode

Im folgenden Kapitel dieser Übersichtsarbeit soll die angewandte Methodik aufgeschlüsselt werden. Dieses systematische Review wurde anhand der PRISMA Guidelines für systematische Reviews durchgeführt (Moher et al., 2011).

2.1 Auswahlkriterien

Da die Forschungsfragen dieser Arbeit kein klinisches Thema behandeln, werden die Auswahlkriterien nicht anhand des PICO-Framework definiert. Kein herkömmliches Framework wurde als passend befunden und daher wurde das PEO-Framework auf das vorliegende Thema angepasst. PEO gliedert sich in Population, Exposure und Outcome. Nach Munn et al. (2018) wird dieses Framework für Reviews verwendet, die sich mit qualitativer Forschung im klinischen Setting auseinandersetzen. Wie das PEO-Framework auf diese Arbeit umgelegt wird, wird im folgenden Abschnitt dargelegt.

2.1.1 Population

Die Population im PEO-Framework beschäftigt sich normalerweise mit jenen Personen, die in die Studie eingeschlossen werden sollen. Umgelegt auf diese Arbeit, schließt die Population alle Personen des Sports, wie Sportler:innen oder Mannschaften, die durch ein Rankingsystem bewertet werden oder durch das Matchmaking in Teams eingeteilt werden, ein. Darunter fallen auch Personen, die im E-Sport Bereich tätig sind.

2.1.2 Exposure

Unter Exposure verstehen Munn et al.(2018) etwa Supplements oder Medikamente, denen die Population ausgesetzt ist. Dieser Punkt des PEO-Frameworks wird umgelegt auf die Themen, die für diese Übersichtsarbeit von Relevanz sind. In diesem Kontext bezieht sich dies darauf, dass die Personen des Sports durch ein Ranking- oder Matchmaking-System bewertet oder eingeordnet werden müssen.

2.1.3 Outcomes

Die Ergebnisse sollen zeigen, wie sich Ranking und Matchmaking-System in den verschiedenen Sportarten zusammensetzen. Weiters sind die mathematischen Hintergründe der einzelnen Systeme von Relevanz.

2.2 Ein- und Ausschlusskriterien

Publikationen, die sich etwa mit Return-to-Sport nach einem Kreuzbandriss beschäftigen, können schon vor dem Literatur-Screening größtenteils ausgeschlossen werden. Aufgrund des im Search-String angewandten Suchwortes „Sports“ sind in den Ergebnissen der Datenbanken sehr viele Studien und Quellen zu finden, die sich mit dem, in der Forschung

sehr großen Bereich der Sportmedizin auseinandersetzen. Dies inkludiert auch die Bewertung oder Einordnung von Sportverletzungen (z.B.: Sports Ankle Rating System). Diese Art von Bewertungssystemen liegen nicht im Fokus dieser Übersichtsarbeit. Ebenso werden Studien exkludiert, die sich mit psychologischen oder trainingswissenschaftlichen (z.B Trainings Load) Inhalten befassen. Weiters ist die Thematik des „Rate of Perceived Exertion“, also das subjektive Bewerten der Belastung, ein Ausschlusskriterium. Ebenso werden Studien, die sich mit Ernährung befassen, nicht inkludiert.

Außerdem wurden Artikel ausgeschlossen, die sich ausschließlich damit beschäftigen, ob anhand der Bewertung oder dem Ranking des Spielers, der Spielerin oder Teams eine Vorhersage über die Leistung getroffen werden kann. Dies betrifft etwa das Markov Chain Modell, mit welchem etwa im Tennis die Wahrscheinlichkeit des Punktegewinn vorhergesagt werden kann (Newton & Aslan, 2006). Weiters fallen auch Studien, die sich mit der Überprüfung der Validation von solchen Bewertungssystemen auseinandersetzen, unter die Ausschlusskriterien. Es gilt jedoch zu erwähnen, dass der Begriff der Vorhersage und Wahrscheinlichkeiten in diesem Themengebiet nie komplett ausgeschlossen werden kann und demnach auch in den eingeschlossenen Studien immer wieder vorkommen wird. Wenn sich eine Studie nur mit den Interpretations- oder Vorhersagemöglichkeiten und nicht mit der Zusammensetzung eines Systems beschäftigt, ist dies ein Ausschlusskriterium.

Eingeschlossen wurden nur Studien in englischer Sprache, die mit spätestens 31.12.2024 veröffentlicht wurden. Arbeiten in deutscher Sprache wurden exkludiert, weil der Search-String in Englisch verfasst worden ist und somit manche relevante Arbeiten in Deutsch dadurch nicht erfasst werden würden. In Betrachtung werden bei einem Titel- und Abstract-Screening sowie in weiterer Folge im Full-Text-Screening nur jene Referenzen, die sich mit einem Ranking- oder Matchmaking-System in einer Team- oder Einzelsportart beschäftigen. Es sollen Ranking- und Matchmaking-Systeme gefunden und untersucht werden, die sich mit der Bewertung von Spielern und Spielerinnen oder Mannschaften basierend auf deren Leistungen, befassen.

2.3 Informationsquellen

Im Rahmen dieser Übersichtsarbeit wurde von Anfang Oktober bis Ende Dezember 2024 eine Literaturrecherche durchgeführt. Um die, für dieses Arbeit relevante Literatur zu erhalten, wurden die Datenbanken SPORTDiscus, SURF und Web of Science zur Literatursuche herangezogen.

2.4 Studiendesign

In der Datenbank SPORTDiscus wurden für die Literatursuche folgende Filter verwendet: Berücksichtigt wurden nur Referenzen, die in englischer Sprache verfasst wurden und als

„Academic Journal“ aufgelistet sind. In allen drei verwendeten Datenbanken wurde nur Literatur in englischer Sprache berücksichtigt. Weiters wurden alle Publikationen durch einen Filter exkludiert, die sich mit Inhalten aus Bereichen der Biomechanik, Trainingswissenschaft, Ernährung oder Sportmedizin auseinandersetzen.

In der Datenbank Web of Science wurde beim Filter „Web of Science Categories“ nur die Kategorie „Sport Science“ ausgewählt und nur jene Art von Dokumenten, die als „Article“ oder „Review Article“ gekennzeichnet sind. Auch in dieser Datenbank wurden durch die Anwendung von Filtern die Publikationen auf die gewünschten Bereiche reduziert. Inkludiert wurden Publikationen von „journal of sport sciences“, „science sports“, „sports“, „European journal of sport science“ und „international journal of performance analysis in sport“.

In der dritten Datenbank SURF wurden nur jene Dokumente inkludiert, die als Zeitschriftenartikel angegeben sind. Weil ein Filtern nach Publikationen, wie in SPORTDiscus oder Web of Science, nicht möglich ist, wurde in diesem Fall nach fünf ausgewählten Schlagwörtern gefiltert, um den Suchbereich einzuschränken. Die ausgewählten Schlagwörter waren: Sportwissenschaft, Bewertung, Ratingskala, Bewertungskriterium und Sportspiel.

2.5 Suchstrategie

In dieser Übersichtsarbeit sollen Ranking- und Matchmaking-Systeme dargestellt, verglichen und kategorisiert werden. Der angewandte Search-String wurde durch die Boolean-Operatoren „AND“ und „OR“ verbunden und in Figure 1 zur Übersicht dargestellt. Der Suchbegriff „Rating“ wurde in den Search-String aufgenommen, um auch individuelle Ranglisten wie die Plus-Minus-Ratings, bei der es zur Bewertung einzelner Athleten und Athletinnen kommt, in die Resultate einzuschließen.

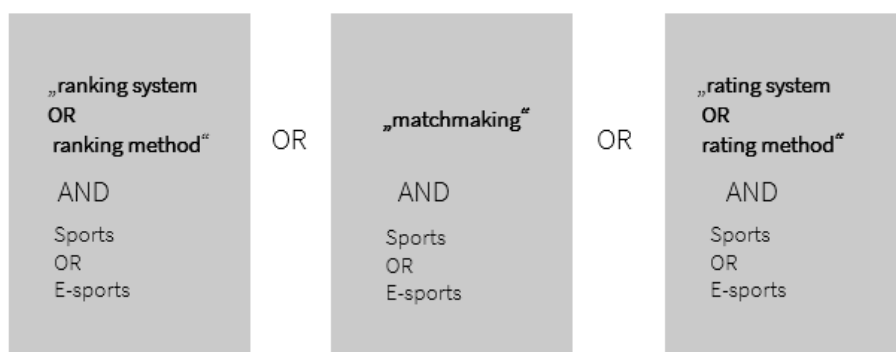


Figure 1 Der angewandte Search-String (Eigene Darstellung)

2.6 Auswahlverfahren

Die ausgewählten Artikel wurden von jeder der drei verwendeten Datenbanken heruntergeladen und in Zotero eingefügt und in weiterer Folge in Rayyan.ai⁴ für das Literatur-Screening importiert. Insgesamt wurden 1636 Referenzen in Rayyan.ai importiert. Davon wurden 803 Studien von der Datenbank „Web of Science“ bezogen, 760 von „SURF“ und 73 Referenzen wurden mittels „SPORTDiscus“ identifiziert. 66 Studien wurden als Duplikat markiert und aussortiert. Nach dem ersten Titel- und Abstract-Screening wurden 1521 Studien exkludiert. 49 Studien wurden im folgenden Full-Text-Screening weiter auf Brauchbarkeit überprüft. Ausgehend von den verwendeten Datenbanken wurde ein Backwards-Search durchgeführt, um weitere relevante Studien zu identifizieren. Dadurch wurde noch eine weitere Studie in die Ergebnisse inkludiert.

Im Laufe des Full-Text-Screenings wurden mittels Rayyan.ai die Studien, die zuvor inkludiert wurden, auf ihre Eignung überprüft. Der Autor dieses Reviews hat das Screening durchgeführt und am Ende 29 Studien für brauchbar befunden. Es werden folglich diese Referenzen für die Beantwortung der Forschungsfragen herangezogen.

2.7 Datenerfassung

Die inkludierten Studien wurden mittels der oben erwähnten Datenbanken erfasst und mittels rayyan.ai identifiziert. Im weiteren Verlauf wurden die Daten aus den Studien mit einer Microsoft EXCEL Tabelle aufbereitet. Jene werden in den Resultaten kategorisiert und zur Übersicht aufbereitet.

2.8 Datenelemente

Die verschiedenen Elemente, die für die Datenerfassung erhoben wurden, waren: Autoren und Autorinnen, Jahr der Veröffentlichung, Population, Exposure und Outcomes Charakteristiken.

2.9 Risk of Bias

Da es sich bei der vorliegenden Arbeit um keine Übersichtsarbeit mit einer klinischen Fragestellung handelt und folglich keine Leitlinienerstellung stattfindet, wird in diesem Rahmen kein Risk of Bias Tool angewandt (Buchberger et al., 2014).

⁴ Rayyan.ai ist eine KI-unterstützte Website, die für das Management des Literaturscreening verwendet wurde. Die KI-Funktion wurde für das Identifizieren von Duplikaten verwendet, jedoch nicht für das Abstract- und Full-Text Screening.

3 Ergebnisse

3.1 Ausgewählte Studien

Anhand eines Prisma Flow Diagramms werden die Ergebnisse des Literatur-Screenings grafisch dargestellt (Figure 2). Durch den anfänglich festgelegten Search-String wurden 1636 Referenzen in Rayyan.ai importiert. 66 Studien wurden als Duplikat entfernt. Die Charakteristik der 29 inkludierten Studien wird folgend aufgeschlüsselt und kategorisiert.

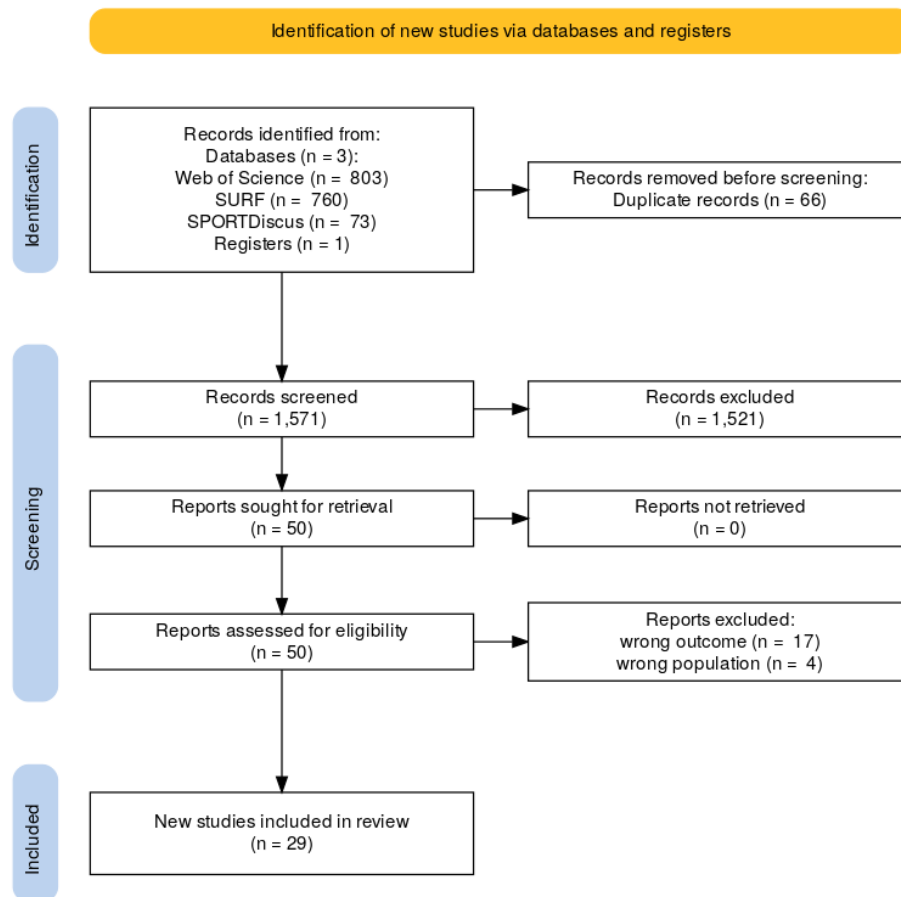


Figure 2 PRISMA FLOW Diagramm zur Übersicht der eingeschlossenen Studien (rayyan.ai)

Williams et al. (2020) behandeln zwar das ATP-Ranking und beschäftigen sich mit ELO basierenden Ratings, jedoch liegt der Fokus auf dem Vorhersagen von Ereignissen und deren Wettquoten. Wunderlich & Memmert (2016) befassen sich mit einem Model, durch welches Spiele bei der Fußballweltmeisterschaft vorhergesagt werden können. Ziv et al. (2010) befassen sich mit leistungsbezogenen Daten, um Teamrankings im Basketball vorhersagen zu können. Somit fallen diese Studien in ein Ausschlusskriterium, weil die Vorhersage der Leistung oder Vorhersagemodelle im Fokus stehen.

Ebenso wurden Studien wie jene von McIntosh et al. (2018) exkludiert, die sich zwar mit Spielerratings bezogen auf Leistung befassen, allerdings liegt der Fokus auf Überprüfung der Validation solcher Bewertungssysteme. Die Studie von Reid et al. (2015) befasst sich mit den Entwicklungsverläufen von professionellen männlichen Tennisprofis in den Top 250 des ATP-Rankings. Da sich die Studie nicht um die Zusammensetzung eines Rankings für Tennisspieler dreht, wurde diese Studie von den Resultaten ausgeschlossen. Bradley und Kollegen (2015) behandeln in ihrer Studie das Ligaranking in der englischen Premier League, doch der Fokus liegt auf der Evaluierung von Match Performance bezogen auf physische und technische Leistungsfaktoren. Dadurch wurde auch diese Studie exkludiert.

3.2 Charakteristik der Studien

Die 29 eingeschlossenen Studien liegen in den Einschlusskriterien und werden für die Beantwortung der Forschungsfragen herangezogen. Im folgenden Abschnitt sollen sie kategorisiert werden und mittels Übersichtstabellen dargestellt werden.

Insgesamt werden elf verschiedene Sportarten behandelt. Zur besseren Übersicht wurden in dieser Arbeit auch die Olympischen Spiele sowie der E-Sport als Sportart klassifiziert. Wenn keine eindeutige Sportart ausgemacht werden konnte, wurde dies mit „All Sports“ markiert. Die am häufigsten vorkommende Sportart ist Basketball, mit fünf Erwähnungen (17% aller Studien), gefolgt von E-Sport und American College Football mit vier (respektive 14%), sowie Fußball drei Erwähnungen (10%). Die Sportarten Schach und Tennis werden je zweimal behandelt (respektive 7%). Je einmal finden Gaelic Football, Biathlon, Golf, Taekwondo und die Olympischen Spiele Erwähnung (respektive 3%). Vier Studien wurden mit der Markierung „All Sports“ versehen (14%).

Ebenso wurden die Studien demnach klassifiziert, ob sich die diskutierten Ranking- und/oder Matchmaking-Systeme an ein Team oder individuelle Sportler:innen richtet. Wenn dieser Sachverhalt nicht eindeutig aus der Studie herausgeht, wurde jene mit „both“ markiert. Unter die Markierung „team“ fallen folglich 17 der 29 Studien (59%), acht Referenzen fallen unter die Kategorie „individual“ (27%). Mit „both“ wurden vier Quellen markiert (14%).

Um die Begriffe des Rankings und des Matchmakings zu trennen, soll auch eine Übersicht gegeben werden, welche Studien sich mit welcher Thematik befassen, oder ob es eine Überschneidung gibt. Knapp über 72% der Studien setzen sich mit einem Ranking- oder Ratingsystem auseinander (21 von 29 Studien). Vier Referenzen setzen sich isoliert mit dem Matchmaking auseinander (14%). Besonders häufig ist dies in Kombination mit dem E-Sport untersucht worden. In vier Studien ist eine klare Trennung der beiden Thematiken nicht klar auszumachen (14%).

Zu erwähnen gilt es weiters das Elo-Ratingsystem (Elo A., 1978). Das Elo-Ratingsystem wird in 13 von 29 Studien von den Autoren und Autorinnen behandelt (45%). Auch andere Bewertungssysteme wie das Bradley-Terry-Model oder das Glicko-System werden in den Resultaten erwähnt (Glickman, 1999).

3.3 Ergebnisse im Detail

Im folgenden Kapitel sollen die Ergebnisse dieser Übersichtsarbeit kategorisiert und dementsprechend aufgelistet werden. Die Resultate werden nach einer Beschreibung der einzelnen Studien auch in einer Übersichtstabelle dargestellt. Die Ergebnisse wurden in folgende vier Kategorien eingeteilt.

- I) Individual-Player-Performance-Ranking/Rating
- II) Team-Performance-Ranking/Rating
- III) Matchmaking-Systeme
- IV) Kombination aus Ranking- und Matchmaking-Systemen

Die Resultate gliedern sich dementsprechend in Studien die sich einerseits mit der Bewertung von Individualsportlern und Individualsportlerinnen (I) oder von Teams (II) beschäftigen sowie andererseits sich mit Matchmaking-Systemen in einer Einzel- oder Teamsportart auseinandersetzen (III). Die vierte Kategorie soll einen Überblick über Studien geben, in welchen sich die Begriffe Ranking und Matchmaking nicht klar voneinander trennen lassen (IV). Die größte Kategorie stellt hierbei die Kategorie „Team-Performance-Ranking/Rating“ dar. Dieser wird 13 Studien zugeordnet.

In den folgenden Abschnitten werden die Ergebnisse mittels Tabellen übersichtlich dargestellt. Eine genauere Beschreibung der einzelnen Studien kann im Anhang nachgelesen werden.

3.3.1 Individual Player Performance Ranking/Rating

In dieser Kategorie versuchen die Autoren und Autorinnen mit den vorgestellten Systemen die Leistung des einzelnen Sportlers oder Sportlerin darzustellen. Dabei haben oft die Bewertungen einzelner Athleten oder Athletinnen einen Einfluss auf die Leistungsbewertung eines gesamten Teams und umgekehrt (Ambrutis & Povilaitis, 2024; Wolf et al., 2021). Dabei wird einerseits von Bottom-up-Ratings gesprochen, wobei beispielsweise Fußballerspieler:innen anhand ihrer Aktionen auf dem Spielfeld bewertet werden, was wiederum einen Einfluss auf die Mannschaft haben kann (Hvattum & Gelade, 2021). Andererseits sprechen Hvattum & Gelade (2021) von Top-down-Ratings, wodurch Kennzahlen der Mannschaftsleistung auf einzelne Spieler:innen verteilt werden. Darunter fällt das sogenannte Plus-Minus-Rating. Unter die Top-down-Ratings fällt auch das Players-impact-Ranking bei Wolf et al. (2021). Figure 3 zeigt ein Ranking nach dem Players-impact anhand von männlichen Fußballspielern aus dem Jahr 2018.

Rank	Name	Current Club	Impact	Rating	Games	Club Rating
1.	MOHAMED SALAH	FC Liverpool	8.43	4858	156	4479
2.	KEVIN DE BRUYNE	Manchester City	6.98	4900	169	4840
3.	NICOLAS OTAMENDI	Manchester City	6.45	4897	155	4840
4.	RAHEEM STERLING	Manchester City	4.97	4879	165	4840
5.	SADIO MANE	FC Liverpool	4.75	4830	138	4779
6.	LIONEL MESSI	FC Barcelona	4.49	4960	180	4923
7.	GERARD PIQUÉ	FC Barcelona	4.23	4974	147	4923
8.	EDERSON	Manchester City	4.21	4854	84	4840
9.	DAVID SILVA	Manchester City	4.15	4896	146	4840
10.	EDINSON CAVANI	Paris St. Germain	4.13	4814	172	4772
11.	MARCO VERRATTI	Paris St. Germain	4.06	4842	127	4772
12.	ERIC BAILLY	Manchester United	4.04	4832	101	4751
13.	PAUL POGBA	Manchester United	3.99	4831	156	4751
14.	ANTOINE GRIEZMANN	Atletico Madrid	3.95	4849	191	4818
15.	DAVID ALBA	Bayern München	3.74	4959	136	4876

Figure 3 Ranking von elitären Fußballspielern anhand des Players-impact (Eigene Darstellung, adaptiert nach Wolf et al., 2021)

Bei Systemen wie jenes von Broadie & Rendleman (2013) steht einzig und allein die tatsächliche Leistung des Sportlers oder der Sportlerin als Bewertungsgrundlage parat. Die Autoren stellen das Official World Golf-Ranking vor, dass die Leistung von Golfprofis anhand von Ergebnissen der letzten zwei Jahre darstellt.

Die gesamten Ergebnisse, die in diese Kategorie eingeordnet werden, sind in der folgenden Übersichtstabelle dargestellt (Tabelle 1).

Tabelle 1 Ergebnisse: Individual-Player-Rating/Ranking (Eigene Darstellung)

Autor	Sport	System	Input	Output
Ambrutis & Povilaitis, 2024	Basketball	Composite-rating-method (CRM)	Match Aktionen einzelner Spieler	Bewertung von Einfluss des Spielers und Teamdynamik
Blackburn, 2013	Tennis	Performance Ranking für weibliche Tennisprofis	Anzahl Siege/gesamten Spiele; Stärke Gegnerin, letzte Spielergebnisse	Ranking nach Winning-Percentage Ranking nach Paarvergleichen
Broadie & Rendleman, 2013	Golf	Official World Golf-Ranking	Turniererfolge Letzten 104 Wochen	Gesamtpunkteanzahl; Jüngsten 13 Wochen werden am höchsten gewertet
Hvattum & Gelade, 2021	Fußball	Top-down & Bottom-up-Rankings	Match Aktionen einzelner Spieler	Plus-Minus-Rating „Value Added by Each Player“-Ratings
Irons et al., 2014	Tennis	Tennis Ranking-System	Punktezahl pro Spiel Ergebnisse werden gewichtet	System basierend auf Wahrscheinlichkeitsfunktion
Kim & Jeon, 2019	Taekwondo	PageRank-Algorithmus	Ergebnisse von Kämpfen zwischen Athleten Vergleich zwischen Leistungen	PageRank für jeden Spieler
Servien, 2022	Biathlon	Biathlon Verfolger-Rankings	Endplatzierungen aus früheren Rennen; Quantil-basierend	Minimierung der Sprintergebnisse;

Wolf et al., 2020	Football	Player-Impact Ranking	Match Aktionen einzelner Spieler	Bewertung Spielstärke und Einfluss der Spieler
--------------------------	----------	-----------------------	----------------------------------	--

3.3.2 Team-Performance-Ranking/Rating

Diese Kategorie fasst jene Studien zusammen, die sich mit der Bewertung und Einordnung von Teams und Mannschaften auseinandersetzen. Dabei stehen als Bewertungsgrundlage besonders Sieg und Niederlage im Fokus. Barrow et al. (2014) erwähnen dabei die Methode nach Winning-Percentage (Gewinnquote), die besonders in populären US-Sportarten, wie NBA oder NHL, häufig angewandt wird. Mit der Colley-Matrix-Methode bei Colley (2002) und Chartier et al. (2011) wird aufbauend auf dieser Methode auch noch die Stärke des Spielplans implementiert, um die Leistungen einer Mannschaft möglichst präzise darzustellen.

Duncan & Parece (2024) versuchen mit ihrem Population-adjusted-national-Ranking die olympischen Medaillenrankings trotz unterschiedlicher Bevölkerungsgröße fairer und vergleichbarer zu gestalten.

Einen anderen Ansatz als die Studien zuvor verfolgen Govan et al. (2009). Sie versuchen Mannschaften im Leistungssport anhand ihrer offensiven und defensiven Stärken und Charakteristiken zu bewerten und einzuordnen.

Tabelle 2 Ergebnisse: Team-Performance-Rankings (Eigene Darstellung)

Autor	Sport	System	Input	Output
Barrow et al., 2013	All Sports	NBA, MLB, NCAA Basketball Rankingsystem	Anzahl Siege/Anzahl Spiele; Gewinnquoten des Gegners	Ranking nach Winning-Percentage Rating-Percentage-Index
Chartier et al., 2011	Basketball	Colley- und Massey-Methode	Anzahl Siege/Anzahl Spiele; Punktedifferenzen Gewichtung von Spielen	Adaptiertes Ranking nach Winning-Percentage Punktedifferenz wird bei Massey widerspiegelt

Colley, 2002	American College Football	Colley-Matrix-Methode	Siege und Niederlagen Stärke des Spielplans	Bewertung anhand Siegen und Niederlagen Bewertung der Gegner wird miteinbezogen
Constantinou & Fenton ,2013	Fußball	Pi-Rating-System	Siege, Niederlagen, Tordifferenzen, Heim- oder Auswärtsspiel	Bewertung der Fähigkeiten von Fußballteams Dynamische Ratings
Duncan & Parece, 2024	Olympia	Population-adjusted-national-Ranking	Medaillenerfolge, Einwohneranzahl	Wahrscheinlichkeit eine Medaille zu gewinnen im Verhältnis zur Einwohnerzahl
Gill & Keating, 2009	American College Football	„weighted/least-square-Methode Zensurmethode	Match Outcome: Sieg und Niederlagen Heim- oder Auswärtsspiel	Wichtigkeit bestimmter Spiele basierend auf z.B.: Winning-Percentage Relative Stärke von Teams
Govan et al., 2009	All Sports	Offense-Defense-Modell	Tore, Punkte, Gegentore Match Aktionen	Bewertung der offensiven und Defensiven Fähigkeiten eines Teams
Jeremic & Radojicic, 2010	Schach	I-Distanzmethode	Sieg, Unentschieden, Niederlage, Matchpoints	Bewertung der Teamleistung anhand diverser Variablen
Lowery et al., 2019	Basketball	Functional-Ratings	Sieg, Niederlage, Punkteunterschiede während dem Spiel	Umwandlung funktionaler Bewertungen in Skalarwerte Stärke- und Schwächephasen

				von Teams werden bewertet
Mangan & Collins, 2016	Gaelic-Football	Modifiziertes Elo-Ratingsystem	Siege, Niederlagen, Gewinnspanne	Dynamische Bewertung von Teams Heimvorteil spielt eine Rolle
Pradhan & Abdourazakou, 2020	E-Sports	Multi-Criteria-Decision-making-Modell	Prozentsatz gewonnener Spiele Anzahl Spiele Anzahl Kills (Treffer)	Bewertung der Stärke eines Teams im Vergleich mit anderen Mannschaften Aktueller Leistungstrend
Steward, 2013	All Sports	Colley-Matrix-Ratingssystem	Siege- und Niederlagen Anzahl gespielter Spiele Stärke des Gegners	Bewertung von Teams Siege gegen ein stärkeres Team werden höher bewertet
Wigness et al., 2010	American College Football	Iterative Rankingmethode	Siege und Niederlagen, direktes Duell, Austragungsort, Gewinnspanne	Bewertung von Teams durch ein Rankingsystem, dass durch iterative Prozesse gesteuert wird

3.3.3 Matchmaking-Systeme

Unter diese Kategorie fallen vier Studien, wobei besonders der E-Sport-Bereich häufig Erwähnung findet. Im E-Sport liegt die Zielsetzung der Systeme darin, die Stärken und Fähigkeiten der Spieler:innen zu verstehen und zu bewerten um diese dann in passende Teams einteilen zu können (Gong et al., 2020; Myślak & Deja, 2015; Wang et al., 2024). Mithilfe von Datenanalysen und maschinellen Lernens sollen Spieler- und Teamdynamiken

besser verstanden werden und so auch passende Teams gefunden werden (Gong et al., 2020). Wang et al. (2024) versuchen mit ihrem System auch das Spielerengagement zu verstehen und zu verbessern.

Sauer et al. (2024) haben als Ziel das Paarungssystem im Schach zu verbessern. Auf der Grundlage von Elo-Ratings, wird der Begriff des Maximum-Weight-Matching (MWM) vorgestellt.

Der Faktor der Fairness nimmt sowohl bei Sauer et al. (2024), als auch bei den anderen Studien in dieser Kategorie einen wichtigen Platz ein (Gong et al., 2020; Myślak & Deja, 2015; Wang et al., 2024).

Tabelle 3 Ergebnisse: Matchmaking-Systeme (Eigene Darstellung)

Author	Sport	System	Input	Output
Gong et al., 2020	E-Sport	OptMatch-System	Frühere Spieldaten, Interaktion zwischen Spielcharakteren Stärke und Schwächeprofile von Charakteren	Bildung von ausgewogenen Teams Spieler:innen werden anhand ihrer Fähigkeiten eingeteilt
Myślak & Deja, 2015	E-Sport	Matchmaking-System für Onlinespiele	Fähigkeiten von Spieler:innen werden als Vektor dargestellt	Einteilung von Personen in Teams anhand ihrer präferierten Spielerrolle Homogene Teams
Sauer et al., 2024	Schach	Matchmaking-System für Schachspieler:innen	Farbbilanzen Punktegruppen	Maximum-Weight-Matching Paarungen werden gewichtet, desto höher die Gewichtung, desto besser die Paarung
Wang et al., 2024	E-Sport	EnMatch-System	Menge an Spieler:innen Fähigkeiten der Spieler:innen Interaktionen untereinander	Durch die Zusammensetzung von fünf Komponenten soll die optimale Teamzusammenstellung gefunden werden

3.3.4 Kombination aus Ranking- und Matchmaking-Systemen

In dieser Kategorie lassen sich die Begriffe Ranking und Matchmaking nicht klar voneinander trennen. Die in den Studien angewandten Systeme richten sich sowohl an das Bewerten und Einordnen von Mannschaften sowie Spielern und Spielerinnen, als auch an die Bestimmung von Setzlisten für zum Beispiel wichtige Turniere im NCAA-Basketball (Matthews et al., 2021).

Mit dem PageRank bei Matthews et al. (2021) werden Teams im amerikanischen College-Basketball eingeordnet und die Qualität der Siege bewertet. Anhand des PageRank-Algorithmus soll in weiterer Folge bestimmt werden, welche Teams an begehrten College-Turnieren teilnehmen dürfen.

Auch Miles et al. (2010) sowie Reinig & Horwitz (2019) nehmen den amerikanischen College-Sport zur Hand. Reinig & Horowitz (2019) stellen das Vier-Quadranten-System vor. Basketball Teams werden zuvor anhand des Ratings-Percentage-Index (RPI) bewertet und dann in das Vier-Quadranten-System eingeteilt. Dadurch soll der Auswahlprozess für das populäre NCAA-Turnier verbessert werden.

Raack et al. (2014) zeigen auf, wie Teams in der Fußballbundesliga, der NHL oder Formel 1 bewertet werden. Die Autoren berechnen in der Folge mit ihrem System, wie viele Punkte ein Team in der NHL benötigen wird, um einen Platz in den Play-Offs zu bekommen.

Tabelle 4 Ergebnisse: Kombination aus Ranking- und Matchmaking-Systemen (Eigene Darstellung)

Author	Sport	System	Input	Output
Matthews et al., 2021	Basketball	PageRank-Algorithmus	Siege und Niederlagen Gewinnspanne Heim- oder Auswärtsspiel	Bewertung der relativen Stärke von Teams Darauf basierend können Teams für wichtige Turniere ausgewählt werden (Matchmaking)

Miles et al., 2010	American College Football	MinV-Modell AccuV-Modell	Siege und Niederlagen Gewinnspanne Heim- oder Auswärtsspiel Winning-Percentage	Bewertung anhand von Integer-Programming Game-Score-Violations sollen minimiert werden Auswahl der besten Teams für wichtige Turniere
Raack et al., 2014	All Sports	Fußballbundesliga, Formel 1 und NHL-Rankings	Siege, Unentschieden und Niederlagen	Teams bekommen ordinalen Rang zugeordnet Durch Integer-Programming wird die Mindestpunktzahl berechnet, die sehr wahrscheinlich benötigt wird, um den Abstieg zu vermeiden.
Reinig & Horowitz, 2019	Basketball	Vier-Quadranten-System	Siege und Niederlagen Winning-Percentage Gewinnquoten der Gegner	Bewertung anhand von Teams anhand des Ratings-Percentage-Index (RPI) Anhand dessen soll bestimmt werden, wer bei wichtigen Turnieren teilnehmen darf

4 Diskussion

4.1 Mathematische Hintergründe und Zusammenhänge

Anhand von Tabelle 5 soll jede Studie und die mathematische Grundlage der verwendeten Systeme übersichtlich dargestellt werden. Anhand der mathematischen Hintergründe lassen sich weitere Unterschiede als auch Gemeinsamkeiten zwischen den Ranking- und Matchmaking-Systemen erkennen.

In den folgenden Abschnitten werden die wichtigsten mathematischen Hintergründe diskutiert.

Tabelle 5 Übersichtstabelle über die mathematische Basis (Eigene Darstellung)

Autor	Basis
Ambrutis & Povilaitis, 2024	Elo
Barrow et al., 2014	Winning-Percentage, Rating-Percentage-Index
Blackburn, 2013	Winning-Percentage
Broadie & Rendleman, 2013	Punktevergabe nach Turnierergebnissen
Chartier et al., 2011	Colley- und Massey-Methode
Colley, 2002	Colley-Matrix-Methode
Constantinou & Fenton, 2013	Pi-Rating
Duncan & Parece, 2024	Wahrscheinlichkeitsmodell
Gill & Keating, 2009	Least-square-Methode
Gong et al., 2020	Elo
Govan et al., 2009	Iterative Prozesse
Hvattum & Gelade, 2021	Elo, Bottom-up und Top-down-Ratings
Irons et al., 2014	Wahrscheinlichkeitsmodell
Jeremic & Radojicic, 2010	Elo, I-Distanz-Methode
Kim & Jeon, 2019	PageRank
Lowery et al., 2019	Least-square-Methode
Mangan & Collins, 2016	Elo
Matthews et al., 2021	PageRank
Miles et al., 2010	Integer-Programming
Myślak & Deja, 2015	Elo, Vektorrechnung

Pradhan & Abdourazakou, 2020	Elo, Rank-Aggregation, iterative Prozesse
Raack et al., 2015	Integer-Programming
Reinig & Horowitz, 2019	Winning-Percentage, RPI
Sauer et al., 2024	Elo
Servien, R., 2022	Quantil-basierte Berechnungen
Steward M., 2013	Colley-Methode
Wang et al., 2024	Elo, Heuristische Methoden
Wigness et al., 2010	Iterative Prozesse
Wolf et al., 2020	Elo

4.1.1 Colley- und Massey Methode

Das Ranking nach Winning-Percentage ist eine der ältesten und einfachsten Methoden, um Mannschaften zu ranken. Dabei wird die Anzahl der gewonnenen Spiele w_i durch die Gesamtanzahl der gespielten Spiele t_i dividiert. In den Ergebnissen bildet diese Methode die Basis der Rankingsysteme bei Barrow et al. (2014), Blackburn (2013) und bei Reinig & Horowitz (2019). Diese Methode wird aufgrund ihrer Einfachheit weltweit verwendet, jedoch wird sie auch kritisiert (Chartier et al., 2011). Denn sie berücksichtigt weder die Stärke des Gegners noch die Stärke des Spielplans, wodurch die Rankings nicht immer die tatsächliche Stärke von Teams widerspiegeln.

Die Colley-Methode richtet sich besonders an den amerikanischen College-Sport. Im Vergleich zu Profisportligen, wie der NBA, NFL oder NHL, spielt nicht jedes Team mindestens einmal gegen denselben Gegner. Wenn demnach, wie im College Football, eine große Anzahl an Teams nur eine kleine Anzahl an Spielen bestreitet, ergibt sich für jede Mannschaft ein unterschiedlicher Spielplan mit unterschiedlicher Stärke (Colley, 2002).

Um dem Problem zu entgegenzuwirken, gibt es die Colley-Methode nach Dr. Wesley Colley. Sie stellt eine Erweiterung der Winning-Percentage Methode dar (Colley, 2002).

Durch die Modifikation ergibt sich folgende Formel:

$$r_i = \frac{1 + w_i}{2 + t_i}$$

Colley (2002) begründet seine Modifikationen auf der LaPlace-Regel. Die Regel ist eine mathematische Methode zur Berechnung von Wahrscheinlichkeiten. Es wird davon ausgegangen, dass, wenn ein Ereignis bestimmt oft vorgekommen ist, es wahrscheinlicher

wieder eintreten wird (Colley, 2002). Die initiale Bewertung eines Teams liegt bei dieser Methode nicht bei $\frac{0}{0}$ sondern startet mit $\frac{1}{2}$. Das Modell soll einen Sieg gegen einen stärkeren Gegner höher belohnen als einen Sieg gegen einen schwächeren Konkurrenten. Das bedeutet, dass die Bewertung eines Teams i von den Ratings des Gegners abhängig ist. Wenn ein Team gewinnt, nimmt das Einfluss auf das Rating eines anderen Teams. Der Colley-Algorithmus kann vereinfacht mit der linearen Gleichung $Cr = b$ dargestellt werden (Langville & Meyer, 2012).

Die Stärke des Spielplans wird bei Colley (2002) folglich so quantifiziert, indem die Bewertungen der Gegner in das Ranking einzelner Teams miteinbezogen werden. Dabei wird ein iteratives Verfahren angewandt, um die Bewertung einer Mannschaft auf der Grundlage der Bewertung der Gegner zu verfeinern.

Die Colley-Methode berücksichtigt jedoch keine Punktedifferenzen am Ende eines Spiels, sondern nur Sieg- und Niederlagenstatistiken. Das ist ein entscheidender Unterschied zu der Massey Methode. Stewart (2013) verwendet eine Kombination dieser beiden Methoden im Kontext des amerikanischen Collegesports (Stewart, 2013).

Die Massey-Methode nach Kenneth Massey (1997) wurde erschaffen um Sportteams bewerten und einordnen zu können. Die Methode basiert auf der Regel, dass der Unterschied zwischen zwei Teams, dargestellt durch $r_i - r_j$, den Punkteunterschied in einem direkten Duell repräsentiert. Ähnliche wie bei Colley (2002), wird auch diese Methode vereinfacht mit einer linearen Gleichung $Mr = p$ dargestellt. Dabei leitet sich die Variabel M aus der Colley-Matrix ab, wodurch die Bewertungen des Teams berechnet werden. Die rechte Seite der Gleichung p_i wird als ein Vektor der kumulativen Punktedifferenz bezeichnet. Dadurch steht p_i für die Gesamtanzahl an Punkten, die ein Team gegen alle seine Gegner erzielt hat, abzüglich der gesamten Gegenpunkte- oder treffer, die eine Mannschaft hinnehmen hat müssen (Chartier et al., 2011; Langville & Meyer, 2012).

Team	Record	Massey Rating	Massey Rank	Colley Rank
Davidson	18-2	13.99	1	1
Charleston	15-5	5.42	2	2
Citadel	15-5	5.32	3	3
Wofford	12-8	0.73	5	4
Chattanooga	11-9	-1.30	8	5
W. Carolina	11-9	2.67	4	6
Samford	9-11	0.00	6	7
App. State	9-11	-0.17	7	8
Elon U	7-13	-4.22	9	9
Georgia Southern	5-15	-7.73	11	10
UNC Greensboro	4-16	-6.57	10	11
Furman	4-16	-8.14	12	12

Figure 4 Bewertung von College Basketball Teams nach der Colley und Massey Methode (Chartier et al., 2011)

Die beiden beschriebenen Methoden sind zwar voneinander abhängig, jedoch können sie zu unterschiedlichen Einstufungen führen. Chartier et al. (2011) zeigen anhand von Figure 4, wie männliche NCAA-College-Basketballteams durch die Colley- und Massey-Methoden differenziert in einem Ranking abschneiden.

Wenn es zu einer Kombination zwischen zwei oder mehr Rankingmethoden kommt, wird von Rank Aggregation gesprochen. Dieser Begriff definiert ein zusammenführen von mehreren Rankings zu einer Rankingliste, wie etwa bei Colley und Massey. In dem Falle, dass sich die Rankings der Teams in den unterschiedlichen Methoden nicht decken, gilt es eine Strategie für das Tie-Breaking festzulegen. Ein weiteres Beispiel für die Aggregation von Rankings ist die Darstellung von Teams anhand eines Graphen. Der PageRank-Algorithmus fällt etwa in diese Kategorie (Langville & Meyer, 2012).

4.1.2 Page-Rank Algorithmus und iterative Prozesse

Der PageRank-Algorithmus wurde 1988 von Page and Brian vorgeschlagen und dann von Google verwendet, um Webseiten nach der Zitationshäufigkeit zu bewerten. Desto höher die Zitationshäufigkeit, umso wichtiger wird die Website angesehen (Page et al., 1999).

Der Algorithmus findet folglich auch Anwendung im Sport, unter anderem im Taekwondo (Kim & Jeon, 2019). Der PageRank bietet sich nur als Möglichkeit für Ranking und Matchmaking an, sondern auch für die Vorhersage von Spielerereignissen.

Bei der Bewertung von Webseiten weist der PageRank jedem einzelnen Element einer Gruppe von durch Hyperlinks verknüpften Dokumenten eine numerische Gewichtung zu. Damit soll die relative Bedeutung der Gruppe gemessen werden. Dabei kommt es zu einer Matrixdarstellung des World Wide Webs, wobei jede Website einen Knoten darstellt und die Links zwischen den Websites als Kanten charakterisiert werden. Im Kontext des Sports beschreiben Matthews et al. (2021) die Knotenpunkte für die Teams und die Spiele als Links zwischen den Knotenpunkten. Die Links können dabei unterschiedlich gewichtet werden, wenn wichtigere Spiele hervorgehoben werden sollen. Das wären am Beispiel des Basketballs etwa Spiele, die mit großen Punkteunterschied gewonnen wurden.

Der Algorithmus ist ein iterativer Prozess der durch das Konzept des „Random-Walker“ erklärt werden kann. Die Random-Walker-Methode, die auch bei Barrow et al. (2013) Erwähnung findet, ist ein stochastisches Verfahren, bei der sich ein „Walker“ durch einen mathematischen Raum bewegt. In diesem Falle bewegt sich der „Walker“ durch das Diagramm des PageRank-Algorithmus und beginnt bei einem Knoten (Team) und folgt dann zufällig den Links zu einem anderen Knoten. Wenn ein Team A gegen ein Team B verliert, entstehen neue Links zwischen den Teams. Durch die Bewegung des „Walkers“

entsteht eine Wahrscheinlichkeitsverteilung, durch die die Gesamtstärke einer Mannschaft bewertet werden kann (Matthews et al., 2021).

Bei Page et al. (1999) wird der PageRank-Algorithmus mit folgender Gleichung dargestellt:

$$PR(A) = (1 - d) + d \left[\frac{(PR(T_1))}{C(T_1)} + \dots + \frac{(PR(T_n))}{C(T_n)} \right]$$

Damit könnte der PageRank von Team A ($PR(A)$) berechnet werden. Anhand von T_i wird ein Team dargestellt, gegen das ein Team A gewonnen hat. $C(T_i)$ steht dabei für die Anzahl an Niederlagen von Team T_i und wird im Graphen durch die Links dargestellt, die dieses Team verlassen. Wenn ein Team A ein Team T_i mit einem hohen PageRank besiegen sollte, erhöht dies den PageRank des Team A (Matthews et al., 2021).

Wie erwähnt wird der PageRank-Algorithmus durch einen iterativen Prozess ermöglicht. Die mathematische Begriff der Iteration findet sich bei mehreren Studien und Systemen wieder (Colley, 2002; Matthews et al., 2021; Wigness et al., 2010). Bei Iterativen Prozessen wird eine gleiche oder ähnliche Handlung mehrmals wiederholt, um sich an ein bestimmtes Ziel oder eine Lösung anzunähern (Bosch & Verborgh, 2016). Der „Walker“ beim PageRank-Algorithmus durchläuft immer wieder den Graphen, um eine Wahrscheinlichkeitsverteilung zu generieren und so ein Rating für ein Team zu bekommen (Matthews et al. 2021). Ursprünglich basiert der PageRank-Algorithmus auf Markow-Ketten, die wiederum auch durch iterative Prozesse bestimmt werden können (Devlin et al., 2020).

Durch Markow-Ketten sollen die Wahrscheinlichkeiten für das Eintreten von bestimmten Ereignissen dargestellt werden. Es wird modelliert, wie sich ein System von einem in den anderen Zustand bewegt. Diese sogenannte Übergangswahrscheinlichkeit (Transition probabilities) soll zeigen, wie sich Teams in einer Markow-Kette bewegen. In diesem Kontext zeigt dies die Wahrscheinlichkeit an, dass ein Team i gegen ein Team j gewinnt. Die Übergangswahrscheinlichkeit kann mit folgendem Ausdruck angegeben werden (Devilin et al., 2020):

$$T_{ij} = \begin{cases} \frac{w_{ij}}{N_j}, & i \neq j \\ \frac{w_i}{N_i}, & i = j \end{cases},$$

Dabei ist w_{ij} die Anzahl an Siegen, die ein Team i gegen Team j gefeiert hat, während w_i die Gesamtanzahl an Siegen von Team i angibt. Der Faktor N gibt die Anzahl an Spielen der beiden Teams an. Die Wahrscheinlichkeit, dass es zu einem Übergang von Team j zu Team i kommt, wird folglich dadurch bestimmt, wie oft Team i gegen j gewonnen hat. Wenn etwa ein Team i gegen Team j aus acht Spielen vier gewinnt, liegt die Übergangswahrscheinlichkeit bei $\frac{4}{8}$ (Devlin et al., 2020).

Die Markow-Ketten beim PageRank-Algorithmus modellieren, wie ein:e Benutzer:in bei Websuchen sich zwischen den Seiten bewegt. So kann festgestellt werden, welche Seiten am wichtigsten sind. Dadurch wird es auch möglich, Teams aufgrund von vergangenen Ergebnissen und Leistungen zu bewerten (BASHARIN et al., 2004; Langville & Meyer, 2012). Devlin et al. (2020) stellen die iterative Markov-Methode vor, bei welcher die Ratings basierend auf neuen Daten stetig aktualisiert werden und dadurch einen dynamischen Charakter haben.

Ein weiteres Modell, in welchem iterative Prozesse angewandt werden, ist das Offense-Defense Modell. Govan et al. (2009) nehmen dieses Modell zur Hand, um die Gesamtstärke von Teams bewerten zu können. Es fließen demnach die offensiven und defensiven Fähigkeiten einer Mannschaft in die Bewertung mit ein. Diese werden danach sukzessive angepasst, damit die relativen Stärken der Mannschaften exakter dargestellt werden können. So eine Verfeinerung passiert beispielsweise, wenn ein Team gegen eine starke Defensive viele Punkte erzielt, wird das Offensivrating dieser Mannschaft dementsprechend angepasst. Dies passiert demnach auch umgekehrt, wenn ein Team viele Punkte des Gegners zulässt, dann verschlechtert sich sein Defensivwert. Die Iteration sorgt für eine dynamische Anpassung der Bewertung anhand der Leistung von Teams (Govan et al., 2009).

4.1.3 Matchmaking

Ein effizientes Matchmaking-System versucht möglichst faire und ausgeglichene Teams zu generieren (Alman & McKay, 2017). Im OptMatch-System bei Gong et al. (2020) lässt sich dieser Prozess als Online-Planning verstehen. In der Praxis werden Spieler:innen aus einem Player-Pool $P_t = \{p_1, \dots, p_n\}$ für ein Spiel ausgewählt und in ein Team zugeteilt. Das System hat als großes Ziel die Spielerzufriedenheit nach einem Spiel „„LOL““. Die Zufriedenheit wird als Utility μ_m bezeichnet. Berechnet wird sie mit folgendem Ausdruck:

$$u_m = \sum_{i=0}^{n_m} u_{m,i}$$

Damit wird die gesamte Zufriedenheit der Spieler aus einem Spiel berechnet. Das Matchmaking hat das Ziel die Utility μ_m so groß wie möglich zu machen (Gong et al. 2020; Wang et al., 2024). Messbar gemacht wird dies auch dadurch, ob ein:e Spieler:in direkt nach einem Match das Spiel verlässt. Es scheint, dass Spieler:innen, die mit dem Matchmaking zufrieden sind, wahrscheinlicher im Spiel verbleiben als Spieler:innen, die mit der Teamzusammenstellung nicht zufrieden sind (Wang et al., 2024).

Um die Zufriedenheit der Spieler:innen zu ermöglichen, versuchen Systeme wie jene von Gong et al. (2020), Wang et al. (2024) oder Myślak & Deja (2015) die Stärken der einzelnen Spieler:innen möglichst präzise zu erfassen. Da in „LOL“ vor Spielantritt eine Rolle (wie etwa Verteidiger:in oder Angreifer:in) ausgewählt werden muss, haben Myślak & Deja (2015) das Ziel, das alle spielenden Personen in ihrer präferierten Rolle spielen können, ohne dass andere Spieler:innen mit derselben Auswahl in ihr Team kommen. Um die Stärke eines Spielers oder einer Spieler:in in verschiedenen Rollen zu bemessen, wird der Skill (Fähigkeiten) als Vektor statt als einfache Zahl (Skalar) dargestellt.

Durch den folgenden Ausdruck können die Fähigkeiten in verschiedenen Rollen von einem Spieler i gemessen werden.

$$\text{Skill}_i(\text{games}_i) = \begin{cases} \text{Skill}(\text{games}_{i1}) \\ \dots \\ \text{Skill}(\text{games}_{ij}) \\ \dots \\ \text{Skill}(\text{games}_{im}) \end{cases}$$

Games_i steht dabei für die Spiele, die Spieler i absolviert hat. Der Vektor gibt folglich an, welche Leistung der Spieler i in verschiedenen Rollen gebracht hat, wodurch bewertet werden kann, wie hoch der Skill jener Person als Angreifer:in oder Verteidiger:in ist. Die Variable j kann dabei die Werte 1 bis m annehmen, wobei m für die Anzahl an Rollen steht, die der Spieler i annehmen kann (Myślak & Deja, 2015).



Figure 5 Tier-based-heuristic-Matchmaking (Wang et al., 2024)

Die bei Wang et al. (2024) verwendete Matchmaking-Methode ist das Tier-based-heuristic-Matchmaking (Figure 5). Anhand von heuristischen Methoden wird versucht, innerhalb kürzester Zeit und mit wenig Rechenaufwand, Lösungen für ein bestimmtes Problem zu finden. Die Spieler:innen werden dabei aufgrund von früheren Leistungen und Fähigkeiten in Stufen eingeteilt. So werden Spieler:innen mit ähnlichen Fähigkeiten in gleiche Gruppen eingeteilt. Es werden folglich eine bestimmte Anzahl von Spielenden aus einem Matching-Pool ausgewählt. Beispielsweise können aus sechs Spieler:innen $\binom{6}{3} = 20$ verschiedene Teamkombinationen für ein 3-vs-3 Spiel generiert werden. Auf Basis von Spielergebnisvorhersagen oder heuristischen Herangehensweisen, können faire Teams gebildet werden (Wang et al., 2024).

Wang et al. (2024) nehmen dieses Tier-based-heuristic-Matchmaking als Basis in dem vorgestellten EnMatch-System. Der adaptierte Heuristic-Operator legt zunächst eine Schätzung fest, welche Spieler:innen zusammengebracht werden sollen. Folglich beginnt er durch „Zerstören“ und „Reparieren“ Änderungen daran vorzunehmen. Dabei werden auf der Suche nach der perfekten Lösung immer wieder Spieler:innen entfernt und hinzugefügt. Dies ist wiederum ein iterativer Prozess, der immer wieder passiert und verschiedene Möglichkeiten sucht, Spieler:innen zusammenzubringen.

Diese Vorgangsweise stützt sich auf der Methode der Large Neighborhood-Search (LNS), erwähnt von Wu et al. (2021). Darunter versteht sich eine Verbesserungsheuristik wobei durch das „Zerstören“ und „Reparieren“ eine Lösung iterativ verbessert werden soll. Das LNS findet meist beim Integer-Programming seine Anwendung.

4.2 Vergleich und Analyse der vier Kategorien

In Kapitel 3 wurden vier Kategorien (I-IV) festgelegt, in welche die 29 inkludierten Studien eingeteilt wurden. Diese Kategorien unterscheiden sich zunächst in ihrer Zielsetzung nach Ranking, Matchmaking oder einer Kombination aus beiden Komponenten. Weiters wurde dabei zwischen Individual- oder Teamranking differenziert. Doch es lassen sich noch weitere Unterschiede und Gemeinsamkeiten zwischen den Kategorien erkennen.

Unterscheidungsmöglichkeiten entstehen durch die unterschiedlichen Sportarten, die verschiedene Anforderungen mit sich bringen und somit teils verschieden Ansätze verfolgen. Dies lässt sich anhand dem Vergleich zwischen Individual- und Teamsportarten nachvollziehen. In Folge bieten sich auch weitere Kategorisierungsmöglichkeiten wie der Unterscheidung zwischen Match-Outcome und Match-Aktion sowie den beiden Ansätzen Top-down und Bottom-up.

Der Faktor, der alle Studien miteinander verbindet, ist die Fairness. Es lässt sich schließen, dass der Grundgedanke aller untersuchten Systeme in einer fairen Bewertung und Einordnung liegt.

4.2.1 Zeitliche Verläufe

Mit Blick auf die Zeitpunkte der Veröffentlichungen lässt sich besonders beim Matchmaking ein noch jüngeres Interessensgebiet erkennen. Dies dürfte besonders am E-Sport liegen, der durch das stetige Fortschreiten der technologischen Möglichkeiten und dem gesteigerten Interesse immer mehr in den Fokus rücken könnte.

Die älteste Arbeit in der Kategorie des Matchmakings stammt von Myślak & Deja (2015). Die Studie versucht die Fähigkeiten von Spielern und Spielerinnen als Vektoren darzustellen und in Folge homogene Teams zu bilden. Auch Gong et al. (2020) sowie Wang et al. (2024), die sich auch auf Gong et al. (2020) beziehen, haben später als Ziel, ausgeglichene Teams zu finden. Dabei geht es auch darum, immer schneller und effektiver Lösungen zu finden. In allen drei Studien finden sich die Grundlagen des Elo-Ratingsystems nach Elo (1978) wieder.

Die Elo-Ratings als Grundlage vieler Systeme hat sich deutlich verankert und wird weiterhin häufig verwendet. Sie finden beispielsweise sowohl bei Jeremic & Radojicic (2010) im Schach als auch bei Ambrutis & Povilaitis (2024) eine Erwähnung.

Wenn ein Blick auf die mathematische Basis der Systeme geworfen wird, lässt sich ab dem Jahre 2011 eine Zunahme der Methodenkombinationen erkennen. Es lässt sich im Verlauf immer weniger einer klaren Basis herausstreichen, sondern vielmehr lassen sich Kombinationen aus verschiedenen Methoden feststellen. Beispielsweise zu erkennen bei Chartier et al. (2011), wo die Colley- und Massey-Methoden miteinander kombiniert werden. Auch bei Pradhan & Abdourazakou (2020) werden die Begriffe der Elo-Ratings, der Rank-Aggregation und der iterativen Prozesse kombiniert, um das vorgestellte Rankingsystem aufbauen zu können.

Generell dürfte das Interesse an Ranking und Matchmaking in den letzten 15 Jahren deutlich zugenommen haben. 26 der 29 inkludierten Studien wurden im Zeitraum zwischen 2010 und 2024 veröffentlicht. Ein besonderer steiler Anstieg lässt sich dabei zwischen 2019 und 2020 verzeichnen (Sieben Studien). Ältere Arbeiten, wie jene von Colley (2002), legen mit der Colley-Methode einen Grundstein für spätere Studien. Auch die Least-Square-Methode wird bei Gill & Keating (2009) und zehn Jahre später bei Lowery et al. (2019) erwähnt. Hingegen werden Modelle wie die heuristischen Methoden bei Wang et al. (2024), erst später genannt.

Weiterhin von Bedeutung dürften die Wahrscheinlichkeitsmodelle zunehmen. Modelle dieser Art die unter anderem bei Irons et al. (2014) und Duncan & Parece (2024) Erwähnung finden, werden häufig verwendet. Da Rankings in Kombination mit Prognosemodellen genutzt werden und dies einen großen Platz in der Forschung in diesem Gebiet einnehmen dürfte. Solche Vorhersage- und oder Prognosemodelle wurden im Vorfeld als Ausschlusskriterium definiert, finden sich aber dennoch in vielen Studien wieder.

Auch der PageRank-Algorithmus hat sich aufgrund seiner vielseitigen Einsatzmöglichkeiten sowohl im Sport als auch in anderen Bereichen durchgesetzt (Kim & Jeon, 2019; Matthews et al., 2021).

Der mehrmals zitierte Rating-Percentage-Index, der besonders im Kontext des amerikanischen College-Sports auftritt (Barrow et al., 2013; Reinig & Horowitz, 2019), wurde laut NCAA.com (2025) durch die NET-Rankings ersetzt. Die NET-Rankings (NCAA-Evaluation-Tool) bewerten NCAA-Basketballteams anhand der Spielergebnisse, Spielort, Spielqualität sowie offensive und defensive Effizienzstatistiken. Diese Rankings dienen folglich als Grundlage, um Teams in das Vier-Quadranten-System einzuteilen. Dieses System, beschrieben bei Reinig & Horowitz (2019), findet daher weiterhin seine Anwendung im amerikanischen College-Basketball.

4.2.2 Die verschiedenen Sportarten und ihre Systeme

Insgesamt wurden in den 29 inkludierten Studien elf verschiedene Sportarten behandelt. Dabei findet besonders Basketball besonders häufig Erwähnung (5 Studien). Die Ziele der Systeme im Basketball liegen nicht nur auf der Bewertung und Einordnung von Spielern und Spielerinnen und Teams, sondern auch darin, wie Setzlisten, demnach das Matchmaking, transparenter gestaltet werden können (Ambrutis & Povilaitis, 2024; Chartier et al., 2011; Lowery et al., 2020). Dies wird mit dem PageRank-Algorithmus und dem NCAA-Vier-Quadranten-System versucht (Matthews et al., 2021; Reinig & Horowitz, 2019). Basketball findet sich in drei der vier Kategorien, die in Kapitel 3 festgelegt worden sind, wieder. Es kommt somit zu einer Kombination aus Individual- und Teamrankings mit dem Matchmaking. Das Setting bietet dafür meist der amerikanische College-Basketball.

Eine weitere Sportart aus dem amerikanischen College-Sport ist der American Football. Dieser wird in vier Studien erwähnt. Diese Sportart findet sich in zwei von vier Hauptkategorien wieder. Bei Colley (2002) sollen Teams anhand ihrer Sieg und Niederlagenstatistik bewertet werden. Sieg und Niederlage sind auch bei Wigness et al. (2010) von Relevanz, dabei sollen Teams anhand iterativer Prozesse eingeordnet werden. Gill & Keating (2009) versuchen mit Methoden wie der least-square-Methode die Wichtigkeit bestimmter Spiele zu betonen und dementsprechend zu werten. Gleichzeitig soll auch die

relative Stärke von Teams bemessen werden. Mit dem Min-V-Modell sollen Game-Score-Violations minimiert werden und in Folge die besten Teams für wichtige NCAA-Turniere ausgewählt werden (Miles et al., 2010).

Die Sportart Fußball wird in drei Studien herangezogen. Mit dem Players-impact und dem Plus-Minus-Ranking lassen sich zwei ähnliche Ansätze hervorheben, die als Ziel haben, den Einfluss eines Spielers oder einer Spielerin auf ein Team und das Spielergebnis zu bemessen (Hvattum & Gelade, 2021; Wolf et al., 2021). Constantinou & Fenton (2013) versuchen Fußballteams anhand ihrer Fähigkeiten mit dem Pi-Rating-System zu bewerten.

Ein weiterer großer Bereich im Rahmen der Ergebnisse war der E-Sport. Drei der vier Studien haben die Optimierung des Matchmaking als Ziel. Die Fähigkeiten der einzelnen Spieler:innen, das Spielerengagement sowie unter anderem die Teamdynamik spielen dabei in den Systemen eine Rolle (Gong et al., 2020; Wang et al., 2024). Myślak & Deja (2015) stellen die Fähigkeiten der Spieler:innen als Vektor dar und das System versucht folglich die Teams anhand der präferierten Spielerrollen der Teilnehmer:innen einzuteilen. Das Multi-Criteria-Decision-making-Modell aus der Kategorie II, hat als Ziel den aktuellen Leistungstrend der E-Sportteams zu bewerten (Pradhan & Abdourazakou, 2020).

In der Kategorie des Matchmakings gibt es auch die Sportart Schach zu erwähnen. Dabei versuchen Sauer et al. (2024), Paarungen im Schach zu gewichten und anhand dessen bessere Paarungen zu ermöglichen. Jeremic & Radojicic (2010) haben als Ziel, mit der I-Distanzmethode die Teamleistung im Schachsport anhand verschiedener Variablen zu bewerten.

Die Sportart Tennis findet zwei Erwähnungen. Blackburn (2013) hat als Ziel, weibliche Tennisprofis anhand von der Gewinnquote (Winning-Percentage) und durch Paarvergleiche in einem Ranking zu vergleichen. Einen anderen Ansätzen wählen Irons et al. (2014), die ein Ranking-System vorstellen, in welchem die Punkte pro Spiel miteinbezogen werden sollen. Das System basiert auf einer Wahrscheinlichkeitsfunktion.

Für die Sportart Golf steht beispielsweise das Official World Golf-Ranking, das als Ziel hat, die Turnierfolge von Golfern und Golferinnen der letzten 104 Wochen darzustellen (Broadie & Rendleman, 2013). Im Biathlon beschreibt Servien (2022) das Pursuit-Ranking, mit welchem die Sprintergebnisse männliche Biathleten minimiert werden sollen und folglich ein transparentes Ranking geschaffen werden soll. In der Sportart Taekwondo sollen anhand des PageRank-Algorithmus die Leistungen der männlichen Athleten bewertet werden (Kim & Jeon, 2019). Die zuletzt genannten Sportarten fallen alle unter die Kategorie I der Individualrankings.

Ebenfalls nur einmal Erwähnung findet Gaelic Football. Die Sportart, in der wenige Spiele gespielt werden, wird bei Mangan & Collins (2016) behandelt und hat als Ziel Teams zu bewerten und die besonderen Anforderungen der Sportart berücksichtigen. Das System wird dabei als eine modifizierte Variante des Elo-Ratingsystems ausgelegt.

Um diese Studie besser einordnen zu können, wurden im Rahmen dieser Arbeit auch die olympischen Spiele als Sportart definiert. Dabei wird das Population-adjusted-national-Ranking angewandt mit dem Ziel, die olympischen Medaillenrankings fairer zu gestalten. In diesem Falle wird die Wahrscheinlichkeit eine Medaille zu gewinnen, im Verhältnis zur Einwohnerzahl, dargestellt. Dieses System fällt unter die Kategorie der Team-Rankings (Duncan & Parece, 2024).

4.2.3 Top-down und Bottom-up-Ratings

Mit den Begriffen Top-down und Bottom-up-Rating lassen sich die unterschiedlichen Studien ein weiteres Mal kategorisieren. Die beiden Begriffe finden in den Ergebnissen bei Hvattum & Gelade (2021) eine Erwähnung.

Die Bottom-up-Ratings basieren auf individuellen Spieleraktionen auf dem Spielfeld. Die tatsächliche individuelle Leistung des Sportlers oder Sportlerin steht dabei im Vordergrund, während die Leistung des Teams dabei nur nebensächlich ist (Hvattum, 2019). Im Gegensatz dazu liegt bei den Top-down-Ratings der Fokus auf der Leistung des gesamten Teams. Wenn eine Mannschaft gute Leistungen bringt, wirkt sich das auf alle Spieler:innen aus, die an dem Spiel teilgenommen haben (Hvattum, 2019).

Hvattum & Gelade (2021) stellen als Beispiel für ein Bottom-up-Rating das Value Added by Each Player-Rating vor. Dabei wird jede einzelne Aktion auf dem Spielfeld bewertet. Als Beispiel für ein Top-down-Rating führen Hvattum & Gelade (2021) das Plus-Minus-Rating (PM) an. Bei diesem Konzept wird die Tordifferenz der Mannschaft einzelnen Spielern oder Spielerinnen zugeordnet.

Ambrutis & Povilatis (2024) behandeln mit der Composite-Rating-Methode (CRM) ein Ranking, um Basketballspieler und -spielerinnen sowie Teams zu bewerten. In diesem System lassen sich sowohl Top-down als auch Bottom-up-Ansatz wiederfinden. Die Leistungsbewertungen für Spielerinnen und Teams dürften sich gegenseitig bedingen.

Bei Wolf et al. (2021) kann der Players-impact, für männliche Fußballspieler, dem Top-down-Ansatz zugeordnet werden. Er zeigt, welchen Wert ein gewisser Spieler für eine Mannschaft haben kann. Der Wert berechnet sich durch die Änderungswerte der Teambewertung, je nachdem ob der Spieler auf dem Feld steht, oder nicht.

Es lässt sich daraus schließen, dass mit Hinblick auf die, in den Ergebnissen festgelegten Kategorien, jene Studien die sich mit individual Bewertungen (Kategorie I) auseinandersetzen eher einem Bottom-up-Ansatz verfolgen. Währenddessen können Teambewertungen (Kategorie II) dem Top-down-Ansatz zugeordnet werden.

Beim Matchmaking dürfte wiederum besonders im E-Sport der Bottom-up-Ansatz verbreitet sein (Gong et al., 2020; Myślak & Deja, 2015; Wang et al., 2024). Dabei geht es um die Aktionen und das Verhalten eines Spielers oder einer Spielerin in einem Onlinespiel. Die Stärke und Fähigkeit eines Spielers oder einer Spielerin ist folgend ausschlaggebend dafür, wie das Matchmaking-System die Teams einteilen wird.

4.2.4 Match Outcome und Match Aktion

Mit Hinblick auf den Vergleich der Ranking- und Matchmaking-Systeme der verschiedenen Sportarten, ergibt sich eine weitere Möglichkeit der Kategorisierung. Dabei können die Studien in Match Aktion und Match Outcome eingeteilt werden. Unter Match Aktion fallen alle Studien, deren Systeme auf einzelne Aktionen in einem Spiel basieren, wie einem Tor, Pass oder der Spielzeit eines Spielers oder einer Spielerin. Match Outcome schließt jene Systeme ein, deren Basis durch den Ausgang eines Events gelegt wird. Darunter fallen etwa Sieg- und Niederlagenverhältnis (Leicht et al., 2017).

Teamsportarten aber auch Systeme im Einzelsport, wie die ATP-Liste oder das OWGR, fallen unter die Kategorie Match Outcome. Sie basieren meist auf Siegen, Niederlagen, Punkteunterschiede am Ende eines Spiels oder dem Siege- und Niederlagenverhältnis. Darunter fallen besonders auch Studien, die auf den Elo-Ratings basieren oder etwa mit der Colley Methode arbeiten (Cartier et al., 2011; Mangan & Collins, 2016).

Bei den Rankings für einzelne Spieler:innen nehmen einzelne Match-Aktionen mehr Bedeutung an. So etwa bei den Plus-Minus Ratings bei Hvattum & Gelade (2021) oder das Players-impact Ranking in der Studie von Wolf und Kollegen (2021). Das Ranking der einzelnen Athleten und Athletinnen wird durch mehrere Match-Aktionen auf dem Spielfeld (Tor, Pässe, Spielzeit) bestimmt (Hvattum, 2019; Wolf et al., 2021).

Nicht alle Modelle aus dem Teamsport, wie das Offense-Defense Modell bei Govan et al. (2009), passen in die Kategorie Match Outcome. Dabei werden Angriffs- (erfolgreiche Pässe, Ballbesitzphasen) und Verteidigungsaktionen (Balleroberungen, Geblockte Schüsse) genutzt, um zu bewerten, wie gut eine Mannschaft sowohl offensiv als auch defensiv spielt. Diese Aktionen sind Komponenten, die in weiterer Folge den Ausgang eines Spiels beeinflussen können, aber das Offense-Defense Modell basiert auf Aktionen in einem Spiel und fällt folglich auch unter die Kategorie Match Aktion (Delgado-Bordonau et al., 2013; Govan et al., 2009).

Es lässt sich dennoch daraus schließen, dass die beiden Kategorien nicht immer klar voneinander zu trennen sind. Die einzelnen Aktionen eines einzelnen Spielers oder eine einzelnen Spielerin, wie etwa beim Player-impact-Ranking, beeinflussen im Laufe des Spiels auch den Match Outcome (Wolf et al., 2020).

4.2.5 Individualsport

Wenn die Individualsportarten, wie Tennis, Taekwondo, Golf oder Biathlon, in Betracht gezogen werden, lassen sich durchaus weitere Unterschiede zu den Teamrankings erkennen. Die Athleten und Athletinnen sind dabei vollständig von ihren eigenen Leistungskennzahlen abhängig. Dies könnten die Leistungen bei großen Turnieren und die darauf bedingten Punktezahlen im Tennis oder Golf, oder die Verfolgungsrangings im Biathlon sein. Die individuelle Leistung steht im Vordergrund und wird nicht durch eine Teambewertung verändert (Blackburn, 2013; Broadie & Rendleman, 2013; Servien R., 2022).

Allerdings werden Faktoren, wie die Stärke des Gegners oder Stärke des Spielplans sowie in vielen Fällen der Heimvorteil, in den diversen Systemen miteinbezogen. So wie etwa bei der Anwendung des PageRank-Algorithmus im Taekwondo (Kim & Jeon, 2019). Die Leistung im Individualsport wird also nicht immer nur anhand von der eigens erbrachten Leistung quantifiziert, sondern kann auch durch äußere Umstände beeinflusst werden. Dies kann in dem Bestreben möglichst transparente Ranking- und Ratingsysteme aufzustellen als durchaus sinnvoll angesehen werden.

4.2.6 E-Sport

Mit Gong et al. (2020) und Wang et al. (2024) finden sich zwei sehr aktuelle Studien unter den Resultaten. Beide setzen sich mit ihren Vorschlägen das Ziel, die Fähigkeitslevel der Spieler:innen genau zu bestimmen und auch die Interaktionen zwischen den einzelnen Spieler:innen festzustellen. Dadurch sollen Teams zusammengestellt werden, dessen Dynamiken gut zusammenpassen und dadurch eine ausgeglichene Mannschaft gebildet werden. Wang et al. (2024) versucht durch die optimalen Teamzusammenstellungen das Spielerengagement zu erhöhen.

Ähnlichkeiten zum „normalen Sport“ gibt es dadurch, dass versucht wird die Spieler:innen anhand ihrer Fähigkeiten und Stärken einzuteilen und das optimale Match zu finden (Gong et al., 2020; Myślak & Deja, 2015; Wang et al., 2024). So wird auch bei Sauer et al. (2024) danach gestrebt, das Matchmaking-System im Schach zu verbessern und fairer zu gestalten. Im Schach wird die Stärke der Spieler:innen anhand des Elo-Ratings bestimmt und eingeordnet. Myślak & Deja (2015) stellen die Fertigkeiten und Stärken einer spielenden Person als Vektor dar, wodurch alle Teilnehmer:innen anhand der Fähigkeit in

einer gewissen Position zu spielen bewertet werden. Das Autorenteam zieht dabei das Beispiel des Spiels „„LOL““ heran, bei dem jeder Spielende einen Champion (Charakter) vor Beginn des Matches auswählen kann (Myślak & Deja, 2015).

Eine Arbeit aus dem E-Sport-Bereich, die sich mit Ranking im Spiel „Dota 2“ beschäftigt, wurde verfasst von Pradhan & Abdourazakou (2020). Das Multi-Criteria-Decision-Making-Tool stellt hier ein Power-Ranking dar, das die aktuelle Stärke eines Teams bewertet und mit anderen vergleicht. Im Basketball oder Eishockey fließen dabei etwa Siege und Niederlagen, individuelle Spielestatistiken, die Stärke des Spielplans oder der aktuelle Leistungstrend mit ein (Barrow et al., 2013). Verglichen mit dem E-Sport und in diesem Kontext mit dem Spiel „Dota 2“ wird das Power-Ranking durch Faktoren wie dem Prozentsatz der gewonnenen Spiele, Anzahl der Treffer („Kills“) oder der Anzahl der gewonnenen Punkte bestimmt (Pradhan & Abdourazakou, 2020).

Der E-Sport und konventionelle Sport haben in ihren Matchmaking- und Rankingsystemen ähnliche Ideen und Vorstellungen. Natürlich unterscheiden sie sich in ihren einzelnen Faktoren, was durch die unterschiedlichen Anforderungen der Sportart bedingt ist. Auch das Matchmaking im E-Sport muss aufgrund des großen Aufkommens an Spielern und Spielerinnen binnen kürzester Zeit noch schneller und effizienter funktionieren als in anderen Sportarten. Es vereint sie jedoch das Streben nach Fairness und ausgeglichenen Systemen, um die Chancengleichheit für alle ermöglichen zu können.

Alman & McKay (2017) beschreiben neben der Fairness auch den Begriff der Einheitlichkeit, der erfüllt sein muss, um ein Onlinespiel informell als ausgewogen zu bezeichnen. Die Autoren versuchen die beiden Faktoren zu quantifizieren und messbar zu machen, damit für alle Spieler:innen Chancengleichheit besteht.

4.3 Fairness und Heimvorteil

4.3.1 Fairness

Die Fairness stellt sicher, dass es für alle Teilnehmer: innen zu gleichen Wettbewerbschancen kommt. Sie sollen das Erlebnis Sport sowohl für Athleten und Athletinnen als auch für Zuschauer:innen verbessern. Dieser Faktor der Fairness ist auch in den diversen Ranking- und Matchmaking-Systemen von Nöten, damit diese im Sinne des Sports funktionieren (Powell, 2023).

Besonders im Matchmaking ist die Fairness von wichtiger Bedeutung. Sie bildet eine ungefähre Basis, die vorliegen muss, um ein erfolgreiches sowie effizientes Matchmaking-System auf den Markt zu bringen. Es geht im Matchmaking per Definition auch darum, möglichst ausgewogene und faire Teams zu bilden, um in der Folge wettbewerbsfähige Spiele zu generieren. Das zielt wiederum darauf ab, auch mit Blickpunkt auf den E-Sport-

Bereich, zufriedenstellende Spielerlebnisse zu schaffen, um dadurch auch Spieler:innen an das Spiel oder die Sportart zu binden. Besonders im E-Sport ist das Matchmaking eine Problemfrage, die immer wieder beantwortet werden muss (Gong et al., 2020; Myślak & Deja, 2015; Wang et al., 2024). Gong et al. (2020) versucht mit dem OptMatch-Matchmaking-System Spielerinteraktionen und Teamdynamiken zu verstehen und dadurch folglich in möglichst kurzer Zeit ausgeglichene Teams zu erstellen. Auch Wang et al. (2024) haben mit ihrem EnMatch System das Ziel, die Attribute und Stärken der Spieler:innen in Onlinespielen zu analysieren um faire Match-ups zu kreieren.

Nicht nur im E-Sport, sondern auch im Schach findet sich der Faktor Fairness in der Literatur wieder. Um die FIDE-Regeln zu befolgen, werden durch das schweizer Schachsystem, beschrieben bei Sauer et al. (2024), mehrere Kriterien festgelegt. Diese sollen die Fairness gewährleisten. Die Kriterien umfassen unter anderem das Einhalten einer ausgewogenen Farbverteilung, die besagt, dass alle Spieler:innen eine annähernd gleiche Anzahl an Spielen mit weiß und schwarz bestreiten. Das Spielen mit weiß wird als Vorteil angesehen. Weiters soll durch die Kriterien auch gesichert sein, dass Spieler:innen in einem Turnier nicht mehr als einmal gegeneinander spielen, um die Abwechslung zu gewähren. Sauer et al. (2024) versucht mit dem Maximum-Weight-Matching-Ansatz die wahre Stärke der Spieler:innen darzustellen. Mit solchen Ansätzen könnte es gelingen die ursprünglichen Kriterien der FIDE bezüglich der Fairness zu übertreffen, indem Spieler:innen mit einem ähnlichen Stärkeprofil zusammengebracht werden (fide.com, 2025).

Der Einfluss des Faktors Fairness dürfte im Kontext der Ranking- und Matchmaking-Systeme sehr groß sein. Anhand der Ergebnisse kann erkannt werden, dass Autoren und Autorinnen versuchen diesem Faktor mit ihren Systemen zurecht zu werden. Das Ziel, die Fairness stetig zu verbessern, ist immer wieder zu erkennen. Ein Beispiel dafür ist das Medaillenranking bei den olympischen Spielen (Duncan & Parece, 2024). Bei den Spielen haben große Länder, die auf einen großen Pool an Athleten und Athletinnen zurückgreifen können, andere Voraussetzungen als sehr kleine Länder. Weil die Medaillenrankings nach der absoluten Anzahl der metallenen Erfolge gereiht werden, stehen dementsprechend die großen Länder an der Spitze. Duncan & Parece (2024) verfolgen dabei den Ansatz Länder anhand ihrer tatsächlichen Fähigkeit Medaillen zu gewinnen einzustufen. Die Länder werden dann anhand der gewonnenen im Vergleich zu den zu erwartenden Medaillen eingereiht. Dies ermöglicht auch verhältnismäßig kleinen Ländern, wie Neuseeland, einen hohen Platz im Ranking bei den Olympischen Spielen 2020 in Tokyo.

4.3.2 Fairness in Onlinespielen

Um feststellen zu können, ob ein Spiel fair ist, kann zunächst eine Messung der Fähigkeiten eines Teams herangezogen werden. Ein:e Spieler:in mit hohen individuellen Fähigkeiten kann möglicherweise einen großen Einfluss auf den Ausgang des Spiels haben. Doch nur weil ein Team über hohe individuelle Fähigkeiten verfügt, muss dies noch keine gut zusammenspielende Mannschaft bedeuten. Wie gut die Spieler:innen zusammenarbeiten, wird über die Teamdynamik definiert. Die generelle Stärke eines Teams kann definiert werden anhand der p -Fähigkeit eines Teams (Alman & McKay, 2017).

$$s_p(X) := \left(\sum_{x \in X} s_x^p \right)^{1/p},$$

Anhand dieser Formel werden die Fähigkeiten aller Spieler:innen summiert und anhand der p -Norm wird die Stärke des Teams ermittelt. Der Parameter S steht dabei für die Fähigkeit oder Stärke der Spieler:innen. Die Komponente p ist eine reelle Zahl, sie ist größer oder gleich eins. Anhand derer können die Fähigkeiten des gesamten Teams berechnet werden. In weiterer Folge soll festgestellt werden, wie fair das Spiel zweier Teams ist (p -fairness). Dies wird ermittelt, indem der Unterschied in den p -Fähigkeiten zwischen beiden Teams berechnet wird (Alman & McKay, 2017):

$$\begin{aligned} d_p(X, Y) &:= |s_p(X) - s_p(Y)| \\ &= \left| \left(\sum_{x \in X} s_x^p \right)^{1/p} - \left(\sum_{y \in Y} s_y^p \right)^{1/p} \right|. \end{aligned}$$

Desto näher die p -Fairness Richtung 0 geht, desto fairer und ausgeglichener ist das Spiel. Anhand der p -Fairness kann es Spielentwicklern gelingen ausgewogene Teams zusammenzustellen und somit gleiche Gewinnchancen für alle zu ermöglichen (Alman & McKay, 2017).

4.3.3 Einheitlichkeit in Onlinespielen

Auch wenn ein Spiel als fair gewertet werden kann, bedeutet dies nicht automatisch, dass für alle Spieler:innen das beste Spielerlebnis garantiert werden kann. Wenn der Unterschied der Fähigkeiten innerhalb eines Teams zwischen den einzelnen spielenden Personen zu groß ist, kann nicht das beste Spielergebnis geboten werden. Im Kontext der Einheitlichkeit soll durch das Zusammenbringen von mit einem ähnlichen Erfahrungs- oder Fähigkeitsniveau die Chancengleichheit innerhalb eines Teams gesteigert werden (Ainsworth et al., 2012).

Mithilfe der Standardabweichung kann berechnet werden, wie stark die Fähigkeiten der Spieler:innen eines Teams voneinander abweichen. Eine niedrige Standardabweichung würde bedeuten, dass die Spieler:innen mit ihren Fähigkeiten dem Durchschnitt sehr nahekommen. Ist der ausgerechnete Wert sehr hoch, spricht das für sehr unterschiedliche Fähigkeitslevel. Der Parameter s_z steht dabei für die Fähigkeiten der einzelnen Personen (Alman & McKay, 2017).

Die Einheitlichkeit wird in Folge als Q-Einheitlichkeit v_q ausgedrückt. Dies ist eine Methode, um zu messen, wie konsistent oder einheitlich eine bestimmte Gruppe von Zahlen ist. Relevant für die Berechnung ist der Parameter q , der eine Zahl darstellt, die größer oder gleich 1 ist. Anhand dieses Parameters können die Unterschiede zum Durchschnitt bemessen werden. Die Formel für die Q-Uniformität wird stellt sich wie folgt dar (Alman & McKay, 2017):

$$v_q(Z) := \left(\frac{1}{|Z|} \sum_{z \in Z} |s_z - \mu|^q \right)^{1/q}.$$

Die Q-Uniformität wird immer als eine nichtnegativ reelle Zahl ausgedrückt. Das Ergebnis kann somit 0 oder eine beliebige positive Zahl sein. Wenn das Ergebnis gleich 0 ist, lässt sich daraus schließen, dass eine Einheitlichkeit im Team besteht und die Leistung der Spieler:innen gleich ist. Desto größer das Ergebnis der Q-Uniformität ausfällt, desto größer sind die Unterschiede innerhalb eines Teams. Mit diesem Instrument im Matchmaking kann es gelingen ein besseres Spielerlebnis für alle Spieler:innen zu generieren (Alman & McKay, 2017).

4.3.4 Heimvorteil

Der Heimvorteil versteht sich als Vorteil, den ein:e Athlet:in oder ein Team hat, wenn es zu Hause antreten darf im Vergleich zu der Leistung die Auswärts erbracht wird. Die Problemstellung liegt in früheren Arbeiten auch darin, dass der Faktor Heimvorteil in den Rankingsystemen verschiedener Sportarten nicht berücksichtigt wurde. Dabei ist auch diese Komponente für ein faires und ausgeglichenes Rankingsystem notwendig (Pollard & Gómez, 2009). Schlussendlich geht es für Sportler:innen bei der Leistungsbewertung um Chancengleichheit. Denn die Endposition in einem Ranking entscheidet in vielen Sportarten, wie etwa im Judo, ob der:die Sportler:in an den Olympischen Spielen teilnehmen darf. Weiters gilt es den Heimvorteil auch bei der Organisation von Wettkämpfen und Turnieren zu beachten, um für Alle die gleichen Voraussetzungen zu schaffen (Ferreira Julio et al., 2013).

Um den Heimvorteil zu quantifizieren, verwendet die Literatur auch erneut das Elo-System. Mangan & Collins (2016) arbeiten mit modifizierten Elo-Ratings, um den Heimvorteil im Gaelic Football auszudrücken. Die World Football Elo Ratings legen den Heimvorteil mit 100 Punkten fest. Das Autorenteam legt diesen Wert in ihrer Studie auf 79 Punkte, die dem Heimteam bei der Bewertung zugeschrieben werden. Auch Wolf et al. (2020) basieren ihr Ratingsystem für männliche Fußballspieler auf dem Elo-Algorithmus. Der Heimvorteil ist in diesem System ein dynamischer Faktor, der sich verändert je nachdem, ob das Heimteam die Erwartungen erfüllt oder nicht. In diesem Beispiel nimmt der Heimvorteil nicht nur einen Einfluss auf die Bewertung des Teams, sondern auch auf jeden einzelnen Spieler. Der Heimvorteil wird auch besonders dann zum Faktor, wenn – um nochmal auf den Punkt der Vorhersagemodelle zurückzugreifen – eine genaue Vorhersage der Leistung eines Teams getroffen werden soll (Wolf et al., 2021).

Der Heimvorteil dürfte auch auf wichtige Spielstatistiken einen Einfluss nehmen. Laut Van Bommel et al. (2021) performen NCAA-Basketballteams (Männer und Frauen) zu Hause besser. Dargestellt wird dies anhand der wichtigsten Box-Score-Statistiken (Punkte, Assists, Blocks, Feldtorversuche). Die Spieler:innen verzeichnen dabei zu Hause etwa mehr Assists oder Blocks und versuchen sich an mehr Würfeln. In diesem Kontext dürfte auch die Anwesenheit der heimischen Zuschauer eine bedeutende Rolle spielen. Im Fußball kann der Zeitpunkt des Heimspiels etwa dann entscheidend sein, wenn das Rückspiel eines KO-Spiels im eigenen Stadion ausgetragen wird (van Bommel et al., 2021). Waquil et al. (2020) analysieren dabei Spiele aus dem brasilianischen Fußball-Cupwettbewerb zwischen 1994 und 2017. Die Ergebnisse zeigen dabei, dass in über 63% der Spiele in KO-Format sich das Team über zwei Spiele durchsetzt hat, welches das Rückspiel zu Hause beschritten hatte. Es ist davon auszugehen, dass Teams in einem solchen Format danach streben, so gesetzt zu werden, sodass das Rückspiel zu Hause gespielt wird. Die Einführung der Auswärtstorregel, durch welche die Tore der Auswärtsmannschaft mehr zählen, minimiert diesen Vorteil (Waquil et al., 2020).

Zurückkehrend zum NCAA-Basketball wird auch im Hinblick auf das 2018 eingeführte Vier-Quadranten-System der Heimvorteil bewertet. Das System bewertet die unterschiedlichen Werte von Siegen und Niederlagen und beurteilt danach die Teams. Es bewertet etwa einen Sieg auswärts gegen ein heimstarkes Team als wertvoller als einen Sieg gegen dasselbe Team zu Hause. Das Quadranten-System hat schließlich eine Auswirkung auf die Setzliste für NCAA-Turniere (Reinig & Horowitz, 2019).

Der Heimvorteil dürfte sich besonders bei neueren Systemen als immer wichtigerer Faktor implementiert haben. Autoren und Autorinnen versuchen in ihren Systemen den Heimvorteil auch zu quantifizieren, besonders auch im Hinblick auf die Fairness. Er kann als

dynamische Komponente in einem Ranking- und/oder Matchmaking-System verwendet werden, um die Bedeutsamkeit eines Heimsieges oder dem Sieg bei einem starken Heimteam zu unterstreichen und in den Bewertungen zu berücksichtigen.

5 Zusammenfassung

Im Rahmen dieser Übersichtsarbeit wurden Ranking- und Matchmaking-Systeme verglichen und kategorisiert. Dabei wurden die Resultate in vier verschiedene Kategorien eingeteilt und deren Charakteristik dargestellt.

Die erste Kategorie behandelt individuelle Spielerbewertungen und Ratingsysteme. Dabei lässt sich erkennen, dass Ratingsysteme für Sportler:innen aus Teamsportarten oft von der Leistung und der Bewertung des gesamten Teams abhängig sind und umgekehrt (Ambrutis & Povilaitis, 2024; Wolf et al., 2021). Andererseits sind die Ratings von den tatsächlichen Leistungen des Einzelsportlers oder der Einzelsportlerin in Sportarten, wie Biathlon oder Tennis, abhängig. Wie weit Spieler:innen in Tennis-Turnieren kommen oder wie oft sie verletzt sind, sind beeinflussende Faktoren (Irons et al., 2014).

Die zweite Kategorie trägt den Titel „Team Performance Ranking/Rating“ und beinhaltet jene Arbeiten, die sich mit der Bewertung von Teamsportarten auseinandersetzen. Dabei liegt insbesondere die Inklusion von Faktoren, wie dem Heimvorteil oder der Stärke des Spielplanes, in die jeweiligen Systeme im Fokus. Die Rankings und die Stärke des Spielplans bedingen sich dabei gegenseitig. Die tatsächliche Stärke eines Teams soll bewertet werden und in das Ranking miteinfließen. So etwa beim Offense-Defense Modell bei Govan et al. (2009).

In der dritten und deutlich kleineren Kategorie dem Matchmaking finden sich drei aus vier Studien im Bereich des E-Sports wieder. Im E-Sport dürfte das Matchmaking von noch wichtigerer Bedeutung sein, weil innerhalb kürzester Zeit tausende von Spielern oder Spielerinnen in passende Match-ups eingeteilt werden müssen. Auch dabei ist es wichtig, die Stärken und Fähigkeiten der Spieler:innen beurteilen zu können (Gong et al., 2020; Myślak & Deja, 2015).

In der vierten und letzten Kategorie der behandelten Studien lassen sich Ranking und Matchmaking nicht klar voneinander trennen lassen. Dabei geht es im NCAA-Basketball etwa darum, das Ranking von Teams transparenter zu gestalten um anhand dieser Bewertungen faire Entscheidungen zu treffen, wer am populären NCAA-Turnier teilnehmen darf (Reinig & Horowitz, 2019).

Eine weitere Möglichkeit die vorliegenden Ranking- und Matchmaking-Systeme zu kategorisieren, ist sie anhand der mathematischen Hintergründe einzuteilen. Häufig werden die Elo-Ratings als Basis für das System verwendet, sowie Markov-Ketten und iterative Prozesse angewandt. Es ergeben sich auch weitere Kategorisierungsmöglichkeiten wie nach den Top-down oder Bottom-up-Ansätzen.

Literaturverzeichnis

- Ainsworth, J., Cass, R. J., & Andersen, A. T. (2012). *Systems, methods, and software arrangements for improving uniformity of assets within an entity*.
- Albers, P. C. H., & Vries, H. de. (2001). *Elo-rating as a tool in the sequential estimation of dominance strengths*.
- Alman, J., & McKay, D. (2017). Theoretical Foundations of Team Matchmaking. *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, 1073–1081.
- Ambrutis, A., & Povilaitis, M. (2024). Composite rating method: Application to European basketball leagues. *J Sports Sci*, 42, 214. <https://doi.org/10.1080/02640414.2024.2326275>
- And the Winner Is? (1987). *International Journal of Sport Biomechanics*, 3(2), 168–168.
- Baerg, A. (2011). Classifying the Digital Athletic Body: Assessing the Implications of the Player-Attribute-Rating System in Sports Video Games. *International Journal of Sport Communication*, 4(2), 133–147. <https://doi.org/10.1123/ijsc.4.2.133>
- 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, 202. <https://doi.org/10.1515/jqas-2013-0013>
- BASHARIN, G. P., LANGVILLE, A. N., & NAUMOV, V. A. (2004). *The life and work of A.A. Markov*. 386, 26.
- Berker, Y. (2014). Tie-breaking in round-robin soccer tournaments and its influence on the autonomy of relative rankings: UEFA vs. FIFA regulations. *European Sport Management Quarterly*, 14, 210. <https://doi.org/10.1080/16184742.2014.884152>
- Blackburn, M. L. (2013). Ranking the performance of tennis players: An application to women's professional tennis. *Journal of Quantitative Analysis in Sports*, 9, 378. <https://doi.org/10.1515/jqas-2013-0006>
- Bosch, R., & Verborgh, R. (2016). Iterative querying of the semantic web. *KWALON (Amsterdam)*, 21. <https://doi.org/10.5117/2016.021.001.004>
- Broadie, M., & Rendleman, R. J. (2013). Are the official world golf rankings biased? *Journal of Quantitative Analysis in Sports*, 9(2), 127–140. <https://doi.org/10.1515/jqas-2012-0013>
- Buchberger, B., von Elm, E., Gartlehner, G., Huppertz, H., Antes, G., Wasem, J., & Meerpohl, J. (2014). Assessment of the Risk of Bias in controlled Studies. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz*, 57, 1438. <https://doi.org/10.1007/s00103-014-2065-6>

- Bührer, W. (2019). Nina Verheyen, Die Erfindung der Leistung, Hanser, Berlin 2018, S. 256, € 23,00. *Zeitschrift Für Unternehmensgeschichte*, 64. <https://doi.org/10.1515/zug-2018-0022>
- Chan, T. C. Y., & Singal, R. (2016). A Markov Decision Process-based handicap system for tennis. *Journal of Quantitative Analysis in Sports*, 12, 188. <https://doi.org/10.1515/jqas-2016-0057>
- Chan, T. C. Y., & Singal, R. (2018). A Bayesian regression approach to handicapping tennis players based on a rating system. *Journal of Quantitative Analysis in Sports*, 14(3), 131–141. <https://doi.org/10.1515/jqas-2017-0103>
- Chan, V. (2022). MSE-optimal K-factor of the Elo rating system for round-robin tournament. *Journal of Quantitative Analysis in Sports*, 18(1), 59–72. <https://doi.org/10.1515/jqas-2021-0079>
- Chartier, T. P., Kreutzer, E., Langville, A. N., & Pedings, K. E. (2011). Sports Ranking with Nonuniform Weighting. *Journal of Quantitative Analysis in Sports*, 7, 1290. <https://doi.org/10.2202/1559-0410.1290>
- Colley, W. N. (2002). *Colley's bias free college football ranking method: The colley matrix explained*. www.Colleyrankings.com
- Constantinou, A. C., & Fenton, N. E. (2013). Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries. *Journal of Quantitative Analysis in Sports*, 9(1), 37–50. <https://doi.org/10.1515/jqas-2012-0036>
- Delgado-Bordonau, J., Doménech-Monforte, C., Guzmán, J., & Mendez-Villanueva, A. (2013). Offensive and defensive team performance: Relation to successful and unsuccessful participation in the 2010 Soccer World Cup. *Journal of Human Sport & Exercise*, 8, Aceptado. <https://doi.org/10.4100/jhse.2013.84.02>
- Devlin, S., Treloar, T., Creagar, M., & Cassels, S. (2021). An iterative Markov rating method. *Journal of Quantitative Analysis in Sports*, 17, 127. <https://doi.org/10.1515/jqas-2019-0070>
- Dingle, N., Knottenbelt, W., & Spanias, D. (2013). *On the (Page) Ranking of Professional Tennis Players* (M. Tribastone & S. Gilmore, Eds.; Vol. 7587, p. 247). https://doi.org/10.1007/978-3-642-36781-6_17
- Duncan, R. C., & Parece, A. (2024). Population-adjusted national rankings in the Olympics. *Journal of Sports Analytics*, 10, 104. <https://doi.org/10.3233/JSA-240874>
- Elo, A. E. (n.d.). *The Rating of Chessplayers, Past and Present*. Arco Pub.

- ESPELAND, W. N., & SAUDER, M. (2007). Rankings and reactivity: How public measures recreate social worlds. *The American Journal of Sociology*, 113, 40. <https://doi.org/10.1086/517897>
- Filipicic, A., Panjan, A., & Sarabon, N. (2014). Classification of top male tennis players. *International Journal of Computer Science in Sport (International Association of Computer Science in Sport)*, 13(1), 2–2.
- Franchini, E., & Ferreira Julio, U. (2015). The Judo World Ranking List and the Performances in the 2012 London Olympics. *Asian Journal of Sports Medicine*, 6(3), 1–3. <https://doi.org/10.5812/asjasm.24045>
- Gebser, M., Glase, T., Sabuncu, O., & Schaub, T. (n.d.). *Matchmaking with Answer Set Programming* (P. Cabalar & T. C. Son, Eds.; p. 347). https://doi.org/10.1007/978-3-642-40564-8_34
- Gill, R., & Keating, J. (2009). Assessing Methods for College Football Rankings. *Journal of Quantitative Analysis in Sports*, 5. <https://doi.org/10.2202/1559-0410.1172>
- Glickman, M. E. (1999). Parameter Estimation in Large Dynamic Paired Comparison Experiments. *Applied Statistics*, 48, 394. <https://doi.org/10.1111/1467-9876.00159>
- Gollub, J. (2021). Forecasting serve performance in professional tennis matches. *Journal of Sports Analytics*, 7, 233. <https://doi.org/10.3233/JSA-200345>
- Gong, L., Feng, X., Ye, D., Li, H., Wu, R., Tao, J., Fan, C., & Cui, P. (2020). *OptMatch: Optimized Matchmaking via Modeling the High-Order Interactions on the Arena*. 2310. <https://doi.org/10.1145/3394486.3403279>
- Govan, A. Y., Langville, A. N., & Meyer, C. D. (2009). Offense-Defense Approach to Ranking Team Sports. *Journal of Quantitative Analysis in Sports*, 5. <https://doi.org/10.2202/1559-0410.1151>
- Gupta, A. A. (2015). A new approach to bracket prediction in the NCAA Men's Basketball Tournament based on a dual-proportion likelihood. *Journal of Quantitative Analysis in Sports*, 11, 67. <https://doi.org/10.1515/jqas-2014-0047>
- Heasman, J., Dawson, B., Berry, J., & Stewart, G. (2008). Development and validation of a player impact ranking system in Australian football. *International Journal of Performance Analysis in Sport*, 8(3), 156–171. <https://doi.org/10.1080/24748668.2008.11868457>
- Heintz, B. (2010). Numerische Differenz. Überlegungen zu einer Soziologie des (quantitativen) Vergleichs. *Zeitschrift Für Soziologie*, 39.

- Helena Hartmann, L., Schulz-Wiemann, C., Spiller, A., & Gertheiss, J. (2014). Weiterentwicklung der Rankingsysteme im Reitsport: Ein Experiment. *Sportwissenschaft*, 44, 115. <https://doi.org/10.1007/s12662-013-0323-7>
- Herbrich, R., Minka, T., & Graepel, T. (2007). *TrueSkill™: A Bayesian skill rating system*. 576.
- Hvattum, L. M. (2019). A comprehensive review of plus-minus ratings for evaluating individual players in team sports. *International Journal of Computer Science in Sport (Sciendo)*, 18(1), 1–23. <https://doi.org/10.2478/ijcss-2019-0001>
- Ingram, M. (2021). How to extend Elo: A Bayesian perspective. *Journal of Quantitative Analysis in Sports*, 17(3), 203–219. <https://doi.org/10.1515/jqas-2020-0066>
- Irons, D. J., Buckley, S., & Paulden, T. (2014). Developing an improved tennis ranking system. *Journal of Quantitative Analysis in Sports*, 10(2), 109–118. <https://doi.org/10.1515/jqas-2013-0101>
- Jeremic, V. M., & Radojicic, Z. (2010). A New Approach in the Evaluation of Team Chess Championships Rankings. *Journal of Quantitative Analysis in Sports*, 6, 1257. <https://doi.org/10.2202/1559-0410.1257>
- Kelley, J. G., & Simmons, B. A. (2015). Politics by Number: Indicators as Social Pressure in International Relations. *American Journal of Political Science*, 59, 70. <https://doi.org/10.1111/ajps.12119>
- Kim, E., & Jeon, M. (2019). Proposal for implementation of a ranking model for Olympic Taekwondo competitions using PageRank. *International Journal of Performance Analysis in Sport*, 19, 235. <https://doi.org/10.1080/24748668.2019.1586506>
- Langville, A. N., & Meyer, C. D. (2012). *Who's #1?: The science of rating and ranking* (1st ed.). <https://doi.org/10.1515/9781400841677>
- Leicht, A. S., Gómez, M. A., & Woods, C. T. (2017). Explaining Match Outcome During The Men's Basketball Tournament at The Olympic Games. *Journal of Sports Science & Medicine*, 16(4), 468–473.
- Lowery, B., Slater, A., & Thies, K. (2020). Functional ratings in sports. *Journal of Quantitative Analysis in Sports*, 16, 191. <https://doi.org/10.1515/jqas-2020-0001>
- Mangan, S., & Collins, K. (2016). A Rating System For Gaelic Football Teams: Factors That Influence Success. *International Journal of Computer Science in Sport (Sciendo)*, 15(2), 78–90. <https://doi.org/10.1515/ijcss-2016-0006>
- Mangan, S., Malone, S., Ryan, M., Mc Gahan, J., Warne, J., Martin, D., O'Neill, C., Burns, C., & Collins, K. (2018). Influence of Team Rating on Running Performance in Elite Gaelic

- Football. *Journal of Strength & Conditioning Research*, 32(9), 2584–2591.
<https://doi.org/10.1519/JSC.0000000000002316>
- Matthews, N. R., McClain, A., Smith, C. M. L., & Tennant, A. G. (2021). Application of PageRank Algorithm to Division I NCAA men's basketball as bracket formation and outcome predictive utility. *Journal of Sports Analytics*, 7, 9. <https://doi.org/10.3233/JSA-200425>
- Miles, W. W., Fowks, G. T., & Coulter, L. O. (2010). AccuV College Football Ranking Model. *Journal of Quantitative Analysis in Sports*, 6(3), 1–17. <https://doi.org/10.2202/1559-0410.1189>
- Munn, Z., Stern, C., Aromataris, E., Lockwood, C., & Jordan, Z. (2018). What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Med Res Methodol*, 18, 5. <https://doi.org/10.1186/s12874-017-0468-4>
- Musante, T. M., & Yellin, B. A. (1979). USTA-equitable computerized tennis ranking system. *Interfaces*, 9(4), 33–37.
- Myślak, M., & Deja, D. (2015). *Developing Game-Structure Sensitive Matchmaking System for Massive-Multiplayer Online Games* (D. McFarland & L. M. Aiello, Eds.; Vol. 8852, p. 208). https://doi.org/10.1007/978-3-319-15168-7_25
- Ncaa.com (2025). College basketball's NET rankings, explained. Zugriff am 18. Juli 2025 unter <https://www.ncaa.com/news/basketball-men/article/2022-12-05/college-basketballs-net-rankings-explained>
- NEL, J. (2013). ESTIMATING SUCCESS PROBABILITY OF A RUGBY GOAL KICK AND DEVELOPING A MEASURE FOR RANKING RUGBY UNION GOAL KICKERS. *South African Journal for Research in Sport, Physical Education & Recreation*, 35(1), 133–142.
- Newton, P. K., & Aslam, K. (2009). Monte Carlo Tennis: A Stochastic Markov Chain Model. *Journal of Quantitative Analysis in Sports*, 5. <https://doi.org/10.2202/1559-0410.1169>
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web. *The Web Conference*. <https://api.semanticscholar.org/CorpusID:1508503>
- Pakaslahti, A. (2019). The use of head-to-head records for breaking ties in round-robin soccer contests. *Journal of the Philosophy of Sport*, 46(3), 355–366. hsi.
- Palmen, M. (2014). *Das Klassifizierungssystem der paralympischen Sportarten* (M. Palmen & E. Hillenbach, Eds.). Bundesinstitut für Sportwissenschaft.

http://www.bisp.de/SharedDocs/Downloads/Publikationen/sonstige_Publikationen_Ratgeber/Klassifizierung.pdf?__blob=publicationFile&v=2

- Paul von Dohlen. (2011). Tweaking the NFL's Quarterback Passer Rating for Better Results. *Journal of Quantitative Analysis in Sports*, 7(3), 1359–1359. <https://doi.org/10.2202/1559-0410.1359>
- Pollard, R., & Gómez, M. A. (2009). Home advantage in football in South-West Europe: Long-term trends, regional variation, and team differences. *European Journal of Sport Science*, 9, 352. <https://doi.org/10.1080/17461390903009133>
- Pollock, N., & D'Adderio, L. (2012). Give me a two-by-two matrix and I will create the market: Rankings, graphic visualisations and sociomateriality. *Accounting, Organizations and Society*, 37, 586. <https://doi.org/10.1016/j.aos.2012.06.004>
- Powell, B. (2023). Generalizing the Elo rating system for multiplayer games and races: Why endurance is better than speed. *Journal of Quantitative Analysis in Sports*, 19, 243. <https://doi.org/10.1515/jqas-2023-0004>
- Pradhan, S., & Abdourazakou, Y. (2020). "Power ranking" professional circuit eSports teams using multi-criteria decision-making (MCDM). *Journal of Sports Analytics*, 6, 73. <https://doi.org/10.3233/JSA-190420>
- Raack, C., Raymond, A., Schlechte, T., & Werner, A. (2014). Standings in sports competitions using integer programming. *Journal of Quantitative Analysis in Sports*, 10, 137. <https://doi.org/10.1515/jqas-2013-0111>
- Ranking College Football Programs Based Upon Athletic Performance and Academic Success. (2014). *Sport Journal*, 1–1.
- Reinig, B. A., & Horowitz, I. (2019). Analyzing the impact of the NCAA Selection committee's new quadrant system. *Journal of Sports Analytics*, 5, 333. <https://doi.org/10.3233/JSA-190337>
- Ringel, L., & Werron, T. (2020). *Where Do Rankings Come From?: A Historical-Sociological Perspective on the History of Modern Rankings* (A. Eppele, J. Grave, & W. Erhart, Eds.; p. 170). <https://doi.org/10.1515/9783839451663-006>
- Rousseau, C., Saint-Aubin, Y., & Stern, M. (2012). Google und der PageRank-Algorithmus. In C. Rousseau & Y. Saint-Aubin (Eds.), *Mathematik und Technologie* (pp. 279–304). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30092-9_9
- SangHyuk Im & Chang-Hoon Lee. (2023). World Tennis Number: The new gold standard, or a failure? *Coaching & Sport Science Review*, 91, 6–12.

- Sauer, P., Cseh, Á., & Lenzner, P. (2024). Improving ranking quality and fairness in Swiss-system chess tournaments. *Journal of Quantitative Analysis in Sports*, 20(2), 127–146. <https://doi.org/10.1515/jqas-2022-0090>
- Schultz, M., Mouritsen, J., & Gabrielsen, G. (2001). Sticky Reputation: Analyzing a Ranking System. *Corporate Reputation Review*, 4, 41. <https://doi.org/10.1057/palgrave.crr.1540130>
- Servien, R. (2022). An alternative ranking system for biathlon pursuit races. *Journal of Sports Analytics*, 8, 148. <https://doi.org/10.3233/JSA-200598>
- Sharp, B. (2007). Tennis is a Sport for All—ITN. *Coaching & Sport Science Review*, 15(42), 8–9.
- Stekler, H. O., & Klein, A. (2012). Predicting the Outcomes of NCAA Basketball Championship Games. *Journal of Quantitative Analysis in Sports*, 8. <https://doi.org/10.1515/1559-0410.1373>
- Stewart, M. J. (2013). A proposed general rating system for which the Colley Matrix Rating System is a special case. *Journal of Quantitative Analysis in Sports*, 9(2), 179–185. <https://doi.org/10.1515/jqas-2012-0039>
- Szczecinski, L., & Roatis, I.-I. (2022). FIFA ranking: Evaluation and path forward. *Journal of Sports Analytics*, 8, 250. <https://doi.org/10.3233/JSA-200619>
- Trent Herda, Eric Ryan, Jason DeFreitas, Pablo Costa, Ashley Walter, Katherine Hoge, Joseph Weir, & Joel Cramer. (2009). Can Recruiting Rankings Predict the Success of NCAA Division I Football Teams? An Examination of the Relationships among Rivals and Scouts Recruiting Rankings and Jeff Sagarin End-of-Season Ratings in Collegiate Football. *Journal of Quantitative Analysis in Sports*, 5(4), 1200–1200. <https://doi.org/10.2202/1559-0410.1200>
- Trono, J. A. (2010). Rating/Ranking Systems, Post-Season Bowl Games, and ‘The Spread’. *Journal of Quantitative Analysis in Sports*, 6(3), 1–20. <https://doi.org/10.2202/1559-0410.1220>
- Ursula, Panissa, V. L. G., Miarka, B., Takito, M. Y., & Franchini, E. (2013). Home advantage in judo: A study of the world ranking list. *J Sports Sci*, 31, 218. <https://doi.org/10.1080/02640414.2012.725855>
- van Bommel, M., Bornn, L., Chow-White, P., & Gao, C. (2021). Home sweet home: Quantifying home court advantages for NCAA basketball statistics. *Journal of Sports Analytics*, 7, 36. <https://doi.org/10.3233/JSA-200450>

- Vaughan Williams, L., Liu, C., Dixon, L., & Gerrard, H. (2021a). How well do Elo-based ratings predict professional tennis matches? *Journal of Quantitative Analysis in Sports*, 17(2), 91–105. <https://doi.org/10.1515/jqas-2019-0110>
- Vaughan Williams, L., Liu, C., Dixon, L., & Gerrard, H. (2021b). How well do Elo-based ratings predict professional tennis matches? *Journal of Quantitative Analysis in Sports*, 17, 105. <https://doi.org/10.1515/jqas-2019-0110>
- Wang, K., Liu, H., Hu, Z., Feng, X., Zhao, M., Zhao, S., Wu, R., Shen, X., Lv, T., & Fan, C. (2024). *EnMatch: Matchmaking for Better Player Engagement via Neural Combinatorial Optimization*. 38, 9106. <https://doi.org/10.1609/aaai.v38i8.28760>
- Waquil, A. P., Horta, E., & Moraes, J. C. (2020). Home advantage and away goals rule: An analysis from Brazil Cup. *Journal of Sports Analytics*, 6, 24. <https://doi.org/10.3233/JSA-200370>
- Werron, T. (2015). Why do we believe in competition? A historical-sociological view of competition as an institutionalized modern imaginary. *Distinktion (Aarhus)*, 16, 210. <https://doi.org/10.1080/1600910X.2015.1049190>
- Wigness, M. B., Williams, C. C., & Rowell, M. J. (2010). A New Iterative Method for Ranking College Football Teams. *Journal of Quantitative Analysis in Sports*, 6. <https://doi.org/10.2202/1559-0410.1242>
- Wolf, S., Schmitt, M., & Schuller, B. (2021). A football player rating system. *Journal of Sports Analytics*, 6, 257. <https://doi.org/10.3233/JSA-200411>
- Wu, Y., Song, W., Cao, Z., & Zhang, J. (2021). Learning Large Neighborhood Search Policy for Integer Programming. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2111.03466>
- Wunderlich, F., & Memmert, D. (2016). Analysis of the predictive qualities of betting odds and FIFA World Ranking: Evidence from the 2006, 2010 and 2014 Football World Cups. *J Sports Sci*, 34, 2184. <https://doi.org/10.1080/02640414.2016.1218040>
- ZRIKE KENNETH L & ZRIKE CHRISTOPHER A. (2010). *SPORTS MATCHMAKER SYSTEMS*.

Abbildungsverzeichnis

Figure 1 Der angewandte Search-String (Eigene Darstellung)	24
Figure 2 PRISMA FLOW Diagramm zur Übersicht der eingeschlossenen Studien (rayyan.ai).....	26
Figure 3 Ranking von elitären Fußballspielern anhand des Players-impact (Eigene Darstellung, adaptiert nach Wolf et al., 2021).....	29
Figure 4 Bewertung von College Basketball Teams nach der Colley und Massey Methode (Chartier et al., 2011)	39
Figure 5 Tier-based-heuristic-Matchmaking (Wang et al., 2024)	43
Figure 6 Individuelles Spieler Ranking anhand vom PPR (Ambrutis & Povilaitis, 2024)...72	
Figure 7 Die Bewertung von Fußballspielern anhand der PM (links) oder VAEP-Ratings (rechts) zum Stand Juli 2019 (Hvattum & Gelade, 2021).....	74
Figure 8 Die Top 15 Spieler nach dem Players-impact-Ranking; Stand Mai 2018 (Eigene Darstellung, adaptiert nach Wolf et al., 2021).....	76
Figure 9 Nationales Medaillen Ranking nach dem Population-adjusted national-Ranking für Olympia 2020 in Tokyo (Duncan & Parece, 2024)	79
Figure 10 Die Klassifikation der 2008 Schach Teammeisterschaften in Serbien mit der I-Distanz Methode im Vergleich zur Endplatzierung (Jeremic & Radojicic, 2009)	82
Figure 11 Das PageRank Diagramm (Matthews et al., 2021)	89

Tabellenverzeichnis

Tabelle 1 Ergebnisse: Individual-Player-Rating/Ranking (Eigene Darstellung).....	30
Tabelle 2 Ergebnisse: Team-Performance-Rankings (Eigene Darstellung).....	31
Tabelle 3 Ergebnisse: Matchmaking-Systeme (Eigene Darstellung)	34
Tabelle 4 Ergebnisse: Kombination aus Ranking- und Matchmaking-Systemen (Eigene Darstellung).....	35
Tabelle 5 Übersichtstabelle über die mathematischen Hintergründe (Eigene Darstellung	37

Abkürzungsverzeichnis

AL= American League

ATP = Association of Tennis Professionals

CMRS= Colley-Matrix-Ratingsystem

FIDE = Fédération Internationale des Échecs, Internationaler Schachverband

FIFA = Fédération Internationale de Football Association; Weltweiter Fußballverband

GRA= Grey-Relational-Analysis-Methode

GSVs= Game Score Violations

HA = Home Advantage; Heimvorteil

„LOL“= League of Legends; Computerspiel; Riot Games, 2009

MCDM= Multi-criteria-decision-making

MLB= Major League Baseball

MWM= Maximum-Weight-Matching

NBA = National Basketball Association

NCAA = National Collegiate Athletic Association

NCAAB= National Collegiate Athletic Association Basketball

NCAAF= National Collegiate Athletic Association Football

NHL= National Hockey League

NL= National League Baseball

OWGR= Official World Golf Ranking

PvP= Player-versus-Player Spiele

RPI= Rating-Percentage-Index

SAW= Simple-additive-Weighting

SOS = Strength of Schedule; Stärke des Spielplans

WTA = Women's Tennis Association

Anhang

Die Studien in der Übersicht

Kategorie	Studien
Kategorie I: Individual Player Performance Ranking/Rating	Ambrutis & Povilaitis, 2024 Blackburn, 2013 Broadie & Rendleman, 2013 Hvattum & Gelade , 2019 Irons et al., 2014 Kim & Jeon, 2019 Servien, 2022 Wolf et al., 2020
Kategorie II: Team Performance Ranking/Rating	Barrow et al., 2013 Chartier et al., 2011 Colley, 2002 Constantinou & Fenton ,2013 Duncan & Parece, 2024 Gill & Keating, 2009 Govan et al., 2009 Jeremic & Radojicic, 2010 Lowery et al., 2019 Mangan & Collins, 2016 Pradhan & Abdourazakou,2020 Steward, 2013 Wigness et al., 2010
Kategorie III: Matchmaking-Systeme	Gong et al.,2020 Myślak & Deja,2015 Sauer et al.,2024 Wang et al.,2024
Kategorie IV: Kombination aus Ranking- und Matchmaking-Systemen	Matthews et al.,2021 Miles et al.,2010 Raack et al.,2014 Reining & Horowitz,2019

Kategorie I: Individual Player Performance Ranking/Rating

Ambrutis & Povilaitis, 2024: Composite rating method: Application to European basketball leagues

Ambrutis & Povilaitis (2024) versuchen in ihrer Studie anhand der Composite-Rating-Method (CRM) die Leistung von Basketballspieler, Basketballspielerinnen und Teams zu bewerten. Es handelt sich dabei um ein duales Bewertungssystem: Anhand der Spielerleistungsbewertung (PPR - Player Performance Rating) wird die individuelle Leistung der Spieler:innen bewertet, während das Team Performance Rating (TPR) die

Gesamtleistung eines Teams einordnet. Weiters bilden der Player Game Quality Score (PQS) und der Team Quality Score (TQS) wichtige Komponenten des CRM. PQS stellt die individuelle Leistungen von Spieler:innen als numerische Zahl dar. Der TQS bewertet den Einfluss der Spieler:innen auf die Teamdynamik, demnach auch Aktionen, die nicht in den persönlichen Statistiken erfasst sind. Nach jedem Spiel werden die Bewertungen der Spieler:innen aufgrund ihrer Leistung aktualisiert. TPR und PPR werden additiv (iterativ) berechnet. Es werden demnach die Bewertungen von Spieler:innen oder Teams angepasst, indem ihre Leistungsindikatoren zu ihren Bewertungen vor dem Spiel addiert oder subtrahiert werden. Dadurch sollen die aktuellen Leistungen wiedergegeben werden. Die Autoren lassen etwa beim PPR den ELO-Algorithmus miteinfließen.

Figure 6 zeigt, wie ein Ranking der besten Spieler der Euroleague mit CRM aus der Saison 22/23 aussehen würde. Dies basiert besonders auf dem PPR und zieht sowohl offensive als auch defensive Beteiligungen in Betracht (Ambrutis & Povilaitis, 2024).

RANK	PLAYER NAME	PPR	TEAM	POSITION
1	Player D	2559	Team B	Center
2	Player E	2311	Team A	Power forward
3	Player C	2300	Team I	Center
4	Player F	2218	Team F	Center/power forward
5	Player G	2185	Team O	Center
6	Player H	2182	Team C	Center/power forward
7	Player I	2164	Team D	Center
8	Player J	2155	Team A	Center
9	Player K	2126	Team K	Center
10	Player L	2115	Team P	Center

Figure 6 Individuelles Spieler Ranking anhand vom PPR (Ambrutis & Povilaitis, 2024)

Blackburn, 2013: Ranking the performance of tennis players: An application to women's professional tennis

In der Studie von Blackburn (2013) wird ein Performance Ranking für weibliche Tennisspielerinnen anhand der WTA-Saison aus dem Jahre 2011 untersucht. Dabei wird das Ranking basierend auf der Winning-Percentage (Gewinnquote) beleuchtet. Es werden die Anzahl der gewonnenen Spiele durch die Gesamtanzahl der gespielten Spiele geteilt und anhand dieses Prozentsatzes werden die Spielerinnen bewertet.

Ein weiterer Ansatz von Blackburn (2013) ist das Ranking nach Paarvergleichen. Die Spielerinnen werden anhand ihrer Leistung in Spielen gegeneinander eingestuft und bewertet. Diese statistische Methode setzt sich aus drei Komponenten zusammen: die letzten Spielergebnisse, Qualität (Stärke) der Gegnerin und die Anzahl der gespielten Spiele einer Spielerin. Der Autor vergleicht in der Studie dieses Ranking mit dem offiziellen WTA-Ranking und es zeigen sich signifikante Unterschiede auf.

Broadie & Rendleman, 2013: Are the official world golf rankings biased?

Broadie & Rendleman (2013) beschreiben in ihrer Studie, wie sich das Official World Golf Ranking zusammensetzt. Die Leistung von Golfprofis wird in diesem Ranking anhand von Ergebnissen der letzten zwei Jahre (104 Wochen) dargestellt. Spieler:innen verdienen Punkte für die Teilnahme an Turnieren. Die Anzahl an Punkten variiert je nach der Endposition in einem Turnier. Ein:e Gewinner:in eines Majors wird mit 100 Punkten belohnt. Vom zweiten bis fünften Platz werden 60, 40, 30 und 24 Punkte respektive verteilt.

Die Ergebnisse der jüngsten 13 Wochen werden in diesem Ranking höher gewichtet. Die Punkte der restlichen 91 Wochen werden schrittweise reduziert. Die Gesamtpunkte werden durch die Anzahl der Turniere geteilt, an denen der Profi teilgenommen hat. Der Divisor erstreckt sich von 40 bis höchstens 52. Das aktuelle Ranking ergibt sich dann anhand der durchschnittlich bereinigten Punktezahl (Broadie & Rendleman, 2013).

Hvattum & Gelade, 2021: Comparing bottom-up and top-down ratings for individual soccer players

In der Studie von Hvattum & Gelade (2021) werden Bottom-up und Top-down-Ratings als Bewertung für Fußballspieler:innen verglichen (Figure 7). Als Beispiel für ein Bottom-up-Rating werden die „Value Added by Each Player“-Ratings herangezogen (VAEP). Diese Methode bewertet den Wert einzelner Aktionen auf dem Spielfeld. Jeder Aktion, die ein:e Spieler:in durchführt wird, ein Wert zugeordnet. Die Gesamtbewertung einer spielenden Person ist die Summe dieser Werte über einen bestimmten Zeitraum. Dieser Wert wird anhand von der Wahrscheinlichkeit geschätzt, mit welcher die Mannschaft des oder der bestimmten Spielers oder Spielerin innerhalb der nächsten zehn Aktionen ein Tor erzielen oder ein Gegentor kassieren wird.

Als Beispiel für ein Top-down-Rating wird das Plus-Minus-Rating (PM) angeführt. Bei diesem Konzept wird die Tordifferenz der Mannschaft einzelnen Spielern oder Spielerinnen zugeordnet. Es werden demnach Kennzahlen der Mannschaftsleistung auf einzelne Spieler:innen verteilt. Für verbesserte Genauigkeit wird unter anderem der Heimvorteil als variabler Faktor angesehen. Durch die PM-Ratings (siehe Figure 7) wird der Einfluss der Sportler:innen auf ein Spiel basierend darauf bewertet, wie das Team performt je nachdem ob die Person auf dem Feld steht oder nicht (Hvattum & Gelade, 2021).

Player	Country	PM	Player	Country	VAEP
1 Ederson	BRA	0.296	1 Lionel Messi	ARG	0.897
2 Thomas Müller	GER	0.269	2 Kylian Mbappe	FRA	0.687
3 Kyle A. Walker	ENG	0.264	3 Arjen Robben	NED	0.615
4 Thiago Alcantara	ESP	0.264	4 C. Ronaldo	POR	0.555
5 Lionel Messi	ARG	0.259	5 Franck Ribery	FRA	0.546
6 Neymar	BRA	0.258	6 Mohamed Salah	EGY	0.534
7 Bernardo Silva	POR	0.252	7 James Rodriguez	COL	0.519
8 Mohamed Salah	EGY	0.231	8 Neymar	BRA	0.513
9 David Alaba	AUT	0.230	9 Z. Misimovic	BIH	0.501
10 Raheem Sterling	ENG	0.223	10 Jadon Sancho	ENG	0.486
11 Danilo	BRA	0.220	11 Eden Hazard	BEL	0.485
12 Toni Kroos	GER	0.219	12 Carlos Tevez	ARG	0.477
13 Aymeric Laporte	FRA	0.215	13 Julio Alvarez	VEN	0.468
14 Sadio Mane	SEN	0.212	14 Kun Aguero	ARG	0.467
15 James Rodriguez	COL	0.211	15 Benito Raman	BEL	0.467
16 R. Lewandowski	POL	0.210	16 O. Dembele	FRA	0.453
17 Ilkay Gundogan	GER	0.207	17 Mohamed Zidan	EGY	0.452
18 Kun Aguero	ARG	0.207	18 Serge D. Gnabry	GER	0.452
19 Fabinho	BRA	0.207	19 Harry Kane	ENG	0.446
20 Jerome Boateng	GER	0.203	20 Dries Mertens	BEL	0.442

Figure 7 Die Bewertung von Fußballspielern anhand der PM (links) oder VAEP-Ratings (rechts) zum Stand Juli 2019 (Hvattum & Gelade, 2021).

Irons et al., 2014: Developing an improved tennis ranking system

Irons und Kollegen (2014) diskutieren ein Rankingsystem für Tennisspieler:innen. Das vorgeschlagene Modell basiert auf einem statistischen Modellierungsrahmen und soll eine genauere Darstellung der Leistung der Spieler:innen ermöglichen. Das System bezieht die Punktezahl, die jede spielende Person in einem Spiel gewonnen hat, mit ein. Dadurch soll eine detailliertere Bewertung der Spielerleistung möglich werden. Dies wird etwa beim ATP-Ranking nicht berücksichtigt. Bei der Modellierung wird eine Maximum-Likelihood-Schätzung verwendet, um für alle Personen eine Spielerbewertung abzuleiten. Berücksichtigt werden dabei Spiele, die vor einem bestimmten Tag gespielt wurden. In Betracht gezogen werden dabei die letzten beiden Jahre, in denen ein:e Spieler:in gespielt hat. Das Modell verwendet eine Wahrscheinlichkeitsfunktion, die auf dem Bradley-Terry-Modell basiert. Das verbesserte Rangsystem verwendet einen ausgeklügelten Ansatz zur Gewichtung der Ergebnisse, der detaillierte Daten zum Spiel - wie Ergebnisse in Sätzen oder Spielen - einbezieht und Flexibilität bei der Erfüllung verschiedener Einschränkungen und Anforderungen ermöglicht.

Kim & Jeon, 2019: Proposal for implementation of a ranking model for Olympic Taekwondo competitions using PageRank

In der Studie von Kim & Jeon (2019) soll der PageRank Algorithmus angewandt werden, um die Olympischen Taekwondo-Rankings zu berechnen. Dadurch soll die Leistung der männlichen Athleten genauer wiedergegeben werden. In der Studie wird dieser Ansatz auf die Gewichtsklasse -58kg der Männer angewandt. Der PageRank-Algorithmus wurde

ursprünglich von Google für das Ranking von Websites verwendet. In diesem Prozess kommen Markov-Ketten und Eigenvektorberechnungen zur Verwendung, um ein Ranking zu generieren. Webseiten werden dann anhand ihrer Zitationshäufigkeit bewertet. Ähnlich wird der Algorithmus auch im Taekwondo eingesetzt. Das Ranking anhand des Algorithmus ergibt sich durch das Erstellen einer Matrix, die die Ergebnisse von Kämpfen zwischen Athleten darstellt. Es wird eine relative Wichtigkeit der einzelnen Athleten ermittelt. Um diese Matrix zu bilden, wird die Leistung eines Athleten gegen die Anderer verglichen, wobei Sieg und Niederlage aufgezeichnet werden. Anhand dieser Matrix wird ein PageRank für jeden Sportler berechnet, der auch die Stärke des Gegners in Betrachtung zieht. Die Daten für eine Beispielsrechnung bezieht die Studie von den Ergebnissen des World Taekwondo Grand Prix.

Servien, 2022: An alternative ranking system for biathlon pursuit races

Servien (2022) diskutiert ein alternatives Rankingsystem für männliche Biathleten, um bei Verfolgungsrennen eine faire und dynamische Rangfolge bieten zu können. Bei Verfolgungsrennen kämpfen Athleten in Echtzeit gegeneinander und nicht gegen die Zeit. Das Ziel des alternativen Rankingsystems ist es, den letzten Rang der Biathleten darzustellen und den Einfluss der Sprintergebnisse zu minimieren. Das alternative Rankingsystem basiert auf einer quantil-basierten Berechnung, bei der für jeden Athleten eine Menge q_{k_i} berechnet wird. Diese Größe wird als Quantil der Verteilung der Endplatzierungen aus früheren Rennen behandelt. Um alle möglichen Rennszenarien zu berücksichtigen, verwendet das System verschiedene Daten von 148 früheren Verfolgungsrennen.

Wenn q_{k_i} gleich Null ist, hat kein Biathlet mit demselben Startrang in den letzten 148 Rennen eine bessere Endplatzierung erreicht. Dies deutet demnach auf eine sehr gute Leistung hin. Umgekehrt dazu würde q_{k_i} gleich Eins bedeuten, dass das aktuelle Ergebnis das Schlechteste aller vorherigen Ergebnisse für diesen Startrang ist (Servien, 2022).

Das alternative Rankingsystem soll die Fairness und Spannung erhöhen, weil Athleten, die bei Sprints meist unterdurchschnittliche Leistungen bringen, bei Verfolgungsrennen dennoch gute Ergebnisse liefern können. Diese Methode korreliert weniger mit dem Ranking zum Start und bietet größere Variabilität (Servien, 2022).

Wolf et al., 2021: A football player rating system

Wolf und Kollegen (2021) stellen in ihrer Studie ein Bewertungssystem für männliche Fußballspieler vor. Das Ziel liegt darin, die Spieler objektiv anhand deren Spielstärke zu bewerten. Das System basiert auf dem Elo-Algorithmus, wodurch die Spielerbewertungen nach jedem Spiel aktualisiert werden, indem die tatsächlichen Ergebnisse mit den

erwarteten Ergebnissen verglichen werden. Anhand der aktuellen Elo-Bewertung wird für jeden Spieler eine erwartete Erfolgschance im Vergleich zum Rating seines Gegenspielers berechnet. Für jeden Spieler wird dabei ein individueller Erwartungswert berechnet. Dadurch kann auch geschätzt werden, wie viel jeder einzelne Spieler zu den Gewinnchancen einer Mannschaft beitragen kann. Um einen Vergleich zwischen dem tatsächlichen und dem zu erwartenden Ergebnis zu ziehen, wird deren Differenz mit dem Faktor k multipliziert. Die Erfahrung, die Anzahl der gespielten Spiele oder das Alter des Spielers beeinflussen diesen Faktor. Der k -Wert von jüngeren Spielern ist höher, damit sich das Rating stärker auf aktuelle Leistungen anpasst.

Auch der Faktor des Heimvorteils spielt eine Rolle, er kann sich auf die Wertung einzelner Spieler auswirken. Dadurch kann es etwa bei Spielern der Heimmannschaft im Vergleich zu der Auswärtsmannschaft zu verschiedenen Wertänderungen kommen. Der Heimvorteil ist ein dynamisches Element in diesem Ratingsystem. Auch die Leistung der gesamten Mannschaft fließt in die Bewertung des Spielers mit ein. Wenn der Spieler ausgewechselt wird, hat dies Auswirkungen auf sein persönliches Rating (Wolf et al., 2021).

Um die Wirkung eines Spielers für ein Team differenzierter betrachten zu können, nimmt der Begriff des Players-impact an Bedeutung zu. Dieser Indikator gibt an welchen Einfluss der Spieler auf das Spiel der gesamten Mannschaft hat. So dürfte etwa Mohamed Salah im Jahr 2018 einen besonders positiven Einfluss auf das Spiel vom Verein FC Liverpool gehabt haben (Figure 8). Der Players-impact berechnet sich durch die Änderungswerte der Teamwertung mit und ohne den Spieler auf dem Feld. Die Spielzeit des Spielers fließt in dieses Ranking mit ein (Wolf et al., 2021).

Rank	Name	Current Club	Impact	Rating	Games	Club Rating
1.	MOHAMED SALAH	FC Liverpool	8.43	4858	156	4479
2.	KEVIN DE BRUYNE	Manchester City	6.98	4900	169	4840
3.	NICOLAS OTAMENDI	Manchester City	6.45	4897	155	4840
4.	RAHEEM STERLING	Manchester City	4.97	4879	165	4840
5.	SADIO MANE	FC Liverpool	4.75	4830	138	4779
6.	LIONEL MESSI	FC Barcelona	4.49	4960	180	4923
7.	GERARD PIQUÉ	FC Barcelona	4.23	4974	147	4923
8.	EDERSON	Manchester City	4.21	4854	84	4840
9.	DAVID SILVA	Manchester City	4.15	4896	146	4840
10.	EDINSON CAVANI	Paris St. Germain	4.13	4814	172	4772
11.	MARCO VERRATTI	Paris St. Germain	4.06	4842	127	4772
12.	ERIC BAILLY	Manchester United	4.04	4832	101	4751
13.	PAUL POGBA	Manchester United	3.99	4831	156	4751
14.	ANTOINE GRIEZMANN	Atletico Madrid	3.95	4849	191	4818
15.	DAVID ALBA	Bayern München	3.74	4959	136	4876

Figure 8 Die Top 15 Spieler nach dem Players-impact-Ranking; Stand Mai 2018 (Eigene Darstellung, adaptiert nach Wolf et al., 2021)

Kategorie II: Team-Performance-Ranking/Rating

Barrow et al., 2014: Ranking rankings: an empirical comparison of the predictive power of sports ranking methods

In der Studie von Barrow et al. (2014) werden verschiedene Rankingmethoden miteinander verglichen und die Zusammensetzung von Team-Rankings in verschiedenen amerikanischen Sportarten beschrieben.

In der NBA werden 82 Spiele in der regulären Saison gespielt, davon je 41 zu Hause und auswärts. Jedes Team spielt mindestens zweimal gegen jedes Team. Pro Spiel gibt es immer einen Sieger, da durch Verlängerungen Unentschieden ausgeschlossen werden. Das Ranking setzt sich anhand der Gewinnquoten der einzelnen Teams zusammen (Winning-Percentage) (Barrow et al., 2013).

Die Major League Baseball (MLB) wird in zwei große Ligen unterteilt: die American League (AL) und die National League (NL). Die Spiele erfolgen seit dem Jahr 1997 auch ligaübergreifend. Durchschnittlich werden 157 Spiele gespielt. Die Anzahl der Spiele wird durch Faktoren, wie Spielerstreiks oder Regen, beeinflusst. Auch in der MLB werden Teams anhand des Prozentsatzes der Spiele, die gewonnen wurden, bewertet (Barrow et al., 2013).

Im amerikanischen College Basketball (NCAAB) erfolgt das Ranking der Teams nach dem Rating-Percentage-Index (RPI). Der RPI berechnet sich aus drei Komponenten: die Gewinnquote einer Mannschaft, die Gewinnquote des Gegners und die Gewinnquoten des Gegners Gegner. Durch diese Methode soll die Stärke des Spielplans in das Ranking miteinfließen und Überblick über die Leistung eines Teams gegeben werden. Die NCAAB verwendet etwa den Home-adjusted-RPI, bei welchem die Heimspiele im Vergleich zu den Auswärtsspielen unterschiedlich gewichtet werden (Barrow et al., 2013).

Im amerikanischen College-Football (NCAAF) werden im Durchschnitt zehn Spiele pro Saison bestritten. Unentschieden gibt es aufgrund von Verlängerung nicht. Die Autoren beschreiben die Rankingmethode der NCAAF als Aggregation von mehreren Rankings. Dazu zählen Methoden, die auf Expertenmeinung basieren als auch computergenerierte Systeme (Barrow et al., 2013).

Chartier et al., 2011: Sports Ranking with Nonuniform Weighting

Chartier et al. (2011) wenden die Colley- und Massey-Methoden an, um Teams im amerikanischen College-Basketball zu ranken.

Die Colley-Methode basiert auf einem Ranking nach Gewinnwahrscheinlichkeit. Unter Anwendung der Laplace-Regel entsteht ein Rating durch das Einbeziehen der Anzahl der

Spiele und der daraus resultierenden Siege. Zu Beginn der Saison erhalten alle Teams eine Bewertung von 0,5, um unvoreingenommen in die Saison starten zu können. Mit einer linearen Gleichung wird die Berechnung vereinfacht und bündig dargestellt. Die Methode wird nicht durch Punktedifferenzen am Ende eines Spieles beeinflusst. Jeder Sieg oder jede Niederlage wird gleich bewertet (Chartier et al., 2011).

Im Vergleich dazu soll bei der Massey-Methode der Unterschied in den Ratings zweier Teams die Punktedifferenz im direkten Duell widerspiegeln. Auch diese Methode wird durch ein lineares System vereinfacht. Die rechte Seite dieser Gleichung gibt den Vektor p an. Der Wert spiegelt die kumulative Punktedifferenz wider: alle Punkte, die ein Team erzielen konnte abzüglich der Summe der von den Gegnern erzielten Punkte. Um Faktoren, wie die Wichtigkeit von Spielen, zu berücksichtigen, wird bei der Massey-Methode eine Gewichtsmatrix verwendet, um gewissen Spielen eine höhere Gewichtung zu verleihen. So versucht das System die aktuellen Spielergebnisse differenziert zu gewichten als den Ausgang älterer Spiele (Chartier et al., 2011).

Colley, 2002: Colley's bias free college football ranking method: The colley matrix explained.

Die Studie von Colley (2002) beschreibt die Colley-Matrix-Methode, durch welche amerikanische College-Football Teams anhand ihrer Leistung bewertet werden können. Sie basiert ausschließlich auf Siegen und Niederlagen, stellt aber auch die Stärke des Spielplans in den Fokus, weshalb sie als Erweiterung der Bewertungsmethode nach Winning-Percentage gesehen werden kann. Sie berücksichtigt allerdings nicht den Faktor der Gewinnspanne.

Die Colley-Methode stellt sich mathematisch mit einem linearen Gleichungssystem auf, wobei eine Matrix aufgestellt wird, in welcher die Siege und Niederlagen einzelner Mannschaften und deren Gegner dargestellt wird. Durch einen iterativen Prozess werden Teams zunächst anhand ihrer Sieg und Niederlagen Bilanz bewertet und fortlaufend wird diese Einordnung auf der Grundlage der Bewertung der Gegner angepasst. Folglich soll ein Ranking entstehen, durch das die Leistung der Teams genau widerspiegelt werden kann (Colley, 2002).

Constantinou & Fenton, 2013: Determining the level of ability of football teams by dynamic ratings based on the relative discrepancies in scores between adversaries.

In der Studie von Constantinou & Fenton (2013) wird das Pi-Ratingsystem zur Bewertung von Fußballteams vorgestellt. Jedes Team bekommt zu Beginn eine Leistungsbewertung von 0. Dieser Wert entspricht der durchschnittlichen Teambewertung verglichen mit anderen Teams. Wenn ein Team die Wertung n erhält, so verliert der Gegner n , um das

Gleichgewicht zu halten. Um die aktuellen Leistungen eines Teams zu erfassen, werden einige Faktoren, wie die Tordifferenz eines Teams, vom System berücksichtigt: Die Bewertungen ändern sich je nachdem ob ein Team zu Hause spielt oder auswärts.

Weiters werden die aktuellen Ergebnisse stärker gewichtet als vergangene, damit die Ratings die jüngsten Leistungen widerspiegeln. Dies wird anhand einer Lernrate λ ermittelt. Die Pi-Ratings verringern auch die Tordifferenzen exponentiell bevor die Ratings aktualisiert werden (Constantinou & Fenton, 2013).

Das System weist im Vergleich zum Elo-Ratingsystem noch genauere Leistungswerte auf, wodurch es zu einer praktikablen und effizienten Methode für die Bewertung von Teams wird (Constantinou & Fenton, 2013).

Duncan & Parece, 2024: Population-adjusted national rankings in the Olympics

Duncan & Parece (2024) haben das Ziel, die olympischen Medaillenrankings trotz unterschiedlicher Bevölkerungsgröße fairer und vergleichbarer zu gestalten. Das Population-adjusted-national-Ranking soll berücksichtigen, wie wahrscheinlich es ist, dass eine Nation aufgrund seiner Einwohnerzahl Medaillen gewinnen wird. Dieses System verwendet den Wahrscheinlichkeitsindex U . Das Wahrscheinlichkeitsmodell nimmt an, dass alle Länder pro Kopf dieselbe Aussicht auf Medaillen haben. Je kleiner die Rangwahrscheinlichkeit P_R ist, desto weniger wahrscheinlich ist ein Medaillengewinn, was für dieses Land auf eine sehr gute Leistung hindeuten würde. Dieses Ranking ermöglicht demnach bevölkerungsarmen Ländern im olympischen Medaillenranking einen hohen Platz einzunehmen, wenn die Athleten und Athletinnen im Verhältnis zur Einwohnerzahl ausgezeichnete Leistungen erbringen. Die Autoren stellen damit eine Möglichkeit vor,

Rank	Country	Population (millions)	Expected Medals	Medals Won	U Index
1	Australia	25.92	3.87	46	33.008
2	Great Britain	67.28	10.05	64	30.381
3	Netherlands	17.50	2.61	36	28.063
4	New Zealand	5.13	0.77	20	21.147
5	Hungary	9.71	1.45	20	15.878
6	USA	337.00	50.35	113	15.310
7	Italy	59.24	8.85	40	14.139
8	Japan	124.61	18.62	58	13.244
9	Cuba	11.26	1.68	15	9.475
10	Jamaica	2.83	0.42	9	9.117
11	Switzerland	8.69	1.30	13	8.889
12	France	64.53	9.64	33	8.712
13	Denmark	5.85	0.87	11	8.623
14	Canada	38.16	5.70	24	8.138
15	Germany	83.41	12.46	37	8.073
16	San Marino	0.03	0.01	3	7.674
17	Georgia	3.76	0.56	8	6.845
18	Croatia	4.06	0.61	8	6.594
19	Czechia	10.51	1.57	11	6.098
20	Norway	5.40	0.81	8	5.677
21	Serbia	7.30	1.09	9	5.667
22	Slovenia	2.12	0.32	5	4.697
23	Sweden	10.47	1.56	9	4.437
24	Ukraine	43.53	6.50	19	4.350
25	Republic of Korea	51.83	7.74	20	3.832

Figure 9 Nationales Medaillen Ranking nach dem Population-adjusted national-Ranking für Olympia 2020 in Tokyo (Duncan & Parece, 2024)

kleinen Ländern bei den Olympischen Spielen für herausragende Leistungen mehr Anerkennung zu bieten. Der Index U wird in diesem Modell als der negative Logarithmus der Rangwahrscheinlichkeit P_R definiert. Die Länder werden nach dem U -Index eingestuft. Desto größer der U -Wert desto besser ist die Leistung anzusehen (Duncan & Parece, 2024). Die Studie testet das Population-adjusted-national-Ranking anhand von Daten der letzten drei Olympischen Sommerspielen (2020 Tokyo, 2016 Rio de Janeiro, 2012 London; Anm. d. Autors: Studie wurde vor Beginn der Olympischen Sommerspiele 2024 in Paris veröffentlicht). Figure 9 zeigt das Nationale Medaillenranking der Olympischen Spiele 2020 in Tokyo anhand der vorgestellten Methode. Anhand dessen zeigt sich, dass auch verhältnismäßig kleine Länder wie Kuba, Jamaika oder Neuseeland, durch das angewandte System einen hohen Platz im Ranking einnehmen können und deren sehr gute Leistungen veranschaulicht wird (Duncan & Parece, 2024).

Gill & Keating, 2009: Assessing Methods for College Football Rankings

Gill & Keating (2009) behandeln in ihrer Studie verschiedene mathematische Methoden, um Teams im amerikanischen College Football zu ranken.

Mit der „weighted-least-squares“ Methode werden Spielen unterschiedliche Gewichtungen zugewiesen. Das Model legt die Wichtigkeit bestimmter Match-ups fest, die für die Gesamtrangliste von Entscheidung sein können. Die Gewichtung kann etwa auf den Gewinnquoten der beteiligten Teams basieren (Gill & Keating, 2009).

Mit den „least-square-methods“ gibt es eine statistische Methode für das Ranking von Teams, die die Summe der quadrierten Unterschiede zwischen den beobachteten und vorhergesagten Ergebnissen minimieren soll. Die Methode kann sich an teamspezifische Faktoren, wie den Heimvorteil, anpassen (Gill & Keating, 2009).

Mit sogenannten „Zensurmethoden“ wird versucht hohe Punkteunterschiede zwischen zwei Teams gegeneinander auszugleichen, weil diese Unterschiede nicht die relative Stärke dieser Teams widerspiegeln. Bei der Berechnung wird ein Schwellenwert festgelegt, wodurch sichergestellt werden soll, dass das Ranking nicht unverhältnismäßig stark durch die Endergebnisse von Spielen beeinflusst wird (Gill & Keating, 2009).

Die von den Autoren vorgestellten Modelle weisen alle einen Weg auf, um die Faktoren des Heimvorteils und die Stärke des Spielplans im Ranking zu berücksichtigen (Gill & Keating, 2009).

Govan et al., 2009: Offense-Defense Approach to Ranking Team Sports

Govan und Kollegen (2009) diskutieren das Offense-Defense-Modell (ODM) als systematischen Ansatz, um Teams im Leistungssport zu ranken. Dabei werden die

offensiven als auch die defensiven Stärken und Charakteristiken jeder Mannschaft miteinbezogen. Die offensive Stärke zeigt, dass ein Team in der Lage ist viele Punkte zu erzielen. Stärke in der Defensiv deutet auf eine starke Verteidigung hin, die es den Gegnern erschwert viele Tore oder Punkte zu erzielen.

Die offensiven und defensiven Fähigkeiten eines Teams werden in einer Matrix widergespiegelt. In einer Punktematrix wird jeder Punkt, den ein Team gegen ein zweites erzielt hat dargestellt. Das System verwendet ein iteratives Verfahren, um die Bewertungen regelmäßig zu aktualisieren. Die Bewertung wird je nach Leistung gegen einen Gegner angepasst. Ein Offensivrating wird etwa erhöht, wenn ein Team gegen eine stark bewertete Verteidigung viele Punkte erzielt. Umgekehrt sinkt das Defensivrating eines Teams, wenn eine schwache Offensive viele Punkte erzielen kann (Govan et al., 2009).

Nach einigen Iterationen sollte sich die relative Reihenfolge der Bewertung stabilisieren, ohne dass es zu einer Anpassung der Bewertungsmatrix kommen muss. Eine Score-Matrix ermöglicht das Normalisieren der Ergebnisse, damit die Offensiv- und Defensivratings mit allgemeinen Kennzahlen übereinstimmen. Als Ergebnis ergibt sich in der Folge eine stochastische Matrix, durch die die Stärken der verschiedenen Mannschaften widergespiegelt werden (Govan et al., 2009).

Jeremic & Radojicic, 2010: A New Approach in the Evaluation of Team Chess Championships Rankings

In der Studie von Jeremic & Radojicic (2010) werden Rankingmethoden für Schachteammeisterschaften diskutiert. Die präsentierten Methoden werden anhand der 2008 und 2009 gespielten Teammeisterschaften in Serbien examiniert. Weiters wird eine Rankingmethode, wie sie auch im Fußball angewandt wird, diskutiert und ein Vergleich gezogen.

Bei der I-Distanzmethode wird von einer metrischen Entfernung in einem n -dimensionalen Raum gesprochen. Anhand dieser Methode soll aus mehreren Variablen ein Indikator erstellt werden, der ein Ranking für Länder oder Teams vergeben kann. Berechnet wird folglich die I-Distanz zwischen zwei Objekten anhand von ausgewählten Objekten. Es handelt sich wieder um eine iterative Methode. Im Schach wird eine Entität als Referenz festgelegt. Das Ranking basiert folglich auf der Entfernung von dieser Entität, berechnet durch die I-Distanzmethode (Jeremic & Radojicic, 2010).

Bei dem Turnier in Serbien im Jahre 2008 haben Teams zwei Punkte für einen Sieg, einen für ein Unentschieden und keinen Punkt bei einer Niederlage erhalten. 2009 wurde das Rankingsystem aus dem Fußball, mit respektive drei, einen und null Punkten, implementiert. Das Elo-Ratingsystem findet auch hier seine Anwendung, weil nicht nur die

Ergebnisse, sondern auch die Stärke des Gegners miteinbezogen wurde (Jeremic & Radojicic, 2010).

Die Fußballrankingmethode wird von Jeremic & Radojicic (2010) als für den Schachsport als unpassend angesehen. Dies begründet sich besonders darin, weil ein Unentschieden häufig vorkommt und wichtig im Verlauf eines Turnieres sein kann. Die I-Distanzmethode bietet sich in diesem Falle eher als genauere und fairere Methode an. Die Tie-Breaks, die Anzahl der Siege und die Matchpoints sind Variablen, die bei der I-Distanzmethode in Betrachtung gezogen werden. Somit spiegelt diese Methode die verschiedensten Aspekte der Teamleistung wider. Im Jahre 2008 gibt es eine hohe Korrelation der I-Distanzmethode mit der endgültigen Wertung (Figure 10). Dies deutet darauf hin, dass die Mannschaftsleistung genauer dargestellt werden konnte.

Rk.	Team	I-distance	Rk. I-distance
1	NSK DDOR NOVI SAD	0.796103	1
2	SK BEOGRAD BEOPUBLIKU	0.243837	3
3	VSK SVETI NIKOLAJ	0.252295	2
4	SK SPARTAK	0.016979	5
5	SK SLOGA AMIGA	0.237741	4
6	SK RADNICKI	-0.027137	6
7	SK NIS	-0.212748	8
8	SK BEOGRADSKI VODOVOD	-0.189925	7
9	SK LASTA	-0.334153	10
10	SK STARA PAZOVA	-0.256229	9
11	SK MLADENOVAC	-0.526763	11

Figure 10 Die Klassifikation der 2008 Schach Teammeisterschaften in Serbien mit der I-Distanz Methode im Vergleich zur Endplatzierung (Jeremic & Radojicic, 2009)

Lowery et al., 2020: Functional ratings in sports

Lowery et al. (2020) zielen darauf ab, eine Methode für das Ranking von Sportteams zu präsentieren. Das Modell wird anhand der Saison 2018/2019 des NCAA-Basketball getestet. Das System soll neben den endgültigen Spielergebnissen auch Punkteunterschiede während eines Spiels erfassen. Anhand von funktionalen Bewertungen kann dies Aufschluss darüber geben, wann ein Team etwa stark ist oder Schwächephasen aufweist. Dies kann einer detaillierten Spielanalyse dienen. Die funktionalen Ratings werden von dem Autorenteam als Kombination der durchschnittlichen Punktedifferenz einer Mannschaft und der Stärke des Spielplanes definiert.

Die Bewertung von Teams funktioniert anhand der Umwandlung von funktionalen Bewertungen in Skalarwerte. Mit einer Gewichtsfunktion werden verschiedene Teile des

Spiele hervorgehoben. Das System berücksichtigt auch den Heimvorteil. Dieser ist zwar statistisch signifikant, variiert aber nicht zwischen den Teams (Lowery et al., 2020).

Mangan & Collins, 2016: A Rating System For Gaelic Football Teams: Factors That Influence Success.

Mangan & Collins (2016) beschreiben in ihrer Studie ein Ratingsystem für Gaelic-Football-Teams, das auf einer modifizierten Variante des Elo-Ratingsystems basiert. Die Studie arbeitet mit Daten aus sechs Saisonen (2010-2015) von Spielen aus der National Football-League oder All-Ireland-Football-Championship.

In der modifizierten Elo-Formel wird der Faktor c auf 10 und der Faktor d auf 400 festgelegt. Der dritte Faktor, der festgelegt werden muss, ist der k -Faktor. Durch diesen wird bestimmt, wie sich das Rating einer Mannschaft nach einer Partie ändert. Da im Gaelic-Football wenig Spiele gespielt werden, wird ein höherer k -Faktor angewandt, um schnellere Anpassungen an das Rating zu ermöglichen. Weiters werden verschiedene Spiele und Wettbewerbe unterschiedlich stark gewichtet (Bsp.: All-Ireland-Championship 100 Punkte; National Football-League 50 Punkte). Dadurch kann das System die Bedeutung der verschiedenen Wettkämpfe und die verschiedenen körperlichen Anforderungen widerspiegeln (Mangan & Collins, 2016).

Auch die Gewinnspanne (Margin of Victory) ist ein entscheidender Faktor in der Elo-Formel. Dadurch soll sichergestellt werden, dass große Punkteunterschiede auch dementsprechend in die Wertung einfließen. Der Multiplikator dafür setzt sich aus dem Verhältnis von erzielten Punkten zu erhaltenen Punkten zusammen. Dieser darf maximal den Faktor 2,5 betragen, um zu große Änderungen in den Ratings zu verhindern (Mangan & Collins, 2016).

Der Heimvorteil im Gaelic Football wurde von Mangan & Collins (2016) analysiert und schließlich anhand von Elo-Wertungspunkten quantifiziert. Der Heimvorteil wurde auf 79 Elo-Wertungspunkte berechnet. Dadurch kann ein Heimteam bei der Kalkulation des erwarteten Ergebnisses einer Partie die eigene Bewertung um 79 Punkte erhöhen. Der Faktor des Heimvorteils ist folglich für die Spielvorhersage von Bedeutung.

Pradhan & Abdourazakou, 2020: “Power ranking” professional circuit eSports teams using multi-criteria decision-making (MCDM)

Pradhan & Abdourazakou (2020) behandelt ein Power-Rankingsystem für den E-sports Bereich mit besonderem Fokus auf das Spiel „Dota 2“ (Entwickler: Valve, USA 2013). Anhand des Power-Rankings sollen Teams gemäß ihrer Stärke bewertet und mit anderen Teams verglichen werden. Um die Teamleistung zu bewerten und einordnen zu können,

verwendet das Autorenteam ein hybrides Modell für Multi-criteria-decision-making (MCDM; Entscheidungsfindung mit mehreren Kriterien).

Dabei werden verschiedene Techniken angewandt. Mit der Grey-Relational-Analysis-Methode (GRA) werden Daten normalisiert und die Beziehung zwischen verschiedenen Leistungskriterien bewertet. Mit dem simple-additive-Weighting (SAW) wird eine gewichtete Summe von Leistungsindikatoren berechnet, um Teams bewerten zu können (Pradhan & Abdourazakou, 2020).

Um folglich den Prozess des Power-Rankings zu vollziehen, verwenden Pradhan & Abdourazakou (2020) das hybride MCDM-Modell, um für jedes Team eine numerische Punktzahl zu berechnen, um ein Ranking zu erstellen. Weiters wichtig ist im Verlauf noch die Datenerfassung von Leistungsdaten aus verschiedenen Turnieren und Spielen, die Identifikation von relevanten Leistungsdaten und die Zuweisung von Gewichtungen für Kriterien, die die Bedeutung des Rankings widerspiegeln. Bezogen auf „Dota 2“ werden etwa der Prozentsatz der gewonnenen Spiele, die Anzahl der gespielten Spiele, die Anzahl der Treffer („Kills“) oder die gesammelten Punkte als Leistungsindikatoren definiert. Die Studie vergleicht das Power-Ranking auch mit herkömmlichen Bewertungsmethoden wie den Elo- oder Glicko-Ratings. Die Autoren beschreiben das MCDM-Modell als valide und fügen hinzu, dass es keine signifikanten Unterschiede im Ranking im Vergleich mit den Elo- und Glicko-Ratings gegeben hat (Pradhan & Abdourazakou, 2020).

Stewart, 2013: A proposed general rating system for which the Colley Matrix Rating System is a special case

Stewart (2013) beschreibt in seiner Studie, wie Mannschaften in Teamsportarten anhand des Colley-Matrix-Ratingsystem (CMRS) bewertet werden können und dadurch ausgeglichene Match-ups geschaffen werden können.

Das CMRS hätte die Möglichkeit Spielstatistiken, wie die Punktedifferenzen, in die Bewertung einfließen zu lassen, jedoch konzentriert sich die in dieser Studie behandelte Variante des Modells nur auf Siege und Niederlagen. Das System benötigt keine Information über die Regeln des Spiels oder welche Sportart gespielt wird. Wichtige Informationen sind allerdings die Anzahl der gespielten Spiele pro Team und der Ausgang der jeweiligen Spiele, insbesondere wer wen besiegt hat (Stewart, 2013).

Das System verleiht jedem Team einen Bewertungswert r . Dieser Wert stellt die Chancen eines Teams dar, ein absolut durchschnittliches Team zu schlagen. Ein durchschnittliches Team hätte einen r -Wert von 0,5, während ein perfektes Team einen r -Wert von 1 zugewiesen bekommt. Anhand der Laplace-Regel wird die Bewertung eines Teams berechnet, wodurch die Anzahl der Siege an die Stärke des Spielplans angepasst wird.

Siege gegen besser bewertete Teams werden höher bewertet und als effektive Siege betitelt. Somit fließt die Stärke des Gegners in das Rating mit ein (Stewart, 2013).

Wigness et al., 2010: A New Iterative Method for Ranking College Football Teams

Wigness et al. (2010) stellen eine iterative Rankingmethode zum Ranking von College-Football-Teams. Die Methode setzt sich zunächst aus einem Initialranking und dann aus dem eigentlich iterativen Ranking zusammen. Das initiale Ranking bildet die Grundlage anhand von Ergebnissen und Gewichtungen der Spiele. Im Folgenden wird für jede Iteration das Ranking der vorigen Iteration verwendet, um die Spielergebnisse stetig anzupassen.

Die Spiele werden aufgrund von verschiedenen Faktoren unterschiedlich gewichtet. Dabei wird in Betracht gezogen, ob es schon ein direktes Match-up zwischen zwei Teams gegeben hat, wo das Spiel stattgefunden hat, der Punkteunterschied am Ende des Spiels und die Zeit in der Saison, zu welcher die Partie ausgetragen wurde. Auch die Stärke des Spielplanes wird in das Iterative Ranking mit einbezogen. Der Algorithmus „lernt“ aus früheren Spielzeiten und der aktuellen Saison. Dadurch soll die Genauigkeit mit der Zeit erhöht werden (Wigness et al., 2010).

Kategorie III: Matchmaking-Systeme

Gong et al., 2020: OptMatch: Optimized Matchmaking via Modeling the High-Order Interactions on the Arena

Gong und Kollegen (2020) diskutieren in ihrer Studie das Matchmaking in Onlinespielen und präsentieren ein System, mit dem Spieler:innen in „Player-versus-Player“-Spiele (PvP) effektiver eingeteilt werden sollen. Mit dem OptMatch-System sollen ausgewogene Teams gebildet werden, die für einen spannenden und fairen Wettbewerb sorgen. Relevant im Kontext des E-Sport ist die Auswahl des Helden oder der Heldin, mit dem der Spieler oder die Spielerin in dem Spiel antritt. Jeder dieser Helden und Heldinnen (Spielcharaktere) hat ein unterschiedliches Stärke- und Schwächeprofil.

Das OptMatch-System ist ein datengesteuertes Matchmaking-Framework, das auf zwei Stufen basiert: dem Offline-Learning und Online-Planning. Das Offline-Learning zielt darauf ab, früherer Spieldaten zu analysieren, um ein Verständnis für Interaktionen und Beziehungen zwischen den Spielenden zu entwickeln. Durch das Erlernen von niedrig dimensionalen Repräsentationen von Helden und Heldinnen, wird versucht zu verstehen, wie verschiedene Charaktere in einem Team miteinander agieren und interagieren. Mithilfe von Datenanalysen und maschinellen Lernens sollen Spieler- und Teamdynamiken besser verstanden werden, was folglich eine Grundlage für die darauffolgende Online-Planning-Phase bildet (Gong et al., 2020).

In der zweiten Phase des OptMatch System, dem Online-Planning, liegt das Ziel darin, für faire und ausgeglichene Match-ups zu sorgen und damit auch die Spielerzufriedenheit zu garantieren. Dabei muss das System innerhalb weniger Sekunden unter dem Druck tausender Spieleranfragen Lösungen finden und Entscheidungen treffen. Durch eine heuristische Methode werden Spieler:innen auf der Basis ihrer Fähigkeiten in mehrere Einheiten aufgeteilt. Dadurch kommt es zu einer organisierten Aufteilung von den spielenden Personen in entsprechende Gruppen. Das Modell verwendet folglich ein Prognosemodell, um für alle Spieler:innen einen Nutzen vorzusagen und trifft darauf dementsprechende Entscheidungen (Gong et al., 2020).

Myślak & Deja, 2015: Developing Game-Structure Sensitive Matchmaking System for Massive-Multiplayer Online Games

Myślak & Deja (2015) beschäftigen sich mit dem Matchmaking in Onlinespielen. Besonderer Fokus liegt in dieser Studie auf dem Spiel League of Legends (kurz „LOL“) (Riot Games, USA 2009), anhand welchem das Matchmaking-System getestet wurde.

Um in „LOL“ in ein Spiel einzusteigen, müssen die Spieler:innen zunächst in eine Warteschlange. Die Spieler:innen wählen dort eine heldenhafte Figur aus, mit der sie spielen wollen. Spieler:innen, die einen höheren Rang im Spiel haben, dürfen zuerst wählen, während niedriger gerankte Teilnehmer:innen meist eher unbeliebtere Charaktere auswählen müssen (Myślak & Deja, 2015).

Um dieser Problemstellung entgegenzuwirken, sollen im vorgeschlagenen System, ein rollenbasiertes Matchmaking die Fähigkeiten der Spieler:innen als Vektor von Fähigkeitsschätzungen für jede Position (Rolle) darstellen. Es soll dadurch alle Spieler:innen anhand ihrer Fähigkeiten in einer gewissen Position zu spielen bewertet werden, um ein ausgeglicheneres Matchmaking zu gestalten. Das System versucht folglich fünf Personen in ein Team zu paaren, deren präferierten Positionen sich nicht überschneiden. Dadurch sollen, gemessen an den Fähigkeiten jedes Spielers und jeder Spielerin, ein homogenes Team entstehen. Die Möglichkeit, den gewünschten Charakter („Champion“) auszuwählen, soll aber auch in diesem System weiter erhalten bleiben (Myślak & Deja, 2015).

Myślak & Deja (2015) merken an, dass dieses System auch in andere Onlinespiele, wie „Dota 2“ oder „Smite“ (Hi-Rez Studios, USA 2014), implementiert werden kann. Die Voraussetzungen dafür sind eine bereits definierte Meta und ein vorhandenes Ratingsystem für Spielerfähigkeiten.

Sauer et al., 2024: Improving ranking quality and fairness in Swiss-system chess tournaments

Sauer et al. (2024) stellen das Paarungssystem der FIDE (Internationaler Schachverband) in Frage und schlagen einen neuen Ansatz für die Paarung von Schachspielern und Schachspielerinnen nach dem schweizerischen System vor. Dabei werden bestehende Paarungssysteme wie Dutch, Monrad, Burnstein oder Random2⁵ in das System implementiert. Das Autorenteam strebt dabei nach einem System, dass versichert, dass Spieler:innen mit ähnlichen Stärken gegeneinander spielen.

Der Ansatz modelliert Faktoren wie die Farbbilanzen und die Punktegruppen anhand von Gewichtungen, die durch „Edges“ (Kanten) in einem Diagramm dargestellt werden. Solche Kantengewichte sind Zahlen, die jenen Kanten zugewiesen werden, um die Wichtigkeit der Paarung zu definieren. Höhere Gewichtungen stehen für bessere Paarungen. Diese Gewichtungen werden auch auf Grundlage von Elo-Ratings oder Punktezahlen erstellt, mit dem Ziel faire Paarungen zu generieren. Dies subsummiert sich unter dem Begriff Maxium-Weight-Matching (MWM) (Sauer et al., 2024).

Bezüglich der Fairness des Matchmaking-Systems ziehen Sauer et al. (2024) den Vergleich zu bestehenden Systemen, wie Burnstein oder Random2. In experimentellen Versuchen hat sich gezeigt, dass das MWM die bestehenden Systeme in gewissen Aspekten der Fairness teilweise sogar übertrifft. Gemessen wird die Fairness dabei an der Ausgewogenheit der Farbverteilung und an der Punktegleichheit der Gegner. Die Spieler:innen werden anhand ihres „Scores“ (Punkte) bewertet und sollten anhand dessen in gleiche Gruppen eingeteilt werden (Punktegleichheit). Durch den Punkt der Farbbalance soll sichergestellt werden, dass alle Spieler:innen eine ähnliche Anzahl an Partien in schwarz und weiß absolviert (Sauer et al., 2020).

Wang et al., 2024: EnMatch: Matchmaking for Better Player Engagement via Neural Combinatorial Optimization

Wang et al. (2024) präsentieren in ihrer Studie ein neues Matchmaking-Framework, das Teamdynamik und individuelle Spielereigenschaften berücksichtigt. Das sogenannte Engagement-oriented-Matchmaking-Framework (EnMatch) soll demnach das Spielerengagement erhöhen und das Spielerlebnis verbessern. Das System richtet sich an *k*-vs-*k* Szenarien, also an Teamzusammenstellungen mit zwei oder mehr Mitgliedern pro

⁵ Die wichtigsten Paarungssysteme im Schach. Dutch ist dabei der Standard in großen Schach-Turnieren. Spieler:innen werden immer anhand ihres Elo-Ratings in ein Ranking eingeteilt. Im Dutch System wird die obere Hälfte gegen die untere Hälfte eines Rankings gepaart (Sauer et al., 2020).

Mannschaft. Das Modell setzt sich aus fünf Komponenten zusammen: Encoder, Masked-Decoder, Matching-Pool, Heuristic-Operator und Engagement-Model.

Der Encoder stellt eine wichtige Komponente dar, weil dadurch alle möglichen Interaktionen zwischen den spielenden Personen mit diversen Fähigkeiten berücksichtigt werden und folglich die Dynamiken zwischen diesen Interaktionen verständlich werden. Dieser Schritt ist demnach für die Bildung ausgewogener Teams unabdingbar. In der Folge stellt der Masked-Decoder autoregressiv Teams zusammen, indem er Spielende anhand der durch den Encoder bereitgestellten Repräsentationen selektiert (Wang et al., 2024).

Der Matching-Pool ist die Menge an Personen, die darauf wartet in ein Team zugelost zu werden. Sobald ein:e Spieler:in in ein Team gelost wird, wird diese Person aus dem Pool entfernt. Es handelt sich demnach um eine stetig dynamische Komponente (Wang et al., 2024).

Das Engagement-Model dient zur Vorhersage des Spielerengagement und legt dadurch eine weitere Grundlage für die Zusammenstellung von Teams. Diese Vorhersage soll aufgrund der Leistungen einer spielenden Person in einem Spiel getroffen und quantifiziert werden. Außerdem bezieht es mit ein, wie sich unterschiedliche Teamzusammenstellungen auf das Spielerengagement auswirken können (Wang et al., 2024).

Durch den Heuristic-Operator soll nach optimalen Teamzusammenstellungen gesucht werden, wodurch wiederum auch das Spielerengagement gesteigert werden soll. Durch heuristische Methoden soll der Matchmaking-Prozess sowohl effizient als auch effektiv ausgestaltet werden. Wang et al. (2024) legen dabei das Tier-based-heuristic Matchmaking als Basis für das EnMatch-System fest.

Kategorie IV: Kombination aus Ranking- und Matchmaking-Systemen

Matthews et al., 2021: Application of PageRank Algorithm to Division I NCAA men's basketball as bracket formation and outcome predictive utility

Matthews et al. (2021) verwenden den PageRank-Algorithmus, um Teams im amerikanischen College-Basketball zu ranken und damit folglich die Setzliste für Turniere zu beeinflussen.

Die Basketballteams und ihre Spiele werden in einem Diagramm dargestellt (Figure 11). Dabei werden die Knotenpunkte für die Teams und die Spiele als Links (Verbindungen) zwischen den Knotenpunkten dargestellt. Der Rang jeder Mannschaft wird durch die Anzahl und die Qualität der Links berechnet. Jedes Team besitzt einen PageRank, der durch den PageRank eines anderen Teams beeinflusst wird, wenn es besiegt wird. Dadurch soll die Qualität von Siegen dargestellt werden. Der PageRank repräsentiert in diesem Diagramm

demnach mehrere Basketballteams sowie deren Anzahl an Niederlagen und Siegen gegeneinander (Matthews et al., 2021).

Der Algorithmus funktioniert mit einem iterativen Prozess, der auf einem Zufallsprinzip aufgebaut ist (Random Walk). Es wird eine Wahrscheinlichkeitsverteilung erstellt, die die relative Stärke der Teams angibt (Matthews et al., 2021).

Die Kanten (Links) im Diagramm können beliebig gewichtet werden, um etwa wichtigere Spiele hervorzuheben. Dies betrifft unter anderem Spiele, die auswärts gewonnen wurden oder Siege mit großen Punkteunterschieden (Matthews et al., 2021).

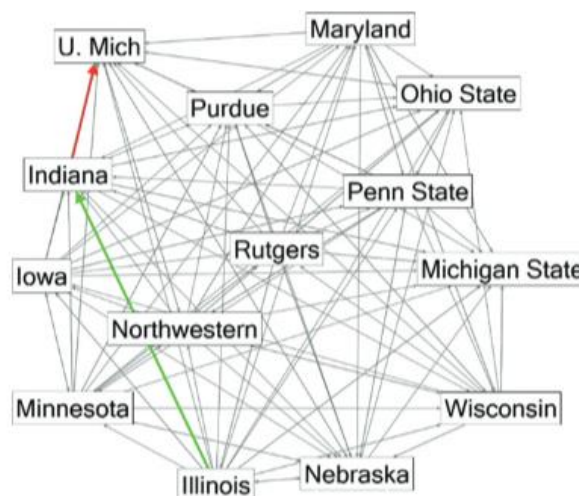


Figure 11 Das PageRank Diagramm (Matthews et al., 2021)

Die Setzliste für das nationale NCAA-Basketballturnier (z.B.: March Madness) wird einerseits durch die Gewinner der respektiven regionalen Turniere (Conference Winner) und andererseits durch ein Komitee festgelegt, deren Entscheidungen oft undurchsichtig sind und teilweise subjektiv getroffen werden. Matthews et al. (2021) wenden den PageRank-Algorithmus auf das Turnier im Jahre 2018 an, woraus sich folglich ein Teilnehmerfeld aus den 64 besten Colleteams ergab. Von diesen 64 Teams wurden 41 ebenso vom Auswahlkomitee bestimmt. Laut der Studie kann das Ranking und folglich die Bestimmung der Setzliste bei NCAA-Turnieren anhand des PageRank-Algorithmus für eine unvoreingenommene und demnach fairere Entscheidungsfindung in Betracht gezogen werden. Die Mannschaften würden folglich anhand ihrer guten Leistungen über die Saison verteilt für das wichtige Turnier ausgewählt.

Miles et al., 2010: AccuV College Football Ranking Model

Miles et al. (2010) diskutieren ein Model zum Ranking von Teams im amerikanischen College-Football. Das präsentierte MinV-Modell soll auch im Prozess zur Erstellung einer Setzliste für kommende Meisterschaftsspiele herangezogen werden.

Miles et al. (2010) stellen als Basismodel das MinV-Modell vor, durch welches Verstöße gegen den Spielstand (GSVs - Game-Score-Violations) minimiert werden sollen. Eine GSV tritt beispielsweise dann auf, wenn Team A ein Team B während der Saison besiegt hat, jedoch Team B im Ranking vor Team A steht. Das Problem der GSVs wird als mixed-integer-problem mit gemischten binären Zahlen formuliert. Die Stärke eines jeden Teams wird als Leistungsindex P_j durch eine kontinuierliche Variable angegeben. Desto höher der Leistungsindex, desto stärker das Team. Die Variable V_i wird verwendet, um ein GSV anzugeben. Wenn V_i gleich eins, dann hat der Leistungsindex des Siegerteams den des verlierenden Teams nicht um eine bestimmte Menge überschritten. Wenn kein GSV vorliegt, ist der Wert gleich null. Das Ziel des MinV Modell ist es die GSVs zu minimieren.

In weiterer Folge stellen Miles et al. (2010) das AccuV-Model vor, welches eine Erweiterung des bestehenden MinV-Systems darstellt. Das verbesserte Modell enthält ein paar Komponenten, durch die Effektivität und die Genauigkeit gesteigert werden sollen. So soll das AccuV-Modell unter anderem die Stärke des Spielplans, die Anzahl der Niederlagen, den Gewinnvorsprung (Margin of Victory), die Winning-Percentage als auch den Heimvorteil berücksichtigen. Der Vorteil, den Teams haben, wenn sie zu Hause spielen, soll in das Ranking mit einfließen.

Die Setzliste von NCAA-Football-Turnieren wird durch eine Kombination aus Computerrankings und Meinungsumfragen bestimmt. Dadurch wird bestimmt, welche zwei Teams um die Meisterschaft spielen werden, wobei es in der Folge um Geldsummen in Millionenhöhe geht. Die Studie kritisiert dieses System und versucht anhand des AccuV-Modells die GSVs auszugleichen damit jene Teams in der BCS-Bowl gesetzt werden, die über die Saison hinweg die besten Leistungen gezeigt haben (Miles et al., 2010).

Raack et al., 2014: Standings in sports competitions using integer programming

Raack und Kollegen (2014) beschreiben ein Rankingsystem anhand von Integer-Programming, durch das die Leistungen von verschiedenen Teams in diversen Sportarten bewertet werden kann. Die Studie wendet das Modell auf die deutsche Fußballbundesliga, die Formel 1 und die National Hockey League (NHL) an.

Bewertet werden die Teams anhand der Leistungen, die sie im Laufe einer Saison erreichen. Am Beispiel der deutschen Fußballbundesliga wären das drei Punkte bei einem Sieg, ein Punkt bei einem Unentschieden und keinem Zähler bei Niederlagen. Jedem Team

wird folglich ein ordinaler Rang zugeordnet, basierend auf der Gesamtpunktezahl. Anhand von Integer-Programming kann das Modell berechnen, wie viele Punkte benötigt werden, um einen bestimmten Rang zu garantieren (Raack et al., 2014).

Somit kann das Modell als Anhaltspunkt verwendet werden, um zu sehen wie viele Punkte ein Team benötigt, um in der kommenden Saison an großen Turnieren teilzunehmen oder den Abstieg in eine niedrigere Liga zu vermeiden (Raack et al., 2014).

In der NHL bekommen Teams für einen Sieg zwei Punkte, bei einer Niederlage nach Verlängerung einen Punkt und bei einer Niederlage nach der regulären Spielzeit null Punkte. Jedes Team spielt pro Saison 82 Spiele, wobei die Liga in zwei Konferenzen (Westen und Osten) aufgeteilt ist. Die 16 besten Teams von allen 30 NHL-Teams (acht je Konferenz) kommen in die Play-offs und spielen um die Meisterschaft. Raack et al. (2014) wenden ihr Modell an, um zu berechnen, wie viele Punkte ein Team mindestens benötigt, um einen solchen Play-off-Platz garantiert zu bekommen. Das Modell schätzt diesen Wert auf 149.

Reinig & Horowitz, 2019: Analyzing the impact of the NCAA Selection committee's new quadrant system

Reinig & Horowitz (2019) präsentieren das von der NCAA 2018 eingeführt Vier-Quadranten System, um College-Basketballteams zu bewerten und den Prozess der Setzliste für das nationale NCAA Division I Turnier zu verbessern.

Im NCAA-Basketball werden Teams anhand des Ratings-Percentage-Index (RPI) bewertet. Dieser setzt sich aus mehreren Komponenten zusammen: der Winning-Percentage, der Gewinnquote des Gegners und der Gewinnquoten des Gegners Gegner. Mit dieser Kennzahl werden Teams nicht nur anhand ihrer Sieg und-Niederlagenstatistik bewertet, sondern auch an der Qualität ihrer Gegner (Reinig & Horowitz, 2019).

Das von Reinig & Horowitz (2019) beschriebene Vier-Quadranten-System besteht aus vier Vierteln (Q1-Q4), wodurch die NCAA die Teamleistung bewertet. Die Siege und Niederlagen, die in Q1 eingestuft werden, werden am wertvollsten angesehen. Dabei spielt der RPI-Rang des Gegners eine Rolle und, ob das Spiel zu Hause oder auswärts stattfindet. Durch dieses System versucht die NCAA den Faktor des Heimvorteils auszugleichen, weil Auswärtssiege grundsätzlich als wertvoller bewertet werden.

Die Setzliste für das NCAA-Turnier wird bestimmt durch die Sieger der einzelnen Konferenzen sowie durch ein Auswahlkomitee, welches bestimmt welche Teams an dem prestigeträchtigen Turnier teilnehmen dürfen. Durch die Einführung des Vier-Quadranten Systems soll die Entscheidung des Komitees vereinfacht werden, indem Teams anhand der Qualität ihrer Siege bewertet werden. Dadurch soll der Auswahlprozess für das Turnier

verbessert werden, um das Turnierfeld ausgeglichen zu gestalten (Reinig & Horowitz, 2019).