Some fun end topics:
(1) Probabilistic sketching: Bloom filters, MinHash, HLL
(2) NP-hardness of video games
(3) NP-hardness of Olympic routine construction

December 6, 2024

15-351/15-650/02-613

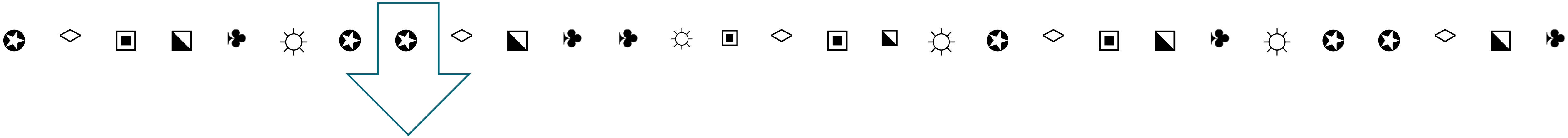
Yun <u>William</u> Yu, Carnegie Mellon University

SKETCHING ALGORITHMS

- Sub-linear space algorithms
- Family of algorithms for representing big data as small probabilistic data structures called "sketches"
- Fast accurate estimates of cardinality, quantiles, frequency distributions, set membership, majority element, etc.
- Widely used: routers, databases, search, etc.

SKETCHES COMPOSE

Data stream



0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

Property A
(e.g. hypertension)

and/or

1	1	0	0
0	0	1	0
0	0	1	0
1	0	0	1

Property B
(e.g. diabetes)

→

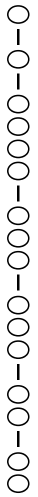
1	1	1	0
0	1	0	0
0	1	1	0
1	0	1	1

Property A+B
(diabetes + hypertension)

Sketches are compressed summaries that can be computed on.

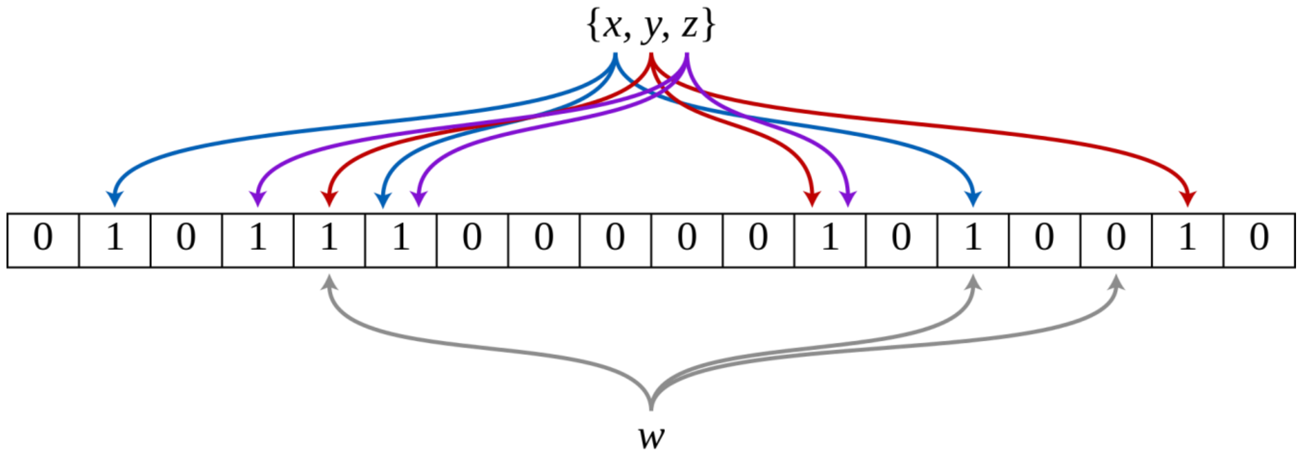
SET MEMBERSHIP PROBLEM

- Given a stream of items, I want to know if a particular item was seen.

Items	Hashes	Truncated Hashes	Bitstring
♣	d41d8cd98f00b204e9800998ecf8427e	d4	
◪	4b0538914e0d1beb795b4f7f1cf6aab2	4b	
▣	734e1e2f75efdec7f1a2fc2966009d0c	73	
◊	262513bfc6076d88414d2118ae98a1a7	26	
✪	337c8770ab0f31871d0769b59161df61	33	
☼	f14d43f49a3b852fedc5c7042a25f910	f1	
Original space	[#items]*[hash size] (n*32 bytes)	[#items]*[hash size] (n*1 bytes)	[# hashed values] (256 bits) = 8 bytes

BLOOM FILTERS

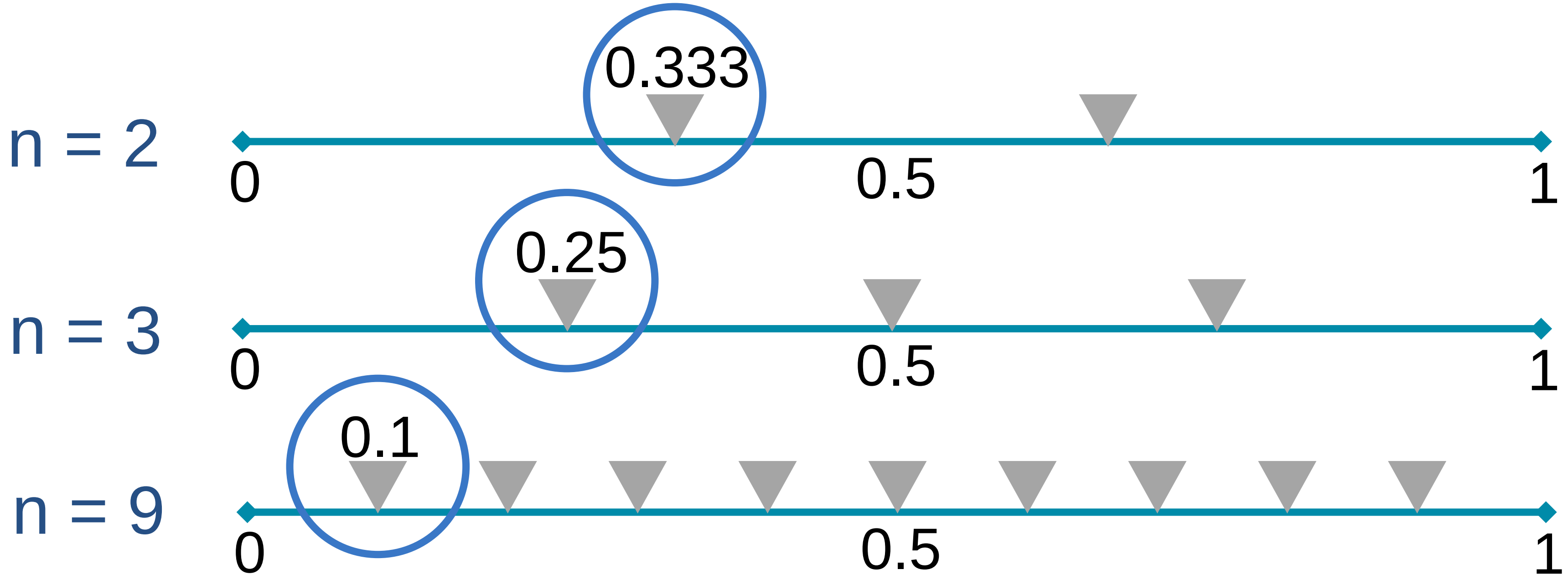
- Modified bitstring, where we use multiple hash functions
- When an item is inserted, set all of the hashed bits
- To query, see if all of the hashed locations are set to 1
- False positives are still possible, but less likely, given the right choice of number of hash fucntions.



https://en.wikipedia.org/wiki/Bloom_filter

COUNT DISTINCT PROBLEM

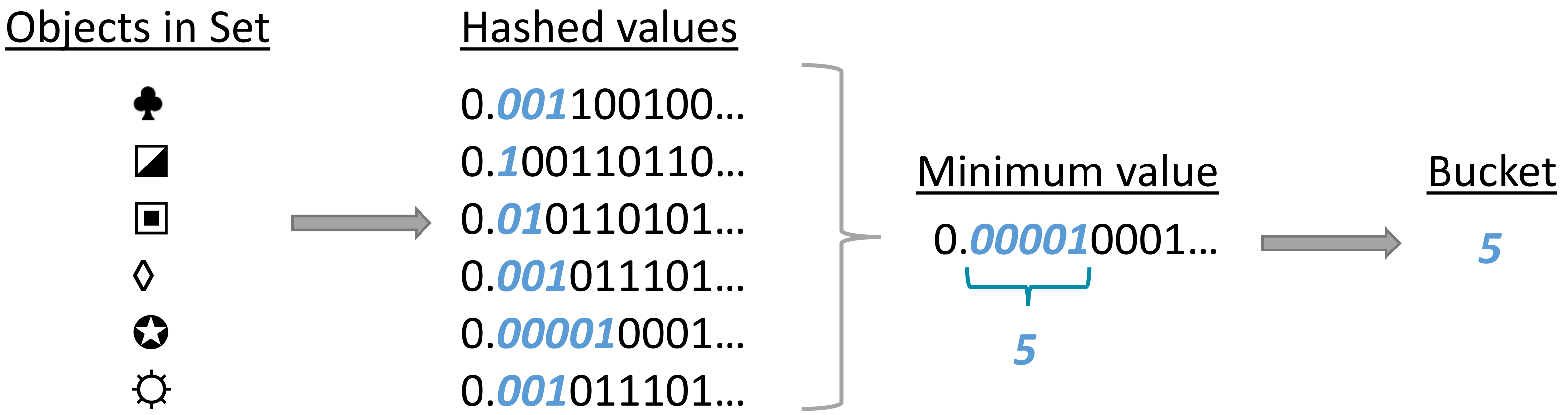
- How many distinct items exist in a list? [Flajolet, Martin, Ziv Bar-Yossef]



- Expected minimum is about $\frac{1}{n+1}$, so we need $O(\log(n))$ bits of storage.

(HYPER)LOGLOG COUNTING [FLAJOLET, ET AL]

- Only need to store the order of magnitude to get a good estimate, so can compress hashed values.



- With some correction terms, get errors that are $O\left(\frac{1}{\sqrt{k}}\right)$, where k is number of buckets / iterations.
- But need only $O(\log\log(n))$ space.

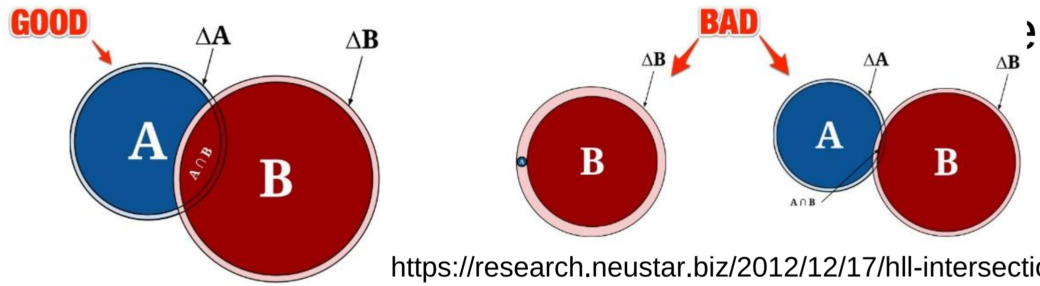
HYPERLOGLOG SET OPERATIONS

- Union cardinality
  - Cardinality of the union of sets is lossless with HLL
  - Determine the largest value for each bucket (iteration)
  - Estimate cardinality using the new sketch

3	6	4	1	5	2	2	4
2	3	3	4	2	3	1	6
5	3	1	0	7	4	3	3

5	6	4	4	7	4	3	6

- Intersection cardinality
  – Use inclusion-exclusion principle: $|A \cap B| = |A| + |B| - |A \cup B|$
  – Only accurate if the union and intersection cardinalities are comparable.
  – ꞌr of sets

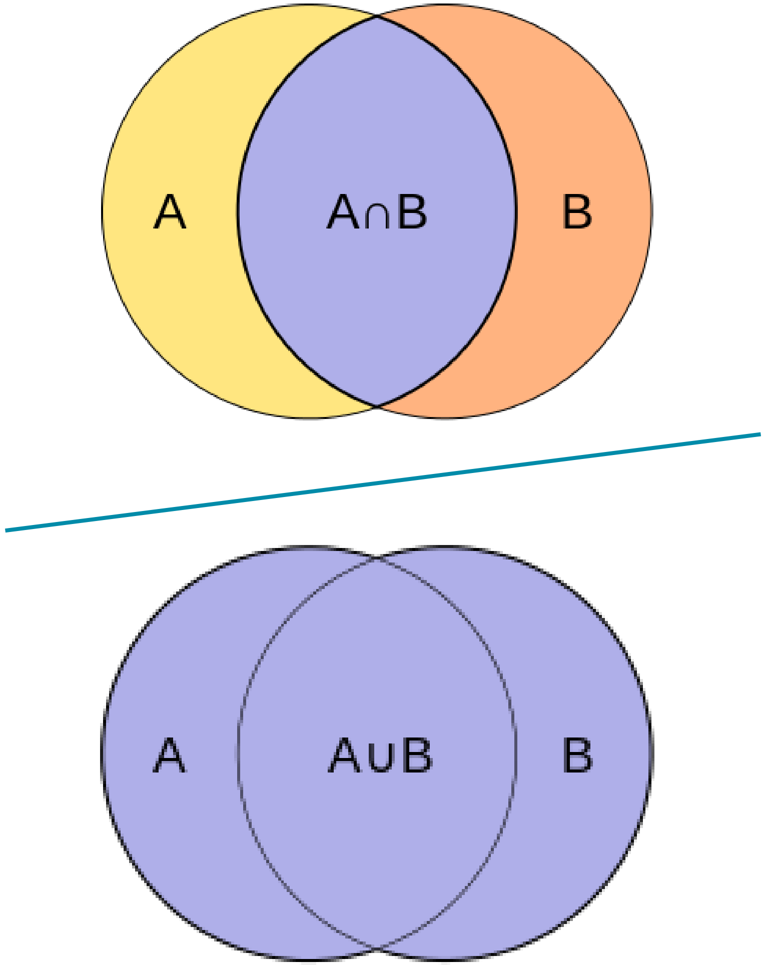


https://research.neustar.biz/2012/12/17/hll-intersections-2/

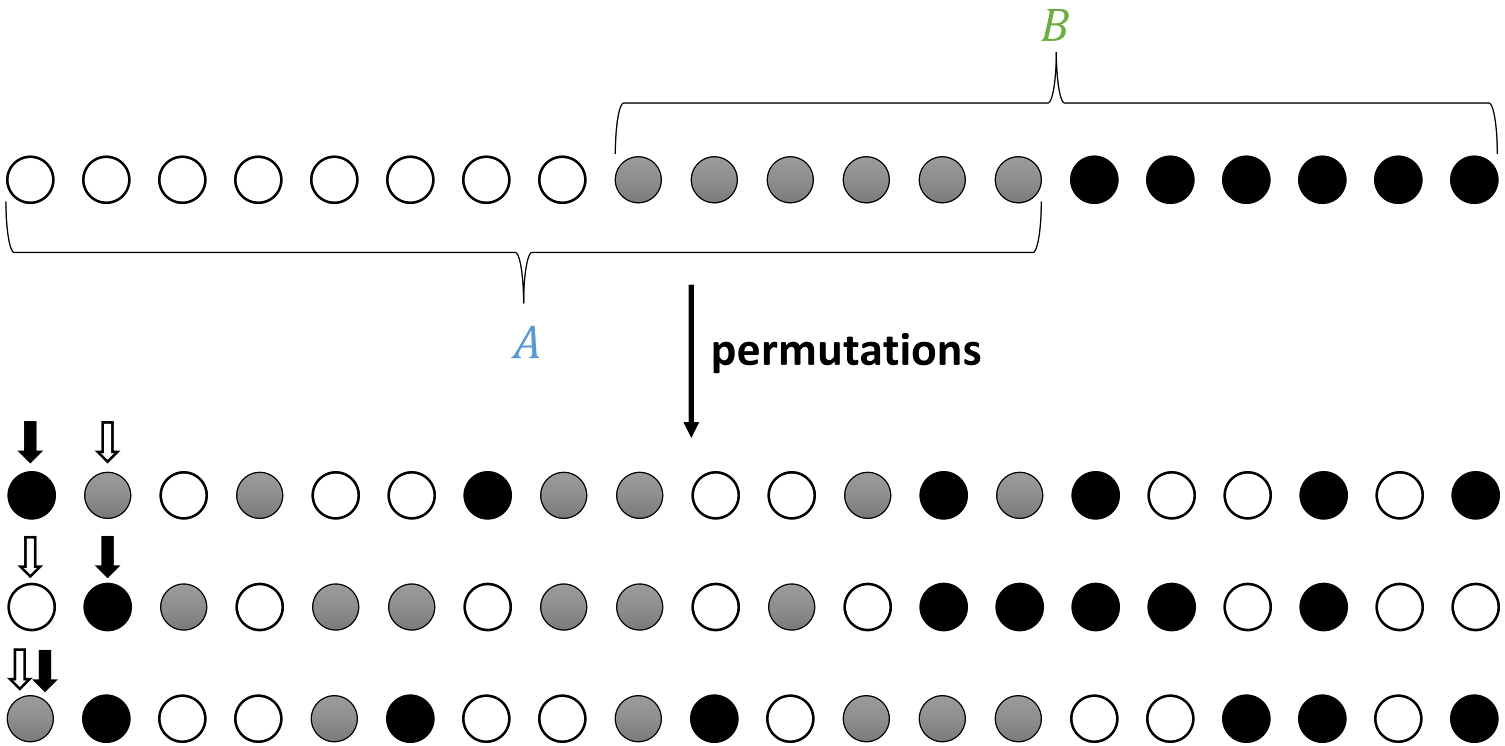
JACCARD INDEX [JACCARD, 1902]

- Measures the similarity between two sets by

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}.$$



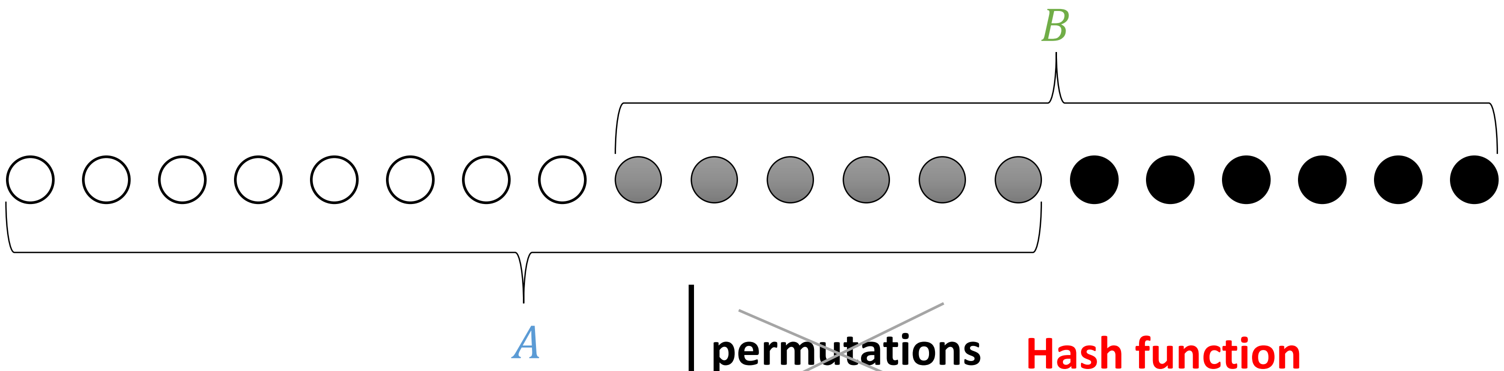
MINHASH [BRODER, 1997]



$$Prob(\min(A) = \min(B)) = \frac{|A \cap B|}{|A \cup B|} = J(A, B)$$

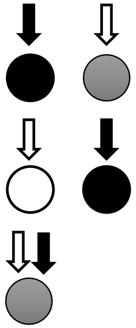
Can estimate Jaccard index from empirical probabilities!

MINHASH [BRODER, 1997]



~~permutations~~ **Hash function**

$O(log(n))\ bits$



← Minimum Hash 1 for sets A and B

← Minimum Hash 2 for sets A and B

← Minimum Hash 3 for sets A and B

$$Prob(\min(A) = \min(B)) = \frac{|A \cap B|}{|A \cup B|} = J(A, B)$$

Can estimate Jaccard index from empirical probabilities!

MINHASH: A WORKED EXAMPLE

Iterations (buckets)

$\|A\| = 5\text{M}$	$\|A \cap B\|$	$\|B\| = 10\text{M}$
0.1548		0.0358
0.1422		0.0657
0.0559	=	0.0559
0.1287		0.0400
0.0811	=	0.0811
0.1208		0.2649
0.1153		0.0120

$$J(A,B) \approx \frac{2}{7}$$

$\|A\| = 5\text{M}$	$\|A \cup B\|$	$\|B\| = 10\text{M}$
0.1548	0.0358	0.0358
0.1422	0.0657	0.0657
0.0559	0.0559	0.0559
0.1287	0.0400	0.0400
0.0811	0.0811	0.0811
0.1208	0.1208	0.2649
0.1153	0.0120	0.0120

Can merge sketches

IPython Notebook (via Google Colab):
http://ywyu.net/hackthevalley

HLL+MINHASH FOR A LARGE CLINICAL DATABASE

- How many distinct patients (cardinality) match a Boolean query?
- Create an HLL and MinHash sketch for each concept (diagnosis, medication, etc.)
- UNION (patients with X or Y)
  - HLL to combine concepts with no accuracy loss
- INTERSECTION (patients with X and Y)
  - HLL and MinHash with some accuracy loss
- Can combine multiple unions and intersections

NATIONAL CLAIMS DATA (70M PATIENTS)

Timing 7 Boolean queries

Query	Accuracy		Performance (Seconds)	
	SQL		SQL	
Diagnoses	51,015,187		628.5	
Diagnoses **OR** Medications	55,719,749		812.0	
Diagnoses **AND** Medications	18,574,415		1234.9	
Hypertension **AND** Diabetes	2,848,997		133.5	
(Age 45-64 **OR** Age 55-64) **AND** Female **AND** Diabetes	757,555		105.8	
Chronic pancreatitis **AND** Ulcerative colitis	1,712		1.5	

NATIONAL CLAIMS DATA (70M PATIENTS)

Timing 7 Boolean queries

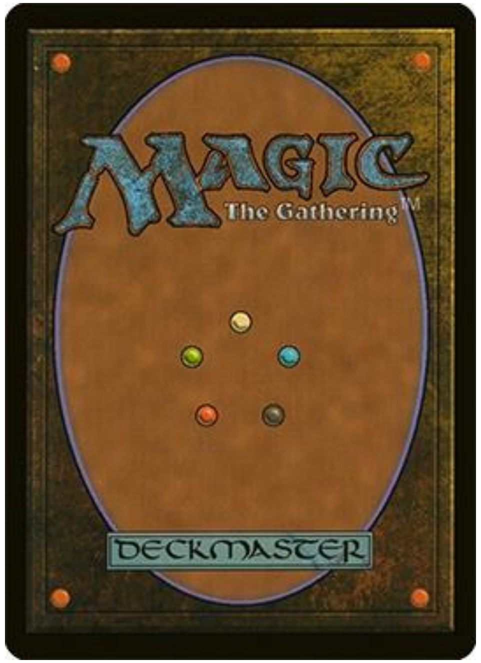
Query	Accuracy			Performance (Seconds)		
	SQL	Sketch	Error	SQL	Sketch	Ratio
Diagnoses	51,015,187	51,300,000	0.56%	628.5	0.033	19,046
Diagnoses **OR** Medications	55,719,749	55,900,000	0.32%	812.0	0.023	35,304
Diagnoses **AND** Medications	18,574,415	18,800,000	1.21%	1234.9	0.030	41,161
Hypertension **AND** Diabetes	2,848,997	2,820,000	1.02%	133.5	0.026	5,133
(Age 45-64 **OR** Age 55-64) **AND** Female **AND** Diabetes	757,555	774,000	2.17%	105.8	0.043	2,459
Chronic pancreatitis **AND** Ulcerative colitis	1,712	1,860	8.64%	1.5	0.030	50
	Database Size (GB):			4,300	5.3	

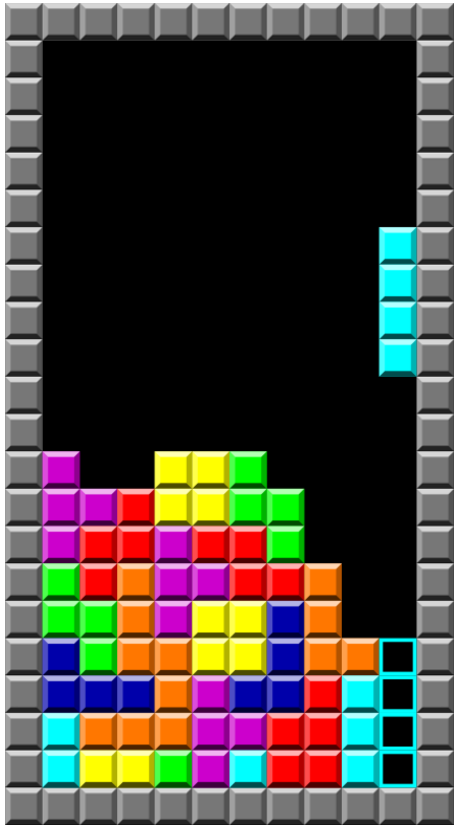
SKETCHES: PROS AND CONS

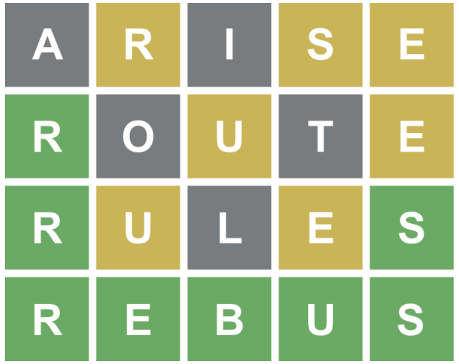
- Pros:
  - Small memory footprint
  - Composable, so good for parallel processing
  - After sketches are created, can run fast computations
- Cons:
  - Initial overhead of computing sketch
  - Can give inexact or wrong answers
  - Sometimes more complicated to implement
  - Have to design a specific sketch for each query type (set membership, item frequency, union cardinality, nearest-neighbors, Jaccard index, etc.)

TYPICAL GAMES AND COMPLEXITY

- Starting point: a popular game.
- Ingredients:
  - Sequence of decisions that have to be made by the player.
  - A score to be optimized.
- Modifications:
  - Need some natural generalization to allow for arbitrarily large input sizes for complexity and hardness to make sense.
  - May sometimes simplify certain rules or assume rational behavior from players.







3SAT $\leq_P$ SUPER MARIO BROS

- 3SAT Elements:
  - Collection of Boolean variables $x_1, \ldots, x_n$
  - Collection of literals $t_1, \ldots, t_m$, where each $t_j = \neg x_i$ or $t_j = x_i$ for some i.
  - Collection of clauses $C_1, \ldots, C_k$, where each $C_k = t_{j_1} \vee t_{j_2} \vee t_{j_3}$.
  - Is there a setting of the Boolean variables that makes all clauses true.
- Reduction strategy given in “Classic Nintendo Games are (Computationally) Hard” by Greg Aloupis, Erik Demaine, Alan Guo, and Giovanni Viglietta, 2015. *Theoretical Computer Science*. Part of Special Issue on Fun With Algorithms conference. (2012 arxiv)
  - Question: given a known Mario map, can you finish the level? (path reachability problem)
  - Strategy: force Mario to set variables by choosing a path through the level that reaches particular points – variable are set if you can reach that side of the variable gadget.

FRAMEWORK FROM [ADGV15]

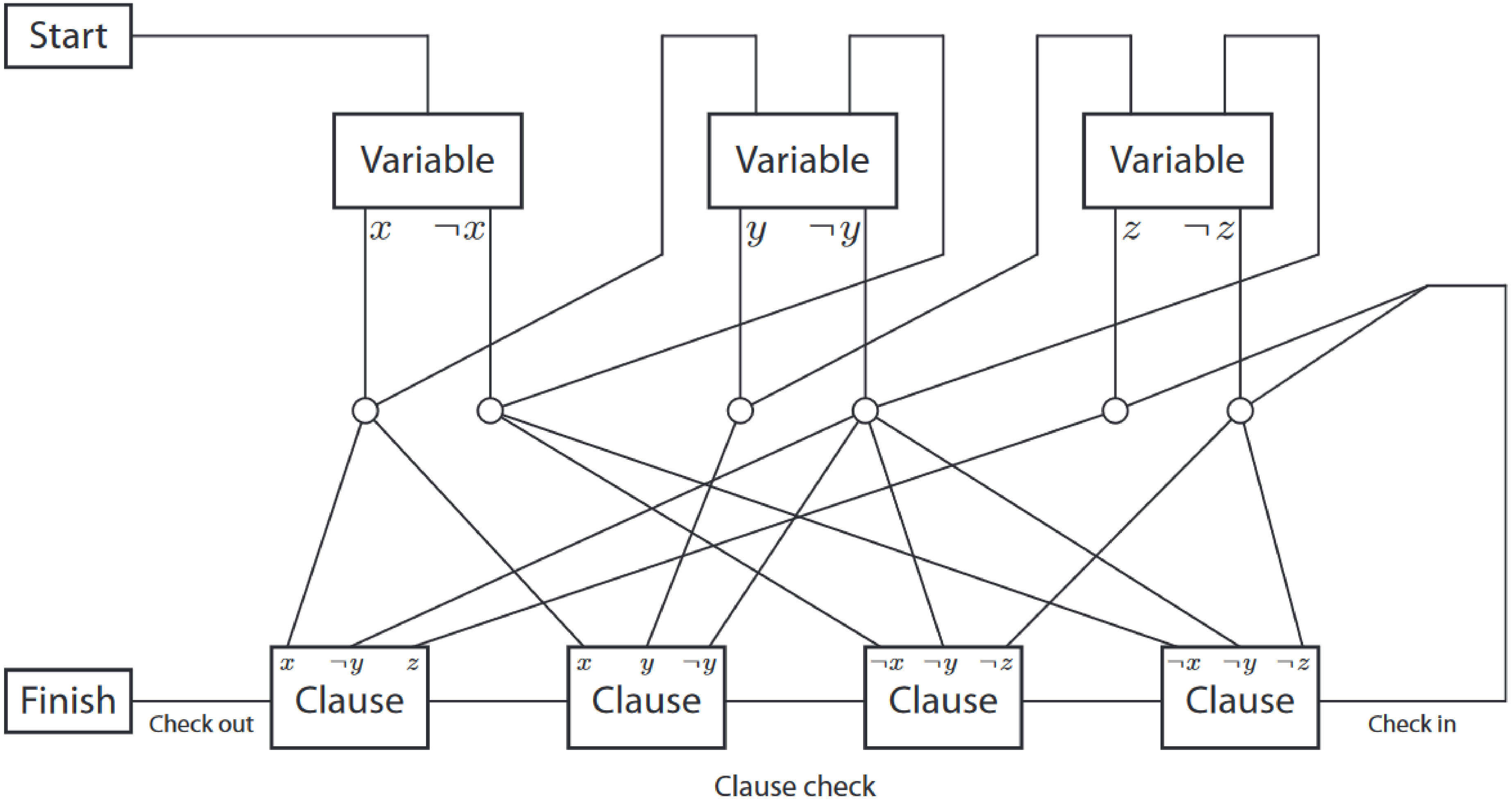
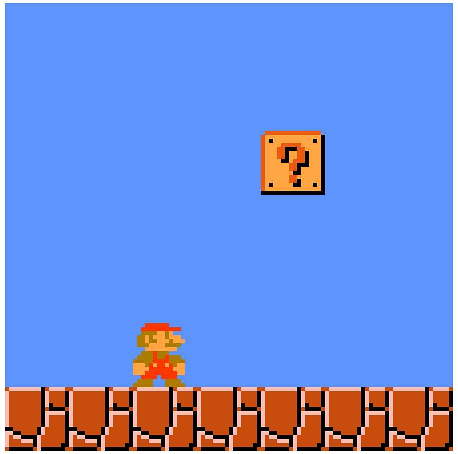
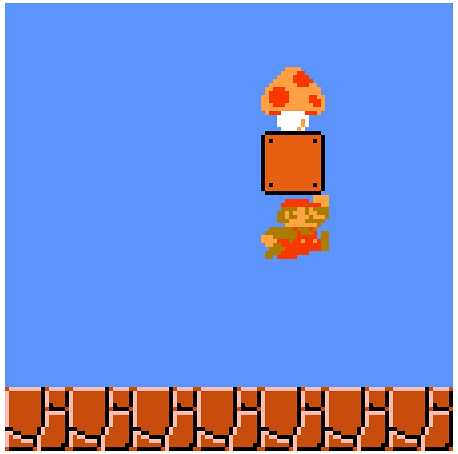


Figure 1: General framework for NP-hardness

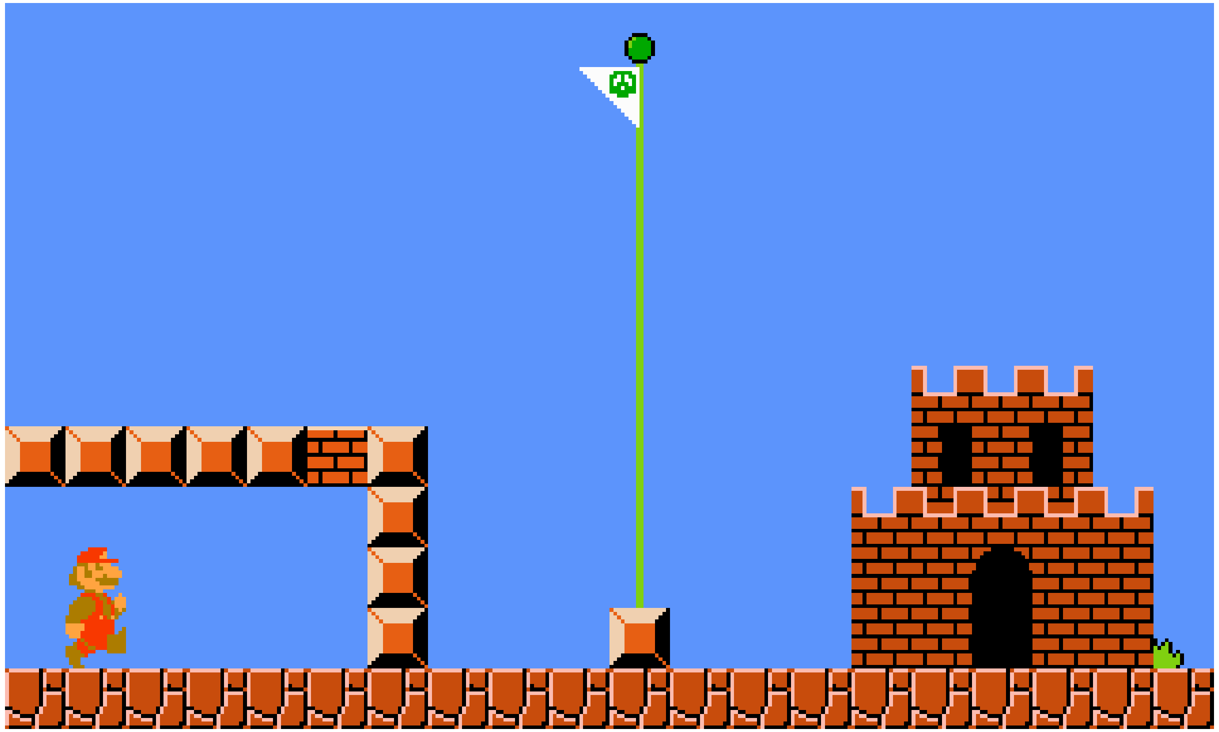
GADGETS

Start Gadget





End Gadget



[ADGV15]

VARIABLE GADGET

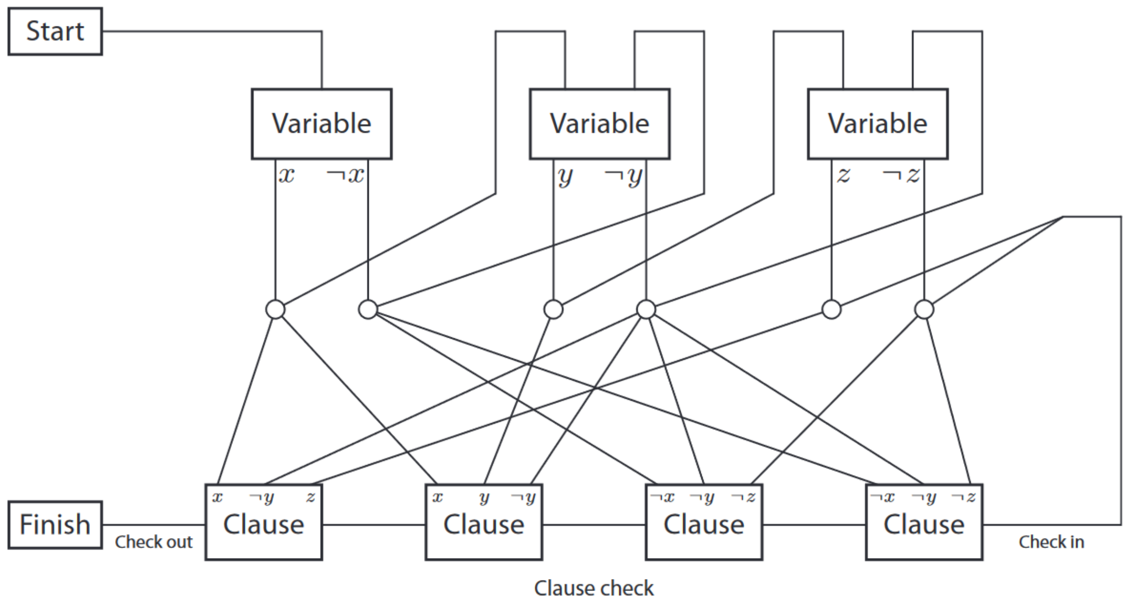
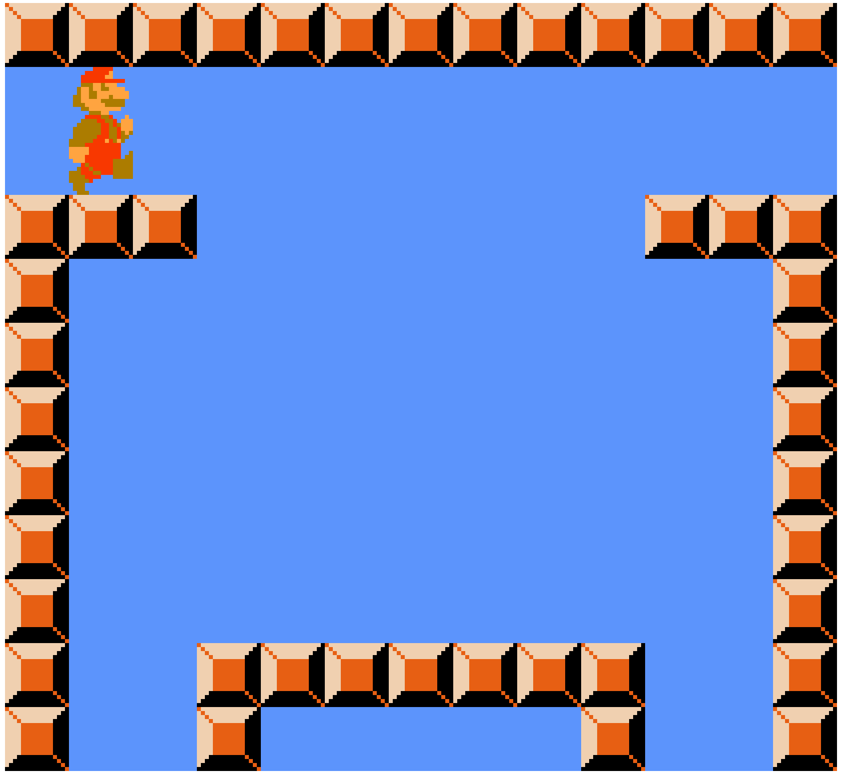


Figure 1: General framework for NP-hardness

Variable Gadget



CLAUSE GADGET

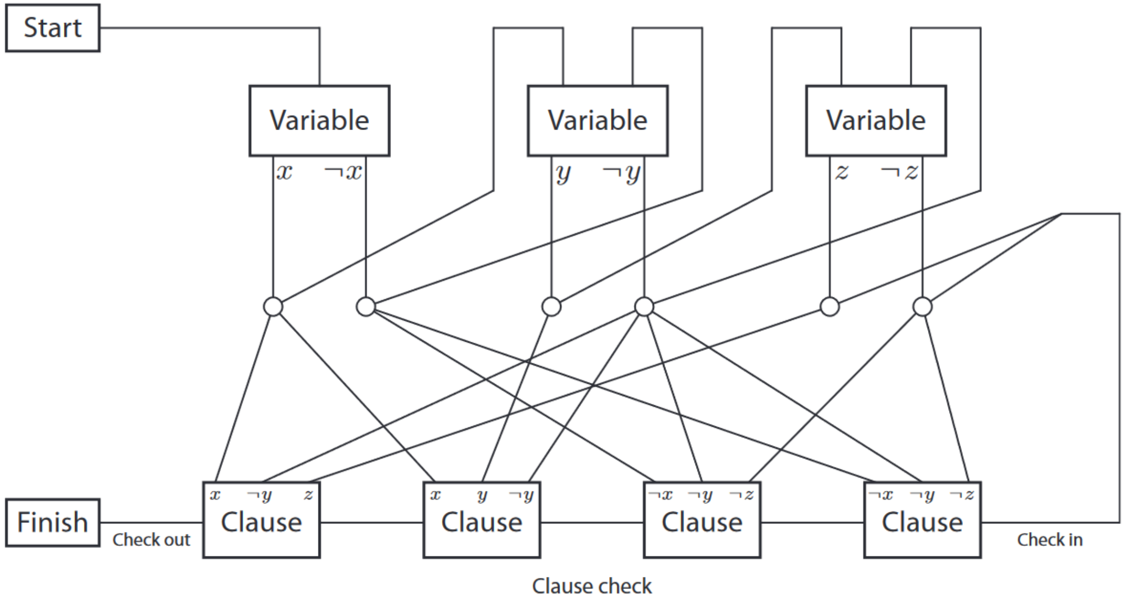
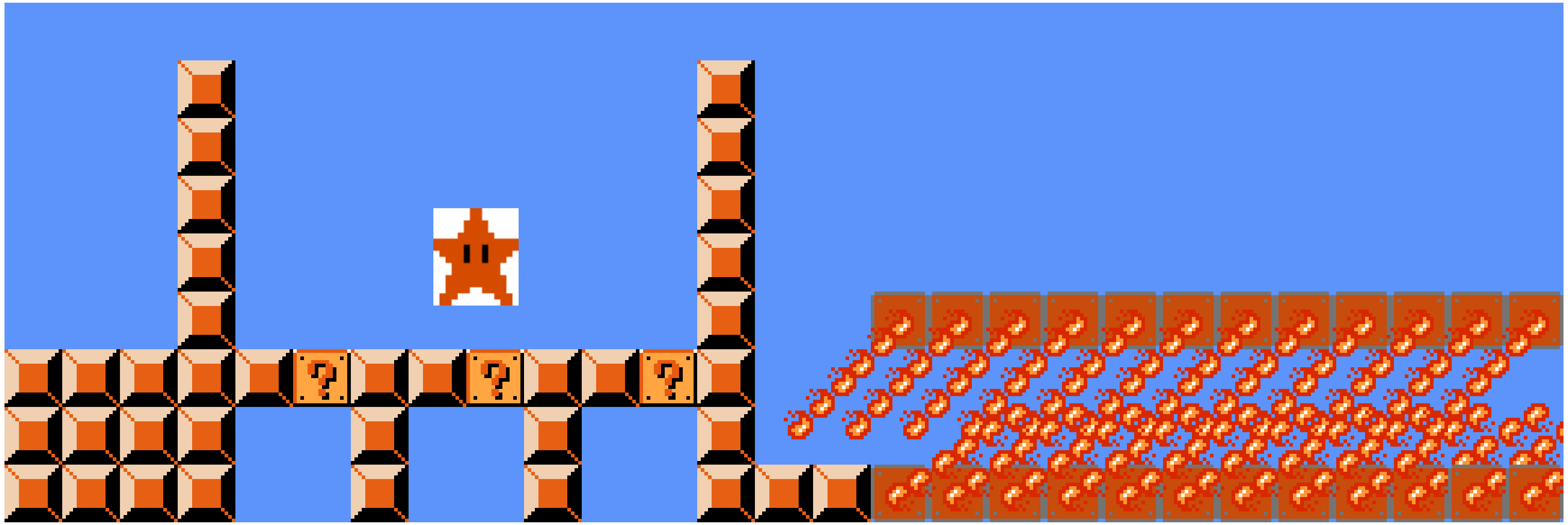


Figure 1: General framework for NP-hardness



FIXING GRAPH PLANARITY

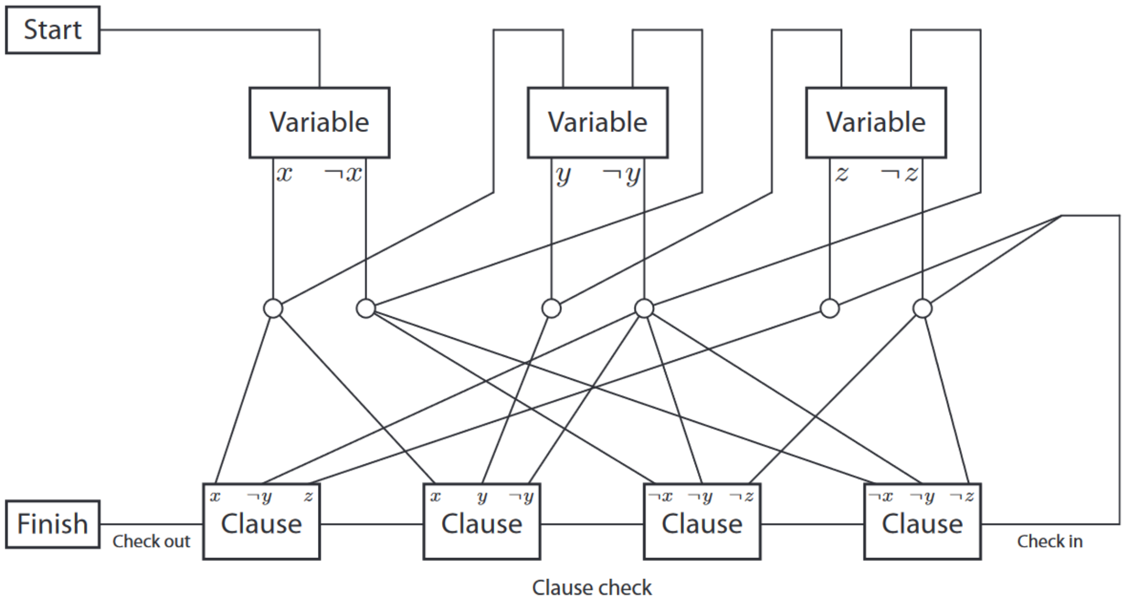
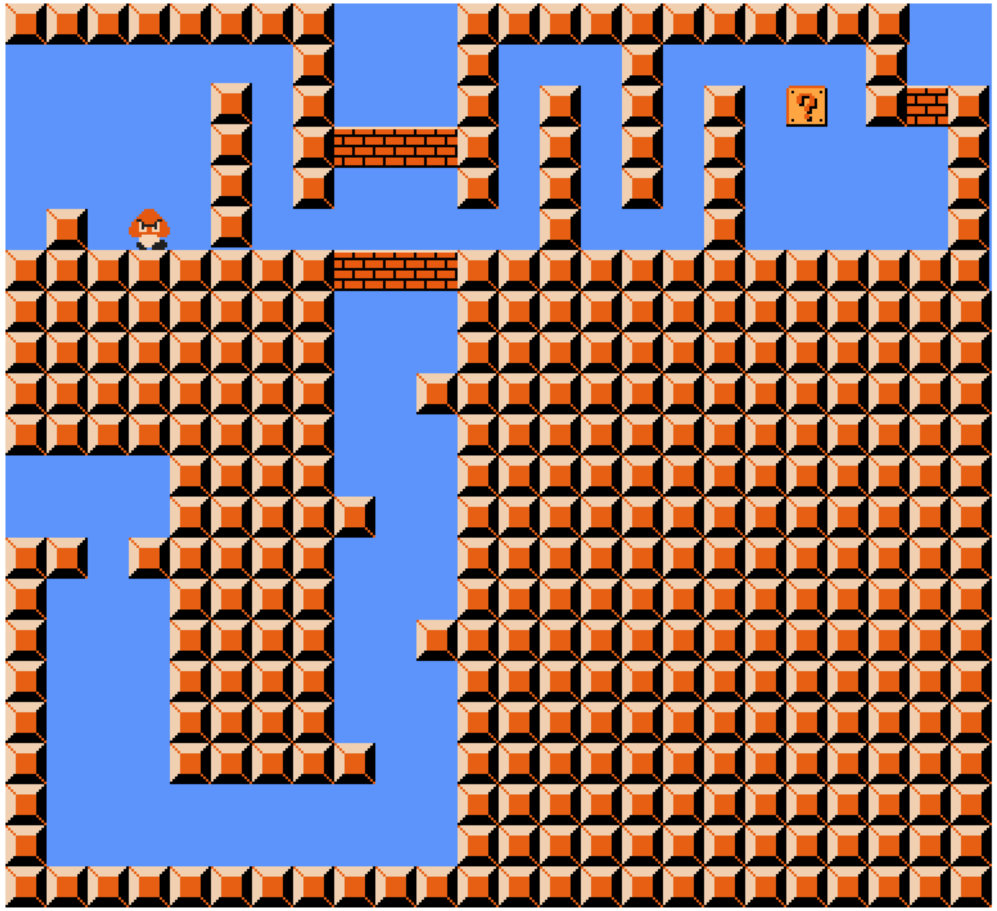


Figure 1: General framework for NP-hardness

Crossover Gadget

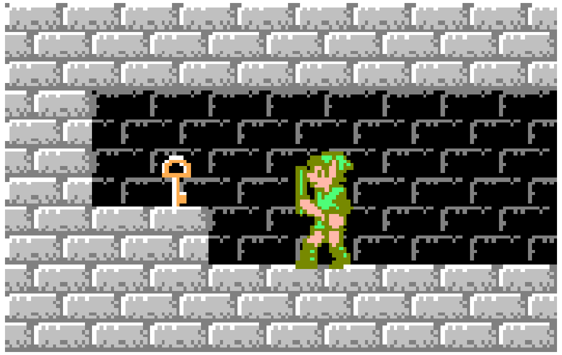


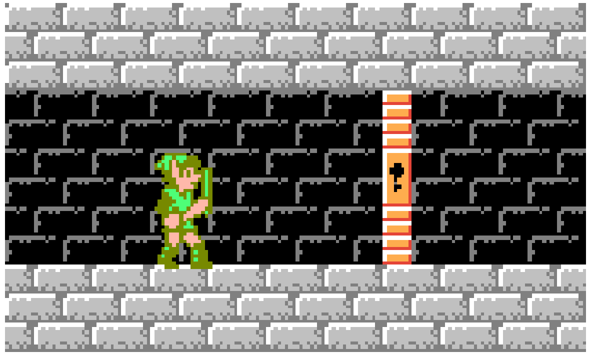
SUPER MARIO BROS IS PSPACE-COMPLETE

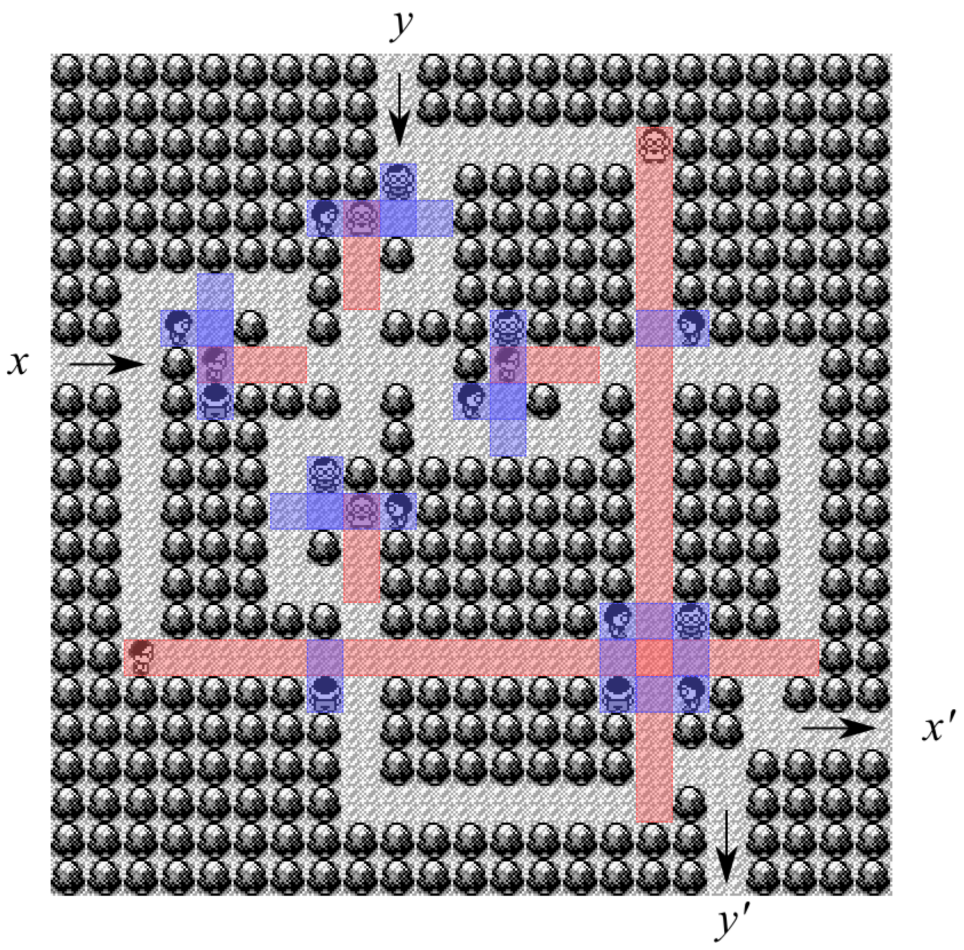
- NP ⊆ PSPACE
- Open question about whether they are equal. Most people think they are not.
- Super Mario Bros is PSPACE-complete. Therefore, it is in NP only if NP=PSPACE.

REACHABILITY METASTRATEGY [ADGV15]

- Proves many generalized games are NP-hard
- Super Marios 1-3, Super Mario World
- Donkey Kong Country 1-3
- Legend of Zelda
- Metroid games
- Pokemon role-playing games







STRATEGY IN OLYMPIC SPORTS

ROUTINES = SEQUENCE OF DECISIONS

- Back handspring
- Back double salto stretched with two turns
- Front salto stretched with full turn
- Front double salto tucked
- Thomas (Arabian) stretched

- Tempo Salto
- Back salto 3/2 turn
- Front salto 3/2 turn
- Split press to Japanese handstand
- Double Arabian pike salto

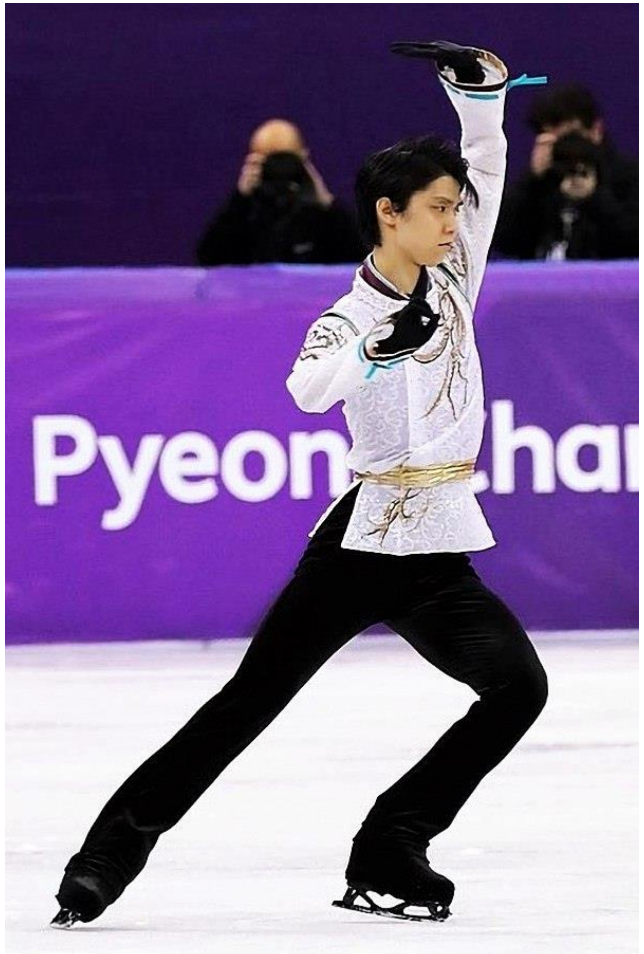
3/2 salto backwards with 3/2 twists named after Kurt Thomas

Natural generalization to larger input sizes in the form of new moves invented by competitors

SCORING A ROUTINE – BASIC COMPOSITIONAL SCORE

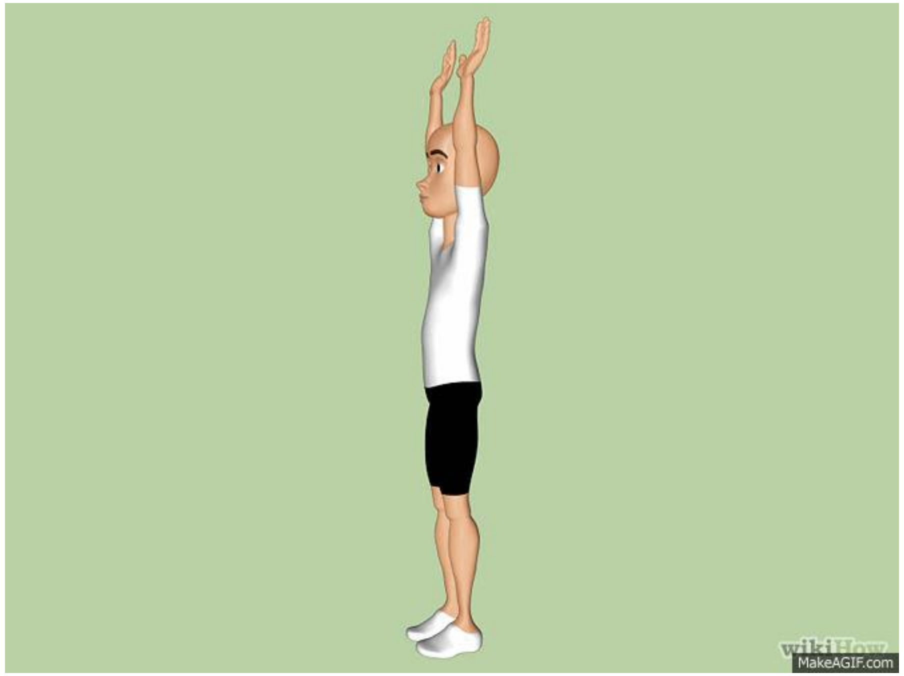
- Compositional scoring assigns each skill a point value, which are added together, as in the below figure skating free skat base value.

#	Executed Elements	Info	Base Value
1	4Lz		11.50
2	4F		11.00
3	4T		9.50
4	3A		8.00
5	CCSp4		3.20
6	StSq4		3.90
7	4T+3T		15.07
8	3Lz+3T		11.11
9	3F+1Eu+3S		11.11
10	ChSq1		3.00
11	FCCoSp4		3.50
12	CCoSp4		3.50
			94.39



OPTIMAL ROUTINE CONSTRUCTION

- Let's construct a routine:
- You know the following skills:
  - Handstand: 1.0pt
  - Front flip: 1.5pt
  - Back flip: 1.6pt
  - Cartwheel: 0.8pt
  - Splits: 0.9pt
- What is the optimal routine of length 10?



Compositional scoring can lead to degenerate routines.

2004 ATHENS GYMNASTICS CONTROVERSY

SUMMER 2004 GAMES -- GYMNASTICS: ALL-AROUND

SUMMER 2004 GAMES -- GYMNASTICS: ALL-AROUND; Judges Suspended for Error, But Hamm Will Keep Gold

Olympic games 2004

Uproar in the gym: Russians complain to IOC

FIG considers gymnastics rule changes

Judges may use video replay in review of start values

Associated Press
updated 5:24 p.m. ET Sept. 24, 2004

CODE OF POINTS

Inhaltsverzeichnis

TEIL I - VORSCHRIFTEN FÜR WETTKAMPFTEILNEHMER

Kapitel 1 - Ziel und Zweck
Kapitel 2 - Regeln für den Turner
Artikel 2.1 - Rechte
Artikel 2.2 - Verantwortlichkeiten
Artikel 2.3 - Pflichten
Artikel 2.4 - Sanktionen
Artikel 2.5 - Eid der Athleten
Kapitel 3 - Regeln für Trainer
Artikel 3.1 - Rechte
Artikel 3.2 - Verantwortlichkeiten
Artikel 3.3 - Sanktionen
Artikel 3.4 - Einsprüche
Artikel 3.5 - Eid der Trainer
Kapitel 4 - Regeln für das Technische Komitee
Artikel 4.1 - MTC Präsident
Artikel 4.2 - MTC Mitglieder
Kapitel 5 - Regeln und Struktur für Kampfgerichten
Artikel 5.1 - Verantwortlichkeiten
Artikel 5.2 - Rechte
Artikel 5.3 - Zusammensetzung des Kampfgerichts am Gerät
Artikel 5.4 - Aufgaben des Kampfgerichts am Gerät
Artikel 5.5 - Aufgaben der Zeitnehmer, Linienrichter und Sekretäre
Artikel 5.6 - Sitzordnungt
Artikel 5.7 - Eid der Kampfrichter
Artikel 5.8 - Anerkennung von Elementnamen

TEIL II - BEWERTUNG DER ÜBUNGEN

Kapitel 6 - Bewertungen der Wettkampfübungen
Artikel 6.1 - Allgemeine Grundsätze
Artikel 6.2 - Bestimmung der Endwertung
Kapitel 7 - Regeln zur D-Note
Artikel 7.1 - Schwierigkeit
Artikel 7.2 - Elementgruppen und Abgangsanforderungen
Artikel 7.3 - Verbindungsbonifikationen
Artikel 7.4 - Bewertung durch das D-Kampfgericht
Artikel 7.5 - Wiederholung
Artikel 7.6 - Übersicht der Bewertung durch das D-Kampfgericht

TABLE OF CONTENTS

PART I - REGULATIONS GOVERNING COMPETITION PARTICIPANTS

Section 1 - Purpose
Section 2 - Regulations governing Gymnasts
Article 2.1 - Rights of Gymnasts
Article 2.2 - Responsibilities of the Gymnast
Article 2.3 - Duties of the Gymnast
Article 2.4 - Penalties
Article 2.5 - Gymnast Oath
Section 3 - Regulations for coaches
Article 3.1 - Rights of the Coaches
Article 3.2 - Responsibilities of the Coaches
Article 3.3 - Penalties
Article 3.4 - Inquiries
Article 3.5 - Coaches' Oath
Section 4 - Regulations for Technical Committee
Article 4.1 - The President of the MTC
Article 4.2 - MTC Members
Section 5 - Regulations and structure of apparatus juries
Article 5.1 - Responsibilities
Article 5.2 - Rights
Article 5.3 - Composition of the Apparatus Jury
Article 5.4 - Function of the Apparatus Jury
Article 5.5 - Functions of Timers, Line judges, and Secretaries
Article 5.6 - Seating Arrangements
Article 5.7 - Judges' Oath
Article 5.8 - Naming of New elements

PART II - EVALUATION OF THE EXERCISES

Section 6 - Evaluation of the Competition Exercises
Article 6.1 - General Rules
Article 6.2 - Determination of the Final Score
Section 7 - Regulations governing the D Score
Article 7.1 - Difficulty
Article 7.2 - Element Groups and Dismount Requirements
Article 7.3 - Connection Points
Article 7.4 - Evaluation by the D Jury
Article 7.5 - Repetition
Article 7.6 - D jury evaluation

СОДЕРЖАНИЕ

ЧАСТЬ I - ПРАВИЛА ДЛЯ УЧАСТНИКОВ СОРЕВНОВАНИЙ

Секция 1 - Цели
Секция 2 - Правила для Гимнастов
Статья 2.1 - Права Гимнастов
Статья 2.2 - Обязанности Гимнастов
Статья 2.3 - Функции Гимнастов
Статья 2.4 - Сбавки
Статья 2.5 - Клятва Гимнастов
Секция 3 - Правила для Тренеров
Статья 3.1 - Права Тренеров
Статья 3.2 - Обязанности Тренеров
Статья 3.3 - Сбавки
Статья 3.4 - Протесты
Статья 3.5 - Клятва тренера
Секция 4 - Правила для Технического Комитета
Статья 4.1 - Президент МТС
Статья 4.2 - Члены МТС
Секция 5 - Правила и состав бригад по снарядам
Статья 5.1 - Обязанности
Статья 5.2 - Права
Статья 5.3 - Состав Бригад по снарядам
Статья 5.4 - Функции Бригад по снарядам
Статья 5.5 - Функции Хронометристов, Судей на Линии и Секретариата
Статья 5.6 - Месторасположение Судей
Статья 5.7 - Клятва Судей
Статья 5.8 - Присвоение имен новым элементам

ЧАСТЬ II - ОЦЕНИВАНИЕ УПРАЖНЕНИЯ

Секция 6 - Оценивание Соревновательного Упражнения
Статья 6.1 - Общие Правила
Статья 6.2 - Определение Окончательной Оценки
Секция 7 - Правила Относительно оценки D
Статья 7.1 - Трудность
Статья 7.2 - Требования к Группам Элементов и Соскокам
Статья 7.3 - Прибавки за Соединения
Статья 7.4 - Оценивание судейской Бригады D
Статья 7.5 - Повторы
Статья 7.6 - Судейство Бригады D

Kapitel 8 - Regeln zur E-Note
Artikel 8.1 - Beschreibung der Übungsausführung
Artikel 8.2 - Berechnung der Abzüge durch das E-Kampfgericht
Artikel 8.3 - Hinweise für den Turner
Kapitel 9 - Technische Richtlinien
Artikel 9.1 - Bewertung durch das E-Kampfgericht
Artikel 9.2 - Bestimmung von Ausführungs- und technischen Fehlern
Artikel 9.3 - Anforderung an die Übungszusammensetzung
Artikel 9.4 - Übersicht der Abzüge durch das E-Kampfgericht

TEIL III - GERÄTE
Kapitel 10 - Boden
Artikel 10.1 - Beschreibung der Übung
Artikel 10.2 - Inhalt und Aufbau
Artikel 10.3 - Spezifische Abzüge
Artikel 10.4 - Elementetabelle
Kapitel 11 - Pauschenpferd
Artikel 11.1 - Beschreibung der Übung
Artikel 11.2 - Inhalt und Aufbau
Artikel 11.3 - Spezifische Abzüge
Artikel 11.4 - Elementetabelle
Kapitel 12 - Ringe
Artikel 12.1 - Beschreibung der Übung
Artikel 12.2 - Inhalt und Aufbau
Artikel 12.3 - Spezifische Abzüge
Artikel 12.4 - Elementetabelle
Kapitel 13 - Sprung
Artikel 13.1 - Beschreibung des Sprungs
Artikel 13.2 - Inhalt und Aufbau
Artikel 13.3 - Spezifische Abzüge
Artikel 13.4 - Elementetabelle
Kapitel 14 - Barren
Artikel 14.1 - Beschreibung der Übung
Artikel 14.2 - Inhalt und Aufbau
Artikel 14.3 - Spezifische Abzüge
Artikel 14.4 - Elementetabelle
Kapitel 15 - Reck
Artikel 15.1 - Beschreibung der Übung
Artikel 15.2 - Inhalt und Aufbau
Artikel 15.3 - Spezielle Abzüge
Artikel 15.4 - Elementetabelle

TEIL IV - Anhang
A. Zusätzliche Erläuterungen und Interpretationen

Section 8 - Regulations governing the E Score
Article 8.1 - Description of Exercise Presentation
Article 8.2 - Calculation of E Jury deductions
Article 8.3 - Instructions for the Gymnasts
Section 9 - Technical Directives
Article 9.1 - Evaluation by the E jury
Article 9.2 - Determination of Execution and Technical Errors
Article 9.3 - Expectations for exercise construction
Article 9.4 - E jury deductions

PART III - Apparatus
Section 10 - Floor Exercise
Article 10.1 - Exercise description
Article 10.2 - Content and Construction
Article 10.3 - Specific Apparatus deductions
Article 10.4 - Table of Elements
Section 11 - Pommel Horse
Article 11.1 - Exercise description
Article 11.2 - Content and Construction
Article 11.3 - Specific Apparatus deductions
Article 11.4 - Table of Elements
Section 12 - Rings
Article 12.1 - Exercise description
Article 12.2 - Content and Construction
Article 12.3 - Specific Apparatus deductions
Article 12.4 - Table of Elements
Section 13 - Vault
Article 13.1 - Description of a Vault
Article 13.2 - Content and Construction
Article 13.3 - Specific Apparatus deductions
Article 13.4 - Table of Elements
Section 14 - Parallel Bars
Article 14.1 - Exercise description
Article 14.2 - Content and Construction
Article 14.3 - Specific Apparatus deductions
Article 14.4 - Table of Elements
Section 15 - Horizontal Bar
Article 15.1 - Exercise description
Article 15.2 - Content and Construction
Article 15.3 - Specific Apparatus deductions
Article 15.4 - Table of Elements

PART IV - Appendices
A. Additional Explanations and interpretations

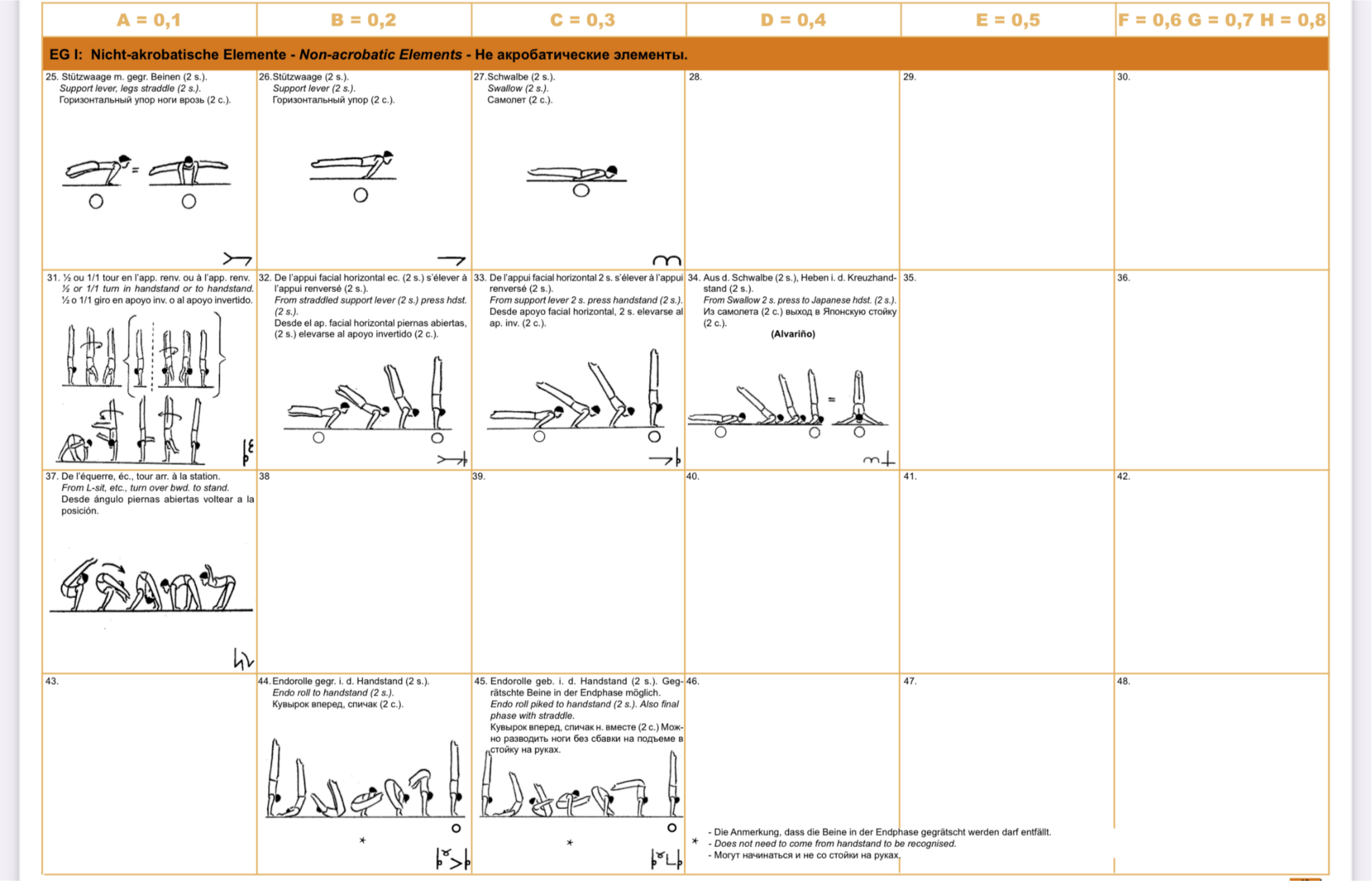
Секция 8 - Правила относительно Оценки E
Статья 8.1 - Описание Преставления Упражнения
Статья 8.2 - Подсчет Сбавок Судейской Бригадой E
Статья 8.3 - Указания Гимнасту
Секция 9 - Технические Директивы
Статья 9.1 - Работа Судейской Бригады E
Статья 9.2 - Определение ошибок исполнения и техники
Статья 9.3 - Требования к построению композиции упражнения
Статья 9.4 - Сбавки производимые бригадой E

ЧАСТЬ III - СНАРЯДЫ
Секция 10 - Вольные Упражнения
Статья 10.1 - Описание Упражниния
Статья 10.2 - Содержание и построение
Статья 10.3 - Специфические сбавки
Статья 10.4 - Таблица элементов
Секция 11 - Конь
Статья 11.1 - Описание Упражниния
Статья 11.2 - Содержание и построение
Статья 11.3 - Специфические сбавки
Статья 11.4 - Таблица элементов
Секция 12 - Кольца
Статья 12.1 - Описание Упражниния
Статья 12.2 - Содержание и построение
Статья 12.3 - Специфические сбавки
Статья 12.4 - Таблица элементов
Секция 13 - Опорный Прыжок
Статья 13.1 - Описание Упражниния
Статья 13.2 - Содержание и построение
Статья 13.3 - Специфические сбавки
Статья 13.4 - Таблица прыжков
Секция 14 - Брусья
Статья 14.1 - Описание Упражниния
Статья 14.2 - Содержание и построение
Статья 14.3 - Специфические сбавки
Статья 14.4 - Таблица элементов
Секция 15 - Перекладина
Статья 15.1 - Описание Упражниния
Статья 15.2 - Содержание и построение
Статья 15.3 - Специфические сбавки
Статья 15.4 - Таблица элементов

ЧАСТЬ IV Приложения
А. Дополнительные объяснения и интерпретации

170 page-long Men's Artistic Gymnastics
2022-2024 code of points

INDEX OF SKILLS IN CODES OF POINTS



A = 0,1	B = 0,2	C = 0,3	D = 0,4	E = 0,5	F = 0,6 G = 0,7 H = 0,8
EG I: Nicht-akrobatische Elemente - *Non-acrobatic Elements* - Не акробатические элементы.					
25. Stützwaage m. gegr. Beinen (2 s.). *Support lever, legs straddle (2 s.).* Горизонтальный упор ноги врозь (2 с.).	26. Stützwaage (2 s.). *Support lever (2 s.).* Горизонтальный упор (2 с.).	27. Schwalbe (2 s.). *Swallow (2 s.).* Самолет (2 с.).	28.	29.	30.
31. ½ ou 1/1 tour en l'app. renv. ou à l'app. renv. *½ or 1/1 turn in handstand or to handstand.* ½ o 1/1 giro en apoyo inv. o al apoyo invertido.	32. De l'appui facial horizontal ec. (2 s.) s'élever à l'appui renversé (2 s.). *From straddled support lever (2 s.) press hdst. (2 s.).* Desde el ap. facial horizontal piernas abiertas, (2 s.) elevarse al apoyo invertido (2 c.).	33. De l'appui facial horizontal 2 s. s'élever à l'appui renversé (2 s.). *From support lever 2 s. press handstand (2 s.).* Desde apoyo facial horizontal, 2 s. elevarse al ap. inv. (2 c.).	34. Aus d. Schwalbe (2 s.), Heben i. d. Kreuzhand-stand (2 s.). *From Swallow 2 s. press to Japanese hdst. (2 s.).* Из самолета (2 с.) выход в Японскую стойку (2 с.). **(Alvariño)**	35.	36.
37. De l'équerre, éc., tour arr. à la station. *From L-sit, etc., turn over bwd. to stand.* Desde ángulo piernas abiertas voltear a la posición.	38	39.	40.	41.	42.
43.	44. Endorolle gegr. i. d. Handstand (2 s.). *Endo roll to handstand (2 s.).* Кувырок вперед, спичак (2 с.). *	45. Endorolle geb. i. d. Handstand (2 s.). Gegrätschte Beine in der Endphase möglich. *Endo roll piked to handstand (2 s.). Also final phase with straddle.* Кувырок вперед, спичак н. вместе (2 с.) Можно разводить ноги без сбавки на подъеме в стойку на руках. *	46.	47.	48.

* - Die Anmerkung, dass die Beine in der Endphase gegrätscht werden darf entfällt.
- *Does not need to come from handstand to be recognised.*
- Могут начинаться и не со стойки на руках.

FOUR TYPES OF MODIFICATIONS

1. Anti repetition within a class (possibly overlapping classes)
2. Element group penalty (non-overlapping element groups)
3. Connection bonuses
4. Graph structure / incomplete graphs

4. Special repetitions:

a) Repeated elements (same Code Identification Number) cannot contribute to the "D" score. On Rings, this rule is extended so that a maximum of 1 final strength position in each EG may be recognized for difficulty. Thus, for example only two cross type elements (regular, L cross, or V cross) or support scale type elements (regular or straddled) are permitted in an exercise for difficulty value (one in Group II and one in Group III).

b) A maximum of 2 Guczoghy type elements can be present in the exercise.

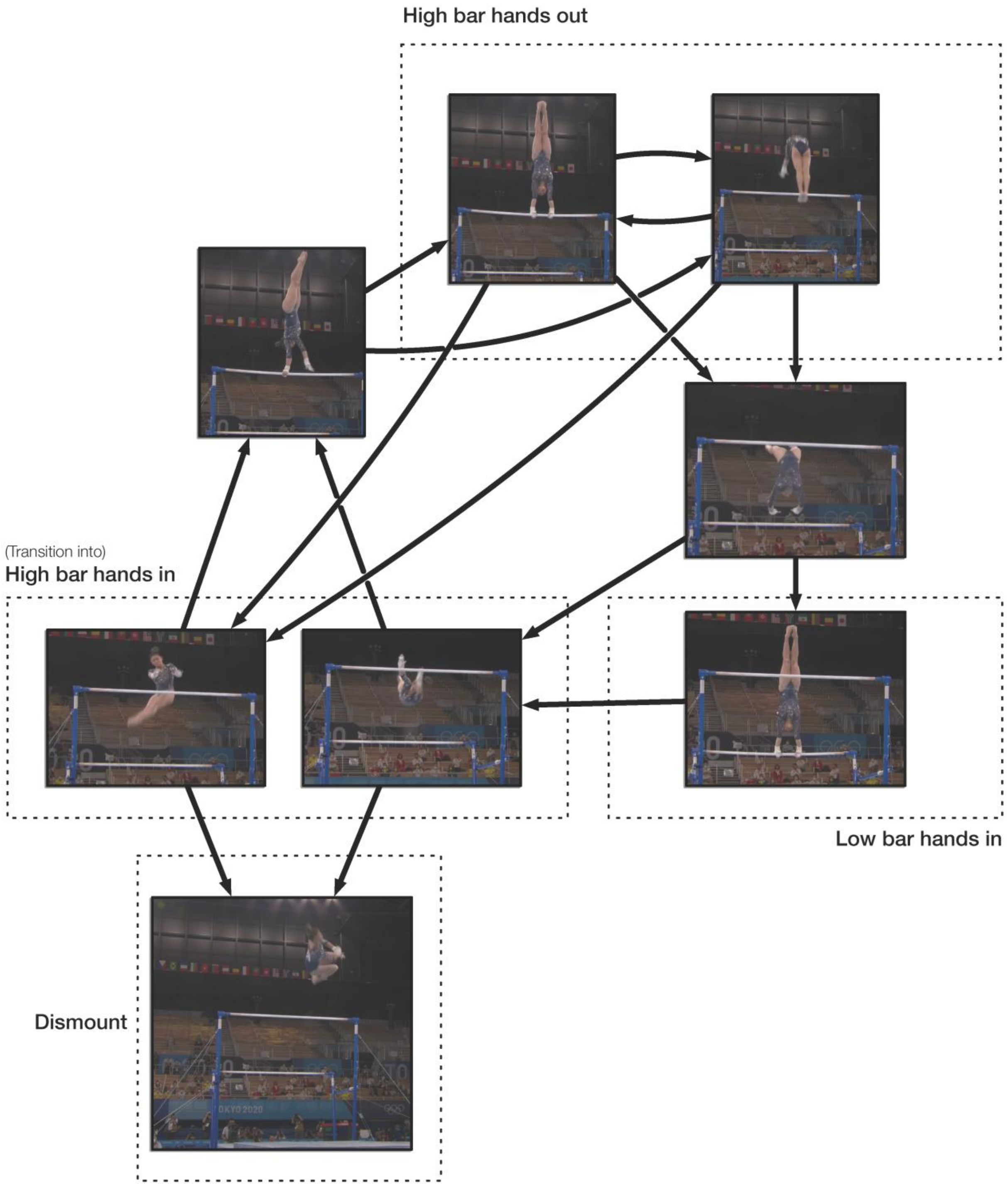
11.3 Composition Requirements (CR) – D-Panel 2.00

1. Flight element from HB to LB	award 0.50
2. Flight element on the same bar	award 0.50
3. Different grips *(not cast, MT or DMT)*	award 0.50
4. Non-flight element with min. 360º turn *(not MT)*	award 0.50

b) Special rule: Elements to one bar in cross support have the same value as done to two bars, except they increase by one value more when connected to Healy type elements (each Healy element also increases by one value) hold is allowed in the one bar handstand.

GRAPH STRUCTURE/ INCOMPLETE GRAPHS

- Physical constraints
- Positions
- Momentum
- May differ between athletes, so can be part of the problem instance.



FORMAL STRUCTURE OF PROBLEM

- Let Σ be the set of possible skills, and $n = |\Sigma|$, the number of skills.
- Let $m = |S|$ be the maximum length of an allowed routine $S \in \Sigma^+$.
- Let q be the total number of scoring rules (of any type, defined in the next section).
- Let $z = n + m + q$, the size of the input.

- The input to the problem is the set of skills an athlete can perform, their point values as defined in the Code of Points, the deductions on each skill the athlete will receive, and the total number of additional scoring rules.

- Decision Problem: given a rational scoring function *f(S)* defined by the scoring rules over the set of possible routines, does there exist a routine scoring at least X points.

- Optimization Problem: what is the highest scoring routine an athlete can perform?

SCORING RULES

Basic compositional scoring on a routine $S = s_1 \ldots s_m$ where $p(s_i)$ is the point value of the skill and $d(s_i)$ is the deductions for improper form on that skill.	$$f_{\text{BASIC}}(S) = \sum_{i=1}^{m} (p(s_i) - d(s_i))$$

SCORING RULES

Basic compositional scoring on a routine $S = s_1 \ldots s_m$ where $p(s_i)$ is the point value of the skill and $d(s_i)$ is the deductions for improper form on that skill.	$$f_{\text{BASIC}}(S) = \sum_{i=1}^{m}(p(s_i) - d(s_i))$$
Anti-repetition rule (ρ, k) where $\rho \subseteq \Sigma$ and k ∈ N. For $S = s_1 \ldots s_m$, define $R = r_1 \ldots, r_m$ as a bitstring specifying if each skill is recognized for points and must satisfy Anti-repetition rules.	$$f_{\text{ANTI-REPETITION}}(S) = \max_{\text{valid } R} \sum_{i=1}^{m}(p(s_i) \cdot r_i - d(s_i))$$

SCORING RULES

Basic compositional scoring on a routine $S = s_1 \ldots s_m$ where $p(s_i)$ is the point value of the skill and $d(s_i)$ is the deductions for improper form on that skill.	$f_{\text{BASIC}}(S) = \sum_{i=1}^{m}(p(s_i) - d(s_i))$
Anti-repetition rule (ρ, k) where $\rho \subseteq \Sigma$ and k ∈ N. For $S = s_1 \ldots s_m$, define $R = r_1 \ldots, r_m$ as a bitstring specifying if each skill is recognized for points and must satisfy Anti-repetition rules.	$f_{\text{ANTI-REPETITION}}(S) = \max_{\text{valid } R} \sum_{i=1}^{m}(p(s_i) \cdot r_i - d(s_i))$
Element group rule (ρ, p_ρ) where $\rho \subseteq \Sigma$ and p_ρ is the point value associated. Let indicator variable I_j specify whether EG rule *j* is satisfied.	$f_{\text{ELEMENTGROUP}} = \sum_{i=1}^{m}(p(s_i) - d(s_i)) + \sum_{j=1}^{q_{eg}} I_j p_{\rho_j}$

SCORING RULES

Basic compositional scoring on a routine $S = s_1 \dots s_m$ where $p(s_i)$ is the point value of the skill and $d(s_i)$ is the deductions for improper form on that skill.	$f_{\text{BASIC}}(S) = \sum_{i=1}^{m}(p(s_i) - d(s_i))$
Anti-repetition rule (ρ, k) where $\rho \subseteq \Sigma$ and k ∈ N. For $S = s_1 \dots s_m$, define $R = r_1 \dots, r_m$ as a bitstring specifying if each skill is recognized for points and must satisfy Anti-repetition rules.	$f_{\text{ANTI-REPETITION}}(S) = \max_{\text{valid } R} \sum_{i=1}^{m}(p(s_i) \cdot r_i - d(s_i))$
Element group rule (ρ, p_ρ) where $\rho \subseteq \Sigma$ and p_ρ is the point value associated. Let indicator variable I_j specify whether EG rule *j* is satisfied.	$f_{\text{ELEMENTGROUP}} = \sum_{i=1}^{m}(p(s_i) - d(s_i)) + \sum_{j=1}^{q_{eg}} I_j p_{\rho_j}$
Connection rule (s_1, s_2, c_{12}) where s_1 and s_2 are the two consecutive skills and c_{12} is the amount of bonus points to be given.	$f_{\text{CONNECTION}}(S) = \sum_{i=1}^{m}(p(s_i) - d(s_i)) + \sum_{i=1}^{m-1} c_{s_i, s_{i+1}}$

SCORING RULES

Basic compositional scoring on a routine $S = s_1 \dots s_m$ where $p(s_i)$ is the point value of the skill and $d(s_i)$ is the deductions for improper form on that skill.	$f_{\textsc{Basic}}(S) = \sum_{i=1}^{m}(p(s_i) - d(s_i))$
Anti-repetition rule (ρ, k) where $\rho \subseteq \Sigma$ and k ∈ N. For $S = s_1 \dots s_m$, define $R = r_1 \dots, r_m$ as a bitstring specifying if each skill is recognized for points and must satisfy Anti-repetition rules.	$f_{\textsc{Anti-Repetition}}(S) = \max_{\text{valid } R} \sum_{i=1}^{m}(p(s_i) \cdot r_i - d(s_i))$
Element group rule (ρ, p_ρ) where $\rho \subseteq \Sigma$ and p_ρ is the point value associated. Let indicator variable I_j specify whether EG rule *j* is satisfied.	$f_{\textsc{ElementGroup}} = \sum_{i=1}^{m}(p(s_i) - d(s_i)) + \sum_{j=1}^{q_{eg}} I_j p_{\rho_j}$
Connection rule (s_1, s_2, c_{12}) where s_1 and s_2 are the two consecutive skills and c_{12} is the amount of bonus points to be given.	$f_{\textsc{Connection}}(S) = \sum_{i=1}^{m}(p(s_i) - d(s_i)) + \sum_{i=1}^{m-1} c_{s_i, s_{i+1}}$
Incomplete Graph rule is an adjacency matrix *C* of a directed graph. All routines must be paths on the adjacency matrix to score non-zero points.	$f_{\textsc{IncompleteGraph}}(S) = \left(\sum_{i=1}^{m}(p(s_i) - d(s_i))\right) \prod_{i=1}^{m-1} c_{s_i, s_{i+1}}$

EASY REDUCTIONS BETWEEN CLASSES

- IncompleteGraph reduces to Connection because you can encode an unweighted incomplete graph as a weighed complete graph with negative infinite weights on missing edges.
- IncompleteGraph+Connection can still be reduced to just Connection.
- ElementGroup reduces to AntiRepetition because element groups have to be non-overlapping. Thus, we can create a copy of all the skills. Within the copy of the skills, cannot repeat any element group for points, so only makes sense to use the "copied" skill once to get the element group bonus.
- In fact, ElementGroup+AntiRepetition can be reduced to just AntiRepetition.

(NON)HIERARCHICAL ANTI-REPETITION

▶ **Definition 3** (*Hierarchical* ANTI-REPETITION structure). *Consider a set of* ANTI-REPETITION *rules* $\{(\rho_1, k_1), \ldots, (\rho_q, k_q)\}$. *If there exists a pair* (ρ_i, ρ_j) *where* $\rho_i \cap \rho_j \neq \varnothing$ *and* $\rho_i \cap \rho_j \neq \rho_i$ *and* $\rho_i \cap \rho_j \neq \rho_j$, *then the* ANTI-REPETITION *rules are Non-hierarchical.*

- Note that we consider *Non-hierarchical* to imply *Hierarchical* because there is always a subset of the *non-hierarchical* rules that is *hierarchical.*
- We make this distinction because this significantly changes the complexity classes, and very often Codes of Points with even complicated anti-repetition rules are hierarchical.
- Aside: we use the distinction *hierarchical* instead of *non-overlapping*, because this allows hierarchical to include prohibiting repeating individual skills, as well as element group bonuses, and a limit on the total number of skills allowed.

HIERARCHICAL ANTI-REPETITION IS IN P

- We solve this by transforming the problem into a minimum-cost maximum-flow problem, which can be solved in polynomial time via linear programming.
- Encode all of the anti-repetition rules into a tree (possible by hierarchy), with the top-level root corresponding to anti-repetition on the class of all skills (i.e. the length of the routine) and the bottom-level leaves being anti-repetition on individual skills. Each node's capacity is the number of times it can be repeated.
- Use standard in- and out- node duplication trick to convert node capacities into edge capacities for flow problem, and assign a negative weight according to the point values.
- We then introduce a source pointing at the leaves and a target coming from the root.

NON-HIERARCHICAL ANTI-REP IS NP-HARD

- Reduction from positive one-in-three-SAT (1-in-3-SAT+), where we are given a family of Boolean variables and a collection of triples. The task is to determine if there exists an assignment of the variables such that each triple has exactly one true variable.
- Our reduction encodes each variable as a skill worth 1 point.
- The clauses are encoded as a pair Anti-repetition rules such that each skill within a clause can only be performed once for credit, but also if no skill from a clause is performed, there is a large penalty.
  - Anti-repetition rule with three skills in a similarity class, allowed to repeat once.
  - Anti-repetition rule with all other skills in a similarity class, but allowed to repeat m-1 times, where m is the length of the routine.
- Routine worth m points thus gives a solution to 1-in-3-SAT+.

CONNECTION IS IN P

- Recall that routines are of fixed length *m.* We can therefore use dynamic programming.
- Consider fully connected directed graph with weights on the nodes corresponding to skills and weights on the edges corresponding to connection values.
- For any given starting node, can find optimal path of length m in $O(n^3m)$ time.

- Corollary: IncompleteGraph is also in P.
- Corollary: Connection+IncompleteGraph is in P.

COMPLEXITY OF RULE CLASSES

	Non-hierarchical ANTI-REPETITION	*Hierarchical* ANTI-REPETITION	CONNECTION	INCOMPLETE GRAPH
Non-hierarchical ANTI-REPETITION	NP-hard Thm 6			
Hierarchical ANTI-REPETITION		In P Thm 5		
CONNECTION			In P Thm 7	
INCOMPLETE GRAPH				In P Cor 8

COMPLEXITY OF RULE CLASSES

	Non-hierarchical Anti-Repetition	*Hierarchical* Anti-Repetition	Connection	Incomplete Graph
Non-hierarchical Anti-Repetition	NP-hard Thm 6	NP-hard Thm 6	NP-hard Thm 6	NP-hard Thm 6
Hierarchical Anti-Repetition		In P Thm 5		
Connection			In P Thm 7	In P Cor 11
Incomplete Graph				In P Cor 8

HIERARCHICAL ANTI-REP+INCOMPLETE GRAPH IS NP-HARD

- Reduction from Hamiltonian path, where the goal is to find a a simple path through a directed graph such that every node is visited exactly once.
- Encode each node as a skill, with point value 1 and deduction 0. Add an anti-repetition rule for every skill so that we only get credit for each skill once.
- Now let's set *m=n.* A routine scoring *m* points thus must go through every skill exactly once, while respecting the incomplete graph.

- Corollary: Hierarchical Anti-rep+Connection is also NP-Hard.

COMPLEXITY OF RULE CLASSES

	Non-hierarchical ANTI-REPETITION	*Hierarchical* ANTI-REPETITION	CONNECTION	INCOMPLETE GRAPH
Non-hierarchical ANTI-REPETITION	NP-hard Thm 6	NP-hard Thm 6	NP-hard Thm 6	NP-hard Thm 6
Hierarchical ANTI-REPETITION		In P Thm 5		
CONNECTION			In P Thm 7	In P Cor 11
INCOMPLETE GRAPH				In P Cor 8

COMPLEXITY OF RULE CLASSES

	Non-hierarchical ANTI-REPETITION	*Hierarchical* ANTI-REPETITION	CONNECTION	INCOMPLETE GRAPH
Non-hierarchical ANTI-REPETITION	NP-hard Thm 6	NP-hard Thm 6	NP-hard Thm 6	NP-hard Thm 6
Hierarchical ANTI-REPETITION		In P Thm 5	NP-hard Cor 10	NP-hard Thm 9
CONNECTION			In P Thm 7	In P Cor 11
INCOMPLETE GRAPH				In P Cor 8

CHOICE OF SPORTS TO CLASSIFY

- Olympic Sports only, otherwise would have to determine how to classify various styles of competitive dance.
- Some sports were excluded for not having objective Codes of Points:
  - Equestrian (Dressage)
  - Snowboarding (Big Air, Halfpipe, Slopestyle)
- Some sports were excluded for only allowing routines of at most 1-2 skills
  - Diving
  - Vault (in artistic gymnastics)
  - Double mini-trampoline

SKIING

- Most skiing events scored on time, so irrelevant for complexity.

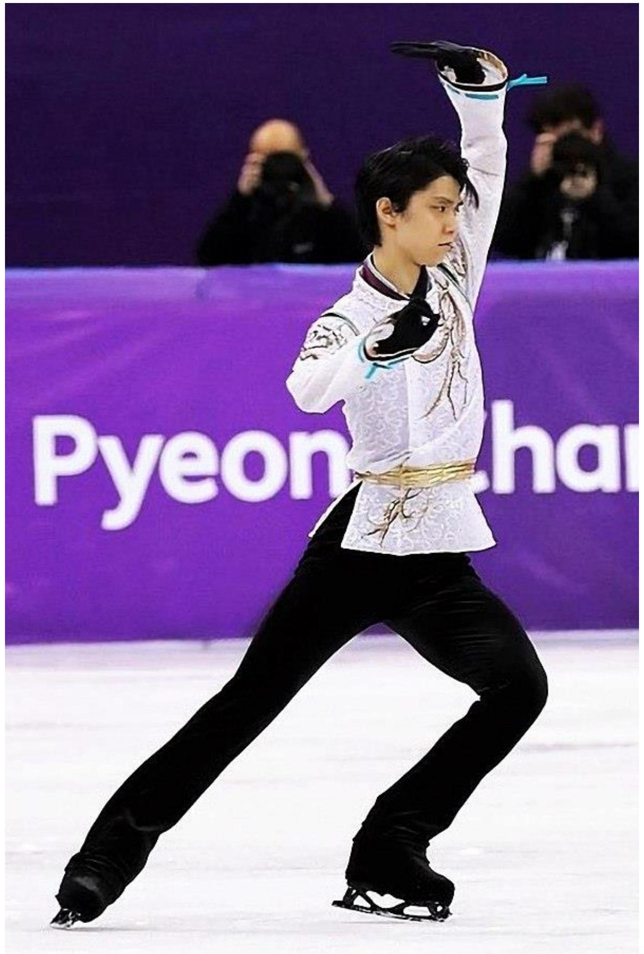
- Left with 5 freestyle events based on aerial tricks:
  - Aerials and Big Air: twists and turns during single jump off ramp
  - Slopestyle and Halfpipe: series of tricks going down a course
  - Mogul: skiing around bumps and performing occasional tricks of ramps (Air component identical to other events)

- The is an Incomplete Graph – e.g. impossible to flip both backwards and forwards in a single jump.

- No real anti-repetition rule – athletes can repeat the same trick
  - Slight complication with Aerials/Big Air because two routines allowed.
  - Athletes “encouraged” to do different tricks in Slopestyle and Halfpipe, but this is not enforced by objective scoring.

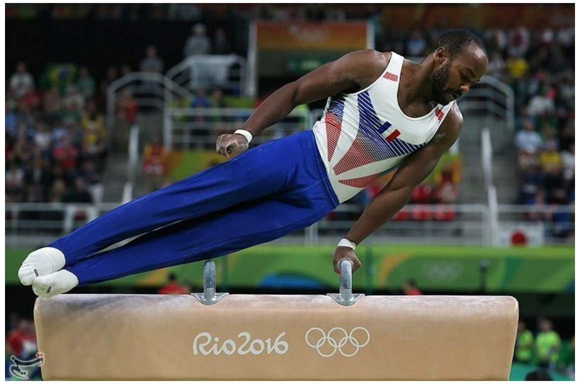
FIGURE SKATING

- Focus on Free Skate (Men’s, Women’s, Pair)
  - Dance events focus on artistry and disallow difficult tricks.
  - Short Program requires performing 7 required skills, so technically Hamiltonian path, but athletes are allowed to reset position so it’s a complete graph.
- We will ignore artistic score and focus only on technical Elements score.
- Jump combinations are treated as higher scoring individual skills, so simply increase the space of skills by a polynomial factor.
- Individual skills may be repeated, but skills within each of four categories can only be repeated a limited number of times. Thus, Hierarchical Anti-Repetition.
- 10% fatigue bonus for skills in second half, but our routine construction rules for Hierarchical Anti-Repetition don’t care about order, so can place higher scoring skills later.



GYMNASTICS

- Lots of different events across rhythmic gymnastics, trampoline, and artistic gymnastics—we only look at the ones with nontrivial routine construction and ignore artistic scores.
- Combinations of anti-repetition, element group, connection, and incomplete graph rules.
- In actuality a lot of additional special case rules that do not fall into those four categories, but luckily does not matter for NP-hardness.

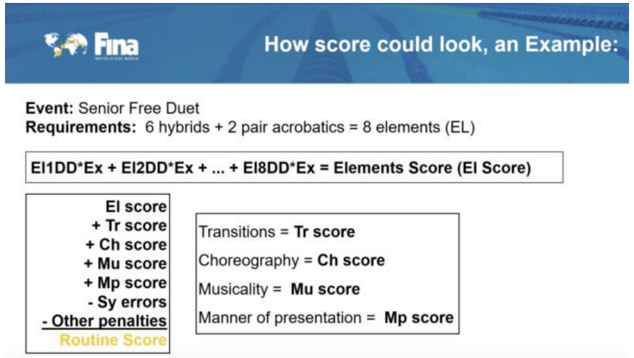


CLASSIFICATION OF ROUTINE SPORTS

Sport/event	*Hierarchical* ANTI-REPETITION	*Non-hierarchical* ANTI-REPETITION	CONNECTION	INCOMPLETE GRAPH	Complexity
Skiing (4 events)	N	N	N	Y	P
Figure skating (Free Skate, Single and Pairs)	Y	N	N	N	P
Rhythmic gymnastics (Individual and Team)	Y	N	(?)	Y	NP-hard
Trampoline	Y	N	N	Y	NP-hard
MAG Floor	Y	N	Y	Y	NP-hard
Pommel Horse	Y	Y	N	Y	NP-hard
Rings	Y	N	N	Y	NP-hard
Parallel bars	Y	N	Y	Y	NP-hard
High Bar	Y	N	Y	Y	NP-hard
WAG Floor	Y	Y	Y	Y	NP-hard
Balance Beam	Y	N	Y	Y	NP-hard
Uneven bars	Y	Y	Y	Y	NP-hard

HONORABLE MENTION: ARTISTIC SWIMMING

- Technical routine has degrees of difficulty, element groups, etc. However, the set and sequence of skills is fixed for all competitors, so there is no sense of routine construction.
- Free routine and highlight routine allow athletes to construct their own routines, but don’t have objective scoring component.
- Proposed 2022 artistic swimming code of points introduces element-based scoring for all events, but not yet in place yet.
- They introduce a new type of scoring rule, the systematic creation of “hybrid” skills. Whether the new rules are NP-hard depends on the interaction of hybrids with the other rules.



SUMMARY

- Olympic sports can be FUN!
- Attempts to formalize human artistic and subjective judgment into a comprehensive set of rules leads to NP-hard problems.
- Gymnastics is hard! But skiing and figure skating are (computationally) easy.