

Skript zur Vorlesung  
**Managing and Mining Multiplayer Online Games**  
im Sommersemester 2016

# Kapitel 10: Kollaboratives und Antagonistisches Verhalten

Skript © 2012 Matthias Schubert

[http://www.dbs.ifi.lmu.de/cms/VO\\_Managing\\_Massive\\_Multiplayer\\_Online\\_Games](http://www.dbs.ifi.lmu.de/cms/VO_Managing_Massive_Multiplayer_Online_Games)

# Kapitelüberblick

---

- Berechnen der Spielstärke aus Siegstatistiken
- ELO-Rankings
- True Skill und das Microsoft-Modell
- Team Skill: Berücksichtigung der Chemie im Team
- Ausblick auf Netzwerkanalyse in Spielen

# Modelle für die Spielstärke

**Idee:** Die Spielstärke kann aus den vergangenen Siegen und Niederlagen ermittelt werden.

**Modell:** Jeder Spieler  $i$  hat eine Spielstärke  $s_i$ .  
Ist  $s_i > s_j$  gewinnt  $s_i$  einen Wettkampf mit hoher Wahrscheinlichkeit.

**Anwendungen:**

- **Matchmaking:** Auswahl von interessanten Gegnern mit vergleichbarer Spielstärke.
- **Ladders/Ranking:** Erstellen von öffentlichen Ranglisten als Ausdruck des Prestiges.  
(Vergleiche Tennis, SC2, WOW-Arena, Halo2, ...)
- **Organisation von Turnieren:** Hilfe bei der Auslosung, Qualifikation, Entscheidung in Streitfällen.

LEAGUES & LADDERS

SEASON 2 - 1v1 GRANDMASTER

BONUS POOL: 0

RANK	NAME	POINTS	WINS	LOSSES
1	alibevla	371	55	28
2	karate	317	54	17
3	Makabite	313	52	20
4	nGendall	312	35	21
5	TASenhu	311	73	21
6	nCOOLde	306	17	1
7	NeuFestung	300	9	8
8	nFlicker	299	32	17
9	REGHulle	297	40	16
10	Wesley	297	31	45
11	nGendall	294	24	27
12	Yonitz	292	50	17
13	NICoway	292	53	20
14	alt	291	32	45
15	Pakun	287	60	35

#	Spieler	Punkte	Win%	Leave%	Total	W-D-L (Leaves)
1.	ku5h	440 VS	74%	0.0%	34	25 - 0 - 9 (0)
2.	KevKev	367 VS	53%	0.0%	43	23 - 2 - 18 (0)
3.	GAMEBUG	343 VS	63%	0.0%	24	15 - 4 - 5 (0)
4.	Scasyy	342 VS	54%	0.0%	39	21 - 1 - 17 (0)
5.	FATAL	337 VS	63%	0.0%	30	19 - 1 - 10 (0)
12.	bueli	278 VS	65%	0.0%	23	15 - 0 - 8 (0)
20.	powerhead	244 VS	56%	0.0%	34	19 - 1 - 14 (0)
12.	bueli	278 VS	65%	0.0%	23	15 - 0 - 8 (0)
41.	random	216 VS	63%	0.0%	16	10 - 1 - 5 (0)
48.	afr0	205 VS	59%	0.0%	29	17 - 0 - 12 (0)

Letzte Aktualisierung 02 Jul 2008

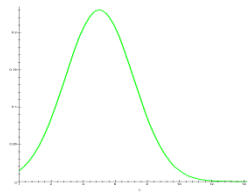
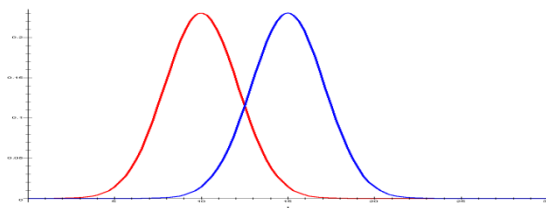
Nächste Veröffentlichung 06 Aug 2008

Rang	Team	P Jul 08	+/- Rang Jun 08	+/- P Jun 08
1	Spanien	1557	3	254
2	Italien	1404	1	-20
3	Deutschland	1364	2	90
4	Brasilien	1344	-2	-169
5	Niederlande	1299	5	188
6	Argentinien	1298	-5	-261
7	Kroatien	1282	8	265
8	Tschechische Republik	1146	-2	-100
9	Portugal	1104	2	10
10	Frankreich	1053	-3	-90

# Das ELO System

---

- Vorgestellt von Arpad Elo 1970 und von der *World Chess Federation* übernommen.
- **Annahme:** Die Performanz  $p_i$  eines Spielers  $i$  ist normalverteilt mit Varianz  $\beta^2$  um seine Spielstärke (Skill)  $s_i$ :  $p_i = N(s_i, \beta^2)$   
 $\Rightarrow s_i > s_j$  bedeutet also nicht zwangsläufig, dass  $i$  gegen  $j$  verliert,  
**sondern:**  $P(i \text{ gewinnt gegen } j) > 50\%$
- **Gesucht:**  $p(p_i > p_j | s_i, s_j)$  (Wahrscheinlichkeit, dass  $i$  besser spielt als  $j$ )  
 $\Rightarrow$  Differenz von 2 normalverteilten Variablen mit gleicher Varianz  $\beta^2$  ist normalverteilt mit Erwartungswert  $s_i - s_j$  und Varianz  $\beta^2$ .



Differenzverteilung von  
 $p_i$  und  $p_j$

- Sei  $\Phi$  die kumulierte Dichtefunktion einer Normalverteilung mit Erwartungswert 0 und Varianz 1, dann gilt:  $P(p_1 > p_2 | s_1, s_2) = \Phi\left(\frac{s_1 - s_2}{\sqrt{2}\beta}\right)$

# Updates beim ELO Ranking

---

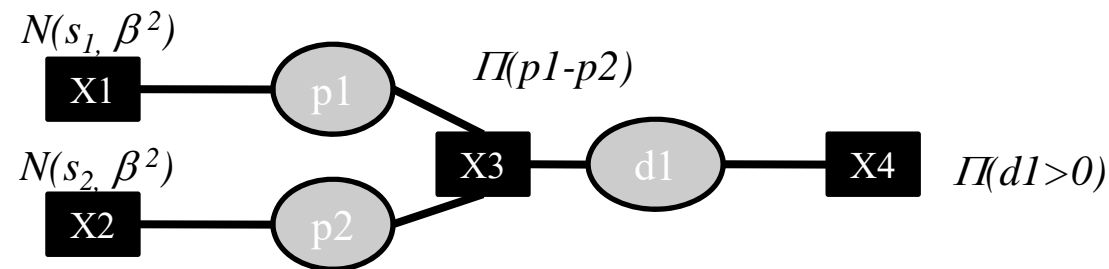
- Positionen müssen bei Bekanntwerden neuer Ergebnisse angepasst werden.
- Änderungen erfolgen nach dem Nullsummenprinzip  $s_1^{new} + s_2^{new} = s_1 + s_2$ .
- Änderung  $\Delta$  soll die Wahrscheinlichkeit der Beobachtung im Modell erhöhen.  
Ergebnis einer Partie:  $y \in \{0, -1, 1\}$  (Sieg:1, Niederlage:-1, Unentschieden:0)

- Update des ELO Scores mit Ergebnis  $y_l$ :
$$\Delta = \alpha\beta\sqrt{\pi}\left(\frac{y_l + 1}{2} - \Phi\left(\frac{s_1 - s_2}{\sqrt{2}\beta}\right)\right)$$

- $\alpha$  : Gewichtungsfaktor für ein Match  $0 < \alpha < 1$  (im Schach ca. 0.07)
- ELO Scores brauchen relativ viele Matches um sich zu stabilisieren. (ca. 20)
- Eigenschaften:
  - Reihenfolge der Updates spielt eine Rolle: Gut bei langen Intervallen zwischen den Messungen, aber bei Turnieren, bei denen man den Skill eines Spielers als konstant annehmen kann eher schlechte Performanz.
  - ELO System erlaubt keine Rückschlüsse auf Einzelleistungen bei Teamspielen.
  - Ergebnisdarstellung ist eingeschränkt. Keine differenzierte Betrachtung bei Wettkämpfen mit Ranking als Ergebnis (z.B. Autorennen, ...).

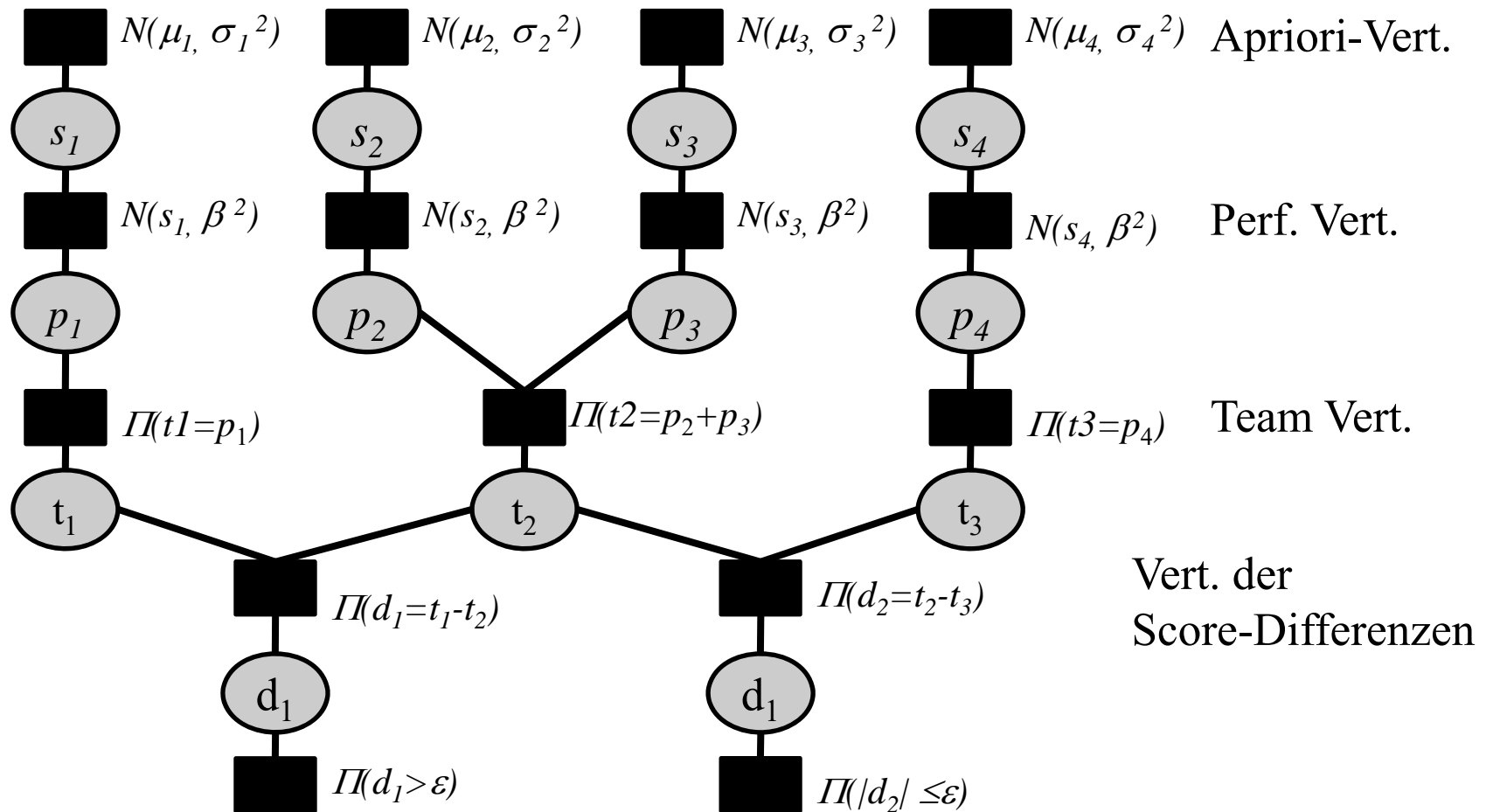
# True Skill

- Darstellung mittels Faktorgraphen:  
Bipartiter Graph mit Faktor und Variablen-Knoten.
  - Variablen-Knoten: Darstellungen von Verteilungsfunktionen
  - Faktor-Knoten: Modellieren das Zusammenwirken von Variablen
  - Kanten: Beschreiben die Variablen ,die für einen Faktor zusammenwirken
  - **Beispiel:** Faktor Graph für ELO System



- **True Skill:** Erweiterung des ELO Systems, das in der XBOX360 Live verwendet wird (z.B. Ranking in HALO2)
- **Berücksichtigt:**
  - Unsicherheit des Skills
  - Erlaubt Rückschlüsse auf Mitglieder in Teamspielen (Additive Performanz  $t_1$  )
  - Ergebnisdarstellung als Ordnung von Spielergebnissen ( $t_1 \geq t_2 \geq \dots \geq t_m$  )

# Faktorgraph für True Skill



**Beispiel:** 4 Spieler, 3 Teams:  $\{(s_1), (s_2, s_3), (s_4)\}$

Ergebnis:  $t_1 > \varepsilon + t_2$ ,  $t_1 > \varepsilon + t_3$ ,  $\varepsilon > |t_2 - t_3|$

# Verwendung des Faktorgraph für True Skill

---

- Faktorgraph stellt die Verteilung für  $Pr(s,p,t/r,A)$  dar
  - **r**: Ergebnisranking, **A**: Teamzusammensetzung
  - **s**: Spieler-Skill, **p**: Spielerperformanz, **t**: Teamwertung
- Gesucht ist die Verteilung des Spieler-Skills  $s$ :

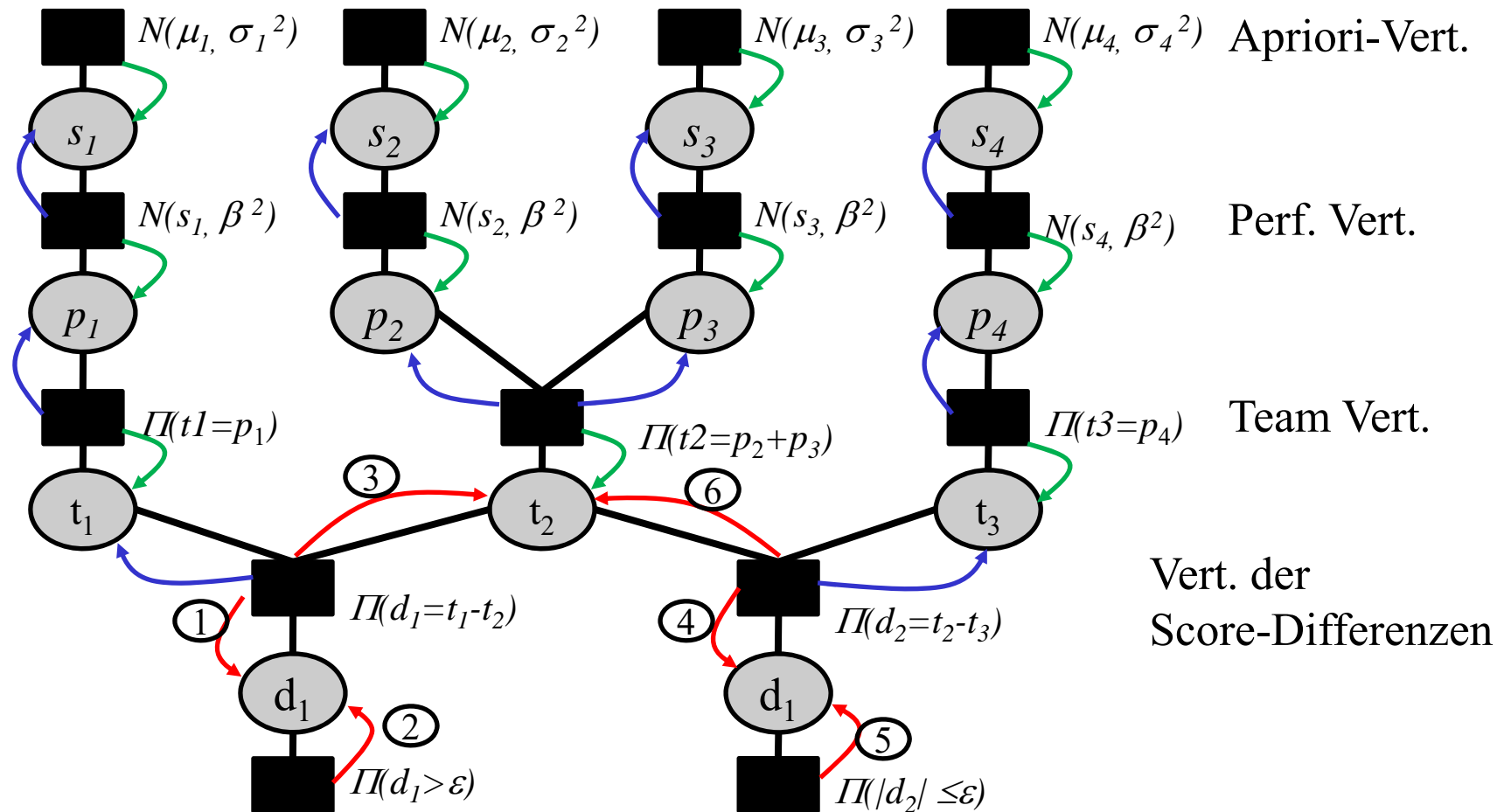
$$Pr(s | r, A) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Pr(s, p, t | r, A) dp dt$$

$s_i$  ist normalverteilt mit Erwartungswert  $\mu_i$  und Standardabweichung  $\sigma_i$

- Mit den gegebenen Faktorgraphen und den aktuellen Werten für  $\mu$  und  $\sigma$  für die an einem Match beteiligten Spieler, lässt sich  $\Pi(d_1 > \varepsilon)$  und  $\Pi(|d_2| \leq \varepsilon)$  abschätzen.
- Vergleicht man jetzt die Vorhersage, mit dem tatsächlichen Ergebnis, dann kann man den Fehler zurück in Richtung  $\mu$  und  $\sigma$  propagieren und so anpassen.
- Das Propagieren von Wahrscheinlichkeiten und Parameter Updates in einem Faktorgraphen nennt man auch Message-Passing oder Belief Propagation.



# Trainingsschemata für True Skill



# Diskussion True Skill

---

- Verbesserung des ELO Systems durch:
  - Verbreiterung der Ergebnisdarstellung
  - Schnellere Konvergenz durch Verwendung von Apriori-Verteilungen für die einzelnen Spieler
  - Bewertung von Teams
- Nachteile von True Skill:
  - Zeitliche Reihenfolge spielt eine Rolle, selbst wenn nicht davon auszugehen ist, dass sich der Skill zwischen 2 Matches verändert hat. (Erweiterung: True Skill Trough Time 2008)
  - Teams werden als frei kombinierbar mit additiver Performanz wahrgenommen.  
(aber Zusammenwirken der Spieler ist in der Realität komplizierter:  
11 Messis  $\neq$  beste Fußballmannschaft der Welt)

# Team Skill

---

**Idee:** Berücksichtige nicht nur individuelle Spielstärke, sondern auch die Chemie im Team.

- => Betrachtung der gemeinschaftlichen Performanz im Vergleich zur durchschnittlichen Einzelperformanz eines Spielers.
- => Manche Spieler zeigen in Kombination mit bestimmten Spielern eine bessere Performanz als in Teams mit anderen.

**Gegeben:** Ein Team  $T = \{p_1, \dots, p_K\}$  mit  $K$  Spielern. Sei  $t_k$  eine  $k$ -elementiges Sub-Team von  $T$ . ( $t_k \subseteq T \wedge |t_k| = k$ ).  $Skill(t_k)$  bezeichnet die Spielstärke des Subteams  $t_k$  (zum Beispiel mit ELO oder True Skill berechnet)

**Gesucht:** Wie stark ist das gesamte Team  $T$  unter Berücksichtigung der Team-Chemie?

**Ansatz:** Bestimmen der Subteamwertungen und bilden des Durchschnitts.

# Team Skill-k

---

- Durchschnittliche Spielstärke im Subteam der Größe k auf K skaliert

$$TS_k(T) = K \cdot \frac{1}{k} \cdot \frac{1}{\binom{K}{k}} \cdot \sum_{i=1}^{\binom{K}{k}} Skill(s_{ki}) = \frac{(k-1)!(K-k)!}{(K-1)!} \cdot \sum_{i=1}^{\binom{K}{k}} Skill(s_{ki})$$

- Beispiel:**

k=1 und K=5

$$TS_k(T) = \frac{5}{1} \cdot \frac{1}{\binom{5}{1}} \cdot \sum_{i=1}^{\binom{5}{1}} Skill(s_{1i}) = \sum_{i=1}^5 Skill(s_{1i})$$

k=2 und K=5

$$TS_k(T) = \frac{5}{2} \cdot \frac{1}{\binom{5}{2}} \cdot \sum_{i=1}^{\binom{5}{2}} Skill(s_{2i}) = \frac{1}{4} \sum_{i=1}^{10} Skill(s_{2i})$$

# Team Skill-AIIK-LS

---

Verbesserungsansätze gegenüber Team Skill  $k$ :

- Bestimmung von  $k$  ist schwierig  $\Rightarrow$  nimm alle möglichen Subteams.
- nicht für alle Subteams gibt es separate Ergebnisse  
 $\Rightarrow$  Verwende nur die Subteams, für die eine verlässliches Ranking vorliegt.

**Idee:** Betrachte alle Subteams für die eine verlässliche Schätzung vorliegt und die selber keine Teilmenge eines verlässlich abgeschätzten Subteams sind.

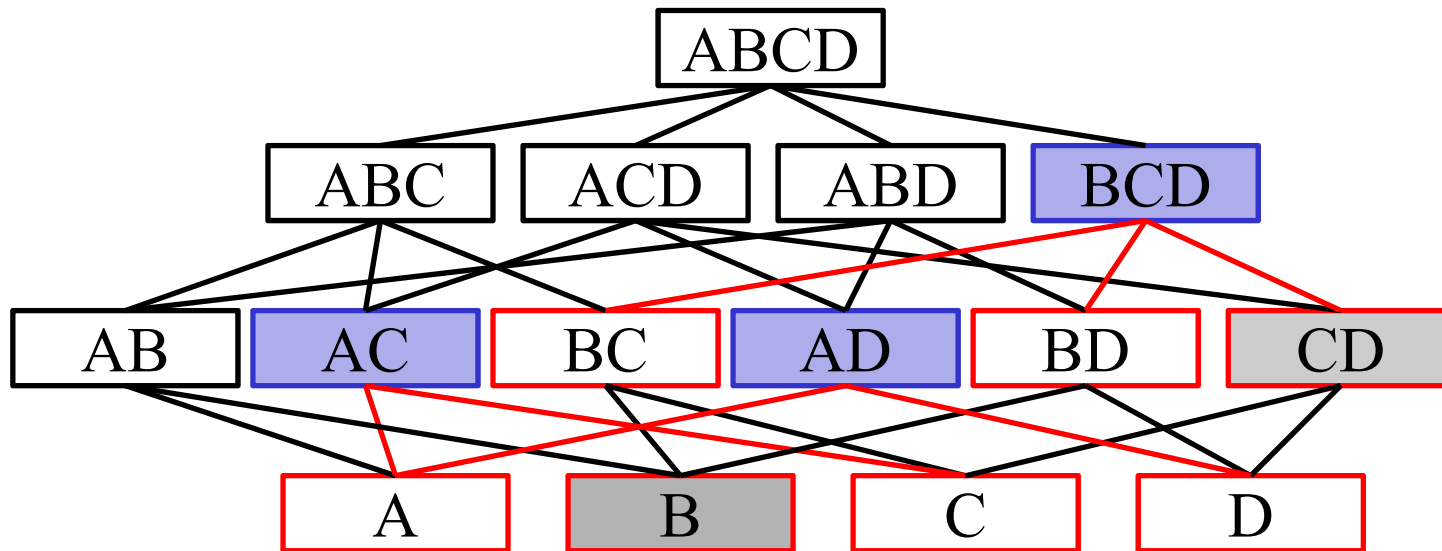
**Vorgehen:** Bestimme alle relevanten Subteams  $t_{k,i}^*$  für die  $Skill(t_{k,i})$  bestimmt werden kann und für die kein Subteam  $t_{k+l,j} \supset t_{k,i}$  existiert.

Berechne die Teamperformanz als  $K$ -fache durchschnittliche Einzelperformanz.

$$TS_{ALL-LS}(T) = \frac{K}{\sum_{m \in \{m | \exists t_m^* \neq \{\}\}} |m|} \left( \sum_{m \in \{m | \exists t_m^* \neq \{\}\}} \left( \frac{1}{l} \cdot \sum_{i=1}^l Skill(t_{m,i}^*) \right) \right)$$

# Beispiel: Team Skill ALL-LS

---



rot: geprunter Bereich, blau: verwendete Subteams, grau: geprunte Subteams.

$$TS_{ALL-LS}(T) = \frac{4}{3+2} \left( Skill(t_{BCD}) + \frac{1}{2} (Skill(t_{AC}) + Skill(t_{AD})) \right)$$

# Fazit

---

- Methode kann den erhöhten Erfolg von Teams mit guter Chemie darstellen.
- Damit Team Skill funktioniert, benötigt es Daten von möglichst vielen unterschiedlichen Teamzusammensetzungen.
- Für echte Bewertung von Teams ist das Modell aber zu einfach, da:
  - Es keine Rollenverteilungen im Team betrachtet
  - Es die Sicherheit der Teilaufgaben nicht berücksichtigt (vgl. Varianz im True Skill)
  - die Zuordnung des Skills auf die Spieler eines Subteams immer gleichverteilt erfolgt.
- Für den Skill in Team Skill, True Skill und ELO gilt eine symmetrische Bewertung zwischen Sieg und Niederlage.  
=> Häufig ist allerdings eine asymmetrische Bewertung (Punkte Sieg > Punkte Niederlage) gewünscht, um die Spieler zu motivieren.

# Alternativer Lösungsansatz

---

- Bewerte Spieler nicht nach Erfolg, sondern danach, ob er sich wie ein erfolgreicher Spieler verhält:
  1. Sammle und beschreibe zeitlich-räumliches Verhalten von Spielern im gesamten Spektrum des Könnens.
  2. Lerne ein Regressionsmodell.
  3. Bewerte Spieler während er spielt aufgrund seiner  $k$  letzten Handlungen.
- Ansatz wird für dynamische Spielstärkenanpassung im PVE verwendet.
- Eignet sich gut, wenn klar ist wie erfolgreiches Verhalten im Spiel aussieht.  
(z.B. Zielgenauigkeit in FPS Games, DPS/HPS Zahlen in MMORPGS)



# Netzwerkanalyse in Spielen

---

- Viele MMO-Spiele enthalten soziale Gefüge die ebenfalls analysiert werden können:

*Wer spielt mit wem und wie lange?*

- Modellierung von Teamstrategien
- Reaktionsschemata auf Gegneraktionen
- Finden krimineller Vereinigungen (z.B. Goldfarmerringe)
- Tools zur Bildung von Zufallsgruppen

# Lernziele

---

- Anwendungsbereiche für Spielerrankings und Matchmaking
- ELO
- True Skill
- Team Skill

# Literatur

---

- A.E. Elo: **The Rating of chess players: Past and present**, Arco Publishing, New York, 1978.
- Pierre Dangauthier, Ralf Herbrich, Tom Minka, Thore Graepel  
**TrueSkill Through Time: Revisiting the History of Chess**,  
In Advances in Neural Information Processing Systems 20 (NIPS), 2008.
- Ralf Herbrich, Tom Minka, and Thore Graepel, **TrueSkill(TM): A Bayesian Skill Rating System**, in Advances in Neural Information Processing Systems 20 , MIT Press, January 2007.
- Colin DeLong, Nishith Pathak, Kendrick Erickson, Eric Perrino, Kyong Shim, Jaideep Srivastava: **TeamSkill: Modeling team chemistry in online multi-player games**, on Proc. of the 15th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD2011), 2011.