

Costruzione di curve di riferimento Lichess stratificate per Elo+Chess

Elo+Chess Note di ricerca

31 maggio 2026

Abstract

I rapporti sui benchmark Elo+Chess sono costruiti da grandi campioni stratificati di Giochi Lichess.org, con il campione di benchmark utilizzato in questa nota contenente 483.974 giochi bullet 1+0, 479.974 giochi blitz 3+0, 925.636 giochi rapidi 10+0, e 660.540 partite rapide più lunghe. Per ciascun tipo di gioco principale, la pipeline si espande ogni gioco negli stati del tabellone prima e dopo la mossa, ne calcola centinaia di variabili intermedie e del gioco del giocatore, aggrega tali variabili in parametri di coaching, quindi calcola la media di tali parametri in base al bucket Elo Lichess. Il campione è deliberatamente stratificato anziché proporzionale al giocatore grezzo pool: lo scopo è quello di stimare curve di comportamento stabili ad ogni livello di rating, compresi i secchi che altrimenti sarebbero scarsi, per non riprodurre il naturale distribuzione del rating. Il risultato è una famiglia di curve benchmark empiriche: per ogni metrica, possiamo chiedere quanto spesso i giocatori a diversi livelli Elo mostrano un determinato valore abitudine o principio nei giochi reali. Questa nota descrive il metodo, la scala di calcolo e diversi esempi di principi di apertura che coinvolgono l'arrocco, lo sviluppo dei pezzi minori, mosse ripetute del pezzo di apertura e movimento iniziale della regina. L'enfasi è metodologico: Elo+Chess non presuppone che i principi degli scacchi siano universali lineari a tutti i livelli di abilità. Invece, le curve di riferimento vengono ispezionate da gamma di valutazione e il prodotto si concentra sulla regione da principiante a avanzato dove le relazioni sono più forti e più interpretabili.

1 Scopo

L'obiettivo del sistema di benchmark è convertire i consigli scacchistici generali in quantità misurabili. Un'affermazione come "sviluppa i tuoi pezzi minori" lo è utile, ma diventa più fruibile quando possiamo dire quanti pezzi minori a il giocatore di solito si sviluppa con la mossa 10, come si confronta con i pari nella stessa fascia di rating e come tale abitudine cambia nello spettro di rating.

Elo+Chess crea quindi una libreria di metriche da giochi reali. Ogni metrica è calcolata per ogni riga del gioco del giocatore, riepilogato per tipo di gioco e gruppo Elo, e utilizzata come curva di riferimento nel rapporto. Il rapporto può quindi inserire un utente la stessa curva e mostra se l'utente si trova sopra, sotto o vicino al comportamento osservato tra giocatori comparabili.

2 Dati di origine e stratificazione

La fonte del benchmark è il database di giochi pubblico Lichess.org. Elo+Chess utilizza campioni stratificati di grandi dimensioni per tipo di gioco e intervallo di valutazione anziché piccoli campioni di convenienza. Gli attuali artefatti del runtime di produzione coprono quattro principali categorie di giochi:

- 1+0 punto elenco,
- 3+0 lampo,
- 10+0 rapido,
- partite rapide di durata superiore a 10 minuti.

L'obiettivo della stratificazione è l'ampiezza tra i livelli di rating. Un casuale proporzionale campione sarebbe efficiente per descrivere il pool complessivo di giocatori, ma lo farebbe essere inefficiente per costruire curve di riferimento. Le categorie di rating più comuni dominerebbero il campione, mentre i bucket di coda a basso supporto ne avrebbero troppo pochi giochi per medie metriche stabili. Elo+Chess quindi campiona per selvaggina media Secchio Elo. Nell'attuale costruzione del campione pulito, i giochi grezzi sono i primi filtrati in giochi con valutazione valida con giocatori conosciuti, valutazioni note, risultati validi, e controlli temporali noti. I giochi vengono quindi filtrati in base ai giocatori con un numero sufficiente di giochi in quella velocità e valutazioni sufficientemente stabili da rappresentare un livello di abilità coerente: per ciascun giocatore, Elo+Chess calcola l'intervallo di valutazione della velocità entro il giocatore ($\max(\text{Elo}) - \min(\text{Elo})$) nel pool di origine e mantiene giochi solo quando entrambi i giocatori hanno almeno il conteggio minimo di giochi configurato e non più di 300 punti Elo di deriva entro la velocità. Il filtro antideriva non è previsto rimuovere il normale miglioramento o la varianza; riduce i casi in cui un giocatore i giochi mescolerebbero livelli di abilità materialmente diversi, come conti provvisori, giocatori che ritornano o account che si sono spostati rapidamente attraverso diversi periodi durante il periodo campionato. Infine, i giochi idonei vengono classificati all'interno di ciascuna media Elo bucket tramite un hash deterministico dell'ID e del seed del gioco e un numero limitato di i giochi vengono conservati da ciascun bucket.

Questo design rende le curve di riferimento molto più utili. Ogni secchio Elo contribuisce con osservazioni sufficienti per stimare il valore medio di ciascuna metrica, e i rari secchi di alta o bassa valutazione non sono soffocati dalla forma naturale della distribuzione del giocatore attivo. Le curve di riferimento utilizzate da Elo+Chess sono poi ristrette alla fascia più attinente al prodotto, grosso modo principiante attraverso il gioco avanzato. Nei report su scala Lichess, l'intervallo primario è 600–1600. Quando un utente visualizza il report sulla scala Chess.com, il i corrispondenti rating bucket sono mappati sullo stesso sottostante Lichess segmenti di benchmark utilizzando il metodo separato di mappatura della valutazione multipiattaforma descritto nel documento allegato: [Estimating Cross-Platform Chess Rating Mappings with Regression modale](#).

3 Distribuzioni delle valutazioni dei giocatori attivi

Se osserviamo la distribuzione delle valutazioni Elo dei giocatori Lichess.org alla fine di marzo 2026, per tipologia di gioco principale, vediamo i seguenti conteggi e percentili. Per questa visualizzazione della distribuzione, ogni giocatore viene conteggiato una volta per tipo di gioco esatto, utilizzando l'ultima valutazione osservata di quel giocatore nel grezzo corrispondente righe di controllo del tempo, dopo aver richiesto almeno cinque partite in quella partita esatta tipo.¹

Questa distribuzione non deve essere confusa con il campione di riferimento stratificato. La distribuzione descrive il pool di giocatori Lichess attivo osservato nel grezzo scansione. Il campione

¹We apply a five-game minimum because lightly active accounts disproportionately remain at the 1500 starting/provisional anchor, creating an artificial spike in all-player latest-rating distributions. Figures 1–4 mostra le distribuzioni con e senza questo filtro.

Game type	Players	Median	75th pct.	80th pct.	% ≤ 1600	Exact 1500	1500 share
1+0 bullet	1,221,146	1,479	1,858	1,951	58.9%	4,756	0.4%
3+0 blitz	1,144,045	1,542	1,845	1,912	54.9%	2,315	0.2%
10+0 rapid	1,679,401	1,381	1,691	1,765	68.3%	3,472	0.2%
>10 rapid	668,255	1,400	1,686	1,753	68.2%	7,969	1.2%

Table 1: Ultima distribuzione della valutazione Lichess del giocatore attivo osservata in base all'esatta tipo di gioco dopo aver richiesto almeno cinque partite in quel tipo di gioco esatto.

di riferimento viene quindi deliberatamente stratificato in modo che Elo venga suddiviso in intervalli avere un supporto sufficiente per curve metriche stabili.

Per mappare da una valutazione Elo Lichess.org per un particolare tipo di gioco a un valutazione Chess.com equivalente, Elo+Chess utilizza la modalità specifica del tipo di gioco equazioni di regressione stimate nel compagno [Elo+Chess rating-mapping paper](#):

$$\begin{aligned}
\hat{R}_{\text{Chess.com,bullet}} &= -531.71 + 0.9871 R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,blitz}} &= -551.54 + 1.0853 R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,10minrapid}} &= -495.04 + 1.0741 R_{\text{Lichess}}, \\
\hat{R}_{\text{Chess.com,>10rapid}} &= -369.02 + 0.9374 R_{\text{Lichess}}.
\end{aligned}$$

Tali equazioni mappano i valori di valutazione, non i ranghi percentili. I ranghi percentili sono definito all'interno di un particolare pool di giocatori e il giocatore Chess.com e Lichess le piscine hanno forme diverse. Nella tabella 2, per il Tipo di gioco rapido da 10 minuti, la prima colonna fornisce un piccolo campione di possibili Chess.com Valori di valutazione Elo. La seconda colonna riporta il corrispondente Classifica percentile Chess.com visualizzata su Chess.com al 31 maggio 2026. A 71,7% il grado su Lichess.org per 10 minuti rapidi corrisponde a una valutazione Elo Lichess di 1,644² e quel valore è mostrato nel file terza colonna. Se applichiamo l'equazione di mappatura Chess.com–Lichess, 1.644 su Lichess equivale a circa 1.271 su Chess.com. Supponendo la stima l'equazione di mappatura è ragionevole, il divario +522 tra la valutazione Chess.com in la colonna 1 e la valutazione mappata nella colonna 4 possono essere spiegate da un numero elevato di giocatori occasionali nel pool di giocatori Chess.com che gonfiano Chess.com rango percentile. In altre parole, trovarsi nel 72° percentile su Lichess.org lo è un risultato più grande che essere nel 72esimo percentile su Chess.com, perché i ranghi percentili non vengono mappati in modo pulito da un sistema all'altro. Elo reale i valori possono essere mappati utilizzando equazioni stimate dai punteggi dello stesso giocatore i due sistemi.

Il motivo delle lacune nella Tabella 2 è che la mappatura stima le valutazioni equivalenti di forza di gioco, mentre un percentile descrive posizione all'interno della popolazione di una piattaforma. Un giocatore del 71,7° percentile in poi Chess.com non è quindi automaticamente un giocatore del 71,7° percentile su Lichess, e viceversa. I numeri osservati sono coerenti con il pubblico di Chess.com pool rapido con una massa molto più grande di giocatori occasionali o con rating inferiore, mentre il pool Lichess a controllo esatto attivo è più concentrato tra le ripetizioni esperte giocatori. Il punto chiave non è che una scala percentile sia "giusta" e l'altra l'altro è "sbagliato"; il fatto è che il rango percentile non è trasferibile su tutte le piattaforme a meno che le distribuzioni dei giocatori sottostanti non abbiano la stessa forma.

²Calculated from the cumulative distribution function of the ZXQPROT3ZXQ 10+0 rapid latest-rating sample after requiring at least five games in that exact game type; see Table 1 and Figure 3.

Chess.com rating	Chess.com pct.	Same-pct. Lichess	Mapped equiv.	Gap
749	71.7%	1,644	1,271	+522
1,012	86.2%	1,871	1,515	+503
1,358	95.4%	2,095	1,755	+397
2,383	99.9%	2,557	2,251	-132

Table 2: Confronto illustrativo tra Chess.com percentile rapido a 10 minuti punti e gli stessi punti percentili nel Lichess 10+0 rapido attivo distribuzione dopo aver richiesto almeno cinque giochi 10+0. “Chess.com mappato equiv.” applica l’attuale Elo+Chess Conversione rapida della scala di valutazione 10+0 nella valutazione Lichess dello stesso percentile. Il le file con la coda più alta sono incluse per mostrare la distinzione della scala, ma quelle montate la linea di valutazione è più affidabile nell’intervallo di coaching principale.

Note on the 1500 spike. The visible jump near 1500 è presente nei dati anziché essere creato mediante livellamento. Nell’ultimo distribuzione delle valutazioni osservata, 1500 valutazioni esatte rappresentano il 13,1% di 1+0 giocatori proiettile, 5,6% di 3+0 giocatori blitz, 9,4% di 10+0 giocatori rapidi e 17,6% dei giocatori rapidi più lunghi. Ciò è coerente con 1500 che funge da comune ancoraggio iniziale o provvisorio nel sistema di rating.

Il picco esatto di 1500 è in gran parte un artefatto di conti poco attivi. A controlla questo, abbiamo ricalcolato le stesse distribuzioni con la valutazione più recente dopo aver richiesto ogni giocatore deve avere almeno cinque partite nello stesso tipo di gioco. Il picco quasi scompare sotto questo filtro: 1500 valutazioni esatte cadono dal 5,6–17,6% di giocatori nella distribuzione per tutti i giocatori allo 0,2–1,2% nelle cinque o più partite distribuzione riportata nella Tabella 1.

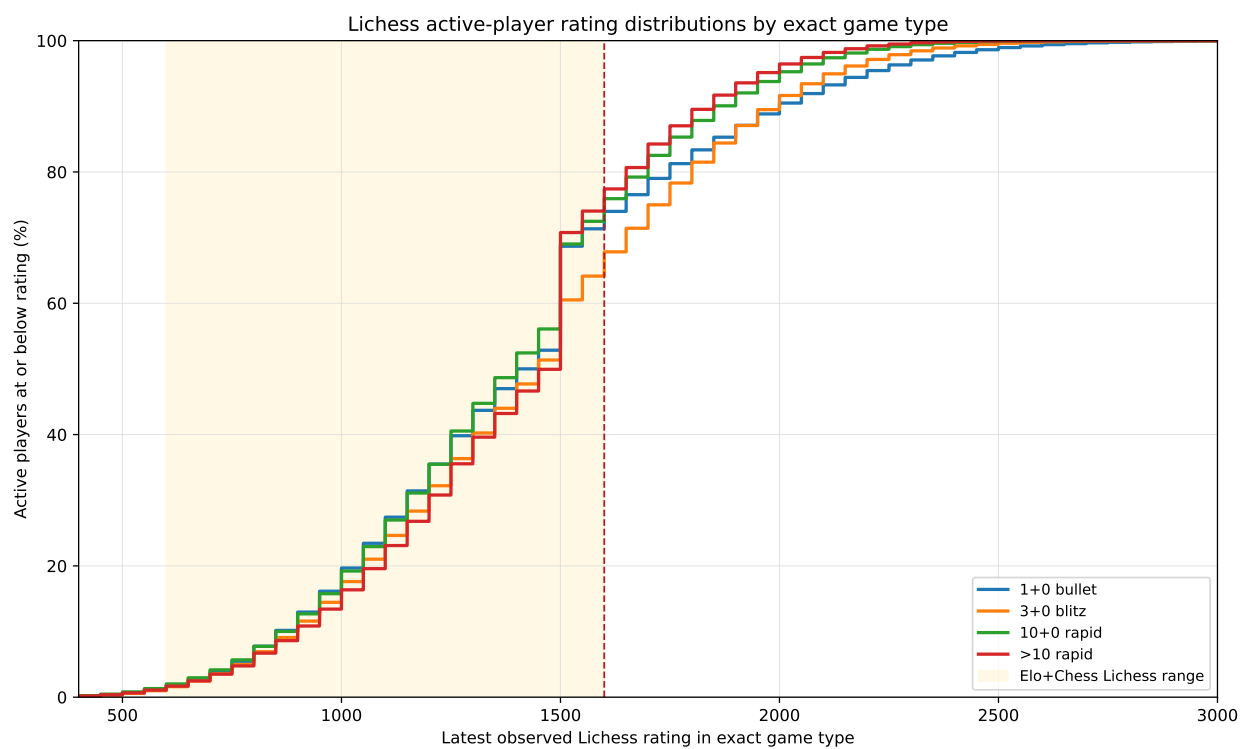


Figure 1: Distribuzioni cumulative della valutazione dei giocatori attivi per tipo di gioco esatto. Il la fascia gialla indica la gamma 600–1600 Lichess utilizzata dal principale Elo+Chess parametri di riferimento dell’allenamento.

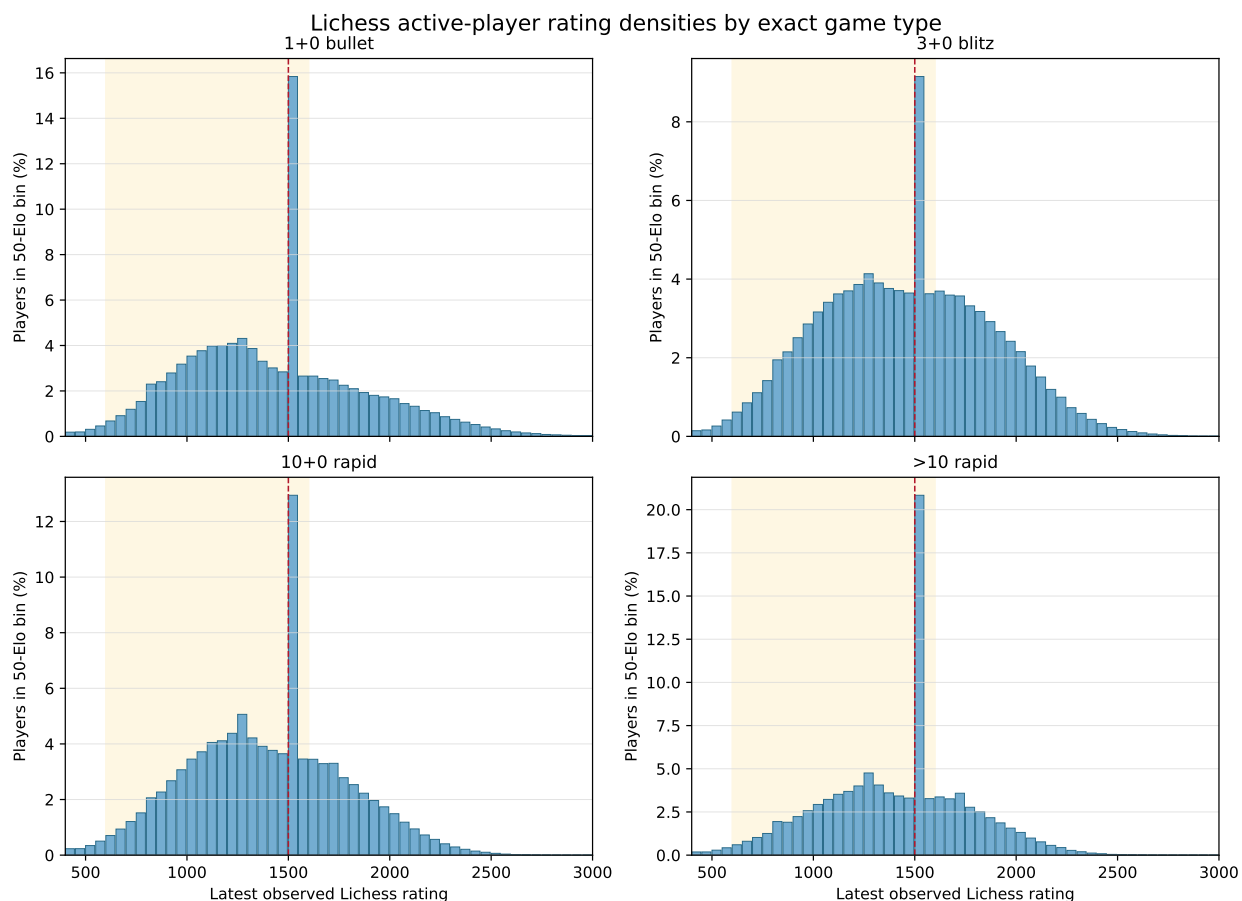


Figure 2: Visualizzazione della densità delle stesse distribuzioni di valutazione dei giocatori attivi. Il la linea tratteggiata segna 1500 Lichess, dove sono presenti le distribuzioni grezze con rating più recente un grande punto di massa di valore esatto. La fascia gialla indica 600–1600 Lichess intervallo utilizzato dai principali report metrici Elo+Chess. Il picco è più visibile in la densità di tutti i giocatori perché molti account leggermente attivi rimangono esattamente allo stesso livello 1500; La figura 4 ripete il grafico della densità dopo aver richiesto almeno cinque giochi nello stesso tipo di gioco, il che rimuove in gran parte questo artefatto.

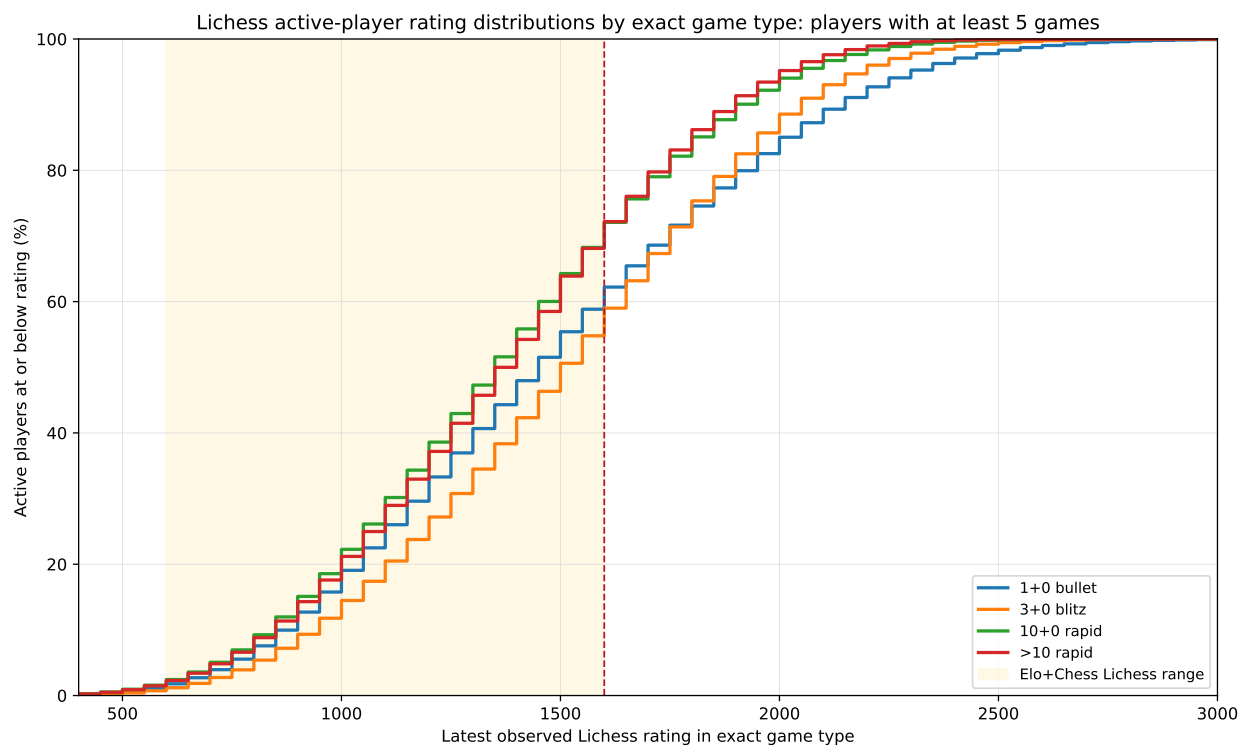


Figure 3: Distribuzioni cumulative delle valutazioni dei giocatori attivi dopo aver richiesto a almeno cinque partite nello stesso tipo di partita. Il salto dei 1500 è notevolmente ridotto.

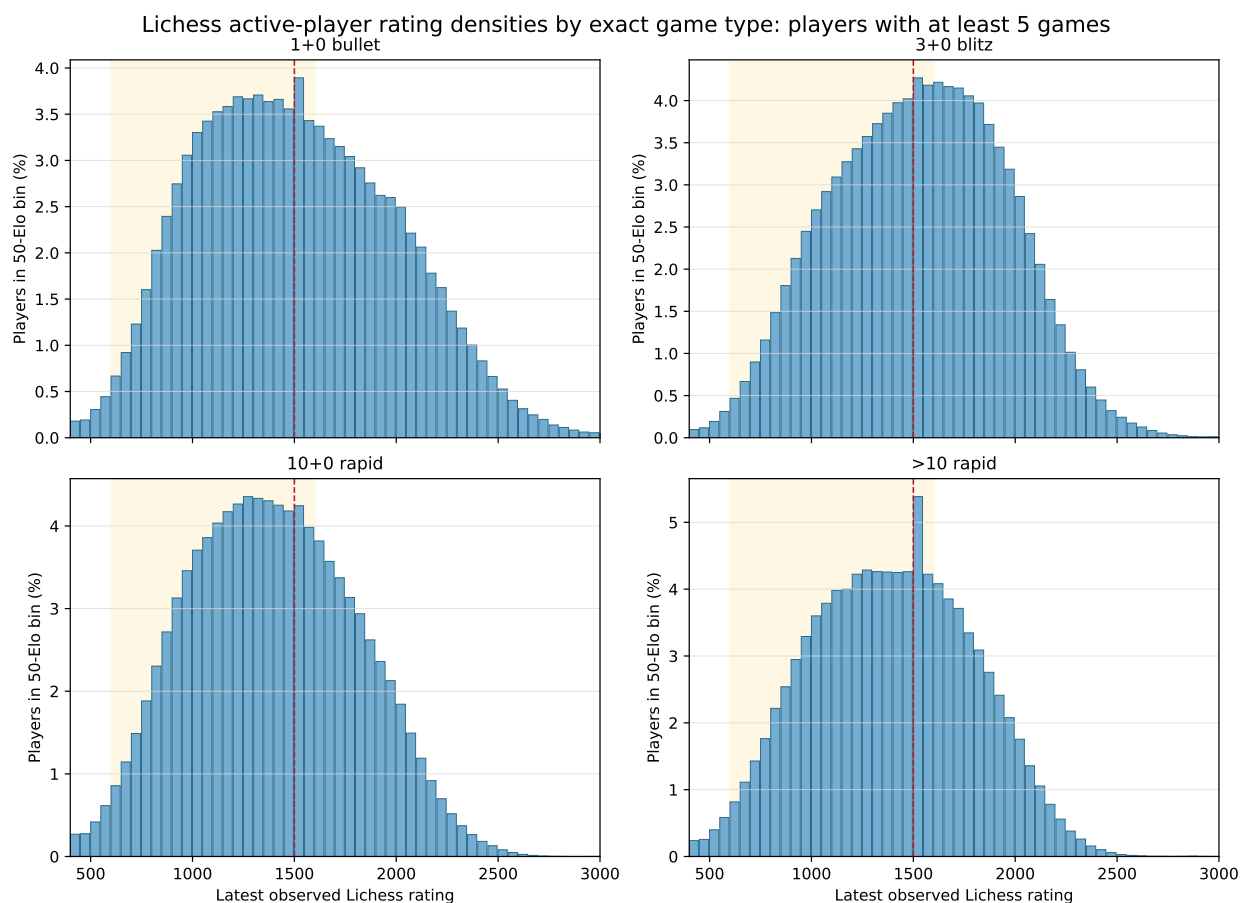


Figure 4: Visualizzazione della densità dopo aver richiesto almeno cinque partite nel gioco esatto tipo. Questa visione conferma che il grande punto di massa 1500 in all-player la distribuzione è per lo più guidata da conti poco attivi. La fascia gialla segna l'intervallo 600–1600 Lichess utilizzato dai principali report metrici Elo+Chess.

4 Scala degli artefatti del benchmark attuale

Il calcolo è intenzionalmente grande. Non è un graffio sul profilo, un piccolo graffio set di PGN scaricati o una raccolta curata manualmente di esempi istruttivi. Il pass Lichess iniziale copriva i file mensili completi da gennaio 2025 fino a marzo 2026. La tabella di scansione non stratificata contiene 1.382.178.087 giochi e 2.764.356.174 righe di valutazione del gioco del giocatore. In quelle righe, 6.623.438 unici i nomi utente dei giocatori appaiono almeno una volta.

Speed	Games in raw scan	Player-game rows	Unique players
Bullet	519,375,423	1,038,750,846	3,077,635
Blitz	839,823,863	1,679,647,726	5,987,634
Rapid	22,978,801	45,957,602	1,535,865
All scanned speeds	1,382,178,087	2,764,356,174	6,623,438

Table 3: Scala di scansione Lichess non stratificata prima del campionamento del benchmark. La scansione copre i file mensili completi della cronologia di gioco da gennaio 2025 a marzo 2026.

Gli artefatti del benchmark di produzione sono molto più piccoli della scansione grezza perché sono selezionati per la costruzione metrica. Sono ancora grandi: il punto di riferimento stage è una pipeline di espansione e aggregazione dello stato di spostamento su centinaia di migliaia di giochi per tipo di gioco.

Game type	Games	Player-game rows	Unique players	Move states	Feature columns
1+0 bullet	483,974	967,948	119,732	22,985,272	277
3+0 blitz	479,974	959,948	129,398	28,138,904	277
10+0 rapid	925,636	1,851,272	229,235	57,554,339	277
>10 rapid	660,540	1,321,080	133,661	38,580,214	277

Table 4: Scala degli artefatti di riferimento stratificati dopo l'allineamento al schema comune delle funzionalità di runtime a 277 colonne. A "Riga giocatore-gioco" indica un gioco dal punto di vista di un giocatore, quindi la maggior parte dei giochi produrre due righe di gioco del giocatore.

Solo per il blitz 3+0, l'attuale benchmark di produzione utilizza 479.974 giochi coinvolgendo 129.398 giocatori unici e 28.138.904 distinte righe di stati di movimento. Il la tabella delle caratteristiche giocatore-gioco del blitz allineato contiene 277 variabili derivate. Il gli artefatti di creazione rapida di funzionalità sono ancora più ampi nella fase dello stato di spostamento: il completo la tabella degli stati di spostamento rapido utilizzata durante la costruzione delle funzionalità contiene 168 colonne dello stato di spostamento prima dell'aggregazione in una funzionalità di gioco del giocatore a 277 colonne tavolo. Gli artefatti del benchmark runtime frontend per tutti e quattro i tipi di gioco sono quindi allineato allo stesso schema di funzionalità giocatore-gioco a 277 colonne, incluso il variabili di breakout espanse della tattica-tassonomia. Da queste tabelle delle caratteristiche, il il report seleziona e visualizza le metriche di coaching più interpretabili.

Questi conteggi sono importanti per interpretare i risultati. Le curve benchmark non sono stime basate su poche centinaia di giochi revisionati manualmente. Invece loro provengono da decine di milioni di stati valutati e più di quattro milioni di righe di giochi per giocatori nelle quattro categorie di giochi di produzione attuali.

5 Panoramica della pipeline

Il percorso di costruzione di riferimento prevede sei fasi principali:

1. Seleziona i giochi Lichess classificati nel tipo di gioco di destinazione e negli strati di valutazione.
2. Analizza ogni mossa del gioco per mossa e memorizza la mossa precedente e successiva stato del consiglio.
3. Calcola le variabili intermedie dello stato di mossa dalla scacchiera, dalla mossa, la prospettiva del giocatore e la prospettiva dell'avversario.
4. Aggrega le variabili dello stato di movimento nelle variabili delle funzionalità del gioco del giocatore.
5. Converti le variabili del gioco del giocatore in metriche di report con un clear direzione dell'interpretazione: più alto è solitamente migliore, più basso è solitamente migliore meglio, o descrittivo.
6. Calcola la media di ogni metrica in base al bucket Elo e controlla manualmente il risultato relazione.

This structure is important because many report metrics are not directly visible in a PGN header. They require reconstructing the position, determining which pieces moved from which starting squares, comparing before-move and after-move material and structure, and then aggregating the result from the move level to the player-game level.

6 Variabili intermedie

Le variabili intermedie sono progettate per descrivere cosa è cambiato sul tabellone e perché è importante dal punto di vista del reporter. Gli esempi includono:

- numero di mossa, numero di mossa del giocatore, lato da muovere, colore del movente e giocatore colore;
- da-casa, a-casa, tipo di pezzo in movimento, tipo di pezzo catturato e se la mossa era una cattura, uno scacco, un arroccamento, una promozione o un matto;
- bilancio materiale prima e dopo il trasloco;
- se un pezzo minore ha lasciato la sua casa iniziale con un dato numero di mosse;
- se lo stesso pezzo è stato spostato su giocatori consecutivi si muove nel apertura;
- se la Regina si è mossa prima della mossa 10;
- lato dell'arrocco, numero della mossa dell'arrocco e se le torri sono diventate collegate dopo l'arrocco;
- pezzi sciolti, pezzi appesi, pedoni raddoppiati, pedoni isolati, passati pedoni e pedoni arretrati;

- stato del file torre, file aperti, file semiaperti e sicurezza del re caratteristiche;
- opportunità tattiche, tattiche eseguite, tattiche di pressione materiale, guadagni successivi garantiti e guadagni realizzati.

The exact intermediate set varies by build stage. Runtime report databases keep only the columns needed to serve reports quickly, while the full feature-build artifacts include wider move-state tables and feature patches used to construct the final player-game feature rows.

7 Esempi di metriche del principio di apertura

I principi di apertura sono esempi utili perché sono facili da comprendere per gli utenti comprendere e facile da collegare ai calcoli dello stato di spostamento.

7.1 Minor Pieces Developed by Move 10

Questa metrica conta quanti dei quattro cavalieri e alfieri di un giocatore sono rimasti le loro caselle di casa entro la decima mossa del giocatore. Il calcolo intermedio utilizza il colore del motore, il tipo di pezzo in movimento, la casella di partenza e la mossa del giocatore numero. Per il Bianco, le caselle di casa sono b1, g1, c1 e f1. Per Black, lo sono b8, g8, c8 e f8. Per ciascun giocatore-gioco, Elo+Chess conta la casa distinta le caselle tra i cavalieri e gli alfieri mossi dal giocatore si muovono 10.

Questa è una metrica più alta è solitamente migliore. Non pretende che ogni minore il pezzo deve sempre muoversi di 10 mosse, ma tra i giocatori in via di sviluppo è empirico La relazione è chiara: i giocatori che raggiungono livelli Elo più alti tendono a svilupparsi altri pezzi minori in precedenza.

7.2 Repeated Piece Moves in the Opening

Questa metrica conta invece le prime mosse extra spese per spostare nuovamente lo stesso pezzo di sviluppare qualcosa di nuovo. Il calcolo intermedio utilizza il precedente casella di destinazione, casella di origine corrente e numero di mossa del giocatore. Se lo stesso pezzo è stato spostato nella mossa precedente del giocatore e viene spostato nuovamente all'interno della mossa prime 10 mosse del giocatore, il conteggio delle mosse ripetute aumenta.

This is a lower-is-usually-better metric. It measures wasted opening tempi in a simple way. Some repeated moves are tactically justified, but at scale the bucket curve captures the general tendency: stronger players at the target levels spend fewer early moves moving the same piece repeatedly.

7.3 Queen Touched Before Move 10

Questa metrica conta le mosse della regina prima della mossa del giocatore 10. L'intermedio il calcolo utilizza il tipo di pezzo in movimento e il numero di mosse del giocatore. Se il pezzo in movimento è la regina e il numero di mosse del giocatore è al massimo 10, il conteggio aumenta.

Questo è anche il valore più basso, solitamente migliore, per l'intervallo target Elo+Chess. Presto Le mosse della regina possono vincere materiale o forzare concessioni, ma molti giocatori in via di sviluppo fai uscire la regina troppo presto, perdi tempo negli attacchi alla regina e ritardi sviluppo dei pezzi minori e sicurezza del re.

7.4 Castling and King Safety

Le metriche di arrocco utilizzano le bandiere di arrocco, il lato dell'arrocco, il numero di mosse dell'arrocco e lo stato del consiglio dopo l'arrocco. Gli esempi includono se il giocatore ha arroccato, se il giocatore ha arroccato con la mossa 10, se il giocatore ha arroccato dal lato del re oppure lato di Donna, se le torri erano collegate dopo l'arrocco e se le file delle torri erano aperti o semiaperti dopo l'arrocco.

Questi parametri non sono tutti interpretati allo stesso modo. Completamento dell'arrocco e l'arrocco con la mossa 10 è generalmente più alto, di solito, migliore nell'intervallo target. L'arrocco a lato della regina o l'arrocco in file aperti può dipendere maggiormente dal contesto. Questo è il motivo per cui Elo+Chess non codifica semplicemente una regola morale; costruisce il secchio curve e controlla quali relazioni sono sufficientemente stabili per il coaching.

8 Curve della benna e ispezione manuale

After player-game metrics are calculated, Elo+Chess groups them by Lichess Elo bucket and computes bucket averages. The bucket average is the raw empirical benchmark point for that metric. A trend line is then used in the report to assign Elo-style grades to user values.

Le curve vengono ispezionate manualmente prima di essere utilizzate. Una metrica è più utile quando il bucket indica una relazione coerente in tutto l'intervallo di rating servito dal prodotto. Una metrica è meno utile quando è rumorosa, piatta, altamente non monotono o guidato da casi speciali difficili da spiegare a un utente.

Questa fase di ispezione manuale è particolarmente importante sopra la tomaia Elo+Chess taglio. I giocatori più forti sanno quando infrangere i principi di apertura e lo fanno in modo efficace. Al di sopra della gamma da principiante a presto avanzato, la relazione tra abitudini semplici e valutazione spesso si appiattisce o cambia forma. Il prodotto sottolinea quindi l'intervallo in cui le relazioni empiriche sono forti e pedagogicamente utile.

9 Esempio di curve di apertura a intervallo completo

La figura 5 mostra quattro curve del principio di apertura blitz 3+0. I punti grigi mostrano la gamma di valori Lichess più ampia. Punti blu e tendenza del blu la linea mostra la gamma 600–1600 Lichess utilizzata dal principale coaching Elo+Chess parametri di riferimento. La linea rossa tratteggiata segna il limite superiore.

Lo stesso schema è visibile nell'intera gamma 10+0 rapid e long-rapid artefatti. La decisione centrale sulla modellazione rimane invariata: utilizzare empiricamente intervallo stabile per le affermazioni del coaching ed evitare di fingere che si tratti di una semplice curva di abitudine rimane ugualmente informativo per sempre.

Lo stesso modello appare numericamente nella Tabella 5. Le piste sono riportato come variazione del valore metrico per 100 punti Elo. Arrocco, pezzo minore lo sviluppo, il movimento ripetuto del pezzo e il movimento iniziale della regina sono tutti chiari movimento tra il 600 e il 1600. Sopra i 1600 in questi le pendenze sono molto minori esempi, il che è coerente con l'idea che i giocatori con un punteggio più alto falliscono principi fondamentali in modo più selettivo e con maggiore successo. Un secondo meccanismo è saturazione. Per alcune abitudini, la metrica raggiunge un tetto pratico o dolce spot: i giocatori forti hanno già imparato l'abitudine, quindi non c'è motivo di farlo la curva continua a muoversi bruscamente verso l'alto. Inoltre, non è sempre

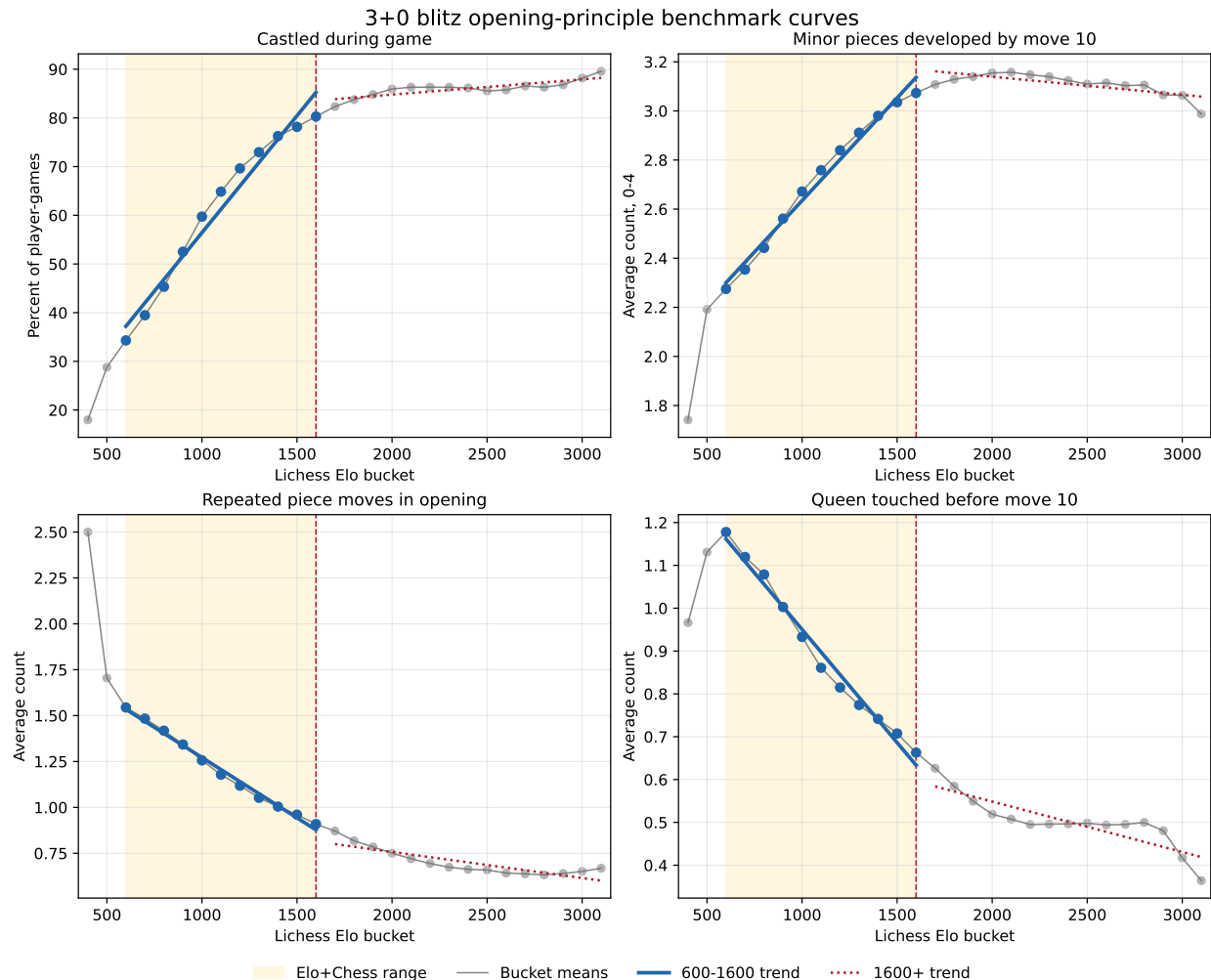


Figure 5: Curve benchmark selezionate del principio di apertura per il blitz 3+0. Il le relazioni sono più forti e più utili nella gamma di coaching Elo+Chess. Al di sopra del limite superiore, diverse relazioni si appiattiscono sostanzialmente.

meglio dopo quel punto. I dati quindi si appiattiscono entrambi perché i giocatori più forti possono rompersi regole in posizioni concrete e perché molte abitudini di base sono già state raggiunte il loro raggio d'azione utile.

10 Perché il limite superiore è importante

L'utente target di Elo+Chess è un giocatore da principiante ad avanzato. Per questi giocatori, le abitudini di base sono fortemente associate al punteggio: sviluppare pezzi, evitare ripetute mosse iniziali dei pezzi, evitare mosse iniziali non necessarie della regina, arroccare nel tempo, collega le torri, gestisci il materiale ed evita di lasciare pezzi sciolti o appeso.

A punteggi più alti, quelle stesse abitudini superficiali diventano meno diagnostiche. Un forte il giocatore può ritardare l'arrocco perché il centro è chiuso, spostare la regina in anticipo perché la posizione contiene una tattica concreta, oppure muovere lo stesso pezzo due volte perché con-

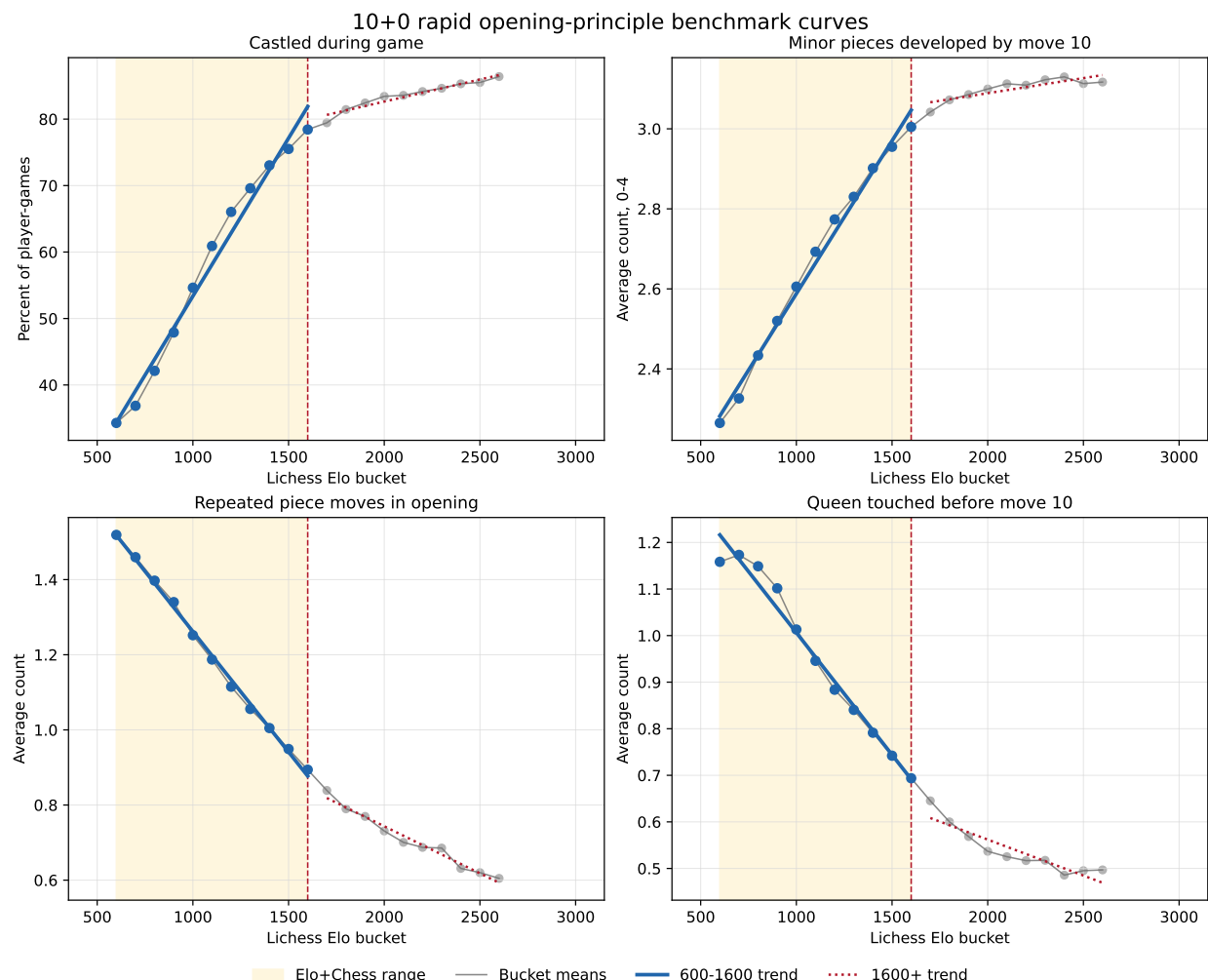


Figure 6: Curve benchmark selezionate del principio di apertura per giochi rapidi 10+0. Il le relazioni sono più forti e più utili nella gamma di coaching Elo+Chess. Al di sopra del limite superiore, diverse relazioni si appiattiscono sostanzialmente.

quista un ritmo o impone una concessione strutturale. L'abitudine ancora cruda ha significato, ma il rapporto tra abitudine e valutazione diventa maggiore condizionale. Questo è il motivo per cui Elo+Chess utilizza un taglio superiore per l'allenamento principale affermazioni e perché gli utenti al di sopra di tale intervallo vengono avvisati della stretta relazione tra abitudini di principio e rating si rompe.

11 Metriche tattiche bilaterali e obiettivi percentili

Alcune metriche hanno la forma opposta rispetto agli esempi del principio di apertura. Forchette sono un esempio utile perché l'evento è intrinsecamente bilaterale. Un giocatore che trova e converte fork più utili dei peer sta facendo qualcosa di buono, ma il tasso di popolazione dei fork convertiti diminuisce con il rating perché più forte gli avversari concedono meno opportunità di fork e si difendono dalle minacce tattiche più attivamente.

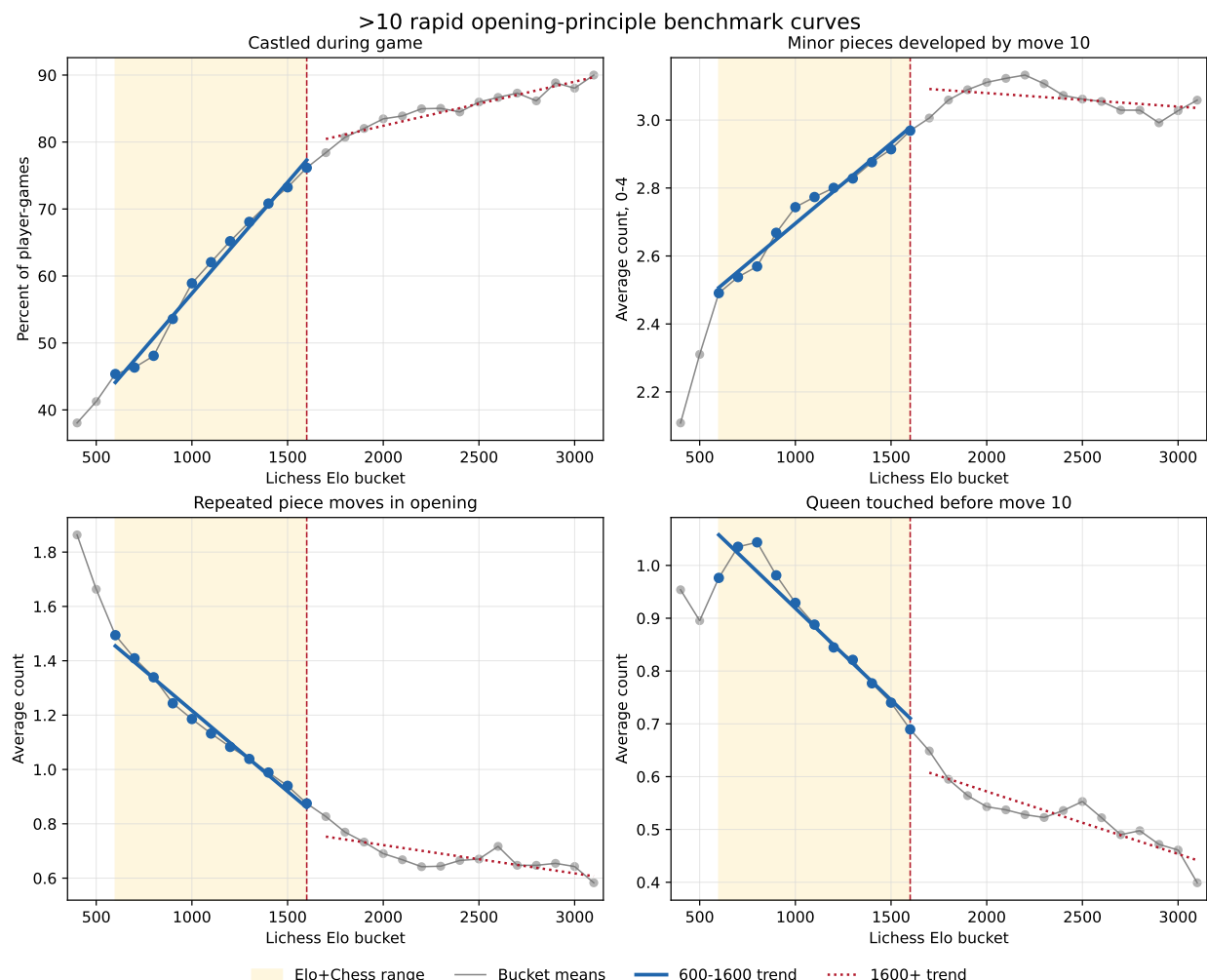


Figure 7: Curve benchmark selezionate del principio di apertura per partite rapide più lunghe.

La figura 8 mostra questo effetto. Entrambe le forcelle convertite da il giocatore di segnalazione e le fork convertite dall'avversario diventano meno frequenti gli ascolti salgono. Ciò non significa che eseguire fork sia sbagliato. Significa l'evento grezzo il tasso è in parte controllato dal tasso di errore dell'avversario e dalla trama del gioco. Per le metriche in questo modo, Elo+Chess utilizza quindi i percentili della peer-history all'interno di l'intervallo Elo del giocatore, anziché fare affidamento solo su una tendenza globale della media dell'intervallo.

Per costruire questi obiettivi percentili, Elo+Chess campiona i giocatori per tipo di gioco e Il bucket Elo utilizza quindi le partite di qualificazione recenti per ciascun giocatore campionato. Il l'attuale set di obiettivi utilizza 11 bucket da 600 a 1600, 200 giocatori ciascuno vecchio e fino a 50 partite di qualificazione recenti per giocatore per ciascuna produzione tipo di gioco. Ciò fornisce circa 2.200 campioni di giocatori per tipo di gioco 110.000 righe di gioco per giocatore per tipo di gioco per percentile all'interno del bucket distribuzioni.

La figura 9 illustra il target risultante distribuzione per il tasso di fork convertito dai giocatori campionati nel blitz 3+0. Ogni scatola è una distribuzione all'interno del bucket tra i giocatori campionati, non una media nel bucket partite individuali. Nell'interfaccia utente del report, il valore di un utente viene confrontato con quello pertinente casella per il bucket Elo mappato dell'utente.

Game type	Metric	600–1600 slope per 100 Elo	1600+ slope per 100 Elo
1+0 bullet	Castled during game	4.137	0.550
	Minor pieces developed by move 10	0.063	0.009
	Repeated piece moves in opening	-0.051	-0.026
	Queen touched before move 10	-0.040	-0.011
3+0 blitz	Castled during game	4.804	0.313
	Minor pieces developed by move 10	0.084	-0.007
	Repeated piece moves in opening	-0.066	-0.014
	Queen touched before move 10	-0.053	-0.012
10+0 rapid	Castled during game	4.753	0.665
	Minor pieces developed by move 10	0.076	0.008
	Repeated piece moves in opening	-0.064	-0.025
	Queen touched before move 10	-0.052	-0.015
>10 minute rapid	Castled during game	3.320	0.659
	Minor pieces developed by move 10	0.047	-0.004
	Repeated piece moves in opening	-0.059	-0.010
	Queen touched before move 10	-0.035	-0.012

Table 5: Pendenze approssimative della curva a secchiello per il principio di apertura dei quattro esempi parametri tra i tipi di giochi di produzione.

Game type	Elo buckets	Players per bucket	Games per player	Player samples
1+0 bullet	11	200	49–50	2200
3+0 blitz	11	200	27–50	2200
10+0 rapid	11	200	37–50	2200
>10 minute rapid	11	200	47–50	2200

Table 6: Set di obiettivi percentili della cronologia peer attuali. Questi sono campioni della cronologia dei giocatori utilizzati per stimare le distribuzioni all'interno del bucket metriche in stile percentile.

Questa è la base per affermazioni come se il tasso di fork convertito di un giocatore è alto o basso rispetto ai suoi pari stesso livello, anche quando il tasso di eventi nella popolazione complessiva diminuisce con la valutazione.

12 Costo di replica

Il metodo è riproducibile in linea di principio, ma non è leggero. Ricostruire le curve benchmark richiedono:

- accesso a file di cronologia di gioco mensili Lichess di grandi dimensioni;
- filtraggio stratificato per tipo di gioco, fascia di valutazione e idoneità regole;
- un analizzatore di mosse legali e un valutatore del consiglio di amministrazione in grado di espandersi decine di milioni di stati spostati;
- logica delle caratteristiche per apertura, materiale, struttura del pedone, sicurezza del re, torre attività, tattica, gestione del tempo e risultati del gioco;
- aggregazione dallo stato di spostamento alle funzionalità del gioco del giocatore;

Two-sided converted-fork rates by Elo bucket

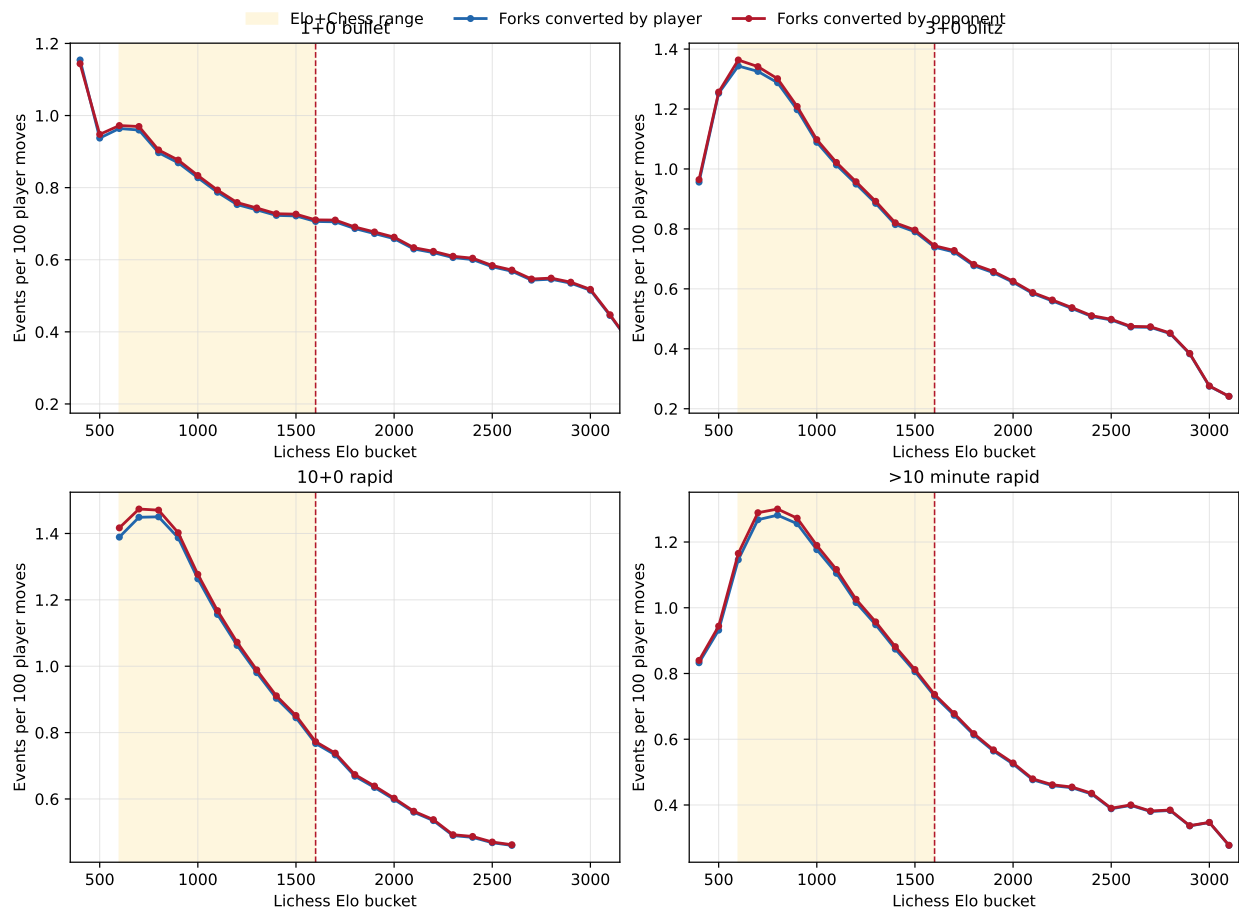


Figure 8: Tassi con forche convertite su due lati mediante benna Elo. Eventi fork convertiti Cadono a valutazioni più alte perché entrambe le parti difendono meglio e con meno forchette pulite sono disponibili conversioni

- riepiloghi a livello di bucket per ogni metrica del report;
- revisione manuale di ogni curva benchmark candidata prima che venga utilizzata un rapporto di coaching.

Per il blitz 3+0 ciò significa quasi mezzo milione di partite, quasi un milione righe del gioco del giocatore e oltre 28 milioni di righe dello stato di movimento. Attraverso i quattro tipi di giochi benchmark, gli attuali artefatti benchmark contengono più di 2,55 milioni di giochi, più di 5,10 milioni di righe di giochi-giocatore e più di 147 milioni di righe di stato di spostamento. La scala è centrale per il valore del benchmark: consente al report di mostrare agli utenti cosa accade effettivamente nei giochi giocati da giocatori vicini al loro livello invece di affidarsi solo a slogan scacchistici generali.

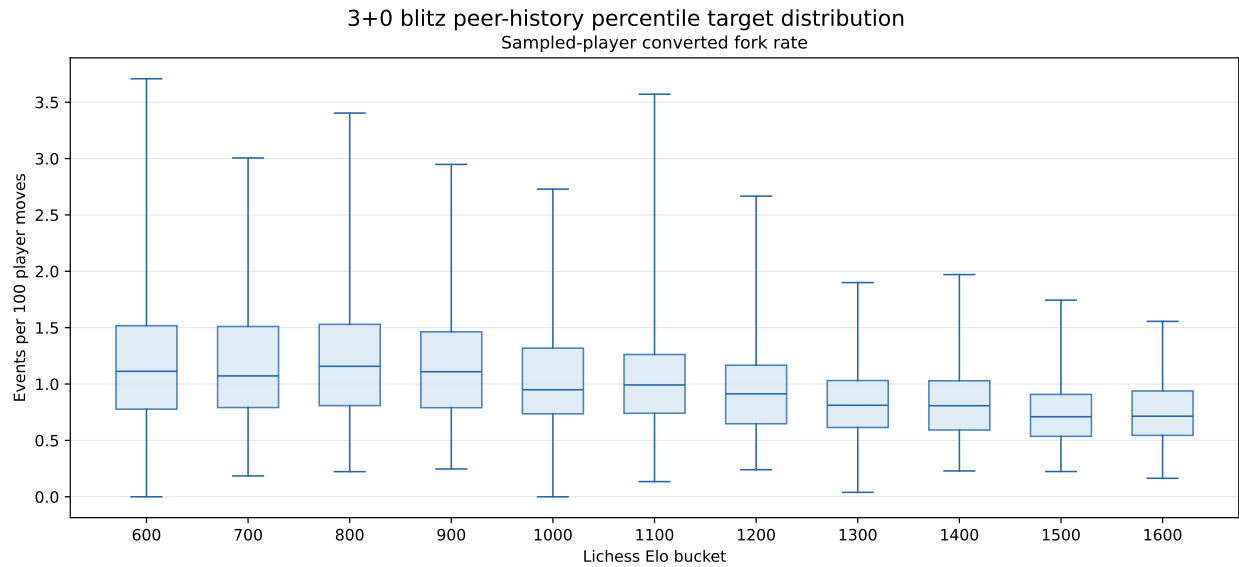


Figure 9: Esempio di distribuzione target della peer-history blitz 3+0 per il giocatore campionato tasso di forcella convertito. I percentili vengono calcolati all'interno del bucket Elo, quindi il rapporto confronta un giocatore con compagni che si trovano ad affrontare opportunità tattiche sostanzialmente simili tariffe.

13 Limitazioni

Le curve di riferimento sono descrizioni empiriche, non prove causali. Un giocatore non ottiene rating semplicemente abbinando una media stagionale. Anche le metriche lo sono riassunti imperfetti di posizioni complesse: una mossa anticipata della regina può essere eccellente in una posizione e spericolato in un'altra. Lo scopo delle curve è quello di identificare le abitudini ricorrenti e confrontarle con il comportamento dei pari, non farlo sostituire il calcolo o il giudizio del coaching.

Le curve sono anche curve in scala Lichess. Quando gli utenti portano il gioco Chess.com storie, Elo+Chess mappa la fascia di rating pertinente sul benchmark Lichess scalare utilizzando il modello di mappatura multiplatforma separato. Quel livello di mappatura è descritto nella nota di ricerca Elo+Chess complementare su Lichess-to-Chess.com conversione del voto: [Estimating Cross-Platform Chess Rating Mappings with Regression modale](#). Questa mappatura non deve essere letta come preservazione del percentile rango. Conserva un'equivalenza stimata tra i due rating sistemi; il rango percentile rimane specifico della piattaforma perché il giocatore attivo le distribuzioni sono diverse.

14 Grafica del report di produzione

Il report di produzione trasforma gli artefatti del benchmark in pannelli a livello di metrica che può essere letto senza conoscere l'intera pipeline di compilazione. Figure 10 and 11 mostra tre esempi dal progetto del report Elo+Chess in tempo reale.

I pannelli del principio di apertura nella Figura 10 utilizzare la presentazione ordinaria della curva benchmark. Ogni carta inizia con il famiglia metrica e nome della metrica, quindi fornisce una breve regola di interpretazione. Il voto grande nell'angolo in alto a destra è intenzionalmente

separato dal grezzo valore metrico: risponde alla domanda posta dall'utente "come valuta questa abitudine". rispetto al benchmark?" mentre le righe sottostanti preservano i valori misurati. Le quattro righe di riepilogo mostrano Durata, Ultimi 10, Vittorie e Perdite. Questo rende il pannello diagnostico piuttosto che meramente decorativo. Se Lifetime e Last 10 sono d'accordo, l'abitudine è probabilmente stabile. Se le vittorie e le perdite divergono, la metrica potrebbe esserlo dipendente dal risultato o legato allo stato del gioco piuttosto che una semplice abitudine.

L'area del grafico utilizza uno sfondo scuro sobrio, una linea di tendenza dorata e colorata segnapunti per le quattro fette. L'annotazione della tendenza indica esplicitamente se il rapporto di riferimento aumenta o diminuisce con Elo. Questo è importante perché alcuni parametri sono "più alto è meglio", come lo sviluppo di pezzi minori, mentre altri sono più bassi è meglio, come il movimento iniziale della regina. La carta quindi non chiede all'utente di dedurre la direzione solo dalla pendenza. La finale Le righe "Che cosa significa" e "Suggerimenti per il coaching" traducono il risultato del benchmark in un'istruzione scacchistica concreta preservando il contesto empirico.

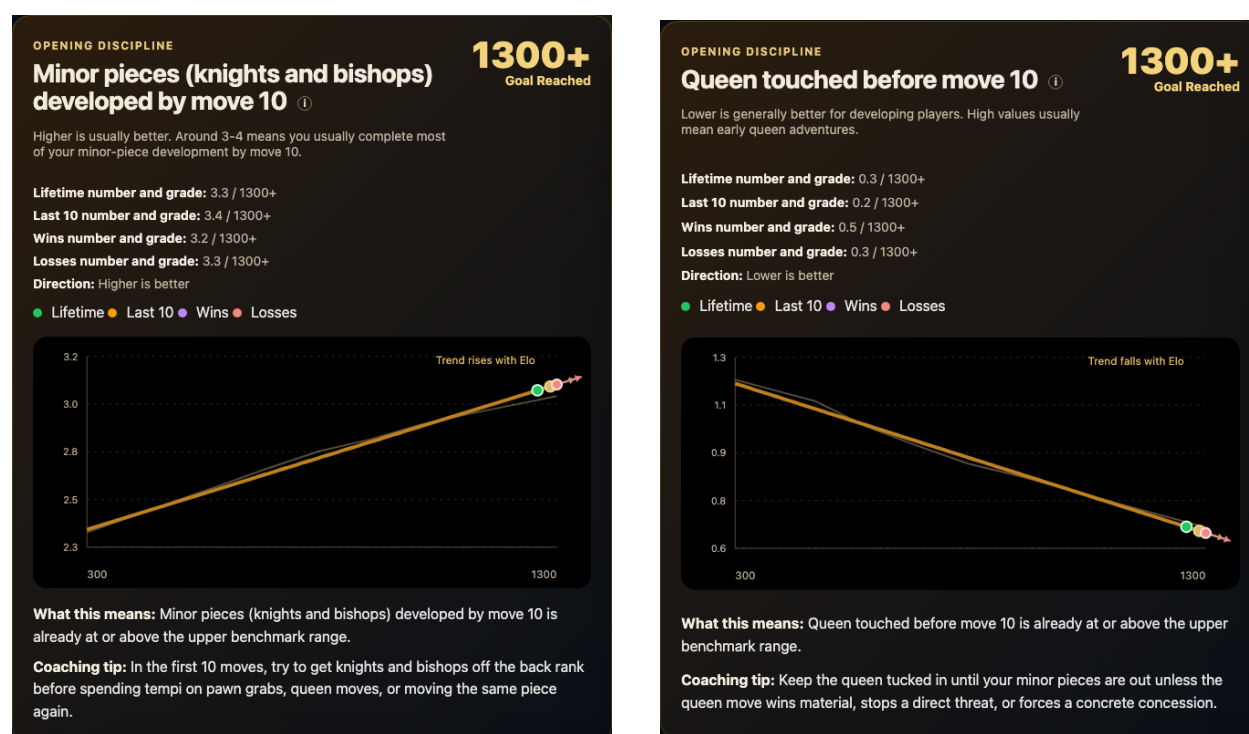


Figure 10: Pannelli con principio di apertura della produzione. La stessa grammatica visiva gestisce a metrica "più alto è meglio", sviluppo dei pezzi minori e metrica "più basso è meglio", primo movimento della regina. Entrambi i pannelli espongono Lifetime, Last 10, Wins e Losses in questo modo il rapporto può separare le abitudini stabili da quelle recenti o dipendenti dai risultati modelli.

La figura 11 mostra lo stile percentile presentazione utilizzata per parametri il cui tasso grezzo non è ben rappresentato da a tendenza globale unica. Le forchette vincenti sono un esempio tattico: il tasso di popolazione dei fork convertiti cambia sia con l'abilità del giocatore che con il tasso di errore dell'avversario. Il card quindi confronta il valore dell'utente con le storie dei giocatori peer nel secchio Elo mappato. Il boxplot mostra la distribuzione dei peer, incluso il intervallo interquartile, mediana e baffi. Il marcatore dell'utente è tracciato sul file stessa scala ed etichettati sia con il valore grezzo che con il rango percentile. Adiacente i contenitori di confronto rimangono

visibili in forma disattivata in modo che l'utente possa vedere che il file il peer bucket è locale, non universale.

Questa progettazione rende verificabili le metriche percentili. L'utente vede il valore grezzo, il percentile, il gruppo peer utilizzato per il confronto e il valore tradotto Mappatura dei bucket da Chess.com a Lichess. La stessa vita, ultime 10, vittorie e Le righe delle perdite vengono conservate in modo che i punteggi percentili tattici possano ancora essere controllati per gli effetti di recency e la dipendenza dai risultati.

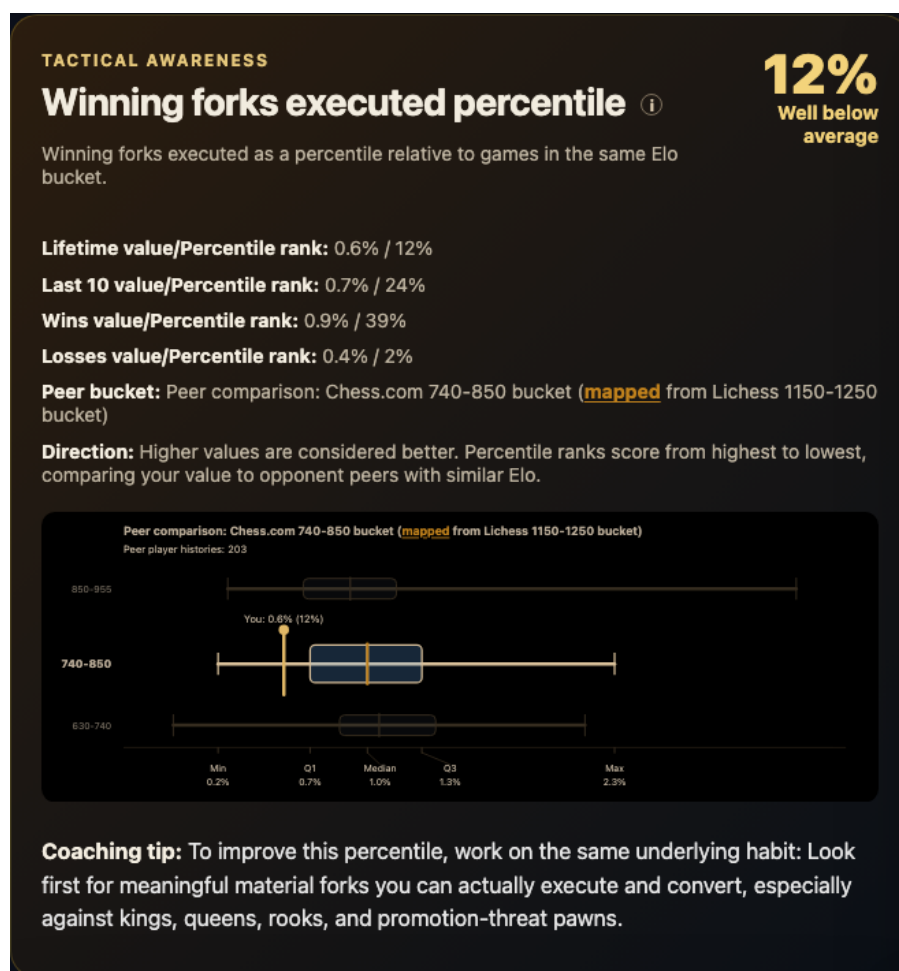


Figure 11: Pannello dei percentili di produzione per le fork vincenti eseguite. La trama a scatola mostra la distribuzione dei peer nello stesso segmento utilizzata dal rapporto, mentre l'oro l'indicatore mostra il valore grezzo e il rango percentile dell'utente.

15 Conclusione

I report sui benchmark Elo+Chess sono costruiti partendo da un ampio calcolo dello stato di spostamento campioni Lichess stratificati. Ogni metrica inizia come un insieme di stati concreti della scheda e le variabili dello stato di movimento, diventa una caratteristica del gioco del giocatore e viene quindi calcolata la media dal bucket Elo per formare una curva benchmark empirica. Il principio di apertura esempi mostrano perché questo approccio è utile: le abitudini scacchistiche

semplici hanno forti, relazioni visibili con la valutazione nell'intervallo target, mentre quelle le relazioni spesso si appiattiscono o diventano più condizionate al di sopra del limite superiore. Questo è il motivo principale per cui Elo+Chess concentra le sue affermazioni di coaching sul rating intervallo in cui i dati li supportano più chiaramente.