

Construindo curvas de referência estratificadas de Lichess para Elo+Chess

Notas de pesquisa Elo+Chess

31 de maio de 2026

Abstract

Os relatórios de benchmark Elo+Chess são criados a partir de grandes amostras estratificadas de Jogos Lichess.org, com o exemplo de benchmark usado nesta nota contendo 483.974 jogos bullet 1+0, 479.974 jogos blitz 3+0, 925.636 jogos rápidos 10+0, e 660.540 jogos rápidos mais longos. Para cada tipo de jogo principal, o pipeline se expande cada jogo nos estados do tabuleiro antes e depois do movimento, calcula centenas de variáveis intermediárias e do jogo do jogador, agrega essas variáveis em métricas de coaching e, em seguida, calcula a média dessas métricas por intervalo Lichess Elo. O a amostra é deliberadamente estratificada em vez de proporcional ao jogador bruto pool: o objetivo é estimar curvas de comportamento estáveis em cada nível de rating, incluindo baldes que de outra forma seriam esparsos, para não reproduzir o natural distribuição de classificação. O resultado é uma família de curvas de referência empíricas: para cada métrica, podemos perguntar com que frequência os jogadores em diferentes níveis Elo mostram um determinado hábito ou princípio em jogos reais. Esta nota descreve o método, a escala de o cálculo e vários exemplos de princípios de abertura envolvendo roque, desenvolvimento de peças menores, movimentos repetidos da peça de abertura e movimento inicial da rainha. A ênfase é metodológico: Elo+Chess não assume que os princípios do xadrez sejam universalmente linear em todos os níveis de habilidade. Em vez disso, as curvas de referência são inspecionadas por faixa de classificação, e o produto se concentra na região do iniciante ao avançado onde as relações são mais fortes e mais interpretáveis.

1 Propósito

O objetivo do sistema de referência é converter conselhos gerais sobre xadrez em quantidades mensuráveis. Uma afirmação como “desenvolva suas peças menores” é útil, mas torna-se mais prático quando podemos dizer quantas peças menores um jogador geralmente se desenvolve no lance 10, como isso se compara com seus pares no mesmo intervalo de classificação e como esse hábito muda em todo o espectro de classificação.

Elo+Chess, portanto, constrói uma biblioteca de métricas a partir de jogos reais. Cada métrica é calculado para cada linha de jogo do jogador, resumido por tipo de jogo e balde Elo, e usado como curva de referência no relatório. O relatório pode então colocar um usuário em a mesma curva e mostrar se o usuário está acima, abaixo ou próximo do comportamento observado entre jogadores comparáveis.

2 Dados de origem e estratificação

A fonte de benchmark é o banco de dados público de jogos Lichess.org. Elo+Chess usa grandes amostras estratificadas por tipo de jogo e classificação, em vez de uma pequena amostra de

conveniência. Os atuais artefatos de tempo de execução de produção cobrem quatro principais categorias de jogos:

- 1+0 bala,
- 3+0 blitz,
- 10+0 rápido,
- jogos rápidos com mais de 10 minutos.

O objetivo da estratificação é a amplitude entre os níveis de classificação. Um aleatório proporcional amostra seria eficiente para descrever o conjunto geral de jogadores, mas seria ser ineficiente para construir curvas de referência. Os grupos de classificação mais comuns dominariam a amostra, enquanto os baldes traseiros de baixo suporte teriam muito poucos jogos para médias métricas estáveis. Elo+Chess, portanto, amostras por jogo médio Balde Elo. Na atual construção de amostra limpa, os jogos brutos são os primeiros filtrado para jogos com classificação válida com jogadores conhecidos, classificações conhecidas, resultados válidos, e controles de tempo conhecidos. Os jogos são então filtrados para jogadores com jogos suficientes nessa velocidade e classificações suficientemente estáveis para representar um nível de habilidade coerente: para cada jogador, Elo+Chess calcula a faixa de classificação dentro da velocidade do jogador ($\max(\text{Elo}) - \min(\text{Elo})$) no pool de origem e mantém jogos somente quando ambos os jogadores tiverem pelo menos a contagem mínima de jogos configurada e não mais que 300 pontos Elo de desvio dentro da velocidade. O filtro de deriva não se destina para remover melhoria ou variação normal; reduz os casos em que o jogador os jogos misturariam níveis de habilidade materialmente diferentes, como contas provisórias, jogadores que retornaram ou contas que passaram rapidamente por vários grupos durante o período amostrado. Por fim, os jogos elegíveis são classificados dentro de cada Elo médio. balde por um hash determinístico do ID do jogo e da semente, e um número limitado de jogos são retidos de cada intervalo.

Este design torna as curvas de benchmark muito mais úteis. Cada balde Elo contribui com observações suficientes para estimar o valor médio de cada métrica, e raros baldes de alta ou baixa classificação não são abafados pela forma natural da distribuição de jogadores ativos. As curvas de benchmark usadas por Elo+Chess são então restrito à faixa mais relevante para o produto, aproximadamente iniciante através do jogo avançado inicial. Nos relatórios da escala Lichess, esse intervalo primário é 600–1600. Quando um usuário visualiza o relatório na escala Chess.com, o os intervalos de classificação correspondentes são mapeados no mesmo Lichess subjacente intervalos de benchmark usando o método de mapeamento de classificação de plataforma cruzada separado descrito no artigo complementar: [Estimating Cross-Platform Chess Rating Mappings with Regressão Modal](#).

3 Distribuições de classificação de jogadores ativos

Se olharmos para a distribuição das classificações Elo do jogador Lichess.org no final de março de 2026, por tipo de jogo principal, vemos as seguintes contagens e percentis. Para esta visualização de distribuição, cada jogador é contado uma vez por tipo de jogo exato, usando a última classificação observada desse jogador na matéria bruta correspondente linhas de controle de tempo, depois de exigir pelo menos cinco jogos naquele jogo exato tipo.¹

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 mostre as distribuições

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: Distribuição de classificação Lichess de jogador ativo mais recente observada por exato tipo de jogo depois de exigir pelo menos cinco jogos nesse tipo de jogo exato.

Esta distribuição não deve ser confundida com a amostra estratificada de referência. A distribuição descreve o pool de jogadores Lichess ativo observado no bruto digitalizar. A amostra de benchmark é então deliberadamente estratificada para que Elo seque têm suporte suficiente para curvas métricas estáveis.

Para mapear de uma classificação Elo Lichess.org para um tipo de jogo específico para um classificação Chess.com equivalente, Elo+Chess usa o modal específico do tipo de jogo equações de regressão estimadas no companheiro [Elo+Chess rating-mapping paper](#):

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

Essas equações mapeiam valores de classificação, não classificações percentuais. As classificações percentuais são definido dentro de um pool de jogadores específico, e os jogadores Chess.com e Lichess piscinas têm formatos diferentes. Na Tabela 2, para o Tipo de jogo rápido de 10 minutos, a primeira coluna fornece uma pequena amostra de possíveis Valores de classificação Chess.com Elo. A segunda coluna fornece o correspondente Classificação percentil Chess.com exibida em Chess.com em 31 de maio de 2026. Um 71,7 classificação em Lichess.org para rápido de 10 minutos corresponde a uma classificação Lichess Elo de 1.644² e esse valor é mostrado no terceira coluna. Se aplicarmos a equação de mapeamento Chess.com–Lichess, 1.644 em Lichess é equivalente a cerca de 1.271 em Chess.com. Assumindo a estimativa equação de mapeamento for razoável, a lacuna +522 entre a classificação Chess.com em coluna 1 e a classificação mapeada na coluna 4 podem ser explicadas por um grande número de jogadores casuais no pool de jogadores Chess.com inflando o Chess.com classificação percentil. Em outras palavras, estar no percentil 72 em Lichess.org é uma conquista maior do que estar no percentil 72 em Chess.com, porque as classificações percentuais não mapeiam claramente de um sistema para outro. Elo real os valores podem ser mapeados usando equações estimadas a partir das pontuações do mesmo jogador em os dois sistemas.

A razão para as lacunas na Tabela 2 é que o mapeamento estima classificações equivalentes de força de jogo, enquanto um percentil descreve posição dentro da própria população de uma plataforma. Um jogador do percentil 71,7 em Chess.com, portanto, não é automaticamente um jogador do percentil 71,7 em Lichess, e vice-versa. Os números observados são consistentes com o público de Chess.com pool rápido tendo uma massa de jogadores casuais ou de classificação

com e sem este filtro.

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: Comparação ilustrativa entre o percentil rápido de 10 minutos Chess.com pontos e os mesmos pontos percentuais no ativo Lichess 10+0 rápido distribuição depois de exigir pelo menos cinco jogos 10+0. “Chess.com mapeado equiv.” aplica o Elo+Chess atual Conversão rápida da escala de classificação 10+0 para a classificação Lichess do mesmo percentil. O as linhas da cauda mais alta são incluídas para mostrar a distinção da escala, mas as linhas ajustadas A linha de classificação é mais confiável na faixa de treinamento principal.

inferior muito maior, enquanto o O pool Lichess de controle exato ativo é mais concentrado entre repetições experientes jogadores. O ponto chave não é que uma escala percentil esteja “certa” e a outro está “errado”; é que a classificação percentual não é portátil entre plataformas a menos que as distribuições de jogadores subjacentes tenham a mesma forma.

Note on the 1500 spike. The visible jump near 1500 está presente nos dados, em vez de ser criado por suavização.

No último distribuição de classificação observada, 1.500 classificações exatas representam 13,1

jogadores bullet, 5,6

17,6

âncora inicial ou provisória no sistema de classificação.

O pico exato de 1.500 é em grande parte um artefato de contas levemente ativas. Para verificar isso, recalculamos as mesmas distribuições de classificação mais recente após exigir cada jogador tenha pelo menos cinco jogos no tipo exato de jogo. O pico quase desaparece sob este filtro: 1.500 avaliações exatas caem de 5,6–17,6

jogadores na distribuição de todos os jogadores para 0,2–1,2

distribuição relatada na Tabela 1.

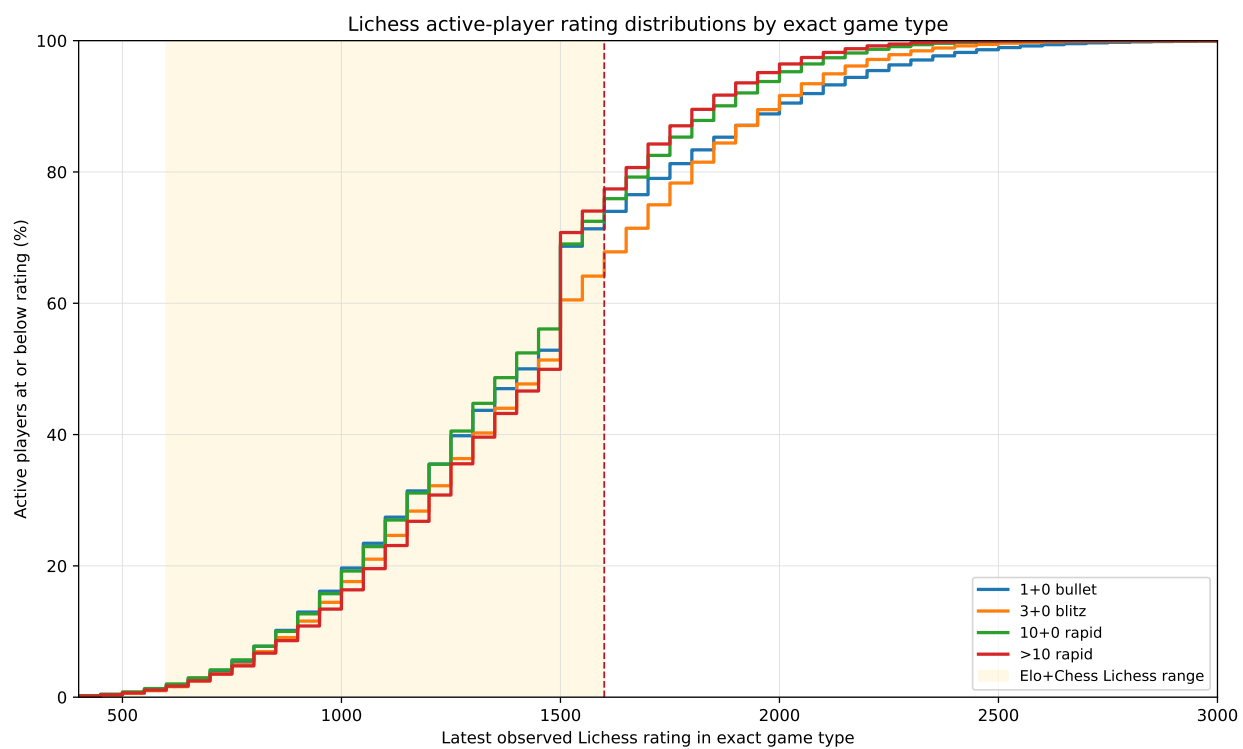


Figure 1: Distribuições cumulativas de classificação de jogadores ativos por tipo de jogo exato. O a faixa amarela marca a faixa 600–1600 Lichess usada pelo Elo+Chess principal referenciais de treinamento.

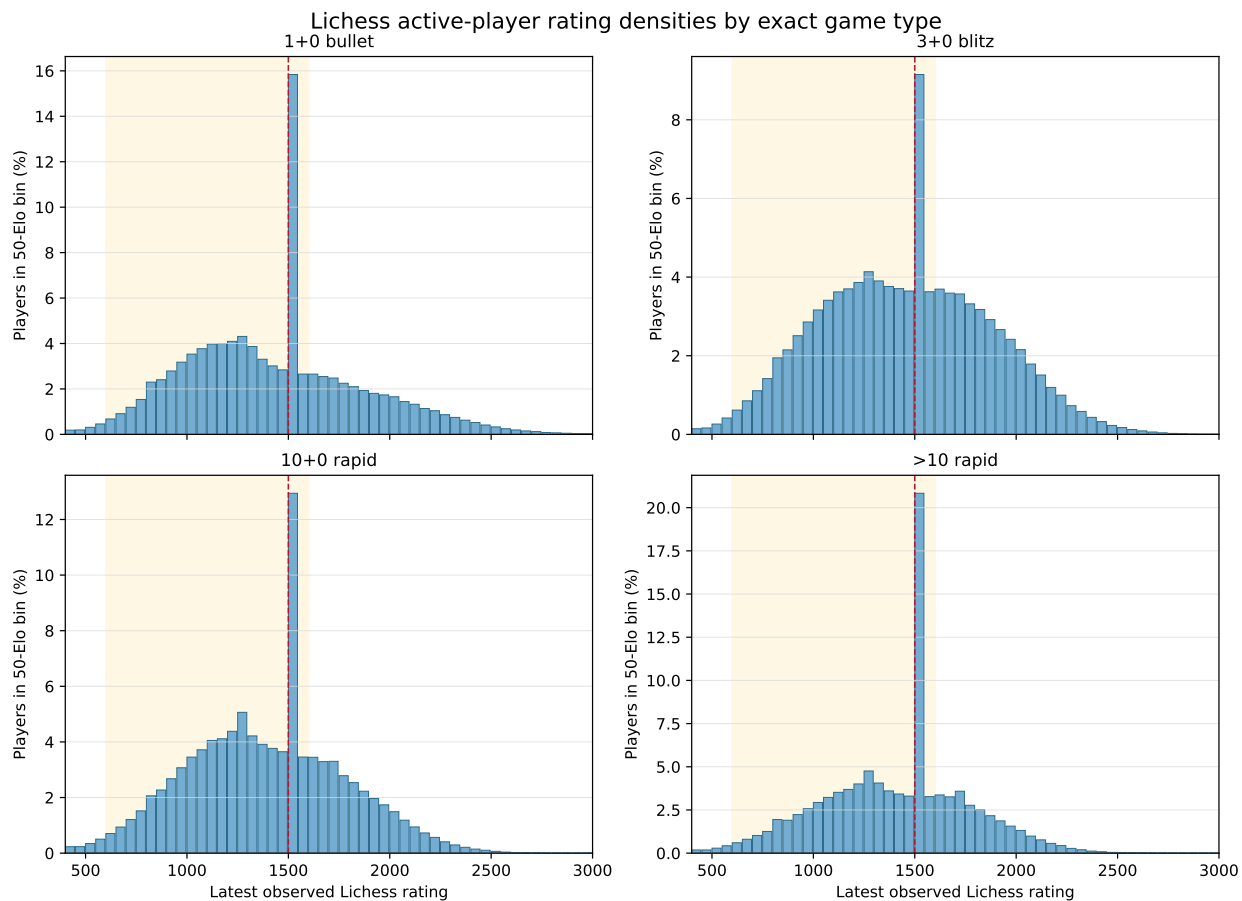


Figure 2: Visualização de densidade das mesmas distribuições de classificação de jogadores ativos. O a linha tracejada marca 1500 Lichess, onde as distribuições brutas de classificação mais recente têm um grande ponto de massa de classificação exata. A faixa amarela marca 600–1600 Lichess intervalo usado pelos principais relatórios de métricas Elo+Chess. O pico é mais visível em a densidade de todos os jogadores porque muitas contas levemente ativas permanecem exatamente em 1500; Figura 4 repete o gráfico de densidade depois de exigir pelo menos cinco jogos no tipo exato de jogo, o que elimina em grande parte este artefato.

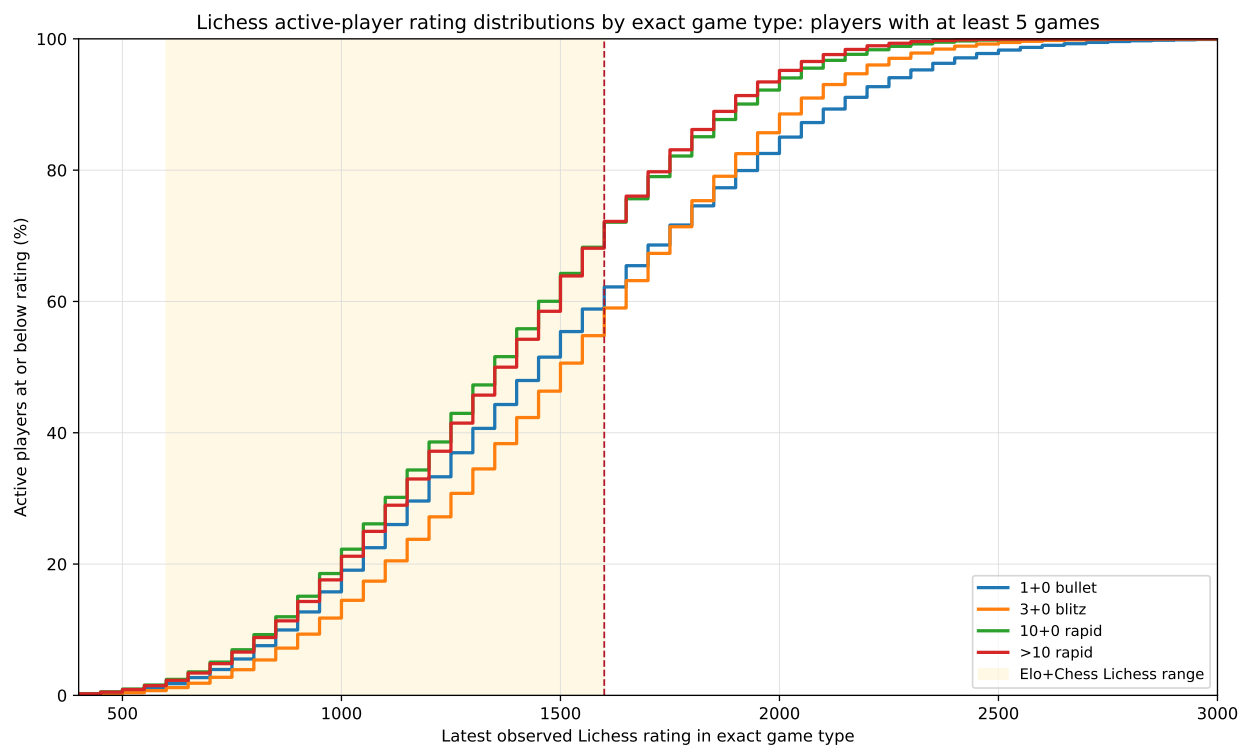


Figure 3: Distribuições cumulativas de classificação de jogadores ativos após exigir pelo menos pelo menos cinco jogos no tipo exato de jogo. O salto de 1500 é bastante reduzido.

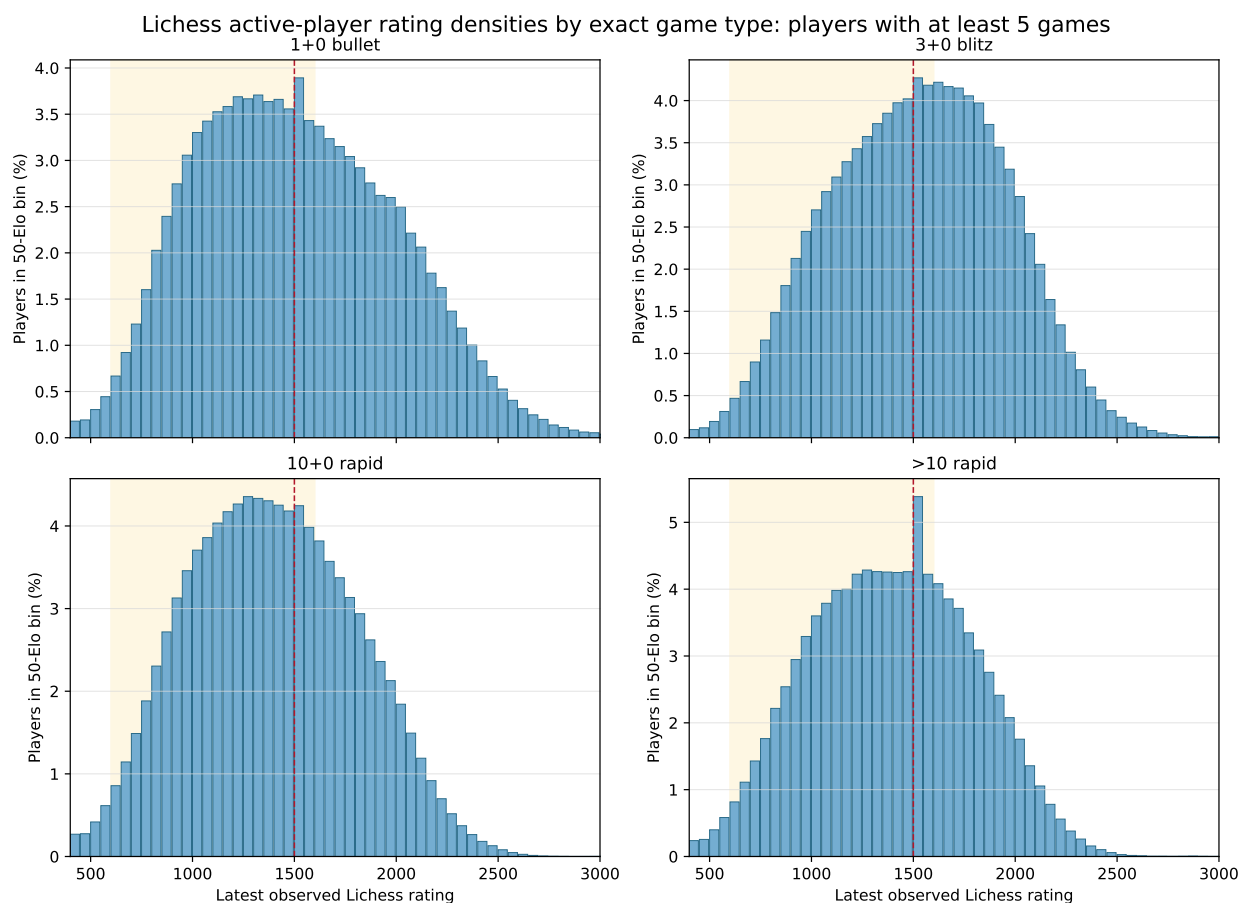


Figure 4: Visualização de densidade após exigir pelo menos cinco jogos no jogo exato tipo. Esta visão confirma que o grande ponto de massa de 1.500 no all-player a distribuição é impulsionada principalmente por contas levemente ativas. A faixa amarela marca a faixa 600–1600 Lichess usada pelos principais relatórios de métricas Elo+Chess.

4 Escala dos artefatos de referência atuais

O cálculo é intencionalmente grande. Não é um arranhão no perfil, é um pequeno conjunto de PGNs baixados ou uma coleção selecionada manualmente de exemplos instrutivos. A passagem Lichess inicial cobriu arquivos mensais completos de janeiro de 2025 até março de 2026. A tabela de varredura não estratificada contém 1.382.178.087 jogos e 2.764.356.174 linhas de classificação de jogos de jogadores. Nessas linhas, 6.623.438 únicos os nomes de usuário dos jogadores aparecem pelo menos uma vez.

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: Escala de varredura Lichess não estratificada antes da amostragem de benchmark. A varredura cobre arquivos mensais completos do histórico do jogo de janeiro de 2025 a março 2026.

Os artefatos de benchmark de produção são muito menores que a varredura bruta porque eles são selecionados para construção métrica. Eles ainda são grandes: a referência estágio é um pipeline de expansão e agregação de estado de movimento em centenas de milhares de jogos por tipo de jogo.

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: Escala dos artefatos de benchmark estratificados após alinhamento com o esquema de recurso de tempo de execução comum de 277 colunas. Um “linha jogador-jogo” significa um jogo da perspectiva de um jogador, então a maioria dos jogos produzir duas linhas de jogo de jogador.

Somente para blitz 3+0, o benchmark de produção atual usa 479.974 jogos envolvendo 129.398 jogadores únicos e 28.138.904 linhas distintas de estado de movimento. O A tabela alinhada de recursos do jogo de blitz contém 277 variáveis derivadas. O artefatos de construção rápida de recursos são ainda mais amplos no estágio de estado de movimento: o completo tabela de estado de movimento rápido usada durante a construção de recursos contém 168 colunas de estado de movimento antes da agregação em um recurso de jogo de jogador de 277 colunas mesa. Os artefatos de benchmark de tempo de execução de front-end para todos os quatro tipos de jogos são em seguida, alinhado ao mesmo esquema de recursos de jogo de jogador de 277 colunas, incluindo o variáveis de ruptura de taxonomia tática expandidas. A partir dessas tabelas de recursos, o O relatório seleciona e exibe as métricas de coaching que são mais interpretáveis.

Essas contagens são importantes para interpretar os resultados. As curvas de referência não são estimativas de algumas centenas de jogos revisados manualmente. Em vez disso, eles vêm de dezenas de milhões de estados avaliados e mais de quatro milhões de linhas de jogos de

jogadores nas quatro categorias atuais de jogos de produção.

5 Visão geral do pipeline

O pipeline de construção de referência tem seis etapas principais:

1. Selecione jogos classificados como Lichess no tipo de jogo alvo e estratos de classificação.
2. Analise cada jogada do jogo por jogada e armazene o antes e o depois do movimento estado do conselho.
3. Calcule variáveis intermediárias de estado de movimento do tabuleiro, o movimento, a perspectiva do jogador e a perspectiva do oponente.
4. Agregue variáveis de estado de movimento em variáveis de recursos do jogo do jogador.
5. Converta variáveis do jogo do jogador em métricas de relatório com uma visão clara direção de interpretação: mais alto geralmente é melhor, mais baixo geralmente é melhor, ou descritivo.
6. Calcule a média de cada métrica por bucket Elo e inspecione manualmente o resultado relação.

Essa estrutura é importante porque muitas métricas de relatório não são diretamente visível em um cabeçalho PGN. Eles exigem reconstruir a posição, determinando quais peças se moveram de quais casas iniciais, comparando antes do movimento e material e estrutura pós-movimentação e, em seguida, agregar o resultado do mover o nível para o nível do jogo do jogador.

6 Variáveis Intermediárias

As variáveis intermediárias são projetadas para descrever o que mudou no quadro e por que isso é importante do ponto de vista do relator. Os exemplos incluem:

- número do movimento, número do movimento do jogador, lado a mover, cor do movimento e jogador cor;
- de quadrado, para quadrado, tipo de peça em movimento, tipo de peça capturada e se o movimento foi uma captura, xeque, castelo, promoção ou mate;
- balanço material antes e depois da movimentação;
- se uma peça menor deixou sua casa inicial em um determinado número de movimento;
- se a mesma peça se moveu em movimentos consecutivos do jogador no abertura;
- se a dama se moveu antes do lance 10;
- lado do roque, número do movimento do roque e se as torres ficaram conectadas depois do roque;

- peças soltas, peças penduradas, peões dobrados, peões isolados, passados peões e peões atrasados;
- status do arquivo rook, arquivos abertos, arquivos semiabertos e segurança do rei características;
- oportunidades táticas, táticas executadas, táticas de pressão material, ganhos de acompanhamento garantidos e ganhos realizados.

O conjunto intermediário exato varia de acordo com o estágio de construção. Os bancos de dados de relatórios em tempo de execução mantêm apenas as colunas necessárias para fornecer relatórios rapidamente, enquanto a criação completa de recursos artefatos incluem tabelas de estado de movimento mais amplas e patches de recursos usados para construir as linhas finais de recursos do jogo do jogador.

7 Exemplos de métricas de princípio de abertura

Os princípios de abertura são exemplos úteis porque são fáceis para os usuários compreensível e fácil de conectar aos cálculos de estado de movimento.

7.1 Minor Pieces Developed by Move 10

Esta métrica conta quantos dos quatro cavalos e bispos de um jogador ainda restam suas casas no 10º lance do jogador. O cálculo intermediário usa a cor do motor, o tipo de peça móvel, o quadrado inicial e o movimento do jogador número. Para as brancas, as casas iniciais são b1, g1, c1 e f1. Para as pretas, elas são b8, g8, c8 e f8. Para cada jogo de jogador, Elo+Chess conta a casa distinta casas entre cavalos e bispos que se moveram pelo lance 10 do jogador.

Esta é uma métrica do tipo "maior é geralmente melhor". Não afirma que todo menor peça deve sempre se mover até o lance 10, mas entre os jogadores em desenvolvimento o empírico a relação é clara: os jogadores que alcançam níveis Elo mais altos tendem a desenvolver mais peças menores anteriormente.

7.2 Repeated Piece Moves in the Opening

Esta métrica conta movimentos iniciais extras gastos movendo a mesma peça novamente de desenvolver algo novo. O cálculo intermediário usa o anterior quadrado de destino, o quadrado de origem atual e o número do movimento do jogador. Se a mesma peça movida na jogada anterior do jogador e é movida novamente dentro do primeiros 10 movimentos do jogador, a contagem de movimentos repetidos aumenta.

Esta é uma métrica do tipo "menor é geralmente melhor". Ele mede tempos de abertura desperdiçados em um maneira simples. Alguns movimentos repetidos são taticamente justificados, mas em escala o a curva do balde captura a tendência geral: jogadores mais fortes no alvo os níveis gastam menos movimentos iniciais movendo a mesma peça repetidamente.

7.3 Queen Touched Before Move 10

Esta métrica conta os movimentos da rainha antes do movimento 10 do jogador. o cálculo usa o tipo de peça móvel e o número do movimento do jogador. Se a peça móvel for a rainha e o número do movimento do jogador for no máximo 10, a contagem aumenta.

Isso também é menor, geralmente melhor para o intervalo de destino Elo+Chess. Cedo movimentos da rainha podem ganhar concessões materiais ou de força, mas muitos jogadores em desenvolvimento trazer a rainha muito cedo, perder tempo para ataques à rainha e atrasar desenvolvimento de peças menores e segurança do rei.

7.4 Castling and King Safety

As métricas de roque usam bandeiras de roque, lado de roque, número de movimento de roque e o estado do tabuleiro após o roque. Os exemplos incluem se o jogador rocou, se o jogador rocou no lance 10, se o jogador rocou na ala do rei ou na ala da dama, se as torres estavam conectadas após o roque e se as fileiras de torres estavam abertos ou entreabertos após o roque.

Essas métricas não são todas interpretadas da mesma maneira. Conclusão do roque e o roque no lance 10 é geralmente maior, geralmente melhor no intervalo alvo. O roque na ala da dama ou em arquivos abertos pode depender mais do contexto. É por isso que Elo+Chess não simplesmente codifica uma regra moral; ele constrói balde curvas e inspeciona quais relacionamentos são estáveis o suficiente para coaching.

8 Curvas da caçamba e inspeção manual

Após o cálculo das métricas do jogo do jogador, Elo+Chess as agrupa por Lichess Elo bucket e calcula as médias do bucket. A média do intervalo é o empírico bruto ponto de referência para essa métrica. Uma linha de tendência é então usada no relatório para atribuir notas no estilo Elo aos valores do usuário.

As curvas são inspecionadas manualmente antes de serem usadas. Uma métrica é mais útil quando o balde significa mostrar um relacionamento coerente em toda a faixa de classificação atendido pelo produto. Uma métrica é menos útil quando é barulhenta, plana e altamente não monotônico ou orientado por casos especiais que são difíceis de explicar ao usuário.

Esta etapa de inspeção manual é especialmente importante acima da parte superior do Elo+Chess corte. Jogadores mais fortes sabem quando quebrar os princípios de abertura e fazem isso efetivamente. Acima da faixa do iniciante ao avançado, o relacionamento entre hábitos simples e classificação muitas vezes se achata ou muda de forma. O produto portanto, enfatiza a faixa onde as relações empíricas são fortes e pedagogicamente útil.

9 Exemplo de curvas de abertura completas

A Figura 5 mostra quatro curvas de princípio de abertura de blitz 3+0. Os pontos cinza mostram a faixa de classificação Lichess mais ampla. Pontos azuis e a tendência azul linha mostra a faixa 600–1600 Lichess usada pelo principal treinamento Elo+Chess benchmarks. A linha vermelha tracejada marca o corte superior.

O mesmo padrão é visível na faixa completa 10+0 rápida e mais rápida artefatos. A decisão central de modelagem permanece inalterada: usar o modelo empiricamente faixa estável para afirmações de coaching e evite fingir que uma simples curva de hábito permanece igualmente informativo para sempre.

O mesmo padrão aparece numericamente na Tabela 5. As encostas são relatado como alteração do valor da métrica por 100 pontos Elo. Roque, peça menor desenvolvimento, movimento repetido da peça e movimento inicial da rainha mostram claramente movimento entre 600-1600.

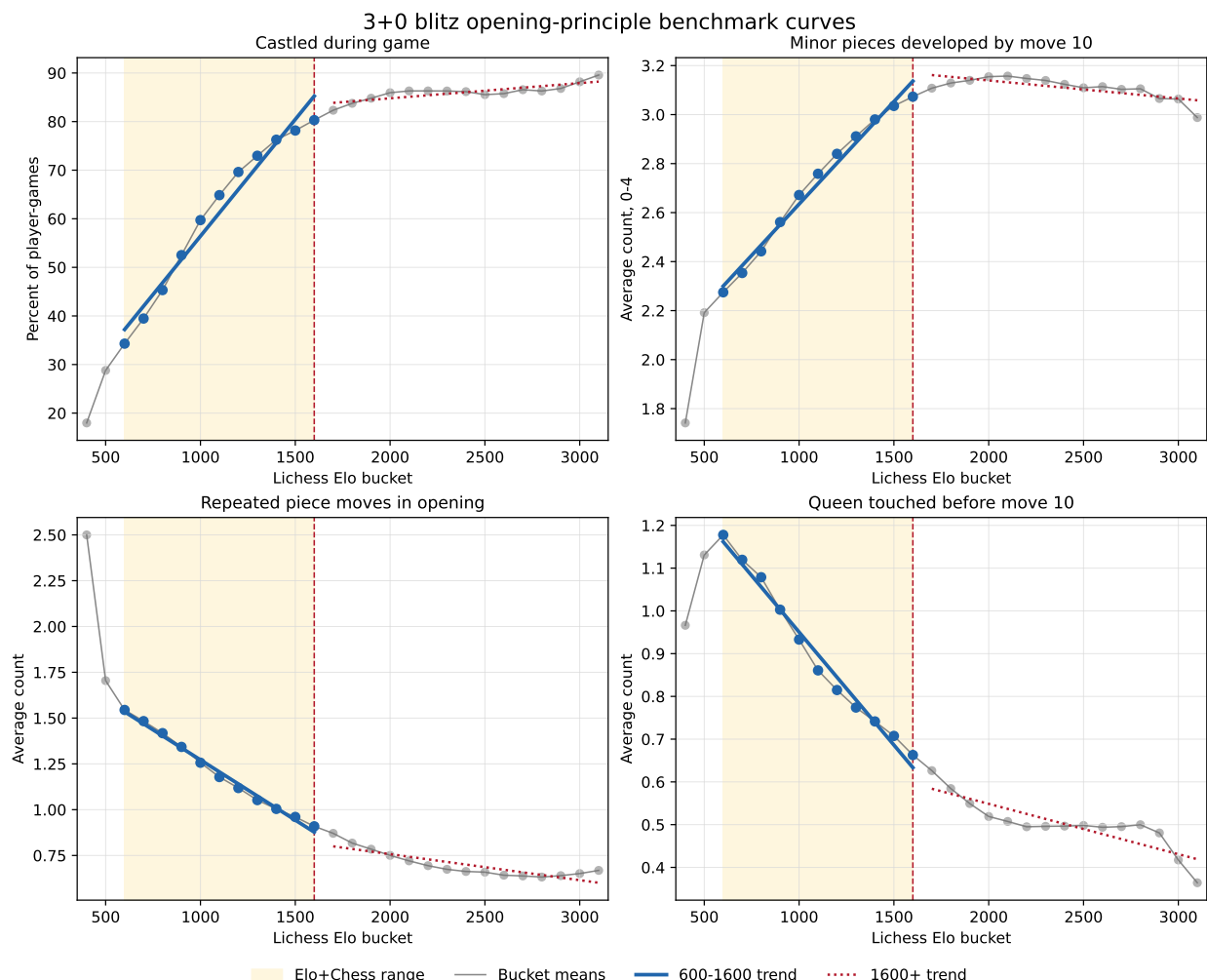


Figure 5: Curvas de benchmark de princípio de abertura selecionadas para blitz 3+0. Os relacionamentos são mais fortes e úteis na faixa de coaching Elo+Chess. Acima do limite superior, vários relacionamentos se estabilizam substancialmente.

Acima de 1600, as encostas são muito menores nestes exemplos, o que é consistente com a ideia de que jogadores com classificações mais altas quebram princípios básicos de forma mais seletiva e com mais sucesso. Um segundo mecanismo é saturação. Para alguns hábitos, a métrica atinge um teto prático ou doce ponto: jogadores fortes já aprenderam o hábito, então não há razão para a curva continue se movendo acentuadamente para cima. Mais também nem sempre é melhor depois esse ponto. Os dados, portanto, ficam achatados porque os jogadores mais fortes podem quebrar regras em posições concretas e porque muitos hábitos básicos já atingiram sua faixa operacional útil.

10 Por que o corte superior é importante

O usuário alvo do Elo+Chess é um jogador do iniciante ao avançado. Para estes jogadores, hábitos básicos estão fortemente associados à classificação: desenvolver peças, evite movimen-

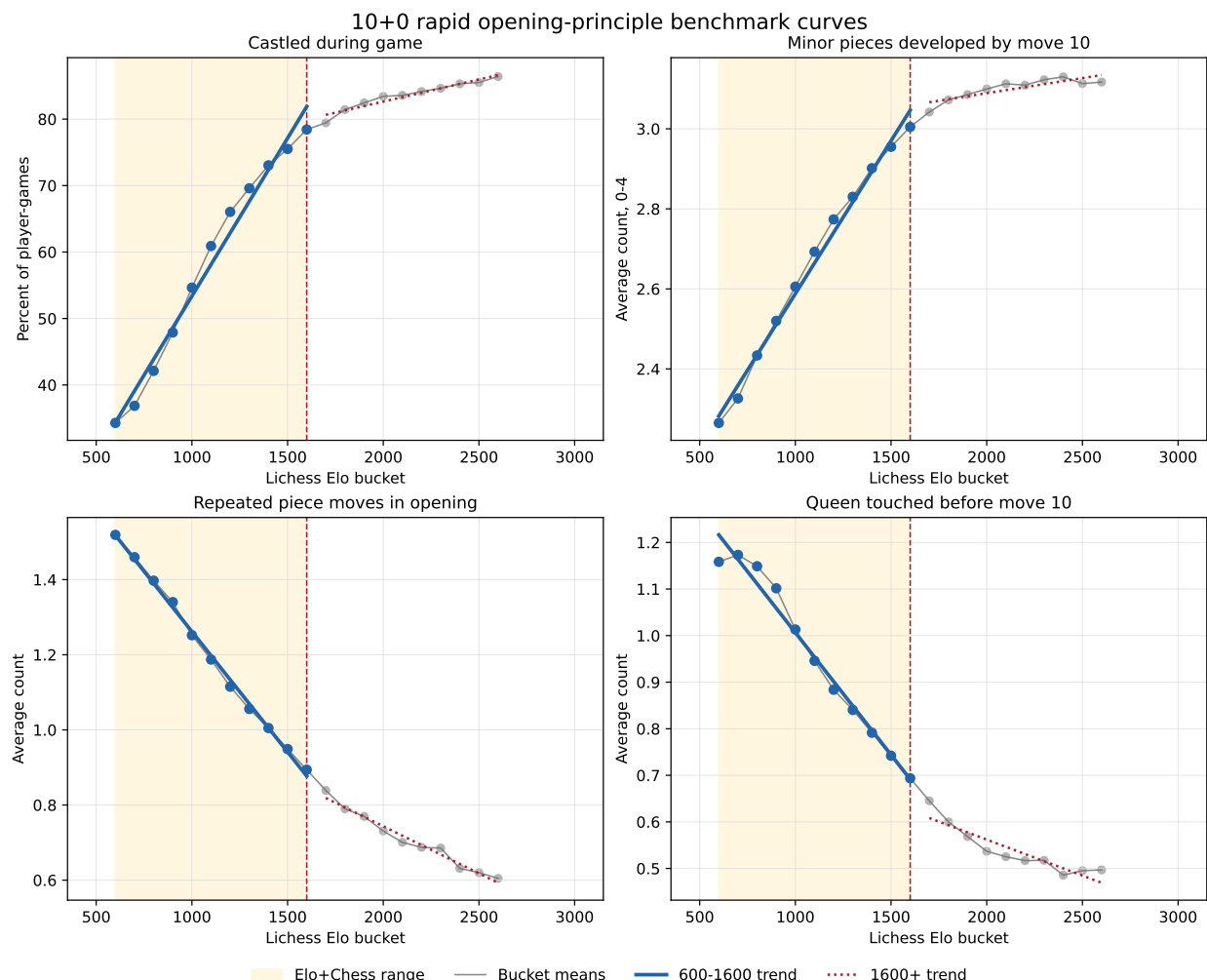


Figure 6: Curvas de benchmark de princípio de abertura selecionadas para jogos rápidos 10+0. O os relacionamentos são mais fortes e úteis na faixa de coaching Elo+Chess. Acima do limite superior, vários relacionamentos se estabilizam substancialmente.

tos repetidos das primeiras peças, evite movimentos desnecessários da rainha, castelo a tempo, conecte as torres, gerencie o material e evite deixar peças soltas ou pendurado.

Em classificações mais altas, esses mesmos hábitos superficiais tornam-se menos diagnósticos. Um forte o jogador pode atrasar o roque porque o centro está fechado, mova a rainha mais cedo porque a posição contém uma tática concreta, ou mova a mesma peça duas vezes porque ganha um ritmo ou força uma concessão estrutural. O hábito cru ainda tem significado, mas a relação entre o hábito e a classificação torna-se mais condicional. É por isso que Elo+Chess usa um corte superior para seu treinamento principal reivindicações e por que os usuários acima dessa faixa são avisados de que o relacionamento estreito entre hábitos de princípios e classificação quebra.

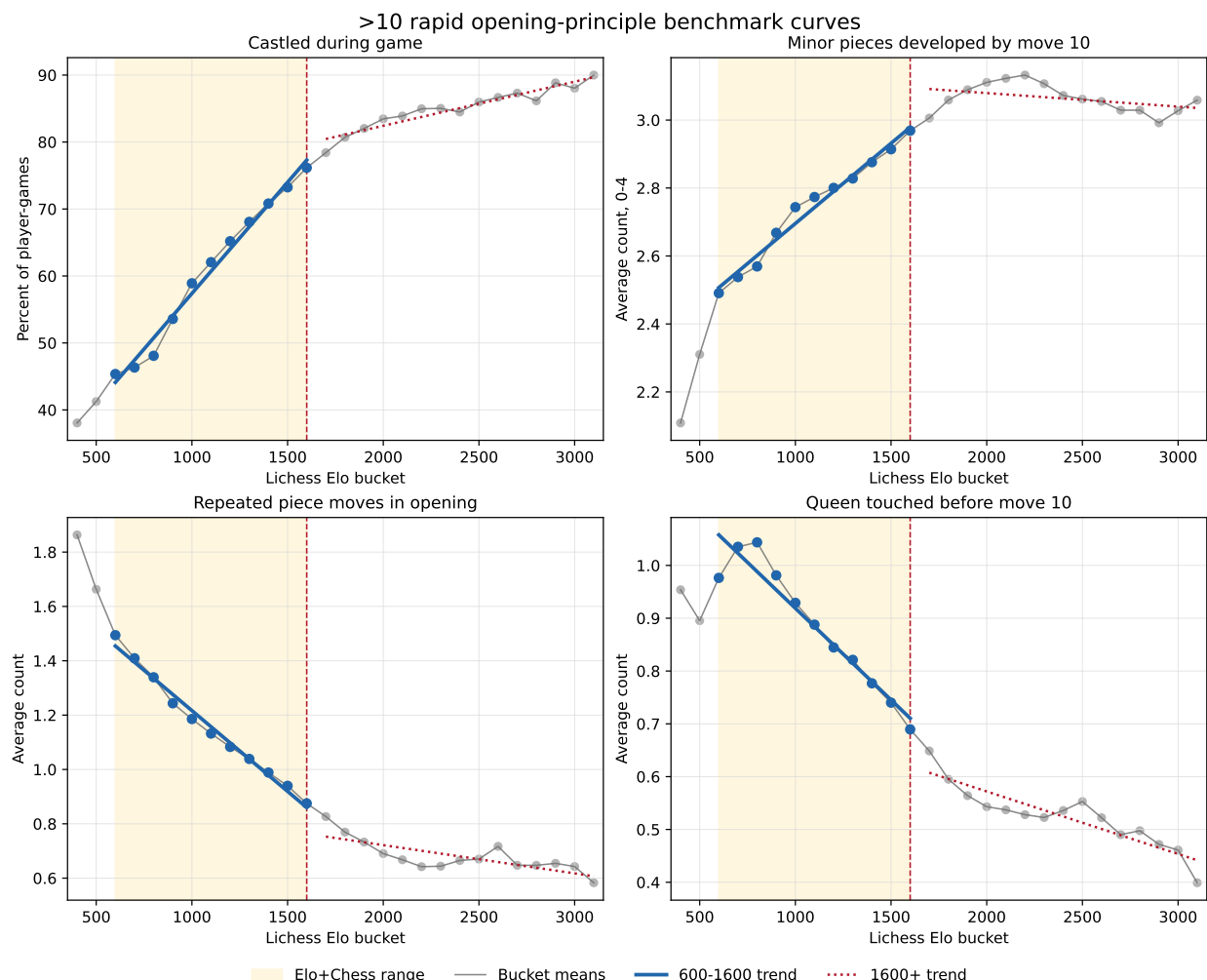


Figure 7: Curvas de benchmark de princípio de abertura selecionadas para jogos rápidos mais longos.

11 Métricas táticas bilaterais e metas percentuais

Some metrics have the opposite shape from the opening-principle examples. Forks are a useful example because the event is inherently two-sided. A player who finds and converts more useful forks than peers is doing something good, but the population rate of converted forks falls with rating because stronger opponents allow fewer fork opportunities and defend against tactical threats more actively.

A Figura 8 mostra esse efeito. Ambos os garfos convertidos por o jogador relator e os garfos convertidos pelo oponente tornam-se menos frequentes à medida que as classificações aumentam. Isso não significa que executar forks seja ruim. Significa o evento bruto a taxa é parcialmente controlada pela taxa de erro do oponente e pela textura do jogo. Para métricas assim, Elo+Chess usa percentis de histórico de pares dentro do balde Elo do próprio jogador, em vez de confiar apenas em uma tendência global média.

Para construir essas metas percentuais, Elo+Chess faz amostras de jogadores por tipo de jogo e Elo bucket usa jogos de qualificação recentes para cada jogador amostrado. O o conjunto de

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: Inclinações aproximadas da curva do balde para os quatro exemplos de princípio de abertura métricas em todos os tipos de jogos de produção.

metas atual usa 11 grupos de 600 a 1.600, 200 jogadores por balde e até 50 jogos de qualificação recentes por jogador para cada produção tipo de jogo. Isso dá 2.200 amostras de jogadores por tipo de jogo e aproximadamente 110.000 linhas de jogo de jogador por tipo de jogo para percentil dentro do intervalo distribuições.

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: Conjuntos atuais de metas percentuais de histórico de pares. Estes são amostras de histórico de jogadores usadas para estimar distribuições dentro do intervalo para métricas de estilo percentil.

Figura 9 ilustra o alvo resultante distribuição para taxa de fork convertida de jogador amostrado em blitz 3+0. Cada caixa é uma distribuição dentro do intervalo entre os jogadores da amostra, e não uma média do intervalo entre jogos individuais. Na IU do relatório, o valor de um usuário é comparado ao valor relevante caixa para o bucket Elo mapeado do usuário. Esta é a base para declarações como se a taxa de fork convertida de um jogador é alta ou baixa em relação aos pares no mesmo nível, mesmo quando a taxa geral de eventos da população diminui com a classificação.

12 Replication Cost

O método é reproduzível em princípio, mas não é leve. Reconstruindo as curvas de referência exigem:

- acesso a grandes arquivos mensais de histórico de jogos Lichess;

Two-sided converted-fork rates by Elo bucket

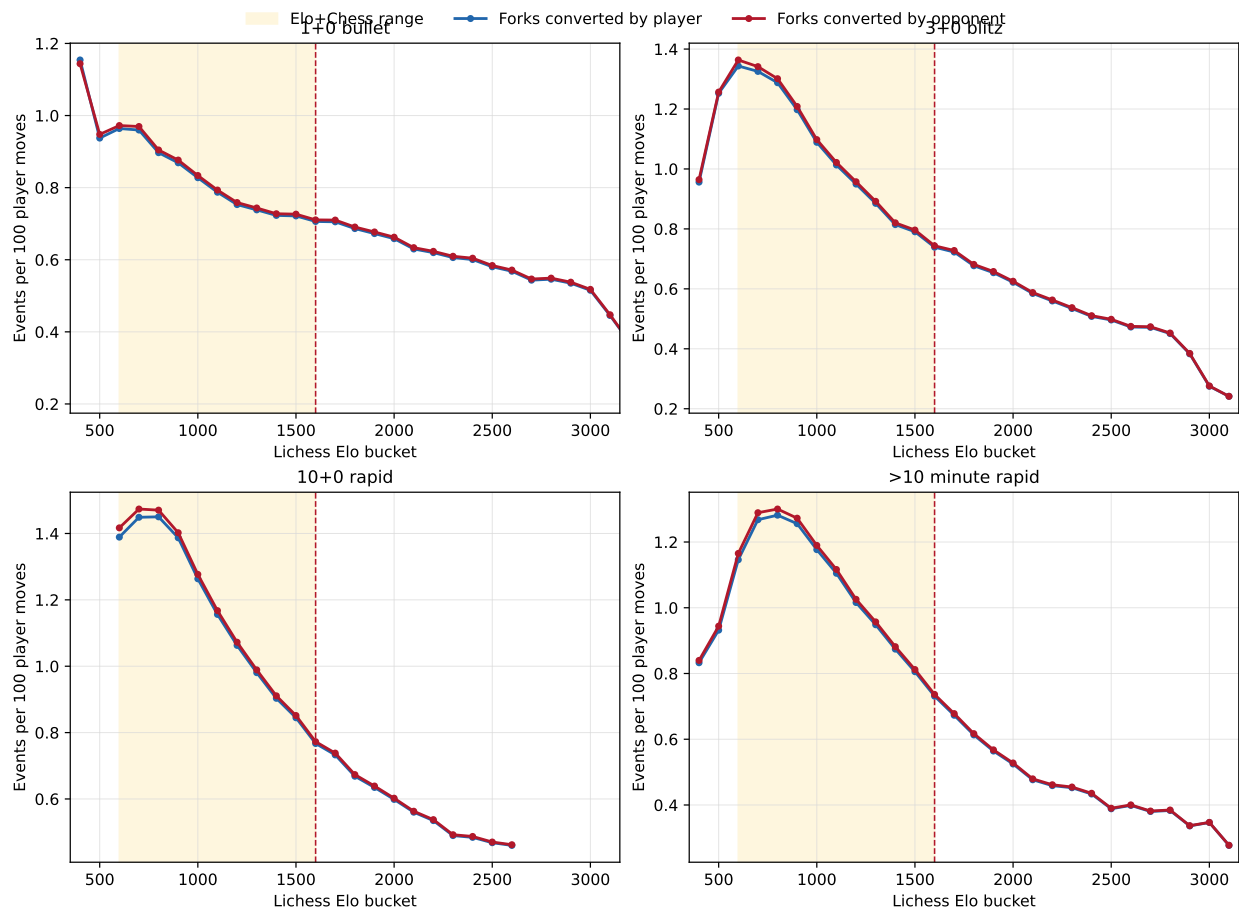


Figure 8: Taxas de garfo convertido bilateral por balde Elo. Eventos de fork convertidos caem em classificações mais altas porque ambos os lados defendem melhor e menos garfo limpo conversões estão disponíveis.

- filtragem estratificada por tipo de jogo, classificação e elegibilidade regras;
- um analisador de movimentos legais e avaliador de estado de conselho capaz de expandir dezenas de milhões de estados móveis;
- apresentam lógica para abertura, material, estrutura de peões, segurança do rei, torre atividade, táticas, gerenciamento de tempo e resultados do jogo;
- agregação do estado de movimento aos recursos do jogo do jogador;
- resumos em nível de intervalo para cada métrica de relatório;
- revisão manual de cada curva de benchmark candidata antes de ser usada em um relatório de treinamento.

Para blitz 3+0, isso significa quase meio milhão de jogos, quase um milhão linhas de jogo de jogador e mais de 28 milhões de linhas de estado de movimento. Ao longo dos quatro tipos de

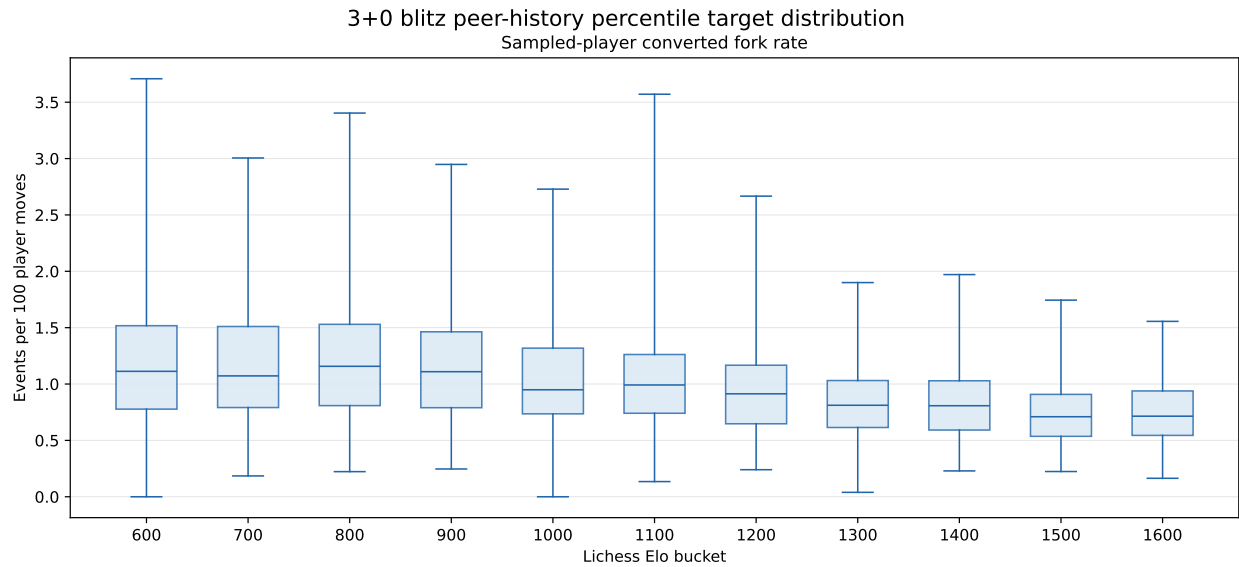


Figure 9: Exemplo de distribuição de alvo de histórico de pares de blitz 3+0 para jogador amostrado taxa de garfo convertida. Os percentis são calculados dentro do bucket Elo para que o relatório compara um jogador com pares que enfrentam oportunidades táticas amplamente semelhantes taxas.

jogos de benchmark, os artefatos de benchmark atuais contêm mais de 2,55 milhões de jogos, mais de 5,10 milhões de linhas de jogos de jogadores e mais de 147 milhões de linhas de estado de movimento. A escala é central para o valor do benchmark: permite que o relatório mostre aos usuários o que realmente acontece nos jogos jogados por jogadores próximos do seu nível, em vez de confiar apenas em slogans gerais de xadrez.

13 Limitações

The benchmark curves are empirical descriptions, not causal proof. A player does not gain rating merely by matching a bucket mean. Metrics are also imperfect summaries of complex positions: an early queen move can be excellent in one position and reckless in another. The purpose of the curves is to identify recurring habits and compare them against peer behavior, not to replace calculation or coaching judgment.

As curvas também são curvas na escala Lichess. Quando os usuários trazem o jogo Chess.com históricos, Elo+Chess mapeia o intervalo de classificação relevante no benchmark Lichess dimensionar usando o modelo de mapeamento de plataforma cruzada separado. Essa camada de mapeamento é descrito na nota de pesquisa Elo+Chess sobre Lichess-to-Chess.com conversão de classificação: [Estimating Cross-Platform Chess Rating Mappings with Regressão Modal](#). Este mapeamento não deve ser lido como preservando o percentil classificação. Ele preserva uma equivalência de classificação estimada entre os dois sistemas; a classificação percentual permanece específica da plataforma porque o jogador ativo distribuições são diferentes.

14 Gráficos de relatório de produção

O relatório de produção transforma os artefatos de benchmark em painéis de nível métrico que pode ser lido sem conhecer o pipeline de construção completo. Figuras 10 and 11 mostre três exemplos do design do relatório Elo+Chess ao vivo.

Os painéis de princípio de abertura na Figura 10 use a apresentação comum da curva de referência. Cada carta começa com o família de métricas e o nome da métrica e, em seguida, fornece uma breve regra de interpretação. O a nota grande no canto superior direito é intencionalmente separada da matéria-prima valor da métrica: responde à pergunta do usuário “como esse hábito é avaliado em relação ao benchmark?” enquanto as linhas abaixo preservam os valores medidos. As quatro linhas de resumo mostram Duração, Últimos 10, Vitórias e Perdas. Isso faz com que diagnóstico do painel em vez de meramente decorativo. Se Lifetime e Last 10 concordarem, o hábito provavelmente é estável. Se as vitórias e as derrotas divergirem, a métrica pode ser dependente do resultado ou vinculado ao estado do jogo, em vez de um simples hábito.

A área do gráfico usa um fundo escuro restrito, linha de tendência dourada e linhas coloridas. marcadores para as quatro fatias. A anotação de tendência indica explicitamente se o relacionamento de referência sobe ou desce com Elo. Isto é importante porque algumas métricas são do tipo quanto maior, melhor, como o desenvolvimento de peças menores, enquanto outros são mais baixos, melhor, como o movimento inicial da rainha. O cartão portanto não pede ao usuário para inferir a direção apenas a partir da inclinação. A final As linhas “O que isso significa” e “Dica de coaching” traduzem o resultado do benchmark em uma instrução concreta de xadrez, preservando o contexto empírico.

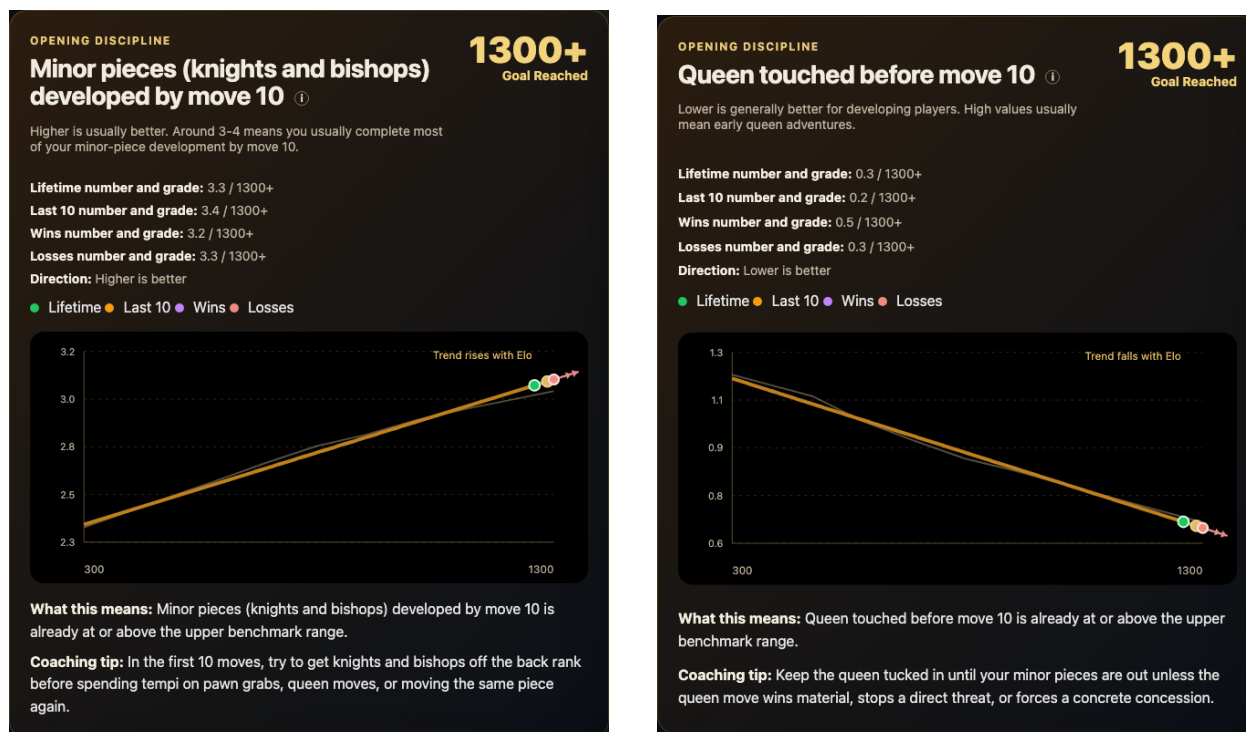


Figure 10: Production opening-principle panels. The same visual grammar handles a higher-is-better metric, minor-piece development, and a lower-is-better metric, early queen movement. Both panels expose Lifetime, Last 10, Wins, and Losses so the report can separate stable habits from recent or result-dependent patterns.

Figura 11 mostra o estilo percentil apresentação usada para métricas cuja taxa bruta não é bem representada por um tendência global única. Os garfos vencedores são um exemplo tático: a taxa de população dos garfos convertidos mudam tanto com a habilidade do jogador quanto com a taxa de erro do oponente. O cartão, portanto, compara o valor do usuário com os históricos dos jogadores pares no balde Elo mapeado. O boxplot exibe a distribuição de pares, incluindo o intervalo interquartil, mediana e bigodes. O marcador do usuário é plotado no mesma escala e rotulados com valor bruto e classificação percentual. Adjacente os intervalos de comparação permanecem visíveis de forma silenciada para que o usuário possa ver que o o peer bucket é local, não uma escala universal.

Esse design torna as métricas percentuais auditáveis. O usuário vê o valor bruto, o percentil, o intervalo de pares usado para comparação e o traduzido Mapeamento de bucket Chess.com para Lichess. O mesmo tempo de vida, últimos 10, vitórias e As linhas de perdas são mantidas para que as pontuações percentuais táticas ainda possam ser verificadas para efeitos de recência e dependência de resultados.

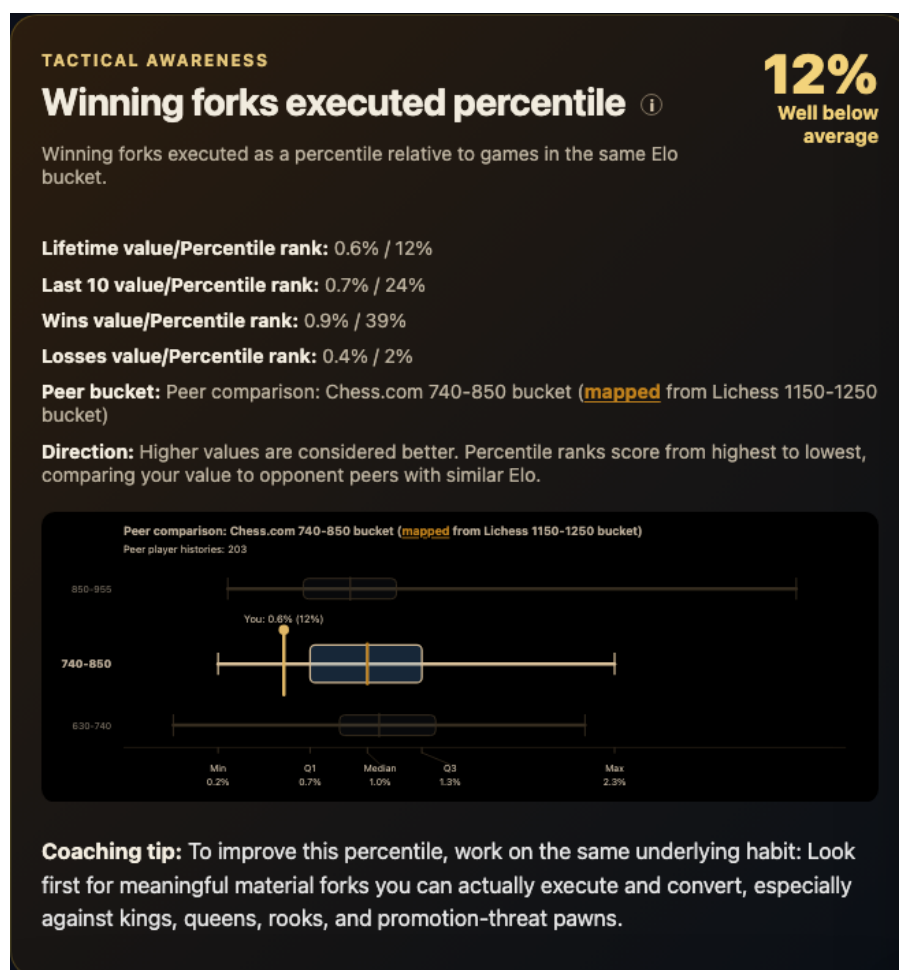


Figure 11: Painel de percentil de produção para garfos vencedores executados. O boxplot mostra a distribuição de pares no mesmo intervalo usada pelo relatório, enquanto o ouro marcador mostra o valor bruto e a classificação percentual do usuário.

15 Conclusão

Os relatórios de benchmark Elo+Chess são construídos a partir de um grande cálculo de estado de movimento ao longo de amostras Lichess estratificadas. Cada métrica começa como um conjunto de estados concretos do quadro e variáveis de estado de movimento, torna-se um recurso do jogo do jogador e é então calculada a média pelo balde Elo para formar uma curva de referência empírica. O princípio de abertura exemplos mostram por que esta abordagem é útil: hábitos simples de xadrez têm fortes, relações visíveis com a classificação na faixa-alvo, enquanto aquelas os relacionamentos muitas vezes se estabilizam ou se tornam mais condicionais acima do limite superior. Essa é a razão central pela qual Elo+Chess concentra suas reivindicações de coaching na classificação faixa onde os dados os apoiam mais claramente.