

Criação de um Dataset de Múltiplos Domínios para a Plataforma de Jogos Online da Epic Games

Samuel S. Gomes¹, Carlos H. G. Ferreira¹, Alexandre M. Sousa¹

¹Departamento de Computação e Sistemas – Universidade Federal de Ouro Preto (UFOP)

samuel.gomes2@aluno.ufop.edu.br, {chgferreira, alexandre.sousa}@ufop.edu.br

Abstract. *The market for online games is growing year by year, but some gaming platforms do not have a structured, organized and public dataset to facilitate the research. Therefore, the goal of this work is to create a dataset that contains information from Epic Games games in a way that incorporates information from different domains such as their respective websites, social networks, and OpenCritic game reviews. The dataset created contains information on nearly one thousand games, 17 thousand reviews, and nearly 1 million of tweets.*

Resumo. *O mercado de jogos online aumenta ano após ano, no entanto, algumas plataformas de jogos não possuem um repositório de dados estruturado, organizado e público para facilitar pesquisas. Dessa forma, o objetivo deste trabalho é criar um dataset que contenha informações dos jogos da Epic Games de maneira que inclua informações de diferentes domínios, tais como do seu respectivo site, de suas redes sociais e avaliações de jogos do OpenCritic. O dataset construído tem informações de quase mil jogos, 17 mil avaliações e quase 1 milhão de tweets.*

1. Introdução

O mercado de jogos digitais vem crescendo significativamente nos últimos anos, impulsionado ainda mais pela pandemia do Covid-19. De acordo com a PWC, o mercado cresceu 32% entre 2019 e 2021 e possui uma taxa de crescimento anual de 8.4% até 2026, tornando-se uma indústria avaliada em 321 bilhões de dólares [PwC 2022]. Com a popularidade dos jogos, o público-alvo agora inclui pessoas de todas as idades. Estudos mostram que 60% dos norte-americanos jogam videogames e jogos de computador pelo menos uma vez ao dia, e 70% dos jogadores são adultos com média de 34 anos de idade [Nyitray 2019]. O aumento da qualidade das ferramentas online incentivou ainda mais jogadores de todo o mundo a se comunicarem no ecossistema criado pelas plataformas de jogos online, gerando uma grande quantidade de dados. Dados esses, que atraem o interesse de grandes empresas para investir na compreensão desses dados e alavancar ainda mais esse mercado.

A Epic Games¹ é uma das maiores empresas de desenvolvimento de jogos do mundo, conhecida por criar um ecossistema digital que fornece uma plataforma de serviços para jogos em larga escala. Através de sua plataforma, a Epic Games disponibiliza diversos jogos e recursos que tornam a experiência do jogador mais completa e interativa. No entanto, a plataforma não disponibiliza uma Interface de Programação de Aplicação (API) pública com os dados de seus jogos e usuários, prática adotada por concorrentes como a Steam², que fornece uma API pública que contém mais de 108 milhões de contas de usuários e

¹<https://store.epicgames.com/pt-BR/>

²https://partner.steamgames.com/doc/webapi_overview?l=brazilian

mais de 384 milhões de jogos proprietários, que tem permitido o desenvolvimento de vários estudos relacionados a análise textual dos comentários dos usuários [O'Neill et al. 2016, Baowaly et al. 2019, Loria et al. 2021]. Isso levanta a necessidade de criar um repositório de dados da Epic Games que seja de fácil acesso e que possa evoluir com o passar dos anos. Neste contexto, o objetivo deste trabalho é criar um dataset estruturado que contenha dados de jogos e jogadores, recuperando dados utilizando técnicas de *Web Scraping* no site da loja da plataforma digital da Epic Games, nos fóruns de discussões, nas avaliações dos jogos e também nas contas de redes sociais vinculadas aos jogos.

O restante deste artigo está estruturado da seguinte maneira. A próxima seção apresenta a importância das plataformas de jogos online como oportunidade de pesquisa. A Seção 3 aborda os dados coletados e seus diferentes domínios. A análise da caracterização preliminar é apresentada na Seção 4. Por fim, as conclusões são apresentadas na Seção 5.

2. Estudos Sobre Plataformas de Jogos Digitais

Os jogos online estão se tornando cada vez mais populares na internet, com aproximadamente 82% dos usuários em 2022 utilizando diferentes plataformas de jogos [Statista 2022]. Empresas líderes de mercado, como Epic Games³, Steam⁴ e Origin⁵, buscam atrair e manter usuários em suas plataformas. A interação dos usuários em plataformas que disponibilizam alguma API, como a Steam, tem sido objeto de estudo na literatura. Por exemplo, O'Neill [O'Neill et al. 2016] coletou dados sobre perfis de usuários, amizades, jogos, horas jogadas e grupos na plataforma Steam para melhor compreender o comportamento dos jogadores. O autor utilizou a API pública fornecida pela Steam para coletar os dados, destacando a importância de repositórios de dados públicos para pesquisas. Vagavolu [Vagavolu et al. 2021] criou um conjunto de dados para armazenar informações de 536 repositórios de código-fonte no GitHub sobre motores de jogos para ajudar os desenvolvedores de jogos. O desenvolvimento de jogos é um processo complexo que requer integração de diversas áreas do conhecimento, portanto, um repositório de dados público é de grande ajuda para facilitar a pesquisa nessa área.

De fato, dados dessa natureza têm sido bastantes úteis para outros estudos. Por exemplo, os autores de [Loria et al. 2021] analisaram como as conexões sociais são formadas em torno de jogos na plataforma Steam e como elas variam dependendo do tipo e mecânica de cada jogo. Para isso, os pesquisadores modelaram a estrutura de amizades entre os usuários da plataforma Steam em torno de 200 jogos em um grafo, dividindo-os em comunidades para comparar suas propriedades de rede e características específicas. Os resultados mostram que as tags definidas pelos usuários da Steam caracterizam melhor as comunidades de jogos do que o próprio gênero do jogo. Além disso, observou-se que jogos multijogadores têm redes mais coesas e agrupadas do que jogos com foco em jogabilidade de um único jogador, indicando que jogar em equipe é mais propenso a produzir capital social (ou seja, amizades na Steam).

Baowaly et al. avaliaram a utilidade das análises de jogos na loja online Steam [Baowaly et al. 2019]. Usando dados de análises de usuários de diferentes gêneros de jogos, os autores propuseram um modelo de classificação para prever se essas análises são úteis ou não. Rizani et al. [Rizani et al. 2023] se concentraram no impacto das conquistas

³<https://store.epicgames.com>

⁴<https://store.steampowered.com>

⁵<https://www.ea.com>

da Steam nas avaliações dos jogos e nos tipos de jogos. O estudo analisou 18.658 jogos listados na Steam e mostrou que o selo (ou seja, conquistas dos jogos) aumenta o envolvimento dos jogadores com a plataforma de publicação com bons dados auxiliares (como tipos, classificação, lançamentos e preços).

Ortogonal a esses esforços, este trabalho tem como principal contribuição fornecer uma base de dados única da plataforma Epic Games, combinada com dados do Twitter e avaliações da OpenCritic, para pesquisas na área de jogos online e redes sociais online. Como não existem trabalhos semelhantes disponíveis para a plataforma Epic Games, acredita-se que o uso desses dados combinados pode levar a uma compreensão mais aprofundada do comportamento dos jogadores e da interação social em torno dos jogos da plataforma. Além disso, a base de dados pode ser valiosa para desenvolvedores de jogos, pesquisadores acadêmicos e profissionais de marketing que buscam entender melhor o mercado de jogos online. Portanto, este trabalho é uma contribuição importante para a comunidade acadêmica e profissional interessada em jogos online e redes sociais, pois oferece novas perspectivas e possibilidades de pesquisa.

3. Metodologia

A Epic Games Store é uma das maiores plataformas de jogos digitais do mundo e sua plataforma é dividida em categorias nas quais os jogos são recomendados. Entre essas categorias estão: (1) jogos em promoção; (2) jogos gratuitos; (3) novos lançamentos; (4) mais vendidos; (5) lançamento em breve; (6) jogos atualizados recentemente; (7) jogos novos na loja; e, por fim, (8) jogos mais populares. A coleta de dados foi realizada no mês de junho de 2022. Para essa coleta de dados via *web scraping* no site da Epic Games, todas essas categorias foram analisadas e, entre as informações consideradas mais úteis, podem ser citadas: identificador do jogo, nome, apelido, preço, data de lançamento, plataforma (e.g., Windows e Mac), descrição, empresa desenvolvedora, empresa publicadora e os gêneros do jogo. Além desses dados, também foram coletados informações de configuração mínima de hardware recomendada para execução do jogo, as quais incluem o sistema operacional, o processador, a memória RAM e recurso gráfico. Por fim, para jogos que apresentam informações de conta de redes sociais online (se houver), também foram coletados: o nome da rede social online (e.g., Twitter, Facebook, dentre outras) e sua respectiva URL para página da conta.

Além dos dados específicos de cada jogo dentro do domínio da plataforma, também foram considerados o site da OpenCritic⁶ e as contas da rede social online dos jogos no Twitter⁷. Por meio da API do Twitter, foram coletadas informações das contas de jogos que estavam disponíveis no momento da coleta, bem como todos os tweets publicados por essas contas. As seguintes informações foram coletadas: identificador da conta, nome do usuário, apelido da conta, descrição, localização, website, data que o usuário ingressou no Twitter, número de *following* e *followers*. Dos tweets das contas foram coletados: o identificador do tweet, URL da mídia (foto ou vídeo), quantidade de curtidas, citações, retweets, respostas, data e hora da publicação do tweet.

Finalmente, por meio da API fornecida pela OpenCritic foram coletadas todas as

⁶OpenCritic é um site que agrega avaliações de jogos online bem como gera pontuações com base em várias análises feitas em outras plataformas por críticos e profissionais da área. Está disponível em: www.opencritic.com.

⁷A escolha do Twitter como rede social online para coleta de dados justifica-se pela facilidade de uso de sua API pública.

Tabela 1. Descrição dos arquivos que constituem o dataset.

Arquivo	Conteúdo
games.csv	game_id, name, game_slug, price, release_date, platform, description, developer, publisher, genres
necessary_hardware.csv	hardware_id, game_id, operational_system, processor, memory, graphics
open_critic.csv	open_critic_id, game_id, rating, comment, company, author, date, description, top_critic
social_networks.csv	social_id, description, url, game_id
twitter_accounts.csv	twitter_account_id, game_id, name, username, bio, location, website, joined_date, following, followers
tweets.csv	tweet_id, twitter_account_id, text, url_media, quantity_likes, quantity_retweets, quantity_quotes, quantity_replies, in_reply_to_user_id, timestamp

avaliações dos jogos existentes na plataforma, lembrando que em cada página de jogo na Epic Games Store existe uma referência ao OpenCritic, o qual possui as avaliações do jogo (em formato de texto). Entre os dados sobre avaliações também estão: identificador da avaliação, nota, comentário, empresa responsável (caso houver), autor, data, descrição do jogo e um atributo que indica se a crítica é de uma empresa com veredicto ou não. A Tabela 1 apresenta a descrição dos arquivos que compõem o dataset em que os atributos de mesma cor conectam diferentes arquivos e formam seu respectivo esquema de dados. Por fim, o dataset está disponível publicamente para download na plataforma Zenodo⁸.

4. Sumarização e Análise Preliminar do Dataset

Tabela 2. Estatísticas básicas por entidade.

Entidade	Total Geral	Média
Jogos	915,00	–
OpenCritic (avaliações)	17.584,00	19,21
Redes Sociais	3.045,00	3,33
Contas Twitter	522,00	0,57
Tweets	989.495,00	1.895,58

A Tabela 2 apresenta as estatísticas básicas para o dataset. A coluna “Entidade” apresenta cada entidade de dados e recebeu um nome intuitivo correspondente ao domínio da fonte da coleta, por sua vez, a coluna “Total Geral” mostra o número total de elementos coletados. Por fim, a última coluna apresenta a média de avaliações por jogo, a média de redes sociais por jogo, a média de contas do Twitter por jogo (ressalta-se que nem todos os jogos possuem uma conta no Twitter), bem como o número médio de tweets por conta do jogo no Twitter, respectivamente.

A classificação de gêneros dos jogos no dataset é bastante diversificada, a Figura 1a apresenta os Top 10 gêneros de jogo com maior quantidade de jogos na Epic Games. Também é apresentado na Figura 1b as Top 10 empresas desenvolvedoras que possuem mais jogos no catálogo da Epic Games. Para se ter uma ideia dos valores de preço dos jogos, uma função de distribuição acumulada (CDF) do atributo *preço* é apresentada na Figura 3a. O gráfico mostra que 60% dos jogos são gratuitos ou tem um valor de até R\$ 20,00 e 95% dos jogos apresentam preço menor ou igual a R\$ 60,00.

Para as avaliações do OpenCritic, a Figura 2 apresenta as Top 10 empresas com mais avaliações. De maneira complementar, na Figura 3b é apresentada a CDF das notas dos jogos nas avaliações, nessa figura é possível observar que 80% das avaliações dos jogos tem uma nota maior que 60 em uma escala de 0 a 100.

⁸Dataset disponível em: <https://zenodo.org/record/7606569>.

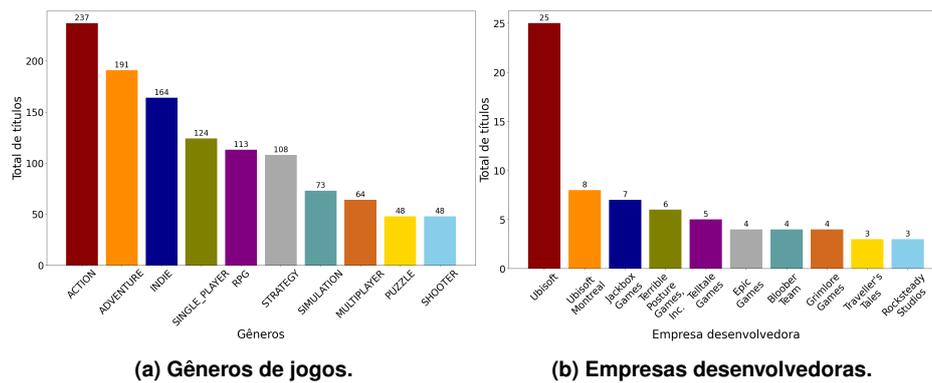


Figura 1. Top 10 gêneros e empresas desenvolvedoras com mais jogos.

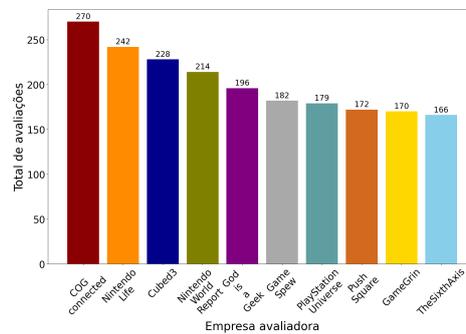


Figura 2. Top 10 empresas com maior número de críticas.

Com relação às redes sociais dos jogos, para aqueles que têm alguma rede social, foi constatada uma preferência pelas redes sociais Twitter (760), Facebook (635) e Discord (444), respectivamente, as quais representam cerca de 60% do total. Como o Twitter é o mais popular entre as redes sociais das empresas de jogos, neste trabalho, isso justifica a escolha dessa rede social como a primeira para realização da coleta.

Com objetivo de verificar quais jogos apresentam maior engajamento com seus seguidores, foram identificadas as contas com o maior número de seguidores: (1º) *Sonic* com 5.990.318, (2º) *Hedgehog* com 5.240.272, (3º) *Fortnite* com 4.895.908 e (4º) *League of Legends* com 4.449.399. Adicionalmente, a Figura 3c mostra que 90% dos jogos seguem no máximo 1000 contas no Twitter, em contraste, o número de seguidores para cada conta de jogos supera em até três ordens de grandeza o número de contas as quais elas seguem (o que é comum no Twitter para contas populares), vide Figura 3d. Além disso, essa última figura mostra que 60% dos jogos tem no máximo 10 mil seguidores e somente 30% dos jogos têm um número de seguidores entre 10 mil e 150 mil seguidores. Essas informações mostram que existem poucas empresas de jogos com contas no Twitter que conseguem engajamento massivo na ordem de milhões de seguidores, tais como as quatro contas mencionadas no início deste parágrafo. Para finalizar a análise sobre o engajamento das contas de jogos no Twitter, a Figuras 4a, 4b, 4c e 4d mostram as Funções de Distribuição Acumulada Complementar (CCDFs) para o número de curtidas, retweets, respostas e citações dos tweets dessas contas. É possível notar que a distribuição dessas ações nos tweets seguem um comportamento semelhante ao da distribuição do número de seguidores por conta no Twitter (Figura 3d), ou seja, poucos jogos possuem um grande engajamento com seus jogadores, o que também é refletido nas ações desses tweets.

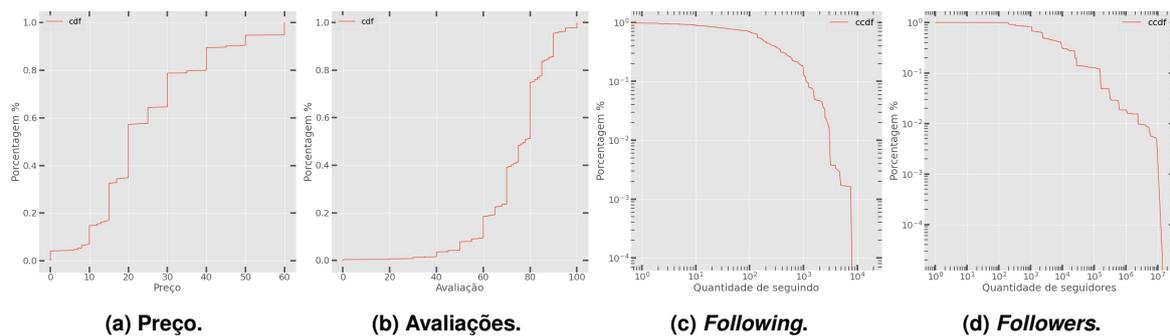


Figura 3. CDF para preço e avaliações e CCDF para *following* e *followers*.

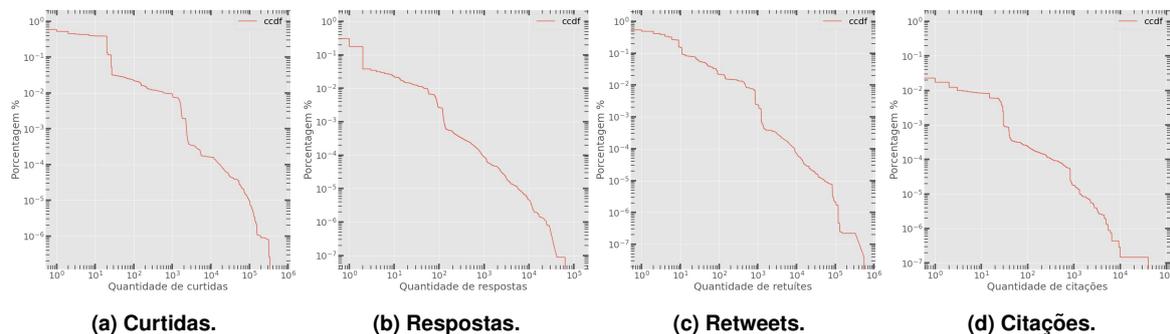


Figura 4. Função de Distribuição Acumulada Complementar (CCDF).

5. Conclusões

Este trabalho teve como objetivo fornecer uma base de dados estruturada da plataforma de jogos online da Epic Games, contribuindo para a área de pesquisa em jogos online e redes sociais online. O dataset resultante é o primeiro conjunto de dados da Epic Games que combina diferentes tipos de domínios. Acredita-se que essa base de dados seja útil para diversos públicos, incluindo desenvolvedores de jogos, pesquisadores acadêmicos e profissionais de marketing que buscam entender melhor o mercado de jogos online. Portanto, este trabalho é uma contribuição importante para a comunidade acadêmica e profissional interessada em jogos online e redes sociais, pois oferece novas possibilidades de pesquisa, como a identificação de tendências de jogabilidade, análise de perfil e comunidades de jogadores, e identificação de jogadores maliciosos. Como trabalhos futuros, espera-se migrar os arquivos do dataset para um banco de dados para garantir sua consistência e constante atualização.

Referências

- Baowaly, M. K., Tu, Y.-P., and Chen, K.-T. (2019). Predicting the helpfulness of game reviews: A case study on the steam store. *Journal of Intelligent & Fuzzy Systems*, 36(5):4731–4742.
- Loria, E., Antelmi, A., and Pirker, J. (2021). Comparing the structures and characteristics of different game social networks—the steam case. In *2021 IEEE Conference on Games (CoG)*, pages 1–8. IEEE.
- Nyitray, K. J. (2019). Game on to game after: Sources for video game history. *Reference and user services quarterly*, 59(1):7.
- O’Neill, M., Vaziripour, E., Wu, J., and Zappala, D. (2016). Condensing steam: Distilling the diversity of gamer behavior. In *Proceedings of the 2016 internet measurement conference*, pages 81–95.
- PwC (2022). Perspectives from the global entertainment & media outlook 2022–2026. <https://www.pwc.com/gx/en/industries/tmt/media/outlook/outlook-perspectives.html>.
- Rizani, M. N., Khalid, M. N. A., and Iida, H. (2023). Application of meta-gaming concept to the publishing platform: Analysis of the steam games platform. *Information*, 14(2):110.
- Statista (2022). Share of internet users worldwide who play video games on any device as of 3rd quarter 2022. <https://www.statista.com/statistics/195768/global-gaming-reach-by-country/>.
- Vagavolu, D., Agrahari, V., Chimalakonda, S., and Venigalla, A. S. M. (2021). Ge526: A dataset of open-source game engines. In *2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR)*, pages 605–609. IEEE.