Mapeando Fandoms: Uma Análise das Fanfictions e suas Dinâmicas no Archive of Our Own

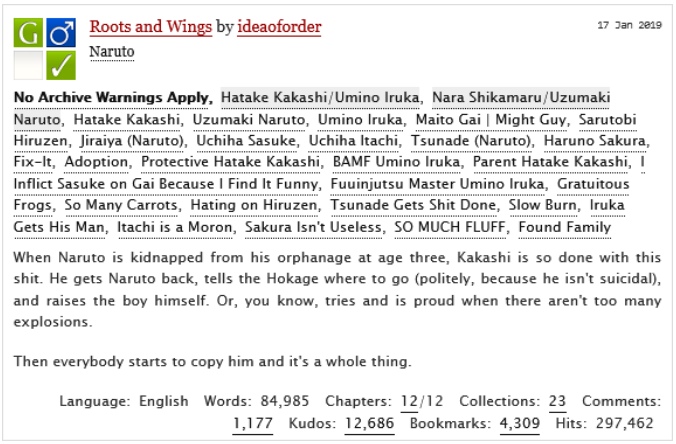
Ana Flávia de Matos Souza
Bacharelado em Ciência da Computação
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

***Resumo*—Fanfiction, or fan-created stories based on existing books, movies or other media, has seen significant growth with the advent of the internet, fostering massive communities around these creations. Archive of Our Own (AO3) stands out as a major platform, hosting over 12,74 million works across 64.330 fandoms. This study aims to map the dynamics of fanfictions within AO3 by creating a graph based on the interconnectedness of these works through content tags. The primary objective is to analyze the structure and characteristics of fanfiction networks, focusing on how different fandoms interact and the patterns within them. The data was collected through web scraping and processed to create graphs that represent the relationships between fanfiction works and their tags. Key analyses include the calculation of network metrics, sentiment analysis of story summaries, and the identification of thematic trends within fandoms.**

I. Introdução

As fanfictions, fanfics (fics) são histórias produzidas por fãs e são baseadas em livros, filmes, seriados, e quadrinhos. Geralmente envolvem os cenários, os personagens e as tramas da obra original, ou ainda fazem o cruzamento de duas ou mais obras, misturando de forma harmônica seus enredos e personagens, para compor uma nova história [1]. Nos últimos anos, houve um crescimento imenso na produção dessas fanfictions, além do surgimento de comunidades de fãs que giram em torno dessas criações. Esse fenômeno levou à criação de plataformas como o Archive of Our Own (AO3)[1], um site que em março de 2024 possuía mais de 12.740.000 obras em 64.330 "fandoms", que podem ser definidos como um grupo de fãs de uma pessoa, time ou obra de ficção, considerados de forma coletiva como uma subcultura [2].

Dentro do AO3, as obras são categorizadas conforme os fandoms aos quais elas pertencem através do uso de "tags", designadas pelos próprios autores. Essas tags podem indicar os relacionamentos, personagens ou conteúdo da história, desempenhando um papel essencial na busca e filtragem das fanfics de acordo com as preferências do leitor. Informações como o número de capítulos, a data de publicação e última atualização, a quantidade de visualizações e a classificação indicativa de cada obra hospedada também são apresentadas. A Fig. 1 mostra um exemplo de uma fanfiction hospedada no Archive of Our Own. No cabeçalho da fanfic temos o título e o autor da obra, com o fandom logo abaixo. À direita do cabeçalho encontramos a data da última atualização do trabalho, e logo abaixo ficam as tags de aviso de conteúdo, relacionamentos, personagens e conteúdo geral. Após as tags podemos encontrar a sinopse e no final temos as informações de idioma, contagem de palavras, número de capítulos, comentários e visualizações.



Roots and Wings by ideaoforder
17 Jan 2019
Naruto

No Archive Warnings Apply, Hatake Kakashi/Umino Iruka, Nara Shikamaru/Uzumaki Naruto, Hatake Kakashi, Uzumaki Naruto, Umino Iruka, Maito Gai | Might Guy, Sarutobi Hiruzen, Jiraiya (Naruto), Uchiha Sasuke, Uchiha Itachi, Tsunade (Naruto), Haruno Sakura, Fix-It, Adoption, Protective Hatake Kakashi, BAMF Umino Iruka, Parent Hatake Kakashi, I Inflict Sasuke on Gai Because I Find It Funny, Fuuinjutsu Master Umino Iruka, Gratuitous Frogs, So Many Carrots, Hating on Hiruzen, Tsunade Gets Shit Done, Slow Burn, Iruka Gets His Man, Itachi is a Moron, Sakura Isn't Useless, SO MUCH FLUFF, Found Family

When Naruto is kidnapped from his orphanage at age three, Kakashi is so done with this shit. He gets Naruto back, tells the Hokage where to go (politely, because he isn't suicidal), and raises the boy himself. Or, you know, tries and is proud when there aren't too many explosions.

Then everybody starts to copy him and it's a whole thing.

Language: English Words: 84,985 Chapters: 12/12 Collections: 23 Comments: 1,177 Kudos: 12,686 Bookmarks: 4,309 Hits: 297,462

Figura 1. Exemplo de uma fanfic no AO3.

Temos atualmente um cenário em que grupos de fãs estão tendo suas tendências de comportamento e consumo cada vez mais intensificadas com a era digital [2], o que também nos fornece um volume muito mais amplo de dados que podem ser explorados. Isso nos proporciona uma base sólida para explorar padrões e características recorrentes nas fanfics de cada fandom, bem como a maneira que esses fandoms diferentes interagem entre si, visando com isso entender melhor o comportamento dessas comunidades.

Diante disso, este trabalho busca construir uma compreensão aprofundada das dinâmicas entre os fãs e as obras que os inspiram online, assim como sobre as preferências e tendências desses indivíduos. Ao examinar os padrões e conexões presentes nas fanfics, nossa intenção é não apenas mapear a rica diversidade desse fenômeno cultural, mas também fornecer uma visão abrangente e inovadora sobre a criatividade e a cultura participativa na era digital.

O objetivo principal deste estudo será realizar a coleta de dados das fanfictions hospedadas no Archive of Our Own, buscando então construir um grafo das obras que evidencie suas conexões e agrupamentos. A análise desse grafo permitirá a análise das características de mais alto nível das fanfics, incluindo o cálculo de métricas específicas de rede e a avaliação da similaridade entre o conteúdo gerado por fandoms distintos.

[1]https://archiveofourown.org/

Tendo isso em vista, as perguntas de pesquisa que orientaram este estudo são:

- Qual é o perfil da rede de fanfics e tags do Archive of Our Own?
- Existe um padrão geral no conteúdo consumido em diferentes fandoms, ou cada um possui um perfil individual?

Os principais resultados encontrados incluem uma variação significativa nas métricas das redes, indicando diferenças estruturais e de comportamento entre diferentes idiomas. A análise de sentimentos mostrou que a maioria das sinopses das histórias possui um tom neutro, sem variações notáveis entre diferentes fandoms. Além disso, a análise de similaridade apresentou que alguns fandoms possuem uma taxa de similaridade baixa com todos os demais fandoms, indicando que eles utilizam um conjunto de tags diferente dos demais.

II. Referencial Teórico

Na literatura, diversos estudos têm se dedicado à análise computacional das fanfictions, com alguns deles concentrados no site Archive of Our Own. No entanto, grande parte dessas pesquisas foca na construção de redes de personagens, ao invés de utilizar as tags de conteúdo para estabelecer as interações nas redes. A seguir, apresentamos alguns trabalhos que analisam e caracterizam os dados deste site.

O estudo realizado por [3] coletou dados do AO3 para investigar 7 redes de personagens em fanfics de crossover, que são aquelas que envolvem mais de um fandom na história. Este trabalho gerou um grafo de tamanho significativo, com mais de 98.000 nós e 2.700.000 arestas, além de identificar padrões nas obras analisadas, como a prevalência de personagens originais independente dos fandoms de cada história.

Por outro lado, em [4] a autora explora a forma como os usuários interagem com as tags do Archive of Our Own e com a plataforma em geral, bem como a qualidade dessas interações. O foco deste estudo reside mais na maneira como os usuários respondem às tags, uma vez que se trata de uma pesquisa voltada para a área de interação humano-computador. Por conta disso, os dados utilizados no artigo baseiam-se em entradas de diário feitas pelos nove participantes do estudo, o que difere da abordagem que será adotada neste projeto.

Ainda no âmbito das fanfictions, mas com um foco em outra plataforma, em [5] os autores conduzem uma análise empírica das obras presentes no site FanFiction[2], utilizando métodos computacionais. Neste estudo as palavras de cada história são separadas em tokens e comparadas com aquelas no material original para identificar diferenças no conteúdo dos dois. Além disso, o artigo tenta prever as reações dos leitores à presença de um elemento na história, utilizando técnicas de processamento de linguagem natural para determinar se as respostas são positivas ou negativas, e passando esses dados classificados para um algoritmo preditivo.

No que se relaciona aos artigos citados, este trabalho utiliza as técnicas de coleta de dados semelhantes às de [3], porém com um foco maior nas tags das fanfictions em vez dos personagens. Semelhante ao que foi feito no trabalho [4], o foco central deste estudo serão as tags do AO3, mas buscando trazer métricas da rede presente na plataforma e realizar análises em cima do conteúdo do site, ao invés de focar na experiência dos usuários. Além disso, serão aplicados métodos semelhantes aos de [5], porém direcionados às tags retiradas do Archive of Our Own, ao invés de utilizar o texto das histórias presentes no site FanFiction.

III. Metodologia

Esta seção descreve as principais etapas da metodologia apresentada neste trabalho. A seção III-A apresenta o conjunto de dados analisado e como ocorreu a coleta deste. A seção III-B traz a modelagem dos dados coletados no formato de grafos e a construção de visualizações dessas estruturas. Por fim, a seção III-C aborda o cálculo das métricas das redes, trazendo explicações sobre o papel de cada um dos cálculos realizados.

A. Coleta de dados

A coleta de dados disponibilizados no A03 foi realizada através de*web scraping* no site, utilizando *webdrivers* de Selenium. O tipo de coleta feito está de acordo com os termos de uso do Archive of Our Own.

O Selenium é um *framework Python*[3] que utiliza *webdrivers* para navegar em uma determinada interface e simular ações de um usuário. Essas ações permitem a reprodução automática e em larga escala de interações que um usuário final realizaria. Assim, usamos as funções disponibilizadas para simular a navegação de um usuário do Archive of Our Own, o que permitiu a realização de uma coleta em *snowball* dos dados das fanfictions disponíveis publicamente no site. Para facilitar a organização dos dados, cada idioma foi coletado separadamente. Os passos gerais da coleta de dados para a construção dos grafos estão definidos a seguir:

1) Acessar a página inicial das fanfictions do idioma selecionado;
2) De forma iterativa, navegar por esta página coletando os dados desejados de cada fanfiction. Esses dados são então armazenados em um arquivo csv;
3) Verificar se existe um botão de próxima página e clicar nele caso ele esteja presente;
4) Repetir os passos 2 e 3 até que não haja mais uma próxima página e todas as fanfictions daquele idioma tenham sido coletadas;
5) Repetir os passos 1-4 para o próximo idioma da lista até que todos os idiomas tenham sido coletados. Essa etapa foi realizada para todos os idiomas presentes no site, exceto aqueles que continham mais de 500.000 fanfictions.

Os metadados de cada fanfiction coletadas foram os seguintes: o ID da obra, título, autor, fandoms, classicação indicativa, avisos de conteúdo, tipo de relacionamento, status de conclusão, data da última atualização, relacionamentos,

[2]https://www.fanfiction.net/

[3]https://www.python.org/

personagens, tags de conteúdo, sinopse, idioma, número de palavras, número de capítulos, número de comentários, número de "kudos", número de "*bookmarks*" e número de acessos.

No final dessa etapa foram coletadas cerca de 270.000 fanfics, distribuídas em 129 idiomas. Desses 129 idiomas, 111 deles tinham 1.000 fanfictions ou menos, o que não permitiria a realização de análises significativas, e um deles possuia mais de 50.000 fanfictions, o que demandaria um poder computacional inviável para este trabalho. Por conta disso, nosso foco principal foi nos 18 idiomas que tinham entre 1.001 e 50.000 fanfics.

B. Modelo Matemático

Neste trabalho, utilizamos o modelo de grafos para representar as fanfictions e as relações entre elas. A estrutura do grafo foi definida de forma que cada nó na rede representa uma obra de fanfiction diferente. Para garantir que não existem nós duplicados, os nós são representados pelo ID de cada obra. Como o objetivo do trabalho é analisar a relação entre as tags de conteúdo e as obras, uma aresta entre dois nós existe caso as duas fanfictions possuem pelo menos uma tag em comum. As arestas tem peso igual ao total de tags que coocorrem em cada par de fanfiction.

Após a definição da estrutura, os grafos foram construídos utilizando a biblioteca *NetworkX*[4], que é um pacote *Python* para a criação, manipulação e estudo de estruturas, dinâmicas e funções de redes complexas. O número de nós e arestas de cada grafo criado estão apresentados na Tabela I.

Tabela I
NÚMERO DE NÓS E ARESTAS PARA CADA GRAFO CRIADO

Idioma	Nós	Arestas
ko	2.446	32.710
ja	2.994	61.125
sv	1.013	76.175
th	2.014	71.403
yue	9.290	76.886
ptPT	1.141	95.999
hu	2.111	129.863
fi	2.071	428.765
tr	3.131	483.377
cs	5.562	714.121
vi	8.155	1.549.669
fil	9.231	5.019.421
pl	16.672	7.351.857
uk	17.751	15.000.972
de	23.923	16.772.335
id	29.499	32.146.143
ptBR	34.972	32.927.493
it	34.989	36.809.452

O próximo passo foi a visualização dos grafos através da ferramenta *Gephi*[5]. Nessa etapa encontramos algumas limitações devido ao significativo de alguns dos grafos, o que não possibilitou que todas as redes tivessem visualizações plotadas. Dos 18 grafos criados neste trabalho, apenas 14 foram capazes de ter visualização. Temos um exemplo de visualização criada na imagem 2, onde cada cor representa um fandom diferente e a espessura das arestas é proporcional ao grau de cada uma delas. Podemos observar na imagem que alguns fandoms formam comunidades mais isoladas, constituíndo blocos formados por uma cor mais predominante, enquanto outros fandoms possuem mais conexões entre si, formando componentes com várias cores próximas umas das outras na visualização.

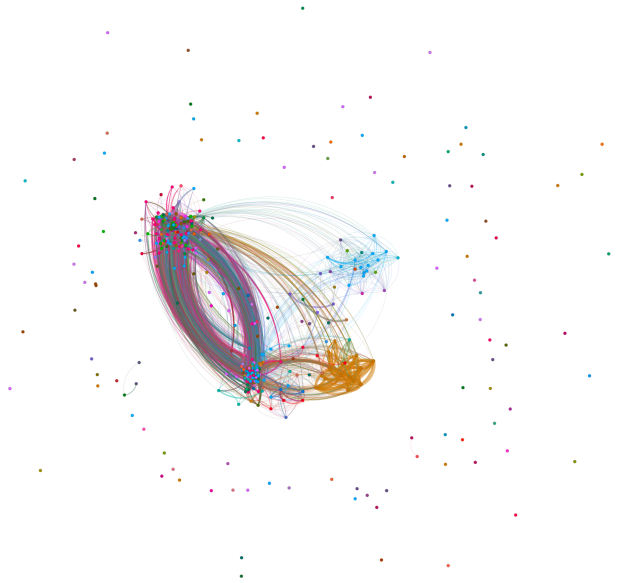


Figura 2. Grafo da rede de fanfictions em português de Portugal (ptPT).

C. Métricas das Redes

Com o objetivo de conhecer melhor os grafos antes da realização de análises mais aprofundadas, foram calculadas algumas métricas para todas as redes. Isso permite que tenhamos uma ideia melhor da estrutura dos grafos, principalmente aqueles que não possuem visualizações plotadas.

As métricas calculadas foram grau médio, densidade, assortatividade, componentes conectados, coeficiente de clusterização, diâmetro, modularidade e número de comunidades. O grau médio se refere à média aritmética dos graus de cada vértice da rede. Assortatividade diz respeito à tendência dos nós a se conectares a outros nós semelhantes no grafo, a chamada homofilia [6]. O número de componentes conectados de um grafo é a quantidade de subgrafos conectados que não fazem parte de nenhum subgrafo conectado maior [7]. O coeficiente de clusterização mede a tendência dos nós de um grafo de se agrupar, sendo que um valor mais elevado indica a presença de um grande número de triângulos na rede [7]. O diâmetro da rede é o maior número de ligações entre dois nós da rede. Modularidade mede a qualidade da divisão de uma rede em comunidades, ou seja, redes com modularidade alta possem muitas conexões entre os nós da mesma comunidade e

[4]https://networkx.org/
[5]https://gephi.org/

poucas conexões entre nós de comunidades diferentes [8]. Por fim, o número de comunidades é a quantidade de partições geradas pelo algoritmo de Louvain, sendo que só foram consideradas as partições com mais de um nó. Vale notar que o número de comunidades foi obtido usando o algoritmo de Louvain com a mesma resolução e o mesmo seed para todos os idiomas. Os resultados obtidos para todas essas métricas estão nas tabelas II e III.

Tabela II
GRAU MÉDIO, DENSIDADE, ASSORTATIVIDADE E NÚMERO DE COMPONENTES CONECTADOS PARA CADA GRAFO CRIADO

Idioma	**Grau Médio**	**Densidade**	**Assortatividade**	Componentes Conectados
ko	13,37	0,011	0,204	1289
ja	20,42	0,014	0,310	1286
sv	75,20	0,149	0,439	135
th	35,45	0,035	0,251	708
yue	8,28	0,002	0,423	7131
ptPT	84,14	0,148	0,277	106
hu	61,52	0,058	0,148	397
fi	207,03	0,200	0,405	246
tr	154,38	0,099	0,206	707
cs	128,39	0,046	0,146	855
vi	190,03	0,047	0,201	1541
fil	543,76	0,118	0,208	1017
pl	440,97	0,053	0,191	2374
uk	845,08	0,095	0,105	1483
de	701,10	0,059	0,173	3654
id	1089,74	0,074	0,282	2993
ptBR	941,54	0,054	0,134	4567
it	1052,03	0,060	0,118	4601

Tabela III
COEFICIENTE DE CLUSTERIZAÇÃO, DIÂMETRO, MODULARIDADE E NÚMERO DE COMUNIDADES PARA CADA GRAFO CRIADO

Idioma	Coeficiente de Clusterização	**Diâmetro**	**Modularidade**	Número de Comunidades
ko	0,310	8	0,439	31
ja	0,443	7	0,611	45
sv	0,625	7	0,223	13
th	0,451	6	0,417	16
yue	0,198	10	0,733	71
ptPT	0,580	6	0,343	9
hu	0,540	6	0,384	19
fi	0,686	5	0,233	13
tr	0,578	7	0,285	15
cs	0,586	7	0,383	18
vi	0,542	7	0,397	44
fil	0,605	5	0,409	24
pl	0,623	8	0,390	27
uk	0,564	6	0,314	14
de	0,536	7	0,361	42
id	0,620	8	0,384	38
ptBR	0,552	7	0,378	66
it	0,594	7	0,356	52

IV. ANÁLISES E RESULTADOS

Com os grafos construídos e informações acerca da estrutura de cada um deles, podemos começar a analisar o conteúdo das redes. Os resultados desta seção analisam o grafo de fanfictions em Português do Brasil.

A. *Nuvens de Palavras*

A primeira análise realizada foi a construção de nuvens de palavras do grafo e de suas comunidades. Para isso foram criados dois tipos de nuvens de palavras, um em cima das palavras presentes nas tags de conteúdo de cada fanfiction e outro baseado nas palavras das sinopses.

Nas nuvens de palavras de tags, as tags foram separadas em palavras, que por sua vez foram utilizadas como tokens. A peculiaridade desse tipo de nuvem é que a maioria das tags são em inglês, independente do idioma da rede. Isso possibilitaria a comparação de comunidades ou fandoms em redes de idiomas diferentes mais facilmente no futuro. Temos um exemplo de uma nuvem desse tipo na Figura 3, que é referente ao grafo inteiro de português brasileiro.

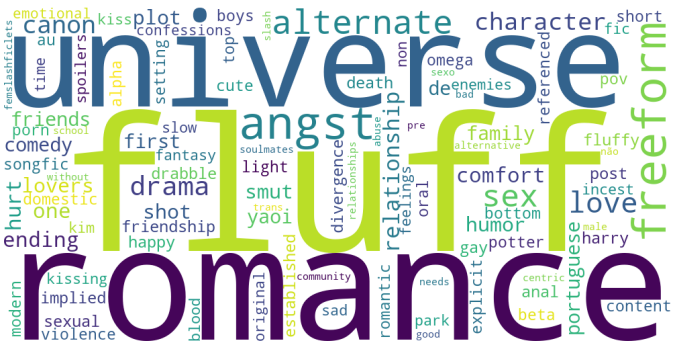


Figura 3. Nuvem de palavras baseada nas tags do grafo de português brasileiro.

Já nas nuvens de palavras de sinopses, as sinopses foram separadas em palavras, que por sua vez foram utilizadas como tokens. A maioria das palavras neste caso estão no idioma da própria rede, o que dificulta a análise comparativa entre redes. Temos um exemplo de uma nuvem desse tipo na imagem 4, que é referente ao grafo inteiro de português brasileiro.

Figura 4. Nuvem de palavras baseada nas sinopses do grafo de português brasileiro.

É interessante notar nas nuvens de palavras das comunidades que enquanto algumas comunidades aparentam ser mistas em relação aos fandoms que contém, outras parecem ser mais específicas. Nesses casos é possível observar também uma grande semelhança entre as nuvens de sinopse e de tags, enquanto nas comunidades mais gerais a relação entre as

nuvens diferentes é mais tênue. Podemos ver um exemplo dessa semelhança entre nuvens nas imagens 5 e 6, onde temos as duas nuvens de palavras da mesma comunidade.

Figura 5. Nuvem de palavras baseada nas sinopses de uma comunidade composta majoritariamente por um único fandom.

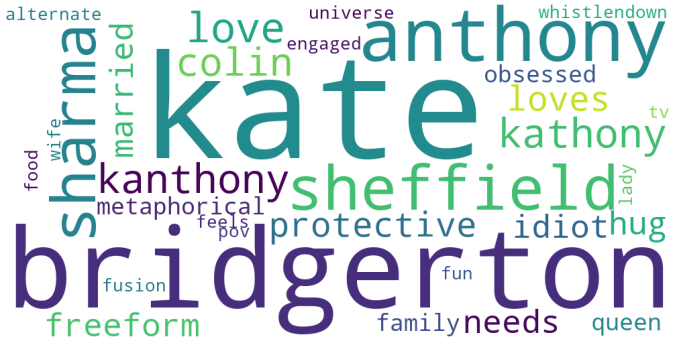


Figura 6. Nuvem de palavras baseada nas tags de uma comunidade composta majoritariamente por um único fandom.

Podemos contrastar as figuras 5 e 6 com as figuras 7 e 8, que são as redes de palavras de uma comunidade formada por vários fandoms diferentes. Essa comparação mostra que em comunidades compostas majoritariamente por um único fandom as palavras mais significativas costumam ser nomes de personagens daquele fandom, enquanto em comunidades mais amplas as palavras destaque tendem a ser referentes aos temas e conteúdos das histórias em si.

Figura 7. Nuvem de palavras baseada nas sinopses de uma comunidade composta por diversos fandoms.

Figura 8. Nuvem de palavras baseada nas tags de uma comunidade composta por diversos fandoms.

B. Análise de Sentimentos

Para tentar entender melhor o perfil de cada fandom, foi feita uma análise de sentimentos em cima dos textos das sinopses de cada fanfiction. Como não há nenhum modelo com fine-tuning para o caso de uso específico deste trabalho, foi utilizado o modelo *twitter-xlm-roberta-base-sentiment-finetunned*[6] por *citizenlab*, que é uma versão do *XLM-Roberta*[7] treinada para análise de sentimento de tuítes brasileiros. Esse modelo retorna três scores para cada fanfiction, referentes à probabilidade daquele texto ser positivo, neutro ou negativo. Para obter os scores dos fandoms foi calculada a média de cada probabilidade para todas as fanfics daquele fandom. A Tabela IV apresentam os resultados da análise de sentimentos.

Tabela IV
NÚMERO DE FANFICS E SCORES DE SENTIMENTO PARA OS 10 MAIORES FANDOMS

Fandom	Fanfics	Negativo	Neutro	Positivo
Harry Potter	2.488	0,137	0,834	0,028
EXO	2.282	0,113	0,862	0,024
Original Work	1.478	0,132	0,842	0,027
Bangtan Boys	1.270	0,106	0,864	0,029
Naruto	1.261	0,124	0,843	0,032
Boku no Hero Academia	994	0,135	0,831	0,034
Haikyuu!!	679	0,113	0,850	0,037
Stray Kids	565	0,120	0,848	0,032
Genshin Impact	548	0,118	0,857	0,025
Quackity SMP	445	0,151	0,825	0,024

Ao analisar os números da tabela, é possível observar que os valores obtidos são extremamente similares. Isso pode implicar que não há diferença significativa nos sentimentos apresentados nas sinopses em fandoms diferentes, ou pode ser devido a dificuldades inerentes à análise de sentimentos de textos em português. Esses desafios estão relacionados a fatores como a quantidade menor de bases de dados e métodos disponíveis em idiomas além do inglês [9], além da dificuldade em identificar sarcasmo e ironia nos textos em português [10]. Seria necessário executar novamente essa análise utilizando modelos diferentes antes de tirarmos conclusões sobre o que está por trás dessa similaridade dos resultados.

[6]huggingface.co/citizenlab/twitter-xlm-roberta-base-sentiment-finetunned
[7]huggingface.co/docs/transformers/model_doc/xlm-roberta

C. Análise de Similaridade

Buscamos analisar também a similaridade entre diferentes fandoms. Isso foi feito em duas tentativas diferentes, uma através da co-ocorrência entre fandoms e uma através da similaridade por cosseno da composição de tags.

Para o cálculo do número de co-ocorrência de fandoms, a ideia é que fandoms similares apareceriam juntos com maior frequência. Isso porque uma fanfiction pode ser associada a mais de um fandom, por exemplo uma fanfic pode envolver personagens da série *Doctor Who* interagindo com personagens do livro *Harry Potter*, levando esses dois fandoms a terem uma co-ocorrência entre si. Dessa maneira, fandoms mais similares apareciam juntos com maior frequência. Na figura 9 temos um heatmap com o número de co-ocorrências entre os 10 maiores fandoms da rede de português brasileiro.

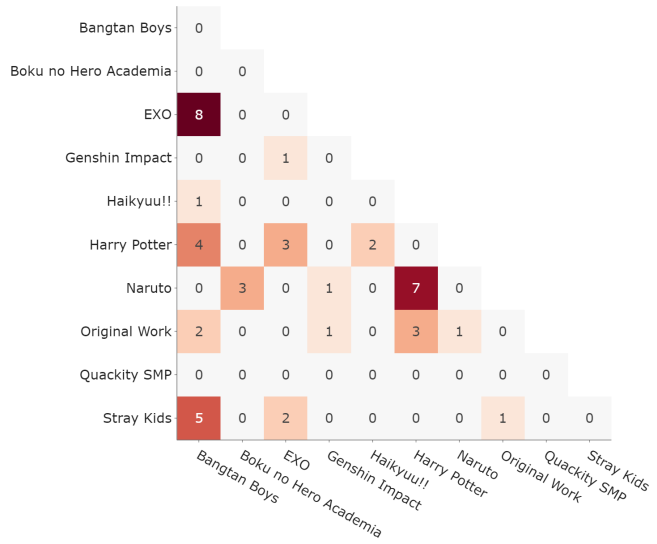


Figura 9. Heatmap com o número de co-ocorrências entre os 10 maiores fandoms.

Ao observar a imagem 9 podemos observar que dois dos maiores números de co-ocorrência foram obtidos entre *EXO* e *Bangtan Boys* e entre *Stray Kids* e *Bangtan Boys*, que são todos fandoms de bandas de K-Pop, indicando que fandoms de temas comuns aparecem juntos com alta frequência. Podemos observar também que *Harry Potter* possui co-ocorrências com 5 dos 9 outros fandoms analisados, o que pode implicar que fãs de *Harry Potter* também fazem parte de outros fandom. Isso pode ser contrastado com o fandom de *Quackity SMP*, que não possui co-ocorrências com nenhum dos fandoms analisados, podendo significar que os membros desse fandom constituem uma comunidade mais isolada.

A maior limitação do uso dessa métrica para quantificar a similaridade dos fandoms é que ela indica mais o número de fãs compartilhados do que a similaridade dos fandoms em si. Não podemos concluir utilizando apenas essa informação que os fandoms em questão são realmente similares, ou se eles aparecem juntos com frequência devido a outros fatores. Por conta disso, buscamos utilizar outra métrica para estimar similaridade, que foi a similaridade por cosseno.

A similaridade por cosseno é uma métrica usada para medir a similaridade entre dois vetores frequentemente utilizada em processamento de linguagem natural [11]. O primeiro passo para obter a similaridade por cosseno dos fandoms foi calcular a porcentagem de fanfics com cada tag para todos os fandoms. Na Tabela V temos as 10 tags mais comuns em um dos fandoms para ilustrar o resultado desta etapa.

Tabela V
10 TAGS MAIS COMUNS NO FANDOM "QSMP / QUACKITY SMP"

Tag	Porcentagem de Fanfics com a Tag
Angst	17,753
Fluff	14,157
guapoduo	11,685
Hurt/Comfort	8,315
Cat Hybrid Rafael Lange / Cellbit	7,191
Alternate Universe - Canon Divergence	6,966
Fluff and Angst	6,742
Hurt No Comfort	6,292
Domestic Fluff	6,067
How Do I Tag	5,618

Esses vetores foram então convertidos para um formato compatível com o cálculo da similaridade por cosseno e a métrica foi calculada para os 10 maiores fandoms, conforme demonstrado na Figura 10.

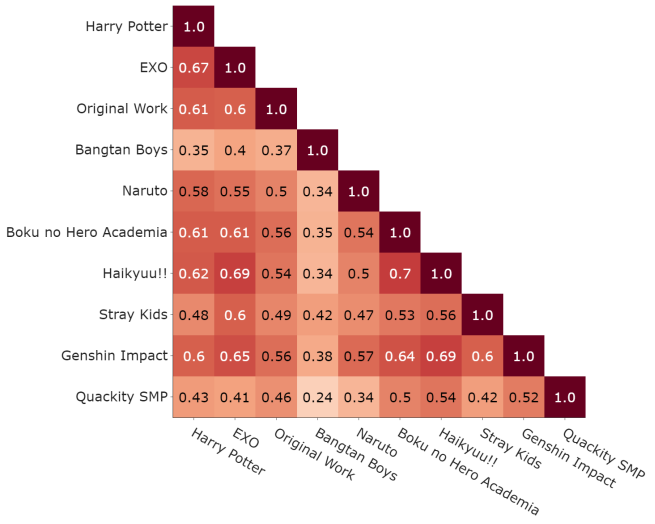


Figura 10. Heatmap com os valores de similaridade entre os 10 maiores fandoms.

Podemos observar no heatmap de similaridade que os valores de similaridade diferem bastante dos resultados encontrados no cálculo de co-ocorrências. É interessante notar que o fandom de *Bangtan Boys* apresenta similaridade baixa com todos os outros, enquanto os fandoms de *Haikyuu!!* e *EXO* possuem a maior similaridade entre os fandoms observados. Ao observar as tabelas V e VI, podemos concluir que isso é devido ao fato de que alguns fandoms utilizam um número maior de tags específicas ao seu próprio fandom, enquanto outros utilizam tags mais genéricas, que podem ser utilizadas por fandoms de diversos temas. Por exemplo, no fandom de *Quackity SMP* apenas 2 das 10 tags mais comuns são referentes à temas exclusivos do próprio fandom, enquanto no fandom de *Bangtan Boys* 5 das 10 tags mais comuns são específicas ao fandom em si.

Tabela VI
10 TAGS MAIS COMUNS NO FANDOM "BANGTAN BOYS"

Tag	Porcentagem de Fanfics com a Tag
2minpjct	26,772
2min - Freeform	20,866
Romance	8,661
Anal Sex	8,661
Top Jeon Jungkook	6,299
Bottom Park Jimin (BTS)	6,220
Fluff	5,591
Friends to Lovers	4,961
Angst	4,961
Bottom Min Yoongi — Suga	4,882

De forma geral, os resultados da similaridade por cosseno diferem bastante daqueles obtidos na análise de co-ocorrências, indicando que ainda que dois fandoms apareçam juntos com frequência, isso não significa que as tags que eles utilizam sejam semelhantes. Isso é exemplificado pelos fandoms de *EXO* e de *Bangtan Boys*, que apresentam um número alto de co-ocorrências, mas baixa similaridade, o que pode significar que eles possuem um número alto de fãs em comum, mas esses fãs costumam utilizar tags separadas para cada comunidade. Isso fica ainda mais evidente ao comparar as tags mais comuns de cada um desses fandoms (tabelas VI e VII), onde é possível ver que o fandom de *Bangtan Boys* costuma utilizar tags mais específicas, enquanto o fandom de *EXO* utiliza tags referentes a temas diferentes.

V. CONCLUSÕES E PRÓXIMOS PASSOS

Este trabalho apresentou uma análise detalhada das fanfictions encontradas no Archive of Our Own, através da construção de grafos baseados nas tags associadas a cada fanfic. Através da coleta e da análise dos dados, conseguimos caracterizar as redes de fanfictions em diversos idiomas, focando principalmente em identificar padrões e características das fanfics e fandoms da rede de português brasileiro.

Os resultados das análises preliminares mostraram uma variação significativa nas métricas de rede, como grau médio, coeficiente de clusterização e modularidade, entre os diferentes idiomas. Essas variações indicam diferenças estruturais e de comportamento entre as comunidades de fãs de diferentes regiões e culturas, o que seria uma oportunidade interessante para aprofundar o trabalho no futuro.

Tabela VII
10 TAGS MAIS COMUNS NO FANDOM "EXO"

Tag	Porcentagem de Fanfics com a Tag
Romance	18,405
Fluff	12,401
EXO - Freeform	8,808
Drama	8,063
Friends to Lovers	7,318
Smut	6,617
One Shot	6,617
Light Angst	6,179
Alternate Universe	6,004
exozone	6,004

A construção das nuvens de palavras também revelou tendências interessantes nas fanfics, destacando temas e elementos recorrentes. Essas nuvens fornecem uma visão mais clara das preferências e tendências dentro das comunidades e dos fandoms, permitindo uma melhor compreensão da dinâmica de criação de conteúdo.

Na análise de sentimentos tivemos resultados menos concretos, pois os números aparentam indicar que não há diferenças significativas entre os sentimentos das sinopses de fandoms diferentes, porém como o modelo utilizado não é próprio para o caso de uso deste estudo não podemos descartar a possibilidade de que o resultado seria diferente caso outro modelo tivesse sido utilizado. Seria interessante rodar novamente essa análise, mas utilizando modelos diferentes para ver se os haveriam alterações nos resultados obtidos.

A análise de similaridade foi capaz de evidenciar fandoms que compartilham tags similares, e com isso apresentam um número maior de temas em comum. Também foi possível perceber que alguns fandoms utilizam tags específicas ao próprio fandom, enquanto outros utilizam tags que se referem a temas mais amplos. Essa abordagem pode nos ajudar a entender como temas e personagens se repetem em fandoms diferentes, sendo uma grande oportunidade de aprofundamento em trabalhos futuros.

Uma das maiores limitações deste estudo que será muito importante de ser trabalhada nos próximos passos é a relação entre as tags no Archive of Our Own. Para este trabalho, se duas tags são escritas de formas diferentes elas são tags separadas, mesmo se elas tiverem o mesmo significado. Já no AO3, as tags são agrupadas, com tags de mesmo significado redirecionando para a tag principal para fins de busca e filtragem, conforme pode ser visto na imagem 11. Creio que o próximo passo para este trabalho será encontrar uma forma de unificar e relacionar as tags coletadas, para com isso trazer conclusões mais definitivas acerca da similaridade entre os fandoms, entre outras análises.

Outra oportunidade muito interessante de aprofundar as análises desenvolvidas até o momento é analisar as interações dos fãs e dos leitores com o conteúdo produzido. As análises

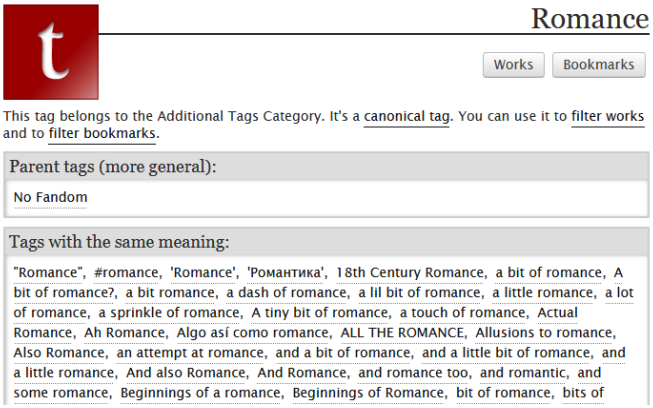


Figura 11. Exemplo de tags com mesmo significado que a tag “Romance”.

trazidas neste trabalho foram focadas no conteúdo produzido pelos fãs, seria relevante tentar entender mais sobre o conteúdo consumido pelos leitores, observando, por exemplo, quais tags possuem fanfics com um número maior de visualizações ou de *bookmarks*.

Em geral, este estudo estabeleceu uma base sólida para a compreensão das fanfictions do Archive of Our Own, mas o tema ainda apresenta muito potencial a ser explorado. Os próximos passos irão focar em expandir e aprofundar as análises realizadas até o momento, buscando oferecer uma visão mais detalhada e abrangente das dinâmicas encontradas nessa comunidade.

Referências

[1] ALENCAR, Daniele Alves ; ARRUDA, Maria Izabel Moreira. Fanfiction: uma escrita criativa na web. Perspectivas em Ciência da Informação, v. 22, n. 2, p. 88–103, 2017. Disponível em: https://www.scielo.br/j/pci/a/bfprW4R7gL9JL7swCZ7hWxs/. Acesso em: 30 mar. 2024

[2] BARBACHAN, Andressa Nadvorny. Levando o fã a sério: Estudos de fandom aplicados às relações públicas. 2016. Disponível em: http://hdl.handle.net/10183/147054. Acesso em: 30 jul. 2024

[3] SCHMIDT, Thomas ; HOFFMANN, Johannes ; WOLFF, Christian. Analyzing character networks in crossover fan fictions of archive of our own, 2022. Disponível em: https://wp.unil.ch/llist/files/2022/06/COMHUM_2022_paper_14.pdf. Acesso em: 01 abr. 2024.

[4] BLACK, Erin. “There’s A Tag for That”: An Exploratory Study of Tag Functions in the Archive of Our Own, 2021. Disponível em: https://cdr.lib.unc.edu/concern/masters_papers/m326mb56w. Acesso em: 01 abr. 2024.

[5] MILLI, Smitha ; BAMMAN, David. 2016. Beyond Canonical Texts: A Computational Analysis of Fanfiction. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, Texas, p. 2048–2053, 2016. Disponível em: https://aclanthology.org/D16-1218/. Acesso em: 02 abr. 2024.

[6] NOLDUS, Rogier ; VAN MIEGHEM, Piet. Assortativity in complex networks. Journal of Complex Networks, v. 3, n. 4, p. 507-542, 2015. Disponível em: https://doi.org/10.1093/comnet/cnv005. Acesso em: 25 jul. 2024.

[7] METZ, Jean et al. Redes Complexas: conceitos e aplicações. 2007. Disponível em: http://repositorio.icmc.usp.br//handle/RIICMC/6720. Acesso em: 25 jul. 2024.

[8] NEWMAN, Mark EJ. Modularity and community structure in networks. Proceedings of the national academy of sciences, v. 103, n. 23, p. 8577-8582, 2006. Disponível em: https://doi.org/10.1073/pnas.0601602103. Acesso em: 25 jul. 2024.

[9] KANSAON, Daniel P. ; BRANDÃO, Michele A. ; DE PAULA PINTO, Saulo A. Análise de Sentimentos em Tweets em Português Brasileiro. Anais do VII Brazilian Workshop on Social Network Analysis and Mining. SBC, 2018. Disponível em: https://sol.sbc.org.br/index.php/brasnam/article/view/3578. Acesso em: 24 jul. 2024.

[10] SILVA, Priscilla de Souza. Avaliação do desempenho de métodos de análise de sentimentos na presença das figuras de linguagem sarcasmo e ironia. Trabalho de Conclusão de Curso (Graduação) - Universidade Federal do Sul e Sudeste do Pará, Campus Universitário de Marabá, Instituto de Geociências e Engenharias, Faculdade de Computação e Engenharia Elétrica, Curso de Bacharelado em Sistemas de Informação, Marabá, 2016. Disponível em: http://repositorio.unifesspa.edu.br/handle/123456789/233. Acesso em: 24 jul. 2024.

[11] LEYDESDORFF, Loet. Similarity measures, author cocitation analysis, and information theory. Journal of the American society for Information Science and Technology, v. 56, n. 7, p. 769-772, 2005. Disponível em: https://doi.org/10.1002/asi.20130. Acesso em: 27 jul. 2024.