UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

JOAQUIM ALVINO DE MESQUITA NETO

Uma análise sobre o comportamento tóxico em jogos on-line baseada em tópicos de conversa

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre em Ciência da Computação

Orientador: Profª. Drª. Karin Becker

Porto Alegre
2019

CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Alvino de Mesquita Neto, Joaquim

Uma análise sobre o comportamento tóxico em jogos on-line baseada em tópicos de conversa / Joaquim Alvino de Mesquita Neto. – Porto Alegre: PPGC da UFRGS, 2019.

134 f.: il.

Dissertação (mestrado) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2019. Orientador: Karin Becker.

I. Becker, Karin. II. Título.

AGRADECIMENTOS

Primeiramente eu agradeço tanto ao meu pai, que sempre tomou a minha educação e de minhas irmãs como prioridade máxima, me apoiou em todos os passos da minha vida acadêmica e nunca me deixou faltar nada, como minha mãe que sempre foi muito atenciosa com a família e nunca deixou de dar suporte emocional nos momentos mais necessários, mesmo como milhares de quilômetros nos separando.

Também gostaria de expressar minha gratidão aos meus amigos, que me propiciaram afeição e momentos de alegria, me ajudando a passar por estes dois anos tão únicos na minha vida. Agradeço ao Matheus Gonzaga e sua família, que me auxiliaram bastante quando eu era só um rapaz perdido em Porto Alegre, amenizando o choque de uma mudança de vida tão súbita. Agradeço também ao Kazuki Yokoyama, por ter me introduzido e ajudado com o tópico de estatísticas, que me era tão estranho até antes dessa dissertação.

Gostaria também de agradecer aos meus amigos e família que eu deixei em Fortaleza, que nunca deixaram de manter contato comigo, e que sempre me recebem de braços abertos toda vez que eu preciso deles.

Gostaria de agradecer minha orientadora Karin Becker, pelo seu suporte, sabedoria e todos os ensinamentos postos nestes dois últimos anos e aos meus colegas do Laboratório 213, pelo suporte e pela paciência.

Finalmente, agradeço ao CNPq, ao Instituto de Informática e a Universidade Federal do Rio Grande do Sul, pelas verbas e infraestrutura necessárias, sem as quais esta pesquisa nunca se tornaria realidade.

RESUMO

Multiplayer Online Battle Arena (MOBA) são jogos competitivos, nos quais a vitória depende do trabalho em equipe entre os jogadores. O comportamento tóxico atrapalha a comunicação entre jogadores e diminui a coesão de uma equipe, provendo um ambiente de jogo pior aos envolvidos. Trabalhos na área focam na detecção automática e na caracterização do comportamento tóxico, através de *features* textuais envolvendo a comunicação entre jogadores. Nós investigamos os padrões de conversa utilizados por jogadores de *League of Legends*, um jogo MOBA popular, e investigamos os efeitos destes padrões sobre o desempenho e a contaminação tóxica destes jogadores, quais as transições mais prováveis entre estes padrões, bem como caracterizamos tais padrões de acordo com os principais sentimentos evocados por estes.

Neste trabalho, buscamos dissecar o comportamento de jogadores em partidas de MOBAs, identificando e validando os tópicos de conversa utilizados por jogadores nestas partidas, tópicos estes que correspondem a diferentes comportamentos adotados por jogadores. Através dos tópicos, nós: a) caracterizamos o comportamento de grupos de jogadores; b) analisamos como os tópicos afetam o desempenho e a contaminação de grupos de jogadores, através de métricas criadas para tal fim; c) descobrimos tendências de como a conversação flui durante uma partida; e d) analisamos como diferentes padrões de conversa associam-se com emoções, através da construção de um léxico de sentimentos voltado a conversas em MOBAs.

Descobrimos que os aliados de um jogador tóxico são, em geral, mais afetados pelo comportamento tóxico do que seus adversários e, que oponentes são mais afetados quando o comportamento tóxico é diretamente direcionado a eles (por exemplo, insultos racistas). Jogadores sem contato significativo com jogadores tóxicos tendem a ser mais positivos, concentrando-se em táticas de jogo e socialização. Também descobrimos que comportamento negativo apresentado por jogadores não-ofensores aparenta ser transitório, podendo voltar a normalidade com relativa facilidade, enquanto jogadores tóxicos recusam-se a colaborar com seu time após algum conflito com este, e que a falta de confiança entre membros de uma equipe, bem como sentimentos de medo, podem servir como estopim para o comportamento tóxico.

Nossos resultados podem servir de porta de entrada para trabalhos mais complexos sobre o estado emocional de jogadores, além poderem ser explorados para um melhor entendimento do comportamento tóxico, bem como para sua detecção através de meios automatizados, até mesmo buscando prevenir tal comportamento durante uma partida.

Palavras-chave: .

A deep analysis of toxic behavior and other kinds of behaviors of MOBA players.

ABSTRACT

Multiplayer Online Battle Arena (MOBA) are competitive games, in which victory depends on efficient teamwork between players. Toxic behavior disrupts communication between players and decreases the cohesion of a team, providing a game ambiance that is worse for those involved. Works in the area focus on the automatic detection of toxic behavior, through textual features involving communication between players. We investigated the patterns of conversation used by players from League of Legends, a popular MOBA game, and investigated the effects of these patterns on the performance and toxic contamination of these players, which are the most likely transitions between these patterns, as well as characterizing these patterns according to the main sentiments evoked by them.

In this work, we try to dissect the behavior of players in MOBAs matches, identifying and validating the topics of conversation used in those matches through in-game chat, topics that correspond to different kinds of behavior adopted by players. Through these topics, we: a) characterize the behavior of groups of players; b) analyze how these topics affect the performance and contamination of groups of players, through metrics created for this purpose; c) discover trends in how conversations flow during a match; and d) analyze how different patterns of conversation are associated with different emotions by building a specific lexicon of emotions for the conversations between players in MOBAs.

We have found that allies of a toxic player are generally more affected by toxic behavior than their opponents, and that opponents are most affected when toxic behavior is targeted at them (e.g. racist insults). Players without significant contact with toxic players tend to be more positive, focusing on game tactics and socialization. We also found that negative behavior in non-offender players appears to be transient, as it can return to normalcy with relative ease, while toxic players refuse to cooperate with their team after some conflict, and that lack of trust between teammates, as well as fear emotions, can serve as a trigger for toxic behavior.

Our results can serve as a gateway to more complex work on the emotional state of players, and can be exploited to better understand toxic behavior as well as to detect it through automated means, even to prevent such behavior during a game .

Keywords: Toxic Behavior, Text Mining, Sentiment Mining, Online Games, League of Legends.

LISTA DE FIGURAS

Figura 2.1 Mapa e bate-papo de *League of Legends*. 17

Figura 5.1 Processo de agrupamento de texto adotado. 40
Figura 5.2 Histograma da duração média das partidas (intervalos de 10 minutos). 49

Figura 6.1 Relação entre desempenho/contaminação e tópicos positivos. 67
Figura 6.2 Relação entre desempenho/contaminação e tópicos relacionados a tática. 68
Figura 6.3 Relação entre desempenho/contaminação e tópicos relacionados ao humor..... 69
Figura 6.4 Relação entre desempenho/contaminação e tópicos negativos. 71
Figura 6.5 Relação entre desempenho/contaminação e tópicos de reclamações 72
Figura 6.6 Relação entre desempenho/contaminação e tópicos de discussões. 74
Figura 6.7 Relação entre desempenho/contaminação e tópicos de insultos. 75
Figura 6.8 Relação entre desempenho/contaminação e tópicos de provocações. 76
Figura 6.9 Relação entre tópicos negativos dos ofensores e as métricas do não-ofensores. 80
Figura 6.10 Relação entre tópicos de reclamações dos ofensores e métricas dos não-ofensores. 81
Figura 6.11 Relação entre tópicos de discussões dos ofensores e métricas dos não-ofensores.83
Figura 6.12 Relação entre tópicos de insultos dos ofensores e métricas dos não-ofensores. .. 84
Figura 6.13 Relação entre tópicos de provocação dos ofensores e métricas dos não ofensores.85
Figura 6.14 Regras de transição entre tópicos, sem repetições - Aliados. 91
Figura 6.15 Regras de transição entre tópicos, sem repetições - Inimigos. 92
Figura 6.16 Transições entre tópicos, sem repetições - Ofensores. 93
Figura 6.17 Distribuição dos valores de emoção para cada tópico. 102
Figura 6.18 Comparação entre os valores de emoção totais para cada tópico. 102
Figura 6.19 Comparação entre os valores de emoções totais para cada emoção. 103
Figura 6.20 Comparação entre os tópicos mais e menos presentes para cada emoção..... 104
Figura 6.21 Comparação entre os tópicos mais e menos presentes na emoção alegria. 105

Figura A.1 Nuvem das 100 palavras mais relevantes do agrupamento 1..... 115

Figura E.1 Nuvem de palavras para as palavras apresentando a emoção de alegria..... 127
Figura E.2 Nuvem de palavras para as palavras apresentando a emoção de antecipação..... 128
Figura E.3 Nuvem de palavras para as palavras apresentando a emoção de confiança. 129
Figura E.4 Nuvem de palavras para as palavras apresentando a emoção de medo..... 130
Figura E.5 Nuvem de palavras para as palavras apresentando a emoção de nojo. 131
Figura E.6 Nuvem de palavras para as palavras apresentando a emoção de raiva..... 132
Figura E.7 Nuvem de palavras para as palavras apresentando a emoção de surpresa. 133
Figura E.8 Nuvem de palavras para as palavras apresentando a emoção de tristeza. 134

LISTA DE TABELAS

Tabela 3.1 Tabela ou Matriz de Confusão 29

Tabela 5.1 Resumo das Abordagens Propostas 37
Tabela 5.2 Avaliação das métricas de desempenho. 46
Tabela 5.3 Proporção de palavras similares entre os tópicos da Seção 5.2.1, e os tópicos descobertos nesta seção 49
Tabela 5.4 Exemplos das divisões de partidas por tempo para 3 partidas e 9 grupos. 50
Tabela 5.5 Palavras removidas do NRC 53
Tabela 5.6 Distribuição das emoções encontradas no léxico voltado a MOBAs 55
Tabela 5.7 Emoções para as 10 palavras mais relevantes no tópico reclamações 56

Tabela 6.1 Tópicos e 10 palavras mais relevantes 57
Tabela 6.2 Distribuição dos tópicos nos grupos 61
Tabela 6.3 Mèdia e Desvio padrão (SD) dos grupos para desempenho e contaminação 61
Tabela 6.4 Comparação das performances médias de grupos com/sem a prevalência de um tópico 64
Tabela 6.5 Correlações (τ de Kendall) entre performance e concentração de tópicos 64
Tabela 6.6 Correlações (τ de Kendall) entre a taxa de uso de tópicos negativos pelo ofensor e contaminação/desempenho de não-ofensores 79
Tabela 6.7 Regras de auto-transição para todos os grupos. 88
Tabela 6.8 Demais regras para Aliados, Inimigos e Ofensores 89
Tabela 6.9 Resultados dos modelos MLP/SVM/RL, para cada classe de sentimento e emoção 96
Tabela 6.10 Micro F-Measure dos modelos construídos 96
Tabela 6.11 Frequências de sentimentos nos léxicos 98
Tabela 6.12 Distribuição da quantidade de emoções em palavras de sentimento nos léxicos . 98
Tabela 6.13 Palavras de sentimento do top500 de cada tópico para cada léxico 99
Tabela 6.14 Resultados da rotulação manual de palavras de sentimento 100
Tabela 6.15 Valores das emoções para cada tópico 101

Tabela A.1 Resumo da interpretação dos Tópicos 117

Tabela C.1 Resultados para execuções do em transações com repetições 121
Tabela C.2 Resultados para experimentos em transações sem repetições 122

Tabela D.1 Micro F-Measure dos testes preliminares 123
Tabela D.2 Resultados dos modelos de classificação para vetores de 100 dimensões 124
Tabela D.3 Resultados dos modelos de classificação para vetores de 200 dimensões 125
Tabela D.4 Resultados dos modelos de classificação para vetores de 300 dimensões 126

LISTA DE ABREVIATURAS E SIGLAS

ADC	Attack Damage Carry
API	Application Programming Interface
CBOW	Continuous Bag of Words
DotA	Defense of the Ancients
ENN	Edited Nearest Neighbors
GloVe	Global Vectors
IQR	Inter-Quartile Range
KDA	Kills-Deaths-Assists
LDA	Latent Dirichlet Allocation
LHS	Left Hand Side
LoL	League of Legends
MOBA	Multiplayer Online Battle Arena
MLP	Multi-Layer Perceptron
NLTK	Natural Language Toolkit
NRC	National Research Council (Canada)
PC	Personal Computer
ReLU	Rectified Linear Unit
RBF	Radial Basis Function
RHS	Right Hand Side
RPG	Role Playing Game
RTS	Real-Time Strategy
SD	Standard Deviation
SMOTE	Synthetic Minority Oversampling TEchnique
SVM	Support Vector Machines
TF-IDF	Term Frequency – Inverse Document Frequency

SUMÁRIO

1 INTRODUÇÃO 11
2 MOBA 15
2.1 MOBAs 15
2.2 *League of Legends* 16
2.3 Dados do Tribunal 17
3 FUNDAMENTAÇÃO TEÓRICA 20
3.1 Representação vetorial de documentos e palavras 20
3.2 Agrupamento de dados textuais 22
3.3 Regras de associação 23
3.4 Análise de sentimentos 24
3.5 Construção automática léxicos de emoções 25
3.6 Algoritmos de Classificação 26
3.6.1 Métricas de Avaliação 28
4 TRABALHOS RELACIONADOS 30
4.1 Definições de comportamento tóxico 30
4.2 Caracterização do comportamento tóxico 31
4.3 Caracterização textual do comportamento de jogadores em MOBAs 34
4.4 Considerações finais 35
5 ANÁLISE DE COMPORTAMENTO TÓXICO E SEUS EFEITOS 36
5.1 Visão Geral 36
5.2 Tópicos de conversa em MOBAs 39
5.2.1 Descoberta de padrões de conversa 40
5.2.2 Interpretação dos tópicos 42
5.2.3 Análise dos jogadores por tópicos 43
5.3 Análise dos Efeitos do Comportamento Tóxico 43
5.3.1 Métricas de Desempenho e Contaminação 44
5.3.2 Análise dos efeitos de tópicos positivos e negativos 46
5.3.3 Análise dos efeitos do comportamento tóxico sobre os demais jogadores 47
5.4 Transições comuns de tópicos ao longo de partidas 47
5.5 Análise das emoções presentes nos tópicos 52
5.5.1 Construção do léxico de emoções 52
5.5.2 Caracterização das emoções de cada tópico 55
6 RESULTADOS 57
6.1 Tópicos de conversação entre jogadores 57
6.1.1 Tópicos Positivos 58
6.1.2 Tópicos Negativos 59
6.2 Tópicos e grupos de jogadores 61
6.3 Efeitos de Tópicos sobre Grupos de Jogadores 65
6.3.1 Efeitos de Tópicos Positivos 65
6.3.2 Efeitos de tópicos negativos 70
6.3.3 Efeitos dos tópicos negativos do ofensor sobre grupos não-ofensores 79
6.4 Relações temporais entre tópicos de conversação 87
6.4.1 Experimentos 1: Todas as transações 88
6.4.2 Experimentos 2: Transações com tópicos distintos 90
6.5 Tópicos e emoções 95
6.5.1 Resultados do modelo de classificação de sentimentos 95
6.5.2 Análise Quantitativa Comparando o NRC e o Léxico de MOBAs 97
6.5.3 Análise Subjetiva Preliminar do Léxico de MOBAs 100

6.5.4 Atribuição de emoções aos tópicos 101
7 CONCLUSÃO E TRABALHOS FUTUROS 107
REFERÊNCIAS 110
APPENDICES 114
APÊNDICEA INSTRUÇÕES DE INTERPRETAÇÃO DE AGRUPAMENTOS 115
A.1 Interpretação dos tópicos 117
APÊNDICEB ESTRUTURA DOS ARQUIVOS DO DATASET 118
APÊNDICEC RESULTADOS DOS EXPERIMENTOS COM PARÂMETROS DO APRIORI 120
C.1 Resultados para experimentos em transações com repetições 121
C.2 Resultados para execuções do apriori em transações sem repetições 122
APÊNDICED RESULTADOS DOS EXPERIMENTOS COM PARÂMETROS DO APRIORI 123
D.1 Determinando o Modelo de classificação a ser utilizado 123
D.1.0.1 Para 100 dimensões 124
D.1.0.2 Para 200 dimensões 125
D.1.0.3 Para 300 dimensões 126
APÊNDICEE PALAVRAS UTILIZADAS PARA A VALIDAÇÃO DO LÉXICO DE EMOÇÕES PARA MOBAS 127
E.1 Alegria 127
E.2 Antecipação 128
E.3 Confiança 129
E.4 Medo 130
E.5 Nojo 131
E.6 Raiva 132
E.7 Surpresa 133
E.8 Tristeza 134

1 INTRODUÇÃO

Jogos online são o passatempo favorito de muitas pessoas. Jogadores ao redor do mundo todo geraram mais de 99 bilhões de dólares em 2016, de acordo com o *Global Games Market Report* [1] e a expectativa é que o mercado cresça para quase 120 bilhões de dólares em 2019.

O gênero de jogos MOBA (Multiplayer Online Battle Area) é um dos responsáveis por este crescimento. Os representantes mais populares deste gênero são *League of Legends*, *DotA 2*, *Smite* e *Heroes of the Storm*, os quais estiveram entre os 20 jogos de PC mais jogados em 2015, representando 30% do tempo total de jogo para jogos de PC neste período[2]. Todos estes jogos também organizam torneios regulares de e-esportes (esportes eletrônicos) que atraem milhares de espectadores, e movimentam grande somas de dinheiro em premiação anualmente.

MOBAs são jogos altamente competitivos, e a vitória depende fortemente de trabalho eficiente em equipe. Jogadores devem se comunicar constantemente para estabelecerem estratégias e decidirem como lidar com obstáculos e jogadores inimigos. Todos os jogos do gênero proveem algum tipo de canal de comunicação, através do qual os jogadores podem interagir via texto, voz e/ou sinais pré-estabelecidos. Diferentes tipos de interações entre jogadores podem emergir dessa comunicação, mas, infelizmente, nem todas elas são saudáveis.

Comportamento tóxico, às vezes referido como desinibição tóxica (SULER, 2004), *griefing* (FOO; KOIVISTO, 2004; LIN; CHUEN-TSAI, 2005) ou trollagem (HARDAKER, 2010), ocorre quando um jogador quebra regras de convivência, agindo de maneira ofensiva. Comportamento tóxico é uma presença constante em jogos online. Um estudo sobre trollagem em jogos online revelou que aproximadamente 80% dos jogadores pesquisados já foram vitimas, ou presenciaram cenas de comportamento tóxico (THACKER; GRIFFITHS, 2012). O anonimato provido pela internet é parcialmente responsável pela prevalência de tal comportamento, já que este permite a dissociação entre o 'eu' real, e um 'eu' digital, que não compartilham responsabilidade pelos seus respectivos atos (SULER, 2004).

Insultos, provocações e culpabilizações são a forma mais clássica de comportamento tóxico, mas outros comportamentos também são considerados ofensivos por jogadores online, como perder uma partida de propósito ou abandonar uma uma partida repentinamente. Como resultado do comportamento tóxico, o humor dos jogadores afetados deteriora, afetando negativamente a experiência de jogo. Combater o comportamento tóxico é importante para reter novos jogadores e para zelar pela reputação do jogo (SHORES et al., 2014).

Comportamento tóxico em jogos já foi estudado no contexto de jogos de interpretação de

[1] https://newzoo.com/solutions/revenues-projections/global-games-market-report/
[2] https://goo.gl/aXLAVh

papéis online (RPGs[3]), e em MOBAs. Alguns estudos analisaram os tipos de jogadores tóxicos no contexto de RPGs e suas motivações (FOO; KOIVISTO, 2004; LIN; CHUEN-TSAI, 2005; BARNETT; COULSON; FOREMAN, 2010), contudo não é simples aplicar esses estudos no contexto de MOBAs, devido à diferenças na mecânica de jogo dos dois gêneros.

Já no contexto de MOBAs, os estudos se focaram em caracterizar e prever comportamento tóxico, usando a comunicação entre jogadores como base. Um previsor de comportamento tóxico usando dados do *League of Legends* (LoL) foi construído, usando, entre outras, *features* textuais extraídas de conversas em partidas (BLACKBURN; KWAK, 2014). Um trabalho complementar compara os vocabulários usados por jogadores tóxicos e 'normais', demonstrando que discrepâncias entre os vocabulários destes jogadores aparecem durante algum ponto da partida (KWAK; BLACKBURN, 2014). Vários efeitos do comportamento tóxico sobre o jogo foram estudados também em LoL (SHORES et al., 2014). Vocabulário dos jogadores e previsão de comportamento tóxico também foram estudados utilizando dados do jogo DotA (MARTENS et al., 2015).

Contudo, estudos existentes que analisam o vocabulário e os efeitos do comportamento tóxico são superficiais quanto à caracterização do comportamento dos demais jogadores, e da influência do jogador tóxico sobre estes, não indo além da dicotomia tóxico/não-tóxico. Uma análise mais aprofundada sobre as consequências do comportamento tóxico, e de como os jogadores comportam-se no geral é necessária para o desenvolvimento de técnicas mais eficientes no combate ao comportamento tóxico, sem prejudicar jogadores não-tóxicos, nem ser demasiado restritivo.

Este trabalho busca descobrir diferentes padrões de conversa usados por jogadores de *League of Legends*, e estudar suas características. Tais padrões são de grande valor para desenvolvedores e gerentes de comunidades de jogos, já que eles permitem uma melhor compreensão do que ocorre em seus jogos, e abrem portas para medidas que visam diminuir comportamentos indesejados.

Para descobrir tais padrões, usamos dados extraídos do tribunal de LoL, um site onde jogadores tóxicos eram julgados por membros da comunidade. A partir desses dados, descobrimos 7 padrões de conversa distintos, usados comumente por jogadores em partida de LoL, e realizamos um estudo aprofundado em cada um deles, respondendo as seguintes questões:

1. Quais tópicos de conversa são comumente utilizados por jogadores em partidas de MOBAs?
2. Como cada um dos tópicos descobertos se associa com diferentes tipos de jogadores,

[3]Do inglês *Role-Playing Games*

divididos em grupos de acordo com sua associação ao jogador tóxico (jogadores tóxicos, seus aliados, e seus inimigos)?

3. Como estes tópicos, para cada um dos grupos citados anteriormente, se relacionam com o desempenho e a contaminação tóxica, que definimos como sendo os efeitos negativos do comportamento tóxico?
4. Existem relações temporais entre estes tópicos, ou seja, regras que determinem como as conversas em uma partida se desdobram ao longo do tempo?
5. Como cada um dos tópicos de conversa descobertos associa-se com emoções derivadas de um modelo de emoções?

Para responder essas perguntas utilizamos técnicas de extração de tópicos em texto, e propusemos métricas específicas para avaliar o desempenho e a contaminação tóxica. Também aplicamos regras de associação entre os tópicos ao longo de uma partida, com o objetivo de descobrir associações temporais entre eles. Finalmente, construímos um dicionário de emoções específico para o domínio de MOBA. Resultados preliminares foram publicados em (NETO; YOKOYAMA; BECKER, 2017) e (NETO; BECKER, 2018).

Ao buscar respostas para estas perguntas, descobrimos que:

- jogadores tóxicos tendem a usar com maior frequência tópicos de conversa considerados negativos;
- aliados de jogadores tóxico (i.e. jogadores do mesmo time do jogador tóxico) são mais afetados pelo seu comportamento negativo do que os seus inimigos;
- a maneira como um grupo de jogadores se comunica durante uma partida se relaciona diretamente com o seu nível de desempenho e contaminação tóxica;
- existe uma transição gradual entre tópicos de conversa considerados saudáveis, e tópicos considerados tóxicos;
- emoções como nojo e raiva são associadas tópicos de conversa considerados mais tóxicos e expressam conflito dentro de uma equipe.

Este trabalho apresenta as seguintes contribuições para a análise de comportamento de jogadores em MOBAs:

- uma descrição de 7 tópicos de conversação distintos usados em partidas de MOBAs, interpretados e validados por um grupo de jogadores experientes;
- uma proposta de métricas para a mensuração de desempenho e de contaminação tóxica em partidas de MOBAs;

- uma análise da relação entre desempenho/contaminação tóxica e tópicos de conversação para cada tipo de jogador;
- uma análise de como os tópicos descobertos relacionam-se entre si temporalmente.
- um dicionário de emoções específico para o domínio de conversas em MOBAs, construído de modo automático, usando o dicionário NRC (MOHAMMAD; TURNEY, 2013) como base;
- a caracterização dos tópicos descobertos a partir das 8 emoções de Plutchik, utilizando o dicionário citado acima.

O restante este trabalho está estruturado como se segue. O Capítulo 3 fala sobre as definições teóricas necessárias para se entender o restante do trabalho. O Capítulo 4 faz uma descrição da área e dos trabalhos relacionados. O Capítulo 5 fala sobre as técnicas de análise que foram aplicadas para construir os resultados e o Capítulo 6 fala sobre os resultados obtidos a partir das análises descritas previamente. Finalmente, o Capítulo 7 apresenta conclusões e fala sobre possíveis trabalhos futuros.

2 MOBA

Este capitulo apresenta o funcionamento geral de MOBAs, e detalhes sobre *League of Legends*, foco deste trabalho. Então, apresentamos um detalhamento da informação presente em nossos dados, extraídos do *site* 'tribunal', aonde jogadores de League of Legends eram julgados sob a acusação de serem tóxicos.

2.1 MOBAs

MOBAs são jogos de equipe velozes e competitivos. Os primeiros MOBAs foram modificações de jogos RTS (*Real Time Strategy*), como DotA, um cenário para o jogo *Warcraft III*. A popularidade alcançada pelo DotA, abriu caminho para *League of Legends*, um jogo independente lançado em 2009 que consolidou o gênero MOBA.

Partidas de MOBAs são compostas por duas equipes, normalmente de 5 jogadores cada. Essas equipes se confrontam em uma arena, buscando destruir uma construção no centro da base do oponente. Estas bases são protegidas por torres e por criaturas denominadas *minions*. Para participar, cada jogador escolhe um entre vários personagens diferentes, que se especializam em um ou mais papéis dentro do jogo. Cada personagem tem um conjunto de habilidades únicas, que são desbloqueadas ao decorrer da partida.

Com poucas exceções, MOBAs incluem uma variedade de itens que podem ser comprados durante uma partida para melhorar um personagem. Esses itens constituem parte vital da força de um personagem, e eles podem ser comprados somente com ouro. Um jogador coleta ouro através de ações como a destruição de torres inimigas, conseguindo abates (quando o jogador mata outro jogador) e assistências, destruindo *minions* inimigos, entre outras. Estes itens, juntamente com o personagem que foi escolhido pelo jogador, definirão qual o papel que o jogador executará durante a partida.

Em uma partida, as equipes estão constantemente barrando umas as outras de obter vantagens dentro do jogo, mantendo um certo equilíbrio entre as forças das equipes. Acumular ouro ajuda jogadores a comprarem itens poderosos, que auxiliam na obtenção de abates. Abates recompensam um jogador com mais ouro, e ajudam a desbalancear a partida, já que o jogador morto só retorna ao jogo depois de um certo tempo, e neste meio tempo, a equipe que conseguiu o abate possui uma vantagem numérica. Assistências também são recompensadas com ouro, e correspondem a situações onde um jogador ajuda outro a conseguir um abate.

Trabalho em equipe é muito importante para vencer uma partida, e jogadores podem

se comunicar utilizando bate-papos textuais ou de voz, e/ou sinalizações providas pelo próprio jogo. A disponibilidade destas ferramentas diferem de jogo para jogo, mas a maioria dos MOBAs provê pelo menos um bate-papo textual. Eles também provêm um sistema de denúncia de jogadores tóxicos, através do qual jogadores que se sentem desconfortáveis de alguma maneira com o comportamento de outro jogador podem enviar uma reclamação.

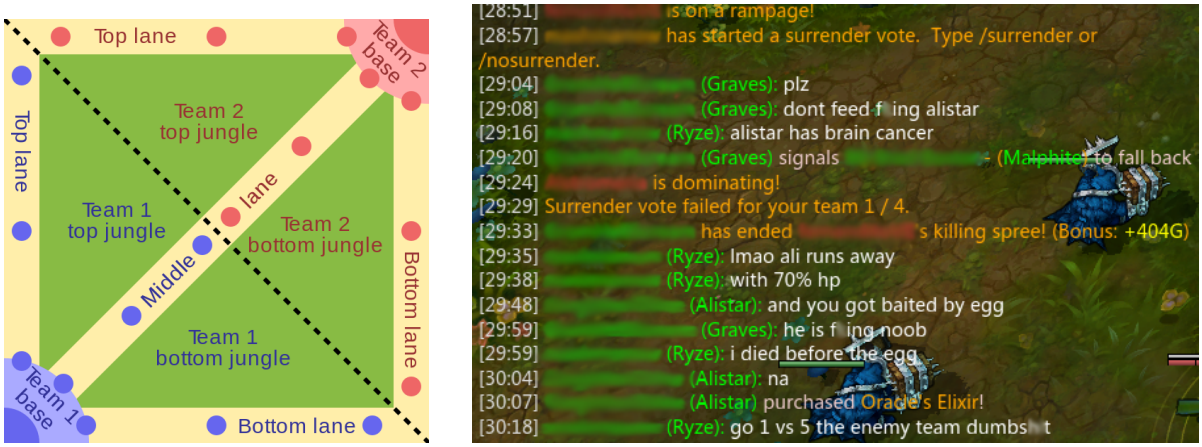
2.2 *League of Legends*

League of Legends (LoL) foi lançado em 2009 pela *Riot Games* como um jogo independente. O acrônimo MOBA foi cunhado pela própria Riot Games, já que não havia um consenso sobre qual expressão utilizar para referir-se ao gênero. O jogo experimentou um enorme crescimento e em 2012 se tornou o jogo de PC mais jogado no mundo.

LoL compartilha todas as características típicas de MOBAs, descritas na Seção 2.1. O mapa utilizado no jogo, comum à maioria dos MOBAs, está ilustrado na Figura 2.1a (em inglês). Ele é dividido em três corredores (topo, meio e baixo), separados por zonas que são referidas como floresta, ou *jungle*. A principal diferença entre LoL e seus competidores de gênero são os itens e personagens disponíveis, duração média de uma partida, e tamanho do mapa.

Em uma partida de LoL, os jogadores são distribuídos em cinco papéis, correspondentes a suas posições no começo do jogo: topo, meio, ADC (*Attack Damage Carry*), suporte e *jungler*. Jogadores do topo e do meio ocupam os corredores do topo e do meio do mapa, respectivamente, enquanto jogadores nos papéis de ADC e suporte jogam juntos no corredor de baixo. *Junglers* vagam pela floresta, aparecendo eventualmente nos corredores para dar apoio aos outros jogadores e ajudá-los a ganhar alguma vantagem sobre seus inimigos.

Uma partida de LoL é normalmente dividida em três momentos distintos: *early*, *mid* e *end* (KWAK; BLACKBURN, 2014). O *early* dura enquanto os jogadores estão fixos em suas lanes, normalmente focados em juntar ouro e derrubar a primeira torre do corredor. Este momento normalmente acaba quando a torre é derrubada, o que normalmente ocorre por volta dos $\approx$10 minutos de jogo. O *mid* é marcado pelos jogadores de um time andando juntos pelo mapa, buscando por falhas do inimigo para explorar. Este momento tem uma duração indeterminada, durando até o quando um dos times tem uma vantagem considerável sobre o outro, ou ambos os times estão fortes o suficiente para conseguir acabar o jogo em cima de uma falha do adversário, o que marca o momento *end*. Este momento normalmente corresponde aos $\approx$10 minutos finais de uma partida, e é marcado pelos times buscando uma última vantagem sobre os adversários para invadir a base destes e encerrar a partida.



(a) Disposição do Mapa. (b) Exemplo de Chat Tóxico.

Figura 2.1: Mapa e bate-papo de *League of Legends*.

Jogadores se comunicam em LoL utilizando bate-papo textual e sinais fornecidos pelo jogo. São providos dois canais de bate-papo: o bate-papo de equipe, e o bate-papo global. O primeiro é restrito aos jogadores de uma mesma equipe, e é utilizado principalmente para organização tática e socialização. O segundo é compartilhado por todos os jogadores. Apesar dele poder ser utilizado para socializar com o time inimigo, o bate-papo global é frequentemente usado de maneira tóxica (LIN, 2013). Já foi mostrado que o bate-papo textual é um componente importante para a identificação do comportamento tóxico (BLACKBURN; KWAK, 2014; MARTENS et al., 2015), já que ele é o principal meio de comunicação entre jogadores em uma partida. A Figura 2.1b mostra um exemplo de conversação tóxica no chat de LoL (em inglês).

2.3 Dados do Tribunal

Os dados usados neste trabalho são oriundos do tribunal do servidor norte-americano de *League of Legends*, uma página da internet onde jogadores denunciados por comportamento tóxico persistente e repetido eram julgados por outros jogadores. O tribunal ficou ativo de 2011 a 2014, quando ele foi substituído por outros métodos de detecção de jogadores tóxicos.

Os dados correspondentes ao julgamento de um jogador são referidos como um caso. Cada caso é associado a até 5 partidas escolhidas aleatoriamente entre as partidas nas quais aquele jogador foi denunciado por comportamento tóxico. Estas partidas tóxicas são usadas como evidência de comportamento tóxico persistente e repetido cometido pelo ofensor (i.e. o jogador denunciado). Os jogadores presentes em uma partida são divididos em grupos de acordo com sua relação com o ofensor: o *grupo aliado* denomina os quatro jogadores no mesmo time

do ofensor e o *grupo inimigo* denomina os cinco jogadores na equipe adversária ao ofensor. Finalmente, o grupo ofensor, ou somente *ofensor* denomina somente o próprio jogador tóxico. Os dados são completamente anonimizados, então não é possível determinar se um mesmo jogador está envolvido em mais de uma partida, com exceção do ofensor, que é o mesmo para todas as partidas em um caso.

Este trabalho explora dados predominantemente na língua inglesa, contendo 2.177.488 partidas, das quais consideramos 1.963.475 como válidas. Uma partida válida contêm dados de desempenho para todos os 10 jogadores, uma indicação do grupo de cada jogador (i.e. inimigo, aliado ou ofensor), e algum bate-papo.

Os casos do tribunal norte-americano estão armazenados em 115.96GB de arquivos .jsons. Cada um destes arquivos representa um caso, contendo um de uma cinco documentos JSON. A estrutura de destes documentos JSON representando uma partida, está disponível no Apêndice B.

Neste trabalho, foram utilizados os dados listados abaixo, extraídos destes documentos.

1. **Metadados da partida:**

   - A duração da partida, em segundos;
   - O resultado da partida para cada jogador, podendo ser: a) vitória, quando o jogador pertence ao time vencedor, b) derrota quando o jogador pertence ao time perdedor, ou c) abandono, quando o jogador sai da partida antes de seu término.

2. **Informações dos jogadores:**

   - Personagem escolhido;
   - Relação entre o jogador e o ofensor, ou seja, se o jogador está no grupo aliado, no grupo inimigo ou se ele é o próprio ofensor.

3. **Informações de performance:**

   - Número de abates, mortes e assistências realizados/sofridos por cada jogador;
   - Total de ouro obtido por cada jogador durante a partida.

4. **Informação de denúncias:**

   - Número de denúncias feitas por jogadores no grupo aliado;
   - Número de denúncias feitas por jogadores no grupo inimigo.

5. **Informações de bate-papo:** Dados referentes a cada linha de bate-papo escrita por algum jogador, contendo:

    - Registro de tempo (*timestamp*);
    - Personagem escolhido pelo jogador que escreveu a linha;
    - Grupo do jogador que escreveu a linha (aliado, inimigo ou ofensor);
    - Bate-papo utilizado (bate-papo do time ou bate-papo global);
    - Conteúdo textual da linha de bate-papo.

Informações que não estão listados acima não foram utilizadas por não serem considerados necessárias a análise, como por exemplo, informações sobre itens comprados e versão do jogo, ou por limitações de escopo, como por exemplo, informações sobre a motivação de denúncia mais utilizada.

O corpus adquirido a partir do conteúdo textual de todos os bate-papos tem um tamanho total de 5.3GB, com 1.176.484.341 termos que consiste em sua grande maioria de palavras na língua inglesa e 6.368.396 termos únicos, com parte relevante destes termos apresentando algum erro de grafia, ou sendo um jargão específico ao jogo.

3 FUNDAMENTAÇÃO TEÓRICA

Este capitulo apresenta a fundamentação teórica necessária a este trabalho. Apresentaremos tecnologias relacionadas a processamento e mineração de dados de texto que foram usadas neste trabalho, a saber: Representação vetorial de palavras e documentos, agrupamento e classificação de dados textuais. Também abordaremos regras de associação e métricas relacionadas. Finalmente, discutiremos brevemente a área de descobertas de emoções em texto, e a construção automatizada de léxicos de emoção.

3.1 Representação vetorial de documentos e palavras

Representar documentos e palavras em um formato vetorial, é uma parte essencial do processamento de dados textuais, visto que sem isso, não seria possível abordar computacionalmente tais valores. Abordagens clássicas para esse fim originaram-se no campo de recuperação de informações e baseiam-se na contagem e discriminação de todos os termos presentes em um documento, com a finalidade de representar o mesmo em um formato vetorial. A ideia por trás destes modelos é que documentos são conjuntos de palavras, e assim, são representados pelas palavras contidas nestes.

Entre os modelos clássicos de representação de documentos, os mais populares são a matriz de contagem, a matriz binária, e a matriz de TF-IDF (*Term Frequency - Inverse Document Frequency*) (AGGARWAL; ZHAI, 2012a). Na matriz de contagem, cada linha representa um documento, e cada coluna uma palavra no vocabulário, com cada célula representando quantas vezes cada termo no vocabulário ocorre em um documento. Já na matriz binária, cada célula é um valor binário, indicando se um dado termo está presente no documento ou não.

A matriz de TF-IDF busca dar um peso maior para palavras que aparecem em poucos documentos, visto que estas são mais representativas dos documentos em que aparecem. Isso é feito utilizando o TF-IDF, um produto entre a frequência de um termo em um documento (tf), e o inverso da frequência do mesmo termo em todos documentos do *corpus* (df), como expresso na Fórmula 3.1.

$$TFIDF = tf * log\left(\frac{1}{df}\right) \tag{3.1}$$

Modelos clássicos de representação de documentos possuem como vantagem sua simplicidade e transparência, e apresentam por desvantagens os fatos de possuírem vetores de di-

mensionalidade muito alta e bastante esparsos, o que encarece o uso computacional destes.

Se documentos são representados pelas suas palavras, então documentos são similares quando possuem palavras em comum. A partir deste princípio, vetores fornecidos pelos modelos de representação de documentos, podem ter sua similaridade calculada através da similaridade dos cossenos (Equação 3.2). Esta fórmula também é válida para calcular a similaridade entre vetores representando palavras, como as *word embbedings*, das quais falaremos a seguir.

$$sim(A, B) = \frac{A \cdot B}{||A|| * ||B||} = cos(A, B) \tag{3.2}$$

As *word embbedings* representam palavras através de vetores densos de números reais, que possuem menor dimensionalidade quando comparados com modelos clássicos de representação de palavras, modelos estes que se assemelham aos de representação de documentos descritos anteriormente. As *embbedings* também capturam a distância entre palavras, onde palavras com contextos similares, ou seja, palavras que aparecem na próximas das mesmas palavras, são similares. Finalmente, elas carregam algum significado semântico, podendo ser até mesmo ser objetos de operações simples com os vetores que representam as palavras (e.g. $king - man + woman = queen$).

Atualmente, dois modelos de construção de *word embbedings* são mais utilizados. O word2vec (MIKOLOV et al., 2013) é baseado em uma rede neural de duas camadas, que após treinada com um corpus, produz representações vetoriais de palavras. O modelo possui duas variações: o *continuous bag-of-words* (CBOW), onde o treinamento se dá a parir de cada termo sendo previsto a partir de seu contexto (i.e. os termos próximos a ele), e o *skip-gram*, onde cada termo é usado para prever o seu próprio contexto. Outro modelo de *word embbedings* bastante utilizado é o GLoVe (*Global Vectors*) (PENNINGTON; SOCHER; MANNING, 2014), que se baseia em aplicar técnicas de fatoração global de matriz, utilizando-se de uma matriz de co-ocorrência de palavras. Ambos modelos permitem a modificação do número de iterações realizadas sobre o corpus no passo de treinamento, bem como o número de dimensões dos vetores de palavras resultantes. Mais iterações e mais dimensões normalmente representam resultados mais precisos, mas a custo de desempenho computacional.

Funcionando de maneira similar ao word2vec, existe o *ParagraphVector* (LE; MIKOLOV, 2014), que utiliza-se do mesmo principio do word2vec para gerar vetores de documentos também densos e de baixa dimensionalidade, e que também tem similaridade calculável via cosseno. Seus autores reportaram resultados superiores aos dos modelos clássicos nas tarefas de recuperação de informações e análise de sentimentos.

Modelos de representação de documentos e palavras são utilizados como entrada para

vários algoritmos diferentes, como de classificação, predição, agrupamento, descoberta de tópicos, entre outros. Neste trabalho utilizamos como modelos de representação de documentos, a matriz de contagem como entrada para o LDA (*Latent Dirichlet Allocation*), um algoritmo de descoberta de tópicos, (BLEI; NG; JORDAN, 2003) e o *ParagraphVector* como entrada para o k-means, um algoritmo de agrupamento (AGGARWAL; ZHAI, 2012b). Já para a representação de palavras, utilizamos o word2vec, na variações de *skip-grams* e CBOW como entrada para uma rede neural, com a função de classificar sentimentos em palavras.

3.2 Agrupamento de dados textuais

A tarefa de agrupar dados textuais tem como o objetivo agrupar documentos e/ou textos similares, possibilitando a melhor interpretação e organização de grandes coleções de documentos. Tradicionalmente, processar um grande volume de dados textuais possui um alto custo computacional. Devido a isso, técnicas de agregação mais simples como *k-means* são preferidas (AGGARWAL; ZHAI, 2012b).

O *k-means* agrupa grupos de dados similares encontrando pontos centrais (centroides) a estes conjuntos de dados, onde cada dado é associado ao centróide mais próximo a ele. O k-means recebe como entrada a quantidade de agrupamentos que ele deve encontrar, a métrica a ser utilizada para estabelecer as distâncias entre os dados, e bem como os dados em forma vetorial. No caso de documentos, a distância normalmente utilizada é a distância de cossenos (Equação 3.3), que nada mais é do que o complemento da similaridade de cossenos. Após a sua execução, o algoritmo descreve os documentos de acordo com o centroide mais próximo a este documento.

$$dist(A, B) = 1 - sim(A, B) \tag{3.3}$$

Outra técnica de agregação de dados, desenvolvida especificamente para dados textuais, é o LDA (BLEI; NG; JORDAN, 2003). O LDA é uma abordagem probabilística, que descobre os tópicos de um conjunto de documentos, e descreve documentos a partir da probabilidade deles pertencerem a algum dos tópicos descobertos. Estes tópicos são descritos como uma distribuição de palavras, por exemplo, o tópico A pode ser representado 40% pela palavra1, 10% pela palavra2, 8% pela palavra3, e assim por diante, para todas as palavras no vocabulário.

O LDA recebe como entrada os documentos no formato de uma matriz de contagem, e o número de tópicos que ele deve encontrar. Após descobertos os tópicos, o LDA descreve

cada documento como uma distribuição dos tópicos encontrados. Por exemplo, dado como parâmetro 3 tópicos, um documento pode ser descrito como sendo representado por 60% do tópico A, 20% do tópico B, e 20% do tópico C. Na prática, os documentos podem ser agregados a partir de seus tópicos majoritários, fazendo o LDA funcionar como um algorítimo de agregação (AGGARWAL; ZHAI, 2012b).

Neste trabalho, testamos tanto o LDA quanto o k-means para descobrir agregações de bate-papos textuais com características similares.

3.3 Regras de associação

Algoritmos de mineração de regras de associação são utilizados para encontrar tendências e relações de co-ocorrência de valores escondidas em grandes conjuntos de dados. Dos algoritmos de mineração de regras, os mais famosos são o apriori (AGRAWAL; SRIKANT et al., 1994), o ECLAT (*Equivalence Class Transformation*) (ZAKI et al., 1997), e o FP-Growth (HAN; PEI; YIN, 2000). Neste trabalho escolhemos o apriori como algoritmo preferencial devido a sua popularidade e facilidade de uso (ZHANG; ZHANG, 2002).

O apriori busca itens frequentes dentro de um conjunto de transações, utilizando uma abordagem *bottom-up*. O algoritmo recebe como entrada um conjunto de transações T, as quais contêm itens de um conjunto I. O algoritmo descobre quais itens/conjunto de itens (*itemsets*) são mais frequentes, e a partir destes *itemsets* frequentes, gera regras de associação entre os mesmos. Estas regras possuem o formato LHS$\rightarrow$RHS (*Left Hand Side* e *Right Hand Side*, respecitvamente), onde LHS e RHS correspondem a *itemsets* contendo itens pertencentes a I.

É comum que muitas regras geradas não sejam relevantes ao problema, ou que elas sejam muito infrequentes para serem válidas. Por isso, ao se executar o apriori é importante se definir métricas para definir-se a relevância das regras. Tradicionalmente, são usados três métricas para tal fim: o suporte, a confiança e o *lift*.

Dada uma regra A $\rightarrow$ B, com $freq(A, B)$ sendo a função que nos dá a frequência com que os *itemsets* A e B aparecem juntos em T, o suporte representa a cobertura desta regra, indicando quantas transações em T correspondem a mesma. O suporte também pode ser interpretado como a probabilidade dos *itemsets* em uma regra aparecerem juntos em T. O suporte é calculado a partir da Equação 3.4.

A confiança de uma regra indica do total de transações onde os itens no lado esquerdo da regra (LHS) aparecem, quantas transações correspondem à regra em questão. Ela também pode ser interpretada como à probabilidade condicional do *itemset* B ocorrer, dada a presença

do *itemset* A em T. Note que a confiança de A $\rightarrow$ B é diferente da confiança de B $\rightarrow$ A. A confiança é dada pela Equação 3.5.

$$suporte(A \rightarrow B) = \frac{freq(A,B)}{|T|} = P(A \cap B) \tag{3.4}$$

$$confianca(A \rightarrow B) = \frac{freq(A,B)}{freq(A)} = P(B|A) \tag{3.5}$$

$$lift(A \rightarrow B) = \frac{confianca(A \rightarrow B)}{P(B)} = \frac{P(B|A)}{P(B)} \tag{3.6}$$

Diferentemente das duas métricas anteriores, o *lift* não é uma probabilidade, mas sim um valor que indica o grau de interdependência entre os *itemsets* em uma regra. Quanto mais distante o valor do *lift* for de 1, maior o grau de dependência entre os *itemsets*, e mais forte é a regra. Um *lift* igual ou muito próximo a 1 indica que os itens nos *itemsets* da regra são independentes entre si, significando que a regra representa uma relação que ocorre de maneira completamente aleatória, e não é interessante ao problema. O *lift* é calculado pela Equação 3.6.

No nosso trabalho, utilizamos o apriori e suas métricas para detectar as tendências de transições entre tópicos em momentos diferentes de uma partida.

3.4 Análise de sentimentos

O termo *análise de sentimentos* é usado para se referir a áreas diferentes, mas relacionadas. Ele é usado mais comumente para designar a tarefa de analisar a polaridade de uma palavra ou expressão. Contudo, existem outras tarefas relacionadas à análise de emoções, como mineração de emoções ou de opiniões (BECKER; MOREIRA; SANTOS, 2017).

Uma maneira simples de se medir sentimentos é utilizar a valência ou polaridade, para representar a orientação de um sentimento, variando entre negativo e positivo. Esta medida pode tanto ser um número real, como um valor discreto (e.g. positivo, negativo ou neutro). Quando este sentimento de polaridade está associado a um alvo, ele é denotado opinião.

O termo *emoção* representa uma demonstração consciente de um sentimento, que pode ocorrer de maneira espontânea ou fingida. As pessoas manifestam emoções para o mundo, e estas podem ter tanto a função de expressão pessoal, como de atender expectativas sociais (SHOUSE, 2007). Emoções podem ser expressas via linguagem e a tarefa de rotular emoções consiste em detectar palavras ou expressões de emoção de acordo com um modelo de

emoções pré-determinado (MUNEZERO et al., 2014).

O modelo de emoções de Plutchik (1984) enumera oito emoções básicas (alegria, antecipação, confiança, medo, nojo, raiva, surpresa, tristeza), com estas emoções não sendo mutuamente exclusivas. Outro modelo de emoção é o VAD (*Valence, Arousal, Domination*), que representa emoções num espaço tri-dimensional formado pelos componentes de valência, excitação e dominância (RUSSELL; MEHRABIAN, 1977).

Um léxico de sentimentos é um conjunto de palavras anotadas com seus respectivos sentimentos (e.g. polaridade e/ou emoção). Um léxico de sentimento pode ser de propósito geral, como o SentiWordNet (BACCIANELLA; ESULI; SEBASTIANI, 2010) para a polaridade, o ANEW (BRADLEY; LANG, 1999) para o modelo VAD, e o NRC (MOHAMMAD; TURNEY, 2013) para emoções de Plutchik e polaridade. O NRC descreve o sentimento de 14.182 palavras comumente utilizadas na língua inglesa de acordo com o modelo de emoções de Plutchik, juntamente com dois valores de polaridade: positivo e negativo. Léxicos de propósito geral normalmente utilizam a semântica do sentimento mais usual em uma língua como base para a sua construção.

Contudo, várias subculturas e áreas específicas de conhecimento utilizam frequentemente jargões específicos, que não estão presentes nestes léxicos de propósito geral, e que só possuem significado dentro do contexto daquela área. Para se minerar sentimentos dentro destas áreas, ou domínios, é necessário um léxico específico àquele domínio. Um exemplo é um léxico especifico a *microblogs* (MOHAMMAD, 2012), que dá significado por exemplo a *emoticons* e abreviaturas usadas nesta mídia.

Como parte deste trabalho, construímos um léxico de emoções específico ao domínio de MOBAs.

3.5 Construção automática léxicos de emoções

Léxicos de emoções podem ser criados manualmente, através de conhecimento de especialistas, como o ANEW, ou através de um processo de crowdsourcing, como o NRC. Abordagens automatizadas para a criação de léxicos de emoções também existem. Uma delas é o uso de palavras semente (BACCIANELLA; ESULI; SEBASTIANI, 2010; MOHAMMAD, 2012; SONG et al., 2015), que são palavras cujo sentimento já é conhecido. A partir destas palavras, o léxico é expandido em um processo gradual, normalmente verificando algum tipo de semelhança ou proximidade (e.g. sinônimos) de outras palavras ainda não rotuladas em relação às sementes, repetindo este processo utilizando as palavras recém rotuladas como novas sementes.

Outra abordagem é a expansão de um léxico pré-existente usando um algoritmo de aprendizado que será treinado com vetores de palavras do léxico pré-existente, gerados a partir de um modelo de representação de palavras alimentado com um corpus de domínio específico. Este modelo de aprendizado treinado irá prever as emoções de palavras novas, presentes em um corpus de domínio específico, mas não-rotuladas no léxico original. Bravo-Marquez et al. (2016) obtiveram resultados satisfatórios na construção de um léxico de emoções para tweets de acordo com esta abordagem. Os autores expandiram o léxico NRC, treinando um modelo de aprendizado a partir das palavras presentes na intersecção de um corpus de *tweets*, de autoria dos autores, com o NRC. Para aprendizado, foi utilizado um algorítmo de Regressão Logistica (DAYTON, 1992). Finalmente, as emoções das palavras do corpus não presentes no léxico foram descobertas a partir do modelo treinado.

Neste trabalho seguiremos seguiremos esta segunda abordagem para construir nosso léxico. Utilizamos um algoritmo de aprendizado para expandir um léxico base, o NRC, com palavras específicas do domínio de MOBAs existentes em nosso corpus de bate-papo. É importante frisar que dados de emoções no NRC são multi-rótulo e desbalanceados, ou seja, uma palavra pode representar mais de um sentimento ao mesmo tempo, e os diferentes rótulos de sentimentos aparecem em proporções desiguais no léxico.

3.6 Algoritmos de Classificação

Algoritmos de classificação são ferramentas utilizadas para definir a que categoria uma nova observação pertence, baseado em um conjunto de observações previamente categorizadas (chamada de conjunto de treinamento) fornecidas ao algoritmo. Algoritmos de classificação são amplamente utilizados em uma multitude de áreas dentro e fora da computação, como computação visual, processamento de linguagem natural, medicina, direito, entre várias outras.

Bravo-Marquez et al. (2016) utilizam-se do algoritmo de Regressão Logística (DAYTON, 1992) para realizar a classificação de sentimentos de palavras. Esse algoritmo, que só funciona com problemas binários, tenta estimar os *odds* de um elemento, ou seja, a razão entre a probabilidade desse elemento pertencer ao rótulo e a probabilidade do elemento não pertencer ao rótulo, encontrando coeficientes e combinando eles linearmente os atributos de cada observação. No final, aplicamos uma função logística nesses *odds* para conseguir a probabilidade de pertinência ao rótulo. Valores maiores do que 0,5 normalmente indicam que o elemento pertence àquele rótulo, e menores indicam o oposto.

Em nosso trabalho exploramos também outros algoritmos de classificação: o SVM (*Sup-*

port Vector Machines), bastante famoso na área de classificação de texto (JOACHIMS, 1998) e o *feedforward* MLP (*Multi-Layer Perceptron*), que demonstrou um bom desempenho ao classificar a polaridade de palavras no NRC (CARDOSO; ROY, 2016).

O SVM é um método matemático que busca a função que melhor define a fronteira entre dois ou mais grupos. A implementação clássica do SVM utiliza-se de hiperplanos para dividir o espaço representado pelos dados de treinamento linearmente, mas existem outras implementações, não lineares, como por exemplo o SVM RBF (*Radial basis function*) (SCHOLKOPF et al., 1997). O SVM permite que sejam fornecidos pesos diferentes para cada um dos rótulos presentes no conjunto de treinamento, de modo a amenizar problemas com desbalanceamento de rótulos. É bom frisarmos que o SVM não fornece suporte a problemas multi-rotulo, sendo necessário abordagens complementares para fornecer este suporte.

O MLP é um tipo de rede neural, composta por três componentes principais:

- Neurônios, que consistem de pesos e uma função de ativação, divididos em três ou mais camadas. Funções de ativação podem ser ajustadas diferentemente para cada camada. Em problemas de classificação, a última camada normalmente utiliza funções como softmax, ou tanh (tangente hiperbólica), enquanto as demais camadas podem usar, além destas, funções como sigmóide ou reLU (SEVERYN; MOSCHITTI, 2015).
- Uma função de perda, que determina o quão longe o algoritmo está de uma solução ótima. Funções de perdas comumente usadas em redes neurais são o erro quadrático médio (MSE), a raiz do erro quadratico médio (RMSE), e a entropia-cruzada.
- Um algoritmo de otimização, cujo objetivo é otimizar a função de perda. Algoritmos de otimização comuns são o gradiente descendente estocástico (SGD), e variações deste, como o ADAM (KINGMA; BA, 2014) e o RMSProp (TIELEMAN; HINTON, 2012).

O treinamento de uma rede neural se dá pela execução repetida do algoritmo de otimização em pedaços diferentes do conjunto de treinamento, por um número especificado de iterações. Cada iteração do algoritmo ajusta os pesos dos neurônios, que são modificados pela função de ativação, de modo a diminuir o valor da função de perda. A escolha tanto da função de perda como do algoritmo de otimização são dependentes do problema.

Entre as camadas de uma rede é possível aplicar uma técnica chamada de *dropout* (SRIVASTAVA et al., 2014), que consiste em zerar a ativação de um conjunto aleatório diferente de neurônios em cada iteração do algoritmo. Esta técnica é usada para amenizar problemas de *overfitting*, que ocorre quando um algoritmo se adapta excessivamente aos dados de treinamento, demonstrando resultados sub-ótimos nos dados reais.

Nem todos modelos de aprendizado dão suporte a uma abordagem multi-rótulo. O algo-

ritmo de relevância binária (LUACES et al., 2012) busca dar esse suporte criando um conjunto de classificadores binários independentes para fazer o trabalho que um só classificador não suporta. Este algoritmo funciona separando o problema multi-rotulado em vários problemas binários, um para cada classe de rótulo, onde cada um destes classificadores binários tem somente que identificar se a entrada pertence a um determinado rótulo, ou não. No final, os resultado de todos os classificadores são mesclados, com cada classificador indicando o resultado da entrada para o seu rótulo.

O algoritmo de cadeia de classificadores é uma extensão do algoritmo de relevância binária, onde cada classificador binário utiliza o resultados de outros classificadores executados anteriormente para melhorar seu resultado, formando assim uma espécie de cascata de classificadores. O algoritmo de cadeia de classificadores tende a mostrar resultados melhores do que o de relevância binária quando os rótulos de um problema multi-rótulo são dependentes entre si (READ et al., 2011).

Conjuntos de dados com desbalanceamento na frequência de aparição dos rótulos no conjunto de treinamento podem afetar seriamente a performance de um modelo de aprendizagem. Para resolver esse problema existem algoritmos de super e sub-amostragem, que adicionam e removem, respectivamente, dados do conjunto de treinamento para produzir classes melhor balanceadas. Um dos algoritmos de super-amostragem mais populares é o SMOTE (*Synthetic Minority Oversampling TEchnique*) (CHAWLA et al., 2002), que gera itens extras utilizando técnicas de interpolação. O ENN (*Edited Nearest Neighbors*) (JANKOWSKI; GROCHOWSKI, 2004) é um algoritmo de sub-amostragem, que remove itens que não são similares o suficiente aos seus vizinhos. Isso é feito através de um algoritmo de vizinhos mais próximos, onde o número de vizinhos a serem considerados é passado como parâmetro.

Nos próximos capítulos utilizaremos tanto SVM como MLP para realizar a classificação de emoções em palavras, complementados pelo algoritmo de relevância binária, que permite a aplicação destes modelos em dados multi-rotulados, e pelo ENN melhora os resultados dos modelos em conjuntos dados desbalanceados.

3.6.1 Métricas de Avaliação

A análise dos resultados de um modelo de aprendizagem é feita através de uma tabela de confusão, ilustrada na Tabela 3.1 para um problema binário. Estas tabelas dividem os resultados de um modelo nas diferentes possibilidades de classificação. Para um problema binário, isto corresponde a quatro diferentes valores: Verdadeiros Positivos (VPs), Verdadeiros Negati-

Tabela 3.1: Tabela ou Matriz de Confusão

		Previsto	
		Positivo	**Negativo**
Real	**Positivo**	VP	FN
	Negativo	FP	VN

vos(VNs), Falsos Positivos (FPs) e Falsos Negativos (FNs). A importância de cada um destes valores varia com o problema em mãos. Para cobrir casos distintos foram criadas as métricas de precisão e *recall*.

A precisão (P), mostrada na Fórmula 3.7, indica, do total dos itens marcados pelo modelo com um rótulo, quantos deles foram rotulados corretamente. Ela é importante quando se deseja que o modelo minimize a taxa de FPs. O *recall* (R), mostrado na Fórmula 3.8, indica, do total dos itens de um rótulo nos dados, quantos foram rotulados corretamente pelo modelo. Ele é importante quando se deseja que o modelo minimize a taxa de FNs. A Medida-F, mostrada na Fórmula 3.9, é uma média harmônica da precisão com o *recall*, e é utilizada quando ambas métricas são importantes para o problema.

Quando é necessário combinar a Medida-F de diferentes rótulos, podemos utilizar várias métodos de combinação, como Macro ou Micro Medida-F, que são diferentes maneiras de se obter uma média entre várias Medidas-F através do cálculo de uma precisão e um *recall* médio. A Macro Medida-F simplesmente calcula a precisão e o *recall* médio a partir da média aritmética das precisões e dos *recalls* de todas as classes. A Micro Medida-F, substituí os VPs, FPs e FNs nos cálculos da precisão/*recall* pela soma dos VPs/FPs/FNs de todas as classes. Em ambas métricas, a precisão e o *recall* calculados são então aplicados na fórmula da Medida-F clássica (Fórmula 3.9).

A Micro Medida-F é utilizada por Bravo-Marquez et al. (2016) para a comparação de resultados, já que ela leva em consideração o desbalanceamento da ocorrência das classes. Pela mesma razão, será também usada no presente trabalho na avaliação dos resultados do modelo de classificação de emoções.

$$Precisao = \frac{VP}{VP + FP} \tag{3.7}$$

$$Recall = \frac{VP}{VP + FN} \tag{3.8}$$

$$MedidaF = 2 * \frac{P * R}{P + R} \tag{3.9}$$

4 TRABALHOS RELACIONADOS

Neste capítulo abordaremos trabalhos da área de caracterização e estudo do comportamento tóxico em jogos, em suas diferentes facetas. Aqui apresentaremos as definições de várias nomenclaturas diferentes dadas ao comportamento tóxico, bem como exploraremos suas motivações, características, e consequências sobre jogadores não tóxicos.

4.1 Definições de comportamento tóxico

Muitos termos são usados para se referir ao comportamento tóxico, que mudam com o passar dos anos e diferem entre comunidades. Suler (2004) pesquisa as causas do comportamento tóxico online e cunha o termo *desinibição tóxica*, em contraponto a uma *desinibição benigna*, para designar pessoas que agem de maneira ofensiva e agressiva em comunidades online, ainda que não necessariamente exibam esse comportamento no seu cotidiano. O autor cita a anonimidade, dissociação, assincronia e invisibilidade resultante da comunicação mediada por computador como motivos dessa desinibição, bem como a falta de figuras de autoridade em meios online e motivos pessoais.

Outros trabalhos dissertam sobre *griefing*, um comportamento relacionado à desinibição tóxica, que acontece quando um jogador se diverte ou ganha vantagens, tanto dentro do jogo, como emocionais, às custas de outros jogadores (FOO; KOIVISTO, 2004). Alguns estudos mostram *griefing* como um comportamento exclusivamente intencional (CHESNEY et al., 2009; ROSS; WEAVER, 2012), mas ele também pode ocorrer de maneiras não intencionais (FOO; KOIVISTO, 2004; LIN; CHUEN-TSAI, 2005; CHESNEY et al., 2009).

Trollagem é outro comportamento que pode ser associado com a desinibição tóxica. Ele é caracterizado por alguém (o *troll*) que se infiltra em um grupo, normalmente demonstrando um 'eu' dissimulado, aparentando sinceridade, com o objetivo de perturbar aquele grupo e causar conflitos, para sua própria diversão (HARDAKER, 2010; GOLF-PAPEZ; VEER, 2017).

Comportamento tóxico é equivalente a *griefing* (BLACKBURN; KWAK, 2014), e como tal pode ser intencional ou não-intencional. Comportamento tóxico intencional ocorre quando um jogador, intencionalmente, assedia ou agride outros jogadores. Neste sentido, *trollagem* no contexto de jogos online é um comportamento tóxico intencional, já que o *troll* intencionalmente assedia outros jogadores para ganho próprio. Comportamento tóxico não intencional é o tipo de comportamento tóxico mais prevalente e ocorre quando um jogador faz ações dentro do jogo que acabam irritando ou agredindo outros jogadores (CHEN; DUH; NG, 2009; LIN;

CHUEN-TSAI, 2005).

No decorrer deste capítulo, buscamos utilizar as nomenclaturas (i.e. *griefing*, *trollagem*, ou comportamento tóxico) adotadas pelos próprios trabalhos, como maneira de distinguir quais trabalhos utilizam quais nomenclaturas.

4.2 Caracterização do comportamento tóxico

Foo and Koivisto (2004) buscam categorizar e definir *griefing* através de perguntas feitas a jogadores *griefers*, não *griefers* e desenvolvedores de jogos. Os autores falam da existência de *grief* intencional e não intencional, e dividem *grief* em RPGs online em quatro categorias: a) assédio, b) imposição de poder, c) estelionato e d) ato ganancioso. Destas, somente o 'ato ganancioso' é considerado como *griefing* não intencional. Então, definem *grief* em RPGs online como: "um estilo de jogo que atrapalha a experiência de jogo de outros jogadores, normalmente com tal intenção. Quando o ato não possui a intenção específica de atrapalhar, mas ainda assim o jogador que faz o ato é o único beneficiário dele, este ato é denominado como ato ganancioso, uma forma sutil de *grief*".

Lin and Chuen-Tsai (2005) estudam uma comunidade de jogadores de Taiwan, procurando descobrir mais detalhes sobre os *griefers*. Os autores: a) mostram que *griefing* é algo ambíguo, subjetivo e que muda com o tempo, b) confirmam a existência da divisão entre *griefing* intencional e não intencional, e c) descobrem que a maioria dos *griefers* se encaixam como não intencionais. *Griefers* intencionais praticam atos tóxicos propositadamente, dizem explicitamente que não respeitam as regras do jogo, agindo de acordo com suas próprias regras, e realizam *grief* de maneira repetida e sistemática. *Griefers* não intencionais, não reconhecem seus atos como tóxicos, dizem respeitar as regras do jogo, e realizam *grief* de maneira ocasional e inconsciente.

Chesney et al. (2009) estudam *griefing* em *Second Life*, uma simulação virtual do mundo real, observando as interações entre os residentes do jogo. Nos seus resultados, eles consideraram somente *griefers* intencionais, e listaram as motivações do comportamento destes como: a) imposição de poder, usando seu conhecimento superior sobre o jogo como um meio de estabelecer uma hierarquia, b) dissociação, onde os jogadores se esquivam das consequências de suas ações por tudo ser 'somente um jogo', e c) anonimidade, que garante maiores chances de que seus atos não gerarão consequências.

Hardaker (2010) utiliza dados de um corpus de conversas em comunidades online para defender que comportamento tóxico nestas comunidades pode ser intencional ou não por parte

do emissor da mensagem. O receptor pode interpretar uma mensagem como tóxica mesmo que esta não tenha sido a intenção original do emissor, reforçando assim a natureza subjetiva do comportamento tóxico. A autora também fala sobre *trolls*, os definido como "usuários de comunicação mediada por computador que aparentam sinceramente desejar pertencer ao grupo em questão, incluindo a profissão ou transmissão de intenções pseudo-sinceras, mas cujas reais intenções é atrapalhar ou criar/exacerbar conflitos para sua diversão pessoal". Esta definição se encaixa na definição de *griefing* intencional provida por Foo and Koivisto (2004), então trollagem será tratada como tal.

Buckels, Trapnell and Paulhus (2014) também estudam *trolls*, através de questionários que perguntam o que as pessoas gostam de fazer enquanto online. Pessoas que afirmaram que gostam de *trollar* acabam mostrando muito mais sinais de sadismo, maquiavelismo, narcisimo e psicopatia do que os outros participantes. Os autores também criam um modelo de regressão para descobrir quais destas quatro variáveis se associam mais fortemente com *trolls*. Seus resultados mostram que apenas sadismo prevê *trolls* com eficácia, sendo esta variável mais associada com *trollagem* das quatro.

Thacker and Griffiths (2012) estudam *trolls* através de dados de um questionário direcionado a jogadores. Eles mostram que comportamento tóxico é comum em jogos online, com apenas 20% dos jogadores participantes afirmando que raramente ou nunca presenciaram tal ato. Entretanto, 55% dos participantes disseram que raramente ou nunca foram vitimas de comportamento tóxico. Também é mostrado que jogadores que se reconhecem como *trolls* (i.e. jogadores tóxicos intencionais) jogam com muito mais frequência do que o jogador médio. Eles também descobrem três formas de *trollagem*: a) *griefing/trollagem* (o uso de mecânicas de jogo para intencionalmente atrapalhar outros jogadores), b) sexismo/racismo, e c) estelionato/enganação. Os autores também descobrem que estes *trolls* são motivados por a) diversão, b) tédio ou c) vingança.

Ross and Weaver (2012) também falam sobre *grief* intencional, através de um experimento onde jogadores jogam um cenário customizados do RPG online *Neverwinter Nights*, criado com o objetivo de estudar as relações entre o *griefer* e sua vítima. Eles concluíram que *grief* influencia o comportamento da vítima negativamente, aumentando seus níveis de agressão e frustração, o que a torna mais suscetível a ser uma fonte de comportamento tóxico. De fato, os autores mostram que jogadores que foram vítimas de *grief* têm mais chances de se tornarem *griefers* em um futuro próximo, indicando que comportamento tóxico é transmissível.

Barnett, Coulson and Foreman (2010) pergunta a jogadores de *World of Warcraft*, outro RPG online, quais fatores do jogo geram os irritam. Eles descobrem que as principais fontes de

raiva entre estes jogadores são: a) *raids*, um modo de jogo altamente dependente da cooperação entre vários jogadores, b) *griefers*, c) situações que aparentam ser uma perda de tempo e d) jogadores anti-sociais. Um tema comum a todo estes cenários é que jogadores se irritam com comportamento tóxico de outros jogadores, independente se este é intencional ou não. Situações onde jogadores têm que cooperar entre si, mostram-se como altamente estressantes, e uma fonte comum de irritação. Finalmente, jogadores irritam-se com comportamento ofensivo em geral por parte de outros, independente deste comportamento estar incluído no contexto do jogo.

Shores et al. (2014) buscam examinar a natureza e os efeitos do comportamento tóxico em LoL, analisando dados de um *add-on* do jogo, criado pela empresa duowan[1]. Baseado em dados específicos do *add-on*, eles criam um índice de toxicidade para medir a intensidade do comportamento tóxico por ofensores. Com isso, eles mostram que toxicidade leva a um índice menor de retenção de novos jogadores, bem como que partidas ranqueadas (i.e. partidas que mudam a classificação geral do jogador), por serem mais competitivas, são mais propensas a comportamento tóxico. Os autores sugerem um foco maior no aspectos cooperativos do jogo, ao invés dos competitivos, como uma maneira de combater comportamento tóxico.

Blackburn and Kwak (2014) utilizam dados do tribunal de LoL para criar um detector de comportamento tóxico para este jogo utilizando técnicas de aprendizado de máquina. O autor utiliza vários dados como *features* na tarefa de construir um classificador capaz de detectar jogadores tóxicos, como: a) o KDA, uma métrica já consagrada nas comunidades de jogos online para medir a performance dos jogadores, juntamente com outras informações relativas a performance dos jogadores, como ouro acumulado e dano causado; b) o número de denúncias feitas por aliados e inimigos, bem como o motivo destas denúncias; c) o bate-papo entre jogadores, tanto em termos de palavras utilizadas, quanto da polaridade destas. Os autores também consideram que um time (i.e. aliado ou inimigo) só é vitima do ofensor se este foi denunciado pelo menos uma vez por algum jogador destes times. O valor de valência das palavras utilizadas pelos ofensores revelou-se como a *feature* que melhor prediz comportamento tóxico, seguido por quantidade de denúncias.

Os trabalhos citados mostram que comportamento tóxico gera raiva e frustração nos jogadores contaminados (ROSS; WEAVER, 2012; BARNETT; COULSON; FOREMAN, 2010). Comportamento tóxico floresce em ambientes altamente competitivos, e de alta tensão (SHORES et al., 2014; BARNETT; COULSON; FOREMAN, 2010). Dificuldades de comunicação entre membros de uma equipe é outra consequência comum do comportamento tóxico, o que pode alimentar mais ainda o mau humor dos jogadores contaminados (HARDAKER, 2010).

[1] www.duowan.com

Isso é perigoso, já que jogadores expostos ao comportamento tóxico podem se tornar tóxicos por causa do ambiente negativo criado, ou por predisposições pessoais (ROSS; WEAVER, 2012; CHEN; DUH; NG, 2009; THACKER; GRIFFITHS, 2012). Uma vez descoberto, comportamento tóxico deve ser rapidamente restrito para evitar contaminação (ROSS; WEAVER, 2012). Pistas para a descoberta do comportamento tóxico incluem o texto utilizado pelo ofensor, e informações dadas pelos outros jogadores, como denúncias (BLACKBURN; KWAK, 2014).

4.3 Caracterização textual do comportamento de jogadores em MOBAs

Martens et al. (2015) anotam manualmente palavras no vocabulário de MOBAs a partir de dados coletados de partidas do jogo DotA. A anotação realizada pelos autores resultaram em 10 categorias, definidas arbitrariamente: não-latim, elogio, ruim (profanidades), riso, *smiley* (emoticons), símbolos, jargões, comandos (comandos do jogo utilizados para ativar certos efeitos), *stopwords* e marcas temporais (*timestamps*) geradas automaticamente. Os autores mostram que certos eventos dentro do jogo levam a uma propensão maior ao uso de profanidades, e mostram que as palavras utilizadas em uma partida predizem bem o vencedor desta.

Kwak and Blackburn (2014) utilizam os mesmos dados utilizados em um trabalho anterior (BLACKBURN; KWAK, 2014) para explorar as diferenças linguísticas entre jogadores tóxicos e não-tóxicos. Para tal fim, cada partida é dividida temporalmente em 100 pedaços de tamanhos iguais. Então todas as partidas são sobrepostas e unificadas, resultando em 100 divisões temporais que contêm os bate-papos de todas as partidas. Então, os uni e bigramas utilizados por estes jogadores em cada um destas divisões são analisados. Os autores mostram que o uso de texto em uma partida não é totalmente uniforme, mostrando picos no inicio e no fim. Estes picos são correspondentes ao momentos *early* e *late* respectivamente, com a quantidade de texto utilizada no *mid* sendo relativamente constante. Também vemos que ofensores param de elogiar e usar comunicação tática em algum ponto de uma partida, transicionado para um comportamento considerado mais tóxico. Os autores também mostram que jogadores tóxicos falam mais, e utilizando palavras mais vulgares do que os jogadores não-tóxicos. Estes utilizam um volume menor de mensagens, preferindo palavras relacionadas a chamadas táticas.

4.4 Considerações finais

Dos trabalhos citados, todos focam na caracterização do jogador tóxico, contrastando-os com os demais jogadores, considerados não tóxicos. Contudo, acreditamos que uma divisão mais complexas de comportamentos de jogadores online pode ser feita, e que tais divisões podem ser analisadas de maneira mais profunda. Martens et al. (2015) dividem o vocabulário utilizado por jogadores em diferentes categorias, contudo, esta divisão é feita de maneira manual, com um número arbitrário de classes, e categorias igualmente arbitrárias, que não são avaliadas por nenhum mecanismo externo.

A importância do conteúdo do bate-papo para analisar o comportamento de jogadores é evidente (BLACKBURN; KWAK, 2014; HARDAKER, 2010; MARTENS et al., 2015; KWAK; BLACKBURN, 2014). Entretanto, quando estes trabalhos categorizam os jogadores, eles não realizam uma análise aprofundada de nenhuma das divisões feitas, nem buscam detalhar comportamentos associados a elas, meramente mostrando as divisões e suas respectivas palavras associadas.

Neste trabalho nós propomos padrões de conversa de jogadores em partidas de LoL, descobertos através de um algoritmo de clusterização de texto e interpretados por jogadores experientes. Também propomos métricas para medir a performance e a contaminação tóxica de grupos de jogadores durante uma partida, um estudo profundo dos padrões de conversa descobertos e a análise de como a mudança entre tais padrões ocorre nas partidas. Além disso, propomos um dicionário de sentimentos voltado ao domínio de MOBAs, juntamente com uma análise demonstrando quais emoções caracterizam melhor diferentes tópicos.

Usando o ferramental proposto, foram realizadas uma série de análises nos padrões descobertos, caracterizando o comportamento dos jogadores de acordo com o padrão de conversa prevalente, dividindo os padrões descobertos em dois grupos, positivos e negativos, que são relatados a como estes padrões afetam as chances de vitória de um time e a sua contaminação tóxica. Também estabelecemos comparações entre diferentes grupos de jogadores, e como cada um deles se sai em frente aos padrões de conversa dominantes nos respectivos grupos, e como eles reagem aos padrões de conversa utilizados pelo ofensor. Estabelecemos relações temporais entre os padrões descobertos, buscando descobrir como se dá a transição entre diferentes tópicos de conversa, e como diferentes grupos de jogadores mudam de tópico durante uma partida. Finalmente, mostramos as emoções evocadas pelos jogadores quando eles utilizam cada um dos padrões de conversa.

5 ANÁLISE DE COMPORTAMENTO TÓXICO E SEUS EFEITOS

Neste capítulo, descrevemos a abordagem proposta para analisar o comportamento tóxico em partidas de LoL e responder nossas questões de pesquisa. Discutiremos os métodos propostos para caracterização de tópicos de conversação prevalentes, análise de seus efeitos nos diferentes tipos de jogadores, descoberta de transições entre os tópicos usados ao longo das partidas, e a caracterização dos tópicos em termos de um modelo de emoções.

5.1 Visão Geral

Vimos que em jogos de equipe como MOBAs, trabalho em equipe é essencial e determina quem ganha ou perde uma partida. A maneira como os jogadores de uma equipe se comportam costuma refletir o grau de entrosamento entre estes jogadores. Outros trabalhos (BLACKBURN; KWAK, 2014; MARTENS et al., 2015; KWAK; BLACKBURN, 2014) mostraram que *features* extraídas de canais de comunicação entre jogadores são importantes para se caracterizar comportamento tóxico. Contudo, estes trabalhos limitam-se a analisar os textos extraídos sob a ótica da dicotomia jogador tóxico/não-tóxico.

Pretendemos ir além desta dicotomia neste trabalho, descobrindo mais variações nos comportamentos dos jogadores e provendo uma analise mais rica do texto presente em chats on-line de partidas de MOBAs. Aqui, utilizamos técnicas de agrupamento de texto para criar e interpretar tópicos de conversação em partidas de MOBAs, e propusemos métricas específicas para avaliar a performance e a contaminação tóxica relacionada a estes tópicos. Também aplicamos regras de associação entre os tópicos ao longo de uma partida, com o objetivo de descobrir quais transições entre tópicos são mais comuns. Finalmente, construímos um dicionário de emoções específico para o domínio de MOBA, buscando analisar as emoções presentes nos tópicos de conversação. Um resumo de nossas propostas para responder às questões de pesquisa deste trabalho estão descrita na Tabela 5.1, junto com a seção onde cada tema será discutido.

Os bate-papo textuais de LoL buscam permitir que jogadores coordenem seus esforços como uma equipe. Contudo, são também um dos principais canais de manifestação de comportamento tóxico (BLACKBURN; KWAK, 2014; KWAK; BLACKBURN, 2014). Identificar os principais padrões de conversação que emergem destes bate-papos, nos permite relacionar estes padrões de conversação com comportamentos específicos. Para responder a pergunta "**quais tópicos de conversa são comumente utilizados por jogadores em partidas de MOBAs?**",

Tabela 5.1: Resumo das Abordagens Propostas

Questão de Pesquisa	**Abordagem Proposta**	**Seção**
Quais tópicos de conversa são comumente utilizados por jogadores em partidas de MOBAs?	• Descoberta de tópicos de conversa prevalentes através de um algoritmo de agrupamento de texto (LDA); • Interpretação destes tópicos utilizando-se do conhecimento de jogadores veteranos.	5.2 (subsecs. 5.2.1 e 5.2.2)
Como cada um dos tópicos descobertos se associa com diferentes tipos de jogadores, divididos em grupos de acordo com sua associação ao jogador tóxico (jogadores tóxicos, seus aliados, e seus inimigos)?	• Rotulação de cada grupo de jogadores com um tópico de conversação; • Estudo da distribuição dos tópicos entre estes grupos de jogadores.	5.2 (subsec. 5.2.3)
Como estes tópicos, para cada um dos grupos citados anteriormente, se relacionam com a performance e a contaminação tóxica, que definimos como sendo os efeitos negativos do comportamento tóxico?	• Proposta de métricas de desempenho e contaminação; • Proposta de análise de agregações de grupos de jogadores, por tópico, de acordo com as métricas criadas. • Análise de grupos com tópicos positivos e negativos, utilizando as agregações propostas. • Análise de grupos não ofensores, a partir dos tópicos negativos utilizados pelos respectivos ofensores, utilizando as agregações propostas.	5.3
Existem relações temporais entre estes tópicos, ou seja, regras que determinem como as conversas em uma partida se desdobram ao longo do tempo?	• Divisão das partidas em intervalos de tempo; • Proposta de uso de regras de associação entre tópicos presentes em divisões consecutivas para descobrir transições entre tópicos; • Análise das regras de associação por grupos de jogadores.	5.4
Como cada um dos tópicos de conversa descobertos associa-se com emoções derivadas de um modelo de emoções?	• Construção automática de um léxico de emoções específico ao domínio de MOBAs; • Descoberta das palavras mais relevantes de cada tópico; • Aplicação do léxico sobre estas palavras mais relevantes em cada tópico.	5.5

aplicamos de técnicas de agrupamento de textos e verificamos a existência de 10 agrupamentos de conversas em partidas de MOBAs. Estes agrupamentos foram interpretados e refinados em 7 tópicos de conversação comumente utilizados por grupos de jogadores em MOBAs, que foram validados por jogadores veteranos a partir de um instrumento de avaliação criado para tal fim.

Depois disso, associamos grupos de jogadores com os tópicos descobertos, para tornar possível a análise do comportamento destes grupos, de acordo com os tópicos utilizados neles. Isso permite que explicitemos as diferenças no uso dos tópicos por diferentes grupos de jogadores, e o que elas significam no contexto do jogo. Para responder "**como cada um dos tópicos descobertos se associa com diferentes tipos de jogadores, divididos em grupos de acordo com sua associação ao jogador tóxico (jogadores tóxicos, seus aliados, e seus inimigos)**", dividimos os grupos em quatro categorias diferentes, rotulando cada grupo de acordo com o tópico de conversa mais prevalente neste. Então, estudamos a distribuição de uso dos tópicos em cada uma destas categorias de grupos.

Com os tópicos associados a grupos de jogadores, podemos avaliar como o uso destes tópicos por diferentes grupos de jogadores impactam as partidas. Para tal, é necessário criarmos métricas, que medirão os níveis de desempenho e contaminação tóxica dos grupos de jogadores. Com tais métricas, vemos como cada tópico relaciona-se com os grupos de jogadores em detalhe, através da verificação de como a taxa de uso de cada tópico pelos grupos de jogadores se modifica com a variação das métricas.

Para descobrir "**como estes tópicos, para cada um dos grupos citados anteriormente, se relacionam com a performance e a contaminação tóxica, que definimos como sendo os efeitos negativos do comportamento tóxico**", criamos métricas para representar o desempenho e a contaminação de grupos de jogadores em uma partida, e então agregamos os grupos de jogadores de acordo com as métricas criadas e os seus respectivos tópicos associados. Com estas agregações, estudamos as taxas de uso dos tópicos em grupos que apresentam prevalência de conversas positivas e em grupos que apresentam prevalência de conversas negativas, em diferentes níveis de desempenho e contaminação. Também analisamos as taxas de uso dos tópicos de jogadores não ofensores, em frente a prevalência de conversas negativas por parte do jogador ofensor.

Kwak and Blackburn (2014) levantam a hipótese de que ofensores iniciam uma partida com um comportamento não tóxico, e em algum momento mudam seu comportamento para tóxico, isto é, que jogadores podem mudar de comportamento ao longo uma partida. Neste trabalho buscamos confirmar esta possibilidade, e através da análise das transições entre tópicos em diferentes momentos da partida, mostraremos se estas ocorrem regularmente. Caso positivo,

detalharemos quais transições entre tópicos são mais frequentes e relevantes. Para descobrir se "**existem relações temporais entre estes tópicos, ou seja, regras que determinem como as conversas em uma partida se desdobram ao longo do tempo**", dividimos as conversas de cada grupo em intervalos de tempo, reaplicamos o processo de descoberta e interpretação de tópicos nas conversas divididas, e então descobrimos e analisamos regras de associação entre os tópicos utilizados nestes intervalos.

O vocabulário usado por jogadores em MOBAs é bastante único, com várias expressões e abreviações que só fazem sentido para quem está incluído nas comunidades dos jogos. Por exemplo, nenhuma das das 4 palavras mais frequentes em nosso corpus (lol, mia, mid, gg) está presente no léxico de emoções NRC. Isso dificulta o uso de léxicos de uso geral para analisar emoções em MOBAs, visto que muitas das palavras mais frequentemente usadas nestes jogos estão presentes em tais léxicos. Possuir as emoções de tais palavras nos ajudará a definir quais categorias de emoções são prevalentes em quais tópicos, nos dando uma melhor descrição de cada um deles. Para responder "**como cada um dos tópicos de conversa descobertos associa-se com emoções derivadas de um modelo de emoções**", construímos um léxico de emoções especifico ao domínio de MOBAs, utilizando como base um léxico pre-existente e um modelo de aprendizado treinado com palavras deste léxico, e então descobrimos as palavras mais relevantes para cada tópico de conversação e desenvolvemos um método para aplicar o dicionário resultante as palavras mais relevantes de cada tópico de conversação.

O restante deste capítulo está estruturado como segue. Na Seção 5.2, discutiremos em detalhes como a descoberta, interpretação dos tópicos de conversa foram feitas, bem como a associação destes com grupos de jogadores. Na Seção 5.3, discutiremos as métricas de desempenho e contaminação propostas, e como foi feita a associação destas métricas com grupos de jogadores e por consequência, tópicos de conversação. Na Seção 5.4, falaremos sobre a descoberta de regras de associações relevantes entre tópicos, e como foi feita a análise destas regras. Na Seção 5.5, falaremos sobre a construção do léxico de emoções específico ao domínio de MOBAs, bem como sobre a aplicação deste léxico sobre nossos tópicos de conversa.

5.2 Tópicos de conversa em MOBAs

Nesta seção descreveremos o passo-a-passo feito para a obtenção de agrupamentos, e então o processo de interpretação destes agrupamentos em tópicos de conversação de jogadores em partidas. A Figura 5.1 (em inglês) descreve o processo de descoberta de tópicos adotado neste trabalho.

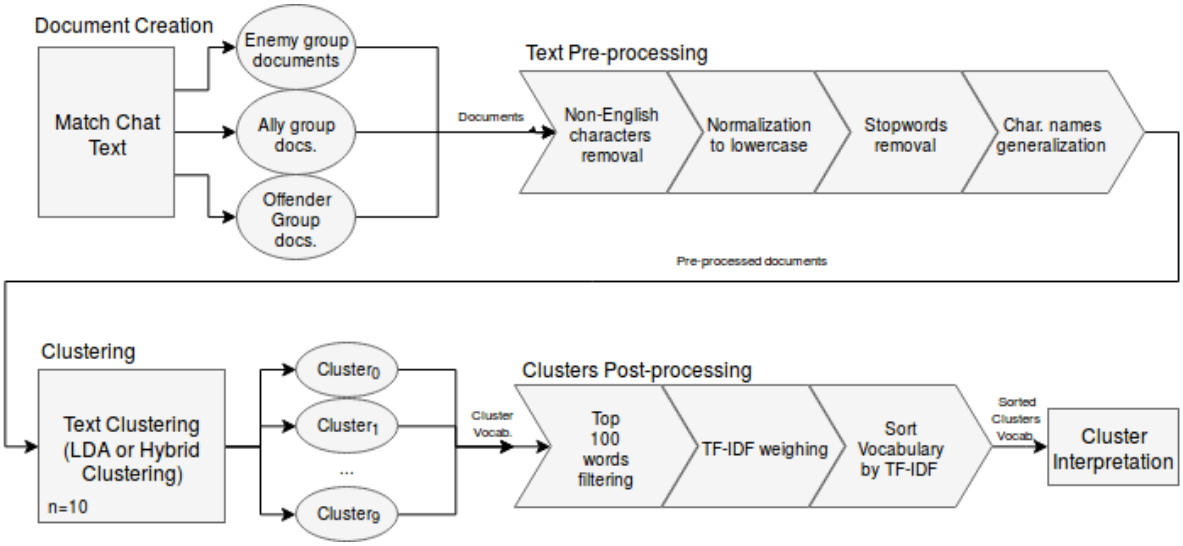


Figura 5.1: Processo de agrupamento de texto adotado.

5.2.1 Descoberta de padrões de conversa

O primeiro passo no sentido de descobrir agrupamentos de bate-papos é criar documentos a partir do corpus de bate-papos para representar os três grupos de jogadores: aliados, inimigos e ofensores. Essa separação permite: a) a análise dos efeitos da contaminação em ambas equipes; b) isolar o comportamento dos ofensores e c) distinguir os efeitos da contaminação de sua fonte tóxica. Para cada partida, três documentos foram criados: um contendo o diálogo entre os quatro jogadores do grupo aliado, outro contendo o diálogo dos 5 jogadores do grupo inimigo e finalmente outro contendo o texto escrito pelo ofensor. Estes documentos foram criados juntando textos do bate-papo de equipe com textos do bate-papo global, englobando todas as conversas que têm como origem os jogadores daquele grupo, visto que o sujeito de nossa análise, primariamente, é o grupo em si e não a destinação das mensagens enviadas por este.

Todos documentos são pré-processados como descrito na Figura 5.1 (em inglês), com o objetivo de melhorar os resultados do nosso agrupamento. Os passos descritos são os seguintes:

1. Normalização para caracteres minúsculos;
2. Remoção de *stop words* inglesas disponíveis no pacote NLTK[1] (BIRD; KLEIN; LOPER,), juntamente de palavras com uma frequência menor do que 800 no corpus, com o objetivo de remover palavras pouco relevantes e conseguir um vocabulário compacto o suficiente (aproximadamente 20.000 palavras) para ser processado em tempo hábil;

[1] http://www.nltk.org/

3. Remoção de caracteres não pertencentes à língua inglesa;
4. Generalização dos nomes dos personagens para um símbolo.

Também foram descartados documentos vazios, que representam grupos onde os jogadores não interagiram textualmente, que correspondem a 2.7% dos documentos.

Para escolher um algoritmo de agrupamento, testamos dois métodos sobre uma amostra aleatória de 30.000 partidas. Os métodos testados foram o LDA e o k-means, descritos na Seção 3.2. Os documentos de entrada para o LDA foram representados em uma matriz de contagem, já que este modelo é a única entrada válida para o método. Os parâmetros utilizados na execução do LDA, com exceção do número de tópicos a serem descobertos, bem como a implementação utilizada foram fornecidos pela API gensim[2] (ŘEHŮŘEK; SOJKA, 2010). Os documentos de entrada para o k-means foram representados a partir de *embbedings* de documentos geradas pelo *ParagraphVector*, treinado com todos os documentos presentes nos dados, devido à baixa dimensionalidade do modelo, o que nos permite uma execução do k-means em tempo hábil, e melhores resultados na agregação. A implementação utilizada para a execução do k-means, bem como os parâmetros utilizados, com exceção do número de agrupamentos, vieram da API scikit-learn[3] (PEDREGOSA et al., 2011).

Para selecionar o número adequado de agrupamentos para ambos algoritmos, levantamos a hipótese da existência de alguns tópicos, que observamos em uma análise preliminar do corpus: conversação positiva, coordenação tática, reclamações, insultos e línguas estrangeiras. Ambos algoritmos foram parametrizados para gerarem dez (10) agrupamentos, com o objetivo de cobrir estes tópicos e também dar margem para possíveis novos tópicos. Agrupamentos que sejam similares a outros, podem ser mesclados manualmente. A execução do k-means e do LDA produziram agrupamentos com vocabulários similares. Baseado nestes testes, escolhemos o LDA, já que a sua execução consome menos recursos de memória, e por isso, apesar do k-means apresentar uma complexidade temporal menor do que o LDA, a execução do LDA acabou sendo consideravelmente mais rápida no equipamento disponível para este experimento.

O passo de pós-processamento busca preparar os resultados do LDA para a interpretação dos agrupamentos resultantes. Cada documento foi atribuído ao tópico do LDA que melhor lhe representa.

A interpretação do significado de cada agrupamento deve ser baseada em suas palavras mais representativas. O LDA provê uma lista de palavras de um tópico ordenada por significância. Contudo, as palavras mais significantes destas listas não se aglutinavam em algum

[2]https://radimrehurek.com/gensim/
[3]http://scikit-learn.org/

assunto ou tópico de conversação específico, não permitindo identificar o contexto de cada tópico. Como alternativa, para distinguir entre palavras que são frequentes em todos os agrupamentos, e palavras que são frequentes em um agrupamento em específico, utilizamos a fórmula do TF-IDF, com a frequência do termo sendo relativa aos termos no agrupamento, e o inverso da frequência dos documentos sendo relativa a todos os documentos (AGGARWAL; ZHAI, 2012b). Foram utilizadas as 100 palavras mais representativas de acordo com o resultado do TF-IDF para a interpretação dos agrupamentos resultantes.

5.2.2 Interpretação dos tópicos

Uma vez que todos os grupos de jogadores estejam identificados por um agrupamento, uma interpretação preliminar destes foi feita pelo autor, um jogador experiente de LoL, com mais de 500 horas de jogo no servidor norte-americano (NA). Para a análise de cada agrupamento, combinamos as 10 palavras com maior peso com uma nuvem de palavras feita com as 100 palavras de maior peso, descobertas usando o processo descrito na Seção 5.2.1.

Essa interpretação preliminar foi então refinada por mais 5 jogadores, com experiência de LoL variando entre 50 e mais de 300 horas de jogo, também no servidor NA. Elaboramos um conjunto de instruções, contidas em um formulário[4] e validadas previamente por um outro jogador experiente em LoL. Estas instruções foram usadas como guia para estes jogadores realizarem suas respectivas interpretações dos agrupamentos. Parte do formulário original, contendo as instruções e as palavras relevantes para um dos agrupamentos é mostrado no Apêndice A.

A interpretação dos tópicos foi feita de acordo com os passos a seguir.

1. Foi pedido a cada jogador que interpretasse o comportamento de um agrupamento baseado nas 10 palavras mais relevantes e na nuvem das 100 palavras mais relevantes, e descrevesse este comportamento através de uma breve descrição.

2. Então as descrições providas por cada jogador foram debatidas individualmente com os mesmos, com o objetivo de conseguir descrições mais claras. Como parte deste debate, foram mostradas as descrições providas pelo outros quatro jogadores. Ao fim deste debate, várias das interpretações foram fornecidas em termos mais claros, não havendo nenhum tipo de mudança no significado destas.

3. Finalmente, mostramos a interpretação original feita pelo autor, e foi perguntado individualmente aos jogadores se eles concordam ou discordam com essa interpretação.

[4] https://goo.gl/forms/4TxxiLiTXFFtp5Eh2

Como resultado deste processo de interpretação, descartamos agrupamentos que não revelaram nenhum padrão do comportamento, e juntamos agrupamentos distintos que representavam o mesmo padrão. Este padrões de conversação foram denominados como *tópicos de conversa*. Quaisquer desacordos existentes no processo de interpretação foram resolvidos através de votações com os cinco jogadores, prevalecendo a interpretação da maioria. Também procuramos por generalizações entre os tópicos, separando-os em duas categorias, positivo e negativo, bem como em outras subcategorias. Os agrupamentos resultantes estão descritos na Seção 6.1.

5.2.3 Análise dos jogadores por tópicos

Após a interpretação e fusão dos agrupamentos, analisamos a distribuição dos tópicos resultantes de acordo com todos os tipos de grupos, mostrando as diferenças entre uso de cada tópico e valores de desempenho/contaminação. Quatro tipos de grupos foram considerados: ofensores, aliados, inimigos contaminados e inimigos não-contaminados.

A divisão dos inimigos em dois tipos de grupo foi motivada pela percepção de que poucos grupos inimigos eram afetados pelo comportamento do ofensor. Assim, assumimos que grupos inimigos que não fizeram nenhuma denúncia são não-contaminados, com o restante sendo contaminados. Fizemos a hipótese de que todos os aliados, independente da existência de denúncias, sofrem de algum nível de contaminação por estarem em contato direto com o ofensor. Assim, consideramos que tal distinção não faria sentido para os aliados. Os resultados são mostrados na Seção 6.2.

5.3 Análise dos Efeitos do Comportamento Tóxico

Nesta seção são descritas as motivações e os detalhes das métricas de contaminação e desempenho propostas neste trabalho. Também verificamos quão adequada cada métrica de desempenho é para representar jogadores individuais e grupos de jogadores. Finalmente, para analisar a variação das métricas de contaminação e desempenho nos tópicos, agregamos grupos de jogadores de acordo com estas métricas.

5.3.1 Métricas de Desempenho e Contaminação

Podemos propor uma métrica de *contaminação* a partir da definição de comportamento tóxico dada pela Riot Games: "qualquer comportamento que impacta negativamente na experiência de jogo de outros jogadores" (SHORES et al., 2014). Inspirada por trabalho relacionando (BLACKBURN; KWAK, 2014), a métrica assume que se a experiência de um jogador é negativa o suficiente para motivá-lo a registrar uma denúncia contra o jogador tóxico, então ele deve ser considerado contaminado. Esta métrica mede a contaminação média somente de grupos de jogadores, já que os dados do tribunal fornecem somente o número total de denúncias por aliados e inimigos.

A métrica de contaminação proposta é definida na Equação 5.1. Para um dado grupo de jogadores $g \in \{ally, enemy\}$ em uma partida m, $num.reports_{g,m}$ é o número de denúncias registradas por aquele grupo, e n é o número de jogadores naquele grupo. A contaminação varia entre 0 e 1, onde 1 é quando todos os jogadores do grupo registraram denúncia. O ofensor não está associado com um valor de contaminação, já que ele é a própria fonte da contaminação.

$$contamination_{g,m} = \frac{num.reports_{g,m}}{n_{g,m}} \tag{5.1}$$

A métrica de *desempenho* é centrada na ideia de descobrir qual a contribuição individual de cada jogador para o resultado da partida, para então agrupar estas contribuições através da média dos desempenhos de todos os jogadores de um grupo. Deste modo, tanto o desempenho quanto a contaminação podem ser medidos pela mesma unidade, grupos de jogadores. O desempenho individual combina a tradicional métrica KDA (do inglês: *Kills-Deaths-Assists*) (BLACKBURN; KWAK, 2014) com a quantidade de ouro ganha durante uma partida, já que o ouro funciona como um medidor de várias pequenas contribuições para o resultado de uma partida que não são cobertas pelo KDA.

O desempenho cumulativo de grupos de jogadores de diferentes tamanhos é calculado a partir da soma das performances dos jogadores no grupo, soma essa que é então normalizada usando o número de jogadores naquele grupo, como descrito na Equação 5.2, onde $P_{m,g}$ é o conjunto de todos os jogadores p em um grupo $g \in \{ally, enemy, ofender\}$ relacionado a uma partida m.

$$desempenho_{g,m} = \frac{\sum_{p \in P_{m,g}} desempenho_{p,m}}{|P_{m,g}|} \tag{5.2}$$

O desempenho de um jogador p em uma partida m é dado pela Equação 5.3, que repre-

senta a média das respectivas porcentagens de KDA e ouro. As equações 5.4 e 5.5 descrevem como estas porcentagens são calculadas por jogador p em uma partida m, onde P_m é o conjunto de todos os jogadores pm na partida m.

$$desempenho_{p,m} = \frac{\%gold_{p,m} + \%KDA_{p,m}}{2} \tag{5.3}$$

$$\%KDA_{p,m} = \frac{KDA_{p,m}}{\sum_{pm \in P_m} KDA_{pm,m}} \tag{5.4}$$

$$\%gold_{p,m} = \frac{gold_{p,m}}{\sum_{pm \in P_m} gold_{pm,m}} \tag{5.5}$$

Para confirmar a adequação das métricas de desempenho individual e de grupo, estabelecemos a hipótese de que as mesmas devem apresentar valores altos para jogadores e grupos vencedores, valores baixos para jogadores e grupos perdedores, e serem igualmente distribuídas entre vencedores e perdedores em valores intermediários. Como consequência, os jogadores vencedores devem apresentar um desempenho médio consideravelmente maior do que jogadores perdedores.

Para validar esta hipótese, e confirmar a adequabilidade das métricas de desempenho, investigamos suas distribuições, dividindo os grupos e os jogadores entre vencedores e perdedores. Os resultados são mostrados na Tabela 5.2. Percebe-se que o desempenho médio de jogadores individuais e de grupos de jogadores são similares. Utilizando uma série de testes-t com 95% de confiança, observamos que há uma diferença significativa entre o desempenho de vencedores (0.129) e o de perdedores (0.071). Também há uma diferença positiva significativa entre o desempenho dos vencedores e o desempenho médio global (0.10). Tal diferença significativa também é encontrada, de maneira negativa, entre a média dos perdedores e o desempenho médio.

A Tabela 5.2 também mostra que grupos e jogadores na área de desempenho baixo (25% menores desempenhos) são, em ambas as métricas, quase totalmente jogadores perdedores (>90%). Grupos e jogadores nos 50% de desempenho centrais (*IQR - Inter-Quartile Range*) são distribuídos balanceadamente entre vencedores e perdedores, também para ambas métricas. Finalmente, grupos e jogadores na área de alto desempenho (25% maiores desempenhos) são vencedores na quase totalidade. Assim, confirmamos que estas métricas são adequadas para medir o desempenho de jogadores e de grupos de jogadores.

Tabela 5.2: Avaliação das métricas de desempenho.

Desempenhos	M	Ven. M	Per. M	Ven. Q Inferior	IQR Ven.	Ven. Q Superior
Player Perf.	0.100	0.129	0.071	6%	49%	92%
Group Perf.	0.100	0.129	0.071	$> 1\%$	50%	$< 99\%$

M: Média, Ven.:Vencedores, Per.:Perdedores, Q:Quartil, IQR:Distância inter-quartil

5.3.2 Análise dos efeitos de tópicos positivos e negativos

Dada a associação de cada grupo com um tópico, e o cálculo dos respectivos desempenho e contaminação, investigamos a relação entre a taxa de uso de cada tópico descoberto (mais generalizações subsequentes) e seus efeitos em cada tipo de grupo. Para investigar tais relações, precisamos estudar a taxa de uso de cada tópico em diferentes contextos para as métricas de desempenho e contaminação. Para criar estes contexto, criamos agregações de grupos.

Considere n o total de partidas em nossos dados, t um tipo de grupo, e um conjunto de grupos $g_1, g_2, ..., g_n \in G_t$ do mesmo tipo t, rotulados com seu tópico correspondente, e ordenados por algum critério (i.e. desempenho ou contaminação).

Cada agregação representa um subconjunto de s grupos consecutivos em G_t. Existe um total de $l = \lceil n/s \rceil$ agregações. Neste trabalho, consideramos $s = 10.000$. Uma agregação a_i, para $i \in [0, l-1]$ é representada por uma tupla $a_i =< G_{t,a_i}; m_{a_i}; \{u_1, \cdots, u_x\} >$, onde:

- $G_{t,a_i} = \{g_{i*s+1}, \cdots, g_{i*s+s}\}$ é o subconjunto de G_t que contem os grupos associados a agregação a_i;
- m_{a_i} é a média aritmética do critério escolhido para ordenar G_t, aplicada aos grupos em G_{t,a_i};
- $\{u_1, \cdots, u_x\}$ é o conjunto das taxas de uso de cada um dos x tópicos existentes, com um u_j sendo a taxa de uso de um tópico $j (j \leq x)$ qualquer, considerados todos os grupos em G_{t,a_i}.

Primeiramente, analisamos o uso de tópicos para cada tipo de grupo, e a suas relações com desempenho e contaminação. Para o desempenho, detalhamos o comportamento de todos os quatro grupos. Assim, organizamos os grupos de acordo com seus valores de desempenho. Isso resultou em 197 agregações de ofensores, 197 agregações de aliados, 132 agregações de inimigos não-contaminados, e 65 agregações de inimigos contaminados.

Para a contaminação, consideramos somente aliados e inimigos, uma vez que os ofensores não estão associados a um valor de contaminação. Os inimigos não foram divididos em contaminados e não-contaminados, porque não faz sentido analisar grupos onde a contaminação é sempre constante (ex. inimigos não-contaminados). Assim, organizamos estes dois grupos

de jogadores de acordo com seus níveis de contaminação, e usando o mesmo tamanho de agregação, produzimos 197 agregações para cada tipo de grupo. Então, analisamos a relação entre cada tópico e desempenho/contaminação, usando as agregações criadas. Resultados da análise destas agregações são discutidos nas Seções 6.3.1 e 6.3.2.

5.3.3 Análise dos efeitos do comportamento tóxico sobre os demais jogadores

Finalmente, analisamos as relações entre o comportamento tóxico do ofensor, representado pelo seu uso de tópicos, e seus efeitos sobre os grupos na mesma partida. Assim, cada grupo não-ofensor (i.e. aliados e inimigos) foi associado com o respectivo ofensor de sua partida. Criamos agregações de maneira similar às criadas previamente, mas $g_1, g_2, ..., g_n \in G_t$ de um certo tipo de grupo estão relacionados com seus respectivos ofensores. Para cada agregação a_i, m_{a_i} ainda representa a média aritmética do desempenho ou contaminação dos grupos em G_{t,a_i}, mas $u_1, ..., u_x$ são as taxas de uso u_j de um tópico qualquer j ($j \leq x$) que foi usado pelos respectivos ofensores naquela agregação.

Novamente utilizando $s = 10.000$, produzimos agregações baseadas em dois critérios de ordenação: desempenho e contaminação. Para o desempenho, isso resultou em 197 agregações para aliados, 132 para inimigos não-contaminados e 65 para inimigos contaminados. Usando a contaminação como critério, temos 197 agregações para aliados e 197 para inimigos. Resultados da análise destas agregações são reportados na Seção 6.3.3.

Utilizamos testes de correlação (τ de Kendall) em nossas análises para mostrar a relação entre o uso de tópicos e o desempenho/contaminação. Consideramos que a correlação entre duas variáveis é forte se ela for maior do que 0.5 (correlação positiva), ou menor do que -0.5 (correlação negativa). Testes-t com 95% de intervalo de confiança foram usados para mostrar as diferenças entre valores de média, quando aplicáveis. Todas as médias comparadas mostraram significância estatística de $p < 2.2 * 10^{-16}$, exceto quando dito o contrário.

5.4 Transições comuns de tópicos ao longo de partidas

Para descobrir relações entre tópicos usados em momentos distintos de uma partida, precisamos dividir as partidas em vários momentos diferentes, e então descobrir quais tópicos de conversação são usados durante estes momentos.

Kwak and Blackburn (2014) dividem temporalmente seu corpus em uma quantidade

igual de pedaços, com alta granularidade (100 pedaços). Contudo, tal granularidade em nosso problema resultaria em uma grande quantidade de documentos vazios, já que ao contrário do trabalho citado, estamos analisando as partidas individualmente ao invés do corpus como um todo, somente dividido por grupos de jogadores. Então, precisamos procurar outra maneira de dividir nossas partidas.

A partir de uma análise das partidas em nossos dados, descobrimos que a duração média de uma partida de League of Legends é de aproximadamente 30 minutos ($media = 33min., SD(DesvioPadrao) = 10min.$), com a duração da maior parte das partidas concentradas no intervalo de 20 a 60 minutos, como mostrado no histograma na Figura 5.2.

Comforme discutimos na Seção 2.2, sabemos que uma partida de League of Legends é dividida em três momentos distintos: *early*, *mid* e *end*. Discutimos na Seção 4.3 que as variações nos volumes de conversas durante uma partida correspondem aos seus três momentos (KWAK; BLACKBURN, 2014), com picos que acontecem durante os $\approx$ 10 minutos iniciais (*early*) e os $\approx$ 10 minutos finais (*end*) de uma partida. Já volume de conversa no intervalo de tempo entre estes picos (*mid*), apresenta pouca variação.

A partir do valor da duração média de uma partida, e considerando que os momentos *early* e *end* têm duração similar, escolhemos o tamanho de nossas *divisões* como sendo de 10 minutos, de modo a representar cada um dos momentos de uma partida. Para uma partida qualquer, temos a primeira divisão representando o *early*, e para representar o *late* a última, ou as duas últimas divisões, no caso de uma partida que não seja múltiplo de 10 (e.g. 42 minutos). Todas as divisões entre aquelas representando o *early* e o *late* representam o *mid* do jogo. Em partidas com duração inferior a 20 minutos, temos divisões representando somente o *early* e o *late*, mas note que estas partidas são bastante raras, e por esta razão são consideradas anomalias.

Para evitar possíveis *outliers*, removemos divisões representando conversas que ocorreram em períodos de tempo acima de 2 desvios padrões acima da média de duração de uma partida (53 minutos). Além de serem prováveis *outliers*, estas divisões apresentam um nível de documentos vazios muito acima do normal, o que poderia se provar um problema para nossa análise. Como cada divisão representa 10 minutos de partida, arredondamos este valor de corte para 60 minutos. As divisões correspondentes a durações acima de 60 minutos correspondem a menos de 5% do total. Também foram removidas divisões correspondentes a partidas com menos de 10 minutos de duração, o que corresponde a dois desvios padrões abaixo da média (13 minutos). Contudo, a quantidade de divisões removidas deste modo é desprezível.

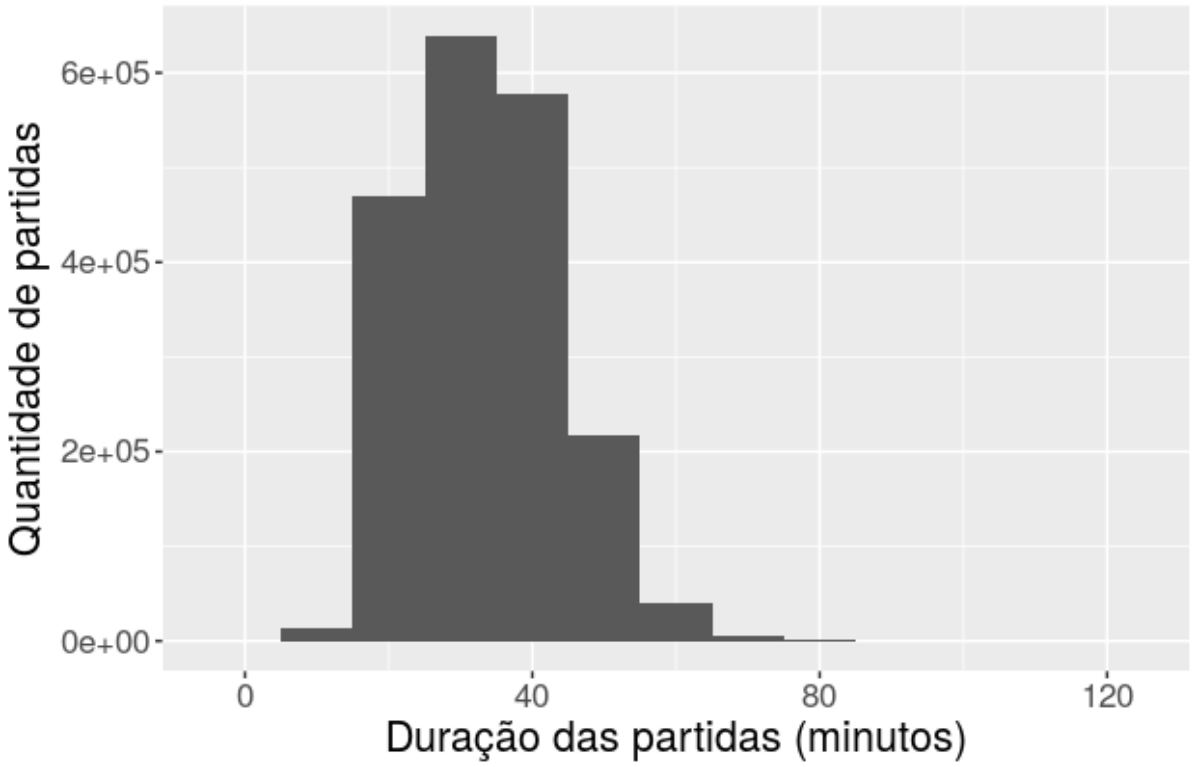


Figura 5.2: Histograma da duração média das partidas (intervalos de 10 minutos).

Tabela 5.3: Proporção de palavras similares entre os tópicos da Seção 5.2.1, e os tópicos descobertos nesta seção

Táticas	**Táticas/Edu.**	**Bate-papo**	**Reclamações**	**Discussões**	**Insultos**	**Provocações**	**Outras Línguas**
0.89	0.92	0.87	0.88	0.87	0.94	0.78	0.76

Repetindo o método de análise descrito na Seção 5.2, separamos cada uma das divisões restantes em três documentos, representando cada um dos três tipos de grupos de jogadores (aliados, inimigos e ofensores). Alguns destes documentos não apresentavam nenhum conteúdo textual, indicando que os jogadores não se comunicaram durante aquela fatia de tempo. Nós rotulamos estes documentos como 'vazio' e não os incluímos no processo de agrupamento para descoberta de tópicos descritos abaixo.

Pré-processamos estes documentos, e os usamos para treinar um modelo de descoberta de tópicos (LDA), utilizando o mesmo processo descrito na Seção 5.2.1, exceto que utilizamos 15 agrupamentos ao invés de 10, com a intenção de descobrir se algum tópico novo, específico a estes documentos, apareceria.

O processo de interpretação dos tópicos foi análogo ao da Seção 5.2.2, seguido da junção dos agrupamentos quando os tópicos fossem similares. Isso resultou em tópicos com as 100 palavras mais relevantes dos tópicos descobertos são bastante similares as palavras dos tópicos do experimento da Seção 5.2.1. A proporção de palavras relevantes similares entre os tópicos descobertos nos dois experimentos está descrita na Tabela 5.3.

Finalmente, juntamos os documentos rotulados por um tópico de conversação com aqueles rotulados com 'vazio' para compor todos os tópicos discutidos pelos grupos ao longo das partidas. Ignorar estes documentos implicaria em partidas contendo documentos sem nenhuma rotulação, dificultando a análise das transições. Um exemplo é mostrado na Tabela 5.4, já separados pelos diferentes grupos de jogadores.

Tabela 5.4: Exemplos das divisões de partidas por tempo para 3 partidas e 9 grupos.

Partida	**Grupo**	**Div. 1**	**Div. 2**	**Div. 3**	**Div. 4**	**Div. 5**	**Div. 6**	**Duração (min.)**
1	Aliados	Táticas	Discussões	Táticas	Discussões	-	-	33
1	Inimigos	Táticas	Outras Línguas	Reclamações	Reclamações	-	-	33
1	Ofensor	Provocações	Provocações	Provocações	Insultos	-	-	33
2	Aliados	Vazio	Táticas/Edu.	Insultos	Insultos	Vazio	-	43
2	Inimigos	Bate-papo	Insultos	Bate-papo	Bate-papo	Vazio	-	43
2	Ofensor	Táticas	Insultos	Insultos	Bate-papo	Vazio	-	43
3	Aliados	Táticas/Edu.	Insultos	Insultos	Insultos	-	-	38
3	Inimigos	Outras Línguas	Outras Línguas	Outras Línguas	Vazio	-	-	38
3	Ofensor	Insultos	Insultos	Provocações	Provocações	-	-	38

Para descobrir padrões de transições temporais entre tópicos, criamos uma tabela de duas colunas a partir de nossos documentos, representando as relações temporais entre ocorridas durante as partidas. O tópico na coluna esquerda (E) é sempre imediatamente anterior ao da coluna direita (D), para todas as linhas da tabela. Então um grupo com os tópicos (t_0, t_1, t_2), ordenados cronologicamente, irá resultar nas transações (t_0, t_1) e (t_1, t_2). Chamamos de *relação*

temporal uma relação entre dois tópicos, t_i e t_{i+1}, associados a documentos consecutivos e distintos de uma mesma partida, aonde t_i é o tópico imediatamente anterior em ordem cronológica a t_{i+1}.

Esta tabela contendo as transações é usada de entrada para o algoritmo apriori, explicado na Seção 3.3, para buscar quais transições entre tópicos são mais relevantes. Sabemos que o apriori não leva em consideração relações temporais, ou seja, as regras descobertas pelo método não necessariamente representam transições da coluna E a coluna D. Resolvemos este problema eliminando quaisquer regras que quebrem essa relação temporal, ou seja, regras representando transições de D a E ($D \rightarrow E$). Ou seja, nossa análise se concentra em saber quais tópicos precedem o uso de outros tópicos.

O valor de suporte escolhido busca englobar a maior quantidade de regras possível, excluindo somente regras extremamente infrequentes. O desbalanceamento entre a frequência de nossos tópicos nos documentos, nos obriga a escolher um valor de suporte baixo, para englobar regras de tópicos infrequentes. Isso implica que o suporte servirá somente como um filtro grosseiro, e não irá ser uma boa métrica de avaliação para as regras.

Experimentalmente, valores de confiança muito altos se demonstraram proibitivos, por deixarem poucas regras passarem pelo seu filtro. Nosso valor de confiança foi definido experimentalmente, sendo o valor mais alto possível para nos dar uma quantidade razoável de regras, que consideramos como sendo $\approx$10 regras, e que estas regras envolvessem a maior parte dos tópicos.

Consideramos um *lift* mínimo de 1.3/0.7 como distante de um (1) o suficiente para representar regras com uma dependência suficientemente alta entre seus itens Usaremos o *lift* e a confiança para determinar com mais precisão a qualidade de nossas regras.

Seguindo o método de análise já definido na Seção 5.3, em nossos experimentos analisamos as transições por tipo de grupo de jogadores, avaliando o comportamento de aliados, inimigos e ofensores separadamente. Essa separação nos permitirá observar as diferenças, se existentes, entre as transições nestes tipos de grupos, e explicitar como grupos de jogadores diferentes interagem com os tópicos de conversação presentes de maneiras diferentes. Também removemos quaisquer regras envolvendo o tópico de Outras Línguas, por motivos similares aos apresentados na Seção 5.2.1. Os experimentos desenvolvidos e as regras resultantes são discutidos na Seção 6.4.

5.5 Análise das emoções presentes nos tópicos

5.5.1 Construção do léxico de emoções

O vocabulário usado por jogadores em MOBAs é bastante único, com várias expressões e abreviações que só fazem sentido a quem está incluído nas comunidades dos jogos. Por exemplo, nenhuma das das 4 palavras mais frequentes em nosso corpus (lol, mia, mid, gg) está presente no léxico de emoções NRC. Isso dificulta o uso de léxicos de uso geral para analisar emoções em MOBAs, visto que estas palavras específicas não estão presentes em tais léxicos. Para resolver este problema, geramos um léxico específico ao domínio de MOBAs.

Dentre as abordagens para construção automática de léxicos apresentadas na Seção 3.5, escolhemos a baseada em aprendizado de máquina, já que devido a linguagem utilizadas nos chats de MOBAs serem bastante únicas ao contexto, seria complicado realizar uma expansão a partir de palavras sementes. Aqui vamos explicar como, através de algoritmos de aprendizagem, construímos um léxico de emoções específico ao domínio de MOBAs, usando como base o léxico de uso geral NRC. O léxico especifico ao domínio resultante possibilita o melhor entendimento do comportamento de jogadores em jogos on-line, e nos é útil para caracterizar as emoções envolvidas em cada tópico, de forma que possamos descrevê-los também em termos emocionais.

Seguindo a proposta de Bravo-Marquez et al. (2016), a construção do léxico novo se deu através seguintes passos:

1. Gerar um modelo de vetores de palavras utilizando o corpus de contexto específico;
2. Selecionar as palavras para a composição do corpus de treinamento;
3. Treinar um modelo de classificação utilizando as palavras selecionadas;
4. Descobrir os rótulos das demais palavras, não pertencentes ao NRC, utilizando o modelo treinado.

Passo 1: O modelo vetorial das palavras foi preparado a partir do corpus das conversas dos jogadores. O pré-processamento deste corpus foi feito de maneira similar à relatada na Seção 5.2.1. A única diferença foi que a remoção de caracteres não-ingleses foi adaptada para não remover *emoticons*, já que estes representam potenciais palavras de emoções.

A partir do corpus pré-processado, foi construído um modelo de vetores de palavras usando o algoritmo word2vec. Além deste algoritmo ter sido aquele usado por Bravo-Marquez et al. (2016), o mesmo demostrou resultados superiores em testes preliminares contra o algoritmo

Tabela 5.5: Palavras removidas do NRC

Palavra Removida	**Significado Original**	**Significado dentro do jogo**
baron	Titulo de nobreza	*Minion* que provê vantagens para jogadores
drag	Arrasto	*Minion* que provê vantagens para jogadores
blue	Cor Azul	*Minion* que provê vantagens para jogadores
red	Cor Vermelha	*Minion* que provê vantagens para jogadores
tower	Torre	Construção dentro do jogo
farm	Fazenda	Acumular Ouro
feed	Alimentar	Morrer constantemente para adversários
lane	Faixa de tráfego	Um dos corredores onde os jogadores se posicionam
mid	Meio; Entre	Denominação ao jogador que se posiciona no corredor do meio
top	Topo	Denominação ao jogador que se posiciona no corredor do topo
tank	Tanque de Guerra	Personagem com mais defesa do que a média
ill	Doente	Escrita abreviada de 'I will' (eu vou)
blitz	Ataque rápido	Abreviado de *Blitzcrank*, um dos personagens do jogo
report	Relatório	Denúncia por comportamento tóxico

GLoVe. Foram testadas as variações *skipgram* e *continuous bag of words*, cada uma em conjunto com 100, 200 e 300 dimensões para os tamanhos dos vetores de palavras gerados pelo modelo, bem como 5, 10 e 15 como valores para o número de iterações realizadas pelo algoritmo. Modelos treinados com *skipgrams* gerando vetores de 300 dimensões renderam melhores resultados, e o corpus aparenta convergir com 10 iterações, já que não há mudanças significativas de resultados entre 10 e 15 iterações. Os resultados mais significativos dos experimentos com diferentes variações estão disponíveis no Apêndice D.

Passo 2: O conjunto de treinamento foi composto pela representação vetorial das palavras presentes no NRC, adquiridas com o mesmo modelo *word2vec* treinado no corpus. Isso é fundamental para nosso modelo, visto que *embbedings* guardam informações do contexto das palavras, e as palavras presentes no NRC precisam ser representadas em um contexto de MOBAs, que é provido pelo modelo treinado no corpus de bate-papo de partidas de MOBAs. Por isto utilizamos somente as palavras do NRC que estão presentes no nosso corpus, em um total de 3.661 palavras. Estas palavras são então representadas por: a) 300 *features* provenientes de suas representações vetoriais; b) dez valores binários, indicando a sua relação com as 8 emoções de Plutchik e as duas polaridades (positiva e negativa).

Após uma análise manual das 100 palavras mais representativas (de acordo com seus respectivos TF-IDFs) em cada um dos tópicos de conversação, foram retiradas do conjunto de treinamento todas palavras do NRC presentes nesta análise que possuíssem um significado dentro do contexto de MOBAs diferente de seu uso geral. No total foram removidas 16 palavras, mostradas na Tabela 5.5, que possuem significados distintos dentro e fora do contexto do jogo, e assim, provavelmente representam emoções diferentes nos dois contextos.

Passo 3: Após experimentos prévios com vários algoritmos, construímos experimentos mais detalhados em cima de três algoritmos de aprendizagem: SVM com kernel RBF, multi-layer perceptron (MLP) e a regressão linear L2. A implementação do SVM utilizada foi a provida pela API scikit-learn (PEDREGOSA et al., 2011). Como o SVM não provê suporte a problemas múlti-rótulo, utilizamos o algoritmo de relevância binária para fornecer este suporte. Os parâmetros utilizados são similares aos parâmetros padrões providos pela API, exceto quanto aos pesos para cada rótulo, que foram estabelecidos como o inverso da frequência deste no léxico, com o objetivo de atacar o problema dos dados desbalanceados. Note que estes pesos são individuais a cada um dos modelos gerados pelo algoritmo de relevância binária.

Já para o MLP, utilizamos a implementação provida pela API keras[5]. Similarmente ao SVM, precisamos contornar os problemas inerentes aos nossos dados, que são multi-rótulo e desbalanceados. Apesar do MLP possuir suporte a múltiplos rótulos, o algoritmo de relevância binária acabou por prover melhores resultados neste problema, então ele foi utilizado sobre uma versão de uni-rotulada do MLP. Para lidar com o problema do desbalanceamento, utilizamos o algoritmo de sub-amostragem ENN. O ENN foi aplicado nas entradas de cada um dos modelos gerados pelo algoritmo de relevância binária.

Os parâmetros utilizados na construção de cada modelo foram escolhidos após uma série de testes com os diferentes parâmetros listados na Seção 3.6. Os parâmetros que geraram melhores resultados considerando a métrica Micro Medida-F estão listados a seguir.

- Três (3) camadas, já que um número maior de camadas não se traduziu em melhores resultados. A estrutura das camadas está listada abaixo.
  - Uma camada de entrada com 300 neurônios, com ativação linear,
  - Uma camada oculta com 100 neurônios, com ativação *ReLU*,
  - uma camada de saída com 2 neurônios, com ativação *softmax*.
- 37.5% de chance de *dropout* entre a camada de entrada e a oculta se mostrou efetivo para evitar problemas com *overfit*.
- A função de perda escolhida foi o erro quadrático médio (MSE).
- O algoritmo de otimização da rede escolhido foi o ADAM.

Os dados foram divididos aleatoriamente em uma proporção 70/10/20 para treino, validação e testes, respectivamente. Apresentamos na Seção 6.5.1 os melhores resultados obtidos, junto com as respectivas métricas.

[5]https://keras.io/

Tabela 5.6: Distribuição das emoções encontradas no léxico voltado a MOBAs

Emoção	Quantidade
Positivo	1681
Negativo	3127
Alegria	367
Antecipação	509
Confiança	639
Medo	1182
Nojo	837
Raiva	976
Surpresa	258
Tristeza	791

Passo 4: Finalmente, utilizamos o modelo definido para descobrir as emoções das palavras desconhecidas ao nosso dicionário base, construindo assim um dicionário de emoções específico ao domínio de MOBAs. As palavras anteriormente presentes no nosso conjunto de treinamento foram mantidas, enquanto as demais palavras do corpus tiveram suas emoções definidas a partir do modelo treinado.

Como resultado deste processo, foram descobertas emoções para 10.777 palavras. A distribuição das diferentes emoções dentro do léxico estão descritas na Tabela 5.6. A Seção 6.5 discute os resultados obtidos, e apresenta uma avaliação preliminar da qualidade deste léxico.

5.5.2 Caracterização das emoções de cada tópico

Com o léxico de emoções específico ao domínio de MOBAs construído conforme descrito na seção anterior, podemos caracterizar as emoções predominantes em cada tópico. Foi considerado que a emoção de cada tópico é representada pelas 100 palavras mais relevantes do mesmo, como descrito na Seção 5.2.1. Então, similarmente ao feito com as palavras dos agrupamentos, foi medida a relevância das palavras presentes nos tópicos, ordenando-as em um *ranking* de acordo com o valor do TF-IDF de cada palavra.

Para medir as emoções presentes em um tópico, é necessário contar as emoções associadas a cada uma das palavras presentes no *ranking* de relevância. Então, para cada emoção, soma-se as frequências de todas as palavras associadas a ela. Contudo, esta abordagem privilegia as palavras mais frequentes em um dado tópico, e não as palavras mais relevantes.

Para refletir que a palavra mais relevante de um tópico possua um peso emocional maior do que a segunda palavra mais relevante, e assim por diante, poderíamos usar o TF-IDF como

Tabela 5.7: Emoções para as 10 palavras mais relevantes no tópico reclamações

Palavra	TF-IDF	**Peso**	**Polaridade**	**Emoções**
Moron	0,290	1,0	Negativo	-
Trash	0,271	0,5	Negativo	Nojo, Tristeza
Useless	0,266	0,33	Negativo	-
Retard	0,265	0,25	Negativo	Medo, Nojo, Tristeza
Dumbass	0,264	0,20	Negativo	Nojo
Worst	0,262	0,17	Negativo	Medo, Nojo, Tristeza
Ganked	0,261	0,14	-	Tristeza
Cs	0,257	0,12	Negativo	-
Retarded	0,256	0,11	Negativo	Medo, Nojo, Tristeza
Blame	0,250	0,10	Negativo	Nojo, Raiva

peso para distinguir as palavras de acordo com sua relevância. Entretanto, o valor do TF-IDF varia de maneira distinta entre os tópicos. No geral, tópicos mais frequentes como táticas ou reclamações tendem a ter uma maior quantidade de palavras com TF-IDFs altos, o que poderia acabar por adicionar um viés em nosso valor de emoção. Por exemplo, a palavra com maior TF-IDF no tópico de táticas tem um TF-IDF de 0,55, enquanto a palavra mais relevante do tópico bate-papo, apresenta um TF-IDF de apenas 0,25.

Para resolver este problema, utilizamos as posições das palavras no *ranking* para calcular nossos pesos, ao invés do TF-IDF. Assim, o peso de cada palavra é o inverso de sua posição no *ranking* do tópico. Um exemplo de *ranking* para as 10 palavras mais relevantes de um dos tópicos (Reclamações) está exposto na Tabela 5.7. Por exemplo, se considerarmos as 10 palavras mais relevantes no tópico de Reclamações, os pesos dados às palavras em relação à emoção medo serão $0,25$ (retard), $0,17$ (worst) e $0,11$ (retarded).

Note que as 10 palavras mais relevantes do tópico Reclamação na Tabela 5.7 diferem daquelas listadas na Tabela 6.1, devido à diferença entre a quantidade de palavras escolhidas como relevantes para o cálculo do ranking de TF-IDF. Contudo, os grupos representados pelos tópicos permanecem os mesmos. Os resultados decorrentes da análise das emoções de cada tópico são relatados na Seção 6.5.

6 RESULTADOS

Neste capítulo, descrevemos os resultados dos experimentos delineados no Capítulo 5, descrevendo os tópicos descobertos e demonstrando suas várias características e relacionamentos entre si e com os grupos de jogadores. O capítulo descreve os resultados da interpretação dos tópicos, bem como sobre suas principais caraterísticas; caracteriza os grupos de jogadores sob a ótica dos tópicos, e como cada tópico se relaciona de maneiras distintas com estes grupos; a relação temporal entre tópicos, e como as transições se dão em diferentes grupos de jogadores; cada tópico de acordo com as 8 emoções de Plutchik.

6.1 Tópicos de conversação entre jogadores

O processo de descoberta e interpretação de tópicos descrito na Seção 5.2 resultou nos 10 agrupamentos descritos na Tabela 6.1, juntos com as respectivas 10 palavras mais relevantes.

Cada tópico então representa o assunto prevalente de conversação em cada agrupamento. Relacionamos estes agrupamentos com sete padrões de comportamento distintos, e um padrão linguístico. Esta interpretação foi validada por mais 5 jogadores, conforme o processo descrito na Seção 5.2.2. No Apêndice A, resumimos a interpretação de cada um dos agrupamentos feita por estes jogadores, junto com o nível de concordância ou discordância com a interpretação apresentada abaixo.

O agrupamento 0 (*Outras Línguas*), contém predominantemente termos não pertencentes à língua inglesa, sendo o único agrupamento que agrega estas palavras. Este agrupamento foi removido do restante da análise, pois ele não representa um padrão de comportamento.

Tabela 6.1: Tópicos e 10 palavras mais relevantes

	Id.	**Tópico**	**10 palavras mais relevantes**
Positivo	0	*Outras Línguas*	mira, vos, weon, ahora, wn, vida, qe, esa, eso, asi
	1	*Táticas*	**drag**, ss, **warded**, **red**, **flash**, **ward**, eve, **wards**, tf, blue
	2	*Táticas*	**group**, together, **baron**, **fight**, **push**, stay, focus, wait, lets, cait
	3	*Táticas/Educação*	brb, mia, np, gj, thx, sry, ty, sorry, lag, mid
	4	*Bate-papo*	op, dat, ap, blitz, ad, build, damage, lol, oh, xd
Negativo	5	*Reclamações*	**cs**, **fucking**, **adc**, **ur**, **support**, **stfu**, **lane**, **fed**, **idiot**, dumb
	6	*Reclamações*	report, ban, troll, trolling, bg, reported, afk, **feeding**, pls, reporting
	7	*Discussões*	talking, people, playing, mad, said, game, say, play, like, cause
	8	*Insultos*	nooob, fu, noob, noobs, vs, ks, idiot, fck, stupid, omg
	9	*Provocações*	nigger, faggot, cunt, mom, fag, dick, bitch, gay, mad, ass

Os agrupamentos restantes representam padrões de comportamento, nomeados *Táticas*, *Táticas/Educação*, *Bate-papo*, *Reclamações*, *Discussões*, *Insultos* e *Provocações*. A análise destes tópicos permite a sua categorização em tópicos *Positivos* e *Negativos*, assim como outras generalizações. No restante desta seção, descreveremos cada tópico, assim como a classificação destes em tópicos mais genéricos.

6.1.1 Tópicos Positivos

Tópicos classificados como *Positivo* são relacionados com interações saudáveis entre jogadores, as quais são associadas com a própria essência do jogo. Estes tópicos focam predominantemente em estabelecer coordenação tática, eventualmente promovendo um bom ambiente de jogo. Assim, eles contribuem com o trabalho em equipe, e criam confiança entre jogadores. Os tópicos específicos dentro desta categoria estão descritos abaixo, seguidos por suas categorizações em tópicos mais genéricos. A menos que explicitamente mencionados, todos os cinco avaliadores concordaram com a descrição provida.

- **Táticas:** este tópico é representado pelas conversas nos agrupamentos 1 e 2, que são caracterizados principalmente por palavras descrevendo chamadas táticas inerentes ao jogo. Estas palavras são relacionadas com chamadas a objetivos (e.g. drag, baron), coordenação entre membros de uma equipe (e.g. push, stay), e chamadas de visão de mapa (e.g. ward, warded). Elas indicam que uma equipe está tentando se coordenar taticamente e agir em conjunto, o que é essencial para se ganhar uma partida. Na Tabela 6.1, palavras em negrito representam as palavras mais relevantes dos agrupamentos 1 e 2 quando considerados juntos. Dois dos cinco avaliadores caracterizaram este tópico como sendo relacionado a chamadas táticas globais, mas todos concordaram que o tópico pode ser generalizado como chamadas táticas sem perda semântica.

- **Táticas/Educação:** as conversas no agrupamento 3 são levemente distintas das que aparecem no tópico Táticas. Este tópico também apresenta palavras referentes a chamadas táticas, referentes à coordenação nos corredores (e.g. brb, mia). Mas este agrupamento também inclui palavras educadas como agradecimentos (e.g. thx, ty), desculpas (e.g. sorry, np) e felicitações (e.g. gj). Estas palavras indicam que, além de estarem focados em chamadas táticas, os membros de uma equipe estão sendo educados uns com os outros.

- **Bate-papo:** as palavras predominantes em conversas no agrupamento 4 são jargões do

jogo utilizado pelos jogadores (e.g. ap, build), e interjeições ou *emoticons* expressando sentimentos durante a partida (e.g. lol, xd). Quatro de cinco avaliadores concordaram que estas expressões revelam discussões sobre elementos de jogo misturadas com tentativas de socialização, criando um melhor ambiente para o trabalho em equipe.

Criamos duas generalizações dos comportamentos descritos acima, que são úteis para entender os efeitos da contaminação tóxica em grupos de jogadores. Generalizamos os tópicos Táticas e Táticas/Educação como um único tópico, denominado *Relacionados a Tática*, pois no final das contas os jogadores nestas equipes estão se focando em coordenação tática e trabalho em equipe.

Também generalizamos os tópicos de Táticas/Educação e Bate-papo como o tópico *Relacionados ao humor*, porque estes tópicos mostram que os jogadores em um grupo estão socializando e criando um ambiente melhor para trabalho em equipe, melhorando assim o humor geral da equipe. Note que os tópicos Relacionados a tática e Relacionados ao humor não representam grupos disjuntos, uma vez que o tópico de Táticas/Educação é incluso em ambos.

6.1.2 Tópicos Negativos

Tópicos classificados como *Negativo* são relacionados a conflitos, culpabilização e ofensas de diferentes intensidades. De uma forma geral, estes tópicos prejudicam o ambiente do jogo, e mostram sinais de comportamento tóxico. Os tópicos Reclamações, Discussões, Insultos e Provocações descritos abaixo, são classificados como pertencentes a esta categoria. Abaixo, a menos que esteja explicitado, todos os cinco avaliadores concordaram com a descrição apresentada.

- **Reclamações:** este tópico é representado pelas conversas nos agrupamentos 5 e 6. Na Tabela 6.1, as palavras em negrito representam as 10 palavras mais relevantes quando estes dois agrupamentos são considerados juntos. Algumas palavras indicam que os jogadores não estão se dando bem (e.g. dumb, stfu), e que eles culpam uns aos outros por possíveis erros durante a partida (e.g. fed, feeding). Existem também algumas expressões que correspondem a ameaças de denúncia por comportamento tóxico (e.g. report, ban), e acusações de comportamento tóxico intencional (e.g. troll, trolling). Outras palavras expressam a insatisfação dos jogadores com o resultado da partida (e.g. bg, afk). No geral, estas palavras representam conflitos e descontentamento entre os jogadores. Ainda que todos os cinco avaliadores tenham concordado que os dois agrupamentos possam ser

generalizados como Reclamações, alguns deles distinguiram-nos: um dos agrupamentos estaria mais relacionado a reclamações sobre o resultado da partida e o comportamento dos jogadores, enquanto que o outro agrupamento estaria como mais relacionado a culpabilizações.

- **Discussões:** o tópico de conversação no agrupamento 7 foi o mais difícil de interpretar. Suas palavras não fazem muito sentidos por si só (e.g. game, people), e podem ser usadas em muitos contextos diferentes. Por essa razão, muitos de nossos avaliadores descreveram inicialmente este tópico como 'conversas em geral'. Entretanto, identificamos algumas expressões utilizadas em discussões (e.g. talking, said, cause), e um dos avaliadores interpretou o significado do agrupamento como 'tentativas agressivas de se conter comportamento percebido como tóxico'. Adicionalmente, análises mostraram que este tópico é similar a outros tópicos negativos no que tange seu relacionamento com o desempenho e a contaminação dos grupos. Esta nova informação levou a maioria de nossos avaliadores a concordarem com a interpretação de 'discussões agressivas, majoritariamente com a intenção de se conter um comportamento percebido como tóxico'.

- **Insultos:** todas as palavras de maior relevância encontradas no agrupamento 8 são associadas a dar vazão a raiva dos jogadores. Tipicamente, os insultos referem-se a alegada falta de habilidade de outros jogadores (e.g. noob e suas variações), ou possíveis equívocos por parte de jogadores (e.g. idiot, stupid). A prevalência de insultos mostra grupos de jogadores que estão em uma situação de alto estresse.

- **Provocações:** As conversas agrupadas no agrupamento 9 envolvem xingamentos e palavras ofensivas. Provocações são expressados principalmente através de palavrões racistas (e.g. nigger), homofóbicos (e.g. faggot) e sexistas (e.g. bitch). Consideramos provocações a forma mais direta de comportamento tóxico, uma vez que ela tende a desestabilizar emocionalmente os outros jogadores. Quatro dos cinco avaliadores concordaram que as palavras neste agrupamento são significativamente mais ofensivas do que as presentes no agrupamento de Insultos. Entretanto, não se chegou a um consenso quanto ao fato de estas palavras estarem sendo empregadas intencionalmente para abalar emocionalmente os outros jogadores (3 votos) ou se elas são empregadas de maneira não-intencional (2 votos).

Tabela 6.2: Distribuição dos tópicos nos grupos

Tópicos	**Todos**	**Inimigo**	**Não Cont.**	**Inimigo Cont.**	**Aliado**	**Ofensor**
Tópicos Positivos	57%	81%	86%	73%	55%	35%
Rel. a Táticas	44%	64%	66%	55%	44%	25%
Táticas	21%	24%	26%	22%	22%	16%
Táticas/Educação	24%	40%	44%	33%	22%	9%
Rel. a Humor	36%	57%	58%	51%	33%	19%
Bate-papo	13%	17%	17%	18%	11%	10%
Tópicos Negativos	43%	18%	14%	27%	45%	65%
Reclamações	19%	6%	6%	8%	20%	30%
Discussões	13%	7%	4%	13%	16%	15%
Insultos	7%	3%	2%	3%	6%	13%
Provocações	4%	2%	1%	3%	3%	7%

Tabela 6.3: Mèdia e Desvio padrão (SD) dos grupos para desempenho e contaminação

	Desempenho		**Contaminação**	
Grupos	**Média**	**SD**	**Média**	**SD**
Inimigo	0,116	0,030	0,094	0,161
Não-Contaminado	0,120	0,026	0,000	0,000
Contaminado	0,107	0,034	0,286	0,156
Aliado	0,084	0,031	0,300	0,218
Ofensor	0,081	0,030	-	-

6.2 Tópicos e grupos de jogadores

A Tabela 6.2 mostra a distribuição dos tópicos de acordo com cada tipo de jogador. Cada entrada indica a taxa de uso de um tópico para cada tipo de grupo. Nesta tabela, *Tópicos Positivos* refere-se à soma de todos os tópicos positivos (i.e. Táticas, Táticas/Educação e Bate-papo). Similarmente, *Tópicos Negativos* refere-se à soma de todos os tópicos negativos (i.e. Reclamações, Discussões, Insultos e Provocações). Finalmente, *Relacionados a tática* contabiliza a soma dos tópicos Táticas e Táticas/Educação, enquanto *Relacionados ao humor* contabiliza a soma dos tópicos Táticas/Educação e Bate-papo.

A Tabela 6.3 mostra o desempenho e a contaminação média para todos os tipo de grupos, juntamente com seus respectivos desvios padrões. Grupos inimigos são divididos em contaminados e não-contaminados. O grupo de ofensores não aparece na tabela já que não estão associados um valor de contaminação.

No restante desta seção, detalhamos as características de cada tipo de grupo, relacionando os dados de contaminação tóxica da Tabela 6.2, com os desempenhos e as contaminações médias reportadas na Tabela 6.3.

- **Grupos Inimigos:** Inimigos mostram a mais alta taxa de uso de tópicos positivos, quando comparado com aliados e ofensores. Então é possível inferir que na maior parte do tempo, eles tendem a manter uma atmosfera de grupo melhor, um maior foco em táticas, e jogam melhor em equipe. Eles são raramente caracterizados por tópicos negativos, particularmente insultos e provocações, que só contabilizam 5% dos tópicos dentro deste grupo. Isso dá origem à nossa alegação de que inimigos tendem a ser menos expostos ao comportamento tóxico. Observa-se a partir da Tabela 6.3 que grupos inimigos apresentam melhor desempenho e contaminação significativamente mais baixa, quando comparados com grupos aliados. Estes fatos embasam nossa afirmação de que uma melhor atmosfera de jogo se traduz em melhores resultados de jogo.

- **Grupos não-contaminados:** A maioria dos inimigos são de fato não-contaminados (67%). Isso reforça nossa alegação de que os inimigos são menos expostos à toxicidade, dada a falta de contato direto com o ofensor, contato este que só ocorre através do chat global. As características dos inimigos são aumentadas em grupos sem contaminação, tanto em termos de desempenho quanto de padrões de comportamentos expostos pelos tópicos. No geral, grupos não-contaminados possuem o melhor desempenho entre seus pares, mostrando um desempenho médio consideravelmente maior do que o dos outros grupos. Grupos não-contaminados também mostram a maior taxa de uso de tópicos positivos: 12pp (pontos percentuais) acima dos inimigos contaminados, e 31pp acima dos grupos aliados. Assim podemos assumir que estes grupos são caracterizados por uma atmosfera de jogo positiva. Consequentemente, eles mostram a menor taxa de tópicos negativos (14%), onde tópicos como insultos e provocações representam apenas 3% das partidas. Estes resultados nos permitem vislumbrar como tópicos positivos beneficiam fortemente o desempenho de jogadores. Mesmo que estes grupos, por definição, não sofram a influência do jogador tóxico, observamos ainda a presença de tópicos negativos, mesmo que muito baixa, o que nos leva a crer que uma pequena presença de tópicos negativos é natural em partidas online.

- **Grupos Inimigos Contaminados:** Somente 33% dos grupos inimigos estão associados com algum nível de contaminação. Inimigos contaminados também mostram uma alta taxa de uso de tópicos positivos, e bom desempenho, ainda que inferior quando comparados com seus irmãos não-contaminados. Entretanto, eles ainda mostram melhor desempenho e uso de tópicos positivos do que seus pares aliados e ofensores. Comparados com aliados, eles apresentam uma taxa de uso de tópicos positivos superior em 18pp. O uso de tópicos negativos em grupos inimigos contaminados praticamente dobra quando comparados com grupos

não-contaminados. O uso de insultos e provocações também duplica, passando de 3% para 6%. Entretanto, essa taxa de uso de tópicos negativos ainda é muito menor do que em grupos aliados, com uma diferença de 18pp. A contaminação média para inimigos contaminados e aliados é virtualmente a mesma, o que poderia nos levar a crer que eles são afetados de maneira similar pelo comportamento do ofensor. Contudo, isso não é verdade, já que os inimigos são menos vulneráveis à contaminação por serem menos expostos ao ofensor. Em contraste, este tem uma miríade de maneiras de afetar seus aliados, tanto por comunicação verbal nos bate-papo geral ou de times, como por ações de jogo. Confirmaremos esta alegação com mais detalhes nas seções 6.3.1 e 6.3.2.

- **Grupos Aliados:** Aliados estão entre os grupos inimigos e ofensores, tanto em termos de uso de tópicos quanto em desempenho. Comparados aos ofensores, eles mostram uma taxa de uso de tópicos positivos significativamente maior (20pp) e um desempenho levemente maior. Por outro lado, eles mostram uma taxa de uso de tópicos positivos menor do que a dos inimigos contaminados (18pp), e desempenho inferior. A diferença nos tópicos negativos segue o mesmo padrão, com os aliados utilizando mais tópicos negativos do que inimigos, mas menos do que ofensores. O uso de tópicos negativos e positivos pelos aliados é razoavelmente balanceado, mas os tópicos positivos ainda predominam com uma diferença de 10pp. Essa menor taxa de tópicos positivos mostra que estes grupos não estão tão concentrados no trabalho em equipe quanto seus inimigos, por causa de níveis mais altos de tópicos negativos, especialmente reclamações e discussões. Isso é uma evidência que grupos aliados sofrem com disputas internas. A contaminação e desempenho médio para os grupos aliados confirmam essa disputa, já que eles apresentam taxas muito menores de desempenho e taxas mais altas de contaminação do que grupos inimigos no geral.

- **Grupos Ofensores:** Ofensores são caracterizados por tópicos essencialmente negativos, e têm o pior desempenho média entre todos os grupos. Estes grupos mostram a menor taxa de uso de tópicos positivos. Comparando com os aliados, há uma redução de 20pp na taxa de uso de tópicos positivos entre os ofensores, com um aumento proporcional no uso de tópicos negativos. Eles também são o único tipo de grupo onde o uso de tópicos negativos é superior ao de positivos. A dominância de reclamações e discussões mostram que estes grupos estão imersos em conflito com seus aliados. Ainda, ofensores mostram uma taxa significativa de uso de insultos e provocações (20%). Isso confirma a natureza tóxica dos ofensores, e que eles são a fonte principal de toxicidade. Isso faz sentido quando assumimos que tópicos negativos são tóxicos, e que conversas sobre tópicos negativos espalham contaminação. Ofensores

Tabela 6.4: Comparação das performances médias de grupos com/sem a prevalência de um tópico

Tópicos	**Todos**	**Não Cont.**	**Inimigo Cont.**	**Aliado**	**Ofensor**
Tópicos Positivos	0,104 / 0,081	0,122 / 0,109	0,113 / 0,093	0,090 / 0,077	0.089 / 0,078
Rel. a Táticas	0,103 / 0,087	0,122 / 0,116	0,113 / 0,102	0,089 / 0,081	0,088 / 0,080
Rel. a Humor	0,107 / 0,087	0,124 / 0,114	0,114 / 0,101	0,093 / 0,080	0,089 / 0,080
Tópicos Negativos	0,081 / 0,104	0,109 / 0,122	0,093 / 0,113	0,077 / 0,090	0,078 / 0,089
Reclamações	0,077 / 0,097	0,104 / 0,121	0,087 / 0,109	0,073 / 0,086	0,076 / 0,084
Discussões	0,082 / 0,095	0,113 / 0,120	0,097 / 0,109	0,078 / 0,085	0,077 / 0,082
Insultos	0,084 / 0,095	0,108 / 0,121	0,093 / 0,108	0,080 / 0,085	0,076 / 0,084
Provocações	0,091 / 0,094	0,118 / 0,120	0,097 / 0,108	0,092 / 0,084	0,084 / 0,081

Tabela 6.5: Correlações (τ de Kendall) entre performance e concentração de tópicos

	Contaminação		**Desempenho**			
Tópicos	**Inimigo**	**Aliado**	**Não-Cont.**	**Inimigo Cont.**	**Aliado**	**Ofensor**
Tópicos Positivos	-0,70	-0,78	0,80	0,84	0,97	0,95
Rel. a Táticas	-0,70	-0,59	0,76	0,82	0,95	0,94
Rel. ao Humor	-0,67	-0,80	0,95	0,95	0,96	0,61
Tópicos Negativos	0,64	0,68	-0,88	-0,92	-0,95	-0,36
Reclamações	0,62	0,77	-0,88	-0,91	-0,96	-0,50
Discussões	0,71	0,81	-0,43	-0,62	-0,92	-0,69
Insultos	0,001	0,13	-0,82	-0,82	-0,44	-0,31
Provocações	0,70	-0,45	-0,46	-0,81	0,58	0,32

mostram a menor média de desempenho entre todos os tipos de grupos. Isso ocorre provavelmente devido às emoções instáveis dos ofensores e ao constante conflito destes com seus aliados.

6.3 Efeitos de Tópicos sobre Grupos de Jogadores

6.3.1 Efeitos de Tópicos Positivos

Nesta seção, examinamos a relação entre a taxa de uso de tópicos positivos e os valores de desempenho/contaminação, detalhando seus efeitos em cada tipo de grupo. Realizamos essa análise considerando primeiramente os grupos relacionados a tópicos positivos como um todo e depois examinamos com mais detalhes as diferenças entre os tópicos Relacionados a tática e Relacionados ao humor. Optamos por analisar os tópicos generalizados, em vez dos específicos, porque estes mostram os diferentes aspectos representados pelos tópicos positivos: chamadas táticas (Relacionados a tática) e ambiente de jogo (Relacionados ao humor).

Os gráficos das Figuras 6.1 a 6.3 mostram a relação de cada tópico analisado com as métricas de desempenho e contaminação. Nos gráficos, cada ponto corresponde a uma agregação de 10.000 grupos, conforme descrito na Seção 5.3.2. O eixo horizontal mostra o desempenho ou a contaminação média de uma agregação, enquanto o eixo vertical mostra a proporção de uso do respectivo tópico nessa agregação. A curva representa a regressão local desses pontos, e as cores diferenciam as agregações e as curvas de regressão para cada tipo de grupo.

A Tabela 6.4 destaca, para cada tópico e tipo de grupo de jogadores, o desempenho médio considerando todos os grupos onde aquele tópico é prevalente e o desempenho médio nos outros grupos (ou seja, grupos onde outros tópicos são prevalentes). O formato de cada célula representa "*desempenho médio do tópico*" /"*desempenho médio dos outros tópicos*". Por exemplo, a primeira célula da tabela permite comparar o desempenho médio de todos os grupos (*Todos*) relacionados a um tópico positivo (0,104) com o desempenho médio de todos os outros grupos (0,081), que neste caso são os relacionados a um tópico negativo. A Tabela 6.5 mostra a correlação τ de Kendall entre os valores da contaminação/desempenho médio e a respectiva taxa de uso de cada tópico.

- **Tópicos positivos**: A Figura 6.1a exibe a relação entre desempenho e o uso médio de tópicos positivos nas agregações. A taxa de uso de tópicos positivos correlaciona-se positivamente com o desempenho, indicando que o aumento no uso de tópicos positivos está associado a aumento no desempenho, conforme detalhado na Tabela 6.5.

  Observa-se na Figura 6.1a que, enquanto os grupos aliados e inimigos apresentam pontos até aproximadamente 0,2 de desempenho, os grupos de ofensores são apresentam pontos até 0,3. No entanto, isso não representa que os ofensores tenham o maior desempenho.

Isso acontece porque esse grupo é o único baseado no desempenho individual. Para todos os outros grupos, o valor corresponde ao desempenho médio dos jogadores desse grupo. Na verdade, o desempenho individual para os aliados e inimigos atinge valores tão elevados quanto o dos ofensores. Este padrão será observado em todos os gráficos de desempenho discutidos ao longo deste trabalho.

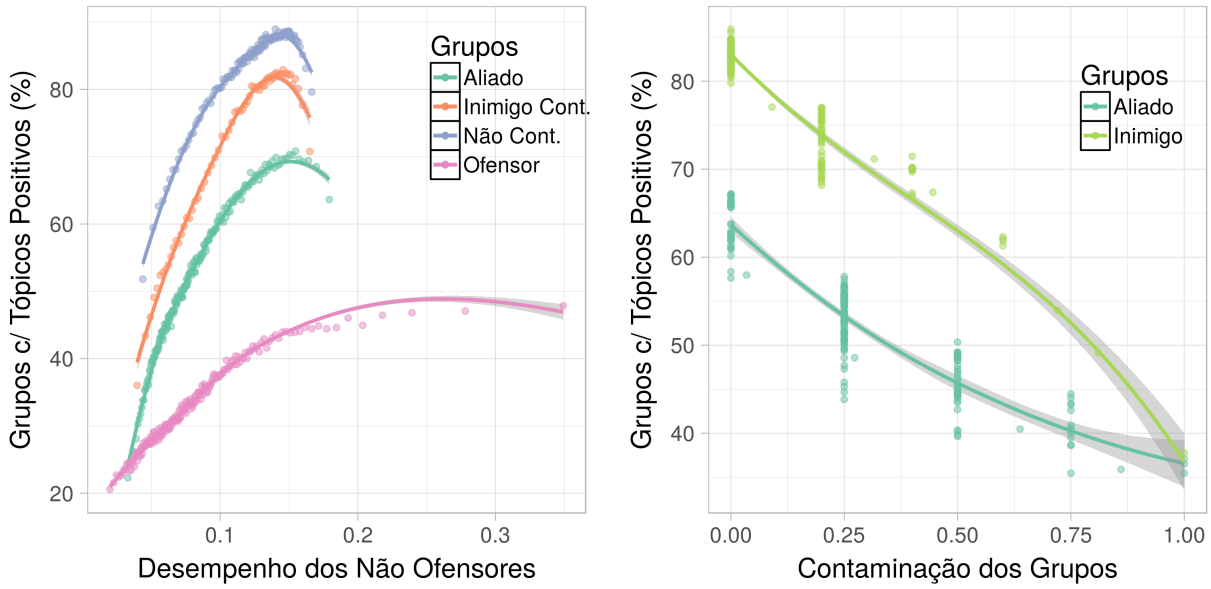
A Tabela 6.4 mostra que as agregações com alta concentração de tópicos positivos contêm grupos que jogam melhor do que os outros, já que seu desempenho médio é consideravelmente superior ao que não estão relacionado a tópicos positivos. Ela também mostra o ranking de desempenho médio para todos os tipos de grupos, onde os grupos não-contaminados têm o melhor desempenho, seguidos pelos grupos inimigos contaminados. Os grupos de aliados vêm em terceiro lugar, quase empatados com grupos de ofensores, que aparecem em último.

Por outro lado, existe uma relação inversa entre contaminação e uso de tópicos positivos. Isso é confirmado pela correlação negativa entre contaminação e uso de tópicos para grupos inimigos e aliados (Tabela 6.5). Para os grupos inimigos, a concentração média de tópicos positivos na metade inferior dos valores de contaminação $(contaminacao < 0,5)$ é 80%, diminuindo para uma média de 55% na metade superior $(contaminacao >= 0,5)$. Para grupos aliados, essas concentrações médias são 56% e 44%, respectivamente.

Observe que, além desta queda, há uma diferença na distribuição de grupos inimigos e aliados sobre o eixo de contaminação, conforme destacado na Figura 6.1b. Vimos que a maioria dos grupos inimigos não são contaminados e, portanto, eles tendem a localizarem-se na metade inferior do intervalo de contaminação. Os grupos aliados, por outro lado, são mais uniformemente espalhados ao longo do eixo, concentrando-se mais entre $0,25$ e $0,50$ de contaminação. Este padrão é observado para todos os tópicos, incluindo os negativos.

- **Tópicos Relacionados a Tática:** Como podemos ver na Tabela 6.2, estes são os tópico mais prevalente entre todos os grupos (44%). Os grupos não-contaminados mostram o maior índice de uso, seguido dos grupos de inimigos contaminados, aliados e ofensores. Observe que apenas 25% dos ofensores se concentram na coordenação tática.

  Os tópicos relacionados a tática constituem a maioria dos tópicos positivos e portanto, é natural que eles compartilhem muitas características em comum: a) relacionam-se com melhor desempenho no jogo (Figura 6.2a), b) a taxa de uso destes tópicos mostra uma



(a) Desempenho x Tópicos positivos.

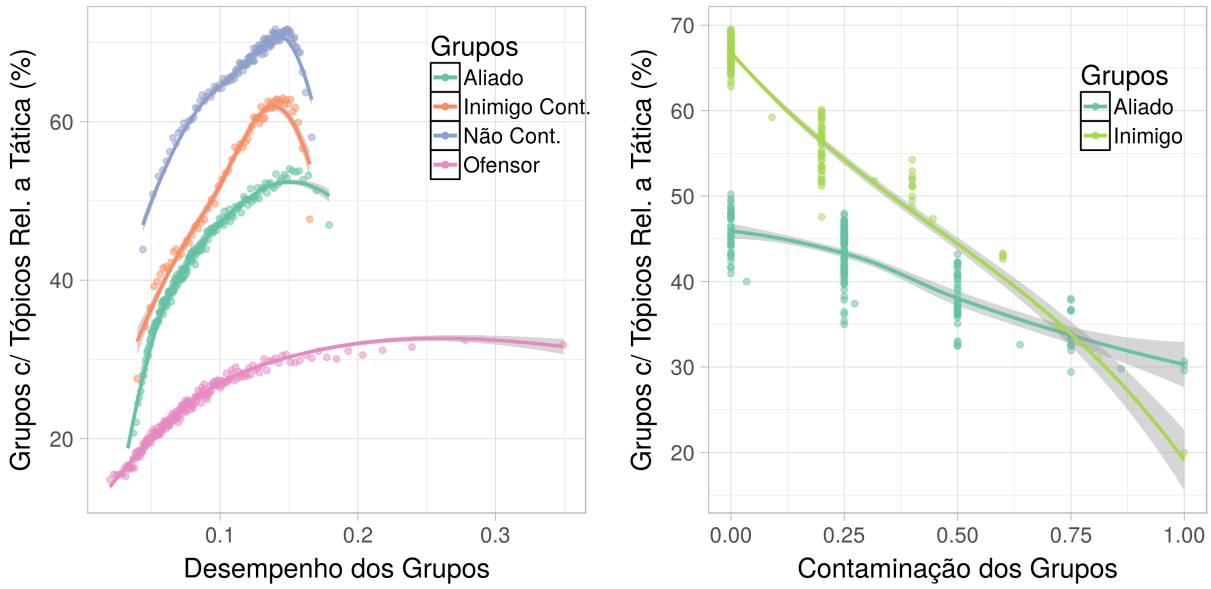
(b) Contaminação x Tópicos positivos.

Figura 6.1: Relação entre desempenho/contaminação e tópicos positivos.

forte correlação positiva com o desempenho de todos os grupos (Tabela 6.5) e c) os grupos que utilizam conversações relacionadas a táticas apresentam melhor desempenho do que os grupos que não as usam (Tabela 6.4), com os desempenhos médios sendo quase iguais aos observados para tópicos positivos em todos os tipos de grupo.

Pode ser observado na Figura 6.2a uma queda acentuada no uso de tópicos relacionados a táticas em níveis de desempenho mais altos para inimigos e aliados. Para grupos não contaminados, ele cai do valor máximo de 72 % para o mínimo de 58 % em $desempenho = 0,14$. Padrões similares são encontrados em grupos de inimigos contaminados (de um máximo de 63% para um mínimo de 48% em $desempenho = 0,145$) e grupos de aliados (de 54% a min. 46% para $desempenho = 0,150$). Os grupos de ofensores não mostram esse comportamento, pois estabilizam o aumento no uso do tópico (cerca de 32%) em níveis de desempenho mais altos. Observe que comportamento semelhante pode ser observado para tópicos positivos (Figura 6.1a). Esta queda caracteriza principalmente grupos que não sentem a necessidade de discutir aspectos táticos pois a vitória é inquestionável, típica em partidas altamente desbalanceadas.

Assim como nos tópicos positivos, a Figura 6.2b mostra uma relação inversa entre tópicos Relacionados a tática e contaminação, bem como uma distribuição semelhante de grupos aliados e inimigos sobre o eixo de contaminação. Fica claro que um maior nível de contaminação está relacionado a uma queda acentuada em tópicos relacionados a tática em aliados e inimigos, conforme mostrado pela correlação negativa entre o uso do tópico e a



(a) Desempenho x Tópicos relacionados a tática. (b) Contaminação x tópicos relacionados a tática.

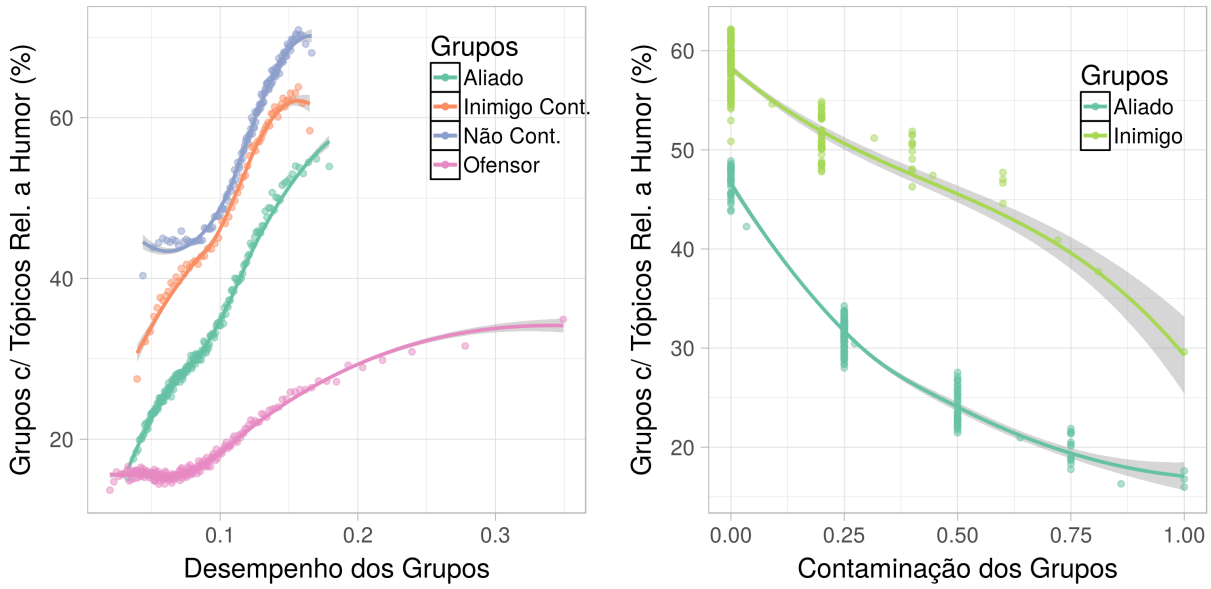
Figura 6.2: Relação entre desempenho/contaminação e tópicos relacionados a tática.

contaminação (Tabela 6.5). Em grupos inimigos, o índice de uso de tópicos Relacionados a tática vai de uma média de 63,1% quando $contaminacao < 0,5$, para uma média de 36,8% quando $contaminacao \geq 0,5$. O mesmo padrão é observado em grupos aliados nos mesmos intervalos de contaminação (de 44,1% a 36,6%).

- **Tópicos relacionados ao Humor:** Tabela 6.2 mostra que o segundo tópico mais frequente é o relacionado a humor (36%). A ordenação da taxa de uso entre os tópicos é semelhante à encontrada para tópicos Relacionados a tática e Positivo, onde os grupos não-contaminados lideram o uso. Novamente, observa-se que os ofensores resistem ao envolvimento em conversas saudáveis (19%).

  A relação entre os tópicos Relacionados ao humor e o desempenho/contaminação é no geral semelhante à existente nos tópicos Relacionados a táticas e nos tópicos positivos (Figura 6.3). A taxa de uso do tópico também se correlaciona positivamente com o desempenho de todos os grupos e negativamente com o nível de contaminação (Tabela 6.5). Os desempenhos médios são quase idênticos aos apresentados nos tópicos Relacionados a tática e Positivo, para todos os grupos (Tabela 6.4).

  No que diz respeito ao nível de desempenho, a diferença mais significativa é que, ao contrário dos tópicos Relacionados à tática, o uso de tópicos Relacionados ao humor não diminui em alto desempenho e, portanto, não se observa uma queda acentuada na Figura 6.3a. Algumas diferenças podem ser observadas na Figura 6.3b no que diz respeito às



(a) Desempenho x tópicos relacionados ao humor. (b) Contaminação x tópicos relacionados ao humor.

Figura 6.3: Relação entre desempenho/contaminação e tópicos relacionados ao humor.

relações entre o uso do tópico e a contaminação. O uso de tópicos Relacionados ao humor por grupos aliados cai rapidamente, de uma média de 35% em menores níveis de contaminação ($contaminacao < 0,5$), para uma média de apenas 22% de níveis mais altos ($contaminacao \geq 0,5$). Os grupos inimigos têm pontuações muito maiores para os respectivos intervalos (56% e 42%, respectivamente).

Discussão: Todos os tópicos positivos estão fortemente correlacionados com o desempenho, o que nos permite concluir que a predominância de conversações positivas dentro de um grupo durante uma partida está fortemente ligada ao bom desempenho desse grupo. Tanto as conversas táticas como as que melhoram o humor associam-se com esse efeito positivo. Este pressuposto nos dá um indicio de que o melhor desempenho dos grupos inimigos, especialmente os não contaminados, pode ser justificado pela sua maior associação a tópicos positivos.

Em todos os grupos relacionados a tópicos positivos, os grupos inimigos não contaminados e contaminados apresentam desempenho significativamente melhor e uma maior concentração de tópicos positivos do que os grupos aliados. Assumimos que essa diferença é justificada por um ambiente de jogo menos contaminado. Apesar da concentração de tópicos positivos para grupos de ofensores também crescer com o desempenho, ela atinge valores baixos (48%), em comparação com grupos aliados (71%), não-contaminados (89%) e inimigos contaminados (83%). Em geral, os ofensores tendem a não se envolver em temas relacionados ao humor, que visam promover um bom ambiente de jogo.

Mostramos que há uma queda no uso do tópico Relacionado a táticas em níveis de alto desempenho para todos os grupos, exceto nos ofensores. Isso ocorre essencialmente em partidas muito desbalanceadas, já que o desempenho médio do grupo adversário nestas partidas é de apenas $0,051$. Apesar disso, a taxa de uso dos tópicos Relacionados a táticas (61%) é ainda muito maior do que a média (43%). Essa queda não é observada em relação aos tópicos Relacionados ao humor, indicando que o foco na coesão interna do time é mantido.

A contaminação crescente está correlacionada com a diminuição no uso de tópicos positivos. No entanto, essa redução ocorre de diferentes maneiras. Em tópicos relacionados à tática, os grupos inimigos têm maior uso de tópicos positivos na faixa de baixa contaminação, com uma queda acentuada na taxa de uso em níveis altos de contaminação, até o ponto em que essa taxa torna-se igual ou menor a à taxa dos grupos aliados. Em tópicos Relacionados ao humor, a relação de uso do tópico dos grupos inimigos permanece sempre maior em comparação com os grupos aliados, mesmo com níveis crescentes de contaminação.

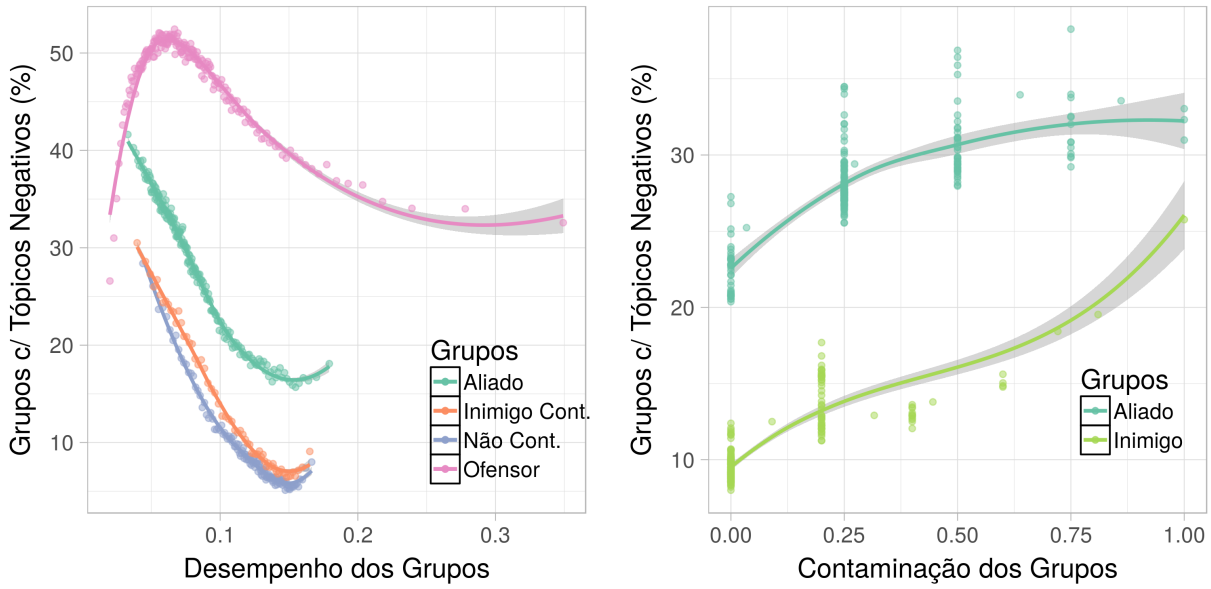
6.3.2 Efeitos de tópicos negativos

A estrutura desta Seção é análoga a anterior: primeiramente analisamos os padrões de desempenho/contaminação relativo aos tópicos negativos como um todo, e então, detalhamos estes efeitos para cada tópico em específico. As Figuras de 6.4 a 6.8 descrevem a relação entre cada tópico e seu respectivo desempenho/contaminação.

- **Tópicos Negativos:** A Figura 6.4a mostra a relação inversa entre taxa de uso de tópicos negativos e desempenho. Conforme apresentado na Tabela 6.5, há uma correlação forte negativa entre a concentração de tópicos negativos e o desempenho, exceto para os ofensores. Como será detalhado no restante desta seção, o desempenho dos ofensores se comporta de maneiras diferentes para cada tópico negativo.

  A Tabela 6.4 mostra que o desempenho médio dos grupos que usam tópicos negativos é pior do que o dos outros grupos. Grupos não-contaminados e grupos inimigos contaminados mantêm os maiores desempenhos médios, mas o desempenho médio dos ofensores é melhor em comparação ao dos grupos aliados. Apesar de os dois últimos estarem quase empatados, a diferença entre eles ainda é estatisticamente significativa.

  No entanto, esse comportamento geral de desempenho não é válido para tópicos negativos específicos. Na verdade, grupos caracterizados pelo uso de Discussões, Insultos e Provocações se desviam deste comportamento, conforme detalhado mais a frente nesta



(a) Desempenho x Tópicos negativos.

(b) Contaminação x Tópicos negativos.

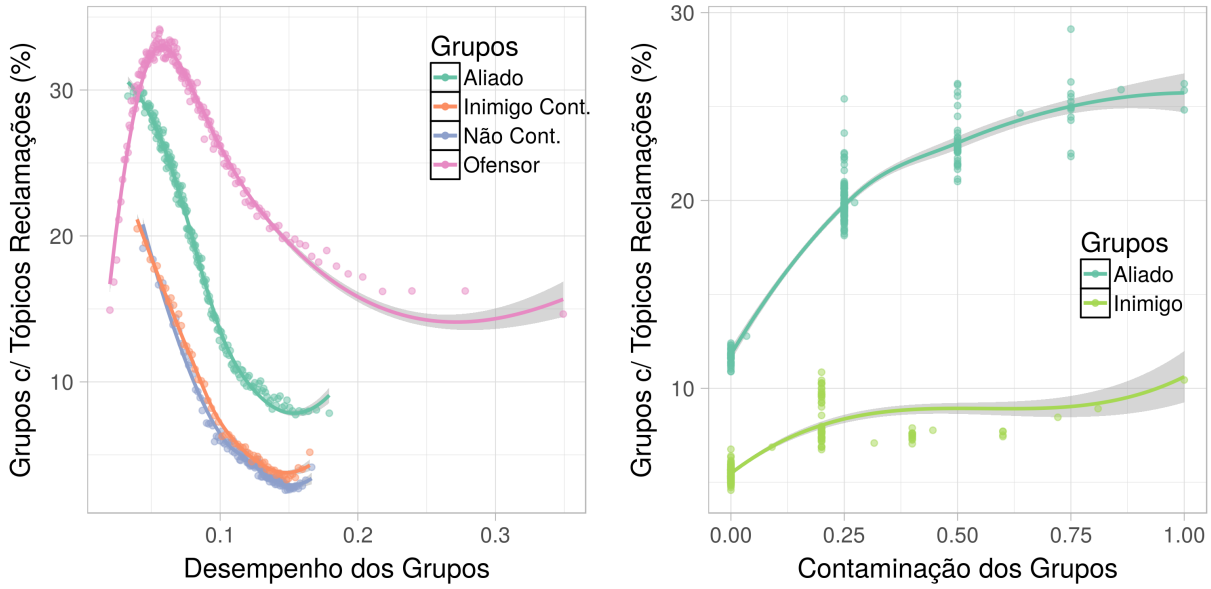
Figura 6.4: Relação entre desempenho/contaminação e tópicos negativos.

seção.

Por outro lado, para os grupos aliados e inimigos, o nível de contaminação correlaciona-se positivamente com a concentração de tópicos negativos, conforme descrito na Figura 6.4b e detalhado na Tabela 6.5. O uso de tópicos negativos por grupos inimigos apresenta um crescimento significativo com à medida que a contaminação cresce, passando de uma proporção média de 11% na metade inferior da contaminação ($contaminacao < 0,5$), para 18% na metade superior da contaminação ($contaminacao \geq 0,5$). O uso de tópicos negativos por grupos aliados é muito maior, em comparação com os grupos inimigos, mas o crescimento de acordo com o nível de contaminação é menor (de uma média de 27 % quando $contaminacao < 0,5$, para 31 %, quando o contrário).

- **Reclamações:** As reclamações são o tópico negativo mais prevalente (19%) e, portanto, este tópico é parcialmente responsável pela manutenção das tendências já observadas para tópicos negativos. Conforme detalhado na Tabela 6.2, seu uso está concentrado em grupos de ofensores e aliados, e é muito menos frequente em inimigos.

  No que diz respeito ao desempenho, também é observada uma correlação negativa entre desempenho e reclamações, bem como o comportamento ligeiramente diferente dos ofensores. Conforme mostrado na Figura 6.5a, os grupos de ofensores diferem em comportamento dos demais, com o uso de reclamações aumentando quando o desempenho



(a) Desempenho x Reclamações.

(b) Contaminação x Reclamações.

Figura 6.5: Relação entre desempenho/contaminação e tópicos de reclamações

.

é muito baixo até atingir um pico em $desempenho = 0,056$. Até este pico, o desempenho dos ofensores se correlaciona positivamente com a taxa do uso de reclamações $(\tau = 0,83)$. Após o pico, o uso do tópico começa a diminuir junto com o desempenho, com correlação negativa $(\tau = -0,88)$. Esta tendência também é observada na Figura 6.4a em relação a tópicos negativos em geral.

Conforme mostrado na Tabela 6.4, o desempenho médio para grupos relacionados a reclamações é menor que o desempenho médio dqueles não relacionados, para todos os tipos de grupos. A ordem média de desempenho por tipo de grupo é a mesma mostrada nos tópicos negativos em geral: inimigos, ofensores e aliados.

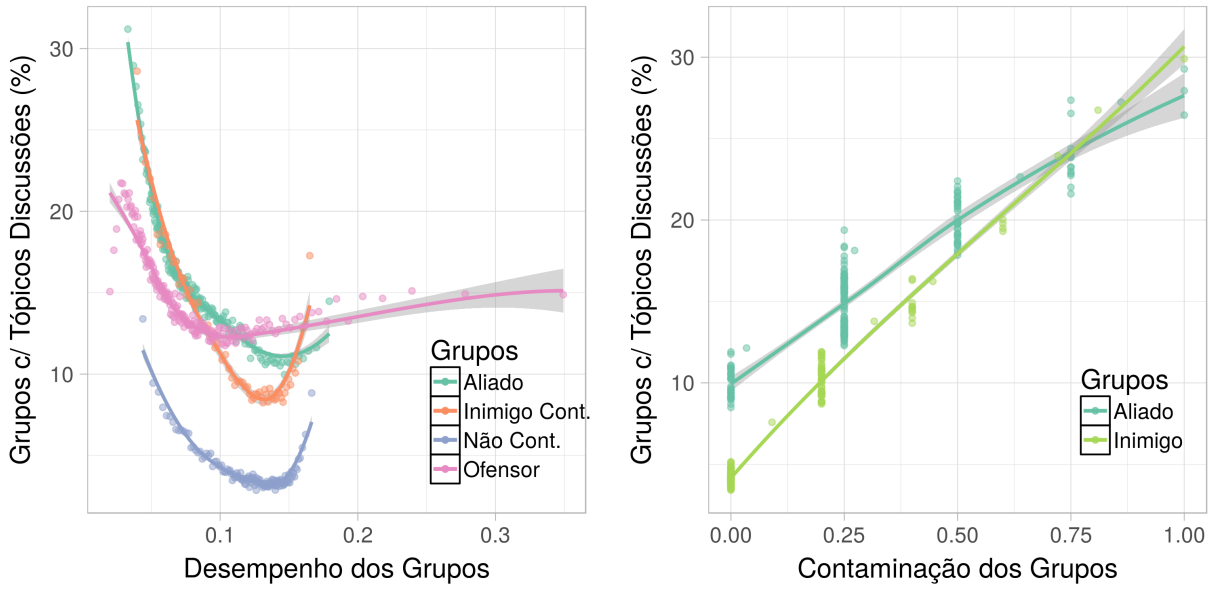
Em relação à contaminação, o comportamento geral e as tendências já destacadas para tópicos negativos como um todo são mantidas (Tabela 6.5). Conforme mostrado na Figura 6.5b, uma média de 18% dos grupos aliados estão envolvidos em reclamações na metade inferior da contaminação, aumentando para uma média de 24% na metade superior. A tendência ascendente relacionando reclamações e contaminação também existe nos grupos inimigos, com uma forte correlação negativa. No entanto, essa tendência é mais suave do que a dos grupos aliados. O uso médio de reclamações por grupos inimigos vai de 6% na menor metade da contaminação, para 8% na metade de contaminação mais alta.

- **Discussões:** A Tabela 6.2 revela que discussões são o segundo tópico negativo mais frequente, adotado por 13% dos grupos. Seu uso é distribuído principalmente entre aliados, ofensores e inimigos contaminados. Este é o único tópico negativo que não é mais prevalente nos grupos dos ofensores, já que aliados e ofensores o adotam em partes iguais (16% e 15%, respectivamente).

  A Tabela 6.4 confirma que os grupos relacionados a discussões mostram um desempenho médio mais baixo do que aqueles que não são relacionados. A diferença entre o desempenho médio dos grupos de aliados e ofensores é pequena, mas significativa. O desempenho e a taxa de uso do tópico apresentam uma forte correlação negativa, exceto para os grupos inimigos não contaminados (Tabela 6.5).

  A Figura 6.6a mostra a relação entre discussões e o desempenho dos diferentes grupos. Essas relações são confusas à primeira vista, pois são representadas por uma curva em forma de "U". As curvas para cada tipo de grupo mostram pontos de inflexão em $desempenho \geq 0,126$ para grupos não-contaminados, $desempenho \geq 0,131$ para grupos inimigos contaminados, $desempenho \geq 0,147$ para grupos aliados e $desempenho \geq 0,137$ para grupos ofensores. Nestes pontos, a tendência da taxa de uso de discussões muda de descendente para ascendente. Examinamos a correlação antes e depois desses pontos de inflexão. Existe uma forte correlação negativa para todos os tipos de grupo antes do respectivo ponto de inflexão, variando de $\tau = -0,82$ para $\tau = -0,93$. Após os respectivos pontos de inflexão, apenas os grupos não contaminados não exibem uma forte correlação positiva entre o índice de uso de discussões e o desempenho ($\tau = 0,43$). Nos outros casos, a correlação existe ($\tau = 0,60$ para inimigos contaminados e $\tau = 0,69$ para aliados e ofensores). Este aumento no uso de tópicos negativos em alto desempenho pode indicar comportamento tóxico em jogos muito desequilibrados, ou algum tipo de confronto feito pelo grupo vencedor, conforme exploraremos mais abaixo.

  Como pode ser visto a partir da Figura 6.6b e da Tabela 6.5, o uso de discussões está relacionado ao crescimento da contaminação tanto para grupos aliados como inimigos. Na metade inferior da contaminação, há um menor uso médio do tópico por grupos inimigos (6%), em comparação com grupos aliados (14%). No entanto, na metade superior da contaminação, a porcentagem média de grupos relacionados ao tópico tópicos de discussões é quase idêntica (23% e 24%, para inimigos e aliados respectivamente). Este é o único subtópico negativo onde as taxas de uso de inimigos e aliados com alto nível de contaminação são semelhantes.



(a) Desempenho x Discussões.

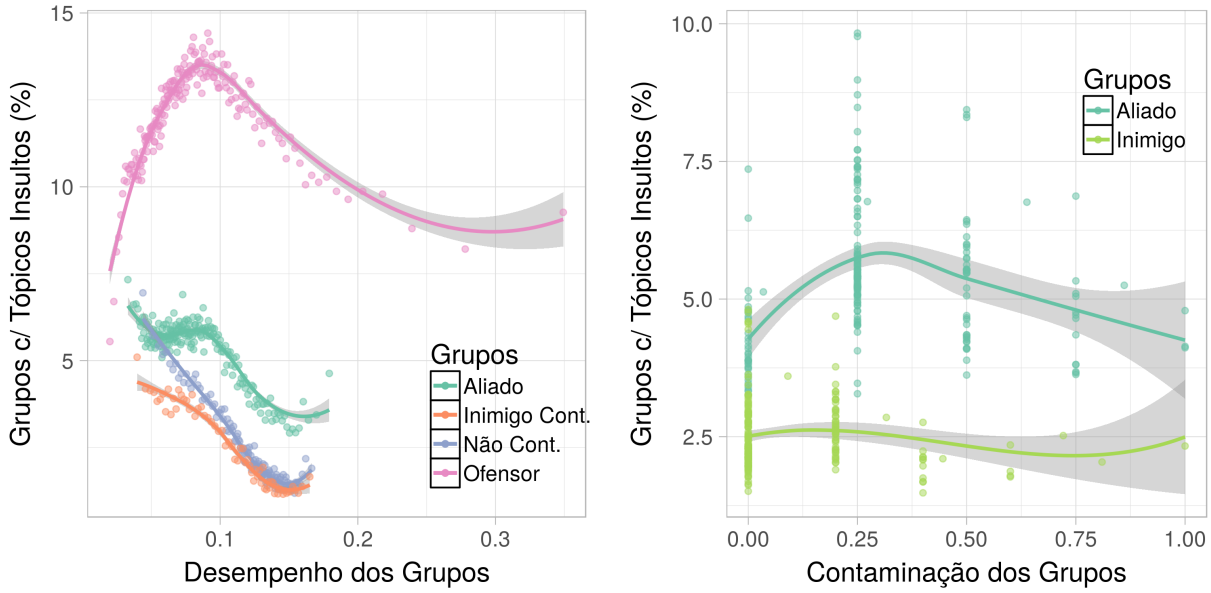
(b) Contaminação x Discussões.

Figura 6.6: Relação entre desempenho/contaminação e tópicos de discussões.

- **Insultos:** Este é o segundo tópico menos prevalente (7%). Conforme detalhado na Tabela 6.2, é principalmente adotado por ofensores (13%), seguido por aliados em menor escala (6%), sendo neglicenciável nos demais tópicos.

  Conforme descrito na Tabela 6.4, o desempenho médio de grupos que usam insultos é menor do que um dos grupos que não o usam. A Figura 6.7a e a Tabela 6.5 detalham a relação entre o uso de insultos e o desempenho por tipo de grupo. Uma forte correlação negativa é observada apenas para grupos inimigos não-contaminados e contaminados, representados pelas tendências descendentes traçadas na Figura 6.7a. A correlação negativa é, no entanto, fraca para aliados e ofensores, conforme inferido pelas tendências apresentadas no gráfico. Em níveis de desempenho mais baixos, inicialmente, há um aumento na concentração de ofensores associados com insultos, até atingir um ponto de inflexão. No mesmo intervalo, o uso de insultos por grupos aliados tende a ser estável, sem correlação. Então, em $desempenho \approx 0,09$, ambas tendências mudam para decrescentes, quando a correlação se torna fortemente negativa ($\tau = -0,8$ para aliados e ofensores).

  Visto a tendência dos outros tópicos negativos, era esperado que a taxa de insultos se correlacionasse com o nível de contaminação. Contudo, isso não ocorre (Tabela 6.5). Conforme ilustrado na Figura 6.7b, a concentração de insultos não varia muito ao longo do eixo de contaminação. O uso médio para grupos aliados e inimigos é 5% e 2,5%, respectivamente, o que mostra que esses grupos não são afetados por insultos dentro da equipe. No entanto, a situação é diferente quando a fonte dos insultos é o ofensor,



(a) Desempenho x Insultos.

(b) Contaminação x Insultos.

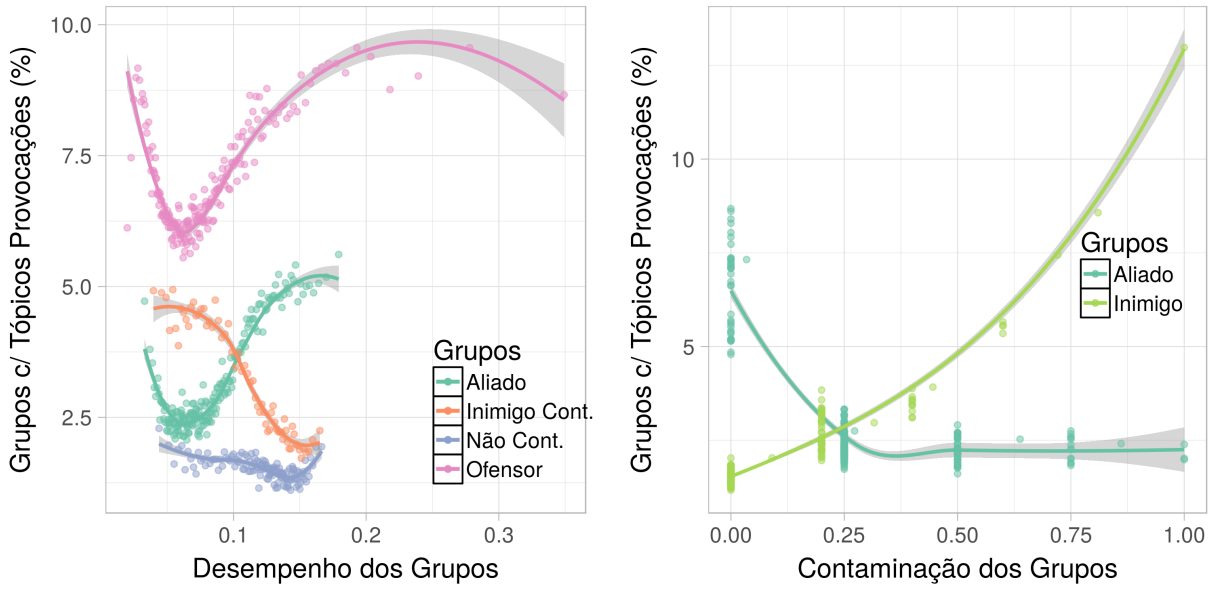
Figura 6.7: Relação entre desempenho/contaminação e tópicos de insultos.

conforme discutido na Seção 6.3.3.

- **Provocações:** Este é o tópico menos usado, caracterizando apenas 4% dos grupos, conforme relatado na Tabela 6.2. Da mesma forma que os insultos, eles são mais usados por ofensores, com aliados e inimigos contaminados mostram índices de uso similares (3 %).

  A Tabela 6.5 e a Figura 6.8a mostram que a relação entre o uso de provocações e desempenho é significativamente diferente para inimigos e aliados/ofensores. Embora se correlacione negativamente com os inimigos, há uma correlação positiva fraca para aliados e ofensores, retratadas no gráfico. Ambos os inimigos não-contaminados e contaminados apresentam uma relação inversa, com formas semelhantes. Os grupos aliados apresentam um comportamento bastante inesperado porque o uso de provocações se correlaciona positivamente com o desempenho. Os grupos de ofensores mostram mais uma vez uma clara curva em forma de "U". Em valores de muito baixo desempenho, a provocação do ofensor diminui com o aumento do desempenho, com uma correlação negativa $(\tau = -0,76)$. No desempenho de $> 0,062$, o uso de provocações começa a aumentar com o crescimento do desempenho, exibindo correlação positiva $(\tau = 0,81)$.

  Um fato interessante sobre os grupos de aliados e ofensores relacionados a este tópico é que seus desempenhos médios são melhores, em comparação com os respectivos grupos não provocadores (Tabela 6.4), e essa diferença é estatísticmente significativa. Este é o único caso em que um comportamento negativo se relaciona a um melhor desempenho.



(a) Desempenho x Provocações.

(b) Contaminação x Provocações.

Figura 6.8: Relação entre desempenho/contaminação e tópicos de provocações.

Já o desempenho médio para grupos inimigos não-contaminados e contaminados usando provocações é menor do que o desempenho médio para os respectivos grupos não provocadores. Discutiremos mais tarde como os ofensores e aliados trabalham juntos para desestabilizar os inimigos através de provocações. Quanto ao ranking de desempenho entre os tipos de grupo, os grupos não contaminados possuem a maior pontuação de desempenho, seguidos por inimigos, aliados e ofensores contaminados.

A Figura 6.8b mostra a relação entre provocações e contaminação, evidenciando a correlação negativa para aliados e positiva para os inimigos. Sem contaminação, os grupos aliados apresentam o maior uso de provocações, contrastando com os grupos inimigos, o que fazem o menor uso. No entanto, no primeiro sinal de contaminação em grupos aliados, o uso da provocação cai para níveis baixos (2,3 %), permanecendo estável apesar da crescente contaminação. Inimigos, por outro lado, experimentam um aumento acentuado no uso de provocações com crescente contaminação.

Discussões: Em geral, os tópicos negativos se correlacionam positivamente com a contaminação em ambos os times (aliados e inimigos). Como todas as denúncias são direcionados aos ofensores, podemos assumir que eles são a principal fonte de comportamento tóxico. Nossos resultados confirmam essa intuição, pois os ofensores sozinhos correspondem por quase 50% do uso de tópicos negativos. O tópico de discussões é o único tópico negativo em que o uso por ofensores empata com aliados, nos outros tópicos negativos a taxa de uso por ofensores é

significativamente maior em comparação com todos os outros grupos. No entanto, a taxa de significativa de discussões por parte dos ofensores, juntamente com a forte correlação entre o uso de discussões e a contaminação, confirmam que este tópico pode ser considerado tóxico, apesar de ser usado com a intenção de combater comportamento tóxico.
As expressões que caracterizam tópicos negativos irritam os outros jogadores em diferentes graus. À medida que a toxicidade aumenta, as discussões parecem ser uma reação comum de aliados e inimigos, enquanto a provocação aumenta principalmente nos grupos inimigos. Os índices de insultos permanecem estáveis em inimigos e aliados, independentemente do nível de contaminação. Podemos assim concluir que os tópicos negativos são manifestações de comportamento tóxico e/ou sua contaminação, já que os jogadores tendem a usar tópicos negativos tanto para gerar toxicidade como para enfrentá-la.

Pode-se pensar que as diferenças entre o uso tópico negativo de grupos não-contaminados e os outros demais grupos dado que eles tendem a ganhar com mais frequência. Contudo, isso é apenas parcialmente verdadeiro. Apesar do seu desempenho superior e índice de vitórias (80%), o uso de tópicos negativos é significativamente menor mesmo entre os perdedores não contaminados. Apenas 25% destes grupos usam tópicos negativos, que são significativamente mais baixo se comparado ao uso por inimigos contaminados (40%), aliados (50%) e ofensores (68%). Assim, podemos concluir que a influência de um jogador tóxico em uma partida tende a ampliar o uso de tópicos negativos e, consequentemente, contaminação da partida.

Apesar das diferenças nas tendências observadas para cada tópico, podemos identificar quanto ao desempenho uma tendência comum até certo ponto para reclamações, discussões e insultos. Para os grupos não-ofensores (ou seja, aliados e inimigos), em geral, o uso de tópicos negativos decai com o aumento do desempenho. O tópico que mais se adapta a esse padrão é o de reclamações, pois o uso de tópicos dos grupos não-ofensores diminui rapidamente com o aumento do desempenho. Discussões e insultos podem exibir algumas peculiaridades, mas ainda assim, seguem esse comportamento em geral.

Analisando grupos que usam discussões, vemos sinais de aumento do uso do tópico em níveis de desempenho mais altos. No entanto, acreditamos que discutir, neste caso, não é um efeito de contaminação, porque o nível de contaminação desses grupos não difere muito da contaminação média de seus respectivos tipos de grupo em geral. Em vez disso, suponhamos que esse comportamento aconteça devido a partidas muito desequilibradas, pois o desempenho médio dos grupos acima mencionados nessas correspondências $(media = 0,1338)$ é muito maior do que o desempenho médio dos grupos opostos correspondentes $(media = 0,060)$. Possíveis explicações para este comportamento são os jogadores de alguma forma minando a equipe

adversária para celebrar uma vitória esmagadora, ou simpatizando com possíveis conflitos na equipe adversária causada pela má performance desta. Note, no entanto, que discussões é o tópico menos claro quanto à interpretação (Seção 6.1), e essas conclusões devem ser consideradas com cuidado.

Usos elevados de provocação por aliados e inimigos caracterizam diferentes comportamentos. Lembre-se de que a relação entre alto desempenho e alto uso de provocações está presente tanto nos grupos de aliados, quanto nos de ofensores. Apesar de um alto uso em níveis de baixo desempenho (ou seja, o quartil de menor desempenho), o uso de provocações para esses grupos tende a aumentar com o desempenho após um certo ponto de inflexão. Ao examinar os níveis de contaminação, concluímos que os grupos aliados tendem a provocar quando apresentam alto desempenho e zero contaminação. Os grupos aliados nesta situação usam mais insultos (7%) em comparação com outras situações (2,5%). Como não há denúncias de jogadores aliados nessas partidas, acreditamos que os aliados estão agindo de forma tóxica, assim como o agressor. Levantamos a hipótese de que os aliados se tornem cúmplices ou contaminados pelo comportamento tóxico do ofensor, empolgados pelo desempenho superior.

Grupos inimigos contaminados, por outro lado, provocam mais quando a contaminação é alta e o desempenho é baixo (9%, versus 2% nos outros casos). Acreditamos que eles fazem isso como uma mistura de frustração resultante das ações do agressor e seu próprio desempenho ruim, o que é um exemplo de como o comportamento tóxico gera mais comportamento tóxico. No entanto, mesmo os grupos inimigos contaminados com baixo desempenho não provocam tanto quanto os ofensores, que ainda mostram um uso muito mais elevado desse tópico (12%).

Os insultos não parecem afetar a contaminação tóxica, apesar das palavras ásperas. Suspeitamos que isso possa ser explicado pelo alvo dos insultos, que pode ser o companheiro de equipe ou o oponente. Para confirmar, analisamos as partidas em que os aliados comunicaram-se apenas através do bate-papo global e comparamos a contaminação média dos grupos inimigos nessas condições com grupos inimigos em geral. No que diz respeito aos aliados que usam insultos nessas condições, a contaminação média dos grupos inimigos é de $0,293$, muito maior do que a contaminação média para grupos inimigos em geral $(media = 0,094)$. Observamos o mesmo em relação às provocações. No entanto, o inverso não é verdade para os inimigos que usam insultos ou provocações nas mesmas condições, pois a contaminação média dos aliados $(media = 0,145)$ é muito menor do que a sua contaminação média geral $(media = 0,30)$. Podemos inferir que os aliados que visam os inimigos com seus insultos/provocações afetam a contaminação destes, mas o inverso não é verdade. De fato, insultos/provocações de grupos inimigos não parecem afetar a equipe oposta (ou seja, aliados/ofensores), mesmo quando dire-

Tabela 6.6: Correlações (τ de Kendall) entre a taxa de uso de tópicos negativos pelo ofensor e contaminação/desempenho de não-ofensores

	Contaminação		**Desempenho**		
Tópicos do Ofensor	**Inimigo**	**Aliado**	**Inimigo Não Cont.**	**Inimigo Cont.**	**Aliado**
Tópicos Negativos	-0,40	0,67	0,79	0,86	-0,91
Reclamações	-0,62	0,69	0,81	0,90	-0,92
Discussões	0,52	0,39	0,57	0,70	-0,48
Insultos	-0,62	0,44	-0,51	-0,40	-0,23
Provocações	0,71	-0,35	-0,50	-0,68	0,73

cionados para eles. Em ambos os casos, os jogadores não parecem afetados por insultos de seus próprios companheiros de equipe.

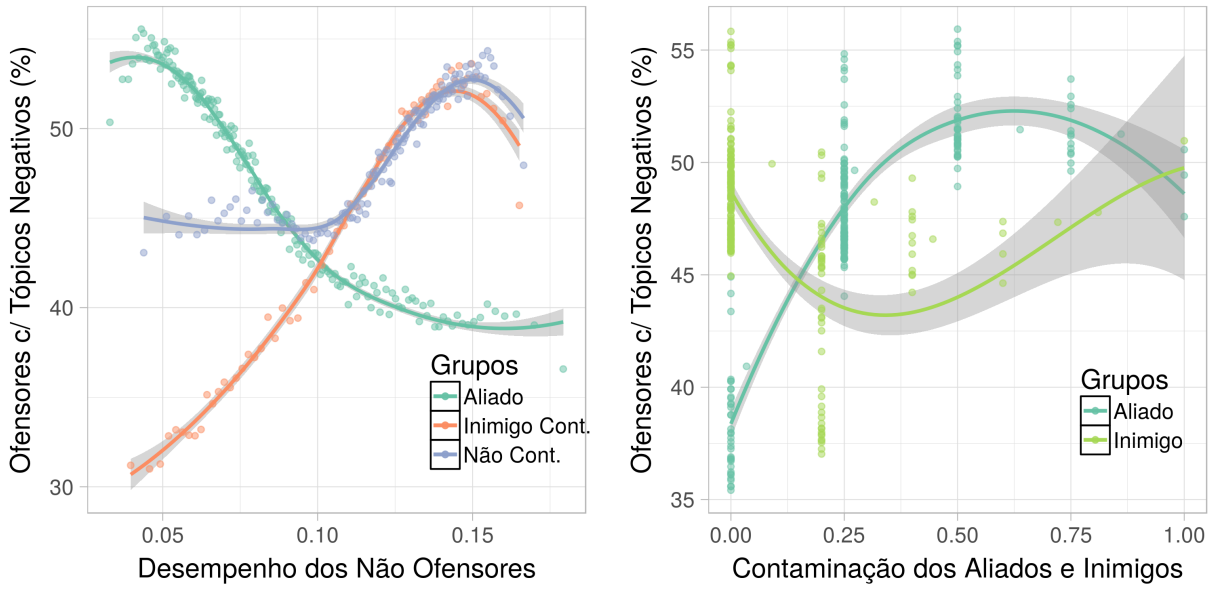
6.3.3 Efeitos dos tópicos negativos do ofensor sobre grupos não-ofensores

Mostramos que os ofensores são a principal fonte de toxicidade em partidas, e que eles usam intensivamente de tópicos negativos. Nesta seção, analisamos os efeitos de tópicos negativos usados pelos ofensores sobre o desempenho/contaminação dos respectivos grupos *não-ofensores*, ou seja, grupos aliados e inimigos relacionados à mesma partida.

Os gráficos nas Figuras 6.9 a 6.13 mostram a relação entre o uso de um tópico por grupos de ofensores e a contaminação ou o desempenho dos respectivos grupos não-ofensores. Vale lembrar que cada ponto corresponde a uma agregação de 10.000 grupos, conforme descrito na Seção 5.3.3, e que as curvas são a regressão local para os pontos mostrados. A Tabela 6.6 mostra a correlação τ de Kendall entre o índice de uso de cada tópico negativo por ofensores e a contaminação/desempenho médio dos respectivos grupos não-ofensores.

Abaixo analisamos a influência de tópicos negativos em geral, seguida de cada tópico negativo em particular.

- **Tópicos Negativos:** A Figura 6.9a mostra a relação entre os tópicos negativos usados pelos ofensores e o desempenho dos respectivos grupos não-ofensores. Podemos ver que o uso de tópicos negativos por ofensores é maior tanto quando seus colegas de equipe (ou seja, grupos de aliados) desempenham mal, quanto quando seus oponentes têm bom desempenho. Por outro lado, níveis mais baixos são observados tanto quando o desempenho dos companheiros se destaca, quanto desempenho dos adversários é fraco. Isso é confirmado pelos dados apresentados na Tabela 6.6, que mostra que o desempenho dos grupos inimigos (não-contaminados e contaminados) apresenta correlação positiva com



(a) Desempenho x tópicos negativos.

(b) Contaminação x tópicos negativos.

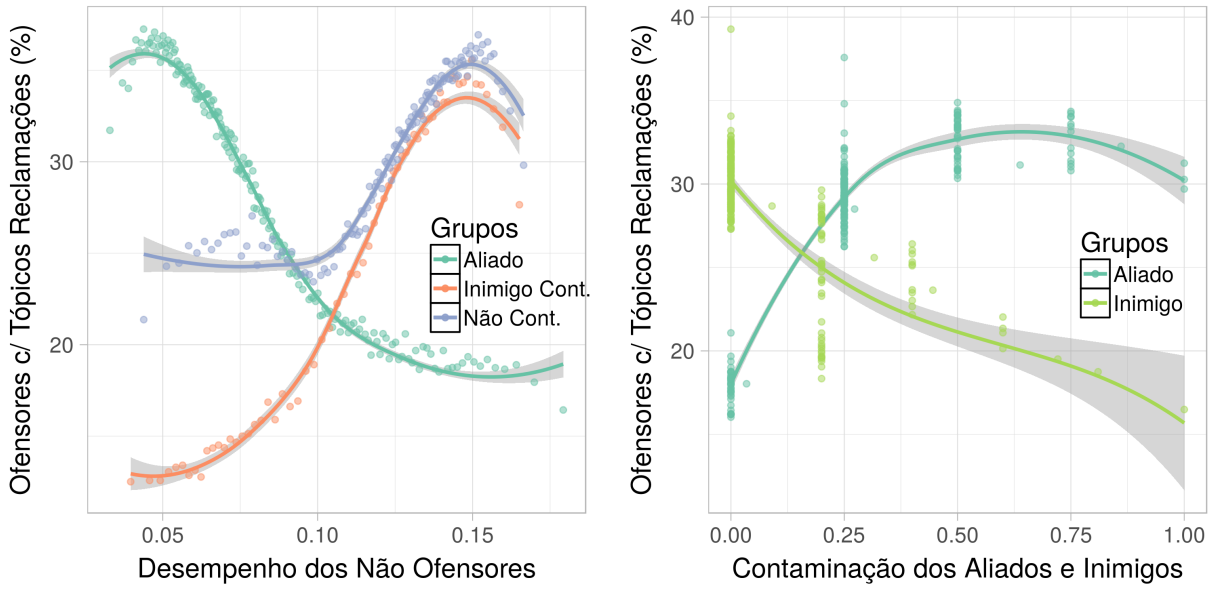
Figura 6.9: Relação entre tópicos negativos dos ofensores e as métricas do não-ofensores.

o uso de tópicos negativos pelos ofensores, enquanto o desempenho dos grupos aliados se correlaciona negativamente com este mesmo uso.

A relação entre o uso de tópicos negativos pelos ofensores e a contaminação de grupos não-ofensores é detalhada na Figura 6.9b e na Tabela 6.6. Verificamos que a contaminação dos aliados aumenta com o uso de tópicos negativos pelos ofensores, uma vez que estes valores se correlacionam positivamente. A relação da contaminação do inimigos com os tópicos negativos dos ofensores mostra um comportamento mais errático, explicado pelas diferenças de comportamento de acordo com o cada tópico negativo, conforme discutido no restante desta seção.

- **Reclamações:** A Figura 6.10a mostra que a relação entre o uso de reclamações por ofensores e o desempenho de não-ofensores é bastante semelhante à observada nos tópicos negativos como um todo. Do mesmo modo, o uso de reclamações por ofensores se correlaciona negativamente com o desempenho dos grupos aliados e, positivamente, com grupos inimigos, como mostrado na Tabela 6.6.

  Analisando com mais detalhe a Figura 6.10a, vemos que as reclamações dos ofensores estão no seu mínimo (13%) quando os inimigos contaminados apresentam desempenho mais baixo. Os inimigos não-contaminados mostram um comportamento curioso, uma vez que o uso de reclamações por ofensores permanece estável para $desempenho < 0,10$



(a) Desempenho x reclamações.

(b) Contaminação x reclamações.

Figura 6.10: Relação entre tópicos de reclamações dos ofensores e métricas dos não-ofensores.

($tau = -0,23; SD(Standard\ Deviation) = 1\%$). Após esse ponto, o uso de reclamações aumenta com o desempenho dos grupos não-contaminados, mostrando valores bastante semelhantes aos dos grupos inimigos contaminados. Também deve-se observar que o uso de reclamações por ofensores é máximo ($media = 30\%, sd = 4\%$) em jogos onde os inimigos não os denunciaram (ou seja, inimigos não-contaminados), em comparação com partidas com denúncias inimigas, ou seja, inimigos contaminados ($media = 24\%, sd = 8\%$).

A Figura 6.10b revela que a contaminação dos grupos aliados e inimigos se relacionam de maneiras opostas ao uso de reclamações pelos ofensores, com correlações positivas e negativas, respectivamente (Tabela 6.6). Quando $contaminacao < 0,5$, existe um índice semelhante de uso de reclamações por ofensores, tanto para grupos aliados (26%) como para inimigos (média de 28%). Depois disso, a situação muda, uma vez que os ofensores reclamam mais em partidas com aliados altamente contaminados (média de 32%) do que com inimigos altamente contaminados (média de 20%).

- **Discussões:** A relação entre o uso de discussões e o desempenho dos ofensores e não-ofensores é exibida na Figura 6.11a. A Tabela 6.6 mostra que, embora o desempenho dos inimigos não-contaminados e contaminados se correlacionem positivamente com o uso de discussões por ofensores, a correlação com o desempenho dos aliados é, por uma pequena margem, não existente.

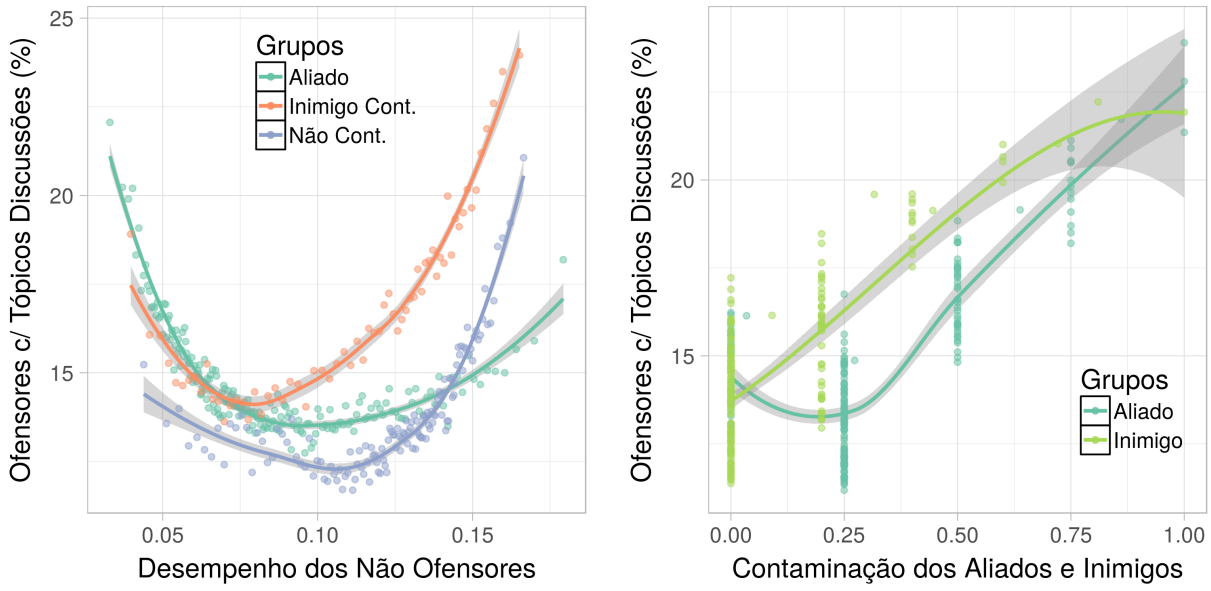
A Figura 6.11a mostra uma curva em formato de "U"para a relação entre uso de discussões e o desempenho. Para ambos tipos de grupos inimigos, podemos ver que o uso de discussões pelos ofensores diminui lentamente em níveis baixos de desempenho, mostrando correlação negativa para os não contaminados ($\tau = -0,54$) e os inimigos contaminados ($\tau = -0,70$), até os respectivos pontos de inflexão ($desempenho = 0,111$ e $desempenho = 0,069$, respectivamente). A partir deste ponto, o uso de discussões pelos ofensores aumenta consideravelmente. Para inimigos contaminados, eleva-se de 12% no menor uso para 21% no nível de desempenho mais alto, com correlação positiva no intervalo ($\tau = 0,81$). Para os inimigos não contaminados, o uso de discussões dos ofensores aumenta de 13% no ponto de inflexão para 24%, também com correlação positiva ($\tau = 0,91$).

Os grupos aliados, por outro lado, apresentam uma queda drástica no uso de discussões pelos ofensores até o ponto de inflexão, já que ele passa de um máximo de 22% no nível de desempenho mais baixo para um mínimo de 13% em $desempenho = 0,095$, com correlação negativa ($\tau = -0,80$) neste intervalo. Após o ponto de inflexão, eles mostram uma tendência ascendente até 18%, com correlação positiva ($\tau = 0,65$).

A Figura 6.11b mostra a relação entre o uso de discussões pelos ofensores e a contaminação de não-ofensores. O uso de discussões pelos ofensores mostra correlação positiva apenas quanto à contaminação dos grupos inimigos (Tabela 6.6). No entanto, a contaminação dos grupos inimigos e aliados aumenta com o do uso de discussões pelos ofensores. Observe que discussões são o único tópico em que o uso aumenta ao longo da contaminação, com aliados e inimigos tendo tendências opostas em todos os outros tópicos negativos.

A falta de correlação entre a contaminação aliada e o uso de discussões pelo ofensor em zero de contaminação pode ser explicada por uma queda média de 1,5% no uso de discussões no intervalo de contaminação de zero a 0,25, como mostrado na Figura 6.11b. Após essa queda, o uso de discussões se correlaciona fortemente com a contaminação ($\tau = 0,72$ para $contaminacao \geq 0,25$).

- **Insultos:** A relação entre o uso de insultos por ofensores e o desempenho de não-ofensores é mostrada na Figura 6.12a. A Tabela 6.6 mostra que apenas o desempenho dos inimigos não-contaminados se correlaciona negativamente com o uso de insultos do ofensor.



(a) Desempenho x Discussões.

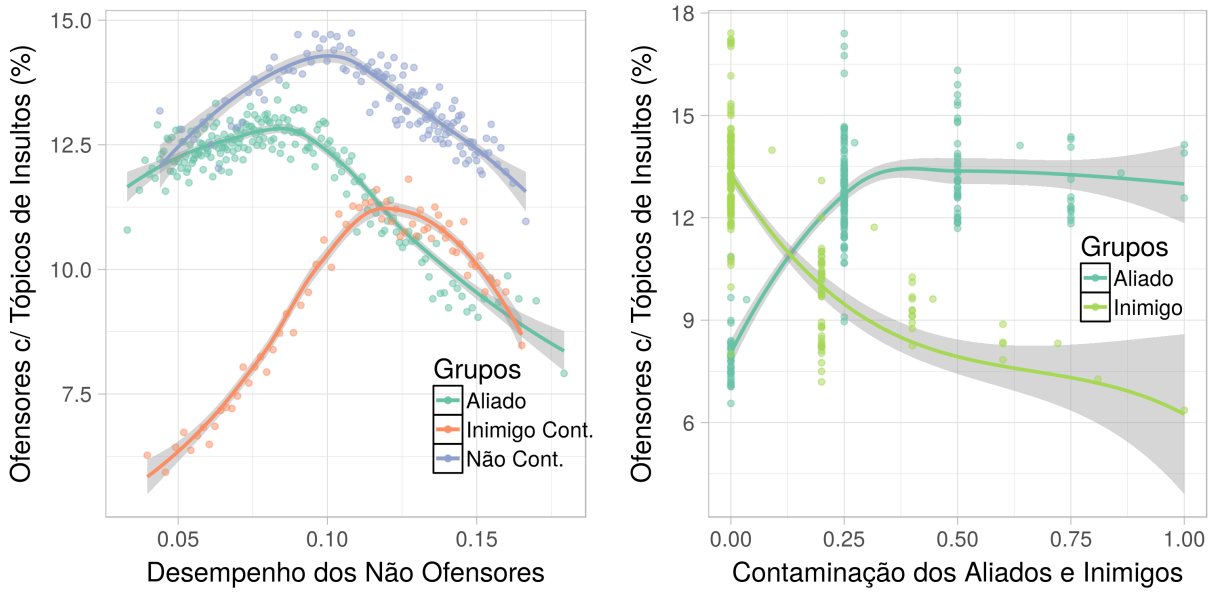
(b) Contaminação x discussões.

Figura 6.11: Relação entre tópicos de discussões dos ofensores e métricas dos não-ofensores.

A Figura 6.12a revela que, para todos os tipos de grupo, o uso de insultos pelos ofensores aumenta do menor nível de desempenho até um ponto de inflexão, quando o uso começa a diminuir com o aumento do desempenho. O uso de insultos por ofensores atinge seu pico em $desempenho = 0,108$ para grupos não-contaminados, $desempenho = 0,127$ para grupos de inimigos contaminados e $desempenho = 0,086$ para grupos aliados. As partidas de inimigos não-contaminados estão relacionadas ao maior uso de insultos pelos ofensores, passando de um valor médio inicial de 12% para 14 % ($\tau = 0,64$). Após o pico, o uso diminui para um valor mínimo de 10%, com forte correlação negativa ($\tau = -0,67$). Os inimigos contaminados mostram um ritmo de crescimento mais rápido no uso de insultos pelos ofensores, passando de 6% de uso médio a 12% no pico ($\tau = 0,85$), quando cai até 8 % ($tau = -0,68$).

Os grupos de aliados mostram um crescimento lento antes do pico (de 11 % de uso médio para 14 %), mas esse crescimento não é suficientemente estável para caracterizar uma correlação positiva $(\tau = 0,41)$. Após o pico, o uso de tópicos de insulto cai rapidamente (8% de uso médio), desta vez com forte correlação negativa $(\tau = -0,81)$.

A Figura 6.12b mostra como o uso de insultos pelos ofensores associa-se a contaminação dos não-ofensores. A correlação, que é negativa, é observada apenas para os inimigos (Tabela 6.6), em um comportamento semelhante às reclamações. No que diz respeito aos aliados, o uso de insultos pelos ofensores cresce de uma média de 8% para 13% em torno de $contaminacao = 0,50$, e nesse intervalo, a contaminação dos aliados e os insultos dos



(a) Desempenho x insultos.

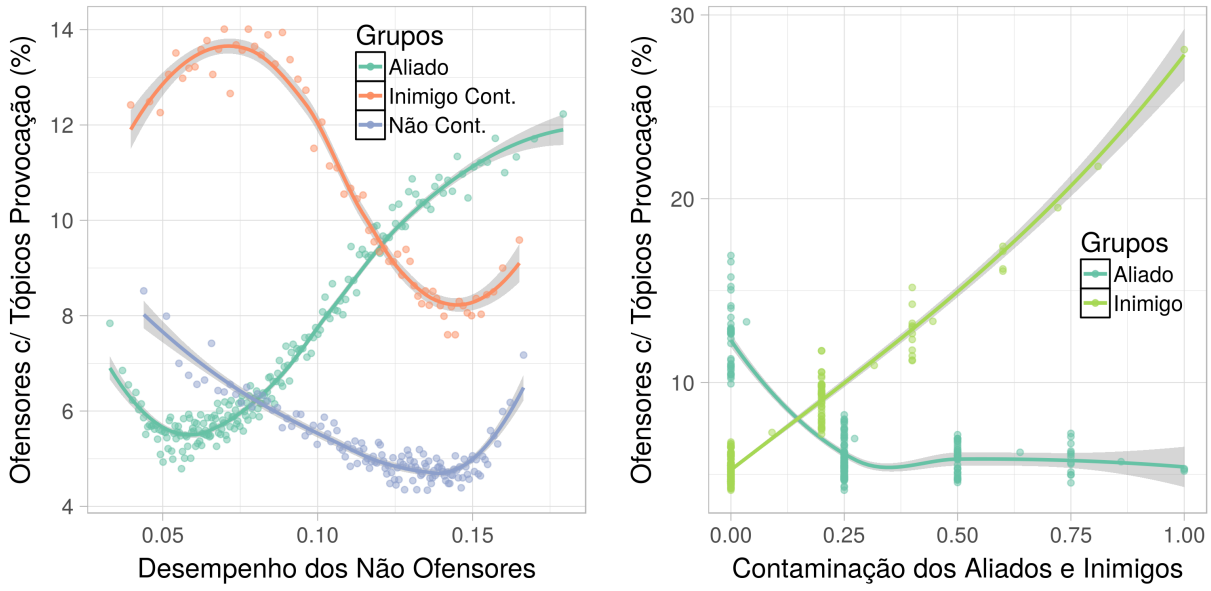
(b) Contaminação x insultos.

Figura 6.12: Relação entre tópicos de insultos dos ofensores e métricas dos não-ofensores.

ofensores estão intimamente correlacionados ($\tau = 0,72$). A partir daí, o uso permanece quase constante apesar da crescente contaminação dos aliados ($tau = 0,20, sd = 1\%$).

- **Provocações:** Finalmente, a relação entre o uso de provocações pelo ofensor e o desempenho de não ofensores é mostrada na Figura 6.13a. A correlação é positiva em relação ao desempenho dos grupos aliados e negativa em relação ao desempenho dos inimigos não-contaminados e contaminados (Tabela 6.6).

  Um olhar mais atento à Figura 6.13a revela algumas peculiaridades, que estão relacionadas a partidas muito desbalanceadas. Estas são representadas por pontos de inflexão em que as relações diferem da tendência geral. Para os inimigos contaminados e não-contaminados, o ponto de inflexão está em uma região de alto desempenho ($desempenho \approx 0,14$), após o qual o uso de provocações pelo ofensor aumenta com o desempenho (3pp e 2pp para inimigos não-contaminados e contaminados, respectivamente). Ao contrário dos demais, o grupo de inimigos contaminados mostram uma curva um pouco distinta, mais se assemelhando a um "S", com outro ponto de inflexão em baixo desempenho ($desempenho = 0,074$), antes do qual o uso de provocações pelos ofensores mostra um ligeiro aumento com crescimento do desempenho (2pp). O ponto de inflexão para grupos aliados ocorre em uma área de baixo desempenho ($performance = 0,056$), resultando em uma diminuição da provocação (2 pp).



(a) Desempenho x provocações.

(b) Contaminação x provocações.

Figura 6.13: Relação entre tópicos de provocação dos ofensores e métricas dos não ofensores.

A Figura 6.13b descreve a relação entre ofensores usando provocações e a contaminação dos grupos não ofensores. A Tabela 6.6 mostra que a contaminação dos grupos inimigos se correlaciona positivamente com provocações de ofensores. Já a tendência para os aliados é bastante distinta. Observa-se um uso muito maior de provocações a contaminação é zero (12 %), mas quando $contaminacao \geq 0,25$, o uso de provocação do defensor permanece constante ($media = 6\%$, $SD = 0,9\%$).

Discussão: As Figuras 6.10 a 6.13 mostram efeitos de toxicidade diferentes em aliados e inimigos de acordo com o tipo de tópico negativo adotado pelo ofensor.

No que diz respeito às reclamações, as tendências descritas na Figura 6.10a, em conjuntos com as respectivas correlações positivas/negativas, revelam que os maiores usos de reclamações pelos ofensores estão relacionados a aliados de baixo desempenho e inimigos de alto desempenho. Supomos que isso aconteça porque os ofensores tendem a reclamar com seus colegas sobre o seu mau desempenho. Da mesma forma, inimigos com alto desempenho podem estimular a raiva dos ofensores, o que também pode resultar em reclamações contra seus colegas de equipe.

A Figura 6.11a mostra que esse mesmo fenômeno é parcialmente verdadeiro para discussões, já que, comparativamente, os maiores usos de discussões estão relacionados tanto com inimigos de alto desempenho quanto com aliados de baixo desempenho. Neste caso, no entanto, as queixas podem ser dissimuladas em discussões contra jogadores supostamente tóxicos que

estão comprometendo a partida. Na mesma figura, também vemos níveis de uso de discussões por ofensores bastante elevados em partidas com aliados de alto desempenho e inimigos contaminados de baixo desempenho. Uma causa provável desse comportamento é um aumento na autoconfiança dos ofensores devido a uma partida desequilibrada a favor de sua equipe.

Para insultos, a relação existente para reclamações é verdadeira apenas para aliados de baixo desempenho. Dado que os insultos podem ser uma forma de liberar o estresse, assumimos que esse comportamento também reflete que os ofensores estão chateados com seus aliados pelo fato de estarem perdendo a partida. No entanto, ao contrário das reclamações e discussões, o alto desempenho inimigo não está relacionados a um uso maior de insultos por parte dos ofensores. Acreditamos que esse comportamento ocorre pelos ofensores estarem conscientes do desempenho ruim de suas equipes, independentemente do desempenho de seus oponentes. Assim, os insultos são limitados apenas à sua equipe. No entanto, a análise de nossos dados não nos permite confirmar esta intuição.

Provocações, por outro lado, apresentam um comportamento oposto quando comparado a reclamações, conforme descrito na Figura 6.13a. Enquanto o uso de provocações aumenta com o desempenho dos aliados, ele diminui com o aumento do desempenho dos inimigos. Em outras palavras, é baixo quando o desempenho aliado é baixo, e alto quando o desempenho do inimigo é baixo. Acreditamos que ambas as situações inflam a autoconfiança dos ofensores, aumentando a probabilidade deles manifestarem essa confiança através de provocações, como forma de afirmar sua superioridade sobre os inimigos. No entanto, esse comportamento não é observado nos respectivos desempenhos extremos baixos/altos, isto é, $desempenho > 0,14$ para inimigos e $desempenho < 0,056$ para aliados. Nessas situações extremas, parece o uso de provocações parece uma forma de externar a própria frustração ou raiva, e não com o objetivo de provocar a equipe inimiga.

A toxicidade representada por reclamações e insultos mostra um padrão diferente quando comparada com provocações. Tanto reclamações como insultos feitos pelos ofensores afetam a contaminação dos aliados, mas não a dos inimigos. A contaminação dos aliados aumenta em conjunto com insultos dos ofensores até $contaminacao = 0,5$, quando permanece constante (Figura 6.12b). Isso indica que existe um limiar de tolerância por parte de grupos aliados aos insultos feitos por seus colegas ofensores. Acima deste limiar, o comportamento do ofensor é considerado inaceitável e propenso a ser reportado por seus colegas de equipe. A contaminação dos aliados aumenta com as reclamações de ofensores de forma bastante direta, como mostrado na Figura 6.10b, o que confirma que os aliados são realmente afetados pelas reclamações dos ofensores. A contaminação dos inimigos, por outro lado, correlaciona-se negativamente com

as reclamações e insultos dos ofensores. Concluímos que esta queda representa que inimigos não são afetados por insultos ou reclamações feitas por ofensores, pois estas são provavelmente direcionadas aos aliados.

Por outro lado, as provocações feitas pelos ofensores tendem a afetar apenas a contaminação dos inimigos. A Figura 6.13b mostra uma relação direta entre o aumento da contaminação em grupos inimigos e do uso de provocações pelos ofensores. Esta relação corrobora nossa conclusão anterior de que os grupos inimigos são a principal vítima da provocação dos ofensores. Os grupos aliados, por outro lado, estão associados a um uso bastante alto de insultos por ofensores, mesmo quando eles não fazem nenhuma denúncia (ou seja, $contaminacao = 0$). Acreditamos que isso também é uma forte evidência de que os aliados acabam usando provocações junto com os ofensores, portanto, toda a equipe acaba agindo de forma tóxica, como já discutido na Seção 6.3.2.

6.4 Relações temporais entre tópicos de conversação

Nesta seção, descrevemos os experimentos realizados para gerar regras de associação e descobrir se existem transições relevates entre os tópicos ao longo das partidas. Para filtrar as regras do algoritmo apriori como descrito na Seção 5.4, utilizamos as métricas de suporte, confiança e *lift*.

Definimos o suporte mínimo como 0,003 (0,3%), devido ao forte desbalanceamento na frequência dos tópicos em nossos dados. Este valor de suporte mínimo é igual a 10% da taxa de aparição do tópico de menor frequência (provocações), que corresponde somente a 3% dos documentos.

O *lift* foi definido como a distância de 0,3 do valor 1, com regras com *lift* acima de 1,3 e abaixo de 0,7 sendo aceitas. Consideramos que este valor seja alto o suficiente para estabelecer uma associação forte entre os *itemsets* da regra.

Já o valor de confiança foi definido empiricamente como 0,10 (10%). Foram testados valores no intervalo entre 0,80 (80%) e 0,10 (10%), buscando o maior valor possível, com o suporte e o *lift* já definidos como os valores especificados acima. Alguns destes testes estão no Apêndice C.

Os experimentos feitos nesta seção estão divididos em dois conjuntos. Cada conjunto contém o mesmo experimento repetido para cada um dos tipos de grupos de jogadores. Todos experimentos utilizaram os mesmos valores de suporte, confiança e *lift*. O primeiro conjunto de experimentos envolve envolve todas as transações, e tem seus resultados listados na Seção

6.4.1.

Os resultados destes primeiros experimentos não revelaram nenhum padrão de transição relevante e motivaram a execução de um segundo conjunto de experimentos, desta vez somente utilizando transações que possuem tópicos distintos nas colunas E e D. Os resultados deste segundo conjunto de experimentos são discutidos na Seção 6.4.2.

6.4.1 Experimentos 1: Todas as transações

O processo de descoberta de regras de transição entre tópicos de conversação foi realizado utilizando todas as 14.199.452 transações, preparadas como descrito na Seção 5.4 das quais 4.813.599 correspondem a tópicos usados por grupos aliados, 4.842.582 correspondem a tópicos usados por grupos inimigos, e 4.543.271 correspondem a tópicos usados por grupos ofensores. Os experimentos realizados resultaram em 8 regras para as transações dos grupos aliados, 9 regras para as transações dos inimigos, e 13 regras para as dos ofensores.

Uma análise sobre estas regras mostra que para quase todos os tópicos, as regras mais prevalentes levam do tópico a si mesmo (e.g. *Táticas* $\rightarrow$ *Táticas*), com valores consideravelmente altos de *lift*. Denominaremos estas regras de *auto-transição*. As regras de auto-transição, juntamente com suas métricas, estão listadas na Tabela 6.7.

As regras descobertas que fogem ao padrão de auto-transição estão listadas na Tabela 6.8 para aliados, inimigos e ofensores.

Tabela 6.7: Regras de auto-transição para todos os grupos.

	Aliados			**Inimigos**			**Ofensores**		
Auto-Transição	**Suporte**	**Confiança**	**Lift**	**Suporte**	**Confiança**	**Lift**	**Suporte**	**Confiança**	**Lift**
Táticas	0,063	0,30	1,95	0,079	0,34	1,97	0,043	0,25	2,00
Táticas/Educação	0,017	0,19	2,98	0,017	0,19	3,34	0,009	0,15	3,17
Bate-Papo	0,034	0,25	2,34	0,045	0,28	2,33	0,020	0,20	2,44
Reclamações	0,042	0,26	1,99	0,027	0,21	2,10	0,042	0,25	2,01
Discussões	0,006	0,12	3,35	0,005	0,11	1,40	0,007	0,13	2,89
Insultos	0,006	0,15	5,21	0,007	0,17	5,58	0,012	0,21	4,81
Provocações	-	-	-	-	-	-	0,004	0,12	4,70

Devemos frisar que não existem padrões com caracterização forte tanto nos experimentos desta seção como nos apresentados mais a frente, devido aos valores baixo de suporte e confiança apresentados por todas as regras. Entretanto, alguns padrões interessantes ainda emergem, e mesmo que não sejam particularmente fortes, serão analisados.

As prevalência de regras com o padrão de auto-transição mostra que o comportamento mais comum dos jogadores é utilizar o mesmo tópico ao longo da partida, mostrando que para os jogadores, há uma certa 'naturalidade' no uso destes tópicos. A única exceção é o tópico de provocações, que só mostra alguma repetição quando usado pelo ofensor, reforçando a toxicidade deste.

Vale também frisar que os *lifts* das regras de auto-transição dos tópicos de insultos e provocações são sensivelmente maiores $(media = 5.07)$ do que os *lifts* dos demais tópicos $(media = 2.41)$. Isso mostra uma tendência maior de de repetição destes tópicos, o que consideramos uma evidência de uma espiral de insultos dentro de uma equipe. Por parte do ofensor, também temos uma espiral de provocações.

Tabela 6.8: Demais regras para Aliados, Inimigos e Ofensores

LHS	**RHS**	**Suporte**	**Confiança**	**Lift**
Aliados				
Táticas	Reclamações	0,033	0,16	1,35
Discussões	Bate-papo	0,006	0,14	1,31
Inimigos				
Táticas	Reclamações	0,031	0,14	1,34
Taticas/Edu.	Bate-papo	0,015	0,17	1,38
Discussões	Reclamações	0,006	0,14	1,40
Ofensores				
Táticas	Reclamações	0,031	0,18	1,43
Taticas/Edu	Táticas	0,011	0,17	1,34
Reclamações	Táticas	0,027	0,16	1,32
Discussões	Reclamações	0,010	0,17	1,34
Discussões	Bate-papo	0,007	0,11	1,44
Provocações	Reclamações	0,006	0,17	1,36

Quanto às regras que não são de auto-transição (Tabela 6.8), vemos um número de regras pequeno para aliados e inimigos (2 e 3 regras, respectivamente). Já os ofensores estão relacionados a uma quantidade maior, totalizando 6 regras. Uma das regras na tabela se destaca (*Táticas* $\rightarrow$ *Reclamações*), pois aparece em todos os grupos, e antecipa uma tendência comum de transição entre estes tópicos.

A regra *Discussões* $\rightarrow$ *Bate-papo* aparece em aliados e ofensores, e mostra um caso de um tópico negativo transicionando para um positivo, possivelmente como caso de sucesso

do apaziguamento realizado ao longo dos bate-papos relacionados ao tópico de discussões. Regras caracterizando transições entre tópicos serão exploradas em mais detalhes no próximo experimento.

6.4.2 Experimentos 2: Transações com tópicos distintos

Enquanto a evidência de que o comportamento mais comum dos jogadores é manter as conversas centradas em torno de um mesmo tópico durante a partida é bastante clara, ainda buscamos saber como as transições entre os tópicos distintos acontecem. Contudo, o grande número de auto-transições acaba ocultando as demais transições na execução do apriori. Para resolver esse problema, elaboramos outro experimento, com parâmetros similares ao primeiro, buscando descobrir os principais padrões de transição entre tópicos distintos. Para isso, foram removidas todas as transações onde os tópicos em t_i e t_{i+1} eram iguais, resultando em 9.697.309 transações. Destas, 3.259.621 correspondem a tópicos usados por grupos aliados, 3.193.237 correspondem a tópicos usados por grupos inimigos, e 3.244.451 correspondem a tópicos usados por grupos ofensores.

Este novo experimento resultou em 8 regras para os aliados, 10 regras para os inimigos e 10 regras para os ofensores, representadas nas Figuras 6.14 a 6.16. Nestas figuras, as setas representam regras de transição entre tópicos, com os valores na parte superior demonstrando o suporte e o *lift* (em laranja) daquela regra. O valor na parte inferior da seta corresponde ao valor da confiança em porcentagem. O percentual abaixo do nome do tópico é a soma das confianças de todas as regras com aquele tópico no LHS. Este valor indica a probabilidade de uma das transições apontadas na figura, entre aquelas que têm por origem aquele tópico, acontecer.

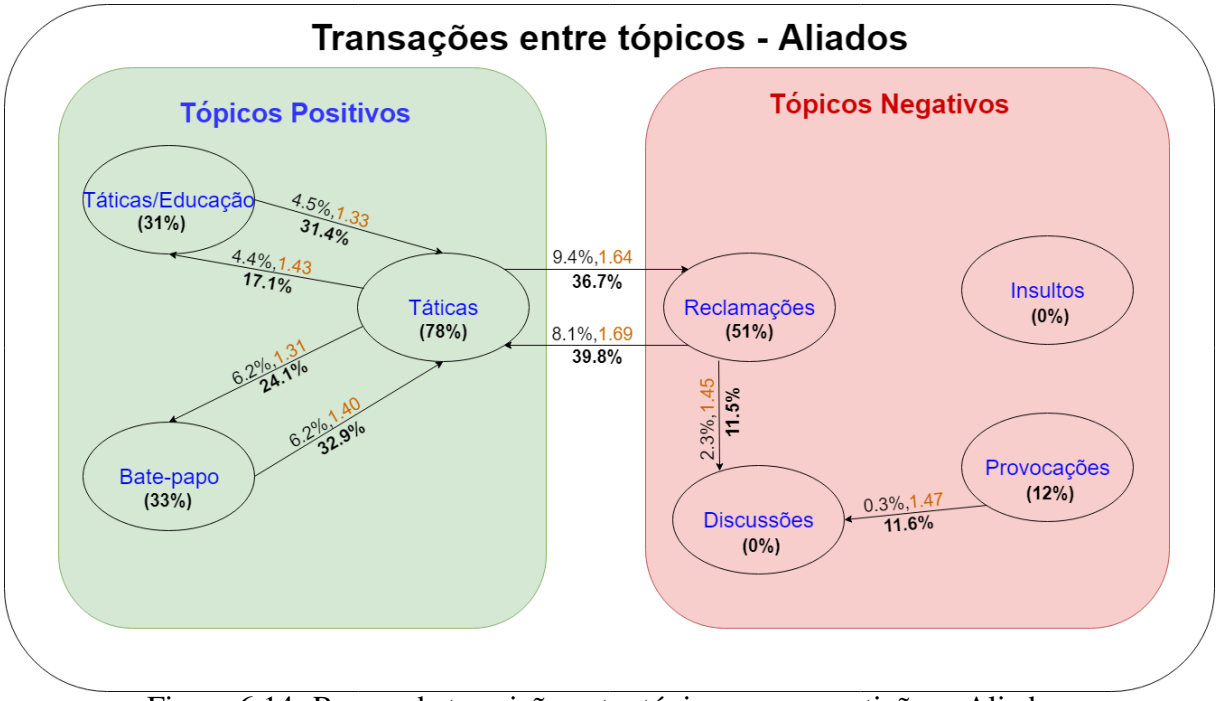


Figura 6.14: Regras de transição entre tópicos, sem repetições - Aliados.

Nas Figuras 6.14 e 6.15 podemos ver que, em grupos não-ofensores, Táticas aparece em regras com todos os outros tópicos positivos. Nestes grupos de jogadores, o tópico de Táticas possui um posicionamento central entre tópicos positivos, e cria também uma ponte entre estes e os tópicos negativos.

O tópico de Táticas é o único ponto de ligação entre os diferentes tópicos positivos, não havendo ligações entre Táticas/educação e Bate-papo, exceto nos grupos inimigos. Isso faz do tópico Táticas um ponto central entre tópicos positivos, com os jogadores alternando entre o foco no jogo, e a promoção de um ambiente favorável.

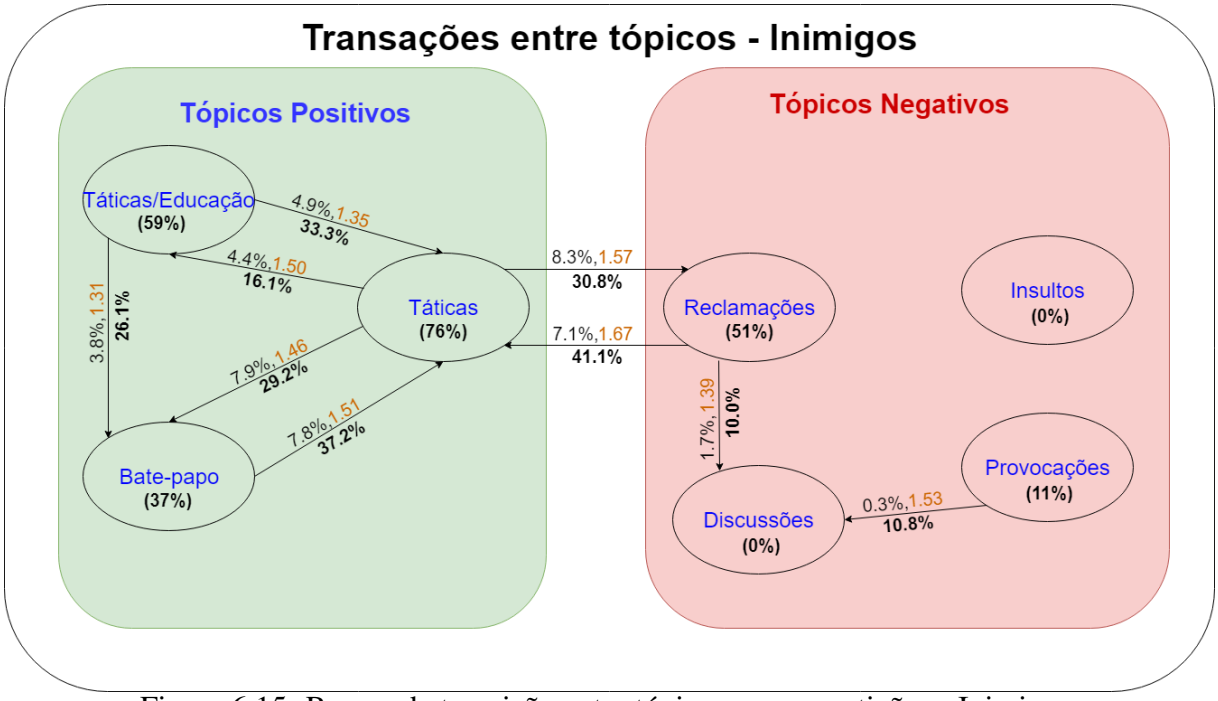


Figura 6.15: Regras de transição entre tópicos, sem repetições - Inimigos.

Também podemos observar o tópico de Táticas como a principal ponte entre tópicos positivos e tópicos negativos, por causa de sua relação com Reclamações. Podemos ver que as regras que ligam os tópicos de Táticas e Reclamações possuem valores médios de confiança $(0,37)$ e *lift* $(1,64)$, valores acima da média geral para todas as regras que são de $0,25$ ($SD = 0,10$) para a confiança e de $1,45$ ($SD = 0,12$) para o *lift*. Isso nos dá indícios de que esse trajeto circular entre Táticas e Reclamações é o tipo de transição entre tópicos mais comum em partidas, com os jogadores alternando entre chamadas táticas, simplesmente jogando o jogo, e reclamações a outros jogadores, provavelmente quando estes cometem algum erro na partida, ou apresentam desempenho abaixo do esperado. Transações contendo Táticas e Reclamações correspondem a 16% do total de todas as transações (incluindo repetições), o que é mais do que o dobro da probabilidade destes tópicos aparecerem na mesma transação ao acaso (7,2%).

No geral, podemos ver Táticas como uma espécie de 'ponto morto', um tópico que é utilizado quando não há nenhum acontecimento como manifestações de comportamento tóxico ou benigno, ou algumas ações feitas por jogadores dentro do jogo, que podem incentivar os jogadores de um grupo utilizar outros tópicos, negativos ou positivos. Nossas observações são reforçadas quando vemos que há uma alta chance de uma transação originada de Táticas corresponder a um dos/das padrões/regras citados anteriormente (78% para inimigos, 76% para aliados).

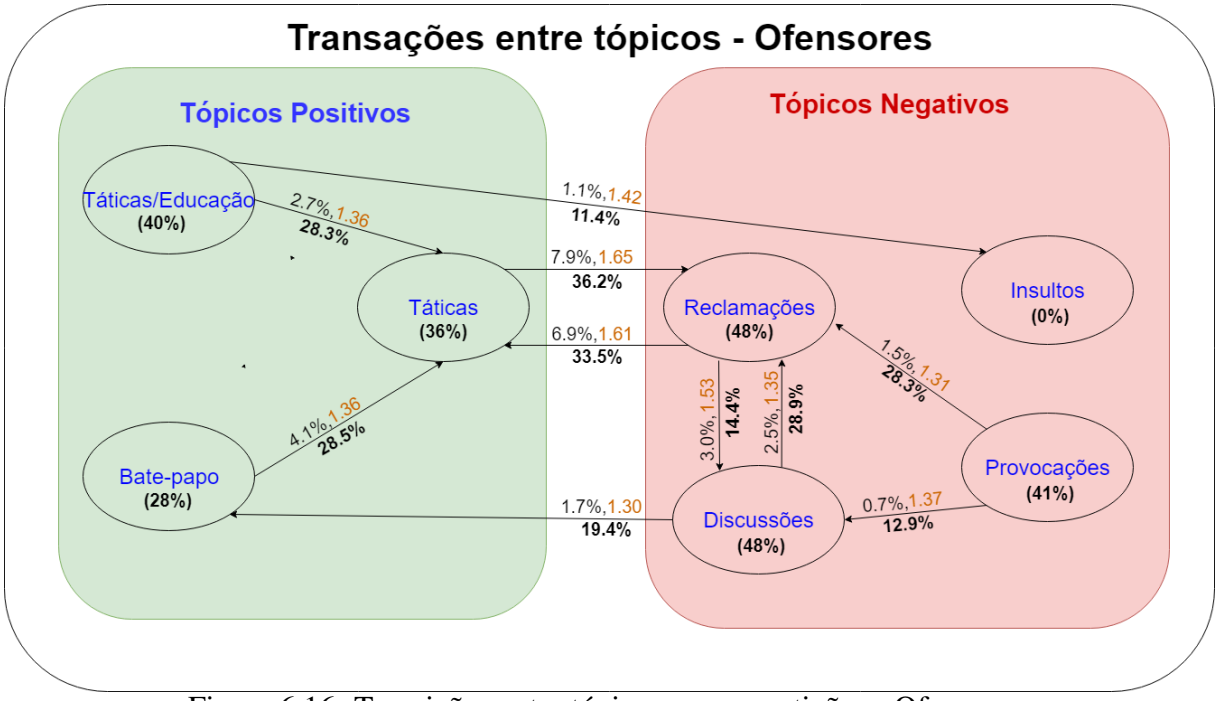


Figura 6.16: Transições entre tópicos, sem repetições - Ofensores.

Já em grupos ofensores, vemos que Táticas não possui este papel central na transição de tópicos tão representativo, como mostrado na Figura 6.16. Não há transições partindo de Táticas aos tópicos positivos, e a quantidade de transições relevantes originadas de Táticas diminui drasticamente para 36%. Ainda vemos que os ofensores mostram uma regra particular: uma ligação entre Discussões e Bate-papo. As motivações por trás desta regra não são claras, mas ela reforça o enfraquecimento da posição central do Táticas em grupos ofensores. O enfraquecimento do tópico de Táticas, com a sua falta de transições a outros tópicos positivos, e o menor número de regras associadas com tópicos positivos no geral, também reforça a ideia de que o ofensor é menos predisposto a tópicos positivos, como já citado na Seção 6.2.

Ainda olhando para as Figuras de 6.14 a 6.15, vemos que, para não ofensores, o tópico de Discussões é antecedido tanto por Reclamações, como por Provocações. Isso sinaliza que as provocações acabam sendo amenizadas, passando a serem discussões. Reclamações, por outro lado, podem acabar agravando-se em discussões.

No caso dos ofensores, como mostrado na Figura 6.16, temos que Discussões também são antecedidas tanto por Provocações como por Reclamações, mostrando o mesmo efeito de amenização. Contudo, devido à transição de Discussões a Reclamações, o ofensor pode acabar preso em um círculo vicioso de comportamento negativo, alternando repetidamente entre Reclamações e Discussões.

Para grupos não ofensores, observamos que não existem padrões quanto ao uso de tópicos que antecedem o uso de Insultos, como mostrado nas Figuras 6.14 e 6.15. Nestes grupos, o

uso de Insultos não aparenta ser motivado por nenhum outro tópico provindo do mesmo grupo em um momento anterior da partida.

Já para os ofensores (Figura 6.16), uma das regras mostra que o uso de Insultos é precedido por Táticas/educação. Uma possível explicação é que, apesar de se mostrar educado através do uso de Táticas/educação, o ofensor pode se irritar facilmente com o que ele entende como outros jogadores não seguindo suas instruções táticas. Isso pode disparar uma reação tóxica por parte dele, insultando os outros membros da equipe pela incapacidade ou falta de vontade destes de seguir suas ordens.

Nas mesmas figuras, vemos que, para grupos não ofensores, não há nenhum padrão de transição para o tópico de Provocações. Isso fortalece a noção da provocação como algo gerado por elementos externos ao grupo, ou da própria malícia do jogador provocando, como citado na Seção 6.3.2.

Já em grupos ofensores, a probabilidade de uma transição originada de Provocações usadas por jogadores ofensores corresponder a um padrão é de 41%, enquanto a mesma probabilidade para aliados e inimigos é de 12%, e 11%, respectivamente. Isso demonstra que este tópico é mais natural aos ofensores. Essa ideia é reforçada pelo fato de que os ofensores são o único grupo onde há uma tendência de uso contínuo do tópico de Provocações, como mostrado na Tabela 6.7.

Nossos resultados mostraram que devido aos valores de suporte e confiança baixos apresentados pelas regras, não existem padrões dr transição entre tópicos que sejam muito representativos. Também vimos que o comportamento mais comum aos jogadores em um grupo é manter a conversa centrada no mesmo tópico durante uma partida, devido aos altos valores de *lift*, em comparação com as outras regras, das regras de auto-transição.

Nas situações em que há transições entre padrões de conversação, vimos que vimos que a mudança de comportamento tóxico para não tóxico realmente ocorre, como mostrado por (KWAK; BLACKBURN, 2014). Contudo, ela aparenta acontecer com uma intensidade muito menor do que a mostrada no estudo. Esta mudança aparenta ocorrer de maneira gradual em grupos não ofensores, que apresentam padrões de transições bastante similares, tendendo a passar pelos tópicos de tática e reclamação, com nenhum padrão de mudança de comportamento brusca, como por exemplo de táticas ou bate-papo para insultos, sendo revelada. Comportamento negativo apresentado por jogadores não-ofensores, aparenta ser transitório, e pode reverter-se para o tópico de táticas com relativa normalidade.

Já em grupos ofensores temos transições menos graduais e centralizadas entre tópicos positivos e negativos. Os padrões de transição dos ofensores mostraram-se mais centrados em

tópicos negativos, com poucos padrões de mudança com destino em tópicos positivos sendo encontrados. Os ofensores também são o único tópico que apresentam algum padrão que leva a Insultos, e apresentam um ciclo entre Reclamações e Discussões que caracteriza um provável ciclo vicioso de comportamento negativo. Ofensores também apresentam alguns padrões de reversão de comportamento negativo, como resultado de um tópico de discussões bem sucedido, ou por utilizar a táticas após reclamações. Contudo, neste segundo caso, o ofensor não parece disposto a colaborar de forma mais ativa com a equipe, já que ele não tende a voltar a usar tópicos como táticas/educação ou bate-papo.

Nenhuma transição que leve ao, ou parta do tópico de Insultos foi encontrada em grupos não ofensores, e somente uma em grupos ofensores, o que faz deste um tópico unicamente isolado, e de uso imprevisível. Isso é um resultado preocupante, já que insultos são uma das formas mais severas de comportamento tóxico, e tal resultado mostra a dificuldade de se prever o uso deste tópico durante uma partida.

6.5 Tópicos e emoções

Nesta seção apresentamos os resultados da construção do léxico de emoções específico ao domínio de MOBAs, buscando validar seus resultados A avaliação foi feita comparando-o com o seu léxico gerador, o NRC, bem como através de um processo de análise manual das emoções encontradas. Após, utilizaremos este léxico para descrever os tópicos de conversação de acordo com as emoções de Plutchik, descobrindo quais emoções são mais presentes nestes tópicos.

6.5.1 Resultados do modelo de classificação de sentimentos

Treinamos nosso modelo utilizando três algoritmos, SVM, MLP e regressão logística, de acordo com a preparação de dados e parâmetros descritos na Seção 5.5.1. Os resultados estão descritos na Tabela 6.9, em termos de precisão, recall e Medida F. Foram incluídos valores médios para os dois tipos de sentimento, calculados como uma média simples das diferentes classe.

Para escolhermos o melhor modelo, estabelecemos uma comparação usando como medida a Micro-F1, e como *baseline* o trabalho publicado por Bravo-Marquez et al. Bravo-Marquez et al. (2016), que utiliza um algoritmo de regressão logística com regularização usando norma

Tabela 6.9: Resultados dos modelos MLP/SVM/RL, para cada classe de sentimento e emoção

-	**MLP**			**SVM**			**Reg. Logistica**		
Emoção	**Precisão**	***Recall***	**Medida-F**	**Precisão**	***Recall***	**Medida-F**	**Precisão**	***Recall***	**Medida-F**
Positivo	0.467	0.531	0.495	0.419	0.581	0.487	0.581	0.319	0.412
Negativo	0.534	0.743	0.621	0.539	0.682	0.602	0.684	0.490	0.571
Polaridades (Média)	0.500	0.637	0.558	0.479	0.631	0.544	0.632	0.404	0.491
Alegria	0.652	0.352	0.458	0.270	0.576	0.368	0.472	0.200	0.280
Antecipação	0.361	0.186	0.246	0.163	0.494	0.245	0.166	0.044	0.069
Confiança	0.465	0.268	0.340	0.259	0.634	0.367	0.363	0.130	0.191
Medo	0.460	0.522	0.489	0.302	0.750	0.431	0.611	0.250	0.354
Nojo	0.390	0.554	0.458	0.262	0.723	0.385	0.461	0.289	0.355
Raiva	0.430	0.550	0.482	0.309	0.702	0.430	0.611	0.369	0.460
Surpresa	0.429	0.138	0.209	0.142	0.508	0.221	0.260	0.092	0.136
Tristeza	0.353	0.427	0.386	0.289	0.690	0.407	0.435	0.245	0.313
Emoções (Média)	0.442	0.374	0.383	0.249	0.634	0.356	0.346	0.202	0.319

Tabela 6.10: Micro F-Measure dos modelos construídos

Modelo (Corpus utilizado)	**Micro Medida-F**
Regressão Logística L2 (tweets)	0.473
Regressão Logística L2 (MOBA)	0.375
SVM (MOBA)	0.447
MLP (MOBA)	0.475

L2. Os resultados dos diferentes modelos, validados usando o conjunto de validação, são mostrados na Tabela D.1. Estão listados tanto o resultado original do *baseline*, feito sobre um corpus de tweets, como o resultado obtido executando o algoritmo do *baseline* sobre nosso corpus, identificado como MOBA, além da Micro-F1 para os dois modelos destacados na Tabela 6.9.

Vemos que o melhor resultado no corpus de conversas MOBA é atingido pelo MLP, seguido pelo SVM, e com o algoritmo do baseline aplicado sobre nosso corpus aparecendo por último. Note que o MLP atinge resultados similares ao *baseline* aplicado sobre o corpus de tweets no trabalho original de Bravo-Marquez et al. (2016). Assim, escolhemos o MLP como nosso modelo para a construção do léxico.

Ainda assim, vemos que os valores de Medida F dos sentimentos não são altos o suficiente para aceitarmos o léxico como válido sem fazer algumas análises extras. Os valores das emoções *antecipação* (0,246) e *surpresa* (0,209) mostram-se especialmente baixos quando comparados os demais. Para nos certificar da qualidade de nosso léxico, e que ele produzirá resultados razoavelmente confiáveis para as análises a seguir, comparamos os rótulos de emoções das palavras presentes neste com os rótulos presentes no NRC, e fizemos uma análise preliminar das palavras de cada emoção buscando por falsos positivos. Esta última análise deve ser considerada preliminar, pois foi feita pelo próprio autor, devendo ser estendida a outras opiniões.

6.5.2 Análise Quantitativa Comparando o NRC e o Léxico de MOBAs

Note que poderíamos justificar quaisquer diferenças entre os léxicos MOBAs e o NRC pelo fato de que os dois léxicos tratam de domínios diferentes. Para diminuir esse viés, utilizamos para essa comparação somente as palavras do NRC que também estão presentes no corpus de bate-papo em LoL.

Para determinar se os sentimentos presentes no léxico de MOBAs variam de alguma maneira anômala face a um léxico já estabelecido e validado como o NRC, comparamos a frequência de palavras rotuladas com sentimentos nestes léxicos, sendo os resultados descritos na Tabela 6.11. A tabela contém, para cada um dos léxicos, a quantidade de palavras rotuladas com cada um dos sentimentos, a proporção destas palavras no total de palavras de emoção, além da diferença em pontos percentuais entre os dois.

Quanto às polaridades, observamos que o número absoluto de palavras aumentou, e que são majoritariamente negativas. Contudo, em termos de proporção, a polaridade negativa aumenta em 8,8 pontos percentuais do NRC para o léxico de MOBAs, sendo que as positivas diminuem na mesma proporção. Essa maior representação pode derivar tanto de uma característica do próprio domínio, como de erros de classificação, hipóteses que só poderiam ser confirmadas com uma análise manual mais criteriosa, fora do escopo deste trabalho. Para garantir a coerência de nossos resultados, optamos por excluir a polaridade da caracterização dos tópicos em termos de sentimento.

Tabela 6.11: Frequências de sentimentos nos léxicos

Sentimento	**# NRC**	**% NRC**	**# MOBA**	**% MOBA**	**%NRC-%MOBA**
Polaridades					
Positivo	692	43,7%	1681	35.0%	8.8pp
Negativo	889	56,3%	3127	65.0%	-8.8pp
Emoções					
Alegria	243	09.39%	367	6.6%	2.8pp
Antecipação	278	10.7%	509	09.1%	1.5pp
Confiança	382	14.7%	639	11.5%	3.2pp
Medo	436	16.8%	1182	21.2%	-4.4pp
Nojo	298	11.5%	837	15.0%	-3.5pp
Raiva	407	15.7%	976	17.5%	-1.8pp
Surpresa	188	07.2%	258	4.6%	-2.6pp
Tristeza	355	13.7%	791	14.2%	-0.5pp

Já a diferença entre as proporções de emoções, também descritas na Tabela 6.11, não aparentam ser muito significativas, com a maior diferença mostrando-se no tópico de *medo*, 4,4 pontos percentuais apenas.

A mineração de emoções utilizando um modelo de emoções básicas é um problema multi-rótulo. Assim, um outro critério de qualidade seria o número de rótulos associado a cada palavra em ambos os dicionários. Assim, para melhor nos certificarmos da qualidade de nossas emoções, também verificamos se a quantidade de emoções por palavras no NRC e no léxico de MOBAs são similares.

Tabela 6.12: Distribuição da quantidade de emoções em palavras de sentimento nos léxicos

Métrica	**NRC**	**MOBAs**
Frequência	1342	3104
Média	1.93	1.79
SD	1.16	1.05
Min.	1.00	1.00
25%	1.00	1.00
Mediana(50%)	2.00	1.00
75%	3.00	2.00
Máximo	8.00	8.00

As distribuições da quantidade de emoções de palavras rotuladas com alguma emoção para cada léxico estão descritas na Tabela 6.12. Um teste t comparando estas distribuições mostrou que elas são diferentes $(t = 3,85;\ p = 0,0001)$. Mas considerando tanto o t resultante, como as médias e desvios padrão destas distribuições, vemos que esta diferença não é grande, sendo que o léxico de MOBAs apresenta uma quantidade levemente menor de emoções por palavras do que o NRC. Assim, podemos concluir que nosso modelo é mais conservador na atribuição de rótulos, e que algumas emoções podem estar levemente sub-representadas, mas não o suficiente para invalidar os resultados.

Tabela 6.13: Palavras de sentimento do top500 de cada tópico para cada léxico

Emoção		**# NRC**	**# MOBA**	**#MOBA-#NRC**		**# NRC**	**# MOBA**	**#MOBA-#NRC**
Alegria	**Todas as palavras**	85	155	70	**Palavras únicas**	19	31	12
Antecipação		140	248	108		26	43	19
Confiança		102	246	144		23	50	28
Medo		187	411	224		30	70	42
Nojo		132	253	121		28	64	37
Raiva		165	301	136		31	57	27
Surpresa		74	83	9		13	16	4
Tristeza		195	345	150		36	62	28

Além disso, para verificarmos quantas palavras novas relevantes o léxico de MOBAs acrescenta sobre o NRC, agrupamos as palavras rotuladas com cada emoção presentes nas top 500 palavras de cada um dos tópicos. Destas, separamos as palavras presentes somente no NRC, as palavras presentes no dicionário de MOBAs (que inclui as palavras do NRC), e as palavras presentes **somente** no dicionário de MOBAs (excluindo as palavras do NRC presentes no dicionário de MOBAs). A quantidade de palavras nestes grupos para cada uma das emoções, incluindo palavras repetidas (i.e. associadas a mais de uma emoção), está descrita na Tabela 6.13, em conjunto com as palavras de sentimento únicas para cada um destes grupos. Vemos que a quantidade de palavras adicionadas para cada emoção é significativa, exceto para a emoção de surpresa. Isso pode ser explicado pelo baixo *recall* de classificação desta emoção, ou pela baixa presença de palavras com a emoção surpresa em partidas de MOBAs.

Tabela 6.14: Resultados da rotulação manual de palavras de sentimento

Emoção	**# Palavras**	**# FPs**	**% FPs**
Alegria	12	1	8,3%
Antecipação	19	7	36,8%
Confiança	28	6	21,4%
Medo	42	11	26,1%
Nojo	37	7	18,9%
Raiva	27	4	14,8%
Surpresa	4	0	0,0%
Tristeza	28	2	7,1%

6.5.3 Análise Subjetiva Preliminar do Léxico de MOBAs

Para conseguir um veredito final sobre a qualidade da classificação do dicionário, analisamos manualmente as palavras presentes somente no dicionário de MOBAs, definindo quais destas palavras associadas a cada emoção são falsos positivos. Os resultados deste experimento estão descritos na Tabela 6.14. Esta análise, como já mencionado é preliminar, pois foi realizada apenas pelo autor. No Apêndice E estão disponíveis, para cada emoção, uma nuvem de palavras contendo as palavras analisadas, bem como as palavras consideradas como falsos positivos.

Verificando os resultados descritos na Tabela 6.14, juntamente com os experimentos da seção anterior, decidimos desconsiderar as emoções de *surpresa* e *antecipação* em nossa caracterização dos tópicos. Estas duas emoções apresentam baixos valores de Medida-F. A emoção *surpresa* tem somente 4 palavras únicas no léxico de MOBAs, o que faz sentido pelo baixo valor de *recall* desta emoção, nos fazendo crer que esta emoção é pouco relevante ao problema. Já a emoção *antecipação* apresenta uma boa quantidade de palavras únicas no léxico, apesar de seu *recall* ser também baixo. Contudo, esta emoção apresenta uma alta taxa de palavras classificadas como falso positivo (36,8%), o que nos leva a crer que o dicionário é pouco confiável quanto a esta emoção, e por esta razão, esta emoção não deve ser considerada em nossa análise.

Finalmente, estes experimentos não substituem a necessidade de uma validação mais completa do léxico de MOBAs por especialistas, nem o tornam completamente confiável. Contudo, elas estabelecem um grau de qualidade mínimo ao léxico, necessário para a utilização deste na caracterização das emoções dos tópicos, que conclui esta seção.

6.5.4 Atribuição de emoções aos tópicos

Os valores de emoções correspondentes a cada tópico, produzidos a partir do processo descrito na Seção 5.5.2 são mostrados na Figura 6.17, no formato de um gráfico em radar para cada tópico. Para cada gráfico, os eixos indicando o valor de cada emoção para aquele tópico.

Os valores das emoções para cada tópico são detalhados na Tabela 6.15, bem como a soma dos valores de todas emoções para todos os tópicos, e a soma de todos os valores referentes a cada emoção. As Figuras 6.18 e 6.19 fazem uma comparação destas somas, para os valores de emoção por tópico e por classe de emoção, respectivamente. Cabe lembrar que tanto a polaridade, quanto as emoções *surpresa* e *antecipação* foram desconsideradas nesta análise.

Embora o valor das emoções não seja muito alto de uma forma geral, é possível notar na Figura 6.17 que existe para a maioria dos tópicos uma emoção mais presente, bem como valores muito baixos de certas emoções. Para os tópicos positivos, vemos a ausência das emoções *raiva*, *nojo* e *tristeza*. As emoções *confiança* e *alegria* são predominantes nos tópicos Táticas e Táticas/Educação. Já o tópico Bate-papo não possui nenhuma emoção predominante. Dentre os tópicos negativos, temos o *nojo*, a *raiva* e a *tristeza* como emoções predominantes nos tópicos Provocações, Insultos e Reclamações, respectivamente. Já as emoções *alegria* e *confiança* são bem baixas nestes tópicos. Por outro lado, o tópico Discussões possui a *confiança* como emoção predominante. Uma discussão mais aprofundada desta prevalência ou ausência de emoções será feita ao longo desta seção.

Considerando os totais do conjunto de emoções de cada tópico, vemos na Figura 6.18 que os valores de emoções se dividem em dois patamares: um mais baixo, com valores de emoção variando entre 2 e 2,5, que corresponde ao total das emoções dos tópicos positivos, e um mais alto, com os valores de emoção variando entre 3 e 4, correspondente ao total das emoções

Tabela 6.15: Valores das emoções para cada tópico

	Alegria	**Confiança**	**Medo**	**Tristeza**	**Nojo**	**Raiva**	**Total Tópicos**
Táticas	0.14	1.25	0.73	0.23	0.08	0.17	2.61
Táticas/Educação	0.71	0.84	0.18	0.25	0.24	0.20	2.42
Bate-papo	0.42	0.43	0.52	0.27	0.18	0.41	2.22
Reclamações	0.19	0.25	0.58	1.22	1.19	0.50	3.91
Discussões	0.58	1.43	0.36	0.59	0.23	0.33	3.51
Insultos	0.31	0.27	0.29	0.39	0.73	1.01	3.00
Provocações	0.34	0.25	0.23	0.41	1.45	0.48	3.16
Total Emoções	2.69	4.73	2.88	3.35	4.09	3.10	-

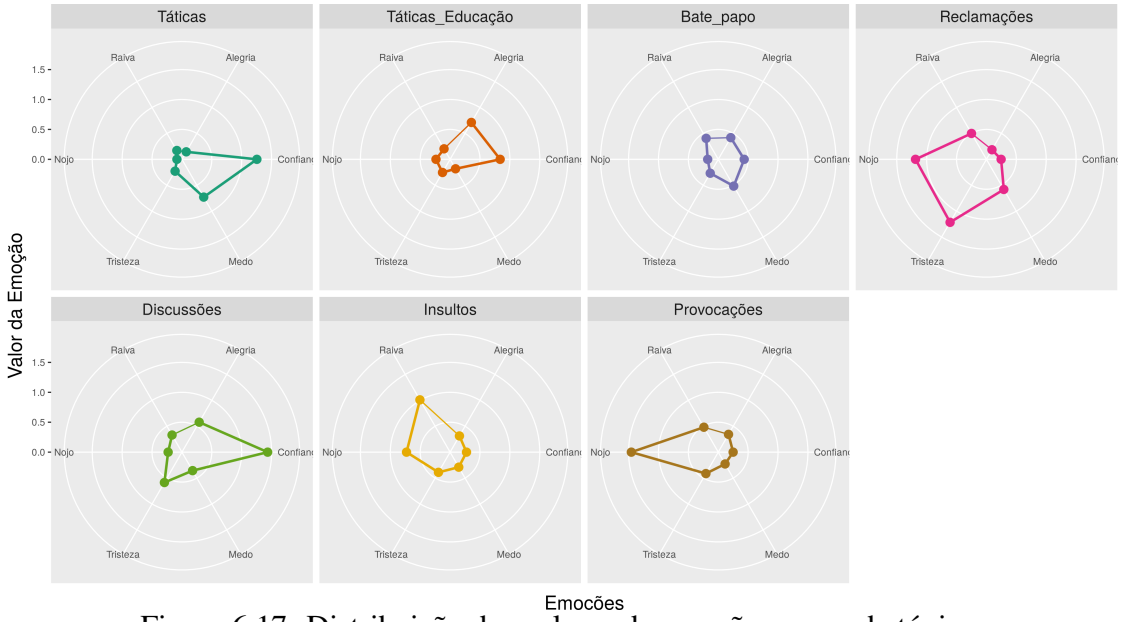


Figura 6.17: Distribuição dos valores de emoção para cada tópico.

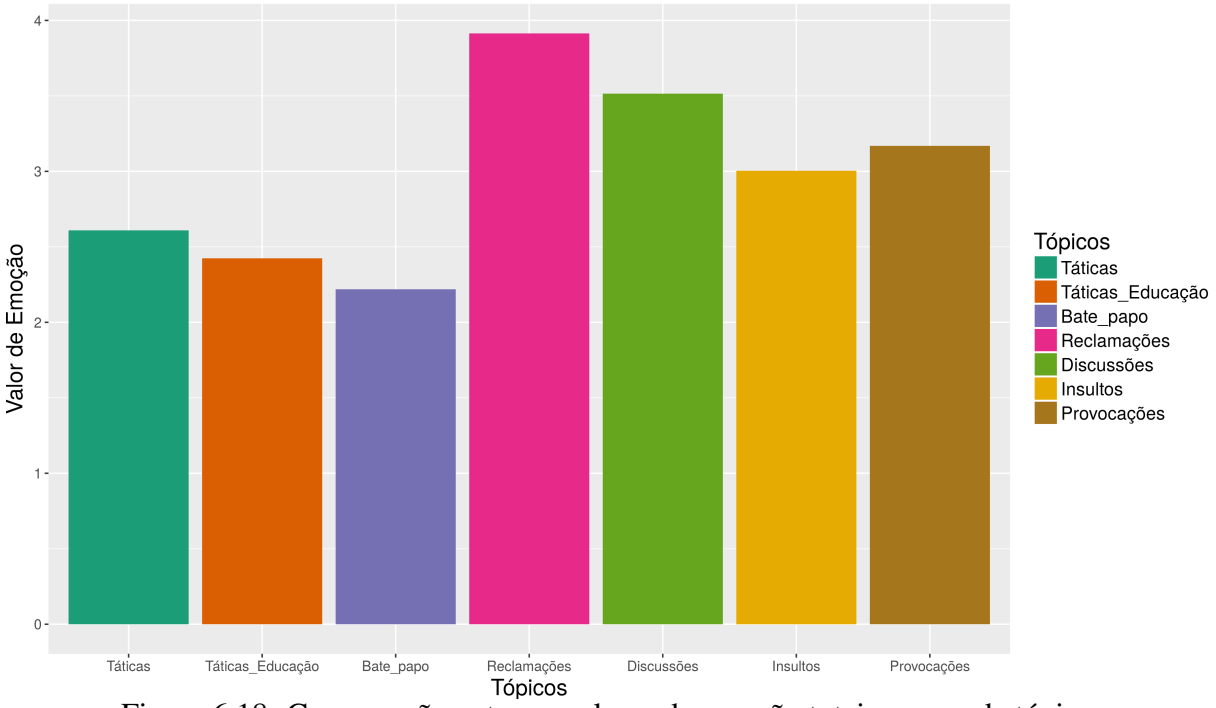


Figura 6.18: Comparação entre os valores de emoção totais para cada tópico.

dos tópicos negativos. Assim vemos que tópicos negativos tendem a ser mais carregados de emoção do que tópicos positivos, o que faz sentido, já que as situações que envolvem tópicos negativos normalmente são mais tensas e conflituosas.

Já considerando os totais de cada emoção, na Figura 6.19 vemos que as emoções com maiores valores nos tópicos são *medo* e *nojo*, que são associadas com tópicos negativos. Em contrapartida, o valor da emoção *alegria* é sensivelmente menor do que as outras. Contudo, ela ainda é bastante descriminatória para alguns tópicos. As emoções *confiança*, *raiva*, e *tristeza*

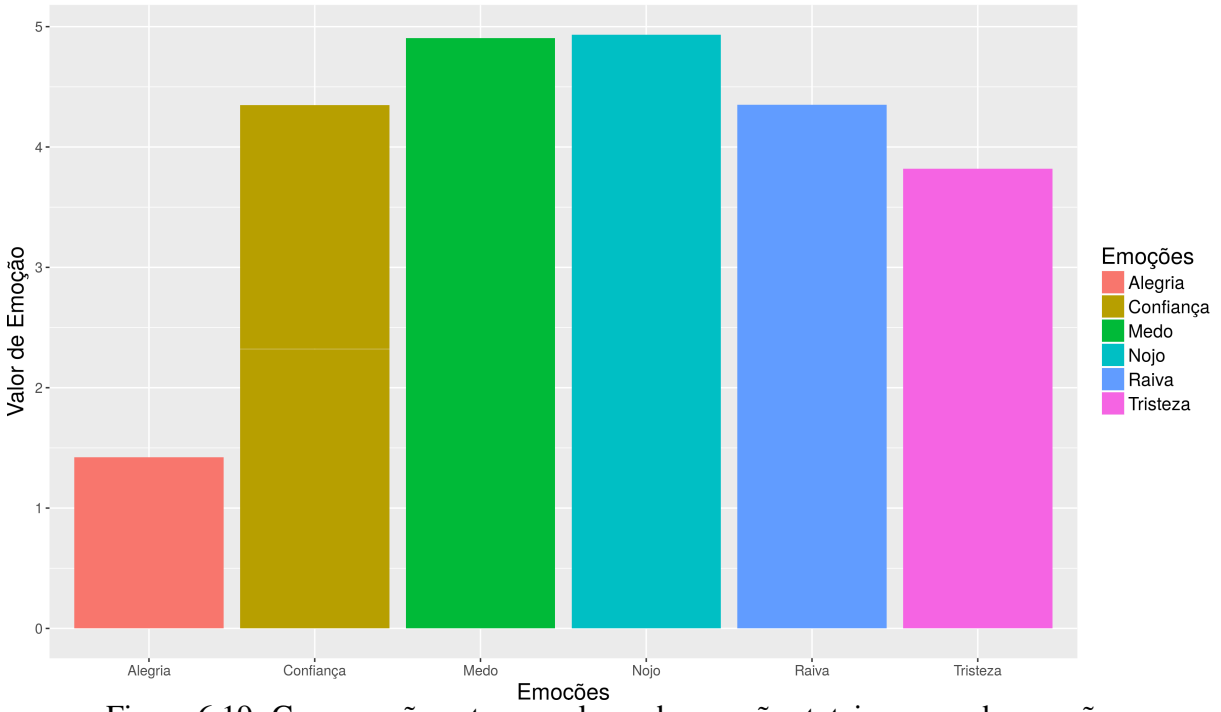


Figura 6.19: Comparação entre os valores de emoções totais para cada emoção.

também apresentam valores totais altos, com *confiança* tendo uma presença bastante forte em tópicos positivos, e as outras duas sendo mais presentes em tópicos negativos.

Finalmente, a Figura 6.20 apresenta vários gráficos em radar, similares aos da Figura 6.17, mas desta vez organizados por emoção ao longo dos diferentes tópicos. Olhando os gráficos correspondentes às emoções de *nojo* e *raiva*, confirmamos que essas são claramente associadas com comportamento tóxico, com os tópicos Reclamações, Provocações e Insultos apresentando altos níveis de ambas emoções. O tópico de Provocações se destaca, apresentando um valor extremamente alto de *nojo* (1,45), mesmo quando comparado com Reclamações (1,22) e Insultos (0,73), que ainda apresentam valores para esta emoção muito mais altos do que os vistos em tópicos positivos ($media = 0.16$).

A emoção *raiva* demonstra um padrão similar ao notado em *nojo*, apresentando valores altos em Provocações, Reclamações e Insultos. Contudo, desta vez, somente Insultos se destaca, apresentando valores de *raiva* (1,01) sensivelmente maiores do que os outros tópicos (Tabela 6.15). Apesar de Reclamações e Provocações apresentarem valores ainda altos de *raiva* (0.50 e 0.48, respectivamente), eles são mais próximos dos valores apresentados por tópicos positivos ($media = 0.26$) do que a emoção de nojo. Acreditamos que a prevalência das emoções *raiva* e *nojo* explicite um conflito entre jogadores, gerado por um certo menosprezo contra outros jogadores, originado provavelmente de jogadores tóxicos, já que estes são os principais usuários

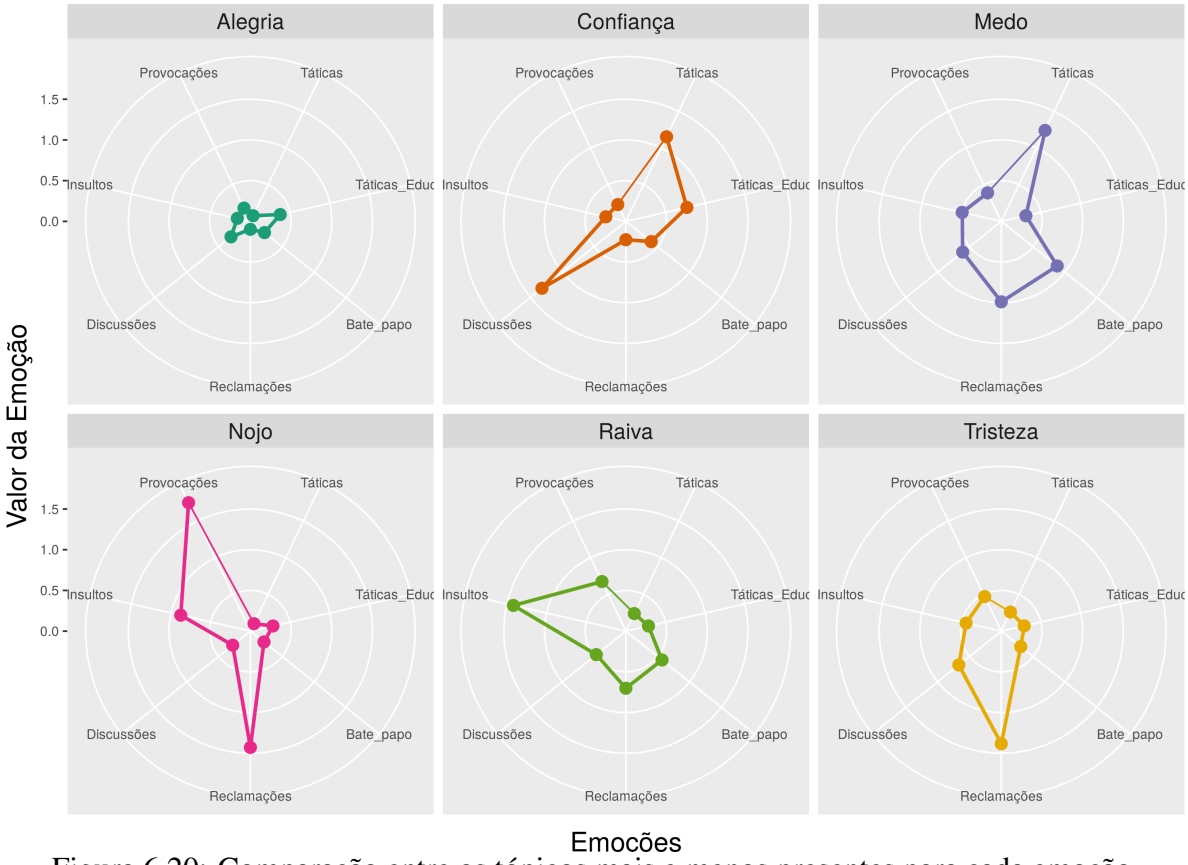


Figura 6.20: Comparação entre os tópicos mais e menos presentes para cada emoção.

de tópicos negativos. Esse conflito é representado pelos tópicos de Reclamações, Insultos e Provocações.

Também vemos na Figura 6.20 uma certa associação da emoção *tristeza* com tópicos negativos, com estes apresentando valores para esta emoção ($media.neg = 0.65$) acima da média dos tópicos positivos ($media.pos = 0.25$) para a mesma emoção, com destaque para Reclamações (1,21). Não encontramos nenhum possível comportamento para associar a este fenômeno, entretanto, esta emoção ainda funciona como um descriminador para tópicos negativos, em especial Reclamações.

Olhando a Figura 6.20 como um todo, vemos que o tópico Discussões tende a acompanhar a tendência apresentada por tópicos positivos nas emoções de *nojo* e *raiva*. Também é estranho que Discussões apresente valores sensivelmente maiores do que os outros tópicos negativos em emoções como *alegria* e *confiança* (($mdia = 0.28$ contra 0.58, e $mdia = 0.26$ contra 1.43, respectivamente), já que estas emoções são associadas a tópicos que representam bom humor, como também mostrado na Figura 6.20. Acreditamos que isso é devido à natureza diferente deste tópico. Enquanto tópicos como Reclamações, Insultos e Provocações tendem a possuir uma natureza de conflito direto com outros jogadores, explicitada pelas emoções de *raiva*, *nojo*, e pela falta da emoção *confiança*, o tópico Discussões busca conter comportamento

tóxico, possivelmente buscando resolver conflitos.

A emoção *confiança* parece estar associada com tópicos positivos. Os tópicos de Táticas (1.25) e Táticas/Educação (0.84) apresentam valores particularmente altos para esta emoção. Bate-papo (0.42) e Discussões (0.58) também apresentam valores sensivelmente altos nesta emoção, com Discussões sendo um tópico negativo "especial", como descrevemos anteriormente. Tópicos que indicamos como representando conflito apresentam valores sensivelmente baixos na emoção de *confiança* ($mdia = 0.26$), o que indica que conflitos entre jogadores podem nascer da falta de confiança entre os jogadores de um grupo.

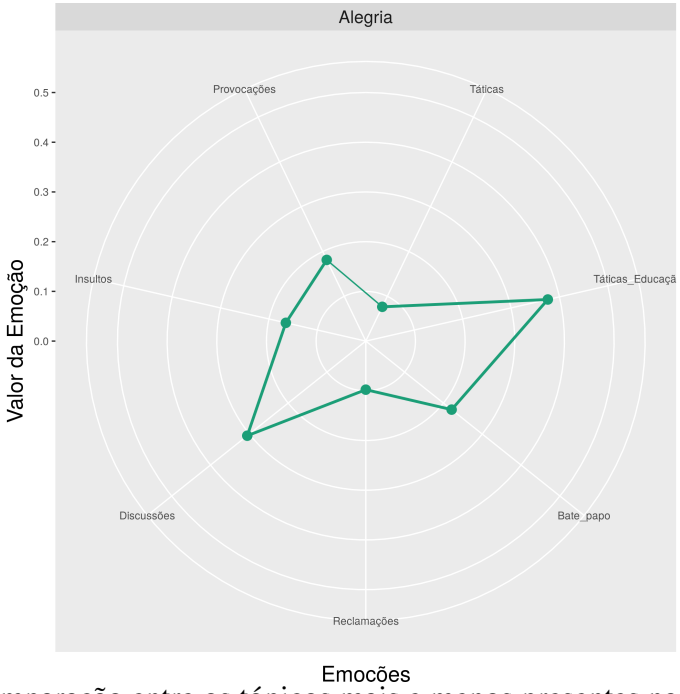


Figura 6.21: Comparação entre os tópicos mais e menos presentes na emoção alegria.

Devido à emoção *alegria* aparecer com uma intensidade muito menor que as outras emoções, os valores dela para diferentes tópicos estão plotados em forma de radar na Figura 6.21, em uma escala menor do que a vista na Figura 6.20. Vemos que a emoção *alegria* associa-se com os tópicos positivos relacionados a humor, como Táticas/educação (0.71) e Bate-papo (0.42), e Discussões (0.58). O tópico de Táticas presenta uma notória ausência de alegria, com tópicos negativos como Insultos (0.31), Provocações (0.34) e Reclamações (0.19) ainda apresentando valores de alegria maiores do que o deste tópico (0.14). A combinação entre *alegria* e *confiança* nos tópicos relacionados a humor mostra um ambiente emocional favorável, reforçando que a tese de que estes tópicos geram um ambiente de partida favorável para os jogadores. Contudo, devido a um peso muito maior das palavras de *confiança*, estas devem ser vista como a principal emoção nesta relação.

Medo também é uma emoção bastante presente nos tópicos positivos de Bate-papo (0.52)

e Táticas (0.73), sendo bastante presente neste último. O Tópico de Reclamações também apresenta níveis altos de medo (0.58). Esse medo em tópicos positivos pode ser justificado por uma preocupação com a partida por parte dos jogadores, mas níveis altos de medo no tópico de Reclamações mostram que essa emoção pode ser uma entrada para comportamentos negativos.

No geral, vemos que há dois principais patamares nos valores de emoções dos tópicos, com positivos apresentando menos emoções, em um patamar mais baixo e negativos apresentando mais emoções. Também vimos que os tópicos de conversação acabam associando-se a alguns tipos de emoções. Tópicos negativos prevalecem em emoções consideradas ruins, como *nojo* e *raiva*, que aparentam ser associadas com conflito. Uma exceção deste padrão é o tópico de Discussões, que em sua maior parte, segue os padrões emocionais apresentados por tópicos positivos. Contudo, ele é o tópico mais proeminente na emoção de *tristeza*, que também é associada com os demais tópicos negativos, por motivos desconhecidos.

Já tópicos positivos prevalecem em emoções como *alegria* e *confiança*, que aparentam promover um bom ambiente de jogo. O tópico de Discussões também aparenta ser bastante proeminente nestas emoções, o que mostra que os jogadores envolvidos na discussão podem estar tentando promover um melhor ambiente de jogo. Já tópicos associados com conflito, como Reclamações, Insultos e Provocações mostram baixos valores de *confiança*, o que indica que a falta desta entre membros de uma equipe pode ser um estopim para o comportamento tóxico. A emoção *medo* também se associa a tópicos positivos, e pode representar tanto uma apreensão pelo resultado da partida, como pode ser uma porta de entrada para comportamento tóxico, como mostrado pelos altos valores de medo presentes no tópico de Reclamações.

Finalmente, é importante destacar que os resultados desta análise devem ser considerados com cautela, uma vez que a validade do léxico ainda carece que uma avaliação mais profunda quanto à sua cobertura e precisão. Contudo, os resultados parecem confirmar e justificar os comportamentos positivos e negativos discutidos ao longo deste capítulo.

7 CONCLUSÃO E TRABALHOS FUTUROS

Neste trabalho, analisamos os padrões comportamentais em LoL com base em tópicos de conversação usados em bate-papos do jogo. Através da análise de 1,9 milhão de partidas tóxicas presentes no tribunal de LoL, demonstramos que estes padrões afetam o desempenho e a contaminação de grupo de jogadores. Nós investigamos em quais situações cada comportamento é predominante, como eles estão associados a cada tipo de jogador e como esses comportamentos estão relacionados ao desempenho e à contaminação. Também exploramos como o uso de tais tópicos por jogadores muda durante uma partida. Finamente, construímos automaticamente um dicionário de sentimentos voltado ao domínio de MOBAs, usando técnicas de aprendizado supervisionado, de forma que os tópicos possam ser caracterizados também pelos sentimentos prevalentes. À medida que exploramos dados típicos de jogos MOBA, todas estas contribuições podem ser aplicadas a outros jogos deste gênero, desde que os dados estejam disponíveis.

Os resultados de nossas pesquisa foram objeto de duas publicações em veículos qualificados, a saber:

- Neto and Becker (2018) - Qualis B1
- Neto, Yokoyama and Becker (2017) - Qualis B1

Confirmamos em nosso trabalho que existem padrões na comunicação textual entre jogadores, e que estes podem estar relacionados a comportamentos específicos dentro do jogo. Observamos que equipes que se concentram em táticas e socialização mostram um desempenho muito melhor, menores níveis de contaminação, bem como níveis mais altos de confiança. Já as equipes relacionadas a tópicos negativos podem ser associadas a comportamentos tóxicos e emoções consideradas ruins, como raiva e nojo, e tendem a ser vítimas de seus efeitos. Os tópicos negativos são frequentemente marcados por palavras e emoções que mostram o estresse, conflito e a tendência de culpabilização dos jogadores devido ao mau desempenho. Esses elementos podem estabelecer um círculo vicioso negativo que pode destruir a experiência do jogo para todos os tipos de jogadores.

Percebemos que o comportamento dos ofensores é frequentemente marcado por discussões negativas e tóxicas, que na maior parte das vezes afetam sua própria equipe. Apenas um terço das equipes adversas sofreram graves efeitos do comportamento tóxico dos ofensores. Além disso, diferentes tipos de discussões negativas podem afetar diferentes equipes de maneiras distintas. Observamos também que os jogos muito desequilibrados tendem a estimular comportamentos mais tóxicos. Os ofensores tendem a expressar mais raiva e frustração quando

seu time está desempenhando mal, particularmente através de queixas, argumentos e insultos. No entanto, identificamos algumas situações em que a equipe do ofensor acaba por replicar este comportamento tóxico de seu companheiro de equipe. Além disso, a provocação parece ser o único tipo de comportamento por parte do ofensor que afeta significativamente os inimigos.

Verificamos também que o comportamento mais comum aos jogadores de um grupo é manter a conversa centrada ao redor de um mesmo tópico, mas quando mudanças de tópicos ocorrem dentro de uma partida, elas tendem a ocorrer de maneiras distintas para jogadores ofensores e não ofensores. Mudanças para jogadores não ofensores tendem a ocorrer de maneira mais gradual, normalmente centradas nos tópicos de tática ou reclamação. Já para jogadores ofensores essa mudança tende a acontecer de maneira mais súbita e tendendo mais fortemente a tópicos negativos. Tópicos fortemente negativos como insultos ou provocações estão pouco ligados a padrões nos grupos de jogadores, e tendem a acontecer de maneira imprevisível, ou de serem motivados por agentes externos à conversa entre os jogadores.

Essas descobertas revelam diferentes formas de comportamento tóxico, que devido a sua definição nebulosa, é difícil de identificar. Nosso resultados podem ser explorados por sistemas automáticos para detectar comportamentos tóxicos para melhorar sua cobertura e precisão, bem como para recomendar o curso apropriado de ações para restringir tal comportamento (por exemplo, alocar partidas mais fáceis, recomendar um modo de jogo que induza menos estresse, ou sugerir uma pequena pausa para jogadores que apresentem sinais de toxicidade). Um sistema de detecção que atue ao vivo, em tempo de jogo, pode verificar métricas de desempenho, tópicos de conversação e a emoções em equipes para adotar medidas para impedir o comportamento tóxico em uma partida antes mesmo que ele aconteça.

Em casos extremos, o jogador pode ser punido por bloqueios no bate-papo de texto, ou mesmo ser banindo, temporariamente ou permanentemente. O problema de amenizar o comportamento tóxico também pode ser abordado na perspectiva de identificar e recompensar os jogadores que apresentem comportamento positivo. Por exemplo, os jogadores que usam tópicos mais positivos do que a média podem ser recompensados com prêmios no jogo e serem alocado com jogadores melhores.

Um possível trabalho futuro seria um estudo mais completo das situações de jogo identificadas neste trabalho, como partidas desequilibradas ou casos em que aliados e infratores são considerados como fonte de comportamento tóxico, bem como casos em que ofensores utilizam tópicos positivos, mas são denunciados por outros jogadores de toda maneira. Podemos utilizar o mesmo processo adotado neste trabalho para investigar o comportamento em outros jogos MOBA, bem como para comparar do comportamento de jogadores de diferentes culturas,

estudar fenômenos culturais como racismo e machismo nestes jogos, entre outros.

A análise temporal feita neste trabalho poderia ser estendida para pesquisar distinções nas palavras utilizadas nos tópicos ao longo dos anos, investigando variações nos conceitos de o que é um tópico positivo/negativo. Outro trabalho interessante seria a previsão do próximo tópico de conversação a ser utilizado em uma partida, a partir dos tópicos de conversas anteriores, junto com dados de desempenho, e informações sobre o comportamento em partidas anteriores dos jogadores envolvidos.

Uma avaliação mais aprofundada, com a presença de especialistas, e possíveis correções e melhorias no léxico de emoções específico ao MOBAs, levando à criação de um léxico mais confiável e que possa ser utilizado por uma gama maior de trabalhos também surge como um trabalho futuro interessante. Um léxico consolidado abriria caminho para a adoção de técnicas mais avançadas para impedir o surgimento da toxicidade em partidas, bem como a análises mais precisas sobre o comportamento dos jogadores.

REFERÊNCIAS

AGGARWAL, C. C.; ZHAI, C. **Mining text data**. [S.l.]: Springer Science & Business Media, 2012. 129–156 p.

AGGARWAL, C. C.; ZHAI, C. A survey of text clustering algorithms. In: **Mining text data**. [S.l.]: Springer, 2012. p. 77–128.

AGRAWAL, R.; SRIKANT, R. et al. Fast algorithms for mining association rules. In: **Proc. 20th int. conf. very large data bases, VLDB**. [S.l.: s.n.], 1994. v. 1215, p. 487–499.

BACCIANELLA, S.; ESULI, A.; SEBASTIANI, F. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: **LREC**. [S.l.: s.n.], 2010. v. 10, n. 2010, p. 2200–2204.

BARNETT, J.; COULSON, M.; FOREMAN, N. Examining Player Anger in World of Warcraft. **Human-Computer Interaction**, Springer London, p. 147–160, 2010.

BECKER, K.; MOREIRA, V. P.; SANTOS, A. G. dos. Multilingual emotion classification using supervised learning: Comparative experiments. **Information Processing & Management**, Elsevier, v. 53, n. 3, p. 684–704, 2017.

BIRD, S.; KLEIN, E.; LOPER, E. **Natural language processing with Python: analyzing text with the natural language toolkit**. [S.l.: s.n.]. 261–286 p.

BLACKBURN, J.; KWAK, H. STFU NOOB! In: **Proceedings of the 23rd international conference on World wide web - WWW '14**. New York, New York, USA: ACM Press, 2014. p. 877–888. ISBN 9781450327442.

BLEI, D. M.; NG, A. Y.; JORDAN, M. I. Latent dirichlet allocation. **Journal of machine Learning research**, v. 3, n. Jan, p. 993–1022, 2003.

BRADLEY, M. M.; LANG, P. J. **Affective norms for English words (ANEW): Instruction manual and affective ratings**. [S.l.], 1999.

BRAVO-MARQUEZ, F. et al. Determining word-emotion associations from tweets by multi-label classification. In: IEEE. **Web Intelligence (WI), 2016 IEEE/WIC/ACM International Conference on**. [S.l.], 2016. p. 536–539.

BUCKELS, E. E.; TRAPNELL, P. D.; PAULHUS, D. L. Trolls just want to have fun. **Personality and individual Differences**, Elsevier, v. 67, p. 97–102, 2014.

CARDOSO, P. M. D.; ROY, A. Sentiment lexicon creation using continuous latent space and neural networks. In: **WASSA@ NAACL-HLT**. [S.l.: s.n.], 2016. p. 37–42.

CHAWLA, N. V. et al. Smote: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321–357, 2002.

CHEN, V. H.-H.; DUH, H. B.-L.; NG, C. W. Players who play to make others cry. In: **International Conference on Advances in Computer Enterntainment Technology**. New York, New York, USA: ACM Press, 2009. p. 341–344. ISBN 9781605588643.

CHESNEY, T. et al. Griefing in virtual worlds: causes, casualties and coping strategies. **Information Systems Journal**, Wiley Online Library, v. 19, n. 6, p. 525–548, 2009.

DAYTON, C. M. Logistic regression analysis. **Stat**, p. 474–574, 1992.

FOO, C. Y.; KOIVISTO, E. M. I. Defining grief play in MMORPGs. In: **Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology - ACE '04**. New York, New York, USA: ACM Press, 2004. p. 245–250. ISBN 1581138822.

GOLF-PAPEZ, M.; VEER, E. Don't feed the trolling: rethinking how online trolling is being defined and combated. **Journal of Marketing Management**, Taylor & Francis, p. 1–19, 2017.

HAN, J.; PEI, J.; YIN, Y. Mining frequent patterns without candidate generation. In: ACM. **ACM sigmod record**. [S.l.], 2000. v. 29, n. 2, p. 1–12.

HARDAKER, C. Trolling in asynchronous computer-mediated communication: from user discussions to theoretical concepts. **Journal of Politeness Research**, v. 6, n. 2, p. 215–242, 2010.

JANKOWSKI, N.; GROCHOWSKI, M. Comparison of instances seletion algorithms i. algorithms survey. In: SPRINGER. **ICAISC**. [S.l.], 2004. p. 598–603.

JOACHIMS, T. Text categorization with support vector machines: Learning with many relevant features. **Machine learning: ECML-98**, Springer, p. 137–142, 1998.

KINGMA, D.; BA, J. Adam: A method for stochastic optimization. **arXiv preprint arXiv:1412.6980**, 2014.

KWAK, H.; BLACKBURN, J. Linguistic Analysis of Toxic Behavior in an Online Video Game. **EGG workshop**, Springer International Publishing, n. Cmc, p. 209–217, 2014.

LE, Q.; MIKOLOV, T. Distributed representations of sentences and documents. In: **Proceedings of the 31st International Conference on Machine Learning (ICML-14)**. [S.l.: s.n.], 2014. p. 1188–1196.

LIN, H.; CHUEN-TSAI, S. The 'White-eyed' Player Culture: Grief Play and Construction of Deviance in MMORPGs. In: **DiGRA 2005 Conference: Changing Views - Worlds in Play**. [S.l.: s.n.], 2005.

LIN, J. The science behind shaping player behavior in online games. In: **Game Developers Conference**. [S.l.: s.n.], 2013.

LUACES, O. et al. Binary relevance efficacy for multilabel classification. **Progress in Artificial Intelligence**, Springer, v. 1, n. 4, p. 303–313, 2012.

MARTENS, M. et al. Toxicity detection in multiplayer online games. In: **2015 International Workshop on Network and Systems Support for Games (NetGames)**. [S.l.]: IEEE, 2015. p. 1–6. ISBN 978-1-5090-0068-5.

MIKOLOV, T. et al. Efficient estimation of word representations in vector space. **arXiv preprint arXiv:1301.3781**, 2013.

MOHAMMAD, S. M. # emotional tweets. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. **Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation**. [S.l.], 2012. p. 246–255.

MOHAMMAD, S. M.; TURNEY, P. D. **Nrc emotion lexicon**. [S.l.], 2013.

MUNEZERO, M. D. et al. Are they different? affect, feeling, emotion, sentiment, and opinion detection in text. **IEEE transactions on affective computing**, IEEE, v. 5, n. 2, p. 101–111, 2014.

NETO, J. A.; YOKOYAMA, K. M.; BECKER, K. Studying toxic behavior influence and player chat in an online video game. In: ACM. **Proceedings of the International Conference on Web Intelligence**. [S.l.], 2017. p. 26–33.

NETO, J. A. de M.; BECKER, K. Relating conversational topics and toxic behavior effects in a moba game. **Entertainment Computing**, Elsevier, 2018.

PEDREGOSA, F. et al. Scikit-learn: Machine learning in python. **Journal of Machine Learning Research**, v. 12, n. Oct, p. 2825–2830, 2011.

PENNINGTON, J.; SOCHER, R.; MANNING, C. Glove: Global vectors for word representation. In: **Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)**. [S.l.: s.n.], 2014. p. 1532–1543.

PLUTCHIK, R. Emotions: A general psychoevolutionary theory. **Approaches to emotion**, v. 1984, p. 197–219, 1984.

READ, J. et al. Classifier chains for multi-label classification. **Machine learning**, Springer, v. 85, n. 3, p. 333, 2011.

ŘEHŮŘEK, R.; SOJKA, P. Software Framework for Topic Modelling with Large Corpora. In: **Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks**. Valletta, Malta: ELRA, 2010. p. 45–50.

ROSS, T. L.; WEAVER, A. J. Shall we play a game? **Journal of Media Psychology**, Hogrefe Publishing, 2012.

RUSSELL, J. A.; MEHRABIAN, A. Evidence for a three-factor theory of emotions. **Journal of research in Personality**, Elsevier, v. 11, n. 3, p. 273–294, 1977.

SCHOLKOPF, B. et al. Comparing support vector machines with gaussian kernels to radial basis function classifiers. **IEEE transactions on Signal Processing**, IEEE, v. 45, n. 11, p. 2758–2765, 1997.

SEVERYN, A.; MOSCHITTI, A. Twitter sentiment analysis with deep convolutional neural networks. In: ACM. **Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval**. [S.l.], 2015. p. 959–962.

SHORES, K. B. et al. The identification of deviance and its impact on retention in a multiplayer game. **Proceedings of the 17th ACM conference on Computer Supported Cooperative Work & Social Computing - CSCW '14**, ACM Press, New York, New York, USA, p. 1356–1365, 2014.

SHOUSE, E. Feeling, emotion, affect. 2007.

SONG, K. et al. Build emotion lexicon from microblogs by combining effects of seed words and emoticons in a heterogeneous graph. In: ACM. **Proceedings of the 26th ACM conference on hypertext & social media**. [S.l.], 2015. p. 283–292.

SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. **Journal of machine learning research**, v. 15, n. 1, p. 1929–1958, 2014.

SULER, J. The online disinhibition effect. **CyberPsychology & Behavior**, Mary Ann Liebert, Inc., v. 7, n. 3, p. 321–326, jun 2004. ISSN 1094-9313.

THACKER, S.; GRIFFITHS, M. D. An exploratory study of trolling in online video gaming. **International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)**, IGI Global, v. 2, n. 4, p. 17–33, 2012.

TIELEMAN, T.; HINTON, G. Lecture 6.5-rmsprop, coursera: Neural networks for machine learning. **University of Toronto, Technical Report**, 2012.

ZAKI, M. J. et al. New algorithms for fast discovery of association rules. In: **KDD**. [S.l.: s.n.], 1997. v. 97, p. 283–286.

ZHANG, C.; ZHANG, S. **Association rule mining: models and algorithms**. [S.l.]: Springer-Verlag, 2002.

Appendices

ApêndiceA INSTRUÇÕES DE INTERPRETAÇÃO DE AGRUPAMENTOS

Abaixo mostramos um pedaço do instrumento de validação de interpretação que foi entregue aos validadores. O instrumento está disponível como um pesquisa online[1].

Instruções de Validação - Agrupamento 1

Para este agrupamento, vamos lhe apresentar as palavras mais importantes:

- Uma nuvem de palavras, onde o tamanho da palavra é proporcional à sua importância para aquele grupo;
- As 10 palavras mais relevantes contidas na imagem.

A partir dessas informações, descreva em poucas palavras qual tipo de conversação em uma partida de *League of Legends* melhor descreve as palavras listadas. Essas palavras buscam descrever o diálogo mais predominantes durante uma partida. Então , busque descrever o tipo de diálogo (ex: "reclamações", "Chamadas de objetivos"), e não o tipo de jogador que normalmente usaria esse tipo de diálogo (ex: jogador tóxico).

Note que:

- Dois ou mais agrupamentos podem ter o mesmo significado, ou significados bastante parecidos. Você pode usar a mesma descrição;
- Você pode desconsiderar agrupamentos onde o idioma predominante não é o inglês. Neste caso, marque-o como 'outras línguas' e desconsidere-o.

Palavras de Validação - Agrupamento 1

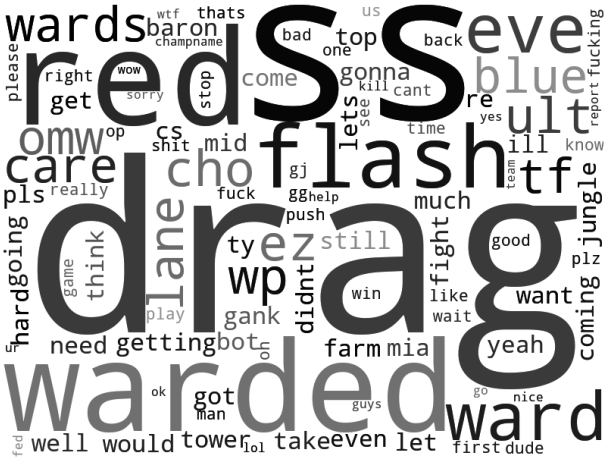


Figura A.1: Nuvem das 100 palavras mais relevantes do agrupamento 1.

[1] https://goo.gl/forms/MzCLYGiboOMCplYx1

As 10 palavras mais relevantes para esse grupo são:
drag, ss, warded, red, flash, ward, eve, wards, tf, blue

Perguntas de Validação - Agrupamento 1

- **Pergunta 1**: Qual tipo de conversação melhor descreve as palavras listadas e a imagem?
- **Pergunta 2**: Você tem mais algum comentário extra sobre as palavras nesse grupo?

A.1 Interpretação dos tópicos

Tabela A.1: Resumo da interpretação dos Tópicos

Agrupamento	Descrição do Avaliador	Descrição Proposta	Nível de Concordância	Razões para discordância
0	Outras Linguas (5)	Outras Linguas	Total	-
1	Chamadas de Objetivo (3) Chamadas de Objetivo (Controle de Mapa) (2)	Táticas	Total	-
2	Chamadas de Objetivo (3) Chamadas de Objetivo (Momento de jogos específico) (2)	Táticas	Total	-
3	Avisos a Equipe (3) Conversa Educada (1) Conversa Genérica (1)	Táticas/Educação	Parcial (4)	Conversa Neutra (1)
4	Conversa sobre elementos do jogo (3) Conversas com o inimigo (1) Conversas Neutras (1)	Bate-papo	Parcial (4)	Conversa Neutra (1)
5	Reclamações (4) Conflito na equipe (1)	Reclamações	Total	-
6	Reclamações (2) Culpabilização (2) Conflito na equipe (1)	Reclamações	Total	-
7	Conversa Genérica (4) Restrição agressiva de comportamento tóxico (1)	Discussões	Total[2]	-
8	Insultos (5)	Insultos	Total	-
9	Insultos Pesados (3) Insultos (1) Discurso de ódio (1)	Provocações	Parcial (4)	Insultos (1)

[2]Depois de mostrar dados de desempenho e contaminação.

ApêndiceB ESTRUTURA DOS ARQUIVOS DO DATASET

Neste apêndice nós apresentamos a estrutura dos arquivos JSON que representam partidas nos nossos dados. Cada entrada descrita abaixo está no seguinte formato: Nome do campo - Tipo do campo - (Detalhes) - [Valores possíveis].

- **Tipo de Partida** - String
- **Registro de bate-papo** - Lista (Número inderterminado de items)
  - ***Timestamp*** - Horário
  - **Bate-papo alvo** - String - [Time1, Time2, Global]
  - **Associação com o ofensor** - String - [Aliado, Inimigo, Ofensor]
  - **Nome do campeão** - String
  - **Mensagem** - String
- **Jogadores** - Lista (10 itens)
  - **Associação com o ofensor** - String - [Aliado, Inimigo, Ofensor]
  - **Nível** - Inteiro
  - **Pontuações** - Objeto
    - **Abates** - Inteiro
    - **Mortes** - Inteiro
    - **Assistências** - Inteiro
  - ***Minions* Mortos** - Inteiro
  - **Dano total causado** - Inteiro
  - **Dano total recebido** - Inteiro
  - **Ouro recebido** - Inteiro
  - **Itens** - Lista - (0 a 6 itens)
    - **ID** - Inteiro
    - **Nome** - String
    - **Descrição** - String
    - **Ícone** - String
    - **Preço** - Inteiro
  - **Resultado** - String - [Vitória, Derrota]
  - **Tempo jogado** - Inteiro
  - **Nome do personagem** - String
  - **Feitiço de invocador 1** - String

  - **Feitiço de invocador 2** - String

- **Motivação de denúncia mais utilizada** - String
- **Numero de denúncias aliadas** - Inteiro
- **Número de denúncias inimigas** - Inteiro
- **Versão do jogo** - String

ApêndiceC RESULTADOS DOS EXPERIMENTOS COM PARÂMETROS DO APRIORI

Neste apêndice mostramos os experimentos realizados para determinar o valor de confiança escolhido no experimento. Os experimentos listados foram realizados utilizando o pacote arules[1] para a linguagem R. Cada linha de cada uma das tabelas mostra os valores mínimos de confiança e *lift* necessários, e o número de regras que foram consideradas relevantes de acordo com estes valores para aliados, inimigos e ofensores. O valor de suporte mínimo é de 0.003 em todos os casos.

[1] https://cran.r-project.org/web/packages/arules/index.html

C.1 Resultados para experimentos em transações com repetições

Tabela C.1: Resultados para execuções do em transações com repetições

Métricas		**Número de Regras**		
Confiança	***Lift***	**Aliados**	**Inimigos**	**Ofensores**
0.80	1.30	0	0	0
0.70	1.30	0	0	0
0.60	1.30	0	0	0
0.50	1.30	0	0	0
0.40	1.30	0	0	0
0.30	1.30	1	1	0
0.20	1.30	3	3	3
0.10	1.30	8	9	13
0.80	1.20	0	0	0
0.70	1.20	0	0	0
0.60	1.20	0	0	0
0.50	1.20	0	0	0
0.40	1.20	0	0	0
0.30	1.20	1	1	0
0.20	1.20	3	5	3
0.10	1.20	10	11	14
0.80	1.10	0	0	0
0.70	1.10	0	0	0
0.60	1.10	0	0	0
0.50	1.10	0	0	0
0.40	1.10	0	0	0
0.30	1.10	1	1	0
0.20	1.10	3	5	3
0.10	1.10	13	13	15

C.2 Resultados para execuções do apriori em transações sem repetições

Tabela C.2: Resultados para experimentos em transações sem repetições

Métricas		**Número de Regras**		
Confiança	***Lift***	**Aliados**	**Inimigos**	**Ofensores**
0.80	1.30	0	0	0
0.70	1.30	0	0	0
0.60	1.30	0	0	0
0.50	1.30	0	0	0
0.40	1.30	0	1	0
0.30	1.30	4	4	2
0.20	1.30	5	6	6
0.10	1.30	8	8	10
0.80	1.20	0	0	0
0.70	1.20	0	0	0
0.60	1.20	0	0	0
0.50	1.20	0	0	0
0.40	1.20	0	1	0
0.30	1.20	4	4	2
0.20	1.20	8	9	6
0.10	1.20	12	12	13
0.80	1.10	0	0	0
0.70	1.10	0	0	0
0.60	1.10	0	0	0
0.50	1.10	0	0	0
0.40	1.10	0	1	0
0.30	1.10	4	4	2
0.20	1.10	11	11	7
0.10	1.10	16	15	17

ApêndiceD RESULTADOS DOS EXPERIMENTOS COM PARÂMETROS DO APRIORI

Neste apêndice nós mostramos os resultados dos experimentos realizados para determinar o classificador utilizado para a construção do léxico de sentimentos específico ao dominio de MOBAs. Os experimentos listados foram feitos utilizando o pacote scikit-learn (Regressão Logística e SVM) e keras (Redes Neurais), ambos para python. Cada linha das tabelas abaixo mostram os resultados do classificador listado para cada um dos sentimentos presentes no léxico NRC.

D.1 Determinando o Modelo de classificação a ser utilizado

Experimentos preliminares com os classificadores naïve-bayes (gaussiano), k-vizinhos, florestas aleatórias, regressão logística, SVM e perceptron de várias camadas (MLP). Para escolher os algoritmos, utilizamos a métrica Micro Medida-F. Os testes preliminares foram realizados com um modelo de *embbedings* gerado pelo algoritmo word2vec (variação skipgram) de palavras de tamanho 300, e com todos os parâmetros padrões da API scikit-learn. Para algoritmos sem suporte a multi-rótulos, utilizamos o método de relevância binária. Dos classificadores, o MLP e o SVM deram os melhores resultados, e foram testados mais a fundo com diferentes parâmetros.

A partir disso, fizemos testes mais avançados com SVM e MLP, que demonstraram os melhores resultados. Variamos o tamanho das *embbedings* do word2vec (entre 100, 200 e 300), tanto como testamos cada um destes nas variantes *skipgram* e *continuous bag of words* do algoritmo. Além disso, testamos os algoritmos multi-rótulos de relevância binária e cadeia

Tabela D.1: Micro F-Measure dos testes preliminares

Modelo	**Micro Medida-F**
Aleatório	0,149
Naïve-Bayes	0,393
K-vizinhos	0,368
Florestas Aleatórias	0,059
Regressão Logística L2	0,385
SVM	0,032
SVM balanceado	0,414
MLP	0,407

de classificadores, bem como os algoritmos de super e sub amostragem SMOTE e ENN, e a variante balanceada do SVM, que dá pesos diferentes para as classes de acordo com suas frequências.

D.1.0.1 Para 100 dimensões

Tabela D.2: Resultados dos modelos de classificação para vetores de 100 dimensões

Modelo	**Micro Medida-F CBOW**	**Micro Medida-F Skipgram**
Relevância Binária		
MLP + ENN	0,412556	0,428571
MLP + SMOTE	0,381316	0,405545
SVM + ENN	0,328976	0,225334
SVM + SMOTE	0,333160	0,439865
SVM balanceado	0,370195	0,380225
SVM balanceado + ENN	0,392341	0,366709
SVM balanceado +SMOTE	0,329100	0,417631
Cadeia de Classificadores		
MLP + ENN	0,399388	0,417044
MLP + SMOTE	0,361446	0,388045
SVM + ENN	0,326099	0,033613
SVM + SMOTE	0,329486	0,403543
SVM balanceado	0,358696	0,344394
SVM balanceado + ENN	0,393596	0,335091
SVM balanceado + SMOTE	0,328638	0,403543

D.1.0.2 Para 200 dimensões

Tabela D.3: Resultados dos modelos de classificação para vetores de 200 dimensões

Modelo	**Micro Medida-F CBOW**	**Micro Medida-F Skipgram**
Relevância Binária		
MLP + ENN	0,439042	0,452572
MLP + SMOTE	0,409621	0,449415
SVM + ENN	0,353185	0,200664
SVM + SMOTE	0,407129	0,447478
SVM balanceado	0,415850	0,401840
SVM balanceado + ENN	0,421654	0,389301
SVM balanceado +SMOTE	0,405556	0,435773
Cadeia de Classificadores		
MLP + ENN	0,431613	0,461095
MLP + SMOTE	0,410401	0,427225
SVM + ENN	0,351366	0,042586
SVM + SMOTE	0,383616	0,419223
SVM balanceado	0,414644	0,378662
SVM balanceado + ENN	0,419009	0,369170
SVM balanceado + SMOTE	0,382368	0,419223

D.1.0.3 Para 300 dimensões

Tabela D.4: Resultados dos modelos de classificação para vetores de 300 dimensões

Modelo	**Micro Medida-F CBOW**	**Micro Medida-F Skipgram**
Relevância Binária		
MLP + ENN	0,428921	0,475262
MLP + SMOTE	0,433632	0,436072
SVM + ENN	0,359326	0,117147
SVM + SMOTE	0,412713	0,447337
SVM balanceado	0,438396	0,402023
SVM balanceado + ENN	0,433765	0,392212
SVM balanceado +SMOTE	0,410898	0,437067
Cadeia de Classificadores		
MLP + ENN	0,432916	0,461538
MLP + SMOTE	0,424009	0,436836
SVM + ENN	0,355841	0,026134
SVM + SMOTE	0,406599	0,428976
SVM balanceado	0,430906	0,380383
SVM balanceado + ENN	0,423837	0,359184
SVM balanceado + SMOTE	0,407913	0,428916

ApêndiceE PALAVRAS UTILIZADAS PARA A VALIDAÇÃO DO LÉXICO DE EMOÇÕES PARA MOBAS

Neste apêndice nós mostramos os dados utilizados para a validação do léxico de emoções. Aqui mostramos as nuvens de palavras contendo palavras exclusivas ao léxico de MOBA, e listamos as palavras consideradas como falsos positivos.

E.1 Alegria

Figura E.1: Nuvem de palavras para as palavras apresentando a emoção de alegria.

Total de Palavras: 12

***Recall* do modelo**: 0,352

Precisão do modelo: 0,652

Falsos Positivos: hi (1/12=8,3% das palavras).

E.2 Antecipação

Figura E.2: Nuvem de palavras para as palavras apresentando a emoção de antecipação.

Total de Palavras: 19

***Recall* do modelo**: 0,186

Precisão do modelo: 0,361

Falsos Positivos: yay, jesus, ga, ty, gj, gl (7/19=36,8% das palavras).

E.3 Confiança

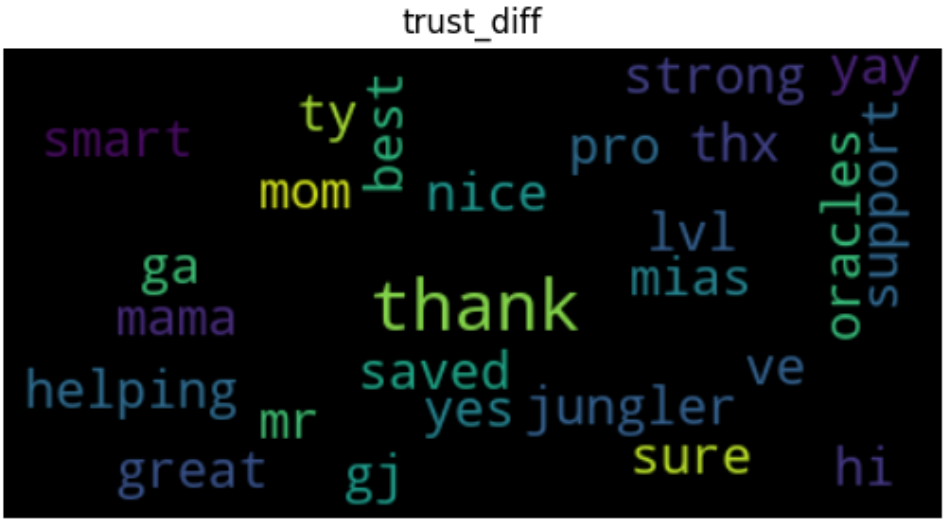


Figura E.3: Nuvem de palavras para as palavras apresentando a emoção de confiança.

Total de Palavras: 28

***Recall* do modelo**: 0,268

Precisão do modelo: 0,465

Falsos Positivos: yay, lvl, yes, mr, ve, hi (6/28=21,4% das palavras).

E.4 Medo

Figura E.4: Nuvem de palavras para as palavras apresentando a emoção de medo.

Total de Palavras: 42

***Recall* do modelo**: 0,522

Precisão do modelo: 0,460

Falsos Positivos: red, strong, taking, drag, worst, oracles, baron, ad, dat, blue, owned (11/42=26,1% das palavras).

E.5 Nojo

Figura E.5: Nuvem de palavras para as palavras apresentando a emoção de nojo.

Total de Palavras: 37

***Recall* do modelo**: 0,390

Precisão do modelo: 0,554

Falsos Positivos: sad, look, piece, string, hijo, ctm, dmg (7/37=18,9% das palavras).

E.6 Raiva

Figura E.6: Nuvem de palavras para as palavras apresentando a emoção de raiva.

Total de Palavras: 27

***Recall* do modelo**: 0,550

Precisão do modelo: 0,430

Falsos Positivos: dmg, taking, vs, pra (4/27=14,8% das palavras).

E.7 Surpresa

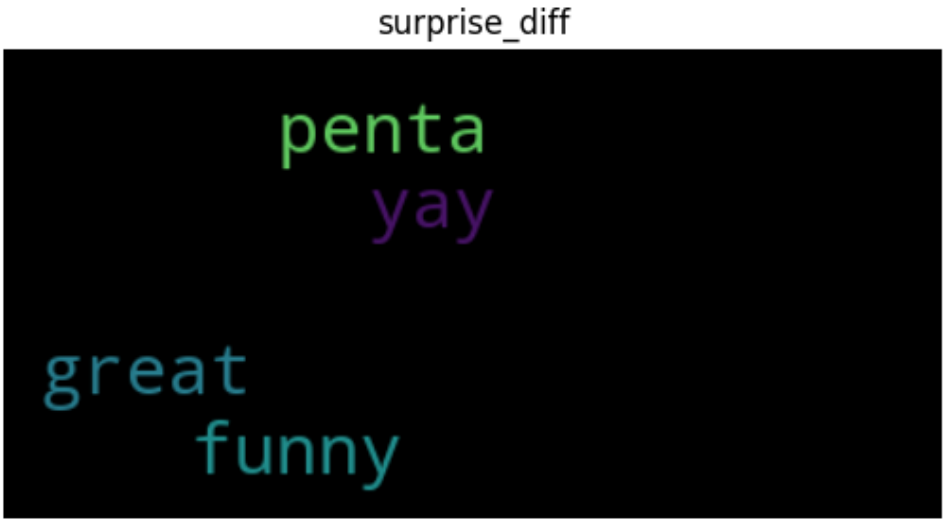


Figura E.7: Nuvem de palavras para as palavras apresentando a emoção de surpresa.

Total de Palavras: 04

***Recall* do modelo**: 0,138

Precisão do modelo: 0,429

Falsos Positivos: None.

E.8 Tristeza

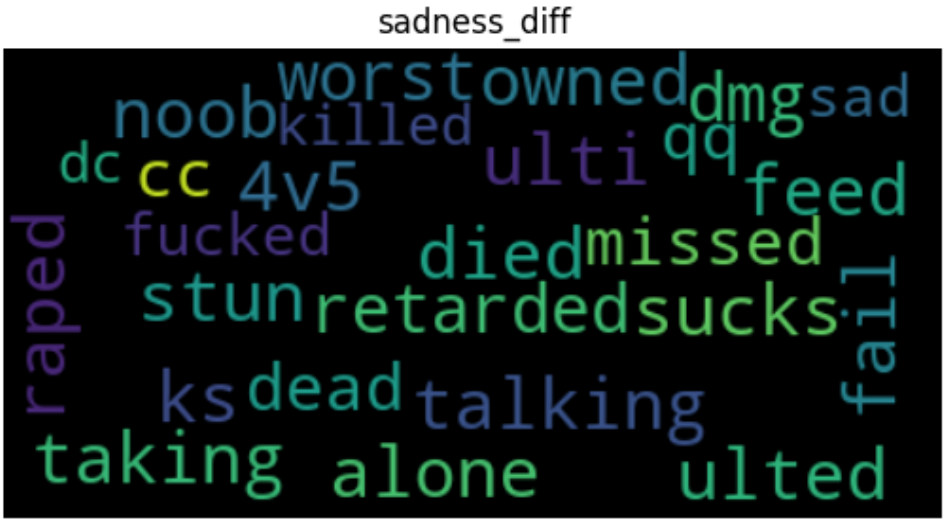


Figura E.8: Nuvem de palavras para as palavras apresentando a emoção de tristeza.

Total de Palavras: 28

***Recall* do modelo**: 0,427

Precisão do modelo: 0,353

Falsos Positivos: retarded, talking (2/28=7,1% das palavras).