



Universidade Federal do Piauí  
Centro de Ciências da Natureza  
Programa de Pós-Graduação em Ciência da Computação

# **Um Estudo sobre a Formação de Equipes em Jogos Virtuais**

**Alexandre Tolstenko Nogueira**

**Teresina-PI, Julho de 2015**

Alexandre Tolstenko Nogueira

## **Um Estudo sobre a Formação de Equipes em Jogos Virtuais**

**Dissertação de Mestrado** apresentada ao Programa de Pós-Graduação em Ciência da Computação da UFPI (área de concentração: Sistemas de Computação), como parte dos requisitos necessários para a obtenção do Título de Mestre em Ciência da Computação.

Universidade Federal do Piauí – UFPI

Centro de Ciências da Natureza

Programa de Pós-Graduação em Ciência da Computação

Orientador: Pedro de Alcântara dos Santos Neto

Teresina-PI

Julho de 2015

FICHA CATALOGRÁFICA  
Serviço de Processamento Técnico da Universidade Federal do Piauí  
Biblioteca Setorial do CCN

N778e Nogueira, Alexandre Tolstenko.

Um estudo sobre a formação de equipes em jogos virtuais / Alexandre Tolstenko Nogueira. – Teresina, 2015.  
93f.: il.

Dissertação (Mestrado) – Pós-Graduação em Ciência da Computação, Universidade Federal do Piauí, 2015.

Orientador: Prof. Dr. Pedro de Alcântara dos Santos Neto.

1. Inteligencia Artificial. 2. Redes Neurais. 3. Formação de Equipes - Games. I. Título.

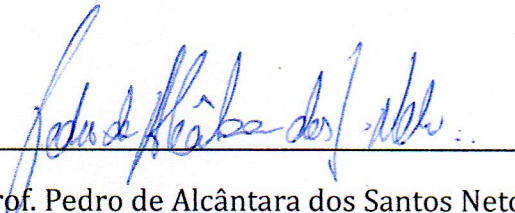
CDD 006.32

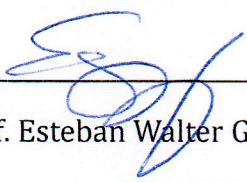
# Um Estudo sobre a Formação de Equipes em Jogos Virtuais

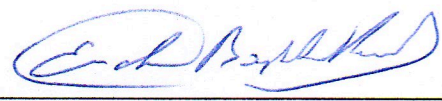
**ALEXANDRE TOLSTENKO NOGUEIRA**

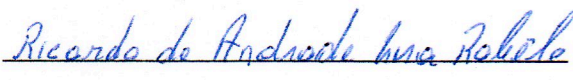
Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Centro de Ciências da Natureza da Universidade Federal do Piauí, como parte integrante dos requisitos necessários para obtenção do grau de Mestre em Ciência da Computação.

Aprovado por:

  
\_\_\_\_\_  
Prof. Pedro de Alcântara dos Santos Neto  
(Presidente da Banca)

  
\_\_\_\_\_  
Prof. Esteban Walter Gonzalez Clua  
(Examinador Externo)

  
\_\_\_\_\_  
Prof. Erick Baptista Passos  
(Examinador Interno)

  
\_\_\_\_\_  
Prof. Ricardo de Andrade Lira Rabêlo  
(Examinador Interno)

Teresina, 23 de julho de 2015

Aos meus amados **pais** Alexandre e Lídyá, pessoas imprescindíveis em minha vida pela educação, formação, trabalho e determinação, responsáveis pelo meu crescimento pessoal e intelectual. Com todo meu amor, minha eterna gratidão!

Às minhas **irmãs** Tatiana, Ludmila pelo afeto e amizade, e em especial à Luciana, que sempre me cuidou e incentivou nos momentos de fraqueza. Obrigado por estarem presentes ao longo de minha vida, responsáveis pelo meu crescimento emocional. Com todo meu carinho, adoro vocês!

À minha **esposa** Marina, pelo carinho, paciência, calma e disposição em ajudar sempre no que esteve em seu alcance. Muito obrigado!

À minha **família** por toda compreensão, carinho e amor a mim dispensados.



# Agradecimentos

***“God used beautiful mathematics in creating the world.”*** (Paul Dirac)

Agradeço a Deus...

Pelo dom da vida e pela cura de uma terrível doença que me tolheu a infância;

Pelas pedras no caminho que me foram impostas grandes o bastantes para me desafiarem e, com a força adquirida, pude ir ao longe;

Por colocar pessoas fantásticas ao meu lado que me estenderam a mão e as muitas que me ajudaram na realização deste trabalho. Desejo profundamente agradecê-las.

***“If I have seen further than others, it is by standing upon the shoulders of giants.”*** (Isaac Newton)

Ao meu orientador, Pedro de Alcântara dos Santos Neto pelo exemplo irrefutável de verdadeiro mestre, minha admiração; pela sua compreensão, paciência, posicionamento e pronta disponibilidade, minha gratidão.

Ao Professor Ricardo Lira por coorientar e tirar minhas dúvidas quando naveguei pela sua área de especialidade, e pela amizade dispensada ao longo da trajetória do mestrado, minha admiração e meus agradecimentos.

***“Education is the most powerful weapon which you can use to change the world.”*** (Nelson Mandela)

Aos professores, promotores de grande expansão de consciência, pelo aprendizado enriquecedor e pela disponibilidade sempre que solicitados. Em especial aos que me desafiavam, me estimulando a ir além dos meus próprios limites. Fizeram-me ver o quanto posso ir longe e que devo sempre confiar em minha capacidade. Meus sinceros agradecimentos.

***“Walking with a friend in the dark is better than walking alone in the light.”*** (Helen Keller)

Aos colegas, uns do Sudeste, outros mais próximos como no Nordeste, e uns mais distantes em outros países, cada qual com cultura e bagagem de vidas diferentes. Todos contribuíram com uma parcela para o meu aprendizado, pela troca enriquecedora pessoal e profissional, com todas as conversas e tirando minhas dúvidas e medos. Meus agradecimentos e votos de sucesso a todos, onde quer que estejam.





*“THERE IS AS YET INSUFFICIENT DATA FOR A MEANINGFUL ANSWER”*  
(Isaac Asimov)



# Resumo

Os jogos eletrônicos ganharam notoriedade nas últimas décadas, assumindo uma proporção sem precedentes. Atualmente, há demanda crescente por jogos eletrônicos em equipes e, em consequência, busca-se a formação de times que ampliem a disputa e tornem as partidas mais interessantes. No entanto, formar equipes equilibradas, de forma a ampliar uma disputa não é tarefa fácil, especialmente em um cenário com milhões de usuários. Neste trabalho foi realizado um estudo sobre a formação de equipes em jogos com vários jogadores. O jogo *Dota 2* do gênero *MOBA* foi utilizado como referência, pois possui uma ampla variedade de heróis, que juntamente com a forma de jogar de cada usuário, torna cada partida singular. Para investigar a formação e balanceamento das equipes, foram analisadas cerca de 11 milhões de partidas para se inferir características dos jogadores. Foram encontrados 12 agrupamentos de jogadores, utilizados para se analisar a composição das equipes de partidas. A partir de tal informação foi possível descobrir que quanto maior o grau de desbalanceamento das equipes, segundo os agrupamentos de usuário encontrados, menor era a duração da partida e maior era a diferença de pontuação da equipe vencedora em relação à perdedora. Isso permitiu que fosse sugerida uma diretriz para formação de equipes baseado na metodologia utilizada neste trabalho e que pode ser usada em outros jogos que envolvam times em uma disputa.

**Palavras-chaves:** *Game Metrics. Game Analytics. Match Making.* Aprendizado não supervisionado e Agrupamento. Jogos Multijogador Massivo em Rede. Mineração de Dados. Aprendizado Não-Supervisionado. Redes Neurais.



# Abstract

Electronic games have gathered a lot of notoriety and assuming an unprecedented position. Nowadays there is growing demand for electronic games in teams and as a result, we need to match teams that broaden the dispute and make the most interesting game. But making balanced teams as to amplify a dispute is not an easy task. Even more so in a scene with millions of users. In this paper a study about matchmaking was done. The game Dota 2 was used as reference since it has an ample variety of heroes which when combined with how each user plays makes each match singular. In order to investigate the formation and balancing of teams about 11 million matches were analysed so as to infer characteristics of players. It were found 12 groupings of players, used to analyse the composition of teams in matches. From such information it was possible to find out that the greater the level of unbalance of teams, according to the groupings found, the shorter was the duration of the match and higher the scoring difference between the winning team and the losing one. This allowed for a suggestion of a directive for teams matchmaking based on the methodology followed by this paper which can be used in other games that have teams divided in matches.

**Keywords:** *Game Metrics. Game Analytics. Match Making. Unsupervised learning and clustering. Massively Multiplayer Online Games. Data Mining. Clustering. Unsupervised Learning. Neural Networks.*



# Lista de ilustrações

Figura 1 – Fluxograma da visão geral do trabalho . . . . .	4
Figura 2 – Exemplo de visualização de <i>SOM</i> em <i>U-Matrix</i> . . . . .	18
Figura 3 – Passos para início de partida no Dota 2 . . . . .	20
Figura 4 – Mapa detalhado do Dota 2 . . . . .	22
Figura 5 – Trilhas e suas dificuldades no mapa do Dota 2 . . . . .	23
Figura 6 – Etapas do Processo de Descoberta de Conhecimento em Banco de Dados (DCBD) utilizado neste trabalho . . . . .	29
Figura 7 – Estrutura de dados típica de um Sumário de Partida: a) dados gerais da partida; b) dados específicos de um jogador. . . . .	30
Figura 8 – Estrutura de dados do descritor de cada usuário para cada partida . . . . .	33
Figura 9 – Vetor de características de um usuário . . . . .	34
Figura 10 – Topologia Escolhida para o <i>SOM</i> . . . . .	40
Figura 11 – Matrizes de coeficientes de correlação . . . . .	46
Figura 12 – Mapa Final de Neurônios Classificados . . . . .	47
Figura 13 – Representações <i>U-Matrix</i> e <i>P-Matrix</i> do <i>SOM</i> . . . . .	48
Figura 14 – Relação entre <i>Denies</i> , <i>HeroDamagePerMin</i> , <i>Votes</i> e <i>Leaver</i> . . . . .	49
Figura 15 – Relação entre <i>Kills</i> , <i>HeroDamagerPercent</i> e <i>GoldSpentPercent</i> . . . . .	49
Figura 16 – Relação entre <i>FirstBloodTime</i> e <i>Leaver</i> . . . . .	50
Figura 17 – Relação entre <i>AssistsDeaths</i> e <i>KillsDeaths</i> . . . . .	50
Figura 18 – Relação entre <i>GoldPerMin</i> e <i>LastHitsPerMin</i> . . . . .	51
Figura 19 – Distribuição de classes ao longo do mapa ESOM. . . . .	51
Figura 20 – Visualizações de Classificação de Regiões com <i>ESOM</i> . . . . .	52
Figura 21 – Métricas Médias das Classes . . . . .	54
Figura 22 – Normalização Relativa das Médias . . . . .	55
Figura 23 – Histograma da Duração da Partida em Segundos . . . . .	60
Figura 24 – Duração por Desbalanço . . . . .	62
Figura 25 – Delta Mortes por Grau de Desbalanço . . . . .	63
Figura 26 – Histograma de Delta Mortes . . . . .	63
Figura 27 – Matriz de Coeficientes de Correlação Pearson . . . . .	78
Figura 28 – Matriz de Coeficientes de Correlação Spearman . . . . .	79
Figura 29 – Todos os Componentes do Mapa ESOM, Antes da Eliminação de Atributos . . . . .	81
Figura 30 – Influência dos Pesos dos Neurônios para cada Componente do Vetor de Características de Usuário na Composição final do <i>SOM</i> . . . . .	82
Figura 31 – Relação de Assassínatos sobre Mortes ( <i>KillsDeaths</i> ) . . . . .	83
Figura 32 – Relação de Assassínatos em Relação ao Time ( <i>KillsPercent</i> ) . . . . .	83
Figura 33 – Relação de Total de Assassínatos em Relação ao outro ( <i>KillsBalance</i> ) . . . . .	83

Figura 34 – Relação de Assistências sobre Mortes ( <i>AssistsDeath</i> ) . . . . .	84
Figura 35 – Relação de Assistências sobre Total de Assistências do Time ( <i>Assists-Percent</i> ) . . . . .	84
Figura 36 – Relação de Mortes sobre Total de Mortes do Time ( <i>DeathsPercent</i> ) . . . . .	84
Figura 37 – Quantidade de Negações ( <i>Denies</i> ) . . . . .	85
Figura 38 – Duração da Partida ( <i>Duration</i> ) . . . . .	85
Figura 39 – Horário do Evento do Primeiro Assassinato ( <i>FirstBloodTime</i> ) . . . . .	85
Figura 40 – Relação de Ganho de <i>Gold</i> por Minuto ( <i>GoldPerMinut</i> ) . . . . .	86
Figura 41 – Relação entre o <i>Gold</i> gasto em Relação ao Total da Equipe ( <i>GoldSpent-Percent</i> ) . . . . .	86
Figura 42 – Quantidade da Dano causado a Oponentes por Minuto ( <i>HeroDamagePerMin</i> ) . . . . .	86
Figura 43 – Quantidade da Dano causado a Oponentes sobre o Total feito pela Equipe ( <i>HeroDamagePercent</i> ) . . . . .	87
Figura 44 – Quantidade de Cura a Aliados por Minuto ( <i>HeroHealingPerMin</i> ) . . . . .	87
Figura 45 – Quantidade entre a Quantidade de Cura a Aliados Sobre o Total feito pela Equipe ( <i>HeroHealingPercent</i> ) . . . . .	87
Figura 46 – Quantidade de Eliminação de <i>Creeps</i> por Minuto ( <i>LastHitsPerMin</i> ) . . . . .	88
Figura 47 – Quantidade de Dano feito a Estruturas ( <i>TowerDamage</i> ) . . . . .	88
Figura 48 – Quantidade de Dano feito a Estruturas em relação ao Total feito pela Equipe ( <i>TowerDamagePercent</i> ) . . . . .	88
Figura 49 – Relação entre Vitórias e Derrotas ( <i>WinsPercent</i> ) . . . . .	89
Figura 50 – Quantidade de Experiência Ganha por Minuto ( <i>XpPerMin</i> ) . . . . .	89
Figura 51 – Abandono de Partidas ( <i>Leaver</i> ) . . . . .	89
Figura 52 – Votos recebidos, positivos e negativos ( <i>Votes</i> ) . . . . .	90
Figura 53 – Distribuição Relativa de Métricas por Classes . . . . .	91
Figura 54 – Normalização das Médias . . . . .	91



# Lista de tabelas

Tabela 1 – Distribuição Percentual de Usuários por Classe . . . . .	53
Tabela 2 – Impacto do Balanço de Equipes no Jogo Dota 2 . . . . .	59
Tabela 3 – Duração por Desbalanço . . . . .	61
Tabela 4 – Delta Mortes por Grau de Desbalanço . . . . .	62
Tabela 5 – Média de Métricas Absolutas por Grupo . . . . .	92
Tabela 6 – Percentual de Média de Métricas por Grupo . . . . .	92
Tabela 7 – Média Padronizada de Métricas por Grupo . . . . .	93



# Lista de abreviaturas e siglas

AD	<i>AssistsDeath</i>
AP	<i>AssistsPercent</i>
AFK	<i>Away From Keyboard</i>
DCBD	<i>Descoberta de Conhecimento em Banco de Dados</i>
Den	<i>Denies</i>
Dur	<i>Duration</i>
DP	<i>DeathsPercent</i>
ESOM	<i>Emergent Self-Organized Maps</i>
ESPORT	<i>Electronic Sport</i>
FBT	<i>FirstBloodTime</i>
GPM	<i>GoldPerMinute</i>
GSP	<i>GoldSpentPercent</i>
HDP	<i>HeroDamagePercent</i>
HDPM	<i>HeroDamagePerMinute</i>
HHP	<i>HeroHealingPercent</i>
HHPM	<i>HeroHealingPerMinute</i>
KB	<i>KillsBalance</i>
KD	<i>KillDeath</i>
KP	<i>KillsPercent</i>
Lea	<i>Leaver</i>
LHPM	<i>LastHitPerMinute</i>
MOBA	<i>Multiplayer Online Battle Arena</i>
PDE	<i>Pareto Density Estimation</i>

RTS	<i>Real-Time Strategy</i>
SOM	<i>Self-Organized Maps</i>
TD	<i>TowerDamage</i>
TDP	<i>TowerDamagePercent</i>
Vot	<i>Votes</i>
XPM	<i>ExperiencePerMinute</i>
WEBAPI	<i>Web Application Program Interface</i>
WP	<i>WinsPercent</i>

# Sumário

<b>Introdução</b>	<b>1</b>	
<b>Motivação</b>	<b>2</b>	
<b>Objetivos</b>	<b>3</b>	
<b>Visão Geral da Proposta</b>	<b>4</b>	
<b>I</b>	<b>PREPARAÇÃO DA PESQUISA</b>	<b>7</b>
<b>1</b>	<b>TRABALHOS RELACIONADOS</b>	<b>9</b>
1.1	Telemetria de Dados de Comportamento de Usuários em Jogos	10
1.2	Play-personas	11
1.3	Game Design Orientado a Métricas	12
1.4	Previsão de Comportamento	13
1.5	Previsão de Balanceamento	13
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>15</b>
2.1	Mapas Auto-Organizáveis de Kohonen	15
2.2	ESOM	16
2.2.1	<i>U-Matrix</i>	17
2.2.2	<i>P-Matrix</i>	18
2.3	<i>Dota 2</i>	19
2.3.1	Papéis dos Heróis	22
2.4	Considerações Finais	25
<b>II</b>	<b>PROPOSTA</b>	<b>27</b>
<b>3</b>	<b>ABORDAGEM PROPOSTA</b>	<b>29</b>
3.1	Coleta de Dados	29
3.1.1	Extração de Dados por Varredura - <i>Crawling</i>	30
3.2	Pré-Processamento - Segregação por Dados de Usuário em Cada Partida	32
3.3	Transformação	34
3.3.1	Estatísticas de Desempenho dos Usuários	34
3.3.2	Eliminação de Atributos	38
3.4	Mineração	39
3.4.1	Treinamento e Classificação com <i>ESOM</i>	39
3.4.2	Classificação de Usuários com <i>ESOM</i>	41

<b>3.5</b>	<b>Interpretação</b> . . . . .	<b>41</b>
3.5.1	Rotulação . . . . .	41
3.5.2	Identificar a composição de equipes . . . . .	41
<b>3.6</b>	<b>Considerações Finais</b> . . . . .	<b>42</b>
<b>III</b>	<b>PARTE FINAL</b>	<b>43</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	<b>45</b>
<b>4.1</b>	<b>Coleta e Tratamento de Dados</b> . . . . .	<b>45</b>
<b>4.2</b>	<b>Treinamento e Classificação</b> . . . . .	<b>46</b>
4.2.1	Treinamento . . . . .	47
4.2.2	Agrupamento e Classificação . . . . .	51
<b>4.3</b>	<b>Interpretação</b> . . . . .	<b>56</b>
<b>4.4</b>	<b>Interpretação da Formação de Equipes</b> . . . . .	<b>58</b>
<b>5</b>	<b>CONCLUSÃO</b> . . . . .	<b>65</b>
<b>5.1</b>	<b>Trabalhos Futuros</b> . . . . .	<b>66</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>69</b>
	<b>APÊNDICES</b>	<b>75</b>
	<b>APÊNDICE A – MATRIZES DE COEFICIENTES DE CORRELAÇÃO ENTRE ATRIBUTOS</b> . . . . .	<b>77</b>
	<b>APÊNDICE B – TODOS OS COMPONENTES DO MAPA ESOM</b>	<b>81</b>
	<b>APÊNDICE C – MATRIZES E TABELAS DAS CARACTERÍSTICAS MÉDIAS DE AGRUPAMENTOS DE USUÁRIOS</b> . . . . .	<b>91</b>

# Introdução

No processo de modernização da sociedade a tecnologia impactou em diversos aspectos do comportamento humano. Um dos comportamentos que mais evoluiu foi a forma na qual as pessoas buscam entretenimento. Dentre as indústrias de entretenimento, a de jogos eletrônicos, tornou-se a mais lucrativa na última década (BOSC et al., 2013).

Uma meta buscada pelos desenvolvedores de jogos modernos é construir algo para ser jogado mais de uma vez. Deve-se prolongar a vida de um produto, dando-lhe um maior valor de repetição (DETERDING et al., 2011). No entanto, os usuários estão se tornando mais seletivos e exigentes sobre quais jogos eles jogam, além de não gostarem de experiências repetitivas ao longo do tempo (KATAVIĆ, 2013). É necessário prover experiências diferentes a cada nova partida.

Jogos eletrônicos são o ramo da indústria de entretenimento que mais cresce, superando o cinema (WOLF, 2001). Atualmente, cerca de 32% do tempo gasto em smartphones concentra-se em jogos (FLURRY, 2014). Um bom exemplo disso pode ser visto no jogo *Dota 2*. Ele possui mais de 750 mil usuários ativos diários (VALVE, 2014b) e quase 10 milhões de usuários únicos mensais (VALVE, 2014a).

O crescimento dos jogos como meio de diversão atingiu patamares impressionantes. Em 2003 ocorreu a primeira “Copa Mundial de Esportes Eletrônicos”, reunindo mais de 150 mil jogadores no mundo todo (TRINCA et al., 2010). Atualmente existem campeonatos que distribuem prêmios generosos aos seus vencedores, como por exemplo, o campeonato mundial de *Dota 2* que distribuiu U\$2.874.381,00 em 2013 (EARNINGS, 2014), o “*League of Legends World Cup*” que distribuiu U\$2.050.000,00 em 2013 (EARNINGS, 2014), o “*Call of Duty Championship 2013*” que distribuiu U\$1.000.000,00 (EARNINGS, 2014). Em 2014, o campeonato “*The International Dota 2 Championships*” se destacou por distribuir aproximadamente o triplo do valor da maior premiação do ano anterior, com premiação de U\$10.931.103,00 (VALVE, 2014d). Além disso, esse campeonato foi transmitido em rede televisiva de largo alcance, via *ESPN* (ESPN, 2014).

Um grande atrativo para jogos é a sua pluralidade. Tantos amadores quanto jogadores casuais podem se destacar e até se tornar profissionais. O fato de ser um esporte em um meio virtual, com jogadores conectados via rede de computadores, permite jogar com e contra pessoas em suas próprias residências e em diferentes cidades, o que facilita o acesso e a criação de uma comunidade em escala global. Entretanto, como muitos dos jogos *online* são cooperativos ou competitivos, é muito importante que se formem equipes coordenadas para que se obtenha um melhor desempenho.

Visando ampliar ainda mais o alcance dos jogos competitivos, é necessário melhorar

a experiência de usuário para os jogadores. A formação de equipes balanceadas é algo que pode contribuir bastante nessa direção. Com equipes balanceadas, as partidas tendem a ser mais disputadas, com chances de vitória para ambos os lados, e ainda potencializando a duração da disputa e até mesmo estimulando que seus usuários adquiram produtos relacionados ao jogo.

Este trabalho visa realizar um estudo sobre a formação de equipes em jogos eletrônicos, na perspectiva de propor uma abordagem para entendimento das características que influenciam no balanceamento das partidas e, a partir disso, sugerir formação de equipes que visem maximizar a disputa. A formação de equipes deve ser feita de maneira a gerar maior nível de tensão em ambos os lados que competem e isso está diretamente relacionado à distribuição de usuários nos times. Para isso ser possível, é necessário conhecer o perfil de cada jogador e com isso sugerir uma composição mais equilibrada.

A principal contribuição deste trabalho consiste na caracterização de usuários de um jogo real e comercial do tipo *MOBA* para entender como as diferentes classes de usuários impactam na qualidade das partidas e, em seguida, sugerir melhoria na formação de partidas, evitando que já comecem desbalanceadas.

Para realizar o trabalho proposto, foi utilizado o jogo *Dota 2* como referência real para o desenvolvimento do estudo. O Dota é muito conhecido, bastante jogado e que disponibiliza uma base de informações sobre o jogo, o que permite seu uso neste tipo de pesquisa.

## Motivação

Para acompanhar o progresso do ramo de tecnologia de jogos, surgiram regras, sistemas de controle, metodologias e algoritmos para garantir o melhor equilíbrio de equipes, bem como aumentar os níveis de diversão e engajamento dos usuários. Esse sistema de registro e controle atua, na maioria das vezes, de forma transparente ao usuário, normalmente via captura e armazenagem de informações relevantes para posterior análise. Tais informações podem ser métricas diretas, como a quantidade de cliques do jogador por minuto, ou derivadas e mais complexas, como o quanto um jogador se movimenta próximo do resto da equipe, que pode ser usada para inferir se prefere jogar sozinho ou em grupo.

A coleta e mineração de dados de como um usuário interage com o jogo é de extrema importância para se obter informações úteis. Esses dados podem e devem ser monitorados e essa prática permite detectar falhas, prever comportamentos, aumentar lucros de venda de produtos dentro do jogo, realizar ajustes finos para remover problemas que atrapalhem a imersão ou desagradem de alguma maneira o usuário, dentre outras possibilidades.



Algumas empresas empreendem tanto esforço no tratamento de dados de telemetria e métricas que chegam a ser rotuladas como companhias de armazém de dados (*Data Warehouse*) mascaradas como empresas de jogos, assim como afirma o vice presidente da *Zynga* (uma das maiores do ramo), Ken Rudino, “*We’re an analytics company masquerading as a games company*” (FISHER et al., 2012).

Da mesma forma que não é agradável jogar um esporte físico em uma equipe desbalanceada em relação à outra, essa mesma sensação também se faz presente em um jogo eletrônico baseado na disputa de equipes. A diversão nesse caso está justamente na tensão gerada pela ansiedade de sair da situação de equilíbrio inicial para começar a pender em direção ao lado que irá ganhar a partida.

Nas diversas comunidades associadas na Internet existem relatos de reclamações de jogadores sobre a formação de equipes. Esses relatos mostram o aborrecimento de um jogador ao disputar um jogo com uma equipe desequilibrada. Por conta disso, a grande maioria dos jogos já incluem um mecanismo de seleção de membros para uma equipe visando equilibrar os times. No entanto, ainda existe espaço para melhorar esses mecanismos. Neste trabalho foi atacado esse problema.

A definição do jogo *Dota 2* como fonte desta pesquisa fundamentou-se em diversas razões. O *Dota 2* é produzido pela empresa *VALVE* e distribuído eletronicamente pela loja virtual *STEAM*. Sua abrangência é mundial. Esse jogo é hoje um dos maiores e mais competitivos do gênero *Multiplayer Online Battle Arena (MOBA)*. Além disso, o *Dota 2* possui uma base de dados de estatísticas de partidas que permitem seu uso para fins acadêmicos.

## Objetivos

### Objetivo Geral

O principal objetivo deste trabalho é a realização de um estudo sobre a formação de equipes em jogos que envolvam a disputa entre equipes, visando identificar características que favoreçam o equilíbrio das partidas e com isso sugerir a composição dos times para maximizar o engajamento dos jogadores e conseqüentemente a qualidade das partidas.

### Objetivos Específicos

Para se alcançar o objetivo geral do trabalho são necessárias diversas ações diretamente ligados ao entendimento sobre a composição de equipes. Para isso, serão objetivos específicos:

- Definição de jogos a serem usados como fonte de pesquisa para o trabalho;

- Coleta de dados de partidas para permitir a análise de características dos jogadores e das partidas;
- Definição de indicadores que favoreçam o entendimento dos aspectos relacionados ao jogo, tanto em nível de jogador quanto ao nível de partida;
- Realização de uma análise dos jogadores para entendimento dos grupos existentes;
- Análise das partidas, visando identificar características de boas partidas e composição das equipes nesses casos;
- Sugestão de diretrizes para formação de equipes com base nas informações geradas.

## Visão Geral da Proposta

Neste trabalho foi utilizada uma metodologia de atuação baseada na Descoberta de Conhecimento em Banco de Dados (DCBD). Essa metodologia é usada para auxiliar na inferência de relações de interesse que não são facilmente observadas por especialistas no assunto, além de servir como forma de auxílio para a validação de conhecimento (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996b) (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996a). A metodologia divide-se em cinco etapas básicas consistindo em Coleta, Pré-Processamento, Transformação, Mineração, Avaliação e Interpretação de Resultados.

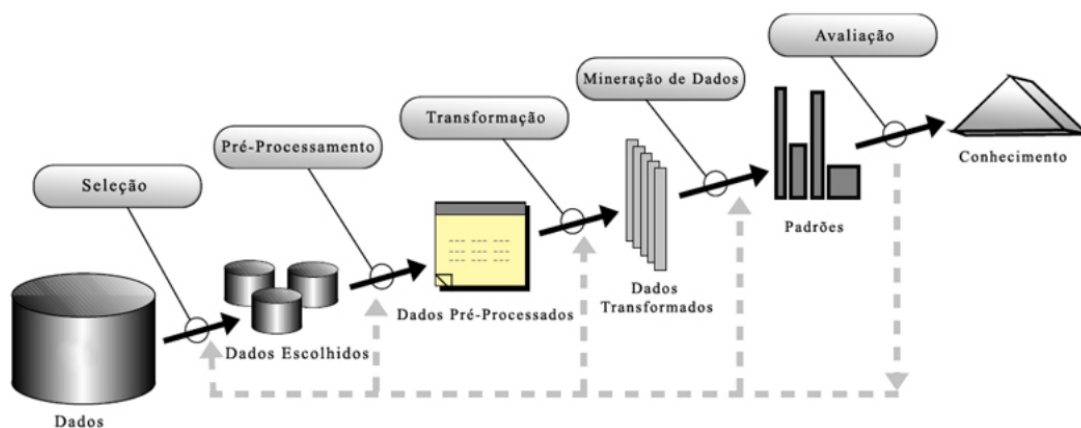


Figura 1 – Fluxograma da visão geral do trabalho

Na etapa de coleta foram adquiridos dados públicos do jogo DOTA 2 referentes a partidas realizadas durante seis meses. Nesse período, os relatórios de fim de partida, também chamados de “Sumários” foram armazenados, contendo dados individuais de desempenho de cada jogador e dados globais de uma partida.

No pré-processamento realizou-se a extração de dados úteis e o tratamento de valores brutos oriundos da base de sumários. A extração consistiu em segregar dados individuais de usuários dos dados globais da partida e, o tratamento, na eliminação de entradas de sumários que não atendiam aos critérios estabelecidos.

Na transformação foi feito um tratamento de dados de partidas e foram criadas estatísticas de como cada usuário joga. Dentre as características extraídas, algumas foram eliminadas usando-se critérios de correlação.

A etapa de mineração consistiu em classificar os vetores de características extraídos da base. A classificação foi feita pelo método *ESOM*.

Na etapa de interpretação realizou-se a descoberta e caracterização dos agrupamentos de jogadores. Foram encontrados 12 agrupamentos e para cada um foi feita uma rotulação e caracterização. Com base nesses dados foram realizadas várias avaliações para se entender a composição das equipes nas partidas mais disputadas, para se chegar a uma diretriz sobre como deveria ser a composição dos times para favorecer a disputa.

## Organização do Documento

Este trabalho está organizado em 4 capítulos. No Capítulo 1 são analisados os trabalhos relacionadas. No Capítulo 2 é feita uma apresentação do referencial teórico utilizado como base do trabalho. No Capítulo 3 é descrita a abordagem proposta, juntamente com a metodologia utilizada e as tarefas realizadas em cada uma das fases. No Capítulo 4 são discutidos os resultados alcançados, as conclusões obtidas, e apresentado um direcionamento para trabalho futuros.



# Parte I

## Preparação da pesquisa



# 1 Trabalhos Relacionados

Neste capítulo são discutidos alguns dos artigos mais influentes na área de *game analytics*, incluindo telemetria, classificação de usuários com *play-personas*, customização, *game design* orientado à métricas, previsão de comportamento e previsão de balanceamento de equipes.

*Game Analytics* é uma área de pesquisa bastante nova, surgiu há cerca de 5 anos, impulsionada por fatores como a demanda por novos modelos de negócios e para isso, expandiu a pesquisa focada em usuários. Foi constatado que ao se analisar, monitorar e sintetizar o comportamento do usuário, pode-se obter uma melhoria na qualidade do produto entregue. É uma área que faz parte de *business intelligence* que é a evolução do que se iniciou com telemetria, porém aplicada a jogos. As ferramentas criadas e empregadas para esta área não são uma receita definitiva do sucesso, contudo certamente garantem um *gameplay* e economia balanceada (EL-NASR; DRACHEN; CANOSSA, 2013a).

Por ser recente, a terminologia empregada em *game analytics* pode parecer confusa e não padronizada. Seus termos e metodologias são geralmente emprestados de outras áreas e adaptados. Além disso, é importante saber que existem muitos tipos de fontes de telemetria e métricas usadas em *game analytics* e inteligência empresarial para o desenvolvimento de jogos.

Atualmente, a maioria dos trabalhos publicados são baseados em métricas de usuário e em especial, telemetria de comportamento de usuário (*user behavior telemetry*). Isto é algo a se esperar, uma vez que os usuários (o termo técnico empregado no contexto de jogos virtuais é *players*) são um fator crucial para o sucesso de um jogo. Jogos são produtos focados primariamente em entregar experiência de usuário. Assim, a análise e interpretação de como os usuários interagem ou atuam nos jogos é a fonte primária de informação para uma efetividade do *game design* em gerar e entregar experiências engajantes.

Alguns termos como *Game Metrics*, *Game Telemetry* e *Game Analytics* são intercambiáveis (EL-NASR; DRACHEN; CANOSSA, 2013b) e, para facilitar o seu entendimento, torna-se necessário ressaltar as suas peculiaridades. *Game Metrics* é a aplicação de *analytics* para desenvolvimento de jogos e pesquisas. *Game Analytics* dá suporte a tomada de decisão, em níveis operacionais, táticos e estratégicos de uma organização, analisando *design*, arte, programação, *marketing* e pesquisa de usuário. *Game Analytics* é a fonte de informação para *business intelligence* em desenvolvimento de jogos e considera jogo tanto como o produto, como o negócio a desenvolver e manter. Nos últimos anos, muitas empresas de jogos, desde *indies* (desenvolvedores independentes) até gigantes da indústria

começaram a coletar dados de telemetria. *Telemetry* refere-se a dados coletados à distância, os quais geralmente podem ser quantitativos, sobre como um usuário joga um jogo, obtidos no cliente e transmitidos para uma coleção de servidores. *Game metrics* são medidas interpretadas de algo relacionado a jogos. Mais especificamente, são medidas quantitativas de atributos de objetos. Uma fonte comum de *Game Metrics* é a telemetria de dados de comportamento de jogador. Esses dados brutos podem ser transformados em métricas, tais como horas totais de jogo e quantidade de usuários ativos diários. Essas medidas descrevem um atributo ou propriedade dos jogadores. *Metrics* são mais do que apenas medidas de comportamento de jogador, entretanto o termo cobre as referências *business intelligence* que operam no contexto de jogos.

Este conhecimento é fundamental para o entendimento de problemas relacionados à formação de equipes em jogos virtuais, uma vez que auxilia a compreensão sobre como foram resolvidos e como a comunidade científica avaliou a qualidade da solução.

## 1.1 Telemetria de Dados de Comportamento de Usuários em Jogos

Telemetria é o processo de medir e armazenar dados remotamente. Nesse sentido, o *Tracking Real-time User Experience (T.R.U.E.)* (KIM et al., 2008) aborda alguns aspectos que são importantes para se trabalhar com telemetria em jogos. A instrumentação clássica faz uso de câmeras para gravar a tela do jogo e também o rosto do jogador, para capturar todas as suas reações. Trata-se de metodologia demorada e onerosa, pois requer um tipo de profissional especializado em testar jogos e outro especializado em observar e detectar as reações em todas as gravações.

A solução apresentada por Kim et.al. (2008) para o problema de captura e armazenamento de dados de uso do jogo baseia-se em questionários que devem ser respondidos em tempos pre-determinados. Esta metodologia foi aplicada em um jogo real produzido pela *Microsoft*, o *Halo 2*. A cada 5 minutos de uso o usuário é interrompido para responder a um questionário que pergunta a opinião do jogador sobre a dificuldade que está enfrentando atualmente. Com isto, é possível a detecção de erros e falhas no *design* do jogo, porém, com prejuízo na imersão.

Apesar de ser uma metodologia bastante intrusiva, interrompendo o jogador constantemente, apresentou resultados relevantes. Contudo, funciona apenas em ambientes controlados, não podendo ser estendida a um produto final. Envolve custos na contratação de profissionais especializados em testes, porém são menores do que os da metodologia tradicional que envolve a observação de gravações de vídeos.

O *T.R.U.E.* é importante, pois é uma das bases para a metodologia de telemetria em jogos. Descreve testes de experiência de jogador em ambientes controlados e explica como fazer diversas tarefas na área de teste de maneira mais automatizada.



Um trabalho recente aborda metodologia diferente para a coleta de dados de telemetria, sem que seja necessário alterar o código do jogo ou *Game Engine* (CLUA et al., 2014). No trabalho, é feita a extração de dados úteis a partir da análise e processamento de imagem do que o usuário está observando. No experimento analisou-se um jogo tradicional chamado “*Super Mario World*” e conseguiu encontrar agentes, entidades e ações. Esta abordagem pode ser estendida para qualquer jogo em 2D se tomadas algumas precauções e pode prover a base para trabalhos futuros em diferentes tipos de jogos.

A coleta de dados de telemetria pode ser utilizada de diversas formas, inclusive dentro do próprio jogo, para ajudar o usuário a entender onde errou. Esta metodologia é conhecida como *Game Provenance* (KOHWALTER; CLUA; MURTA, 2012). É usada para melhorar o processo de aprendizado no contexto de *Serious Games*. Funciona coletando dados da sessão de jogo para posterior análise, ajudando a entender quais decisões foram feitas ao longo do tempo. Estas decisões são exibidas ao jogador para que ele as analise e identifique os pontos mais importantes e os seus impactos em cada uma de suas decisões.

## 1.2 Play-personas

*Play-personas* são *clusters* de interações preferenciais e atitudes temporalmente expressas que se aglutinam ao redor de diferentes formas de interação, providos pelos *game designers* (CANOSSA; DRACHEN, 2009). Trata-se de uma abordagem que busca solucionar o problema de como simular o comportamento humano em interações com jogos a fim de criar uma inteligência artificial que seja compatível com a habilidade do usuário ou simular um outro humano (CARDAMONE; LOIACONO; LANZI, 2009).

Alguns trabalhos descrevem a criação de perfis de como os usuários jogam, ou *play-personas* (TYCHSEN; CANOSSA, 2008) (CANOSSA, 2009) (CANOSSA; DRACHEN, 2009) (HOORN et al., 2009) (CARDAMONE; LOIACONO; LANZI, 2009) (ORTEGA et al., 2013) (CANOSSA, 2013). O que há de comum nestes trabalhos é que sempre usa-se uma base de dados de métricas para definir e avaliar padrões de comportamento e tomada de decisão dos usuários. Com isso, criam-se modelos de jogadores que podem ser usados para melhorar a experiência ao conseguir balancear o grau de dificuldade mais adequado.

Um trabalho bastante influente nesta área utilizou algoritmos evolucionários multi-objetivos para modelar o comportamento do usuário a fim de simular uma inteligência artificial que seja sempre desafiadora (HOORN et al., 2009). A referida abordagem mostrou-se capaz de generalização quando inserida em ambientes novos, apresentando um balanço entre a simulação do comportamento humano e propiciando um maior engajamento ao jogar. Contudo, o autor reitera que a abordagem utilizada de um aproximador de sequências juntamente com algoritmos de aprendizagem de máquina não é suficiente para modelar o comportamento humano como esperado.

O conceito de *play-personas* pode ser usado para oferecer conteúdos customizados que mais agradem ao jogador, permitindo ajustar dinamicamente a dificuldade e o comportamento dos inimigos, para sempre deixar o usuário com um elevado nível de tensão (YANNAKAKIS; TOGELIUS, 2011). Os autores apresentam uma forma de criar motoristas artificiais que desenvolvem o seu comportamento por meio do aprendizado por imitação, utilizando métodos de aprendizado não-supervisionados. Ao invés de prever o *input* puro que um humano faria, a abordagem faz uma previsão das ações tomadas em alto nível, que consegue alcançar em alguns casos, uma diferença de menos do que 15% da eficiência do melhor motorista artificial, o que é algo bastante positivo.

Para o presente trabalho, *play-personas* é importante para ser usado como referência no planejamento de como se modela o perfil de comportamento de usuário, a partir de dados de telemetria e métricas de desempenho.

### 1.3 Game Design Orientado a Métricas

Como já foi abordado, muitas tarefas da produção de jogos são complexas e demoradas, principalmente a de *game design*. Buscando automatizar e otimizar a produção deste setor, é possível evoluir as regras que definem o jogo (TOGELIUS; SCHMIDHUBER, 2008) (HARRISON; ROBERTS, 2013).

Foi realizado um estudo de caso envolvendo a criação de uma ferramenta *online* para analisar o comportamento de jogadores dentro de uma sessão de jogo, visando ajudar no processo de ajuste de conteúdo (MEDLER; JOHN; LANE, 2011). Foi detectado que ao se acoplar a produção de uma ferramenta de *game analytics* juntamente com a produção de um jogo, no caso o *Dead Space 2*, despertou-se o interesse dos times de desenvolvedores e permitiu que o jogo continuasse em produção mesmo depois de lançado, fazendo com que sempre evoluísse e melhorasse. Ao descrever a ferramenta utilizada, *Data Cracker*, é apresentado um guia para desenvolver ferramentas semelhantes, as arquiteturas utilizadas e como os dados são processados de maneira instantânea ou, como realizar observações de longo prazo. Esta ferramenta provou ser capaz de detectar diversas falhas (*bugs*) no produto, também para melhorar o game design ao ajustar parâmetros.

Um trabalho recente (HARRISON; ROBERTS, 2014) usa esta abordagem para adaptar o conteúdo oferecido a fim de melhorar a retenção do usuário por sessão. Utiliza-se de conceitos de *game analytics* para reduzir o escopo do problema e tornar o problema mais tratável. Modela-se a retenção do usuário em um espaço e ajusta-se parâmetros de forma a influenciar o comportamento do usuário. Os experimentos mostraram que a versão adaptativa do jogo conseguiu abranger uma quantidade maior de estados de jogo e também diminuiu a taxa de abandono. A validação foi feita usando o teste exato de Fisher (UPTON, 1992).

A contribuição adquirida com customização no presente trabalho situa-se no campo de sistemas de recomendação, tomando como oferta de conteúdo, a formação de grupos com maior potencial competitivo. A importância desta abordagem consiste na possibilidade de usar as métricas para modificar as regras do jogo, fazendo-o evoluir, até mesmo evitar ou corrigir falhas de design ou balanceamento.

## 1.4 Previsão de Comportamento

Além de criar *play-personas*, é possível categorizar os usuários para melhor atender o que cada tipo de usuário busca, podendo ajustar parâmetros do jogo ou *design* para agradar especificamente um determinado tipo de jogador (LOPES; BIDARRA, 2011). Alguns trabalhos foram feitos para analisar um banco de dados de como jogadores atuam e categorizá-los de acordo com a forma em que eles avançam em uma sessão de jogo (DRACHEN; CANOSSA; YANNAKAKIS, 2009) (HARRISON; ROBERTS, 2012). De posse dessa informação, os *game designers* possuem a liberdade e a capacidade de modificar elementos do jogo para melhor atingir o público esperado.

Para classificar como os jogadores atuam bem como seu comportamento durante a partida, foi realizado um trabalho relevante que é capaz de prever quando o jogador vai parar de jogar e encerrar uma sessão de jogo (MAHLMANN et al., 2010). De posse deste conhecimento, é possível tomar medidas que ofereçam opções no sentido de evitar o abandono do usuário, ou por exemplo, sugerir a opção de pausar e pedir para voltar em seguida.

No presente estudo busca-se formar equipes que minimizem a taxa de abandono de uma partida e, a partir da informação gerada pela categorização, formar grupos de usuários que possuam um comportamento semelhante na forma de jogar e, sejam compatíveis entre si para estarem na mesma equipe.

## 1.5 Previsão de Balanceamento

Em jogos *MOBA* é muito importante ranquear fatores de sucesso dos times para eleger jogadores e distribuí-los de tal forma que se encontre o equilíbrio entre as equipes. Existem trabalhos realizados para tentar solucionar este problema usando também o jogo *Dota 2* (POBIEDINA et al., 2013a) (POBIEDINA et al., 2013b).

Os trabalhos citados utilizam análises estatísticas de dados sociais como rede de amigos e jogos comprados, caracterizando jogadores a partir de dados globais do perfil do usuário providos pela *STEAM* (plataforma *online* de suporte aos jogos vendidos pela *VALVE*). Os autores usam como características definidoras do jogador, o desempenho em

outros jogos e a quantidade de amigos, o que pode causar uma contribuição duvidosa no perfil de desempenho no jogo estudado.

Um trabalho recente (YANG; HARRISON; ROBERTS, 2014), que também aborda jogos do gênero *MOBA*, identifica táticas de combate que podem contribuir para o sucesso de um time. O combate foi modelado como uma sequência de grafos como métricas (KARZANOV, 1998). Experimentos mostram que os padrões extraídos alcançaram cerca de 80% de precisão de previsão em testes com partidas novas.

Um trabalho aborda a distribuição dos integrantes das equipes em jogos *MOBA* estudando como o comportamento varia em função do nível dos times, usando 4 categorias que vão de novato até profissional (DRACHEN et al., 2014). Utilizou-se uma abordagem de clusterização temporal de distanciamento entre avatares a cada segundo. O desafio da metodologia foi encontrar uma função de similaridade entre as séries temporais, sendo para isso utilizada a Distribuição de Permutação (*Permutation Distribution*). Como consequência, foi possível mostrar diferentes formas nas quais o comportamento varia em cada agrupamento de nível de times. Este resultado pode ser usado para balancear equipes na formação de uma partida.

No presente trabalho, toma-se como foco as estatísticas de desempenho de cada jogador na partida do *Dota 2*, ao contrário do primeiro autor mencionado (POBIEDINA et al., 2013a) (POBIEDINA et al., 2013b), o que confere uma riqueza maior de detalhes de características. Já o último autor (DRACHEN et al., 2014) usou uma abordagem bastante complexa e custosa para tratar este mesmo problema observando não apenas contadores de fim de partidas, mas dados coletados no seu decorrer.

## 2 Referencial Teórico

Neste capítulo descreve-se uma classe de algoritmos de inteligência artificial e suas extensões que serão utilizadas para extrair conhecimento da base de dados. Além disso é apresentado o funcionamento do jogo *Dota 2*, mostrando o fluxo de conexão, os papéis dos avatares, como se ganha a moeda financeira dentro do jogo e as métricas estatísticas de cada jogador.

### 2.1 Mapas Auto-Organizáveis de Kohonen

O nome *Self-Organizing Map (SOM)* significa uma classe de algoritmos de redes neurais na categoria de aprendizado de máquina não supervisionado (KOHONEN, 1990). Esta classe de algoritmos é utilizada para extrair regularidades estatísticas (padrões) presentes em conjuntos de dados (KOHONEN, 1977) (KOHONEN; TEUVO, 1995) para posteriormente serem usadas na classificação. No *SOM* cada ponto do mapa representa um neurônio da rede.

Uma das propriedades *SOM* é a formação de uma projeção não linear de múltiplos dados de alta dimensionalidade em uma estrutura regular de baixa dimensionalidade, usualmente em uma grade (que apresenta uma topologia particular, que pode ser retangular, hexagonal etc) de duas dimensões. Mapas de dimensões maiores também são possíveis, porém são de mais difícil aplicação e compreensão. Essa projeção reflete o agrupamento do espaço de dados e as relações métricas-topológica dos itens de dados. O *SOM* pode ser utilizado com itens tomados como vetores, tais como os descritores de dados estatísticos ou medições que descrevem um processo. O *SOM* é o mapa sobre o qual cada um dos itens de dados pode ser exibido separadamente utilizando a escala de cinza ou cores. Este tipo de visualização combinada é muito útil para a compreensão das dependências mútuas entre as variáveis, assim como as estruturas do conjunto de dados.

Nessa classe de algoritmos, o processo de aprendizagem é baseado no aprendizado competitivo, possui natureza de busca local e as modificações dos pesos sinápticos são confinadas à vizinhança (região elegível a ter atualização dos pesos sinápticos) do neurônio ativado (PRIDDY; KELLER, 2005). A ordem global, ou seja, o equilíbrio da rede surge de diversas iterações locais. Os neurônios de saída competem entre si para serem ativados, de forma que apenas um neurônio seja considerado “vencedor” - “*The winner takes it all*” (O vencedor leva tudo).

No algoritmo são apresentados os padrões de entrada à rede, onde cada padrão tem uma região de atividade na grade. A localização e natureza de uma determinada região

variam de um padrão de entrada para outro. Assim sendo, todos os neurônios da rede devem ser expostos a um número suficiente de diferentes padrões de entrada, garantindo que o processo de auto-organização ocorra de forma apropriada. Os neurônios recebem estímulos (padrões de entrada) ao longo do processo competitivo de aprendizado e serão considerados “vencedores” os neurônios que mais se assemelharem ao padrão de entrada, sendo que para esta comparação são utilizadas medidas de distâncias entre vetores de pesos dos neurônios (normalmente utiliza-se a distância euclidiana) (HAYKIN, 2001) (ULTSCH, 1999) (HULLE, 2012) (KOHONEN, 2013).

O algoritmo de atualização de pesos podem ser resumidos em três etapas (KOHONEN, 1988) :

1. Encontrar o neurônio (*Best Matching Unit*) cujos vetor de pesos mais se assemelhem ao vetor de entrada;
2. Aumentar o acoplamento (*matching*) neste neurônio e em suas vizinhança ajustando seu vetor de pesos. Vizinhança é uma função de distância em relação a outros neurônios que preserve as propriedades topológicas do espaço de entrada.
3. Decrescer gradualmente o tamanho do raio de busca da vizinhança (ou área de influência) e a quantidade mudança dos pesos quando as iterações do aprendizado progredirem.

As áreas de pesquisa que mais utilizam esta classe de algoritmos são (KOHONEN, 1990) (MAREN; HARSTON; PAP, 2014):

- Mineração de dados em larga escala, em especial, visualização de dados estatísticos e coleções de objetos;
- Análise de processos, diagnósticos, monitoramento e controle;
- Aplicações biomédicas, incluindo métodos de diagnóstico e análise de dados de bioinformática;
- Análise de dados em comércio, indústria (como a de jogos), macro economia e setor financeiro.

## 2.2 ESOM

*Emergent Self-Organizing Maps*, abreviado como *ESOM*, é uma extensão do tradicional *SOM* com algumas modificações que lhe conferem características emergentes. A emergência é a capacidade de um sistema desenvolver estruturas de alto nível pela

cooperação de muitos processos elementares. Em sistemas auto-organizáveis, as estruturas evoluem de maneira espontânea. As estruturas emergentes oferecem uma descrição mais abstrata de um sistema complexo, constituído de indivíduos de baixo nível. Um exemplo popular de um fenômeno emergente é a chamada onda “la ôla” em um estádio de esportes. Um grande número de pessoas realiza a simples tarefa de levantar-se e acenar por um breve momento. Desta forma, uma onda é formada e segue no meio da multidão. A onda só é visível à distância uma vez que precisa ter a visão das ações de diversos indivíduos (ULTSCH; MÖRCHEN, 2005).

Transferir os princípios de auto-organização para análise de dados pode ser feito deixando pontos de dados multivariados organizando-se em grupos homogêneos. Uma ferramenta bem conhecida para esta tarefa, que incorpora os princípios mencionados é o *SOM*. O *SOM* iterativamente ajusta à distância estruturas num espaço dimensional elevado e dá uma projeção de baixa dimensionalidade que preserva a topologia do espaço de entrada tão bem quanto for possível.

Neste trabalho foi utilizado o algoritmo de Mapas de Kohonen com parâmetros que o faz ter características emergentes. É válido ressaltar que na evolução do conhecimento, muitos autores usavam mapas com um número reduzido de neurônios, o que dificultava a observação de características emergentes, que só se expressam mediante mapas maiores (LI; YU; CASTELLI, 1998) (FU et al., 2001) (PAMPALK; RAUBER; MERKL, 2002).

Enquanto na abordagem clássica de algoritmos *SOM* utilizam-se mapas planares com bordas, no *ESOM* busca-se evitar efeitos de borda para valorizar ainda mais características emergentes. Para tanto, uma abordagem bastante comum é usar um mapa de neurônios no qual as bordas superiores e inferiores são conectadas, assim como as laterais direita e esquerda, o que gera uma estrutura semelhante a um toróide.

### 2.2.1 *U-Matrix*

O *ESOM* utiliza a *U-Matrix* que é uma estrutura de distâncias locais para visualização e interpretação da estrutura emergente da rede neural *SOM*. Cada neurônio é apresentado como um valor de altura, criando uma paisagem em 3D como um mapa de topografia do espaço de dados de alta dimensão. A altura é calculada como a soma das distâncias a todos os vizinhos imediatos normalizados. Este valor será maior em áreas onde nenhum ou poucos pontos de dados residem, criando cadeias de montanhas para limites do agrupamento. A soma será pequena em áreas de alta densidade e assim, aglomerados são retratados como vales.

A *U-Matrix* indica uma topologia das relações de distância dos dados de entrada no espaço de dados. A *U-Matrix* apresenta as seguintes propriedades (ULTSCH, 2003b):

- A posição das projeções dos pontos de entrada de dados reflete a topologia do espaço

de entrada, que é herdada do algoritmo *SOM* subjacente;

- Vetores de pesos dos neurônios com grandes inversões de altura estão muito distantes de outros vetores no espaço de dados;
- Vetores de pesos dos neurônios com baixa altura são cercados por outros vetores no espaço de dados;
- Projeções dos pontos de dados de entrada são normalmente encontrados em depressões;
- Valores aberrantes ou anômalos no espaço de entrada são encontrados em “funís”, “picos” ou “montanhas”.
- “Serras” indicam possíveis limites de agrupamentos;
- “Vales” indicam possíveis centros de agrupamentos.

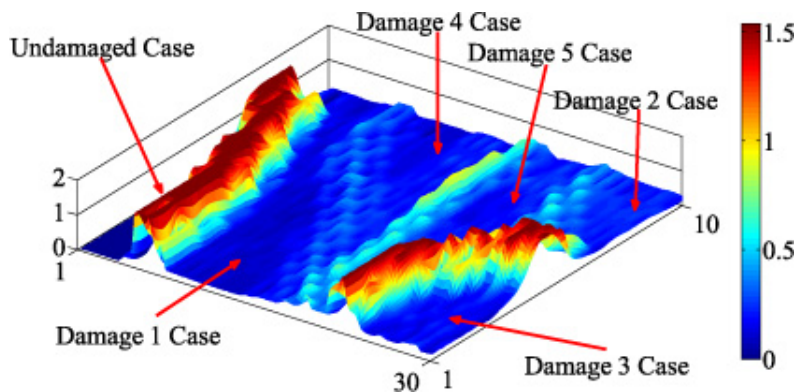


Figura 2 – Exemplo de visualização de *SOM* em *U-Matrix*

A Figura 2 mostra um exemplo da visualização *U-Matrix* do *SOM* na qual se observa a escala de cores para representar a topografia e a altura tridimensional (ANAYA et al., 2014). As cores quentes significam altos valores e as cores frias, o contrário. O autor apresenta 6 agrupamentos que são separados por cadeias de montanhas. Pode-se notar que existem separações que são muito fortes e outras mais leves, isto advém do quanto uma classe é diferente da outra.

## 2.2.2 *P-Matrix*

Enquanto os métodos baseados em distância comumente funcionam bem para agrupamentos claramente separados, podem ocorrer problemas com grupos que mudam lentamente as densidades e os que possuem sobreposição. Métodos baseados em densidade medem mais diretamente a densidade do espaço de dados amostrados nos vetores dos protótipos.



A *P-Matrix* exibe as medidas locais de densidade com a *Pareto Density Estimation (PDE)*, uma estimativa ótima da densidade *kernel* (ULTSCH; MÖRCHEN, 2005). O *PDE* apresenta a estimativa de densidade em um ponto  $x$  como sendo o número de pontos dentro da hipersfera (*Pareto Sphere*) ao redor de  $x$ . Foi demonstrado que o *PDE* leva a uma estimativa de densidade significativa e se ajusta perfeitamente no *ESOM* e *U-Matrix* (ULTSCH, 2003a).

Assim como a *U-Matrix*, a *P-Matrix* também usa o mapa *ESOM* como entrada. A *U-Matrix* revela a estrutura de distâncias (locais), enquanto a *P-Matrix* usa uma abordagem para exibir a estrutura de densidades de um grupo de dados de alta dimensionalidade. Os elementos da *P-Matrix* são os números de pontos dentro da esfera *pareto* (ULTSCH, 2003b).

As propriedades da *P-Matrix* são (ULTSCH, 2003b):

- A localização da projeção no mapa *ESOM* reflete a topologia do espaço de entrada, que é herdado do algoritmo *SOM*;
- Neurônios com elevado valor de altura estão situados em regiões de alta densidade do espaço de dados;
- Neurônios com baixa altura, estão em uma região de baixa densidade do espaço de dados;
- Valores Aberrantes ou Anômalos no espaço de entrada são encontrados em “funís”, “picos” ou “montanhas”;
- Valas ou poços indicam limites de agrupamentos;
- Planaltos indicam regiões de iguais densidades;

## 2.3 Dota 2

“*Dota is a combination of football - soccer for the Americans - and chess.*” (*Dota* é uma combinação de futebol e xadrez) [Su-Leo Liu] (*Star Commentator*) (VALVE, 2014c).

O *Dota 2* é um jogo do gênero arena de batalha *online* multijogador (*MOBA*) que derivou do tipo estratégia em tempo real (*Real-Time Strategy - RTS*) pela *VALVE Corporation* e lançado oficialmente em Julho de 2013 (GAO et al., 2013). Cada partida é composta por dois times de cinco jogadores. Cada jogador controla um avatar com a meta de derrotar o time adversário.

Para uma partida ser considerada válida, todos os jogadores devem se conectar à partida no seu início e nenhum jogador deve abandoná-la ou desconectá-la por mais de

cinco minutos. Caso a partida seja considerada inválida, nenhum jogador terá seu perfil atualizado com o desempenho obtido na partida, além disso, haverá uma penalidade extra para os jogadores desertores, fazendo-os ter baixa prioridade na fila de espera para iniciar novas partidas.

Os jogadores enfrentam-se conectados via internet ou rede local. No entanto, o início de uma partida exige que seja executada uma série de passos e podem ocorrer problemas de comunicação. A Figura 3 descreve o fluxo:

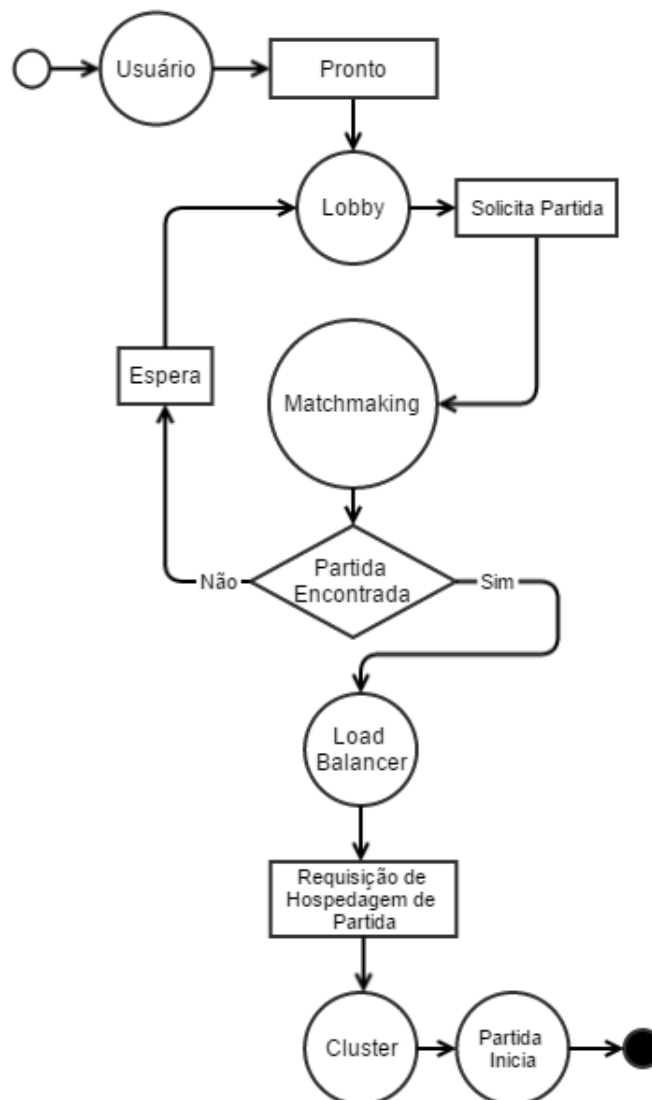


Figura 3 – Passos para início de partida no Dota 2

1. O usuário avisa ao servidor gerenciador da fila de partidas, denominado servidor de *Lobby*, que quer iniciar uma partida;
2. O servidor de *Matchmaking* analisa constantemente a fila presente no servidor de *Lobby* e seleciona os jogadores presentes no banco que comunicaram que estão

- querendo iniciar uma partida;
3. O servidor de *Lobby* solicita ao servidor *Load Balancer* a indicação de qual servidor do *Cluster* de servidores de partida pode hospedar a partida a ser iniciada e envia a resposta para que todos do grupo se conectem ao servidor que foi indicado;
  4. Uma vez conectados, cada um dos jogadores que já foram alocados a um time, escolhe um avatar que deseja controlar, ou como o jogo os chama, *Hero* (Herói), dentro de uma tabela contendo mais de 100 opções;
  5. A partida se inicia.

Cada herói possui características únicas que o diferenciam em categorias como papel ou função exercida na batalha, tipos de ataque e atributos primários, tais como inteligência, agilidade e força, além de atributos secundários, como pontos de vida, velocidade de movimentação, dentre outros. Os heróis possuem também um conjunto de habilidades que podem ser usadas em inimigos, ou em aliados, ou mesmo em elementos do cenário, conferindo assim uma mecânica de jogo única para cada jogador.

Um elemento central do jogo é a moeda financeira que pode ser usada para comprar itens ou diminuir o tempo de espera para renascer. Essa moeda no jogo é chamada de *gold*. Os itens que podem ser vendidos servem para obtenção de habilidades adicionais ou para incrementar atributos do herói. É possível ganhar *gold* no jogo das seguintes maneiras:

- Matando *creeps* ativos, que são inimigos de ambos os times e que defendem uma trilha que leva à base inimiga;
- Matando *creeps* neutros, que vivem nas florestas;
- Matando heróis inimigos ou participando e ajudando a matar heróis inimigos;
- Destruindo construções e torres de defesa dos inimigos;
- Ao longo do tempo, via bonificação regular dada pelo jogo.

Durante o curso do jogo, além de acumular *gold*, o jogador também acumula pontos de experiência. Ao chegar em determinados valores, o jogo permite que o herói incremente o seu nível, o que confere uma pontuação (*skill points*) que pode ser trocada para evoluir alguma de suas habilidades, tornando-a mais forte, intensa ou abrangente.

O jogo é baseado em partidas *multiplayer online*, com o objetivo de destruir a estrutura principal da base inimiga, que é o trono, localizado no centro de cada base, e, proteger a sua base. O jogo acaba quando o trono é destruído, concedendo a vitória a quem o destruiu. A Figura 4 exibe um exemplo de um mapa do terreno com diversos

elementos comuns do cenário (MILLENIUM, 2013). Para atravessar de uma base até a outra existem diversos caminhos que são protegidos por estruturas e ondas de *creeps* (personagens não controláveis). Porém, três dos caminhos se destacam e são chamados de *Top Lane*, *Middle Lane* e *Bottom Lane* que podem ser vistos na Figura 5. Cada *Lane* (caminho primário) possui uma característica específica que pode ser melhor aproveitada por determinados tipos de heróis. Além das *Lanes*, existem as florestas onde podem-se encontrar *creeps* neutros e é onde costumam ocorrer batalhas e ataques-surpresa chamados pela comunidade de *gank*.



Figura 4 – Mapa detalhado do Dota 2

### 2.3.1 Papéis dos Heróis

No Dota 2 existem papéis e responsabilidades que os jogadores podem exercer ao longo da partida e que são importantes para o decorrer do jogo. Esses papéis e responsabilidades são de domínio público e podem ser encontrados em fóruns da comunidade de usuários do Dota 2. Há um consenso de que um herói pode ter um ou mais papéis (rótulos) dentre os descritos:

1. **Carregador ou *Carry***: Herói que se torna mais importante com o decorrer do jogo. É fraco no início. Depende muito de itens e níveis. Torna-se bastante forte ao fim da partida. São geralmente os heróis que o time tem que proteger para que no



Figura 5 – Trilhas e suas dificuldades no mapa do Dota 2

fim se tornem fortes o bastante para assegurar ou “carregar” o time em direção à vitória;

- a) **Carregadores difíceis ou *Hard Carry***: São carregadores que são extremamente fracos no início.
  - b) **Semi carregadores ou *Semi Carry***: São carregadores que são menos fracos no início.
2. **Desativador ou *Disabler***: São os heróis que possuem uma ou mais habilidades que anulam ou debilitam de alguma maneira o oponente em uma batalha. Por exemplo, podem possuir habilidades que controlam a movimentação dos adversários temporariamente, assegurando que fiquem imóveis (*stun*) ou lentos (*slow*), impedindo os adversários de usarem habilidades (*silence*), enfraquecendo-os (*debuff*) o que os torna alvos mais fáceis. São bons em controle de multidão (*crowd control*);
  3. **Iniciador ou *Initiator***: São heróis que são bons em iniciar ou participar de batalhas. Geralmente possuem habilidades em área com uma grande amplitude de dano ou *stun*. Podem possuir uma habilidade suprema (*ultimate*) devastadora que demora muito para recarregar. Podem enfrentar de assalto (*gank*) heróis desprevidos e vencer. Uma mecânica comum é servir de isca para os oponentes o atacarem e tirar o foco dos demais, fazendo o time adversário sair de sua formação defensiva;
  4. **Suporte ou *Support***: Heróis que ajudam o time a partir do bom uso de suas habilidades, não precisando focar em acumular *gold* ou experiência. Não há necessidade inicial de concorrer com os aliados para acumular recursos, permitindo-lhes crescerem mais facilmente. São úteis durante toda a partida e em especial no começo.

Geralmente possuem habilidades de regeneração de status como vida ou mana e são mais úteis no início da partida;

- a) **Suporte na Trilha ou *Lane Support***: Heróis que protegem e dão suporte à aliados especificamente na *lane*, permitindo que os aliados acumulem recursos mais facilmente. Geralmente possuem habilidades de controle de inimigos, protegendo de ataques surpresas ou melhorando temporariamente alguma característica de um aliado;
5. **Resistente ou *Durable***: São os que possuem alguma habilidade ou característica que permite durar ou sobreviver por mais tempo em batalhas. Por exemplo, podem possuir mais pontos de vida, proteção, ou habilidades que os protejam em assaltos;
    - a) **Escapista ou *Escape***: Heróis que dentre os que sobrevivem mais em batalhas, possuem alguma forma de escapar da morte de forma rápida. Por exemplo, possuem alguma habilidade como correr, ou saltar alguma distância ou ficar invisível;
  6. **Bombardeador ou *Nuker***: Heróis que podem matar adversários facilmente no início da partida usando magias com alto dano, ou com baixo tempo de recarga ou *stun*. Geralmente tomam a trilha do meio para evoluir mais rapidamente ou andam acompanhados de aliados, que os protejam ou trazem inimigos para sua área de influência, permitindo aproveitar a oportunidade para atacar. Movimentam-se bastante pelo mapa e não costumam ficar em uma só trilha para arquitetar emboscadas (*ganks*);
  7. **Empurrador ou *Pusher***: Heróis que podem atacar e destruir torres e estruturas rapidamente. Possuem habilidades que permitam invocar servos para servirem de isca para os *creeps* inimigos, ou que causem dano em uma ampla área permitindo que os *creeps* aliados avancem mais facilmente;
  8. **Caçador ou *Jungler***: Heróis que são efetivos na eliminação de *creeps* neutros na floresta durante o início do jogo, o que permite que mais um aliado possua controle uma trilha (*lane*) sozinho. Geralmente se destacam por possuírem habilidades que lhes permitem sobreviver sozinhos, absorvendo ataques diretos, porém de baixo dano, ou com proteções ou escudos;

## 2.4 Considerações Finais

Observando a variedade de mecânicas de jogo e riqueza de detalhes, nota-se que o jogo *DOTA 2* não é simples, exige estratégia e reflexos rápidos. Se uma partida for formada de maneira desbalanceada, o nível de tensão e diversão será baixo, acarretando até mesmo a desistência de alguns jogadores. Disso surge a importância de se criar um mecanismo que permita a formação de partidas mais competitivas e, para tanto, foi utilizado o algoritmo emergente *ESOM* para extrair conhecimento da base e sugerir a formação de equipes balanceadas. Este capítulo apresentou o jogo Dota 2, que será o alvo desta pesquisa, bem como o ferramental teórico a ser usado para se entender melhor a formação de equipes. No próximo capítulo apresenta-se a abordagem proposta neste estudo.





Parte II

Proposta



## 3 Abordagem Proposta

Nesta seção apresenta-se o conjunto de passos que guiou a realização do estudo sobre a formação de equipes.

A realização deste trabalho utilizou como base a metodologia de Descoberta do Conhecimento em Base de Dados (DCBD) (FAYYAD *et al.*, 1996) (PIATESKI; FRAWLEY, 1991). Essa metodologia é dividida em cinco etapas. O diagrama da Figura 6 apresenta todas as etapas do processo de DCBD utilizado como base neste trabalho.

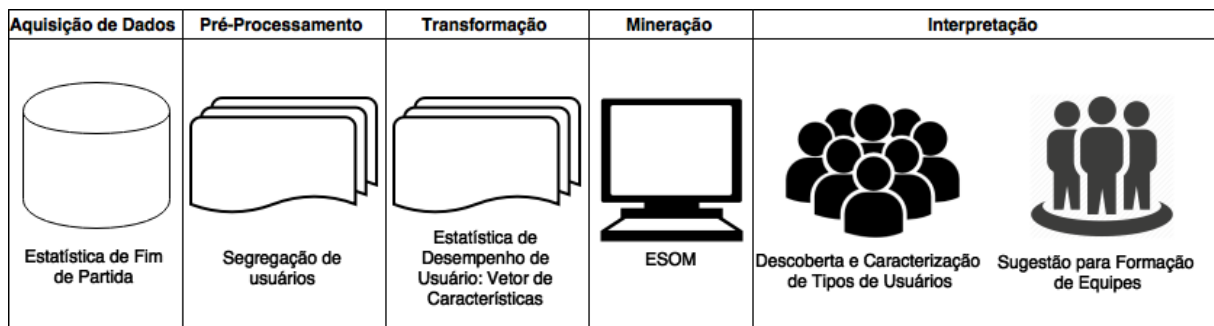


Figura 6 – Etapas do Processo de Descoberta de Conhecimento em Banco de Dados (DCBD) utilizado neste trabalho

### 3.1 Coleta de Dados

A *VALVE* possui um sistema de gravação (*replays*) que armazena dados ao final de uma partida. É possível assistir a uma partida concluída ou coletar dados estatísticos sobre sua execução. Isso inclui os dados sobre os jogadores que fizeram parte de uma partida, contendo informações que mostram sua eficiência no jogo, quantidade de pontos obtidos, morte a inimigos, dentre outros, além de informações gerais da partida, como por exemplo, quem ganhou, qual o tempo de duração, quem eram os jogadores dos times, etc.

Os dados acima mencionados são bastante úteis para uma avaliação dos jogadores bem como das equipes que participam dos jogos. A partir dessas informações é possível realizar diferentes avaliações e com isso propor mecanismos para ajudar a uma melhoria no jogo. Por conta disso, tais dados foram coletados para permitir a análise de diversos indicadores para auxiliar o entendimento sobre a formação das equipes no Dota 2. Os dados estatísticos são encontrados dentro de um arquivo de *replay*, no campo sumário (*overview*) (SKADISTATS, 2014).

### 3.1.1 Extração de Dados por Varredura - *Crawling*

Neste trabalho foi feita a coleta de dados de partidas de maneira simplificada, capturando apenas as suas estatísticas ao fim da partida, sem observar a forma como o jogador joga ao longo do tempo. A estrutura dos dados coletados é em forma de árvore, que pode ser esparsa. A Figura 7 mostra um sumário de fim de partida extraído que, de maneira geral, possui dois tipos de dados, os relacionados aos dados globais e os relacionados a cada jogador:

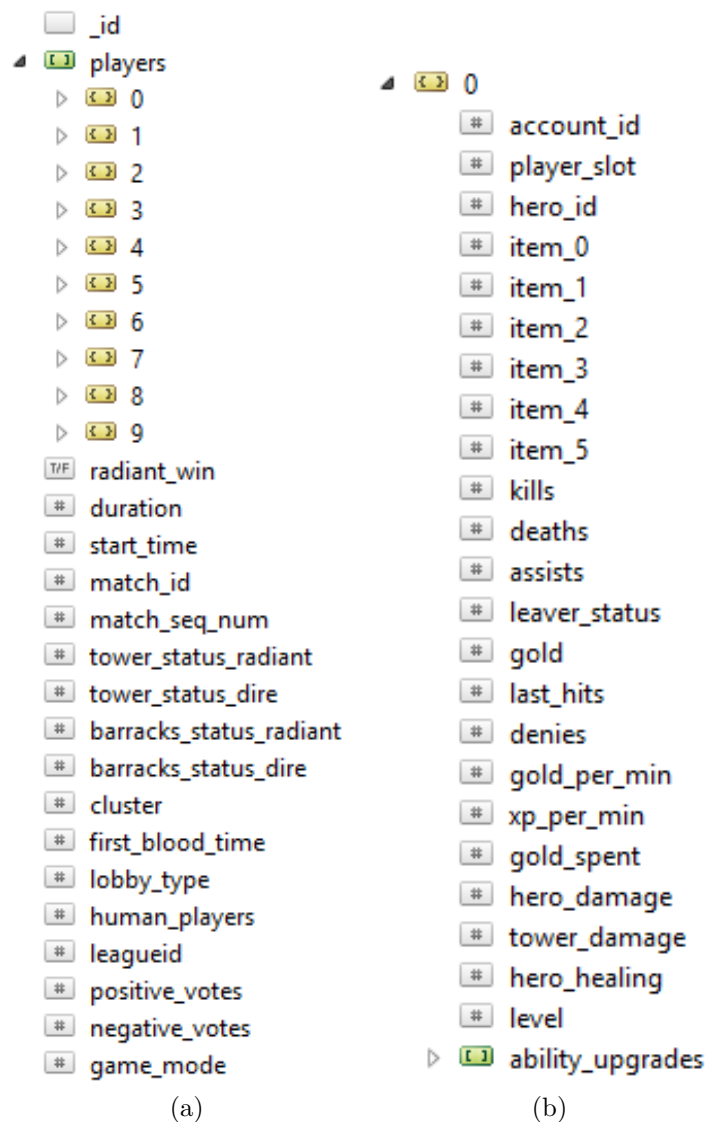


Figura 7 – Estrutura de dados típica de um Sumário de Partida: a) dados gerais da partida; b) dados específicos de um jogador.

- Dados globais da partida: identificação da partida, qual time ganhou, duração, quando começou, número de sequência de partida, status de torres, status de estruturas, tempo da primeira morte, localização de onde foi armazenada a partida (*cluster*), tipo de *lobby* (*matchmaking* público, treinamento, cooperativo com *bots*, *matchmaking*

entre times já formados, jogos solo, e desafios 1x1), número de jogadores humanos e tipo de jogo;

- Dados de cada jogador: identificação de usuário, *slot*, avatar, itens ao fim da partida, finalizações ou assassinatos de adversários, mortes, assistências em assassinatos, status de abandono (*leaver status*), *gold* gasto, *gold* ao fim da partida, *gold* ganho por minuto, *creeps* mortos, negação de *creeps*, experiência ganha, dano a heróis, cura a heróis, nível máximo, horário de evolução de cada nível;

Para esta captura de dados, a *STEAM* (plataforma de distribuição de software digitalmente via internet criada pela *VALVE*) oferece uma interface aberta para desenvolvedores, chamada de *WebAPI* <<http://steamcommunity.com/dev>>.

A metodologia de captura de dados consistiu em executar de forma repetida, os seguintes passos:

1. Capturar o histórico das últimas 100 partidas e os jogadores presentes em cada uma das partidas, para coletar dados dos últimos usuários ativos;
2. Para cada jogador presente na listagem anterior, capturar e armazenar os sumários de no máximo 100 últimas partidas nas quais ele participou, gerando assim a média móvel simples instantânea dos últimos estados de desempenho. Este limitante evita a coleta de dados de partidas muito antigas que não refletem o atual perfil de desempenho do jogador. Para calcular a média móvel, foram utilizadas as Equações 3.1 e 3.2;

$$MMS = \frac{A_M + A_{M-1} + \dots + A_{M-(n-1)}}{n} \quad (3.1)$$

Onde:

*MMS* : Média móvel simples

*A* : Vetor de dados

*M* : Identificador do último elemento do vetor

*n* : Quantidade de amostras usadas

$$MMS_{t+1} = \frac{\sum_{i=t+1-n}^t A_i}{n} \quad (3.2)$$

Onde:

$MMS_{t+1}$  : Média móvel simples no final do período  $t$

$A_i$  : Elemento  $i$  do Vetor de Dados  $A$

$n$  : Quantidade de amostras usadas

Como o desempenho de um usuário considerado neste estudo é o instantâneo, utilizou-se os conceitos de estatística básica para a análise de séries. Tomou-se a média móvel simples do desempenho dos jogadores para expressar o estado atual, observado na Equação 3.1. A média móvel é uma técnica usada para analisar dados em um intervalo de tempo permitindo fazer previsão do próximo valor, expresso pela Equação 3.2. O uso da média móvel simples fornece o valor médio do desempenho do usuários dentro de um determinado período (SAID; DICKEY, 1984) (MCCONVILLE; MILOSEVIC, 2014).

## 3.2 Pré-Processamento - Segregação por Dados de Usuário em Cada Partida

Uma vez que os dados de cada sumário de partida são estruturados de maneira hierárquica, contendo um vetor de dados de desempenho de cada um dos jogadores, é preciso que se extraia a informação do usuário individual para se fazer a estatística de desempenho na próxima etapa.

Após a captura de dados, foi realizado a segregação da informação por usuário, individualizando cada jogador, para assim facilitar as buscas e a geração de métricas. A Figura 8 mostra a estrutura de um descritor intermediário gerado a partir da segregação de cada jogador dentro de uma partida. Alguns atributos foram gerados artificialmente, como os que começam com o texto “total”, pois referem-se ao total daquele atributo para a equipe que participou da partida na qual a informação do usuário foi segregada.

No Dota 2 são considerados desistentes os usuários que se ausentam de uma partida por mais de 300 segundos (5 minutos). O campo *Leaver Status* contém a informação relacionada à desistência do jogo. Existem sete *status* de jogadores, diretamente relacionadas à sua participação durante o jogo:

0. Não desertor;
1. Sofreu uma desconexão que durou menos que 5 minutos;
2. Desconectado por mais de 5 minutos;
3. Usuário confirmou o abandono da partida;

4. Longe do teclado, não efetuando interações com o jogo (*Away From Keyboard - AFK*);
5. Nunca se conectou (partidas com duração menor que 5 minutos);
6. Nunca se conectou (partidas com duração de mais de 5 minutos).

## account_id	## start_time
## player_slot	## match_id
## hero_id	## match_seq_num
## item_0	## tower_status_radiant
## item_1	## tower_status_dire
## item_2	## barracks_status_radiant
## item_3	## barracks_status_dire
## item_4	## cluster
## item_5	## first_blood_time
## kills	## lobby_type
## deaths	## human_players
## assists	## leagueid
## leaver_status	## positive_votes
## gold	## negative_votes
## last_hits	## game_mode
## denies	## totalassists
## gold_per_min	## totalkills
## xp_per_min	## totaldeaths
## gold_spent	## total_last_hits
## hero_damage	## total_denies
## tower_damage	## total_healing
## hero_healing	## total_damage
## level	## total_tower
## ability_upgrades	## total_gold_spent
## radiant_win	## valid
## duration	

Figura 8 – Estrutura de dados do descritor de cada usuário para cada partida

Neste trabalho foram contabilizadas os 5 primeiros rótulos (de 0 a 4). Partidas encontradas com usuários com rótulos “Nunca se conectou” no campo *Leaver Status* foram desconsideradas uma vez que o Dota 2 atualizou o seu sistema de formação de partidas e estes casos não apareceram mais na base. As partidas com duração menor que 300 segundos (5 minutos) foram descartadas por apresentarem usuários desertores logo no início, uma vez que partidas tão curtas podem atrapalhar a medição, pois os seus valores não expressam o desempenho real de um jogador.

## 3.3 Transformação

### 3.3.1 Estatísticas de Desempenho dos Usuários

O sumário de fim de partida contém diversos atributos que são capturados ao longo da partida, contudo, alguns deles precisam ser trabalhados a fim de se tornarem métricas a serem usadas em algoritmos de mineração de dados. Nesta etapa foram criados descritores para cada usuário, a partir da média dos dados individuais de desempenho em cada partida. Foi criado um vetor de características contendo diversas métricas a serem descritas a seguir. A maioria das métricas foram oriundas dos dados brutos fornecidas pela *WebAPI* da *VALVE*, contudo, algumas tiveram de ser criadas artificialmente para abranger melhor as características de uma partida.

```
# account_id
# killsDeath
# killsPercent
# assistsDeaths
# assistsPercent
# deathsPercent
# denies
# deniesPercent
# duration
# first_blood_time
# gold_per_min
# gold_spent_percent
# gold_spent_per_minute
# hero_damage_per_min
# hero_damage_percent
# hero_healing_per_minute
# hero_healing_percent
# last_hits_per_min
# last_hits_percent
# tower_damage
# tower_damage_percent
# wins_percent
# xp_per_min
# leaver
# positive_votes
# negative_votes
# votes
```

Figura 9 – Vetor de características de um usuário

**Identificador de Usuário ou *AccountID*:** É usado para identificar qual é o jogador;

**Assassinatos/Mortes ou *KillsDeaths*:** É uma métrica composta que mostra a relação entre a quantidade de assassinatos sobre a quantidade de mortes de um usuário em uma partida. Indica o quanto um jogador realiza assassinatos e evita a morte;

**Assassinatos/Total de Assassinatos da Equipe ou *KillsPercent*:** É uma métrica composta que determina a relação entre a quantidade de assassinatos sobre a



quantidade total de assassinatos do time do usuário. Indica o desempenho de um dado jogador em matar em relação ao resto de seu time, exibindo o quanto ele exerce o papel de eliminador de adversários dentro do seu time;

**Assassinatos do Time por Assassinatos Totais ou *KillsDelta*:** É uma métrica composta que mostra o equilíbrio de assassinatos entre os times. Indica o quanto que a presença do usuário causa desequilíbrio no resultado da contagem final de assassinatos entre os times.

**Assistências/Mortes ou *AssistsDeaths*:** É uma métrica composta que representa a relação entre a quantidade de assistências (quem contribuiu no assassinato de um oponente) sobre a quantidade de mortes de um usuário em uma partida. Indica o desempenho em contribuir em batalhas sem morrer;

**Assistências/Total de Assistências da Equipe ou *AssistsPercent*:** É uma métrica composta que representa a relação entre a quantidade de assistências (foi quem contribuiu no assassinato de um oponente) sobre a quantidade total de assistências do time do usuário. Indica o desempenho em contribuir nas batalhas em relação o time, mostra o quanto é capaz de exercer o papel de assistente dentro do seu time;

**Negação ou *Denies*:** É uma métrica simples e absoluta de uma mecânica secundária bastante importante no início da partida e consiste na quantidade de vezes que um jogador impede o adversário de ganhar experiência e dinheiro ao evitar que um *creep* aliado seja morto pelas mãos do adversário. Este impedimento é feito quando um usuário mata um *creep* aliado antes que o adversário desfira o ultimo golpe. Indica o quanto o jogador é capaz de atrapalhar a evolução do adversário;

**Negação/Total de Negações da Equipe ou *DeniesPercent*:** É uma métrica composta que significa qual o percentual de *denies* o jogador fez em relação ao resto de seu time. Indica o quanto o jogador se destaca no trabalho de atrapalhar a evolução do adversário;

**Duração da Partida ou *Duration*:** É uma métrica simples e mostra o impacto da presença de um dado usuário na duração da partida. É a média da duração da partida nas quais o usuário participa. Um usuário pode ser defensivo e causar que as partidas demorem mais que o normal;

**Horário da Primeira Morte ou *FirstBloodTime*:** É uma métrica simples que aponta o tempo que se passou até a primeira morte. Geralmente em partidas desequilibradas um jogador experiente consegue obliterar um adversário inexperiente pouco tempo depois da partida iniciar;

**Ouro por Minuto ou *GoldPerMinute*:** É uma métrica simples extraída diretamente do sumário que contabiliza quanto de dinheiro um usuário consegue fazer por minuto

de partida. Indica o quanto que o usuário se destaca em eliminar *creeps*, adversários e estruturas;

**Ouro Gasto sobre o Total de Ouro Gasto Pelo Time ou *GoldSpentPercent*:** É uma métrica composta que contabiliza quanto de dinheiro um usuário consegue efetivamente converter em itens relativo aos demais de seu time. Indica a efetividade em gastar o dinheiro adquirido em relação aos seus aliados;

**Ouro Gasto por Minuto ou *GoldSpentPerMinute*:** É uma métrica composta que contabiliza quanto de dinheiro um usuário consegue efetivamente converter em itens, pois em cada morte perde-se dinheiro. Indica o quanto o usuário consegue gastar o dinheiro adquirido sem perder em mortes;

**Dano a Heróis por Minuto ou *HeroDamagePerMinute*:** É a métrica composta que indica a quantidade de dano efetivamente feito a um herói adversário. Indica o quanto o usuário é bom em atacar os adversários;

**Dano a Heróis pelo Total de Dano a Heróis do Time ou *HeroDamagePercent*:** É a métrica composta que mostra o desempenho do jogador em relação ao time no quesito quantidade de dano efetivamente feito a heróis adversários. Indica o quanto um usuário é bom em exercer o papel de atacar adversários em relação ao seu time;

**Cura por Minuto ou *HeroHealingPerMinute*:** É uma métrica composta que mostra a quantidade de pontos de vida restaurados em heróis aliados por minuto. Indica o quanto que um usuário é efetivo no suporte a aliados curando-os;

**Cura pelo Total de Cura do Time ou *HeroHealingPercent*:** Métrica composta que mostra a quantidade de cura feito por um usuário em relação aos demais do time. Indica o quanto o usuário se destaca dentro do time no papel de suporte a aliados curando-os;

**Finalizações de *Creeps* Por Minuto ou *LastHitsPerMinute*:** É uma métrica composta que mostra a quantidade de finalizações de *creeps* por minuto por usuário. Indica o quanto o usuário domina a mecânica de acertar o último ataque a um *creep*;

**Finalizações de *Creeps* pelo Total do Time ou *LastHitsPercent*:** É uma métrica composta que mostra a quantidade de finalizações de *creeps* por minuto por usuário em relação ao seu time. Indica o quanto o usuário se destaca dentro do time no papel de eliminar *creep*;

**Dano a Torres ou *TowerDamage*:** É uma métrica simples que conta a quantidade de dano feito pelo usuário a estruturas inimigas. Indica o quanto o usuário é efetivo no seu papel tático de avançar em direção à base inimiga;

**Percentual de Dano a Torres ou *TowerDamagePercent*:** É uma métrica composta que conta a quantidade de dano feito pelo usuário a estruturas inimigas em relação ao seu time. Indica o quanto o usuário é efetivo no seu papel tático de avançar em direção à base inimiga em relação aos demais do seu time;

**Taxa de Vitórias ou *WinsPercent*:** É uma métrica composta que conta a quantidade de vitórias do usuário. Indica o quanto o usuário participa de partidas no qual sai vitorioso;

**Dano a Torres pelo Total causado pelo Time ou *TowerDamagePercent*:** É uma métrica composta que mostra a quantidade de dano feito pelo usuário a estruturas inimigas em relação ao seu time. Indica o quanto o usuário se destaca dentro do time no papel tático de avançar;

**Experiência por Minuto ou *XpPerMinute*:** É uma métrica simples que conta a média da quantidade de experiência ganha por minuto de um usuário. Indica se o usuário consegue efetivamente evoluir dentro da partida.

**Votos Positivos ou *PositiveVotes*:** É uma métrica simples que conta a quantidade de votos positivos que a partida recebeu da plateia que assistia em tempo real. Indica se a comunidade está gostando do que está vendo;

**Votos Negativos ou *NegativeVotes*:** É uma métrica simples que conta a quantidade de votos negativos que a partida recebeu da platéia que assistia em tempo real. Indica se a comunidade não está gostando do que está vendo;

**Votos ou *Votes*:** É uma métrica derivada da soma do total de votos positivos e negativos que a partida recebeu da plateia que assistia em tempo real. Indica se a comunidade está ou não interessada na partida;

**Desertor ou *Leaver*:** É uma métrica que foi criada usando como base o estado de abandono de partida *LeaverStatus* descrito na Seção 3.2. É o resultado da contagem de quantas vezes o usuário atrapalhou uma partida por meio de abstinência ou desconexão dividido pelo total de partidas que participou. Das diferentes formas de apresentação da métrica original do Dota 2, considerou-se que o usuário atrapalhou a partida quando seu *LeaverStatus* apresenta os valores:

1. Desconexão rápida por menos de 5 minutos;
2. Desconexão por mais de 5 minutos;
3. Abandono proposital, o usuário informa ao jogo que abandonou a partida intencionalmente;

4. Abstinência de interação do usuário com o jogo por mais de 5 minutos. O usuário para de interagir com o jogo, por exemplo alternando para outra aplicação enquanto a partida não terminou.

Um usuário que desconecta por menos de 5 minutos é visto como um problema pontual. Contudo, todos os demais casos atrapalham significativamente a partida, caso haja algum usuário que siga pelo menos uma das condições descritas, o mesmo será punido com um maior tempo de espera na formação de partidas posteriores e a partida corrente se tornará imediatamente *safe-to-leave*. Este é o termo dado a uma partida na qual os demais usuários podem sair sem serem penalizados, uma vez que a partida será injusta por ter pelo menos um desfalque.

A métrica *Leaver* considera o caso 0 do *LeaverStatus* como o único que não atrapalha o andamento saudável da partida. O componente do descritor *Leaver* é uma estatística que representa a porcentagem de vezes que o usuário atrapalhou uma partida devido a abandonos ou falhas de conexão, ou seja, é a soma das estatísticas de quantas vezes o usuário foi rotulado como classes de 1 a 6 sobre o total de partidas que jogou.

Algumas informações do sumário de fim de partida não foram consideradas para a geração de métricas pelo fato de serem referentes ao estado momentâneo e transiente do final da partida, não representando, portanto, graus evolutivos ou de desempenho. Desse modo, a métrica *gold* que indica a quantidade de dinheiro ao finalizar a partida e também todas as relativas a *item* que se referem aos itens presentes no fim da partida foram desprezadas.

### 3.3.2 Eliminação de Atributos

Devido ao alto custo computacional dos algoritmos de mineração de dados e para melhoria da qualidade de resultados, foi realizada uma análise de correlação entre os vetores com o intuito de remover atributos desnecessários. A correlação é uma técnica estatística que pode mostrar quando e como dois pares de atributos estão relacionados. Construiu-se uma tabela com os valores de correlação entre todos os atributos e os atributos mais correlatos foram eliminados ou agregados em um só. A análise foi feita usando o coeficientes de correlação de *Pearson* (LAWRENCE; LIN, 1989) (ADLER; PARMRYD, 2010) e também o de *Spearman* (LYERLY, 1952) (CROUX; DEHON, 2010).

O coeficiente de correlação é um número entre  $-1$  e  $+1$  que mede o grau de associação entre dois atributos, por exemplo,  $X$  e  $Y$ . Um valor positivo de correlação implica uma associação positiva. Neste caso, valores elevados em  $X$  tendem a ser associados com valores elevados em  $Y$  e valores baixos de  $X$  tendem a serem associados com valores baixos em  $Y$ . Um valor negativo de correlação implica uma associação inversa, neste caso

valores elevados de  $X$  tendem a se associarem com valores baixos em  $Y$  e vice-versa.

$$Pearson = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_i^2 - (\sum y_i)^2]}} \quad (3.3)$$

$$Spearman = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (3.4)$$

Estes operadores são usados para criar a matriz de correlação para todos os atributos. Como o objetivo é eliminar atributos semelhantes, toma-se o valor absoluto dos elementos da tabela e remove-se aqueles com um valor acima de um determinado limite.

## 3.4 Mineração

Nesta etapa deve ser realizada a análise dos dados de estatística de jogadores visando gerar informações úteis para se entender a formação de equipes.

Durante o processo de estudo de quais métodos de aprendizado não-supervisionados seriam mais adequados para o problema em questão, foram experimentados *k-means*, clusterização hierárquica e redes neurais. Redes neurais apresentaram resultados mais interessantes, sendo capazes de segregar e identificar uma maior quantidade de classes dentre os jogadores. Dentre as redes utilizadas, escolheu-se *ESOM* por ser usado pela comunidade de maneira bem sucedida para caracterizar jogadores (DRACHEN; CANOSSA; YANNAKAKIS, 2009) (GRUNZ; MEMMERT; PERL, 2012).

A clusterização hierárquica e *k-means* sugeriram que haveria cerca de 2 ou 3 classes, um valor considerado baixo para o *Dota 2* e considerado não representativo para este trabalho, uma vez que não ressaltava características singulares dos grupos. Assim, tentou-se explorar mais técnicas e foi possível descobrir que o uso de outros algoritmos geravam resultados mais interessantes do ponto de vista de agrupamento de jogadores.

### 3.4.1 Treinamento e Classificação com *ESOM*

De posse de uma base dados com descritores de desempenho de jogadores, foi executado o algoritmo *ESOM* (ULTSCH; MÖRCHEN, 2005) com alguns valores modificados em relação aos sugeridos na ferramenta “*Databionics ESOM Tools*” produzida pelos autores. Dentre eles destacam-se:

- A topografia da rede neural: Toróide como na Figura 10a. Esta escolha foi tomada para evitar efeitos de borda que poderiam levar a uma classificação errada dos neurônios. Um toróide possui continuidade vertical e horizontal, facilitando o entendimento de

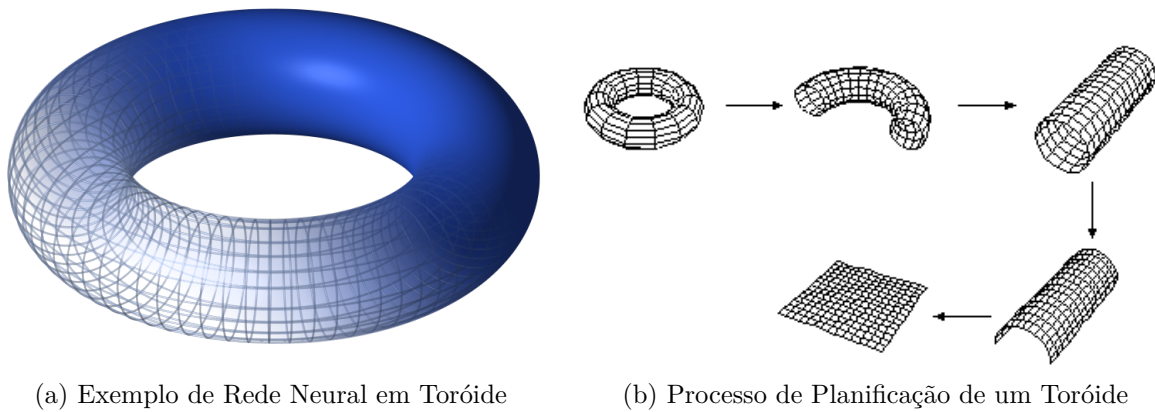


Figura 10 – Topologia Escolhida para o *SOM*

classes correlatas. A transformação em mapa pode ser observado pela Figura 10b e ajuda a entender como essa topografia auxilia a observação dos resultados;

- A largura e altura da rede são respectivamente de 150 x 123 neurônios. Esses valores correspondem a multiplicação dos valores sugeridos pela referência, que são 100 x 82, por uma vez e meia. Este incremento se justifica uma vez que se busca observar o caráter emergente da rede que só é obtido com uma maior quantidade de neurônios;
- Foram usadas 300 épocas de treinamento, valor maior que as 20 sugeridas pela referência, com a finalidade de dar mais oportunidade da rede se organizar melhor.
- Raio de busca inicial foi tomado como 15, menor que o sugerido 24, escolha feita para realizar mais buscas locais e compensar o aumento da quantidade de épocas. Quanto menor o raio inicial, as buscas tornam-se mais locais.
- O algoritmo de distâncias e raios de busca foi o comum da literatura: distância euclidiana;
- Todos os decaimentos e resfriamentos foram os mesmos da referência, lineares em função da iteração da época de treinamento;
- A função de pesos de vizinhança adotada foi o mesmo da referência, a gaussiana;
- O algoritmo de treinamento de neurônios utilizado foi o mesmo adotado pela literatura de referência, o *Online Training*.

Vale ressaltar que para chegar aos parâmetros descritos foram executados sucessivos treinamentos utilizando o *ESOM*, até obter um mapa que apresentasse características satisfatórias como a mesma topologia em dois testes consecutivos (convergência), granularidade e resolução suficiente para a classificação dos neurônios.

### 3.4.2 Classificação de Usuários com *ESOM*

A partir do mapa *SOM* treinado, foi feito o agrupamento de neurônios utilizando os conceitos de *U-Matrix* e *P-Matrix* descritos na Seção 2.2. Cada agrupamento de neurônios classificou um determinado tipo de usuário. Em seguida todos os vetores de características dos usuários foram categorizados em classes usando o classificador treinado.

Para facilitar a etapa de agrupamento de neurônios, foram realizadas modificações no código fonte da ferramenta “*Databionics ESOM Tools*”. As modificações foram similares às empregadas em processamento de imagens denominadas limiarização e detecção de rampa. Objetivou-se encontrar proceduralmente as regiões da *U-Matrix* e *P-Matrix* que são cercadas por cadeias de montanhas e segmentá-las. A característica principal destas regiões é a separação por um gradiente, ou como em paralelo à imagens, são regiões nas quais há uma grande mudança de cor.

## 3.5 Interpretação

### 3.5.1 Rotulação

Após o agrupamento de neurônios usando o *ESOM*, foi realizada a classificação da base de dados. Para cada classe de usuários, fez-se a média das estatísticas do desempenho do agrupamento gerando um centróide. Ao se comparar o desempenho médio de cada centróide com as demais, foram identificados os padrões de comportamento comuns aos indivíduos dos grupos. Cada agrupamento recebeu como rótulo um nome que sintetiza as suas características.

### 3.5.2 Identificar a composição de equipes

Visando propor um estudo para formação de equipes, é preciso entender suas composições e como cada uma impacta em relação a parâmetros que avaliem o quanto uma partida é divertida ou que causa um maior grau de tensão.

Por conta do exposto, foi feito um estudo para analisar métricas que poderiam indicar uma melhor disputa. No caso do Dota 2, foram utilizadas duas métricas que se mostraram bastante adequadas para a indicação de uma partida disputada: a quantidade de assassinatos realizados por um time e a duração da partida. Esse é um ponto importante nessa etapa da avaliação: descobrir as métricas que mais estejam associadas a partidas disputadas.

A partida é mais divertida quanto mais ela for equilibrada. O equilíbrio entre times implica em empate de desempenho médio dos times. Para o caso do futebol, uma boa métrica é o placar. No entanto, nem sempre o placar indica realmente como foi a partida.

Existem jogos que um time ganha, mas a posse de bola e quantidade de chances de gol do time perdedor é muito maior. Isso normalmente é considerada uma exceção. Mas acontece.

No caso do Dota, a duração da partida é uma fonte de informação sobre sua disputa. Uma partida que rapidamente é concluída indica que houve um desbalanceamento entre equipes. Da mesma forma, uma partida em que o número de assassinatos é muito grande em relação a outra equipe, normalmente impacta no vencedor. No entanto, foram encontradas partidas em que o vencedor assassinou menos que o perdedor, de forma similar ao que foi descrito sobre o futebol.

Uma vez que tenham sido definidos as métricas para se avaliar boas partidas, é necessário analisar a formação das equipes nesses casos. Ao se analisar essa questão no Dota 2, notou-se que quanto mais equilibrados eram os times, em termos das classes de jogadores encontrados na mineração, melhor era a partida, no que se refere ao tempo de duração e assassinatos por time.

### 3.6 Considerações Finais

Neste capítulo foi apresentada a metodologia empregada para se estudar a formação de equipes. A metodologia foi baseada no processo de descoberta de conhecimento em banco de dados e foi executada em uma série de etapas. Essas etapas incluíram a coleta de dados, pré-processamento, transformação e a efetiva mineração, usando algoritmos de inteligência artificial.

A partir dos resultados da mineração será possível identificar agrupamentos de usuários mais significativos. Após a análise de um especialista, os grupos serão rotulados, visando identificar as principais características ressaltadas nos agrupamentos.

De posse dos grupos de usuários identificados, serão realizadas análises das partidas, especialmente aquelas com bons indicadores de competitividade, para tentar se inferir como deveria ser a formação de times para maximizar o engajamento e diversão. Para isso, será necessário definir que indicadores poderão ser usados para se avaliar uma partida disputada.

A análise das partidas disputadas, especialmente abordando a composição das equipes, baseadas nos grupos aqui identificados, poderão ressaltar critérios para formação de equipes que maximizem a disputa entre times.

O próximo capítulo demonstra, passo a passo, a execução da abordagem descrita neste capítulo, bem como as conclusões sobre a formação de equipes no Dota 2, além de uma tentativa de generalização dessa abordagem, visando torná-la independente de jogo e aplicável em qualquer disputa envolvendo times.



Parte III

Parte Final



## 4 Resultados e Discussão

### 4.1 Coleta e Tratamento de Dados

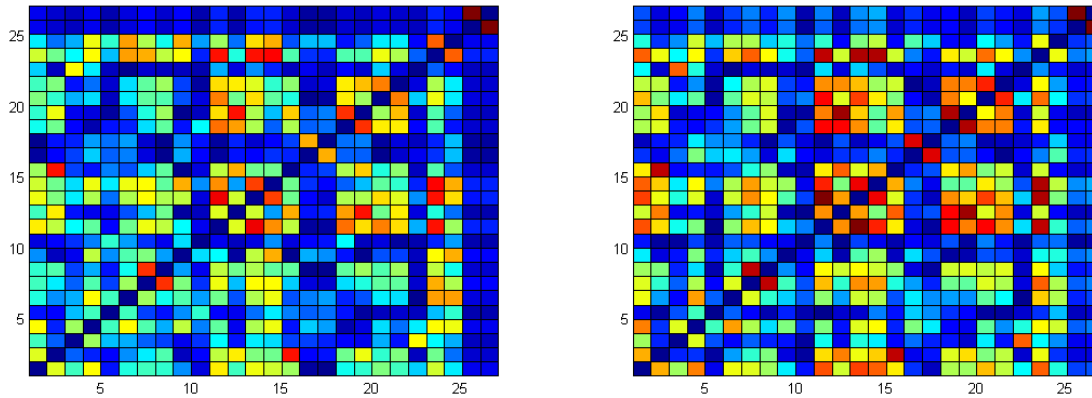
Para a realização do presente estudo foram coletadas informações de cerca de 11 milhões de partidas do Dota 2, entre março e maio de 2014, na versão 6.80, alcançando aproximadamente 140GB de dados de sumários de fim de partida. Foram descartadas as partidas que possuíam uma quantidade diferente de 10 usuários presentes, medida tomada para remover as partidas que já iniciavam desfalcadas. Da mesma forma, foram descartadas as partidas com duração menor que 5 minutos, uma vez que esse é o tempo limite utilizado pelo jogo para caracterizar uma deserção além de ser uma duração improvável para um jogo envolvendo uma disputa real entre dois times.

O pré-processamento consistiu em fazer a estatística de desempenho de jogador apenas para os que estiveram presentes em pelo menos 10 partidas, dentre as coletadas, considerando os critérios anteriormente mencionados. Tal etapa permitiu calcular a média móvel (SAID; DICKEY, 1984) (MCCONVILLE; MILOSEVIC, 2014) simples do desempenho do jogador e minimizar desvios estatísticos que não representem o seu desempenho frequente.

Em seguida foi criado o vetor de característica para cada usuário. Ele é composto pelas médias de componentes de cada métrica aritmética dentro das partidas que cada usuário participou. Este processo de seleção e tratamento resultou em cerca de 125 mil descritores de usuários, ou vetores de características de desempenho de jogadores distintos.

Realizou-se a análise de correlação entre os atributos utilizando os coeficientes de *Pearson* e de *Spearman*. Removeu-se aqueles que possuíam uma grande similaridade, que são representados pelos valores com coeficiente de correlação acima de 0.8 em absoluto. Os atributos removidos foram: *PositiveVotes*, *NegativeVotes*, *LastHitsPercent*, *HeroDamagePercent*, *GoldSpentPerMinute* e *DeniesPercent*.

A Figura 11 representa uma matriz quadrada de 27 linhas. Cada linha está associada a um dos atributos utilizados na caracterização de um jogador. Os atributos mostram a representação das tabelas das matrizes de correlação variando em escala de cores. A escala de cores varia de vermelho (*RGB* 255, 0, 0) que representa valores de correlação próximos a 0%, passa pela cor verde (*RGB* 0, 255, 0) para valores próximos a 50%, até o azul (*RGB* 0, 0, 255) para os próximos a 100%. As demais cores representam valores intermediários dentro da escala. Os valores da tabela são representados no Apêndice A. Em sequência, os atributos são: *KillsDeath*, *KillsPercent*, *KillsBalance*, *AssistsDeaths*, *AssistsPercent*, *DeathsPercent*, *Denies*, *DeniesPercent*, *Duration*, *FirstBloodTime*, *GoldPerMin*, *GoldSpentPercent*,



(a) Representação da Matriz Coeficientes de Correlação de Pearson (b) Representação da Matriz Coeficientes de Correlação de Spearman

Figura 11 – Matrizes de coeficientes de correlação

*GoldSpentPerMinute, HeroDamagePerMin, HeroDamagePercent, HeroHealingPerMin, HeroHealingPercent, LastHitsPerMin, LastHitsPercent, TowerDamage, TowerDamagePercent, WinsPercent, XpPerMin, Leaver, Votes, PositiveVotes e NegativeVotes.*

Para exemplificar a remoção dos atributos, destaca-se o exemplo dos dois últimos atributos: *PositiveVotes* e *NegativeVotes*, ambos apresentam alta taxa de correlação com os demais atributos, observada pela grande quantidade de elementos na cor de azul na Figura 11 alcançando extremos que ultrapassam o limiar estabelecido. No lugar de ambos, usou-se a soma dos votos (*Votes*) pois a quantidade de positivos era bastante correlata com a de negativos além de que poucas partidas recebiam votos. Logo, a somatória de votos positivos e negativos é uma métrica que indica se a partida é popular entre os telespectadores ou não. Outro exemplo é a remoção de algumas métricas compostas que relacionam duas informações tais como *HeroDamagePercent, GoldSpentPerMinute e DeniesPercent*, que possuem correlação com outros componentes. Ao se detectar a existência de dois componentes correlacionados, o critério de escolha para a remoção de um em detrimento do outro recaiu sobre o que apresentou maior correlação com os demais.

## 4.2 Treinamento e Classificação

Os valores de ajuste do algoritmo *ESOM* descritos na Seção 3.4 foram os que permitiram que a estrutura de organização de neurônios conseguisse convergir e gerar sempre a mesma topografia de localização dos neurônios, criando um mapa estável no qual a vizinhança dos neurônios é a mais ajustada que o algoritmo encontrou. Este critério é cientificamente aceito na literatura (ULTSCH; MÖRCHEN, 2005) (DRACHEN; CANOSSA; YANNAKAKIS, 2009) para verificar se os parâmetros escolhidos estão adequados. Foram testados diversos parâmetros, para observar o quanto a rede se modificava em relação

a cada um deles, até que os mapas gerados consecutivamente apresentassem a mesma topologia.

Neste sentido, o parâmetro de tamanho da rede foi testado com diferentes valores visando encontrar o que gerasse mapas com a mesma topografia e com granularidade suficiente para segmentar os agrupamentos e descobriu-se que este valor teria de ser superior ao da referência. Além disso, O parâmetro épocas de treinamento também foi modificado para dar mais oportunidade para os agrupamentos que se movem lentamente poderem se ajustar melhor.

A Figura 12 mostra o resultado da classificação do mapa de neurônios, no qual se pode notar uma riqueza de detalhes e contornos para separar os agrupamentos. A seguir serão abordados com mais detalhes as etapas que levaram até este resultado e a interpretação de cada classe.

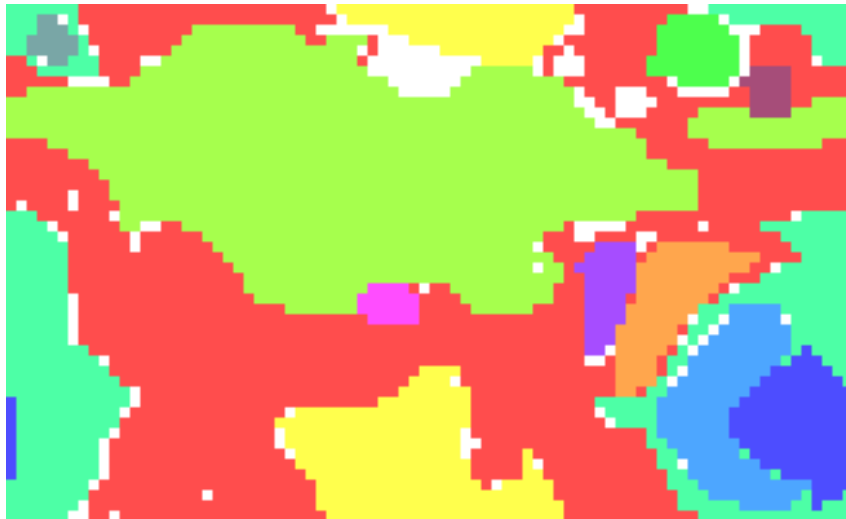


Figura 12 – Mapa Final de Neurônios Classificados

### 4.2.1 Treinamento

A execução do algoritmo de treinamento do *ESOM* resultou em um mapa de neurônios que podem ser visualizado pelas representações em *U-Matrix* e *P-Matrix* a seguir. Eles mostram a topologia para cada neurônio exibido como um valor de altura 3D de uma paisagem de alta dimensionalidade reduzida a um espaço bidimensional. A descrição do emprego da metodologia *ESOM* para agrupar e classificar os neurônios é evidenciada na Seção (4.2.2).

A estrutura da rede neural adotada foi a toroidal, conforme descrito na Seção 3.4.1. Dessa maneira, o *SOM* possui as características que podem ser observadas nos mapas em *P-Matrix* e *U-Matrix* tais como a continuidade na topologia nos sentidos verticais e horizontais.

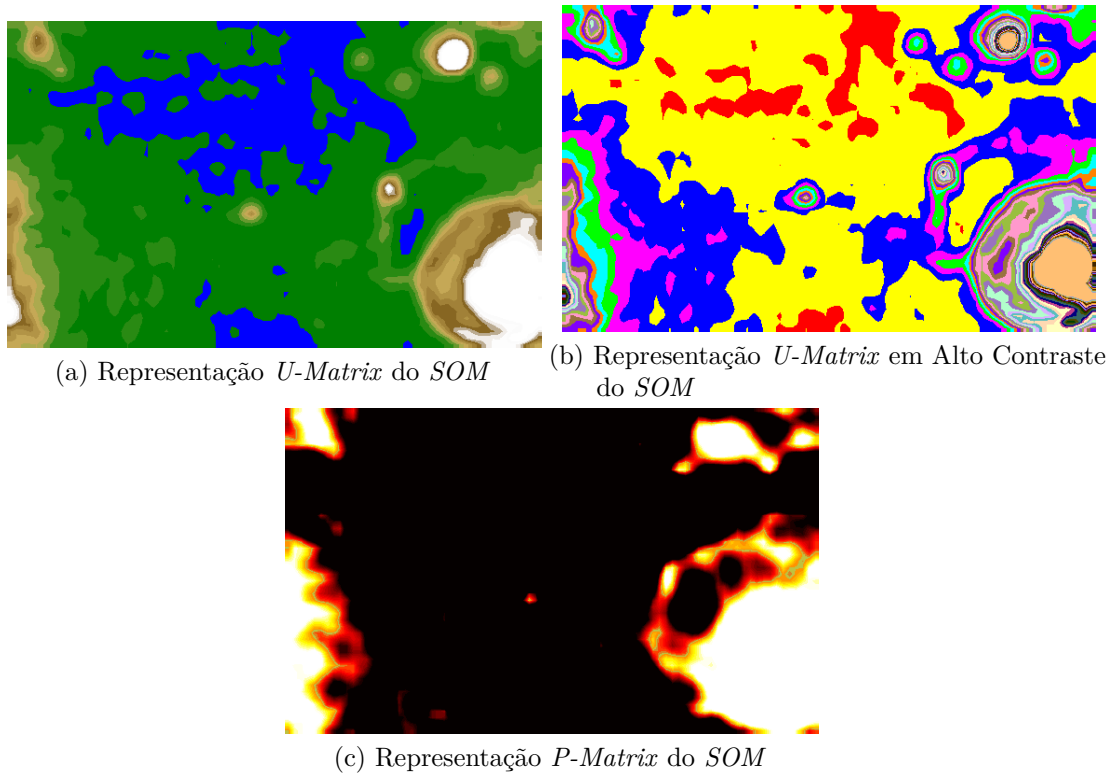


Figura 13 – Representações *U-Matrix* e *P-Matrix* do *SOM*

A Figura 13a exibe a representação dos neurônios treinados do *SOM* como uma matriz de distâncias *U-Matrix* e para melhorar a visualização e entendimento, a Figura 13b exibe-a em alto contraste. A Figura 13c exibe o mesmo mapa *SOM* porém, seguindo os conceitos de densidade pareto (*P-Matrix*). Uma vez que ambas são geradas usando o mesmo *SOM*, percebe-se que existe uma grande relação entre as duas formas de se observar o mapa. Isso sugere que tanto a forma de visualização do mapa em *P-Matrix* e *U-Matrix* possuem uma topologia bastante semelhante. Ao se comparar as duas, pode-se notar que existem regiões e estruturas semelhantes, como por exemplo a porção do canto inferior direito das duas representações e uma região bem definida, cercada por montanhas próximo à região descrita. A limiarização exibida na Figura 13b permite uma observação mais facilitada dos contornos das curvas de nível da topologia.

Nas figuras a seguir, são apresentadas as relações entre cada um dos componentes do vetor de características, evidenciando como os neurônios se agrupam em função das métricas para alcançar a organização final do mapa *SOM*. O apêndice B contém a compilação de todas as influências individuais. A escala de cores destas imagens varia de vermelho para azul. Quanto mais vermelho for um neurônio, mais positivo é o valor da influência, quanto mais azul, mais negativo é o peso do neurônio para esse componente. A imagem é importante para detectar as relações entre os componentes e, se a informação representada por um componente possui ou não associações com os demais.

Uma quádrupla associação pode ser observada na Figura 14 e diz respeito à

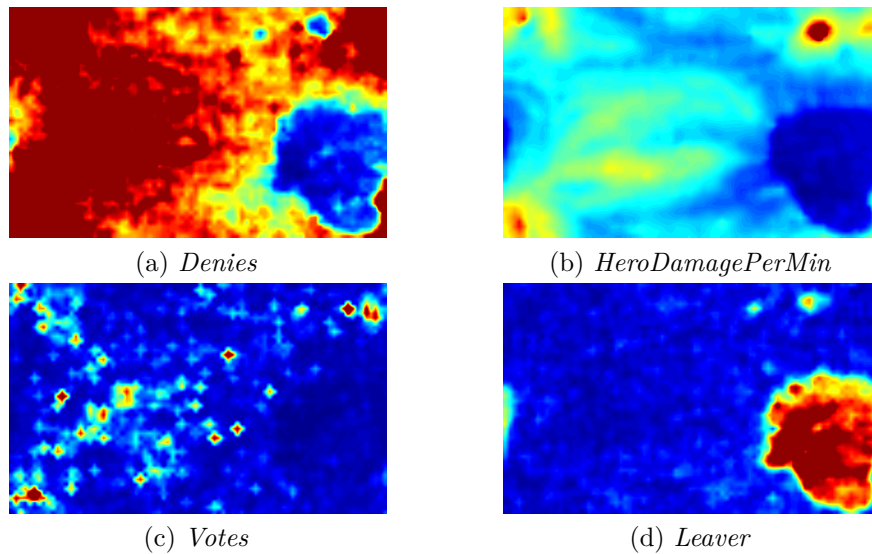


Figura 14 – Relação entre *Denies*, *HeroDamagePerMin*, *Votes* e *Leaver*

localização da região colorida de azul em *Denies*, representando a negação de *creeps*, com o que está em azul na *HeroDamagePerMinute*, que representa o dano feito a heróis por minuto, com o que está em azul escuro na *Votes*, que representa o quanto que uma partida é assistida e, com o que está colorido de vermelho na *Leaver*, que representa a taxa de abandono. A partir dessa associação pode-se afirmar que usuários que costumam abandonar as partidas, são os mesmos que possuem baixo rendimento em negação de *creeps*, são os que conseguem causar menos danos aos inimigos durante uma partida e são os que participam de partidas nas quais a comunidade não gosta de assistir. Observando ainda a *Votes* e *HeroDamagePerMinute* representado pelos valores em vermelho, percebe-se que a comunidade prefere assistir jogos com a presença de usuários com alta taxa de dano a heróis. Nota-se que a maioria dos outros componentes se organizam respeitando o formato do contorno da *Leaver* e, portanto, pode-se inferir que é uma métrica de grande importância.

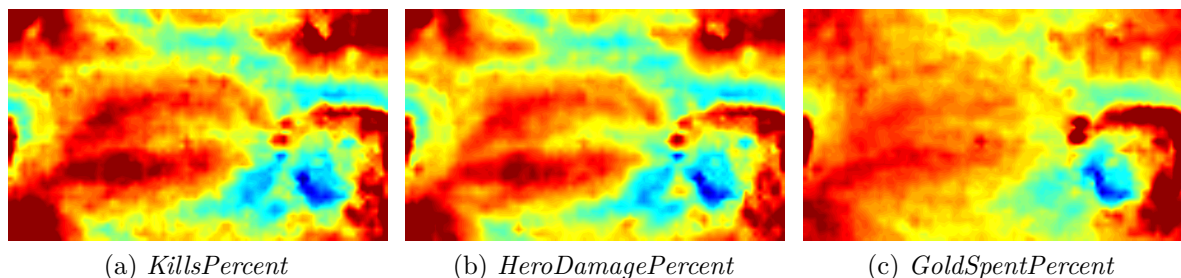


Figura 15 – Relação entre *Kills*, *HeroDamagePercent* e *GoldSpentPercent*

A Figura 15 apresenta outra associação importante entre três componentes, *KillsPercent*, *HeroDamagePercent* e *GoldSpentPercent*. O formato geral da topografia do mapa na *KillsPercent* representa o desempenho da finalização de inimigos em relação ao seu time, e *HeroDamagePercent* mostra o quanto um usuário causa dano a inimigos, enquanto

a *GoldSpentPercent* indica o quanto um usuário conseguiu gastar em relação aos demais do seu time. Essa associação indica que o usuário que costuma ser efetivo na finalização de inimigos, também é o usuário que mais gasta *gold* do seu time, e é o que mais causa dano em inimigos.

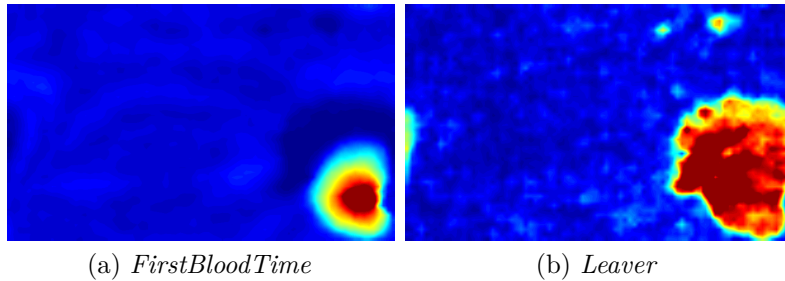


Figura 16 – Relação entre *FirstBloodTime* e *Leaver*

Ao observar na Figura 16 o componente *FirsBloodTime* que representa o momento do primeiro assassinato e a *Leaver* que mostra a taxa de abandono, percebe-se que usualmente os usuários desertores, são os mesmos que participam de partidas nas quais ocorrem assassinatos logo no começo. Isso indica que um usuário que é assassinado justamente no início da partida, é o mesmo que tende a abandoná-la.

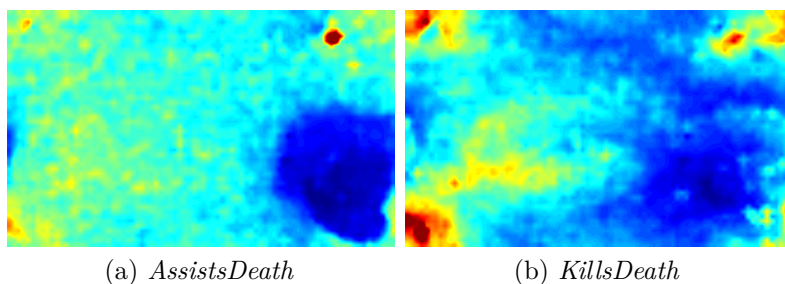


Figura 17 – Relação entre *AssistsDeaths* e *KillsDeaths*

Existe uma semelhança na Figura 17 na região específica em vermelho do componente *AssistsDeaths*, e o componente *KillsDeaths*. O *AssistsDeaths* representa a relação entre a quantidade de contribuições ou assistências em assassinatos de inimigos sobre o número de vezes que o usuário morreu, enquanto que o componente *KillsDeaths*, representa a taxa de assassinatos de inimigos sobre a quantidade de vezes que o usuário morreu. A partir desta associação, pode-se afirmar que o mesmo usuário que costuma ter um alto desempenho em assistências, também demonstra ter um desempenho semelhante em assassinatos. Contudo, os demais dados divergem ou não possuem associação clara, evidenciando que são jogadores que diferem uns dos outros por focar no desempenho maior em assassinatos, logo devem ser classificados de maneira diferente.

Ao se comparar as topografias dos mapas da Figura 18, percebe-se que os componentes *GoldPerMin* e *LastHitsPerMin* são bastante semelhantes em seu formato, mudando apenas a escala das cores. O *GoldPerMin* representa a taxa de acúmulo de *gold* por minuto,



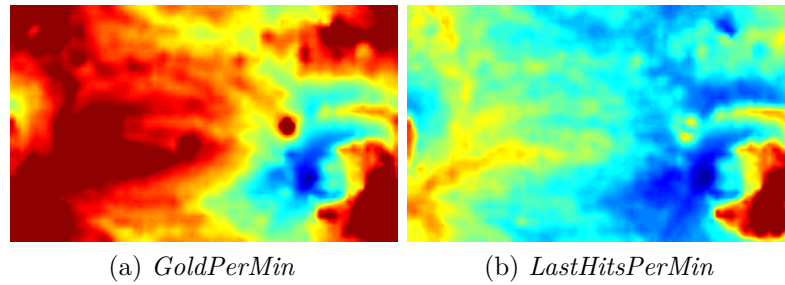


Figura 18 – Relação entre *GoldPerMin* e *LastHitsPerMin*

que pode ser proveniente de assassinatos, finalizações de *creeps* ou outros meios, enquanto que o *LastHitsPerMin*, representa apenas a quantidade de finalizações de *creeps* por minuto. Isso indica que os usuários que possuem um determinado rendimento em um dos dois componentes do vetor de características, também possuem um rendimento semelhante na taxa do outro componente. Portanto, ambos os componentes mapeiam o mesmo padrão semelhante de ação ou decisão de usuário e ambos se complementam.

#### 4.2.2 Agrupamento e Classificação

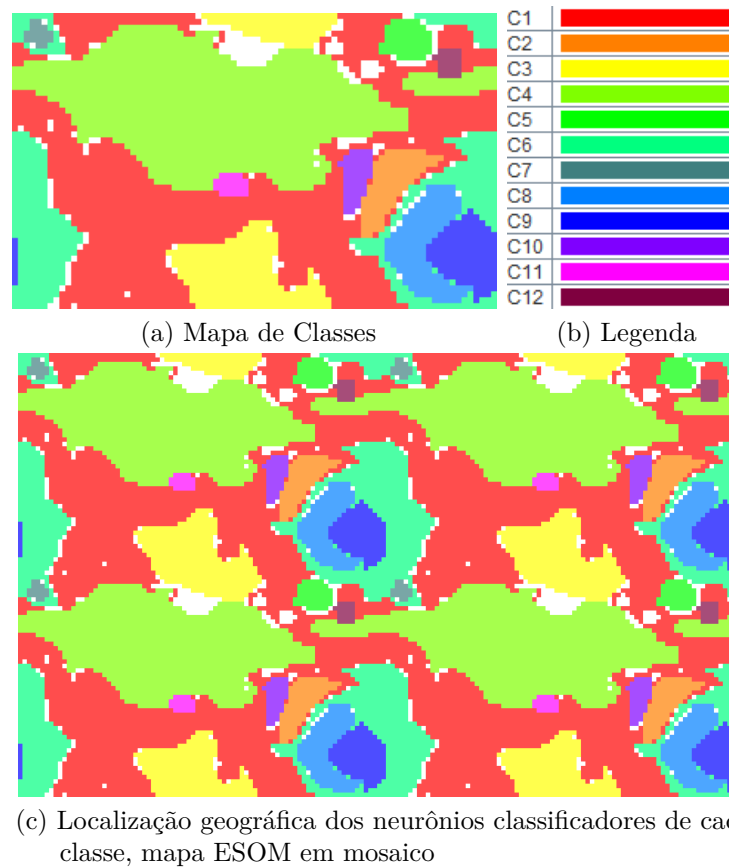


Figura 19 – Distribuição de classes ao longo do mapa ESOM.

Após o treinamento dos neurônios do *SOM*, a etapa seguinte consistiu no agrupamento de neurônios utilizando os conceitos descritos na Seção 2.2.1 e Seção 2.2.2 para

posterior classificação da base. O *SOM* foi segmentado considerando a representação *U-Matrix* seguindo os quesitos de vales ou planícies, separados por cadeia de montanhas e picos, e observando a representação *P-Matrix* seguindo o conceito de planaltos no mapa de densidades.

Dentro da Figura 19, a Figura 19a mostra a silhueta das classes de usuários que serão abordadas a seguir, e, de maneira análoga, a Figura 19c mostra-a em mosaico. Como o objeto de estudo foca-se em melhorar a permanência do usuário na partida e o balanceamento, pode-se observar a influência do componente *LeaverStatus* descrito anteriormente na Figura 14 com a segregação dos neurônios em classes na Figura 19a, em especial com as classes que mais apresentam usuários desertores frequentes.

De posse deste mapa, particionou-se os usuários em 12 classes diferentes. Usando a base de dados de usuários classificados, buscou-se analisar e avaliar a formação de partidas e cujo resultado foi condizente com o adotado pelo jogo *Dota 2*.

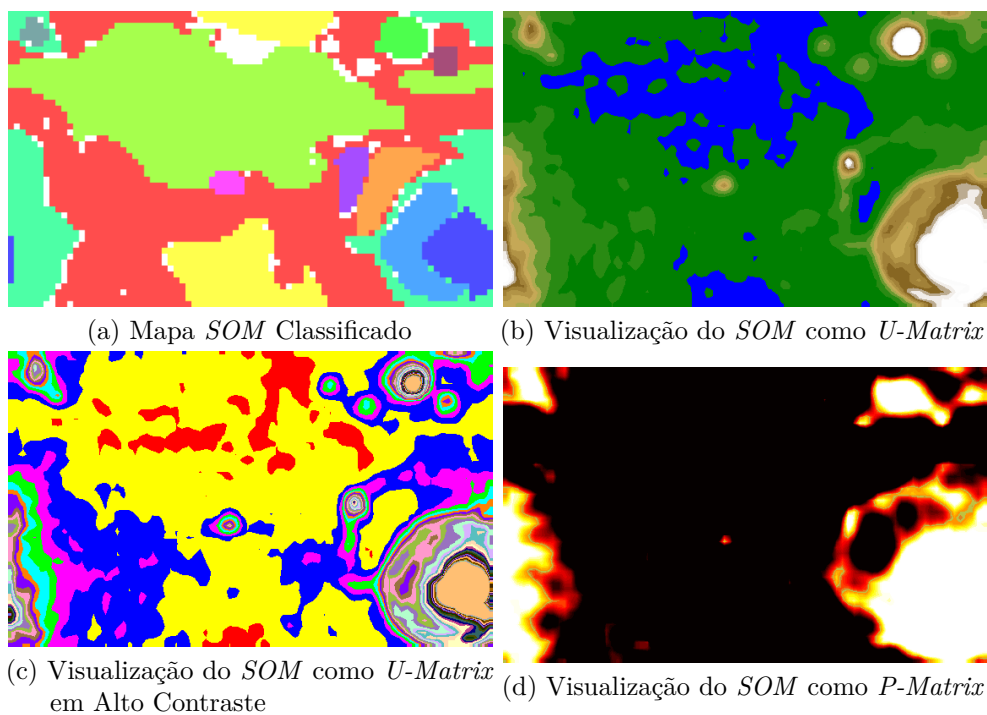


Figura 20 – Visualizações de Classificação de Regiões com *ESOM*

A Figura 20 coloca lado a lado todas as visualizações do mesmo *SOM* usadas no algoritmo do *ESOM* para agrupar os neurônios classificadores. A Figura 20a mostra o mapa já classificado para ser usado como referência; a 20b, a visualização *U-Matrix*; a 20c, a *U-Matrix* em alto contraste e a 20d, a *P-Matrix*. Uma forma de se entender o funcionamento do agrupamento e classificação usando *ESOM* é observar a classe *C2* (uma das classes denominadas *Common*) vista na Legenda 19b representada pela região em cor laranja na Figura 20a, formada a partir da separação por meio de cadeias de montanhas, observáveis nas demais figuras. Ao se observar o agrupamento de neurônios feito na Figura

Tabela 1 – Distribuição Percentual de Usuários por Classe

Classe de Usuários	%
C1 - <i>CommonA</i>	32.52
C2 - <i>CommonB</i>	36.49
C3 - <i>Leaver</i>	4.40
C4 - <i>Experience</i>	0.35
C5 - <i>Tactical</i>	9.54
C6 - <i>Supporter</i>	10.29
C7 - <i>Losers</i>	1.05
C8 - <i>SemiLeaver</i>	1.56
C9 - <i>Wealthy</i>	0.38
C10 - <i>TopPlayers</i>	0.17
C11 - <i>Speedy</i>	0.80
C12 - <i>FullLeaver</i>	2.45
<b>Total</b>	<b>100.00</b>

20a e comparar com as outras formas de visualização, nota-se que obedece fielmente a topografia exibida pelas visualizações em *U-Matrix* e *P-Matrix*.

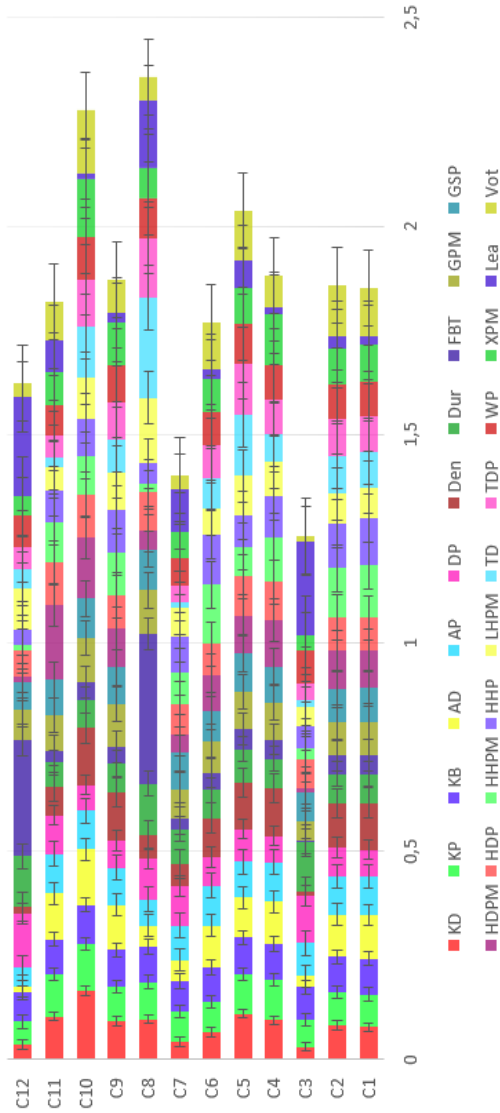
Com as classes definidas, realizou-se a estatística que mostra a distribuição percentual de usuários em cada classe, conforme a Tabela 1. Pode-se notar que as classes mais comuns são as *Common* (C1 e C2), *Supporter* (C6) e *Tactical* (C5). O significado do nome de cada rótulo de classe é abordado na Seção 4.3.

A Figura 21a exibe o valor médio de cada uma das 22 métricas criadas neste estudo relativas a cada uma das 12 classes de usuários indicadas na Tabela 1, especificando o erro padrão associado.

A Figura 21b mostra os dados utilizados para a geração da Figura 21a em escala de cores variando de verde a vermelho. Quanto mais vermelho for uma determinada média, melhor é a métrica em relação às demais classes, e de maneira oposta, quanto mais verde for a média, pior é a métrica. A observância de quais valores eram melhores, piores e intermediários de cada métrica levou a caracterização das classes. Por exemplo, ao observar a Figura 21b, consegue-se perceber que a classe C10 possui diversas métricas com os melhores valores em métricas bastante importantes, o que indica que são os usuários com o melhor desempenho.

O significado das siglas encontram-se descritos na Seção 3.3.1. Para favorecer a compreensão das tabelas a seguir, apresentam-se as siglas e os seus respectivos nomes: **KD**: *KillsDeaths*; **KP**: *KillsPercent*; **KB**: *KillsBalance*; **AD**: *AssistsDeaths*; **AP**: *AssistsPercent*; **DP**: *DeathsPercent*; **Den**: *Denies*; **Dur**: *Duration*; **FBT**: *FirtstBloodTime*; **GPM**: *GoldPerMinute*; **GSP**: *GoldSpentPercent*; **HDPM**: *HeroDamagePerMinute*; **HDP**: *HeroDamagePercent*; **HHPM**: *HeroHealingPerMinute*; **HHP**: *HeroHealingPercent*; **LHPM**: *LastHitsPerMinute*; **TD**: *TowerDamage*; **TDP**: *TowerDamagePercent*; **WP**: *WinsPercent*; **XpM**: *ExperiencePerMinute*; **Lea**: *LeaverStatus*; **Vot**: *Votes*.

### MÉTRICAS MÉDIAS DAS CLASSES



(a) Métricas Médias das Classes

	KD	KP	KB	AD	AP	DP	Den	Dur	FBT	GPM	GSP	HDP	HHPM	HHP	LHPM	TD	TDP	WP	XPM	Lea	Vot	
<b>C1</b>	1,536	0,198	0,503	2,455	0,199	0,209	5,784	37,887	140,960	389,437	0,201	275,956	0,197	10,649	0,164	2,847	1080,030	0,191	0,506	447,747	0,047	0,187
<b>C2</b>	1,611	0,204	0,497	2,347	0,198	0,222	5,564	38,056	139,576	381,616	0,202	276,967	0,205	9,992	0,153	2,779	1167,979	0,197	0,502	426,384	0,078	0,192
<b>C3</b>	0,617	0,162	0,460	0,648	0,170	0,371	0,465	65,082	6,989	225,381	0,173	34,866	0,173	2,155	0,078	1,747	238,274	0,087	0,489	173,281	0,565	0,020
<b>C4</b>	1,866	0,240	0,513	2,394	0,200	0,207	5,905	38,422	150,438	421,704	0,217	337,654	0,231	9,197	0,142	3,147	846,820	0,180	0,507	615,869	0,046	0,118
<b>C5</b>	2,162	0,240	0,513	2,249	0,185	0,255	5,744	43,810	153,079	424,755	0,235	274,884	0,236	6,113	0,107	3,643	1905,246	0,267	0,583	433,077	0,167	0,192
<b>C6</b>	1,290	0,182	0,491	2,354	0,205	0,220	4,908	36,580	134,420	352,903	0,186	260,496	0,189	12,053	0,174	2,344	943,186	0,175	0,480	398,524	0,057	0,180
<b>C7</b>	0,854	0,183	0,425	1,189	0,176	0,309	2,798	44,821	78,792	338,589	0,221	130,038	0,185	6,520	0,122	2,688	160,641	0,092	0,399	310,547	0,257	0,055
<b>C8</b>	1,900	0,227	0,499	1,192	0,135	0,315	2,968	66,991	1160,609	491,967	0,250	134,547	0,233	1,882	0,068	6,069	3103,575	0,313	0,588	363,611	0,406	0,090
<b>C9</b>	1,803	0,212	0,517	2,510	0,191	0,210	6,054	37,320	127,059	486,588	0,225	282,225	0,204	8,680	0,148	3,479	994,899	0,197	0,538	520,316	0,056	0,128
<b>C10</b>	3,254	0,280	0,559	3,126	0,197	0,203	6,997	36,787	141,248	503,749	0,238	440,135	0,265	7,844	0,129	3,710	1595,153	0,245	0,641	687,028	0,037	0,240
<b>C11</b>	2,037	0,253	0,500	2,590	0,201	0,303	3,565	31,422	91,543	408,650	0,212	541,930	0,257	8,375	0,106	2,209	304,163	0,114	0,441	405,284	0,188	0,149
<b>C12</b>	0,723	0,136	0,414	0,356	0,094	0,422	0,806	66,877	900,738	337,679	0,169	39,270	0,157	1,210	0,050	3,847	618,561	0,117	0,467	229,523	0,591	0,052

(b) Destaque dos valores médios das métricas

Figura 21 – Métricas Médias das Classes

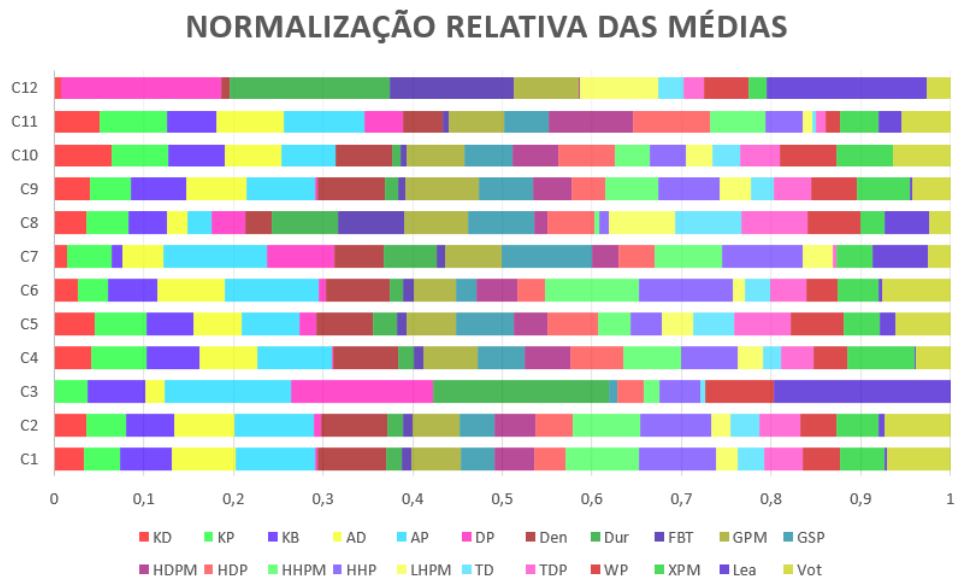


Figura 22 – Normalização Relativa das Médias

A Figura 22 é uma variação das duas figuras anteriores e mostra a silhueta das médias das métricas de cada classe. Para gerar esta silhueta executou-se uma normalização dos valores seguindo a Equação 4.1. Essa normalização gerou como resultado a atribuição do valor 1 ao grupo cujo componente apresentou a maior média para determinada métrica dentre as classes observadas e, 0 ao menor valor da média, enquanto que aos demais, foi aplicada uma função linear para ajustar ao intervalo entre 0 e 1. Em alguns casos, como em *DeathsPercent*, que representa a quantidade de vezes que o jogador morreu em relação ao total de vezes que os seus aliados morreram, essa normalização aconteceu de maneira inversa para representar o sentido do quanto menor, melhor. Em seguida, os valores foram escalados utilizando a Equação 4.2 para coincidirem e, portanto, permitir a comparação da distribuição relativa entre cada componente. Este cálculo foi feito para ressaltar o quanto que cada componente presente no vetor de características médio impacta relativamente em relação a outras classes. Para maiores detalhes e aprofundamento, o Apêndice C contém as demais tabelas e figuras intermediárias.

$$\frac{ValorCorrente - ValorMínimo}{ValorMáximo - ValorMínimo} \quad (4.1)$$

$$\frac{ValorCorrente}{SomatóriaDeTodosOsValoresDaClasse} \quad (4.2)$$

### 4.3 Interpretação

As classes foram rotuladas considerando os valores médios de cada um dos componentes. O *ESOM* conseguiu diferenciar 12 classes de usuários. Para efeito didático, os tipos de usuários foram agrupados em 3 superclasses com características comuns: *Leavers*, *TopPlayers* e *Common Users* e, 4 outras independentes, descritas a seguir:

***Leavers*** : jogadores cuja principal característica que os diferenciam é serem desertores frequentes. Este agrupamento foi caracterizado observando a componente *Leaver* na Figura 21 (valores em verde) e, principalmente, a Figura 22, na qual é possível perceber que essa métrica se destaca em relação as de outras classes. Na base de dados quando uma partida contava com a presença de um usuário desta classe, em pelo menos 25% dos casos ela se tornava inválida (*safe-to-leave*). Cerca de 8% dos usuários possuem estas características indesejáveis. Usualmente, em razão do abandono e da baixa interação com o jogo, as partidas com usuários desta superclasse tendem a demorar muito (*Duration*), devido ao baixo engajamento (deserção e *AFK*) de ambos os lados, restando poucos usuários para destruir a base adversária.

***FullLeaver (C12)***: possui quase todas as piores métricas, como por exemplo, a maior taxa de abandono (cerca de 60%), conforme pode ser observado na Figura 21b pelos valores em verde. Além disso, apresenta o menor valor da métrica *FirstBloodTime*, o que indica que jogadores dessa classe devam ser descuidados, normalmente morrendo logo no início do jogo, fato esse que deve incentivá-los a desistirem rapidamente;

***Leaver (C3)***: subclasse bastante semelhante ao *FullLeaver*, contudo, a componente *FirstBloodTime* é discrepantemente maior que os demais, o que indica que o abandono é tão grande em ambos os lados que nem chega a ocorrer uma morte na partida. Isto acontece quando um usuário abandona logo no começo e os demais o acompanham. Isto pode ser observado mais facilmente na Figura 22 ao observar a componente *FBT*;

***SemiLeaver (C8)***: subclasse dos desertores que mais se destacam em destruir torres e estruturas inimigas, fato destacado pela métrica *TowerDamage*. Correspondem ao subgrupo dos *Leavers* que possui as melhores métricas, embora ainda baixas, fato evidenciado na Figura 21b.

***Losers (C7)***: jogadores cujas principais características são de possuírem a menor taxa vitória (*WinsPercent*) e de dano a estruturas (*TowerDamage*), conforme pode ser observado na Figura 21b.

***TopPlayers*** jogadores que possuem as melhores médias de métricas de maneira geral, com excelente desempenho em partidas. Evidenciado na Figura 21b pelas cores mais

próximas ao vermelho.

**TopPlayer (C10):** são os melhores jogadores, de maneira geral, uma vez que possuem as melhores métricas de cada atributo, destacando-se pela maior taxa de vitórias (*WinsPercent*), observada na Figura 21b. Detectou-se na base de dados que a equipe que possuía um usuário desta classe ganhou cerca de 64% das partidas;

**Tactical (C5):** jogadores com bom desempenho nas métricas e, de maneira geral, possuem características mais táticas, preferindo atuar estrategicamente em estruturas (*TowerDamage*) do que em heróis. Também apresentam uma maior duração de partida (*Duration*), o que indica que sabem defender bem. Estes dados ficam mais claros ao se analisar essas métricas na Figura 21b;

**Common Users:** jogadores que formam a maior parcela de usuários, com desempenho baixo em todas as métricas. A observação da diferença entre as duas subclasses desse agrupamento é melhor observada na Figura 22.

**CommonA (C1):** Dentre os usuários comuns, são aqueles cuja principal diferença para o outro grupo de usuários comuns (CommonB) está na realização de cura (*HeroHealingPerMinute* e *HeroHealingPercent*) aos seus aliados e menos dano aos seus adversários (*HeroDamagePerMinute* e *HeroDamagePercent*). Além disso, são menos desertores (*Leaver*);

**CommonB (C2):** Dentre os usuários comuns, são aqueles cuja principal diferença é fazerem menos cura à aliados (*HeroHealingPerMinute* e *HeroHealingPercent*), mais dano a adversários (*HeroDamagePerMinute* e *HeroDamagePercent*) em relação ao outro grupo comum (CommonA) e serem mais desertores (*Leaver*);

**Speedy (C11):** Usuários que se destacam por proporcionarem que as partidas durem pouco. São os que possuem a menor média na componente *Duration* observada na Figura 21b;

**Supporter (C6):** Usuários que se destacam por possuírem os maiores valores nas taxas de cura (*HeroHealingPerMinute* e *HeroHealingPercent*) e de assistências em assassinatos (*AssistsDeaths* e *AssistsPercent*) observados na Figura 21b. O que indica que preferem se distanciar mais das batalhas e focam em curar os demais ou, dar suporte para que estes evoluam;

**Wealthy (C9):** Usuários que se destacam por possuírem elevados valores de arrecadação de *Gold* (*GoldPerMinute* e *GoldSpentPercent*) dentre as classes não-desertoras. As taxas não são tão elevadas, o que indica que possivelmente sejam usuários que passam boa parte do tempo na floresta e são os que usualmente acumulam mais dinheiro

dentro do seu time. Estes detalhes são melhor observados nas respectivas métricas exibidas na Figura 21b;

**Experience (C4):** Usuários cuja principal característica é a aquisição de elevada taxa de experiência (*eXperiencePerMinute*), perdendo apenas para a classe *TopPlayers*, conforme pode ser visualizado na Figura 21b.

## 4.4 Interpretação da Formação de Equipes

Neste trabalho foi realizado um estudo sobre a formação de equipes no jogo DOTA 2. Esse estudo foi baseado na classificação dos jogadores, oriunda de métricas e na avaliação de partidas observando a composição das equipes.

Uma vez tendo os diversos agrupamentos existentes entre jogadores, partimos para o último passo da metodologia empregada: analisar métricas de boas partidas e avaliar como essas partidas eram compostas em termos de jogadores.

Para se analisar quais métricas indicariam boas partidas, foi utilizado o conhecimento de um especialista no jogo, que indicou as métricas de duração e a diferença no número de assassinatos como sendo decisivas para se entender o nível de disputa de uma partida.

A duração é algo diretamente ligado a esse tipo de jogo. Quanto mais um jogo durar, significa que houve mais disputa e nenhum dos times conseguiu sobrepujar de forma fácil o adversário. A diferença entre assassinatos é importante pois foca em um aspecto normalmente associado à definição do ganhador e que reflete a interação entre jogadores. Ela representa uma forma de medir o quanto um usuário atua em outro. Neste sentido, pode-se fazer uma analogia com a métrica de chutes a gol de um jogo de futebol. Embora essa métrica reflita diretamente quem mais atuou no jogo e que teve mais chances, nem sempre quem ganha é quem mais cria chances de gol, contudo é uma boa métrica para indicar se a partida está disputada.

A partir da definição das métricas de boas partidas a serem usadas foram analisadas como essas partidas eram compostas em termos de jogadores, usando as classes encontradas neste trabalho e como isso influenciava nas métricas que indicavam partidas equilibradas.

Percebeu-se que a existência de um desbalanceamento das classes levava a um desequilíbrio no jogo, ou seja, quanto mais balanceadas eram as equipes, seguindo as classes aqui identificadas, mais equilibrada era a partida. A Tabela 2 mostra o impacto do quanto uma formação de times desbalanceada implica na qualidade de entretenimento de uma partida.

Foi definida a métrica grau de desbalanço, visando com isso encontrar seu impacto nas métricas associadas a partidas disputadas. O grau de desbalanço é uma métrica de



Tabela 2 – Impacto do Balanço de Equipes no Jogo Dota 2

Grau de Desbalanço	Duração Média da Partida	Média de Assassinatos do time do Vencedor	Proporção de Partidas
0 – 1	42m55s ± 14m21s	61.87% ± 9.95%	32,24%
2	42m01s ± 13m06s	62.24% ± 8.92%	39,66%
3	37m49s ± 12m36s	63.42% ± 8.38%	22,72%
4	31m06s ± 10m24s	66.07% ± 8,31%	4,85%
5	21m26s ± 7m40s	68.49% ± 10.80%	0,53%

partida que mede a diferença entre o quanto um perfil de classes de um time difere do outro. A Equação 4.3 mostra como essa métrica é calculada.

$$\text{Grau de Desbalanço} = \frac{\sum_{i=1}^n |T1_i - T2_i|}{2} \quad (4.3)$$

Onde:

$i$  : Identificador da  $i$ -ésima classe

$n$  : Número de classes

$T1_i$  : Quantidade de usuários da classe  $i$  presentes no Time 1

$T2_i$  : Quantidade de usuários da classe  $i$  presentes no Time 2

Para exemplificar o funcionamento da Equação 4.3, considere uma partida formada por dois times, com a seguinte composição: Time 1 -  $[A, A, B, B, C]$  e, Time 2 -  $[A, A, A, B, C]$ . A classe “A” tem a diferença de proporção entre os times de uma unidade, e, da mesma forma, a classe “B”. Entretanto, a classe “C” possui diferença nula apresentando a mesma quantidade de usuários em cada time, resultando no grau de desbalanço 1. De maneira análoga, caso os times fossem compostos por  $[A, B, B, E, F]$  e  $[B, C, G, H, H]$  ter-se-ia o valor de desbalanço 4, uma vez que existe apenas 1 usuário de cada time em equilíbrio de classes.

A Tabela 2 mostra que quanto maior é o grau de desbalanço entre os times, algumas métricas diretamente relacionadas ao nível de entretenimento são afetadas. Uma partida equilibrada, com desbalanço de 0 a 3 tende a demorar mais tempo, considerando que há um desafio maior em sobrepujar o oponente. Pode-se notar também que a proporção de assassinatos do time vencedor pelo total de assassinatos é tanto maior quanto é o desbalanço. Isso indica que um lado sobrepuja o outro mais facilmente, levando a um estado de pouco desafio para a equipe vencedora e também o sentimento de impossibilidade de reverter a situação, por parte do lado perdedor.

A análise relativa à proporção de partidas pelo nível de desbalanço nos indica que o jogo já possui um mecanismo de formação de equipes equilibradas, uma vez que a maioria das partidas se concentra nos menores níveis de desbalanço. Mais de 70% das partidas possuem um grau de desbalanço de até dois elementos. Apenas 5% das partidas possuem grau de desbalanço maior ou igual a quatro, o que indica que o jogo já faz algum algoritmo inteligente de *matchmaking*. No entanto, as conclusões obtidas aqui podem ajudar a melhorar ainda mais a competitividade das partidas.

É interessante registrar que para iniciar um jogo, os pretensos jogadores ficam em uma fila de espera, até a formação da partida. Isso deve ser feito de forma rápida, para não causar desistências. Dessa forma, o jogo deve ter um mecanismo fácil de formação. Com base nos resultados aqui obtidos, seria fácil criar um algoritmo de otimização para formar a maior quantidade de partidas com o mínimo de desbalanço entre elas.

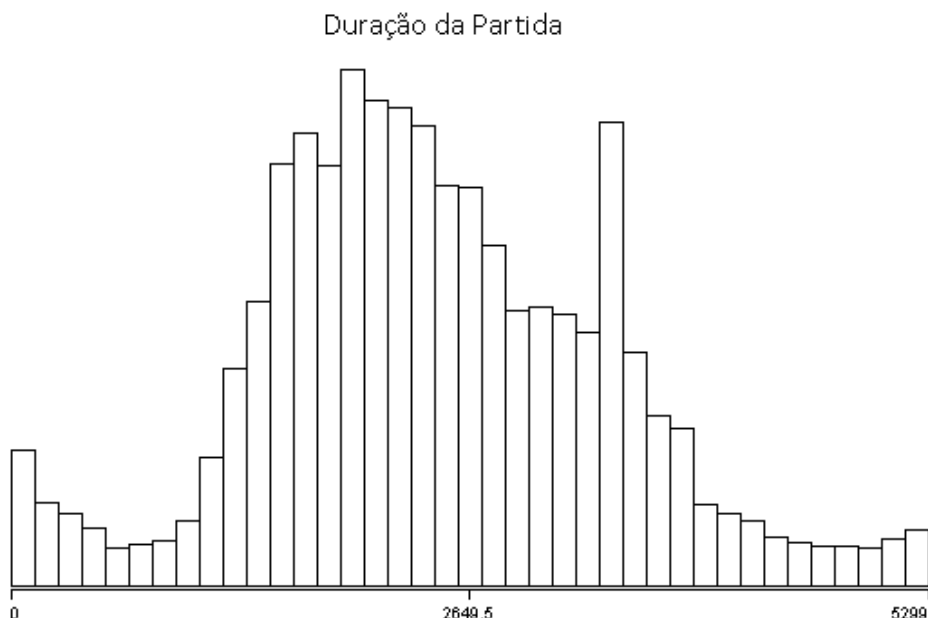


Figura 23 – Histograma da Duração da Partida em Segundos

A Figura 23 mostra o histograma de distribuição da duração da partida. Pode-se perceber que há uma considerável quantidade de partidas que duram pouco, muitas delas devem-se ao fato de que no início da partida o usuário sofre uma morte e se sente injustiçado ou envergonhado e abandona, e isto encoraja os demais a desistirem. Este tipo de assassinato geralmente ocorre quando usuários experientes enfrentam novatos.

Depois do alto valor local inicial de partidas com baixíssima duração (barra inicial do histograma), observa-se que poucas partidas terminam com uma duração baixa, o que é de se esperar, uma vez que o jogo não possui meios para uma equipe sobrepujar a outra em tão pouco tempo a ponto de causar uma vitória. Em seguida os valores de duração de partida passam a seguir uma distribuição semelhante a uma curva *lognormal* uma vez que representa a capacidade dos jogadores de um time de acumular suficientemente uma

diferença de vantagens em relação ao outro. O valor aberrante da curva deve-se a algum mecanismo de proteção dentro do jogo que envolve uma desconexão forçada pelo sistema. Este tipo de observação deixa claro o quanto é importante formar partidas com usuários com características semelhantes.

Tabela 3 – Duração por Desbalanço

<b>Duração (minutos)</b>	<b>Desbalanço Médio</b>	<b>Número de Partidas</b>
5	2.1885	2960
10	2.1209	1059
15	2.0507	2010
20	2.1271	5491
25	2.0742	9631
30	1.8990	8261
35	1.8622	8630
40	1.8310	7029
45	1.8539	5360
50	1.8185	4258
55	1.8531	4233
60	1.8593	2303
65	1.8199	1066
70	1.7920	697
75	1.7820	500
80	1.7382	424

Os dados contidos na Tabela 3 detalham a Figura 24 e mostram a relação entre a duração da partida e o desbalanço. Ao se observar as partidas ordenando-as segundo a duração de maneira crescente, nota-se que há uma tendência mostrando que quanto mais a partida dura, menor é o desbalanço ou seja, maior equilíbrio entre os times. Ou seja, o balanceamento das equipes, em termos das classes aqui descobertas, favorece a criação de partidas mais equilibradas.

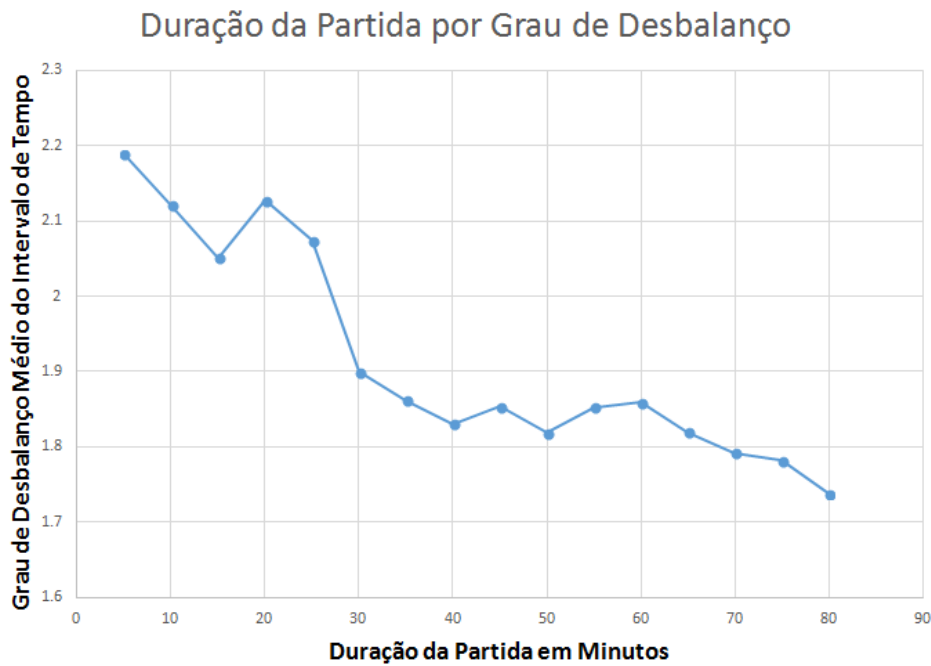


Figura 24 – Duração por Desbalanço

$$\Delta Mortes = \frac{AssassinatosDoTimeVencedor - AssassinatosDoTimePerdedor}{TotalDeAssassinatos} \quad (4.4)$$

Para complementar a análise da utilidade do Grau de Desbalanço, foi observada a relação com a métrica Delta Mortes a ser definida pela Equação 4.4. Esta métrica mostra o quanto que um time vencedor é efetivo em fazer assassinatos sobre o time perdedor.

Tabela 4 – Delta Mortes por Grau de Desbalanço

Delta Mortes(%)	Desbalanço Médio	Número de Partidas
-40	1,9713	349
-30	1,9716	739
-20	2,0299	937
-10	2,0735	1061
0	2,1462	14990
10	2,1326	4222
20	2,1566	5446
30	2,1752	5074
40	2,2690	3907
50	2,2791	7365
60	2,2546	2123

Foi possível observar, na Tabela 4 e evidenciada pela Figura 25, uma certa tendência de crescimento. Quanto mais positivo for o Delta Mortes, maior é a indicação de que

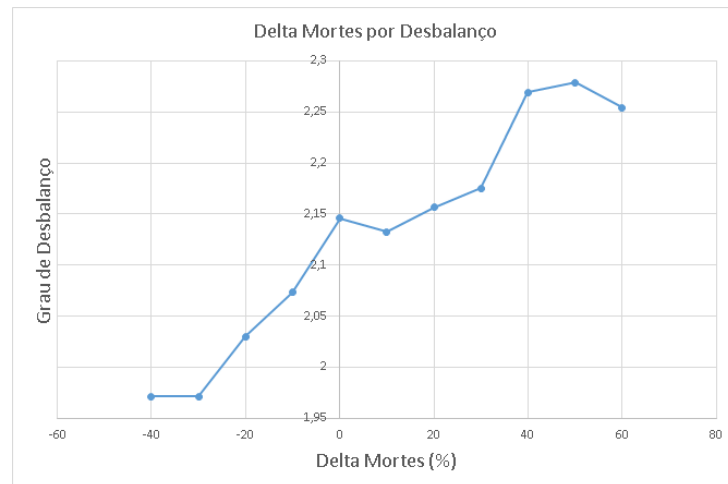


Figura 25 – Delta Mortes por Grau de Desbalanço

os times foram formados de maneira desbalanceada (grau de desbalanço elevado). Ao contrário, quanto mais negativo for o Delta Mortes, ou seja, o time vencedor foi o que causou menos assassinatos, tem-se partidas equilibradas, evidenciado pelo baixo Grau de Desbalanço. Ainda neste sentido, pode-se interpretar que quanto maior é o desbalanço, maior é a diferença entre assassinatos do time vencedor em relação ao perdedor, ou que um time sobrepõe o outro com mais facilidade.

#### Delta Mortes por Quantidade de Partidas

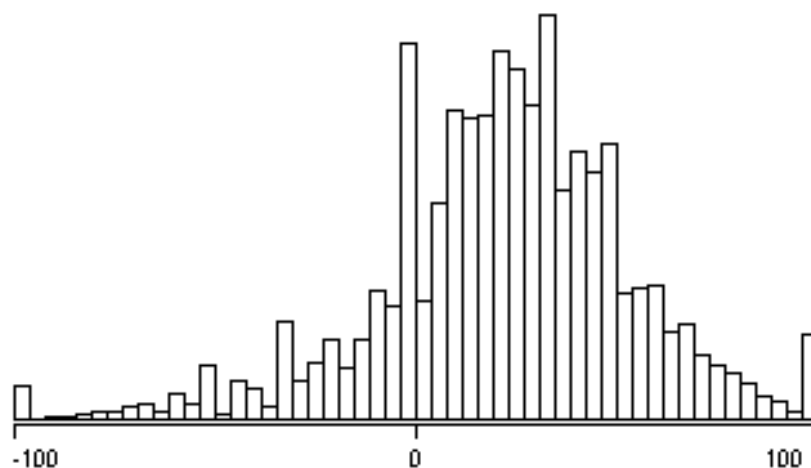


Figura 26 – Histograma de Delta Mortes

Para concluir a interpretação do Delta Mortes, a Figura 26 mostra com mais granularidade o histograma dessa métrica. Percebe-se que o time vencedor tende a possuir uma maior taxa de assassinatos em relação ao time adversário, que pode ser observado como a concentração maior das frequências no lado direito. Este comportamento era de se esperar, contudo existem partidas nas quais os vencedores são os que menos fizeram

assassinatos e esses dados estão relacionados com o grau de desbalanço como foi mostrado anteriormente.

Todas as métricas auxiliares aqui descritas foram desenvolvidas com o intuito de se comprovar a coerência da métrica Grau de Desbalanço e como ela pode ser usada para avaliar a formação de equipes.

## 5 Conclusão

Este trabalho apresentou um estudo sobre a formação de equipes em jogos virtuais. A formação de equipes é um problema pouco abordado pela literatura e também um tema constante em fóruns da comunidade de jogadores, uma vez que existem muitas reclamações sobre o tema.

Neste trabalho foi utilizada uma metodologia para abordar o tema, baseada no processo de descoberta de conhecimento em banco de dados. Essa metodologia é dividida em várias etapas, incluindo a coleta de dados, pré-processamento, transformação e a efetiva mineração, usando algoritmos de inteligência artificial. A partir dos resultados da mineração foi possível identificar agrupamentos de usuários bem significativos. Após a análise de um especialista, os grupos foram rotulados, visando refletir as principais características associadas a cada grupo.

Foram identificados 12 agrupamentos de usuários. Esses agrupamentos foram ainda organizados em 3 superclasses com características comuns: *Leavers*, *TopPlayers* e *Common Users*, além de serem identificados 4 outras classes. Cada uma das classes refletia um grupo de usuários com características em comum.

De posse dos grupos de usuários identificados, foi possível analisar dados das partidas, especialmente aquelas com bons indicadores de competitividade, para tentar se inferir como deveria ser a formação de times para maximizar o engajamento e diversão. Foram utilizadas duas métricas para se avaliar as partidas: duração e diferença de assassinatos. Essas métricas foram sugeridas por um especialista no jogo por refletirem bem o nível de disputa de uma partida.

Foi possível descobrir que as partidas com maior equilíbrio entre os grupos de usuários identificados na mineração eram partidas com maior duração e menor diferença de assassinatos entre vencedor e perdedor. Por conta dessa descoberta, notou-se que um bom critério para formação de times no Dota 2 seria equilibrar a quantidade de jogadores de cada uma das 12 classes encontradas na mineração realizada, uma vez que isso favorece a disputa em uma partida.

O estudo realizado permitiu a identificação de uma macro abordagem para se tratar com a formação de equipes em qualquer jogo, que é uma outra contribuição deste trabalho. Essa macro abordagem é dividida em etapas, tais quais as etapas realizadas neste trabalho:

1. Coleta de dados de partidas. Nessa etapa deve-se coletar os dados de partidas, tais como a duração, equipe vencedora, quantidade de pontos, jogadores participantes, etc.

2. Obtenção de métricas de jogadores. Nessa etapa deve-se gerar métricas para cada jogador presente nas partidas coletadas na etapa anterior. A definição de métricas para caracterizar os jogadores deve ser feita em conjunto com um especialista no jogo, de forma a gerar dados que sejam muito relevantes para a classificação dos jogadores.
3. Identificação de agrupamentos de jogadores. Nessa etapa devem ser identificados agrupamentos de usuários, utilizando como base as métricas geradas na etapa anterior. O principal objetivo é identificar jogadores com características em comum.
4. Identificação de métricas de disputa de partidas. Nessa etapa devem ser identificadas as métricas que melhor exprimem uma partida equilibrada. Essa etapa também exige a participação de um especialista no jogo em questão.
5. Avaliação da composição das equipes nas partidas. Nessa etapa deve ser realizada uma avaliação da composição das equipes nas partidas, analisando em especial as métricas identificadas na etapa anterior. Essa avaliação deve identificar padrões de equipe associados a certos indicadores de partidas.
6. Sugestão de uma diretriz para composição de times. Com base na avaliação realizada na etapa anterior, será possível gerar uma diretriz para a formação de equipes. Essa sugestão pode ser incorporada em algum mecanismo automático de formação ou ser usada de forma manual.

É importante ressaltar que este trabalho é inovador na área de pesquisa em *Game Metrics*, *Analytics* e *Telemetry*, uma vez que poucos estudos analisam o perfil de usuários visando prever desconexões (MAHLMANN et al., 2010) ou formar equipes (POBIEDINA et al., 2013a) (POBIEDINA et al., 2013b). Trata-se de conhecimento que as empresas não costumam compartilhar.

## 5.1 Trabalhos Futuros

A realização deste trabalho mostrou que ainda existe muito a ser feito na indústria de jogos naquilo que se refere à formação de equipes. Por conta disso, uma linha de continuidade direta desta pesquisa é a aplicação da proposta aqui identificada a outros jogos que envolvam disputas entre times. Uma sugestão de jogo a ser abordado é o *League of Legends*, desenvolvido pela *Riot Games*, e que é um concorrente direto do *Dota 2*. A principal ideia associada a esse trabalho é tentar validar a macro abordagem proposta aqui em um outro contexto.

A exploração de outros dados para o agrupamento de usuários também poderia ser uma outra vertente a ser explorada. Neste trabalho foram utilizados dados relacionados



---

à visão estática obtida ao final da partida, mas não foi considerado nada relativo à forma como um jogador joga uma partida. Diversos jogos fornecem essa informação. Ela contém aspectos não contemplados no sumário de uma partida e pode gerar uma visão completamente diferente daquela usada aqui.

Um outro trabalho bastante interessante seria embutir a diretriz de formação de equipes aqui identificada no Dota 2. Dessa forma, a formação de equipes utilizaria a sugestão identificada neste trabalho. Isso permitiria avaliar se essa diretriz gera melhores resultados, permitindo ainda uma comparação da estratégia utilizada pela Valve e a estratégia aqui proposta. No entanto, isso exige a participação direta da Valve, que é a responsável pelo jogo.



# Referências

- ADLER, J.; PARMRYD, I. Quantifying colocalization by correlation: the pearson correlation coefficient is superior to the mander's overlap coefficient. *Cytometry Part A*, Wiley Online Library, v. 77, n. 8, p. 733–742, 2010. Citado na página 38.
- ANAYA, M. et al. Data-driven methodology to detect and classify structural changes under temperature variations. *Smart Materials and Structures*, IOP Publishing, v. 23, n. 4, p. 045006, 2014. Citado na página 18.
- BOSC, G. et al. Strategic Pattern Discovery in RTS-games for E-Sport with Sequential Pattern Mining. p. 1–11, 2013. Disponível em: <<http://liris.cnrs.fr/Documents/Liris-6365.pdf>>. Citado na página 1.
- CANOSSA, A. *Play-Persona: Modeling Player Behavior in Computer Games*. [S.l.]: Danmarks Designskole, 2009. Citado na página 11.
- CANOSSA, A. Meaning in gameplay: Filtering variables, defining metrics, extracting features and creating models for gameplay analysis. In: *Game Analytics*. [S.l.]: Springer, 2013. p. 255–283. Citado na página 11.
- CANOSSA, A.; DRACHEN, A. Patterns of play: Play-personas in user-centred game development. In: *Breaking New Ground: Innovation in Games, Play, Practice and Theory: Proceedings of the 2009 DiGRA Conference, London, Brunel University (September 2009)*. [S.l.: s.n.], 2009. Citado na página 11.
- CARDAMONE, L.; LOIACONO, D.; LANZI, P. L. Learning drivers for torcs through imitation using supervised methods. In: IEEE. *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*. [S.l.], 2009. p. 148–155. Citado na página 11.
- CLUA, E. et al. A non-intrusive approach for game design analysis based on provenance data extracted from game streaming. In: IEEE. *Games and Digital Entertainment (SBGAMES), 2014*. [S.l.], 2014. Citado na página 11.
- CROUX, C.; DEHON, C. Influence functions of the spearman and kendall correlation measures. *Statistical Methods & Applications*, Springer, v. 19, n. 4, p. 497–515, 2010. Citado na página 38.
- DETERDING, S. et al. Gamification. using game-design elements in non-gaming contexts. In: ACM. *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. [S.l.], 2011. p. 2425–2428. Citado na página 1.
- DRACHEN, A.; CANOSSA, A.; YANNAKAKIS, G. N. Player modeling using self-organization in tomb raider: Underworld. In: IEEE. *Computational Intelligence and Games, 2009. CIG 2009. IEEE Symposium on*. [S.l.], 2009. p. 1–8. Citado 3 vezes nas páginas 13, 39 e 46.
- DRACHEN, A. et al. Skill-based differences in spatio-temporal team behaviour in defence of the ancients 2 (dota 2). 2014. Citado na página 14.

- EARNINGS e S. *e-Sports Earnings*. 2014. <<http://www.esportsearnings.com/tournaments>>. Accessed: 2014-12-03. Citado na página 1.
- EL-NASR, M. S.; DRACHEN, A.; CANOSSA, A. *Game analytics: Maximizing the value of player data*. [S.l.]: Springer, 2013. Citado na página 9.
- EL-NASR, M. S.; DRACHEN, A.; CANOSSA, A. Introduction. In: EL-NASR, M. S.; DRACHEN, A.; CANOSSA, A. (Ed.). *Game Analytics*. Springer London, 2013. p. 3–12. ISBN 978-1-4471-4768-8. Disponível em: <[http://dx.doi.org/10.1007/978-1-4471-4769-5\\_1](http://dx.doi.org/10.1007/978-1-4471-4769-5_1)>. Citado na página 9.
- ESPN. *The International Dota 2 Championship Main Event: Grand Finals*. 2014. <[http://espn.go.com/watchespn/index/\\_/id/1934071/](http://espn.go.com/watchespn/index/_/id/1934071/)>. Accessed: 2014-08-11. Citado na página 1.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine*, v. 17, n. 3, p. 37, 1996. Citado na página 4.
- FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, ACM, v. 39, n. 11, p. 27–34, 1996. Citado na página 4.
- FAYYAD, U. M. et al. *Advances in knowledge discovery and data mining*. the MIT Press, 1996. Citado na página 29.
- FISHER, D. et al. Interactions with big data analytics. *interactions*, ACM, v. 19, n. 3, p. 50–59, 2012. Citado na página 3.
- FLURRY. *Flurry Five-Year Report: It's an App World. The Web Just Lives in It*. 2014. <<http://www.flurry.com/bid/95723/Flurry-Five-Year-Report-It-s-an-App-World-The-Web-Just-Lives-in-It>>. Accessed: 2014-10-14. Citado na página 1.
- FU, T.-c. et al. Pattern discovery from stock time series using self-organizing maps. In: CITESEER. *Workshop Notes of KDD2001 Workshop on Temporal Data Mining*. [S.l.], 2001. p. 26–29. Citado na página 17.
- GAO, L. et al. Classifying dota 2 hero characters based on play style and performance. In: . [S.l.: s.n.], 2013. Citado na página 19.
- GRUNZ, A.; MEMMERT, D.; PERL, J. Tactical pattern recognition in soccer games by means of special self-organizing maps. *Human movement science*, Elsevier, v. 31, n. 2, p. 334–343, 2012. Citado na página 39.
- HARRISON, B.; ROBERTS, D. L. Analytics-driven dynamic game adaption for player retention in scrabble. In: IEEE. *Computational Intelligence in Games (CIG), 2013 IEEE Conference on*. [S.l.], 2013. p. 1–8. Citado na página 12.
- HARRISON, B. E.; ROBERTS, D. Analytics-driven dynamic game adaption for player retention in a 2-dimensional adventure game. In: *Tenth Artificial Intelligence and Interactive Digital Entertainment Conference*. [S.l.: s.n.], 2014. Citado na página 12.
- HARRISON, B. E.; ROBERTS, D. L. When players quit (playing scrabble). In: . [S.l.: s.n.], 2012. Citado na página 13.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2. ed. [S.l.]: Prentice Hall, 2001. Citado na página 16.

HOORN, N. V. et al. Robust player imitation using multiobjective evolution. In: IEEE. *Evolutionary Computation, 2009. CEC'09. IEEE Congress on*. [S.l.], 2009. p. 652–659. Citado na página 11.

HULLE, M. M. V. Self-organizing maps. In: *Handbook of Natural Computing*. [S.l.]: Springer, 2012. p. 585–622. Citado na página 16.

KARZANOV, A. V. Minimum 0-extensions of graph metrics. *European Journal of Combinatorics*, Elsevier, v. 19, n. 1, p. 71–101, 1998. Citado na página 14.

KATAVIĆ, T. Using metrics to create a dynamic level. 2013. Citado na página 1.

KIM, J. H. et al. Tracking real-time user experience (true): a comprehensive instrumentation solution for complex systems. In: ACM. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. [S.l.], 2008. p. 443–452. Citado na página 10.

KOHONEN; TEUVO. Springer series in information sciences. *Self-organizing maps*, Springer, v. 30, 1995. Citado na página 15.

KOHONEN, T. *Self-organizing Maps*. [S.l.]: Springer, 1977. ISBN 3-540-62017-6. Citado na página 15.

KOHONEN, T. An introduction to neural computing. *Neural networks*, Elsevier, v. 1, n. 1, p. 3–16, 1988. Citado na página 16.

KOHONEN, T. The self-organizing map. *Proceedings of the IEEE*, v. 78, n. 9, p. 1464–1480, Sep 1990. ISSN 0018-9219. Citado 2 vezes nas páginas 15 e 16.

KOHONEN, T. Essentials of the self-organizing map. *Neural Networks*, Elsevier, v. 37, p. 52–65, 2013. Citado na página 16.

KOHWALTER, T. C.; CLUA, E.; MURTA, L. Provenance in games. In: *Brazilian Symposium on Games and Digital Entertainment (SBGAMES)*. [S.l.: s.n.], 2012. Citado na página 11.

LAWRENCE, I.; LIN, K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, JSTOR, p. 255–268, 1989. Citado na página 38.

LI, C.-S.; YU, P. S.; CASTELLI, V. Malm: a framework for mining sequence database at multiple abstraction levels. In: ACM. *Proceedings of the seventh international conference on Information and knowledge management*. [S.l.], 1998. p. 267–272. Citado na página 17.

LOPES, R.; BIDARRA, R. Adaptivity challenges in games and simulations: a survey. *Computational Intelligence and AI in Games, IEEE Transactions on*, IEEE, v. 3, n. 2, p. 85–99, 2011. Citado na página 13.

LYERLY, S. B. The average spearman rank correlation coefficient. *Psychometrika*, Springer, v. 17, n. 4, p. 421–428, 1952. Citado na página 38.

- MAHLMANN, T. et al. Predicting player behavior in tomb raider: Underworld. In: IEEE. *Computational Intelligence and Games (CIG), 2010 IEEE Symposium on*. [S.l.], 2010. p. 178–185. Citado 2 vezes nas páginas 13 e 66.
- MAREN, A. J.; HARSTON, C. T.; PAP, R. M. *Handbook of neural computing applications*. [S.l.]: Academic Press, 2014. Citado na página 16.
- MCCONVILLE, K. M.; MILOSEVIC, M. Active video game head movement inputs. *Personal and ubiquitous computing*, Springer-Verlag, v. 18, n. 1, p. 253–257, 2014. Citado 2 vezes nas páginas 32 e 45.
- MEDLER, B.; JOHN, M.; LANE, J. Data cracker: developing a visual game analytic tool for analyzing online gameplay. In: ACM. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. [S.l.], 2011. p. 2365–2374. Citado na página 12.
- MILLENIUM. *Dota 2, Introduction*. 2013. <<http://www.millenum.org/dota-2/accueil/guides/dota-2-introduction-presentation-et-details-du-jeu-moba-72934>>. Accessed: 2014-10-14. Citado na página 22.
- ORTEGA, J. et al. Imitating human playing styles in super mario bros. *Entertainment Computing*, Elsevier, v. 4, n. 2, p. 93–104, 2013. Citado na página 11.
- PAMPALK, E.; RAUBER, A.; MERKL, D. *Using smoothed data histograms for cluster visualization in self-organizing maps*. [S.l.]: Springer, 2002. Citado na página 17.
- PIATESKI, G.; FRAWLEY, W. *Knowledge discovery in databases*. [S.l.]: MIT press, 1991. Citado na página 29.
- POBIEDINA, N. et al. *On successful team formation*. [S.l.], 2013. Citado 3 vezes nas páginas 13, 14 e 66.
- POBIEDINA, N. et al. Ranking factors of team success. In: INTERNATIONAL WORLD WIDE WEB CONFERENCES STEERING COMMITTEE. *Proceedings of the 22nd international conference on World Wide Web companion*. [S.l.], 2013. p. 1185–1194. Citado 3 vezes nas páginas 13, 14 e 66.
- PRIDDY, K. L.; KELLER, P. E. *Artificial neural networks: an introduction*. [S.l.]: SPIE Press, 2005. v. 68. Citado na página 15.
- SAID, S. E.; DICKEY, D. A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, Biometrika Trust, v. 71, n. 3, p. 599–607, 1984. Citado 2 vezes nas páginas 32 e 45.
- SKADISTATS. *Clarity: Comically fast, almost complete Dota 2 "demo"(aka "replay") parser written in java*. [S.l.]: GitHub, 2014. <<https://github.com/skadistats/clarity>>. Accessed: 2014-07-16. Citado na página 29.
- TOGELIUS, J.; SCHMIDHUBER, J. An experiment in automatic game design. In: IEEE. *Computational Intelligence and Games, 2008. CIG'08. IEEE Symposium On*. [S.l.], 2008. p. 111–118. Citado na página 12.
- TRINCA, D. et al. Jogos online. *Revista de Informática Aplicada/Journal of Applied Computing*, v. 2, n. 1, 2010. Citado na página 1.

- TYCHSEN, A.; CANOSSA, A. Defining personas in games using metrics. In: ACM. *Proceedings of the 2008 Conference on Future Play: Research, Play, Share*. [S.l.], 2008. p. 73–80. Citado na página 11.
- ULTSCH, A. Data mining and knowledge discovery with emergent self-organizing feature maps for multivariate time series. *Kohonen maps*, Citeseer, v. 46, p. 33–46, 1999. Citado na página 16.
- ULTSCH, A. *Optimal density estimation in data containing clusters of unknown structure*. [S.l.]: Univ., 2003. Citado na página 19.
- ULTSCH, A. *U\*-matrix: a tool to visualize clusters in high dimensional data*. [S.l.]: Fachbereich Mathematik und Informatik, 2003. Citado 2 vezes nas páginas 17 e 19.
- ULTSCH, A.; MÖRCHEN, F. Esom-maps: tools for clustering, visualization, and classification with emergent som. Univ., 2005. Citado 4 vezes nas páginas 17, 19, 39 e 46.
- UPTON, G. J. Fisher's exact test. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, JSTOR, p. 395–402, 1992. Citado na página 12.
- VALVE. *Dota 2 Blog*. 2014. <<http://blog.dota2.com/>>. Accessed: 2014-08-11. Citado na página 1.
- VALVE. *Dota 2 Steam Chart*. 2014. <<http://steamcharts.com/app/570>>. Accessed: 2014-07-16. Citado na página 1.
- VALVE. Motion picture, *Free to Play: The Movie*. 2014. <<http://www.freetoplaythemovie.com/>>. Accessed: 2014-07-16. Citado na página 19.
- VALVE. *The International Championships*. 2014. <<http://www.dota2.com/international/>>. Accessed: 2014-07-16. Citado na página 1.
- WOLF, M. J. *The medium of the video game*. [S.l.]: University of Texas Press, 2001. Citado na página 1.
- YANG, P.; HARRISON, B.; ROBERTS, D. L. Identifying patterns in combat that are predictive of success in moba games. 2014. Citado na página 14.
- YANNAKAKIS, G. N.; TOGELIUS, J. Experience-driven procedural content generation. *Affective Computing, IEEE Transactions on*, IEEE, v. 2, n. 3, p. 147–161, 2011. Citado na página 12.





# Apêndices



# APÊNDICE A – Matrizes de Coeficientes de Correlação entre Atributos

KD	KP	KB	AD	AP	DP	Den	DenP	Dur	FBT	GpM	GSP	GSpm	HDpM	HDP	HHpM	HHP	LHpM	LHP	TD	TDP	WP	XpM	Lea	Vot	PV	NV	
KD	1,000	0,572	0,428	0,614	0,123	0,396	0,425	0,401	0,258	0,083	0,625	0,422	0,584	0,560	0,516	0,023	0,010	0,377	0,413	0,487	0,415	0,315	0,567	0,338	0,099	0,097	0,093
KP	0,572	1,000	0,209	0,148	0,112	0,254	0,339	0,431	0,058	0,117	0,546	0,625	0,478	0,468	0,843	0,114	0,134	0,472	0,586	0,401	0,487	0,056	0,466	0,222	0,073	0,072	0,068
KB	0,428	0,209	1,000	0,509	0,307	0,281	0,209	0,170	0,095	0,154	0,353	0,136	0,352	0,327	0,184	0,173	0,152	0,110	0,105	0,285	0,128	0,614	0,365	0,287	0,059	0,059	0,053
AD	0,614	0,148	0,509	1,000	0,422	0,600	0,433	0,279	0,544	0,235	0,455	0,106	0,532	0,588	0,137	0,307	0,270	0,034	0,085	0,332	0,216	0,344	0,602	0,592	0,093	0,093	0,084
AP	0,123	0,112	0,307	0,422	1,000	0,420	0,179	0,138	0,259	0,399	0,110	0,102	0,218	0,338	0,238	0,202	0,224	0,126	0,028	0,051	0,071	0,056	0,302	0,414	0,044	0,045	0,037
DP	0,396	0,254	0,281	0,600	0,420	1,000	0,508	0,379	0,437	0,262	0,472	0,190	0,529	0,586	0,198	0,235	0,266	0,139	0,197	0,346	0,323	0,001	0,699	0,692	0,087	0,089	0,077
Den	0,425	0,339	0,209	0,433	0,179	0,508	1,000	0,807	0,481	0,163	0,587	0,299	0,603	0,609	0,312	0,079	0,071	0,332	0,387	0,453	0,455	0,072	0,679	0,536	0,109	0,110	0,099
DenP	0,401	0,431	0,170	0,279	0,138	0,379	0,807	1,000	0,270	0,102	0,549	0,474	0,534	0,460	0,446	0,001	0,013	0,448	0,520	0,442	0,518	0,054	0,545	0,383	0,073	0,073	0,066
Dur	0,258	0,058	0,095	0,544	0,259	0,437	0,481	0,270	1,000	0,328	0,362	0,026	0,541	0,687	0,006	0,280	0,247	0,135	0,011	0,190	0,176	0,047	0,604	0,678	0,118	0,121	0,105
FBT	0,083	0,117	0,154	0,235	0,399	0,262	0,163	0,102	0,328	1,000	0,098	0,052	0,089	0,256	0,083	0,111	0,129	0,355	0,048	0,078	0,026	0,033	0,144	0,265	0,037	0,037	0,033
GpM	0,625	0,546	0,353	0,455	0,110	0,472	0,587	0,549	0,362	0,098	1,000	0,579	0,869	0,717	0,537	0,023	0,020	0,751	0,642	0,735	0,640	0,279	0,837	0,519	0,142	0,142	0,130
GSP	0,422	0,625	0,136	0,106	0,102	0,190	0,299	0,474	0,026	0,052	0,579	1,000	0,563	0,272	0,684	0,132	0,146	0,705	0,843	0,466	0,625	0,091	0,396	0,222	0,044	0,043	0,042
GSpm	0,584	0,478	0,352	0,532	0,218	0,529	0,603	0,534	0,541	0,089	0,869	0,563	1,000	0,792	0,467	0,112	0,074	0,534	0,523	0,632	0,577	0,251	0,836	0,631	0,138	0,139	0,124
HDpM	0,560	0,468	0,327	0,588	0,338	0,586	0,609	0,460	0,687	0,256	0,717	0,272	0,792	1,000	0,462	0,167	0,122	0,212	0,297	0,448	0,405	0,149	0,854	0,680	0,147	0,149	0,132
HDP	0,516	0,843	0,184	0,137	0,238	0,198	0,312	0,446	0,006	0,083	0,537	0,684	0,467	0,462	1,000	0,147	0,172	0,528	0,646	0,398	0,499	0,053	0,432	0,191	0,068	0,068	0,064
HHpM	0,023	0,114	0,173	0,307	0,202	0,235	0,079	0,001	0,280	0,111	0,023	0,132	0,112	0,167	0,147	1,000	0,687	0,201	0,207	0,010	0,066	0,105	0,155	0,298	0,026	0,027	0,023
HHP	0,010	0,134	0,152	0,270	0,224	0,266	0,071	0,013	0,247	0,129	0,020	0,146	0,074	0,122	0,172	0,687	1,000	0,217	0,236	0,026	0,081	0,060	0,149	0,305	0,022	0,023	0,018
LHpM	0,377	0,472	0,110	0,034	0,126	0,139	0,332	0,448	0,135	0,355	0,751	0,705	0,534	0,212	0,528	0,201	0,217	1,000	0,800	0,626	0,627	0,116	0,442	0,124	0,075	0,074	0,071
LHP	0,413	0,586	0,105	0,085	0,028	0,197	0,387	0,520	0,011	0,048	0,642	0,843	0,523	0,297	0,646	0,207	0,236	0,800	1,000	0,516	0,669	0,064	0,444	0,209	0,048	0,047	0,045
TD	0,487	0,401	0,285	0,332	0,051	0,346	0,453	0,442	0,190	0,078	0,735	0,466	0,632	0,448	0,398	0,010	0,026	0,626	0,516	1,000	0,727	0,307	0,596	0,343	0,083	0,084	0,075
TDP	0,415	0,487	0,128	0,216	0,071	0,323	0,455	0,518	0,176	0,026	0,640	0,625	0,577	0,405	0,499	0,066	0,081	0,627	0,669	0,727	1,000	0,081	0,535	0,342	0,069	0,069	0,065
WP	0,315	0,056	0,614	0,344	0,056	0,001	0,072	0,054	0,047	0,033	0,279	0,091	0,251	0,149	0,053	0,105	0,060	0,116	0,064	0,307	0,081	1,000	0,200	0,117	0,046	0,047	0,041
XpM	0,567	0,466	0,365	0,602	0,302	0,699	0,679	0,545	0,604	0,144	0,837	0,396	0,836	0,854	0,432	0,155	0,149	0,442	0,444	0,596	0,535	0,200	1,000	0,767	0,140	0,142	0,123
Lea	0,338	0,222	0,287	0,592	0,414	0,692	0,536	0,383	0,678	0,265	0,519	0,222	0,631	0,680	0,191	0,298	0,305	0,124	0,209	0,343	0,342	0,117	0,767	1,000	0,099	0,102	0,086
Vot	0,099	0,073	0,059	0,093	0,044	0,087	0,109	0,073	0,118	0,037	0,142	0,044	0,138	0,147	0,068	0,026	0,022	0,075	0,048	0,083	0,069	0,046	0,140	0,099	1,000	0,982	0,948
PV	0,097	0,072	0,059	0,093	0,045	0,089	0,110	0,073	0,121	0,037	0,142	0,043	0,139	0,149	0,068	0,027	0,023	0,074	0,047	0,084	0,069	0,047	0,142	0,102	0,982	1,000	0,871
NV	0,093	0,068	0,053	0,084	0,037	0,077	0,099	0,066	0,105	0,033	0,130	0,042	0,124	0,132	0,064	0,023	0,018	0,071	0,045	0,075	0,065	0,041	0,123	0,086	0,948	0,871	1,000

Figura 27 – Matriz de Coeficientes de Correlação Pearson

KD	KP	KB	AD	AP	DP	Den	DenP	Dur	FBT	GpM	GSP	GSpM	HDpM	HDP	HHpM	HHP	LHpM	LHP	TD	TDP	WP	XpM	Lea	Vot	PV	NV	
KD	1,000	0,639	0,475	0,675	0,086	0,554	0,535	0,481	0,319	0,117	0,718	0,527	0,733	0,705	0,576	0,153	0,055	0,495	0,474	0,615	0,499	0,334	0,707	0,359	0,168	0,185	0,116
KP	0,639	1,000	0,134	0,153	0,018	0,243	0,375	0,445	0,076	0,027	0,585	0,675	0,552	0,603	0,850	0,088	0,126	0,537	0,598	0,450	0,503	0,058	0,548	0,145	0,078	0,084	0,062
KB	0,475	0,134	1,000	0,530	0,178	0,111	0,194	0,132	0,127	0,016	0,351	0,091	0,369	0,314	0,110	0,211	0,142	0,118	0,070	0,321	0,118	0,706	0,339	0,240	0,098	0,104	0,075
AD	0,675	0,153	0,530	1,000	0,394	0,650	0,493	0,301	0,444	0,169	0,453	0,118	0,538	0,558	0,139	0,451	0,314	0,108	0,088	0,414	0,239	0,357	0,568	0,527	0,186	0,214	0,105
AP	0,086	0,018	0,178	0,394	1,000	0,195	0,114	0,036	0,141	0,039	0,009	0,061	0,069	0,218	0,121	0,285	0,238	0,149	0,106	0,020	0,017	0,063	0,144	0,276	0,045	0,064	0,011
DP	0,554	0,243	0,111	0,650	0,195	1,000	0,537	0,384	0,298	0,196	0,459	0,255	0,524	0,498	0,179	0,324	0,240	0,249	0,235	0,423	0,341	0,001	0,603	0,492	0,132	0,158	0,052
Den	0,535	0,375	0,194	0,493	0,114	0,537	1,000	0,843	0,388	0,181	0,614	0,384	0,659	0,653	0,359	0,245	0,142	0,431	0,414	0,536	0,489	0,075	0,702	0,487	0,193	0,218	0,109
DenP	0,481	0,445	0,132	0,301	0,036	0,384	0,843	1,000	0,232	0,107	0,563	0,514	0,568	0,515	0,460	0,082	0,018	0,520	0,534	0,501	0,529	0,056	0,562	0,297	0,114	0,131	0,063
Dur	0,319	0,076	0,127	0,444	0,141	0,298	0,388	0,232	1,000	0,087	0,290	0,037	0,406	0,471	0,053	0,300	0,185	0,022	0,037	0,174	0,169	0,063	0,342	0,368	0,206	0,236	0,139
FBT	0,117	0,027	0,016	0,169	0,039	0,196	0,181	0,107	0,087	1,000	0,225	0,013	0,166	0,093	0,037	0,112	0,080	0,179	0,016	0,222	0,116	0,025	0,187	0,201	0,023	0,036	0,043
GpM	0,718	0,585	0,351	0,453	0,009	0,459	0,614	0,563	0,290	0,225	1,000	0,658	0,912	0,760	0,577	0,073	0,017	0,787	0,674	0,794	0,660	0,301	0,854	0,377	0,199	0,218	0,134
GSP	0,527	0,675	0,091	0,118	0,061	0,255	0,384	0,514	0,037	0,013	0,658	1,000	0,654	0,445	0,705	0,154	0,175	0,785	0,876	0,541	0,657	0,075	0,517	0,102	0,051	0,055	0,042
GSpM	0,733	0,552	0,369	0,538	0,069	0,524	0,659	0,568	0,406	0,166	0,912	0,654	1,000	0,804	0,535	0,156	0,053	0,659	0,604	0,738	0,625	0,290	0,864	0,452	0,210	0,235	0,134
HDpM	0,705	0,603	0,314	0,558	0,218	0,498	0,653	0,515	0,471	0,093	0,760	0,445	0,804	1,000	0,636	0,223	0,095	0,416	0,418	0,572	0,483	0,185	0,866	0,499	0,233	0,260	0,151
HDP	0,576	0,850	0,110	0,139	0,121	0,179	0,359	0,460	0,053	0,037	0,577	0,705	0,535	0,636	1,000	0,135	0,174	0,581	0,652	0,445	0,513	0,058	0,525	0,110	0,070	0,076	0,060
HHpM	0,153	0,088	0,211	0,451	0,285	0,324	0,245	0,082	0,300	0,112	0,073	0,154	0,156	0,223	0,135	1,000	0,823	0,178	0,189	0,108	0,003	0,113	0,227	0,400	0,135	0,153	0,072
HHP	0,055	0,126	0,142	0,314	0,238	0,240	0,142	0,018	0,185	0,080	0,017	0,175	0,053	0,095	0,174	0,823	1,000	0,199	0,220	0,030	0,050	0,063	0,116	0,297	0,088	0,099	0,042
LHpM	0,495	0,537	0,118	0,108	0,149	0,249	0,431	0,520	0,022	0,179	0,787	0,785	0,659	0,416	0,581	0,178	0,199	1,000	0,867	0,672	0,678	0,125	0,577	0,110	0,111	0,112	0,090
LHP	0,474	0,598	0,070	0,088	0,106	0,235	0,414	0,534	0,037	0,016	0,674	0,876	0,604	0,418	0,652	0,189	0,220	0,867	1,000	0,559	0,681	0,059	0,520	0,093	0,049	0,052	0,041
TD	0,615	0,450	0,321	0,414	0,020	0,423	0,536	0,501	0,174	0,222	0,794	0,541	0,738	0,572	0,445	0,108	0,030	0,672	0,559	1,000	0,768	0,324	0,686	0,341	0,144	0,160	0,086
TDP	0,499	0,503	0,118	0,239	0,017	0,341	0,489	0,529	0,169	0,116	0,660	0,657	0,625	0,483	0,513	0,003	0,050	0,678	0,681	0,768	1,000	0,085	0,567	0,251	0,106	0,118	0,064
WP	0,334	0,058	0,706	0,357	0,063	0,001	0,075	0,056	0,063	0,025	0,301	0,075	0,290	0,185	0,058	0,113	0,063	0,125	0,059	0,324	0,085	1,000	0,232	0,126	0,067	0,071	0,059
XpM	0,707	0,548	0,339	0,568	0,144	0,603	0,702	0,562	0,342	0,187	0,854	0,517	0,864	0,866	0,525	0,227	0,116	0,577	0,520	0,686	0,567	0,232	1,000	0,544	0,220	0,247	0,131
Lea	0,359	0,145	0,240	0,527	0,276	0,492	0,487	0,297	0,368	0,201	0,377	0,102	0,452	0,499	0,110	0,400	0,297	0,110	0,093	0,341	0,251	0,126	0,544	1,000	0,152	0,183	0,058
Vot	0,168	0,078	0,098	0,186	0,045	0,132	0,193	0,114	0,206	0,023	0,199	0,051	0,210	0,233	0,070	0,135	0,088	0,111	0,049	0,144	0,106	0,067	0,220	0,152	1,000	0,902	0,752
PV	0,185	0,084	0,104	0,214	0,064	0,158	0,218	0,131	0,236	0,036	0,218	0,055	0,235	0,260	0,076	0,153	0,099	0,112	0,052	0,160	0,118	0,071	0,247	0,183	0,902	1,000	0,506
NV	0,116	0,062	0,075	0,105	0,011	0,052	0,109	0,063	0,139	0,043	0,134	0,042	0,134	0,151	0,060	0,072	0,042	0,090	0,041	0,086	0,064	0,059	0,131	0,058	0,752	0,506	1,000

Figura 28 – Matriz de Coeficientes de Correlação Spearman



# APÊNDICE B – Todos os Componentes do Mapa ESOM

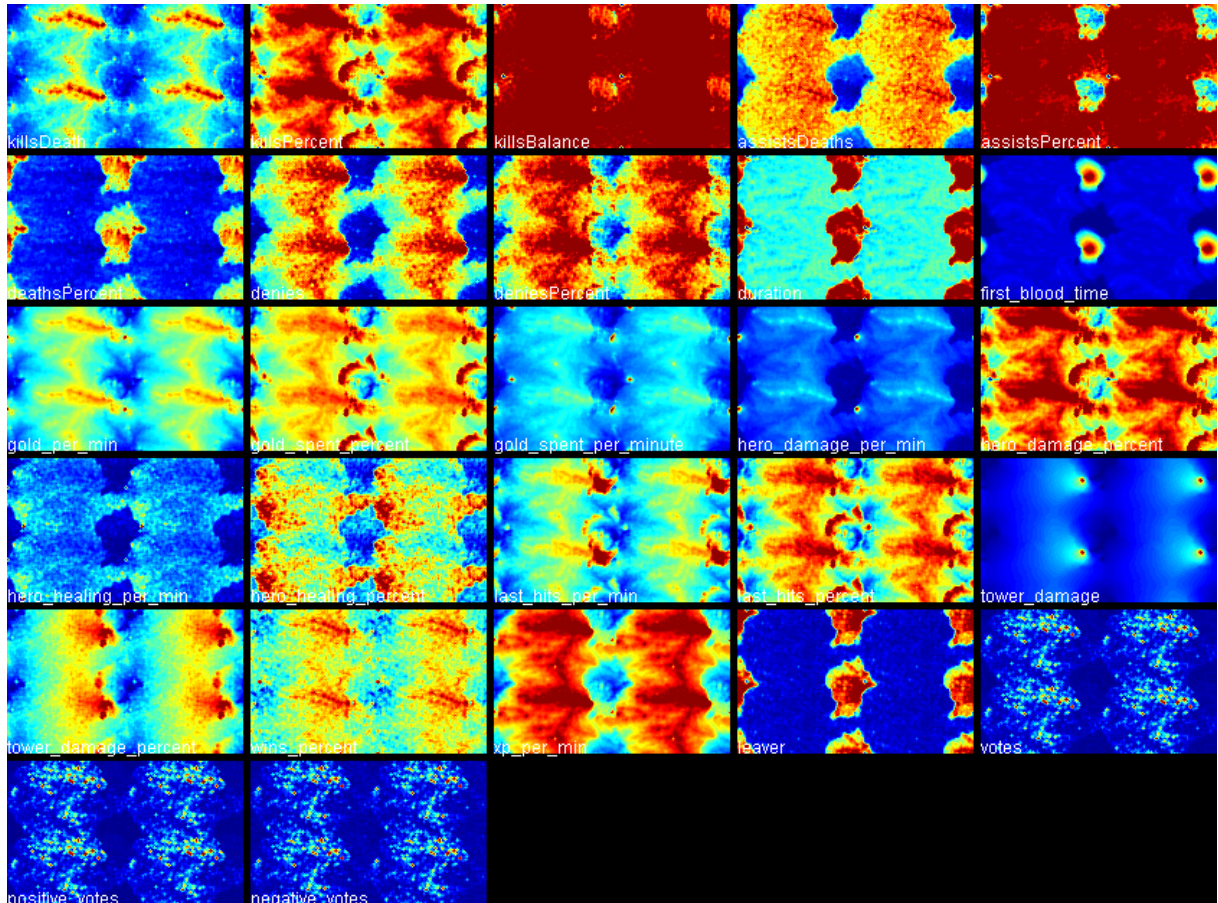


Figura 29 – Todos os Componentes do Mapa ESOM, Antes da Eliminação de Atributos

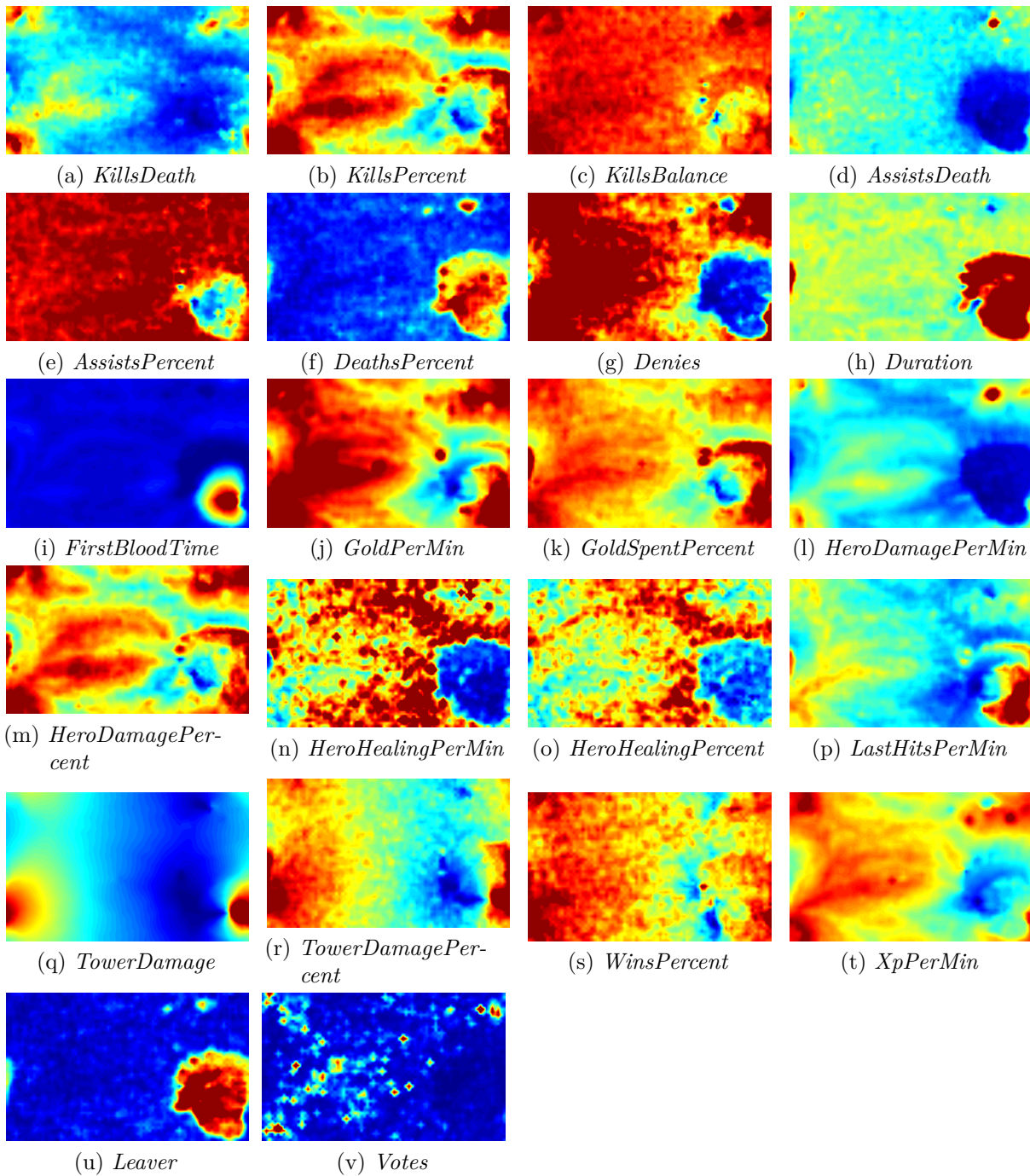


Figura 30 – Influência dos Pesos dos Neurônios para cada Componente do Vetor de Características de Usuário na Composição final do SOM



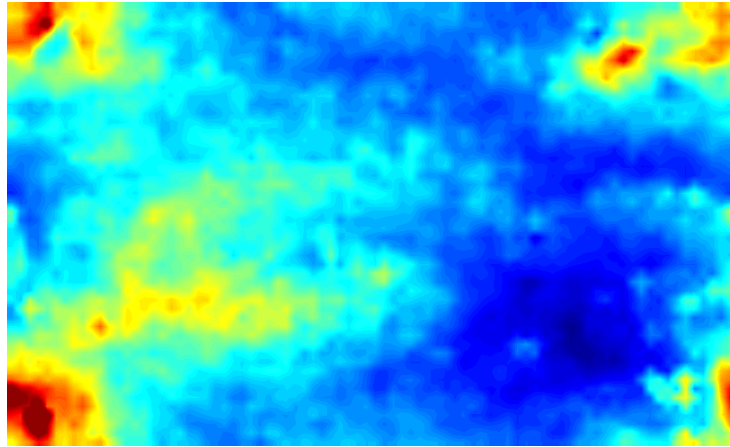


Figura 31 – Relação de Assassinatos sobre Mortes (*KillsDeaths*)

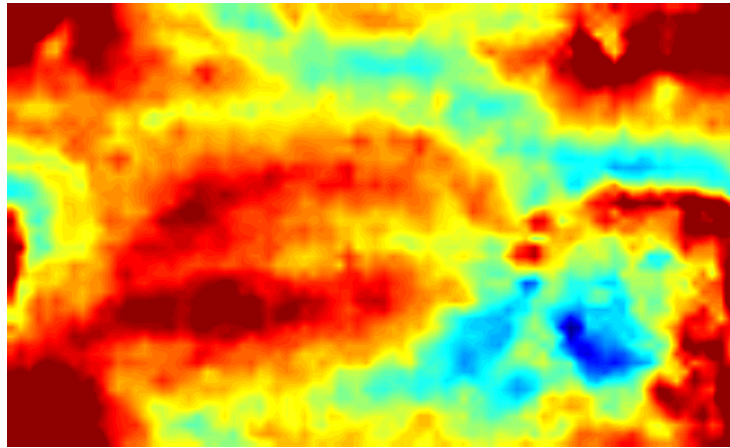


Figura 32 – Relação de Assassinatos em Relação ao Time (*KillsPercent*)

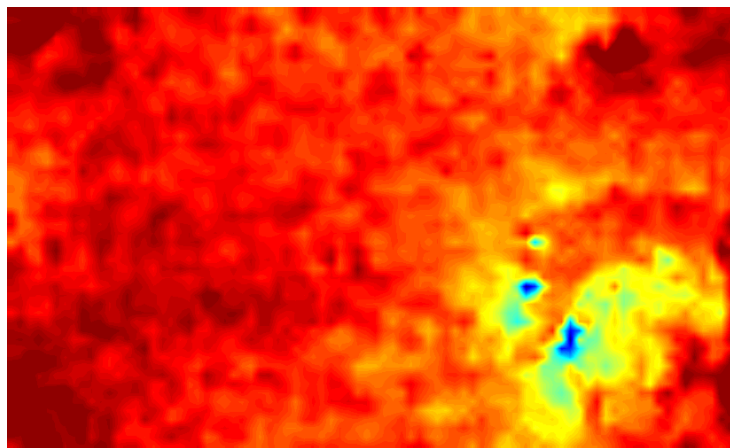


Figura 33 – Relação de Total de Assassinatos em Relação ao outro (*KillsBalance*)

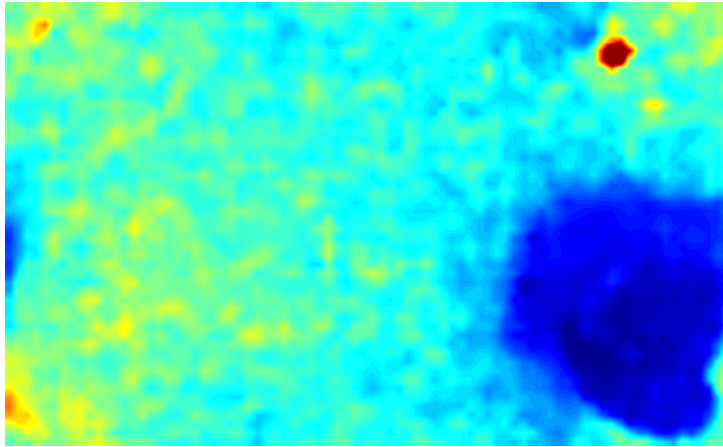


Figura 34 – Relação de Assistências sobre Mortes (*AssistsDeath*)

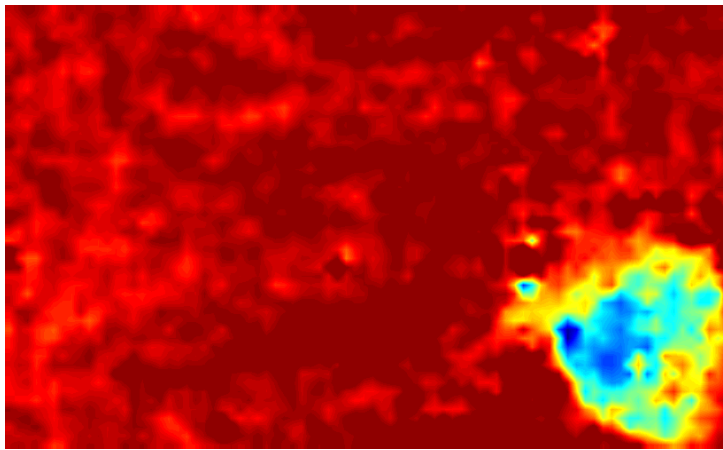


Figura 35 – Relação de Assistências sobre Total de Assistências do Time (*AssistsPercent*)

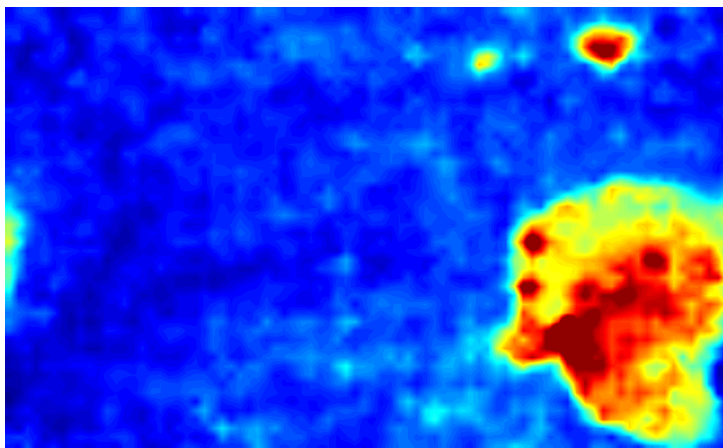


Figura 36 – Relação de Mortes sobre Total de Mortes do Time (*DeathsPercent*)

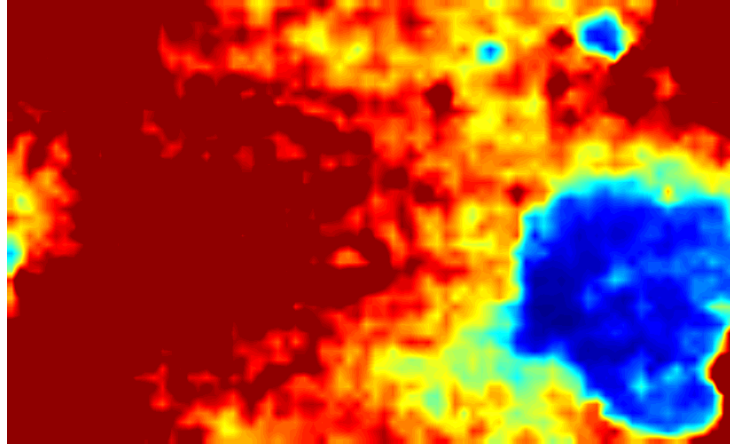


Figura 37 – Quantidade de Negações (*Denies*)

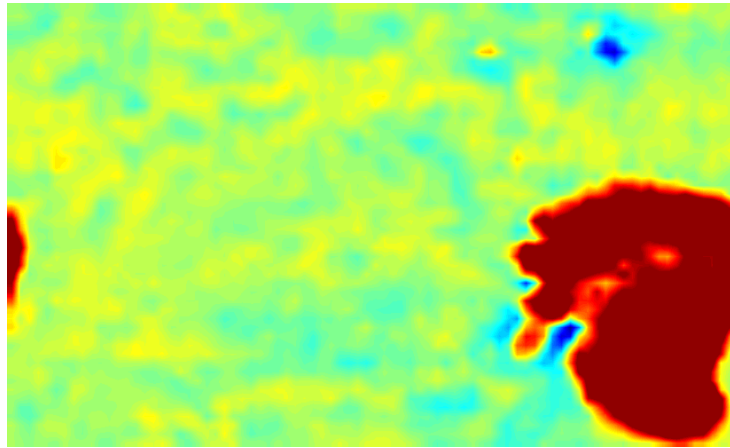


Figura 38 – Duração da Partida (*Duration*)

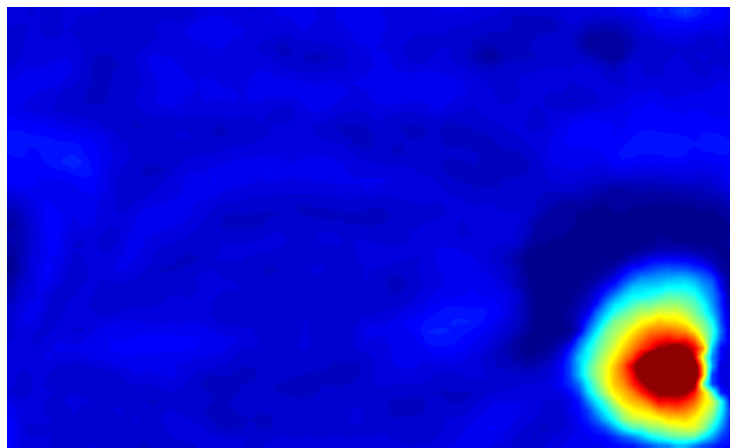


Figura 39 – Horário do Evento do Primeiro Assassinato (*FirstBloodTime*)

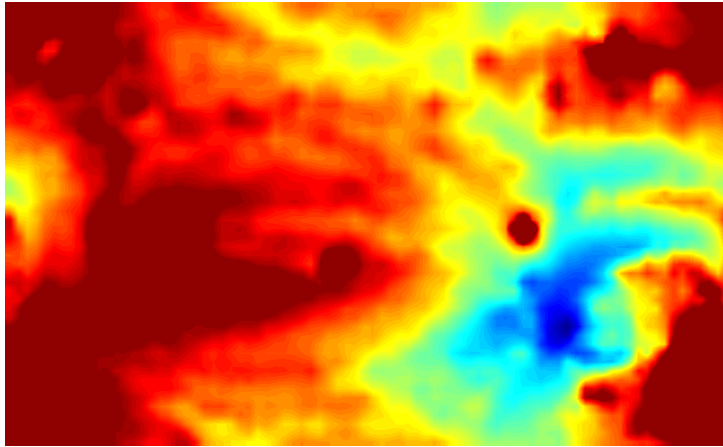


Figura 40 – Relação de Ganho de *Gold* por Minuto (*GoldPerMinut*)

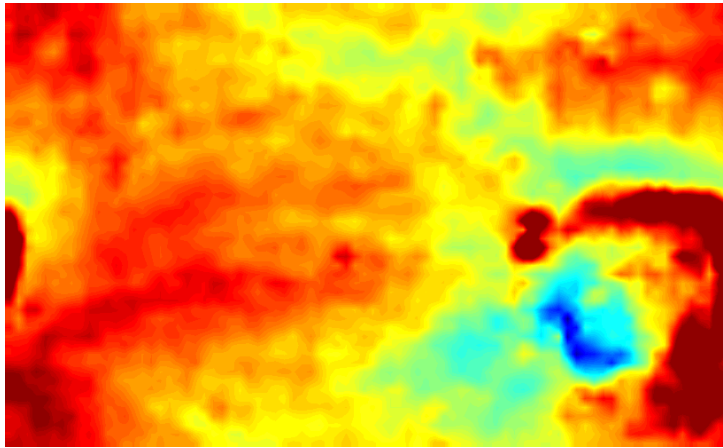


Figura 41 – Relação entre o *Gold* gasto em Relação ao Total da Equipe (*GoldSpentPercent*)

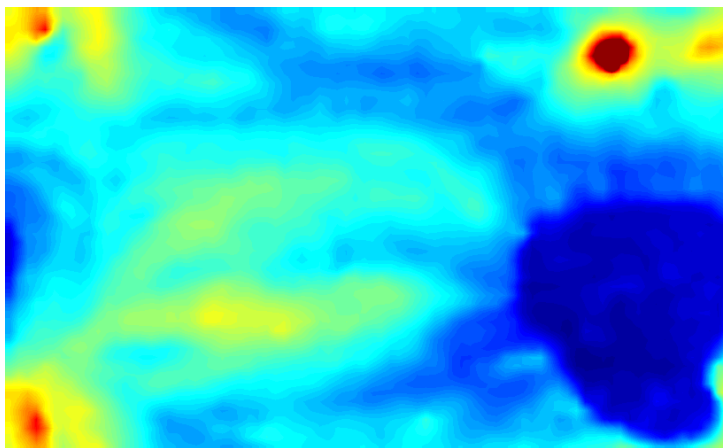


Figura 42 – Quantidade da Dano causado a Oponentes por Minuto (*HeroDamagePerMin*)

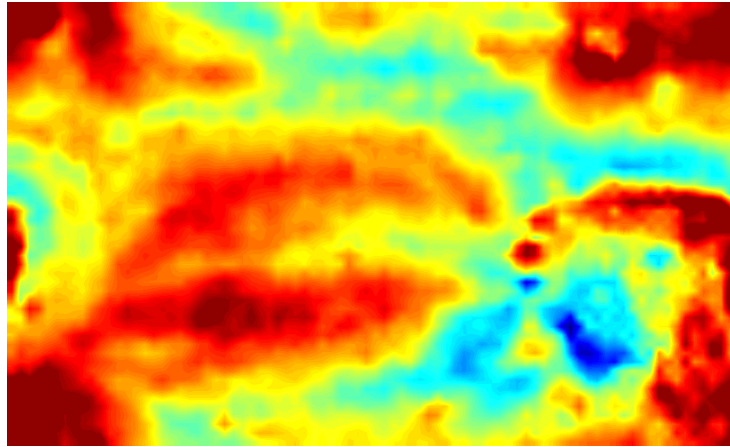


Figura 43 – Quantidade da Dano causado a Oponentes sobre o Total feito pela Equipe (*HeroDamagePercent*)

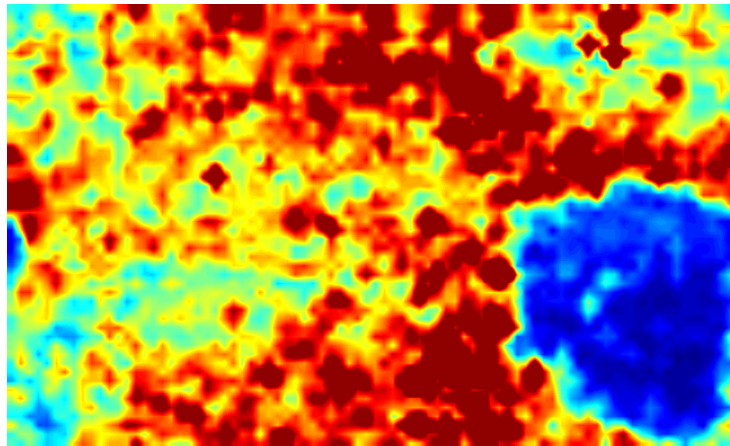


Figura 44 – Quantidade de Cura a Aliados por Minuto (*HeroHealingPerMin*)

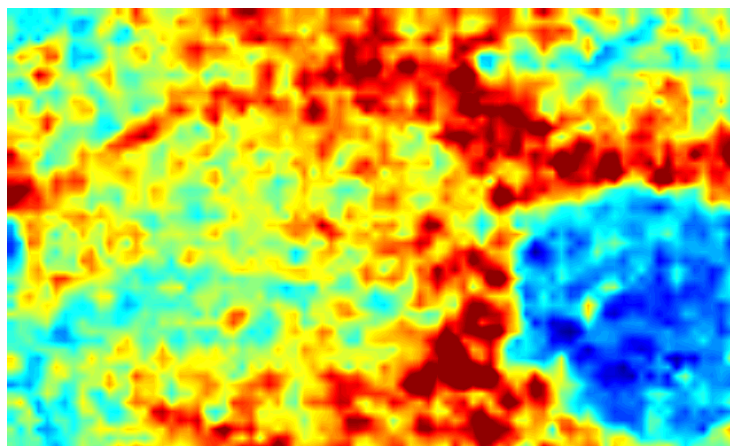


Figura 45 – Quantidade entre a Quantidade de Cura a Aliados Sobre o Total feito pela Equipe (*HeroHealingPercent*)

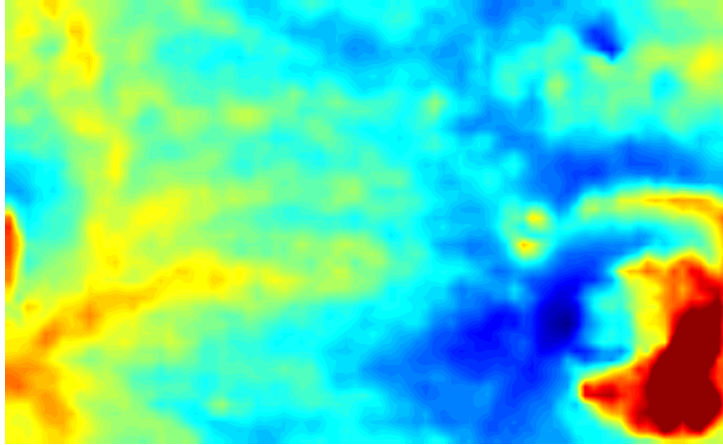


Figura 46 – Quantidade de Eliminação de *Creeps* por Minuto (*LastHitsPerMin*)

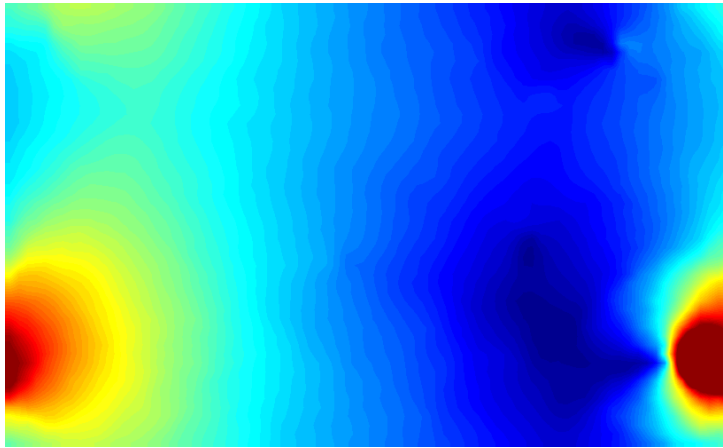


Figura 47 – Quantidade de Dano feito a Estruturas (*TowerDamage*)

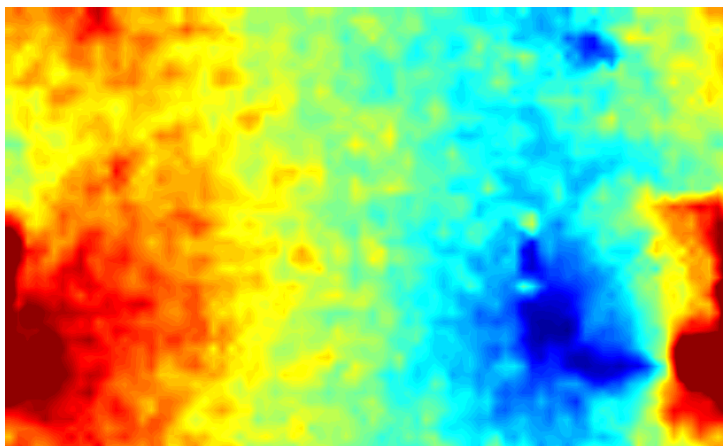


Figura 48 – Quantidade de Dano feito a Estruturas em relação ao Total feito pela Equipe (*TowerDamagePercent*)

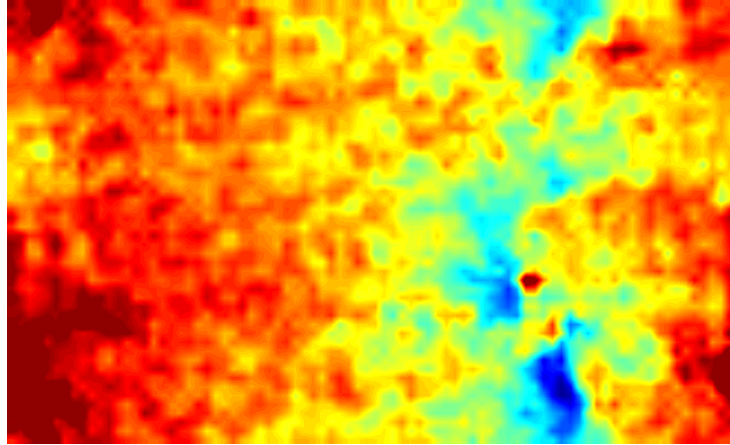


Figura 49 – Relação entre Vitórias e Derrotas (*WinsPercent*)

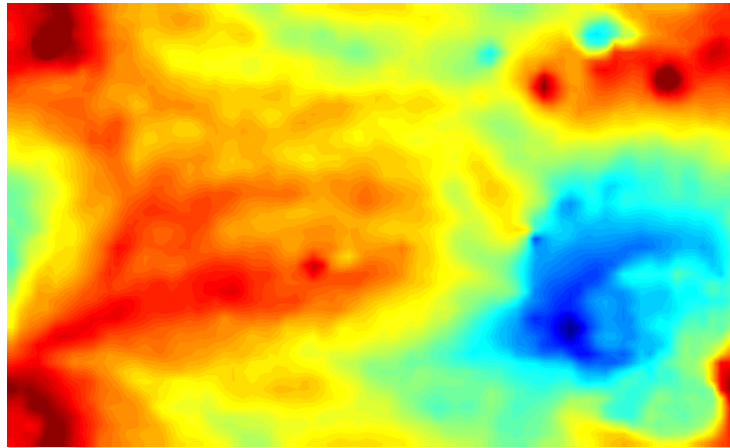


Figura 50 – Quantidade de Experiência Ganha por Minuto (*XpPerMin*)

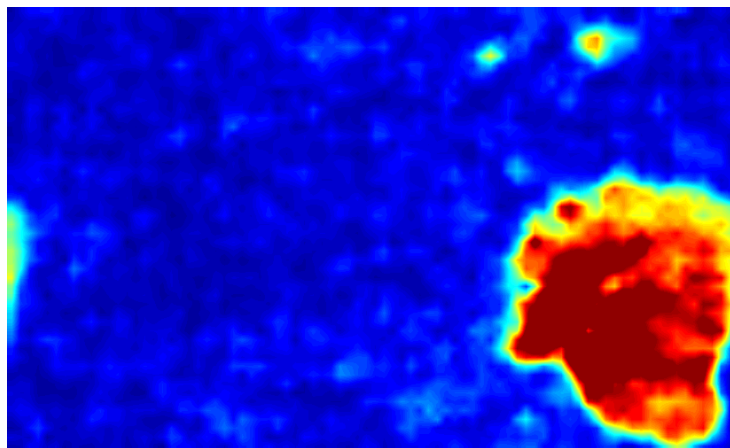


Figura 51 – Abandono de Partidas (*Leaver*)

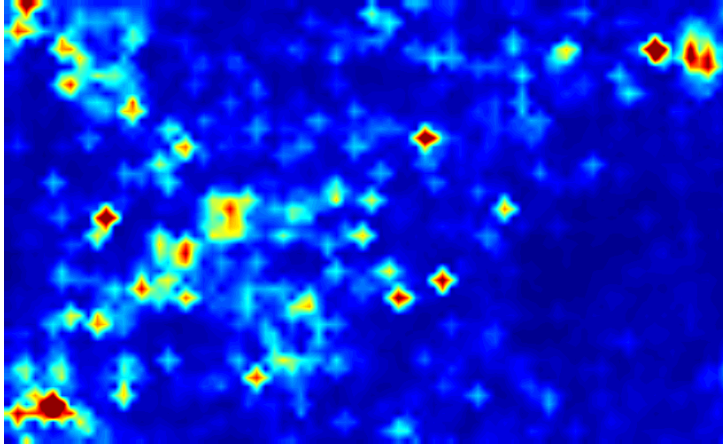


Figura 52 – Votos recebidos, positivos e negativos (*Votes*)



# APÊNDICE C – Matrizes e Tabelas das Características Médias de Agrupamentos de Usuários

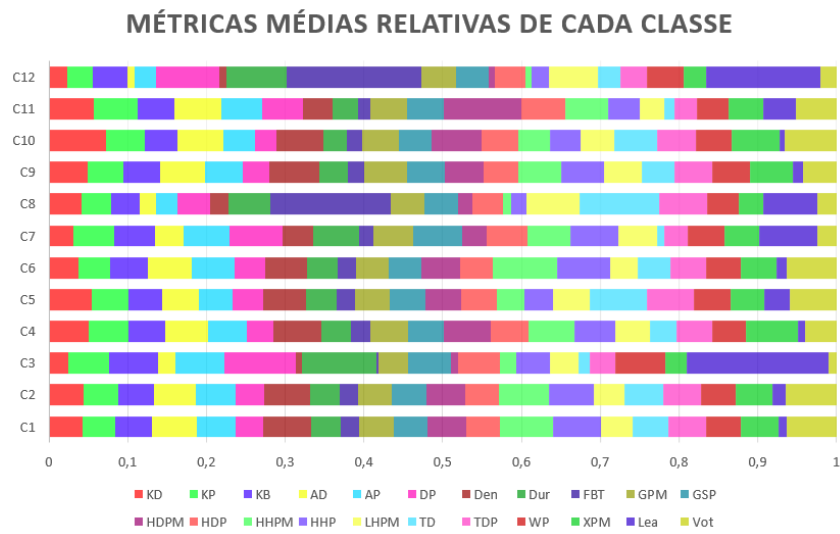


Figura 53 – Distribuição Relativa de Métricas por Classes

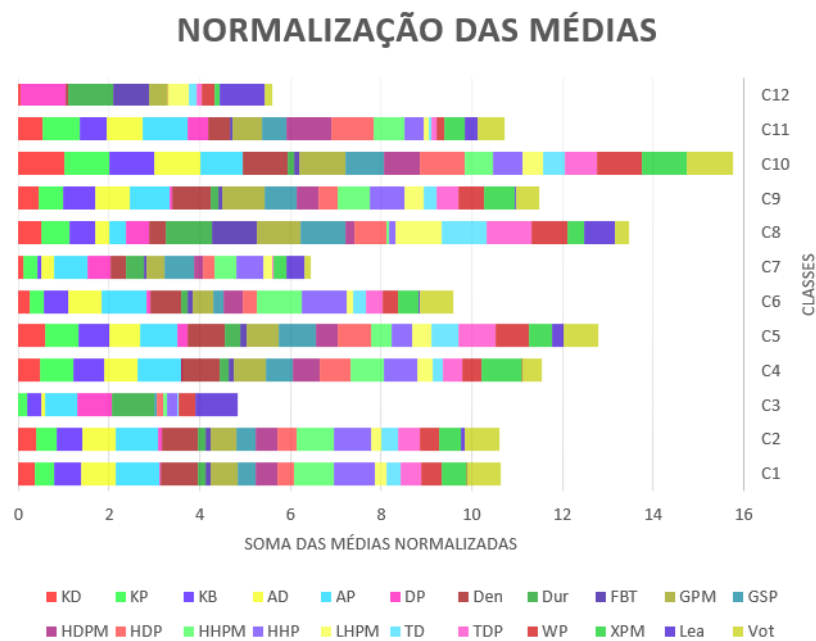


Figura 54 – Normalização das Médias

Tabela 5 – Média de Métricas Absolutas por Grupo

	KD	KP	KB	AD	AP	DP	Den	Dur	FBT	GPM	GSP	HDPM	HDP	HHPM	HHP	LHPM	TD	TDP	WP	XPM	Lea	Vot
C1	1,5362	0,1975	0,5031	2,4554	0,1994	0,2092	5,7843	37,8872	140,9600	389,4365	0,2010	275,9558	0,1968	10,6487	0,1639	2,8470	1080,0302	0,1909	0,5057	447,7465	0,0474	0,1867
C2	1,6114	0,2035	0,4971	2,3470	0,1979	0,2218	5,5637	38,0557	139,5764	381,6158	0,2022	276,9668	0,2046	9,9918	0,1533	2,7785	1167,9790	0,1965	0,5019	426,3836	0,0784	0,1920
C3	0,6166	0,1618	0,4600	0,6479	0,1695	0,3707	0,4645	65,0816	6,9887	225,3811	0,1725	34,8655	0,1726	2,1552	0,0778	1,7467	238,2741	0,0871	0,4888	173,2808	0,5649	0,0197
C4	1,8664	0,2398	0,5133	2,3942	0,2003	0,2074	5,9051	38,4219	150,4378	421,7040	0,2169	337,6540	0,2309	9,1970	0,1415	3,1468	846,8198	0,1800	0,5066	615,8685	0,0461	0,1181
C5	2,1621	0,2404	0,5131	2,2494	0,1848	0,2554	5,7436	43,8099	153,0792	424,7549	0,2352	274,8843	0,2363	6,1126	0,1072	3,6425	1905,2455	0,2668	0,5833	433,0774	0,1670	0,1915
C6	1,2898	0,1817	0,4906	2,3541	0,2046	0,2204	4,9079	36,5802	134,4195	352,9027	0,1860	260,4963	0,1891	12,0532	0,1744	2,3444	943,1856	0,1748	0,4799	308,5242	0,0574	0,1796
C7	0,8539	0,1832	0,4248	1,1893	0,1762	0,3091	2,7977	44,8207	78,7924	338,5892	0,2214	130,0381	0,1852	6,3203	0,1219	2,6881	160,6411	0,0923	0,3991	310,5472	0,2569	0,0550
C8	1,8999	0,2266	0,4985	1,1924	0,1347	0,3145	2,9675	66,9908	1160,6086	491,9666	0,2496	134,5468	0,2330	1,8819	0,0678	6,0693	3103,5747	0,3125	0,5878	363,6105	0,4056	0,0903
C9	1,8027	0,2119	0,5173	2,5101	0,1911	0,2097	6,0537	37,3204	127,0593	486,5883	0,2247	282,2254	0,2035	8,6797	0,1476	3,4787	994,8988	0,1971	0,5381	520,3161	0,0564	0,1279
C10	3,2538	0,2802	0,5588	3,1256	0,1970	0,2030	6,9965	36,7873	141,2479	503,7493	0,2379	440,1350	0,2647	7,8443	0,1291	3,7099	1595,1528	0,2448	0,6408	687,0276	0,0370	0,2404
C11	2,0366	0,2526	0,4996	2,5899	0,2012	0,3034	3,5652	31,4217	91,5429	408,6500	0,2116	541,9301	0,2572	8,3747	0,1061	2,2088	304,1625	0,1136	0,4413	405,2837	0,1883	0,1494
C12	0,7233	0,1363	0,4144	0,3557	0,0941	0,4217	0,8055	66,8765	900,7382	337,6790	0,1688	39,2704	0,1573	1,2104	0,0503	3,8471	618,5612	0,1165	0,4665	229,5233	0,5908	0,0517

Tabela 6 – Percentual de Média de Métricas por Grupo

	KD	KP	KB	AD	AP	DP	Den	Dur	FBT	GPM	GSP	HDPM	HDP	HHPM	HHP	LHPM	TD	TDP	WP	XPM	Lea	Vot
C1	0,3487	0,4256	0,6143	0,7580	0,9522	0,0283	0,8144	0,1818	0,1161	0,5893	0,3986	0,4755	0,3683	0,8705	0,9151	0,2546	0,3124	0,4606	0,4408	0,5342	0,0188	0,7570
C2	0,3772	0,4668	0,5729	0,7189	0,9386	0,0862	0,7807	0,1865	0,1149	0,5613	0,4132	0,4775	0,4402	0,8099	0,8298	0,2387	0,3423	0,4856	0,4253	0,4927	0,0749	0,7805
C3	0,0000	0,1772	0,3160	0,1055	0,6820	0,7668	0,0000	0,9463	0,0000	0,0000	0,0462	0,0000	0,1426	0,0871	0,2218	0,0000	0,0264	0,0000	0,3711	0,0000	0,9532	0,0000
C4	0,4739	0,7196	0,6849	0,7359	0,9605	0,0204	0,8329	0,1968	0,1243	0,7053	0,5953	0,5971	0,6850	0,7366	0,7349	0,3239	0,2332	0,4125	0,4446	0,8615	0,0165	0,4461
C5	0,5860	0,7236	0,6839	0,6837	0,8203	0,2398	0,8082	0,3483	0,1266	0,7162	0,8217	0,4733	0,7356	0,4521	0,4579	0,4386	0,5928	0,7973	0,7619	0,5057	0,2347	0,7783
C6	0,2553	0,3153	0,5278	0,7215	1,0000	0,0795	0,6803	0,1450	0,1105	0,4581	0,2125	0,4450	0,2959	1,0000	1,0000	0,1383	0,2659	0,3891	0,3341	0,4384	0,0369	0,7243
C7	0,0900	0,3258	0,0718	0,3009	0,7430	0,4854	0,3572	0,3767	0,0622	0,4067	0,6512	0,1877	0,2603	0,4897	0,5765	0,2178	0,0000	0,0233	0,0000	0,2672	0,3971	0,1601
C8	0,4866	0,6273	0,5823	0,3021	0,3675	0,5102	0,3832	1,0000	1,0000	0,9577	1,0000	0,1966	0,7051	0,0619	0,1406	1,0000	1,0000	0,7805	0,7805	0,3705	0,6657	0,3201
C9	0,4498	0,5255	0,7125	0,7778	0,8776	0,0307	0,8557	0,1658	0,1041	0,9384	0,6919	0,4878	0,4302	0,6889	0,7841	0,4007	0,2835	0,4880	0,5750	0,6755	0,0350	0,4903
C10	1,0000	1,0000	1,0000	1,0000	0,9310	0,0000	1,0000	0,1508	0,1164	1,0000	0,8562	0,7992	1,0000	0,6118	0,6345	0,4542	0,4874	0,6997	1,0000	1,0000	0,0000	1,0000
C11	0,5384	0,8083	0,5905	0,8066	0,9691	0,4594	0,4747	0,0000	0,0733	0,6584	0,5296	1,0000	0,9296	0,6607	0,4493	0,1069	0,0488	0,1176	0,1745	0,4516	0,2733	0,5877
C12	0,0405	0,0000	0,0000	0,0000	0,0000	1,0000	0,0522	0,9968	0,7747	0,4034	0,0000	0,0087	0,0000	0,0000	0,0000	0,4859	0,1556	0,1307	0,2787	0,1095	1,0000	0,1450

Tabela 7 – Média Padronizada de Métricas por Grupo

	KD	KP	KB	AD	AP	DP	Den	Dur	FBT	GPM	GSP	HDPM	HDP	HHPM	HHP	LHPM	TD	TDP	WP	XPM	Lea	Vot
C1	0,3487	0,4256	0,6143	0,7580	0,9522	0,0283	0,8144	0,1818	0,1161	0,5893	0,3986	0,4755	0,3683	0,8705	0,9151	0,2546	0,3124	0,4606	0,4408	0,5342	0,0188	0,7570
C2	0,3772	0,4668	0,5729	0,7189	0,9386	0,0862	0,7807	0,1865	0,1149	0,5613	0,4132	0,4775	0,4402	0,8099	0,8298	0,2387	0,3423	0,4856	0,4253	0,4927	0,0749	0,7805
C3	0,0000	0,1772	0,3160	0,1055	0,6820	0,7668	0,0000	0,9463	0,0000	0,0000	0,0462	0,0000	0,1426	0,0871	0,2218	0,0000	0,0264	0,0000	0,3711	0,0000	0,9532	0,0000
C4	0,4739	0,7196	0,6849	0,7359	0,9605	0,0204	0,8329	0,1968	0,1243	0,7053	0,5953	0,5971	0,6850	0,7366	0,7349	0,3239	0,2332	0,4125	0,4446	0,8615	0,0165	0,4461
C5	0,5860	0,7236	0,6839	0,6837	0,8203	0,2398	0,8082	0,3483	0,1266	0,7162	0,8217	0,4733	0,7356	0,4521	0,4579	0,4386	0,5928	0,7973	0,7619	0,5057	0,2347	0,7783
C6	0,2553	0,3153	0,5278	0,7215	1,0000	0,0795	0,6803	0,1450	0,1105	0,4581	0,2125	0,4450	0,2959	1,0000	1,0000	0,1383	0,2659	0,3891	0,3341	0,4384	0,0369	0,7243
C7	0,0900	0,3258	0,0718	0,3009	0,7430	0,4854	0,3572	0,3767	0,0622	0,4067	0,6512	0,1877	0,2603	0,4897	0,5765	0,2178	0,0000	0,0233	0,0000	0,2672	0,3971	0,1601
C8	0,4866	0,6273	0,5823	0,3021	0,3675	0,5102	0,3832	1,0000	1,0000	0,9577	1,0000	0,1966	0,7051	0,0619	0,1406	1,0000	1,0000	1,0000	0,7805	0,3705	0,6657	0,3201
C9	0,4498	0,5255	0,7125	0,7778	0,8776	0,0307	0,8557	0,1658	0,1041	0,9384	0,6919	0,4878	0,4302	0,6889	0,7841	0,4007	0,2835	0,4880	0,5750	0,6755	0,0350	0,4903
C10	1,0000	1,0000	1,0000	1,0000	0,9310	0,0000	1,0000	0,1508	0,1164	1,0000	0,8562	0,7992	1,0000	0,6118	0,6345	0,4542	0,4874	0,6997	1,0000	1,0000	0,0000	1,0000
C11	0,5384	0,8083	0,5905	0,8066	0,9691	0,4594	0,4747	0,0000	0,0733	0,6584	0,5296	1,0000	0,9296	0,6607	0,4493	0,1069	0,0488	0,1176	0,1745	0,4516	0,2733	0,5877
C12	0,0405	0,0000	0,0000	0,0000	0,0000	1,0000	0,0522	0,9968	0,7747	0,4034	0,0000	0,0087	0,0000	0,0000	0,0000	0,4859	0,1556	0,1307	0,2787	0,1095	1,0000	0,1450