

**UNIVERSIDADE FEDERAL DE ALFENAS
DEPARTAMENTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

Wilson Sasaki Junior

**PREDIÇÃO DE RELACIONAMENTO PROFISSIONAL
CARACTERIZADO ATRAVÉS DE INTERAÇÕES NA REDE
SOCIAL FACEBOOK**

Alfenas, __ de Janeiro de 2014.

UNIVERSIDADE FEDERAL DE ALFENAS
DEPARTAMENTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**PREDIÇÃO DE RELACIONAMENTO PROFISSIONAL
CARACTERIZADO ATRAVÉS DE INTERAÇÕES NA REDE
SOCIAL FACEBOOK**

Wilson Sasaki Junior

Monografia apresentada ao Curso de Bacharelado em
Ciência da Computação da Universidade Federal de
Alfenas como requisito parcial para obtenção do Título
de Bacharel em Ciência da Computação.

Orientador: Prof^a. Dr^a. Maria Regina Martinez
Coorientador: Prof. Douglas Donizeti de Castilho Braz

Alfenas, __ de Janeiro de 2014.

Wilson Sasaki Junior

**PREDIÇÃO DE RELACIONAMENTO PROFISSIONAL
CARACTERIZADO ATRAVÉS DE INTERAÇÕES NA REDE
SOCIAL FACEBOOK**

A Banca examinadora abaixo-assinada aprova a monografia apresentada como parte dos requisitos para obtenção do título de Bacharel em Ciência da Computação pela Universidade Federal de Alfenas.

Prof. Dr. Ricardo Menezes Salgado
Universidade Federal de Alfenas

Prof. Dr. Humberto César Brandão de Oliveira
Universidade Federal de Alfenas

Prof. Douglas Donizeti de Castilho Braz
Universidade Federal de Alfenas

Prof^a. Dr^a. Maria Regina Martinez
Universidade Federal de Alfenas

Alfenas, __ de Janeiro de 2014.

À minha família

AGRADECIMENTO

Agradeço a todos os participantes voluntários da pesquisa, sem os quais esta não teria sido possível.

Agradeço aos meus pais que sem o apoio e a ajuda constante eu jamais teria realizado esse trabalho.

Agradeço a todos que estiveram ao meu lado me apoiando e me motivando constantemente. Agradeço a minha namorada Daniele de Almeida Dias por sempre acreditar em mim e me motivar, sem ela eu não teria alcançado meus objetivos e nem me tornado quem sou hoje. Agradeço ao Júnio César Rosa por se demonstrar um verdadeiro amigo e ter me acompanhado em toda a minha trajetória pela universidade.

Agradeço a todos os meus professores que contribuíram imensamente nessa jornada, tanto de forma intelectual, quanto de forma amigável atendendo sempre a todas as minhas dúvidas e questionamentos e possibilitando um ponto de partida quando eu estava sem rumo. Em especial agradeço a minha orientadora Profa. Dra Maria Regina Martinez e ao meu coorientador Douglas Donizeti de Castilho Braz pela paciência, dedicação e por todo conhecimento confiado a mim no desenvolvimento desse trabalho.

"Temos de nos tornar na mudança que queremos ver."

Mahatma Gandhi

RESUMO

Afeição, neutralidade ou rejeição podem definir os relacionamentos interpessoais que se manifestam a partir de um conjunto de interações que são dinâmicas e complexas. Estes relacionamentos podem ser utilizados para definir preferências sociais, como, por exemplo, parcerias profissionais ou amizades e, assim, criar redes sociais. Apesar redes sociais de amizades e profissionais serem distintas, este trabalho estuda a sua intersecção para encontrar métricas que permitam classificar preferências profissionais como afetivas, neutras ou de rejeição a partir das características sociais de amizades.

Palavras-Chave: Facebook, Redes Sociais, Predição de Link.

ABSTRACT

Affection, neutrality or rejection may define the interpersonal relationships that are manifested from a set of interactions that are dynamic and complex. Relationships are used to define social preferences, such as, professional partnerships or friendships and thus create social networks. Although social and professional are distinct networks, this work studies their intersection to find metrics that allow to classify professional partnerships as affective, neutral or rejection from the social characteristics of friendships.

Keywords: Facebook, Social Network, Link Prediction.

LISTA DE FIGURAS

FIGURA 1 – DESDOBRAMENTOS DA SOCIONOMIA.....	28
FIGURA 2 – GRÁFICO DE BARRAS DA PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM NEGATIVAS E NÃO NEGATIVAS.....	54
FIGURA 3 – GRÁFICO DE BARRAS DA PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM INDIFERENTES E NÃO INDIFERENTES.	55
FIGURA 4 – GRÁFICO DE BARRAS DA PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM POSITIVAS E NÃO POSITIVAS.	56
FIGURA 5 – GRÁFICO DE BARRAS DA PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM NEGATIVAS E INDIFERENTES.....	57
FIGURA 6 – GRÁFICO DE BARRAS DA PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM NEGATIVAS E POSITIVAS.	58
FIGURA 7 – GRÁFICO DE BARRAS DA PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM INDIFERENTES E POSITIVAS.	59

LISTA DE TABELAS

TABELA 1 – PSEUDOCÓDIGO DO APLICATIVO PARA COLETA DE DADOS NO FACEBOOK.....	39
TABELA 2 – EXEMPLO DE MATRIZ DE MAPEAMENTO ENTRADA E SAÍDA.....	43
TABELA 3 - PERCENTUAL DAS ARESTAS NO GRAFO TURMA1.	45
TABELA 4 - PERCENTUAL DAS ARESTAS NO GRAFO TURMA2.	45
TABELA 5 - PERCENTUAL DAS ARESTAS NEGATIVAS E NÃO NEGATIVAS NO GRAFO TURMA1.....	46
TABELA 6 - PERCENTUAL DAS ARESTAS NEGATIVAS E NÃO NEGATIVAS NO GRAFO TURMA2.....	46
TABELA 7 - PERCENTUAL DAS ARESTAS INDIFERENTES E NÃO INDIFERENTES NO GRAFO TURMA1.	46
TABELA 8 - PERCENTUAL DAS ARESTAS INDIFERENTES E NÃO INDIFERENTES NO GRAFO TURMA2.	46
TABELA 9 - PERCENTUAL DAS ARESTAS POSITIVAS E NÃO POSITIVAS NO GRAFO TURMA1.....	47
TABELA 10 - PERCENTUAL DAS ARESTAS POSITIVAS E NÃO POSITIVAS NO GRAFO TURMA2.....	47
TABELA 11 - PERCENTUAL DAS ARESTAS NEGATIVAS E INDIFERENTES NO GRAFO TURMA1.....	47
TABELA 12 - PERCENTUAL DAS ARESTAS NEGATIVAS E INDIFERENTES NO GRAFO TURMA2.....	47
TABELA 13 - PERCENTUAL DAS ARESTAS NEGATIVAS E POSITIVAS NO GRAFO TURMA1.....	48
TABELA 14 - PERCENTUAL DAS ARESTAS NEGATIVAS E POSITIVAS NO GRAFO TURMA2.....	48
TABELA 15 - PERCENTUAL DAS ARESTAS INDIFERENTES E POSITIVAS NO GRAFO TURMA1.	48
TABELA 16 - PERCENTUAL DAS ARESTAS INDIFERENTES E POSITIVAS NO GRAFO TURMA2.	48
TABELA 17 - PERCENTUAL DAS ARESTAS NEGATIVAS E NÃO NEGATIVAS NO GRAFO BALANCEADO TURMA1.....	50
TABELA 18 - PERCENTUAL DAS ARESTAS NEGATIVAS E NÃO NEGATIVAS NO GRAFO BALANCEADO TURMA2.....	50
TABELA 19 - PERCENTUAL DAS ARESTAS POSITIVAS E NÃO POSITIVAS NO GRAFO BALANCEADO TURMA1.....	51
TABELA 20 - PERCENTUAL DAS ARESTAS POSITIVAS E NÃO POSITIVAS NO GRAFO BALANCEADO TURMA2.....	51
TABELA 21 - PERCENTUAL DAS ARESTAS NEGATIVAS E INDIFERENTES NO GRAFO BALANCEADO TURMA1.....	52
TABELA 22 - PERCENTUAL DAS ARESTAS NEGATIVAS E INDIFERENTES NO GRAFO BALANCEADO TURMA2.....	52
TABELA 23 - PERCENTUAL DAS ARESTAS NEGATIVAS E POSITIVAS NO GRAFO BALANCEADO TURMA1.	52
TABELA 24 - PERCENTUAL DAS ARESTAS NEGATIVAS E POSITIVAS NO GRAFO BALANCEADO TURMA2.	53
TABELA 25 – PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM NEGATIVAS E NÃO NEGATIVAS.....	54
TABELA 26 – PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS INDIFERENTES E NÃO INDIFERENTES.....	55
TABELA 27 – PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM POSITIVAS E NÃO POSITIVAS.....	55
TABELA 28 – PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM NEGATIVAS E INDIFERENTES.....	57
TABELA 29 – PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS BALANCEADAS EM NEGATIVAS E POSITIVAS.....	58
TABELA 30 – PORCENTAGEM DE ACERTO NA FASE DE TESTE, COM ARESTAS INDIFERENTES E POSITIVAS.	59
TABELA 31 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS NEGATIVAS E NÃO NEGATIVAS.....	60

TABELA 32 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS NEGATIVAS E INDIFERENTES.	60
TABELA 33 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS NEGATIVAS E POSITIVAS.....	60
TABELA 34 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS INDIFERENTES E POSITIVAS.	61
TABELA 35 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS INDIFERENTES E NÃO INDIFERENTES.	61
TABELA 36 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS POSITIVAS E NÃO POSITIVAS.....	61
TABELA 37 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS NEGATIVAS E NÃO NEGATIVAS.	62
TABELA 38 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS INDIFERENTES E NÃO INDIFERENTES.	62
TABELA 39 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS INDIFERENTES E POSITIVAS.	63
TABELA 40 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS POSITIVAS E NÃO POSITIVAS.....	63
TABELA 41 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS NEGATIVAS E INDIFERENTES.	63
TABELA 42 – ATRIBUTOS PARA O GRAFO COMPOSTO DE ARESTAS NEGATIVAS E POSITIVA.	63

LISTA DE ABREVIATURAS E SIGLAS

MLP	- <i>Multi Layer Perceptron</i>
RF	- <i>Random Forest</i>
DT	- <i>Decision Table</i>
TCLE	- Termo de Consentimento Livre e Esclarecido
DAP	- Declaração de Aceite de Participação
IGAE	- <i>Info Gain Attribute Evaluation</i>
CSAE	- <i>Chi Square Attribute Evaluation</i>
API	- <i>Application Programming Interface</i>
T1T1	- Treinamento com a Turma1 e Teste com a Turma1
T1T2	- Treinamento com a Turma1 e Teste com a Turma2
T2T1	- Treinamento com a Turma2 e Teste com a Turma1
T2T2	- Treinamento com a Turma2 e Teste com a Turma2

SUMÁRIO

1 INTRODUÇÃO	22
1.1 JUSTIFICATIVA E MOTIVAÇÃO.....	23
1.2 PROBLEMATIZAÇÃO.....	24
1.3 OBJETIVOS.....	25
1.3.1 Gerais.....	25
1.3.2 Específicos.....	26
1.4 ORGANIZAÇÃO DA MONOGRAFIA.....	26
2 FUNDAMENTAÇÃO TEÓRICA	27
2.1 REDES SOCIAIS E SUA ANÁLISE.....	27
2.2 MODELOS DE PREDIÇÃO DE COMPORTAMENTO EM REDES SOCIAIS.....	31
2.2.1 Weka.....	33
2.2.1.1 <i>Decision Table</i>	34
2.2.1.2 J48.....	34
2.2.1.3 <i>Multi Layer Perceptron</i>	35
2.2.1.4 <i>Random Forest</i>	35
2.2.1.5 <i>Info Gain & Chi Square Attribute Evaluation</i>	36
3 METODOLOGIA	37
3.1 COLETA DE DADOS.....	37
3.1.1 Aspectos Éticos da Pesquisa.....	38
3.1.2 Facebook.....	38
3.1.3 Identificação do Usuário.....	39
3.1.4 Listando os ID's que Aceitaram em Participar da Pesquisa.....	40
3.1.5 Estado do ID.....	40
3.1.6 Teste Sociométrico.....	41
3.2 PRÉ PROCESSAMENTO DA INFORMAÇÃO.....	41
3.2.1 Transcrição dos Dados.....	42
4 PREVISÃO DE LINKS	43
4.1 ORGANIZAÇÃO DA BASE DE DADOS.....	44
4.1.1 Base de Dados Experimental.....	44
4.1.1.1 Grafo de Arestas Negativas e Não Negativas.....	45
4.1.1.2 Grafo de Arestas Indiferentes e Não Indiferentes.....	46
4.1.1.3 Grafo de Arestas Positivas e Não Positivas.....	46
4.1.1.4 Grafo de Arestas Negativas e Indiferentes.....	47
4.1.1.5 Grafo de Arestas Negativas e Positivas.....	47
4.1.1.6 Grafo de Arestas Indiferentes e Positivas.....	48
4.1.2 Balanceamento dos Grafos.....	48
4.1.2.1 Balanceamento das Arestas Negativas e Não Negativas.....	49
4.1.2.2 Balanceamento das Arestas Positivas e Não Positivas.....	50
4.1.2.3 Balanceamento das Arestas Negativas e Indiferentes.....	51
4.1.2.4 Balanceamento das Arestas Negativas e Positivas.....	52

4.2 RESULTADOS.....	53
4.2.1 Predição de Arestas Negativas e Não Negativas	53
4.2.2 Predição de Arestas Indiferentes e Não Indiferentes	54
4.2.3 Predição de Arestas Positivas e Não Positivas	55
4.2.4 Predição de Arestas Negativas e Indiferentes.....	56
4.2.5 Predição de Arestas Negativas e Positivas	58
4.2.6 Predição de Arestas Indiferentes e Positivas	59
4.3 RANKING DOS ATRIBUTOS DAS ARESTAS	60
4.3.1 Atributos das Arestas do Grafo Turma1	60
4.3.2 Atributos das Arestas do Grafo Turma2	62
5 CONCLUSÕES E TRABALHOS FUTUROS.....	65
6 REFERÊNCIAS BIBLIOGRÁFICAS	66
7 APÊNDICE.....	69
7.1 APÊNDICE I.....	69
7.2 APÊNDICE II.....	70
7.3 APÊNDICE III.....	71

1 Introdução

Redes sociais têm sido o foco de estudo de diversos autores [Garton *et al.*, 1997], [Leskovec *et al.*, 2010], [Oliveira *et al.*, 2010]. Redes sociais são compostas por objetos que interagem de algum modo pré-determinado [Schaefer 2007] e, computacionalmente, redes sociais podem ser representadas como um grafo $G(V, A)$, sendo V o conjunto de vértices que representam os objetos da rede e A o conjunto de arestas que representam a(s) interação(ões) entre os vértices. Várias interações entre dois objetos podem ser caracterizadas como uma rede social, como, por exemplo, redes de roteadores [Furuta *et al.*, 2006], interações proteicas [Pellegrini *et al.*, 2004] e pessoas [Schaeffer, 2007], [Oliveira *et al.*, 2010], [Fortunato, 2010].

Interações sociais na Internet envolvem tanto relações positivas como negativas. Pessoas criam elos positivos para indicar amizade, suporte e/ou confiabilidade, mas também criam elos que significam desconfiança, inimizade ou rejeição. Mesmo que relações positivas e negativas sejam importantes no estudo de redes sociais, a maior parte das pesquisas sobre redes sociais *online* consideram apenas relações positivas [Newman, 2003].

Nos últimos anos têm surgido pesquisas que estudam relações positivas e negativas em redes sociais *online*. Por exemplo, usuários da rede Wikipedia podem votar a favor ou contra outros para nomeação de administrador [Burke & Kraut, 2008], usuários da rede Epinions podem expressar confiança ou desconfiança de outros [Guha *et al.*, 2004], [Massa & Avesani, 2005], e participantes da rede Slashdot podem declarar outros como sendo amigos ou inimigos [Brzozowski *et al.*, 2008], [Kunegis *et al.*, 2009], [Lampe *et al.*, 2007].

Para um dado link em uma rede social, seu sinal pode ser definido como positivo, indiferente ou negativo dependendo do seu comportamento dentro da

rede. Um questionamento a respeito dos tipos de relacionamentos é saber se os tipos de *links* possuem características distintas nas interações entre objetos de modo que seja possível classificar um conjunto de interações como sendo um *link* positivo, indiferente ou negativo. Quando se encontram respostas para este questionamento, elas ajudam na concepção de aplicações computacionais que focam em inferir relacionamentos não existentes em uma rede social [Leskovec *et al.*, 2010].

Um estudo recente [Mansson & Myers, 2011] analisou como os estudantes da *Atlantic University* expressavam afeições do mundo real dentro da rede social Facebook. O estudo detectou 51 tipos de interações entre usuários, dos quais 29 foram considerados importantes, ou seja, caracterizam uma amizade do mundo real através do Facebook.

Como Facebook é a rede social que mais agrega usuários atualmente, para corporações que desenvolvem dinâmicas em grupos, conseguir detectar se duas pessoas quaisquer possuem ou não afinidade, indiferença ou rejeição uma com a outra pode minimizar riscos e aumentar a produtividade do grupo.

1.1 Justificativa e Motivação

A técnica de predição de arestas para uma rede social é possível através das análises das características dos vértices ou das arestas dentro do grafo topológico. Utilizando-se da técnica de predição que podem alcançar grande precisão, experimentos biológicos que possuem alto valor agregado podem ter seus custos reduzidos [Lü & Zhou, 2010]. Outra aplicação relevante é a predição de tendências políticas através das características de campanha e direcionamento político nos Estados Unidos [Lü *et al.*, 2009].

Em contextos empresariais detectar corretamente o relacionamento

positivo, indiferente ou negativo, entre as pessoas pode aumentar a confiabilidade e a satisfação de um grupo de trabalho minimizando, assim, os riscos de falha de concluir uma atividade [Oliveira *et al.*, 2010]. Detectar o tipo de relacionamento entre duas pessoas quaisquer é uma tarefa difícil, pois as dinâmicas de um relacionamento são complexas.

Para encontrar o tipo de relacionamento entre duas pessoas a aplicação do teste sociométrico, desenvolvido por Moreno, pode gerar desconforto para as pessoas e o resultado pode não ser confiável, pois o conhecimento prévio do teste pode permitir que as respostas sejam manipuladas [Figueiredo, 2011].

Neste contexto, o uso de técnica de predição de *link* fornece os relacionamentos internos em um grupo de pessoas, evitando, assim, o emprego do teste sociométrico. A confiabilidade do grafo tipológico seria mais confiável, pois para um usuário manipular informações das redes sociais para favorecer um resultado torna-se uma tarefa muito mais complexa.

1.2 Problematização

Segundo Bustos, 1979, a sociometria é uma ferramenta que permite representar o grau de um relacionamento de forma métrica (positivo quando há afeição, indiferente se não há preferência social ou negativo representando rejeição), bem como sua organização psicossocial. Seu objetivo é identificar padrões de relacionamento bilateral identificando e mapeando vínculos do grupo em forma de sociograma.

Um sociograma é um meio de representar graficamente redes sociais. O Facebook é uma grande rede social que está presente em inúmeros ambientes de pesquisa, desenvolvimento, estudo, lazer, etc. É possível identificar pequenas redes sociais internas que possuem suas próprias características,

como assuntos, gostos, interesses, entre outros, que por sua vez, podem ser representados como um grafo $G(V,A)$, onde uma aresta é composta pela combinação das interações entre pessoas dentro da rede social.

Assim como mencionado em Oliveira [2010], dentro destas pequenas redes internas, os relacionamentos das pessoas são dinâmicos, sendo possível identificar alterações em suas configurações. A manipulação de informações de gostos e interesses, por parte de um indivíduo, tentando se projetar em algum grupo social ou favorecer um relacionamento (positivo, indiferente ou negativo) com outro indivíduo é muito difícil devido à quantidade e complexidade de fluxo de conteúdo e interatividade dentro do Facebook.

Devido a complexidade de fluxo de conteúdo e iteratividade dentro do Facebook, este trabalho tem como foco o seguinte questionamento:

É possível identificar qual o tipo de relacionamento profissional entre pessoas (positivo, indiferente, negativo) a partir de um conjunto de informações dinâmicas coletadas no Facebook?

1.3 Objetivos

1.3.1 Gerais

Este trabalho teve como principal objetivo avaliar a capacidade de predição de preferências em parcerias profissionais para a execução de atividades colaborativas a partir de interações sociais entre pessoas dentro da rede social Facebook. As preferências profissionais previstas neste trabalho são classificadas como positivas, indiferentes ou negativas e serão validadas com a aplicação de um teste sociométrico.

1.3.2 Específicos

Como objetivo secundário pretendeu-se desenvolver um aplicativo para ser usado no Facebook que colete automaticamente um conjunto de interações sociais entre usuários, de modo controlado, assegurando que terceiros não tenham seus dados coletados, por não comporem a amostra.

Pretendeu-se, também, mapear dados coletados automaticamente no Facebook de uma dada rede social com os dados obtidos a partir da aplicação de um teste sociométrico e investigar se é possível prever as características relacionais de uma rede social a partir de dados obtidos em interações sociais mediadas pelo Facebook.

1.4 Organização da Monografia

O Capítulo 2 é dedicado a mostrar toda a fundamentação teórica e conceitos que proporcionaram o desenvolvimento do projeto. O Capítulo 3 apresenta toda a metodologia, onde são apresentados os passos do desenvolvimento do projeto. O Capítulo 4 apresenta todos os resultados e as discussões a cerca de cada um. O Capítulo 5 aponta as conclusões obtidas e trabalhos futuros. O Capítulo 6 apresenta as referências bibliográficas que guiaram o desenvolvimento do projeto. O Capítulo 7 apresenta os apêndices, documentos legais para a pesquisa.

2 Fundamentação Teórica

Este capítulo tem por finalidade descrever os seguintes assuntos. Na Seção 2.1 fala sobre redes sociais e suas análises. A Seção 2.2 apresenta modelos de predição de links que serão utilizados na pesquisa.

2.1 Redes Sociais e sua Análise

As redes sociais expressam o mundo em movimento que é um mundo que não entendemos [Molina & Aguilar, 2004]. Isto porque redes sociais referem-se a um conjunto de pessoas (ou organizações ou outras entidades sociais) conectadas por relacionamentos sociais, motivados pela amizade e por relações de trabalho ou compartilhamento de informações e que, por meio dessas ligações, vão construindo e reconstruindo a estrutura social.

Essas redes são sistemas compostos por “nós” e conexões entre eles que, nas ciências sociais, são representados por sujeitos sociais (indivíduos, grupos, organizações etc.) conectados por algum tipo de relação. De forma genérica, pode-se estudar esses sistemas visando apenas a entender como eles se comportam e como as conexões influenciam esse comportamento [Borgatti, 2009], [Watts, 1999].

Os estudos e a concepção contemporânea sobre as redes sociais tiveram seu início a partir das descrições de Jacob Levi Moreno, um médico que se tornou analista social e desenvolveu o conceito de sociometria.

A partir das suas primeiras observações em um campo de refugiados durante a Primeira Guerra Mundial, pesquisando os relacionamentos das pessoas que conviviam em cada barraca, entre as barracas e nos diferentes setores de trabalho existentes no campo, Moreno descreveu o sentido e a importância da pertinência aos grupos e das afinidades e incompatibilidades entre as pessoas como elementos articuladores da vida social.

Fundamentado em suas observações, começou a transferir famílias nos

alojamentos com base nas suas afinidades, promovendo aumento da cooperação e ajuda mútua e diminuição dos desajustes entre os diferentes grupos familiares [Marineau, 1889].

Segundo Moreno, o homem deve eleger e estabelecer sua vida junto a quem também o elege, não ficando amarrado a determinismos biológicos e sociais. Seu sonho era criar uma sociedade dinâmica com reformulações contínuas, uma sociedade ideal, onde cada pessoa pudesse encontrar sua família sociométrica [Moreno, 1934].

A abordagem teórica desenvolvida por Moreno, conhecida amplamente como Psicodrama, foi denominada de Socionomia que se desdobra em sociometria, sociodinâmica e a sociatria (FIGURA 1). Socionomia veio da junção da palavra latina *sociu*, que significa companheiro, grupo e da palavra grega *nomos*, que significa regra, lei. O desenvolvimento da Teoria Socionômica traduz a tentativa de construção de uma nova teoria de estudo e tratamento das regras e da dinâmica que regem o comportamento social e grupal.

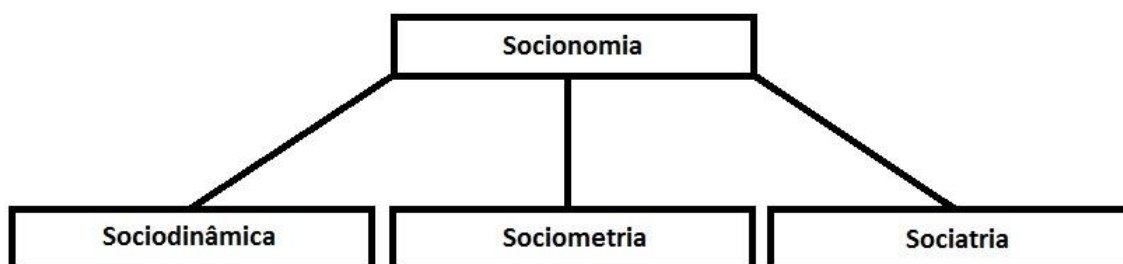


Figura 1 - Desdobramentos da Socionomia

Moreno buscava uma compreensão do campo social partindo de seu interior, estudando os átomos sociais e a maneira como estes se interligavam formando redes sociais. Nesta teoria, Moreno postula que a socionomia seria o estudo amplo da sociedade, e um dos seus campos de estudo seria a Sociometria [Moreno, 1934].

A Sociometria é a representação métrica das relações interpessoais dentro de um grupo social, visando desvendar as características dos vínculos interpessoais entre os seus membros, bem como seu modo de funcionamento e sua organização psicossocial. Segundo Moreno a teoria das relações interpessoais baseia-se na díade primária, a ideia e experiência do encontro de

dois atores, o evento concreto-situacional preliminar a todas as relações interpessoais, propondo que se pode discernir padrões de atração, repulsa e indiferença no limite entre indivíduos e grupos. Esses padrões formam uma configuração dinâmica dentro de cada grupo (desde os menores grupamentos, até as comunidades e a sociedade como um todo) [Moreno, 1934].

O objetivo final da Sociometria é, portanto, a partir da identificação dos padrões de relacionamento bilateral que compõem uma configuração grupal, definir sua estrutura, mapeando-se sociometricamente os vínculos dos indivíduos do grupo, para que se possa compreender seu *modus operandi* [Moreno, 1934].

Em uma pesquisa com os internos da prisão de Sing Sing, a partir de entrevistas e questionários, Moreno foi capaz de definir algumas das variáveis sociais, culturais e psicológicas daquele grupo. De posse dessas informações, caracterizou cada detento e recomendou uma nova organização prisional, agrupando nas celas os prisioneiros que pudessem ser compatíveis e terapêuticos uns para com os outros, com o objetivo de favorecer sua recuperação [Marineau, 1889].

Na intervenção de Sing Sing, os métodos utilizados não permitiram a participação ativa dos detentos nas escolhas sociométricas. Moreno e seus colaboradores avaliavam as características psicossociais de cada um e definiam quem devia ficar com quem, de forma que a influência entre si fosse a mais favorável possível para a sua reabilitação. Mais tarde realiza um estudo com 1.853 jovens estudantes, de 4 a 14 anos, em uma escola pública, com o objetivo de avaliar se há conflito entre a configuração oficial existente entre os alunos e a organização realmente desejada por eles. Neste estudo, diferentemente do conduzido em Sing Sing, os participantes têm papel ativo nas escolhas sociométricas, pois ao invés de Moreno inferir as preferências relacionais a partir do questionamento das características individuais, ele utiliza como método de pesquisa a aplicação do teste sociométrico [Moreno, 1934].

Pelo teste sociométrico os alunos escolhem, por escrito, o menino ou a menina que desejam ter como companheiros de carteira e são avisados que

estão elegendo os colegas que poderão realmente se sentar ao lado no próximo semestre. Pela análise dos resultados, Moreno conseguiu concluir como se dão as estruturas relacionais, as escolhas entre os sexos, as escolhas recíprocas e o isolamento em função da idade.

O teste sociométrico foi concebido como um caminho para a reformulação positiva de grupos e comunidades. É um método de investigação que objetiva facilitar a compreensão das redes de vínculos que configuram a estrutura dos grupos humanos [Bustos, 1979].

O teste sociométrico se situa nos vínculos entre as pessoas que compõem um grupo, a fim de estabelecer o sinal através do qual se vinculam: positivo (aceitação), negativo (rejeição) e indiferente (neutro). A partir da sua aplicação pode-se construir o sociograma, que nada mais é do que o mapeamento da rede social de um grupo [Bustos, 1979].

No sociograma são registrados os relacionamentos entre todas as pessoas do grupo interagindo entre si. Os sociogramas são registrados que representam como se os processos acontecessem em um espaço-social - não em um espaço físico real - e em um universo atemporal [Moreno, 1934].

O relacionamento social identificado e mapeado por Moreno pode ser encontrado e mapeado em redes sociais mediadas pelo computador [Wellman, 2001], sendo possível identificar as mesmas relações de afinidade, neutralidade ou rejeição entre os indivíduos. Redes sociais como Twitter e Facebook não possuem declaração específica de arestas que signifiquem afinidade, neutralidade ou rejeição, já redes como WikiElections, Slashdot e Epinions as arestas que definem o relacionamento são explícitas no sociograma da rede.

Foi a partir do estudo de Moreno que diversas análises em vários tipos de redes sociais foram surgindo, sendo o principal objetivo de se analisar uma rede social estabelecer relação entre vértices. Embora os estudos tenham surgido a partir das interações entre pessoas, as técnicas utilizadas para estudo de redes sociais possuem grande abrangência podendo ser aplicadas em redes biológicas - para entender, por exemplo, o comportamento entre proteínas, micro organismos e metabolismo celular - e em redes colaborativas onde seria

inviável detectar informações e inferir relacionamentos entre os vértices através de análises manuais [Fortunato, 2010].

Há um interesse crescente no desenvolvimento potencial que pode ser oferecido pela análise das redes sociais, principalmente nas interações grupais mediadas pelo computador [Wellman, 2001].

A análise da rede social vai para além dos atributos específicos de indivíduos, considerando as relações e as trocas entre os diferentes atores sociais. Os analistas de redes sociais procuram descrever as redes de relações, pormenorizando, tanto quanto possível, os padrões de relacionamento que se destacam nessas redes. São descritos os padrões relacionais e como a dinâmica da rede social ajuda a explicar o envolvimento e o comportamento individual na própria rede [Garton *et al.*, 1997].

A análise das interações sociais podem incluir medidas de existência, frequência e direcionamento. O conteúdo das interações entre os indivíduos também é relevante para o entendimento da dinâmica da rede social e permitem a análise da qualidade das relações. Em redes sociais mediadas pelo computador, este conteúdo pode ser exemplificado por mensagens trocadas, fotos compartilhadas, sistemas de reforço comportamental (representado pelas “curtidas”), etc [Garton *et al.*, 1997].

2.2 Modelos de Predição de Comportamento em Redes Sociais

As redes sociais – em particular redes sociais *online*, tais como Eppinions, WikiElections e Slashdot – continuamente agregam novos usuários. Um procedimento adotado por uma grande parte das redes sociais *online* existentes (Facebook, Hi5, Twitter, etc) é a sugestão de arestas para os usuários [Alsaleh *et al.*, 2012] através, por exemplo, de técnicas de mineração de grafos usando informações dos vértices, detecção de comunidades, vizinho mais próximo e clusterização [Terveen *et al.*, 2005].

A partir da análise da estrutura e características, tanto globais como locais, de um grafo que represente a topologia de uma rede social, Symeonidis *et al.* [2010] apresenta uma metodologia para prever novas conexões e recomendar novas amizades para usuários em redes sociais. A partir de dois testes, o primeiro consistindo apenas de relacionamentos positivos no grafo da topologia da rede e um segundo consistindo de arestas positivas e negativas, verificou-se que a metodologia desenvolvida pela pesquisa foi superior aos métodos de predição atualmente utilizados nas redes sociais Facebook e Hi5 - considerando que só existem arestas positivas - e a rede social Epinions com arestas negativas.

Zhao *et al.* [2012], propõem um novo foco sobre os estudos em redes sociais. Diferentemente da predição e recomendação de arestas, seus estudos têm como objetivo estabelecer a força de um relacionamento entre dois indivíduos através de um *framework* desenvolvido, em particular, para a rede social Facebook. Os dados coletados retratam as interações sociais entre dois indivíduos, tais como esportes e entretenimento, os dados do perfil e atividades em comum. Assim, através do *framework* desenvolvido, é possível estimar a força de uma relação. Em um grafo que represente a topologia da rede social, a força (a intensidade de um relacionamento positivo) de uma relação é identificada como o peso de uma aresta entre dois vértices e estes representam os usuários que compõem a rede social.

Todos os dados extraídos das redes sociais - como, por exemplo, dados do perfil, lista de amigos, interesses, entre outros - servem unicamente para estabelecer relações entre os usuários, ou seja, estabelecer arestas (ponderadas ou não através de interpretação dos dados coletados) para cada par de vértices existente no grafo que representa a topologia da rede social. Deixando de lado o significado dos vértices e das arestas do grafo que representa a topologia da rede social, é possível fazer uma análise sobre a estrutura de grafos.

A Teoria do Equilíbrio Estrutural [Facchetti *et al.*, 2011], que foi proposta, inicialmente, por Heider [1946], com o objetivo de compreender como relações de amizade e hostilidade surgem em redes sociais com grandes escalas. A

Teoria do Equilíbrio Estrutural permite a inferência de novas arestas, positivas e negativas, com a tese de que “o amigo do meu amigo é meu amigo”, “o amigo do meu inimigo é meu inimigo”, “o inimigo do meu amigo é meu inimigo” e “o inimigo do meu inimigo é meu amigo”.

A importância de identificar uma aresta positiva, indiferente ou negativa, está diretamente relacionada à utilização desta informação.

Em contexto de redes sociais, detectar uma aresta positiva pode influenciar o direcionamento de propagandas com maior probabilidade de interesse do usuário, podem-se recomendar novas amizades. A aresta negativa auxilia em não permitir que um sistema de recomendação automática desagrade aos usuários com propagandas fora do seu perfil ou com direcionamentos de amizades incompatíveis. Já arestas indiferentes podem ser utilizadas para explorar novas características e proporcionar novas experiências que ajudem a definir o perfil do usuário.

Para contextos empresariais, os relacionamentos interpessoais de cada pessoa podem influenciar as dinâmicas corporativas. A afinidade entre pessoas pode aumentar a colaboração interpessoal dos trabalhadores, já a inimizade pode representar o oposto. E a neutralidade pode ser estudada para modelar afeição através da exploração de características sociais.

2.2.1 Weka

Weka é um *software* gratuito registrado sobre a licença GNU *General License* e foi desenvolvido pela Universidade de Waikato, Nova Zelândia. Este *software* representa uma coleção de algoritmos para análises de dados com mais de cem modelos de aprendizagem de máquina, além de métodos para análises estatísticas [Hall *et al.*, 2009].

Dos mais de cem modelos, serão apresentados apenas quatro modelos de aprendizagem de máquina e dois modelos para análises estatísticas, os quais foram escolhidos para serem utilizados: *Decision Table* (DT) [Kohavi, 1995], J48 [Quinlan, 1993], *Multi Layer Perceptron* (MLP) [Hall *et al.*, 2009], *Random Forest*

(RF) [Breiman, 2001], *Info Gain Attribute Evaluation* (IGAE) e *Chi Square Attribute Evaluation* (CSAE) [Hall et al., 2009].

2.2.1.1 Decision Table

A *Decision Table*, ou Tabela de Decisões, determina, através de um conjunto de condições, qual ação deve ser tomada. Essa técnica consiste de quatro componentes:

- Condição – Identificação da lista de condições
- Ações – Identificação da(s) ação(ões) para uma condição
- Regras – Os dados que compõem o conjunto de condições
- Decisão – Ação decidida através das regras

O *software* Weka traz o modelo proposto por Kohavi. Seu estudo focou em uma tabela de decisão muito simples e constrói a(s) regra(s) de decisão(ões) com a DTM (*Decision Table Majority* – Tabela de Decisão por Maioria), que é o seu mecanismo de aprendizado inteligente.

Dada uma instância cuja ação é desconhecida, a Tabela de Decisão procura por uma regra. Note que como pode não haver correspondência exata, a instância pode ser compatível com mais de um conjunto de regras, logo o conjunto de decisões pode ser mais que um e nestes casos a DTM é responsável por priorizar em qual regra a instância se classifica [Kohavi, 1995].

2.2.1.2 J48

A J48 é uma árvore de decisões probabilísticas encontrada na obra de Quinlan "*C4.5: Programs for Machine Learning*".

A principal característica deste modelo é que para cada nó folha existe um grau de confiabilidade, chamado de entropia, contido no intervalo [0, 1] calculado automaticamente. A cada folha encontrada durante a tomada de decisão, o algoritmo verifica, através da entropia, se pode ou não confiar em criar novas ramificações a partir da folha atual.

A confiabilidade desejada é um parâmetro passado para o algoritmo, também no intervalo [0, 1]. Quando a confiabilidade é 1, isso significa que a

árvore não deve fazer nenhum corte durante a sua verificação de decisão a ser tomada. Quando a confiabilidade é 0, isso significa que a árvore deve fazer todos os cortes, implicando em não conseguir decidir.

A aprendizagem do método está relacionada ao processo de construção da árvore de decisão, baseando-se na entropia de cada atributo para criar ramificações e encontrar a decisão a ser tomada.

2.2.1.3 Multi Layer Perceptron

O MLP (*Multi Layer Perceptron* - Perceptron de Múltiplas Camadas) é um modelo de rede neural artificial que mapeia conjuntos de dados de entrada em um conjunto de saídas apropriadas. A MLP consiste de várias camadas de nós em um grafo direcionado, com cada camada totalmente ligada a próxima, exceto para os nós de entrada, pois cada nó é um neurônio (ou elemento de processamento), com uma função de ativação não linear.

O modelo de MLP do software Weka utiliza o aprendizado supervisionado de retropropagação, conhecida como *backpropagation*, para treinamento da rede [Hall *et al.*, 2009].

2.2.1.4 Random Forest

A *Random Forest*, Floresta Aleatória, é uma técnica baseada em Algoritmos Genéticos. Um algoritmo genético é uma técnica de otimização com fundamentação teórica de algoritmos evolutivos que usam técnicas inspiradas pela biologia evolutiva como hereditariedade, mutação, seleção natural e recombinação para achar soluções aproximadas em problemas de otimização e busca [Hall *et al.*, 2009].

O *software* Weka apresenta a implementação de Floresta Aleatória proposta por Breiman [Breiman, 2001]. Neste modelo a Floresta Aleatória é uma combinação de árvores classificadoras de tal modo que cada árvore depende dos valores de um vetor aleatório de forma independente e com a mesma distribuição para todas as árvores. Ainda, o modelo possui mecanismos para calcular o erro de convergência que tende a um limite quando o número de

árvores na floresta torna-se grande. O erro generalizado de uma árvore classificadora depende da força de uma árvore individual e da correlação entre elas e o mecanismo de aprendizagem dá-se pelas operações baseadas em programação genética, através de mutações em nós e folhas e cruzamento entre árvores.

2.2.1.5 Info Gain & Chi Square Attribute Evaluation

Info Gain e *Chi Square Attribute Evaluation* são métricas que avaliam a importância de um atributo dentro dos padrões de treinamento dos métodos de aprendizagem de máquina.

A importância de um atributo simboliza a sua relevância para classificar um padrão. Quanto maior o valor de um atributo, maior é o seu poder de influenciar na classificação de um padrão. Para o modelo *Info Gain*, o valor está contido no intervalo $[0, 1]$ e para o modelo *Chi Square*, o valor está contido no intervalo $[0, \infty[$. Já o menor valor possível é zero, indicando que o atributo em análise não possui relevância alguma para classificar o padrão.

3 Metodologia

Este capítulo fala sobre todas as etapas de coleta dos dados e transcrição das informações para o anonimato.

3.1 Coleta de Dados

A pesquisa faz uso de duas redes sociais, sendo uma *online* e outra *offline*. Uma rede social *online* é uma rede cujas interações entre os usuários são mediadas pelo uso da internet, já uma rede social *offline* não é mediada pela internet.

As redes sociais *online* e *offline*, foco deste estudo, são o Facebook e o teste sociométrico e dada a natureza das redes sociais, optamos por dividir a coleta de dados em alguns momentos distintos.

I. Apresentação da pesquisa.

➤ Apresentar a proposta da pesquisa, bem como a hipótese e os métodos de análise. Apresenta-se, também, o Termo de Consentimento Livre e Esclarecido (TCLE) e a Declaração de Aceite de Participação (DAP) que resguardam os direitos dos participantes respeitando a Resolução 466/12 do Conselho Nacional de Saúde (CNS) (BRASIL, 2012).

II. Assinatura do Termo de Consentimento Livre e Esclarecido (TCLE – Apêndice I).

➤ Entregar o TCLE para que todos os participantes o assinem concordando em participar da pesquisa.

III. Coleta dos User's ID.

➤ Todos os participantes, que concordaram com o TCLE e o assinaram, têm seu ID coletado para que fique assegurado a não coleta de dados de qualquer usuário que não assinou o TCLE.

IV. Coleta de Dados do Facebook.

➤ Todos os participantes, que tiveram seu ID coletado, acessam o Facebook e o aplicativo disponibilizado para a coleta de dados automática.

V. Aplicação do Teste Sociométrico.

Todos os participantes, após acessarem o aplicativo do Facebook, foram chamados para responder a um teste sociométrico impresso, respondendo para todos os participantes da pesquisa a pergunta: *Com quais das pessoas listadas você gostaria de trabalhar em equipe?*

3.1.1 Aspectos Éticos da Pesquisa

O presente estudo foi realizado respeitando-se os preceitos estabelecidos pela Resolução 466/12, de 12/12/2012, do Conselho Nacional de Saúde (CNS) (BRASIL, 2012).

O TCLE (Apêndice I) junto ao DAP (Apêndice II) oficializou a decisão do sujeito em participar do estudo, de maneira livre, espontânea, esclarecida e voluntária, podendo desistir, se assim o desejasse. Estes termos foram assinados em três vias, sendo uma via entregue ao participante e só após a assinatura dos mesmos que se procedeu a coleta de dados.

Foi respeitado, também, o princípio do anonimato, assegurando a privacidade e a proteção da imagem, garantindo que as informações obtidas não serão utilizadas em prejuízo, de qualquer natureza para os respondentes.

Em atendimento à Resolução CNS/MS No 466/12, esta pesquisa foi aprovada sob o número do parecer 270.706 pelo Comitê de Ética em Pesquisa da Universidade Federal de Alfenas – Unifal/MG.

3.1.2 Facebook

O aplicativo para o Facebook foi desenvolvido com o único propósito de coletar algumas das inúmeras formas de interações entre os participantes que concordaram em disponibilizar seus dados para coleta, assim, o controle dos participantes da pesquisa foi feito manualmente pelo desenvolvedor para que pessoas que não compõem a amostra não tenham o ID coletado.

Todos participantes da pesquisa, que concordaram em assinar o TCLE, tiveram seu ID coletado manualmente e essa informação foi armazenada em

um Banco de Dados (BD) que foi automaticamente acessado pelo aplicativo, assim, este fez uso dessa informação para identificar se o ID que acessou o aplicativo consentiu ou não em ter seus dados coletados e, como o único modo de ter seu ID inserido no BD é através da assinatura do TCLE e da coleta manual pelo desenvolvedor do sistema, ficou, assim, garantido que apenas os participantes da pesquisa, de fato, tiveram seus dados coletados.

A razão para que o controle foi feita através do ID de cada usuário do Facebook é que essa é uma informação pública e é como o Facebook identifica cada um de seus milhares de usuários, isso significa que é impossível encontrar dois usuários quaisquer que tenham o mesmo ID, assim, essa propriedade foi utilizada para garantir que o aplicativo coletasse informações único e exclusivamente de ID's contidos no BD.

Tabela 1 - Pseudocódigo do aplicativo para coleta de dados no Facebook.

1.	Begin
2.	<i>id</i> ← usuário_atual();
3.	<i>lista_ids</i> ← lista_ids();
4.	if verifica_id() = true ;
5.	then coletar_dados(<i>id</i> , <i>todos_ids</i>);
6.	else mensagem();
7.	end if
8.	End

3.1.3 Identificação do Usuário

Na rede social Facebook todo usuário está associado a um número chamado ID. Este é único e de caráter público, ou seja, qualquer entidade pode ter acesso a esta informação sem consentimento ou aviso.

Quando um usuário qualquer acessa o aplicativo, com o auxílio da *Application Programming Interface* (API) (um conjunto de funcionalidades prontas disponibilizadas pelo Facebook para que os desenvolvedores de aplicativos possam acessar os dados dos usuários mediante requisição feita pelo próprio aplicativo) que o Facebook disponibiliza, o seu ID é a primeira informação acessada para validar se o usuário é ou não um participante voluntário da pesquisa.

3.1.4 Listando os ID's que Aceitaram em Participar da Pesquisa

Sempre que um usuário qualquer acessa o aplicativo, este precisa recuperar todos os ID's armazenados no BD para verificar se o ID que está acessando o aplicativo faz ou não parte do conjunto de ID's que assinaram o TCLE.

Caso o ID esteja presente no BD, indicando que o usuário é um participante voluntário da pesquisa, o aplicativo inicia a coleta de informações. Caso contrário, o aplicativo apresenta uma mensagem ao usuário informando que ele não faz parte do conjunto de pessoas voluntárias e, assim, não coleta as informações do ID corrente.

3.1.5 Estado do ID

Os dados coletados são, em sua grande maioria, interações de todos participante da pesquisa para todos os outros, então, a função “coletar_dados” deve receber os parâmetros que indica o usuário atual e uma lista indicando os participantes da pesquisa para coletar as seguintes informações:

1. Amizades – ID é ou não amigo dos ID's no Facebook.
2. Comentário em Foto – ID comentou em fotos pertencente aos ID's.
3. Comentários em *Links* – ID comentou em *links* publicado pelos ID's.
4. Comentários em *Status* – ID comentou em *status* publicado pelos ID's.
5. Comentário em Álbuns – ID comentou em álbuns pertencentes aos ID's.
6. Conversas – Número de mensagens trocadas entre o ID e os ID's.
7. Filmes – Filmes do ID.
8. Fotos Curtidas – Quantas vezes o ID clicou em *Like* nas fotos pertencentes aos ID's.
9. Grupos – Todos os grupos que o ID participa.
10. Interesses – Todos os interesses preenchidos no perfil do ID.

11. *Links Curtidos* – Quantas vezes o ID clicou em *Like* nos *links* publicado pelos ID's.
12. *Marcações* – Quantas vezes o ID foi marcado em fotos pertencente aos ID's.
13. *Músicas* – Todas as músicas do ID.
14. *Status Curtidos* – Quantas vezes o ID curtiu os *status* pertencente aos ID's.

Todas as informações coletadas são armazenadas em Banco de Dados.

3.1.6 Teste Sociométrico

O Teste Sociométrico consiste de um questionário técnico, ou seja, uma folha de respostas que deve ser preenchida pelos participantes voluntários da pesquisa. A pergunta feita aos participantes “*Com quais das pessoas listadas você gostaria de trabalhar em equipe?*”, teve como opção de resposta as alternativas “*SIM*”, “*INDIFERENTE*” e “*NÃO*” para que único e exclusivamente uma resposta fosse escolhida para cada um dos participantes (Apêndice III).

Para identificar quem é o autor da resposta, o Teste solicita do participante que ele faça uma marcação em seu nome e assine o documento.

3.2 Pré Processamento da Informação

Por questões éticas, o professor responsável pela pesquisa foi o responsável por mapear os dados coletados pelo aplicativo no Facebook e as respostas obtidas pelo teste sociométrico.

O mapeamento realizado consiste em esconder do pesquisador acadêmico as informações que possam identificar cada um dos participantes voluntários para que tanto o indivíduo quanto suas informações privadas não sejam identificáveis, ficando resguardada, assim, a privacidade dos participantes.

3.2.1 Transcrição dos Dados

Dos dados coletados, há alguns cujas características não são quantitativas, mas sim, qualitativas como, por exemplo, filmes, músicas, interesses e grupos. Para utilizar essas informações, torna-se necessário estabelecer algum critério de relacionamento entre os usuários.

O critério escolhido para tornar os atributos qualitativos em quantitativos entre usuários foi a similaridade entre os dados coletados. Então, para dois usuários quaisquer, temos os seguintes atributos quantitativos:

- A quantidade de grupos em comum
- A quantidade de gêneros musicais em comum
- A quantidade de gêneros cinematográficos em comum
- A quantidade de áreas de interesse em comum

4 Previsão de *Links*

Este capítulo apresenta os resultados e discussões de todos os experimentos feitos.

Para avaliar a capacidade de predição de arestas os experimentos foram executados para cada tipo de grafo:

- Grafo de arestas negativas e não negativas
- Grafo de arestas indiferentes e não indiferentes
- Grafo de arestas positivas e não positivas
- Grafo de arestas negativas e indiferentes
- Grafo de arestas negativas e positivas
- Grafo de arestas indiferentes e positivas

Tabela 2 - Exemplo de matriz de mapeamento Entrada e Saída

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15
B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15
...

A Tabela 2 é um exemplo simplificado de matriz de mapeamento de padrões de Entrada e Saída para os modelos de aprendizagem de máquina.

Cada linha da matriz representa a associação de uma pessoa qualquer i para uma j , para todo $i \neq j$. A matriz é composta de quinze atributos, sendo os primeiros quatorze as relações presentes no Facebook. O décimo quinto atributo é a resposta mapeada no teste sociométrico da pessoa i para a pessoa j .

Cada um dos experimentos feitos utilizaram os grafos Turma1 e Turma2 combinando-os entre padrões de treinamento e padrões de teste:

- Treinamento com o grafo Turma1 e teste com o grafo Turma1.
- Treinamento com o grafo Turma1 e teste com o grafo Turma2.
- Treinamento com o grafo Turma2 e teste com o grafo Turma1.
- Treinamento com o grafo Turma2 e teste com o grafo Turma2.

4.1 Organização da Base de Dados

Os dados coletados do Facebook e do teste sociométrico são componentes de um grafo, então a partir deste ponto, serão utilizadas as terminologias vértices e arestas para fazer referência aos participantes da pesquisa e seus dados coletados no Facebook e no teste sociométrico.

Os dados foram divididos em dois grafos, representando originalmente dois grupos de participantes. Os grafos topológicos obtidos são grafos completos e direcionados, ou seja, há uma aresta de todos os vértices para todos os outros. Cada aresta é constituída por quinze atributos, sendo quatorze deles as interações coletadas pelo aplicativo do Facebook e o décimo quinto representa a resposta dada no teste sociométrico, logo, o que determina o tipo de aresta é a resposta dada pelo teste sociométrico.

O primeiro grafo topológico, chamado Turma1, contém um total de 31 vértices e 930 arestas, pois não existe aresta de um vértice qualquer para ele mesmo. Das 930 arestas, 120 são negativas, 441 são indiferentes e 369 são positivas.

O segundo grafo topológico, chamado Turma2, contém um total de 19 vértices e 342 arestas, pois não existe aresta de um vértice qualquer para ele mesmo. Das 342 arestas, 39 são negativa, 167 são indiferentes e 136 são positivas.

4.1.1 Base de Dados Experimental

Para a execução dos experimentos e verificação da capacidade de predição de relacionamentos profissionais através de informações provenientes do Facebook torna-se necessário a obtenção de um conjunto de dados para treinamento, ou seja, para que os algoritmos aprendam padrões ou métricas que permitam dizer quando uma aresta será positiva, indiferente ou negativa, dado um conjunto de dados como entrada, até então, desconhecido.

Tabela 3 - Percentual das arestas no grafo Turma1.

Aresta	Quantidade	Porcentagem
Negativa	120	12,90%
Indiferente	441	47,42%
Positiva	369	39,68%
Total	930	100,00%

Tabela 4 - Percentual das arestas no grafo Turma2.

Aresta	Quantidade	Porcentagem
Negativa	39	11,40%
Indiferente	167	48,83%
Positiva	136	39,77%
Total	342	100,00%

Para avaliar se os modelos de aprendizagem de máquina são capazes de distinguir um tipo de arestas das demais, foram criados novos grafos, tendo como base os grafos Turma1 e Turma2 originais, para verificar qual é o tipo de aresta que os modelos conseguem identificar.

Os novos grafos gerados, tanto para o grafo Turma1 como para o grafo Turma2, são compostos por arestas:

- Arestas negativas e arestas não negativas
- Arestas indiferentes e arestas não indiferentes
- Arestas positivas e arestas não positivas
- Arestas negativas e arestas indiferentes
- Arestas negativas e arestas positivas
- Arestas indiferentes e arestas positivas

4.1.1.1 Grafo de Arestas Negativas e Não Negativas

Um grafo composto de arestas negativas e arestas não negativas é um grafo com dois tipos de arestas: negativas e não negativas.

O tipo de aresta negativa é a original proveniente do grafo Turma1 ou do grafo Turma2. Já uma aresta não negativa é a junção de arestas indiferentes e positivas em um tipo único, seja considerando arestas indiferentes como positivas ou arestas positivas como sendo indiferentes. Note que em ambos os casos, a diferença se dá, exclusivamente, no décimo quinto atributo da aresta.

Tabela 5 - Percentual das arestas negativas e não negativas no grafo Turma1.

Aresta	Quantidade	Porcentagem
Negativa	120	12,90%
Não Negativa	810	87,10%
Total	930	100,00%

Tabela 6 - Percentual das arestas negativas e não negativas no grafo Turma2.

Aresta	Quantidade	Porcentagem
Negativa	39	11,40%
Não Negativa	303	88,60%
Total	342	100,00%

4.1.1.2 Grafo de Arestas Indiferentes e Não Indiferentes

Um grafo composto de arestas indiferentes e arestas não indiferentes é um grafo com dois tipos de arestas: indiferentes e não indiferentes.

O tipo de aresta indiferente é a original proveniente do grafo Turma1 ou do grafo Turma2. Já uma aresta não indiferente é a junção de arestas negativas e positivas em um tipo único, seja considerando arestas negativas como positivas ou arestas positivas como sendo negativas. Note que em ambos os casos, a diferença se dá, exclusivamente, no décimo quinto atributo da aresta.

Tabela 7 - Percentual das arestas indiferentes e não indiferentes no grafo Turma1.

Aresta	Quantidade	Porcentagem
Indiferente	441	47,42%
Não Indiferente	489	52,58%
Total	930	100,00%

Tabela 8 - Percentual das arestas indiferentes e não indiferentes no grafo Turma2.

Aresta	Quantidade	Porcentagem
Indiferente	167	48,83%
Não Indiferente	175	51,17%
Total	342	100,00%

4.1.1.3 Grafo de Arestas Positivas e Não Positivas

Um grafo composto de arestas positivas e arestas não positivas é um grafo com dois tipos de arestas: positivas e não positivas.

O tipo de aresta positiva é a original proveniente do grafo Turma1 ou do

grafo Turma2. Já uma aresta não positiva é a junção de arestas negativas e indiferentes em um tipo único, seja considerando arestas negativas como indiferentes ou arestas indiferentes como sendo negativas. Note que em ambos os casos, a diferença se dá, exclusivamente, no décimo quinto atributo da aresta.

Tabela 9 - Percentual das arestas positivas e não positivas no grafo Turma1.

Aresta	Quantidade	Porcentagem
Positiva	369	39,68%
Não Positivas	561	60,32%
Total	930	100,00%

Tabela 10 - Percentual das arestas positivas e não positivas no grafo Turma2.

Aresta	Quantidade	Porcentagem
Positiva	136	39,77%
Não Positiva	206	60,23%
Total	342	100,00%

4.1.1.4 Grafo de Arestas Negativas e Indiferentes

Um grafo composto de arestas negativas e arestas indiferentes é um grafo com dois tipos de arestas: negativas e indiferentes.

O tipo de aresta negativa e indiferente é a original proveniente do grafo Turma1 ou do grafo Turma2. A aresta do tipo positiva, para este grafo, é descartada, ou seja, considera-se que ela não existe.

Tabela 11 - Percentual das arestas negativas e indiferentes no grafo Turma1.

Aresta	Quantidade	Porcentagem
Negativa	120	21,39%
Indiferente	441	78,61%
Total	561	100,00%

Tabela 12 - Percentual das arestas negativas e indiferentes no grafo Turma2.

Aresta	Quantidade	Porcentagem
Negativa	39	18,93%
Indiferente	167	81,07%
Total	206	100,00%

4.1.1.5 Grafo de Arestas Negativas e Positivas

Um grafo composto de arestas negativas e arestas positivas é um grafo com dois tipos de arestas: negativas e positivas.

O tipo de aresta negativa e positiva é a original proveniente do grafo Turma1 ou do grafo Turma2. A aresta do tipo indiferente, para este grafo, é descartada, ou seja, considera-se que ela não existe.

Tabela 13 - Percentual das arestas negativas e positivas no grafo Turma1.

Aresta	Quantidade	Porcentagem
Negativa	120	24,54%
Positiva	369	75,46%
Total	489	100,00%

Tabela 14 - Percentual das arestas negativas e positivas no grafo Turma2.

Aresta	Quantidade	Porcentagem
Negativa	39	22,29%
Positiva	136	77,71 %
Total	175	100,00%

4.1.1.6 Grafo de Arestas Indiferentes e Positivas

Um grafo composto de arestas indiferentes e arestas positivas é um grafo com dois tipos de arestas: indiferentes e positivas.

O tipo de aresta indiferente e positiva é a original proveniente do grafo Turma1 ou do grafo Turma2. A aresta do tipo negativa, para este grafo, é descartada, ou seja, considera-se que ela não existe.

Tabela 15 - Percentual das arestas indiferentes e positivas no grafo Turma1.

Aresta	Quantidade	Porcentagem
Indiferente	441	54,44%
Positiva	369	45,56%
Total	810	100,00%

Tabela 16 - Percentual das arestas indiferentes e positivas no grafo Turma2.

Aresta	Quantidade	Porcentagem
Indiferente	167	55,12%
Positiva	136	44,88%
Total	303	100,00%

4.1.2 Balanceamento dos Grafos

Como observado nas tabelas de 4 a 15, que indicam a porcentagem que cada tipo de aresta ocupa em seus respectivos grafos, o volume de arestas não é

proporcional em vários grafos, como, por exemplo, no grafo composto por arestas positivas e não positivas.

Quando a proporção das arestas é muito diferente, os métodos de aprendizagem de máquina podem alcançar métricas triviais de previsão de arestas. Note que se um método de aprendizagem de máquina, aplicado ao grafo composto de arestas “negativas e não negativas”, diz que todas as arestas são não negativas, de acordo com as tabelas Tabela 4 e Tabela 5, a porcentagem de acerto será de 87,10% no grafo Turma1 e 88,60% no grafo Turma2.

Para evitar que os métodos de aprendizagem de máquina alcancem métricas triviais, os grafos criados compostos de arestas “negativas e não negativas”, “positivas e não positivas”, “negativas e indiferentes”, “negativas e positivas”, foram balanceados para que a proporção de suas arestas seja de 50% de cada tipo.

Os grafos que tiveram suas arestas balanceadas foram:

- Arestas negativas e não negativas
- Arestas positivas e não positivas
- Arestas negativas e indiferentes
- Arestas negativas e positivas

4.1.2.1 Balanceamento das Arestas Negativas e Não Negativas

O balanceamento do grafo composto de arestas negativas e não negativas consiste em criar um novo grafo em que 50% de suas arestas sejam negativas e 50% sejam arestas não negativas.

Como observado na Tabela 2, do total de 930 arestas do grafo Turma1, 120 arestas são negativas, 441 arestas são indiferentes e 369 são arestas positivas. Neste caso, 50% das arestas negativas do grafo balanceado serão representadas pelas 120 arestas negativas do grafo original. Os outros 50% de arestas, somando 120 arestas não negativas, serão representadas por 25% de arestas indiferentes, totalizando 60 arestas indiferentes do grafo original, e 25% de arestas positivas, totalizando 60 arestas positivas do grafo original.

Tabela 17 - Percentual das arestas negativas e não negativas no grafo balanceado Turma1.

Aresta	Quantidade	Porcentagem
Negativa	120	50,00%
Não Negativa	120	50,00%
Total	240	100,00%

Da Tabela 3, do total de 342 arestas do grafo Turma2, 39 são arestas negativas, 167 são arestas indiferentes e 136 são arestas positivas. Neste caso, 50% das arestas negativas do grafo balanceado serão representadas pelas 39 arestas negativas do grafo original. Os outros 50% de arestas, somando 39 arestas não negativas, serão representadas por 25% de arestas indiferentes, totalizando 20 arestas indiferentes do grafo original, e 25% de arestas positivas, totalizando 19 arestas positivas do grafo original.

Tabela 18 - Percentual das arestas negativas e não negativas no grafo balanceado Turma2.

Aresta	Quantidade	Porcentagem
Negativa	39	50,00%
Não Negativa	39	50,00%
Total	78	100,00%

Para que a escolha das arestas indiferentes e positivas, que formarão o conjunto de arestas não negativas, fosse válida, a seleção de arestas foi feita aleatoriamente num total de dez vezes, dando origem, assim, a dez grafos distintos compostos de arestas negativas e não negativas para eliminar a possibilidade de uma escolha boa ou ruim das arestas.

4.1.2.2 Balanceamento das Arestas Positivas e Não Positivas

O balanceamento do grafo composto de arestas positivas e não positivas consiste em criar um novo grafo em que 50% de suas arestas sejam positivas e 50% sejam arestas não positivas.

Como observado na Tabela 2, do total de 930 arestas do grafo Turma1, 120 arestas são negativas, 441 arestas são indiferentes e 369 são arestas positivas. Neste caso, 50% das arestas positivas do grafo balanceado serão representadas por 240 arestas positivas do grafo original. Os outros 50% de arestas, somando 240 arestas não positivas, serão representadas por 25% de arestas negativas, totalizando 120 arestas negativas do grafo original, e 25% de

arestas indiferentes, totalizando 120 arestas indiferentes do grafo original.

Tabela 19 - Percentual das arestas positivas e não positivas no grafo balanceado Turma1.

Aresta	Quantidade	Porcentagem
Positivas	240	50,00%
Não Positivas	240	50,00%
Total	480	100,00%

Da Tabela 3, do total de 342 arestas do grafo Turma2, 39 são arestas negativas, 167 são arestas indiferentes e 136 são arestas positivas. Neste caso, 50% das arestas positivas do grafo balanceado serão representadas por 78 arestas positivas do grafo original. Os outros 50% de arestas, somando 78 arestas não positivas, serão representadas por 25% de arestas negativas, totalizando 39 arestas negativas do grafo original, e 25% de arestas indiferentes, totalizando 39 arestas indiferentes do grafo original.

Tabela 20 - Percentual das arestas positivas e não positivas no grafo balanceado Turma2.

Aresta	Quantidade	Porcentagem
Positivas	78	50,00%
Não Positivas	78	50,00%
Total	156	100,00%

Para que a escolha das arestas negativas e indiferentes, que formarão o conjunto de arestas não positivas, fosse válida, a seleção de arestas foi feita aleatoriamente num total de dez vezes, dando origem, assim, a dez grafos distintos compostos de arestas positivas e não positivas para eliminar a possibilidade de uma escolha boa ou ruim das arestas.

4.1.2.3 Balanceamento das Arestas Negativas e Indiferentes

O balanceamento do grafo composto de arestas negativas e indiferentes consiste em criar um novo grafo em que 50% de suas arestas sejam negativas e 50% sejam arestas indiferentes.

Como observado na Tabela 2, do total de 930 arestas do grafo Turma1, 120 arestas são negativas e 441 arestas são indiferentes. Neste caso, 50% das arestas negativas do grafo balanceado serão representadas por 120 arestas negativas do grafo original e os outros 50% de arestas indiferentes do grafo balanceado serão representadas por 120 arestas indiferentes do grafo original.

Tabela 21 - Percentual das arestas negativas e indiferentes no grafo balanceado Turma1.

Aresta	Quantidade	Porcentagem
Negativas	120	50,00%
Indiferentes	120	50,00%
Total	240	100,00%

Da Tabela 3, do total de 342 arestas do grafo Turma2, 39 são arestas negativas e 167 são arestas indiferentes. Neste caso, 50% das arestas negativas do grafo balanceado serão representadas por 39 arestas negativas do grafo original e os outros 50% de arestas indiferentes do grafo balanceado serão representadas por 39 arestas indiferentes do grafo original.

Tabela 22 - Percentual das arestas negativas e indiferentes no grafo balanceado Turma2.

Aresta	Quantidade	Porcentagem
Negativas	39	50,00%
Indiferentes	39	50,00%
Total	78	100,00%

Para que a escolha das arestas indiferentes fosse válida, a seleção de arestas foi feita aleatoriamente num total de dez vezes, dando origem, assim, a dez grafos distintos compostos de arestas negativas e indiferentes para eliminar a possibilidade de uma escolha boa ou ruim das arestas.

4.1.2.4 Balanceamento das Arestas Negativas e Positivas

O balanceamento do grafo composto de arestas negativas e positivas consiste em criar um novo grafo em que 50% de suas arestas sejam negativas e 50% sejam arestas positivas.

Como observado na Tabela 2, do total de 930 arestas do grafo Turma1, 120 arestas são negativas e 369 arestas são positivas. Neste caso, 50% das arestas negativas do grafo balanceado serão representadas por 120 arestas negativas do grafo original e os outros 50% de arestas positivas do grafo balanceado serão representadas por 120 arestas positivas do grafo original.

Tabela 23 - Percentual das arestas negativas e positivas no grafo balanceado Turma1.

Aresta	Quantidade	Porcentagem
Negativas	120	50,00%
Positivas	120	50,00%
Total	240	100,00%

Da Tabela 3, do total de 342 arestas do grafo Turma2, 39 são arestas negativas e 136 são arestas positivas. Neste caso, 50% das arestas negativas do grafo balanceado serão representadas por 39 arestas negativas do grafo original e os outros 50% de arestas positivas do grafo balanceado serão representadas por 39 arestas positivas do grafo original.

Tabela 24 - Percentual das arestas negativas e positivas no grafo balanceado Turma2.

Aresta	Quantidade	Porcentagem
Negativas	39	50,00%
Positivas	39	50,00%
Total	78	100,00%

4.2 Resultados

Para cada grafo criado foram realizados quatro sequências de testes para avaliar a predição de *links*.

- Treinamento com a Turma1 e Teste com a Turma1 (T1T1)
- Treinamento com a Turma1 e Teste com a Turma2 (T1T2)
- Treinamento com a Turma2 e Teste com a Turma1 (T2T1)
- Treinamento com a Turma2 e Teste com a Turma2 (T2T2)

Para os casos em que o treinamento e o teste foram feitos com o mesmo grafo, como em T1T1 e T2T2, a base de dados foi dividida em duas partes. A primeira parte foi composta de 66% do total de arestas e foi utilizada para a fase de treinamento dos modelos, já a segunda parte, com os 34% restantes, foi responsável pela fase de teste.

4.2.1 Predição de Arestas Negativas e Não Negativas

Os experimentos realizados indicam que em alguns casos é possível distinguir arestas negativas e não negativas para um grupo social, ou seja, é possível prever suas ocorrências, mas que essa distinção não pode ser aplicada para grupos sociais distintos.

Tabela 25 - Porcentagem de acerto na fase de teste, com arestas balanceadas em negativas e não negativas.

	J48	DT	RF	MLP
T1T1	63,4146%	63,4146%	65,8537%	57,3171%
T1T2	46,1988%	42,6901%	46,7836%	50,2924%
T2T1	31,3978%	57,4194%	40,3333%	32,2473%
T2T2	62,2223%	54,0741%	58,5186%	68,1482%

Pode-se observar, pela Tabela 24, que todos os modelos conseguem distinguir, com acerto acima de 50%, as arestas para os experimentos T1T1 e T2T2. Isso é um indicativo de que as arestas negativas e não negativas de um grupo de pessoas tendem a ter características em comum.

Por outro lado, os experimentos T1T2 e T2T1 mostram que, em geral, os modelos não conseguem extrapolar a detecção de arestas negativas e não negativas para diferentes grupos sociais. Como o DT conseguiu acertar 57.4194%, isso é um indicativo de que há alguma regra que classifique corretamente as arestas negativas e não negativas para o experimento T2T1, que os outros modelos não conseguiram encontrar.

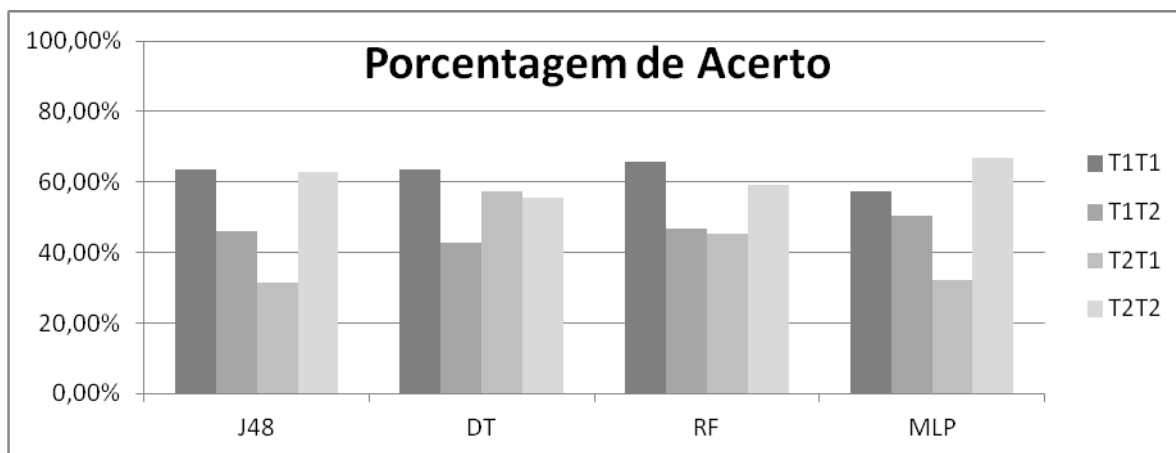


Figura 2 - Gráfico de barras da porcentagem de acerto na fase de teste, com arestas balanceadas em negativas e não negativas.

4.2.2 Predição de Arestas Indiferentes e Não Indiferentes

Os experimentos realizados indicam que é possível distinguir arestas indiferentes e não indiferentes dentro de um grupo social e entre grupos distintos.

Tabela 26 - Porcentagem de acerto na fase de teste, com arestas indiferentes e não indiferentes.

	J48	DT	RF	MLP
T1T1	71,5190%	71,8354%	68,0380%	71,2025%
T1T2	71,0526%	68,4211%	79,8246%	68,7135%
T2T1	57,9570%	73,2258%	66,2366%	66,5591%
T2T2	60,3448%	60,3448%	67,2414%	61,2069%

Na Tabela 25 pode-se observar que todos os modelos acertam acima de 50%, com destaque para os modelos DT e RF que acertam 73,2258% e 79,8246% respectivamente, que são as duas maiores taxas de acerto para os experimentos. Em média os modelos acertam cerca de 70%.

Como o RF conseguiu acertar 79,8246%, isso é um indicativo de que há alguma relação dos atributos que classifique corretamente as arestas indiferentes e não indiferentes para o experimento T1T2, que os outros modelos não conseguiram encontrar.

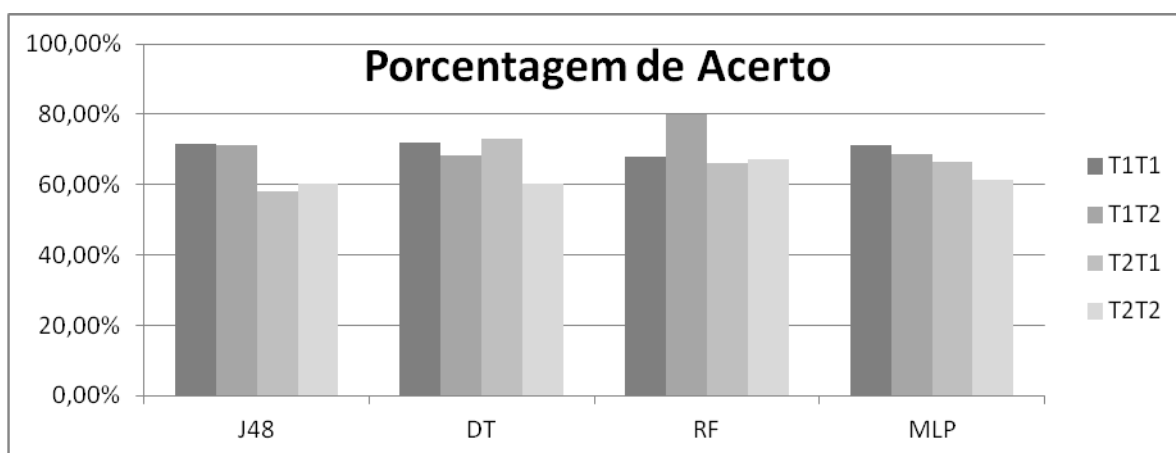


Figura 3 - Gráfico de barras da porcentagem de acerto na fase de teste, com arestas balanceadas em indiferentes e não indiferentes.

4.2.3 Predição de Arestas Positivas e Não Positivas

Os experimentos realizados indicam que é possível distinguir arestas positivas e não positivas dentro de um grupo social e entre grupos distintos.

Tabela 27 - Porcentagem de acerto na fase de teste, com arestas balanceadas em positivas e não positivas.

	J48	DT	RF	MLP
T1T1	57,6687%	60,1227%	60,1227%	58,8957%
T1T2	68,4211%	66,6667%	66,0819%	66,0819%
T2T1	66,3441%	68,9247%	66,4516%	61,0968%
T2T2	66,0377%	62,2642%	67,9245%	67,9245%

Na Tabela 26 pode-se observar que todos os modelos acertam acima de 50%, mas diferente do que aconteceu nos experimentos com arestas “negativas e não negativas” e “indiferentes e não indiferentes”, nenhum modelo possui taxa de acerto muito diferente dos demais (Figura 3). Em média os modelos acertam cerca de 64,4380%.

O comportamento estável dos modelos é um indicativo de que as relações positivas e não positivas são facilmente detectadas e tendem a ser comum para diversos grupos sociais, diferente das arestas negativas e não negativas cujos testes indicaram que elas tendem a ser diferentes entre grupos sociais. Comparando com as arestas indiferentes e não indiferentes, a taxa de acerto é menor, indicando que, apesar das arestas positivas e não positivas serem mais estáveis para os modelos de predição de *links*, as arestas indiferentes e não indiferentes são caracterizadas com maior facilidade.

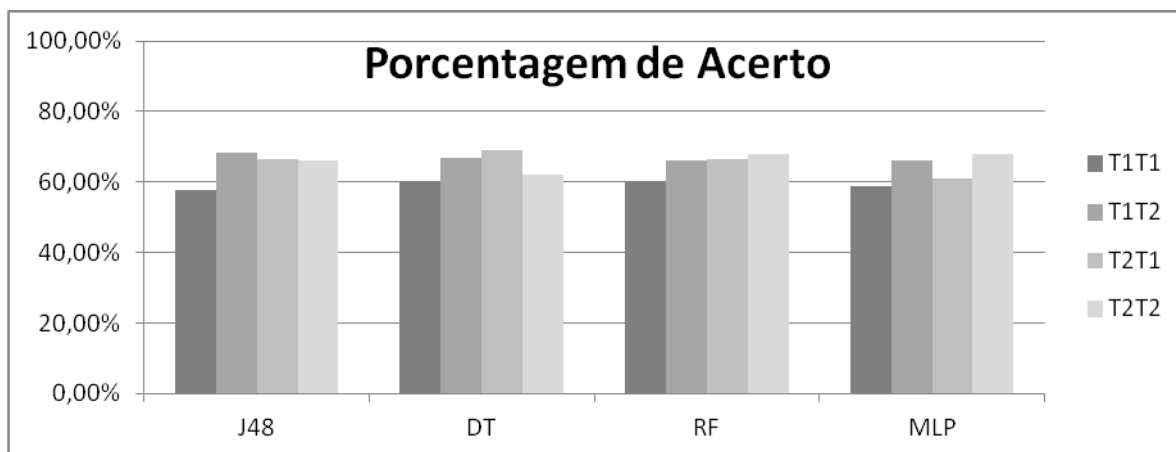


Figura 4 - Gráfico de barras da porcentagem de acerto na fase de teste, com arestas balanceadas em positivas e não positivas.

4.2.4 Predição de Arestas Negativas e Indiferentes

Os experimentos realizados indicam que em alguns casos é possível distinguir arestas negativas e indiferentes para um grupo social, ou seja, é possível prever suas ocorrências, mas que essa distinção não pode ser aplicada, também em alguns casos, para grupos sociais distintos.

Tabela 28 - Porcentagem de acerto na fase de teste, com arestas balanceadas em negativas e indiferentes.

	J48	DT	RF	MLP
T1T1	61,2500%	62,5000%	63,7500%	65,0000%
T1T2	56,7961%	61,1650%	57,2816%	56,7961%
T2T1	38,1462%	63,1016%	46,1141%	46,4528%
T2T2	51,8519%	40,7407%	59,2593%	48,1481%

Pode-se observar, pela Tabela 27, que todos os modelos conseguem distinguir, com acerto acima de 60%, as arestas para o experimento T1T1, Isso é um indicativo de que as arestas negativas e indiferentes de um grupo de pessoas tendem a ter características em comum, contudo este comportamento não é observado nos experimentos T2T2, pois apenas o modelo RF conseguiu ter taxa de acerto próximo a 60%, já os modelos J48 e MLP ficam próximos aos 50% e o DT acerta apenas 40%.

Por outro lado, o experimento T1T2 mostra que os modelos conseguem extrapolar a detecção de arestas negativas e indiferentes para diferentes grupos sociais, mesmo que, para este caso, a taxa de acerto seja no máximo 61,1650%, Já o experimento T2T1 mostra que os modelos, em geral, não conseguem extrapolar a detecção das arestas para diferentes grupos sociais com exceção do DT que, como na predição das arestas negativas e não negativas, teve taxa de acerto próximo a 63%.

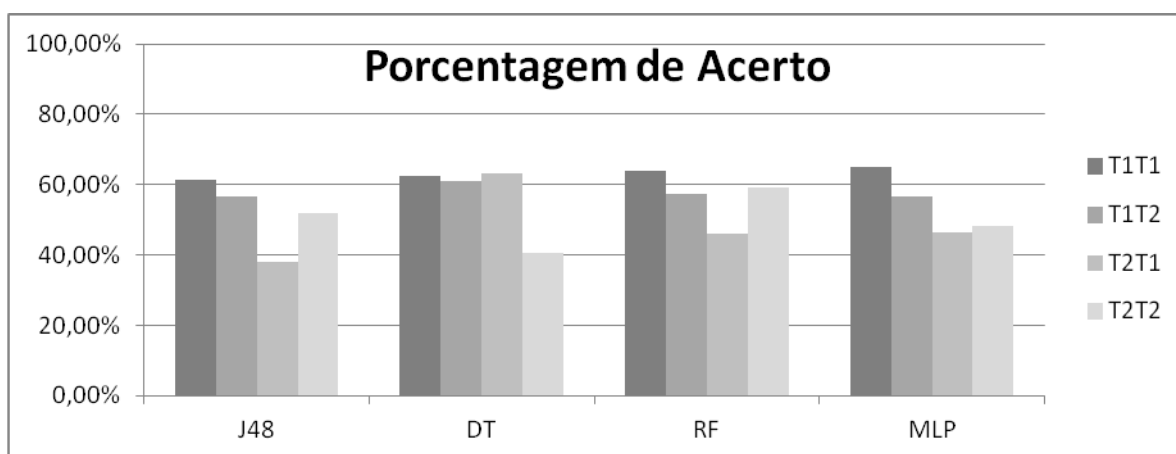


Figura 5 - Gráfico de barras da porcentagem de acerto na fase de teste, com arestas balanceadas em negativas e indiferentes.

4.2.5 Predição de Arestas Negativas e Positivas

Os experimentos realizados indicam que é possível distinguir arestas negativas e positivas para um grupo social, ou seja, é possível prever suas ocorrências, mas que essa distinção não pode ser aplicada para grupos sociais distintos.

Tabela 29 - Porcentagem de acerto na fase de teste, com arestas balanceadas em negativas e positivas.

	J48	DT	RF	MLP
T1T1	56,0976%	52,4390%	59,7561%	57,3171%
T1T2	46,2857%	36,0000%	50,8571%	51,4286%
T2T1	45,1943%	68,3027%	50,2045%	43,9469%
T2T2	70,3704%	62,9630%	77,7778%	81,4815%

Pode-se observar, pela Tabela 28, que todos os modelos conseguem distinguir, com acerto acima de 50%, as arestas para os experimentos T1T1 e T2T2. Isso é um indicativo de que as arestas negativas e positivas de um grupo de pessoas tendem a ter características em comum.

Por outro lado, os experimentos T1T2 e T2T1 mostram que, em geral, os modelos não conseguem extrapolar a detecção de arestas negativas e positivas para diferentes grupos sociais. Analogamente aos cenários de previsão de arestas “negativas e não negativas” e “negativas e indiferentes”, o modelo DT foi o único que conseguiu taxa de acerto acima de 50%, atingindo 68,3027%.

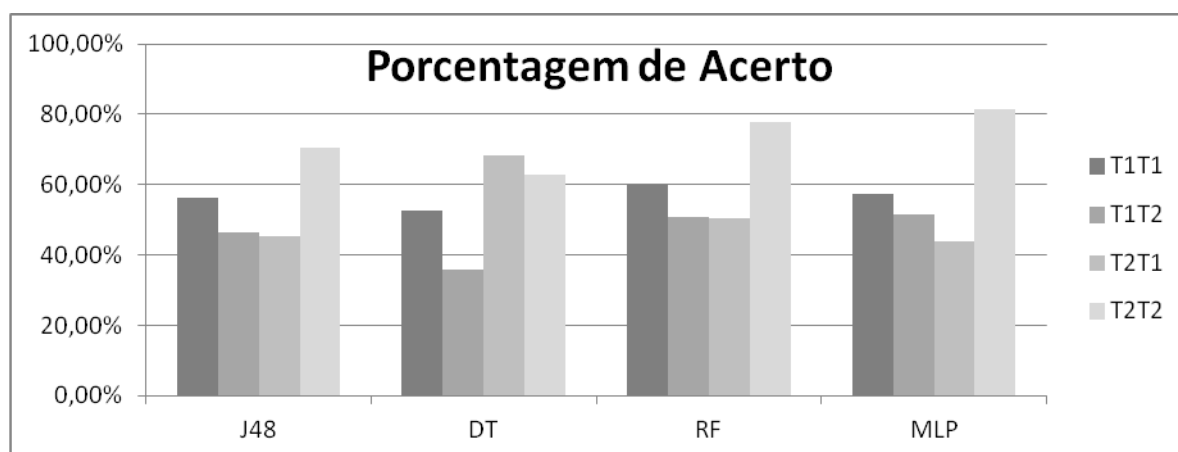


Figura 6 - Gráfico de barras da porcentagem de acerto na fase de teste, com arestas balanceadas em negativas e positivas.

4.2.6 Predição de Arestas Indiferentes e Positivas

Os experimentos realizados indicam que é possível distinguir arestas indiferentes e positivas dentro de um grupo social e entre grupos distintos.

Tabela 30 - Porcentagem de acerto na fase de teste, com arestas indiferentes e positivas.

	J48	DT	RF	MLP
T1T1	77,0909%	72,7273%	71,6364%	75,6364%
T1T2	69,6370%	69,9670%	65,0165%	65,0165%
T2T1	68,7654%	68,2716%	71,3580%	65,4321%
T2T2	73,7864%	65,0485%	72,8155%	63,1068%

Na Tabela 29 pode-se observar que todos os modelos acertam acima de 60%, analogamente ao comportamento da previsão de arestas positivas e não positivas, mas com uma porcentagem de acerto maior. Note que nenhum modelo possui taxa de acerto muito diferente dos demais (Figura 3), Em média os modelos acertam cerca de 70%.

O comportamento estável dos modelos é um indicativo de que as relações indiferentes e positivas são facilmente detectadas e tendem a ser comum para diversos grupos sociais, diferente das arestas negativas e não negativas cujos testes indicaram que elas tendem a ser diferentes entre grupos sociais. Comparando com as arestas indiferentes e não indiferentes, a taxa de acerto é maior, indicando que, apesar das arestas indiferentes e positivas serem mais estáveis para os modelos de predição de *links*, as arestas indiferentes e positivas são caracterizadas com maior facilidade.

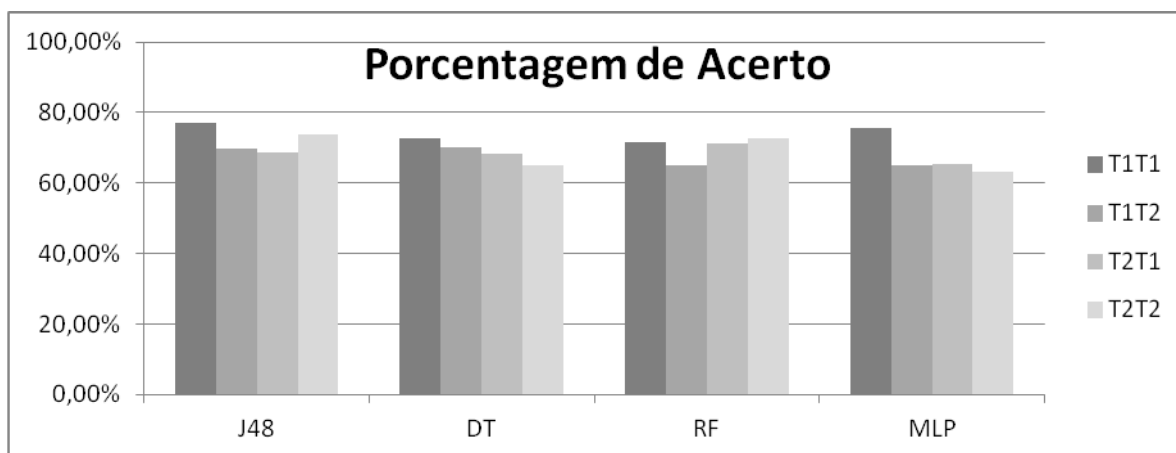


Figura 7 - Gráfico de barras da porcentagem de acerto na fase de teste, com arestas balanceadas em indiferentes e positivas.

4.3 Ranking dos Atributos das Arestas

4.3.1 Atributos das Arestas do Grafo Turma1

O ranking das arestas representa, em ordem decrescente de importância, o quanto relevante um atributo foi no processo de classificação das arestas.

Tabela 31 - Atributos para o grafo composto de arestas negativas e não negativas.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Grupos	1º	Grupos
2º	Interesses	2º	Interesses
3º	Amizades	3º	Amizades

Tabela 32 - Atributos para o grafo composto de arestas negativas e indiferentes.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Grupos	1º	Grupos
2º	Amizades	2º	Interesses
3º	Comentários em Status	3º	Comentários em <i>Stauts</i>
4º	<i>Status</i> Curtidos	4º	<i>Status</i> Curtidos
5º	Comentários em <i>Links</i>	5º	Comentários em <i>Links</i>

Tabela 33 - Atributos para o grafo composto de arestas negativas e positivas.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Grupos	1º	Grupos
2º	Comentários em Fotos	2º	Comentários em Fotos
3º	Marcações	3º	Marcações
4º	Amizades	4º	Amizades
5º	Comentários em Álbuns	5º	Comentários em Álbuns

Como observado nas tabelas Tabela 34, Tabela 35 e Tabela 36, o atributo mais relevante para a classificação das arestas “negativas e não negativas”, “negativas e indiferentes” e “negativas e positivas” é “Grupos”, que representa a quantidade de grupos em comum entre dois vértices quaisquer. Observe que, de acordo com o tipo de aresta que se analisa, uma quantidade diferente de atributos são utilizados e a ordem de importância são alterados, isso indica que um dado atributo é importante para classificar uma aresta, mas não será necessariamente igualmente importante para classificar outros tipos de arestas.

Tabela 34 - Atributos para o grafo composto de arestas indiferentes e positivas.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Amizades	1º	Amizades
2º	Grupos	2º	Grupos
3º	Conversas	3º	Conversas
4º	Comentários em Fotos	4º	Comentários em Fotos
5º	Comentários em <i>Status</i>	5º	Comentários em <i>Status</i>
6º	Marcações	6º	Marcações
7º	Comentários em <i>Links</i>	7º	Comentários em <i>Links</i>
8º	<i>Status</i> Curtidos	8º	<i>Links</i> Curtidos
9º	<i>Link</i> Curtidos	9º	<i>Status</i> Curtidos
10º	Comentários em Álbuns	10º	Comentários em Álbuns

Tabela 35 - Atributos para o grafo composto de arestas indiferentes e não indiferentes.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Amizades	1º	Amizades
2º	Conversas	2º	Conversas
3º	Grupos	3º	Grupos
4º	Comentários em Fotos	4º	Comentários em <i>Status</i>
5º	Comentários em <i>Status</i>	5º	Comentários em Fotos
6º	Marcações	6º	Comentários em <i>Links</i>
7º	Comentários em <i>Links</i>	7º	Marcações
8º	<i>Links</i> Curtidos	8º	<i>Links</i> Curtidos
9º	<i>Status</i> Curtidos	9º	<i>Status</i> Curtidos
10º	Interesses	10º	Interesses
11º	Comentários em Álbuns	11º	Comentários em Álbuns

Tabela 36 - Atributos para o grafo composto de arestas positivas e não positivas.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Amizades	1º	Amizades
2º	Grupos	2º	Grupos
3º	Conversas	3º	Conversas
4º	Comentários em Fotos	4º	Comentários em Fotos
5º	<i>Links</i> Curtidos	5º	<i>Links</i> Curtidos
6º	Filmes	6º	Marcações
7º	Marcações	7º	Filmes
8º	Comentários em Álbuns	8º	Comentários em Álbuns

Nas tabelas Tabela 37, Tabela 38 e Tabela 39, o atributo mais relevante para a classificação das arestas “indiferentes e não indiferentes”, “indiferentes e positivas” e “positivas e não positivas” é “Amizades”, que indica se dois vértices quaisquer do grafo são amigos ou não. Analogamente às tabelas Tabela

34, Tabela 35 e Tabela 36, de acordo com o tipo de aresta que se analisa, uma quantidade diferente de atributos são utilizados e a ordem de importância são alterados, isso indica que um dado atributo é importante para classificar uma aresta, mas não será necessariamente igualmente importante para classificar outros tipos de arestas.

Os atributos que não aparecem nas tabelas Tabela 34, Tabela 35, Tabela 36, Tabela 37, Tabela 38 e Tabela 39, não são relevantes no processo de classificação, isso significa que utilizá-los ou não é indiferente para o resultado.

4.3.2 Atributos das Arestas do Grafo Turma2

Pelos resultados obtidos podemos observar que o comportamento dos atributos do grafo Turma2 é diferente quando comparado ao grafo Turma1.

Tabela 37 - Atributos para o grafo composto de arestas negativas e não negativas.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	<i>Status Curtidos</i>	1º	<i>Status Curtidos</i>
2º	<i>Amizades</i>	2º	<i>Amizades</i>

Como observado na Tabela 40, o atributo mais relevante para a classificação das arestas “negativas e não negativas” é “Status Curtidos”, que representa quantos Status o vértice de origem curtiu do vértice de destino.

Tabela 38 - Atributos para o grafo composto de arestas indiferentes e não indiferentes.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	<i>Conversas</i>	1º	<i>Amizades</i>
2º	<i>Amizades</i>	2º	<i>Conversas</i>
3º	<i>Links Curtidos</i>	3º	<i>Links Curtidos</i>
4º	<i>Comentários em Status</i>	4º	<i>Comentários em Status</i>
5º	<i>Status Curtidos</i>	5º	<i>Status Curtidos</i>
6º	<i>Comentário em Fotos</i>	6º	<i>Comentários em Fotos</i>
7º	<i>Grupos</i>	7º	<i>Comentários em Links</i>
8º	<i>Comentários em Links</i>	8º	<i>Filmes</i>
9º	<i>Filmes</i>	9º	<i>Marcações</i>
10º	<i>Marcações</i>	10º	<i>Grupos</i>

Tabela 39 - Atributos para o grafo composto de arestas indiferentes e positivas.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Amizades	1º	Conversas
2º	Conversas	2º	Amizades
3º	<i>Links</i> Curtidos	3º	<i>Links</i> Curtidos
4º	Comentários em <i>Status</i>	4º	Comentários em <i>Status</i>
5º	<i>Status</i> Curtidos	5º	<i>Status</i> Curtidos
6º	Comentário em Fotos	6º	Comentário em Fotos
7º	Comentários em <i>Links</i>	7º	Comentários em <i>Links</i>
8º	Marcações	8º	Marcações
9º	Grupos	9º	Grupos

Nas tabelas Tabela 41 e Tabela 42 o atributo mais relevante para a classificação das arestas “indiferentes e não indiferentes” e “indiferentes e positivas” não é bem definido, pois ele varia entre os métodos *Info Gain* e *Chi Square*. Na Tabela 41 o atributo mais relevante para a classificação das arestas é “Conversas” para o *Info Gain* e “Amizades” para o *Chi Square*, De maneira similar, a Tabela 42 tem comportamento semelhante.

Tabela 40 - Atributos para o grafo composto de arestas positivas e não positivas.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Amizades	1º	Amizades
2º	Conversas	2º	Conversas
3º	<i>Status</i> Curtidos	3º	<i>Status</i> Curtidos
4º	<i>Links</i> Curtidos	4º	Comentários Fotos
5º	Comentários em Fotos	5º	Comentários em <i>Status</i>

Tabela 41 - Atributos para o grafo composto de arestas negativas e indiferentes.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Amizades	1º	Amizades

Tabela 42 - Atributos para o grafo composto de arestas negativas e positiva.

<i>Info Gain</i>		<i>Chi Square</i>	
1º	Amizades	1º	Amizades
2º	<i>Status</i> Curtidos	2º	Conversas
3º	Conversas	3º	<i>Status</i> Curtidos
4º	Músicas	4º	Músicas
5º	Comentários em Fotos	5º	Comentários em Fotos

As tabelas Tabela 43, Tabela 44 e Tabela 45 mostram quem o atributo mais relevante para a classificação das arestas “positivas e não positivas”, “negativas e indiferentes” e “negativas e positivas” é “Amizades”.

Note que para o caso de arestas “negativas e indiferentes”, Tabela 44, o único atributo relevante para classificá-las é “Amizades”. Isso pode explicar porquê os modelos tiveram dificuldade em avaliar arestas negativas.

5 Conclusões e Trabalhos Futuros

Este capítulo apresenta as conclusões obtidas através das análises e predições de arestas;

Os testes feitos indicam que a capacidade de previsão de arestas depende em parte do método utilizado para predição de dados e em parte do tipo de dado que se quer prever, bem como a qualidade destes dados.

Os resultados mostraram que os modelos de aprendizagem de máquina possuem taxa de acerto entre 70% e 80% para a maior parte dos casos estudados das arestas indiferentes e positivas, já as arestas negativas se demonstraram extremamente complexas de serem previstas, principalmente quando a previsão dava-se tendo grafos diferentes para as fases de treinamento e teste dos métodos.

Também foi possível concluir, através de todos os experimentos feitos, que a utilização dos métodos de aprendizagem de máquina influenciam o resultado das predições de arestas. Como observado na detecção de arestas negativas e indiferentes tendo como base de treinamento o grafo Turma2 e teste no grafo Turma1, apenas a DT foi capaz de obter taxa de acerto superior a taxa de erro. Outro exemplo de sucesso do método DT foi na predição de arestas negativas e positivas tendo como base de treinamento o grafo Turma2 e teste no grafo Turma1.

Em trabalhos futuros espera-se ampliar a quantidade de atributos provenientes do Facebook para promover melhor caracterização dos relacionamentos profissionais, bem como avaliar se, de fato, há ou não relevância social para formar equipes de trabalho.

6 Referências Bibliográficas

Alsaleh, S.; Nayak, R. & Xu, Y. (2012), Grouping people in social networks using a weighted multi-constraints clustering method, *in* 'Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on'.

Borgatti, S. P.; Mehra, A.; Brass, D. J. & Labianca, G. (2009), 'Network analysis in the social sciences', *science* **323**(5916), 892--895.

Breiman, L. (2001), 'Random Forests', *Machine Learning* **45**(1), 5-32.

Brzozowski, M. J.; Hogg, T. & Szabo, G. (2008), Friends and foes: ideological social networking, *in* 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', pp. 817--820.

Burke, M. & Kraut, R. (2008), Mopping up: modeling wikipedia promotion decisions, *in* 'Proceedings of the 2008 ACM conference on Computer supported cooperative work', pp. 27--36.

Bustos, D. M. (1979), *O teste sociométrico*, Brasiliense.

Facchetti, G.; Iacono, G. & Altafini, C. (2011), 'Computing global structural balance in large-scale signed social networks', *Proceedings of the National Academy of Sciences* **108**(52).

Figueiredo, T. F. (2011), 'Haffner: Um Sistema de Informação para auxiliar gestores na formação de equipes de trabalho cooperativo.' XXIV Congresso Internacional de Administração'.

Fortunato, S. (2010), 'Community detection in graphs', *Physics Reports* **486**(3).

Furuta, T.; Sasaki, M.; Ishizaki, F.; Suzuki, A. & Miyazawa, H. (2006), 'A new cluster formation method for sensor networks using facility location theory', Technical report, Tech. Rep. NANZAN-TR-2006-01, Nanzan Academic Society Mathematical Sciences and Information Engineering, Nagoya, Japan.

Garton, L.; Haythornthwaite, C. & Wellman, B. (1997), 'Studying Online Social Networks', *Journal of Computer-Mediated Communication*.

Guha, R.; Kumar, R.; Raghavan, P. & Tomkins, A. (2004), Propagation of trust and distrust, *in* 'Proceedings of the 13th international conference on World Wide Web', pp. 403--412.

Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P. & Witte, I. H.11, ed. (2009), *The WEKA Data Mining Software: An Update*, SIGKDD Explorations.

Heider, F. (1946), 'Attitudes and cognitive organization', *The Journal of psychology*

21(1).

Kohavi, R. (1995), The Power of Decision Tables, in '8th European Conference on Machine Learning', Springer, , pp. 174-189.

Kunegis, J.; Lommatzsch, A. & Bauckhage, C. (2009), The slashdot zoo: mining a social network with negative edges, in 'Proceedings of the 18th international conference on World wide web', pp. 741--750.

Lü, L.; Jin, C.-H. & Zhou, T. (2009), 'Similarity index based on local paths for link prediction of complex networks', *Physical Review E* **80**(4), 046122.

Lü, L. & Zhou, T. (2011), 'Link prediction in complex networks: A survey', *Physica A: Statistical Mechanics and its Applications* **390**(6), 1150--1170.

Lampe, C. A.; Johnston, E. & Resnick, P. (2007), Follow the reader: filtering comments on slashdot, in 'Proceedings of the SIGCHI conference on Human factors in computing systems', pp. 1253--1262.

Leskovec, J.; Huttenlocher, D. & Kleinberg, J. (2010), Signed networks in social media, in 'Proceedings of the SIGCHI Conference on Human Factors in Computing Systems', pp. 1361--1370.

Leskovec, J.; Huttenlocher, D. & Kleinberg, J. (2010), Predicting positive and negative links in online social networks, in 'Proceedings of the 19th international conference on World wide web', pp. 641--650.

Mansson, D. H. & Myers, S. A. (2011), 'An initial examination of college students' expressions of affection through Facebook', *Southern Communication Journal* **76**(2), 155--168.

Marineau, R. F. (1992), *Jacob Levy Moreno, 1889-1974 pai do psicodrama, da sociometria e da psicoterapia de grupo.*, Editora Agora.

Massa, P. & Avesani, P. (2005), Controversial users demand local trust metrics: An experimental study on epinions. com community, in 'Proceedings of the National Conference on artificial Intelligence', pp. 121.

Molina, J. & Aguilar, C. (2004), 'Redes sociales y antropología: Discursos étnicos y redes personales entre jóvenes de Sarajevo', *Larrea, KC, Estrada i Bonell, F.(eds.). Antropologia en un mundo en transformación. Barcelona: Universitat de Barcelona.*

Moreno, J. L. (1934), 'Who shall survive?: A new approach to the problem of human interrelations.', .

Newman, M. E. (2003), 'The structure and function of complex networks', *SIAM review* **45**(2), 167--256.

de Oliveira, H. C. B.; Martinez, M. R.; Salgado, R. M.; da Silva, L. E. & Figueiredo, T. F. (2010), 'Clusterização em Redes Sociais Através do Simulated Annealing Não

Monotônico', *Simpósio Brasileiro de Pesquisa Operacional*.

Pellegrini, M.; Haynor, D. & Johnson, J. M. (2004), 'Protein interaction networks', *Expert review of proteomics* **1**(2).

Quinlan, R. (1993), *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA.

Schaeffer, S. E. (2007), 'Graph clustering', *Computer Science Review* **1**(1).

Symeonidis, P.; Tiakas, E. & Manolopoulos, Y. (2010), Transitive node similarity for link prediction in social networks with positive and negative links, in 'Proceedings of the fourth ACM conference on Recommender systems', pp. 183--190.

Terveen, L. & McDonald, D. W. (2005), 'Social matching: A framework and research agenda', *ACM Transactions on Computer-Human Interaction (TOCHI)* **12**(3).

Watts, D. J. (1999), *Small worlds: the dynamics of networks between order and randomness*, Princeton university press.

Zhao, X.; Li, G.; Yuan, J.; Chen, X. & Li, Z. (2012), 'Relationship strength estimation for online social networks with the study on Facebook', *Neurocomputing*.

7 Apêndice

7.1 Apêndice I

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

1. NOME DA PESQUISA: Utilização da Sociometria como Ferramenta de Clusterização através de Redes sociais

2. PESQUISADORES RESPONSÁVEIS: Prof^a Dra, Maria Regina Martinez e Prof Dr, Humberto César Brandão de Oliveira, Graduando do curso de Bacharel em Ciência da Computação da Universidade Federal de Alfenas - Unifal-MG: Wilson Sasaki Junior

3. INSTITUIÇÃO PROMOTORA: Universidade Federal de Alfenas - UNIFAL/MG

4. OBJETIVO, JUSTIFICATIVA E PROCEDIMENTO DE COLETA DE DADOS: O objetivo desta pesquisa é coletar um sociograma que será utilizado como base de conhecimento para a tentativa de identificar padrões de dados na rede social Facebook a fim de construir o mesmo sociograma com informações exclusivamente obtidas do Facebook. Esses sociogramas possibilitarão a criação de um *software* voltado para a construção de agrupamentos harmônicos de pessoas. Mediante assinatura de termos de sigilo e confidencialidade, o *software* por esta pesquisa desenvolvido poderá ser auditado em qualquer fase do seu desenvolvimento. A participação é voluntária e a coleta de dados se dará pela resposta de um questionário e acesso a um aplicativo disponibilizado no Facebook, esta participação não implicará em qualquer risco, despesa ou desconforto. O sigilo da identidade dos participantes será assegurado. Dados de terceiros disponibilizados pelo Facebook não serão coletados, pois o *software* será programado para a filtragem desses dados, de modo que somente serão coletados dados dos sujeitos que tiverem aceitado

participar da pesquisa. A coleta de dados digitais obedecerá a Política de Privacidade prevista no Facebook,

5. BENEFÍCIOS E DIVULGAÇÃO DOS RESULTADOS: Não é previsto nenhum benefício financeiro ou de outra natureza aos sujeitos da pesquisa. Os autores reservam-se o direito de publicar e apresentar os dados em meios de divulgação científica como meio de gerar informações importantes para o desenvolvimento da área de conhecimento,

Pesquisadores responsáveis:

Profa Dra. Maria Regina Martinez

Prof Dr. Humberto César Brandão de Oliveira

Acadêmico Wilson Sasaki Junior

7.2 Apêndice II

Declaração de Aceite de Participação

DECLARO para fins de participação em pesquisa, na condição de sujeito da mesma, que fui devidamente esclarecido do projeto intitulado: UTILIZAÇÃO DA SOCIOMETRIA COMO FERRAMENTA DE CLUSTERIZAÇÃO ATRAVÉS DE REDES SOCIAIS, **sob a responsabilidade dos pesquisadores Maria Regina Martinez, Humberto César Brandão de Oliveira e Wilson Sasaki Junior**, quanto aos seguintes aspectos:

- Justificativa, objetivos e procedimentos que serão utilizados na pesquisa;
- Desconfortos e riscos possíveis;
- Liberdade de me recusar a participar da pesquisa, sem penalização alguma e sem prejuízo;
- Garantia de sigilo quanto aos dados confidenciais envolvidos na pesquisa assegurando-me absoluta privacidade;
- Esclarecido que, talvez, não terei benefício direto com a pesquisa, porém poderei contribuir com outras pessoas, pois se espera que este estudo traga informações importantes e que o conhecimento produzido possa ser divulgado em eventos e revistas científicas;

Após convenientemente esclarecido, consisto, voluntariamente, em participar desta pesquisa.

NOME: _____

RG: _____

ASSINATURA: _____

7.3 Apêndice III

Teste Sociométrico

Assinale com um círculo a resposta para a pergunta: *Com quais das pessoas listadas você gostaria de trabalhar em equipe?*

*ASSINALE COM UMA SETA O SEU NOME

Nome do Voluntário	Sim	Não	Indiferente
Voluntário 1			
Voluntário 2			
Voluntário 3			