

Mineração de Dados em Redes Sociais e Textuais para Identificação de Papéis em Sala de Aula

Viviane Soares Rodrigues Silva

Programa de Pós Graduação em Informática -
Universidade Federal do Rio de Janeiro - PPGI/UFRJ
- RJ - Brasil
vivianerodrig@ppgi.ufrj.br

Robson Costa de Castro

Colégio Pedro II – Campus Tijuca II
CEP 20550-012 – Rio de Janeiro – RJ – Brasil
prof.robinho@gmail.com

José Otávio Pompeu e Silva

Programa de Pós Graduação em Informática -
Universidade Federal do Rio de Janeiro - PPGI/UFRJ
- RJ - Brasil
pompeuesilva@nce.ufrj.br

Claudia L. R. Motta

Programa de Pós Graduação em Informática -
Universidade Federal do Rio de Janeiro - PPGI/UFRJ
- RJ - Brasil
claudiam@nce.ufrj.br

ABSTRACT

The social and textual networks topology can reveal patterns or trends that help us to understand its dynamic. This article describes an ongoing research about dynamic relationships in classroom using tools for network analysis. We describe an experiment that generates two kinds of networks, social and textual, through exchanges of messages during a dynamic in the classroom. Data mining techniques are applied on those networks in order to identify roles played by students during conversations.

RESUMO

A topologia de sistemas como as redes sociais e as redes textuais pode revelar padrões ou tendências que auxiliam na compreensão de sua dinâmica. Este artigo é parte de uma pesquisa em andamento que estuda as dinâmicas relações em salas de aula presenciais utilizando as ferramentas para análise de redes. Este trabalho traz a descrição de um experimento que, através de trocas de mensagens durante uma dinâmica em sala de aula presencial, gera duas redes, social e textual. Uma técnica de mineração de dados sobre medidas extraídas das redes visa a identificação de papéis desempenhados pelos alunos.

Descritor de Categorias e Assuntos

D.3.3 [Computers and Education]: Computers use in education – collaborative learning, social networks

General Terms

Experimentation

Palavras Chave

Mineração de Dados, Redes Sociais, Redes Textuais, Sala de Aula

1. INTRODUÇÃO

As questões trazidas neste trabalho surgiram durante uma calorosa discussão entre os professores de uma escola pública do Rio de Janeiro, enquanto participavam de um Conselho de Classe. A questão polêmica era o uso de aparelhos celulares e equipamentos eletrônicos na sala de aula.

Na ocasião, destacou-se que tem crescido o número de alunos que carregam consigo smartphones de última geração e que, naquele ano, esta mesma escola iniciou a prática de distribuir *tablets* para todos os alunos do Ensino Médio. Se por um lado, havia o entendimento de que existia a necessidade de inclusão de recursos tecnológicos em suas metodologias de ensino, por outro, o uso de tais dispositivos móveis causava preocupação aos professores quanto à dificuldade que encontravam em manter a atenção e a participação dos alunos.

De fato, a realidade na escola ainda hoje é esta: nossos alunos fazem uso de tecnologia o dia inteiro. Fazem parte de redes sociais, trocando todo tipo de informação com dezenas de pessoas diferentes. Conhecem e fazem uso de aplicativos que facilitam suas vidas. Dedicam horas a quebrar desafios, cumprir missões e alcançar novos índices nos mais diversos jogos eletrônicos. A escola, por sua vez, oferece ferramentas tecnológicas como *tablets*, laboratórios de informática e acesso a internet, no entanto, na hora da aula este mesmo aluno precisa desligar tudo, pegar seu caderno e ficar atento ao que está sendo dito.

Diante dos novos desafios propostos pela atividade de ensino no século XXI, a sala de aula na modalidade presencial, necessita de uma renovação em sua dinâmica. Os alunos de hoje, nativos da era digital, estão ávidos por aulas dinâmicas, atrativas, interativas e enriquecedoras, mas em geral encontram as tradicionais aulas em frente a um quadro branco onde, na maior parte do tempo, o professor é o emissor e o aluno o receptor das informações [10]

Somado a isto, é comum encontrarmos alunos não interagindo ou não participando ativamente das aulas, seja por vergonha ou medo de falar em público, com receio de receber críticas ou ser motivo de

piadas por parte dos seus colegas ou até mesmo do professor [9]. Esta situação pode ser um dos motivos que levam o aluno ao desestímulo e a dificuldade em enxergar a verdadeira aplicação dos conteúdos no seu cotidiano [10].

Neste contexto, as TIC (Tecnologias da Informação e Comunicação), assim como o aumento exponencial da informação, induzem o professor a uma nova organização do trabalho, sendo necessário estar aberto aos novos conhecimentos e estar preparado a incentivar a colaboração entre alunos. Logo, a instituição de ensino precisa mudar o seu papel frente a essas possibilidades. Deste modo, com as TIC, um conjunto de novas atividades pode ser promovido a fim de enriquecer o repertório em sala de aula, unindo dados científicos e culturais de diversas naturezas [22]

O uso de TIC na educação, de acordo a Unesco [33] e o Ministério da Educação [25], representa um benefício para os alunos, pois ao longo da sua educação eles se sentirão à vontade com a tecnologia na escola, bem como nas suas futuras áreas de trabalho. Além disso, como uma quantidade significativa do alunado tem aproximação com o uso de tecnologias, a compreensão dos materiais usados nas disciplinas são melhores compreendidos.

No sentido de fazer o uso destas ferramentas tecnológicas uma prática cotidiana em sala de aula, é necessário que se proponha alguma solução que atraia o interesse do professor. Que não exija tanto de seu tempo para que aprenda sobre a tecnologia a ser introduzida, e que traga retorno positivo e concreto para o ensino, incluindo os alunos, o professor e a escola.

Neste cenário, é interessante que sejam encontradas soluções que estimulem trocas e incentivem a colaboração, auxiliando o professor a identificar diferentes potenciais de contribuição que seus alunos possuem [31].

Diante do exposto, este trabalho propõe experimentar o uso de uma ferramenta para bate-papo durante uma aula presencial na escola. Alunos e professor estariam livres para conversar uns com os outros, via ferramenta, a fim de resolver exercícios propostos pelo professor de uma maneira conjunta. No entanto, devido a limitações de acesso a internet via rede wifi do colégio, neste primeiro momento, a atividade aconteceria no laboratório de informática, onde a maioria dos alunos utilizariam desktops e outros alunos, através de acesso via 3G, fariam uso de seus *smartphones e tablets*.

Acredita-se que, com esta simples mudança na dinâmica da aula, promove-se a introdução de tecnologia, cria-se uma oportunidade para um número maior de alunos participarem da resolução da tarefa proposta, abre-se um canal propício para alunos ensinarem uns aos outros além de estimular o uso dos dispositivos e máquinas oferecidos pela escola.

Além disso, acontece de forma automática, o registro de informações úteis e não facilmente observáveis: todo o conteúdo transmitido naquele dia, como se deu a participação individual dos alunos quando utilizaram o recurso de bate papo naquela aula específica e como fica a rede de relacionamentos estabelecidos durante as trocas realizadas.

Desta forma, através de uma análise sobre a rede social formada, torna-se possível identificar alunos que mais facilmente iniciam novos laços de relacionamento, aqueles que são mais requisitados por outros alunos, aqueles que têm a habilidade em atuar como pontes ligando grupos distintos de amigos, ou ainda aqueles que

têm a tendência em incentivar que dois de seus amigos se tornem amigos também. Revelando assim algumas características sobre o comportamento dos alunos durante aquela aula, e auxiliando ao professor e a escola a desenvolverem um olhar um pouco mais individualizado para o aluno.

Outra questão que surge é se seria possível identificar o papel que cada aluno exerce na turma. Neste caso, a hipótese é que as características particulares sobre o comportamento dos alunos durante as trocas de mensagens podem ser fonte de dados úteis para a separação destes alunos em grupos distintos. Havendo grupos distintos estes podem representar papéis desempenhados pelos alunos presentes naquela turma.

Este artigo apresenta uma proposta de modelagem da dinâmica de trocas de mensagens através dos conceitos de redes. Considerando a possibilidade de dois tipos de sistemas modelados como uma rede: uma rede social construída através das conexões que os alunos fazem quando enviam ou recebem uma mensagem, e as redes textuais, resultantes da organização dos textos das mensagens como uma rede de palavras.

Após a separação dos alunos em grupos, de acordo com as medidas extraídas das redes, são feitas a análise de cada conjunto de alunos e uma avaliação se representam papéis diferentes. Considerando também o conteúdo dos textos trocados e a percepção dos professores sobre a participação dos alunos durante suas aulas.

Este artigo está organizado da seguinte maneira: esta seção introdutória que traz a motivação, a hipótese e objetivo apresentados neste trabalho. A Seção 2 aborda o tema Mineração de Dados Educacionais, trazendo trabalhos relacionados a este nas áreas de redes sociais aplicadas na educação e também trabalhos relacionados em redes textuais. A Seção 3 traz a metodologia empregada na proposta. A Seção 4 apresenta os resultados do experimento realizado e analisa a validade da hipótese levantada. Por fim, a Seção 5 conclui e aponta trabalhos futuros.

2. MINERAÇÃO DE DADOS EDUCACIONAIS

Nos últimos anos têm-se visto um crescente interesse nas pesquisas que utilizam técnicas de Mineração de Dados na Educação. Este campo emergente, chamado Mineração de Dados Educacionais, concentra-se na descoberta de conhecimento através dados originados em ambientes educacionais [21].

Baker *et. al.* [5], trouxeram um panorama a respeito dos métodos e aplicações que vêm influenciando a pesquisa e a prática da educação em vários países. Em particular, revela que, apesar dos esforços de seus pesquisadores, essa área ainda é pouco explorada no Brasil.

Dentre as possíveis técnicas de mineração de dados para classificação, Baradwaj *et. al.* [7] optaram pela árvore de decisão como o método utilizado. Informações a respeito da presença, avaliações em aula, apresentação de seminários e alguns atributos foram extraídos de um sistema de gerenciamento das atividades dos estudantes inscritos num certo curso. O objetivo era prever um possível resultado sobre o desempenho daqueles estudantes ao final do semestre.

De acordo com a Conferência Internacional sobre Mineração de Dados Educacionais (*International Conference on Educational Data Mining*), as aplicações primárias da MDE são [35]: predição

de desempenho estudantil, modelagem do estudante, detecção de comportamentos indesejáveis dos alunos, análise e visualização de dados, fornecimento de feedback para apoiar os instrutores, construção de material didático, planejamento e programação, recomendações para estudantes, categorização de alunos, análise de redes sociais, desenvolvimento de mapas conceituais.

2.1 Redes Sociais

A Ciência das Redes fornece ferramentas de medida e representação dos relacionamentos de um sistema, onde cada elemento da relação pode ser representado por um vértice e a relação estabelecida entre dois vértices pode ser representada por uma aresta [29]. Desta forma, tem provido uma estrutura unificadora capaz de colocar diferentes sistemas sob o olhar de uma mesma lente conceitual [6].

Dentre os benefícios obtidos com a integração de tecnologias de redes sociais à educação para alcançar objetivos de aprendizagem, estão incluídos: conhecimento e desenvolvimento de habilidades no manuseio de tecnologias, motivação, incentivo à conectividade entre alunos e professores facilitando a passagem de conhecimento, familiaridade com as ferramentas, conveniência para que o processo de aprendizagem continue após as aulas, incentivo para que o aluno expresse seus pensamentos e opiniões e fomento ao trabalho em equipe e à cooperação [31].

No trabalho de Gomes [18] ocorreu uma convergência entre duas tendências da área de Informática na Educação: a Mineração de Dados Educacionais e Análise de Redes Sociais Educacionais (RSE). A compreensão das estruturas sociais em uma RSE pôde auxiliar na navegação da rede, visualização e análise, tornando possível um aprendizado direcionado aos grupos, assim como a realização de repasse de conteúdos adaptáveis.

Um trabalho posterior, descrito Gomes e Prudêncio [19], também utilizou diferentes técnicas de mineração de dados aplicadas sobre os dados gerados através das interações numa rede social educacional. O objetivo do trabalho era extrair atributos de cada usuário individualmente e verificar se os membros dos grupos realmente possuíam interesses e/ou características em comum que os diferenciavam do restante da rede. Com este resultado, foi possível delinear os perfis respectivos a cada grupo identificado e assim entender o processo de formação da comunidade em plataformas educacionais online.

O trabalho de De Brito *et. al.* [15] traz uma aplicação de algoritmos de aprendizagem de máquina na identificação de estudantes com risco de evasão, com base em suas notas nas matérias do vestibular e disciplinas do primeiro semestre do curso. Para estudar a viabilidade do modelo na predição da situação final dos estudantes, utilizaram-se dados do histórico de alunos do curso de Ciência da Computação, da Universidade Federal da Paraíba (UFPB). Os resultados obtidos, com taxas de acerto de até 86,9%, indicaram ser uma abordagem viável para a detecção de grupos de risco.

Podemos verificar no trabalho de Wegner *et. al.* [34] os resultados de pesquisa envolvendo a técnica de agrupamento na identificação e classificação do Estilo de Cognitivo de Aprendizagem de estudantes que interagem em um Ambiente Virtual de Aprendizagem.

Delgado e Castro [16], com o objetivo de tirar proveito do fato de que os atuais estudantes do ensino superior participam ativamente

de comunidades virtuais, onde compartilham conteúdo e atuam colaborativamente, discutiram sobre o uso de ferramentas de suporte as redes sociais para *e-learning* e apresentou uma experiência de uso do Facebook para o suporte ao ensino universitário. Tal experiência melhorou o entendimento a respeito de possíveis estratégias de gerência do conhecimento dos alunos para fins didáticos, e também reforçou que quanto maior a compreensão de como acontecem as dinâmicas sociais, mais proveito pode se ter com o uso da ferramenta em estudo. Acrescentam que a captura do interesse do aluno propicia as interações entre eles.

No estudo de Maia *et. al.* [23], alunos de um curso de graduação e as disciplinas nas quais estavam inscritos foram modelados como nós e arestas de redes e foram feitas simulações como redes complexas. Os resultados obtidos trouxeram uma boa precisão nas notas simuladas e possibilitaram discussões sobre características do curso e seus alunos.

Os trabalhos de Oliveira *et. al.* e Brito *et. al.* [10, 30] exploraram a utilização de redes sociais como o Twitter e o Google Groups a fim de definir o grau de afinidade entre alunos de uma mesma classe, seus conhecimentos individuais relacionados a determinadas matérias e o grau de prestígio de um aluno com relação aos demais utilizando métricas de Análise de Redes Sociais. O primeiro define e analisa o grau de afinidade entre os alunos utilizando uma fórmula composta pelo número de *Twitters* e *retwitters* ocorridos entre pares de alunos. Analisa também o grau de intermediação dos alunos (vértices da rede) e relata a existência de falhas na disseminação da informação e interação entre eles, concluindo sobre a necessidade de intervenção do professor no suprimento de tais falhas. O segundo disponibiliza uma ferramenta própria para as trocas de mensagens e analisa o grau de impacto da colaboração de cada aluno utilizando tal ferramenta. O cálculo do grau de impacto considera a relevância da mensagem, o reenvio da mesma e quantas vezes esta mensagem foi repassada por outros colegas.

Dentro do contexto de combinação de técnicas, o trabalho de Miyata *et. al.* [26] utiliza a Mineração de Texto com Análise de Redes Sociais para identificar de forma automática a área de atuação de um pesquisador ou o assunto principal de um artigo. Estes podem servir, por exemplo, de base para a análise de redes acadêmicas visando avaliar grupos interdisciplinares ou identificar tendências. A característica baseada em mineração de textos utilizou a frequência relativa das palavras dos títulos dos artigos publicados em periódicos. As duas características extraídas utilizando-se métricas da Análise de Redes Sociais foram a porcentagem dos vizinhos pertencentes a cada grande área (ou área ou subárea) utilizando-se o primeiro nível de vizinhos (apenas vizinhos diretos, no caso, coautores) e utilizando-se a vizinhança nível dois (vizinhos e vizinhos dos vizinhos). Verificou-se que a combinação simples da técnica de mineração de textos utilizada com a análise da vizinhança de nível dois obteve melhores resultados do que aqueles produzidos pelas técnicas usadas individualmente.

De modo semelhante, este trabalho utiliza métricas de Análise de Redes Sociais para estudar a dinâmica de uma turma, no entanto, além do grau de intermediação, analisa o coeficiente de agrupamento, alunos que promovem novas conexões entre aqueles que já são seus amigos e os graus de entrada e saída dos alunos, quantificando suas conexões e o tipo de participação na aula.

O presente trabalho encontra-se neste contexto proporcionando o incentivo a realização de trocas de conteúdos acadêmicos através de uma rede social que acontece enquanto os alunos estão participando de discussões em sala de aula presencialmente. Medidas feitas sobre a rede social gerada, combinadas às características da rede textual pertinente a cada aluno participante, são a fonte de dados submetidos a técnicas de mineração de dados, com enfoque de agrupamento e classificação. Quanto ao objetivo, diferencia-se dos demais trabalhos encontrados na literatura, na tentativa de identificar grupos e mapear papéis desempenhados pelos alunos durante as dinâmicas presenciais em aula.

2.2 Redes Textuais

Na última década, tem crescido o número de pesquisas onde a representação das interações entre as palavras utilizam as ferramentas da Teoria dos Grafos, gerando uma rede textual. Desta forma, possibilitando o emprego de propriedades matemáticas e a formulação de leis gerais que emergem das relações pertinentes a estas redes e revelando padrões semelhantes aos de outros sistemas complexos [11, 24, 28 e 32].

Dependendo do contexto e objetivos de uma aplicação, a rede textual pode ser construída conectando palavras de acordo com relações semânticas entre elas. Quando o estudo, baseado numa rede textual, visa buscar o estilo ou estrutura topológica da escrita, a construção desta rede deve levar em consideração a relação sintática, numa análise a respeito da ordem e frequência na qual as palavras aparecem [2]. A similaridade entre a análise sintática de um texto e o estudo a respeito da frequência e posição das palavras acontece pelo fato de que, geralmente conexões sintáticas acontecem entre palavras vizinhas [12].

Existem, portanto, diferentes maneiras de se modelar uma rede textual, e dentre estes estão os que avaliam os textos como uma rede de adjacência entre as palavras. Neste tipo de representação, palavras relevantes contendo significado são modeladas como nós e os relacionamentos de adjacência são usados para estabelecer as conexões. Com o uso deste modelo, diferentes características de textos e linguagens tem sido deduzidas através de uma análise estatística realizadas sobre a estrutura das redes [2].

Amancio [1], estuda a influência do tamanho da amostra de texto para fins de reconhecimento de autoria e conclui que uma análise topológica local, relativa a grandes textos, pode revelar características importantes para a classificação do texto como um todo. Seus resultados provam o princípio de que pequenos textos podem ser analisados à luz das redes complexas e servem como embasamento teórico para esta pesquisa, corroborando com a hipótese de que é válida a análise topológica dos textos gerados via *chat*, durante uma dinâmica de sala de aula, direcionada ao ensino-aprendizado de uma disciplina específica.

No modelo de redes textuais, baseado na adjacência das palavras, utilizando os elementos da Teoria dos Grafos, as palavras são representadas por vértices, e arestas são estabelecidas entre palavras adjacentes, considerando a sequência na qual são escritas [2]. Cada par de palavras subsequentes define uma aresta direcionada e, associada a esta, haverá um valor para o peso desta aresta, que significa a frequência de associações entre tais palavras [4].

Antes de mapear uma rede textual baseada na adjacência das palavras, se faz necessária uma fase de pré-processamento do texto submetido a esta análise. Para alguns tipos de aplicações, palavras

com pouco significado, tais como preposições, conjunções, pronomes e artigos, devem ser retiradas do texto. Tais palavras são conhecidas como *stopwords*. Este procedimento tem sido adotado em diferentes estudos [1, 2, 24]. Numa segunda etapa, o texto resultante deve passar por um processo de *lemmatization*, ou seja, cada palavra deve ser convertida para sua forma canônica, nomes e verbos são convertidos para sua forma infinitiva e singular [2,3]. Podendo assim uma mesma palavra, em contexto e significados diferentes, ser representada pelo mesmo vértice [4].

Até este momento, no estudo em tela, os textos extraídos dos *chats* foram convertidos em uma rede de palavras considerando a adjacência entre elas, no entanto, os textos não foram submetidos aos tipos de pré-processamentos mencionados acima.

2.3 A Matemática das Redes

A Tabela 1 reúne as definições das medidas e termos que serão utilizados nas análises sobre as redes consideradas neste trabalho.

Tabela 1 - Definições de Conceitos Importantes para a Análise de Redes.

Número total de Vértices	Quantidade de diferentes indivíduos (rede social)/palavras(rede textual)
Número total de Arestas	Quantidade total de conexões estabelecidas entre os vértices da rede.
Grau de centralidade um vértice	Número de arestas conectadas diretamente ao vértice
Grau de Entrada de um vértice	Numa rede direcionada, é o número de arestas que incidem no vértice.
Grau de Saída de um vértice	Numa rede direcionada, é o número de arestas que partem do vértice.
Distância	Quantidade de arestas que existem entre dois vértices.
Caminho Mínimo	Caminho que contém o menor número de arestas entre dois vértices.
Caminho Mínimo Médio da Rede	Média dos caminhos mínimos relativos a cada vértice = CM_{rede}
Centralidade	Indica a importância do vértice
Coefficiente de Agrupamento	A probabilidade de que dois vizinhos de um vértice serão vizinhos entre si.
Coefficiente de Agrupamento Médio da Rede	Média dos coeficientes de agrupamentos dos vértices da rede. A probabilidade que uma rede tem na formação de grupos (CA_{rede}).
Densidade da Rede	Relação entre o número de conexões existentes e o número Máximo total de conexões que poderiam ser constituídas.
Diâmetro da Rede	Maior caminho mínimo entre dois vértices
Grau de Intermediação	Quanto um vértice está no caminho entre os outros vértices numa rede. Quanto maior for o valor deste grau, significa que este vértice é uma passagem obrigatória entre as menores

	distâncias entre os vértices de um grafo.
Hubs	Vértices com maiores graus de centralidade

3. METODOLOGIA

Seis etapas configuram as atividades metodológicas descritas neste artigo: seleção da amostra, coleta, armazenamento e tratamento dos dados, construção das redes social e textual, construção do vetor de características dos indivíduos, agrupamento e análises dos dados.

Seleção da amostra: direcionada a um ambiente educacional presencial, uma parceria com a disciplina de Física foi feita, abordando os conteúdos de Mecânica, considerado de difícil entendimento pela maioria dos alunos, resultando no baixo desempenho das turmas. Idealizou-se um experimento que acontecesse durante uma das aulas tradicionais da disciplina. O professor explanaria sobre o assunto daquele encontro e, através de uma ferramenta para bate-papo, alunos e professor poderiam debater as questões relativas ao conteúdo em pauta tendo como orientação a realização de alguns exercícios propostos pelo professor.

Coleta, armazenamento e tratamento dos dados: a coleta dos dados foi feita com o uso da ferramenta de bate-papo. O aplicativo consistiu de um conjunto de scripts compondo uma interface simples, utilizando as linguagens HTML e PHP para a conexão com o banco de dados do servidor MySQL. Ao receber os textos digitados pelos participantes da discussão, a ferramenta registrou, além dos textos propriamente ditos, de onde e para quem partiram as mensagens. A Figura 1 traz um esquema representativo desta metodologia.

Nesta etapa, os arquivos dos textos individuais passaram por um pré-processamento antes de serem transformados em redes de conexões de palavras. Até o momento da aplicação deste primeiro experimento, a ferramenta não disponibilizava uma formatação especial para a digitação de fórmulas matemáticas. Além disso, as palavras do tipo *stopwords* foram mantidas.

Considera-se, no entanto, para a continuação desta pesquisa, que esta fase de pré-processamento das palavras dos textos das mensagens seja modelada e executada de acordo com uma avaliação sobre os requisitos necessários para o emprego de alguma técnica de mineração de dados a ser testada e selecionada.

Construção das redes textuais: o texto respectivo a cada indivíduo foi transformado numa rede textual, onde as palavras foram representadas por vértices e uma aresta direcionada conectou as palavras de acordo com a sequência na qual apareceram no texto. Para esta conversão de texto em rede foi utilizada a ferramenta *Speech Graph* [27]. O aplicativo recebe um arquivo do tipo “txt” como entrada retornando a imagem de uma rede das palavras contidas no arquivo e, num outro arquivo também do tipo “txt”, o conjunto dos resultados de medidas matemáticas aplicadas à rede.

Construção da rede social: Ainda nesta etapa, acontece o uso de outra ferramenta de visualização e medição de grafos denominada Gephi [20]. A ferramenta recebe como arquivo de entrada uma planilha contendo as colunas “de” e “para”, relacionando os nomes dos indivíduos que trocaram mensagens ao longo da discussão. Com base neste dado, a ferramenta constrói uma rede, transformando os nomes dos indivíduos em vértices e as relações

(as linhas da planilha) em arestas. Feito isso, gera uma visualização da rede como um todo e fornece os resultados das medidas necessárias para a análise do comportamento e busca de padrões nesta rede.

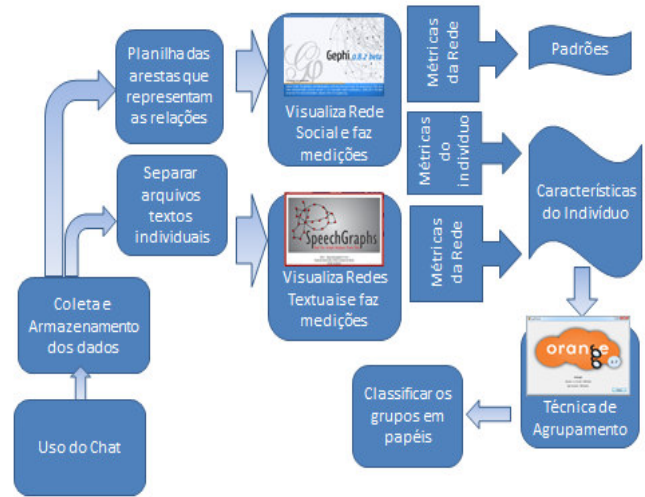


Figura 1 - Esquema representativo da metodologia adotada

Construção do vetor de características: combinar as características que cada indivíduo carrega numa rede e na outra simultaneamente. A composição dos vetores de características acontece pela junção das medidas calculadas sobre o a rede textual do indivíduo mais as medidas calculadas sobre a rede social que construída durante a atividade, como nos esquemas da Figura 2, Figura 3 e Figura 4.

Agrupamento: agrupar os participantes da atividade através do emprego de alguma técnica de Mineração de Dados. De posse das características dos indivíduos, um terceiro aplicativo para mineração de dados: Orange Data Mining [17] é utilizado para agrupamento.

Análise dos dados: Uma análise a respeito da centralidade dos vértices mostra se há indivíduos que se destacaram durante a discussão (*hubs*), seja pela quantidade de conexões que fizeram (o grau de centralidade de cada vértice), seja pelas intermediações que proporcionaram (grau de intermediação) ou se são indivíduos com a tendência de formar triângulos de relacionamentos (coeficiente de agrupamento).

De posse dos grupos indicados pela ferramenta de mineração de dados, segue-se uma avaliação do tipo de participação que mais caracteriza cada grupo. Nesta avaliação deve-se levar em consideração a quantidade de indivíduos que se destacaram numa mesma métrica somada à percepção do professor sobre o comportamento dos alunos em sala de aula.

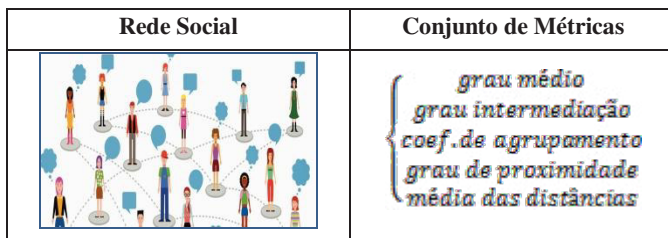


Figura 2 - Métricas extraídas da Rede Social considerada.

de Informática do colégio. A maioria dos alunos utilizou desktops para o bate-papo, mas alguns puderam fazer uso de seus *smartphones e tablets*. A intensão, num próximo experimento, é que alunos e professor possam utilizar apenas dispositivos móveis e não precisem se deslocar da sala de aula convencional.

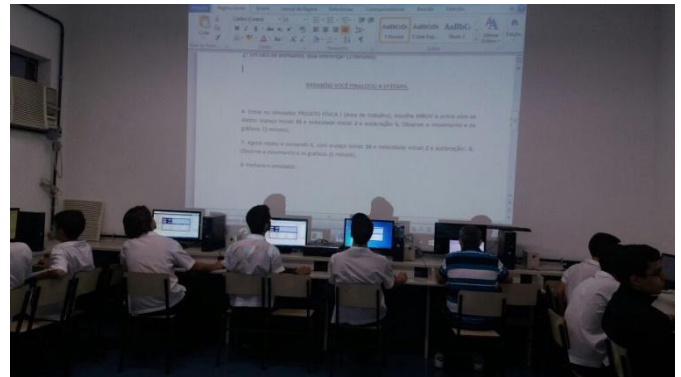


Figura 5 - Alunos e Professor utilizando uma sala de bate-papo durante a aula.

A rede social construída durante as trocas de mensagens ficou composta por 23 vértices, dos quais um foi chamado de “Todos” representando todas as vezes que alguém dirigia uma mensagem para os demais, sem um destinatário específico.

A Tabela 2 traz os resultados numéricos calculados. Mostra que 155 mensagens foram trocadas, que é o número de arestas. Que os participantes estabeleceram conexão com uma média de 6,7 pessoas diferentes e o tamanho do caminho mínimo médio é 1,7. Estes números mostram a proximidade da turma como um todo para a resolução dos exercícios propostos e confirmam a hipótese de que mais alunos se sentiriam a vontade para emitirem suas opiniões durante a discussão. O coeficiente de agrupamento no valor de 0,4 indica que a rede tem espaço para a formação de novos triângulos de amizade, que atingiria o máximo de conexões deste tipo se esse valor fosse 1.

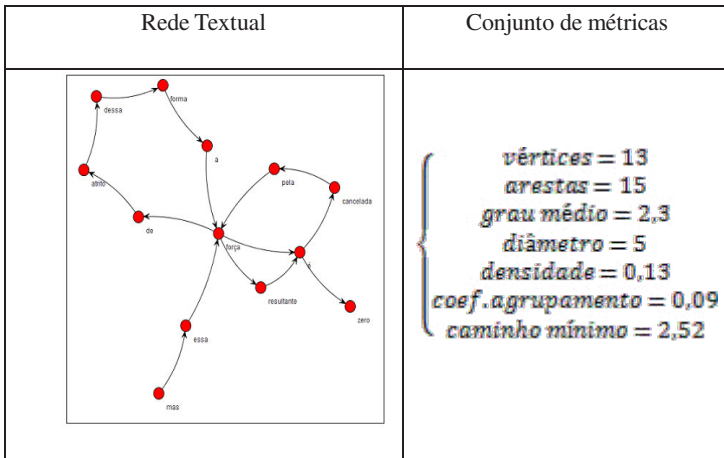


Figura 3 - Métricas extraídas das Redes Textuais consideradas. Exemplo de Rede textual construída para o texto: “*mas essa força é cancelada pela força de atrito. Dessa forma, a força resultante é zero*”

Tabela 2 - Rede social das trocas de mensagens entre alunos e professor

Número Vértices	Número Arestas	Grau Médio	Caminho Mínimo Rede	Coefficiente Agrupamento Rede
23	155	6,7	1,7	0,4

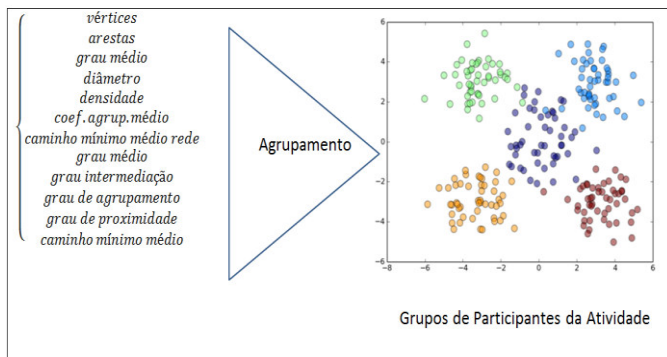


Figura 4 - Proposta para a construção do vetor de características de um indivíduo.

4. RESULTADOS E ANÁLISE DO EXPERIMENTO

Para verificação da proposta apresentada neste trabalho, realizou-se um experimento no dia 19 de abril de 2016, com uma das turmas que cursa o 2º ano do Ensino Médio, numa escola pública na cidade do Rio de Janeiro. Neste dia houve a participação de 21 alunos (4 meninas e 17 meninos) e o professor regente da disciplina de Física. A Figura 5 traz o momento onde alunos e professor trocam mensagens diante da tarefa de resolver os exercícios propostos para a dinâmica daquele dia. Como pode ser visualizada na Figura 5, a dinâmica ocorreu num dos laboratórios

A Figura 6 mostra as conexões da rede social destacando aqueles indivíduos que estabeleceram maior número de conexões, medido pelo seu grau (*hubs*). Destacam-se os vértices: 3, 6, 9, 10, 11, 12, 14, 18, 19, 21. A saber, o vértice de número 18 é o professor.

A Figura 7 traz os vértices que se destacaram com maiores graus de intermediação. Os indivíduos que mais atuaram como pontes entre outros grupos de pessoas foram: 1, 6, 7, 9, 10, 11, 12, 18, 19, 21.

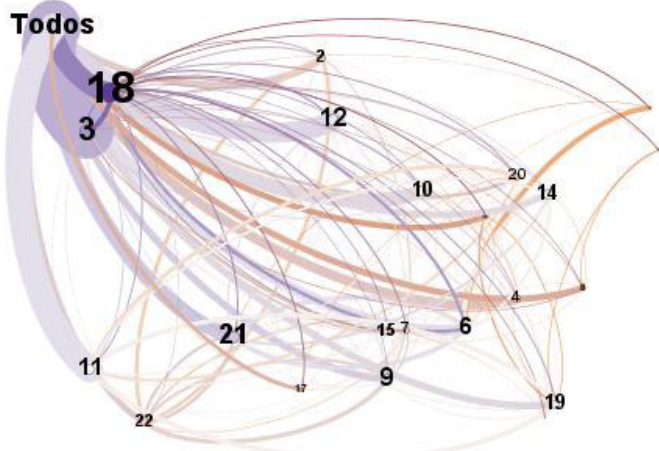


Figura 6 - Rede social resultante da dinâmica de troca de mensagens. Destaque para os vértices com maiores graus.

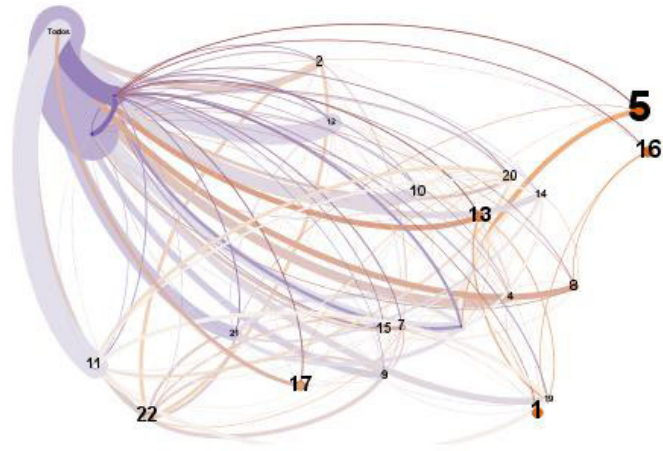


Figura 8 - Rede social resultante da dinâmica de troca de mensagens. Destaque para os vértices com maiores coeficientes de agrupamento.

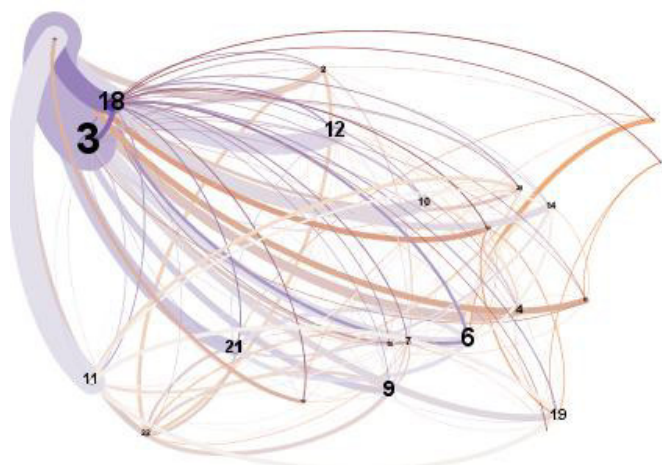


Figura 7 - Rede social resultante da dinâmica de troca de mensagens. Destaque para os vértices com maiores graus de intermediação.

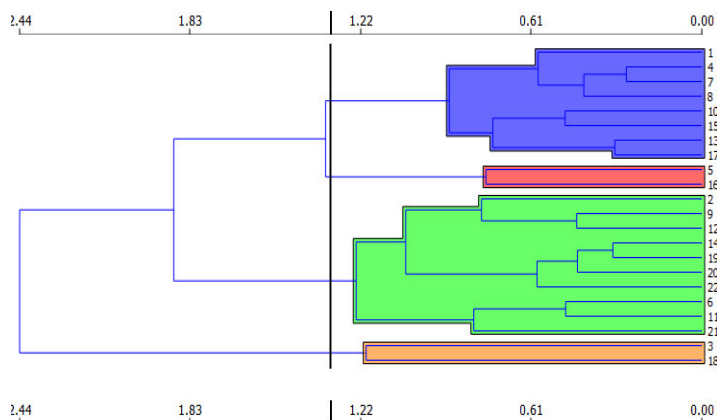


Figura 9 - Distribuição em Grupos: Algoritmo de agrupamento hierárquico com base na distância Euclidiana entre os vértices. (Fonte: Orange 2.7.8.dev-7887.ad2)

A Figura 8 mostra os vértices classificados de acordo com o coeficiente de agrupamento que atingiram. Aqueles que atingiram maiores coeficientes foram: 2, 4, 5, 7, 8, 10, 13, 16, 17, 22.

A Figura 9 mostra como ficou a distribuição em grupos após a aplicação do algoritmo de agrupamento hierárquico, baseado na distância euclidiana entre os vértices. Como já dito anteriormente, cada vértice foi representado por um vetor de características, contendo os resultados numéricos extraídos da rede social e da rede de palavras relativas a cada indivíduo.

Foram registradas as impressões dos professores regentes de diferentes disciplinas durante um Conselho de Classe, ocorrido em junho deste mesmo ano. Faz parte da dinâmica deste conselho a indicação de alunos que se destacam positiva ou negativamente. De um modo geral, considerando, dentre outros quesitos, se o aluno é participativo nas aulas, se é responsável com suas tarefas acadêmicas, se é assíduo e se tem bom rendimento em suas notas.

Em paralelo, solicitou-se que o professor de Física, o mesmo que participou do experimento, fizesse também uma avaliação sobre o tipo de participação percebida por ele em suas aulas tradicionais. A análise foi direcionada para que considerasse algumas características específicas em sua observação, segundo a Tabela 3.

Tabela 3 - Critérios para classificação sobre a participação dos alunos na atividade proposta.

Auxiliadores (A)	Oferecem ajuda aos demais espontaneamente.
Interessados (I)	Faz perguntas, dá retorno sobre o que entendeu, participa ativamente dos debates.
Distratores (D)	Fazem brincadeiras e piadas ao longo das discussões.
Apático (Ap)	Não se envolve nas discussões, nem atrapalha a dinâmica.

As colunas da Tabela 4 trazem a classificação feita pelo professor de Física, os alunos que receberam indicação de destaque pelo Conselho de Classe e o grupo ao qual pertence, de acordo com o algoritmo de agrupamento executado.

Tabela 4 - Classificação dos alunos quanto à participação em aula e o grupo ao qual pertencem após o algoritmo de agrupamento.

IdAluno	Professor Física	Conselho de Classe	Agrupamento Hierárquico
A1	I	destaque positivo	Azul
A4	I		Azul
A7	AP		Azul
A8	AP		Azul
A10	AP		Azul
A13	AP	destaque positivo	Azul
A15	D	destaque negativo	Azul
A17	D		Azul
A5	AP	destaque positivo	Rosa
A16	D	destaque negativo	Rosa
A2	I		Verde
A6	A		Verde
A9	AP		Verde
A11	I	destaque positivo	Verde
A12	I		Verde
A14	I		Verde
A19	D		Verde
A20	AP	destaque positivo	Verde
A21	A	destaque positivo	Verde
A22	I		Verde
A3	A	destaque negativo	Laranja
A18		Professor	Laranja

É possível ver na Figura 9, representada também na Tabela 4, através de diferentes cores, a separação da turma em quatro grupos distintos. Numa primeira análise, os dois grupos menores, aquele que abrange os vértices 5 e 16 e o outro que contém os vértices 3 e 18 chamam a atenção, pelo fato de reunirem pessoas que

receberam algum tipo de destaque pela percepção dos professores do Conselho de Classe.

No primeiro caso, um destaque positivo e um destaque negativo. No entanto, durante a dinâmica revelou que participaram de modo semelhante. Estes alunos são aqueles que atingiram maiores valores para o coeficiente de agrupamento na rede social. De fato, analisando trechos das mensagens trocadas por estes alunos verificam-se termos como¹:

Aluno A5: “>>**concordo A6 pra sair do repouso tem aceleração translação** >> o A3 explicou que é rotação n. >> é A2, pq a Terra n faz um círculo perfeito >> **sim A6, já a translação e a terra em seu próprio eixo >> sim boa A6**”

Aluno A16: “>>**vamos trocar** >> depende da situação, o ponteiro tem velocidade constante >> **boa A3; vdd A20**”

Mostrando que fizeram comentários pertinentes ao conteúdo discutido, envolveram outros alunos nas discussões usando palavras de incentivo.

Já o grupo formado pelo aluno A3 e o professor (A18) também evidência a participação deste aluno numa postura semelhante a do professor. Ambos aparecem na rede social com altos valores para o número de conexões (grau) e para as intermediações que fizeram (grau de intermediação). A análise dos textos corrobora este resultado. O Aluno A3, incentiva os outros colegas a participarem da discussão, faz comentários a respeito das mensagens que postam, e registra o que, pessoalmente, está compreendendo sobre a matéria discutida. Além disso, faz comentários descontraídos ao longo da conversa,

Aluno A3: “>>”Pronto, começou e parei de zoar. >> ‘tava’ falando do espaço inicial. >> **‘Bora’ Galera, já passou pro 5.** >> A velocidade muda em uma função linear, de forma constante. Sempre em $2m/s$ ($2m/s^2$) >> **Boa A20, 2 é repouso também, A11.** >> E a aceleração é 0, então... velocidade aumenta de forma constante. >> Interessante variação. >> **Palmas** >> Esse comando 3 é o mais difícil até agora hein?>> Desculpa. >> 1) Um é uma diagonal para cima, a outra, para baixo.>>; **ah, boa, A4** “

No grupo azul, dos oito alunos, seis alcançaram os maiores valores para o coeficiente de agrupamento. Esta métrica indica a tendência a formação de triângulos de relacionamento.

No grupo verde, dos dez alunos, oito se destacaram acumulando altos índices para o grau de intermediação e para o grau de centralidade. De acordo com a classificação do professor, em maioria, alunos interessados e que auxiliam outros.

5. CONCLUSÃO E TRABALHOS FUTUROS

Este trabalho apresentou a proposta de utilização de uma rede social e redes textuais individuais, construídas com o uso de um bate papo durante uma aula presencial, como ferramentas para registro e análise da dinâmica. A proposta teve como objetivo identificar o tipo de papel que alunos assumem quando participam de discussões provocadas pelo professor.

¹ Os nomes originais foram substituídos pelos respectivos identificadores utilizados no texto.

Um experimento foi realizado com uma turma do Ensino Médio. As expectativas do professor e alunos foram alcançadas tendo em vista que puderam experimentar uma dinâmica diferente e durante o tempo de aula. Utilizou-se *smartphones*, *tablets* e *desktop* enquanto alunos puderam ensinar uns aos outros.

Os registros das conexões realizadas e representadas em termos de uma rede social e de redes textuais mostraram, além do conteúdo da aula naquele dia, a participação individual dos alunos. Os relacionamentos estabelecidos revelaram características particulares sobre a colaboração de alguns alunos naquela aula. Através do emprego de um algoritmo de agrupamento hierárquico verificou-se a existência de quatro grupos na turma.

Também foi possível verificar que os grupos identificados apontam diferentes tendências de colaboração e consequentemente papéis.

De posse destas observações foi possível desenvolver um olhar individualizado para o aluno, evidenciando alguns que naturalmente não se destacavam.

Como trabalhos futuros pretende-se expandir a investigação para outras disciplinas. Também pretende-se analisar como seria a evolução no tempo após alguns outros experimentos e verificar a existência ou não de padrões clássicos encontrados em redes complexas.

6. REFERÊNCIAS

- [1] Amancio, D. R. Probing the Topological Properties of Complex Networks Modeling Short Written Texts. *Plos One*, N. 10(2), 2015A.
- [2] Amancio, D. R. A Complex Network Approach to Stylometry, 2015B.
- [3] Amancio, D. R.; Jr, O. N. O.; Costa, L. F. Identification of Literary Movements Using Complex Networks to Represent Texts. *New Journal of Physics*, N. 14, 2012.
- [4] Antikeira, L. *et.al.* Strong Correlations Between Text Quality and Complex Networks Features. *Research Gate*, 2005.
- [5] Baker, R.; Isotani, S.; Carvalho, A. Mineração de Dados Educacionais: Oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, V. 19, N. 2, 2011.
- [6] Barabasi, A. The Network Takeover. *Nat. Phys*, V. 8(1), P. 14, 2011.
- [7] Baradwaj, B. K. . & P. S.. Mining Educational Data to Analyze Students' Performance. *Arxiv Preprint*, P. 1201.3417, 2012.
- [8] Baronchelli, A. *et.al.* Networks In Cognitive Science. *Trends In Cognitive Sciences*, V. 17, P. 348-357, 2013.
- [9] Brito, A. V. Simulação Baseada em Atores como uma Ferramenta Educacional. *XX Simpósio Brasileiro de Informática na Educação (SBIE)*, 2009.
- [10] Brito, A. V. *et.al.* Estudo Da Utilização de Redes Sociais Como Ferramenta de Avaliação de Participação. *XXI Simpósio Brasileiro de Informática na Educação, SBIE*. João Pessoa, 2010.
- [11] Cancho, R. F.; Solé, V. R. The Small World of Human Language. *Proceedings of The Royal Society of London. Series B: Biological Sciences*, 2001. 2261-2265.
- [12] Cancho, R. F. I.; Solé, R. V.; Köhler, R. Patterns in Syntactic Dependency Networks. *Physical Review*, N. 69, 2004.
- [13] Costa, L. D. F. *et.al* . Correlations Between Structure and Dynamics in Complex Networks. *Arxiv Preprint Physics/0611247*, 2006.
- [14] Costa, L. D. F. *et.al.* Analyzing and Modeling Real-World Phenomena with Complex Networks: A Survey of Applications. *Advances in Physics*, P. 60(3), 2011.
- [15] De Brito, D. M., Lemos, M. O., Pascoal, T. A., Do Rêgo, T. G., & Araújo, J. G. G. D. O. (2015). Identificação de Estudantes do Primeiro Semestre com Risco de Evasão Através de Técnicas de Data Mining.
- [16] Delgado, C.; Castro, D. G. D. O Uso De Uma Plataforma De Redes Sociais Para Elearning. *Anais Do Brazilian Workshop On Social Network Analysis And Mining (Brasnam)*, 2012.
- [17] Demsar J, Curk T, Erjavec A, Gorup C, Hocevar T, Milutinovic M, Mozina M, Polajnar M, Toplak M, Staric A, Stajdohar M, Umek L, Zagar L, Zbontar J, Zitnik M, Zupan B (2013) Orange: Data Mining Toolbox In Python. *Journal Of Machine Learning Research* 14(Aug):2349–2353.
- [18] Gomes, J. E. A. Caracterização de Grupos para Entendimento das Redes Sociais Educacionais. *Dissertação Mestrado*, Universidade Federal de Pernambuco, 2013.
- [19] Gomes, J. E. A.; Prudêncio, R. B. C. Educational Social Network Group Profiling: An Analysis Of Differentiation-Based Methods. *Iii Brazilian Workshop On Social Network Analysis And Mining (Brasnam)*, 2014.
- [20] Graph Visualization And Manipulation Software Gnu General Public License Version 0.8.2. Gephi Graph Visualization and Manipulation Software GNU General Public License Version 0.8.2., 2015. Disponível Em: <<https://gephi.org/>>. Acesso Em: 05 Agosto 2015.
- [21] Jiawei, H.; Kamber, M.; Pei., J. *Data Mining: Concepts and Techniques*. [S.L.]: Elsevier, 2011.
- [22] Lemos, E. C.; Oliveira, W. Utilizando o Ambiente Virtual de Aprendizagem Moodle como Apoio às Disciplinas Presenciais. *Anais do MoodleMoot 2010*. Buenos Aires, 2010.
- [23] Maia, R.; Spina, E.; Shimizu, S. (2010), Sistema de Previsão de Desempenho de Alunos para Auxílio a Aprendizagem e Avaliação de Disciplinas. *Anais do XXI Simpósio Brasileiro de Informática na Educação – UFPB- Paraíba*, 2010.
- [24] Masucci, A. P.; Rodgers., G. J. Network Properties of Written Human Language. *Physical Review E* 74.2, 2006.
- [25] Ministério da Educação. Brasil. *Caderno: O uso de tecnologias na Educação*. Brasília: Junho, 2008.
- [26] Miyata, B. K. O.; Kano, V. Y.; Digiampietri, L. A. Combinando Mineração de Textos e Análise de Redes Sociais para a Identificação das áreas de atuação de pesquisadores. In *Second Brazilian Workshop On Social Network Analysis and Mining (Brasnam)*, 2013.

- [27] Mota, N. B. *et al.* Graph Analysis of Dream Reports is Especially Informative About Psychosis. *Science Reports*, V. 4, P. 3691, 2014.
- [28] Motter, A. E. *et al.* Topology of the Conceptual Network Of Language. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. *Physical Review E* 65.6 065102, 2002. 65(6).
- [29] Newman, M. *Networks, An Introduction*. New York, Eua: Oxford University Press Inc, 2010.
- [30] Oliveira, R. F. *et. al.* Estratégia para Avaliação do Grau de Participação e Afinidade de Alunos Através de Microblogging. *Anais Do XXII Sbie - XVII Wie*, 2011.
- [31] Purcell, M. A. *The Networked Library: A Guide for the Educational Use of Social Networking Sites*. New York, Usa. 2010.
- [32] Solé, R. V. *et.al* Language Networks: Their Structure, Function, And Evolution. *Complexity*, V. 15(6), P. 20-26, 2010.
- [33] Unesco. Tics na Educação do Brasil. 2015. Disponível Em: [Http://Www.Unesco.Org/New/Pt/Brasilia/Communication-Andinformation/Access-To-Knowledge/Ict-In-Education/](http://www.unesco.org/new/pt/brasilia/communication-andinformation/access-to-knowledge/ict-in-education/). Acesso Em: 24/08/2016
- [34] Wegner, W., Frozza, R., Bagatini, D. D., & Da Cruz, M. E. J. Técnica De Clusterização Para Determinar O Estilo Cognitivo De Usuários Em Um Ambiente Virtual De Aprendizagem. *Conferencia Internacional de Informática na Educação (TISE)*, V.11, P. 444-449, 2015
- [35] Yacef, K. *et.al.* Proceedings of the 5TH International Conference on Educational Data Mining. 2012.