

Mineração de Dados em publicações de Fóruns de Discussões do Moodle como geração de Indicadores para aprimoramento da Gestão Educacional

Leandro A. Silva, Douglas Trindade, Cristiane de Paula, Simone Nobrega Pinto

Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie
São Paulo - Brasil

prof.leandro.augusto@mackenzie.br, douglas_san-trindade@hotmail.com,
crisdepaulasp@hotmail.com, si.timackenzie@gmail.com

***Abstract.** Social networks have acquired great relevance in recent years and for this reason have been studied from many different areas like business, justice, education, politics and etc. Data mining can help data-driven processes of use of the data from the social network once it permits the discovery of information to those areas promoting new knowledge. It will be showed in this work the use of data mining applied in the discussing forum of the tool of teaching called as moodle, which is considered here as a pseudo social network, using the data clustering technique known as K-means. Results provided by this work may be used as monitoring indexes to improve educational management.*

***Resumo.** As redes sociais têm adquirido grande relevância nos últimos anos e por esta razão tem sido objeto de estudo das mais diversas áreas, como negócios, justiça, educação, política e etc. A mineração de dados auxilia no aproveitamento dos dados das redes sociais, pois proporciona a descoberta de informações, provendo as áreas de novos conhecimentos. Neste trabalho apresenta-se o uso de mineração de dados em fóruns de discussões da ferramenta de apoio de ensino Moodle, o qual aqui será considerado como uma pseudo rede social, através da técnica de agrupamento de dados conhecida por K-médias. Por fim, os resultados podem ser utilizados como indicadores para o monitoramento da qualidade da gestão educacional.*

1. Introdução

A Gestão de Instituições de Ensino Superior (IES) é uma difícil tarefa geralmente executada por um Diretor e um ou mais Coordenador de Curso. O grande desafio dos gestores em administrar uma Faculdade ou um Curso constitui-se de aspectos acadêmico, com atualização de projetos pedagógicos, gestão de professores e etc; gerencial, através da supervisão das salas de aula, instalações físicas e etc; e muitos outros desafios institucionais e até mesmo políticos [Tachizawa & Andrade, 1999; Tanaka & Pessoni, 2011]. A responsabilidade da administração pode interferir na qualidade do ensino como um todo e até mesmo se tornar motivo do aluno em se evadir de um curso [Baggi & Lopes, 2011].

Contudo, quando algum problema gerencial acontece, muitas vezes os gestores acabam sendo as últimas pessoas a ficarem sabendo, e isso eleva a dificuldade na administração de conflitos. O ideal nesse caso é o gestor saber do problema no momento que ele surge, ou próximo à ocorrência.

Com o advento das redes sociais e a facilidade em seu acesso, muitos problemas acabam se externando por esse tipo de mídia. Quando se faz uma publicação em uma rede social sobre uma específica ocorrência, o problema que poderia ser resolvido talvez de forma simples toma-se outras dimensões que o torna, por vezes, difícil de ser resolvido. Não se trata, entretanto, de uma publicação na rede social ter a exposição de um sentimento e, que as empresas estão coletando essas opiniões para avaliar produtos, monitorar marcas e estudar comportamentos de seus usuários [Agarwal et. al, 2011]. O fato é que por vezes, algum descontentamento individual e pessoal que pode não ser a opinião de todos acaba tomando proporções desnecessárias. E ainda, sem considerar as notícias falsas e irresponsáveis que são publicadas para denegrir a imagem de uma pessoa e até uma marca.

Pensando nisto, este projeto apresenta uma proposta que usa um recurso do Moodle (*Modular Object-Oriented Dynamic Learning Environment*) chamado Fórum, onde os alunos de uma IES devidamente cadastrados nesse ambiente tem a possibilidade de publicar as ocorrências vivenciadas em sua rotina escolar. Dessa maneira, neste trabalho será adotada a definição de Recuero et. al (2009) que define o Moodle como uma *pseudo* rede social, pois conecta um grupo de pessoas, nesse caso alunos, professores e gestores, para uma discussão coletiva. A escolha do Moodle se deu pelo fato dessa ser a ferramenta de apoio à aprendizagem adotada na IES onde esta pesquisa se insere.

Para analisar as publicações na pseudo rede social será feito o uso de técnicas de Mineração de Dados [Silva, 2015]. Ou seja, as publicações representadas em formato de texto, *corpus*, serão tratadas para a construção de uma coleção de palavras (do inglês, *bag-of-words*) e, conseqüentemente, a construção de um *corpus* representado numericamente pela importância de cada palavra. O *corpus* representado, conseqüentemente, será agrupado por similaridade de conteúdo, com o uso do algoritmo *k*-médias (Silva, 2015). Dessa maneira, e aqui está a contribuição mais importante do trabalho, os *corpus* agrupados serão usados para que se recupere a respectiva postagem, que manualmente será classificada em algum tipo de ocorrência como, por exemplo, burocracia, infraestrutura, avaliação, planejamento, comunicação e etc. O resultado classificado e a análise de ocorrência de cada grupo serão apresentados como dois importantes indicadores que poderão ser usados pela IES para definir prioridades de solução e monitoramento dos problemas descobertos.

O trabalho está organizado como segue. No Capítulo 2 discute-se, de forma introdutória, redes sociais, da qual, assume-se aqui, o Moodle sendo uma pseudo rede social. No Capítulo 3, apresenta-se uma introdução sobre Mineração de Dados e a descrição do algoritmo usado neste trabalho. A metodologia experimental e os resultados estão no Capítulo 4. Por fim, Capítulo 5, apresenta-se uma discussão e os resultados deste estudo.

3. Rede Social

Com a evolução tecnológica, as redes sociais físicas passaram para o âmbito eletrônico e hoje temos as mídias sociais também conhecidas como redes sociais, que é um

conjunto de pessoas ou entidades interligados uns aos outros através da internet, ou seja, quando uma rede de computadores conecta uma rede de pessoas e organizações, é, segundo Recuero et. al (2009) uma rede social.

Atualmente existem diversas redes sociais na internet, cada uma abrangendo um nível específico da vida social, tais como: redes de relacionamentos, redes profissionais, redes comunitárias, redes políticas, redes educacionais. Alguns exemplos de redes sociais são:

- **Facebook:** é uma rede social de relacionamento onde o usuário cria seu perfil, adiciona amigos, posta fotos, vídeos e links que são compartilhados com todos os amigos.

- **Twitter:** é uma rede social de relacionamento onde o usuário cria um perfil e a partir disso pode acompanhar outros perfis de acordo com seu interesse; e também poderá ser acompanhado pelos usuários interessados.

- **Youtube:** é uma rede social destinada a vídeos onde o usuário cria seu perfil e atrela a ele seus vídeos favoritos e publicados. Todos esses vídeos ficam disponíveis em uma página chamada de canal.

- **LinkedIn:** é uma rede social profissional onde o usuário cria seu perfil com todos os dados profissionais e acadêmicos com o objetivo de se relacionar com outros profissionais. O uso dessa rede ainda permite indicar perfis à vagas de empregos.

As redes sociais possuem uma grande capacidade de geração de dados, principalmente pelo fato da quantidade de usuário que as acessam. Entender as vantagens dessa nova forma de comunicação, que possui um alcance imenso, e utilizar todas estas informações geradas diariamente, a fim de gerar conhecimentos, poderá trazer grandes benefícios e até mesmo ganhos financeiros. A utilização da Mineração de Dados em redes sociais é uma forte tendência para descoberta de conhecimento neste tipo de bases de dados.

O ambiente de auxílio ao aprendizado Moodle será considerado aqui neste trabalho como uma pseudo rede social, pois segundo Recuero et. al (2009), ele proporciona conexões para as pessoas, mas em última análise, são as pessoas que constroem as redes. Portanto, a rede social é criada através da interação entre as pessoas que utilizam a ferramenta e criam conexões entre si e não somente pela existência do sistema Moodle. Por esta mesma razão, Consolo (2008) também define o Moodle como uma pseudo rede social, pois dentro dessa plataforma troca-se informações aumentando, assim, o conhecer de cada um, passando e recebendo, de alguns usuários, dicas muito importantes.

No Moodle existe o recurso chamado fórum, onde os alunos, participantes desta pseudo-rede social, publica assuntos diversos acerca de um assunto de aula ou mesmo da Instituição de Ensino Superior (IES). Por essa razão, o objetivo aqui é verificar os principais assuntos discutidos na atividade fórum do Moodle de uma IES e, com isso, descobrir conhecimento que ajudará na tomada de decisão.

No entanto, as postagens se dão de forma textual e não estruturada. Para um processo de descoberta de conhecimento, este tipo de dados precisa ser transformado para um formato numérico para que os algoritmos de mineração de dados possam ser aplicados. Na seção seguinte será explicado com mais detalhes como se faz esta transformação.

3. Mineração de Dados

Mineração de Dados (MD) é uma área de pesquisa multidisciplinar, envolvendo basicamente Banco de Dados, Estatística e Aprendizado de Máquina. A MD é parte principal de um processo que tem como entrada uma Base de Dados e como saída um Conhecimento [Fayyad et al., 1996]. Ela é dividida em tarefas como predição, clusterização e associação que devem ser escolhidas de acordo com análises exploratórias inicialmente feitas sobre os dados [Han et al., 2006; Tan et al., 2009; Faceli et al., 2011; Silva, 2015]. A MD tem sido amplamente utilizada em diferentes áreas, inclusive na educação. Quando se usa dados educacionais, define-se como Mineração de Dados Educacionais ou MDE [Silva & Silva, 2014; Romero & Ventura, 2010].

3.1 Análise de Agrupamento

A descoberta de grupos ou *clustering* trata-se de um modelo de MD que procura encontrar exemplares ou objetos de um conjunto de dados com atributos semelhantes. Portanto, a segmentação da base em grupos é feita a partir de medidas de similaridade. Em problemas que envolvem agrupamento não se encontra disponível o atributo especial classe ou rótulo. Por esta razão, dizemos que o aprendizado realizado pelos algoritmos de agrupamento é não-supervisionado [Jain & Dubes, 1988; Han et al., 2006; Tan et al., 2009; Jain, 2010; Faceli et al., 2011; Witten et al., 2001; Silva, 2015].

O aprendizado não-supervisionado é de grande desafio, pois não se tem o objetivo que se deseja alcançar, o que significa não conhecer o número de grupos da base de dados. E ainda, os exemplares estão distribuídos em um espaço de dimensão elevada com diferentes formatos e separação. Estes aspectos indicam a grande dificuldade para lidar em problemas de agrupamento de dados.

Em um contexto geral, o que o agrupamento faz é a descoberta de perfil. Mais especificamente em educação, a descoberta poderia ser útil para descobrir estilos de aprendizado dos alunos, disciplinas de interesses comuns e muitas outras aplicações.

Os algoritmos de agrupamento também podem ser divididos em Aprendizado de Máquina e Inteligência Computacional. Algoritmos típicos de Aprendizado de Máquina são o Agrupamento Hierárquico, k-Médias (k-Means), Agrupamento Espacial baseado em Densidade ou DBScan (*Density Based Spatial Clustering of Applications with Noise*). Por outro lado, como abordagem baseada em Redes Neurais temos o Mapa Auto-Organizável ou SOM (*Self-Organizing Map*) [Jain & Dubes, 1988; Han et al., 2006; Tan et al., 2009; Haykin, 2009; Jain, 2010; Faceli et al., 2011; Witten et al., 2001]. A seguir será apresentado em detalhes o K-Médias, algoritmo a ser utilizado neste estudo.

3.1.1 O algoritmo k-Médias

O k-Médias ou, em inglês, k-Means é o principal algoritmo de agrupamento particional (*Partitional Clustering*) [Silva, 2015]. O objetivo é encontrar particionamentos nos exemplares do conjunto de dados dentro de k grupos disjuntos, sendo k um valor inteiro maior que um. No entanto, o número de grupos representado por k deve ser dado como um parâmetro de entrada do algoritmo, como também a medida de distância - geralmente a Euclidiana. Com os parâmetros de entrada definidos é feita uma escolha aleatória de k distintos valores para centroides de grupos. A partir destes centroides, o

algoritmo iterativo gera as partições na base, fornecendo como resultado final a base de dados agrupada. A Figura 1 ilustra um exemplo da operação do algoritmo k-Médias. Na Figura 1a os exemplares são representados por dois atributos quaisquer (a_1 e a_2), constituindo o conjunto de dados (círculos pequenos e sem preenchimento); nessa figura ainda apresenta-se os centroides (círculos grandes e com preenchimento), cujas posições iniciais serão assumidas por escolhas aleatórias. O exemplar é comparado com os centroides através de uma medida de distância, sendo que a menor distância indica uma associação do exemplar ao centróide, veja Figura 1b. Por fim, Figura 2c, o centróide é atualizado, a partir da média dos atributos dos exemplares nele agrupado.

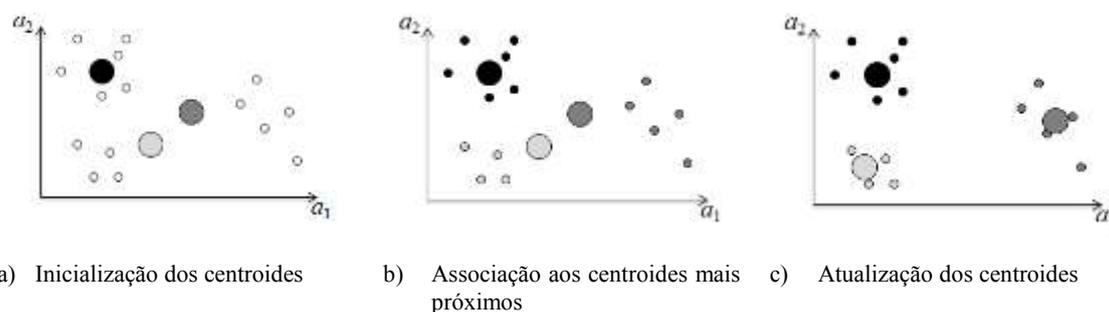


Figura 1 – Exemplo de operação do algoritmo k-Médias

O algoritmo k-Médias explicado nesta seção espera que os exemplares parem de fazer troca de centroides em alguma iteração, o que pode não acontecer. Por esta razão, pode-se considerar uma função objetivo para que seja minimizada (ou maximizada) até um valor aceitável como critério de parada do algoritmo ou que ao menos se defina um número máximo de iterações para que o algoritmo se estabilize [Tan et. al, 2011].

Em problemas de clusterização não há metodologia para validação de resultados. No entanto, ainda assim é possível medir desempenho. Em agrupamento a medida de desempenho é chamada de índice de validação de agrupamento (em inglês, *cluster validity indices*) e os métodos se dividem em interno, relativo e externo [Jain & Dubes, 1988]. Os índices internos baseiam-se na estrutura gerada pelo algoritmo de clusterização. Portanto, a grande maioria dos quantificadores mensuram a coesão e o isolamento dos grupos. Exemplos de índices internos são Coeficiente de Silhueta, Índice de Dunn, Índice de Davies-Bouldin entre outros.

3.2. Mineração de Texto

Para aplicação do algoritmo de agrupamento nos dados oriundos da ferramenta Fórum do Moodle é necessário utilizar um modelo de representação, que transforma as palavras de cada postagem em um peso. O resultado disso é chamado de corpus, e a partir desse, é que o algoritmo de agrupamento é executado. Esse processo ocorre da seguinte maneira [Göker, Davies, 2009]:

i. *Tokenization*: a partir do caractere espaço, as palavras da postagem são separadas em *tokens*. Os caracteres especiais como vírgulas (","), pontos de exclamação ("!") e de interrogação ("?"), entre outros são removidos, assim como números.

ii. *Stopwords*: palavras como preposições, pronomes, artigos, advérbios que são comuns a ambos os vocabulários são removidas do processo;

iii. *Stemming*: as palavras resultantes das etapas anteriores passam por uma normalização ortográfica para que sejam reduzidas ao radical. Este processo é importante, pois permite que palavras com o mesmo radical sejam consideradas como semelhantes.

Ao final deste processo gera-se um vocabulário de palavras chamado de *bag-of-words*. O *corpus* representado será gerado com base nesse vocabulário. A representação do *corpus* pode ser feita, usando um modelo chamado de vetorial [Göker, Davies, 2009]. O modelo vetorial busca representar o *corpus* em forma de vetor, atribuindo pesos aos termos (palavras) que indicam importância. Uma das formas de se fazer a ponderação é combinando a frequência que um termo aparece em um *corpus* ou *tf* (do inglês, *term of frequency*) com a frequência invertida do documento ou *idf* (do inglês, *inverse document frequency*). A combinação de $tf \times idf$ define a importância (peso) de um termo dentro do *corpus* [Göker, Davies, 2009].

Assim, depois de calculado o peso de cada termo do documento, tem-se como resultado uma base de dados numérica formada por vetores de documentos e, assim, conseqüentemente, aplica-se o algoritmo de agrupamento de dados.

5. Metodologia e Resultados Experimentais

Para a realização dos experimentos, utilizou-se como base de dados postagens publicadas no período de dois anos em um fórum do Moodle de uma IES chamado de críticas e sugestões.

A ferramenta utilizada nos experimentos foi a RapidMiner, que possui diversos tipos de recursos para Mineração de Dados e também de texto. Os dados foram exportados da base do Moodle em formato XML, para que todas as postagens do fórum fossem carregadas na ferramenta RapidMiner.

O processo iniciou-se com seleção das 1.000 publicações no fórum do Moodle. Através da ferramenta Rapid Miner realizaram-se inicialmente alguns passos da preparação dos dados, como: o *tokenize* que separa cada palavra de um post em um termo e onde se fez o tratamento para retirar as tags HTML; extração de *stopwords* que elimina termos como preposições, verbos e todos aqueles que podem ser comuns a outros posts. Após análise desse primeiro processo, notou-se a necessidade de inclusão da etapa de *stemming* que extrai de cada termo o seu radical. Isso se deu pelas diferentes formas de tratamento das publicações, como por exemplo, em relação ao coordenador como “prof.” ou “professor”. A importância destas operações no processo de preparação de documentos relaciona-se também com o número de termos existentes, que por sua vez estão vinculados diretamente com a dimensão dos valores ponderados ($tf \times idf$).

Na tarefa de Mineração de Dados, embora o algoritmo de descobrir grupos aqui utilizado, *k*-Médias, ter grande simplicidade na parametrização, como discutido anteriormente, a tarefa de agrupamento de dados não é trivial. A escolha do valor de grupos a segmentar, *k*, passa a ser um problema quando não se sabe quantos grupos existem na base de dados. Na prática, o que se faz é variar o valor de grupos, iniciando *k* com valor maior ou igual a 2 e, a cada variação, avalia-se a qualidade do grupo através de um índice. Na literatura existem diversos índices para medir a qualidade de um ou mais grupos gerados. Dentre estes índices foi escolhido o chamado soma quadrática, que avalia a coesão dos objetos de um grupo, considerando como referência o centro do grupo [Jain et. al, 1999].

A Figura 2 mostra o resultado da investigação do número de grupos. Nota-se que, a partir de $k \geq 5$, a soma quadrática começa a reduzir lentamente, sugerindo uma estabilização do número de grupos. No entanto, é importante ressaltar que a decisão por oito grupos ($k = 8$), além de apresentar o menor resultado, que é o valor ótimo dessa análise e, por isso é suficiente como critério, a análise de cada valor de k permitiu concluir que para valores menores a 8 haveriam particionamentos das postagens em mini grupos, os quais resultariam em um volume pequeno de até duas postagens. É importante lembrar que a quantidade de postagens em cada grupo, assim como o relacionamento entre grupos são fatores importantes para a geração de conhecimento.

Para melhor analisar os resultados, as postagens mapeadas em cada grupo foram analisadas através de uma investigação manual, e a partir de então, foram criados temas para cada grupo de postagens. Ou seja, o grupo 1, G1, tem o tema Comunicado referente à avisos da Faculdade sobre as parcerias estabelecidas com empresas e, conseqüentemente, atribuiu-se aos demais grupos os seguintes temas: G2: Organização e Burocracia, G3: Comunicado de cursos, bolsas e descontos, G4: Planejamento, G5: Respostas à comunicados, G6: Avaliação, G7: Prova Unificada e G8: Infraestrutura. Os resultados estão representados nas Figuras 3 e 4 que permitem, de modo proporcional, analisar cada grupo, o tema que o mesmo abrange e a relação destes grupos entre si. É importante notar que esta intersecção entre os grupos, representam postagens que abordam mais de um tema.

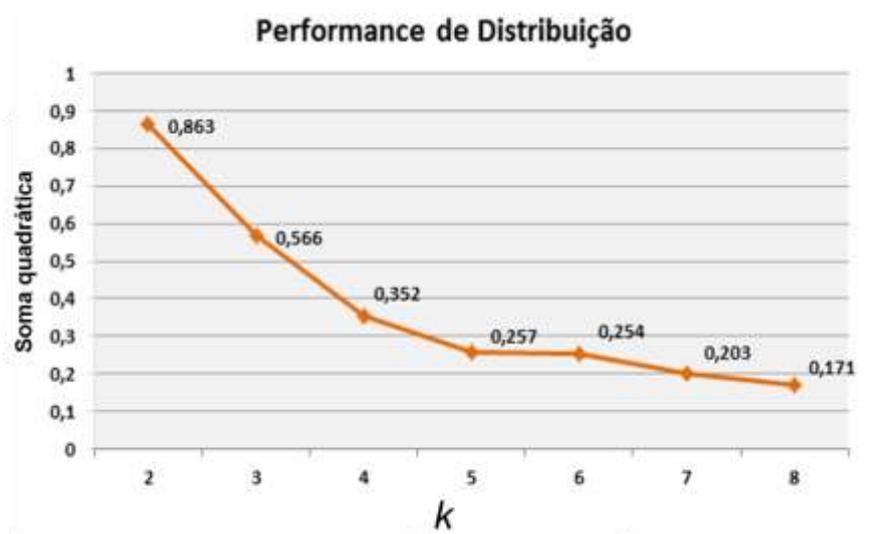


Figura 2 - Gráfico que mostra a variação do índice de qualidade de agrupamento, soma quadrática, para os valores de k analisados.

Antes de especificar as informações contidas e analisadas dentro de cada um dos oito grupos, abordam-se os dados quantitativos dos mesmos na Figura 3. Nesta, pode-se perceber o volume de postagens em cada grupo, bem como permite que se faça uma ordenação. Isso servirá para, após a geração do conhecimento, definir estratégias de ações, como, por exemplo, a prioridade que cada assunto será abordado.

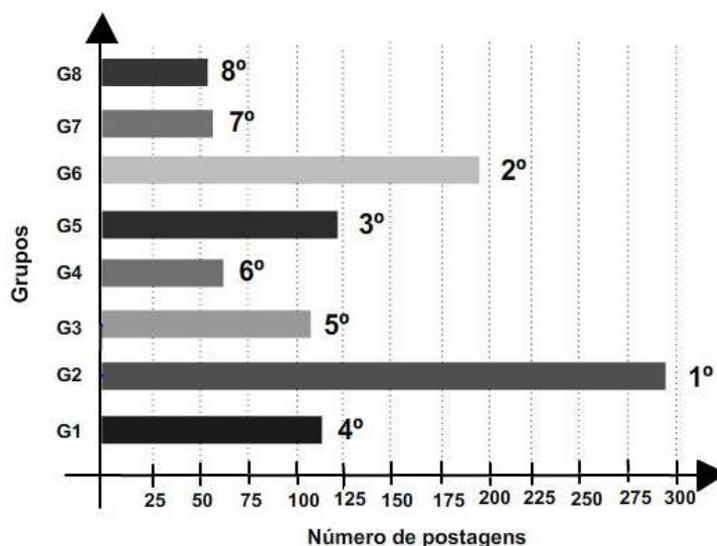


Figura 3 – Volume de postagens em cada grupo descoberto pelo k-médias

A partir da análise dos grupos, Figura 4, é possível se ter ideia do volume de postagens e suas interações (áreas comuns) e, conseqüentemente, pode-se obter conhecimento as críticas, sugestões, ideias, pensamentos, impressões e constatações das principais publicações dos alunos. As postagens do grupo G2, maior deles, são relacionadas à organização e burocracia. O grupo se relaciona com os grupos G1 e G2, com postagens sobre comunicados e, também G8 com postagens de infraestrutura. Por sua vez, o grupo G8 se relaciona com postagens sobre planejamento, G4, prova unificada, G7 e avaliação, G6. Por fim, o grupo G5, postagens sobre respostas à comunicados, se relaciona com os grupos G3 e G6.

A partir deste conhecimento gerado com base nas postagens do Moodle, pode-se definir um plano de ação para os temas das postagens dos alunos, permitindo, inclusive, a definição de prioridades.

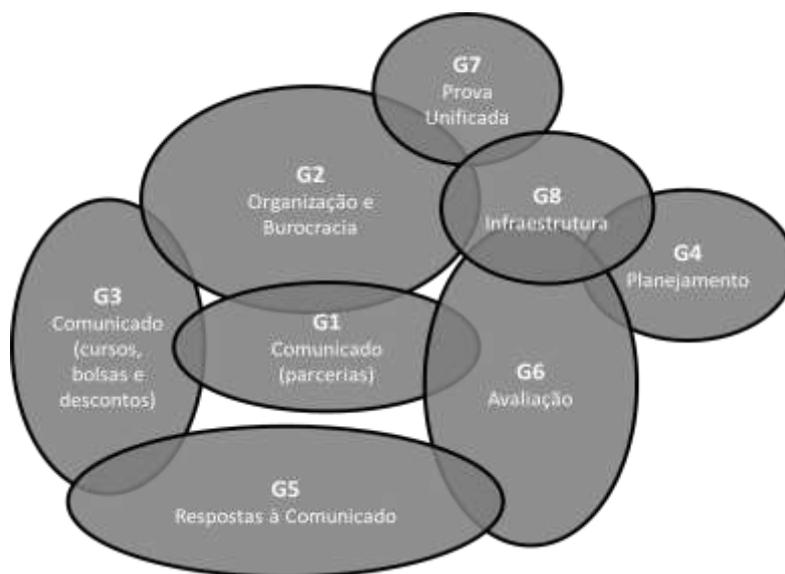


Figura 4 – Representação Gráfica dos grupos descobertos que mantêm o volume e relações.

6. Conclusão

Empresas, governos, instituições das mais diversas, incluindo de Ensino Superior, aplicam técnicas de Mineração de Dados para extrair informações de redes sociais que lhes podem ser úteis. A Mineração de Dados além de permitir revelar conhecimentos antes não imaginados, trará também uma redução no tempo dispensado para obtenção dessas informações, pois se comparado às técnicas tradicionais, de pesquisa e obtenção de opiniões, a Mineração de Dados nas redes sociais é mais rápida, visto que nos métodos tradicionais há a necessidade de criação de formulário de pesquisa, distribuição, preenchimento por parte dos pesquisado, computação dos dados obtidos e por fim a análise, todas essas etapas requerem muito tempo. Já na Mineração de Dados nas redes sociais a base de dados está disponível, sendo necessária apenas a aplicação do processo para descoberta de conhecimento. Vale ressaltar que nas pesquisas convencionais muitas vezes podem não abranger os assuntos de maior interesse pelos alunos como, por exemplo: em uma pesquisa de satisfação à alunos, se o pesquisador não souber de antemão, quais são os pontos mais preocupantes, então esse não poderá o incluir na pesquisa e perder essa informação. Por outro lado, na mineração de dados em redes sociais, os alunos poderão se expressar da maneira que melhor lhe convier sobre os diversos assuntos expor seu ponto de vista, e após a mineração de dados o conhecimento obtido ilustrará cada detalhe que pode ser omitido em formulário pré-fixados de pesquisa de opinião.

Neste trabalho, a aplicação de Mineração de Dados em publicações no Moodle se tornou um processo importante por permitir a descoberta de grupos e, conseqüentemente, gerar índices que podem ser monitorados pelos gestores com o intuito de melhorar a qualidade da Gestão de uma Instituição de Ensino Superior. Como trabalho futuro pretende-se pesquisar estratégias para a descoberta dos temas em cada grupo, bem como a geração automática dos agrupamentos preservando as interseções de assuntos.

Referencias Bibliográficas

- Agarwal, A.; XIE, Boyi; Vovsa, I. & Passonneau, R. (2011). *Sentiment Analysis of Twitter Data*. Columbia University. In Proceedings of the Workshop on Languages in Social Media (pp. 30-38). Association for Computational Linguistics.
- Baggi, C., A. S., & Lopes, D. A. (2011). Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. *Avaliação*, Campinas; Sorocaba, SP, v. 16, n. 2, p. 355-374.
- Consolo, A. Rede Social. Disponível em: <<http://www.moodlelivre.com.br/categoria/57-artigos/30-rede-social-e-o-moodle.html/>> acessado em 24 de maio de 2015.
- Faceli, K., Lorena, A. C., Gama, J., & Carvalho, A. C. P. L. F. (2011). *Inteligência Artificial: Uma Abordagem de Aprendizagem de Máquina*, LCT.
- Fayyad, U.; Piatetski-Shapiro, G.; Smyth, P (1996). From Data Mining to Knowledge Discovery: An Overview. In *Advances in Knowledge Discovery and Data Mining*. California. p. 1-34.
- Göker, A., & Davies, J. (2009). *Information retrieval: searching in the 21st century*. Chichester U.K.: John Wiley & Sons, Inc.p.6-7.
- Han, J. & Kamber, M. (2006): *Data Mining Concepts and Techniques*. Editora Elsevier. Morgan Kaufmann, 3rd edition.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. New Jersey: Prentice-Hall.

- Jain, A.K.; Murty, M.N. & Flynn, P.J. (1999). Data clustering: a review. *ACM Comput Surv*, New York, vol. 31, No 3, 264–323.
- Recuero, R. *Redes Sociais na Internet*. São Paulo: Meridional, 2009.
- Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 40(6), 601-618.
- Silva, L. A. *Mineração de dados: uma abordagem introdutória e ilustrada*. Editora Mackenzie (Coleção conexão inicial, v. 11), 1. ed., São Paulo, 2015.
- Silva, L. A., & Silva, L. (2014). Fundamentos de Mineração de Dados Educacionais. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação* (Vol. 3, No. 1).
- Tachizawa, T., & de Andrade, R. O. B. (1999). *Gestão de instituições de ensino*. FGV Editora.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2009). *Introdução ao datamining: mineração de dados*. Rio de Janeiro: Ciência Moderna.
- Tanaka, V. R. D. S., & Pessoni, L. M. D. L. (2011). A gestão do ensino superior: o gestor e seu papel. *Anais do Seminário sobre Docência Universitária*, 1(1).
- Witten, I. H., Frank, E., & Mark, A. (2011). Hall (2011)." *Data Mining: Practical machine learning tools and techniques*.