

Ciência de dados aplicada na análise de processos cognitivos em grupos sociais: um estudo de caso

Viviani R. Kwecko¹, Fernando P. de Tolêdo², Fernanda P. Mota¹,
Gabriel da S. Almeida², Sam da S. Devincenzi¹, Fabiana Z. Ferreira¹,
Alessandro de L. Bicho³, Silvia S. da C. Botelho³

¹PPG em Educação em Ciências: química da vida e saúde (PPGEC)

²PPG em Modelagem Computacional (PPGMC)

³Centro de Ciências Computacionais (C3)

Universidade Federal do Rio Grande (Furg) – Rio Grande/RS – Brasil

viviani.kwecko@riogrande.ifrs.edu.br

Abstract. *The new technologies appear as facilitators in the process of interaction between development agents in the territory and virtual knowledge systems, from which emerges the concept of Smart City. However, we realize that this intelligence of the city disregards in its scope the factors associated with the construction of knowledge generated by public opinion. In this study, we sought to discuss and present, through a case study, how Data Science can become a powerful tool capable of revealing emerging processes of shared cognition among members of a social group, as well as the implications revealed by socio-educational order contained in the generated knowledge.*

Resumo. *As tecnologias emergentes surgem como facilitadoras no processo de interação entre os agentes de desenvolvimento no território e os sistemas de conhecimento virtuais, dos quais surge o conceito de Cidade Inteligente. Contudo, percebemos que essa inteligência da cidade desconsidera em seu escopo os fatores associados à construção de saberes gerados pela opinião pública. Neste estudo, buscamos discutir e apresentar por meio de um estudo de caso como a Ciência de Dados pode transformar-se em uma potente ferramenta, capaz de revelar processos emergentes de cognição partilhada entre os membros de um grupo social, bem como as implicações reveladas pela ordem sócio-educacional contida no conhecimento gerado.*

1. Introdução

O desenvolvimento tecnológico que caracteriza as últimas décadas de progresso econômico das sociedades ocidentais vêm transformando a cidade, outrora industrial, numa cidade da informação e do conhecimento. As tecnologias emergentes surgem como facilitadoras no processo de interação entre os agentes de desenvolvimento no território e os sistemas de conhecimento virtuais, dos quais emerge o conceito de território inteligente. Segundo a União Europeia [Manville et al. 2014], *Smart Cities* são sistemas de pessoas interagindo e usando energia, materiais, serviços e financiamento para catalisar o desenvolvimento econômico e a melhoria da qualidade de vida. Esses fluxos de interação

são considerados inteligentes por fazerem o uso estratégico de infraestrutura e serviços, de informação e comunicação com planejamento e gestão urbana para dar resposta às necessidades sociais, educacionais e econômicas da sociedade.

Algumas abordagens vem sendo utilizadas para analisar o nível de inteligência das cidades. São exemplos: *i) European Smart Cities (2016)* [Khatoun and Zeadally 2016], metodologia que descreve a inteligência através de seis características (economia inteligente, pessoas inteligentes, governança inteligente, mobilidade inteligente, meio ambiente inteligente, vida inteligente), 31 fatores e 74 indicadores; *ii) Urban Systems* [Systems 2018], apresenta o Ranking *Connected Smart Cities* que considera a verificação em 11 eixos, (mobilidade, urbanismo, meio ambiente, energia, tecnologia e inovação, economia, educação, saúde, segurança, empreendedorismo e governança) e 73 indicadores; *iii) IESE Cities in Motion* [Gomyde 2017]; *iv) o projeto "Brasil 2030: Cidades inteligentes e humanas"* [Gomyde 2017], o qual considera a verificação em 4 eixos (Governança e Arquitetura, Urbanismo e Antropologia, Tecnologia e Segurança) e 205 indicadores; e os estudos do brasileiro [Afonso et al. 2013] que propõe, para além dos indicadores, a constituição de um índice de maturidade para cidades inteligentes.

Contudo, percebemos que estas abordagens de análise da inteligência da cidade não consideram em seu escopo os fatores associados à construção de saberes gerados pela população acerca destes indicadores. De acordo com [Lévy 2007], essa inteligência está em constante processo de construção e todos os saberes devem ser considerados. Assim, neste estudo buscamos refletir e discutir questões acerca da mediação tecnológica como extensão de um processo comunicativo de co-construção de significados. Consideramos como significados o conjunto de saberes relacionais (conjunto de opiniões) oriundos das interações cidadão-cidade virtual. Diante deste fato, nosso objeto de investigação está centrado nas falas e/ou discursos realizados em contextos de interações digitais¹. O foco principal são aprendizagens construídas em processos sociais coletivos, participativos, onde o saber não é gerado em estruturas formais de ensino, mas no campo da educação não formal. Falar sobre as relações entre aprendizagem e saber, fora do sistema regular de ensino, implica em participar do amplo debate epistemológico sobre a produção de conhecimento no mundo contemporâneo.

De modo a validar nossas hipóteses desenvolvemos um estudo de caso no qual aplicamos um processo de Mineração de Conhecimento e uma Análise Estatística Implícitiva. Através da mineração buscamos identificar e analisar as relações impressas por um conjunto de mensagens extraídas do site da rede social Facebook[®] com os eixos da inteligência da cidade, bem como o sentimento da população a ele relacionado. E com a análise estatística nos propomos a responder as seguintes implicações: *i)* que ordem sócio-educacional está envolvida nesse conhecimento gerado na mineração de dados? *ii)* estamos a repetir o tradicional meio de extrair normas e médias da diversidade social e humana, uma vez que se trata de determinar padrões e regularidades? Assim organizamos este artigo apresentando na Seção 2 a fundamentação teórica acerca dos modelos metodológicos selecionados para essa pesquisa. A Seção 3 apresentamos os caminhos metodológicos traçados que possibilitam os resultados apresentados. Na Seção 4, os re-

¹Segundo [Belloni 2001], interatividade é a característica técnica que significa a possibilidade de o usuário interagir com uma máquina. Já a interação quer dizer ação recíproca entre dois ou mais atores onde ocorre intersubjetividade, isto é, encontro de dois sujeitos – que pode ser direta ou indireta.

sultados da avaliação qualitativa são apresentados e discutidos. Finalmente, na Seção 5 a conclusão explicita nossa compreensão diante dos resultados obtidos, bem como aponta para nossas perspectivas de trabalho futuro.

2. Descoberta de Conhecimento em Banco de Dados e Análise Estatística Implicativa (A.S.I.)

Várias técnicas fazem parte do conjunto de ferramentas, procedimentos e algoritmos conhecidos como Mineração de Dados (*Data Mining*), que por sua vez faz parte de um processo conhecido como Descoberta de Conhecimento em Bancos de Dados (*Knowledge Discovery in Databases - KDD*). O procedimento de Mineração de Dados é um processo que pode ser aplicado a qualquer problema que requeira a identificação de padrões em dados. Entretanto, apesar das definições levarem a crer que o processo de extração de conhecimento se dá de forma automática, de fato isso não é verdade. Particularmente, o processo encontra-se associado à aplicação de algoritmos capazes de identificar e extrair padrões relevantes em um conjunto de dados informados, mas cujos resultados precisam de uma análise humana [Hand 2007].

Tradicionalmente, os métodos de mineração de dados são divididos em aprendizado supervisionado (preditivo) e não-supervisionado (descritivo) [Han et al. 2011]. Apesar desse limite, [Wang 2005] propõe abordagens semi-supervisionadas. A diferença entre os métodos de aprendizado supervisionados e não-supervisionados reside no fato de que os métodos não-supervisionados não precisam de pré-categorização para os registros, ou seja, não é necessário um atributo alvo. Tais métodos geralmente usam alguma medida de similaridade entre os atributos [McCue 2014]. As tarefas de agrupamento e associação são consideradas como não-supervisionadas. Já no aprendizado supervisionado, os métodos são providos com um conjunto de dados que possuem uma variável alvo pré-definida e os registros são categorizados em relação a ela. As tarefas mais comuns de aprendizado supervisionado são a classificação (que também pode ser não-supervisionado) e a regressão [McCue 2014].

Durante o processo de mineração, diversas técnicas podem ser testadas e combinadas a fim de selecionar a melhor (ou uma combinação de técnicas) [McCue 2014]. Suas análises centram-se na transformação de um mapeamento de dados brutos em arranjos de informação compactos (relatórios, gráficos, etc) que possibilitam a identificação de novos padrões, válidos e potencialmente úteis para a configuração de referenciais preditivos de ações futuras [Fayyad et al. 1996]. A partir da metodologia desenvolvida por [Fayyad et al. 1996], podemos descrever de forma genérica o processo a partir 6 fases.

A fase inicial de todo processo ocorre com base no levantamento de requisitos, em que o pesquisador busca determinar e entender as intencionalidades do processamento dos dados. A partir dessa definição dos objetivos do processamento, são selecionados o conjunto de dados-alvo, as técnicas e as ferramentas a serem aplicadas ao longo do transcurso. Algumas técnicas de mineração contêm parâmetros que são utilizados em seu funcionamento, também faz parte desta etapa encontrar essas propriedades, para que o método possa ser o mais preciso e ágil possível.

A segunda fase é caracterizada pela organização dos aspectos, sendo composta pelas etapas: (i) extração dos dados; (ii) pré-processamento, remoção dos erros e exemplos fora do padrão pré-estabelecido; (iii) definição de estratégias para valores faltosos;

e (iv) formatação dos dados de acordo com os requisitos da ferramenta de mineração. Já na terceira fase ocorre a transformação dos dados em um conjunto de atributos. Somente após essa etapa é que serão aplicadas as técnicas de mineração para alcançar os objetivos definidos na etapa inicial. Sendo assim, na quarta fase é escolhida uma técnica de mineração cujos objetivos podem ser descritos como tarefas de classificação, regressão, agrupamento, predição, etc. Na quinta fase adentramos na etapa de interpretação e validação dos padrões identificados. Este passo inclui visualizar os padrões extraídos que resumem a estrutura das informações presentes nos dados. Além da visualização, são utilizadas medidas tanto técnicas (precisão, erro médio, erro quadrático e taxas de falsos positivos e falsos negativos) quanto subjetivas (contexto e complexidade dos padrões) para avaliar os modelos extraídos. Na última fase do processo teremos conhecimento em forma de padrões que podem ser utilizados associados ou não a outros sistemas.

Segundo [Damasceno 2015] a fase de validação é construída durante o processo de seleção dos dados, sendo recomendado dividir o conjunto inicial de dados em 3 (três) diferentes agrupamentos destinados ao treinamento, a validação e ao teste, sendo a proporção de 70% para treinamento, 20% para validação e 10% para testes. Quando dividido em 2 (dois), geralmente a divisão é de 70% do conjunto para treinamento e 30% para testes. Além disso, a aprendizagem do sistema ocorre de 2 (dois) modos: supervisionada ou não supervisionada.

A Análise Estatística Coesitiva, Hierárquica e Implicativa (A.S.I.) é um método estatístico multidimensional, aplicado à pesquisas qualitativas que buscam uma análise hierárquica de similaridade ou semelhanças [Almeida et al. 2000] entre um conjunto de dados. Segundo [Couturier et al. 2004], essa abordagem permite ao pesquisador (i) extrair de um conjunto de dados regras de associação entre variáveis; (ii) fornecer um índice de qualidade de associação; e (iii) representar uma estruturação das variáveis obtida por meio destas regras. Para tanto utilizamos o *software* de Classificação Hierárquica Implicativa e Coesiva (*Classification Hierarchique Implicative et Cohesitive - CHIC*).

No mapeamento traçado pelo *CHIC*, o pesquisador desenvolve suas interpretações e significações construídas a partir de: oposições ou aproximações; semelhanças, proximidades ou afastamentos; contradições ou repetições, que pela sua trajetória ou estrutura revelam as concepções profundas, mais autênticas dos indivíduos. Possibilita a análise dos dados coletados em três eixos: construção de árvores hierárquicas, árvores de similaridade e grafos implicativos. O índice de similaridade foi definido por [Lerman and Manski 1981] e permite construir uma hierarquia ascendente. De forma similar, a intensidade da implicação pode ser utilizada para construir uma árvore de interação hierárquica. Além disso, *CHIC* oferece a possibilidade de gerar um grafo implicativo, através do qual são estabelecidas as regras de associação entre as variáveis selecionadas [de Queiroz et al. 2016]. A possibilidade de agrupar diferentes elementos perceptivos em um mesmo eixo temático para levantar os atributos que definem a preferência de um indivíduo por um ambiente são características que tornam a estatística multivariada adequada ao objetivo pretendido. A análise hierárquica coesitiva e implicativa, segundo [Gras and Almouloud 2002], fornece uma ferramenta prestante a visualizar, organizar, construir modelos e explicar fenômenos associados aos dados, evidenciando o comportamento ou as imagens mentais de sujeitos em diferentes situações.

3. Metodologia

Essa investigação assume o estudo de caso como estratégia de pesquisa [King et al. 1994] cuja concepção metodológica objetiva uma análise quanti-qualitativa/quali-quantitativa ou métodos mistos a partir da tipologia exploratória, proposta por [Creswell 2007] e [Creswell and Clark 2015], cujos os resultados quantitativo contribuem para o desenvolvimento do subseqüente método qualitativos e vice-versa. Para uma efetiva condução da pesquisa, organizamos o planejamento operacional em dois momentos sequenciais e complementares: no primeiro executamos o processo KDD; e com os resultados obtidos implementamos a A.S.I., detalhados nos itens abaixo.

Nossa proposta foi aplicada aos dados extraídos de 33 grupos do site de rede social Facebook, no período que compreende a primeira mensagem de criação de cada grupo até o limite de julho de 2017, sendo que o post mais antigo é datado em março de 2012. Especificamente foram obtidos um total de 53.355 mil posts, desconsiderando todo e qualquer comentário associado a ele.

Na etapa de pré-processamento dos dados, desenvolvemos uma ferramenta Web para o treinamento do sistema, através da qual as mensagens foram anotadas/classificadas quanto: (i) aos 11 eixos de inteligência da cidade sendo, eles mobilidade, urbanismo, meio ambiente, energia, tecnologia e inovação, economia, educação, saúde, segurança, empreendedorismo e governança; (ii) a polaridade da opinião, cuja classificação determinou o sentimento expresso em cada mensagem como *post* positivo, *post* negativo, *post* neutro; (iii) a sintaxe adjetivas e substantivas das palavras. Obtemos um total de 11.624 posts anotados/classificados, sendo que 1050 destas obtiveram anotações realizadas por no mínimo 5 pessoas, alcançando uma média de 3 diferentes votos no mesmo eixo. Quanto a validação do sistema, 210 *posts* anotados (20% da totalidade) foram reservados, restando um total de 840 utilizados durante o treinamento.

A fim de procedermos com o aprendizado supervisionado (preditivo), solicitamos a um grupo de discentes e docentes da Universidade Federal do Rio Grande - FURG e do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - IFRS, que indicassem expressões chaves que julgassem corresponder aos eixos da inteligência da cidade. É importante observarmos nossa opção pela não utilização de especialistas na área para indicação desse conteúdo, isso justifica-se pela característica dos dados que analisamos: a opinião de senso comum de cidadãos em grupos no Facebook. Foram anotadas manualmente uma média de 150 palavras para cada eixo.

Para a terceira etapa o método utilizado para a Mineração de Conhecimento utilizando o método *Expectadores mutual information measure*² (método desenvolvido em Stanford), de modo a obter o delineamento de um agrupamento por padrões. Para definir o relacionamento entre os padrões, podem ser aplicadas, também, aos grupos anteriormente definidos, técnicas de Visualização de Informações como, por exemplo *núvem de tags*³ Ao final obtemos um conjunto de 33 mensagens classificadas (para cada eixo um conjunto de positivo, negativo e neutro), um conjunto de substantivos relacionado e três conjuntos de adjetivos (positivo, negativo e neutro).

Completando nossa abordagem metodológica passamos a utilizar o processo de

²<https://nlp.stanford.edu/IR-book/html/htmledition/mutual-information-1.html>

³<http://vispublica.gov.br/vispublica/publico/catalogo/google/wordCloud.jsp>

Análise Estatística Implicativa (A.S.I.). A.S.I. foi realizada neste trabalho com o auxílio do *software* de Classificação Hierárquica Implicativa e Coesiva - *CHIC* (Versão 6.0, 2012), cuja sigla significa *Classification Hierarchique Implicative et Cohesitive*. Por exigência do *software CHIC*, a estruturação dos dados foi realizada de forma binária, por meio de planilhas em MS Excel, com extensão csv, que separa os dados por vírgulas. A partir dos resultados obtidos na etapa da Mineração de Dados foram elaboradas três planilhas distintas para cada um dos 11 eixos que compõem a inteligência da cidade. As tabelas foram organizadas de forma a apresentar na primeira coluna os números de registro (ids) de cada post armazenado no banco de dados da pesquisa e a primeira linha apresenta os termos representativos que emergiram após o processo KDD. Para o *CHIC*, a frequência dos termos foi analisada em cada mensagem considerando a palavra de maior regularidade para a composição do valor entre 0 e 1. Neste trabalho, limitaremos nossa análise ao eixo Educação Substantivo-positivo cujos dados encontram-se distribuídos em 193 mensagens classificadas pelo sistema como substantivo-positivos e 729 termos substantivos representativos indicados pelo KDD.

4. Discussão dos Resultados

No pós-processamento realizamos a análise e tratamento dos 729 termos representativos agrupando a partir de sua semântica, 31 conjuntos nas seguintes categorias: Gestor, Cidade, Comunicação, Secretarias, Didática, Capacitação; Escola, Universidade, Estudantes, Gênero, Atividades, Apoio, Avaliação, Educador, Infraestrutura, Cultura, Esporte, Divulgação, Currículo, Motivação, Tradição, Cultura Popular, Saúde, Ambiente, Inscrição, ONGs, Tempo, Cidadania, Festejos e Fomentos. Com base na frequência dos termos em cada post, o *CHIC* organizou as variáveis em grupos a partir das diferenças e semelhanças entre as classes. Os resultados indicam que a taxonomia e o conhecimento não revelam um conjunto de características intrínsecas à cidade, mas sim percepções de fatores que constituem determinadas circunstâncias. Os aspectos de uma cidade apresentam-se como padrões que ao mesmo tempo ordenam e objetivam a multiplicidade da convivência humana, legitimando formas de experienciá-la, habitá-la ou governá-la. A observação das diferentes dimensões que compõem uma cidade encarnam múltiplas micro-regularidades entre inúmeras variáveis heterogêneas e, de modo algum, apresentam-se como regulamentos.

Para a análise dos dados com o uso do *CHIC* selecionamos a opção “nós significativos”, “cálculo longo”. O tipo de implicação foi segundo a teoria clássica. O tipo de lei foi a lei binominal. Estas escolhas orientam os cálculos e estão adequadas a amostra. Segundo [Martins and Théophile 2009] quanto mais próximo a 1,0 e acima de 0,7 melhor a confiabilidade dos dados, assim para a análise de similaridade consideramos os valores iguais ou superiores a 0,762. Foram identificadas no *CHIC* 36 classificações. Contudo, acima de 0,70 se apresentam 23 classificações que foram agrupadas em 5 grades de classes, sendo que os nós significativos encontram-se nos níveis: 1, 6, 12, 15, 17, 20, 24, 26, 29, conforme apresentado na figura 1. Interpretamos a similaridade dos dados de forma a identificar no conjunto de dados proximidade de respostas comuns ou comportamentos semelhantes. A figura 1 apresenta as classes selecionadas pelo software, sendo que para essa análise serão consideradas somente aquelas contêm elementos pertinentes para essa etapa de nossa pesquisa.

Destacamos inicialmente a Classe A (Cidade, Secretarias) com maior índice de

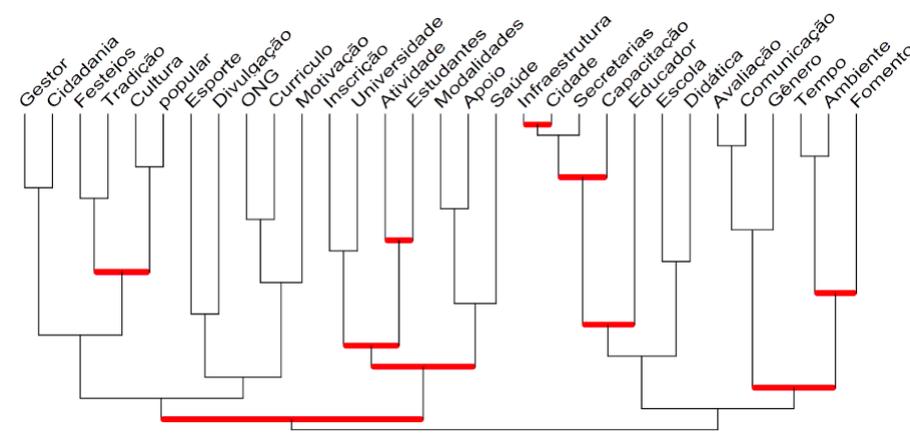


Figura 1. Gráfico de similaridade com base nos dados tratados.

similaridade igual a um (1). O conteúdo dessas postagens relacionam o conceito de educação ao trabalho desenvolvido pelas secretarias municipais da cidade do Rio Grande/RS, sem no entanto referir-se somente Secretaria Municipal de Educação que foi mencionada em 64 posts. Sendo assim, podemos supor que para a população todas as ferramentas de gestão municipal são percebidas como elementos positivos para a promoção da educação. Em síntese, esta seria uma essa cidade educadora no qual o território é representado por uma profusão de histórias, como uma espécie de Wikipédia⁴, em que toda a população pode construir livremente, acessar, relacionar, opinar, compartilhar e se questionar continuamente a partir do seu capital de conhecimento. Outro índice importante revela uma similaridade de 0.999823 entre a Classe B (Infraestrutura, Cidade, Secretarias e Capacitação). Os posts relacionados relatam valores expressos pela população quanto à formação e qualificação. Porém, para apurar nossa análise observamos essa indicação considerando a frequência dos posts que referem processos de Inscrição em programas, que é mínima. Das 193 mensagens de educação positiva, cerca de 3% expõem referências a matrículas em cursos.

Se a população destaca valor para a capacitação, por que esta percepção não encontra-se refletida em posts que indiquem um aumento de procura por cursos? A distinção entre educação, entendida enquanto instrução, e educação entendida como produção-formação de homens é uma velha distinção que encontra uma longa e tensa tradição teórica e prática. A hipótese que norteia nossa reflexão concebe a construção desse ato educativo na sociedade como um movimento através do qual as pessoas explicitam seu desejo em aprender em locais que não se reduzem apenas a espaços institucionalizados. Recentemente, o coletivo [Zemos 1998] cunhou o termo “Educação Expandida”, para significar a indissociação entre a educação e a comunicação, principalmente em redes de aprendizagem geradas fora das instituições educativas. A questão central não é tanto o que se aprende, mas sim como, onde, quando e com quem se aprende, ou seja quais são os novos espaços, pessoas e redes com os quais as pessoas querem e pode partilhar saberes [Zemos 1998].

Observamos que essa educação expandida também relaciona-se com a Classe C

⁴Espécie de enciclopédia colaborativa, universal e multilíngue estabelecido na internet

(Comunicação e Gênero) que tem uma similaridade 0,99966. Posts problematizam questões sobre a identidade e a diferença para o tema de gênero. E esse debate é fundamental, na medida em que reconhece o outro social enquanto diferente e singular, em que a comunicação passa a ser um fenômeno estruturador das relações sociais e a educação a responsável pela construção desse saber social. Outras Classes que chamam a atenção são a Classe C (Estudantes Modalidades) com índice de similaridade 0,9627, a Classe D (Universidade e Atividade) com índice de similaridade 0,9567 e a Classe E (Didática Avaliação) com índice de similaridade 0,9425. Verificamos que a população, mesmo em meio ao território mais banal como é o das mídias sociais, problematiza aquilo que [Zabalza 2004] vai indicar como as três condições fundamentais para a aprendizagem: eliminar os obstáculos que dificultam a aprendizagem; motivar e elucidar o sentido e os benefícios das novas aprendizagens; e superar a tendência inata a inércia. Precisamos debater as modalidades de ensino, a necessidade dos processos avaliativos e principalmente novas formas de motivar a participação ativa dos estudantes. O que percebemos a partir dessa experiência de escuta é o processo de internalização do conceito de educação para a população dessa localidade, que por meio de reflexões coletivas, promove a aproximação entre o sentido pessoal (de cada post) e o significado social da palavra (implicação).

Nossa análise abordou apenas algumas classes indicadas pela Árvore de Similaridade do *CHIC*, devido a limitação desse exercício de escrita. Outro tratamento proporcionado pela metodologia da A.S.I. é a análise coesitiva. Esta foi realizada a partir da associação entre variáveis ou grupo de variáveis a partir de meta-regras não simétricas do tipo "se a, então provavelmente b". Isso significa que a coesão de uma classe considera a quantidade de informação proporcionada por um conjunto de variáveis: o índice pode ser interpretado como a quantidade de informação que uma variável proporciona sobre outra. Na figura 2 apresentamos a análise por agrupamentos organizadas em um dendrograma, construído pelo software *CHIC* a partir de dois critérios: a coesão máxima dentro de cada classe e o maior grau de implicação entre duas classes. Para determinar o valor máximo o software determina o índice de coesão e o índice de implicação associando as variáveis ou agrupamentos com outras variáveis ou agrupamentos, sempre dois a dois, destacando o maior valor observado.

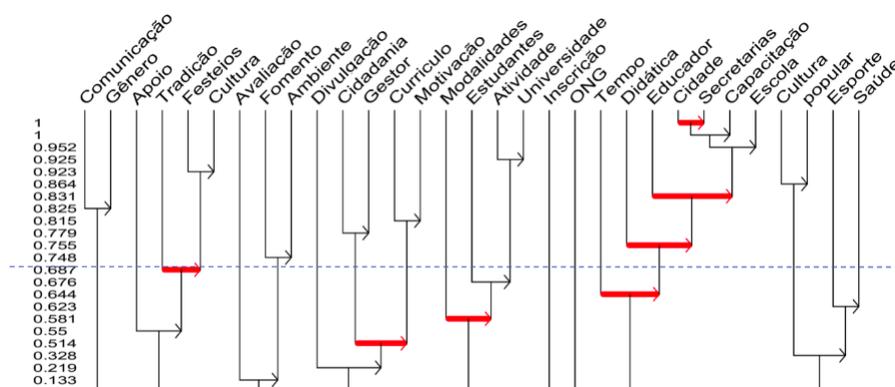


Figura 2. Árvore Coesitiva.

Observa-se que a positividade relacionada ao contexto educacional reúne os seguintes elementos da Classe A (Cidade, Secretarias, Capacitação, Escola e Educador),

Tabela 1. Itens associados na Classe A

Categories	Items
Cidade	Município
Secretarias	SMED, SJN-RG, SMTP, SMTTEL, MCRG, SMGA, SMCAS, CRAS
Capacitação	alfabetização, formação, orientação, qualificação, aperfeiçoamento
Escola	colégio, particulares, Turmas, aulas, vagas, rede, pública, gratuito.
Educador	professor, psicopedagoga, aposentadoria, arte-educador.

provavelmente porque a cidade, desde 1999, está inserida em um sistema de educação completo no qual a Secretaria de Município da Educação exerce as atribuições de órgão gestor e o Conselho Municipal de Educação de órgão normativo e fiscalizador. O Plano Municipal de Educação Municipal referente ao período entre 2015 - 2015 revela um quantitativo de 77 escolas de Educação Infantil, com 3.338 matrículas, envolvendo 540 docentes; os anos iniciais do Ensino Fundamental, possui 96 escolas, com 27.862 matrículas, envolvendo 1.616 docentes, já o Ensino Médio, apresenta 21 instituições, com 7.116 matrículas e 219 docentes. Além disso, a cidade possui uma universidade pública, faculdades particulares e uma rede de Ensino Médio Técnico composta por instituições estaduais e federais. O nó de nível 7 (que agrupa as variáveis citadas) apresentado como significativo, possui um nível de coesão de 0.831 (probabilidade de ocorrência da meta-regra). Pode-se inferir que as mensagens classificadas apresentam a problemática da educação como tema central da opinião registrada. A Tabela 1 apresenta os termos representativos associados as categorias que compõem a Classe A. Além da gestão o município possui 90,4% de seus da Educação Básica com formação em nível superior sendo 85,4% dos profissionais estão na rede privada e 91,6% na pública. Destes, 78,6% possuem licenciatura e a pós-graduação atinge 44,7% dos professores, sendo que dos pós-graduados 51,8% atuam na rede municipal, 39% na estadual e 31,8% na rede privada. É importante destacar que o maior percentual possui o nível de Especialização, representando 39,9% do total.

5. Conclusões

Descobrir a opinião pública é um fator importante para o processo de tomada de decisões, principalmente de forma a orientar o poder público em sua relação com a participação cidadã. Entretanto, à grande quantidade de conteúdo subjetivo existente em documentos textuais na Web torna complexa a atividade de recuperação e de interpretação desse tipo de informação, agravada quando é necessário extrair e classificar a subjetividade em diferentes fontes de dados. Desta forma contribuimos como uma abordagem que caracteriza uma cidade inteligente como aquela com capacidade de integrar sistemas de inteligências coletivas, visando o enfrentamento dos problemas locais; produzir valores públicos, com enfoque e elaboração na co-participação, como instrumento capaz de fornecer conhecimento e aprendizagens estruturantes para um projeto de comunidade. Como trabalhos futuros estão sendo desenvolvidos grafos implicativos que auxiliem a compreensão das relações estabelecidas.

Referências

- Afonso, R. A., Silva, W. D., Tomas, G., Gama, K., Oliveira, A., Alvaro, A., and Garcia, V. C. (2013). Br-scmm: Modelo brasileiro de maturidade para cidades inteligentes. *Simpósio Brasileiro De Sistemas De Informação*.

- Almeida, M. E. B. T. M. et al. (2000). O computador na escola: contextualizando a formação de professores- praticar e teoria, refletir a prática. *Pontifícia Universidade Católica de São Paulo*.
- Belloni, M. L. (2001). Educação a distância: coleção educação contemporânea.
- Couturier, R., Bodin, A., and GRAS, R. (2004). A classificação hierárquica implicativa e coesiva. *Manual Curso CHIC versão, 2*.
- Creswell, J. W. (2007). Procedimentos de métodos mistos. *Creswell, JW Projeto de Pesquisa: Métodos qualitativo, quantitativo e misto. 2ª ed. Porto Alegre: Artmed*, pages 211–227.
- Creswell, J. W. and Clark, V. L. P. (2015). *Pesquisa de Métodos Mistos-: Série Métodos de Pesquisa*. Penso Editora.
- Damasceno, M. (2015). Introdução a mineração de dados utilizando o weka. [Online; acessado setembro, 2015].
- de Queiroz, C., Coutinho, S., and da Silva Santos, A. (2016). Educação financeira para alunos de um curso de engenharia de produção financial education for students of a production engineering course. *BoEM-Boletim online de Educação Matemática*, 4(7):234–253.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37.
- Gomyde, A. (2017). Cidades inteligentes e humanas. *Boletim de Conjuntura*, 2:7–9.
- Gras, R. and Almouloud, S. A. (2002). A implicação estatística usada como ferramenta em um exemplo de análise de dados multidimensionais. *Educ Mat Pesqui*, 4(2):75–88.
- Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Hand, D. J. (2007). Principles of data mining. *Drug safety*, 30(7):621–622.
- Khatoun, R. and Zeadally, S. (2016). Smart cities: concepts, architectures, research opportunities. *Communications of the ACM*, 59(8):46–57.
- King, N., Cassell, C., and Symon, G. (1994). Qualitative methods in organizational research: A practical guide. *The Qualitative Research Interview*, 17.
- Lerman, S. and Manski, C. (1981). On the use of simulated frequencies to approximate choice probabilities. *Structural analysis of discrete data with econometric applications*, 10:305–319.
- Lévy, P. (2007). *Inteligência coletiva (A)*. Edições Loyola.
- Manville, C., Cochrane, G., Cave, J., Millard, J., Pederson, J. K., Thaarup, R. K., Liebe, A., Wissner, M., Massink, R., and Kotterink, B. (2014). Mapping smart cities in the eu. *European Union*.
- Martins, G. and Theóphilo, C. R. (2009). Metodologia da investigação científica. *São Paulo: Atlas*.
- McCue, C. (2014). *Data mining and predictive analysis: Intelligence gathering and crime analysis*. Butterworth-Heinemann.
- Systems, U. (2018). Urban systems. [Online; acessado abril, 2018].
- Wang, J. (2005). *Encyclopedia of data warehousing and mining*. IGI Global.
- Zabalza, M. A. (2004). O ensino universitário: seu cenário e seus protagonistas. porto alegre: Artmed. *Resolução CNE/CES n. 8*, page 16.
- Zemos (1998). Educación expandida. [Online; acessado abril, 2018].