

SmartCluster: Utilizando Dados Públicos para Agrupar Cidades Inteligentes por Domínios

Alternative Title: SmartCluster: Using Public Data to Group Smart Cities by Domains

Ricardo Alexandre Afonso^{1,2}
¹UFAL - Univ. Federal de Alagoas
Arapiraca, Alagoas, Brasil
ricardo.afonso@arapiraca.ufal.br

Clóvis Holanda do Nascimento²
Vinicius Cardoso Garcia²
²UFPE - Univ. Federal de Pernambuco
Recife, Pernambuco, Brasil
chn2@cin.ufpe.br
vcg@cin.ufpe.br

Alexandre Álvaro²³
³UFSCar - Univ. Federal de São Carlos
Sorocaba, São Paulo, Brasil
alexandre.alvaro@gmail.com

RESUMO

Atualmente existem vários domínios e indicadores ao redor do mundo que servem para categorizar Cidades Inteligentes, entretanto, não há estudos suficientes sobre a comparação destas cidades no Brasil. As bases de dados públicos dispõem de dados sobre os mais variados indicadores e domínios, e esses dados precisam ser normalizados e agrupados para permitir uma comparação entre as cidades inteligentes brasileiras. Realizar uma análise baseada no agrupamento de similaridade de indicadores de cidades inteligentes pode trazer aos gestores municipais uma maior compreensão das possibilidades estratégicas de otimização de recursos.

Palavras-chave

Cidades Inteligentes, Governança, Dados Públicos.

ABSTRACT

There are currently several domains and indicators around the world that serve to categorize Smart Cities, however, there are not enough studies on the comparison of these cities in Brazil. The public databases have data on various indicators and domains, and these data need to be standardized and clustered to allow a comparison between the smart cities. Make an analysis based on clustering similarity smart cities indicators can bring to the municipal managers a better understanding of the strategic possibilities of resource optimization.

Categories and Subject Descriptors

J.1 [Computer Applications]: Administrative Data Processing – Government

General Terms

Measurement, Documentation, Economics, Standardization.

Keywords

Smart Cities, Government, Public Data.

1. INTRODUÇÃO

Segundo Brousell [1] o conceito de Cidades Inteligentes (CI)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SBSI 2015, May 26–29, 2015, Goiânia, Goiás, Brazil.
Copyright SBC 2015

ainda continua sem uma definição clara. Enquanto o autor sugere o uso de cenários relacionados à transporte, tecnologia, infraestrutura, sustentabilidade e governança, outros autores como [2,3] se referem ao termo CI, utilizando diferentes nomenclaturas, contextos e significados. A forma como os recursos destas cidades são gerenciados e como são escolhidos estrategicamente os domínios a serem desenvolvidos impacta diretamente na vida de seus municípios. No Brasil, existem atualmente 5570 municípios dos quais vinte e sete representam as capitais de seus respectivos estados. Neste trabalho foram analisadas todas as capitais brasileiras pelos seus indicadores e comparadas entre si, na tentativa de identificar possíveis agrupamentos por semelhanças (*cluster*) de domínios. Fica, portanto definido neste trabalho que os autores utilizarão o termo Cidade Inteligente para definir aquelas cujos domínios e indicadores básicos constantes na Seção 3 deste trabalho sejam de alguma forma quantificados.

A principal motivação deste trabalho é apresentar uma metodologia de obtenção de dados públicos para a comparação e criação de agrupamentos de cidades por indicadores de desempenho, indicando assim, diferentes níveis e agrupamentos de CI. Para isso, a Seção 2 deste trabalho apresenta os conceitos e desafios sobre domínios e indicadores de CI. Na Seção 3 é demonstrada a metodologia para obtenção e agrupamento de dados. Na Seção 4 é realizada uma discussão dos resultados encontrados. A Seção 5 conclui este trabalho com as considerações finais.

2. GOVERNANÇA INTELIGENTE

O trabalho de Vaz [4] indica vários desafios para a governança eletrônica, entre eles, a falta de uma agenda de dados governamentais que deixe de ser uma iniciativa local e passe a se espelhar na agenda dos países desenvolvidos que já fazem bom uso de dados abertos.

Cohen [5] desenvolveu um conceito e o chamou de “A Roda das Cidades Inteligentes” contendo seis fatores de inteligência: Pessoas, Economia, Meio Ambiente, Governo, Vida e Mobilidade.

Chourabi [6] apresenta um framework dividido em oito eixos: gestão e organização, tecnologia, governança, contexto político, as pessoas e as comunidades, economia, infraestrutura construída, e o ambiente natural.

Mais recentemente, em 2013, a Norma ISO 37120 criada pela *International Standards Organization* (ISO) [8] fornece cem diferentes indicadores para Cidades Sustentáveis.

A próxima seção apresenta a metodologia utilizada por este trabalho para levantar os domínios e indicadores que possam auxiliar no entendimento desses dados disponíveis em bases de dados públicos.

3. METODOLOGIA

Para criar o agrupamento de cidades inteligentes por domínios, foram seguidos os passos de obtenção de dados, normalização, definição e por fim, o agrupamento. As subseções seguintes detalham essa metodologia.

3.1 Selecionando fontes de dados

Os indicadores para cada domínio obedecem a dois critérios simples: a disponibilidade de dados públicos para medição e a possibilidade de comparação desses dados locais com o mesmo conjunto de dados levantados em outros municípios ao redor do mundo. Com base no trabalho de levantamento de dados públicos e artigos científicos [26], foi criada a Tabela 1 com estes domínios.

Tabela 1. Fontes de dados públicos para Cidades Inteligentes

Domínios	Indicadores	Fontes de dados públicos
Água	Água encanada	[10,16,17]
Educação	IDH-Educação	[9,10,13]
Energia	Acesso à energia	[10,18]
Governança	IDH-Emprego	[19,20,21]
Moradia	Casa própria	[10,14,22]
Ambiente	Coleta de lixo	[10,23]
Saúde	IDH-Saúde	[10,14,15]
Segurança	Taxa Homicídios	[10,11,12]
Tecnologia	Computador/casa	[10]
Transporte	Transporte público	[24,25]

Após a definição dos dez domínios e seus indicadores, foi realizada uma busca pelas fontes de dados públicos para o processo de mineração destes dados. Esse processo de mineração de dados foi realizado de forma manual, e deverá futuramente fazer uso de ferramentas para automatizar a busca de dados. Embora estes dados sejam estatísticos e sejam divulgados em intervalos mínimos de dois anos, ainda assim, necessitam de maior velocidade na obtenção das fontes.

3.2 Normalizando dados

Os indicadores disponíveis para mensurar os domínios especificados para CI possuem diferentes metodologias de cálculos. Dessa forma, foi necessário normalizar os dados para que fosse possível compara-los dentro dos seus respectivos domínios. A normalização foi necessária para que os dados mantenham uma mesma ordem de grandeza. (Equação 1).

Equação 1. Cálculo da normalização de atributos numéricos.

$$X_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

Onde X_i representa cada um dos indicadores mensurados dentro do intervalo normalizado de ($x_i = 0$) até o valor máximo ($x_{\max} = 5$).

3.3 Definindo os Domínios e Indicadores

Dentre os dez domínios pré-definidos na seção anterior, este trabalho utilizou os indicadores minerados nas bases de dados públicas de (A) Educação, (B) Saúde e (C) Segurança.

3.3.1 (A) Educação – IDEB

O IDEB (Índice de Educação Básica) foi criado em 2007 e fornece dados sobre a qualidade incremental da educação básica nos municípios. Este indicador foi criado e é mantido pelo MEC (Ministério da Educação) [9], é medido a cada dois anos e possui a mesma métrica utilizada em municípios de outros países.

3.3.2 (B) Saúde – IDH-m

O IDH-m (Índice do Desenvolvimento Humano Municipal) é composto basicamente pela equação que envolve Educação, Longevidade e Renda. Este indicador é utilizado em todo o mundo para medir e comparar os avanços referentes ao desenvolvimento humano em cidades de diferentes portes. Para calcular o IDH-m é utilizada a renda per capita municipal disponibilizada pelas fontes de dados do IBGE [10].

3.3.3 (C) Segurança – Homicídios por 100hab.

O indicador de Segurança é determinado pelo cálculo do número de mortes por mil habitantes e recebe o nome de Risco de Homicídio. Para obter os dados desse indicador, foram consultadas as fontes de dados públicos de [10,11]. Segundo a OMS (Organização Mundial de Saúde) os estudos em Risco de Homicídio podem ser qualificados por faixa etária, gênero ou raça. [12].

3.4 Criando os agrupamentos

Além de normalizar os dados se faz necessário quantificar os diferentes indicadores para compor níveis ou camadas de inteligência. Diante deste desafio, a principal questão de pesquisa que este trabalho levanta é:

- **(QP1):** Existe uma forma de agrupar as cidades com base na composição heterogênea de seus indicadores?

Uma forma de encontrar a resposta pode ser utilizando-se um coeficiente de correlação ou a distância euclidiana simples entre séries temporais. Para isso, foram considerados os três domínios selecionados (Educação, Saúde e Segurança) e calculou-se a similaridade entre todas as combinações dos indicadores obtidos no conjunto de dados das capitais brasileiras. Esse processo de agrupamento por similaridades (*cluster*) deu origem a uma matriz de similaridade que permitiu compor um mapa de calor (*heatmap*) para visualizar o resultado do cálculo de agrupamento hierárquico de forma mais clara. (Figura 1)

Município	(A)	(B)	(C)	Média	Município	(A)	(B)	(C)	Média
Florianópolis	5,00	4,81	4,03	4,6	Rio Branco	4,64	4,15	3,45	4,1
São Paulo	5,00	4,63	4,13	4,6	Teresina	4,28	4,21	3,59	4,0
Brasília	5,00	4,64	3,33	4,3	Aracaju	4,55	4,37	3,06	4,0
Boa Vista	4,87	4,28	3,72	4,3	Belo Horizonte	4,83	4,61	2,53	4,0
Rio de Janeiro	5,00	4,63	3,22	4,3	Fortaleza	4,44	4,32	2,99	3,9
Campo Grande	4,92	4,48	3,39	4,3	São Luís	4,31	4,28	3,08	3,9
Goiânia	4,90	4,58	3,27	4,2	Salvador	4,57	4,43	2,54	3,8
Manaus	5,00	4,26	3,38	4,2	Porto Velho	4,87	4,20	2,44	3,8
Macapá	5,00	4,25	3,39	4,2	Palmas	4,73	4,40	2,22	3,8
Cuiabá	4,94	4,52	3,06	4,2	Joao Pessoa	4,36	4,31	2,17	3,6
Belém	4,74	4,43	3,29	4,2	Vitoria	4,88	4,71	1,23	3,6
Curitiba	5,00	4,71	2,73	4,1	Recife	4,46	4,38	0,63	3,2
Porto Alegre	5,00	4,76	2,64	4,1	Maceió	4,17	4,06	0,13	2,8
Natal	4,46	4,33	3,59	4,1	Média capitais	4,74	4,43	2,86	4,0

Figura 1. Mapa de calor (heatmap) dos domínios (A) Educação, (B) Saúde e (C) Segurança.

3.4.1 Distância euclidiana simples

A distância Euclidiana entre os dois pontos, a e b, com as dimensões de k é dada pela Equação 2. A distância Euclidiana possui por definição sempre valores superiores ou iguais à zero. Esses valores se aproximam de zero absoluto somente quando os pontos são idênticos e se tornam maiores quando se tornam muito distantes.

Equação 2. Distância Euclidiana Simples.

$$\sqrt{\sum_{j=1}^k (a_j - b_j)^2}$$

Este trabalho utilizou a apresentação dos resultados medidos por esta equação para a construção de dendrogramas de similaridade. Este dendrograma agrupa os indicadores de Educação, Saúde e Segurança das vinte e sete capitais brasileiras obtidas por média simples (Figura 2).



Figura 2. Dendrograma das Capitais Inteligentes

Quando agrupados por similaridades de indicadores, esses dendrogramas podem apresentar uma visão mais detalhada da relação que estas cidades possuem. Na seção seguinte são discutidas essas similaridades.

4. ANÁLISE DOS RESULTADOS

A técnica de mineração de dados públicos possibilita a obtenção de equações e gráficos comparativos e um melhor entendimento dos indicadores de inteligência escolhidos por cada município para apoiar a gestão de recursos públicos. É apresentado na Tabela 2 o agrupamento (*cluster*) por similaridade das capitais brasileiras, suas médias e o intervalo de desvio entre as amostras.

É possível identificar, por exemplo, através da similaridade entre os indicadores das cidades do *Cluster #1* (São Paulo e Florianópolis) que os resultados do domínio de Educação são iguais (5,0), enquanto que o domínio de Saúde apresenta a maior diferença entre indicadores (4,626 e 4,813 respectivamente). Ainda assim, a média obtida pelos três domínios isola estas duas cidades na classificação geral de CI.

No *Cluster #2* (Porto Alegre, Curitiba e Belo Horizonte) a Segurança se configura como maior desvio entre os três domínios mensurados. Por características regionais, os dois primeiros municípios localizados na região sul do país possuem uma tendência de maior controle da violência urbana em relação a outras regiões do país.

Embora o *Cluster #3* seja composto por oito capitais de três diferentes regiões do país (norte, centro oeste e sudeste) a diferença registrada entre as médias de domínios não é tão expressiva ($\pm 0,32$). Em uma classificação genérica, seria possível dividir essas capitais em dois agrupamentos menores, respeitando as suas características regionais (norte/centro oeste e sudeste). Entretanto, os valores obtidos, provam que os resultados obtidos com os domínios mensurados independem de tais características.

Os *Cluster #4* e *#5* apresentam a maior variação individual entre os domínios. Ainda assim, no processo de agrupamento por similaridade a média obtida com os domínios mensurados coloca essas capitais juntas. Em ambos os casos, se destaca a diferença obtida no domínio de Educação, uma vez que as capitais mais distantes da capital do país parecem apresentar indicadores menores em relação aos demais.

O *Cluster #6* assim como o primeiro, apresenta um desvio muito pequeno e é composto por três capitais nordestinas (São Luís, Aracaju e Fortaleza). Neste caso as características políticas que norteiam estas capitais são historicamente muito parecidas, o que reflete no resultado do levantamento de dados.

Por fim, o *Cluster #7* composto pelas capitais: Maceió, Recife e Vitória apresentam um cenário curioso. As duas primeiras capitais situam-se na região nordeste onde os indicadores relacionados a Saúde e Educação sistematicamente vem apresentando aumentos nos levantamentos do IBGE, entretanto, os indicadores de Segurança não apresentam o mesmo desempenho. Com isso a capital Vitória, se distancia das demais capitais do sudeste neste indicador, pois embora seus indicadores de Saúde e Educação estejam próximos das demais capitais do sudeste, seus números em relação à segurança são equiparáveis aos das capitais mais violentas do país.

Tabela 2. Clusters de Cidades por domínios

Cluster #1 – média 4,599 ± 0,04		
Florianópolis	São Paulo	
Cluster #2 – média 4,088 ± 0,74		
Porto Alegre	Curitiba	Belo Horizonte
Cluster #3 – média 4,232 ± 0,32		
Belém Manaus Goiânia	Campo Grande Rio de Janeiro Cuiabá	Macapá Brasília
Cluster #4 – média 3,767 ± 1,14		
João Pessoa Palmas	Salvador	Porto Velho
Cluster #5 – média 4,131 ± 1,26		
Natal Rio Branco	Teresina	Boa Vista
Cluster #6 – média 3,933 ± 0,27		
São Luís	Aracaju	Fortaleza
Cluster #7 – média 3,184 ± 1,67		
Maceió	Recife	Vitória

5. CONCLUSÃO

O objetivo principal deste trabalho foi responder a questão de pesquisa (QP1) estabelecida nesse trabalho sobre a possibilidade de agrupamento de cidades por similaridade de indicadores. Este objetivo foi alcançado pelos *clusters* que agrupam os indicadores e suas diferenças entre os domínios levantados.

Embora o objetivo deste trabalho tenha sido apresentar uma forma de transformar os dados públicos existentes em indicadores que apontem similaridades entre as capitais brasileiras, é importante que esses indicadores sirvam como base para a tomada de decisão por parte dos governantes municipais. Os dados de agrupamento regional podem e devem ser utilizados para compreender particularidades e potencialidades capazes de compor um *ranking* de CI nacional.

Este trabalho fez uso de bases de dados públicos, ferramentas de cálculo e geração de gráficos para as capitais brasileiras, mas como trabalho futuro, devem ser criados novos agrupamentos que possibilitem um catálogo mais amplo em tempo real e interativo com todas as cidades brasileiras.

6. AGRADECIMENTOS

Este trabalho foi parcialmente financiado pela FACEPE (Fundação de Apoio à Tecnologia e Ciência do Estado de Pernambuco), concessão IBPG-0499-1.03/11.

7. REFERENCIAS

- [1] Brousell, Lauren. Five Things You Need to Know About Smart Cities. 2012. Disponível em: <http://goo.gl/XMnAtD> acessado em 06 de março de 2015.
- [2] Hollands, R.G. (2008). Will the real smart city please stand up? *City*, 12(3), 303-320
- [3] Boulton, A., (2011). Cyberinfrastructures and “smart” world cities. Disponível em: <http://goo.gl/t6DGDv> acessado em 06 de março de 2015.
- [4] Vaz, José Carlos, Manuella Maia Ribeiro, e Ricardo Matheus. "Desafios para a Governança Eletrônica e Dados Governamentais Abertos em Governos Locais." WTRANS13-Workshop de Transparência em Sistemas. 2013.
- [5] Cohen, Boyd. What exactly is a Smart City? 2013. Disponível em: <http://goo.gl/YNuyvi>, acessado em 06 de março de 2015.
- [6] Chourabi, H.; Taewoo Nam; Walker, S.; Gil-Garcia, J.R.; Mellouli, S.; Nahon, K.; Pardo, T.A.; Scholl, H.J.; , "Understanding Smart Cities: An Integrative Framework," *System Science (HICSS)*, 2012. doi: <http://10.1109/HICSS.2012.615>
- [7] Afonso, R.A.; Silva, W. M. Da; Tomas, G.H. R. P.; Gama, K.; Oliveira, A.; Alvaro, A.; Garcia, V.C. Br-Scmm: Modelo Brasileiro De Maturidade Para Cidades Inteligentes. In: Simpósio Brasileiro De Sistemas De Informação, 2013, João Pessoa.
- [8] McCarney, P.L., (2013), “Why Cities? Why Metrics? – Tracking Cities in the 21st Century” Global Cities Institute Working Papers, University of Toronto.
- [9] IDEB. Índice da Educação Básica do Brasil. MEC (Ministério da Educação). 2015 Disponível em: <http://goo.gl/yv7FPS>, acessado em 06 de março de 2015.
- [10] IBGE, Instituto Brasileiro de Geografia e Estatística. 2015. Disponível em <http://www.ibge.gov.br/home> acessado em 06 de março de 2015.
- [11] Ministério da Justiça. Portal do Cidadão. 2015. Disponível em: <http://www.justica.gov.br/portaldpadrao> acessado em 06 de março de 2015.
- [12] United Nations Office on Drugs and Crime. 2015. Disponível em: <http://www.unodc.org> acessado em 06 de março de 2015.
- [13] Fundo Nacional de Desenvolvimento da Educação. Acesso a informação. 2015. Disponível em: <http://www.fnde.gov.br> acessado em 06 de março de 2015.
- [14] Portal ODM, Acompanhamento municipal dos objetivos do milênio. 2015 Disponível em: <http://www.portalodm.com.br> acessado em 06 de março de 2015.
- [15] Empresa de Tecnologia e Informações da Previdência Social. 2015. Disponível em: <http://portal.dataprev.gov.br> acessado em 06 de março de 2015.
- [16] Instituto TrataBrasil. Ranking do Saneamento 2015. Disponível em: <http://www.tratabrasil.org.br> acessado em 06 de março de 2015.
- [17] Organização das Nações Unidas para a educação, a ciência e a cultura. Website. 2015. Disponível em: <http://en.unesco.org> acessado em 06 de março de 2015.
- [18] Agencia Nacional de Energia Elétrica. 2015. Disponível em: <http://www.aneel.gov.br> acessado em 06 de março de 2015.
- [19] IPEA, Instituto de Pesquisa Econômica Aplicada. 2015. Disponível em: <http://www.ipea.gov.br> acessado em 06 de março de 2015.
- [20] Fundação Dom Cabral. Desenvolvimento de executivos e empresas. 2015. Disponível em: <http://www.fdc.org.br> acessado em 06 de março de 2015.
- [21] Serviço Brasileiro de Apoio as Micro e Pequenas Empresas. Portal Sebrae. 2015. Disponível em: <http://www.sebrae.com.br> acessado em 06 de março de 2015.
- [22] Fundação SEADE. Fundação do Sistema Estadual de Análise de Dados. 2015. Disponível em: <http://www.seade.gov.br> acessado em 06 de março de 2015.
- [23] MMA. Ministério do Meio Ambiente. 2015. Disponível em: www.mma.gov.br acessado em 06 de março de 2015.
- [24] MT. Ministério dos Transportes. 2015. Disponível em: <http://www.transportes.gov.br> acessado em 06 de março de 2015.
- [25] Agencia Nacional de Transportes Públicos. 2015. Disponível em: <http://www.antp.org.br> acessado em 06 de março de 2015.
- [26] Afonso, R. A. ; Brito, K. S. ; Nascimento, C. ; Alvaro, Alexandre ; Garcia, V. C. . Brazilian Smart Cities: Using a Maturity Model to Measure and Compare Inequality in Cities. In: International Conference on Digital Government Research, 2015, Phoenix, Arizona, USA. 16th Annual International Conference on Digital Government Research, 2015. DOI: <http://dx.doi.org/10.1145/2757401.2757426>