

COMP: uma métrica para avaliar a compatibilidade entre as informações fornecidas pelos *rankings* e aquelas percebidas pelos usuários

Rodrigo de Magalhães Cunha ^[1], Alexandre Linhares ^[2]

[1] rodrigo.cunha@ifsudestemg.edu.br. IF Sudeste MG campus Juiz de Fora/Departamento de Educação e Tecnologia (DET).

[2] alexandre.linhares@fgv.br. FGV/EBAPE

RESUMO

Este estudo propôs a criação de uma métrica – chamada *COMP* – para mensurar a relação entre as informações disponibilizadas pelos *rankings* e aquelas percebidas pelos usuários. Para isso, estudos de psicologia foram utilizados para dar sustentação teórica às propriedades definidas para a métrica. Matematicamente, foram utilizados métodos determinísticos e probabilísticos no cálculo, entre eles média ponderada, distância de *Mahalanobis*, o método de *clustering Affinity Propagation* e a medida de correlação *Tau de Kendall*. Os resultados mostraram-se válidos e confiáveis através da aplicação em *rankings* simulados. *Rankings* reais foram utilizados para exemplificar aplicações práticas da métrica. Entende-se que essa proposta de mensuração contribui para evidenciar um aspecto até então pouco considerado em contextos de *rankings*.

Palavras-chave: Informações dos *rankings*. Percepção dos usuários. Categorias top. Métrica. Cluster.

ABSTRACT

This study proposes a metric – named COMP – to measure the relationship between information available in rankings and those perceived by users. To do it, psychological studies provide theoretical support to definition of metric's properties. Mathematically, deterministic and probabilistic methods were used, among them weighted mean, Mahalanobis' distance, Affinity Propagation clustering method and Kendall's Tau correlation's coefficient. The results were valid and reliable through the application in simulated rankings. Real rankings were used to illustrate practical applications. It is understood that this measurement proposal contributes to evidence an aspect that until then was little considered in contexts of rankings.

Keywords: Information of rankings. Users' perceptions. Top categories. Metric. Cluster.

1 Introdução

Os *rankings* têm influenciado os usuários nas suas escolhas. A maioria dessas escolhas é enviesada para as entidades pertencentes às categorias *top* – formadas pelas entidades mais bem ranqueadas (*top-5*, *top-10*...) –, com exceção daqueles casos em que os usuários têm interesse num contexto particular. Dada à importância das categorias *top*, as entidades nelas inseridas – ou entidades *top* – têm sido analisadas em alguns estudos (BUSH *et al.*, 2011; CARARE, 2012; MARWELL, 2012; SEHGAL, 2010; SHARMA; GIVENS-SKEATON, 2010; WILSON, 2009; YANG; TAO, 2012) e têm sido usadas para suportar o desenvolvimento de outros (HORN; HENDEL; FRY, 2007; NOLEN, 2009).

Da psicologia, sabe-se que os usuários, geralmente, criam mentalmente as categorias *top* e percebem as entidades *top* como sendo mais similares entre si do que são em relação às outras entidades de um *ranking* (BRENNER; ROTTENSTREICH; SOOD, 1999; ISAAC; SCHINDLER, 2013; LECLERC; HSEE; NUNES, 2005). Entretanto, a base matemática e estatística utilizada na construção da maioria dos *rankings* – com indicadores agregados em índices –, não suporta integralmente essa ideia, conforme já identificado no contexto acadêmico (NAUDÉ; HENNEBERG; JIANG, 2009).

Complementarmente, sabe-se que os métodos de *clustering* são destinados a criar grupos – ou *clusters* – de entidades baseados nas similaridades dos seus indicadores (HAIR *et al.*, 1998; JAIN, 2010; NARDO *et al.*, 2005). Como os usuários, em geral, percebem as entidades *top* como similares, e estas entidades são representadas, nos *rankings*, pelos seus próprios indicadores, é possível que os métodos de *clustering* possam ser usados para analisar a compatibilidade entre as informações percebidas e as de fato fornecidas pelos *rankings*.

Na prática, pode-se notar que muitos dos *rankings* com maior divulgação na mídia têm usado *clusters* tanto para fornecer alguma classificação para as entidades ranqueadas quanto para embasar algumas análises qualitativas, como o *Economic Freedom of the World* 2012 (EFW) (FRASER INSTITUTE, 2012), o *Human Development Index* 2013 (HDI) (UNDP, 2013a, 2013b) e o *Index of Economic Freedom* 2013 (IEF) (MILLER; HOLMES; FEULNER, 2013). Entretanto, os *clusters* fornecidos por eles são normalmente baseados nos próprios índices ou em classificações

tradicionais – como aspectos geográficos ou de renda. Portanto, esses *clusters*, não necessariamente são equivalentes àqueles construídos a partir das similaridades entre os indicadores de um *ranking*.

Um estudo sugeriu que a representação de indicadores suportada por medidas de similaridade é uma promissora alternativa no contexto de *rankings* (HOYLAND; MOENE; WILLUMSEN, 2012). Nesse sentido, conforme destacado no referencial teórico, *clusters* construídos a partir de indicadores de *rankings* têm sido usados em inúmeros contextos.

Entretanto, apesar desses estudos, não existe nenhuma métrica proposta para mensurar o nível no qual um *ranking* representa as similaridades entre as entidades *top*. O objetivo desse artigo é, portanto, encarar esse *gap* e construir a métrica. Dada a intenção da métrica de avaliar a compatibilidade entre as informações fornecidas pelos *rankings* e aquelas percebidas pelos usuários, ela será chamada *COMP*, como uma alusão à “compatibilidade” entre as informações que serão confrontadas.

2 Referencial teórico

Para o embasamento teórico deste estudo, uma questão precisa ser respondida: como os usuários estão percebendo as informações disponibilizadas pelos *rankings*? Nesse sentido, estudos de psicologia das últimas décadas mostraram que as pessoas interpretam conjuntos de dados em pequeno número de categorias (KAUFMAN *et al.*, 1949; KUBOVY; WAGEMANS, 1995; LASKI; SIEGLER, 2007; MILLER, 1956; SHEPARD; KILPATRIC; CUNNINGHAM, 1975; VAN OEFFELEN; VOS, 1982) e que a existência de limites entre essas categorias influencia na percepção de que os itens de categorias distintas são diferentes (ALLEN; WILDER, 1979; LOCKSLEY; ORTIZ; HEPBURN, 1980; MADDOX *et al.*, 2008; MAKI, 1982; MISHRA; MISHRA, 2010; OTTEN, 2002; TVERSKY, 1992). Por outro lado, sabe-se que as pessoas percebem itens dentro de uma categoria como mais similares entre si do que itens entre categorias (BRENNER; ROTTENSTREICH; SOOD, 1999; LECLERC; HSEE; NUNES, 2005). Em espaços unidimensionais, como os *rankings*, uma das formas já identificadas para a criação mental desses limites é através de números arredondados, múltiplos de 5 ou 10 (COUPLAND, 2011; DAHAENE; MEHLER, 1992; DE LUSIGNAN *et al.*, 2004; HORNIK; CHERIAN; ZAKAY, 1994; JANSEN; POLLMANN, 2001; KAUFMAN *et al.*, 1949;

POPE; SIMONSOHN, 2011; TARRANT; MANFREDO, 1993).

Os estudos supracitados foram relacionados e o “efeito top-10” foi identificado nos *rankings* (ISAAC; SCHINDLER, 2013). Mostrou-se que os consumidores: mentalmente dividem os *rankings* num pequeno grupo de categorias; percebem diferenças entre entidades de categorias distintas; percebem similaridades entre entidades de uma mesma categoria; e geram essas categorias através de números arredondados (0 ou 5) – referentes às posições das entidades nos *rankings*. Ou seja, identificou-se que os consumidores, ou usuários dos *rankings*, criam mentalmente as categorias *top*.

Esses estudos oferecem resultados compatíveis com os encontrados por Hoyland, Moene e Willumsen (2012) os quais sugerem, após a análise de alguns *rankings* de países, que seria “melhor alocar países em grupos com performances similares” com base no “critério de que a diferença entre quaisquer dois membros desses grupos não seja muito grande”. Embora os autores não tenham embasado seu argumento na percepção dos usuários, mas na incerteza inerente aos índices, este corrobora o anterior e implicitamente direciona que estudos de *rankings* poderiam ser complementados através de análises de *clusters*.

Sucintamente, a análise de *clusters* é o estudo dos métodos de agrupamento – ou *clustering* – de objetos baseados nas similaridades (JAIN, 2010). Esta metodologia tenta minimizar a variância dentro dos grupos e maximizá-la entre eles (HAIR *et al.*, 1998) gerando *clusters* com objetos mais similares entre si do que quando comparados com objetos de outros *clusters* (NARDO *et al.*, 2005). Ou seja, são gerados grupos heterogêneos com elementos homogêneos (HAIR *et al.*, 1998).

Deve-se considerar que vários estudos têm proposto relacionar *clusters* e *rankings*, indicando o interesse de alguns pesquisadores nessa relação. Por exemplo, *clusters* já foram usados: como parte de métodos propostos para ranquear entidades (ÖNSEL *et al.*, 2008; RAD; NADERI; SOLTANI, 2011; VALADKHANI; VILLE, 2010); para agrupar indicadores de índices – não utilizados em *rankings* – em estudos sobre regimes de bem-estar (ABU SHARKH; GOUGH, 2010), desenvolvimento socioeconômico (FERNANDES, 2013), pesquisas (TAMMI, 2009; VALADKHANI; VILLE, 2010), competitividade de nações (ÖNSEL *et al.*, 2008), etc; e para agrupar indicadores de *rankings*

(NAUDÉ; HENNEBERG; JIANG, 2009; RENDE; DONDURAN, 2011). Como destaque, Naudé, Henneberg e Jiang (2009), analisando o FT1, mostraram que aspectos diferentes – representados pelos indicadores – levaram as universidades às primeiras posições no *ranking*. Na verdade, embora cada uma dessas ferramentas – *clusters* e *rankings* – tenha seu potencial de aplicação, o uso delas em conjunto, foi reconhecido como uma boa prática (CAVUSGIL; KIYAK; YENIYURT, 2004). Ou seja, elas foram entendidas como complementares.

Entretanto, apesar dos *clusters* terem sido propostos para representar similaridades e tenham potencial de aplicação em *rankings*, *clusters* “tradicionais” também têm sido explorados. Estes, aqui chamados de tradicionais, são aqueles construídos a partir de critérios pré-estabelecidos, e comumente utilizados, como localização geográfica, produto interno bruto (PIB) ou renda. Para ilustrar: *clusters* baseados em aspectos geográficos foram aplicados para comparar os *rankings* ARWU 2 e THE 3 (JÖNS; HOYLER, 2013); e as entidades responsáveis por alguns *rankings* têm explorado, nos relatórios nos quais os *rankings* são divulgados, análises qualitativas baseadas nesses tipos de *clusters*, como o EFW, o HDI e o IEF.

Entretanto, não pode-se afirmar que esses *clusters* “tradicionais” sejam capazes de agrupar conjuntamente entidades com altos níveis de similaridade entre seus indicadores. Esse aspecto foi encontrado: num estudo que explorou indicadores de água e saneamento a partir do agrupamento baseado em aspectos geográficos e de renda com critérios estabelecidos pelo Banco Mundial (ONDA *et al.*, 2014); e num estudo que questionou a utilização do PIB para agrupar países (ÖNSEL *et al.*, 2008).

Adicionalmente, outras formas de criação de *clusters* amplamente usadas em *rankings* são aquelas suportadas pelo índice. Isso sugere que os responsáveis pelos *rankings* têm se preocupado com a atribuição das entidades em grupos de acordo com suas performances, ou ordens nos *rankings*. Esses grupos são normalmente gerados através de dois métodos: comparando o índice com valores pré-

1 *Financial Times top-100 full-time MBA programmes* de 2008

2 *Academic Ranking of World Universities* (CWCU, 2012)

3 *Times Higher Education World University Rankings* (THOMSON REUTERS, 2013)

–definidos para encontrar o grupo de cada entidade, como no IEF; ou distribuindo as entidades em grupos de mesmo tamanho, como no HDI.

Entretanto, como já exposto, sabe-se que as entidades *top* não são, necessariamente, mais similares entre si do que entre elas e as outras (NAUDÉ; HENNEBERG; JIANG, 2009; RENDE; DONDURAN, 2011). Desse modo, é pertinente enfatizar que os grupos gerados dessas formas estão agrupando entidades de uma forma não totalmente alinhada com a percepção dos usuários.

Nesse contexto, para a construção da métrica, dado o alinhamento dos métodos de *clustering* com a percepção dos usuários, um desses métodos foi usado para criar agrupamentos tendo como base os indicadores – e não os critérios tradicionais nem os índices comumente explorados. Entende-se, dessa forma, que o nível no qual cada *ranking* representa as similaridades entre as entidades de acordo com a percepção dos usuários pôde ser avaliado.

3 Método da pesquisa

3.1 A escolha do método de *clustering*

Existem vários métodos de *clustering*. No geral, eles podem ser distinguidos entre hierárquicos e *partitionings* (JAIN, 2010; JAIN; MURTY; FLYNN, 1999). Em síntese, enquanto os primeiros recursivamente encontram *clusters* aninhados os últimos encontram simultaneamente todos os *clusters* sem impor nenhuma estrutura hierárquica (JAIN, 2010). No geral, esses tipos de métodos também podem ser diferenciados em função dos resultados: enquanto os *partitionings* fornecem *clusters* válidos a cada interação, os hierárquicos os fornecem somente ao término do processamento (BANDYOPADHYAY; SAHA, 2013). Apesar das inúmeras possibilidades, os mais usados, conforme Valadkhani e Ville (2010), são o *partitioning K-means* (MCQUEEN, 1967) e o hierárquico *Ward* (WARD, 1963). Como referência, alguns estudos sobre *rankings* e/ou indicadores já exploraram *K-means* (RAD; NADERI; SOLTANI, 2011; TAMMI, 2009), *Ward's* (ONDA *et al.*, 2014; ÖNSEL *et al.*, 2008; VALADKHANI; VILLE, 2010), métodos hierárquicos não declarados (FERNANDES, 2013; NAUDÉ; HENNEBERG; JIANG, 2009; RODE; COLL, 2011), e uma mescla de métodos hierárquicos não declarados e *K-means* (ABU SHARKH; GOUGH, 2010; MAVRI, 2013).

Apesar da “soberania” desses métodos, um estudo recente propôs um método alternativo de *clustering* chamado *Affinity Propagation* (AP) (FREY; DUECK, 2007) que considera cada objeto – ou entidade – como um nó de uma rede. Ele recursivamente transmite mensagens pelas arestas da rede entre os nós até os exemplares – ou representantes – e seus *clusters* possam emergir.

O AP possui características a serem destacadas. Primeiro, ao considerar todos os pontos como exemplares potenciais dos *clusters* que serão criados, o maior problema do método *K-means* foi resolvido: o fato da solução ser dependente da seleção inicial e aleatória dos exemplares dos *clusters* (BANDYOPADHYAY; SAHA, 2013; JAIN, 2010; JAIN; MURTY; FLYNN, 1999). Segundo, o AP solucionou um grande problema dos métodos hierárquicos: aquele no qual dois pontos não similares podem ser agrupados em conjunto por uma infeliz sequência de comparações pareadas em função desses métodos não requerem um centro simples de referência. Dessa forma, o AP encontra *clusters* com muito menos erros e numa quantidade de tempo menor do que um centésimo do tempo gasto pelos outros métodos (FREY; DUECK, 2007).

Além disso, o AP já foi usado com bons resultados em áreas como biologia (APELTSIN *et al.*, 2011; HONG-WEI, 2011; KIDDLE *et al.*, 2010; LEONE; SUMEDHA; WEIGT, 2007) biomecânica (XIAO; GUO, 2009; ZHANG *et al.*, 2011), biblioteconomia (CHEN, 2008), redes sociais (PITSILIS; ZHANG; WANG, 2011), entre outras. Portanto, embora não existam aplicações do AP no contexto de *rankings*, dado o seu potencial e a abrangência das aplicações já conhecidas, ele foi selecionado como método de *clustering*.

3.2 A escolha da medida de similaridade

Alguns métodos de *clustering* lidam com matrizes de similaridades ao invés dos dados brutos (JAIN; MURTY; FLYNN, 1999), como o AP. Para isso, medidas de similaridade, em geral relacionadas ao cálculo de distâncias, têm sido usadas. Nelas, distâncias pequenas equivalem a grandes similaridades (NARDO *et al.*, 2005).

A importância da escolha da medida de similaridade para a obtenção de *clusters* tem recebido destaque (BANDYOPADHYAY; SAHA, 2013; MILLIGAN; MARTHA, 1987). Sabe-se que essa escolha deve ser realizada em função da natureza dos dados que servirão de base para a construção dos *clusters*

(BANDYOPADHYAY; SAHA, 2013; LEONE; SUMEDHA; WEIGT, 2007).

Em muitos estudos sobre *rankings* e índices, apesar da sua importância, essa medida não foi reportada (ABU SHARKH; GOUGH, 2010; FERNANDES, 2013; MAVRI, 2013). Em outros, a Distância Euclidiana (DE), ou alguma variação da mesma, foi utilizada (NAUDÉ; HENNEBERG; JIANG, 2009; ÖNSEL *et al.*, 2008; RAD; NADERI; SOLTANI, 2011; RENDE; DONDURAN, 2011; RODE; COLL, 2011; TAMMI, 2009). De fato, a medida de distância para dados contínuos – aqueles mais comuns em *rankings* – com maior aplicação na criação de *clusters* é a DE (JAIN; MURTY; FLYNN, 1999).

Um problema da DE é a sua sensibilidade às escalas (JAIN; MURTY; FLYNN, 1999; NARDO *et al.*, 2005). Este problema deve ser naturalmente resolvido nos *rankings* que normalizam os indicadores para não distorcer o índice. Entretanto, a DE ignora a estrutura dos dados (WANG; BENSMAIL, 2013). Como os indicadores normalmente estão correlacionados, este é um aspecto a ser considerado. Para ilustrar, as correlações de *Pearson* entre pares de indicadores variam entre: 0,549 e 0,816 no HDI; 0,505 a 0,932 no FSI⁴; 0,414 a 0,822 no RGI⁵. Nota-se que o nível dessas correlações não pode ser desprezado.

Dessa forma, a DE, embora muito usada, parece não ser a medida de similaridade mais apropriada aos estudos de *rankings*. Neste cenário, a DM (MAHALANOBIS, 1936) é uma alternativa. Esta medida considera a correlação dos dados e é igual à DE nos casos em que a correlação é zero (JAIN; MURTY; FLYNN, 1999; MAESSCHALCK; JOUAN-RIMBAUD; MASSART, 2000; MCLACHLAN, 1999). Matematicamente, dado um *ranking* e considerando duas das suas entidades $a (a_1, a_2, \dots, a_n) \in R^n$ e $b (b_1, b_2, \dots, b_n) \in R^n$, a DM entre elas é definida pela Equação 1:

$$DM_S(a, b) = \sqrt{(a - b)^t S^{-1} (a - b)} \quad (1)$$

Onde: n é o número de indicadores do *ranking*; t denota uma matriz transposta; S^{-1} é a matriz de covariância inversa ($n \times n$); $S \in R^{n \times n}$; e S é definida positiva.

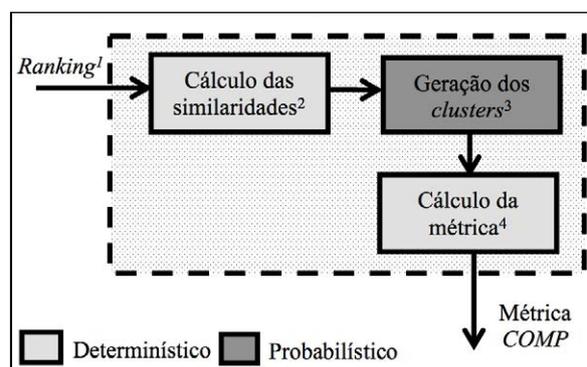
Embora a DM seja útil quando existe correlação, a computação da matriz variância-covariância pode causar problemas quando os dados têm um grande número de indicadores e/ou o número de entidades não é maior que o de indicadores (MAESSCHALCK; JOUAN-RIMBAUD; MASSART, 2000). Entretanto, essas condições não devem ser encontradas nos *rankings* que, em geral, têm poucos indicadores diretamente relacionados com o cálculo do índice – normalmente menos de 20 – bem como o número de entidades maior do que o de indicadores.

Além disso, uma desvantagem da DM em relação à DE é o maior tempo de cálculo, principalmente em função do processamento da matriz de covariância (BANDYOPADHYAY; SAHA, 2013). Entretanto, como o número de indicadores é normalmente menor do que 20 (gerando uma matriz de covariância com, no máximo, 400 células) e o AP é um método mais rápido do que os tradicionais (FREY; DUECK, 2007), este aspecto é minimizado. Pelo exposto, a DM foi escolhida.

3.3 Framework para a construção da métrica

O processo de construção da métrica foi constituído por procedimentos determinísticos e probabilísticos e está exposto na Figura 1.

Figura 1 – Framework para construção da métrica



Notas: 1. matriz de indicadores e entidades dos rankings; 2. através da DM de modo determinístico; 3. através do método AP de modo probabilístico; 4. através das Equações 4 e 5.

Fonte: elaborada pelos autores.

Destaca-se que, por ser probabilístico, heurísticas são usadas no processo de criação de *clusters* no AP (PSIG, 2014). Isso significa que, dada uma matriz de similaridades e o número de *clusters* que deverão ser criados, é possível que o AP não encontre a mesma solução em duas execuções distintas.

4 Failed States Index 2012 (TFFF, 2012; 2014a; 2014b)

5 Resource Governance Index 2013 (RWI, 2013)

3.4 Algumas características exploradas na construção da métrica

A métrica proposta pode ser entendida como uma medida de dispersão, isto é, uma medida capaz de representar a dispersão das entidades *top* pelos *clusters*. Dessa forma, duas situações extremas podem ser identificadas: cada entidade *top* pode ser atribuída a um *cluster* diferente; e todas as entidades *top* podem ser atribuídas ao mesmo *cluster*. Pode-se perceber que, se a primeira situação acontecesse, as percepções dos usuários estariam totalmente equivocadas. Por sua vez, no caso da segunda, os usuários estariam percebendo os dados exatamente da forma como eles são. Portanto, a faixa de valores da métrica deve ser limitada por essas duas situações. Por conveniência, arbitrou-se o valor zero (0) para a primeira situação e o valor um (1) para a segunda. Ou seja, a métrica varia de 0 a 1.

Entretanto, o número de *clusters* nos quais as entidades *top* estão distribuídas não é o único parâmetro a ser considerado. Pode-se perceber que várias soluções podem ser encontradas mesmo estando as entidades *top* distribuídas em apenas dois *clusters*, por exemplo: 9 entidades no primeiro *cluster* e apenas uma no segundo; 5 entidades em cada *cluster*; entre outras. Nesse contexto, apesar das entidades *top* estarem distribuídas em apenas dois *clusters*, o impacto dessas dispersões sobre a percepção dos usuários é diferente. No primeiro caso, se um par de entidades *top* fosse escolhido ao acaso, a probabilidade dessas duas entidades pertencerem ao mesmo *cluster* seria de 80% enquanto que, no segundo, de 50%. Portanto, a métrica distingue esses tipos diferentes de solução.

Adicionalmente, considerando que categorias *top* diferentes podem existir em cada *ranking*, é possível que soluções diferentes sejam encontradas para cada uma delas. Por exemplo, imagine que as 12 primeiras entidades estão atribuídas ao primeiro *cluster* e que as entidades de 13 a 16 ao segundo. Nesse caso, o *ranking* estaria aderente à percepção dos usuários nos contextos das categorias *top-5* e *top-10*, mas não no da *top-15*. A métrica foi proposta para considerar essa questão.

Sabe-se, também, que os usuários veem altos níveis de similaridade entre as entidades *top* independente dos índices de cada uma delas, tendo como base as suas posições no *ranking* (ISAAC; SCHINDLER, 2013). Portanto, justifica-se que os índices não devam ser considerados na construção da métrica,

apenas os indicadores. Como o uso de dados sintéticos – como os índices – é considerado uma prática pobre nos métodos de *clustering* (KETTENRING, 2006), a decisão de utilizar apenas os indicadores parece ainda mais oportuna.

Avançando, como os usuários estão baseando suas escolhas apoiados por vários *rankings* com diferentes características, a métrica deve ser universal. Ou seja, ela deve ser aplicável a todos os *rankings* para que possa ser comparável, oferecendo um valor relativo. Para isso, os dois cenários extremos já descritos precisam ser identificados e considerados para cada *ranking*.

Além disso, a métrica deve considerar que mais de uma entidade pode estar na mesma posição de um *ranking*. Nesses casos, considerando a percepção dos usuários, é esperado que essas entidades estejam atribuídas ao mesmo *cluster*. Entende-se, que, caso a solução do AP não tenha esse padrão, alguma penalização deva ser aplicada ao valor da métrica.

Além desses aspectos, o maior desafio é determinar o número de *clusters* a ser considerado. Como mais *clusters* tendem a representar as entidades *top* de um modo mais disperso, a métrica é sensível ao número de *clusters*. Ou seja, para ser comparável, a solução deve considerar o mesmo número de *clusters* para todos os *rankings*.

Em função da larga subjetividade envolvida, considera-se a identificação do número de *clusters* o maior problema na análise de *clusters* (NARDO *et al.*, 2005). De fato, estudos já mostraram essa dificuldade (DUBES, 1987; ÖNSEL *et al.*, 2008) e enfatizaram a arbitrariedade dessa seleção (RODE; COLL, 2011). Neste estudo, dado que a métrica é orientada aos usuários, a capacidade deles para processar informações (KAUFMAN *et al.*, 1949; MILLER, 1956) pode ser vista como uma alternativa. Dessa forma, com base nesses estudos, 6 *clusters* foram considerados.

Entretanto, o número de entidades e categorias *top* influencia na métrica. Para ilustrar, considere dois *rankings*: A com 100; e B com 180 entidades. Caso fossem gerados os 6 *clusters*, cada um deles teria: em A, 16,67 entidades em média; e, em B, 30. Isso significa que B, apenas por ter mais entidades, teria maior probabilidade do que A de atribuir todas as entidades *top* no mesmo *cluster*. Destaca-se que, embora o número de entidades por *cluster* possa variar entre eles, pode-se considerar a média como um *baseline* para análise. Dessa forma, construiu-se a métrica compensando esse aspecto e permitindo

que ela seja comparável entre *rankings* sem vieses gerados pelo método escolhido.

Pode-se entender, considerando os aspectos destacados, que a métrica representa quão bem os *rankings* disponibilizam informações de acordo com a percepção dos usuários. Para os cálculos, o *Python* foi selecionado como plataforma de *software*.

4 Resultados da pesquisa

4.1 A métrica COMP

A métrica *COMP* foi definida, inicialmente, pela Equação 2 da seguinte forma:

$$COMP'_k = \sum_{i=1}^{nt} \frac{\left(\frac{1}{i}\right) \cdot Ptop_i}{n^{(1-Ptop_i)} \cdot \sum_{j=1}^{nt} \frac{1}{j}}, \text{ com } n > 5 \quad (2)$$

Onde: k representa o *ranking*; $COMP'_k$ a métrica não normalizada de k ; i uma categoria *top* de k ; n o número de entidades de k ; nt o número de categorias *top* de k ; e $Ptop_i$ a probabilidade de duas entidades de i , selecionadas ao acaso, pertencerem ao mesmo *cluster*.

Para a interpretação de (2), deve-se notar que as parcelas descritas para o somatório mais externo representam as parcelas de uma média aritmética ponderada. Destaca-se, também, os seguintes aspectos. Primeiro, como as parcelas associam-se a i , cada uma delas representa a contribuição de uma categoria *top* diferente na composição da métrica.

Segundo, nt é uma função de n . No caso, utilizou-se o maior número inteiro menor que $\frac{n}{5}$ (função piso de $\frac{n}{5}$). O número 5 foi utilizado, pois representa uma forma de captar o número máximo de categorias *top* que poderiam ser criadas mentalmente pelos usuários, as quais, conforme já mencionado, são delimitadas pelas entidades cujas posições terminam em 0 ou 5. Exemplificando, um *ranking* com 21 entidades ($nt=4$) teria 4 categorias *top* consideradas para cálculo da métrica: *top-5* ($i=1$), *top-10* ($i=2$), *top-15* ($i=3$) e *top-20* ($i=4$).

Terceiro, cada categoria *top* i tem a sua própria probabilidade $Ptop_i$, a qual é dependente dos *clusters*. Ela representa a essência da métrica, pois capta o conceito associado à percepção dos usuários de que duas entidades *top* são mais similares entre si do que entre elas e as outras.

Quarto, $\frac{1}{i}$ representa o peso que a probabilidade de cada categoria *top* $Ptop_i$, assume na média aritmética ponderada. Uma função hipérbole foi selecionada para privilegiar as primeiras categorias *top* – normalmente utilizadas pelos usuários nos processos de tomada de decisão – e minimizar as diferenças dos pesos entre as últimas – às quais, em geral, são menos importantes aos usuários. Ou seja, isso significa que a queda de importância da *top-5* para a *top-10* é maior do que a da *top-10* para a *top-15* e assim por diante.

Quinto, $\sum_{j=1}^{nt} \frac{1}{j}$ representa o somatório dos pesos para todas as categorias *top*. Esse elemento faz parte do cálculo da média aritmética ponderada e é fixo para todas as parcelas do somatório.

Sexto, $\frac{1}{n^{(1-Ptop_i)}}$ representa uma penalização em função do número de entidades n do *ranking*. Dado que 6 *clusters* serão gerados, como já sinalizado, *rankings* com maior n têm maior probabilidade de posicionar as entidades *top* no mesmo *cluster*. Isso não significa que eles estejam mais alinhados com a visão dos usuários. Esse “falso” alinhamento pode estar sendo obtido por uma imposição do método – ou seja, pela própria escolha dos 6 *clusters*. Para exemplificar a importância desse elemento, considere dois *rankings*: A com 30 e B com 200 entidades. Imagine que, quando uma determinada categoria *top* é analisada, a probabilidade $Ptop_i$ seja igual para eles. Nesse caso, o elemento vai garantir que A tenha uma parcela do somatório maior do que B, minimizando o viés imposto pelo método e permitindo que os *rankings* possam ser comparados.

Sétimo, para que a métrica gerasse resultados entre 0 e 1, foi necessário normalizá-la. Pode-se perceber que, sem o processo de normalização, os limites estabelecidos teoricamente para a métrica seriam utópicos. Ou seja, pelo próprio método, nenhum *ranking* conseguiria atingi-los.

Para realizar a normalização, utilizou-se os valores mínimo e máximo possíveis para cada *ranking*. Esses valores dependem do número de entidades e estão associados às duas situações extremas já explicadas. Nesse contexto, o cálculo dos valores máximo e mínimo aproximado para uma dada categoria *top* i de um *ranking* com n entidades estão escritos, respectivamente, nas Equações 3 e 4:

$$P_{top_MAX_{n,i}} \begin{cases} 1, & \text{se } i \neq nt \\ \frac{n^2 - 11n + 30}{n^2 - n} & \text{se } i = nt \end{cases} \quad (3)$$

$$P_{top_MIN_{n,i}} \begin{cases} 0, & \text{se } i = 1; \\ 0,089, & \text{se } i = 2 \\ 0,0238 * \ln ni + 0,0555, & \text{se } i > 2 \end{cases} \quad (4)$$

Onde: n representa o número de entidades do ranking; i uma categoria *top* de rankings com n entidades; nt – na Equação 3 – o número de categorias *top* de rankings com n entidades; e ni – na Equação 4 – o número de elementos de i .

Com as probabilidades calculadas nas Equações 3 e 4, a substituição desses valores na Equação 2 fornece, respectivamente, os valores máximo MAX_n e mínimo aproximado MIN_n da métrica para rankings com n entidades. Como referência, deve-se considerar que a Equação 4 pode apresentar pequenos resíduos para $i > 2$ – próximos a $\pm 0,001$ em média para cada i . Apesar disso, optou-se por apresentar essa fórmula para fins de simplificação. Neste estudo, entretanto, para determinação do valor mínimo, o cálculo exato foi utilizado.

Dessa forma, de posse dos valores mínimo e máximo possíveis, a métrica inicialmente obtida foi normalizada com base na Equação 5:

$$COMP_k = \frac{(COMP'_k - MIN_n)}{(MAX_n - MIN_n)} \quad (5)$$

Onde: k representa o ranking; $COMP_k$ a métrica normalizada de k ; $COMP'_k$ a métrica não normalizada de k ; n o número de entidades de k ; e MAX_n e MIN_n , respectivamente, os valores mínimo e máximo possíveis para a métrica para rankings com n entidades.

4.2 Como os resultados da métrica foram validados?

A validade relaciona-se com a confiança nos resultados de um teste dado seu objetivo (COOK; BECKMAN, 2006). Neste estudo, como o propósito da métrica é mensurar quão bem os rankings estão representando os dados de acordo com a percepção dos usuários, entende-se que este deve ser o aspecto principal da validação. Para isso, rankings simulados foram utilizados.

Para a construção deles alguns parâmetros relacionados às características dos rankings foram utilizados e atribuídos aleatoriamente. São eles: número de entidades (de 20 a 200); número de indicadores (de 2 a 15); e método de agregação (soma e médias aritmética, geométrica e ponderada). Destaca-se que, em todos os rankings, o valor de cada indicador para cada entidade foi aleatório assim como o peso de cada indicador no caso de rankings baseados em médias ponderadas. Além disso, apesar do método de normalização poder variar (por exemplo, de 0 a 1, de 0 a 100), ele não foi considerado como um parâmetro, pois a DM não é uma medida sensível à escala.

O último parâmetro foi o nível de correlação entre os indicadores dos rankings. Considerando a relação entre esses níveis e as similaridades entre as entidades (NARDO *et al.*, 2005), esse parâmetro foi manipulado. Para isso, como os coeficientes de correlação (CC) são medidas contínuas, para fins de interpretação, oito faixas foram utilizadas: $CC < 0,3$; $0,3 \leq CC < 0,4$; $0,4 \leq CC < 0,5$; $0,5 \leq CC < 0,6$; $0,6 \leq CC < 0,7$; $0,7 \leq CC < 0,8$; $0,8 \leq CC < 0,9$; e $CC \geq 0,9$. Essas faixas foram aplicadas a todos os possíveis pares de indicadores de cada ranking. Gerou-se 500 rankings por faixa.

Destaca-se, ainda, que o CC Kendall's Tau (τ) foi escolhido, pois, em relação a outros: independe da premissa da normalidade, ao contrário do CC de Pearson's (r) (FIELD; 2009); e possui menores requerimentos de tamanho da amostra para a mesma precisão (BONETT; WRIGHT, 2000). Essas características são importantes, pois: para garantir a aleatoriedade, não se restringiu a normalidade; e alguns rankings foram gerados com pequeno número de entidades. Entende-se que esses parâmetros forneceram uma ampla e comparável amostra de rankings.

Em seguida, submeteu-se cada ranking ao cálculo da métrica. Desses, a métrica foi calculada para 3.930, ou seja, em 70 rankings (1,75%) o método AP não convergiu.

Em sequência, um modelo de regressão linear com correção de heteroscedasticidade, detalhado na Equação 6, foi desenvolvido para validação dos resultados. Para isso, utilizou-se os 3.930 rankings para os quais a métrica pôde ser calculada.

$$COMP_k = \beta_0 + \beta_1 C_k + \beta_2 E_k + \epsilon_k \quad (6)$$

Onde: k representa o *ranking*; $COMP_k$ a métrica de k ; C_k a média das correlações entre os pares de indicadores de k ; E_k o número de entidades de k ; ϵk o erro do modelo para k ; β_0 o intercepto; β_1 o coeficiente de C_k ; e β_2 o coeficiente de E_k . A Tabela 1 expõe os resultados.

Tabela 1 – Coeficientes do modelo de regressão para validação da métrica

Variável	Coefficiente	Valor p
β_0	- 0,039	0,000*
β_1	0,372	0,000*
β_2	0,000	0,442

Notas: $R^2 = 0,195$; *significantes ao nível 0,01; $n = 3930$.

Fonte: elaborada pelos autores.

Pelo modelo, a moderação entre C_k e E_k foi desprezada. Deve-se considerar que não existe base teórica para suportar que este relacionamento poderia fornecer diferentes níveis de similaridade e, portanto, diferentes valores para a métrica.

Como as correlações influenciam as similaridades de modo diretamente proporcional (NARDO *et al.*, 2005) e a métrica é dependente das medidas de similaridade, era esperado $\beta_1 > 0$. O valor 0,371 é compatível e sugere que, quanto maior a média das correlações, maior a métrica.

Entretanto, dado que a métrica não deveria ser enviesada pelo número de entidades, esperava-se ter $\beta_2 \approx 0$, como obtido. Destaca-se que, embora esse resultado fosse esperado, E_k foi usada no modelo para que a validade pudesse ser verificada.

Por fim, entende-se que essas evidências sugerem que a métrica é válida, ou seja, que ela reflete o quanto os *rankings* estão representando os dados de acordo com a percepção dos usuários.

4.3 Como a confiabilidade dos resultados da métrica foi avaliada?

Numa definição simples, a confiabilidade associa-se à reprodutibilidade de dados ao longo das ocasiões (DOWNING, 2004) podendo ser mensurada de várias formas (COOK; BECKMAN, 2006), sendo que a importância de cada uma delas é dependente de cada contexto (DOWNING, 2004). Para checar a reprodutibilidade dos resultados – ou seja, a estabi-

lidade dos *clusters* que serão gerados pelo AP – um procedimento similar ao aplicado por Tegarden, Tegarden e Sheetz (2007) foi realizado neste estudo. Dessa forma, cinco *rankings* de cada nível de correlação, dentre aqueles cuja métrica foi obtida, foram selecionados aleatoriamente. Cada um foi submetido ao cálculo da métrica 1.000 vezes.

Mesmo considerando que o processo de criação da métrica tem um aspecto probabilístico, sabe-se que o AP sempre converge quando existe uma solução “estável” (PSIG, 2014). De fato todos os 40 *rankings* obtiveram o mesmo valor em todos os cálculos. Dessa forma, a métrica pode ser considerada confiável.

4.4 Alguns resultados da métrica para *rankings* reais e simulados

A Tabela 2 mostra estatísticas descritivas da $COMP$. Destaca-se que: os valores médios sugerem uma relação positiva entre a métrica e a faixa de correlação; contrastando o valor da métrica máxima com o número de casos em que $COMP > 0,5$, nota-se que, embora não comuns, valores altos da métrica (ex: $COMP > 0,75$) podem ser encontrados em quase todas as faixas de correlação; o percentual de casos em que $COMP > 0,5$ foi superior a 20% apenas nas últimas faixas de correlação (τ).

Tabela 2 – Estatísticas descritivas da métrica

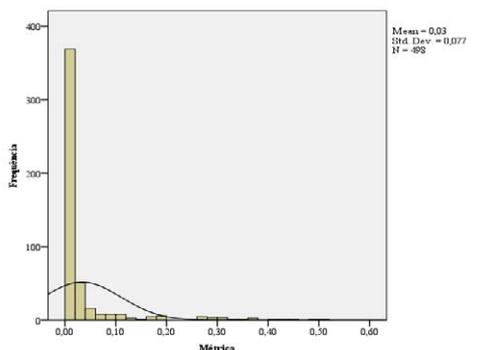
Faixa de correlação	n	COMP			
		μ	σ	Max	>0,5 (%)
$0,3 > \tau$	498	0,03	0,77	0,52	0,20
$0,3 \leq \tau < 0,4$	490	0,05	0,10	0,78	0,41
$0,4 \leq \tau < 0,5$	499	0,07	0,14	0,93	2,61
$0,5 \leq \tau < 0,6$	493	0,09	0,16	0,82	4,06
$0,6 \leq \tau < 0,7$	490	0,18	0,25	0,92	15,51
$0,7 \leq \tau < 0,8$	492	0,23	0,28	0,93	20,94
$0,8 \leq \tau < 0,9$	481	0,46	0,33	0,95	54,89
$\tau \geq 0,9$	487	0,65	0,29	0,97	77,00

Notas: o valor mínimo foi 0,01 nas duas últimas faixas de correlação e 0,00 nas demais; valores arredondados na segunda casa decimal.

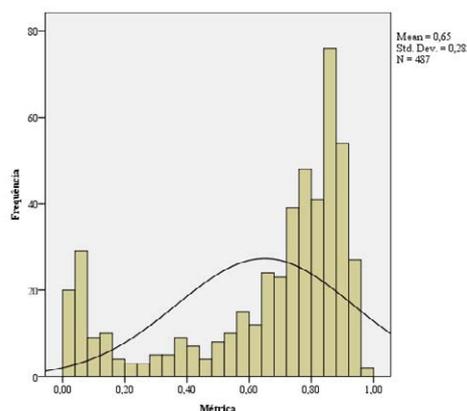
Fonte: elaborada pelos autores.

Para oferecer uma visão da distribuição da métrica por faixa de correlação a Figura 2 foi fornecida. Destaca-se que, mesmo quando $\tau \geq 0,9$, não foram raros os casos em que a $COMP < 0,5$.

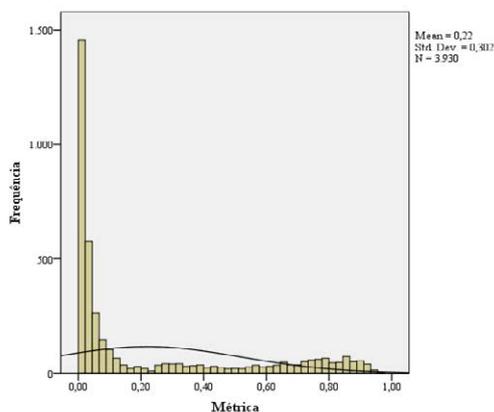
Figura 2 – Alguns histogramas das métricas obtidas para os rankings simulados



(a)



(b)



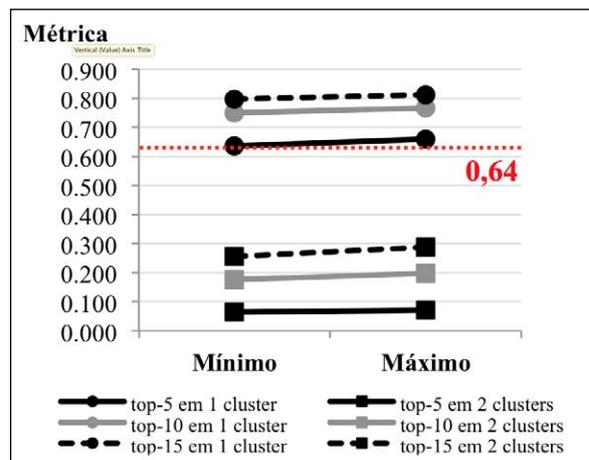
(c) todas as faixas de correlação

Fonte: elaborada pelos autores.

Para auxiliar na avaliação da relação entre a métrica e a dispersão das entidades *top* pelos *clusters*, apresentou-se a Figura 3. Nota-se que: independente da categoria *top*, quando as entidades *top* aparecem dispersas em 2 *clusters* o valor da métrica, se com-

parado com 1 *cluster*, é bem menor; *rankings* com *COMP* < 0,64 possuem maior probabilidade de terem as entidades *top-5* dispersas em mais de um *cluster*. Condições similares podem ser observadas para valores abaixo de: 0,75 para a *top-10*; e 0,80 para a *top-15*.

Figura 3 – Relacionamento entre a métrica e a dispersão das entidades top pelos *clusters*

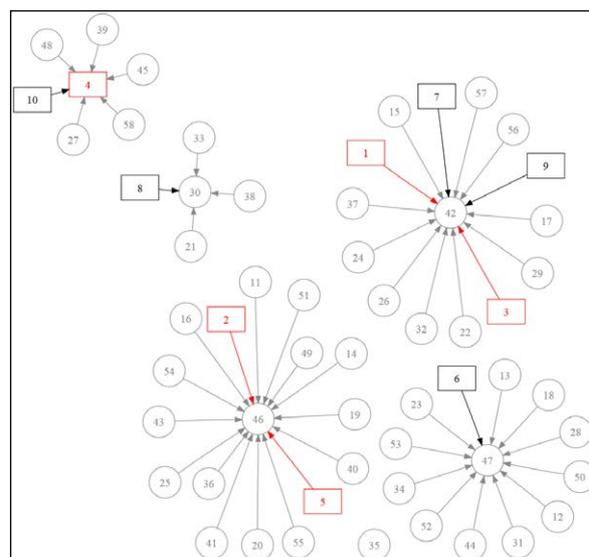


Nota: os valores mínimo e máximo sinalizam, com 95% de confiança, o intervalo da métrica; *n* = 3.930.

Fonte: elaborada pelos autores.

Para ilustrar uma aplicação num *ranking* real, a Figura 4 exibe dados do RGI. Percebe-se que o valor da métrica (0,006) é muito pequeno, o que pode ser justificado pela alta dispersão das entidades *top-5* e *top-10* pelos *clusters*.

Figura 4 – Clusters gerados para o RGI 2013



Notas: entidades representadas pelas posições no ranking.

Fonte: elaborada pelos autores.

Para fins de comparação, cálculos da métrica para outros *rankings* reais foram sumarizados na Tabela 3. Pode-se observar que: o FSI obteve o maior valor da métrica (0,531) e foi o único *ranking* com as *top-10* no mesmo *cluster*; os valores da métrica, no geral, estão compatíveis com a dispersão das entidades pelos *clusters*.

Tabela 3 – Características de alguns rankings reais

<i>Ranking</i>	<i>COMP</i>	$\tau_{\text{médio}}$	top-5*	top-10*
ARWU **	0,228	0,302	2	3
EFW **	0,466	0,228	1	2
FSI **	0,531	0,550	1	1
HDI ***	0,121	0,600	2	2

Notas: *número de *clusters* nos quais as entidades top estão distribuídas; **2012; ***2013.

Fonte: elaborada pelos autores.

Nesse ponto, alguns aspectos podem ser destacados para efeitos de referência sobre o padrão dos resultados obtidos através do AP. Observou-se que, nos *rankings* simulados: o maior *cluster* tinha de 18% a 97,7% das entidades; em 19,5% dos *rankings* não foi encontrado nenhum *cluster* com apenas uma entidade; e em 1,5% dos casos, $n - 5$ entidades estavam representadas no mesmo *cluster* (onde n é o número de entidades).

5 Conclusão/Considerações

Em termos práticos a métrica *COMP* representa um avanço em relação à interpretação das informações fornecidas pelos *rankings*. Se divulgada conjuntamente com o próprio *ranking*, ela poderá oferecer, aos usuários, uma visão até então ocultada pelas suas próprias percepções. De modo sucinto, entende-se que ela poderá indicar situações em que análises mais aprofundadas poderão ser necessárias para um correto julgamento da realidade das entidades. Obviamente a decisão de buscar ou não essas análises dependerá de cada usuário e de cada contexto. Entretanto, sugere-se que *rankings* com $COMP < 0,64$ – que dificilmente estariam expondo informações compatíveis com as percepções dos usuários – deveriam ser analisados em detalhes.

Para a realização dessas análises, desencoraja-se a utilização de agrupamentos de entidades até então comumente utilizadas. Essas classificações, que

não baseiam-se, necessariamente, nas similaridades entre as entidades, são aquelas dependentes dos índices ou de critérios como localização geográfica e PIB.

Dessa forma, outros métodos, principalmente aqueles que permitem a análise da similaridade entre entidades, como recomendado (HOYLAND; MOENE; WILLUMSEN, 2012), podem ser usados. Defende-se, considerando a própria aderência ao contexto de *rankings*, a utilização do AP com matriz de DM para a geração de *clusters* de entidades para fins dessas análises. Entretanto, entende-se que outros métodos de *clustering* e/ou outras medidas de similaridade poderiam ser explorados, dependendo da natureza dos dados.

Nesse cenário, outros métodos também poderiam trazer contribuição prática. Por exemplo, o método baseado em conceitos de cognição que busca encontrar a melhor forma estrutural para representar um conjunto de dados, desenvolvido por Kemp e Tenenbaum (2008), poderia ser explorado. Neste método, não só as entidades seriam agrupadas em *clusters*, mas as relações entre eles poderiam ser identificadas. Um estudo que explora o potencial dessa ferramenta no contexto de *rankings*, especialmente no FSI, pode ser encontrado (MIGON; CUNHA, 2013).

Entende-se, pelo exposto, que a métrica *COMP* pode auxiliar diretamente no processo de tomada de decisão dos usuários dos *rankings*. Ela pode ser interpretada como uma “lente” que oferece novos subsídios para suportar as decisões.

Apesar das vantagens, a métrica apresenta limitações de utilização às quais, na maioria, podem ser vistas como possibilidades de estudos futuros. Primeiro, por utilizar um método probabilístico, existem casos de não convergência em que a métrica não pode ser calculada.

Segundo, a métrica não foi testada para *rankings* com dados faltantes, aspecto este comum em estudos de índices (NARDO *et al.*, 2005). Nesses casos, testes adicionais e eventuais ajustes precisariam ser realizados.

Terceiro, por limitações de cálculo da DM, um grande número de indicadores e/ou um número de indicadores maior do que o de entidades poderiam causar problemas para a computação da matriz variância-covariância. Entretanto, essas condições dificilmente são observadas nos *rankings*.

Quarto, o processo de construção da métrica não é trivial, envolvendo diversos cálculos. Dessa forma, recursos computacionais podem auxiliar na criação dos *clusters*, na determinação da matriz de similaridades, no cálculo e na normalização da métrica. Ou seja, habilidades em programação são úteis para a manipulação da ferramenta o que pode impor limitações para a sua utilização.

Quinto, a solução apresentada é dependente de aspectos definidos com certa arbitrariedade, como: a medida de similaridade; o método de *clustering*; e o número de *clusters*. Dessa forma, mesmo mantendo os conceitos da psicologia utilizados, variando esses itens, a métrica poderia ser alterada.

Sexto, a base conceitual utilizada na criação da métrica provém da psicologia (foco em pessoas físicas). Ou seja, pessoas jurídicas – como, por exemplo, empresas e países –, que podem ser usuários dos *rankings*, não foram consideradas. Portanto, a compatibilidade entre as informações “percebidas” por elas e aquelas disponibilizadas pelos *rankings* pode não estar sendo representada.

Por fim, entende-se que as limitações apresentadas não diminuem a importância da métrica para a avaliação de *rankings*. Ao contrário, acredita-se que essa proposta de mensuração contribui para evidenciar um aspecto até então pouco considerado em contextos de *rankings*.

REFERÊNCIAS

- ABU SHARKH, M.; GOUGH, I. Global welfare regimes: a cluster analysis. **Global Social Policy**, 23 mar. 2010. v. 10, n. 1, p. 27–58.
- ALLEN, V.L.; WILDER, D.A. Group Categorization and Attribution of Belief Similarity. **Small Group Behavior** v. 10, n. 1, p. 73–80, 1979.
- APELTSIN, L. *et al.* Improving the quality of protein similarity network clustering algorithms using the network edge weight distribution. **Bioinformatics (Oxford, England)**, 2011. v. 27, n. 3, p. 326–33.
- BANDYOPADHYAY, S.; SAHA, S. **Unsupervised classification - similarity measures, classical and metaheuristic approaches, and applications**. [S.l.]: [s.n.], 2013.
- BRENNER, L.; ROTTENSTREICH, Y.; SOOD, S. Comparison, Grouping, and Preference. **Psychological Science** v. 10, n. 3, p. 225–229, 1999.
- BONETT, D. G.; WRIGHT, T. A. Sample size requirements for estimating pearson, kendall and spearman correlations. **Psychometrika**, 2000. v. 65, n. 1, p. 23–28.
- BRENNER, L.; ROTTENSTREICH, Y.; SOOD, S. Comparison, grouping, and preference. **Psychological Science**, 1999. v. 10, n. 3, p. 225–229.
- BUSH, R. A. *et al.* Role of reputation in top pediatric specialties *rankings*. **Pediatrics**, 2011. v. 128, n. 6, p. 1168–1172.
- CARARE, O. The impact of bestseller rank on demand: evidence from the app market. **International Economic Review**, 2012. v. 53, n. 3, p. 717–742.
- CAVUSGIL, S. T.; KIYAK, T.; YENIYURT, S. Complementary approaches to preliminary foreign market opportunity assessment: country clustering and country ranking. **Industrial Marketing Management**, 2004. v. 33, n. 7, p. 607–617.
- CENTER FOR WORLD-CLASS UNIVERSITIES OF SHANGHAI JIAO TONG UNIVERSITY (CWCU). Academic ranking of world universities – 2012. **ShanghaiRanking Consultancy**, [S.l.], 2012.
- CHEN, C.-M. Classification of scientific networks using aggregated journal-journal citation relations in the journal citation reports. **Journal of the American Society for Information Science and Technology**, 2008. v. 59, n. 14, p. 2296–2304.
- COOK, D. A; BECKMAN, T. J. Current concepts in validity and reliability for psychometric instruments: theory and application. **The American journal of medicine**, 2006. v. 119, n. 2, p. 166.e7–16.
- COUPLAND, Nikolas. How frequent are numbers? **Language & Communication** v. 31, n. 1, p. 27–37, 2011.
- DAHAENE, Stanislas; MEHLER, Jacques. Cross-linguistic regularities in the frequency of number words. **Cognition** v. 43, p. 1–29, 1992.
- DE LUSIGNAN, S *et al.* End-digit preference in blood pressure recordings of patients with ischaemic heart disease in primary care. **Journal of human hypertension** v. 18, n. 4, p. 261–5, 2004.
- DOWNING, S. M. Reliability : on the reproducibility of assessment data. **Medical Education**, 2004. v. 38, p. 1006–1012.

- DUBES, R. C. How many clusters are best? - an experiment. **Pattern Recognition**, 1987. v. 20, n. 6, p. 645–663.
- FERNANDES, S. An empirical approach of the distinctive aspects for socioeconomic development. **International Journal of Social Economics**, 2013. v. 40, n. 11, p. 956–970.
- FIELD, A. **Discovering statistics using spss**. 3 ed. [S.l.]: SAGE, 2009.
- FRASER INSTITUTE. **Economic freedom of the world. Dental economics - oral hygiene**. [S.l.]: [s.n.], 2012.
- FREY, B. J.; DUECK, D. Clustering by passing messages between data points. **Science (New York, n.Y.)**, 2007. v. 315, n. 5814, p. 972–6.
- HAIR, J. F. *et al.* **Multivariate data analysis**. 5th. ed. New Jersey.: Prentice-Hall, 1998.
- HONG-WEI, L. Community detection by affinity propagation with various similarity measures. in. [S.l.]: IEEE, 2011. p. 182–186.
- HORN, A. S.; HENDEL, D. D.; FRY, G. W. Ranking the international dimension of top research universities in the united states. **Journal of Studies in International Education**, 2007. v. 11, n. 3-4, p. 330–358.
- HORNIK, J.; CHERIAN, J.; ZAKAY, D. The influence of prototypic values on the validity of studies using time estimates. **Journal of the Market Research Society**, 1994.
- HOYLAND, B.; MOENE, K.; WILLUMSEN, F. The tyranny of international index rankings. **Journal of Development Economics**, 2012. v. 97, n. 1, p. 1–14.
- ISAAC, M. S.; SCHINDLER, R. M. The top-ten effect: consumers' subjective categorization of ranked lists. **Journal of Consumer Research**, 2013.
- J. MCQUEEN. Some methods for classification and analysis of multivariate observations. [S.l.]: [s.n.], 1967. p. 281–297.
- JAIN, A. K. Data clustering: 50 years beyond k-means. **Pattern Recognition Letters**, 2010. v. 31, n. 8, p. 651–666.
- JAIN, A. K.; MURTY, M. n.; FLYNN, P. J. Data clustering : a review. **ACM Computing Surveys**, 1999. v. 31, n. 3, p. 264–323.
- JANSEN, C.J.M.; POLLMANN, M.M.W. On Round Numbers : Pragmatic Aspects of Numerical Expressions. **Journal of Quantitative Linguistics** v. 8, n. 3, p. 187–201, 2001.
- JÖNS, H.; HOYLER, M. Global geographies of higher education: the perspective of world university rankings. **Geoforum**, 2013. v. 46, p. 45–59.
- KAUFMAN, E. L. *et al.* The discrimination of visual number. **The American Journal of Psychology**, 1949. v. 62, n. 4, p. 498–525.
- KEMP, C.; TENENBAUM, J. B. The discovery of structural form. **Proceedings of the National Academy of Sciences of the United States of America**, 2008. v. 105, n. 31, p. 10687–92.
- KETTENRING, J. R. The practice of cluster analysis. **Journal of Classification**, 2006. v. 23, n. 1, p. 3–30.
- KIDDLE, S. J. *et al.* Temporal clustering by affinity propagation reveals transcriptional modules in arabidopsis thaliana. **Bioinformatics (Oxford, England)**, 2010. v. 26, n. 3, p. 355–62.
- KUBOVY, Michael; WAGEMANS, Johan. Grouping by Proximity and Multistability in Dot Lattices: A Quantitative Gestalt Theory. **Psychological Science** v. 6, n. 4, p. 225–234, 1995.
- LASKI, Elida V; SIEGLER, Robert S. Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. **Child development** v. 78, n. 6, p. 1723–43, 2007.
- LECLERC, F.; HSEE, C. K.; NUNES, J. C. Narrow focusing: why the relative position of a good in its category matters more than it should. **Marketing Science**, 2005. v. 24, n. 2, p. 194–205.
- LEDOLTER, J.; ABRAHAM, B. Parsimony and its importance in time series forecasting. **Technometrics**, 1981. v. 23, n. 4, p. 411–414.
- LEONE, M.; SUMEDHA; WEIGT, M. Clustering by soft-constraint affinity propagation: applications to gene-expression data. **Bioinformatics (Oxford, England)**, 2007. v. 23, n. 20, p. 2708–15.
- LOCKSLEY, Anne; ORTIZ, Vilma; HEPBURN, Christine. Social categorization and discriminatory behavior: Extinguishing the minimal intergroup

discrimination effect. **Journal of Personality and Social Psychology** v. 39, n. 5, p. 773–783, 1980.

MADDOX, Keith B. *et al.* Social influences on spatial memory. **Memory & Cognition** v. 36, n. 3, p. 479–494, 2008.

MAESSCHALCK, R. DE; JOUAN-RIMBAUD, D.; MASSART, D. L. The mahalanobis distance. **Chemometrics and Intelligent Laboratory Systems**, 2000. v. 50, n. 1, p. 1–18.

MAHALANOBIS, P. C. On the generalized distance in statistics. Calcutta: [s.n.], 1936. V. 2, p. 49–55.

MAKI, R. H. Why do categorization effects occur in comparative judgment tasks? **Memory & cognition** v. 10, n. 3, p. 252–64, 1982.

MARCOT, B. G. Metrics for evaluating performance and uncertainty of bayesian network models. **Ecological Modelling**, 2012. v. 230, p. 50–62.

MARWELL, G. Departmental demography and reputational success: the fall and rise of top sociology departments 1950–1980. **The American Sociologist**, 2012. v. 43, n. 3, p. 294–309.

MAVRI, M. Classifying greek banks based on bank ranking index (bri). **Benchmarking: An International Journal**, 2013. v. 20, n. 5, p. 607–624.

MCLACHLAN, G. F. Mahalanobis distance. **Resonance**, 1999. n. June, p. 20–26.

MIGON, E. X. F. G.; CUNHA, R. M. Descobrimos a forma estrutural: uma apreciação empírica da (in)segurança africana sob a perspectiva da ciência cognitiva. In: Seminário Brasileiro de Estudos Estratégicos Internacionais, 2013, Porto Alegre. **O Atlântico Sul como Eixo da Inserção Internacional do Brasil**. Porto Alegre: [s.n.], 2013.

MILLER, G. A. The magical number seven, plus or minus two: some limits on our capacity for processing information. **Psychological Review**, 1956. v. 63, p. 81–97.

MILLER, T.; HOLMES, K. R.; FEULNER, E. J. **2013 index of economic freedom**. Washington, D.C.; New York: [s.n.], 2013.

MISHRA, Arul; MISHRA, Himanshu. Border bias: the belief that state borders can protect against disasters. **Psychological science** v. 21, n. 11, p. 1582–6, 2010.

NAUDÉ, P.; HENNEBERG, S.; JIANG, Z. Varying routes to the top: identifying different strategies in the mba marketplace. **Journal of the Operational Research Society**, 2009. v. 61, n. 8, p. 1193–1206.

NOLEN, A. L. The content of educational psychology: an analysis of top ranked journals from 2003 through 2007. **Educational Psychology Review**, 2009. v. 21, n. 3, p. 279–289.

ONDA, K. *et al.* Country clustering applied to the water and sanitation sector: a new tool with potential applications in research and policy. **International journal of hygiene and environmental health**, 2014. v. 217, n. 2-3, p. 379–85.

ÖNSEL, Ş. *et al.* A new perspective on the competitiveness of nations. **Socio-Economic Planning Sciences**, 2008. v. 42, n. 4, p. 221–246.

OTTEN, S. “Me and us” or “us and them”? The self as a heuristic for defining minimal ingroups. **European review of social psychology** v. 13, n. 1, p. 1–33, 2002.

PITSILIS, G.; ZHANG, X.; WANG, W. Clustering recommenders in collaborative filtering using explicit trust information. [S.l.]: Springer Berlin Heidelberg, 2011. p. 82–97.

POPE, Devin; SIMONSOHN, Uri. Round numbers as goals: evidence from baseball, SAT takers, and the lab. **Psychological science** v. 22, n. 1, p. 71–9, 2011.

PROBABILISTIC AND STATISTICAL INFERENCE GROUP (PSIG). Affinity propagation faq. **University of Toronto**, [S.l.], 2014. Disponível em: <<http://www.psi.toronto.edu/affinitypropagation/faq.html>>.

RAD, A.; NADERI, B.; SOLTANI, M. Clustering and ranking university majors using data mining and ahp algorithms: a case study in iran. **Expert Systems with Applications**, 2011. v. 38, n. 1, p. 755–763.

RENDE, S.; DONDURAN, M. Neighborhoods in development: human development index and self-organizing maps. **Social Indicators Research**, 2011. v. 110, n. 2, p. 721–734.

REVENUE WATCH INSTITUTE (RWI). **The 2013 resource governance index - a measure of**

transparency and accountability in the oil, gas and mining sector. [S.l.]: [s.n.], 2013.

RODE, M.; COLL, S. Economic freedom and growth. which policies matter the most? **Constitutional Political Economy**, 2011. v. 23, n. 2, p. 95–133.

SEHGAL, A. R. The role of reputation in u. s. news & world report's *rankings* of the top 50 american hospitals. [S.l.]: [s.n.], 2010. p. 521–525.

SHARMA, R.; GIVENS-SKEATON, S. Ranking the top 100 firms according to gender diversity. **Advancing Women in Leadership Journal**, 2010. v. 30, n. 3.

SHEPARD, Roger n; KILPATRIC, Dan W; CUNNINGHAM, James P. The Internal Representation of Numbers. **Cognitive Psychology** v. 7, p. 82–138, 1975.

TAMMI, T. The competitive funding of university research: the case of finnish science universities. **Higher Education**, 2009. v. 57, n. 5, p. 657–679.

TARRANT, M. A.; MANFREDO, M. J. Digit preference, recall bias, and nonresponse bias in self reports of angling participation. **Leisure Sciences** v. 15, n. 3, p. 231–238, 1993.

TEGARDEN, D. P.; TEGARDEN, L. F.; SHEETZ, S. D. Cognitive factions in a top management team: surfacing and analyzing cognitive diversity using causal maps. **Group Decision and Negotiation**, 2007. v. 18, n. 6, p. 537–566.

THE FUND FOR PEACE (TFFF). **Failed states index 2012.** Washington, DC, USA: [s.n.], 2012.

_____. **Conflict assessment framework manual (cast).** Washington, D.C.: [s.n.], 2014a.

_____. Reviews and studies of the cast methodology and the failed states index. [S.l.], 2014b. Disponível em: <<http://ffp.statesindex.org/reviews>>.

THOMSON REUTERS. Times higher education world university *rankings* 2012-2013. **Thomson Reuters**, [S.l.], 2013. Disponível em: <<http://www.timeshighereducation.co.uk/world-university-rankings/2012-13/world-ranking>>

TVERSKY, Barbara. Distortions in Cognitive Maps. **Geoforum** v. 23, n. 2, p. 131–138, 1992.

UNITED NATIONS DEVELOPMENT PROGRAMME (UNDP). **Human development report 2013**

- the rise of the south: humans progress in a diverse world. New York: [s.n.], 2013a.

_____. **Human development report 2013 - technical notes.** New York: [s.n.], 2013b.

VALADKHANI, A.; VILLE, S. Ranking and clustering of the faculties of commerce research performance in australia. **Applied Economics**, 2010. v. 42, n. 22, p. 2881–2895.

VAN OEFFELEN, M P; VOS, P G. Configurational effects on the enumeration of dots: counting by groups. **Memory & cognition** v. 10, n. 4, p. 396–404, 1982.

WANG, J. J.-Y.; BENSMAIL, H. Unified framework for representing and ranking. **Pattern Recognition**, 2013. v. 47, p. 2293–2300.

WARD, J. H. J. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, 1963. v. 58, p. 236–244.

WILSON, A. **Women trailblazers: strategies to secure top ranking positions often dominated by men.** [S.l.]: Capella University, 2009.

XIAO, H.; GUO, P. Iris image analysis based on affinity propagation algorithm. [S.l.]: Springer Berlin Heidelberg, 2009. p. 943–949.

YANG, P.; TAO, L. Perspective: ranking of the world's top innovation management scholars and universities. **Journal of Product Innovation Management**, 2012. v. 29, n. 2, p. 319–331.

ZHANG, J. *et al.* Analysis of activity in fmri data using affinity propagation clustering. **Computer methods in biomechanics and biomedical engineering**, 2011. v. 14, n. 3, p. 271–281.