

Algoritmo para agrupamento automático de descontinuidades baseado no método *Fuzzy K-means*

André Monteiro Klen¹; Pedro Gabriel de Carvalho²

Resumo – O reconhecimento e a associação das descontinuidades de um maciço rochoso em agrupamentos e posteriormente a definição de suas orientações médias é uma das tarefas mais importantes para se estabelecer as relações de estabilidades do maciço. Nesse artigo é apresentado o algoritmo Nébula, desenvolvido para o agrupamento automático de descontinuidades. O Nébula é baseado no método *Fuzzy K-means* e ameniza as subjetividades relacionadas à interpretação do diagrama de densidade de polos como por exemplo: tamanho da área da célula de contagem e identificação do número de agrupamentos. Para verificar a validade do algoritmo, primeiramente, ele é aplicado em um banco de dados artificial, e em seguida seu desempenho é comparado com a de outros algoritmos de agrupamentos, com a utilização de um banco de dados real descrito na literatura. Em ambos os casos o Nébula apontou resultados consistentes e compatíveis com o de outros métodos, na associação das descontinuidades aos agrupamentos e na determinação do número de conjuntos.

Abstract – The recognition and association of discontinuities of a rock mass in clusters and later definition of its average orientations, is one of the most important tasks to establish the relations of stability of the rock mass. This paper presents the Nébula algorithm, developed for the automatic grouping of discontinuities. The Nébula is based on the Fuzzy K-means method and minimizes the subjectivities related to the interpretation of the pole density diagram, such as: size of the count cell area and identification of the number of potential clusters. To check the validity of the algorithm, it is first tested on an artificial database, and then its performance is compared with that of other clustering algorithms, using a real database described in the literature. In both cases, the Nébula showed consistent and compatible results with those of other methods in the association of discontinuities with the clusters and in the determination of the number of sets.

Palavras-Chave – Agrupamentos de descontinuidades; Análise de agrupamentos; Método *Fuzzy K-means*;

¹ Gerais - Campus Ouro Preto, Eng., PhD. Instituto Federal de Minas (31) 987616806, andre.klen@ifmg.edu.br

² Eng. Estudante. Universidade Federal de Ouro Preto. (31)988384146, pedro.gabriel@aluno.ufop.edu.br

1. INTRODUÇÃO

A definição das descontinuidades em agrupamentos e a caracterização de suas atitudes médias é um dos aspectos mais importantes em aplicações de engenharia geotécnica, uma vez que, esses conjuntos definem um sistema que rege os mecanismos e as geometrias de ruptura dos maciços rochosos.

As descontinuidades podem controlar toda a estabilidade do maciço rochoso, pois conferem a ele um comportamento em termos de deformabilidade, resistência ao cisalhamento e permeabilidade substancialmente diferente do material que o compõem. (KLEN, LANA, 2014; SONG *et al.*, 2017).

Conforme Gao *et al.* (2019), a definição dos agrupamentos de descontinuidades envolve a interpretação visual do diagrama de polos. Essa técnica, conhecida como método clássico, apesar de ser amplamente empregada é influenciada por fatores subjetivos, os quais podem conduzir a resultados divergentes. Os principais fatores são: a experiência profissional, o tamanho da célula de contagem, a presença de sobreposição e *outliers* entre os conjuntos e a definição do número de agrupamentos.

Assim, de acordo com Jimenez e Sitar (2006) e Song *et al.* (2017), a interpretação visual do diagrama de polos não é satisfatória em muitos casos, o que conduz ao desenvolvimento de técnicas matemáticas e estatísticas para a identificação automática dos conjuntos de descontinuidades que minimizem os fatores de subjetividade.

Entre essas técnicas a análise de agrupamentos se destaca, ela é capaz de agrupar as descontinuidades sem nenhuma informação, a priori, do conjunto de dados. Para isso, é necessário um algoritmo de agrupamento e os mais indicados para essa tarefa são o *K-means* e a sua variação o *Fuzzy K-means* (FKM). (JIMENEZ, SITAR, 2006; LIU *et al.*, 2017).

O FKM se destaca em relação ao *K-means*, porque ele considera a sobreposição entre os agrupamentos e os *outliers* pouco influenciam no resultado final. Contudo, os dois métodos não sofrem interferência do tamanho da célula de contagem. Principalmente, após os estudos de Hammah e Curran (1998), o FKM e o *K-means* têm sido aplicados por diversos autores entre os quais se destacam os trabalhos de Jimenez e Sitar (2006), Klen e Lana (2014), Li *et al.* (2015) e mais recentemente Cui e Yan. (2020).

Dessa forma, o artigo tem como objetivo apresentar o algoritmo Nébulas, desenvolvido por Klen e Lana (2014), e os seus resultados alcançados no agrupamento automático de descontinuidades. O Nébulas é baseado no método *Fuzzy K-means* e desenvolvido no *software* livre de computação numérica Scilab.

O artigo se divide da seguinte forma: a sessão 2, apresenta os detalhes e parâmetros do método *Fuzzy K-means* e, conseqüentemente, do algoritmo Nébulas. Na sessão 3, o Nébulas é testado em relação a uma base de dados artificial, para depois ter seu desempenho comparado com a de outros algoritmos de agrupamentos, usando uma base de dados descrita na literatura. Finalmente, na sessão 4, as conclusões são apresentadas.

2. METODOLOGIA E PARÂMETROS DO FUZZY K-MEANS PARA O AGRUPAMENTO DE DESCONTINUIDADES.

Nos tópicos seguintes são apresentados a metodologia, as etapas e os parâmetros do algoritmo FKM para o agrupamento de descontinuidades. Essas informações são baseadas nos trabalhos de Klen e Lana (2014) e Hammah e Curran (1998).

O objetivo do método FKM é particionar o conjunto de descontinuidades em K agrupamentos formados por elementos similares. Para isso, ele busca por regiões com alta densidade de dados, minimizando a distância entre as descontinuidades e o centro dos agrupamentos através da função objetiva Equação (1).

$$\min: J = \sum_{i=1}^N \sum_{j=1}^K u_{ij}^m d^2(X_i, V_j) \quad (1)$$

Onde: N é o número de descontinuidades do conjunto de dados, K é o número de agrupamentos, u_{ij}^2 é o grau de inclusão e $d^2(X_i, V_j)$ é a distância entre a descontinuidade X_i e o centro do agrupamento V_j .

2.1. Representação das descontinuidades

As descontinuidades de um maciço rochoso, geralmente, são consideradas como sendo uma estrutura planar e suas orientações espaciais podem ser expressas em termos de *dip direction*, direção de mergulho, ($0^0 \leq \alpha \leq 360^0$) e *dip angle*, mergulho, ($0^0 \leq \beta \leq 90^0$). A orientação de uma descontinuidade qualquer i também pode ser representada por um vetor unitário normal, o qual pode ser determinado pelo cosseno diretor $X_i = (x_i, y_i, z_i)$, Equação (2).

A Figura 1 mostra o vetor unitário normal X_i da descontinuidade i no sistema de coordenadas Cartesianas. O eixo positivo x é horizontal a direção norte, o eixo positivo y é horizontal a direção leste e o eixo positivo z é direcionado para cima.

$$X_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{cases} \cos\alpha \sin\beta \\ \sin\alpha \sin\beta \\ \cos\beta \end{cases} \quad (2)$$

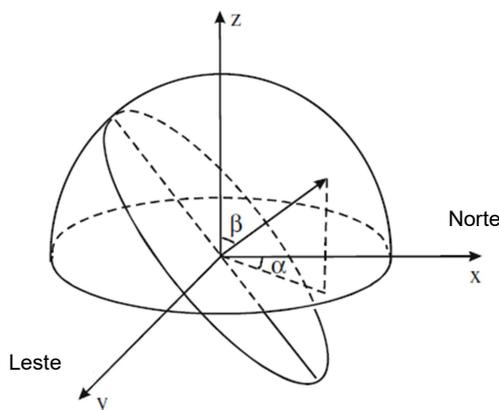


Figura1: Representação da orientação da descontinuidade. Adaptado de Shanley e Mahtab (1976).

2.2. Inicialização do algoritmo

O método FKM é sensível aos centros iniciais, isso quer dizer que, diferentes centros iniciais podem conduzir a agrupamentos finais, também diferentes. Para remediar essa questão existem diversos algoritmos como, por exemplo, os trabalhos de: Jimenez e Sitar (2006), Klen (2015), Li et al. (2015).

Nesse caso, aplica-se o algoritmo proposto por Klen (2015) que utiliza a distribuição de Fisher para dados esféricos, na geração aleatória dos vetores unitários dos centros iniciais. Esse método possibilita que os centros sejam formados, de tal maneira que, a menor distância entre eles seja a maior possível. Essa restrição garante que o FKM seja corretamente inicializado.

2.3. Medida de distância

A medida de distância é um importante parâmetro no agrupamento das descontinuidades, pois ela mensura a similaridade entre os elementos. A medida mais utilizada nesse caso é o valor do seno ao quadrado do ângulo agudo entre os vetores unitários normais das descontinuidades. (Hammah, Curran, 1998; Gao et al., 2019).

Essa é a medida aplicada no algoritmo, a Figura 2, mostra o conceito que estabelece a similaridade através do valor do seno ao quadrado do ângulo agudo entre as descontinuidades i e j representadas respectivamente por seus vetores unitários normais n_i e n_j .

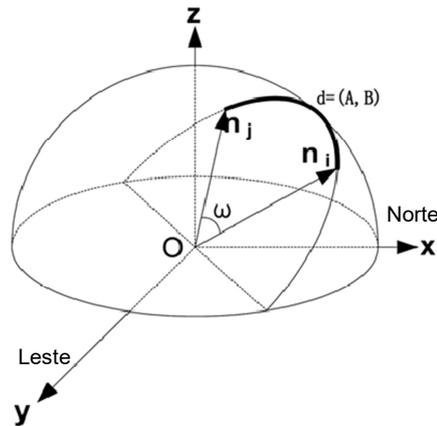


Figura 2: Representação do ângulo entre duas descontinuidades. Adaptado de Gao *et al.* (2019).

Considere que, o ângulo ω entre os vetores unitários n_i and n_j é calculado como:

$$\cos \omega = |n_i \cdot n_j^T| = |x_i x_j + y_i y_j + z_i z_j| \quad (3)$$

Onde \cdot é o produto escalar entre dois vetores.

De acordo com a Equação (3), a medida do valor do seno ao quadrado pode ser escrita como:

$$\text{dist}(A, B) = \sin^2 \omega = 1 - \cos^2 \omega \quad (4)$$

Assim, a distância entre a descontinuidade X_i e o centro V_j , de qualquer agrupamento é:

$$d^2(X_i, V_j) = 1 - (X_i \cdot V_j)^2 \quad (5)$$

2.4. Grau de inclusão

O grau de inclusão, u_{ij} , permite que o FKM avalie a sobreposição entre os agrupamentos e identifique os *outliers*. Ele pode ser entendido como a probabilidade da descontinuidade pertencer a um determinado conjunto, e $u_{ij} \in [0, 1]$, sendo inversamente proporcional à distância das amostras aos centros dos agrupamentos. Ele é definido pela Equação (6).

$$u_{ij} = \frac{\left[\frac{1}{d^2(X_i, V_j)} \right]^{\frac{1}{m-1}}}{\sum_{t=1}^K \left[\frac{1}{d^2(X_i, V_t)} \right]^{\frac{1}{m-1}}} \quad (6)$$

Onde m é o grau *fuzzy* e assume $m=2$ esse é o melhor valor para esse tipo de aplicação. (Cui, Yan, 2020; Hammah e Curran, 1998).

2.5. Cálculo dos centros dos agrupamentos

Para se determinar os centros dos agrupamentos é necessário definir os autovalores e autovetores da matriz de orientação S de cada agrupamento.

$$S_j = \begin{bmatrix} \sum_{i=1}^N (u_{ij})^m x_i x_i & \sum_{i=1}^N (u_{ij})^m x_i y_i & \sum_{i=1}^N (u_{ij})^m x_i z_i \\ \sum_{i=1}^N (u_{ij})^m x_i y_i & \sum_{i=1}^N (u_{ij})^m y_i y_i & \sum_{i=1}^N (u_{ij})^m z_i y_i \\ \sum_{i=1}^N (u_{ij})^m x_i z_i & \sum_{i=1}^N (u_{ij})^m y_i z_i & \sum_{i=1}^N (u_{ij})^m z_i z_i \end{bmatrix} \quad j = 1, 2 \dots K$$

Onde (x_i, y_i, z_i) é o cosseno diretor da i -ésima descontinuidade i .

Os autovalores de S_j representados por $(\tau_{j1}, \tau_{j2}, \tau_{j3})$, ordenados de tal forma que $(\tau_{j1} < \tau_{j2} < \tau_{j3})$ estão relacionados aos seus respectivos autovetores normalizados $[\xi_{j1}, \xi_{j2}, \xi_{j3}]$. Portanto, o autovetor associado ao maior autovalor, neste caso τ_{j3} e ξ_{j3} , é um excelente estimador do centro do agrupamento K . (Cui, Yan, 2020; Hammah e Curran, 1998).

2.6. Medida de Validação

As medidas de validação são um método efetivo para se determinar o número ideal de agrupamentos K , pois na maioria das vezes essa informação é desconhecida. De forma geral, o número de agrupamentos varia entre 2 e 6 grupos, pois este é o intervalo no qual, normalmente, o conjunto de descontinuidades é dividido na prática. (CUI, YAN, 2020; LIU *et al.*, 2017)

Uma das medidas de validação mais aplicadas no agrupamento de descontinuidades é a medida de Xie e Beni (1991), Equação (7). Ela identifica como ideal o particionamento, no qual, os grupos formados são compactos e distantes entre si. (CUI, Yan, 2020; KLEN, LANA, 2014; XU *et al.*, 2012).

$$XB = \frac{\sigma}{N \times sep_{min}} = \frac{\sum_{j=1}^K \sum_i^N u_{ij}^2 d^2(X_i, V_j)}{N \times \min_{i \neq j} [d^2(X_i, V_K)]} \quad (7)$$

Onde: a relação $\frac{\sigma}{N}$ é entendida como a compacidade e expressa a variabilidade do particionamento e sep_{min} indica a menor distância entre os centros dos agrupamentos.

Para se determinar o número ideal de agrupamentos o algoritmo Nébulas deve ser executado para os valores de K dentro do intervalo $[K_{min} = 2; K_{max} = 6]$ e os índices de XB calculados e comparados. O menor valor de XB indica o melhor particionamento e conseqüentemente, o número de agrupamentos ideal.

As iterações se encerram quando o critério de parada, Equação (8), é alcançado, em seguida, cada descontinuidade é associada aos agrupamentos de acordo com o seu maior grau de inclusão u_{ij} .

$$\max_{ij} = [|u_{ij} - \hat{u}_{ij}|] < \varepsilon \quad (8)$$

Onde: u_{ij} e \hat{u}_{ij} são os valores dos graus de inclusão das iterações t e $t + 1$ respectivamente e $\varepsilon = 0,01$.

Assim, após a apresentação dos conceitos e parâmetros do método *Fuzzy K-means*, o algoritmo Nébulas pode ser resumido nos seguintes passos:

Parâmetros: $K_{min} = 2$; $K_{máx} = 6$; $\varepsilon = 0,01$;

Enquanto $K_{min} \leq K_{máx}$ faça:

- 1- Calcular os vetores unitários X_i das descontinuidades Eq.(1).
- 2- Gerar os K centros iniciais V_j .
- 3- Definir a distância entre cada descontinuidade X_i e os centros dos agrupamentos V_j , Eq.(5).
- 4- Calcular os graus de inclusão u_{ij} , Eq.(6).
- 5- Atualizar os centros dos agrupamentos V_j , matriz S_j .
- 6- Repetir os passos 3, 4 e 5 até atingir o critério de parada Eq.(8).
- 7- Assinalar as descontinuidades nos agrupamentos com base no maior grau de inclusão.
- 8- Calcular a medida de Validação Eq.(7).

3. Análise e Discussão

3.1. Estudo de caso 1: Banco de dados artificial.

Para avaliar o desempenho do algoritmo Nébulas, utiliza-se uma base de dados artificial, gerada a partir da distribuição de Fisher para dados esféricos, através do método de simulação de Monte Carlo, proposto por Zheng *et al.* (2014), Equação (9).

Essa base permite que o algoritmo seja testado em uma situação complexa, uma vez que, o conjunto é formado por 437 descontinuidades, classificadas em cinco agrupamentos que possuem diferentes concentrações e sobrepõem entre si. Os parâmetros de cada grupo e o diagrama de polos são apresentados respectivamente na tabela 1 e Figura 3(a).

$$f(\theta) = \frac{k \sin \theta \quad k \cos \theta}{e^k - e^{-k}} \quad (9)$$

Onde: k é o parâmetro de concentração, θ é a variação angular em relação à orientação média.

Tabela 1: Parâmetros para geração dos dados artificiais.

Agrupamentos	Número de descontinuidades	Orientação Média (Dip / Dip Direction)	Parâmetros (k)
1	122	71 / 341	27
2	89	82 / 290	13
3	63	68 / 208	17
4	93	10 / 173	10
5	70	65 / 40	35

Ao analisar a Figura 3-a, percebe-se a complexidade em definir o número de conjuntos e as suas estruturas através da inspeção visual do diagrama de polos. Isso é devido ao elevado número de descontinuidades, grupos com diferentes concentrações de dados e da interseção entre os agrupamentos.

A figura 3-b demonstra a base de dados artificial com os agrupamentos identificados e evidencia as regiões de sobreposição. Como descrito na Tabela 2, e na Figura 3-c, os resultados do algoritmo são muito similares aos parâmetros da base de dados artificial. Pode-se explicar que, as diferenças entre os resultados são oriundas da realocação das descontinuidades que se encontram nas regiões de sobreposição.

Por exemplo, o agrupamento sub-horizontal 4 possui descontinuidades que se interceptam com o grupo subvertical 3, destaque na Figura 3-b. Essas apresentam o Dip com valor mais

próximo de 90° do que 0°, e por isso, estão perto do centro do agrupamento 3. Note na Figura 3-c, que o Nébula assinala corretamente essas discontinuidades ao grupo subvertical, revelando que o algoritmo consegue identificar as estruturas e as orientações médias dos conjuntos de forma consistente.

Para avaliar o número de agrupamentos e a qualidade dos conjuntos formados é aplicada a medida de Xie Beni (XB), Tabela 3. O resultado do menor índice de XB aponta que o Nébula identifica corretamente o número de conjuntos e os grupos formados são compactos e separados entre si.

Tabela 2: Comparação dos resultados obtidos pelo Nébula e o banco de dados original

Agrupamentos	Nébula		Dados artificiais	
	Dip / Dip direction	Número de discontinuidades	Dip / Dip direction	Número de discontinuidades
1 - ■	71/340	131	71 / 341	122
2 - ▲	85/288	79	82 / 290	89
3 - ▼	63/207	66	68 / 208	63
4 - +	12/175	76	10 / 173	93
5 - x	66/39	85	65 / 40	70

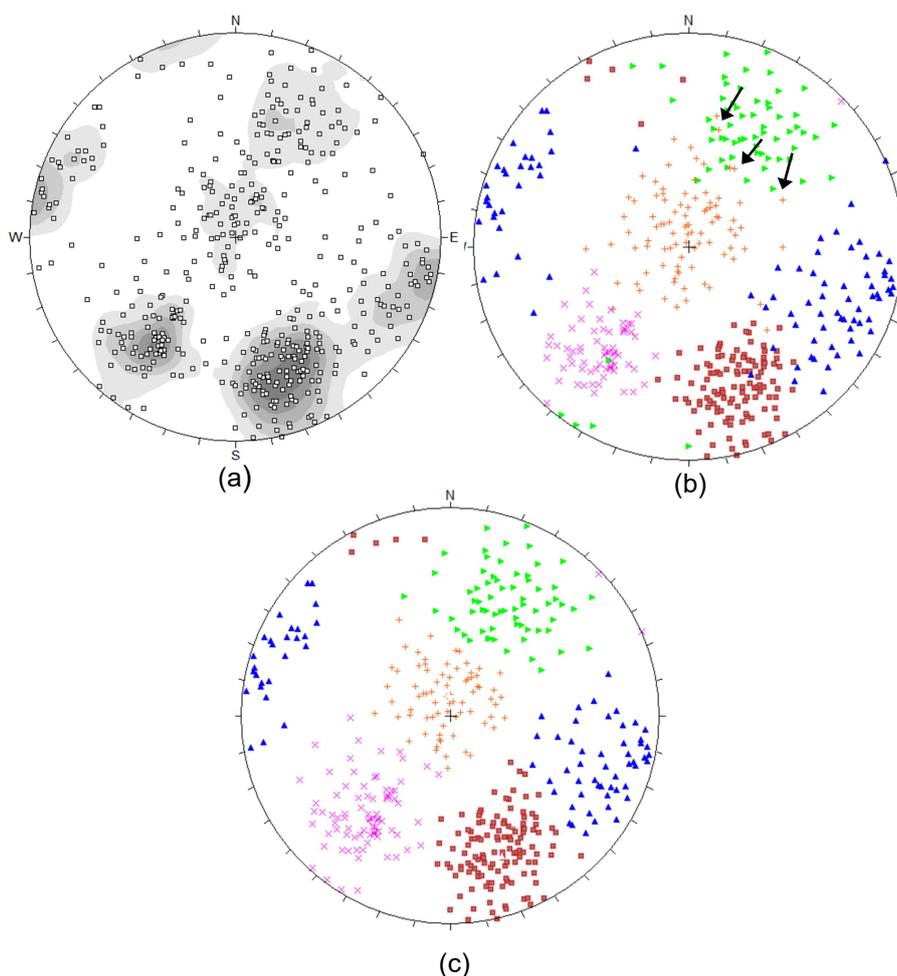


Figura 3: Diagramas de polos do banco de dados artificiais: (a) sem separação; (b) agrupamentos identificados; (c) resultado do Nébula. Projeções estereográficas de igual ângulo e hemisfério inferior.

Tabela 3: Índice Xie-Beni para os agrupamentos gerados pelo Nébulas aplicado a base de dados artificial.

Agrupamentos (K)	2	3	4	5	6
Xie-Beni	0,252	0,118	0,131	0,082	0,353

3.2. Estudo de caso 2: Banco de dados de Shanley e Mahtab(1976)

O banco de dados de Shanley e Mahtab (1976) representa um conjunto de 286 descontinuidades, divididas em três agrupamentos, amostradas na Mina de cobre de San Manuel situada no Arizona, Estados Unidos da América. Esse banco de dados é importante, pois foi usado pelos autores supracitados, na apresentação do primeiro método para o agrupamento automático de descontinuidades.

O método de Shanley e Mahtab (1976), e o banco de dados são usados como referência na análise da performance de diversos algoritmos, uma vez que o número de agrupamentos e as descontinuidades associadas aos mesmos são bem conhecidas. Dentre os algoritmos testados através dessa base de dados destacam-se: Cui e Yan (2020), Hammah e Curran (1998), Jimenez e Sitar (2006), Li *et al.* (2015), Liu *et al.* (2017).

Assim, nesse estudo de caso, o Nébulas é aplicado ao conjunto de descontinuidades e seu desempenho, é comparado aos resultados alcançados pelos algoritmos de Shanley e Mahtab (1976), Jimenez e Sitar (2006) e Li *et al.* (2015). Esses são utilizados como referências, pois são frequentemente testados por diversos autores como, Cui e Yan (2020) e Song *et al.* (2017)

O algoritmo de Shanley e Mahtab (1976) é o primeiro método a apresentar critérios matemáticos para dividir as descontinuidades em agrupamentos. Já o método de Jimenez e Sitar(2006), *Spectral method*, baseia-se na representação das descontinuidades em um novo espaço amostral R^K . Nesse, os K agrupamentos são bem separados e de fácil identificação, assim o algoritmo *K-means* é aplicado para classificar os conjuntos.

Já o algoritmo de Li *et al.* (2015), aplica o método estocástico de otimização *Particle Swarm* para minimizar a função objetivo do *K-means* e definir os grupos de descontinuidades. Essa técnica busca pelo ótimo global da função, o que elimina, na maioria dos casos, a sensibilidade do FKM à escolha dos centros iniciais. A Figura 4 mostra os diagramas de polos para os três agrupamentos resultantes do particionamento do banco de dados pelos algoritmos.

A tabela 4 compara a atitude média e o número de descontinuidades dos agrupamentos definidos pelos métodos: Nébulas, Jimenez e Sitar (2006), Li *et al.*(2015) e Shanley e Mahtab (1976). Já a tabela 5 aponta os resultados da medida de validação Xie Beni e mostra que o algoritmo identifica corretamente o número de agrupamentos, pois o menor valor de XB é para K=3.

Tabela 4: Comparação entre os resultados do Nébulas e dos outros algoritmos

Agrupamentos	Nébulas		Jimenez e Sitar (2006) ¹		Li <i>et al.</i> (2015) ¹		Shanley e Mahtab (1976) ¹	
	Dip / Dip direction	Número	Dip / Dip direction	Número	Dip / Dip direction	Número	Dip / Dip Direction	Número
1 – ■	79/251	115	77/252	124	80/251	108	75/251	127
2 – ▲	80/347	115	81/347	107	77/344	116	75/343	112
3 – ►	17/129	56	12/135	55	16/151	62	8/123	47

¹Dados extraídos de Liu *et al.* (2017).

Tabela 5: Índice Xie-Beni para os agrupamentos gerados pelo Nébulas.

Agrupamentos (K)	2	3	4	5	6
Xie-Beni	0,1902	0,1074	0,2615	0,2102	0,2207

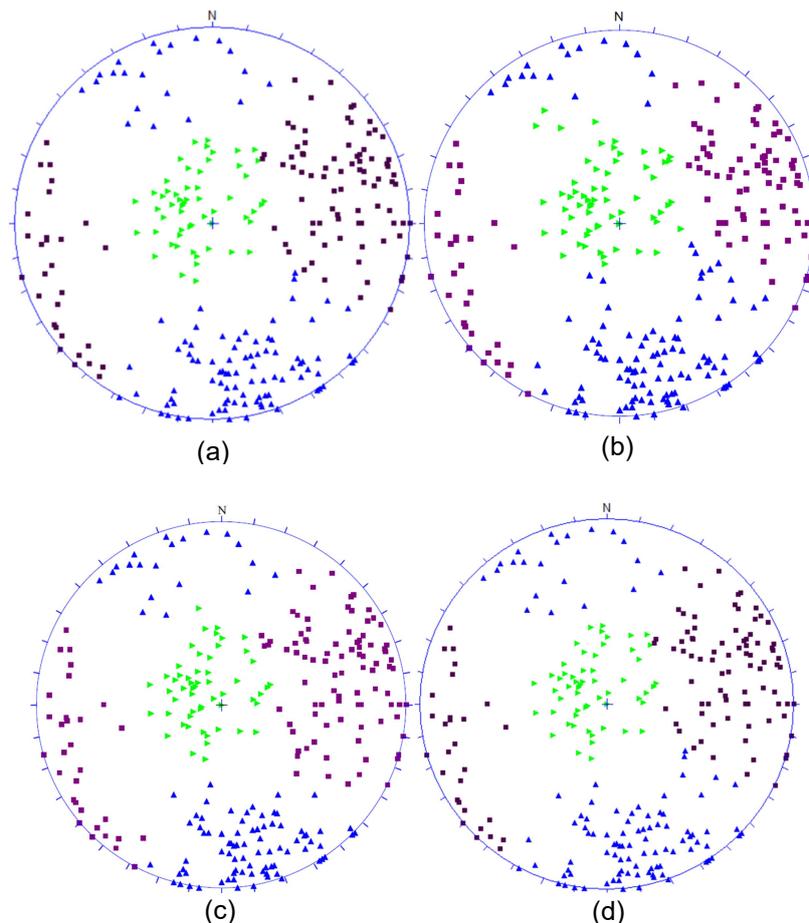


Figura 4. Comparação entre os resultados dos algoritmos para o banco de dados de Shanley e Mahtab (1976): (a) Shanley e Mahtab (1976); (b) Li *et al.* (2015); (c) Nébulas; (d) Jimenez e Sitar (2006). Projeções estereográficas de igual ângulo e hemisfério inferior.

Analisando os resultados, percebe-se que os algoritmos apresentam performances muito parecidas e que a diferença entre o número de descontinuidades e, conseqüentemente, a atitude média dos agrupamentos, tem sua origem na técnica de otimização da função objetivo e nos elementos que estão na região de sobreposição. A forma com a qual cada método lida com esses dois pontos altera a associação das descontinuidades aos agrupamentos.

Conforme Gao *et al.* (2019) e Liu *et al.* (2017), os métodos empregados por Jimenez e Sitar (2006) e Li *et al.* (2015) são complicados em teoria e precisam de parâmetros que são difíceis de estabelecer. Nesse ponto é que o Nébulas se destaca, pois por ser baseado no FKM, ele não requer ajuste de parâmetros ou inserção de nenhum outro método de otimização, como é o caso do algoritmo de Li *et al.* (2015). Isso garante, facilidade de implementação, processamento ágil e consistência aos resultados.

4. CONCLUSÃO

O Nébulas é baseado no algoritmo de particionamento *Fuzzy K-means* e desenvolvido por Klen e Lana (2014) para o agrupamento de descontinuidades. As principais vantagens do algoritmo são: rápida convergência, o fato dele não sofrer influência dos elementos que geram subjetividade no método clássico, classificar automaticamente as descontinuidades e identificar o número de agrupamentos. O algoritmo ainda permite que, os dados sejam acessados através de arquivos com extensão, .txt, e os resultados plotados diretamente em *softwares* de projeção estereográfica.

Os resultados do Nébulas no primeiro estudo de caso, apontam que o algoritmo consegue classificar corretamente as descontinuidades mesmo em conjunto de dados complexos. Essa complexidade está presente, no número de descontinuidades, na quantidade de agrupamentos e nas diversas regiões de sobreposição entre os grupos.

No segundo estudo de caso, o Nébula tem seus agrupamentos confrontados em relação aos resultados dos principais algoritmos estudados na literatura. Para isso, é aplicado o banco de dados de Shanley e Mahtab (1976). Esse conjunto é a referência mais importante quando se discute o desempenho de algoritmos para o agrupamento de descontinuidades. A performance do Nébula é similar aos resultados dos demais algoritmos, demonstrando que ele é uma opção confiável para a definição automática dos agrupamentos de descontinuidades.

Dessa forma, o algoritmo apresentado é uma importante ferramenta na tarefa de se identificar os agrupamentos de descontinuidades. Ela permite que os possíveis mecanismos de ruptura dos maciços rochosos sejam avaliados com maior precisão e conhecimento estatístico.

AGRADECIMENTOS

Os autores agradecem ao Instituto Federal de Minas Gerais - *Campus* Ouro Preto pelo apoio ao fomento à pesquisa e às bolsas de pesquisa.

REFERÊNCIAS

CUI, X.; YAN, E. Fuzzy C-means cluster analysis based on variable length string genetic algorithm for the grouping of rock discontinuity sets **KSCE Journal of Civil Engineering**, v.24, n.11, p. 3237-3246, 2020. DOI: 10.1007/s12205-020-2188-2.

GAO, F.; CHEN, D.; ZHOU, K.; NIU, W.; LIU, H. A Fast Clustering Method for Identifying Rock Discontinuity Sets. **KSCE Journal of Civil Engineering**, v. 23, n. 2, p. 556-566, 2019.

HAMMAH, R. E.; CURRAN, J. H. Fuzzy cluster algorithm for the automatic identification of joint sets. **International Journal of Rock Mechanics & Mining Sciences**, v. 35, n. 7, p. 889-905, 1998.

JIMENEZ, R. R.; SITAR, N. A spectral method for clustering of rock discontinuity sets. **International Journal of Rock Mechanics & Mining Sciences**, n. 43, p. 1052-1061, 2006.

KLEN, A., M. **Algoritmo para agrupamento de descontinuidades em famílias baseado no método Fuzzy K-means**. 2015, 142 f. Tese (Doutorado Geotecnia aplicada à mineração). Universidade Federal de Ouro Preto. Escola de Minas. Núcleo de Geotecnia da Escola de Minas. Ouro Preto, 2015.

KLEN, A. M.; LANA, M. S. Fuzzy algorithm of discontinuity sets. **REM: R. Esc. Minas**, Ouro Preto, 67(4), p. 439-445, oct. dec. 2014. DOI: 10.1590/0370-44672014670178

LI, Y.; WANG, Q.; CHEN, J.; XU, L.; SONG, S. K-means algorithm based on particle swarm optimization for the identification of rock discontinuity sets. **Rock Mechanics and Rock Engineering**, 48 p. 375-385, 2015. DOI: 10.1007/s00603-014-0569-x.

LIU, J.; ZHAO, X. D.; XU, Z. H. Identification of rock discontinuity sets based on modified affinity propagation algorithm. **International Journal of Rock Mechanics & Mining Sciences**, v. 94, p. 32-42, 2017. DOI: 10.1016/j.ijrmms.2017.02.012.

SHANLEY, R. J.; MAHTAB, M. A. Delineation and analysis of clusters in orientation data. **Journal of the International Association for Mathematical Geology**, v. 8, n. 1, p. 9-23, 1976.

SONG, S.; WANG, Q.; CHEN, J.; LI, Y.; ZHANG, W.; RUAN, Y. Fuzzy C-means clustering analysis based on quantum particle swarm optimization algorithm for the grouping of rock discontinuity sets. **KSCE Journal of Civil Engineering**, v.21, n.4, p. 556-566, 2017.

XIE, X. L.; BENI G. A validity measure for fuzzy clustering. **IEEE Transactions on pattern analysis and machine intelligence**, v. 13, n. 8, p. 841-847, 1991.

XU, L. M.; CHEN, J. P.; WANG, Q.; ZHOU, F. J. Fuzzy C-means cluster analysis based on mutative scale chaos optimization algorithm for the grouping of discontinuity sets. **Rock Mechanic and Rock Engineering**, Technical note, 2012.

ZHENG, J.; DENG, J.; YANG, X.; WEI, J. A improved Monte Carlo simulation method for discontinuity orientations based on Fisher distribution and its program implementation. **Comp. Geotech.**, v. 62, p.266-276, 2014.