

# Agrupamento Espectral Aglomerativo: Uma Proposta de Algoritmo

Luciano G. Garcia<sup>1</sup>, Leonardo R. Emmendorfer<sup>2</sup>

<sup>1</sup>Instituto de Matemática, Estatística e Física – Universidade Federal do Rio Grande (FURG)  
Caixa Postal 474 – 96.201-900 – Rio Grande – RS – Brazil

<sup>2</sup>Centro de Ciências Computacionais – Universidade Federal do Rio Grande (FURG)  
Caixa Postal 474 – 96.201-900 – Rio Grande, RS–Brazil

lucianogarim@gmail.com, leonardo.emmendorfer@gmail.com

**Abstract.** *In this paper we present the spectral clustering based on an agglomeration step in the  $k$ -smallest eigenvectors from Laplacian matrix, that represents the dataset in the graph. The proposed algorithm is applied in three datasets of different geometric formats and the results are compared to the  $k$ -means method clustering and the  $k$ -means spectral clustering method. To measure the performance of the algorithms, the  $F$ -measure is used and the results are presented in table and graph form. Finally, a discussion is made about the sensitivity of the spectral method in varying the number of neighbours of an element from the dataset.*

**Resumo.** *Neste artigo é apresentado o agrupamento espectral baseado em uma etapa de aglomeração dos  $k$ -menores autovalores da matriz Laplaciana que representa o conjunto de dados no grafo. O algoritmo proposto é aplicado em três conjunto de dados de formatos geométricos distintos e os resultados são comparados aos agrupamentos obtidos pelo método  $k$ -médias e o método de agrupamento espectral via  $k$ -médias. Para medir a performance dos algoritmos é utilizada a medida- $F$  e os resultados são apresentados em forma de tabela e gráfico. Por fim, é feita uma discussão a respeito da sensibilidade do método espectral na variação do número de vizinhos de um elemento do conjunto de dados.*

## 1. Introdução

A tarefa de agrupar dados de acordo com a sua estrutura organizacional tem impulsionado diversas pesquisas na área de Aprendizagem de Máquina. Um dos principais desafios é propor uma metodologia determinística ou heurística que seja capaz descobrir grupos em um conjunto de dados de modo automático. Neste sentido, métodos de agrupamento espectral vem sendo utilizados com frequência devido à sua fácil implementação e a característica de obter um estudo global das propriedades dos dados. Esta visão global dos algoritmos baseados no método espectral é justificada pelo fato de que a similaridade entre os dados são medidas par a par, desse modo, é gerada uma matriz de similaridade que contém as informações de todo conjunto.

A metodologia de agrupamento espectral proposta por [A. Ng and Y.Weiss 2001] utiliza o algoritmo  $k$ -médias para estabelecer um agrupamento nos  $k$ -menores autovetores da matriz Laplaciana. Com isto, ao invés de agrupar um conjunto de pontos no  $\mathbb{R}^n$

o método permite agrupar os dados em um conjunto no  $\mathbb{R}^k$  onde  $k$  é o número de grupos correspondente aos  $k$ -menores autovetores e  $k < n$ . Um problema decorrente neste mapeamento é que nem sempre os autovetores possuem uma geometria simples tal que o  $k$ -médias possa identificar facilmente os grupos corretos. Neste sentido, a utilização de um algoritmo baseado em aglomeração surge como proposta de contornar o problema em estudo.

### Trabalhos relacionados

A estrutura de alguns conjuntos de dados em função de sua escala de tamanho ou densidade não permite obter um resultado satisfatório do método espectral via  $k$ -médias [Nadler and Galun 2006]. Diversos trabalhos encontrados na literatura utilizam o  $k$ -médias como algoritmo de agrupamento no método espectral [D. Yan and Jordan 2009], [L. Wang and Bezdek 2009], [Chen and Cai 2011], [Sakai and Imiya 2009]. O  $k$ -médias é um dos algoritmos de agrupamento mais comumente usados, porém não funciona bem em conjuntos com *outliers* ou de formatos não-esféricos [L. Ertoz and Kumar 2003]. Para contornar o problema de escalabilidade do conjunto é proposto em [Manor and Perona 2004] o cálculo de um parâmetro de escala local na medida de similaridade entre pontos vizinhos do conjunto. Além disso, em [P. Yang and Huang 2011] surge a proposta de uma medida de similaridade que leva em consideração a sensibilidade de densidade de uma determinada região, ou seja, estabelecendo uma medida distinta da função de similaridade Gaussiana que até então era utilizada na literatura. Por fim, em [N. Tremblay and Vandergheynst 2016] é apresentada uma proposta de compressão do método espectral utilizando *bandlimited graph-signals*, que tem por objetivo diminuir a complexidade do  $k$ -médias. Já neste trabalho o algoritmo proposto foi escrito de maneira autoral utilizando os autovetores como informação prévia para a aplicação do método aglomerativo.

## 2. Algoritmo Espectral com K-médias

Considere os elementos de um conjunto de dados como os vértices de um grafo não dirigido  $G(V, E)$ . O peso de cada aresta representa a similaridade entre os elementos de acordo com alguma medida [Pentney and Meila 2005]. O objetivo geral é particionar o grafo em dois ou mais conjuntos disjuntos, removendo as suas arestas. Para tal tarefa, a ideia proposta é o mapeamento dos dados originais para os primeiros  $k$ -autovetores da matriz Laplaciana obtida em função do grafo de representação dos dados, em seguida, aplica-se um algoritmo de agrupamento padrão, como o  $k$ -médias sobre estas novas coordenadas [Saerens 2004].

De acordo com [N. Abreu and Vinagre 2013] o segundo menor autovalor do Laplaciano de um grafo  $G$ ,  $\mu_{n-1}$ , é chamado conectividade algébrica do grafo  $G$ , sendo assim, diz-se também que um grafo é conexo se, e somente se, o seu segundo menor autovalor Laplaciano é positivo. Desse modo, o agrupamento feito pelo  $k$ -médias será elaborado de acordo com o segundo menor autovetor da matriz de autovetores, ou sobre os primeiros  $k$  autovetores. Assim, um conjunto de dados com dimensão  $n$  é reduzido para uma dimensão menor na qual será feito o agrupamento, reduzindo o número de iterações do  $k$ -médias se comparado à sua aplicação no conjunto original.

O parâmetro de escala  $\sigma^2$  (desvio padrão) no Algoritmo (1) controla a rapidez com que a matriz de similaridade  $W$  decresce levando em consideração a distância entre  $x_i$  e

---

**Algoritmo 1:** Agrupamento Espectral via K-médias  
[A. Ng and Y.Weiss 2001]

---

**Entrada:** Conjunto de pontos  $X = \{x_1, \dots, x_n\}$

1. Forme a matriz de similaridade  $W$  definida por:  $W_{ij} = e\left(-\frac{1}{2\sigma^2}d^2(x_i, x_j)\right)$ .

2. Construa a matriz Laplaciana normalizada simétrica:

$$L = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}.$$

3. Encontre os  $k$  autovetores de  $L$  (escolhidos para serem ortogonais entre si no caso de autovalores repetidos), e forme matriz  $U$  colocando os autovetores

em colunas:  $U \begin{bmatrix} u_1 & \dots & u_k \end{bmatrix} \in \mathbb{R}^{n \times k}$

4. Forme a matriz  $Y$  a partir de  $U$  normalizando cada linha de  $U$  para ter valores unitários.

5. Considere cada linha de  $Y$  como um ponto em  $\mathbb{R}^k$  e classifique-os em  $k$  grupos via k-médias.  $Y_{ij} = \frac{U_{ij}}{\left[\sum_{j=1}^n U_{ij}^2\right]^{\frac{1}{2}}}$ .

6. Coloque os pontos originais  $x_i$  no grupo  $j$ , se e somente se, a linha  $i$  da matriz  $Y$  for colocada no grupo  $j$ .

---

$x_j$  [A. Ng and Y.Weiss 2001]. A matriz  $D$  é obtida calculando-se o peso de cada aresta incidente em um determinado vértice, desse modo, forma-se uma matriz diagonal, a qual é utilizada na obtenção da matriz Laplaciana.

### 3. Algoritmo Espectral Aglomerativo

A proposta de elaboração de um algoritmo de agrupamento espectral sem a etapa de mapeamento dos autovetores utilizando o k-médias é motivado pela ineficácia do método em conjuntos de dados de formato não esférico. Apesar da convergência do k-médias ocorrer em poucas iterações, uma má inicialização dos centroides é um problema que afeta o resultado do agrupamento. Isto ocorre porque os centroides são inicializados aleatoriamente, por exemplo, se existem  $k$  grupos reais, então a probabilidade de selecionar um centroide para cada grupo é relativamente pequena [P.N. Tan and Kumar 2009]. Além disso, esta metodologia é sensível ao parâmetro de vizinhança estabelecido na construção do grafo de representação dos dados, ver Seção 6. Desse modo, uma abordagem por aglomeração na etapa de mapeamento dos autovetores pode se tornar um método alternativo para contornar os problemas existentes com a abordagem do k-médias.

A busca pelos número de grupos que acontece nos autovetores é feita de maneira aglomerativa dentro desse novo conjunto de informações. A informação que os autovetores carregam do conjunto de dados é utilizada como um rótulo de agrupamento, ou seja, cada elemento de um autovetor possui um número associado a ele, que é utilizado no processo de aglomeração. Uma matriz de distâncias é determinada de acordo com as coordenadas dos autovetores, e logo após é determinada uma medida de relação entre estas informações, por exemplo, a média das distâncias. São inicializados três vetores contendo o número inicial de rótulos de acordo com o tamanho da dimensão dos autovetores, o vetor de distância média entre os vizinhos calculado de acordo com a média das distâncias e o vetor número de vizinhos inicializado com todas posições iguais

---

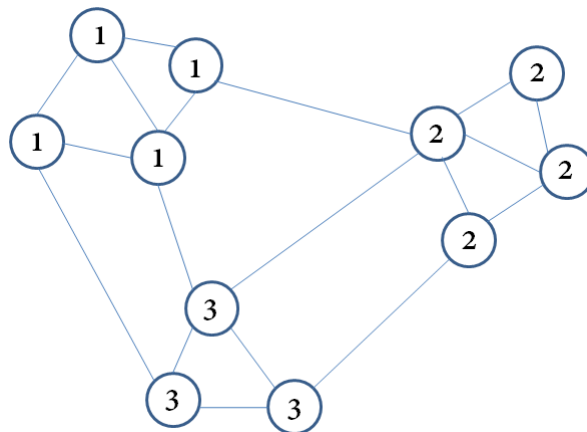
**Algoritmo 2:** Passo aglomerativo

---

**Entrada:** k-menores autovetores: Matriz [U]

1. Cálculo da matriz de distância dos elementos de  $D = dist(u_i, u_j)$ ;
  2. Cálculo da média das distâncias da matriz U, ou seja,  $\bar{u} = \sum_{i=1}^n \frac{1}{n} \cdot u_i$ ;
  3. Inicialização dos rótulos, números de vizinhos e distância entre os vizinhos;
  4. Atualiza os itens do passo 3 de acordo com a distância média entre os grupos(rótulos);
  5. Para se o número de vizinhos de um grupo for maior que o número de vizinhos no conjunto de dados, caso contrário, volta ao passo 4.
- 

a um. São selecionadas duas posições aleatórias do conjunto dos autovetores, onde cada posição passará por um teste de distância no grupo de acordo com o seu rótulo, o número de vizinhos próximos e distância entre a outra posição aleatória. Se o rótulo da posição "a" for diferente de zero e o número de vizinhos multiplicados pela distância entre as posições "a" e "b" forem menores que a distância dos vizinhos de "a", então o rótulo da posição "b" torna-se o mesmo da posição "a", o número de vizinhos de "b" incrementa em uma unidade assim como a distância entre os vizinhos de "b" incrementa de acordo com a distância entre "a" e "b". Como critério de parada do processo de aglomeração é utilizada uma comparação entre o número de vizinhos da posição "b" e o número de vizinhos do conjunto de dados atual (autovetores). Se a quantidade de vizinhos da posição "b" extrapolar o número de vizinhos do conjunto de dados, então o processo de aglomeração naquela posição termina.



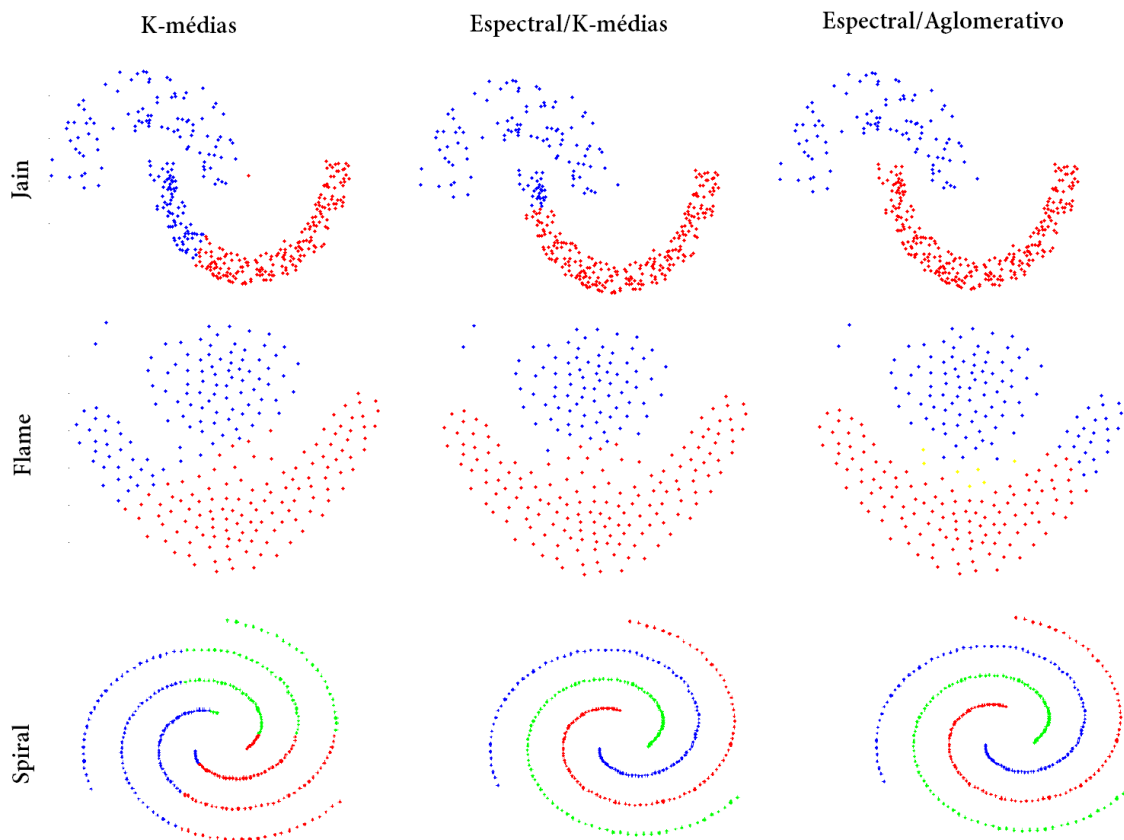
**Figura 1. Grupos rotulados**

Após o processo de aglomeração cada elemento de um determinado grupo possui um rótulo, que é a sua característica em relação a todo o conjunto. Na Figura 1 é apresentado um exemplo no grafo da distribuição dos rótulos em três grupos distintos.

#### **4. Experimentos**

Para a aplicação dos métodos de agrupamento discutidos aqui, foram escolhidos alguns conjuntos de dados disponíveis em [Fu and Medico 2007], [Jain and Law 2005] e

[Chang and Yeung 2008]. Houve a preocupação em trabalhar com conjuntos de dados de diferentes formatos geométricos, sendo específicos para a finalidade de analisar os resultados de cada algoritmo aplicado. Em cada teste é previamente informado pelo usuário a quantidade de grupos a serem formados, uma tabela contendo as coordenadas  $x$  e  $y$  de cada elemento do conjunto e o número fixo de vizinhos mais próximos para a construção do grafo.



**Figura 2. Resultados dos Experimentos**

No primeiro teste foi utilizado o conjunto de dados [Jain and Law 2005] que possui o formato de duas meia luas conforme a Figura 2. Pode-se inferir visualmente que o conjunto possui dois grupos bem definidos e bem separados no plano bidimensional, ou seja, o número de grupos que os algoritmos irão estabelecer será de  $k = 2$ . Conforme a visualização da Figura 2 é possível notar que o algoritmo k-médias não obteve um agrupamento coerente de acordo com a estrutura dos dados. Isto se deve ao fato de que a utilização de centróides como parâmetro de agrupamento não funciona de maneira eficaz em conjuntos de dados que possuem uma determinada curvatura. Como a distribuição dos dados no plano não acompanha uma distribuição Gaussiana, então o uso do k-médias neste caso não é o mais recomendado. Já os grupos formados pelo algoritmo espectral via k-médias gerou um resultado melhor que o seu antecessor. Apesar disso, a informação proveniente da matriz Laplaciana não foi suficiente para estabelecer um agrupamento coerente com a geometria do conjunto, isto ocorre porque os autovetores desta matriz não possuem formato hiper-esférico. Pode-se notar que os resultados do último algoritmo de agrupamento foram satisfatórios, respeitando a estrutura organizacional dos dados. A

explicação para este fato é decorrente de que o método utilizado não é baseado em um conjunto de centroides, mas em uma relação entre os elementos de acordo com a sua média dentro do conjunto de autovetores.

No segundo experimento foi utilizado o conjunto de dados [Fu and Medico 2007], que por sua vez possui características geométricas interessantes, composta por uma distribuição esférica de elementos e outra concentração no formato de meia lua conforme Figura 2. No agrupamento estabelecido pela metodologia k-médias, os grupos se formaram de maneira incorreta, agrupando parte da meia lua juntamente com a concentração superior acima dela. Isto ocorre devido a separação de forma linear dos grupos que o algoritmo utilizado obtém. No segundo algoritmo o resultado do agrupamento foi obtido com sucesso, pois visualmente os grupos formaram-se corretamente respeitando a distribuição dos elementos no plano. É interessante ressaltar que o desvio padrão dos dados é moderadamente diferente entre os dois grupos, o que pode dificultar a resposta de vários algoritmos baseados neste parâmetro. No processo de aglomeração dos autovetores o resultado obtido foi satisfatório, enfatizando em boa parte a distinção entre o grupo de formato esférico e a meia lua.

Por fim, é utilizado o conjunto de dados [Chang and Yeung 2008] que também possui um formato de curva, neste caso o número de grupos existentes no conjunto é de  $k = 3$ , conforme é possível analisar na Figura 2. Novamente o número de grupos informado como entrada em ambos os algoritmos é o mesmo, desse modo, será possível estabelecer uma conclusão em relação ao resultado obtido por cada um deles. No experimento com o uso do k-médias o resultado do agrupamento não foi satisfatório, novamente o problema ocorre devido a curvatura da distribuição dos elementos no plano. Diferente de muitos casos, neste conjunto nem uma boa inicialização do centroide irá estabelecer corretamente os grupos, sendo assim, o algoritmo é ineficaz. Em contrapartida, os algoritmos baseados na metodologia de particionamento espectral apresentaram bons resultados. Pode-se notar na Figura 2 o agrupamento nos autovetores ocorreu de maneira esperada, onde o mapeamento no conjunto de dados foi o melhor possível. Neste teste não houve diferença entre o método espectral via k-médias e o método espectral aglomerativo, pois a estrutura organizacional dos autovetores estava bem definida, ou seja, estava de acordo com uma distribuição Gaussiana, o que difere do conjunto de dados anterior.

#### 4.1. Medida de Qualidade de Agrupamento

Com o intuito de medir a performance dos algoritmos utilizados em cada conjunto de dados foi utilizado a medida-F. Dada uma predição  $h(x) = (h_1(x), \dots, h_m(x)) \in Y$  de um vetor de rótulo binário  $y = (y_1, \dots, y_m)$ , a medida-F é definida a seguir:

$$F(y, h(x)) = \frac{(1 + \beta^2) \sum_{i=1}^m y_i h_i(x)}{\beta^2 \sum_{i=1}^m y_i + \sum_{i=1}^m h_i(x)} \in [0, 1], \quad (1)$$

onde  $\frac{0}{0} = 1$ . Esta medida essencialmente corresponde a uma média harmônica ponderada de precisão e exaustividade [K. Dembczynski and Waegeman 2013]. Para utilizar esta medida foram considerados os resultados corretos dos agrupamentos encontrados na literatura. Para cada teste foram definidas as variáveis *TrueCluster* e *PredictedCluster*, ou

seja, a informação dos grupos corretos e dos grupos formados pelos algoritmos respectivamente.

**Tabela 1. Resultados da Medida-F**

Resultados da Medida-F			
Conjunto de dados	K-médias	Espectral/K-médias	Espectral/Aglomerativo
<i>Jain</i>	0.8288	0.9446	1
<i>Flame</i>	0.806	0.9825	0.8731
<i>Spiral</i>	0.2695	1	1

O principal objetivo de utilizar a medida-F é relacionar os resultados obtidos pelo agrupamento espectral aglomerativo com um parâmetro estatístico. Os resultados de performance obtidos pelos outros dois algoritmos servem apenas para comparação, enfatizando em qual situação são mais eficazes ou não.

A aplicação da métrica referente ao conjunto de dados *Jain* mostrou que o melhor desempenho foi obtido pelo algoritmo aglomerativo, obtendo o valor máximo da medida conforme a Tabela(1). Os resultados obtidos pelo outros dois algoritmos foram razoáveis, ficando com valores acima de 0.8 da medida. Em relação ao conjunto de dados *Flame* nenhum dos algoritmos alcançou 100%, porém o melhor resultado foi obtido pelo algoritmo espectral. Finalmente, no último conjunto de dados tanto o método espectral via k-médias quanto o aglomerativo obtiveram o valor máximo, em contrapartida, o método k-médias ficou com valor abaixo de 0.3, o qual foi o pior resultado obtido.

## 5. Complexidade de Computacional

Para estudar a complexidade de tempo do algoritmo espectral aglomerativo é necessário apenas considerar as medidas de complexidade obtidas pelo Algoritmo (1) até a etapa de obtenção dos autovetores da matriz Laplaciana e soma-las às medidas obtidas pela etapa de aglomeração.

Sabe-se da análise feita anteriormente que o processo que exige mais tempo computacional é dado pelo cálculo dos autovetores, gerando uma complexidade de tempo na ordem de  $O(n^3)$ . Após esta etapa ocorre a aplicação do algoritmo de aglomeração aplicado ao conjunto de autovetores de tamanho  $n \times k$ , onde  $n$  é a quantidade de elementos do conjunto original de dados e  $k$  é o número de menores autovetores da matriz Laplaciana. Quando o agrupamento ocorre apenas no segundo menor autovetor (Vetor de Fidler) o conjunto de autovetores irá medir  $n \times 1$ .

A complexidade de tempo do agrupamento aglomerativo é:

$$O(n^2 \log n). \quad (2)$$

Primeiro calcula-se todas as distâncias  $n^2$  para os grupos iniciais, e aglomera-se os elementos em grupos (tempo:  $O(n^2 \log n)$ ). Em cada uma das  $O(n)$  iterações, identifica-se o par de grupos com a coesão mais alta em  $O(n)$ ; Agrupa-se o par; e atualiza-se as médias dos grupos. Para cada grupo, também atualiza-se a lista ordenada de elementos, excluindo os dois grupos já aglomerados. Assim, cada iteração toma  $O(n \log n)$ . A complexidade do tempo geral é então  $O(n^2 \log n)$ .

Estudando o número de operações executadas no passo de aglomeração conclui-se que no geral a complexidade do método ainda continua sendo  $O(n^3)$ , devido ao esforço computacional exigido no cálculo dos autovetores de matriz Laplaciana. Em geral, para aplicação do método em grandes bancos de dados o algoritmo necessitaria de uma otimização nas etapas que possuem alto custo de execução.

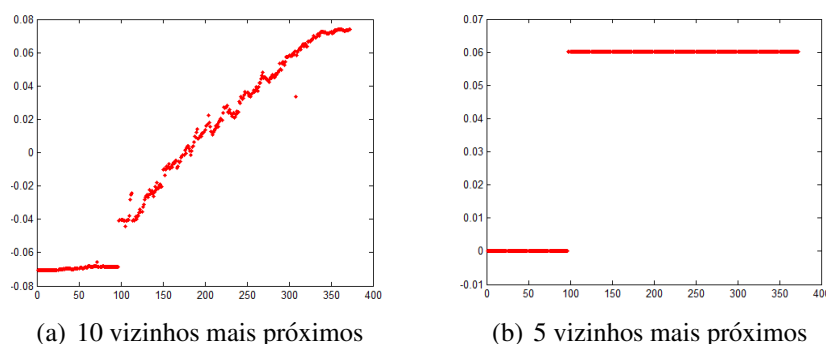
## 6. Discussão

No método espectral o conjunto de dados pode ser visualizado na forma de um grafo, onde cada elemento do conjunto pode ser representado pelo vértice e a sua similaridade entre os demais por arestas incidentes nele. Cada valor de uma aresta entre um vértice e outro pode ser calculado por uma função de similaridade, assim, é possível obter um grafo que represente todo o conjunto de dados. Porém, existem alguns métodos que estabelecem como este grafo pode ser construído. O mais geral é o grafo completo, que estabelece uma relação de similaridade entre todos os indivíduos do conjunto gerando um grande número de arestas, o que pode deixar grande a matriz de distâncias a ser calculada. Uma maneira alternativa é construir o grafo utilizando um raio de similaridade entre um elemento em relação ao demais, desse modo, é possível estudar a semelhança entre um conjunto de elementos menor do que o conjunto global. Esta abordagem traz algumas dificuldades de interpretação quando utilizada para representar conjuntos de dados com formato de curva, um exemplo seria o conjunto de dados *Spiral* visto anteriormente. Outra possibilidade a ser considerada é de construir o grafo a partir de um elemento e seus  $k$ -vizinhos mais próximos, assim como a abordagem anterior esta possibilidade ajuda a estudar localmente a similaridade em um conjunto menor do que o global.

Quando utilizada a abordagem de construção do grafo usando o método dos  $k$ -vizinhos mais próximos ocorre a possibilidade de alternar o número de vizinhos de um elemento. Com esta alternância, o algoritmo espectral pode ser aplicado ao conjunto de dados com a possibilidade de estudar a similaridade localmente ou expandir sua visão dentro do conjunto. Consequentemente, o resultado do agrupamento será sensível a este fato, podendo gerar diferentes grupos para determinados valores de  $k$ . Deste modo, o número de  $k$ -vizinhos está diretamente relacionado ao conjunto dos  $k$ -menores autovetores da matriz Laplaciana, onde ocorre a execução de outro algoritmo de agrupamento, no caso geral, do  $k$ -médias.

O que muitas vezes pode gerar um resultado ruim no agrupamento é quando os  $k$ -menores autovetores possuem uma curvatura, nesse caso, aplicação do  $k$ -médias não é recomendável. No experimento realizado com o conjunto de dados *Jain* tomou-se os dez vizinhos mais próximos para a construção do grafo de similaridade, gerando como resultado intermediário a representação dos autovetores conforme na Figura 4(a). Nota-se que os autovetores não tem um comportamento linear, assim, o  $k$ -médias não irá obter bom resultado quando aplicado neste conjunto. Já na Figura 4(b) pode-se inferir que existem dois grupos bem definidos, isto porque o número de vizinhança mais próxima estabelecida foi de cinco vizinhos. Isto mostra a sensibilidade do método espectral em relação a vizinhança de elementos de um conjunto, ficando a trabalho do usuário acertar uma vizinhança correta para a construção do grafo. Todavia, o método de agrupamento por aglomeração torna-se uma opção para contornar o problema com a geometria dos autovetores, obtendo uma sensibilidade menor em relação a vizinhança do que o método proposto por [A. Ng and Y.Weiss 2001].





**Figura 3. k menores autovetores do conjunto de dados Jain**

Assim, uma das vantagens do método por aglomeração é a menor sensibilidade ao número de  $k$ -vizinhos escolhidos para montar o grafo de similaridade. Conforme o crescimento do número de vizinhos o método por aglomeração também pode variar os resultados do agrupamento, visto que a dependência da geometria do conjunto de dados influencia muito neste contexto.

## 7. Conclusão e Trabalhos Futuros

Neste trabalho foi proposto um algoritmo de aglomeração nos  $k$  menores autovetores da matriz Laplaciana obtida pelo método espectral para agrupamento de dados. Foram apresentados experimentos com dados disponíveis na literatura para aplicação dos algoritmos espectrais e  $k$ -médias. Os resultados obtidos foram comparados através da medida-F. Pode-se notar que o algoritmo proposto obteve bons resultados, mostrando-se promissor para futuras aplicações. Para trabalhos futuros pretende-se aplicar o algoritmo a outros conjuntos de dados e estudar sua performance de acordo com a sua geometria. Além disso, é interessante analisar os resultados de agrupamento em conjuntos de formato hiper-esférico, no qual o  $k$ -médias é o algoritmo mais recomendado.

## Referências

- A. Ng, M. J. and Y. Weiss (2001). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems 14*.
- Chang, H. and Yeung, D. (2008). Robust path-based spectral clustering. In *Pattern Recognition*, pages 191–203.
- Chen, X. and Cai, D. (2011). Large scale spectral clustering with landmark-based representation. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*.
- D. Yan, L. H. and Jordan, M. (2009). Fast approximate spectral clustering. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 907–916.
- Fu, L. and Medico, E. (2007). Flame, a novel fuzzy clustering method for the analysis of dna microarray data. In *BMC bioinformatics*.
- Jain, A. and Law, M. (2005). Data clustering: A user's dilemma. In *Lecture Notes in Computer Science*, pages 1–10.

- K. Dembczynski, A. Jachnik, W. K. I. and Waegeman, W. (2013). Optimizing the f-measure in multi-label classification: Plug-in rule approach versus structured loss minimization. In *International Conference on Machine Learning*, pages 1130–1138.
- L. Ertoz, M. S. and Kumar, V. (2003). Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data. In *SDM*, pages 47–58.
- L. Wang, C. Leckie, K. R. and Bezdek, J. (2009). Approximate spectral clustering. In *Advances in Knowledge Discovery and Data Mining*, pages 134–146.
- Manor, L. Z. and Perona, P. (2004). Self-tuning spectral clustering. In *Advances in Neural Information Processing Systems (NIPS17)*.
- N. Abreu, R. Del-Vecchio, V. T. and Vinagre, C. (2013). Teoria espectral de grafos-uma introducao.
- N. Tremblay, G. Puy, R. G. and Vandergheynst, P. (2016). Compressive spectral clustering. In *arXiv preprint arXiv:1602.02018*.
- Nadler, B. and Galun, M. (2006). Fundamental limitations of spectral clustering. In *Advances in neural information processing systems*, pages 1017–1024.
- P. Yang, Q. Z. and Huang, B. (2011). Spectral clustering with density sensitive similarity function. In *Knowledge-Based Systems*, pages 621–628.
- Pentney, W. and Meila, M. (2005). Spectral clustering of biological sequence data. In *Association for the Advancement of Artificial Intelligence*, pages 845–850.
- P.N. Tan, M. S. and Kumar, V. (2009). Introduction to data mining. In *Ciencia Moderna*.
- Saerens, M. (2004). The principal components analysis of a graph, and its relationships to spectral clustering. In *European Conference on Machine Learning*, pages 371–383.
- Sakai, T. and Imiya (2009). Fast spectral clustering with random projection and sampling. In *Machine Learning and Data Mining in Pattern Recognition*, pages 372–384.