

# Using Graph Spectral to solve Change Point Detection Problems

Luis Gustavo C. Uzai<sup>1</sup>, André Y. Kashiwabara<sup>1</sup>

<sup>1</sup>Universidade Tecnológica Federal do Paraná (UTFPR)  
Av. Alberto Carazzai, 1.640 – 86.300-000 – Cornélio Procópio – PR – Brasil

uzai@alunos.utfpr.edu.br, kashiwabara@utfpr.edu.br

**Abstract.** *Time series are sequence of values distributed over time. Analyzing time series is important in many areas including medical, financial, aerospace, commercial and entertainment. Change Point Detection is the problem of identifying changes in meaning or distribution of data in a time series. This article presents Spec, a new algorithm that uses the graph spectrum to detect change points. The Spec was evaluated using the UCR Archive which is a large database of different time series. Spec performance was compared to the PELT, ECP, EDM, and gSeg algorithms. The results showed that Spec achieved a better accuracy compared to the state of the art in some specific scenarios and as efficient as in most cases evaluated.*

**Resumo.** *Séries temporais são sequência de valores distribuídos ao longo do tempo. Analisar séries temporais é importante em várias áreas, incluindo médica, financeira, aeroespacial, comercial e entretenimento. Detecção de Pontos de Mudança é o problema de identificar mudanças no significado ou na distribuição de dados em uma série temporal. Este artigo apresenta o Spec, um novo algoritmo que utiliza o espectro de grafo para detectar pontos de mudança. O Spec foi avaliado utilizando o UCR Archive que é uma grande base de dados de diferentes séries temporais. A performance do Spec foi comparado com os algoritmos PELT, ECP, EDM, e gSeg. Os resultados mostraram que o Spec alcançou uma exatidão melhor em comparação ao estado da arte em alguns cenários específicos e tão eficiente quanto na maioria dos casos avaliados.*

## 1. Introdução

A análise de séries temporais é importante em muitas áreas, incluindo médica, financeira, aeroespacial, empresarial e meteorológica [Aminikhanghahi and Cook 2016]. Séries temporais são um conjunto de valores dispersos ao longo do tempo, que descrevem um comportamento particular de um sistema dinâmico. Em pontos de tempo arbitrários uma série temporal pode alterar seu comportamento em relação ao restante da série já observada, esses pontos são chamados de *Change Points* (Pontos de Mudança).

*Change Point Detection* (Detecção de Ponto de Mudança/CPD) é o problema em identificar a mudança na distribuição ou significado dos dados em uma série temporal [Basseville and Nikiforov 1993] [Aminikhanghahi and Cook 2016]. Com o passar dos anos, diversas técnicas para detecção de pontos de mudança foram desenvolvidas [Page 1955], entretanto, com o surgimento de diferentes maneiras de capturar uma grande

quantidade de dados o interesse por esse problema vem aumentando consideravelmente [Basseville and Nikiforov 1993, Li et al. 2015].

Um desafio na detecção de pontos de mudança é no cenário de *Big Data*, visto que, frequentemente os dados possuem anomalias e a maioria dos algoritmos atuais não são resilientes a anomalias [James et al. 2016]. Anomalias são observações fora do padrão dos dados, como rúidos ou valores irregulares a distribuição da série que dificultam a identificação de padrões estatísticos da série temporal [James et al. 2016]. Outro problema é que frequentemente os dados não possuem uma distribuição normal, o que implica na limitação da aplicação de diversos algoritmos que mostram eficiência em contextos paramétricos [Killick et al. 2012].

Existem várias abordagens para detecção de pontos de mudança de categorias distintas. Cada método possui estratégias diferentes para contornar os problemas descritos anteriormente e detectar pontos de mudança com maior precisão [Aminikhanghahi and Cook 2016]. A seleção de um método deve considerar primordialmente a natureza dos dados analisados. Algoritmos produzidos na última década tem como objetivo principal a detecção de múltiplos pontos de mudança em grandes conjuntos de dados em baixo tempo [Killick et al. 2012] [Matteson and James 2014] [James et al. 2016]. Entretanto, conjuntos de dados de alta dimensionalidade são problemáticos para a maioria dos algoritmos atuais, já que, a medida que aumenta a dimensionalidade a precisão dos algoritmos diminui e o tempo para realizar o cálculo aumenta consideravelmente [Chen and Zhang 2015].

A detecção de pontos de mudança utilizando grafos foi proposto por Chen [Chen and Zhang 2015] com o propósito de detectar assertivamente pontos de mudança em bases de dados de alta dimensionalidade sem grande perda de performance. O método *gSeg* proposto por Chen apresentou a mesma assertividade que o estado da arte em bases de baixa dimensionalidade, contudo, obteve melhor assertividade que o estado da arte em bases de média para alta dimensionalidade. O método de Chen não explora a topologia dos grafos gerados a partir das séries temporais, este trabalho apresenta a hipótese de que é possível utilizar teorias fundamentais dos grafos para recolher maiores informações da série temporal e conseguir detectar pontos de mudança de forma mais assertiva sendo mais resiliente a anomalias apresentadas na série temporal.

Este trabalho faz uso da teoria espectral do grafo com a finalidade de obter o conhecimento da topologia dos grafos gerados a partir das séries temporais multivaloradas. A teoria espectral do grafo é descrita pelo estudo dos autovalores e autovetores de matrizes associadas a grafos [Spielman 2007]. O método criado nesse trabalho *Spec* (abreviação de Graph Spectrum) utiliza o espectro do grafo a fim de separar os nós de forma que o ponto limite das separações seja o ponto de mudança. O uso do espectro do grafo para agrupamento e separação de dados se mostrou mais eficiente que os métodos tradicionais em diversos cenários [Ng et al. 2002], visto que, métodos tradicionais como o K-means tendem a falhar quando os grupos não correspondem a regiões convexas. Apesar do uso do espectro do grafo para agrupamento ser comum, no melhor do nosso conhecimento, não existem trabalhos descrevendo o uso do espectro dos grafos para detecção de pontos de mudança.

A eficiência do método proposto foi comparada com o estado da arte na detec-

ção de pontos de mudança, PELT, ECP, EDM e gSeg o qual obteve maior precisão em diferentes cenários, como verificado na base de dados *Meat* onde o resultado máximo atingido pelos outros algoritmos é de 0.64 e o *Spec* obteve resultado médio de 0.80. Além disso, atingiu a mesma média de precisão (0.60) quando comparado aos outros métodos. Este trabalho demonstra que o espectro do grafo é uma alternativa válida para detecção de múltiplos pontos de mudança.

## 2. Trabalhos Relacionados

Detecção de pontos de mudança tem diversas aplicações, conforme descrito abaixo:

**Detecção e diagnóstico de falhas:** Detectar e diagnosticar falhas com antecedência pode ser de grande vantagem principalmente quando considerado sistemas dinâmicos, pois, pode gerar ganho no aumento de confiabilidade do equipamento, assim como, redução de risco de paradas não programadas de produção reduzindo perdas materiais e acidentes de trabalho [Moreira 2011].

**Controle de qualidade:** uma das primeiras aplicações envolvendo pontos de mudança é no controle de qualidade e monitoramento da produção [Basseville and Nikiforov 1993]. É possível exemplificar um uso hipotético de uma fábrica de bolachas e uma câmera que analisa a cor das bolachas recém assadas em uma esteira. Naturalmente, algumas bolachas estarão mais escuras e outras mais claras, porém, caso a média das bolachas esteja mais escura que o padrão anteriormente registrado (ponto de mudança encontrado) é possível que o forno precise diminuir a temperatura. O inverso também é verdadeiro caso a média das bolachas esteja mais clara o forno possivelmente precisará aumentar a temperatura. O exemplo hipotético mencionado acima pode ser generalizado para uma série de cenários, por muitas razões é compreensível que atributos de determinado produto estejam dentro de padrões com desvios mínimos para a garantia da qualidade [Lai 2014].

**Previsão climática:** Nas últimas décadas cada vez mais informação climática vem sendo armazenada tornando-se necessária técnicas para minerar e entender os dados [Itoh and Kurths 2010]. Assim, pontos de mudança são utilizados para conhecer as alterações de padrões em séries temporais de informação climática e, categorizar motivos pelos quais essas mudanças ocorrem [Yonetani and McCabe 1994] [Itoh and Kurths 2010]. Monitorar dados climáticos é de grande interesse acadêmico já que, uma grande diversidade de eventos climáticos possuem relação com as mudanças de temperatura constatadas nas últimas décadas [Bonsal et al. 2001]. Nesse sentido, diversas técnicas de detecção de pontos mudanças são aplicadas para identificar oscilações de padrões climáticos, inclusive, são comparados entre si com a finalidade de indicar o melhor algoritmo para cada cenário. [Reeves et al. 2007] [Itoh and Kurths 2010] [Ducré-Robitaille et al. 2003].

**Monitoramento médico:** Para garantir a saúde de um paciente frequentemente é necessário o monitoramento de uma série de sinais vitais que indicam a evolução de um quadro clínico. Os sinais vitais são variáveis fisiológicas como: frequência cardíaca, eletrocardiograma, eletroencefalograma e entre outros, podendo ser transpostos em séries temporais [Aminikhangahi and Cook 2016]. O monitoramento médico necessita de métodos que fornecem suporte rápido ao detectar mudanças sendo aplicado em casos críticos como medição cardíaca de crianças anestesiadas [Yang et al. 2006]. E também, em análise mais longas e detalhadas, por exemplo, ao verificar a variação cardíaca durante o sono

é necessárias horas de coletas de dados para que seja possível obter resultados válidos [Malladi et al. 2013] e entre outras aplicações [Staudacher et al. 2005] [Bosc et al. 2003].

## 2.1. Métodos comparados

Nessa subseção serão descritos os métodos que foram comparados com o novo método proposto sendo detalhado sua natureza, contexto de aplicação e vulnerabilidades. Todos os algoritmos apresentados são open-source e podem ser adquiridos no repositório CRAN ou GitHub, a linguagem utilizada em todos os algoritmos é primordialmente R (alguns métodos internos foram implementados em C++ para ganhar performance). O fato desses pacotes ainda estarem em evolução levanta o risco dos métodos sejam alterados após a data de publicação deste trabalho e conseqüentemente, o tempo computacional ou vulnerabilidades variem.

O algoritmo **PELT**, abreviação de Pruned Exact Linear Time, da categoria de razão de verossimilhança e de complexidade  $\mathcal{O}(n^2)$  no entanto, em caso de uma penalidade linear  $f(k) = k$ , a complexidade do algoritmo também será linear  $\mathcal{O}(n)$ . Proposto por [Killick et al. 2012] como uma solução para o problema de identificar múltiplos pontos de mudança em grandes conjuntos de dados offline, porém, pode ser facilmente usado no contexto de detecção de pontos de mudança online graças à sua alta precisão, como demonstrado em [James et al. 2016].

O algoritmo E-Divisive **ECP** foi projetado para detectar pontos de mudança online [Matteson and James 2014] sendo não-paramétrico e aplicável a quase todas as séries temporais. O *ECP*, possui a finalidade de detectar múltiplos pontos de mudança sem a necessidade de informar a total quantidade ou aproximação de pontos na série temporal, bem como, a possibilidade de trabalhar com dados de múltiplos valores o que geralmente é problemático para a maioria dos algoritmos.

E-Divisive utiliza uma nova metodologia baseada em *U-statistics cluster based* (baseado em agrupamento), o algoritmo gera uma matriz contendo a distância entre todas as observações e divide a série temporal em dois grupos cada um contendo a maior distância entre eles, este processo se repete até que todas as séries temporais tenham sido percorridas, logo, a complexidade do método é  $\mathcal{O}(n^2)$ .

O algoritmo **EDM**, abreviatura de E-Divisive with Medians da categoria *cluster based*, propõe uma melhoria no algoritmo E-Divisive pela técnica estatística de *Moving Median* [James et al. 2016]. O algoritmo foi desenvolvido pela equipe do Twitter para ser aplicado em suas bases em produção na nuvem, pois, o número alto de anomalias no cenário é frequente (cloud big data), portanto, trata-se de um cenário de detecção online, onde, o objetivo é identificar o ponto de mudança real o mais cedo possível, algoritmos desenvolvidos até então não possuíam assertividade suficiente para serem aplicados nesse contexto.

O algoritmo **gSeg**, abreviação de Graph Segmentation, da categoria *Graph Based* (Baseado em Grafos), propõe a estimativa de localização de pontos de mudança utilizando estatísticas de varredura em uma sequência de dados. A varredura faz uso de grafos que representam a verossimilhança entre cada observação da série [Chen and Zhang 2015].

### 3. Fundamentação Teórica

O cálculo para obtenção dos valores do espectro de um grafo se dá pela matriz de adjacência. Em um grafo não direcionado, simples e sem pesos nas arestas é representado por uma matriz de zeros e uns que se constrói naturalmente a partir das relações de adjacência entre os vértices do grafo. Nomeando uma matriz adjacente de um grafo qualquer como  $A$ , consideramos  $a_{ij}$  a relação entre os vértices, sendo 1 caso  $v_i$  e  $v_j$  sejam adjacentes e 0 caso contrário [de Abreu et al. 2014]. A matriz de adjacência é uma forma de representação do grafo. O espectro do grafo são os conjuntos dos autovalores da matriz de adjacência [de Abreu et al. 2014] [Fritscher 2011].

Com a investigação das propriedades espectrais dos gráficos, foram introduzidos métodos estatísticos para a seleção do modelo, estimativa de parâmetros e testes de hipóteses para discriminar amostras, a identificação dos conjuntos de correlação e o fluxo de informação entre gráficos de dados genéticos [Vidal et al. 2017] e neuroimagens [Fujita et al. 2017].

#### 3.1. Agrupamento Espectral

O objetivo do agrupamento espectral é agrupar dados que estão conectados, mas, não necessariamente agrupados em limites convexos. Para realizar o agrupamento espectral assim como no agrupamento comum, é necessário uma medida de similaridade ou afinidade  $s(x, y)$  para determinar o quão próximo os pontos  $x$  e  $y$  são um do outro [Luxburg 2007].

Denotando a matriz de similaridade de um grafo qualquer como  $S$ , sendo que  $S_{ij} = s(x_i, x_j)$  dada a semelhança entre as observações  $x_i$  e  $x_j$ . Usando uma medida de similaridade dos pontos como a Gaussiana, que, quanto mais próximo os valores mais tendem a 1 e quanto mais distantes mais tendem a 0, computando a matriz de similaridade  $S$  e, gerando uma matriz baseado na afinidade de  $S$  denominada  $A$ , sendo  $A$  positiva e simétrica [Neto 2013].

Possuindo a matriz de afinidade, o agrupamento é substituído por um problema de partição de grafo, onde componentes conectados do grafo são interpretados como *cluster* (grupos). O grafo deve ser particionado de tal forma que as arestas que conectam diferentes grupos devem ter pesos baixos e as bordas dentro do mesmo grupo deve ter valores altos.

A próxima etapa é computar o grafo laplaciano, que é o resultado da subtração de uma matriz de grau onde cada valor diagonal é o grau do vértice respectivo e todas as outras posições são zero da matriz de afinidade anteriormente calculada, existem variantes da matriz laplaciana que podem ser aplicadas nesse caso [Luxburg 2007].

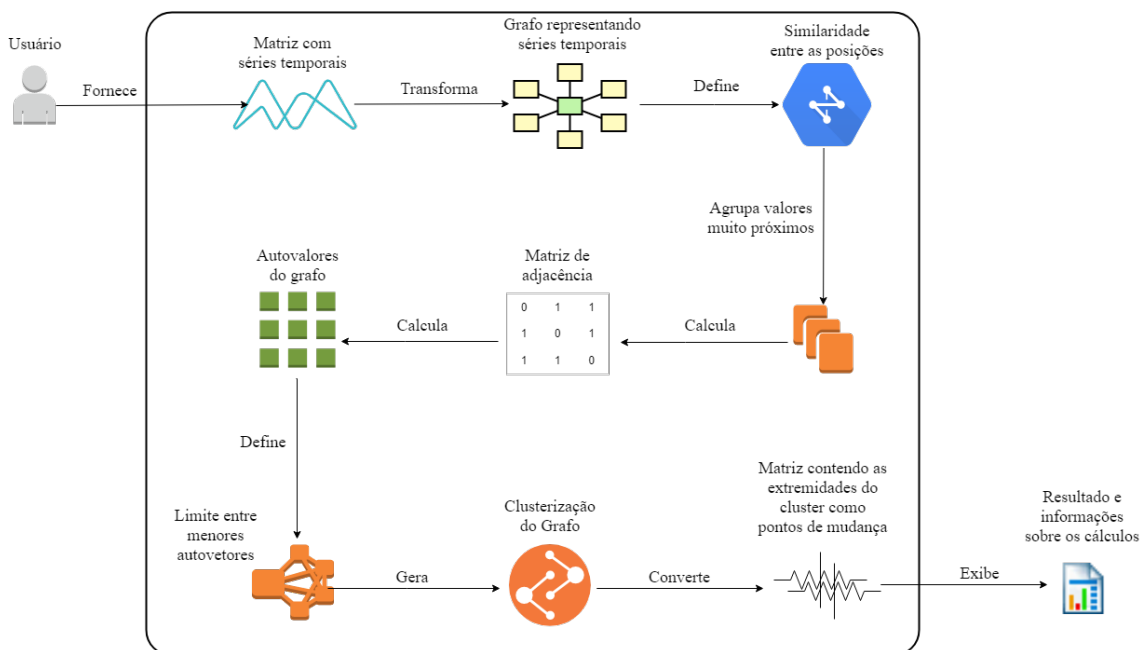
Assumindo que pretende-se identificar  $n$  grupos, a próxima etapa é encontrar os  $n$  menores autovetores (ignorando o autovetor constante trivial). O espectro do autovalor possui uma lacuna que fornece o valor de  $n$ . Em [Neto 2013] é possível verificar o exemplo do agrupamento da série que dificilmente seria agrupada pelos métodos tradicionais (k-Means e derivados).

O espectro do grafo é frequentemente utilizado para agrupamento em várias áreas e tem obtido resultados promissores na segmentação de imagens [Casaca 2014]. Entretanto, no melhor do nosso conhecimento não há trabalhos que relatem o agrupamento

espectral para detecção de pontos de mudança. O uso de grafos no contexto de pontos de mudança é recente e tem sido pouco explorado [Chen and Zhang 2015].

## 4. Metodologia

Nesta seção, será apresentada a metodologia para Detecção de Ponto de Mudança com o Espectro do Grafo. A Figura 1 faz um resumo da metodologia utilizada para detectar pontos de mudança utilizando o espectro do grafo.



**Figura 1. Explicação gráfica do processo de detecção de pontos de mudança com o espectro do grafo**

O algoritmo proposto é não-paramétrico e aplicável a quase todas as séries temporais, uma vez que, a medida de similaridade relatada no espaço amostral pode ser definida. É necessário fornecer ao algoritmo quais estatísticas serão utilizadas na varredura baseada em grafos, bem como medidas de similaridade e alternativas de intervalo.

### 4.1. Transformação dos dados

O primeiro passo do método proposto é focado em transformar séries temporais multivariadas em um grafo. Inicialmente, cada observação de cada medida representa um nó no grafo e o valor entre as arestas representa a distância entre cada observação. Existem várias distâncias métricas que podem ser utilizadas. Neste trabalho foi utilizada a distância euclidiana tradicional que é consolidada e trivial [Gower 1982].

Após a conversão, uma matriz de similaridade é formada sobre todas as observações do grafo. O processo é computacionalmente pesado e uma grande quantidade de dados precisa ser armazenada em memória em caso de múltiplas medidas ou em séries muito longas. A alternativa adotada para diminuir esse entrave de alto consumo de memória e tempo, bem como reduzir o ruído entre os valores é agrupar os nós muito próximos, assim, apenas os nós agrupados são armazenados em memória.

Sendo assim, se um nó A estiver relativamente próximo de um nó B, ambos os nós poderão ser substituídos por um único nó C. A distância mínima pode ser parametrizada ou calculada com base na distância média entre todos os nós da mesma série temporal. Este processo pode ser repetido até que permaneçam apenas distâncias relevantes entre os nós.

O processo de agrupamento de nós também pode ser considerado como uma suavização dos dados reduzindo o ruído. No entanto, a estrutura da base deve ser bem conhecida para que informações importantes não sejam perdidas. Para os testes realizados nesse trabalho foram agrupados os nós onde a distância é menor que 0,01 da distância máxima entre nós na mesma série e também, é menor que 0,02 da distância média na mesma série.

O agrupamento de nós leva em consideração o armazenamento de marcações que representam as observações originais, de modo que, a transformação do gráfico em séries temporais pode ser feita diretamente por referência após a detecção dos pontos de mudança.

## 4.2. Detecção de Pontos de Mudança

Usando o agrupamento espectral como descrito na sessão 3, as observações da série temporal são divididas em grupos distintos, onde  $1 < n_0 \leq \tau \leq n_1 < n$ , sendo  $n_0$  e  $n_1$  dois grupos distintos e  $\tau$  o valor limite entre os grupos, portanto,  $\tau$  é o ponto de mudança. Essa definição pode ser extrapolada para vários grupos na série temporal até que todos os pontos de mudança sejam determinados.

Tendo determinado a posição inicial dos pontos de mudança e grupos da série temporal é plausível verificar se os grupos realmente representam uma alteração no significado e na variância, ou seja, executar um teste de verificação afim de identificar falsos positivos. Várias técnicas estatísticas podem ser aplicadas para averiguar se dois grafos distintos possuem diferença de variação e significância, para este trabalho, foram utilizados os mesmos testes estatísticos de varredura aplicados no *gSeg*.

Com base nos resultados dos testes estatísticos os grupos podem ser reagrupados e os pontos de mudança desconsiderados. A tolerância dos testes deve ser definida levando em conta a natureza da distribuição dos dados bem como, o conhecimento prévio do domínio de aplicação, já que, uma tolerância muito alta podem aumentar a incidência de falsos positivos e uma tolerância baixa pode causar perda de pontos de mudança reais.

O Algoritmo 1 demonstra a detecção de pontos de mudança utilizando o *Spec*, ilustrando tanto a transformação de dados como o processo de detecção. O Algoritmo 1 é executado em *loop* para todas as séries temporais existentes no conjunto.

## 5. Resultados Experimentais

Experimentos foram implementados em R 3.4.3, sendo utilizados apenas os pacotes dos métodos para comparação com o método proposto, sem nenhum pacote de suporte.

### 5.1. Bases de Dados

Todas as bases de dados foram utilizadas a partir do repositório Arquivo de Classificação de Séries Temporais UCR [Chen et al. 2015]. A seleção do repositório foi dada pelos

---

**Algoritmo 1** Detecção de Mudanças com Spec

---

**Entrada:**  $S$  = série temporal multivalorada,  $L$  = Limite para Agrupamento

**Saída:** Conjunto de posições representando os pontos de mudança

**início**

**(\*Etapa 1: Transformação de dados\*)**

**para** cada  $s_i \in S$  **faça**

$O = \text{MarcaValoresOriginais}(s_i)$

$D = \text{CalculaDistanciaEntrePontos}(\text{"Euclidiana"}, S, s_i)$

**fim**

$SN = \text{AgrupaValoresProximos}(s_i, S, D, L, O)$

$MS = \text{CalculaMatrizSimilaridade}(D, SN)$

**(\*Etapa 2: Detecção por Agrupamento Espectral \*)**

$MA = \text{CalculaMatrizAfinidade}(\text{"Gaussiana"}, SN, MS)$

$A = \text{CalculaMatrizAdjacencia}(SN, MA)$

$AV = \text{CalculaAutoValores}(A)$

$G = \text{AgrupaConjuntoDados}(\text{"KMeans"}, AV, G)$

**para** cada  $g_i \in G$  **faça**

$GN = \text{TestaValidadeGrupo}(G_1, G)$

**fim**

**para** cada  $gn_i \in GN$  **faça**

$P = \text{PosicaoPontoMudanca}(O, SN, gn_i)$

**fim**

**retorna**  $P$

**fim**

---

critérios de adaptação dos algoritmos, pois alguns algoritmos possuem alta complexidade e tempo de processamento em séries temporais com mais de 10.000 observações o que tornaria as comparações inviáveis.

Para a seleção das séries temporais também foi considerado a segurança e integridade dos dados, desse modo, todas as séries apresentam as seguintes características: sem valores não preenchidos, rotuladas e já utilizadas em trabalhos recentes em benchmarks [Keogh and Kasetty 2003] [Rakthanmanon et al. 2013].

## 5.2. Avaliação de Desempenho

Para avaliar o desempenho dos algoritmos é necessário ter métricas objetivas que sejam aplicadas igualmente a todos os algoritmos. Portanto, é necessário utilizar bases rotuladas ou semi-rotuladas. Cada observação de uma série temporal é rotulada em uma classe Booleana, "Sim" onde representa que a observação é um ponto de mudança e "Não" se a observação é normal e natural na distribuição [Cook and Krishnan 2015].

- VP = Ponto de Mudança real predito como real;
- VN = Ponto de Mudança falso predito como falso;
- FP = Ponto de Mudança falso predito como real;
- FN = Ponto de Mudança real predito como falso;

$$Acc = \frac{VP + VN}{VP + FP + FN + VN}$$



**Tabela 1. Descrição das Bases de Dados**

Nome	Autor	Pontos de Mudança	Observações	Series
ECG	Olszewski	6	300	60
Yoga	Xi	2	3000	426
BeetleFly	Tony Bagnall	2	20	512
Computers	Tony Bagnall	2	250	720
Strawberry	Tony Bagnall	2	613	235
RefrigDevices	Tony Bagnall	3	375	720
OliveOil	Tony Bagnall	4	30	570
LgKitchenApp	Tony Bagnall	3	375	720
Meat	Tony Bagnall	3	60	448
BirdChicken	Tony Bagnall	2	20	512

$$Sp = \frac{VN}{VN + FP}$$

$$Re = \frac{VP}{VP + FN}$$

$$Pc = \frac{VP}{VP + FP}$$

$$F_1 = \frac{2 * Pc * Re}{Pc + Re}$$

$$Media = \frac{Acc + Sp + F_1}{3}$$

O fato de pesos e tolerâncias não terem sido aplicados faz com que a média dos resultados seja menor que a encontrada no estado da arte [Cheboli 2010]. Mudanças nas configurações e parâmetros de entrada dos algoritmos também podem alterar consideravelmente os resultados, atualmente, os parâmetros e posições foram ajustados para bases de dados, porém, foi feito o esforço para deixá-los o mais próximo possível da configuração padrão. O método proposto neste trabalho é denotado por *Spec* (abreviação de Spectral Clustering).

## 6. Conclusão

O desenvolvimento deste trabalho combinou-se em um novo pacote R capaz de identificar pontos de mudança com alta precisão em vários cenários. O uso do Espectro do Grafo já se mostrou eficiente para segmentação [Casaca 2014] e utilizá-lo na detecção de pontos de mudança se mostrou uma alternativa válida abrindo também novas possibilidades, já que, o estudo dos grafos em análise de séries temporais ainda não é muito explorado.

Em três conjunto de séries, o *Spec* obteve um resultado significativamente melhor quando comparado aos outros algoritmos:

**Tabela 2. Comparação - Média dos Resultados por Método**

	<b>PELT</b>	<b>ECP</b>	<b>EDM</b>	<b>gSeg</b>	<b>Spec</b>
ECG	0,42	0,44	0,47	0,47	0,47
Yoga	0,46	0,43	0,44	0,46	0,46
BeetleFly	0,69	0,63	0,69	0,69	0,69
Computers	0,58	0,65	0,71	0,71	0,71
Strawberry	0,49	0,46	0,46	0,53	0,53
RefrigDevices	0,57	0,62	0,55	0,65	<b>0,71</b>
OliveOil	0,44	0,55	0,59	0,59	<b>0,63</b>
LgKitchenApp	0,53	0,60	0,65	0,65	0,65
Meat	0,64	0,60	0,64	0,64	<b>0,80</b>
BirdChicken	0,58	0,63	0,69	0,69	0,69
<b>Média</b>	<b>0,54</b>	<b>0,56</b>	<b>0,59</b>	<b>0,61</b>	<b>0,63</b>

- Meat;
- OliveOil;
- RefrigDevices;

Explorar as características dessas séries deve ajudar a entender quando o uso do Espectro é mais vantajoso. Mesmo considerando os conjunto de séries que o algoritmo não apresentou uma melhora na média ainda se mostrou com performance similar aos outros algoritmos, mostrando sua viabilidade.

Este trabalho demonstrou que a transformação de séries em grafo torna possível aplicar o Espectro de Grafo para resolver o problema de pontos de mudança.

## Referências

- Aminikhanghahi, S. and Cook, D. J. (2016). A survey of methods for time series change point detection. *Knowledge and Information Systems*, pages 1–29.
- Basseville, M. and Nikiforov, I. (1993). *Detection of abrupt changes. Theory and application*, volume 2.
- Bonsal, B. R., Zhang, X., Vincent, L. A., and Hogg, W. D. (2001). Characteristics of daily and extreme temperatures over Canada. *Journal of Climate*, 14(9):1959–1976.
- Bosc, M., Heitz, F., Armspach, J. P., Namer, I., Gounot, D., and Rumbach, L. (2003). Automatic change detection in multimodal serial MRI: Application to multiple sclerosis lesion evolution. *NeuroImage*, 20(2):643–656.
- Casaca, W. C. d. O. (2014). *Graph laplacian for spectral clustering and seeded image segmentation*. PhD thesis, Universidade de São Paulo.
- Cheboli, D. (2010). Anomaly detection on time series. *2010 IEEE International Conference on Progress in Informatics and Computing*, 1:603–608.
- Chen, H. and Zhang, N. (2015). Graph-based change-point detection. *Annals of Statistics*, 43(1):139–176.

- Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., and Batista, G. (2015). The ucr time series classification archive. [www.cs.ucr.edu/~eamonn/time\\_series\\_data/](http://www.cs.ucr.edu/~eamonn/time_series_data/).
- Cook, D. and Krishnan, N. (2015). *Activity Learning*, volume 1.
- de Abreu, N. M. M., Del-Vecchio, R., Trevisan, V., and Vinagre, C. T. M. (2014). Teoria Espectral de Grafos - Uma Introdução. page 201.
- Ducré-Robitaille, J. F., Vincent, L. A., and Boulet, G. (2003). Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology*, 23(9):1087–1101.
- Fritscher, E. (2011). Propriedades espectrais de um grafo.
- Fujita, A., Takahashi, D. Y., Balardin, J. B., Vidal, M. C., and Sato, J. R. (2017). Correlation between graphs with an application to brain network analysis. *Computational Statistics and Data Analysis*, 109:76–92.
- Gower, J. C. (1982). Euclidean distance geometry. *Math. Sci*, 7(1):1–14.
- Itoh, N. and Kurths, J. (2010). Change-Point Detection of Climate Time Series by Nonparametric Method. In *Proceedings of the World Congress on Engineering and Computer Science*, volume I, pages 20–23.
- James, N. A., Kejariwal, A., and Matteson, D. S. (2016). In *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016*, pages 3499–3508.
- Keogh, E. and Kasetty, S. (2003). On the need for time series data mining benchmarks: a survey and empirical demonstration. *Data Mining and knowledge discovery*, 7(4):349–371.
- Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107(500):1590–1598.
- Lai, T. L. (2014). Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society*, 47(3):429–437.
- Li, S., Xu, L. D., and Zhao, S. (2015). The internet of things: a survey. *Information Systems Frontiers*, 17(2):243–259.
- Luxburg, U. V. (2007). A Tutorial on Spectral Clustering. pages 1–32.
- Malladi, R., Kalamangalam, G. P., and Aazhang, B. (2013). Online Bayesian change point detection algorithms for segmentation of epileptic activity. In *Conference Record - Asilomar Conference on Signals, Systems and Computers*, pages 1833–1837.
- Matteson, D. S. and James, N. A. (2014). *Journal of the American Statistical Association*, 109(505):334–345.
- Moreira, F. d. S. (2011). Detecção de pontos de mudança em séries temporais utilizando uma formulação Neural/Fuzzy/Bayesiana: Aplicação na detecção de falhas. *Dissertação Mestrado - UFMG*, page 69.
- Neto, J. (2013). Spectral clustering. <http://www.di.fc.ul.pt/~jpn/r/spectralclustering/spectralclustering.html>.

- Ng, A. Y., Jordan, M. I., and Weiss, Y. (2002). On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856.
- Page, E. S. (1955). A Test for a Change in a Parameter Occurring at an Unknown Point. *Biometrika*, 42(3/4):523.
- Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., and Keogh, E. (2013). Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(3):10.
- Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46(6):900–915.
- Spielman, D. A. (2007). Spectral graph theory and its applications. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on*, pages 29–38. IEEE.
- Staudacher, M., Telser, S., Amann, A., Hinterhuber, H., and Ritsch-Marte, M. (2005). A new method for change-point detection developed for on-line analysis of the heart beat variability during sleep. *Physica A: Statistical Mechanics and its Applications*, 349(3-4):582–596.
- Vidal, M. C., Sato, J. R., Balardin, J. B., Takahashi, D. Y., and Fujita, A. (2017). Anocva in r: A software to compare clusters between groups and its application to the study of autism spectrum disorder. *Frontiers in neuroscience*, 11:16.
- Yang, P., Dumont, G., and Ansermino, J. M. (2006). Adaptive change detection in heart rate trend monitoring in anesthetized children. *IEEE Transactions on Biomedical Engineering*, 53(11):2211–2219.
- Yonetani, T. and McCabe, G. J. (1994). Abrupt changes in regional temperature in the conterminous United States, 1895-1989. *Climate Research*, 4:13–23.