

Spektralna teoria grafów w analizie danych

Sławomir T. Wierzchoń

<http://www.ipipan.waw.pl/~stw/sem2111>

Instytut Podstaw Informatyki PAN



Warszawa, 21 listopada 2016

Treść

- 1 Wprowadzenie
 - Krótka charakterystyka
 - Zasadnicze kroki procedury
- 2 Narzędzia formalne
 - Laplasjan grafu
 - Wartość, wektor, porządek Fiedlera
- 3 Grupowanie spektralne
 - Kryteria rozcinania grafu
- 4 Uczenie częściowo nadzorowane
- 5 Signal Processing on graphs
 - Wprowadzenie
 - Szybkie grupowanie spektralne
 - Active semi-supervised learning

Czym jest skupienie (grupa)?

„Skupienie to **maksymalnie koherentny** zbiór obserwacji”.

- V. Estivill-Castro. “Why so many clustering algorithms: A position paper”. *SIGKDD Explor. Newsl.*, 4(1):65-75, Jun. 2002.
- M. Pelillo. “What is a cluster?”. A NIPS 2009 Workshop “Clustering: Science or Art? Towards Principled Approaches”.

Trudno jednoznacznie odpowiedzieć na to pytanie.

Czym jest skupienie (grupa)?

„Skupienie to **maksymalnie koherentny** zbiór obserwacji”.

- V. Estivill-Castro. “Why so many clustering algorithms: A position paper”. *SIGKDD Explor. Newsl.*, 4(1):65-75, Jun. 2002.
- M. Pelillo. “What is a cluster?”. A NIPS 2009 Workshop “Clustering: Science or Art? Towards Principled Approaches”.

Trudno jednoznacznie odpowiedzieć na to pytanie.

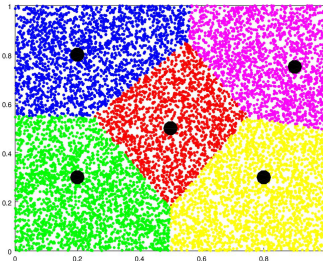
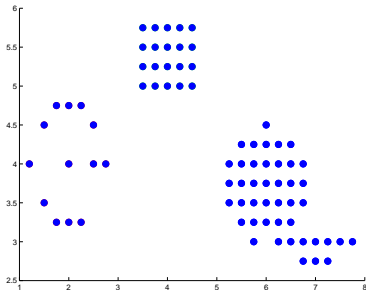
Czym jest skupienie (grupa)?

„Skupienie to **maksymalnie koherentny** zbiór obserwacji”.

- V. Estivill-Castro. “Why so many clustering algorithms: A position paper”. *SIGKDD Explor. Newsl.*, 4(1):65-75, Jun. 2002.
- M. Pelillo. “What is a cluster?”. A NIPS 2009 Workshop “Clustering: Science or Art? Towards Principled Approaches”.

Trudno jednoznacznie odpowiedzieć na to pytanie.

Skupienie – „klasyczna” odpowiedź



Rysunek: Skupienia to zwarte i dobrze separowalne grupy obiektów.

Prekursorzy

Ad 1

- 1 H. Steinhaus (1956): Sur la division des corp matériels en parties. *Bull. Acad. Polon. Sci*, Vol.1 (1956), pp. 801-804
- 2 S.P. Lloyd (1957): Least squares quantization in PCM. Bell Telephone Labs Memorandum, Murray Hill, NJ. Reprinted in: *IEEE Trans. Information Theory*, IT-28 (1982), vol. 2, 129-137

Ad 2

- 1 J. Czekanowski. *Zarys metod statystycznych w zastosowaniu do antropologii*. Towarzystwo Naukowe Warszawskie, 1913
- 2 P.L. Graham, P. Hell. On the history of the minimum spanning tree problem, *IEEE Ann. of the History of Computing*, 7(1),43-57, 1985

Jan Czekanowski (1882 - 1965)



(a)



(b)

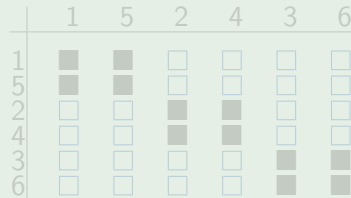
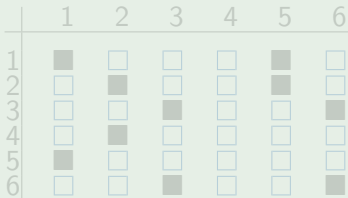
Rysunek: (a): pierwszy z prawej (b) pierwszy z lewej.

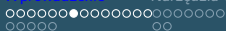
Algorytm Czekanowskiego

- Wyznacz elementy macierzy podobieństwa/odmienności,
 $W: X \times X \rightarrow \mathbb{R}$.
- Ustal dyskretyzację wartości w_{ij} , a przynależność w_{ij} do konkretnego przedziału oznacz innym symbolem graficznym (diagram Czekanowskiego).
- Uporządkuj wiersze i kolumny macierzy celem wizualizacji wewnętrznej struktury.

Example

	1	2	3	4	5	6
1	1.0	0.2	0.1	0.3	0.8	0.4
2	0.2	1.0	0.1	0.9	0.3	0.3
3	0.1	0.1	1.0	0.2	0.2	0.7
4	0.3	0.9	0.2	1.0	0.4	0.1
5	0.8	0.3	0.2	0.4	1.0	0.2
6	0.4	0.3	0.7	0.1	0.2	1.0



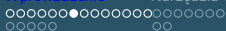


Example

	1	2	3	4	5	6
1	1.0	0.2	0.1	0.3	0.8	0.4
2	0.2	1.0	0.1	0.9	0.3	0.3
3	0.1	0.1	1.0	0.2	0.2	0.7
4	0.3	0.9	0.2	1.0	0.4	0.1
5	0.8	0.3	0.2	0.4	1.0	0.2
6	0.4	0.3	0.7	0.1	0.2	1.0

	1	2	3	4	5	6
1	■	□	□	□	■	□
2	□	■	□	□	■	□
3	□	□	■	□	□	■
4	□	■	□	□	□	□
5	■	□	□	□	□	□
6	□	□	■	□	□	■

	1	5	2	4	3	6
1	■	■	□	□	□	□
5	■	■	□	□	□	□
2	□	□	■	■	□	□
4	□	□	■	■	□	□
3	□	□	□	□	■	■
6	□	□	□	□	■	■



Example

	1	2	3	4	5	6
1	1.0	0.2	0.1	0.3	0.8	0.4
2	0.2	1.0	0.1	0.9	0.3	0.3
3	0.1	0.1	1.0	0.2	0.2	0.7
4	0.3	0.9	0.2	1.0	0.4	0.1
5	0.8	0.3	0.2	0.4	1.0	0.2
6	0.4	0.3	0.7	0.1	0.2	1.0

	1	2	3	4	5	6
1	■	□	□	□	■	□
2	□	■	□	□	■	□
3	□	□	■	□	□	■
4	□	■	□	□	□	□
5	■	□	□	□	□	□
6	□	□	■	□	□	■

	1	5	2	4	3	6
1	■	■	□	□	□	□
5	■	■	□	□	□	□
2	□	□	■	■	□	□
4	□	□	■	■	□	□
3	□	□	□	□	■	■
6	□	□	□	□	■	■

Metody grafowe a metody spektralne

- Metody rozcinania grafu, oparte na optymalizacji pewnej funkcji celu, są na ogół *NP*-trudne. Metody spektralne polegają na relaksacji dyskretnego zadania optymalizacyjnego i w czasie wielomianowym zwracają rozwiązania (które trzeba binaryzować).
- von Luxburg, Belkin, & Bousquet (2008): “The reason why the spectral relaxation is so appealing is not that it leads to particularly good solutions. Its popularity is mainly due to the fact that it results in a standard linear algebra problem which is simple to solve.”

Metody grafowe a metody spektralne

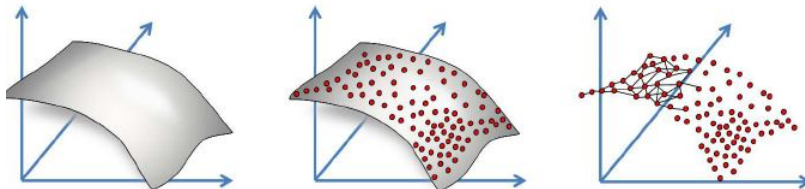
- Metody rozcinania grafu, oparte na optymalizacji pewnej funkcji celu, są na ogół *NP*-trudne. Metody spektralne polegają na relaksacji dyskretnego zadania optymalizacyjnego i w czasie wielomianowym zwracają rozwiązania (które trzeba binaryzować).
- von Luxburg, Belkin, & Bousquet (2008): “The reason why the spectral relaxation is so appealing is not that it leads to particularly good solutions. Its popularity is mainly due to the fact that it results in a standard linear algebra problem which is simple to solve.”

Zastosowania: Segmentacja obrazów



Rysunek: Oryginalne zdjęcie i jego segmentacja

Zastosowania: “Manifold learning”

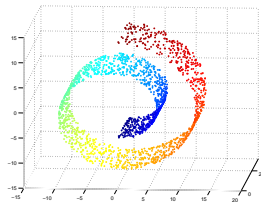


Rysunek: Gładka rozmierność w \mathbb{R}^3 , jej dyskretny odpowiednik uzyskany drogą próbkowania losowego, oraz struktura grafowa reprezentująca lokalne sąsiedztwa.

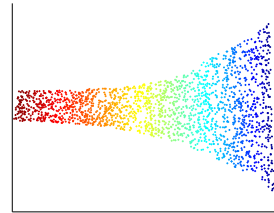
Zastosowania: LLE – Roweiss & Saul (2000)



(a)



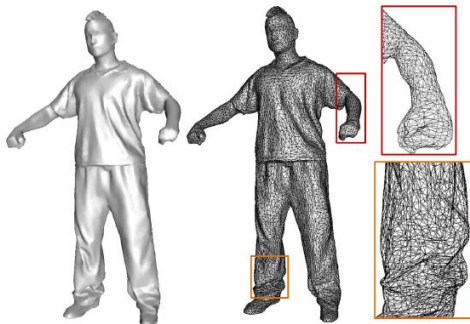
(b)



(c)

Rysunek: (a): “swiss-roll”, (b) model abstrakcyjny, (c): wynik końcowy.

Zastosowania: Grafika komputerowa



Rysunek: Dyskretna reprezentacja gładkiej 2D rozmierności Riemanna zanurzonej w 3-wymiarowej przestrzeni euklidesowej [A. Sharma, 2012].

Literatura

- F. Chung. *Spectral Graph Theory*. AMS, Providence, RI, 1997
- strona: [Applications of spectral graph theory](#)
- U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395-416, 2007
- H. Jia, *et al.* The latest research progress on spectral clustering. *Neural Comput. & Applic.*, 24(7-8):1477-1486, 2014
- M. Filippone, *et al.* A survey on spectral and kernel methods for clustering. *Pattern Recognition*, 41(1):176-190, 2008

Zakres stosowalności grupowania spektralnego

D.A. Spielman and S.-H. Teng. Spectral partitioning works: Planar graphs and finite element meshes. *Linear Algebra and its Applications*, 421(2-3), 284–305, 2007. DOI: 10.1016/j.laa.2006.07.020:

(...) spectral partitioning methods work well on bounded-degree planar graphs and finite element meshes – the classes of graphs to which they are usually applied.

Zasadnicze kroki: (1) Wybór funkcji wagowej

- Typowy wybór:

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$$

Uzasadnienie podają Belkin i Niyogi (2003). Ale:

(a) **podobieństwo zależy tylko od odległości**, (b) **jak wybrać σ ?**

- Niech x_i^K – K -ty najbliższy sąsiad obserwacji x_i i niech $\sigma_i = \|x_i - x_i^K\|$. Wtedy

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j}\right)$$

gdzie $K \approx \lceil \ln m \rceil$ – por. (Zelnik-Manor & Perona, 2000).

Zasadnicze kroki: (1) Wybór funkcji wagowej

- Typowy wybór:

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$$

Uzasadnienie podają Belkin i Niyogi (2003). Ale:

(a) **podobieństwo zależy tylko od odległości**, (b) **jak wybrać σ ?**

- Niech x_i^K – K -ty najbliższy sąsiad obserwacji x_i i niech $\sigma_i = \|x_i - x_i^K\|$. Wtedy

$$w_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma_i \sigma_j}\right)$$

gdzie $K \approx \lfloor \ln m \rfloor$ – por. (Zelnik-Manor & Perona, 2000).



Zasadnicze kroki: (2) Konstrukcja grafu podobieństwa

WE: $W \in \mathbb{R}^{m \times m}$ – symetryczna macierz podobieństwa,
 $\text{diag}(W) = \mathbf{0}$, $w_{ij} \geq 0$.

WY: $S = [s_{ij}]$ – uogólniona (ważona) macierz sąsiedztw
 reprezentująca graf $G = (V, E)$, $V \equiv X$.

- Graf ϵ -sąsiedztw: $\{v_i, v_j\} \in E \Leftrightarrow s_{ij} > \epsilon$. Ponadto: $s_{ij} = 1$ gdy $\{v_i, v_j\} \in E$, $s_{ij} = 0$ w p.p.
- Graf K -najbliższych sąsiadów (K -NN). Niech $N_K(v_i)$ – zbiór K najbliższych (najbardziej podobnych) węzłów do v_i . Wtedy

$$\{v_i, v_j\} \in E \Leftrightarrow v_i \in N_K(v_j) \wedge v_j \in N_K(v_i)$$

Ponadto $s_{ij} = w_{ij}$ gdy $\{v_i, v_j\} \in E$, $s_{ij} = 0$ w p.p. $\{v_i, v_j\} \in E$ gdy węzły należą do skupień o podobnej gęstości (von Luxburg, 2007).



Zasadnicze kroki: (2) Konstrukcja grafu podobieństwa

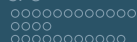
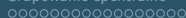
WE: $W \in \mathbb{R}^{m \times m}$ – symetryczna macierz podobieństwa,
 $\text{diag}(W) = \mathbf{0}$, $w_{ij} \geq 0$.

WY: $S = [s_{ij}]$ – uogólniona (ważona) macierz sąsiedztw
 reprezentująca graf $G = (V, E)$, $V \equiv X$.

- Graf **ϵ -sąsiedztw**: $\{v_i, v_j\} \in E \Leftrightarrow s_{ij} > \epsilon$. Ponadto: $s_{ij} = 1$ gdy $\{v_i, v_j\} \in E$, $s_{ij} = 0$ w p.p.
- Graf **K -najbliższych sąsiadów (K -NN)**. Niech $N_K(v_i)$ – zbiór K najbliższych (najbardziej podobnych) węzłów do v_i . Wtedy

$$\{v_i, v_j\} \in E \Leftrightarrow v_i \in N_K(v_j) \wedge v_j \in N_K(v_i)$$

Ponadto $s_{ij} = w_{ij}$ gdy $\{v_i, v_j\} \in E$, $s_{ij} = 0$ w p.p. $\{v_i, v_j\} \in E$,
 gdy węzły należą do skupień o podobnej gęstości (von Luxburg, 2007).



Zasadnicze kroki: (2) Konstrukcja grafu podobieństwa

WE: $W \in \mathbb{R}^{m \times m}$ – symetryczna macierz podobieństwa,
 $\text{diag}(W) = \mathbf{0}$, $w_{ij} \geq 0$.

WY: $S = [s_{ij}]$ – uogólniona (ważona) macierz sąsiedztw
 reprezentująca graf $G = (V, E)$, $V \equiv X$.

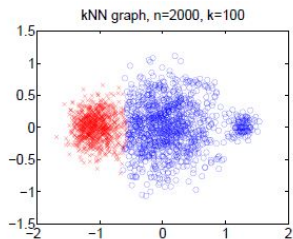
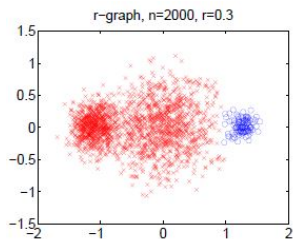
- Graf **ϵ -sąsiedztw**: $\{v_i, v_j\} \in E \Leftrightarrow s_{ij} > \epsilon$. Ponadto: $s_{ij} = 1$ gdy $\{v_i, v_j\} \in E$, $s_{ij} = 0$ w p.p.
- Graf **K -najbliższych sąsiadów** (K -NN). Niech $N_K(v_i)$ – zbiór K najbliższych (najbardziej podobnych) węzłów do v_i . Wtedy

$$\{v_i, v_j\} \in E \Leftrightarrow v_i \in N_K(v_j) \wedge v_j \in N_K(v_i)$$

Ponadto $s_{ij} = w_{ij}$ gdy $\{v_i, v_j\} \in E$, $s_{ij} = 0$ w p.p. $\{v_i, v_j\} \in E$,
 gdy węzły należą do skupień o podobnej gęstości (von Luxburg, 2007).



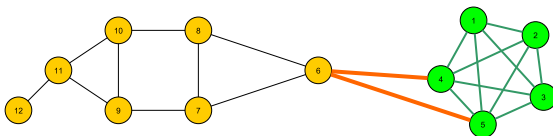
Który graf wybrać?



Rysunek: Maier, von Luxburg, & Hein (2011): Kryteria rozcinania grafu zależą od sposobu konstrukcji grafu podobieństwa

Zasadnicze kroki: (3) Kryteria jakości rozcięcia grafu

$$(a) : cut(A, B) = \sum_{i \in A, j \in B} s_{ij}, \quad (b) : assoc(A, A) = \sum_{i, j \in A} s_{ij}$$



Rysunek: Kolor czerwony – *cut*, zielony – *assoc*

Szukamy podziału (A, B) o minimalnym koszcie *cut*, lub maksymalnym *assoc*

Minimalizacja kosztu rozcięcia

Niech:

- χ – wektor charakterystyczny podziału (A, B) , tzn.: $\chi_j = +1$ gdy $v_j \in A$ oraz $\chi_j = -1$ gdy $v_j \in B$.

$$\begin{aligned} \chi^T D \chi &= \sum_{j=1}^m d_j \chi_j^2 = \left(\sum_{\substack{i \in A \\ j \in A}} s_{ij} \right) + \left(\sum_{\substack{i \in A \\ j \in B}} s_{ij} \right) + \left(\sum_{\substack{i \in B \\ j \in B}} s_{ij} \right) + \left(\sum_{\substack{i \in B \\ j \in A}} s_{ij} \right) \\ &= \text{assoc}(A, A) + \text{assoc}(B, B) + 2\text{cut}(A, B) \end{aligned}$$

$$\chi^T S \chi = \sum_{i=1}^m \sum_{j=1}^m s_{ij} \chi_i \chi_j = \text{assoc}(A, A) + \text{assoc}(B, B) - 2\text{cut}(A, B)$$

$$\chi^T (D - S) \chi = 4\text{cut}(A, B), \quad \mathbf{L} = \mathbf{D} - \mathbf{S}$$

Minimalizacja kosztu rozcięcia

Niech:

- χ – wektor charakterystyczny podziału (A, B) , tzn.: $\chi_j = +1$ gdy $v_j \in A$ oraz $\chi_j = -1$ gdy $v_j \in B$.

$$\begin{aligned} \chi^T D \chi &= \sum_{j=1}^m d_j \chi_j^2 = \left(\sum_{\substack{i \in A \\ j \in A}} s_{ij} \right) + \sum_{\substack{i \in A \\ j \in B}} s_{ij} + \left(\sum_{\substack{i \in B \\ j \in B}} s_{ij} \right) + \sum_{\substack{i \in B \\ j \in A}} s_{ij} \\ &= \text{assoc}(A, A) + \text{assoc}(B, B) + 2\text{cut}(A, B) \end{aligned}$$

$$\chi^T S \chi = \sum_{i=1}^m \sum_{j=1}^m s_{ij} \chi_i \chi_j = \text{assoc}(A, A) + \text{assoc}(B, B) - 2\text{cut}(A, B)$$

$$\chi^T (D - S) \chi = 4\text{cut}(A, B), \quad \mathbf{L} = \mathbf{D} - \mathbf{S}$$

Minimalizacja kosztu rozcięcia

Niech:

- χ – wektor charakterystyczny podziału (A, B) , tzn.: $\chi_j = +1$ gdy $v_j \in A$ oraz $\chi_j = -1$ gdy $v_j \in B$.

$$\begin{aligned}\chi^T D \chi &= \sum_{j=1}^m d_j \chi_j^2 = \left(\sum_{\substack{i \in A \\ j \in A}} s_{ij} \right) + \left(\sum_{\substack{i \in A \\ j \in B}} s_{ij} \right) + \left(\sum_{\substack{i \in B \\ j \in B}} s_{ij} \right) + \left(\sum_{\substack{i \in B \\ j \in A}} s_{ij} \right) \\ &= \text{assoc}(A, A) + \text{assoc}(B, B) + 2\text{cut}(A, B)\end{aligned}$$

$$\chi^T S \chi = \sum_{i=1}^m \sum_{j=1}^m s_{ij} \chi_i \chi_j = \text{assoc}(A, A) + \text{assoc}(B, B) - 2\text{cut}(A, B)$$

$$\chi^T (D - S) \chi = 4\text{cut}(A, B), \quad \mathbf{L} = \mathbf{D} - \mathbf{S}$$

Minimalizacja kosztu rozcięcia

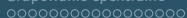
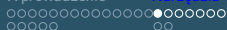
Niech:

- χ – wektor charakterystyczny podziału (A, B) , tzn.: $\chi_j = +1$ gdy $v_j \in A$ oraz $\chi_j = -1$ gdy $v_j \in B$.

$$\begin{aligned}\chi^T D \chi &= \sum_{j=1}^m d_j \chi_j^2 = \left(\sum_{\substack{i \in A \\ j \in A}} s_{ij} \right) + \left(\sum_{\substack{i \in A \\ j \in B}} s_{ij} \right) + \left(\sum_{\substack{i \in B \\ j \in B}} s_{ij} \right) + \left(\sum_{\substack{i \in B \\ j \in A}} s_{ij} \right) \\ &= \text{assoc}(A, A) + \text{assoc}(B, B) + 2\text{cut}(A, B)\end{aligned}$$

$$\chi^T S \chi = \sum_{i=1}^m \sum_{j=1}^m s_{ij} \chi_i \chi_j = \text{assoc}(A, A) + \text{assoc}(B, B) - 2\text{cut}(A, B)$$

$$\chi^T (D - S) \chi = 4\text{cut}(A, B), \quad \mathbf{L} = \mathbf{D} - \mathbf{S}$$



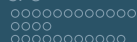
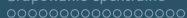
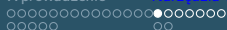
Laplasjan grafu

Oznaczenia:

- $d_i = \sum_{j=1}^m s_{ij}$ – stopień („typowość”) wężła v_i ;
 $D = \text{diag}(d_1, \dots, d_m)$ – macierz stopni.

Definicje:

- $L = D - S$ – laplasjan kombinatoryczny
- $\mathcal{L} = D^{-1/2} L D^{-1/2}$ – laplasjan normalizowany
- $\Delta = D^{-1} L = \mathbb{I} - D^{-1} S = D^{-1/2} \mathcal{L} D^{1/2}$ – dyskretny laplasjan
 ($P = D^{-1} S$ – macierz przejść)



Laplasjan grafu

Oznaczenia:

- $d_i = \sum_{j=1}^m s_{ij}$ – stopień („typowość”) wężła v_i ;
 $D = \text{diag}(d_1, \dots, d_m)$ – macierz stopni.

Definicje:

- $L = D - S$ – laplasjan kombinatoryczny
- $\mathcal{L} = D^{-1/2} L D^{-1/2}$ – laplasjan normalizowany
- $\Delta = D^{-1} L = \mathbb{I} - D^{-1} S = D^{-1/2} \mathcal{L} D^{1/2}$ – dyskretny laplasjan
 ($P = D^{-1} S$ – macierz przejść)



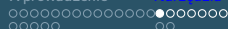
Laplasjan grafu

Oznaczenia:

- $d_i = \sum_{j=1}^m s_{ij}$ – stopień („typowość”) wężła v_i ;
 $D = \text{diag}(d_1, \dots, d_m)$ – macierz stopni.

Definicje:

- $L = D - S$ – laplasjan kombinatoryczny
- $\mathcal{L} = D^{-1/2} L D^{-1/2}$ – laplasjan normalizowany
- $\Delta = D^{-1} L = \mathbb{I} - D^{-1} S = D^{-1/2} \mathcal{L} D^{1/2}$ – dyskretny laplasjan
 ($P = D^{-1} S$ – macierz przejść)



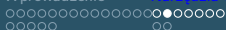
Laplasjan grafu

Oznaczenia:

- $d_i = \sum_{j=1}^m s_{ij}$ – stopień („typowość”) wężła v_i ;
 $D = \text{diag}(d_1, \dots, d_m)$ – macierz stopni.

Definicje:

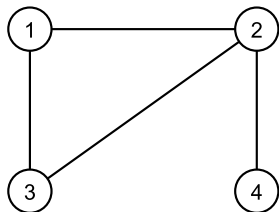
- $L = D - S$ – laplasjan kombinatoryczny
- $\mathcal{L} = D^{-1/2} L D^{-1/2}$ – laplasjan normalizowany
- $\Delta = D^{-1} L = \mathbb{I} - D^{-1} S = D^{-1/2} \mathcal{L} D^{1/2}$ – dyskretny laplasjan
 $(P = D^{-1} S$ – macierz przejść)



Przykład: Graf i jego laplasjan (1)

$$S = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{d} = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

$$L = D - S = \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

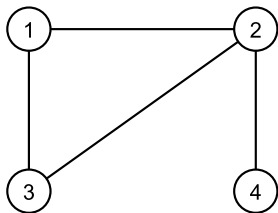


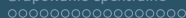
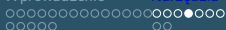


Przykład: Graf i jego laplasjan (2)

$$S = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{d} = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 1 \end{bmatrix}$$

$$P = \mathbb{I} - \Delta = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 \\ 1/3 & 0 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$





Własności kombinatorycznego laplasjanu (1)

- 1 L jest kwadratową macierzą symetryczną o elementach

$$l_{ij} = \begin{cases} d_i - s_{ij} & \text{gdy } i = j \\ -s_{ij} & \text{gdy } i \neq j \text{ oraz } \{v_i, v_j\} \in E \\ 0 & \text{w p.p.} \end{cases}$$

- 2 $Le = \mathbf{0}$,

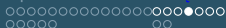
- 3 jeżeli $f: V \rightarrow \mathbb{R}$, to

$$(a) \quad (Lf)_j = d_j f_j - \sum_{e \neq j} l_{je} f_e = \sum_{i=1}^m s_{ij} (f_j - f_i)$$

$$(b) \quad f^T Lf = \sum_{i,j} s_{ij} (f_i - f_j)^2 \geq 0$$

tnz. L jest macierzą dodatnio półokreśloną





Własności kombinatorycznego laplasjanu (1)

- 1 L jest kwadratową macierzą symetryczną o elementach

$$l_{ij} = \begin{cases} d_i - s_{ij} & \text{gdy } i = j \\ -s_{ij} & \text{gdy } i \neq j \text{ oraz } \{v_i, v_j\} \in E \\ 0 & \text{w p.p.} \end{cases}$$

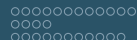
- 2 $Le = \mathbf{0}$,

- 3 jeżeli $f: V \rightarrow \mathbb{R}$, to

$$(a) \quad (Lf)_j = d_j f_j - \sum_{e \neq j} l_{je} f_e = \sum_{i=1}^m s_{ij} (f_j - f_i)$$

$$(b) \quad f^T Lf = \sum_{i,j} s_{ij} (f_i - f_j)^2 \geq 0$$

tnz. L jest macierzą dodatnio półokreśloną



Własności kombinatorycznego laplasjanu (1)

- 1 L jest kwadratową macierzą symetryczną o elementach

$$l_{ij} = \begin{cases} d_i - s_{ij} & \text{gdy } i = j \\ -s_{ij} & \text{gdy } i \neq j \text{ oraz } \{v_i, v_j\} \in E \\ 0 & \text{w p.p.} \end{cases}$$

- 2 $L\mathbf{e} = \mathbf{0}$,

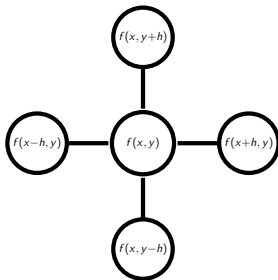
- 3 jeżeli $f: V \rightarrow \mathbb{R}$, to

$$(a) \quad (Lf)_j = d_j f_j - \sum_{\ell \neq j} l_{j\ell} f_\ell = \sum_{i=1}^m s_{ij} (f_j - f_i)$$

$$(b) \quad f^T Lf = \sum_{i,j} s_{ij} (f_i - f_j)^2 \geq 0$$

tzn. L jest macierzą dodatnio półokreśloną

Interpretacja (*5-stencil formula*)



Rysunek:

$$(\nabla^2 f)(x, y) = -\frac{1}{h} [4f(x, y) - f(x-h, y) - f(x+h, y) - f(x, y-h) - f(x, y+h)] = -Lf$$

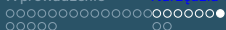
Interpretacja, cd

L. Grady, J.L. Polimieni: *Discrete Calculus. Applied Analysis on Graphs for Computational Science*, Springer 2010.

Sekcja „Discretized, vs. Discrete calculus”:

In the first case, the goal is to compute a solution to some problem on a continuous space. But, an analytic solution is too difficult to find and so a discretization strategy is employed that allows for a computer to produce an approximate solution.

*In contrast, **discrete calculus** treats a discrete domain (e.g., a graph) as entirely its own entity with no reference to an underlying continuum. E.g., a social network is not associated with any continuous space in the sense that the network is not viewed as a discretization or sampling of an underlying continuum.*

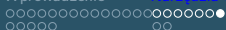


Własności kombinatorycznego laplasjanu (2)

Definition

Jeżeli $Ly = \lambda y$, to y jest wektorem własnym, λ – wartością własną laplasjanu

- 1 L symetryczna: $\Rightarrow \lambda \in \mathbb{R}$
- 2 L dodatnio półokreślona: $\Rightarrow \lambda_i \geq 0, i = 1, \dots, m,$
- 3 $Le = \mathbf{0}$: $\lambda_1 = 0$ – minimalna wartość własna; $y_1 = e/\sqrt{m}$
- 4 konwencja: $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m \leq m$

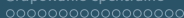
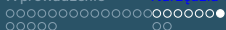


Własności kombinatorycznego laplasjanu (2)

Definition

Jeżeli $Ly = \lambda y$, to y jest wektorem własnym, λ – wartością własną laplasjanu

- 1 L symetryczna: $\Rightarrow \lambda \in \mathbb{R}$
- 2 L dodatnio półokreślona: $\Rightarrow \lambda_i \geq 0, i = 1, \dots, m,$
- 3 $Le = \mathbf{0}$: $\lambda_1 = 0$ – minimalna wartość własna; $y_1 = e/\sqrt{m}$
- 4 konwencja: $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m \leq m$

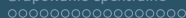
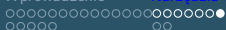


Własności kombinatorycznego laplasjanu (2)

Definition

Jeżeli $Ly = \lambda y$, to y jest wektorem własnym, λ – wartością własną laplasjanu

- 1 L symetryczna: $\Rightarrow \lambda \in \mathbb{R}$
- 2 L dodatnio półokreślona: $\Rightarrow \lambda_i \geq 0, i = 1, \dots, m,$
- 3 $L\mathbf{e} = \mathbf{0}$: $\lambda_1 = 0$ – minimalna wartość własna; $y_1 = \mathbf{e}/\sqrt{m}$
- 4 konwencja: $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m \leq m$

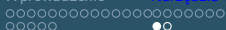


Własności kombinatorycznego laplasjanu (2)

Definition

Jeżeli $Ly = \lambda y$, to y jest wektorem własnym, λ – wartością własną laplasjanu

- 1 L symetryczna: $\Rightarrow \lambda \in \mathbb{R}$
- 2 L dodatnio półokreślona: $\Rightarrow \lambda_i \geq 0, i = 1, \dots, m,$
- 3 $Le = \mathbf{0}$: $\lambda_1 = 0$ – minimalna wartość własna; $y_1 = \mathbf{e}/\sqrt{m}$
- 4 konwencja: $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m \leq m$



Wartość i wektor Fiedlera (λ_2, y_2)

$$0 \leq \lambda_2 \leq m;$$

Fiedler (1973,1975) udowodnił, że

- $\lambda_2 > 0$ w.t.w. $G = (V, E)$ jest grafem spójnym.
- Jeżeli $G' = (V, E')$, $E' \subset E$, to $\lambda_2(G') \leq \lambda_2(G)$. Niskie wartości λ_2 przemawiają za istnieniem wewnętrznej struktury w zbiorze X . Np. gdy graf planarny, to $\lambda_2 \leq \frac{8d_{max}}{m}$ (Spielman & Teng, 2007).
- Wektory własne y_i odpowiadające wartościom własnym λ_i są ortogonalne. Zatem $y_1^T y_2 = 0 \Rightarrow \sum_j y_{j2} = 0$.
- Jeżeli G jest spójnym i nieskierowanym grafem oraz $A = \{v_i: y_{i2} \geq 0\}$, $B = V \setminus A = \{v_i: y_{i2} < 0\}$, to podgraf $G_B = (B, E_B)$ jest spójny.

Wartość i wektor Fiedlera (λ_2, y_2)

$$0 \leq \lambda_2 \leq m;$$

Fiedler (1973,1975) udowodnił, że

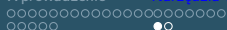
- $\lambda_2 > 0$ w.t.w. $G = (V, E)$ jest grafem spójnym.
- Jeżeli $G' = (V, E')$, $E' \subset E$, to $\lambda_2(G') \leq \lambda_2(G)$. Niskie wartości λ_2 **przemawiają** za istnieniem wewnętrznej struktury w zbiorze X . Np. gdy graf **planarny**, to $\lambda_2 \leq \frac{8d_{max}}{m}$ (Spielman & Teng, 2007).
- Wektory własne y_i odpowiadające wartościom własnym λ_i są ortogonalne. Zatem $y_1^T y_2 = 0 \Rightarrow \sum_j y_{j2} = 0$.
- Jeżeli G jest spójnym i nieskierowanym grafem oraz $A = \{v_i : y_{i2} \geq 0\}$, $B = V \setminus A = \{v_i : y_{i2} < 0\}$, to podgraf $G_B = (B, E_B)$ jest spójny.

Wartość i wektor Fiedlera (λ_2, y_2)

$$0 \leq \lambda_2 \leq m;$$

Fiedler (1973,1975) udowodnił, że

- $\lambda_2 > 0$ w.t.w. $G = (V, E)$ jest grafem spójnym.
- Jeżeli $G' = (V, E')$, $E' \subset E$, to $\lambda_2(G') \leq \lambda_2(G)$. Niskie wartości λ_2 **przemawiają** za istnieniem wewnętrznej struktury w zbiorze X . Np. gdy graf **planarny**, to $\lambda_2 \leq \frac{8d_{max}}{m}$ (Spielman & Teng, 2007).
- Wektory własne y_i odpowiadające wartościom własnym λ_i są ortogonalne. Zatem $y_1^T y_2 = 0 \Rightarrow \sum_j y_{j2} = 0$.
- Jeżeli G jest spójnym i nieskierowanym grafem oraz $A = \{v_i : y_{i2} \geq 0\}$, $B = V \setminus A = \{v_i : y_{i2} < 0\}$, to podgraf $G_B = (B, E_B)$ jest spójny.



Wartość i wektor Fiedlera (λ_2, y_2)

$$0 \leq \lambda_2 \leq m;$$

Fiedler (1973,1975) udowodnił, że

- $\lambda_2 > 0$ w.t.w. $G = (V, E)$ jest grafem spójnym.
- Jeżeli $G' = (V, E')$, $E' \subset E$, to $\lambda_2(G') \leq \lambda_2(G)$. Niskie wartości λ_2 **przemawiają** za istnieniem wewnętrznej struktury w zbiorze X . Np. gdy graf **planarny**, to $\lambda_2 \leq \frac{8d_{max}}{m}$ (Spielman & Teng, 2007).
- Wektory własne y_i odpowiadające wartościom własnym λ_i są ortogonalne. Zatem $y_1^T y_2 = 0 \Rightarrow \sum_j y_{j2} = 0$.
- Jeżeli G jest spójnym i nieskierowanym grafem oraz $A = \{v_i : y_{i2} \geq 0\}$, $B = V \setminus A = \{v_i : y_{i2} < 0\}$, to podgraf $G_B = (B, E_B)$ jest spójny.

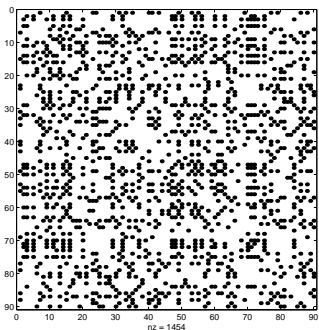
Wartość i wektor Fiedlera (λ_2, y_2)

$$0 \leq \lambda_2 \leq m;$$

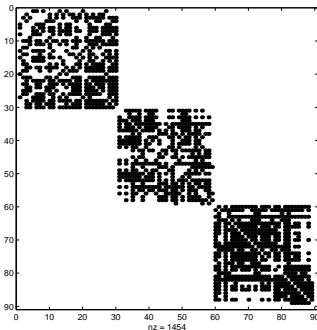
Fiedler (1973,1975) udowodnił, że

- $\lambda_2 > 0$ w.t.w. $G = (V, E)$ jest grafem spójnym.
- Jeżeli $G' = (V, E')$, $E' \subset E$, to $\lambda_2(G') \leq \lambda_2(G)$. Niskie wartości λ_2 **przemawiają** za istnieniem wewnętrznej struktury w zbiorze X . Np. gdy graf **planarny**, to $\lambda_2 \leq \frac{8d_{max}}{m}$ (Spielman & Teng, 2007).
- Wektory własne y_i odpowiadające wartościom własnym λ_i są ortogonalne. Zatem $y_1^T y_2 = 0 \Rightarrow \sum_j y_{j2} = 0$.
- Jeżeli G jest spójnym i nieskierowanym grafem oraz $A = \{v_i : y_{i2} \geq 0\}$, $B = V \setminus A = \{v_i : y_{i2} < 0\}$, to podgraf $G_B = (B, E_B)$ jest spójny.

Porządek Fiedlera

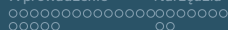


(a)



(b)

Rysunek: macierz L : (a) oryginalna, (b) po posortowaniu względem wartości y_2



Inne kryteria rozcinania

$$Rcut(C_1, \dots, C_k) = \sum_{j=1}^k \frac{cut(C_j, \bar{C}_j)}{|C_j|}, \quad (\text{Hagen \& Kahng, 1992})$$

$$Ncut(C_1, \dots, C_k) = \sum_{j=1}^k \frac{cut(C_j, \bar{C}_j)}{\text{vol } C_j}, \quad (\text{Shi \& Malik, 2000})$$

$$Mcut(C_1, \dots, C_k) = \sum_{j=1}^k \frac{cut(C_j, \bar{C}_j)}{\text{assoc}(C_j, C_j)}, \quad (\text{Ding } et al., 2001)$$

gdzie, $\text{vol } C_j = \sum_{v_i \in C_j} d_i$.

Relaksacja $Rcut$ (1)

Niech $h_i = (h_{1i}, \dots, h_{ji}, \dots, h_{m_i})^T$ wektor o składowych

$$h_{ji} = \begin{cases} 1/\sqrt{|C_i|} & \text{gdy } j \in C_i \\ 0 & \text{w p.p.} \end{cases}, \quad H = (h_1, \dots, h_k)$$

$\|h_i\| = h_i^T h_i = 1$ oraz $h_i^T h_j = 0$ gdy $i \neq j$, czyli macierz macierz H spełnia warunek $H^T H = \mathbb{I}$. Ponadto,

$$h_i^T L h_i = \frac{cut(C_i, \overline{C_i})}{|C_i|} = (H^T L H)_{ii}$$

Zatem

$$Rcut(C_1, \dots, C_k) = \sum_{i=1}^k h_i^T L h_i = \sum_{i=1}^k (H^T L H)_{ii} = \text{tr}(H^T L H)$$

Relaksacja R_{cut} (2)

$$\min_{(C_1, \dots, C_k) \in \mathcal{C}^k} \text{tr}(H^T L H), \text{ gdzie } H \text{ jak wyżej}$$

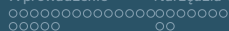
Relaksacja:

$$\min_{\substack{H \in \mathbb{R}^{n \times k} \\ H^T H = \mathbb{I}}} \text{tr}(H^T L H)$$

Na mocy tw. Fan'a (1947), por. (Mirzal & Furukawa, 2010)

$$\min_{\substack{H \in \mathbb{R}^{n \times k} \\ H^T H = \mathbb{I}}} \text{tr}(H^T L H) = \sum_{j=1}^k \lambda_j \quad \text{jeżeli } H = (y_1, \dots, y_k) R$$

gdzie y_i - wektor własny odpowiadający i -tej minimalnej wartości własnej λ_j laplasjanu L , oraz $R \in \mathbb{C}^{k \times k}$ - macierz unitarna.



Relaksacja $Ncut$ (1)

Niech $h_i = (h_{1i}, \dots, h_{ji}, \dots, h_{mi})^T$ wektor o składowych

$$h_{ji} = \begin{cases} 1/\sqrt{\text{vol } C_i} & \text{gdy } j \in C_i \\ 0 & \text{w p.p.} \end{cases}, \quad H = (h_1, \dots, h_k)$$

$h_j^T Dh_j = 1$, oraz $h_i^T h_j = 0$ gdy $i \neq j$, czyli $H^T DH = \mathbb{I}$. Ponadto,

$$h_i^T Lh_i = \frac{\text{cut}(C_i, \overline{C_i})}{\text{vol } C_i} = (H^T LH)_{ii}$$

Zatem

$$Ncut(C_1, \dots, C_k) = \sum_{i=1}^k h_i^T Lh_i = \sum_{i=1}^k (H^T LH)_{ii} = \text{tr}(H^T LH)$$

Relaksacja $Ncut$ (2)

Relaksacja zadania $NCut$ przyjmuje postać

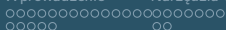
$$\min_{\substack{H \in \mathbb{R}^{n \times k} \\ H^T D H = \mathbb{I}}} \text{tr}(H^T L H)$$

Niech $Q = D^{1/2} H$. Wówczas

$$\min_{\substack{Q \in \mathbb{R}^{n \times k} \\ Q^T Q = \mathbb{I}}} \text{tr}[Q^T (D^{-1/2} L D^{-1/2}) Q] = \min_{\substack{Q \in \mathbb{R}^{n \times k} \\ Q^T Q = \mathbb{I}}} \text{tr}(Q^T \mathcal{L} Q)$$

i można skorzystać z tw. Fan'a (1949), tzn. rozwiązaniem jest

$$Q = (y_1, \dots, y_k) R, \text{ gdzie } y_i \text{ to } i\text{-ty wektor własny macierzy} \\ \mathcal{L} = D^{-1/2} L D^{-1/2}.$$



Normalizowany laplasjan $\mathcal{L} = D^{-1/2}LD^{-1/2}$ (1)

- $(\forall x \in \mathbb{R}^m): x^T \mathcal{L} x = \sum_{i=1}^m \sum_{j=1}^m s_{ij} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2$

- $\mathcal{L} = D^{-1/2}LD^{-1/2}$ jest macierzą symetryczną i dodatnio-półokreśloną o elementach

$$\mathcal{L}(i, j) = \begin{cases} (d_i - s_{ij})/d_i & \text{gdy } i = j \\ -s_{ij}/(\sqrt{d_i d_j}) & \text{gdy } i \neq j \text{ oraz } \{v_i, v_j\} \in E \\ 0 & \text{w p.p.} \end{cases}$$

- jeżeli $y_1 = D^{1/2} \mathbf{e}$, to $\mathcal{L} y_1 = D^{-1/2} L D^{-1/2} (D^{1/2} \mathbf{e}) = D^{-1/2} (L \mathbf{e}) = 0 \Rightarrow (0, y_1)$ para własna \mathcal{L} .
- ogólnie $0 = \lambda_1 \leq \dots \lambda_m \leq 2$.



Normalizowany laplasjan $\mathcal{L} = D^{-1/2}LD^{-1/2}$ (1)

- $(\forall x \in \mathbb{R}^m): x^T \mathcal{L}x = \sum_{i=1}^m \sum_{j=1}^m s_{ij} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2$
- $\mathcal{L} = D^{-1/2}LD^{-1/2}$ jest macierzą symetryczną i dodatnio-półokreśloną o elementach

$$\mathcal{L}(i, j) = \begin{cases} (d_i - s_{ij})/d_i & \text{gdy } i = j \\ -s_{ij}/(\sqrt{d_i d_j}) & \text{gdy } i \neq j \text{ oraz } \{v_i, v_j\} \in E \\ 0 & \text{w p.p.} \end{cases}$$

- jeżeli $y_1 = D^{1/2}e$, to $\mathcal{L}y_1 = D^{-1/2}LD^{-1/2}(D^{1/2}e) = D^{-1/2}(Le) = 0 \Rightarrow (0, y_1)$ para własna \mathcal{L} .
- ogólnie $0 = \lambda_1 \leq \dots \lambda_m \leq 2$.



Normalizowany laplasjan $\mathcal{L} = D^{-1/2}LD^{-1/2}$ (1)

- $(\forall x \in \mathbb{R}^m): x^T \mathcal{L}x = \sum_{i=1}^m \sum_{j=1}^m s_{ij} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2$
- $\mathcal{L} = D^{-1/2}LD^{-1/2}$ jest macierzą symetryczną i dodatnio-półokreśloną o elementach

$$\mathcal{L}(i, j) = \begin{cases} (d_i - s_{ij})/d_i & \text{gdy } i = j \\ -s_{ij}/(\sqrt{d_i d_j}) & \text{gdy } i \neq j \text{ oraz } \{v_i, v_j\} \in E \\ 0 & \text{w p.p.} \end{cases}$$

- jeżeli $y_1 = D^{1/2}\mathbf{e}$, to $\mathcal{L}y_1 = D^{-1/2}LD^{-1/2}(D^{1/2}\mathbf{e}) = D^{-1/2}(L\mathbf{e}) = 0 \Rightarrow (0, y_1)$ para własna \mathcal{L} .
- ogólnie $0 = \lambda_1 \leq \dots \lambda_m \leq 2$.



Normalizowany laplasjan $\mathcal{L} = D^{-1/2}LD^{-1/2}$ (1)

- $(\forall x \in \mathbb{R}^m): x^T \mathcal{L}x = \sum_{i=1}^m \sum_{j=1}^m s_{ij} \left(\frac{x_i}{\sqrt{d_i}} - \frac{x_j}{\sqrt{d_j}} \right)^2$
- $\mathcal{L} = D^{-1/2}LD^{-1/2}$ jest macierzą symetryczną i dodatnio-półokreśloną o elementach

$$\mathcal{L}(i, j) = \begin{cases} (d_i - s_{ij})/d_i & \text{gdy } i = j \\ -s_{ij}/(\sqrt{d_i d_j}) & \text{gdy } i \neq j \text{ oraz } \{v_i, v_j\} \in E \\ 0 & \text{w p.p.} \end{cases}$$

- jeżeli $y_1 = D^{1/2}\mathbf{e}$, to $\mathcal{L}y_1 = D^{-1/2}LD^{-1/2}(D^{1/2}\mathbf{e}) = D^{-1/2}(L\mathbf{e}) = 0 \Rightarrow (0, y_1)$ para własna \mathcal{L} .
- ogólnie $0 = \lambda_1 \leq \dots \lambda_m \leq 2$.



Normalizowany laplasjan $\mathcal{L} = D^{-1/2}LD^{-1/2}$ (2)

1

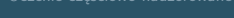
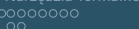
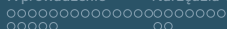
$$\mathcal{L} = D^{-1/2}LD^{-1/2} = \mathbb{I} - D^{-1/2}SD^{-1/2} = \mathbb{I} - \mathcal{S}$$

Jeżeli (λ_i, ψ_i) para własna \mathcal{L} , to $(\bar{\lambda}_i = 1 - \lambda_i, \psi_i)$ – para własna macierzy \mathcal{S} . Wystarczy więc wyznaczyć k wektorów własnych odpowiadających **maksymalnym** wartościom własnym macierzy $\bar{\mathcal{L}}$: $1 = \bar{\lambda}_1 \geq \dots \geq \bar{\lambda}_m \geq -1$.

2

$$\mathcal{L}\psi = \lambda\psi \Rightarrow D^{-1/2}(D^{-1/2}L(D^{-1/2}\psi)) = \lambda(D^{-1/2}\psi)$$

tzn. $(\lambda, D^{-1/2}\psi)$ jest parą własną macierzy Δ .

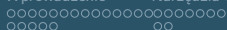


Prosty algorytm grupowania (MATLAB)

```

X = load('data.dat');
utwórz macierz wag W;
przekształć W w macierz powiązań S;
wyznacz laplasjan Ł (normalizowany ew. nie-);
[Y, Lambda]=eigs(Ł,'SM',2);
binaryzuj wektor y = Y(:,2)
  
```

Ciekawy wariant binaryzacji przedstawili Tolliver, & Miller (2006).
 Ewentualnie: powtarzaj algorytm (wybierając podgraf o minimalnej wartości λ_2) do chwili uzyskania podziału na k grup.



Prosty algorytm grupowania (MATLAB)

```
X = load('data.dat');  
utwórz macierz wag W;  
przekształć W w macierz powiązań S;  
wyznacz laplasjan  $\mathbf{L}$  (normalizowany ew. nie-);  
[Y, Lambda]=eigs(L,'SM',2);  
binaryzuj wektor  $y = Y(:,2)$ 
```

Ciekawy wariant binaryzacji przedstawili Tolliver, & Miller (2006).
Ewentualnie: powtarzaj algorytm (wybierając podgraf o minimalnej wartości λ_2) do chwili uzyskania podziału na k grup.

Prosty algorytm grupowania (MATLAB)

```
X = load('data.dat');
utwórz macierz wag W;
przekształć W w macierz powiązań S;
wyznacz laplasjan Ł (normalizowany ew. nie-);
[Y, Lambda]=eigs(Ł,'SM',2);
binaryzuj wektor y = Y(:,2)
```

Ciekawy wariant binaryzacji przedstawili Tolliver, & Miller (2006).
Ewentualnie: powtarzaj algorytm (wybierając podgraf o minimalnej wartości λ_2) do chwili uzyskania podziału na k grup.

Algorytm NJW (Ng, Jordan, & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ij} = 0$

- 1 wyznacz $\mathcal{S} = D^{-1/2}SD^{-1/2}$
- 2 wyznacz $Y = (y_1, \dots, y_k)$; y_j - j -ty dominujący wektor
- 3 rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \dots, m$, $j = 1, \dots, k$ o elementach

$$z_{ij} = y_{i,j} / \left(\sum_{\ell=1}^k y_{i,\ell}^2 \right)^{-1/2}$$

- 4 wyznacz finalny podział danych reprezentowanych przez macierz Z .

Algorytm NJW (Ng, Jordan, & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ij} = 0$

- 1 wyznacz $\mathcal{S} = D^{-1/2}SD^{-1/2}$
- 2 wyznacz $Y = (y_1, \dots, y_k)$; y_j - j -ty dominujący wektor
- 3 rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \dots, m$, $j = 1, \dots, k$ o elementach

$$z_{ij} = y_{i,j} / \left(\sum_{\ell=1}^k y_{i,\ell}^2 \right)^{-1/2}$$

- 4 wyznacz finalny podział danych reprezentowanych przez macierz Z .

Algorytm NJW (Ng, Jordan, & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ij} = 0$

- 1 wyznacz $\mathcal{S} = D^{-1/2}SD^{-1/2}$
- 2 wyznacz $Y = (y_1, \dots, y_k)$; y_j - j -ty dominujący wektor
- 3 rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \dots, m$, $j = 1, \dots, k$ o elementach

$$z_{ij} = y_{i,j} / \left(\sum_{\ell=1}^k y_{i\ell}^2 \right)^{-1/2}$$

- 4 wyznacz finalny podział danych reprezentowanych przez macierz Z .

Algorytm NJW (Ng, Jordan, & Weiss, 2002)

WE: symetryczna macierz S taka, że $s_{ij} = 0$

- 1 wyznacz $\mathcal{S} = D^{-1/2}SD^{-1/2}$
- 2 wyznacz $Y = (y_1, \dots, y_k)$; y_j - j -ty dominujący wektor
- 3 rzutuj wiersze macierzy Y na sferę jednostkową, tzn. utwórz macierz $Z = [z_{ij}]$, $i = 1, \dots, m$, $j = 1, \dots, k$ o elementach

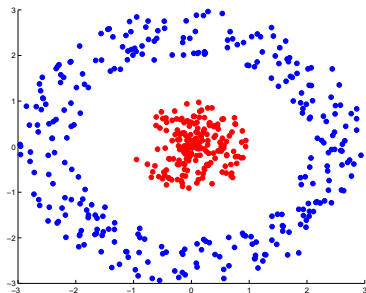
$$z_{ij} = y_{i,j} / \left(\sum_{\ell=1}^k y_{i\ell}^2 \right)^{-1/2}$$

- 4 wyznacz finalny podział danych reprezentowanych przez macierz Z .

Interpretacja algorytmu NJW

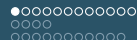
Krok (2) algorytmu to projekcja węzłów grafu na k -wymiarową przestrzeń. Zatem **grupowanie spektralne = projekcja + k -means**

Deep Learning: dobór właściwej reprezentacji



(a)

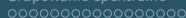
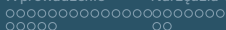
(b)



Sygnaly na grafach

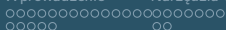
$\Gamma = (V, E)$ – nieskierowany, ważony graf o m węzłach.

Odwzorowanie $f: V \rightarrow \mathbb{R}$, takie że $f(i)$ jest wartością przypisaną węzłowi v_i nazywamy sygnałem. f utożsamia się z wektorem $\mathbf{f} \in \mathbb{R}^m$, ale trzeba pamiętać, że $\mathbf{f} = (f_1, \dots, f_m)^T$ nie jest zwykłą listą; elementy f_i wektora są powiązane zależnościami (podobieństwo, korelacja, ...) zakodowanymi w grafie.



Sygnaly na grafach – literatura

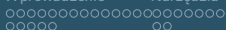
- 1 A. Sandryhaila, J.M.F. Moura: [Big data analysis with signal processing on graphs](#)
- 2 D. Shuman *et al.* [The emerging field of signal processing on graphs](#)
- 3 D.K. Hammond *eta al.* [Wavelets on graphs via spectral graph theory](#)
- 4 I. Pesenson: [Sampling in Paley-Wiener spaces on combinatorial graphs](#)



Prosta i odwrotna grafowa transformata Fouriera

$\mathcal{L} = \Psi \Lambda \Psi^T$ - normalizowany laplasjan grafu Γ .

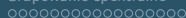
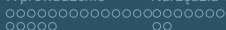
- 1 $\hat{\mathbf{f}} = \Psi^T \mathbf{f}$ – transformata Fouriera (GFT), czyli projekcja sygnału na wektory własne laplasjanu; $\hat{f}_i = \langle \mathbf{f}, \boldsymbol{\psi}_i \rangle$
- 2 $\mathbf{f} = \Psi \hat{\mathbf{f}}$ – odwrotna transformata Fouriera
- 3 $\langle \mathbf{f}, \mathbf{g} \rangle = \langle \hat{\mathbf{f}}, \hat{\mathbf{g}} \rangle$ i tw. Parsewala $\|\mathbf{f}\|^2 = \|\hat{\mathbf{f}}\|^2$.



Prosta i odwrotna grafowa transformata Fouriera

$\mathcal{L} = \Psi \Lambda \Psi^T$ - normalizowany laplasjan grafu Γ .

- 1 $\hat{\mathbf{f}} = \Psi^T \mathbf{f}$ – transformata Fouriera (GFT), czyli projekcja sygnału na wektory własne laplasjanu; $\hat{f}_i = \langle \mathbf{f}, \boldsymbol{\psi}_i \rangle$
- 2 $\mathbf{f} = \Psi \hat{\mathbf{f}}$ – odwrotna transformata Fouriera
- 3 $\langle \mathbf{f}, \mathbf{g} \rangle = \langle \hat{\mathbf{f}}, \hat{\mathbf{g}} \rangle$ i tw. Parsewala $\|\mathbf{f}\|^2 = \|\hat{\mathbf{f}}\|^2$.



Prosta i odwrotna grafowa transformata Fouriera

$\mathcal{L} = \Psi \Lambda \Psi^T$ - normalizowany laplasjan grafu Γ .

- 1 $\hat{\mathbf{f}} = \Psi^T \mathbf{f}$ – transformata Fouriera (GFT), czyli projekcja sygnału na wektory własne laplasjanu; $\hat{f}_i = \langle \mathbf{f}, \boldsymbol{\psi}_i \rangle$
- 2 $\mathbf{f} = \Psi \hat{\mathbf{f}}$ – odwrotna transformata Fouriera
- 3 $\langle \mathbf{f}, \mathbf{g} \rangle = \langle \hat{\mathbf{f}}, \hat{\mathbf{g}} \rangle$ i tw. Parsewala $\|\mathbf{f}\|^2 = \|\hat{\mathbf{f}}\|^2$.



Twierdzenie Whittaker-Nyquist-Kotelnikow-Shannona

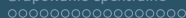
Kiedy z sygnału dyskretnego $\check{x}(t)$ złożonego z próbek danego sygnału ciągłego $x(t)$ można wiernie odtworzyć sygnał $x(t)$?

Jeżeli:

- 1 widmo $\hat{x}(q)$ sygnału $x(t)$ jest ograniczone do pewnego przedziału częstotliwości $[-\omega_N, +\omega_N]$, tzn. $|\hat{x}(q)| = 0$ dla $|q| > \omega_N$, oraz
- 2 częstotliwość próbkowania $q_s \geq 2\omega_N$

Aby tego dokonać potrzebny jest filtr o transmitancji:

$$h_{q_s}(q) = \begin{cases} 1 & \text{gdym } q \leq q_s \\ 0 & \text{w p.p.} \end{cases}$$



Kilka definicji

- 1 Próbkowanie sygnału grafowego \mathbf{f} na zbiór $Z \subset V$ polega na ograniczeniu sygnału do tego zbioru, $\mathbf{f}(Z) = \mathcal{D}_Z \mathbf{f}$, gdzie $\mathcal{D}_Z \in \{0, 1\}^{n \times m}$, $n = |Z|$.
Np. gdy $V = \{1, \dots, 4\}$, a $Z = \{1, 3\}$ to

$$\mathcal{D}_Z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathcal{D}_Z \mathbf{f} = \begin{bmatrix} f_1 \\ f_3 \end{bmatrix}$$

- 2 Sygnał dolnopasmowy (gładki) \mathbf{f} – taki, że $\hat{f}_i = 0$ dla $\lambda_i > \omega$.
Przestrzeń ω -ograniczonych (*bandlimited*) sygnałów to przestrzeń Paley-Wienera, $PW_\omega(\Gamma)$.
- 3 $L_2(Z^c) = \{\mathbf{f} \in \mathbb{R}^m : \mathbf{f}(Z) = 0\}$.

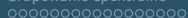
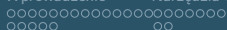


Kilka definicji

- 1 Próbkowanie sygnału grafowego \mathbf{f} na zbiór $Z \subset V$ polega na ograniczeniu sygnału do tego zbioru, $\mathbf{f}(Z) = \mathcal{D}_Z \mathbf{f}$, gdzie $\mathcal{D}_Z \in \{0, 1\}^{n \times m}$, $n = |Z|$.
Np. gdy $V = \{1, \dots, 4\}$, a $Z = \{1, 3\}$ to

$$\mathcal{D}_Z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathcal{D}_Z \mathbf{f} = \begin{bmatrix} f_1 \\ f_3 \end{bmatrix}$$

- 2 Sygnał dolnopasmowy (gładki) \mathbf{f} – taki, że $\hat{f}_i = 0$ dla $\lambda_i > \omega$.
Przestrzeń ω -ograniczonych (*bandlimited*) sygnałów to przestrzeń Paley-Wienera, $PW_\omega(\Gamma)$.
- 3 $L_2(Z^c) = \{\mathbf{f} \in \mathbb{R}^m : \mathbf{f}(Z) = 0\}$.

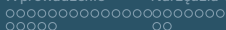


Kilka definicji

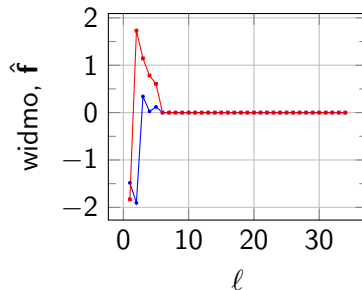
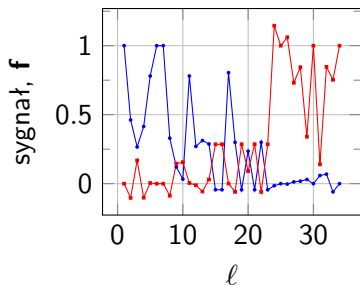
- 1 Próbkowanie sygnału grafowego \mathbf{f} na zbiór $Z \subset V$ polega na ograniczeniu sygnału do tego zbioru, $\mathbf{f}(Z) = \mathcal{D}_Z \mathbf{f}$, gdzie $\mathcal{D}_Z \in \{0, 1\}^{n \times m}$, $n = |Z|$.
Np. gdy $V = \{1, \dots, 4\}$, a $Z = \{1, 3\}$ to

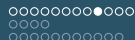
$$\mathcal{D}_Z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \mathcal{D}_Z \mathbf{f} = \begin{bmatrix} f_1 \\ f_3 \end{bmatrix}$$

- 2 Sygnał dolnopasmowy (gładki) \mathbf{f} – taki, że $\hat{f}_i = 0$ dla $\lambda_i > \omega$.
Przestrzeń ω -ograniczonych (*bandlimited*) sygnałów to przestrzeń Paley-Wienera, $PW_\omega(\Gamma)$.
- 3 $L_2(Z^c) = \{\mathbf{f} \in \mathbb{R}^m : \mathbf{f}(Z) = 0\}$.



Przykład sygnału dolnopasmowego





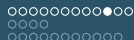
Filtr dolnoprzepustowy

„Transmitancja” filtru: $H = h(\mathcal{L}) = \Psi \text{diag}(h(\lambda_1), \dots, h(\lambda_m)) \Psi^T$,
gdzie $h: [0, 2] \rightarrow \mathbb{R}$ funkcja ciągła. Filtrowany sygnał s jest równy Hs .

Filtr dolnoprzepustowy

$$h_\omega(\lambda) = \begin{cases} 1 & \text{gd}y \lambda \leq \omega \\ 0 & \text{w p.p.} \end{cases}$$

tnzn. $H_\omega = \Psi_\kappa \Psi_\kappa^T$, $\Psi_\kappa = (\psi_1, \dots, \psi_\kappa)$, $\kappa = |\{\lambda: \lambda \leq \omega_c\}|$.

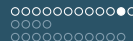


Filtr dolnoprzepustowy, $H_\omega = \Psi_\kappa \Psi_\kappa^T$ – Własności

- 1 $H_\omega^T = H_\omega$,
- 2 $\sigma(H_\omega) \{0^{m-\kappa}, 1^\kappa\} \Rightarrow \text{tr}(H_\omega) = \kappa \Rightarrow \mu_{[\lambda_i, \lambda_j]} = \text{tr}(\sum_{\lambda_i \leq \lambda_\ell \leq \lambda_j} \psi_\ell \psi_\ell^T)$,
- 3 $H_\omega \psi_i = \psi_i$ gdy $i \leq \kappa$, $H \psi_i = \mathbf{0}$ gdy $i > \kappa$

Zastosowanie (2) w:

E. Di Napoli, *et al.* Efficient estimation of eigenvalue counts in an interval. *arXiv*, abs/1308.4275, 2013



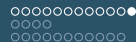
Filtr dolnoprzepustowy – aproksymacja

$$\tilde{h}(\lambda) = \sum_{\ell=0}^p \alpha_{\ell} \lambda^{\ell} \approx h(\lambda)$$

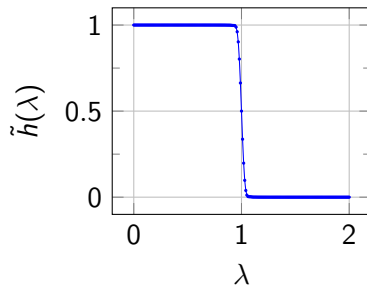
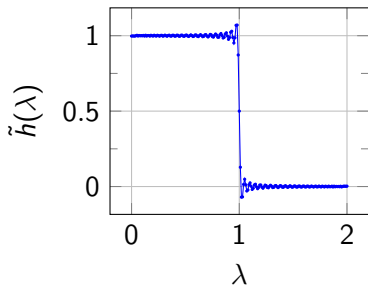
$$\tilde{H} = \tilde{h}(\mathcal{L}) = \Psi \text{diag}(\tilde{h}(\lambda_1), \dots, \tilde{h}(\lambda_m)) \Psi^T = \sum_{\ell=0}^p \alpha_{\ell} \mathcal{L}^{\ell}$$

$$\tilde{H}\mathbf{s} = \sum_{\ell=0}^p \alpha_{\ell} \mathcal{L}^{\ell} \mathbf{s} = \alpha_0 \mathbf{s} + \alpha_1 \mathcal{L} \mathbf{s} + \alpha_2 \mathcal{L}(\mathcal{L} \mathbf{s}) + \alpha_3 \mathcal{L}(\mathcal{L}(\mathcal{L} \mathbf{s})) + \dots$$

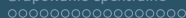
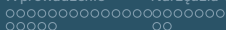
wymaga p -krotnego mnożenia macierzy \mathcal{L} przez wektor przeplatanego sumowaniem wektorów. **Nie trzeba diagonalizować macierzy \mathcal{L} .**



Filtr dolnoprzepustowy – aproksymacja



Rysunek: Aproksymacja Czebyszewa (lewy rys.) oraz Czebyszewa-Jacksona (prawy)



Wady algorytmu NJW

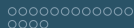
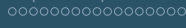
- 1 Konieczność wyznaczenia k dominujących wektorów własnych macierzy $\mathcal{S} = \mathbb{I} - \mathcal{L}$.
- 2 Algorytm k -średnich na m obserwacjach $\mathbf{y}_1, \dots, \mathbf{y}_m$.

Prace dotyczące tego tematu

- 1 D. Ramasamy, U. Madhow: [Compressive spectral embedding: sidestepping the SVD](#),
- 2 N. Tremblay, *et al.*: [Compressive spectral clustering](#),
- 3 G. Puy, *et al.*: [Random sampling of bandlimited signals on graphs](#)

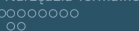
Problemy

- P1 Częstotliwość odcięcia (Nyquista): Dla danego $Z \subset V$ znaleźć częstotliwość odcięcia ω taką, że dowolny sygnał $\mathbf{f} \in PW_\omega(\Gamma)$ można zrekonstruować znając $\mathbf{f}(Z)$.
- P2 Optymalny zbiór próbkowania: dla danej ω znaleźć najmniejszy zbiór Z_{min} taki, że każdy sygnał $\mathbf{f} \in PW_\omega(\Gamma)$ można zrekonstruować znając $\mathbf{f}(Z)$.
- P3 Rekonstrukcja: mając $\mathbf{f}(Z)$ znaleźć $\mathbf{f}(Z^c)$.

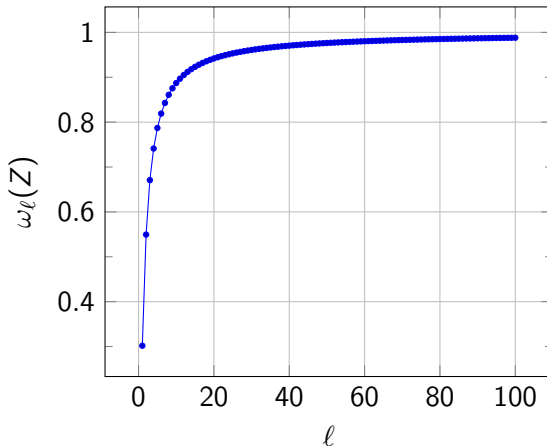


P1: Częstotliwość odcięcia

- 1 A. Anis, A. Gadde, A. Ortega: [Towards a sampling theorem for signals on arbitrary graphs](#)
- 2 Momenty spektralne: $\omega_k(\mathbf{f}) = \left(\frac{\mathbf{g}^T \mathcal{L}^k \mathbf{g}}{\mathbf{g}^T \mathbf{g}} \right)^{1/k}$.
 $\forall \mathbf{f}, k_1 < k_2: \omega_{k_1}(\mathbf{f}) \leq \omega_{k_2}(\mathbf{f}), \omega(\mathbf{f}) = \lim_{k \rightarrow \infty} \omega_k(\mathbf{f})$.
- 3 $\Omega_k(Z) = \inf_{\mathbf{g} \in L_2(Z^c)} \omega_k(\mathbf{g})$
- 4 $\omega_c(Z) \geq \lim_{k \rightarrow \infty} \Omega_k(Z) \geq \omega_{k_2}(\mathbf{Z}) \geq \omega_{k_1}(\mathbf{Z}), k_2 > k_1$
- 5 Niech $(\lambda_{1,k}, \psi_{1,k})$ para własna $\mathcal{L}^k(Z^c, Z^c)$. Wtedy $\Omega_k(Z) = \lambda_{1,k}, \mathbf{g}^*(Z^c) = \psi_{1,k}, \mathbf{g}^*(Z) = 0$.

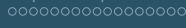


Momenty spektralne rzędu k sygnału f



Rysunek: 22

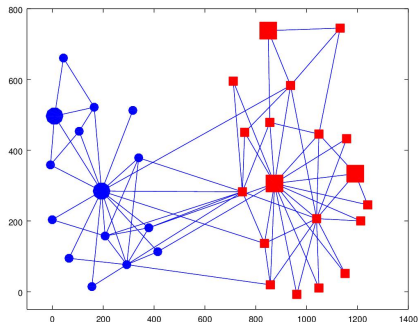




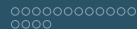
Wyznaczanie zbioru Z o liczebności z

- 1 $Z = \emptyset$
- 2 while $|Z| \leq z$ do
 - wyznacz najgładszy sygnał $\mathbf{g}_k^* \in L_2(Z^c)$
 - $v \rightarrow \arg \max |g_k^*(j)|$
 - $Z = Z \cup \{v\}$
- 3 end while
- 4 zwróć Z oraz $\omega = \lambda_{1,k}$

Wybór węzłów w sieci karate



Rysunek: Węzły wybrane do zamarkowania przez algorytm



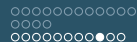
Rekonstrukcja sygnału (1)

- 1 S.K. Narang *et al.*: [Localized Iterative Methods for Interpolation in Graph Structured Data](#)

- 2 $\psi_i(Z) = (\psi_{z_1,i}, \dots, \psi_{z_3,i})^T$, $\Psi_\kappa(Z) = (\psi_1(Z), \dots, \psi_\kappa(Z))$

- 3
$$\begin{bmatrix} \mathbf{f}(Z) \\ \mathbf{f}(Z^c) \end{bmatrix} = \begin{bmatrix} \psi_1(Z) & \dots & \psi_\kappa(Z) \\ \psi_1(Z^c) & \dots & \psi_\kappa(Z^c) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_\kappa \end{bmatrix} = \begin{bmatrix} \Psi_\kappa(Z) \\ \Psi_\kappa(Z^c) \end{bmatrix} \alpha$$

- 4 $\mathbf{f}(Z) = \Psi_\kappa(Z)\alpha \Rightarrow \alpha = \Psi_\kappa(Z)^\dagger \mathbf{f}(Z)$
 $\mathbf{f}(Z^c) = \Psi_\kappa(Z^c) [(\Psi_\kappa(Z)^T \Psi_\kappa(Z))^{-1} \Psi_\kappa(Z)^T] \mathbf{f}(Z)$



Rekonstrukcja sygnału (2)

- 1 POCS (Projection onto Convex Sets):
- 2 zbiory wypukłe to $C_1 = \{\mathbf{x}: \mathcal{D}_Z \mathbf{x} = \mathcal{D}_Z \mathbf{f} = \mathbf{f}(S)\}$,
 $C_2 = PW_\omega(\Gamma)$
- 3 Niech $\mathfrak{P}_Z \mathbf{x} = \mathbf{x} + \mathcal{D}_Z^T (\mathbf{f}(Z) - \mathcal{D}_Z \mathbf{x})$
- 4 Wtedy: $\mathbf{f}_{t+1} = H_\omega \mathfrak{P}_Z \mathbf{f}_t$, $t = 0, 1, \dots$ oraz $\mathbf{f}_0 = H_\omega \mathfrak{P}_Z \mathbf{f}(Z)$

Algorytm

- 1 Wyznacz maksymalną i k -tą wartość własną laplasjanu.
- 2 Znajdź wielomianową aproksymację $\tilde{h}(\lambda)$ stopnia p
- 3 Wygeneruj macierz R
- 4 Filtracja $F = \left(\sum_{\ell=0}^p \alpha_{\ell} L^{\ell} \right) R$
- 5 Grupuj dane F (ew. grupowaniu poddawany jest podzbiór $F' \in \mathbb{R}^{n \times k}$, $n \ll m$)

Dziękuję za uwagę