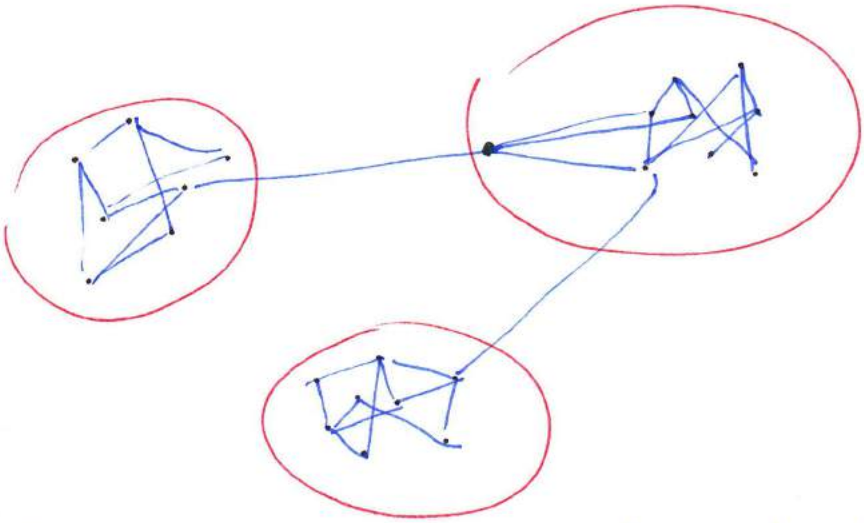
Clustering II



"Community Detection"

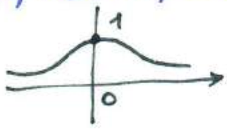
Objekte im Netzwerk

Linkage / Gewichtung der Verbindung

z.B.
- Personen im sozialen Netzwerk

  w_{ij} = Anzahl von Kontakten zwischen i und j
- Objekte mit einem Ähnlichkeitsmaß (Waren, Zeilen, Zeitreihen etc.)

  $w_{ij} = s(i,j)$ → similarity

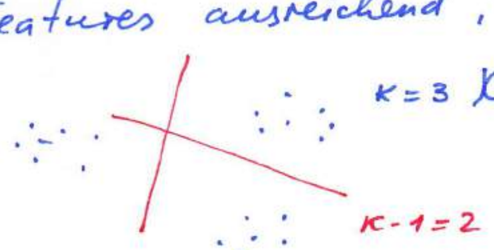


- Daten $x_i \in \mathbb{R}^p$, $i = 1, \dots, n$ (p → Anzahl von Merkmalen)

$$w_{ij} = K(d(x_i, x_j)) \underset{\text{z.B.}}{=} e^{-\frac{\|x_i - x_j\|^2}{2 \cdot \sigma}} \quad : \quad \begin{array}{l} x_i = x_j \Rightarrow w_{ij} = 1 \\ x_i \neq x_j \Rightarrow w_{ij} < 1 \end{array}$$

K: Kernel, d: Dissimilarity, Gauß-Kernel

$\xrightarrow{\quad}$ k-Clustering: $(k-1)$ Features ausreichend, aber p Merkmale vorhanden



Ansatz: Informationsverbreitung im Netzwerk

↪ Zwei Personen sind ähnlich, wenn sie in etwa den selben Informationszustand im Netzwerk über die längere Zeit auslösen

$\boxed{\text{Spectral Clustering}}$

Informationsverbreitung

$W := (w_{ij}) \in \mathbb{R}^{n \times n}$ - symmetrische Gewichtungsmatrix

$D := \operatorname{diag}\Big(\underbrace{\sum_{i=1}^{n} w_{ij}}_{\text{Grad von } j \,=\, \text{Gesamtgewichtung von } j},\ j = 1, \dots, n\Big)$

$P := W \cdot D^{-1} = \left(\dfrac{w_{ij}}{\sum_{i=1}^{n} w_{ij}} \right)$ ← Wahrscheinlichkeit des Übergangs von j nach i

↑ Übergangsmatrix

$\Rightarrow \underbrace{e^T \cdot P = e^T, \ P \geq 0}$

P - stochastische Matrix, d.h. Spaltensumme ist 1.

$x(0) = e_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ ← j-te Stelle "Information wird von der Person j ausgelöst"

$x^j(t) = P \cdot x(t-1) = P^2 x(t-2) = \dots = P^t \cdot x(0) = P^t \cdot e_j$

Information zum Zeitpunkt t — Information zum Zeitpunkt $t-1$

$\Rightarrow x^j(t) = P^t \cdot e_j$ ← Informationszustand

P^t: t-te Potenz von P

Struktur von P^t:

- $S := D^{-1/2} \cdot W \cdot D^{-1/2}$ - symmetrisch $\Rightarrow S v_i = d_i \cdot v_i, \ i = 1, \dots, n$

  (d_i: Eigenwerte; v_i: Eigenvektoren sind paarweise orthogonal, d.h. $v_j^T v_i = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$)

Also: $v_j^T \cdot S \cdot v_i = v_j^T \cdot d_i \cdot v_i = \begin{cases} d_i, & i = j \\ 0, & i \neq j \end{cases}$

$\Rightarrow V^T \cdot S \cdot V = \underbrace{\Lambda}$

$\left(\begin{array}{l} = \operatorname{diag}(|d_1| \geq \dots \geq |d_n|) \\ \text{o.B.d.A. monoton fallend sortiert} \end{array} \right)$

Mit $V = (v_1, \dots, v_n)$ gilt:

- $V^T V = I \Rightarrow V^T \underbrace{V \cdot V^{-1}}_{I} = V^{-1}$
- $V \cdot V^T = V \cdot V^{-1} = I$

$\Rightarrow \boxed{S = V \cdot \Lambda \cdot V^T}$

- $P = W \cdot D^{-1} = D^{1/2} \cdot \underbrace{D^{-1/2} \cdot W \cdot D^{-1/2}}_{= S} \cdot D^{-1/2} = \underbrace{D^{1/2} \cdot V}_{= \Phi} \cdot \Lambda \cdot \underbrace{V^T \cdot D^{-1/2}}_{= \Psi}$

$\Rightarrow \boxed{P = \Phi \cdot \Lambda \cdot \Psi}$

$\Phi = \begin{pmatrix} | & \dots & | \\ \varphi_1 & \dots & \varphi_n \end{pmatrix}$ Spalten, $\quad \Psi = \begin{pmatrix} \text{—} \ \psi_1^T \\ \vdots \\ \text{—} \ \psi_n^T \end{pmatrix}$ Zeilen

und $\underbrace{\Psi \cdot \Phi} = V^T \cdot D^{-1/2} \cdot D^{1/2} V = V^T V = I$

bi-orthonormale Matrizen, d.h. $\psi_i^T \cdot \varphi_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$

$$\cdot\ P^t = \Phi\cdot\Lambda\cdot\underbrace{\Psi\cdot\Phi}_{=I}\cdot\Lambda\cdot\underbrace{\Psi\cdot\ \ldots\ \cdot\Phi}_{I\qquad I}\cdot\Lambda\cdot\Psi = \Phi\cdot\Lambda^t\cdot\Psi$$ 3

$$\Rightarrow\ x^j(t) = P^t\cdot e_j = \Phi\cdot\Lambda^t\cdot\Psi\cdot e_j = \Phi\cdot\begin{pmatrix}\lambda_1^t & & 0\\ & \ddots & \\ 0 & & \lambda_n^t\end{pmatrix}\cdot\begin{pmatrix}\psi_{1j}\\ \vdots\\ \psi_{nj}\end{pmatrix} = \Phi\cdot\underbrace{\begin{pmatrix}\lambda_1^t\psi_{1j}\\ \vdots\\ \lambda_n^t\psi_{nj}\end{pmatrix}}_{=:F^j(t)}$$

$$\Psi = \begin{pmatrix}\psi_1^T & \text{—}\\ \vdots & \\ \psi_n^T & \text{—}\end{pmatrix}$$

j-te Spalte von Ψ

Insgesamt: $x^j(t) = \Phi\cdot F^j(t)$ "Diffusion Map"

$$\|D^{-1/2}\cdot\underbrace{(x^j(t)-x^l(t))}\|^2 = \|D^{-1/2}\cdot\Phi\cdot(F^j(t)-F^l(t))\|^2 =$$

Gewichtung der Norm

Unterschied im Informationszustand ausgelöst durch j bzw. l

$z^T\cdot z = \|z\|^2$ – Euklidische Norm

$$= \|\underbrace{D^{-1/2}\cdot D^{1/2}}_{=I}\cdot V\cdot\underbrace{(F^j(t)-F^l(t))}_{=a}\|^2 = (V\cdot a)^T\cdot V\cdot a = a^T\cdot\underbrace{V^T\cdot V}_{=I}\cdot a =$$

$$= \|a\|^2 = \underbrace{\|F^j(t)-F^l(t)\|^2}.$$

"Diffusion Distance"

Reduktion der Dimension

Eigenwerte / Eigenvektoren

Eigenwert von P ; Eigenvektor von P

$$\cdot\ P\cdot\varphi_j = \Phi\cdot\Lambda\cdot\underbrace{\Psi\cdot\varphi_j}_{=e_j} = \Phi\cdot e_j\cdot\lambda_j = \lambda_j\cdot\varphi_j,\ j=1,\ldots,n$$

$\Rightarrow$ (ü) $|\lambda_j| \le 1$, insbesondere $\lambda_1 = 1$ und φ_1 – Google-Ranking

"alle Eigenwerte einer stochastischen Matrix sind betragsmäßig kleiner als 1 oder gleich 1"

$$\cdot\ \psi_i^T\cdot P = \underbrace{\psi_i^T\cdot\Phi}_{e_i^T}\cdot\Lambda\cdot\Psi = \lambda_i\cdot e_i^T\cdot\Psi = \lambda_i\cdot\psi_i^T \Rightarrow \lambda_1 = 1 \text{ und } \psi_1^T = e^T.$$

$e^T\cdot P = e^T$

$\Longrightarrow$

- erste Komponente von $F^j(t)$ ist 1, also irrelevant:

$$\lambda_1^t\cdot\psi_{1j} = 1^t\cdot 1 = 1 \quad \forall j=1,\ldots,n$$

- letzte Komponenten von $F^j(t)$ sind klein, also vernachlässigbar:

$$\underbrace{\lambda_{k+1}^t}_{\in(0,1)}\cdot\psi_{k+1\,j}\to 0,\ \ldots,\ \underbrace{\lambda_n^t}_{\in(0,1)}\cdot\psi_{nj}\to 0 \quad \underbrace{\text{für } t\to\infty}$$

k-Clustering

langfristige Informationsverbreitung

Reduktion auf $\quad k\text{-}F^j(t) = \begin{pmatrix} \lambda_2^t \cdot \psi_{2j} \\ \vdots \\ \lambda_k^t \cdot \psi_{kj} \end{pmatrix} \in \mathbb{R}^{k-1}$

Anzahl von Features

k-Diffusion Map

Spectral Clustering

① Wähle t groß genug und konstruiere k-Diffusion Map:

$$\underbrace{\begin{array}{ccc} 1 & & n \\ \downarrow & \dots & \downarrow \\ k\text{-}F^1(t) & & k\text{-}F^n(t) \end{array}}_{} \in \mathbb{R}^{k-1}$$

② Wende auf n Datensätze in $\mathbb{R}^{k-1}$ k-Means an.

(bzgl. der euklidischen Norm)

$\approx$ "Diffusion Distance"