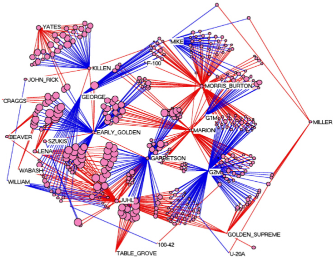
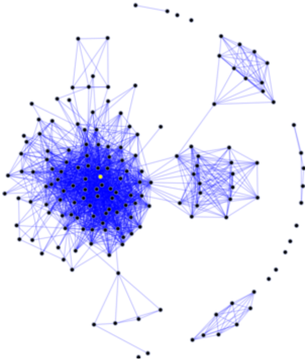
Partitioning/Coarse Graining and Embedding of Networks

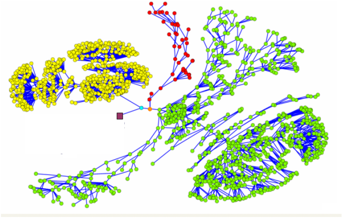
Christof Schuette

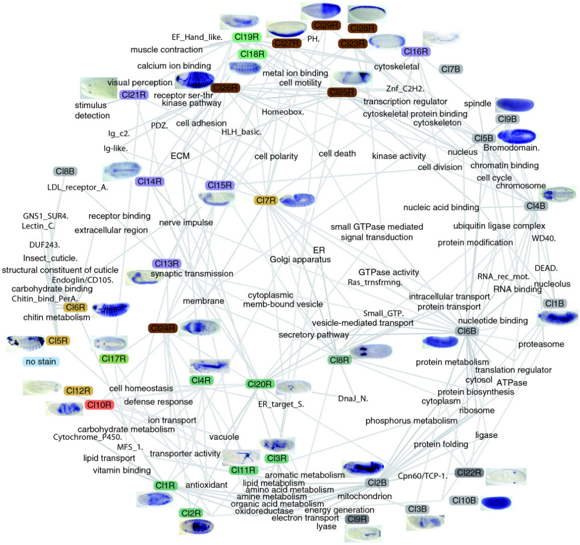
DFG Research Center MATHEON
Mathematics for key technologies

Networks


YATES
KILLEN
MIKE
F-100
JOHN_RICK
MORRIS_BURTON
CRAGGS
GEORGE
G1M
MILLER
BEAVER
EARLY_GOLDEN
MARION
SZUKIS
LENA
GARRETSON
WABASH
G2M
WILLIAM
JUHL
GOLDEN_SUPREME
100-42
TABLE_GROVE
U-20A






EF_Hand_like.
Cl19R
muscle contraction
Cl18R
Cl27R
PH.
Cl29R
Cl28R
Cl23R
Cl16R
calcium ion binding
metal ion binding
cell motility
cytoskeletal
Cl7B
visual perception
Cl26R
Cl25R
Znf_C2H2.
Cl21R
receptor ser-thr
kinase pathway
stimulus
detection
Homeobox.
transcription regulator
cytoskeletal protein binding
cytoskeleton
spindle
Cl9B
PDZ.
cell adhesion
HLH_basic.
Ig_c2.
Ig-like.
Cl5B
Bromodomain.
ECM
cell polarity
cell death
kinase activity
nucleus
cell division
chromatin binding
cell cycle
chromosome
Cl8B
LDL_receptor_A.
Cl14R
Cl15R
Cl7R
nucleic acid binding
Cl4B
GNS1_SUR4.
Lectin_C.
receptor binding
extracellular region
nerve impulse
small GTPase mediated
signal transduction
ubiquitin ligase complex
protein modification
WD40.
DUF243.
Insect_cuticle.
structural constituent of cuticle
Endoglin/CD105.
carbohydrate binding
Chitin_bind_PerA.
Cl13R
synaptic transmission
ER
Golgi apparatus
GTPase activity
Ras_trnsfrmng.
DEAD.
RNA_rec_mot.
RNA binding
nucleolus
Cl6R
chitin metabolism
membrane
cytoplasmic
memb-bound vesicle
intracellular transport
protein transport
Cl1B
Small_GTP.
vesicle-mediated transport
nucleotide binding
Cl6B
Cl24R
secretory pathway
proteasome
Cl5R
no stain
Cl17R
Cl8R
protein metabolism
translation regulator
cytosol
ATPase
Cl12R
Cl4R
Cl20R
cell homeostasis
Cl10R
defense response
ER_target_S.
DnaJ_N.
protein biosynthesis
cytoplasm
ribosome
ion transport
carbohydrate metabolism
Cytochrome_P450.
phosphorus metabolism
vacuole
MFS_1.
lipid transport
transporter activity
Cl3R
protein folding
ligase
vitamin binding
Cl11R
aromatic metabolism
Cl2B
Cpn60/TCP-1.
Cl22R
Cl1R
antioxidant
lipid metabolism
amino acid metabolism
amine metabolism
mitochondrion
Cl10B
organic acid metabolism
oxidoreductase
energy generation
Cl3B
Cl2R
electron transport
Cl9R
lyase

Contents

Contents

Edge Weights and Transition probabilities

Undirected, connected network: Graph $G = (V, E)$, $V = \{1, \dots, n\}$, $E \subset V \times V$

Edge weights:

$$w(x,y) \begin{cases} > 0 & \text{if} \quad (x,y) \in E \\ = 0 & \text{if} \quad (x,y) \notin E \end{cases}$$

Transition Probabilities:

$$d(x) = \sum_y w(x,y), \qquad p(x,y) = \frac{w(x,y)}{d(x)}.$$

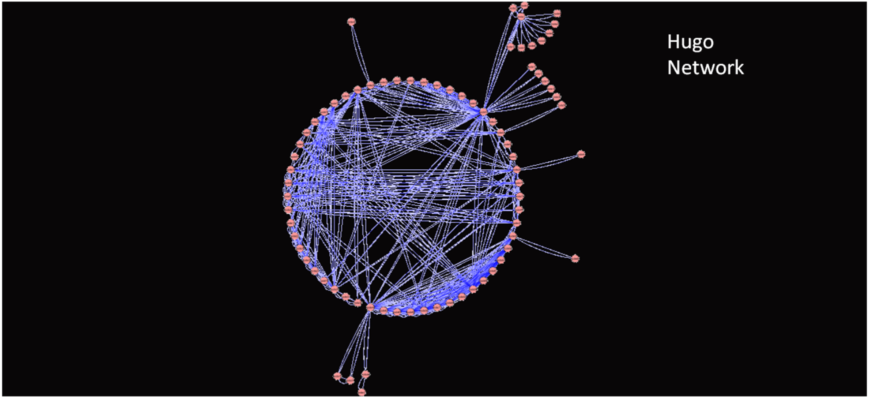
Random Walker on the Network!

Victor Hugo: Les Miserables.

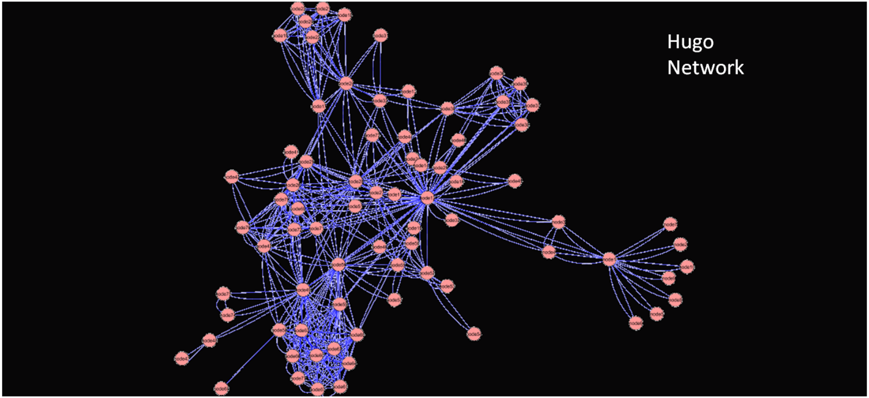
Weighted network given by coappearances of characters in Victor Hugo's novel "Les Miserables".

Each character is represented by a node in the graph and we put an edge connecting two nodes, if the corresponding characters appear in the same chapter. Moreover each edge is weighted by the number of coappearances. The network consists of 77 nodes and 254 weighted edges.

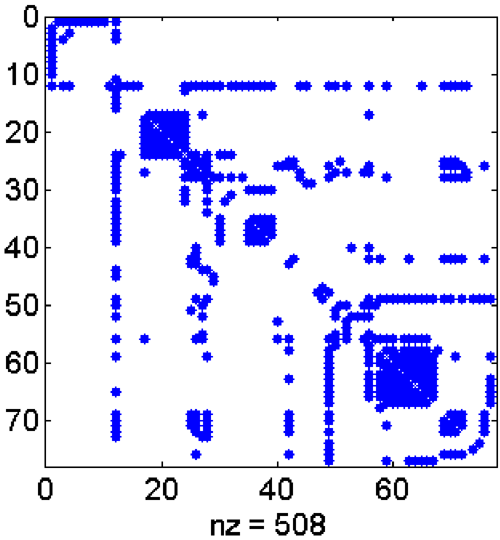
Hugo Network II



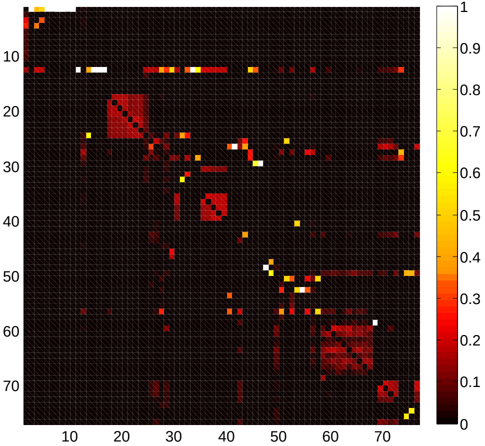
Hugo Network III



Example: Hugo-Network IV



$$\lambda = \ldots, -0.7269, 0.7746, 0.8337, 0.8862, 0.9270, 1.0000$$



Invariant measure

$$\mu(x) = \frac{d(x)}{\sum_x d(x)} \quad \Rightarrow \quad \sum_x \mu(x)p(x,y) = \mu(y).$$

Undirected network:

$$w(x,y) = w(y,x) \quad \Rightarrow \quad \mu(x)p(x,y) = \mu(y)p(y,x).$$

Markov chain $p(x,y)$ is reversible!

Markov Chain Aspects II

Propagation of probability distributions $u(x), x \in V$

$$(u^T P)(y) = \sum_x u(x)p(x, y)$$

Scalar product

$$\langle u, v \rangle_\mu = \sum_x u(x)v(x)\mu(x).$$

Reversibility implies symmetry of P:

$$\langle u, Pv \rangle_\mu = \langle Pu, v \rangle_\mu.$$

Eigenvalue of P are real-valued!

$$\sigma(P) = \{1 = \lambda_1, \lambda_2, \ldots, \lambda_n\} \subset (-1, 1]$$

Left and right eigenvectors

$$\psi_k^T P = \lambda_k \psi_k, \quad P\phi_k = \lambda_k \phi_k$$

Diagonalization of P

$$p(x, y) = \sum_{k=1}^{n} \lambda_k \phi(x) \psi_k(y)$$

Simple algebra

$$\psi_k(y) = \mu(y)\phi(y)$$

Consequence

$$p(x, y) = \sum_{k=1}^{n} \lambda_k \phi(x) \phi_k(y) \mu(y)$$

Transition probabilities for larger times t

$$p_{t+1}(x, y) = \sum_z p(x, z) p_t(z, y), \quad p_1(x, y) = p(x, y).$$

Consequence

$$p_t(x, y) = \sum_{k=1}^{n} \lambda_k^t \phi(x) \phi_k(y) \mu(y)$$

Since $\phi_1 = \mathbf{1}$ and $1 = |\lambda_1| > \lambda_k$ for all $k > 1$:

$$p_t(x, y) \to \mu(y), \quad \text{for } t \to \infty$$

Metric on Networks: Frobenius Norm

Definition of Frobenius Norm of transition probabilities $p(x, y)$

$$\|p\|_\mu^2 = \sum_{xy} \frac{\mu(x)}{\mu(y)} |p(x, y)|^2$$

Diagonalization implies

$$\|p\|_\mu^2 = \sum_k \lambda_k^2$$

Distance of some other networks p_2 from our given network p

$$E(p_2) = \|p_2 - p\|_\mu^2$$

Contents

N-set partitions:

$$S_1, \dots, S_N \subset V, \quad \cup_j S_j = V, \quad i \neq j \Rightarrow S_i \cap S_j = \emptyset$$

Set of all N-set partitions: $\mathcal{S}_N$
Coarse grained state space: $\hat{V} = \{1, \dots, N\}$

Let $\hat{p}$ be any stochastic matric on $\hat{V}$.

Lift of $\hat{p}$ to V (reversible if $\hat{p}$ reversible)

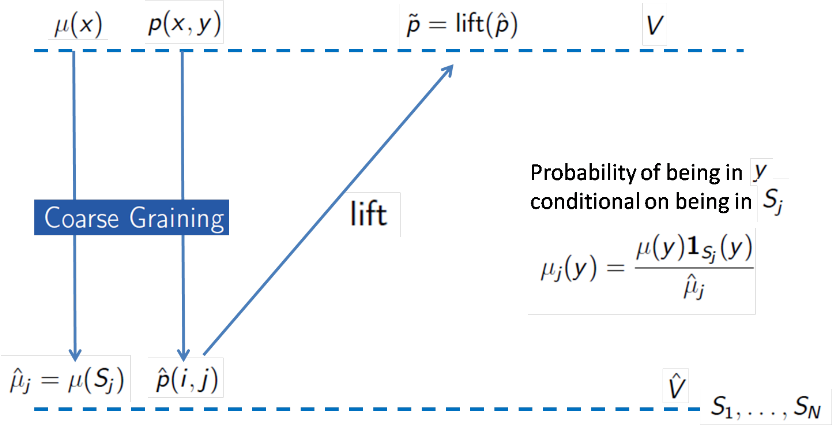
$$\tilde{p} = \text{lift}(\hat{p}), \quad \tilde{p}(x, y) = \sum_{ij} \mathbf{1}_{S_i}(x) \hat{p}(i, j) \mu_j(y)$$

Choose y from S_j due to invariant measure

$$\mu_j(y) = \frac{\mu(y) \mathbf{1}_{S_j}(y)}{\hat{\mu}_j}, \quad \hat{\mu}_j = \mu(S_j) = \sum_{x \in S_j} \mu(x)$$

Coarse Graining / Partitioning II

$$\tilde{p} = \mathrm{lift}(\hat{p}), \quad \tilde{p}(x,y) = \sum_{ij} \mathbf{1}_{S_i}(x)\hat{p}(i,j)\mu_j(y)$$



Optimal $\hat{p}$ on $\hat{V}$ for given p on V

Energy functional

$$E(\hat{p}) = \|p - \text{lift}(\hat{p})\|_\mu^2$$

Optimal coarse grained transition matrix = lift has smallest distance to p

$$\hat{p}^* = \text{argmin}_{\hat{p}} E(\hat{p})$$

with *constraints*

1. $\hat{p}$ stochastic matrix, i.e., $\hat{p}(i,j) \geq 0, \qquad \sum_j \hat{p}(i,j) = 1$
2. $\hat{\mu}$ invariant measure of $\hat{p}$, i.e., $\sum_i \hat{\mu}(i)\hat{p}(i,j) = \hat{\mu}(j)$

Explicit solution

$$\hat{p}^*(i,j) \;=\; \sum_{x \in S_i, y \in S_j} \mu_i(x) p(x,y) = \frac{\langle P\mathbf{1}_{S_i}, \mathbf{1}_{S_j}\rangle_\mu}{\langle \mathbf{1}_{S_i}, \mathbf{1}_{S_i}\rangle_\mu}$$

Optimal Aggregation for given network p on V

Functional

$$E(S_1,\dots,S_N) = E(\hat{p}^*[S_1,\dots,S_N]) = \|p\|_\mu^2 - \|\hat{p}^*[S_1,\dots,S_N]\|_{\hat{\mu}}^2$$

Optimal Partition

$$[S_1^*,\dots,S_N^*] = \mathrm{argmin}_{[S_1,\dots,S_N]} E(S_1,\dots,S_N)$$

Iterative computation via a variant of K-means!

Definition:
$p(x, y)$ is (perfectly) lumpable wrt. $S_1, \ldots, S_N$ iff the random walker associated with $p(x, y)$ is Markov on $S_1, \ldots, S_N$

Minimal energy: Minimization over ALL rank-N $n \times n$-matrices $\tilde{p}$

$$E^* = \min_{\tilde{p}} E(\tilde{p}) = \min_{\tilde{p}} \|p - \tilde{p})\|_\mu^2$$

Theorem:

1. $E(\hat{p}^*[S_1, \ldots, S_N]) \geq E^*$
2. $E(\hat{p}^*[S_1, \ldots, S_N]) = E^*$ iff $p(x, y)$ is lumpable wrt. $S_1, \ldots, S_N$

$n = 4$ and transition matrix (eigenvalues: $\lambda = 1, 0, 0, -1$)

$$P = \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix}$$

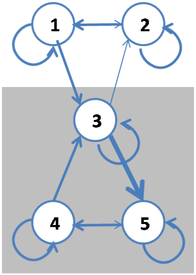
Optimal coarse graining for $N = 2$: $S_1 = \{1, 3\}$ and $S_2 = \{2, 4\}$

$$\hat{P}^* = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

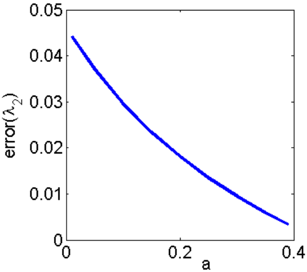
$E(\hat{p}^*) = E^*$, i.e., perfectly lumpable into twp PERIODIC clusters.

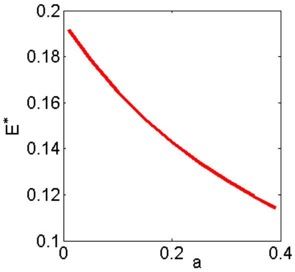
$$p = \begin{pmatrix} 0.5 & 0.4 & 0.1 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & a & 0.2 & 0 & 0.8-a \\ 0 & 0 & 0.1 & 0.4 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix},$$

$$\begin{aligned} \|p\|_\mu^2 &\approx 2 \\ S_1, S_2 &= \{1,2\},\{3,4,5\} \\ \mathrm{err}(\lambda) &= \lambda_2(p) - \lambda_2(\hat{p}) \\ E^* &= \min \|p-\hat{p}\|_\mu^2 \end{aligned}$$



Diagram





Deterministic affiliation functions:

$$W(x,i) = \begin{cases} 1 & \text{if} \quad C(x) = i \quad x \in S_i \\ 0 & \text{if} \quad C(x) \neq i \quad x \notin S_i \end{cases}$$

"Fuzzy" affiliation functions: $C : V \to V^*$ is a random variable

$$W(x,i) = \mathrm{Prob}(C(x) = i)$$

Therefore:

$$\begin{aligned} W(x,i) &\geq 0 \\ \sum_i W(x,i) &= 1 \end{aligned}$$

Generalization: Fuzzy Partition II

Coarse grained transition matrix on $\hat{V}$: $\hat{p} = \hat{p}(i, j)$

Lift of $\hat{p}$ wrt. fuzzy affiliation function W

$$\begin{aligned}\text{lift}_W(\hat{p})(x, y) &= \text{Expect}\left(\hat{p}(C(x), C(y)) \cdot \frac{\mu(y)}{\hat{\mu}(C(y))}\right) \\ &= \sum_{ij} W(x, i)\hat{p}(i, j) W(y, j) \frac{\mu(y)}{\hat{\mu}(j)}\end{aligned}$$

with

$$\hat{\mu}(i) = \sum_x \mu(x) W(x, i)$$

Lift of $\hat{p}$ wrt. fuzzy affiliation function W

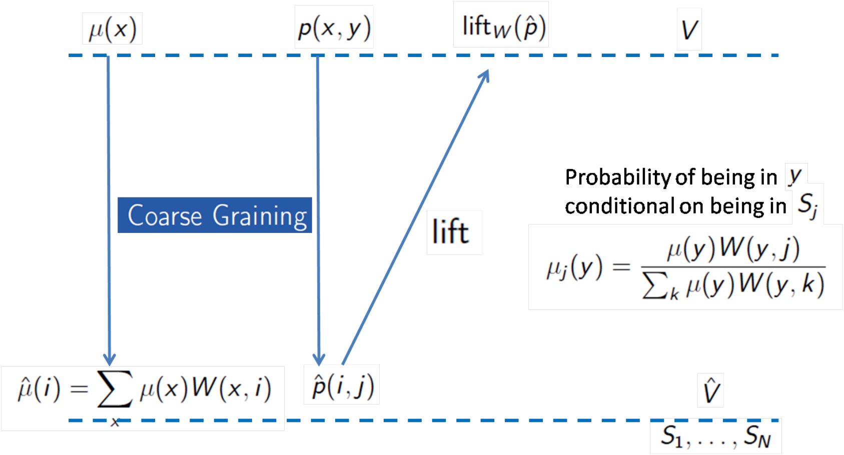
$$\text{lift}_W(\hat{p})(x,y) = \sum_{ij} W(x,i)\hat{p}(i,j)W(y,j)\frac{\mu(y)}{\hat{\mu}(j)}$$

1. Being in state x, pick i wrt. probability $W(x,\cdot)$
2. Pick j wrt. probability $\hat{p}(i,\cdot)$
3. Pick y wrt. the equilibirum probability distribution of the original chain conditional on $C(y) = j$,

$$\mu_j(y) = \frac{\mu(y)W(y,j)}{\sum_k \mu(y)W(y,k)}$$

Fuzzy Partition - Interpretation

$$\mathrm{lift}_W(\hat{p})(x,y) = \sum_{ij} W(x,i)\hat{p}(i,j)W(y,j)\frac{\mu(y)}{\hat{\mu}(j)}$$



Optimal Fuzzy Aggregation

Energy functional

$$E(\hat{p}, W) = \|p - \mathrm{lift}_W(\hat{p})\|_\mu^2$$

Optimal coarse grained transition matrix = lift has smallest distance to p

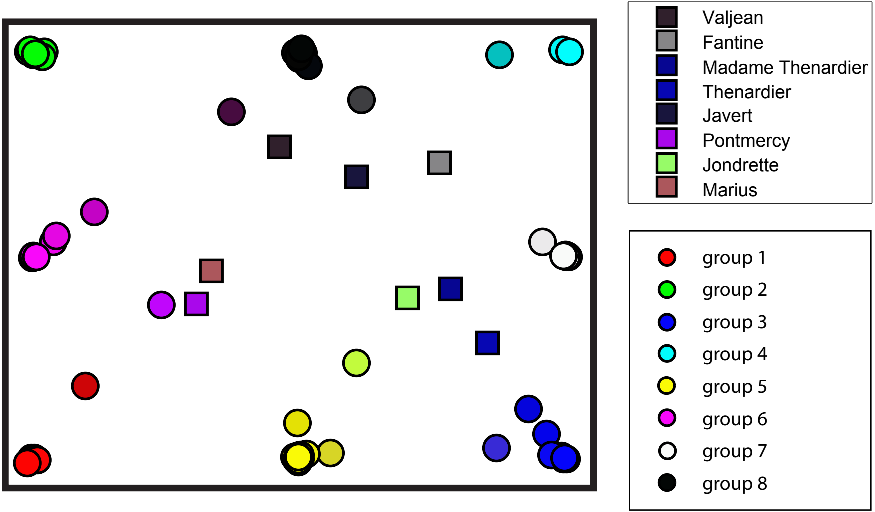
$$(\hat{p}^*, W^*) = \mathrm{argmin}_{\hat{p},W} E(\hat{p}, W)$$

with *constraints*

1. Fuzzy affiliation functions: $W(x,i) \geq 0, \quad \sum_i W(x,i) = 1$
2. $\hat{p}$ stochastic matrix, i.e., $\hat{p}(i,j) \geq 0, \quad \sum_j \hat{p}(i,j) = 1$
3. $\hat{\mu}$ is coarse graining of μ: $\hat{\mu}(i) = \sum_x W(x,i)\mu(x)$

No explicit solution!

Constrained (projected) gradient descent algorithm!


Valjean
Fantine
Madame Thenardier
Thenardier
Javert
Pontmercy
Jondrette
Marius
group 1
group 2
group 3
group 4
group 5
group 6
group 7
group 8

group 1:
Bamatabois
Judge
Champmathieu
Brevet
Chenildieu
Coch epaille

group 2:
Myriel
Napoleon
Mlle Baptistine
Mme Magloire
CountessDeLo
Geborand
Champtercier
Cravatte
Count
OldMan

group 3:
Boulatruelle
Eponine
Anzelma
Gueulemer
Babet
Claquesous
Montparnasse
Brujon

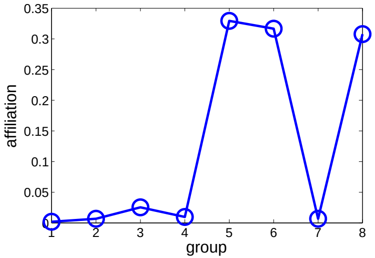
group 4:
Fauchelevent
MotherInnocent
Gribier

group 5:
Gillenormand
Magnon
Mlle Gillenormand
Mme Pontmercy
Mlle Vaubois
Lt Gillenormand
Baroness T

group 6:
Mme Burgon
Gavroche
Mabeuf
Enjolras
Combeferre
Prouvaire
Feuilly
Courfeyrac
Bahorel
Bossuet
Joly
Grantaire
Mother Plutarch
Child1
Child2
Mme Hucheloup

group 7:
Tholomyes
Listolier
Fameuil
Blacheville
Favourite
Dahlia
Zephine

group 8:
Labarre
Marguerite
Mme De R
Isabeau
Gervais
Cosette
Perpetue
Simplice
Scaufflaire
Woman1
Woman2
Toussaint



Example: Hugo-Network IX

group 1:
Bamatabois
Judge
Champmathieu
Brevet
Chenildieu
Coch epaille

group 5:
Jondrette
MmeBurgon
Gavroche
Mabeuf
Enjolras
Combeferre
Prouvaire
Feuilly
Courfeyrac
Bahorel
Bossuet
Joly
Grantaire
MotherPlutarch
MmeHucheloup

group 2:
Myriel
Napoleon
MlleBaptistine
MmeMagloire
CountessDeLo
Geborand
Champtercier
Cravatte
Count
OldMan
Marguerite
Tholomyes
Listolier
Fameuil
Blacheville
Favourite
Dahlia
Zephine
Fantine

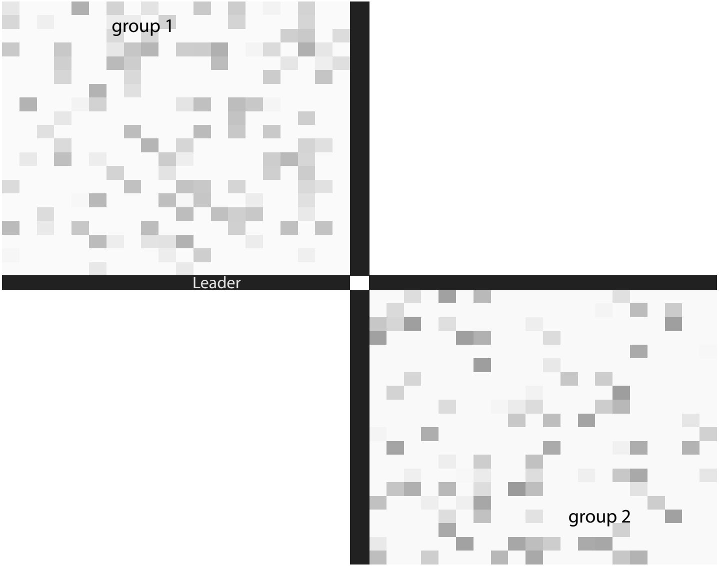
group 6:
Child1
Child2

group 3:
MmeThenardier
Javert
Anzelma
MotherInnocent
MmePontmercy
Marius
Claquesous
Brujon

group 7:
Valjean
Perpetue
Pontmercy
Boulatruelle
Eponine
Gribier
Gillenormand
MlleVaubois
LtGillenormand
Babet
Toussaint

group 4:
Labarre
MmeDeR
Isabeau
Gervais
Thenardier
Cosette
Fauchelevent
Simplice
Scaufflaire
Woman1
Woman2
Magnon
MlleGillenormand
BaronessT
Gueulemer
Montparnasse

Example: Hugo-Network X



Functional using

$$\mathcal{D}(S_1, \ldots, S_N) = \sum_{i=1}^{N} \hat{p}^*(i, i)$$

with

$$\hat{p}^* = \hat{p}^*[S_1, \ldots, S_N] = \sum_{x \in S_i, y \in S_j} \mu_i(x) p(x, y)$$

Optimal Partition

$$[S_1^*, \ldots, S_N^*] = \text{argmax}_{[S_1, \ldots, S_N]} \mathcal{D}(S_1, \ldots, S_N).$$

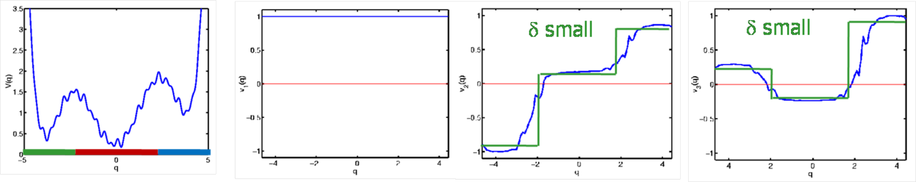
Iterative computation via a variant of K-means!

Diffusion maps + our estimates

Sets: $S_1, \ldots, S_N$

Q: Projection onto $D_N = \text{span}\{\chi_{S_1}, \ldots, \chi_{S_N}\}$

$\delta_j = \|(Id - Q)\phi_j\|_{2,\mu}$: non-steplikeness of eigenvectors



Theorem:

For ANY N aggregation sets $S_1, \ldots, S_N$

$$(1-\delta_n)^2 \cdot \lambda_n + \ldots + \lambda_1 + c \leq \sum_{i=1}^{N} \hat{p}^*(i,i) \leq \lambda_n + \ldots + \lambda_1$$

with

$$c = -\lambda_{N+1}(\delta_n^2 + \ldots + \delta_2^2)$$

To find optimal aggregation sets minimize $\delta = \max_j \delta_j$.

Diffusion distance and eigenvectors

$$d^2(x,y) = \sum_{z \in V} \frac{(p(x,z) - p(y,z))}{\mu(z)} = \sum_k \lambda_k^2 \cdot (\phi_k(x) - \phi_k(y))^2$$

Diffision map $\Phi : V \to \mathbf{R}^n$ introduces new "coordinates"

$$\Phi(x) = (\lambda_1 \phi_1(x), \ldots, \lambda_n \phi_n(x))$$

Centroid of subsets $S_i \subset V$

$$\phi_k(S_i) = \frac{1}{\mu(S_i)} \sum_{y \in S_i} \mu(y) \phi_k(x) \quad (1)$$

$$\Phi(S_i) = (\lambda_1 \phi_1(S_i), \ldots, \lambda_n \phi_n(S_i)) \quad (2)$$

Diffusion maps - better known version II

Average diffusion distance from centroid within S_i

$$\bar{d}^2(S_i) = \sum_{x \in S_i} \mu(x) \|\Phi(x) - \Phi(S_i)\|^2$$

Diffusion map minimization

$$\min_{[S_1,\ldots,S_N]} \sum_{i=1}^{N} \bar{d}^2(S_i)$$

minimization via k-means (only) if eigenvector ϕ_k are known!

Diffusion map minimization is equivalent to

$$\max_{[S_1,\ldots,S_N]} \sum_{x \in S_i, y \in S_j} \mu_i(x) p(x, y).$$

Nodes are associated with coordinates

$$V = (x_1, \ldots, x_n)$$

Edge weights

$$w(x_i, x_j) = \begin{cases} \exp(-\frac{\|x_i - x_j\|^2}{2\epsilon^2}) & \text{if} \quad \|x_i - x_j\| \leq \epsilon \text{ and } (x_i, x_j) \in E \\ 0 & \text{if} \quad \|x_i - x_j\| > \epsilon \text{ or } (x_i, x_j) \notin E \end{cases}$$

as usual for the transition probabilities

$$p(x_i, x_j) = \frac{w(x_i, x_j)}{\sum_k w(x_i, x_k)}$$

Often used with diffusion maps!

Definition of Network Laplacian based on edge weights $w(x, y)$:

$$L(x, y) = \begin{cases} w(x, y) & \text{if} \quad x \neq y \\ -\sum_z w(x, z) & \text{if} \quad x = y \end{cases}$$

and

$$D(x, y) = \begin{cases} 0 & \text{if} \quad x \neq y \\ \sum_z w(x, z) & \text{if} \quad x = y \end{cases}$$

Laplacian is symmetric!

Eigenvalues and -vectors $(0 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n)$

$$Lu_k = \lambda_k Du_k$$

Find map $\xi : V \to \mathbf{R}^m$ such that nodes with high weights stay close together!
Minimize

$$\frac{1}{2}\sum_{x,y}(\xi(x)-\xi(y))^2 w(x,y) = \xi^T L \xi, \quad \xi = (\xi(1),\dots,\xi(n))^T$$

with constraints

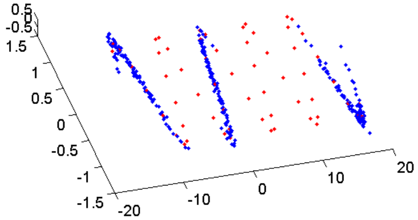
1. Avoid trivial solution $\langle \mathbf{1}, D\xi\rangle = 0$
2. Removing arbitrary scaling $\langle \xi, D\xi\rangle = 1$

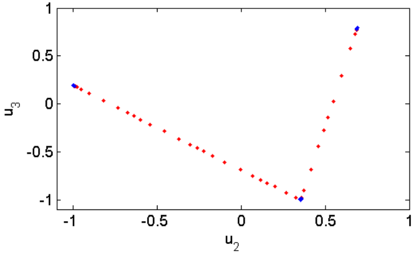
Explicit solution given by eigenvectors

$$\xi(x) = (u_2(x),\dots,u_{m+1}(x))$$

Example: Laplacian Eigenmaps

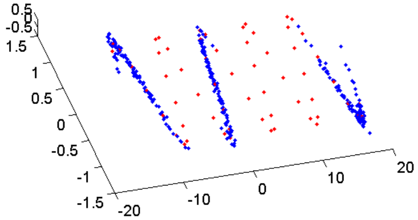
330 data points in dim = 300, almost lying on a sine-like 1d-manifold
Eigenmaps with $\epsilon = 1$

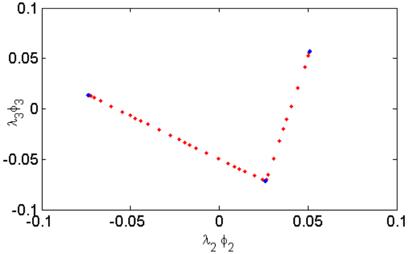




λ_1	λ_2	λ_3	λ_4	...
0.0000	0.0001	0.0003	0.0389	...

330 data points in dim $= 300$, almost lying on a sine-like 1d-manifold
Diffusion maps with $\epsilon = 1$





λ_1	λ_2	λ_3	λ_4	...
1.0000	0.9999	0.9997	0.9701	...

Question

What does the Laplacian approximate in the limit for $n \to \infty$?

Problem

Find map $f : \mathcal{M} \to \mathbb{R}^m$ such that points close together on the manifold are mapped closely together.

Solution

$$f(x) = (\eta_1(x), \ldots, \eta_m(x))^T$$

where $\eta_1, \ldots, \eta_m$ are the eigenfunctions of the first m non-trivial eigenvalues $0 < \lambda_1 \leq \ldots \leq \lambda_m$ of the eigenvalue problem,

$$\mathcal{L}\eta = \lambda\eta,$$

where $\mathcal{L}$ is the Laplace-Beltrami operator:

$$\mathcal{L}f \stackrel{\text{def}}{=} \Delta_{\mathcal{M}} f, \quad f \in C^2(\mathcal{M})$$

Method	*Partitioning*	*Low-dim. embedding*
Optimal aggregation	YES, deterministic $\min \|p - \text{lift}(\hat{p})\|_\mu^2$	NO
Optimal Fuzzy Aggregation	YES, fuzzy $\min \|p - \text{lift}_W(\hat{p})\|_\mu^2$	NO
Diffusion maps	YES, deterministic $\max \sum_i \sum_{x \in S_i, y \in S_i} \mu_i(x) p(x, y)$	YES, if spectr. gap $(\lambda_2\phi_2, \ldots, \lambda_m\phi_m)$
Laplacian eigenmaps	NO	YES, if spectr. gap $(u_2, \ldots, u_m)$