Entropic Optimal Transport and Wasserstein Barycenters in Random Graphs

Nicolas Keriven

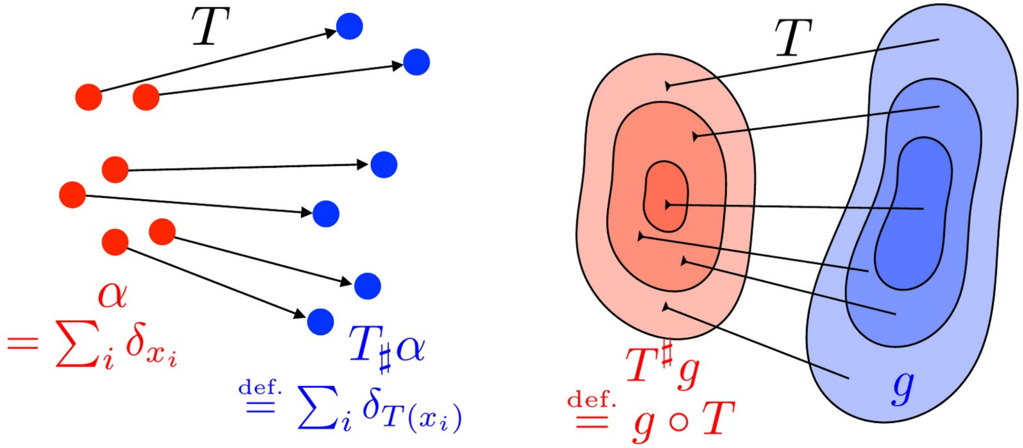
CNRS, Gipsa-lab, IRISA

Joint work with Marc Theveneau

(Ecole polytechnique)

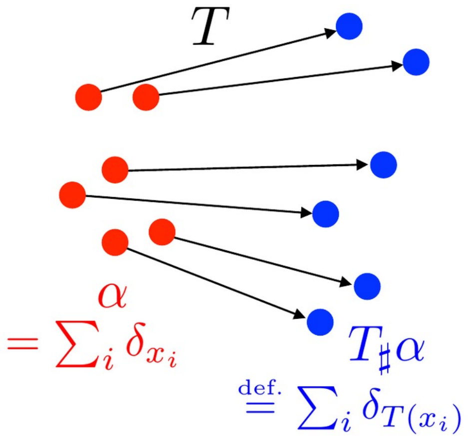
(Optimal) Transport in Graphs

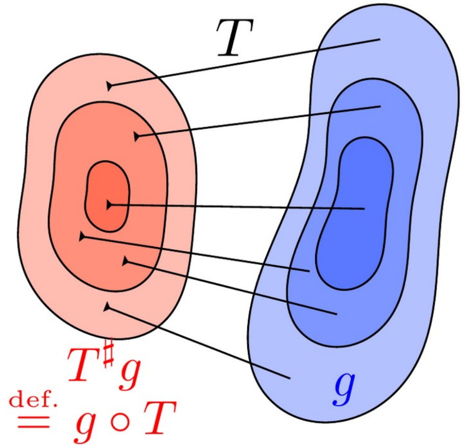
Optimal Transport (OT): "optimal" way to transport "mass" between several locations. Defines a (family of) metric(s) between probability distributions.



(Optimal) Transport in Graphs

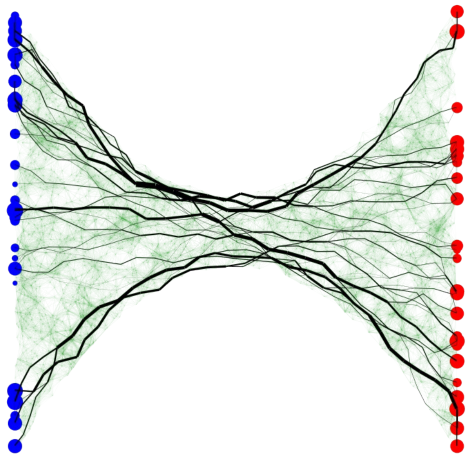
Optimal Transport (OT): "optimal" way to transport "mass" between several locations. Defines a (family of) metric(s) between probability distributions.



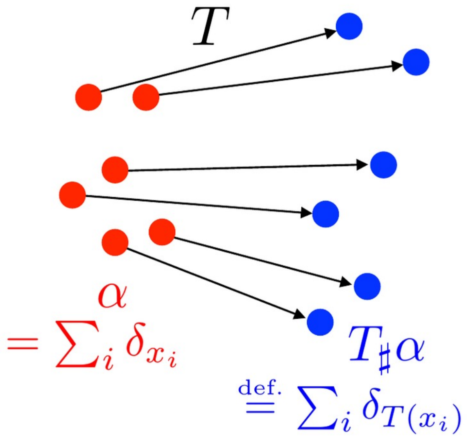


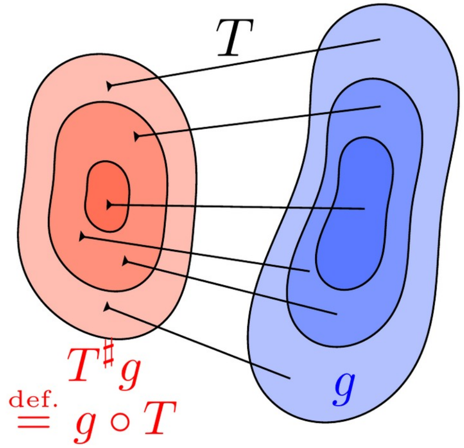
On "graphs"?

- Usually transporting mass along the edges



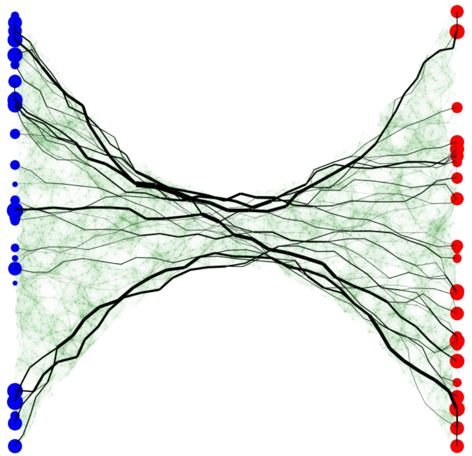
Optimal Transport (OT): "optimal" way to transport "mass" between several locations. Defines a (family of) metric(s) between probability distributions.





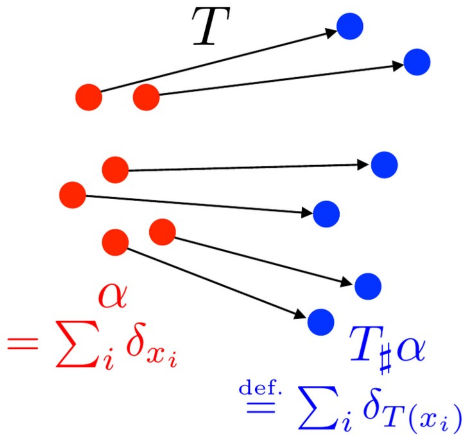
On "graphs"?

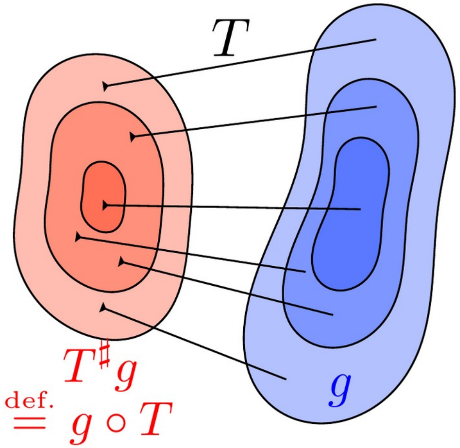
- Usually transporting mass along the edges
- Interpretable metrics between groups of (weighted) nodes are also interesting
  - Non-existing edges can be inferred (ie, nodes are "close" in some sense)



(Optimal) Transport in Graphs

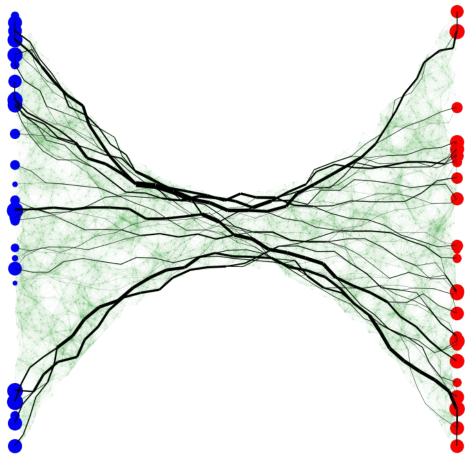
Optimal Transport (OT): "optimal" way to transport "mass" between several locations. Defines a (family of) metric(s) between probability distributions.





On "graphs"?

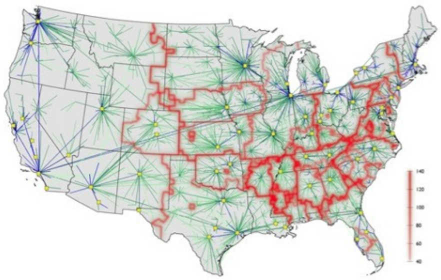
- Usually transporting mass along the edges
- Interpretable metrics between groups of (weighted) nodes are also interesting
  - Non-existing edges can be inferred (ie, nodes are "close" in some sense)
- Here, target nodes are **given** (user- or algorithm-chosen)



(Optimal) Transport in Graphs

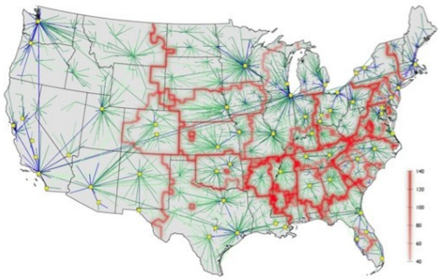
(Optimal) Transport in Graphs

- How to transport stuff on a road/computer/etc network...?



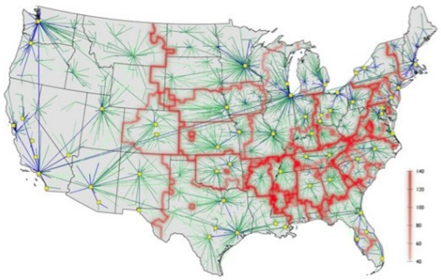
(Optimal) Transport in Graphs

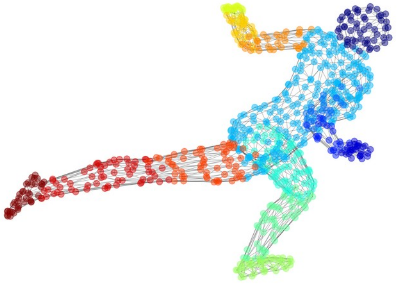
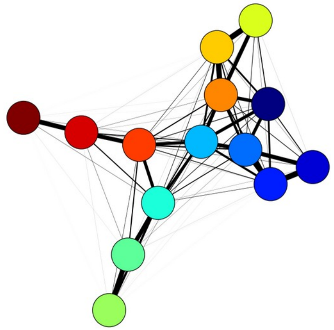
- How to transport stuff on a road/computer/etc network...?
- How different two groups of people are in a social network? (w.r.t. unobserved preferences)



(Optimal) Transport in Graphs

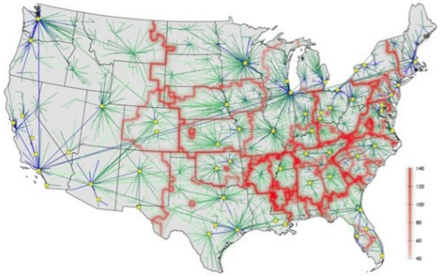
- How to transport stuff on a road/computer/etc network...?
- How different two groups of people are in a social network? (w.r.t. unobserved preferences)
- How "far" apart are different regions of a manifold? (w.r.t. geodesic distance)

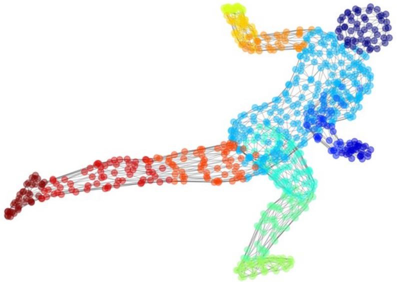
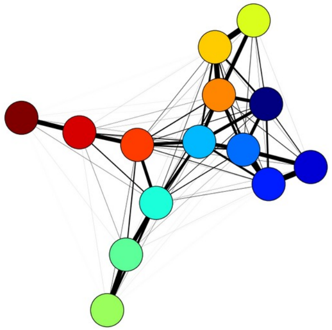
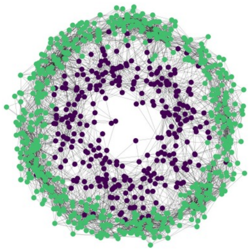
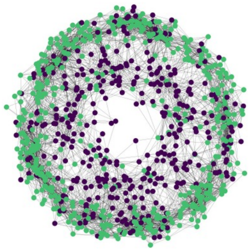


(Optimal) Transport in Graphs

- How to transport stuff on a road/computer/etc network...?
- How different two groups of people are in a social network? (w.r.t. unobserved preferences)
- How "far" apart are different regions of a manifold? (w.r.t. geodesic distance)
- What is a good criterion to evaluate the "quality" of clustering algorithms?



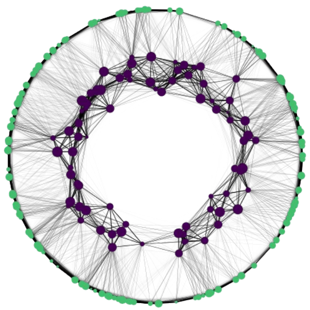
Entropic OT... in random graphs

Distributions	**Cost Matrix**
$\alpha \in \Delta_n^+$ $\beta \in \Delta_m^+$	$C \in \mathbb{R}_+^{n \times m}$

Entropic OT... in random graphs

Distributions	**Cost Matrix**
$\alpha \in \Delta_n^+$ $\beta \in \Delta_m^+$	$C \in \mathbb{R}_+^{n \times m}$

$$\{x_1, \ldots, x_n\}$$
$$\{x_{n+1}, \ldots, x_{n+m}\}$$
$$C = [c(x_i, x_{n+j})]_{ij}$$



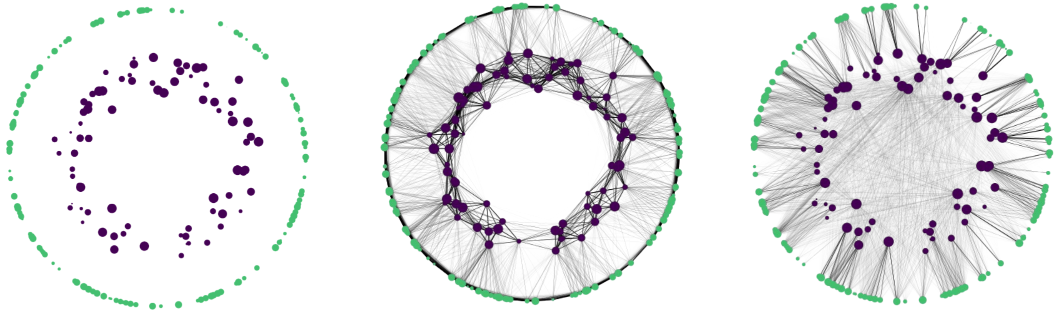
Entropic OT... in random graphs

Distributions	**Cost Matrix**
$\alpha \in \Delta_n^+$ $\beta \in \Delta_m^+$	$C \in \mathbb{R}_+^{n \times m}$

$\{x_1, \dots, x_n\}$

$\{x_{n+1}, \dots, x_{n+m}\}$

$C = [c(x_i, x_{n+j})]_{ij}$



Entropic-regularized OT: [Cuturi 2013]

$$\mathcal{W}_\epsilon^C(\alpha, \beta) = \min_{P \in \Pi(\alpha,\beta)} \langle C, P \rangle + \epsilon KL(P|\alpha \otimes \beta)$$

*NB: **Sinkhorn's algorithm** only uses* $K = e^{-C/\epsilon}$

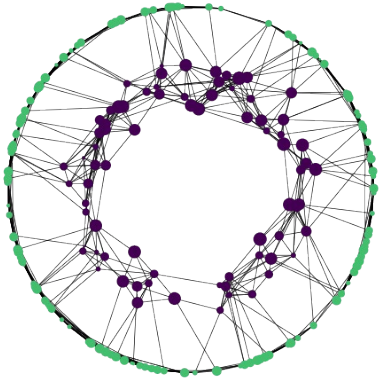
Entropic OT... in random graphs

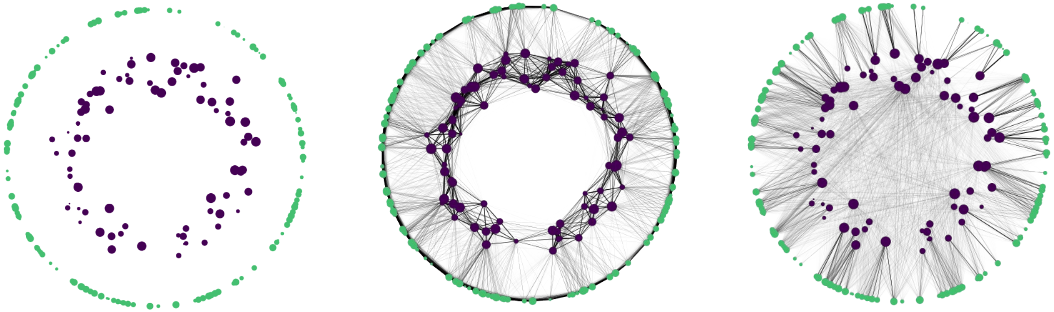
Distributions	**Cost Matrix**
$\alpha \in \Delta_n^+$ $\beta \in \Delta_m^+$	$C \in \mathbb{R}_+^{n\times m}$

$\{x_1, \ldots, x_n\}$

$\{x_{n+1}, \ldots, x_{n+m}\}$

$C = [c(x_i, x_{n+j})]_{ij}$





Random graph/edges

$$a_{ij} \sim \mathrm{Ber}(w_n(x_i, x_j))$$

Entropic-regularized OT: [Cuturi 2013]

$$\mathcal{W}_\epsilon^C(\alpha,\beta) = \min_{P\in\Pi(\alpha,\beta)} \langle C, P\rangle + \epsilon KL(P|\alpha\otimes\beta)$$

*NB: **Sinkhorn's algorithm** only uses* $K = e^{-C/\epsilon}$

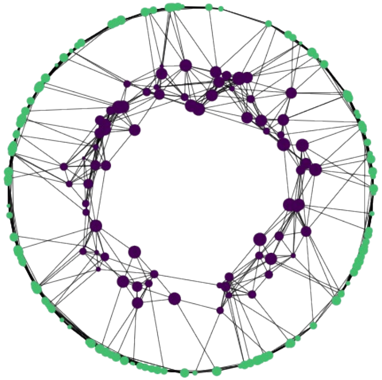
Entropic OT… in random graphs

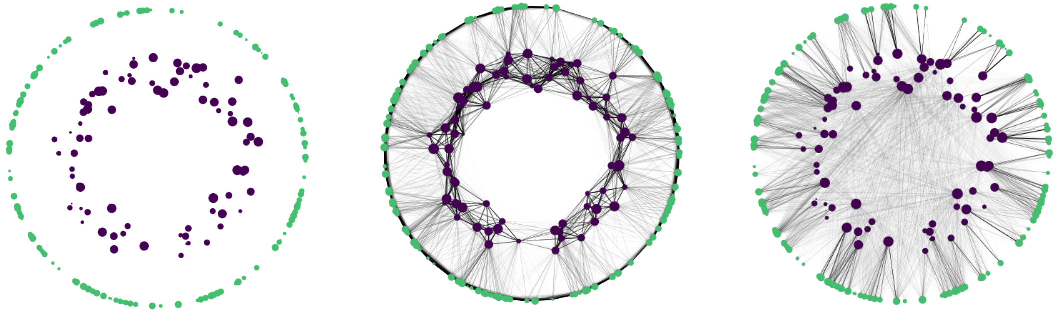
Distributions	Cost Matrix
$\alpha \in \Delta_n^+$ $\beta \in \Delta_m^+$	$C \in \mathbb{R}_+^{n\times m}$

$\{x_1, \dots, x_n\}$

$\{x_{n+1}, \dots, x_{n+m}\}$

$C = [c(x_i, x_{n+j})]_{ij}$





Entropic-regularized OT: [Cuturi 2013]

$$\mathcal{W}_\epsilon^C(\alpha,\beta) = \min_{P\in\Pi(\alpha,\beta)} \langle C, P\rangle + \epsilon KL(P|\alpha\otimes\beta)$$

*NB: **Sinkhorn's algorithm** only uses* $K = e^{-C/\epsilon}$

Random graph/edges

$$a_{ij} \sim \mathrm{Ber}(w_n(x_i, x_j))$$

- Estimate $\hat{C}$

Entropic OT... in random graphs

Distributions

$\alpha \in \Delta_n^+$

$\beta \in \Delta_m^+$

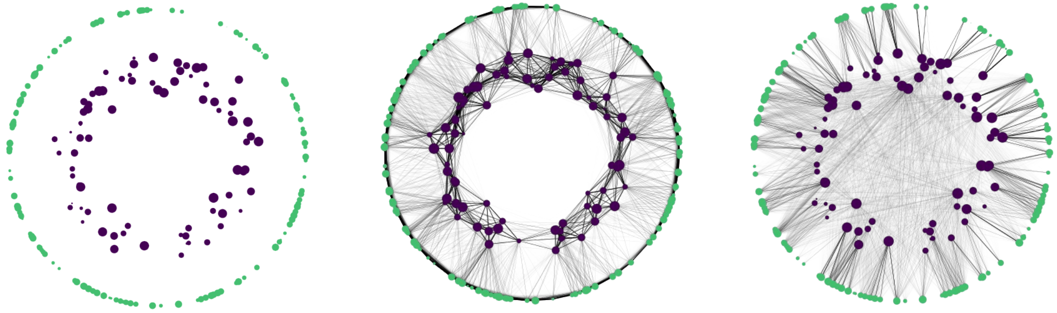
Cost Matrix

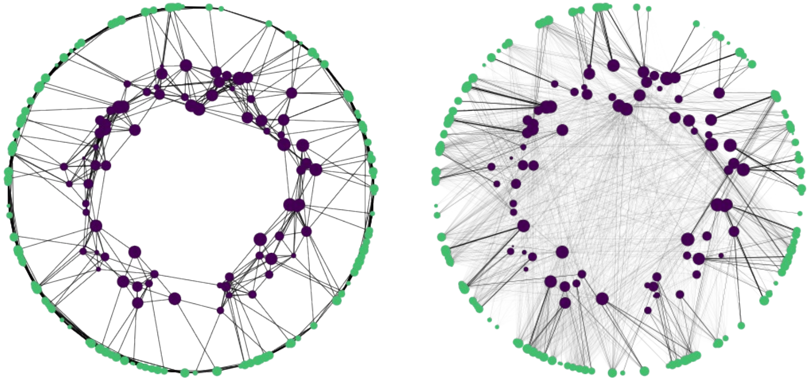
$C \in \mathbb{R}_+^{n \times m}$

$\{x_1, \ldots, x_n\}$

$\{x_{n+1}, \ldots, x_{n+m}\}$

$C = [c(x_i, x_{n+j})]_{ij}$





Random graph/edges

$$a_{ij} \sim \mathrm{Ber}(w_n(x_i, x_j))$$

- Estimate $\hat{C}$
- **How close is** $\mathcal{W}_\epsilon^{\hat{C}}(\alpha, \beta)$ **?**

Entropic-regularized OT: [Cuturi 2013]

$$\mathcal{W}_\epsilon^C(\alpha, \beta) = \min_{P \in \Pi(\alpha, \beta)} \langle C, P \rangle + \epsilon KL(P | \alpha \otimes \beta)$$

*NB: **Sinkhorn's algorithm** only uses* $K = e^{-C/\epsilon}$

Outline

1. Stability of OT to inexact cost
2. Application to RG with “local” kernels
3. Application to RG with “non-local” kernel
4. Wasserstein Barycenters *(w/ Marc Theveneau)*

Stability to inexact cost

Stability to inexact cost matrix?

Immediate: $\forall \epsilon \geq 0 \quad |\mathcal{W}_\epsilon^C(\alpha, \beta) - \mathcal{W}_\epsilon^{\hat{C}}(\alpha, \beta)| \leq \sup_P |\langle P, C - \hat{C} \rangle| \leq \|C - \hat{C}\|_\infty$

Stability to inexact cost

Stability to inexact cost matrix?

Immediate: $\forall \epsilon \geq 0 \quad |\mathcal{W}_\epsilon^C(\alpha, \beta) - \mathcal{W}_\epsilon^{\hat{C}}(\alpha, \beta)| \leq \sup_P |\langle P, C - \hat{C} \rangle| \leq \|C - \hat{C}\|_\infty$

May not be sufficient! E.g., obviously $\|A - \mathbb{E}A\|_\infty$ does not converge...

Stability to inexact cost

Stability to inexact cost matrix?

Immediate: $\forall \epsilon \geq 0 \quad |\mathcal{W}_\epsilon^C(\alpha,\beta) - \mathcal{W}_\epsilon^{\hat{C}}(\alpha,\beta)| \leq \sup_P |\langle P, C - \hat{C}\rangle| \leq \|C - \hat{C}\|_\infty$

May not be sufficient! E.g., obviously $\|A - \mathbb{E}A\|_\infty$ does not converge...

Theorem (K.): If $\ell \leq C_{ij}, \hat{C}_{ij} \leq L$ and $\alpha_i \lesssim \frac{1}{n}, \beta_j \lesssim \frac{1}{m}$

$\forall \epsilon > 0$

$$|\mathcal{W}_\epsilon^C(\alpha,\beta) - \mathcal{W}_\epsilon^{\hat{C}}(\alpha,\beta)| \lesssim \epsilon e^{(2L-\ell)/\epsilon} \frac{\|e^{-C/\epsilon} - e^{-\hat{C}/\epsilon}\|}{\sqrt{nm}}$$

$$\lesssim e^{2(L-\ell)/\epsilon} \frac{\|C - \hat{C}\|_F}{\sqrt{nm}}$$

Stability to inexact cost

Stability to inexact cost matrix?

Immediate: $\forall \epsilon \geq 0 \quad |\mathcal{W}_\epsilon^C(\alpha,\beta) - \mathcal{W}_\epsilon^{\hat{C}}(\alpha,\beta)| \leq \sup_P |\langle P, C-\hat{C}\rangle| \leq \|C-\hat{C}\|_\infty$

May not be sufficient! E.g., obviously $\|A - \mathbb{E}A\|_\infty$ does not converge...

Theorem (K.): If $\ell \leq C_{ij}, \hat{C}_{ij} \leq L$ and $\alpha_i \lesssim \frac{1}{n}, \beta_j \lesssim \frac{1}{m}$

$\forall \epsilon > 0$

$$|\mathcal{W}_\epsilon^C(\alpha,\beta) - \mathcal{W}_\epsilon^{\hat{C}}(\alpha,\beta)| \lesssim \epsilon e^{(2L-\ell)/\epsilon} \frac{\|e^{-C/\epsilon} - e^{-\hat{C}/\epsilon}\|}{\sqrt{nm}}$$

$$\lesssim e^{2(L-\ell)/\epsilon} \frac{\|C-\hat{C}\|_F}{\sqrt{nm}}$$

- Invariant to translating $C, \hat{C}$
- Exponential in ϵ
- First bound stronger, second bound more "usable"
- Proof: classical, bound the dual potentials

Stability of OT plan

Using strong convexity, we can obtain stability of the OT **plan:**

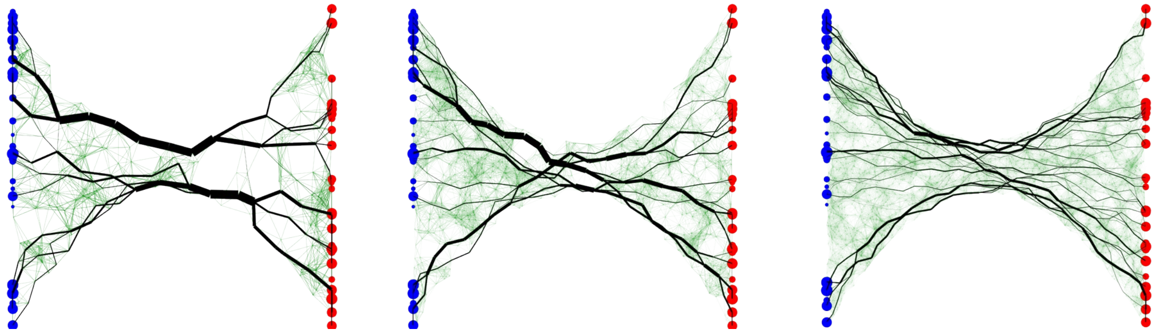
Stability of OT plan

Using strong convexity, we can obtain stability of the OT **plan:**

Theorem (K.): If $\ell \leq C_{ij}, \hat{C}_{ij} \leq L$ and $\alpha_i \lesssim \frac{1}{n}, \beta_j \lesssim \frac{1}{m}$

$\forall \epsilon > 0$

$$\mathrm{KL}(P^C | P^{\hat{C}}) \lesssim \epsilon^{-1} e^{2(L-\ell)/\epsilon} \frac{\|C - \hat{C}\|_F}{\sqrt{nm}} + e^{(4L-7\ell/2)/\epsilon} \sqrt{\frac{\|e^{-C/\epsilon} - e^{-\hat{C}/\epsilon}\|}{\sqrt{nm}}}$$



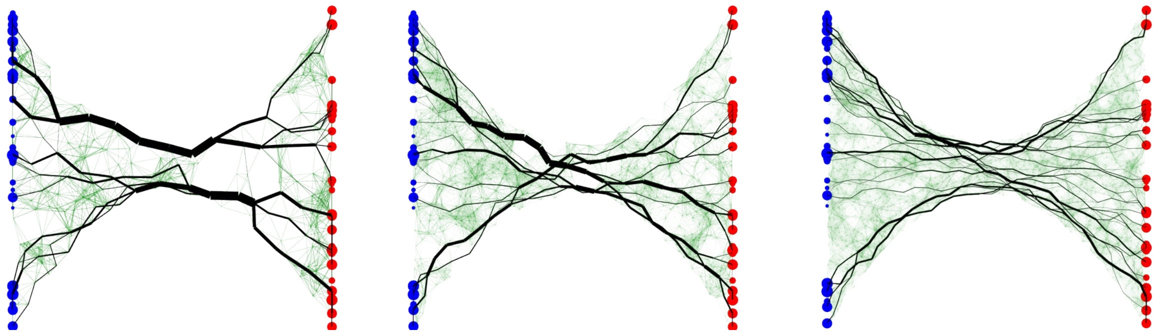
Stability of OT plan

Using strong convexity, we can obtain stability of the OT **plan:**

Theorem (K.): If $\ell \leq C_{ij}, \hat{C}_{ij} \leq L$ and $\alpha_i \lesssim \frac{1}{n}, \beta_j \lesssim \frac{1}{m}$

$\forall \epsilon > 0$

$$\mathrm{KL}(P^C | P^{\hat{C}}) \lesssim \epsilon^{-1} e^{2(L-\ell)/\epsilon} \frac{\|C - \hat{C}\|_F}{\sqrt{nm}} + e^{(4L - 7\ell/2)/\epsilon} \sqrt{\frac{\|e^{-C/\epsilon} - e^{-\hat{C}/\epsilon}\|}{\sqrt{nm}}}$$



- Still invariant by cost shift
- Includes both norms
- Slower rate than convergence of the metric itself

Outline

1. Stability of OT to inexact cost
2. Application to RG with "local" kernels
3. Application to RG with "non-local" kernel
4. Wasserstein Barycenters *(w/ Marc Theveneau)*

Geodesics on manifolds

RGs with "**local kernels**": close nodes are connected, radius decreases when #nodes increases

*Known: **weighted shortest paths** converge to geodesic distance*

[Bernstein et al. 2000]

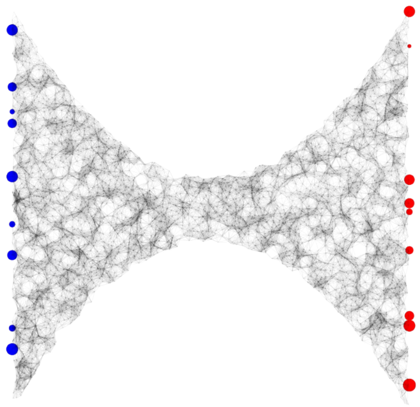
Geodesics on manifolds

RGs with "**local kernels**": close nodes are connected, radius decreases when #nodes increases

- k-manifold $\mathcal{M}_k \subset \mathbb{R}^d$ with geo. dist. $d_\mathcal{M}(x,y)$
- Fixed $\{x_1, \ldots, x_{n+m}\} \subset \mathcal{M}_k$
- Nodes $\{x_{n+m+1}, \ldots, x_N\} \overset{iid}{\sim} \nu$ with $N \to \infty$
- Kernel $w_N(x,y) = 1_{\|x-y\| \leq h_N}$ with $\frac{\log(1/h_N)}{N h_N^k} \to 0$

*Known: **weighted shortest paths** converge to geodesic distance*

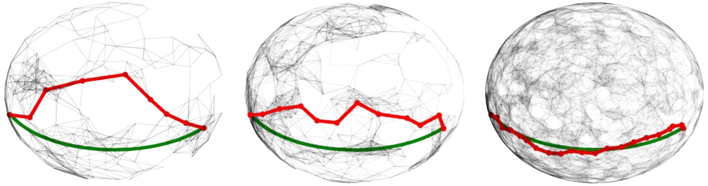
[Bernstein et al. 2000]



Geodesics on manifolds

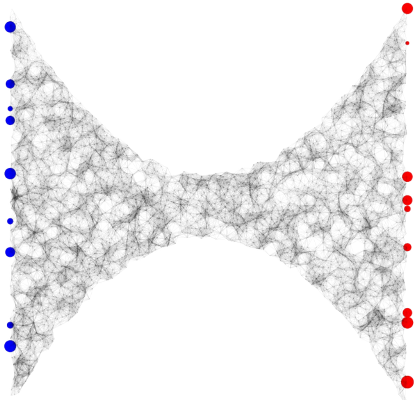
RGs with "**local kernels**": close nodes are connected, radius decreases when #nodes increases

- k-manifold $\mathcal{M}_k \subset \mathbb{R}^d$ with geo. dist. $d_\mathcal{M}(x,y)$
- Fixed $\{x_1, \ldots, x_{n+m}\} \subset \mathcal{M}_k$
- Nodes $\{x_{n+m+1}, \ldots, x_N\} \overset{iid}{\sim} \nu$ with $N \to \infty$
- Kernel $w_N(x,y) = 1_{\|x-y\| \leq h_N}$ with $\frac{\log(1/h_N)}{N h_N^k} \to 0$



Known: ***weighted shortest paths*** *converge to geodesic distance*

[Bernstein et al. 2000]



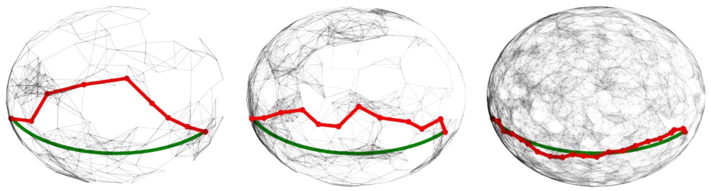
Theorem (K.): if ν has a lower-bounded density, whp

$$h_N \mathrm{SP}(v_i, v_{n+j}) = d_\mathcal{M}(x_i, x_{n+j}) + \mathcal{O}\left(\left(\frac{\log 1/h_N}{N h_N^k}\right)^{\frac{1}{k}}\right)$$

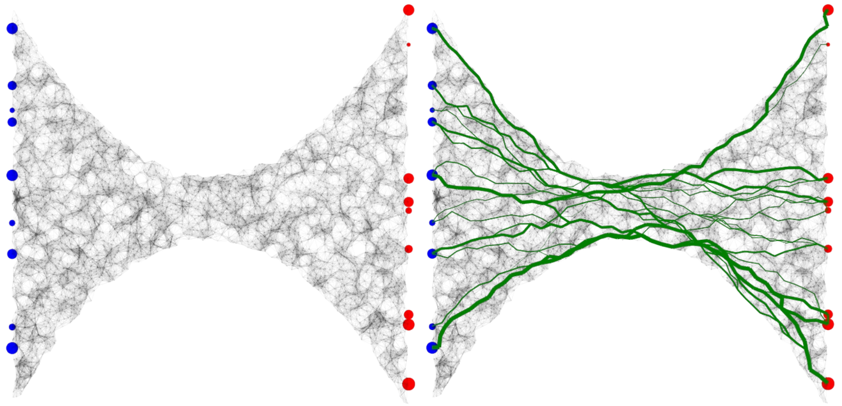
Geodesics on manifolds

RGs with "**local kernels**": close nodes are connected, radius decreases when #nodes increases

- k-manifold $\mathcal{M}_k \subset \mathbb{R}^d$ with geo. dist. $d_{\mathcal{M}}(x, y)$
- Fixed $\{x_1, \ldots, x_{n+m}\} \subset \mathcal{M}_k$
- Nodes $\{x_{n+m+1}, \ldots, x_N\} \overset{iid}{\sim} \nu$ with $N \to \infty$
- Kernel $w_N(x, y) = 1_{\|x-y\| \leq h_N}$ with $\frac{\log(1/h_N)}{N h_N^k} \to 0$



*Known: **weighted shortest paths** converge to geodesic distance*
[Bernstein et al. 2000]



Corollary

$C_{ij} = f(d_{\mathcal{M}}(x_i, x_{n+j}))$

leads to

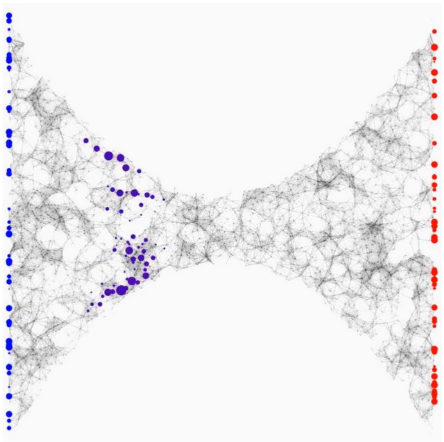
$\|\hat{C}_{\mathrm{SP}} - C\|_\infty \to 0$

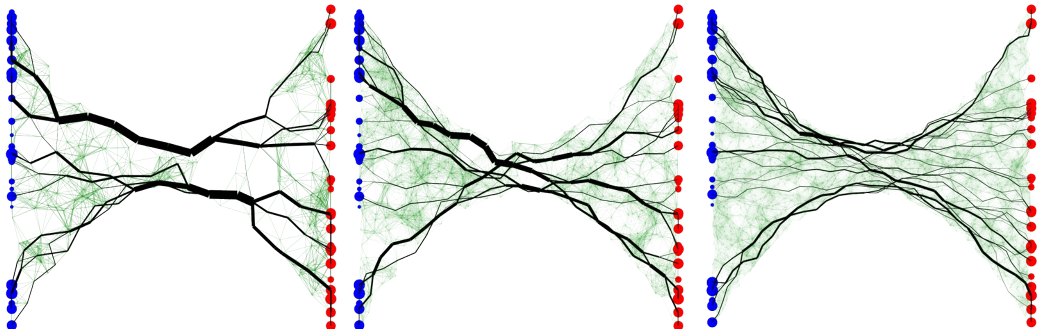
Theorem (K.): if ν has a lower-bounded density, whp

$$h_N \mathrm{SP}(v_i, v_{n+j}) = d_{\mathcal{M}}(x_i, x_{n+j}) + \mathcal{O}\left(\left(\frac{\log 1/h_N}{N h_N^k}\right)^{\frac{1}{k}}\right)$$

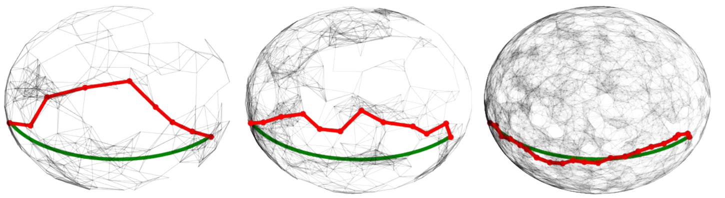
Illustration

Some numerical illustrations...

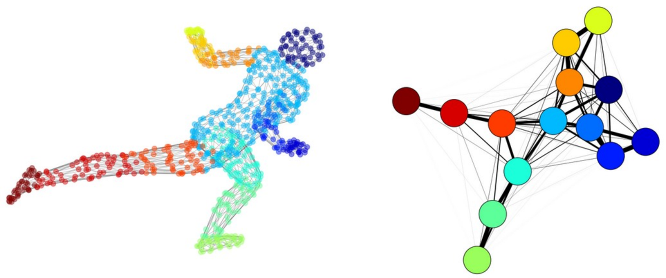




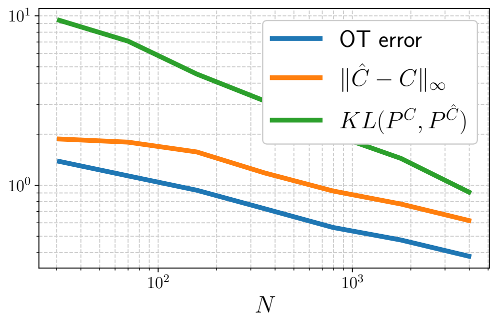
Convergence of OT plan



Convergence of shortest path



MDS embedding



Convergence rates

Outline

1. Stability of OT to inexact cost
2. Application to RG with “local” kernels
3. Application to RG with “non-local” kernel
4. Wasserstein Barycenters *(w/ Marc Theveneau)*

USVT estimator

RGs with "nonlocal kernels": fixed kernel, multiplying factor decreases when #nodes increases

- Nodes $\{x_1, \ldots, x_{n+m}\}$
  with $n \sim m \to \infty$
- Kernel $w_n(x,y) = \rho_n w(x,y)$
  with $\rho_n \gtrsim (\log n)/n$
  and psd kernel
- Cost $c(x,y) = f(w(x,y))$
  with Lipschitz f

USVT estimator

RGs with "nonlocal kernels": fixed kernel, multiplying factor decreases when #nodes increases

- Nodes $\{x_1, \dots, x_{n+m}\}$
  with $n \sim m \to \infty$
- Kernel $w_n(x,y) = \rho_n w(x,y)$
  with $\rho_n \gtrsim (\log n)/n$
  and psd kernel
- Cost $c(x,y) = f(w(x,y))$
  with Lipschitz f

[Lei&Rinaldo 2015]

Pbm: $\frac{1}{n}\|A/\rho_n - W\| \lesssim (n\rho_n)^{-\frac{1}{2}}$

but $\frac{1}{n}\|A/\rho_n - W\|_F \not\to 0$

USVT estimator

RGs with "nonlocal kernels": fixed kernel, multiplying factor decreases when #nodes increases

- Nodes $\{x_1, \dots, x_{n+m}\}$
  with $n \sim m \to \infty$
- Kernel $w_n(x,y) = \rho_n w(x,y)$
  with $\rho_n \gtrsim (\log n)/n$
  and psd kernel
- Cost $c(x,y) = f(w(x,y))$
  with Lipschitz f

[Lei&Rinaldo 2015]

Pbm: $\frac{1}{n}\|A/\rho_n - W\| \lesssim (n\rho_n)^{-\frac{1}{2}}$

but $\frac{1}{n}\|A/\rho_n - W\|_F \not\to 0$

Universal Singular Value Thresholding (USVT)

[Chatterjee 2015]

- Diagonalize $A = \sum_i \sigma_i a_i a_i^\top$

$$\hat{W}_\gamma = \mathrm{cut}_{[w_{\min}, w_{\max}]}(\rho_n^{-1} \textstyle\sum_{\sigma_i \geq \gamma\sqrt{\rho_n n}} \sigma_i a_i a_i^\top)$$

USVT estimator

RGs with "nonlocal kernels": fixed kernel, multiplying factor decreases when #nodes increases

- Nodes $\{x_1, \dots, x_{n+m}\}$
  with $n \sim m \to \infty$
- Kernel $w_n(x,y) = \rho_n w(x,y)$
  with $\rho_n \gtrsim (\log n)/n$
  and psd kernel
- Cost $c(x,y) = f(w(x,y))$
  with Lipschitz f

[Lei&Rinaldo 2015]

Pbm: $\frac{1}{n}\|A/\rho_n - W\| \lesssim (n\rho_n)^{-\frac{1}{2}}$

but $\frac{1}{n}\|A/\rho_n - W\|_F \not\to 0$

Universal Singular Value Thresholding (USVT)

[Chatterjee 2015]

- Diagonalize $A = \sum_i \sigma_i a_i a_i^\top$

$$\hat{W}_\gamma = \mathrm{cut}_{[w_{\min}, w_{\max}]}(\rho_n^{-1} \textstyle\sum_{\sigma_i \geq \gamma\sqrt{\rho_n n}} \sigma_i a_i a_i^\top)$$

Theorem (K.): for all $r > 0$, there is γ_r such that, with proba $1 - n^{-r}$, $\frac{1}{n}\|\hat{W}_{\gamma_r} - W\|_F \lesssim (n\rho_n)^{-1/4}$

USVT estimator

RGs with "nonlocal kernels": fixed kernel, multiplying factor decreases when #nodes increases

- Nodes $\{x_1, \ldots, x_{n+m}\}$ with $n \sim m \to \infty$
- Kernel $w_n(x,y) = \rho_n w(x,y)$ with $\rho_n \gtrsim (\log n)/n$ and psd kernel
- Cost $c(x,y) = f(w(x,y))$ with Lipschitz f

[Lei&Rinaldo 2015]

Pbm: $\frac{1}{n}\|A/\rho_n - W\| \lesssim (n\rho_n)^{-\frac{1}{2}}$

but $\frac{1}{n}\|A/\rho_n - W\|_F \not\to 0$

Universal Singular Value Thresholding (USVT)

[Chatterjee 2015]

- Diagonalize $A = \sum_i \sigma_i a_i a_i^\top$

$$\hat{W}_\gamma = \mathrm{cut}_{[w_{\min}, w_{\max}]}(\rho_n^{-1} \textstyle\sum_{\sigma_i \geq \gamma\sqrt{\rho_n n}} \sigma_i a_i a_i^\top)$$

Theorem (K.): for all $r > 0$, there is γ_r such that, with proba $1 - n^{-r}$, $\quad \frac{1}{n}\|\hat{W}_{\gamma_r} - W\|_F \lesssim (n\rho_n)^{-1/4}$

Corollary:

$$|\mathcal{W}_\epsilon^{\hat{C}_{\gamma_r}}(\alpha,\beta) - \mathcal{W}_\epsilon^{C}(\alpha,\beta)| \lesssim e^{2(L-\ell)/\epsilon}(\rho_n n)^{-1/4}$$

$$\mathrm{KL}(P^C | P^{\hat{C}}) \lesssim \epsilon^{-1/2} e^{4(L-\ell)/\epsilon}(\rho_n n)^{-1/8}$$

"Fast" rate

When $w(x,y) = e^{-\frac{\|x-y\|^p}{\sigma}}$, the matrix W is **directly** the "Sinkhorn" matrix $K = e^{-C/\sigma}$

when $\epsilon = \sigma$ **and** $c(x,y) = \|x-y\|^p$

"Fast" rate

When $w(x,y) = e^{-\frac{\|x-y\|^p}{\sigma}}$, the matrix W is **directly** the "Sinkhorn" matrix $K = e^{-C/\sigma}$

when $\epsilon = \sigma$ **and** $c(x,y) = \|x-y\|^p$

Theorem (K.):

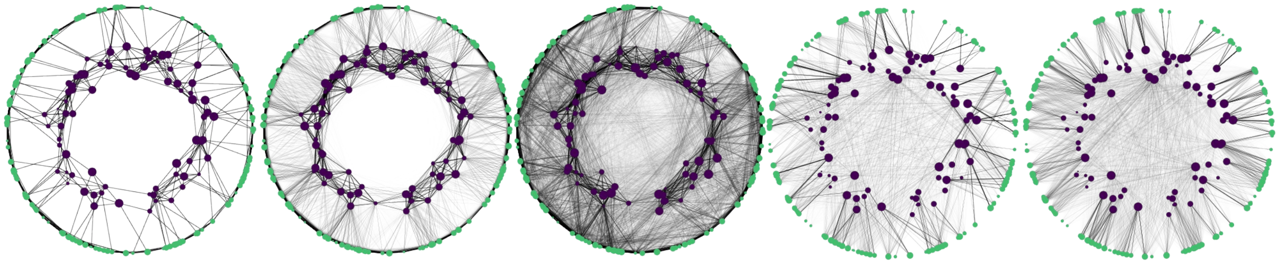
Defining $\mathcal{L}_\epsilon^K(\alpha,\beta) = \max_{f,g} f^\top\alpha + g^\top\beta - \epsilon(e^{\frac{f}{\epsilon}} \odot \alpha)^\top K(e^{\frac{g}{\epsilon}} \odot \beta) + \epsilon$

the dual OT cost with matrix K , whp

(plus some bounding conditions on the potentials)

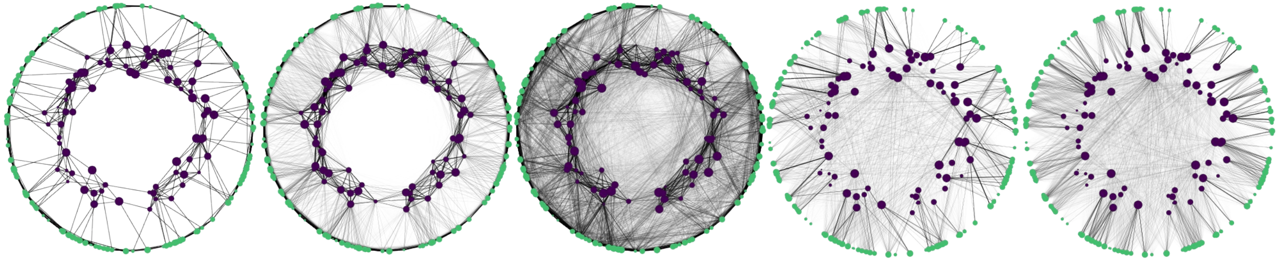
$$|\mathcal{L}_\sigma^{A/\rho_n}(\alpha,\beta) - \mathcal{W}_\sigma^C(\alpha,\beta)| \lesssim (n\rho_n)^{-1/2}$$

Illustration

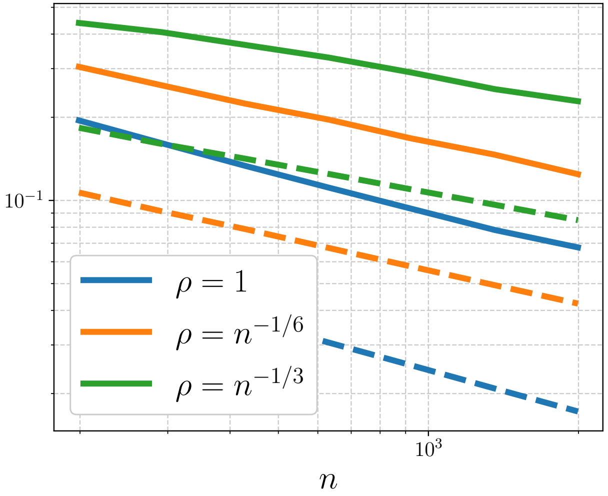


Observation, true kernel, USVT, true OT plan, estimated OT plan.

Illustration



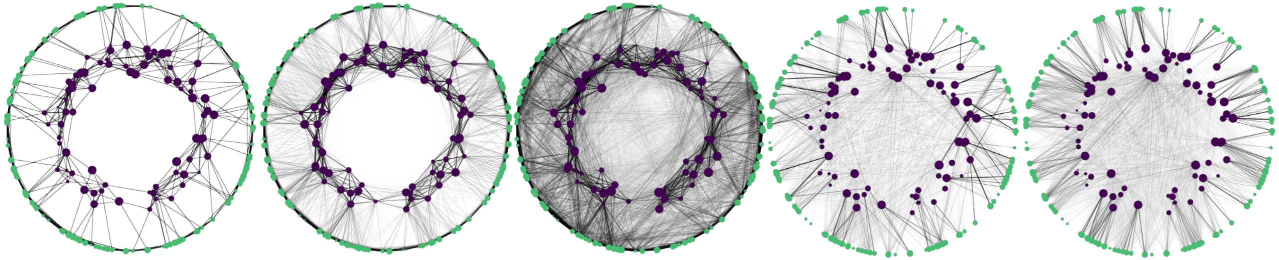
Observation, true kernel, USVT, true OT plan, estimated OT plan.



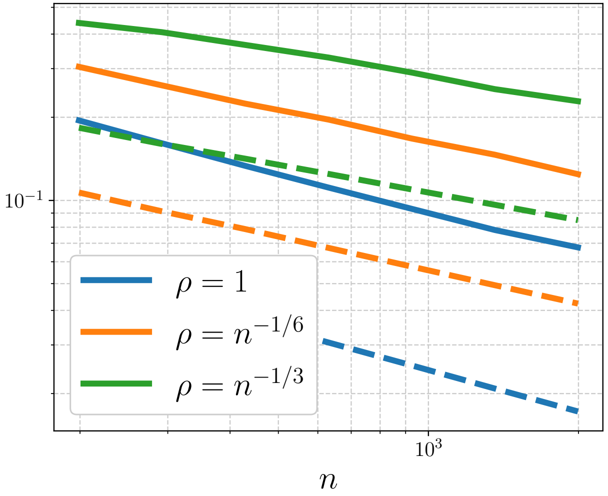
Convergence of OT distance.

Dotted: fast estimator

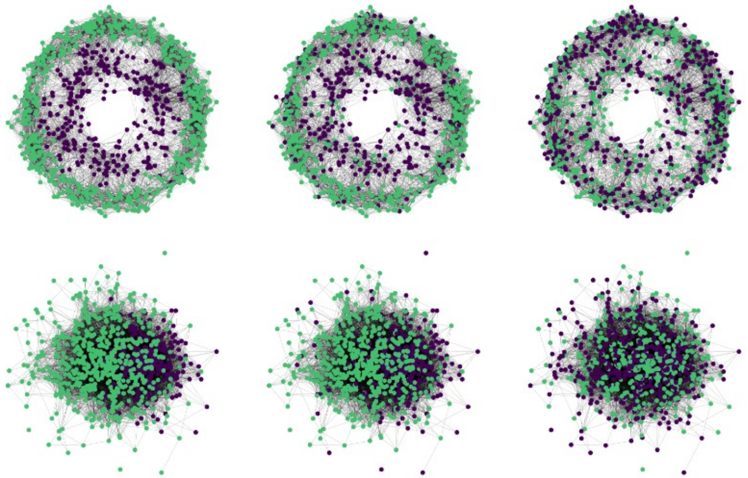
Illustration



Observation, true kernel, USVT, true OT plan, estimated OT plan.



Convergence of OT distance.
Dotted: fast estimator



	Cond.	Cut	Cov.	Perf.	Mod.	OT
Circ.	0.78	0.95	0.96	0.91	0.96	**0.97**
GMM	0.71	0.95	0.95	0.92	0.95	**0.97**

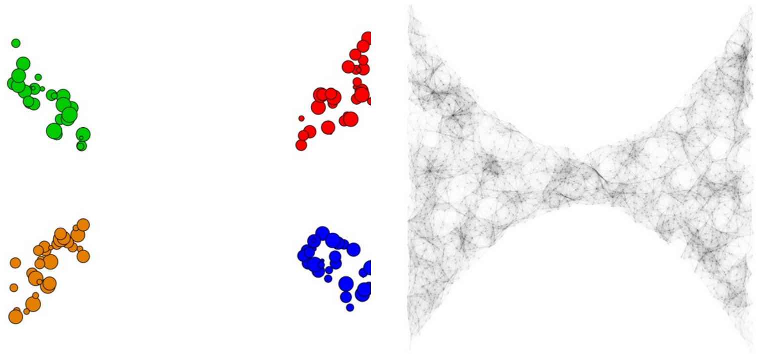
***Clustering quality**: correlation between quality metrics and increasingly noisy clustering*

Outline

1. Stability of OT to inexact cost
2. Application to RG with “local” kernels
3. Application to RG with “non-local” kernel
4. Wasserstein Barycenters *(w/ Marc Theveneau)*

Entropic Wasserstein Barycenters

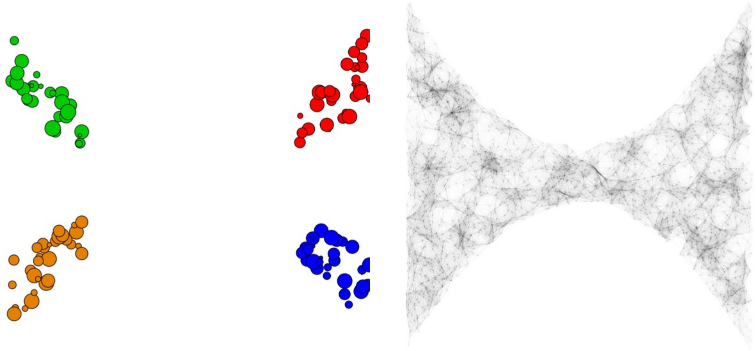
S distributions	S cost matrices **to a common space**
$\beta_s \in \Delta_{m_s}^+$	$C_s \in \mathbb{R}_+^{n \times m_s}$



Entropic Wasserstein Barycenters

S distributions $\beta_s \in \Delta^+_{m_s}$

S cost matrices **to a common space** $C_s \in \mathbb{R}_+^{n \times m_s}$

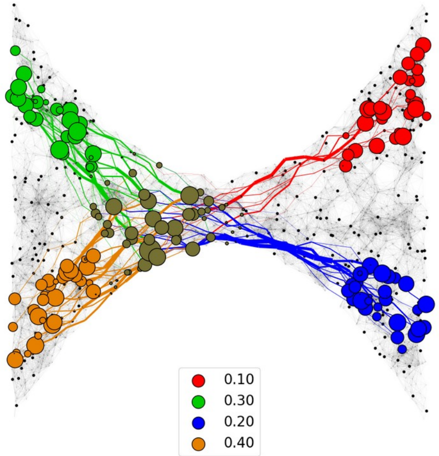


Wasserstein Barycenters [Agueh Carlier 2011]

Given nonnegative weights $\sum_s \lambda_s = 1$

$$\alpha^C = \arg \min_{\alpha \in \Delta_n^+} B^C(\alpha) = \sum_s \lambda_s \mathcal{W}_\epsilon^{C_s}(\alpha, \beta_s)$$

NB: a variant of ***Sinkhorn's algorithm*** *only uses* $K_s = e^{-C_s/\epsilon}$



WB stability to inexact cost

Immediate: $\forall \epsilon \geq 0 \quad |B_\epsilon^C(\alpha_\epsilon^C) - B_\epsilon^{\hat{C}}(\alpha_\epsilon^{\hat{C}})| \leq \sum_s \lambda_s \|C_s - \hat{C}_s\|_\infty$

WB stability to inexact cost

Immediate: $\forall \epsilon \geq 0 \quad |B_\epsilon^C(\alpha_\epsilon^C) - B_\epsilon^{\hat{C}}(\alpha_\epsilon^{\hat{C}})| \leq \sum_s \lambda_s \|C_s - \hat{C}_s\|_\infty$

We are more interested in the stability of the barycenters α^C !

WB stability to inexact cost

Immediate: $\forall \epsilon \geq 0 \quad |B_\epsilon^C(\alpha_\epsilon^C) - B_\epsilon^{\hat{C}}(\alpha_\epsilon^{\hat{C}})| \leq \sum_s \lambda_s \|C_s - \hat{C}_s\|_\infty$

We are more interested in the stability of the barycenters α^C !

Theorem (T,K): If $\ell \leq C_{sij}, \hat{C}_{sij} \leq L$

$\forall \epsilon > 0$

$$\|\alpha_\epsilon^C - \alpha_\epsilon^{\hat{C}}\|_2^2 \lesssim \epsilon e^{3(L-\ell)/\epsilon} \sum_s \lambda_s \|C_s - \hat{C}_s\|_\infty$$

WB stability to inexact cost

Immediate: $\forall \epsilon \geq 0 \quad |B_\epsilon^C(\alpha_\epsilon^C) - B_\epsilon^{\hat{C}}(\alpha_\epsilon^{\hat{C}})| \leq \sum_s \lambda_s \|C_s - \hat{C}_s\|_\infty$

We are more interested in the stability of the barycenters α^C !

Theorem (T,K): If $\ell \leq C_{sij}, \hat{C}_{sij} \leq L$

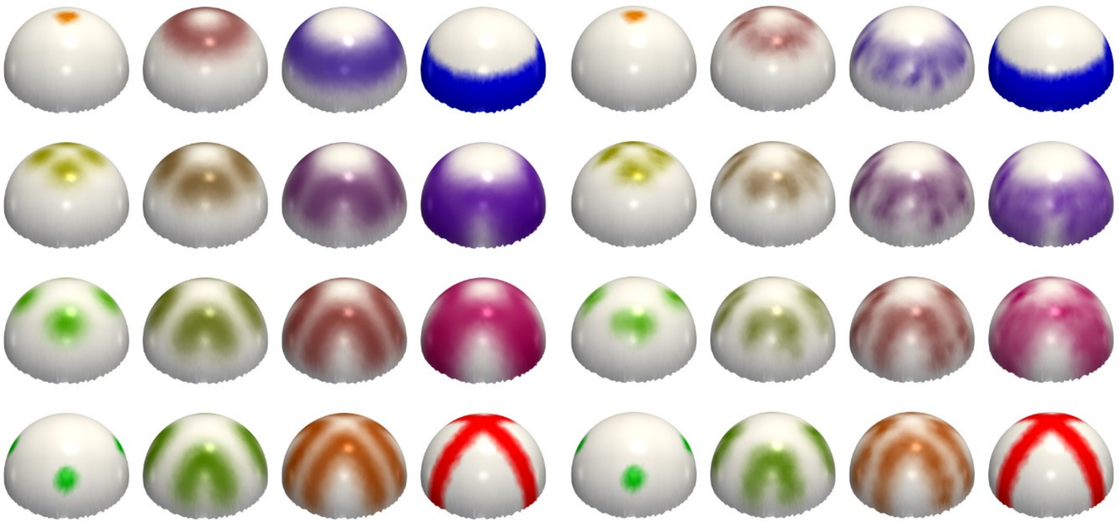
$\forall \epsilon > 0$

$$\|\alpha_\epsilon^C - \alpha_\epsilon^{\hat{C}}\|_2^2 \lesssim \epsilon e^{3(L-\ell)/\epsilon} \sum_s \lambda_s \|C_s - \hat{C}_s\|_\infty$$

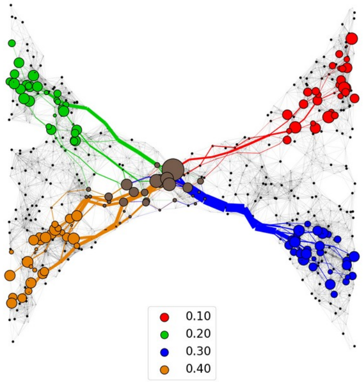
- Invariant to translating $C, \hat{C}$
- Exponential in ϵ
- Only supremum norm, Frobenius still open
- Proof: classical, bound the dual potentials
- More recent results with different approach: see Chizat 2023

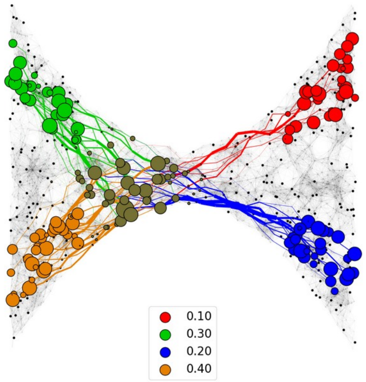
Illustration

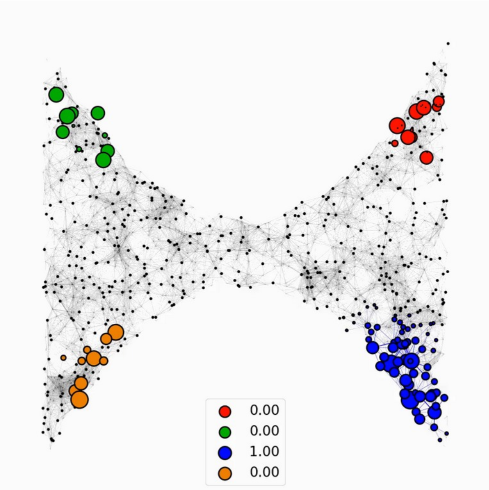
Immediately lead to convergence for **local kernels on manifolds** (non-local still open)

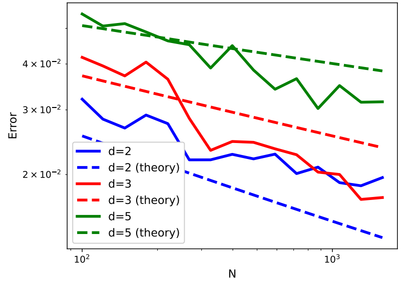


(a) Barycenters with the true geodesics (known for the sphere). (b) Barycenters with the shortest paths in a random graph.









Conclusion

- OT and WB can be done when the cost matrix is not known exactly
- *Maybe "reinventing the wheel" a bit, but* interesting results in the context of random graphs
- First steps, many **outlooks**:
  - More integrated, data-driven way of estimating the cost?
  - WB with non-local kernels
  - Other applications?

Keriven N. **Entropic Optimal Transport in Random Graphs.** a*rXiv:2201.03949*
Theveneau M., Keriven N. **Stability of Entropic Wasserstein Barycenters and application to random geometric graphs.** arXiv:2210.10535

nkeriven.github.io