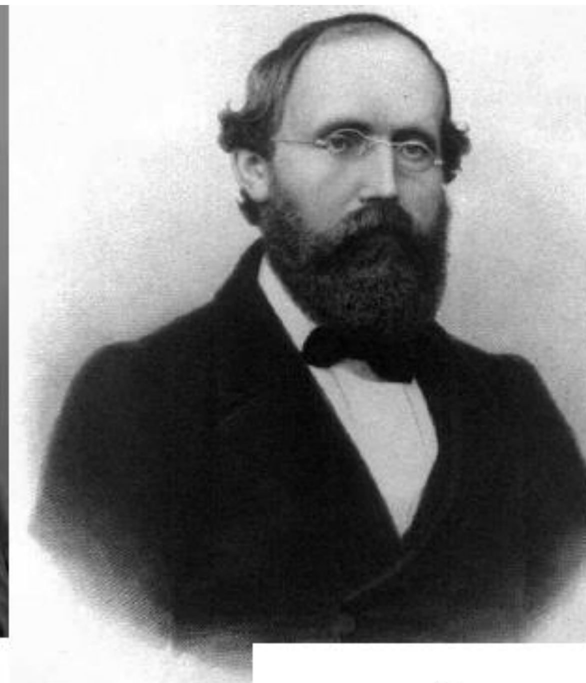
Optimal Mass Tranport and the Robustness of Complex Networks

Allen Tannenbaum
Distinguished Professor, CS/Applied Mathematics, SBU

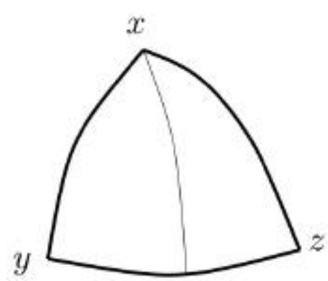
Collaborators/References

- Many: Romeil Sandhu, Yi Gao, Yongxin Chen, Tryphon Georgiou, Liangjia Zhu, Ivan Kolesov, Steve Haker, Eldad Haber, Sigurd Angenent, Wilfrid Gangbo, Ron Kikinis
- Matrix-valued: Similar approaches by Carlen-Maas, Mittnenzweig-Mielke.
- Curvature analysis for other networks: C.Wang, E. Jonckheere, R. Banirazi; R.P Sreejith, K. Mohanraj, J. Jost, E. Saucan, A. Samal; C.-C.Ni, Y.-Y. Lin, J. Gao, X. Gu, E. Saucan
- Papers available on arxiv.

Our Theme

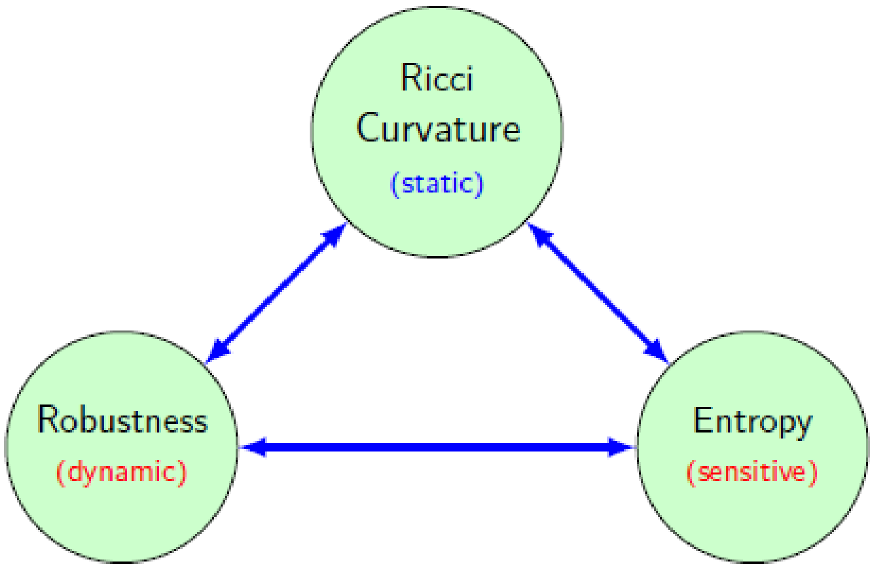


Min $\Sigma c(x, y)$



$S = -\int \varrho \log \varrho$

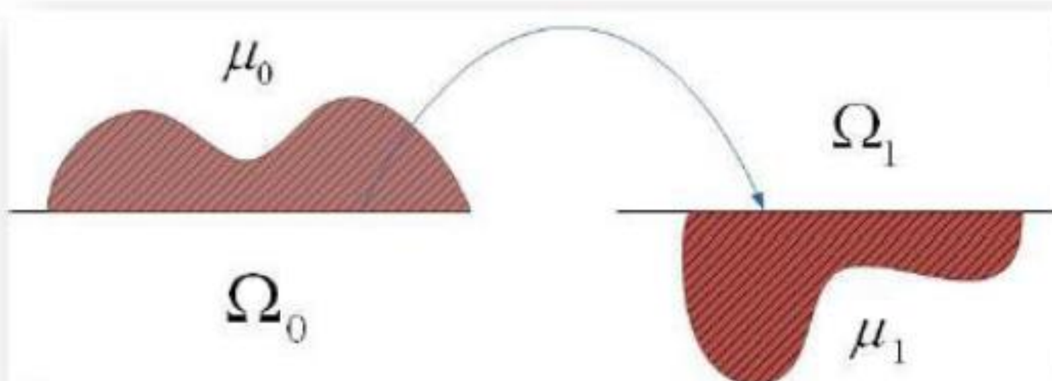
The Connection



All are connected via Optimal Mass Transport.

Optimal Mass Transport
Monge Transportation Cost (1781)

- Considers the engineer's problem of transporting a pile of soil or rubble to an excavation with the least amount of work.



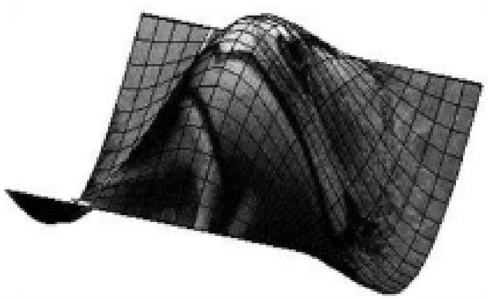
Optimal Mass Transport (MKW)

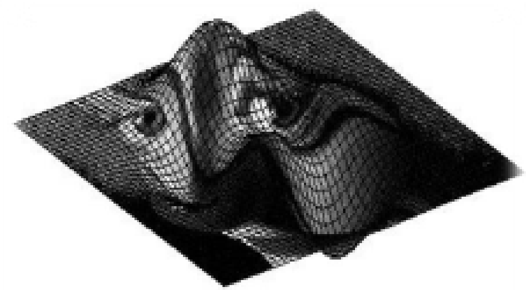
Given two oriented Riemannian manifolds

Ω_0 and Ω_1

with corresponding density functions

μ_0 and μ_1





and the same

amount of total mass:

$$\int_{\Omega_0} \mu_0(x)\,dx = \int_{\Omega_1} \mu_1(x)\,dx$$

Transportation Cost
Modern Formulation – Monge Kantorovich (MK)

Construct a smooth mapping:

$$u : (\Omega_0, \mu_0) \rightarrow (\Omega_1, \mu_1)$$

With *mass preserving (MP)* constraint:

$$\mu_0 = \det(\nabla u)\, \mu_1(u) \quad \textit{(Jacobian equation)}$$

so as to minimize the cost function:

$$M(u) = \int_{\Omega_0} \Phi(x, u(x))\, \mu_0(x)\, dx$$

$\Phi(x, u(x))$ is a positive twice differentiable convex function.

Kanterovich-Wasserstein Metric

Jacobian problem has many solutions. Want **optimal** one (**Lp-Kantorovich-Wasserstein metric**)

$$d_p(\mu_0, \mu_1)^p := \inf_u \int |u(x) - x|^p \mu_0(x) dx$$

Optimal map (when it exists) chooses a map with preferred geometry (like the Riemann Mapping Theorem) in the plane.

Solution of L2 M-K and Polar Factorization

Specializing to quadratic cost:

$$\Phi(x) = \frac{|x|^2}{2}$$

leads to the following "non-local" gradient descent equation:

$$\tilde{u}_t = -1/\mu_0 \nabla \tilde{u}(\tilde{u} - \nabla \Delta^{-1} \operatorname{div}(\tilde{u}))$$

Motivation for the approach:

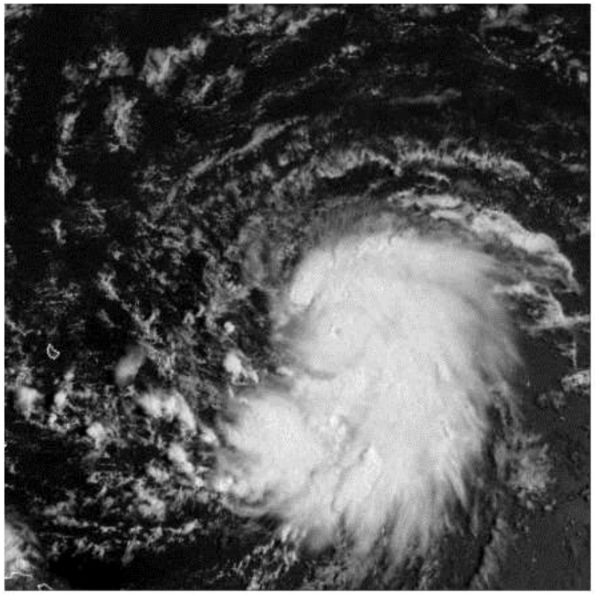
$$\tilde{u} = u \circ s^{-1} = \nabla w + \chi, \quad \operatorname{div}(\chi) = 0 \quad \textbf{Helmholtz decomp.}$$

The key idea is to push the fixed initial map u (thought of as a vector field) using the one-parameter family of MP maps in order to remove the divergence-free part!

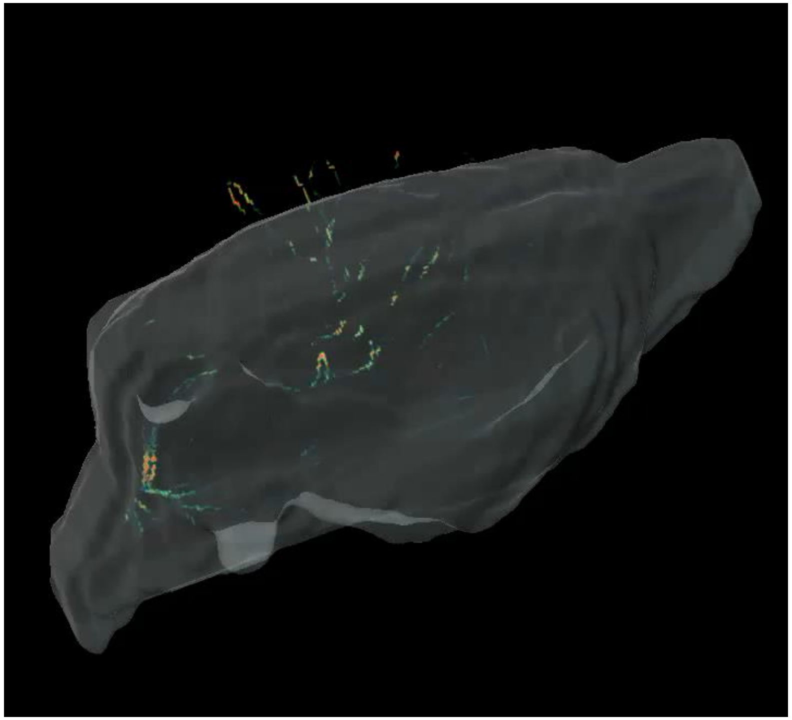
$$u = \nabla w \circ s \quad \text{Polar factorization}$$

Optimal Mass Transport Applications

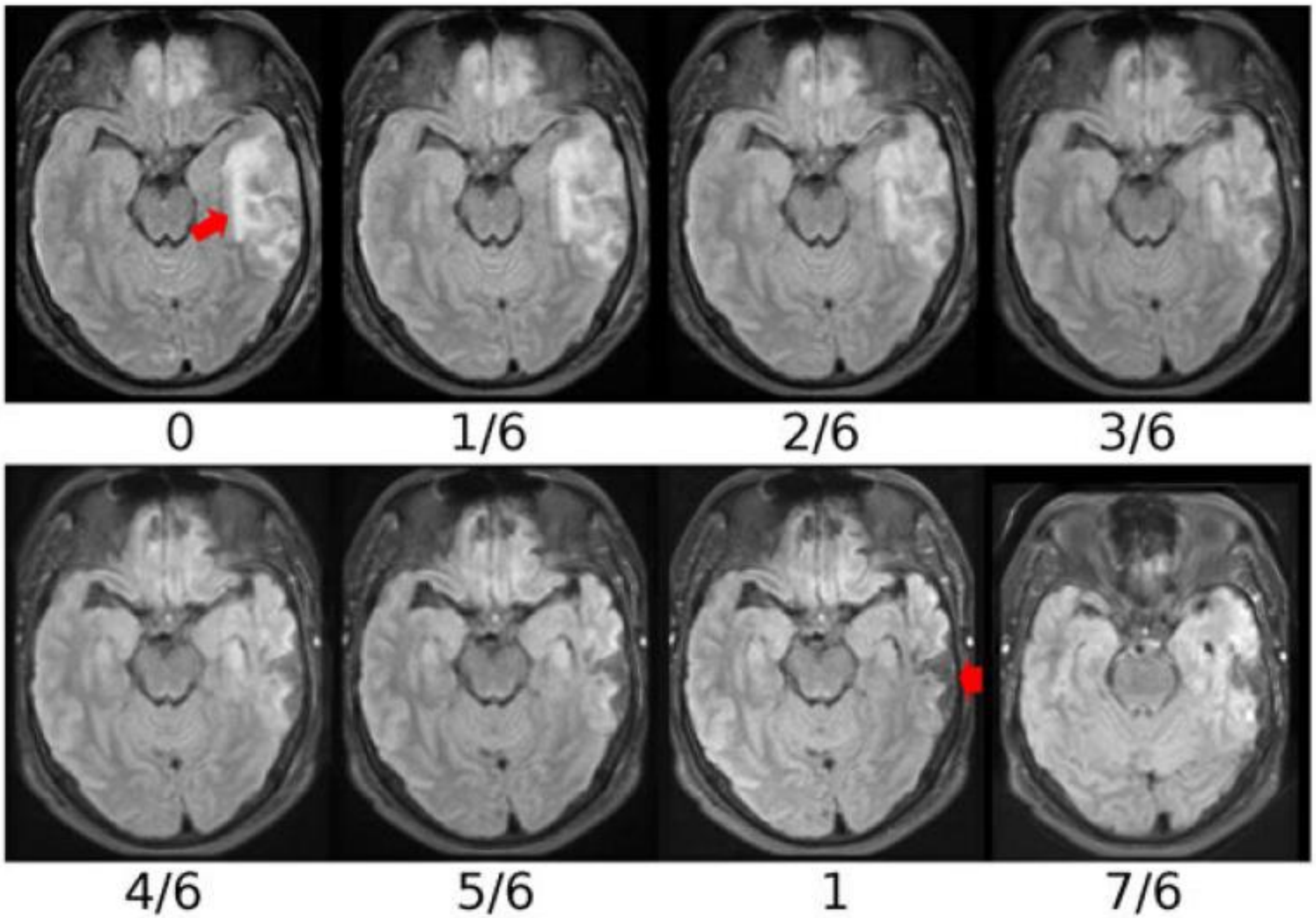
- Econometrics, fluid dynamics, automatic control, statistical physics, shape optimization, expert systems, meteorology, spectral analysis, time-series analysis, and many more fields.



Glymphatic System



Interpolation and Prediction: TBI



Motivation: Cancer Network as Robust System

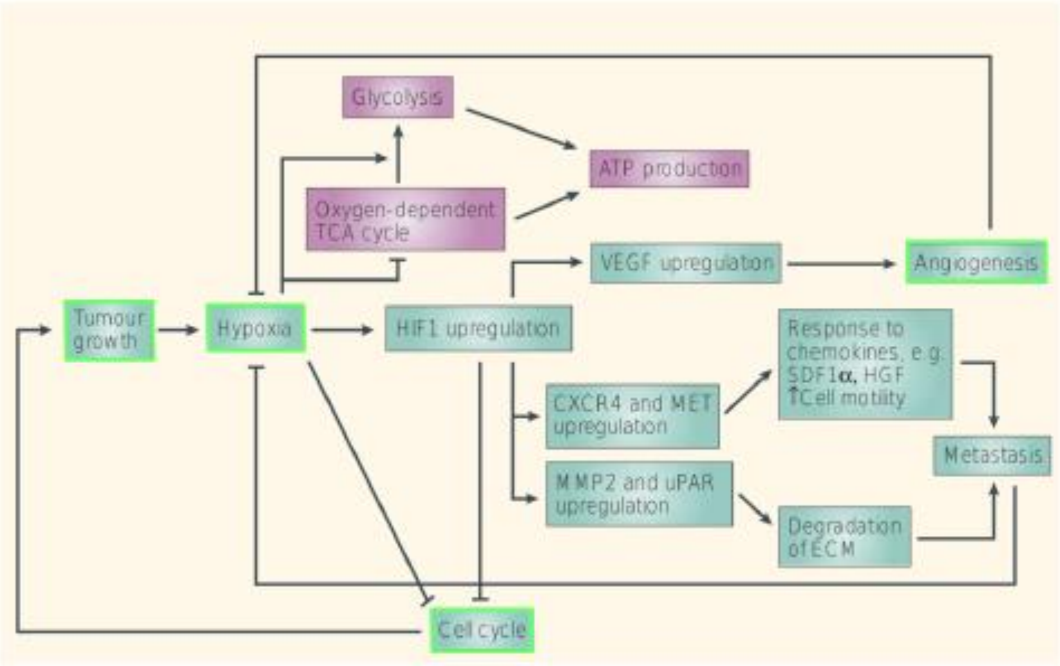


Figure : Feedback loops for hypoxia responses of tumor cells [4]

- *How to* ***quantitatively*** *measure the robustness?*

[4] Kitano, Cancer as a robust system: implications for anticancer therapy, *Nature Reviews Cancer*, 2004

Robustness & Fragility

Network Robustness & Fragility:

- If node/edge x is perturbed, how does the network react to such a change. A highly robust network continues to operate in a similar manner with respect to its functionality.

Quantitively:

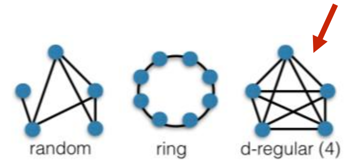
- Consider a network perturbation (fluctuation) that will result in a deviation of an observable from its unperturbed value. How quickly will this return to equilibrium (e.g., decay rate)?
- Let $p_e(t)$ denote the probability that the mean deviates by more than e at time t (with $p_e(t) \to 0$ as $t \to \infty$), then

$$R := \lim_{t\to\infty}\left(-\frac{1}{t}\log p_e(t)\right)$$

measures the decay rate [1].

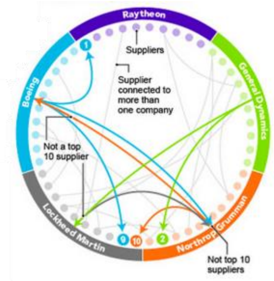
- Robustness is measured as the ability to withstand perturbations (noise) or stochastic fluctuations to a network yet still allow for ``information to be passed'' in a reliable manner.

I. GRAPH THEORY

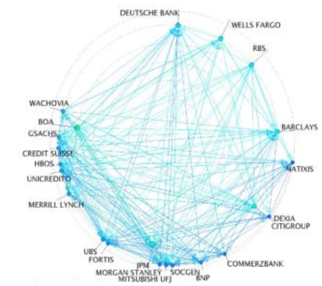


Which is hardest to "bring down"

II. SUPPLY CHAIN AND FINANCIAL NETWORKS - CRISIS

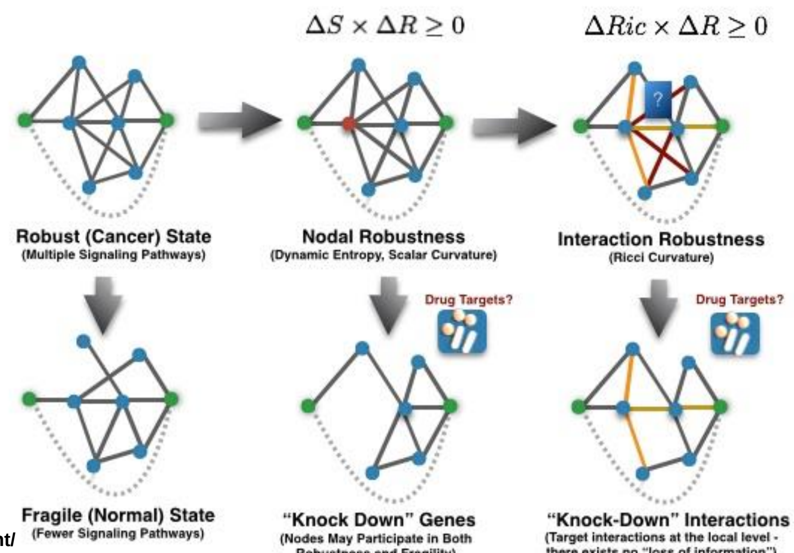


DoD Supply Chain [2]
(risk propagates through primes and subs)



2008 Financial Crisis [3]
(small/large banks - risk exposure)

III. BIOLOGICAL NETWORKS - CANCER



[1] Albert, R. et al. Statistical mechanics of complex networks. Reviews of Modern Physics. 74, 47 (2002).
[2] http://about.bgov.com/bgov200/bgov-analysis/competition-cooperation-among-defense-contractors-bgov-insight/
[3] Battiston, S. et al. DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk. Scientific Reports 2 (2012).

Wasserstein Distance

Wasserstein 1-Metric:

Let μ_1 and μ_2 now be two discrete distributions with same total mass over *n* points, respectively, and let $d(x,y)$ represent the distance between such samples (for the case of graphs, this is simply taken to be the hop distance). Then, $W_1(\mu_1, \mu_2)$ may be described as follows:

$$W_1(\mu_1, \mu_2) = \min \sum_{i,j=1}^{n} d(x_i, x_j)\mu(x_i, x_j)$$

where $\mu(x, y)$ is a coupling (or flow) subject to the following constraints:

$$\begin{aligned} \mu(x,y) &\geq 0, \\ \sum_{j=1}^{n} \mu(x, y_j) &= \mu_1(x), \ \forall x, \\ \sum_{i=1}^{n} \mu(x_i, y) &= \mu_2(y), \ \forall y. \end{aligned}$$

The cost above finds the optimal coupling of moving a set of mass from distributions μ_1 to μ_2 with minimal "work" [4].

[4] **Rubner, Y et. al. The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision. 42 (2000)**

Generalities on Ricci Curvature

Curvature:

- Curvature, in the broad sense, is a measure by which a geometrical object deviates from being flat, and is defined in varying manners given context [5].

Sectional Curvature:

- For M an n-dimensional Riemannian manifold, $x \in M$, let $T_x M$ denote the tangent space at x, and $u_1, u_2 \in T_x M$ orthonormal vectors. Then for geodesics $\gamma_i(t) := \exp(tu_i)$, i = 1,2, the *sectional curvature* $K(u_1,u_2)$ measures the deviation of geodesics relative to Euclidean geometry, i.e.,

$$d\left(\gamma_1(t), \gamma_2(t)\right) = \sqrt{2}t\left(1 - \frac{K(u_1,u_2)}{12}t^2 + O(t^4)\right)$$

Ricci Curvature:

- The Ricci curvature is the average sectional curvature. Namely, given a (unit) vector $u \in T_x M$, we complete it to an orthonormal basis $u, u_2, \ldots, u_n$. Then the Ricci *curvature* is defined by

$$Ric(u) := \frac{1}{n-1}\sum_{i=2}^{n} K(u,u_i)$$

Where we note there might be several scaling factors and it may be extended to the quadratic form, yielding the so-called Ricci *curvature* tensor. Ricci curvature is also strongly related to the Laplace-Beltrami operator and in geodesic normal coordinates, we have

$$R_{ij} = -3/2\Delta g_{ij}$$

where g_{ij} denotes the metric tensor on *M.*

[5] **DoCarmo, M. Riemannian Geometry (Birkhauser, 1992).**

Generalities on Ricci Curvature

Ricci Curvature (con't):

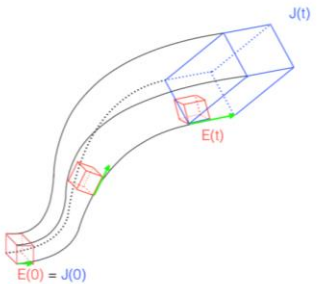
- We can alternatively describe Ricci curvature as the spreading of geodesics. Let γ denote a geodesic and γ_s a smooth one parameter family of geodesics with $\gamma_0 = \gamma$. Then a *Jacobi field* may be defined as

$$J(t) = \frac{d\gamma_s(t)}{ds}\Big|_{s=0}$$

It may be regarded as an infinitesimal deformation of the given geodesic. Then it is standard that *J(t)* (essentially the Jacobian of the exponential map) satisfies the Jacobi equation:

$$\frac{D^2}{dt^2}J(t) + R(J(t),\dot{\gamma}(t))\dot{\gamma}(t) = 0,$$

where $\frac{D}{dt}$ denotes covariant derivative, and *R* is the Riemann curvature tensor.



Curvature in Terms of Jacobian

Discrete Spaces:

- We want to extend these notions to **discrete graphs and networks** - ordinary differentiability does not apply. A nice argument (due to Villani)[6] approaches this problem through convexity. More precisely, let $f: R^n \to R$. Then if f is C^2, convexity may be characterized as $\nabla^2 f(x) \geq 0$ for all x. One may also define convexity in a synthetic manner:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

Following this, one may define a **synthetic notion of Ricci curvature** in terms of so-called *displacement convexity* inherited from the Wasserstein geometry on probability measures.

[6] **Lott, J. & Villani, C. Ricci curvature for metric-measure spaces via optimal transport. Annals of Mathematics. 169, 903-991 (2009).**

Explaining Curvature to Boltzmann

- Displacement convexity/concavity

$$S(\mu_t) \geq tS(\mu_0) + (1-t)S(\mu_1) + \frac{Kt(1-t)}{2} W(\mu_0, \mu_1)^2, 0 \leq t \leq 1 \tag{19}$$

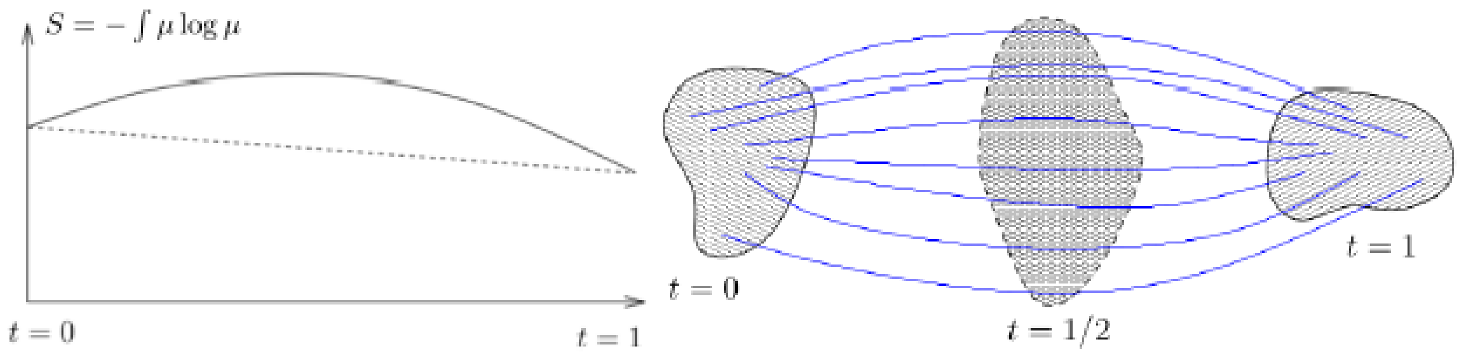


Figure : The lazy gas experiment. [6]

[6] Villani. Optimal Transport. Old and New. 2008.

Ricci Curvature and Entropy

Lott & Villani [6]:

Let (X,d,m) denote a geodesic space, and set:

$$\begin{aligned} P(X,d,m) \quad &:= \quad \{\mu \geq 0 : \int_X \mu \, dm = 1\}, \\ P^*(X,d,m) \quad &:= \quad \{\mu \in P(X,d,m) : \lim_{\varepsilon \searrow 0} \int_{\mu \geq \varepsilon} \mu \log \mu \, dm < \infty\}. \end{aligned}$$

We define

$$H(\mu) := \lim_{\varepsilon \searrow 0} \int_{\mu \geq \varepsilon} \mu \log \mu \, dm, \text{ for } \mu \in P^*(X,d,m),$$

Which is the negative of the *Boltzmann entropy* $S_e(\mu) := -H(\mu)$; note concavity of S_e is equivalent to the convexity of H. Then we say that X has *Ricci curvature bounded from below by k* if for every $\mu_0, \mu_1 \in P(X)$ there exists a constant speed geodesic μ_t with respect to the Wasserstein 2-metric connecting μ_0 and μ_1 such that

$$S_e(\mu_t) \geq t S_e(\mu_0) + (1-t) S_e(\mu_1) + \frac{kt(1-t)}{2} W(\mu_0, \mu_1)^2, \qquad 0 \leq t \leq 1$$

This indicates the **positive correlation** of entropy and curvature that we will express as

$$\Delta S_e \times \Delta Ric \geq 0$$

We now need to connect Ricci curvature and entropy to the notion of robustness (next slide) as well as define appropriate notions of curvature/entropy for discrete spaces (graphs).

[6] **Lott, J. & Villani, C. Ricci curvature for metric-measure spaces via optimal transport. Annals of Mathematics. 169, 903-991 (2009).**

Curvature and Robustness

Recall Definition of Robustness:

- If we let $p_e(t)$ denote the probability that the mean deviates by more than ϵ at time t (with $p_e(t) \to 0$ as $t \to \infty$), then

$$R := \lim_{t \to \infty} \left(-\frac{1}{t} \log p_e(t) \right)$$

measures the decay rate.

Fluctuation Theorem:

- In thermodynamics, it is well-known that entropy and rate functions from large deviations are closely related.

The Fluctuation Theorem is a realization of this fact for networks and can be expressed as:

$$\Delta S_e \times \Delta R \geq 0$$

This can now be further extended to be

$$\Delta Ric \times \Delta R \geq 0.$$

- The Fluctuation Theorem has consequences for just about any type of network: **biological**, **communication**, **social**, or **neural**. In rough terms, it means that the ability of a network to maintain its functionality in the face of perturbations (internal or external), can be quantified by the correlation of activities of various elements that comprise the network.

Network Entropy & Curvature:

- Given a Markov chain , $\mu = (\mu_x), \sum_y \mu_x(y) = 1,$

Network Entropy can be defined as

$$\bar{S}_e = \sum_x p_x S_e(x) \qquad S_e(x) = -\sum_y \mu_x(y) \log \mu_x(y)$$

- We now need an **appropriate definition of Ricci curvature** for a network.

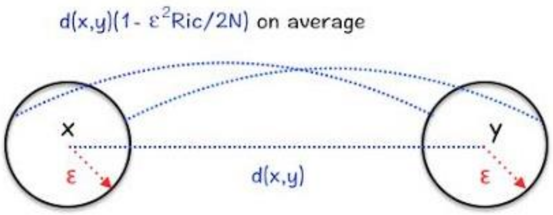
Ollivier-Ricci Curvature

Motivation:

- We employ the notion of Ollivier-Ricci curvature motivated by adopting coarse geometric properties [7]

- Two very close points x and y with tangent vectors w and w′, in which w′ is obtained by a parallel transport of w, the two geodesics will get closer if the curvature is positive.

- Distance between two small (geodesic balls) is less than the distance of their centers. Ricci curvature along direction x-y reflects this, averaged on all directions w at x.



Pictorial Motivation for Ollivier Ricci Curvature

Definition:

Formally, we define for (X,d) a metric space equipped with a family of probability measures $\{\mu_x : x \in X\}$, the *Ollivier-Ricci curvature* $k(x,y)$ along the geodesic connecting x and y via

$$W_1(\mu_x,\mu_y) = (1 - k(x,y))d(x,y)$$

where W_1 denotes the Wasserstein 1-metric defined previously and *d(x,y)* is the geodesic (hop) distance on a graph. For the case of weighted graphs, we set

$$d_x = \sum_y w_{xy}$$

$$\mu_x(y) := \frac{w_{xy}}{d_x}$$

and the sum is taken over all neighbors of x where w_{xy} denotes the weight of an edge connecting *x* and *y* (it is taken as zero if there is no connecting edge between *x* and *y*). The measure μ_x may be regarded as the distribution of a one-step random walk starting from *x*.

[7] Ollivier, Y. Ricci curvature of metric spaces. C. R. Math. Acad. Sci. Paris. 345, 643-646 (2007)

Ornstein-Uhlenbeck (OU) Process

Very informative to consider the relationship of (Ollivier-)Ricci curvature and robustness via a simple example. Consider the following OU process:

$$dX_t = -\alpha X_t dt + \sigma dW_t, \quad X(0) = x_0$$

where W is Brownian motion (Wiener process), and we take x_0 to be deterministic. We treat the 1-dimensional case for simplicity. Everything goes through in higher dimensions as well. The corresponding Fokker-Planck equation is

$$\frac{\partial p}{\partial t} = \alpha \frac{\partial xp}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 p}{\partial x^2},$$

where $p = p(x,t|x_0,0)$ is the transition probability of the underlying Markov process. One may show that $p(x,t|x_0,0)$ is a Gaussian process with mean and variance given by $\langle X(t) \rangle = x_0 e^{-\alpha t}, \mathrm{var}\, X(t) = \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha t}\right).$ We see that we get transition probabilities of mean $x_0 e^{-\alpha t}$ and variance independent of x_0. Since all the transitions $p(x,t|x_0,0)$ have the same variance (and are Gaussian) the 1-Wasserstein distance

$$W_1(p(x,t\,|\,x_0,0), p(x,dt\,|\,x_1,0)) = \left| x_0 e^{-\alpha t} - x_1 e^{-\alpha t} \right|.$$

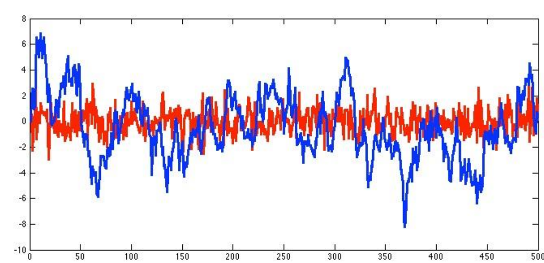
Finally,

$$\kappa(x_0,x_1) = 1 - \frac{W_1(p(x,t\,|\,x_0,0), p(x,t\,|\,x_1,0))}{|x_0 - x_1|} = 1 - e^{-\alpha t}.$$

This implies

$$\Delta Ric \times \Delta\alpha \geq 0$$

Larger α corresponds to larger curvature κ and this corresponds to how quickly the systems returns to equilibrium, that is to the mean going to 0.



Simulation of Two OU Process

[7] Ollivier, Y. Ricci curvature of metric spaces. C. R. Math. Acad. Sci. Paris. 345, 643-646 (2007)

Invariant Distribution and Triangle Density

Convergence to Invariant Distribution:

Larger Ollivier-Ricci curvature indicates greater robustness via rate of convergence to the invariant (equilibrium) distribution. Specifically, suppose $\kappa(x,y) \geq k > 0$. Then there exists a unique invariant probability measure ν. Moreover, for any x,

$$W_1(\mu_x^{*t}, \nu) \le \frac{W_1(\delta_x, \mu_x)}{k}(1-k)^t.$$

Here,

$$\mu_x^{*t}(y) := \sum_{z \in X} \mu_x^{*(t-1)} \mu_z(y), \qquad \mu_x^{*1} := \mu_x.$$

Note that $W_1(\delta_x, \mu_x)$ represents the jump of the random walk at *x*. On a connected graph *X* with diameter *D* (defined as the longest graph geodesic), this yields the following estimate for the mixing time [7]:

$$\frac{1}{2} \sum_{y \in X} \left| \mu_x^{*t}(y) - \nu(y) \right| \le D(1-k)^t.$$

The relationship of robustness to the Ollivier-Ricci curvature is again seen for the case of Markov chains.

"Triangle Density":

On an unweighted graph, the lower bound for the Ricci curvature k(x, y) for x adjacent to y becomes:

$$k(x,y) = -\left(1 - \frac{1}{d_x} - \frac{1}{d_y} - \frac{\sharp(x,y)}{d_x \wedge d_y}\right)_+ - \left(1 - \frac{1}{d_x} - \frac{1}{d_y} - \frac{\sharp(x,y)}{d_x \vee d_y}\right)_+ + \frac{\sharp(x,y)}{d_x \vee d_y} + \frac{c(x)}{d_x} + \frac{c(y)}{d_y},$$

where $\#(x,y) := \sum_{x_1 \in N_{xy}} 1$ is the number of triangles containing x,y, c(x)=0 or 1 is the number of loops at x [8].

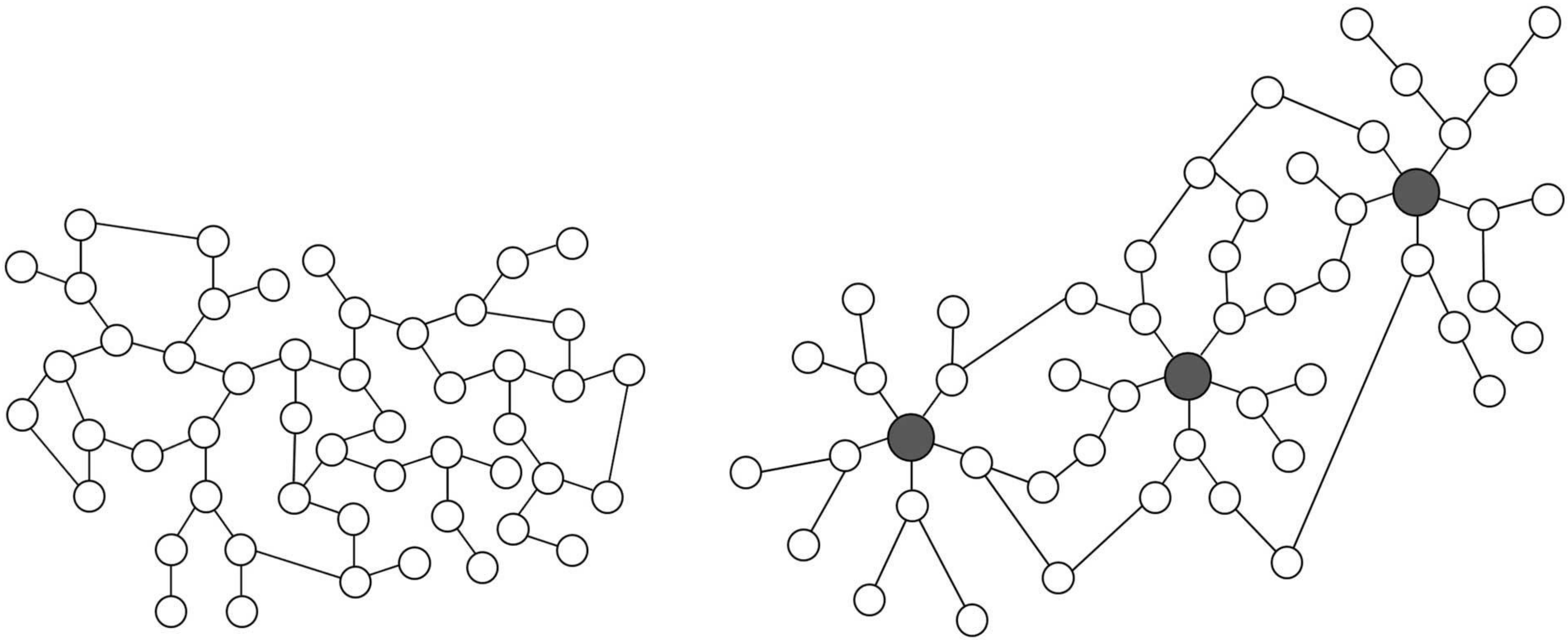
This indicates multiple signaling pathways correlates to Ricci curvature (robustness)

[7] Ollivier, Y. Ricci curvature of metric spaces. C. R. Math. Acad. Sci. Paris. 345, 643-646 (2007)

[8] Bauer, F., Jost, J. & Liu, S. Ollivier-Ricci curvature and the spectrum of the normalized graph Laplacian operator. http://arxiv.org/abs/1105.3803 (2011)

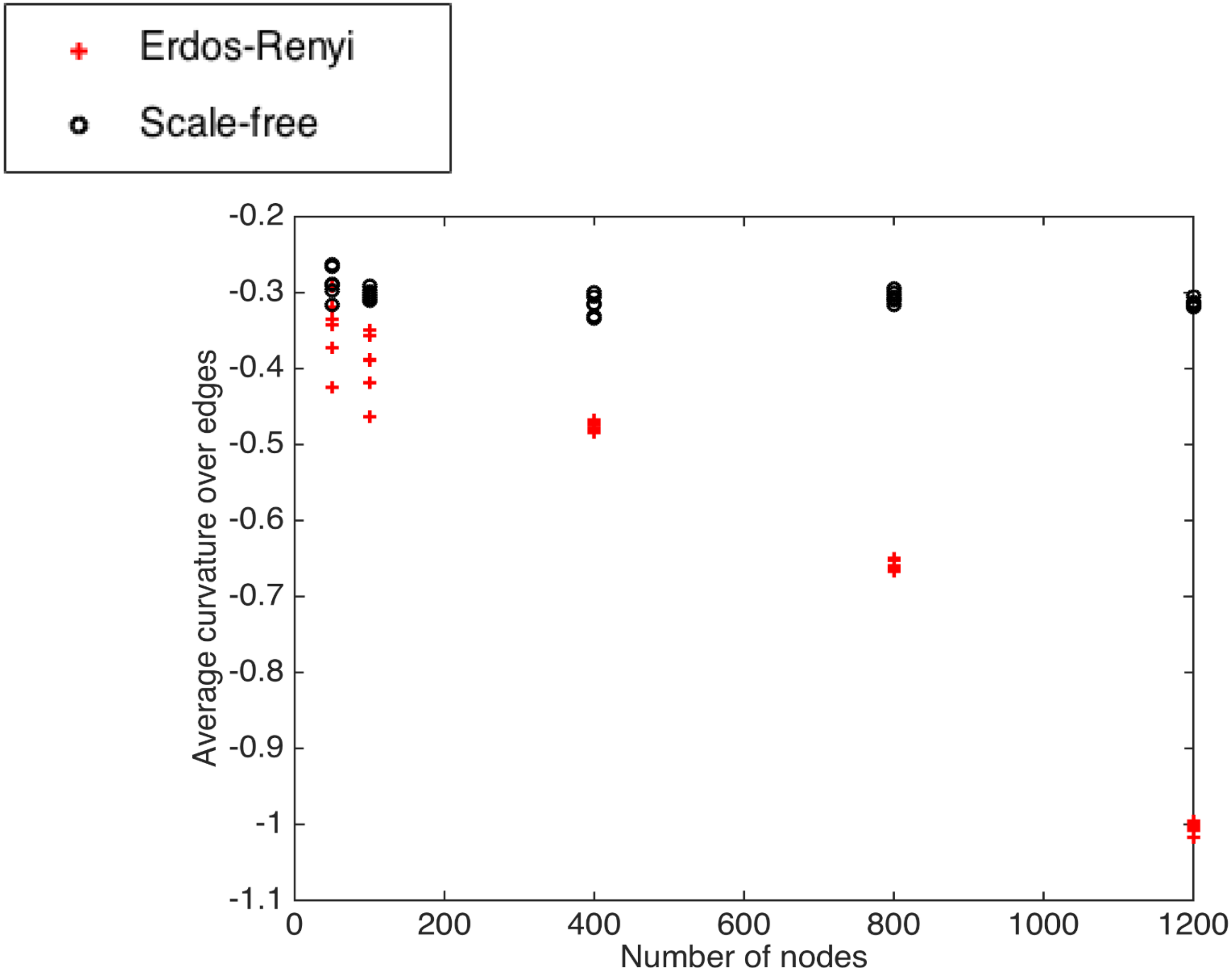
Erdos-Renyi network vs Scale-free network

Poisson vs power law



(A) Random network (B) Scale-free network

https://www.spandidos-publications.com



The average curvature of scale-free network is larger

Curvature: Cancer Hallmark?

Is Curvature a Cancer Hallmark?: Analysis

Comparison of Curvatures

Cancer Type	Δ Average OR Curvature	Δ Average BER Curvature	Δ Average Forman Curvature
Breast Carcinoma	0.012	0.182	13.022
Head/Neck Carcinoma	0.004	0.116	9.100
Kidney Carcinoma	0.010	0.217	7.711
Liver Carcinoma	0.008	0.227	3.136
Lung Adenocarcinoma	0.013	0.320	7.898
Prostate Adenocarcinoma	0.009	0.179	7.368
Thyroid Carcinoma	0.006	0.133	2.969

- All three notions of Ricci curvature have a higher average value in all seven cancer networks compared to the complementary normal networks.

Drug Targets: Ewing Sarcoma

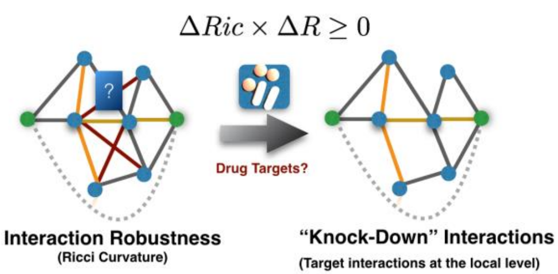
Highlights of a novel approach based on Network Curvature

- Represent biological systems as complex dynamically evolving networks
- Quantify the ability to withstand perturbations by a mathematical notion of "robustness"
- Utilize the recently discovered relation between Ricci Curvature ("Ric") and robustness ("R"):

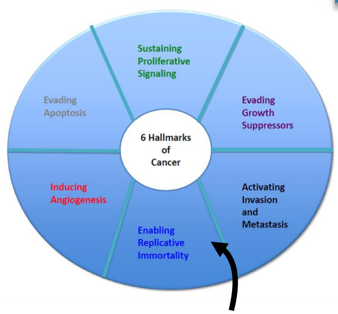
$$\Delta Ric \times R \geq 0$$

- **Goal**: Systematically uncover "targets of opportunity" in an adaptive manner

Biological Systems as Weighted Graph

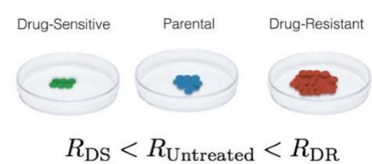


Cancer Hallmark?

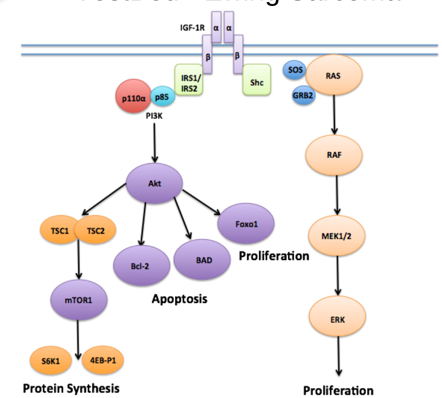


Ricci Curvature, Cancer Hallmark?

Drug Resistance



TestBed - Ewing Sarcoma



<u>After static treatments, what gives rise to modes of resistance and ability for system (disease) to adapt?</u>

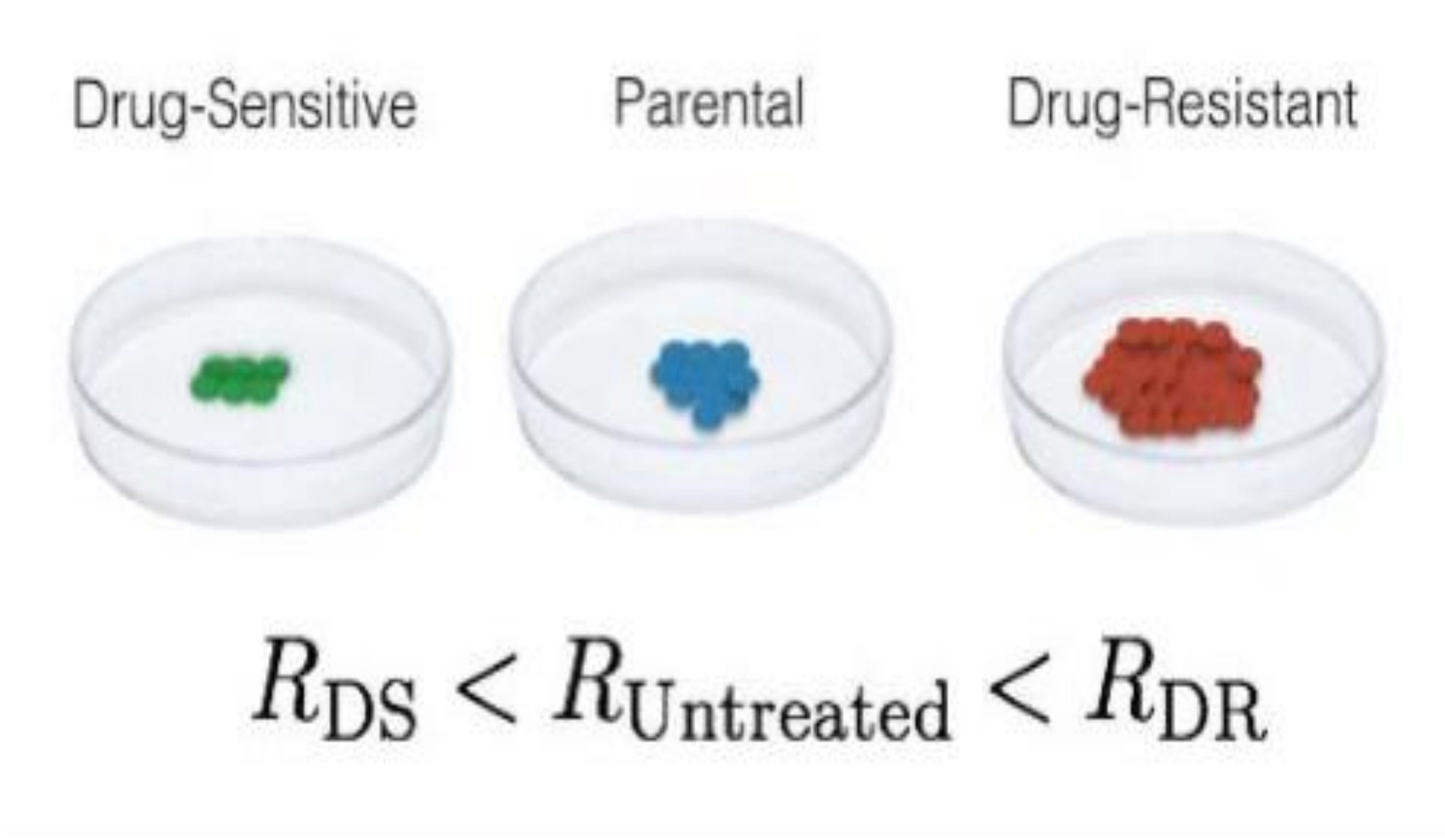
Preliminary Results: Cancer vs Normal Tissue

- Our studies performed showing cancer tissues has larger curvature (increase robustness or ability to adapt) than normal tissue
- Studies focused on variety of cancer types: Breast, Lung, Liver, Head/Neck, Kidney, Thyroid
- Refer the reader to a recently published manuscript - "Graph Curvature for Differentiatng Cancer Networks"

Significance to RFI

- Employing network curvature to dynamically understand sensitivity/resistance in cancer (i.e., pre-treatment, post-treatment)
- Approach can be utilized for vaccine treatment, drug persistence– the key is ability to quantify fragility and robustness of system
- Safety and Efficacy Procedures are stringently in place with collaborators at MSKCC and MD Anderson (e.g.,xenograft mice, cell lines)

Drug Resistance



Initial Preliminary Results

Global Network Fragility via Ricci Curvature:

	72-Hour	Untreated	Resistant
Average Curvature[1]	0.0071	0.0329	0.0367
5 % Left Tail (Avg.)	-0.4619	-0.4428	-0.4123
1% Left Tail (Avg.)	-0.8535	-0.8246	-0.7822
Min Curvature	-1.9906	-1.7771	-1.4625

Notes:

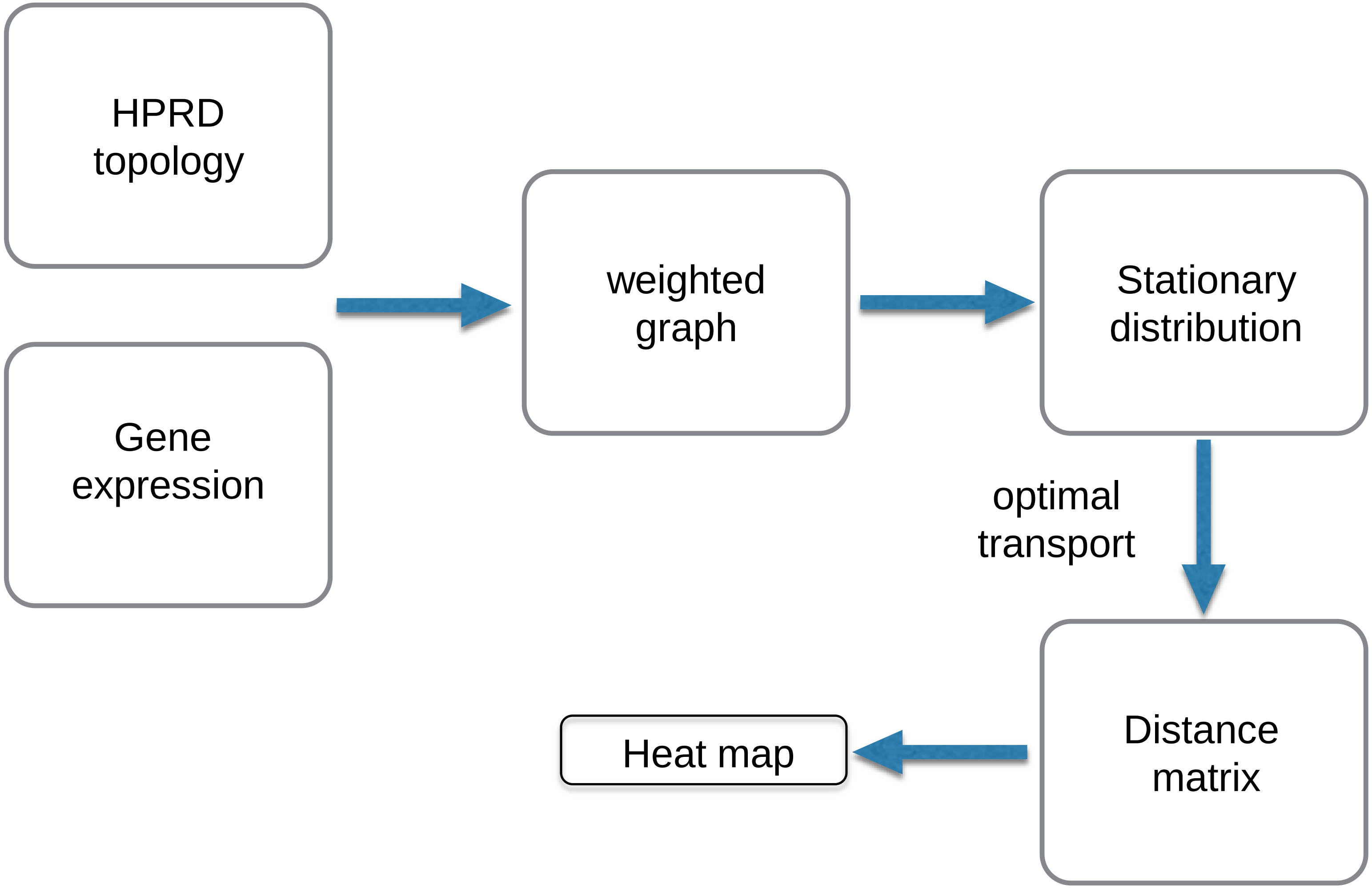
- We quantify that resistant tumors are more robust 72-hour/untreated via curvature.
- The most fragile case is the 72-Hour
- This coincides with our initial hypothesis and with our previous cancer studies

Local Protein Interaction Fragility via Scalar Curvature:

	72-Hour	Untreated	Resistant
mTor	1.7404	3.5812	0.4081
MEK	1.2259	1.7511	1.8042

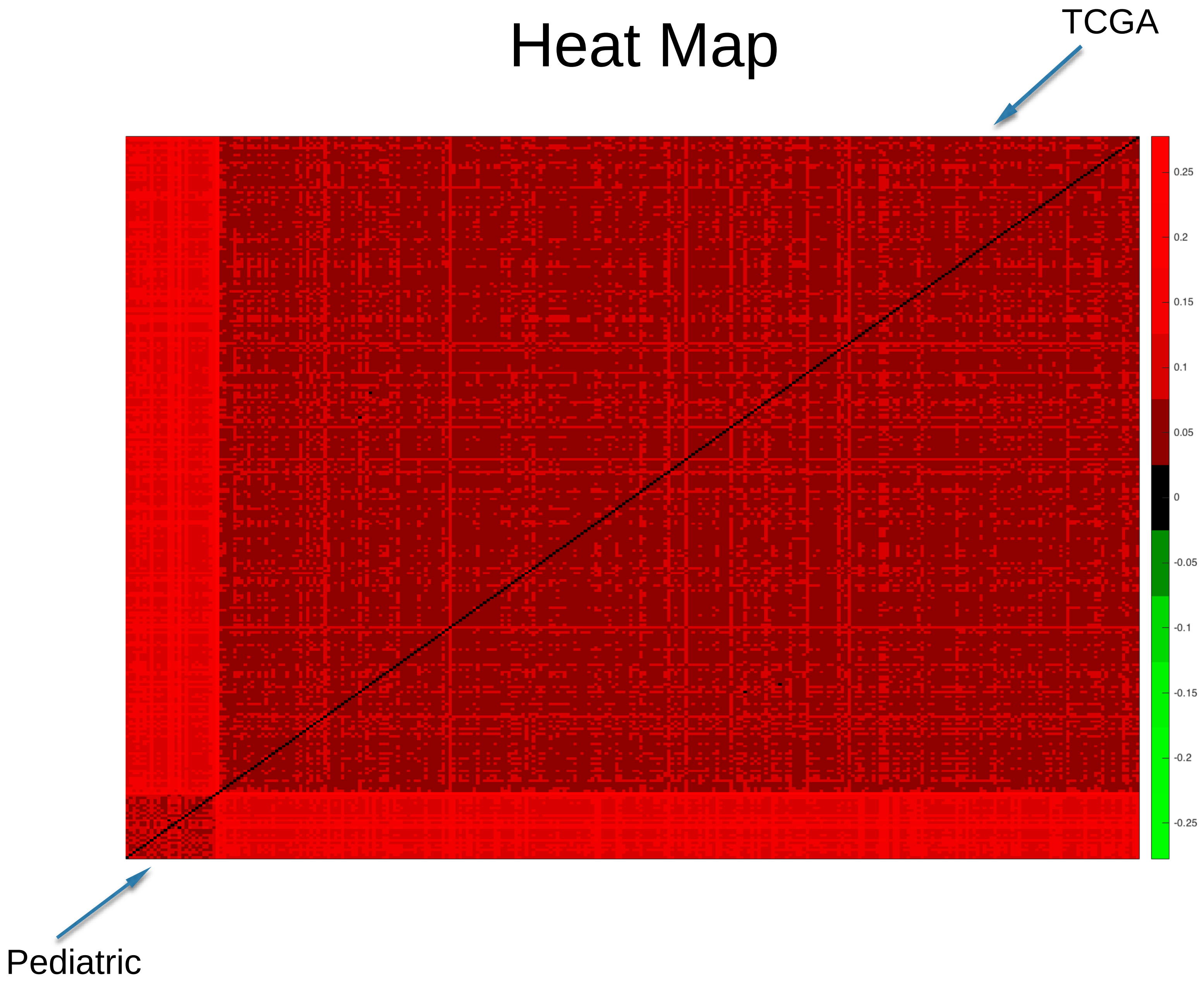
Notes:

- We noticed all (direct/indirect) pathways to mTor become “fragile” during resistant and 72-hour case
- MEK pathways becomes more robust in resistant case
- We caution these local results are too preliminary to draw convulsive evidence



Sketch of Pipeline for Sarcoma Data Clustering

Heat Map

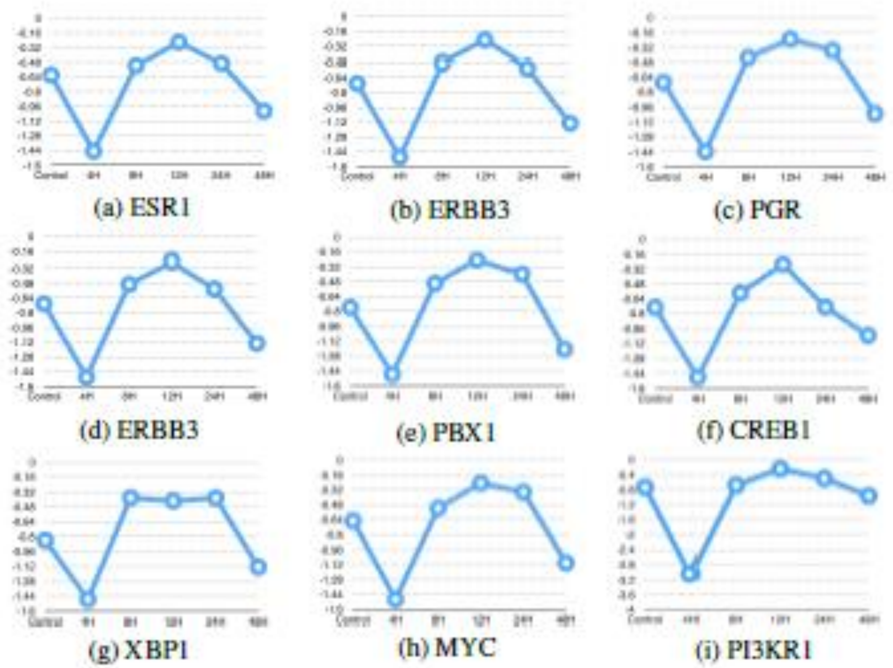


Breast Cancer

(Collaborator: MSKCC - Baselega Group)

ER Positive Breast Cancer

- PI3K inhibition induces Estrogen Receptor (ER) activity
- ER-related genes are mechanisms of resistance
- Time-Varying Treatment with BYL710 (PI3K Inhibitor)
- Measured expression at 4H, 8H, 12H, 24H, 48H
- Goal: Uncover Targets



(a) ESR1 (b) ERBB3 (c) PGR (d) ERBB3 (e) PBX1 (f) CREB1 (g) XBP1 (h) MYC (i) PI3KR1

Ricci Curvature: During initial treatment, activity of genes exhibits fragility prior to building resistance and then subsides. This is in line with gene expression data where maximal expression of ER-related genes is seen at the 24 hour mark and then subsides. Effect seems to be greatest on PI3KR1, which makes biological sense since we are considering the effect of a PI3K inhibitor. In this case, we see a very large increase in fragility at 4 hours (exhibited by large negative curvature). Results are presented minus scaling factor of 1e3 for figures (a)-(h) and 1e5 for figure (i).

Autism: Connectome

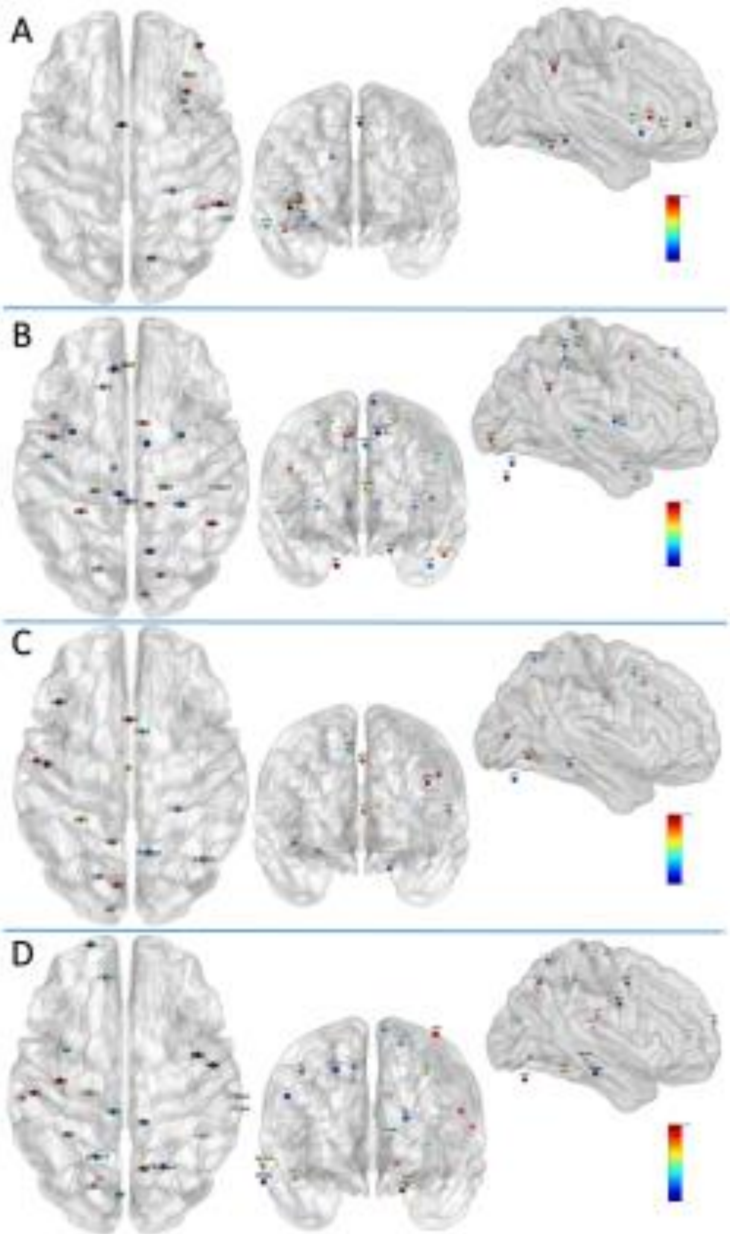
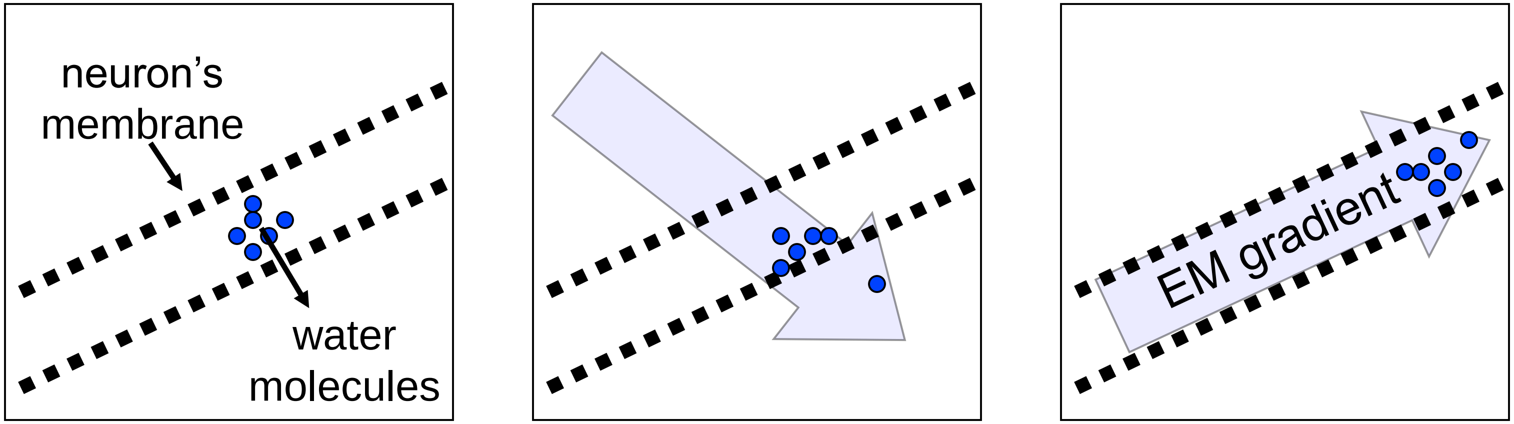


Figure 2: Structural network nodes with significant changes in ASD cohort as compared to TD. Node color shows p-value. (A) Node Curvature. (B) Node Strength. (C) Betweenness Centrality. (D) Clustering Coefficient. **Node Name Abbreviations:** Prefix R = Right Hemisphere, L = Left Hemisphere; AG = Angular-Gyrus; CC = Cuneal-Cortex; CGad = Cingulate-Gyrus-anterior-division; CGpd = Left-Cingulate-Gyrus-posterior-division; COC = Central-Opercular-Cortex; FOC = Frontal-Orbital-Cortex; Frontal-Orbital-Cortex; FP = Frontal-Pole; IC = Insular-Cortex; IcC = Intracalcarine-Cortex; IT-Gtp = Inferior-Temporal-Gyrus-temporooccipital-part; JLC = Juxtapositional-Lobule-Cortex; LG = Lingual-Gyrus; LOCid = Lateral-Occipital-Cortex-inferior-division; LOCsd = Lateral-Occipital-Cortex-superior-division; MTGad = Middle-Temporal-Gyrus-anterior-division; MTGpd = Middle-Temporal-Gyrus-posterior-division; OFG = Occipital-Fusiform-Gyrus; OP = Occipital-Pole; P = Putamen; PaG = Paracingulate-Gyrus; PC = Precuneous-Cortex; PG = Postcentral-Gyrus; PGad = Parahippocampal-Gyrus-anterior-division; PGpd = Parahippocampal-Gyrus-posterior-division; PrG = Precentral-Gyrus; SFG = Superior-Frontal-Gyrus; SGpd = Supramarginal-Gyrus-posterior-division; SPL = Superior-Parietal-Lobule; T = Thalamus; TOF = Temporal-Occipital-Fusiform-Cortex.

Application:
Diffusion MRI tractography

- Diffusion MRI measures the diffusion of water molecules in the brain
- Neural fibers influence water diffusion
- Tractography: “recovering probable neural fibers from diffusion information”



fMRI and DTI for IGS

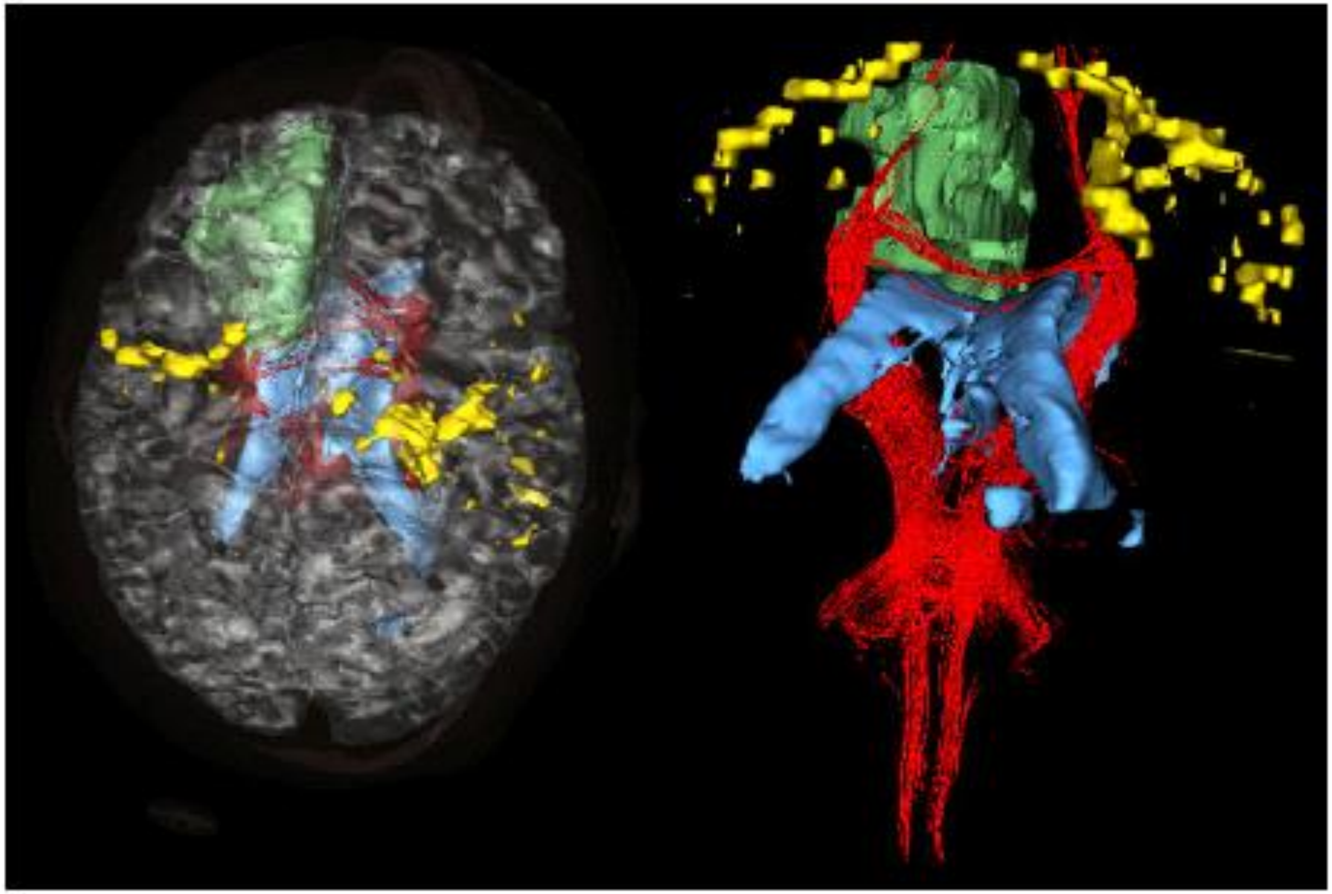


Figure 8.4.6-1. Retrospective Example of fMRI for Neurosurgical Application
62-year-old female patient with left frontal hyperintense non-enhancing mass lesion
Skin, Brain, Ventricles (blue) and Tumor (green) models from conventional MRI; fMRI activations (yellow) from pre-operative finger-taping experiment. Fiber tract indications (red) from Diffusion Tensor MRI.
Imaging suggests that the tumor is in front of motor strip with involvement of supplementary motor area, with fibers from SMA piercing tumor in its posterior aspect.

Benamou-Brenier Framework

- extend the **Benamou-Brenier** framework to transport of
  - Hermitian matrices (Quantum density matrices)
  - matrix-valued distributions

I.e., formulate for matrices...

$$\inf \int \int_0^1 \rho(t,x) \|v(t,x)\|^2 \, dt \, dx$$

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) \ = \ 0,$$

$$\rho(0,\cdot) = \rho_0, \ \rho(1,\cdot) = \rho_1$$

Quantum continuity equation

Starting point: Lindblad equation (in "diagonal form" $L_k = L_k^*$)

$$\dot{\rho} = -[iH, \rho] + \sum_{k=1}^{N} (L_k \rho L_k - \frac{1}{2}\rho L_k L_k - \frac{1}{2} L_k L_k \rho),$$

Notation:

$\mathcal{H}$ and $\mathcal{S}$ the set of $n \times n$ Hermitian and skew-Hermitian matrices

$\mathcal{H}_+$ and $\mathcal{H}_{++}$ nonnegative and positive-definite matrices

$\mathcal{D}_+ := \{\rho \in \mathcal{H}_{++} \mid \mathrm{tr}(\rho) = 1\}$ "density matrices"

$\mathcal{S}^N$, $\mathcal{H}^N$ block-column vectors with matrix-entries

Some calculus

Note for functions:

$$f(x) \ : \ g(x) \mapsto f(x)g(x)$$

$$\partial_x \ : \ g(x) \mapsto \partial_x g(x)$$

$$[\partial_x, f(x)] \ : \ g(x) \mapsto \partial_x f(x)g(x) - f(x)\partial_x g(x) = (\partial_x f(x))g(x)$$

For matrices:

$$\partial_{L_i} X = [L_i, X] = [L_i X - X L_i]$$

define the *gradient operator* for $L \in \mathcal{H}^N$

$$\nabla_L : \mathcal{H} \to \mathcal{S}^N, \quad X \mapsto \begin{bmatrix} L_1 X - X L_1 \\ \vdots \\ L_N X - X L_N \end{bmatrix}$$

Some calculus

∇_L is a derivation

$$\begin{aligned}\nabla_L(XY+YX) = (\nabla_L X)Y + X(\nabla_L Y) \\ + (\nabla_L Y)X + Y(\nabla_L X), \quad \forall X, Y \in \mathcal{H}.\end{aligned}$$

dual is an analogue of the (negative) *divergence operator*:

$$\nabla_L^* : \mathcal{S}^N \to \mathcal{H},\ Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_N \end{bmatrix} \mapsto \sum_k^N L_k Y_k - Y_k L_k.$$

$$\langle \nabla_L X, Y \rangle = \langle X, \nabla_L^* Y \rangle$$

Continuity equation

$$\dot{\rho} = \nabla_L^* M_\rho(v),$$

with $M_\rho(v)$ a *"multiplication"* between ρ and v

momentum field "ρv" $= M_\rho(v) \in \mathcal{S}^N$.

choices of non-commutative multiplication:

i) $\frac{1}{2}(\rho v + v\rho)$ ("anti-commutator")

ii) $\int_0^1 \rho^s v \rho^{1-s} ds$ (Kubo-Mori)

iii) $\rho^{1/2} v \rho^{1/2}$

Case i) “anti-commutator”

Problem i):

$$\boldsymbol{W}_{2,a}(\rho_0,\rho_1)^2 := \min_{\rho\in\mathcal{D}_+, v\in\mathcal{S}^N} \int_0^1 \mathrm{tr}(\rho v^* v)\,dt,$$
$$\dot{\rho} = \frac{1}{2}\nabla_L^*(\rho v + v\rho),$$
$$\rho(0)=\rho_0,\quad \rho(1)=\rho_1,$$

Note: $v^*v = \sum_{k=1}^N v_k^* v_k$ and $v \in \mathcal{S}^N$.

Duality

$\boldsymbol{\lambda}(\cdot) \in \boldsymbol{\mathcal{H}}$ Lagrangian multiplier

$$\boldsymbol{\mathcal{L}}(\rho, v, \boldsymbol{\lambda}) = \int_0^1 \left\{ \frac{1}{2} \mathrm{tr}(\rho v^* v) - \mathrm{tr}(\boldsymbol{\lambda}(\dot{\rho} - \frac{1}{2} \boldsymbol{\nabla}_L^* (\rho v + v\rho))) \right\} dt$$

Point-wise minimization $\Rightarrow$

$$v_{opt}(t) = -\boldsymbol{\nabla}_L \boldsymbol{\lambda}(t).$$

Duality

If $\lambda(\cdot) \in \mathcal{H}$:

$$\dot{\lambda} = \frac{1}{2}(\nabla_L \lambda)^*(\nabla_L \lambda) = \frac{1}{2}\sum_{k=1}^{N}(\nabla_L \lambda)_k^*(\nabla_L \lambda)_k$$

and

$$\dot{\rho} = -\frac{1}{2}\nabla_L^*(\rho \nabla_L \lambda + \nabla_L \lambda \rho)$$

matches the marginals $\rho(0) = \rho_0, \rho(1) = \rho_1$,

then (ρ, v) with $v = -\nabla_L \lambda$ solves Problem i)

Riemannian structure

$$\delta_j \in \mathrm{TangentSpace}_\rho = \{\delta \in \mathcal{H} \mid \mathrm{tr}(\delta) = 0\}, \text{ for } j = 1, 2$$

"Poisson" equation: δ's $\Leftrightarrow$ λ's

$$\delta_j = -\frac{1}{2}\nabla_L^*(\rho\nabla_L\lambda_j + \nabla_L\lambda_j\rho)$$

and

$$\langle\delta_1, \delta_2\rangle_\rho = \frac{1}{2}\mathrm{tr}(\rho\nabla\lambda_1^*\nabla\lambda_2 + \rho\nabla\lambda_2^*\nabla\lambda_1)$$

Note: given δ, then $-\nabla_L\lambda$ is the unique minimizer of $\mathrm{tr}(\rho v^* v)$ over $v \in \mathcal{S}^N$ satisfying

$$\delta = \frac{1}{2}\nabla_L^*(\rho v + v\rho).$$

Matrix transport with added spatial component

$$\mathcal{D} = \{\rho(\cdot) \mid \rho(x) \in \mathcal{H}_+ \text{ such that } \int_{\mathbb{R}^m} \mathrm{tr}(\rho(x))dx = 1\}.$$

Continuity equation: $w \in \mathcal{H}$ along space dimension

$$\frac{\partial \rho}{\partial t} + \frac{1}{2}\nabla_x \cdot (\rho w + w\rho) - \frac{1}{2}\nabla_L^*(\rho v + v\rho) = 0.$$

Metric:

$$\begin{aligned}
W_{2,a}(\rho_0, \rho_1)^2 := \min & \int_0^1 \int_{\mathbb{R}^m} \{\mathrm{tr}(\rho w^* w) + \gamma\, \mathrm{tr}(\rho v^* v)\}\, dxdt, \\
& \rho \in \mathcal{D}_+, w \in \mathcal{H}^m, v \in \mathcal{S}^N \\
& \frac{\partial \rho}{\partial t} + \frac{1}{2}\nabla_x \cdot (\rho w + w\rho) - \frac{1}{2}\nabla_L^*(\rho v + v\rho) = 0, \\
& \rho(0, \cdot) = \rho_0, \quad \rho(1, \cdot) = \rho_1
\end{aligned}$$

Gradient flow of Entropy

$$\begin{aligned}\frac{dS(\rho(t))}{dt} &= \dots \\ &= -\operatorname{tr}((\nabla_L \log\rho)^* \int_0^1 \rho^s v \rho^{1-s} ds),\end{aligned}$$

$\Rightarrow$ greatest ascent direction $v = -\nabla_L \log\rho$.

---

non-commutative analog of: $\partial_x \rho = \rho\, \partial_x(\log\rho)$):

$$\nabla_L \rho = \int_0^1 \rho^s (\nabla_L \log\rho)\rho^{1-s} ds$$

Gradient flow:

$$\dot\rho = -\nabla_L^* \int_0^1 \rho^s (\nabla_L \log\rho)\rho^{1-s} ds = -\nabla_L^* \nabla_L \rho = \Delta_L \rho,$$

Linear heat equation (now Lindblad) just as in the scalar case!

Vector-Valued Densities

A vector-valued density $\rho = [\rho_1, \rho_2, \cdots, \rho_\ell]^T$ on $\mathbb{R}^N$ is a function from $\mathbb{R}^N$ to $\mathbb{R}^\ell_+$ such that

$$\sum_{i=1}^{\ell} \int_{\mathbb{R}^N} \rho_i(x)dx = 1.$$

The set of all vector-valued densities and its interior are denoted by $\mathcal{D}$ and $\mathcal{D}_+$ respectively.

Different ρ_i may represent densities of different types (or particles) but there is a possibility for the mass to transfer from one type to another, that is, we allow mass transfer between different densities. Therefore, the change of density is determined by two factors: the mass flow within the same type of density and the mass transfer between different types of densities. This dynamics may be described by the following ***continuity equation***:

$$\frac{\partial \rho_i}{\partial t} + \nabla_x \cdot (\rho_i v_i) - \sum_{j \neq i} (\rho_j w_{ji} - \rho_i w_{ij}) = 0, \quad \forall i = 1, \ldots, \ell. \tag{1}$$

Here v_i is the velocity field of particles i and $w_{ij} \geq 0$ is the transfer rate from i to j. In (1), all types of densities are treated equally. More generally, the mass transfer between the ℓ types of particles can be modeled by a graph $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1, \mathcal{W}_1)$. Equation (1) corresponds to the case where $\mathcal{G}_1$ is a complete graph with all weights equal to 1. For simplicity, we will stay with the special case (1).

With this continuity equation, given $\mu, \nu \in \mathcal{D}_+$, we can formulate the optimal mass transport problem

$$W_{2,b}(\mu,\nu)^2 := \inf_{\rho \in \mathcal{D}_+, v, w} \int_0^1 \int_{\mathbb{R}^N} \left\{ \sum_{i=1}^{\ell} \rho_i(t,x) \|v_i(t,x)\|^2 + \gamma \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \rho_i w_{ij}^2(t,x) \right\} dx dt \tag{2a}$$

$$\frac{\partial \rho_i}{\partial t} + \nabla_x \cdot (\rho_i v_i) - \sum_{j \neq i} (\rho_j w_{ji} - \rho_i w_{ij}) = 0, \quad \forall i = 1, \ldots, \ell \tag{2b}$$

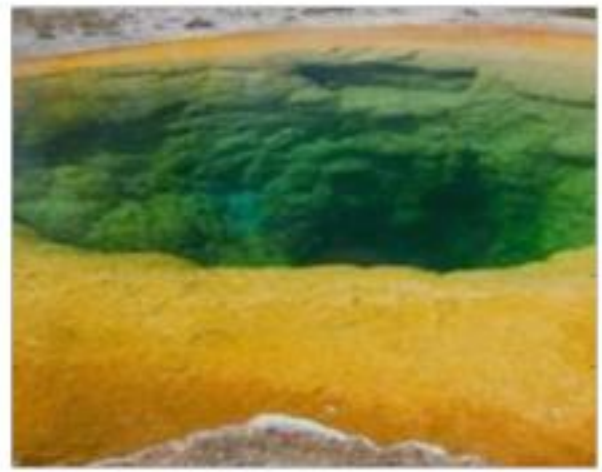
$$w_{ij}(t,x) \geq 0, \quad \forall i, j, t, x \tag{2c}$$

$$\rho(0,\cdot) = \mu(\cdot), \;\; \rho(1,\cdot) = \nu(\cdot) \tag{2d}$$

The coefficient $\gamma > 0$ is a tradeoff parameter between the transport cost for the same type of particles and mass transfer cost between different types of particles. When γ is large, the solution tends to reduce the mass transfer.

Example: Color Images

(a) ρ_0



(b) ρ_1

The marginal distributions shown in Figure are color images (256 by 256) of two geothermal basins in Yellowstone Park, where bacterial growth give them distinctly different colors and hues.

Interpolation of Images

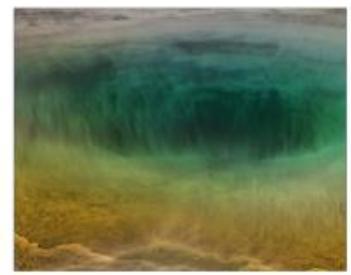
(a) $t = 0.1$

(b) $t = 0.2$

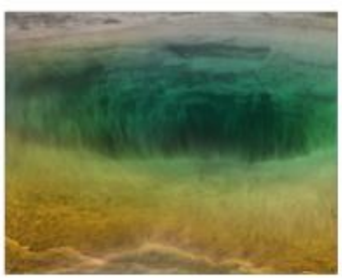
(c) $t = 0.3$

(d) $t = 0.4$

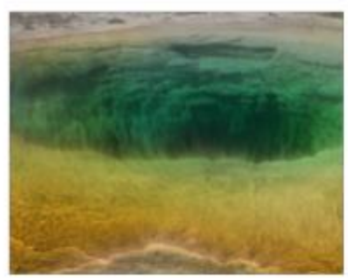
(e) $t = 0.5$



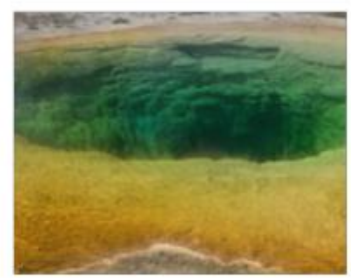
(f) $t = 0.6$



(g) $t = 0.7$



(h) $t = 0.8$



(i) $t = 0.9$

Transcription-Harmonic Oscillator

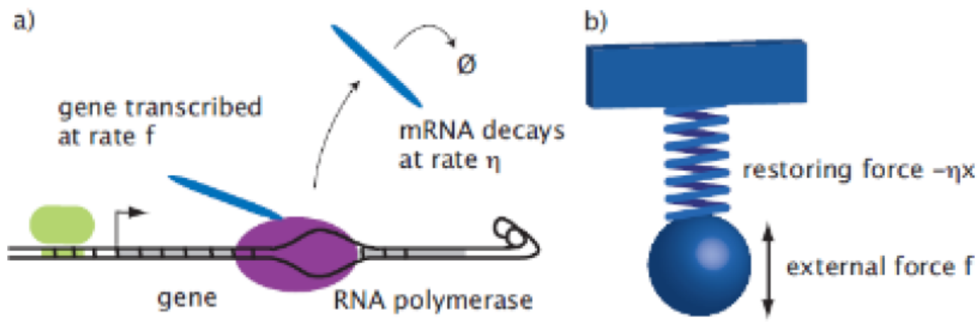


Fig. 1: Transcription and mRNA decay. a) Transcription of a gene is controlled by the binding of transcription factors (left, shown in green) to the regulatory region of a gene. Transcription of a gene leads to the production of mRNA molecules at some rate f. mRNA molecules decay at a rate η per molecule. b) The resulting dynamics of mRNA concentration x can be mapped onto an harmonic oscillator subject to a restoring force $-\eta x$ and an external force f driving the system out of equilibrium.

Jarzynski Fluctuation

Equation (2) is the Langevin equation of the Ornstein-Uhlenbeck process describing the motion of an over-damped particle with position x in a quadratic potential $V(x) = \frac{(\eta x - f)^2}{2\eta}$. Free energy of equilibrium distribution (Einstein relation) is given by

$$e^{-\beta F} = \int e^{-\beta V(x)} dx = \sqrt{\frac{\sigma^2 \pi}{\eta}}.$$

Now let us look at the work. The external force f is a function of t. Changes in external force lead to changes in the potential, and thus changes in the total work performed between the initial and final time points of the system. Accordingly,

$$\Delta W = (\frac{\partial V}{\partial f})_x \Delta f = \frac{-(\eta x - f)}{\eta} \Delta f.$$

Notice the $\Delta F = 0$. Clausius tells that that $\langle W \rangle \geq \Delta F = 0$. Jarzynski, tells us that

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F} = 1.$$

This uses all the information.

Summary

Key Points:

- Riemannian geometry, entropy, and network robustness via OMT
- Quantum mechanics for matrix-valued OMT
- Data interpolation and prediction
- Examined biological and financial networks
- Results and methods are generalizable to other systems