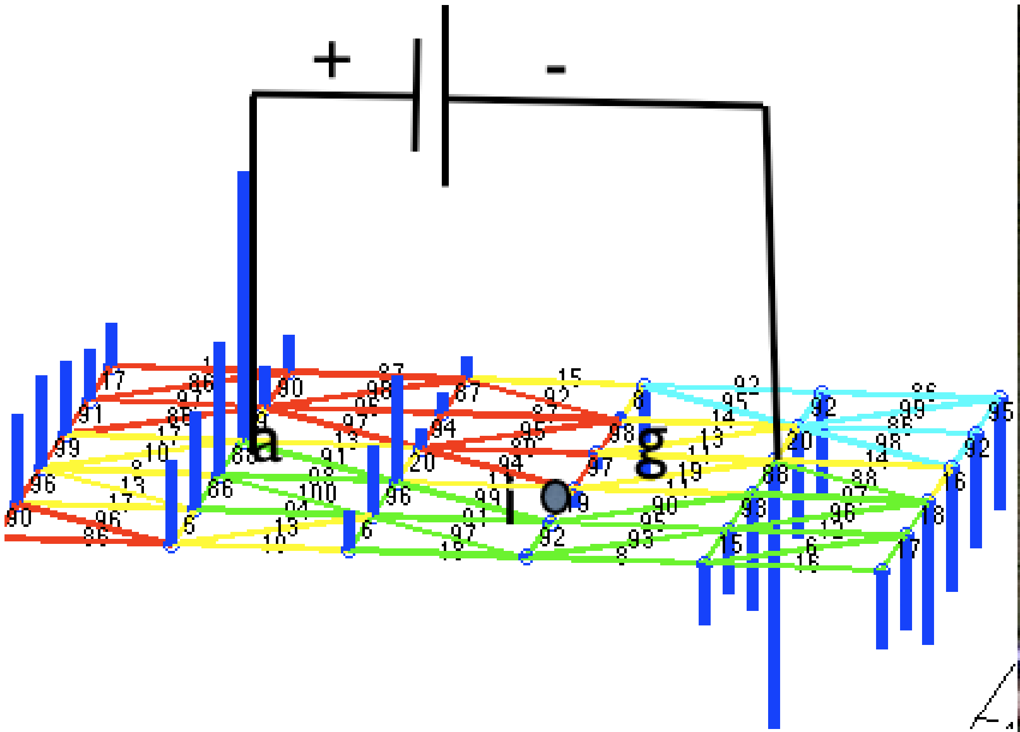
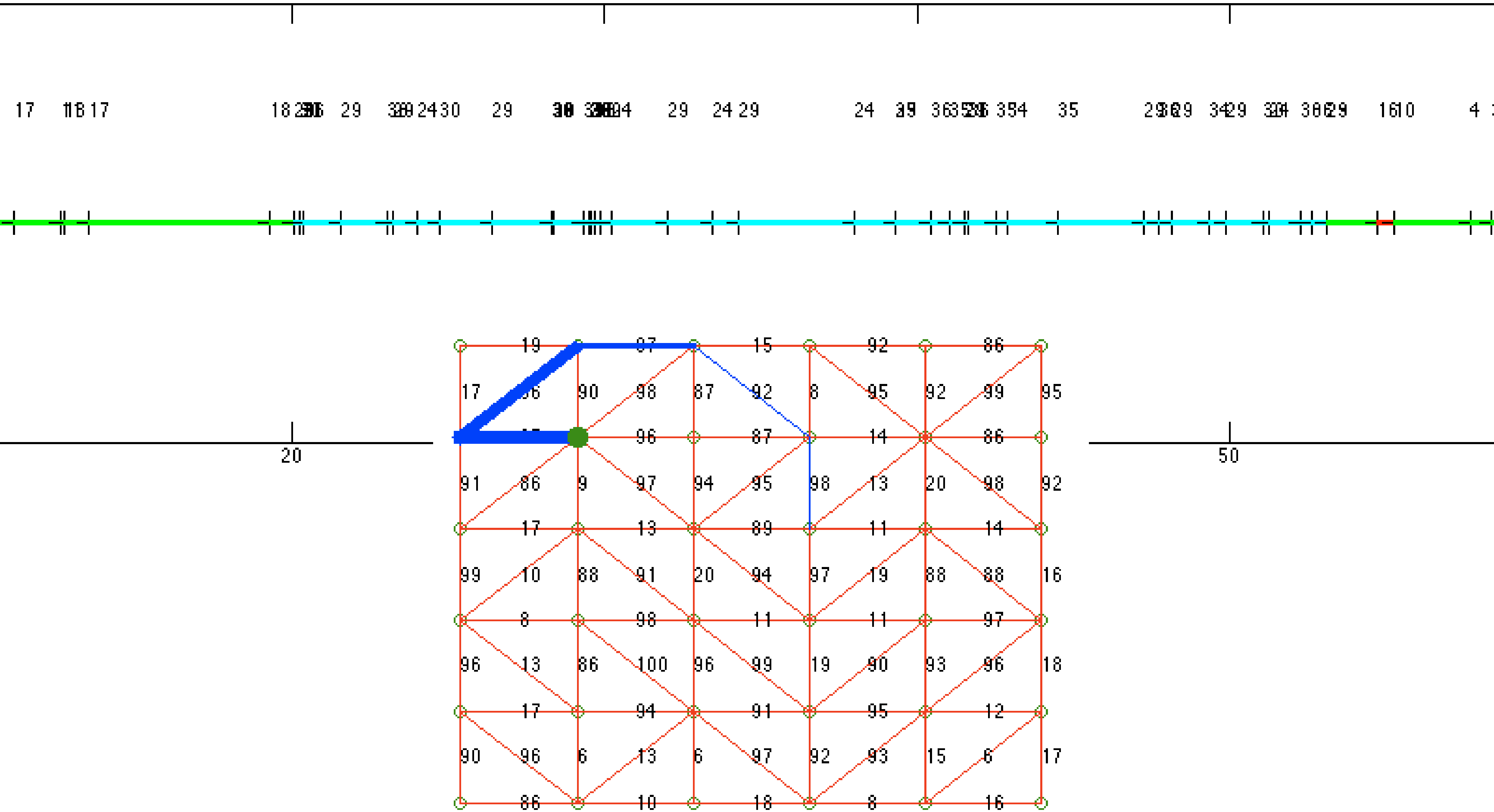
Statistical learning, complex systems, and pickles

Lecture 1, CSSS10



Greg Leibon
Memento, Inc
Dartmouth College

Markov Process: X_t




```
states=Snake(500,5,'No');
```

Formed from

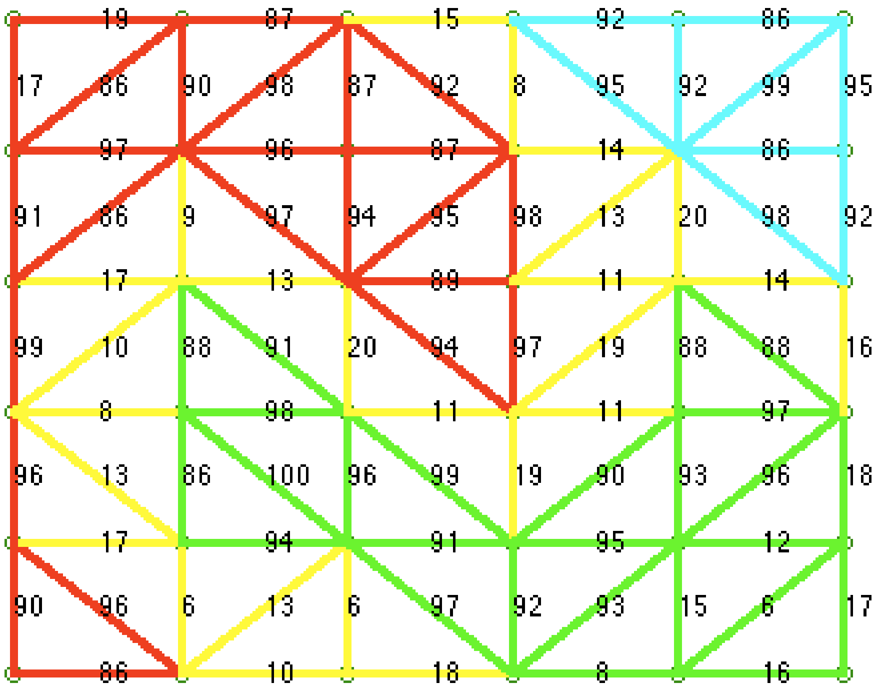
State transitions $P_{ij} = P(S_{n+1} = j \mid S_{n+1} = i)$

Expected waiting time τ_i

Key object

$$P^t = P(X_t = j \mid X_0 = i)$$

Toy Example



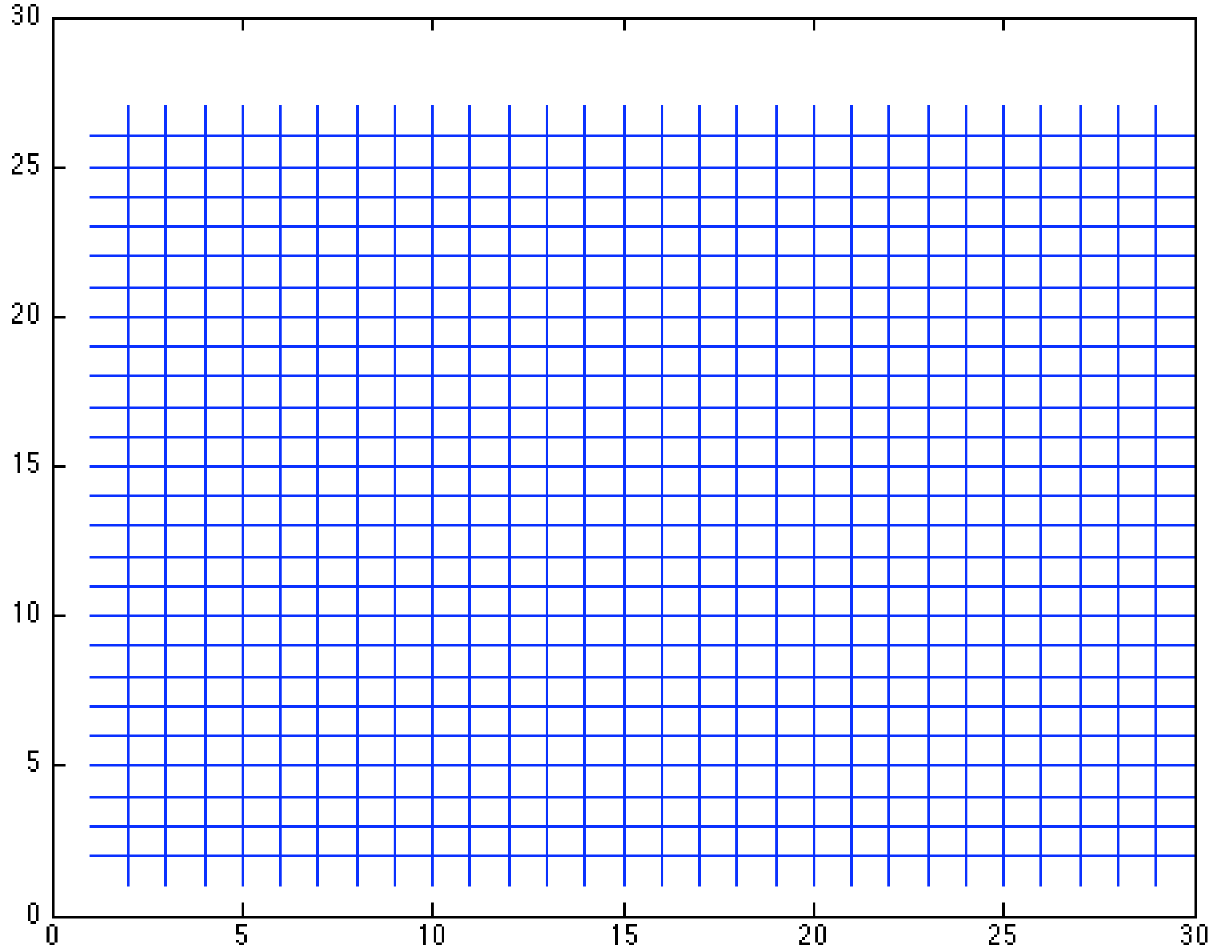
$$P = [W\mathbf{1}]^{-1}W$$

```
N=200;  K=10; [states]=PlotTheState(N,K);
```

W=

	1	2	3	4	5
1	0	86	0	0	
2	86	0	10	0	
3	0	10	0	18	
4	0	0	18	0	
5	0	0	0	8	
6	0	0	0	0	1
7	90	96	0	0	
8	0	6	0	0	
9	0	13	6	97	
10	0	0	0	92	
11	0	0	0	93	1
12	0	0	0	0	
13	0	0	0	0	
14	0	0	0	0	
15	0	0	0	0	
16	0	0	0	0	
17	0	0	0	0	
18	0	0	0	0	
19	0	0	0	0	

Uniform Example



Use adjacency matrix as weight matrix

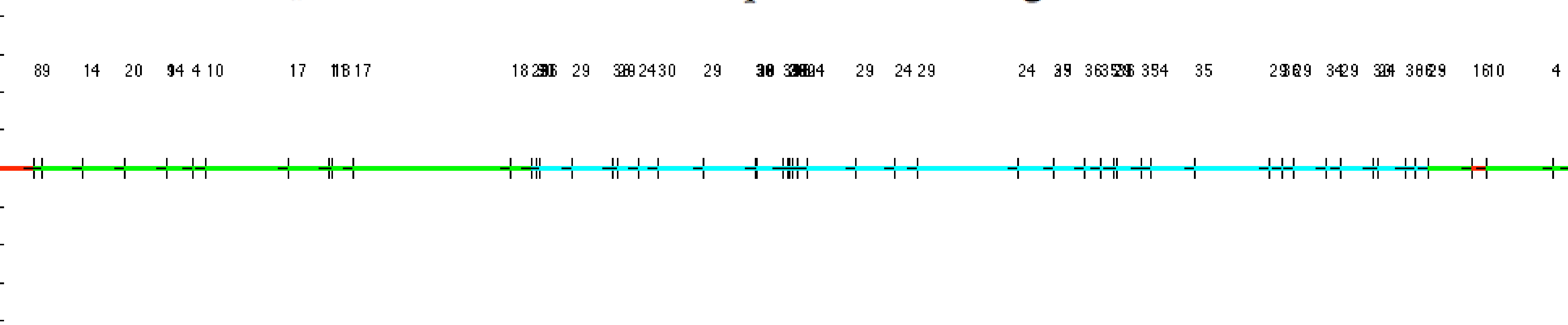
Assume our *chains* satisfy...

1. $p_{ii} = 0$

2. *Ergodic:* It it is possible to get from any state to any other state.

discrete equilibrium measure: Ergodic implies there exists π such that $\pi^{tr} P = \pi^{tr}$

π_a is the fraction of states equal to a in a long state list



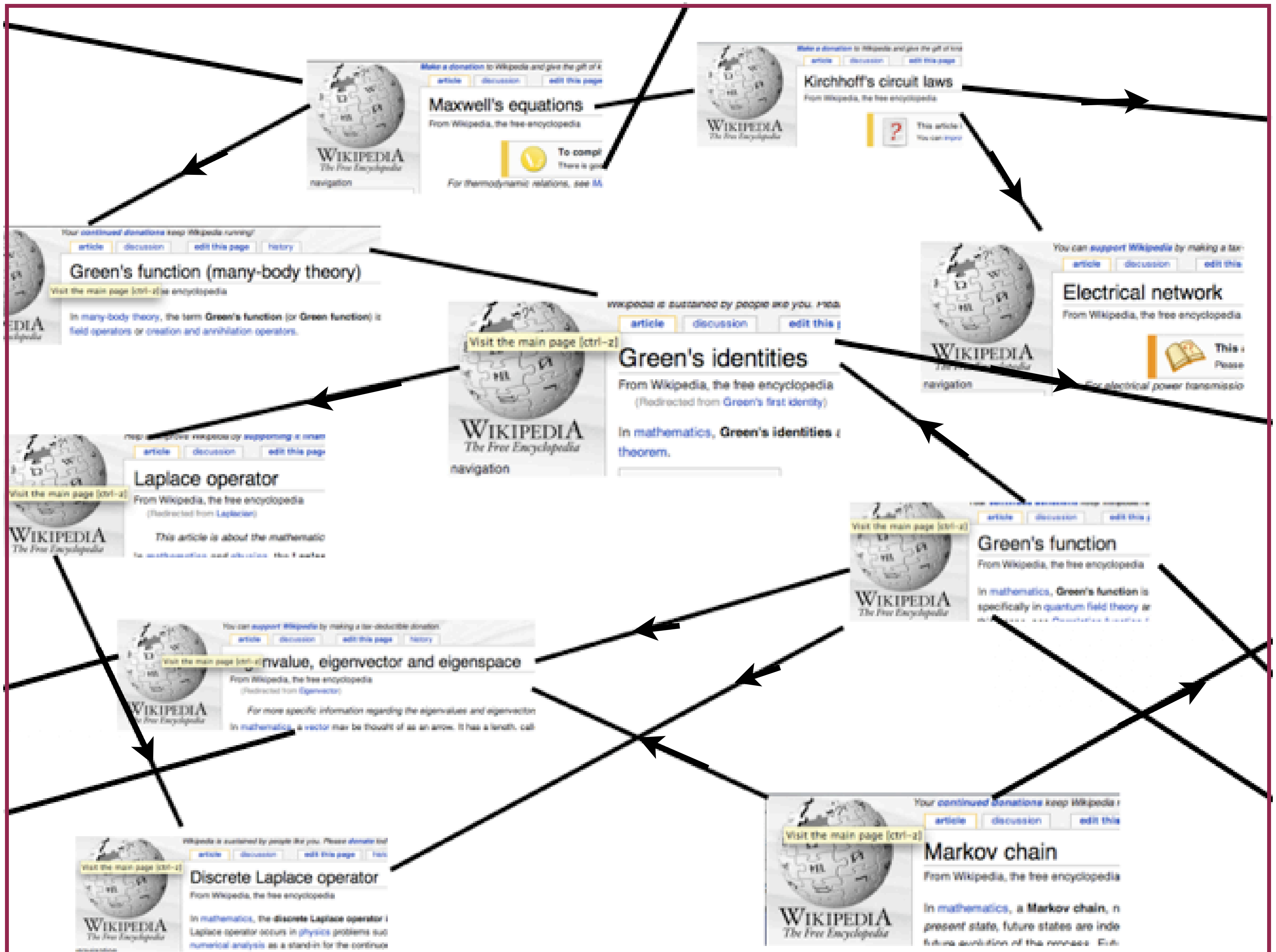
equilibrium measure: notice $\tau_a \pi_a$ is the expected time spent in tae a, hence

$$\omega_a = \frac{\tau_a \pi_a}{\sum \tau_a \pi_a}$$

is the asymptotic fraction of time X_t spends in state a.

3. $\sum \tau_a \pi_a = 1$

Space Of Mathematics



roughly 18,000 mathematics pages on Wikipedia's *list_of_mathematics_articles*

Memoryless

Given $X_{t_0} = i$, let T be the first time one transitions from i.

Memoryless: implies for $s > t_0$, we have $P(T > (t+s) \mid T > s) = P(T > t)$

$$\frac{dP(T>t)}{dt} = \lim_{dt \to 0} \frac{P(T > (t+dt)) - P(T>t)}{dt}$$

$$= \lim_{dt \to 0} \frac{P(T > t+dt \mid T > t)P(T>t) - P(T>t)}{dt}$$

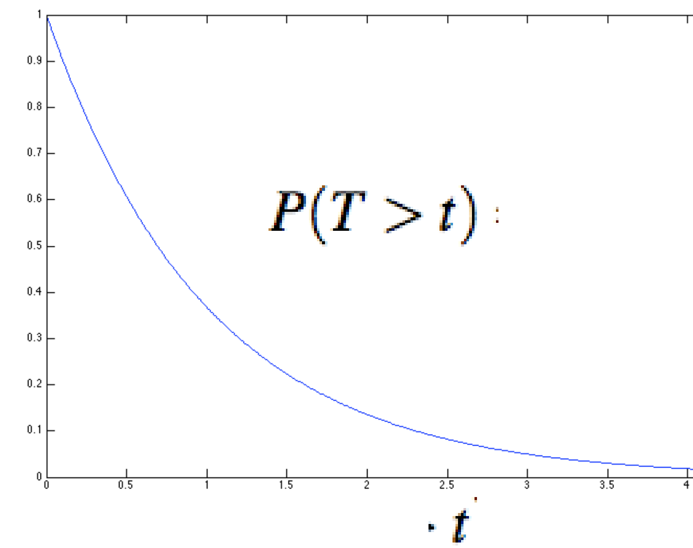
$$= \left(\lim_{dt \to 0} \frac{P(T > dt \mid T > 0) - 1}{dt} \right) P(T > t)$$

$$= CP(T > t)$$

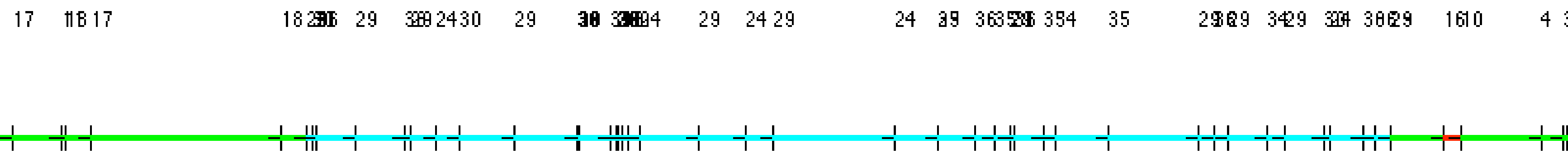
Hence $\quad P(T > t) = e^{ct}$

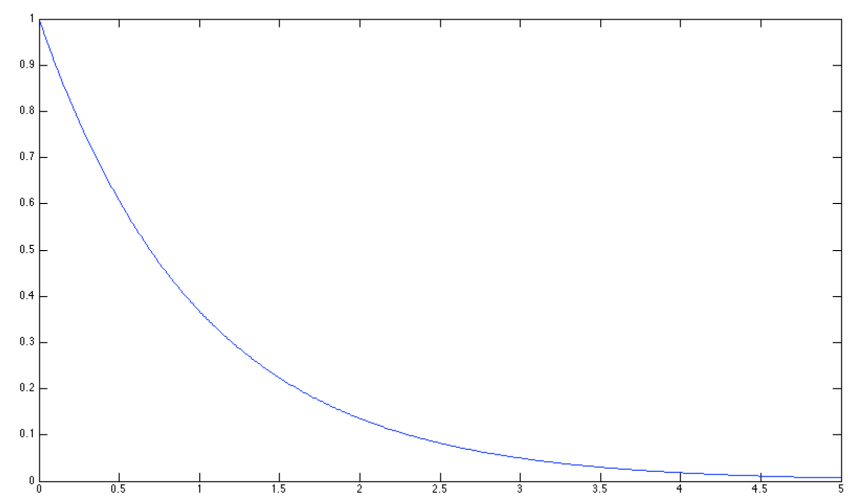
and since $E(T) = \int_0^\infty t e^{ct} dt = \tau_i$

$$P(T > t) = e^{-\frac{1}{\tau_i} t}$$




CAPITAL METRO




1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5

In general...

$$\frac{d}{dt}P_t\bigg|_{t=0} = -\Delta_\omega$$

$$\Delta_\omega = [\tau]^{-1}(I - P)$$

$$P^t = e^{-t\Delta_\omega}$$

$e^z = \sum_{n=1}^{\infty} \frac{z^n}{n!}$ converges when A is plugged in

MECHANIQUE CELESTE

OF

P. S. LAPLACE,

Member of the Institute and of the Bureau of Longitude of France, &c. &c.

When the motions are very small; we may neglect the squares and the products of u, v, and v; the equation (H) then becomes

$$\delta V-\frac{\delta p}{\rho}=\left(\frac{du}{dt}\right).\delta x+\left(\frac{dv}{dt}\right).\delta y+\left(\frac{d\mathrm{v}}{dt}\right).\delta z;$$

therefore in this case $u.\delta x+v.\delta y+\mathrm{v}.\delta z$ is an exact variation, if, as we have supposed, p be a function of ρ; by naming this differential $\delta\varphi$, we shall have

$$V-\int\frac{\delta p}{\rho}=\left(\frac{d\varphi}{dt}\right)^{*};$$

and if the fluid be homogeneous, the equation of continuity will become

$$0=\left(\frac{d^2\varphi}{dx^2}\right)+\left(\frac{d^2\varphi}{dy^2}\right)+\left(\frac{d^2\varphi}{dz^2}\right).$$

These two equations contain the whole of the theory of very small undulations of homogeneous fluids.

Circe 1799

The Laplacian

$$\Delta f = f(x) - av_{S(x)}(f)$$

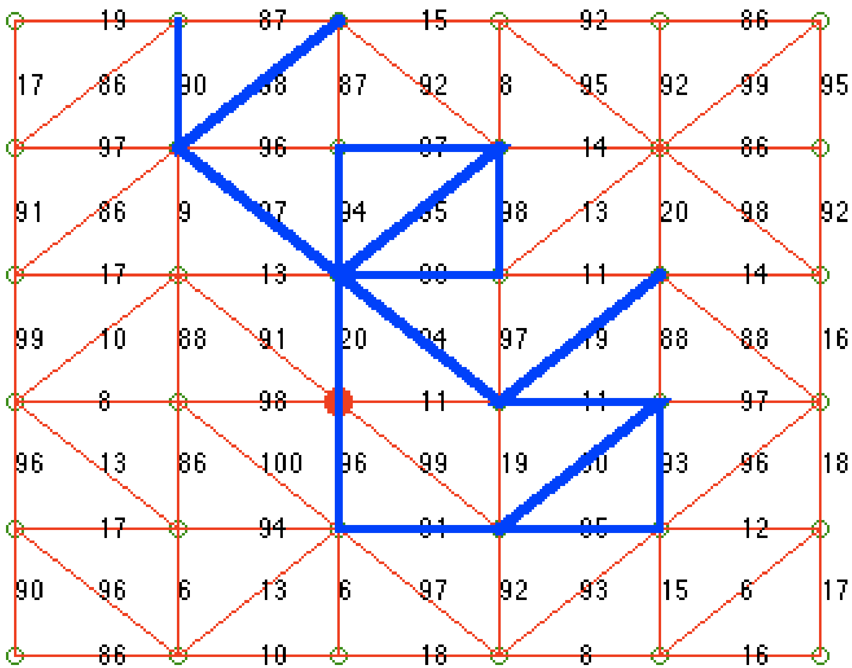
$$\Delta f = \lim_{\varepsilon\to 0}\frac{1}{\varepsilon^2}\left(f(x)-\frac{1}{4\pi\varepsilon^2}\int_{S^2(x,\varepsilon)} f(y)\,dA\right)$$

$$= -\left(\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}+\frac{\partial^2 f}{\partial z^2}\right)$$

$$\Delta f = (I-P)f$$

Conformal Invariance

We define different choices of τ with the same P as *conformal changes*



Let R be the return time random variable

```
Loop(15,1000);
```

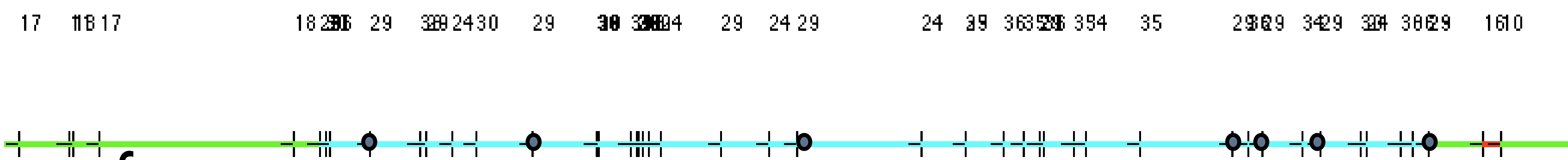
For a discrete chain, the first return time equals $1/\pi_a$

Conformal Invariance of First Return Time: The expected first return time at a is conformally invariant, and hence equal to $1/\pi_a$.

Conformal Invariance

We define different choices of τ with the same P as *conformal changes*

Conformal Invariance of First Return Time: The expected first return time at a is conformally invariant, and hence equal to $1/\pi_a$.



proof:

Suppose we visit roughly M states in time T.

Since $\sum \tau_a \pi_a = 1$

after a conformal change we still visit roughly M dates in time T.

Notice, the state sequence is the same,
so the number of loops N after visiting M sates is the same.

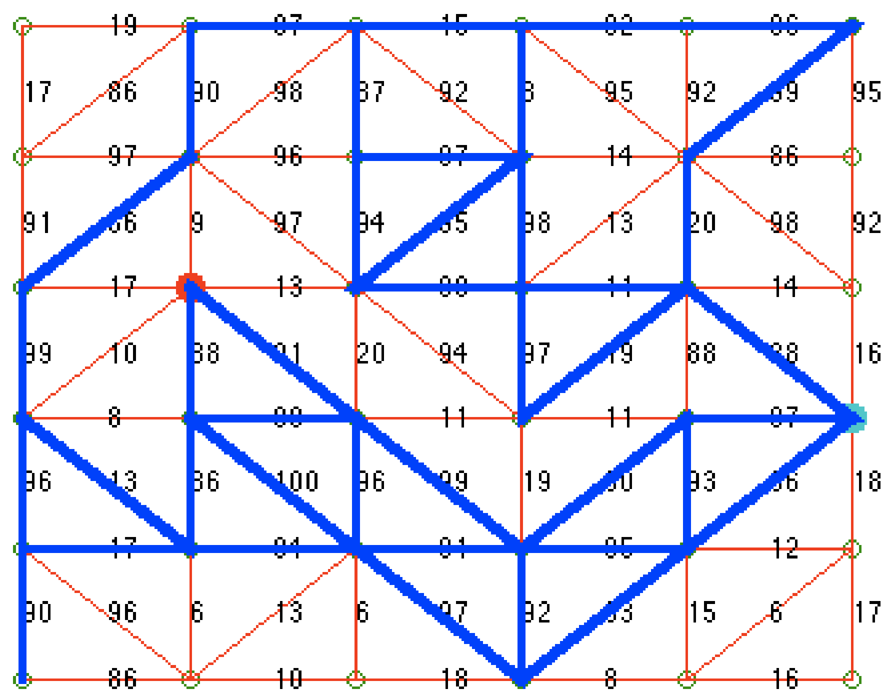
Since the loop (basically) cover [0,T] once
E(R)=T/N is conformally invariant.

QED

Expected Commute Time

Conformal Invariance Of the Expected Commute Time: The expected commute time depends only the underlying discrete structure.

proof: same argument



116

```
[states Time]=Commute(15,25,1000);
```

MATHEMATICAL PAPERS

OF THE LATE

GEORGE GREEN,

FELLOW OF GONVILLE AND CAIUS COLLEGE, CAMBRIDGE.

AN ESSAY

ON THE APPLICATION OF MATHEMATICAL ANALYSIS

TO THE THEORIES

OF ELECTRICITY AND MAGNETISM.*

1828

To compute $E(T_{ab})$

Potential from the charge distribution: Kirchhoff Operator

$$K = [\omega]\Delta_\omega$$

$$KV = \rho$$

Back to Green's Paper

(1.) THE function which represents the sum of all the electric particles acting on a given point divided by their respective distances from this point, has the property of giving, in a very simple form, the forces by which it is solicited, arising from the whole electrified mass.—We shall, in what follows, endeavour to discover some relations between this function, and the density of the electricity in the mass or masses producing it, and apply the relations thus obtained, to the theory of electricity.

Firstly, let us consider a body of any form whatever, through which the electricity is distributed according to any given law, and fixed there, and let x', y', z', be the rectangular co-ordinates of a particle of this body, ρ' the density of the electricity in this particle, so that $dx'dy'dz'$ being the volume of the particle, $\rho'dx'dy'dz'$ shall be the quantity of electricity it contains: moreover, let r' be the distance between this particle and a point p exterior to the body, and V represent the sum of all the particles of electricity divided by their respective distances from this point, whose co-ordinates are supposed to be x, y, z, then shall we have

$$r' = \sqrt{(x'-x)^2 + (y'-y)^2 + (z'-z)^2},$$

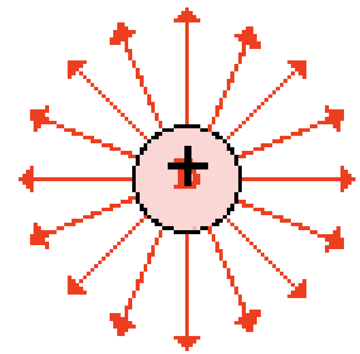
and

$$V = \int \frac{\rho' dx' dy' dz'}{r'};$$

the integral comprehending every particle in the electrified mass under consideration.

2—2

$\mathbb{R}^2$ $\qquad V_a(z) = \frac{1}{2\pi} \log |z-a|$



.... with a second source at infinity

$\mathbb{R}^3$

$$G(x,y) = \frac{1}{4\pi |x-y|}$$

LAPLACE has shown, in his Méc. Céleste, that the function V has the property of satisfying the equation

$$0 = \frac{d^2V}{dx^2} + \frac{d^2V}{dy^2} + \frac{d^2V}{dz^2},$$

and as this equation will be incessantly recurring in what follows, we shall write it in the abridged form $0 = \delta V$; the symbol δ being used in no other sense throughout the whole of this Essay.

In order to prove that $0 = \delta V$, we have only to remark, that by differentiation we immediately obtain $0 = \delta \frac{1}{r'}$, and consequently each element of V substituted for V in the above equation satisfies it; hence the whole integral (being considered as the sum of all these elements) will also satisfy it. This reasoning ceases to hold good when the point p is within the body, for then, the coefficients of some of the elements which enter into V becoming infinite, it does not therefore necessarily follow that V satisfies the equation

$$0 = \delta V,$$

although each of its elements, considered separately, may do so.

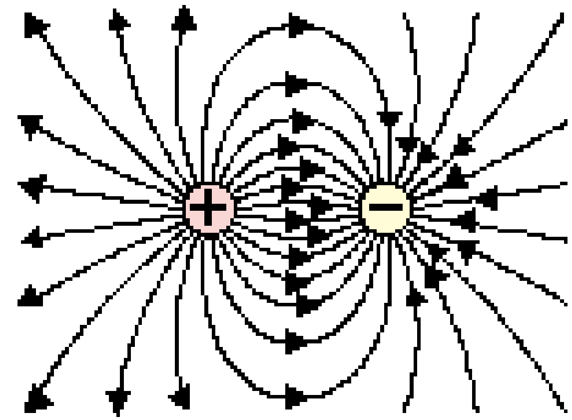
...

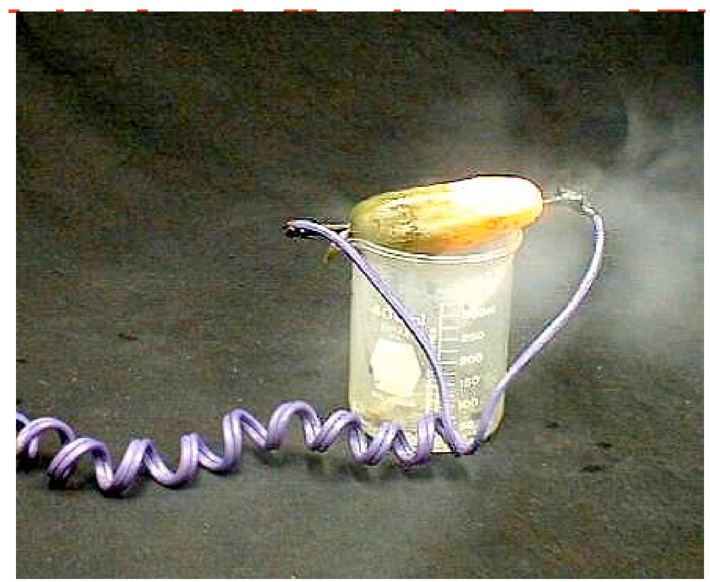
Hence, throughout the interior of the mass

$$0 = \delta V + 4\pi\rho;$$

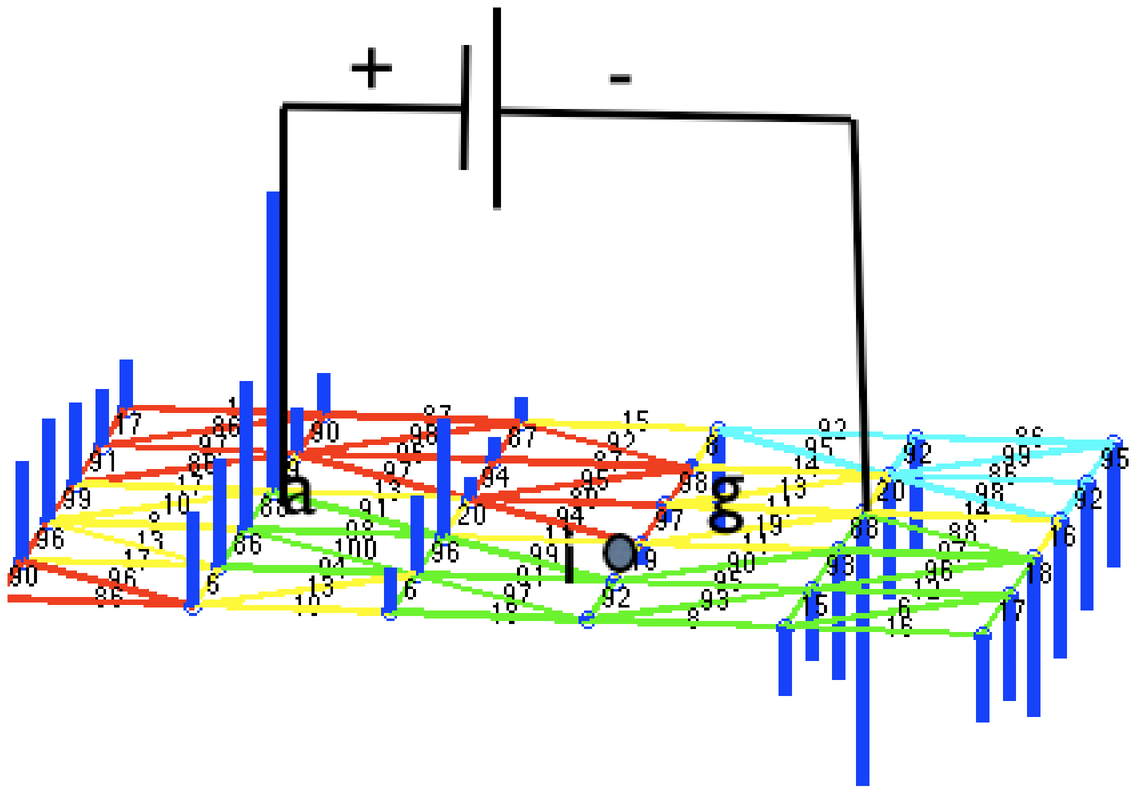
of which, the equation $0 = \delta V$ for any point exterior to the body is a particular case, seeing that, here $\rho = 0$.

$$V^d_{c,S^2}(z) = \frac{1}{2\pi}\log\left|\frac{z-c}{z-d}\right|$$





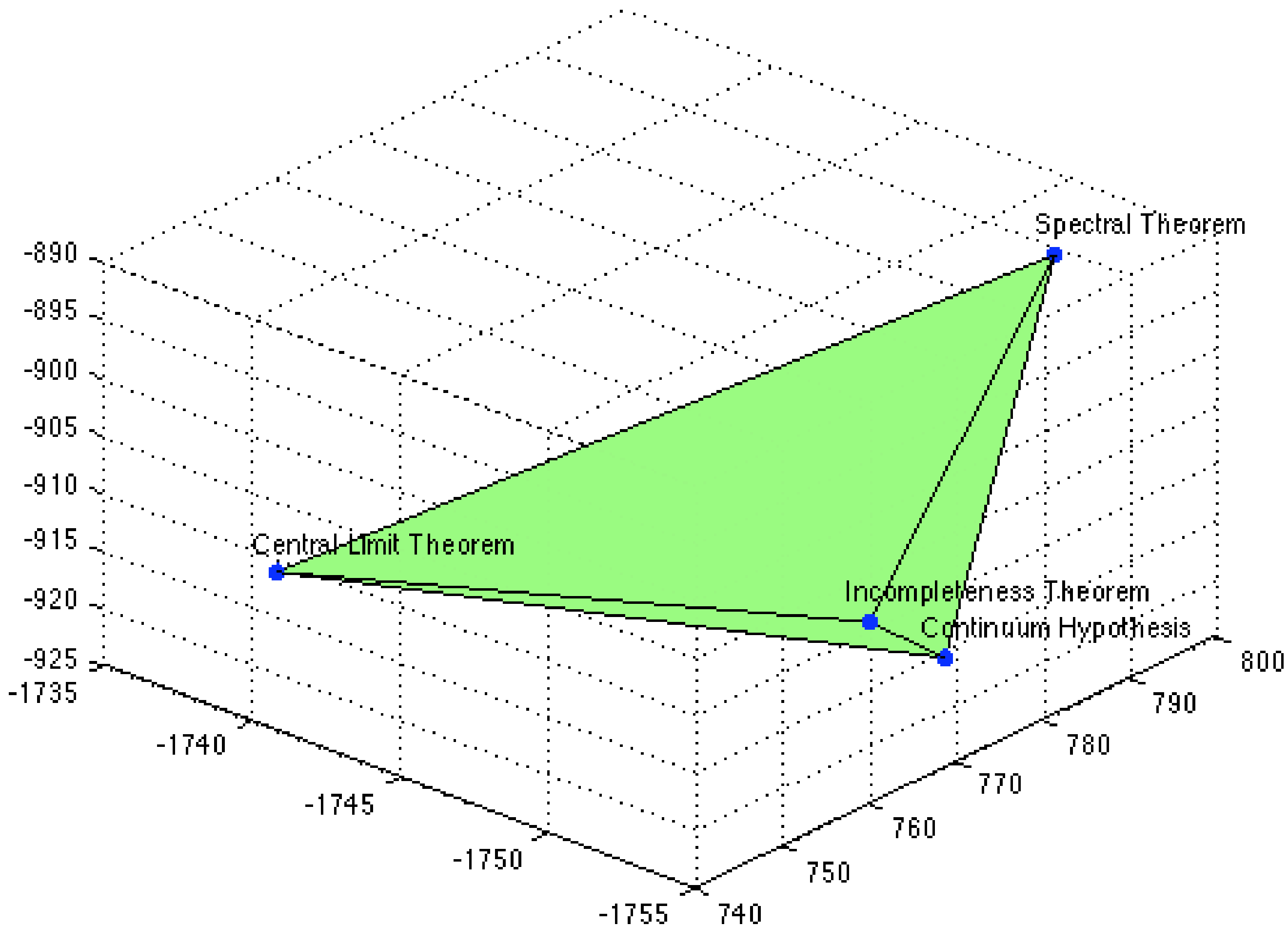
$$KV_a^b = \delta_a - \delta_b$$



Commute Theorem:

$$E(T_{ab}) = V_a^b(b) - V_a^b(a)$$

... the effective resistance



Space Of Mathematics Example: We find that the commute time between the *Fundamental Theorem of Calculus* and *Stokes Theorem* is roughly 5287., while the commuute time between the *Fundamental Theorem of Calculus* and the *Gauss Bonnet Theorem* is roughly 34809. In figure 1.3 we see the embedding of some of our favorite theorems using the distance $\sqrt{E(T_{ab})}$. The use of this particular distance is explained in section 1.3.

WHY? $d(a,b) = \sqrt{E(T_{ab})}$ WHY?

Green's Identity: $< f, g >_\omega = f^{tr}[\omega]g$

$$< f, \Delta_\omega f >_\omega = \frac{1}{2}\sum_{i,j} \pi_i p_{ij}(f_i - f_j)^2$$

$$
\begin{aligned}
< f, \Delta_\omega f >_\omega &= \textstyle\sum_{i,j} f_i \omega_i \frac{1}{\tau_i}(I_{ij} - P_{ij}) f_j && \text{(def.)}\\
&= \textstyle\sum_{i,j} f_i \pi_i (I_{ij} - P_{ij}) f_j && (\omega_i = \tau_i \pi_i)\\
&= \tfrac{1}{2}\left(\textstyle\sum_i \pi_i f_i^2 - 2\sum_{i,j} f_i \pi_i P_{ij} f_j + \sum_j \pi_j f_j^2\right) && \text{(foil)}\\
&= \tfrac{1}{2}\left(\textstyle\sum_{i,j} \pi_i P_{ij} f_i^2 - 2\sum_{i,j} f_i \pi_i P_{ij} f_j + \sum_j \pi_j f_j^2\right) && \text{(rowsum P is 1)}\\
&= \tfrac{1}{2}\left(\textstyle\sum_{i,j} \pi_i P_{ij} f_i^2 - 2\sum_{i,j} f_i \pi_i P_{ij} f_j + \sum_{i,j} \pi_j P_{i,j} f_j^2\right) && (\pi \text{ is } P \text{ equi. vec.})\\
&= \tfrac{1}{2}\textstyle\sum_{i,j} \pi_i P_{ij}(f_i^2 - 2 f_i f_j + f_j^2) && \text{(factor)}\\
&= \tfrac{1}{2}\textstyle\sum_{i,j} \pi_i P_{ij}(f_i - f_j)^2 && \text{(unfoil)}
\end{aligned}
$$

"Do not imagine that mathematics is hard and crabbed, and repulsive to common sense. It is merely the etherialisation of common sense."

Dirichlt form $< f,g >_{Dir} = < f, \Delta_\omega g >_\omega$

Make Euclidean $< f,g >_{SymDir} = \frac{1}{2}(< f,g >_{Dir} + < g,f >_{Dir})$

Green's Embedding: $Green(a) = V_g^a$

$$||Green(a) - Green(b)||_{SymDir}^2 = E(T_{ab})$$

$$||Green(a) - Green(b)||_{SymDir}^2 = ||V_g^a - V_g^b||_{SymDir}^2 \qquad \text{(def.)}$$

$$= ||V_g^a - V_g^b||_{Dir}^2 \qquad \text{(unsym.)}$$

$$= ||V_a^b||_{Dir}^2 \qquad (\ \Delta \text{ linearity, see 1.3ex.3})$$

$$= < V_a^b, \Delta V_a^b >_\omega \qquad (\text{def. } Dir)$$

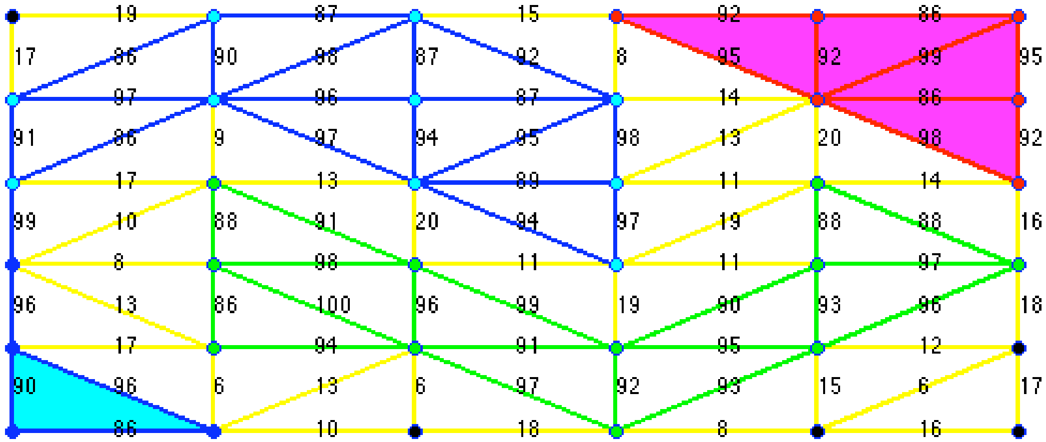
$$= (V_a^b)^{tr} K V_a^b \qquad (\text{def. } K\ )$$

$$= (V_a^b)^{tr}(\delta_a - \delta_b) \qquad (\text{def. } V_a^b)$$

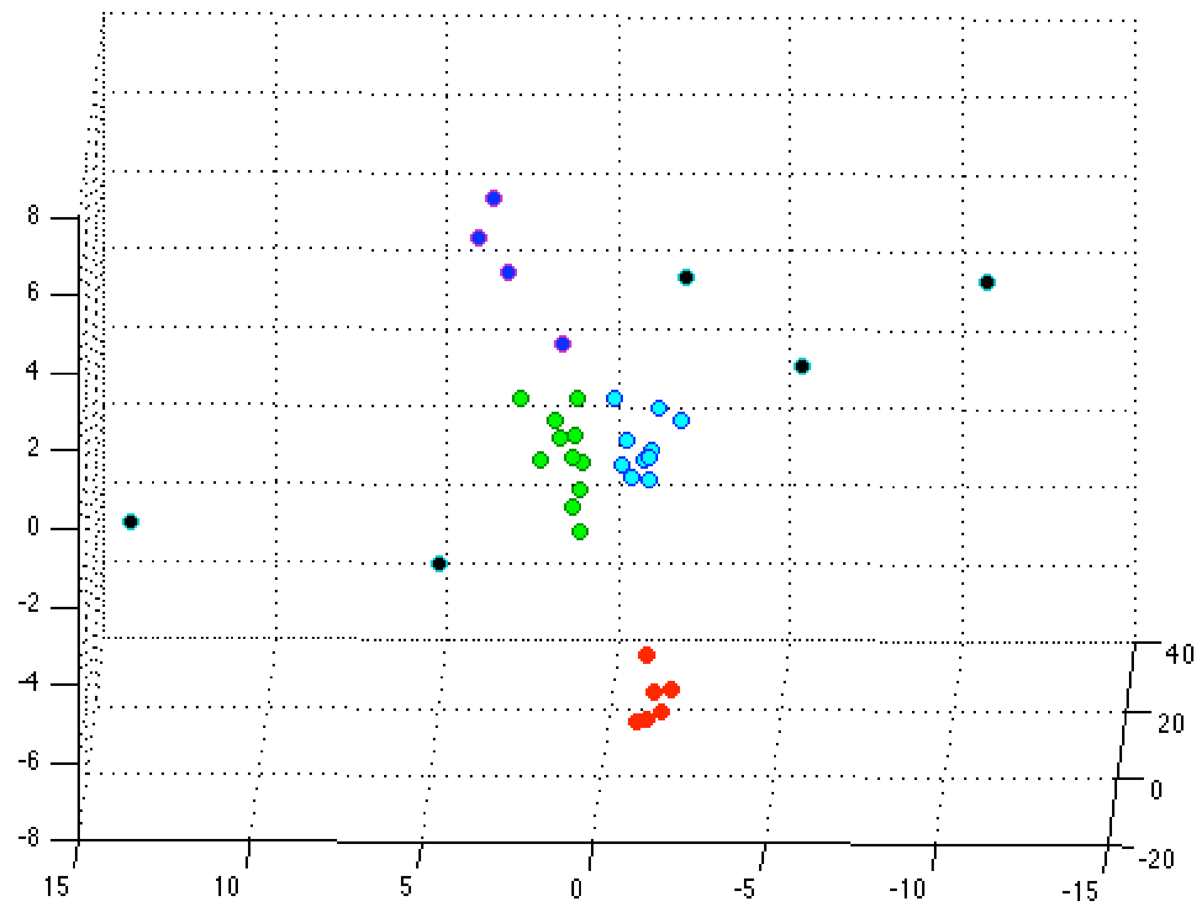
$$= V_a^b(a) - V_a^b(b) \qquad \text{(evaluate)}$$

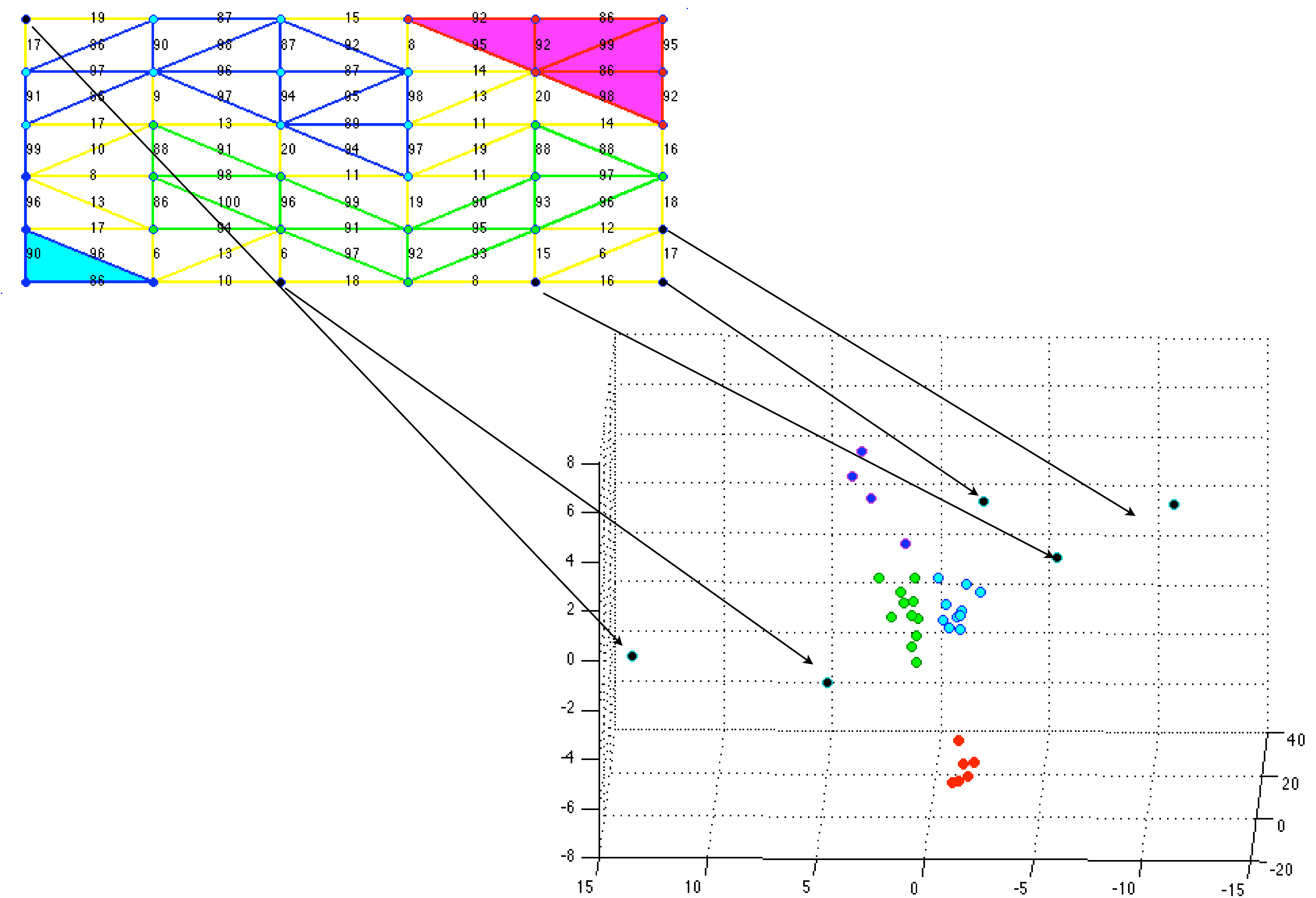
$$= E(T_{ab}) \qquad \text{(Commute Theorem)}$$

q.e.d



Green's Embedding (and MDS)





8
6
4
2
0
-2
-4
-6
-8
15
10
5
0
-5
-10
-15
40
20
0
-20

MDS algorithm

Simply, attempt to minimize a positive loss function that would be zero if for all (x_i, x_j)

$$d(x_i, x_j) = d_{Euc}(f(x_i), f(x_j))$$

Example Raw Stress

$$stess = \sum_{i>j} (d_{Euc}(f(x_i), f(x_j)) - d(x_i, x_j))^2$$

Minimization Techniques: Gradient Decent, Newton Raphson, Iterative Majorization, Tabu Search, Genetic Algorithms, Simulated Annealing....

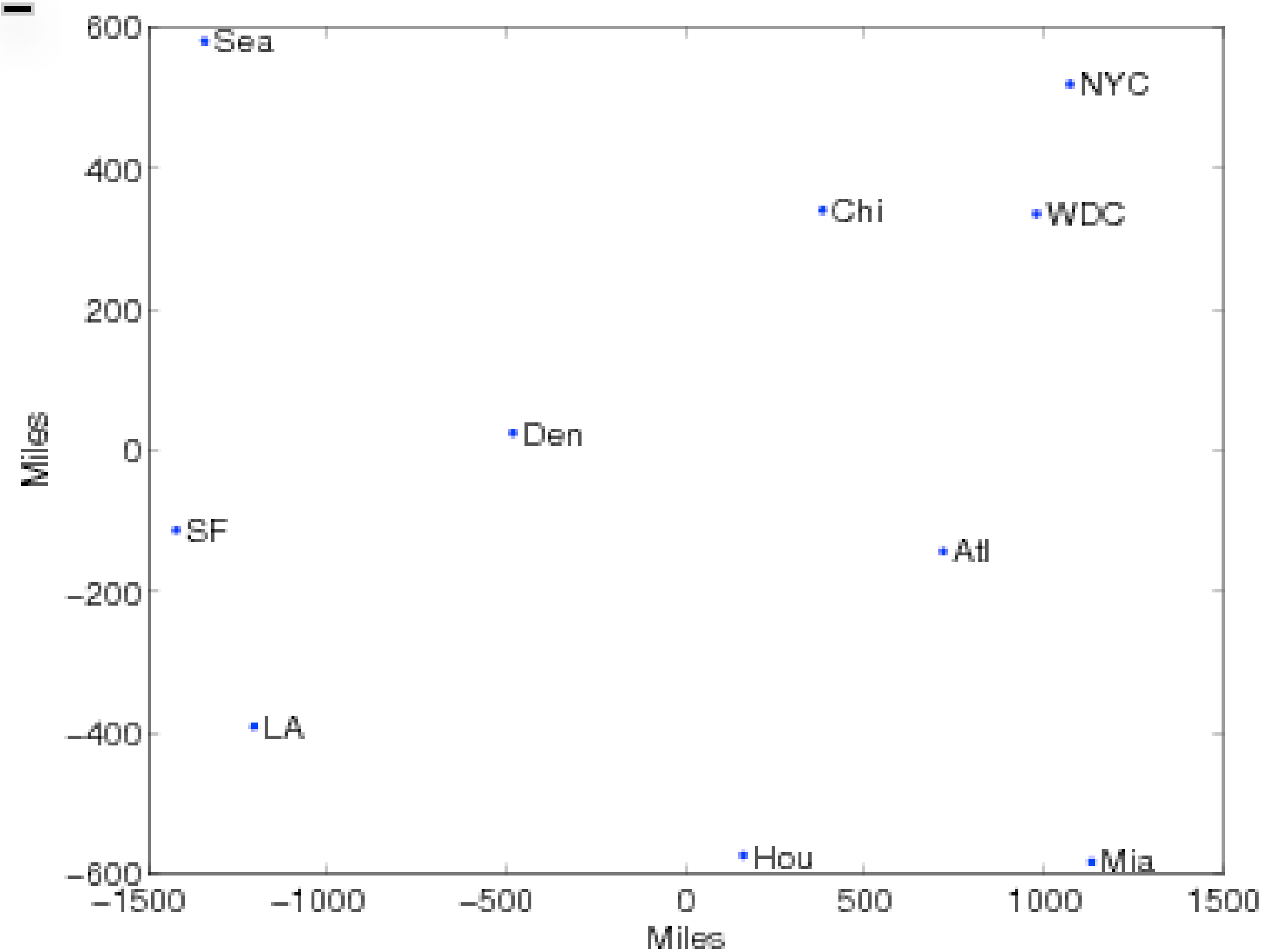
Himalayas

2-d Example

```
cities =
{'Atl','Chi','Den','Hou','LA','Mia','NYC','SF','Sea','WDC'};
D = [    0  587 1212  701 1936  604  748 2139 2182   543;
       587    0  920  940 1745 1188  713 1858 1737   597;
      1212  920    0  879  831 1726 1631  949 1021  1494;
       701  940  879    0 1374  968 1420 1645 1891  1220;
      1936 1745  831 1374    0 2339 2451  347  959  2300;
       604 1188 1726  968 2339    0 1092 2594 2734   923;
       748  713 1631 1420 2451 1092    0 2571 2408   205;
      2139 1858  949 1645  347 2594 2571    0  678  2442;
      2182 1737 1021 1891  959 2734 2408  678    0  2329;
       543  597 1494 1220 2300  923  205 2442 2329     0];
```

Example taken from MatLab's help

MDS 2-d

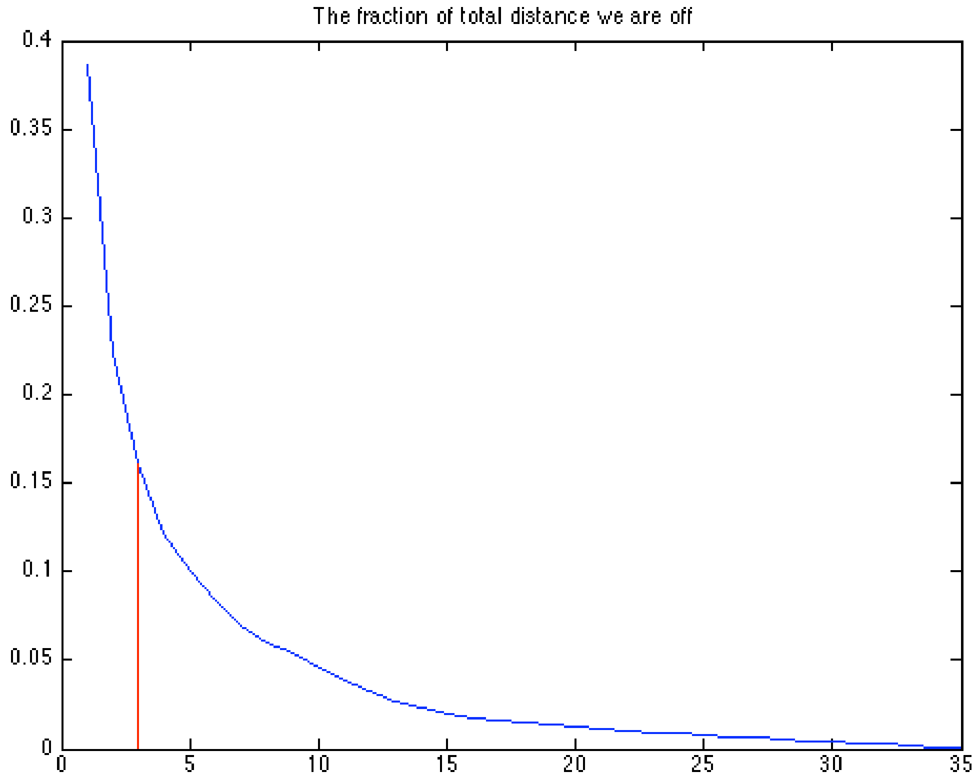


600
400
200
0
−200
−400
−600
−1500
−1000
−500
0
500
1000
1500
Miles
Miles
Sea
NYC
Chi
WDC
Den
SF
Atl
LA
Hou
Mia

Optimal Dimension

Call the embedding f and use the Euclidean distance d(f(X),f(Y))

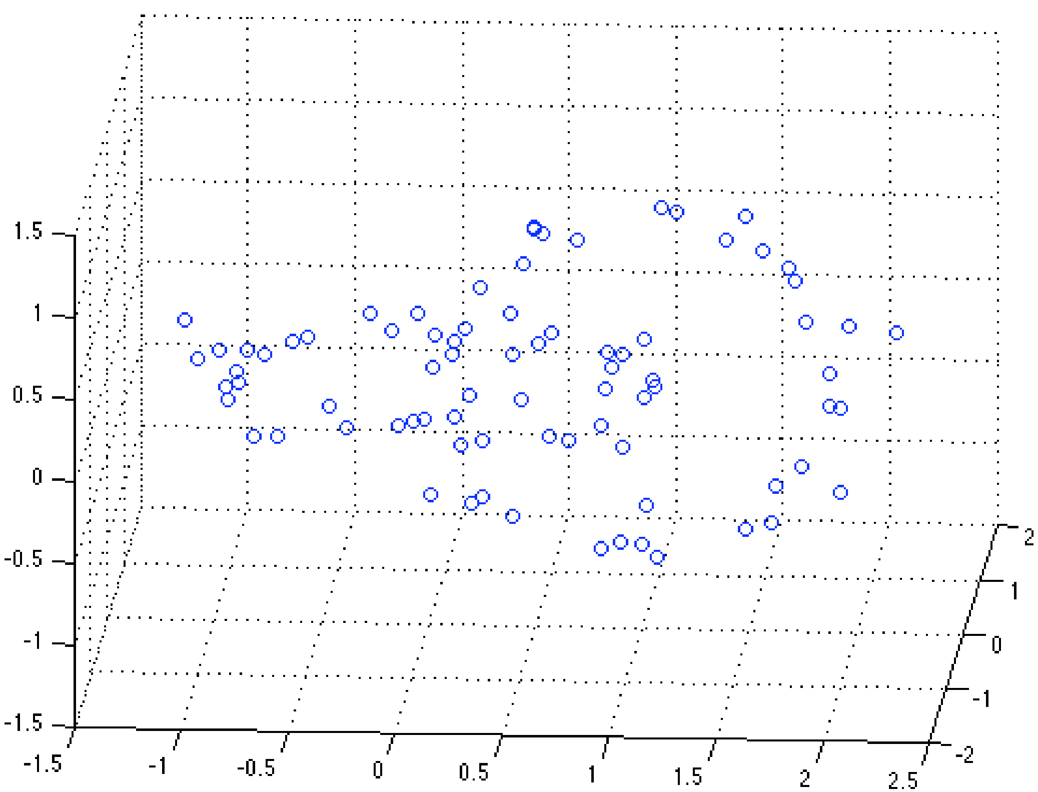
$$\frac{\sum \left| d(X,Y) - d(f(X), f(Y)) \right|}{\sum \left| d(X,Y) \right|}$$

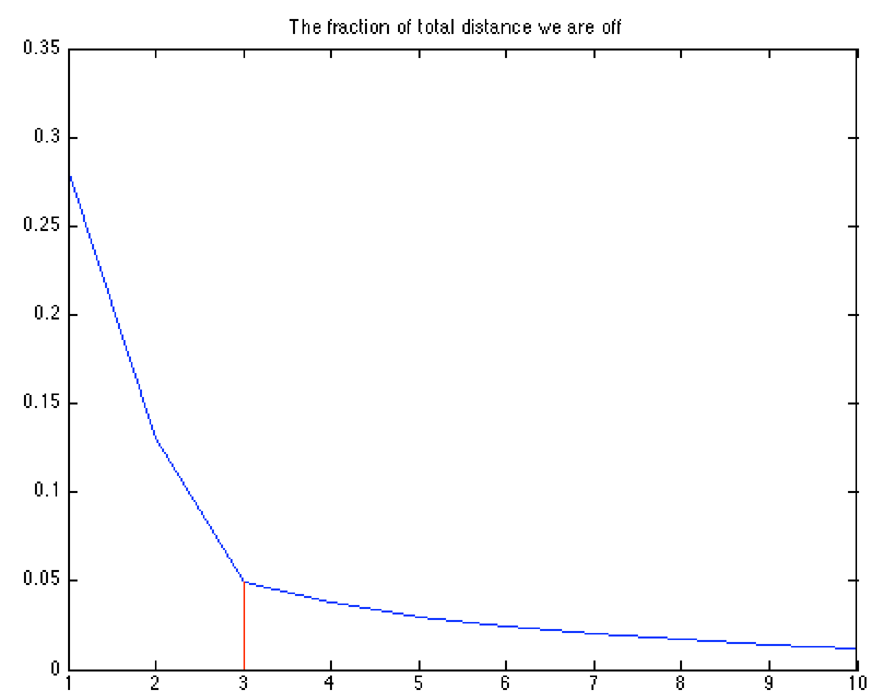


Optimal Dimension

Call the embedding f and use the Euclidean distance d(f(X),f(Y))

$$\frac{\sum \left| d(X,Y) - d(f(X), f(Y)) \right|}{\sum \left| d(X,Y) \right|}$$





Weighted MDS algorithm

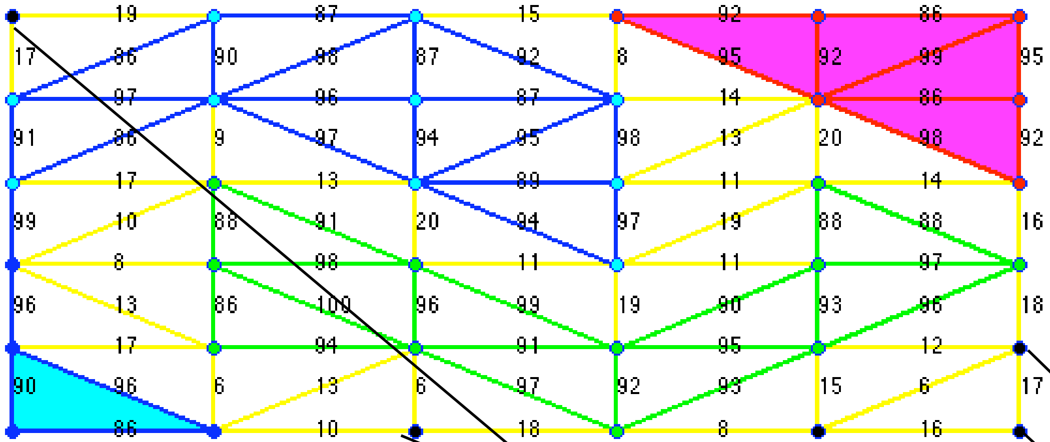
Simply, attempt to minimize the

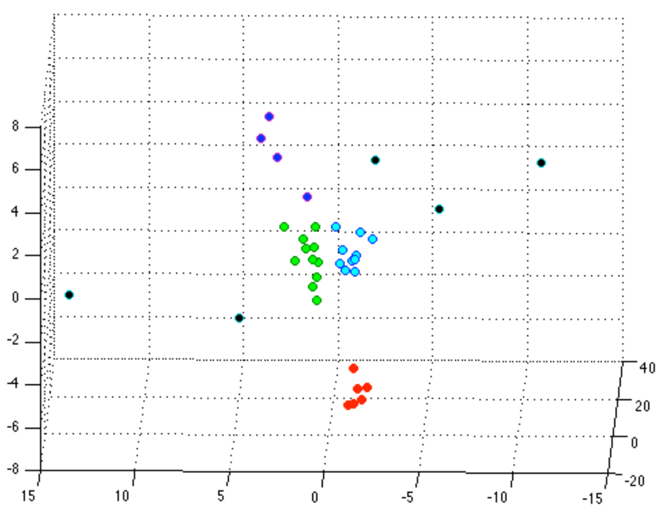
Weighted Raw Stress

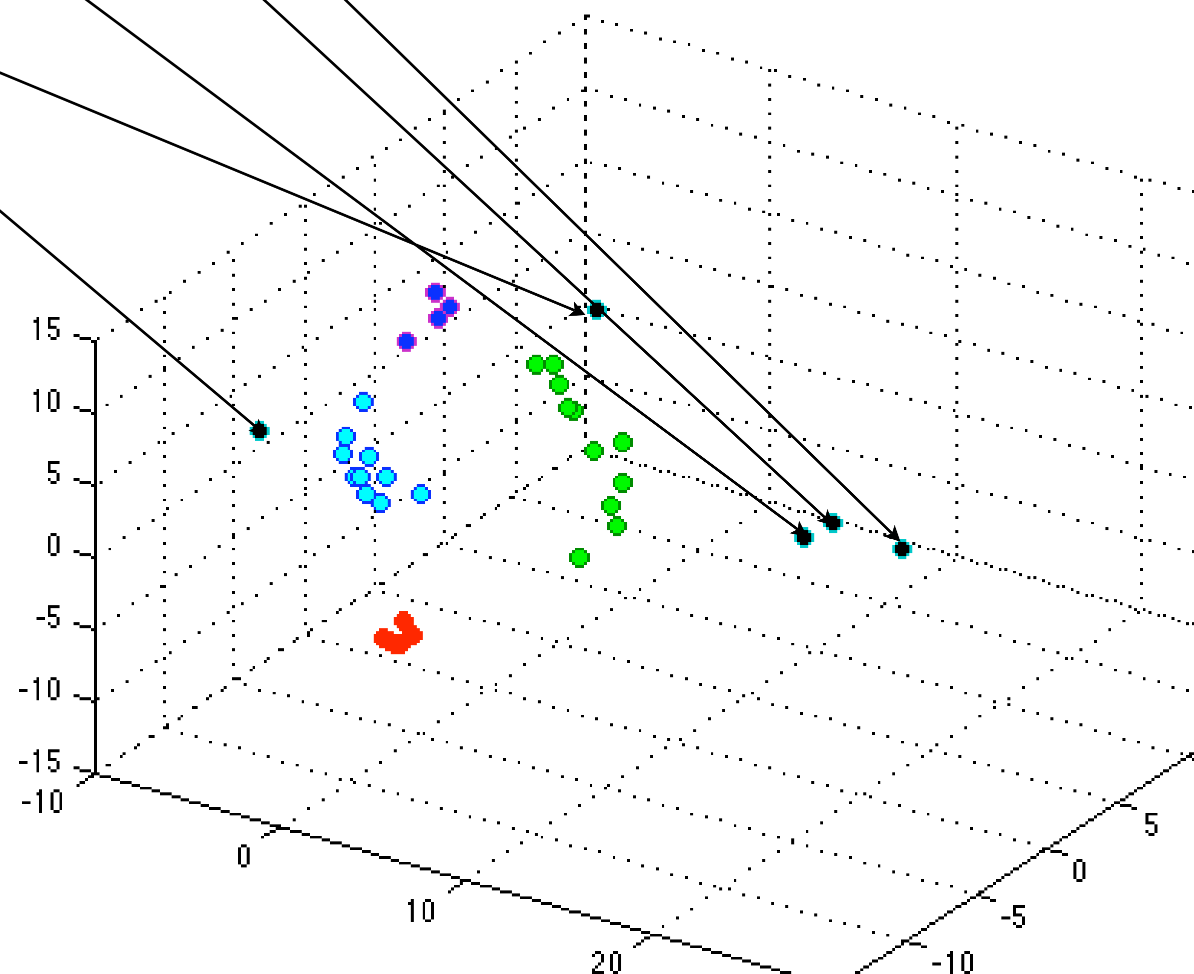
$$stessWeighted = \sum_{i>j} W_{i,j}\left(d_{Euc}(f(x_i), f(x_j)) - d(x_i, x_j)\right)^2$$

Weighted MDS

$$W_{ij} = g(\omega_i)g(\omega_j)$$







MatLab

```
load W.mat;  % The conductance of our Toy Chain
P = diag(1./sum(W,2))*W; % Discrete transition probabilities
n = length(W); % Number of states
I = speye(n); % Identity matrix
A = I-P+sparse(n,n,1); % Make I-P nonsingular.
w0 = I(n,:)/A; % Find a row vector killed by I-P.
w = w0/sum(w0); % Equilibrium measure
M = hit(P);  % Compute hitting times
CommutTimes = M+M';  % Compute commute times
DistEuc = sqrt(Com); % Compute Green's distances
MDS3 = mdscale(DistEuc,3,'criterion','metricsstress');
   % Compute 3-d MDS
Weights = (sqrt(w))'*(sqrt(w)); % Weights fro MDS
MDS3weighted = mdscale(DistEuc3,'Weights',Weights,'criterion','metricsstress')
   % Compute 3-d MDS
 figure(1) % Graph these 3-d embedding
subplot(2,1,1)
scatter3(MDS3(:,1),MDS3(:,2),MDS3(:,3))
title('Un-weighted MDS of Green distances')
subplot(2,1,2)
scatter3(MDS3weighted(:,1),MDS3weighted(:,2),MDS3weighted(:,3))
title('Weighted MDS of Green distances')
```

Appendix: Proof of the

Commute Theorem:

$$E(T_{ab}) = V_a^b(b) - V_a^b(a)$$

Why we are doing it.... to do some mathematics (yeah!)....

- Visit the almighty *maximum principle*
- Dwell on the *fundamental mysteries of probability theory*
- Directly interact with the single most important fact about infinity

Harmonic

A harmonic function is a function f which satisfies $\Delta f = 0$ off a specified set call the *boundary*

a specified set call the *boundary*.

An example

$$q(x) = \frac{V_a^b(x) - V_a^b(a)}{V_a^b(b) - V_a^b(a)}$$

q is clearkly harmonic with boundary $\partial = \{a, b\}$

Recall:

$$\Delta_\omega = [\tau]^{-1}(I - P)$$

$$K = [\omega]\Delta_\omega.$$

$$KV_a^b = \delta_a - \delta_b$$

Another example...

$p(x)$: the probability of starting at of reaching b before hitting a

away from ∂, that if I take one step then $p(x)$ is equal to the average of its neighbors.

Hence both p and q are harmonic with $\partial = \{a, b\}$ and furthermore $p(a) = q(a) = 0$ and $p(b) = q(b) = 1$

... or is it?

Maximum principle

We look at the function $f = p - q$. By the linearity of Δ this function is harmonic and zero on the boundary. So $p - q = 0$ follows from the following principle:

Maximal Principle: A harmonic function is constant or takes on its maximum on the boundary.

Let's evaluate the Laplacian!

Notice $(\Delta_\pi p)(a) = p_e$ is the probability that we escape from a, in other words leaving a we reach b before returning to a.

Since $KV_a^b = \delta_a - \delta_b$ we have $(\Delta_\pi q(a)) = \frac{1}{\pi_a(V_a^b(b) - V_a^b(a))}$

So we now see that

$$V_a^b(b) - V_a^b(a) = \frac{1}{\pi_a p_e}$$

Lemma: $E(T_{ab}) = \frac{1}{\pi_a p_e}$

Notice that starting at a I can view my vouage back to a via b as a sequence of M returns to a, each involving $R_{a,i}$ steps, together with one last one where I go via a. Hence $T_{a,b} = \sum_{i=1}^{M} R_{a,i}$ and

$$
\begin{aligned}
E(T_{a,b}) &= E(\textstyle\sum_{i=1}^{M} R_{a,i}) && \text{(just observed)}\\
&= E(E(\textstyle\sum_{i=1}^{M} R_i \mid M = m)) && \text{(basic expectation property)}\\
&= \textstyle\sum_{i=1}^{m} E(\sum_{i=1}^{m} R_i) P(M = m) && \text{(def. Expected val.)}\\
&= \textstyle\sum_{i=1}^{m} E(\sum_{i=1}^{m} R_i) p_e (1-p_e)^{m-1} && \text{(escaped m-1 times)}\\
&= \textstyle\sum_{i=1}^{m} m E(R_i) p_e (1-p_e)^{m-1} && \text{(linearity exp. val.)}\\
&= p_e E(R_i) \textstyle\sum_{i=0}^{m} m (1-p_e)^{m-1} && \text{(rearrangement)}\\
&= \frac{p_e}{\pi_a} \textstyle\sum_{i=0}^{m} m (1-p_e)^{m-1} && \text{(expected return time, lem. 8.?)}\\
&= \frac{p_e}{\pi_a} \frac{1}{p_e^2} && \text{(geometric series, Ex. 8.??)}\\
&= \frac{1}{\pi_a p_e}
\end{aligned}
$$

SFMPT

FFMPT

SMIFA∞

FFMPT (first fundamental mystery of probability theory): E is linear

SFMPT (first fundamental mystery of probability theory): E(E(X|Y))=E(X)

SMIFA ∞ (single most important fact about infinity): Geometric Series

Some References

Seminal paper:

A. K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, and P. Tiwari. The electrical resistance of a graph captures its commute and cover times. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pages 574–586, Seattle, May 1989.

Awesome Introduction to main ideas:

Peter G. Doyle and J. Laurie Snell. *Random Walks and Electric Networks.* The Mathematical Association of America, 1984, arXiv:math/0001057v1

Applications:

Klein, D. and Randic, M. (1993). Resistance distance. *Journal of Mathematical Chemistry*, *12*, 81-95.

Francois Fouss, Alain Pirotte, Jean-Michel Renders, Marco Saerens: Random-Walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation. IEEE Trans. Knowl. Data Eng. 19(3): 355-369 (2007)

Y. Ollivier and P. Senellart, Finding Related Pages Using Green Measures: An Illustration with Wikipedia. In *Proc. AAAI*, pp. 1427– Conference article ver, Canada, July 2007. (pdf | slides | slides, short | website | bib)

Pre-printS

Will post on CSSS10 Wiki, with Lecture Notes

Conformal geometry of Markov chains

Peter G. Doyle Gregory Leibon Jean Steiner

Markov chains in a shoebox

Peter G. Doyle Gregory Leibon Jean Steiner