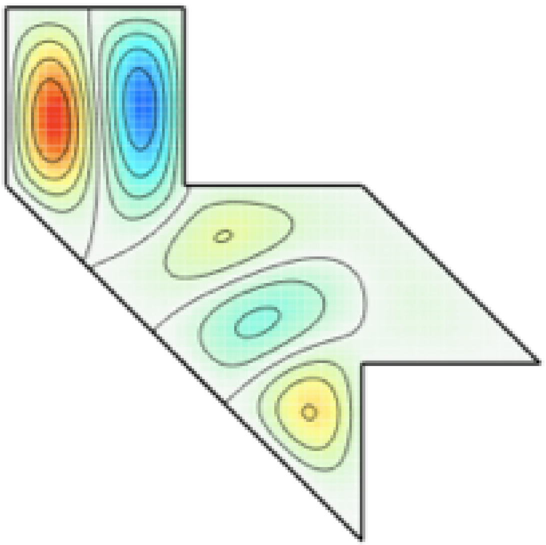
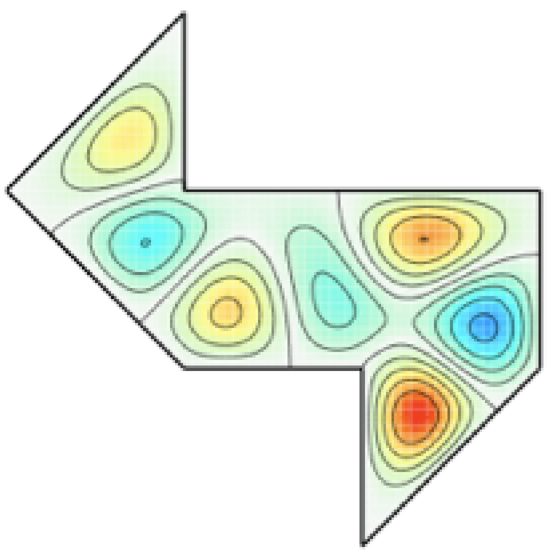
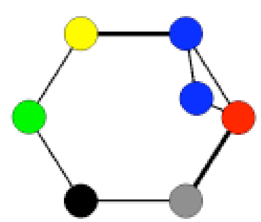
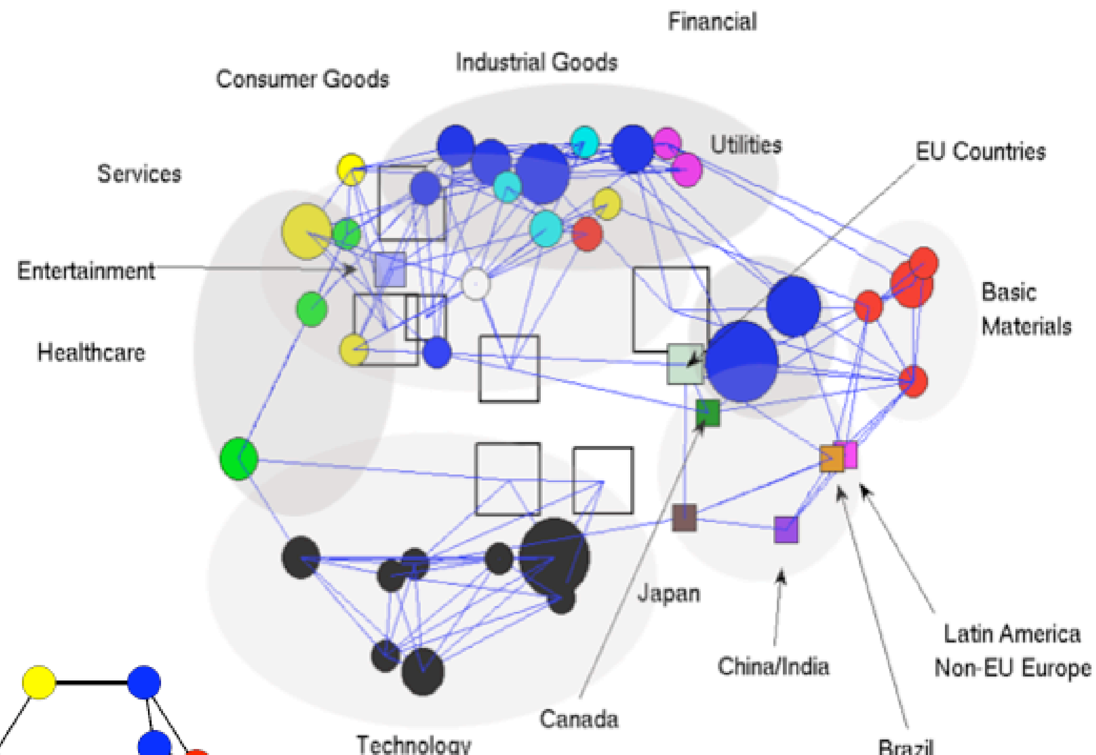
Unsupervised Learning in Complex Systems
Part B: Can you hear the shape of the market?
CSSS 2008







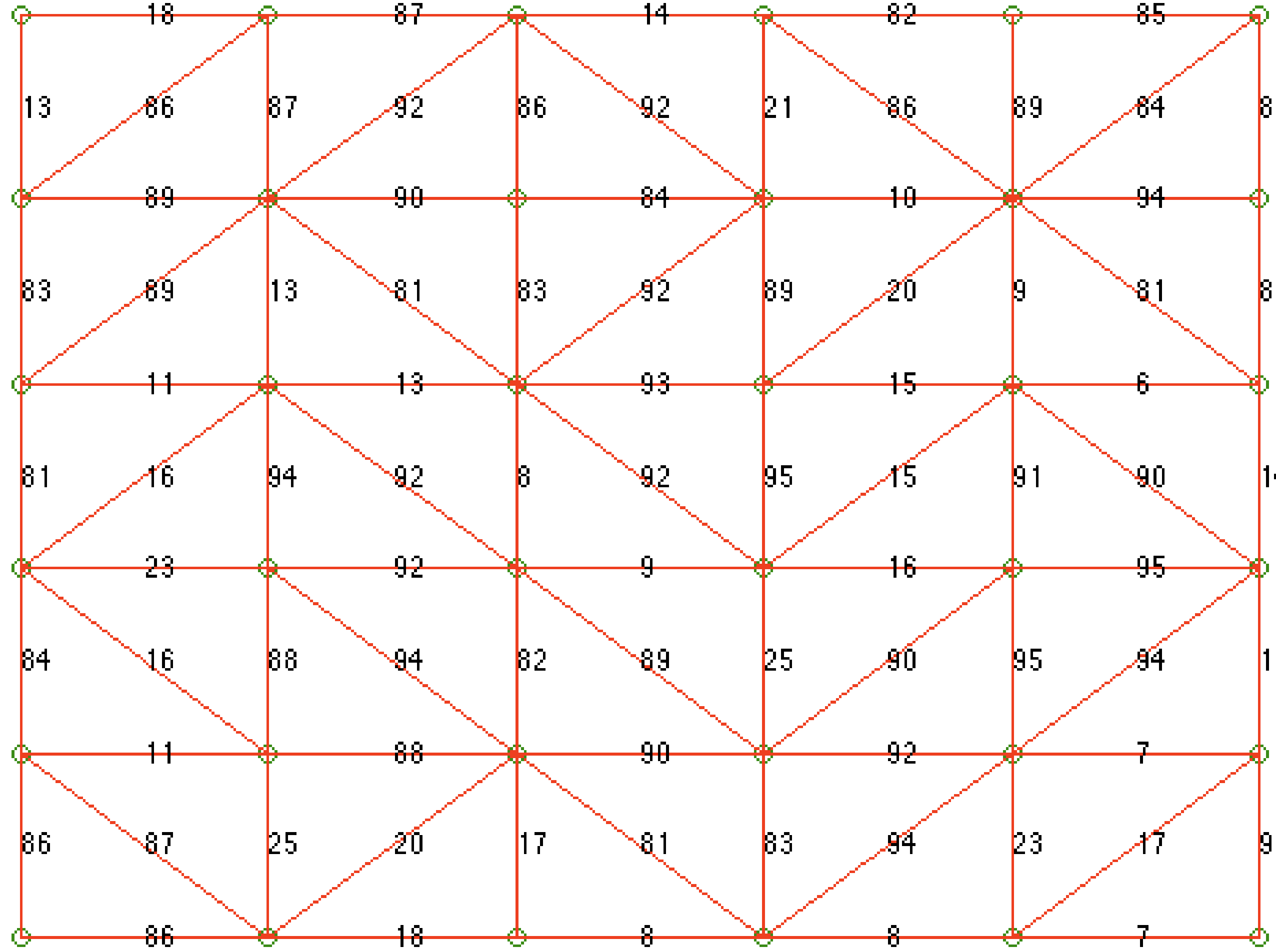


Gregory Leibon
Memento Security & Dartmouth College

What we do...

- Introduce random walks on a network and describe Vagabond Clustering
- We then “relax” Vagabond Clustering and describe Spectral Clustering
- Examine the topological properties of Spectral Clustering
- Listen to the Spectrally Clustered Market and 109th congress.

Conductance (or similarity) Network.



W=

0	86	0	0	0	0	90	0	0	0	0	0	0	0	0	0
86	0	10	0	0	0	96	6	13	0	0	0	0	0	0	0
0	10	0	18	0	0	0	0	6	0	0	0	0	0	0	0
0	0	18	0	8	0	0	0	97	92	93	0	0	0	0	0
0	0	0	8	0	16	0	0	0	0	15	6	0	0	0	0
0	0	0	0	16	0	0	0	0	0	0	17	0	0	0	0
90	96	0	0	0	0	0	17	0	0	0	0	96	0	0	0
0	6	0	0	0	0	17	0	94	0	0	0	13	86	0	0
0	13	6	97	0	0	0	94	0	91	0	0	0	100	96	0
0	0	0	92	0	0	0	0	91	0	95	0	0	0	99	19
0	0	0	93	15	0	0	0	0	95	0	12	0	0	0	0
0	0	0	0	6	17	0	0	0	0	12	0	0	0	0	0
0	0	0	0	0	0	96	13	0	0	0	0	0	8	0	0
0	0	0	0	0	0	0	86	100	0	0	0	8	0	98	0
0	0	0	0	0	0	0	0	96	99	0	0	0	98	0	11
0	0	0	0	0	0	0	0	0	19	0	0	0	0	11	0
0	0	0	0	0	0	0	0	0	90	93	0	0	0	0	11

Form the Transition Matrix

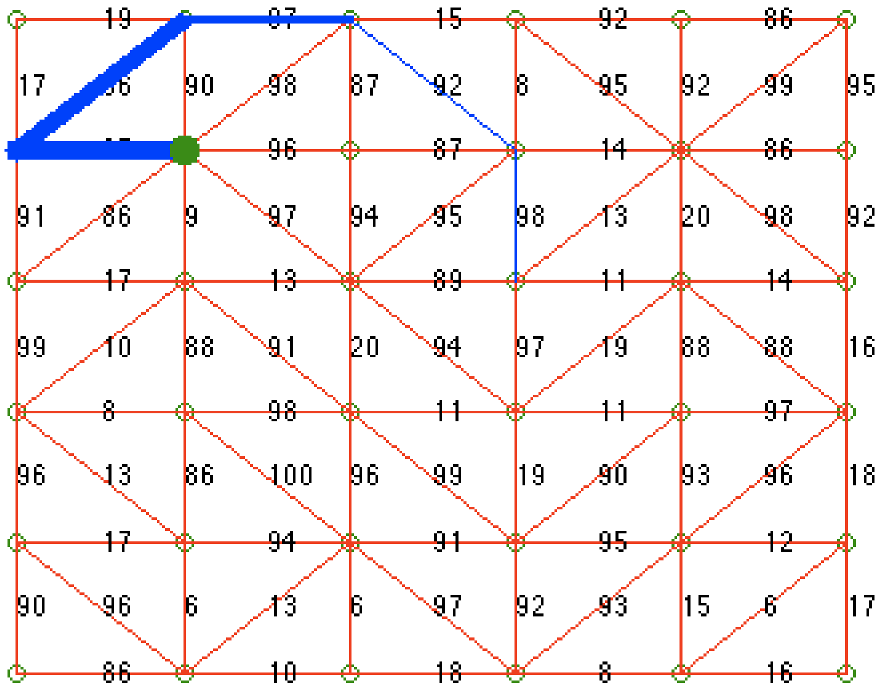
normalize	`W=(1/sum(sum(W)))*W;`
probability vector	`mu=sum(W,1);`
transition matrix	`T=diag(1./mu)*W;`

$$diag(\mu)T = W$$

Useful Observations:

$$\mu T = \mu$$

states=Snake(W,S,'No',N,K);




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

Interpret as Random

$$T_{ij} = P(X_{t+1} = j \mid X_t = i)$$

States

S(i)=[mod(i,6),floor(i/6)]

....26 32 26 19 25 32 26 21 28 21 28 22 28 33 26 21 15 16 21 16 21 26 27 33 26 25 26 33 32 33 27 26 32 33 26 20 14 20 15 9 4 9 8 9 15 20 14 9 10 15 9 14 8 14 15 9 4 3 2 7 13 7 2 9 8 9 8 9 15 9 4 11 18 17 23 16 22 21 16.....

MatLab:

```
E=repmat(1/36,36,36);
[seq,states] = hmmgenerate(N,T,E);
```

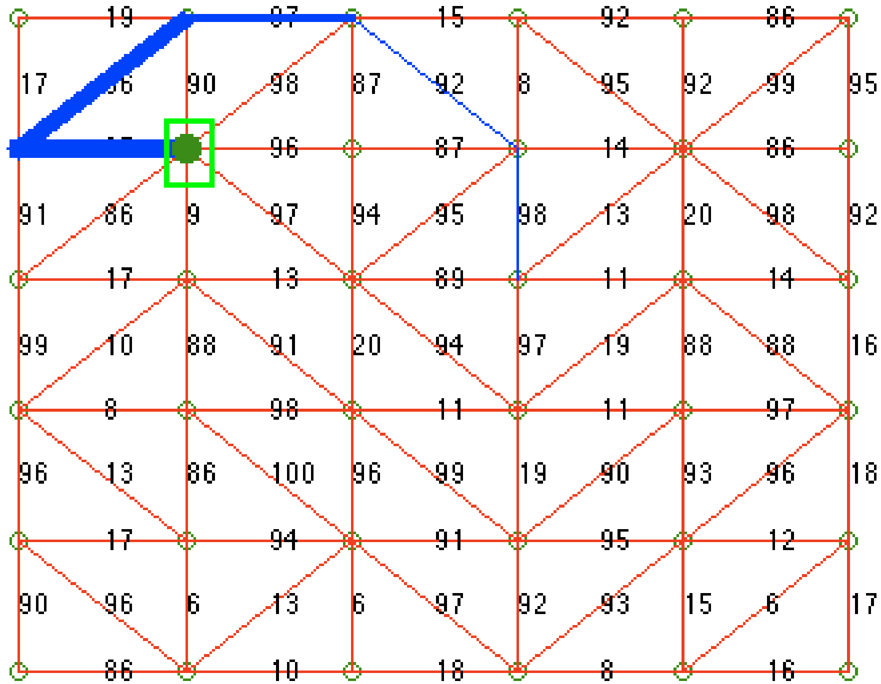
Called a Markov Chain

Question: Given I'm in state i what is the probability that I'm in state j after 2 steps?

$$P(X_{t+2} = j \mid X_t = i) = \sum_{k=1}^{|S|} P(X_{t+2} = j \mid X_{t+1} = k) P(X_{t+1} = k \mid X_t = i)$$

Hey! that's matrix multiplication! In general....

$$P(X_{t+N} = j \mid X_t = i) = \left(T^N\right)_{ij}$$



Equilibrium measure is the average

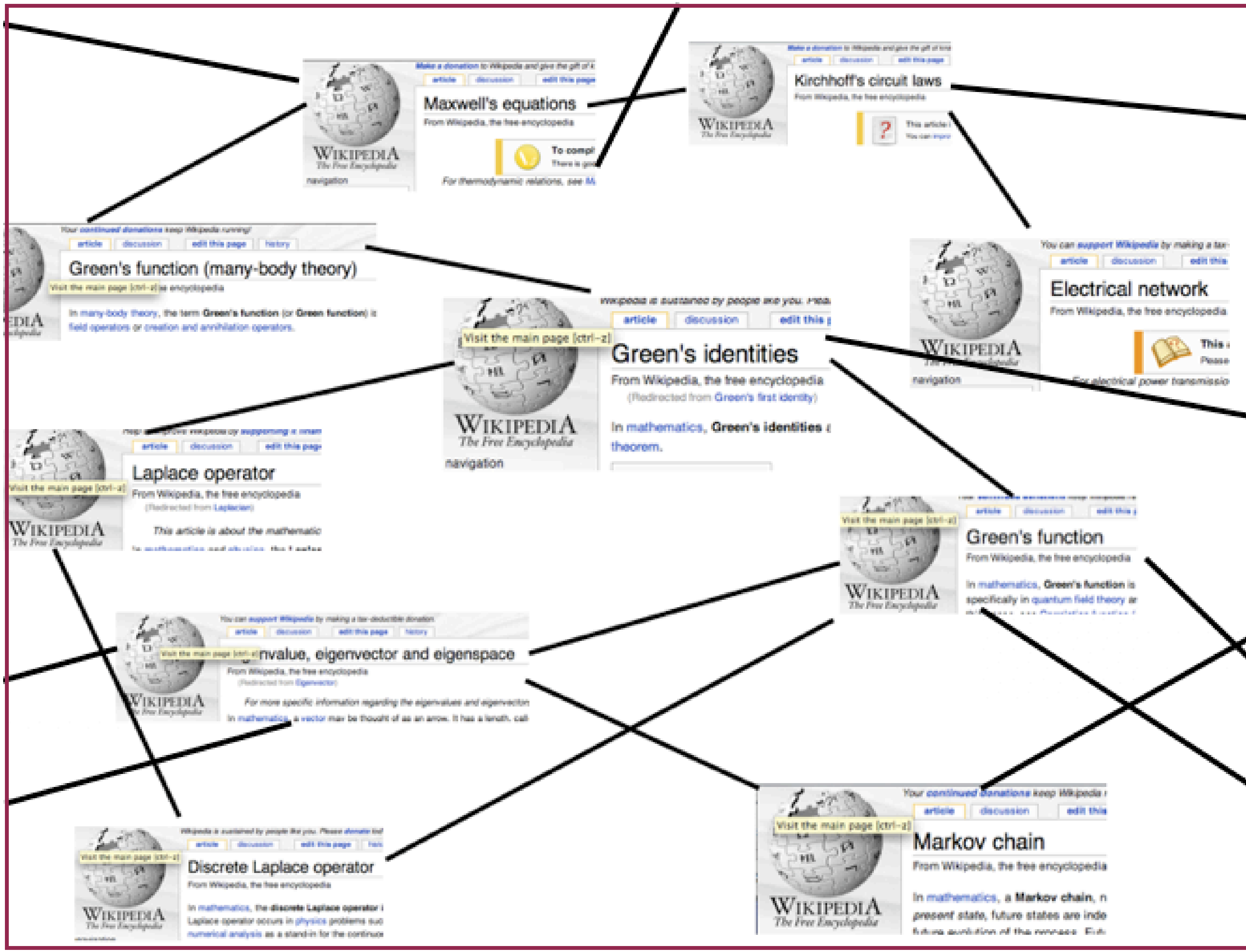
$$\lim_{T \to \infty} \frac{\sum_{k=1}^{T} 1_k(i)}{T} = \mu_i$$

Law of Large Numbers

$$1_k(i) = \begin{cases} 1 & X_k = i \\ 0 & X_k \neq i \end{cases}$$

....26 32 26 19 25 32 26 21 28 21 28 22 28 33 26 21 15 16 21
16 21 26 27 33 26 25 26 33 32 33 27 26 32 33 26 20 14 20 15
9 4 9 8 9 15 20 14 9 10 15 9 14 8 14 15 9 4 3 2 7
13 7 2 9 8 9 8 9 15 9 4 11 18 17 23 16 22 21 16.....

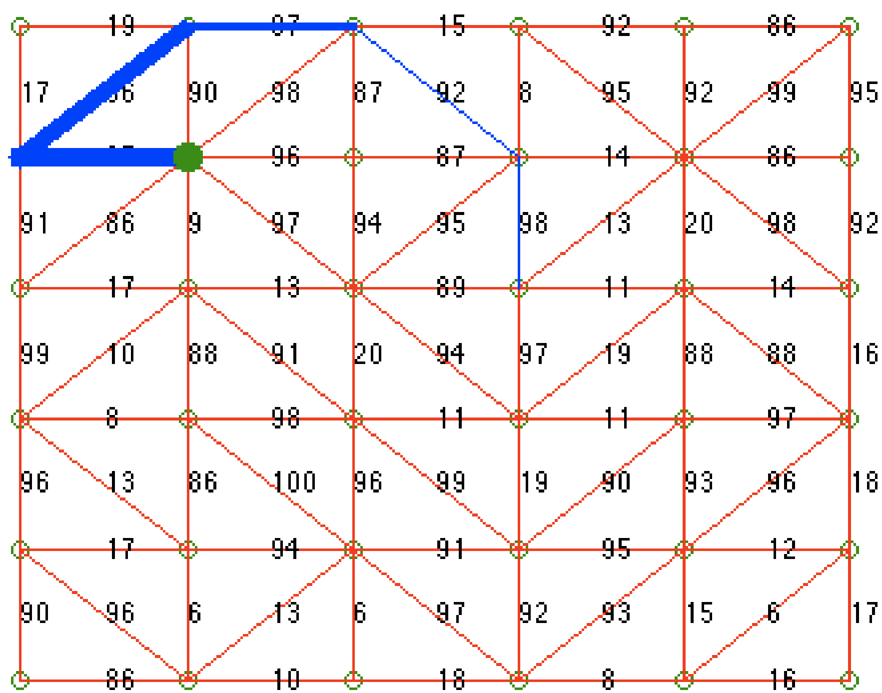
A Wild Wiki Adventure!



Let us cluster idea space....

"Vagabond Clustering": find a partition $\{A_k\}_{k=1}^{K}$ that minimizes

$$Vagabondliness = \sum_k P(X_{t+1} \in \bar{A}_k \mid X_t = A_k)$$

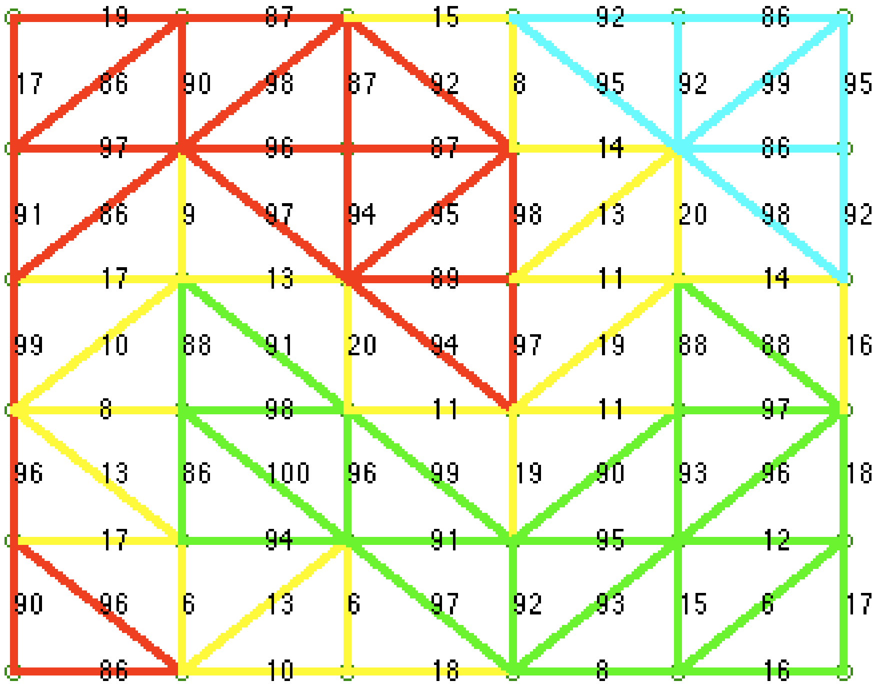


states=Snake(500,5,'No');

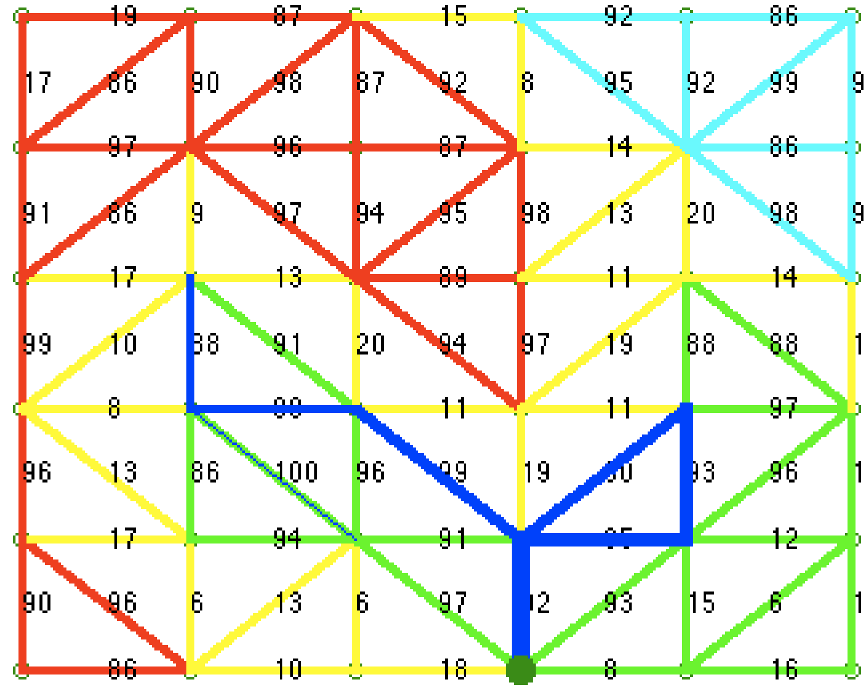
states=Snake(W,S,'No',N,K);

Finding this partition is usually called "NCut" and it is NP hard.

Stoer, M. and Wagner, F. (1997). A simple min-cut algorithm. *J. ACM*, *44*(4), 585–591.



Partition



states=Snake(500,10,'Vg');

Since finding this partition is NP Hard, to find the solution we will need to "relax..."

Key Theorem: $P(X_{t+1} \in \bar{A} \mid X_t \in A) = \langle \chi_A, \Delta \chi_A \rangle$

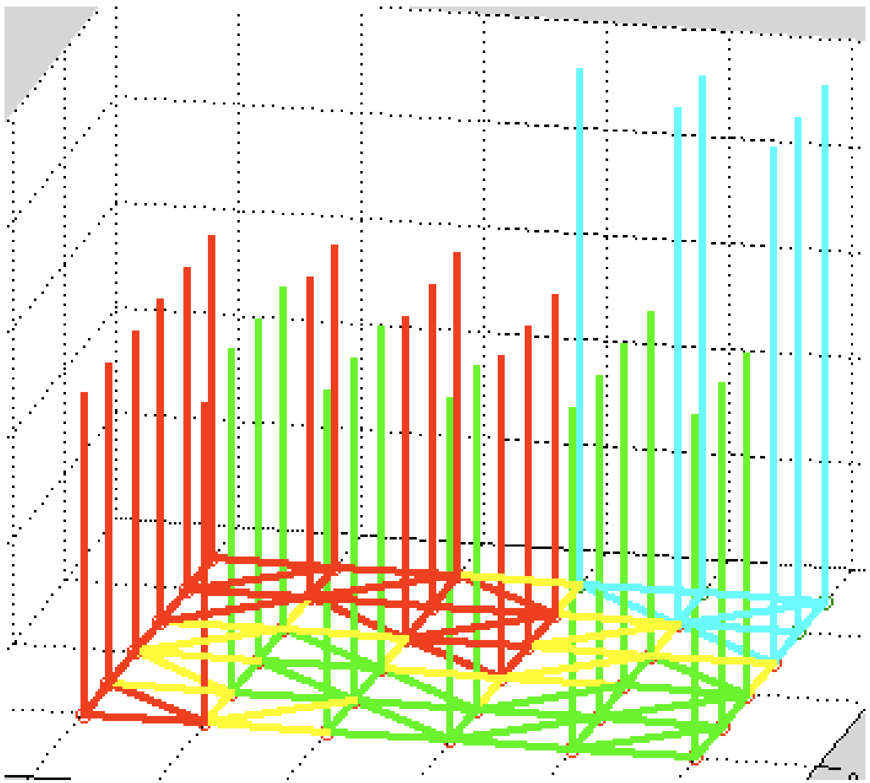
$$\langle f, g \rangle = \sum_k f_k g_k \mu_k$$

$$\chi_A(i) = \begin{cases} \frac{1}{\sqrt{\mu(A)}} & i \in A \\ 0 & i \notin A \end{cases}$$

for A and B disjoint

$$\langle \chi_A, \chi_B \rangle = 0$$

$$\langle \chi_A, \chi_A \rangle = 1$$



Inspiration, L^2 $\langle f, g \rangle = \int_M f(x) g(x) dVol(x)$

The Laplacian

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

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

therefore in this case $u.\delta x + v.\delta y + \text{v}.\delta z$ 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

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

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

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

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

Key property 1

$$< f, \Delta g >=< \Delta f, g >$$

proof:

$$\langle f, \Delta g \rangle = f^{tr} diag(\mu)(I - T)g$$

KEY: weights
are
symmetric

$$= f^{tr}(diag(\mu) - W)g$$

$$= ((diag(\mu) - W)f)^{tr} g$$

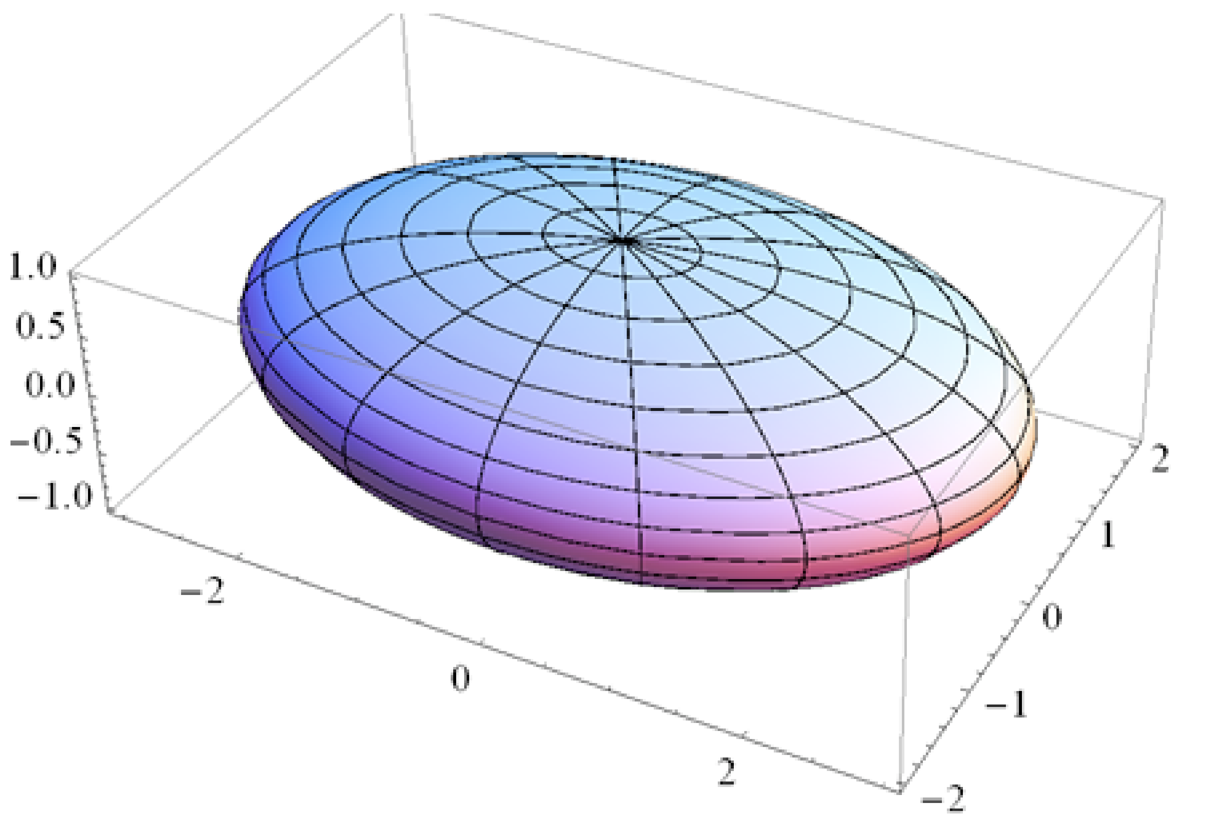
$$= (diag(\mu)(I - W)f)^{tr} g$$

$$= ((I - W)f)^{tr} diag(\mu) g = \langle \Delta f, g \rangle$$

Q.E.D

$$< f,\Delta g >=< \Delta f, g >$$

Spectral Theorem (Real): If <,> is an inner product and <Av,w>=<v,Aw>, then there is an orthonormal basis of A eigenvectors.



Find this basis with MatLab....

```
    %Create Laplacian and find eigenbasis
Lap=diag(ones(length(T),1)) - T;
[B Eig]=eig(Lap);
    % (Check: Lap=B*Eig*B^(-1))
    %Note: better algorithms for symmetric matrix
                              O=[];
for i=1:length(Lap)
    norm=sqrt((B(:,i))'.*mu*B(:,i));
    O=[O B(:,i)/norm ];
end
    % (Check:  O'*diag(mu)*O)
```

Key Property 2: Non-negative, with swanky formula!

$$\langle f, \Delta f \rangle = \frac{1}{2} \sum_{i,j} \mu_i T_{ij} (f_i - f_j)^2 \quad >=0$$

Proof: $$\langle f, \Delta f \rangle = \sum_{i,j} f_i \mu_i (I_{ij} - T_{ij}) f_j$$

$$= \frac{1}{2} \sum_i \mu_i f_i^2 - \sum_{i,j} f_i \mu_i T_{ij} f_j + \frac{1}{2} \sum_j \mu_j f_j^2$$

$$= \frac{1}{2} \left(\sum_{ij} \mu_i T_{ij} f_i^2 - 2 \sum_{i,j} \mu_i T_{ij} f_i f_j + \sum_{i,j} \mu_j T_{ij} f_j^2 \right)$$

$$= \frac{1}{2} \sum_{ij} \mu_i T_{ij} (f_i^2 - 2 f_i f_j + f_j^2)$$

row sum is one

is equillbrium measure

Key Theorem: $P(X_{t+1} \in \bar{A} \mid X_t \in A) = \langle \chi_A, \Delta \chi_A \rangle$

Approximate the Vagabond Clustering

$$\min_{partitions\{A_k\}} \sum_k \langle \chi_{A_k}, \Delta \chi_{A_k} \rangle$$

by "relaxing" and solving

$$\min_{\{v_i \mid \langle v_i, v_j \rangle = \delta^i_j\}} \sum_i \langle v_i, \Delta v_i \rangle$$

By the Spectral Theorem and positivity, to solve this problem we are searching for the N eigenvectors of the Laplacian with the N smallest eigenvalues.

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

~Lord Kelvin

Proof: $$P(X_{t+1} \in \bar{A} \mid X_t \in A) = \frac{P(X_{t+1} \in \bar{A}, X_t \in A)}{P(X_t \in A)}$$

definition conditional probability

$$= \frac{\sum_{j \in \bar{A}, i \in A} P(X_{t+1} = j, X_t = i)}{P(X_t \in A)}$$

$$= \frac{\sum_{j \in \bar{A}, i \in A} P(X_{t+1} = j \mid X_t = i) P(X_t = i)}{P(X_t \in A)}$$

$$= \frac{\sum_{j \in \bar{A}, i \in A} T_{ij} \mu_i}{\mu(A)}$$

Swanky formula

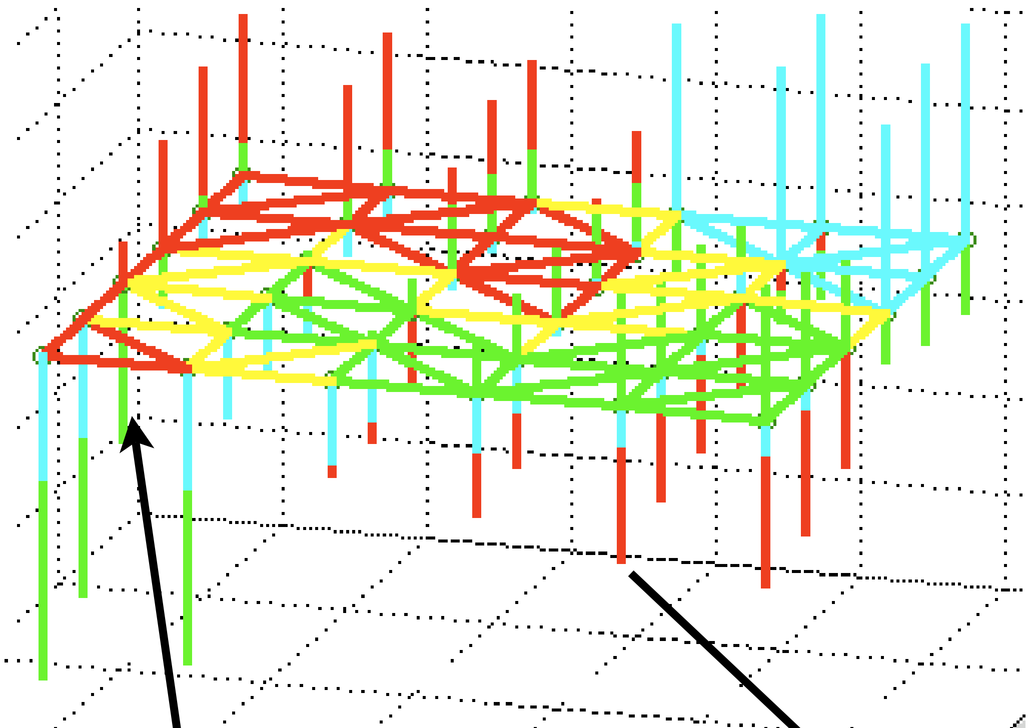
$$= \sum_{j \in \bar{A}, i \in A} T_{ij} \mu_i \left(\chi_A(i) - \chi_A(j)\right)^2$$

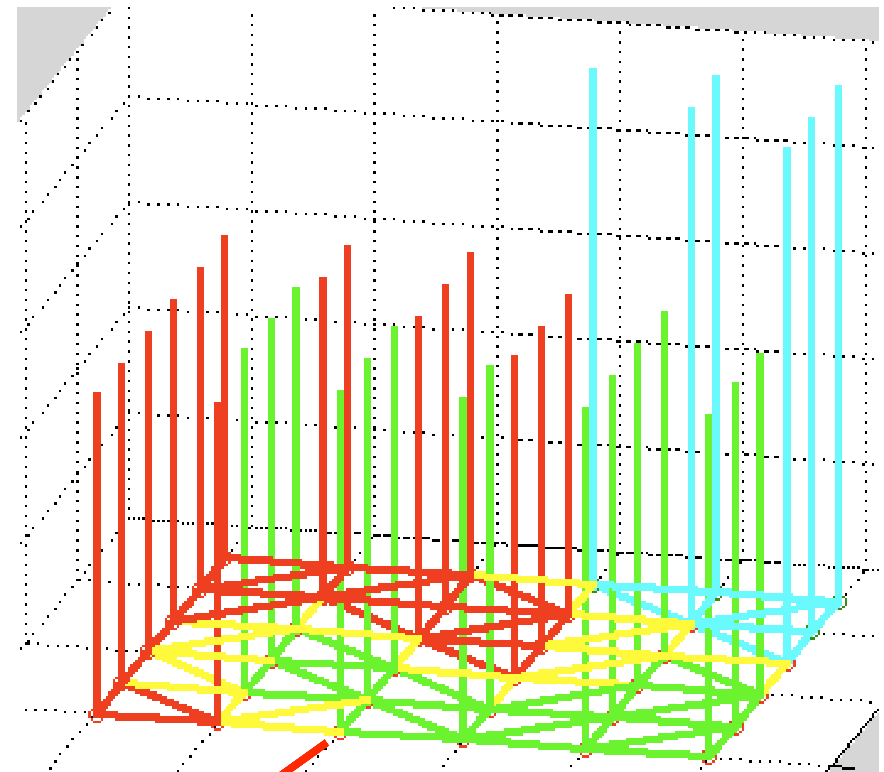
$$= \frac{1}{2} \sum_{i,j} T_{ij} \mu_i \left(\chi_A - \chi_A\right)^2 = \langle \chi_A, \Delta \chi_A \rangle$$

Q.E.D.

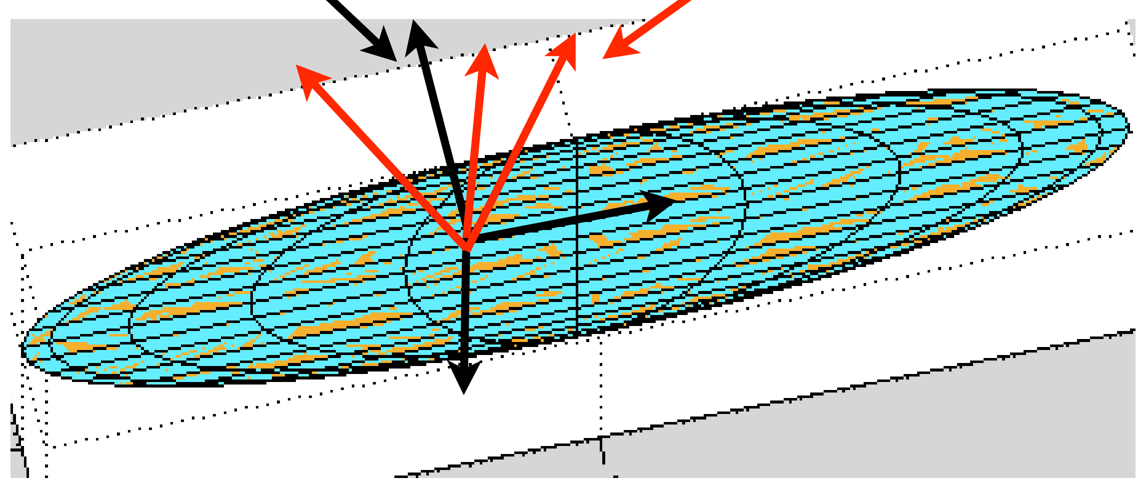
First three (non-trivial) Eigenfunctions

The Vagabond functions

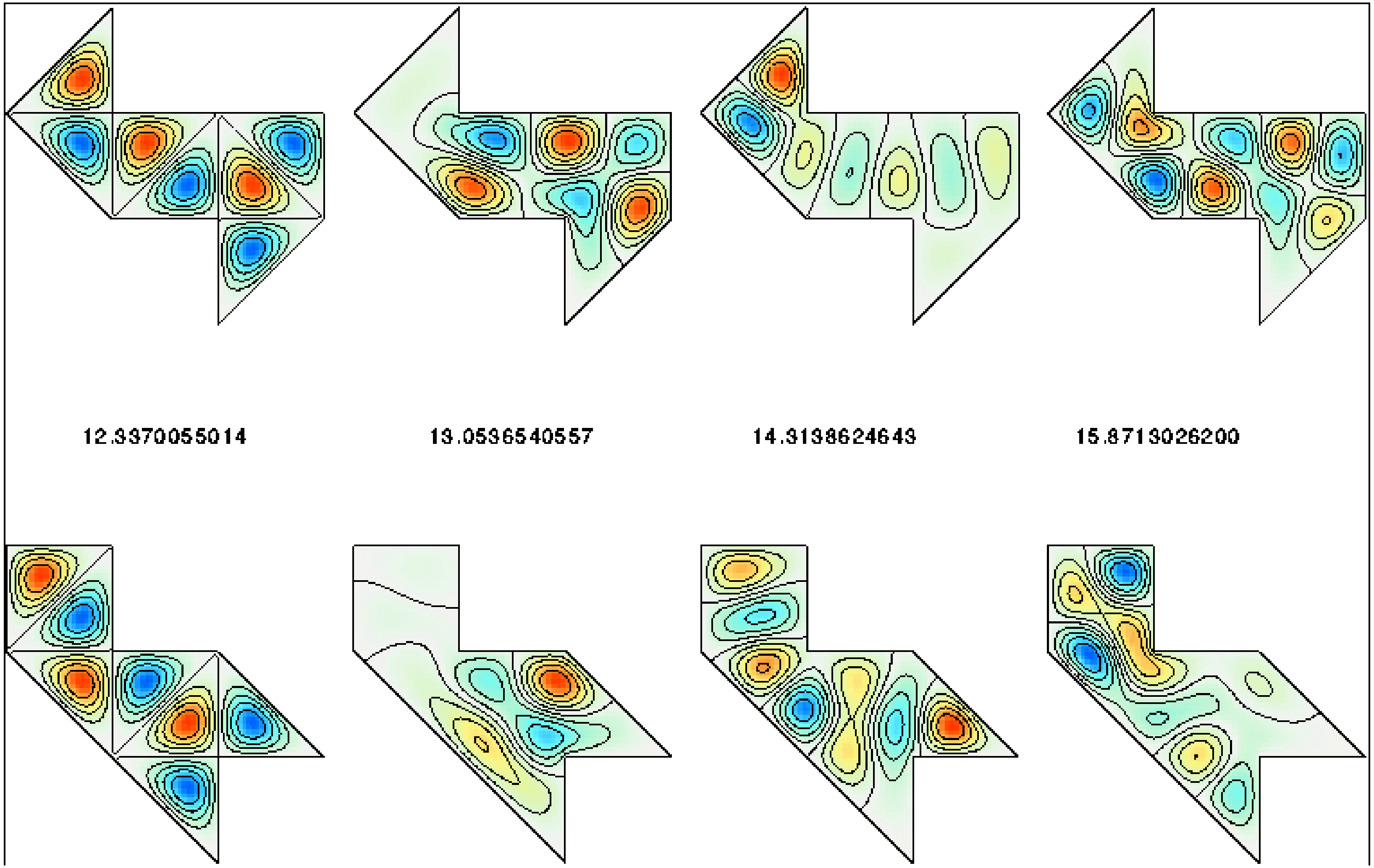




Notice, they are not localized



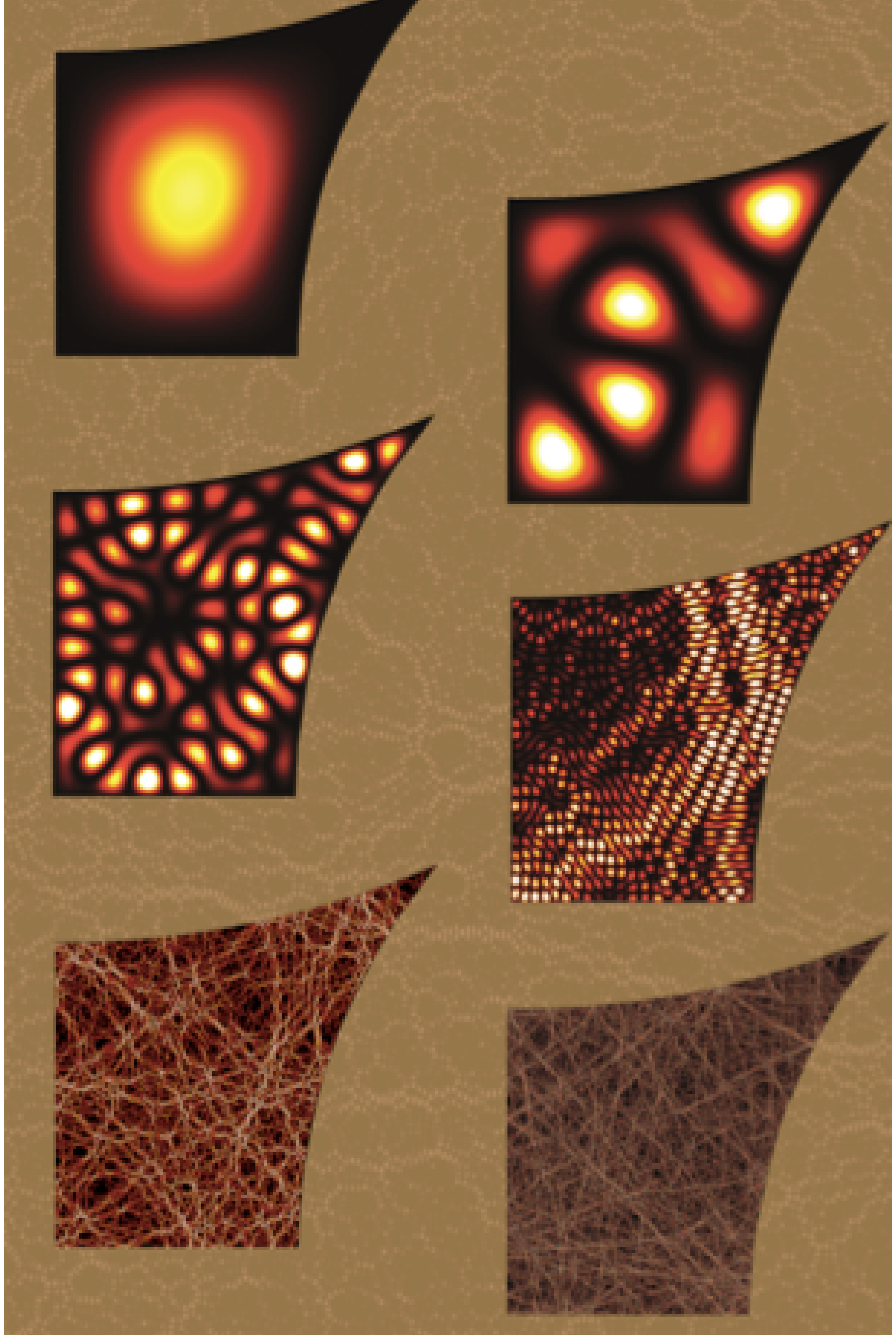
Can you hear the shape of a drum?



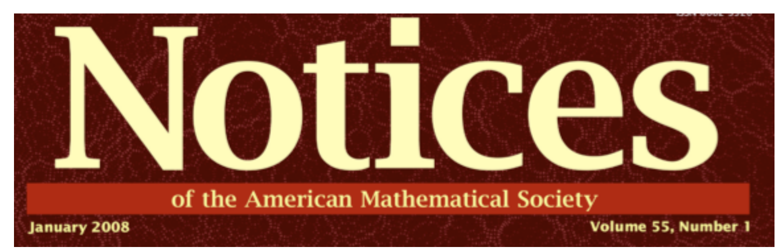
T. A. Driscoll. Eigenmodes of isospectral drums. *SIAM Review* 39, pp. 1-17, 1997.

C. Gordon, D. Webb, and S. Wolpert, *Isospectral plane domains and surfaces via Riemannian orbifolds*, Invent. Math., 110 (1992), pp. 1–22.

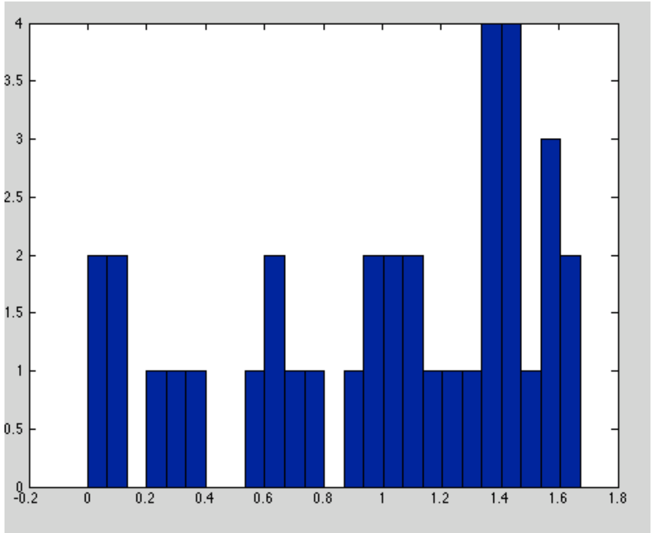
M. Kac, *Can one hear the shape of a drum?*, Amer. Math. Monthly, 73 part II (1966), pp. 1–23.



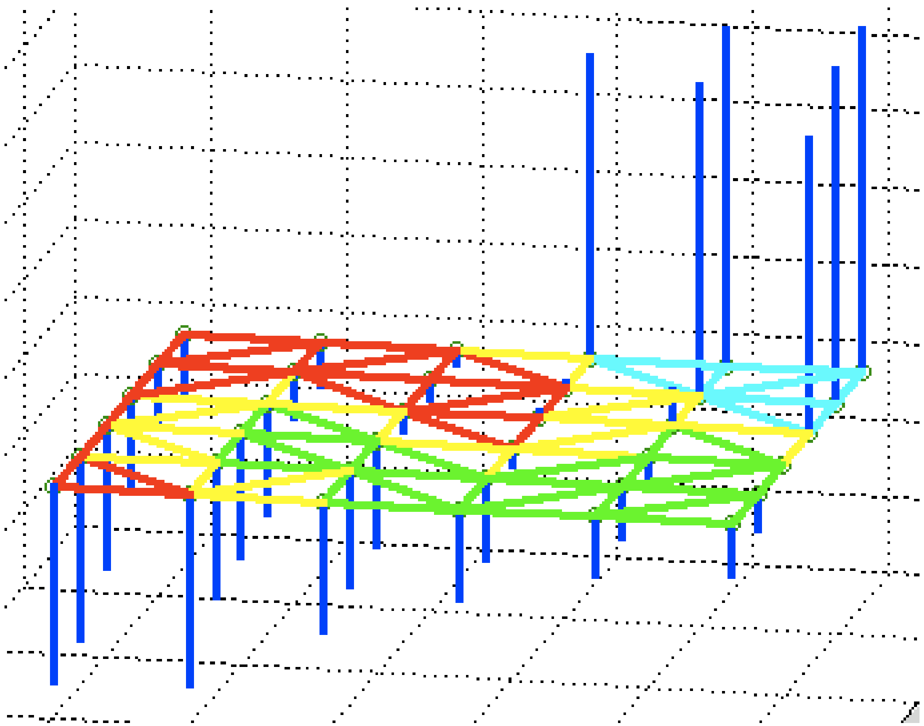
Eigenvalues are the "frequency of oscillation" of the eigenfunctions.



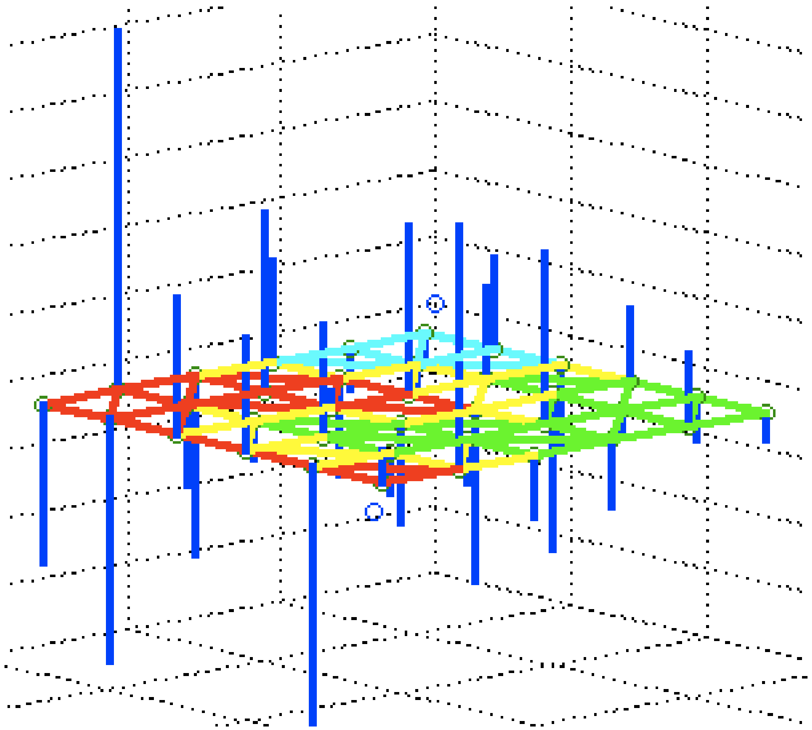
Alex Barnett



Eigenvalues are the frequency of oscillation

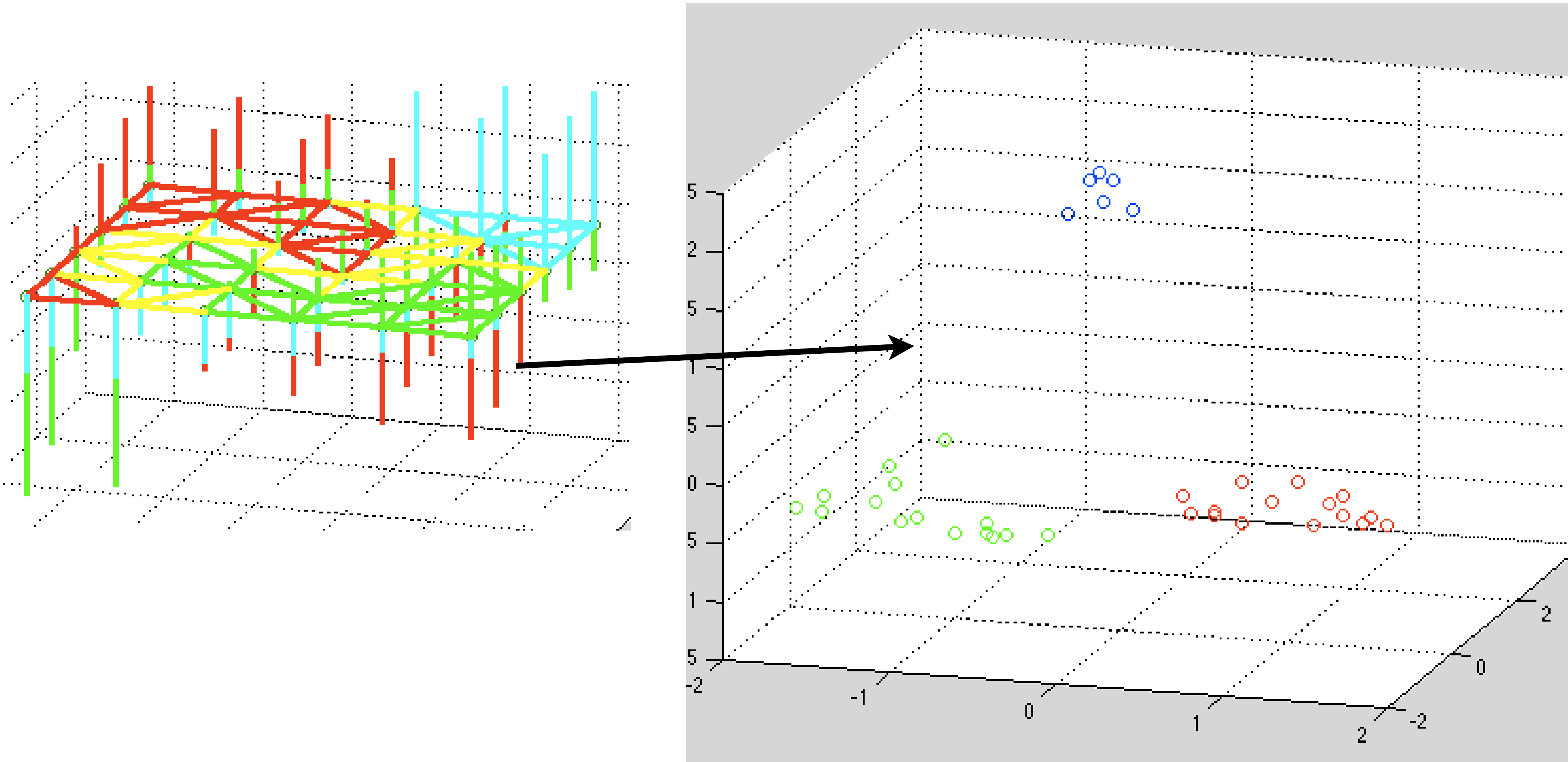


λ_2



λ_{30}

Can use our eigenfunctions to embed our states in Euclidean space



Now we can use K means!

Spectral Clustering Algorithm

1. Find the M>=K orthonormal eigenvectors corresponding to the M smallest eigenvalues.

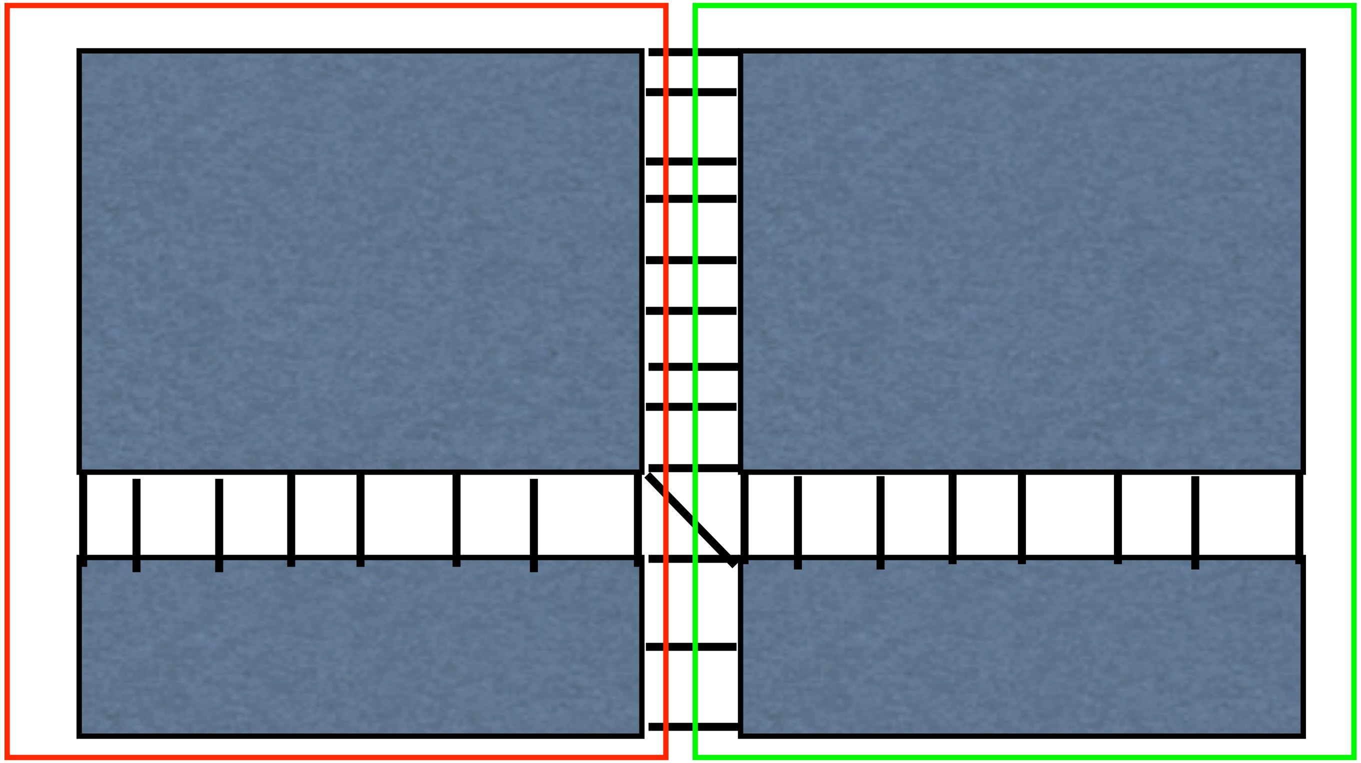
2. Using these eigenfunctions, embed our states into Euclidean space and then apply K-means.

Many Variations...

See: On spectral clustering: Analysis and an algorithm. A. Y. Ng, M. I. Jordan, and Y. Weiss. In T. Dietterich, S. Becker and Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems (NIPS) 14*, 2002.

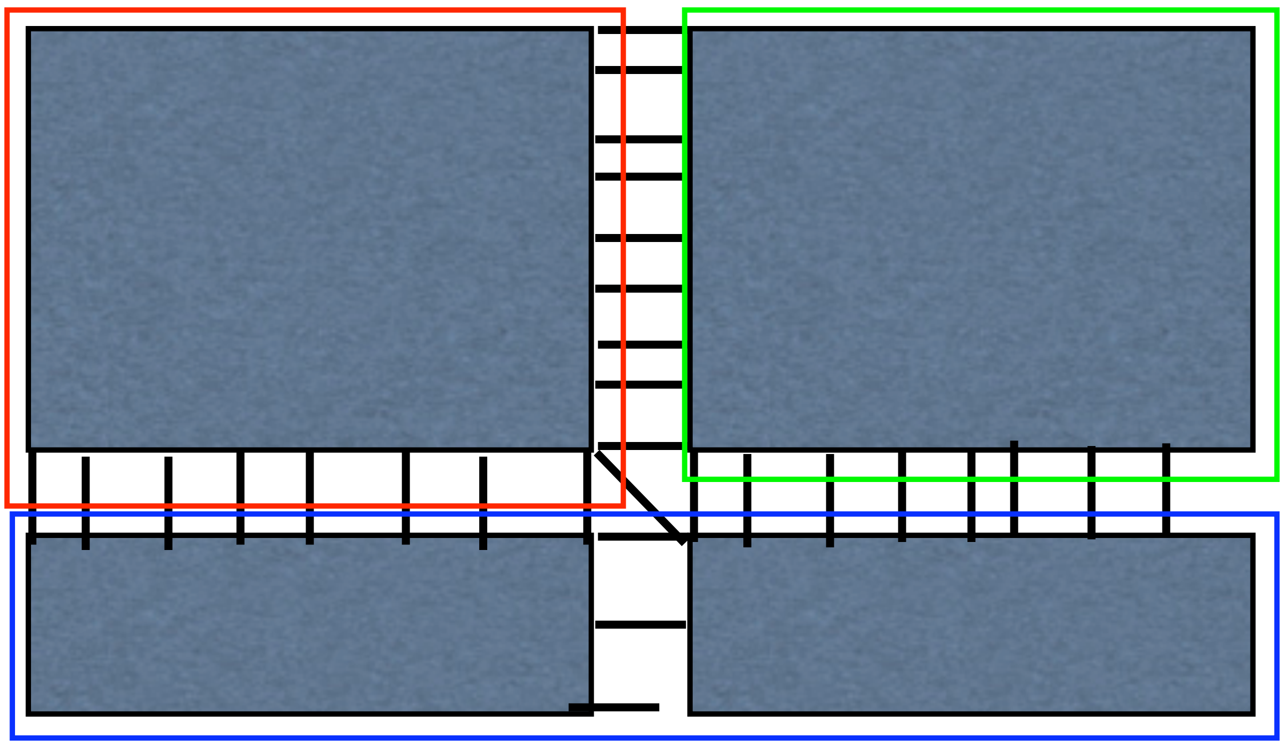
Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8), 888-905.

von Luxburg, U.: A Tutorial on Spectral Clustering. Statistics and Computing **17(4)**, 395-416 (12 2007)



Typically M should be GREATER than K

Here we see describe he vagabond clustering for both N=2 and N=3 requires the span of 4 Vagabond functions.

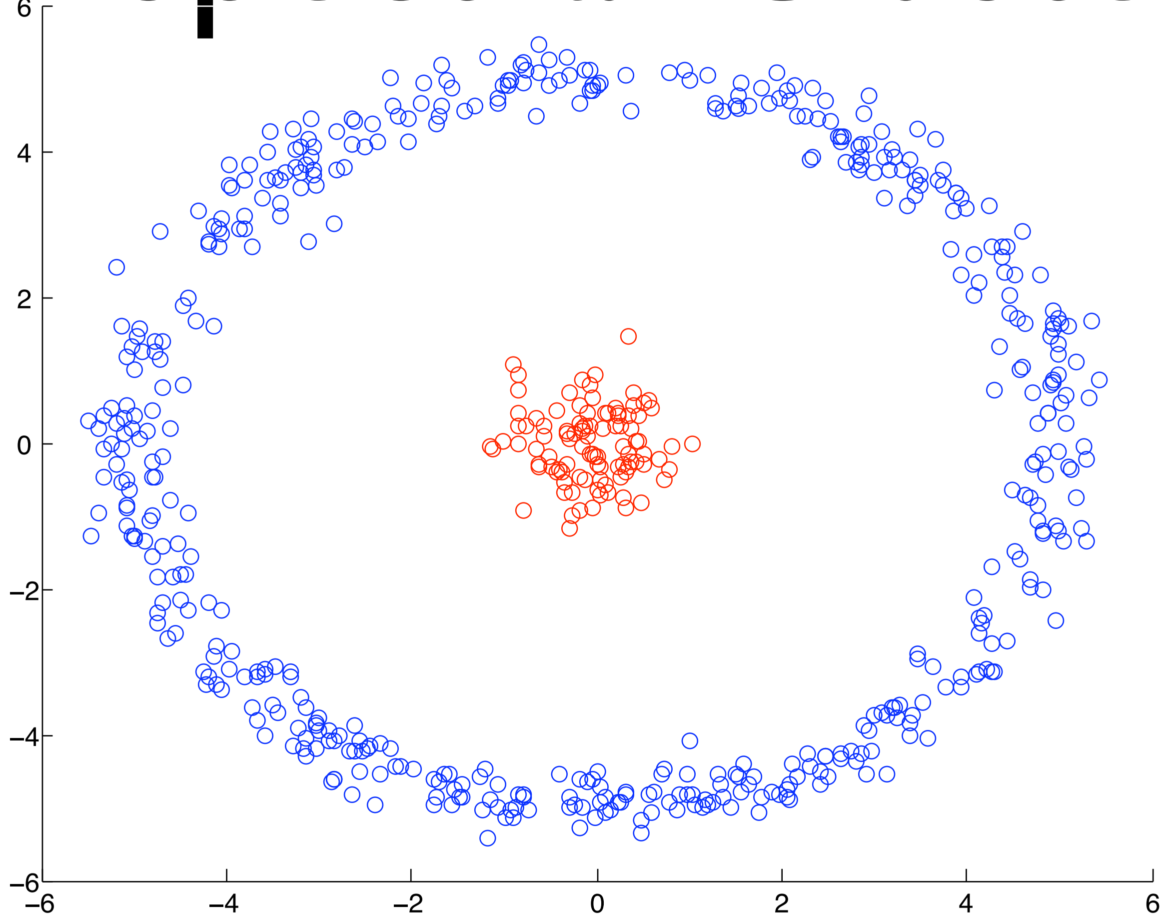


The Laplacian is trying to do MANY things at once (in particular it is the solution to a variety of discrete relaxations.)

In Mat Lab.

```
   %Sort by eigenvalue the eigenbasis [Eig,O]
Eig=diag(Eig);
[Eig Srt]=sort(Eig);
O=O(:,Srt);
    %Now apply K-means 'Rep' times
Emb=O(:,2:N);
[IDX,C,sumd,D] = kmeans(Emb,K,'emptyaction','drop');
for i=1:Rep
    [IDX0,C,sumd0,D] = kmeans(Emb,K,'emptyaction','drop');
    if (sum(sumd0)<sum(sumd))
        IDX=IDX0;
        sumd=sumd0;
    end;
end;
```

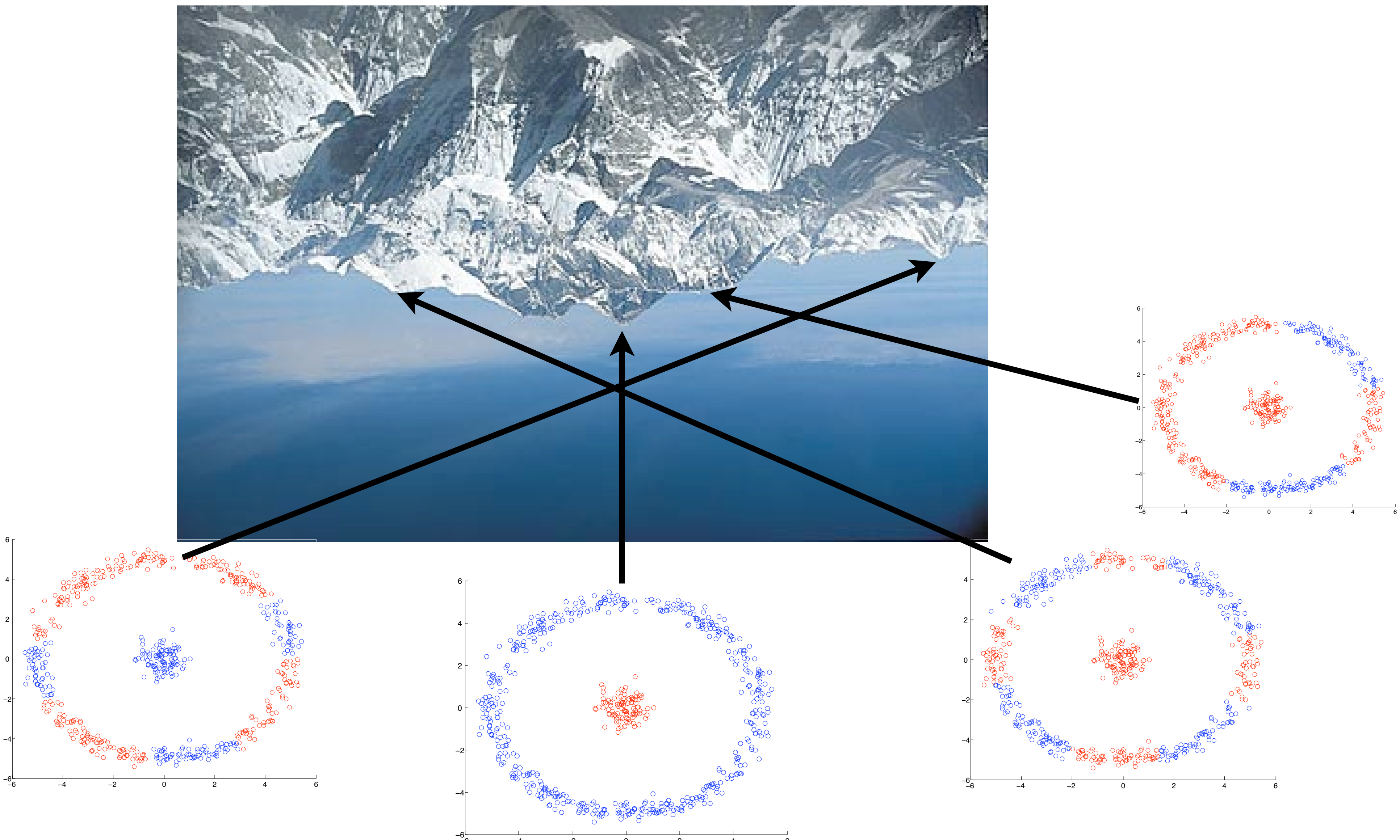
Spectral Clustering



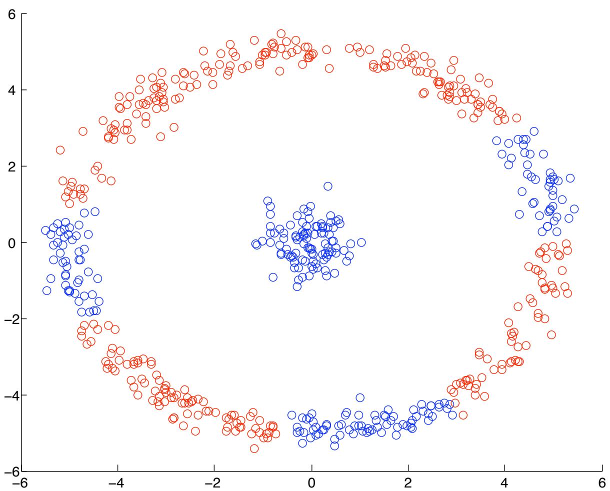
Manifold Learning

$$L = \sum_{k=1}^{N} \sum_{x_i \in C_k} d(x_i, \mu_k) \approx 488$$

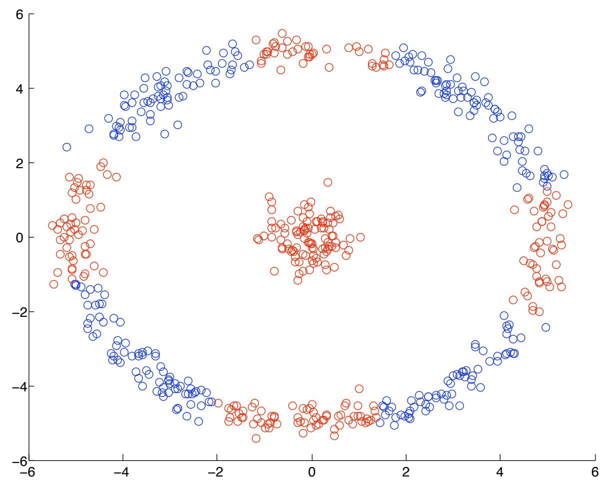
Himalayas



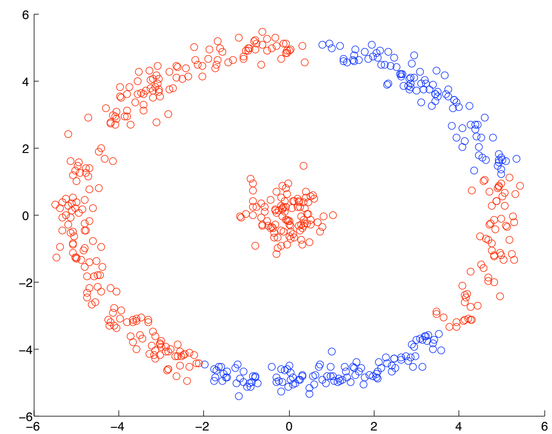
Watch Out! 100 convergent runs of the k-means algorithm were performed.



$$L = \sum_{k=1}^{N} \sum_{x_i \in C_k} d(x_i, \mu_k) \approx 536$$



$$L = \sum_{k=1}^{N} \sum_{x_i \in C_k} d(x_i, \mu_k) \approx 538$$



$$L = \sum_{k=1}^{N} \sum_{x_i \in C_k} d(x_i, \mu_k) \approx 534$$

Can you hear the shape of the market?

$$\hat{X} = \frac{X - \langle X \rangle}{\sqrt{\left\langle (X - \langle X \rangle)^2 \right\rangle}}$$

$$\rho(X, Y) = \hat{X} \cdot \hat{Y}$$

$$d(X, Y) = 2\sin(\theta/2) = \sqrt{2\left(1 - \rho(X, Y)\right)}$$

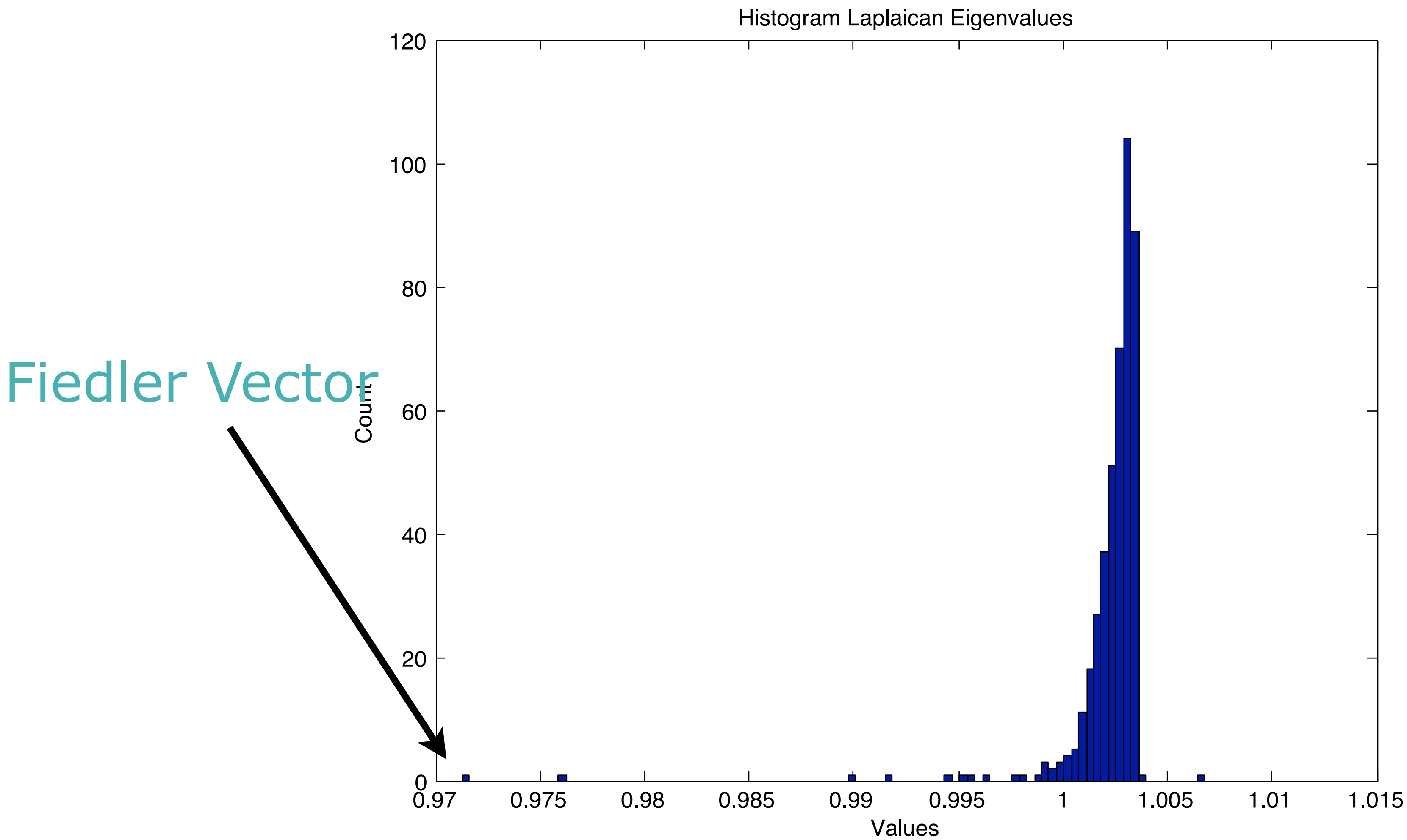
Candidate Weights

$$Conductance_s = e^{\frac{-d(X,Y)^2}{\sigma^2}}$$

$$Conductance_e = \frac{1}{d(X, Y)}$$

...and zero on diagonal, force a move.

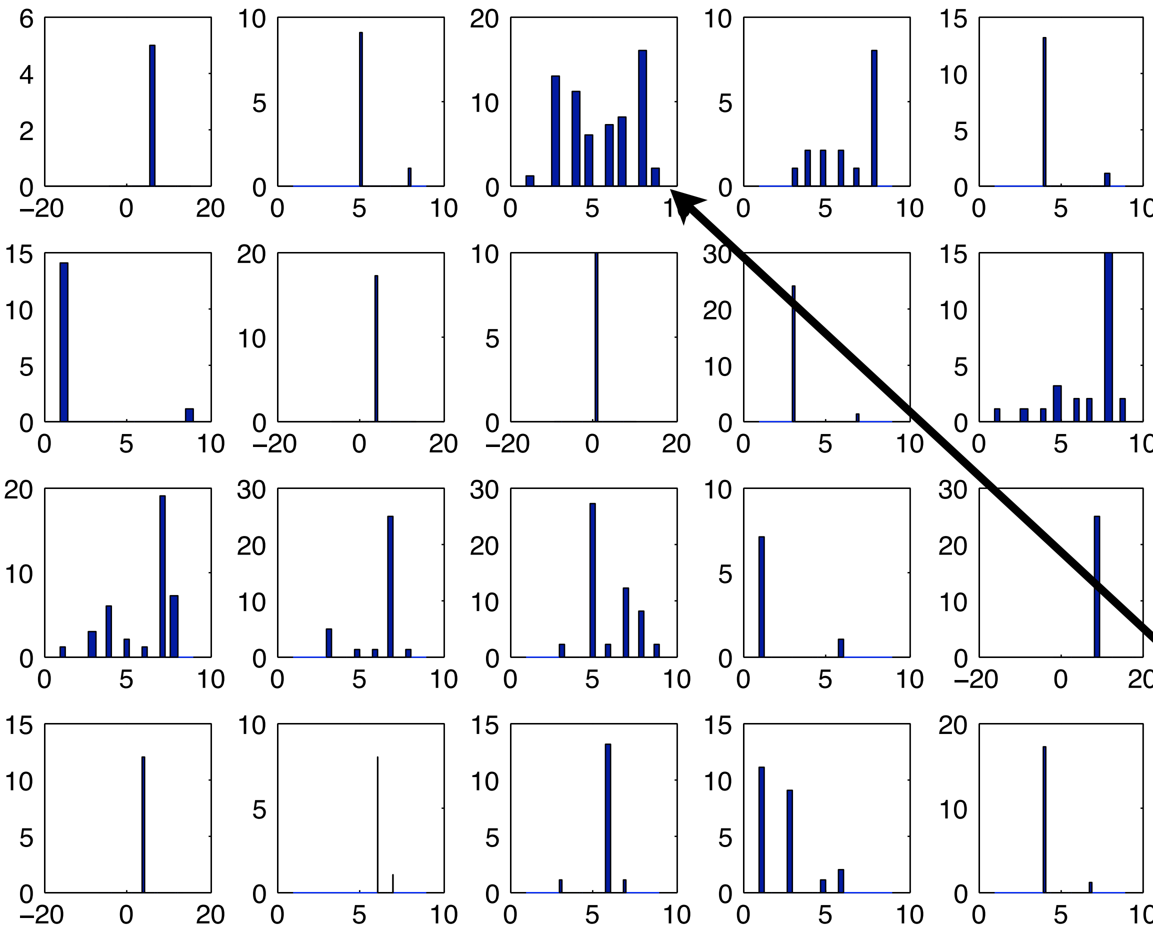
Eigenvalues>0



Major clusters correspond to the "outlier" eigenvalues....

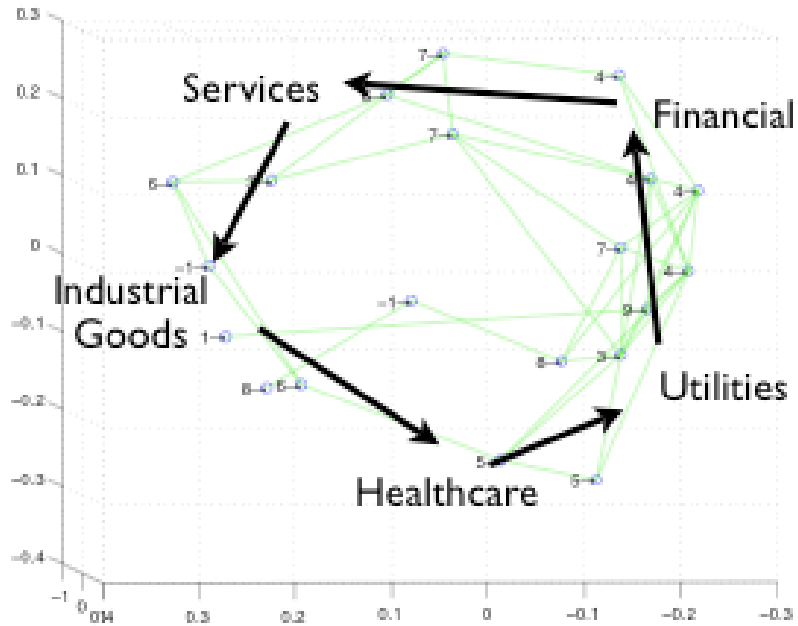
...though this localization is "lucky"

The Spectral Market

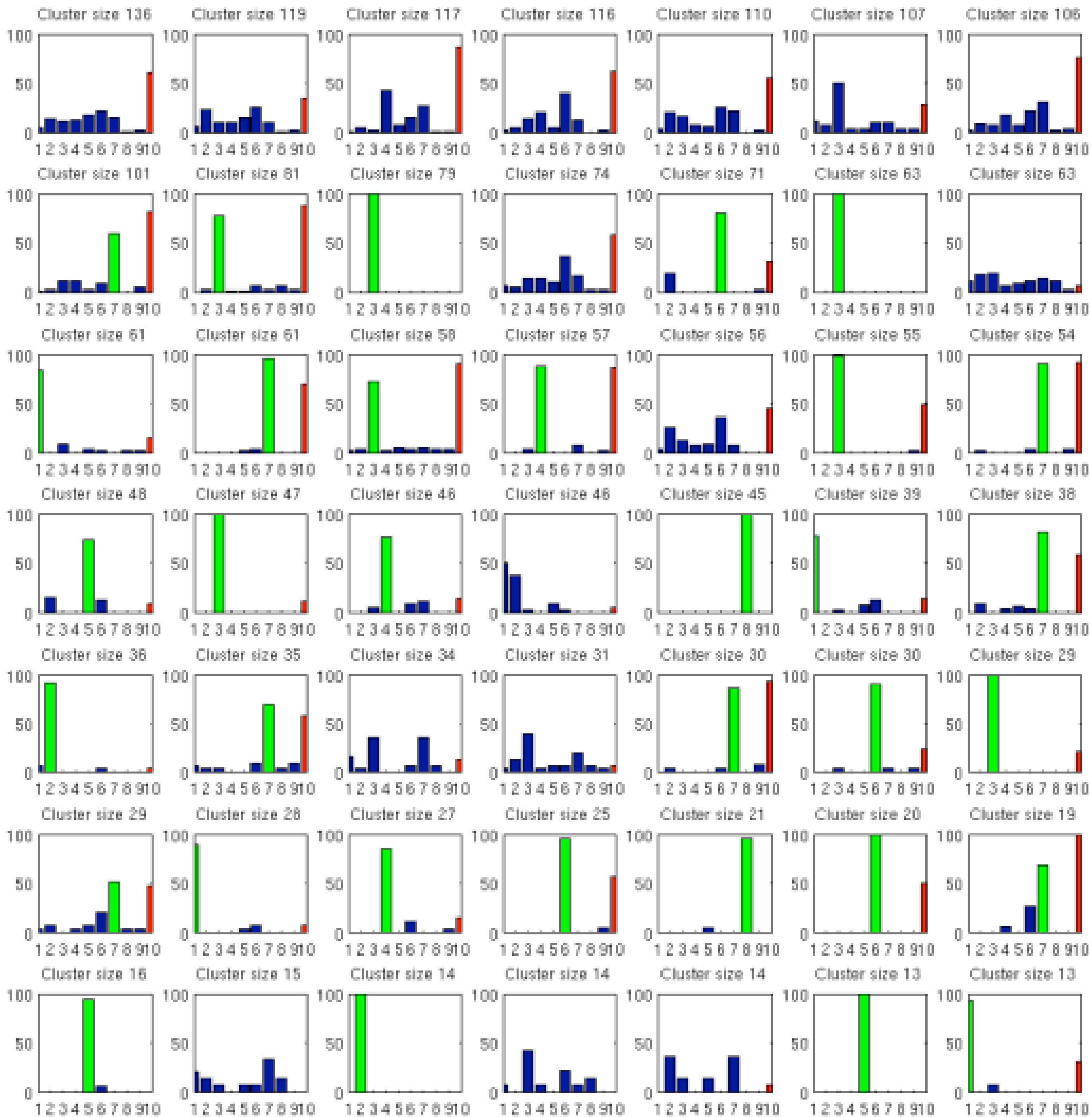


1 Basic Materials
2 Conglomerates
3 Consumer Goods
4 Financial
5 Healthcare
6 Industrial Goods
7 Services
8 Technology
9 Utilities

Guess who that is?

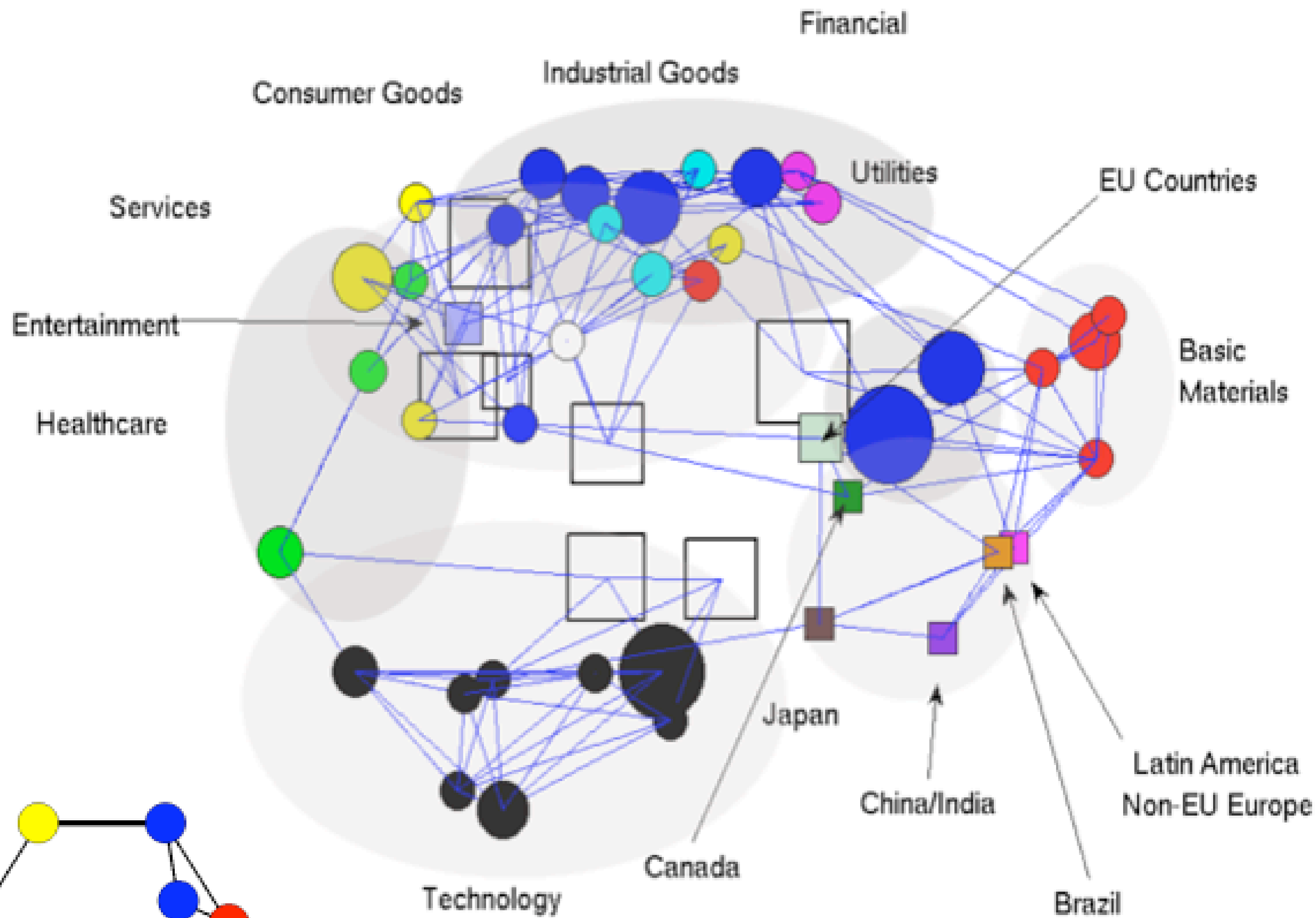


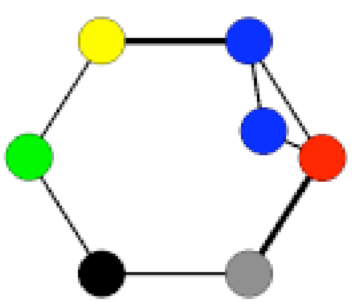
The Whole Market....



Cluster size 136
Cluster size 119
Cluster size 117
Cluster size 116
Cluster size 110
Cluster size 107
Cluster size 106
Cluster size 101
Cluster size 81
Cluster size 79
Cluster size 74
Cluster size 71
Cluster size 63
Cluster size 63
Cluster size 61
Cluster size 61
Cluster size 58
Cluster size 57
Cluster size 56
Cluster size 55
Cluster size 54
Cluster size 48
Cluster size 47
Cluster size 46
Cluster size 46
Cluster size 45
Cluster size 39
Cluster size 38
Cluster size 36
Cluster size 35
Cluster size 34
Cluster size 31
Cluster size 30
Cluster size 30
Cluster size 29
Cluster size 29
Cluster size 28
Cluster size 27
Cluster size 25
Cluster size 21
Cluster size 20
Cluster size 19
Cluster size 16
Cluster size 15
Cluster size 14
Cluster size 14
Cluster size 14
Cluster size 13
Cluster size 13

The Spectral Market



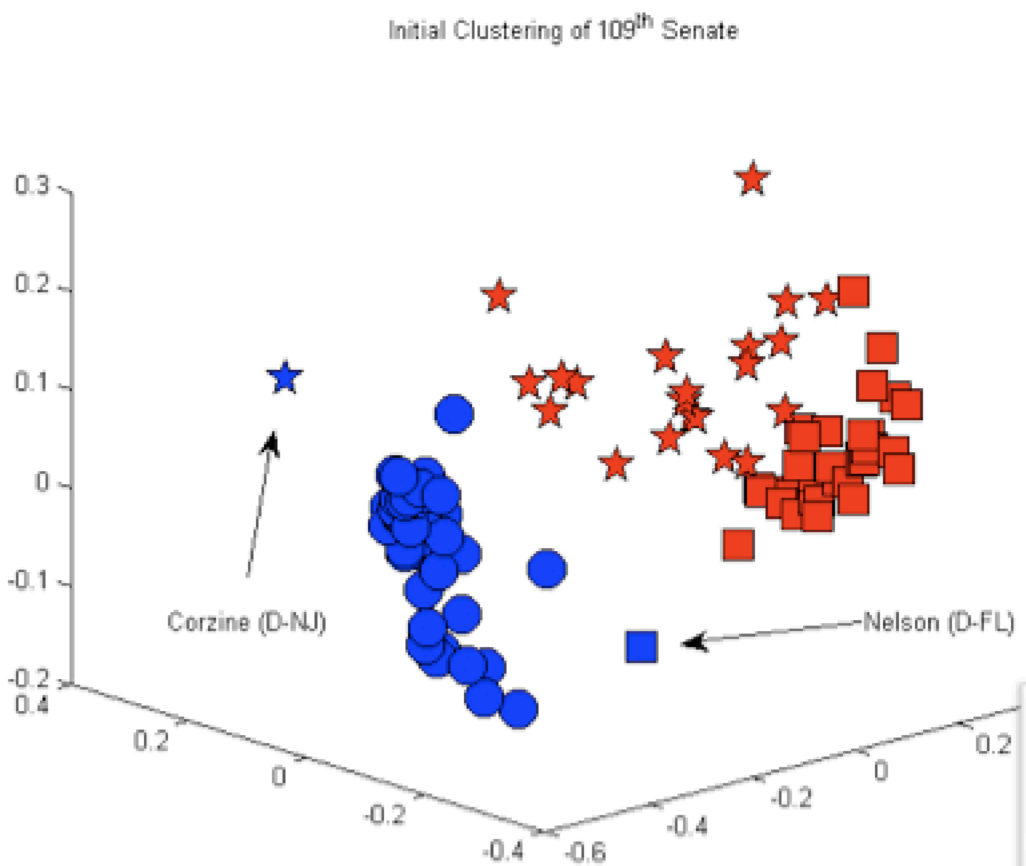


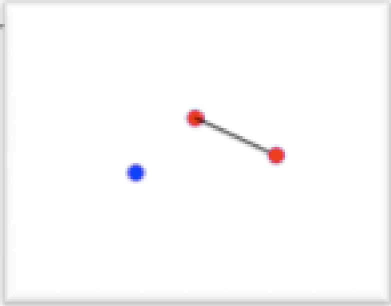
Greg Leibon, Scott D. Pauls, Daniel N. Rockmore, Robert Savell
Topological structures in the equities market network
http://arXiv.org/abs/0805.3470

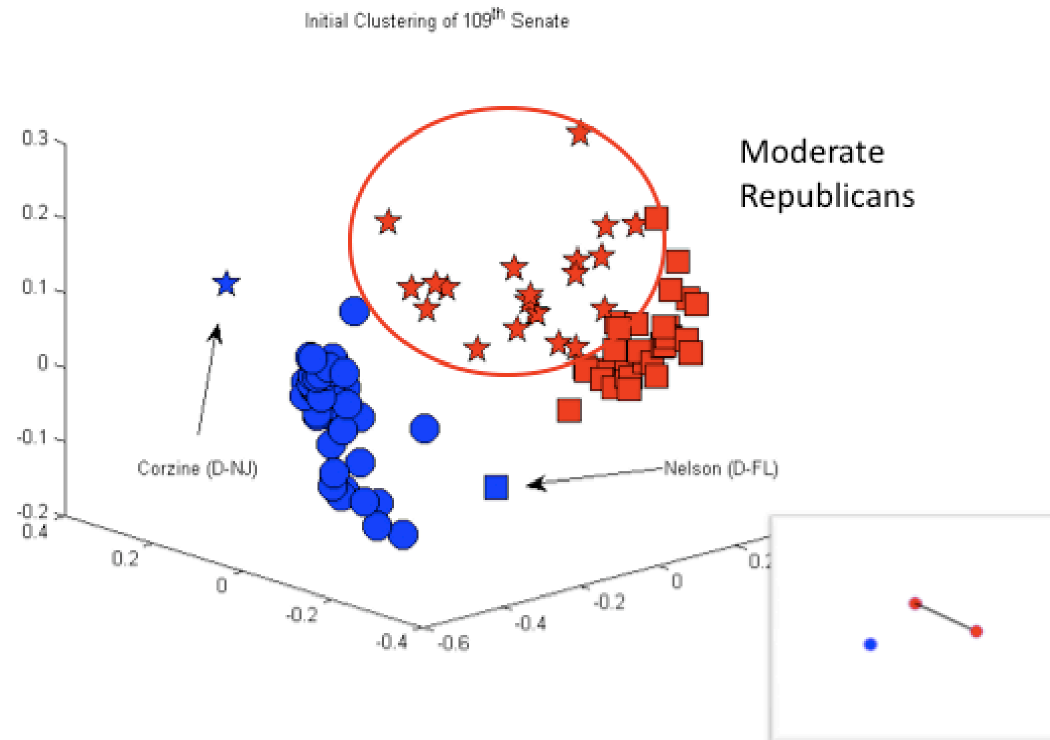
Example 2: 109th Congress

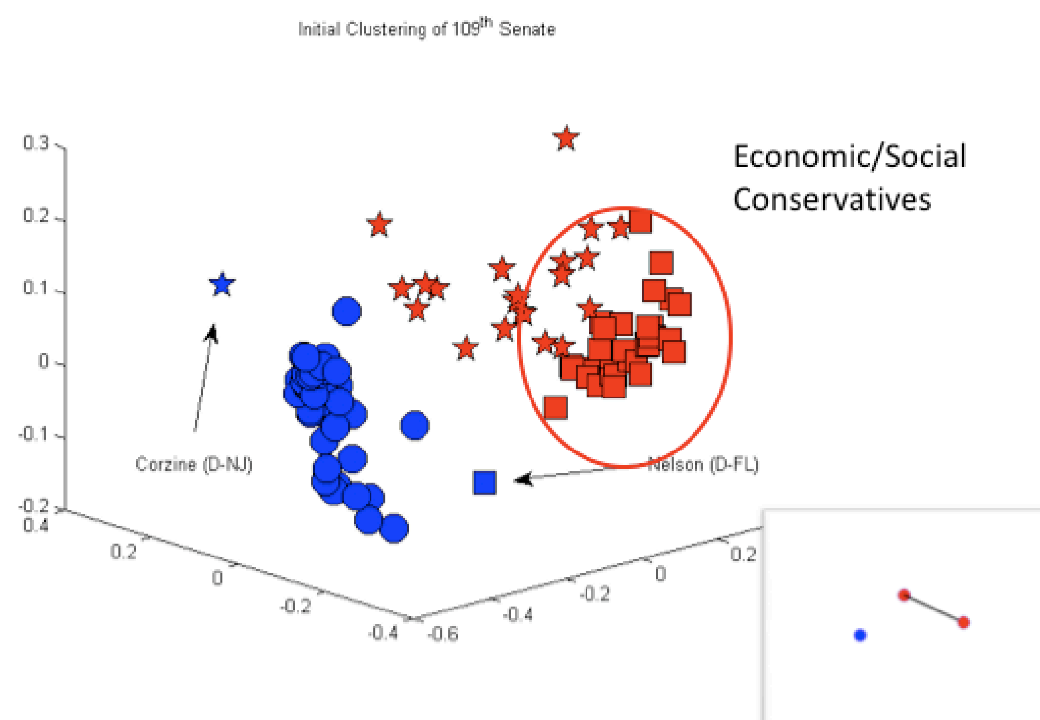
Remove unanimous votes, create similarity matrix using percent of votes in common.

Spectral cluster with three clusters









Validates well and
Looks like
“usual view”
(Poole-Rosenthal)