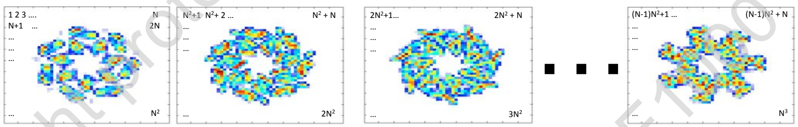
A hierarchical elastic network model for unsupervised EM density map segmentation

Virginia Burger[1,2], Ivet Bahar[2], and Chakra Chennubhotla[2]
[1]Joint CMU-Pitt Ph.D. Program in Computational Biology
[2]Department of Computational and Systems Biology, University of Pittsburgh School of Medicine
vmb12@pitt.edu bahar@pitt.edu chakracs@pitt.edu

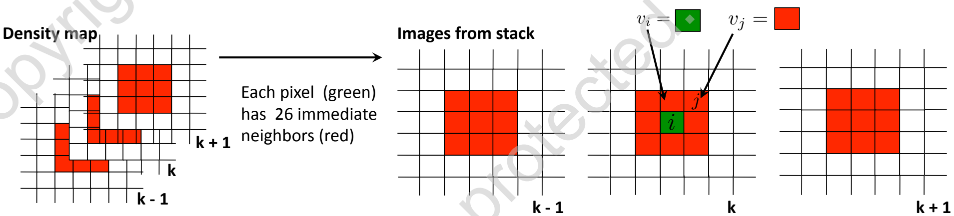
Hierarchical Elastic Network Model

We propose an unsupervised hierarchical elastic network model (hENM) based on a Markov diffusion process [1] to segment electron density maps into meaningful sub-regions, thereby highlighting different levels of structural and functional organization found in Cryo-EM density maps [2,3].

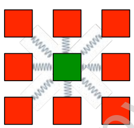
1 Consider each cross-section of the $N \times N \times N$ density map as an image with $N \times N$ pixels and a total of $n_{pixel} = N^3$ pixels.



Let $\vec{x}_i$ be the location of pixel i in Cartesian coordinates, and v_i be its intensity value.



2 Connect each pixel to its 26 neighbors with a spring.

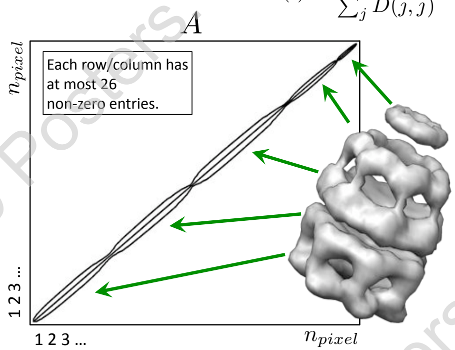


3 Give each spring an *affinity* (spring constant) to denote the likelihood of two pixels belonging to the same region [4,5]:

$$a_{ij} = \frac{1}{||\vec{x}_i - \vec{x}_j||} e^{-\beta(v_i - v_j)^2}$$

where β is proportional to the median of the distribution $|v_i - v_j|$.

4 Form symmetric $n_{pixel} \times n_{pixel}$ affinity matrix $A = (a_{ij})$, degree matrix D, Markov transition matrix M (see next step), and the stationary distribution $\vec{\pi}$ of the Markov chain: $\vec{\pi}(i) = \frac{D(i,i)}{\sum_j D(j,j)}$.



5 Hierarchically segment with hENM [6,7,8]:

Let $t = 0$, $A_0 = A$, $n_0 = n_{pixel}$, $D_0 = D$, and $\vec{\pi}_0 = \vec{\pi}$.

While $n_t > 1$:

$$D_{t-1}(i,j) = \begin{cases} \sum_{j=1}^{n_{t-1}} A_{t-1}(i,j) & i = j \\ 0 & i \neq j \end{cases}$$ degree matrix

$$M_{t-1} = A_{t-1} D_{t-1}^{-1}$$ Markov transition matrix

$$M_{t-1}^d = M_{t-1} \times M_{t-1}$$ diffusion

Choose K_t, a subset of n_t columns of M_{t-1}^d based on $\vec{\pi}_{t-1}$

$$\vec{\pi}_{t-1} = K_t \vec{\pi}_t$$ solve for next stationary distribution

$$A_t = \text{diag}(\vec{\pi}_t) K_t^T \text{diag}(K_t \vec{\pi}_t)^{-1} \text{diag}(\vec{\pi}_t)$$ next affinity matrix

$n_t =$ number of nodes in A_t

$t = t + 1$

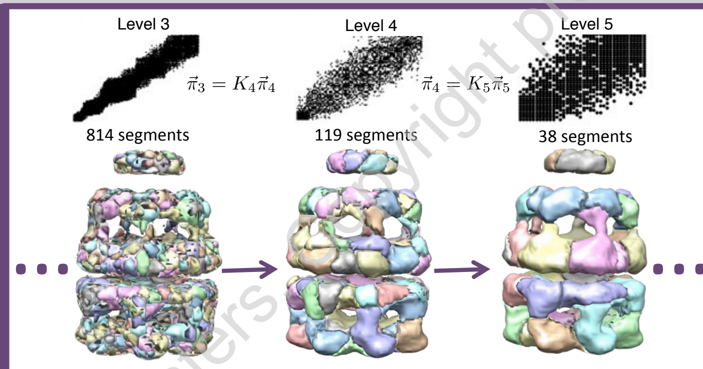
End.

6 Perform soft clustering at each level of hierarchy by computing $W_t(i,j)$, which gives the probability that pixel i belongs to segment j at level t of the hierarchy:

$$W_t(i,j) = \frac{K_t(i,j)\vec{\pi}_t(j)}{\sum_{k=1}^{n_t} K_t(i,k)\vec{\pi}_t(k)}.$$

References:
1. C. Chennubhotla, I. Bahar (2007) "Markov Methods for Hierarchical Coarse Graining of Large Protein Dynamics" Journal of Computational Biology 14:765-76.
2. F. Tama, O. Miyashita, C. L. Brooks III. (2004) "Flexible multi-scale fitting of atomic structures into low-resolution electron density maps with elastic network normal mode analysis" Journal of Molecular Biology 337: 985-999.
3. G. Pintilie, J. Zhang, T. Goddard, W. Chiu, D. Gossard (2010) "Quantitative analysis of cryo-EM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions" Journal of Structural Biology 170:427-438.
4. C. Chennubhotla, A. Jepson (2003) "Half-Lives of Eigenflows for Spectral Clustering" Advances in Neural Information Processing Systems (NIPS) 15, S. Becker, S. Thrun and K. Obermeyer eds, MIT Press, Cambridge, MA, 689-696.
5: L. Grady, M. Jolly (2008) "Weights and Topology: A Study of the Effects of Graph Construction on 3D Image Segmentation" Proceedings Medical Image Computing and Computer Assisted Intervention (MICCAI) 1: 153-161.
6. C. Chennubhotla, A. Jepson (2005) "Hierarchical Eigensolver for Transition Matrices in Spectral Methods" Advances in Neural Information Processing Systems (NIPS) 17, L. Saul, Y. Weiss and L. Bottou eds, MIT Press, Cambridge, MA, 273-280.
7. C. Chennubhotla, I. Bahar (2007) "Markov Methods for Hierarchical Coarse Graining of Large Protein Dynamics" Journal of Computational Biology 14(6), 765-776.
8. C. Chennubhotla, I. Bahar (2007) "Signal Propagation in Proteins and its Relation to Equilibrium Fluctuations" PLoS Computational Biology 3(9),1716-1726.
9.* F. Pettersen, T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, T. Ferrin (2004) "Chimera UCSF Chimera - a visualization system for exploratory research and analysis" Journal of Computational Chemistry 13:1605-12.
10.* S. Nickell, F. Forster, A. Linaroudis, W. Del Net, F. Beck, R. Hegerl, W. Baumeister, J. Plitzko (2005) "TOM software toolbox: acquisition and analysis for electron tomography" Journal of Structural Biology, 149(3):227-234.
* Used for visualization

e.g. Hierarchical Segmentation of 23.5Å GroEL+GroES map



Results: Segmentation of 2010 Cryo-EM Modeling Challenge Targets