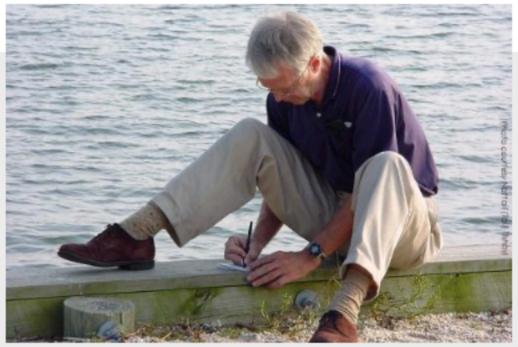
Mean Field Annealing for Graph Bisection

MFA for constrained optimization

- Mean field annealing
- Overviews
- Graph bisection problem
- Traveling salesman problem



John J. Hopfield

Howard A. Prior Professor in the Life Sciences, Professor of Molecular Biology, Associated Faculty in PNI, Emeritus

Email
hopfield@princeton.edu

Office
150 Princeton Neuroscience Institute

Related Links:

- Now What?
- Videos
- Curriculum vitae
- Publications

Related News

PNI's John Hopfield receives Nobel Prize in physics

Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 2554–2558, April 1982
Biophysics

Neural networks and physical systems with emergent collective computational abilities

(associative memory/parallel processing/categorization/content-addressable memory/fail-soft devices)

J. J. HOPFIELD

Division of Chemistry and Biology, California Institute of Technology, Pasadena, California 91125; and Bell Laboratories, Murray Hill, New Jersey 07974

Contributed by John J. Hopfield, January 15, 1982

ABSTRACT **Computational properties of use to biological organisms or to the construction of computers can emerge as collective properties of systems having a large number of simple equivalent components (or neurons). The physical meaning of content-addressable memory is described by an appropriate phase space flow of the state of a system. A model of such a system is given, based on aspects of neurobiology but readily adapted to integrated circuits. The collective properties of this model produce a content-addressable memory which correctly yields an entire memory from any subpart of sufficient size. The algorithm for the time evolution of the state of the system is based on asynchronous parallel processing. Additional emergent collective properties include some capacity for generalization, familiarity recognition, calized content-addressable memory or categorizer using extensive asynchronous parallel processing.**

The general content-addressable memory of a physical system

Suppose that an item stored in memory is "H. A. Kramers & G. H. Wannier *Phys. Rev.* **60**, 252 (1941)." A general content-addressable memory would be capable of retrieving this entire memory item on the basis of sufficient partial information. The input "& Wannier, (1941)" might suffice. An ideal memory could deal with errors and retrieve this reference even from the input "Vannier, (1941)". In computers, only relatively simple forms of content-addressable memory have been made in hard-

ABSTRACT Computational properties of use to biological organisms or to the construction of computers can emerge as collective properties of systems having a large number of simple equivalent components (or neurons). The physical meaning of content-addressable memory is described by an appropriate phase space flow of the state of a system. A model of such a system is given, based on aspects of neurobiology but readily adapted to integrated circuits. The collective properties of this model produce a content-addressable memory which correctly yields an entire memory from any subpart of sufficient size. The algorithm for the time evolution of the state of the system is based on asynchronous parallel processing. Additional emergent collective properties include some capacity for generalization, familiarity recognition, categorization, error correction, and time sequence retention. The collective properties are only weakly sensitive to details of the modeling or the failure of individual devices.

Complex Systems **2** (1988) 59–89

Neural Networks and NP-complete Optimization Problems; A Performance Study on the Graph Bisection Problem

Carsten Peterson *
James R. Anderson †
Microelectronics and Computer Technology Corporation,
3500 West Balcones Center Drive, Austin, TX 78759-6509, USA

Abstract. The performance of a mean field theory (MFT) neural network technique for finding approximate solutions to optimization problems is investigated for the case of the minimum cut graph bisection problem, which is NP-complete. We address the issues of solution quality, programming complexity, convergence times and scalability. Both standard random graphs and more structured geometric graphs

Abstract. The performance of a mean field theory (MFT) neural network technique for finding approximate solutions to optimization problems is investigated for the case of the minimum cut graph bisection problem, which is NP-complete. We address the issues of solution quality, programming complexity, convergence times and scalability. Both standard random graphs and more structured geometric graphs are considered. We find very encouraging results for all these aspects for bisection of graphs with sizes ranging from 20 to 2000 vertices. Solution quality appears to be competitive with other methods, and the effort required to apply the MFT method is minimal. Although the MFT neural network approach is inherently a parallel method, we find that the MFT algorithm executes in less time than other approaches even when it is simulated in a serial manner.

Peterson & Soderberg

- Carsten Peterson – Homepage
- Mean field annealing
- Hopfield Neural Networks

MFA for constrained optimization

Problem modeling

A mathematical framework
- An objective function
- Constraints

Energy function

Drive an discrete energy function

Mean field equations

Interactive dynamics for minimizing the energy function

Software or hardware implementation

Graph bisection

- mean field annealing
- MFA Optimization

JM Wu, MFA Optimization For Graph Bisection Problem

Abstract

This work derives the mean field approximation to the mean configuration of a stochastic Hopfield neural network under the Boltzmann assumption. The new approximation is realized by two sets of interactive mean field equations, respectively estimating mean activations subject to mean correlations and mean correlations subject to mean activations. The two sets of interactive dynamics are derived based on two dual mathematical frameworks. Each aims to optimize the objective quantified by a combiation of the Kullback-Leibler (KL) divergence and the correlation strength between any two distinct fluctuated variables subject to fixed mean correlations or activations. The new method is applied to the graph bisection problem. By numerical simulations, we show that the new method effectively improves in both performance and relaxation efficiency against the naive mean field equation

Travelling Salesman Problem

- MFA optimization

  PottsTSP.pdf
  JM Wu, Potts models with two sets of interactive dynamics

Abstract

In this work, we develope Potts models with two sets of interactive dynamics. We derive the mean field annealing for a new energy function which contains two sets of neural variables, one for combinatorial constraints and the other for internal geometrical representations. Two sets of interactive dynamics are further developed for both sets of neural variables. The obtained Potts model thus possesses two sets of dynamics and exactly fits the requirement of the hairy model proposed by Hzu. We call the new Potts model as hairy Potts neural network. The elastic ring method proposed by Durbin and Willshaw and the Potts model of Peterson and Söderberg, and the Hopfield's TSP modeling are shown to be special forms of the hairy Potts neural network. We explore various energy functions for the hairy Potts neural network and test the new network with simulations. The results are encouraging.

Traveling salesman problem

- Mean field annealing
  - Spin models (Hopfield & Tank)
  - Potts models (Peterson & Soderberg)
- Simulated annealing (Kirkpatrick)
- Elastic ring (Durbin & Willshaw)
- Self-organization (Kohonen)

Fundamental tasks

- Combinatorial Optimization
  - Graph bisection
  - Traveling salesman problem
  - Scheduling
- Unsupervised learning
  - Clustering analysis
  - Principle component analysis
  - Self-organization
    - Kohonen Self-organization Map
    - Elastic nets

Fundamental tasks

- Supervised learning
  - Classification
  - Function approximation
  - Recursive function approximation
- Independent component analysis
- Blind source separation
- Density estimation and conditional density estimation

Integer Programming

- Integer programming
  - Discrete variables
  - Methods: Mean field annealing, simulated annealing

Unconstrained optimization

- Continuous variables
- Gradient based approaches:
  - gradient method
  - Newton method
  - Newton-Gauss method
  - Levenberg-Marquardt method

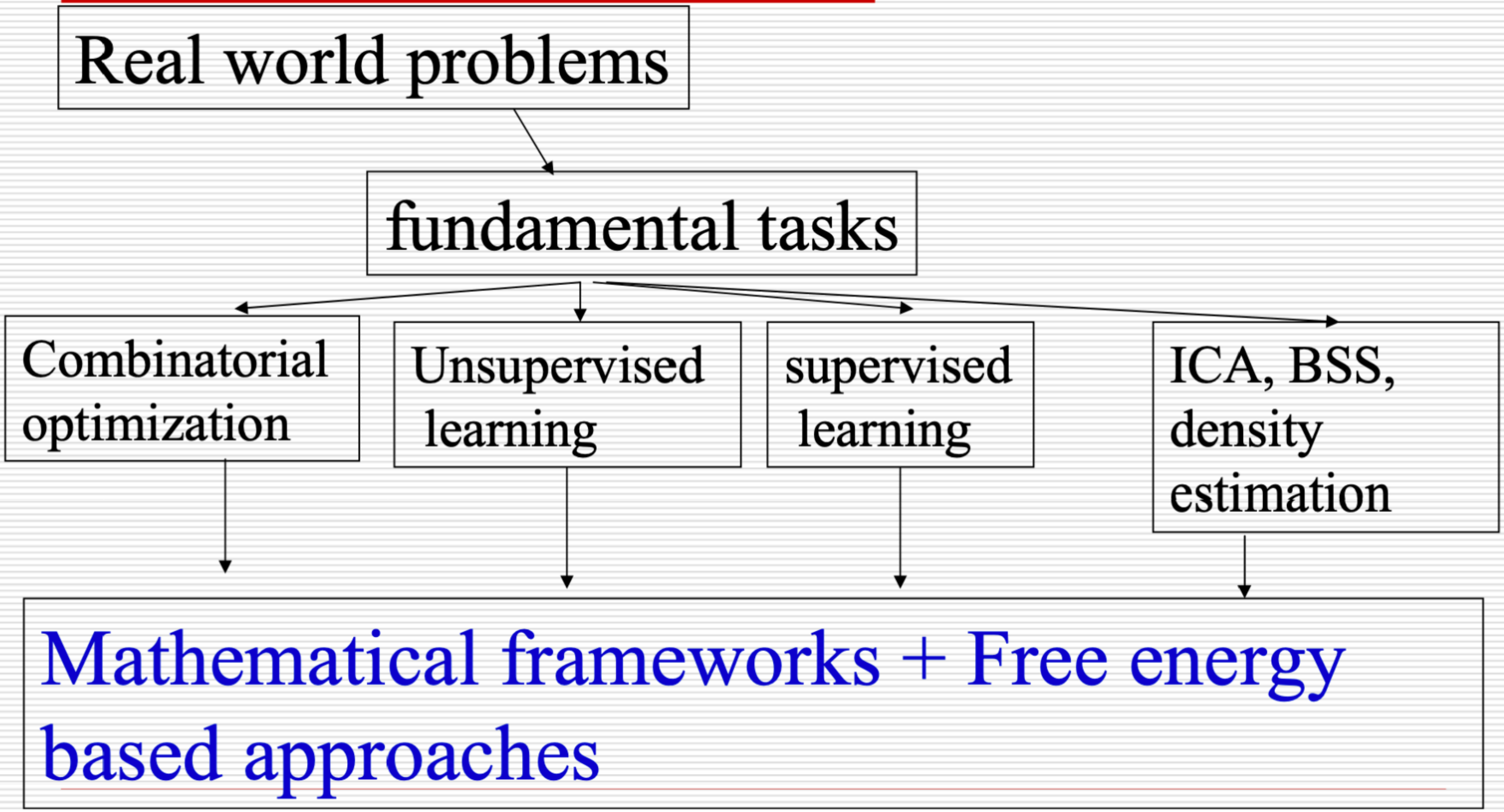
Mixed integer programming

- Continuous and discrete variables
- Method:
  - A hybrid of mean field annealing and gradient descent method
  - Free-energy based learning
  - Annealed Free-energy based learning

Hopfield Neural Networks

- Solving Combinatorial optimization
- Methods
  - mean field annealing

Free Energy Based Approach



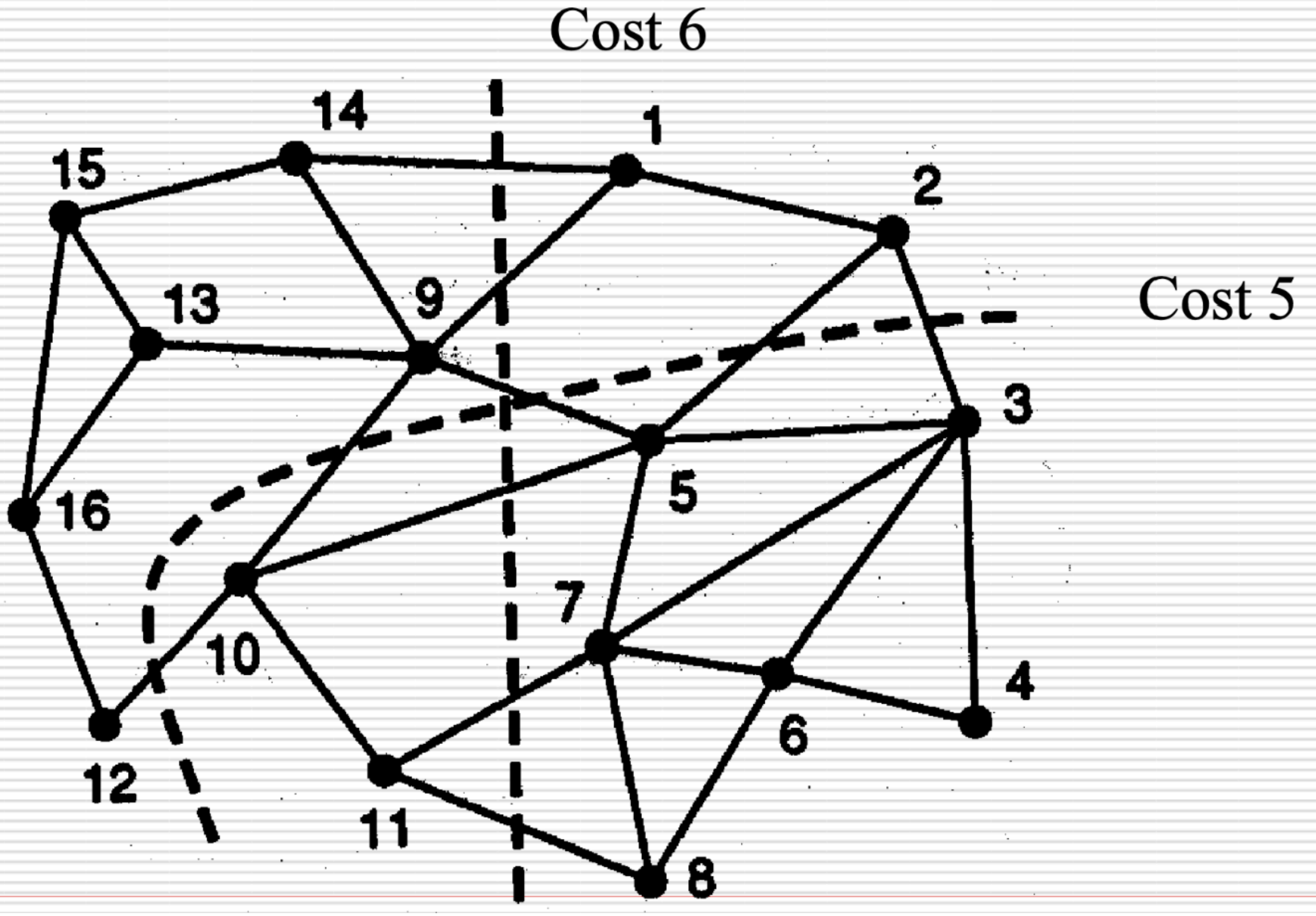
Graph bisection

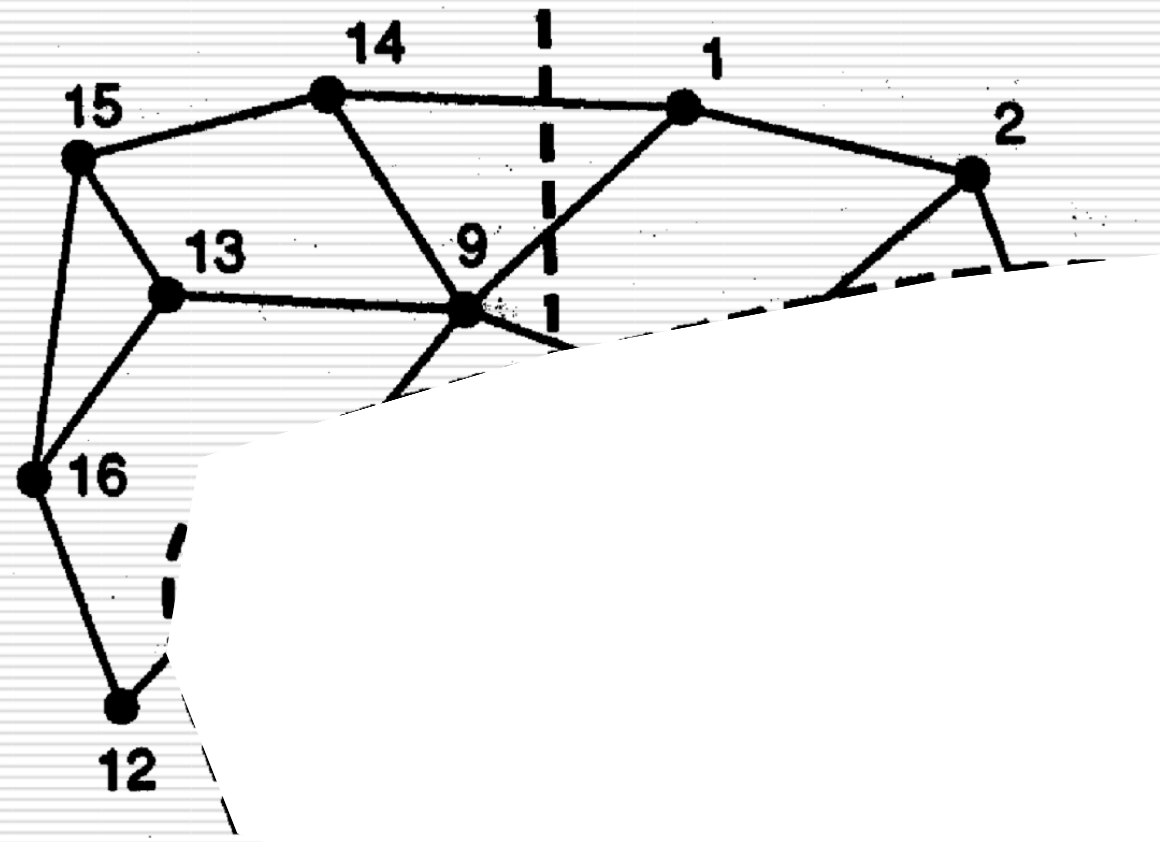
- Problem statement
- Mathematical framework
- Methods
- Numerical results

Graph generation

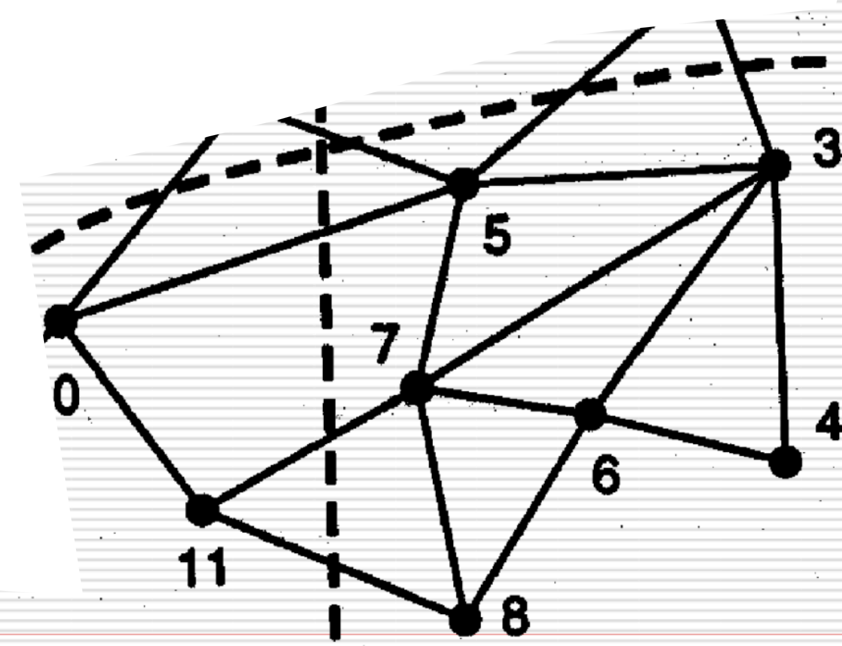
```
T = zeros(N,N);
for i = 1:N
for j = 1:N
        if rand(1,1) > 0.5 & j~=i
                T(i,j) = 1;
                T(j,i)=1;
        end
end
end
```

Graph Bipartition





Cost 5



Cost 5

Graph bisection

- Operation: partition all nodes to two sets
- Objective
  - Minimization of cut size
  - Cut size means the number of connections between two sets
- Constraints
  - Two sets have equal size

Representations of edges and memberships

- $G_{ij} = 1$ if vertices i and j are connected
- $G_{ij} = 0,$ if they are not connected
- Define a variable S_i at each vertex
- $S_i = 1$, if site i is in one set and $S_i = -1$ if it is in the other

Mathematical framework

$C_{ij} = G_{ij}$

- Minimize. $$L = -\sum_{i \neq j} C_{ij} S_i S_j$$
- Subject to $$\sum_i S_i = 0$$

S_i	S_j	C_{ij}	$-S_i C_{ij} S_j$
1	0	1	0
1	0	0	0
1	1	1	-1
1	1	0	0

S_i	S_j	C_{ij}	$-S_i C_{ij} S_j$
1	-1	1	1
1	-1	0	0
1	1	1	-1
1	1	0	0

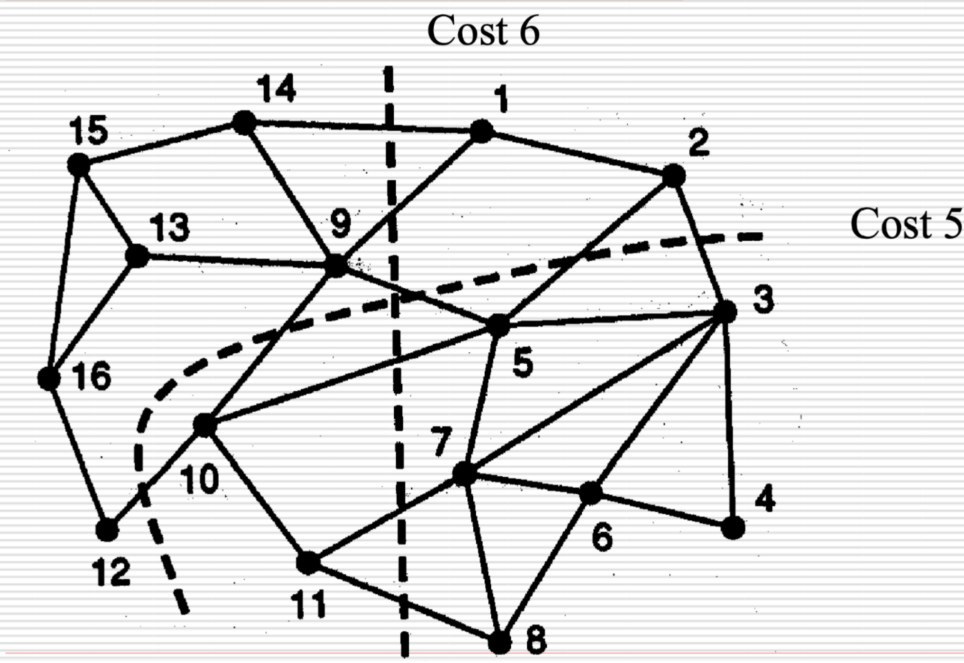
$$H(S) = -\sum_{i \neq j} C_{ij} S_i S_j + \mu\left(\sum_i S_i\right)^2$$

$$H(S) = N\mu - \sum_{i \neq j} w_{ij} S_i S_j$$

$$\frac{\partial H}{\partial S_i}$$

$$w_{ij} = C_{ij} - 2\mu$$

Graph Bipartition



Spin model

$S_i \in \{+1, -1\}$

2^N

2^{10000}

- Graph bisection

$$E(S) = -\sum_{i=1}^{N} \sum_{j \neq i}^{N} w_{ij} S_i S_j$$

$$\min_{\{S\}} E(S)$$

A physical-like random system

- Boltzmann assumption
  - S is regarded as a random vector

$$\Pr(S) \propto \exp(-\beta E(S))$$

- Free energy

$$F = \langle E(S) \rangle - \frac{1}{\beta} H(S)$$

Entropy

- Entropy of the whole system

$$|\{S\}| = 2^N \qquad H(S) = -\sum_{\{S\}} \Pr(S) \log \Pr(S)$$

Computationally intractable

- Sum of individual entropies

$$H(S) \approx \sum_i \mathrm{H}(\mathrm{S_i})$$

Computationally tractable

Mean field approximation

- Individual entropy

$$H(S_i) = -\sum_{S_i=\pm 1} \Pr(S_i) \log \Pr(S_i)$$

$$Pr(S_i) = Ce^{\beta u_i S_i}$$

$$\Pr(S_i) \propto \exp(\beta u_i S_i)$$

$$\Pr(S_i) = \frac{\exp(\beta u_i S_i)}{\exp(\beta u_i) + \exp(-\beta u_i)}$$

$$Pr(S_i = 1) + Pr(S_i = -1) = Cexp(\beta u_i) + Cexp(-\beta u_i)) = 1$$

$$C = \frac{1}{exp(\beta u_i) + exp(-\beta u_i))}$$

$$\Pr(S_i) \propto \exp(\beta u_i S_i)$$

$$\Pr(\mathrm{S_i}) = \frac{\exp(\beta u_i S_i)}{\exp(\beta u_i) + \exp(-\beta u_i)}$$

$$< S_i > = Pr(S_i = 1) - Pr(S_i = -1)$$

$$1 = Pr(S_i = 1) + Pr(S_i = -1)$$

$$\langle S_i \rangle = \frac{\exp(\beta u_i) - \exp(-\beta u_i)}{\exp(\beta u_i) + \exp(-\beta u_i)} = \tanh(\beta u_i)$$

$$\Pr(\mathrm{S_i} = 1) = \frac{\langle S_i \rangle + 1}{2}, \Pr(\mathrm{S_i} = -1) = \frac{1 - \langle S_i \rangle}{2}$$

$$\langle S_i \rangle = \frac{\exp(\beta u_i) - \exp(-\beta u_i)}{\exp(\beta u_i) + \exp(-\beta u_i)}$$

$$\Pr(\mathrm{S_i} = 1) = \frac{\langle S_i \rangle + 1}{2}, \Pr(\mathrm{S_i} = -1) = \frac{1 - \langle S_i \rangle}{2}$$

$$\Downarrow$$

$$\mathrm{H(S_i)} = -\sum_{S_i = \pm 1} \Pr(S_i) \log \Pr(S_i)$$

$$= -\Pr(S_i = 1) \log \Pr(S_i = 1) - \Pr(S_i = -1) \log \Pr(S_i = -1)$$

$$= -\frac{\langle S_i \rangle + 1}{2} \log \frac{\langle S_i \rangle + 1}{2} - \frac{1 - \langle S_i \rangle}{2} \log \frac{1 - \langle S_i \rangle}{2}$$

Mean field approximation

$$F = \langle E(S) \rangle - \frac{1}{\beta} H(S)$$

$$\approx E(\{\langle S_i \rangle\}) - \frac{1}{\beta} \sum_i H(S_i)$$

$$= -\sum_{i=1}^{N} \sum_{j \neq i}^{N} w_{ij} \langle S_i \rangle \langle S_j \rangle - \frac{1}{\beta} \sum_i H(S_i)$$

$$\mathrm{F} \approx -\sum_{i=1}^{N}\sum_{j\neq i}^{N} w_{ij}\langle S_i\rangle\langle S_j\rangle$$

$$+\frac{1}{\beta}\sum_{i}\left\{\frac{\langle S_i\rangle+1}{2}\log\frac{\langle S_i\rangle+1}{2}+\frac{1-\langle S_i\rangle}{2}\log\frac{1-\langle S_i\rangle}{2}\right\}$$

$$v_i \equiv \langle S_i\rangle$$

$$\frac{1}{2}\log\frac{1+\langle s_i\rangle}{2}{+1}-\frac{1}{2}\log\frac{1-\langle s_i\rangle}{2}{-1}$$

$$\frac{\partial \mathrm{F}}{\partial v_i} = -\sum_{j\neq i}^{N} w_{ij}v_j + \frac{1}{\beta}\tanh^{-1}(v_i) = 0$$

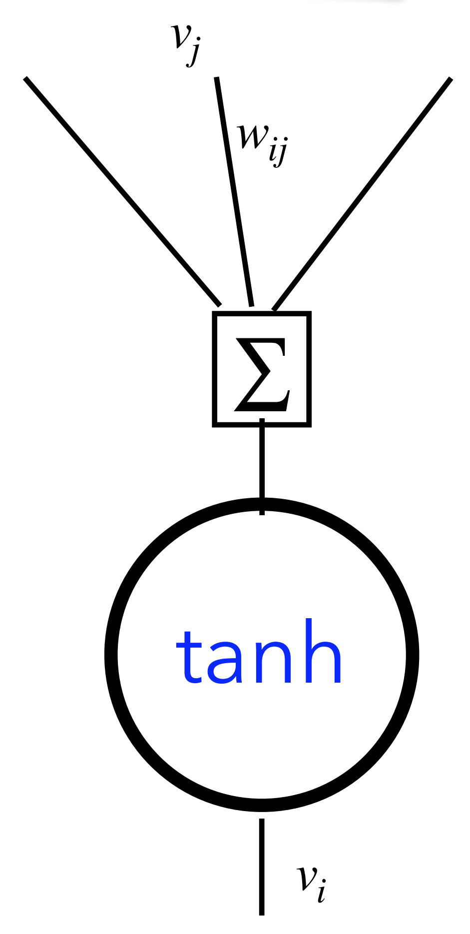
$$v_i = \tanh(\beta\sum_{j\neq i}^{N} w_{ij}v_j)$$

Mean field equation

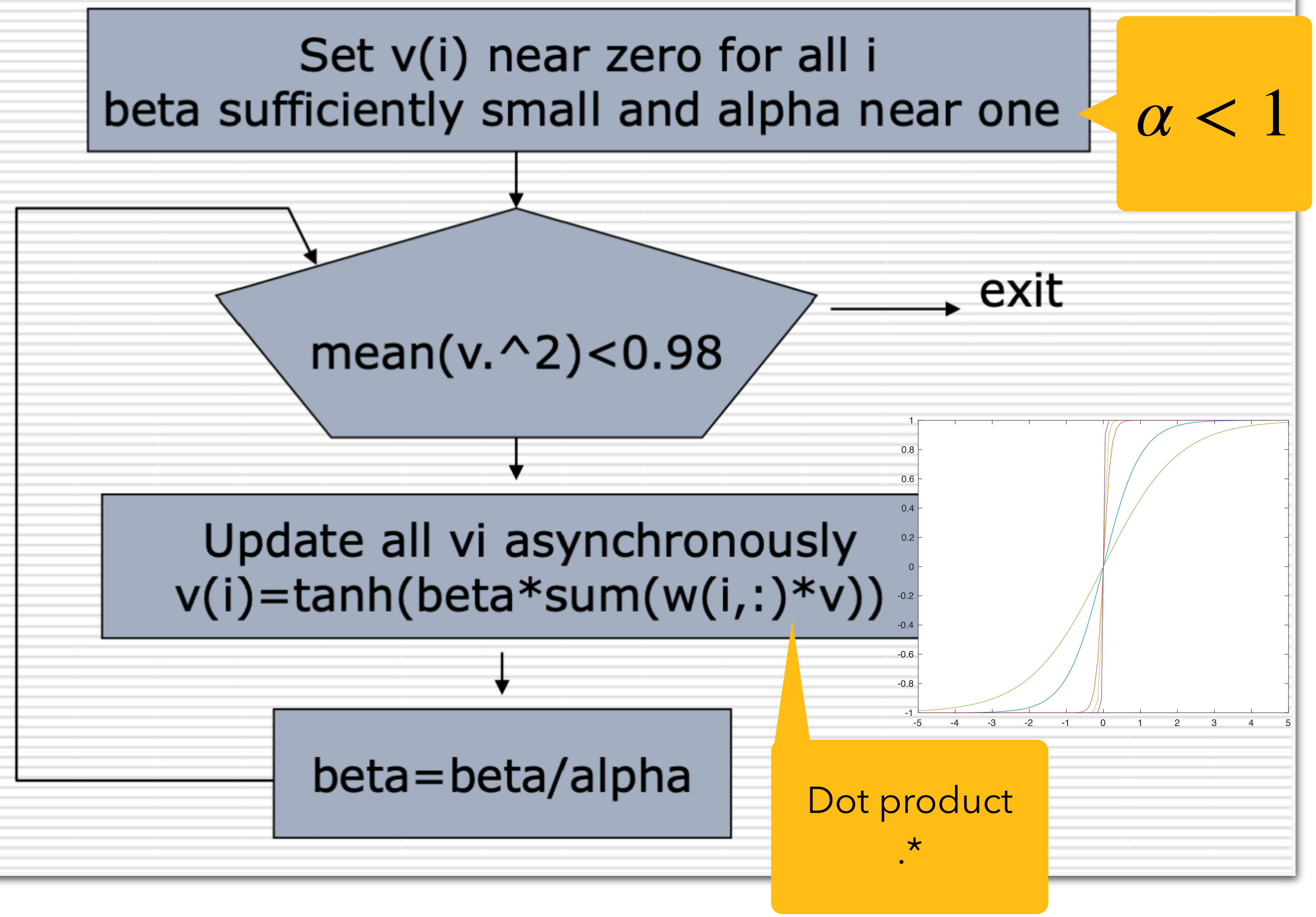
$$v_i = \tanh(\beta \sum_{j \neq i}^{N} w_{ij} v_j)$$

$$u_i = \sum_{j \neq i}^{N} w_{ij} v_j$$

$$v_i = tanh(\beta u_i)$$



Flow chart



Set v(i) near zero for all i
beta sufficiently small and alpha near one
$\alpha < 1$
mean(v.^2)<0.98
exit
Update all vi asynchronously
v(i)=tanh(beta*sum(w(i,:)*v))
beta=beta/alpha
Dot product
.*

Halting condition

- Mean of squares of all $<s_i>$ exceeds a predetermined threshold value

$$v \text{ is a vector that collects all } \langle \mathrm{s_i} \rangle$$

$$\text{halting condition :}$$

$$\text{mean}(\text{v.\^{}2}) > 0.99$$

Annealing schedule

- Set beta to a sufficiently small value
- Increase beta carefully

Exercise

1. Implement MFA for solving graph bisection
2. How to improve MFA for constrained optimization?

Multilayer Neural Network

- Fundamental Tasks
  - Data driven classification
  - Data driven function approximation
- Methods
  - Gradient method
  - Newton-Gauss method
  - Leveberg-Marduardt method

KL(Kullback-Leberler) Divergence

- Boltzmann distribution

$$P_x(x) \propto \exp(-\beta H_x(x))$$

- Normalization

$$P_x(x) = \frac{\exp(-\beta H_x(x))}{Z_x}$$

Intractable Partition function

$$Z_x = \sum\nolimits_{\{x\}} \exp(-\beta H_x(x))$$

Factorial distribution

- An approximation

$$Q_x(x) = \prod_i q_i(x_i)$$

KL divergence

Cross entropy

- Semi-distance between two pdfs

$$\mathrm{KL}(Q_x \,||\, P_x) = \sum_{\{x\}} Q_x(x) \ln \frac{Q_x(x)}{P_x(x)}$$

$$\mathrm{KL}(Q_x \,||\, P_x) = \sum_{\{x\}} Q_x(x) \ln \frac{Q_x(x)}{P_x(x)}$$

Step 1

$$\mathrm{KL}(Q_x \,||\, P_x) = \sum\nolimits_{\{x\}} Q_x(x) \ln Q_x(x) + \beta \sum\nolimits_{\{x\}} Q_x(x) H_x(x) + \ln Z_x.$$

$$P_x(x) = \frac{\exp(-\beta H_x(x))}{Z_x}$$

Step 2

$$S[Q_x] \triangleq -\sum_{\{x\}} Q_x(x) \ln Q_x(x)$$
$$= -\sum_i \left(\frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right)$$

Step 3

$$\text{and} E_{Q_x}[H(x)]$$
$$\triangleq \sum_{\{x\}} Q_x(x) H_x(x)$$
$$= -\frac{1}{2} \sum_{i,j \neq i} J_{ij} m_i m_j.$$

$$F \approx -\sum_{i=1}^{N}\sum_{j\neq i}^{N} w_{ij}\langle S_i\rangle\langle S_j\rangle$$

$$+\frac{1}{\beta}\sum_{i}\left\{\frac{\langle S_i\rangle+1}{2}\log\frac{\langle S_i\rangle+1}{2}+\frac{1-\langle S_i\rangle}{2}\log\frac{1-\langle S_i\rangle}{2}\right\}$$

Step 4

$$\begin{aligned} F(m) &\triangleq -S[Q_x] + \beta E_{Q_x}[H(x)] \\ &= \sum_i \left(\frac{1+m_i}{2}\ln\frac{1+m_i}{2} + \frac{1-m_i}{2}\ln\frac{1-m_i}{2} \right) \\ &\quad - \frac{\beta}{2}\sum_{i,j\neq i} J_{ij} m_i m_j. \end{aligned} \tag{10}$$