Interaction Models of Probabilistic Networks Suggested by Statistical Mechanics

(Quantitative Methods in Defense and National Security 2007)

John E. Gray
Code Q-23
Electromagnetic & Sensor Systems Department
Naval Surface Warfare Center Dahlgren Division,
Dahlgren, VA 22448

2/8/2007

Outline

1. Introduction

2. Discrete Probability Notation (Optional)

3. Maximum Entropy Procedure
   - 3.1. Example Normalization Only (Optional)
   - 3.2. Partition Function (Optional)
   - 3.3. Maximum Entropy Applied to Random Graphs
   - 3.4. Graph Entropy
   - 3.5. General Model
   - 3.6. General Model Observations
   - 3.7. Graph Characteristics (Optional)

4. Concluding Comments

1 Introduction

- In its most elemental form, physics of the individual is governed by Newton's Laws, change in motion is determined by determining the forces acting on a the individual (mass)

$$\overrightarrow{F} = m\overrightarrow{a}. \tag{1}$$

- Two approaches to solving this differential equations: data based or theory based.

- The theory based approach is to explicitly formulate the force in terms potentials, $\overrightarrow{F} = -\nabla V(x)$, and then solve the differential equation

$$\frac{d^2\overrightarrow{x}}{dt^2} + \frac{\nabla V(x)}{m} = 0, \tag{2}$$

one then solves this ordinary differential equation to get $\overrightarrow{x} = \overrightarrow{x}(t)$ which solves the **direct problem of individual behavior** for the particular mass. When

one has observational data about the trajectory, one has $\overrightarrow{x}_i = \overrightarrow{x}(t_i)$ which often allows one to infer a particular trajectory $\overrightarrow{x}(t)$.

- Newton solved many of the interesting problems that related trajectory class to potential that could produce them, but lacked the mathematics to solve the inverse problems.

- For two body interaction, the situation became more complicated. Failure of the analytical approach for the few body problem lead to another approach to be formulated for many body interactions.

- Methods for dealing with collective behavior of large number of interacting particles, statistical mechanics, allows prediction of statistical properties of a large number of interacting units based on knowledge of the individual units two body interactions.

- Generalized notion of partition function of the interaction

2 Discrete Probability Notation (Optional)

- A collection of numbers p_i that collectively satisfy the two conditions of

$$p_i \geq 0 \qquad \text{(Positivity)}$$

and

$$\sum_{i=1}^{n} p_i = 1. \qquad \text{(Normalization)}$$

are termed probabilities for state i.

DEFINITION: A column vector $|p\rangle$ is a *probability vector* if its components satisfy both the positivity and normalization conditions

$$|p\rangle = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ . \\ . \\ p_n \end{bmatrix} \qquad \text{(probability vector)}$$

DEFINITION: The *transpose* of a probability vector is the row vector $\langle p|$ which obeys (^{t} denotes transpose) $\langle p| = |p\rangle^t$.

DEFINITION: The inner product of two vectors $|p\rangle$ and $|q\rangle$ is

$$\langle p|q\rangle = \sum_{i=1}^{n} p_i q_i \qquad \text{(inner product)}$$

DEFINITION: The *characteristic vector* $|e_i\rangle$ is defined as

$$|e_i\rangle = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ . \\ . \\ 0 \end{bmatrix} \right. - \text{i}^{th} \text{ row}$$

DEFINITION: The *probability of a state* i of a probability vector $|p\rangle$ is

$$p_i = \langle p|e_i\rangle$$

DEFINITION: The *mean* of a probability vector $|p\rangle$ is

$$\overline{p} = \langle i \rangle = \langle p|i\rangle \,. \qquad \text{(mean)}$$

DEFINITION: The *variance* of a probability vector $|p\rangle$ is

$$\sigma_p^2 = \left\langle p|i^2 \right\rangle - \langle p|i\rangle^2 \,. \qquad \text{(variance)}$$

DEFINITION: The *n-th moment* is

$$\langle i^n \rangle = \langle p|i^n\rangle \qquad \text{(n-th moment)}$$

DEFINITION: The *expected value* of a vector $|f(x)\rangle$ is

$$\langle f(x)\rangle = \langle p|f(x)\rangle = \sum_{i=1}^{n} p_i f(x_i) \quad \text{(expected value)}$$

3 Maximum Entropy Procedure (Optional)

DEFINITION: The "surprise value" of a probability vector $|p\rangle$ (introduced by Hartley) is -$|\ln p\rangle$. (Note this definition is consistent with our intuition, the less probable a state i occurs, the more "surprised" we are when it does.

DEFINITION: Degree of uncertainty (which is equivalent to the surprise value) in the information is defined as

$$S[|p\rangle\,; n] = -k \sum_{i=0}^{n} p_i \ln p_i = -k \langle p| \ln p\rangle\,; \quad \text{(Entropy)}$$

- Most applications do not permit one to measure the probabilities p_i associated with a physical variable $f(x_i)$, instead the expected value $\langle f(x)\rangle$ is measured.

- Probabilities are connected to expected values by noting the formula

$$\langle f(x)\rangle = \sum_{i=0}^{n} p_i f(x_i) = \langle p|f(x)\rangle$$

(Expected Value)

- Goal to find an assignment to the probability vector $|p\rangle$ subject to available information.

- Problem is to find an assignment of probabilities that maximizes the entropy. For example, we always know

  $|p\rangle$ is normalized

$$\langle 1|p\rangle = \sum_{i=0}^{n} p_i = 1, \qquad \text{(Condition 1)}$$

  Known assignment of the means ($r = 1, 2, 3...m$)

$$\langle p|g_r\left(|x\rangle\right)\rangle = \langle g_r(|x\rangle)\rangle = \sum_{i=1}^{n} p_i g_r(x_i) = \alpha_r.$$

(Condition 2)

- Method of Lagrange undetermined multipliers is a standard method for solving an optimization problem subject to constraints.

  The function one wants to maximize is the entropy, so it is adjoined to the free parameters (Lagrange undetermined multipliers) times the equations for the constraints to form a **Lagrangian**.

  These two condition the principle "*The distribution* $|p\rangle$ *that maximizes the uncertainty in the expected value subject to the constraint of the available information*" provides such an assignment that gives the probability assignment.

3.1 Example: Normalization Only (Optional)

- **Example (assignment of probabilities with no known information):** Lagrangian for entropy maximization is

$$L = -\langle p|\ln p\rangle - \lambda_0\left(\langle 1|p\rangle - 1\right)$$

1. To minimize with respect to the probabilities, one minimizes the Lagrangian with respect to a particular state $p_i = \langle p|e_i\rangle$, so computing $\frac{\partial L}{\partial\langle p|e_i\rangle}$gives

$$\frac{\partial L}{\partial\left\langle p|e_i\right\rangle} = 0 = \ln p_i + 1 + \lambda_0 \qquad (3)$$

2. Solving for p_i gives

$$p_i = \langle p|e_i\rangle = e^{-(\lambda_0+1)}$$

(probability assignment)

3. Substituting probabilities into the normalization

condition

$$\sum_{i=1}^{n} e^{-(\lambda_0+1)} = ne^{-(\lambda_0+1)} = \sum_{i=1}^{n} \langle p|e_i\rangle = 1. \tag{4}$$

4. Lagrange undetermined multiplier as

$$e^{-(\lambda_0+1)} = \frac{1}{n}. \tag{5}$$

5. Thus the least informative assignment of the probabilities when only the normalization condition is known is the assignment that the probabilities are uniformly distributed:

$$p_i = \langle p|e_i\rangle = \frac{1}{n}. \tag{6}$$

3.2 Partition Function (Optional)

- Define the partition function $Z(\mu)$, μ is a conveniently chosen scaling parameter or the inverse temperature for a physical system, for a collection of data $\{f(x_i)\}$ with a probability vector $|p\rangle$ as

$$Z(\mu) = \sum_i \exp\left(-\mu p_i f(x_i)\right) = \exp\left(-\mu \langle p|f(x)\rangle\right). \tag{7}$$

- Take first derivative of $Z(\mu)$ with respect to μ

$$-\frac{\partial}{\partial \mu}[Z(\mu)] = \langle p|f(x)\rangle\, Z(\mu), \tag{8}$$

- Rearrangement of this expression gives the definition of the expected value

$$-\frac{1}{Z(\mu)}\frac{\partial}{\partial \mu}[Z(\mu)] = \langle f\rangle \tag{9}$$

which is the definition of mean of the data set $\{f(x_i)\}$.

- Definition of the **free energy** as

$$F_\mu = -\ln\left(Z(\mu)\right), \tag{10}$$

so

$$\langle f \rangle = \frac{dF_\mu}{d\mu} \tag{11}$$

- Expression for **variance** is the second derivative of the free energy:

$$\frac{d^2 F_\mu}{d\mu^2} = \left\langle f^2 \right\rangle - \langle f \rangle^2 . \tag{12}$$

- Variance, σ_f^2, can be used to define the **heat capacity** of the network function f is defined as

$$C_f = \mu^2 \sigma_f^2. \tag{13}$$

- Note, there are a variety of physics based variables that can be defined as derivatives of the partition function including pressure, average energy, capacity, temperature, entropy, etc.

3.3 Maximum Entropy Applied to Random Graphs

3.3.1 Random Graph

- Random graphs date back to the seminal work by Erdos and Renyi.

- Random graph models have been interpreted as ensemble models that reflect assumptions about the PDF's over a collection of networks.

- Specific graph G is drawn from a collection of graphs $\mathcal{G}$ and one can associated a probability $P(G)$ with that graph.

- $P(G)$ is such that expected value of our graph observable $\{x_i\}$ is equal to the expected value.

- Vastly undetermined problem since the PDF has many more degrees of freedom than that obtained by knowledge of a specific instance of a graph.

- Random graph can be defined in a variety of fashions with the randomness use to condition some aspect of the graph.

- Assume a fixed set of vertices with a rule for assigning edges to the vertices.

- Graph with vertex set $V = \{1, 2, ..., n\}$ and edge state $A = \{i : X(i), i = 1, 2..., n\}$ where $X(i)$ are independent random variables with a probability rule for assignment of an edge ($j = 1, 2, ..., n$):

$$P\{X(i) = j\} = f(j) \tag{14}$$

3.4 Graph Entropy (Introduction)

- One can define the **Gibbs entropy** of a graph as

$$S^{\mathcal{G}} = -\sum_{G\in\mathcal{G}} P(G)\ln(P(G)) \tag{15}$$

subject to the two constraints:

$$\sum_{G\in\mathcal{G}} P(G) = 1, \tag{16}$$

and

$$\sum_{G\in\mathcal{G}} P(G)x_i(G) = \langle x_i\rangle\,. \tag{17}$$

- Using MEP formalism gives

$$\ln(P(G)) + 1 + \lambda_0 + \sum_{i=1}^{n}\lambda_i x_i(G) = 0; \tag{18}$$

- Probability assignment for graph

$$P(G) = \frac{\exp(-H(G))}{Z(G)}, \tag{19}$$

- The **Hamiltonian of the graph** is

$$H(G) = \sum_{i=1}^{n} \lambda_i x_i(G), \tag{20}$$

- The **Partition Function of the graph** is

$$Z(G) = \exp(1 + \lambda_0) = \sum_{G} \exp(-H(G)) \tag{21}$$

- Interpretation of an instance of a graph G is that is a sample drawn from an ensemble $\mathcal{G}$.

- Definition equivalent to a physical system defined over its microstate, provided the temperature is finite.

- It is always possible to compute expected value of any graph property x over the particular model as

$$\langle x \rangle = \sum_{G} P(G)x(G) \tag{22}$$

- Definition of the Hamiltonian is a way of specifying the interaction model for our network.

- *Do not need to compute the statistical averages over the graph probability distributions directly, instead can work from definition of the partition function of our network.*

3.4.1 Model 1

- The simplest non-trivial model is when adjacency matrix $A(G)$ has components a_{ij}

$$a_{ij} = \begin{cases} 1 & \text{if } i \text{ is connected to } j \\ 0 & \text{if } i \text{ is not connected to } j \end{cases} \tag{23}$$

- The number of edges m is

$$m(G) = \sum_{i<j} a_{ij}(G) \tag{24}$$

- Hamiltonian is

$$H(G) = \mu \sum_{i<j} a_{ij}(G) \quad \text{(Hamiltonian Model 1)}$$

- The partition function (PF) is

$$\begin{aligned} Z_\mu(G) &= \sum_G \exp(-H(G)) \\ &= \sum_{a_{ij}} \exp(-\mu \sum_{i<j} a_{ij}(G)) \\ &= \prod_{i<j} \sum_{a_{ij}=0}^{1} \exp(-\mu a_{ij}(G)) \\ &= \prod_{i<j} (1+\exp(-\mu)) \\ &= (1+\exp(-\mu))^{\binom{n}{2}} \text{ (PF Model 1)} \end{aligned}$$

- Model 1 free energy is

$$F_\mu^{(\text{Model 1})} = -\ln\left(Z_\mu^{(1)}(G)\right) = -\binom{n}{2}\ln(1+\exp(-\mu)), \quad \text{(Free Energy)}$$

- Mean is

$$\langle f\rangle^{(\text{Model 1})} = \frac{dF_\mu}{d\mu} = \frac{\binom{n}{2}}{(1+\exp(\mu))}, \quad \text{(Mean)}$$

- Standard deviation is

$$\sigma_f^2\,(\text{Model 1}) = -\frac{\binom{n}{2}\exp(\mu)}{(1+\exp(\mu))^2},$$

(Standard Devivation)

- Heat capacity is

$$C_f^{(\text{Model 1})} = -\mu^2\frac{\binom{n}{2}\exp(\mu)}{(1+\exp(\mu))^2}.$$

(Heat Capacity)

3.4.2 Model 1 Equivalent probability model

- Instead of interpreting $\lambda_i x_i(G)$ as the Lagrange undetermined multiplier times some property of the graph,

- Can be interpreted as interaction potential between components i and j.

- The λ_i interpreted as scaling parameter μ, a field coupling parameter, or the inverse temperature.

- Possible to posit a variety of interaction models, work out their consequences, work out the equivalent probability distributions.

- To compare the interaction model to the underlying graph probability model, we reparameterize p as

$$p = \frac{1}{(1 + \exp(\mu))}, \tag{25}$$

so

$$\exp(\mu) = \left(\frac{1-p}{p}\right) \tag{26}$$

and therefore

$$p = \frac{\exp(-\mu)}{(1+\exp(-\mu))} \tag{27}$$

Thus, solving for $(1-p)$

$$p\exp(\mu) = (1-p) = \frac{1}{(1+\exp(-\mu))}. \tag{28}$$

- The first moment is

$$\langle f \rangle^{(1)} = \binom{n}{2} p, \tag{29}$$

- The standard deviation is

$$\sigma_f^2(1) = \binom{n}{2}\left(p - p^2\right); \tag{30}$$

- Heat capacity is

$$C_f^{(1)} = \binom{n}{2} \left(\ln \left(\frac{1}{p} - 1 \right) \right)^2 \left(p - p^2 \right). \qquad (31)$$

- Partition function is

$$\begin{aligned} Z_\mu(G) &= \exp\left(-F_\mu\right) = \left(1 + \exp\left(-\mu\right)\right)^{\binom{n}{2}} \\ &= (1-p)^{\binom{n}{2}} \end{aligned} \qquad (32)$$

- Probability, $P(G)$ of graph in the ensemble is

$$\begin{aligned} P(G) &= \frac{\exp\left(-H\right)}{Z} = \frac{\exp(-\mu m)}{\left(1 + \exp\left(-\mu\right)\right)^{\binom{n}{2}}} \\ &= \left(\frac{1-p}{p} \right)^{-m} (1-p)^{\binom{n}{2}} \\ &= p^m (1-p)^{\binom{n}{2} - m}. \end{aligned} \qquad (33)$$

This is the probability for a graph for which each of the $\binom{n}{2}$ possible edges appear with probability p, which is the well known Bernoulli random graph model studied by Erdos and Renyi.

3.5 General Model

- More general model is to replace μ_i with μ_{ij} so the Hamiltonian is

$$H(G) = \sum_{i<j} \mu_{ij} a_{ij}(G) \qquad \text{(Hamiltonian)}$$

- Partition function

$$Z_{\vec{\mu}}^{(\text{Model})}(G) = \prod_{i<j} \left(1 + e^{-\mu_{ij}}\right). \qquad \text{(Partition Function)}$$

- Free energy is

$$F_{\mu}^{(\text{Model})} = -\sum_{i<j} \ln\left(1 + e^{-\mu_{ij}}\right) \quad \text{(Free Energy)}$$

- First moment is

$$\langle f \rangle^{(\text{Model})} = \frac{dF_{\mu}^{(\text{Model})}}{d\mu} = \frac{1}{(1 + e^{\mu_{ij}})} \qquad \text{(Mean)}$$

- Standard deviation is

$$\sigma_f^2\,(\text{Model}) = \frac{d^2 F_\mu^{(\text{Model})}}{d\mu^2} = -\frac{e^{\mu_{ij}}}{(1+e^{\mu_{ij}})^2}$$
(Standard Deviation)

- Capacity is

$$C_f^{(\text{Model})} = -\mu_{ij}^2 \frac{e^{(\mu_{ij})}}{\left(1+e^{(\mu_{ij})}\right)^2}. \quad \text{(Heat Capacity)}$$

- Reparameterize p as

$$p_{ij} = \frac{1}{(1+e^{\mu_{ij}})}, \tag{34}$$

3.6 General Model Observations

- In general, can specify interaction Hamiltonian as

$$H = \sum_{i<j} \Lambda_{ij}\sigma_{ij} \quad \text{(General Hamiltonian Model)}$$

where Λ_{ij} is parameter that couples each edge together and the degree sequence satisfies $\nu_i = \sum_j \sigma_{ij}$.

- Partition function is

$$Z = \prod_{i<j} \left(1 + e^{-\Lambda_{ij}}\right), \tag{35}$$

- Free energy is

$$F = -\sum_{i<j} \ln\left(1 + e^{-\Lambda_{ij}}\right). \tag{36}$$

- Note probability of an edge between edge i and j is the expected number of the degree sequence, then

$$p_{ij} = \left\langle \sigma_{ij} \right\rangle = \frac{\partial F}{\partial \Lambda_{ij}} = \frac{1}{\left(1 + e^{\Lambda_{ij}}\right)} \tag{37}$$

- Underlying basis for the parameterization that gives us Bernoulli model with

$$P(G) = p_{ij}^{m}(1 - p_{ij})^{\binom{n}{2}-m}. \tag{38}$$

- Term $\sum_i C_i\theta_i$ added to Hamiltonian without changing probability model; relates $\langle C_i \rangle$ to characteristic of graph interaction.

- *Any interaction model that can be cast into the above form is a hidden Markov model.*

3.7 Graph Characteristic (Optional)

- Any graph characteristic C, can find expected value from the formula

$$\langle C \rangle = Z^{-1}(G) \sum_{G} C(G) \exp\left(-H((G))\right). \quad (39)$$

- To measure parameter C, one simply adjoins it as a term added to the Hamiltonian

$$H(G) = \sum_{i=1}^{n} \mu_i \nu_i(G) + \theta C \quad (40)$$

- Calculate the expected value, $\langle C \rangle$, for a linear interaction potential from the free energy as

$$\langle C \rangle = \left. \frac{dF}{d\theta} \right|_{\theta=0}. \quad (41)$$

- This generalizes to enable one to calculate correlation functions of the graph

$$\left\langle \nu_i \nu_j ... \nu_l \right\rangle = Z^{-1} \left[\frac{\partial}{\partial \mu_i} \frac{\partial}{\partial \mu_j} ... \frac{\partial}{\partial \mu_l} \right] Z, \quad (42)$$

- Correlation between two elements as

$$\frac{\partial^2 F}{\partial \mu_i \partial \mu_j} = \left\langle \nu_i \nu_j \right\rangle - \left\langle \nu_i \right\rangle \left\langle \nu_j \right\rangle , \qquad (43)$$

  while the extension to higher orders is straight forward.

4 Concluding Comments

- In principle, any probability model for random graphs can be thought of as a physical interaction model, so one can bring to bear the full power of statistical physics on it.

- This approach applies to any type of network that can be modeled as a graph with a normalized or zero/one entries for the adjacency matrix.

- Many design problems can be viewed as instances of random graph models
  - Social Networks
  - Physics problems
  - System engineering
  - Politics

- The translation: **Interaction Model** $\Leftrightarrow$ **Probability Model** provides a richer interpretation of "physics". This allows us to consider more general types of interactions that the *statistical mechanics crank* can be brought to bear.

- Time evolving graphs have the potential to be a huge problem area with many different types of domain applications as well.

- This technique can be applied to flow networks either by using non-symmetric matrix or by using negative probabilities. Each could be interesting in terms of the problems raised as well as the problems solved.

- Phase Transitions for networks are possible for social networks in analogy with Ising Model.

References

[1] Frank, O. and Strauss, D., "Markov Graphs", *Journal of the American Statistical Association*, 1Vol. 81, 832-842, 1986.

[2] E. T. Jaynes, in *Maximum Entropy and Bayesian Methods*, edited by C.R. Smith, G. J. Erickson, and P.O. Neudorfer, Kluwer Academic, 1992.

[3] E. T. Jaynes, *E. T. Jaynes: Papers on Probability, Statistics, and Statistical Physics*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1989.

[4] Park, J. and Newman, M. E. J., "The Statistical Mechanics of Networks", arXiv:cond-mat/0405566v1 25 May, 2004.

[5] Shannon, C. E. and Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1964. (Originally published in 1949, this contains a reprint of Shannon's 1948 papers that appeared in the *Bell System Technical Journal*.)

[6] Skvoretz, J., "Complexity Theory and Models for Social Networks", *Complexity*, Vol. 8, No. 1, 2003.

[7] Strauss, D., "On a General Class of Models for Interaction", *SIAM Review*, Vol. 28, No. 4, 1986.