Comparing Structured and Unstructured Models

Dheeraj Mysore Nagaraj

Lab for Information and Decision Systems, MIT
(joint work with Guy Bresler)

dheeraj@mit.edu

January 31, 2018
Suppose there are households on a 2D grid and people living near each other tend to have similar political opinions.

In bipartisan voting, the outcome is 60:40 and the data about who voted for whom is available.

Accusation of rigging: that someone assigned votes randomly to people to get this result.

Can we decide if the voting data is fraudulent or legitimate?

(a) voting households
A Story about Voting

(a) probably legitimate

(b) probably fraudulent

(c) South Carolina 2008 (courtesy: NBC)
A Story about Voting: Key Features

- Only one sample available.
- Underlying graph structure is known. (Structured Model)
- Exact extent of interaction (people's influence on neighbors) unknown.
- Test against data coming from distribution without the said graph structure. (Unstructured Model)
- The data generated is intimately linked to the structure of the underlying network.
- We will use Ising model to for voter behavior.
Data is generated by networks or network itself is data.

- Ising model: graphical model for binary data over networks.
- Erdős-Rényi graphs, Exponential Random Graphs, Stochastic Block Model etc. are used to model social networks.
Ising Model

Graph $G = (V, E)$. Random variable along each vertex $\in \{-1, 1\}$. Each node interacts only with the neighbors in the graph.

$$p(x) \propto \exp \sum_{(i,j) \in E} J_{ij} x_i x_j = \exp \frac{1}{2} x^T J x$$

Curie-Weiss Model

Ising model over complete graph. ‘Inverse temperature’ $\beta$,

$$p(x) \propto \exp \frac{n\beta}{2} m^2$$

Where $m(x) := \frac{\sum_{i=1}^{n} x_i}{n}$
Interaction Diagram

(a) Curie Weiss Interaction

(b) 3 Regular Graph Interaction

Figure: Interaction Diagram
Vote Rigging Model

$x \in \{-1, 1\}^n$: $-1$ is vote for first party, $+1$ is vote for second party.

**H$_0$: Legitimate voting**

$$p(x) \propto \exp \beta (A(x) - D(x))$$

$A(x) = \#$ agreements between neighbors

$D(x) = \#$ disagreements between neighbors

$\beta$ unknown. $d$ neighbors for everyone $\implies$ $d$ regular Ising model.

Intuition: Promotes agreement between neighbors.

**H$_1$: Fraudulent voting**

Curie-Weiss model or everyone votes independently with some partisan bias. Unknown inverse temperature or bias.
Random Graph Models

Erdős-Rényi Model

$G(n, p)$: Edges are independently chosen with probability $p$.

$$
\mu(G) = p^{|E(G)|} (1 - p)^{\binom{n}{2} - |E(G)|} 
\propto e^{|E(G)| \log \left( \frac{p}{1-p} \right)} 
$$

Exponential Random Graph Model

$$
p(G) \propto \exp \left[ \beta_1 |E(G)| + \frac{\beta_2}{n} N_G(\cdot) + \ldots \right]
$$

Where $\beta \in \mathbb{R} \times (\mathbb{R}^+)^{K-1}$. Intuition: Promotes the appearance of sub-graphs in addition to edges.
Curie-Weiss model has a lot of symmetries.

Fraudulent voting data from Curie-Weiss model: it doesn’t consider the ‘community structure’ where neighbors influence each other.

\( G(n, p) \) assigns the same probability to graphs with the same number of edges, irrespective of the structural properties.

We concentrate on Ising model to illustrate our method. Similar methods can be used for ERGMs.
Interactions are small enough and no long range order.

A node cannot influence nodes very far away.

We can still distinguish structured and unstructured models in high temperature regime!

Main result gives sharp thresholds above which it is possible to distinguish them and below which it is impossible to distinguish them.
Hypothesis Testing Problem

We observe \( x \in \{-1, 1\}^n \)

**H\(_0\)**: Data from Curie-Weiss model at unknown temperature/Independence model with unknown bias.

**H\(_1\)**: Data from \( d \)-regular Ising model, with known graph but unknown interaction strength \( \beta \).
Our Theorem: Ising Model

- Let $d = o(n)$. Statistical test to distinguish d-regular Ising model from Curie-Weiss/Independent voting model.
- Can distinguish with one sample with high probability when $\beta \gg \frac{1}{\sqrt{nd}}$, where $\beta$ interaction along each edge in the $d$ regular model.
- When $\beta \ll \frac{1}{\sqrt{nd}}$, there is no test to do this with constant number of independent samples: Our test is optimal!
- High temperature regime is $\beta \leq \Theta(\frac{1}{d})$. The statistical threshold lies deep inside it.
Let $A_m$ be the set of all states such that $\sum_i \frac{x_i}{n} = m$.

Curie-Weiss model/ Independent voting model: $p(. | x \in A_m)$ uniform on $A_m$ because of symmetries in them.

d regular Ising model: $q(x) \propto e^{\beta(A(x) - D(x))}$. Higher probability to states with higher number of agreements.

We decide $H_1$ when $A(x) - D(x)$ ‘is large enough’.

This test is optimal!
Figure: distribution of $A(x) - D(x)$ under uniform measure over $A_m$
An Intuitive View

- To prove that this test works, we need to show that the right tail is large enough: Use central limit theorem! (We apply Stein’s method to show this)
- To prove the converse, we need exponential concentration inequalities
Consider ERGM(\(\beta_1, \beta_2\)) with \(K = 2\), with \(H_2\) being the \(V\) graph (\(\vee\)).

Same framework to distinguish ERGM(\(\beta_1, \beta_2\)) from \(G(n, p)\).

Statistical test: Count the number of (\(\vee\)) graphs!

When \(\beta_2 \gg \frac{1}{\sqrt{n}}\), our test can distinguish them using one sample with high probability.

When \(\beta_2 \ll \frac{1}{\sqrt{n}}\), there is no test to do this: Our test is optimal!

Correlation decay regime threshold is \(\beta \leq \Theta(1)\). The statistical threshold lies deep inside the correlation decay regime.
Abstract Result: Conditions

We consider a sequence of probability spaces \((\Omega_n, \mathcal{F}_n, p_n)\), \(n \in \mathbb{N}\). Define measure \(q\) as:

\[
\frac{dq}{dp} = \frac{e^{\beta g}}{\mathbb{E}_p[e^{\beta g}]}
\]

Let the following conditions hold.

- We can partition \(\Omega = \bigcup_{m=1}^{M} A_m\) with disjoint sets \(A_m\) such that \(p(A_m) > 0 \ \forall \ i\).

- There exists \(S_n \subset [M]\) such that \(p_n(\bigcup_{m \in S_n} A_m) \geq 1 - \alpha_n\) for some sequence \(\alpha_n \to 0\).

- Let \(p^{(m)} = p(\cdot | x \in A_m)\). Let \(X_m \sim p^{(m)}\). Let \(e_m(g) := \mathbb{E}[g(X_m)]\) and \(\sigma^2_m(g) := \text{var}[g(X_m)]\). For all \(m, m' \in S_n\),

\[
0 < c \leq \frac{\sigma^2_m(g)}{\sigma^2_{m'}(g)} \leq C
\]
Abstract Result: Conditions

- \( \sup_{m \in S_n} d_{KS} \left( \mathcal{L} \left( \frac{g(X_m) - e_m(g)}{\sigma_m} \right), \mathcal{N}(0, 1) \right) < \tau_n \) for some sequence \( \tau_n \to 0 \). We let \( \sigma_n \) be any number such that \( c \sigma_m \leq \sigma_n \leq C \sigma_m \) for some constants \( c \) and \( C \) for every \( m \in S_n \).

- Let \( \text{var}(g(X)) = O(\sigma_n^2) \). Further, let

\[
\log \mathbb{E} \left[ e^{\beta(g(X) - \mathbb{E}g(X))} \right] \leq \frac{\beta^2 \nu}{1 - |\beta| c}
\]

\( \forall |\beta| < \frac{1}{c} \) such that \( c \leq A \sigma_n \) and \( \nu \leq B \sigma_n^2 \) for absolute constants \( A, B \) independent of \( n \).
Canonical Statistical Test

Function $m(x)$ such that $m(x) = m_0$ iff $x \in A_{m_0}$.

$H_0 : X \sim p$

$H_1 : X \sim q$ for some $\beta_n$

We call the following statistical test to distinguish $H_0$ and $H_1$ the canonical test with parameter $T \geq 0$:

**Definition (Canonical Test)**

Given a sample $X$, we define the following decision function $D(X) \in \{H_0, H_1\}$:

1. if $m(X) \not\in S_n$ then $D(X) = H_1$
2. if $m(X) \in S_n$ and $\frac{g(X) - e_m(x)(g)}{\sigma_m(x)} \geq T$ then $D(X) = H_1$
3. otherwise $D(X) = H_0$

The test doesn’t depend on $\beta_n$
Abstract Result

Theorem (Main Result)

If the conditions above hold,

\[
\lim_{n \to \infty} d_{TV}(p_n, q_n) = 1 \quad \text{if } \beta_n \sigma_n \to \infty \tag{1}
\]

\[
\lim_{n \to \infty} d_{TV}(p_n, q_n) = 0 \quad \text{if } \beta_n \sigma_n \to 0 \tag{2}
\]

If \( \beta_n \sigma_n \geq L_n \) for a known sequence \( L_n \to \infty \), (\( \beta_n \) being possibly unknown) the canonical test can distinguish between \( p \) and \( q \) with high probability with a single sample for a particular choice \( T_n \to \infty \) depending only on \( L_n \) and \( \tau_n \). The probability of error can be bounded above by a function of \( \alpha_n \), \( T_n \) and \( L_n \) tending to 0.
To compare $d$-regular Ising model over graph $G$ at temperature $\beta$ with Curie-Weiss model at temperature $\beta'$, take $p(\cdot)$ to be the Curie-Weiss model.

$A_{m_0} := \{x : m(x) = m_0\}$ and $g(x) = x^\top J x - \frac{\beta'}{\beta} n m^2$ where $J$ is the adjacency matrix of $G$.

$\sigma_n = \Theta(\sqrt{nd})$ uniformly $\forall m \in [-1 + \delta, 1 - \delta] =: S_n$

$\sup_{m \in S_n} d_{KS} \left( \mathcal{L} \left( \frac{g(X_m) - e_m(g)}{\sigma_m} \right), \mathcal{N}(0, 1) \right) < C \sqrt[4]{\frac{d}{n}}$
To compare ERGM($\beta_1, \beta_2$) with $G(n, p)$, take $p(\cdot) \sim G(n, p)$

$A_{m_0} := \{ G = (V, E) : |E| = m_0 \}$

$g(G) = \left( n \left( \frac{\beta_1 - \frac{1}{2} \log \frac{p}{1-p}}{\beta_2} \right) E(G) + N_2(G) \right)$.

$\sigma_n = \Theta(n^{3/2})$ uniformly $\forall m \in \left[ \frac{\delta}{2} N, N(1 - \frac{\delta}{2}) \right] =: S_n$ where $N = \binom{n}{2}$

$\sup_{m \in S_n} d_{KS} \left( \mathcal{L} \left( \frac{g(X_m) - e_m(g)}{\sigma_m} \right), \mathcal{N}(0, 1) \right) < C \sqrt{\frac{1}{n}}$

Proof of both central limit theorems rely on application of Stein’s method using suitable exchangeable pairs.
Questions ?