Belief Propagation, Robust Reconstruction and Optimal Recovery of Block Models
Elchanan Mossel, Joe Neeman & Allan Sly

Kristen Vaccaro, Ismini Lourentzou, Casey Hanson

University of Illinois, Urbana-Champaign

Overview

Purpose

Purpose: community detection

Graph with

- nodes = individuals divided into communities
- edges = connections between individuals
  more likely in-class than between groups

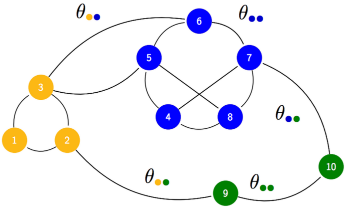

Figure: Denser connections within communities than between them

Purpose

Use connection information to infer community affiliation

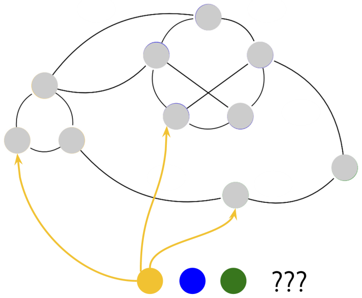

Figure: Goal: recover community affiliations

Previous Work

Recover exact communities
For dense graphs
Require a, b poly(n)

- Dyer and Frieze
- Snijders and Nowicki
- Condon and Karp

Spectral algorithm, modularity maximization
Require a, b O(log(n))

- McSherry
- Bickel and Chen

Previous Work (2)

O(log(n)) important!

If average degree:	**Recover communities:**
log(n)	exactly whp
<log(n)	impossible to recover exactly
Constant	cannot beat constant fraction correct
Conjectured threshold	cannot beat random guessing

Constant average degree

Motivation:
Many real networks sparse (most avg degree ≤ 20)
More realistic to expect imperfect recovery

- Only one algorithm guaranteed to do anything
- Coja-Oghlan (2010)
- Produced communities which have a better-than-50% overlap with the true communities

Sparse Block Models

Block model as described above

- nodes = n nodes divided into communities
- edges = drawn independently at random

  a/n = probability of an in-class edge will appear
  b/n = probability of a between-class edge will appear

Here:
a, b fixed
n tends to infinity
Only two classes (+, -)
classes (+, -) are roughly equal size
Sparsest non-trivial case

Problem Definition

In a sparse block model, how accurately one can recover the underlying communities?

- upper bound on the recovery accuracy (previous work)
- upper bound is tight
  when the signal to noise ratio is sufficiently high
- give an algorithm* which performs as well as the upper bound

*The algorithm leverages belief propagation as an initial guess at the communities and then tries to locally improve that initial guess.

Connection between the block model and broadcasting on trees

Core idea:

- Neighborhood in G looks like a tree with branching number $d = (a+b)/2$
- Labels on the neighborhood look like they came from a broadcast process with $\eta = \frac{b}{a+b}$
- $\theta^2 d = (1-2\eta)^2 d = \frac{(a-b)^2}{2(a+b)}$
- Threshold for community reconstruction is the same as the proven threshold for tree reconstruction

Solution outline

Show that:

1. Graph cluster problem is **harder** than tree reconstruction problem (smaller optimal accuracy)
2. Graph cluster problem is **easier** than robust tree reconstruction problem (higher accuracy)
3. Tree reconstruction **as hard as** robust tree reconstruction

Broadcasting on Trees

- Information transmitted from root
- Send down edges
- Each edge:
  - bit reversed w/ prob ϵ
  - errors occur independently

Common question:
What correlation btw inferred root label from leaves and true label?

Belief Propagation

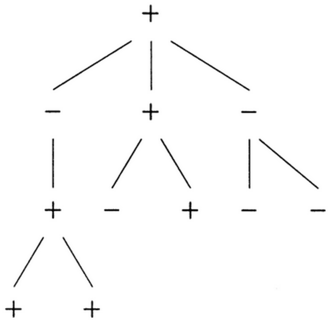

Figure: Infer root label from leaf labels, given some probability of transmission error between nodes

Signal to noise ratio

Theorem

If $\frac{|a-b|}{\sqrt{a+b}} > \sqrt{2}$ *then*
node labels can be inferred with accuracy better than 50%

When the signal-to-noise ratio is sufficiently high, then adding extra noise on the leaves of a large tree does not hurt our ability to guess the label of the root given the labels of the leaves.

$$\frac{|a-b|}{\sqrt{a+b}} > \sqrt{2} \Leftrightarrow \frac{|a-b|}{\sqrt{2(a+b)}} > 1 \Leftrightarrow \frac{(a-b)^2}{2(a+b)} > 1 \Leftrightarrow d\theta^2 > 1$$

Broadcasting on trees

Definition

Given $\eta \in [0, \frac{1}{2})$ and a tree T, the broadcasting process on T is defined as follows:
let σ_ρ be $+$ or $-$ with a probability $\frac{1}{2}$.
Then, for each u such that σ_u is defined and for each $v \in L_1(u)$, let $\sigma_v = \sigma_u$ with probability $1 - \eta$ and $\sigma_v = -\sigma_\rho$ otherwise

where $L_k(u)$ are the kth-level decedents of u
and $i \in C(u)$ iff $ui \in L_1(u)$ (u's children)

Broadcasting on trees

Theorem (Tree reconstruction threshold)

Let $\theta = 1 - 2\eta$ and d be the branching number of T. Then $\mathbb{E}[\sigma_\rho | \sigma_u : u \in L_k(\rho)] \to 0$ as $k \to \infty$ iff $d\theta^2 \leq 1$

If $d\theta^2 > 1$ then $\forall k$ there is an algorithm which guesses σ_ρ given $\sigma_{L_k}(\rho)$ and succeeds with probability bounded away from $\frac{1}{2}$

Noisy Broadcasting on trees

Definition

Given $\delta \in [0, \frac{1}{2}]$ and a broadcasting process σ on a tree T, the noisy broadcasting process on T is defined by independently taking $\tau_u = -\sigma_u$ with probability δ and $\tau_u = \sigma_u$ otherwise

For a certain range of parameters, the presence of noise at leaves does not affect the accuracy with which the root can be reconstructed.

Belief Propagation and Tree Reconstruction

Optimal estimator of σ_ρ given $\sigma_{L_k}(\rho)$: $sgn(X_{\rho,k})$ where
$X_{\rho,k} = 2Pr(\sigma_\rho = +|\sigma_{L_k(\rho)}) - 1$
expected gain $= \mathbb{E}|Pr(\sigma_\rho = +|\sigma_{L_k(\rho)}) - \frac{1}{2}|$

Definition (Tree reconstruction accuracy)

The probability of correctly inferring σ_ρ given the labels at infinity
$p_T(a, b) = \frac{1}{2} + \lim_{k\to\infty} \mathbb{E}|Pr(\sigma_\rho = +|\sigma_{L_k(\rho)}) - \frac{1}{2}|$ where $\eta = \frac{b}{a+b}$

Graph reconstruction and Tree Reconstruction

Definition (Graph reconstruction accuracy)

Consider the block model on n vertices with parameters a,b where $a + b > 1$.
Let $p_{G,n}(a,b) = \frac{1}{2} + \sup_f \mathbb{E}|\frac{1}{n}\sum_u 1(f(u,G) = \sigma_u) - \frac{1}{2}|$
to be the best reconstruction probability for the cluster.

Let $p_G = \limsup_n p_{G,n}$
p_G is the optimal fraction of nodes that can be reconstructed correctly.

Theorem

$p_G(a,b) \le p_T(a,b)$ *(Proof in previous work)*

Solution outline

Show that:

1. **Graph cluster problem is harder than tree reconstruction problem (smaller optimal accuracy)**
2. Graph cluster problem is **easier** than robust tree reconstruction problem (higher accuracy)
3. Tree reconstruction **as hard as** robust tree reconstruction

Graph reconstruction and Robust Tree Reconstruction

Definition (Robust tree reconstruction accuracy)

Consider the noisy tree broadcast process with an additional $\delta \in [0, \frac{1}{2}]$ and some extra variables $\{\tau_u : u \in T\}$, which are independent conditioned on $\{\sigma_u : u \in T\}$ and satisfy $Pr(\tau_u = \sigma_u) = 1 - \delta$

Define the robust reconstruction probability as:
$\tilde{p}_T(a, b) = \frac{1}{2} + \lim_{\delta \to \frac{1}{2}} \lim_{k \to \infty} \mathbb{E}|Pr(\sigma_\rho = +|\tau_{L_k(\rho)}) - \frac{1}{2}|$

- noise in σ propagates down the tree
- noise in τ does not propagate, therefore not important

What is left to show

Consider an algorithm for reconstructing the block models which satisfies that with high probability it labels $\frac{1}{2} + \delta$ of the nodes accurately. Then the algorithm can be used in a black box manner to provide an algorithm whose reconstruction accuracy (with high probability) is $\tilde{p}_T(a, b)$

Solution outline

Show that:

1. **Graph cluster problem is harder than tree reconstruction problem (smaller optimal accuracy)**
2. **Graph cluster problem is easier than robust tree reconstruction problem** (higher accuracy)
3. Tree reconstruction **as hard as** robust tree reconstruction

What is left to show (2)

Theorem (1.)

There exists a constant C such that if $(a-b)^2 \geq C(a+b)$ then $p_G(a,b) = p_T(a,b)$

Theorem (2.)

There exists a constant C such that if $(a-b)^2 \geq C(a+b)$ then $\tilde{p}_T(a,b) = p_T(a,b)$

Solution outline

Show that:

1. **Graph cluster problem is harder than tree reconstruction problem (smaller optimal accuracy)**
2. **Graph cluster problem is easier than robust tree reconstruction problem (higher accuracy)**
3. **Tree reconstruction as hard as robust tree reconstruction**

Magnetization

Definition

$X_{u,k} = Pr(\sigma_u = + \mid \sigma_{L_k(u)}) - Pr(\sigma_u = - \mid \sigma_{L_k(u)})$
$x_k = \mathbb{E}(X_{u,k} \mid \sigma_u = +)$

$X_{u,k}$ is the magnetization of u given $\sigma_{L_k(u)}$

$\frac{(1+x_k)}{2}$ is the probability of estimating σ_ρ correctly given $\sigma_{L_k(\rho)}$

Noisy Magnetization

Definition

$Y_{u,k} = Pr(\sigma_u = + \mid \tau_{L_k(u)}) - Pr(\sigma_u = - \mid \tau_{L_k(u)})$
$y_k = \mathbb{E}(Y_{u,k} \mid \sigma_u = +)$

$Y_{u,k}$ is the magnetization of u given $\sigma_{L_k(u)}$

$\frac{(1+y_k)}{2}$ is the probability of estimating σ_ρ correctly given $\tau_{L_k(\rho)}$

Proof of Theorem 1

Basic assumptions

Consider the broadcast process on the infinite $\frac{a+b}{2} =$ d-ary tree with parameter $\eta = \frac{a}{a+b}$. Set $\theta = 1 - 2\eta$. For any $0 < \theta^* < 1$, there is some d^* such that if $\theta \geq \theta^*$ and $d \geq d^*$ then $\tilde{p}_T(a, b) = p_T(a, b)$

Under these assumptions:
$lim_{k\to\infty} x_k = lim_{k\to\infty} y_k$

Outline of proof

First show that when $\theta^2 d$ is large, both the exact reconstruction and the noisy reconstruction do quite well.

Studying the estimator of the most common label among $\sigma_{L_k(\rho)}$

Simple majority method

Suppose $d\theta^2 > 1$.

Define:

$$S_{u,k} = \sum_{v \in uL_k} \sigma_v$$
$$\widetilde{S}_{u,k} = \sum_{v \in uL_k} \tau_v.$$

Goal: estimate σ_ρ by $\mathrm{sgn}(S_{\rho,k})$ or $\mathrm{sgn}(\widetilde{S}_{\rho,k})$

Simple majority method (2)

First moment

Lemma (1)

$$\mathbb{E}^+ S_{\rho,k} = \theta^k d^k$$
$$\mathbb{E}^+ \widetilde{S}_{\rho,k} = (1-2\delta)\theta^k d^k$$

Second moment

Lemma (2)

$$Var^+ S_{\rho,k} = 4\eta(1-\eta) d^k \frac{(\theta^2 d)^k - 1}{\theta^2 d - 1}$$
$$Var^+ \widetilde{S}_{\rho,k} = 4 d^k \delta(1-\delta) + 4(1-2\delta)^2 \eta(1-\eta) d^k \frac{(\theta^2 d)^k - 1}{\theta^2 d - 1}$$

Simple majority method (3)

Lemma (3)

If $d\theta^2 > 1$ then

$$\frac{Var^+ S_k}{(\mathbb{E}^+ S_k)^2} \to \frac{4\eta(1-\eta)}{\theta^2 d} \text{ as } k \to \infty$$

$$\frac{Var^+ \widetilde{S}_k}{(\mathbb{E}^+ \widetilde{S}_k)^2} \to \frac{4\eta(1-\eta)}{\theta^2 d} \text{ as } k \to \infty$$

Simple majority method (4)

- The estimators $sgn(S_k)$ and $sgn(\widetilde{S}_k)$ succeed with probability at least $[1 - \frac{4\eta(1-\eta)}{\theta^2 d}]$ as $k \to \infty$
- The optimal estimator of σ_ρ given $\sigma_{L_k(\rho)}$ is $sgn(X_{\rho,k})$ with success probability $\frac{(1+x_k)}{2}$
- $\frac{(1+x_k)}{2}$ must be larger than the success probability of $sgn(S_k)$

Simple majority method (5)

Lemma (4)

If $d\theta^2 > 1$ *then* $x_k \geq [1 - \frac{10\eta(1-\eta)}{\theta^2 d}]$ *for large* k

(Similarly for y_k *and* $sgn(\widetilde{S}_k)$*)*

Since $X_{u,k} \leq 1$ and $x_k = \mathbb{E}(X_{u,k} \mid \sigma_u = +)$ we can apply Markov's inequality to show that $X_{u,k}$ is large with high probability.

Connection

$$\forall v \in V^i \qquad Pr(v_*^+) \to P_T(a,b)$$

$$Pr(Y_{v,R}(\epsilon) = \sigma_v) = P_T(a,b)$$

Algorithm

Algorithm 1 Optimal graph reconstruction algorithm

$R \leftarrow \lfloor \frac{1}{10\log(2(a+b))} \log n \rfloor$
$W_G^+, W_G^- \leftarrow \texttt{Partition}(G)$
$W_*^+, W_*^- \leftarrow \varnothing$
for all $v \in V$ **do**
  $W_v^+, W_v^- \leftarrow \texttt{Partition}(G \setminus B(v, R-1))$
  relabel W_v^+, W_v^- so that $|W_v^+ \Delta W_G^+| \leq n/2$.
  define $\xi \in \{+,-\}^{S(v,R)}$ by $\xi_u = i$ if $u \in W_v^i$
  add v to $W_*^{Y_{\rho,R}(\xi)}$
end for
return W_*^+, W_*^-

Assign initial guess for all node labels

Loop:

- Sample a node
- Delete local neighborhood
- Use nodes at distance R & belief propagation to hypothesize label