Compressive Sensing and Clique Identification in Social Networks

Yuan Yao

Peking University

Selected Topics in Advanced Statistics
Nov. 13, 2009

Example I: Basket ball teams

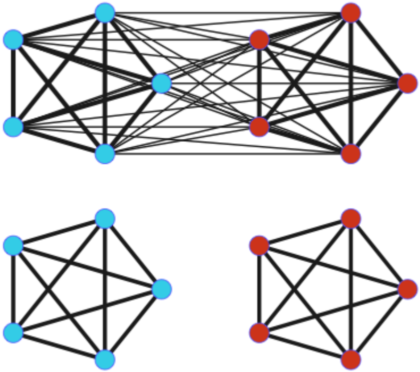


Figure: Two teams in a virtual Basketball Game, with large intra-team interaction and noisy cross-team interaction.

Example II: Social Network of Les Miserables

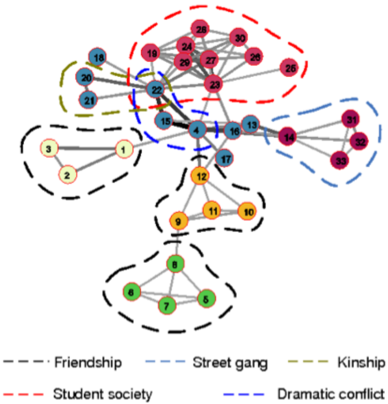


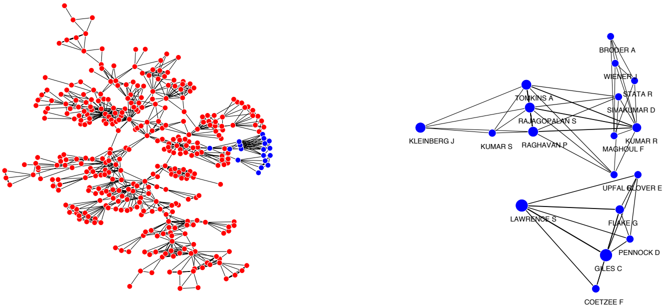
Figure: Cliques in the social network of Les Miserables, by Victor Hugo (data courtesy to Knuth'93).

Example II continued: Cliques in Les Miserables

Table: Cliques in The Social Network of *Les Miserables*

Cliques	Names of Characters	Relationships
$\{1,2,3\}$	{Myriel, Mlle Baptistine, Mme Magloire}	Friendship
$\{4,12,16\}$	{Valjean, Fantine, Javert}	Dramatic Conflicts
$\{4,13,14\}$	{Valjean, Mme Thenardier, Thenardier}	Dramatic Conflicts
$\{4,15,22\}$	{Valjean, Cosette, Marius}	Dramatic Conflicts
$\{20,21,22\}$	{Gillenormand, Mlle Gillenormand, Marius}	Kinship
$\{5,6,7,8\}$	{Tholomyes, Listolier, Fameuil, Blacheville}	Friendship
$\{9,10,11,12\}$	{Favourite, Dahlia, Zephine, Fantine}	Friendship
$\{14,31,32,33\}$	{Thenardier, Gueulemer, Babet, Claquesous}	Street Gang

Example III: Coauthorship in Network Science



(a) (b)

Figure: Coauthorship in Network Science: (a) coauthorship relations between scientists working on network theory (Newman'06); (b) A close-up around Jon Kleinberg

Example IV: Jester Dataset

- In Jester data set, there are 24, 000 users rating over 100 jokes, partially.
- From the data we can count votes on all top-3 jokes (or just the best joke).
- Can we infer which 5-tuple is the first tier group?

Problem

- These examples observe low order (pairwise) interactions, which are often governed by high order cliques (complete subgraphs: teams, first tier groups)
- Cliques may have overlaps, where traditional partition-based clustering such as spectral clustering fails here
- Can we find a mathematical framework for detecting such cliques?

(Yes!)

Compressive Sensing + algebraic Radon basis

Look for a representation

Given n nodes, labeled from $1, \ldots, n$.

- Permutation Group: The $n!$ rankings make up of the permutation group S_n
- Homogeneous Space: cosets $H_k := \{S_n/S_k \times S_{n-k}\}$ can be identified as all k-subsets of $\{1, \ldots, n\}$.

Fact

Inferring high order cliques from low order interactions can be regarded as a mapping between functions on homogeneous spaces $H_i^* \mapsto H_k^*$ $(i < k)$.

Inferring High Order Cliques from Low Order Interactions

Example

2-cliques	Frequency
{1 2}	10
{1 3}	7
{1 4}	3
{1 5}	6
⋮	⋮

Example

3-cliques	Frequency
{1 2 3}	?
{1 2 4}	?
{1 2 5}	?
{1 2 6}	?
⋮	⋮

Radon Basis

- Interpret the function on 2-subsets as interaction frequency
- A 2-subset is randomly from some k-cliques (teams) included
- Assume inherent frequency function on k-cliques (teams) is sparse.
- Build matrix A as following:

	1 2 3	1 2 4	1 2 5	1 3 4	1 3 5	···	···	3 4 5
1 2	1	1	1	0	0	···	···	···
1 3	1	0	0	1	1	···	···	···
1 4	0	1	0	1	0	···	···	···
1 5	0	0	1	0	1	···	···	···
2 3	1	0	0	0	0	···	···	···
.	···	···	···	···	···	···	···	···
.	···	···	···	···	···	···	···	···
4 5	0	0	0	0	0	···	···	1

Radon Basis

- Such a matrix is an example of Radon basis
- In general, there is a canonical Radon Transform in algebraic combinatorics (Diaconis'88) which maps functions on k-subsets to j-subsets ($j \leq k$)

$$(R^{k,j})u(\tau) = \sum_{\sigma \subset \tau} u(\sigma), \quad \tau \in H_k, \sigma \in H_j$$

- Radon basis is just the transpose of Radon Transform, upto a scaling factor

Radon Basis Pursuit Formulation

Suppose x_0 is a sparse function on k-cliques. To reconstruct this sparse function based on low order observation data b, consider the following linear programming first known as Basis Pursuit

$$\mathcal{P}_1: \quad \min \quad \|x\|_1 \quad \text{subject to} \quad Ax = b$$

which is a convex relaxation of original NP-hard problem

$$\mathcal{P}_0: \quad \min \quad \|x\|_0 \quad \text{subject to} \quad Ax = b$$

A Result from KKT-Condition for $\mathcal{P}_1$

Suppose A is a M-by-N matrix and x_0 is a sparse signal. Let $T = \text{supp}(x_0)$, T^c be the complement of T, and A_T (or A_{T^c}) be the submatrix of A where we only extract column set T (or T^c, respectively).

Theorem (Exact Recovery Theorem, Candes-Tao'05)

Assume that $A_T^* A_T$ is invertible and there exists a vector $w \in R^M$ such that
(1) $A_T^* w = sgn(x_0)|_T$,
(2) $\|A_{T^c}^* w\|_\infty < 1$,
where $*$ denote matrix transpose and $sgn(x_0)|_T$ is the restriction of $sgn(x_0)$ on T. Then x_0 is the unique solution for $\mathcal{P}_1$. The conditions are also necessary.

Proof Ideas

1. Consider equivalently

$$\begin{array}{ll} \min & 1^*\xi \\ \text{subject to} & Ax = b, \quad -\xi \leq x \leq \xi, \quad \xi \geq 0 \end{array}$$

2. Lagrangian is
$L(x,\xi;\gamma,\lambda,\mu) = 1^*\xi + \gamma^*(Ax-b) - \lambda_+^*(\xi - x) - \lambda_-^*(\xi + x) - \mu^*\xi$
3. Karush-Kuhn-Tucker (KKT) condition gives
   - $A^*\gamma = -(\lambda_+ - \lambda_-) \Rightarrow A_T^*\gamma = -\operatorname{sign}(x_0)|_T$
   - $1 - (\lambda_+ + \lambda_-) - \mu = 0 \Rightarrow |A_{T^c}^*\gamma| = 1 - \mu < 1$

Irrepresentable Condition

Searching w satisfying ERT is equivalent to solve the dual problem of $\mathcal{P}_1$, hence one often consider the special case that $w \in \mathrm{im}(A_T)$. Then ERT can be simplified to the following

$$\|A_{T^c}^* A_T (A_T^* A_T)^{-1} sgn(x_0)_T\|_\infty < 1$$

whose sufficient condition is easy to check

(Irrepresentable Condition (IRR), Yu-Zhao'06)

$$\|A_{T^c}^* A_T (A_T^* A_T)^{-1}\|_\infty < 1$$

Random Design

Candes-Romberg-Tao shows in a series of papers that when A is a random matrix, such as

- random Fourier transform
- Berrnoulli matrix
- Gaussian matrix

and when $|T| < O(M/log(N))$, with high probability IRR holds. This leads to Uniform Recovery such that for any s-sparse signal ($|T| \leq s$), one may recover it by $\mathcal{P}_1$ with high probability.

Restricted Isometry Property

This is a result due to the Restricted Isometry Property (RIP, Candes-Tao'05, Candes'08) for random matrices.

(Restricted Isometry Property)

For every set of columns T with $|T| \leq s$, there exists a certain universal constant $\delta_s \in [0, 1)$ such that

$$(1 - \delta_s)\|x\|_{l_2}^2 \leq \|A_T x\|_{l_2}^2 \leq (1 + \delta_s)\|x\|_{l_2}^2, \quad \forall x \in R^s.$$

This is generalized to other Restricted Eigenvalue conditions (e.g. Bickel-Ritov-Tsybokov'07, Zhang'08)

Fixed Design

- However many deterministic A in fixed design, RIP fails
- This in particular includes Radon basis defined above
- In our basis construction of matrix $A = R^{j,k}$, RIP is not satisfied unless $s < \binom{k+j+1}{k}$ which cannot scale up with n.
- Universal recovery is impossible unless for extremely sparse signals
- But one can look for those T such that IRR etc. holds.

Exact Recovery Theorem: A lemma

Let $A = R^{j,k}$, given data b on all $j-$subsets, we wish to infer common interest groups on all $k-$subsets. Suppose x_0 is a sparse signal on all $k-$subsets.

Lemma

Let $T = supp(x_0)$, and $j \geq 2$. Suppose that for any $\sigma_1, \sigma_2 \in T$, there holds $|\sigma_1 \cap \sigma_2| \leq r$.

- *If $r = j - 2$, then $\|A^*_{T^c} A_T (A^*_T A_T)^{-1}\|_\infty < 1$;*
- *If $r = j - 1$, then $\|A^*_{T^c} A_T (A^*_T A_T)^{-1}\|_\infty \leq 1$ where equality holds with certain examples;*
- *If $r = j$, there are examples such that $\|A^*_{T^c} A_T (A^*_T A_T)^{-1}\|_\infty > 1$.*

Exact Recovery Theorem in Radon Basis Pursuit

Theorem

Let $T = supp(x_0)$, if we allow overlaps among k-cliques to be no larger than r, then the maximum r that can guarantee Irrepresentable Condition is $j - 2$.

- It says that when cliques have small overlaps, then exact recovery for sparse signals will hold.
- In practice, when overlaps are larger than $j - 2$, you may possibly find exact recovery by $\mathcal{P}_1$; as the theorem simply says there exists an example in this case which fails $\mathcal{P}_1$, but you might not meet it.

Sparse Approximation

- In real case, low order information b can be written as $b = Ax_0 + z$, where z accounts for bounded noises. In this case, we solve:

$$\mathcal{P}_{1,\delta}: \quad \min \quad \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_\infty \leq \delta$$

- For Gaussian noise, one may consider BPDN (Chen-Donoho-Saunders'99), close to Lasso

$$\mathcal{P}_{BPDN}: \quad \min \quad \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \delta$$

Regularization Path

In our applications, we choose bounded noise assumption which seems more natural.

Definition

A regularization path of $\mathcal{P}_{1,\delta}$ refers to the map $\delta \mapsto x_\delta$ where x_δ is a solution of $\mathcal{P}_{1,\delta}$.

A natural theoretical question asks: when the true signal x_0 lies on a unique regularization path?

A Result from KKT-Condition for $\mathcal{P}_{1,\delta}$

Theorem (Exact Recovery in Noisy Case)

Assume that A_T is of full column-rank. Then $\mathcal{P}_{1,\delta}$ has a unique solution x_0 if and only if there exists a $w \in R^N$ such that
(1) $A_T^* w = sgn(x_0)|_T$,
(2) $\|A_{T^c}^* w\|_\infty < 1$.
In other words, x_0 must lie on a unique regularization path.

Stable Recovery Theory in Noisy Case

Theorem

Using the same notation as before, assume that $\|z\|_\infty \leq \epsilon$, $|T| = s$, *and the Irrepresentable condition*

$$\|A_{T^c}^* A_T (A_T^* A_T)^{-1}\|_\infty \leq \alpha < \frac{1}{s}.$$

Then the following error bound holds for any solution $\hat{x}_\delta$ *of* $\mathcal{P}_{1,\delta}$,

$$\|\hat{x}_\delta - x_0\|_1 \leq \frac{2s(\epsilon + \delta)}{1 - \alpha s} \|A_T (A_T^* A_T)^{-1}\|_1.$$

Proof Ideas

1. Small tail bound: $\|h_{T^c}\|_1 \leq \|h_T\|_1$ where $h = \hat{x}_\delta - x_0$, i.e. $\|h_T\|_1 = \|x_0 - \hat{x}_\delta|_T\|_1 \geq \|x_0\|_1 - \|\hat{x}_\delta|_T\|_1 \geq \|\hat{x}_\delta\|_1 - \|\hat{x}_\delta|_T\|_1 = \|\hat{x}_\delta|_{T^c}\|_1 = \|h_{T^c}\|_1$, by $\|\hat{x}_\delta\|_1 \leq \|x_0\|_1$
2. Lower bound: (let $A_T^\dagger = A_T(A_T^* A_T)^{-1}$)

$$
\begin{aligned}
|\langle Ah, A_T^\dagger h_T\rangle| &= |\langle A_T h_T, A_T^\dagger h_T\rangle + \langle A_{T^c} h_{T^c}, A_T^\dagger h_T\rangle| \\
&\geq \|h_T\|_2^2 - \|h_{T^c}\|_1 \|A_{T^c}^* A_T^\dagger h_T\|_\infty \\
&\geq \frac{1}{s}\|h_T\|_1^2 - \alpha \|h_{T^c}\|_1 \|h_T\|_\infty \\
&\geq \frac{1}{s}\|h_T\|_1^2 - \alpha \|h_{T^c}\|_1 \|h_T\|_1 \\
&\geq \left(\frac{1}{s} - \alpha\right)\|h_T\|_1^2, \quad (\|h_{T^c}\|_1 \leq \|h_T\|_1)
\end{aligned}
$$

Proof Ideas: continued

3. Given $\|A\hat{x}_\delta - b\|_\infty \leq \delta$ and $z = Ax_0 - b$ with $\|z\|_\infty \leq \epsilon$. Then $\|Ah\|_\infty = \|A\hat{x}_\delta - Ax_0\|_\infty = \|A\hat{x}_\delta - b + b - Ax_0\|_\infty \leq \|A\hat{x}_\delta - b\|_\infty + \|z\|_\infty \leq \delta + \epsilon$.
4. Upper bound: (let $A_T^\dagger = A_T(A_T^* A_T)^{-1}$) $|\langle Ah, A_T^\dagger h_T\rangle| \leq \|Ah\|_\infty \|A_T^\dagger h_T\|_1 \leq (\delta + \epsilon)\|A_T^\dagger\|_1 \|h_T\|_1$
5. Combining lower and upper bounds gives

$$\|h_T\|_1 \leq \frac{s(\delta + \epsilon)}{1 - \alpha s}\|A_T(A_T^* A_T)^{-1}\|_1,$$

and the theorem follows from $\|h\|_1 \leq 2\|h_T\|_1$.

Stability Theory

Corollary

Assume that $k = j + 1, |T| = s$, *and overlap* $|\sigma_1 \cap \sigma_2| \leq j - 2$ *for any* $\sigma_1, \sigma_2 \in T$. *Then there holds*

$$\|A^*_{T^c} A_T (A^*_T A_T)^{-1}\|_\infty \leq 1/(j+1)$$

and the following error bound for solution $\hat{x}_\delta$ *of* $\mathcal{P}_{1,\delta}$,

$$\|\hat{x}_\delta - x_0\|_1 \leq \frac{2s(\epsilon + \delta)}{1 - \frac{s}{j+1}} \sqrt{j+1}, \qquad s < j + 1.$$

Practical Concerns: Mixed Cliques

- Stagewise algorithm: solving $P_{1,\delta}$ with different basis matrices ($A = R^{j,k}$ with the same j but different k) to detect cliques of different sizes.
- Concatenating different basis matrices $A = R^{j,k}$ together, solve for all cliques at the same time.
- Both actually work in practice.

Practical Concerns: Scalability

The basis matrix $R^{j,k}$ is of size $\binom{n}{j}$ by $\binom{n}{k}$ which makes it impossible to solve the linear programming $\mathcal{P}_1$ or $\mathcal{P}_{1,\delta}$ for all but very small n. Possible ways to deal with that

- Down-sample columns of A
- Divide-and-Conquer: use spectral clustering to pre-cluster the data, followed by Radon Basis Pursuit
- Iterative algorithms to solve LP

Divide-and-Conquer in coauthorship network

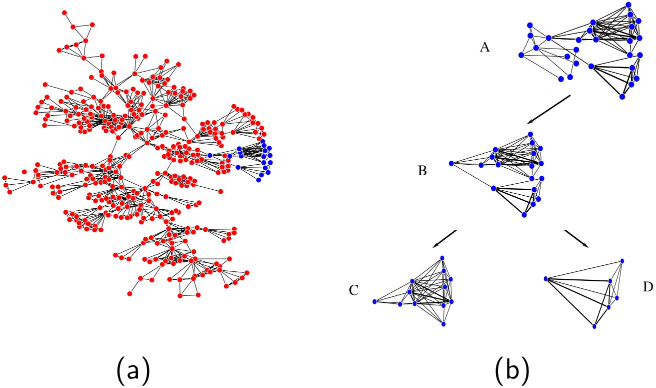


Figure: (a) coauthorship relations between scientists working on network theory (Newman'06); (b) Binary spectral clustering tree with Radon Basis Pursuit

Conclusions

- Radon Basis Pursuit provides a novel approach for clique identification in social networks, with possible overlaps where traditional partition-based clustering fails
- Its shortcoming lies in the combinatorial explosion in basis size, which however can be alleviated with the aid of spectral clustering preprocessing, etc.
- Can we exploit random design in this problem?

Acknowledgement

Collaborators:

- Xiaoye Jiang, Stanford ICME
- Leo Guibas, Stanford CS

Thanks to:

- Persi Diaconis
- Risi Kondor
- Minyu Peng