Ramanujan Graphs and the Matrix Completion Problem

M. Vidyasagar FRS

SERB National Science Chair and Distinguished Professor
IIT Hyderabad
m.vidyasagar@iith.ac.in, www.iith.ac.in/∼m_vidyasagar/

Gonit Sora 20 March 2021

Outline

Outline

Outline

Singular Values and SVD

Suppose $M \in \mathbb{R}^{m \times n}$ with (say) $m \leq n$. Then there exist matrices $U \in \mathbb{R}^{m \times m}, V \in \mathbb{R}^{n \times m}$ and nonnegative numbers $\sigma_1, \cdots, \sigma_m$ such that

$$U^\top U = V^\top V = I_m, M = \sum_{i=1}^{m} \sigma_i u_i v_i^\top .$$

This is called the **singular value decomposition (SVD)** of M. Here $\sigma_1, \cdots, \sigma_m$ are the eigenvalues of $MM^\top$, and are called the **singular values** of M. The rank of M is the number of nonzero singular values.

Singular values are less sensitive to errors in measuring M than eigenvalues.

Some Basic Graph Theory

A **Graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of "vertices" $\mathcal{V}$, and a set of "edges" $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. It is said to be **undirected** if, whenever $(v_i, v_j) \in \mathcal{E}$, also $(v_j, v_i) \in \mathcal{E}$. A graph is **simple** if (i) there are no self-loops (edges of the form (v_i, v_i)), and no multiple edges between any pair of vertices. A graph is **bipartite** if we can partition $\mathcal{V}$ as $\mathcal{V}_r \cup \mathcal{V}_c$ such that

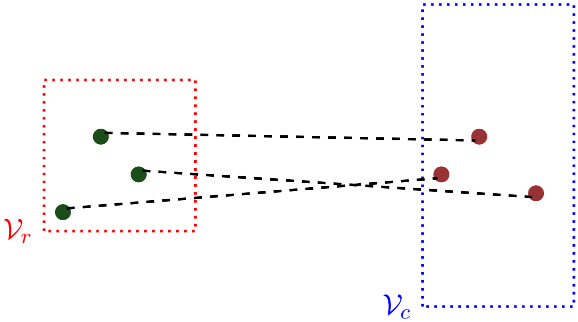
$$\mathcal{E} \cap (V_r \times V_r) = \emptyset, \mathcal{E} \cap (V_c \times V_c) = \emptyset.$$

Equivalently

$$\mathcal{E} \subseteq ((\mathcal{V}_r \times \mathcal{V}_c) \cup (\mathcal{V}_c \times V_r)).$$

A bipartite graph is **balanced** if $|\mathcal{V}_r| = |\mathcal{V}_c|$, **unbalanced** otherwise.

Depiction of a Bipartite Graph



No edges between two vertices in $\mathcal{V}_r$, or between two vertices in $\mathcal{V}_c$.

Incidence Matrix and Regularity

The **incidence matrix** $A \in \{0,1\}^{|\mathcal{V}| \times |\mathcal{V}|}$ has $a_{ij} = 1$ if $(i,j) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise.

The incidence matrix of a bipartite graph looks like

$$A = \begin{bmatrix} 0 & B \\ B^\top & 0 \end{bmatrix},$$

where $B \in \{0,1\}^{|\mathcal{V}_r| \times |\mathcal{V}_c|}$ is called the **biadjacency matrix**.

A graph is d-**regular** if every vertex has degree d. A bipartite graph is (d_r, d_c)-**biregular** if all vertices in $\mathcal{V}_r$ have degree d_r and all vertices in V_c have degree d_c (which also implies that $d_r|\mathcal{V}_r| = d_c|\mathcal{V}_c|$).

Spectra of Graphs

Fact: Suppose $A \in \{0,1\}^{n \times n}$ is the adjacency graph of a connected undirected d-regular graph. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ denote the eigenvalues of A. Then

1. $\lambda_i \in [-d, d]$ for all i.
2. $\lambda_1 = d$ with associated eigenvector $\mathbf{1}_n$, and $\lambda_i < d$ for $i \geq 2$.
3. $\lambda_n = -d$ if and only if the graph is bipartite.

Spectra of Bipartite Graphs

Fact: If a bipartite graph is (d_r, d_c)-biregular, then

1. The spectrum of A is $\{\pm\sigma_1, \ldots, \pm\sigma_l\}$, where σ_i are the singular values of B, together with the required number of zeros.
2. $\sigma_1 = \sqrt{d_r d_c}$ is the largest singular value of B.

IIT Hyderabad

Outline

IIT Hyderabad

Definition of a Ramanujan Graph

Definition

A d-regular graph $\mathcal{G}$ is said to be a **Ramanujan graph** if the *second largest eigenvalue* λ_2 of A satisfies

$$|\lambda_2| \leq 2\sqrt{d-1}.$$

Theorem

(Alon-Boppana Bound (1986)) In any d-regular graph with n vertices, we have that

$$\liminf_{n \to \infty} |\lambda_2| \geq 2\sqrt{d-1}.$$

Definition of a Ramanujan Bigraph

Definition

Suppose a bipartite graph is (d_r, d_c)-biregular. Then it is said to be a **Ramanujan bigraph** if

$$|\sigma_2| \leq \sqrt{d_r - 1} + \sqrt{d_c - 1}.$$

Theorem

(Feng-Li (1996)) If we fix d_r, d_c and let $n_r, n_c \to \infty$ (subject of course to $n_r d_r = n_c d_c$), then

$$\liminf_{\min\{n_r, n_c\} \to \infty} |\sigma_2| \geq \sqrt{d_r - 1} + \sqrt{d_c - 1}.$$

Prevalence of Ramanujan Graphs

Theorem

(Friedman (2008)) Suppose $d \geq 4$ is even and let $\epsilon > 0$ be arbitrary. Consider a d-regular n-vertex graph formed by $d/2$ uniform and independent permutations on $[n]$. Then

$$\max\{|\lambda_2|, |\lambda_n|\} \leq 2\sqrt{d-1} + \epsilon \text{ w.p. } 1 - O(n^{-r}),$$

where

$$r = [(\sqrt{d-1} + 1)/2] + 1.$$

Therefore

$$\max\{|\lambda_2|, |\lambda_n|\} \leq 2\sqrt{d-1} + \epsilon \text{ a.a.s.}$$

Simply put: Almost all d-regular graphs “almost” satisfy the Ramanujan property.

Prevalence of Ramanujan Bigraphs

No formal statement. Theorem due to Decker et al. (2018): An analogous statement holds if we study (d_r, d_c)-regular graphs with (n_r, n_c) vertices, fix d_r, d_c and let $n_r, n_c \to \infty$ (subject of course to $n_r d_r = n_c d_c$).

Where Are the Ramanujan Graphs Hiding?

"Randomly generated" d-regular graphs almost satisfy the Ramanujan property with a probability that approaches 1 as $n \to \infty$. (Ditto for biregular graphs.)

Ramanujan graphs are everywhere, but how to construct them *explicitly?*

Thus far there are just a handful of *explicit* constructions of Ramanujan graphs, and not a single *explicit* construction of a Ramanujan bigraph!

Outline

Outline

Cayley Graphs

Suppose G is a group (not necessarily Abelian), and $S \subseteq G$ satisfies $a \in S \implies a^{-1} \in S$. Such a set is called "symmetric." Then the **Cayley graph** of G generated by S has the elements of G as its vertices, and the edge set $\{(x, xa), x \in G, a \in S\}$.

We deal only with *finite* groups.

If there is an edge (x, xa), then there is also an edge $(xa, xaa^{-1}) = (xa, x)$. Hence a Cayley graph is $|S|$-regular and undirected.

Example of a Cayley Graph

Suppose $G = \mathbb{Z}_8$, the set of integers with addition modulo 8 as the group operation, and suppose that $S = \{3, 5\}$. Note that S is symmetric.

The associated Cayley graph is 2-regular. It is shown both as a list of vertices and associated edges, and also in pictorial form on the next slide. Note that the notation $\mathcal{N}(i)$ denotes the set of neighbors of vertex i.

i	0	1	2	3	4	5	6	7
$\mathcal{N}(i)$	$(3,5)$	$(4,6)$	$(5,7)$	$(6,0)$	$(7,1)$	$(0,2)$	$(1,3)$	$(2,4)$

Example of a Cayley Graph (Cont'd)

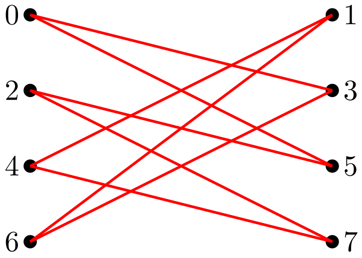


Figure: Cayley graph of $\mathbb{Z}_8$ with $S = \{3, 5\}$. The graph is connected and bipartite.

A Negative Result

(Due to Friedman-Murty-Tillich, (2003)) In any Cayley graph where the group G is Abelian, we have that

$$|\lambda_2| \geq d - O(dn^{-4/d}),$$

where $d = |S|$.

Note that

$$d - O(dn^{-4/d}) \to d^- \text{ as } n \to \infty.$$

Therefore, as $n \to \infty$, the second largest eigenvalue of a Cayley graph not only goes past $2\sqrt{d-1}$, but in fact approaches d from below! Hence it *fails* to be a Ramanujan graph.

Conclusion: If we want to construct Ramanujan graphs with fixed d and arbitrarily large n using the Cayley method, then *the underlying group G cannot be Abelian!*

Outline

General and Projective Linear Groups

Suppose p is a prime number. Then $\mathbb{F}_q := \{0, 1, \cdots, q-1\}$ is a *field* if addition and multiplication are modulo q. Define

$$GL(n, \mathbb{F}_q) := \{M \in \mathbb{F}_q^{n \times n} : \det(M) \neq 0\},$$

the set of $n \times n$ matrices with elements in $\mathbb{F}_q$, whose determinant is nonzero. This is called the **general linear group** of order n over $\mathbb{F}_q$.

General and Projective Linear Groups (Cont'd)

The **projective general linear group** of order n, denoted by $PGL(n, \mathbb{F}_q)$, is obtained by defining $A \sim B$ if $A = cB$ for some $c \neq 0$. Here $A, B \in GL(n, \mathbb{F}_q)$ and $c \in \mathbb{F}_q$.

Fact:

$$|GL(n, \mathbb{F}_q)| = \prod_{i=0}^{n-1} (q^n - q^i),$$

$$|PGL(n, \mathbb{F}_q)| = \frac{1}{q-1} \prod_{i=0}^{n-1} (q^n - q^i).$$

$$|GL(2, \mathbb{F}_q)| = (q^2 - 1)(q^2 - q),$$

$$|PGL(2, \mathbb{F}_q)| = \frac{(q^2 - 1)(q^2 - q)}{q - 1} = q(q^2 - 1).$$

The Lubotzky-Phillips-Sarnak (LPS) Construction

(Lubotzky-Phillips-Sarnak (1988)). Choose p, q to be distinct prime numbers that are $\equiv 1 \bmod 4$. Find all $p+1$ solutions of

$$p = a_0^2 + a_1^2 + a_2^2 + a_3^2,$$

where a_0 is odd and positive, and a_1, a_2, a_3 are even.

Example: Let $p = 5$. Then the solutions are

$$(1, \pm 2, 0, 0), (1, 0, \pm 2), (1, 0, 0, \pm 2).$$

IIT Hyderabad

The LPS Construction (Cont'd)

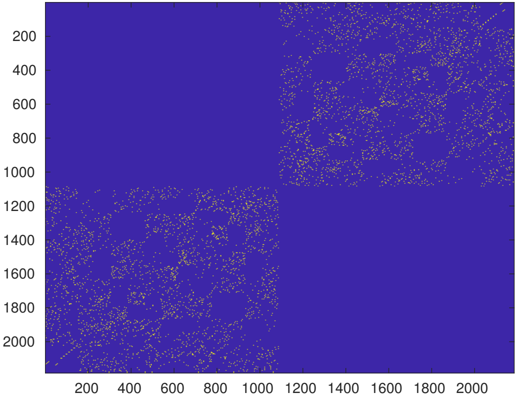
Let the group G be $PGL(2, \mathbb{F}_q)$. Choose $\mathbf{i}$ such that $\mathbf{i}^2 \equiv -1 \mod q$. Define $S = \{M_j\}_{j=1}^{p+1}$, where

$$M_j = \begin{bmatrix} a_{0j} + \mathbf{i}a_{1j} & a_{2j} + \mathbf{i}a_{3j} \\ -a_{2j} + \mathbf{i}a_{3j} & a_{0j} - \mathbf{i}a_{1j} \end{bmatrix} \mod q, j = 1, \ldots, p+1,$$

Then the resulting Cayley graph satisfies the Ramanujan property of degree $p+1$ with $(q(q^2-1))/2$ vertices.

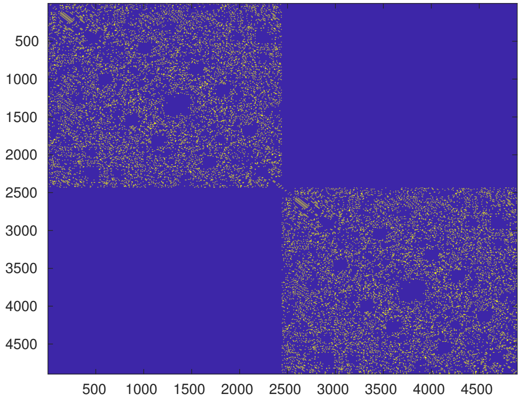
If $p \equiv x^2 \mod q$ for some x (p is a quadratic residue of q), then the graph consists of two disconnected components. Otherwise it is a balanced bipartite graph.

LPS Construction with $p = 37, q = 13$



The graph is bipartite, because 37 it not a quadratic residue of 13.

LPS Construction with $p = 257, q = 17$



$36 \equiv 257 \bmod 17$, so the graph is disconnected.

Outline

Our Construction of Ramanujan Bigraphs

Until now, there has not been a single *explicit* construction of a Ramanujan bigraph!

Let q be a prime number, and let P denote the "right shift" permutation. Define the "array code" matrix

$$B(q,l) = \begin{bmatrix} I_q & I_q & I_q & \cdots & I_q \\ I_q & P & P^2 & \cdots & P^{(q-1)} \\ I_q & P^2 & P^4 & \cdots & P^{2(q-1)} \\ I_q & P^3 & P^6 & \cdots & P^{3(q-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ I_q & P^{(l-1)} & P^{3(l-1)} & \cdots & P^{(q-1)(l-1)} \end{bmatrix}.$$

The associated bipartite graph is (q,l)-biregular.

Our Construction of Ramanujan Bigraphs (Cont'd)

(Doctoral research of Shantanu Prasad Burnwal.)

Theorem

Suppose $l \leq q$. *The matrix* $B(q,l)$ *has a singular value of* $\sqrt{lq}$, $l(q-1)$ *singular values of* $\sqrt{q}$, *and* $l-1$ *singular values of* 0. *Therefore, whenever* $2 \leq q-1$, $B(q,l)$ *defines a Ramanujan bigraph. With* $l = q$, *this defines a (new class of) Ramanujan graphs.*

Our Construction of Ramanujan Bigraphs (Cont'd)

Theorem

Suppose $l > q$.

1. *When $l \bmod q = 0$, in addition B has $(q-1)q$ singular values of $\sqrt{l}$ and $q-1$ singular values of 0.*
2. *When $l \bmod q \neq 0$, let $k = l \bmod q$. Then B has, in addition, $(q-1)k$ singular values of $\sqrt{l+q-k}$ $(q-1)(q-k)$ singular values of $\sqrt{l-k}$, and $q-1$ singular values of 0.*

Therefore, whenever $l > q$, $B(q,l)$ represents a Ramanujan bigraph.

Reprise

We have constructed two explicit classes of Ramanujan bigraphs: Of dimensions $lq \times q^2$ where q is a prime number and $2 \leq l \leq q$, and $q^2 \times lq$ where q is a prime number and $l > q$.

This is the first such construction.

Outline

Outline

Statement of the Matrix Completion Problem

Suppose $X \in \mathbb{R}^{m \times n}$ is unknown, but an upper bound r on its rank is known.

Question: By measuring only $s \ll mn$ elements of X, can we "complete" the remaining elements?

Let

$$\Omega = \{(i_1, j_1), \ldots, (i_s, j_s)\} \subseteq [m] \times [n]$$

denote the "sampling set." (Note: $[n] = \{1, \ldots, n\}$.)

Question: By measuring $X_{ij}, (i, j) \in \Omega$, can we determine X uniquely?

Solution via Nuclear Norm Minimization

Possible Approach: Solve

$$\hat{X} = \operatorname*{argmin}_{Z} \operatorname{rank}(Z) \text{ s.t. } Z_{ij} = X_{ij}, \ \forall (i,j) \in \Omega.$$

This problem is *NP-hard!*

Change the problem to

$$\hat{X} = \operatorname*{argmin}_{Z} \|Z\|_N \text{ s.t. } Z_{ij} = X_{ij}, \ \forall (i,j) \in \Omega,$$

where $\|Z\|_N$ is the **nuclear norm** of Z, i.e., the sum its singular values. $\|\cdot\|_N$ is the "convex relaxation" of the rank – the largest convex function that is everywhere dominated by the rank.

This problem is easy to solve. But when does its solution equal the unknown matrix X?

Sketch of Solution

Choose Ω to be the edge set of a (d_r, d_c)-biregular Ramanujan bigraph. That is, after constructing the graph, sample X_{ij} if and only if vertex $v_i \in \mathcal{V}_r$ is connected to $v_j \in \mathcal{V}_c$.

So only $d_r/m = d_c/n$ elements are sampled.

Then, under suitable conditions, nuclear norm minimization recovers the unknown matrix *exactly* provided $\min\{d_r, d_c\} = O(r^3)$.

Conservatism of Solution

This is *the first result* on matrix completion using a deterministic sampling pattern Ω.

This sufficient condition is *awfully conservative!* How close is it to being necessary?

Answer: Not at all close! Numerical simulations with randomly generated low-rank mmatrices show that

$$r \approx (1/3) \min\{d_r, d_c\}$$

is the true condition. But no proofs as yet!

Outline

Motivation: Missing Measurements

When applying sampling methods to "real" data, we sometimes find that some measurements are "missing."

After choosing Ω to be the edge set of a Ramanujan bigraph, what to do if some elements are "missing"?

Define the set of "missing" measurements $M \in \{0,1\}^{m \times n}$, such that if $M_{ij} = 1$, then it is not possible to measure X_{ij}.

Question: Can we "perturb" the original Ramanujan graph so as to avoid any pairs (i, j) where $M_{ij} = 1$, and still remain a Ramanujan graph with the same degrees?

Main Result

Theorem

Suppose that every row and every columm of M has no more than p entries of one. Define $E \in \{0,1\}^{m \times n}$ to be the biadjacency matrix of the Ramanujan bigraph, and let θ_c denote the maximum inner product between any two columns of E. Then it is possible to perturb the Ramanujan bigraph provided

$$2p \leq n - d_r, d_c - \theta_c.$$

So, provided not too many elements are "missing" in any one row or one column, it is possible to perturb a Ramanujan graph into another one.

IIT Hyderabad

Questions?