Seminar 4 : Ramanujan Graphs and the Matrix Completion Problem.

Saturday March 20 06:30 – 07:45.

Speaker: M. Vidyasagar FRS

Outline:

1. Ramanujan Graphs
   - Some Basic Graph Theory.
   - Ramanujan Graphs and Ramanujan Bigraphs.
2. Construction of Ramanujan Graphs.
   - Cayley Graphs
   - LPS construction of Ramanujan Graphs.
3. Construction of Ramanujan Bigraphs.
4. The Matrix Completion Problem.
   - Problem Statement and Solution.
   - Fault-Tolerant Ramanujan Graphs.

Singular Values and SVD

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

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

This is called the singular value decomposition (SVD) of M. Here $\sigma_1, --, \sigma_m$ are the eigenvalues of MM^T, 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 $G = (V, E)$ consists of a set of "vertices" V, and a set of "edges" $E \subseteq V \times V$. It is said to be undirected, if whenever $(v_i, v_j) \in E$, also $(v_j, v_i) \in E$, A graph is simple if (i) there are no self-loops (edges of the form

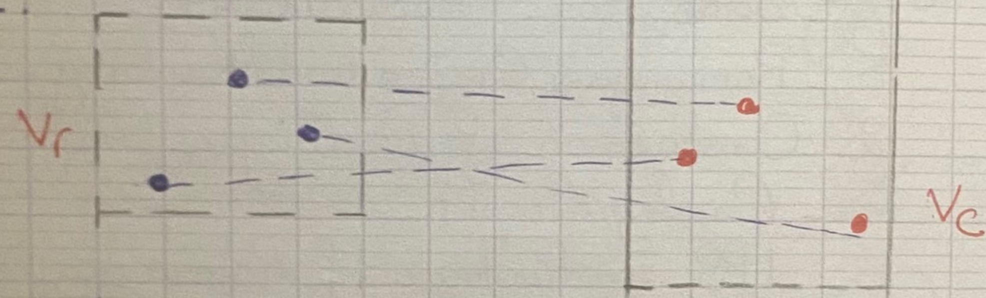
(v_i, v_i)), and no multiple edges between any pair of vertices.
A graph is bipartite if we can partition $\mathcal{V}$ as $\mathcal{V}_r \cup \mathcal{V}_c$ such that:

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

Equivalently,

$$\mathcal{E} \subseteq ((V_r \times V_c) \cup (V_c \times V_r))$$

A bipartite graph is balanced if $|V_r| = |V_c|$, unbalanced otherwise.



No edges between two vertices in V_r, or between two vertices in V_c.

Incidence Matrix and Regularity

The incidence matrix $A \in \{0,1\}^{|V| \times |V|}$ has $a_{ij} = 1$ if $(i,j) \in \mathcal{E}$, and $a_{ij} = 0$ otherwise.
The incidence matrix of bipartite graph looks like
$A = \begin{bmatrix} 0 & B \\ B & 0 \end{bmatrix}$ where $B \in \{0,1\}^{|V_r| \times |V_c|}$ 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 V_r has degree d_r, and all vertices in V_c have degree d_c (which also implies that $d_r |V_r| = d_c |V_c|$).
No edges between two vertices in V_r, or btw two vertices in V_c

Spectra of Graphs

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

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

- If a bipartite graph is (d_r, d_c)-biregular, then
  1) The spectrum of A is $\{\pm\sigma_1, \dots, \pm\sigma_l\}$, 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.

Definition of a Ramanujan Graph

Def: A d-regular graph 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-Boppna Bound (1986)) In any d-regular graph with n vertices, we have that

$$\lim_{n\to\infty} \inf |\lambda_2| \geq 2\sqrt{d-1}$$

Definition of a Ramanujan Bigraph

Def: 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

$$\lim_{\min(n_r, n_c)\to\infty} |\sigma_2| \geq \sqrt{d_r - 1} + \sqrt{d_c - 1}$$

Prevelance of Ramanujan Graphs

Theorem: (Friedman (2008)) Suppose $d \geq 4$ is even and let $\varepsilon > 0$ be arbitrary. Consider a d regular n-vertex graph formed by $d/2$ uniform and independant permutation on $[n]$.
Then $\max\{|\lambda_2|, |\lambda_n|\} \leq 2\sqrt{d-1} + \varepsilon$ w.p. $1 - O(n^{-r})$.
where $r = [(\sqrt{d-1} + 1)/2] + 1$.
Therefore $\max\{|\lambda_2|, |\lambda_n|\} \leq 2\sqrt{d-1} + \varepsilon$ a.a.s.
Simply put: Almost all d-regular graphs "almost" satisfy the Ramanujan property.
No formal Statement: Theorem due to Decker et al. (2008): 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!

Cayley Graphs

Suppose G is a group (not neccessarily Abelian), and $S \subseteq G$ satisfies $a \in S \Rightarrow a^{-1} \in S$. Such that 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.

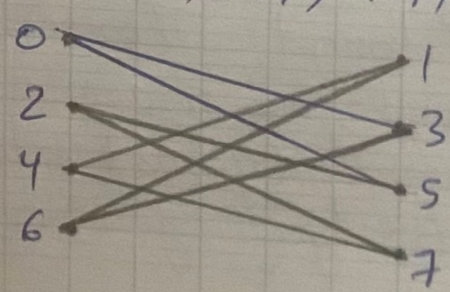
If there is an edge (x, xa), then there is also an edge $(xa, xaa^{-1}) = (xa, x)$. Hence a Cayley graph is $|S|$-regular and undirected.

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 page. Note that the notation $N(i)$ denotes the set of neighbors of vertex i.

i	0	1	2	3	4	5	6	7
$N(i)$	(3,5)	(4,6)	(5,7)	(6,0)	(7,1)	(0,2)	(1,3)	(2,4)



Cayley graph of $\mathbb{Z}_8$ with $\{3,5\}$. This graph is connected & bipartite.

A Negative Result.

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

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

Note that $d - O(dn^{-4/d}) \longrightarrow d^-$ as $n \longrightarrow \infty$

Therefore, as $n \to \infty$, the second largest eingenvalue 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 graph with fixed d and arbitrarily large n using the Cayley method, then the underlying group G cannot be Abelian.

General and Projective Linear Groups.

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

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

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 $\mathbb{F}_q$.

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

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

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

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

The Lubotzky-Phillips-Sarnak (LPS) Construction.

(Lubotzky-Phillips-Sarnak (1998)). 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
$(1, \pm 2, 0, 0), (1, 0, \pm 2), (1, 0, 0, \pm 2)$

The LPS Construction

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

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

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

If $p \equiv x^2 \bmod q$ 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 is not a quadratic residue of 13.

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

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

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

Our Construction of Ramanujan Bigraphs.

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

Suppose $\ell > q$

- When $\ell \bmod q = 0$, in addition B has $(q-1)q$ singular values of $\sqrt{\ell}$ and $q-1$ singular values of 0.
- When $\ell \bmod q \neq 0$, let $k = 1 \bmod q$. Then B has, in addition, $(q-1)k$ singular values of $\sqrt{1+q-k}$ $(q-1)(q-k)$ singular values of $\sqrt{1-k}$, and $q-1$ singular values of 0. Therefore, whenever $\ell > q$, $B(q,\ell)$ represents a Ramanujan bigraph.

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 column 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.

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?

Possible Approach: Solve (Solution via Nuclear Norm Minimization)

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

This problem is NP-hard!

Change the problem to $\hat{X} = \underset{Z}{\operatorname{argmin}} \|Z\|_N$ s.t. $Z_{ij} = X_{ij}$, $\forall (i,j) \in \Omega$.

where $\|Z\|_N$ is the nuclear norm of Z, i.e. the sum of 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 to the unknown matrix X?

Sketch of the 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 $v_i \in V_r$ is connected to $v_j \in V_c$. 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 neccessary? Answer: Not at all close! Numerical simulations with randomly generated low-rank mmatrices show that $r \gtrsim (1/3) \min\{d_r, d_c\}$ is the true condition. But no proofs as yet!