Spectral gap in bipartite biregular graphs and applications

Ioana Dumitriu

Department of Mathematics
University of Washington (Seattle)

Joint work with Gerandy Brito and Kameron Harris

ICERM workshop on Optimal and Random Point Configurations
February 27, 2018

1. Intro: expanders and bipartite graphs
2. Random Bipartite Biregular Graphs are almost Ramanujan
3. Applications

Definitions

- A simple bipartite graph consists of a set of vertices partitioned into two classes, and a set of edges which occur solely between the classes.
- Sometimes denoted as $G = (X, Y, E)$, where X, Y are vertex classes and E is the set of edges.
- Notation: $|X| = m$, $|Y| = n$.

Definitions

- A biregular bipartite graph has the property that all vertices in the same class have the same degree
- Notation: $|X| = m$, $|Y| = n$, d_1 for the common degree of class X, d_2 for the degree of class Y.
- Note that $md_1 = nd_2$.

Importance; applications

- A number of important and interesting classes of graphs are bipartite and some are biregular (trees, even cycles, median graphs, hypercubes).

Importance; applications

- A number of important and interesting classes of graphs are bipartite and some are biregular (trees, even cycles, median graphs, hypercubes).
- Applications include projective geometry (Levi graphs), coding theory (yielding factor codes and Tanner codes, more on that later), computer science (Petri nets, assignment problems, community detection), signal processing (matrix completion).

Adjacency matrix

- For a bipartite graph, the adjacency matrix A with $A_{ij} = \delta_{i \sim j}$ looks like
$$A = \left[\begin{array}{cc} 0 & X \\ X^T & 0 \end{array} \right] .$$

Adjacency matrix

- For a bipartite graph, the adjacency matrix A with $A_{ij} = \delta_{i \sim j}$ looks like

$$A = \begin{bmatrix} 0 & X \\ X^T & 0 \end{bmatrix} .$$

- As a consequence, their spectrum is symmetric.

Expanders

- Graphs with high connectivity and which exhibit rapid mixing; sparse analogues to the complete graph

Expanders

- Graphs with high connectivity and which exhibit rapid mixing; sparse analogues to the complete graph
- Of particular interest in CS and coding theory (from mixing to design of error-correcting codes)

Expanders

- Graphs with high connectivity and which exhibit rapid mixing; sparse analogues to the complete graph
- Of particular interest in CS and coding theory (from mixing to design of error-correcting codes)
- Random regular graphs (uniformly distributed) are classical (and best-known) examples of such expanders; expanding properties characterized by the *spectral gap* of the adjacency matrix.

Expanders

- Graphs with high connectivity and which exhibit rapid mixing; sparse analogues to the complete graph
- Of particular interest in CS and coding theory (from mixing to design of error-correcting codes)
- Random regular graphs (uniformly distributed) are classical (and best-known) examples of such expanders; expanding properties characterized by the *spectral gap* of the adjacency matrix.
- Uniform distribution important in making assertions like “almost all regular graphs”

Random regular graphs

- For an (n, d) random regular graph (n vertices, each of degree d), the eigenvalues denoted $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$,

Random regular graphs

- For an (n, d) random regular graph (n vertices, each of degree d), the eigenvalues denoted $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$,
- $\lambda_1 = d$ (trivial)

Random regular graphs

- For an (n, d) random regular graph (n vertices, each of degree d), the eigenvalues denoted $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$,
- $\lambda_1 = d$ (trivial)
- Quantity of interest is the *second largest eigenvalue*, defined as $\eta = \max\{|\lambda_2|, |\lambda_n|\}$.

Random regular graphs

- McKay ('81) calculated asymptotical empirical spectrum distribution (Kesten-McKay law); yields lower bound on η of $2\sqrt{d-1} - o(1)$

Random regular graphs

- McKay ('81) calculated asymptotical empirical spectrum distribution (Kesten-McKay law); yields lower bound on η of $2\sqrt{d-1} - o(1)$
- Work on lower bounding η also by Alon-Boppana ('86), upper bounding η by Friedman ('03). Uniformly random regular graphs are *almost Ramanujan*, i.e.,

$$\eta \in [2\sqrt{d-1} - o(1), 2\sqrt{d-1} + \epsilon] \, .$$

a.a.s. as $n \to \infty$.

Random regular graphs

- McKay ('81) calculated asymptotical empirical spectrum distribution (Kesten-McKay law); yields lower bound on η of $2\sqrt{d-1} - o(1)$
- Work on lower bounding η also by Alon-Boppana ('86), upper bounding η by Friedman ('03). Uniformly random regular graphs are *almost Ramanujan*, i.e.,

$$\eta \in [2\sqrt{d-1} - o(1), 2\sqrt{d-1} + \epsilon] \ .$$

  a.a.s. as $n \to \infty$.
- Recently, Bordenave ('15) tightened Friedman's proof to $\eta = 2\sqrt{d-1} + o(1)$.

Work on bipartite biregular graphs

- Bipartite biregular graphs not quite expanders; mixing keeps track of class, but can mix quickly within class

Work on bipartite biregular graphs

- Bipartite biregular graphs not quite expanders; mixing keeps track of class, but can mix quickly within class
- Studied in most contexts where regular graphs appear

Work on bipartite biregular graphs

- Bipartite biregular graphs not quite expanders; mixing keeps track of class, but can mix quickly within class
- Studied in most contexts where regular graphs appear
- Again, uniform distribution important.

Work on bipartite biregular graphs

- Largest eigenvalue $\lambda_1 = \sqrt{d_1 d_2}$, matched by $\lambda_n = -\sqrt{d_1 d_2}$.
- Godsil and Mohar ('88) calculated asymptotical empirical spectrum distribution when $m/n = d_2/d_1 \to \gamma \in [0, 1]$ (Marčenko-Pastur-like);

Work on bipartite biregular graphs

- Largest eigenvalue $\lambda_1 = \sqrt{d_1 d_2}$, matched by $\lambda_n = -\sqrt{d_1 d_2}$.
- Godsil and Mohar ('88) calculated asymptotical empirical spectrum distribution when $m/n = d_2/d_1 \to \gamma \in [0, 1]$ (Marčenko-Pastur-like);
- Their work shows lower bound on $\lambda_2 \geq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} - o(1)$

Work on bipartite biregular graphs

- Largest eigenvalue $\lambda_1 = \sqrt{d_1 d_2}$, matched by $\lambda_n = -\sqrt{d_1 d_2}$.
- Godsil and Mohar ('88) calculated asymptotical empirical spectrum distribution when $m/n = d_2/d_1 \to \gamma \in [0, 1]$ (Marčenko-Pastur-like);
- Their work shows lower bound on $\lambda_2 \geq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} - o(1)$
- Feng and Li ('96) and Li and Sole ('96) also worked on lower bound

Work on bipartite biregular graphs

- Largest eigenvalue $\lambda_1 = \sqrt{d_1 d_2}$, matched by $\lambda_n = -\sqrt{d_1 d_2}$.
- Godsil and Mohar ('88) calculated asymptotical empirical spectrum distribution when $m/n = d_2/d_1 \to \gamma \in [0, 1]$ (Marčenko-Pastur-like);
- Their work shows lower bound on $\lambda_2 \geq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} - o(1)$
- Feng and Li ('96) and Li and Sole ('96) also worked on lower bound
- Matching upper bound: work by Brito, D., Harris (2018).

Random bipartite, biregular graphs (RBBG)

- Would like to study the uniform distribution on RBBG, but it's hard to work with

Random bipartite, biregular graphs (RBBG)

- Would like to study the uniform distribution on RBBG, but it's hard to work with
- Instead, use the configuration model (Bender, Canfield '78, Bollobas '80)
  - "asymptotically uniform" (contiguous to the uniform one), anything happening a.a.s. in configuration model happens a.a.s. in the uniform one

Main Result

- Let $G(d_1, d_2, m, n)$ be a random bipartite graph generated with the configuration model.
- Largest modulus eigenvalues are $\pm\lambda = \pm\sqrt{(d_1 - 1)(d_2 - 1)}$. What is the third largest?

Main Result

- Let $G(d_1, d_2, m, n)$ be a random bipartite graph generated with the configuration model.
- Largest modulus eigenvalues are $\pm\lambda = \pm\sqrt{(d_1-1)(d_2-1)}$. What is the third largest?

Theorem (BDH'18)

$\lambda_3 \leq \sqrt{d_1-1} + \sqrt{d_2-1} + o(1)$, with high probability.

Main Result

- Let $G(d_1, d_2, m, n)$ be a random bipartite graph generated with the configuration model.
- Largest modulus eigenvalues are $\pm\lambda = \pm\sqrt{(d_1-1)(d_2-1)}$. What is the third largest?

Theorem (BDH'18)

$\lambda_3 \leq \sqrt{d_1-1} + \sqrt{d_2-1} + o(1)$, *with high probability.*

- Note sum instead of product.

Main Result

- Let $G(d_1, d_2, m, n)$ be a random bipartite graph generated with the configuration model.
- Largest modulus eigenvalues are $\pm\lambda = \pm\sqrt{(d_1-1)(d_2-1)}$. What is the third largest?

Theorem (BDH'18)

$\lambda_3 \leq \sqrt{d_1-1} + \sqrt{d_2-1} + o(1)$, *with high probability.*

- Note sum instead of product.
- Proof follows in the footsteps of Bordenave ('15)

Random bipartite, biregular graphs (RBBG)

- Bounding second eigenvalues general idea:
    - If v is the norm-one eigenvector for λ_1, subtract vv^T from A to make λ_2 largest eigenvalue; $\tilde{A} = A - vv^T$

Random bipartite, biregular graphs (RBBG)

- Bounding second eigenvalues general idea:
  - If v is the norm-one eigenvector for λ_1, subtract vv^T from A to make λ_2 largest eigenvalue; $\tilde{A} = A - vv^T$
  - If A positive definite find upper bound on $||\tilde{A}||$ by $||\mathrm{Tr}(\tilde{A}^m)^{1/m}||$ as m grows large

Random bipartite, biregular graphs (RBBG)

- Bounding second eigenvalues general idea:
  - If v is the norm-one eigenvector for λ_1, subtract vv^T from A to make λ_2 largest eigenvalue; $\tilde{A} = A - vv^T$
  - If A positive definite find upper bound on $||\tilde{A}||$ by $||\mathrm{Tr}(\tilde{A}^m)^{1/m}||$ as m grows large
  - If $\tilde{A}$ not positive definite, examine $\tilde{A}^l(\tilde{A}^*)^l$ instead (in some form)

Random bipartite, biregular graphs (RBBG)

- Bounding second eigenvalues general idea:
  - If v is the norm-one eigenvector for λ_1, subtract vv^T from A to make λ_2 largest eigenvalue; $\tilde{A} = A - vv^T$
  - If A positive definite find upper bound on $||\tilde{A}||$ by $||\text{Tr}(\tilde{A}^m)^{1/m}||$ as m grows large
  - If $\tilde{A}$ not positive definite, examine $\tilde{A}^l(\tilde{A}^*)^l$ instead (in some form)
  - Applied in many contexts, with success

Random bipartite, biregular graphs (RBBG)

- Bounding second eigenvalues general idea:
  - If v is the norm-one eigenvector for λ_1, subtract vv^T from A to make λ_2 largest eigenvalue; $\tilde{A} = A - vv^T$
  - If A positive definite find upper bound on $||\tilde{A}||$ by $||\mathrm{Tr}(\tilde{A}^m)^{1/m}||$ as m grows large
  - If $\tilde{A}$ not positive definite, examine $\tilde{A}^l(\tilde{A}^*)^l$ instead (in some form)
  - Applied in many contexts, with success
  - Not here. Sadly, $\tilde{A}^m$ is too hard to work with (too much "chaff")

Non-backtracking matrix

- Idea: Examine instead the "non-backtracking" matrix B, whose rows/columns indexed by *edges*, and $B_{ef} = 1$ iff $e = (v_1, v_2)$, $f = (v_2, v_3)$ with $v_1 \neq v_3$. Non-symmetric!

Non-backtracking matrix

- Idea: Examine instead the "non-backtracking" matrix B, whose rows/columns indexed by *edges*, and $B_{ef} = 1$ iff $e = (v_1, v_2)$, $f = (v_2, v_3)$ with $v_1 \neq v_3$. Non-symmetric!
- Can relate the eigenvalues of B to those of the adjacency matrix A via the Ihara-Bass formula

$$\det(B - \lambda I) = (\lambda^2 - 1)^{|E|-n} \det(D - \lambda A + \lambda^2 I) ,$$

with $|E|$ = number of edges, D the diagonal matrix of degrees.

Non-backtracking matrix

- Idea: Examine instead the "non-backtracking" matrix B, whose rows/columns indexed by *edges*, and $B_{ef} = 1$ iff $e = (v_1, v_2)$, $f = (v_2, v_3)$ with $v_1 \neq v_3$. Non-symmetric!
- Can relate the eigenvalues of B to those of the adjacency matrix A via the Ihara-Bass formula

$$\det(B - \lambda I) = (\lambda^2 - 1)^{|E|-n} \det(D - \lambda A + \lambda^2 I) ,$$

  with $|E|$ = number of edges, D the diagonal matrix of degrees.
- Spectral gap for B may yield spectral gap for A (works here).

RBBG spectral gap proof ideas

- Show that B has spectral gap. (Easier to do so than for A; yet very technical.)

RBBG spectral gap proof ideas

- Show that B has spectral gap. (Easier to do so than for A; yet very technical.)
- Subtract off a "centering" matrix that has the effect of zeroeing the two largest eigenvalues to get $\bar{B}$. Bound highest eigenvalue of $\bar{B}$ by

$$\mathbb{E}\left(||\bar{B}^{\ell}||^{2k}\right) \leq \mathbb{E}\left(\operatorname{Tr}\left((\bar{B}^{\ell})(\bar{B}^{\ell})^*\right)^k\right) .$$

RBBG spectral gap proof ideas

- Show that B has spectral gap. (Easier to do so than for A; yet very technical.)
- Subtract off a "centering" matrix that has the effect of zeroeing the two largest eigenvalues to get $\bar{B}$. Bound highest eigenvalue of $\bar{B}$ by

$$\mathbb{E}\left(||\bar{B}^{\ell}||^{2k}\right) \leq \mathbb{E}\left(\mathrm{Tr}\left((\bar{B}^{\ell})(\bar{B}^{\ell})^*\right)^k\right) .$$

- Linear algebra (Bordenave '15) yields

$$\lambda_2 \leq \max_{x \perp v_1, v_2, ||x||_2 = 1} \left(||B^l x||\right)^{1/l} .$$

RBBG spectral gap proof ideas

- Show that B has spectral gap. (Easier to do so than for A; yet very technical.)
- Subtract off a "centering" matrix that has the effect of zeroeing the two largest eigenvalues to get $\bar{B}$. Bound highest eigenvalue of $\bar{B}$ by

$$\mathbb{E}\left(||\bar{B}^{\ell}||^{2k}\right) \leq \mathbb{E}\left(\mathrm{Tr}\left((\bar{B}^{\ell})(\bar{B}^{\ell})^*\right)^k\right) .$$

- Linear algebra (Bordenave '15) yields

$$\lambda_2 \leq \max_{x \perp v_1, v_2, ||x||_2 = 1} \left(||B^l x||\right)^{1/l} .$$

- Show that $B^l = \tilde{B}^l(1 + o(1))$

RBBG spectral gap proof ideas

- Show that B has spectral gap. (Easier to do so than for A; yet very technical.)
- Subtract off a "centering" matrix that has the effect of zeroeing the two largest eigenvalues to get $\bar{B}$. Bound highest eigenvalue of $\bar{B}$ by

$$\mathbb{E}\left(||\bar{B}^{\ell}||^{2k}\right) \leq \mathbb{E}\left(\operatorname{Tr}\left((\bar{B}^{\ell})(\bar{B}^{\ell})^*\right)^k\right) .$$

- Linear algebra (Bordenave '15) yields

$$\lambda_2 \leq \max_{x \perp v_1, v_2, ||x||_2 = 1} \left(||B^l x||\right)^{1/l} .$$

- Show that $B^l = \tilde{B}^l(1 + o(1))$
- The rest is (roughly) sophisticated circuit-counting.

RBBG spectral gap proof ideas

- We call a graph l-tangle-free if all vertex neighborhoods of size up to l contain at most one cycle.

RBBG spectral gap proof ideas

- We call a graph l-tangle-free if all vertex neighborhoods of size up to l contain at most one cycle.
- Roughly speaking, this means that cycles are far from each other.

RBBG spectral gap proof ideas

- We call a graph l-tangle-free if all vertex neighborhoods of size up to l contain at most one cycle.
- Roughly speaking, this means that cycles are far from each other.
- We can show $G(m, n, d_1, d_2)$ is $c \log_d n$-tangle-free with high probability ($d = \max\{d_1, d_2\}$).

RBBG spectral gap proof ideas

- We call a graph l-tangle-free if all vertex neighborhoods of size up to l contain at most one cycle.
- Roughly speaking, this means that cycles are far from each other.
- We can show $G(m, n, d_1, d_2)$ is $c \log_d n$-tangle-free with high probability ($d = \max\{d_1, d_2\}$).
- This, together with non-backtracking feature, helps with circuit-counting.

RBBG spectral gap proof ideas

- Dominant terms come from circuits that are exactly trees traversed once forward and once backward.

RBBG spectral gap proof ideas

- Dominant terms come from circuits that are exactly trees traversed once forward and once backward.
- Like in the moment method proof of the semicircle law.

RBBG spectral gap proof ideas

- Dominant terms come from circuits that are exactly trees traversed once forward and once backward.
- Like in the moment method proof of the semicircle law.
- There, existence of an edge repeated more than once brought one less choice of vertex and edges appearing once cancelled term

RBBG spectral gap proof ideas

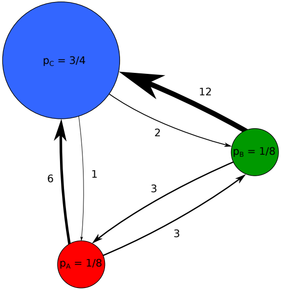
- Dominant terms come from circuits that are exactly trees traversed once forward and once backward.
- Like in the moment method proof of the semicircle law.
- There, existence of an edge repeated more than once brought one less choice of vertex and edges appearing once cancelled term
- Same here about multiple repetitions, but no exact cancellation for edges appearing only once; finer estimates needed due to lack of independence. Still, doable.

□

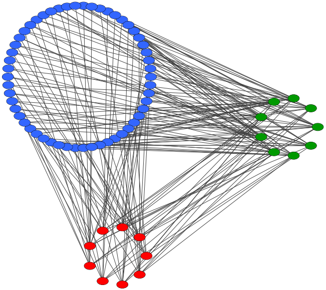
Applications for RBBG: community detection

– frame graphs: given a small, edge-weighted graph, use it to define community structure in a larger, random graph. Each graph is represented by a vertex, the weights in the frame define the number of edges between classes. Quasi-regular.

A Frame



B Random regular frame graph



Applications for RBBG: community detection

– Such graphs are known as *equitable* graphs, as per Mohar '91, Newman & Martin '10, Barucca '17, Meila & Wan '15. Objects of study: **community detection** (with lots of assumptions).

Applications for RBBG: community detection

- Such graphs are known as *equitable* graphs, as per Mohar '91, Newman & Martin '10, Barucca '17, Meila & Wan '15. Objects of study: **community detection** (with lots of assumptions).
- Using a very general theorem of Meila '15 (under certain conditions, the highest eigenvalues of the random graphs are those of the frame), we concluded that **community detection is possible in such graphs** (removing assumptions).

Applications for RBBG: community detection

- Such graphs are known as *equitable* graphs, as per Mohar '91, Newman & Martin '10, Barucca '17, Meila & Wan '15. Objects of study: **community detection** (with lots of assumptions).
- Using a very general theorem of Meila '15 (under certain conditions, the highest eigenvalues of the random graphs are those of the frame), we concluded that **community detection is possible in such graphs** (removing assumptions).
- Conditions not optimal, but a starting point for further study.

Applications for RBBG: expander codes

- Expander codes (Tanner codes) introduced in Tanner, ’62
- Linear error-correcting codes whose parity-check matrix encoded in an expander graph

Applications for RBBG: expander codes

- Expander codes (Tanner codes) introduced in Tanner, '62
- Linear error-correcting codes whose parity-check matrix encoded in an expander graph
- Using Tanner '81, Janwa and Lal '03, one may construct codes with decent relative minimum distance and rate by using bipartite biregular graphs.

Applications for RBBG: matrix completion

– Idea: given Y a large matrix with "low complexity" (e.g. sparse, low-rank, etc.) observe some of Y's entries, and based on them find Y' such that $||Y - Y'||$ is small (or even 0) in some norm $|| \cdot ||$. (Netflix problem; Amazon, etc.)

Applications for RBBG: matrix completion

– Idea: given Y a large matrix with “low complexity” (e.g. sparse, low-rank, etc.) observe some of Y’s entries, and based on them find Y' such that $||Y - Y'||$ is small (or even 0) in some norm $|| \cdot ||$. (Netflix problem; Amazon, etc.)
– Matrix version of compressed sensing (Candès and Plan ’10, Candès and Tao, ’10).

Applications for RBBG: matrix completion

- Idea: given Y a large matrix with "low complexity" (e.g. sparse, low-rank, etc.) observe some of Y's entries, and based on them find Y' such that $||Y - Y'||$ is small (or even 0) in some norm $||\cdot||$. (Netflix problem; Amazon, etc.)
- Matrix version of compressed sensing (Candès and Plan '10, Candès and Tao, '10).
- Recent idea: sample entries according to a random regular graph (Heiman et al '14, Bhojanapalli and Jain '14, Gamarnik et al '17).

Applications for RBBG: matrix completion

- Idea: given Y a large matrix with "low complexity" (e.g. sparse, low-rank, etc.) observe some of Y's entries, and based on them find Y' such that $||Y - Y'||$ is small (or even 0) in some norm $||\cdot||$. (Netflix problem; Amazon, etc.)
- Matrix version of compressed sensing (Candès and Plan '10, Candès and Tao, '10).
- Recent idea: sample entries according to a random regular graph (Heiman et al '14, Bhojanapalli and Jain '14, Gamarnik et al '17).
- If one uses a RBBG instead (simple-mindedly), improvement in bounds by a factor of 2 (as compared to Heiman et al. '14; studying Gamarnik et al. '17). Possibly more?...