RESEARCH STATEMENT FOR LINH V. TRAN

1. Introduction

I am interested in probabilistic and additive combinatorics. My primary interest is probabilistic methods for investigating properties of random structures , focusing on the models of random matrices, random graphs and hypergraphs. I am also interested in some additive number theory problems, in particular ones involving long arithmetic progression in sumset. Below , there is a brief introduction to topics that I've been working on.

- **Random matrices** In a joint work with V. Vu and K. Wang, we proved the semi-circular law for random d-regular graph model in the case d tends to infinity as n does. Our result complements the McKay law [15], which applied for the case d is an absolute constant. We also obtained a few upper bounds on the infinity norm of eigenvectors of Erdös-Renyi random graph $G(n, p)$ where $p = \omega(\log n/n)$.
- **Random hypergraph**. Take n random boxes with axis-parallel edges inside the unit cube $[0, 1]^d$, the piercing number $\tau_d(n)$ is the minimum number of points needed to pierce all boxes. Using hypergraph setting, I [16] was able to prove a near sharp estimation for $\tau_d(n)$: for any $d \geq 2$

  $$\Omega_d(\sqrt{n}(\log n)^{d/2-1}) = \tau_d(n) = O_d(\sqrt{n}(\log n)^{d/2-1} \log \log n).$$

  The proof was based on an existing result about perfect matching in deterministic hypergraph and a bounded martingale concentration inequality. Further improvement is in progress.
- **Additive number theory** Given a set A of integers, $S_A := \{\sum_{b \in B} : B \subset A\}$ be the set of all sums of distinct elements of A. In a joint work with V. Vu and P. Matchett Wood [18], we proved a 20-year-old conjecture due to N. Alon [1] about S_A, using a strutural theorem of A. Sárközy. I am interested in further applications of this structural approach.

2. Random matrices

I am currently interested in the universality of the global and local eigenvalue statistics of several random $n \times n$ Hermitian matrix models. Typically the entries of the considered matrix are complex random variables which have the same first few moments but not necessarily are independent or have identical distribution. Famous examples of these models are the Wigner random matrices (which include the Gaussian Unitary Ensemble GUE) whose entries have mean zero and variance 1, and the adjacency matrix of Erdös-Renyi random graph $G(n,p)$ or random regular graph $G_{n,d}$. The eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$ thus are real random variables.

2.1. **Global properties.** One famous example of **global properties** of the eigenvalues is the Wigner semi-circular law. Define the empirical spectral distribution (ESD) of the matrix M as

$$F_n^{\mathbf{M}}(x) = \frac{1}{n}|\{1 \leq j \leq n : \lambda_j(M) \leq x\}|,$$

then if M_n is Wigner random matrices

$$\lim_{n \tilde{o} \infty} F_n^{\mathbf{M_n}}(x) = \int_{-2}^{x} \rho_{sc}(y)dy,$$

in the sense of probability, where

$$\rho_{sc} = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2} & |x| \leq 2 \\ 0 & |x| > 2 \end{cases}$$

The semi-circular law can be extended without difficulty for the adjacency matrix of Erdös Renyi random graph $G(n,p)$, except for the case $np = O(1)$, using the same moment method.

Theorem 2.2. *For $p = \omega(\frac{1}{n})$, the empirical spectral distribution (ESD) of the matrix $\frac{1}{\sqrt{n}\sigma}A_n$ converges in distribution to the semicircle distribution which has a density $\rho_{sc}(x)$ with support on $[-2,2]$,*

$$\rho_{sc}(x) := \frac{1}{2\pi}\sqrt{4-x^2}.$$

This method, however, does not apply for the adjacency matrix of regular random graph $G_{n,d}$, since the entries of $G_{n,d}$ are not independent. About 30 years ago, McKay [15] proved that if d is fixed, and $n \to \infty$, then the limiting density function is

$$f_d(x) = \begin{cases} \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)}, & \text{if } |x| \leq 2\sqrt{d-1}; \\ \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to verify that as $d \to \infty$, if we normalize the variable x by $\sqrt{d-1}$, then the above density converges to the semicircle distribution on $[-2,2]$. It is thus natural to conjecture that Theorem 2.2 holds for $G_{n,d}$ with $d \to \infty$. Let A'_n be the adjacency matrix of $G_{n,d}$, and set

$$M'_n = \frac{1}{\sqrt{\frac{d}{n}(1-\frac{d}{n})}}(A'_n - \frac{d}{n}J).$$

Conjecture 2.3. *If $d \to \infty$ then the ESD of $\frac{1}{\sqrt{n}}M'_n$ converges to the standard semicircle distribution.*

Recently, Dimitriu and Pal [6] showed that this conjecture is true for d tending to infinity slowly, $d = n^{o(1)}$. Their method does not extend to larger d. In a joint work with V. Vu and K. Wang [17], I was able to prove the conjecture in full generality.

Theorem 2.4. *If d tends to infinity with n, then the empirical spectral distribution of $\frac{1}{\sqrt{n}}M'_n$ converges in distribution to the semicircle distribution.*

Our method is very different from that of [6]. Theorem 2.4 is a direct consequence of the following stronger result, which shows convergence at small scales. For an interval I let N'_I be the number of eigenvalues of M'_n in I.

Theorem 2.5. *(Concentration for ESD of $G_{n,d}$). Let $\delta > 0$ and consider the model $G_{n,d}$. If d tends to ∞ as $n \to \infty$ then for any interval $I \subset [-2, 2]$ with length at least $\delta^{-4/5} d^{-1/10} \log^{1/5} d$, we have*

$$|N'_I - n \int_I \rho_{sc}(x)dx| < \delta n \int_I \rho_{sc}(x)dx$$

with probability at least $1 - O(\exp(-cn\sqrt{d}\log d))$.

2.6. **Local properties.** My long-term project is investigating the universality phenomenon for sparse Hermitian random matrix models. The universality conjecture, which is supported by enormous numerical evidence, states that the local behaviour of eigenvalues of random matrices does not depend on the distribution of the entries, as long as the distribution satisfies some simple properties. It means that we can obtain some common local properties for a large class of random matrices. The universality conjecture has been a major topic in random matrices theory, and it was proved for many special properties and matrix classes. Recently, Tao and Vu proved a breakthrough result, the Four Moment Theorem, which states that for Wigner random matrices (whose entries has mean zero and variance one) the first four moments of the entries decide the local behaviours of the eigenvalues. I plan to generalize this result for a wider class of sparse random matrices, namely the one whose entries have variance tends to zero, in particular the matrices associated with random graph models. The sparsity causes problem of controlling the magnitude of normalized elements, which often exceed some upper bounds required in the proof of the original Four moment theorem.The proof will contain a few key steps that needed to be adjusted. In a joint work with V. Vu and K. Wang [17], we proved one of those key steps: the delocalization of eigenvectors for the Erdös-Renyi random graph.

Theorem 2.7. *(Infinity norm of eigenvectors) Let $p = \omega(\log n/n)$ and let A_n be the adjacency matrix of $G(n, p)$. Then there exists an orthonormal basis of eigenvectors of A_n, $\{u_1, \dots, u_n\}$, such that for every $1 \le i \le n$, $||u_i||_\infty = o(1)$ almost surely.*

In the near future this project will be my main focus.

3. Random hypergraphs

3.1. **Background.** The model of random hypergraph is a natural extension of the Erdös-Rényi model of random graph. In a k-uniform random graph $G_k(n,p)$, the vertex set has n vertices, and any subset of k vertices forms an edge independently and randomly with probability p. Random hypergraph is ideal for modeling high dimensional random discrete structures.

Let $\mathbf{U}_d$ be the unit hypercube in $\mathbf{R}^d$, where d is a constant: $\mathbf{U}_d := [0,1]^d$. We generate a d-dimensional random box by taking the product of d independent random sub-intervals of $[0,1]$, where each random interval is determined by two random (end-) points, each chosen independently with respect to the uniform measure on $[0,1]$. Consider a family of n (independently) random boxes. Our investigation is motivated by the following basic questions:

What is the size of the largest sub-family of pairwise disjoint boxes (or packing number, denoted by $\nu_d(n)$)?

What is the minimum number of points one needs to pierce all the boxes (or piercing number, denoted by $\tau_d(n)$)?

These quantities are of fundamental interest and have been studied for a large variety of hypergraphs, deterministic and random alike. It is useful to notice that the piercing number is at least the matching number, by definition. There is a conjecture, made by Beck and other about 20 years ago.

Conjecture 3.2. *Let $\mathcal{F}$ be finite family of axis parallel boxes in $\mathbf{R}^d$. Then the matching and piercing numbers of $\mathcal{F}$ are the same, up to a constant factor (which may depend on the dimension).*

The conjecture is proved directly for $d = 1$. However, for high dimensional cases, it is better to estimate $\tau_d(n)$ and $\nu_d(n)$ separately.

3.3. **Recent work and further directions.** About ten years ago, Coffman, Lueker, Spencer and Winkler studied $\nu_d(n)$ [5]. They showed

Theorem 3.4. *For $d = 2$, $\nu_2(n) = \Theta(\sqrt{n})$. For $d \geq 3$*

$$\Omega(\sqrt{n}) \leq \nu_d(n) \leq O(\sqrt{n \log^{d-1} n}). \tag{1}$$

In [16], I was able to prove a nearly sharp estimate for the piercing number

Theorem 3.5. *For any fixed $d \geq 2$, we have, almost surely, that*

$$\Omega_d(\sqrt{n}(\log n)^{d/2-1}) = \tau_d(n) = O_d(\sqrt{n}(\log n)^{d/2-1}\log\log n). \tag{2}$$

Our approach was trying to classify the boxes in to classes according to volume and apply appropriate treatment for each class. The proof was based on a result on large matching of hypergraph by V. Vu [19] and the bounded martingale concentration inequality.

Since $\tau(n) \geq \nu(n)$, Theorem 3.5 improved (the general case of) Theorem 3.4.

Corollary 3.6. *For $d \geq 3$*

$$\Omega_d(\sqrt{n}) = \nu_d(n) = O_d(\sqrt{n}(\log n)^{d/2-1}\log\log n). \tag{3}$$

The extra factor $\log\log n$ in (2) looks redundant and I believe it can be removed. On the other hand, if Conjecture 3.2 is true, then we should be able to improve the lower bound for $\nu_d(n)$. The way that the lower bound in (1) was proved is to show that a set of random boxes are disjoint because (piecewise) they are contained in a set of disjoint deterministic boxes. I am trying to find a way to guarantee the disjointness by probabilistic construction, maybe by using independent set in random hypergraph. All in all, the following conjecture looks plausible.

Conjecture 3.7. *For any fixed $d \geq 2$, almost surely,*

$$\nu_d(n) = \tau_d(n) = \Theta_d(\sqrt{n}(\log n)^{d/2-1}).$$

4. Additive number theory: On Alon's conjecture

4.1. **Background.** For n a large positive integer and m an integer between n and n^2, we define $f(n,m)$ to be the maximum cardinality of a set $A \subset \{1, 2, \dots, n\}$ such that no subset $B \subset A$ satisfies the condition $\sum_{b\in B} b = m$. In 1986, Erdős and Graham [7] observed that $f(n,m) \geq \left(\frac{1}{2} + o(1)\right)\frac{n}{\log n}$.

For s a positive integer not dividing m, it is clear that $f(n,m) \geq \lfloor \frac{n}{s} \rfloor$, since any sum of elements of the set $\left\{s, 2s, 3s, 4s, \dots, \lfloor \frac{n}{s} \rfloor s\right\}$ cannot divide (and hence cannot equal) m. Letting $\operatorname{snd}(m)$ denote the smallest positive integer that does not divide m, we thus have

$$\lfloor \frac{n}{\operatorname{snd} m} \rfloor \leq f(n,m). \tag{4}$$

By the prime number theorem, we know that $\operatorname{snd}(m) \le (2+o(1))\log n$, and so Inequality (4) matches the lower bound observed by Erdős and Graham [7]. Alon later made the following conjecture, which essentially states that the lower bound is asymptotically sharp.

Conjecture 4.2. [1] *If $n^{1.1} \le m \le n^{1.9}$, then*

$$f(n,m) = (1+o(1))\frac{n}{\operatorname{snd}(m)}$$

as $n \to \infty$.

4.3. **Recent work and further directions.** The previous best results towards Conjecture 4.2 were due to Alon and Freiman [2], who proved the conjecture for $3n^{5/3+\epsilon} < m < \frac{n^2}{20\log^2 n}$ and also due to Lipkin [14] proved for $cn\log^6 n < m < \frac{n^{3/2}}{\log^3 n}$. Therefore the gap $\frac{n^{3/2}}{\log^3 n} < m < 3n^{5/3+\epsilon}$ is still open to be filled.
In [18], Van H. Vu, Phillip Matchett Wood and I proved the Conjecture 4.2 in full using elementary methods along with a theorem of Sárközy on arithmetic progressions in iterated sumsets (see [20]). Our main result is the following, which implies both Alon and Freiman's result and Lipkin's result:

Theorem 4.4. *For any constants $c > 0$ and $\epsilon > 0$, if n is a sufficiently large positive integer and if m and n satisfy*

$$cn(\log n)^{1+\epsilon} \le m \le \frac{n^2}{9\log^2 n},$$

then we have

$$f(n,m) = (1+o(1))\frac{n}{\operatorname{snd}(m)}.$$

In fact, the methods we used also suffice to prove the "inverse result" below (Theorem 4.5), which shows that if $|A| \le f(n,m)$ (but is not too much smaller than $f(n,m)$) and if m cannot be represented as a sum of distinct elements of A, then A must contain primarily integers congruent to 0 mod d, where d is an integer that does not divide m. This essentially says that the example giving the lower bound in Inequality (4) is the only way for a reasonably large subset of $\{1,2,3,\dots,n\}$ to avoid containing a set of distinct elements that sum to m.

Theorem 4.5. *Let c, δ, ϵ_1, and ϵ_2 be positive constants such that $0<\epsilon_1<\epsilon_2$, and let m and n be integers satisfing*

$$cn(\log n)^{1+\epsilon_2} \le m \le \frac{\delta^2 n^2}{8(\log n)^{2+2\epsilon_1}},$$

where we assume that n is sufficiently large. If

$$\frac{\delta n}{(\log n)^{1+\epsilon_1}} \le |A| \le f(n,m)$$

and if no subset $B \subset A$ satisfies $\sum_{b\in B} b = m$, then A contains at least $(1+o(1))\,|A|$ elements that are congruent to $0 \bmod d$, where d is an integer that does not divide m.

The result of Sárközy on arithmetic progression in iterated sumsets is closely related to recent results of Szemerédi and Vu and I am interested in inverstigating how these structural results can be applied to other problems.

References

[1] N. Alon, *Subset sums*, J. Number Theory **27** (1987), no. 2, 196–205.

[2] N. Alon and G. Freiman, *On sums of subsets of a set of integers*, Combinatorica **8** (1988), no. 4, 297–306.

[3] C. Cooper, A. Frieze, M. Molloy and B. Reed, *Perfect matchings in random r-regular, s-uniform hypergraphs*, Combin. Probab. Comput. 5 (1996), 114.

[4] C. Cooper, A. Frieze and B. Reed, *Random regular graphs of non-constant degree: connectivity and Hamiltonicity*, Combin. Probab. Comput. 11 (2002), no. 3, 249–261.

[5] E. G. Coffman, G.S. Lueker, J. Spencer and P.M. Winkler, *Packing random rectangles.* Probab. Theory Relat. Fields 120, 585-599 (2001).

[6] I. Dumitriu and S. Pal. *Sparse regular random graphs: spectral density and eigenvectors.* Arxiv preprint arXiv:0910.5306, 2009.

[7] Paul Erdős, *Some problems and results on combinatorial number theory*, Graph theory and its applications: East and West (Jinan, 1986), Ann. New York Acad. Sci., vol. 576, New York Acad. Sci., New York, 1989, pp. 132–145.

[8] A. Frieze, L. Tran, V. Vu, *Sandwiching random graphs: Treatment for high degree regular random graph*, in preparation.

[9] J. H. Halton, *On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals.* Numer. Math. 2 1960 84–90.

[10] J.H. Kim, B. Sudakov and V. H. Vu, *One the asymmetry of random graphs and random regular graphs*, Random Structures and Algorithms 21(2002), 216-224.

[11] J.H. Kim and V.H. Vu, Sandwiching random graphs: Universality between random graph models, Advances in Mathematics 188 (2004) 444-469.

[12] J.H. Kim, V. H. Vu, *Concentration of multi-variate polynomials and its applications*, Combinatorica, 20 (3) (2000), 417-434.

[13] M. Krivelevich, B. Sudakov, V. H. Vu and N. Wormald, *Random regular graphs of high degree*, Random Structures Algorithms 18 (2001), no. 4, 346363.

[14] E. Lipkin, *On representation of rth powers by subset sums*, Acta Arith. **52** (1989), no. 4, 353–365.

[15] B.D. McKay. *The expected eigenvalue distribution of a large regular graph.* Linear Algebra and its Applications, 40:203216, 1981.

[16] Linh V. Tran *Piercing random boxes*, To appear in Random Strutures and Algorithms (2010).

[17] Linh V. Tran, Van H. Vu, Ke Wang, *Sparse random graphs: Eigenvalues and Eigenvectors*, submitted.

[18] Linh V. Tran, Van H. Vu, Phillip Matchett Wood, *On a conjecture of Alon*, J. Num. Theory. Vol. 129, Issue 11, (2009) 2801-2807.

[19] V. H. Vu, *New bounds on nearly perfect matchings in hypergraphs: higher codegrees do help*, Random Structures and Algorithms, 17 (1)(2000), 29-63.

[20] A. Sárközy, *Finite addition theorems. II*, J. Number Theory **48** (1994), no. 2, 197–218.

[21] N.C. Wormald, *Models of random regular graphs*, In Surveys in Combinatorics, 1999, J.D. Lamb and D.A. Preece, eds, pp. 239298.

Department of Mathematics, Rutgers, Piscataway, NJ 08854

E-mail address: linhtran@math.rutgers.edu