Finding Structures in Large-scale Graphs

Sang Peter Chin† Eizabeth Reilly† Linyuan Lu‡

†Johns Hopkins University Applied Physics Laboratory
‡University of South Carolina

ABSTRACT

One of the most vexing challenges of working with graphical structures is that most algorithms scale poorly as the graph becomes very large. The computation is extremely expensive even for polynomial algorithms, thus making it desirable to devise fast approximation algorithms. We herein propose a framework using advanced tools[1–6] from random graph theory and spectral graph theory to address the quantitative analysis of the structure and dynamics of large-scale networks. This framework enables one to carry out analytic computations of observable network structures and capture the most relevant and refined quantities of real-world networks.

1. INTRODUCTION

In this information age that we live in, there is a growing need to respond to the challenges to make sense of large-scale graphs that often arise from various communication networks, sensor networks, social networks, etc. In response to such challenges, we herein propose a framework using advanced tools[1–6] from random graph theory and spectral graph theory to address the quantitative analysis of the structure and dynamics of large networks. This framework enables one to carry out analytic computations of observable network structures and capture the most relevant and refined quantities of real-world networks.

One of the most vexing challenges of working with graphical structures is that most algorithms scale poorly as the graph becomes very large. The computation is extremely expensive even for polynomial algorithms, thus making it desirable to devise fast approximation algorithms. From the view of information theory, one would like to capture the essential structure of the network and discard information that amounts to random noise.

Our approach is based on the celebrated *Szemerédi regularity lemma*, which has proved to be an important tool by asserting the existence of certain subgraphs in any sufficiently large graph. It roughly states that every large enough graph can be divided into subsets (or *partitions*) of about the same size so that the edges between different subsets behave almost randomly; in other words, any sufficiently large/dense graph can be approximated by a general random graph $G(n, P)$. In recent years, the regularity lemma and its variations have emerged as potentially powerful tools to solve numerous algorithmic and combinatorial problems, *e.g.*, approximation of Max-Cut in dense graphs,[27] property testing in graphs,[24] and fast Boolean matrix multiplication.[25] These problems are closely related to algorithmic questions involving the regularity lemma, such as efficient regularity testing and regular partitioning.

If these regular partitions of a given graph can be found efficiently, it would then imply efficient (and often parallel and distributed among partitions) methods to compute myriad graph properties of interest, *e.g.*, number of triangles, maximum cliques, *etc.*, just to name a few. Unfortunately, there is no known computationally efficient algorithm to find such partitions that scale for large graphs. In fact, Alon, *et al.*[22] observed that the problem of regularity testing is *co*-NP. However, it is precisely here, that we believe ideas from spectral graph theory can help. We will show that the investigation of the dominant eigenvalues of the adjacency matrix of a graph can often lead to computationally efficient and scalable methods to approximate these partitions and suggest a few applications.

2. TECHNICAL APPROACH

2.1 The Origin of Our Approach

Let us start with some well-known concepts from the general random graph theory. For a fixed m, a probability matrix $P = (p_{ij})$ is an $m \times m$ symmetric matrix satisfying $0 \leq p_{ij} \leq 1$ for $1 \leq i \leq j \leq m$. Let $[n] := \{1, 2, \ldots, n\}$ denote the set of the first n numbers. Given a partition $\mathcal{P}$ of $[n]$ into m sets: $[n] = V_1 \cup V_2 \cup \ldots \cup V_m$, we define the general random graph $G(n, P, \mathcal{P})$ on the vertex set $[n]$ as follows. For any $u \in V_i$ and $v \in V_j$, a pair (u, v) is an edge of $G(n, P, \mathcal{P})$ with probability p_{ij} independently. When $\mathcal{P}$ is an equitable partition, *i.e.*, the sizes to two parts differ by at least 1, we simply write $G(n, P)$ for short.

If the number of blocks, m, is 1, then this model reduces to the Erdős-Renyi random graph model $G(n, p)$ (see[6]). If $m = 2$ and $p_{11} = p_{22} = 0$, then it is a model for a random bipartite graph. If $m = n$ and $p_{ij} = w_i w_j \frac{1}{\sum_{i=1}^{n} w_i}$, then it is the model $G(w_1, \ldots, w_n)$ of a random graph with expected degree sequence. The properties of $G(w_1, \ldots, w_n)$ have been studied extensively.[1–5] It is a good model for capturing power law graphs, but less flexible for modeling general graphs. However, the techniques used in $G(w_1, \ldots, w_n)$ can still be applied to the general random graph model.

Let Δ (or δ) be the maximum (or minimum) expected degree of the random graph $G(n, P, \mathcal{P})$ respectively. Let A be the adjacency matrix and D be the diagonal matrix of degrees. The (normalized) Laplacian is given by $L = I - D^{-1/2} A D^{-1/2}$. We have the following proposition.

PROPOSITION 2.1. *Let $G := G(n, P, \mathcal{P})$ be the general random graph associated to the partition $\mathcal{P}$ and the probability matrix P. Then with probability $1 - o(1)$, we have*

1. *If $\Delta \gg \ln^4 n$, then all but at most m eigenvalues of the adjacency matrix A have absolute values at most $(2 + o(1))\sqrt{\Delta}$.*
2. *If $\delta \gg \max\{m, \ln^4 n\}$, then all but almost m eigenvalues λ_i of the Laplacian L satisfy $|\lambda_i - 1| \leq (2 + \sqrt{m} + o(1))\frac{1}{\sqrt{\delta}}$.*

Proof: This is a corollary of Lu-Peng's result on edge-independent random graphs. Note that in random graph $G(n, P, \mathcal{P})$ edges are mutually independent. Let $\bar{A} = \mathrm{E}(A)$ the expectation of A. Lu-Peng[32] proved that if $\Delta \gg \ln^4 n$ then

$$\|\lambda_i(A) - \lambda_i(\bar{A})\| \leq (2 + o(1))\sqrt{\Delta}.$$

Observe that $\bar{A}$ is a rank m matrix since $\bar{A}_{uv} = p_{ij}$ if $u \in V_i$ and $v \in V_i$. All but at most m of $\lambda_i(\bar{A})$'s are zeros. Thus, all but at most m eigenvalues of $\lambda_i(A)$'s are at most $(2 + o(1))\sqrt{\Delta}$.

The proof of the statement of Laplacian is similar. Let $\bar{D}$ be the diagonal matrix of the expected degrees and $\bar{L} = I - \bar{D}^{-1/2} \bar{A} \bar{D}^{-1/2}$. Lu-Peng[32] proved that

$$|\lambda_i(L) - \lambda_i(\bar{L})| \leq (2 + \sqrt{m} + o(1))\frac{1}{\sqrt{\delta}}.$$

Since $I - \bar{L}$ has rank k, all but almost m eigenvalues λ_i of the Laplacian L satisfy $|\lambda_i - 1| \leq (2 + \sqrt{m} + o(1))\frac{1}{\sqrt{\delta}}$. □

Now, assume $n \gg m$. The model $G(n, P, \mathcal{P})$ is complicated but one can still determine the graph's properties such as connectivity, diameter, maximal cliques, the number of triangles, *etc.* using the probability matrix P and the partition $\mathcal{P}$. Put another way, from the view of information theory, the general random graph $G := G(n, P, \mathcal{P})$ can be reduced to the pair $(P, \mathcal{P})$, and because m is much smaller than n, any algorithm that runs on $(P, \mathcal{P})$ will be significantly faster than if it were to run on G itself. The pair $(P, \mathcal{P})$ can be viewed the backbone of G. *One of the key enabling ideas that allows us to formulate our approach is that Proposition 2.1 shows that all but m eigenvalues (of the adjacency matrix) are $O(\sqrt{n})$.* There are two obvious questions to ask:

Question 1: *Is this model general enough to capture a variety of graphs?*

Question 2: *How can we get $(P, \mathcal{P})$ from a given graph G?*

Szemerédi regularity lemma: The first question can be answered by the celebrated *Szemerédi regularity lemma*. The Szemerédi regularity lemma roughly states that every sufficiently large graph can be divided into subsets of about the same size so that the edges between different subsets behave almost randomly; in other words, any large/dense graph can be approximated by a general random graph $G(n, P)$.

More precisely, let $G = (V, E)$ be a simple graph. For two disjoint vertex sets X and Y, the edge density of the pair (X, Y) is

$$d(X, Y) := \frac{|E(X, Y)|}{|X|\,|Y|}$$

where $E(X, Y)$ denotes the cut set, i.e., edges having one end vertex in X and one in Y. For any $\epsilon > 0$, a pair of vertex sets X and Y is called *ϵ-regular* if for all subsets A of X and B of Y satisfying $|A| \geq \varepsilon|X|$ and $|B| \geq \varepsilon|Y|$, we have

$$|d(X, Y) - d(A, B)| \leq \epsilon.$$

A partition of the vertex set of G into k sets $V_0, V_1, \ldots, V_k$ is called an *ϵ-regular partition*, if $|V_0| \leq \epsilon n$, $|V_1| = |V_2| = \cdots = |V_k|$, and all but ϵk^2 of the pairs V_i, V_j, $1 \leq i < j \leq k$, are ϵ-regular. The Szemerédi regularity lemma Regularity lemma can be stated as the following:

LEMMA 2.1. *For every $\epsilon > 0$ and positive integer m there exists an integer M such that if G is a graph with at least M edges, there exists an integer k in the range $m \leq k \leq M$ and an ϵ-regular partition of the vertex set of G into k sets.* Tao[11] proved two extensions of the regularity lemma—the probabilistic version and the information-theoretic version. Both versions are more powerful than the original version; for example, they can be used to prove the regularity lemma for hypergraphs.[8]

Now, for the second question above, the bound M for the number of parts in the partition of the graph given by Szemerédi's regularity lemma is very large (it is about a $1/\epsilon^C$-level iterated exponential of m for some absolute constant C.) Gowers[7] found examples of graphs for which M does indeed grow very fast and is at least as large as a $log(1/\epsilon)$-level iterated exponential of m. *This makes the algorithm associated to its proof intractable in practice. New methods are needed to find a similar partition of a given graph G.*

2.2 A New Way Forward via Spectral Graph Theory

Consider a family of graphs $\{G_n\}$, where G_n is a graph on n vertices. List all eigenvalues of the adjacency matrix of G_n such that the absolute values are in decreasing order:

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|.$$

An increasing function $f(n)$ is a *coarse spectral bound* of G_n if

$$|\{i \colon |\lambda_i(G_n)| \geq f(n)\}| = o(n).$$

The coarse spectral radius is not unique. Here are some properties.

Monotone: If $f(n) \geq g(n)$ for sufficiently large n and $g(n)$ is a coarse spectral bound of G_n, then $f(n)$ is also a coarse spectral bound of G_n.

Continuity: Suppose two graphs G_n and G'_n only differ by $o(dn)$-edges. If $f(n)$ is a coarse spectral bound of G, then $f(n) + o(d)$ is a coarse spectral bound of G'.

The "least" coarse spectral bound is called *essential spectral radius* of G. Roughly speaking, an essential spectral radius is the absolute maximum of all but $o(n)$ eigenvalues of G. By the monotonicity property, the essential spectral radius is well-defined up to a lower order additive term. By Proposition 2.1, the general random graph $G(n, P, \mathcal{P})$ has essential spectral radius $O(\sqrt{n})$, provided $m = o(n)$. We have the following theorem.

THEOREM 2.2. *For any slowly increasing function $f(n)$ (growing to infinity) and any sequence of graphs $\{G_n\}$, the essential spectral radius of G_n is at most $f(n)\sqrt{n}$.*

Proof: Let A be the adjacency matrix of G_n and $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of A in the decreasing order of absolute values. We have

$$\sum_{i=1}^{n} \lambda_i^2 = \text{trace}(A^2) = 2|E(G)|. \tag{1}$$

For any integer $k \leq n$, we have

$$k|\lambda_k|^2 \leq \sum_{i=1}^{k} \lambda_i^2 < 2|E(G)| < n^2. \tag{2}$$

Hence

$$|\lambda_k| < \frac{n}{\sqrt{k}}.$$

Choosing $k = \frac{n}{f^2(n)} = o(n)$, we have $\lambda_k < f(n)\sqrt{n}$. In other words, the number of eigenvalues exceeding $f(n)\sqrt{n}$ in absolute value is $o(n)$. The proof of theorem is finished. □

The upper bound of essential spectral radius in Theorem 2.2 cannot be replace by $C\sqrt{n}$ for $C < 2$. The essential spectral radius of Erdős-Renyi random graph $G(n,p)$ is at least $(2 + o(1))\sqrt{np}$, which is greater than $(2-\epsilon)\sqrt{n}$ as p approaches 1. In fact, we make the following conjecture.

CONJECTURE 1. *For any family of graphs $\{G_n\}$, the essential spectral radius of G_n is at most $(2+o(1))\sqrt{n}$.*

If the conjecture is true, it is then best possible.

The eigenvalues within the essential spectral radius are *less significant* while those outside the essential spectral radius are *significant*. The subspace spanned by the eigenvectors corresponding to the less significant eigenvalues are corresponding to the "random part" of the graph; which are less important. One should focus on significant eigenvalues and their associated eigenvectors. In general, it is hard to determine the essential spectral radius. However, if our aforementioned conjecture is true, then significant eigenvalues separate from less significant eigenvalues. One can compute the most significant eigenvalue; and then the second most significant eigenvalue; and so on until the big gap is reached. Let $\lambda_1, \ldots, \lambda_m$ be the m significant eigenvalues where m is determined by the algorithm. Let α_i $(1 \leq i \leq m)$ be the corresponding eigenvectors. Let Q be the matrix defined as

$$Q = \sum_{i=1}^{m} \lambda_i \alpha_i \alpha_i'.$$

The matrix Q has rank $m \ll n$. *We believe that Q contains the essential information of G.* We may group the similar entries of Q together to get the partition $\mathcal{P}$ of vertices and the probability matrix P. Once we obtain a partition $\mathcal{P}$ and the probability matrix P, we can use them to estimate many graph parameters of G efficiently.

The efficiency of the approach above depends on the rank of Q, equivalently, the number of significant eigenvalues. Roughly speaking, Semeredi's regularity Lemma implies that any dense graphs can be approximated by some random graph $G(n, P, \mathcal{P})$ with a finite number of partitions. Base on this observation, we made the following conjecture.

CONJECTURE 2. *For any $\epsilon > 0$, there is a constant $C = f(\epsilon)$ such that the following holds. For any dense graph G_n with at least $\epsilon\binom{n}{2}$ edges, the number of significant eigenvalues is at most C.*

There do exist versions of the regularity Lemma for sparse graphs, however, they are less powerful. Our approach should still work for sparse graphs, in principle, but we have to be careful where to cut the significant eigenvalues versus the less significant ones. We have the following conjecture on the essential spectral radius for sparse graphs.

CONJECTURE 3.

For any family of sparse graphs $\{G_n\}$ with maximum degree Δ_n, the essential spectral radius of G_n is $O(\sqrt{\Delta_n})$ if $\Delta \gg \ln n$.

Alternatively, we can use a matrix other than the adjacency matrix A. For example, the Laplacian matrix for undirected graphs is defined as

$$\mathcal{L} = I - D^{-1/2} A D^{-1/2}.$$

Here D is diagonal matrix of degrees in G. Another example is using the page-rank kind of matrix

$$\frac{c}{n}J + (1-c)D^{-1/2}AD^{-1/2}.$$

2.3 Simulation of our new Approach

To illustrate our idea, we first generate a general random graph $G := G(n, P)$ as follows. Let $n = 180$ and P be the (3×3)-matrix defined as

$$P = \begin{pmatrix} 0.2 & 0.4 & 0 \\ 0.4 & 0.4 & 0.4 \\ 0 & 0.4 & 0.6 \end{pmatrix}.$$

The graph G can be viewed as a (random) blow-up of the weighted graph of Figure 1.

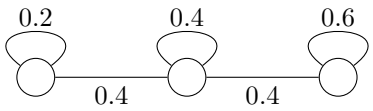
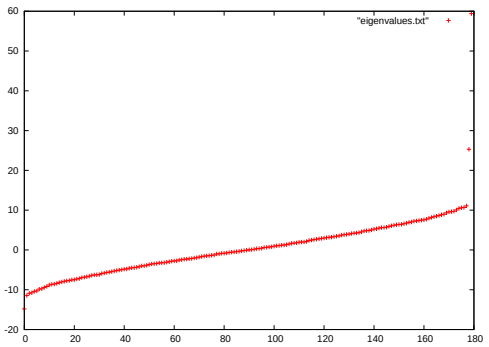
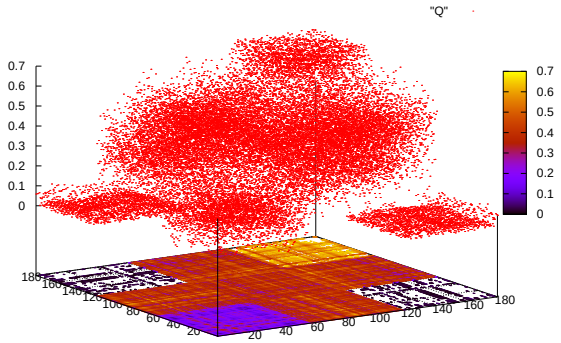


Figure 1. The probability matrix P is viewed as a weighted graph.

The eigenvalues of the adjacency matrix of G are computed by *Maple* and then are drawn using *Gnuplot* (see Figure 2(a)).



(a) Eigenvalues of G.



(b) The matrix Q.

Figure 2. The distribution of eigenvalues of G, (a), and the distribution of all entries in the matrix Q, (b).

From Figure 2(a), it is clear that there are only three significant eigenvalues: $\lambda_1 = 59.407153, \lambda_2 = 25.300653, \lambda_3 = -14.809906$. The rest of eigenvalues have absolute values all less than 12. For $i = 1, 2, 3$, let α_i be the eigenvector associated to λ_i. We computed $Q := \sum_{i=1}^{3} \lambda_i \alpha_i \alpha_i'$. The matrix Q is plotted in Figure 2(b).

From Figure 2(b), it is clear that the vertices are partitioned into three groups, roughly agreeing with the original partition $\mathcal{P}$. This partition divides the matrix Q into 9 blocks (see the contour map at the bottom in Figure 2(b).) By averaging over each block, we get the following matrix $\bar{P}$, which almost recovers the matrix P.

$$P = \begin{pmatrix} 0.198167 & 0.395086 & 0.002750 \\ 0.395086 & 0.383684 & 0.390717 \\ 0.002750 & 0.390717 & 0.595066 \end{pmatrix}.$$

Now we consider the problem of counting the number of triangles in G. In the case that each block in partition has the same size t, using equation (4) from Section 3, this can be simplified as

$$\#\text{ of triangles } \approx \frac{t^3}{6}\text{trace}(P). \tag{3}$$

In our example, $t = 60$. We use $\bar{P}$ to approximate P. We get $\text{trace}(P) \approx \text{trace}(\bar{P}) = 0.9982612835$ and conclude the number of triangles is about 35937. By heuristic search, the actual number of triangles in G is 35058. The relative error is about 2.5%.

3. APPLICATIONS

Application 1: Estimating the number of triangles in G. The number of triangles in G can be approximated using the probability matrix and the partition $\mathcal{P}$ as follows.

$$\#\text{ of triangles } \approx \frac{1}{6}\sum_{i,j,k}|V_i||V_j||V_k|p_{ij}p_{jk}p_{ik}. \tag{4}$$

This is an $O(m^3)$ algorithm. This is efficient since $m \ll n$.

Application 2: Find the maximum clique in G. There are many ways that a clique K_k can intersect the blocks of $\mathcal{P}$; for each intersection pattern, we can compute the probability. If all probabilities are small for all patterns, then decrease the value of k until we find the intersection pattern with the maximum and non-trivial probability. For this particular pattern, we go back to the subgraph of G on the blocks which have nontrivial intersections in this pattern. Then search the clique K_k as usual. Note that $m \ll n$, the first step is efficient while the running time of the second step is also reduced because the subgraph we considered is usually much smaller than G.

Justification: The most costly step in the algorithm is to find the (absolute) largest eigenvalues/eigenvectors. Computing the pairs of eigenvalues and eigenvectors is a well-studied problem. For a single pair, it can be computed using $O(n)$-time. We can use parallel algorithms to speed up the computation. (See[12–15]).

Application 3: Dot Product Graph and Partitions

Next in order for us to compute more non-trivial features of social networks, we consider using *dot product representations* to obtain a low dimensional representation of a simple graph, G, or where G is one of the partitions we obtain using our approach described above. Specifically, the goal of dot product representations is, given G, to assign a vector to each vertex of G such that the set of vectors capture the most important and essential structure of the graph. Scheinerman & Tucker develop and analyze an algorithm for doing this in.[18]

procedure ITERATIVE-EIGENVALUE-METHOD(A, d, t) of[18]
Require: A, the adjacency matrix of a given graph G with n vertices. Dimension d of the representation. Threshold t for the stopping condition.
  Let $D := 0$ be an $n \times n$ diagonal matrix. $X := 0$, a $d \times n$ matrix.
  while $\left|\left|A + D - X^T X\right|\right| > t$ **do**
    Compute the singular value decomposition $(A + D) = U^T \Lambda U$. Define $\hat{U}$ to be the the first d rows of U and $\hat{\Lambda}$ to be the first d rows and columns of Λ. Let $X = \hat{\Lambda}^{1/2}\hat{U}$.
    For $j = 1, \ldots, n$, update the j, j entry of D to be $\mathbf{x}_j \cdot \mathbf{x}_j$ where $\mathbf{x}_j$ is the j^{th} column of X.
  end whilereturn Columns of X, $\mathbf{x}_1, \ldots, \mathbf{x}_n$ as dot product representation of vertices 1 through n of G.
end procedure

Random dot product graphs were first introduced by Kraetzel, *et al.*, in.[16] This model, which randomly selects vectors and uses their dot products to construct a graph, has been studied and extended by several others. In,[18] the reverse approach is taken. Given a graph, $G = (V, E)$, where edge $ij \in E$, $i \neq j$, has weight w_{ij}, we look for the set of n vectors in dimension d that would approximately generate edges with these weights. In other words, we want to find vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ such that $\mathbf{x}_i \cdot \mathbf{x}_j$ is as close to w_{ij} as possible. In,[18] Scheinerman & Tucker use an iterative eigenvalue method to discover an optimal vector representation in $\mathbb{R}^d$.

There is a natural way to apply dot product representations to the Szemerédi partition found in the previous section. The iterative eigenvalue method is efficient for any graph with an adjacency matrix of a size allowing for fast matrix-vector multiplication (as discussed later). Thus, it is often the case that the method is applied to large, sparse graphs. However, the Szemerédi partition results in an $m \times m$ probability matrix P where m is small. Thus, the method could also be applied to this dense matrix P. The diagonal of P could be viewed as an intelligent "guess" of what D should be and the rest of P would be the adjacency matrix A. By applying the iterative eigenvalue method to $P = A + D$, one would obtain vectors $\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_m}$. Every vertex in V_1 would have vector $\mathbf{v_1}$ as its representation, every vertex in V_2 would have vector $\mathbf{v_2}$ as its representation, and so on. This geometric representation would allow for the application of fast geometric algorithms to discover properties of the graph.

Also note that the current theory supporting the discovery of the Szemerédi partition above applies to dense graphs. Therefore, dot product representations serve two additional purposes in our approach. First, because the methods are efficient for large, sparse graphs, they can be used in those cases as an alternative to the Szemerédi partition for discovering the backbone of G. Second, when the regularity lemma theory is extended for sparse graphs, we may compare the Szemerédi partition with the partition discovered using dot product representations and an angular k-means algorithm. Discovery of similar partitions would validate both methods. Because of these additional purposes, we will study dot product representations in their own right, such as what dimension d is necessary to capture the "important" information of the graph, as well as in the context of the Szemerédi partition. We seek to extend the work in[18] to further explore how this geometric graph representation improves the scalability of graph algorithms and analysis. Minimally, we will consider the information lost through such a low dimensional representation and how to adapt the algorithm in[18] to include the dynamic nature of graphs.

4. CONCLUSIONS

In this paper we have presented a graph partition methodology based on spectral-theoretic understanding of Szemerédi's lemma with simulation and testing that scales to massive streaming data sets. This is a first step towards establishing a distributed system for automatically identifying various structures from large-scale graphs. Ultimately, we believe this will allow for greater and more complex automatic analyses of various networks (social networks, communication networks, etc.).

REFERENCES

[1] F. Chung and L. Lu, Average distances in random graphs with given expected degree sequences, *Internet Mathematics* **1** (1), 2003, 91-114.

[2] F. Chung, *Spectral Graph Theory*, AMS Publications, 1997.

[3] F. Chung, L. Lu, and V. Vu, The spectra of random graphs with given expected degrees, *Internet Mathematics* **1** (2004) 257–275.

[4] F. Chung, and L. Lu, Connected components in random graphs with given expected degree sequences, *Annals of Combinatorics* **6**, (2002), 125–145,

[5] F. Chung, and L. Lu, *Complex Graphs and Networks*, AMS Publications, 2006.

[6] P. Erdős and A. Rényi, On Random Graphs I, *Publ. Math Debrecen* **6**, (1959), 290-297.

[7] Gowers, W. T. , Lower bounds of tower type for Szemerdi's uniformity lemma, *Geometric and Functional Analysis* **7** (2), (1997) 322337.

[8] Gowers, W. T., Hypergraph regularity and the multidimensional Szemerdi theorem,*Annals of Mathematics. Second Series* **166** (3), (2007) 897946,

[9] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Polska Akademia Nauk. Instytut Matematyczny. Acta Arithmetica* **27**,(1975) 199245.

[10] E. Szemerédi, Endre, Regular partitions of graphs, Problèmes combinatoires et théorie des graphes, *Colloq. Internat. CNRS*, **260**, Paris: CNRS, (1978) pp. 39940.

[11] Tao, Terence, A variant of the hypergraph removal lemma, *Journal of Combinatorial Theory. Series A* **113** (7), (2006) 12571280.

[12] J. W. Demmel, M. T. Heath, and H. A. van der Vorst, Parallel numerical linear algebra, *Acta Numerica* **2** (1993), 111-197.

[13] P. J. Eberlein and H. Park, Efficient implementation of Jacobi algorithms and Jacobi sets on distributed memory architectures, *J. Parallel Distrib. Comput.*, **8** (1990), 358-366.

[14] G. W. Stewart, A Jacobi-like algorithm for computing the Schur decomposition of a non-Hermitian matrix, *SIAM J. Sci. Stat. Comput.* **6** (1985) 853-864.

[15] G. W. Stewart, A parallel implementation of the QR algorithm, *Parallel Comput.* **5** (1987) 187-196.

[16] M. Kraetzl and C. Nickel, and E. Scheinerman, *Random dot product graphs: a model for social networks*, (2009), (submitted).

[17] D. Marchette and C. Priebe, *Predicting unobserved links in incompletely observed networks*, Computational Statistics and Data Analysis, **52** (2008), pp. 1373–1386.

[18] E. Scheinerman and K. Tucker, *Modeling graphs using dot product representations*, Computational Statistics **25** (2010), pp. 1–16.

[19] K. Tucker, *Exact and asymptotic dot product representations of graphs*, PhD Thesis, Johns Hopkins University (2007).

[20] S.J. Young, *Random dot product graphs: a flexible model for complex networks*, PhD Thesis, Georgia Institute of Technology (2008).

[21] S.J. Young and E. Scheinerman, *Directed Random Dot Product Graphs*, Internet Mathematics **5** (2008), pp. 91–112.

[22] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster. The algorithimic aspects of the regularity lemma. *Journal of Algorithms*, volume 16, no. 1, 1994, 80–109.

[23] N. Alon, A. Coja-Oghlan, H. Hán, M. Kang, V. Rödl, and M. Schacht. Quasi-Randomness and Algorithmic Regularity for Graphs with General Degree Distributions. *SIAM J. on Computing*, Volume 39, no. 6, 2010, 2336–2362.

[24] N. Alon and A. Shapira. Homomorphisms in graph property testing. *In Topics in discrete mathematics*, volume 26 of *Algorithms Combin.*, Springer, Berlin, 2006, 281–313.

[25] N. Bansal and R. Williams. Regularity Lemmas and Combinatorial Algorithms. *In EEE Symposium on Foundations of Computer Science*, 2009, 745–754.

[26] B. Csaba and A. Pluhŕ. Weighted Regularity Lemma with Applications. *preprint*, arXiv:0907.0245v2.

[27] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, volume 19, no. 2, 1999, 175–220.

[28] A. Frieze and R. Kannan. A simple algorithm for constructing Szemerédis regularity partition. Electronic Journal of Combinatorics, volume 6, no. 1, 1999.

[29] Y. Kohayakawa, V. Rödl, and L. Thoma. An optimal algorithm for checking regularity. *SIAM J. Comput*, volume 32, no. 5, 2003, 1210–1235.

[30] J. Komloś, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. The regularity lemma and its applications in graph theory. *Theoretical aspects of computer science*, Lecture Notes in Comput. Sci., volume 2292, Springer, Berlin, 2002, 84–112.

[31] Blow up lemma. 1997. Combinatorica.

[32] L. Lu and X. Peng, Spectra of edge-independent random graphs, preprint, http://arxiv.org/pdf/1204.6207v1.