Pseudorandom and Quasi-random Graphs

Yi Sun

January 13, 2008

Abstract

This paper presents an overview of the theory of pseudorandom and quasi-random graphs. We explore Thomason's concept of jumbled graphs as a model for pseudorandom graphs. We then discuss the more general quasi-random graphs given by Chung, Graham, and Wilson and use these two models of pseudorandom graphs to describe some basic properties of the resulting graphs.

1 Introduction

The study of random graphs provides a method to understand the behavior of a so-called typical graph. Traditionally, the use of the standard graph model $G(n, p)$ allows us to obtain various existence proofs and general bounds; these are properties that apply, in some sense, to all graphs. Here, we are reducing possibly complex statements about large classes of concrete graphs to simple propositions about random graphs. In a similar vein, we can reverse the process and attempt to find concrete graphs that are realizations of the nice properties of random graphs. This is often done by constructing specific classes of graphs that somehow mimic randomness. Constructions of this type are known as *pseudo-random* graphs; examples include the Paley graph P_Q based on the quadratic residue character, the Cayley graph $\Gamma(G, S)$ derived from the generating set of a finite group, and the projective norm graph $NG_{p,t}$ [3].

These families of graphs provide good approximations to the properties of standard random graphs, but they give no measure of how closely any specific type of pseudorandom graph resembles the model $G(n, p)$. The first approach to this issue, by Thomason, based such a comparison on the edge distribution of the graph via the definition of a jumbled graph [6]. This was later generalized by Chung, Graham, and Wilson to the more general notion of quasi-randomness, which they proved to be equivalent to a large and varied number of conditions [2]. In this paper, we give an overview of these approaches and present some typical properties of such pseudorandom graphs.

In Section 2, we give a more precise definition of a pseudorandom graph and discuss its connection with Thomason's jumbled graphs. In Section 3, we present Chung, Graham, and Wilson's generalization to quasi-random graphs and discuss their seven equivalent conditions. In Section 4, we give a typical connection between these relatively new models of pseudorandom graph and the more traditional examples.

2 Pseudorandom and Jumbled Graphs

Given a graph $G = (V, E)$, call it (p, α)-*jumbled* if for every induced subgraph $U \subset V$, we have

$$\left| e(U) - p\binom{|U|}{2} \right| \leq \alpha |U|. \tag{1}$$

This concept of jumbled graphs, first introduced by Thomason in [6], provides a measure of how uniformly the edges of G are distributed. Here, the parameter p varies with the number of edges in the graph, while α is a measure of how uniform the distribution is. First, we note that this idea of jumbledness is compatible with some common operations on graphs. If G is (p, α)-jumbled, all induced subgraphs $\bar{U}$ of the complement $\bar{G}$ satisfy

$$\left| e(\bar{U}) - (1-p)\binom{|\bar{U}|}{2} \right| = \left| \binom{|U|}{2} - e(\bar{U}) - p\binom{|U|}{2} \right| = \left| e(U) - p\binom{|U|}{2} \right| \leq \alpha |U| = \alpha |\bar{U}|,$$

meaning that $\bar{G}$ is $(1-p, \alpha)$-jumbled. Further, if G is (p, α)-jumbled, then it is clear that all subgraphs of G are also (p, α)-jumbled.

The criterion for jumbledness might initially appear to require quite strong global knowledge about each induced subgraph. However, in the following two theorems, we provide first a weaker global condition and then a local condition for a graph to be jumbled. Both of these are from [6].

Theorem 1 (Thomason)**.** Suppose we have $0 < p < 1$ and $h, m, n \in \mathbb{N}$ with $2 \leq h \leq n-2$. Then, given a graph G on n vertices such that all induced subgraphs H of size h satisfy

$$\left| e(H) - p\binom{h}{2} \right| \leq m$$

we have

$$\left| e(H) - p\binom{|H|}{2} \right| \leq 80m \frac{1}{\eta^2(1-\eta)^2},$$

for any induced subgraph H of G, where $\eta = h/n$.

Proof. In the following proof, we will write $y = B(x)$ to mean that $|y| \leq x$. Note in particular that a graph is (p, α)-jumbled if for every induced subgraph $U \subset V$, we have

$$e(U) - p\binom{|U|}{2} = B\left(\alpha |U|\right).$$

We consider three cases. First, suppose $|H| = k \geq h$. Then, note that there are $\binom{k}{h}$ induced subgraphs of order h in H. Further, each edge in H appears in $\binom{k-2}{h-2}$ possible subgraphs, so we have

$$\begin{aligned} e(H) = \frac{1}{\binom{k-2}{h-2}} \sum_{L \subseteq H} e(L) = \frac{1}{\binom{k-2}{h-2}} \sum_{L \subseteq H} p\binom{h}{2} + B(m) = \frac{\binom{k}{h}}{\binom{k-2}{h-2}} \left[p\binom{h}{2} + B(m) \right] \\ = p\binom{k}{2} + B\left(\frac{k(k-1)}{h(h-1)} m \right) = p\binom{k}{2} + B\left(\frac{2k^2}{h^2} m \right) \end{aligned}$$

and therefore

$$\left|e(H) - p\binom{k}{2}\right| \leq \frac{2k^2}{h^2}m \leq \frac{2n^2}{h^2}m = \frac{2m}{\eta^2} \leq \frac{80m}{\eta^2(1-\eta)^2},$$

as desired.

We now consider $|H| = k \leq \min\{h, n-h\}$. In this case, consider subgraphs F of $G - H$ of order h and subgraphs L of H of some fixed order l with $1 \leq l \leq k$ to be determined later. Then, we see that

$$e(H \cup F) = p\binom{k+h}{2} + B\left(\frac{2(k+h)^2}{h^2}m\right)$$

and

$$e(L \cup F) = p\binom{l+h}{2} + B\left(\frac{2(l+h)^2}{h^2}m\right)$$

by the case with $k \geq h$. We now count the edges between H and F by considering all L as follows:

$$\sum_L e(L \cup F) = \sum_L e(L) + e(L, F) + e(F) = \binom{k-2}{l-2}e(H) + \binom{k-1}{l-1}e(H, F) + \binom{k}{l}e(F),$$

since each edge in H, (H, F), and F is counted the corresponding number of times. But we also have

$$\sum_L e(L \cup F) = \binom{k}{l}\left[p\binom{l+h}{2} + B\left(\frac{2(l+h)^2}{h^2}m\right)\right],$$

so we obtain

$$\frac{l(l-1)}{k(k-1)}e(H) + \frac{l}{k}e(H, F) + e(F) = p\binom{l+h}{2} + B\left(\frac{2(l+h)^2}{h^2}m\right).$$

From before, we have

$$e(H \cup F) = e(H) + e(H, F) + e(F) = p\binom{k+h}{2} + B\left(\frac{2(k+h)^2}{h^2}m\right),$$

so we can solve using the fact that $e(F) = p\binom{h}{2} + B(m)$ to obtain

$$\frac{l-k}{k-1}e(H) = \frac{k}{l}p\binom{l+h}{2} - p\binom{k+h}{2} - \frac{k-l}{l}\binom{h}{2} + B\left(\frac{2k(l+h)^2}{lh^2}m + \frac{2(k+h)^2}{h^2}m + \frac{k+l}{l}m\right)$$

meaning that

$$\begin{aligned} e(H) &= p\binom{k}{2} + B\left[\frac{k-1}{k-l}\left(\frac{2k(l+h)^2}{lh^2}m + \frac{2(k+h)^2}{h^2}m + \frac{k+l}{l}m\right)\right] \\ &= p\binom{k}{2} + B\left[\frac{m}{1-\lambda}\left(\frac{2(1+\lambda)^2}{\lambda} + 4 + \frac{1}{\lambda} + 1\right)\right] \end{aligned}$$

where $\lambda = \frac{l}{k}$ and we use the fact that $k \leq h$. Since $k \geq 2$, we can then pick $\frac{k}{3} \leq l \leq \frac{k}{2}$, meaning that $\frac{1}{3} \leq l \leq \frac{1}{2}$, so we find that

$$e(H) = p\binom{k}{2} + B(64m) = p\binom{k}{2} + B\left(\frac{80m}{\eta^2(1-\eta)^2}\right).$$

Finally if we have a subgraph H of order k such that $n-h \geq k \leq h$, then we count the number of edges of H in two different ways by summing over all subgraphs L of H of order $n-h$. Then, we obtain

$$e(H) = \frac{\binom{k}{n-h}}{\binom{k-2}{n-h-2}}\left[p\binom{n-h}{2} + B(m)\right] = p\binom{k}{2} + B\left(\frac{k(k-1)}{(n-h)(n-h-1)}m\right),$$

which implies the desired by an argument similar to the first case. □

Note that while the constant 80 was chosen here largely for convenience, the bound given is in fact asymptotically optimal.

In additional to this weaker global condition for jumbledness, there is a more local condition that relies only on the common neighbors of each pair of vertices, allowing us to approach this property from different directions.

Theorem 2 (Thomason)**.** Let G be a graph on n vertices with minimum degree pn. If no pair of vertices of G has more than $p^2n + l$ common neighbors for some $l \geq 1$, then G is $\left(p, \frac{1}{2}\left[\sqrt{n(p+l)} + p\right]\right)$-jumbled.

Proof. Take a induced subgraph $H = \{x_1, \ldots, x_k\}$ of G with k vertices and let d_i be the degree of x_i (with edges considered within H only). Take d to be the average of the degrees d_i. Now, let $e_1, \ldots, e_{n-k}$ to be the number of total edges between H and each of the $n-k$ vertices in $G-H$. By our condition, each vertex x_i in H has degree at least pn (within G), so x_i has edges to at at least $pn - d_i$ vertices in $G-H$. Therefore, counting the number of edges between H and $G-H$ in two different ways, we have that

$$\sum_{i=1}^{n-k} e_i \geq \sum_{i=1}^{k}(pn - d_i) = kpn - kd.$$

Now, we count the common neighbors of the $\binom{k}{2}$ different pairs of vertices in H. Note that each common neighbor of a pair of vertices in H is the common vertex of two edges within H or two edges leaving H; counting these, we find that there are exactly

$$\sum_{i=1}^{k}\binom{d_i}{2} + \sum_{i=1}^{n-k}\binom{e_i}{2}$$

such common neighbors. But our condition implies that each pair of vertices has at most $p^2n + l$ common neighbors and there are $\binom{k}{2}$ such pairs. Thus, we have

$$\binom{k}{2}(p^2n + l) \geq \sum_{i=1}^{k}\binom{d_i}{2} + \sum_{i=1}^{n-k}\binom{e_i}{2} \geq k\binom{d}{2} + (n-k)\binom{\frac{k}{n-k}(pn-d)}{2},$$

by the convexity of $\binom{x}{2}$, which we rearrange to get

$$\begin{aligned}\frac{n-k}{n}[(k-1)l + np(1-p)] &\geq \frac{n-k}{n}[(k-1)(p^2n+l) - d(d-1)] \\ &\geq \frac{(pn-d)(kpn-kd-n+k)}{n} \geq (pk-d)^2.\end{aligned}$$

This means that

$$n(p+l) \geq (k-1)l + np(1-p) \geq \frac{n}{n-k}[(k-1)l + np(1-p)] \geq (pk-d)^2$$

so we multiply through by k^2 to obtain

$$k^2 n(p+l) \geq (pk-d)^2 \iff \sqrt{n(p+l)}k \geq |pk-d| \iff \left| e(H) - p\binom{|H|}{2} \right| \leq \frac{k}{2}\left[\sqrt{n(p+l)} + p\right],$$

so G is indeed $\left(p, \frac{1}{2}(\sqrt{n(p+l)} + p)\right)$-jumbled, as desired. □

These two conditions are simpler than the definition of jumbleness and hence provide methods to test if a given graph is jumbled. In particular, we are now able to compare jumbled graphs to true random graphs $G(n,p)$ and evaluate how similar their properties are. The definition of jumbledness was crafted to mimic the edge distribution of $G(n,p)$, and this property is indeed satisfied, as shown by the following theorem.

Proposition 3. Take $p < 1-\varepsilon$. Then, we have almost surely in $G(n,p)$ that for any subset $U \subset V$ of the vertex set V we have

$$\left| e(U) - p\binom{|U|}{2} \right| = O\left(|U|\sqrt{np}\right).$$

Proof. (sketch) Note that $e(U)$ is the sum of $\binom{u}{2}$ Bernoulli random variables with probability p. Standard tail probability estimates bound the deviation of $e(U)$ from its mean $p\binom{|U|}{2}$ by $O\left(\sqrt{p\binom{|U|}{2}}\right) = O(|U|\sqrt{np})$. □

Proposition 3 shows that an element of $G(n,p)$ is almost surely $(p, O(\sqrt{np}))$-jumbled, so jumbled graphs do share an edge distribution with our random graph model. This correspondence of edge distributions leads to other similarities; for instance, jumbled graphs have a similar distribution of induced subgraphs to $G(n,p)$ as a consequence of the following theorem.

Theorem 4 (Thomason)**.** Let G be a (p,α)-jumbled graph with $p \leq \frac{1}{2}$, and let H be a graph of order $h \geq 3$ and m edges. Then, if $\varepsilon^2 p^h n \geq 42\alpha h^2$ for some $0 < \varepsilon < 1$, the number of times $N_G^*(H)$ that H occurs as an induced subgraph of G satisfies

$$(1-\varepsilon)^h p^m (1-p)^{\binom{h}{2}-m} n^h \leq N_G^*(H) \leq (1+\varepsilon)^h p^m (1-p)^{\binom{h}{2}-m} n^h.$$

Proof. (sketch) The idea of the proof, which we do not present in its entirety here, is to construct the desired induced subgraphs by sequentially choosing vertices with a tightly bounded number of neighbors in the subgraph, which is guaranteed by the following lemma:

Lemma 1. Let G be a (p,α)-jumbled graph with $|G| = n$ and take $0 < \varepsilon < 1$. If H is an induced subgraph of order k, then at least $n - \varepsilon k$ vertices in G have between $pk - 21\frac{\alpha}{\varepsilon}$ and $pk + 21\frac{\alpha}{\varepsilon}$ neighbors in H.

We can bound the number of ways to construct such a sequence using Lemma 1 repeatedly to look at neighbors of the portion of our graph already constructed. □

In the model $G(n,p)$, there are $\binom{n}{h}$ possible induced subgraphs with h vertices, and each of these has the induced edges of H with probability about $p^m(1-p)^{\binom{h}{2}-m}$, for a total of roughly $\binom{n}{h}p^m(1-p)^{\binom{h}{2}-m} \approx p^m(1-p)^{\binom{h}{2}-m}n^h$ induced copies of H. Therefore, Theorem 4 does provide the desired correspondence with the behavior of $G(n,p)$ that Thomason's jumbledness property seeks to capture.

3 Quasi-random Graphs

While the idea of a jumbled graph works well to determine if a given graph approximates the behavior of a random graph, it alone does not account as well for the number of vertices in the graph. Note that if a graph G is $(p, \Theta(np))$-jumbled, we have only that

$$\left| e(U) - p\binom{|U|}{2} \right| \leq \Theta(np) \cdot |U|,$$

which gives quite a weak bound on the variation in edge distribution, since $\Theta(np) \cdot |U|$ has the same order as $p\binom{|U|}{2}$, the expected number of edges for $G(n,p)$. The notion of a (p,α)-jumbled graph is therefore non-trivial only when $\alpha = o(np)$ and thus $\alpha|U| = o(n^2)$; this suggests that we should attempt to examine this particular case further, culminating in the idea of *quasi-randomness* introduced by Chung, Graham, and Wilson [2]. We use $p = \frac{1}{2}$ in our discussion for convenience, but this idea can be easily extended to all p.

To find a more exact analogue of the graph model $G(n, 1/2)$, we instead consider sequences (G_n) of graphs. In this case, we say that the sequence G_n has a property P if, for sufficiently large n, the graph G_n satisfies P. As suggested by the fact that jumbledness is non-trivial only when $\alpha|U| = o(n^2)$, we call a graph *quasi-random* if, for all induced subgraphs U of G, we have that

$$e(U) = \frac{1}{4}|U|^2 + o(n^2). \tag{2}$$

Intuitively, this condition means that, as n becomes large, the graph G_n becomes $(1/2, o(n/2))$-jumbled and therefore has an edge distribution close to that of $G(n, 1/2)$. Remarkably, Chung, Graham, and Wilson were able to show that this defining condition for a quasi-random graph is equivalent to six other properties in the following theorem.

Theorem 5 (Chung, Graham, and Wilson)**.** For a sequence of graphs (G_n), the following properties are equivalent for $s, t \geq 4$:

1. For every graph H with s vertices, the number of copies $N^*_{G_n}$ of H in G_n as an induced subgraph satisfies
$$N^*_{G_n} = (1 + o(1))n^s 2^{-\binom{s}{2}}$$

2. Let t be even. The number of copies $N_{G_n}(C_t)$ of the cycle C_t with t vertices as a subgraph (not necessarily induced) of G_n satisfies

$$N_{G_n}(C_t) \leq (1+o(1))\left(\frac{n}{2}\right)^t$$

and the number of edges satisfies $e(G_n) \geq (1+o(1))\frac{n^2}{4}$.

3. The two eigenvalues $|\lambda_1| \geq |\lambda_2|$ of maximal modulus of the adjacency matrix of G_n satisfy

$$\lambda_1 = (1+o(1))\frac{n}{2} \text{ and } \lambda_2 = o(n)$$

and the number of edges satisfies $e(G_n) \geq (1+o(1))\frac{n^2}{4}$.

4. For all induced subgraphs U of G_n, we have

$$e(U) = \frac{1}{4}|U|^2 + o(n^2).$$

5. For all induced subgraphs U of G_n with $|U| = \lfloor n/2 \rfloor$, we have

$$e(U) = \frac{1}{4}|U|^2 + o(n^2).$$

6. The number of vertices $a_n(u,v)$ connected to either both or none of vertices $u, v \in G_n$ satisfies

$$\sum_{u,v \in V} \left| a_n(u,v) - \frac{n}{2} \right| = o(n^3).$$

7. The number of common neighbors $b_n(u,v)$ of vertices $u, v \in G_n$ satisfies

$$\sum_{u,v \in V} \left| b_n(u,v) - \frac{n}{4} \right| = o(n^3).$$

The proof of this theorem is quite long, so we omit it in favor of a number of comments. Note that Property 4 is the definition of a quasi-random graph, so this theorem provides many equivalent ways of checking that a sequence of graphs (G_n) is quasi-random. This is somewhat surprising, since each of these conditions originates from the seemingly weak bound on edge distribution given by Property 4.

We also note that Property 1 is the analogue of Theorem 4 for quasi-random graphs, and the equivalence of Properties 4 and 5 is the analogue for Theorem 1, providing some continuity between our previous concept of jumbledness and the generalization we now provide. It can also be noted that each of our properties holds for the sequence of graphs $G(n, 1/2)$; in some sense, they provide a fairly complete knowledge of the behavior of $G(n, 1/2)$, allowing us to again treat a quasi-random graph as a realization of a random graph model.

A generalization of Property 2 in Theorem 5 gives the following stronger condition on induced subgraphs, which we will state without proof.

Theorem 6. For a fixed $0 < p < 1$, if $N_{G_n}(H) = (1 + o(1))n^h p^{e(H)}$ for all graphs H of order h, then G_n is quasi-random.

The result was further improved by Simonovits and Sós using Szemerédi's lemma in [5] to allow consideration of only a single sample graph.

Theorem 7. For some fixed graph H, if the number of copies of H as a subgraph (not necessarily induced) satisfies $N_F(H) = |F|^h p^{e(H)} + o(n^{|H|})$ for every induced subgraph F of G_n, then G_n is quasi-random.

This improvement does require that every induced subgraph must now be checked for this property, but it provides yet another link between quasi-random graphs and $G(n, p)$. It should be noted, however, that this correspondence is not perfect. In fact, Simonovits and Sós recently proved in [4] that if a similar condition holds for $N_F^*(H)$, then it is also possible for the sequence of graphs G_n to be the union of two distinct quasi-random graphs.

4 A Classical Example

Both jumbled and quasi-random graphs originated as generalizations of more specific types of pseudo-random graphs. We give here a connection between our new ideas and the traditional construction of a pseudo-random Paley graph.

Example 8. Let $q \equiv 1 \pmod 4$ be a prime power and let χ be the quadratic residue character on $\mathbb{F}_q$. Define the Paley graph P_q to be the graph with vertex set $\mathbb{F}_q$, the finite field of order q such that $x, y \in \mathbb{F}_q$ are connected with an edge if and only if $\chi(x - y) = 1$.

The Paley graph is one of the most canonical pseudo-random graphs. Below, we see that, when interpreted as a graph sequence, it becomes a quasi-random graph, as expected.

Proposition 9 (Chung, Graham, and Wilson)**.** The Paley graph P_q is quasi-random.

Proof. For $x, y \in P_q$, note that z is connected to both or neither exactly when $\chi(z - x) = \chi(z - y)$, so $\chi\left(\frac{z-x}{z-y}\right) = 1$ and $\frac{z-x}{z-y}$ must be a quadratic residue modulo q. But this associates each quadratic residue modulo $q \neq 1$ to a unique z, so there are exactly $\frac{q-3}{2}$ values for z and thus

$$\sum_{x,y \in P_q} \left| a_n(x, y) - \frac{q-1}{2} \right| = \frac{1}{2} n^2 = o(n^3),$$

satisfying Property 6 in Theorem 5 and making the Paley graph a quasi-random graph. □

This result also makes P_q a pseudo-random graph, since Property 4 in Theorem 5 shows that

$$\left| e(U) - \frac{1}{4}|U|^2 \right| = o(n^2),$$

so P_q is $(1/2, k)$ jumbled for some constant k. These two correspondences both stem from the relatively uniform distribution of quadratic residues, which results in uniform edge distribution in P_q.

5 Conclusion

In this paper, we have sketched some of the basic results on pseudo-random and quasi-random graphs and given the basic framework for their use. We have attempted to give some motivation for the basic definitions and approaches in the context of comparing graphs with random graph models. In some sense, this comparison gives us a greater understanding of the nature of "typical" graphs and provides some intuition for exactly which conditions render a random graph model representative of the properties of a typical graph.

Many more advanced results in this area have been found, but they require technical approaches that are quite involved and beyond the scope of this paper. This may indicate the inherent complexity of attempting to model random properties with deterministic ones. Some first steps in this direction are given in [2], [6], and [7], and the interested reader can find an example of more sophisticated approaches in [4] and [5].

References

[1] Bollobás, Béla: *Random Graphs*. 2nd Ed. Cambridge University Press, 2001: Chp. 2, 13.

[2] Chung, F. R. K.; Graham, R. L.; Wilson, R. M: Quasi-random graphs. *Combinatorica* **9** (1989), no. 4, 345–362.

[3] M. Krivelevich and B. Sudakov: Pseudo-random graphs. *More sets, graphs and numbers*, E. Gyori, G. O. H. Katona and L. Lovasz, Eds., Bolyai Society Mathematical Studies Vol. 15, 199–262.

[4] Simonovits, Miklós and Sós, Vera T.: Hereditarily extended properties, quasi-random graphs and induced subgraphs. Preprint. Available at `http://www.renyi.hu/~miki/quasiind.pdf`.

[5] Simonovits, Miklós and Sós, Vera T.: Hereditarily extended properties, quasi-random graphs and not necessarily induced subgraphs. *Combinatorica* **17** (1997), no. 4, 577–596.

[6] Thomason, Andrew: Pseudo-random graphs. *Random Graphs '85* (Poznán, 1985), 307–331, North-Holland Math. Stud., 144, North-Holland, Amsterdam, 1987.

[7] Thomason, Andrew, Random graphs, strongly regular graphs and pseudorandom graphs. *Surveys in Combinatorics 1987* (New Cross, 1987), 173–195, London Math. Soc. Lecture Note Ser., 123, Cambridge Univ. Press, Cambridge, 1987.