A CAUCHY-DAVENPORT TYPE RESULT FOR ARBITRARY REGULAR GRAPHS

Peter Hegarty

Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, SE-41296 Gothenburg, Sweden
hegarty@chalmers.se

Abstract

Motivated by the Cauchy-Davenport theorem for sumsets, and its interpretation in terms of Cayley graphs, we prove the following main result: There is a universal constant $\epsilon > 0$ such that, if $\mathcal{G}$ is a connected, regular graph on n vertices, then either every pair of vertices can be connected by a path of length at most three, or the number of pairs of such vertices is at least $1 + \epsilon$ times the number of edges in $\mathcal{G}$. We discuss a range of further questions to which this result gives rise.

1. Introduction and Statement of Results

Let A be a subset of an abelian group G, written additively, and h a positive integer. The h-fold sumset hA is defined as

$$hA = \{g \in G : g = a_1 + \cdots + a_h \text{ for some } a_1, \ldots, a_h \in A\}.$$

We say that A is a *basis* for G if $hA = G$ for some h. The cardinality of a set S will be denoted $|S|$. The following is a (special case of a) fundamental result in the theory of sumsets:

Theorem 1. (Cauchy-Davenport) *Let p be a prime and A a subset of $\mathbb{Z}_p$. Then $|hA| \geq \min\{p, h|A| - (h-1)\}$.*

There is a well-known generalisation of this result to arbitrary abelian groups, due to Kneser [3], but that is not what is of primary interest to us here. Instead, we are interested in interpreting the Cauchy-Davenport result in terms of graphs. First, recall the following definition:

Definition 2. Let G be an abelian group and S a subset of G. The *Cayley graph*[1] $\mathcal{G} = \mathcal{G}(G, S)$ is the directed graph whose vertices are the elements of G and whose edges consist of the ordered pairs (g_1, g_2) such that $g_2 - g_1 \in S$.

---

[1] Throughout this paper, the letter G will be reserved to denote an abelian group and graphs will be denoted by the scripted letter $\mathcal{G}$.

Note that the graph $\mathcal{G}(G, S)$ is strongly connected if and only if S is a basis for G. For simplicity, let us assume this, plus that the set S is symmetric, i.e., $S = -S$, and contains the identity element of G. Then we can think of the Cayley graph as being undirected, with a loop at every vertex. In this case, let $\mathcal{G}_0(G, S)$ be the part of $\mathcal{G}(G, S)$ with all the loops removed. For the rest of the paper, we shall only deal with undirected, loopless graphs. Now consider the following definition:

Definition 3. Let $\mathcal{G}$ be a graph on the vertex set V and h a positive integer. We denote by $h\mathcal{G}$ the graph on vertex set V such that $\{v_1, v_2\}$ is an edge in $h\mathcal{G}$ if and only there is a path joining v_1 to v_2 in $\mathcal{G}$ of length at most h. We shall call $h\mathcal{G}$ the *h-fold sumgraph*[2] of $\mathcal{G}$ and denote its set of edges by $h\mathcal{E} = \mathcal{E}(h\mathcal{G})$.

Then the following is an immediate consequence of the Cauchy-Davenport theorem:

Theorem 4. *Let p be a prime and let A be a subset of $\mathbb{Z}_p$ such that $0 \in A$ and $A = -A$. Let $\mathcal{G} = \mathcal{G}_0(\mathbb{Z}_p, A)$ be the Cayley graph of A, minus all loops. Then for every positive integer h,*

$$|h\mathcal{E}| \geq \min\left\{ \binom{p}{2}, h|\mathcal{E}| \right\}.$$

The question which motivated this paper is whether anything like this result is true for more general connected graphs. More precisely, the feature of Theorem 4 that we are interested in generalising is the fact that the (edge) sizes of the graphs $h\mathcal{G}$ grow at least linearly in h, as long as $\mathcal{G}$ isn't already too dense. As we shall show below, it is hopelessly optimistic to hope for anything like this phenomenon in arbitrary connected graphs. However, Cayley graphs have the very important property that they are regular. Our main result is the following partial generalization of Theorem 4:

Theorem 5. *There is a universal constant $\epsilon > 0$ such that if $\mathcal{G}$ is a regular, connected graph on n vertices, then*

$$|3\mathcal{E}| \geq \min\left\{ \binom{n}{2}, (1 + \epsilon)|\mathcal{E}| \right\}.$$

In fact, we can take ϵ to be the unique positive root of the equation

$$\epsilon = \frac{1}{4}(1 - \sqrt{\epsilon})^3, \quad (\epsilon \approx 0.087\dots). \tag{1}$$

---

[2]This definition is well-known in the literature, though it is standard to use multiplicative notation instead, which is natural when one thinks in terms of the adjacency matrix of the graph. So what we are calling the h-fold sumgraph $h\mathcal{G}$ is usually referred to as the *h-th power of $\mathcal{G}$* and denoted $\mathcal{G}^h$. Observe that if we add a loop at each vertex of $\mathcal{G}$ and let A be the adjacency matrix of the resulting graph, then $|\mathcal{E}(\mathcal{G}^h)|$ is just half the number of non-zero off-diagonal entries in A^h.

For the remainder of this note we shall retain our additive notation and terminology for graphs so as to emphasise the connection to sumsets. The reader should also be aware, however, that the term 'sum graph' (two words) has also been used in a different context by F. Harary [2].

We were surprised by the simplicity and elegance of this result, which is why we considered it worth mentioning. Of course, it is unsatisfactory in many respects so some detailed remarks are in order:

1. The obvious problem with our result is that it cannot be used recursively to obtain estimates for the growth of h-fold sumgraphs for arbitrary h. This is because, even if the graph $\mathcal{G}$ is regular, then the graphs $h\mathcal{G}$ need not be, for any $h > 1$ (note that regularity is preserved for Cayley sum-graphs). Thus it remains to obtain a generalization of Theorem 5 to h-fold sumgraphs for arbitrary h. Note that, for fixed degree, the sumgraphs $h\mathcal{G}$ grow at least linearly 'on average' until the graph becomes complete. This is a trivial observation, but a more precise result is contained in the next proposition. Recall that the *diameter* of a graph is the smallest $\delta > 0$ such that any pair of vertices are connected by a path of length at most δ. In the notation of Definition 3, the diameter of a graph $\mathcal{G}$ on n vertices is the smallest h such that $h\mathcal{G} = K_n$, the complete graph. Now we have

Proposition 6. *Let $\mathcal{G}$ be a connected graph on n vertices and of minimal degree d. Then*

$$diam(\mathcal{G}) \leq \frac{3n - (d + 3)}{d + 1}. \tag{2}$$

2. However, the growth of sumgraphs can certainly be irregular. In particular, and this is the most natural thing to ask about, there is no constant $\epsilon' > 0$ such that the analogue of Theorem 5 holds for 2-fold sumgraphs. To see this consider the following example.

Example 7. Fix $d > 0$ and let n be a multiple of $d + 1$, say $n = m(d + 1)$. Let $\mathcal{G} = \mathcal{G}_{d,m}$ be the following graph on n vertices: Partition the vertex set V into m disjoint subsets of size $d + 1$, say $V_1, V_2, \ldots, V_m$. For each $i = 1, \ldots, m$ pick two vertices $v_{i1}, v_{i2} \in V_i$. Now the graph $\mathcal{G}_{d,m}$ contains the following edges:

(i) for each $i = 1, \ldots, m$, insert all edges among the vertices of V_i, except the edge $\{v_{i1}, v_{i2}\}$.

(ii) for each $i = 1, \ldots, m - 1$, insert the edge $\{v_{i1}, v_{(i+1),2}\}$, and then finally add the edge $\{v_{m1}, v_{12}\}$.

Clearly, this graph is connected and d-regular, so

$$|\mathcal{E}_{d,m}| = \left(\frac{d}{2}\right) n.$$

However, one easily checks that

$$|2\mathcal{E}_{d,m} \backslash \mathcal{E}_{d,m}| = \left(\frac{2d - 1}{d + 1}\right) n, \tag{3}$$

so that $|2\mathcal{E}_{d,m}| = (1 + o_d(1))|\mathcal{E}_{d,m}|$. Note also that for this graph one may check that

$$|3\mathcal{E}_{d,m}\backslash\mathcal{E}_{d,m}| = \left(\frac{d^2 + 4}{d+1}\right)n, \qquad (4)$$

so that $|3\mathcal{E}_{d,m}| = (3 - o_d(1))|\mathcal{E}_{d,m}|$.

Considering this example naturally leads one to ask for more precise extremal results. We believe that the graphs $\mathcal{G}_{d,m}$ are essentially extremal for 2-fold sumgraphs, but these latter objects are still somewhat mysterious to us. Motivated by (3), we are prepared at this stage to conjecture the following:

Conjecture 8. Let d, n be positive integers. If $\mathcal{G}$ is a d-regular, connected graph on n vertices, then either $2\mathcal{G} = K_n$ or $|2\mathcal{E}\backslash\mathcal{E}| \geq (2 - o_d(1))\,n$.

Note that, in the notation of this conjecture, if $n \geq d+2$ then trivially $|2\mathcal{E}\backslash\mathcal{E}| \geq n/2$, since every vertex is connected to at least one non-neighbor by a path of length two. Hence there is a factor of four separating (asymptotically) the trivial lower bound for $|2\mathcal{E}\backslash\mathcal{E}|$ and what we conjecture to be the truth.

Neither is it clear to us whether the graphs $\mathcal{G}_{d,m}$ are essentially extremal for 3-fold sumgraphs. The question here is what is the best-possible choice of the constant ϵ in Theorem 5? By (4), we cannot take $\epsilon > 2$. Indeed, the same conclusion could be drawn by considering the Cayley graph of an arithmetic progression.

Also, note that the graphs $\mathcal{G}_{d,m}$ are certainly not close to being extremal sumgraphs in general. This is because it is easy to see that $\mathcal{G}_{d,m}$ has diameter $m + 1 = \frac{n}{d+1} + 1$, whereas from the proof of Proposition 6 we will easily be able to construct examples which show that the upper bound in (2) is essentially best-possible, even for regular graphs (see Remark 9). Hence, we suspect that the extremal problem for sumgraphs in general might be quite hard.

3. Finally, note that there doesn't seem to be any hope of obtaining meaningful generalizations of our results to graphs which are not regular. For example, let n be a positive integer and let $\mathcal{G}_n$ be the graph on n vertices which is the union of a complete subgraph on $\lfloor n^{3/4} \rfloor$ vertices and a path of length $n - \lfloor n^{3/4} \rfloor$ which is joined to the complete subgraph at one vertex. This graph is connected and contains $\Theta(n^{3/2})$ edges but, for any fixed h, the h-fold sumgraph contains only $\Theta_h(n)$ additional edges.

The rest of the paper is organised as follows. Sections 2 and 3 are devoted to the proofs and discussion of Theorem 5 and Proposition 6 respectively. Section 4 contains a quick recap of unresolved issues.

2. Proof of Theorem 5

Notation. If $\mathcal{G}$ is a graph and $X \subseteq V(\mathcal{G})$, then $N(X)$ will denote the set of all neighbors of the vertices in X. If X is a singleton set, say $X = \{x\}$, then we simply write $N(x)$.

Let d, n be positive integers and let $\mathcal{G}$ be a connected, d-regular graph on n vertices. Let $\epsilon > 0$ be the solution of (1) and suppose that $|3\mathcal{E}| < (1 + \epsilon)|\mathcal{E}|$. We must show that $3\mathcal{G} = K_n$. Since $2\mathcal{E} \subseteq 3\mathcal{E}$, we can first of all deduce that $|2\mathcal{E}| < (1+\epsilon)|\mathcal{E}|$. We present the argument in a sequence of steps.

Step 1: Set $\epsilon_1 := \sqrt{\epsilon}$. For each $v \in V(\mathcal{G})$, let

$$T_v := \{w \in V(\mathcal{G}) : \{v, w\} \in 2\mathcal{E}\backslash\mathcal{E}\}$$

and let

$$V_1 := \{v \in V(\mathcal{G}) : |T_v| < \epsilon_1 d\}.$$

Since, by assumption,

$$\epsilon dn > 2 \times |2\mathcal{E}\backslash\mathcal{E}| = \sum_v |T_v|,$$

it follows easily that $|V_1| > (1 - \epsilon_1)n$.

Step 2: Let $v \in V_1$. Set $A_v := N(v)$, $B_v := N(N(v))$ and $C_v := B_v\backslash(\{v\} \cup A_v)$. If the set C_v were empty then, since the graph is connected, it would imply that $V(\mathcal{G}) = \{v\} \cup A_v$ and hence that $2\mathcal{G} = K_n$. So we may assume that C_v is non-empty. If $c \in C_v$ then there is a path $v \to a \to c$ in $\mathcal{G}$, for some $a \in A_v$, hence $\{v, c\} \in 2\mathcal{E}$. By definition of the set V_1, it follows that

$$|C_v| < \epsilon_1 d. \tag{5}$$

Set $D_v := V(\mathcal{G})\backslash(\{v\} \cup A_v \cup C_v)$. Suppose this set were empty. Since $\mathcal{G}$ is connected, it would imply that $V(\mathcal{G}) = \{v\} \cup A_v \cup C_v$. We claim that, in this case, $3\mathcal{G} = K_n$. We need to show that any two vertices in $\{v\} \cup A_v \cup C_v$ can be connected by a path of length at most 3. This is obvious unless both vertices lie in C_v. Consider a pair of such vertices, say c_1 and c_2. Our assumptions say that $N(c_1) \cup N(c_2) \subseteq A_v \cup C_v$. But by (5), d-regularity and the fact that $\epsilon_1 < 1/2$ (see the statement of Theorem 5), it follows that c_1 and c_2 must have a common neighbor in A_v. Hence $\{c_1, c_2\} \in 2\mathcal{E}$, in fact.

So we may now assume that the set D_v is non-empty. There must be at least one edge between C_v and D_v. For any such edge, say $\{c_v, d_v\}$, we know by (5) that at least $(1 - \epsilon_1)d$ of the neighbors of c_v lie in $A_v \cup D_v$. Let $\mathcal{C}_v$ be the set of vertices in C_v with at least one neighbor in D_v and set

$$\alpha_v := \frac{1 - \epsilon_1}{d} \times \max\{|N(c_v) \cap A_v| : c_v \in \mathcal{C}_v\}.$$

In the steps to come, we consider the following two cases, at least one of which must obviously apply:

> `Case 1`: For at least half of all $v \in V_1$, one has $\alpha_v \leq \frac{1}{2}$.
> `Case 2`: For at least half of all $v \in V_1$, one has $\alpha_v > \frac{1}{2}$.

`Step 3`: Suppose Case 1 holds. Let $V_2 := \{v \in V_1 : \alpha_v \leq \frac{1}{2}\}$. For each $v \in V_2$, pick any vertex $c_v \in \mathcal{C}_v$. Then there are at least $\frac{1}{2}(1 - \epsilon_1)d$ choices for an edge $\{c_v, d_v\}$ such that $d_v \in D_v$. Notice that, for any choice of d_v, there is a path $v \to a_v \to c_v \to d_v$ in $\mathcal{G}$, for some $a_v \in A_v$. Hence $\{v, d_v\} \in 3\mathcal{E}\backslash\mathcal{E}$. Summing over all $v \in V_2$ and noting that any given pair of vertices is counted at most twice, it follows that

$$|3\mathcal{E}\backslash\mathcal{E}| \geq \frac{1}{2} \cdot |V_2| \cdot \frac{1}{2}(1 - \epsilon_1)d \geq \frac{1}{2}(1 - \epsilon_1)^2|\mathcal{E}| > \epsilon|\mathcal{E}|,$$

contradicting our assumptions.

`Step 4`: Suppose Case 2 holds. Let $V_3 := \{v \in V_1 : \alpha_v > \frac{1}{2}\}$, so that $|V_3| \geq \frac{1}{2}|V|$. Let $v \in V_3$ and fix a choice of a vertex $c_v \in \mathcal{C}_v$ such that c_v has at least $\frac{1}{2}(1 - \epsilon_1)d$ neighbors inside A_v. Let d_v be any neighbor of c_v inside D_v. Observe that all the neighbors of d_v lie inside $C_v \cup D_v$. Hence, by (5), there are at least $(1 - \epsilon_1)d$ choices for a vertex $e_v \in N(d_v) \cap D_v$. For any such vertex e_v and any vertex $a_v \in N(c_v) \cap A_v$, there is a path in the graph $a_v \to c_v \to d_v \to e_v$. Hence $\{a_v, e_v\} \in 3\mathcal{E}\backslash\mathcal{E}$. Therefore, if we set

$$\mathbb{S} := \sum_{v \in V_3} \#\{\{a_v, e_v\} \in 3\mathcal{E}\backslash\mathcal{E} : a_v \in A_v, \ e_v \in D_v\},$$

then we have

$$|\mathbb{S}| \geq |V_3| \times \frac{(1 - \epsilon_1)d}{2} \times (1 - \epsilon_1)d \geq \left(\frac{(1 - \epsilon_1)^3}{2}\right)\frac{d^2 n}{2}.$$

On the other hand, since since $v \in N(a_v)$ always, any pair of vertices can appear in the sum at most $2d$ times. It follows that

$$|3\mathcal{E}\backslash\mathcal{E}| \geq \frac{1}{4}(1 - \epsilon_1)^3|\mathcal{E}| \geq \epsilon|\mathcal{E}|,$$

which again contradicts our assumptions, and completes the proof of the theorem.

3. Proof of Proposition 6

Let $\mathcal{G}$ be a connected graph on n vertices of minimal degree d. Let δ be the diameter of $\mathcal{G}$ and let v, w be a pair of vertices such that a shortest path between them has length exactly δ. Let such a path be

$$v_1 = v \to v_2 \to \cdots \to v_\delta \to v_{\delta+1} = w. \tag{6}$$

Let A be the set of vertices along the path and $B := V(\mathcal{G})\backslash A$. Using the fact that there is no shorter path in $\mathcal{G}$ between v and w, we shall count in two ways the number $e(A, B)$ of edges in $\mathcal{G}$ between A and B. On the one hand, this fact implies that there are no edges between the vertices along the path other than those in the path itself. Since $\mathcal{G}$ has minimal degree d, it follows that

$$e(A, B) \geq (\delta - 1) \cdot (d - 2) + 2 \cdot (d - 1) = (\delta + 1)(d - 2) + 2. \qquad (7)$$

On the other hand, the absence of a shorter path between v and w means that no vertex in B can be joined to more than three vertices of A (and if it joined to exactly three of them, then they must be adjacent along the path (6)). Hence,

$$e(A, B) \leq 3|B| = 3(n - \delta - 1).$$

From (6) and (7) one easily deduces (2).

Remark 9. The proof just given can be easily adapted to construct explicit examples of graphs which show that the upper bound in (2) is essentially best-possible. Let $d \geq 5$ be odd for simplicity and choose a non-negative integer k. Let

$$a := 3(k + 1), \quad b := k(d - 2) + 2(d - 1) = (k + 2)(d - 2) + 2, \quad n := a + b.$$

We construct a d-regular graph on n vertices as follows. The vertices of $\mathcal{G}$ are partitioned into two disjoint sets A and B such that $|A| = a$, $|B| = b$. Denote

$$A := \{v_1, \ldots, v_a\}, \quad B := \{w_1, \ldots, w_b\}.$$

The graph $\mathcal{G}$ will contain the following edges:

Type 1: The edges of the path $v_1 \rightarrow v_2 \rightarrow \cdots \rightarrow v_a$.

Type II: All edges $\{v_i, w_j\}$ such that $1 \leq i \leq 3$ and $1 \leq j \leq d-1$, except the edges $\{v_2, w_{d-1}\}$ and $\{v_3, w_1\}$.

Type III: All edges $\{v_{(a+1)-i}, w_{(b+1)-j}\}$ such that $1 \leq i \leq 3$ and $1 \leq j \leq d - 1$, except the edges $\{v_{a-1}, w_{b+2-d}\}$ and $\{v_{a-2}, w_b\}$.

Type IV: All edges $\{v_{3r+s}, w_{(d-2)r+1+t}\}$ such that $1 \leq r \leq k$, $1 \leq s \leq 3$ and $1 \leq t \leq d - 2$.

Type V: The complete subgraph on the vertices $w_1, \ldots, w_{d-1}$, minus a perfect matching on the $d - 3$ vertices $w_2, \ldots, w_{d-2}$.

Type VI: The complete subgraph on the vertices $w_{b+2-d}, \ldots, w_b$, minus a perfect matching on the $d - 3$ vertices $w_{b+3-d}, \ldots, w_{b-1}$.

Type VII: For each $1 \leq r \leq k$, the complete subgraph on the vertices $w_{r(d-2)+2}, \ldots, w_{(r+1)(d-2)+1}$.

One can readily check that this graph is indeed d-regular and, for sufficiently large k, of diameter $a - 1 := \delta$. Moreover, for $d \geq 5$ one has

$$\delta = \left\lfloor \frac{3n - (d+3)}{d+1} \right\rfloor.$$

For further known results on the maximum diameter of regular (di)graphs, the interested reader is referred to [5], for example.

4. Concluding Remarks

There are two obvious directions in which the results of this paper need to be improved upon. The first is to generalise them to h-fold sumgraphs for arbitrary h, and in particular to understand better the most natural case when $h = 2$. The second is to sharpen them, in particular to obtain the best-possible constant ϵ in Theorem 5. Both directions naturally lead in turn to Freiman-type inverse problems, where one wishes to say something about the 'structure' of regular, connected graphs whose sumgraphs grow slowly.

Finally, we note that, while studying Cayley graphs purely from the point of view of sumgraph growth appears to be a new idea, a related property of such graphs - that of possessing short cycles - has been previously studied in connection with the so-called Caccetta-Häggkvist conjecture. See, for example, [1] and [4].

References

[1] Y.O. Hamidoune, *An application of connectivity theory in graphs to factorisations of elements in groups*, European J. Combin. **2**, no. 4. (1981), 349–355.

[2] F. Harary, *Sum graphs and difference graphs*, Congr. Numer. **72** (1990), 101–108.

[3] M. Kneser, *Abschätzung der asymptotischen Dichte von Summenmengen*, Math. Z. **58** (1953), 459–484.

[4] M.B. Nathanson, *The Caccetta-Häggkvist conjecture and additive number theory*, Analytic Number Theory, pp. 347–358, Cambridge Univ. Press, Cambridge (2009).

[5] J. Soares, *Maximum diameter of regular digraphs*, J. Graph Theory **16**, no. 5 (1992), 437–450.