Received: 23 August 2022 | Revised: 1 June 2023 | Accepted: 19 July 2023

DOI: 10.1002/jgt.23024

ARTICLE

WILEY

The average order of a connected induced subgraph of a graph and union-intersection systems

Andrew Vince

Department of Mathematics, University of Florida, Gainesville, Florida, USA

Correspondence
Andrew Vince, Department of Mathematics, University of Florida, Gainesville, FL, USA.
Email: avince@ufl.edu

Funding information
Simons Foundation,
Grant/Award Number: 322515

Abstract

Because connectivity is such a basic concept in graph theory, extremal problems concerning the average order of the connected induced subgraphs of a graph have been of notable interest. A particularly resistant open problem is whether or not, for a connected graph G of order n, all of whose vertices have degree at least 3, this average is at least $n/2$. It is shown in this paper that if G is a connected, vertex transitive graph, then the average order of the connected induced subgraphs of G is at least $n/2$.

The extremal graph theory problems mentioned above lead to a broader theory. The concept of a Union-Intersection System (UIS) $\mathcal{S} := (P, \mathcal{B})$ is introduced, P being a finite set of points and $\mathcal{B}$ a set of subsets of P called blocks satisfying the following simple property for all $A, B \in \mathcal{B}$: if $A \cap B \in \mathcal{B}$, then $A \cup B \in \mathcal{B}$. To generalize results on the average order of a connected induced subgraph of a graph, it is conjectured that if a UIS is, in various senses, "connected and regular," then the average size of a block is at least half the number of points. We prove that if a union-intersection set system is regular, completely irreducible, and nonredundant, then the average size of a block is at least half the number of points.

KEYWORDS

average order, connectedness, extremal, set system

MATHEMATICS SUBJECT CLASSIFICATION 2010

05C30, 05C35, 05C40, 05D05

1 | INTRODUCTION

The average value of various graph invariants is currently a major research topic. In particular, the average order of a connected induced subgraph of a graph has received considerable attention recently; see, for example [2, 5–7, 10–14, 16–18] and references therein. Extremal problems concerning the average order of a connected induced subgraph of a graph originated in the 1983 and 1984 papers of Jamison [8, 9] on the average order of a subtree of a tree. For a tree, a connected induced subgraph is simply a subtree. Jamison proved that the average order of a subtree, over all trees of order n, is minimized by the path P_n.

For a graph G of order n, call a set U of vertices *connected* if U induces a connected subgraph of G. Let $\mathcal{C} := \mathcal{C}(G)$ denoted the set of all connected sets of G, and let

$$av(G) := \frac{\sum_{U\in\mathcal{C}}|U|}{|\mathcal{C}|} \quad \text{and} \quad d(G) := \frac{av(G)}{n}$$

denote the average size of a connected set and the proportion of vertices in an average size connected set. The parameter $d(G)$ is called the *density* of G and allows for the comparison of the average size for graphs of different orders. For example, if $d(G) = 1/2$, then the average size of a connected induced subgraph is half the order of G. With this notation, the result of Jamison is that $av(T) \geq (n+2)/3$ for every tree T of order n with equality if and only if $T = P_n$. Therefore $d(T) > 1/3$ for every tree T. It had been open since Jamison's work, and conjectured formally by Kroeker, Mol, and Oellermann in 2018 [10], that P_n minimizes the average size of a connected set over all connected graphs of order n. This conjecture was proved in 2022:

Theorem 1.1 (Haslegrave [6] and Vince [15]). *If G is a connected graph, then $d(G) > \frac{1}{3}$.*

With respect to the 1/3 bound in Theorem 1.1, vertices of degree 2 play a special role. It is therefore natural to ask about a lower bound for graphs with no vertex of degree 2. For trees, the following result appeared in 2010.

Theorem 1.2 (Vince and Wang [13]). *For a tree T with no vertex of degree 2 we have $d(T) > \frac{1}{2}$, the bound being tight.*

Concerning a lower bound for graphs in general, we posed the following conjectures in 2020. In the first, the hypothesis is combinatorial, in the second group theoretic.

Conjecture 1 (Vince [14]). *If G is a connected graph, all of whose vertices have degree at least 3, then $d(G) \geq \frac{1}{2}$.*

Conjecture 2. *If G is a connected, vertex transitive graph, then $d(G) \geq \frac{1}{2}$.*

Although Conjecture 1 remains open, Conjecture 2 is confirmed in this paper. It is statement (2) of Theorem 1.4. The proof of Theorem 1.4, which is provided in Section 3, relies on a strong connection between the density $d(G)$ and the maximum of the ratio of the number of connected sets of G containing a given vertex v to the total number of connected sets. With

$$N(G) = |\mathcal{C}(G)| \quad \text{and} \quad N(G,v) = |\{U \in \mathcal{C}(G) : v \in U\}|,$$

this maximum ratio is

$$R(G) := \max_{v \in V(G)} \frac{N(G,v)}{N(G)}.$$

The connection between $d(G)$ and $R(G)$ is made precise in Proposition 1.3, stated in a more general setting.

A *set system* $(P, \mathcal{B})$ consists of a finite nonempty set P of *points* and a nonempty collection $\mathcal{B}$ of subsets of P called *blocks*. Call $n := |P|$ the *order* of the set system. Call a set system *regular* if each point is contained in the same number of blocks. Extend notation from a graph to a set system $\mathcal{S}$ of order n as follows. Let

$$av(\mathcal{S}) := \frac{\sum_{X \in \mathcal{B}} |X|}{|\mathcal{B}|} \quad \text{and} \quad d(\mathcal{S}) := \frac{av(\mathcal{S})}{n} \tag{1}$$

denote the average size of a block and the proportion of points in an average size block. Further, for $x \in P$, define

$$\begin{aligned} N(\mathcal{S}) &= |\mathcal{B}|, \\ N(\mathcal{S}, x) &= |\{A : x \in A \in \mathcal{B}\}|, \\ R(\mathcal{S}) &= \max_{x \in P} \frac{N(\mathcal{S}, x)}{N(\mathcal{S})}. \end{aligned} \tag{2}$$

In applying Proposition 1.3 to a graph G with vertex set V, the relevant set system is

$$\mathcal{S}(G) = (V, \mathcal{C}(G)).$$

Therefore $N(G) = N(\mathcal{S}(G))$ and $N(G, v) = N(\mathcal{S}(G), v)$. Statement (1) in Proposition 1.3 can be restated as follows: there is a point x such that the probability that a randomly selected block contains x is at least α.

Proposition 1.3. *For a regular set system $\mathcal{S} = (P, \mathcal{B})$ of order n and a real number α, $0 \le \alpha \le 1$, the following statements are equivalent:*

(1) $R(\mathcal{S}) \ge \alpha$.
(2) $N(\mathcal{S}, x)/N(\mathcal{S}) \ge \alpha$ *for every* $x \in P$.
(3) $d(\mathcal{S}) \ge \alpha$.

Proof. Because every point of $\mathcal{S}$ is contained in the same number of blocks, statements (1) and (2) are equivalent.

Let $x_0 \in P$ be an arbitrary point of $\mathcal{S}$ and count the number of elements in the set $\{(x, X) : x \in X \in \mathcal{B}\}$ in two ways to obtain

$$\sum_{X \in \mathcal{B}} |X| = |\{(x, X) : x \in X \in \mathcal{B}\}| = \sum_{x \in P} N(\mathcal{S}, x) = nN(\mathcal{S}, x_0).$$

Therefore

$$d(\mathcal{S}) = \frac{\sum_{X \in \mathcal{B}} |X|}{nN(\mathcal{S})} \geq \alpha \quad \text{if and only if} \quad \frac{N(\mathcal{S}, x_0)}{N(\mathcal{S})} \geq \alpha.$$

□

Theorem 1.4. *Let G be a connected graph. Then*

(1) $R(G) \geq \frac{1}{2}$;
(2) *if G is vertex transitive, then* $d(G) \geq \frac{1}{2}$;
(3) *there is an edge of G that is contained in at least half of all connected (not necessarily induced) subgraphs of G of order* ≥2.

Remark 1.5. The 1/2 lower bound on $R(G)$ in statement (1) of Theorem 1.4 is the best possible. For example, for the path P_n of even and odd orders n, it is not difficult to show that for a "central" vertex v

$$\frac{N(P_n, v)}{N(P_n)} = \frac{1}{2}\left(1 + \frac{1}{n-1}\right) \quad \text{and} \quad \frac{N(P_n, v)}{N(P_n)} = \frac{1}{2}\left(1 + \frac{1}{n}\right),$$

respectively.

The 1/2 lower bound in statement (2) of Theorem 1.4 is the best possible because it is attained, for example, by complete graphs and by cycles.

The 1/2 lower bound in statement (3) is also the best possible. An example is the cycle of order n. There are $2\binom{n}{2}$ connected subgraphs of order ≥2, and each edge is contained in exactly $\binom{n}{2}$ of them.

For a tree with no vertex of degree 2, Theorem 1.6 is a much stronger result than statement (1) of Theorem 1.4. Statement (1) of Theorem 1.4 states that $R(G) \geq 1/2$ for any connected graph G. Theorem 1.6 states that, for a tree T with no vertex of degree 2, the ratio $R(T) = 1 - o(1)$. The proof of Theorem 1.6 appears in Section 3.

Theorem 1.6. *If T is a tree of order n with no vertex of degree* 2, *then*

$$R(T) > 1 - \frac{9}{2n + 10}.$$

The lower bound on $R(T)$ in Theorem 1.6 is not the best possible. We pose the following conjecture concerning the extremal graph. An affirmative answer to Conjecture 3 would imply that there is a constant c such that, for every tree T of order n with vertex set V and with no vertex of degree 2,

$$R(T) \geq 1 - \frac{c}{2^{n/4}}.$$

Conjecture 3. *For every tree of T of order n with no vertex of degree* 2, *the ratio* $R(T)$ *is minimized when T is one of the caterpillars in Figure* 1, *depending on n modulo* 4.

2 | A BROADER CONTEXT: UISs

If two connected sets of vertices of a graph have nonempty intersection, then their union is also a connected set. That simple fact is the motivation for Definition 1.

Definition 1. A set system $(P, \mathcal{B})$ is a *UIS* if the following property holds for all $A, B \in \mathcal{B}$:

- $A \cup B \in \mathcal{B}$ whenever $A \cap B \neq \emptyset$.

Example 1 (Graph UIS)**.** For a graph G with vertex set V and collection $\mathcal{C}(G)$ of connected vertex sets, the set system $\mathcal{S}(G) = (V, \mathcal{C}(G))$ is a UIS, called a *graph UIS*.

Example 2 (Hypergraph UIS)**.** A hypergraph $H = (V, E)$ consists of a nonempty set V of vertices and a set E of nonempty subsets of V called edges. See [1, 3] as references on hypergraphs. Given a hypergraph $H = (V, E)$, a hypergraph UIS $\mathcal{S}(H)$ can be defined in a manner analogous to a graph UIS. Given $U \subseteq V$, the *hypergraph* $H(U)$ *induced by* U is the hypergraph with vertex set U and edge set $\{e \cap U : e \in E\}$. A hypergraph H is *connected* if, for any pair x, y of vertices, there is a path which connects them, a path being a vertex-edge alternating sequence $x = x_1, e_1, x_2, e_2, \ldots, x_s, e_s, x_{s+1} = y$ such that $x_i, x_{i+1} \in e_i$ for $i = 1, 2, \ldots, s$. A set $U \subseteq V$ is said to be a *connected set of vertices* or simply a *connected set* if the induced hypergraph $H(U)$ is connected. If $\mathcal{C}(H)$ denotes the collection of all connected vertex sets of H, then the UIS $\mathcal{S}(H) = (V, \mathcal{C}(H))$ will be referred to as a *hypergraph UIS*.

Remark 2.1 (Not every UIS is a hypergraph UIS)**.** Not every UIS, not even every completely irreducible UIS, is a hypergraph UIS. A simple example is $\mathcal{S} = (P, \mathcal{B})$, where $P = [5]$ and

$$\mathcal{B} = \{\{1, 2, 3\}, \{3, 4, 5\}, \{1, 2, 3, 4, 5\}\}.$$

The UIS $\mathcal{S}$ is not a hypergraph UIS because if there exists a hypergraph H such that $\mathcal{S} = \mathcal{S}(H)$, then the sets $\{1, 2, 3\}$ and $\{3, 4, 5\}$ would have to be edges of H. Hence the set $\{2, 3, 4\}$ would be a connected set in H and hence a block of $\mathcal{S}$.

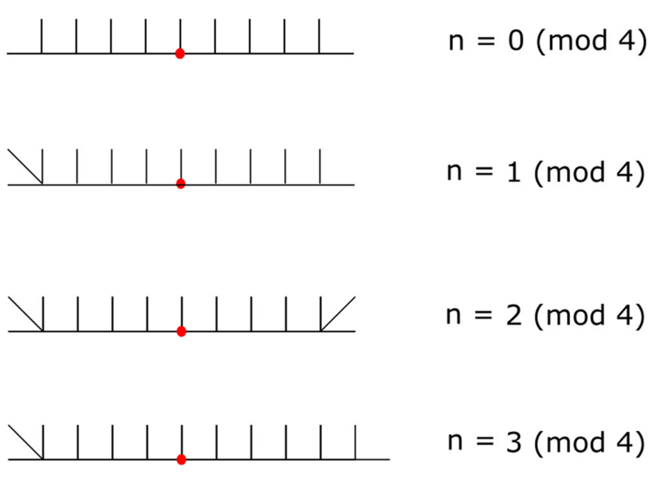

FIGURE 1 Conjectured extremal trees. The maximizing vertex v is colored red. [Color figure can be viewed at wileyonlinelibrary.com]

Call a set system $(P, \mathcal{B})$ of order n *trivial* if all the blocks have size either n or 1. The notation $[n] := \{1, 2, ..., n\}$ will be used. Recall that a set system is *regular* if each point is contained in the same number of blocks. In the graph case, it was always assumed that G is a connected graph. This motivates the following generalization.

Definition 2. A nontrivial set system $\mathcal{S} = (P, \mathcal{B})$ is called *reducible* if there is a partition (P_1, P_2) of P into two nonempty parts such that if $A \in \mathcal{B} \setminus \{P\}$, then $A \subseteq P_1$ or $A \subseteq P_2$. For example, if $\mathcal{S}_1 = (P, \mathcal{B}_1)$ and $\mathcal{S}_2 = (P, \mathcal{B}_2)$, where $P = [4]$ and

$$\mathcal{B}_1 = \{\{1, 2\}, \{3, 4\}\} \quad \text{and} \quad \mathcal{B}_2 = \{\{1, 2\}, \{3, 4\}, \{1, 2, 3, 4\}\},$$

then both $\mathcal{S}_1$ and $\mathcal{S}_2$ are reducible. If $\mathcal{S}$ is not reducible, it is called *irreducible*.

If $Q \in \mathcal{B}$ and $\mathcal{B}(Q) = \{A \in \mathcal{B} : A \subseteq Q\}$, then $(Q, \mathcal{B}(Q))$ is called the *subsystem induced by block* Q. A set system $\mathcal{S}$ is called *completely irreducible* if $\mathcal{S}$ itself and all subsystems induced by its blocks are irreducible.

The *automorphism group* of a set system $(P, \mathcal{B})$ is the group of all bijections $g : P \to P$ such that $A \in \mathcal{B}$ if and only if $g(A) \in \mathcal{B}$. Call a set system *transitive* if its automorphism group acts transitively on P. Note that a transitive set system is regular. Call a set system *primitive* if its automorphism group acts primitively on P. Recall that a permutation group Γ acting on a set P of size n is called *primitive* if the only partitions of P that are preserved by the Γ-action are the trivial partitions into either a single part of size n or into n parts of size 1. Primitive permutation groups are transitive; otherwise orbits of the group Γ form a nontrivial partition preserved by Γ.

Proposition 2.2. *Let $G = (V, E)$ be a graph and $\mathcal{S}(G) = (V, \mathcal{C}(G))$ its UIS. Then*

(1) *G is a connected graph if and only if $\mathcal{S}(G)$ is irreducible;*
(2) *the graph G is vertex transitive if and only if $\mathcal{S}(G)$ is transitive; and*
(3) *the automorphism group of G acts primitively on the set of vertices of G if and only if $\mathcal{S}(G)$ is primitive.*

Proof. Concerning statement (1), assume that G is not connected. Let V_1 be the set of vertices in a connected component of G and $V_2 = V \setminus V_1$. If A is a connected set of vertices of G, that is, a block of $\mathcal{S}(G)$, then $A \subseteq V_1$ or $A \subseteq V_2$. Hence $\mathcal{S}$ is reducible.

Conversely assume that $\mathcal{S}$ is reducible in $\mathcal{S}(G)$. Then there is a partition (V_1, V_2) of V into two nonempty parts such that if $A \neq V$ is a connected vertex set then $A \subseteq V_1$ or $A \subseteq V_2$. Therefore, V_1 and V_2 are unions of connected components of G and hence G is not connected.

Statements (2) and (3) follow because a bijection $g : V \to V$ preserves adjacency in G if and only if g preserves connectedness of subsets of vertices of G, that is, if and only if it takes blocks to blocks in $\mathcal{S}(G)$. □

Proposition 2.3. *If $\mathcal{S} = (P, \mathcal{B})$ is a nontrivial, primitive UIS, then $\mathcal{S}$ is transitive and irreducible.*

Proof. It has already been mentioned that a primitive permutation group must be transitive. By way of contradiction, assume that $\mathcal{S}$ is reducible. Because $\mathcal{S}$ is nontrivial, there is a block of size at least 2 properly contained in P. Let $A_1, A_2, ..., A_k$ be the maximal blocks of $\mathcal{S}$ in $\mathcal{B}\backslash\{P\}$. If $\mathcal{S}$ is not transitive, then $\mathcal{S}$ is not primitive, and we are done. Hence, assume that $\mathcal{S}$ is transitive, which implies that $k \geq 2$ and that every point is contained in a block A_i for some i. By Definition 1 and the fact that the A_i are maximal, either (1) the A_i are pairwise disjoint or (2) there is a pair $A_i, A_j, i \neq j$, such that $A_i \cap A_j \neq \emptyset$ and $A_i \cup A_j = P$. In case (2), $\mathcal{S}$ cannot be reducible, a contradiction. In case (1), $(A_1, A_2, ..., A_k)$ is a partition of P that is left invariant by the automorphism group of $\mathcal{S}$. Hence $\mathcal{S}$ is not primitive. $\square$

Remark 2.4. That a UIS $\mathcal{S}$ is primitive does not necessarily imply that $\mathcal{S}$ is completely irreducible. An example is $\mathcal{S}(C_n)$, where C_n is an n-cycle. The dihedral group acts primitively on the set of vertices of C_n. However, if u and v are adjacent vertices of C_n, then $\{u\}, \{v\}, \{u, v\}$ are blocks of $\mathcal{S}(C_n)$, and the partition $(\{u\}, \{v\})$ shows that the block $\{u, v\}$ is reducible, hence $\mathcal{S}(C_n)$ is not completely irreducible.

The density d and parameter R were extended from graphs to set systems in Equations (1) and (2). It is natural to ask what is true for $d(\mathcal{S})$ and $R(\mathcal{S})$ in this more general setting. We formulate the following conjectures.

Conjecture 4. *If $\mathcal{S}$ is a nontrivial, primitive UIS, then $R(\mathcal{S}) \geq 1/2$.*

Conjecture 5. *If $\mathcal{S}$ is a completely irreducible UIS, then $R(\mathcal{S}) \geq 1/2$.*

Conjecture 6. *If $\mathcal{S}$ is a completely irreducible UIS, then $d(\mathcal{S}) > \frac{1}{3}$.*

Conjecture 7. *If $\mathcal{S}$ is a nontrivial, primitive UIS, then $d(\mathcal{S}) \geq \frac{1}{2}$.*

Conjecture 8. *If $\mathcal{S}$ is a completely irreducible and regular UIS, then $d(\mathcal{S}) \geq \frac{1}{2}$.*

Remark 2.5. In Conjecture 4, assuming only that the UIS $\mathcal{S}$ is nontrivial and transitive is not sufficient to conclude that $R(\mathcal{S}) \geq 1/2$. An example is $\mathcal{S} = ([2n], \mathcal{B}_1 \cup \mathcal{B}_2)$ for $n \geq 2$, where $\mathcal{B}_1$ is the set of all subsets of $[n]$ and $\mathcal{B}_2$ is the set of all subsets of $\{n+1, n+2, ..., 2n\}$. Each point is contained in only 2^{n-1} of the $2(2^n - 1) > 2 \cdot 2^{n-1}$ blocks.

Remark 2.6. Conjecture 5 may fail if $\mathcal{S}$ is not completely irreducible. An example is the UIS $\mathcal{S} = (P, \mathcal{B})$ with $P = [5]$ and

$$\mathcal{B} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{1, 2\}, \{1, 2, 5\}, \{3, 4\}, \{3, 4, 5\}, \{1, 2, 3, 4\}, \{1, 2, 3, 4, 5\}\}.$$

In this example, $\mathcal{S}$ itself is irreducible, but the subsystem induced by the block $\{1, 2, 3, 4\}$ is not. Each point of $\mathcal{S}$ is contained in at most 5 out of the $11 > 2 \cdot 5$ blocks.

Remark 2.7. The statements of Conjectures 4 and 5 are similar in structure to that of the well-studied, but still open, Union-Closed Sets Conjecture posed by Peter Frankl in 1979; see the survey [4] and references therein. This paper, however, does not address the Union-Closed Sets Conjecture.

Remark 2.8. The 1/3 lower bound on the density in Conjecture 6 is the best possible. With the notation $[a, b] := \{a, a+1, a+2, ..., b\}$ for integers $a < b$, an example is the UIS $([n], \mathcal{B})$, where

$$\mathcal{B} = \{[a, b] : 1 \le a < b \le n\}.$$

In this example, the density is $\frac{1}{3}(1 + \frac{4}{n})$.

Remark 2.9. Conjecture 7 is, according to Proposition 1.3, equivalent to Conjecture 4. Conjecture 5, if correct, implies Conjecture 8 via Proposition 1.3.

Remark 2.10. Assuming that Conjectures 7 and 8 are correct, the 1/2 lower bound on the density is asymptotically best possible. An example is the UIS $\mathcal{S} = ([n], \mathcal{B})$, where $\mathcal{B}$ consists of all nonsingleton subsets of $[n]$. The density is $d(\mathcal{S}) = \frac{1}{2}\left(1 + \frac{n-2}{2^n - n}\right)$.

Although Conjectures 4, 5, 6, 7, and 8 remain open, several closely related results are obtained. These results, Theorems 2.11 and 2.12, are stated below and proved in Section 3. Moreover, if G is a graph, then Conjectures 4, 5, 6, and 7 are true for $\mathcal{S}(G)$. Statement (1) of Theorem 1.4, together with Propositions 2.2 and 2.3, implies the graph versions of Conjectures 4 and 5. Theorem 1.1 and Proposition 2.2 imply the graph version of Conjecture 6. Statement (2) of Theorem 1.4, together with Propositions 2.2 and 2.3, imply the graph version of Conjecture 7. Statements (1) and (2) of Theorem 1.4 are actually stronger than the graph versions of the relevant conjectures because the assumption that the UIS $\mathcal{S}(G)$ is completely irreducible is not required, only that $\mathcal{S}(G)$ be irreducible, that is, that G be connected.

Theorem 2.11. *If H is a connected hypergraph, then*

(1) *There is a vertex that is contained in at least half of the connected vertex sets of H, and*
(2) *if H is vertex transitive, then $d(H) \ge \frac{1}{2}$.*

For a UIS $\mathcal{S} = (P, \mathcal{B})$, let $\mathcal{M} := \mathcal{M}(\mathcal{S})$ denote the set of all minimal blocks, where minimal is with respect to containment. If $\mathcal{M}' \subseteq \mathcal{M}$, then define

$$\overline{\mathcal{M}'} = \bigcup\{A : A \in \mathcal{M}'\} \subseteq P.$$

Call $\mathcal{S}$ *redundant* if there are distinct subsets $\mathcal{M}_1$ and $\mathcal{M}_2$ of $\mathcal{M}$ such that $\overline{\mathcal{M}_1} = \overline{\mathcal{M}_2}$; otherwise *nonredundant*.

Theorem 2.12. *If a UIS $\mathcal{S}$ is completely irreducible and nonredundant, then*

(1) *there is a block of $\mathcal{S}$ that is a subset of at least half the blocks of $\mathcal{S}$, and*
(2) *if, in addition, $\mathcal{S}$ is regular, then $d(\mathcal{S}) \ge \frac{1}{2}$.*

Theorem 2.12 gives the conclusions of Conjectures 5 and 8, but requires the additional assumption that $\mathcal{S}$ is nonredundant. The conclusion of the statement (1) of Theorem 2.12 is somewhat stronger than the conclusion of Conjecture 5 in that a whole block, not just a single point, is contained in half of the blocks of $\mathcal{S}$.

3 | PROOFS OF THE THEOREMS

This section contains the proofs of Theorems 1.4, 1.6, 2.11, and 2.12, in that order.

3.1 | Theorem 1.4

The following notation is used in the proof, where G is a graph and U is a set of vertices of G.

$$\begin{aligned} N = N(G) &= \text{ the number of connected sets in } G, \\ N(U) = N(G, U) &= \text{ the number of connected sets in } G \text{ containing } U, \\ \overline{N}(U) = \overline{N}(G, U) &= \text{ the number of connected sets in } G \text{ disjoint from } U, \\ N(U \neg U') &= \text{ the number of connected sets in } G \text{ containing } U \text{ but disjoint from } U'. \end{aligned}$$

Lemma 3.1. *If G is a connected graph of order n, then for every k, $1 \leq k \leq n$, there is a connected set U of size k such that $N(U) \geq \overline{N}(U)$.*

Proof. Let G be a connected graph of order n. The proof is by backward induction on k. The statement is clearly true for $k = n$, even counting the empty set as a connected set. Assuming that the lemma is true for k with $k \geq 2$, we will prove that it is true for $k - 1$.

Let U be the connected set of size k such that $N(U) \geq \overline{N}(U)$. Let $v_1, v_2 \in U$ be such that both $U_1 := U \setminus \{v_1\}$ and $U_2 := U \setminus \{v_2\}$ remain connected sets. For example, take v_1, v_2 to be any two leaves in a spanning tree of the subgraph of G induced by U. We claim that $N(U_1) \geq \overline{N}(U_1)$ or $N(U_2) \geq \overline{N}(U_2)$, which would conclude the proof of the lemma.

To prove the claim, first note that, for $i = 1, 2$, we have

$$\begin{aligned} N(U_i) &= N(U) + N(U_i \neg v_i), \\ \overline{N}(U_i) &= \overline{N}(U) + N(v_i \neg U_i). \end{aligned}$$

Subtracting the two equalities and using the induction hypothesis yields

$$\begin{aligned} N(U_i) - \overline{N}(U_i) &= N(U) - \overline{N}(U) + (N(U_i \neg v_i) - N(v_i \neg U_i)) \\ &\geq N(U_i \neg v_i) - N(v_i \neg U_i). \end{aligned}$$

It is now sufficient to show that $N(U_i \neg v_i) \geq N(v_i \neg U_i)$ for $i = 1$ or $i = 2$. By way of contradiction, assume that $N(v_i \neg U_i) > N(U_i \neg v_i)$ for $i = 1, 2$. Summing yields

$$N(v_1 \neg U_1) + N(v_2 \neg U_2) > N(U_1 \neg v_1) + N(U_2 \neg v_2). \tag{3}$$

For $i = 1, 2$, let $\mathcal{A}_i$ be the collection of all sets containing v_i but disjoint from U_i; thus $|\mathcal{A}_i| = N(v_i \neg U_i)$. Let $\mathcal{B}_i$ be the collection of all sets containing U_i but disjoint from v_i; thus $|\mathcal{B}_i| = N(U_i \neg v_i)$. Define a map $\phi_1 : \mathcal{A}_1 \to \mathcal{B}_2$ by $\phi_1(A) = A \cup U_2 \in \mathcal{B}_2$, and define a map $\phi_2 : \mathcal{A}_2 \to \mathcal{B}_1$ by $\phi_2(A) = A \cup U_1 \in \mathcal{B}_1$. Note that $\phi_1(A) \in \mathcal{B}_2$ because $v_1 \in A \cap U_2$; similarly $\phi_1(A) \in \mathcal{B}_1$. Because both ϕ_1 and ϕ_2 are injections, we have

$$N(U_2 \neg v_2) \ge N(v_1 \neg U_1) \quad \text{and} \quad N(U_1 \neg v_1) \ge N(v_2 \neg U_2),$$

which contradicts inequality (3). □

Proof of Theorem 1.4. Concerning statement (1), by Lemma 3.1 with $k = 1$, there is a vertex u such that $N(u) \ge \overline{N}(u)$. Therefore $2N(u) \ge N(u) + \overline{N}(u) = N$ and $\frac{N(u)}{N} \ge \frac{1}{2}$.

Concerning statement (2), by statement (1) we have $R(G) \ge 1/2$. By Proposition 2.2, G vertex transitive implies that $\mathcal{S}(G)$ is transitive, hence also regular. By Proposition 1.3 with $\alpha = 1/2$ and Proposition 2.2 this implies that $d(G) = d(\mathcal{S}(G)) \ge \frac{1}{2}$.

Concerning statement (3), let G^* be the line graph of G. The vertices of G^* are the edges of G and two vertices of G^* are adjacent if the corresponding edges of G are incident. Since G is connected, G^* is also connected. For a connected vertex set H^* of G^*, let H be the corresponding (not necessarily induced) subgraph of G. Then the mapping $H^* \mapsto H$ is a bijection between the collection of connected vertex sets of G^* and the connected subgraphs of G of order ≥2. By statement (1) of Theorem 1.4 there is a vertex of G^*, that is, an edge of G, that is contained in at least half of the connected vertex sets of G^*, that is, half of the connected subgraphs of G of order ≥2. □

3.2 | Theorem 1.6

The following notation is used in the proof. For a vertex v of degree k in T let $T_1, T_2, \ldots, T_k$ be the connected components of $T \setminus v$, and let v_i be the vertex of T_i adjacent to v. For $1 \le i \le k$, denote by N_i, $\overline{N_i}$, and n_i the number of subtrees of T_i containing v_i, the number of subtrees of T_i not containing v_i, and the order of T_i, respectively.

Lemma 3.2. *Let T be a tree of order n rooted at a vertex v of degree $k \ge 2$. If T has no vertex of degree 2 except possibly the root, then $N(v) \ge \overline{N}(v) + \frac{n+1}{2}$.*

Proof. The proof is by induction on the order n of the tree. If $n = 1$ then $N(v) = 1, \overline{N}(v) = 0$ and the statement is true. Assume the statement true for all such trees of order less than n. For a tree of order n we have

$$N(v) = \prod_{i=1}^{k}(N_i + 1) \ge 2\sum_{i=1}^{k} N_i \ge \sum_{i=1}^{k}(N_i + \overline{N_i}) + \sum_{i=1}^{k} \frac{n_i + 1}{2} \ge \overline{N}(v) + \frac{n + k - 1}{2}$$
$$\ge \overline{N}(v) + \frac{n+1}{2},$$

where the second inequality is by the induction hypothesis. The first inequality, for $k = 2$, is equivalent to $(N_1 - 1)(N_2 - 1) \ge 0$. For $k \ge 3$, the inequality is easily proved by induction on k. □

Proof of Theorem 1.6. That there exists a vertex that satisfies the first inequality below is [13, Lemma 5]. The second inequality is from Lemma 3.2; and the last inequality is because $\overline{N} \geq n-1$.

$$
\begin{aligned}
N(v) &= \prod_{i=1}^{k}(N_i+1) \geq \frac{4}{9}\left(\sum_{i=1}^{k} N_i\right)^2 = \frac{1}{9}\left(\sum_{i=1}^{k}(N_i+N_i)\right)^2 \geq \frac{1}{9}\left(\sum_{i=1}^{k}\left(N_i+\overline{N}_i+\frac{n_i+1}{2}\right)\right)^2 \\
&= \frac{1}{9}\left(\overline{N}(v)+\frac{n+k-1}{2}\right)^2 \geq \frac{1}{9}\left(\overline{N}(v)+\frac{n+2}{2}\right)^2 \geq \frac{1}{9}\left(\overline{N}(v)^2+(n+2)\overline{N}(v)+\frac{(n+2)^2}{4}\right) \\
&\geq \frac{1}{9}\left((n-1)\overline{N}(v)+(n+2)\overline{N}(v)+\frac{(n+2)^2}{4}\right) = \frac{1}{9}\left((2n+1)\overline{N}(v)+\frac{(n+2)^2}{4}\right).
\end{aligned}
$$

Now

$$
\begin{aligned}
\frac{1}{9}(2n+10)N(v) &= N(v)+\frac{1}{9}(2n+1)N(v) \\
&= \frac{1}{9}\left((2n+1)\overline{N}(v)+(2n+1)N(v)+\frac{(n+2)^2}{4}\right) \\
&= \frac{1}{9}\left((2n+1)N+\frac{(n+2)^2}{4}\right).
\end{aligned}
$$

Therefore

$$
N(v) \geq \left(1-\frac{9}{2n+10}\right)N+\frac{(n+2)^2}{8(n+5)}.
$$

□

3.3 | Theorem 2.11

Given a hypergraph H, the hypergraph UIS $\mathcal{S}(H)$ was defined in Section 2. One might expect that the collection of graph UISs is properly contained in the collection of hypergraph UISs. This is not the case, as implied by Lemma 3.3. Two UISs are *isomorphic* if there is a bijection between their point sets that preserves blocks. Isomorphism of $\mathcal{S}_1$ and $\mathcal{S}_2$ will be denoted by $\mathcal{S}_1 = \mathcal{S}_2$.

Lemma 3.3. *For every hypergraph H there is a simple graph G with the same vertex set as H and such that $\mathcal{S}(H) = \mathcal{S}(G)$. In particular, H is connected if and only if G is connected.*

Proof. Given a hypergraph $H = (V, E)$, define a graph $G = (V, \widetilde{E})$, where $\widetilde{E} = \{\{u, v\} : u, v \in e$ for some $e \in E\}$. It is now sufficient to show that H and G have the same collection of connected vertex sets.

If C is a connected set in H and $x, y \in C$, then there is a path $x = x_1, e_1, x_2, e_2, \ldots, x_s, e_s, x_{s+1} = y$ in the hypergraph $H(C)$ induced by C. For each edge e_i in the induced graph hypergraph, there is an edge $e_i' \in E$ such that $e^i \subseteq e_i'$. Since there is an edge e_i' in the hypergraph H containing x_i and x_{i+1}, there must also be an edge between x_i and x_{i+1} in the associated graph G. Therefore C is a connected set in G.

If C is a connected set in G and $x, y \in C$, then there is a path $x = x_1, e_1, x_2, e_2, \dots, x_s, e_s, x_{s+1} = y$ in G, where $e_i \in \widetilde{E}$ for all i. Each $e_i \subseteq e_i' \cap C$ for some edge $e_i' \in E$. Let $e_i'' = e_i' \cap C$, which is an edge in $H(C)$. Now $x = x_1, e_1'', x_2, e_2'', \dots, x_s, e_s'', x_{s+1} = y$ is a path joining x and y in $H(C)$. Therefore C is connected in $H(C)$ and hence a connected set in H. □

Proof of Theorem 2.11. Concerning statement (1), let G be the graph associated with H as in Lemma 3.3. By that lemma, G is connected because H is connected. By Theorem 1.4 there is a vertex v in G that is contained in at least half of the connected vertex sets of G. By Lemma 3.3 the collection of connected sets of G is the same as the collection of connected sets of H. Therefore, there is a vertex of H that is contained in at least half the connected vertex sets of H.

Concerning statement (2), since H is vertex transitive, so is the graph G. By Theorem 1.4 $d(G) \geq 1/2$. Since the collection of connected sets of G is the same as the collection of connected sets of H, we have $d(H) = d(G) \geq 1/2$. □

3.4 | Theorem 2.12

Lemma 3.4. *Let $\mathcal{S} = (P, \mathcal{B})$ be a UIS. If $\mathcal{S}$ is irreducible, then P is a block.*

Proof. By way of contradiction, assume that P is not a block. Since $\mathcal{B}$ is nonempty by the definition of a set system, there is at least one block. Let A be a maximal block. Let $C \in \mathcal{B}$. If $C \cap A \neq \emptyset$ and $C \nsubseteq A$, then $C \cup A$ contradicts the maximality of A. Therefore, either $C \subseteq A$ or $C \subseteq P \backslash A$, implying that $\mathcal{S}$ is reducible □

Let $(P, \mathcal{M})$ be a set system. There is a unique minimal, with respect to containment, collection $\langle \mathcal{M} \rangle$ of subsets of P such that $\mathcal{M} \subseteq \langle \mathcal{M} \rangle$ and $\mathcal{S}(\mathcal{M}) \coloneqq (P, \langle \mathcal{M} \rangle)$ is a UIS. This unique minimum is given by

$$\langle \mathcal{M} \rangle = \bigcap \{\mathcal{B} : \mathcal{M} \subseteq \mathcal{B} \subseteq 2^P \text{ and } (P, \mathcal{B}) \text{ is a UIS}\}.$$

Call $\mathcal{S}(\mathcal{M})$ the UIS *spanned* by $\mathcal{M}$. This spanned UIS can also be obtained recursively by the following simple routine:

Algorithm M

Input: A set $\mathcal{M}$ of subsets of P

Output: $\langle \mathcal{M} \rangle$

Initialize: Set $\langle \mathcal{M} \rangle = \mathcal{M}$

While there exist $A_1, A_2 \in \langle \mathcal{M} \rangle$ with $A_1 \cap A_2 \neq \emptyset$ and $A_1 \cup A_2 \notin \langle \mathcal{M} \rangle$

  $\langle \mathcal{M} \rangle \leftarrow \langle \mathcal{M} \rangle \cup \{A_1 \cup A_2\}$

If $\mathcal{S} = (P, \mathcal{B})$ is a UIS and $\mathcal{M} \subseteq \mathcal{B}$, then $\mathcal{S}$ is said to be *spanned by* $\mathcal{M}$ if $\mathcal{S} = \mathcal{S}(\mathcal{M})$. If, in addition, there is no set $X \in \mathcal{M}$ such that $X \in \langle \mathcal{M} \backslash \{X\} \rangle$, then call $\mathcal{M}$ a *basis* for $\mathcal{S}$.

Proposition 3.5. *For a UIS $\mathcal{S} = (P, \mathcal{B})$, let $\mathcal{M}(\mathcal{S})$ be the set of all minimal blocks of $\mathcal{S}$. Then $\mathcal{M}(\mathcal{S})$ is a basis for $\mathcal{S}$ if and only if $\mathcal{S}$ is completely irreducible.*

Proof. By definition, the set $\mathcal{M}(\mathcal{S})$ of minimal blocks satisfies the property that there is no set $X \in \mathcal{M}(\mathcal{S})$ such that $X \in \langle \mathcal{M}(\mathcal{S}) \backslash \{X\} \rangle$.

Assume, by way of contradiction, that $\mathcal{M} := \mathcal{M}(\mathcal{S})$ does not span $\mathcal{S}$. Let B be a block of $\mathcal{S}$ that is minimal among those such that $B \notin \langle \mathcal{M} \rangle$. Consider the set system $\mathcal{S}(B) = (B, \mathcal{B}(B))$ induced by the block B (see Definition 2). Since $\mathcal{M} \subseteq \langle \mathcal{M} \rangle$, the block $B \notin \mathcal{M}$. Therefore, there is a block in $\langle \mathcal{M} \rangle$ that is properly contained in B. Let A be a maximal block in $\langle \mathcal{M} \rangle$ that is properly contained in B, and consider the partition of B into A and $B \backslash A$. If C is a block in $\mathcal{M}$ properly contained in B, then $C \in \langle \mathcal{M} \rangle$; otherwise C would contradict the minimality of B. Also, either (1) $C \subseteq A$ or $C \subseteq B \backslash A$, or (2) $C \cap A \neq \emptyset$ and $C \cap (B \backslash A) \neq \emptyset$. In case (1) $(B, \mathcal{B}(B))$ is reducible, hence $\mathcal{S}$ is not completely irreducible, and we are done. In case (2) either the set $A \cup C \in \langle \mathcal{M} \rangle$ is properly contained in B, contradicting the maximality of A, or $A \cup C = B$, contradicting $B \notin \langle \mathcal{M} \rangle$.

Conversely, assume that $\mathcal{S}$ is not completely irreducible, that $\mathcal{S}$ has a block A such that $\mathcal{S}(A) = (A, \mathcal{B}(A))$ is reducible. Then $\mathcal{B}(A) \neq \emptyset$ and there is a partition (A_1, A_2) of A into two nonempty parts such that if $C \in \mathcal{B}(A) \cap \langle \mathcal{M} \rangle$, then $C \subseteq A_1$ or $C \subseteq A_2$. It is then impossible that $A \in \langle \mathcal{M} \rangle$, that is, $\mathcal{S}$ is not spanned by $\mathcal{M}$. □

The *intersection graph* $G(\mathcal{S})$ of a UIS $\mathcal{S} = (P, \mathcal{B})$ is the graph whose vertex set is $\mathcal{M}(\mathcal{S})$ and where adjacency is defined as follows. If $A \in \mathcal{M}(\mathcal{S})$ is a minimal block of $\mathcal{S}$, denote the corresponding vertex in $G(\mathcal{S})$ by $v(A)$. Two vertices $v(A)$ and $v(B)$ of $G(\mathcal{S})$ are adjacent if $A \cap B \neq \emptyset$.

If U is a set of vertices of $G(\mathcal{S})$, let U^* be the corresponding subset of $\mathcal{M}(\mathcal{S})$, and let

$$\overline{U} := \bigcup \{A : A \in U^*\} \subseteq P.$$

If $U = \{u\}$, a single vertex, then we simplify the notation to u^* and $\bar{u}$.

Lemma 3.6. *Let $\mathcal{S} = (P, \mathcal{B})$ be a completely irreducible UIS and $G(\mathcal{S})$ the corresponding intersection graph.*

(1) *If U is a connected vertex set of $G(\mathcal{S})$, then $\overline{U} \in \mathcal{B}$.*
(2) *If $A \in \mathcal{B}$ then there exists a connected vertex set U of $G(\mathcal{S})$ such that $\overline{U} = A$.*

Proof. Concerning statement (1), let $U \subseteq V$ be a connected vertex set of $G(\mathcal{S})$. The set U can be obtained recursively, starting from a single vertex, adding one adjacent vertex at each step, always leaving the subgraph connected, and terminating with all vertices of $G(\mathcal{S})$. Since each vertex in $G(\mathcal{S})$ corresponds to a minimal set in $\mathcal{M}$ and adjacent vertices indicate nonempty intersection, $\overline{U}$ can be obtained using Algorithm M. This shows that $\overline{U} \in \langle \mathcal{M}(\mathcal{S}) \rangle = \mathcal{B}$, the equality from Proposition 3.5.

Concerning statement (2), let $A \in \mathcal{B}$. We will prove by induction on $k := |A|$ that there exists a connected vertex set U such that $\overline{U} = A$. If $A \in \mathcal{M}(\mathcal{S})$, then there is a single vertex v in $G(\mathcal{S})$ such that $\bar{v} = A$ and a single vertex is connected. So assume that the statement (2) of the lemma is true for all blocks A' with $|A'| < k$, and assume that $|A| = k$. Since $A \in \mathcal{B}$, Proposition 3.5 implies that $A \in \langle \mathcal{M}(\mathcal{S}) \rangle$. Therefore A can be obtained by Algorithm M. Thus there exists blocks A_1, A_2 such that $A = A_1 \cup A_2$, $A_1 \cap A_2 \neq \emptyset$, $|A_1| < k$, and $|A_2| < k$. By the induction hypothesis, there are connected vertex sets U_1, U_2 of $G(\mathcal{S})$ such that (1)

$\overline{U_1} = A_1$ and $\overline{U_2} = A_2$, and (2) there exists vertices $u_1 \in U_1$ and $u_2 \in U_2$ such that $u_1^* \cap u_2^* \neq \emptyset$ (because $A_1 \cap A_2 \neq \emptyset$). Let $U = U_1 \cup U_2$. Now

$$\overline{U} = \overline{U_1 \cup U_2} = \overline{U_1} \cup \overline{U_2} = A_1 \cup A_2 = A.$$

By the definition of the intersection graph, since $u_1^* \cap u_2^* \neq \emptyset$ the vertices $u_1 \in U_1$ and $u_2 \in U_2$ are adjacent in $G(\mathcal{S})$. Since U_1 and U_2 are connected vertex sets, and u_1 and u_2 are adjacent, then also $U = U_1 \cup U_2$ is a connected vertex set. □

Corollary 3.7. *If $\mathcal{S} = (P, \mathcal{B})$ is a completely irreducible UIS, then $G(\mathcal{S})$ is a connected graph.*

Proof. Given a completely irreducible UIS $\mathcal{S} = (P, \mathcal{B})$, let $\mathcal{C}$ denote the set of all connected vertex sets of $G(\mathcal{S})$. By Lemma 3.6, there is a surjective map $p : \mathcal{C} \to \mathcal{B}$ defined by $U \mapsto \overline{U}$. By Lemma 3.4, the set P is a block of $\mathcal{S}$. By statement (2) of Lemma 3.6, there is a connected vertex set U such that $\overline{U} = P$. Therefore U is a connected spanning set (subgraph) of $G(\mathcal{S})$, that is, $G(\mathcal{S})$ is a connected graph. □

Proof of Theorem 2.12. We will prove statement (1); statement (2) then follows from statement (1) and Proposition 1.3. By Corollary 3.7, $G(\mathcal{S})$ is connected. By statement (1) of Theorem 1.4, there is a vertex v of $G(\mathcal{S})$ that is contained in at least half of the connected vertex sets of $G(\mathcal{S})$.

Let $p : \mathcal{C} \to \mathcal{B}$ be the surjective map used in the proof of Corollary 3.7. If p is also injective, then the correspondence $U \leftrightarrow \overline{U}$ is a bijection between the collection of connected vertex sets of $G(\mathcal{S})$ and the blocks of $\mathcal{S}$. In this case, $v^* \in \mathcal{M}$, where v is the vertex in the preceding paragraph, is contained in half of the blocks of $\mathcal{S}$, proving the theorem.

To see that p is injective, let $U_1, U_2 \in \mathcal{C}$ and assume that $p(U_1) = p(U_2)$. Then $\overline{U_1} = p(U_1) = p(U_2) = \overline{U_2}$. Therefore, by the definition of nonredundant $U_1^* = U_2^*$ from which $U_1 = U_2$ follows. □

ACKNOWLEDGMENTS

This work was partially supported by a grant from the Simons Foundation (322515 to Andrew Vince).

ORCID

Andrew Vince http://orcid.org/0000-0002-1022-1320

REFERENCES

1. M. A. Bahmanian and M. Sajna, *Connection and separation in hypergraphs*, Theory Appl. Graphs. **2** (2015), 1–24.
2. K. J. Balodis, L. Mol, and R. Oellermann, *On the mean order of connected induced subgraphs of block graphs*, Australasian J. Combin. **76** (2020), 128–148.
3. A. Bretto, *Hypergraph theory, an introduction*, Springer, New York, 2013.
4. H. Bruhn and O. Schaudt, *The journey of the union-closed sets conjecture*, Graphs. Combin. **31** (2015), 2043–2074.
5. A. J. Chin, G. Gordon, K. J. MacPhee, and C. Vincent, *Subtrees of graphs*, J. Graph. Theory. **89** (2018), 413–438.
6. J. Haslegrave, *The path minimises the average size of a connected set*, Discrete Math. **345** (2022), 7.

7. J. Haslegrave, *The number and average size of connected sets in graphs with degree constraints*, Graph. Theory. **100** (2022), 1–13.
8. R. Jamison, *On the average number of nodes in a subtree of a tree*, J. Combin. Theory Ser. B. **35** (1983), 207–223.
9. R. Jamison, *Monotonicity of the mean order of subtrees*, J. Combin. Theory Ser. B. **37** (1984), 70–78.
10. M. E. Kroeker, L. Mol, and O. Oellermann, *On the mean connected induced subgraph order of cographs*, Australasian J. Combin. **71** (2018), 161–183.
11. L. Mol and O. Oellermann, *Maximizing the mean subtree order*, J. Graph. Theory. **91** (2018), 326–352. https://doi.org/10.1002/jgt.22434
12. A. M. Stephens and O. Oellermann, *The mean order of sub-k-trees of k-trees*, J. Graph. Theory. **88** (2018), 61–79.
13. A. Vince and H. Wang, *The average order of a subtree of a tree*, J. Combin. Theory Ser. B. **100** (2010), 161–170.
14. A. Vince, *The average size of a connected vertex set of a graph—Explicit formulas and open problems*, J. Graph. Theory. **97** (2020), 82–103.
15. A. Vince, *A lower bound on the average size of a connected vertex set of a graph*, J. Combin. Theory Ser. B. **152** (2022), 153–170.
16. A. Vince, *The average size of a connected vertex set of a k-connected graph*, Discrete Math. **344** (2021), 4.
17. S. Wagner and H. Wang, *On the local and global means of subtree orders*, J. Graph. Theory. **81** (2016), 154–166.
18. W. Yan and Y. Yeh, *Enumeration of subtrees of trees*, Theoret. Comput. Sci. **369** (2006), 256–268.

How to cite this article: A. Vince, *The average order of a connected induced subgraph of a graph and union-intersection systems*, J. Graph Theory. (2023), 1–15. https://doi.org/10.1002/jgt.23024