THE AVERAGE SIZE OF NONSINGULAR SETS IN A GRAPH

Wilfried IMRICH[1], Norbert SAUER[2] and Wolfgang WOESS[3]

ABSTRACT. An nonsingular set in a finite graph is defined as the vertex set of an induced subgraph which has no isolated point. If G is a graph without isolated points and with at least two vertices and B is a connected subgraph, then the average size of those nonsingular sets in G which contain B is at least $(|G|+|B|)/2$. This result is used to prove the following: if $\mathcal{F}$ is a family of sets which is closed with respect to union, and if none of the generating sets in $\mathcal{F}$ has more than two elements, then the average size of a set in $\mathcal{F}$ is at least half of the size of the largest set in $\mathcal{F}$.

1980 Math. Subject Classification: 05C99, 04A20

Key words: Union closed family of sets, subset in a graph, isolated vertex.

1. INTRODUCTION AND STATEMENT OF RESULTS

Let G be a finite graph (unoriented, without loops or multiple edges). Subsets A of $V(G)$ are identified with induced subgraphs; $|A|$ denotes the number of vertices. We call a subset A of G *nonsingular*, if it has no isolated vertex. If B is a subgraph of G, then we define

$$\mathcal{F}(G,B) = \{A \subseteq G \mid B \subseteq A \text{ and } A \text{ is nonnsingular}\}\,.$$

In this note, we are interested in the average size of an element in $\mathcal{F}(G,B)$. To this end, we introduce a generating function:

$$f(t|G,B) = \sum_{A\in\mathcal{F}(G,B)} t^{|A|}\,.$$

Then the average size in question is given by $f'(1|G,B)/f(1|G,B)$, where $f'(t|G,B)$ denotes the derivative with respect to t.

Theorem 1. *If G is a nonsingular graph (with at least two vertices) and B is a connected subgraph (possibly void), then*

$$\frac{f'(1|G,B)}{f(1|G,B)} \geq \frac{|G|+|B|}{2}\,.$$

Observe that $\mathcal{F}(G,B)$ is a family of sets which is closed with respect to set union. Thus, our result can be related with the study of union closed families.

Let $\mathcal{F}$ denote an arbitrary finite family of finite sets which is closed with respect to set union. P. Frankl [1], [2] has conjectured the following.

[1] Institut für Mathematik und Angewandte Geometrie, Montanuniversität, A-8700 Leoben, Austria.

[2] Department of Mathematics, University of Calgary, Calgary, Alberta T2N 1N4, Canada.

[3] Dipartimento di Matematica, Università di Milano, Via C. Saldini, 50,
20133 Milano, Italy.

Conjecture. *There is an element which occurs in at least half of the sets of $\mathcal{F}$.*

In other words, if V is the largest set in $\mathcal{F}$ and, for x in V, $\mathcal{F}_x = \{A \in \mathcal{F} \mid x \in A\}$, then one is looking for an x such that $|\mathcal{F}_x| \geq |\mathcal{F}|/2$. In discussions of the Conjecture at the 1985 Banff meeting, R. Graham and several others observed the following.

Lemma 1. *If $\mathcal{F}$ contains a nonvoid set A with no more than two elements, then $|\mathcal{F}_x| \geq |\mathcal{F}|/2$ for some $x \in A$.*

As no proof of this observation seems to have appeared in print, we include a proof in this paper. At the same meeting in Banff, R. Graham claimed to have an example of a union closed family $\mathcal{F}$ with the following properties: there is an $A \in \mathcal{F}$ with three elements, such that $|\mathcal{F}_x| < |\mathcal{F}|/2$ for every $x \in A$, but the Conjecture is true for this $\mathcal{F}$.

Consider the following statement.

$$\textit{The average size of a set in } \mathcal{F} \textit{ is at least half of the size of the largest set in } \mathcal{F}. \tag{1}$$

It is easy to see that statement (1) implies the validity of the Conjecture for $\mathcal{F}$: assume that

$$\sum_{A \in \mathcal{F}} |A| \big/ |\mathcal{F}| \geq |V|/2\,,$$

Suppose that $|\mathcal{F}|/2 > |\mathcal{F}_x|$ for every x in V. Then

$$|V| \cdot |\mathcal{F}|/2 > \sum_{x \in V} |\mathcal{F}_x| = \sum_{A \in \mathcal{F}} |A|\,,$$

a contradiction.

Thus, in particular, the Conjecture is true for $\mathcal{F}(G, B)$ as in Theorem 1. (However, we remark that Lemma 1 is needed in the proof of Theorem 1.) A family of *generators* $\mathcal{E}$ of $\mathcal{F}$ is a subfamily of $\mathcal{F}$ with the following properties:

(i) every set in $\mathcal{F}$ is a finite union of sets in $\mathcal{E}$,

(ii) $\mathcal{E}$ is minimal with respect to (i).

It is easy to see that $\mathcal{E}$ is unique and no set in $\mathcal{E}$ is empty. Theorem 1 can be applied to the case when no generator has more than two points.

Theorem 2. *If $|A| \leq 2$ for every generator A of $\mathcal{F}$, then statment (1) is true.*

In general it is not true that the average size of a set in a finite union closed family $\mathcal{F}$ is at least half of the size of the largest set in $\mathcal{F}$. Actually, the proportion between the average size and the largest size of a set in $\mathcal{F}$ may be arbitrarily small. Here is an example: let F be a set with $|F| = 2^n$ and let $E \subset F$ with $|E| = 2n$, and define

$$\mathcal{F} = \{A \subset F \bigm| |A| \geq 2^n - 1 \text{ or } A \subseteq E\}.$$

In this case, the average size of a set in $\mathcal{F}$ is asymptotically equal to $n + 1$.

In a preliminary version of this note, we raised the question whether statement (1) was true for every union closed family $\mathcal{F}$ where all generators have the same size. This has been disproved recently by P. Wójcik [3].

2. PROOFS

We start with Lemma 1, as discussed at Banff.

Proof of Lemma 1. If $A = \{x\}$ for some $x \in V$, where V is the largest set in $\mathcal{F}$, then $B \mapsto B \cup \{x\}$ is injective from $\mathcal{F} \setminus \mathcal{F}_x$ into $\mathcal{F}_x$, so $|\mathcal{F}_x| \geq |\mathcal{F}|/2$.

Now assume that $A = \{x, y\}$ with $|\mathcal{F}_x| \geq |\mathcal{F}_y|$. Let $a = |\{B \in \mathcal{F} \mid x \in B, y \notin B\}|$, $b = |\{B \in \mathcal{F} \mid y \in B, x \notin B\}|$, $c = |\{B \in \mathcal{F} \mid A \subseteq B\}|$ and $d = |\{B \in \mathcal{F} \mid A \cap B = \emptyset\}|$. Then $|\mathcal{F}| = a + b + c + d$, $|\mathcal{F}_x| = a + c$ and $|\mathcal{F}_y| = b + c$. Hence $a \geq b$, and (as in the first case) $c \geq d$, so again $|\mathcal{F}_x| \geq |\mathcal{F}|/2$. ∎

Now we turn to the study of $\mathcal{F}(G, B)$, where G is a graph. Note that

$$f(t|G, \emptyset) = f(t|G, \{x\}) + f(t|G \setminus \{x\}, \emptyset) . \tag{2}$$

In particular, if G is a graph without isolated vertices, then Lemma 1 implies that for some x in G,

$$f(1|G, \{x\}) \geq f(1|G \setminus \{x\}, \emptyset) . \tag{3}$$

Proof of Theorem 1. We first assume that G is connected.

We use induction on $n = |G|$ and, for each n, induction on $k = |G| - |B|$.

For $n = 2$ the verification is trivial.

Assume that $|G| = n$ and that the statement of the theorem is true for every graph $|G'|$ with $|G'| = m$, $2 \leq m < n$. First, let $k = 0$. Then $B = G$, and $f(t|G, B) = t^{|G|}$. We have

$$f'(1|G, B) = |G| = \frac{|G| + |B|}{2} f(1|G, B) ,$$

so the result follows. Now let $k = |G| - |B|$ be positive and assume also that the statement is true for all B' such that $0 \leq |G| - |B'| < k$.

Case 1. B is nonvoid. As G and B are connected and $G \setminus B \neq \emptyset$, there is an x in $G \setminus B$ such that

$$B' = B \cup \{x\}$$

is connected.

Case 2. $B = \emptyset$. By (3), there is a vertex x such that $f(1|G, \{x\}) \geq f(1|G \setminus \{x\}, \emptyset)$. Set

$$B' = \{x\} .$$

In both cases,

$$f(t|G, B) = f(t|G \setminus \{x\}, B) + f(t|G, B') . \tag{4}$$

We intend to differentiate and to obtain inequalities for the right-hand side.

The graph $G \setminus \{x\}$ may be disconnected, and it may have isolated points. Let $C = C(x, B)$ be the set of those isolated points of $G \setminus \{x\}$ which are not in B. Then, for every y in C, the only neighbour of y in G is x. In particular,

$$G' = G \setminus C$$

is connected. No point of C can be an element of a nonsingular set of $G \setminus \{x\}$. Hence

$$f(t|G \setminus \{x\}, B) = f(t|G' \setminus \{x\}, B) . \tag{5}$$

Now write $G' \setminus \{x\}$ as the union of its connected components: $G' \setminus \{x\} = G_1 \cup G_2 \cup ... \cup G_s$, listed so that $B \subseteq G_1$ and $|G_i| \geq 2$ for $i = 2, ..., s$. A set A is in $\mathcal{F}(G' \setminus \{x\}, B)$ if and only if $A = A_1 \cup A_2 \cup ... \cup A_s$, where $A_1 \in \mathcal{F}(G_1, B)$ and $A_i \in \mathcal{F}(G_i, \emptyset)$ for $i = 2, ..., s$. This and (5) yield

$$f(t|G \setminus \{x\}, B) = f(t|G_1, B) \cdot f(t|G_2, \emptyset) \cdot ... \cdot f(t|G_s, \emptyset) .$$

Note that we may have $|G_1| = 1$ only if $|B| = 1$. In this case, $f(t|G_1, B) = f'(t|G_1, B) = 0$. Otherwise, we may apply the induction hypothesis (for n) to (G_1, B). In both cases

$$f'(t|G_1, B) \geq \frac{|G_1| + |B|}{2} f(t|G_1, B) .$$

The induction hypothesis also applies to $(G_i, \emptyset)$, giving

$$f'(t|G_i, \emptyset) \geq \frac{|G_i|}{2} f(t|G_i, \emptyset) \,, \; i = 2, ..., s .$$

This, (5) and the product rule for differentiation give

$$f'(1|G \setminus \{x\}, B) \geq \frac{|G| - |C| - 1 + |B|}{2} f(1|G \setminus \{x\}, B) . \tag{6}$$

On the other hand, B' is a connected subgraph of G'. For any $U \subseteq C$, considering those nonsingular sets in $\mathcal{F}(G, B')$ which contain U and do not contain $C \setminus U$, we get

$$f(t|G, B') = \sum_{U \subseteq C} f(t|G' \cup U, B' \cup U) .$$

Differentiating and applying the induction hypothesis (for k, if $U = C$, and for n otherwise) to each of the summands, we obtain

$$f'(1|G, B') \geq \sum_{U \subseteq C} \frac{|G| - |C| + |B| + 1 + 2|U|}{2} f(1|G' \cup U, B' \cup U) . \tag{7}$$

Combining inequalities (6) and (7), we obtain from (4)

$$f'(1|G, B) \geq \frac{|G| + |B|}{2} f(1|G, B) + \frac{\Phi}{2} \,, \tag{8}$$

where

$$\Phi = \sum_{U\subseteq C} \big(|U| - |C\setminus U| + 1\big) f(1|G'\cup U, B'\cup U) \ - \ \big(|C|+1\big) f(1|G\setminus\{x\}, B)\,.$$

The proof will be complete if we can show that $\Phi \geq 0$. We observe that if $U \subseteq C$ is nonvoid, then $A \in \mathcal{F}(G'\cup U, B'\cup U)$ if and only if $A\cup(C\setminus U) \in \mathcal{F}(G, B'\cup C)$. Hence, $f(1|G'\cup U, B'\cup U) = f(1|G, B'\cup C)$ whenever $\emptyset \neq U \subseteq C$. (If B is nonvoid, this also holds if $U = \emptyset$.) Consequently,

$$\begin{aligned}\Phi &= \Big(\sum_{U\subseteq C} |U| - |C\setminus U| + 1\Big) f(1|G, B'\cup C)\\ &\quad + \big(|C|-1\big)\Big(f(1|G, B'\cup C) - f(1|G', B')\Big) \ - \ \big(|C|+1\big) f(1|G\setminus\{x\}, B)\\ &= 2^{|C|} f(1|G, B'\cup C) \ - \ \big(|C|+1\big) f(1|G\setminus\{x\}, B)\\ &\quad + \big(|C|-1\big)\Big(f(1|G, B'\cup C) - f(1|G', B')\Big)\,.\end{aligned}$$

If $A \in \mathcal{F}(G', B')$ then $x \in A$, and $A\cup C$ is nonsingular: $A \in \mathcal{F}(G, B'\cup C)$. Thus, $f(1|G, B'\cup C) \geq f(1|G', B')$, and

$$\Phi \geq \big(|C|+1\big)\Big(f(1|G, B'\cup C) - f(1|G\setminus\{x\}, B)\Big)\,,$$

since $2^{|C|} \geq |C|+1$. There are two cases to consider.

Case 1. B is nonvoid. If $A \in \mathcal{F}(G\setminus\{x\}, B) = \mathcal{F}(G'\setminus\{x\}, B)$, then $A\cup\{x\}$ is nonsingular, as x has a neighbour in $B \subseteq A$. In other words, $A\cup\{x\}\cup C$ is nonsingular (even if $C = \emptyset$), and is an element of $\mathcal{F}(G, B'\cup C)$. Thus, $f(1|G\setminus\{x\}, B) \leq f(1|G, B'\cup C)$, so $\Phi \geq 0$.

Case 2. $B = \emptyset$. We have to consider two subcases.

a) C is nonvoid. In this case, $\{x\}\cup C$ is nonsingular, and if $A \in \mathcal{F}(G\setminus\{x\}, \emptyset) = \mathcal{F}(G'\setminus\{x\}, \emptyset)$ then $A\cup\{x\}\cup C = A\cup B'\cup C \in \mathcal{F}(G, B'\cup C)$. Again, $f(1|G\setminus\{x\}, B) \leq f(1|G, B'\cup C)$, and $\Phi \geq 0$.

b) $C = \emptyset$. Then $f(1|G, B'\cup C) - f(1|G\setminus\{x\}, B) = f(1|G, \{x\}) - f(1|G\setminus\{x\}, \emptyset)$, which is nonnegative by the choice of x, and $\Phi \geq 0$.

In all three cases, $\Phi \geq 0$, and by (8), proof is completed (for connected G).

Now suppose in general that G is nonsingular. Then we can use the formula

$$f(t|G, B) = f(t|G_1, B)\cdot f(t|G_2, \emptyset)\cdot \ldots \cdot f(t|G_s, \emptyset)\,,$$

where $G_1, \ldots, G_s$ are the connected components of G and $B \subseteq G_1$. Differentiation and application of the result for connected graphs yield the theorem. ∎

Now there may be several ways to derive Theorem 2. Our proof uses the full strength of Theorem 1.

Proof of Theorem 2. We may assume that $\emptyset \in \mathcal{F}$ (adding the empty set to $\mathcal{F}$ makes the average size smaller). If all generators of $\mathcal{F}$ are singletons, then Theorem 1 is trivially true. Hence, we assume that $\mathcal{E}$ contains two-element sets and maybe also singletons. Observe that there may be points x, y such that $\{x\}$ and $\{x, y\}$ are generators. In this case, $\{y\}$ cannot be a generator.
Let V be the largest set in $\mathcal{F}$, and set $B = \big\{x \mid \{x\} \in \mathcal{E}\big\}$. Choose a new set $\bar{B} = \{\bar{x} \mid x \in B\}$ in one-to-one correspondence with B, disjoint from V. We define a graph G with vertex set $V \cup \bar{B}$; the edges are (i) all $\{x, y\} \in \mathcal{E}$, (ii) all $\{x, \bar{x}\}$, where $x \in B$, and (iii) all $\{\bar{x}, \bar{y}\}$, where $\bar{x}, \bar{y} \in \bar{B}$, $\bar{x} \neq \bar{y}$.

Thus, $\bar{B}$ induces a complete subgraph of G which is connected to the rest of G by the edges $\{x, \bar{x}\}$, $x \in B$. (If B is empty, then so is $\bar{B}$.) Now, G has no isolated point, and we may apply Theorem 2 and Remark A to $(G, \bar{B})$:

$$f'(1|G, \bar{B}) \geq \frac{|G| + |B|}{2} f(1|G, \bar{B})\,.$$

By our construction, we obviously have $A \in \mathcal{F}$ if and only if $A \cup \bar{B} \in \mathcal{F}(G, \bar{B})$. Hence, $f(1|G, \bar{B}) = |\mathcal{F}|$, and

$$f'(1|G, \bar{B}) = \sum_{A \in \mathcal{F}} |A \cup \bar{B}| = |B| \cdot |\mathcal{F}| + \sum_{A \in \mathcal{F}} |A|\,.$$

We obtain

$$\sum_{A \in \mathcal{F}} |A| \geq \frac{|G| + |B|}{2} |\mathcal{F}| - |B| \cdot |\mathcal{F}| = \frac{|V|}{2} |\mathcal{F}|\,. \qquad \blacksquare$$

REFERENCES

[1] I. RIVAL, Ed.: Graphs and Order. Problem Session 1, p 525. Reidel, Dordrecht - Boston, 1985.

[2] R.P. STANLEY: Enumerative Combinatorics, Vol. I. Exercise 39a, p 161. Woodsworth & Brooks / Cole Math. Series, Monterey, Calif., 1986.

[3] P. WÓJCIK: Density of union-closed families. Preprint, Adam Mickiewicz University, Poznań, Poland.