arXiv:1812.02422v1 [math.CO] 6 Dec 2018

GRAPHS AND UNICYCLIC GRAPHS WITH EXTREMAL CONNECTED SUBGRAPHS

AUDACE A. V. DOSSOU-OLORY

Abstract. Over all graphs or unicyclic graphs of a given order, we characterise all graphs (or unicyclic graphs) that minimise or maximise the number of connected subgraphs or connected induced subgraphs. For each of these classes, we find that the minimal graphs for the number of connected induced subgraphs coincide with those that are known to maximise the Wiener index (the sum of the distances between all unordered pairs of vertices) and vice versa. For every k, we also determine the connected graphs that are extremal with respect to the number of k-vertex connected induced subgraphs. We show that, in contrast to the minimum which is uniquely realised by the path, the maximum value is attained by a rich class of connected graphs.

1. Introduction and first results

Counting and understanding graph structures with particular properties has many applications, especially to network theory, computer science, biology and chemistry. For instance, graphs can represent biological networks at the molecular or species level (protein interactions, gene regulation, etc.) [18]. The topological structure of an interconnection network is a *connected* graph where, for example, vertices are processors and edges represent links between them [3]. In chemical networks, vertices are atoms and edges represent their bonds. An important question is to find all matches of a specific motif within a larger network (the *subgraph* isomorphism problem [19], or the *induced subgraph* isomorphism problem [23]). Both cases are known to be in general NP-complete [21] although in some instances (such as planar graphs), efficient algorithms are available [22]. A step to these problems usually consists of enumerating all possible subgraphs or induced subgraphs of the network [20]. This paper discusses the total number of subgraphs or induced subgraphs of a finite and simple (no loops, no parallel edges, undirected) graph with a particular emphasis on induced subgraphs of a connected graph.

Let G be a simple graph consisting of a finite (but not empty) set $V(G)$ of vertices and a finite set $E(G)$ of edges. A graph H such that $\emptyset \neq V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ is called a subgraph of G (so we do not consider the empty graph!). A graph formed from G by taking a nonempty subset of vertices of G and all edges incident with them is called

2010 *Mathematics Subject Classification.* Primary 05C30; secondary 05C35, 05C05.

Key words and phrases. induced subgraphs, connected graphs, unicyclic graphs, tadpole graphs, extremal graph structures, path, complete graph.

The author was supported by Stellenbosch University in association with the African Institute for Mathematical Sciences (AIMS) South Africa.

an induced subgraph of G. In this paper, we are concerned with the extremal problem of determining the minimum and maximum number of subgraphs or induced subgraphs of G, and also characterising the extremal graphs. This problem will be considered from certain types of graphs all sharing the same number of vertices.

It is trivial that every graph of order n (the number of vertices) has precisely $2^n - 1$ induced subgraphs since every graph has only one k-vertex induced subgraph for every choice of k between 1 and n. A graph with no edge is called *edgless*, while a graph with an edge between any two distinct vertices is called *complete*. It is clear that adding an edge e to a graph G creates at least one new subgraph (namely the subgraph e). This implies that among all graphs having n vertices, precisely the edgeless graph E_n has the minimum number of subgraphs, and the maximum is uniquely attained by the complete graph K_n. In fact, all subgraphs of E_n are induced subgraphs, namely the graphs E_k, each of them counted precisely $\binom{n}{k}$ times; so E_n also has the minimum number of k-vertex subgraphs (for $k > 1$, E_n is the only minimal graph in this case). The complete graph K_n has $\binom{n}{2}$ edges and thus $2^{\binom{k}{2}}$ subgraphs of order k obtained by destroying between 0 and $\binom{k}{2}$ edges of K_k as all induced subgraphs of K_n are again complete subgraphs. In particular, K_n is the only graph having the maximum number of $k(> 1)$-vertex subgraphs.

Distinct vertices $u, v \in V(G)$ are said to be connected in G if there is a path from u to v in G. The graph G is connected if and only if any two distinct vertices of G are connected in G. The edgless graph and the complete graph are the only extremal graphs for the total number of connected induced subgraphs:

Proposition 1. *Every graph of order n has at least n connected induced subgraphs (with equality for the edgless graph E_n only) and at most $2^n - 1$ connected induced subgraphs (with equality for the complete graph K_n only).*

Proof. Every single vertex of a graph G is a connected induced subgraph of G. Clearly, the only connected induced subgraphs of the edgless graph are its vertices. If G is a graph of order n which is not E_n, then G has at least one edge e; so e is a connected induced subgraph of G. This proves the case of minimum.

For the maximum, it is clear that every induced subgraph of the complete graph K_n is connected; so K_n has $2^n - 1$ connected induced subgraphs. If G is a graph of order n which is not K_n, then G has at least two nonadjacent vertices u, v: the subgraph induced by $\{u, v\}$ is not connected, which proves the case of maximum. □

Note that the notions of subgraph and induced subgraph coincide for the edgeless graph only. Adding an edge e in a graph increases the number of connected subgraphs by at least one (namely, the graph e). Thus, the complete graph remains the only graph having the maximum number of connected subgraphs. The problem becomes more interesting when one considers the total number of connected subgraphs of a connected graph. In fact, the various graphs used as models in different applications are connected. Tittmann et al. [13] enumerated the number of connected components in induced subgraphs by means of a generating function approach. Yan and Yeh [14] gave a linear-time algorithm for counting the sum of weights of subtrees of a tree (a connected acyclic graph). In their paper [14], they also asked for methods to enumerate connected subgraphs of a connected graph. Very recently, Kroeker et al. [7] investigated the extremal structures for the mean order of connected induced subgraphs among so-called cographs (graphs containing no induced path of order 4). Our main interest in this paper is to know the minimum and maximum number of connected subgraphs that a connected graph of a given order can contain; the approach we use does not involve generating functions.

The complete graph remains maximal among all connected graphs of a given order, while the edgeless graph is no longer minimal.

Corollary 2. *A connected graph G of order n has at most $2^n - 1$ connected induced subgraphs with equality if and only if $G = K_n$.*

Note that if $e \in E(G)$, then $G - e$ is the graph obtained from G by removing edge e in G (but leaving the two vertices incident with e in G). We note that the notions of connected subgraph and induced connected subgraph coincide for trees (connected acyclic graphs) only. This is because every connected graph of order n has at least $n - 1$ edges with equality if and only if the graph is a tree. In the next theorem, we show that the n-vertex path P_n has very few connected induced subgraphs, and this is in fact the only minimal graph among all connected graphs of order n.

Theorem 3. *The path P_n which has $\binom{n+1}{2}$ connected induced subgraphs, is the only minimal graph among all connected graphs of order n.*

Proof. Let G be a connected graph of order n which contains a cycle C and e be an edge of C. Clearly, every connected induced subgraph of $G - e$ is again a connected induced subgraph of G. However, the converse is not true: the two vertices that are incident with e in G induce a disconnected graph in $G - e$. This immediately implies that a minimal graph must be a tree. Futhermore, it is well known [11] that a tree T of order n has at least $\binom{n+1}{2}$ connected induced subgraphs with equality if and only if T is the path P_n. □

For a connected graph G, we shall denote by $N_k(G)$ the total number of k-vertex connected induced subgraphs of G. By deleting a vertex $u \in V(G)$, we mean removing u and all edges incident with u in G. So the subgraph induced by a (nonempty) set $W \subseteq V(G)$ is obtained by deleting in G all vertices that do not belong to W. From this point onwards, G is always a connected graph. A very basic observation is that $N_1(G) = |V(G)|$, $N_2(G) = |E(G)|$ and $N_{|V(G)|}(G) = 1$. It is important to note that all induced subgraphs (not necessarily connected) of order $n-1$ are easily established by deleting one vertex from G, giving $N_{n-1}(G) \leq n$. A vertex of degree 1 in G is called a *pendent* vertex of G. Thus, the subgraph obtained by deleting a pendent vertex of a connected G is always connected. We shall see that $N_{n-1}(G)$ is at least the sum of the number of pendent vertices of G and the number of vertices of G whose all neighbors are contained in the same cycle of G (a precise interpretation of $N_{n-1}(G)$ is given in Proposition 7). We denote by $N(G)$ the total number of connected induced subgraphs of G.

If $v_0, v_1, \ldots, v_{k-1}$ are vertices of $V(G)$, then we write $G - \{v_0, v_1, \ldots, v_{k-1}\}$ to mean the induced subgraph obtained from G by deleting vertices $v_0, v_1, \ldots, v_{k-1}$ in G. The cycle of length $k \geq 3$ will be denoted by $C_k = (v_0, v_1, \ldots, v_{k-1}, v_k = v_0)$, where v_j is adjacent to v_{j+1} for every $j \in \{0, 1, \ldots, k-1\}$.

The rest of the paper is organised as follows. The next section (Section 2) carries a study of the connected graphs that are extremal with respect to the total number of k-vertex connected induced subgraphs $N_k(G)$. We first introduce two lemmas and use them thereafter to show that the path (resp. complete graph) provides the minimum (resp. maximum) number of k-vertex connected induced subgraphs. In Section 3, we restrict the study to unicyclic graphs (connected graphs having exactly one cycle) and investigate the problem of finding the extremal unicyclic graphs, given the order. It will be shown, after discussing a series of auxiliary results, that the so-called tadpole graph (obtained by merging a vertex of a cycle to a pendent vertex of a path) is minimal, while the connected graph obtained by adding one edge between two pendent vertices of the star is maximal.

It occurs very often that a certain tree is extremal with respect to several graph invariants (the number of subtrees and the Wiener index, for instance) within a given class of trees. This also holds in our current context: the unicyclic graphs that are found to be extremal for the number of connected induced subgraphs were previously shown to be extremal for the Wiener index and the energy (among others).

From now on, the term 'subgraph' always means 'induced subgraph', unless otherwise specified.

2. The extremal graphs for $N_k(G)$

The focus in this section is on the graph parameter $N_k(G)$, the total number of connected subgraphs of order k of a connected graph G of order n. Sharp upper and lower bounds are determined for $N_k(G)$ in terms of order. In order to prove our results, we shall require the following two lemmas.

Lemma 4. *Let G be a connected graph of order at least 2. Then G contains a vertex v^* whose removal yields a connected subgraph of G.*

Proof. If G has a pendent vertex v, then it is clear that $G - \{v\}$ remains connected (so $v^* = v$). Assume then that G does not have a pendent vertex. It follows that G contains a cycle as it is well known (and easy to prove) that every tree of order at least 2 has two or more leaves (pendent vertices). If G contains a cycle C_k and a vertex $u \in V(C_k)$ such that all neighbors of u in G belong to $V(C_k)$, then deleting u in G immediately yields a connected subgraph of G (so $v^* = u$).

We are left with the situation in which G does not satisfy any of the above two assumptions. In this case, let $C_{k_1} = (v_0, v_1, \ldots, v_{k_1-1}, v_{k_1} = v_0)$ be a cycle in G. Since every vertex $v_i \in V(C_{k_1})$ has at least one neighbor $w_i \notin V(C_{k_1})$, we deduce that $|V(G)| \geq k_1 + 1$. Moreover, by assumption none of the w_i is a pendent vertex of G or belongs to a cycle C_{k_2} ($V(C_{k_2}) \neq V(C_{k_1})$) of G such that all neighbors of w_i in G lie in $V(C_{k_2})$ (otherwise, we are done immediately!). Therefore, $w_i \in V(C_{k_2})$ for some $V(C_{k_2}) \neq V(C_{k_1})$ and also has at least one neighbor $x_i \notin V(C_{k_2}) \cup V(C_{k_1})$: this implies that $|V(G)| \geq k_1+2$. By assumption, none of the x_i is a pendent vertex of G or belongs to a cycle C_{k_3} $(V(C_{k_3}) \neq V(C_{k_1}), V(C_{k_2}))$ of G such that all neighbors of x_i in G lie in $V(C_{k_3})$.

Since G is a finite graph (i.e., $|V(G)| < \infty$), this search process can not be repeated indefinitely. Hence, the legitimate existence of vertex v^*, which completes the proof of the lemma. □

The proof of Lemma 4 immediately implies that v^* can always be chosen to be either a pendent vertex of G or to have all its neighbors belonging to the same cycle of G. Hence, $N_{n-1}(G)$ is at least the sum of the number of pendent vertices of G and the number of vertices of G whose all neighbors are contained in the same cycle of G. The next lemma shows that G has a connected subgraph of every order less than or equal to $|V(G)|$.

Lemma 5. *Let G be a connected graph of order n. Then G has a connected subgraph of every order k between 1 and n.*

Proof. Let T be a spanning tree of G. By repeatedly removing leaves from T, we obtain subtrees (of T) of every order k between 1 and n. Now add to each of these subtrees all the missing edges between their vertices in G. This completes the proof of the lemma. □

It is important to note from the proof of Lemma 5 that G has, in particular, a connected subgraph of every order $k \leq |V(G)|$ that contains a given vertex v of G. This is easily seen by considering the rooted version of a spanning tree T of G (T is rooted at v). We combine Lemmas 4 and 5 to prove the following theorem.

Theorem 6. *Every connected graph of order n has at least $n-k+1$ connected subgraphs of order k, with equality holding (in the case $2 < k < n$) only for the path P_n.*

Proof. The cases $k \in \{1,2\}$ are essentially trivial since $N_1(G) = n$ and it is well known (and easy to prove) that every connected graph of order n has at least $n-1$ edges ($N_2(G) \geq n-1$) with equality if and only if G is a tree. Assume $k > 2$ and let us prove the statement of the theorem by induction on n. The case $n = k$ is trivial ($N_n(G) = 1$).

Consider a connected graph G of order $n > k$. By Lemma 4, let v be a vertex of G whose removal in G yields the connected subgraph $G - \{v\}$. The number of k-vertex connected subgraphs of G that do not involve v is therefore $N_k(G - \{v\})$ and by the induction hypothesis, we have $N_k(G - \{v\}) \geq n - k$ (the order of $G - \{v\}$ is $n - 1$) with equality if and only if $G - \{v\} = P_{n-1}$. On the other hand, G also has at least one connected subgraph of order k that contains vertex v (see the proof of Lemma 5). It follows that $N_k(G) \geq n - k + 1$, and equality holds if and only if $G - \{v\} = P_{n-1}$ and G has precisely only one connected subgraph of order k that involves v.

It remains to show that G is indeed a path in this case. To this end, consider a vertex w adjacent to v. So w lies on the path P_{n-1} since $G - \{v\} = P_{n-1}$. Let $w_1, w_2, \ldots, w_{n-1}$ be all vertices of P_{n-1} such that w_i is adjacent to w_{i+1} for every $i \in \{1, 2, \ldots, n-2\}$. We have $w = w_j$ for some $j \in \{1, 2, \ldots, n-1\}$.

- If $2 \leq j \leq n-k+1$, then G has at least two distinct v-containing subgraphs of order k, namely the subgraphs induced by
$$\{v, w_j, w_{j+1}, \ldots, w_{j+k-2}\} \quad \text{and} \quad \{v, w_{j-1}, w_j, w_{j+1}, \ldots, w_{j+k-3}\}$$
for instance.
- If $n-k+2 \leq j \leq n-2$, then G also has at least two distinct v-containing subgraphs of order k, namely the subgraphs induced by
$$\{w_{n-k+1}, w_{n-k+2}, \ldots, w_{j-1}, v, w_j, w_{j+1}, w_{j+2}, \ldots, w_{n-1}\}$$
and
$$\{w_{n-k}, w_{n-k+1}, \ldots, w_{j-1}, v, w_j, w_{j+1}, w_{j+2}, \ldots, w_{n-2}\}$$
for instance.

Thus, we must have $j \in \{1, n-1\}$ if exactly one v-containing subgraph of order k is to be obtained. Now we claim that in either situation $j = 1$ or $j = n-1$, vertex v must be a pendent vertex. We can assume (without loss of generality) that $j = 1$.

Suppose (for contradiction) that v has at least two neighbors w_1 and w_l for some $l > 1$.

- If $l \leq n-k+1$, then G has at least two distinct v-containing subgraphs of order k, namely the subgraphs induced by
$$\{v, w_1, w_2, \ldots, w_{k-1}\} \quad \text{and} \quad \{v, w_l, w_{l+1}, \ldots, w_{l+k-2}\}.$$
- If $l \geq n-k+2$, then G also has at least two distinct v-containing subgraphs of order k, namely the subgraphs induced by
$$\{v, w_1, w_2, \ldots, w_{k-1}\} \quad \text{and} \quad \{w_{n-k+1}, w_{n-k+2}, \ldots, w_{l-1}, v, w_l, w_{l+1}, w_{l+2}, \ldots, w_{n-1}\}.$$

This is a contradiction: hence $l = 1$, proving that G must be a path if exactly one v-containing subgraph of order k is to be obtained (which is indeed the case). This completes the proof of the theorem. □

Proposition 7. *Let G be a connected graph of order n. Then $N_{n-1}(G)$ is precisely the number of vertices of G that are leaves of a spanning tree of G.*

Proof. By Lemma 4, let v be a vertex of G such that $G - \{v\}$ is connected. Let T be a spanning tree of $G - \{v\}$ and w a neighbor of v in G. Let T^+ be the tree obtained from T and v by adding an edge between v and w. Then T^+ is a spanning tree of G and v is a leaf of T^+. □

The maximum (analogue of Theorem 6) can be attained for a rich class of connected graphs. We recall that $N_1(G) = |V(G)|$, $N_2(G) = |E(G)|$, $N_{|V(G)|}(G) = 1$ and $N_{|V(G)|-1}(G) \leq |V(G)|$ with equality holding for the cycle and the complete graph, for instance.

Denote by $\mathcal{G}_n^l$ the set of all inequivalent graphs that result from removing exactly $l \geq 1$ independent edges (edges sharing no common vertex) in the complete graph K_n ($n \geq 3$). It is not difficult to see that every graph in $\mathcal{G}_n^l$ is of order n and connected. In general, we have the following:

Proposition 8. *For every connected graph G of a fixed order $n \geq 3$ and every $k \in \{3, 4, \ldots, n\}$, the number of k-vertex connected subgraphs of G is at most $\binom{n}{k}$. Equality holds for all graphs $G \in \mathcal{G}_n^l$ for every l.*

Proof. For a connected graph G of order n, we have $N_k(G) \leq \binom{n}{k} = N_k(K_n)$ for every k. Equality holds if and only if every subset of k vertices of G induces a connected subgraph. Therefore, we have $N_k(G) = \binom{n}{k}$ for every $G \in \mathcal{G}_n^l$ and every $k \geq 3$. □

For a vertex u of G, we denote by $\mathcal{N}(u)$ the set of all neighbors of u in G, and $|\mathcal{N}(u)|$ its size (the degree of u in G).

3. The extremal unicyclic graphs

This section is concerned with a particularly well-studied class of tree-like structure as a sole subject: we consider (connected) unicyclic graphs of a given order and investigate which unicyclic graphs minimise or maximise the total number of connected subgraphs. A *unicyclic* graph is a connected graph which contains exactly one cycle. The number of unicyclic graphs of a fixed order $n > 2$ begins

$$1, 2, 5, 13, 33, 89, 240, 657, 1806, 5026, 13999, 39260, 110381, 311465, 880840, \ldots ;$$

see the sequence A001429 in [17] for more information. It is clear that the complete graph is no longer extremal among unicyclic graphs of a given order $n > 3$ ($K_3 = C_3$ is the only unicyclic graph of order 3).

A number of different graph invariants were studied in various subclasses of unicyclic graphs. This includes the sum of the absolute values of the eigenvalues (also known as the energy of a graph) and two closely related parameters, namely the number of independent sets (Merrifield-Simmons index) and the number of matchings (Hosoya index).

Hou [5] determined the unicyclic graphs with minimal energy, given the order. Li and Zhou [8] found the graphs with minimal energy among all unicyclic graphs in terms of order and diameter. Hou, Gutman and Woo [6] characterised the unicyclic bipartite graphs (that are not cycle) with maximal energy, given the order. Andriantiana [1] determined all unicyclic bipartite graphs with maximal energy in terms of order. Andriantiana and Wagner [2] found the non-bipartite unicyclic graphs with the largest energy.

Pedersen and Vestergaard [10] determined sharp upper and lower bounds for the Merrifield-Simmons index in a unicyclic graph in terms of order. They also found the maximal unicyclic graphs for the Merrifield-Simmons index in terms of order and girth. Ou [9] characterised both the unicyclic graphs that have the largest and the second-largest Hosoya index, given the order. Zhu and Chen [16] determined the maximal unicyclic graphs for the Merrifield-Simmons index, given girth and number of pendent vertices.

Among unicyclic graphs of a given order (and potentially other structural restrictions), the largest and second-largest energies are usually attained by cycles and so-called tadpole graphs (obtained by merging a vertex of a cycle to a pendent vertex of a path). Among all unicyclic graphs of order $n \geq 6$, the minimum energy is attained by the graph that results from connecting two leaves of a star by an edge. These extremal graphs will also play an important role in our current context of determining the number of connected subgraphs of a unicyclic graph, given the order. The following lemma will aid in proving our next results.

Lemma 9. *The cycle C_n has n^2-n+1 connected subgraphs.*

Proof. Let $C_n = (v_0, v_1, \dots, v_{n-1}, v_n = v_0)$ be the cycle of order n. Then a subset S of k elements of $V(C_n)$ induces a connected subgraph if and only the vertices in S can be arranged in the unique form

$$(v_j, v_{(j+1) \mod n}, v_{(j+2) \mod n}, \dots, v_{(j+k-1) \mod n})$$

for some $j \in \{0,1,\dots,k-1\}$ such that $v_{(j+i) \mod n}$ is adjacent to $v_{(j+i+1) \mod n}$ for every $i \in \{0,1,\dots,k-2\}$. This representation fails to be unique if and only if vertex $v_{(j+k-1) \mod n}$ is adjacent to vertex v_j: this only happens when $k=n$. Therefore, C_n has precisely n connected subgraphs of every order k between 1 and $n-1$, while it has only one connected subgraph of order n (the subgraph C_n). This proves the lemma. □

The tadpole graph $G_{p,q}$ is the connected graph obtained by identifying a vertex of the cycle C_p with a pendent vertex of the path P_{q+1}. So the order of $G_{p,q}$ is $p+q$.

Lemma 10. *The tadpole graph $G_{p,q}$ has*

$$\binom{p}{2} + \binom{q+1}{2} + \frac{(q+1)(p^2-p+2)}{2}$$

connected subgraphs.

Proof. Consider the tadpole graph $G_{p,q}$ as depicted in Figure 1. We distinguish between connected subgraphs of $G_{p,q}$ that contain vertex v_0 and connected subgraphs of $G_{p,q}$ that do not contain v_0.

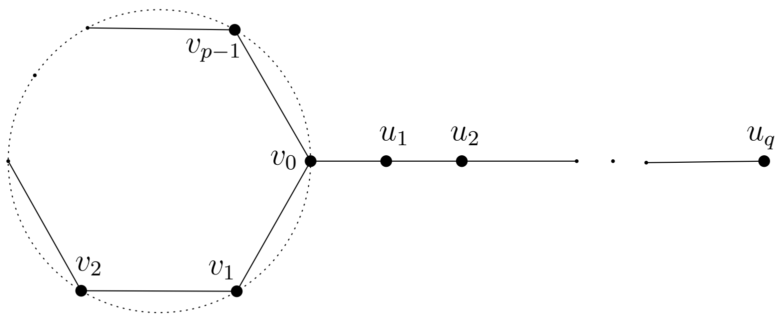

FIGURE 1. The tadpole graph $G_{p,q}$.

Since deleting v_0 in $G_{p,q}$ yields the two connected components P_{p-1} and P_q, we deduce by Theorem 3 that

$$N(P_{p-1}) + N(P_q) = \binom{p}{2} + \binom{q+1}{2}$$

gives the number of connected subgraphs of $G_{p,q}$ that do not contain v_0. On the other hand, every v_0-containing connected sugraph of $G_{p,q}$ is uniquely determined by merging a v_0-containing connected subgraph of the cycle C_p and a v_0-containing connected subgraph of the path P_{q+1} at vertex v_0 (a fixed pendent vertex of P_{q+1}): thus, there are

$$\frac{(q+1)(p^2-p+2)}{2}$$

of them. Indeed, a path of order n contains exactly n subtrees containing a fixed pendent vertex of P_n, while by the proof of Lemma 9, the number of v_0-containing connected subgraphs of C_p is given by $p^2-p+1-\binom{p}{2}$. This completes the proof of the lemma. □

We shall also need the following simple lemma about trees.

Lemma 11. *Let T be a rooted tree. The number of root-containing subtrees of T is at least the order of T, with equality if and only if T is a path rooted at one of its pendent vertices.*

Proof. By induction on the order n of a tree. The case $n=1$ is trivial. For the induction step, let $n>1$ and consider the r branches $T_1, T_2, \ldots, T_r$ of T (all connected components that remain after deleting the root u of T) endowed with their natural roots $u_1, u_2, \ldots, u_r$ (all the neighbors of u in T). Then the number $N(T)_u$ of root-containing subtrees of T is given by

$$N(T)_u = \prod_{j=1}^{r}(1+N(T_j)_{u_j}) \geq 1 + \sum_{j=1}^{r} N(T_j)_{u_j}\,.$$

This is established by noticing that every root containing subtree of T must involve the root or the empty set of a branch of T. By applying the induction hypothesis to each of the $N(T_j)_{u_j}$, we obtain

$$N(T)_u \geq 1 + \sum_{j=1}^{r} |V(T_j)| = |V(T)| = n\,.$$

Moreover, equality can only hold if $r=1$ at every induction step (in which case T is indeed a path). This proves the lemma. □

Our next result shows that the tadpole graph $G_{3,n-3}$ is the minimal graph with respect to the total number of connected subgraphs. It is interesting to point out that the tadpole graph $G_{3,n-3}$ is also known to maximise the Wiener index (the sum of the distances between all unordered pairs of vertices) among all unicyclic graphs of a given order; see [12, 15, 4].

For a graph G and two vertices u, v of G, we shall denote by $N(G)_u$ (resp. $N(G)_{u,v}$) the total number of connected subgraphs of G that contain u (resp. both u and v).

Theorem 12. *Among all unicyclic graphs of order n, only the tadpole graph $G_{3,n-3}$ has $(n^2+3n-4)/2$ connected subgraphs and this is the mininmum possible.*

Proof. The specialisation $p = 3$ in Lemma 10 yields

$$N(G_{3,n-3}) = \frac{(n-1)(n+4)}{2} = \frac{n^2+3n-4}{2}.$$

The statement is true for $n = 3$ as $C_3 = G_{3,0}$ is the only unicyclic graph of order 3 and $N(C_3) = 7$. By Lemma 9, we have $N(C_n) = n^2 - n + 1$ and so it is easy to see that

$$n^2 - n + 1 = N(C_n) > \frac{n^2+3n-4}{2} = N(G_{3,n-3})$$

provided that $n \neq 3$. For $n > 3$, let $G \neq C_n$ be a unicyclic graph of order n. It is easy to see that G has at least one pendent vertex (otherwise, G is a cycle).

Let v^* be a pendent vertex of G . In this case, $G - \{v^*\}$ is a unicyclic graph of order $n-1$. We induct on n to prove that $N(G) \geq (n-1)(n+4)/2$ with equality holding only for $G_{3,n-3}$. Consider the unique cycle $C_k = (v_0, v_1, v_2, \ldots, v_{k-1}, v_k = v_0)$ of G. For every $j \in \{0, 1, 2, \ldots, k-1\}$, the subgraph T_j of G depicted in Figure 2 is a tree rooted at vertex v_j.

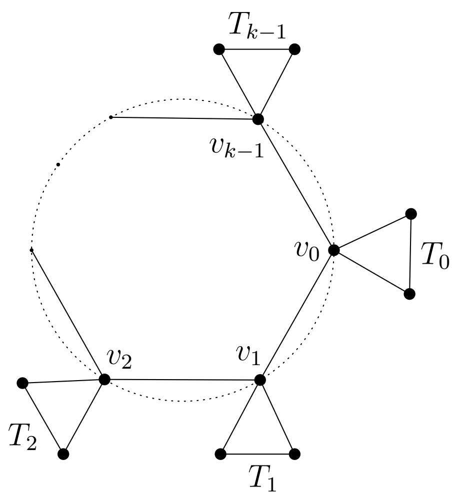

FIGURE 2. The general shape of a (connected) unicyclic graph.

Denote by a_j the number of v_j-containing subtrees of T_j and by $S_{j,1}, S_{j,2}, \ldots, S_{j,a_j}$ all the corresponding a_j subtrees. Let $S_{0,1}, S_{0,2}, \ldots, S_{0,b_0}$ be those subtrees of T_0 that contain both v^* and v_0 (we definitely assume that v^* is a leaf of T_0). It follows immediately that the subgraphs induced by

(1)
$$\begin{aligned}
&V(S_{0,1}), V(S_{0,2}), \ldots, V(S_{0,b_0}),\\
&V(T_0)\cup V(S_{1,1}), V(T_0)\cup V(S_{1,2}), \ldots, V(T_0)\cup V(S_{1,a_1}),\\
&V(T_0)\cup V(T_1)\cup V(S_{2,1}), V(T_0)\cup V(T_1)\cup V(S_{2,2}), \ldots,\\
&V(T_0)\cup V(T_1)\cup\cdots\cup V(T_{k-2})\cup V(S_{k-1,1}), V(T_0)\cup V(T_1)\cup\cdots\cup V(T_{k-2})\cup V(S_{k-1,2}),\\
&\ldots, V(T_0)\cup V(T_1)\cup\cdots\cup V(T_{k-2})\cup V(S_{k-1,a_{k-1}})
\end{aligned}$$

are all connected subgraphs of G that contain both v^* and v_0: thus, their number is precisely $b_0+a_1+\cdots+a_{k-1}$. Since $k\geq 3$, vertices v_1 and v_{k-1} are distinct; so the subgraph induced by $V(T_0)\cup V(S_{k-1,1})$ also contains both v^* and v_0.

Now consider those subtrees of T_0 that contain v^* but not v_0 (T_0 is chosen in such a way that v^* and v_0 are distinct vertices – this is clearly possible since G is not a cycle): thus, their number is $N(T_0)_{v^*}-b_0$, where $N(T_0)_{v^*}$ stands for the number of v^*-containing subtrees of T_0. Hence, the number of connected subgraphs of G that contain v^* is at least

$$b_0+a_1+\cdots+a_{k-1}+1+N(T_0)_{v^*}-b_0\,.$$

By Lemma 11, $N(T_0)_{v^*}\geq V(T_0)$ and $a_j\geq V(T_j)$ for every $j\in\{0,1,\ldots,k-1\}$. Hence, the number of connected subgraphs of G that contain v^* is at least

$$V(T_1)+\cdots+V(T_{k-1})+1+V(T_0)=n+1\,.$$

From the above discussion (including the types of subgraphs in (1)), equality can only hold if and only if

$$a_1=a_2=\cdots=a_{k-1}=1,\ k=3\quad\text{and}\quad T_0=P_{|V(T_0)|}\,.$$

On the other hand, since $G-\{v^*\}$ is a unicyclic graph of order $n-1$, we obtain

$$N(G-\{v^*\})\geq\frac{(n-2)(n+3)}{2}$$

by the induction hypothesis. Equality holds if and only if $G-\{v^*\}$ is the tadpole graph $G_{3,n-4}$. It follows that

$$N(G)\geq\frac{(n-2)(n+3)}{2}+n+1=\frac{(n-1)(n+4)}{2}\,.$$

Equality holds if and only if $G-\{v^*\}=G_{3,n-4}$, $k=3$, $|V(T_1)|=|V(T_2)|=1$ and $T_0=P_{n-2}$. In this case, we have $G=G_{3,n-3}$. This completes the proof of the theorem. $\square$

Before we can state our next and final theorem (the analogue of Theorem 12), we need to start with a few definitions and auxiliary results. Recall that the star S_n is the unique connected graph of order $n>2$ that has $n-1$ pendent vertices (S_1 is defined to P_1 and $S_2=P_2$). For $n>2$, its unique vertex of degree at least 2 is called the center of S_n. Denote by Q_n the connected graph obtained by adding one edge between two pendent vertices of the star S_n ($n>2$). Then Q_n contains only one cycle and its length is 3. We are going to show that the graph Q_n is maximal with respect to the total number of connected subgraphs of a unicyclic graph. Again, it is worth mentioning that the graph Q_n is known to minimise the Wiener index among all unicyclic graphs of a given order [12, 15, 4].

It is important to note that since $Q_4 = G_{3,1}$ and there are only two unicyclic graphs of order 4, the cycle C_4 is therefore the maximal graph. On the other hand, it is not difficult to see that out of the five unicyclic graphs of order 5 (see Figure 3), only the cycle C_5, the graph Q_5 and the so-called banner graph B_5 (obtained by dropping a pendent edge from a vertex of C_4) are maximal: $N(C_5) = N(Q_5) = N(B_5) = 21$. Also, recall that $Q_3 = C_3$ is the only unicyclic graph of order 3.

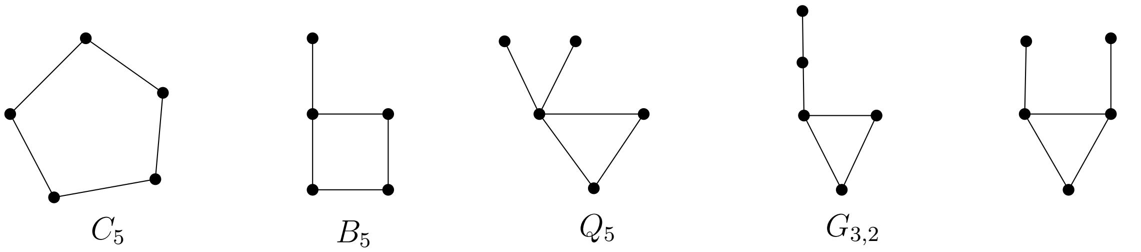

FIGURE 3. All the unicyclic graphs of order 5.

We begin with a counterpart of Lemma 11.

Proposition 13. *Let T be a rooted tree of order n whose root is v. Then the number of v-containing subtrees of T is at most 2^{n-1}. Equality holds if and only if T is the star S_n.*

Proof. We go by induction on the order n of the tree. The case $n = 1$ is trivial. Let T be a rooted tree of order $n > 1$ whose root is v. Denote by $T_1, T_2, \ldots, T_r$ all the branches of T endowed with their natural roots $v_1, v_2, \ldots, v_r$ (all neighbors of v). As in Lemma 11, every v-containing subtree of T must involve the root or the empty set of a branch of T: this yields

$$N(T)_v = \prod_{j=1}^{r} (1 + N(T_j)_{v_j}) .$$

Note that $n \geq r+1$ since T has precisely r branches for some $r \geq 1$. The induction hypothesis implies that

$$
\begin{aligned}
N(T)_v \leq & \prod_{j=1}^{r}(1+2^{|V(T_j)|-1}) \\
&= 1 + 2^{-1}\sum_{j=1}^{r} 2^{|V(T_j)|} \ + \ 2^{-2} \sum_{1\leq i_1<i_2\leq r} 2^{|V(T_{i_1})|+|V(T_{i_2})|} \ + \ \cdots \\
&+ 2^{-(r-1)} \sum_{1\leq i_1<\cdots<i_{r-1}\leq r} 2^{|V(T_{i_1})|+\cdots+|V(T_{i_{r-1}})|} \ + \ 2^{-r}\cdot 2^{\sum_{j=1}^{r}|V(T_j)|} \\
&\leq 1 + 2^{-1}\sum_{j=1}^{r} 2^{n-1-(r-1)} \ + \ 2^{-2} \sum_{1\leq i_1<i_2\leq r} 2^{n-1-(r-2)} \ + \ \cdots \\
&+ 2^{-(r-1)} \sum_{1\leq i_1<\cdots<i_{r-1}\leq r} 2^{n-1-(r-(r-1))} \ + \ 2^{-r}\cdot 2^{n-1-(r-r)}
\end{aligned}
$$

as $|V(T_1)|+\cdots+|V(T_r)| = |V(T)| = n-1$. It follows that

$$
\begin{aligned}
N(T)_v &\leq 1 + 2^{-1}\binom{r}{1}2^{n-r} \ + \ 2^{-2}\binom{r}{2}2^{n-(r-1)} \ + \cdots \\
&+ 2^{-(r-1)}\binom{r}{r-1}2^{n-2} \ + \ 2^{-r}\binom{r}{r}2^{n-1} \\
&= 1 + 2^{n-r-1}\sum_{i=1}^{r}\binom{r}{i} = 1 - 2^{n-r-1} + 2^{n-1} \leq 2^{n-1}
\end{aligned}
$$

and equality holds if and only if $n = r+1$, in which case T is indeed a star. This completes the proof of the proposition. ☐

In general, we define the banner graph B_n of order $n \geq 4$ to be the connected graph constructed from C_4 by dropping $n-4$ pendent edges from the same vertex of C_4.

Lemma 14. *The banner graph B_n has precisely $2+n+7\cdot 2^{n-4}$ connected subgraphs.*

Proof. The statement is true for $n = 4$ as $B_4 = C_4$ and $N(C_4) = 13$ by Lemma 9. For $n > 4$, let v be the neighbor of a pendent vertex of B_n (see Figure 4). Deleting v in B_n yields $n-4$ copies of the one vertex graph and one copy of the path P_3. So $N(B_n - \{v\}) = n-4+N(P_3) = n-2$ as $N(P_n) = \binom{n+1}{2}$ by Theorem 3.

On the other hand, every v-containing connected subgraph of B_n decomposes naturally into a v-containing connected subgraph of the cycle C_4 and a connected subgraph of the star S_{n-3} whose root is v: this gives $N(B_n)_v = N(C_4)_v \cdot N(S_n)_v = 7\cdot 2^{n-4}$ by Proposition 13. Thus, $N(B_n) = 2+n+7\cdot 2^{n-4}$, completing the proof of the lemma. ☐

We are now ready to state our next theorem.

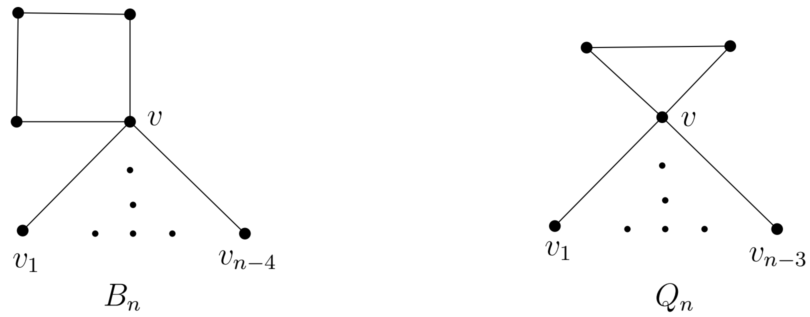

FIGURE 4. The unicyclic graphs B_n (left) and Q_n (right) of order n.

Theorem 15. *Among all unicyclic graphs of a fixed order $n > 5$, only the graph Q_n has $n + 2^{n-1}$ connected subgraphs and this is the maximum possible.*

To prove the theorem, we shall need one more auxiliary result.

Proposition 16. *Let T be a rooted tree of order $n > 1$ whose root is v. If $l \neq v$ is a leaf of T, then the number of subtrees of T that involve both v and l is at most 2^{n-2}. Equality holds if and only if T is the star S_n.*

In a certain sense, Proposition 16 parallels a result of Székely and Wang [11] who proved that the star maximises the number of subtrees of T that contain at least one leaf of T among all trees T of a given order.

Proof of Proposition 16. We prove the statement by induction on n. The case $n = 2$ is trivial since the tree in this case is the path P_2. For $n > 2$, let $T_1, T_2, \ldots, T_r$ be all the branches of T endowed with their natural roots $v_1, v_2, \ldots, v_r$, respectively. If $r = 1$, then the number $N(T)_{v,l}$ of subtrees of T that involve both v and l is precisely $N(T_1)_{v_1,l}$ and since $v_1 \neq l$, we can apply the induction hypothesis to T_1: this gives

$$N(T)_{v,l} = N(T_1)_{v_1,l} \leq 2^{|V(T_1)|-2} = 2^{|V(T)|-3} < 2^{|V(T)|-2}$$

and in this case, we are done. Otherwise $r \geq 2$: If T is a star, then the number of subtrees of T that involve both v and l is precisely $N(S_{n-1})_v$. By Proposition 13, $N(S_{n-1})_v = 2^{n-2}$ and in this case, we are done as well. Otherwise, we can assume without loss of generality that $|V(T_1)| \geq 2$ and let l be a leaf of T_1. Denote by D the tree whose branches are $T_2, \ldots, T_r$ (the root of D is v). With this decomposition, $N(T)_{v,l}$ is immediately given by

$$N(T)_{v,l} = N(T_1)_{v_1,l} \cdot N(D)_v$$

as every subtree of T containing both v, l induces a subtree of T_1 containing both v_1, l and a subtree of $T - V(T_1) = D$ containing v. Thus, the induction hypothesis yields

$$N(T)_{v,l} \leq 2^{|V(T_1)|-2} \cdot N(D)_v \leq 2^{|V(T_1)|-2} \cdot 2^{|V(D)|-1} = 2^{V(T)-3}\,,$$

where the last inequality follows from Proposition 13. This completes the induction hypothesis and thus the proof of the proposition. ∎

We can now give a proof of Theorem 15.

Proof of Theorem 15. For $n > 3$, let v be the unique vertex of Q_n whose degree is at least 3 (see Figure 4). By identifying v with a pendent vertex of the path P_2 and a vertex of the cycle C_3, all connected v-containing subgraphs of Q_n are uniquely determined by taking $n-3$ connected subgraphs of P_2 that contain v, a connected subgraph of C_3 that contain v and merging them at v. Thus, their number is $4 \cdot 2^{n-3} = 2^{n-1}$. On the other hand, deleting vertex v in Q_n yields $n-3$ copies of the single vertex graph and one copy of P_2. Thus, the number of connected subgraphs of Q_n that do not involve v is $n-3+3 = n$. This proves that $N(Q_n) = 2^{n-1} + n$.

It is obvious that if G is the banner graph B_n of order $n > 5$, then by Lemma 14, we have

$$N(G) = N(B_n) = 2 + n + 7 \cdot 2^{n-4} < n + 2^{n-1} = N(Q_n)\,.$$

For the rest of the proof, we assume that G is not a banner graph. Let G be a unicyclic graph of a fixed order $n \geq 5$. Let $C_k = (v_0, v_1, \ldots, v_{k-1}, v_k = v_0)$ be the unique cycle of G. Then G has precisely the shape depicted in Figure 2, where $T_0, T_1, \ldots, T_{k-1}$ are all trees rooted at vertices $v_0, v_1, \ldots, v_{k-1}$, respectively. Assume G has the maximum number of connected subgraphs among all unicyclic graphs of order n.

Claim 1: *Each of the trees $T_0, T_1, \ldots, T_{k-1}$ is a star.*

For the proof of the claim, suppose (without loss of generality) that T_0 is not a star. Let us first derive a formula for the number $N(G; T_0)$ of connected subgraphs of G that contain a subtree of T_0. Every such subgraph is

(1) either a subtree of T_0 only;
(2) or contains all vertices of C_k;
(3) or involves v_0 and a left sequence $v_1, v_2, \ldots, v_l$ of consecutive vertices of C_k such that $l < k-1$;
(4) or involves v_0 and a right sequence $v_{k-1}, v_{k-2}, \ldots, v_{k-r}$ of consecutive vertices of C_k such that $k - r > 1$;
(5) or involves v_0, a (left) nonempty sequence $v_1, v_2, \ldots, v_l$ of consecutive vertices of C_k and a (right) nonempty sequence $v_{k-1}, v_{k-2}, \ldots, v_{k-r}$ of consecutive vertices of C_k such that $l + 1 < k - r$.

Thus, we have

$$N(G; T_0) = N(T_0) + \prod_{j=0}^{k-1} N(T_j)_{v_j} \; + N(T_0)_{v_0} \sum_{l=1}^{k-2} \prod_{j=1}^{l} N(T_j)_{v_j} \; + N(T_0)_{v_0} \sum_{r=1}^{k-2} \prod_{j=1}^{r} N(T_{k-j})_{v_{k-j}}$$
$$+ \sum_{l=1}^{k-3} \sum_{r=1}^{k-l-2} \left(\prod_{j=1}^{l} N(T_j)_{v_j} \cdot N(T_0)_{v_0} \cdot \prod_{j=1}^{r} N(T_{k-j})_{v_{k-j}} \right),$$

where every single summand corresponds to the cases distinction in this order.

Construct from G a new unicyclic graph G' by replacing T_0 with the star $T_0' = S_{|V(T_0)|}$ centered at vertex v_0. Thus, we also have

$$N(G';T_0') = N(T_0') \ + N(T_0')_{v_0} \prod_{j=1}^{k-1} N(T_j)_{v_j} \ + N(T_0')_{v_0} \sum_{l=1}^{k-2} \prod_{j=1}^{l} N(T_j)_{v_j}$$
$$+ N(T_0')_{v_0} \sum_{r=1}^{k-2} \prod_{j=1}^{r} N(T_{k-j})_{v_{k-j}} \ + \sum_{l=1}^{k-3} \sum_{r=1}^{k-l-2} \left(\prod_{j=1}^{l} N(T_j)_{v_j} \cdot N(T_0')_{v_0} \cdot \prod_{j=1}^{r} N(T_{k-j})_{v_{k-j}} \right)$$

by a simple substitution. By taking the difference, we obtain

$$N(G';T_0') - N(G;T_0) = N(T_0') - N(T_0) + (N(T_0')_{v_0} - N(T_0)_{v_0}) \Bigg(\prod_{j=1}^{k-1} N(T_j)_{v_j}$$
$$+ \sum_{l=1}^{k-2} \prod_{j=1}^{l} N(T_j)_{v_j} \ + \sum_{r=1}^{k-2} \prod_{j=1}^{r} N(T_{k-j})_{v_{k-j}} \ + \sum_{l=1}^{k-3} \sum_{r=1}^{k-l-2} \left(\prod_{j=1}^{l} N(T_j)_{v_j} \cdot \prod_{j=1}^{r} N(T_{k-j})_{v_{k-j}} \right) \Bigg).$$

By Proposition 13, we have $N(T_0')_{v_0} - N(T_0)_{v_0} > 0$, while it is also known that $N(T_0') - N(T_0) > 0$ (see Szekely and Wang [11, Theorem 3.1]). Hence, $N(G';T_0') > N(G;T_0)$. This contradicts the optimality of G as both G and G' have the same number of connected subgraphs avoiding a subtree of T_0 or T_0'. The claim is proved.

In the following, we assume that each of the trees $T_0, T_1, \dots, T_{k-1}$ is a star.

Claim 2: *We have* $k = 3$.

For the proof of this claim, suppose (for contradiction) that $k > 3$. Let us first derive a formula for the number $N(G; T_0 \cup T_{k-1})$ of connected subgraphs of G that contain a subtree of T_0 or T_{k-1}. The number of connected subgraphs of G that contain

(1) a subtree of T_0 but not a subtree of T_{k-1} is given by

$$N(T_0) \ + N(T_0)_{v_0} \sum_{l=1}^{k-2} \prod_{j=1}^{l} N(T_j)_{v_j} \,;$$

(2) a subtree of T_{k-1} but not a subtree of T_0 is given by

$$N(T_{k-1}) \ + N(T_{k-1})_{v_{k-1}} \sum_{r=2}^{k-1} \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} \,;$$

(3) a subtree of both T_0 and T_{k-1} is given by

$$\begin{aligned} & N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}} + \prod_{j=0}^{k-1} N(T_j)_{v_j} + N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}} \sum_{l=1}^{k-3} \prod_{j=1}^{l} N(T_j)_{v_j} \\ & \quad + N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}} \sum_{r=2}^{k-2} \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} \\ & \quad + \sum_{l=1}^{k-4} \sum_{r=2}^{k-l-2} \left(\prod_{j=1}^{l} N(T_j)_{v_j} \ (N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}}) \ \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} \right). \end{aligned}$$

Combining all cases, we obtain

(2)

$$\begin{aligned} N(G;T_0 \cup T_{k-1}) &= N(T_0) + N(T_{k-1}) + N(T_0)_{v_0} \sum_{l=1}^{k-2} \prod_{j=1}^{l} N(T_j)_{v_j} \\ & + N(T_{k-1})_{v_{k-1}} \sum_{r=2}^{k-1} \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} + N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}} \Bigg(1 + \prod_{j=1}^{k-2} N(T_j)_{v_j} \\ & + \sum_{l=1}^{k-3} \prod_{j=1}^{l} N(T_j)_{v_j} + \sum_{r=2}^{k-2} \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} + \sum_{l=1}^{k-4} \sum_{r=2}^{k-l-2} \prod_{j=1}^{l} N(T_j)_{v_j} \ \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} \Bigg) \end{aligned}$$

for the number of connected subgraphs of G that contain a subtree of T_0 or T_{k-1}.

Construct from G a new unicyclic graph G'' by deleting all vertices of T_{k-1} except v_{k-1}, then contracting vertex v_{k-1} and finally replacing T_0 with the star $T_0'' = S_{|V(T_0)|+|V(T_{k-1})|}$ centered at vertex v_0. By distinguishing cases as we did in Claim 1, we obtain

(3)

$$\begin{aligned} N(G;T_0'') &= N(T_0'') + N(T_0'')_{v_0} \prod_{j=1}^{k-2} N(T_j)_{v_j} + N(T_0'')_{v_0} \sum_{l=1}^{k-3} \prod_{j=1}^{l} N(T_j)_{v_j} \\ & + N(T_0'')_{v_0} \sum_{r=2}^{k-2} \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} + \sum_{l=1}^{k-4} \sum_{r=2}^{k-l-2} \left(\prod_{j=1}^{l} N(T_j)_{v_j} \ N(T_0'')_{v_0} \ \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} \right) \end{aligned}$$

for the number $N(G;T_0'')$ of connected subgraphs of G'' that contain a subtree of T_0''. The difference (3)-(2) is given by

$$\begin{aligned}
&N(G;T_0'') - N(G;T_0 \cup T_{k-1}) = N(T_0'') - N(T_0) - N(T_{k-1}) - N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}}\\
&+ N(T_0'')_{v_0} - N(T_0)_{v_0}(1 + N(T_{k-1})_{v_{k-1}}) \sum_{l=1}^{k-2} \prod_{j=1}^{l} N(T_j)_{v_j}\\
&+ N(T_0'')_{v_0} - N(T_{k-1})_{v_{k-1}}(1 + N(T_0)_{v_0}) \sum_{r=2}^{k-2} \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}}\\
&+ (N(T_0'')_{v_0} - N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}}) \sum_{l=1}^{k-5} \sum_{r=2}^{k-l-2} \left(\prod_{j=1}^{l} N(T_j)_{v_j} \prod_{j=2}^{r} N(T_{k-j})_{v_{k-j}} \right)\\
&+ (N(T_0'')_{v_0} - N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}} - N(T_{k-3})_{v_{k-3}} N(T_{k-1})_{v_{k-1}}) \prod_{\substack{j=1\\ j\neq k-3}}^{k-2} N(T_j)_{v_j}
\end{aligned} \tag{4}$$

after some basic manipulations. By Proposition 13 and Theorem 3.1 in [11], we have

$$\begin{aligned}
&N(T_0'') - N(T_0) - N(T_{k-1}) - N(T_0)_{v_0} \cdot N(T_{k-1})_{v_{k-1}}\\
&= \big(2^{|V(T_0)|+|V(T_{k-1})|-1} + |V(T_0)| + |V(T_{k-1})| - 1\big) - \big(2^{|V(T_0)|-1} + |V(T_0)| - 1\big)\\
&- \big(2^{|V(T_{k-1})|-1} + |V(T_{k-1})| - 1\big) - 2^{|V(T_0)|-1} \cdot 2^{|V(T_{k-1})|-1}\\
&= \big(2^{|V(T_{k-1})|-1} - 1\big)\big(2^{|V(T_0)|-1} - 1\big) \geq 0
\end{aligned}$$

and

$$\begin{aligned}
N(T_0'')_{v_0} - N(T_0)_{v_0}\big(1 + N(T_{k-1})_{v_{k-1}}\big) &= 2^{|V(T_0)|+|V(T_{k-1})|-1} - 2^{|V(T_0)|-1}\big(1 + 2^{|V(T_{k-1})|-1}\big)\\
&= 2^{|V(T_0)|-1}\big(2^{|V(T_{k-1})|-1} - 1\big) \geq 0\,.
\end{aligned}$$

Likewise,

$$\begin{aligned}
N(T_0'')_{v_0} - N(T_{k-1})_{v_{k-1}}\big(1 + N(T_0)_{v_0}\big) &= 2^{|V(T_0)|+|V(T_{k-1})|-1} - 2^{|V(T_{k-1})|-1}\big(1 + 2^{|V(T_0)|-1}\big)\\
&= 2^{|V(T_{k-1})|-1}\big(2^{|V(T_0)|-1} - 1\big) \geq 0
\end{aligned}$$

and

$$\begin{aligned}
N(T_0'')_{v_0} - N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}} &= 2^{|V(T_0)|+|V(T_{k-1})|-1} - 2^{|V(T_0)|-1} \cdot 2^{|V(T_{k-1})|-1}\\
&= 2^{|V(T_0)|+|V(T_{k-1})|-2} > 0\,.
\end{aligned}$$

Also,

$$\begin{aligned}
&N(T_0'')_{v_0} - N(T_0)_{v_0} N(T_{k-1})_{v_{k-1}} - N(T_{k-3})_{v_{k-3}} N(T_{k-1})_{v_{k-1}}\\
&\qquad = 2^{|V(T_{k-1})|-2}\big(2^{|V(T_0)|} - 2^{|V(T_{k-3})|}\big) \geq 0
\end{aligned}$$

as T_0 can be chosen to have the maximum order among the trees $T_0, T_1, \ldots, T_{k-1}$. The following conclusions about (4) can be derived immediately:

- If $k \geq 6$ then $N(G;T_0'') > N(G;T_0 \cup T_{k-1})$;
- If $k = 5$ then

$$\begin{aligned}0 \leq N(G;T_0'') - N(G;T_0 \cup T_4) &= \big(2^{|V(T_4)|-1} - 1\big)\big(2^{|V(T_0)|-1} - 1\big)\\ &+ 2^{|V(T_0)|-1}\big(2^{|V(T_4)|-1} - 1\big)\sum_{l=1}^{3}\prod_{j=1}^{l} N(T_j)_{v_j}\\ &+ 2^{|V(T_4)|-1}\big(2^{|V(T_0)|-1} - 1\big)\sum_{r=2}^{3}\prod_{j=2}^{r} N(T_{5-j})_{v_{5-j}}\\ &+ 2^{|V(T_4)|-2}\big(2^{|V(T_0)|} - 2^{|V(T_2)|}\big)N(T_1)_{v_1}N(T_3)_{v_3}\,.\end{aligned}$$

  Thus, $N(G;T_0'') > N(G;T_0 \cup T_{k-1})$ as soon as $|V(T_0)| > 1$. This is indeed the case since T_0 was chosen to have the maximum order among the trees $T_0, T_1, \ldots, T_{k-1}$: we have $|V(T_0)| = 1$ if and only if $G = C_5$ (the cycle of order 5).
- If $k = 4$ then

$$\begin{aligned}0 \leq N(G;T_0'') - N(G;T_0 \cup T_3) &= \big(2^{|V(T_3)|-1} - 1\big)\big(2^{|V(T_0)|-1} - 1\big)\\ &+ 2^{|V(T_0)|-1}\big(2^{|V(T_3)|-1} - 1\big)\sum_{l=1}^{2}\prod_{j=1}^{l} N(T_j)_{v_j}\\ &+ 2^{|V(T_3)|-1}\big(2^{|V(T_0)|} - 1 - 2^{|V(T_1)|-1}\big)N(T_2)_{v_2}\,.\end{aligned}$$

  It is easy to see that $N(G;T_0'') - N(G;T_0 \cup T_3) = 0$ if and only if $|V(T_0)| = |V(T_1)| = |V(T_3)| = 1$.

Now, observe that the number of connected subgraphs of G that contain neither a subtree of T_0, nor a subtree of T_{k-1} is the same as the number of connected subgraphs of G'' that do not contain a subtree of T_0''. Altogether, we conclude that $k = 3$. This completes the proof of the claim.

The specialisation $k = 3$ in equation (2) yields

$$\begin{aligned}N(G;T_0 \cup T_2) = N(T_0) + N(T_2) &+ N(T_0)_{v_0}N(T_1)_{v_1} + N(T_2)_{v_2}N(T_1)_{v_1}\\ &+ N(T_0)_{v_0}N(T_2)_{v_2}(1 + N(T_1)_{v_1})\end{aligned}$$

for the number of connected subgraphs of G that contain a subtree of T_0 or T_2. We assume that T_0 has the maximum order among the trees T_0, T_1, T_2.

Claim 3: *We have* $|V(T_1)| = |V(T_2)| = 1$.

To see this, suppose (for contradiction) that $|V(T_2)| > 1$. Construct from G a new unicyclic graph G''' by replacing both T_2 with the star $T_2''' = S_{|V(T_2)|-1}$ centered at vertex v_2, and T_0 with the star $T_0''' = S_{|V(T_0)|+1}$ centered at vertex v_0. Thus, the number $N(G;T_0''' \cup T_2''')$ of connected subgraphs of G''' that contain a subtree of T_0''' or T_2''' is given by

$$\begin{aligned}N(G;T_0''' \cup T_2''') = N(T_0''') + N(T_2''') &+ N(T_0''')_{v_0}N(T_1)_{v_1} + N(T_2''')_{v_2}N(T_1)_{v_1}\\ &+ N(T_0''')_{v_0}N(T_2''')_{v_2}(1 + N(T_1)_{v_1})\end{aligned}$$

which implies that

$$\begin{aligned} N(G;T_0''' \cup T_2''') - N(G;T_0 \cup T_2) &= N(T_0''') + N(T_2''') - N(T_0) - N(T_2) \\ &+ (N(T_0''')_{v_0} + N(T_2''')_{v_2} - N(T_0)_{v_0} - N(T_2)_{v_2})N(T_1)_{v_1} \\ &+ (N(T_0''')_{v_0}N(T_2''')_{v_2} - N(T_0)_{v_0}N(T_2)_{v_2})(1 + N(T_1)_{v_1})\,. \end{aligned}$$

Again Proposition 13 along with Theorem 3.1 in [11] gives

$$N(T_0''') + N(T_2''') - N(T_0) - N(T_2) = 2^{|V(T_0)|-1} - 2^{|V(T_2)|-2}\,,$$
$$N(T_0''')_{v_0} + N(T_2''')_{v_2} - N(T_0)_{v_0} - N(T_2)_{v_2} = 2^{|V(T_0)|-1} - 2^{|V(T_2)|-2}$$

and

$$N(T_0''')_{v_0}N(T_2''')_{v_2} - N(T_0)_{v_0}N(T_2)_{v_2} = 0\,.$$

Therefore, we get

$$N(G;T_0''' \cup T_2''') - N(G;T_0 \cup T_2) = (2^{|V(T_0)|-1} - 2^{|V(T_2)|-2})(1 + N(T_1)_{v_1}) > 0$$

which completes the proof of the claim. Hence, $|V(T_1)| = |V(T_2)| = 1$ and this also completes the proof of the theorem. ☐

References

[1] E. O. D. Andriantiana. Unicyclic bipartite graphs with maximum energy. *MATCH Commun. Math. Comput. Chem*, 66(3):913–926, 2011.

[2] E. O. D. Andriantiana and S. Wagner. Unicyclic graphs with large energy. *Linear Algebra and its Applications*, 435(6):1399–1414, 2011.

[3] Y-C. Chen and Y.-J. Tian. Induced subgraphs with maximum number of edges on crossed cubes. *Pervasive Computing (JCPC), 2009 Joint Conferences on.*, 829–832, IEEE, 2009.

[4] Y. Hong, H. Liu, and X. Wu. On the wiener index of unicyclic graphs. *Hacettepe J. Mathematics and Statistics*, 40(1):63–68, 2011.

[5] Y. Hou. Unicyclic graphs with minimal energy. *J. Mathematical Chemistry*, 29(3):163–168, 2001.

[6] Y. Hou, I. Gutman, and C.-W. Woo Unicyclic graphs with maximal energy. Special issue on algebraic graph theory (Edinburgh, 2001). *Linear Algebra and its Applications*, 356(1-3):27–36, 2002.

[7] M. E. Kroeker, L. Mol, and O. R. Oellermann. On the mean connected induced subgraph order of cographs. *Australas. J. Combin*, 71, 161–183, 2018.

[8] F. Li and B. Zhou. Minimal energy of unicyclic graphs of a given diameter. *J. Mathematical Chemistry*, 43(2):476–484, 2008.

[9] J. Ou. On extremal unicyclic molecular graphs with maximal Hosoya index. *Discrete Applied Mathematics*, 157(2):391–397, 2009.

[10] A. S. Pedersen and P. D. Vestergaard The number of independent sets in unicyclic graphs. *Discrete Applied Mathematics*, 152(1-3):246–256, 2005.

[11] L. A. Székely and H. Wang. On subtrees of trees. *Advances in Applied Mathematics*, 34(1):138–155, 2005.

[12] Z. Tang and H. Deng. The (n, n)-graphs with the first three extremal Wiener indices. *J. Mathematical Chemistry*, 43(1):60–74, 2008.

[13] P. Tittmann, I. Averbouch, and J. A. Makowsky. The enumeration of vertex induced subgraphs with respect to the number of components. *European J. Combinatorics*, 32(7):954–974, 2011.

[14] W. Yan and Y.-N. Yeh. Enumeration of subtrees of trees. *Theoretical Computer Science*, 369(1-3):256–268, 2006.

[15] G. Yu and L. Feng. On the Wiener index of unicyclic graphs with given girth. *Ars Comb*, 94:361–369, 2010.

[16] Z. Zhu and G. Chen. On Merrifield-Simmons index of unicyclic graphs with given girth and prescribed pendent vertices. *Australasian J. Combinatorics*, 50:87–95, 2011.

[17] The On-Line Encyclopedia of Integer Sequences. Published electronically at https://oeis.org, 2018.

[18] N. W. Lemons, B. Hu, and W. S. Hlavacek. Hierarchical graphs for rule-based modeling of biochemical systems. *BMC Bioinformatics*, 12(1):p45, 2011.

[19] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. A (sub) graph isomorphism algorithm for matching large graphs. *IEEE transactions on pattern analysis and machine intelligence*, 26(10):1367–1372, 2004.

[20] S. Wernicke and F. Rasche. FANMOD: a tool for fast network motif detection. *Bioinformatics*, 22(9):1152–1153, 2006.

[21] M. R. Garey. A Guide to the Theory of NP-Completeness. *Computers and intractability*, Freeman and Company, 1979.

[22] D. Eppstein. Subgraph isomorphism in planar graphs and related problems. *J. Graph Algorithms And Applications*, 3(3):1–27, 1999.

[23] Y. Chen, M. Thurley, and M. Weyer. Understanding the complexity of induced subgraph isomorphisms. *International Colloquium on Automata, Languages, and Programming*, 587–596, 2008.

Audace A. V. Dossou-Olory, Department of Mathematical Sciences, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa

E-mail address: audace@aims.ac.za