

**On Pitfalls in Computing the
Geodetic Number of a Graph**

P. Hansen
N. van Omme

G-2005-89

November 2005

Revised: September 2006

Les textes publiés dans la série des rapports de recherche HEC n'engagent que la responsabilité de leurs auteurs. La publication de ces rapports de recherche bénéficie d'une subvention du Fonds québécois de la recherche sur la nature et les technologies.

On Pitfalls in Computing the Geodetic Number of a Graph

Pierre Hansen

*GERAD and HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7
pierre.hansen@gerad.ca*

Nikolaj van Omme

*CRT and École Polytechnique de Montréal
C.P. 6079, Succ. Centre-ville
Montréal (Québec) Canada H3C 3A7
nikolaj@crt.umontreal.ca*

November 2005
Revised: September 2006

Les Cahiers du GERAD

G-2005-89

Copyright © 2006 GERAD

Abstract

Given a simple connected graph $G = (V, E)$ the geodetic closure $I[S] \subset V$ of a subset S of V is the union of all sets of nodes lying on some geodesic (or shortest path) joining a pair of nodes $v_k, v_l \in S$. The geodetic number, denoted by $g(G)$, is the smallest cardinality of a node set S^* such that $I[S^*] = V$. In “The geodetic number of a graph”, *Mathematical and Computer Modelling* 17 (June 1993) 89–95, F. Harary, E. Loukakis and C. Tsouros propose an incorrect algorithm to find the geodetic number of a graph G . We provide counterexamples and show why the proposed approach must fail. We then develop a 0-1 integer programming model to find the geodetic number. Computational results are given.

Key Words: Graph, Geodetic number, Maximal Geodetic closure, Algorithm.

Résumé

Étant donné un graphe $G = (V, E)$ simple et convexe, la fermeture géodésique $I[S] \subset V$ d'un sous-ensemble S de V est l'union de tous les ensembles de sommets situés sur une géodésique (ou plus court chemin) joignant une paire de sommets $v_k, v_l \in S$. Le nombre géodétique, noté $g(G)$, est la plus petite cardinalité d'un ensemble de sommets S^* tel que $I[S^*] = V$. Dans l'article “The geodetic number of a graph”, *Mathematical and Computer Modelling* 17 (juin 1993) 89–95, F. Harary, E. Loukakis et C. Tsouros proposent un algorithme incorrect pour déterminer le nombre géodétique d'un graphe G . Nous présentons des contre-exemples et montrons pourquoi l'approche proposée ne peut réussir. Nous développons ensuite un modèle de programmation entière en variables 0-1 pour déterminer le nombre géodétique et donnons des résultats de calcul.

Mots clés : Graphe, nombre géodétique, fermeture géodésique maximale, algorithme.

1 Introduction

We start by giving some definitions. Let $G = (V, E)$ be a connected graph with node set V and edge set E . Denote the number of nodes by $n = |V|$ and the number of edges by $m = |E|$. A *geodesic* between two nodes u and v is a shortest path between these two nodes. A *maximal geodesic* is a geodesic that cannot be extended. In other words, it is a geodesic that is not properly contained in any other geodesic. Denote by the closed interval $I[u, v]$ the set of all nodes of G lying on any geodesic between the nodes u and v . In a similar way, denote by $I[S]$ the union of all $I[u, v]$ for all u and v in $S \subset V$. $I[S]$ is called the *geodetic closure* of the set $S \subset V$. So,

$$I[S] = \{x \in V \mid \exists u, v \in S, x \in I[u, v]\}.$$

When the geodetic closure of a set S is V itself, S is a *geodetic set*. The smallest cardinality of a geodetic set is the *geodetic number* $g(G)$ of the graph G :

$$g(G) = \min\{|S| \mid S \subset V \text{ and } I[S] = V\}.$$

The set S itself, i.e., one with the property that $|S| = g(G)$, is called a *minimum geodetic set*.

The geodetic number was first introduced in 1990 (see [4]) and is further studied in [1, 2, 3, 5, 6, 7, 8] and [9]. Several classes of graphs and their geodetic number are presented in [8]. To the best of our knowledge, only one algorithm for finding the geodetic number has been proposed, by F. Harary, E. Loukakis and C. Tsouros [8], but no computational results were given.

The complexity of the problem of finding the geodetic number has been studied in [1] where Atici has shown that the corresponding GEODETIC SET decision problem, namely “*given a nontrivial connected graph $G = (V, E)$ and an integer $k \leq |V|$, is there a set $S \subset V$ with $|S| = k$ and such that $I[S] = V$?*”, is NP-Complete.

Several generalizations were made. In a natural way, the definition extends to the directed case which is studied in [7]. Let $D = (V, A)$ be a connected, directed graph with node set V and arc set A . For two nodes u and v of V , let $I[u, v]$ denote the set of nodes either on a directed $u - v$ geodesic or on a directed $v - u$ geodesic. Extension of the definitions of geodetic closure $I[S]$ of $S \subset V$, geodetic sets, minimum geodetic set and geodetic number to the directed case is straightforward. Do different orientations of a graph lead to different geodetic numbers? The answer is yes and the set of all the values $g(D)$ for all possible orientations is called the *geodetic spectrum* of a graph. Among them, the minimum and the maximum received special attention and are called the *lower orientable geodetic number* $g^-(D)$ and the *upper orientable geodetic number* $g^+(D)$ respectively. The geodetic spectrum of a graph is further studied in [5].

Many graph concepts have an edge and a node version. Atici defined and studied the edge version of the geodetic number in [2] and in [3] with Vince. Define $\varepsilon(u, v)$ as the set

of all edges lying on some $u - v$ geodesic of G . For a non-empty subset $S \subset V$ define

$$\varepsilon(S) = \bigcup_{u,v \in S} \varepsilon(u, v).$$

A subset $S \subset V$ is called an *edge geodetic set* if $\varepsilon(S) = E$. An edge geodetic set of minimum cardinality is called a *minimum edge geodetic set* and this cardinality is the *edge geodetic number* of G , denoted by $g_e(G)$.

A number of lower and upper bounds have been proposed. Of course, $2 \leq g(G) \leq n$. The set of the two end-nodes of a path P_n is a geodetic set and so $g(P_n) = 2$ while the only graph G with $g(G) = n$ is K_n . Because every simplicial node, i.e., a node all pairs of neighbors of which are adjacent, has to be in any geodetic set ([10]), the number of simplicial nodes gives a lower bound on $g(G)$. On the other hand, every cutnode, i.e., a node whose deletion disconnects G , has to be excluded from any minimum geodetic set ([4], page 137), so we have an upper bound given by the number of nodes that are not cutnodes. Chartrand et al. show in [6] that $g(G) \leq n - d + 1$ where d is the diameter of the graph G , i.e., the largest distance between a pair of nodes of G . Atici and Vince follow a completely different approach in [3] to find a lower bound on $g_e(G)$, i.e., $g_e(G) \geq \lceil 3 \log_3(\omega(G)) \rceil$ where $\omega(G)$ is the clique number of G , i.e., the cardinality of a maximum complete induced subgraph.

Here is the outline of this paper. We first define *maximal geodetic closures* in Section 2 because it is a key concept in the algorithm of Harary et al. Then, in Section 3, we show by counter-examples that and why this algorithm may fail to find the geodetic number. Moreover, we indicate why the proposed approach itself cannot yield a correct algorithm. We then give, in Section 4, a 0–1 integer programming model for determining the geodetic number together with computational results. Brief conclusions are drawn in Section 5.

2 Maximal geodetic closures

To compute the geodetic number of a graph, Harary et al. [8] introduce the concept of *maximal geodetic closure* induced by a pair of nodes of V . A pair of nodes $\{x, y\}$ induces a maximal geodetic closure denoted by $MGC(x, y)$ if the geodetic closure of $\{x, y\}$ is maximal, i.e., it is not properly contained in any other geodetic closure of a pair of nodes.

On Figure 1 we can see, on the one hand, that $\{v_1, v_4\}$ does not induce a maximal geodetic closure because $I(\{v_1, v_4\}) = \{v_1, v_2, v_4\} \subsetneq I(\{v_1, v_3\}) = \{v_1, v_2, v_3, v_4, v_5\}$. On the other hand, $\{v_1, v_3\}$ gives a maximal geodetic closure and so we can write $MGC(v_1, v_3) = \{v_1, v_2, v_3, v_4, v_5\}$. All other pairs that induce a maximal geodetic closure are $\{v_1, v_3\}$, $\{v_1, v_6\}$, $\{v_1, v_8\}$, $\{v_4, v_5\}$, $\{v_3, v_6\}$ and their symmetrical counterpart starting with v_8 in place of v_1 .

In the cited article, Harary et al. propose a simple but redundant way to produce all maximal geodetic closures. They introduce the set $MGC(x)$ of all maximal geodetic closures containing node x . For example, on the graph of Figure 1, $MGC(v_3) =$

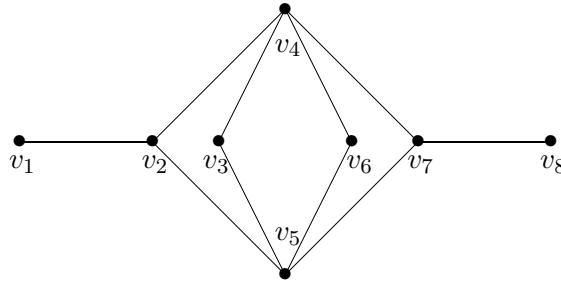


Figure 1: A graph G to illustrate the maximal geodetic closure concept.

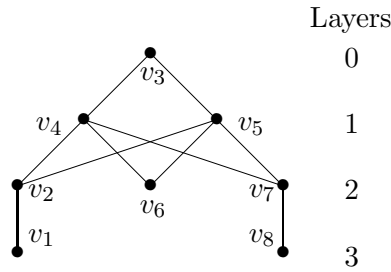


Figure 2: The layered graph $G(v_3)$ of the graph of Figure 1. $TV(v_3) = \{v_1, v_6, v_8\}$.

$\{I[v_1, v_3], I[v_3, v_8]\}$. To compute $MGC(x)$, they transform the initial graph G into an isomorphic layered graph $G(x)$. Only the node x is in layer 0 and every other node y at a distance $k > 0$ from x is placed in layer k for $k = 1, 2, \dots$. The edges of $G(x)$ are the same as the edges of G . Figure 2 shows the layered graph $G(v_3)$ rooted at node v_3 of the graph of Figure 1. The terminal nodes of $G(x)$ are denoted by $TV(x)$.

Harary et al. state that “ $MGC(x)$ is the set of $MGC(x, y)$ where y is a terminal node of $G(x)$ ” ([8] page 90). This is not always true because not all nodes x can be matched with another node y to induce a maximal geodetic closure. For example, on Figure 2, $\{v_3, v_6\}$ does not induce a maximal geodetic closure because $I[v_3, v_6] = \{v_3, v_4, v_5, v_6\} \subsetneq I[v_4, v_5] = \{v_2, v_3, v_4, v_5, v_6, v_7\}$. Figure 3 shows the layered graph $G(v_2)$ of the graph of Figure 1 but $\{v_2, v_8\}$ does not induce a maximal geodetic closure since $I[v_2, v_8] = \{v_2, v_4, v_5, v_7, v_8\} \subsetneq I[v_1, v_8] = \{v_1, v_2, v_4, v_5, v_7, v_8\}$.

If a pair $\{x, y\}$ of nodes induces a maximal geodetic closure then the two nodes x and y must be the end nodes of a maximal geodesic, otherwise one can extend the geodesic which implies that the pair $\{x, y\}$ is not inducing a maximal geodetic closure. This is the case for the pair $\{v_2, v_8\}$ of Figure 3 because geodesic $v_2 - v_8$ can be extended to the geodesic $v_1 - v_8$. For the same reason, $\{v_2, v_1\}$, $\{v_2, v_6\}$ and $\{v_2, v_3\}$ do not induce maximal geodetic closures. Note that maximality of the geodesic $x - y$ is a necessary but not sufficient condition to obtain a maximal geodetic closure. To find all maximal geodetic closures and them only, one can compute $G(x)$ for all $x \in V$ and then compare two by two

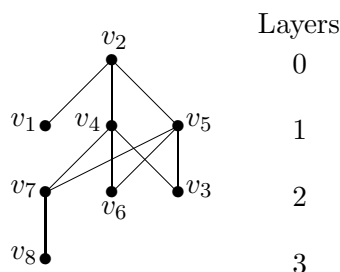


Figure 3: The layered graph $G(v_2)$ of the graph of the Figure 1.

the resulting geodetic closures of $\{x, y$ a terminal node of $G(x)\}$ for all x and y , keeping only maximal ones.

3 Counter-examples for the algorithm of Harary et al.

Further in the same article, Harary et al. state that “From the structure of $G(x)$, it follows that the geodetic closure of $\{x\} \cup TV(x)$ is the node set V of G ” ([8] page 91). In our example of Figure 3, $I[\{v_2\} \cup \{v_1, v_3, v_6, v_8\}] = V$. From there, they conclude hastily that “we can reduce the problem of determining $g(G)$ into that of finding a minimum dominating set in the bipartite graph $G = (V_1, V_2, A)$ ” ([8] page 91). This bipartite graph is constructed as follows. Let V_1 be the node set V of G . Take for V_2 the set of all node pairs $\{x, y\}$ which induce a maximal geodetic closure. Define a relation A between V_1 and V_2 that maps each node pair $\{x, y\}$ to all the nodes of the induced maximal geodetic closure:

$$A : V_2 \rightarrow V_1 : \{x, y\} \mapsto w \in MGC(x, y)$$

For each $x \in V$ there is an edge going from every element of $MGC(x)$ to x , i.e., each pair that induces a maximal geodetic closure is connected to each of the nodes of V it geodetically covers. To illustrate this bipartite graph concept, consider the small graph depicted on Figure 4.

There are only 4 pairs inducing maximal geodetic closures: $\{v_1, v_3\}$, $\{v_1, v_5\}$, $\{v_3, v_5\}$ and $\{v_2, v_4\}$. Since $I[\{v_2, v_4\}] = V$, the geodetic number for this graph is 2. Figure 5 shows the bipartite graph of the graph of Figure 4.

Extend the definition of the relation A to whatever subset of nodes of V_2 :

$$\forall S \subset V_2 : A(S) = \bigcup_{\{x, y\} \in S} MGC(x, y).$$

A subset $S \subset V_2$ is called a *dominating set* of the bipartite graph (V_1, V_2, A) if $A(S) = V_1 = V$. Harary et al. propose then “that the problem of determining the geodetic number of a graph is reduced to that of determining a smallest size subset of V_2 which dominates V_1 ” ([8] page 92). Unfortunately, this is not always true. Figure 6 depicts a counter-example.

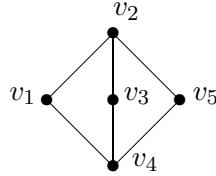


Figure 4: A small graph to illustrate the bipartite graph (V_1, V_2, A) concept.

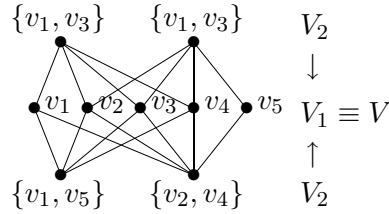


Figure 5: The bipartite graph of Figure 4.

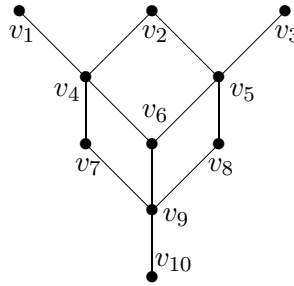


Figure 6: A counter-example for the algorithm of Harary et al.

There are 6 pairs that induce maximal geodetic closures: $\{v_1, v_3\}$, $\{v_1, v_{10}\}$, $\{v_3, v_{10}\}$, $\{v_1, v_8\}$, $\{v_3, v_7\}$ and $\{v_2, v_{10}\}$. The corresponding bipartite graph (V_1, V_2, A) is represented on Figure 7.

We have a subset of size 2 of V_2 that dominates V_1 : $S = \{\{v_2, v_{10}\}, \{v_1, v_3\}\}$ and because none of the elements of V_2 alone dominates V_1 , this is an optimal solution following the algorithm of Harary et al. There are two more “optimal” solutions: $\{\{v_1, v_8\}, \{v_3, v_{10}\}\}$ and $\{\{v_3, v_7\}, \{v_1, v_{10}\}\}$. All of them consist in a solution of 4 different nodes for a minimum geodetic set and so, following the algorithm of Harary et al. we have $g(G) = 4$. But $S = \{v_1, v_3, v_{10}\}$ also covers geodetically V and has only 3 elements which shows that $g(G) \leq 3$.

One might object that a modified version of the algorithm of Harary et al. would have considered $\{\{v_1, v_3\}, \{v_1, v_{10}\}, \{v_3, v_{10}\}\}$ and would have found the solution $\{v_1, v_3, v_{10}\}$. But as the counter-example on Figure 8 shows, the maximal geodetic closure approach cannot in itself lead to a correct algorithm.

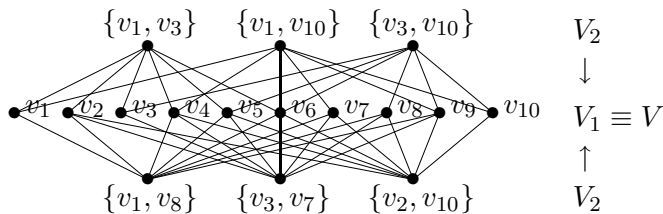
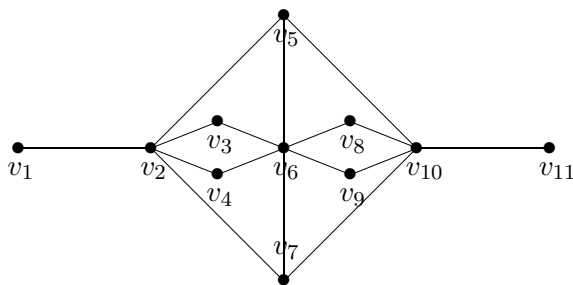
Figure 7: Bipartite graph (V_1, V_2, A) for the graph of Figure 6.

Figure 8: Another counter-example. One cannot only consider pairs that induce maximal geodetic sets.

Node v_6 does not belong to any pair that induces a maximal geodetic closure. To see this, we only need to consider the pairs $\{v, v_6\}$ with $v = v_1, v_2, v_3$ and v_5 by symmetry. In the three last cases, it is easy to see that $\{v, v_6\}$ does not induce a maximal geodetic closure. For the first one, $I[\{v_1, v_6\}] = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\} \subsetneq I[\{v_1, v_8\}] = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_{10}\}$.

Nevertheless we need node v_6 to construct the smallest geodetic set, i.e., $S = \{v_1, v_6, v_{11}\}$ with $g(G) = 3$.

4 A 0-1 integer programming model to find $g(G)$

We have developed a 0-1 integer programming model to find the geodetic number of a graph. To do this, we simply construct a minimum geodetic set S . For each node $v_k \in V$, let us characterize all the geodesics going through that node v_k by collecting their pairs of initial and end nodes in a set P_k :

$$P_k = \{(v_i, v_j) \mid d(v_i, v_k) + d(v_k, v_j) = d(v_i, v_j)\}.$$

The model is the following:

$$\min z = \sum_{k=1}^n x_k$$

subject to

$$\forall v_k \quad 1 - x_k \leq \sum_{(v_i, v_j) \in P_k} y_{ij} \quad (i < j) \quad (1)$$

$$i = 1..n \quad y_{ij} \leq x_i \quad (i < j) \quad (2)$$

$$j = 1..n \quad y_{ij} \leq x_j \quad (i < j) \quad (3)$$

$$i, j = 1..n \quad x_i + x_j - 1 \leq y_{ij} \quad (i < j) \quad (4)$$

and all variables are binary variables.

The main decision variables are the Boolean indicators x_k for each node:

$$x_k = \begin{cases} 1 & \text{if } v_k \in S \\ 0 & \text{if } v_k \notin S. \end{cases}$$

They are used in combination with secondary variables that are completely determined when the x_k are chosen:

$$y_{ij} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \in S \\ 0 & \text{otherwise.} \end{cases}$$

We can restrict ourselves to $i < j$.

In the objective function we want to minimize the number of elements in the geodetic set. Inequality (1) ensures that all nodes of V are covered in a geodetic sense by the subset S . Inequalities (2), (3) and (4) force y_{ij} to be 1 or 0 depending on the values of x_i and x_j .

Table 1 shows the performance of CPLEX 7.1 (<http://www.ilog.com>) in solving the model on a Ultra Sparc III with a 1,2 Ghz processor and 2 Gb of RAM. For each n , 90 graphs were randomly generated, i.e., 15 graphs for each size of 20%, 40%, 50%, 60%, 80%, 95% of the maximum number of edges. The computational results were compiled as follow: the average ($\mathbf{t_ave}$), the maximum ($\mathbf{t_max}$) and minimum ($\mathbf{t_min}$) computing time of CPLEX 7.1 are given in the second column and the average ($\mathbf{z_ave}$), the maximum ($\mathbf{z_max}$) and minimum ($\mathbf{z_min}$) of the objective function in the third column. Times do not include the fraction of a second necessary to generate the model. All instances are solved to optimality. One can see that augmenting the number of nodes by 5 multiplies approximatively the computing time by 5. This illustrates the fact that the problem of finding the geodetic number is NP-Hard. No tuning has been made to improve the results.

Table 2 presents the case $n = 30$ in more detail. One interesting result that we could observe in all cases is that the geodetic number is dropping down on average with the augmentation of the number of edges. Even with 95% of the edges ($m = 413$), the average geodetic number is 2.6, i.e., close to 3 and the maximum geodetic number seen in this experiment was 3 ! But it is known ([8]) that a complete graph has all its nodes in a geodetic set and so $g(K_n) = n$.

Table 1: Results from the processing of the 0-1 integer programming model using CPLEX. For each n , 90 graphs (15 graphs with 20%, 40%, 50%, 60%, 80%, 95% of the maximum number of edges) were randomly generated. In the second column, the time is measured in seconds.

n	$m_{\min} \leq m \leq m_{\max}$	t_ave	t_max	t_min	z_ave	z_max	z_min
10	$9 \leq m \leq 45$	0.1465	0.25	0.02	3.8	7	2
15	$14 \leq m \leq 105$	1.2295	2.38	0.09	4.0113	7	2
20	$19 \leq m \leq 190$	6.3875	16.34	0.2	4.089	7	2
25	$24 \leq m \leq 300$	29.0235	143.74	0.58	4.4333	8	2
30	$29 \leq m \leq 435$	99.2343	341.51	1.42	4.6778	7	2
35	$34 \leq m \leq 595$	433.3908	2136.8	2.61	4.6888	8	2
40	$39 \leq m \leq 780$	1939.0728	12901.46	15.56	4.9333	8	2

Table 2: Results from the processing of the 0-1 integer programming model using CPLEX on 90 graphs with $n = 30$. Each line is representing the compilation of data for 15 randomly generated graphs corresponding respectively to 20%, 40%, 50%, 60%, 80%, 95% of the maximum number of edges. In the second column, the time is measured in seconds.

$29 \leq m \leq 435$	t_ave	t_max	t_min	z_ave	z_max	z_min
87	163.933	341.51	72.94	6.4	7	5
174	145.627	240.24	80.69	5.867	6	5
217	157.629	272.1	72.09	5.267	6	5
261	87.433	140.89	31.42	4.6	5	4
348	27.279	39.63	14.11	3.333	4	3
413	13.505	23.49	1.42	2.6	3	2

5 Conclusions

Counter-examples are provided for the only algorithm proposed in the literature for computing the geodetic number of a graph. Moreover, it is shown that maximal geodetic sets do not suffice to build a correct algorithm. A 0 – 1 integer programming model is proposed, which can solve to optimality, using CPLEX 7.1, instances with up to 40 nodes for all densities. If the density is very small or very large, a few further experiments show that instances with $n \leq 70$ can also be solved to optimality in moderate computing time.

This last approach generalizes in a straightforward way to determination of the edge geodetic number and the directed geodetic number.

References

- [1] M. Atici. Computational complexity of geodetic set. *International Journal of Computer Mathematics*, 79(5):587–591, 2002.
- [2] M. Atici. On the edge geodetic number of a graph. *International Journal of Computer Mathematics*, 80(7):853–861, 2003.
- [3] M. Atici and A. Vince. Geodesics in graphs, an extremal set problem, and perfect hash families. *Graphs and Combinatorics*, 18:403–413, 2002.
- [4] F. Buckley and F. Harary. *Distance in graphs*. Addison-Wesley, Redwood City, 1990.
- [5] G. J. Chang, L.-D. Tong, and H.-T. Wang. Geodetic spectra of graphs. *European Journal of Combinatorics*, 25:383–391, 2004.
- [6] G. Chartrand, F. Harary, and P. Zang. On the geodetic number of a graph. *Networks*, 39(1), 2002.
- [7] G. Chartrand and P. Zang. The geodetic number of an oriented graph. *European Journal of Combinatorics*, 21:181–189, 2000.
- [8] F. Harary, E. Loukakis, and C. Tsouros. The geodetic number of a graph. *Mathematical and Computer Modelling*, 17:89–95, June 1993..
- [9] C. Hernando, T. Jiang, M. Mora, I. M. Pelayo, and C. Seara. On the steiner, geodetic and hull numbers of graphs. *Discrete Mathematics*, 293:139–154, 2005.
- [10] Pelayo I. On convexity in graphs. <http://personal.telefonica.terra.es/web/ipm/research/Definitions.pdf>, October 2003.