Alternating paths*

M. Abellanas*, A. García**, F. Hurtado*** and J. Tejel**
* mabellanas@fi.upm.es
Universidad Politécnica de Madrid
Facultad de Informática. Departamento de Matemática Aplicada
28660—Boadilla del Monte (Madrid, España)
** {olaverri,jtejel}@posta.unizar.es
Universidad de Zaragoza
Fac. Ciencias. Departamento de Métodos Estadísticos
Pedro Cerbuna, 12; 50009—Zaragoza (España)
*** ferran.hurtado@upc.es
Universitat Politècnica de Catalunya
Fac. de Matematiques i Estadistica. Dept. de Matem. Aplicada II
Pau Gargallo, 5; 08028—Barcelona (España)

Abstract

Given two disjoint point sets R y B in the plane (the *red* points and the *blue* points, respectively), an *alternating path* is a non-crossing path of line segments visiting alternatively red and blue points. In this work we study the properties of the longest such paths and give bounds on their length. In particular, we prove that for points in convex position there are configurations of points such that the longest alternating path has length less or equal than $(4/3 - (4/3)\sqrt{3a^2 - 3a + 1})|R \cup B|$, where $a = |R|/|R \cup B|$.

Keywords: Geometric graphs, Hamiltonian paths, matchings.

1 Introduction

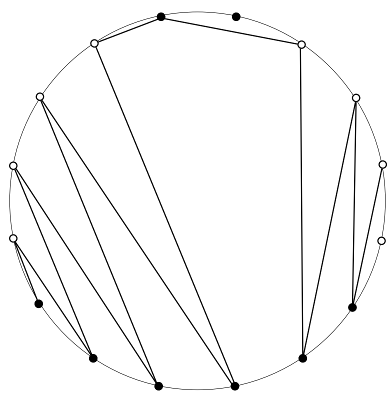

Figure 1: A configuration without any alternating Hamiltonian path.

*Work partially supported by MCyT TICC02-4486-C02-1, CAM 07T/0014/2001, MEC-DGES-SEUID PB98-0933, MCYT-FEDER BFM2002-0557 y Gen. Cat 2001SGR00224.

Let R and B be two disjoint sets of points on the plane in general position, whose points will be called *red* points and *blue* points, respectively, and such that $|R| \leq |B|$. An *alternating path* is a sequence of vertices $v_1, \ldots, v_h$ alternatively red and blue, such that the set of segments connecting vertices consecutive in the path is non-crossing. It is not always possible achieve alternating paths that visit all the points, as can be seen in the configuration shown in Figure 1. When crossings are allowed, Kaneko and Kano prove in [1] that one can always find alternating Hamiltonian non-crossing paths with no more than $n-1$ crossings when $|R| = |B| = n$.

A special case happens when the sets R and B are separable by a line and, additionally, $|R| = |B| = n$. In [2], Abellanas et al. proved that in this case there is always an alternating path through all the points in $R \cup B$.

The study of alternating paths was considered in [3], where Akiyama and Urutia give an algorithm for deciding whether a set of n blue points and n red points in convex position admits or not a Hamiltonian alternating path. In this paper we consider bounds on the length of the longest alternating path. We focus especially on point sets in convex position, and prove that there are configurations of points such that the longest alternating path has length less or equal than $(4/3 - (4/3)\sqrt{3a^2 - 3a + 1})|R \cup B|$, where $a = |R|/|R \cup B|$.

2 General sets

From the algorithm in [2] the two following results are easily derived:

Lemma 1 *If $|R| = |B| = n$ there is an alternating path with length greater or equal than n starting at any given vertex of the convex hull $CH(R \cup B)$.*

Proof. Simply observe than through any vertex v of $CH(R \cup B)$ there is a line splitting the whole set into two nearly-equal parts, thus obtaining at least sets of $n/2$ points of each color which are line separable. □

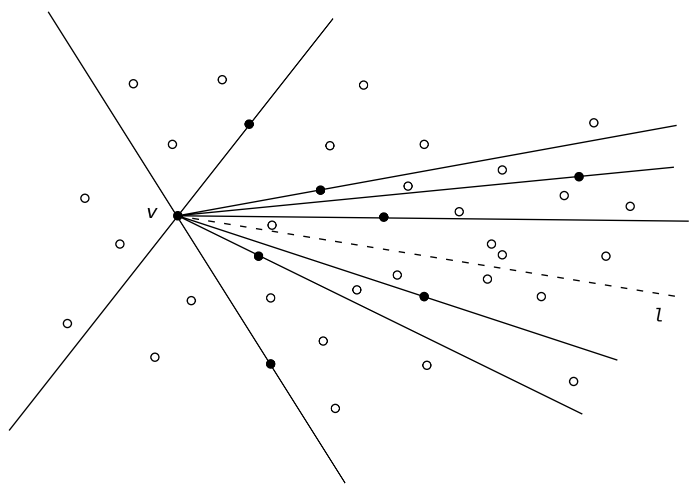

Figure 2: Dividing the plane into n regions.

Lemma 2 *If $|R| = n$ and $|B| \geq n^2 - 2n + 1$, then there is an alternating path visiting all the points in R.*

Proof. Let v be any vertex of $CH(R)$. If we take v as apex of rays through all other points in R we obtain $n-1$ regions, all convex but one. We can consider the non-convex region as union of the two halfplanes obtained by extending the limiting rays (see Figure 2). As we have n regions one of them has to contain at least $\lceil (n^2 - 2n + 1)/n \rceil \geq n - 1$ blue points. If the region is one of the halfplanes the algorithm in [2] gives the solution. If that region lies between two consecutive rays from v, leaving i points from R to the left and $n - i - 1$ to the right, we can draw a new ray l from v splitting the bluer points in the region into at least i to the left and at least $n - i - 1$ to the right. Again the technique from [2] applied to the halfplanes defined by l, starting at v, gives the path we were looking for. □

On the other hand, if we have points in convex position whose counterclockwise distribution is $\frac{n}{2} + 2$ red points, $\frac{n}{2}$ blue points, 2 red points, $\frac{n}{2}$ blue points, 2 red points, $\frac{n}{2}$ blue points, ..., 2 red points, $\frac{n}{2}$ blue points, we obtain a configuration that does not admit any alternating path visiting the n points in R, and we have $|B| = \frac{n^2}{8}$.

This improves on the results obtained by Kaneko and Kano in [4], where they show that there is a configuration of points without alternating path visiting the n points in R, with $|B| = \frac{n^2}{16} + \frac{n}{2} - 1$, and that for the existence of an alternating path visiting all the points in R at least $2n^2 - 2n - 3$ blue points are required.

3 Points in convex position

One aspect of the interest of studying configurations of points in convex position is that, apparently, they achieve the worst possible case, i. e., the situation in which the the longest alternating path is shortest. Obviously alternating paths for points in convex position do not depend on the exact position of the points but only in the cyclic sequences $r_1, b_1, r_2, b_2, \ldots r_k, b_k$, indication that the configuration has r_1 red points,followed by b_1 blue points, followed by r_2 red points, and so on.

For points in convex position the following algorithm, based on the dynamic programming paradigm, gives the longest alternating path..

Let us denote by $v_1, \ldots, v_N$ the N points of $R \cup B$ described clockwise, where v_1 has been picked arbitrarily. Let us define the rank $[v_i, v_j]$ to be the point set $v_i, v_{i+1}, \ldots, v_j$, where we suppose that $v_{N+1} = v_1$. We denote by $l^+(i, j)$ and $l^-(i, j)$ the length of the longest alternating path using only vertices from the rank $[v_i, v_j]$, according to whether we start at v_i or at v_j, respectively.

The algorithm computes all the values $l^+(i, j)$ y $l^-(i, j)$, in increasing order of the number points in the rank $[v_i, v_j]$. For ranks with only one point (i.e., when $i = j$) $l^+(i, j)$ and $l^-(i, j)$ are obviously 0. For greater ranks of points:

$$l^+(i,j) = \begin{cases} \max(l^+(i+1,j), l^-(i+1,j)) & v_i, v_{i+1}, v_j \text{ same color} \\ l^-(i+1,j) + 1 & v_j \text{ color different from } v_i, v_{i+1} \\ l^+(i+1,j) + 1 & v_{i+1} \text{ color different from } v_i, v_j \\ \max(l^+(i+1,j), l^-(i+1,j)) + 1 & v_i \text{ color different from } v_{i+1}, v_j \end{cases}$$

and $l^-(i, j)$ is obtained with similar formulas. The maximum among $l^+(i, j), l^-(i, j)$ for ranks with N points will give the length of the longest alternating path. The exact sequence of points is obtained in the usual way of this approach, using the preceding formulas. The complexity of the algorithm is $\Theta(N^2)$.

Let us now introduce what we call a *linear matching*: a set of disjoint segments, each one connecting one red point with one blue point, such that there exists a line crossing all the segments..

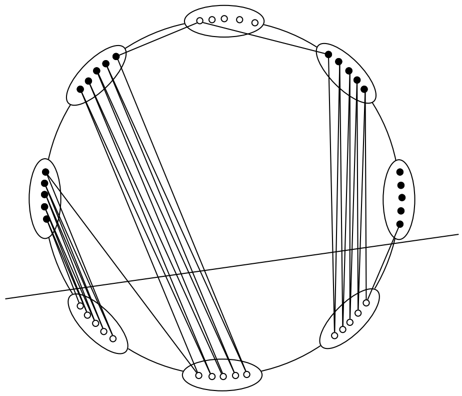

Figure 3: Relation between alternating paths and linear matchings.

If there are m segments in the linear matching, we say that *it covers* $2m$ *points.* In Figures 1 and 4 we show configurations that do no admit linear matchings covering all points in R. The largest linear matchings for points in convex position can easily found using an extension of the preceding algorithm which requires time $\Theta(N^2)$.

In general, finding properties of the maximum linear matching is somehow simpler than studying similar properties of the longest alternating path.Moreover, for points in convex position, finding bounds on the size of the maximum linear matching happens to be equivalent to finding bounds for the longest alternating, in the sense that we precise in the following lemma.

Lemma 3 *Given N_1 red points and $N_2 \geq N_1$ blue points in convex position, let us consider the values $N = N_1 + N_2$ and $a = N_1/(N_1 + N_2)$. Suppose that there is a configuration of points where the maximum linear matching covers less than $\alpha a N$ points; then there exists a configuration with N' points, with the same proportion a of red points, where the longest alternating path covers $\leq \alpha a N'$ points. Reversely, if there is a configuration with an alternating path covering less than $\alpha a N$ points, then there is another configuration with N'points and the same proportion of red points where the maximum linear matching covers less than $\alpha a N'$ points.*

Proof. The idea of the proof consists of replacing every point in the configuration by $r > 0$ copies of itself, and then check that the longest alternating path of the new configuration is essentially twice the size of the maximum linear matching, up to a constant (see Figure 3). □

4 Extremal configurations in convex position

Assume that $|R| = |B| = n$ and that $n = 6k-4$. Let us consider the following symmetrical configuration of points: v_1 is red, then come $k-2$ points $v_2, \ldots, v_{k-1}$ which are alternatively blue and red (in this way v_{k-1} is red when k is even, and blue when k is odd); in the same way blue and red alternate among the $k-2$ vertices $v_{6k-4}, \ldots, v_{5k-1}$ preceding v_1. The vertex v_{k-1} is followed by a group of k vertices $v_k, \ldots, v_{2k-1}$ with color opposite to the color of v_{k-1} (thus blue when k is even); similarly before v_{5k-1} there is another group of

k vertices, $v_{5k-2}, \ldots, v_{4k-1}$. Finally, the remaining $2k-1$ points $v_{2k}, \ldots, v_{4k-2}$ have color opposite to the color of v_{2k-1} (red when k is even). This configurations, for the values $k = 2, 4$, are shown in Figure 4.

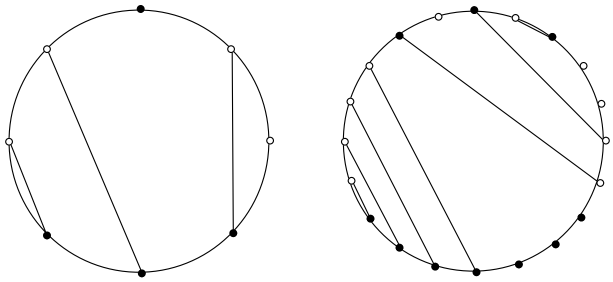

Figure 4: Maximal linear matchings.

If we call *zone* any maximal sequence of consecutive points having the same color, our figure has $2k$ zones: one of size $2k-1$, two of size k, and $2k-3$ of size 1. Our next result gives the maximum linear matching for these figures.

Lemma 4 *The maximum linear matching starting with the edge (v_i, v_{i+1}) with $1 \leq i \leq v_k$, or starting with the edge (v_{2k-1}, v_{2k}) contains exactly $2k-1$ edges, and a maximum matching can be achieved starting that way, the last edge being included is (v_{4k-2}, v_{4k-1}).By symmetry, the maximum linear matching starting at the points (v_i, v_{i+1}) with $5k-1 \leq i \leq 6k-4 = n$, or starting at (v_{4k-2}, v_{4k-1}) contains $2k-1$ edges, and can be finished with the edge (v_{2k-1}, v_{2k}).*

Proof. By case analysis. The key observation is that once the edge (v_i, v_j) is included in a maximal matching with a given start, one must include either the edge (v_{i+1}, v_{j_1}) (where j_1 indicates the first vertex preceding v_j with color different to the color of v_{i+1}) or the edge (v_{i_1}, v_{j-1}) (where v_{i_1} indicates the first vertex following v_i with color different to the color of v_{j-1}). Therefore, if v_{i+1} and v_{j-1} have different color, the edge joining them must be added to the maximum matching. □

The number of vertices covered by a maximum linear matching in our figure, or in the one obtained by replacing every point by r copies of itself, is $(4k-2)/(6k-4)$ multiplied by the number of points. This value, $(4k-2)/(6k-4)$ (which has limit $\frac{2}{3}$ when $k \to \infty$), is $1, 3/4, 5/7, 7/10, \ldots$, when $k = 1, 2, 3, 4, \ldots$, and the number of zones for the corresponding figures is $2k$. It is straightforward that in any configuration with only two zones ($k = 1$ in our configuration), there is a linear matching covering all the points. It is easy to see that for every configuration with only 4 zones there is linear matching covering 3/4 of the points. Similarly, an analysis quite long-winded of cases proves that for 6 zones ($k = 3$), there are linear matchings covering 5/7 of the points.

We conjecture that for points in convex position this configuration achieves the smallest maximum linear matching. More precisely, we believe that for any configuration of $N = 2n$ points in convex position, n of them colored red and n of them colored blue, there is a linear matching having at least $\lceil N/3 \rceil$ edges, i.e., covering $\lceil 2N/3 \rceil$ vertices. The preceding example shows that for every E $\alpha > (1/3)$, there are configurations where the maximum linear matching has less than αN edges.

5 Continuous models

We have seen that a maximum linear matching lineal doesn't change essentially if every point is replaced by r copies, and that the configuration is defined by the number of points $r_1, b_1, \ldots, r_k, b_k$ in each zone. We have also seen that what matters us is the proportion of points covered by the maximum linear matching.

We call *continuous configuration* with $2k$ zones a cyclic sequence $r_1, b_1, r_2, b_2, \ldots, r_k, b_k$ of real numbers in the rank $[0, 1]$, adding up to 1, where $a = \sum_{i=1}^{k} r_i \leq 1/2$.

Let $f^2(r_1, b_1) = \min(r_1, b_1)$ and $f^{2k}(r_1, b_1, \ldots, r_k, b_k)$ be a function from R^{2k} to R^+ recursively defined as follows:

$$f^{2k}(r_1,b_1,\ldots,r_k,b_k)=\begin{cases} b_1+\max(f^{2(k-1)}(b_k,r_2,b_2,\ldots,r_k),f^{2(k-1)}(r_1-b_1,b_2,r_3,\ldots,b_k)) & if \quad r_1\geq b_1 \\ r_1+\max(f^{2(k-1)}(b_k,r_2,b_2,\ldots,r_k),f^{2(k-1)}(r_k,b_1-r_1,r_2,b_2,\ldots,b_{k-1})) & if \quad r_1\leq b_1 \end{cases}$$

Notice that were $r_1, b_1, \ldots, r_k, b_k$ integer numbers describing a convex configuration of points $R \cup B$, the preceding formula gives the maximum number of edges in a linear matching starting with the edge limiting region r_1 with region b_1, therefore: $\bar{f}^{2k}(r_1,b_1,\ldots,r_k,b_k)=\max(f^{2k}(r_1,b_1,\ldots,r_k,b_k),f^{2k}(b_1,r_2,b_2,\ldots,b_k,r_1),\ldots,f^{2k}(b_k,r_1,b_1,\ldots,r_k))$ gives the maximum number of edges in a linear matching. The number of points that can be covered is $2\bar{f}^{2k}(r_1, b_1, \ldots, r_k, b_k)$, and the number of red points left uncovered is $|R| - \bar{f}^{2k}(r_1, b_1, \ldots, r_k, b_k)$. For a continuous configuration we call *loss* the size of uncovered red zone, which is $a - \bar{f}^{2k}(r_1, b_1, \ldots, r_k, b_k)$.

For fixed k, it can be proved by induction that the functions f^{2k} and $\bar{f}^{2k}$ are continuous and piecewise linear (in the variables $r_1, b_1, \ldots, r_k, b_k$). Moreover, in the formulas defining the functions, the coefficients can only be 0 or 1, i.e., $\bar{f}^{2k}(r_1, b_1, \ldots, r_k, b_k) = \alpha_1^i r_1 + \beta_1^i b_1 + \ldots + \alpha_k^i r_k + \beta_k^i b_k$ when $(r_1, b_1, \ldots, r_k, b_k)$ belongs to some region R^i, where the coefficients α_j^i and β_j^i can only be 0 or 1.

We are interested in finding values $r_1, b_1, \ldots, r_k, b_k$ such that $\bar{f}^{2k}$ is minimized, with the additional conditions $0 \leq r_i, b_i \leq 1, \forall i$, $\sum r_i = a \leq \sum b_i = 1 - a$, and we have special interest in the case $a = 1/2$. We denote by $f(a, k)$ such minimum value, and by $p(a, k) = a - f(a, k)$ the maximum loss.

The function $\bar{f}^{2k}$ is piecewise linear in a compact domain, hence it will reach a minimum in the domain. Moreover, it can be proved that the minimum is reached in an interior point. As $\bar{f}^{2k}$ is piecewise linear, it has to be in a point of the domain where the linear formulas that give the value of f^{2k} coincide. Therefore, once a is fixed, given the fact that there are $2k - 2$ free parameters, the point in the domain must satisfy $2k - 1$ equalities having the form $\alpha_1^{i_1} r_1 + \beta_1^{i_1} b_1 + \ldots + \alpha_k^{i_1} r_k + \beta_k^{i_1} b_k = \ldots = \alpha_1^{i_{2k-1}} r_1 + \beta_1^{i_{2k-1}} b_1 + \ldots + \alpha_k^{i_{2k-1}} r_k + \beta_k^{i_{2k-1}} b_k$. These equalities determine which values of r_j, b_j give the minimum value. This observations lead to the following lemma.

Lemma 5 *For fixed k, the function $f(a, k)$ is continuous, concave and piecewise linear in the variable a. Moreover, its values is a, if* $0 \leq a \leq 1/(k + 1)$.

Proof. As the number of possible systems of equations we have is finite, we have a finite number of points in the domain, and $f(a, k)$ is the smallest value of $\bar{f}^{2k}$ in these points. Therefore, fixed k, we see that $f(a, k)$ is also a piecewise linear function, which is concave because it is the maximum of other functions.

When $a \leq 1/(k + 1)$, the biggest b_i is $\geq a$, hence all the red zones can be "inserted" in a zone b_i, giving loss 0. □

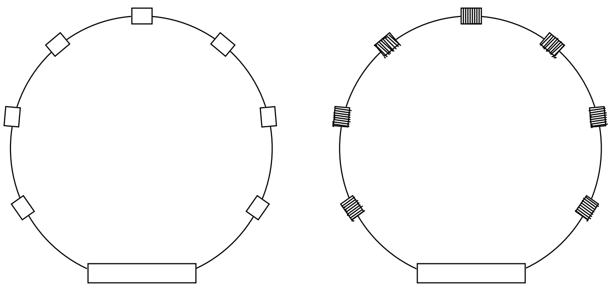

Figure 5: To the left, the first configuration achieving loss when $a > \frac{1}{k+1}$. To the right, an example of configuration having loss $\epsilon = a - 2/3 + (2/3)\sqrt{3a^2 - 3a + 1}$.

Let us observe that $p(a, k)$ is piecewise linear, convex, and its value is 0 if $a \leq 1/(k+1)$. Analyzing the points of the domain where $\bar{f}^{2k}$ is minimized, we find $p(a, 2)$ and $p(a, 3)$. The resulting formulas are as follows:

$$p(a,2) = \begin{cases} 0 & \text{si } a < \frac{1}{3} \\ \frac{(3a-1)}{4} & \text{si } \frac{1}{3} \leq a \leq \frac{1}{2} \end{cases} \qquad p(a,3) = \begin{cases} 0 & \text{si } a < \frac{1}{4} \\ \frac{(4a-1)}{9} & \text{si } \frac{1}{4} \leq a \leq \frac{11}{26} \\ \frac{(6a-2)}{7} & \text{si } \frac{11}{26} \leq a \leq \frac{1}{2} \end{cases}$$

Our next lemma tells us which is the behavior of $p(a, k)$ for values of a slightly greater than $1/(k + 1)$.

Lemma 6 *If $1/(k + 1) \leq a \leq 1/k$, the maximum loss is obtained when $b_1 = b_2 = \ldots b_k = (1 - a)/k$, $r_1 = b_1 + \epsilon$, $r_2 = \ldots = r_k = \epsilon$, with $\epsilon = ((k + 1)a - 1)/k^2$. Moreover, the loss is ϵ.*

Proof. Given any configuration such that $1/(k + 1) \leq a \leq 1/k$, the proof is based in checking that any two zones b_i and b_j are sufficient for covering all the red zone. In this situation, a case analysis shows that the configuration given in the lemma is the one that maximizes the loss function. □

The configuration of the lemma is shown in the left part of Figure 5. The red zones have been represented by rectangles.

For any given a, we will be interested in the value of $p(a, k)$, when k tends to infinite. Therefore, for fixed a, we can compute the loss in the preceding configuration for different values of k satisfying $a \geq 1/(k + 1)$. The maximum is achieved for $k = 2(1 - a)/a$ and the loss is $a^2/(4(1 - a))$.

Nevertheless, for fixed a, there is a family configurations that in the limit produce a greater loss. The configuration is obtained form the preceding configuration, assuming that each of the regions $r_2, \ldots, r_k$ is subdivided into r regions of size r_i/r, and that between any two subregions we insert blue zones having size r_i/r too. Thus we have $2(1 + (k - 1)r)$ zones: k big blue zones, say of size b_1, plus $(r - 1)(k - 1)$ small blue zones of size ϵ/r, plus $r(k - 1)$ small red zones of size ϵ/r, and one big red zone (sayr_1), with size $b_1 + \epsilon$. It must happen that $b_1 + k\epsilon = a$ and $kb_1 + (r - 1)(k - 1)(\epsilon/r) = 1 - a$ and again the loss of this figure is ϵ. A configuration of this type is shown in figure 5.

When r tends to infinite, the second formula becomes $kb_1 + (k - 1)\epsilon = a$, and therefore the loss is $\epsilon = ((k + 1)a - 1)/(k^2 - k + 1)$ with $b_1 = (a + k(1 - 2a))/(k^2 - k + 1)$.

If we fix a and look for the value of k which minimizes the loss, we obtain $k = (1 - 2a)/(1 - a - \sqrt{3a^2 - 3a + 1})$, and the corresponding loss is $\epsilon = a - 2/3 + (2/3)\sqrt{3a^2 - 3a + 1}$.

As this continuous model can be approximated with a discrete model of points, we deduce the following result:

Lemma 7 *Given any proportion $a \leq (1/2)$ of red points, for every α such $0 < 2/3 - (2/3)\sqrt{3a^2 - 3a + 1} < \alpha$ there are configurations of red and blue points in such proportion, where the size of any maximum linear matching is less than $\alpha|R \cup B|$; equivalently, any alternating path has length less than $2\alpha|R \cup B|$.*

Acknowledgement

The authors are grateful to Jesús García, Gregorio Hernández, Marc Noy, Pedro Ramos and Jorge Urrutia for helpful discussions at the early stages of this work, as well as for providing pointers to related problems.

References

[1] A. Kaneko, M. Kano, K. Yoshimoto, "Alternating Hamilton cycles with minimum number of crossings in the plane", *International Journal of Computational Geometry and Applications*, Vol. 10, N. 1, pp. 73-78, 2000.

[2] M. Abellanas, J. García, G. Hernández, M. Noy, P. Ramos, "Bipartite embeddings of trees in the plane", *Discrete Applied Mathematics*, Vol. 93, pp. 141-148, 1999.

[3] J. Akiyama, J. Urrutia, "Simple alternating path problem", *Discrete Mathematics*, Vol. 84, pp. 101-103, 1990.

[4] A. Kaneko, M. Kano, "On paths in a Complete Bipartite Geometric Graph", *Lecture Notes in Computer Science*, 2098, pp. 187-191, 2001.