



HAL
open science

A new algorithm for graph center computation and graph partitioning according to the distance to the center

Frédéric Protin

► **To cite this version:**

Frédéric Protin. A new algorithm for graph center computation and graph partitioning according to the distance to the center. 2019. hal-02304090

HAL Id: hal-02304090

<https://hal.science/hal-02304090>

Preprint submitted on 2 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A new algorithm for graph center computation and graph partitioning according to the distance to the center

Frédéric Protin

2 October 2019

Abstract: We propose a new algorithm for finding the center of a graph, as well as the rank of each node in the hierarchy of distances to the center. In other words, our algorithm allows to partition the graph according to nodes distance to the center. Moreover, the algorithm is parallelizable. We compare the performances of our algorithm with the ones of Floyd-Warshall algorithm, which is traditionally used for these purposes. We show that, for a large variety of graphs, our algorithm outperforms the Floyd-Warshall algorithm.

1 Introduction

The *center* of a non-oriented graph, also called its Jordan center, is the set of all vertices of minimum eccentricity, that is, the set of all vertices for which the greatest distance to other vertices is minimal. Equivalently, it is the set of vertices with eccentricity equal to the graph's radius. Vertices belonging to the center are called central nodes. (See e.g. [5] for precise definitions).

Finding the center of a graph is useful e.g. in facility location problems, where the goal is to minimize the worst-case distance to the facility. For example, placing a fire station at a central point reduces the longest distance the fire truck has to travel. [9] showed that the center of a graph related to a rumor spread is a good estimator for the rumor source (see also [7]). We are also interested in partitioning the graph according to nodes distance to the center. This partitioning is useful, for example, for visualization or reporting, as showed in Figure 1 below, where the graph represents the cable network between electronic devices, as proposed in the Connect'it project of the enterprise Activus-group. Vincent Viton of Activus-group ([12]) suggested to use the nodes hierarchy in order to improve visualization.

For a non-oriented graph $G = (V, E)$, where V is the set of nodes and E the set of edges, Bellman-Ford algorithm [2, 8, 10] computes the distance from a fixed node to every other node. Its complexity is $O(|V||E|)$, where $|\cdot|$ denotes

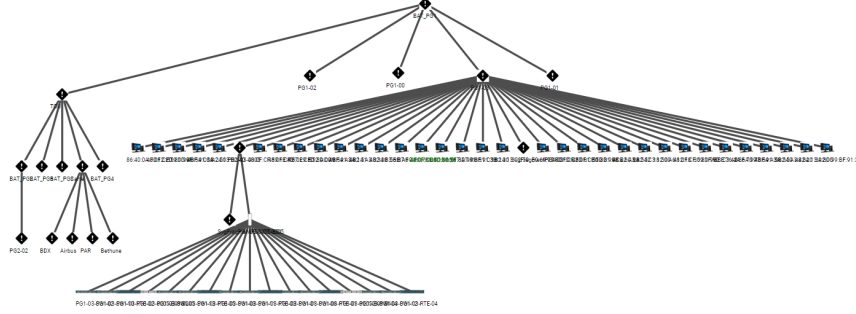


Figure 1: Application of nodes hierarchization of a graph to visualization. In this example the graph is a tree.

the cardinal of a set. Therefore, the complexity in the worse case, i.e. for a dense graph, is $O(|V|^3)$. Dijkstra's algorithm [3], applied to the same graph, returns in addition the paths from the fixed node to every other node. When applied to our context of interest, these algorithms return more than the required information, at the price of being slower than the algorithm we propose in the present work. Note that another algorithm was proposed in [6] for trees.

2 Principles of the new algorithm

Let $G = (V, E)$ be a connected graph, with V being the set of nodes (or vertices), and E the set of edges. Recall that the *center* of the graph G is the set of nodes that minimizes the greatest distance to the other nodes. Here we consider the distance $\text{dist}(\cdot, \cdot) : V \times V \rightarrow \mathbb{N}$, where the distance (i, j) between two nodes i and j is the smallest number of edges in a path joining them. Therefore, the center of G is the set

$$\mathcal{C}_0 := \left\{ i \in V : \max_j \text{dist}(i, j) = \min_i \max_j \text{dist}(i, j) \right\}.$$

Besides determining the center of a graph, we also want to obtain a partition of the graph according to nodes distance to the center :

$$G = \bigsqcup_{n \in \mathbb{N}} \left\{ i \in V : \max_j \text{dist}(i, j) = \min_i \max_j \text{dist}(i, j) + n \right\} =: \bigsqcup_{n \in \mathbb{N}} \mathcal{C}_n.$$

Let A denote the adjacency matrix of the graph G . We denote by \tilde{A} the matrix obtained from A by changing the diagonal coefficients into 1, i.e. $\tilde{A}_{ij} =$

A_{ij} for $i \neq j$, and $\tilde{A}_{ii} = 1$ for every i . In other words, \tilde{A} is the adjacency matrix of the graph \tilde{G} , obtained from G by imposing the fact that every node has a single edge connecting it to itself.

Note that, for every $i, j \in V$, the coefficient $(\tilde{A}^n)_{ij}$ of the n th power of the matrix \tilde{A} indicates the number of paths of length at most n going from node i to node j .

Note also the fact that, if for given $i, j \in V$, n is the smallest integer such that $(\tilde{A}^n)_{ij} \neq 0$, then the shortest path from i to j has length equal to n .

We deduce the following method for computing the center of the graph and for partitioning it.

The main idea of the algorithm

Compute the successive powers \tilde{A}^n of the matrix \tilde{A} , for $n \geq 1$, until at least one row gets filled with non-null coefficients. More precisely, let us denote by n_0 the smallest power such that the matrix \tilde{A}^{n_0} contains at least one row with no null coefficients:

$$n_0 := \min \left\{ n \geq 1 : \exists i \in V \text{ s.t. } (\tilde{A}^n)_{ij} \neq 0, \forall j \in V \right\}.$$

Note that the indexes of those rows in \tilde{A}^{n_0} correspond to the center nodes of the graph and hence form the set \mathcal{C}_0 :

$$\mathcal{C}_0 := \left\{ i \in V : (\tilde{A}^{n_0})_{ij} \neq 0, \forall j \in V \right\}.$$

In order to completely partition the graph and determine \mathcal{C}_k , for every $k \geq 1$, we continue computing successively the powers \tilde{A}^n , for $n = n_0 + k$, $k \geq 1$, until all the rows of \tilde{A}^n get filled with non-null coefficients.

Note that for any $k \geq 1$, the set \mathcal{C}_k is formed by the indexes of the rows having at least one null coefficient in \tilde{A}^{n_0+k-1} , but no null coefficients in \tilde{A}^{n_0+k} :

$$\mathcal{C}_k := \left\{ i \in V \setminus \mathcal{C}_{k-1} : (\tilde{A}^{n_0+k})_{ij} \neq 0, \forall j \in V \right\}.$$

Note that, since the graph is connected, we have $\mathcal{C}_k \neq \emptyset$ for $k = 0, \dots, r$, and $\mathcal{C}_k = \emptyset$ for $k > r$, where r is the radius of the graph.

An improved version

The method presented above solves the problem of finding the center of a graph and of partitioning the graph, but it is expensive in time. We present in the following few improvements which significantly diminish the computing time, resulting in a quite powerful algorithm.

Improvement 1. Note that the only important thing to know about the coefficients of \tilde{A}^n , when applying the previous method, is if they vanish or not.

Therefore, the method still works if the non-null coefficients in the matrix \tilde{A} , as well as in its powers, are replaced by 1. Let us formalize this remark, and show how we can use it in order to accelerate the previously described algorithm.

Denote by $M \mapsto R(M)$ the application which replaces each of the coefficients $m_{ij} \in \mathbb{N}$ of a matrix M by $\max(m_{ij}, 1)$. Note that this operation commutes with the matrix product. Consequently, if we want to compute $R(AB)$ for two matrices A and B with coefficients in \mathbb{N} , we can instead prefer to compute the product $R(A)R(B)$, which gives the same result and faster. Indeed, when we multiply a row (ℓ_1, \dots, ℓ_k) by a column $(c_1, \dots, c_k)^T$, we can stop the recursive computation of the scalar product as soon as $\ell_i = c_i = 1$ for some index i , and return the value 1.

Improvement 2. Since the coefficient $(\tilde{A}^n)_{ij}$ of \tilde{A}^n equals the number of paths of length at most n from node i to node j , the sequence $((\tilde{A}^n)_{ij})_n$ is increasing. Thus $(\tilde{A}^{n-1})_{ij} \geq 1$ entails $(\tilde{A}^n)_{ij} \geq 1$. It is therefore useless to compute a coefficient $(\tilde{A}^n)_{ij}$ when $(\tilde{A}^{n-1})_{ij} \geq 1$.

Improvement 3. A naive way to determine n_0 is to compute successively the terms of the sequence $(\tilde{A}^n)_n$, until one row gets filled with non-null coefficients. We propose here a faster way to determine n_0 , as follows.

We compute recursively the sequence of powers $\tilde{A}, \tilde{A}^2, \tilde{A}^4, \tilde{A}^{16}, \dots, \tilde{A}^{2^n}, \dots$, until at least one row gets filled with non-null coefficients, say for some \tilde{A}^{2^m} , with $m \geq 1$. We form a list with all these powers of \tilde{A} until $\tilde{A}^{2^{m-1}}$.

We then construct recursively a subset \mathcal{A} of $\{1, 2, \dots, m-1\}$, in the following manner. We first initialize $\mathcal{A} := \{m-1\}$. We then consider one by one the numbers $m-2, m-3, \dots, 2, 1$ in this order, and add to the current subset \mathcal{A} those numbers k which satisfy the fact that the product $\tilde{A}^{2^k} \cdot \prod_{\ell \in \mathcal{A}, \ell > k} \tilde{A}^{2^\ell}$ does

not contain any row filled with non-null coefficients.

It is straightforward that

$$\tilde{A}^{n_0-1} = \prod_{\ell \in \mathcal{A}} \tilde{A}^{2^\ell},$$

and $n_0 - 1 = \sum_{\ell \in \mathcal{A}} 2^\ell$, which is the unique binary decomposition of $n_0 - 1$.

Improvement 4. Since the graph is non oriented, the matrix \tilde{A}^n is symmetric and thus the computing time of \tilde{A}^n is reduced by half.

A detailed description of the final algorithm using the above improvements is given in the next section.

3 Description of the algorithm

In this section we present in detail the final algorithm, which takes into account the main ideas and the improvements previously presented.

Improvements 1, 2 and 4 lead to the following function (called *multiply*) for computing $R(M \cdot M2)$, the product of two matrices M and $M2$, composed with the application R described in Improvement 2, which replaces each coefficient by its maximum with 1. This function will be used for computing the powers A^n . The variable *width* represents the number of rows of the matrix M .

Algorithm 1: def multiply(M : matrix, M2 : matrix, width : int):

```

M3 ← M2
for ℓ ← 0 to width do
  for c ← ℓ + 1 to width do
    if M3[ℓ][c] = 0 then
      for i ← 0 to width do
        if M[ℓ][i] ≥ 1 and M2[i][c] ≥ 1 then
          M3[ℓ][c] = 1
          M3[c][ℓ] = 1
          break
        end
      end
    end
  end
end
end
return M3

```

We will also use the following variant (called *multiplyIns*) which performs the same multiplication as before, but also inserts the sets $\mathcal{C}_k, k \geq 0$ into a dictionary, as the algorithm progresses. The keys of this dictionary are the nodes of the graph, and the values are the corresponding distances to the center of the graph.

Algorithm 2: def multiplyIns(M : matrix, M2 : matrix, width : int, round : int, rows : list of int):

```

M3 ← M2
for ℓ ← 0 to width do
  for c ← ℓ + 1 to width do
    if M3[ℓ][c] = 0 then
      for i ← 0 to width do
        if M[ℓ][i] ≥ 1 and M2[i][c] ≥ 1 then
          M3[ℓ][c] = 1
          M3[c][ℓ] = 1
          /* rows[ℓ] is the number of non-null
             coefficients of the ℓth row of M3. */
          rows[ℓ] ← rows[ℓ] + 1
          rows[c] ← rows[c] + 1
          if rows[ℓ] = width then
            dictNodes[nodes[ℓ]] ← round ; // insert data.
          end
          if lines[c] = width then
            dictNodes[nodes[c]] ← round ; // insert data.
          end
        end
      end
    end
  end
end
return M3

```

The last function we use (called *testRow*) checks, for a given matrix M , if it contains any row filled with non-null coefficients.

Algorithm 3: def testRow(M : matrix, width : int):

```

Result: def testRow(M, width):
output ← FALSE
for ℓ ← 0 to width do
  s ← |{x ∈ M2[ℓ] : x ≠ 0}|; // Number of coefficients ≠ 0.
  if s = width then
    output ← TRUE
    break
  end
end
return output

```

The final algorithm (called *Partitioning*) is then the following. Here M_{Adj} denotes the adjacency matrix of the graph, and *width* its number of rows, which are given as input. Identity matrix of order n is denoted $Identity(n)$.

Algorithm 4: Partitioning(MAdjacency : matrix, width : int)

```

/* Initialization. */
M ← MAdjacency + Identity(width); // The matrixList  $\tilde{A}$ .
M2 ← M
matrixList ← [M]; // list initialized with the matrix M.
/* Computation of the list of successive powers  $\tilde{A}^{2^n}$ . */
for round ← 0 to width do
    product ← multiply(matrix[-1], matrix[-1], width)
    if (product, width); // Check if a row is filled.
    then
        | break
    end
    matrixList.add(product)
end
/* Computation of  $\tilde{A}^{n_0-1}$ . */
if length(matrixList) ≥ 2 then
    encode ← length(matrixList) - 1
    aux ← matrixList[-1]
    for round ← 0 to width do
        U ← multiply(aux, matrixList[encode-1], width)
        if not testRow(U, width); // Check if a row is filled.
        then
            | aux ← U
        end
        encode ← encode - 1
        if encode = 0 then
            | break
        end
    end
    M2 ← aux
end
/* Counting the number of non-null coefficients in each row */
for ℓ ← 0 to width do
    rows[ℓ] ← |{x ∈ M2[ℓ] : x ≠ 0}|; // Nb. of coefficients ≠ 0.
    /* If the matrix  $\tilde{A}$  contains a row without null
    coefficients, fill the dictionary with center nodes.
    */
    if rows[ℓ] = width then
        | dictNodes[nodes[ℓ]] = 0
    end
end
/* Computation of the complete hierarchy of nodes. */
for round ← 0 to width do
    if length(dictNodes) = width then
        | break
    end
    M2 = multiplyIns(M, M2, width, round, rows)
end
print(dictNodes); // print the final result

```

Scalability. Each step described in Section 1 is obviously parallelizable. The proposed algorithm is therefore parallelizable in a straightforward manner.

4 Experimental results and conclusions

An explicit derivation of the computational complexity of the algorithm is a quite tedious question, because it depends highly on the geometry of the graph.

We choose instead to compare experimentally the algorithm we propose with the Floyd-Warshall algorithm. Note that, by construction, the computing time of the algorithm we propose decreases when the connectivity of the graph increases, or equivalently, when the depth of the graph decreases. The worse case for the algorithm we propose is when the graph is linear. In this case, the Floyd-Warshall algorithm is faster than our new algorithm, but still the computing times are of the same order.

The experiments were performed using the library *Networkx* in Python. We generated random graphs with a given morphology (number of nodes (N), number of edges (NA) and depth (P)). Recall that the depth of a graph is the maximal distance from a node to the center.

In Table 1 we report an extract of our experiments. The computing times for the Floyd-Warshall algorithm and for the algorithm we propose are reported in seconds. Unfavorable cases for our proposed algorithm are highlighted in red, and improvements of a factor ≥ 10 for the computing time with respect to the Floyd-Warshall algorithm are highlighted in green.

Conclusion. The experiments performed suggest that the algorithm we propose is faster than the Floyd-Warshall algorithm, except for graphs with a high ratio P/N .

Note also that the Improvement 3 is optional, but beneficial. We have experimentally noted that, as we could expect, this improvement slows down the algorithm a little bit on shallow graphs, but accelerates it a lot on very deep graphs.

References

- [1] S. Warshall, A theorem on Boolean matrices, *Journal of the ACM* (1962), 9(1), pp. 11–12.
- [2] R. Bellman, On a routing problem, *Quarterly of Applied Mathematics* (1958), 16, pp. 87–90.
- [3] E.W. Dijkstra, A Short Introduction to the Art of Programming, Technische Hogeschool Eindhoven (1971), pp. 67–73.

Table 1: Comparison of computing time (in seconds) of the new algorithm and the Floyd-Warshall algorithm, for different graph morphologies.

heightGraph morphology	Floyd-Warshall	New algorithm
N= 500 , NA= 995, P = 15	189	114
N = 500 , NA = 1495, P = 17	198	113
N = 500 , NA = 1491, P = 14	191	97
N = 500 , NA = 1981, P = 14	195	82
N = 500 , NA = 2477, P = 14	201	89
N = 500 , NA = 2962 , P = 14	194	76
N= 500 , NA= 5365 , P = 13	186	62
N= 500 , NA= 14511, P = 13	196	55
N= 500 , NA= 22840 , P = 13	186	49
N= 500 , NA= 40824, P = 13	198	52
N= 500 , NA= 14632, P = 3	203	13
N= 500 , NA= 14515, P = 8	191	26
N= 500 , NA= 14569, P = 6	193	18
N= 500 , NA= 14556, P = 3	193	10
N= 500 , NA= 14550, P = 2	204	8
N= 500 , NA= 14553, P = 1	189	10
N= 500 , NA= 14402, P = 25	198	147
N= 500 , NA= 22563, P = 25	192	132
N= 500 , NA= 2964, P = 26	192	157
N= 500 , NA= 2963, P = 4	196	54
N= 200 , NA= 1161, P = 3	12	2
N= 200 , NA= 2915, P = 8	12	3
N= 200 , NA= 1473, P = 75	13	51
N= 400 , NA= 15075, P = 75	98	383
N= 400 , NA= 17412, P = 25	97	90
N= 400 , NA= 4247, P = 26	99	85
N= 400 , NA= 4276, P = 16	97	43
N= 400 , NA= 30467, P = 15	97	48
N= 400 , NA= 31636, P = 2	100	1
N= 400 , NA= 31329, P = 6	100	10
N= 400 , NA= 17807, P = 7	98	12
N= 400 , NA= 7963, P = 7	100	13
N= 400 , NA= 7964, P = 3	102	6
N= 400 , NA= 4270, P = 5	100	15
N= 600 , NA= 6460, P=6	338	60
N= 800 , NA= 8674, P =6	794	141
N= 800 , NA= 311332, P=5	793	37
N= 1000 , NA= 39301, P = 4	1503	81
N= 1000 , NA= 39164, P = 25	1403	558
N= 1000 , NA= 10868, P = 26	1510	727
N= 200 , NA= 1998, P=26	11	17
N= 300 , NA= 3145, P = 26	41	42

- [4] R.W. Floyd, Algorithm 97: Shortest Path, *Communications of the ACM* (1962), 5(6), pp. 345.
- [5] L. Gross and J. Yellen, *Graph Theory and Its Applications*, CRC Press (2019).
- [6] S.M. Hedetniemi, E.J. Cockayne and S.T. Hedetniemi, Linear algorithms for finding the Jordan center and path center of a tree, *Transportation Science* (1981), 15(2), pp. 98-114.
- [7] R. Krithika, A.K. Mohan and M. Sethumadhavan, Jordan center segregation: rumors in social media networks, *in* : *Security in Computing and Communications*, 5th International Symposium, SSCC (2017), Manipal, India.
- [8] R. Lester and Jr. Ford, *Network Flow Theory*, Santa Monica, California, RAND Corporation (1956), coll. "Paper P-923".
- [9] W. Luo, W.P. Tay, M. Leng and M.K. Guevara, On the universality of the Jordan center for estimating the rumor source in a social network, *2015 IEEE International Conference on Digital Signal Processing (DSP)*, Singapore (2015), pp. 760-764.
- [10] E.F. Moore, The shortest path through a maze, *in* *Proc. Internat. Sympos. Switching Theory* (1957), Part II, pp. 285–292, Harvard Univ. Press.
- [11] B. Roy, Transitivité et connexité, *C. R. Acad. Sci. Paris* (1959), 249, pp. 216–218.
- [12] V. Viton (2019), Personal communication.