Sequence Matching Hamilton Path
NP-hardness proof

Nil Mamano

Barcelona School of Informatics
[redacted]

Abstract. In the Sequence Matching Hamilton Path problem we are given an $n \times n$ matrix of 1's and 0's and a sequence of n^2 0's and 1's, and we are asked whether there is a path through adjacent matrix entries, covering each entry exactly once, with values matching the sequence. We prove it is NP-hard by giving a reduction from *Hamilton Path on Grid Graphs*, a known NP-complete problem.

Keywords: graph theory, Hamilton paths, grid graphs, NP-complete problem

1 Introduction

In the Sequence Matching Hamilton Path *(SMHP)* problem we are given an $n \times n$ matrix of 1's and 0's and a sequence of n^2 0's and 1's, and we are asked whether there is a path through adjacent matrix entries, covering each entry exactly once, with values matching the sequence. We say that two matrix entries $M_{i,j}$ and $M_{i',j'}$ are adjacent when $|i - i'| + |j - j'| = 1$.

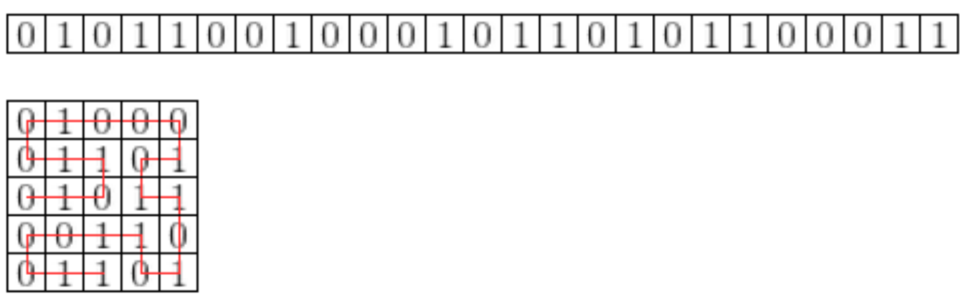

Fig. 1. A sequence and a matrix with a matching path highlighted

As defined in [1], the infinite grid is the infinite graph whose vertex set consists of all points of the plane with integer coordinates and in which two vertices are adjacent if and only if the distance between them is 1.

A grid graph is a vertex-induced finite subgraph[1] of the infinite grid.

The Hamilton Path on Grid Graphs *(HP)* problem is the Hamilton Path problem restricted to Grid Graphs. The NP-completeness proof of *HP* is established in [1].

2 Reduction

Assume G is a graph from the *HP* problem. In this section, we show how to obtain a corresponding input for the *SMHP* problem: a square matrix M and a sequence S.

2.1 Constructing the matrix

A *cluster* is a set of 2×2 matrix entries with the same value.

The matrix M is made of clusters. It is constructed by applying the following transformation to G:

The vertices are replaced by a cluster of 1's. The edges are replaced by a cluster of 0's placed between the clusters of 1's corresponding to the incident vertices in G.

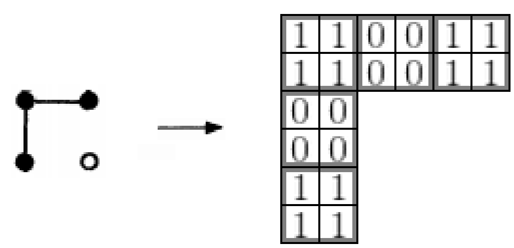

Fig. 2. Correspondence between vertices and edges in G and clusters in M

This construction is then embedded into the smallest possible square matrix by filling the remaining positions with clusters of 0's. Finally, this matrix is surrounded by an extra layer of clusters of 0's.

The obtained matrix preserves the structure of the original graph in the sense that two clusters of 1's are *close* (can be reached through a sequence of two 0's) if, and only if, the corresponding vertices in G are adjacent. For non-adjacent vertices, at least five 0's are required.

[1] A subgraph induced by a subset of the vertices has as set of edges all edges with both endpoints in the subset.

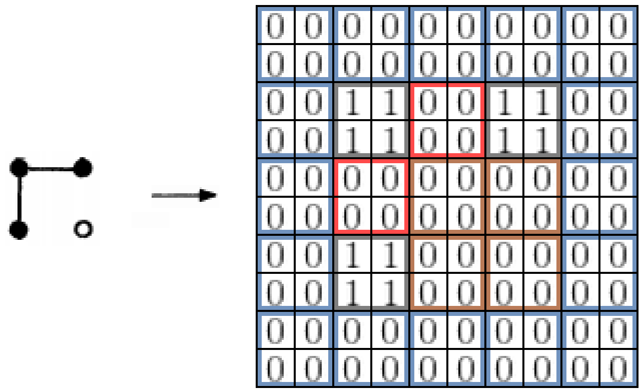

Fig. 3. Result of the transformation for this particular graph. Clusters corresponding to vertices of G are in grey, clusters corresponding to edges of G are in red, the remaining clusters of the smallest square matrix are in brown, and the clusters from the extra layer are in blue.

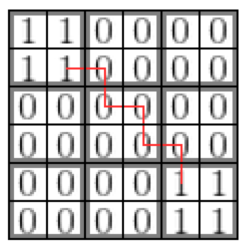

Fig. 4. The shortest path between Clusters of 1's corresponding to non-adjacent vertices of G has at least five 0's.

2.2 Constructing the sequence

The sequence is conceptually divided in two phases.

As we will see, the first phase emulates the hamilton path through G (in case there is one). Let k be the number of vertices in G. Then, S_1 is made of k sequences of four 1's intercalated with $k-1$ sequences of four 0's.

$$S_1 = 11110000111100001111...1111$$

The second phase consists of the remaining 0's. The exact length of S_2 can be deduced by the size and number of 1's in M.

$$S_2 = 000...0$$

$$S = S_1 S_2$$

2.3 Correctness proof

Claim. G has a hamilton path if, and only if, M contains a path without repeated entries that matches S_1.

Proof. $\Rightarrow$) Let $v_1 v_2 \cdots v_k$ be the hamilton path in G. The path in M that matches S_1 can be constructed as follows: the first four 1's match the cluster corresponding to v_1. They are traversed in an order such that the last 1 is adjacent to the cluster of 0's corresponding to the edge between v_1 and v_2. The following four 0's match the cluster corresponding to the edge between v_1 and v_2, being careful that the last 0 is adjacent to the 1's in the cluster of v_2. Similarly, the following four 1's match the cluster corresponding to v_2, and so on.

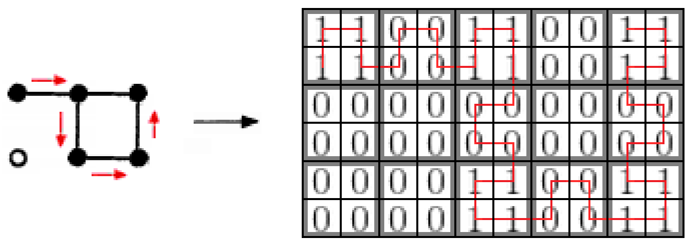

Fig. 5. S_1 is matched by traversing the clusters of 1's and 0's corresponding to vertices and edges in the path in the same order as the vertices and edges in the path

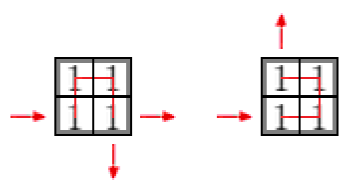

Fig. 6. Regardless of the *entry point* of the sequence in a cluster of 1's, we can traverse it in some order such that the last 1 is adjacent to any of the clusters corresponding to edges incident to the vertex.

Since the hamilton path $v_1 v_2 \cdots v_k$ does not repeat vertices nor edges, we will always find their corresponding clusters available. □

$\Leftarrow$) Suppose M contains a path that matches S_1. Since clusters of 1's are isolated by construction of M, the sequences of four 1's must correspond to all the 1's of a cluster corresponding to a vertex in the original graph. This guarantees that a cluster corresponding to a vertex in not visited twice. As we have seen, with sequences of four 0's we can only reach other clusters of 1's that correspond to vertices adjacent in the original graph. Hence, if we can match S_1, which has as many subsequences of 1's as vertices are in G, the vertices in the order in which their corresponding clusters are visited form a hamilton path in the original graph. □

The previous claim is sufficient to prove that the reduction preserves the negative answer. Indeed, when there is not a hamilton path in G, the resulting matrix won't even be able to match S_1. On the other hand, to prove that the positive answer is preserved, we have to show that once S_1 has been traversed, all the 0's of S_2 can *always* be matched.

To do this, we visualize the remaining matrix entries as a grid graph G_M. Each matrix entry correspond to a vertex, and there are edges between vertices corresponding to adjacent matrix entries. We have to show that G_M has a hamilton path. Interestingly, this is *HP*, the problem we are reducing from, but in this particular case our graph has some key properties that imply that there must always be a path.

We extend the notion of clusters to grid graphs. A cluster in a grid graph is a set of four vertices in adjacent positions in square shape. Two clusters are *adjacent* when there are two edges with endpoints in both clusters. A graph is *clustered* if it is a grid graph and its vertices are arranged in clusters, and those clusters are well aligned (any two clusters are either adjacent or not, but do not have a single edge between vertices in each one).

Claim. G_M is a clustered graph.

Proof. We constructed M in a way that it was divisible in clusters. When we traversed S_1 we only visited full clusters, so we had left a set of whole clusters. Those correspond one to one to the clusters in G_M. □

Claim. G_M is connected.

Proof. When we built M, we added a layer of clusters of 0's around the matrix. Those clusters do not correspond to any edge in G, and thus were not used to traverse S_1. This guarantees that the whole perimeter is connected.

In this situation, there is a set of isolated clusters if and only if it is surrounded by clusters corresponding to vertices or edges of the hamilton path in G, but this would imply that there is a cycle in the hamilton path. □

We say two edges v_1v_2, w_1w_2 of a clustered graph are *neighbors* if v_1 and v_2 are in the same cluster, w_1 and w_2 are in an adjacent cluster, v_1 and w_1 are adjacent and v_2 and w_2 too. We say an edge is a *frontier edge* if it doesn't have a neighbor.

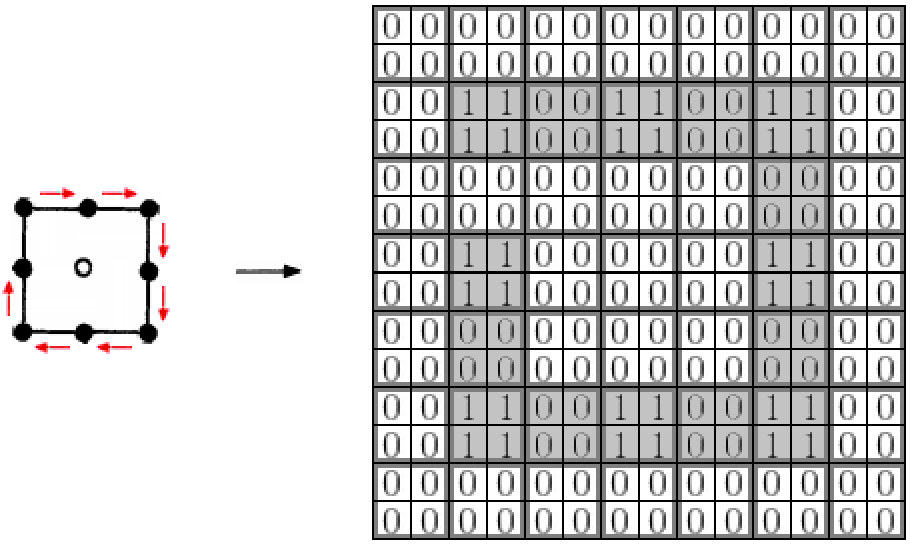

Fig. 7. An example of graph and its corresponding matrix. The matrix entries that are part of the path matching S_1 have been shadowed. The remaining 0's form a connected clustered graph.

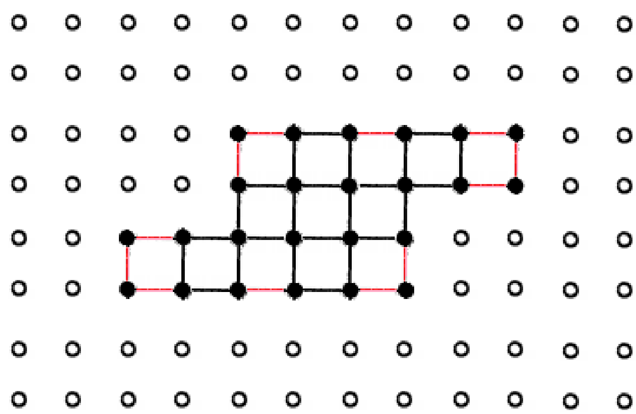

Fig. 8. A clustered graph with the frontier edges highlighted.

Theorem 1. *A connected clustered graph G has a hamilton circuit that contains all frontier edges of G.*

Proof. by induction on the number n of clusters.

Base case: $n = 1$. The graph has a single cluster, which consists of four vertices and four frontier edges. All the edges form a circuit. □

Inductive step: $n > 1$. Pick some cluster c from G. Let S be the set of c and its adjacent clusters. Partition G into sets of connected clusters such that each one contains one element of S. This can be done because G is connected, so every cluster can be reached from c or one of its adjacent vertices. In particular, the set containing c is consists of c alone. The subgraphs of G induced by the vertices on each set fulfill the induction hypothesis, and hence have a hamilton circuit. By properly rewiring some edges in c and the adjacent clusters, as illustrated in figure 9, we get a hamilton cycle through whole G.

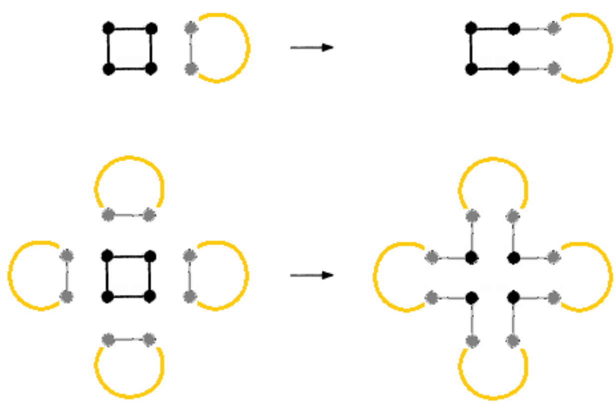

Fig. 9. Edge rewiring in the inductive step. c is the black cluster. In the first case, c has 3 frontier edges in G, whereas in the second it has none, although other scenarios are possible. The gray edges on the left side pictures are frontier edges of the induced subgraphs, as they don't have a neighbor (it would be an edge in c). The yellow hoops represent the remainder of the hamilton circuits of the subgraphs. Notice that in both cases, the right side picture displays a single circuit through all the edges.

We also have to see that every frontier edge of G is part of the obtained hamilton path. Since we reused most part of the paths through the subgraphs, this condition is compromised if either

1. G has some frontier edge that was not in the subgraphs and it is not part of the path
2. one of the frontier edges of one of the subgraphs has been removed from the path and it is still a frontier edge in G

Case 1 does not happen because the only edges in G that were not in any of the subgraphs are those between c and the rest of subgraphs, and they are not frontier edges because their endpoints are in different clusters.

Case 2. does not happen either because the only edges that have been removed from the paths in the subgraph are those that were removed to connect c with the rest of the subgraphs, and they are not frontier in G because they have an edge in c as neighbor. □

Let v be the last matrix entry used to match S_1. Then, v is adjacent to some matrix entry u not already visited with value 0. Due to the previous theorem, u is part of a hamilton circuit through all the remaining matrix entries. By disconnecting u from one of its neighbors in the path and connecting it to v, we form a path that matches S_2. This completes the correctness proof of the reduction.

Acknowledgements

I would like to thank Guillem Godoy and Carles Creus for their useful insight.

References

1. Itai, A., Papadimitriou, C., Szwarcfiter, J.: Hamilton Paths in Grid Graphs September 22, 1980