MLG:
Machine Learning with Graphs (*Strojno učenje z grafi*)

Assignment: Homework 3

Submission time: 20:30 **and date:** 10. 11. 2021

Submission Fill in and include this cover sheet with each of your assignments. It is an honor code violation to write down the wrong date and/or time. Assignments are due at 9:00am and should be submitted through Gradescope and eUcilnica. Students should check Piazza for submission details.

Late Periods Each student will have a total of *two* free late periods. *One late period expires the morning on the day before the next class.* (Assignments are usually due on Fridays, which means the first late period expires on the following Tuesday at 9:00am.) Once these late periods are exhausted, any assignments turned in late will be penalized 50% per late period. However, no assignment will be accepted more than *one* late period after its due date.

Honor Code We strongly encourage students to form study groups. Students may discuss and work on assignments in groups. However, each student must write down their solutions independently, i.e., each student must understand the solution well enough in order to reconstruct it by him/herself. Students should clearly mention the names of all the other students who were part of their study group. Using code or solutions obtained from the web (GitHub, Google, previous years etc.) is considered an honor code violation. We check all submissions for plagiarism. We take the honor code very seriously and expect students to do the same.

Name: Sara Bizjak

Email: sarabizjak97@gmail.com **SID:** 27202020

Study group: Maruša Oražem, Vid Stropnik

I acknowledge and accept the Honor Code.

(Signed)

Machine Learning with Graphs

CS224W Homework 3

Sara Bizjak — 27202020

November 2021

1 GraphRNN

1.1

We are given the following graph.

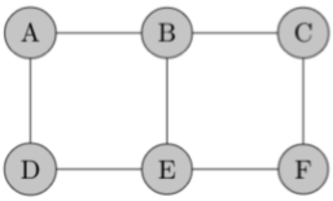

Figure 1: Example grid graph.

The order of nodes, added in BFS ordering staring from A is: A, B, D, C, E, F. Or more clearly, as presented in the figure below.

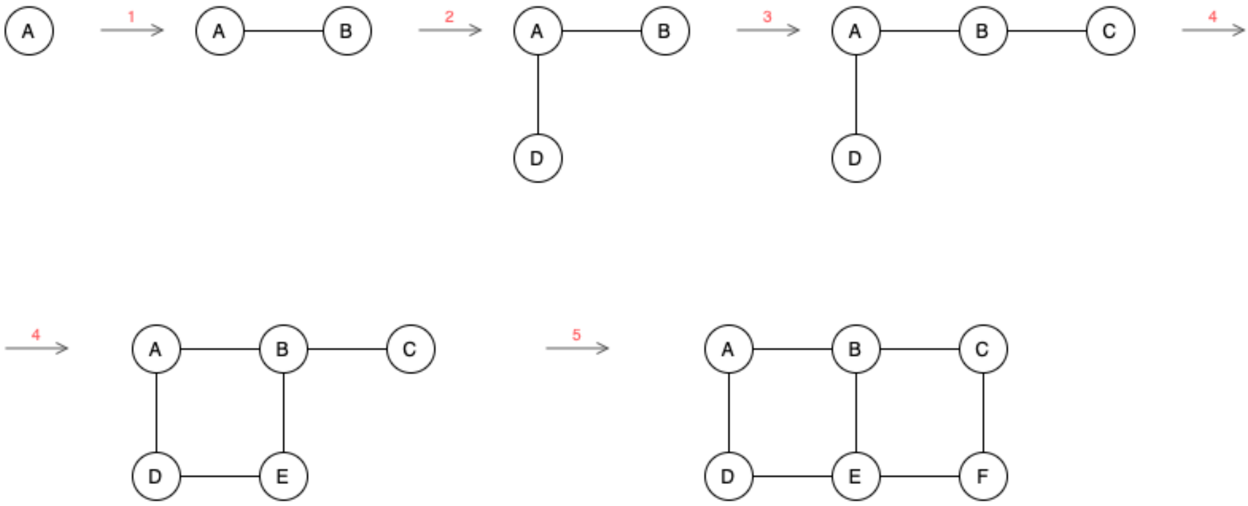

Figure 2: The order of adding nodes in BFS.

The edge level predictions for each node, respectively as steps in BFS ordering (steps 1-5 in the figure):

- 1: $S^{\pi}_{B} = [1]$
- 2: $S^{\pi}_{D} = [1, 0]$
- 3: $S^{\pi}_{C} = [0, 1, 0]$
- 4: $S^{\pi}_{E} = [0, 1, 1, 0]$
- 5: $S^{\pi}_{F} = [0, 0, 0, 1, 1]$

However, for S^{π}_{E} and S^{π}_{F}, it would be enough to predict only the edges between the last three added nodes, since the shortest-path between two the farthest nodes is of distance three (following Corollary 1 from attached GraphRNN paper). Therefore, we could write $S^{\pi}_{E} = [1, 1, 0]$ and $S^{\pi}_{F} = [0, 1, 1]$.

1.2

Two advantages of graph generation using BFS ordering of nodes, as apposed to generating with a random ordering of nodes in the graph are:
Less options for node orderings. The number of all possible node permutations is $n!$, which is also the worst case for BFS ordering (star graphs), however, the number of observed BFS ordering on many real-world graphs is a lot smaller.
Less edge predictions. As seen in the example above when following Corollary 1, when added a new node, only last M steps are enought to make edge predictions (last 3 steps in the example). Therefore, we have only M possible edges on every step, thus time complexity is $\mathcal{O}(Mn)$, where n is the number of added nodes, instead of $\mathcal{O}(n^2)$ when checking "every node with every node".

2 Subgraphs and Order Embeddings

2.1 Transitivity

Graph A is a subgraph of graph B and B is a subgraph of graph C, we then prove that A is a subgraph of C.
Following the subgraph isomorphism definition:

$\exists$ bijective $f : V_A \to U_B \subseteq V_B$ and subgraph of B induced by $\{f(v)|v \in V_A\}$ is graph-isomorphic to A.
$\exists$ bijective $g : V_B \to U_C \subseteq V_C$ and subgraph of C induced by $\{g(v)|v \in V_B\}$ is graph-isomorphic to B.

Therefore, we need to find a bijective mapping, denote it by h, such that $h : V_A \to U_C \subseteq V_C$ and prove that subgraph of C induced by $\{h(v)|v \in V_A\}$ is graph-isomorphic to A.
A bijective mapping h could, clearly, be a compositum of mappings g and f,

$$h = g \circ f : V_A \to U_C \subseteq V_C$$

Assume that a_1 and a_2 are two nodes from graph A, so $a_1, a_2 \in V_A$. Since

$$\begin{aligned} f(a_1) &= b_1 \in U_B \subseteq V_B \\ f(a_2) &= b_2 \in U_B \subseteq V_B \\ g(b_1) &= c_1 \in U_C \subseteq V_C \\ g(b_2) &= c_2 \in U_C \subseteq V_C \end{aligned}$$

using the mapping h on a_1 and a_2 will give

$$h(a_1) = g(f(a_1)) = c_1 \in U_C \subseteq V_C$$

and

$$h(a_2) = g(f(a_2)) = c_2 \in U_C \subseteq V_C$$

Thus, $\forall a_1, a_2 \in V_A : a_1 \sim_A a_2 \Leftrightarrow h(a_1) \sim_C h(a_2)$, so $h = (g \circ f)$ is isomorphism and A is a subgraph of C.

2.2 Anti-symmetry

Graph A is a subgraph of graph B, and graph B is subgraph of graph A. Prove that A and B are graph-isomorphic.
From $|V_A| \leq |V_B|$ and $|V_B| \leq |V_A|$ it follows that $|V_A| = |V_B|$, so we can rewrite the subgraph isomorphic definition as

$$\begin{aligned} &\exists \text{ bijective } f : V_A \to V_B \\ &\exists \text{ bijective } g : V_B \to V_A \end{aligned}$$

since subgraph of B induced by $\{f(v)|v \in V_A\}$ and subgraph of A induced by $\{g(v)|v \in V_B\}$ in this case equals B and A, respectively. In other words, mappings between all nodes from A and all nodes from B and vice versa are bijective.
Therefore, by the definition, we can conclude that graphs A and B are isomorphic.

2.3 Common subgraph

Prove that graph X is a common subgraph of A and B if and only if $z_x \preceq \min\{z_A, z_B\}$, where min denotes the element-wise minimum of the two embedding vector.
The proof goes in both ways (equivalence).
Firstly, we assume that X is a common subgraph of A and B.
For X to be a subgraph of A, the following must be true:

$$z_x \preceq z_A$$

And similar for X to be a subgraph of B:

$$z_x \preceq z_B$$

To make the presentation of the two possibilities easier, let us take a look at the figure below, presenting both cases.

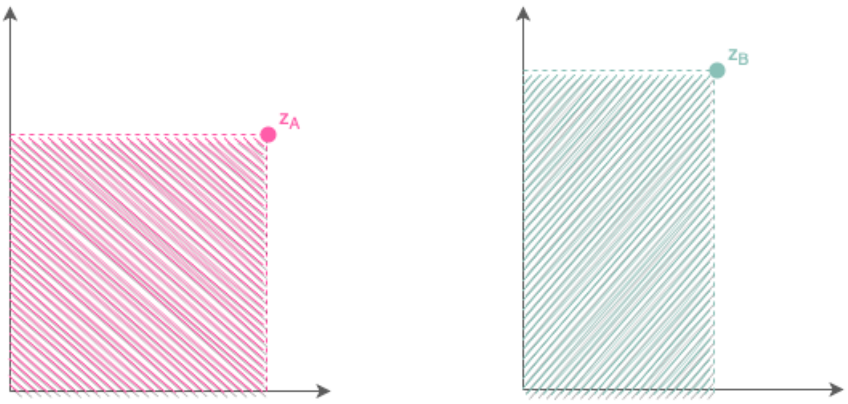

Figure 3: Embedding of X must lie in the colored areas for X to be a subgraph of A in the left and B on the right graph.

For X to be a subgraph of both, we join the two conditions, so

$$\begin{aligned} & z_x \preceq z_A \quad \text{and} \quad z_x \preceq z_B \\ \Rightarrow\ & z_x \preceq \min\{z_A, z_B\} \end{aligned}$$

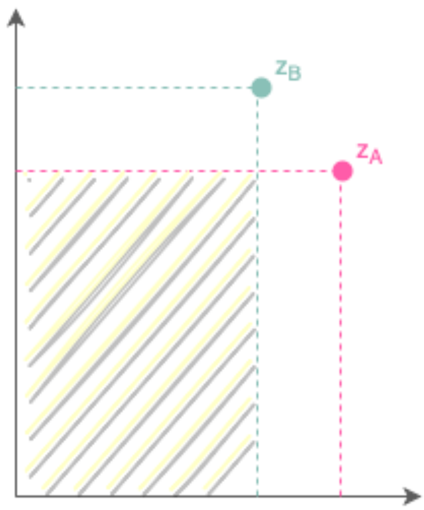

Figure 4: Embedding of X must lie in the colored area for X to be a subgraph of both A and B.

Secondly, we assume $z_x \preceq \min\{z_A, z_B\}$. The condition is presented on the Figure 4. If the embedding of X lies in the colored area, the graph is clearly a subgraph of both A and B.

2.4 Order embeddings for graphs that are not subgraphs of each other

We have the graphs A, B, C that are not subgraphs of each other. We embed them into 2-dimensional space and the condition $z_A[0] > z_B[0] > z_C[0]$ hold for first dimension. The vizualization of the given conditions is presented in the figure below.

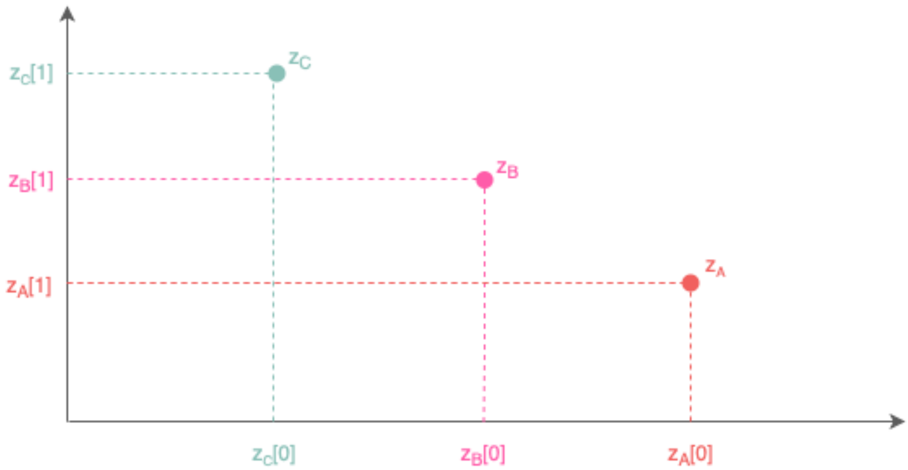

Figure 5: Embeddings of A, B, C such that $z_A[0] > z_B[0] > z_C[0]$ hold.

Thus, the situation presented in the figure clearly implies $z_A[1] < z_B[1] < z_C[1]$.

2.5 Limitation of 2-dimensional order embedding space

Here we show that 2-dimensional order embedding space is not sufficient to perfectly model subgraph relationships.
Suppose graphs A, B, C are non-isomorphic graphs that are not subgraphs of each other. Construct an example of graphs X, Y, Z such that they are each subgraphs of one or more graphs in {A, B, C}. Corresponding embeddings should satisfy $z_x \preceq z_Y$ and $z_X \preceq z_Z$.

Following the additional clarifications, such an example, so that each of the constructed graph is a supgraph of exactly two graphs in {A, B, C}, would be

- X is a subgraph of A and C,
- Y is a subgraph of A and B,
- Z is a subgraph of B and C.

From the subgraph relations above it is obvious that graph X is not a subgraph of Y and Z (because it is not subgraph of all three graphs in {A, B, C}). Since without loss of generality we can suppose $z_A[0] > z_B[0] > z_C[0]$ (the situation is presented in the Figure 2.4), there obviously exist such embeddings for X, Y and Z, such that conditions $z_x \preceq z_Y$ and $z_X \preceq z_Z$ hold, meaning X is a common subgraph of Y and Z. Therefore, we really need higher dimensional embedding space to perfectly model subgraph relationships.