# Similarity Measure of Graphs

https://doi.org/10.3991/ijes.v5i2.7251

Amine Labriji
University Hassan II Casablanca, Morocco
labriji73@gmail.com

Salma Charkaoui
University Hassan II Casablanca, Morocco
charkaoui.salma@gmail.com

Issam Abdelbaki
University Hassan II Casablanca, Morocco
i.abdelbaki@gmail.com

Abdelouhaed Namir
University Hassan II Casablanca, Morocco
a.namir@yahoo.fr

El Houssine Labriji
University Hassan II Casablanca, Morocco
labriji@yahoo.fr

**Abstract—**The topic of identifying the similarity of graphs was considered as highly recommended research field in the Web semantic, artificial intelligence, the shape recognition and information research. One of the fundamental problems of graph databases is finding similar graphs to a graph query. Existing approaches dealing with this problem are usually based on the nodes and arcs of the two graphs, regardless of parental semantic links. For instance, a common connection is not identified as being part of the similarity of two graphs in cases like two graphs without common concepts, the measure of similarity based on the union of two graphs, or the one based on the notion of maximum common sub-graph (SCM), or the distance of edition of graphs. This leads to an inadequate situation in the context of information research. To overcome this problem, we suggest a new measure of similarity between graphs, based on the similarity measure of Wu and Palmer. We have shown that this new measure satisfies the properties of a measure of similarities and we applied this new measure on examples. The results show that our measure provides a run time with a gain of time compared to existing approaches. In addition, we compared the relevance of the similarity values obtained, it appears that this new graphs measure is advantageous and offers a contribution to solving the problem mentioned above.

**Keywords—** Graph, ontology, similarity measure, semantic web, user profile.

# 1 Introduction

The importance of graphs has been increasing in the complex structured data modeling in several real and recent applications (Bioinformatics, pattern recognition, XML, chemistry, user profile, information research etc.) [1], [2], [3], [4], [5]. All of those applications show the importance and the wide use of the Graph Databases paradigm (BDGS) [3], [6], [7]. The similarity between graphs is an issue that has been at the heart of several research studies. Usually, we can classify the requests to a BDG into two categories:

- Search graphs by inclusion relation.
- Search graph by similarity.

The first category consists of two sub-problems:

1. Search sub-graphs.

Let D = {g1, g2, . . . , gn}, a BDG and q a graph query(said a request subgraph) it is to search all graphs gi of database, D, such that q is a subgraph of gi (it means, q ⊂ gi).

2. Search super graphs. Let D = {g1, g2, . . . , gn}, a BDG and q a graph query(called super graph query) This is to search all gi graphs, of the graph database such as q is a super graph gi (it means, gi ⊂ q).

[5], [8], [9], [10]).“ As for the second category (that is to say search graphs by similarity) it consists in seeking all the graphs of GDB which are structurally similar to the graph of the current request and has emerged as a new trend [11], [12]).

In recent years, a number of approaches have been proposed to meet the research graphs similarity queries (or simply, queries by similarity). For example

- (Yan and al., [13], He and Singh [14] and Tale (Tian and Patel [11] have proposed techniques to respond to requests subgraphs by means of an approximate matching.
- hang et al. [15] proposed a technique to find an answer to the super-graph queries using a similarity search.
- oth approaches, C-Tree and Tale, use the edit distance to measure the similarity between graphs.
- he work of (Yan and al., [13] and (Shang and al. 2010) [15], [16] use the concept of maximal common subgraph for the calculation of such similarity.

As can be seen, the proposed approaches to respond to queries by similarity they rely solely on the concepts of the two graphs disregarding fathers concepts. That way, we cannot preserve information about the similarity of each feature in the comparison of two graphs. In this article, we propose a new definition of the similarity between graphs using the notion of conceptual similarity [17], [18].

## 2 Existing Measure of Graphs Similarities

In this section we will first recall some basic notions on graphs and measurements of existing similarities between graphs and then we will recall the definition of the similarity measure between nodes that will be used to define our new measure of similarity between graphs. In this section we will first recall some basic notions on graphs and measurements of existing similarities between graphs and then we will recall the definition of the similarity measure between nodes that will be used to define our new measure of similarity between graphs.

### 2.1 Somme basic notions about graphics

Graph A graph g is defined by a quadruplet (V, E, L, l) where V is all the knots, E is all the bones, L all the tabs is and l the function of labelling is which puts in correspondence every knot or bone with a tab of L.

Let us note that different knots can have the same tab and size of g is defined as follows:

|g| = |E(g)| (i.e., the size of a graph is the number of its bones).

**Isomorphism of graph:** Either two graphs g = (V, E, L, 1) and g' = (V', E', L', l'), g an isomorphic in g' (denoted by g ≈ g' ) if there is a bijection f : V → V' , such as

1. ∀v ∈ V, f(v) ∈ V' and l(v) = l'(f(v)).

2. ∀(v, v') ∈ E, (f(v), f(v')) ∈ E' and l(v, v' ) = l'(f(v), f(v'))

**Isomorphism of sub-graph:** Given two graphs g = (V, E, L, 1) and g' = (V', E', L', 1'), g an isomorphic of sub-graph in g' if there is an injection fr : V → V', such as

1. ∀v ∈ V, f(v) ∈ V' and l(v) = l'(f(v)).

2. (v, v') ∈ E, (f(v), f(v')) ∈ E' and l(v, v' ) = l'(f(v), f(v'))

**Sub-graph super-graph:** Given two graph g = (V, E, L, 1) and g' = (V', E', L', 1'), g is said sub-graph of g' (or g' is a super-graph g), denoted by g ⊂ g' if there is an isomorphism of sub-graph g to g'.

**Common Maximal Subgraph, CMS:** Definition (Common Maximal Subgraph, CMS).

Let g 1 and g 2 be the largest common subgraph of g 1 and g 2 is the largest connected subgraph of g 1 which is isomorphic to g 2, denoted by g '= SCM (g 1, g 2).

### 2.2 The similarity between graphs.

**Similarity based on the notion CMS:** Bunke and Shearer (1998) [3] developed a type of similarity measures between graphs which is based on the maximum common

subgraph (CMS). Consider two groups g1 and g2, the similarity-based CMS is defined as follows,

$$Sim_{CMS}(g_1, g_2) = \frac{|CMS(g_1, g_2)|}{\max\{|g_1|, |g_2|\}}$$

where |max(g1, g2)| = max(|g1|, |g2|) and |CMS(g1, g2)| denotes the number of edges in CMS(g1, g2). Clearly, if the CMS of two graphs is wide, then their similarity is high. The Measurement, $Sim_{CMS}$, is standardized ($0 \leq Sim_{CMS}(g1, g2) \leq 1$) car $|CMS(g1, g2)| \leq |max(g1, g2))|$.

**Similarity based on the union of two graphs, $Sim_{UG}$:** The distance measurement based on the union of graphs (UG), proposed by Wallis et al. (2001) [7], is based on the union of graphs. This distance is used to model the size of the problem. Definition Given two graphs g1 and g2, the similarity of graphs based on the union of graphs is defined as follows,

$$Sim_{UG}(g_1, g_2) = \frac{|\mathrm{CMS}(g_1, g_2)|}{|g_1| + |g_2| - |\mathrm{CMS}(g_1, g_2)|}$$

Where the denominator is the size of the union of two graphs as a set-view. This similarity measure is normalized and its behavior is similar to that of SimSCM. It is easy to see that $Sim_{UG}(g1, g2) < Sim_{CMS}(g1, g2)$ ( meaning that $Sim_{UG}$ is more demanding as $Sim_{CMS}$).

## 3 Similarity Between Nodes of a Graph

Our new measure of similarity between graphs is based essentially on the similarity between the nodes. In this section we will describe the similarity measurements between the most used nodes. The most intuitive similarity measure of the nodes of a graph is their distance [19]- [20], [17]. This similarity is evaluated by the distance that separates the nodes in the graph. In each graph, distance is characterized by the shortest path that involves a common ancestor or the smallest generalizing, potentially connecting two objects across common descendants. Among the works classified under this banner are:

### 3.1 Measure of rada and al

This measure [19] is based on the fact that we can compute the similarity based on the "est-to" hierarchical links. To calculate the similarity of the two nodes in the graph, we must calculate the number of minimal arcs that separate them. Intuitively, this measure is based on the following principle A node C1 is considered more similar to a node C2 than to a node C3 if the distance from C1 to C2 within the graph is shorter than that of C1 to C3 Rada and al. [19] Considers this distance, denoted distedge(c1, c2), as the length of the shortest path between two concepts. The similarity between c1, c2 is defined by

$$Sim(c_1, c_2) = \frac{1}{dist_{edge}(c_1, c_2)}$$

### 3.2 Wu and Palmer measure

The principal of calculating similarity based on the count edge method is defined as follow; considering the graph Ω formed by a set of nodes and a root node (R) (Fig. 1). C1 and C2 represent two elements of the graph of which we will calculate the similarity. The principle of

calculating similarity is based on the distance (n1 and n2) from nodes C1 and C2 to the closest common ancestor (CS) and the distance, n, from the closet common ancestor (CS) of C1 and C2 to the root node.

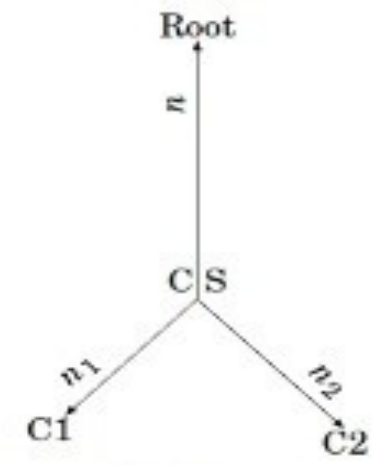


**Fig. 1.** Two concepts of a graph

The similarity measure of Wu and Palmer is defined by the following expression:

$$Sim_{wp}(c1, c2) = \frac{2.n}{n_1 + n_2 + 2.n}$$

This measure has been improved and the new measure is

$$Sim(c1, c2) = \frac{2n}{n_1 + n_2 + 2.n + n.n_1.n_2}$$

## 4 Similarity of Graphs Based on the Similarity of Concepts

A weighted graph is defined by

$$\boldsymbol{G = \{(C_i, \alpha_i),\ \ i = 1 \ldots n\}}$$

Where Ci is a concept and $\boldsymbol{\alpha_i}$ is its weight.

### 4.1 Similarity between concept and graph

*Definition:*

Let G be a graph, we pose $\bar{G}$, the set of all the parents of all concepts graph G.

Let C be a concept, a graph G and Sim(., .) a similarity function between concept. We define the similarity between the concept C and the graph G, by

$$Sim(C, G) = max_{g \in G}\{\text{sim}(C, g)\}$$

*Proposition*

Let C a concept and a graph G, we have

$$C \in G \Leftrightarrow Sim(C, G) = 1$$

2.

$$Sim(C, G) = 0 \Leftrightarrow \overline{\{C\}} \cap \bar{G} = \emptyset$$

*Proof*

1. $C \in G \Leftrightarrow Sim(C, G) = max_{g \in G}\{sim(C, g)\} = 1.$

So

$$\exists\, f \in G / max_{g \in G}\{sim(C, g)\} = sim(C, f) = 1$$

Based on the properties of the similarity function, we deduce that

$$C = f.$$

Therefore, C ∈ G.

2. $Sim(C, G) = 0 \Leftrightarrow max_{g \in G} Sim(C, g) = 0$

So

$$\forall g \in G, Sim(C, g) = 0$$

According to the definition of the similarity function see [18], we deduce that C concept and all concepts of G does not have parents common. So

$$\overline{\{C\}} \cap \bar{G} = \emptyset$$

### 4.2 Similarity of a graph with respect to another

*Definition*

let G1 = {($g_{1,i}$, $\alpha_i$), i = 1, 2, . . .} and G2 = {($g_{2,j}$ , $\beta_j$ ), j = 1, 2, . . .} two weighted graphs. We define similarity graph G1 compared to G2, which notes

$$sim(G1/G2) = \frac{\sum_i \alpha_i max_{g2,j} sim(g_{1,i}, g_{2,j})}{\sum_i \alpha_i}$$

*Proposition*

Let G1 and G2 are two weighted graphs
1. sim(G1/G2) = 1 ⇒ G1 ⊂ G2.
2. sim(G1/G2) = 0 ⇒ G1 ∩ G2 = ∅.

*Proof*

1. We have

$$Sim(G1/G2) = \frac{\sum_i \alpha_i max_{g2,j} Sim(g_{1,i}, g_{2,j})}{\sum_i \alpha_i} = 1$$

Therefore

$$\sum_i (\alpha_i max_{g2,j} sim(g_{1,i}, g_{2,j}) - \alpha_i) = 0$$

So,

$$\forall\, g_{1,i} \in G1, max_{g2,j} sim(g_{1,i}, g_{2,j}) = 1$$

because coefficients αi are different from zero. Therefore

$$\exists g_{2,k} \in G2, max_{g2,j}\, Sim(g_{1,i}, g_{2,j}) = \; Sim(g_{1,i}, g_{2,K}) = 1$$

So

$$g_{1,2} = g_{2,k} \;\; and\; g_{1,i} \in G2.$$

It follows that

$$G1 \subset G2.$$

2. Since

$$\forall\, g_{1,i} \in G1, max_{g2,j} sim(g_{1,i}, g_{2,j}) = 0$$

Thereafter

$$\forall\, g_{1,i} \in G1, \forall g_{2,j} \in G2 \;\; sim(g_{1,i}, g_{2,j}) = 0$$

That is, the concepts of the two graphs G1 and G2 do not have common parents. So

$$G1 \cap G2 = \emptyset.$$

*Proposition*

Let G1 = {(g1,i, αi), i = 1, 2, . . .}, G2 = {(g2,i, γi), i = 1, 2, . . .} and G3 = {(g3,j , βj ), j = 1, 2, . . .} three weighted graphs. We have

$$Sim(G1/G2) \leq Sim(G1/G3) + Sim(G3/G2)$$

*Proof*

By definition

$$Sim(G1/G2) = \frac{\sum_i \alpha_i max_{g2,j} Sim(g_{1,i}, g_{2,j})}{\sum_i \alpha_i}$$

Let, $g_{3,k} \in G3$, according to the properties of the similarity between two concepts, it was

$$Sim(g_{1,i}, g_{2,j}) \leq Sim(g_{1,i}, g_{3,k}) + Sim(g_{3,k}, g_{2,j}).$$

Therefore

$$Sim(G1/G2) = \frac{\sum_i \alpha_i max_{g2,j} Sim(g_{1,i}, g_{2,j})}{\sum_i \alpha_i}$$

$$\leq \frac{\sum_i \alpha_i max_{g2,j} Sim(g_{1,i}, g_{3,k}) + \sum_i \alpha_i max_{g2,j} Sim(g_{3,k}, g_{2,j})}{\sum_i \alpha_i}$$

$$\leq Sim(G1/G3) + Sim(g_{3,k}, g_{2,j})$$

$$\leq Sim(G1/G3) + \frac{\sum_j \beta_j max_{g2,j} Sim(g_{3,k}, g_{2,j})}{\sum_j \beta_j}$$

So, we deduce

$$Sim(G1/G2) \leq Sim(G1/G3) + Sim(G3/G2)$$

### 4.3 Similarity between two graphs

*Definition*

G1 and G2 are two weighted graphs. It poses as similarity of graphs G1 and G2 the following expression

$$Sim(G1, G2) = \frac{Sim(G1/G2) + Sim(G2/G1)}{2}$$

*Proposition*

Let G1, G2 and G3 three weighted graphs, we have

1. $Sim(G1, G2) = 1 \Leftrightarrow G1 = G2.$
2. $Sim(G1, G2) = Sim(G2, G1)$
3. $Sim(G1, G2) \leq Sim(G1, G3) + Sim(G3, G2)$

*Proof*

1. If

$$Sim(G1, G2) = 1 \Rightarrow \frac{Sim(G1/G2) + Sim(G2/G1)}{2} = 1$$

Since Sim(G1/G2) and Sim(G2/G1) are less than or equal to 1. It follows that Sim(G1/G2) = 1 and Sim(G2/G1) = 1 we know that

$$Sim(G1/G2) = 1 \Leftrightarrow G1 \subset G2$$

$$Sim(G2/G1) = 1 \Leftrightarrow G2 \subset G1$$

We can deduce

$$G1 = G2$$

2.

$$Sim(G1,G2) = \frac{Sim(G1/G2) + Sim(G2/G1)}{2}$$

$$= \frac{Sim(G2/G1) + Sim(G1/G2)}{2}$$

$$= Sim(G2,G1)$$

3. Based on the properties of the similarity of a graph with respect to another, it has been

$$Sim(G1,G2) = \frac{Sim(G1/G2) + Sim(G2/G1)}{2}$$

$$\leq \frac{Sim(G1/G3) + Sim(G3/G1)}{2} + \frac{Sim(G3/G2) + Sim(G3/G1)}{2}$$

$$\leq Sim(G1,G3) + Sim(G3,G2)$$

consequently

$$Sim(G1, G2) \leq Sim(G1, G3) + Sim(G3, G2)$$

## 5 An Illustrative Example

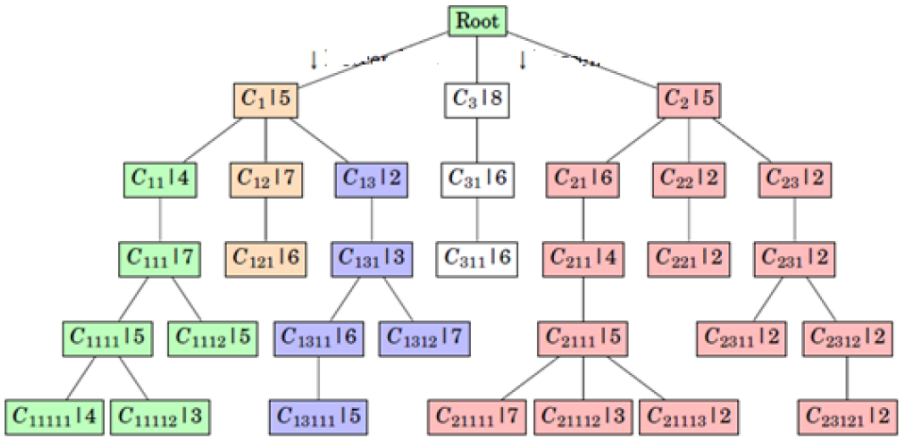


**Fig. 2.** Example of graph

Considering weighted graphs given by

1. G1 = {($C_1$, 5),($C_{11}$, 4),($C_{12}$, 7),($C_{111}$, 7),($C_{121}$, 6),($C_{1111}$, 5),($C_{1112}$, 5),($C_{11111}$, 4), $C_{11112}$, 3)}.
2. G2 = {($C_1$, 5),($C_{13}$, 2),($C_{12}$, 7),($C_{13}$, 2),($C_{121}$, 6),($C_{1311}$, 6),($C_{1312}$, 7),($C_{14111}$, 5)}.
3. G3={($C_{11}$, 4),($C_{111}$, 7),($C_{1111}$, 5),($C_{1112}$, 5),($C_{11111}$, 4), $C_{11112}$, 3)}.
4. G4 = {($C_{13}$, 2),($C_{13}$, 2),($C_{1311}$, 6),($C_{1312}$, 7),($C_{13111}$, 5)}.
5. G5 = {($C_2$, 5),($C_{21}$, 6),($C_{22}$, 2),($C_{23}$, 2),($C_{211}$, 4), . . . ,($C_{23121}$, 2)}.

To compare the new graph similarity measure with existing measure Simscm(Gi, Gj) and simUG(Gi, Gj), we will calculate

$$Sim(Gi, Gj) \text{ for } i = 1 \ldots 4 \text{ and } j = 1 \ldots 4.$$

$$SimSCM(Gi, Gj) \text{ for } i = 1 \ldots 4 \text{ and } j = 1 \ldots 4.$$

$$SimUG(Gi, Gj) \text{ for } i = 1 \ldots 4 \text{ and } j = 1 \ldots 4.$$

### 5.1 The new measure of similarity between graphs

In the following table the element ai,j = sim(ci , cj ) represents the similarity between the concepts ci and cj.

**Table 1.** Similarity between the concepts

| G1 / G2 | $C_1\|5$ | $C_{11}$ \|4 | $C_{12}$\| 7 | $C_{121}$ \|6 | $C_{111}$ \|7 | $C_{1111}$ \|5 | $C_{1112}$ \|5 | $C_{11111}$ \|4 | $C_{11112}$ \|3 |
|---|---|---|---|---|---|---|---|---|---|
| $C_1$\| 5 | 1 | 2/3 | 2/3 | ½ | ½ | 2/5 | 2/5 | 1/3 | 1/3 |
| $C_{12}$\|7 | 2/3 | 2/5 | 1 | 4/5 | 2/7 | 2/9 | 2/9 | 2/11 | 2/11 |
| $C_{121}$ \|6 | ½ | 2/7 | 4/5 | 1 | 1/5 | 2/13 | 2/13 | 2/16 | 2/16 |
| $C_{13}$\|2 | 2/3 | 2/5 | 2/5 | 2/7 | 2/7 | 2/9 | 2/9 | 2/11 | 2/11 |
| $C_{131}$\|3 | ½ | 2/7 | 2/7 | 1/5 | 2/5 | 2/13 | 2/13 | 1/8 | 1/8 |
| $C_{1311}$\|6 | 2/5 | 2/9 | 2/9 | 2/13 | 2/13 | 2/17 | 2/17 | 2/21 | 2/21 |
| $C_{1312}$\|7 | 2/5 | 2/9 | 2/9 | 2/13 | 2/13 | 2/17 | 2/17 | 2/21 | 2/21 |
| $C_{13111}$\|5 | 1/3 | 2/11 | 2/11 |  | 1/8 | 2/21 | 2/21 | 1/13 | 1/13 |

The similarity, Sim(G1/G2), of G1 compared to G2 is given by

$$Sim(G1/G2) = \frac{\sum_i \alpha_i max_{C_j \in G2} Sim(C_i, C_j)}{\sum_i \alpha_i} = \frac{1*5+7*1+6*1+2*2/5+3*2/5+6*2/9+7*2/9+5*2/11}{41} = 0.58$$

The similarity, sim(G2/G1), of G2 compared to G1 is given by

$$Sim(G1/G2) = \frac{\sum_j \beta_j max_{g_{1,i}} Sim(g_{1,i}, g_{2,j})}{\sum_j \beta_j} = \frac{5+4*2/5+7*1+6*1+7*2/5+5*2/9+5*2/9+4*2/11+3*2/11}{41} 0.3414$$

We deduce the similarity of two graphs G1 and G2.

Sim(G1, G2) = (0, 58 + 0, 34)/2 = 0.46

We propose to calculate the similarity between the graphs G3 and G4.

**Table 2.** Similarity between the concepts of graphs G3 and g4

| G3 / G4 | ($C_{11}$\|4) | ($C_{111}$\|7) | ($C_{1111}$\|5) | ($C_{1112}$\|5) | ($C_{11111}$\|4) | ($C_{11112}$\|3) |
|---|---|---|---|---|---|---|
| ($C_{13}$,2) | 2/5 | 2/7 | 2/9 | 2/9 | 2/11 | 2/11 |
| ($C_{131}$,2) | 2/7 | 2/10 | 2/13 | 2/13 | 2/16 | 2/16 |
| ($C_{1311}$,6) | 2/9 | 2/13 | 2/17 | 2/17 | 2/21 | 2/21 |
| ($C_{1312}$,7) | 2/9 | 2/13 | 2/17 | 2/17 | 2/21 | 2/21 |
| ($C_{13111}$,5) | 2/11 | 2/16 | 2/21 | 2/21 | 2/26 | 2/26 |

We can deduce

$$Sim(G4/G3) = (0.8 + 0.57 + 1.33 + 1, 55 + 0.91)/22 = 0.2345$$

$$Sim(G3/G4) = (0.88 + 2 + 1.11 + 1.11 + 0.88 + 0.54)/28 = 0.2328$$

Following the similarity between the two graphs G1 and G2 is

$$\mathrm{Sim}(\mathrm{G3}, \mathrm{G4}) = (0.2345 + 0.2328)/2 = 0.2336$$

*In the followi*ng array we show the values of the new measure of similarity between graphs.

**Table 3.** Similarity between the concepts of graphs

| Sim(Gi;Gj) | G1 | G2 | G3 | G4 |
|---|---|---|---|---|
| **G1** | 1 | 0,46 | 0,89 | 0,36 |
| **G2** | 0,46 | 1 | 0,33 | 0,88 |
| **G3** | 0,89 | 0,33 | 1 | 0,23 |
| **G4** | 0,36 | 0,88 | 0,23 | 1 |

### 5.2 The similarity measure between graphs based on CMS

In applying the definition of similarity based on SCM, we find the following results:

$$Sim_{SCM}(G1, G2) = \frac{|SCM(G1, G2)|}{|max(G1, G2)|} = \frac{3}{9} = 0,33$$

$$Sim_{SCM}(G3, G4) = \frac{|SCM(G3, G4)|}{|max(G3, G4)|} = 0.6 = 0$$

In the following array we show, $\mathrm{Sim_{CMS}}$(Gi, Gj), the values of the measure of similarity, based on SCM, between graphs

**Table 4.** Similarity, $\mathrm{Sim_{CMS}}$, between graphs

| $\mathrm{Sim_{SCM}}$ | G1 | G2 | G3 | G4 |
|---|---|---|---|---|
| G1 | 1 | 0,33 | 0,66 | 0 |
| G2 | 0,33 | 1 | 0 | 0,625 |
| G3 | 0,66 | 0 | 1 | 0 |
| G4 | 0 | 0,625 | 0 | 1 |

### 5.3 The similarity measure between graphs based on UG

In applying the definition of similarity based on UG, we find the following results:

$$Sim_{UG}(G1, G2) = \frac{|SCM(G1, G2)|}{|G1| + |G2| - |SCM(G1, G2)|} = \frac{3}{8 + 9 - 3} = 0.21$$

$$Sim_{UG}(G3, G4) = \frac{|SCM(G3, G4)|}{|G3| + |G4| - |SCM(G3, G4)|} = \frac{0}{8 + 9 - 3} = 0$$

In the following array we show, $\mathrm{Sim_{UG}}$(Gi, Gj), the values of the measure of simi-

larity, based on UG, between graphs.

**Table 5.** Similarity, Simug, between graphs

| $Sim_{UG}$ | G1 | G2 | G3 | G4 |
|---|---|---|---|---|
| G1 | 1 | 0,21 | 0,66 | 0 |
| G2 | 0,21 | 1 | 0 | 0,625 |
| G3 | 0,66 | 0 | 1 | 0 |
| G4 | 0 | 0,625 | 0 | 1 |

Interpretation of results According to the definition of similarity measures $Sim_{UG}(., .)$ et $Sim_{CMS}(., .)$, the graphs that have no nodes in common, have a measure of similarity zero. This is the case for graphs G4 and G1 or G4 and G3. Both measures do not take into account a possible existence of common parent concepts to concepts of the two graphs. This is the case for graphs G3 and G4.

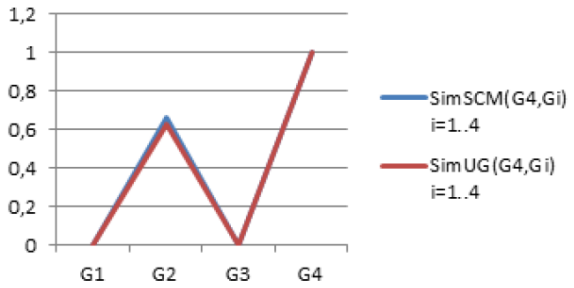


**Fig. 3.** Comparison of $Sim_{UG}$ and $Sim_{CMS}$

From these results, it is likely that these graphs have nothing in common, but that is not the case. Our new similarity measure takes into account a possible existence of parent concepts in common.

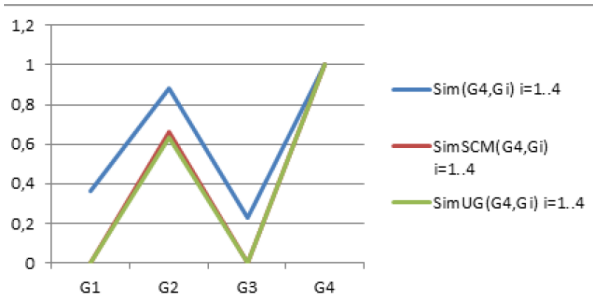


**Fig. 4.** [2]Comparison of our similarity, $Sim_{UG}$ and $Sim_{CMS}$

For example Sim(G4, G3) = 0, 2336. These one hand. On the other hand, we note that G3 is a part of G1, therefore, there is a strong similarity between these two graphs. the latter is confirmed by our measurement and less for other measures.

We have G3 is a subset of G1.

$$\mathrm{Sim}_{\mathrm{UG}}(\mathrm{G3}, \mathrm{G1}) = 0,66,$$
$$\mathrm{Sim}_{\mathrm{MCS}}(\mathrm{G3}, \mathrm{G1}) = 0,66,$$
$$\mathrm{Sim}(\mathrm{G3}, \mathrm{G1}) = 0,8412.$$

Note that the graphs G1 and G5 do not have in common nodes and have no parent nodes in common. We deduce that

$$\mathrm{Sim}(\mathrm{G1}, \mathrm{G5}) = \mathrm{Sim}_{\mathrm{CMS}}(\mathrm{G1}, \mathrm{G5}) = \mathrm{Sim}_{\mathrm{UG}}(\mathrm{G1}, \mathrm{G5}) = 0$$

## 6 Conclusion

In this work we presented a new measure of graph similarity. We compared it with the similarity measurements of graphs based on "CMS" and "UG" considered as the most used. We have demonstrated in the Evaluation section that the new measure improves the results and eliminates the problems of existing measures. Since information retrieval systems do not use a semantic similarity measure, we intend to use it in ours to increase the relevance of their results.

## 7 References

[1] S. Borzsonyi, D. Kossmann and K. Stocker. The skyline operator. In Proc. of ICDE, pp. 421 – 430, 2001. https://doi.org/10.1109/icde.2001.914855

[2] H. Bunke. On a relation between graph edit distance and maximum common subgraph. Pattern Recogn. Letters, 18 (9), 689-697, 1997. https://doi.org/10.1016/S0167-8655(97)00060-3

[3] H. Bunke, and K. Shearer. A graph distance metric based on the maximal common. subgraph. Pattern Recogn. Letters, 19 (3-4), 255-259, 1998. https://doi.org/10.1016/S0167-8655(97)00179-7

[4] D. Cai, Z. Shao, X. He, X. Yan, and J. Han. Community mining from multirelational networks. In Proc. of PPKDD, pp. 445-452, 2005.

[5] C. Chen, X. Yan, P. S. Yu, J. Han, D.-Q. Zhang, and X. Gu (2007). Towards Graph Containment Search and Indexing. In Proc. of VLDB, Vienna, Austria, pp. 926-937.

[6] K. Abbaci, A. Hadjali, L. Li′etard, D. Rocacher. Interrogation de bases de données de graphes : Une approche basée sur un skyline par similarité. Conférence Maghrébine sur l'Extraction et la Gestion des Connaissances (EGC-M), Algeria. pp.46-57, 2010. https://hal.archives-ouvertes.fr/hal-00670672/document

[7] W. D. Wallis, P. Shoubridge, M. Kraetz, et D. Ray. Graph distances using graph union. Pattern Recogn. Letters, 22 (6-7), 701-704, 2001. https://doi.org/10.1016/S0167-8655(01)00022-8

[8] X. Yan, P. S. Yu, et J. Han. Substructure similarity search in graph databases. In Proc. of ACM SIGMOD, pp. 766-777, 2005. https://doi.org/10.1145/1066157.1066244

[9] S. Zhang, M. Hu, et J. Yang. Treepi: A novel graph indexing method. In Proc. OfICDE, pp. 966-975, 2007. https://doi.org/10.1109/icde.2007.368955

[10] S. Zhang, J. Z. Li, H. Gao, et Z. Zou. A novel approach for efficient supergraph query Processing on graph databases. In Proc. of EDBT, pp. 204-215, 2009. A. Hadjali, O. Pivert, and H. Prade. Possibilistic contextual skylines with incomplete preferences. In Proc. of SoCPaR, Paris, France, 2010.
[11] Y. Tian, et J. M. Patel. Tale: A tool for approximate large graph matching. In Proc. of ICDE, Cancun, Mexico, pp. 963-972, 2008. https://doi.org/10.1109/icde.2008.4497505
[12] E.G.M. Petrakis, and C. Faloutsos. Similarity searching in medical image databases. Proc. of TKDE 9 (3), 435 -447, 1997. H. He, and A. K. Singh. Closure-tree: An index structure for graph queries. In Proc. of ICDE, pp. 38-54, 2006.
[13] X. Yan, P. S. Yu, and J. Han. Substructure similarity search in graph databases. In Proc. of ACM SIGMOD, pp. 766-777, 2005. https://doi.org/10.1145/1066157.1066244
[14] H. He, et A. K. Singh. Closure-tree : An index structure for graph queries. In Proc. of ICDE, pp. 38-54, 2006.
[15] H. Shang, K. Zhu, X. Lin, Y. Zhang, and R. Ichise. Similarity search on supergraph containment. In Proc. of ICDE, pp. 637-648, 2010.
[16] ] G. Salton and M. J.McGill, Introduction to modern information retrieval. McGraw Hill Book Co. New York, 1983.
[17] A. Labriji, I. Abdelbaki, N. Reddahi, A. Namir, A. Aboudou. Journal of Theoretical, Applied Information Technology, Vol. 83 Issue 2, p291-298. 8p, 2016.
[18] R. Rada, H. Mili, E. Bichnell et M. Blettner, Development and application of a metric on semantic nets. IEEE Transaction on Systems, Man, and Cybernetics, pp 17-30. 1989. https://doi.org/10.1109/21.24528
[19] Z. Wu et M. Palmer. Verb semantics and lexical selection. In Proceedings of the 32nd Annual Meeting ofthe Associations for Computational Linguistics, pp 133-138. 1994. https://doi.org/10.3115/981732.981751

## 8 Authors

**Amine Labriji** is a PhD Student at the TIM laboratory of Faculty of Science Ben M'sik. University Hassan II Casablanca Morocco and is a Software Technical Consultant.

**Salma Charkaoui** is a PhD Student at the TIM laboratory of Faculty of Science Ben M'sik. University Hassan II Casablanca Morocco

**Issam Abdelbaki** has a PhD degree .He is a member of the TIM laboratory and is a Software Technical Consultant at Casablanca Morocco.

**Labriji El Houssine** is a professor at the Faculty of Science Ben M'Sik. Unversity Hassan II Casablanca Morocco. Responsible for the Mathematics and Computer science and member of the TIM laboratory

**Abdelouhaed Namir** is a professor at the Faculty of Science Ben M'Sik. Unversity Hassan II Casablanca Morocco and member of the TIM laboratory

Article submitted 04 April 2017. Published as resubmitted b the authors 10 June 2017.