7th Annual International IEEE EMBS Conference on Neural Engineering
Montpellier, France, 22 - 24 April, 2015

A novel algorithm for measuring graph similarity: application to brain networks

A. Mheich, M. Hassan, V. Gripon, M. Khalil, C. Berrou, O. Dufor and F. Wendling

Abstract—**measuring similarity among graphs is a challenging issue in many disciplines including neuroscience. Several algorithms, mainly based on vertices or edges properties, were proposed to address this issue. Most of them ignore the physical location of the vertices, which is a crucial factor in the analysis of brain networks. Indeed, functional brain networks are usually represented as graphs composed of vertices (brain regions) connected by edges (functional connectivity).**

In this paper, we propose a novel algorithm to measure a similarity between graphs. The novelty of our approach is to account for vertices, edges and spatiality at the same time. The proposed algorithm is evaluated using synthetic graphs. It shows high ability to detect and measure similarity between graphs. An application to real functional brain networks is then described. The algorithm allows for quantification of the inter-subjects variability during a picture naming task.

I. Introduction

Brain functions result from the interactions between different and separated brain regions [1] often referred to as functional brain networks. These networks are usually represented by weighted graphs that consist in sets of vertices (brain regions) interconnected by edges. The weights are given by some functional connectivity measure [2]. In this context, graph theory based analysis is the best candidate to reveal the properties of these networks.

Graph theory has enormously developed in the last decades and extensive literature exist about the plethora of methods proposed to characterize graph properties [3].These measures fall into two categories. Some are related to vertex properties, such as the degree, the strength and hub. Others related to the global features of the graph such as density and modularity [4].

As compared with the large number of methods aiming at characterizing graph properties, less attention has been devoted to the methods aimed at comparing different graphs and fewer algorithms have been proposed to measure similarity. From literature review, two main categories can be identified. The first category includes algorithms aiming at detecting vertex similarities. The most obvious method in this category uses a graph-based distance [5]. Other algorithms are based on the edit distance method [6] or graph isomorphism [7]. The second category is based on edge similarities. Those can be measured using the Levenshtein distance [8] or the DeltaCon framework[9] . Here the approach is to study the similarity between two graphs with prior about vertices correspondence. However, the literature does not mention algorithms that exploit both vertex and edge similarities. Such a joint design is an essential contribution of our proposed algorithm.

In the context of brain networks, the location of vertices is a key factor for the comparison of graphs. Indeed, two graphs with identical properties but interconnecting different brain areas should be considered to have low similarity. Conversely, two graphs with dissimilar properties but interconnecting similar brain regions should be considered to be closer. A very recent study showed the importance of taking into account the 3D coordinates of the brain vertices when comparing graphs representing brain networks [10].

In this paper, we propose a novel algorithm for computing the similarity between two graphs. It uses the fact that the topological property represented by the physical location of the vertices is a crucial parameter.

On the one side, this algorithm combines the vertices (based on the edit distance method) and edges (based on Levenshtein distance) similarities and on the other side it takes into account the physical locations (the 3D coordinates) of the brain vertices. The performance of the proposed algorithm is analyzed on synthetic graphs. Then, the performance is illustrated in a real application. This application consists in detecting inter-subject variability among brain networks identified from the EEGs recorded in people who performed the same picture naming task.

II. Materials And Methods

A. Definitions

Let us consider undirected, simple, weighted graphs. A graph G is a pair of sets $G=(V,E)$, where V is the set of vertices (with known Cartesian coordinates) and $|V|$ is the order of the graph (number of vertices). $E \subseteq V \times V$ defines

* This work was supported by AZM and SAADE Association (Tripoli, Lebanon) and the Rennes University Hospital (COREC Project named conneXion, 2012-14). The work has also received a French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the "Investing for the Future" program under reference ANR-10-LABX-07-01. This work was also supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n° 290901.

A. Mheich is with AZM center-EDST, Lebanese university, and the Laboratoire de Traitement du Signal et de L'image (LTSI), UMR Inserm, Université de Rennes 1, France (corresponding author: Tel: +33 2 23 23 56 05 - Fax : +33 2 23 23 69 17, e-mail: ahmad.mheich@univ-rennes1.fr)

M. Hassan and F. Wendling are with the Laboratoire de Traitement du Signal et de L'image (LTSI), UMR Inserm, Université de Rennes 1, France (e-mail: mahmoud.hassan@univ-rennes1.fr, fabrice.wendling@univ-rennes1.fr)

M. Khalil is with AZM center-EDST, Lebanese university (e-mail: mohamad.khalil@ul.edu.lb).

V. Gripon , O. Dufor and C. Berrou are with Télécom Bretagne (Institut Mines-Télécom) and UMR CNRS Lab-STICC Brest, France, e-mail: (vincent.gripon@telecom-bretagne.eu , olivier.dufor@telecom-bretagne.eu, claude.berrou@telecom-bretagne.eu,).

the edges . The graph is said to be simple if there is no edge linking a vertex with itself. We denote by $W_{n,m}^{G}$ the weight of the edge between vertices n and m in graph G . The graph is said to be undirected if the adjacency matrix is symmetric. The similarity measure between two graphs G_1 and G_2 is denoted by $sim(G_1,G_2)$.

We introduce the matrix S ($u \times u$) as the Euclidian distance matrix between all the vertices of a graph, where u is the maximal number of vertices that can participate in one graph. Lines and columns in S represent the vertices. The values of S are the Euclidian distances between vertices coordinates, S is symmetric with zero values in the diagonal.

B. *Proposed similarity algorithm*

Let us consider two graphs $G_1(V_1,E_1)$ and $G_2(V_2,E_2)$ with possibly distinct numbers of vertices and distinct edges (see Fig. 1). Our objective is to propose a method able to provide a "distance" between G_1 and G_2. It should satisfy the following properties:

- Identity : $sim(G_1,G_1)=1$
- Symmetry: $sim(G_1,G_2)=sim(G_2,G_1)$
- Zero similarity: $sim(G_1,G_2)\to 0$ for $V\to\infty$, where G_1 is the complete graph and G_2 is the empty graph.

The proposed method is based on estimating the minimum number of transformations (deletion, insertion, substitution of vertices and edges) that maps G_2 to G_1 . More precisely, we decompose the problem into two steps:

First step: "vertex distance" between G_1 and G_2, this part can be summarized as following:

- Calculate the Euclidian distances between each pair of vertices in the grid to obtain the matrix of distances S
- Initialize the vertices distance dv between G_1 and G_2
- Compute the intersection of the sets of vertices $C=V_1\cap V_2$
- Define $V_1^{'}=V_1-C$ and $V_2^{'}=V_2-C$ as the sets of vertices that belong to one graph and do not belong to the other graph.
- Define a sphere with radius R
- Route the sphere on each vertex n of $V_1^{'}$ in G_2
- Substitute the vertex m of $V_2^{'}$ with the vertex n of $V_1^{'}$ in G_2 , if m located in the sphere, in other words the distance between n and m is less than R, the cost of substitution is equal to the Euclidian distance between n and m.
- Delete the rest of vertices that belongs to G_2 and do not substitute with any other vertex that belongs to G_1 .
- Insert vertices in G_2 that belong to G_1 and do not belong to G_2 .

The cost of substitution between two vertices is equal to the Euclidian distance between these two vertices. This cost is less than the cost of insertion or deletion of a vertex which is equal to a constant value. In our study, the cost of insertion or deletion of a vertex is equal to the maximal distance between two vertices, which is the maximal value in the matrix S .

Second step: edge distance between G_1 and G_2 :

In this step, the distance between the edges of G_1 and G_2 is computed. It consists in calculating the weight difference between two edges into two different graphs. In the example of figure 1, the graphs are unweighted, that is, their adjacency matrices are binary. We thus consider that the weight of an edge equals 1 if it exists and 0 if it does not exist.

We use the equation (1) to calculate the edge distance, the $diff(W_{n,m}^{G_1},W_{n,m}^{G_2})$ score $\in[0,1]$ where 1 means that an edge exists just in one graph and does not exist in the other, while 0 means an edge exists for both graphs between the same vertices.

$$diff(W_{n,m}^{G_1},W_{n,m}^{G_2})=|W_{n,m}^{G_1}-W_{n,m}^{G_2}| \qquad (1)$$

The distance between two graphs is then calculated by combining the vertices distance *(dv)* and the edges distance:

$$D(G_1,G_2)=dv+\sum_{n=1}^{|V|}\sum_{m=2}^{|V|}diff(W_{n,m}^{G_1},W_{n,m}^{G_2}) \qquad (2)$$

We convert the distance $D(G_1,G_2)$ to the similarity measure $sim(G_1,G_2)$ via the formula $sim=(1/(1+D))$. The similarity score $\in[0,1]$ where 0 means that G_1 and G_2 are totally dissimilar (no common vertex, no common edge), while 1 means that G_1 and G_2 are identical.

C. *Real data*

In order to assess performance of our proposed method, we used real data obtained from EEG signals measured on subjects when performing a picture naming task.

Twenty one subjects were shown pictures (n=74) on a screen using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA). They were asked to name the displayed objects. The 148 images were selected from a database of 400 pictures standardized for French [11]. The brain activity was recorded using an hr-EEG system (256 electrodes, EGI, Electrical Geodesic Inc.). EEG signals were collected at a 1 kHz sampling frequency and were band-pass filtered between 3 and 45 Hz (see [12-14] for more details). Each trial was visually inspected, and epochs contaminated by eye blinking, movements or any other noise source were rejected and excluded from the analysis performed using the EEGLAB open source toolbox. This study was approved by the National Ethics Committee for the Protection of Persons (CPP), (*conneXion* study, agreement number 2012-A01227-36, promoter: Rennes University Hospital). We excluded the electrodes located on the face as well as the few electrodes showing too high impedance.

Hassan *et al*. [2] proved that the combination of weighted Minimum Norm Estimator (wMNE) for source localization method and the Phase Locking Value (PLV) as connectivity method applied on dense EEG provides the best performance to study functional connectivity at source level.

In this application the graph G is defined as a set of vertices V representing the brain regions segmented from a Destrieux Atlas [15] and the edges E represent the functional connectivity between the EEG reconstructed sources.

III. Results

A. Synthetic Test

The performance of the algorithm is first illustrated on very simple synthetic 2D graphs. In this toy example, we generate three graphs G, G_1 and G_2 whose vertices are located on a 5x5 grid (see Fig.1). Graphs G and G_2 have 5 vertices and 4 edges and graph G_1 has 6 vertices and 5 edges. The radius R of the sphere is equal to 1, which corresponds to the minimum value of the Euclidian distance between two vertices. The similarity is equal to 0.656 between G and G_1, 0.412 between G and G_2, and 0.327 between G_1 and G_2. This can be explained by the spatial difference between G and G_2 as well between G_1 and G_2. Results show also that adding one vertex has an influence on the similarity value.

To test the sensitivity of the algorithm to variations in the location of vertices, vertices coordinates were shuffled and matrix S was recalculated. The similarity between the graphs was then obtained at each noise level. Briefly, the noise level is defined as the value added to the vertex coordinates. High value indicates a distant vertex position from the original one explained by high level noise.

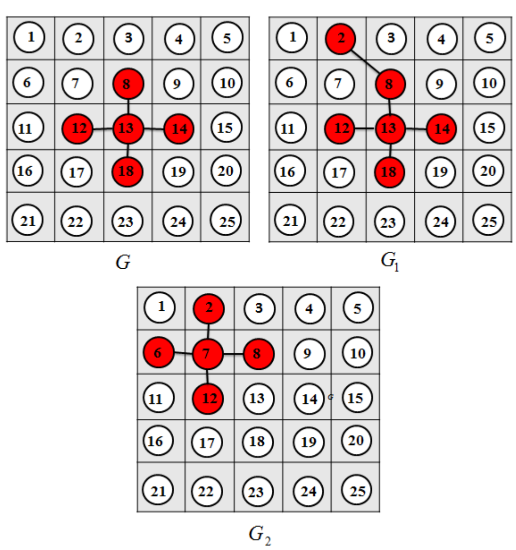

Fig 1: The three simulated graphs G, G_1 *and* G_2. Vertices of the three graphs are represented in red color and projected onto a 5x5 grid.

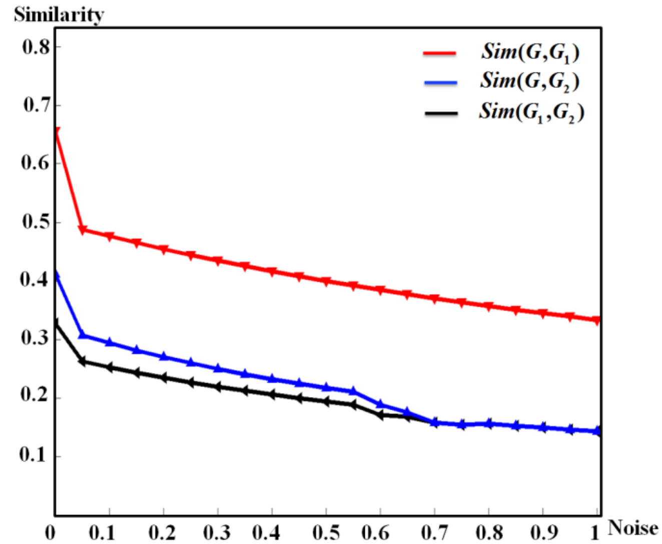

Fig 2: Variation of similarity values between G, G_1 *and* G_2 with the level of noise added to the vertices coordinates.

In Fig. 2, for each pair of graphs, the similarity is represented as a function of noise level. The curve shows a decrease in the similarity values for increasing noise level.

As the spatial position of vertices goes away from the initial locations, the similarity value decreases.

The similarity between G and G_1 represented by the red color, starts with a high similarity value 0.656 and decreases by adding noise to the vertices of graph G_1. The black color represents the similarity between G_1 and G_2, it starts with a value 0.327 and decreases by adding noise to the vertices of graph G_2.

The similarity between G and G_2 represented by the blue color, starts with a similarity value 0.412 and decreases by adding noise to the vertices of graph G_2. This explains clearly the sensibility of the algorithm to the physical position of the vertices.

B. Application to brain networks

We will now consider using our method on brain graphs computed from the 21 subjects performing a picture naming task. These graphs were calculated from High-Resolution EEG at two different periods of the process 120:200 ms (corresponding to the visual recognition) and 200:620 ms (corresponding to the semantic processing and motor response programming) as described in [2]. We used the algorithm to compute the similarity between all the graphs in order to investigate the inter-subject variability at each period. Results are illustrated in Fig 3-A-. The results show relatively high similarity values and low inter-subjects variability at 120-200 ms. The highest values were between subject 13 and 14, and between subjects 19, 20 and 21. However, results show higher inter-variability at 200:620 ms, the similarity values are very low; the highest value is equal to 0.4 between subjects 15 and 17. Typical example of brain networks with high similarity at 120:200 ms is represented in Fig.3–B- showing qualitatively high similarity between the graphs. Vertex's size represents the strength value.

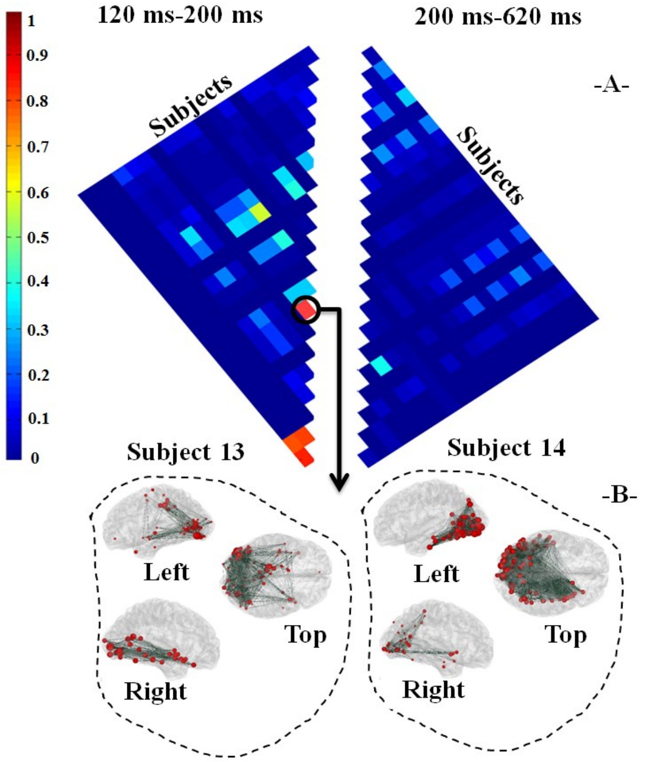

Fig 3:A- Inter-subject variability of graph similarities: the right triangle represents the similarity value of connectivity graphs between 200ms and 620 ms. The left triangle represents the similarity value of connectivity graphs between 120 ms and 200 ms. B- the connectivity graph 3D representation for subject 14 and 13 between 120ms and 200ms with different view (Left, Right and Top).

IV. Discussion

In this paper, preliminary results were presented regarding the performance of a new algorithm aiming at measuring the similarity between graphs in a context where the network topology is a key factor. One of the questions faced in the study was to specify the optimal value of the radius *R* of the sphere used to specify the zone of substitution between two vertices. Increasing the radius value may increase the similarity between two graphs. There is a compromise to find between the value of *R* and the value of the similarity between the graphs. One solution (used here) is taking the minimum distance between vertices in *S* as the *R* value. Efforts will be done for a more optimal choice of this crucial parameter. Regarding the application to brain networks, the distance between vertices was assumed to be Euclidian. However, this distance doesn't fit perfectly with the brain surface which consists of sulci and gyri (folded brain surface). The length of the shortest path between two vertices on the cortical surface (geodesic distance) would likely be more appropriate in this case. For this reason, the algorithm will be improved to use the geodesic distance instead of the Euclidian distance between vertices to measure the similarity between brain graphs. Finally, our ongoing work is to compare our algorithm with the other existed approaches. Then apply it to compare between brain connectivity graphs under different stimuli conditions. Typically, in the picture naming task, these conditions may correspond to different types of pictures (animals vs tools). The proposed method might help to assess brain categorization.

V. Conclusion

In this article, a new algorithm was proposed to detect similarity among graphs. It accounts for vertices, edges and physical location of the vertices. The performance of the proposed algorithm on synthetic graphs showed high ability to detect small shifting of the vertices location and robustness to noise added to the vertex location. Also, this new algorithm showed a high capacity to detect inter-individual variability among functional brain networks obtained from HR-EEG in subjects who performed the same picture naming task.

References

[1] McIntosh, A. R., "Towards a network theory of cognition," *Neural Networks,* vol. 13, pp. 861-870, 2000.

[2] Hassan, M., Dufor, O., Merlet, I., Berrou, C., and Wendling, F., "EEG source connectivity analysis: from dense array recordings to brain networks," *PloS one,* vol. 9, p. e105041, 2014.

[3] Costa, L. d. F., Rodrigues, F. A., Travieso, G., and Villas Boas, P. R., "Characterization of complex networks: A survey of measurements," *Advances in Physics,* vol. 56, pp. 167-242, 2007.

[4] Bullmore, E. and Sporns, O., "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nature Reviews Neuroscience,* vol. 10, pp. 186-198, 2009.

[5] Watts, D. J., *Small worlds: the dynamics of networks between order and randomness*: Princeton university press, 1999.

[6] Gao, X., Xiao, B., Tao, D., and Li, X., "A survey of graph edit distance," *Pattern Analysis and applications,* vol. 13, pp. 113-129, 2010.

[7] Ullmann, J. R., "An algorithm for subgraph isomorphism," *Journal of the ACM (JACM),* vol. 23, pp. 31-42, 1976.

[8] Cao, B., Li, Y., and Yin, J., "Measuring Similarity between Graphs Based on the Levenshtein Distance," *Appl. Math,* vol. 7, pp. 169-175, 2013.

[9] Koutra, D., Vogelstein, J. T., and Faloutsos, C., "DeltaC on: A Principled Massive-Graph Similarity Function," *SDM,* pp. 162-170, 2013.

[10] Pineda-Pardo, J. Á., Martínez, K., Solana, A. B., Hernández-Tamames, J. A., Colom, R., and del Pozo, F., "Disparate Connectivity for Structural and Functional Networks is Revealed When Physical Location of the Connected Nodes is Considered," *Brain topography,* pp. 1-10, 2014.

[11] Alario, F.-X. and Ferrand, L., "A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition," *Behavior Research Methods, Instruments, & Computers,* vol. 31, pp. 531-552, 1999.

[12] Mheich, A., Hassan, M., Dufor, O., Khalil, M., Berrou, C., and Wendling, F., "Spatiotemporal Analysis of Brain Functional Connectivity," in *6th European Conference of the International Federation for Medical and Biological Engineering*, 2015, pp. 934-937.

[13] Hassan, M., Mheich, A., Wendling, F., Dufor, O., and Berrou, C., "Graph-based analysis of brain connectivity during spelling task," in *Advances in Biomedical Engineering (ICABME), 2013 2nd International Conference on*, 2013, pp. 191-194.

[14] A. Mheich, M. Hassan, M. Khalil, C. Berrou, and F. Wendling, "A new algorithm for spatiotemporal analysis of brain functional connectivity," *Journal of Neuroscience Methods,* vol. 242, pp. 77-81, 2015.

[15] Destrieux, C., Halgren, E., Dale, A., Fischl, B., and Sereno, M., "Variability of the human brain studied on the flattened cortical surface," in *Abstr.-Soc. Neurosci*, 1998, p. 1164.