Fuzzy Logic-Based estimation of random graph generation models as an approximation for drosophila medulla structural connectome

Ksenia Shilova
Department of Computer Science
Higher School of Economics
Moscow, Russia
kashilova@edu.hse.ru

***Abstract*—Approximation of graphs obtained from real life is a very acute problem that can be solved by models of random graph generation based on some pre-known properties. One of the most essential biological graphs is the structural connectome of a fly (drosophila melanogaster), namely its visual part called optic lobe. Structural connectome is a complete description of the structure of connections in the nervous system of the organism. In this paper, some models for generating random graphs based on the properties of a structural connectome are considered. Then the random graphs are compared to the reference graph of the structural connectome according to some parameters, which make it possible to identify the most suitable of the models. In general, the parameters for evaluation can be selected depending on the interest of researchers, but the approach using fuzzy logic allows to identify the graph generation model as close as possible to the graph of the structural connectome.**

***Index Terms*—drosophila, structural connectome, medulla, random graph, fuzzy logic**

I. Introduction

Modelling and approximation of graphs obtained from real life is a huge problem for the science of our time, when complex biological graphs began to appear. A graph of the structural connectome of the drosophila fly has recently been published (December, 2020). The graph was obtained using layer-by-layer microscopy and machine learning methods for detecting the components of neurons and the connections between them [1]. Janelia Research Campus and their project FlyEm is the latest and precise development of the structural connectome. The presence of the structural connectome of such a complex insect as the drosophila fly helps to explore various properties of the brain and try to find dependencies. The geometric properties of the graph structure can greatly help to describe the structure of neurons and synapses, which is necessary in the study of various neuron systems, such as the visual one. The drosophila fly has a well-studied visual system, and understanding the work of neurons will demonstrate patterns common to many creatures.

The visual system of the drosophila fly consists of few parts, neuropiles, called retina, lamina, medulla, lobula and lobula plate. Medulla is the first neuropile where motion from the local field is detected [2,3]. That is why the study of the medulla region is critically significant for understanding the work of the visual system and particularly the motion detection circuit. Many structural properties of this region have already been studied, which makes it possible to generate random models based on them and compare their dynamic functioning or static structure with the connectome reference graph. This will make it possible to distinguish a random component from a pattern in the process of transmitting a nerve impulse. There are a lot of models for generating a random graph based on some of its previously known properties. In this work, we have selected the five most common models: Erdos-Renyi random graph, Geometric random graph, Barabasi-Albert random graph, Chung Lu random graph and Geometric Chung Lu random graph.

Fuzzy logic allows to evaluate models by some pre-selected parameters and make an aggregated assessment of the model in comparison with the connectome reference graph. On the one hand, fuzzy approach will simplify the decision-making process: which model should be used to compare the dynamic and static properties of a connectome to a random graph. On the other hand, the estimate obtained as a result of using fuzzy logic will be sufficiently aggregated and will take into account many parameters that seem to be important to the researcher.

In this paper, section 2 presents the basic definition related to fuzzy logic used as the tool of the research. Section 3 describes the parameters for estimating the similarity of graphs, as well as section 4 illustrates the concept of generating random graphs of different considered models, that were mentioned above. Section 5 provides some information about obtaining the structural connetome of drosophila medulla, that is the reference graph. Sections 6 and 7 illustrate how the experement was conducted by applying fuzzy logic. Conclusions and the expectations are provided in section 8.

II. Basic definitions

A. Linguistic variable

A linguistic variable is characterized by a 5-tuple $\langle L_V, T(L_V), U, R_{syn}, R_{sem}\rangle$, where parameter L_V is the name of the variable, $T(L_V)$ is the set formed by labels of variable's linguistic values $l_1, ..., l_n$. These names are generated using syntactic rule R_{syn}, whereas the meaning $R_{sem}(l_i)$ is associated with each value $l_i, i = \overline{1,n}$ from $T(L_V)$ by means of semantic rule R_{sem}; $R_{sem}(l_i)$ is a fuzzy set (respective membership function) defined on a universe of discourse U [4].

B. Triangular membership function

Triangular membership function is defined by a triple (a_1, a_2, a_3) in the following way:

$$\mu(x) = \begin{cases} 0, x \leq a_1 \\ \frac{1}{a_2 - a_1}x + \frac{a_1}{a_1 - a_2}, a_1 \leq x \leq a_2 \\ \frac{1}{a_2 - a_3}x + \frac{a_3}{a_3 - a_2}, a_2 \leq x \leq a_3 \\ 0, a_3 \leq x \end{cases}$$

The triangular membership function has the core of a_2 point on the axis (degree is equal to 1) and the support of the segment $[a_1; a_3]$ (not null degree value).

C. Mamdani algorithm

The Mamdani algorithm can be describe in the following way:
Firstly, The degrees of truth for the assumptions of each rule are found. Then "cut-level" for the prerequisites of each of the rules (using the min operation) are determined. Using the max operation, the found truncated functions are combined, which leads to the final fuzzy subset for the output variable with the membership function. Then, the next step is defuzzification [5].

D. Deffuzification

The process of forming a crisp number from a fuzzy function is called defuzzification. If $\mu(x)$ is an output fuzzy function, then Center Of Area method of defuzzification is defined in the following way:

$$CrispResult = \frac{\int_U \mu(x)x dx}{\int_U \mu(x) dx}$$

III. Estimated parameters

To compare two graphs we provide some characteristics that can be computed for graphs. There are parameters that are related to topological structure of a graph, that is they do not depend on weights of the edges. The main parameters independent of distances are edge density, degrees vector, maximum degree, number of triangles, clusterization coefficient and page rank vector. The characteristics that provides some information about distances between nodes are the eigenvalues of adjacency matrix and the weights distribution.

A. Edge density

The parameter called an edge density illustrates the density of graph edges. The maximum number of edges (full graph) is equal to $n(n-1)/2$, where n - number of nodes. Therefore, edge density is equal to $2|E|/(n(n-1))$, where E is the set of edges. This parameter reflects general structure of a graph and does not depend on the weights of edges.

B. Connected components

The parameter illustrates the number of connected components of a graph. This characteristic is the simplest way to concern the topological properties of a graph.

C. Degrees vector

A vector of degrees is a vector with the length n (number of nodes), where each of coordinates gives the degree of the corresponding node. Degrees distribution is an extremely important parameter that allow to build integrated estimation of a graph including edge density, is it appropriate to use a geometric component or not (that is a vertex with a higher degree is more likely to connect to a new vertex) and other properties.

D. Triangles

The number of triangles of a graph gives the information about possible clusterization process and the ability of nodes to connect with neighbours. The parameter shows a general structure of a network.

E. Clustering coefficient

Clusters are particularly significant objects related to connectomes, since connectomes can be divided into large regions with the neurons which have similar functions.

F. PageRank

The PageRank vector illustrates the importance of nodes. This parameter reflects the probability to be in the node after random walk.

G. Eigenvalues

The eigenvalues of adjacency matrices strongly correspond to the local and global properties of the network such as degree, clustering coefficient and so on. The eigenvalues provides global and complex comparing characteristic of a graph.

H. Weights

The vector of weights of the graph edges can illustrate the maximum, minimum, the most frequent and other number from the set of the weights, distances between the nodes.

IV. Description of algorithms

A. Erdos-Renyi model

The Erdos-Renyi model for generating a random graph has the input of the number of nodes and the edge density. Then, the points are randomly placed in a space and the probability that two chosen nodes have an edge is equal to the edge density [6].

B. Geometric model

The random geometric graph is the concept of N randomly distributed nodes in a space, where each two node is adjacent if the distance between them do to exceed some given threshold. The input of model is N (number of nodes) and t (threshold value)[7].

C. Barabasi-Albert model

The Barabasi-Albert model for generating a random graph get as an initial value N (number of nodes). Then, the selected M ($2 \leq M < N$, usually $M = 2$) nodes are placed in a space and each new node becomes connected with M existed nodes, besides the probability to be connected with a vertex is proportional to degree of this vertex and equal to the degree divided by the sum of degrees. It is the concept called "the rich get richer"[8].

D. Chung Lu model

The Chung-Lu model required as an input the vector of expected degrees of the nodes. Then each node is assigned a weight from this vector and the two selected node get a connection edge with the probability proportional to the product of their weights, namely, is equal to the product of weights divided by the sum of the weights[9].

E. Geometric Chung Lu model

The Simple Geometric Chung Lu model illustrate how to combine the random geometric graph model and the Chung Lu model. The input of the model is the vector of expected degrees of the nodes. The process of the graph building: points are placed independently and randomly in the space, then they are associated to the element of the nodes weights vector. Then, the probability of the existing connection between two nodes can be calculated as the product of the weights divided by the sum of the weights vector elements normalized by the distance between them[10].

V. Drosophila Medulla connectome

The reference graph of the structural drosophila medulla connectome was obtained by the python access to "neuPrint" project[1, 11-12]. It was decided to approximate the positions of the neurons by finding the average coordinates of all synapses associated with the neuron. This approximation provides sufficient accuracy, but nevertheless there is no reason to believe that the position of the neuron is reliable. In addition, the graph involves synapses located in adjacent regions of the regions of the visual system. This allows you to design the Medulla region and its connections with other regions as accurately as possible.

VI. Estimations to linguistic labels

To construct the fuzzy model it is necessary to define the correspondence of the computational estimate and the linguistic label. All of metrics, that is presented in the section 3, can be used to demonstrate the relative error of the approximation of other graph. That is we compare two graphs, however one of them is reference. Consequently, the linguistic variable indicate the quality of the graph approximation estimated by certain parameter. This is an open question, how to define the number of the linguistic terms, nevertheless in this paper we consider 7 labels as an optimal number providing sufficient accuracy. The set of labels is defined as the similarity is {extremely high, very high, high, medium, low, very low, extremely low}. As a membership function, it is convenient to define a piecewise-defined triangular function for each linguistic term. In this case, trapezoidal membership functions is the unnecessary complication of charts, because there are enough linguistic labels to get the result.

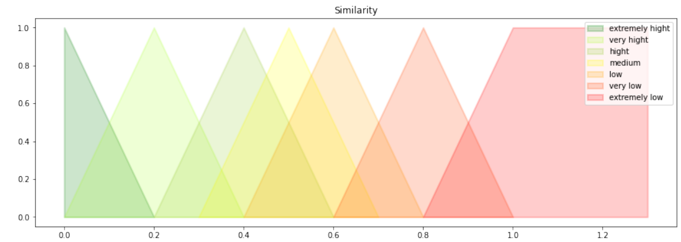

Fig. 1. Membership functions.

In fact, the relative error is measured as a float number. Then it is mandatory to define all linguistic terms corresponding to the linguistic labels.
To make a assumptions, we put that

- the similarity is **extremely** high has the core of 0 and the support from 0 to 0.2.
- the similarity is **very high** has the core of 0.2 and the support from 0 to 0.4
- the similarity is **high** has the core of 0.4 and the support from 0.2 to 0.6
- the similarity is **medium** has the core of 0.5 and the support from 0.3 to 0.7
- the similarity is **low** has the core of 0.6 and the support from 0.4 to 0.8
- the similarity is **very low** has the core of 0.8 and the support from 0.6 to 1
- the similarity is **extremely low** has the core of the segment $[1; +\infty]$ and the support from 0.8 to $+\infty$. (the only trapezoidal membership function that illustrates that all of a quite large relative errors demonstrate the extremely low similarity. In the following steps, all parameters greater than 1 will be equated to 1, that is it will be the triangular function.)

VII. Experiment

The five different models mentioned above were considered during the experiment. To start with the model generated a graph was described with python code, as well as the calculation of parameters for estimating. Then, the relative error of each approximation was calculated and fixed. There were 100 iterations of the random graph construction, which were averaged. As the main result obtained by this step, it can be considered a vector of 8 averaged estimated relative errors (for each parameter) for all generation models (Table 1).

The problem is to estimate the model according to eight pre-prepared criteria. To get the enough correct result it is needed to determine the system of rules. The linguistic labels for output estimation are the same as for the input ones.

TABLE I
AVERAGED ESTIMATED RELATIVE ERRORS

Model,Parameter	Edge Density	Connected Components	Degrees	Triangles	Clustering	PageRank	Eigenvalues	Weights
Erdos-Renyi	0.03	0.5	0.47	0.85	0.67	0.19	11.09	11.05
Geometric	0.75	3.48	0.81	0.97	0.14	0.14	0.47	1.72
Barabasi-Albert	0.7	0.5	0.73	0.99	0.67	0.54	5.55	10.47
Chung Lu	0.07	0.49	0.09	0.33	0.09	0.3	10.89	11.04
Geometric Chung Lu	0.65	0.5	0.57	1.33	0.38	0.25	12.38	10.15

Rules:
IF similarity of any parameter is **extremely high** THEN graph similarity is **extremely high**.
IF similarity of any parameter is **very high** THEN graph similarity is **very high**.
IF similarity of any parameter is **high** THEN graph similarity is **high**.
IF similarity of any parameter is **medium** THEN graph similarity is **medium**.
IF similarity of any parameter is **low** THEN graph similarity is **low**.
IF similarity of any parameter is **very low** THEN graph similarity is **very low**.
IF similarity of any parameter is **extremely low** THEN graph similarity is **extremely low**.

Then, let us provide the results for each model separately. Using the Mamdani model, it is necessary to determine the system of rules, that is what membership functions activate with the dependence of input estimation value (determined above). Then the max-min rule is applied. In other words, the minimum of membership functions degrees is chosen to be a potential the cut-off level of the activated output function. Then, the maximum of the minimums is projected on the output membership function to construct the output function. The next step is to defuzzificate the function using ”Center of Area” method (COA).

A. Erdos-Renyi model

It can be clearly seen from the charts below (Fig. 2 and Fig. 3) that the model has been interpreted as ”medium” and ”low” level of similarity with the reference graph. Speaking in natural language, the model is more likely to have ”low” similarity with the connectome graph.

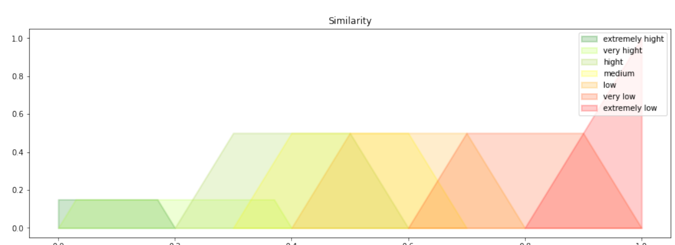

Fig. 2. Erdos-Renyi model output.

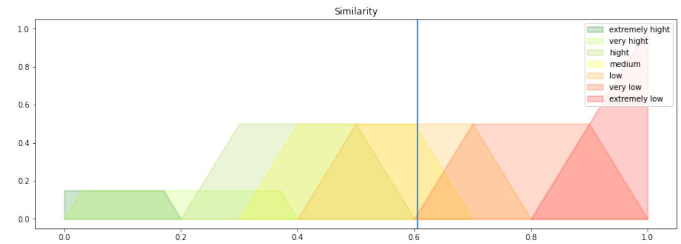

Fig. 3. Erdos-Renyi model defuzzification value 0.61.

B. Geometric model

Geometric model provides the graph with more likely ”medium” similarity with the reference graph. It also can be interpreted as ”high” and ”low” (Fig. 4 and Fig. 5).

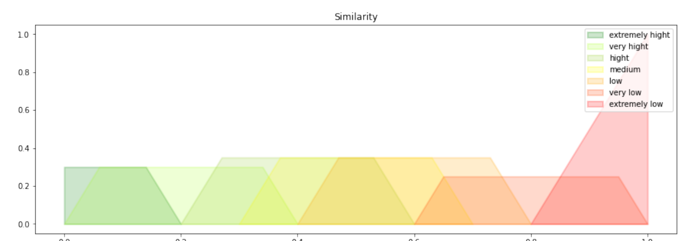

Fig. 4. Geometric model output.

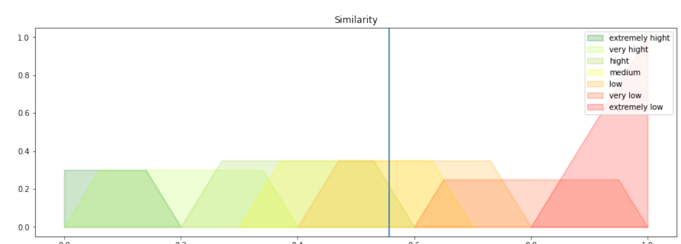

Fig. 5. Geometric model defuzzification value 0.56.

C. Barabasi-Albert model

Barabasi-Albert graph has more likely ”low” similarity with the connectome graph. The result estimation is related to ”medium” and ”very low” too (Fig. 6 and Fig. 7).

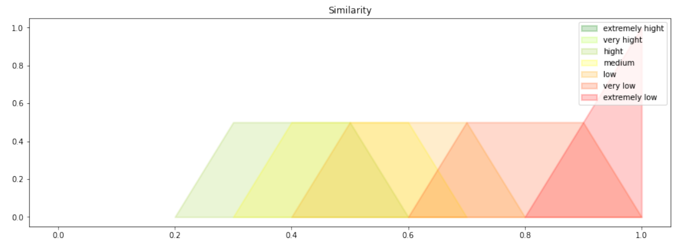

Fig. 6. Barabasi-Albert model output.

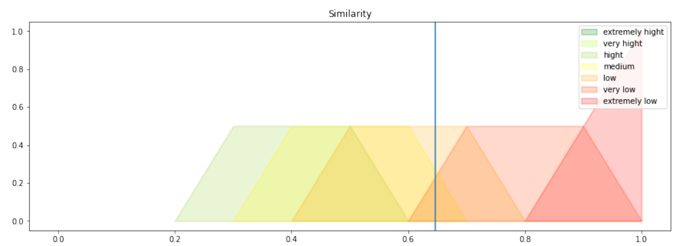

Fig. 7. Barabasi-Albert model defuzzification value 0.65.

D. *Chung Lu model*

Chung Lu model provides graph with ”high” similarity with the drosophila connectome. Moreover, it intersect with ”medium” area and slightly with ”low” area (Fig. 8 and Fig. 9).

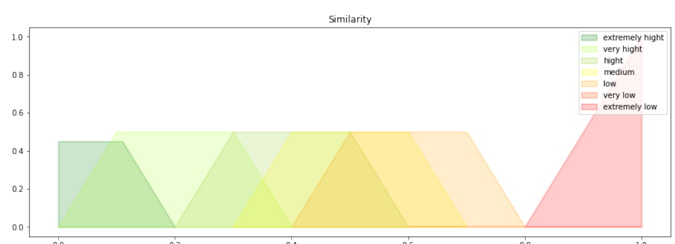

Fig. 8. Chung Lu model output.

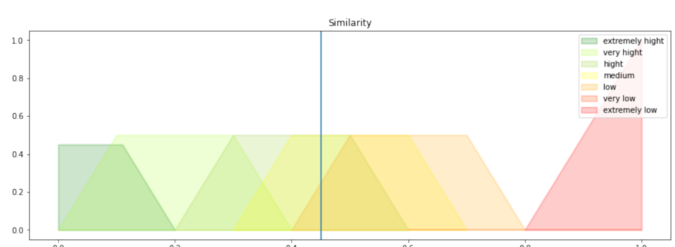

Fig. 9. Chung Lu model defuzzification value 0.45.

E. *Geometric Chung Lu model*

As well as with the geometric model, geometric Chung Lu graph can be reflect ”medium” similarity and ”high” and ”low” level of similarity (Fig. 10 and Fig. 11).

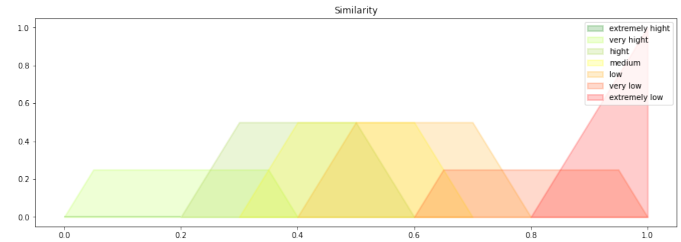

Fig. 10. Geometric Chung Lu model output.

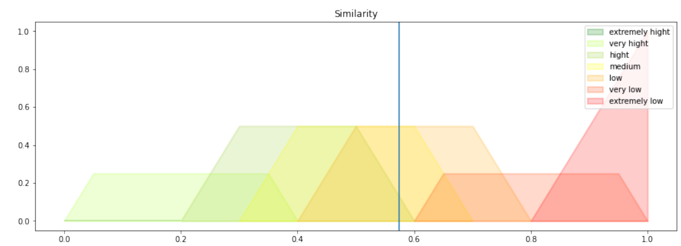

Fig. 11. Geometric Chung Lu model defuzzification value 0.57.

VIII. Conclusion

The structural connectomes and other natural graphs are an extremely significant graphs that are needed to be approximated by program modelling and properties comparing. There are large number of models to generate random graphs. Many of them have been developed based on some properties of natural graphs derived from real life, however there is no fully suitable universal solution for this problem. In this paper was considered five models: Erdos-Renyi random graph, Geometric random graph, Barabasi-Albert random graph, Chung Lu random graph and Geometric Chung Lu random graph. Applying the fuzzy logic to problem of comparing the derived graphs to the reference graph of the drosophila medulla connectome (one of the most recent obtained connectomes), the models were compared. The most suitable model according our approach is the Chung Lu model (summary defuzzification relative error is 45%). Geometric graph and Geometric Chung Lu model produced approximately the same results of 56% and 57% for relative error. Erdos-Renyi and Barabasi-Albert random graphs have the aggregate relative errors of 61% and 65% respectively.

Talking about the expectations of the results, it was clear that the Chang Lu and Geometric Chang Lu models would struggle with each other. This was expected, because the models have repeatedly shown their superiority for generating graphs close to real ones. Generally speaking, it has been shown that the distance between nodes affects the weakening of the probability of an edge appearing between them, but for some reason geometric graphs with this property turned out to be slightly worse than the simple Chang Lu model. It is possible that if only topological characteristics were evaluated, the geometric graph would demonstrate its efficiency, but in the case of complex estimates, this did not occur. Erdos-Renyi and Barabasi-Albert fulfilled expectations and were not precise enough.

REFERENCES

[1] Shan Xu, Michal Januszewski,Zhiyuan Lu, Shin-ya Takemura, 2020. A Connectome of the Adult Drosophila Central Brain. BiorXiv, the preprint server for biology.

[2] Shin-ya Takemura, Arjun Bharioke, Zhiyuan Lu, 2013. A visual motion detection circuit suggested by Drosophila connectomics. Nature.

[3] Shin-ya Takemura, 2014. Connectome of the fly visual circuitry. Microscopy.

[4] Zadeh L.A., 1996. Fuzzy logic = Computing with Words. IEEE Trans. Fuzzy Systems, 103–111.

[5] E.H. Mamdani and S. Assilian, 1975. An experiment in linguistic synthesis with a fuzzy logic controller. International Journal of Man-Machine Studies, 1-13.

[6] P.Erdos and A.Renyi. On random graphs. Publicationes Mathematicae,290-297.

[7] Penrose, Mathew, 2003. Random geometric graphs. Oxford: Oxford University Press.

[8] Barabasi, Albert-Laszlo; Albert, Reka, 1999. Emergence of scaling in random networks. Science, 509–512.

[9] Karl Bringmann, Ralph Keusch and Johannes Lengler, 2019. Geometric inhomogeneous random graphs. Theoretical Computer Science.

[10] Susama Agarwala, Franklin Kenter, 2021. A Geometric Chung Lu model and the Drosophila Medulla connectome.

[11] Janelia Research Campus/FlyEm Hemibrain: [web-site]. URL: https://www.janelia.org/project-team/flyem/hemibrain.

[12] Janelia Research Campus/Neuroprint: [web-site]. URL: https://neuprint.janelia.org.