Definition of artificial neural networks with comparison to other networks
Erkam Guresen a*, Gulgun Kayakutlu b.

*Department of Business Administration, Okan University, Istanbul 34959, Turkey
b Department of Industrial Engineering, Istanbul Technical University, Istanbul 34367, Turkey

Abstract
Definition of Artificial Neural Networks (ANNs) is made by computer scientists, artificial intelligence experts and mathematicians in various dimensions. Many of the definitions explain ANN by referring to graphics instead of giving well explained mathematical definitions; therefore, misleading weighted graphs (as in minimum cost flow problem networks) fit the definition of ANN. This study aims to give a clear definition that will differentiate ANN and graphical networks by referring to biological neural networks. The proposed definition of ANN is a mathematical definition, from the point of graph theory which defines ANN as a directed graph. Then differences between ANNs and other networks will be explained by examples using proposed definition.

Keywords: Artificial Neural Network (ANN); graph theory

1. Introduction

In literature there are no such clear and good definitions of Artificial Neural Networks (ANNs). Many of the definitions explain ANN by referring to figures instead of giving well explained mathematical definitions. The most widely definitions used in application refer to the processing units as in [1,2,3,4] and learning algorithm as in [1,2,3,4,5]. In [1,2,3,4] learning is defined as modifying synaptic weights to capture information. In [1,3] it is also added that ANNs can modify their own topology while learning.

This study aims to give a new definition that will emphasize the unique features of ANN by referring to biological neural networks. After having discussed the existing definitions in the next section, the proposed definition of ANN is demonstrated in mathematical terms in the third section. Though the definition benefits the graph theory, there is a need to differentiate ANNs and other graphs based on examples given in the fourth section. The last section is reserved for conclusion and suggestions for further research.

2. Existing definitions of ANN

A good definition of ANN, is given by Haykin [1] describing ANN as a massively parallel combination of simple
processing unit which can acquire knowledge from environment through a learning process and store the knowledge in its connections.

ANN definitions [2, 3, 4, 5] emphasize Processing Elements (PE) and the learning algorithm. Rojas [4] noted that we still do not fully understand the computing mechanism of a biological neuron and that is why we prefer PE (or computing units) instead of artificial neuron. Learning is defined as modifying synaptic weight to capture information in [1, 2, 3, 4]. In [1] and [3] it is also noted that ANN can modify its own topology.

PEs are inspired by the existing neurons in animal nerve system. Real neurons get stimulus and change them via synaptic weights, combine them, and lastly produce a single response (output) dissimilar to the combination. In a similar way, [1] identified three basic elements of a PE as synaptic weight, summing function to combine input with respect to corresponding weight, and an activation function to produce an output. In [2] output of a PE is defined as a function of function, in which a summation is performed to combine the input, than an activation function is used to calculate the output. Similarly, in [4] it is noted that four structures (dendrites, synapses, cell body and axon) of biological neuron shape the minimal structure we would adopt from biological models.

In [4] a pragmatic definition is used for ANN as follows:

“ANNs are distributed, adaptive, generally nonlinear learning machines built from many different processing elements (PEs). Each PE receives connections from other PEs and/or itself. The interconnectivity defines the topology. The signals flowing on the connections are scaled by adjustable parameters called weights, wij. The PEs sum all these contributions and produce an output that is a nonlinear (static) function of the sum. The PEs’ outputs become either system outputs or are sent to the same or other PEs.”

To make things more comprehensible, definition of graph is taken as the starting point. Geometrically a graph is a set of points (vertices or nodes) in space which are interconnected by a set of lines (edges or links) [7]. And the weighted graph can be defined as follows: a weighted graph is a graph in which a number is assigned to for any edge e [7]. A directed graph or shortly digraph is a graph whose edges have a direction and edges are represented by arrows showing the direction [7]. And lastly we should define connectivity in digraphs as follows: \( v_i \) is connected to \( v_j \) if there is a path from \( v_i \) to \( v_j \) [7].

In [1] ANN is mentioned as a directed graph, in which three graphical representations are used for defining a neural network:

- with a block diagram to describe the network functionally,
- with signal-flow graph to describe signal flow in the network,
- with the architectural graph to describe the network layout.

Description of ANN in [1] as a directed graph is not complete as long as it excludes the learning process, input and output sets (number of input or output records and number of attributes of inputs and outputs) and parallel structure.

An interesting and highly mathematical definition we found in [5]. In [5] ANN is defined from the point of graphs as follows:

“A neural network model is defined as a directed graph with the following properties:
- A state variable \( n_i \) is associated with each node \( i \),
- A real valued weight \( w_{ik} \) is associated with each link \( (ik) \) between two nodes \( i \) and \( k \),
- A real valued bias \( v_i \) is associated with each node \( i \),
- A transfer function \( f(n_i, w_{ik}, v_i, (i\neq k)) \) is defined for each node \( i \), which determines the state of the node as a function composed of its bias, the weights of incoming links, and the states of nodes connected to it.”

In [5] input nodes are defined as the nodes with no incoming link and output nodes as the nodes with no outgoing links. This definition contains some problems like not containing recurrent neural networks, in which output of each neuron can be its input. Thus definition in [5] cannot point out input and output nodes clearly.

Another problem with definition in [5] is that it does not contain input and output nodes (or layers) in the definition of ANN. Clearly an ANN should have some input neurons and some output neurons with specific features that are not applicable for all other graphs. Also in [5] parallel distribution of nodes is not mentioned and similar to definition in [1] it does not mention about a learning algorithm. Those will cause confusions with other graphs.

3. Proposed Artificial Neural Network Definition

Common character of all the ANN definitions in literature is the comparison with biological neural networks [1,
The inspirations of similarities are summarized in Table 1. It is noted that receptors are specialized neurons for gathering specific information from environment and that the neural net generally refers to brain where effectors are the specialized neurons for evoking the specific tissues.

### Table 1. Similarities between biological neural networks and artificial neural networks

| Biological Neural Networks | Artificial Neural Networks |
|---------------------------|---------------------------|
| Stimulus                  | Input                     |
| Receptors                 | Input Layer               |
| Neural Net                | Processing Layer(s)       |
| Neuron                    | Processing Element        |
| Effectors                 | Output Layer              |
| Response                  | Output and an entry       |

Activities of biological neurons and processing elements of ANN can be compared as in Table 2. Briefly, synapses act like a weight of the incoming stimulus and inspired the weights of ANN; dendrites accumulate the incoming weighted stimulus, inspired the summing function of ANN; cell body, that causes conversion of summed stimulus to a new stimulus, inspires activation function; axon, which distributes the new stimulus to the corresponding neurons, inspires the output and output links; and lastly, threshold value with a role of activating or inactivating increase and decrease of the stimulus, inspires the bias. All four structures mentioned in [4] (dendrites, synapses, cell body and axon) are necessarily contained in processing elements.

### Table 2. Similarities of neurons and Processing Elements (PEs)

| Neurons     | Processing Elements (PEs) |
|-------------|---------------------------|
| Synapses    | Weights                   |
| Dendrites   | Summing Function          |
| Cell Body   | Activation Function       |
| Axon        | Output                    |
| Threshold value | Bias                     |

In the light of above characterization and inspiration we can enrich the definition by mentioning a network which is made up of massively parallel processors with connections. A clear definition of processors will differentiate an artificial neural network with unique features. In general, nodes in a graph could be considered as PEs with identity function, which returns the same input as output. A complete definition of ANN from the point of graphs is suggested to include the features give in the following definitions.

**Definition 1.** A directed graph is called an *Artificial Neural Network (ANN)* if it has
- at least one *start node* (or *Start Element; SE*),
- at least one *end node* (or *End Element; EE*),
- at least one *Processing Element (PE)*,
- all the nodes used must be *Processing Elements (PEs)*, except *start nodes* and *end nodes*,
- a state variable $n_i$ associated with each node $i$,
- a real valued weight $w_{kj}$ associated with each link $(ki)$ from node $k$ to node $i$,
- a real valued bias $b_i$ associated with each node $i$,
- at least two of the multiple PEs connected in parallel,
- a *learning algorithm* that helps to model the desired output for given input.
- a flow on each link $(ki)$ from node $k$ to node $i$, that carries exactly the same flow which equals to $n_i$ caused by the output of node $k$,
- each start node is connected to at least one end node, and each end node is connected to at least one start node,
- no parallel edges (each link $(ki)$ from node $k$ to node $i$ is unique).
The definition of Artificial Neural Networks will be complete when we define Start Element (SE), End Element (EE), Processing Element (PE) and Learning Algorithm as follows.

**Definition 2.** Start Element (SE) \( k \) is a node in a directed graph, which gets an input \( I_k \) from the input matrix \( I=[I_{ij}; i=1,2,\ldots,n, j=1,2,\ldots,m] \) of \( n \) attributes of \( m \) independent records, and starts a flow in the graph.

**Definition 3.** End Element (EE) \( i \) is a node in a directed graph, which produces an output \( O_i \) from the output matrix \( O=[O_{ij}; i=1,2,\ldots,n, j=1,2,\ldots,m] \) of \( n \) desired outputs of \( m \) independent input records and ends a flow in the graph.

**Definition 4.** Let \( G \) be a directed graph with the following properties;
- a state variable \( n_i \) is associated with each node \( i \),
- a real valued weight \( w_{ki} \) is associated with each link \( (k,i) \) from node \( k \) to node \( i \),
- a real valued bias \( b_i \) is associated with each node \( i \),
- has no parallel edges (links).

Let \( f[n_i,w_{ki},b_i(i\neq k)] \) be the following function in graph \( G \) for node \( i \);

\[
f_i(u_i,b_i)=n_i=q(u_i+b_i)
\]  \( (I) \)

where \( q(.) \) is the activation function and \( u_i \) is as follows;

\[
u_i = \sum_{j=1}^{m} w_{ji} n_j
\]  \( (2) \)

where \( j \) is the node which has a link to node \( i \) and hence node \( i \) is called a Processing Element (PE).

**Corollary 1.** In a directed graph, each node can be considered as PE with (if not specially assigned) \( w_{ki}=1, b_i=0 \) and \( \phi(.)=I(.) \), where \( I(.) \) is the identity function. With these properties flow does not change at nodes.

Hence, we can briefly define PE as a node with functions, constructed by the state of the node, weights of its incoming links, and the bias of weights.

**Definition 5.** Learning Algorithm in an ANN is an algorithm which modifies weights of the ANN to obtain desired output(s) for given input(s).

**Hint 1.** Desired output can be exactly known values, number of desired classes or some pattern expectations for certain input sets. Therefore, “desired output” term contains output of supervised, unsupervised or reinforced learning.

**Hint 2.** Note that every element \( k \) (SE, PE and EE) can generate only and only one output value at a time. But every element \( k \) can send the same output value to another element \( i \) with no restriction if there is a link (edge) from \( k \) to \( i \).

The suggested definition of an ANN is different from any other kind of graph and is strong enough to avoid previous conflicts. First of all, it is a network which has specific starting and ending nodes. This new start and end node (element) definitions do not contradict with recurrent neural networks as in [5]. By differentiating the nodes as PEs, SEs or EEs, components of an ANN are clarified. By describing the variables and parameters associated with each node and link contained in a graph, confusions are void.

Besides, massively parallel structure makes an ANN more biological based than computer based. Structures containing some SEs and EEs with one or more PEs connected serially cannot be referred as an ANN because it will lose the power of parallel computing and starts to act more like existing computers than a brain. Good explanations of parallel and serial computing can be found in [1]. Shortly it can be said that parallel computing is powerful in complex calculations and mappings; serial computing is powerful for arithmetic calculations [1]. Also serial structure cannot contain fault tolerance property. Thus a damage or corruption in serial structure will cause catastrophic failures but in biological neural networks death of a neuron does not cause catastrophic failures of the neural network.

**Definition 6.** Adapitivity is the ability to adapt changes in the environment.

Adapitivity refers to retraining of ANN with only the new data set when it is available. Thus once an ANN model is formed, there is no need to built a new ANN model with the modified new data. (it can be referred as
A graph that does not have the ability to learn (or to be corrected through its weights) cannot be mentioned as an ANN, because it cannot contain adaptivity property. If a graph cannot be taught, environmental changes cause the necessity for a new graph to represent the environment instead of modifying the existing graph through learning.

The flow in an ANN is specific to the network designed since every outgoing link carries the same flow \( n_i \) produced as output of the node \( i \). For this reason proposed definition contains details of flow to avoid confusion. To be clear let us take an example of a node in a specific ANN; when a flow comes to node \( k \) (can be SE, EE or PE), node \( k \) generates an output \( n_k \). The output \( n_k \) is send to all nodes \( i \) if an edge exists from node \( k \) to node \( i \). In other words each edge from node \( k \) to node \( i \), duplicates the output value \( n_k \) and carries as a flow. But in many other graphs each edge from node \( k \) to node \( i \) carries some portion of the output value \( n_k \) as a flow, in such a way that their sum is equal to \( n_k \).

The proposed definition also contains restrictions about connectivity of an ANN, through which the input and the output are mapped. This subject will be explained as the output must be obtained from given input variables. If an SE is not connected to at least one of the EEs, than this means it does not have an effect on output. In a similar way if an EE is not connected to at least one of the SEs, than this means it is not effected by any of the input.

4. Evaluation of ANN Definitions

In this section ANN definitions of Principe et al. [6], Muller et al. [5] and the proposed ANN definition will be used to evaluate whether the existing graph is an ANN or not. These definitions are chosen for evaluation because of the clearity as well as widely accepted respect.

Consider the following example; a graph with its structure given in Fig 1, it is a directed graph with ten nodes. Let us assume that we have a suitable input data set and a suitable learning algorithm, and each node has a suitable function to process the coming flow which carries exactly the same flow on each link \((k_i)\) from node \( k \) to node \( i \), equals to \( n_k \) caused by the output of node \( k \). In such case the graph in Fig 1 is an artificial neural network according to the proposed definition, Principe et al.’s (1999) definition and Muller et al.’s (1996) definition. Clearly it is an ANN with one input layer, one output layer and two hidden layers.

![Fig. 1. Example graph structure](image)

When we change the rule for the flow in this graph as sum of the links \((k_i)\) from node \( k \) to node \( i \), equals to \( n_k \) caused by the output of node \( k \), than the graph in Fig 1 cannot be mentioned as an ANN according to the proposed definition but it is still an ANN according to [5, 6].
If we remove the nodes 1, 3, 4, 6, 7, 9 and the corresponding links from the example given in Fig 1, we will get the graph given in the Fig 2. This is a serial connected graph whose output is generated by the following function:

\[ f(I) = \varphi_{10}(u_{10} + b_{10}) = \varphi_{10}(\varphi_{8}(u_{8} + b_{8}) + b_{10}) = \varphi_{10}(\varphi_{5}(u_{5} + b_{5}) + b_{7} + b_{8} + b_{10}) \]  

(3)

where \( \varphi(.) \) is the activation function, \( b_i \) is the bias and \( u_i \) is the weighted sum of incoming flow defined as in Eq. (2). This is a single function, which behaves more like a computer processor than an ANN. Parallel distributions of nodes are important for having ANN properties such as fault tolerance, pointed in [1]. From the point of proposed ANN definition it is not an ANN because it has no parallel nodes, but it is an ANN according to [5] and [6].

For the example given in Fig 1, if no suitable learning algorithm exists, then it is not an ANN according to the proposed definition and [6] but it is an ANN according to [5]. But if the given example has a suitable learning algorithm but has no suitable input data set it is an ANN according to [5] and [6] but not an ANN according to the proposed definition.

Consider the case if the input layer and the corresponding links are removed from the example given in Fig 1. In such case the example is still an ANN according to [5] and [6] but not an ANN according to the proposed definition. But instead of removing the input layer, if the output layer is removed, the underlying graph is not an ANN according to [6] and the proposed definition but it is an ANN according to [5] since no constraint mentioned about having outputs in definition.

To clarify the proposed definition, consider the following example inspired from structure in [10]; let Bosphorus City (BC) have 3 electricity supplies; Saint Village Power Station (SVPS), Hook Village Power Station (HVPS), and Borealis Village Power Station (BVPS). Electricity comes to Bosphorus City through energy lines, crossing one of the three villages: Middle Village (MV), Black Village (BV) or Wolf Village (WV). Energy lines only transfers energy only one way. Possible routes and lengths for energy lines are shown in the Fig 4. Currently no energy line exists but each energy line will be constructed from a metal mixture of copper and silver. Copper is cheaper but cause much energy lost through the lines. So governor of Bosphorus City will decide the mixture of each energy line according to the city budget constraints. But in future, this mixture will be a constant weight for energy amount transferred through each energy line.
Each of SVPS, HVPS, BVPS, MV, BV, WV and BC has transformer system with constant operating cost of $b_i$. Since the loss of energy depends on the metal mixture of energy line and its length (weight - $w_{ij}$) and amount of the energy. Thus cost of water in each node is a function of weighted sum of incoming energy and the constant operating cost like given in equation X1.

The underlying graph of this problem has nodes, which produces water-cleaning costs as output water amount as flow, pipelines as links, metal mixture and length of pipelines as weights. Adjusting weights (mixture of metal for each pipeline) can be considered as learning since no restriction exists in the definition. Thus according to [5] and [6], this graph is an artificial neural network. Clearly it is not an artificial neural network and does not fit the proposed ANN definition in many ways. First off all it does not have specific SEs which gets an input from input data set and starts a flow in graph. Secondly it does not have specific EEs, which produces a desired output for each input record. Existing nodes do not contradict with existing PE definition since they produce outputs from a function of weighted sums. Weights, state variables and bias can be assumed to exist. Some of the nodes are distributed in parallel so this does not contradict with the proposed definition. But water flow is not a flow as described in the proposed definition. Another point is that output of each node is a cost but these costs are not carried as flow, instead, water amount is the flow in the graph. Since [5] and [6] does not specify the flow in the definition, it seems to be normal. But in the proposed definition it is specified that flow should be equal to output of the node and all the links beginning from that node should carry exactly the same output. This underlying graph of the Bosphorus City water supply problem is not an artificial neural network at all, and the proposed ANN definition clearly differentiates it from ANNs.

5. Conclusion

This study is driven by a conflict of mathematical definitions on ANN. In depth analysis of literature on some newly developed ANN architectures has lead us through the advantages of this new methods with respect to the statistical or heuristic methods. Nevertheless, the confusion arose by conflicting utilization and inconsistent findings based on different properties of ANN. In order to continue research a robust definition of all the features had to be accumulated and the observations were to be clearly demonstrated.

This paper is prepared as a guide to researchers who would like to use ANN mathematics; understand the structure of ANN and differentiate it from other methods. That is why, each essential condition is handled in detail by comparing the definitions by distinguished scientists. Integrating a variety of defined features of ANN is demonstrated to include graph theory, mathematics and statistics.

Once the concept is clearly defined, examples of casual utilization of ANNs will be easier to understand. It is our hope that respected ANN researchers benefit the definition and use as a step for the future studies. Evolutions on the ANN can be studied now.
References

1. S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, New Jersey, 1999.
2. R. Eberhart and Y. Shi, Computational Intelligence, Morgan Kaufmann 2007.
3. K. Gurney, An Introduction to Neural Networks, CRC Press, 2003.
4. R. Rojas, Neural Networks: A Systematic Introduction, Springer-Verlag Berlin Heidelberg, 1996.
5. B. Muller, J. Reinhardt, M.T. Strickland, Neural Networks an Introduction, Springer-Verlag Berlin Heidelberg, 1995.
6. J.C. Principe, N.R. Euliano, W.C. Lefebvre, Neural and Adaptive Systems: Fundamentals Through Simulations, John Wiley & Sons, New York, 1999.
7. A. Gibbons, Algorithmic Graph Theory, Cambridge University Press, 1999.
8. P.J. Braspennin, F. Thuijsman, A.J.M.M Weijters, Artificial Neural Networks, Springer-Verlag Berlin Heidelberg, 1995.
9. DARPA Neural Network Study, AFCEA International Press, 1992.
10. D. Çınar, G. Kayakutlu, T. Daim, Energy 35:4 (2010) 1724-1729.