

A Survey on Complex Networks

Angel Garrido

Faculty of Sciences, National University of Distance Education, Madrid, Spain
Paseo Senda del Rey, 9, 28040, Madrid, Spain
algbmv@telefonica.net

Abstract

Our purpose with this paper may be clearly established: to show many essential aspects of such active topic of research. For this objective, previously we may analyze here some relevant features strongly supported on Graph Theory (the modern Theory of Graphs and of Complex Networks).

Keywords: artificial intelligence, complex networks, graph theory, discrete mathematics

1. Introduction

Complex Networks are everywhere [1, 3]. This can be considered an important tenet of current science. And the language of such networks is precisely Graph Theory [16], which represents by nodes the different states or situations, and by edges connecting them their mutual relationships. Therefore, it is essential to learn these new codes and rules, allowing us to understand the new scientific challenges.

Almost all the tools necessary to represent and solving problems by Graph Theory shown appealing features, very appropriated to develop in student's mathematical abilities, by reasoning and developing very intuitive and useful methods. So, enhancing their motivation and increasing their enthusiasm with respect to Mathematics. Because it provides a fascinating sample of beautiful and visual mathematical objects, disposable to any people with a minimal theoretical basis. It will be very important in problem solving.

In the words of Georges Pólya "we expect that any symmetry found in the data and conditions of the problem will be mirrored by the solution" [20].

2. Graphs

Graphs are mathematical objects [14, 16] very frequently used in Computer Science, and other important fields. Because often we can reduce a real-world problem to a mathematical statement about graphs. Hence, if the graph problem is solved, then the real-world problem it is also solved.

A graph is a pair, $G = (V, E)$, where V is a set of points, called nodes, and E will be the set of their edges, or links, between nodes. Sometimes, it is denominated simple graph. Note that a simple graph represents a symmetric relation, R , according which two nodes, a and b , being connected by an edge are related to each other in both directions, i.e. $a R b$ and $b R a$.

Many times, we said Undirected Graph (UG, in acronym), instead of simple graph.

V and E are usually taken to be finite, because some of the well-known results can fail for infinite graphs. It is due that many times the arguments are not applicable in such infinite case.

The *order* of a graph is the number of their nodes. That is, the cardinality of the set V , sometimes denoted by $|V|$, also by $\text{card}(V)$, or $c(V)$.

The *size* of a graph is the number of their edges. That is, the cardinality of the set E , sometimes denoted by $|E|$.

The *degree* of a graph is the number of edges that connect to it.

Note that an edge that connects to the node at both ends (therefore, a loop) is counted twice.

Let $G = (V, E)$ be a graph. Then, we see as its complement a new graph, with the same set of nodes, but its set of edges contains the complement of E . So, we consider G as a graph.

It will be very interesting to give some previous concepts, as that of line graph, vertex-transitivity, edge-transitivity, and so on.

The so-called *line graph* of G will be the graph whose set of nodes is E (therefore, it coincides with the set of edges of G), and whose edges connect all pairs of E which have one common end (or extremity) in G . Usually, it is abridged by $L(G)$. Hence, the Line Graph of G is another graph that represents the adjacencies between edges of G . The Line Graph is sometimes called Adjoint Graph, Interchange Graph, Edge Graph, Derived Graph of G , and so on.

A graph, G , is said to be *node-transitive* (or vertex-transitive), if for any two of its nodes, there is an automorphism which acts between them.

A simple graph, G , is said to be *edge-transitive* (or link-transitive), if for any two of its edges, e and e' , there is an automorphism which maps e into e' .

A simple graph, G , is said to be symmetric, when it is both, node-transitive and edge-transitive.

But a simple graph, G , which is edge-transitive, but not node-transitive, is said *semi-symmetric*.

Obviously, such a graph will be necessarily a bipartite graph.

An auxiliar concept, which many times appears at Graph Theory problems, will be the Chordality.

Let G be an undirected graph (UG). We says that G is *chordal*, if every cycle of length greater than three possesses a "chord". This name ("chord") means an edge joining two non-consecutive nodes of the cycle. Therefore, an UG will be *chordal*, if it does not contain an induced subgraph isomorphic to the C_n , when $n > 3$.

The *chordality* result a hereditary property, because all the induced subgraphs of a chordal graph will be also *chordal*. For instance, the interval graphs are chordal.

3. Graph Symmetry

As we known, Symmetry into a system means invariance of its elements under a group of transformations [25]. When we take Network Structures [23, 24], it means invariance of adjacency of nodes under the permutations on node set.

Let G and H be two graphs [16]. An *isomorphism* from G to H will be a bijection between the node sets of both graphs, i. e. an application, $f: G \rightarrow H$, such that any two nodes, u and v , of G are *adjacent* in G if and only if $f(u)$ and $f(v)$ are also adjacent in H . Usually, it is called "edge-preserving bijection".

If an isomorphism exists between two graphs, G and H , then such graphs are called *Isomorphic Graphs*.

The graph isomorphism is an equivalence, or equality, as relation on the set of graphs. Therefore, it partitions the class of all graphs into equivalence classes.

The underlying idea of isomorphism is that some objects have the same structure, if we omit the individual character of their components.

A set of graphs isomorphic to each other is denominated an isomorphism class of graphs.

An *automorphism* of a graph, $G = (V, E)$, will be an isomorphism from G onto itself. So, a graph-automorphism of a simple graph, G , is simply a permutation on the set of its nodes, $V(G)$, $f: G \rightarrow G$, such that the image of any edge of G is always an edge in G . That is, if $e = \{u, v\} \in V(G)$, then

$$f(e) = \{f(u), f(v)\} \in V(G)$$

Either expressed in group theoretical way, we have

$$u \sim v \Leftrightarrow ug \sim vg \Leftrightarrow u^g \sim v^g$$

being ug and vg the corresponding images of u and v under the permutation g .

The family of all automorphisms of a graph G is a permutation group on $V(G)$. The inner operation of such group is the composition of permutations. Its name is very well-known, the *Automorphism Group* of G , and abridgedly, it is denoted by $\text{Aut}(G)$.

And conversely, all groups may be represented as the automorphism group of a connected graph.

The automorphism group is an algebraic invariant of a graph. So, we can say that an automorphism of a graph is a form of symmetry in which the graph is mapped onto itself while preserving the edge-node connectivity. Such automorphic tool may be applied both on Directed Graphs (DGs) and on Undirected Graphs (UGs).

About another interesting concept in Mathematics, the word "genus" has different, but very related, meanings. So, in Topology, it depends on to consider orientable or non-orientable surfaces.

In the case of connected and orientable surfaces, it is an integer that represents the maximum number of cuttings, along closed simple curves, without rendering the resultant manifold disconnected. For this reason, it is the number of "handles" on it. Usually, it is denoted by the letter g .

It will be also definable through the Euler number, or Euler Characteristic, denoted by χ . Such relationship will be expressed, for closed surfaces, by

$$\chi = 2 - 2g$$

When the surface has b boundary components, this equation transforms to

$$\chi = 2 - 2g - b$$

which obviously generalizes the above equation. For example, a sphere, an annulus, or a disc have genus $g = 0$. Instead of this, a torus has $g = 1$.

In the case of non-orientable surfaces, the genus of a closed and connected surface is a positive integer, representing the number of cross-caps attached to a sphere.

Recall that a cross-cap is a two-dimensional surface that is topologically equivalent to a Möbius string.

As in the precedent analysis, it can be expressed in terms of the Euler characteristic, by

$$\chi = 2 - 2k$$

being k the non-orientable genus. For example, a projective plane have non-orientable genus $k = 1$. And a Klein bottle has a non-orientable genus $k = 2$.

Turning to graphs [16], its corresponding genus will be the minimal integer, n , such that the graph can be drawn without crossing itself on a sphere with n handles. So, a planar graph has genus $n = 0$, because it can be drawn on a sphere without self-crossing.

In the non-orientable case, the genus will be also the minimal integer, n , such that the graph can be drawn without crossing itself on a sphere with n cross-caps.

If we pass now to topological graph theory, we will define as *genus of a group*, G , the minimum genus of any of the undirected and connected Cayley graphs for G .

From the viewpoint of the Computational Complexity, *the problem of "graph genus" is NP-complete*.

We will say either graph invariant or graph property, when it depends only of the abstract structure, not on graph representations, such as particular labeling or drawings of the graph. So, we may define a graph property as every property that is preserved under all its possible isomorphism of the graph. Therefore, it will be a property of the graph itself, not depending on the representation of the graph.

The semantic difference also consists in its character quantitative or qualitative. For instance, when we said "the graph does not possess directed edges", this will be a property, because it is a qualitative statement. While when we said that "the number of nodes of degree two in such graph", this would be an invariant, because it is a quantitative statement.

From a mathematically strict viewpoint, a graph property can be interpreted as a class of graphs, composed by the graphs that have in common the accomplishment of some conditions.

Hence, also can be defined a graph property as a function whose domain would be the set of graphs, and its range will be the bivalued set composed by two options, true and false, $\{T, F\}$, according which a determinate condition is either verified or violated for the graph.

A graph property is called hereditary, if it is inherited by its induced subgraphs. And it is additive, if it is closed under disjoint union.

For example, the property of a graph to be planar is both additive and hereditary. Instead of this, the property of being connected is neither.

The computation of certain graph invariants may be very useful, with the purpose to discriminate when two graphs are isomorphic, or rather non-isomorphic.

The support of these criteria will be that for any invariant at all, two graphs with different values cannot be isomorphic between them. But however, two graphs with the same invariants may or may not be isomorphic between them. So, we will arrive to the notion of completeness.

Let $I(G)$ and $I(H)$ be invariants of two graphs, G and H . It will be considered complete, if the identity of the invariants ever implies the isomorphism of the corresponding graphs, i. e.

$$I(G) = I(H) \Rightarrow G \approx H.$$

A directed graph, or digraph, is the usual pair $G = (V, E)$, but now with an additional condition: it have at most one directed edge from node i to node j , being $1 \leq i, j \leq n$. We add the term acyclic when there are no cycles of any length. Usually, we use the acronym *DAG* to denote an *acyclic directed graph*.

A very important result may be this:

For each n , the cardinality of the n -DAGs, or DAGs with n labeled nodes, is equal to the number of $(n \times n)$ -matrices of 0's and 1's whose eigenvalues are positive real numbers, i. e.

$$\lambda_i \in \mathbf{R}^+, \text{ with } 1 \leq i \leq n$$

A previous result said that a digraph contains no cycle if and only if all eigenvalues of its adjacency matrix are equal to zero.

It is possible to prove that every group is the automorphism group of a graph. If the group is finite, the graph may be taken to be finite.

And George Pólya observed that not every group is the automorphism group of a tree.

4. Enumerating Digraphs

Among the different types of graphs [15], Bayesian Networks (BNs, in acronym) are the most successful class of models to represent uncertain knowledge. But the representation of conditional independencies (CIs, also in acronym) does not have uniqueness. The reason is that probabilistically equivalent models may have different representations.

And this problem is overcome by the introduction of the concept of *Essential Graph*, as unique representation of each equivalence class. They represent Conditional Independence Models by graphs. Such mathematical and graphical tools containing both, directed or/and undirected edges; hence, producing respectively Directed Graphs (DGs), in particular acyclic elements, or Directed Acyclic Graphs (DAGs), either Undirected Graphs (UGs), or Chain Graphs (CGs), in the mixed case. So, DAG models are generally represented as Essential Graphs (EGs).

Knowing the ratio of EGs to DAGs is a valuable tool, because through this information we may decide in which space to search. If the ratio is low, we may prefer to search the space of DAG models, rather than the space of DAGs directly, as it was usual until now.

The most common approach to learning DAG models is that of performing a search into the space of either DAGs or DAG models (EGs).

It is preferable, from a mathematical point of view, to obtain the more exact solution possible, studying its asymptotic behavior. But also it is feasible to propose a Monte Carlo Chain Method (MCMC) to approach the ratio, avoiding the straightforward enumeration of EGs. And a many more elegant construct, if very difficult, through the Ihara Zeta function for counting graphs.

Recall that a DAG, G , is essential, if every directed edge of G is protected. So, an Essential Graph (EG) is a graphical representation of a Markov equivalence class.

In relation with the Essential Graph, each directed edge would have the same direction in all the graphs that belongs to its equivalence class. There is a bijective correspondence (one-to-one) among the set of Markov equivalence classes and the set of essential graphs, their representatives.

The labeled or unlabeled character of the graph means whether its nodes or edges are distinguishable or not. For this, we will say that it is vertex-labeled, vertex-unlabeled, edge-labeled, or edge-unlabeled. The labeling will be considered as a mathematical function, referred to a value

or name (label), assigned to its elements, either nodes, edges, or both, which makes them distinguishable.

5. Complex Networks

A system can be defined as a set of components functioning together as a whole. A systemic point of view allows us to isolate a part of the world, and so, we can focus on those aspects that interact more closely than others.

Network Science is a new scientific field that analyzes the interconnection among diverse networks, as for instance, on Physics, Engineering, Biology, Semantics, and so on. Between its developers, we may remember to Duncan Watts, with the Small-World Network, and Albert-László Barabasi, which developed the Scale-Free Network [2, 4, 8, 22]. About its work, this author has found that the websites that form the network (of the WWW) have certain mathematical properties.

Network Theory is a quickly expanding area of Network and Computer Sciences, and also may be considered a part of Graph Theory [6].

Many phenomena in nature can be modeled as a network, as

- Brain structures. The brain as a network of neurons (as nodes) connected by synapses (their edges).
- Social interactions or the World Wide Web (WWW). All such systems can be represented in terms of nodes and edges.
- On Internet, the nodes can represent routers, and the edges by wires, or physical connections between them.
- In transport networks, the nodes can represent the cities and the edges the roads that connect them. These edges can have weights.

These networks are not random. The topologies of different networks are sometimes very close.

We may distinguish four structural models. So Regular Networks, Random Networks, Small-World Networks and Scale-free Networks.

In the Regular Network, each node is connected to all other nodes. I.e. they are fully connected. Because such type of structure, they have the lowest path length (L), and the lowest diameter (D), being

$$L = D = 1$$

Also they have the highest clustering coefficient. So, it holds $C = 1$.
Furthermore, the highest possible number of edges

$$\text{card}(E) = ((n(n-1))/2) \sim n^2$$

About *Random Graphs* (RGs), we can say that each pair of nodes is connected with probability p . They have a low average path length, according to

$$L \approx ((\ln n) / (\ln \langle k \rangle)) \sim \ln n$$

for $n \gg 1$.

It is because the total network may be covered in $\langle k \rangle$ steps, from which

$$n \sim \langle k \rangle^L$$

Also they possess a low clustering coefficient, when the graph is sparse. Thus,

$$C = p = ((\langle k \rangle) / n) \ll 1$$

given that the probability of being connected each pair of neighboring nodes is precisely equal to p .

The Small-World effect is observed on a network when it has a low average path length, i.e.

$$L \ll n, \text{ for } n \gg 1$$

Recall the now famous "*six degrees of separation*", also called the small-world phenomenon [4, 6]. This hypothesis was firstly formulated by the Hungarian thinker Frigyes Karinthy, in 1929. Then, it was tested by Stanley Milgram, in 1967. The subjacent idea is that two arbitrarily selected

people may be connected by only six degrees of separation (in average, and it is not too larger than this value). Therefore, the diameter of the corresponding graph is not much larger than six. For instance, on social connections.

In the case of the Watts-Strogatz (in acronym, WS) *small-world model* (1998), it represent a hybrid case between a Random Graph and a Regular Lattice. So, Small-World models share with RGs some common features, as may be

- *Poisson or Binomial degree distribution*, near to Uniform distribution.
- Network Size: *it does not grow*.
- Each node has approximately the same number of edges, i.e. *it shows a homogeneous nature*.

WS-models shows the low average path length typical of Random Graphs, i.e.

$$L \sim \ln n, \text{ for } n \gg 1$$

And also such models the usual high clustering coefficient of Regular Lattices, giving

$$C \approx 0.75, \text{ for } k \gg 1$$

In consequence, WS-models have a small-world structure, being well clustered. Whereas the Random Graphs coincides on the small-world structure, but they are poorly clustered. This model (WS) has a peak degree distribution, of Poisson type.

In reference to the last model, called Scale-Free Network, it appears when the degree distribution follows a Power-Law, i.e.

$$P(k) \sim k^{-\gamma}$$

In such case, there exist a small number of highly connected nodes, called *hubs*, being the tail of the distribution. Whereas many of their nodes have sparse connections. They represent the head of such distribution. Such model [2, 5] was introduced by A.-L. Barabási and R. Albert in 1999.

Some of their most notable *features* may be

- *Non-homogeneous nature*, in the sense that some (few) nodes have many edges from them, and the remaining nodes only have very few edges, or links.
- About the network size, *it continuously grows*.
- And on the connectivity distribution, *it obeys a Power-Law distribution*.

A very notable example of Scale-Free Network may be the World Wide Web (WWW, in acronym). As we known [2, 8, 21, 22], it is a collection of many sub-networks. In such case the nodes corresponding to their pages, whereas the edges to hyperlinks.

6. Searching Communities

One of the most interesting research points in Computer Science is the analysis of the structure and properties of Complex Networks. To understanding the function and the dynamics of a network [2, 4], the first step will be known the topology of such network.

Hierarchical Clustering of Networks is one classical method for finding community structures in a Complex Network. The technique establishes a hierarchy of groups according to an associated weight function. So, the data set may be represented by a “dendrogram”, which is a typical tree structure.

Another technique may be the Girman-Newman (expressed GN by acronym) algorithm [18].

Many real networks contain parts in which the nodes are more connected to each other than to the rest of the network. Such parts are called Communities, Clusters, Modules, or Cohesive Groups. Graph Clustering is an essential problem in the analysis of relational data. It is now described as the problem of partitioning networks into communities.

A new and interesting index for Graph Clustering is the so-called *Modularity*, which was proposed by Newman and Girvan, in 2004. It is a quality index for clustering [16, 18]. From then, many algorithms of heuristic type have been proposed attempting to optimize such quality measure [13, 18].

Among the lines of research in the more recent times has been also the aforementioned possible decomposition of every network into independent subsets called "*communities*". A *community* would be a group of nodes that is densely connected internally, but sparsely connected externally.

To sociologists, this concept is also named a "*cohesive subgroup*". So, a typical community [20] consists of several subgraphs, fully connected, with trend to share many of their nodes.

A more precise name would be "*k-clique community*", being defined as a union of all *k*-cliques that can be reached from each other, through a series of adjacent *k*-cliques. Recall that a *k-clique* means a complete subgraph of size *n*.

In a network, it is possible to make groups of elements, by equivalence classes. It will be according the role played by such elements [4, 5]. The assignment of roles, in Social Networks, permits us establish two new concepts, as may be *Structural Equivalence* and the *Regular Equivalence*.

Two nodes are *structurally equivalent*, if they have the exact same neighbors.

While the *regular equivalence* need that the nodes plays their roles exactly.

The study of the community connection of a network helps us to understand its global structure, but also the distribution of agents and activities.

Newman [17] says that "the development of methods for finding communities within networks is a thriving sub-area of the field, with an enormous number of different techniques under development methods for understanding what the communities mean after you find them are by contrast still quite primitive". This will be a promising line of advance.

7. Conclusions

Social Network graphs have patterns that can be useful to characterize them, and may be very interesting to accelerate algorithms, in Optimization. They have also an essential clustering tendency, jointly with a community structure [18].

Observe that the Web is a fractal, considered by a graph analysis [8, 9, 15]. In fact, it is composed of highly cohesive sub-regions, each one of them showing similar characteristics as the complete Web. Such self-similarity may offer the way [7, 9, 19] to construct models about the evolution of the WWW.

Great attention has been focused [10-12, 20] on the problem of making a network of coupled dynamical systems synchronize on a common evolution, as on empirical network analysis and simulations. Networked Control Systems have been an attractive area. Complex Dynamic Networks [4], and Cellular Neural Networks are also studied, with many applications, as solving PDEs, or on pattern recognition. The analysis of synchronization problems is other interesting current field of research [7, 13].

So, we conclude this overview on some of the more interesting and subtle tools of the current mathematics, very interrelated with many other fields, as may be Artificial Intelligence, Sociology, Biocomputing, and so on.

Because *Symmetry* connects between different mathematical branches, as may be Geometry (Euclidean and non-Euclidean), Algebra, Probability Calculus and the very deep underlying concepts, distribution functions and density functions, with many different associated models, Mathematical Analysis (for instance, on fuzzy measures [6, 14]), and so on.

For this purpose, we must recall the claim of Dreyfus and Eisenberg about that the teachers need to develop awareness of the importance of symmetry (not only at high school, but also for undergraduate and graduate university level) for mathematical problem solving. They support this claim by their finding that many teachers (too much) did not recognize symmetry as an essential tool that can simplify the way to reach the solutions. The support of graph theoretical examples may be surely the better aid to overcome such difficulty.

8. References

- [1] Albert, R., Barabási, A.-L. (2002). Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74: 47-66.
- [2] Barabási, A.-L., Bonabeau (2003): Scale-Free Networks. *Scient. Am.* May 2003: 50-59.
- [3] Barabási, A.-L. (2004). *Linked: How Everything is Connected to Everything Else*. Plume Publ.
- [4] Barrat, A. et al. (2008). *Dynamical processes in Complex Networks*. Cambridge University Press.
- [5] Bocaletti, S. et al. (2006). Complex Networks: Structure and Dynamics. *Phys. Rep.*, 424: 175-308.
- [6] Bornholdt, S., Schuster, H. G. (editors) (2003). *Handbook of Graphs and Networks: From the Genome to the Internet*.
- [7] Broder, A. et al. (2000). Graph structure in the Web. *Proc. 9th. WWW*, pp. 309-320.
- [8] Calderelli, G. (2007). *Scale-Free Networks*. Oxford University Press.
- [9] Dill, S. et al. (2002). Self-similarity in the Web. *ACM Trans. Intr. Techn.* 2(3): 205-223.
- [10] Dorogotsev, S. N., Mendes, J. F. F. (2002). Evolution of Networks. *Adv. Phys.*, 51: 1079-1082.
- [11] Dorogotsev, S. N., Mendes, J. F. F. (2003). *Evolution of Networks: From biological networks to the Internet and WWW*. Oxford University Press.
- [12] Dorogotsev, S. N. , Goltsev, A. V., Mendes, J. F. F. (2008). Critical phenomena in Complex networks. *Rev. Mod. Phys.*, 80: 1275-1284.
- [13] Dreyfus, T., Eisenberg, T. (1986). On visual versus analytical thinking in Mathematics. *Proc. of the Tenth International Conference on the Psychology of Mathematics*, pp.153-158.
- [14] Garrido, A. (2009). Symmetry of Complex Networks. *EIJ-AMO (Electronic International Journal-Advanced Modeling and Optimization)*. 11(4): 615-624.
- [15] Girvan, M., Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Nat. Acad. Sci. USA* 99: 7821-7826.
- [16] Harary, F. (1967). *Graph Theory*. Academic Press.
- [17] Newman, M. (2003). The structure and function of Complex Networks. *SIAM Review*, 45: 167-256.
- [18] Newman, M. (2004). Fast algorithm for detecting community structures in networks. *Phys Rev. E* 69: 066133.
- [19] Pastor-Satorras, R., Vespignani, A. (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press.
- [20] Pólya, G. (1981). *Mathematical Discovery*. John Wiley and Sons, New York, p. 161.
- [21] Strogatz, S. H. (2001). Exploring Complex Networks. *Nature*, 410: 268-276.
- [22] Watts, D. J., Strogatz, S. H. (1998). Collective Dynamics of "small-world" networks. *Nature*, 393: 440-442.
- [23] Watts, D. J. (2003). *Six Degrees: The Science of a Connected Age*. Norton and Co.
- [24] Watts, D. J. (2003). *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press.
- [25] Weyl, H. (1952). *Symmetry*. Princeton University Press.