Network Science

Amaro, Alan.

1. Introduction

What is network science? Network science is an interdisciplinary field that studies networks, which are structures composed of nodes (or vertices) connected by edges (or links). These networks can represent various systems across different domains, including social interactions, biological systems, communication networks, and more. The primary goal of network science is to understand the structure and dynamics of networks, analyze their properties, and apply this knowledge to solve real-world problems.

In order to represent certain networks we'll be using a **Colaboratory Notebook**

1.1. Key Concepts in Network Science

1.1.1. Nodes and Edges

- **Nodes** (or vertices): Represent entities or points in a network.
- **Edges** (or links): Represent the connections or relationships between nodes.

1.1.2. Network Types

- **Social Networks**: Represent relationships and interactions between people, such as friendships, collaborations, and communication.
- **Biological Networks**: Include networks of biological entities like genes, proteins, and metabolic pathways.
- **Communication Networks**: Describe networks of communication between devices or entities, such as the internet or telecommunication systems.
- **Transportation Networks**: Represent the connections in transportation systems, like roads, railways, and flight routes.

1.1.3. Graph Theory

Network science heavily relies on graph theory, which provides mathematical tools to analyze network properties. Key concepts include:

- **Degree**: The number of edges connected to a node.
- **Path**: A sequence of edges connecting nodes.
- **Connectivity**: How well nodes are connected within the network.
- **Centrality**: Measures of the importance or influence of a node within the network, including degree centrality, betweenness centrality, and closeness centrality.

1.1.4. Network Properties

- **Topology**: The arrangement of nodes and edges in a network, including random networks, small-world networks, and scale-free networks.
- **Clustering**: The tendency of nodes to form tightly-knit groups.
- **Community Structure**: The organization of nodes into groups where nodes within the same group are more densely connected to each other than to nodes outside the group.
- **Network Dynamics**: The study of how networks evolve over time, including the addition or removal of nodes and edges.

1.1.5. Applications

- **Social Network Analysis**: Understanding social dynamics, influence, and information spread.
- **Epidemiology**: Modeling the spread of diseases through populations.
- **Infrastructure**: Optimizing and managing communication and transportation systems.
- **Data Science**: Analyzing relationships and patterns in large datasets.

1.1.6. Tools and Techniques

- **Network Visualization**: Techniques to visually represent and analyze network structures.
- **Algorithms**: Algorithms for detecting communities, finding shortest paths, and analyzing network robustness.
- **Simulation**: Simulating network dynamics to study behaviors and properties under different conditions.

1.2. Example of a Simple Network

Consider a social network where nodes represent people, and edges represent friendships. This network can be analyzed to identify central individuals, detect groups of friends (communities), and study how information or diseases spread through the network.

Network science provides a framework for understanding complex systems by focusing on the interactions and relationships between components. Its applications span many disciplines, making it a powerful tool for analyzing and solving various real-world problems.

1.3. About the Brideges of Konigsberg

Graph theory, a foundational aspect of network science, originated from a problem in 18th-century Königsberg, now Kaliningrad, Russia. In 1735, mathematician Leonard Euler tackled the challenge of finding a path across seven bridges without crossing any bridge more than once. Euler represented the city's layout as a graph, identifying that the path was impossible due to the graph's structure having four nodes with an odd number of connections. This insight established that some problems can be simplified through graph representation, revealing inherent properties of networks that dictate their behavior. Euler's work laid the groundwork for graph theory, which explores network structures and their characteristics.[1]

2. Mathematics of networks

2.1. Types of nets and adjacency matrix

A network or graph correspond to a set of n nodes $\{n_1, \ldots\}$ along with L edges between node pairs $\{n_i, n_j\}$. N denotes the size of the network.

We define a sparse matrix as matrix which satisfies that $L \ll L_{max}$

Figura 1: Network representation

We define the adjacency matrix $(n \times m)$ as a matrix that represents the number of connections in a network. Each entry A_{ij} is defined as 1 if there's a link between the node i and j, and 0 if there's not connection. The convention corresponds to take the columns as the origin and the rows and the target.

There's not connection of a node with itself

$$A_{ii} = 0 \tag{1}$$

Which type of network does we have? A **simple network** satisfies:

$$A_{ij} = A_{ji} \tag{2}$$

In this network, the edges don't have a defined direction. The adjacency matrix is binary and represents the existence of connection between the nodes.

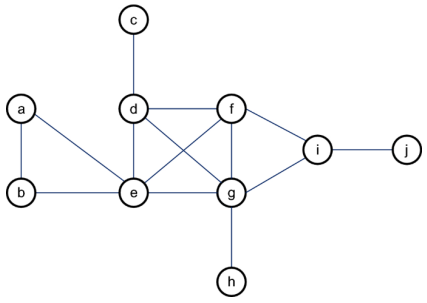

Figura 2: Simple network

We may also find network with a defined direction in the edges, this kind of network is denoted as **directed network** (DiGraph). The edges may be uni or bi-directional

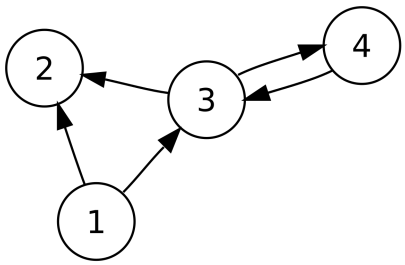

Figura 3: Directed network

In all this networks the nodes are connected with only one edge, we can define multiple networks with more than one edge between nodes, as example, we can connect node a with node b with 2 edges. In this case the adjacency matrix it's not binary as $A_{ij} \in \mathbb{N}$. In a pure multiple network there's not direction.

We can define weighted networks, in this case the edges have a real value, so $A_{ij} \in \mathbb{R}$.

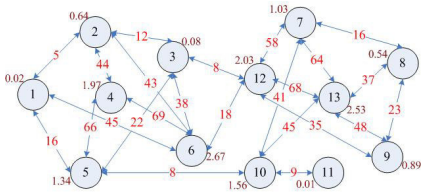

Figura 4: Weighted Network

What's a bipartita network? A **bipartite network** (or **bipartite graph**) is a type of graph in which the set of vertices can be partitioned into two disjoint sets such that no two vertices within the same set are adjacent. Formally, a graph $G = (V, E)$ is bipartite if there exists a partition of V into two disjoint sets U and V such that every edge in E connects a vertex in U to a vertex in V.

In other words:

$$V = U \cup V$$

$$E \subseteq \{(u, v) \mid u \in U, v \in V\}$$

where U and V are the two disjoint sets of vertices, and E is the set of edges connecting vertices from U to vertices in V.

Bipartite graphs are useful in various applications, such as:

- Modeling relationships between two different classes of objects. For example, in a social network, one set might represent people and the other set might represent events, with edges connecting people to events they attend.
- Recommender systems, where one set could be users and the other set could be products, with edges representing user preferences or interactions with products.

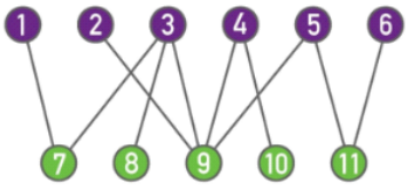

Figura 5: Bipartita Graph

2.2. Degree

What's the degree of a node? The degree of a node corresponds to a intrinsic characteristic of each node, denoted as k_i, it's the number of connections (in or out) this node have. In the directed network we defined k^{in}, k^{out} as the input and output degree.

We obtain this vector from the adjacency matrix through the sum of rows or columns

$$k_i = \sum_{j=i}^{n} A_{ij} \tag{3}$$

In the directed networks, the sum of rows corresponds to the input vector and the sum of columns to the output vector.

$$k^{in} = \sum_{j=1}^{j} A_{ij} \tag{4}$$

$$k^{out} = \sum_{i=1}^{j} A_{ij} \tag{5}$$

2.3. Adjacency matrix properties

The total amount of edges in a graph corresponds to:

$$L = \frac{1}{2}\sum_i k_i = \frac{1}{2}\sum_{i,j} A_{ij} \tag{6}$$

In the directed graph, the bi-edges must be consider as two links:

$$L = \sum_{i=i} k_i^{out} + \sum_{i=i} k_i^{in} \tag{7}$$

The degree vector may be interpret as a propagation measure of information. We consider important to consider the next properties:

- $A \cdot \hat{1}$: This operation represents the total amount of connections.
- $\frac{\hat{1}^T \cdot A \cdot \hat{1}}{2}$: This operation represents the total amount of edges in the graph
- $\frac{Tr(A^3)}{6}$: This operation represents the total amount of cycle paths of length 3
- $A^2 \cdot \hat{1}$: This operation represents the total amount of paths of length 2 in the whole graph

How can I represent the adjacency matrix of a bipartite network? Given a bipartite network with two disjoint sets of vertices U and V, the adjacency matrix A is typically arranged in the following block matrix form:

$$A = \begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$$

Here's a breakdown of what this means:

- U **and** V are the two disjoint sets of vertices. Let $|U| = m$ and $|V| = n$.
- **The matrix** A is an $(m+n)\times(m+n)$ matrix.
- B is an $m \times n$ matrix representing the edges between vertices in U and vertices in V.
- B^T is the transpose of B, an $n \times m$ matrix.

Block Matrix Format

1. **Top-Left Block (Zero Matrix)**: An $m \times m$ matrix of zeros. It represents no edges between vertices within the set U.
2. **Top-Right Block (B)**: An $m \times n$ matrix where each entry B_{ij} indicates an edge between the i-th vertex in U and the j-th vertex in V. Typically, $B_{ij} = 1$ if there is an edge and 0 otherwise.
3. **Bottom-Left Block (B^T)**: The transpose of B, an $n \times m$ matrix, where each entry $(B^T)_{ji}$ mirrors B_{ij}, indicating the presence of an edge from the j-th vertex in V to the i-th vertex in U.
4. **Bottom-Right Block (Zero Matrix)**: An $n \times n$ matrix of zeros. It represents no edges between vertices within the set V.

Example
Consider a bipartite graph with two disjoint sets:

- Set $U = \{U_1, U_2\}$
- Set $V = \{V_1, V_2, V_3\}$

With the following edges:

- U_1 is connected to V_1 and V_2
- U_2 is connected to V_2 and V_3

The adjacency matrix A of this bipartite graph is a 5×5 matrix, partitioned into blocks as follows:

$$A = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Where:

- The top-left 2×2 block is zero, representing no edges within set U.
- The top-right 2×3 block shows the connections from U to V.
- The bottom-left 3×2 block is the transpose of the top-right block.
- The bottom-right 3×3 block is zero, representing no edges within set V.

2.4. Edges properties

Let's consider a simple network, what is the total amount of edges if we consider n nodes full connected?

$$L_{max} = \frac{1}{2}\sum_{i=1}^{N} k_{max} = \frac{1}{2}\sum_{i=1}^{N}(N-1) = \frac{N(N-1)}{2} \tag{8}$$

Full connected means all nodes connected with all its neighbors. And what is the minimum value? For this case we have two options; If the network is a

tree graph then the nodes in the network have only one edge:

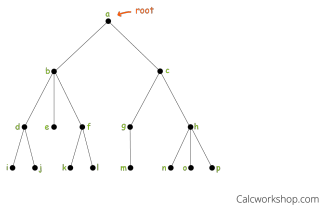

Figura 6: Tree network

And then the minimum corresponds to: $L_{min} = N - 1$. And the second type corresponds to disconnect network where $L_{min} = 0$, in this case there's not connection.

Let's define the average degree o density $\bar{K}$ as:

$$\bar{k} = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N} \tag{9}$$

The average degree varies depending of the network kind.

2.5. Path and distance between nodes

Path is the set of edges that share one node with the *next element*, as example: $\{(1,4),(4,3),(3,1),(1,2)\}$. And we define length as the number of edges that join this nodes. From these definitions we have the next type of paths:

- Geodesic: Shortest path between two nodes.
- Eulerian: Cross each node just one time.
- Hamiltonian: Visit each node just one time.
- Cycle: Start and finish in the same node.

And **distance** is the minimum length between nodes.

Between node i and j, we have a path of length 1 if $A_{ij} = 1$. In a similar way, we have a path of length 2 if there's a node k such as:

$$A_{ik} = 1 \wedge A_{kj} = 1 \rightarrow A_{ik}A_{kj} = 1 \tag{10}$$

From these we can express the number of paths of length d as:

$$N_{ij}^{d} = [A^{d}]_{ij} \tag{11}$$

This number represents the amount of paths of length d between the node i and j

3. Measures of node centrality

Measures of centrality is a way to determine how important a node is in a network. The types of measures correspond to: *Degree centrality, Closeness, Betweenness, Eigenvector, Katz, Page-Rank*

The degree measure, closeness, and betweenness are particularly useful in a start network. In a star network, the most connected node is the central node, where the distance between the central node and any other node is 1, and the distance between any two nodes is 2. From these we can denote that the central node is the intermediate node.

3.1. Degree centrality

Degree's centrality $C_D(i)$ corresponds to the degree of a given node. In a simple network, it corresponds to the node's connectivity, so the more central it is, the more connect it will be.

In a directed network, we define the input and output degree. We label the input degree as the node's reception and the output degree as the node's transmission or influence. From these, we can define the normalized degree as follows:

$$C_D(i) = \frac{k_i}{N-1} \tag{12}$$

Where N is the amount of nodes in the network. The degree's normalization allows to a better understanding of the node's importance.

3.2. Closeness centrality

To analyze this topic is fundamental to review the following concepts:

- Path: It is an ordered set of links with a common node $(A,B)(B,C)\ldots$
- Path's length: It is the amount of elements in the path's set.
- Geodesic Path: Given all the possible paths between two nodes, the geodesic path is the path with the lowest distance.
- Distance between nodes: Given two nodes, the distance between them is the length of the geodesic path between them.

From these concepts, we introduce the idea of the closeness' centrality. This metric measures the relevance of each node based on its closeness to other nodes. First, let's consider the distance between the node i and j, we define the average distance of the node i as follows:

$$l_i = \frac{1}{N-1}\sum_j d_{ij} \tag{13}$$

Since the distances d_{ij} may be significant bigger among them, we consider the inverse of l_i:

$$C_C = \frac{1}{l_i} = \frac{N-1}{\sum_j d_{ij}} \tag{14}$$

If two nodes are disconnected, we define the distance between them as ∞. This implies certain problems with the given definition since the centrality would be zero. In order to solve this problem, we consider the following definition.

$$C_C^* = \frac{1}{N-1}\sum_j \frac{1}{d_{ij}} \tag{15}$$

This definition solves the problematic, however, the most used expression corresponds to:

$$C_C = \frac{1}{l_i} = \frac{N-1}{\sum_j d_{ij}} \tag{16}$$

3.3. Betweenness Centrality

Consider the amount of geodesic paths between the node s and t that cross the node i as g_{st}^i. We define the betweenness centrality as follows:

$$C_B(i) = \sum_{s,t} g_{st}^i \tag{17}$$

This metric represents how importance a node is based on the information flux. The most important node in a network corresponds to the node in which the information will necessary cross it.

Consider the next definitions: $g_{i,i}^i = 1$, $g_{s,s}^i = \delta_{i,s}$. And the path g_{is}^i consider the external nodes. In not directed networks we consider $g_{st}^i = g_{ts}^i$. From these concepts we consider the betweenness centrality as follows:

$$C_B(i) = \sum_{s,t} \frac{g_{st}^i}{g_{st}} \tag{18}$$

From the given definitions we can transform this expression into:

$$\begin{aligned} C_B(i) &= \sum_{s,t} \frac{g_{st}^i}{g_{st}} \\ &= \sum_s g_{ss}^i + \sum_{s\neq t} \frac{g_{st}^i}{g_{st}} \\ &= 1 + 2\sum_{s<t} \frac{g_{st}^i}{g_{st}} \end{aligned} \tag{19}$$

Where the last step is only true in simple networks. In figure **??** we presented an example. If we required a normalized value we used the following expression:

$$C_B'(i) = \frac{C_B(i)}{N^2} \tag{20}$$

3.4. Eigenvector centrality

The eigen-vector centrality consider the node's relevance based on the neighbours' relevance. This corresponds to the importance or quality of its connections. To calculate this value, we consider the adjacency matrix A_{ij} and the greater eigenvalue λ, then the eigenvector centrality is given by:

$$x_i = \frac{1}{\lambda}\sum_j A_{ij}x_j \tag{21}$$

Where x_j is the centrality score of node j. Some benefits of this centrality corresponds to a global measure of the network.

3.5. Katz's Centrality

' Katz's centrality is a improved to betweenness centrality. This centrality take into account the diverse paths between two nodes, it assign to the closest paths a bigger importance than the larger paths. This importance is defined with the $\alpha < 1$ parameter and the following expression:

$$S_n = I + \alpha A + (\alpha A)^2 + (\alpha A)^3 + \ldots + (\alpha A)^n \tag{22}$$

In this expression we are indicating that the most larger paths are more likely probable to transfer information than the shortest paths. If we consider the following expression:

$$x < 1 \rightarrow \lim_{n\to\infty} S_n = \frac{1}{1-x} \tag{23}$$

It is easy to obtain for our case:

$$\alpha < \rho \rightarrow \lim_{n\to\infty} S_n = (I - \alpha A)^{-1} \tag{24}$$

Where ρ denotes the spectral ratio of the adjacency matrix. This limit represents the weighted sum of all paths between i and j according to their length. Hence, Katz's centrality represents the influence of a node into the network and the influence of the network into the node.

$$[S_\infty]_{ij} = (I - \alpha A)^{-1}_{ij} \tag{25}$$

To calculate the influence of a node i into the network we consider the sum over all the nodes in the net:

$$\sum_{j=1}^{N}[S_\infty]_{ij} = \sum_{j=1}^{N}(I - \alpha A)^{-1}_{ij} \tag{26}$$

Or with the usual notation:

$$C_k(i) = (I - \alpha A)^{-1} \cdot 1 \tag{27}$$

3.6. Page-Rank Centrality

Page-Rank centrality is a centrality measure originally developed by Larry Page and Sergey Brin to rank websites in Google's Search Engine. It builds upon the idea of eigenvector centrality, but applies it to directed networks.

It measures the importance of a node in a directed network by considering not just the number of incoming links, but also the quality or importance of the nodes providing those links. This model includes a probability-based model that reflects the behavior of a 'random surfer' on the web. This is:

$$PR(i) = \frac{1-d}{N} + d \sum_{j \in M(i)} \frac{PR(j)}{L(j)} \tag{28}$$

Where d is the damping factor (0,85), $M(i)$ ins the set of nodes that link to node i, $L(j)$ is the number of outbound links from node j.

4. Node Group Measurements

4.1. Clustering coefficient

Accumulation coefficient or clustering is a measurement of the number of links between adjacent nodes with respect to a given node. Consider a node i with degree k_i, the accumulation coefficient is then a proportion between the number of neighbors' links of the node i and the possible maximum number of neighbors' links of i.

If we define the number of neighbors' links of i as L_i:

$$L_i = \frac{1}{2}[A^3]_{ii} \tag{29}$$

This expression comes from the fact that when to three nodes are connected, they form a path of length 3, this path can be cross by 2 ways. The maximum number of links we can form corresponds to the maximum number of different pairs, which is: $\frac{k_i(k_i-1)}{2}$. By these definitions, we define the *clustering* of i as follows:

$$C_i = \frac{2L_i}{k_i(k_i-1)} \tag{30}$$

This coefficient may be interpreted as: From the connection of the neighbors from a given node, how connected the neighbors are among them.

4.2. Node Groups

A wide variety of networks divide into groups or communities. In this section, we present a series of concepts related to how we cluster a network.

- Cliques: A *clique* (Fig 8) is a maximal subset of vertices of nodes in a undirected network, such that every member of the set is connected by an edge to every other. Where maximal refers that there is no other node that can be added while preserving the properties of the set.

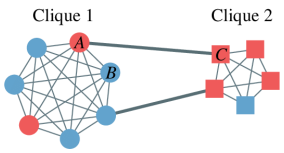

Figura 8: Clique representation

The present of clique in a network is an indicator of a highly cohesive subgroup.

- Plexes: From the definition of *clique* we may notice that in network we encounter almost perfect cliques, we define a k-plex (Fig 9), corresponding a maximal subset of n nodes where each node is connected to at least $n - k$ of the others. One important concept about *cliques* and *plexes* is that they can overlap each

other, so a node can belong to more than one clique or plex.

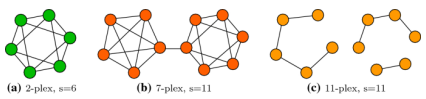

Figura 9: Plex representation

Usually, smaller values for k can be meaningful for smaller groups, while large groups tend to reflect meaningful results.

- Cores: In contrast with the definition of a plex, a k-core of size n is a maximal subset where each node has at least k connections.
- k-clique: It corresponds to a maximal subset of nodes where each node is no more than a distance k from any other node. If we restrict this definition to a set where each node in a k-clique belongs to the same path form by the nodes, we get a k-clan or a k-club.
- Components and k-Components: A component is a maximal subset of nodes where each one is reachable by some path from each other. Then, a k-component is a maximal subset of nodes where each node is reachable from the other by at least k independent paths. For cases $k = 2$ and $k = 3$, the components are said to be *bicomponents* or *tricomponets*. The number k is a presentation of the robustness of the network.

  We define independent paths to a pair of paths that only share the initial and final nodes.

4.3. Modularity

Modularity is a measurement of how many communities are present in a network. In contrast to other metrics, it is not possible to extract it from the adjacency matrix. It is more related to the topology of the network itself.

Modularity measure the number of links between nodes in the same class, consider C_i as the node's i class, n_c the total number of classes in a network. We define modularity Q as follows:

$$Q = \frac{1}{2L}\left[\sum_{i,j}(A_{ij} - \frac{k_i k_j}{2L})\delta(c_i, c_j)\right]$$

The first term represents the total number of links among nodes in the same class. The second term represents the probability that node i is connected to node j. Note that we consider the probability proportional to the degree of each node. The delta function indicates belonging to the same class.

Similarly, the second term can be interpreted as the number of links between same-class nodes if the nodes are randomly connected, as an average value. This election of metric Q is arbitrary, but it gives the best results.

Therefore, Q is a comparison between the number of links in the same class and the average value if the links are random distributed. A positive value of Q indicates a strong grouping into the same class. A negative value of Q indicates a strong grouping into different classes. The factor $1/2L$ is to standardize the value between $[-1, 1]$

4.4. Maximizing Modularity

When considering a phenomena and analyzing its components it is possible to define a network. Inside a network we may find communities, the associate metric to measure them corresponds to Modularity.

How can we get the modularity of a network? Blondel et al[2] published a iterative algorithm that allow us to get the modularity in two steps.

1. Consider a network with no classes. For each node define a different class, calculate its modularity for each neighbor. We change the modularity if it is greater than the last one and we assign them then the same class for the pairs of nodes. We iterate over the network until reach a stable Q value.

2. In this step, the network will have stable communities, we convert these communities into nodes connected into themselves, the among of links between each community convert into one link with the value of sub-links. We repeat the step one until reach a stable Q value.

Each of this steps will give us information about the network, such as hierarchy, interconnections, and so on.

5. Global properties of the Network

In this section we mention the fundamental properties of a network. These corresponds to the most useful metrics involving the whole set, such as number of nodes, links, etc.

- Network's average degree $\langle k \rangle$
- Clustering or Average accumulation coefficient $\langle C \rangle = \frac{1}{N} \sum C_i$: This coefficient involves the node's distribution (homogeneous or in-homogeneous)
- Average distance $\langle d \rangle = \frac{1}{N} \sum l_i = \frac{1}{N(N-1)} \sum_{i,j} d_{ij}$.
- Eccentricity $E(i) = \text{máx}\{d_{ij} : j \in \text{nodes}\}$
- Network's diameter: This is the eccentricity's maximum for all nodes.
- Network's radius: This is the eccentricity's minimum for all nodes.

5.1. Degree Distribution

Consider a histogram of the number of links N_k with degree k. We define the probability of pick a random node with degree k as follows:

$$P_k = \frac{N_k}{N} \tag{31}$$

This probability contains information related to the *distribution's momentums*: average value, variance, These momentums are defined as: $\langle k^n \rangle = \sum^{\infty} k^n P_k$

6. Random Networks

6.1. What is a Random Network?

A random network the assignment of links between nodes is done randomly. From these we have two possible definitions:

1. $G(N, L)$: This model fixed the number of nodes N and the number of links L between them. The correlation between nodes is then set randomly, we select L random pairs of nodes among the $\frac{N(N-1)}{2}$ possibles.
2. $G(N, p)$: This model fixed the number of nodes N and assigns links between nodes with a probability of p. Here $p = 1$ corresponds to a fully connected network.

6.2. Binomial distribution

The binomial distribution expression corresponds to:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Where: - $P(X = k)$ is the probability of getting exactly k successes in n trials. - $\binom{n}{k}$ is the binomial coefficient, calculated as $\frac{n!}{k!(n-k)!}$. This coefficient represents the number of subsets with k elements in a set with n elements - p is the probability of success on a single trial. - $1 - p$ is the probability of failure. - n is the number of trials. - k is the number of successes.

6.3. Links number

As random networks are intrinsically related to probability processes, we need to implement a statistical analysis in the set of networks (ensemble). Consider a random network $G(N, p)$, what is the probability that G has L links?

1. Consider that the number of links must be a positive number and it must lower that the maximum:

$$0 \leq L \leq L_{max} = \frac{N(N-1)}{2} \tag{32}$$

2. The probability that the network has L links corresponds to p^L
3. From this, the probability that the network does not have L links corresponds to $(1-p)^{L_{max}-L}$
4. Finally, we must consider the binomial distribution since we have a set with $\frac{N(N-1)}{2}$ maximum links and L links.
5. From these considerations, we obtain:

$$P(X = L) = \binom{L_{max}}{L} p^L (1-p)^{L_{max}-L}$$

If we derivative the Newton binomial expression and consider $(p+q) = 1$ (Because we're dealing with probabilities), we obtain:

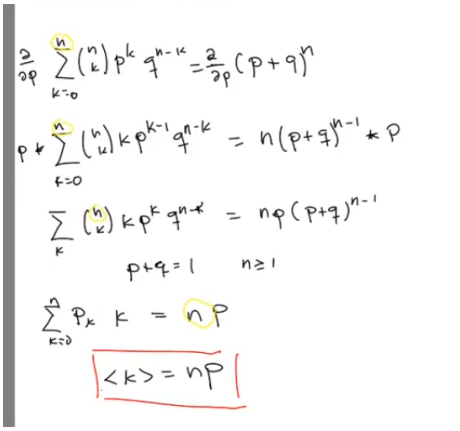

Figura 10: Deduction of the average value

And using this results we get:

$$\langle L\rangle = p\left[\frac{N(N-1)}{2}\right] \tag{33}$$

6.4. Average degree and clustering

From the definition of degree as: $2L/N$; Now we consider the average value:

$$\langle k\rangle = \frac{2}{N}p\left[\frac{N(N-1)}{2}\right] = p(N-1) \tag{34}$$

This is:

$$\langle k\rangle = p(N-1) \tag{35}$$

And for the clustering we consider the definition as $C_i = \frac{2L_i}{k_i(k_i-1)}$, k_i is the number of neighbors links. If we consider the maximum of possible connections among our neighbors, we will get:

$$\langle L_i\rangle = p\left[\frac{k_i(k_i-1)}{2}\right] \tag{36}$$

If we add this result to the C_i definition, we obtain the following expression:

$$\langle C_i\rangle = p \tag{37}$$

So, the clustering coefficient is independent of the nodes configuration.

6.5. Degree distribution

What the probability that a node would have a degree of k? If we consider the probability that the network has k degree as p^k, and the probability that the network does not have k degree as $(1-p)^{N-1-k}$, we obtain the following:

$$P_k = \binom{N-1}{k}p_k(1-p)^{N-k-1} \tag{38}$$

6.6. Poisson's Distribution

In random networks that satisfies the following condition:

$$\frac{\langle k\rangle}{N} \ll 1 \tag{39}$$

The binomial distribution resembles to Poisson's Distribution:

$$P_k = e^{-\langle k\rangle}\frac{\langle k\rangle^k}{k!} \tag{40}$$

The previous condition is equivalent to: $\langle k\rangle \ll N$, this condition represents the *sparce of the network*. As mentioned above, real networks are often sparse.

Referencias

[1] A.L. Barabási. *Network Science*. Cambridge University Press, Glasgow, United Kingdom, 2016. Interactive version available online. URL: `http://networksciencebook.com/`.

[2] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, oct 2008.