**Centrality measures (CM)** has been identified as an appropriate method for identifying influencers in social media. Although plenty of centrality measures have been developed, **four basic concepts have been recognized – degree, closeness, betweenness, and eigenvector**.

**Degree**

Degree centrality, which counts how many direct connections each individual has with other actors within the same network. It is a local measure of the importance of a node. It is a local measure of the importance of a node.

we could define **followers as the in-degree relationship** or in simple words, as a number of connections (users) pointing to a node (a member). Whereby **friends refer to the relationships; namely, out-degree** (from a node to other links).

**Closeness**

Closeness centrality shows how each node in the network is independent of the others. In this situation, the duration time needed for the source to spread information will be reflected. It corresponds to the time taken for the source of the vertex to spread the information. The more central a node is, the lower its total distance to all other nodes.

Closeness can be regarded as a measure of how long it will take to spread information from v to all other nodes sequentially

**Betweenness,**

Betweenness centrality measures the number of paths that pass through a node to reach any other node in the network with respect to the total number of paths in the network that allow these same nodes to connect, showing the extent to which, an actor is on the shortest path between two other nodes. It was introduced as a measure for quantifying the control of a human on the communication between other humans in a social network by Linton Freeman

It can be thought of as a “bridge” within the network, in which the actor takes a strategic position in the flow of communication between different groups. An individual with a high betweenness centrality can present the shortest path for the diffusion of a message through an extensive network. It is also a global measure of the importance of that actor in

the network.

**Eigenvector**

Eigenvector centrality, also ponders the quality of these connections: an actor with the same degree centrality that another can be more influential within a network if its connections are with actors who are in turn well connected. Information about the whole structure of the network is required for its estimation as eigenvalues and eigenvectors associated to the network adjacency matrix are estimated, obtaining information on the direct and indirect importance of each node in particular. It is a global measure of the importance of that actor in the network.

**Betweenness and eigenvector** centralities have very desirable properties for the location of an influencing potential. A combination of both features would simultaneously include those actors who connect dispersed groups through highly connected actors. A priori, the potential for diffusion is very large.

Members of the network that simultaneously meet the highest values of both betweenness and eigenvector centrality should be classified as potential influencers.

The more important the node in a network the higher the eigenvector score obtained by the node in the network, and, hence, should be the one considered as an influential node. The basic factor in determining the most influential nodes in a network with centrality measurement using eigenvector centrality is the eigenvector value among its adjacency

Example Comparison among centrality measures

Rank Degree Closeness Betweenness Eigenvector PageRank

1 MEDICI MEDICI MEDICI MEDICI MEDICI

2 GUADAGNI RIDOLFI GUADAGNI STROZZI GUADAGNI

3 STROZZI ALBIZZI ALBIZZI RIDOLFI STROZZI

Deciding which are most appropriate for a given application clearly requires consideration of the context

Create a data frame from all the measurements

User the Pearson correlation to understand correlation among centrality measures

Generally, the 3 centrality types will be positively correlated. When they are not (low) correlated, it probably tells you something interesting about the network

Create a matrix that represents a two-dimensional scatter plot in which the individual

components of a network are plotted according to their betweenness centrality (x-axis) and

their eigenvector centrality (y-axis). Then, according to a relevant criterion to determine

thresholds in each dimension, this plane is divided into quadrants, allowing classifying

every actor into four different groups:

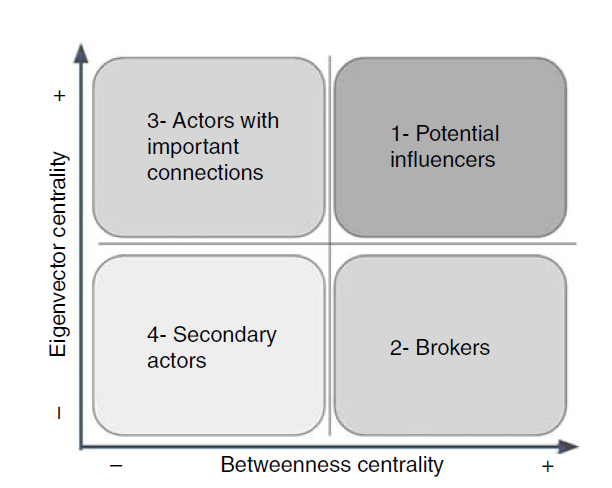
(1) potential influencers with a high degree of betweenness and eigenvector centrality;

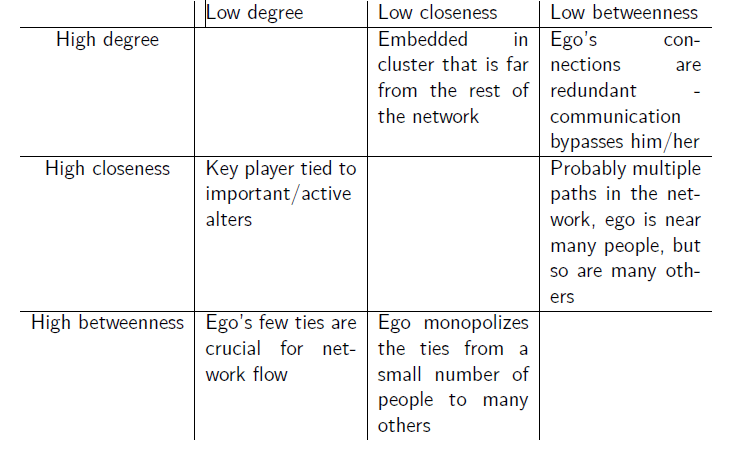
(2) brokers or individuals with high betweenness centrality and low eigenvector centrality;

(3) actors with important connections, their low score in betweenness centrality

suggesting a limited outreach to groups outside their local community; and

(4) secondary actors.





An important node (a node is important if it is connected to other important nodes) makes more people depend on them to make a connection with others and vice versa. The more important the node, the more central and closer that node is to the others.

[**Centrality measures in SNA**](https://rampages.us/jameson/2017/10/02/centrality-measures-in-sna/)

The four types of centrality used in social network analysis, degree, betweenness, closeness, and Eigenvector centrality, have a variety of different applications. Briefly, degree centrality is the total number of direct connects one node has with another, closeness considers the average distance from one node to the other nodes in the network, betweenness considers a node’s position within the totality of all shortest paths between any two given nodes within a network, and Eigenvector centrality rates the importance of the other nodes connected to the specific node in question. The application of these modes of centrality depends on the research question. For example, if one is interested in specifying individuals within a network that act as gatekeepers, betweenness might be the most appropriate metric, as it considers a node’s position in important paths within the network. Alternatively, if one is interested in the spread of a pathogen, Eigenvector centrality might help identify node that could compromise the entire network. If one is interested in determining who’s on the “inside” and who is relatively isolated within a network, closeness centrality would be a good choice. Finally, there’s degree centrality, which is somewhat descriptively weak as it doesn’t consider the broader network, but if a node exhibits a high level of degree centrality, it’s probably fairly important within its network.

Measures of centrality don’t feature prominently within every analysis section that utilize social network analysis, but researchers frequently use these metrics to describe a network. For example, in Grandjean’s (2016)⁠ article, “A social network analysis of Twitter: Mapping the digital humanities community,” Grandjean constructs a directional network using Twitter following/er relationships, and ranks them according to degree centrality, betweenness centrality, and Eigenvector centrality – note, Grandjean uses in-degree and out-degree centrality measures, which indicate the direction of the relationships which the degree centrality ranks, because Grandjean’s data is directional in nature. Simiarly, Murthy & Lewis (2014)⁠ also construct social networks based off of social media data. They utilize a survey instrument which also informs much of their analysis, but one of the metrics which they utilize to find “hubs” within their network is degree centrality – they also analyze this population of individuals with high degree centrality and find that they are more likely to be women. Personally, I find this methodology surprising – given the authors’ interest in finding hubs, ranked centrality metrics might bolster the conclusions made from their survey analysis, and I think that it is a very practical form of analysis for social media projects.

## Section 1 – Quick Introduction to Network Analysis

Top of Form

## Lesson 1 – Let’s Sync

In this course you will learn how to create beautiful and useful visualizations for your network analysis needs.

We created this introductory session for those who are not very familiar with network analysis. In this lesson we will provide a quick overview of different types of network analysis followed by introduction of the key concepts that are used in this field. If things like SNA, “betweenness centrality” and “modularity” sound familiar and you just want to get to the practical side of things, simply skip this lesson and the next lesson and go directly to [**Installing Gephi and InfraNodus**](https://noduslabs.com/courses/network-visualization-and-analysis-with-gephi/units/section-0-15-minute-overview/page/2/?try). It is then followed by Section 2 where we will guide you through the process of creating your own network visualization from scratch.

For those who need a quick introduction to the basic concepts and types of network analysis, read on.

## Lesson 2 – Social Network Visualization & Analysis

Whether you’re a student, a researcher, or marketing specialist chances are you’re studying people, groups and their interactions. Social network analysis (SNA) offers a very powerful toolbox for that and below I will demonstrate how it can be useful.

Consider we have a group that we want to study. Maybe it’s a group of politicians, or your customers, or the people in a classroom. SNA can help us discover some very valuable insights about the group: what are the subgroups it consists of, who are the most influential players, what are the most efficient communication strategies, etc.

**[](https://noduslabs.com/wp-content/uploads/2012/05/Screen-Shot-2012-05-29-at-15.12.16.png)**

The way it works is very simple. Each individual is considered to be a node and their interactions are the relations between them. As a result we get a network graph of individuals, a snapshot of social dynamics. We can then use various methods and tools from graph theory to perform qualitative and quantitative study of this group to find the information we need.

Visualization is an important part of this process. Our eyes are trained to see patterns, so visualizing a social network will give you a global snapshot of its inner structure and help you ask the right questions.

It is also a great method to communicate a certain idea and to prove your point. A picture is worth a thousand words, after all…

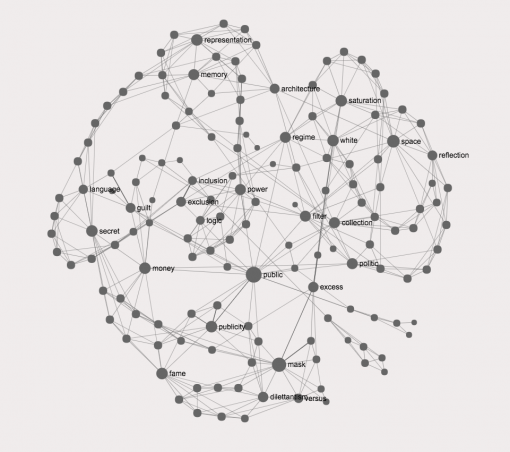
Here’s **a quick how-to list for you to perform your first SNA operation:**

1) **Identify** the **network (group)** you’re interested in  
2) Identify **what insights you’d like to have**  
3) Encode **the right data (e.g. “friendships”? collaborations?)**  
4) **Visualize** them as a graph  
5) Apply **various metrics** and **layout**  
6) **Observe, find patterns**  
7) **Reiterate**  
 

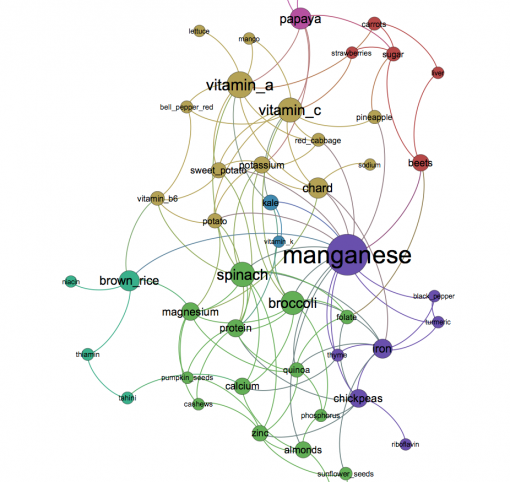
## Lesson 3 – Knowledge Network Visualization and Analysis

Network analysis can also be applied to concepts, knowledge, and texts.

For instance, in the context of a conference we could create a map of all the concepts that were mentioned in all the different talks to see how they relate to one another. This would allow us to have a comprehensive visual overview of the discourse produced as a result of the conference, and also discover the topical clusters as well as the “blind spots” – areas that were not covered.

**[](https://noduslabs.com/wp-content/uploads/2015/07/text-network-visualization-conference-sessions.png)**

Linking various foods to the vitamins and minerals that they contain, may help us find out what are the most nutritious foods and how to combine them to get the most wholesome diet.

**[](https://noduslabs.com/wp-content/uploads/2016/05/food-nutrition-vitamins-network-graph.png)**

Text network visualization will provide a good visual overview of the text and show how different concepts link together – a brief visual summary of content, which can be inspiring both for writers and for researchers.

In all these cases the strategy is similar to the one we use with social networks:

1) **Identify** the **data** you’re interested in (e.g. foods and nutrition information)  
2) Identify **what insights you’d like to have**(what are the most nutritious foods?)  
3) Identify what represents the **nodes** in the network (these could be the concepts, objects, words)  
4) Identify what represents the **edges** (connections) in the network (could be co-occurrence of objects/concepts/words)  
5) Encode the **data as a graph**  
6) Apply **basic metrics and layout**, to make it **readable**  
7) Understand the **emerging patterns**, find insights as identified in 2)  
8) **Reiterate**

## Lesson 4 – Key Network Analysis Concepts

Before we get to the practical side of things it is important to introduce some of the basic concepts used in network science. In this lesson we will provide a brief overview of those concepts and we will look into them in more detail later when we will be working on a real case study. Having a brief overview before starting is, nevertheless, very useful, because it will allow us to better formulate our goals and intentions when creating our first graph visualization.

It can be helpful to think of those concepts as **the different points of view** provided by network visualizations onto the complexities that you will be dealing with.

Those include:

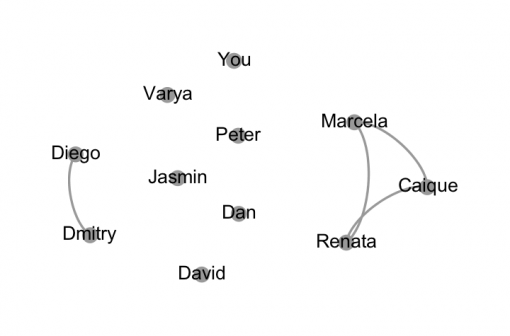
– nodes and edges (what the networks are made of)  
– clusters (groups of nodes that are related)  
– degree (number of connections that a node has)  
– betweenness centrality (how influential a node is)  
– modularity (community structure)

There are many more concepts used in network analysis, but those are the key ones and we will look into each of them in more details in the lessons that follow.

## Lesson 5 – Nodes and Edges

Any network consists of the nodes and the connections between them, the edges. The nodes represent more or less solid entities that do not change over time, the edges represent the relations, interactions, transactions or any other temporary connections that occur between the nodes over a period of time.

For example, if we were to come into a room where there are 11 people and you don’t know anyone, the network representation of this situation would look something like this:

**[](https://noduslabs.com/wp-content/uploads/2016/05/networks-nodes-edges-graph.png)**

There are 6 people who don’t know each other (including you), 3 people who know each other (the nodes Marcela, Caiqui and Renata on the right) and 2 other people who know each other (Diego and Dmitry).

This representation helps you have a good overview of the situation. Studies have shown that graph representations encourage people to make new connections in places where there are structural gaps between the nodes. Therefore, thinking in terms of the nodes and the edges does not only make you more aware of your surroundings, but also encourages you to link people up.

In this particular scenario the edges represent some friendships that have been established before the group came together in this particular constellation. So if you are to make more connections you could either follow the same logic or introduce a different **connectivity basis**, which will establish completely new constellations between the nodes. We will talk about it in more detail in the next lesson.

## Lesson 6 – Connectivity Basis

In the previous lesson we introduced the concept of **nodes** and **edges** that network graphs are comprised of. The nodes represent more or less stable entities, the edges represent the connections between them: **friendships, proximity, transactions, exchanges,** and any other temporary **links** between stable entities that happen with a certain **frequency**.

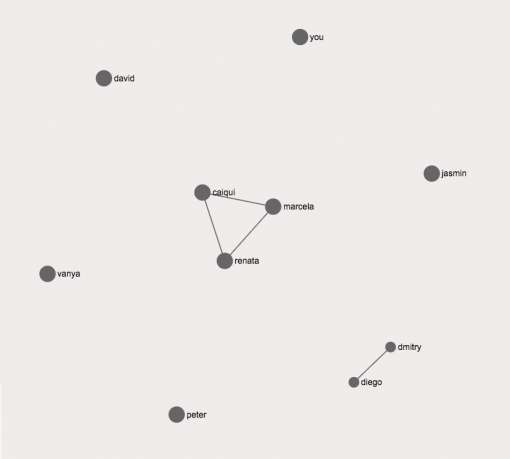
Edges are crucial to network analysis because they represent the **connectivity basis** that you will be using to get your insights about the complexity that you study.

For example, if you’re interested to know the most popular friends or to identify friend-groups in a social network, your connectivity basis is the friendship between the different nodes you’re studying. Alternatively, if you want to know which people in a group are most similar in terms of the preferences that they have and who are the most influential individuals in a group, your basis will be similarity and not friendship between them.

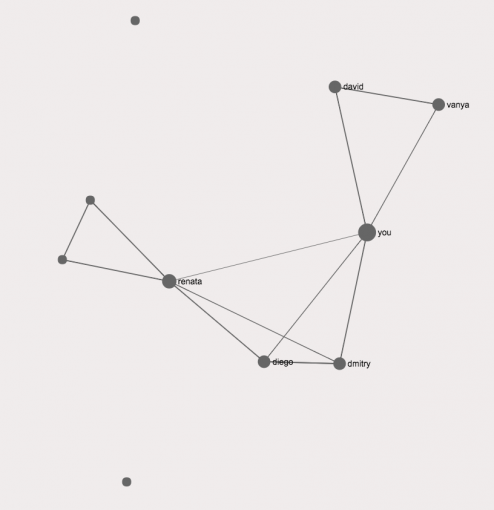
Similarly, if you’re studying a network of concepts and want to know how related they are semantically, you would use semantic proximity as the basis for connectivity when building your network.  However, if you wanted to identify the most influential concepts in a given context (e.g. political speech or a specialized conference), you would use co-occurrence as the connectivity basis – which would be much more related to the situation that you are studying.

In general, **connectivity basis**is **your point of view on the network**. What’s interesting is that by choosing a different point of view and using it to analyze and interact with a network, you may **influence the network itself** by proposing it a different connectivity basis (the observer effect).

For instance, consider the social network from the previous lesson where connectivity basis was the friendships between the people:

**[](https://noduslabs.com/wp-content/uploads/2016/05/networks-nodes-edges-graph-SNA.png)**

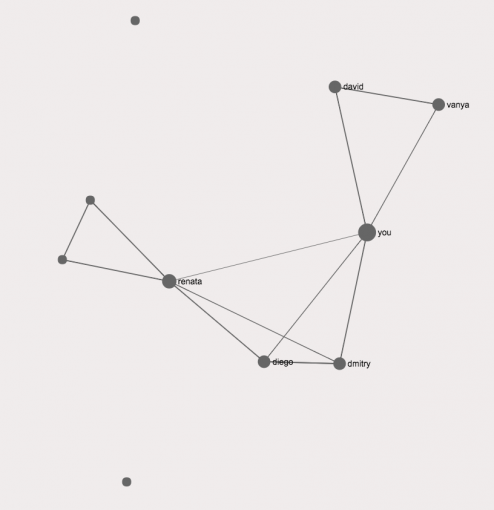
This network is pretty disconnected, so it’s difficult for all the different people (nodes) to communicate with one another, because the basis for connectivity is “friendship”, which takes time to establish.

If, however, we approach this network with a **different basis for connectivity**, such as a shared **ideology**, then the network configuration will be very different. Consider an example where we find out that we realize that “you” share marxist views with “david” and “vanya” as well as with “diego”, “dmitry” and “renata”. Then our network will look differently:   
   
**[](https://noduslabs.com/wp-content/uploads/2016/05/connecting-nodes-edges-network-graph.png)**

In the next lesson we will look at the notion of giant component, which is very important for proliferation of information inside a network.

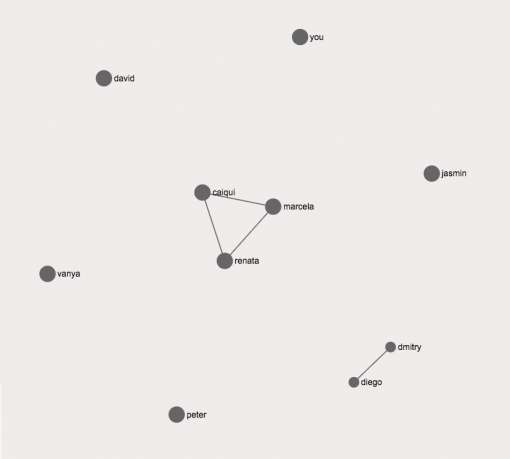
## Lesson 7 – Giant Component

Giant component is an important notion in network analysis. It’s an interconnected constellation that includes most of the nodes in a network. In the example from the previous lesson we established a giant component based on shared ideology:

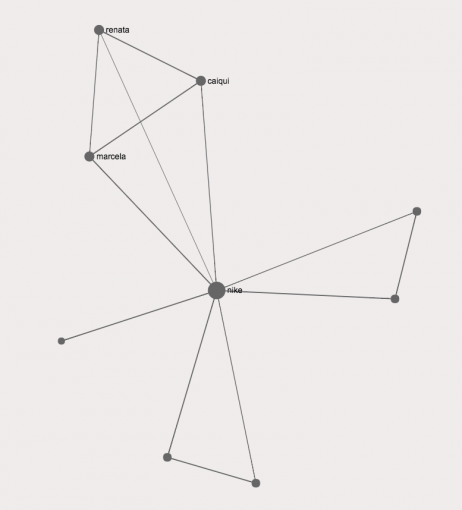
**[](https://noduslabs.com/wp-content/uploads/2016/05/connecting-nodes-edges-network-graph.png)**

When giant component exists within a network it is much easier to proliferate information through it. This can be beneficial or detrimental to the whole network, depending on the nature of content. For instance, if the shared ideology is a common desire to help each other in case of a natural disaster, the presence of giant component is of benefit to the network (e.g. Twitter users who can inform each other about an approaching hurricane). But if the same shared ideology is based on  hatred and fear, then such groups become very susceptible to external influence and manipulation (as is the case with ISIS networks, for instance).

Digital social networks use such formations for helping advertisers deliver their messages across a wide range of communities. For example, in the example above we could take the initial social network formation:

**[](https://noduslabs.com/wp-content/uploads/2016/05/networks-nodes-edges-graph-SNA.png)**

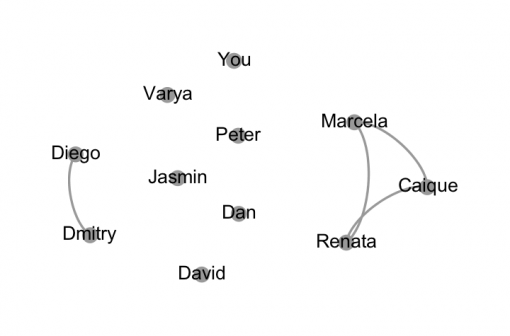
…and then create a giant component by introducing two or three popular brands into that constellation. Then all the nodes become interconnected through those different brands into a giant component (e.g. by “liking” the page of the brand on Facebook) making those brands central elements that can be used for disseminating content to different parts of the network:

**[](https://noduslabs.com/wp-content/uploads/2016/05/giant-component-branding-marketing-graph-networks-sna.png)**

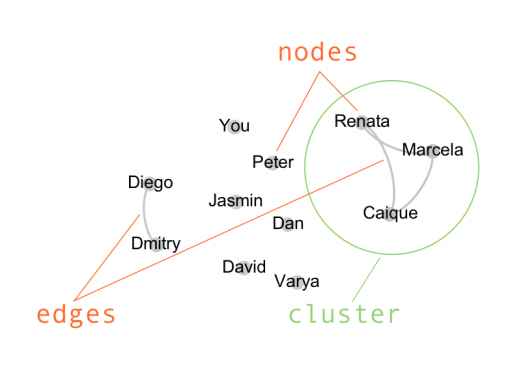
## Lesson 8 – Clusters

Clusters are the constellations of nodes that are more densely connected together than with the rest of the nodes in the network. Clusters represent different subgroups within a group and can be used to identify various subcategories that are present within.

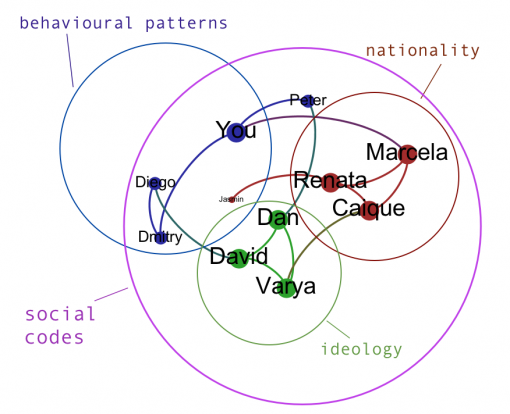
To use the example from the previous lesson, let’s consider a network of friends where only a few people know each other:

**[](https://noduslabs.com/wp-content/uploads/2016/05/networks-nodes-edges-graph.png)**

This network already has two clusters: “Marcela” – “Caiqui” and “Renata” as well as “Diego” and “Dmitry”

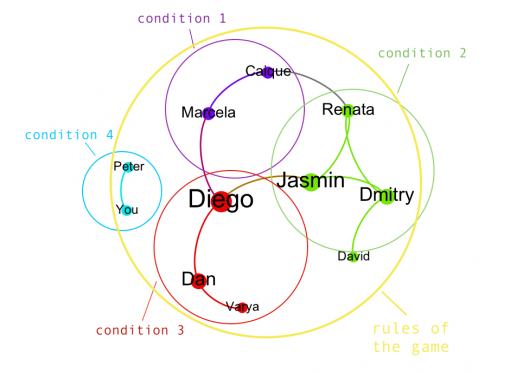
**[](https://noduslabs.com/wp-content/uploads/2012/05/Screen-Shot-2012-05-31-at-14.23.06.png)**

If we then add different connectivity bases, we will have several different clusters forming:

**[](https://noduslabs.com/wp-content/uploads/2012/05/Screen-Shot-2012-05-31-at-15.19.49.png)**

## Lesson 9 – Degree

A node’s degree indicates how many connections it has to the other nodes in the network. The more degree a node has, the more “connected” it is, which indicates its relative influence in the network.

**[](https://noduslabs.com/wp-content/uploads/2012/05/Screen-Shot-2012-05-31-at-19.24.53.png)**

In the example above, Diego, Jasmin and Dmitry are the nodes with the highest degree because they have the most number of connections to the rest of the nodes of the network. This means that they’ll play central role in transmitting information within the network to other nodes that are less connected.

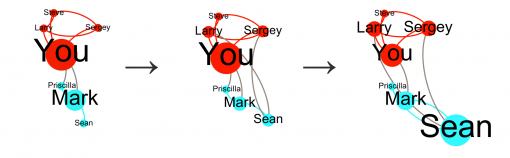
Identifying the nodes with the highest degree (also called “hubs”) is an important part of network analysis as it helps identify the most crucial parts of the network. This knowledge can then later be used both to improve network’s connectivity (by linking the hubs together) or disrupt it (by removing the nodes).

## Lesson 10 – Betweenness Centrality

Betweenness centrality is another important measure of the node’s influence within the whole network. While degree simply shows the number of connections the node has, betweenness centrality shows how often the node appears on the shortest path between any two randomly chosen nodes in a network. Thus, betweenness centrality is a much better measure of influence because it takes the whole network into account, not only the local connectivity that the node belongs to.

A node may have **high degree but low betweenness centrality**. This indicates that it’s well-connected within the cluster that it belongs to, but not so well connected to the rest of the nodes that belong to the other clusters within the network. Such nodes may have high local influence, but not globally over the whole network.

Alternatively, other nodes may have **low degree but high betweenness centrality**. Such nodes may have fewer connections, but the connections they do have are linking different groups and clusters together, making such nodes influential across the whole network. In fact, many efficient networkers and politicians will often trade some degree for betweenness centrality as it dramatically reduces their load while maintaining their central position within the network.

**[](https://noduslabs.com/wp-content/uploads/2012/05/Influence.png)**

In the example above “Sean” is a node with a low betweenness centrality, while “You” has a high betweenness centrality. “You” is connected to all the different nodes within a group, except for Sean, while Sean is only connected to Mark. However, if Sean then makes links both to Sergey and to Larry in the first cluster, connects to Priscilla, and maintains his link to Mark , he will have higher betweenness centrality than You, because he connects all the different groups that exist within the network, even though You has more connections than Sean.

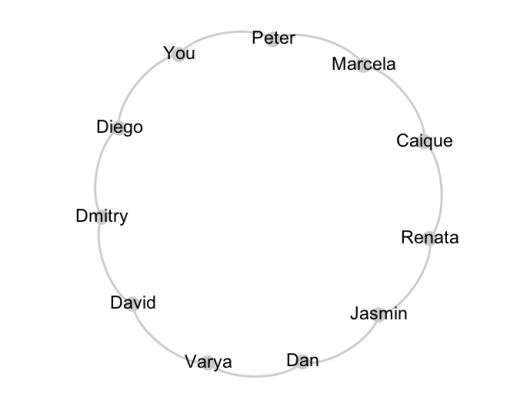
In network visualization we often**range the node sizes** by their degree or betweenness centrality to indicate the most influential nodes, as shown above.

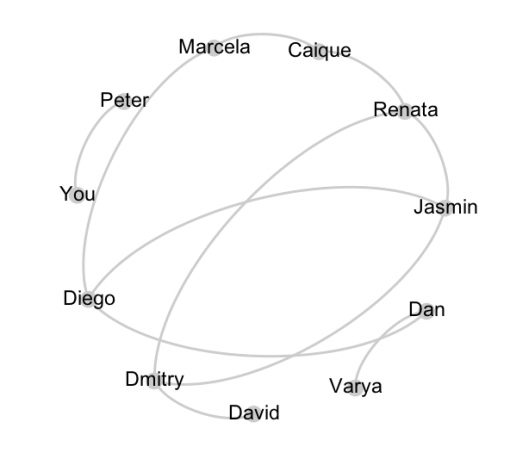
## Lesson 11 – Topology: Small World vs Regular vs Random

Network topology is an important element of network analysis. If we analyze networks on the structural basis, we will discover many differences among them.

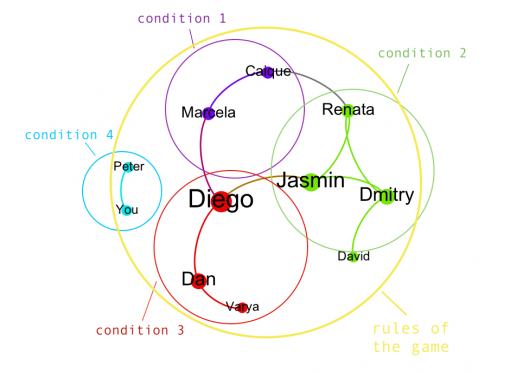
In general, there are 3 main types of networks:

– **Regular**, or highly ordered graphs (often indicating that those networks were constructed artificially)

**[](https://noduslabs.com/wp-content/uploads/2012/05/Screen-Shot-2012-05-31-at-18.43.40.png)**

– **Randomized** networks (most connections between the nodes are random):  
**[](https://noduslabs.com/wp-content/uploads/2012/05/Screen-Shot-2012-05-31-at-18.53.20.png)**

– **Small-world** networks (network consists of densely interconnected clusters, which are also connected globally):

**[](https://noduslabs.com/wp-content/uploads/2012/05/Screen-Shot-2012-05-31-at-19.24.53.png)**

It has been shown in numerous studies that the more densely connected a network is, the easier it is to proliferate information through it. However, the life-time of this information within the network will be short.

Alternatively, regular networks take the longest time to proliferate information and often it will not be able to reach all the nodes on time.

Small-world networks have been shown to have an optimal combination between their ability to proliferate and to retain information. This network topology is a typical strategy for self-organized constellations that are subject to external limitations: from societies to the human brain. Small-world networks consist of densely connected clusters, which can also interact on the global level. Such network topology allows society to have different points of views on the same issues (thus, increasing diversity) and it also makes it possible for the brain to be aware of several things at once while weaving them all together into a coherent experience.

When performing network analysis and visualization it is important to **classify the topology of the network**. This can be done through **quantitative** analysis of **degree distribution** among the nodes and / or through **qualitative** analysis using **various visual graph layouts**.

**Degree distribution** can be a good indicator of the network’s topology. If most of the nodes in the network have exactly the same degree, the network is more of a **regular** one (it may also indicate the presence of tree-like hierarchical system within the network). If most of the nodes have an average number of connections that is the same and then some of the nodes have more and some of the nodes have less (normal bell-curve distribution of degree), we’re dealing with a**randomized** network. Finally, if there’s a small, but significant number of nodes with a high degree and then degree distribution follows a long tail towards a gradual decline (scale-free distribution), this is a small-world network, where there’s a significant number of well-connected hubs, which are surrounded by less connected satellites, which form clusters. Those clusters are connected to one another via the hubs and the nodes that belong to several communities at once.

**Graph layout** a qualitative measure for identifying topology of a network. A very useful type of layout is Force Atlas, where the most connected nodes with the highest degree are pushed apart from each other, while the nodes that are connected to them but have lower degree are grouped around those hubs. After several iterations this sort of layout produces a very readable representation of a network, which can be used to better understand its structural properties and identify the most influential groups, differences between them, and structural gaps within networks.

## Lesson 12 – Network Motifs

Network motifs are the different types of constellations that emerge within network graphs. They can provide a lot of useful information about the structural nature of networks.

For example, some networks may be comprised of diads or pairs of nodes (which indicates that the level of overall connectivity is quite low). Some other networks can have a high proportion of triads, which usually indicate the presence of feedback loops, which makes the resulting network formations much more stable. More complex formations include groups of four nodes that can be connected as a sequence or between each other, forming interconnected clusters that can encode certain levels of complexity that go beyond simple triad feedback constellations.

It is important to take notice of the network motifs that emerge within a network because it will provide a very good indication of the level of complexity and thus the capacity of the network you study.

## Lesson 13 – Modularity

Modularity is a quantitative measure that indicates the presence of distinct communities within a network. If the network’s modularity is high, it means it has a pronounced community structure, which, in turn, means that there’s a space for plurality and diversity inside. If the modularity is too high, however, it might also indicate that the network consists of many disconnected communities, which are not globally connected, making it much less efficient than an interconnected one.

Modularity works through an iterative algorithm, which identifies the nodes that are more densely connected to each other than to the rest of the nodes in the network. It will then calculate the measure of modularity for the network at large. The higher this measure is, the more distinct those communities of densely connected nodes are. If the modularity measure is 0.4 or above it means that the community structure in the network is quite pronounced. If it’s less it means that there are no big differences between the different clusters and most of the nodes are equally densely connected to each other across the whole network.

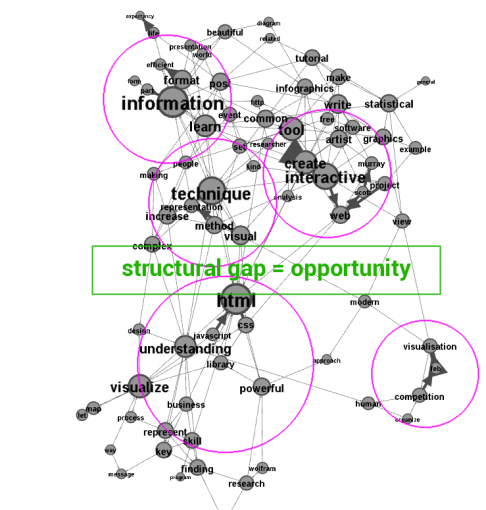
## Lesson #14 – Structural Gaps

So far, we’ve looked at the different measures of connectivity that exist within networks and that help us identify the most influential nodes, clusters, and deduce some basic functional properties of the networks we study.

However, one of the most important aspects of network graphs is that they also let you see the gaps, empty blank spaces, between the islands. Those gaps are usually referred to as “structural gaps” and it has been shown that bridging those gaps can spur innovation, create most interesting collaborations, and give rise to new, unexpected ideas.

**In other words, “structural gaps” is where creativity and potential is hidden within the network.**

Therefore, when visualizing a network, it is important to identify those structural gaps and to devise different actions that could help bridge different nodes and clusters across those empty spaces within the graph in order to spur creativity and innovation.

**[](https://noduslabs.com/wp-content/uploads/2015/03/structural-gap-text-network-data-visualization.png)**

Bottom of Form

Steps in Gephi

1. Load the data.
2. On left panel (Layout) select ForceAtlas 2 and Run. This seems to be a good layout. I stopped it after a minute or so because it will run long.
3. Select the Expansion Layout and play with it to give you a better visibility
4. Right panel, select Statistics tab and run:
   1. Average Degree / Average Weighted Degree (degree, in-degree and out-degree)
5. Right panel, Statistics tab, run (directed and normalize)
   1. Network Diameter (betweenness centrality, closeness centrality & eccentricity
6. Right panel, Statistics tab, run eigenvector centrality (directed)

Begin the analysis:

Right panel, Filters tab, Topology->Degree Range….

Shows that the degree is between 128 and 254, we will look at nodes that have degree higher than 200 to see highly connected nodes:

With 200 – 254 range I still get a very dense network with 540 modes and more than 30000 edges

220 – 254 range for degree reduces the network to 85 nodes which is a much more manageable set of strongly connected nodes.

Letf panel, Appearance->Nodes->Ranking and pick your color scheme and Apply