Università degli Studi di Roma Tre

Dipartimento di Matematica e Fisica

Corso di Laurea Magistrale in Scienze Computazionali

Tesi di Laurea Magistrale - CAPITOLO 0

Axioms for Centrality

Relatore:
Dr. Marco Liverani

Candidata:
Sara Prevete

Correlatore:
Dr. Stefano Guarino

Anno Accademico 2019/2020

Contents

0.1 Introduction

Complex and large networks have generated much interest over the last few years, as they can be used to describe several biological, social and technological systems.

There are different ways to analyze a network.
Given a network it is often important to understand which of its nodes is the most central.
The concept of centrality has been used several times in sociology, psychology and computer science. Measuring the centrality of a given node means understanding its role and importance within a specific network.
Different analyses of a network have discovered diverse properties of nodes. For each property, a specific measure of centrality has been formulated: some are based on positional aspects of the nodes, how their position affects their role in network connectivity. Other properties, instead, study the mediation force of a node, its influence on neighbor nodes, and its importance in the flow of information.

Despite all these multiple proposals, a satisfying definition of "what makes a central node?" has not yet been given.
The only clear thing, on which everyone agrees, is that centrality is a construct at the level of the nodes.
In this regard, we could set ourselves other objectives:
Understanding what characterizes the category;
Looking for what all these measures of centrality have in common;
Finding, if any, structural properties of the nodes that are not measures of centrality;

In this work we would like to provide a strong investigation of the most important classical centrality measures and propose an axiomatic approach to verify if the above measures are working for what they are designed for.
In other words, we will try to define the centrality of graphs through a system of axioms. Therefore, we will investigate the axioms that identify centrality, study the measures that validate these axioms, and finally understand what has prevented us from providing a valid definition until now.

0.2 Centrality in Graphs

0.2.1 Notions of graph theory

In the study of social networks, and therefore in our case in the search for centrality, a valid starting point is given by the theory of graphs which provides us with a mathematical approach and a language for the description of the networks and their characters. In fact, it provides us with a basis of concepts and theorems that adapt and can represent the essential characteristics of social networks. So let's see some basic elements of graph theory.

A graph $G = (V, E)$ is a pair of disjoint sets: V is a non-empty, discrete, finite set, whose elements are the vertices of the graph, sometimes also called nodes or points, while E is the set of edges of the graph, also called arcs or sides, which is a subset of the Cartesian product $V \times V$, or a set of pairs of vertices of the graph. Often, for greater clarity, $V(G)$ and $E(G)$ denote respectively the set of vertices and edges of graph G. Conventionally, the letters n and m indicate the cardinality of V and E, that is the number of vertices and edges of the graph: $n = |V(G)|$ and $m = |E(G)|$.
The graph $G = (V, E)$ with $V(G) = v$ and $E(G) = \emptyset$ is the null graph.

If $e \in E(G)$ and $e = (u, v)$, with $u, v \in V(G)$, then the vertices u and v are said to be adjacent to each other and constitute the ends of the edge e; at the same time we will say that the edge e is incident on the vertices u and v. An edge (u, u) from a vertex u in itself is a loop.
If for each pair of vertices $u, v \in V(G)$ there is at most only one edge $(u, v) \in E(G)$ we will say that the graph G is simple. On the contrary, a multigraph is a graph in which there are two or more distinct edges that have the same pair of vertices as ends.

The graph $G = (V, E)$ is oriented if the edges are considered as ordered pairs of vertices: in this case $(u, v) \neq (v, u)$. If $e = (u, v) \in E(G)$ is an edge of the oriented graph G, we will say that the edge is outgoing from u and is entering in v.
Vice versa if the edges of the graph are considered as unordered pairs, that is if $(u, v) = (v, u)$ for each edge of the graph, then we will say that the graph is not oriented.
However in the following, for simplicity, we will indicate with (u, v) both the non-oriented edges, and those with a direction, belonging to an oriented graph, explaining where it should not be clear from the context, if it is oriented edges or not oriented. It is good to specify that an oriented graph G

such that $(u, v), (v, u) \in E(G)$ is not a multigraph, since the two edges come out and enter in different vertices.

Given an oriented graph $G = (V, E)$, it is possible to define the transposed graph G^T, placing $V(G^T) = V(G)$ and $E(G^T) = \{(u, v) : u, v \in V(G)$ and $(v, u) \in E(G)\}$; in fact it is the graph obtained by G reversing the direction of the edges.
It is often very useful to produce a graphical representation of a graph using a drawing. Graphs can be drawn by representing vertices as points and edges as segments (oriented or non-oriented) or curves that join pairs of vertices.
If $E(G) \equiv V(G) \times V(G)$ then we will say that the graph is complete: in this case, for each pair of distinct vertices there is an edge that connects them. We will denote the complete graph with n vertices with the notation K_n.

A graph with "many" edges with respect to the number of vertices is called dense, while on the contrary a graph with "few" edges is called sparse; in general we can say that a graph is scattered if the number of edges is of the same order of magnitude as the number of vertices: $\mid E(G) \mid = O(\mid V(G) \mid)$.

A weighted graph is a graph with which weights have been associated with each of its vertices or with each of its edges.
In the case of graphs weighted on the vertices the weight of a vertex is a function $w : V(G) \rightarrow R$; in the case of graphs weighed on the edges, the weight function w is defined as $w : E(G) \rightarrow R$.
In an undirected graph, the degree of a vertex v, $deg(v)$, is given by the number of edges incident on it; if $deg(v) = 0$ then we will say that v is an isolated vertex, while if $deg(v) = n - 1$, that is if v is adjacent to every other vertex of the graph, v is universal.
In a graph we define the maximum degree, $\Delta(G)$, given by the maximum degree of its vertices, and the minimum degree, $\delta(G)$, given by the minimum degree of its vertices.
A graph is regular if the vertices all have the same degree ($\Delta(G) = \delta(G)$); in particular if $deg(v) = p$ for every $v \in V(G)$ then G is a p-regular graph.
It will be helpful to represent a graph in terms of its adjacency matrix A, in which a $ij = 1$ if (i, j) is in E.

Nodes that are not adjacent may nevertheless be reachable from one to the other. A walk from node u to node v is a sequence of adjacent nodes that begins with u and ends with v. A trail is a walk in which no edge is repeated. A path is a trail in which no node is visited more than once.
The length of a walk is defined as the number of edges it contains, and the

shortest path between two nodes is a known as a geodesic.
The length of a geodesic path between two nodes is known as the geodesic or graph-theoretic distance between them. We can represent the graph theoretic distances between all pairs of nodes as a matrix D in which d_{ij} gives the length of the shortest path from node i to node j.
A geodesic is a shortest path between two nodes.

Given an undirected graph $G(V, E)$, con $|V| = n$, $|E| = m$ $x_i \in V(G) \forall i = 1, ..., n$, let $dist(x_i, x_j)$ is the distance between the nodes x_i and x_j ;$g_{jk}(i)$ is the number of shortest paths between node x_j and x_k passing through node x_i (geodesic).

A cycle is a path that has the same starting and ending node.
A graph is said to be connected if there exists a path between every two nodes in the graph.
A tree is a connected graph that does not contain any cycles.
A component of a graph is a maximal connected subgraph of the graph. The component to which i belongs is denoted $C_g(i)$.

0.2.2 A short historical account

As we know, centrality is a fundamental concept for the study of network analysis.
We briefly summarize the development of the concept of centrality by following [1, 4]. As early as the late 1940s Bavelas (1948,1950) and Leavitt (1951) realized they could use centrality to explain different communication performances and network members work on a number of variables including problem solving time, number of mistakes, leadership perception, efficiency and job satisfaction.
Their research has led to a great deal of experimental, theoretical and implications of the network structure especially in the context of organizations.

Consequently, many studies followed to use centrality to analyze influence in interorganizational networks (Laumann and Pappi, 1973; Marsden and Laumann, 1977; Galaskiewicz, 1979), power (Burt, 1982; Knoke and Burt, 1983), in exchange networks Marsden, 1982), competence in formal organizations (Blau, 1963), job opportunities (Granovetter, 1974), adoption of innovation (Coleman et al., 1966), corporate interlocks (Mariolis, 1975; Mintz and Schwartz, 1985; Mizruchi, 1982), power in organizations (Brass, 1984, and differential growth rates between medieval cities (Pitts, 1979).

Although many centrality measures were immediately proposed, each suitable for analyzing a specific quality of the network, the category itself has never been well defined.
As we have already mentioned, however, the one thing everyone agrees on is that centrality is a construct at the node level. But what specifically defines the category? What do all centrality measures have in common? Are there structural properties of nodes that are not measures of centrality?

Sabidussi, in 1966, tries to provide a mathematical answer to these questions. He suggested a number of criteria that measures must meet in order to qualify as centrality measures. For example, he felt that adding an edge to a node should always increase the centrality of the node and that adding an edge anywhere in the network should never reduce the centrality of any node.
These requirements seem reasonable, and it's easy to see the value of separating measures that do well from those that don't.
Approaching more actively to the method given by Sabidussi, we immediately notice some things that are not good.
First, it appears that its criteria eliminate the most well-known measures of centrality, including the centrality between.
This is unfortunately a counter productive result.
Furthermore, while its criteria provide some desirable, prescriptive characteristics for a centrality measure, they do not actually attempt to explain what centrality is.

Later, in 1979, Freeman provided another approach to answer the question "what is centrality?".
He examined a number of published measures and reduced them to three basic concepts for which he provided canonical formulations.
These were the degree, the proximity and the between.
He noted that all three reach their maximum values for the center of a star-shaped network, consequently argued that this property is the defining characteristic of centrality measures.

Borgatti (2005) has recently proposed a dynamic, model-based view of centrality that focuses on outcomes for nodes in a network where something flows from node to node across edges. He argues that the fundamental questions that must be asked about individual nodes in the context of dynamic flow are:
(a) how often traffic flows through a node,

(b) how long does it take to get to a node.
Once these questions are set, it becomes easier to construct theoretical measures of graphs based on the network structure that predict the answers to these questions.
Therefore, in this approach, the centrality measures are expressed as predictive models of specific properties of the network flows.

0.3 Axioms for Centrality

Various are the centrality measures that have been studied in the literature, each of which is used to characterize a different individual property.

0.3.1 The Centrality measures

Geometric measure

We call *geometric* those measures assuming that importance is a function of distances.

Degree measure The *degree centrality* indicates the potential communication activity of a node: attaches high value to individuals who have a great influence on their neighbors.

$$c_D(x_i) = \frac{deg(x_i)}{n-1} \tag{1}$$

Decay measure

$$M^{\delta}(i,g) = \sum_{j \neq i} \delta^{l^g(i,j)} \tag{2}$$

Where for a given network $G = (N, g), i \in N$ and $0 < \delta < 1$,

Note that as δ gets close to 0, $M\delta$ approaches M^{deg} and as δ gets close to 1, $M\delta$ counts the total number of nodes in the component to which i belongs. For intermediate values of δ, $M\delta$ is similar to M^{close} and measures how close i is to other nodes on average.

Closeness measure The *closeness* of x is defined by

$$\frac{1}{\sum_y d(x,y)} \tag{3}$$

It indicates the potential of a node in communication control: emphasizes players who have easy contact with everyone else.
Closeness is that nodes that are more central have smaller distances, and thus a smaller denominator, resulting in a larger centrality.

Assuming that nodes with an empty coreachable set have centrality 0 by definition.

Lin's index Nan Lin tired to repair the *definition of closeness for graphs with infinite distances* by weighting closeness using the square of the number of coreachable nodes; his definition for the centrality of a node x with a nonempty coreachable set is:

$$\frac{|\{y|d(y,x) < \infty\}|^2}{\sum_{d(y,x)<\infty} d(y,x)} \tag{4}$$

Harminic centrality Marchiori and Latora propose to replace the average distance with the harmonic mean of all distances.

Indeed, in case a large number of pairs of nodes are not reachable, the average of finite distances can be misleading: a graph might have a very low average distance while it is almost completely disconnected. The harmonic mean has the useful property of handling cleanly.
In general,for each graph-theoretical notion based on arithmetic averaging or maximization there is an equivalent notion based on the harmonic mean.

We thus define the *harmonic centrality* of x as:

$$\sum_{x\neq y} \frac{1}{d(y,x)} = \sum_{d(y,x)<\infty, y\neq x} \frac{1}{d(y,x)} \tag{5}$$

Spectral measure

Spectral measures compute the left dominant eigenvector of some matrix derived from the graph, and depending on how the matrix is modified before

the computation we can obtain a number of different measures.
Existence and uniqueness of such measures is usually derivable by the theory of nonnegative matrices.

The left dominant eigenvector The *left dominant eigenvector* of the plain adjacency matrix. Can be thought as the fixed point of an iterated computation in which every node starts with the same score, and then replaces its score with the sum of the scores of its predecessors.
The vector is then normalized, and the process repeated until convergence.
Dominant eigenvectors fail to behave as expected on graphs that are not strongly connected.

Seeley's index The dominant eigenvector rationale can be slightly amended. Each has a reputation and is giving its reputation to its successors so that they can build their own.
It is more reasonable to divide equally our reputation among our successors.
From a linear-algebra viewpoint, this corresponds to normalizing each row of the adjacency matrix using the l_1 norm.
The matrix resulting from the l_1-normalization process is stochastic, so the score can be interpreted as the stationary state of a Markov chain.
Also *Seeley's index* does not react very well to the lack of strong connectivity.

Katz's index Katz introduced his celebrated index using a summation over all paths coming into a node, but weighting each path so that the summation would be finite.
Katz's index can be expressed as

$$k = 1 \sum_i \beta^i A^i \tag{6}$$

The attenuation factor β must be smaller than $1/\lambda$, where λ is the dominant eigenvalue of A.
Katz immediately noted that the index was expressible using linear algebra operations: $k = 1(1 - \beta A) - 1$.
Katz's index is the left dominant eigenvector of aperturbed matrix $\beta\lambda A + (1 - \beta\lambda)e^T 1$.

e is a right dominant eigenvector of A such that $1e^T = \lambda$.

The normalized limit of Katz's index when $\beta \to 1/\lambda$ is a dominant eigenvector.

We can see the Katz measures also as a weighted count of generic walks which can also go over the same nodes. Here the extent to which the weights decrease with length is an arbitrary parameter b

$$c_i = \sum_j wijw_{ij} = ba_{ij} + b^2(a^2)_{ij} + ... = \sum_{k01} b^k(a^k)_{ij} \tag{7}$$

where k is the length of the walk.

PageRank The *PageRank* is one of the most discussed and quoted spectral indices. PageRank is the unique vector p satisfying:

$p = \alpha p A^{-} + (1-\alpha)v$,

where A^{-} is again the $l1$-normalized adjacency matrix of the graph, $\alpha \in [0..1)$ is a damping factor, and v is a preference vector.

Brinand Page themselves propose a different but essentially equivalent line an applicant in the style of *Hubbell's index*, and acknowledge that A^{-} can have null rows, in which case the dominant eigenvalue of A^{-} could be smaller than one, and the solution might need to be normalized to have unit $l1$ norm.

Equation is of course solvable even without any patching, giving:

$p = (1-\alpha)v(1-\alpha\overline{A})^{-1}$

and finally: $p = (1-\alpha)v\sum_i \alpha^i \overline{A}^i$

HITS The *HITS algorithm* using the web metaphor of "mutual reinforcement": a page is authoritative if it is pointed by many good hubs, and a hub is good if it points to authoritative pages.

This process converges to the left dominant eigenvector of the matrix $A^T A$.

Path-based measures

The *path-based measures* exploit not only the existence of shortest paths but actually take into examination all shortest paths (or all paths) coming into a node. We remark that indegree can be considered a path-based measure, as it is the equivalent to the number of incoming paths of length one.

Betweenness The *betweenness centrality measure* computes the probability that a random shortest path passes through a given node:

$$c_B(x_i) = \frac{2\sum\sum g_{jk}(i)}{(n-1)(n-2)} \tag{8}$$

The intuition behind betweenness is that if a large fraction of shortest paths passes through x, then x is an important junction point of the network. It is the index of the independence of a node: emphasizes individuals through whom it is easy to pass information.

Spectral measures as path-based measures The *spectral measures* can be interpreted as path-based measures and in both cases we can express these algebraic operations in terms of suitable paths.
The left dominant eigenvector of a nonnegative matrix can be computed with the power method by taking the limit of $1A^k/||1A^k||$
for $k \to \infty$. Analogously, Seeley's index can be computed by taking the limit of $1\overline{A}^k$: it assigns to each x the sums of the weights of the paths coming in to x.

0.3.2 A family of centrality measures

It is also interesting (as Manuj Garg does in his article "Axiomatic foundations of centrality in networks", 2009) [3] to axiomatize the main centrality measures in distant families and study the characteristics of family areas.

In his article we will see the degree, closeness and decay measures which all belong to the same family of measures.
The feature that links these measures is the limitation for breadth first search tree research.
Another common feature is that the centrality of a node, defined by the first three measures, is additively separable in all the other nodes.
These measures assign the same centrality to symmetric nodes and the maximum centrality to the central ones of a star.
He goes so far as to say that the degree, closeness and decay centrality measures belong to the same family of measure.

For each of these three measures it also defines a series of axioms that characterize it, which we will briefly discuss below.

There are four axioms completely characterize the degree centrality measure.

Axiom 0.3.2.1 (Isolation)

$$Ng(i) = \emptyset \Rightarrow M(i,g) = 0 \tag{9}$$

Axiom 0.3.2.2 (Simmetry) *If $i, j \in N$ are symmetric in g, then $M(i, g) = M(j, g)$.*

Axiom 0.3.2.3 (Additivity) *Let $Pg = \{g1, ..., gk\}$ be a subnetwork-partition of g, then $M(i, g) = \sum_{l=1,...,k} M(i, g_l)$.*

Axiom 0.3.2.4 (Star Maximization) *$M()$ should be such that $\{(c*, g*)\} \in argmax_{(i,g)\in N\times G(N)} M(i, g)$.*

Where he have considered the follow definition:

Definition 0.3.2.1 (Symmetric Nodes) *For a given network $G = (N, g)$, nodes $i, j \in N$ are symmetric in g if $\exists\pi : N \longrightarrow N$, a permutation on N, s.t. $\pi(i) = j, \pi(j) = i$ and $g^\pi = g$; where $g^\pi = \{\pi(i)\pi(j)|ij \in g\}$.*

It is easy to see that $M^{deg}()$ satisfies the above axiom, two symmetric nodes always have the same number of neighbors.

Definition 0.3.2.2 (Subnetwork-Partition) *$P_g = \{g_1, ..., g_k\}$ is a* subnetwork-partition *of g if*

a $Gl = (N, g_l)$ is a subnetwork of $G = (N, g)\forall l \in \{1, ..., k\}$.

b $\cup_{l\in\{1,...,k\}} g_l = g$

c $g_l \cap g_l = \emptyset \forall l, l' \in \{1, ..., k\}$

Where the subnetwork-partition is a partition of g into distinct subsets of itself such that their union is g. The crucial thing is that for each $l, N(g_l) = N$,each g_l is defined over the entire set of nodes.

And the last axiom is based on star network. Formally, a star network is a network $G* = (N, g*)$ such that $g* = \{c * j|j \in N \ c^*\}$, where the node c^* is called the center or hub of the network.

Then he presents the characterization of the closeness centrality measure.

Axiom 0.3.2.5 (Breadth-First Search) *Given $G = (N, g)$ and $i \in N$, for any $G_B(i) = (N, g_B(i)) \in Ti$, $M(i, g) = M(i, g_B(i))$*

Axiom 0.3.2.6 (C-Additivity) *For any $G_B(i) \in Ti$, and the corresponding subnetwork-partition, $P_{g_B}(i) = \{g_1(i), ..., g_K(i)\}$, then*

- $M(i, g_B(i)) = \sum_{k=1,\dots,K} M(i, g_k(i))$.

- $M(i, gk(i)) = \sum_{j^k \in T_k(i)} f(j^k)$ *for some function* $f()$.

Axiom 0.3.2.7 (Closeness) *For each* $j^k \in T_k(i)$*, then*

$$f(j^k) = \frac{M(j^k, \{j^k j^{k-1})\}}{k} \tag{10}$$

where $j^k \in N_{j^{k-1}}(g)$ *and* $j^{k-1} \in T_{k-1}(i)$*;* $j^0 \equiv i$.

For for this characterization we need some background on Breadth-First Search Trees.
A cycle is a path that has the same starting and ending node.
A network is said to be connected if there exists a path between every two nodes in the network. A tree is a connected network that does not contain any cycles. A component of a network is a maximal connected subnetwork of the network.

Formally, $C = (N', g')$ is a component of $G = (N, g)$ if

- $N' \subset N, g' \subset g$;

- (N', g') is connected;

- if $i \in N'$ and $ij \in g$, then $j \in N'$ and $ij \in g'$.

The component to which i belongs is denoted $Cg(i)$.

Breadth-First Search Trees: Given a network G and a node $i \in G$, we can define a Breadth-First Search Tree (BFST) in G rooted at i as follows: begin at node i, which we call the root node, and explore for all its neighboring nodes. Then for each of those neighboring nodes, explore for their unexplored neighboring nodes, and so on, until there are no more nodes left to explore – either because all the nodes of the network have been exhausted or all remaining nodes are disconnected from i, they are in a different component of the network than i belongs to, $Cg(i)$.

For the purpose of the second axiom, define, for a fixed $G_B(i), G_k(i) = (N, g_k(i))$ as the level-k-search-tree. Note that in $G_k(i)$, each node in $T_{k-1}(i)$ is connected with at least one node in $T_k(i)$.
All other nodes are disconnected with these nodes and each other, they are isolated in $G_k(i)$. In the notation of definition, $P_{g_B}(i) = \{g_1(i), ..., g_K(i)\}$.

Axiom of C-Additivity captures two important additivity (or separability) features of the measure:
First, it says that i derives its centrality independently from each level of $G_B(i)$.
Second, within each level, i's centrality is additively separable in each node of the level.

The next axiom can now be stated.

Axiom 0.3.2.8 (Up-Closure) *For any $G_B(i) \in T_i$ and a given $0 < \delta < 1$, then*

$$M(i, g_B(i)) = \sum_{j^1 \in T_1(i)} \delta[M(j^1, g_B(i)|\lfloor j^1 \rfloor) + 1] \tag{11}$$

This expression is easier to interpret: i's centrality is δ times the sum of the centrality of all the nodes in its neighborhood in their respective up-closure trees corresponding to a given BFST, plus a δ for each of those neighboring nodes.

Thus far, we have a characterization for degree, closeness and decay centrality measures.
Going through those characterizations, it seems that there isn't much common between these measures.
I show that these three measures are a closely related family.
By only varying Axiom of Closeness we can get the other two measures.
So, let us define two new axioms which will give us this family of measures.

The modification of Axiom of Closeness we need to characterize $M^{deg}()$ is the following.

Axiom 0.3.2.9 (Degree) *For any $G_B(i) \in T_i$, for each $j^k \in T_k(i)$, then*

$$f(j^k) = \frac{M(j^k, \{j^k j^{k-1}\})}{k} \boldsymbol{1}_{k=1} \tag{12}$$

where $j^k \in N_{j^{k-1}}(g)$ for some $j^{k-1} \in T_{k-1}(i)$;
$j_0 \equiv i$ and $\boldsymbol{1}$ is the indicator function.

This axiom clearly shows that $M^{deg}()$ is a special case of $M^{close}()$. According to this axiom, the centrality of a node i does not depend on any node other those in it's neighborhood. This is the essential feature of $M^{deg}()$ that we emphasized earlier as well, but capture it in a different way here.

The modification of Axiom of Closeness to characterize $M\delta()$ is as follows:

Definition 0.3.2.3 (Axiom of Decay) *For any $G_B(i) \in T_i$ and a given $0 < \delta < 1$, for each $j^k \in T_k(i), f(j^k) = \delta f(j^{k-1})$ where $j^k \in Nj^{k-1}(g)$ for some $j^{k-1} \in T_{k-1}(i)$; $k \geq 2$.*

Note that Axiom of Decay above is silent about $f(j^1)$

We can now show that the three measures belong to the **same family**.

Proposition 1 *Suppose $M()$ satisfies Axioms of Symmetry, Star Max, BFS and C-Additivity. Then:*

- *$M() = M^{close}()$ if and only if $M()$ satisfies Axiom Closeness as well.*
- *$M() = M^{deg}()$ if and only if $M()$ satisfies Axiom Degree as well.*
- *$M() = M^{\delta}()$ if and only if $M()$ satisfies Axiom Decay as well.*

It is useful to note here that betweenness and eigenvector centrality measures both violate Axiom BFS. This is the key distinguishing feature between these measures and the measures characterized in this paper. Additionally, the centrality of a node when defined by either of these measures is not additively separable in other nodes.

The aim of his work is to axiomatize the standard centrality measures into distinct families.
characterizations of three important centrality measures – degree, closeness and decay – and establishing them as part of the same family.
We can now redefine degree centrality in terms of the out-degree of a node.
Such axiomatizationes lend structure to, and provide a better understanding of, the vast array of centrality measures that exist. Moreover, it is easier to distinguish between different measures once we know their precise structural bases.

0.3.3 Closeness centrality VS harmonic centrality

An alternative to the proximity centrality index - harmonic centrality index - which give comparable results and present a possible interpretation on a disconnected graph without closeness centrality is study by Even Yannch

Rochart in work "Closeness centrality extended to unconnected graph: The harmonic centrality index" (2009), [5] where he finds in the measure of harmonic centrality something more.

For each node, to calculate the closeness centrality index, we need the distance between all the pairs of vertices, i.e. the geodesic distance. To do this we go to write a matrix. If the graph is disconnected the distances between the vertices of two different components is infinite. In this case the closeness centrality index is useless.

We study so how can you use the closeness centrality index for any graph. A proposal is to write the infinite distance between two vertices as two distinct components of the number of vertices of the graph: the minimum maximum path in a graph with n vertices $n - 1$. I can generalize

$$c_{\alpha}(x_i) = \frac{n-1}{\sum_{i \neq j} dist(x_i, x_j) + m\alpha} \tag{13}$$

where $m = |E|$ and $\alpha \in R, \alpha$ costant $\geqslant$ the diameter of the graph.

The innovative proposal of this article is the index of harmonic centrality.

$$\frac{1}{1-n} \sum_{i \neq j} \frac{1}{dist(x_i, x_j)} \tag{14}$$

This index attaches greater importance to well-connected vertices.

If the graph is disconnected I will still have lower values. and this reflects the inability of individuals of different components to communicate with each other, and the maximum value is in the center of the stars where the index is $\frac{1}{n-1}(1 + (n-2)\frac{1}{2}) = \frac{n}{2(n-1)}$

Three types of networks are used in these studies:

- random networks
- real networks.
- scale-free networks,

For each node we calculate the two indices (harmonic and closeness) and compare the ranks using the Sperman ρ correlation.

During the simulation the standard deviation and the ρ were calculated.

For real graphs it is calculated only at the ρ.

For random graphs the generalization starts with unconnected vertices and for each pair of vertices is created a edge with fixed probability.

For scale-free scratches at each step we add both vertices and sides proportionally. The probability of two nodes being tied is proportional to their degree of connection.
The computational complexity of harmonic centrality is $(n|E|)$ for closeness centrality.

When the degree is connected these two centralities are very similar:
The nodes that are close to the node we are interested in will improve their size (i.e. if a node is in a dense cluster, even if small, it will have a high value index); in this calculation the harmonic centrality is compatible with the closeness centrality.
If instead we have a graph not connected with the harmonic centrality I have a value other than zero. This does not mean that the node can communicate with everyone, but represents its role in the graph. Being in a small component does not imply having a small harmonic centrality.
Harmonic centrality with high values, respect for closeness centrality, even for disconnected and scattered graphs.

0.3.4 What do centrality measures measure?

Starting from what Freeman, Borgatti and Everett did in their article "A graph-theoretic perspective on centrality" (2005), [1] they analyze the measures of Betweenness, Closeness, and Degree and many of their variants and study it in relation to four characteristics: types of walks considered (geodesics, disjoint sides, etc...), types of summary (average or sum), the properties of walk (length and volume), position of the node involved (radial or medial).

It is apparent in this review of measures that all of the measures evaluate a node's involvement in the walk structure of a network. That is, they evaluate the volume or length of walks of some kind that originate, terminate, or pass through a node. Furthermore, all are based on the marginals of an appropriately constructed node-by-node matrix, although the method of calculating marginals can vary from simple sums to averages and weighted averages to harmonic means, and so on. Thus four basic dimensions distinguish between centrality measures: the types of walks considered (called Walk Type, such as geodesic or edge-disjoint), the properties of walks measured (called Walk Property, namely volume or length), the type of nodal involvement (called Walk Position, namely radial or medial), and type of summarization (called Summary Type, such as sum or average).
The Walk Type dimension concerns the restrictions that some measures im-

	RADIAL	MEDIAL
VOLUME	Freeman degree, Sade k-path, Bonacich eigenvector, Katz status, Hubbell status, Hoede status, Doreian iterated Hubbell, Markovsky et al. GPI, Friedkin TEC, Coleman power, Bonacich power Burt prestige	Anthonisse rush Freeman betweenness, Freeman et al. flow between, Friedkin MEC
LENGTH	Freeman closeness, Stephenson-Zelen information Friedkin IEC	Borgatti FD

pose on the kind of walks considered, such only geodesics, only true paths, limited length walks, and so on. The Walk Property dimension distinguishes between measures that evaluate the number of walks a node is involved in from measures that evaluate the length of those walks. The Walk Position dimension distinguishes between measures that evaluate walks emanating from a node from measures that evaluate walks passing through a node.

The choice between radial and medial measures can be seen in terms of the distinct roles played by nodes in the network.

A radial measure of volume counts the number of these paths in which a given node serves as an endpoint.

A medial measure counts the number of these paths in which the node serves as an interior point.
Together, the radial and the medial add up to the total number of paths that a node is involved with in any role. In this sense, we can speak of decomposing a node's total involvement in the paths of a network into radial and medial portions.

If so, radial and medial measures are complementary and both are needed to deliver a complete picture of a node's contribution to the network .

Total Involvement = Radiality + Mediality

In conclusion we can say that following Sabidussi (1966), we have described the notion of centrality in purely graph-theoretic terms: what all measures of centrality do is assess a node's involvement in the walk structure of a network. This is the graph-theoretic answer to the question 'What do centrality measures measure?' We have suggested that centrality measures differ along four key dimensions: choice of summary measure, type of walk considered, property of walk assessed, and type of involvement.

The choice of summary dimension has the least variance, consisting mostly of simple sums and averages, along with a few exemplars of weighted sums (e.g., eigenvectors) and centroids. The type of walk dimension distinguishes measures based on edges, geodesics, paths, trails and walks. The property of walk dimension distinguishes between volume and length measures. The type of involvement dimension distinguishes between radial and medial measures.

It can be seen that the single distinction made by Borgatti between frequency and time can be derived as a collapsing of the property of walk and type of involvement dimensions.

The medial measures essentially measure the impact of the presence of a node on the dyadic cohesion among all pairs of nodes. In other words, they measure the change in cohesion that would result from removing a given node.

As such, medial measures do not depend on core/periphery structures for interpretability, and in fact are particularly useful when networks have "clumpy" structures characterized by wide variation in local density. At a general level, we note the relationship of centrality concepts with the concepts of graph cohesion and cohesive subgroups. The key underlying concept is that of dyadic cohesion—the social proximity of pairs of actors in a network.

Dyadic cohesion is what is measured by the W matrix that undergirds all measures of centrality. There are two fundamental ways of analyzing cohesion. One is to seek regions of the network that are more cohesive than others — a focus on the pattern of cohesion.
The other is to attribute to individual nodes their share of responsibility for the cohesion of the network — a focus on the amount of cohesion.

0.4 Thesis Layout

Follow the work of Vigna and Boldi, [7], we can analyze eleven centralities measure and three axioms of centrality for

0.4.1 The work of Vigna and Boldi for the centrality

They try to provide a mathematically survey of the most important classic centrality measures known from the literature and propose an axiomatic approach to establish whether they are actually doing what they have been designed for.
Surprisingly, only a new simple measure based on distances, harmonic centrality, turns out to satisfy all axioms. The harmonic centrality is a correction to Bavelas's classic closeness centrality designed to take into account unreachable nodes.

One of the most important notions that researchers have been trying to capture in such networks is "node centrality": every node has some degree of influence or importance within the social domain under consideration, and one expects such importance to surface in the structure of the social network.

Centrality is a quantitative measure that aims at revealing the importance of a node.
Among the types of centrality that have been considered in the literature, many have to do with distances between nodes.
Take, for instance, a node in an undirected connected network: if the sum of distances to all other nodes is large, the node under consideration is *peripheral*; this is the starting point to define *Bavelas's closeness centrality*, which is the reciprocal of peripherality.
The role played by *shortest paths* is justified by one of the most well-known features of complex networks, the so called *small-world phenomenon.*
A small-world network is a graph where the average distance between nodes is logarithmic in the size of the network.
The purpose of this paper is to pave the way for a formal well-grounded assessment of centrality measures, based on some simple guiding principles; we seek notions of centrality that are at the same time robust and understandable.
We shall present and compare the most popular and well known centrality measures .
The comparison will be based on a set of axioms.
We compare the measures we discuss in an information-retrieval setting.
The results suggest that simple measures based on distances, *harmonic centrality, can give better results than some of the most sophisticated indices used in the literature.*

As already mentioned the centrality is a fundamental tool in the study

of social networks: the first efforts to define formally centrality indices were put forth in the late 1940s by the Group Networks Laboratory at MIT directed by Alex Bavelas, those concluded that centrality was related to group efficiency in problem-solving.
In the following decades, various measures of centrality were employed in a multitude of contexts.
We can certainly say that the problem of singling out influential individuals in a social group is what that sociologists have been trying to capture.

Freeman acutely remarks that the central node of a star should be deemed more important than the other vertices.

In fact, *the center of a star* is at the same time :

1. the node with largest degree;
2. the node that is closest to the other nodes ;
3. the node through which most shortest paths pass;
4. the node with the largest number of incoming paths of length k, for every k;
5. the node that maximizes the dominant eigenvector of the graph matrix;
6. the node with the highest probability in the stationary distribution of the natural random walk on the graph.

Degree is probably the oldest measure of importance ever used.

The most classical notion of *closeness*, instead, was introduced by Bavelas for undirected, connected networks as the reciprocal of the sum of distances from a given node.
Centrality indices based on the count of shortest paths were formally developed independently by Anthonisse and Freeman, who introduced betweenness as a measure of the probability that a random shortest path passes through a given node or edge.

Katz's index is based instead on a weighted count of all paths coming into a node.

Another line of research studies spectral techniques to define centrality.

Jon Kleinberg defined another centrality measure called: *HITS* .
The idea is that every node of a graph is associated with two importance indices: one measures how reliable a node is, and other measures how good the node is in pointing to authoritative nodes, with the two scores mutually reinforcing each other. The result is again the dominant eigenvector of a suitable matrix.

A set N of n nodes and a set $A \subseteq N \times N$ of arcs.
An arc with the same source and target is called a loop.
The transpose of a graph is obtained by reversing all arc directions .
A *symmetric graph* is a graph such that $x \rightarrow y$ whenever $y \rightarrow x$.
A successor of x is a node y such that $x \rightarrow y$, and a predecessor of x is a node y such that $y \rightarrow x$.
The outdegree $d^{+}(x)$ of a node x is the number of its successors, and the indegree $d^{-}(x)$ is the number of its predecessors.
A *path* (of length k) is a sequence $x_0, x_1, ..., x_{k-1}$, where $x_j \rightarrow x_{j+1}, 0 \leq j < k$.
A *walk* (of length k) is a sequence $x_0, x_1, ..., x_{k-1}$, where $x_j \rightarrow x_{j+1}$ or $x_{j+1} \rightarrow x_j, 0 \leq j < k$.

A *connected component* of a graph is a maximal subset in which every pair of nodes is connected by a walk .
A graph is *connected* if there is a single connected component.
A *strongly connected componen*t is terminal if its nodes have no arc towards other components.
The distance $d(x, y)$ from x to y is the length of a shortest path from x to y, or ∞ if no such path exists.
The nodes reachable from x are the nodes y such that $d(x, y) < \infty$.
The nodes coreachable from x are the nodes y such that $d(y, x) < \infty$.
A node has trivial (co)reachable set if the latter contains only the node itself.
The notation A^{-}, where A is a non negative matrix, will be used throughout the paper to denote the matrix obtained by $l1$-normalizing the rows of A, that is, dividing each element of a row by the sum of the row (null rows are left unchanged).

The three axioms

Sometimes, the attitude was actually to provide evidence that different measures highlight different kinds of centralities and are, therefore, equally incomparably interesting.
While it is clear that the notion of centrality,in its vagueness, can be interpreted differently giving rise to many good but incompatible measures, we will provide evidence that some measures tend to reward nodes that are in

no way central.

We propose to understand (part of) the behavior of a centrality measure using a set of axioms. It is reasonable to set up some necessary axioms that an index should satisfy to behave predictably and follow our intuition.

Defining such axioms is a delicate matter.
First of all, the semantics of the axioms must be very clear. Second, the axioms must be evaluable in an exact way on the most common centrality measures. Third, they should be formulated avoiding the trap of small, finite (counter) examples, on which many centrality measures collapse. We assume from the beginning that the centrality measures under examination are invariant by isomorphism, that is, that they depend just on the structure of the graph, and not on particular labeling chosen for each node.
To meet these constraints, we propose to study the reaction of centrality measures to change of size, to (local) change of density and to arc additions.

To do so, we need to try something entirely new—evaluating exactly (i.e., in algebraic closed form) all measures of interest on all nodes of some representative classes of networks.
A good approach to reduce the amount of computation is using strongly connected vertex-transitive graphs as basic building blocks: these graphs exhibit a high degree of symmetry, which should entail a simplification of our computations.
Finally, since we want to compare density, a natural choice is to pick the densest strongly connected vertex-transitive graph, the clique, and the sparsest strongly connected, the directed cycle.

Consider a graph made by a k-clique and a p-cycle. Because of invariance by isomorphism, all nodes of the clique have equal score, and all nodes of the cycle have equal score, too; but which nodes are more important? Probably everybody would answer that if $p = k$ the elements on the clique are more important, and indeed this axiom is so trivial that is satisfied by almost any measure of which we are aware, but we are interested in assessing the sensitivity to size, and thus we state our first axiom as follows:

Axiom 0.4.1.1 (Size axiom) *Consider the graph $S_{k,p}$ made by a k-clique and a directed p-cycle. A centrality measure satisfies the* size axiom *if for every k there is a P_k such that for all $p \geq P_k$ in $S_{k,p}$ the centrality of a node of the p-cycle is strictly larger than the centrality of a node of the k-clique, and if for every p there is a K_p such that for all $k \geq K_p$ in $S_{k,p}$ the centrality* *of a node of the k-clique is strictly larger than the centrality of a node of the p-cycle.*

Connect a node x of the k-cycle with a node y of the p-cycle through a bidirectional arc, the bridge.
If $k = p$, the vertices x and y are symmetric, and thus must necessarily have the same score.
Now, we increase the density of the k-cycle as much as possible, turning it into a k-clique.

We are thus strictly increasing the local density around x, leaving all other parameters fixed, and in these circumstances we expect that the score of x increases.

Axiom 0.4.1.2 (Density axiom) *Consider the graph $D_{k,p}$ made by a k-clique and a p-cycle connected by a bidirectional bridge $x \leftrightarrow y$,where x is a node of the clique and y is a node of the cycle. A centrality measure satisfies the density axiom if for $k = p$ the centrality of x is strictly larger than the centrality of y.*

Finally, we propose an axiom that specifies strictly monotonic behavior upon the addition of an arc:

Axiom 0.4.1.3 (Score-Monotonicity Axiom) *A centrality measure satisfies the score-monotonicity axiom if for every graph and every pair of nodes x, y such that $x \nrightarrow y$,when we add $x \rightarrow y$ to G the centrality of y increases.*

In some sense, this axiom is trivial: it is satisfied by essentially all centrality measures we consider on strongly connected graphs.Thus, it is an excellent test to verify that a measure is able to handle correctly partially disconnected graphs.

Recapitulate

We are considering 11 centralities (harmonic, indegree, closeness, betweenness, Katz, Lin, dominant eigenvector, Seeley'sindex, HITS, SALSA, pagerank) and 3 axioms (size, density, score-monotonicity). We can to verify 33 statements.

For simplicity all our results are summarized in this Table where we distilled them into simple yes/no answers to the question:
Does a given centrality measure satisfy the axioms?

Centrality	Size	Densisty	Score monotonicity
Degree	only k	yes	yes
Harnomic	yes	yes	yes
Closeness	no	no	no
Lin	only k	no	no
Betweenness	only p	no	no
Dominant	only k	yes	no
Seeley	no	yes	no
Katz	only k	yes	yes
Pagerank	no	yes	yes
HITS	only k	yes	no
SALSA	no	yes	no

It was surprising that only harmonic centrality satisfies all axioms. All spectral centrality measures are sensitive to density. Row-normalized spectral centrality measures are insensitive to size, whereas the remaining ones are only sensitive to the increase of k (or p in the case of betweenness).
All non-attenuated spectral measures are also non-monotone. Both Lin's and closeness centrality fail density tests.
Closeness has, indeed, the worst possible behavior, failing to satisfy all our axioms. While this result might seem counter intuitive, it is actually a consequence of the known tendency of very far nodes to dominate the score, hiding the contribution of closer nodes, whose presence is more correlated to local density.
All centralities satisfying the density axiom have no watershed: the axiom is satisfied for all $p, k \geq 3$. The watershed for closeness (and Lin's index) is $k \leq p$, meaning that they just miss it, whereas the watershed for betweenness is a quite pathological condition ($k \leq (p^2 + p + 2)/4$): one needs a clique whose size is quadratic in the size of the cycle before the node of the clique on the bridge becomes more important than the one on the cycle (compare this with closeness, where $k = p + 1$ is sufficient).
We remark that our results on geometric indices do not change if we replace the directed cycle with a symmetric (i.e., undirected) cycle, with the additional condition that k > 3. It is possible that the same is true also of spectral centralities, but the geometry of the paths of the undirected cycle makes it extremely difficult to carry on the analogous computations in that case.

We have presented a set of axioms that try to capture part of the intended behavior of centrality measures. We have proved or disproved all our axioms for ten classical centrality.

Nonetheless, we believe we have made the important point that geometric measures are relevant not only to social networks, but also to information retrieval.

Bibliography

[1] Stephen P Borgatti and Martin G Everett. A graph-theoretic perspective on centrality. *Social networks*, 28(4):466–484, 2006.

[2] Ulrik Brandes, Stephen P Borgatti, and Linton C Freeman. Maintaining the duality of closeness and betweenness centrality. *Social Networks*, 44:153–159, 2016.

[3] Manuj Garg. Axiomatic foundations of centrality in networks. *Available at SSRN 1372441*, 2009.

[4] Mitri Kitti. Axioms for centrality scoring with principal eigenvectors. *Social choice and welfare*, 46(3):639–653, 2016.

[5] Yannick Rochat. Closeness centrality extended to unconnected graphs: The harmonic centrality index. Technical report, 2009.

[6] Gert Sabidussi. The centrality index of a graph. *Psychometrika*, 31(4):581–603, 1966.

[7] Paolo Boldi Sebastiano Vigna. Axioms for centrality. *arXiv preprint arXiv:1308.2140*, 2013.