**mathematics**

MDPI

*Article*

# Redefining Centrality Measures in Weighted Causal Graphs

**Cristina Puente [1,*], Javier Rodrigo [2] , Mª Dolores López[3], Álvaro Arrieta-Puente[4]**

1 Computer Science Dept, ICAI School of Engineering, Comillas Pontifical University, Spain; cristina.puente@comillas.edu;
2 Mathematics Dept, ICAI School of Engineering, Comillas Pontifical University, Spain; jrodrigo@comillas.edu
3 Universidad Politécnica de Madrid; marilo.lopez@upm.es
4 Industrial Engineering and Innovation Sciences, Technische Universiteit Eindhoven, Eindhoven, Holland, a.arrieta.puente@student.tue.nl
* Correspondence: cristina.puente@comillas.edu

**Abstract:** Causal graphs are powerful instruments for structuring information and analyzing the extent to which an observed effect can be attributed to a given cause. In this work, we present how edge-weighted causal graphs can be used to quantify whether a sentence acts as a direct or indirect cause of an effect. We introduce methods to identify the causal path with the highest overall strength between two concepts and, given a specific context, to determine the pair of concepts that exhibits the strongest causal connection. In addition, we propose centrality measures that incorporate causality scores and graph weights, enabling the ranking of sentences by their causal importance and highlighting those that are most relevant from a causal perspective.

**Keywords:** Weighted causal graphs, causality, causal model, causal representation, centrality measures.

**Citation:** Lastname, F.; Lastname, F.; Lastname, F. Title. *Mathematics* **2021**, *9*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date



## 1. Introduction

Causality is an important notion in every field of science. In empirical sciences, causality is a useful way to generate knowledge and provide for explanations. When a quantum physicist calculates the probability of an atom absorbing a photon, he analyses this event as the cause of the atom's jump to an excited energy level; that is, he tries to establish a cause-effect relationship [1].

Causation is a type of relationship between two entities: cause and effect. The cause provokes an effect, and the effect is a consequence of the cause. Causality can be a direct process when A causes B and B is a direct effect of A, or an indirect one when A causes C through B, and C is an indirect effect of A.

The typical form of causality is A causes B and the classic form of conditionality is If A then B. Causality and conditionality are not only restricted to these formats. Synonyms of cause or effect may indicate causality. Statements like B is due to A, A produces B, etc., are some other ways of expressing causality, as well as there are some other forms of expressing conditionality, like B if A, or A if only B. Therefore, in order to study causality, these forms need also to be taken into account.

The use of causal graphs as a way to represent information has been very present in literature, as Pearl [2], Spirtes [3], or Sobrino et al. [4] exemplifies. These representations usually have a qualitative ponderation in the edges to represent causal intensity like always, can, sometimes…. On the other hand, there are studies about causality that use a numerical degree to weight edges in a graph, which supposed and (creo que es: an) advance in the study of causal graphs. One of these studied (studies) is the one presented by López et.al. [5] to obtain the causality degree of several causal paths linking two nodes. In this paper we present as novelty the application of such weighted causal graphs to the detection of new centrality measures related to causality. These measures are focused in quantifying the idea of finding the most central vertex in a graph, for example taking into account the length of the paths derived from a node. If we have a weighted graph according to causality measures, we will be able to adapt these weights to get the central vertex in a graph and predict the strongest set of effects produced by a cause for example.

In section 2 we will explain these new definitions of centrality measures depicted by a practical example.

## 2. Weighted Centrality Measures

In this paper, we propose change the centrality measures of causal graph's vertices defined in [5] by using the weights of the incoming edges of such vertices. This would create the idea of dynamic graph, as edges ponderation would not be the same though it would be calculated in base to their causality degree, so centrality would be calculated from a causal point of view.

With these premises, we establish the following definitions:

Definition 1 (see [5]): We define a causal graph associated to a set of concepts as a weighted, directed graph $G=(V, E)$, where the vertices of $V$ represent to the concepts selected and $(a, b) \in E$ if the vertex $a$ causes directly the effect of vertex $b$ (without any intermediate vertex). The weight of an edge (a, b) , which will be named as w($a$, $b$), is a number in the interval (0, 1] indicating the degree on which the vertex $a$ causes directly the effect on vertex $b$

Definition 2 (see [5]): given a causal graph with vertices set $V=\{v_1,...,v_n\}$, its weighting matrix is $M=(a_{ij})_{n\times n}$, with:

$$a_{ij}=\begin{cases} w(v_i,v_j) \text{ if } (v_i,v_j)\in E \\ 0 \text{ if } (v_i,v_j)\notin E \end{cases}$$

Definition 3 (see [5]): given a causal graph with set of vertices: $V=\{v_1,...,v_n\}$, we define the matrix $A=(a_{ij})_{n\times n}$, where:

$$a_{ij}=\begin{cases} w(v_i,v_j)v_j \text{ if } (v_i,v_j)\in E \\ 0 \text{ if } (v_i,v_j)\notin E \end{cases}$$

Definition 4 (see [5]): given a causal graph with $n$ vertices: $G=(V=\{v_1,...,v_n\},E)$, we define recursively the matrix $A_k$, $k=1,...,n-1$ as:

$$A_1=(a_{ij})_{n\times n}, \text{ with } a_{ij}=\begin{cases} w(v_i\, v_j)v_i\, v_j \text{ if } (v_i,v_j)\in E \\ 0 \quad \text{if } (v_i,v_j)\notin E \end{cases}, \quad A_k=A_{k-1}A, \quad k=2,...,n-1.$$

The product (symbolic) of matrices is calculated without commute factors.

Definition 5: given a causal graph of $n$ vertices, the weighted output degree of a vertex is the addition of all weights that come out of the vertex and the input weighted degree of a vertex is the addition of all the weights of the edges linking the vertex.

Definition 6: given a causal graph of $n$ vertices, the weighted centrality in the output degree of a vertex is the weighted output degree of such vertex divided by n-1, and the weighted centrality in the input degree of a vertex is the weighted input degree of such vertex divided by n-1.

Remarks:

1. To calculate the weighted centrality output degree of a vertex i, we just have to calculate the addition of the elements of the row i of the weighting matrix divided by the order of such matrix -1.
2. To calculate the weighted centrality input degree of a vertex i, we just have to calculate the addition of the elements of the column i of the weighting matrix divided by the order of such matrix -1.

Definition 7: given a causal graph of *n* vertices, the weighted proximity centrality in the output degree of a vertex is the addition of the degrees of all causal paths going from that vertex to the rest, divided by *n*, where the degree of a path is the product of the weights of the edges contained in the path, and a causal path is a path in a causal graph.

Definition 8: given a causal graph of *n* vertices, the weighted proximity centrality in the input degree of a vertex v is the addition of the degrees of all causal paths going from all vertices to v, divided by *n*.

Remarks:

1. To calculate the weighted proximity centrality output degree of a vertex i, we just have to calculate the maximum of the coefficients of the element (i, j) of the matrix $A_1 + \ldots + A_{n-1}$ and add these maximum elements from j=1 till n, dividing the result by *n*.
2. To calculate the weighted proximity centrality input degree of a vertex j, we just have to calculate the maximum of the coefficients of the element (i, j) of the matrix $A_1 + \ldots + A_{n-1}$ and add these maximum elements from i=1 till n, dividing the result by n.

In all cases we are looking for the vertex or vertices with a highest centrality measure.

Example:

With these definitions, we propose a practical example based in the graph of figure 1 and calculate the weighted centrality of its vertices. We have 9 vertices, v1,…,v9, and the following edges: $(v_1, v_2)$, $(v_1, v_4)$, $(v_1, v_9)$, $(v_2, v_3)$, $(v_3, v_4)$, $(v_4, v_5)$, $(v_4, v_6)$, $(v_5, v_8)$, $(v_6, v_7)$, $(v_7, v_9)$, $(v_2, v_3)$, and the following weights on each edge:

w(v1,v2)=0.95, w(v1,v4)=0.9, w(v1,v9)=0.85, w(v2,v3)=0.95, w(v3,v4)=0.5, w(v4,v5)=0.6, w(v4,v6)=0.6, w(v5,v8)=0.6, w(v6,v7)=0.6, w(v7,v9)=0.95, w(v8,v9)=0.6.

So, the weighting matrix of this graph is:

$$M = \begin{pmatrix} 0 & \frac{19}{20} & 0 & \frac{9}{10} & 0 & 0 & 0 & 0 & \frac{17}{20} \\ 0 & 0 & \frac{19}{20} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{5} & \frac{3}{5} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{5} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{19}{20} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{3}{5} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

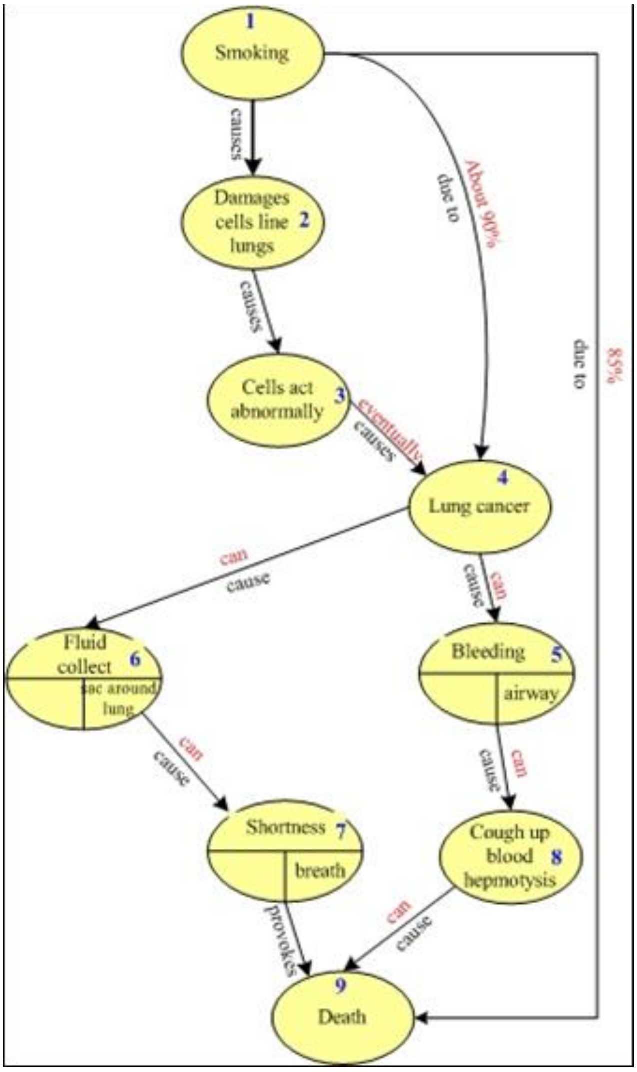


**Figure 1.** Causal graph to answer the question How smoking causes death? Automatically extracted from [4], [6].

If we apply the first observation of definition 6, we will obtain the following weighted centralities in the output degree:

v1: $\frac{1}{8}\left(\frac{19}{20}+\frac{18}{20}+\frac{17}{20}\right)=\frac{27}{80}$, v2: $\frac{1}{8}\frac{19}{20}=\frac{19}{160}$, v3: $\frac{1}{8}\frac{1}{2}=\frac{1}{16}$, v4: $\frac{1}{8}\left(\frac{3}{5}+\frac{3}{5}\right)=\frac{3}{20}$, v5: $\frac{1}{9}\frac{3}{5}=\frac{1}{15}$, v6: $\frac{1}{8}\frac{3}{5}=\frac{3}{40}$, v7: $\frac{1}{8}\frac{19}{20}=\frac{19}{160}$, v8: $\frac{1}{8}\frac{3}{5}=\frac{3}{40}$, v9: 0.

So v1 is the vertex with highest weighted centrality degree in the output degree.

Applying remark 2 of definition 6, we will obtain the following weighted centralities in the input degree:

v1: 0, v2: $\frac{1}{8}\frac{19}{20}=\frac{19}{160}$, v3: $\frac{1}{8}\frac{19}{20}=\frac{19}{160}$, v4: $\frac{1}{8}\left(\frac{9}{10}+\frac{1}{2}\right)=\frac{7}{40}$, v5: $\frac{1}{8}\frac{3}{5}=\frac{3}{40}$, v6: $\frac{1}{8}\frac{3}{5}=\frac{3}{40}$, v7: $\frac{1}{8}\frac{3}{5}=\frac{3}{40}$, v8: $\frac{1}{8}\frac{3}{5}=\frac{3}{40}$, v9: $\frac{1}{8}\left(\frac{17}{20}+\frac{19}{20}+\frac{3}{5}\right)=\frac{3}{10}$.

So v9 is the vertex with highest weighted centrality degree in the input degree.

The matrix $A_1+\ldots+A_{n-1}$ of this graph is:

{{0.,0.95v1v2,0.9025v1v2v3,0.9v1v4+0.45125v1v2v3v4,0.54v1v4v5+0.27075v1v2v3v4v5,0.54v1v4v6+0.27075v1v2v3v4v6,0.324v1v4v6v7+0.16245v1v2v3v4v6v7,0.324v1v4v5v8+0.16245v1v2v3v4v5v8,0.85v1v9+0.3078v1v4v6v7v9+0.1543275v1v2v3v4v6v7v9+0.1944v1v4v5v8v9+0.09747v1v2v3v4v5v8v9},
{0.,0.,0.95v2v3,0.475v2v3v4,0.285v2v3v4v5,0.285v2v3v4v6,0.171v2v3v4v6v7,0.171v2v3v4v5v8,0.16245v2v3v4v6v7v9+0.1026v2v3v4v5v8v9},
{0.,0.,0.,0.5v3v4,0.3v3v4v5,0.3v3v4v6,0.18v3v4v6v7,0.18v3v4v5v8,0.171v3v4v6v7v9+0.108v3v4v5v8v9},
{0.,0.,0.,0.,0.6v4v5,0.6v4v6,0.36v4v6v7,0.36v4v5v8,0.342v4v6v7v9+0.216v4v5v8v9},
{0.,0.,0.,0.,0.,0.,0.,0.6v5v8,0.36v5v8v9},
{0.,0.,0.,0.,0.,0.,0.6v6v7,0.,0.57v6v7v9},
{0.,0.,0.,0.,0.,0.,0.,0.,0.95v7v9},
{0.,0.,0.,0.,0.,0.,0.,0.,0.6v8v9},
{0.,0.,0.,0.,0.,0.,0.,0.,0.}}

Applying observation 1 of definitions 7 and 8, we will obtain the following proximity weighted centralities in the output degree:

v_1: 1/9 (0.95+0.9025+0.9+0.54+0.54+0.324+0.324+0.85)=0.5923,
v2: 1/9 (0.95+0.475+2×0.285+2×0.171)=0.2597,
v3: 1/9 (0.5+2×0.3+2×0.18)=0.1622,
v4: 1/9 (2×0.6+2×0.36)=0.2133, v5 and
v6: 0.6/9=0.0667, v7, v8 and v9: 0.

So v1 is the vertex with a highest proximity weighted centrality in the output degree.

Applying observation 2 of definitions 7 and 8, we will obtain the following proximity weighted centralities in the input degree:

v1: 0, v2: 0.95/9=0.1056, v3: 1/9 (0.9025+0.95)=0.2058, v4: 1/9 (0.9+0.475+0.5)=0.2083,
v5: 1/9 (0.54+0.285+0.3+0.6)=0.1917,
v6: 1/9 (0.324+0.171+0.18+0.36+0.6)=0.1817,
v7: 1/9 (0.324+0.171+0.18+0.36+0.6)=0.1817,
v8: 1/9 (0.324+0.171+0.18+0.36+0.6)=0.1817,
v9: 1/9 (0.85+0.16245+0.171+0.342+0.36+0.57+0.95+0.6)=0.4451

So v9 is the vertex with a highest proximity weighted centrality in the input degree.

## 5. Conclusions

The problem in causal graphs of selecting the most causal path linking two nodes has been largely discussed and is not trivial. In this paper we have proposed a new approach by applying centrality weights to select those nodes with highest degrees and in accordance, obtain the 'best' causal path between two nodes. This approach takes into account the relationship of a node with the ones surrounding him and the input and output edges, providing better results than the ones that we used before (see [4]). In addition, for future works, it will serve us for three main goals:

The first one is to select the most important nodes according to its causal weight when creating a summary. The second, would be when asking a question, select the causal path that links two nodes with the highest degree of causality to include those nodes in the answer of the question. The third use would be to remove redundant nodes in a causal graph. For instance, in the graph included in [6] we had "tobacco use" and "smoking". With this measurement we are able to select the node with a highest weight to work with.

**Author Contributions:** hay que cumplimentar esto

**Acknowledgments:**

**Conflicts of Interest:** "The authors declare no conflict of interest."

## References


1. Bunge M., Causality: the place of the causal principle in modern science. Cambridge: Harvard University Press. (1959)
2. Pearl J., Causality, models, reasoning, and inference. Cambridge University Press. (2000).
3. Spirtes P., Glymour C., Scheines R., Causation, Prediction and Search. Second edition, MIT Press, Cambridge, Massachusetts. (2000).
4. Sobrino, A., Puente, C., & Olivas, J. A. (2014). Extracting answers from causal mechanisms in a medical document. Neurocomputing, 135, 53-60.
5. Mª Dolores López González, Cristina Puente Águeda, Javier Rodrigo Hitos, José Ángel Olivas Varela. Weighted graphs to model causality. ICAI 2017 (International Conference on Artificial Intelligence), ISBN: 1-60132-460-X, pp. 297-301. Las Vegas, 17-20 de julio de 2017. Actas editadas por CSREA Press
6. Puente, C., Sobrino, A., Olivas, J. A., & Merlo, R., Extraction, analysis and representation of imperfect conditional and causal sentences by means of a semi-automatic process. In Fuzzy Systems (FUZZ), 2010 IEEE International Conference, pp. 1-8. Barcelona, Spain (2010).