# Quantifying External Information in Social Network Analysis: an Application to Comparative Mythology

Javier Fumanal-Idocin, *Member, IEEE*, Oscar Cordón *Fellow, IEEE*, Graçaliz Dimuro, Antonio-Francisco Roldán López-de-Hierro, Humberto Bustince, *Fellow, IEEE*

***Abstract*—Social network analysis is a popular tool to understand the relationships between interacting agents by studying the structural properties of their connections. However, this kind of analysis can miss some of the domain-specific knowledge available in the original information domain and its propagation through this existing network. In this work, we develop an extension of classical social network analysis to incorporate external information from the original source of the network. With this extension we propose a new centrality measure, the semantic value, and a new affinity function, the semantic affinity, that establishes fuzzy-like relationships between the different actors in the network. We also propose a new heuristic algorithm based on the shortest capacity problem to compute this new function. As an illustrative case study, we use the novel proposals to analyze and compare the gods and heroes from three different classical mythologies: Greek, Celtic and Nordic. We study the relationships of each individual mythology and those of the common structure that is formed when we fuse the three of them. We also compare our results with those obtained using other existing centrality measures and embedding approaches. In addition, we test the proposed measures on a classical social network, the Reuters terror news network. We found that the novel method obtains more meaningful comparisons and results than previous existing approaches.**

***Index Terms*—Social Network Analysis; Centrality measures; Semantic value; Word embeddings; Mythology; Comparative mythology.**

## I. Introduction

Network science has become an important tool to study systems composed of interacting agents, such as proteins or human societies [1], [2], [3], [4]. One of the key ideas in social sciences is that living beings are embedded by our own social nature in a complex web of social relations and interactions. This social fabric that we form has been traditionally modelled as a network, where each person is represented as a node that is connected to others according to some criteria. Some of the most popular tools to perform social network analysis are the centrality measures, which ponder the importance of each node in the network according to its structural properties [3], [5]. Social network analysis stands as an appropriate tool to understand important characteristics of the human behaviour, as it seems that many of us are deeply affected by the social structure in which we take part [6].

Networks are a popular way to model and represent knowledge, but they cannot be directly used in many popular machine learning algorithms [7]. This is a common issue in other domains such as natural language processing (NLP) which is usually tackled using different embedding methods. The aim of those techniques is to construct a continuous representation space for the data taking into account their possible interactions in the original domain. The most popular kind of these representations are word-embeddings, in which words are encoded in vectors according to their co-occurrences with other words [8]. This process is supposed to encode the meaning of each word because of the frequency in which each word appear near others. One of the most popular word embedding approaches is the family of word2vec architectures [9]. They consist of a neural network that learns the context of each word given a text corpus and have been successfully applied in text classification [10], entity recognition [11], and sentiment analysis [12]. In a similar fashion, nodes in a network can be embedded in a vectorial representation using word2vec approaches [13], [14], [15].

Word embeddings have been a very successful line of research because they make possible to use standard machine learning techniques on non-structured data [7]. They are also suitable to capture latent knowledge in the original material [16], which is not always possible in the structural analysis performed in social network analysis and graph theory. However, word embedding approaches still present some limitations [17]. They require large datasets to learn the proper context for each word. Besides, some word meanings depend heavily on the context, which might not be reflected in the original text corpus. Comparison between different word embeddings is usually performed using the cosine similarity, which is not necessarily explainable and can produce unreasonable results in the original domain [18].

Because of the limitations of these methods, there is not a standard way to fuse the information obtained from one phenomenon with a network layout when available. Besides, when comparing the elements of such phenomenon, the results of these comparisons lose the original conveniences of a graph structure when using the cosine distance between embeddings. Hence, our aim in this paper is double:

- To develop a new method to relate the structural properties and domain-specific properties from the data that originated the network.

Javier Fumanal-Idocin, Graçaliz Pereira Dimuro and Humberto Bustince are with the Departamento de Estadistica, Informatica y Matematicas, Universidad Publica de Navarra, Campus de Arrosadia, 31006, Pamplona, Spain. emails: javier.fumanal@unavarra.es, gracalizdimuro@furg.br, bustince@unavarra.es

Oscar Cordón is with the Dept. of Computer Science and Artificial Intelligence, Andalusian Research Institute “Data Science and Computational Intelligence” (DaSCI), University of Granada, 18071 Granada, Spain. email: ocordon@decsai.ugr.es

Antonio Francisco Roldán López de Hierro is with the Department of Mathematics, University of Jaén, Campus las Lagunillas s/n, 23071, Jaén, Spain, email: afroldan@ujaen.es

- To establish a method to compare two actors that can be interpreted using both the network architecture and the specific knowledge from the studied phenomena.

We achieve these goals by defining three novel concepts: the intrinsic, extrinsic and semantic values of an actor in a social network. Then, we show how these values can be used in the context of social network analysis as metrics to weight the importance of each concept. Subsequently, we present a new heuristic algorithm to perform interpretable comparisons between one or more actors using their semantic values. Finally, we use these tools to study three different compilations of traditional stories and myths (from the Greek, Nordic, and Celtic traditions) as a case study. We characterize the gods and heroes in these tales according to the importance of their semantic value and their semantically closest terms. We show how gods from different cultures relate to each other, and which kind of entities does each culture give more importance to. We have also explored the use of our proposed tools in the popular Reuters terror news network.

The rest of the manuscript is organized as follows. Firstly, in Section II we review some concepts regarding social network analysis, affinity functions, and embedding methods. In Section III we present our formalization for the semantic value of an actor and how to use it as a centrality measure. Then, in Section IV we present the semantic affinity concept and we detail the algorithm proposed to compute it. Subsequently, in Section V we show the results obtained when analyzing different mythology networks using the semantic value and the semantic affinity, other centrality measures, and word2vec. In Section VI we study the application of the proposed concepts to the Reuters terror attack word association network, and we compare how the semantic affinity performs with respect to other affinity functions in a community detection problem. Finally, in Section VII we discuss the obtained results, we present some conclusions for this work, and we establish the guidelines for future research.

## II. Background

In this section we discuss some basic notions about centrality measures and the novel concept of affinity functions in social network analysis. We also discuss some of the existing node and word embedding techniques.

### A. Centrality measures in social network analysis

In graph theory and network science, centrality measures indicate how relevant each node is in a structure [19], [20], [5]. Some very well known centrality measures are:

- Degree centrality: the number of edges incident upon a node. In the case of directed networks, the degree is the sum of the number of edges incident to the node (in-degree) and the number of edges salient to the node (out-degree).
- Betweenness centrality: the betweenness of a node is the number of times that such node is in the shortest path of other two nodes. It measures the brokerage ability of the node in the network's information flow.
- Closeness centrality: the closeness centrality of a node is the reciprocal of the average length of the shortest path between that node and the rest of the nodes in the network. It measures the overall location of the node in the network, establishing a center-periphery difference.
- Eigenvector centrality: it assigns a relative score to each node in the network based on the idea that connections to well connected nodes should ponder more than connections to poorly connected ones. The famous PageRank algorithm uses a variant of this centrality measure for directed networks [21].

### B. Affinity functions

Affinity functions were defined in [22] as a way to measure the relationship between a pair of actors in a social network by capturing the nature of their local interactions. "Affinities" are defined as functions over the set of actors of a given social network assigning a number between 0 and 1 to every pair of actors $x, y$ that is, if we denote by $\mathcal{A}$ the set of actors, then:

$$F_C : \mathcal{A} \times \mathcal{A} \to [0, 1] \tag{1}$$

Usually, in order to get this mapping, the adjacency matrix $C$ is used. Recall that each entry $C_{x,y}$ of the adjacency matrix $C$ quantifies the strength of the relationship for the pair of actors $x, y$ in a weighted network, composed of by a finite set of actors. The affinity between two actors shows how strongly they are connected according to different criteria, depending on which aspect of the relationship we are taking into account.

In the following, we recall two definitions of affinity functions that we will use in our subsequent developments:

- Best friend affinity: it measures the importance of a relationship with an agent $y$ for the agent $x$, in relation to all the other relationships of $x$:

$$F_C^{BF}(x, y) = \frac{C_{x,y}}{\sum_{a \in \mathcal{A}, x \neq a} C_{x,a}}. \tag{2}$$

- Machiavelli affinity: it computes how affine two actors $x$ and $y$ are based on how similar is the social structure that surrounds them:

$$F_C^{Mach}(x, y) = 1 - \frac{\mid I_x - I_y \mid}{\max\{I_x, I_y\}}, \tag{3}$$

where $I_a = \sum_{z \in Z(a)} D(z)$, where $Z(a)$ is the set of actors where $C_{a,z} > 0, \forall z \in Z$, and $D(z)$ is the degree centrality of $z$.

A 0 affinity value means that no affinity has been found at all while an 1 value means that there is a perfect match according to the analyzed factors. Since affinities are not necessarily symmetrical, the strength of this interaction depends on who the sender and receiver are, as it happens in human interactions e.g. unrequited love.

### C. Word and node representation in continuous spaces

Word representation using continuous spaces has been a popular research topic for the past decades [23]. One of the main advantages of this representation of words is that these codifications can model the context of a word, and they are more suitable to perform different operations on them, like analogy searching.

Neural networks have been the most successful tools for this task. In [24] the authors proposed a feedforward neural network with a linear projection layer and a non-linear hidden layer in order to learn each word representation. Many others followed this work [25], [26]. Recurrent Neural networks have also been very popular for this task [27], [28].

Mikolov et al. proposed in [9] to use a deep learning architecture that learns each word context using skip-grams, which obtained state-of-the-art results for different natural language processing tasks. This research was very popular and further iterations of this idea have been successfully applied to language translation [29], text classification [12], and medical records [30].

Network embedding is also a popular research topic [31]. A similar idea to word2vec has been applied to learn features for each node in a network [13]. This proposal, commonly called node2vec, works by generating a series of random walks of a predefined length from each node in the network. Then, these nodes are fed to a word2vec architecture that treats each path as if it were a sentence. There are other types of graph embeddings, using spectral clustering [32] or deep learning [14], [33]. Some methods are also capable of embeddings attributed networks [15]. These approaches usually consist of computing similarities between the attributions and exploiting their homophily.

## III. Semantic value in social network analysis

In this section we introduce the proposed formalization of the semantic value of an actor in a network. The objective of this formalization is to quantify information from the original domain of the network, also considering its propagation through the network connections. In this way, we characterize each actor according to the information it possessed in the original network and the information that it received through the propagation of the other actors in the network.

### A. A formal definition of the semantic value

We define Ж, the semantic value of an actor in a social network, as the union of its intrinsic and extrinsic values:

$$Ж(x) = \cup(⅄(x), ϒ(x)). \tag{4}$$

First we define $⅄(x)$ the intrinsic value of an actor $(x)$ in a social network, as the property that is unique and inherent to him, not necessarily deducible from the network structure or topology. If the actor $x$ were removed from the network, then all the present $⅄(x)$ in the network would also be removed from all the semantic values in the network. For example, if we are working with a network of routers, the real-world condition of each router location is an important part of each router ⅄.

Then, we define the extrinsic value of $x$, denoted as $ϒ(x)$, as the property that represents the information in the local interactions of $x$, considering the relationships as a way of information or resource transmission, as usual in social network analysis [3]. Being $X = X_1, X_2, ..., X_a$ the vector of nodes connected to $x$, of dimension $a$, and $F_C$ an affinity function, the extrinsic value, $ϒ(x)$, is the set defined as:

$$\begin{split} ϒ(x) = \bigcup_{i=1}^{a} \{F_C(X_i, x)Ж(X_i) - \\ \cup_{j \in J_i}\{\cap(F_C(X_i, x)Ж(X_i), F_C(X_j, x)Ж(X_j))\} \end{split} \tag{5}$$

where $J_i = \{j \in \{1, \dots, a\}, i \neq j\}$. With this expression, we establish that the extrinsic value is the union of the received semantic values from the rest of the actors. Each actor in $X$ sends its own semantic value to $x$, modulated in each case by the affinity for that relationship, and erasing the redundancies from other relationships. Since the semantic value has a recursive definition, the result of Eq. (4) is a set of intrinsic values.

For the sake of simplicity, we can shorten Eq. (5) by using $V_x(b) = F_C(X_b, x)Ж(X_b)$:

$$ϒ(x) = \bigcup_{i=1}^{a} \{V_x(i) - \cup_{j \in J}\{\cap(V_x(i), V_x(j))\} \tag{6}$$

The present recursivity in the definition of the semantic value is unavoidable for this definition and it is similar to those present in other centrality measures like PageRank [21]. In order to compute the expression in Eq. 6 we set the initial values of $Ж(X_i)$ and $Ж(X_j)$ as $⅄(X_i)$ and $⅄(X_j)$ respectively.

Let's take as a toy example a network where different actors are discussing their opinion about a topic. Each actor's own opinion is their intrinsic value, and their connections are modeled using an affinity function. In the beginning, each actor's own semantic value is equal to its intrinsic semantic value. When an actor $x$ talks to another actor $y$, he expresses his opinion and tries to persuade the receptor about it. The receptor listens to that actor, and agrees with a fraction of the sender, quantified using the affinity function $F(y, x)$. So, after this interaction the semantic value of $y$ is $⅄(y)F_C(y, x)⅄(x)$, which represents the current opinion of $y$ about that topic. Then, when the number of interactions grows, the extrinsic values of of all the actors will become more significant.

### B. Computing the semantic value

In order to give a computable version of the semantic value for a node $x$, it is first required to give a computable version of the intrinsic and extrinsic values, $⅄(x)$ and $ϒ(x)$. We denote the computable version of ⅄ as $I$, the computable version of ϒ as $E$, and the computable version of Ж as $S$.

Due to inherent fuzziness of the concept of intrinsic value, we cannot give an exact mathematical formula to compute it. Depending on the context and the application, we can use a function to transform this abstract idea into a number. In order to choose a proper $I$ function, we must take into account that depending on the application, the criterion to obtain this function can vary significantly. As a general rule, $I$ should grow according to its relevance in its original domain. Following the router network example, if we consider the fitness of each of these real-world conditions for the signal transmission task, we could use this fitness as $I$. If we are studying a financial trading network, the economic value of each agent financial assets is also an appropriate $I$.

The computation of the extrinsic value requires choosing an affinity function to quantify the relationships. Since in this case we have numbers instead of sets we use the summation instead of the union. The intersection of the received semantic value from $X_i$ and $X_j$ to $x$ is approximated as the result of propagating $I(X_i)$ through $X_j$ to $x$. So, the expression that approximates the extrinsic value is:

$$E(x) = \sum_{i=1}^{a} \max \Bigg( F_C(X_i, x) I(X_i) - \sum_{j \in J} F_C(X_i, X_j) I(X_i) F_C(X_j, x), 0 \Bigg) \tag{7}$$

where $J = \{j \in \{1, \ldots, a\}, i \neq j\}$.

Finally, we can compute $S(x)$ with the analogous formula to Eq. (4):

$$S(x) = I(x) + E(x). \tag{8}$$

In order to compute the semantic value for an actor we need to access the affinity values for that node and its neighbours. So, the computational cost to compute the semantic value for each actor in a network scales quadratically with the number of edges in that network.

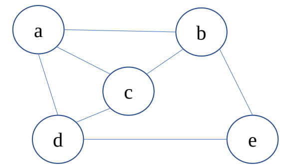
### C. Semantic value as a centrality measure

Once we have computed $S$ for each node in a network, we can analyze the results just as with any other centrality measure. In order to have a high value of $S$, an actor $x$ must have a high value of $I$ and $E$ in comparison with the rest of the actors in the network.

To have a high $I$ value, $x$ must be important in the original domain from which we constructed the network. The function $I(x)$ does not tell much about the network structure directly but it can reinforce the importance of $x$ if $x$ has also high values in other centrality measures. If $x$ does not have high values in the remaining centrality measures but it has a high $I(x)$, this is revealing that $x$ was important in the original domain in a way that has not been taken into account when building the network.

To own a high $E$ value, the actor $x$ must have connections with other nodes with high $I$ that are not connected among them. Low values of $E$ indicate that the actor relations are not very important in the network dynamics, it is part of a small community or its connections are few or redundant.

For instance, in the specific case of a word association network in a text, a domain-specific term is not likely to have very high semantic value. As it points out to a very narrow concept, it will not have a high $I$ value. Besides, since those domain-specific terms relate mostly to other specific terms of the same domain, they will not have many connections with other actors, so the possible extrinsic value is very limited.




| Actor | I | S | Eigencentrality | Degree centrality |
|---|---|---|---|---|
| a | 1 | 2.16 | **1.00** | **3** |
| b | 1 | **2.91** | 0.92 | **3** |
| c | 1 | 2.16 | **1.0** | **3** |
| d | 1 | **2.91** | 0.92 | **3** |
| e | 1.5 | 2.58 | 0.65 | 2 |

Fig. 1: Example network where we computed the semantic value, the eigencentrality and the degree centrality.

On the contrary, high semantic values indicate a very general concept.

Consider the following example, shown in figure 1. This Figure shows a simple social network structure in which different actors share their opinion about a topic. Based on their knowledge of this topic, we can ponder and quantize the value of each actor's opinion, which we will set as the intrinsic value of each actor. In this network structure we have computed two classical centrality measures, the node degree centrality and the eigencentrality, as well as the semantic value. Based on the degree centrality, all nodes are pondered as equally important except $e$, which has less connections. Eigencentrality ponders the importance of each node according to the importance of its connected nodes, which favours $a$ and $c$. Finally, the semantic value can also take into account the value of each actorś opinion, which favours $b$ and $d$ because both of them have good connections and are connected to the most valuable opinion in the network.

## IV. Semantic affinity

The semantic affinity of two actors $x$ and $y$ measures the affinity between them based on the idea of how notably we need to change $S(x)$ to convert it to into $S(y)$. Terms that are similar in meaning should have high values of semantic affinity and non-related terms should have a very low semantic affinity. For example, the semantic affinity between "water" and "ice" should be high because these are very close terms and in real life we only need to freeze water below 0ºC to obtain ice. However, the semantic affinity between "water" and "earth" should be lower, as the difference in real life between those concepts is higher.

The advantage of the semantic affinity is that it is capable of using external information to the network, while existing affinity functions only use the structural information one. For the case of "water" and "ice", the semantic affinity takes into account the nature of both substances, while the rest of the affinities could only deduce this information if it was imprinted in the network structure somehow.

We can compute the semantic affinity based on how efficient it is to propagate $S(x)$ into $S(y)$, using the Pipe algorithm, detailed in the next section.

### A. Pipe algorithm to compute the semantic affinity between two actors in a network

The Pipe algorithm computes the semantic affinity based on the idea of modelling $S(x)$ as a liquid we need to carry from $x$ to $y$. Each actor $x$ has a capacity equal to its own semantic value $S(x)$ and each edge $x \rightarrow y$ can carry up to $F_C(x, y) \cdot S(y)$ of that liquid. So, each edge is treated as a "pipe" where the liquid goes and each actor as a bifurcation in the path. Then, we need to carry all the liquid from the source actor to the destination actor using the best possible path. To compute the final semantic affinity value, we will take into account three different aspects of this transportation process: how "good" were the taken paths, the difference in magnitude between $S(x)$ and $S(y)$ and the average affinity values of the emisor.

There are different possibilities to define "best path" in this setting. In this work we have denoted the best path as the one with the highest average affinity value, as the other alternatives could result in counter-intuitive results in some cases.

In order to compute the best path we need to take the capacity of each actor and the affinity in each edge into account. This problem is quite similar to the Shortest Capacitated Path Problem [34], which consists of finding a set of edge-disjoint paths that connects all the nodes in a graph, but in our case we are only taking one path into account, from $x$ to $y$. The classical "shortest path" optimization problem between a pair of nodes [35] is also very related to the desired task. We can reformulate the problem by rescaling the affinity values so that the shortest path is actually the one with the highest average affinity value.

In this way, we obtain the shortest possible path using all the nodes and edges that are not yet "full". Usually, it is required to use more than one path to carry all the semantic value from the source to the destination. So, we have to compute a new shortest path with the available nodes and edges every time the current path has already met its capacity limit. Considering $P(x, y)$ as the list of affinity values in the edges used in the paths to carry the semantic value in the Pipe algorithm, and $|P(x, y)|$ as the cardinal of such set, $B(x, y)$ is the average affinity value of that path:

$$B(x, y) = \frac{\sum P(x, y)}{|P(x, y)|} \tag{9}$$

We also take into account the numerical difference in the semantic value of $x$ and $y$. For example, if $S(x) = 20$ and $S(y) = 100$, and we are computing the semantic affinity between $x$ and $y$ then no matter what we do, the $80\%$ of the $S(y)$ can not be "filled". On the contrary, if we are computing the semantic affinity between $y$ and $x$, then at least the $80\%$ of the $S(y)$ will not reach $x$. To take this difference in magnitude between $x$ and $y$ semantic values into account, $M(x, y)$ we use the following expression:

$$M(x, y) = \left(1 - \frac{|S(x) - S(y)|}{\max(S(x), S(y))}\right) \tag{10}$$

Since some actors naturally have low affinity values, for example when they have a lot of connections, the expected value of the average affinity values of a path can be deceptively low. In order to better compare the different semantic affinities that originate from actor $x$, we rescale the result by the maximum semantic affinity that $x$ emits. Considering $N(x)$ as the set of neighbours of $x$, we take this into account using this expression:

$$R(x) = \frac{1}{\max_{n \in N(x)} F_C(x, n)} \tag{11}$$

Finally, the expression of the semantic affinity is the product of the three aspects that we took into account:

$$A(x, y) = B(x, y) M(x, y) R(x) \tag{12}$$

In order to compute the semantic affinity in our experimentation, we have used a combination of best friend and Machiavelli affinities as the $F_C(x, y)$. Using this mix of affinity functions we can characterize each edge based on the importance of the pairwise relationship between $x$ and $y$, and also take into account the homophily between them. We do this for two reasons:

1) In high degree actors, the best friend affinity values are necessarily low, which will result in artificially low semantic values.
2) In the texts we are studying, there are many characters with similar roles, although not directly connected. Using the Machiavelli affinity we take into account this similarity when computing the semantic affinity.

We have combined both affinity functions using a convex combination, so the value of each edge is $90\%$ the best friend affinity value and $10\%$ the Machiavelli affinity. This mixing parameters are fixed to these values because the Machiavelli affinity tend to be much higher that the best friend affinity. In this way, the contribution of personal relationships and homophily to each edge is approximately similar.

The complexity of the semantic affinity is determined by the step of finding the shortest path, which can be solved in $O(|\mathcal{E}|+|\mathcal{A}|\log|\mathcal{A}|)$ time [35], where $\mathcal{E}$ is the number of edges in the network.

## V. Comparative Mythology Analysis Using the Semantic Value and the Semantic Affinity

One social phenomena that has been widely studied due to its importance in human history is religion. Religious practices are as old as society. They are one of the pillars in which society and modern culture place their roots. Each civilization has had its own share of heroes, myths and gods that have helped shape the spirit and mind of the youth and the elder alike [36]. Although nowadays some of the most obscure gods and traditions are only known to scholars, very popular heroes and gods are still alive in the collective consciousness of modern societies and they still echo in the characters present in many popular films and media [37]. There is a long tradition in the study of the ancient myths and deities from a scientific point of view [38], [39], [40]. The interest in these stories has not faded over time and is still a very researched topic in the humanities community [41], [42], [43]. The comparison and syncretism among the different gods that have populated the ancient mythologies has also been profoundly studied [44],

TABLE I: **Centrality measures in *Greek Myths* network**. For the top 10 most repeated entities in the associated texts, sorted according to their $S$ value.

| | $S$ | $E$ | Freq. ($I$) | Degree | Betweenness | Closeness | Eigencentrality |
|---|---|---|---|---|---|---|---|
| Heracles | 239.49 | 89.49 | 150 | 364 | 0.48 | 0.52 | 0.53 |
| Theseus | 84.69 | 29.69 | 55 | 118 | 0.10 | 0.44 | 0.21 |
| King | 84.15 | 27.15 | 57 | 55 | 0.03 | 0.42 | 0.13 |
| Jason | 76.93 | 27.93 | 49 | 88 | 0.06 | 0.37 | 0.13 |
| Apollo | 64.83 | 20.83 | 44 | 86 | 0.08 | 0.40 | 0.13 |
| Psyche | 64.72 | 15.72 | 49 | 91 | 0.07 | 0.39 | 0.13 |
| Eurystheus | 57.54 | 25.54 | 32 | 77 | 0.01 | 0.39 | 0.16 |
| Zeus | 57.27 | 21.27 | 36 | 75 | 0.09 | 0.43 | 0.14 |
| Perseus | 45.30 | 16.30 | 29 | 54 | 0.04 | 0.36 | 0.08 |
| Pelias | 43.76 | 21.76 | 22 | 48 | 0.02 | 0.37 | 0.09 |

TABLE II: **Centrality measures in *Celtic Wonder-Tales* network**. For the top 10 most repeated entities in the associated texts, sorted according to their $S$ value.

| | $S$ | $E$ | Freq. ($I$) | Degree | Betweenness | Closeness | Eigencentrality |
|---|---|---|---|---|---|---|---|
| Lugh | 130.73 | 55.73 | 75 | 258 | 0.13 | 0.51 | 0.27 |
| Ireland | 107.93 | 56.93 | 51 | 256 | 0.24 | 0.56 | 0.32 |
| Conary | 101.44 | 48.44 | 53 | 198 | 0.11 | 0.49 | 0.13 |
| King | 79.73 | 40.73 | 39 | 133 | 0.07 | 0.50 | 0.17 |
| Son | 78.22 | 41.22 | 37 | 176 | 0.06 | 0.49 | 0.18 |
| Balor | 73.10 | 38.10 | 35 | 186 | 0.08 | 0.49 | 0.20 |
| Gobhaun | 71.95 | 23.95 | 48 | 147 | 0.04 | 0.46 | 0.17 |
| Ethaun | 63.92 | 27.92 | 36 | 100 | 0.04 | 0.44 | 0.09 |
| Fomor | 56.07 | 24.07 | 32 | 100 | 0.03 | 0.47 | 0.17 |
| Turann | 54.17 | 22.17 | 32 | 105 | 0.03 | 0.44 | 0.11 |

TABLE III: **Centrality measures in *The Younger Edda* network.** For the top 10 most repeated entities in the associated texts, sorted according to their $S$ value.

| | $S$ | $E$ | Freq. ($I$) | Degree | Betweenness | Closeness | Eigencentrality |
|---|---|---|---|---|---|---|---|
| Odin | 215.00 | 106.00 | 109 | 1113 | 0.47 | 0.75 | 0.37 |
| Thor | 176.63 | 44.63 | 132 | 508 | 0.14 | 0.61 | 0.26 |
| Loki | 100.58 | 34.58 | 66 | 291 | 0.06 | 0.56 | 0.22 |
| King | 53.63 | 13.63 | 40 | 79 | 0.02 | 0.49 | 0.07 |
| Frey | 51.01 | 20.01 | 31 | 167 | 0.02 | 0.53 | 0.17 |
| Har | 50.24 | 16.24 | 34 | 116 | 0.03 | 0.52 | 0.09 |
| Sigurd | 45.02 | 19.02 | 26 | 178 | 0.02 | 0.52 | 0.13 |
| Balder | 43.69 | 14.69 | 29 | 151 | 0.02 | 0.52 | 0.14 |
| Freyja | 28.78 | 10.78 | 18 | 154 | 0.01 | 0.51 | 0.16 |
| Norse | 24.06 | 4.06 | 20 | 37 | 0.00 | 0.46 | 0.04 |

TABLE IV: **Centrality measures in the fusion network.** For the 10 most repeated entities in every text analyzed, sorted according to their $S$ value.

| | $S$ | $E$ | Freq. ($I$) | Degree | Betweenness | Closeness | Eigencentrality |
|---|---|---|---|---|---|---|---|
| Heracles | 244.33 | 94.33 | 150 | 368 | 0.15 | 0.43 | 0.07 |
| Odin | 227.31 | 118.31 | 109 | 1121 | 0.17 | 0.48 | 0.33 |
| King | 222.12 | 86.12 | 136 | 269 | 0.09 | 0.48 | 0.13 |
| Thor | 178.63 | 46.63 | 132 | 502 | 0.05 | 0.44 | 0.24 |
| Lugh | 131.88 | 56.88 | 75 | 259 | 0.04 | 0.41 | 0.08 |
| Son | 116.97 | 77.97 | 39 | 414 | 0.11 | 0.49 | 0.22 |
| Ireland | 116.43 | 65.43 | 51 | 261 | 0.08 | 0.43 | 0.10 |
| Conary | 104.54 | 51.54 | 53 | 196 | 0.03 | 0.40 | 0.05 |
| Loki | 103.06 | 37.06 | 66 | 293 | 0.02 | 0.43 | 0.20 |
| Theseus | 80.19 | 25.19 | 55 | 116 | 0.02 | 0.40 | 0.04 |

[45]. Besides the very-well known equivalence that bounds the Roman and Greek gods, there are many other studies that show evidence of similarities between gods and myths from distant cultures [38], [46], [47]. These relationships can be studied from a computational point of view considering, for example, agent-based social simulation [48] and social network analysis [49], [50], [51], [52], [53], [54].

In this section we discuss the different steps required to perform our mythology analysis:

1) The results for the centrality measures and the semantic value in each network.
2) The affinity values for the semantic, best friend, and Machiavelli affinities for important characters in their respective mythologies.
3) A comparison with the results obtained using word embedding methods.

The construction process for each network is shown in Appendix A.

### *A. Analysis of the semantic value and centrality measures in the myth networks*

We computed the word association networks for each of the studied mythologies and then we fused the networks into a single one. The four resulting networks are shown in Figure 2. We computed the intrinsic, extrinsic and semantic value for each actor. We also computed other common centrality measures in social network analysis [5] for comparison purposes.

In order to compute the semantic value as a numerical value we approximate the intrinsic value as the frequency of the word in the original collection of texts. We did so because it is logical to think that the authors of those tales wrote more about things they considered important.

Table I shows the result for the most important actors according to semantic value in the *Greek Myths* network. We found that "Heracles" is the most important actor in the network, according to all the measures taken. There are other important heroes in this list like "Theseus", "Jason", and "Perseus". All of them are somewhat the embodiment of bravery and authority, so it is not surprising that "King" has also a high semantic value. There are more human characters than gods: "Apollo" and "Zeus" are the only ones which appear at the top, with similar $S$ values, but not as high as the other Greek heroes here present. Regarding the classical centrality measures studied, the betweenness gives the highest value to "Heracles" by a large margin and penalizes specially "King" compared to the other metrics. The closeness does not show such a big gap between "Heracles" and the other actors, and similarly to the betweenness prefers "Zeus" over the human actors that posses more semantic value than him. This also happens in a smaller scale in the eigencentrality, that also preferred "Eurystheus" over the rest of the heroes. In general terms, classic centrality measures preferred gods, while the semantic value highlighted human and heroic figures.

Table II shows the result for the most important actors according to semantic value in the *Celtic Wonder-Tales* network. We found "Lugh", the most prominent god of the Irish pantheon, to own the highest $S$ value, followed by "Ireland". The third actor in $S$ value, "Conary", is an important mythical king of Ireland whose reign ends when he breaks three sacred oaths. The concept of "King" also has a high $S$, just as in the *Greek tales* case. Regarding the classical centrality measures, all of the them rated most highly "Ireland". Betweenness and closeness seem to be quite correlated in this case, showing the same top 3, but betweenness values quickly decrease after that. The eigencentrality significantly highlights "Balor" compared to the other metrics computed as it ranks in the third position. Contrary to most classical centrality measures, the semantic value favored the mythical embodiments of kingship like

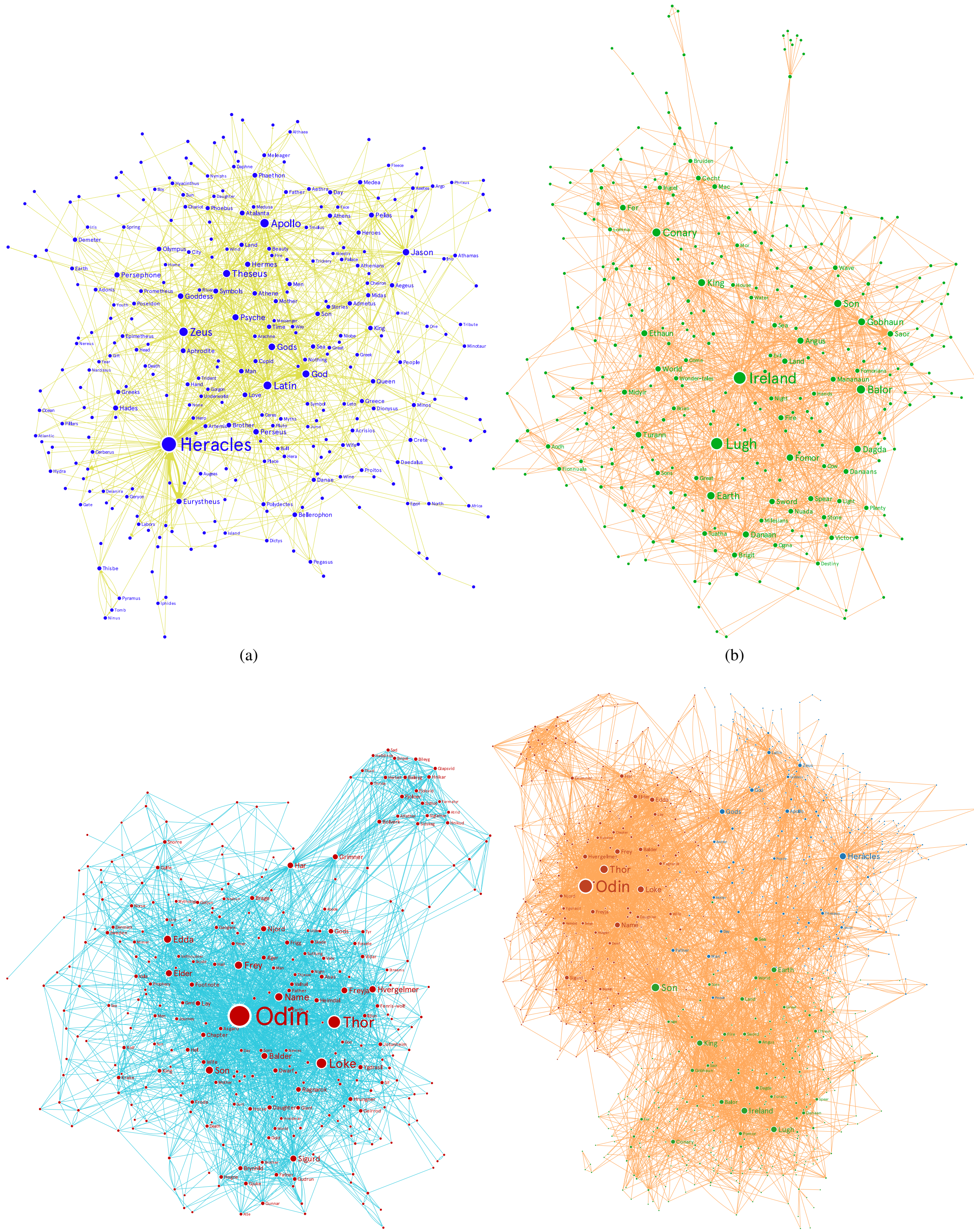



Fig. 2: **Word co-occurrence networks**. Each network is formed using the 300 most repeated entities in each corpus. We consider a connection between two words every time they appear less than 10 words apart from each other in one of the analyzed texts. **a.** *Greek Myths* **b.** *Celtic Wonder-Tales* **c.** *The Younger Edda* **d.** Fusion network of the three cultures. Node size is directly proportional to the in-degree measure and the layout algorithm considered is Force Atlas 2 [55].

"Lugh", "Conary", "Balor", and "Ethaun". This fact, alongside the high $S$ value of "Ireland", indicates a strong connection in this compilation between these mythical figures and the sovereign of the country (Table II). Such bond was not found in the other two mythologies.

Table III shows the result for the most important actors according to semantic value in the *The Younger Edda* network. We found "Odin", one of the main gods of the Germanic pantheon, to be the most important actor in terms of $S$. Being the father of all the Æsir, but also wise in the ways of magic and divination, the strength of these two different attributions might be the origin of such high $S$ value. Following "Odin", there is "Thor", another character with many attributions in his tales. A total of 6 gods populate this ranking, which shows that the gods themselves are more important in this mythology than in the other two. Regarding the four classical centrality measures, the top 3 is the same than in the semantic value. After the top 3, all the metrics prefer characters rather than concepts, but the semantic value puts "King" as the top 4 value, in consonance with the other compilations, which is an important difference.

Finally, Table IV shows the result for the most important actors according to semantic value in the fused network. We can see how the three fused structures can be recognized in the final structure but also numerous bridges have appeared to join them. As expected, these bridges are mostly general concepts, such as "Son", "Gods", and "Father", which connect the specific deities for each mythology. When analyzing the semantic values, "Odin" is again the one with the highest value. However, in this case the correlation of $I$ and $S$ seems to be less important. "King" presented a specially significant growth in $S$ value in the fuse network with respect to the individual networks. This growth is due to the fact that idea of kingship and authority is a key idea in the compilations studied, but they illustrate this idea with different attributions.

Comparing the semantic value to the remaining classic centrality measures, "Odin" generally gets the top value for them, which again reinforces the tendency of the semantic value to prefer human and heroic figures over deities. The eigencentrality and the degree centrality generally favored actors from the *Younger Edda*, probably because they form a more densely connected structure. The closeness put "Son" as the top value, since "Son" is an important bridge between the Nordic and Celtic entities.

### B. Semantic affinity analysis in the myth networks

In Fig. 3 we show how the 10 most repeated entities in each network, ordered according to semantic value, relate to each other in terms of semantic affinity and showcase some of the most interesting actors to study. There are some relationships to remark:

- In *Greek Myths* (Figure 3a), "Psyche", the impersonation of the human soul and lover of "Eros", receives significant semantic affinity from "Persephone", who is the wife of "Hades". There is a direct connection between "Persephone" and "Psyche", as they both appear in the same story, and both are the wife of a god and both are connected to the underworld. "Zeus" is connected more strongly to human characters than other gods and its most important connection is with "Psyche", the personification of the human soul.
- In *Celtic Wonder-Tales* (Figure 3b), "Dagda", the sun god, emits most semantic affinity to "King"and "Ireland", which suggests a relationship between earthly and divine mandates. Besides, "Dagda" is heavily entwined with "Ireland" but not with "Earth", which implies a negative connotation for "Earth". This might be in line with the idea that good things are "heavenly" things and "bad" things are more "earthly". "Earth" emits a lot of semantic value to "King", reinforcing again the bond between the earth and the ruler.
- In *Younger Edda* (Figure 3c), "Freyja" emits and receives significant affinity from "Loki". "Loki" is the responsible for the death of the almost invincible god "Baldr", who is also "Freyja"'s son. "Freyja" is considered the leader of the Valkyries and takes half of the fallen to her own afterlife field. This high affinity value here might indicate that the death theme is in fact a very important bond between them. "Frey", one of the most important Vanir gods, sends the most affinity to the actor "Son". "Hvergelmer" is the fountain in Nifelheim, the reign of the dead from which all rivers are born, and it is mostly associated with chaos. "Hvergelmer" sends and receives a significant amount of affinity from "Freyja" and "Loki", and both of them showed certain relationship with death. Besides, just as in the case of "Heracles" and "Hydra", the most emitted semantic affinity is to "Thor", which is considered to be associated to order.

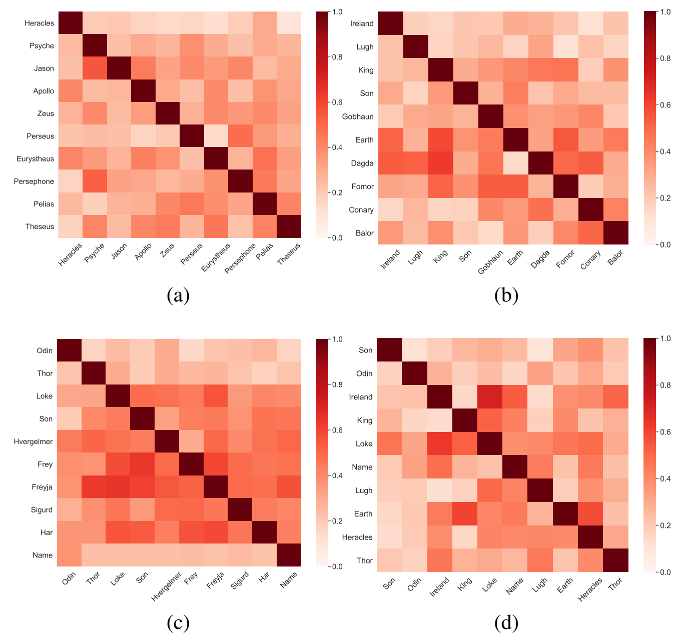



Fig. 3: **Semantic affinities in all the networks studied**. We chose the 10 most repeated entities in each text to compare themselves. **a** *Greek myths* network. **b** *Celtic Wonder-Tales* network. **c** *Younger Edda* network. **d** Fused myths network.

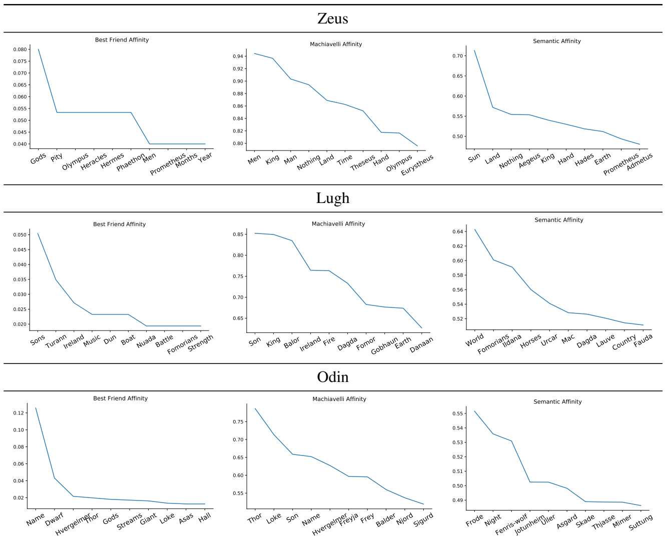



Fig. 4: **Study of affinities for three key characters.** Top 10 affinity values for the best friend, Machiavelli, and semantic affinity for "Zeus", "Lugh" and "Odin" in their respective networks.

- In the Fused myths network (Figure 3d), the highest affinity of "Lugh" is "Loki". This is a remarkable result, as there have been many studies discussing a possible a relationship between these two gods [56]. We also found a strong semantic affinity between "Ireland" and "Loki", in both directions, and between "Ireland" and "Thor", to a lesser extent. This might be due to "Ireland" being notably close to "Lugh", who is a god closely entwined to both "Odin" and "Loki", and because all of them are symbols related to authority in their original stories. "Earth" is notably affine to "King", which means that the strong tie between the land and the ruler present in the *Celtic Wonder-Tales* network is also present in the other two.

An in-depth analysis of some of the most relevant semantic affinities found is shown in Appendix B.

### C. Semantic affinity compared to other affinities

To complete our analysis, we have compared the results of different affinity functions in relevant characters in the original material. We have studied 3 different characters, chosen according to their relevance, showcasing their most important best friend, Machiavelli, and semantic affinities in Figure 4.

In the case of "Zeus", we can see that the best friend affinity includes mostly other Olympic gods. However, it is interesting to note that the results for the Machiavelli and semantic affinities do not show the same gods. This means that although "Zeus" appears repeatedly with other gods in his stories, he plays a different role than them in their stories. His highest semantic values reveal that he is mostly affine with general concepts, such as "Sun", "Land", "Nothing", and "Time", which indicates a connection between the world state and this god. The high affinities to "Sun" and "King" reinstate the connection of this god with the idea of authority.

"Lugh" is one of the most important gods in the Irish mythology and also a member of the "Tuatha Dé Danann". He is the maternal grandson of "Balor", the leader of the Fomorians, which makes him a descendant of both tribes of gods in this mythology. His best friend affinity values show that he is indeed tightly connected to the "Tuatha Dé Danann" and other authoritarian symbols such as "Ireland". His Machiavelli affinities show that indeed the structure of actors formed around him is similar to actors that wield authority, such as "King" and "Balor". The semantic affinity reveals that the top value is "World", which reflects how wide the attributions and roles for this god are. "Ildana", "Lauve" and "Fauda" are other words to refer to "Lugh".

"Odin"'s most important best friend affinities are "Name", as he is introduced many times in *The Younger Edda* with different titles, and "Dwarf", as the tribes of dwarves also appear repeatedly in the presence of "Odin" and share many attributions. The two most important Machiavelli affinities are his son "Thor" and "Loki". The relationship of "Odin" and "Loki" is very complex and they both take a central role in the many stories of the corpus. We also found that his highest semantic affinity is "Frode". "Frode" is another name for "Frey", the leader of the other tribe of gods in *The Younger Edda*, the Vanir. "Frode" is associated with authority and sacral kingship, which explains why these gods are so affine in this case. We also found that "Night" has a very high semantic affinity with "Odin". This is indeed quite an abstract connection but it is true that "Odin" is highly associated with the Wild Hunt, a repeated folklorical motif in which a group of supernatural hunters lead by a mythical figure chase the skies in the night [57]. He is also considered to be a god of the dead, as he greets fallen warriors in the Nordic afterlife, and he is also capable to raise dead out of the earth. He is also affine to "Fenris the wolf" and "Jotunheim", which are enemies to "Odin" and key characters in the developing of the Ragnarök.

### D. Comparison with other word embedding methods

In this section we compare our results with those obtained using methods to compute continuous representations of words. We have also considered transformer models, such as BERT [58], which are very popular in NLP tasks. However, these models generate a different word embeddings depending on the context, and in this work we have focused on singular representations for each concept.

First, we tried to study the word2vec architecture using the pre-trained embeddings with the glove-25 dataset [59]. Training the model using our texts resulted in trivial results due to the small size of the text corpus. However, we found that pre-trained model using the "glove-wiki-gigaword-300" dataset [59] contained all the required vocabulary for the *Greek Myths* texts. Figure 5 shows the resulting cosine similarities for the 10 most repeated entities. In this case, the most prominent similarity is "Zeus" and "Heracles" followed by "Heracles" and "Theseus". These similarities can be explained

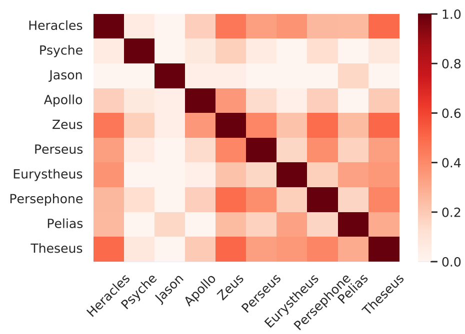



Fig. 5: **Cosine similarities using word2vec features for selected actors in the *Greek myths* text**. Word embeddings pre-trained using "glove-wiki-gigaword-300" dataset [59].

by the popularity of these characters, which should appear in generalist texts dealing with Greek mythology. Further connections and relationships are not present in these results, unlike the case of the semantic affinity.

In Appendix C we did a similar study for node2vec model [13].

## VI. Semantic value and semantic affinity in the Reuters terror word association network

The Reuters terror network comprises the words used in all pieces of news released during 66 consecutive days by the news agency Reuters, regarding the September 11 terrorist attack on the U.S [60]. The vertices of a network correspond to 13332 different words, joined by 243447 edges. Two words are connected in this network if they appeared together in a piece of news. In order to study this network, we have selected the 300 most repeated words.

We have considered as intrinsic value the bias induced by each word. In order to do so, we have used the SentiWordNet dictionary, that assigns a positive and a negative score to each word [61]. As intrinsic value, we have used the sum of the absolute value of both scores for each word. We have also computed the weighted degree, the closeness and the eigen centralities. We also considered the betweenness, but the results were all near 0 and almost indistinguishable from one another.

In this case, there is less correlation among the classical centrality measures and the semantic value (see Table V). The former ones focus on the structural properties of each word, while the latter one is also considering that some of these words transmit more information than others. For example, "death" and "attack" have a clear bias and are used to influence the readers' perspective. This is a noteworthy discrepancy between the semantic value and the classical centrality measures, as these two words have the lowest values of the table in Degree, Closeness and Eigencentrality. The contrary also happens in "fear". This word is the one with the highest classical centrality measures, yet its semantic value is not that high.

We have computed the Semantic affinity of some of these words and in Table VI we have reported the highest values for the incoming and outgoing semantic affinities for five words that were specially important in the news reports. Regarding the incoming affinities, "Death" is related to "Test" and "Investigation", which denotes that this word has been used in the context of the investigation of the origins of the attack. In contrast, we can see that for "Attack", the top words have most to do with institutions, like "Washington", and political figures, such as "President". This makes sense as the concept of attack has more political connotations than the rest of the chosen words. "Fear" receives most semantic affinity from "Night", which can have a psychological meaning. "Threat" receives from words related to personal issues, like "Work" and "Home", which are important things that people can feel threatened. It is also strongly connected to "News" and "Network", which are means of communication which can express the threat itself. Finally, "Time" receives from "Week" which is a unit of time, and from information that is connected to spatial locations, like "New York" and "American", which designate the target of the attack, and "Taliban", which identifies the ideology of the terrorists. Outgoing affinities are fairly similar to the incoming ones in all the words. However, a noteworthy difference is the presence of "Pakistani" in the "Threat" top outgoing semantic affinities, which is revealing about the perception of this nationality in these reports.

TABLE V: **Centrality measures in Reuters terror network**. For the top 10 most repeated words in the dataset, sorted according to their $S$ value.

| Word | S | Degree | Closeness | Eigencentrality |
|---|---|---|---|---|
| death | 0.75 | 129 | 0.54 | 0.21 |
| attack | 0.64 | 131 | 0.56 | 0.27 |
| good | 0.64 | 110 | 0.61 | 0.42 |
| concern | 0.63 | 152 | 0.55 | 0.25 |
| time | 0.55 | 291 | 0.64 | 0.51 |
| effort | 0.55 | 119 | 0.57 | 0.29 |
| fear | 0.54 | 1468 | 0.75 | 0.75 |
| threat | 0.46 | 591 | 0.69 | 0.61 |
| security | 0.42 | 483 | 0.63 | 0.48 |
| force | 0.40 | 199 | 0.57 | 0.33 |

TABLE VI: **Semantic affinity in Reuters terror network.** Top 5 entities that send or received most semantic affinity from the ones in the header, chosen as their relevancy in the original material.

| **Death** | **Attack** | **Fear** | **Threat** | **Time** |
|---|---|---|---|---|
| Incoming Semantic Affinity | | | | |
| Control | People | Night | Center | Week |
| Front | Official | Rule | Network | Afghanistan |
| Health | President | Concern | News | American |
| Test | Country | System | Work | New York |
| Investigation | Washington | Raid | Home | Taliban |
| Outgoing Semantic Affinity | | | | |
| Control | People | Rule | Center | Afghanistan |
| Health | Official | Night | News | Week |
| Test | President | Concern | Home | American |
| Front | USA | System | Network | New York |
| Local | Country | Pakistani | Service | Taliban |

## VII. Discussion and Conclusions

In this work we have studied the quantification of external information in a social network and how it propagates through its connections. Based on this formalization, we have proposed a new centrality measure to ponder the importance of each actor in the network, called the semantic value. We have also proposed a new way to compare the semantic value of two actors in the network, called the semantic affinity. Using our proposal, we model the way in which actors, words in our experimentation, enrich their own meanings by connecting with others. We have compared our proposals with other centrality measures and representation learning methods.

As a case study, we have considered the social networks resulting from three different mythological text collections, using the introduced proposals, existing social network analysis tools and different word embedding methods. Results in the three compilations have shown a mix of both historical and psychological relationships. Roughly speaking, we have found that gods are very close to kingship and authoritarian concepts, and particularly in the case of the Celtic myths, also to the land. We have also found that gods serve as a common nexus between the different topics that appeared in each of the compilation tales, and the Nordics and Celts prefer to focus their stories in gods, while the Greeks did so on humanoid and heroic figures. When comparing semantic affinities, we have found a strong bond between kingship and the earth. This has been previously studied in many traditions thorough the world [38] and it seems that it left its footprint in these tales as well. We have also found the traces of historical connections between the Nordic gods "Loki" and "Odin" and the Celtic myths. This has been hypothesized before in [56], and the connection is clear in this analysis.

Future lines of our research shall study the possible uses of the semantic value and the semantic affinity where important information available is outside of the network structure, as well as the reduction of the computational complexity of the semantic value. For instance, we intend to use it in trading and economic networks, where the financial assets of each agent can be used to construct the intrinsic value. We also aim to use the semantic value to study influence in different artistic works. Finally, we also intend to use the semantic value as a base to construct explainable features in order exploit the actor properties from outside the network.

## Appendix A
## Building the mythology networks

In this section we show how we built the network for each mythology. We discuss which books were used to form each network, some statistics regarding word counts, and how we processed the text to obtain the desired networks.

*1) Processing the texts:* We have chosen three of the ancient mythologies to perform the comparative study: Greek, Nordic, and Celtic. We have opted for these three due to their well-known interest and the existence of available compilations of tales translated to English, which makes the text processing for each book much easier. We have selected the following books as a basis for our analysis:

- *Celtic Wonder-Tales* by Ella Young (1867-1956) [65]. Originally written in 1910, it is a collection of Celtic traditional tales translated to modern English.
- *Greek Myths* by Olivia Collidge (1908-2006) [66]. Is a compilation of various stories regarding the classical Greek pantheon in modern English.
- *The Younger Edda* by Snorri Sturluson (1179-1241) [67]. *The Prose Edda* or *The Younger Edda* is a medieval Icelandic compilation of mythical texts, made by the Icelandic historian Snorri Sturluson.

Since we have the plain text files, it is easy to extract each chapter/tale in each book. We then parse each of them following the standard procedure [68] using a pre-trained multilayer perceptron in the Python Natural Language ToolKit [69]. We purge every word that is not a noun, since we only want to model interaction between entities and concepts. In Table VII we report the size of each book and the number of entities found.

TABLE VII: **Report of the size of each mythology.** Number of words, chapters and entities for each book in this work.

| Mythology | Book | Chapters | Words | Entities |
|---|---|---|---|---|
| Celt | *Celtic Wonder-Tales* | 13 | 41613 | 5114 |
| Greek | *Greek Myths* | 27 | 61246 | 5985 |
| Nordic | *The Younger Edda* | 21 | 65388 | 7521 |

*2) Obtaining the networks:* Once we have extracted the nouns from the text, what we have is a series of stemmed tokens. To obtain a network, we need the nodes and the edges to form it. In the case of the nodes, we will make a bijective association, so that one noun will correspond to one node, and *vice versa*. There are different ways to compute the edges in terms of noun co-occurrence. We have decided to create an edge every time a word appears in a $k$-distance or less from another in the text, choosing $k$ as 10.

*3) Fusing the networks:* Given the network for each tale, we can fuse them to obtain a "global" network containing the information from all the different networks referring to each tale. There are no problems of scale in this context, since all stories range from 2 to 7 pages long only. So, we simply add up all the edges into a single network. When an edge between two actors is repeated in various of the networks, we take the highest value.

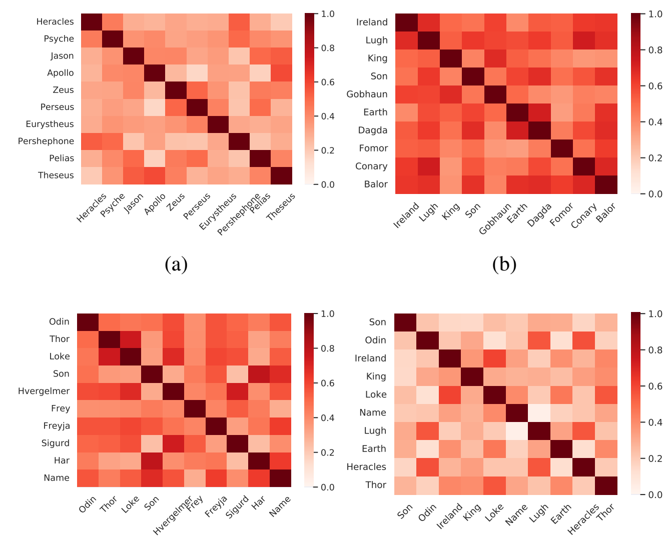
## Appendix B
## Explanation of relevant semantic affinities found

In this section we have studied some of the most interesting obtained results using semantic affinities, because they were unexpected, particularly relevant, or not easily explainable. To do so, we have studied the actors contained in the paths used in the Pipe algorithm in each case, in order to find which connections resulted in those affinities:

- "Ireland" and "Loki": "Loki" received the most semantic affinity from "Ireland" in the fused network of the three corpora. It was a surprising result, as they are concepts from different mythologies, and they are not clearly connected. The Pipe algorithm visited 16 actors to compute this semantic affinity. They mostly refer to general concepts present in all the compilations of texts, like “Fire”, “King” and “Land”, as well as other gods. This attributions are very affine to both “Loki” and “Ireland”, which explains the high affinity value. Besides, “Ireland” exhibits a high homophily with respect to the Nordic gods.

- “Odin” and “Lugh”: these gods come from different cultures, however, their origins are intertwined [56]. The Pipe algorithm indeed computed a high semantic affinity in this case, visiting a total of 26 actors. Most of these actors are not specific to one mythology but are common concepts like “Music”, “Battle”, and “Fire”. This explains why this affinity is high: they are both strongly related to a set of common concepts in both compilation of tales, which reveal that these gods have similar attributions.
- “Freyja” and “Loki”: “Freyja” and “Loki” exhibit a strong connection. This was unexpected, as we expected “Freyja” to have more semantic ties with other Vanir gods. The Pipe algorithm used a total of 30 actors for this semantic affinity. In this list, there are some common words to all Nordic gods like “Æsir”, “Asgard”, and “Gods”. There are also concepts related to the death theme, like “Wolf” (referred to Fenris the Wolf), “Night”, and “Death”. There also three other important female deities in the visited actors, “Frigg”, “Sif”, and “Asynjes” (the female word for “Aesir”). These female connections are relevant, because they connect “Loki”’s deceptive capabilities with the magic and witchcraft attributions that are exclusively linked to female figures in this culture. In fact, in one of the stories, “Loki” transforms himself into a mare, giving birth to Odin’s horse. These findings indicate that this high affinity value is due to the godly condition of both characters, their strong relationship with fatality, and the resemblance between “Loki”’s wit and transfiguration powers with “Freyja”’s divination capabilities.

## APPENDIX C
## COMPARISON WITH NODE2VEC




(a) (b) (c) (d)

Fig. 6: Cosine similarities using node2vec features for selected actors in the three mythological texts. We chose the 10 most repeated entities in each text to compare themselves. **a** *Greek myths*. **b** *Celtic Wonder-Tales*. **c** *Younger Edda*. **d** Fused myths network.

In this section we computed the results obtained with a node2vec model, displayed in Figure 6. In this case we obtained better results for the three mythologies than in the case of pure word2vec models. However, the conclusions are similar: most of the similarities can be explained by mere co-occurrences. For the case of *Greek myths* we found that the two important similarities: “Apollo” and “Theseus”, and “Heracles” and “Persephone”, which can be explained due to their co-appearences in some tales. In the *Celtic-Wonder tales* network we found a particularly high average similarity value with no apparent reason. We found a specially high value between “Dagda” with “Earth” and ‘Balor”, and between “Conary” with “Balor” and “Lugh”. “Dagda” appears together many times with both terms which explains such value. For the case of “Conary”, we did not find any clear criterion to connect it to the other characters, besides they were all authority figures in their respective tales. In the *Younger Edda* network, “Thor” and “Loki” present a very high similarity, because they appear in many different stories together. Finally, in the fused network we found that “Loki” and “Ireland”, and “Odin” and “Lugh” are very similar, a result that we also found with the semantic affinity. “Odin” and “Heracles” are significantly similar, because they have so many attributions that some of them inevitably connect them. Again, we also found a strong similarity between “Thor” and “Loki”, for the same reasons as in the *Young Edda* network.

## APPENDIX D
## ACKNOWLEDGMENTS

Javier Fumanal Idocin and Humberto Bustince's research has been supported by the project PID2019-108392GB I00 (AEI/10.13039/501100011033).

Oscar Cordón research has been funded by the government of Spain, EXASOCO (PGC2018-101216-B-I00), including European Regional Development Funds (ERDF).

Antonio Francisco Roldán López de Hierro research has been funded by the project Project PID2020-119478GB-I00 FEDER Project A-FQM-170-UGR20.

## REFERENCES

[1] J. Scott, "Social network analysis," *Sociology*, vol. 22, no. 1, pp. 109–127, 1988.

[2] A.-L. Barabási, "Network science," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1987, p. 20120375, 2013.

[3] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge university press, 1994, vol. 8.

[4] D. Camacho, Á. Panizo-LLedot, G. Bello-Orgaz, A. Gonzalez-Pardo, and E. Cambria, "The four dimensions of social network analysis: An overview of research methods, applications, and software tools," *Information Fusion*, vol. 63, pp. 88–120, 2020.

[5] M. Newman, *Networks*. Oxford university press, 2018.

[6] C. S. Fischer, *To dwell among friends: Personal networks in town and city*. University of Chicago Press, 1982.

[7] M. Grohe, "word2vec, node2vec, graph2vec, x2vec: Towards a theory of vector embeddings of structured data," in *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, 2020, pp. 1–16.

[8] O. Levy and Y. Goldberg, "Dependency-based word embeddings," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2014, pp. 302–308.

[9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[10] J. Lilleberg, Y. Zhu, and Y. Zhang, "Support vector machines and word2vec for text classification with semantic features," in *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*. IEEE, 2015, pp. 136–140.

[11] M. Seok, H.-J. Song, C.-Y. Park, J.-D. Kim, and Y.-s. Kim, "Named entity recognition using word embedding as a feature," *International Journal of Software Engineering and Its Applications*, vol. 10, no. 2, pp. 93–104, 2016.

[12] J. Acosta, N. Lamaute, M. Luo, E. Finkelstein, and C. Andreea, "Sentiment analysis of twitter messages using word2vec," *Proceedings of Student-Faculty Research Day, CSIS, Pace University*, vol. 7, pp. 1–7, 2017.

[13] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 855–864.

[14] X. Shen and F.-L. Chung, "Deep network embedding for graph representation learning in signed networks," *IEEE transactions on cybernetics*, vol. 50, no. 4, pp. 1556–1568, 2018.

[15] Z. Yu, G. Zhang, J. Chen, H. Chen, D. Zhang, Q. Yang, and J. Shao, "Toward noise-resistant graph embedding with subspace clustering information," *IEEE Transactions on Cybernetics*, 2021.

[16] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, and A. Jain, "Unsupervised word embeddings capture latent knowledge from materials science literature," *Nature*, vol. 571, no. 7763, pp. 95–98, 2019.

[17] G. Di Gennaro, A. Buonanno, and F. A. Palmieri, "Considerations about learning word2vec," *The Journal of Supercomputing*, pp. 1–16, 2021.

[18] J. Ye, "Improved cosine similarity measures of simplified neutrosophic sets for medical diagnoses," *Artificial Intelligence in Medicine*, vol. 63, no. 3, pp. 171–179, 2015.

[19] A. Landherr, B. Friedl, and J. Heidemann, "A critical review of centrality measures in social networks," *Business & Information Systems Engineering*, vol. 2, no. 6, pp. 371–385, 2010.

[20] T. Martin, X. Zhang, and M. E. Newman, "Localization and centrality in networks," *Physical review E*, vol. 90, no. 5, p. 052808, 2014.

[21] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[22] J. Fumanal-Idocin, A. Alonso-Betanzos, O. Cordón, H. Bustince, and M. Minárová, "Community detection and social network analysis based on the italian wars of the 15th century," *Future Generation Computer Systems*, vol. 113, pp. 25 – 40, 2020.

[23] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[24] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of machine learning research*, vol. 3, no. Feb, pp. 1137–1155, 2003.

[25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[26] E. Arisoy, T. N. Sainath, B. Kingsbury, and B. Ramabhadran, "Deep neural network language models," in *Proceedings of the NAACL-HLT 2012 Workshop: Will We Ever Really Replace the N-gram Model? On the Future of Language Modeling for HLT*, 2012, pp. 20–28.

[27] T. Mikolov, S. Kombrink, L. Burget, J. Černockỳ, and S. Khudanpur, "Extensions of recurrent neural network language model," in *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2011, pp. 5528–5531.

[28] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[29] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *arXiv preprint arXiv:1607.01759*, 2016.

[30] E. Choi, M. T. Bahadori, E. Searles, C. Coffey, M. Thompson, J. Bost, J. Tejedor-Sojo, and J. Sun, "Multi-layer representation learning for medical concepts," in *proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1495–1504.

[31] P. Cui, X. Wang, J. Pei, and W. Zhu, "A survey on network embedding," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 5, pp. 833–852, 2018.

[32] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," *Advances in neural information processing systems*, vol. 24, pp. 1413–1421, 2011.

[33] S. Pan, R. Hu, S.-f. Fung, G. Long, J. Jiang, and C. Zhang, "Learning graph embedding with adversarial training methods," *IEEE transactions on cybernetics*, vol. 50, no. 6, pp. 2475–2487, 2019.

[34] M.-C. Costa, A. Hertz, and M. Mittaz, "Bounds and heuristics for the shortest capacitated paths problem," *Journal of Heuristics*, vol. 8, no. 4, pp. 449–465, 2002.

[35] A. V. Goldberg and C. Harrelson, "Computing the shortest path: A* search meets graph theory," in *Proc. 16th ACM-SIAM Symposium on Discrete Algorithms, 2005*, vol. 5, 2005, pp. 156–165.

[36] M. E. Stith and M. Leach, *The Beginning: Creation Myths around the World*. New York: Funk & Wagnalls, 1958.

[37] J. Campbell, *The hero with a thousand faces*. New World Library, 2008, vol. 17.

[38] J. G. Frazer, *The Golden Bough*. Palgrave Macmillan UK, 1922.

[39] J. E. Cirlot, *Diccionario de símbolos (in Spanish)*. Siruela, 2004.

[40] S. S. Elliott, *Reinventing Religious Studies: Key Writings in the History of a Discipline*. Routledge, 2014.

[41] T.-S. B. Liew, *Present and Future of Biblical Studies: Celebrating Twenty-five Years of Brill's Biblical Interpretation*. Brill, 2018.

[42] A. Lassen, "Old norse mythology—comparative perspectives," *Journal of English and Germanic Philology*, vol. 118, no. 2, pp. 298–300, 2019.

[43] J. Bengtson, "Iarl and Iormun-; Arya-and Aryaman-: A study in indo-european comparative mythology," *Comparative Mythology*, vol. 2, no. 1, pp. 33–67, 2016.

[44] Y. Wen, "Norse mythology and chinese mythology in comparison: The origin of the cosmos, time and space," *Germanic Myths in the Audiovisual Culture*, vol. 5, p. 183, 2020.

[45] W. Burkert, *Oriental and Greek mythology: the meeting of parallels*. Verlag nicht ermittelbar, 1987.

[46] B. G. Purzycki, "The minds of gods: A comparative study of supernatural agency," *Cognition*, vol. 129, no. 1, pp. 163–179, 2013.

[47] C. S. Littleton, *The new comparative mythology: an anthropological assessment of the theories of Georges Dumézil*. Univ of California Press, 1973.

[48] J. Dow, "Is religion an evolutionary adaptation?" *Journal of Artificial Societies and Social Simulation*, vol. 11, no. 2, p. 2, 2008.

[49] D. Chartash, "Maths meets myths: Quantitative approaches to ancient narratives," *Folklore*, vol. 130, pp. 315 – 317, 2019.
[50] J. Yose, *A network scientific approach to the quantitative analysis of epic texts*. Coventry University, 2017.
[51] J. Nescolarde-Selva, J. Usó-Doménech, and M. Lloret-Climent, "Mythical systems: mathematic and logical theory," *International Journal of General Systems*, vol. 44, no. 1, pp. 76–97, jan 2015.
[52] R. Kenna and P. MacCarron, "A networks approach to mythological epics," in *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*. Springer, 2017, pp. 21–43.
[53] P. J. Miranda, M. S. Baptista, and S. E. de Souza Pinto, "The odyssey's mythological network," *PLoS One*, vol. 13, 2018.
[54] P. Mac Carron and R. Kenna, "Universal properties of mythological networks," *EPL (Europhysics Letters)*, vol. 99, no. 2, p. 28002, 2012.
[55] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, "Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software," *PloS One*, vol. 9, no. 6, p. e98679, 2014.
[56] T. Ewing, "The birth of Lugh—Odin and Loki among the Celts," *Sinsear*, vol. 8, p. 119, 1995.
[57] C. Lecouteux, *Phantom armies of the night: the wild hunt and the ghostly processions of the undead*. Simon and Schuster, 2011.
[58] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," *arXiv preprint arXiv:1810.04805*, Oct 2018.
[59] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.
[60] V. Batagelj and A. Mrvar, "Density based approaches to network analysis-analysis of reuters terror news network," *Proceedings of the Workshop on Link Analysis for Detecting Complex Behavior (LinkKDD2003)*, 2003.
[61] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining," in *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, 2010.
[62] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
[63] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, oct 2008.
[64] J. Fumanal-Idocin, O. Cordón, M. Minárová, A. A. Betanzos, and H. Bustince, "Combinations of affinity functions for different community detection algorithms in social networks," in *Annual Hawaii International Conference on System Sciences*. IEEE, 2022, pp. 2107–2114.
[65] E. Young, *Celtic Wonder-Tales*. Maunsel and Co., 1923.
[66] O. E. Coolidge, *Greek Myths*. Houghton Mifflin Company, 1948.
[67] S. Sturlson, *The Prose Edda*. New York, The American-Scandinavian Foundation, 1916.
[68] J. J. Webster and C. Kit, "Tokenization as the initial phase in NLP," in *Proceedings of the 14th Conference on Computational Linguistics - Volume 4*, ser. COLING '92. Association for Computational Linguistics, 1992, p. 1106–1110.
[69] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc., 2009.

**Javier Fumanal Idocin** holds a B.Sc in Computer Science at the University of Zaragoza, Spain (2017) and a M.Sc in Data Science and Computer Engineering at the University of Granada, Spain (2018). He is now a PhD Student of the Public University of Navarre, Spain in the department of Statistics, Informatics and Mathematics. His research interests include machine intelligence, fuzzy logic, social networks and brain-computer interfaces.

**Oscar Cordón** is Full Professor with the University of Granada (UGR), Spain. He was the founder and leader of its Virtual Learning Center (2001-05) and the Vice-President for Digital University (2015-19). He also was one of the founding researchers of the European Centre for Soft Computing (2006-2011), being later contracted as Distinguished Affiliated Researcher (2011-15). He has published 400 peer-reviewed scientific publications including a research book on genetic fuzzy systems and 119 JCR-SCI-indexed journal papers (71 in Q1, 41 in D1), advised 20 Ph.D. dissertations, and coordinated 41 research projects and contracts with an overall amount of 10M€. His publications had received 15558 citations (h-index=61), and he has been included in the 1% of most-cited researchers in the world according to Clarivate Analytics' Web of Science and in the 2% of most-cited researchers in the world in Artificial Intelligence (source: 'Ranking of the World Scientists: World's Top 2% Scientists', University of Stanford). He was a member of the Working Group that developed the Spanish R+D Strategy for Artificial Intelligence by the Spanish Ministry of Science, Innovation and Universities, 2018-19.

Prof. Cordón received the UGR Young Researcher Career Award in 2004; the IEEE Computational Intelligence Society (CIS) Outstanding Early Career Award in 2011; the IFSA Award for Outstanding Applications of Fuzzy Technology in 2011; the National Award on Computer Science ARITMEL by the Spanish Computer Science Scientific Society in 2014, and was recognized by the Spanish Association for Artificial Intelligence in 2020. He was elevated to IEEE Fellow in 2018 and received the IFSA Fellow in 2019.

**Graçaliz Pereira Dimuro** received the Undergraduate degree from Universidade Católica de Pelotas, Pelotas, Brazil, and the M.Sc. and Ph.D. degrees from the Instituto de Informática of Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. She is currently an Adjunct Professor with the Centro de Ciências Computacionais, Universidade Federal do Rio Grande, Rio Grande, Brazil and a Researcher of the Brazilian Research Funding Agenc and CNPQ.

**Antonio Francisco Roldán López de Hierro** received the M.Sc. and Ph.D. degrees in mathematics from the University of Granada, Spain, in 1998 and 2003, respectively, and the Ph.D. degree in statistics from the University of Jaen, Spain, in 2012. He is currently a Teacher with a secondary school and as a Lecturer with the Department of Quantitative Methods for Economics and Business, University of Granada. He is a coauthor of approximately 45 original research articles. His research interests include fuzzy sets theory, fuzzy regression, fuzzy ranking, decision making, and fixed point theory.

**Humberto Bustince** received the Graduate degree in physics from the University of Salamanca in 1983 and Ph.D. in mathematics from the Public University of Navarra, Pamplona, Spain, in 1994. He is a Full Professor of Computer Science and Artificial Intelligence in the Public University of Navarra, Pamplona, Spain where he is the main researcher of the Artificial Intelligence and Approximate Reasoning group. He has led 11 I+D public-funded research projects, at a national and at a regional level. He is currently the main researcher of a project in the Spanish Science Program and of a scientific network about fuzzy logic and soft computing. He has taken part in two international research projects. He has authored more than 210 works in conferences and international journals. Moreover, five of these works are also among the highly cited papers of the last ten years, according to Science Essential Indicators of Web of Science. Dr. Bustince is the Editor-in-Chief of the online magazine Mathware & Soft Computing of the EUSFLAT and of the Axioms journal. He is an Associated Editor of the IEEE Transactions on Fuzzy Systems Journal and a member of the editorial board of the Journals Fuzzy Sets and Systems, Information Fusion, International Journal of Computational Intelligence Systems and Journal of Intelligent & Fuzzy Systems. Honorary Professor at the University of Nottingham, National Spanish Computer Science Award in 2019 and EUSFLAT Excellence Research Award in 2019.