Analysis of Semantic Networks using Complex Networks Concepts

Ortiz-Arroyo, Daniel

Published in:
Flexible Query Answering Systems

DOI (link to publication from Publisher):
10.1007/978-3-642-40769-7_12

Publication date:
2013

Analysis of Semantic Networks using Complex Networks Concepts

Daniel Ortiz-Arroyo

Computational Intelligence and Security Laboratory
Department of Electronic Systems, Aalborg University
Niels Bohrs Vej 8, Esbjerg, Denmark
do@es.aau.dk

Abstract. In this paper we perform a preliminary analysis of semantic networks to determine the most important terms that could be used to optimize a summarization task. In our experiments, we measure how the properties of a semantic network change, when the terms in the network are removed. Our preliminary results indicate that this approach provides good results on the semantic network analyzed in this paper.

Keywords: Complex Networks, Semantic Networks, Information Theory

1 Introduction

Automatic text summarization is a computer processing task that consists in selecting those sentences within a text that best represent its contents. One way to perform summarization is by assigning a score to each of document's sentences, according to its importance.

Many approaches have been explored in the past to perform automatic text summarization. Among these are the application of TF-IDF[1] to assign importance scores or the use of more elaborated algorithms based on fuzzy logic, genetic algorithms, neural networks, semantic role labeling, and latent semantic analysis.

Automatic text summarization can be applied not only to full documents but also to a group of phrases or sentences contained in a document. The goal is to extract those keywords or terms that best summarize sentences' contents. After these sentences have been extracted from a document, they can be represented as a semantic network.

In general a semantic network is one form of knowledge representation that depicts how terms or concepts are inter-related. Different types of semantic networks are used for different purposes. For instance, semantic networks can be used in defining concepts, representing beliefs or causality, or in performing inferences.

[1] Term frequency-inverse document frequency

We use a broad definition of what semantic networks are to represents not only relationships between concepts but how words or terms used in phrases or sentences are inter-related. In particular we use certain word properties, such as their position within a sentence or their frequency of co-occurrence with other words. Other properties that can be used to create a semantic network from sentences are its syntactical structure, or the grammatical category to which the words in them belong.

In these semantic networks the words within sentences are the nodes in the graph and the syntactical or grammatical relationship existing between words represent the edges. This type of semantic network is described in [5] and will be used in this paper.

Previous studies [6] have shown that semantic networks have some of the properties that complex networks possess.

Complex networks are networks that are neither random [2] nor regular. Complex networks have some non-trivial topological properties that differentiate them from random and regular networks. [3] The discovery of these properties has produced an exponential growth of interest in these networks during the last years.

Some of the well known properties of complex networks are *scale-free* degree distribution and *small-world effect*. In a scale-free network, the degree distribution of the nodes follows a power-law. This basically means that a few nodes in the network have connections to many other nodes, but most nodes in the network have just a few connections with the rest of the nodes. An example of a network with scale-free degree distribution is the Internet. Its scale-free property explains why the Internet network is resilient to the random failures that may occur in some of the nodes. The probability that a random failure occurs in one of the few of the nodes that have a large number of connections is smaller compared to the probability that a node with few connections fails.

The power-law describes probability distributions that also commonly occur in other phenomena in nature and society. An example is the Pareto distribution. This distribution describes how wealth is distributed within society i.e. that a few percentage of a population owns most of wealth of a country and that most population owns little of that wealth.

Another property that characterizes complex networks is the *small-world* effect. This effect characterizes complex networks that have a high global clustering coefficient. This means that nodes in a complex network tend to lay at relatively short geodesic distances [4] between each other, compared to how nodes are clustered in a random network. In social networks this property commonly occurs in the form of closing triads that describe fact that "the friends of my friends are also commonly my friends".

[2] Random Networks are also called Erdos-Renyi networks

[3] In this paper we will use as synonyms the terms graph and network, node and vertex, and edge and link.

[4] Geodesic distances are also called shortest paths

The small-world effect is also known as the “six-degrees” of separation, a metric that describes the average number of links that separates two persons in a social network. A similar effect has been observed in networks extracted from bibliographic cites in mathematical papers (called the Erdos number) or from movie actors (called the Bacon number). In these networks the average degree of separation between authors or actors is even smaller than six.

The scale-free power-law distribution can be used to build synthetic models of complex networks, using a preferential attachment process. In the preferential attachment process, a network is built iteratively by connecting new nodes with higher probability to nodes in the network that are already highly connected.

Complex networks have multiple applications in a wide variety of fields such as the Internet, energy, traffic, sociology, neural networks, natural language etc.

Interestingly, the distribution of words in natural languages show some of the known properties of complex networks. For instance, the well known Zipf’s law, states that the frequency of words follows a power-law distribution. This fact has been used to compress text documents efficiently by assigning shortest codes to most frequently used words.

In this paper we use semantic networks extracted from sentences and methods from complex networks to find the terms within these sentences that best summarize its contents. We compare the experimental results obtained by applying two different methods from complex networks. Our preliminary results indicate that this approach shows good results in the experiment we have performed.

This paper is organized as follows. Section 2 presents a brief summary of related work. Section 3 describes the methods we used and the intuitions behind them. Section 4 presents the preliminary results of our experiments and section 5 concludes the paper and describes future work.

2 Related Work

There is a plethora of research work in automatic summarization systems and complex networks. In this section we will provide a brief summary of the research work that is directly related to the approach presented in this paper.

Many approaches have been proposed in the literature to perform automatic summarization. Among these are supervised and unsupervised machine learning-based methods.

In [7] both methods were applied to the summarization task. Classifiers were constructed using supervised methods such as J48, Naive Bayes, and SVM [5]. In the same work, classifiers were also induced using the HITS algorithm in an unsupervised way. Results of the experiments reported in [7] show that supervised methods work better when large labeled training sets are available, otherwise unsupervised methods should be used.

In [5] an approach to extract keyphrases from books is presented. Phrases are represented as semantic networks and centrality measures are applied to extract

[5] Support Vector Machine

those phrases that are the most relevant. The method employs an unsupervised machine learning method and the concepts of *betweeness centrality* and *relation centrality* as feature weights to extract keyphrases. Relation centrality measures dynamically, the contribution of a node to the connectedness of the network. Relation centrality counts statically, how many routes betweeness centrality is actually shortening.

On the side of complex networks, the communication efficiency of a network is defined in [3] as a function that is inversely proportional to the length of the shortest path between any two nodes. The effect that one node has on the overall efficiency of a network is found by calculating how the network's efficiency changes when that node is removed. Those nodes that have a larger, detrimental effect on network's communication efficiency, are considered the most important since their removal will force network's communication to happen through larger paths. This approach was employed to find the importance of the members of a terrorist organization in [3].

In [1] an approach to find sets of key players within a social network was presented. The method consists in selecting simultaneously k players via combinatorial optimization.

In [6] it was shown that several types of semantic networks have a small-world structure with sparse connectivity. Authors found that these semantic networks have short average path lengths between words, and a strong local clustering that is typical in structures that have the small-world property. The distribution of the number of connections observed, indicates that these networks follow a scale-free pattern of connectivity.

In a related work described in [4], we found that the concept of entropy can be applied to find sets of key-players within a social network. This approach works well in networks that have a sparse number of edges. The reason is that the removal of a node in dense networks will still keep the network very dense, making the changes in entropy very small.

Shannon's definition of entropy used as a metric to identify important nodes in a network has been previously reported in a diversity of research work. For instance, in analyzing social networks extracted from a corpus of emails [2], in finding key players in social networks [4], and in other very different application domains such as city planning [8]. However, the definition of the probability distributions used in these works to calculate entropy changes slightly. For instance in [8] the probability distribution employs all shortest paths that pass through certain node and [4] includes all shortest paths that originate from a node.

3 Finding the Most Important Terms in a Semantic Network

The main objective of this paper is to determine if some of the concepts applied in complex networks and social network analysis are useful to find the most important terms within the phrases or sentences of a document, that best summarize its content.

In this approach, the terms used in phrases are represented as a semantic network. The semantic network may be obtained in different ways. One way is by using the relative position of words within a phrase or group of phrases. Other methods analyze the syntactic relation of the terms among each other and/or using the grammatical category to which they belong.

In our experiments we have used the semantic network that represents the phrases extracted from a book that best represent its content as is described in [5]. The method used in that work to generate the semantic network, employs neighboring relations and the co-occurrence of terms within phrases.

In our analysis we have used the concept of centrality entropy. Centrality entropy represents the uncertainty that nodes could be able to reach other nodes in the network through shortest paths when a node is removed from the network.

Centrality entropy can be calculated using Shannon's definition of entropy:

$$C_e(G) = -\sum_{i=1}^{n} p_g(i) log(p_g(i)) \tag{1}$$

where $C_e(G)$ is the centrality entropy of graph G and $p_g(i)$ represents the probability distribution of the shortest paths from node i to all other nodes in the network. This probability distribution is defined as:

$$p_g(i) = \frac{g_p(i)}{\sum_{j=1}^{n} g_p(j)} \tag{2}$$

where the numerator $g_p(i)$ is the number of shortest paths that communicate node i with all other nodes in the network and the denominator is the total number of shortest paths that exist in the network. Note that the actual length of the shortest paths [6] is not used to calculate centrality entropy. Entropy, defined in this way, changes as nodes are deleted from the graph, disconnecting some nodes and reducing as a consequence, the number of shortest paths available in the network to communicate the rest of the nodes in the graph.

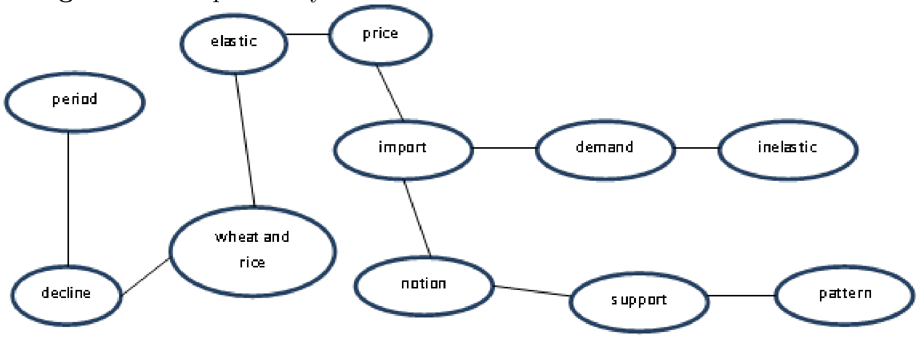
A similar method has been proposed in [3] to detect important nodes in a network. The method determines how the communication efficiency of a network changes when nodes are removed. In this case communication efficiency may be interpreted as how important a node is to establish a semantic link between the terms in the network. Communication efficiency is measured using the equation described in [3]:

$$E(G) = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{ls_{ij}} \tag{3}$$

where $E(G)$ is the efficiency of graph G, n the total number of nodes in the graph, and ls_{ij} is the length of the shortest path between nodes i and j. The equation shows that communication efficiency is inversely proportional to the length of the shortest path.

[6] Shortest paths are also called *geodesic paths*

Fig. 1. An example of a syntactic semantic network extracted from 2 sentences



To procedure used in both cases, to measure the efficiency of a network and to find centrality entropy, consists in disconnecting nodes one by one and measuring the efficiency or entropy of the resulting network.

4 Experimental Results

In our preliminary experimental results we used two sentences that were extracted from a book and analyzed syntactically as is described in [5]. The sentences are:

"The import price elasticities remain less than one for both wheat and rice and decline over the entire period. This pattern again tends to support the notion that import demand is inelastic"

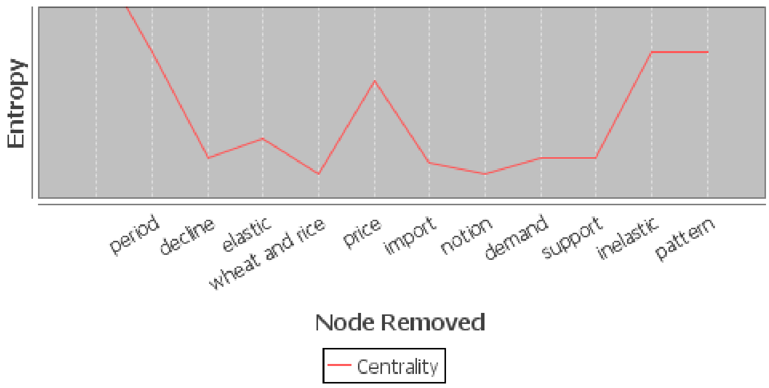
Arguably the main subject of these two sentences is *"the notion of how import of wheat and rice behaves"*. Therefore, we could conclude that the terms in the semantic network that may be used to summarize the main topic of these two sentences are {*notion, import, wheat and rice*}

The semantic network generated from these two sentences was taken from [5] and is shown in Fig. 1. As is described in [5], the sentences were pre-processed using stop word removal and stemming. Then, sentences were selected and the network was created using an unsupervised machine learning method that employs as feature weights, two different centrality measures.

We apply our method to analyze how the entropy of the semantic network changes when nodes in the semantic network are removed. First, we calculate the total entropy of the network using Eq. 1. Afterwards, nodes are removed one by one, recalculating in each iteration, the probability distributions and the total entropy of the graph.

Using this method we obtained a plot that shows how entropy changes in Fig. 2.

Fig. 2. Drop in total entropy when nodes in the semantic network are removed one by one



The entropy defined in Eq. 1 provides a measure of the probability that a node could be reached from any other node in the graph through shortest paths. In a fully connected graph, the probability is 1 since a node can reach any other node in the graph through a single edge. Hence, no matter which node is removed the entropy will be the same since the remaining nodes will still keep the graph fully connected.

As graphs become more sparse, some nodes could be reached through shortest or non-shortest paths. However, in centrality entropy only shortest paths are used since we are interested in finding the nearest related terms. In the semantic network that we will analyze, the shortest paths represent how semantically close are the terms in the network.

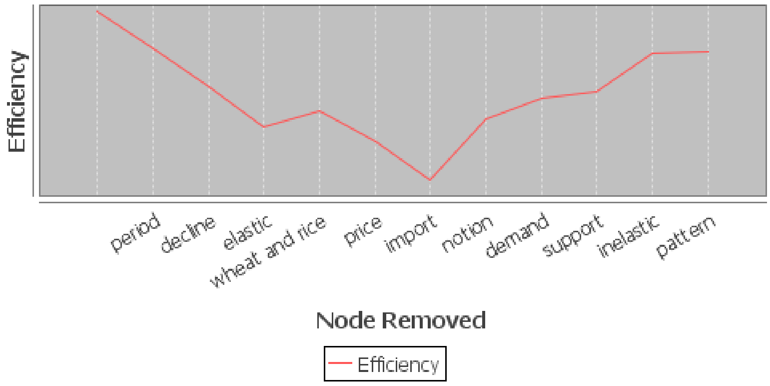
When nodes are removed from the graph, these nodes that produce the largest drop in entropy are considered the most important since their removal will reduce the number of shortest paths that the remaining nodes in a graph could use to reach the rest of nodes in the graph.

A threshold value can be used to determine how many of these important terms will be included in the summarization task.

The centrality entropy drop graph obtained in Fig. 2, indicates that the nodes that have most effect, when removed from the network are {*notion, import, wheat and rice,* } and to a lesser degree {*decline, period, elastic, price, demand, support, inelastic, pattern*}. By changing the threshold value more or less terms could be included as the most important ones.

Interestingly, the {*price*} term was not detected as an important term by the centrality entropy calculation. This is firstly due to the fact that, as can be seen in Fig. 2, the {*import*} term works as a hub for terms {*notion*} and {*demand*}, making it important since its removal will reduce the number of shortest paths that will be available in the graph compared to the effect that the term {*price*} may produce on entropy when removed.

Fig. 3. Drop in network's communication efficiency when nodes in the semantic network are removed one by one



We could ask why our method finds that terms such as {*wheat and rice*} are more important than other terms such as {*elastic*} or {*decline*}. These terms seem to have similar importance judged by their position in the network.

The reason is that when the term {*decline*} is removed, the node {*period*} becomes isolated from the graph and the number of shortest paths available in the graph decreases proportionally for the rest of the nodes. However, that single isolated node does not contribute to the total shortest paths available.

When node {*elastic*} is removed, the original graph is split into two graphs. The one containing nodes { *period, decline, wheat and rice*} and the one containing {*price, import, demand, inelastic, notion, support, pattern*}. In this case when node {*elastic*} is removed, the number of shortest paths will be reduced since the larger graph will not be able to reach the smaller graph. However, the smaller graph with 3 nodes still provides some local shortest paths to reach these local nodes i.e. there will be 6 shortest paths within the smaller network.

Finally, when node {*wheat and rice*} is removed, the graph is again split into two graphs, but in this case the smaller graph consisting of only two nodes {*period, decline*}, provides only 2 shortest paths in the smaller network, decreasing the total amount of shortest paths available and with this the probability that some node in the network could reach any other node.

Fig. 3 shows how network's communication efficiency, defined in Eq. 3, changes when nodes are removed one by one from the network. The plot shows that the terms that produce the maximum drop in efficiency are firstly {*import, price*} and then {*elastic, wheat and rice, notion*} with the rest of terms having a lesser degree on the drop in efficiency. Arguably, these first two terms {*import, price*}, do not fully capture the *"the notion of how import of wheat and rice behaves"*. However, if we increase the threshold value, other terms such as {*elastic, wheat and rice, notion*} will be included in the set of most important terms.

Our preliminary results indicate that centrality entropy is a metric that produces good results when applied to select the most important terms in a semantic network. These terms can be used to summarize the content of the two sentences used in the example.

Given that the terms in the semantic network were selected as the most important ones in the phrase extraction phase described in [5], our method can be used to perform a further optimization by selecting from the terms in the semantic network, those that best summarize the contents of a group of phrases or sentences.

5 Conclusions

We have presented some preliminary results on the usefulness of applying graph entropy to summarize the subject of a group of phrases or sentences. The semantic network used in our experiments was obtained from [5].

Our method's results depend on the structure of the semantic network used. Therefore, in future work we plan to investigate efficient ways to extract semantic networks from documents, additionally to a more extensive set of experiments to evaluate the real potential of our approach, comparing its results with other summarization systems.

Finally, a more extensive analysis of semantic networks using other methods from complex networks is also planned in future work.

6 Acknowledgements

The research leading to the results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 312651.

References

1. Borgatti, S. P.: Identifying sets of key players in a network. In: Computational, Mathematical and Organizational Theory, 12(1): pp. 21–34 (2006)
2. Shetty J. and Adibi, J.: Discovering important nodes through graph entropy the case of enron email database. In: LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery, pp. 74–81, ACM, New York, NY, USA, (2005)
3. Latora, V. and Marchiori, M.: How the science of complex networks can help developing strategies against terrorism. In: Chaos, Solitons and Fractals, 20(1), pp. 69–75, (2003)
4. Ortiz-Arroyo, D.: Discovering Sets of Key Players in Social Networks. In: Abraham, A., Aboul-Ella, H., and Vaclav S. (eds.) In: Computational Social Network Analysis: Computer Communications and Networks, pp. 27-47. Springer, London (2010)
5. Huang, C.,Tian Y., Zhou, Z., Ling, C. X., Huang, T.: Keyphrase Extraction using Semantic Networks Structure Analysis. In: ICDM '06. Sixth International Conference on Data Mining, pp. 275–284. IEEE (2006)

6. Steyvers M., Tenenbaum, J. B.: The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. In: Cognitive Science vol. 29, pp. 4178(2005)
7. Litvak M., Last M.: Graph-Based Keyword Extraction for Single-Document Summarization: In: Coling 2008: Proceedings of the workshop on Multi-source Multilingual Information Extraction and Summarization, pp. 1724(2008)
8. Volchenkov, D., Blanchard, P.: Discovering Important Nodes Through Graph Entropy Encoded in Urban Space Syntax: In: eprint arXiv:0709.4415, ARXIV(2007)