

The Extraction from News Stories a Causal Topic Centred Bayesian Graph for Sugarcane

Brett Drury
ICMC, University of São Paulo
São Carlos, Brazil
brett.drury@gmail.com

Conceição Rocha
LIAAD-INESC-TEC
Porto, Portugal

Maria-Fernanda Moura
EMBRAPA
Campinas – SP, Brasil

Alneu de Andrade Lopes
ICMC, University of São Paulo
São Carlos, Brazil

ABSTRACT

Sugarcane is an important product to the Brazilian economy because it is the primary ingredient of ethanol which is used as a gasoline substitute. Sugarcane is affected by many factors which can be modelled in a Bayesian Graph. This paper describes a technique to build a Causal Bayesian Network from information in news stories. The technique: extracts causal relations from news stories, converts them into an event graph, removes irrelevant information, solves structure problems, and clusters the event graph by topic distribution. Finally, the paper describes a method for generating inferences from the graph based upon evidence in agricultural news stories. The graph is evaluated through a manual inspection and with a comparison with the EMBRAPA sugarcane taxonomy.

Categories and Subject Descriptors

I.2.4 [ARTIFICIAL INTELLIGENCE]: Knowledge Representation Formalisms and Methods; I.2.7 [ARTIFICIAL INTELLIGENCE]: Natural Language Processing

Keywords

Text Mining, Causal Relations, Bayesian Graph, Sugarcane

1. INTRODUCTION

Sugarcane is an important product in Brazil because it is a primary ingredient of ethanol, which is used as a gasoline alternative. It is also a food product. The identification of events / factors that can influence the yield of sugarcane may assist farmers in making informed decisions concerning sugarcane cultivation. A natural media for this information is news because it contains "time sensitive" information.

The events that affect sugarcane can have a number of interactions. For example, one of the side effects of soy produc-

tion is deforestation which can reduce rainfall which reduces the sugarcane harvest. These interactions can be modelled in a Causal Bayesian Network (CBN). A CBN can be used to make inferences about the effect of new information upon a given event. Manual construction of CBNs can be a time consuming process requiring skilled labour.

This paper will describe a process to automatically construct a topic centred Causal Bayesian Network from news stories that models the interaction of events that influence the sugarcane harvest. The paper will describe the: construction, probability calculations, automated query generation as well as graph evaluation. The remainder of the paper will discuss: 1. Related Work, 2. Proposed Strategy, 3. Causal Relations, 4. Graph Construction, 5. Probability Calculation and 6. Evaluation. The results demonstrate a relationship between "graph quality" and accuracy of the relations encoded into the edges. In addition when measured against a pre-existing resource the final graph has a "good" representation of the domain.

2. RELATED WORK

The literature search covered the following areas: causality, causal relations, causation representation in graphs, and the automatic construction of Bayesian Graphs from text.

There are a number of differing opinions on the nature of causation. This paper follows the work of [15] that states that the building blocks of causation expressed in text are events. Causation in text is often referred to as causal relations and typically contain: a cause event, causal linkage (verb or continuation) and an effect event. The cause event is "wholly or partially responsible for the effect event" [1]. Strategies for extracting causal relations can be grouped into two main strategies: 1. supervised learning and 2. linguistic. The supervised learning techniques are predicated upon sequential classifiers such as Conditional Random Fields (CRF). The CRF is trained upon labelled data, and identifies start and end points of causal relations as well as their cause and effect events [8]. The linguistic methods rely upon lexical resources such as WordNet to provide "causal clues". The causal clues are then used in pre-defined extraction patterns to extract causal relations [5].

The interactions of causal events can be modelled in a directed graph where events are nodes, and edges represent the flow of causation between events. A node's parents are

its causes and its children are its effects. A specific type of graph, Bayesian Network, can model causation because it allows the representation of the likelihood of an event or a combination of events of producing a specific effect.

There are a small number of papers that describe the automatic construction of Causal Bayesian Networks (CBN) from text. The construction media for all of these papers are causal relations. The causal relations are then aggregated into a CBN. An early example was proposed by [14] who used linguistic resources and ad-hoc rules to create Bayesian Networks. [11] used a "literature mining" approach to create a causal Bayesian Network. The causal relations were used to "enrich" an existing Bayesian Network. A similar approach was used by [12] who developed a system called AutoBayesian.

It should be noted that the papers surveyed for Bayesian Network construction often did not detail fundamental issues such as: 1. graph cycles resolution, 2. probability calculations or 3. query generation. In addition the previous research did not seem to represent specific domains and used only English texts.

3. PROPOSED STRATEGY

The proposed strategy has two distinct parts: construction and query / inference generation. The proposed graph construction strategy creates a topic centred CBN by initially creating an event graph from causal relations extracted from Brazilian-Portuguese agricultural news stories. The event graph is: 1. pruned to remove irrelevant events, 2. equivalent nodes are merged to reduce sparsity, and 3. cycles are removed. The graph is clustered using topic distribution of event (node) titles. The probability of event occurrence and the likelihood of the sentiment orientation (positive, negative or neutral) of a future event in the presence of previous events is calculated.

The proposed query strategy generates evidence for a query by fitting new information against the centroids of the node cluster. A target cluster against which inferences will be made is identified through keywords. A probability distribution of sentiment: positive, negative or neutral, is returned for a given query.

3.1 Corpus Acquisition

The corpus for this paper currently has 95585 stories which have an average of 407.60 ± 343.08 words per document. The stories were obtained from agricultural news resources such as "Noticias Agrícolas" and "Canal do Produtor", and were written in Brazilian-Portuguese. The documents in the corpus spanned from 1997 - 2016 The Corpus is available on request from the first author.

4. CAUSAL RELATIONS

The causal relation extraction process has two steps: causal relation extraction, and relation processing. The extraction process extracts causal relations, and labels the cause and effect events. The relation processing step removes irrelevant information.

The causal relation extraction strategy (CRE) relies upon lexical clues. The lexical clues are pre-labelled causal verbs and continuations that are available from [2]. The lexical clues provide a label and a sentiment orientation. For example, the verb, *causar* (to cause), has a "cause" label and

a neutral sentiment orientation.

The extraction process relied upon the following linguistic extraction patterns: 1. *NP V NP* (Pattern 1) and 2. *NP C NP* (Pattern 2), where *NP* is a noun phrase, *V* is a causal verb and *C* is a causal continuation. A causal continuation is a continuation such as *because* that is a causal linker between cause and effect events.

The flow of causation for Pattern 1 is left to right where the first *NP* is a cause event and the second *NP* is the effect event. In Pattern 2 the flow of causation is right to left. The flow of causation in Pattern 1 can be reversed in the presence of a preposition before or after the causal verb. An example of a causal relation extracted by Pattern 1 is: *chuva* (rain) *causa* (causes) *problemas com a safra* (problems with the harvest), and Pattern 2 is: *problemas com a safra* (problems with the harvest) *por conta da* (because of) *chuva* (rain). In each case the relation is an example, of a neutral statement (rain) causing a negative event (problems with the harvest).

4.1 Causal Relation Processing

A problem with graphs created from text is sparsity. Sparse graphs contain relatively large number of nodes compared to edges. Graphs constructed from text are prone to sparsity because of the ability of natural language to represent the same information in numerous different ways. The aforementioned causal relations extracted by Patterns 1 and 2 illustrate this property of text.

An initial step to limit sparsity was to remove: 1. stop-words (*de*, *da*, etc), 2. location information, 3. time information and 4. named entities. The stop word list was an amalgam of a number of sources including: NLTK, Linguatca (<http://www.linguatca.pt/chave/stopwords/>) and ranks (<http://www.ranks.nl/stopwords/portugese>). The named entity and location information was detected using an approach proposed by [13]. The time information was identified by a hard coded list of time related words.

In addition the text was normalized. Text normalization removed accents and replaced the cedilla, *ç*, with a *c*. In addition all text was changed to lowercase. In the aforementioned examples both cause events would be normalized to *problemas safra*. The nodes that were normalized to the same node title were merged, along with their edges.

5. GRAPH CONSTRUCTION

The graph construction process initially built an event graph. Structural and relevance issues were then resolved. The event graph was then clustered to produce a topic centred graph.

5.1 Event Graph Construction

The event graph was constructed by converting the *NP* into nodes, and the edges between the nodes are created from a causal linker, which is either a: causal verb or continuation. The edge information contained: 1. the frequency of edge assertion in the corpus, 2. the latest date the edge has been asserted and 3. the frequency of the causal linker. The direction of the edge represents the flow of causation, i.e the parent of a node is one of its causes.

5.2 Causation Generalization

Causation generalization is the process of grouping together nodes that represent the same event. The equivalent

nodes and their edges are merged. This process reduces sparsity by reducing the node count, while retaining the same number of edges. There are two main approaches for causation generalization: 1. linguistic with ad-hoc rules, that rely on lexical resources such as WordNet [14], and 2. similarity measures [3]. Approach 1 was not possible because lack of: 1. available suitable linguistic resources for Brazilian-Portuguese and 2. comprehensive set of rules. The second approach, that proposed a combination of string and topic distribution similarity to identify equivalent nodes was chosen. The similarity thresholds and number of topics suggested by [3] were used.

The merging process was a recursive process that merged nodes by comparing node titles. Initially a node is compared with all of the remaining nodes. Node pairs that passed the similarity threshold proposed by [3] were merged together along with their edges. This process was repeated for all of the nodes in the graph. The merged nodes kept the original node titles (lists of node titles). The merged nodes were then compared with each other by means of the node title lists, and if they passed the aforementioned threshold the merged nodes were merged. This process continued until no more nodes can be merged. Examples of the merge candidates can be found in [3].

The complexity of the node similarity algorithm is $O(n^2)$, because each node was compared with all remaining nodes. The conversion of the node title to a probability distribution and the string similarity measure was a computationally expensive operation. A naive implementation took in excess of 48 hours to run. A number of measures were taken to ensure that the graph was created in a reasonable period of time. These were: 1. caching node probability distributions, and 2. caching previous calculation, consequently the first node in the algorithm was compared to $n - 1$ nodes, where n is the number of nodes in the graph, the last was compared to $n - n$ nodes because all the calculations were completed in previous iterations. Finally, the string comparison was computationally less expensive than the topic similarity measures, consequently the string similarity was executed first, and if the similarity measure was low enough to ensure that the aforementioned threshold could not be reached, the topic similarity measure was not calculated. These optimization drastically reduced the execution time.

5.3 Removal of Irrelevant Nodes

The graph contained a large number of nodes that had no relationship with sugarcane, and consequently this characteristic may inhibit robust inference in the final Bayesian Graph. Relevant nodes were identified by identifying direct paths to nodes that describe an event concerning sugar or a related product such as ethanol. Nodes that represented a sugarcane related event were identified using keywords supplied by a domain expert. These words were: *cana-de-açúcar* (sugarcane), *açúcar* (sugar), *etanol* (ethanol), *canaviais* (sugarcane), *sucroalcooleiro* (sugarcane), *alcooolquímica* (alcohol chemistry) and *álcool* (alcohol). It is accepted that is an ad-hoc process, but there are no authoritative keywords list for sugarcane. Each node that did not contain a "keyword" was tested to see if it had a direct path to one of the nodes that had a "keyword". A node that did not have a direct path was removed.

5.4 Graph Cycles Resolution

The graph had cycles. A cycle is a group of nodes that are linked in a closed chain, consequently in a cycle an event becomes a cause of one or more of its own causes. Cycles were identified with Johnson's algorithm [7]. The cycles were resolved by comparing the dates of the in-edges with out-edges of an node. If an in-edge was asserted after an out-edge it was deleted. This process continued until the cycle was removed. The rational behind this cycle resolution strategy is that a cause of a event could not occur after an effect of the event, consequently the temporal order of cause and effect was incorrect. The removal of the cause that occurs after the effect resolves the temporal ordering and the graph cycle.

5.5 Clustering

The event graph at this stage had in excess of 4000 nodes. A CBN generated from this event graph would not be able to provide responses to queries quickly and it may not be possible to generate probability tables for nodes with a large number of in-edges because the number of possible combinations grows exponentially. An experiment that queried a BN created from this graph was not able to produce an inference because the queries never terminated.

A solution was to generalize about the events in the graph by clustering the nodes in the event graph. There are a number of differing clustering techniques, and for these experiments the following were used: 1. DBScan [4], 2. K-means [6] and 3. Spectral over nearest neighbour [10]. These were selected because they represented a number of differing techniques.

The distance measure was a topic distribution similarity between node titles. The topic distribution of a node title was inferred using a topic model that was trained using LDA, and 2000 topics, a threshold that was identified by experiments for [3] as an optimal level for causation generalization. The topic distribution similarity was calculated using a Hellinger Distance, which can be represented as: $H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2}$, where p_i is a probability for state i from probability distribution p and q_i is a probability for state i from probability distribution q .

Clustering techniques 2 and 3 require a pre-set number of clusters. The number of clusters were calculated by performing of number of clustering experiments with $k \geq 2$ and $k \leq 100$, with the number of clusters increased by 1 with each iteration. The quality of the clusters was calculated with Davis-Bouldin Index (DBI). DBI favours "tight" clusters where the centroid of each cluster is separated by a large as distance as possible. The DBI can be calculated as $DBI = \frac{1}{N} \sum_N D_i$, where N is the number of clusters, and $D = \max(R_{ij})$ where i and j represent two different clusters, and $R = \frac{S_i + S_j}{M_{ij}}$ where S is the spread of a cluster and M is the distance between the centroids of $Cluster_i$ and $Cluster_j$. The optimal number of clusters for techniques 2 and 3 was 30. DbScan identified the optimal number of clusters as 24. Centroids in this paper were assumed to be the event node that had the smallest average Hellinger distance to the remaining event nodes in the same cluster.

Edges between clusters were calculated by locating nodes in disparate clusters that were linked by an edge. The edges were calculated by aggregating the edges of nodes in different clusters. The edge information described on page was preserved and aggregated.

Cycles in the clustered event graph could not be removed with the date strategy because of the aggregation of edges. In this case a simple count of the edge frequency was used, and the least frequent edge was removed.

6. PROBABILITY CALCULATION

A Causal Bayesian Network (CBN) requires two types of probability calculation: 1. event cluster probability and 2. event cluster probability in the presence of one or more parent (causes). In the CBN created for this paper, each node had three attributes: positive, negative and neutral. The probability of these attributes would be the response to a query given specific evidence.

6.1 Event Cluster Probability

The probability of an event in a cluster occurring was computed by identifying the number of days a related news story to the cluster occurred in a given period of time. It can be represented as $P(C_n) = \frac{NDE}{TD}$, where $P(C_n)$ is the probability of an event in a cluster occurring, NDE is the number of days that a related news story directly related to Cluster C_n was published and TD is the total number of days in the time period.

A related news story was identified by comparing the topic distribution of the centroid of each cluster with a topic distribution of given news story on a specific day. The comparison was by Hellinger distance. It was assumed that the event cluster with the centroid closest to the news was to have occurred that day.

6.2 Probability Tables

As stated earlier this Causal Bayesian Graph not only represents the flow of causation from one event cluster to another, but the type of causal relationship. The type of causal relationship could be: negative, neutral or positive. Consequently, each node cluster had these sentiment attributes.

Each node cluster had a probability table that stated the likelihood of each state in the presence of parent cluster nodes (causes). A simple example with two causes: rain and presence of pests, is presented in Table 1.

Rain	Pests	Prob. Neg.	Prob. Pos.	Prob. Neu.
1	1	1.0	0.0	0.0
0	1	0.5	0.0	0.5
1	0	0.5	0.0	0.5
0	0	0.0	0.5	0.5

Table 1: Simple Probability Table

The probability calculation relied upon the sentiment of the relationship between parent and child clusters. The sentiment of relationship was calculated by identifying sentiment in the causal linker for resultative causal verbs or the effect event in the event graph. The sentiment indicators for resultative verbs from [2] were used. Sentilex [9] was used to identify the sentiment of an effect event. This information was added to the cluster edges. At this time the cluster edges have the frequency of: negative, positive and neutral edges between individual clusters.

A probability was calculated for the sentiment of each edge for each specific cluster by dividing the frequency of a sentiment in-edge by the total number of in-edges. For

example, the likelihood of a negative causal relationship between cluster n and cluster n_1 can be represented by $P(C_{n(neg)n_1}) = \frac{TNE}{TE}$, where $P(C_{n(neg)n_1})$ is the probability of a negative causal relationship, and TNE is a count of negative edges from Cluster n to Cluster n_1 , and TE is the total number of in-edges to Cluster n_1 . The probability for each sentiment state was calculated by summing the probabilities of each state from the combinations of parent nodes. It is possible the probability may exceed 1. This is because the causes are independent of each other. For example, 1. a serve drought and 2. a large infestation of pests, may each independently have a probability of 1.0 of having a negative causal relationship with the sugarcane harvest. In combination they may have a serve impact on the harvest, but they can't increase the probability of a negative outcome to more than 1. A limit of 1 was placed on any particular probability calculation for any given state.

6.3 Causal Bayesian Graph

The graph structure and probability information was then exported to the JSON format specified by libpgm (<http://pythonhosted.org/libpgm/>), a python library for Bayesian Network manipulation and querying.

6.4 Query Generation

A query is a method of generating inferences given evidence. Evidence in this case is the state of a random variable (node) which can be either: positive, negative or neutral. Output nodes are random variables against which a query is made to determine the likelihood of a given state. The selection of the output node is arbitrary, but a possible method could be to select output nodes by keywords for terms such as: "harvest" (*safr*) or "production" (*produção*). It should be noted that this a subjective process which is reliant upon the intuition of the user.

Evidence is gathered for the query by the same "fitting process" used in the probability calculation. News stories are gathered for a pre-determined time period. The centroids are fitted against the news story, and the centroids that "win" the fitting process are used as evidence in the query. The "state" of the evidence is calculated by: 1. computing a sentiment categorization for each successfully fitted news story and 2. counting the classifications for each node. The majority sentiment state is determined to be the state for that random variable. For example, in a simplified system with four nodes which are represented by the following centroids: 1. production cost, 2. weather, 3. government policies and 4. harvest, and the harvest is the output node, and each of the remaining nodes are fitted to news stories that have the following sentiment categories: negative: 4 stories, positive: 3 stories and neutral: 2 stories. A hypothetical query would assume that the evidence: nodes 1 to 3 would be in the negative state.

7. EVALUATION

The evaluation was not a simple process because of a lack of competing strategies and comprehensive resources for sugarcane. The evaluation relied upon a partially complete resource: 1. EMBRAPA Sugarcane Taxonomy (www.embrapa.br) and 2. manual evaluation. It should be noted that manual evaluations, particularity for causal relations can be subjective.

The EMBRAPA Sugarcane Taxonomy is a simple taxonomy that is used by EMBRAPA to demonstrate the major influences upon the sugarcane crop. Terms were extracted from the taxonomy, and stop words were removed. The terms were treated as a 'bag of words'. The same process was applied to the initial event graph. An intersection of terms in the initial event graph with the EMBRAPA taxonomy was calculated. There were: 69.48% of the terms in the EMBRAPA Sugarcane Taxonomy where in the initial event graph.

The manual evaluation evaluated: the event graph and the graphs created by the clustering technique. The event graph was evaluated by three experts (2 agricultural economists and 1. agricultural researcher) supplied by EMBRAPA. The experts evaluated the 495 most frequently asserted edges. The experts were supplied with a cause event (CE) and an effect event (EE), and had to decide if the CE could cause the EE. There were three choices: 1. correct, 2. incorrect or 3. unsure. The decision was by majority vote. The confidence interval, at 95%, for the total edge population was 1.18. The results were: Correct: $42.58\% \pm 1.18$, Incorrect: $4.17\% \pm 1.18$ and Unsure: $52.85\% \pm 1.18$.

The cluster graphs were evaluated by a single expert (agricultural researcher). The expert was supplied with edges in the aforementioned format. The CE and the EE were represented by the most frequent words in the cluster. The expert had the same one of three choices to make. In addition to the manual evaluation results a "graph quality" (average degree centrality-ADC) measure was calculated. The higher the score the larger number of edges to number of nodes. The results are in Table 7.

Tech.	Numb. Nodes	ADC	Correct (%)	Incorrect (%)	Unsure (%)
Spect.	30	10.62	40.48	19.04	40.48
DB.	24	13.23	49.69	17.28	33.02
K-m.	30	24.17	65.34	9.33	25.33

Table 2: Graph Evaluation Results, where Spect.= Spectral, DB. = DBScan and K-m = K-means

The results show that ADC is a good indicator of a manual evaluation. The higher the ADC for a graph the lower the number of incorrect and unknown edges in the manual evaluation.

8. CONCLUSION

This paper shows a technique that allows the automatic construction of a Causal Bayesian Network that represents the sugarcane domain. The graph demonstrates the interconnections with the sugarcane domain as well as the sources of sentiment. The evaluation demonstrates that the graph contains terms relevant to the sugarcane domain, and that the causal relationships are "reasonable". The evaluation, in particular the manual could be seen as subjective, but there was no authoritative alternative evaluation techniques. Consequently, the future work will concentrate upon testing the inferences made by the various networks by trading sugar futures.

A sugar future has a price for delivery of sugar at a given date in the future. This price represents all known information about sugar. A trading strategy that is able to trade

successfully and beat the market, will have access to information that is not incorporated into the price. It is the hypothesis of this work that the Causal Bayesian Graph has this information. The evaluation will gather evidence from news stories that are published before the exchange is open, and make inferences about the direction of the market from this evidence. The inferences will then be incorporated into a traditional trading strategy that: 1. shorts on positive sentiment (prices will go down with increased sugar supply) and 2. buys on negative sentiment (prices go up with restricted supply).

In addition to improved evaluation, future work will evaluate the influence of: 1. structure learning, 2. alternate cycle resolution techniques and 3. link prediction, on the ability of the network to make robust inferences.

9. ACKNOWLEDGMENTS

This research was partially supported by Grant Numbers: 11/20451-1, 2011/22749-8, 15/14228-9 (from FAPESP), 302645/2015-2 (from CNPq, Brazil); and by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF) within project "NORTE-01-0145-FEDER-000020".

10. REFERENCES

- [1] Bengt Altenberg. Causal linking in spoken and written English. *Studia Linguistica*, 38(1):20–69, 1984.
- [2] Brett Drury, Paula C.F. Cardoso, Janie M. Thomas, and Alneu de Andrade Lopes. Lexical resources for the identification of causative relations in portuguese text. In *Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish*. DBDComp, 2014.
- [3] Brett Drury, Jorge Valverde-Rebaza, and Alneu de Andrade Lopes. Causation generalization through the identification of equivalent nodes in causal sparse graphs constructed from text using node similarity strategies. In *Proceedings SIMBIG*, volume 1318 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2015.
- [4] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.
- [5] Roxana Girju. Automatic detection of causal relations for question answering. In *Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering - Volume 12*, pages 76–83. Association for Computational Linguistics, 2003.
- [6] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.
- [7] Donald B Johnson. Finding all the elementary circuits of a directed graph. *SIAM Journal on Computing*, 4(1):77–84, 1975.
- [8] S. Mehrabi, A. Krishnan, E. Tinsley, J. Sligh, N. Crohn, H. Bush, J. Depasquale, J. Bandos, and M. Palakal. Event causality identification using conditional random field in geriatric care domain. In *Proceedings - ICMLA 2013*, volume 1, pages 339–343, 2013.

- [9] Paula Carvalho Mário J. Silva and Luís Sarmento. Building a sentiment lexicon for social judgement mining. In *International Conference on Computational Processing of Portuguese (PROPOR)*, 2012.
- [10] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.
- [11] S. Raghuram, Yuni Xia, M. Palakal, J. Jones, D. Pecenka, E. Tinsley, J. Bandos, and J. Geesaman. Bridging text mining and bayesian networks. In *Network-Based Information Systems*, pages 298–303, 2009.
- [12] Sandeep Raghuram, Yuni Xia, Jiaqi Ge, Mathew Palakal, Josette Jones, Dave Pecenka, Eric Tinsley, Jean Bandos, and Jerry Geesaman. Autobayesian: developing bayesian networks based on text mining. In *Database Systems for Advanced Applications*, pages 450–453, 2011.
- [13] C. Rocha, A.M. Jorge, M. Oliveira, P. Brito, J. Gama, and C Pimenta. From entity extraction to network analysis: a method and an application to a portuguese textual source”. Technical report, Working Paper number 32, Edições Húmus & OBEGEF - Observatório de Economia e Gestão de Fraude, Porto, Portugal., 2014.
- [14] Olivia Sanchez-Graillet and Massimo Poesio. Acquiring bayesian networks from text. In *LREC*, 2004.
- [15] Zeno Vendler. Causal relations. *The Journal of Philosophy*, 64(21):704–713, Nov 1967.