Towards the Automatic Extraction of Policy Networks using Web Links and Documents

Theodosis Moschopoulos, Elias Iosif, *Student Member, IEEE,* Leeda Demetropoulou, Alexandros Potamianos, *Senior Member, IEEE*, and Shrikanth Narayanan, *Fellow, IEEE*

Abstract—Policy networks are widely used by political scientists and economists to explain various financial and social phenomena, such as the development of partnerships between political entities or institutions from different levels of governance. The analysis of policy networks demands a series of arduous and time consuming manual steps including interviews and questionnaires. In this paper, we estimate the strength of relations between actors in policy networks using features extracted from data harvested from the web. Features include webpage counts, outlinks, and lexical information extracted from web documents or web snippets. The proposed approach is automatic and does not require any external knowledge source, other than the specification of the word forms that correspond to the political actors. The features are evaluated both in isolation and jointly for both positive and negative (antagonistic) actor relations. The proposed algorithms are evaluated on two EU policy networks from the political science literature. Performance is measured in terms of correlation and mean square error between the human rated and the automatically extracted relations. Correlation of up to 0.74 is achieved for positive relations. The extracted networks are validated by political scientists and useful conclusions about the evolution of the networks over time are drawn.

Index Terms—policy networks, social networks, relatedness metrics, similarity metrics, web search, policy actors, link analysis

1 Introduction

Modern democratic governance reflects a shift away from the traditional notions of hierarchy towards more co-operative forms of public policy making. Within this context, the term 'network' is often used to describe clusters of different types of actors who are related in the political, social and economic spheres. In [1], the term 'policy network' is defined as "a cluster of actors, each of which has an interest, or 'stake' in a given policy sector and the capacity to help determine policy success or failure." Political scientists use policy networks to investigate social and financial phenomena, especially, the evolution of relations between actors and the effectiveness of policies towards the formation of partnerships among actors. This is achieved by reference to the structure of networks in a given policy field at different phases of policy development (planning, implementation and evaluation). A policy network can be described by its actors, their linkages and its boundary [2]. Policy networks consist of a set of public and private actors and a number of linkages between them that serve as channels for communication and the exchange of information, expertise, trust and other policy resources. The network boundaries are not primarily determined by formal institutions but rather by functional relevance and structural embeddedness [2].

Typically, policy networks are identified through a manual procedure performed by experts. Identifying actors, links and boundaries, i.e., analyzing a policy network's structure, requires refined techniques and extensive and time-consuming manual collection of data through interviews and questionnaires. During the manual identification of networks, many subjective factors may be present, since this procedure relies strongly on the human subjects that participate in the interviews. Such factors include personal opinions, the person's willingness to participate and even cultural issues. Overall, policy network identification currently requires a "large scale investment" that does not always "lead to breathtaking empirical and theoretical results" [2]. When lacking the resources for data collection and network analysis, political scientists often revert to qualitative analysis or construct the network topology using their intuition, significantly limiting the evidence-based validation of their results.

In this work, we propose an algorithm for the automatic extraction (or validation) of policy networks using information collected from the web. Specifically the degree of relatedness (strength of link) between policy actors in a network is computed using three types of features on documents or snippets downloaded by web search engines, namely: 1) the frequency of co-occurrence for each pair of actors (in web documents), 2) the lexical contextual similarity between snippets of web documents in which the actors appear, and 3) the co-occurrence of hyperlinks present in web documents that contain the actors. For each type of feature and for their combinations, a variety of similarity metrics are used

- *T. Moschopoulos, E. Iosif, and A. Potamianos are with the Dept. of Electronics and Computer Engineering, Technical Univ. of Crete, Chania 73100, Greece, Email: {tmos,iosife,potam}@telecom.tuc.gr.*
- *L. Demetropoulou is with the Dept. of Environmental Engineering, Technical Univ. of Crete, Chania 73100, Greece, Email: ldemetro@gmail.com.*
- *S. Narayanan is with the Dept. of Electrical Engineering Univ. of Southern California, Los Angeles, USA, Email: shri@sipi.usc.edu.*

in order to estimate the link strength for each pair of actors. The proposed algorithm is not intended to substitute expert knowledge, but rather it should be viewed as a low-cost, semi-automated computational tool that can significantly support and enhance policy network analysis. The proposed method aims to be efficient and reduce human biases.

A policy network can be considered as a type of social graph in which the nodes represent the actors involved in a given policy field, while their relations are represented by the edges. In social networks, nodes usually correspond to persons and edges represent relations among them built on a ground of mutual understanding, which can take several forms, such as friendship and co-authorship. In policy networks, actors can be organizations or even groups or unions of variable size and degree of formal organization. The relations among such actors usually signify the development of partnerships rather than a lax social relation. Furthermore, relations in policy networks depend on external factors, such as economic policies and funding at the local, national or supra-national level. Relations between policy actors (much like economic actors) can also be antagonistic rather than co-operative, or follow a more complex pattern of both co-operation and competition (sometimes referred to by economists as coompetition). Often policy networks are studied at their infancy when the links between actors are just emerging and might not be directly observable through common action or direct communication. All these subtleties imply that established features and algorithms for social network analysis might not be directly applicable to policy network extraction.

To our knowledge, this is the first comprehensive research effort towards the automatic extraction of policy networks. A variety of features extracted from web documents or snippets are proposed for estimating the relations between policy actors; these features are motivated by recent research in the fields of information retrieval and natural language processing. Another important contribution is that the proposed features and algorithms are evaluated against actual policy networks identified by expert political scientists. It is shown that the automatically extracted policy networks are capable of capturing the main relations between policy actors and are in broad agreement with networks built manually. In some cases, the automated method is shown to offer a deeper understanding of the relations between actors, especially as this pertains to the evolution of a network over time and policy outcomes.

The rest of this paper is organized as follows. In Section 2, we briefly present prior work conducted in the fields of computational methods in politics, social network extraction and semantic relatedness computation between words or terms. In Section 3, we describe the metrics used to measure the degree of relatedness between actors, namely, text-, hit- and link-based metrics computed from web documents or snippets. The proposed algorithm used to estimate the link strengths between actors and construct the policy network is detailed in Section 4. In Section 5, we describe the experimental methodology, the policy networks used for evaluation purposes and the evaluation metrics. In Section 6, the performance of the proposed relatedness metrics is presented, and the resulting policy networks are visualized and compared with the networks identified by experts. The results are interpreted in Section 7, while in Section 8 we conclude our work and offer directions for future research.

2 Related Work

The use of computational analysis of large amounts of data by political analysts has flourished in the past few decades facilitating the study of group connections. Two computational methods have been widely used in political science namely text analysis and social network analysis. More specifically, political analysts have used text mining to analyze electoral campaigns, identify voters' profiles, determine ideological positions, code political interaction and detect political conflict's content [3], [4]. Textual data mostly consist of political manifestos, but transcribed speeches and political statements are also used. In [5], [6], the WORDSCORES system is proposed that extracts economic and social policy dimensions based on word frequencies from manifestos. Similarly, the WORDFISH system [7] mines policy dimensions of parties and estimates their uncertainty over time using word frequencies from manifestos. Opinion mining is an active research area that is also relevant to political scientists. Opinions can be mined from text, blogs or from transcribed speech, e.g., [8]. Important research questions include the selection of lexical features (words and terms), the scores assigned to each term, as well as, the computational model used to combine the evidence, e.g., [9]. In [10], lexical features are combined with social information extracted from blog to classify political sentiments during the 2008 U.S Presidential election. In [11], opinion mining techniques (including lexical feature selection) are applied to the analysis of political conflicts.

Regarding social network analysis, political analysts have used network analysis to study formal and informal interactions. Policy networks extraction can be considered as a special type of social networks extraction, an active research area. The major steps in the extraction of social networks are *relation identification* [12], [13], [14], [15], i.e., to identify whether two actors are related, *relation labeling* [16], [17], assign an existing relation to a category and the *estimation of strength* [18], i.e., identify whether an existing relation is weak or strong. The most common feature used to identify a relation is the frequency of co-occurrence of the related pair of terms in web documents, but other features, such as, lexical context, keyphrases, log files and e-mail information are also used. In [12], a network of experts with respect to certain topics is constructed by estimating similarity of users according to the frequency of co-occurrence of their names in web documents. Similarly, in [19], [20], web co-occurrence of entities is used for creating a network of research communities. In [13], [14], [15], web co-occurrences are used for the extraction of social network of conference participants; a machine learning approach is used to classify each relation from a predefined set of relation types. In [16], automatically extracted keyphrases are used in order to describe the relations between entities. E-mail contacts are used as features in [21], [22] to create personal and professional relationship networks. In [23], social networks are extracted and updated over time using monolingual or multilingual news from articles. In [17], social networks of entities are extracted using posts from the blogosphere and the lexical context of entity pairs is used to automatically label the relations. In [24], quoted phrases from novels are used to extract the social network of the novel's characters. In [25], the log files of shared workspaces are used to extract user-oriented and object-oriented social networks. Some of the aforementioned systems also apply generic relatedness metrics used in other fields, such as Natural Language Processing and Information Retrieval. For example, metrics that are based on web co-occurrence and lexical features are extensively used for the computation of semantic similarity between words and terms [26], [27], [28], [29].

3 Relatedness metrics

In this section three types of relatedness metrics for the computation of the ties between political actors are presented, namely: 1) page-count-based, 2) text-based, and 3) link-based metrics. Each metric explores different features, capturing different perspectives of web information. Page-count-based metrics use co-occurrence of the (name or acronym of the) actors in web documents or snippets. Web co-occurrence captures a variety of relations among terms ranging from similarity to association. Text-based metrics compute lexical similarity between the context in which the political actors appear in web documents or snippets. Contextual lexical similarity is a popular metric to measure semantic similarity; in our case, we expect political actors with similar function to share high lexical similarity scores. For link-based metric computation, we examine the number of shared hyperlinks (outlinks) among the web documents that contain the terms of interest. Common outlinks indicate that political actors share the same interests or point to common links in the network, thus are directly or indirectly related. Finally, we propose linear combinations of the normalized values of the three metrics.

3.1 Page-count-based metrics

The degree of relatedness between actors is estimated as an association ratio that is a function of the co-occurrence frequency of actors[1] in web documents. The assumption here is that *related actors tend to co-occur in web documents*; co-occurrence implies that both actors deal with common policy issues or serve similar policy functions. The set of all documents indexed by a search engine is denoted as $\{D\}$, and the cardinality of this set is denoted as $|D|$. For the set of documents that are indexed by an actor a_i we use the notation $\{D_{a_i}\}$. In similar fashion, the set of documents that contain two actors, a_i and a_j, is denoted as $\{D_{a_i,a_j}\}$ with cardinality $|D_{a_i,a_j}|$. We employ the four page-count-based similarity metrics defined next.

Jaccard coefficient: Generally this coefficient computes the similarity between sets. In our case, we consider the sets of web documents that are indexed by the actors of interest. The Jaccard coefficient S_J^P between actors a_i and a_j is defined as follows:

$$S_J^P(a_i,a_j) = \frac{|D_{a_i,a_j}|}{|D_{a_i}| + |D_{a_j}| - |D_{a_i,a_j}|}. \quad (1)$$

For identical actors the Jaccard coefficient assigns the maximum similarity score of 1. For unrelated actors a_i, a_j that never co-occur the Jaccard coefficient is 0.

Dice coefficient: This coefficient is closely related to the Jaccard coefficient and it is defined as:

$$S_D^P(a_i,a_j) = \frac{2|D_{a_i,a_j}|}{|D_{a_i}| + |D_{a_j}|}. \quad (2)$$

As before, (2) is equal to 1 and 0, for absolute similarity and dissimilarity, respectively.

Mutual information: Assuming that $|D_{a_i}|$, $|D_{a_j}|$ are random variables, then their pointwise mutual information reflects the dependence between the occurrence of a_i and a_j, as follows [28]:

$$S_I^P(a_i,a_j) = \log \frac{\frac{|D_{a_i,a_j}|}{|D|}}{\frac{|D_{a_i}|}{|D|}\frac{|D_{a_j}|}{|D|}}. \quad (3)$$

For identical actors the mutual information equals an unbounded positive value. If two actors never co-occur, (3) is undefined, and their similarity is 0.

Google-based semantic relatedness: The "normalized Google distance" is another page-count-based similarity metric that was proposed in [30], [31], defined as follows:

$$S_R^P(a_i,a_j) = \frac{\max\{\log|D_{a_i}|, \log|D_{a_j}|\} - \log|D_{a_i,a_j}|}{\log|D| - \min\{\log|D_{a_i}|, \log|D_{a_j}|\}}. \quad (4)$$

This metric is a dissimilarity measure, i.e., as the distance between two actors increases the metric takes smaller values. The scores assigned by (4) are unbounded, ranging from 0 to ∞. In [29], a variation of the normalized Google distance was used, proposing a bounded similarity measure called "Google-based semantic relatedness", defined as:

$$S_G^P(a_i,a_j) = e^{-2S_R^P(a_i,a_j)}, \quad (5)$$

1. Henceforth, by actor we mean all names, terms or acronyms that are used to refer to a policy actor in web documents or snippets.

$S_R^P(a_i, a_j)$ is computed according to (4). The Google-based semantic relatedness is bounded in [0,1].

3.2 Text-based metrics

The proposed text-based metric computes the strength of relation between actors by examining the lexical context in web documents where such actors are mentioned. The fundamental assumption here is that related actors have similar syntactic, semantic and topical features, e.g., if two actors share political activities it is expected that such activities will be mentioned in the lexical vicinity of the actors. In order to extract the lexical features for actor a_i, text-based metrics apply a contextual window W (containing W words preceding and W words following the actor), i.e.,

$$[f_{W,L} \;\; ... \;\; f_{2,L} \;\; f_{1,L}] \;\; a_i \;\; [f_{1,R} \;\; f_{2,R} \;\; ... \;\; f_{W,R}],$$

where $f_{j,L}$ and $f_{j,R}$ represent the j^{th} feature (in this case, word) that exist to the left and to the right context of a_i, respectively. Given a fixed value of W, a feature vector for a_i is built as $V_{a_i,W} = (v_{a_i,1}, v_{a_i,2}, ..., v_{a_i,N})$, where $v_{a_i,j}$ is a non-negative integer and W is the context window size. The feature vector has N elements, where N is the vocabulary size. The feature value f_j corresponds to the occurrence of vocabulary word v_j within the left or right context window W of a_i. The value of $v_{a_i,j}$ can be a function of the frequency of occurrence of v_j in the context of a_i. More specifically the value of $v_{a_i,j}$ can be defined according to (i) binary (B) scheme: $v_{a_i,j} = 1$ if $c(f_{a_i,j}) > 0$, and (ii) logarithm of term frequency (LTF) scheme: $v_{a_i,j} = \frac{\log(c(f_{a_i,j}))}{\log(c(a_i))}$ if $c(f_{a_i,j}) > 0$, where $c(.)$ denotes counts. Note that the value of $v_{a_i,j}$ is set by considering all the occurrences of a_i in the corpus. Once a weighting scheme is selected, the context-based metric S_W^T computes the similarity between two actors, a_i and a_j, as the cosine of their feature vectors, $V_{a_i,W}$ and $V_{a_j,W}$, as follows:

$$S_W^T(a_i, a_j) = \frac{\sum_{l=1}^{N} v_{a_i,l} v_{a_j,l}}{\sqrt{\sum_{l=1}^{N} (v_{a_i,l})^2} \sqrt{\sum_{l=1}^{N} (v_{a_j,l})^2}}, \quad (6)$$

where W is the context window length and N is the vocabulary size. The cosine similarity metric assigns 0 similarity score when a_i, a_j share no common context (completely dissimilar actors), and 1 for identical actors (or actors sharing the same contexts).

3.3 Link-based metrics

In this section, we define a link-based relatedness metric for computing the degree of association between actors. This metric exploits the hyperlinks of the downloaded web documents, usually referred to as "outlinks" [32], as features. It is expected that hyperlinks will point to topically relevant web sites and documents [33]. The link-based relatedness metric assumes that two actors are related to the extent they share common topics of interest as indicated by full or partial match of outlink web addresses. The outlinks are being used in two different forms, the full form where the whole path is specified (excluding the actual document specified in the outlink), e.g., www.ypes.gr/el/MediaCenter/Minister/, or the base form where only the main website address is used, e.g., www.ypes.gr.

For each actor a_i, we consider the set of (full or base) outlinks $\{O_{a_i}\}$ that appear in web documents where this political actor is mentioned. The similarity between two actors a_i and a_j is computed according to the overlap between the members of their outlink sets. For this computation, variations of (5) and (6) are employed.

Google-based semantic relatedness using outlinks (S_G^L): We apply the metric of (5), using the set of outlinks, instead of the document sets. Specifically,

$$S_R^L(a_i, a_j) = \frac{\max\{\log|O_{a_i}|, \log|O_{a_j}|\} - \log|O_{a_i,a_j}|}{\log|O| - \min\{\log|O_{a_i}|, \log|O_{a_j}|\}}, \quad (7)$$

where $\{O_{a_i}\}$, $\{O_{a_j}\}$ and $\{O_{a_i,a_j}\}$ are the set of outlinks for actors a_i, a_j and jointly for both a_i and a_j, respectively, i.e., $\{O_{a_i,a_j}\}$ is the intersection of $\{O_{a_i}\}$ and $\{O_{a_j}\}$. We then normalize S_R^L into S_G^L using (5).

Cosine similarity using outlinks (S_T^L): Alternatively, for each actor a feature vector is built using the members of the set of outlinks. The relation strength between two actors is computed as the cosine of their feature vectors in the same fashion as (6) (here the window size parameter W is irrelevant and is not specified). The feature values can be set according to the weighting schemes defined in Section 3.2.

3.4 Linear fusion of relatedness metrics

Each of the aforementioned metrics uses different features to estimate relatedness, i.e., actor co-occurrence for page-count metrics, lexical contextual similarity for text-based metrics and outlink similarity for link-based metrics. Here we propose to combine these features using late integration, i.e., combine the relatedness scores from the three types of metrics. For linear fusion the composite relatedness score S between actors a_i and a_j is defined as:

$$S(a_i, a_j) = \lambda_P S^P(a_i, a_j) + \lambda_T S^T(a_i, a_j) + \lambda_L S^L(a_i, a_j), \quad (8)$$

where S^P, S^T and S^L refers to the proposed page-count, text and link-based metrics, respectively, and λ_P, λ_T, λ_S are the corresponding weights. Two cases are investigated: equal weights (that sum up to 1) and inverse variance weighting (informative fusion). For informative fusion, the weights for each type of metric set equal to the inverse variance, e.g., $\lambda_P = 1/\sigma_P^2$. The variance is computed across the relatedness scores for all actor pairs and a specific metric.

4 Computing Actor Relations

The estimation of relatedness scores between political actors involves three steps: 1) the actors are lexicalized and acronymized manually, 2) web data is downloaded as required by each of the similarity metrics, and 3) the relatedness metrics are computed using the equations defined in Section 3. Next, we provide a brief description for each of these steps.

Step 0: Lexicalization of actors. A crucial step for the successful extraction of policy network is the derivation of the lexicalized forms that describe each of the actors. Lexicalizations are usually multi-word terms or abbreviations, e.g., actor "Industrial Development Authority" is also lexicalized as "Industrial Development Agency" or abbreviated as "IDA". Using only the official (long) names of actors often returns very few relevant documents (hits), while certain lexicalizations (especially abbreviations) can be overly general, returning many irrelevant documents. In order to tackle both the data sparseness and term ambiguity problems, a number of lexicalized forms (both multi-word terms and abbreviations) is manually selected for each actor in collaboration with political scientists (see also Section 5.2). The regular expression that contains all possible lexicalizations of actor a_i is referred to as $R(a_i)$, while A is the set of actors.

Step 1: Retrieval of web data. Once the set of actors is created we search the web in order to retrieve data, using the Yahoo! Search API[2]. For this purpose we use two different query types: 1) individual (IND), e.g., "a_i", and 2) conjunctive (AND), e.g., "a_i AND a_j". The IND type concerns individual actors, while the AND type requires the co-existence of the two actors in the returned data. We consider three different types of information returned by the search engine: 1) page counts, 2) URLs of web documents, and 3) their corresponding snippets. In order to acquire the outlinks of the web documents, we employ a further downloading step using the returned URLs. The outlinks are extracted using HTML::SimpleLinkExtor[3].

Step 2: Computation of relatedness. Relatedness scores are computed according to the metrics defined in Section 3: 1) page-count-based (S^P), 2) text-based (S^T), and 3) link-based (S^L). For the S^P metrics, we use the page counts that are returned by IND and AND queries. The S^T metric is applied over snippets that are retrieved using either IND or AND queries. The S^L metric operates on the outlinks of documents that are downloaded using IND queries. In addition, the above metrics are linearly combined as in (8).

5 METHODOLOGY

The proposed algorithms and metrics have been evaluated on policy networks that were manually mapped by political scientists using questionnaires and interviews. Documents, snippets and number of hits were mined from the web for all policy actor pairs in these networks and relatedness metrics were computed, as outlined in the previous section. The manually identified and automatically extracted networks were then compared in terms of correlation of relation strength for all actor pairs, as well as, the mean square difference between the scores. Finally, the networks were visualized as graphs and political scientists were asked to compare and evaluate them qualitatively and quantitatively.

5.1 Policy Network Corpus

Two policy networks from the political science literature were used to evaluate our approach. Both networks examine the patterns of adaptation and institutional policy learning, in two EU country members, namely, Ireland and Greece. The networks were extracted through a time and effort consuming manual process based on interviews and questionnaires collected during the Fifth Framework Project ADAPT (EU Enlargement and Multi-level Governance in European Regional and Environmental Policies). The same (translated) questionnaires were used for the analysis of the transformation of regional development policy-making procedures and institution building in Ireland and Greece.

The first network is based on the research conducted by Rees et al. [34], and includes the main governmental and non-governmental political actors involved in regional policy-making in Ireland and specifically in the Mid-West Region. The network consists of 37 public and private actors representing institutions at the local, regional and national levels. Relations among institutions are undirected; thus the network is represented by a 37×37 symmetric matrix. Each matrix element denotes the strength of the relation between the corresponding actors. Not all possible relations were investigated by the political scientists[4]. Each examined relation is rated with a score of "1","2" or "3" corresponding to a weak, medium or strong relation. According to [34], actors in the network matrix were clustered in blocks, so that actors in the same block of matrix have a "positive" relation, while actors of different blocks have a "negative" (antagonistic) relation. The relation strength ranges from "1" to "3" for both positive and negative relations. In our work, we present separately results for positive and negative relations.

The second network is based on the study by Getimis and Demetropoulou [35] that focuses on the South Aegean region in Greece. The objectives of this research are very similar to those of the Irish case. The Aegean network consists of 21 political private and public actors from the local, regional and national levels. As in the Irish case, relations are assumed symmetric and the network is represented by a 21×21 symmetric matrix. Each

2. http://search.cpan.org/~timb/Yahoo-Search-1.11.3/lib/Yahoo/Search.pm
3. http://search.cpan.org/~bdfoy/HTML-SimpleLinkExtor-1.23/lib/SimpleLinkExtor.pm

4. This is the common practice in the political science literature. Only those actor pairs that are judged by the experts to be related are examined formally.

element denotes the strength of relation between the corresponding actors using the same "1" to "3" (weak to strong) scale. Unlike [34], only positive relations among actors have been measured in the Aegean network.

5.2 Experimental Set Up

For the Ireland case study, the network contains 37 political actors and 226 rated actor pairs. In this work, we have focused on a subset of the network containing only 24 actors for which all relatedness metrics can be effectively computed, i.e., each actor generates an adequate number of web hits[5]. In the corresponding 24×24 submatrix there are 85 rated relations corresponding to 19 positive relations (denoted henceforth as "pos") and 66 negative relations (denoted as "neg"). Similarly for the Aegean case study, there are 21 political actors and 145 rated relations in the policy network. Using the same criterion as above, 3 actors were excluded; for the remaining 18 actors we examine the 109 rated relations (all "pos"). The same policy network extraction algorithm was applied to both networks.

As discussed in Section 4, actors might appear with different names or abbreviations (e.g., acronyms) in web documents. An initial list of actor lexicalizations were proposed by political scientists and then refined using web queries. For each actor lexicalization in the initial list, an individual web query was posed and the top-20 returned documents were inspected for alternative wordforms of this actor [6]. Each candidate lexicalization was then tested via a follow-up query to verify the relevance of the returned documents. The goal is to select a list of names, abbreviation and acronyms for each actor that is not overly ambiguous. At the end, each actor name was represented as a regular expression with the list of alternative names connected via OR conjunctions. Despite our best efforts to select a list of unambiguous actor names, there are issues related to organizations in different countries that share the same name or acronym. This is especially true for the case of Ireland where confusion arises with similar names in the US or UK. In order to reduce ambiguity we have included the pragmatic constraint "Ireland" via an AND conjunction in the actor's regular expression.

For the computation of hit-based metrics, we used the returned hit counts from AND and IND type queries using the regular expressions for each actor, as presented above. Similarly for text-based metrics, we used the snippets returned by these AND and IND type queries. Specifically, our search engine was requested to retrieve the 500 top-ranked URLs for each IND and AND query. Snippets (characteristic portions of the document as selected by the search engine containing the actor name) were downloaded for each URL. In our experiments, we report results using the top 100, 200 or 500 snippets for AND or IND query. A window $W = 10$, i.e., ten words to the left and ten words to the right of the actor, was used. Stop words[7] were excluded from the list of contextual features. For the computation of the link-based metrics the base outlinks extracted from the downloaded documents from IND queries were used.

In the experiments that follow, we evaluate the performance of our metrics by keeping all the relations (1,2,3) denoted as '3-levels' or by keeping only the weak and strong relations (1,3) denoted as 'high-low'. The Ireland network includes 19 positive/66 negative relations for '3-levels', and 14 positive/40 negative relations for 'high-low'. The Aegean network includes 109 and 62 relations for '3-levels' and 'high-low', respectively.

5.3 Evaluation Metrics

Let $H = (h_1, h_2, ..., h_M)$ and $K = (k_1, k_2, ..., k_M)$ be the vectors of human rated and automatically computed relatedness scores, respectively, where M is the total number of relations. Scores k_i may be computed by any of the relatedness metrics presented in Section 3 or their fusion. In order to match the range of human ratings all relatedness scores are linearly scaled as follows:

$$e_i = \frac{2(k_i - k_{min})}{k_{max} - k_{min}} + 1, \tag{9}$$

where k_{min}, k_{max} is the min and max scores (for a specific metric), respectively, and e_i is the normalized relatedness score that takes continuous values in [1,3].

To measure the correlation between the human ratings and normalized relatedness scores we use the Pearson correlation coefficient defined as:

$$r_{H,E} = \frac{\sum_{i=1}^{M}(h_i - \bar{H})(e_i - \bar{E})}{\sqrt{\sum_{i=1}^{M}(h_i - \bar{H})^2 \sum_{i=1}^{M}(e_i - \bar{E})^2}}, \tag{10}$$

where $\bar{H}$ and $\bar{E}$ denote the sample mean of H and E respectively, and $E = (e_1, e_2, ..., e_M)$ is the vector of values produced by (9). In addition, the Mean Square Error (MSE) is used to measure the distance between human ratings and normalized relatedness averaged over all investigated relations, as follows:

$$MSE = \frac{1}{M} \sum_{i=1}^{M} (h_i - e_i)^2. \tag{11}$$

Note that MSE values range between 0 and 4.

A widely used measure in social network analysis is the degree of centrality that indicates the importance of

5. For the text-based metrics we have set a requirement that at least 500 web documents should refer to each actor. Although, relations between actors could be inferred with fewer hits, the main goal of this paper is to evaluate the policy extraction methodology and compare the relatedness metrics.

6. We also investigated a semi-automatic approach for the Aegean network for identifying alternative lexicalizations, where all documents were grouped together into a single corpus. Then, the C-value/NC-value term extraction algorithm [36] was applied to the corpus and the top-200 multi-word terms were selected obtaining 40% recall. Specifically, 27 new lexicalizations were identified for all existing actors, while 12 new actors were discovered.

7. For a stop word list for the Greek language see [37].

Ratings	Metric	Correlation	MSE
Page-count-based metrics			
IRELAND (pos/neg)			
3-levels	Jaccard (S_J^P)	0.29/0.28	1.77/1.14
	Dice (S_D^P)	0.29/**0.29**	1.75/1.12
	Mutual Info (S_I^P)	**0.61**/0.09	**0.42**/0.77
	Google (S_G^P)	0.49/0.17	0.69/**0.70**
high-low	Jaccard (S_J^P)	0.30/0.34	2.20/1.38
	Dice (S_D^P)	0.30/**0.35**	2.18/1.36
	Mutual Info (S_I^P)	**0.66**/0.10	**0.54**/1.19
	Google (S_G^P)	0.56/0.19	0.92/**1.08**
AEGEAN			
3-levels	Jaccard (S_J^P)	0.35	0.53
	Dice (S_D^P)	**0.37**	**0.51**
	Mutual Info (S_I^P)	0.24	1.14
	Google (S_G^P)	0.35	0.91
high-low	Jaccard (S_J^P)	0.41	0.55
	Dice (S_D^P)	0.43	**0.53**
	Mutual Info (S_I^P)	0.44	0.81
	Google (S_G^P)	**0.52**	0.61

TABLE 1

an actor in a network [38]. The degree of centrality for each actor a_i is defined

$$DC_{a_i} = \frac{1}{(A-1)} \sum_j w_{i,j}, \tag{12}$$

where A is the number of actors (vertices in the network), and $w_{i,j}$ is the weight (rating) of the relation (edge) between actors a_i, a_j. The degree of centrality is computed for both the original and extracted networks. The two centrality vectors (extracted vs. original) are compared in terms of correlation and MSE using (10) and (11), respectively.

6 Evaluation

Next we present evaluation results for page-count, text and link based relatedness metrics, as well as, their fusion on the Ireland and Aegean corpora. The human rated and automatically extracted relatedness scores are compared in terms of correlation and average mean square error. The differences between the manually created and automatically extracted networks are also visualized using graphs.

6.1 Page-count-based metrics

The performance of the four page-count-based metrics (Jaccard S_J^P, Dice S_D^P, mutual information S_I^P, Google-based relatedness S_G^P) is shown in Table 1 in terms of correlation and average MSE for the Ireland and Aegean policy networks. Results are shown separately for '3-levels' (all pairs included) and 'high-low' (only pairs with scores 1 or 3 included). For the case of Ireland where negative (antagonistic) relations also exist in the network, results are shown separately for positive and negative relations.

For Ireland and the positively related actor pairs, the Google S_G^P and mutual information S_I^P metrics outperform the Jaccard S_J^P and Dice S_D^P metrics both in terms of correlation and (especially) MSE. The highest correlation of 0.61 is achieved by S_I^P (0.66 for high-low ratings). For the negatively related actor pairs, the results are relatively poor, with all correlations being below 0.34. The Jaccard and Dice metrics achieve somewhat higher correlations here, although their MSE is higher than S_G^P and S_I^P. As expected, higher correlation scores in achieved for the 'high-low' experiment rather than the '3-levels' experiments, however, the MSE is usually higher for the 'high-low' experiment. Overall, good correlation is achieved for positive relations using page-count metrics (especially for S_I^P, S_G^P), however, page-count metrics perform poorly for negative relations.

For the case of Aegean, all four metrics achieve similar performance in terms of correlation, while in terms of MSE the Jaccard and Dice metrics outperform the Google and mutual information metrics. Note that for the 'high-low' experiment better correlation scores are achieved (compared to the '3-levels' experiment) and the average MSE is lower (S_I^P, S_G^P) or stays at about the same levels (S_J^P, S_D^P). Overall, correlation results are lower than those achieved for the (positive) relations in the Ireland network and reach the maximum value of 0.52 for the 'high-low' experiment using the Google metric. In terms of average MSE similar conclusions can be reached; for the Aegean case study the Dice metric achieves the minimum MSE at about 0.51 compared to 0.42 for the mutual information in the Ireland case study ('3-levels' experiment).

6.2 Text-based metrics

Next, we present the performance of text-based metric using snippets downloaded from the web using conjunctive queries containing both actors (AND) or individual queries for each actor (IND). Various context window sizes (W) were evaluated experimentally and best results were achieved around window size $W = 10$, i.e., ten words to the left and ten words to the right of the term of interest. Results are reported in Table 2 as a function of number of snippets (100, 200 or 500), type of web query (AND, IND), and cosine similarity weighting scheme (binary, log term-frequency).

For the Ireland case study and positive relations, text-based metrics perform relatively poorly especially for the '3-levels' experiment. AND queries outperform IND queries consistently, especially in terms of correlation. The binary (B) weighting scheme outperforms somewhat the LTF scheme but the differences in performance are small. Better performance is achieved for more snippets, especially for the 'high-low' experiment, although the improvement going from 100 to 500 snippets is modest (from 0.36 to 0.42 at best). The highest correlation of 0.42 is achieved for the 'high-low' experiment when using AND queries, the binary weighting scheme and 500 snippets. For negative relations, similar but somewhat higher correlation scores are achieved, up to 0.45. Here the best results are achieved when using individual (IND) queries. Also there is little or no performance

Text-based metrics									
Ratings	Number of snippets	Correlation				MSE			
		Weighting schemes							
		B		LTF		B		LTF	
		Query types							
		AND	IND	AND	IND	AND	IND	AND	IND
IRELAND (pos/neg)									
3-levels	100	**0.29**/0.29	0.06/**0.33**	0.26/0.31	0.10/0.30	0.94/0.65	1.06/**0.55**	0.95/0.67	**0.81**/0.59
	200	0.30/0.28	**0.33**/0.32	0.26/0.29	0.29/**0.34**	**0.95**/0.66	1.51/**0.57**	0.96/0.69	1.39/0.58
	500	**0.29**/0.29	0.09/0.34	0.26/0.30	0.13/**0.35**	0.99/0.68	1.17/0.70	**0.97**/0.69	0.98/**0.64**
high-low	100	**0.36**/0.40	-0.08/**0.43**	0.31/0.41	-0.04/0.40	1.09/0.89	1.34/**0.81**	1.10/0.87	**1.03**/0.86
	200	**0.39**/0.39	0.29/0.41	0.33/0.39	0.20/**0.43**	**1.10**/0.91	1.86/**0.85**	1.10/0.93	1.72/0.86
	500	**0.42**/0.39	0.04/0.44	0.36/0.40	0.04/**0.45**	1.12/0.91	1.47/0.99	**1.09**/**0.89**	1.60/0.93
AEGEAN									
3-levels	100	0.20	0.37	0.19	**0.40**	0.53	**0.41**	0.57	0.43
	200	0.18	0.36	0.19	**0.38**	0.54	0.46	0.59	**0.41**
	500	0.17	0.32	0.19	**0.36**	0.52	0.46	0.58	**0.42**
high-low	100	0.19	0.49	0.16	**0.56**	1.05	0.70	1.00	**0.52**
	200	0.16	0.44	0.17	**0.49**	1.16	0.73	1.14	**0.63**
	500	0.14	0.41	0.18	**0.48**	1.10	0.81	1.12	**0.70**

TABLE 2

difference between the B and LTF weighting schemes. Note that although the correlation scores for negative relations are low they are higher than those achieved using page-count metrics (see Table 1) or link-based metrics (see Table 3 that follows). Similar conclusions can be drawn from the average MSE scores.

For the Aegean case study, slightly higher correlation scores are achieved, up to 0.56 for the 'high-low' experiment. The best results are obtained for the individual (IND) queries; however the performance for conjunctive queries (AND) here is very poor (correlation below 0.2 is achieved throughout). Also the performance does not improve when a larger number of snippets is used and the best correlation and MSE results are obtained (mostly) for 100 or 200 snippets (a sign of data sparseness). Log term frequency weighting outperforms binary weighting especially for the 'high-low' experiment, although the differences are small. Overall, moderate correlation scores are achieved using text-based metrics for the Aegean case study, at the same level or better than those achieved using page-count metrics (see Table 1).

6.3 Link-based metrics

The performance of link-based metrics using outlinks at *base* form is shown in Table 3 in terms of correlation and MSE for Ireland and Aegean. The following metrics are evaluated: Google based semantic relatedness using outlinks S_G^L and cosine similarity using outlinks S_T^L (with binary B and log term frequency LTF weighting). For the case of Ireland and for positive relations, very good correlation performance is achieved especially for the 'high-low' experiment at 0.85. Cosine similarity achieves good performance for the 'high-low' experiment at 0.62 (less so for the '3-levels' experiment). There is no major performance difference between the B and LTF weighing schemes. For the negative relations, very poor results are achieved, throughout, with the binary cosine similarity metric achieving the best performance at 0.25. Overall, the outlinks perform the best out of all evaluated metrics for positive relations in the Ireland network, but fail to identify negative relations.

For the Aegean network, results are not as impressive. Good performance is achieved only for the cosine metric (using binary weighting), up to 0.46 for the 'high-low' experiment, while the Google outlink metric performs poorly (unlike Ireland). Note that in terms of average MSE performance cosine similarity using outlinks S_T^L performs the best out of all metrics (page-count and text-based). Overall, outlinks produce good results for both case studies; for Ireland S_G^L performs the best, while for Aegean S_T^L provides the best results.

6.4 Combination of metrics

Next we investigate the performance for the linear combination of the three types of metrics, namely, page-count, text and link-based metrics. For each case study, we have selected the metric that performs best in terms of correlation. Specifically for the Ireland case study and for positive relations, we have selected mutual information S_I^P as the best performer among the page-count-based metrics, binary weighting using the 200 top-ranked snippets (AND queries) as the best text-based metric and the S_G^L as the best link-based metric. For negative relations, we have selected the Dice S_D^P page-count metric, LTF weighting using the 500 top-ranked snippets (IND queries) from the text-based metrics and the S_T^L with B scheme as link-based metric. Similarly for the Aegean case study, we have selected the Google page-count metric S_G^P, the LTF weighted text-based metric using 100 snippets (IND queries), and the binary weighted cosine similarity S_T^L link-based metric, respectively. The results are presented in Table 4 for the two networks, using equal weights. First the performance of the individual metrics is shown (first three rows), then their two-way combinations are shown with equal weights or inverse of variances (next three lines) and finally the three way combination results are shown (using equal or inverse variance weights).

Ratings	Link-based metrics: Correlation S_G^L	Correlation S_T^L B	Correlation S_T^L LTF	MSE S_G^L	MSE S_T^L B	MSE S_T^L LTF
IRELAND (pos/neg)						
3-levels	**0.62**/0.01	0.34/**0.21**	0.36/0.18	**0.36**/0.83	0.79/0.71	0.79/**0.69**
high-low	**0.85**/0.001	0.61/**0.25**	0.59/0.22	**0.25**/1.24	0.79/1.04	0.84/**1.02**
AEGEAN)						
3-levels	0.27	**0.36**	0.25	0.84	**0.41**	0.57
high-low	0.23	**0.46**	0.27	1.34	**0.80**	1.20

TABLE 3

Ratings	Weights λ_P	Weights λ_T	Weights λ_L	IRELAND (pos/neg) Ratings Correlation	IRELAND Ratings MSE	IRELAND Degree of Centrality Correlation	IRELAND Degree of Centrality MSE$\times10^{-2}$	AEGEAN Ratings Corr.	AEGEAN Ratings MSE	AEGEAN Degree of Centrality Corr.	AEGEAN Degree of Centrality MSE$\times10^{-2}$
3-levels	1	0	0	0.61/0.29	0.42/1.12	0.97/0.97	0.6/5.4	0.35	0.91	0.90	28.4
	0	1	0	0.30/0.35	0.95/0.64	0.89/0.98	2.7/1.1	0.40	0.43	0.90	3.9
	0	0	1	0.62/0.21	0.36/0.71	0.96/0.97	0.5/1.3	0.36	0.41	0.89	2.1
	0	0.5	0.5	0.51/0.35	0.57/0.58	0.94/0.98	1.3/**0.6**	0.42	**0.38**	0.90	**1.8**
	0.5	0	0.5	**0.74**/0.27	**0.26**/0.84	**0.98**/0.97	**0.2**/2.8	0.39	0.52	0.91	8.8
	0.5	0.5	0	0.63/**0.42**	0.51/**0.57**	0.97/**0.98**	1.1/0.9	**0.45**	0.41	**0.92**	4.1
	0.33	0.33	0.33	0.68/0.37	0.36/0.68	0.98/0.98	0.6/1.7	0.44	0.42	0.91	4.2
(SVM)	*0.48/–*	*-0.02/–*	*0.54/–*	*0.74/–*	*0.28/–*	*0.99/–*	*0.31/–*	*0.43*	*0.54*	*0.92*	*11*
high-low	1	0	0	0.66/0.35	0.54/1.36	0.97/0.95	0.7/2.7	0.52	0.61	0.92	3.3
	0	1	0	0.39/0.45	1.10/0.93	0.91/**0.97**	2.6/1.0	0.56	**0.52**	0.91	**2.5**
	0	0	1	0.85/0.25	**0.25**/1.04	0.98/0.95	0.5/0.9	0.46	0.80	0.92	6.9
	0	0.5	0.5	0.71/0.39	0.60/0.86	0.96/0.97	1.4/0.5	0.57	0.59	0.93	4.1
	0.5	0	0.5	**0.86**/0.31	0.26/1.13	**0.99**/0.95	**0.3**/1.5	0.53	0.70	0.92	5.2
	0.5	0.5	0	0.68/**0.46**	0.66/**0.83**	0.97/0.97	1.3/**0.5**	**0.62**	0.63	0.93	4.9
	0.33	0.33	0.33	0.81/0.40	0.43/0.95	0.98/0.96	0.9/0.9	0.60	0.62	**0.93**	4.7
(SVM)	*0.21/–*	*-0.09/–*	*0.88/–*	*0.88/–*	*0.22/–*	*0.99/–*	*0.37/–*	*0.61*	*0.50*	*0.93*	*2.6*

TABLE 4

For the Ireland case study and for positive relations, the two-way combination of the page-count and link-based metrics achieves the highest correlation on positive relations both for the '3-levels' and 'high-low' experiments at 0.74 and 0.86, respectively. The three-way combination with equal linear weights performs somewhat worse, which is expected due to the poor baseline performance of text-based metrics. For negative relations and for the '3-levels' experiment, the two-way combination of page-count-based and text-based metrics achieves the highest correlation at 0.42, followed closely by the three-way combination at 0.37. The results are very similar also for 'high-low' experiment with correlation up to 0.46. Overall, simple linear fusion outperforms the individual metrics and achieves very good performance for positive relations and acceptable performance for negative relations.

For the Aegean case study and for equal weights, the combination of page-count and text-based metrics achieves the best performance in terms of correlation, while the three-way combination is a close second. All metric combinations achieve a consistent performance improvement in terms of correlation over the baseline, however, this is not always the case in terms of average MSE[8]. Overall, the performance of the combined metrics is good and achieves correlation of up to 0.62 for the 'high-low' experiment. Unlike Ireland where the link-based metrics perform the best for positive relations, here the text-based metric is the best performer and combinations that contain it achieve the highest correlation.

The degree of centrality results are also shown in Table 4 for the two networks, using individual metrics and their linear fusion with equal weights. The agreement between the original and extracted networks is very good both in terms of correlation and MSE. For the Ireland case study, correlation of up to 0.99 is achieved. The lowest correlation of 0.89 is for the text-based metric and for positive relations; all other individual and combined metrics score over 0.94. For negative relations, agreement between the original and extracted network centralities is excellent (over 0.97) for all metrics and their combinations. The results are also very good for the Aegean case study, achieving correlations between 0.90 and 0.93. There are no significant differences in performance between metrics or their combinations. Overall, all metrics (with the possible exception of the text-based metric for positive relations in Ireland) perform equally well for the degree of centrality computation and provide very good to excellent correlation results.

We have also investigated linear fusion using inverse variance weighting. The results are not shown here since they are very similar to equal weight fusion. In fact, the variance of each of the metrics is very similar (with the exception of negative relations in the Ireland case study) resulting in very similar correlation scores for inverse variance and equal weighting. For the three-way combination the correlation results are $0.80/0.40$ and 0.59 for

8. MSE results are significantly lower than chance, i.e., randomly assigning scores to actor pairs. Specifically, for the Ireland case study, MSE is reduced approximately 5-10 times for positive relations compared to chance, while for negative relations and the Aegean network MSE reduction is 2-4 times.

the inverse variance weights for the positive/negative Ireland and Aegean case studies, respectively ('high-low' experiment). This is within 0.01 of the equal weighting scores. Finally, we computed the optimal linear weights through exhaustive grid search, i.e., the weights that maximize correlation with human ratings. For the Ireland case study ('high-low') experiment the best correlation obtained for positive relations was 0.88 (compare this with 0.86 in the table) for weights $(0.3, 0, 0.7)$, while for the Aegean case correlation of 0.62 was achieved for weights $(0.3, 0.5, 0.2)$, a small improvement over the equal weighting scheme. Lastly, we used linear regression and Support Vector Machines (SVM) for the fusion of the three metrics. The estimated weights obtained by both approaches were found to be quite similar. The results for the SVM algorithm (in italics) are presented in Table 4 for both networks (only positive relations for Ireland). The performance is comparable to the scores achieved by the fusion scheme using equal weights.

6.5 Network Visualization

In this section, the manually annotated and automatically extracted networks for both case studies are displayed as graphs. The graphs were created using the NEATO program, http://www.graphviz.org, that implements the graph creation algorithm proposed in [39]. In Fig. 1 the graphs of the original and extracted policy networks are shown for Ireland and Aegean, respectively. The nodes are labeled using the acronyms of the actors supplied by political scientists. We used the relatedness scores from the three-way linear combination of all metrics using equal weights (see '3-levels' experiment in Table 4). In the graphs, the line thickness for each edge is proportional to the relatedness score of the corresponding actor pair, i.e., greater line thickness denotes a stronger relation (positive or negative). We use five levels of line thickness each corresponding to a sub-interval of [1,3]; the sub-intervals are selected so that relatedness scores are uniformly distributed in the five value ranges. In order for the original and extracted graph to be directly comparable we use the same topology (location of the actors) in both networks. Only the subset of relations for which we have computed a score automatically are shown in the graphs.

The graphs for the positive relations of the Ireland network are shown in Fig. 1(a), (b) for the original (a) and extracted networks (b). Each sub-graph corresponds to one of the diagonal blocks (positive relations) of the relatedness matrix. Overall, there is good agreement between Fig. 1 (a) and (b) in terms of strength of relations; only SEREGA in block 1 appears somewhat less connected to the rest of the extracted network (compared to the original). The negative relations that appear in the off-diagonal blocks are lumped together in a single network shown in Fig. 1(c), (d). Despite the very low correlation scores achieved for negative relations, the original and extracted graphs look reasonably similar, e.g., the actors in the {DOE, LIMCOCO, CLCOCO} clique are strongly interconnected in both graphs.

A qualitative analysis of the Aegean graphs in Fig. 1(e), (f) reveals very similar connectivity patterns for most of the actors in the original and extracted network. For example, the actors {CDA, DTEDK, DPR, RS, CC} have high connectivity and are central in both graphs, while the actors {UA, CTUC, MC, RCC} have weaker relations and are peripheral (again in both). However, some actors have increased their strength of relations and connectivity, and have become more central in the extracted network, e.g., {DC, CTEDK}. Overall, the qualitative analysis of the extracted graphs shows good agreement with those from political scientists.

7 Discussion

Important parameters that affect the quality of the automatically extracted network (in addition to the relatedness metrics used) include data sparseness, lexical ambiguity for actors, language and type of network relations. When comparing manually annotated policy networks with automatically extracted ones, human biases also come into play, e.g., cultural biases and scaling of the scores by political scientists.

Policy networks often involve small and medium size actors with limited web presence. The data sparseness problem is especially pronounced for metrics that require the co-occurrence of actor terms in the same document, i.e., page-count metrics and text-based metrics that use conjunctive (AND) queries. Between the two case studies, Aegean is the one suffering the most from data sparseness. This is evident from the small number of web hits for some Aegean actors, as well, as by the poor performance of text-based metrics that use AND queries. In general, AND queries should outperform IND queries for relatedness computation (see Ireland case and [27]) but here the quality of the snippets downloaded using AND queries is too poor. Note that data sparseness is also an issue for link-based metrics in the Aegean network; outlinks appear much less often in documents that contain Greek actors. There is no obvious solution to this problem other than working with metrics that are more robust to sparse data (such as the text-based IND metric) and carefully selecting actor lexicalizations to improve the relevance of web harvested data. Also, the relatedness scores computed by different types of metrics have different distributions that affect the fusion of metrics. However, the effect of the different distributions to the computation of the final relatedness is not obvious.

An important issue that has not been researched much in this work is the lexicalization of actors. Ideally, one would like to create a generative grammar of actor names that contains all possible wordforms of the actor with no overlap with other actor names. Unfortunately, this is rarely possible especially for actor acronyms, resulting in significant ambiguity. In our data, this has been

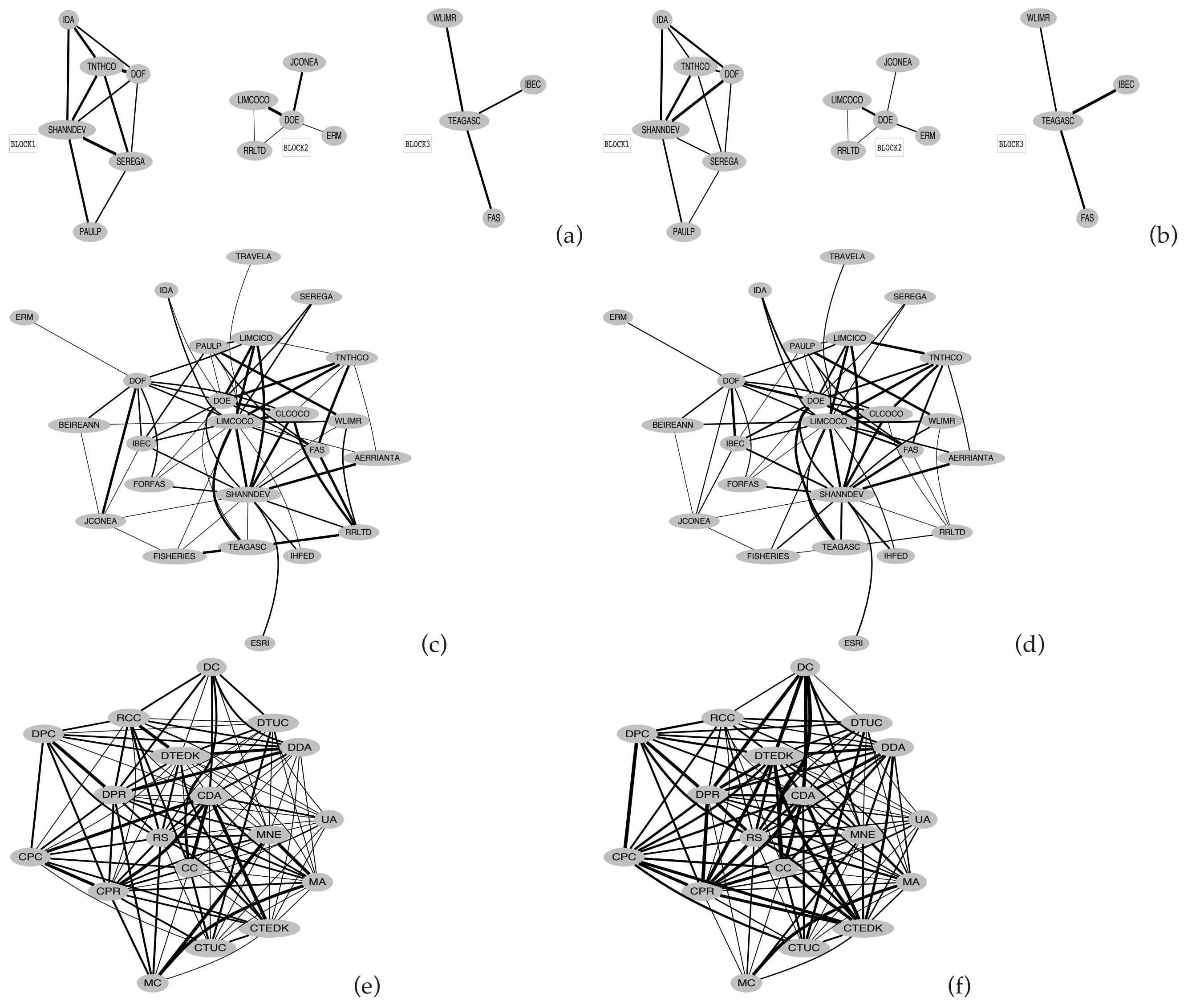

Fig. 1. Ireland-positive relations: (a) original and (b) extracted networks. Ireland-negative relations: (c) original and (d) extracted networks. Aegean: (e) original and (f) extracted networks.

a problem mainly for Ireland where many competing actor names exist in other English speaking countries. As a simple solution, we have constrained queries using additional terms, e.g., “Ireland”. Thus, it remains an open problem how to select the appropriate lexicalizations automatically (given a baseform), as well as, how to define the constraints that can pre- or post-filter the relevant web documents (see also Section 7.1).

In the Ireland case study, it is much harder to identify negative relations among actors than positive ones. Also, page-count, link-based metrics and text-based metrics using AND queries that work well for positive relations, perform relatively poorly for negative ones. It is interesting to note that all the aforementioned metrics use some notion of co-occurrence (be it actor, context or actor-outlink co-occurrence). Co-occurrence is a good feature for measuring positive relations (or similarity) because it reduces ambiguity and identifies the common ground among two actors. This does not seem to be the case for negative relations. In fact, text-based metrics using IND queries perform the best for this task. A possible explanation is that text-based metrics are good at identifying similarity [27], and actors that have similar policy roles often compete.

Overall, the following conclusions can be drawn on the relative performance of metrics: 1) For positive relations link-based metrics work the best, while for negative relations context-based metrics work the best. 2) Ambiguity in actor lexicalization significantly affects the relevance of web harvested data. Certain metrics are less robust to actor ambiguity, especially, text-based metrics that use IND queries and (less so) link-based metrics. 3) Data sparseness mostly affects the performance of metrics that use co-occurrence as a feature, i.e., page-count metrics and text-based metrics that use conjunctive (AND) queries. Extensive experimentation on various policy networks from different countries is required to further validate these claims.

Some of the observed differences between the two case studies might also reflect methodological differences between the two data collection efforts[9]. When comparing the Ireland and Aegean networks it is clear that although the range of scores used is the same, the score distribution is different. For the Aegean case the majority of scores are 1's and mainly 2's, with very few 3's. For Ireland, the scores are more uniformly distributed. This indicates not so much a fundamental difference between the two networks but rather a different cognitive scaling mechanism for the relatedness scores by the political scientists. The automatic method eliminates such scaling differences and normalization issues. Note that web harvested data might also be biased, since web data is also generated by humans. The proposed method can better identify (but cannot fully eliminate) such biases by focusing on specific types of web sources (e.g., blogs, news) or time periods.

An important question is the applicability of the proposed methods and metrics to the automatic extraction of other types of social networks. Specifically, we investigated the correlation between actual and extracted relations for a flight traffic and a scientific co-citation network. Overall, results were poor for the flight network; only metrics using co-occurrence as a feature provided positive correlation (up to 0.34 correlation for page-count metrics and 0.28 for text-based metrics). For the co-citation network results were more encouraging: up to 0.64 correlation was achieved by the text-based metrics using AND queries and 500 web snippets. Overall, it is hard to draw general conclusions about the generalizability of the proposed metrics; performance depends on the type of relation that the network encodes.

7.1 Policy network analysis

Next, we investigate the relevance of our results for political scientists. Compared to the typical top-down approach of mapping actors and their relations via interviews and questionnaires, the web-based approach provides a significant advantage in terms of discovering missing actors and linkages (in a bottom-up fashion). We have investigated the potential of the proposed methodology for extracting relations that have escaped political scientists in the Aegean network. From the top-15 automatically extracted relations (for pairs that had been manually assigned a score of 0), political scientists rated three pairs as “probably related” and nine pairs as “maybe related”. Similarly for actor discovery we have identified the top-200 terms extracted using C-value/NC-value [36] from 500 downloaded documents using AND queries. Political scientists identified five new stakeholders as potential actors for the Aegean network.

Researchers investigate policy network characteristics such as centrality and structural equivalence as indicators of administrative restructuring, sub-national mobilization, devolution and decentralization. The automatically extracted policy networks provide political analysts with a valuable tool to validate results, verify assumptions and more thoroughly comprehend the *dynamics of network governance* in a rapidly changing world. We attempt to capture the evolution of the two networks over time in Fig. 2. For this purpose we harvest web data for specific dates using a simple heuristic, i.e., adding the year at the end of each IND and AND query. The data from each year between 1996 and 2010 was processed independently and a separate network was extracted for each year. In Fig. 2, the correlation between the degree of centrality of the original network (a single data point from data collected over the time period 2001-2003) and each of the networks for years 1996-2010 is shown. The similarity scores from the page-count-based metrics (S_I^P for Ireland and S_G^P for Aegean) were used to compute the degree of centrality of the extracted networks. In both case studies, correlation scores were smoothed using a three year moving average window. Correlation of the degree of centrality is computed using the ‘3-levels’ ratings. For the Ireland case study, correlation results are shown for all relations (we do not separate between positive and negative relations). The results for Ireland show very good agreement between the original and extracted networks throughout the time period under examination. For the Aegean network, there is better agreement (higher correlation) in the 2001-2003 time-frame when the political scientists performed their analysis. The evolution of the automatically extracted Aegean network over the 1996-2010 time-frame was judged to provide valuable information by political scientists. An example of the centrality evolution per year for regional and central actors is presented in Table 5. The evolution percentage was estimated using the centralities of actors for the period 1996-2010. This is an example of our generic observation suggesting that the political influence of central actors over regional actors

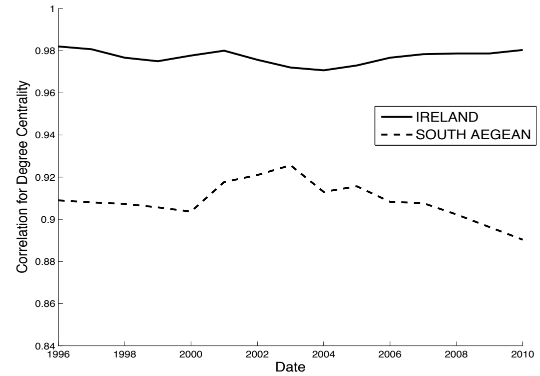

Fig. 2. Evolution of centrality correlation over time.

Network	IRELAND		AEGEAN	
Actor	TNTHCO (regional)	DOF (central)	CTEDK (regional)	MNE (central)
Centrality evol.	0.9%	-0.7%	0.4%	-0.6%

TABLE 5

9. Anecdotal evidence from the political scientists involved in the two studies verify this, e.g., Greek participants would often confuse social and professional relations between staff members of actors.

becomes more weak as the network of the latter evolves positively over time. For example it was expected that the European funding opportunities would allow local actors such as CTEDK to increase networking and participation in development projects; this is consistent with the increased centrality of CTEDK over time in the network (not shown here). Last but not least, the web-based method can provide intuition on different *subjective views of network*, e.g., exploiting government vs private sector web data sources.

Overall, the proposed method can produce high-quality results, offer new perspectives into the data and lead to significant savings in effort and time required to map policy networks. Note, however, that face-to-face interactions with the interviewees will reveal information that might not be accessible on the web and reflects positive or negative feelings, attitudes and beliefs that might influence the way networks evolve. All in all, a mixed approach where the web is used as a starting point to map actors and their relations, followed by a limited set of interviews with key actors is judged as the optimal procedure by political scientists.

8 Conclusions

In this work, we have shown that it is possible to automatically compute the strength of relations between actors in order to automatically create policy networks. A variety of features were proposed and evaluated that used information automatically extracted from the world wide web. Specifically, we investigated the use of page-counts, lexical context and outlinks, as well as, their fusion, as potential features for estimating relatedness between actor pairs. The proposed method was evaluated on two case studies with good results achieving correlation of up to 0.74 for positive relations. However, it was shown that it is much harder to extract negative relations, only moderate success was achieved for this task. Among the metrics there was not a clear winner. A variety of parameters such as data sparseness, actor name ambiguity, language and relation type affect the performance of the relatedness metrics. The automatically extracted networks were also validated by political scientists and useful conclusions about the evolution of the networks over time were drawn.

This work is a first step towards creating algorithms and tools useful to policy network analysts. Future work should involve: 1) automatically identifying actors participating in policy networks and their lexicalizations, 2) investigating machine learning algorithms for selecting the most informative metrics, as well as, for their fusion, 3) filtering web data based on relevance and type of source (e.g., non-profit, government, corporate, blogs, news), and 4) investigating the applicability of the proposed metrics for other types of social networks.

References

[1] J. Peterson and E. Bomberg, *Desicion-Making in the European Union*. Basingstoke and New York: Palgrave Macmillan, 1999.

[2] P. Kenis and V. Schneider, *Policy Networks and Policy Analysis: Scrutinizing a New Analytical Toolbox*. Westview Press, 1991, pp. 25–59.

[3] L. Zhu. Computational Political Science Literature Survey. (last accessed 28/2/11). [Online]. Available: http://www.personal.psu.edu/luz113

[4] B. Monroe and P. Schrodt, "Introduction to the Special Issue: The Statistical Analysis of Political Text," *Political Analysis*, vol. 16, no. 4, pp. 351–355, 2008.

[5] M. Laver, K. Benoit, and J. Garry, "Extracting Policy Positions from Political Texts Using Words as Data," *American Political Science Review*, vol. 97, no. 2, pp. 311–331, 2003.

[6] K. Benoit and M. Laver, "Estimating Irish Party Policy Positions Using Computer Wordscoring: The 2002 Elections - A Research Note," *Irish Political Studies*, vol. 18, no. 1, pp. 97–107, 2003.

[7] Jonathan B. Slapin and Sven-Oliver Proksch, "A Scaling Model for Estimating Time-Series Party Positions from Texts," *American Journal of Political Science*, vol. 52, no. 3, pp. 705–722, 2008.

[8] M. Thomas, B. Pang, and L. Lee, "Get Out the Vote: Determining Support or Opposition from Congressional Floor-debate Transcripts," in *Proc. of the 2006 Conf. on Empirical Methods in Natural Language Processing*, 2006, pp. 327–335.

[9] B. Chen, L. Zhu, D. Kifer, and D. Lee, "What Is an Opinion About? Exploring Political Standpoints Using Opinion Scoring Model," in *Proc. of the 24th AAAI Conf. on Artificial Intelligence*, 2010, pp. 1007–1012.

[10] W. Gryc and K. Moilanen, "Leveraging Textual Sentiment Analysis with Social Network Modelling: Sentiment Analysis of Political Blogs in the 2008 U.S Presidential Election," in *Proc. of the "From Text to Political Positions" Workshop*, 2010.

[11] B. Monroe, M. Colaresi, and K. Quinn, "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict," *Political Analysis*, vol. 16, no. 4, pp. 372–403, 2008.

[12] H. Kautz, B. Selman, and M. Shah, "The Hidden Web," *AI Magazine*, vol. 18, no. 2, pp. 27–36, 1997.

[13] Y. Matsuo, J. Mori, and M. Hamasaki, "POLYPHONET: An Advanced Social Network Extraction System from the Web," in *Proc. of the 15th International World Wide Web Conf.*, 2006, pp. 397–406.

[14] H. Tomobe, Y. Matsuo, and K. Hasida, "Social Network Extraction of Conference Participants," in *Proc. of the 12th Int'l World Wide Web Conf.*, 2003.

[15] Y. Jin, Y. Matsuo, and M. Ishizuka, "Extracting Social Networks among Various Entities on the Web," in *Proc. of the European Conf. on The Semantic Web: Research and Applications*, 2007, pp. 251–266.

[16] J. Mori, T. Tsujishita, Y. Matsuo, and M. Ishizuka, "Extracting Relations in Social Networks from the Web Using Similarity Between Collective Contexts," in *Proc. of the 5th International Semantic Web Conference*, 2006, pp. 487–500.

[17] F. Mesquita, Y. Merhav, and D. Barbosa, "Extracting Information Networks from the Blogosphere: State-of-the-Art and Challenges," in *Proc. of the 4th International Conf. on Weblogs and Social Media, Data Challenge Workshop*, 2010.

[18] R. Xiang, J. Neville, and M. Rogati, "Modeling Relationship Strength in Online Social Networks," in *Proc. of the 19th International World Wide Web Conference*, 2010, pp. 981–990.

[19] P. Mika, "Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks," *Journal of Web Semantics*, vol. 3, no. 2, pp. 211–223, 2005.

[20] ——, "Ontlogies Are Us: A Unified Model of Social Networks and Semantics," *Journal of Web Semantics*, vol. 5, no. 1, pp. 5–15, 2007.

[21] A. Culotta, R. Bekkerman, and A. Mccallum, "Extracting Social Networks and Contact Information from Email and the Web," in *Proc. of the 1th Conference on Email and Anti-Spam*, 2004.

[22] A. Gruzd and C. Haythornthwaite, "Automated Discovery and Analysis of Social Networks from Threaded Discussions," in *Proc. of the International Network of Social Network Analysis*, 2008.

[23] B. Pouliquen, R. Steinberg, and J. Belyaeva, "Multilingual Multi-document Continuously-updated Social Networks," in *Proc. of the International Conference in Recent Advances in Natural Language Processing*, 2007, pp. 25–32.

[24] David K. Elson, Nicholas Dames, and Kathleen R. McKeown, "Extracting Social Networks from Literary Fiction," in *Proc. of the 48th Annual Meeting of the Association of Computational Linguistics*, 2010, pp. 138–147.

[25] P. Nasirifard, V. Peristeras, C. Hayes, and S. Decker, "Extracting and Utilizing Social Networks from Log Files of Shared Workspaces," in *Proc. of the 10th IFIP Working Conference on Virtual Enterprises*, 2009, pp. 643–650.

[26] E. Iosif and A. Potamianos, "Unsupervised Semantic Similarity Computation using Web Search Engines," in *Proc. of the IEEE/WIC/ACM Int'l Conf. on Web Intelligence*, 2007, pp. 381–387.

[27] ——, "Unsupervised Semantic Similarity Computation between Terms Using Web Documents," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 11, pp. 1637–1647, 2010.

[28] D. Bollegala, Y. Matsuo, and M. Ishizuka, "Measuring Semantic Similarity between Words Using Web Search Engines," in *Proc. of the 16th International World Wide Web Conference*, 2007, pp. 757–766.

[29] J. Gracia, R. Trillo, M. Espinoza, and E. Mena, "Querying the Web: A Multiontology Disambiguation Method," in *Proc. of the 6th International Conference on Web Engineering*, 2006, pp. 241–248.

[30] R. Cilibrasi and P. Vitanyi, "The Google Similarity Distance," *IEEE Trans. Knowledge and Data Eng.*, vol. 19, no. 3, pp. 370–383, 2007.

[31] P. Vitanyi, "Universal Similarity," in *Proc. of Information Theory Workshop Coding and Complexity*, 2005, pp. 238–243.

[32] P. Calado, M. Cristo, E. Moura, N. Ziviani, B. Ribeiro-Neto, and M. A. Gonçalves, "Combining Link-Based and Content-Based Methods for Web Document Classification," in *Proc. of the 12th International Conf. on Information and Knowledge Management*, 2003, pp. 394–401.

[33] S. Chakrabarti, *Mining the Web: Discovering Knowledge from Hypertext Data*. Morgan-Kauffman, 2002.

[34] N. Rees, B. Quinn, and B. Connaughton, "Ireland's Pragmatic Adaptation to Regionalization: the Mid-West Region," *Regional and Federal Studies*, vol. 14, no. 3, pp. 379–404, 2004.

[35] P. Getimis and L. Demetropoulou, "Europeanization towards New Forms of Regional Governance in Greece," *Regional Studies Association International Conference*, 2003.

[36] K. Frantzi, S. Ananiadou, and H. Mima, "Automatic Recognition of Multi-Word Terms: The C-Value/NC-Value Method," *International Journal on Digital Libraries*, vol. 3, no. 2, pp. 115–130, 2000.

[37] F. Lazarinis, "Engineering and Utilizing a Stopword List in Greek Web Retrieval," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 11, pp. 1645–1652, 2007.

[38] M. E. J. Newman, "Analysis of Weighted Networks," *Physical Review E*, vol. 70, no. 5, p. 056131, 2004.

[39] T. Kamada and S. Kawai, "An Algorithm for Drawing General Undirected Graphs," *Information Processing Letters*, vol. 31, no. 1, pp. 7–15, 1989.

Theodosis Moschopoulos received the Diploma and the MS degrees from the ECE Dept., Tech. Univ. of Crete, in 2008 and 2012 respectively.

Elias Iosif (StM'08) received the Diploma and the MS degrees from the ECE Dept., Tech. Univ. of Crete, in 2005 and 2007 respectively. Since 2007, he is a research assistant and a Ph.D. candidate at the ECE Dept. Tech. Univ. of Crete. His current research interests include natural language processing and semantic web. He is a member of the Cyprus Scientific and Technical Chamber since 2005.

Leeda Demetropoulou received the BA (Hons) degree from the Centre for Language & Communication of the Univ. of Wales in 1995, and the MA and Ph.D. degrees from the Centre for European Studies of the Univ. of Exeter in 1996 and 2001, respectively. From 1999 to 2002 she was a research assistant at the Polytechnic School of the Aristotle Univ. From 2002 to 2005 she was a research assistant at the Institute of Human Resources and Urban Development of Panteion Univ. In 2008 she was a Fulbright Visiting Scholar at the Viterbi School of Engineering, Univ. of Southern California. She associated with the Tech. Univ. of Crete in 2006 and has been working as a research associate at the Dept. of Environmental Engineering since 2009. Her research interests include European governance, policy networks and public participation, regional development and spatial planning, sustainability and environmental management.

Alexandros Potamianos (M'92, SM'10) received the Diploma in Electrical & Computer Engineering from the National Tech. Univ. of Athens, Greece in 1990. He received the MS and Ph.D. degrees in Engineering Sciences from Harvard University, USA in 1991 and 1995, respectively. From 1991 to June 1993 he was a research assistant at the Harvard Robotics Lab, Harvard University. From 1993 to 1995 he was a research assistant at the Digital Signal Processing Lab at Georgia Tech. From 1995 to 1999 he was a Senior Technical Staff Member at the Speech and Image Processing Lab, AT&T Shannon Labs, Florham Park, NJ. From 1999 to 2002 he was a Technical Staff Member and Technical Supervisor at the Multimedia Communications Lab at Bell Labs, Lucent Technologies, Murray Hill, NJ. From 1999 to 2001 he was an adjunct Assistant Professor at the Dept. of Electrical Engineering of Columbia University, New York. In the spring of 2003, he joined the ECE Dept., Tech. Univ. of Crete, Greece as an associate professor. His current research interests include speech processing, analysis, synthesis and recognition, dialog and multimodal systems, nonlinear signal processing, natural language understanding, data mining, artificial intelligence and multimodal child-computer interaction. Prof. Potamianos has authored or co-authored over 80 papers in professional journals and conferences. He is the co-author of the paper "Creating conversational interfaces for children" that received a 2005 IEEE Signal Processing Society Best Paper Award. He holds four patents. He is a member of the IEEE Signal Processing Society since 1992 and he is currently serving his second term with the IEEE Speech Technical Committee.

Shrikanth (Shri) Narayanan (StM'88, M'95, SM'02, F'09) is Andrew J. Viterbi Professor of Engineering at the University of Southern California (USC), and holds appointments as Professor of Electrical Engineering, Computer Science, Linguistics and Psychology and as the founding director of the Ming Hsieh Institute. Prior to USC he was with AT&T Bell Labs and AT&T Research from 1995–2000. At USC he directs the Signal Analysis and Interpretation Laboratory (SAIL). His research focuses on human-centered information processing and communication technologies with a special emphasis on behavioral signal processing and informatics. Prof. Narayanan is a Fellow of the Acoustical Society of America and the American Association for the Advancement of Science (AAAS) and a member of Tau Beta Pi, Phi Kappa Phi, and Eta Kappa Nu. He is also an Editor for the Computer Speech and Language Journal and an Associate Editor for the IEEE TRANSACTIONS ON MULTIMEDIA, IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, and the JASA. He was also previously an Assoc. Editor of the IEEE TRANSACTIONS OF SPEECH AND AUDIO PROCESSING (2000–2004) and the IEEE SIGNAL PROCESSING MAGAZINE (2005–2008). He is a recipient of a number of honors including Best Paper awards from the IEEE Signal Processing Society in 2005 (with A. Potamianos) and in 2009 (with C. M. Lee) and selection as an IEEE Signal Processing Society Distinguished Lecturer for 2010–2011. Papers co-authored with his students have won awards at Interspeech'11 Speaker State Challenge, Interspeech'10, Interspeech'09-Emotion Challenge, IEEE DCOSS'09, IEEE MMSP'07, IEEE MMSP'06, ICASSP'05 and ICSLP'02. He has published over 475 papers and has twelve granted U.S. patents.