Original Research Article

Big Data & Society
January–June 2015: 1–28

DOI: 10.1177/2053951715572916
bds.sagepub.com

# Automated analysis of the US presidential elections using Big Data and network analysis

Saatviga Sudhahar[1], Giuseppe A Veltri[2] and Nello Cristianini[1]

**Abstract**

The automated parsing of 130,213 news articles about the 2012 US presidential elections produces a network formed by the key political actors and issues, which were linked by relations of support and opposition. The nodes are formed by noun phrases and links by verbs, directly expressing the action of one node upon the other. This network is studied by applying insights from several theories and techniques, and by combining existing tools in an innovative way, including: graph partitioning, centrality, assortativity, hierarchy and structural balance. The analysis yields various patterns. First, we observe that the fundamental split between the Republican and Democrat camps can be easily detected by network partitioning, which provides a strong validation check of the approach adopted, as well as a sound way to assign actors and topics to one of the two camps. Second, we identify the most central nodes of the political camps. We also learnt that Clinton played a more central role than Biden in the Democrat camp; the overall campaign was much focused on economy and rights; the Republican Party (Grand Old Party or GOP) is the most divisive subject in the campaign, and is portrayed more negatively than the Democrats; and, overall, the media reported positive statements more frequently for the Democrats than the Republicans. This is the first study in which political positions are automatically extracted and derived from a very large corpus of online news, generating a network that goes well beyond traditional word-association networks by means of richer linguistic analysis of texts.

**Keywords**

Big Data, network analysis, structural balance, computational social science, mediascape, subject-verb-object

## Introduction

In all mature democracies, political elections are the arena of enormous mobilisation of material and symbolic resources, where the mobilisation of symbolic resources is defined as the strategic use of words, images and concepts to either persuade or influence public opinion. The American presidential elections are among the most interesting campaigns because of the sheer amount of resources deployed and the influence of the United States on global governance.

This study presents a large-scale analysis of mass media coverage of the 2012 US presidential elections, combining automatic corpus linguistics methods and network analysis to obtain a network representation of the entire campaign coverage by the news media. Mapping the full extent to which an electoral campaign is represented by media offline and online constitutes a very difficult challenge for researchers, given the large amount of data and the multitude of sources available in advanced democracies. In addition, whenever the full coverage has been analysed, the core method used so far has been traditional content analysis (in either the automatic or manual coding variant), which allows us to identify the most salient issues in each candidate's campaign (e.g. Conway et al., 2012).

[1]Department of Computer Science, University of Bristol, Bristol, UK
[2]Department of Media and Communication, University of Leicester, Leicester, UK

**Corresponding author:**
Saatviga Sudhahar, Department of Computer Science, University of Bristol, Merchant Venturer's Building, Woodland Road, Bristol BS8 1UB, UK.
Email: saatviga.sudhahar@bristol.ac.uk

The traditional lines of investigation of the social sciences, such as the ideological position of candidates, their political communication strategy and the social representations (Moscovici and Duveen, 2000) of the election in the media remain salient, but require new conceptual and methodological approaches. We propose a Big Data approach, based on automated information extraction.

Our study is based on the automated analysis of 130,213 news articles related to the US presidential elections from 719 news outlets, based on state-of-the-art Natural Language Processing (NLP) and Artificial Intelligence (AI) techniques, to extract information about the key actors and their relations in the media narrative of the US elections. These relations are in the form of subject-verb-object (SVO) triplets extracted by a parser, and are organised as a network, or semantic graph, which can then be analysed with mathematical tools. This approach goes well beyond traditional word-association networks, producing not only directed links, but links whose semantic nature is retained and understood. Much of the following analysis would not be possible without these features, which ultimately resulted from the use of a parser to identify noun phrases and verbs to form SVO triplets. An example of one such triplet is 'Obama criticised Romney'.

This article is divided into different sections: first, we discuss the theoretical framework that informed our study, stating the importance of applying a semantic graph approach to texts; second, we outline the novel methodology employed to automate the analysis of news media content. Third, we present our findings; last, we conclude by reflecting on both theoretical and methodological aspects of studying the network of actors and actions.

## Big Data analysis of news coverage

In this study, we analyse texts by an automatic identification of 'semantic triplets', formed by key actors, objects and their relationships. In particular, we consider SVO triplets, following the approach developed by Quantitative Narrative Analysis or QNA (Earl et al., 2004; Franzosi, 1987). The core idea of QNA is that SVO triplets form the fundamental units of narration. Our methodological contribution to that literature has been to automate the extraction of triplets, by identifying subjects and objects with noun phrases and using a parser. The set of triplets obtained is naturally organised into a network, which we call a 'semantic graph'. The semantic graph encodes the relations among all actors in the corpus. Note that sometimes the expression 'semantic graph' is used to indicate networks formed by linking co-occurring words, which is a purely statistical approach that makes no use of more detailed linguistic information. Instead, our process of automatically extracting the SVO triplets from the corpus is based on the use of a syntactical parser, and is described in detail in the Appendix.

In the context of a news corpus about a political election, we introduce a slightly different interpretation of SVO triplets based on the observation that candidates define their political positioning through a set of claims or actions through which they either endorse or oppose issues and actors. This interpretation will be discussed in the next section. In this sense, SVO triplets can be used as a source of political positioning information. Extracting triplets and adding them in two coherent semantic graphs provides a map of the candidates' strategic positioning across a wide range of issues. Manually generated semantic graphs have been applied in several different collections of texts. For example, Moretti (2013) decomposed novels into network structures in which actors are linked by verbs. In this approach, the properties of a text can be revealed by the structural and topological features of the novel network. Similarly, Mac Carron and Kenna (2012) analysed three classical mythological narratives comparing the topological properties of their corresponding narrative networks.

All these studies were necessarily limited in size for being reliant on manual coding. However, some progress has recently been made in the automatic creation of networks, the closest approach to ours being Diesner et al. (2012) where statistical co-occurrence of named entities is used to form networks. Also in this case, however, the semantics of the link in the network is only provided by statistical co-occurrence between words and not by linguistic analysis.

Compared to these previous studies, we introduce several methodological innovations that are characteristic of a beneficial Big Data approach. This is a crucial point: we not only 'went big' by increasing the 'resolution' of the data collected but this is also the first time such methodology has been applied on a Big Data scale, by making use of syntactical parsers and a fully automated text analysis pipeline. We have increased the quality and resolution of the data obtained by automatic extraction because this method goes well beyond the traditional creation of networks based on word co-occurrences. We have retained semantic information by extracting SVO triplets and classifying relationships in a signed network. It is the first time that noun phrases are used rather than words. Our automated infrastructure produced the network of actors and their relations without human intervention, starting from the corpus of 130,213 articles that we have collected.

The mass media representation of the campaign is decomposed into a signed network structure that reveals relationships between the main actors – e.g.

Obama and Romney and the other objects (e.g. the economy, Iraq, etc.). Besides analysing the overall campaign network, we also extract the networks of each of the candidates. The relationships between Obama and Romney and other objects constitute their networks of claims; the sign (positive or negative) of the edge indicates the nature of the relationship.

## Ideographs, rhetorical space and issues ownership

In the public sphere (Habermas et al., 1974), there are competing definitions of public issues, within what is a complex game played for semantic control (Gaskell et al., 1998). This is particularly evident during the context of political elections. Candidates define their political positioning through a complex set of claims through which they either endorse or oppose issues and actors (both individuals and/or institutions). To identify and map this multifaceted ideological and political representative space, we start from the notion of 'ideographs' introduced by rhetorical theorists (McGee, 1980). Ideographs are ideological stances, expressed in a few words, which define an actor's political identity. They are the basic structural elements, or building blocks, of ideology. Thus, they may be thought of as 'ideographs' for, like Chinese symbols, they signify and contain a unique ideological commitment.

Interestingly, this observation complements the approach commonly taken in QNA, where the emphasis is on detecting 'who did what to whom': here, by viewing actions (and claims) as position statements, we can identify SVO triplets with political positioning claims (where we only consider triplets linked by a verb that expresses support or opposition for the object), and provide a rigorous and effective representation of the rhetorical and political space in which the candidates' positioning takes place – this is achieved by embedding the resulting network of triplets into a space as shown in the 'Network partitioning and embedding' section.

As part of our approach, we can also create graphical descriptions of the overall positioning of an actor, by operationalising and expanding McGee's notion of ideograph into a network perspective to include any graphs encoding all positive and negative statements made by actors towards objects (i.e. other actors or political issues). In this context, it is an important observation that the decomposition of the network of triplets leads to a meaningful distinction of two subnetworks roughly corresponding to the 'Democratic Party' and 'Republican Party' (Grand Old Party (GOP)) semantic fields (see Figure 4). The embedding of the network into a space can also be used to visualise the political space in which positioning takes place (see Figure 5). Subsets of the triplets centered on a specific candidate can also be used to create candidate-specific 'ideographs', summarising the political position of an actor (as further demonstrated in the Appendix).

The second important point is that our network is constructed out of media reporting rather than from direct quotations from candidates. In that respect, we present how the online news 'mediascape' has represented the positioning of Obama and Romney regarding all issues that were deemed newsworthy.

The ideographs constructed on endorsement and opposition claims can reveal if 'issue ownership' occurs. The theory of issue ownership rests on the proposition that due to long-standing party reputations, candidates are more likely to be perceived as having credibility over certain issues 'owned' by their party (Campbell et al., 1966; Hamill et al., 1985). For instance, voters see Republicans as being better able to handle foreign policy and government management issues, while the Democrats are perceived as more adept on civil rights and social welfare issues. Prior research suggests that these partisan issue associations are consistent and well defined (e.g. Hayes, 2008; Petrocik, 1996). However, candidates do occasionally venture into 'enemy territory'. Specifically, Downs (1957) argues that to appeal to the largest segment of voters, candidates in a two-party system should cast some policies into the other's territory in order to convince voters that their net position is closer to them. Thus, instances of issue trespassing should be expected as a function of normal campaign strategy (Hayes, 2008). The construction of a semantic graph, as previously described, which represents the campaign and the two candidates' issue networks, has the additional benefit of allowing the identification of issue trespassing.

In this study, we formulate a number of typical research questions from a social scientific perspective within the context of our network approach to text and media data. The questions that inform this study are:

1. How did the media 'as a whole' report and represent the campaign and the political positions of Obama and Romney? How are the candidates positioned in a political space by their actions of support and opposition?
2. What do the semantic signed graphs of the two candidates reveal to us about their communication strategies and positioning?
3. What can the structure of an election's semantic graph reveal to us, both in general and in this particular instance?

## Topological properties of a semantic graph

The set of SVO triplets that we have extracted from the news corpus can best be analysed when linked together to form a network, which we call a 'semantic graph'. Graph partitioning methods can be deployed to embed actors in a political space based on their relations with other actors. The methodology is described in Appendix A8. It is a remarkable observation that the key political fault-lines are readily detected by this approach.

We analyse other global properties, such as structural balance, assortativity and hierarchy, as well as the centrality of nodes. The former provide insights about the structural features of semantic graphs. The latter reveal the roles played in the election by specific actors or issues in the political 'fields'.

There are a number of well-known measures of centrality of nodes in a network including: betweenness centrality, closeness centrality, out-degree, in-degree, pagerank, hubs and authorities. We select some of these in order to capture certain aspects of narrative centrality. Different measures of centralities help us in identifying nodes and their roles in the candidates' semantic graphs. For example, degree centrality reveals the most salient issues or actors; betweenness centrality identifies issues or actors that have a bridging function between clusters of issues/actors; and in-closeness centrality helps us identify the nodes that act as 'contextualisers'. We will discuss the interpretation of centralities in further detail in the section entitled 'Detecting the central actors in the semantic graphs'.

## Methodology and data

Our software pipeline (Sudhahar et al., 2013) detects election-related articles in large corpora, parses them, extracts SVO triplets, and identifies key actors, objects and actions. The triplets are then used to form a network and the network is analysed. A detailed description of the pipeline architecture is available in Appendix A5. The two guiding principles were for us to be able to re-use existing tools where possible, and to build a system that can be scaled up to large corpora. The overall methodology represents an innovative combination of existing techniques and tools applied to a novel research aim that extracts meaningful networks out of large textual datasets.

### Dataset

The dataset used for this study is obtained using our infrastructure, which is a media monitoring system (Flaounas et al., 2011) that gathers and analyses news items from the web from a large number of news outlets. English-language news articles related to the 2012 US elections were identified using an automatic text classifier. The classifier was a support vector machine (Cristianini and Shawe-Taylor, 2000) trained and validated using the election news feed from Yahoo!, where it achieved 83.46% precision (correct classification) and 73.29% recall. The dataset comprises 490 US and international media outlets in English for a total of 81,023 articles during the primaries (1 January 2012 to 1 August 2012); and 444 US and international media outlets in English for a total of 49,190 articles during the election (1 August 2012 to 1 November 2012). In total, 130,213 articles were analysed from January to November 2012 using the SVO triplets method.

Figure 1(a) shows the types of outlets and the number of outlets in each type that were captured in our analysis. The types of news outlets captured were newspapers, broadcast media, blogs and online media and magazines. Figure 1(b) shows the number of articles collected from January to November by the 2012 Elections tagger.

The plot shows peaks of media attention for the elections in January (during the Republican victory in the primary contests in Iowa), September (during the Democrat conventions), October (during the presidential debates) and November (during the election day and Obama's victory).

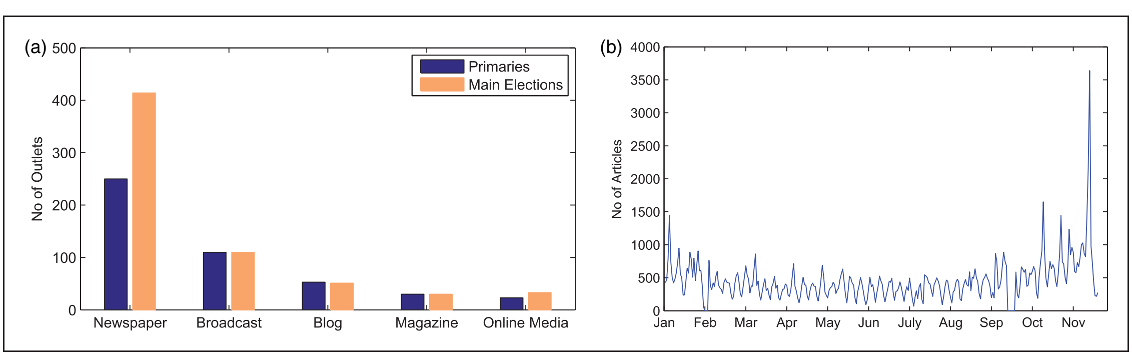


**Figure 1.** Plot (a) shows the number of outlets covered in each type during the 2012 election primaries and main elections. Plot (b) shows the number of articles collected from January to November.

### Filtering of relevant and reliable entities, actions and triplets

As each step of our pipeline may be subject to error, we rely on the large amount of data we collect in order to maintain a clean dataset and correct classification/ extraction. This is done by filtering away uncertain results at various stages of processing and introducing the need to quantify the reliability and relevance of a triplet, as well as that of the entities and their actions. Appendix A4 discusses in detail how this step is performed.

### Positive and negative relations between entities

An important part of our study is that the verbs linking entities are classified into those that show a positive or a negative attitude (the remaining triplets are removed as they are not relevant for our goal of detecting political positions). So once we have identified a set of reliable triplets, we can also assign a weight that quantifies the relation of support or opposition between these two entities. These claims or actions can then be interpreted as 'positioning statements' and be used to define the political position of the actors involved.

We do this by assembling two lists of verbs, one signalling actions compatible with a positive attitude and the other signalling a negative attitude. Verb lists denoting political support/opposition are manually created by going through actions in triplets that were extracted from the elections corpus. Synonyms of these verbs were then added to the lists using an online thesaurus dictionary. Using the verb lists we could count a triplet as a vote in favour of a positive or negative attitude, and calculate a weight for each of the two possible relations. The verb lists are available in Appendix A9.

In order to define the extent to which one actor $a$ supports/opposes an object or another actor $b$, we need to combine the number of positive and negative statements observed in the data connecting node $a$ to node $b$. We model this as the probability of observing a positive or a negative verb, when $a$ is seen as acting on $b$, therefore reducing this to the estimation of a Bernoulli random variable. Bearing in mind the uncertainty introduced by using finite (and possibly small) samples of observations, we consider also the confidence of our estimation of the parameter of the Bernoulli distribution associated to each oriented pair of actors $(a,b)$. This is done by using the Wilson score coefficient, a standard approach in statistics, as described in Appendix A2.

Once the entities and their relations are quantified with the above methods we assemble them in a network in which nodes represent entities/objects and edges represent positive/negative weights.

## Results and discussion: Analysis of networks

At the end of the software pipeline described above, we have obtained 295,164 distinct triplets formed by 41,739 distinct actors and 3677 distinct verbs. After filtering the actors and the triplets to identify the most relevant and reliable ones, we are left with only 2275 distinct triplets (of course, observed multiple times) and we kept only the 406 most significant actors and 609 (positive and negative) links.

The most reliable actors or links are obtained by comparing their relative frequency in a topic-specific corpus with that of a separate, static background corpus or by their absolute number of mentions in the text (explained in Appendix A4). However, this measure of relevance is not appropriate in all domains. When key entities in the topic-specific corpus are also very widely used in the background corpus, the relevance weighting will fail to give greater weighting to the most important entities. In this case, key entities can be identified by their absolute number of mentions in the text as we do for this study. Therefore, we use the top 100 most frequent entities in the 2012 elections corpus as key entities. Triplets containing these key entities are considered reliable if they have been seen in more than $k$ independent documents. The decision on $k$ is explained in Appendix A6. Finally, triplets containing these key entities are used to create the network and this network contained the most significant actors and links.

Figures 2 and 3 illustrate the complete signed and directed networks obtained from our pipeline during the primaries phase (January to August) and the main elections phase (August to November) in 2012. The nodes in the network represent key entities and edges represent positive/negative weights. The green/red colour on the edges indicates positive/negative relations and the size of the node represents the degree of the node.

During the primaries we can see many hubs showing different campaigns taking place within the Republican camp ('Romney', 'Santorum', 'Gingrich', 'Perry', 'Cain') besides the Obama campaign, while after the conventions we see only the two huge campaigns by 'Obama' and 'Romney', who were the dominant players.

## Network partitioning and embedding

As discussed above, candidates define their political positioning through a complex set of claims through which they either endorse or oppose issues and actors, which are captured in the network. Their positioning takes place in the public sphere and is reported by the media. We use this reported 'political space' to map such positioning, and the main challenge is whether

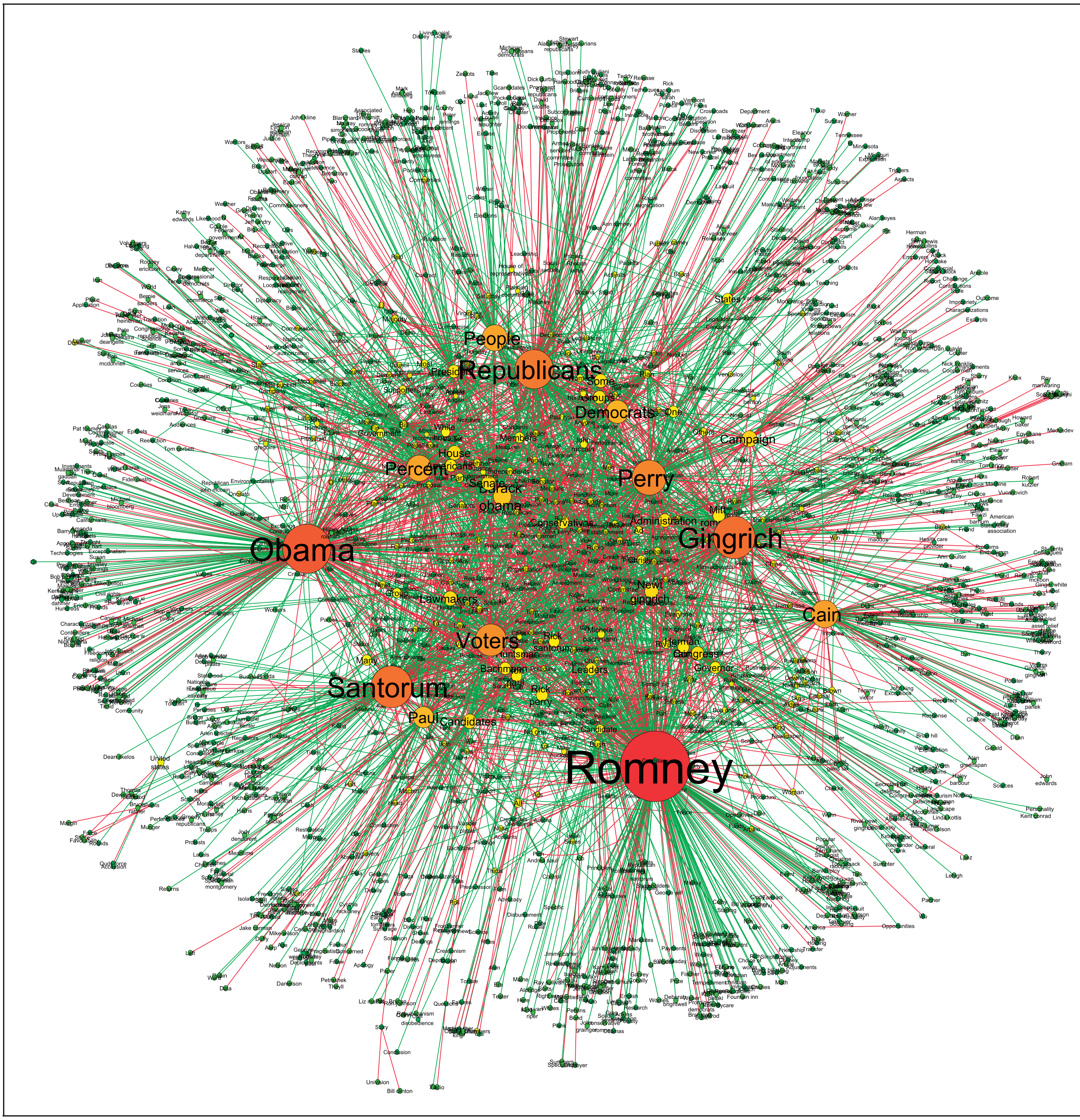


**Figure 2.** Network with positive and negative edges between entities (US presidential election data: Primaries 2012).

we can gain insights about this 'space' from the network of triplets.

A first analysis that can be performed on the network is its partitioning by means of spectral decomposition (see Appendix A8), which also has the effect of assigning each node a 'degree of membership' to one of the two partitions. The results show that the two candidates and the two parties are naturally separated by this method, a finding that provides a powerful sanity check for the entire pipeline, as well as providing an embedding of actors into a 'political space' where proximity is correlated to the amount of support (directly and indirectly) linking them (Figures 4 and 5). This also allows us to assign topics to one of the two camps: for example, 'Israel' was portrayed as being closer to the Republican camp in 2012, while 'Abortion' was closer to the Democratic camp. The sub-networks resulting from the partition are shown in Figures 6 and 7.

### *Assortativity, hierarchy and structural balance*

The next step is to analyse the overall topological properties of the entire network. Assortativity refers to a preference for a network's nodes to attach to others

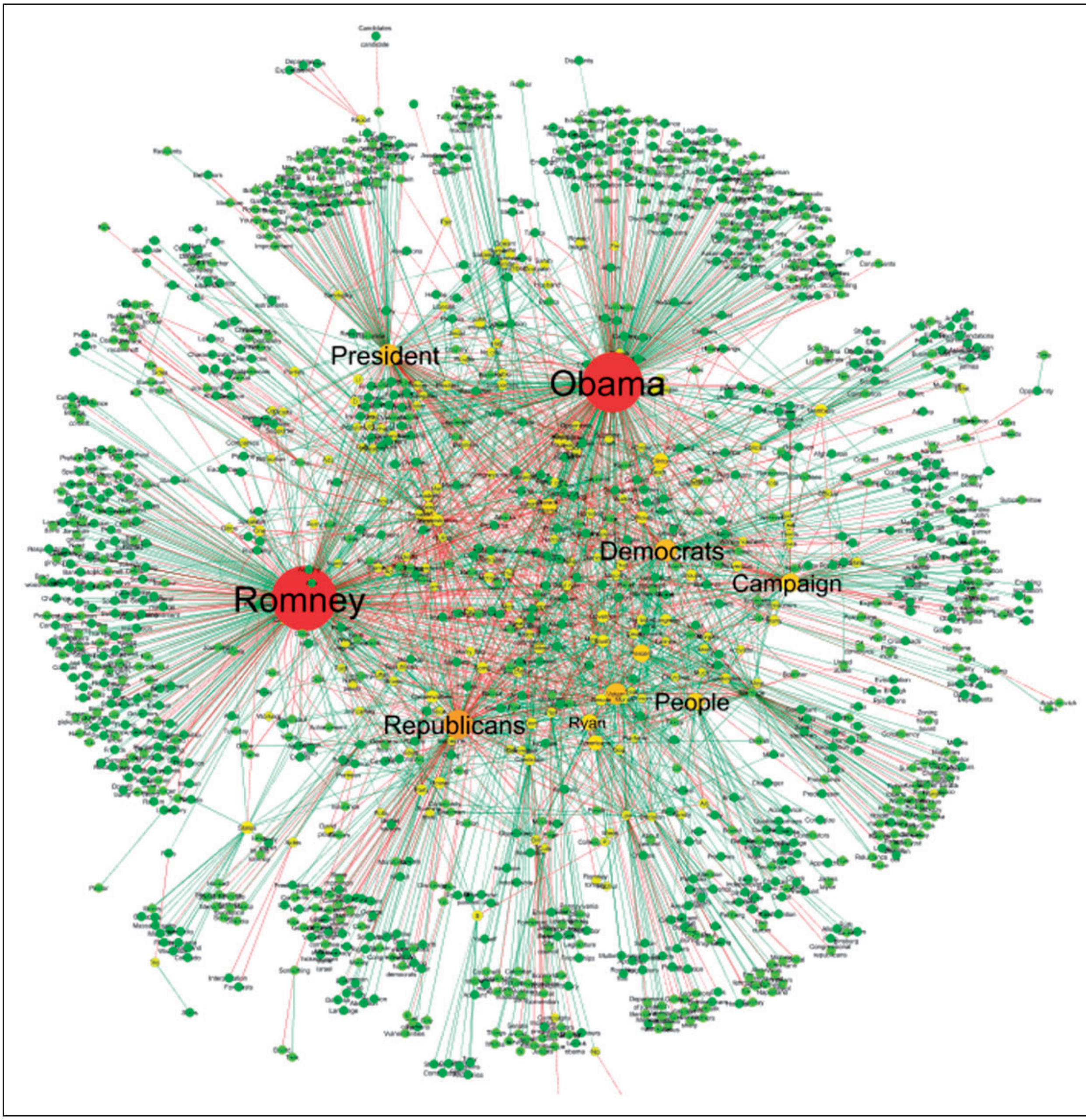


**Figure 3.** Network with positive and negative edges between entities (US presidential election data: After conventions 2012).

that are in some way similar. Though the specific measure of similarity may vary, network theorists often examine assortativity in terms of a node's degree. Correlations between nodes of similar degree are often found in the mixing patterns of many observable networks. For instance, in social networks, highly connected nodes tend to be connected with other high degree nodes. On the other hand, technological and biological networks typically show disassortative mixing, or dissortativity, as high-degree nodes tend to attach to low-degree nodes. We measured assortativity and hierarchy, finding that the network is not assortative but hierarchical. The interpretation of such structural features is open because there are no clear pre-existing studies about semantic networks and their assortative mixing. We further develop this point in the discussion, while the technical steps carried out in our testing are reported in the Appendix.

We also analysed the structural balance of the network (Cartwright and Harary, 1956; Davis, 1967). This notion (defined in more detail in the Appendix) attempts to capture the natural property of networks

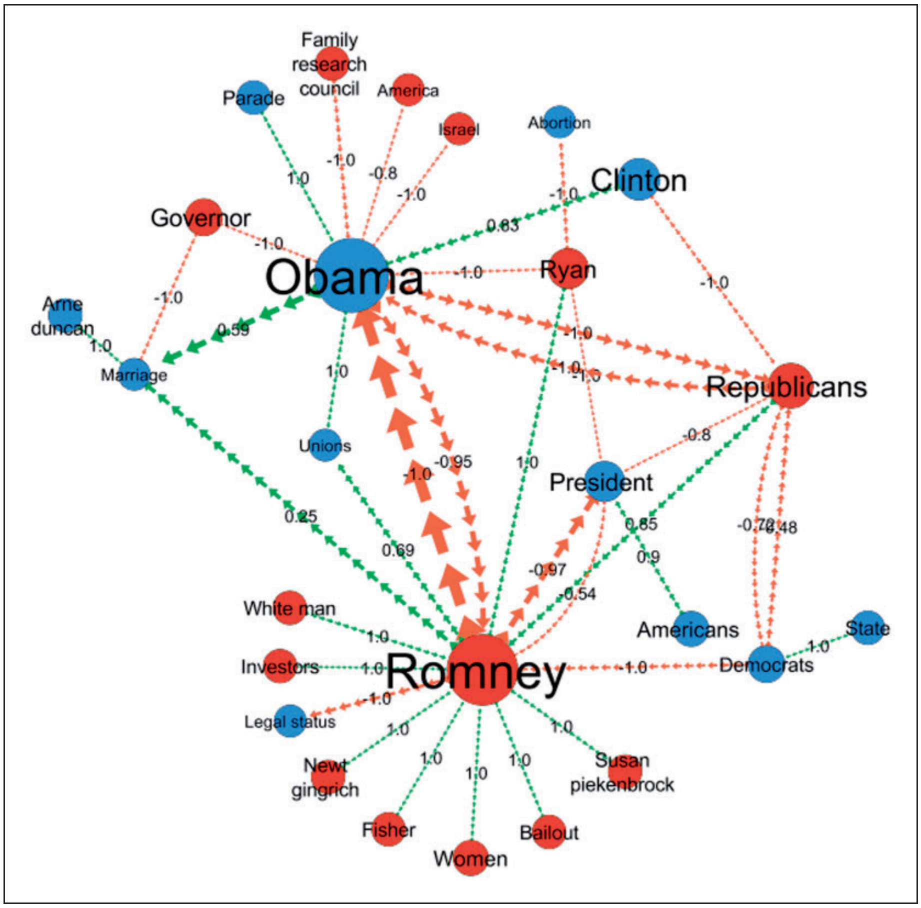


**Figure 4.** A subset of the election network, coloured by partitioning it via the first eigenvalue of the symmetrised adjacency matrix (see Appendix A8). Note that the split captures well the expected distinction between the Republican (red) and Democratic (blue) camps. The orange and green links show negative and positive relations between entities.

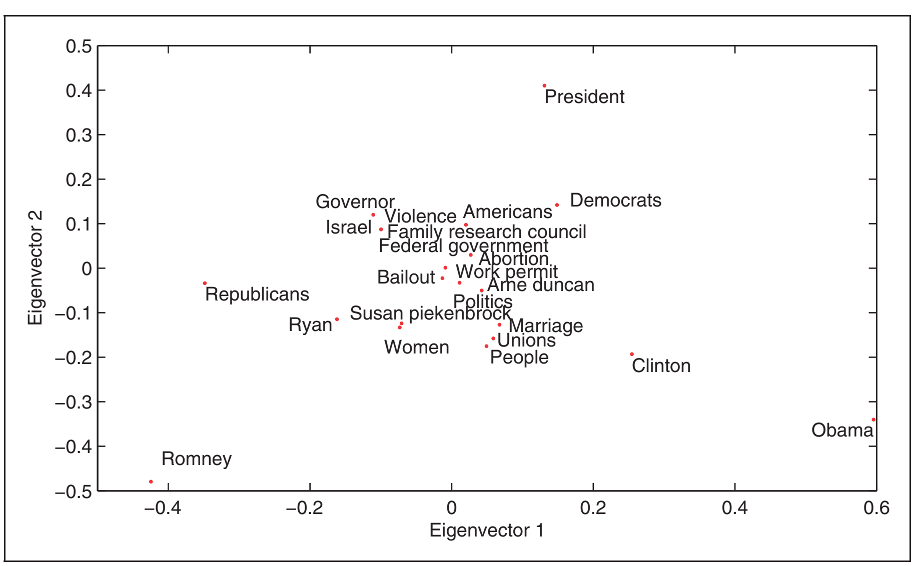


**Figure 5.** The embedding of actors in the political space generated by the network of statements.

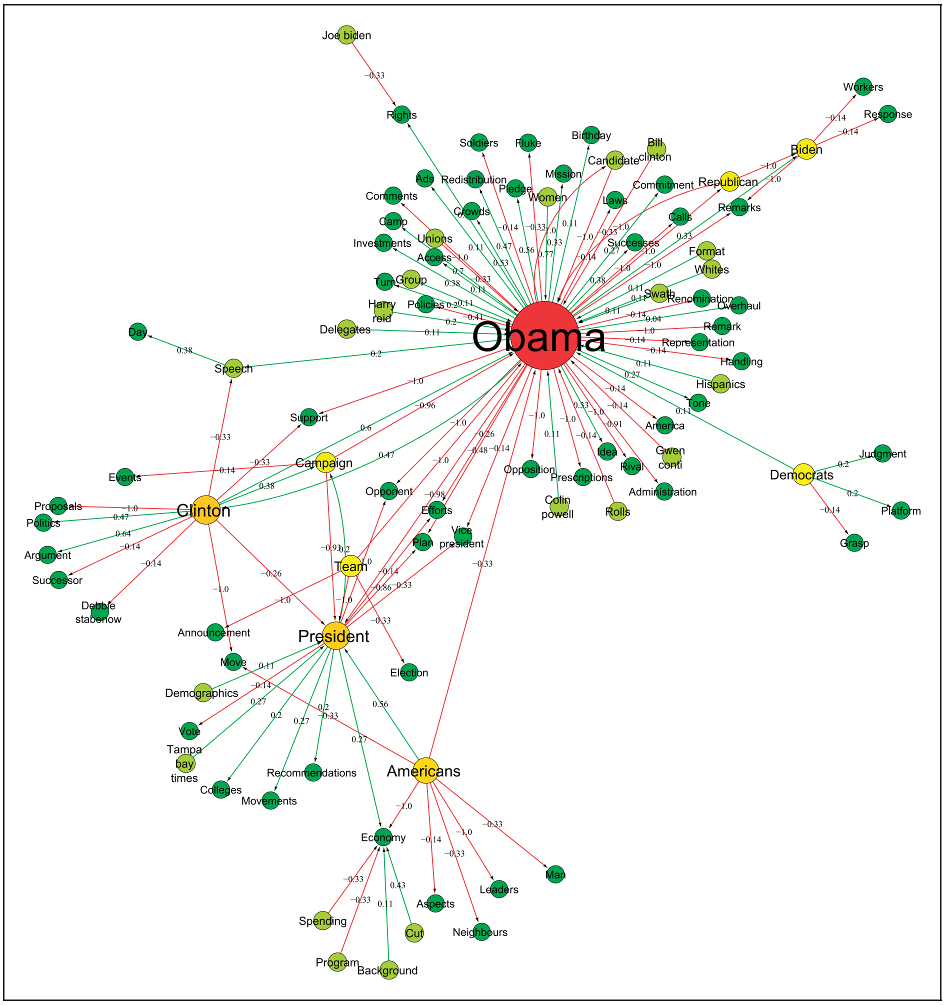


**Figure 6.** Obama's network after the conventions. Dark colours on nodes indicate high degree and light colours indicate low degree.

expressing the attitude (positive or negative) among nodes: reciprocity (if A likes B then B should like A) and transitivity (if A likes B, and if B likes C, we would expect that A likes C).

The results show that while at the level of dyads (pairs of nodes) there is strong balance (or reciprocity), at the level of triads (three nodes) there is significant imbalance. One of the main reasons for this is the presence of many triads where the two candidates profess to support the same issue, while still remaining critical of each other (the technical Appendix reports on the detailed definitions, findings and statistical tests associated with structural balance).

### *Detecting the central actors in the semantic graphs*

The next step is to analyse each of the two candidates' networks. One of the most important questions is the identification of the 'central' issues or actors in each political camp during the campaign as reported by the media data that we have collected. Given the complete network of all actors and their relations, we explored various measures to quantify the centrality of actors.

The transposing and interpreting of centrality measures into semantic graphs is still a work in progress and there is no completely validated way of doing this. This is particularly true in our case, in which we go beyond the use of simple co-occurrences to generate the links between nodes. We selected a number of centralities that have been used before in the context of semantic graphs (Doerfel, 1998; Kim, 2011; Yuan et al., 2013). We used the degree of a node as the basic measure of prominence, as it indicates the number of other actors or issues a given actor directly interacts with (Freeman, 1979). This often – but not necessarily – correlates with the frequency with which the actor is mentioned in the corpus.

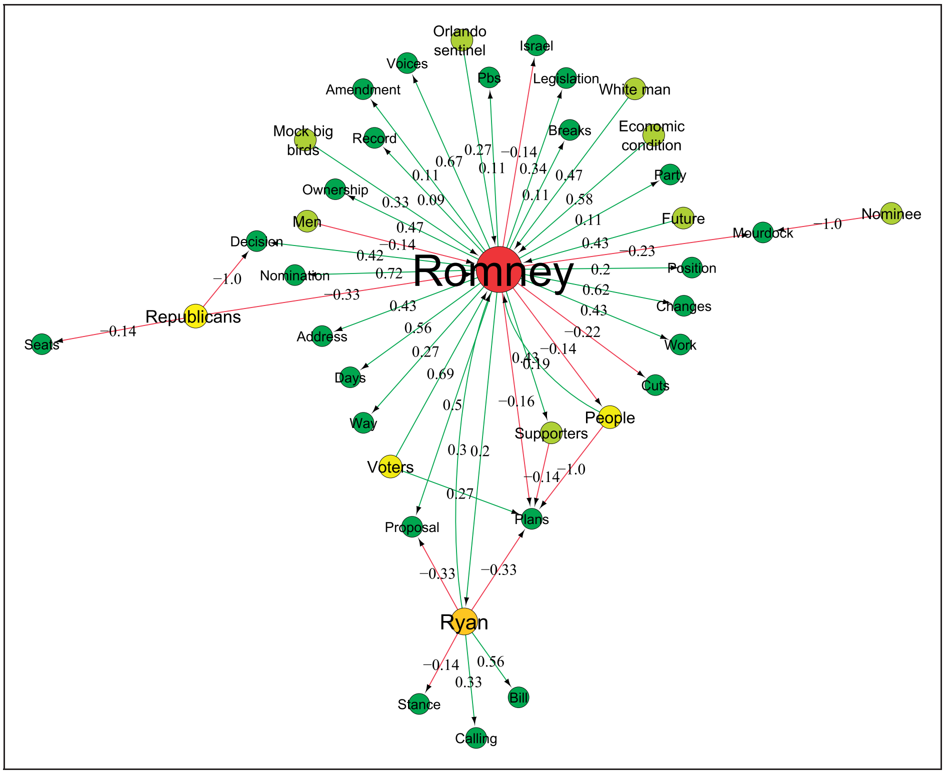


**Figure 7.** Romney's network after the conventions. Dark colours on nodes indicate high degree and light colours indicate low degree.

Because we have a directed network, we can compute the *in-closeness centrality* of nodes (Carrington et al., 2005), which measures the degree to which a node can be easily reached 'from' other nodes (i.e. using edges coming in towards the node) where 'easily' means the shortest distance. In a semantic graph, a node with a high in-closeness positions itself closest to the centre of reference playing a 'connotative' role or, in other words, a 'qualifier' or 'contextualiser'. *Betweenness* has been measured to identify those nodes that are crucial in holding together the network and connecting different clusters of issues or actors (the nodes with high betweenness centrality are translated as active mediators in the communication).

Last, we used the notion of *hubs and authorities*: a high authority actor receives links from many good hubs (highly connected actors). In the context of our network, we interpret authority as a proxy for an agenda-setting item in both active and passive terms[1] (further technical details are available in the Appendix).

We can also apply the tools described above to analyse the two sub-networks obtained by partitioning the overall election network (see Appendix for the partitioning method). The detailed results of that analysis are presented in the Appendix A10, but the main findings are as follows.

*Democratic camp.* The highest degree objects in the Democratic campaign are: Obama, President, Clinton, Americans, Economy and Campaign. This indicates that the economy was the most salient issue upon which the campaign focused, with the discussion of a 'Plan' to recover the US economy. The objects with the highest *betweenness* centralities in the Democratic network were: Obama, President, Campaign, Clinton, Americans, Biden and Speech, showing how these entities were the bridges linking the otherwise separate aspects of the electoral narrative. The objects with the highest *in-closeness* centrality in Obama's network include: 'Campaign', 'Law', 'Redistribution', 'Pledge', 'Ads', 'Overhaul', 'Biden', 'Investments', etc. These objects are a mixture of policy and economic tools (law, redistribution, overhaul of a policy, investments) and contexts (Campaign, Ads, Pledge). Objects with high *authority* can help us to define the agenda-setting efforts of each of the two parties. In the Democratic network, we can find: Obama/President, Economy, Plan, Rights, Proposals, Efforts, Biden/Vice-President, Speech, Remarks, Election, Policies and Campaign. Besides the expected presence of Obama and Biden, the other objects are indicative of a campaign strongly focused on the economy and on actions and tools (plan, proposals, efforts, policies) combined with an emphasis on rights. In the above discussion about issue ownership during US presidential campaigns, the latter topic has traditionally been a Democrat theme. It appears that the 2012 campaign was very much focused on the 'Economy' and 'Rights' for the Democrats – the agenda-setting effort was directed towards these two issues.

*Republican camp.* The top two objects in terms of *degree centrality* in the Republican network are Mitt Romney and Paul Ryan, followed by two nouns: People and Plans, which refer to the Republican initiative of prescriptions for the US economy. In the case of the Republican network, the objects with high *betweenness* centralities are all actors: 'Romney', 'Ryan', 'People' and 'Republicans'. It was also the case for the Democratic network that actors were the most bridging objects. However, it appears that for the Republican network there are the two candidates – Romney and Ryan – and then the 'People' or the Republican Party. In the Republican network, we find high in-closeness centrality for objects such as: 'Amendment', 'Work', 'Proposal' and 'Breaks'. Some of them come from the Republican campaigns but others are forced upon them. The highest object in terms of *authority* is 'Romney'. This is followed by 'Plans', 'People', 'Decision' and 'Proposal', all of which are keywords of Romney's campaign and language. However, there are two objects suggestive of a reaction rather than an initiative in terms of agenda setting: 'Mourdock', 'Bill' and 'Public Broadcasting'. All three objects are related to attacks from Obama's campaign (see Appendix). 'Israel' is also present among the objects with high authority and refers to Romney's speech in Israel about US foreign policy in the Middle East. The remaining objects with high authority are what now with hindsight appear to be important issues in the 2012 Republican campaign: tax breaks (Breaks), the amendment to the federal marriage law (Amendment) and the work requirements for welfare benefits (Work).

## Positive and negative statements, and divisive issues

Actors can choose to signal their position in the political space by either making negative statements about what they do not stand for, or positive statements about what they do stand for. We have considered how often an actor is reported as making positive or negative statements about another actor or issue. Similarly, we have considered how often an actor is the target of positive or negative statements or actions. In other words, we have measured how often actors position themselves by their opposition to or support for something or someone. We also measured the extent to which an object is 'divisive' – in that it attracts very different sets of statements – and therefore how important it is for political positioning by both parties.

*The most divisive subjects and objects in the network.* Since we have positively and negatively signed networks, it is possible to assess the divisiveness of subjects or objects in the network. This will examine the target or source nodes of diverse sign links (positive and negative) or same sign links (positive or negative) in the network. To identify the most divisive objects, for each node we take into account all the incoming edges and count how many were positive (p) and negative (n). Treating this as a Bernoulli random variable, we use its variance to measure how divisive the object is. More details on this are given in the Appendix A3.

Table 1 shows the most divisive subjects and objects in our network. Among the subjects we find the main actors of American politics and of the 2012 campaign. The presence of the GOP (the Republican Party) at the top of the list confirms what was discussed previously about how much more divisive the Republicans had to be by comparison to the Democrats. We detect that the most divisive objects in the campaign concern either the

**Table 1.** Most divisive subjects and objects.

| Subjects | Objects |
|---|---|
| GOP | Access |
| Parade | Breaks |
| Congress | Clinton |
| Debate | Economy |
| Governor | Law |
| People | Public Broadcasting Service |
| President | Platform |
| Speech | Proposals |
| Obama | Supporters |
| Clinton | Biden |
| Romney | Campaign |
| Democrats | Rights |
| Ryan | Speech |
| Biden | President |
| Americans | Decision |
| Republicans | Election |
| ... | Obama |
| ... | Plans |
| ... | Message |
| ... | Romney |
| ... | Ryan |

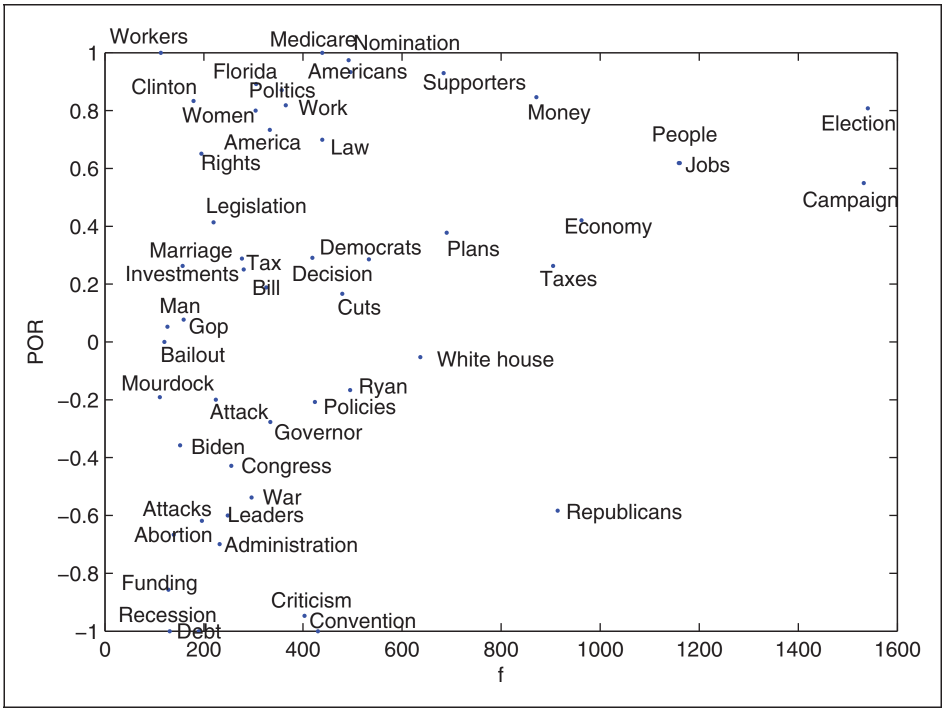


**Figure 8.** Overall positive objects ratio vs. frequency.

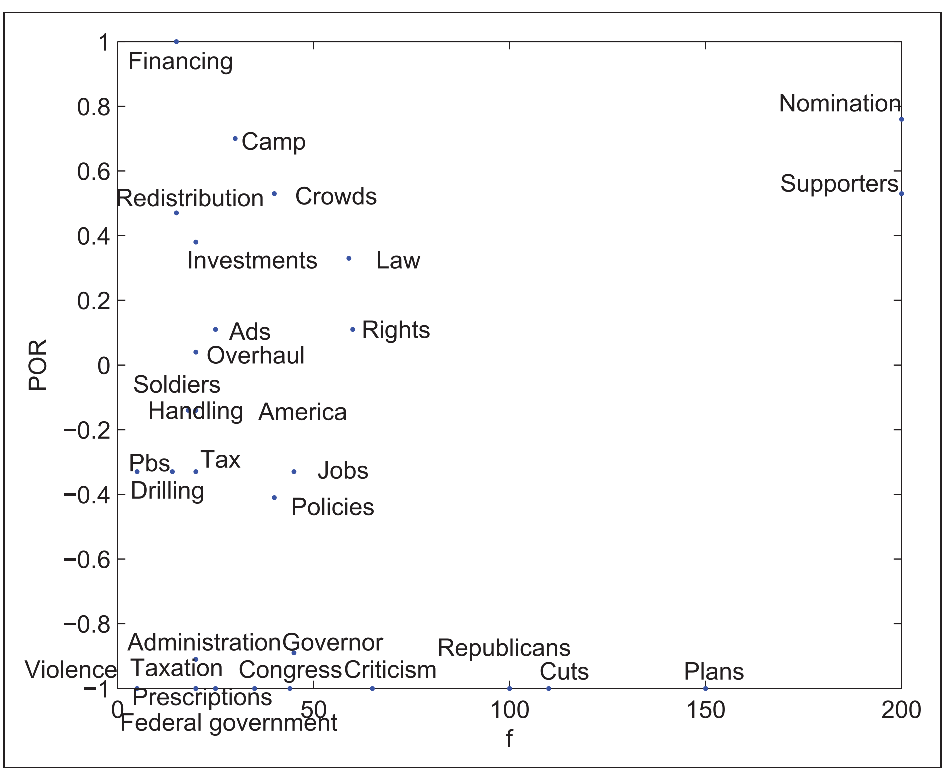


**Figure 9.** Candidate Obama: overall positive objects ratio vs. frequency.

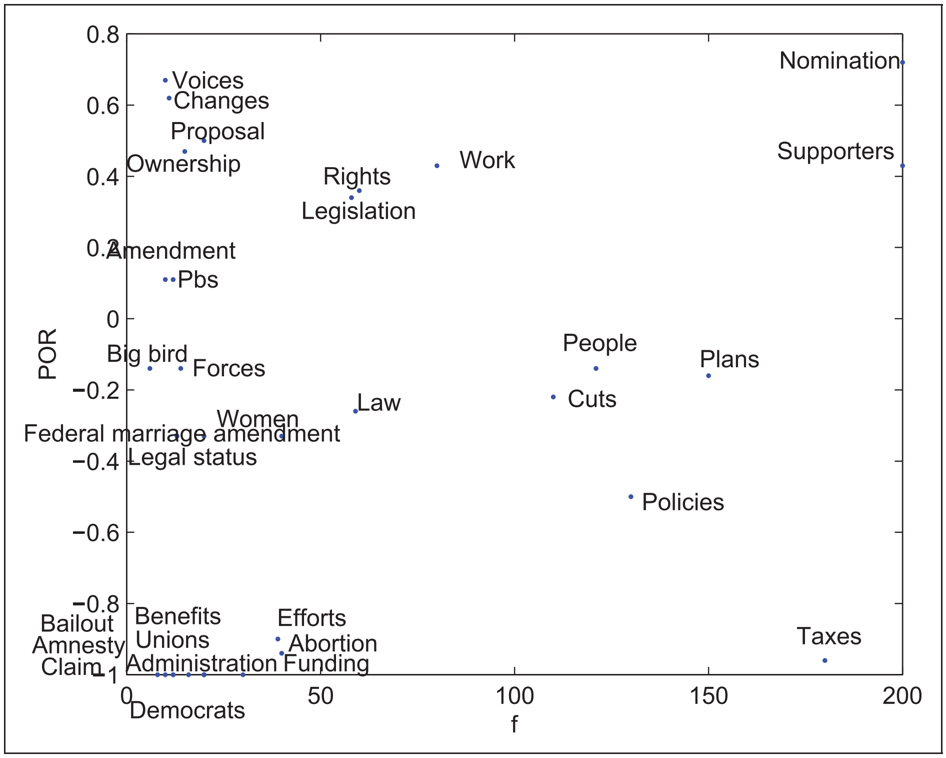


**Figure 10.** Candidate Romney: overall positive objects ratio vs. frequency.

economic split between the two camps (tax breaks, economy) or the social split between them (gay marriage), while it is also confirmed that the Clintons remain very polarising figures.

*Positive object ratio analysis of the election network.* We experiment with positive object ratios (PORs) for Obama and Romney, which could reveal their campaign representations in the media dataset. We also analyse this for the overall network. The POR for a candidate is computed using equation (1). $P_o$ refers to the number of positive statements/expressions made by the candidate about an object and $N_o$ refers to the number of negative statements/expressions made by the candidate towards the same object. The overall POR is computed by taking into account the total number of positive/negative statements towards an object.

$$POR = \frac{P_o - N_o}{P_o + N_o} \tag{1}$$

From Figure 8, a number of interesting questions can be answered. For example, was the media reportage of the 2012 electoral campaign biased towards one of the two parties? The overall positive objects ratio plot in Figure 8 provides the answer, albeit within the limitations of the methodology applied in this study. This quantity reflects the number of times an actor has been the object in a SVO triplet with a positive verb. The two objects 'Democrats' and 'Republicans' are present in different quadrants: the 'Democrats' are reported as being (slightly) more often the object of positive statements or actions, while the Republicans are more often the object of negative statements or actions.

From Figure 8, we also learn that both vice-president candidates, Biden and Ryan, had mildly negative reporting. Above all, the chart confirms the extreme importance played by the status of the US economy as the main issue towards which the campaign gravitated. 'Economy', 'Money', 'Jobs' and 'Taxes' are the main focus of the reporting and also of both candidates.

## Comparing the two campaigns

*Democratic camp.* Barack Obama's ideograph provides us with Figure 9: on the Y axis we have the POR of objects linked to Obama (the weight vector for outgoing links only) and on the X axis their frequency. The two top items that were most positively referred to by Obama are 'Nomination' and 'Supporters' and this is not particularly surprising. The other objects present in Figure 9 reveal a more interesting picture.

As discussed previously, issue ownership in US presidential elections traditionally assigns an advantage to the Republican Party in using the economy as a campaign issue (Campbell et al., 1966; Hamill et al., 1985). However, the 2012 elections saw an 'issue trespassing' strategy with President Obama taking the initiative on the 'economy'. This is evident from the fact that objects such as 'financing', 'redistribution' and 'investments' have a high POR.

The other defining issue related to the economy was jobs creation: the object 'jobs' has a mildly negative ratio indicating that the media reported Obama outcomes in jobs creation less effectively. Further, two important and defining issues of the campaign, represented in Figure 9, are 'Rights' and 'Law'. 'Rights' refer to the legal recognition of same-sex marriage. 'Law' also refers to the amendment to the federal marriage law. These issues were important for President Obama but they were also an important focus of Mitt Romney's representation in the media.

On the negative part of the chart, we found a rather frequent disapproval of 'plans'. This is due to the topic of the Republican 'economic plans'. The US economy and economic policy is also highly relevant in the negative part of the chart, with a perhaps obvious negative link to the recession and with the Republican recipe for federal government budget cuts.

*Republican camp.* The two top positive objects in Romney's signed network according to Figure 10 are 'Nomination' and 'Supporters', as is the case for Obama's signed network. The other very frequent and positive objects are: 'Work', 'Rights', 'Legislation', 'Ownership' and 'Proposal'. 'Work' refers to the debate over the work requirements necessary to access welfare benefits: Romney, who strongly supported this measure, accused Obama of wanting to remove such requirements.

'Changes' is also a highly positive object linked to Romney in the mass media data. It refers to one of the mottos of Romney's campaign about changing the current status of the US and trying to re-appropriate the concept of change from Obama that was used as a slogan in his first presidential campaign ('Change').

At the opposite end of the chart we find the item most negatively linked to Romney: Taxes. This relationship reflects the great emphasis given by Romney and the Republican Party to tax cuts and breaks as the economic solution to recession, something confirmed by the presence of other relevant objects in Figure 10 ('Benefits', 'Cuts', etc.).

Romney is negatively linked to the object 'Unions', perhaps unsurprisingly, because of his critical comments about US labour unions and in particular of his opposition to the Employee Free Choice Act. The object 'Bailout' refers to Romney's statements against the car industry federal bailout in Michigan (his home state), which was a major moment during the campaign as his opposition created problems for his own campaign.

The object 'Amnesty' refers to the amnesty given to illegal immigrants by the Obama administration. This issue is negative and linked to Romney because his position on this issue was confused during the campaign; from an initial critique he later declared that he would not revoke the amnesty. Media data widely reported his ambivalence on the topic.

Among the non-economic issues, three topics have been identified in the negative quadrant: abortion, the amendment to the marriage law and women. The first two are issues about civil rights and are negatively linked to Romney. The difference between Obama and Romney here is evident: Obama continued the 'issue ownership' of the Democrats on rights in addition to his personal history while Romney re-affirmed traditional Republican positions on abortion and marriage.

## Conclusions

Analysing the media coverage of the US election using our Big Data approach revealed the strategic positioning of actors around issues in the 'mediascape' of online English language news. We used a semantic graph obtained by parsing the text, identifying the noun phrases and the verbs and linking them and used SVO triplets as building blocks for a network. This approach is innovative and has never been applied to a real-world dataset of this scale. Using media data and transposing signed relationships in a graph, we uncovered the unique mixture of endorsements and disapproval that constitute the Republican and Democratic camps. It is a remarkable finding that the spectrum of political positions can be reliably recovered from the set of claims attributed to each actor by the media reporting. The split of the network into the two main camps provides strong evidence that the main political relations can be extracted by this approach.

Results suggest that the 2012 campaign was characterised by Obama's strategy of defending his record on the economic policy and 'owning the issue', by his being 'on the attack' about a number of issues and forcing Romney and the Republicans to be defensive about salient issues. The set of triplets show that the 2012 campaign was very much focused on the issues of the US economy and civil rights for the Democrats. The agenda-setting effort was directed towards these two areas. Obama challenged the traditional Republican issue ownership about the economy. Other major actors who played an important role for Obama and Romney were Bill Clinton and Paul Ryan.

Overall, media reporting contained more frequently positive statements about the Democrats than the Republicans. Overall, the Republicans were more frequently the object of negative statements either by Democrats or other actors. The Republican Party (GOP) is the most divisive subject in the campaign, and is portrayed in a more negative fashion than the Democrats.

The topological features of these semantic graphs are less straightforward – and perhaps more interesting to interpret. The network's structurally unbalanced nature marks a difference with social networks. The presence of unclosed triplets indicates that latent and unrealised ties might be common in semantic graphs in which there is less obligation to be 'coherent'. Most likely, this is because of the specific event under investigation: a political campaign. Another structural characteristic was the disassortative nature of the network, with high-degree nodes connecting with low-degree nodes and vice versa. In disassortative networks, well-connected nodes join a much larger number of less well-connected nodes. This is typical for biological and technological networks. Real-world social networks are usually assortative because in real life[2] everyone in a society would like to connect with elites, while the elites would rather communicate with those who share the same social status. In our case, we studied a political campaign and proposed that a disassortative network might represent the network translation of a mobilisation campaign operated by the candidates.

This study was made possible by the use of a computational infrastructure for the collection and analysis of 'Big Data' representing an example of the application of techniques to gain insight about societal phenomena. AI and text mining techniques were used to detect election-related articles, parse their content, extract the key actors in a narration and their relations, forming a network whose topology has then been analysed. We believe that our methodology represents a substantial step forward in the linguistic analysis of texts by means of extracting relational data. The future will include making better use of the information coming from the parser, which can go well beyond the simple SVO structure of sentences and focus also on automating the creation of positive/negative verb lists using the tool OpinionFinder (Wilson et al., 2005).

A network approach to text based on semantic graphs is a promising approach to analysing large corpora of texts and, with a clear application in mapping 'mediascapes', in cases of specific events and potentially also longitudinally.

### Declaration of conflicting interests

The authors declare that there is no conflict of interest.

### Funding

This work was supported by EU-funded research projects CompLACS (FP7-ICT 270327); and ThinkBig (FP7-IDEAS-ERC 339365).

## Notes

1. The rationale behind this is that highly connected nodes – in our case mostly the two candidates – make constant reference to these nodes (objects). In the context of a campaign, this may be interpreted as the outcome of an agenda-setting effort. However, some objects might appear as having high authority because of the need for a candidate to respond to a direct attack. In other words, authority could be considered a proxy of both active and passive (forced upon) agenda setting.
2. Online social networks can be disassortative (Hu and Wang, 2009).
3. JUNG: http://jung.sourceforge.net
4. JAMA: http://math.nist.gov/javanumerics/jama/
5. See for details: http://thecaucus.blogs.nytimes.com/2012/10/24/republicans-struggle-to-contain-mourdock-comments/
6. It refers to the debate generated during the presidential election generated by Romney's statement about 'Sesame Street', a PBS children's television show

### References

Annis DH (2005) Permutation, parametric, and bootstrap tests of hypotheses. *Journal of the American Statistical Association* 100(472): 1457–1458.

Campbell A, Converse PE, Miller WE, et al. (1966) *Stokes 1960. The American Voter*. New York, NY: Wiley.

Carrington PJ, Scott J and Wasserman S (2005) *Models and Methods in Social Network Analysis*. Cambridge: Cambridge University Press.

Cartwright D and Harary F (1956) Structural balance: A generalization of Heider's theory. *Psychological Review* 63(5): 277.

Conway LG III, Gornick LJ, Burfeind C, et al. (2012) Does complex or simple rhetoric win elections? An integrative complexity analysis of us presidential campaigns. *Political Psychology* 33(5): 599–618.

Cristianini N and Shawe-Taylor J (2000) *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press.

Cunningham H (2002) Gate, a general architecture for text engineering. *Computers and the Humanities* 36(2): 223–254.

Davis JA (1967) Clustering and structural balance in graph. *Human Relations* 20: 181–187.

De Fazio G (2013) *Political radicalization in the making*: *The civil rights movement in Northern Ireland, 1968-1972*. PhD Thesis, Emory University, Atlanta, GA.

Diesner J, Carley KM and Tambayong L (2012) Extracting socio-cultural networks of the Sudan from open-source,

large-scale text data. *Computational & Mathematical Organization Theory*. pp. 1–12.
Doerfel ML (1998) What constitutes semantic network analysis? A comparison of research and methodologies. *Connections* 21: 16–26.
Downs A (1957) *An Economic Theory of Democracy*. New York, NY: Harper.
Earl J, Martin A, McCarthy JD, et al. (2004) The use of newspaper data in the study of collective action. *Annual Review of Sociology*. pp. 65–80.
Erdös P and Rényi A (1960) On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5: 17–61.
Flaounas I, Ali O, Turchi M, et al. (2011) *Noam: News outlets analysis and monitoring system*. ACM SIGMOD International Conference on Management of Data, Athens, Greece.
Franzosi R (1987) The press as a source of socio-historical data: Issues in the methodology of data collection from newspapers. *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 20(1): 5–16.
Freeman LC (1979) Centrality in social networks conceptual clarification. *Social Networks* 1(3): 215–239.
Gaskell G, Bauer M and Durant J (1998) The representation of biotechnology: policy, media and public perception. *Biotechnology in the Public Sphere*. pp. 3–8.
Good PI (2005) *Permutation, Parametric and Bootstrap Tests of Hypotheses*, 3rd ed. New York: Springer.
Habermas J, Lennox S and Lennox F (1974) The public sphere: An encyclopedia article (1964). *New German Critique* 3: 49–55.
Hamill R, Lodge M and Blake F (1985) The breadth, depth, and utility of class, partisan, and ideological schemata. *American Journal of Political Science*. pp. 850–870.
Hassan A, Abu-Jbara A and Radev D (2012) Extracting signed social networks from text. *Workshop of textgraphs-7 on graph-based methods for natural language processing,* Jeju, Republic of Korea, 13 July, pp. 6–14.
Hayes D (2008) Party reputations, journalistic expectations: How issue ownership influences election news. *Political Communication* 25(4): 377–400.
Heider F (1946) Attitudes and cognitive organization. *The Journal of Psychology* 21(1): 107–112.
Horn RA and Johnson CR (2012) *Matrix Analysis*. Cambridge: Cambridge University Press.
Hu H and Wang X (2009) Disassortative mixing in online social networks. *EPL (Europhysics Letters)* 86(1): 18003.
Kim L (2011) Denotation and connotation in public representation: Semantic network analysis of Hwang supporters internet dialogues. *Public Understanding of Science* 22(3): 335–350.
Leskovec J, Huttenlocher D and Kleinberg J (2010) Signed networks in social media. *SIGCHI conference on human factors in computing systems,* 10–15 April, 2010, Atlanta, Georgia, USA, pp. 1361–1370.
Lin D (2003) Dependency-based evaluation of minipar. *Treebanks* 47: 317–329.
Mac Carron P and Kenna R (2012) Universal properties of mythological networks. *EPL (Europhysics Letters)* 99(2): 28002.
McGee MC (1980) The ideograph: A link between rhetoric and ideology. *Quarterly Journal of Speech* 66(1): 1–16.
Moretti F (2013) *Distant Reading*. London: Verso Books.
Moscovici S and Duveen G (2000) *Social Representations: Explorations in Social Psychology*, 41st ed. Cambridge: Polity Press.
Petrocik JR (1996) Issue ownership and presidential elections. *American Journal of Political Science* 40: 825–850.
Petrocik JR, Benoit WL and Hansen GJ (2003) Issue ownership and presidential campaigning, 1952–2000. *Political Science Quarterly* 118: 599–626.
Shannon P, Andrew M, Owen O, et al. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11): 2498–2504.
Sudhahar S, Fazio GD, Franzosi R, et al. (2013) Network analysis of narrative content in large corpora. *Natural Language Engineering*. Available on CJO2013. DOI:10.1017/S1351324913000247.
Wasserman S (1994) *Social Network Analysis: Methods and Applications*, 8th ed. Cambridge: Cambridge University press.
Welch WJ (1990) Construction of permutation tests. *Journal of the American Statistical Association* 85(411): 693–698.
Wilson EB (1927) Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association* 22(158): 209–212.
Wilson T, Hoffmann P, Somasundaran S, et al. (2005) Opinionfinder: A system for subjectivity analysis. *HLT/EMNLP on interactive demonstrations,* Vancouver, 6–8 October, pp. 34–35.
Yuan EJ, Feng M and Danowski JA (2013) Privacy in semantic networks on Chinese social media: The case of Sina Weibo. *Journal of Communication*. DOI: 10.1111/jcom.12058.

# Appendix

## A1 – Data sources

Figure 11 shows the number of news outlets from each state during primaries and after the conventions.

## A2 – Estimation of Bernoulli parameter using Wilson coefficient

A possible approach to quantifying the weight of a relation between entity a and entity b is to consider a confidence interval around our estimate of the value of that relation. This will relate to the estimation of the parameter of a Bernoulli distribution, so that we can then calculate the confidence interval around this

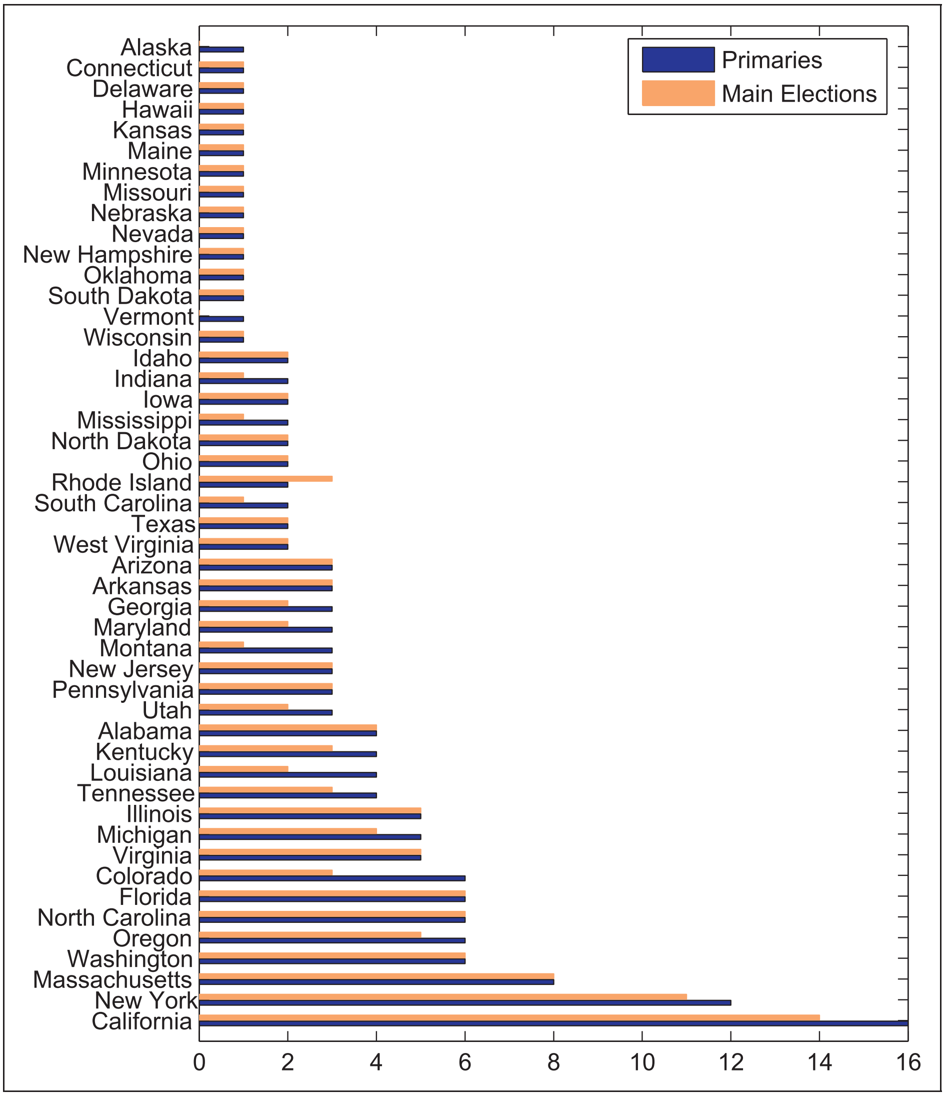


**Figure 11.** Number of outlets covered in each state of the US during 2012 primaries and main elections.

estimate by using standard methods. We use the Wilson score confidence interval (Wilson, 1927) for a Bernoulli parameter given by,

$$w = \left(\hat{p} + \frac{z^2_{\alpha/2}}{2n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2_{\alpha/2}}{4n^2}}\right) \Big/ \left(\frac{z^2_{\alpha/2}}{n} + 1\right) \quad (2)$$

Here $\hat{p}$ is the fraction of positive observations, $z_{\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution and $n$ is the total number of observations. The final score of our links should be associated to a function of the proportion of positive triplets, which is possible in this case since: $(P-N)/(P+N) = 2P/(P+N) - 1$; then observing that $P/(P+N)$ is the rate of positive mentions, we can treat the estimation of this quantity like the estimation of parameters of a Bernoulli distribution. This score is computed using the lower bound of the Wilson interval. For a confidence level of 95% the value for $z_{\alpha/2}$ is 1.96. This can be approximated to 2 and a simplified version of the Wilson score interval can be obtained by considering the number of positive ($P$) and negative ($N$) triplets found between any two entities $a$ and $b$. The final weight on the links becomes,

$$S = 2\left(\frac{P+2}{P+N+4} - \frac{2\sqrt{\frac{P.N}{P+N}+1}}{P+N+4}\right) - 1 \quad (3)$$

As we can see, this range consists of two parts: the mean $m = \frac{P+2}{P+N+4}$ and the interval $i = \frac{2\sqrt{\frac{P.N}{P+N}+1}}{P+N+4}$ on either side of the mean. When a positive/negative relation is supported by many independently generated triplets (from $k$ different documents), $i$ becomes smaller and the resulting network would contain the most reliable information. Hence we can introduce a threshold to the percentage of $i$ and accept relations only if $i$ lies below this threshold. We provide details on how we select this threshold in the validation section.

## A3 – Variance of Bernoulli as measure of divisiveness in objects

We compute the variance score $S_v$ for each node as shown in equation (2). Sorting the nodes according to the score would give us a ranking of objects from the most divisive to the less divisive. The same could be done to find out the most divisive subjects by looking at the outgoing links of nodes.

$$S_v = \left(\frac{p}{p+n}\right) \cdot \left(\frac{n}{p+n}\right) \tag{4}$$

## A4 – Relevance and reliability of entities, actions and triplets

There is a need to quantify the reliability and the relevance of a triplet, as well as entities and actions to reduce the amount of errors in our output. To tackle this problem, we first define both the weight of an entity and action. These weights may differ for different applications and tasks, but allow for the ranking and selection of the most relevant or reliable information to include in the network. Relevance of entities or actions to a specific topic can be gauged by comparing their relative frequency in a topic-specific corpus with that of a separate, static background corpus. Equation (5) shows the weighting method,

$$w_i = \frac{f(i, D_1)}{f(i, D_2)} \tag{5}$$

Where $w_i$ refers to the weight of the entity/action $i$; $f(i, D_1)$ and $f(i, D_2)$ refer to the frequency of the entity/action in a given corpus $D_1$ and a background corpus $D_2$, respectively. The highest weighted entities/actions can be considered as key entities/actions in the given corpus.

However, this measure of relevance is not appropriate in all domains when key entities in the topic-specific corpus are also very widely used in the background corpus, the relevance weighting will fail to give greater weighting to the most important entities. In this case, key entities can be identified by their absolute number of mentions in the text as we do for the elections study. Therefore we use the top 100 most frequent entities in the 2012 elections corpus as key entities. In this study we are interested in analysing networks of positive/negative relationships between entities and hence we use all actions denoting political support or opposition identified in the corpus for the analysis. Triplets containing key entities are considered reliable if they have been seen in more than $k$ independent documents.

## A5 System pipeline

We describe here the software pipeline that we have used in our experiments. The two guiding principles were for us to re-use existing tools where possible, and to make a system that can scale to large corpora. Figure 12 shows the system pipeline. Each component of the pipeline is explained in detail.

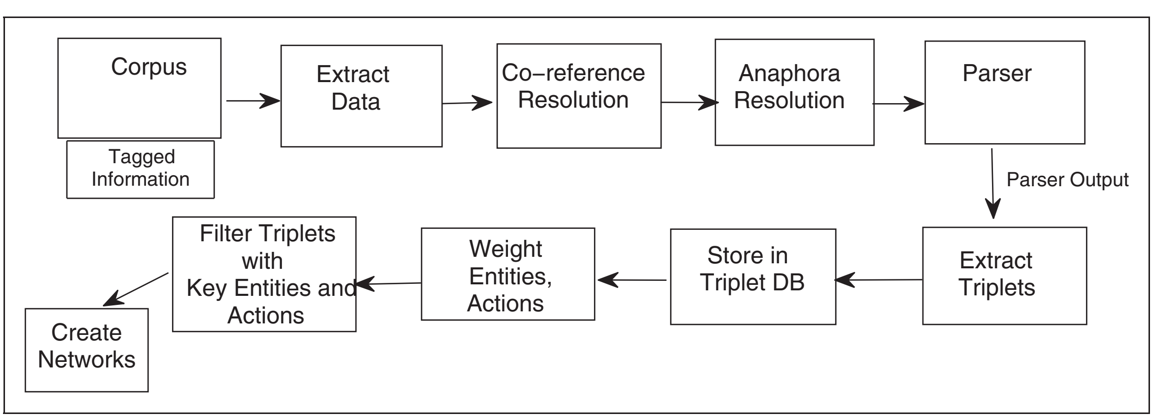


**Figure 12.** System pipeline.

- News corpus – The system uses articles contained in an available news corpus to perform the task.
- Extract data – We could first extract the content from articles that are specific to a domain which is of interest to the analysis, e.g. crime, elections, sports, etc.
- Co-reference resolution – The text in every individual article is processed for named entity co-reference resolution. The Orthomatcher module in ANNIE Information extraction system in General Architecture for Text Engineering (GATE) (Cunningham, 2002) distribution is used to perform this task.
- Anaphora resolution – Once the co-references have been resolved the Pronominal resolution module in ANNIE is used to perform anaphora resolution. The system solves pronouns in all forms that are identified by GATE.
- Minipar parser – We use the parser Minipar (Lin, 2003) to parse the above processed text. The parser tags each word of the sentence with its grammatical relation to it. Minipar has its own limitations since it cannot parse sentences more than 1024 characters long. On the other hand, we found that this length exceeds the size of a typical sentence in the news which is made of approximately 500 characters.
- Extract triplets – From the Minipar parser output we extract words tagged with s (subject), i (verb) and obj (object of the verb) relations. A SVO triplet is formed out of these words if the s, i, obj relations are found in the sentence in this chronological order.
- Store in triplet DB – All extracted triplets are stored in the Triplets database along with the article information from which they were extracted. This includes, article date, title, content and article feed URL. We also store the Minipar parser output for each article.
- Weight entities and actions – Entities (subject/object of triplet) and actions are weighted according to equation (2) and this weight is used to rank and select the highest ranking candidates as key entities and key actions.
- Filter triplets with key entities and actions – We then filter the triplets that have key entities as subjects/objects and key actions as verbs.
- Create networks – Directed networks are created with the triplets where the nodes are entities and the edges are actions linking them. To create networks we use Cytoscape (Shannon et al., 2003) which is a general platform for complex network analysis and visualization. We also used JUNG[3] for automatically generating networks and analysing network properties.
- Weight positive and negative relations between entities – Positive and negative relations indicate friendship or hostility between actors like mentioned before. In order to identify the strength of these relations we introduce a weighting method which is shown in equation (3). This would result in entities linked by a positive/negative link with weights.
- Create signed networks – We create signed networks where nodes are entities and edges are the positive/negative links with weights.
- Spectral graph partitioning – Signed networks are partitioned using spectral graph partitioning methods to assess the degree to which actors/objects belong to or are in favor of one of two parties, in the assumption that the networks are naturally organised into two main communities. We used the JAMA[4] matrix package for java to perform this task.
- List of entities showing partitions – Once the network is partitioned we obtain a list of entities which shows the association of them to one of the two communities in the network.

### A6 Validation

Estimating the performance of our methodology is a difficult but important task since there are no accessible corpora that have been annotated in terms of SVO triplets that we can use in order to measure the precision and recall of our method, and there are no other networks of actors/objects that have been generated by hand, based on a corpus. Nonetheless, we can measure various aspects, in order to increase our confidence in the method, and obtain a rigorous statistical estimate of performance.

### Validation of triplets

We used a corpus covering the Civil Rights movement in Northern Ireland for which 72 manually extracted triplets were available (De Fazio, 2013). We applied our methodology, without filtering 'reliable' triplets due to the limitations of the data. Our method extracted 66 triplets, of which 41 were correct and 31 were missed. This gives us 62% precision and 57% recall in the very unfavourable case when we cannot use any filtering for reliable triplets. This means that there is a 38% probability of a triplet being incorrect, if it has been seen just once. If we use this figure as the error rate for triplets seen once, we can use it in a model for the probability of error in triplets seen more than $k$ times, which would be $0.38^k$. This is true under the assumption that the triplets seen more than $k$ times are independently generated. By only selecting triplets that are seen at least three times we achieve a 5% error rate. But we use higher values for $k$ according to the Wilson score model.

We also analysed by hand 75 triplets coming from the 2012 US election campaign, and checked how many were actually present in the articles that were indicated by our pipeline as supporting them. Of 75, 72 were actually present in the article achieving 96% precision. This gives us a clear estimate of precision after our filtering step, but no estimation of recall, which we expect to be low.

## *Validation of the networks*

Our next validation method was intended to measure the extent to which our analysis captures the division into two camps of the US political actors/objects. We use permutation testing (Annis, 2005; Good, 2005; Welch, 1990), a central part of non-parametric testing. It directly obtains the distribution of the test statistic under the null hypothesis by calculating all possible values of the test statistic under rearrangements of the treatments (labels) on the observed data points. The one-sided $p$ value of the test is calculated as the proportion of sampled permutations where the test statistic is greater than or equal to that in the original dataset.

We obtain the eigenvalue distance $d$ between the 'Republicans' and the 'Democrats' in the entity list and use it as the test statistic. We compare it with distance $d_1$ obtained by taking the distance between the same actors from 100 randomised networks. Here we use two random network models, Erdös–Rényi (Erdös and Rényi, 1960) and Random re-wiring to generate the random networks.

Finally we check for the number of times $r$ that $d_1 \geq d$ in the 100 random networks to calculate the $p$ value which is $r/100$. Table 2 shows the resulting $p$ values obtained for experiments performed on the 2012 and the previous election cycles. It shows that the $p$ value is very low for the same signal appearing by chance. Therefore we could say that the network topology contains meaningful information.

We also checked the effect on $p$ values when different thresholds are introduced to the confidence interval $i$ in our weighting equation (3) which is, $i = \frac{2\sqrt{\frac{P.N}{P+N}+1}}{P+N+4}$. This helped us to select an optimal threshold for $i$. We expect that $p$ value of the resulting network should be $\leq 0.01$ and the network should contain more information as possible. Table 3 shows the $p$ values obtained for different thresholds of $i$ according to the two randomisation models. According to our selection criteria 35% is considered optimal since in both cases the $p$ value is $\leq 0.01$ and also the network is large enough and contains more information.

We also note that the $p$ values are not significant for values of $i$ in two different cases. For example in Table 3 when $i$ is low (13%, 17%, 20%, 23%) we see that the $p$ values are insignificant and this could be because the network contains less information. On the other hand, when values for $i$ are very high (43%) the network contains more noise, which in turn causes the signal to disappear and therefore leads to insignificant $p$ values. Networks for $i = 32\%$ and 35% are the most significant according to Table 3 and 35% fits our criterion.

The statistical significance study on the elections network and the precision estimates performed on the triplets we have extracted point in the same direction: that our system extracts very precise information that represents the true relations among the entities in the corpus. Estimating the recall would be harder, but since we work in the setting of very large datasets, we choose to focus on obtaining high precision rather than high recall.

**Table 2.** $p$ values for distance $d_1 \geq d$ over 100 random networks.

| Year | $p$ value (random rewiring) | $p$ value (Erdös-Rényi) |
|---|---|---|
| 2012 | 0 | 0 |
| 2008 | 0 | 0 |
| 2004 | 0.01 | 0 |
| 2000 | 0.05 | 0.01 |
| 1996 | 0.06 | 0 |
| 1992 | 0.05 | 0 |
| 1988 | 0 | 0 |

**Table 3.** $p$ values for distance $d_1 \geq d$ over 100 random networks for different values of $i$.

| Threshold ($i$) | No of nodes ($n$) | $p$ value (random rewiring) | $p$ value (Erdös-Rényi) |
|---|---|---|---|
| <13% | 37 | 0.01 | 0.08 |
| <17% | 70 | 0.06 | 0.1 |
| <20% | 92 | 0.06 | 0.06 |
| <23% | 142 | 0.12 | 0.07 |
| <28% | 159 | 0.01 | 0.01 |
| <32% | 218 | 0 | 0.02 |
| <35% | 406 | 0 | 0.01 |
| <38% | 414 | 0.01 | 0.02 |
| <43% | 1047 | 0.55 | 0.92 |

### A7 – Properties of networks

A number of statistical properties have been developed to capture features of social networks (Mac Carron and Kenna, 2012). Some structural properties are quantified using the characteristic path length $l$, the longest geodesic $l$max and the clustering coefficient $C$. In this section we will discuss various properties of networks which are prevalent in real world networks and compare it with the results obtained for election networks.

*Hierarchical networks.* Complex networks have a modular structure implying that groups of nodes organise in a hierarchical manner into increasingly larger groups. Hierarchical networks are characterised by a power law dependency of the clustering coefficient $C(k)$ on the node degree according to equation (6). This implies that while nodes with a small degree are part of highly cohesive, densely interlinked clusters, the hubs in the network are not, as their neighbors have a small chance of linking to each other. Therefore the hubs play the important role of bridging the many small communities of clusters into a single, integrated network.

$$C(k) \sim \frac{1}{k} \tag{6}$$

We tested for hierarchical structure in the networks by plotting the average clustering coefficient $C(k)$ vs. degree $k$. Figure 13a shows the plot in a log–log scale. The nodes with a smaller degree present higher clustering than those with a larger degree and the decay approximately follows equation (6). This indicates that the nodes with a low degree are part of densely interlinked clusters, while the nodes with a high degree, which are the hubs, bring together the many small communities of clusters into a single, integrated network.

Candidates' ideograms are attempts at creating an organic network of issues (objects) and actors (subjects). Moreover, the hierarchical structure of the semantic graph represents the hierarchical nature of concepts indicating the presence of overarching hubs that represent semantic domains (for example, the 'economy', 'civil rights', etc.).

*Assortativity.* Assortativity is the measure of the preference of nodes to attach to other similar or dissimilar nodes. Assortativity is high when high-degree nodes tend to connect to other high-degree nodes; it is low when high-degree nodes are linked to low-degree nodes. Assortativity can also be tested by plotting the degree of the neighbours $k_{nn}$ of a vertex as a function of its degree $k$. A positive slope indicates assortativity and a negative slope indicates disassortativity.

Figure 13b plots the average $\langle k_{nn} \rangle$ of the degrees of the neighbours of vertices of degree $k$. It shows the disassortative nature of the network where high-degree nodes tend to connect with low-degree nodes and vice versa. This could represent a network expression of a 'mobilisation' of symbolic resources. Real-world

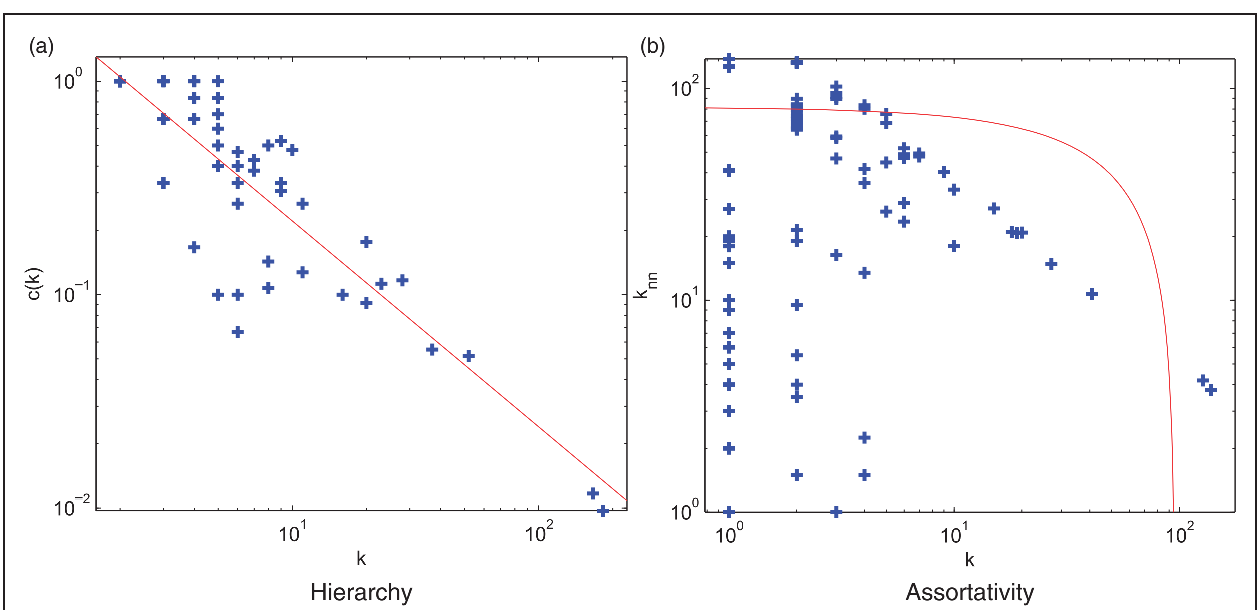


**Figure 13** Plot (a) shows the average clustering coefficient $c(k)$ vs. degree $k$ and plot (b) shows the average $\langle k_{nn} \rangle$ of the degrees of the neighbors of vertices of degree $k$. (a) Hierarchy and (b) assortativity.

social networks are usually assortative because in real life everyone would like to connect with the elites in a society; whereas the elites would rather communicate with those who share their own social status. The only times when elites appeal to the masses are special occasions such as political elections. In a semantic graph, the disassortative nature of the network can be interpreted as the mobilisation of candidates on many different issues, even marginal ones, to attract the largest amount of voters possible. In other words, it represents the semantic equivalent of political rallies in small towns that, nevertheless might have some significance for the overall electoral result.

*Structural balance of networks.* In addition to the contribution to communication theory and political science literature, we aim to contribute to the current debate on structural balance theory in social networks. Structural balance theory originated in social psychology during the mid-twentieth century. As formulated by Heider (1946) in the 1940s, and subsequently cast in graph-theoretic language by Cartwright and Harary (1956), structural balance considers the possible ways in which triangles on three individuals can be signed and posits that triangles with three positive signs (three mutual friends) and those with one positive sign (two friends with a common enemy) are more plausible and hence should be more prevalent in real networks than triangles with two positive signs (two enemies with a common friend) or none (three mutual enemies). Balanced triangles with three positive edges exemplify the principle that 'the friend of my friend is my friend', whereas those with one positive and two negative edges capture the notions that 'the friend of my enemy is my enemy', 'the enemy of my friend is my enemy' and 'the enemy of my enemy is my friend'.

Even though balance theory has been intended for undirected networks, it has also been applied to directed networks by ignoring direction. The work by Leskovec et al. (2010) analysed theories of balance and status in the context of social media sites using the triad configurations of their networks, investigating the extent to which each theory helped explain the linking behavior of users on these sites. Hassan et al. (2012) compare the predictions of network edge signs made by their system to the structural balance theory by counting the frequencies of different types of triads in the predicted network.

We study structural balance in our election networks based on closed triads, which are acyclic or cyclic and mutual dyads. The acyclic triad depicts transitivity. The triad involving actors i, j and k is transitive if whenever $i \rightarrow j$ and $j \rightarrow k$ then $i \rightarrow k$ (Wasserman, 1994). A mutual dyad occurs when $i \rightarrow j$ and $j \rightarrow i$. According to Heider, we know the rules for balance in transitive triads and mutual dyads, but in the case of cyclic triads we introduce rules for balance.

There are six possible closed triads for transitivity in a network and four possible closed triads for the cyclic case. Figure 14 shows the six different configurations of transitive triads and four configurations of cyclic triads with the number of transitive and cyclic triples in each

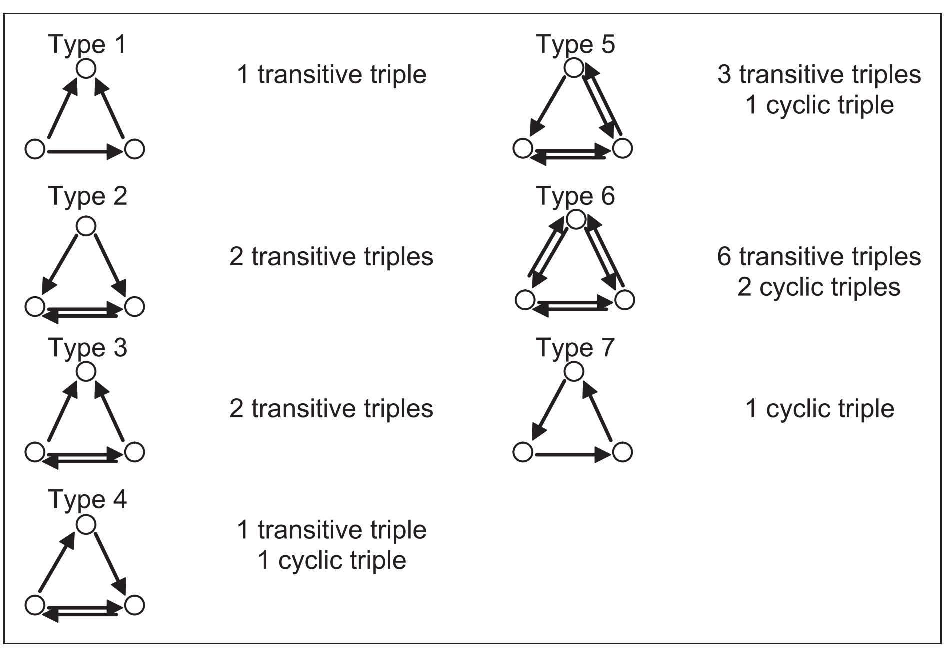


**Figure 14.** Configurations of transitive and cyclic triads in a network.

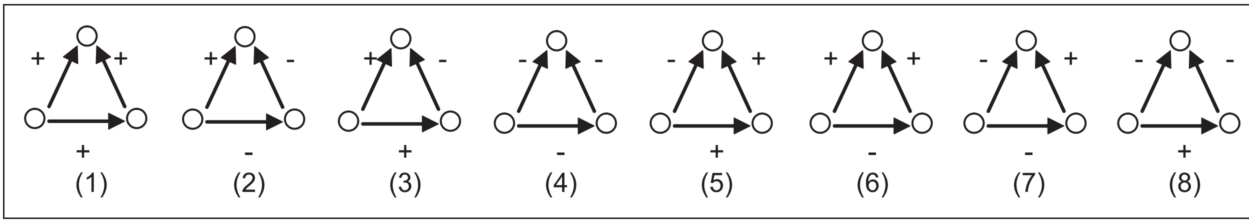


**Figure 15.** Eight types of positive and negatively signed transitive triads.

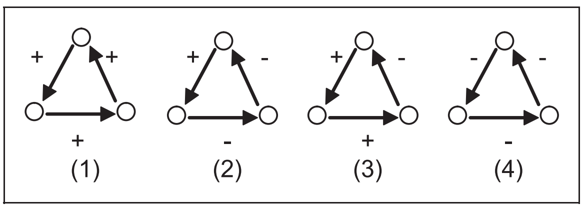


**Figure 16.** Four types of positive and negatively signed cyclic triads.

case. We analyse structural balance in transitive and cyclic triads for all these configurations separately.

When considering positive and negative signs on triads there could be eight types for a single transitive triad as shown in Figure 15. According to the balance theory cases 1, 2, 7 and 8 are balanced. For the cyclic case there could be four types as Figure 16 shows. We consider a cyclic triad balanced when the product of the signs is positive. Since $(-1)^n$ is $= +1$ for $n =$ even and equal to $-1$ for $n =$ odd, any network path containing an even number of opposition statements indicates a support from the root to the leaf of the path. Since we assume that any agent should not be critical of itself, we expect that all closed cycles will contain only an even number of negative edges, in order to be self consistent. Hence cases 1 and 2 are balanced. In a mutual dyad there could be two positives (1), two negatives (2) or one positive and one negative (3). In this case (1) and (2) are balanced.

We present our results on structural balance for each type of triad shown in Figure 14. This includes the analysis of structural balance in transitive, cyclic triads and mutual dyads in networks of different sizes from the US elections.

In all the configurations in Figure 14 we analyse the transitive, cyclic triads and the mutual dyads separately for structural balance. Our experiments are based on 11 election networks of different sizes $N$. We generate these networks by using different confidence thresholds for Wilson score shown in equation (3). For each network we calculate the fraction of balanced triads $T_b$ and then recalculate the same value $T_r$ for 100 random networks. We generate 100 random networks by removing signs from triads and producing them randomly but keeping the fraction of positive links $f_p$ in the network the same. We count the number of times $m$ that $T_r > T_b$, hence our $p$ value $p_T$ is estimated as follows:

$$p_T = \frac{m}{100} \tag{7}$$

**Table 4.** Structural balance in all transitive triads in total.

| N | fp | Tn | Tb | pT |
|---|---|---|---|---|
| 37 | 0.5106 | 7 | 0.1429 | 0.93 |
| 70 | 0.5556 | 16 | 0.3125 | 0.88 |
| 92 | 0.5328 | 26 | 0.2692 | 0.99 |
| 142 | 0.5606 | 52 | 0.2693 | 1.0 |
| 159 | 0.4764 | 72 | 0.2917 | 0.99 |
| 218 | 0.3383 | 131 | 0.3817 | 0.86 |
| 392 | 0.1913 | 251 | 0.3507 | 0.79 |
| 406 | 0.1827 | 264 | 0.3409 | 0.89 |
| 1047 | 0.067 | 838 | 0.278 | 0.98 |

Table 4 shows the scores for $T_b$ and $p_T$ for the total number of transitive triads found in all types of triads in Figure 14. In Table 5 we report the scores for each triad type separately. $T_n$ refers to the number of transitive triads in each case. According to Table 4 we see that $p_T$ is very high and $14\% < T_b < 38\%$, which indicates very poor structural balance overall. In Table 5 we see that type 1 and type 3 triads contribute significantly to the overall imbalance since the majority of imbalanced transitive triads lie in these configurations.

When carefully looking at the signs in triad type 1 and 3 we find that the most dominating transitive triad is the one with three negative links (–,–,–). For triad type 3 in the majority of the cases the signs on the mutual edge were negative. Therefore according to this triad type 2 entities $A$ and $B$ have some sort of disagreement between each other, while they both also disagree about another entity $C$.

It shows that when candidates both disapproved on one issue usually they still disapproved of each other, therefore not creating any common ground. Hence,

**Table 5.** Structural balance of transitive triads in the six different triad type configurations.

| N | Tn | Tb | pT | N | Tn | Tb | pT |
|---|---|---|---|---|---|---|---|
| Type 1 | | | | Type 2 | | | |
| 37 | 1 | 0 | 0 | 37 | 4 | 0.25 | 0.71 |
| 70 | 5 | 0.2 | 0.7 | 70 | 6 | 0.3334 | 0.67 |
| 92 | 8 | 0.125 | 0.93 | 92 | 8 | 0.25 | 0.86 |
| 142 | 14 | 0.1429 | 0.98 | 142 | 12 | 0.1667 | 0.97 |
| 159 | 26 | 0.2692 | 0.98 | 159 | 13 | 0.1538 | 0.99 |
| 218 | 45 | 0.3333 | 0.94 | 218 | 14 | 0.2143 | 0.97 |
| 392 | 86 | 0.3257 | 0.68 | 392 | 27 | 0.2963 | 0.83 |
| 406 | 86 | 0.3257 | 0.62 | 406 | 27 | 0.3333 | 0.66 |
| 1047 | 225 | 0.2489 | 0.36 | 1047 | 130 | 0.2154 | 0.62 |
| Type 3 | | | | Type 4 | | | |
| 37 | 1 | 0 | 0.46 | 37 | 0 | 0 | 0 |
| 70 | 3 | 0 | 0.89 | 70 | 0 | 0 | 0 |
| 92 | 8 | 0.25 | 0.82 | 92 | 0 | 0 | 0 |
| 142 | 18 | 0.2222 | 1 | 142 | 3 | 0.3333 | 0.48 |
| 159 | 24 | 0.2083 | 0.97 | 159 | 3 | 0.3333 | 0.47 |
| 218 | 48 | 0.2917 | 0.81 | 218 | 7 | 0.4286 | 0.38 |
| 392 | 86 | 0.2674 | 0.96 | 392 | 20 | 0.4 | 0.24 |
| 406 | 93 | 0.2581 | 0.97 | 406 | 20 | 0.35 | 0.34 |
| 1047 | 294 | 0.2448 | 0.73 | 1047 | 66 | 0.1971 | 0.53 |
| Type 5 | | | | Type 6 | | | |
| 37 | 1 | 0 | 0 | 37 | 0 | 0 | 0 |
| 70 | 2 | 0.5 | 0.13 | 70 | 0 | 0 | 0 |
| 92 | 2 | 0.5 | 0.14 | 92 | 0 | 0 | 0 |
| 142 | 5 | 0.8 | 0.01 | 142 | 0 | 0 | 0 |
| 159 | 6 | 0.8333 | 0.05 | 159 | 0 | 0 | 0 |
| 218 | 16 | 0.5625 | 0.1 | 218 | 1 | 1 | 0 |
| 392 | 28 | 0.5 | 0.05 | 392 | 4 | 0.25 | 0.18 |
| 406 | 30 | 0.4001 | 0.33 | 406 | 8 | 0.375 | 0.26 |
| 1047 | 94 | 0.3617 | 0.62 | 1047 | 29 | 0.138 | 0.69 |

there might be a substantial latent ‘convergence space’ on issues that is not expressed because of the competitive nature of the campaign.

Table 6 shows the scores for $T_b$ and $p_T$ for the total number of cyclic triads found. Here we see that $33\% < T_b < 79\%$ while values for $p_T$ remain low in most cases other than the largest network. We see that structural balance fails in cases where there is less information in the networks ($N \leq 70$) or there is more noise ($N = 1047$). Also we found that many balanced cyclic triads were found in triad type 5 according to Figure 14 with negative signs on mutual edges and a positive sign on the remaining edge. An increased number of this type had caused the improvement in balance in the cyclic triads.

Table 7 shows $D_b$ (fraction of balanced dyads) and $p_D$ scores for mutual dyads found in the network. $p_D$ refers to the $p$ value of the number of times that $D_r > D_b$ in 100 random networks. $D_r$ is the fraction of balanced dyads in a random network.

It is interesting to see that $81\% < D_b < 100\%$ while $p_D$ is 0 for all cases, indicating that they are significantly balanced. We found that the dyad type most likely to contribute to the balance was the dyad with two negative (–,–) links. Hence the network shows very high balance in the dyad level.

From the experiments we can say that election networks do not show balance at the triad level while they do at the cyclic level and dyad level. Triadic closure and transitivity appears not to be characteristics of our

**Table 6.** Structural balance in all cyclic triads in total.

| N | fp | Tn | Tb | pT |
|---|---|---|---|---|
| 37 | 0.5106 | 1 | 0 | 0 |
| 70 | 0.5556 | 3 | 0.6666 | 0.17 |
| 92 | 0.5328 | 4 | 0.75 | 0.02 |
| 142 | 0.5606 | 9 | 0.7777 | 0.05 |
| 159 | 0.4764 | 10 | 0.7999 | 0.02 |
| 218 | 0.3383 | 24 | 0.75 | 0.02 |
| 392 | 0.1913 | 54 | 0.5556 | 0 |
| 406 | 0.1827 | 59 | 0.5423 | 0.01 |
| 1047 | 0.067 | 192 | 0.3385 | 0.21 |

**Table 7.** Structural balance in all dyads.

| N | Dn | Db | pD |
|---|---|---|---|
| 37 | 3 | 1 | 0 |
| 70 | 5 | 1 | 0 |
| 92 | 6 | 1 | 0 |
| 142 | 11 | 0.9091 | 0 |
| 159 | 12 | 0.9167 | 0 |
| 218 | 21 | 0.9048 | 0 |
| 392 | 36 | 0.8333 | 0 |
| 406 | 44 | 0.8182 | 0 |
| 1047 | 101 | 0.8317 | 0 |

semantic graphs and this can be intuitively interpreted as a less 'coherent' set of signed ties. This intriguing finding suggests that semantic graphs might be characterised by a low level of 'coherence' and the unbalanced nature can be explained in the context of what the semantic graphs represent. In the case of the representation of the US election, in the context of competition among political actors, coherent positioning on issues does not translate into a similar endorsement of other actors that might share the same views. However, at the dyad level, balance is present and especially in the negative form (− −), suggesting tested here is a great deal of common ground regarding what is disapproved of that binds the two parties and candidates.

## A8 – Network partition method

The network should be partitioned into two classes such that nodes in the same class are linked by positive edges and nodes in different classes are linked by negative edges. The division of a network into two parts can be a computationally expensive step, but the optimisation problem can be relaxed, reducing to a simple algebraic task by introducing the approximation that the adjacency matrix of the network is symmetric and positive definite, an assumption that can be readily satisfied. We lose information at this stage by ignoring the direction, but it is done only for the partitioning of the network. For the rest of the analysis we use the actual directed network.

Given a network with its adjacency matrix $A$, we make it symmetric by adding it to its transpose resulting in matrix $M = A + A^T$.

In matrix $M$ where $M_{ij} \in \{-1, +1\}$ we want to assign each node to one of the two classes $\{-1, +1\}$ as mentioned. This leads to the following optimisation problem.

$$\underset{y \in \{-1,+1\}^m}{\operatorname{argmax}} \sum_{ij} M_{ij} y_i y_j$$

We relax this problem (which is NP hard) by allowing the membership function of each node to assume values in $\Re( y \in \Re)$ while keeping the norm of $y$ fixed to avoid trivial solutions.

$$\underset{y \in \Re, ||y||=m}{\operatorname{argmax}} \sum_{ij} M_{ij} y_i y_j$$

The problem now reduces to the following optimisation problem.

$$\underset{y}{\operatorname{argmax}} \frac{y^T M y}{y^T y}$$

This is equivalent to the eigenvalue problem, $My = \lambda y$ by Rayleigh quotient (Horn and Johnson, 2012) since $M$ is symmetric and positive definite by construction, and therefore is efficiently solvable.

The real value assigned to each node in the eigenvector can be interpreted as the degree to which it belongs to one of the two classes. Each eigenvector corresponds to a possible bi-partitioning of the graph, with the quality of the partition being represented by the corresponding eigenvalue. Therefore it is natural to make use of the first eigenvector, possibly looking at the second one when the eigenvalues are very similar.

## A9 – List of verbs

See Table 8.

## A10 – Narrative centrality by candidate

Once we have partitioned the network into two components, by analysing the first eigenvector of the adjacency matrix, we can then analyse separately the two resulting sub-networks, one relative to the Obama camp and the other relative to the Romney camp. In this section we will report the network centrality and their interpretation in Obama's and Romney's ideographs separately. The networks are shown in the main text.

**Table 8.** Verb lists showing positive and negative verbs used for the analysis.

| Positive verbs | | Negative verbs | |
|---|---|---|---|
| Admire | Favor | Accuse | Dislike |
| Adore | Glorify | Bash | Exclude |
| Appreciate | Honor | Blame | Eliminate |
| Applaud | Like | Chide | Insult |
| Approve | Laud | Condemn | Mock |
| Accept | Praise | Criticize | Oppose |
| Attest | Support | Castigate | Repel |
| Acclaim | Vote | Damn | Reject |
| Cheer | | Denounce | Renounce |
| Canvass | | Deprecate | Scold |
| Celebrate | | Deride | Tease |
| Commend | | Decline | Threaten |
| Defend | | Deny | Taunt |
| Exalt | | Disagree | |
| Endorse | | Disapprove | |
| Eulogize | | Disbelief | |

*Obama's network.* Table 9 presents the most prominent object defined by a set of network metrics in the Democratic part of the network. Let us consider four main network metrics: degree, betweenness, closeness centralities and authority.

Degree. Degree centrality indicates how prominent an object is in the network by 'brute force' of number of incoming and outgoing links. As shown in Table 9, in the case of the Democratic network, the highest degree objects are: Obama, President (in this case acting as a synonym for Obama), Clinton, Americans, Economy, Campaign, etc. Several items are of particular interest here: Economy, Plan, Rights, Proposals and GOP. The economy was the most salient issue upon which the campaign focused, with an indication of a clear plan to recover the US economy. Hence, it is not surprising that 'economy' has a high degree centrality in the Democrats' network. The presence of both 'Plan' and 'Proposals' suggests the Obama campaign was focused on presenting defined courses of action rather than principles. This is in line with Conway et al. (2012) who analysed Obama's strategy during the 2012 campaign in terms of the complexity of its rhetoric. Interestingly enough, the object 'GOP' (Grand Old Party) or the Republican Party is among those with highest degree centrality.

Betweenness. The objects with the highest betweenness centralities in the Democratic network were: Obama, President, Campaign, Clinton, Americans,

**Table 9.** Network centrality measures for the Democrats' sub-network.

| Betweeness centrality | In-closeness centrality | Degree | Hub | Authority |
|---|---|---|---|---|
| Obama | Campaign | Obama | Obama | Obama |
| President | Law | President | President | President |
| Campaign | Redistribution | Clinton | Speech | Economy |
| Clinton | Pledge | Americans | Campaign | Plan |
| Americans | Ads | Economy | Biden | Rights |
| Biden | Overhaul | Campaign | Candidate | Proposals |
| Speech | Biden | Biden | Clinton | Efforts |
| ... | Investments | Team | Program | Vice President |
| ... | Camps | Democrats | Gwen Conti | Speech |
| ... | Rights | Plan | Unions | Remarks |
| ... | Commitments | Rights | Biden | Election |
| ... | Mission | Proposals | Hispanics | Policies |
| ... | Politics | Candidate | Group | Campaign |
| ... | Obama | Vice President | Harry Reid | Biden |
| ... | Clinton | GOP | Delegates | Debbie Stabenow |

**Table 10.** Network centrality measures for the Republicans' sub-network.

| Betweeness centrality | In-closeness centrality | Degree | Hub | Authority |
|---|---|---|---|---|
| Romney | Amendment | Romney | Romney | Romney |
| Ryan | Work | Ryan | People | Plans |
| People | Proposal | Plans | Supporters | People |
| Republicans | Breaks | People | Ryan | Mourdock |
| ... | Supporters | Republicans | Men | Decision |
| ... | Ownership | Mourdock | Orlando Sentinel | Proposal |
| ... | Party | Voters | Mock Big Birds | Bill |
| ... | Nomination | Decision | White Man | Israel |
| ... | Decision | Proposal | Future | Amendment |
| ... | Position | Supporters | Voters | Breaks |
| ... | PBS | Bill | Economic condition | PBS |
| ... | Legislation | Israel | Republicans | Ownership |
| ... | Changes | Men | ... | Work |
| ... | Voices | Stance | ... | Legislation |
| | Bill | Amendment | ... | Cuts |

Biden and speech. This is unsurprising because Obama's policies and standpoints on a wide range of different issues were the main link bringing about these nodes. The same can be said about Campaign, Clinton and Biden. The other bridging node is 'speech', because the links between issues were made in Obama's (and Clinton's or Biden's) speeches. Thus the speech is the 'bridge' between the different issues in the sense that they bring the issues together. The other interesting item is 'Americans'. This is the object that Obama's campaign used as the noun for American citizens. It marks the difference with Romney's high betweenness item 'People' (as we will see in the next Section), which has a Republican connotation.

In-closeness. Table 9 reports 'in-closeness' centrality. Considering the objects with the highest in-closeness centrality in Obama's network, we find objects such as: 'Campaign', 'Law', 'Redistribution', 'Pledge', 'Ads', 'Overhaul', 'Biden', 'Investments', etc. These objects are a mixture of policy and economic tools (law, redistribution, overhaul of a policy, investments) and contexts (Campaign, Ads, Pledge). The vice-president Biden can also be considered as a context in the sense of being the context of a given Obama statement. This centrality measure gives an idea of the Democrats' domains of statements that can be compared to the Republicans'.

Authority. As discussed previously, identifying objects with high authority can help us to define the agenda-setting efforts of each of the two parties. Using Table 9 to navigate the objects with the highest authority in the Democratic network, we can find: Obama/President, Economy, Plan, Rights, Proposals, Efforts, Biden/Vice-President, Speech, Remarks, Election, Policies, Campaign and Debbie Stabenow. The latter refers to the Democratic Senate candidate in Michigan and the related Republican 'Asian Ad', an unfortunate television advert for the Republican candidate Peter Hoekstra. Beside the expected presence of Obama and Biden, the other objects are indicative of a campaign strongly focused on the economy and on actions and tools (plan, proposals, efforts, policies) combined with an emphasis on rights. In the above discussion about issue ownership during US presidential campaigns, the latter topic has traditionally been a Democrat theme. It appears that the 2012 campaign was very much focused on the 'Economy' and 'Rights' for the Democrats, and the agenda-setting effort was directed towards these two issues.

*Mitt Romney's network.* Table 10 presents the most prominent objects defined by a set of network metrics in the GOP of the network. Let us consider four main network metrics: degree, betweenness, closeness centralities and authority.

Degree centrality. The top two objects in terms of degree centrality in the Republican network are obviously related to the presidential 'ticket': Mitt Romney and Paul Ryan. The two main actors of the Republican campaign are followed by two nouns: people and plans. We have mentioned earlier the historical and symbolic importance of this term of the GOP. It is, therefore, unsurprising that it is among the objects with a very high degree in the network of claims. Another object with high degree centrality is 'Plans', which refers to the Republican initiative of prescriptions for the US economy.

'Voters' and 'Supporters' also have a high degree centrality as they constitute the main targets of the campaign. The other target of many Republican statements was the object 'People' (we have previously discussed the symbolic value of this term in the Republican political vocabulary).

There are, however, two objects that are related to controversies and attacks from the Democrats. The first is 'Mourdock'. This refers to the controversial public statement on abortion and rape made by the Republican senator Richard Mourdock.[5]

The second issue related to an 'attack' from Obama's campaign is the object 'Bill'. It refers to the Republican vice-presidential nominee Paul Ryan 'blocking' a farm bill that Obama claimed would 'provide relief and certainty to US farmers and ranchers' during the worst drought in more than 50 years. Another interesting object with high degree centrality is 'Israel'. Candidate Romney travelled to Israel during the campaign and expressed his strong support for the Israeli state and his policy in the Middle East. This topic generated a substantial amount of discussion in the media.

Betweenness. In the case of the Republican network, the objects with high betweenness centralities are all actors: 'Romney', 'Ryan', 'People' and 'Republicans'. Also in the case of the Democratic network, actors were the most bridging objects, but it appears that for Republicans there are the two candidates – Romney and Ryan – and then the 'People' or the GOP. By comparison, in Obama's network there is the presence of Bill Clinton in addition to Joe Biden, there are the 'Americans' rather than 'People', and there is no party but rather Obama's 'speeches'.

In-closeness. As discussed in the case of the Democrats' network, these are objects that in virtue of their high in-closeness centrality are related to context as 'qualifiers'. In the Republican network, we find objects such as: 'Amendment', 'Work', 'Proposal' and 'Breaks'. Some of these objects have been explained before, but to summarise: 'Amendment' refers to the amendment to the federal marriage law to allow same-sex marriage; 'Work' refers to the work requirements necessary to receive welfare benefits; 'Breaks' is related to tax breaks and 'proposal' to the Republican anti-recession economic policies; 'Bill' to Obama's attack on Paul Ryan about the Farms Bill. Another relevant object is 'PBS', which refers to Romney's comments on public broadcasting framed by the 'Sesame Street' comment.[6] What is evident is that some of the objects with high in-closeness centrality come from the Republican campaigns but others are forced upon them.

Authority. As mentioned before, authority centrality represents a proxy of agenda-setting efforts. In this respect, the objects with the highest authority centrality in the Republican network have been already encountered in the above description. Not surprisingly, the highest object in terms of authority is 'Romney'. This is followed by 'Plans', 'People', 'Decision' and 'Proposal', all of which are keywords of Romney's campaign and language. However, there are three objects indicating a reaction rather than an initiative in terms of agenda setting: 'Mourdock', 'Bill' and 'Public Broadcasting'. As discussed above, all three objects are related to attacks from Obama's campaign. 'Israel' is also present among the objects with high authority and it refers to Romney's speech in Israel, about US Middle East foreign policy.

The remaining objects with high authority are what now appear to be important issues in the 2012 Republican campaign: tax breaks (Breaks), the amendment to the federal marriage law (Amendment), the work requirements for welfare benefits (Work).