

Computational identification of media frames: strengths, weaknesses, and opportunities

Tom Nicholls*¹ and Pepper D. Culpepper¹

Abstract

With the availability of large volumes of electronic communications data and the increasing sophistication of computational techniques, the development of automated approaches for different kinds of framing analysis is an important goal of researchers. There is as yet no standard method for the ‘unsupervised’ inductive identification of frames based upon the content of articles. Three groups of core approaches underlie a wide range of work in this area, and we compare three techniques based on these approaches against each other and against manual human analysts. The three techniques are a k-means clustering algorithm together with a sophisticated natural language processing (NLP)-based feature selection process; evolutionary factor analysis (EFA), a factor analysis approach; and the structural topic model (STM). We use two datasets – one very focused and one extremely broad – as examples of the kind of frame analysis problems readers may wish to attempt. Even in a highly targeted dataset, we find some distance between the frames generated by computational analysis and those manually produced by our human analysts. The details of each method have a substantial impact on frame quality and interpretation. We find that the STM approach is the most effective for our narrow-scope dataset, but that it returns definite topics, and not frames, when working on our very broad dataset. We also show that we can get results emphasizing different parts of the framing problem by combining parts of different methods as a multi-stage process, rather than viewing available methods as simple plug-and-play models.

Keywords: framing, computational communication science, political communication, methods

*: Corresponding author. tom.nicholls@bsg.ox.ac.uk.

¹: Blavatnik School of Government, University of Oxford, United Kingdom.

This article has been accepted for publication in Political Communication, published by Taylor & Francis.

<https://doi.org/10.1080/10584609.2020.1812777>

Introduction

Media frames are a widely used concept in political communication research; in addition to understanding what is being covered in the media, scholars are interested in how issues are covered, what angles are being taken in the coverage, and the dynamics of how these things change (e.g. Entman, 2003). Consequently, various kinds of framing analysis are a mainstay of communications scholarship, the traditional approach being a multiply-coded manual analysis using an a priori codebook based either on generic frames (e.g. conflict, attribution of responsibility, human interest, morality, and economic consequences in An & Gower, 2009) or specific theoretical expectations about the issue context studied.

With the availability of large volumes of electronic communications data, the development of automated approaches is an important goal of researchers (Matthes & Kohring, 2008).

Although there is as yet no standard method, much work has been put towards the ‘unsupervised’ inductive identification of frames based upon the content of the article texts¹ (e.g. Baden & Tenenboim-Weinblatt, 2017; Burscher et al., 2016; Vlieger & Leydesdorff, 2012). Nevertheless, there is a debate about the extent to which the constructs generated by these computational tools are actually frames, in the sense used by manual frame analysts, and therefore whether the new techniques are fully fit-for-purpose. This is an important subject for scholarly debate – but it is only a relatively narrow slice of the potential issues involved in trying to use unsupervised methods to assess media framing. Our goal in this article is to broaden this conversation and render salient and visible a couple of its problematic assumptions.

We have several findings that may be of interest to scholars thinking about applying computational analysis to substantive problems in various fields. Firstly, it matters how

¹ This can be contrasted with the ‘supervised’, deductive, labeling of document corpora given a set of frames decided in advance. This can be done with supervised classification tools (see, for example, Burscher et al. 2014) and is quite a different problem.

targeted the dataset is: as computational frame extraction techniques rely on differences in language use, they return quite different ‘frames’, depending on the breadth of the subjects covered in the initial dataset; we analyze two different datasets, one focused and one very broad.

Secondly, we show that, even in a highly targeted dataset, there is still some distance between the frames generated by computational analysis and those manually produced by our human analysts. The details of each method have a substantial impact on frame quality and interpretation.

Thirdly, we find that a topic modeling approach using the structural topic model (STM) is by some margin the most effective for our narrow-scope dataset, but that it returns definite topics, and not frames, when working on our very broad dataset. We also show that we can get results emphasizing different parts of the framing problem by combining parts of different methods as a multi-stage process, rather than viewing available methods as simple plug-and-play models, and by using deliberate manual intervention in different parts of the process to get more clearly frame-like outcomes.

Our enterprise is to make people think about where these methods can be applied most plausibly and to have an open discussion about the benefits of various computational methods and some of the trade-offs between computational and manual methods, including how they can usefully be combined.

In the next section we analyze the basic logic of framing and the computational methods proposed to identify and code frames. In the third section, we introduce our datasets and present three methods for applying computational approaches to them as well as a conventional manual approach for comparison. We also present a further method which draws on the combined strengths of different parts of these methods as well as the researchers’ own knowledge. In the fourth section we analyze the results and explore what

they reveal about the nature of framing and computational approaches to it. Finally, we recommend productive avenues for future research and draw conclusions.

Framing Theory and Computational Analysis

Computational frame analysis faces a host of inter-related challenges. The most basic, and contested, of these is, ‘what is a frame?’ On this fundamental point scholars differ, and our goal is not to resolve the difference. It is rather to show both the promise and limits of how computational methods can contribute to the analysis of frames, variously defined, in media. We use unsupervised methods to see if we can ‘let the data speak’. But we also acknowledge explicitly the challenges of, and the assumptions required in, this act of computational ventriloquism.

The first challenge facing computational frame analysts is simple: can it even be done? A prominent strand of framing literature (e.g. Reese, 2010) is critical of overly-mechanistic approaches to frame analysis and directly skeptical of the possibilities of computational analysis, seeing frame packages as too complex to be reduced to a single indicator or topic.

We sympathize with this view. To the extent that texts use ‘packages’ of words and symbols to connect to inter-subjectively shared meanings that are not part of the texts themselves, the current state of computational approaches does not allow us to infer those ideas themselves from analysis. This is the area of theoretical interpretation, not computer coding². Insofar as frames that constitute coherent packages of meaning can arise from computational frame analyses, it must be because these packages are statistically associated with something that can be measured: typically, the lexical choices made by authors.

² Computational identification of higher-level constructs is certainly feasible in principle – there are techniques in the wider Natural Language Processing literature such as automated discourse analysis and argument mining – but these are not built for the identification of frames specifically, do not allow a direct understanding of actors’ intentions beyond those manifest in the text itself, and are still relatively underdeveloped.

Having said that, if a coherent *textual* package of combined symbols (with more or less emphasis) is actually coherent, then they are potentially modellable as such, distinct from other packages. Some other possible aspects of the framing ‘package’ may also be computationally tractable: moral rhetoric, for example, has dictionary-based detection methods available (Sagi & Dehghani, 2014); affective content can be detected using sentiment analysis (Young & Soroka, 2012), and it may in principle be possible to detect recommendatory language using Natural Language Processing (NLP) tools. Most existing work on computational frame detection works at the level of words rather than concepts, though.

A second challenge inheres in the reliability and reproducibility of framing, though this is a challenge that confronts both manual and computational methods. This is partly because scholars disagree about the extent to which frames can be empirically distinguished between the related concepts of ‘priming’ and ‘agenda-setting’ (Scheufele & Tewksbury, 2007) and the competing merits of equivalence frames (e.g. Kahneman & Tversky, 1984; Scheufele & Iyengar, 2017) and emphasis frames (e.g. Druckman & Nelson, 2003; Leeper & Slothuus, 2018). Much of the media framing literature in political science focuses on emphasis frames, and that is the approach we follow (Busby et al., 2018; Chong & Druckman, 2007). But the more important point for purposes of this article is to recognize that neither computational nor manual methods can draw on a consistent conceptualization of frames, because divergent conceptualizations remain a point of scholarly dispute and frames are, in practice, difficult to conceptualize consistently even if a common definition is chosen (Vliegenthart, 2012). Consequently, “many scholars have pointed to threats to reliability and validity in frame analysis.” (Matthes & Kohring, 2008, p. 274)

Consistent with much of the emphasis framing literature, we follow the widely cited definition by Robert Entman (1993). He defines framing as “[selecting] some aspects of perceived reality and make them more salient in a communicating text, in such a way as to

promote a particular definition of a problem, causal interpretation, moral evaluation, and/or treatment recommendation for the item described” (p. 52). In our analysis we focus in particular on the features of problem definition and the diagnosis of causes as two minimal features of the frames in communication that we find in media.

These differences in conceptualization find their analogues in the consistency of results found in framing studies. A particularly striking finding comes as a by-product of an analysis of supervised machine learning methods by Burscher et al. (2014), who tested a new supervised approach against human coders, using ‘standard’ generic news frames. Their system performed well against the humans, but critically the human inter-coder reliability was poor, with Krippendorff’s alpha scores between .21 and .58 for the four frames they coded.

This has two implications for the computational enterprise. One is that frames are difficult to define even for humans and therefore may not be entirely coherent as a construct, even given very concrete guidance for identifying frames. The other is that as a result we do not have sets of truly gold standard human-coded data against which to develop and test computational approaches,. Despite this, and the consequently increased importance of validating and testing computational methods against each other, there are limited numbers of comparative studies analyzing the effects of different methodological choices on frame extraction (e.g. Burscher et al., 2014; David et al., 2011).

Supervised computational methods can be used to automate the coding of predetermined frames (Matthes & Kohring, 2008). They are able to draw on theoretical insights to structure expectations about frames, which can generate powerful insights. Yet, by importing these assumptions, supervised approaches are unable to offer insight into what frames might exist in a dataset, which theory has not already anticipated. This is most important when dealing with very large collections of texts from diverse news sources, where it is equally difficult for the researchers to manually analyze a large enough part of the collection to provide candidate frames which are representative of the whole dataset.

Here, there is a danger of falling into the ‘drunk under the streetlight’ syndrome, in which we focus on the points we can ‘see’ under the streetlight of supervised methods, and failing to explore the darker terrain of unconsidered and unexpected potential frames. By coding our intuitions (either directly in manual methods or via a training dataset in supervised computational ones), we embed our assumptions about the frames we expect.

Trying to develop effective unsupervised methods is a promising way to break out of this problem, analogously to the two-stage inductive/deductive process advocated in qualitative news frame analysis by Van Gorp (2010). In this conception, an exploratory and inductive process is the first part of the analysis, focusing on identifying potential frames and their associated packages of framing devices, before moving on to a strictly deductive second stage where the elements of those frame packages are used to code frame prevalence in a controlled and reproducible way. Applied to the computational project, this suggests using unsupervised analyses to identify patterns of commonality in the documents that might represent frames, before moving on to an exhaustive coding – whether using these identified frames directly, or by using them as a conceptual input prior to the researcher’s downstream process of coding frames using a hand-coded training set and supervised methods. Baden notes that unsupervised studies can best identify “constellations of framing devices *possibly* indicative of distinct frames. The only way to reconnect such indicator sets to those frames they express, however, is to - *very subjectively* - *interpret the possible meanings of the detected framing devices*” (Baden, 2018, pp. 13–14, our emphasis). This is in keeping with a healthy tradition of hybrid methods encompassing both computational and manual elements and drawing strength from both (e.g., Muddiman et al., 2019; Shahin, 2016).

In this paper we focus on inductive, unsupervised, methods, aiming to identify frames rather than simply code them. The risk of using unsupervised methods is the inverse of that of supervised methods – with little theoretical superstructure on which to build, the patterns found may have little connection to the most conceptually interesting issues in framing.

Unsupervised methods are an engine of discovery, but one in which we might expect a higher ratio of noise to signal than in supervised methods whose strength (and weakness) is to look for patterns that we have reason to believe are of theoretical interest³.

There is no standard method for unsupervised, inductive, computational frame identification. Indeed, there are a multiplicity of different methods published in a variety of disciplines; computational frame analysts work from disciplines including communications science and political science, and draw on insights from a huge range of sources – home-grown computational communications research directly, the NLP, information retrieval, and machine learning fields within computer science, and classical and Bayesian statistics.

There are, nevertheless, three groups of core approaches which underlie a wide range of work in this area. These are the sources of the representative methods we will test in this paper.

The first are vector space clustering methods. These models normally represent each document as a vector of word counts, with each vector as long as the number of words in the corpus vocabulary list. By treating vectors for each document as points in several-hundred-dimensional vector space, documents can be clustered using standard vector space tools (Manning et al., 2008, pp. 266–267). The second are factor analysis techniques, designed to extract latent variables from high dimensional data. By modeling documents as sets of semantic or lexical variables and then applying factor analysis techniques such as PCA, Latent Semantic Analysis, or Evolutionary Factor Analysis, latent semantic variables can be extracted which represent aspects of the documents under study. Thirdly, topic modeling methods such as Latent Dirichlet Allocation (LDA, Blei et al., 2003) and the structural topic model (STM, Roberts et al, 2014) have been used. While each of the three approaches substantially varies, each is based upon the underlying insight that variations in the choice of

³ Our discussion of supervised versus unsupervised methods is stylized, with most real-world applications falling between the two; no theoretical schema is designed without knowledge of the real world, and no inductive approach can avoid all theoretical assumptions. Even the most rigid supervised method relies on the data to support or not the proposed coding, and even the most data-driven unsupervised method necessarily embeds a particular conception of framing within it. For manual methods, these are the implications of the instructions for open coding. For computational approaches, it is also within the design of the method.

words within articles, and particularly patterns of word co-occurrence, are revealing about its semantic content.

We find it helpful to conceptualize inductive frame extraction techniques as a modular process, where data pre-processing and the subsequent identification of frames are independent and replaceable steps. The output from the first stage can be conceptualized as a feature matrix, fed into a second stage to extract the frames. This is analogous to the way that computer scientists and especially data scientists approach similar text analysis problems, where the selection and processing of features is treated as a discrete problem to the selection or development of a classifier to reach the desired end state.

In the next section, we will outline three complete published methods using each of these core approaches for inductive computational frame identification, and how we have applied them to our data.

Methods

To examine more concretely the strengths and weaknesses of computational approaches to inductive frame analysis, we implement three methods from the wide range available in the literature and examine them comparatively. Each of the three methods selected is representative of one of the key approaches in computational frame extraction identified above. We also include a manual frame analysis, carried out by two human coders, following the process set out in Chong & Druckman (2007). Our three computational methods are these:

- A k-means clustering algorithm together with a sophisticated NLP-based feature selection process to extract latent frames in collections of articles (Burscher et al., 2016)
- Evolutionary Factor Analysis, which aims to model the over-time changes in the content and meaning of discourse (Motta & Baden, 2013)

- The Structural Topic Model (STM; Roberts et al., 2014), identifying latent topics within a corpus of articles

The first two have been published specifically as computational frame detection methods. STM is methodologically adjacent to frame detection but developed to extract different constructs (topics, defined as probability distributions over words, not frames). Nevertheless, its approach – identifying groups based on term co-occurrence – is in abstract similar to the other techniques. Topic models have previously been used to measure frames (DiMaggio et al., 2013) and Gilardi et al. (2020, p. 3) “consider the topics uncovered by these models as an operationalization of policy frames”. Each of these methods, and our implementation of them, is outlined in more detail below.

We come to this project with a defined substantive interest connected to our large research project on financial regulation: understanding the ways banking has been discussed in the media in different countries since the 2008 financial crisis. The reaction against banks and bankers, as representatives of the elite, has been a fundamental component of post-crisis populism in many countries (Norris & Inglehart 2019; Rodrik, 2018). To understand how this perception varies both across countries, and within countries over time, we want to examine broad trends of media coverage of banks across several countries.

We do not expect that all political communication scholars share our substantive interest in banking regulation. But many *do* share our interest in trying to understand trends in media coverage of a particular subject of political import on which pre-existing analyses do not offer substantial empirical clarity and for which this analysis can provide a model.

Our main dataset for testing the methods is a highly-focused collection of 2,064 Australian articles discussing banks and scandals in the context of the Australian Royal Commission into Misconduct in the Banking, Superannuation and Financial Services Industry⁴. This was

⁴ We searched Australian newspapers using Nexis, using the selection criteria (`hayne OR "royal commission"`) AND (`bank! or "financial services"`) and subsequently filtered to include only articles including ‘scandal’. The date range is 2017/06/01-2018/11/20. Newspapers selected are The Advertiser/

an extremely prominent inquiry into banking scandals in Australia, led by a well-respected judge (Kenneth Hayne). The Commission was initially politically controversial, featured senior bankers being publicly challenged by the Commission's counsel, and highlighted significant wrongdoing by banks in multiple areas. It was front-page news in Australia throughout 2018.

We also have a supplementary dataset, deliberately much wider in scope than the Australian Royal Commission dataset and reflecting a more abstract subject for framing analysis. This consists of news articles from USA Today about banks, bankers, and banking published between 2007 and 2010⁵. We analyze two separate versions of this USA Today dataset. The first is the full collection of 4,653 news articles, covering all aspects of banking (and, indeed, containing some articles about non-bank related entities such as Red Bank (New Jersey), Tyra Banks, and the overflowing banks of various rivers). The second is a smaller 2,394 article subset produced by aggressively filtering out articles from the full dataset that match keywords associated with non-bank entities and subjects⁶. Thus, we have one narrow dataset (Australia Royal Commission) and one broad dataset (banking in USA Today), with the latter having both a filtered and unfiltered version.

The value of using two contrasting kinds of data in this context is that they reflect two ends of a range of possibilities for the frame analyst. On the one hand, the Australian corpus is a fairly conventional tightly demarcated group of articles reflecting a single (extended) story. On the other, the USA Today corpus is much broader. It is the kind of dataset that would

Sunday Mail, The Age (Melbourne), The Australian, Australian Financial Review, Canberra Times, The Courier Mail and The Sunday Mail, Daily Telegraph & Sunday Telegraph (Sydney), Herald Sun & Sunday Herald Sun (Melbourne), Hobart Mercury/Sunday Tasmanian, Sydney Morning Herald, and The West Australian (Perth).

⁵ We searched USA Today data using Nexis, with the following selection criteria: ATLEAST2(bank OR banker OR banking). The date range is 2007/01/01-2010/12/31.

⁶ The filter keywords (case insensitive) were identified inductively from the data: west bank, red bank, outer banks, food bank, tyra banks, lloyd banks, world bank, phone bank, river bank, jobs bank, blood bank, bank robber, bankrupt, golf, sports, sport's, film, holiday, greek, iraqi, unemployment, S&P, Nasdaq, and oil.

allow an analysis of how framing of an entity or subject has changed generally, over longer periods of time.

Both datasets are clean in the sense demanded by Maier et al. (2018) in that they have reliable article texts and metadata, and are generally free of the kinds of boilerplate and navigation items that might be generated by a raw web crawl. But the Australian dataset, and the filtered USA Today one, are much cleaner than the unfiltered USA Today dataset in that they focus more tightly on the precise subject of interest, and have fewer extraneous articles which are unrelated, or only tangentially related, to the real subject of interest.

It is reasonable to expect that the unfiltered USA Today dataset will perform less well than the Australia one, but in practice very few datasets are completely clean in the second sense: there are large numbers of homographs in English, and so keyword-selected corpora are inherently messy on collection. Effective filtering by reading each article is infeasible for large-N datasets, while statistically assisted techniques for filtering inapt articles, or keyword techniques such as that deployed on our filtered USA Today dataset, are necessarily not perfectly reliable.

In each case, the methods' underlying model of text is as a bag of words: models input a matrix where each row represents a document and each column represents a feature (normally a single word). This is known as a Document Term Matrix (DTM). The matrix data are the counts ("term frequency") or a weighted function of them ("tf-idf"⁷). Word order is not considered.

Differences between the frames inductively identified are difficult to evaluate computationally as, unlike most computer science work on unsupervised classification algorithms, there is no gold standard against which to compare. Although we have statistical

⁷ tf-idf scoring modifies the term frequencies, down-weighting each by a function of the number of documents each term appears in. The intention is to make words which appear in large numbers of documents in the corpus have much less weight than those which are more discriminating (see Manning et al., pp.108–109).

metrics for evaluation of models, they do not, alone, measure how well the methods map to the original concept:

The likelihood function and common evaluation metrics reward models which are predictive of unseen words, but our interest is rarely in predicting the words in a document; we want a model which provides a semantically coherent, substantively interesting summary of the documents (Roberts et al., 2016, p. 10)

We therefore evaluate the different frame identification methods subjectively, according to the quality of the frames they have selected. There are no quantitative measures of each of these criteria; they are abstract theoretical qualities about which we make judgments. We use three criteria for quality, with the latter two drawing on Gerring (2001) and Roberts et al. (2014). First, the extent to which the frames identified by the methods fit with our understanding of what a media frame *is* – an interpretive lens that includes the definition of a problem and a diagnosis of its causes (Entman, 1993, p. 52). Second, the extent to which each identified frame is internally coherent. Third, the extent to which each identified frame is externally differentiable from other potential frames.

For each of these methods, the number of frames to identify, K , is a free parameter. This is a conceptual, as well as a practical, problem for frame analysts. Conceptually, in aiming to computationally induct the frames that exist in the dataset, we wish to minimize our imposition on the data of our prior view on that data as to how many frames to expect. Nevertheless, we *do* have expectations, which come from the nature of our conception of emphasis framing, which suggests individual frames with distinct problem definition and models of causal interpretation. There are small number of finite frames in communication that possess distinct causal interpretation of specific issue areas – about two to four is typical of the work in the field (e.g., Valenzuela et al., 2017; Burscher et al., 2014). There is no established law governing the number of possible frames that can emerge in an issue area, but past manual and computational analysis suggests that we should be looking in the single

digits. There are not, logically, 50 sharply distinct ways of defining problem definition and causal interpretations of most issue areas. This observation does not sit easily with the standard computational approaches for optimizing cluster numbers. For topic models, for example, the number of distinct topics in a large corpus is potentially unbounded, and a reasonable model for a large corpus may well have over 100.

There are a wide range of metrics for selecting K in the topic modeling literature. The STM authors recommend the combined use of held-out likelihood and a residuals dispersion test after Taddy (2012), and provide a ‘searchK’ function in the stm package to assist this.

However, they also report that selecting STM parameters purely using metrics can lead to a mild decrease in interpretability and will often not lead to the same model as human selection (Roberts et al, 2016). This is a fortiori true for selecting a number of frames, where human analysts would likely identify far fewer frames than a metrics-fitted topic model will return.

Nevertheless, we fitted models for a wide range of values of K between 2 and 200 on our datasets, in order to see where the ‘natural’ value of K identified by the metrics would fall⁸. For the STM models, this fell around 75 for the Australia data, and 150 for the wider USA Today data. As expected, these models were impossible to interpret as frames, being far too narrow to offer distinct problem definitions or causal interpretations. They were simply topics, as you would expect from a topic model.

Scree plots of metrics against K were fairly uninformative, showing smooth curved slopes with no obvious elbows or inflection points which would represent prominent values for K smaller than the metric optimum. We therefore chose a starting point for K of around 5-10 largely on theoretical grounds: manual framing analyses tend to try to extract a handful of key frames and we should try to replicate this as closely as possible. We might reasonably expect the Australian articles (being much narrower in scope) to have fewer frames than the USA Today articles, so for STM and k-means, we then fit models for each dataset for values

⁸ For k-means we used the proportion of sum-of-squares that was within-cluster and for STM held-out likelihood and residuals dispersion.

of K from 5 to 10 (Australia) and 8 to 13 (USA Today) . For the EFA analysis, such pre-selection is not necessary: the EFA frames are the product of a PCA-like factor analysis where factors 1 ... K are constant even as factor K+1 is added. Consequently, we fit a single K=13 model for each dataset.

For each method and dataset, we then manually examined each set of clusters, selecting a final value of K for which the frames produced appeared in our judgment to be the most internally coherent and externally differentiable. The manual analysis of a range of models with similar values of K in order to select a final K is, in practice, a standard approach (e.g. Barnes & Hicks, 2018; Gilardi et al., 2020).

Implementation code for each of these analyses is available online (Nicholls & Culpepper, 2020).

k-means

Burscher et al. (2016), proposed a method for frame identification using a sophisticated NLP feature selection stage coupled to a simple k-means clustering model (Manning et al., 2008, p. 338)⁹. The approach is specific and parsimonious in selecting features for inclusion as variables to be clustered on, using natural language processing (NLP) tools. A named entity recognition system is used to remove peoples' and organizations' names, and an automated parts-of-speech tagger to allow the words used to be restricted to only nouns, adjectives, and adverbs. This latter has theoretical support, given that adjectives can be seen as giving *attributes* to objects (e.g. people, issues) within news stories, and those attributes can reflect framing differences when modeled as a second level of agenda setting (Maher, 2010).

Numbers, times, stopwords and other assumed noise are also rejected. The resulting bag of

⁹ k-means is one of the baseline clustering models across many academic fields. The model uses high-dimensional vector space, where each (word) feature is a dimension. Each article is modeled as a point in that vector space, with the adjusted number of times each word appears in the document taken as the distance from the origin in that dimension. K cluster centers are then calculated, minimizing the sum of the square of the distances between each document and its nearest center (which, in our case, becomes its assigned frame).

word counts is then filtered by overall frequency, retaining terms appearing in between 0.5% and 40% of articles, then tf-idf weighted to downweight common terms, and then made into a DTM for clustering. We implemented the NLP parts of this model using CoreNLP (Manning et al., 2014) and clustered using R's standard 'kmeans' function (R Core Team, 2019).

There are two properties of the k-means clustering algorithm which are particularly important for its use in frame identification. Firstly, it produces single-membership clustering: each article is assigned exactly one frame, rather than being treated as a mixture of multiple frames in different proportions. This is distinct from both the other methods used, which are multiple-membership. Secondly, k-means is sensitive to outliers (as its loss function is based on sum-of-squared distances) and so it tends to split off outlying groups where possible. In our case, we might expect a number of small frames reflecting relatively unusual features of some input articles, rather than frames of relatively even size.

EFA

Evolutionary Factor Analysis (Motta & Baden, 2013) is a PCA-type factor analysis technique which was specifically developed for frame analysis. Its distinctive contribution over other factor analysis approaches is a novel technique for identifying time-varying latent semantic structures. Similarly to PCA it uses a formal transformation of the input data matrix to create a set of variables which are a linear combination of the input features, but in the case of EFA the factor loadings are generated such that they are smoothly time-varying and the particular terms which make up the repertoire of a frame can evolve over the time period studied.

The original EFA process involves a manual ontology mapping process to reduce the dimensionality of input data by filtering out “non-meaningful” lexical expressions and combining conceptually-equivalent ones into a single token (Motta & Baden, 2013, p. 61). To maintain some commonality with the other methods, and to reduce the influence of the authors' own intuition in this inductive process, we present here the application of the EFA

dimension-reduction process on top of much rawer input data: an ordinary term-frequency weighted bag of words with conventional pre-processing¹⁰. This allows us to test the comparative effect of the second stage, but with the cost of being unfriendly to the original authors' intent: the results here are based upon only half of the original method. The EFA analysis itself is implemented using the original authors' own code (Motta & Baden, 2015), in GNU Octave (Eaton et al., 2020).

STM

The structural topic model (Roberts et al., 2014) is a topic model popular in political science, which has also been used extensively for computational frame analysis, despite doubts as to whether topics can be fully equated with frames, and under what circumstances¹¹. It is similar in approach to Latent Dirichlet Allocation (Blei et al., 2003), the standard topic model in most academic disciplines, but with the additional ability to condition the topics on covariates. This is particularly helpful for frame analysis because, analogously to EFA, it allows us to model frame language as having variable prevalence over time, as well as allowing us to try to control for newspaper-specific language variation. We fit the models using the 'stm' package in R (Roberts et al., 2018).

¹⁰ Documents were converted to lowercase; stopwords, punctuation, and numbers removed; stemmed (words reduced to their base forms – democrat, democracy, and democratize to democra, for example); and filtered to a minimum sparsity of 10%. General text handling was done using the tm, tm.plugin.lexisnexis, and quanteda packages for R (Feinerer et al., 2008; Bouchet-Valat, 2018; Benoit et al, 2018).

¹¹ The basic aim of STM, and other topic models, is to estimate 'topics' as probability distributions over groups of words such that each topic is a coherent and semantically interpretable group. The key challenge for using topic models for frame identification is that they are built to extract topics, not frames – which define an issue as a 'problem' with an associated causal diagnosis. One classic test is to analyze the '20Newsgroups' corpus of users' posts collected from 20 discussion groups on different subjects (e.g. Atheism, Apple Macintosh computer hardware, Cryptography, and Christianity). Given a setting of K=20, the topic model should separate out fairly reliably the group of origin. Given an instruction to find K=100 topics, though, subtler variations will be extracted – possibly on subtopics within the group. When analyzing a set of articles all about the same subject or event, we might expect the differences between different framings of that subject or event to be a major source of linguistic variation; that is part of the intuition for its use in frame identification.

We produced two STM models, each using the month of publication and the news outlet as prevalence covariates and the outlet as a content covariate¹². The first uses conventional document pre-processing steps similar to the EFA models¹³. The aim here is to reflect a baseline STM approach, with the topics intended to be directly interpretable as frames.

Because we found the STM the most promising of the methods tried, we subsequently built a final model aiming to combine the strengths of the Burscher et al. (k-means) feature selection stage, the STM clustering stage with a large number of topics, and a final manual analysis stage of selecting and merging the resulting topics to identify frames in a semi-supervised way which draws on both the latent structure of the data and our domain-specific knowledge and understandings of framing.

For this second, modified STM model, we select the headline and lead paragraph of each article, then use automated parts-of-speech tagging and named entity recognition to keep nouns, adjectives, and adverbs, while discarding peoples' and organizations' names, numbers, times, and stopwords. We then extend the feature selection stage by modeling each document as a combined bag of single words and bi-grams¹⁴, to better model specific phrases used in multiple texts (e.g. Morstatter et al., 2018). We took the resulting term vectors and combined them into a document-term matrix for the whole corpus, filtering out terms that appeared in fewer than 0.5% or greater than 40% of documents. The terms list was then manually processed to merge semantically equivalent words and filter terms with no particular semantic meaning for this analysis, treating them as corpus-specific stopwords (this

¹² As there is only one outlet in the USA Today dataset, only the publication month is used for that model.

¹³ We reduce the minimum sparsity to 1%, and eliminate stemming. This reflects that stemming does not appear to improve topic quality and may adversely affect topic stability (Schofield and Minno, 2016). Except for the decision to eliminate stemming, steps are consistent with the standard pre-processing procedures for topic modeling in communication research (Maier et al., 2018). We also carried out a grid search using the Australian data to optimize the main potentially relevant hyper-parameter, the sigma prior. There was a minimal, and inconsistent, effect on held-out likelihood so we leave this at its default value of 0.

¹⁴ A bigram is any pair of words occurring together. The previous sentence therefore includes "A bigram", "bigram is", "is any" etc. Because only features containing in > 0.5% of documents are used, though, only very prominent bigrams end up being coded. Its main advantage is in bringing out two-word terms such as "New York" and "Chief Executive", which are more meaningful than their component parts separately.

was inspired by EFA's manual ontology step). The final feature selection stage was to tf-idf weight the matrix to downweight common terms.

Once the feature matrix was complete, we then fitted STMs for a wide range of values of K , using the standard metrics in STM's 'searchK' function, selecting a final value of K to maximize model fit before fitting a final STM. As a final step, we characterized each topic according to its keywords and the documents which load on it most heavily, then had two coders independently identify irrelevant topic clusters and attempt to combine those which remained into coherent frame packages (inspired by Walter & Ophir, 2019, and their combination of large- K STM groupings for frame identification).

Human-guided frame identification

We also carried out a manual analysis on the Australian dataset, for comparison with the computational methods. The basic structure of the analysis was based upon the procedure recommended by Chong & Druckman (2007, pp. 106–108). In the first stage, the authors specified the event for which frames would be sought; in this case, the Australian banking Royal Commission. In the second stage, the frame-defining attitudes were chosen to be attitudes to the banking scandals. For the third stage, two researchers (the first author and a research assistant well-versed in the Australian banking scandals) worked separately after being briefed on the event and attitudes, and also the Chong & Druckman definition of emphasis frames. Each was issued an identical 300-article randomly ordered random sample of the Australian data, and instructed to inductively identify a framing schema.

We do not claim that the results of this manual analysis are “the” gold standard set of frames against which to measure the computational approaches. Given the challenges of operationalizing framing and the low reliability of human coding as a result of researcher effects (Matthes and Kohring, 2008), we offer them as one plausible set of frames, not as the

only one. However, they do stand up as reasonably distinct statements of problem definition and causal interpretation of Australian banking scandal of 2018.

Results

Because the output of these models is lengthy, we have summarized the content of the frame clusters for each method in the appendix, labeling them with their keywords. As each of the clustering methods is based on a different approach we used different, conventional, keywording methods. For k-means, we found the 20 most prototypical document features with the highest means. For EFA, we selected the features which loaded most heavily onto each frame. For STM modeling, we adopted FREX scoring as implemented in the ‘stm’ package (Roberts et al., 2018), which is the harmonic mean of the rank of frequency and exclusivity¹⁵. The headlines of a set of exemplar articles are also shown. For the single-membership k-means method, these are the articles closest to the cluster center – in some sense those most prototypical of the group. For the mixed-membership models, it is the articles which load most heavily onto each frame¹⁶.

Manual analysis

In the manual analysis of the Australia dataset, each analyst identified six frames. These were broadly comparable, though the two differed in detail. These frames, and the compatibility between them, are shown in Table 1.

¹⁵ “ $FREX = \left((w/F) + ((1-w)/E) \right)^{-1}$ where F is the frequency score given by the empirical CDF of the word in its topic distribution. Exclusivity $[E]$ is calculated by column-normalizing the beta matrix (thus representing the conditional probability of seeing the topic given the word). Then the empirical CDF of the word is computed within the topic. Thus, words with high values are those where most of the mass for that word is assigned to the given topic.” (Roberts, Stewart, and Tingley 2018, `calcfx()` documentation). A James-Stein type shrinkage estimator is then applied to depress the scores of the rarest words.

¹⁶ Neither of these measures is perfect and the exemplars should be treated with some caution. In particular, where an article contains only a few features, perhaps because it is very short or many of its words are rare, then it may load very heavily onto a single frame whilst being somewhat atypical of the wider dataset.

Table 1: Manual analysis results

First researcher	Degree of compatibility	Second researcher
Responsibility for regulatory failure	High	Regulator or government to blame
Banking system failure --cultural problems --ethics	Medium – First researcher identified potential cultural problems vs ethics subframes, while the second researcher felt that the abstract reporting of ‘misconduct’ and moralistic reporting of disgraceful behavior were an important top-level distinction	Misconduct occurred
		Banks have behaved scandalously/criminally/greedily
Role of banks in Australia	High: the first researcher’s frame essentially combines the two from the second researcher.	Impact on markets, profitability, and bank investors
		Banking will be better in the future
Victims of banking misconduct	High	Focus on individual victims
Decline in trust	None	No equivalent
Regulators and politicians are tough on banks	None	No equivalent

Recall that a frame as defined by Entman has a problem definition and causal diagnosis. The first frame, regulatory failure, identifies the problem as banks breaking the law and the cause of that failure as regulatory laxness. The second frame is differently labeled by the two analysts, which is why they are classified as medium compatibility. But they both define the problem of banks breaking the law, with the cause of that failure seen as internal to the logic of banks: variously, because they are greedy or because they lack ethical reasoning. The third frame – defined by researcher 1 as “role of banks in Australia” changes the problem definition to problems of the Australian economy caused by market reactions to the scandal. The fourth frame – victims of banking misconduct – emphasizes the problem as the individual losses suffered by particular people, with the cause being banking misconduct. This frame is notably episodic, portraying an issue in terms of the specificity of an individual (Iyengar, 1991). Our first researcher also identified two additional frames – decline in trust and regulators/politicians are tough on banks – but these were not reflected in frames identified by the second researcher.

k-means

Using Burscher et al.'s feature selection with k-means clustering has proved only very modestly successful for frame identification on our datasets; we conclude that the groups identified are not obviously frames, and are neither entirely coherent or clearly externally differentiable. The best Australia model was that where $K=6$ (Table A1). Cluster 6 contains 71% of articles, and our interpretation of the frame cannot be more precise than 'banking'. Four others are somewhat more promising: Cluster 1 is a small grouping focusing on banking culture, its consequences, and the need for accountability; it might be interpretable as a frame. Clusters 2, 4, and 5 cover a pending class action against the bank AMP, scandals affecting the banks' wealth management arms, and critiques of bankers' bonuses. These are most readily interpretable as topic clusters rather than emphasis frames emphasizing a problem and a causal diagnosis. The final group, cluster 3, is about the Banking Royal Commission itself, but without a particular framing emphasized. As a whole, these results have reasonable coherence and differentiability, but only cluster 1 appears as an emphasis frame. We would not be content to treat this output, directly, as a frame set against which to code.

The analysis of the comparator USA Today dataset also does not fare well, either the baseline corpus (best model $K=10$, Table A5) or when filtered for relevance ($K=8$, Table A8). In both cases a single bank cluster absorbs the majority of the articles (62% and 60% respectively) without presenting a specific emphasis, and the remaining clusters (such as entertainment, central banks, investment funds, economic news, stock markets, consumer banking, politics, banking regulation, food banks, and mortgages) are topics and do not coherently present a problem or interpretation.

We believe that the key issues for both datasets are tied into features of the k-means clustering, rather than of the NLP-based feature selection that preceded it. The most

important of these is that k-means is single-membership: it models articles as being in precisely one cluster, rather than as a mixture of clusters. Consequently, the groupings are unsubtle. Where a cluster covers politics, for example, if the co-occurrence of politics-related words is stronger than those indicating the politicians' particular views on banking, then the article will get grouped with politics as a whole. Multiple membership clusterings, by contrast, allow both sets of affinities to come out, with an article capable of being coded as political, but also of critiquing banking culture.

EFA

For the EFA model (best K=8, Table A2), one of the factors for the Australian data is wildly diverse, covers 40% of all content, identifies common uninformative words as being key, and has no coherent explanation. The intuition here is that the underlying PCA-type transformation has extracted a large common factor from the textual data: ordinary language such as “one”, “just”, and “like”, used in many articles. By contrast, the next two factors were both more coherent, more externally differentiated, and better reflecting our understanding of frames. The first is a ‘legal’ frame, featuring both the quasi-judicial proceedings of the Commission itself and other legal challenges facing the banks. The second gathers a ‘business as usual’ package, focusing on market movements and the impacts of scandal on banks’ investors. Unfortunately, this coherence is not reflected in the following four frames, which seem to group several angles into each. The final frame, 8, is a politics frame and collects a number of primarily legislative terms such as “labor”, “parliamentari”, and “senat”.

While we, again, would not code these directly, the potential ‘legal’ frame was not picked up in the same way by any of the other methods or our manual analysis, and so the analysis generated some potential insight. The EFA analysis did seem fairly well attuned to picking up moralistic and emotive language, with the exemplar headlines bringing out emotional qualities in a way which the other techniques did not.

For the USA Today analyses (baseline $K=10$, Table A6; filtered $K=7$, Table A9) we found the results very poor. Factors 5 and 6 in the baseline model and 4 and 5 in the filtered model all cover the mortgage market, but are difficult to differentiate from each other and have no clear frame interpretation. The other factors were a mixture of subjects we found impossible to coherently interpret even as topics.

STM

We started by running a standard STM model for both the Australia data ($K=7$, Table A3), and for the USA Today dataset (baseline $K=11$, Table A7; filtered $K=8$, Table A10). The Australia results look promising, with Topics 1, 2, and 4 most clearly interpretable as being coherent and having a problem definition and causal diagnosis. The others cover general misconduct, and are split by major bank in a topic-like manner rather than strictly one of emphasis. Nevertheless, these are workable groupings with the potential to inform an analyst using them. For the USA Today datasets, on the other hand, the topic model heritage of the STM technique shows through clearly. Each of the groups are coherent, all are clearly differentiated from each other in content, and none reflects a frame – that is, an interpretive lens that gives a problem definition and causal interpretation of the problem (Entman, 1993). Rather, each is a division of the source material into component topics such as central banking, consumer banking, and sports results. As with the k-means results, the filtered and baseline models differ most significantly only in that the ‘extraneous’ topics such as television were largely removed from the filtered model entirely, whereas the baseline model identifies and splits them out itself.

Both sets of results appear constrained by the choice of a relatively small K . Although we set this low in part so that the resulting frames could be used as-is when our theoretical expectation is that there are relatively few frames, it seems clear that for our data the models

are both separating things which are not frames (which suggests lowering K) but also combining things which we believe should be separate (which suggests raising K).

Treating the selection and filtering of candidate frames as a third stage of the process would allow the introduction of a more theory- and expertise-driven third part of the process in a way which is potentially generative as well as more useful. By allowing an increase in the value of K in the initial part of the pipeline, a greater set of potential insights can be examined for value, while allowing a reduction in those coded to minimize the noise in the final dataset. We therefore followed up the standard STM with the second STM approach outlined in the methods section, combining what we considered to be the strengths of each of the different methods, together with the use of other techniques from the literature including direct researcher expertise.

In some ways, this final method is an analogous process to that of a manual frame analysis, but on a carefully feature-selected and then dimensionally-reduced large dataset rather than directly on a sample of articles. The raw results are in Table A4 (K=29)¹⁷, with a summary comparison to the manual analysts' frames depicted in Table 2.

Table 2: Combined and semi-manual STM results

First researcher [R1]	Second researcher [R2]	Combined and semi-manual STM	
Responsibility for regulatory failure	Regulator or government to blame	Financial regulator failure [12] – though some terms refer to regulator without necessarily including regulatory failure	
Banking system failure (sub-frames: cultural problems vs ethics)	Misconduct occurred	Banks' bad behavior [clusters 5, 10, 14, 18, 23, 26] – comparable to both R2 frames and the ethics sub-frame from R1	Banks' culture and need for reform [8, 21, 24] – comparable to the R1 cultural problems sub-frame
	Banks have behaved scandalously/criminally/greedily		

¹⁷ We calculated models using the new feature matrix over a wide range of values for K and evaluated them based upon the same held-out likelihood and residuals metrics from before, but with the constraint that the final K had to represent a plausible number of distinct frames relaxed. Consequently, a value for K closer to the optimum number of topics was identified, c. K=30, and a manual analysis of models from K=28 to K=32 resulted in the selection of K=29 for the final model.

First researcher [R1]	Second researcher [R2]	Combined and semi-manual STM
Role of banks in Australia	Impact on markets, profitability, and bank investors	Banks role in markets, and the interests of their investors [9, 28]
	Banking will be better in the future	-
Victims of banking misconduct	Focus on individual victims	-
Decline in trust	-	-
Regulators and politicians are tough on banks	-	Banks under pressure [16, 19, 29] – comparable to R1 frame, but with focus on difficulties for banks rather than the impact of the regulators
-	-	The difficulties of individual banks and their CEOs [2, 3, 15]

The two human analysts produced overlapping but not identical frames when selecting and merging the STM clusters. Two of the frames consistently identified by our manual researchers – banking system failure and role of banks in Australia – are also identified using the modified STM model. Both those frames have the characteristic that the problem definition and the causal diagnosis are quite conceptually close to each other. In the first, there is banking misconduct because of failures (of culture or ethics) internal to banks; in the second, there is an economic problem (stock market woes, interest rate rises) causally related to the banking scandal. The modified STM also came up with a frame we call ‘banks under pressure’, which is close to the ‘tough on banks’ frame of R1, but regulators are largely outside the frame of pressure – the problem is of pressure on CEOs and banks from economic actors, such as boards or investors, and the cause is the banking scandal.

The dog that almost did not bark in the night in our modified STM are the two manual analysis frames that we – based on an admittedly qualitative assessment – would judge to be the most politically salient. The responsibility for regulatory failure frame, which was in fact a prominent theme in the governing Liberal Party’s response to the crisis (Culpepper and Lee, 2019), was present in only a single topic identified (Topic 12), and even then some of the

terms identified referred to regulators without connoting failure. The episodic focus on the individual victims of the scandal, which was a central plank of the opposition Labor Party's response to the Australian scandal (albeit a relatively uncommon framing within the newspapers' coverage), was entirely missed by our modified STM. Both frames link banks with distinctly non-bank subjects: regulators on the one hand, individual victims of those banks on the other. This is merely an observation, which may be idiosyncratic to our study. But it is also possible that our modified STM ran into difficulty linking inter-subjective meanings that weren't in the text – but that could be spotted by our human coders. Much more research would be required to say anything more definitive.

We believe that the STM results, both in their plain low-K and final semi-manual high-K forms, highlight something important about the computational framing enterprise. Critics of topic models in frame analysis are wholly supported by the USA Today analysis, while proponents should be buoyed by the Australia case, especially in the results of the final model. The difference is in the scope of the data given. For the Australia data, the coverage is tight and thus the intrinsic variation between the different articles is lower than for the USA Today data, whether keyword-filtered or not. Consequently, when building topics to differentiate groups of articles from each other using STM, the USA Today data tends towards clusters based on gross characteristics – the very different subjects being discussed in the broad USA Today dataset – while the Australian model is able to extract subtler variations based more closely on tone, emphasis, and repertoire.

Analysis

Only in the case of the STM results would we be potentially happy directly coding the Australian results to articles as frames and then using the results in some downstream

analysis. While the other methods clearly extracted inter-article linguistic variation, they did not identify what we consider to be useful frames. In the case of both STM and EFA, we believe that the multiple-membership nature of the clustering allowed better results for frame analysis than the simpler single-membership outcome of k-means clustering.

We also found, however, that the opportunity to look through the lenses provided by the different analysis systems gave us new insights into our data which we did not get from the manual coding exercise we carried out. Our prior belief was that differences in corpus scope and scale would be important in the behavior of these models. That belief has been amply demonstrated by the substantially worse performance of all methods on the wider and noisier USA Today dataset, whether filtered or not for extraneous articles. Put simply, the results on this dataset simply look less like frames than they do on the narrower Australian one. On the other hand, we found few important differences between the filtered and unfiltered USA Today datasets: the clustering algorithms which were most successful on the USA Today dataset were equally capable of separating the extraneous articles into their own group. One key message from our results, therefore, is the importance of selecting an appropriately scaled dataset¹⁸.

For the broad USA Today dataset, we believe that the linguistic variation between articles on different kinds of news articles featuring banks (e.g. coverage of retail banking vs investment returns) is overwhelming the subtler variation caused by different approaches to similar topics. It is easy to see how models based upon maximizing bag of words differences between groups could extract topics based on gross differences between vocabulary where the articles are very diverse and on quite separate subjects, while extracting something more akin to a frame based on more subtle variations in word co-occurrence if given a set of articles tightly grouped around one topic or story. –It has been previously noted that co-

¹⁸ Here it is worth noting that we did not even try to submit the broad datasets to our human coders. We were hoping the computational techniques could make sense of broad scope data that we judged extremely challenging for human coders.

occurrence patterns of particular words can represent issues, events, writing styles, or frames, depending on their content (Jacobi et al., 2016, pp. 90–91). We might therefore reasonably expect these computational frame identification methods to work less well the higher up the ladder of abstraction they are applied (‘bank’) and better the lower down (for example, a carefully-collected comparable series of opinion pieces about a discrete event relating to banking). The nuances of each method also matter. The methods we used were published as unitary systems, each combining various pre-processing and clustering techniques into standalone and separate packages. It is common in machine learning work to separate the feature selection and classification steps, allowing for separate selection and optimization of each stage. That is a practice we also endorse in computational frame analysis work.

Future directions

The multifaceted and scale-varying nature of frames means that the computational tools for analyzing media frames are, in fact, adjacent to many other techniques in computational text analysis from other disciplines. Depending on the exact scale we are interested in, or the particular conceptualization of framing that we are pursuing, some of these may have traction. Even if not, the ways in which computer scientists and computational linguists conceptualize and operationalize this kind of text analysis can be of value. Although computational inductive frame detection is now well established, it remains an active area of research.

More radically, rather than using the bag-of-words representation of articles, improvements may be possible by using document embedding models. These train deep learning models to build representations of a document in n-dimensional vector space, such that semantically similar documents have representations which are close in distance. A well-known approach

for this purpose is the doc2vec¹⁹ model (Le & Mikolov, 2014). These vectors could then be used in place of a DTM as a document representation for downstream clustering tasks.

As another possible approach, the underlying computational problem could be seen as one of separating a lexical co-occurrence network into related sections, which is the subject of substantial research in other academic fields. Network analysts, for example, partition networks into clusters using several different statistical techniques (Lancichinetti & Fortunato, 2009). A recent computational communication science paper takes a slightly different approach, applying network partitioning to the output of topic models, which again looks like an excellent way forward (Walter & Ophir, 2019).

Conclusions

There is significant value in understanding how different computational methods can give insight into possible media frames, but we should not simply expect them to produce frames that directly replicate those of human coders: in our analyses there were substantial differences between each other and the humans.

In many ways this mirrors the diversity among manual analysts in conceptions of framing and in the consequent high degree of inter-coder variation. Even with careful instructions and a shared understanding on quite a focused task, our two manual frame coders emphasized somewhat different sets of frame packages and produced distinct (though recognizably similar) potential frames for coding.

The STM approach was effective, but only when dealing with a narrow scope of articles. Neither STM nor any of the other techniques produced frame-like groups on the broader USA Today dataset. The question then is not whether computational methods can inductively derive frames, but on what kinds of input data are they likely to be successful: scale matters.

¹⁹ Called ‘Paragraph Vectors’ in the original Le & Mikolov (2014) paper; this is an extension of the word2vec word embedding model to model chunks of text rather than single words.

We have shown in the second STM analysis of the Australia data that researchers have the opportunity to mix-and-match theoretically promising aspects of different methods to combine these insights. . The addition of researcher expertise to computational approaches is valid and useful, and should be explored where it seems warranted.

Overall, computational techniques provide both the opportunity to look for data-driven insights into framing problems and potentially the generation of candidate frames for semi-supervised analyses which build on the strengths of both computational and manual approaches. As computational framing methods are relatively young, it remains an active research area and there are several potential avenues for significantly improved performance.

Acknowledgements

We would like to thank Hayley Pring for her research assistance, the editors and reviewers for their incisive and productive critiques, and the members of the ICA Computational Methods interest group who provided useful comments on an earlier version of this paper at the 2019 ICA Conference. We would also like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work.

<http://dx.doi.org/10.5281/zenodo.22558>

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 787887). This paper reflects only the author's views; the European Research Council Executive Agency is not responsible for any use that may be made of the information it contains.

References

- An, S.-K., & Gower, K. K. (2009). How do the news media frame crises? A content analysis of crisis news coverage. *Public Relations Review*, 35(2), 107–112.
<https://doi.org/10.1016/j.pubrev.2009.01.010>
- Baden, C. (2018). Reconstructing frames from intertextual news discourse: A semantic network approach to news framing analysis. In P. D'Angelo (Ed.), *Doing news framing analysis II: empirical and theoretical perspectives* (2nd ed., pp. 3–26). Routledge.
- Baden, C., & Tenenboim-Weinblatt, K. (2017). Convergent news? A longitudinal study of similarity and dissimilarity in the domestic and global coverage of the Israeli-Palestinian conflict. *Journal of Communication*, 67(1), 1–25.
<https://doi.org/10.1111/jcom.12272>
- Barnes, L., & Hicks, T. (2018). Making austerity popular: The media and mass attitudes toward fiscal policy. *American Journal of Political Science*, 62(2), 340–354.
<https://doi.org/10.1111/ajps.12346>
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, 3(30), 774. <https://doi.org/10.21105/joss.00774>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.
- Bouchet-Valat, M. (2018). *tm.plugin.lexisnexis: Import articles from 'LexisNexis' using the 'tm' text mining framework*. <https://github.com/nalimilan/R.TeMiS>
- Burscher, B., Odijk, D., Vliegthart, R., Rijke, M. de, & de Vreese, C. H. (2014). Teaching the Computer to Code Frames in News: Comparing Two Supervised Machine Learning Approaches to Frame Analysis. *Communication Methods and Measures*, 8(3), 190–206. <https://doi.org/10.1080/19312458.2014.937527>

- Burscher, B., Vliegenthart, R., & de Vreese, C. H. (2016). Frames beyond words: Applying cluster and sentiment analysis to news coverage of the nuclear power issue. *Social Science Computer Review*, 34(5), 530–545.
<https://doi.org/10.1177/0894439315596385>
- Busby, E., Flynn, D. J., & Druckman, J. N. (2018). Studying framing effects on political preferences: Existing research and lingering questions. In P. D'Angelo (Ed.), *Doing news framing analysis II: empirical and theoretical perspectives* (2nd ed.). Routledge.
- Chong, D., & Druckman, J. N. (2007). Framing theory. *Annu. Rev. Polit. Sci.*, 10, 103–126.
<https://doi.org/10.1146/annurev.polisci.10.072805.103054>
- Culpepper, P.D., & Lee, T. (2019). Framing the Bankers: Media and Public Attitudes on the Australian Banking Scandal. Paper presented at APSA annual meeting, 2019.
- David, C. C., Atun, J. M., Fille, E., & Monterola, C. (2011). Finding Frames: Comparing Two Methods of Frame Analysis. *Communication Methods and Measures*, 5(4), 329–351. <https://doi.org/10.1080/19312458.2011.624873>
- de Vreese, C. H. (2005). The Spiral of Cynicism Reconsidered. *European Journal of Communication*, 20(3), 283–301. <https://doi.org/10.1177/0267323105055259>
- DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41(6), 570–606.
<https://doi.org/10.1016/j.poetic.2013.08.004>
- Druckman, J. N., & Nelson, K. R. (2003). Framing and deliberation: How citizens' conversations limit elite influence. *American Journal of Political Science*, 47(4), 729–745.
- Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. (2020). *GNU Octave version 5.2.0 manual: A high-level interactive language for numerical computations*.
<https://www.gnu.org/software/octave/doc/v5.2.0/>

- Entman, R. M. (2003). Cascading Activation: Contesting the White House's Frame After 9/11. *Political Communication*, 20(4), 415–432.
<https://doi.org/10.1080/10584600390244176>
- Feinerer, I., Hornik, K., & Meyer, D. (2008). Text mining infrastructure in R. *Journal of Statistical Software*, 25(5), 1–54.
- Gerring, J. (2001). *Social science methodology: A criterial framework [electronic resource]*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815492>
- Gilardi, F., Shipan, C. R., & Wüest, B. (2020). Policy Diffusion: The Issue-Definition Stage. *American Journal of Political Science*. <https://doi.org/10.1111/ajps.12521>
- Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, 21(3), 267–297.
<https://doi.org/10.1093/pan/mps028>
- Iyengar, S. (1991) *Is Anyone Responsible?* University of Chicago Press.
- Jacobi, C., van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89–106.
<https://doi.org/10.1080/21670811.2015.1093271>
- Kahneman, D., & Tversky, A. (1984). Choices, frames and values. *The American Psychologist*, 39(4), 341–350.
- Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E*, 80(056117), 1–11.
<https://doi.org/10.1103/PhysRevE.80.056117>
- Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. *International Conference on Machine Learning*, 1188–1196.
- Leeper, T. J., & Slothuus, R. (2018). *How the News Media Persuades: Framing Effects and Beyond*. <https://s3.us-east-2.amazonaws.com/tjl-sharing/assets/HowTheNewsMediaPersuade.pdf>

- Maher, T. M. (2010). Framing: An emerging paradigm or a phase of agenda setting? In S. D. Reese, O. H. Gandy, & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 83–94). Routledge.
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2–3), 93–118.
<https://doi.org/10.1080/19312458.2018.1430754>
- Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. <https://nlp.stanford.edu/IR-book/information-retrieval-book.html>
- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 55–60.
- Matthes, J., & Kohring, M. (2008). The Content Analysis of Media Frames: Toward Improving Reliability and Validity. *Journal of Communication*, 58(2), 258–279.
<https://doi.org/10.1111/j.1460-2466.2008.00384.x>
- Morstatter, F., Wu, L., Yavanoglu, U., Corman, S. R., & Liu, H. (2018). Identifying Framing Bias in Online News. *ACM Transactions on Social Computing*, 1(2), 1–18.
<https://doi.org/10.1145/3204948>
- Motta, G., & Baden, C. (2013). Evolutionary Factor Analysis of the Dynamics of Frames: Introducing a Method for Analyzing High-Dimensional Semantic Data with Time-Changing Structure. *Communication Methods and Measures*, 7(1), 48–82.
<https://doi.org/10.1080/19312458.2012.760730>
- Motta, G., & Baden, C. (2015). *MATLAB program for the analysis of Semantic Frames*.

- Muddiman, A., McGregor, S. C., & Stroud, N. J. (2019). (Re)claiming our expertise: Parsing large text corpora with manually validated and organic dictionaries. *Political Communication*, 36(2), 214–226. <https://doi.org/10.1080/10584609.2018.1517843>
- Nicholls, T., & Culpepper, P. D. (2020). Replication code for ‘Computational identification of media frames: strengths, weaknesses, and opportunities’. <https://doi.org/10.17605/OSF.IO/F5H8A>
- Norris, P., & Inglehart, R. (2019) *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge University Press.
- R Core Team. (2019). *R: A language and environment for statistical computing*. <https://www.R-project.org/>
- Reese, S. D. (2010). Framing public life: A bridging model for media research. In S. D. Reese, O. H. Gandy, & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the social world* (pp. 7–31). Routledge.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2016). Navigating the local modes of big data. In R. M. Alvarez (Ed.), *Computational Social Science: Discovery and Prediction* (pp. 51–97). Cambridge University Press.
- Roberts, M. E., Stewart, B. M., & Tingley, D. (2018). *stm: R Package for Structural Topic Models* (Version 1.3.3) [Computer software]. <https://cran.r-project.org/web/packages/stm/index.html>
- Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for open-ended survey responses. *American Journal of Political Science*, 58(4), 1064–1082. <https://doi.org/10.1111/ajps.12103>
- Rodrik, D. (2018). Populism and the economics of globalization. *Journal of International Business Policy*, 1(1), 12–33. <https://doi.org/10.1057/s42214-018-0001-4>
- Sagi, E., & Dehghani, M. (2014). Measuring Moral Rhetoric in Text. *Social Science Computer Review*, 32(2), 132–144. <https://doi.org/10.1177/0894439313506837>

- Scheufele, D. A., & Iyengar, S. (2017). The state of framing research: A call for new directions. In K. Kenski & K. H. Jamieson (Eds.), *The Oxford Handbook of Political Communication* (pp. 619–632). Oxford University Press.
- Scheufele, D. A., & Tewksbury, D. (2007). Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. *Journal of Communication*, 57(1), 9–20. <https://doi.org/10.1111/j.0021-9916.2007.00326.x>
- Schofield, A., & Mimno, D. (2016). Comparing apples to apple: The effects of stemmers on topic models. *Transactions of the Association for Computational Linguistics*, 4, 287–300.
- Shahin, S. (2016). When scale meets depth: Integrating natural language processing and textual analysis for studying digital corpora. *Communication Methods and Measures*, 10(1), 28–50. <https://doi.org/10.1080/19312458.2015.1118447>
- Taddy, M. (2012). On estimation and selection for topic models. In N. D. Lawrence & M. Girolami (Eds.), *Proceedings of the fifteenth international conference on artificial intelligence and statistics* (Vol. 22, pp. 1184–1193). PMLR. <http://proceedings.mlr.press/v22/taddy12/taddy12.pdf>
- Van Gorp, B. (2010). Strategies to take subjectivity out of framing analysis. In P. D'Angelo & J. A. Kuypers (Eds.), *Doing news framing analysis: Empirical and theoretical perspectives* (pp. 100–125). Routledge.
- Vliegenthart, R. (2012). Framing in Mass Communication Research – An Overview and Assessment. *Sociology Compass*, 6(12), 937–948. <https://doi.org/10.1111/soc4.12003>
- Vlieger, E., & Leydesdorff, L. (2012). Visualization and Analysis of Frames in Collections of Messages: Content Analysis and the Measurement of Meaning. In M. Mora, O. Gelman, & M. S. Raisinghani, *Research Methodologies, Innovations and Philosophies in Systems Engineering and Information Systems* (pp. 322–340). Information Science Reference. <https://arxiv.org/pdf/1112.6286>

- Walter, D., & Ophir, Y. (2019). News Frame Analysis: An Inductive Mixed-method Computational Approach. *Communication Methods and Measures*, 13(4), 248–266.
<https://doi.org/10.1080/19312458.2019.1639145>
- Young, L., & Soroka, S. (2012). Affective News: The Automated Coding of Sentiment in Political Texts. *Political Communication*, 29(2), 205–231.
<https://doi.org/10.1080/10584609.2012.671234>

Appendix: Tables of identified frames

Table A1: Australia: Burscher k-means

Frame	Prop	Key words	Exemplar headlines
1	0.02	culture cba governance report damning inquiry risk pay review result accountability bank remuneration company final key poor cricket executive australian	Hayne urged to break up the big banks; CULTURE 'Fiduciary duties ... Hayne urged to break up the big banks; CULTURE Fiduciary duties ... CBA's scorecard horrifying, but not totally unexpected Here we go yet again at the CBA; Page Two Culture key for companies to avoid wrath Review of CBA culture set to be released tomorrow Culling poor bank culture is going to need more time Man at the heart of a malignant banking culture Damning report is a must read Review of CBA culture set to be released tomorrow; BANKING
2	0.02	action class amp court suit lawsuit firm legal wealth law group embattled lawyer scandal-hit value shareholder damage threat market fee	The jury's out on class actions AMP's third class action Bidding war in AMP class action AMP faces class action triple play Fourth class action against AMP lodged with court AMP class action under wraps AMP to face class actions Court to rule on who gets a piece of the action Firms face off in AMP class action Malaysians sue ANZ over 1MDB saga
3	0.18	inquiry commission royal bank banking loan sector financial victim big service government industry scandal home week report hearing misconduct year	ROYAL COMMISSIONS THE ROYAL COMMISSION Why are banks so scared of a commission of inquiry? Smaller banks on alert over inquiry Start with the banks, PM Inquiry can be a 'circuit breaker' The royal commission is painti... Greens on mission to 'break up the banks' No error in delaying royal commission: O'Dwyer New PM should start with the banks
4	0.05	advice wealth arm management financial business mlc profit compo bank ban amp former key scandal- prone exit planning poor scandal chief	Fancy flats keep Cath in clover Stop tinkering with financial advice: IOOF ASIC missed shot at stopping the rot in advice years ago; BANKING ... AMP unveils 74pc profit plunge Perpetual motion: Lloyd tolead NAB wealth spin-off Customers flood CBA boss with advice of their own Evans Dixon boss stands by vertically integrated products CBA to pay \$3m over advice fees Macquarie to merge advice, private bank 'WE GOT IT WRONG'
5	0.02	bonus pay cut executive top boss bank exec chief	NAB plans to change criteria for bonuses 'Clawback' for AMP bonuses \$14m CEO bonus cloud

Frame	Prop	Key words	Exemplar headlines
		still banker bad excessive table board big remuneration scheme nab option	Bank exec bonuses targeted How bad must it get for banks to cut bonus? Bonuses for bank bosses could be scarce this year Customer pain to hit NAB bonuses Crisis at CBA sees bonuses binned How bad must it get for banks to cut bonus?; COMMENT Crackdown on big banks' exec bonuses
6	0.71	bank scandal amp super board fee chief financial new big executive corporate fund investor banking customer regulator year service cba	ANZ slides on \$824m red flag ALP 'plays politics' on banks tour Rising star falls What you can buy for a lazy \$70m Ex-chair's unexploded ordnance Rich-listers rally behind Brenner Joyce, Campion welcome baby boy; Family IMMIGRATIONADD IT UPHigher GDP... Fancy flats keep Cath in clover Shorten the pocket battleship Peats proving fit for Gold Coast captaincy CITY BEAT

Table A2: Australia: EFA

Frame	Prop	Key words	Exemplar headlines
1	0.40	one time just like also way now made get right think make peopl well come across australia say want look	BYRES' MARKET; Rising regulator Big business? Bigger crunch coming; PageTwo BEWARE THE BEAR; Supervision RBA's plea to banks: front up; Banking The big end of town has no one but itself to blame; ... THE DEMOCRACY BLUES THE DEMOCRACY CANBERRA SCAPEGOATS AGL Trust, technology hits bank oligopoly BCA is pushing policy irrelevance; Advocacy
2	0.19	counsel commiss document royal asic evid lawyer hayn kenneth fee mislead refer law result legal global commission act conduct servic	Gillard lover told lawyer of fraud NAB's straw man simply 'loose language' Banking inquiry flags big changes 'Liar loans' come into sharp focus; ANALYSIS NAB's sad tale of fraud and dodgy loans NAB's sad tale of dishonesty and delays ASIC to give its story to Hayne CSL just the tonic as index surges to a decade high Banks don't impress in failing first test; COMMENT Banks fail first test and it doesn't impress Commissioner Hayne; DAY ONE
3	0.12	lend analyst cent per loan sharehold amp share billion investor join investig price cover minist stock anz bank cba million	Big banks still raking it all in ANZ's Elliot a safe choice by bankers AMP board in weekend crisis meeting ANZ accused over pushing wealth advice Banks reap \$30b from tough year Right royal minting machines NAB staff feel the pain of hard year Triple class-action threat puts more heat on AMP Big four bank \$29.5m for year

Frame	Prop	Key words	Exemplar headlines
			Pressure builds on AMP to sack chairwoman
4	0.09	turnbul minist westpac complianc half leader parti elect earn fall prime profit senat labor parliament malcolm analyst investor share sharehold	New CBA chief executes purge of the Narev executive old guard MARKET WRAP Comyn people? Not this senior CBA trio Labor offers to work with Nationals to set up bank probe; Royal ... BUZZ'S TOP 50 Energy, materials lead the gains CBA's new chief clears the decks CBA's new chief clears the decks; BANKING Reporting lines under review Why there'll be no royal commission, yet; analysis MARKET WRAP RETAILERS, MINERS KEEP BOURSE ON A ROLL
5	0.07	execut matt malcolm team live comyn michael andrew share poor superannu turnbul amp appoint paid industri narev nsw board chief	Self-confident Turnbull overreaching to the end COALITION IN CRISIS BUZZ'S TOP 50 Bourse retreats in quiet trade Morrison warns all options open; CBA board cuts \$16m in executive bonuses ... Rear Window Hayne commission offers chance to cash in: BKI Experts urge CBA to go further than bonus cuts Arbitration the talk of the town Bankers' \$16m bonus cut
6	0.05	advic super countri among announc octob feder limit place banker appoint turnbul minist cent dollar per done april malcolm ian	Banking royal commission details due Scramble to find new ASIC chair as top pick withdraws Case for a federal ICAC is growing CBA in for rough trot, says Citi; FINANCE Royal Commission hearings ... CBA in for rough trot, says Citi Banks to file embarrassing dirty laundry Bourse shrugs off regional falls Move on planners' 10yr rule LET RBA WATCH OVER BANKS Future Fund flexes its muscles
7	0.05	list super sector superannu fund annual labor mark industri march nation left prime minist already earlier nab team explain boss	Westpac finance exec in hot seat over rate-rig trial Bank bosses take a hit Top funds and regulators in the firing line Arbitration the talk of the town CBA in for rough trot, says Citi CBA in for rough trot, says Citi; FINANCE Royal Commission hearings ... How did insurers get it so wrong? Rising star falls Bonuses for bank bosses could be scarce this year Hayne demands fresh AMP, bank documents ahead of super probe
8	0.04	labor opposit present parliamentari fund penalti establish littl senat life along forward malcolm parti push nation super due fine incent	Unfair loans face 'perfect storm' UBS warns on household debt blowout Banks to file embarrassing dirty laundry Shorten faces split as Dastyari departs AMP to sell off life insurance as shares plunge Super savers face gouge of \$80m Banking royal commission details due Gen Y advisers shy away from big banks Sad tale of rising star senator who flew too close to the ... Watchdog accuses embattled CBA of rigging interest rate; FRESH SCANDAL

Table A3: Australia: STM

Frame	Prop	Key words	Exemplar headlines
1	0.16	prime coalition shorten party minister cricket labor politics turnbull church liberal political office dastyari tax care joyce election senator great	'I've offered to MC my gay relative's wedding' FIGHT TO UNITE A QUESTIONABLE WASTE OF TIME SAM'S GONE, HIS DASH DONE Time to end the Coalition: Christensen Self-confident Turnbull overreaching to the end Plots lost and found in a topsy-turvy year; A roller-coaster year has ... Sad tale of rising star senator who flew too close to the ... COALITION IN CRISIS ScoMo floating away
2	0.16	earnings growth costs dividend net prices cent per housing rates higher analysts profit market stocks capital lower revenue asset businesses	Profit season - the good, bad and ugly Profit season -the good, bad and ugly Good, bad and the ugly Mum and dad shareholders still miss out ASX rout punches hole in returns Westpac sticking with wealth as compo payouts hit profits; BANKING Westpac sticking with wealth as compo payouts hit profits Time to cash out? Argo wary of volatility a ft er half-year lift Analysts give CSL a shot in the arm
3	0.09	directors boards reelection vote director murray kramer wallace wilkins board women gonski female agm nonexecutive shareholders gender shareholder joined harmos	Women bear the brunt of scandal AMP's all-female exits reflect 'cultural issues'; governance AMP's all-female exits reflect 'cultural issues'; GOVERNANCE AMP's all-female exits reflect 'cultural issues' Riding the waves of boy's club bias Brenner to leave Coca-Cola board next year; Directorship Brenner to leave Coca-Cola board next year Brenner to leave Coca-Cola board next year; DIRECTORSHIP 'Cultural issues' behind backlash on women Chairman's light notes fall fall flat at AGM
4	0.20	powers code institutions consumers industry regulation inquiry reforms apra reference shipton regime standards penalties system conflicts legislation behaviour odwyer sector	Push to extend accountability laws New bank laws should go further: committee; REGULATION Widening of new bank laws urged Banking watchdog's threat to name 'egregious' culprits Dishonesty must stop, says ASIC's Shipton Bank inquiry lashed over shortfall of advice Inquiry's shortage of advice attacked Lift super outcomes now, says ASIC boss ASIC's banks penalties fall short: committee COMMISSION IS 'FLAWED'
5	0.09	utz clayton salter amp class regan meller amps independent firm actions misleading mised craig counsel slater brenner	Revealed: the frantic days that followed AMP's fall from grace Brenner's changes 'substantive'; AMP ROW Emails tell story Brenner's changes 'substantive' Revealed: the email trail that caused a financial giant to unravel AMP may face full force of federal law Finance titan AMP may face criminal penalties AMP class actions in one court Third class action in works for AMP

Frame	Prop	Key words	Exemplar headlines
		mavrakis jack gordon	AMP LAWSUITS ROLLED INTO ONE AMP court case tweak
6	0.13	comyn austrac narev laundering cba cbas matt livingstone transactions allegations bonuses moneylaundering deposits ian elliott deposit antimoney accounts incentives staff	Laundering failures risk \$1bn fine CBA 'failed on money laundering' Fake names, drug lords, suspicious trading and CBA; Comment Fake names, drug lords, suspicious trading and CBA; COMMENT Fake names, drug lords, suspicious trading and CBA Things will change, vows CBA's Comyn Things will change, vows CBA's Comyn; BANKS Incoming CEO faces Austrac questions Royal commission call on banks grows louder; COMMENT CBA accused of ignoring warnings in fraud scandal
7	0.16	nab nabs hagger trustee documents hodge ioof hearings fee thorburn loans loan compensation brokers trustees round orr fraud commissions nulis	NAB staff 'took bribes' Claims staff in alleged bribery ring took cash The letter that caught NAB red-handed in fee-for-no-service scandal NAB fee grab slammed by the watchdog NAB staff 'took cash bribes'; ROYAL COMMISSION ASIC sues NAB over fee claims NAB felt wrath of watchdog More civil breaches for NAB NAB fee scandal goes back to 2004, ASIC report says; Royal commission Dirt file airs ASIC ire at NAB's fees for no service

Table A4: Australia: Combined and Semi-Manual STM

Frame	Prop	Key words	Exemplar headlines	Handling
1	0.03	new focus probe choice new_chief blow deputy broader tough insider term prime_minister rule need financial_institution comyn scrutiny standard well earlier	New blow as ANZ to face cartel probe My fight to unite: new Deputy PM 'I've offered to MC my gay relative's wedding' Comyn faces challenge of new course; CHANGEOVER People first focus needed New lending crackdown threatening confidence FIGHT TO UNITE 'I've offered to MC my gay relative's wedding ' I'll MC gay wedding Witness collapses as probe turns Orwellian Cathing a new wave	Drop - coherence
2	0.02	just letter policy customer front question commissioner yesterday lack cba_boss little last test failing australia even bank_boss cba never australia_financial	Customer letters swamp CBA boss LETTERS TO THE EDITOR Self pity? That's pathetic Letters to the Editor LETTERS Hillary: "Just promise the che... Customers flood CBA boss with advice of their own Chief faces the wrath of customers Letters Sackfuls of 'complaints and compliments', CBA	Group with 3 and 15 – individual banks' and executives' issues
3	0.06	exec ceo chief cbd senior_executive senior incoming pay incoming_chief seat jail former_chief top bank_boss accountability parliamentary grilling back comyn watch	Execs in hot seat as pillar feels pinch Things will change, vows CBA's Comyn; BANKS CBA BEEFS UP ITS EXECUTIVE Things will change, vows CBA's Comyn CBD; Year in Review Looking back on a year disrupted CBD TAMING THE PAY MONSTER BUZZ'S TOP 50 Move quickly: Lessons for Comyn from Barclays veteran	Group with 2 and 15 - individual banks and executives' issues
4	0.01	rear rear_window window reason woman election profit dead chair cba canberra probe scandal enough just amp	Rear Window The deeper reason we hate banks; We need banks, but they make ... Dual win gives PM reason to re-Joyce Women bear the brunt of scandal The land of the unfair go	Drop – markers of AFR's 'Rear Window' and similar opinion columns

Frame	Prop	Key words	Exemplar headlines	Handling
		bank_financial back peak loss	Populist backlash 'no reason to stop tax cuts' 'Cultural issues' behind backlash on women ACSI hails progress on board diversity The economy might be getting stronger - but so what? Besieged ACBF sued over \$40k funeral scheme	
5	0.04	life fee_scandal fee insurance insurer life_insurance nab corporate_watchdog watchdog customer_fee hundred claim compensation document rip- service_never corporate court thousand lawsuit	Life insurers' \$6bn fee grab in crosshairs NAB 'hid compo from ASIC' ASIC launches court action against NAB over super fees ASIC prepares court bid over 'fee for no service' NAB felt wrath of watchdog ASIC accuses NAB of \$100m super rip-off ASIC drags NAB to court over super fees scandal NAB boss quits over fee scandal Ban failed to stop NAB fund fee bonanza Watchdog takes NAB to court over fees for no service	Group with 10, 14, 18, 23, and 26 - banks' poor behavior - superannuation fees
6	0.03	change right day overhaul public banker party month face option position past revelation future treasurer word disastrous group far week	Shorten faces split as Dastyari departs Sorry won't cut it when bankers face Hayne BATTLE FOR POLL POSITION GOOD WEEK BAD WEEK Can a change of the guard fix ASIC's image? 'Bankers' is now a dirty word for lobby group Bankers' charm offensive starts with a lick of paint System shaken to its core by rogues and bad planners 'Bankers' officially now dirty word for lobbyists Populism does not have to be a dirty word	Drop - coherence
7	0.04	big nation talk town deal end short australia short-term institution damage control enough bank financial_institution worst fund_manager battle canberra country	Arbitration the talk of the town The big end of town has no one but itself to blame; ... Big swing a boost for Joyce Fundies wary on CBA's dip; Equities Fundies split on CBA's prospects Mind the gender gap: small company boards lag big end of town We name the biggest laggards Big banks scrap hated ATM fees Super funds slam 'short term' banks Secrecy as big banks seek deal with ASIC	Drop - coherence
8	0.03	problem return cultural global exclusive key	The great Australian blindness; The Nation Culture key for companies to avoid wrath; VALUES	Group with 21 and 24 - Culture of banking and consequences

Frame	Prop	Key words	Exemplar headlines	Handling
		global_financial financial_crisis review crisis poor giant aussie worst possible process clear decade ago business	Culture key for companies to avoid wrath Reckoning is coming for complacency ALP to move on business corruption as leaked records reveal building giant's ... CRICKET AUSTRALIA HAS A CORPORATE CULTURE PROBLEM After day of drama, AMP rues dark night of the soul Problems ignored as stellar returns flowed ASX rout punches hole in returns A potential solution to our housing problem	for the future
9	0.02	rate major interest_rate major_bank budget credit mortgage interest card lending funding attack practice low back recent loss rise session key	Banks, energy sector drag ASX lower for fourth session; EQUITIES Rates on the rise: Westpac takes first shot with interest hike Banks, energy sector drag ASX lower for fourth session Banks, energy sector drag ASX lower; Equities Banks rush back into rate hikes as ACCC probe ends Savers lose in CBA deposit switch Lending curbs not holding back credit growth: Byres Henderson tipped for top economics post amid ongoing scrutiny of the major ... Watchdog accuses embattled CBA of rigging interest rate; FRESH SCANDAL Budget to determine legacy	Group with 28 - banks as market participants
10	0.02	client cash people long year dead shock threat cost shocking shocking_revelation revelation bill game never level last still bank_financial explosive	ANZ took 'too long' returning client cash Plots lost and found in a topsy-turvy year; A roller-coaster year has ... Bank pays for disrespecting clients NAB slashes 1900 jobs ASIC should not be a cash cow EEL BILL SHOCK Shock at the level of AMP's lies Bank pays for disrespecting clients; COMMENT Storm-lashed grower's date with destruction More jobs to go as NAB continues a restructure	Group with 5, 14, 18, 23, and 26 - banks abusing client funds
11	0.04	risk power compliance high trust cost business community ahead close also failure fresh senator political legal fine network profit even	High time to weigh up whether to stick with blue chips or ... No Coalition gain in final poll Risk in wider snooping powers THE DEMOCRACY THE DEMOCRACY BLUES Newspoll: no Coalition gain Sad tale of rising star senator who flew too close to the ... Power behind the throne: Gonski's network of influence ScoMo floating away	Drop - coherence

Frame	Prop	Key words	Exemplar headlines	Handling
			Shanghai Sam is on Santa's naughty list	
12	0.06	financial_service service service_industry financial corporate_regulator regulation cop act penalty misconduct corporate corporate_cop misconduct_financial breach service_giant sector federal tough crime service_company	Who's in line for financial services swamp? NAB breaches may be criminal acts NAB kept breaches secret: regulator; ROYAL COMMISISON Need for tougher corporate cop NAB faces probe on 'criminal breaches' Take over the running of Norfolk Island? Who would want to do ... As Australia reels over revela... Financial watchdog has failed, says MP New laws need tough regulator, experts say; CORPORATE CRIME Whistleblowers need protection	Group with 16, 19, and 29 – ASIC enforcement action
13	0.02	member hit former minister now retirement savings coalition retirement_savings prime_minister word leader witness manager week account conduct street team name	Mahathir eats own words for campaign Self-confident Turnbull overreaching to the end ASX clamp on ethics of all its members Time to end the Coalition: Christensen WESTPAC HIT IN NAME OF SIMPLICITY Conflicts hurt super members The Coalition of chaos Privacy rights are now crucial BT to cut fees and profits The missing minister in team ScoMo	Drop - coherence
14	0.06	bank behaviour australia_bank hard bad rort bad_behaviour gouging notice bill reform competition bank_big australia need practice worker account aussie payout	Will you have to pay for bank's bad behaviour? When our wants become needs, it's time for a serious discussion about ... Banks to foot \$5b bill Banks facing \$5bn payout Banks face Kodak moment Banks scramble to meet Hayne deadline Banks miss first downpayment Pay revamp makes it harder to reward bad behaviour How the Australian banks must wish we had Trump; Washington observed Trump's ... Banks' bad behaviour blowout	Group with 5, 10, 18, 23, and 26 - banks' bad behavior
15	0.05	director meeting board chair story annual role pressure annual_meeting resignation crisis next amp executive_bonus shareholder general revolt	Failure to hustle only exacerbates crisis for directors Coke director's last drinks call Call for Brenner to leave CCA board AMP in crisis meeting Coke in fizz over ex-AMP head Pressure on AMP board intensifies	Merge with 2 and 3 – individual banks' and executives' issues

Frame	Prop	Key words	Exemplar headlines	Handling
		responsibility next_month company	Shareholders pressure Brenner to quit CCA Murray tapped for AMP role AMP DIRECTORS OPT OUT Pressure on AMP as board gets ultimatum	
16	0.10	commission fallout inquiry worse banking misconduct round dishonesty hearing submission public_hearing finally statement commissioner lender bank_inquiry far public opportunity scrutiny	Digital disrupters a bigger threat to lenders than the royal commission THE ROYAL COMMISSION No error in delaying royal commission: O'Dwyer Everything you need to know about THE ROYAL COMMISISON Banking royal commission back in play ROYAL BLUE ROYAL COMMISSIONS Seven lessons we've learnt from royal commission COMMISSION IS 'FLAWED' Are scam victims finally being heard?	Group with 12, 19, and 29 - Royal commission as institution
17	0.03	pain comment nothing political bank_inquiry book world conflict inquiry turnbull_government interest move proposal government reform full election best_interest concern next	Dispute referee now a political football; Comment Sheep ship of shame causing Turnbull some pain Reforms not full solution MOVE ON BANKS DANGEROUS Nothing to see here THREE-RING CIRCUS Banks girding loins for full force of election campaign Conspiracy theories drive bank inquiry Banks brace for election attack Dispute referee is now a political football; COMMENT	Drop – markers of opinion articles
18	0.04	financial_advice home_loan advice home victim fraud dodgy loan plan bank_bad mortgage_broker financial scheme compensation remediation financial_scandal whistleblower arm number dishonesty	'Inexcusable' greed and dishonesty in financial advice; Royal Commission FINANCIAL PLANNING Home loan fraud, brokers in spotlight Dodgy scheme fostered bribes and fraud Mortgage brokers 'bypassing' NAB's living-expense check; LOAN APPROVALS Mortgage brokers in revolt Move on planners' 10yr rule Basic checks 'bypassed' by brokers Abuse victims' payout scheme in danger as states balk on costs National payout scheme for abuse victims in danger of falling apart 'Inexcusable' greed and dishonesty in financial advice; ROYAL COMMISSION FINANCIAL PLANNING	Group with 5, 10, 14, 23, and 26 - banks' bad behavior – financial advice
19	0.03	class_action law_firm	AMP class action under wraps	Group with 12, 16, and 29 –

Frame	Prop	Key words	Exemplar headlines	Handling
		action firm amp_class amp lawsuit independent suit court independent_report law legal embattled group supposedly_independent market_value supposedly counsel revelation_misconduct	Court to rule on who gets a piece of the action AMP class actions in one court Third class action in works for AMP AMP to face class actions Firms face off in AMP class action AMP faces class action triple play MONEY TALK Three lawsuits looming for AMP Third suit as lawyers line up to bash AMP	class actions against AMP for superannuation fees
20	0.03	tax cut wrong tax_cut union earlier company bottom leader consumer federal soon remuneration opposition test backlash revolt government sign best	Company tax cuts to reward the wrong people Union ties unhealthy It's the ties that bind Backlash for sponsors who jumped early Revenue more vital than tax cuts NAB's \$755m job cuts bill Spotlight on company cop Consumers better watch their step with this bunch Herald Sun Populist backlash 'no reason to stop tax cuts'	Drop – general tax cuts
21	0.05	code reputation lender small banking_scandal conduct man job code_conduct run small_business banking way run_scandal banking_industry series head banking_regulator series_scandal trust	More hearings set over finance industry scandals More hearings planned on finance industry scandals Trust in newspapers rises while social media falls Code revamp a small step in regaining trust; analysis KPIs: leading to too much temptation? 'Brave' call to appoint insider to top job Breach of trust: mystery surrounds NAB scandal Regulators need to lift fear factor, put jail on the table There are devious cricketers - then there are banks Banks play us as bunch of patsies	Group with 8 and 24 - Need for banking culture to be reformed and code of conduct
22	0.03	case almost order house economy local detail fall finance_sector yet melbourne_sector high confidence call many child gain decade sharemarket	Energy sector fuels rise in stocks Energy, materials lead the gains Case for a federal ICAC is growing Bourse shrugs off regional falls PM pins ministry in call to order CSL just the tonic as index surges to a decade high Gains in financials a boost for bourse Herald Sun	Drop - coherence

Frame	Prop	Key words	Exemplar headlines	Handling
			CBA chief applauds fall in house prices Pulling down the house	
23	0.04	wealth wealth_arm crime charge management exit manager arm embattled_wealth mlc amp organisation embattled business ban model giant trouble job corporate	Matt Comyn's fast road to the top of CBA; Banking Brains, charm ... Murray called in to steer AMP out of trouble Ban on wealth arms 'needed' NAB placed on criminal charge alert Ex-Perpetual chief to run MLC as NAB prepares exit; WEALTH MANAGEMENT Ban on wealth arms 'needed'; EXCLUSIVE Murray faces a big job at wealth giant Murray must shake up AMP A wealth of church secrets ASIC wasn't focused on job: Murray	Group with 5, 10, 14, 18, and 26 - banks' bad behavior – wealth management
24	0.02	bonus staff target sale bank_boss bank_account asset incentive branch product account still worth banker nab scam excessive pay call performance	ACCC boss hits out at bonuses Incentives encouraged risks, NAB boss admits NAB plans to change criteria for bonuses NAB chief: Bonuses led to 'excessive' risk Incentives more than money, too Meddling with kids' accounts not on Meddling with kids' accounts is not on Bank bosses cop a hit in the bonus Dollarmites bites: Scandal behind bank scheme CBA readies for another battering	Group with 8 and 21 - Bankers' excessive bonuses
25	0.03	fund super_fund super attention investment price great bit industry state investment_bank line fund_manager poor opportunity whistleblower structure rule system result	'Missed opportunity': super funds in whistleblower blast; DISCLOSURE Super fund morphs into world-beater Pay attention to super or let fee-gougers run amok No Hayne pain for industry super funds Super funds blast whistleblower policies; REGULATION 'Fix super before giving more wages to funds' Independent directors vital to super funds THE LAST GREAT IDEOLOGICAL BATTLE; Banking Cornered into a royal ... What have the industry super funds got to hide? Super in sick state	Drop – articles about super funds but without a consistent frame
26	0.02	money laundering real allegation greed warning shame terrorism systemic serious committee game cba hot player financing	It's real money to real people: ASIC; COMMENT It's real money to real people: ASIC; Comment It's real money to real people: ASIC Narev admits CBA's 'mistakes'; - Vows to restore trust over money laundering ... CBA boss will face difficult questions; COMMENT	Group with 5, 10, 14, 18, and 23 – banks' bad behavior – money laundering

Frame	Prop	Key words	Exemplar headlines	Handling
		breach pressure failure apra	CBA boss must prepare for tough questions CBA accused of ignoring warnings in fraud scandal String of jailings linked to CBA money laundry DIRTY MONEY GAMES Fake names, drug lords, suspicious trading and CBA; COMMENT	
27	0.02	analyst late report interim interim_report prudential time cba prudential_regulator likely damning lot spotlight week final governance regulation centre decision cultural	Damning report is a must read; Comment Damning report is a must read Culture at CBA back in spotlight CBA inquiry may be too little, too late; Comment CBA inquiry at risk of being too little, too late; COMMENT Governance takes centre stage CBA's damning report is the new must read for any business leader Reputation tops Comyn's CBA to-do list Upstart raids its MST-have analysts Analysts tipping radical change ahead for disclosure	Drop - coherence
28	0.04	share investor market shareholder value woe earnings stock share_price low profit full-year result trouble turmoil yesterday company still backlash aussie	Mum and dad shareholders still miss out Bourse retreats in quiet trade Profit season - the good, bad and ugly Profit season -the good, bad and ugly Investors brace for CBA results; EARNINGS Full-year figures may disappoint Why Bendigo Bank investors are sitting on their hands; Due Diligence Another \$600m wiped off value Good, bad and the ugly Woes pile up for investors Investors brace for CBA results	Group with 9 – investors' perspectives
29	0.01	issue heat scandal finance banking finance_sector hearing sector round financial already commission regulation management wake year plan bank_financial bank advice	Heat turns on planning scandals; banking advice issues probed Heat turns on planning scandals Finance sins to be aired APRA issues enforceable undertaking defence ROME FEELS HEAT FROM PELL'S HELL Coalition feels heat over delay 'With the benefit of hindsight' ... Coalition feels heat Heat on McMaster as Dover pulls plug Lift gag so banking truths can be told Morrison mulls next move in CBA scandal	Group with 12, 16, and 19 - external pressure on banks

Table A5: USA Today: Burscher et al. k-means

Frame	Prop	Key words	Exemplar headlines
1	0.04	card fee credit debit consumer bank overdraft customer rule account issuer practice retailer new rate charge debt law transaction balance	Walmart to offer prepaid card; The reloadable Bluebird to come with few ... Schools rely on college-linked card revenue; 'It is a big chunk of ... Thieves turn simple strip into cutting-edge tool; They alter the 'magstripe' ... Banks reduce credit card rewards; Financial pinch trims freebies Credit cards: Riskiest users still out of luck? 'Scary stuff' of debt, credit cards resonates in 'Maxed Out' Prepaid card fees can hurt; If you don't choose carefully, they can ... The downside to debit cards; They're convenient, but they lack consumer ... Over-limit fees ending for 2 cards; Discover, AmEx are doing away with ... Tips for gift card givers and getters
2	0.14	stock market investor europe fund dow gain rally news debt new wall street index york worry loss bear year fear	Investor caution gives pause to rally Passion dims for stock buybacks; Some investors might prefer innovation Stocks' pullback grows more painful Battered blue chips prove none are risk-free; 'It's brutal out there,' as ... Fed's caution note takes breeze out of market's sails The opening bell may have a nervous ring to it; Last week's ... World Bank report hammers stocks Stocks avoiding seasonal swoon? Markets in wait-and-see mode remain relatively flat Jittery stock investors suffer relapse
3	0.05	mortgage foreclosure home homeowner loan lender subprime borrower payment housing program buyer plan crisis aid help year market rate effort	Few get complete loan modifications; Only 12% have been converted after ... Maintain contract integrity Mortgage modifications faulted Charities get more donated homes Deadlines for mortgage servicers' reviews Military foreclosure rate up 32% over 2008; 20,000 servicemembers ... Mortgage refinancing takes off as HARP kicks in Financial WMD Joblessness drove foreclosures in 2010 Help for mortgage holders in trouble
4	0.02	tax swiss client account u.s. break bank case american plan bill federal cut credit americans charge transaction asset name deal	Rushing to do taxes? Beware of mistakes Soak the rich? Obama should start with tax evaders Yacht broker pleads guilty to hiding \$3.4M from IRS Financial adviser charged in tax case Kodak plans to snap up \$1B of its own stock; Such big ... IRS pays UBS whistle-blower \$104 million Tax records: Do you save them or shred them?; Some paperwork you ... Filing taxes on their own; Despite complex U.S. code, more are going ... Geithner contrite on tax failures; Treasury pick repaid the IRS Americans should start seeing tax rebates in May; IRS website helps ...
5	0.62	bank financial new economy u.s. year bailout economic job crisis time money usa today government nation street plan talk company	Check It Out; Goldman transformation How the senators voted Check it out Watch, Listen & Read Who's saying what, and what's behind it In theaters this weekend TECHNOLOGY LIVE Airport Check-in Watch, listen & read

Frame	Prop	Key words	Exemplar headlines
			Commissioner's departure could leave all-GOP SEC
6	0.02	sale home price market housing buyer retail estate real low foreclosure new year property u.s. short recovery drop mortgage loan	Columbia, Md.; Sales may have fallen, but prices haven't Home sales slide slightly in May Bargain hunters boost home sales in some markets; For first-timers ... Dodgers sale isn't win-win for MLB; McCourt out, but bankruptcy route ... Spokane prices hold up; Tough winter hits sales, but diversified economy helps Home sales rose a sweet 5.1% in Feb.; Surprising jump could signal ... Home prices fall 12% from '08 In Holy Land, no Western Christmas; Black Friday sales, 'Jingle Bells' don't ... Sales of new homes rise 2.7%; Reflects poor August figures Pending home sales give stocks a lift
7	0.05	rate interest cut low inflation central economy bank mortgage credit market economist growth loan washington consumer higher bond forecast jobless	Fed's rate cut fails to invigorate stocks Fed leaves rate near 0%; Central bank sees progress Odds favor smaller rate cut from Fed Fed leaves rates unchanged, sees jobs improvement What to ask before you buy an annuity; Fixed rate or variable? ... Suicide rates in Army, Marines No rate change expected from Fed; Economy pulling in separate directions Fed sticks to pledge to keep rates low; Expect figure to hover ... Rate cut 'prudent,' say Fed minutes Economy's pickup could mean Fed holds rates steady
8	0.01	fraud former security charge investor exec fund case mortgage lawsuit investigation york firm official ceo new federal possible regulator thousand	'King of music business' loses his throne; Search is on for mogul ... JPMorgan Chase to pay \$154M to settle fraud case Interagency task force to target financial fraud; Justice, Treasury ... Reserve Management accused of fund fraud; Complaint seeks fine, restitution for investors SEC charges missing fund manager with fraud; Investigators track Nadel ... Fewer suits charge securities fraud; Plaintiffs' lawyers find it harder 'to ... While he served abroad, his credit was under siege; Crooks exploit ... After raid, fraud trial underway; Iowa plant saw hundreds of deportations Wells Fargo accused of fraud; U.S. government sues bank over mortgages Morgan Keegan to pay \$200M to settle fraud charges
9	0.03	business small stress test bank loan big firm owner u.s. government money capital bad financial failure economy result today new	Bernanke urges more lending to small businesses How to tell if a bank's worthy of your money Banks try to boost business through generous perks Fancy fruit bouquets grew from flower shop; Edible Arrangements' owners ... Small businesses make Square deals; Diminutive credit card-processing ... Creative sources of seed money; Small businesses get crafty, bypass banks Businesses fight to reopen in tornado-stricken states Economists worry as businesses tighten spending Europe's banks jump at \$639B in ECB loans; Businesses should benefit Business roots; Entrepreneurial urges often take form when someone else ...
10	0.02	ceo company executive board pay chief loss subprime money former new job many package exec bad personal work street top	Corbat passed up NFL; New Citigroup CEO's passion was business Can Obama be both CEO, cheerleader? Why 'sorry' isn't in many CEOs' vocabularies; No. 1: Prison. No. 2: ... Pimco CEO sees an imbalance in economic revival CEO calls Zynga his 'crowning achievement' You can bet on MGM Mirage CEO; He's helping to usher in ... Niederauer promoted to CEO of NYSE Euronext

Frame	Prop	Key words	Exemplar headlines
			Relative newcomer named Citi's CEO; Pandit must extract bank from ... CEO polishes Goldman's image; Blankfein tries to head off bonus backlash BofA CEO Lewis plans to step down; Bank under fire over Merrill ...

Table A6: USA Today: EFA

Frame	Prop	Key words	Exemplar headlines
1	0.49	year say now even mani one make get also time like much still last less just month take can come	A very public display of non-affection; Young fans of Brown, Rihanna get ... Supplier looks beyond auto parts; Feeling industry's pain, Schmalde seeks to ... Recession finds fertile field of egg, sperm donors; Decent pay, but screening ... Simpson gossip spotlights the burden of celebrity; Experts question fuss over ... Clive Owen counts his lucky stars; 'I'm where I want to be' A year later, small cuts have changed to big ones; Families adjust ... A time of fear and pessimism -- but not for all Americans; ... Top medical trauma center still swamped by hurricane; Ike's destruction leaves ... Griffin good to go; Quarterback has game plan for NFL combine Students get a crash course in economics; Colleges open food banks, freeze ...
2	0.11	stock sampp investor market index gain growth dow economi poor fee charg confid card check central account law invest point	How to cope as the credit card environment continues to change USA TODAY's 2007 All-USA College Academic Team A new record for a new era; Go-go days long gone; in ... Fed chief suggests rate cut awaits; Credit woes, slower spending seen as ... Fed lowers outlook for growth in 2008; Notes on October meeting show ... Earnings news, Europe woes extend stocks' losing streak What to watch; Fed's QE3 fails to ignite stock market The rate cut's impact on consumers Stocks up as fall in jobless claims indicates labor market is healing S&P 500 continues to climb despite UPS' low profit growth
3	0.08	card fee credit stock democrat account republican parti payment consum polit low index investor purchas leader charg custom limit committe	New scoring formula means building a credit history will get harder A strong credit score can help you score a lower rate on ... BofA plans \$5 debit card fee; Monthly charge to hit if card ... Forge ahead with fee-avoidance strategies; More banks charging customers for ... Law passed last year makes gift cards more consumer-friendly; Penalties ... Financial reform bill clears Senate; 'We're finally restoring law and order to ... Next to go: Free checking accounts; Banks are adding on more charges Wolfowitz quits World Bank; Board accepts assertion he acted in good faith The downside to debit cards; They're convenient, but they lack consumer ... Panel says Wolfowitz violated rules, contract; World Bank committee rejects ...
4	0.07	said committe treasuri chairman financi billion bank card critic secretari congress propos bailout regul stock republican democrat investor polit senat	Trailing lead pack but still in THE hunt; Nos. 6-30 have shot ... Debit cards can derail vacationers' plans to spend wisely; Hotels, gas stations ... Developing nations poised to challenge USA as king of the hill; World ... Will credit card deal affect fees? Congressional hearings raise foreclosure concerns; Witness: 10% of ... Federal crackdown nets Capital One card holders \$150M; Consumer ... Details on the \$25 billion settlement Will regulators rein in big banks? 'We'll find out'; Fed's Bernanke and ... Agency yields mixed results; Consumer bureau has growing pains Credit card rates headed for federal slap down?; White House, Congress get ...
5	0.05	mortgag sale economist foreclosur home retail	How the senators voted Next wave of foreclosures looms; Moratoriums, overwhelmed lenders have ... Customers flee banks over fees; Survey: New costs prompt more to change ...

Frame	Prop	Key words	Exemplar headlines
		lender default said borrow unemploy slow declin spend report fund congress loan fed stock	A crash course on racetracks in Chase Benefits vary under \$25B mortgage settlement; Terms allow different relief ... Inflation fears rise as price index gains 0.2%; Concerns discussed at Fed's ... Encouraged Fed leaves interest rates alone; Economy growing; inflation fears ... \$25B mortgage settlement with banks is official; \$20 billion to help struggling ... Foreclosures take biggest dive in years in November Deutsche Bank sued in mortgage dispute; U.S. attorney points to reckless ...
6	0.04	mortgag foreclosur borrow lender loan home default hous spend afford report fed economist buyer unemploy payment lend retail propos june	USA TODAY Super 25 rankings; High school football USA TODAY SUPER 25 baseball rankings Encouraged Fed leaves interest rates alone; Economy growing; inflation fears ... Don't vilify mortgage investors Most investors say they'll stick with foreign stocks; They don't plan changes Foreclosure procedures undergo reviews; Courts could see wave of challenges ... Bernanke doesn't expect interest rate change soon; Fed chief expresses some ... Ohio attorney general sues lender; Lawsuit alleges GMAC Mortgage ... Let the market heal itself Foreclosures not halted nationwide; But more companies implement, expand ...
7	0.03	foreclosur mortgag home democrat told sale obama known activ bank communiti republican spend administr lender york file exchang loss along	Regulator cites weak JPMorgan risk controls Requests may get mixed reception Index funds make the most 'Common Sense' For the record: Tracking the medals race Dimon blames complacency for loss; JPMorgan Chase CEO says plan to ... As ETF funds grow, focus can narrow; At some point, ETFs go ... USA TODAY Super 25 rankings; High school girls basketball Online banks heat up competition with savings rates of 5% or more Few respond to foreclosure review offer; Borrowers' cases are checked; errors ... Wave of deals propels stocks higher
8	0.03	tax friday point total file america foreclosur largest washington central effort firm billion practic play mortgag final wall asset fund	USA TODAY Super 25 rankings; High school girls basketball USA TODAY SUPER 25 Rankings; Girls Down to the final 30 Legal gaps help hide assets from creditors, IRS; Court records track alleged ... Over 14,700 take IRS up on leniency; Offshore tax deal means billions ... Moratorium slows foreclosures; But filing rate could rise despite new laws ... Foreclosures soar despite efforts to help; July repossessions nearly 3 times ... Deadlines for mortgage servicers' reviews 150 targeted in Swiss bank case; UBS set to identify more clients ... IRS asks to seize \$10M in allegedly hidden assets; Nevada firm suspected ...
9	0.03	fund predict appear held global foreclosur largest final activ possibl cover report data central free agre bank revenu inform file	USA TODAY SUPER 25 Rankings; Girls Wolfowitz apologizes for securing raises for girlfriend at World Bank Barclays makes \$91B offer for ABN Amro; Biggest banking deal ever could ... Banks branch out to lure family, friends; Wachovia joins others offering gifts ... World Bank panel leaves Wolfowitz's fate in limbo; 'We have to ensure' turmoil at World Bank Turn customers into guests, hotelier suggests Fees are a deterrent Dems to push subway safety; Transit systems 'ripe, easy targets' USA TODAY Super 25 Expert Rankings; High school girls basketball
10	0.02	intern step firm ceo michael jone latest appear largest examl busi report quarter street fail	Down to the final 30 Legal gaps help hide assets from creditors, IRS; Court records track alleged ... Roundup Foreign roads can be deadly for travelers; Crashes, especially in developing ... Exotic funds win big, lose big; Extreme specialization means they're probably ...

Frame	Prop	Key words	Exemplar headlines
		white exchange obama friday analyst	Why Clinton is right to stay in the race Late-night talk TV talk highlights Time to turn Fatah into model of success in Mideast Tour de France leader ousted; Own team kicks Rasmussen out

Table A7: USA Today: STM

Frame	Prop	Key words	Exemplar headlines
1	0.11	man etpt film father movie story life police stars family actor son wife friends love men talk woman book actress	The characters in 'People Like Us' aren't anything like real people; Mawkish ... In theaters this weekend 'Porno' arouses too few laughs; Familiar crude jokes, romance don't mix Late-night talk TV talk highlights 'People' beams Pine up to the next level; 'Star Trek' actor adds ... New this weekend Motto for 'Justified' writers: What would Elmore do?; Elmore Leonard, that ... Critic's Corner A call from Mick brings satisfaction
2	0.10	fed european bernanke central greece feds economists inflation cuts economy unemployment economist eurozone euro europe recession cut rate greek economic	Fed chief vows 'substantive' action; Central bank could cut rates again when ... Fed cuts economic growth forecast; Central bankers expect contraction into ... Fed lowers outlook for growth in 2008; Notes on October meeting show ... Fed chief suggests rate cut awaits; Credit woes, slower spending seen as ... Encouraged Fed leaves interest rates alone; Economy growing; inflation fears ... Fed hints at cheerier future, keeps rates steady; Central bank sees jobless ... Consumer confidence bounces back a bit; Fed, meanwhile, cuts its growth ... Fed trims economic growth forecast; But it may be done with rate ... Rates to stay near zero, Fed says; Economy healing at 'moderate pace' Fed raises its expectations for growth and new jobs; Forecasts unemployment ...
3	0.08	fund funds china bonds dollar bond gold yields chinese yield invest dividends portfolio investments mutual chinas emerging money foreign value	Why do you own a money market fund?; They're good for reducing ... Income investors might try bond ETFs; Few low-risk investment choices ... To get 6% return now requires taking a risk Investors aren't chasing returns; Record amount of money goes into bond funds Need income off your savings? Oy vey.; Today it's nearly impossible, but ... Want to invest in gold? You have options Low interest rate dilemma; Q: With the Federal Reserve holding interest rates ... Those who dived into bond funds due for splash of reality; Rising ... What should you do about investments?; Check asset allocation, diversify, try ... Scared to wade back into stocks?; 3 baby steps to get you ...
4	0.04	sales construction district estate gas northern southern fla manufacturers manufacturing activity commercial ariz prices areas demand production retail tourism gasoline	Beige book reports modest growth in economy Economy ticks up a bit, but jobs still lagging Economic contraction slows in some regions Fed sees sluggish auto sales, 'weak' housing market Economy shows some life; Fed report reveals progress across USA Hardly anything's coming up roses, according to report Some Fed districts report moderating pace of decline Economy shows widespread pain Fed report shows spry economy despite gas prices Fed report shows steady growth across much of USA

5	0.11	dow sampp earnings poors stocks rally strategist index nasdaq jones cents standard amp fell industrial composite investors points bear shares	Earnings gallop in with flags waving; Growth isn't just from cost cutting ... Despite forecasts, growth persists; S&P 500 firms see increase in profits Credit woes crunch profits; Corporate earnings may drop for first time in ... Investors swoop in to buy bargains Profit stall raises issues; Have investors factored in drop? Profits ride to Wall Street's rescue; Stocks rally as strong results start ... Investors look for clues in earnings; Is economy as dour as it ... Dow sets a new bull market high; As election dust settles, index ... Markets mixed as bad news piles up Dow hits another record as oil drops
6	0.09	card fees debit cards fee accounts irs overdraft account issuers customers online checking transactions consumers charge checks unions credit ubs	Weigh before you pay: Debit or credit? How you pay for stuff could soon determine how much you pay Debit card fees are coming; How to avoid them; Banks ding shoppers ... Overdraft protection comes in many forms; You don't have to let the ... Forge ahead with fee-avoidance strategies; More banks charging customers ... Overdraft fees can be too easy to rack up; Consumers need to ... How to cope as the credit card environment continues to change Law passed last year makes gift cards more consumer-friendly; Penalties ... BofA plans \$5 debit card fee; Monthly charge to hit if card ... Finding info on bank fees may take digging; Study reveals 'maze of ...
7	0.10	democrats republicans republican democratic obama israel campaign obamas bush presidential political palestinian senate clinton voters election mccain vote gop party	America's hand in Middle East peace Meeting revives Bush's vision of Palestinian state Direct talks provide best path toward Mideast peace Following Palestinian bid, more talk about talks; Quartet: No preconditions ... Israel warns against U.N. bid by Palestinians Rice: U.S. 'absolutely committed' to Mideast Rice on mission to Mideast in latest U.S. bid for peace Netanyahu talks of compromises; But Palestinians dismiss proposed peace ... Clinton pushes for sanctions against Iran 'that will bite'; She urges Israel ... Recent rift adds tension to AIPAC meeting; U.S. administration's tough ...
8	0.13	sec bonuses goldman lehman aig executives regulators assets compensation institutions paulson merrill firms shareholders billion securities bailout commission derivatives stearns	Judge shoots down BofA settlement; Says investors shouldn't be punished if ... BofA to repay its TARP funds; Returning \$45B bailout would end extra ... Bank of America charged with fraud; Cuomo also targets former CEO and ... Bank of America fined \$33M; SEC says bank lied about Merrill bonuses Bank of America ousts Thain; He's the third top Merrill Lynch executive ... Banks can repay \$68B in assets for bailout; Obama: Return of taxpayers' ... 9 banks pay \$33 billion in bonuses; Compensation paid despite Wall Street ... Lewis: BofA pushed to Merrill deal; CEO says job threatened unless merger ... BofA shareholders oust Lewis from chairman role JPMorgan Chase to pay \$154M to settle fraud case
9	0.11	students food college workers airlines schools education school small community travel employees technology blood jobs local business businesses university marketing	Pantries tackle campus hunger; Students, others turn to food banks for help ... Recession finds fertile field of egg, sperm donors; Decent pay, but screening ... Good deeds fall victim to bad times; Many volunteers busy tending to ... Blood banks for dogs, cats see an increased demand; Facilities replace the ... Health care costs worry near-retirees; Workers underestimate expenses under ... Tech tools bring big success for small firms Food banks can't meet demand; In sour economy, pantries see many ... Harvesting help to Make A Difference; The apples this day keep hunger ... Food banks adapting for record demand; Donations are up -- but outstripped ... Haiti relief less than Katrina, 9/11; Recession may have been factor
10	0.08	borrowers homeowners lenders foreclosure mortgage	Foreclosure filings rise in spite of aid program's efforts; 'We're at the ... About 40% of borrowers leave mortgage aid program Help on home loans makes progress; Mortgage modifications still have long ...

		loan loans mortgages foreclosures lender servicers payments score freddie subprime modifications fannie refinance housing mac	4 programs trying to work problems out with homeowners Help arrives for 2nd mortgages; Homeowners getting first home loan ... BofA will cut loan balances; Goal is to reduce home foreclosures Few get complete loan modifications; Only 12% have been converted after ... Help is out there for homeowners; Many get modified home loans without ... Modification of troubled mortgages on slow pace; Since March, program has ... Trials drag on for loan help; Many 3-month plans are hitting 6 ...
11	0.05	championship tour golf woods game fedexcup players win games won team playoffs playoff pga tournament sports saturday race wins season	Weekend TV Rivals don't mind Woods' absence Roundup Stricker looks to make another FedExCup run Format change might produce drama swings FedExCup redux: Kuchar, Hoffman, Johnson stake claims USA TODAY Super 25 rankings; High school girls basketball USA TODAY Super 25 rankings; High school football Simpson rallies for victory; Second win in three weeks puts Simpson at ... Others to watch

Table A8: USA Today (filtered): Burscher et al. k-means

Frame	Prop	Key words	Exemplar headlines
1	0.06	card fee credit debit overdraft bank consumer customer rule account issuer new practice retailer charge rate transaction law limit use	Walmart to offer prepaid card; The reloadable Bluebird to come with few ... Banks reduce credit card rewards; Financial pinch trims freebies About the Social Security debit card Tips for gift card givers and getters SunTrust, Regions cancel debit card fees Over-limit fees ending for 2 cards; Discover, AmEx are doing away with ... Schools rely on college-linked card revenue; 'It is a big chunk of ... Credit card protections come with some holes; Rates and fees wouldn't be ... New rules mean new fees Will credit card deal affect fees?
2	0.11	stock investor market street wall company share big york financial new dow europe dividend loss year bank rally earnings bull	Battered blue chips prove none are risk-free; 'It's brutal out there,' as ... The opening bell may have a nervous ring to it; Last week's ... Buffett chips in \$3B to bolster GE Street begins to hear 'buy' rumblings; Pummeled stocks could be called cheap Berkshire believes in 5 stocks Chain reaction stuns some investors; Fannie, Freddie downgraded, financial stocks pummeled STORY STOCKS Stocks avoiding seasonal swoon? Merrill Lynch case brought Eliot Spitzer to prominence Kodak plans to snap up \$1B of its own stock; Such big ...
3	0.04	fund money bailout test market stress mutual bank investor cash bond plan government subprime return confidence yield	Don't undermine money funds Some new funds that hold promise; Experienced managers make all the ... Investors are evacuating emerging markets funds What do you do when your cute, cuddly money fund sprouts fangs? Money market funds at risk?; Bank of America plans a \$600 million ... Panel sets voluntary guides for hedge funds Money markets not unscathed; Historically stable fund industry a bit shaky

Frame	Prop	Key words	Exemplar headlines
		financial lending fee	Federal Reserve to shore up money market funds Hedge fund group wants looser short-selling rules Senate votes to release rest of bailout funds; Obama acknowledges that ...
4	0.02	deal merger cash offer mortgage student school college cost good bank settlement taxpayer company bailout site bad foreclosure biggest show	Platform-sharing deal leads GM to agree to sell its Opel division A raw deal for consumers Wall Street sees three debt-limit scenarios; Markets are banking on a ... More companies say, 'Let's make a deal'; Mergers and acquisitions make a ... Cash for clunker truck deals are catching on; Clean-air programs help ... Small businesses make Square deals; Diminutive credit card-processing ... Assistants' deals contain novel clauses; Bonuses tied to grades, more Oracle is triumphant in pursuit of BEA Systems; \$8.5B deal comes after ... Paulson endures tough day on Hill; He admits he pressured to get ... Student lenders scramble to save their sweet deal
5	0.05	rate interest low cut mortgage credit loan market inflation economist central higher economy bank buyer washington savings increase bond consumer	What to ask before you buy an annuity; Fixed rate or variable? ... Fed holds rates steady, but what next?; Statement's new language open for ... Economy's pickup could mean Fed holds rates steady 30-year rates fall to 4.71%, elicit mortgage applications Will Fed cut rates? Analysts divided No rate change expected from Fed; Economy pulling in separate directions Wall Street hopes for another Fed rate cut Investors sweat out Fed rate increases; But their trepidation may be premature When interest rates hit 32%, there ought to be a law Fed bumps up discount rate for banks; Move takes away a crutch, ...
6	0.09	home foreclosure mortgage sale homeowner loan housing buyer price borrower modification market lender year program house payment plan value property	Charities get more donated homes Mortgage refinancing takes off as HARP kicks in Military foreclosure rate up 32% over 2008; 20,000 servicemembers ... Home sales slide slightly in May Foreclosures not halted nationwide; But more companies implement, expand ... Crisis at Fannie, Freddie bares fundamental flaws Fed chief: Home slump shouldn't hurt economy; Strengths outweigh ... Help arrives for 2nd mortgages; Homeowners getting first home loan ... New plan to save homes; Could revamp loans for 3 million Home prices fall 12% from '08
7	0.60	bank financial new u.s. business economy company year crisis money government loan economic plan consumer bailout job washington ceo time	Check It Out; Goldman transformation How the senators voted Check it out TECHNOLOGY LIVE Airport Check-in Commissioner's departure could leave all-GOP SEC Watch, Listen & Read Norton's Crowdrise goes the distance; Micro-giving at the core of actor's charity Recap of the 2009 FedEx playoffs Airport check-in After Sandy, two Manhattans
8	0.03	tax swiss client offshore u.s. account evasion bank break american wealthy secret case probe charge federal americans name transaction datum	Yacht broker pleads guilty to hiding \$3.4M from IRS Soak the rich? Obama should start with tax evaders IRS pays UBS whistle-blower \$104 million Tax records: Do you save them or shred them?; Some paperwork you ... U.S. seeks tax evaders at second bank Use smartphone app to file taxes; TurboTax's SnapTax works for those who ... IRS asks to seize \$10M in allegedly hidden assets; Nevada firm suspected ...

Frame	Prop	Key words	Exemplar headlines
			Eager for your tax rebate? You can make sure you get it ... Feds pounce on tax evaders; Prosecutions up 25% since 2001 Where's my tax rebate? Here's why you might not have gotten it ...

Table A9: USA Today (filtered): EFA

Frame	Prop	Key words	Exemplar headlines
1	0.39	say year even mani time make also get one now like back take can still much less financi month recent	Corporate owners hide assets, identities; A mini-industry promotes some U.S. ... Reforms? Not for rates on private student loans; Critics' proposals wouldn't ... How students borrow for college could soon change; Cuomo's inquiry into ... 3 lenders up student loan discounts; They call it competition; critics point ... Sallie Mae changes its student loan ways; Deal aims to make process ... Food banks can't meet demand; In sour economy, pantries see many first-time ... Hope builds for stock recovery in new year; But strategists warn that ... Boom in Australia goes bust in record time as global slowdown hits 5; questions 5 questions for Eugene Ludwig
2	0.15	fee stock card custom charg investor account use check dow crisi market index jone ralli invest strategist credit sampp pay	Debit card fees are coming: How to avoid them; Banks ding shoppers ... Finding info on bank fees may take digging; Study reveals 'maze of ... Chase, Wells Fargo drop plans to charge customers a debit card fee; ... Bank fees a pain? Switching is, too; Some things to consider before ... Small banks capitalize on rivals' fees; They're using consumer anger to gain ... Online banks heat up competition with savings rates of 5% or more Banking on success as a one-stop shop; Kovacevich's Wells Fargo wins with ... Forge ahead with fee-avoidance strategies; More banks charging customers ... It's doubtful banks will reconsider debit card fees BoFA plans \$5 debit card fee; Monthly charge to hit if card ...
3	0.09	said stock foreclosur index homeown mortgag committe amp fail home regul averag secretari gain card taxpay point propos higher administr	Congressional hearings raise foreclosure concerns; Witness: 10% of ... Legal gaps help hide assets from creditors, IRS; Court records track alleged ... Foreclosure proceedings downshift; More homeowners hire lawyers, while Another chance for home loan help; Some who lost homes can seek ... How you pay for stuff could soon determine how much you pay Iran's economic conditions deteriorate; Country grows more isolated in its ... Bull is aging, but it's still going; 'This is typical': The higher ... Feds charge 13 with inside trading; Officials say ring netted \$8 million ... 'King of music business' loses his throne; Search is on for mogul ... Pay the penalty
4	0.07	home mortgag homeown foreclosur fed reserv wall street earn lender asset banker washington regul largest afford case file earlier loss	Few respond to foreclosure review offer; Borrowers' cases are checked; errors ... Investment banks benefit most from M& A mania; Buying their stocks is ... Lenders want to help Banking on returns? Look at financial services sector Talks may tiptoe around jobs bank; Automakers choose to focus on slashing ... Cashing in on immigrants; Bank of America's campaign to offer credit cards ... What's in store for money funds?; They're still relatively safe places to ... Fed official says subprime mess hasn't hurt economy Corrections & Clarifications Proposed money market mutual fund changes draw fire; Industry opposes ...
5	0.06	million foreclosur	To boost your stock performance, pit analysts vs. short sellers; A combination ...

Frame	Prop	Key words	Exemplar headlines
		billion action home homeown money offic mortgag bank fargo estim support quarter target washington control plan congress file	Over 14,700 take IRS up on leniency; Offshore tax deal means billions ... Savers' Scoreboard Predatory mortgages labeled 'crisis'; Loans called 'greatest threat' ever to ... Wells Fargo settles lending bias case; Bank denies putting creditworthy ... Wolfowitz apologizes for securing raises for girlfriend at World Bank Hackers target ID, credit, e-mail data World Bank panel leaves Wolfowitz's fate in limbo; 'We have to ensure' ... Barclays makes \$91B offer for ABN Amro; Biggest banking deal ever could ... Look who wants protection fromforeign rivals now
6	0.05	commiss billion firm appear ceo total consid money took homeown thing exchang account drop fail mortgag case million largest american	Overdrafts rise, and banks are making more on fees; Banks say they're ... Look who wants protection fromforeign rivals now Source: Clients' money tapped; Troubles mount at MF Global MF Global showed appetite for big risks; Ratings agencies and others raised ... What might have happened in \$2.3B UBS case?; By analyzing basic facts, ... Wachovia-A.G. Edwards deal may inspire more mergers Con artists take advantage of check rules Corzine's MF Global declares bankruptcy; Reports: Regulators probe if ... Rogue trader at UBS; Swiss bank hit with \$2B loss Wal-Mart regains 'Fortune' top spot; Profits for 500 up an 'awesome' 29% ...
7	0.04	countri case fund deposit presid product agre american world stop million addit money transact reduc reform buy possibl tax power	If issuing bank fails, brokered CDs may be inaccessible for a while 150 targeted in Swiss bank case; UBS set to identify more clients ... U.S., UBS reach agreement over tax evasion U.S. officials indict 2 Swiss citizens in UBS case; Men accused of ... List of 'problem banks' grows to 117; FDIC says most of the ... UBS must release data on 4,500 clients; IRS expects up to 10,000 ... IRS may cast wider net in tax probe Supreme Court rules states can't regulate subsidiaries of national banks Don't blame the banks; OPPOSING VIEW Bush plan would raise grants to low-income students; Lender subsidies would ...

Table A10: USA Today (filtered): STM

Frame	Prop	Key words	Exemplar headlines
1	0.10	card debit cards fees overdraft fee consumers customers checking credit issuers unions charge consumer union account retailers atm transactions deposit	Debit card fees are coming: How to avoid them; Banks ding shoppers ... Overdraft protection comes in many forms; You don't have to let the ... How to cope as the credit card environment continues to change Forge ahead with fee-avoidance strategies; More banks charging customers ... Weigh before you pay: Debit or credit? Debit card fees are like weeds: Chop one and 10 new ones ... Banks changing how they handle overdrafts; New rule requires them to get ... FDIC study criticizes banks' overdraft-fee policies Get ready to get spanked for staying out of credit card debt Bank fees get more aggressive; Today is consumers' last chance to comment ...
2	0.11	fund funds shares yields bond stock dividends portfolio bonds cds investors yield dividend mutual investments invest securities	Why do you own a money market fund?; They're good for reducing ... Income investors might try bond ETFs; Few low-risk investment choices offer ... Need income off your savings? Oy vey.; Today it's nearly impossible, but ... Remember risk when chasing dividends; Q: If you had \$1,000 to invest ... Before you buy a stock, consider your sell point Low interest rate dilemma; Q: With the Federal Reserve holding interest rates ... Low yields deflate money funds; \$1.1 trillion has bled out, pouring into ... Scared to wade back into stocks?; 3 baby steps to get you ...

		stocks investment gold	What's a saver to do these days?; Generating income with low risk ... What do you do when your cute, cuddly money fund sprouts fangs?
3	0.11	foreclosure homeowners mortgage lenders foreclosures homes buyers borrowers freddie loan loans mortgages fannie housing home servicers modifications mac lender mae	Number of short sales on the rise; Homes can retain value better ... About 40% of borrowers leave mortgage aid program Few get complete loan modifications; Only 12% have been converted after ... Pace of short sales increases Foreclosed homes sell at big discounts; Prices more than 35% below regular ... Help on home loans makes progress; Mortgage modifications still have long ... Foreclosures take a big jump; Filings in March 46% higher than last ... Lenders embrace more short sales Sales of foreclosure properties on the rise; Worsening problem's a major drag ... Refinancing soars on falling rates; Lenders struggle to meet demand
4	0.14	care education students workers college parents children businesses schools technology family women school work small business student health jobs age	Business roots; Entrepreneurial urges often take form when someone else on ... Entrepreneurs hire close to home: Their moms; Family link can be both ... Tech tools bring big success for small firms Hey, digital media, Take a look At Current TV; Cable channel could ... Color app gives an instant peek at photos Health care costs worry near-retirees; Workers underestimate expenses under ... Leaving corporate life behind; Entrepreneurs take a chance to be own boss Faith at the office. Why not?; You don't have to leave your ... Author: Promotions, raises and benefits make bad motivators; They inspire ... Green-conscious students can save energy and cash; Energy-efficient ...
5	0.08	fla irs court ubs filed fraud criminal justice swiss prosecutors security offshore legal attorney enforcement file scams investigation documents fbi	Spring training schedules 150 targeted in Swiss bank case; UBS set to identify more clients ... 3rd American pleads guilty to tax charges linked to UBS; U.S. seeks ... Man pleads guilty in tax case; Homann says he hid \$5M from ... U.S., UBS reach agreement over tax evasion IRS asks to seize \$10M in allegedly hidden assets; Nevada firm suspected ... UBS wants to keep working toward a settlement; U.S., Swiss bank ask ... U.S. seeks tax evaders at second bank HSBC alerts U.S. clients amid tax evasion probe; Disclosure says bank ... IRS may cast wider net in tax probe
6	0.19	obama republicans committee democrats republican bailout congress senate bush sen bonuses aig lawmakers geithner political obamas compensation secretary presidential goldman	Obama to renominate Bernanke to lead Fed Senate vote on financial reform blocked; GOP says bill goes too far, ... Obama tells Wall Street to join in; President calls for support of ... Financial reform bill clears Senate; 'We're finally restoring law and order to ... Bernanke denies pressuring Bank of America; Lawmakers want to know if ... McCain, Obama heading to White House summit; On TV, Bush warns 'entire ... Geithner: Financial reform due; Says recession proves need, provides impetus Wall Street rumbling felt in elections; Congressional candidates duel on ... Paulson endures tough day on Hill; He admits he pressured to get ... Deal or no deal?; House GOP defies Bush on bailout hours after ...
7	0.17	growth european central fed dow economy global inflation recession cut strategist markets europe china cuts economists earnings	Earnings gallop in with flags waving; Growth isn't just from cost cutting ... Consumer prices nudge up in July; Inflation was a mere 0.1%, giving ... Rate fears rein in markets' charge; Once-zooming Dow falls more than 200 ... Economy's pickup could mean Fed holds rates steady 4th quarter usually hot for stocks; Bulls expect same this year, despite ... Investors look for clues in earnings; Is economy as dour as it ... Fed holds rates steady, but what next?; Statement's new language open for ... Fed again leaves key interest rate unchanged; Still cites inflation as top ...

		<p>feds chinese rally</p>	<p>Double-dip fears fade, stocks rally; 'Near-miss' recession could lead to bigger ... Wall Street presses for rate cut; Consumer confidence, Fed minutes add to ...</p>
8	0.10	<p>race tour team win game season won town championship players games river feet play black rock series city miles park</p>	<p>Down to the final 30 Format change might produce drama swings NCAA lacrosse championships Phillies not accustomed to trailing; Yankees pitchers make life tough on Philly ... Simpson rallies for victory; Second win in three weeks puts Simpson at ... Intermediate tracks have big say in the Chase; Five playoff races run ... Freshman class might rate with greats A crash course on racetracks in Chase Coyotes beat Kings, avoid West sweep Senators confident they're not out of Finals</p>