Temporal Multi-layer Network Construction from Major News Events

Borut Sluban, Miha Grčar, and Igor Mozetič

Jožef Stefan Institute,
Jamova 39, 1000 Ljubljana, Slovenia
borut.sluban@ijs.si

Abstract. Good news should answer the following questions: *'Who?'*, *'Where?'*, *'When?'*, *'What?'*, and possibly *'Why?'*. We present an approach which extracts interesting events from thousands of daily news. We construct a time-varying, three-layer network where the nodes are entities of interest in the news. The temporal aspect of the network answers the *'When?'* question. The layers are: 1) the co-occurrence of entities which answers the *'Who?'* or *'Where?'*, 2) the summary layer which answers the *'What?'*, and 3) the sentiment layer which labels the links as 'good' or 'bad' news. We demonstrate the news network evolution over a period of four years in an interactive web portal.

Keywords: multi-layer networks, temporal networks, sentiment, summarization

1 Introduction

News inform people about interesting events around the world. We monitor a large number of news web sites around the globe and analyze the structure and the contents of the news. The paper addresses the following question: how to characterize and extract the 'unusual', highly publicized events? We apply a set of network analysis, text mining, sentiment analysis and visualization methods to extract and highlight major news.

The theory of complex networks characterizes systems in the form of entities (nodes) connected by some interactions (links) [1, 3, 11]. A special case of networks extracted from the data are co-occurrence networks, used in diverse fields, such as linguistics [7], bioinformatics [5], ecology [10], scientometry [21], and socio-technological networks [4]. Co-occurrence networks are defined as networks in which nodes represent some entities (for example persons, companies, countries, etc.), and links represent an observation that these entities exist together in some data collection (for example database, news article, etc.). For textual sources, it is important to extract the links between the entities that represent real relations, and are not created by chance.

In our previous work, we have developed a method to estimate the significance of co-occurrences, and a benchmark model against which their robustness is evaluated [14]. The method was applied to analyze the contents of financial news in comparison to empirical networks, constructed from other data sources, like geographical proximity, trade volumes, and correlations between financial indicators [19].

The above co-occurrence detection method models well the persistent, 'everyday' contents of news. However, one is often interested in unusual events, reported by multiple media sources in high news volume. In this paper, we report on a method which detects days with news peaks, and construct a time-varying multi-layer network. In particular, at the daily time resolution, we construct a three-layer network with the co-occurrence layer, the summary layer, and the sentiment layer. The co-occurrence layer consists of all the links at peak days, when the news volume is significantly higher than in the past. The summary layer consists of top news for the peak days, where the top news are summarized by the most distinguished titles. The sentiment layer might have a longer time span. We aggregate the sentiment of the top news over all the peak days within a time period. Finally, we are concerned with the presentation of such a temporal multi-layer network. The network evolution over time, with drill-down inspection of details, is demonstrated in a public, interactive web portal at `http://newsstream.ijs.si/occurrence/major-news-events-map`. The portal facilitates access to over 35 million news, predominantly financial, collected from 170 English news sites, over a period of the last four years.

The paper is organized as follows. In Section 2 we describe the entity recognition in news, detection of days with news peaks, and identification of distinguishing topics which summarize major events at peak days. We also describe a lexicon-based approach to sentiment analysis in the news, and the network construction method. In Section 3 we give details about the financial news collected, and illustrate the detection of significant events. Some interesting topics recently reported in the news are highlighted, together with the estimated sentiment. We compare the sentiment distribution of all the news, peak news, and top news, and show that there are small, but statistically significant differences. Finally, we show the network visualization implemented in our web portal. We conclude in Section 4 with ideas for future work.

2 Methods

We describe a multi-stage approach to construct a multi-layer network of major news events. The stages consist of entity recognition (which identify nodes of the network), event detection (which identify links between the nodes and the co-occurrence layer), content identification (the summary layer), and sentiment analysis (the sentiment layer).

2.1 Entity recognition

News are about events related to politicians, countries, companies, etc., which we call entities. The process of identifying entities in textual documents requires three components: an ontology of entities and terms, gazetteers of the possible appearances of the entities in the text, and a semantic annotation procedure that finds and labels the entities. We describe the entity recognition approach, as implemented in our NewsStream portal (`http://newsstream.ijs.si`) [12].

The ontology we use for information extraction consists of three main categories: financial entities, financial terms, and geographical entities. Most of the ontology is automatically constructed from various data sources. The geographical entities (continents, countries, cities, organizations) were extracted from GeoNames (`http://www.geonames.org`). MSN Money (`http://money.msn.com`) was used to organize stock indices and link them to the companies that issue these stocks. The hierarchy of financial terms related to the financial crisis was developed in collaboration with financial experts.

Each entity in the ontology has associated a gazetteer, which is a set of rules that specify the lexicographic information about possible appearances of the entity in text. For example, 'The United States of America' can appear in text as 'USA', 'US', 'the United States', etc. The rules include capitalization, lemmatization, POS tag constraints, must-contain constraints (i.e., another gazetteer must be detected in the document or in the sentence) and followed-by constraints.

Finally, a semantic annotation procedure recognizes the entities of interest. It traverses each document and searches for entities from the ontology. The gazetteers of the entities in the ontology provide information required for the disambiguation of different appearances of the observed entities.

2.2 Event detection—peak days

The next step of content analysis of news is detection of relevant events in the news. We use the daily volume of news articles as a proxy for identifying exceptional events in the news. Given a set of entities of interest $E = \{e_1, \ldots, e_l\}$, we identify all events related to all pairs of entities (e_i, e_j). We monitor the volume of news about these pairs and construct a network of exceptional events between the observed entities.

In [14] we proposed to establish a co-occurrence link between a pair of entities (e_i, e_j) when the number of observed co-occurrences is significantly grater than expected by chance. The probability of a random co-occurrence was estimated from the observed individual occurrences. In this paper, we propose an alternative approach, where we compare the daily number of observed co-occurrences to a longer time period.

We construct a time series of co-occurrence volumes $\mathbf{v}_{ij} = \{v_{ij}(t)\}$ for a pair (e_i, e_j) and a time period T, $t \in T$. At a given time point t, we consider a window $W_h(t) = \{v_{ij}(t-h-1), \ldots, v_{ij}(t-1)\}$ of length h as a historical baseline, from which we calculate the expected volume at the time point t.

We assume, for a pair of entities, that the volume of their co-occurrences fluctuates around the average volume for a given time period, and that the fluctuations have Gaussian distribution around the average. As the value of the average changes through time, we use a sliding window W_h to compute a moving average.

Given the co-occurrence volume time series $\mathbf{v}_{ij}$, and the size h of the sliding window, we calculate the mean co-occurrence volume $\bar{v}_{ij}(t)$ in $W_h(t)$ and its standard sample deviation $\sigma_{ij}(t)$. Let $z_{ij}(t)$ denote the multiple of $\sigma_{ij}(t)$-deviations from the mean $\bar{v}_{ij}(t)$:

$$z_{ij}(t) = \frac{v_{ij}(t) - \bar{v}_{ij}(t)}{\sigma_{ij}(t)} .$$

For a given Z_0, we say that the co-occurrence volume, such that $z_{ij}(t) > Z_0$, is unexpected and represents an exceptional event between the entities e_i and e_j at day t. Such day t is named a *peak* day. The co-occurrence links between the entities at peak days constitute the *co-occurrence layer* of the constructed network.

2.3 Identification of relevant topics—top news

The goal of the next stage in network construction is to attribute a shallow semantics to the links. The semantics is actually a summary of the top news at peak days, in the form of the most relevant titles.

First we select all the news related to a particular link on a particular day. For example, to attribute semantics to the link between the U.S. and China on a particular day, we consider only the news that contain both these two entities and were published on that day. All the titles of these news are merged into a single text document. One such merged document is created for each day in the past two months (excluding weekends). We apply the standard text preprocessing approach to compute the bag-of-words (BOW) vectors of these documents [9]. In this process, we employ tokenization, stop word removal, stemming, and the TF-IDF weighting scheme [17], standard in text mining. The TF (term frequency) weight, $TF_{d,k}$, denotes the number of times the word k occurs in the document d. The TF-IDF weight is a combination of the TF weight and the IDF weight, where IDF stands for inverse document frequency. IDF of the word k is computed as follows:

$$IDF_k = log \frac{|D|}{n_k} ,$$

where n_k is the number of documents in the collection D that contain the word k. The TF-IDF weight is then:

$$TFIDF_{d,k} = TF_{d,k} \times IDF_k .$$

The TF-IDF scheme weights a word higher if it occurs often in the same document (the TF component), and if it occurs in only a few documents from the corpus (the IDF component).

The BOW vector for the current day contains information about how important a certain word is with respect to the most relevant events on that day. Instead of showing the top-ranked words, we propagate the weights to the news titles and thus rank the titles by their relevance. The weight-propagation formula computes the average of the word-weights in a title c. The weight of the title, w_c, is computed as follows:

$$w_c = \frac{1}{|c|} \sum_{k \in c} TFIDF_{d^*,k} \ \ ,$$

where k denotes the words in the title c, and d^* represents the merged document for the day in question. Note that the weight w_c penalizes long titles since it is inversely proportional to the title length $|c|$. In our case, this is a desirable property because we would like to find short and to-the-point titles that best describe the most important events. The most distinguished titles at peak days represent the *summary layer* of the constructed network. This layer enriches the co-occurrence links between a pair of entities, by summarizing the news published at the peak days.

2.4 Lexicon-based sentiment analysis

The final stage of the network construction attributes sentiment to the links. We construct the *sentiment layer* of the news network by detecting sentiment orientation and strength of news articles which mention pairs of entities. The sentiment attached to a link between two entities indicates whether the news were 'good' or 'bad' for a given day. However, in contrast to the co-occurrence and summary layers, which have daily time granularity, it is often convenient to aggregate the sentiment links over a longer time period, encapsulating all the top news at peak days.

A sentiment polarity is calculated by a lexicon-based approach. The sentiment polarity of a document is computed from the counts of predefined sentiment terms (positive and negative) in the document. The sentiment terms are from the Harvard-IV-4 sentiment dictionary [22]. For a document d, the sentiment polarity s is calculated by the following formula:

$$s_d = \frac{pos_d - neg_d}{pos_d + neg_d} \,,$$

where *pos* and *neg* are the numbers of positive and negative dictionary terms found in the document d, respectively. The sentiment polarities of a set of documents can then be aggregated. An aggregate sentiment for a pair of entities (e_i, e_j) is computed from the top news documents d at peak day t, and from several peak days in a period T:

$$s_{ij}(T) = \frac{1}{N} \sum_{t \in T} n(t) \times s_{ij}(t) \ \ , \ \ s_{ij}(t) = \frac{1}{n(t)} \sum_{d} s_d \ \ , \ \ \text{where: } (e_i, e_j) \in d \,,$$

where N is the total number of documents selected at peak days t in the time period T.

2.5 Network construction parameters

The temporal news network consists of nodes and three layers of links, at daily resolution. The nodes are entities of interest, $E = \{e_1, \ldots, e_l\}$. The links are pairs of entities (e_i, e_j), with different properties attached at each layer.

The *co-occurrence layer* links entities detected during unusual events, i.e., the peak days. A link (e_i, e_j) is created when the volume of documents containing e_i and e_j significantly exceeds the average volume observed in the previous h days. Assuming the volume of entity co-occurrences in news has Gaussian distribution around the average for a given time period, the significance threshold is set to $Z_0 = 3$, and $h = 44$ (the number of weekdays in the past two months). In a document, each entity must occur at least three times, and at least one entity must occur in the document title. These constraints eliminate most of the noise due to the entity occurrence in a boilerplate. The *summary layer* links the two entities by extracting the most relevant news contents at peak days, thus providing a shallow semantics of the links. A summary link consists of the titles of the top three news articles. The *sentiment layer* presents the emotional attitude of the top news, in terms of the balance between the positive and negative words used. A sentiment link value ranges between -1 and $+1$, where -1 denotes the 'bad' news, 0 the neutral or balanced news, and $+1$ the 'good' news.

3 Results and Discussion

We have been collecting articles from 170 English financial news and blog sites, from November 2011. On average, there are about 35,000 articles per day, a total of over 35 million articles collected until September 2015. This data holds information about temporal relations between different types of entities, such as people, companies, stocks, countries, etc. In this paper, we describe how to detect major events involving different countries, and the construction of the corresponding temporal network. The network captures the major news events detected over the previous four years, and reveals the semantics of the relations between the countries in terms of the contents and sentiment.

3.1 Significant events

The NewsStream portal provides an API to collect all the news about any pair of entities. We detect significant events by comparing the daily news volume to the volume of the past two months. If on a particular day the news volume exceeds the average volume of the past two months by more than three standard deviations, i.e., $\bar{v}_{ij} + 3 \cdot \sigma_{ij}$, this day is identified as a significant event day for the observed pair of countries (e_i, e_j).

Figure 1 shows the volume of the news involving 'China' and 'United Stated' from November 2011 to September 2015. The significant increases in the news volume are identified as volume peaks above the gray line.

In the period between January 2012 and September 2015, 17,702 significant events between 217 countries were detected. We analyze these events in terms of the most relevant content and the associated sentiment.

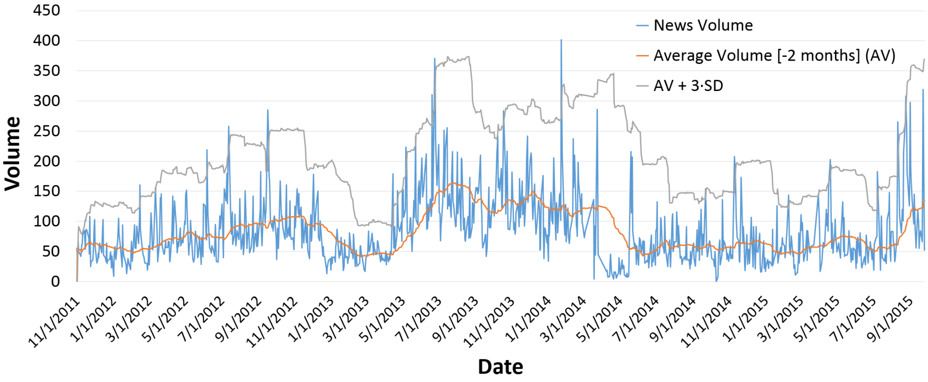


Fig. 1. Volume of news articles about 'China' and 'United States' (blue line). The orange line denotes the moving average volume over a two months window, and the gray line is three standard deviations above the average. Significant events occur at days peaking above the gray line.

3.2 Most relevant contents and its sentiment

We focus on the news related contents published at the peak days. We identify the most relevant and distinguishing topics for each significant event day, as described in Section 2.3. We compared our top news results to the major news timeline of *Europe Media Monitor* (EMM, `http://emm.newsexplorer.eu/NewsExplorer/timelineedition/en/timeline.html`), and reached an overlap of 45% with all EMM major news, and 60% overlap with major news topics mentioning at least two countries in the topic title. These differences are mostly due to the following reasons. The major news events of EMM are not limited to country relations (links), therefore they include also news events mentioning only one country or none at all. Topics persisting for several days with little development are avoided by our approach as we are looking for significant new events. Some countries that are involved in certain topics may be overlooked in the evaluation process due to unresolved indirect mentioning, like 'Merkel' or 'VW' instead of Germany.

For each news article at a peak day we also compute its sentiment, as described in Section 2.4. Some significant event days in August and September 2015 for three country pairs are in Table 1. Each event is characterized by the top news headlines and the associated sentiment.

Table 1 illustrates three examples of breaking news about a pair of countries. The first two examples are about the events concerning French-built warships, which were not delivered to Russia, but were later sold to Egypt. The third news example highlights the 'emissions scandal' of a German automobile producer VW, which broke out in the United States.

Table 1. Content and sentiment of the most relevant news on significant event days. Shown are significant links between France and Russia, France and Egypt, and between Germany and the United States, in August and September 2015.

Link (Sentiment)	**Day**	**News**	**Sentiment**
FR - RU (0.265)	Aug 6 2015	France to pay Russia under $1.31 billion over warships	0.286
		France to pay Russia under 1.2 billion euros over warships	0.256
		France says several nations interested in Mistral warships	0.254
FR - EG (-0.072)	Sep 23 2015	France sells 2 disputed warships to Egypt	-0.091
		France sells warships to Egypt after Russia deal scrapped	-0.020
		France to sell warships to Egypt after Russia deal scrapped	-0.103
DE - US (-0.015)	Sep 21 2015	VW rocked by US emissions scandal as stock slides 17 percent	0.039
		VW Rocked by U.S. Emissions Scandal as Stock Slides 17%	-0.036
		VW shares plunge on emissions scandal US widens probe	-0.026
	Sep 24 2015	Will Volkswagen scandal tarnish Made in Germany image?	0.007
		After year of stonewalling Volkswagen stunned U.S. regulators with confession	-0.042
		Insight - After year of stonewalling Volkswagen stunned U.S. regulators with ...	-0.030

3.3 Sentiment distribution

We examine the differences in the sentiment distribution over different sets of news articles. The goal is to compare the sentiment distribution of ‘everyday’ news with the sentiment at peak days, and with the top news at the peak days. Figure 2 shows the three sentiment distributions.

All three sentiment distributions are approximately Gaussian, and very similar. There is a minor positive sentiment bias in all news, while the peak news are slightly negative. The top news at peak days also seem relatively less positive than the all news. However, the top news contain proportionally more extremely positive and extremely negative news articles. The statistics are in Table 2.

Table 2. The sentiment distributions for different sets of news relating pairs of countries. $\bar{s}$ is sentiment mean, *SD* standard deviation, and *SEM* standard error of the mean.

	Documents	$\bar{s}$	*SD*	*SEM*
All news	1,567,396	0.028	0.239	0.0002
Peak news	274,806	-0.003	0.231	0.0004
Top news	48,097	0.010	0.249	0.0011

We test the null hypothesis that a pair of news populations (all, peak, or top) has equal mean sentiment. We apply Welch's t-test [23] which is robust for skewed distributions [8]. The results are in Table 3. With t values > 10, the degrees of freedom $\gg 100$, and the p-values $\ll 0.0001$, the null hypothesis can be rejected

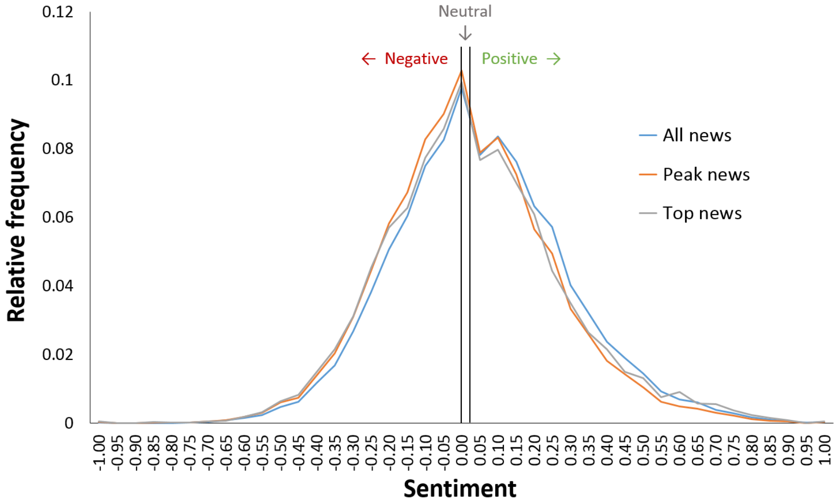


Fig. 2. Comparison of the sentiment distribution of all news articles, peak day articles, and top news, i.e., most relevant articles at peak days.

for all pairs of news populations. We conclude, with high confidence, that the three populations of news have significantly different sentiment means, though these differences are small. Of course, with such large samples one always detects differences. Nevertheless, the results are useful to hint at meaningful differences which are exploited by introducing the neutral zone.

Table 3. The results of t-tests for comparison of sentiment means. DF is the estimated degrees of freedom.

	t	DF
All news vs. Peak news	65.55	385,256
Top news vs. Peak news	10.67	63,425
All news vs. Top news	15.63	50,852

To distinguish 'bad' news from 'good' news, we introduce a neutral zone around the sentiment mean. The $\bar{s}$ value is the sample mean, and the population mean is in the interval $\bar{s} \pm 9\textsc{sem}$ with very high confidence. We take this interval band around $\bar{s}$ as the neutral zone. We classify the sentiment of the top news into three discrete classes: *negative* if $-1 \leq s < 0$, *neutral* if $0 \leq s \leq +0.02$, and *positive* if $+0.02 < s \leq +1$. The neutral zone is very narrow, as shown in Figure 2, and is used just to clearly distinguish between the negative and positive sentiment of top news. This classification is used to label the links in the network visualization with different colors.

3.4 Network visualization

A network visualization offers a unique way to understand and analyze complex systems by enabling the user to easily inspect and comprehend relations between individual units and their properties [16]. In addition to single layer network visualization [2], also multi-layer visualization is increasingly popular [6, 13].

We have implemented a spatio-temporal visualization of the country co-occurrence network, constructed from the detected major news events, their most relevant content, and the associated sentiment. The visualization facilitates the inspection of various aspects of the network: time dimension, news content, news sentiment, and geography. We have embedded the network into the world map and included functionality to explore the different aspects of the network. Figures 3 and 4 show two instances of the network in time and space. The visualization is an extension of the NewsStream portal, publicly accessible at `http://newsstream.ijs.si/occurrence/major-news-events-map`.

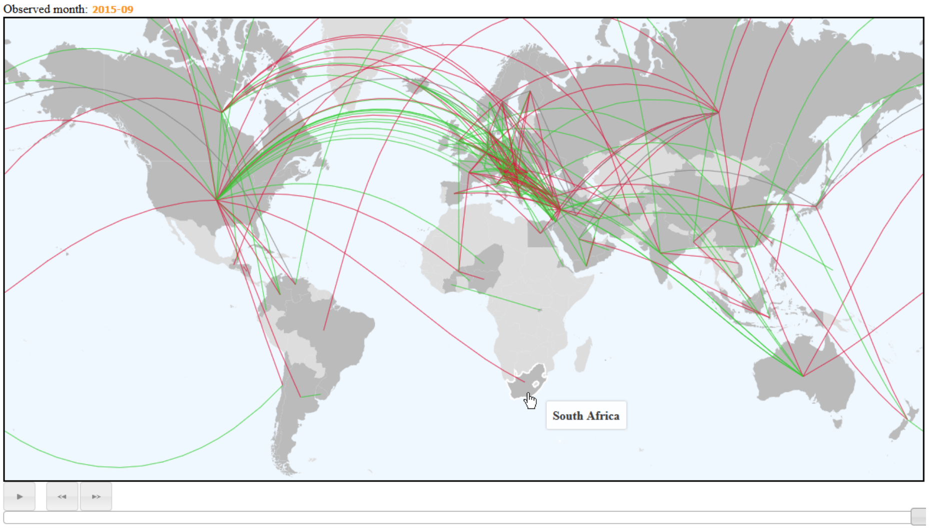
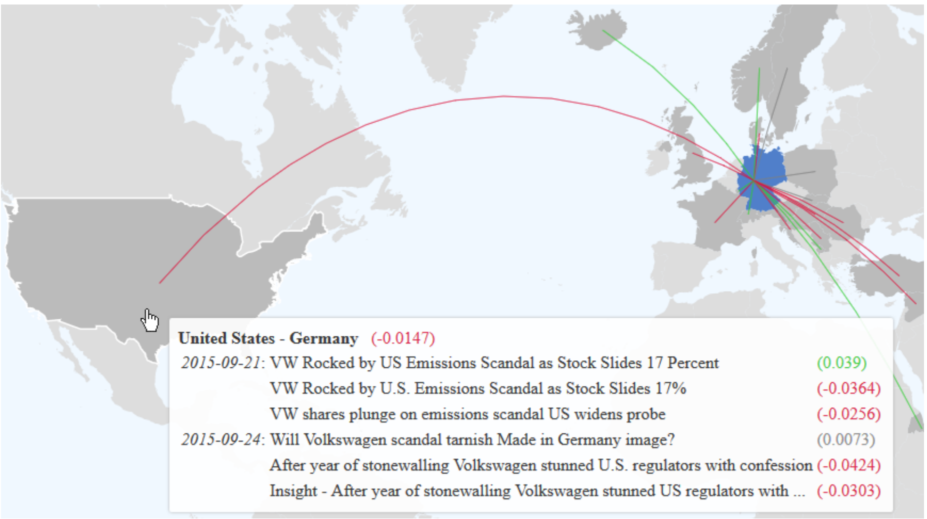


Fig. 3. Temporal country co-occurrence network of major news events during Sep 2015.



United States - Germany (-0.0147)

2015-09-21:	VW Rocked by US Emissions Scandal as Stock Slides 17 Percent	(0.039)
	VW Rocked by U.S. Emissions Scandal as Stock Slides 17%	(-0.0364)
	VW shares plunge on emissions scandal US widens probe	(-0.0256)
2015-09-24:	Will Volkswagen scandal tarnish Made in Germany image?	(0.0073)
	After year of stonewalling Volkswagen stunned U.S. regulators with confession	(-0.0424)
	Insight - After year of stonewalling Volkswagen stunned US regulators with ...	(-0.0303)

Fig. 4. The most significant news about Germany and the United Stated in Sep 2015.

4 Conclusions

We describe a methodology for the construction of temporal multi-layer news networks. The network captures the links between entities during major news events, summarizes the relations between them, and assigns the sentiment to them. In an experimental setup, we have constructed a time-varying network of countries mentioned in the news over the past four years. We have detected 17,702 major news events involving 217 countries, in the period from January 1, 2012 until September 30, 2015. The interactive visualization of the network supports the spatio-temporal exploration of the major news events.

One of the weaknesses of this approach is a simple, lexicon-based sentiment analysis. We have already implemented much more sophisticated sentiment classification approaches, based on the SVM models, and applicable to short texts in different languages and in various domains [18, 24, 15, 20]. In the future, we plan to focus on key sentences around the entities of interest, and combine the model-based and lexicon-based approaches to sentiment classification for evaluating longer news articles.

Another direction of future research is to study the role of news in the policy making process. In the context of policy debates about complex global issues, such as climate change, financial crises, sustainable development, or migrations, there is an antagonism between the public and private interests. The news play an important role in shaping the policy debates but they might be influenced by the ownership structure of media companies and industrial corporations. We plan to analyze the news and create multi-layer networks of corporations and legal issues, or corporations and environmental issues, and study their sentiment leaning towards different issues. At the same time we will take into consideration the (in)direct ownership structure of the media companies, and analyze how this influences the reported news.

Acknowledgments. This work was supported in part by the European Commission FP7 projects MULTIPLEX (no. 317532) and SIMPOL (no. 610704), the H2020 FET project DOLFINS (no. 640772), and by the Slovenian ARRS programme Knowledge Technologies (no. P2-103).

References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Reviews of modern physics 74(1), 47 (2002)
2. Bastian, M., Heymann, S., Jacomy, M.: Gephi: An open source software for exploring and manipulating networks (2009)
3. Caldarelli, G.: Scale-free networks: complex webs in nature and technology. Oxford University Press (2007)
4. Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V.D., Loreto, V., Hotho, A., Grahl, M., Stumme, G.: Network properties of folksonomies. AI Communications 20(4), 245–262 (2007)

5. Cohen, A.M., Hersh, W.R., Dubay, C., Spackman, K.: Using co-occurrence network structure to extract synonymous gene and protein names from medline abstracts. BMC bioinformatics 6(1), 103 (2005)
6. De Domenico, M., Porter, M.A., Arenas, A.: MuxViz: a tool for multilayer analysis and visualization of networks. Journal of Complex Networks (2014)
7. Edmonds, P.: Choosing the word most typical in context using a lexical co-occurrence network. In: Proc. 35th Annual meeting of ACL. pp. 507–509. Association for Computational Linguistics (1997)
8. Fagerland, M.W.: t-tests, non-parametric tests, and large studiesa paradox of statistical practice? BMC Medical Research Methodology 12(78) (2012)
9. Feldman, R., Sanger, J.: Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, New York, NY, USA (2006)
10. Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., Ruppin, E.: The large-scale organization of the bacterial network of ecological co-occurrence interactions. Nucleic acids research 38(12), 3857–3868 (2010)
11. Jackson, M.O.: Social and economic networks. Princeton University Press (2010)
12. Kralj Novak, P., Grčar, M., Sluban, B., Mozetič, I.: Analysis of financial news with NewsStream. Tech. Rep. IJS-DP-11965, arXiv:1508.00027 (2015)
13. Piškorec, M., Sluban, B., Šmuc, T.: MultiNets: Web-Based Multilayer Network Visualization. In: Proc. European Conf. on ML and KDD. LNCS, vol. 9286, pp. 298–302. Springer (2015)
14. Popović, M., Štefančić, H., Sluban, B., Kralj Novak, P., Grčar, M., Puliga, M., Mozetič, I., Zlatić, V.: Extraction of temporal networks from term co-occurrences in online textual sources. PLoS ONE 9(12), e99515 (2014)
15. Ranco, G., Aleksovski, A., Caldarelli, G., Grčar, M., Mozetič, I.: The effects of Twitter sentiment on stock price returns. PLoS ONE 10(9), e138441 (2015)
16. Rossi, L., Magnani, M.: Towards effective visual analytics on multiplex and multilayer networks. Chaos, Solitons & Fractals 72(0), 68–76 (2015)
17. Salton, G.: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1989)
18. Sluban, B., Smailović, J., Battiston, S., Mozetič, I.: Sentiment leaning of influential communities in social networks. Computational Social Networks 2(9), 1–21 (2015)
19. Sluban, B., Smailović, J., Mozetič, I.: Understanding financial news with multilayer network analysis. In: Proc. European Conf. on Complex Systems, ECCS-14. Springer (2015)
20. Smailović, J., Kranjc, J., Grčar, M., Žnidaršič, M., Mozetič, I.: Monitoring the Twitter sentiment during the Bulgarian elections. In: Proc. IEEE Intl. Conf. on Data Science and Advanced Analytics. IEEE (2015)
21. Su, H.N., Lee, P.C.: Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in technology foresight. Scientometrics 85(1), 65–79 (2010)
22. Tetlock, P.C., Saar-Tsechansky, M., Macskassy, S.: More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance 63(3), 1437–1467 (2008)
23. Welch, B.L.: The generalization of "Student's" problem when several different population variances are involved. Biometrika 34(1-2), 28–35 (1947)
24. Zollo, F., Kralj Novak, P., Del Vicario, M., Bessi, A., Mozetič, I., Scala, A., Caldarelli, G., Quattrociocchi, W.: Emotional dynamics in the age of misinformation. PLoS ONE 10(9), e138740 (2015)