

Understanding Financial News with Multi-layer Network Analysis

Borut Sluban, Jasmina Smailović, and Igor Mozetič

Jožef Stefan Institute, Ljubljana, Slovenia

Abstract. What is in the news? We address this question by constructing and comparing multi-layer networks from different sources. The layers consist of the same nodes (hence multiplex networks), but links are constructed from textual news on one hand, and empirical data on the other hand. Nodes represent entities of interest, recognized in the news. From the news, links are extracted from significant co-occurrences of entities, and from strong positive and negative sentiment associated with the co-occurrences. In a case study, the observed entities are 50 countries, extracted from more than 1.3 million financial news acquired over a period of two years. The empirical network layers are constructed from the geographical proximity, the trade connections, and from correlations between financial indicators of the same countries. Different network comparison metrics are used to explore the similarity between the news and the empirical networks. We examine the overlap of the most important links in the constructed networks, and compare their structural similarity by node centrality and main k -cores. The comparative analysis reveals that the co-occurrences of countries in the news most closely match their geographical proximity, while positive sentiment links most closely match the trade connections between the countries. Correlations between financial indicators have the lowest similarity to financial news.

Keywords: multiplex, multi-layer networks, temporal networks, sentiment analysis, network comparison

1 Introduction

Methods developed in the fields of mathematics, computer science and statistical physics have contributed to the emergence of the theory of complex networks. The theory mathematically characterizes systems in the form of entities (nodes) connected by some interactions (links) [2]. The analyses of complex networks strongly influenced and advanced research in social media, biology, and economics [4,10]. A special type of networks extracted from the data are co-occurrence networks, used in diverse fields, such as linguistics [7], bioinformatics [6,22,18], ecology [8], scientometry [19,12], and socio-technological networks [5,23,9]. Co-occurrence networks are loosely defined as networks in which nodes represent some entities (for example persons, companies, genes, etc.), and links represent the fact that these entities exist together in some collection (for example database, article,

etc.). For textual sources, it is of high importance to extract the links between the entities that represent a real relationship and are not created by chance [15].

In this paper we investigate the relation between the networks extracted from online texts and the networks drawn from empirical data. Both networks, the news network and the empirical network, are multi-layer, but with the same nodes, hence multiplex networks. In our case study, we analyze two years of financial news from 170 major English-language web sites. The nodes in all network layers are 50 countries recognized in the news. The news network has a co-occurrence layer and the sentiment (positive or negative) layer (associated with the co-occurrence links between the countries). The empirical network has three layers: a geo layer (corresponds to geographical proximity between the countries), a trade layer (volume of trade between the countries), and a financial layer (correlations between the Credit Default Swaps of the countries). See Figure 1. Most of the layers, except geo and trade, vary in time. For the period of two years, we constructed monthly snapshots of the time-varying layers.

The goal of this research is to shed some light on the contents of online news. The research question we attempt to answer is: What are the similarities between the news and the empirical network layers? The results indicate that the co-occurrences of countries in financial news most closely match their geographical proximity. When we take the news sentiment into account, the positive sentiment layer is most similar to the trade layer. Somehow surprisingly for financial news, the financial layer of the empirical network has the lowest similarity to any news layer.

The paper is organized as follows. In Section 2 we describe entity recognition in news, and the construction of the news network layers in terms of co-occurrences, and by sentiment analysis. The empirical network layers are constructed from the geographical proximity, the trade volumes, and the correlations between financial indicators. We then propose several metrics to compare layers of multiplex networks. In Section 3, we first describe the acquisition and processing of textual data used in our case study. We then present the most interesting results of comparison between the news and empirical network layers. We conclude in Section 4.

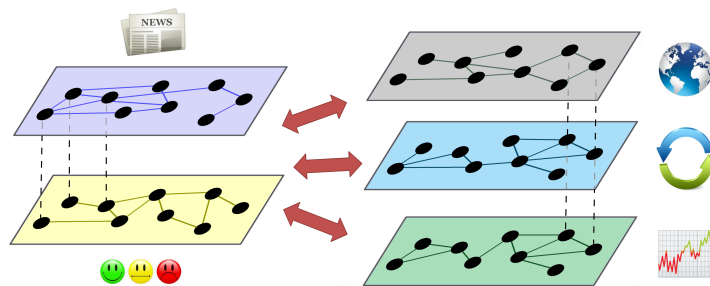


Fig. 1. Comparison of multiplex networks representing different types of relations between the same entities: significant co-occurrences and sentiment extracted from news (at left), versus geographical proximity, high trade, and high correlation of financial indicators (at right, top to bottom).

2 Methods

In general, when processing financial news, one first has to identify different entities, such as financial institutions, countries, or persons. Then one can construct layers of the news network, consisting of significant entity co-occurrences and their associated sentiment. We compare the news network to the empirical network, in order to discover connections between entity relations in different layers of both networks.

2.1 Entity recognition

Financial news are about events related to companies, stocks, countries or persons, which we call financial entities. The process of identifying financial entities in textual documents requires three components: an ontology of financial entities and terms, gazetteers of the possible appearances of entities in the text, and a semantic annotation procedure that finds and labels the entities. We describe the entity identification approach as implemented in our NEWSSTREAM portal [11]¹.

The ontology we use for information extraction constitutes of three main categories: financial entities, financial terms, and geographical entities. The ontology also includes a dictionary of positive and negative words for dictionary-based sentiment analysis². Most of the ontology is automatically induced from various data sources. The geographical entities (continents, countries, cities, organizations) were extracted from GeoNames³. The IDMS database and MSN Money⁴ were 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 experts in economics. It includes the main European politicians, Central banks and other financial institutions, rating agencies, fiscal and monetary policy terms.

Each entity in the ontology has associated gazetteers, which are sets of rules that specify the lexicographic information about the possible appearances of entities 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, the so-called semantic annotation procedure annotates the entities of interest. It traverses each document and searches for entities from the financial ontology. The gazetteers of the entities in the ontology provide information required for the disambiguation of different appearances of the observed entities, resulting in the correct uniform annotation of entities.

¹ <http://newstream.ijs.si/>

² Harvard-IV-4 sentiment dictionary [20,21]

³ GeoNames: <http://www.geonames.org/>

⁴ MSN Money: <http://money.msn.com/>

2.2 Network construction

For a particular set of entities $E = \{e_1, \dots, e_n\}$ we construct networks that are obtained from different data sources. We distinguish two networks: *News network* – constructed from entities and relations appearing in financial news, and *Empirical network* – with the same entities, but linked by relations extracted from other information sources or databases. Each network consists of several layers, and each layer is constructed for a particular time period, resulting in a series of snapshots for each network.

News network. Financial entities identified in a single news document can be connected with various types of relations. One of the simplest is their common appearance in the document, referred to as the co-occurrence of entities. Hence, for a selected set of entities $E = \{e_1, \dots, e_n\}$ we construct a layer of entity co-occurrences within a particular time frame – the *Co-occurrence layer*. Each link in such a layer represents a significant co-occurrence relation between two entities. We use the Significance algorithm proposed in [15] to assess whether the co-occurrence of two entities is significant.

The number of all documents with at least two entities from E is N . Let A and B be two entities that occur with at least one other entity from E in N_A and N_B documents, respectively. Let N_{AB} denote the number of actual A and B co-occurrences. Then the expected number of co-occurrences is given by

$$\mathbb{E}(N_{AB}) = \frac{N_A N_B}{N}. \quad (1)$$

According to [15], the standard deviation is

$$\sigma_{AB} = \sqrt{\frac{N_A N_B}{N} \left(\frac{N^2 - N(N_A + N_B) + N_A N_B}{N(N-1)} \right)} \quad (2)$$

and hence the standard significance score of the co-occurrence N_{AB} from the data is

$$Z_{AB} = \frac{N_{AB} - \mathbb{E}(N_{AB})}{\sigma_{AB}}. \quad (3)$$

For a selected threshold Z_0 , one can distinguish significant $Z_{AB} > Z_0$ and non-significant $Z_{AB} < Z_0$ co-occurrence relations between the two entities.

A set of entities can be linked also by other types of relations, e.g., based on expressed sentiment in documents which discuss the entities. We construct a *Sentiment layer* of the news network by detecting sentiment orientation and strength of financial news articles which mention pairs of entities in a specific time period. A sentiment link between two entities in the layer exists if its sentiment value is higher than a predefined threshold.

In order to calculate sentiment between entities, we use the sentiment analysis implementation of the NEWSSTREAM portal. The implementation is dictionary-based, meaning that sentiment polarity of a document is based on the count of predefined sentiment terms (positive and negative) in the document. The

implementation relies on the Harvard-IV-4 sentiment dictionary [20,21]. For each document it calculates the overall sentiment polarity by applying the following formula:

$$polarity = \frac{pos - neg}{pos + neg} \quad (4)$$

where pos is the number of positive and neg is the number of negative dictionary terms found in the document.

Using the NEWSSTREAM portal we obtain the sentiment results calculated on the level of a document and aggregated by summing the results for each day. Moreover, since we are interested in making a snapshot of a network over a longer time period of T days, we further aggregate the obtained results for T days. Based on the analysis of the sentiment distribution, we determine the thresholds p_0 and n_0 for the creation of positive and negative sentiment links.

Empirical network. We observe the same set of entities E as in the ‘News network’, but the information regarding their mutual interactions is not acquired from the news. In particular, we explore three data sources to construct the empirical network layers: the geographical proximity of the financial entities, correlations between their financial indicators, and their direct interaction in terms of mutual trade. We use these layers as the underlying empirical representation of the complex relations between financial entities, from which we try to understand the dynamics of entity appearance in financial news.

The simplest among the ‘empirical network’ layers is the geographical proximity, or short the *Geo layer*. Each financial entity has a predominant geographical location, place of residence, headquarters address, area, country, continent of trade or market influence. A financial entity is assigned a geographical entity and hence a link between two entities is established if a certain proximity measure is above a given geographically feasible threshold. Examples of proximity measures include geographical distance $d(A, B)$, inverse distance $\frac{1}{d(A, B)}$, or inverse squared distance $\frac{1}{d(A, B)^2}$.

Constructing a layer from trade interaction data is also fairly straightforward. Consider the interaction between financial entities as the amount of mutual trade. Each entity $e_i \in E$ engages in $c(e_i)$ of trade with all other entities, therefore

$$c(e_i) = \sum_{e_j \in E \setminus \{e_i\}} c(e_i, e_j) \quad (5)$$

where $c(e_i, e_j)$ is how much e_i trades to e_j . Notice the implied direction of the trade. In our experiment we use an undirected *Trade layer*, and therefore define $t(e_i, e_j) = c(e_i, e_j) + c(e_j, e_i)$ as the cumulative trade exchange between e_i and e_j . A trade link between two entities e_i and e_j is established if any of the relative amounts $\frac{t(e_i, e_j)}{c(e_i)}$ or $\frac{t(e_i, e_j)}{c(e_j)}$ is above a given threshold t_0 .

Important financial entities have also an associated time-varying financial indicator (e.g., price, trade volume, confidence index), which is represented as

a time series. A basic approach to measure similar trends in the movement of financial indicators is the Pearson correlation [14] between time series s_i and s_j of the entities e_i and e_j , over a period of K time points:

$$\rho_{i,j} = \frac{\sum_{k=1}^K (s_{i,k} - \bar{s}_i)(s_{j,k} - \bar{s}_j)}{\sqrt{\sum_{k=1}^K (s_{i,k} - \bar{s}_i)^2 \sum_{k=1}^K (s_{j,k} - \bar{s}_j)^2}}, \quad (6)$$

where \bar{s}_i and \bar{s}_j stand for the average (arithmetic mean) value of the respective series. A *Financial layer* can hence be constructed using a threshold value c_0 , which determines whether the indicator time-series of two entities are sufficiently correlated ($\rho_{i,j} > c_0$) to form a link between them.

2.3 Network comparison

We described the construction of the two-layer *News network* and the three-layer *Empirical network*. Both networks share the same set of nodes, i.e., entities $E = \{e_1, \dots, e_n\}$, which are in each layer connected by a different type of relation. As a whole, we are considering a *multi-layer network* of the same nodes, also called a *multiplex network*.

We try to understand the ‘news network’ by comparison to the ‘empirical network’. In the ‘empirical network’ we construct a link between two entities when there is a ‘strong’ empirical relation between the entities. Although all layers contain the same entities, in the comparison isolated nodes are not considered. The layers that we compare are constructed from the sets of strongest links for a particular relation type.

The most straightforward comparison of the network layers $\mathcal{L} = \{L_1, \dots, L_m\}$ is done by measuring the size of link overlap between the layers. Let $l(L_i)$ and $l(L_j)$ be the sets of links in layers L_i and L_j , where a link is defined as a pair of nodes it connects, e.g., (e_u, e_v) , then

$$o(L_i, L_j) = \frac{|l(L_i) \cap l(L_j)|}{|l(L_j)|} \quad (7)$$

is the size of their link overlap relative to layer L_j .

Considering for each layer not only the links that indicate the strength of a relation above a certain threshold, but also their weight – strength of the relation, then a comparison of top strongest links in each layer can be performed. Let $sl(L_i)$ and $sl(L_j)$ be lists of links from layers L_i and L_j , ordered descending by their weights, and let $sl_k(L)$ denote the list of first k element of list $sl(L)$, then *precision-at-k* [16] is defined as:

$$r_k(L_i, L_j) = \frac{|sl_k(L_i) \cap sl_k(L_j)|}{k}. \quad (8)$$

If for all pairs of layers L_i and L_j , $i, j \in \{1, \dots, m\}$, the same k is selected, then a meta-network can be constructed with nodes representing layers L_i , $i \in$

$\{1, \dots, m\}$ and links representing the relation between layers, where $r_k(L_i, L_j)$ values are weights of the links, indicating the magnitude of the relationship.

Other comparisons of the network layers induced on the ‘strongest’ links for a particular relation type, are based on the most important nodes in each layer. In one approach, we measure the importance of nodes in terms of their *centrality*, as denoted by the *eigenvector centrality measure* [3]. Let A be the adjacency matrix of nodes e_1, \dots, e_n in the network, then the components of the eigenvector of the largest eigenvalue λ solving the equation $A\mathbf{x} = \lambda\mathbf{x}$ hold the centrality values of the corresponding nodes. Nodes connected to better-connected nodes get higher centrality values. This measure is used to compare which are the most central nodes between pairs of layers.

Another approach to identify the most important nodes of a network is the *k-core decomposition* [17]. This is an iterative process pruning all nodes with degree smaller than k , and the remaining part of the network which holds only nodes with degree greater or equal to k is called the k -core. The core with the largest k is called the main core of the network. Comparing the main cores of different network layers will be used to assess the similarity between layers.

3 Experiments

Financial news cover a wide range of topics, they include numerous entities of interest, and are influenced by different factors. We describe the data we use to show the presence and interaction of entities in the news, and the empirical data to model the real-world context that may shape the news. Next, we present the comparison results between the network layers.

3.1 Data

We acquire news articles and blogs from 2,503 RSS feeds from 170 English language web sites (14,567 domains), covering the majority of web news in English and focusing on financial news and blog sources. We collect data from the main news providers and aggregators (like yahoo.com, dailymail.co.uk, nytimes.com, bbc.co.uk, wsj.com) and also from the main financial blogs (like zerohedge.com). The fifty most productive web sites account for 80% of the collected documents.

The documents used in our experiment cover the period from November 1st, 2011 until December 31st 2013. A total of 18 million documents were filtered for strictly financial news, resulting in 1.3 million documents. From these documents we extract relevant entities, and construct the ‘news network’ layers in monthly time windows. For our analyses, we select 50 countries as entities of interest. Snapshots of the resulting co-occurrence layer, positive sentiment layer, and negative sentiment layer are shown in Figures 2, 3 and 4, respectively.

On the other hand, the construction of the empirical network, which should reflect the real-world context of the news, was done using data from different sources. For the ‘Geo layer’ we simply used the is-a-neighbour-of relation to link the selected countries. To the links representing common terrestrial borders we

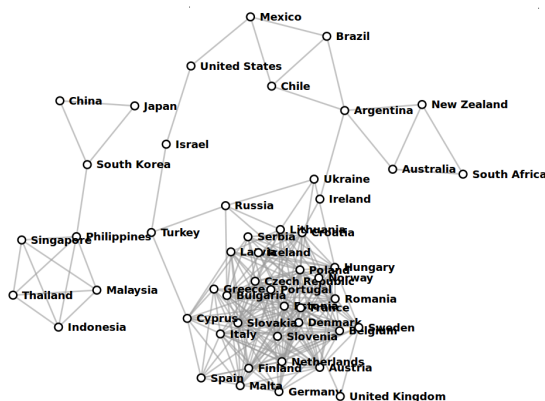


Fig. 2. Snapshot of the country co-occurrence network layer (October 2012).

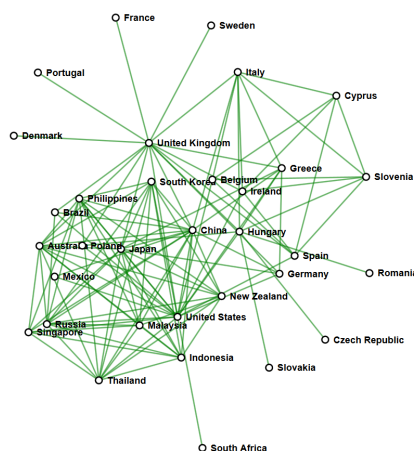


Fig. 3. Positive sentiment network layer (Jan 2012).

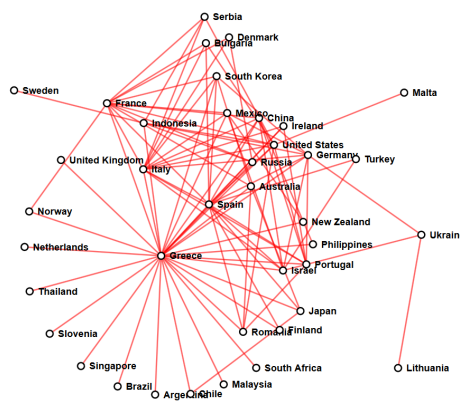


Fig. 4. Negative sentiment network layer (May 2012).

added also a few links between countries that are considered relatively adjacent in the local geographical context, such as Australia and New Zealand, South Korea and Japan, or Italy and Malta.

Trading relations between the countries were obtained from the *UNCTAD website*⁵, the United Nations statistics data center, providing yearly aggregations of trade data. Our ‘Trade layer’ was constructed from trade links that present relatively important trade relations (greater than 10%, i.e. $t_0 = 0.1$) for at least one of the connected countries.

In our experiments we consider 50 countries that issue sovereign bonds, and which are insured by Credit Default Swaps (CDS), i.e., an insurance for the case when the bond issuer ‘defaults’ and is unable to repay the debt. To construct the ‘financial layer’ we used the time series of the countries’ CDS prices, which are often considered a good proxy for the risk of default of a financial institution

⁵ <http://unctadstat.unctad.org>

issuing bonds [1,13]. We create links between countries whose correlation between their CDS time series is above 0.9 ($c_0 = 0.9$). In order to ensure enough data for reliable correlation results, we use a three-months time window for each snapshot, which is assigned to the last month (e.g., Nov-Dec-Jan for the ‘Jan’ snapshot). A snapshot of the ‘Financial/CDS layer’ and the ‘Trade layer’ are presented in Figures 5 and 6, whereas the presentation of the ‘Geo layer’ is well known and therefore omitted.

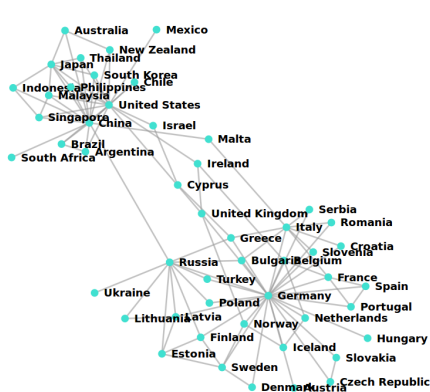


Fig. 5. Trade layer (2012–2013).

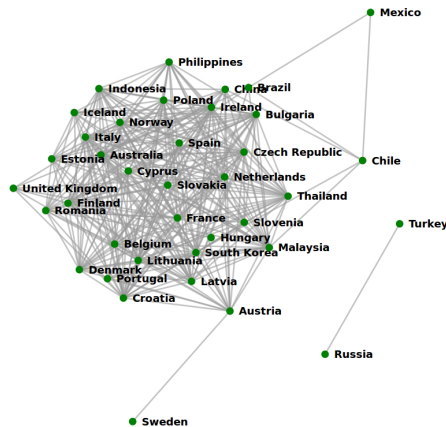


Fig. 6. Financial/CDS layer (October 2012).

3.2 Results

The results are presented for a multiplex network with 50 country nodes for the time period of two years in monthly steps. The Geo layer is static, as well as the Trade layer — we used the yearly aggregated trade data from 2012 also for 2013.

First, we present the analysis of overlapping links between the network layers L_{CO} , L_{Geo} , L_{Tr} , and L_{CDS} . We are interested in the number of links from the ‘empirical network’ that appear in the financial news as country co-occurrences over time. The relative overlaps $o(L_{CO}, L^*)$ for $L^* \in \{L_{Geo}, L_{Tr}, L_{CDS}\}$ are presented in Figure 7. We see that most of the Geo layer links coincide with the country co-occurrences in the news, whereas on average less than half of the links between the countries in the Trade and CDS layers also appear in the co-occurrence layer.

Next, we investigate how is the sentiment associated with the country co-occurrences related to the empirical network. Using the sentiment analysis approach presented in Section 2.2 we find that there is a strong bias towards positive sentiment in the financial news. We set thresholds n_0 and p_0 to two standard deviations apart from the average sentiment polarity in the documents, thus selecting only links that reflect the most negative and most positive sentiment in the context of two countries. The negative sentiment layer turns out to be predominantly small, even for a slightly less restrictive threshold n_0 (at 90% st. dev. from the average) and therefore has mostly low overlap with the empirical layers. On the other hand, the comparison of the positive sentiment layer L_P

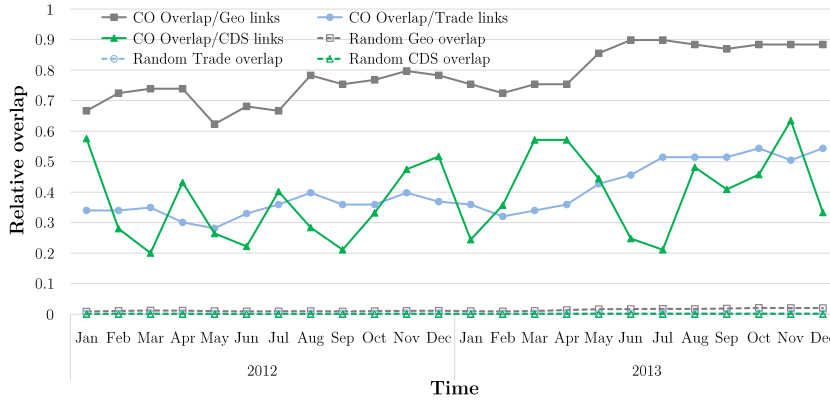


Fig. 7. Relative size of the empirical layer links present in the co-occurrence layer.

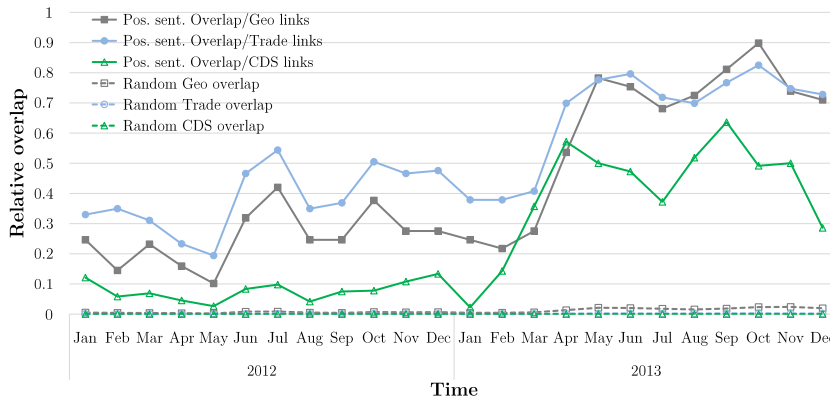


Fig. 8. Relative size of the empirical layers' links present in the positive sentiment layer.

with the empirical layers results in a larger number of common links, as shown in Figure 8. Positive sentiment between the countries has the largest overlap with the trade relations, followed by geographical proximity and to the smallest extent by the correlation between the CDS time series.

Comparison of the most important nodes in each layer shows similar results. The comparison of main k -cores results in the largest overlap between the co-occurrence and Geo layer cores, and positive sentiment and the Trade layer cores. The co-occurrence layer cores overlap with the Geo layer cores in central European countries, and with the Trade layer cores in western European countries. The overlaps between the co-occurrence and CDS layers show common presence of some eastern European countries in 2012, but no regular presence in 2013. Several countries regularly appear in the core overlap between the positive sentiment and the Trade layers (CN, DE, US, UK, JP, BR, FR, and AU). Germany is also almost all the time (23 months) in the core overlap of the positive sentiment and Geo layers.

Most central nodes of the co-occurrence layer coincide with the Geo layer in central European countries (AT, CZ, HU, SK, SI), with the CDS layer in few eastern European countries, and with the Trade layer only Finland appears often

among the top ten most central nodes. For the positive sentiment layer, the common most central nodes are Germany and Russia for the Geo layer, and some of the largest economies (CN, DE, FR, JP, RU, US) for the Trade layer.

Finally, we use the *precision-at-k* method to measure the link overlap of the strongest relations in each layer. Limited by the number of links in the Geo layer, k was set to 69. Figure 9 illustrates the relations between layers weighted by the precision-at-‘69’ values.

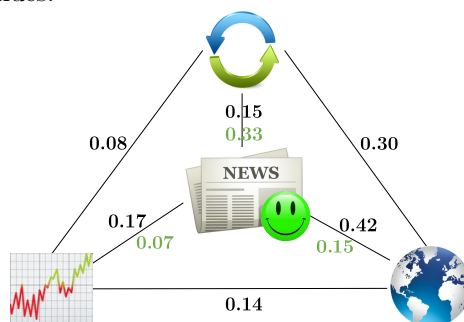


Fig. 9. A meta network between the news and empirical network layers.

4 Conclusions

In the paper we present methods to extract nodes and layers of multiplex networks. The emphasis is on the network construction from large textual streams, where entity recognition, co-occurrence extraction, and sentiment analysis are performed. We propose several metrics to compare different layers of the same network. In the case study, we compare the news layers to the empirical layers, and conclude that financial news mostly reflect geographical proximity and trade relations, and much less financial relations. These results provide the background, long-term characterization of the news. We speculate that the underlying reason is the method to calculate the significant co-occurrences, which computes the expected co-occurrences from individual frequencies. Alternatively, one could compare actual co-occurrences in a short time window to a longer time window, and so detect potentially interesting deviations. In our future work, we intend to investigate this alternative approach and evaluate how the news layers constructed in such a way compare to different empirical layers.

Acknowledgments. This work was supported in part by the EC FP7 projects SIMPOL (no. 610704) and MULTIPLEX (no. 317532), and by the Slovenian ARRS programme Knowledge Technologies (no. P2-103).

References

1. Aizenman, J., Hutchison, M., Jinjark, Y.: What is the risk of European sovereign debt defaults? Fiscal space, CDS spreads and market pricing of risk. *Journal of International Money and Finance* (2012)

2. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. *Reviews of modern physics* 74(1), 47 (2002)
3. Bonacic, P.: Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology* 2, 113–120 (1972)
4. Caldarelli, G.: *Scale-free networks: complex webs in nature and technology*. Oxford University Press (2007)
5. 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)
6. 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)
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. Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., Ruppim, E.: The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic acids research* 38(12), 3857–3868 (2010)
9. Ghoshal, G., Zlatić, V., Caldarelli, G., Newman, M.: Random hypergraphs and their applications. *Physical Review E* 79(6), 066118 (2009)
10. Jackson, M.O.: *Social and economic networks*. Princeton University Press (2010)
11. Kralj Novak, P., Grčar, M., Mozetič, I.: Analysis of Financial News with NewsStream. Tech. Rep. IJS-DP-11892, Jožef Stefan Institute, Ljubljana, arXiv:1508.00027 (2015)
12. Mane, K.K., Börner, K.: Mapping topics and topic bursts in PNAS. *Proc. National Academy of Sciences* 101(Suppl 1), 5287–5290 (2004)
13. Pan, J., Singleton, K.J.: Default and recovery implicit in the term structure of sovereign CDS spreads. *The Journal of Finance* 63(5), 2345–2384 (2008)
14. Pearson, K.: Note on regression and inheritance in the case of two parents. *Proc. Royal Society of London* 58, 240–242 (1895)
15. Popović, M., Štefančić, H., Sluban, B., Kralj Novak, P., Grčar, M., Mozetič, I., Zlatić, V.: Extraction of temporal networks from term co-occurrences in online textual sources. *PLoS ONE* 9(12), e99515 (2014)
16. Raghavan, V., Bollmann, P., Jung, G.S.: A critical investigation of recall and precision as measures of retrieval system performance. *ACM Transactions on Information Systems* 7(3), 205–229 (1989)
17. Seidman, S.B.: Network structure and minimum degree. *Social Networks* 5(3), 269–287 (1983)
18. Shalgi, R., Lieber, D., Oren, M., Pilpel, Y.: Global and local architecture of the mammalian microRNA–transcription factor regulatory network. *PLoS computational biology* 3(7), e131 (2007)
19. 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)
20. Tetlock, P.C.: Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance* 62(3), 1139–1168 (2007)
21. 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)
22. Wilkinson, D.M., Huberman, B.A.: A method for finding communities of related genes. *Proc. National Academy of Sciences* 101(Suppl 1), 5241–5248 (2004)
23. Zlatić, V., Ghoshal, G., Caldarelli, G.: Hypergraph topological quantities for tagged social networks. *Physical Review E* 80(3), 036118 (2009)