Special Collection: Social and Collaborative Networks and their Applications

INTERNATIONAL JOURNAL OF
ENGINEERING BUSINESS
MANAGEMENT

International Journal of Engineering
Business Management
Volume 9: 1–12
© The Author(s) 2017
DOI: 10.1177/1847979017720040
journals.sagepub.com/home/enb
SAGE

Big data analysis of economic news: Hints to forecast macroeconomic indicators

Mohammed Elshendy and Andrea Fronzetti Colladon

Abstract

We propose a novel method to improve the forecast of macroeconomic indicators based on social network and semantic analysis techniques. In particular, we explore variables extracted from the Global Database of Events, Language, and Tone, which monitors the world's broadcast, print and web news. We investigate the locations and the countries involved in economic events (such as business or economic agreements), as well as the tone and the Goldstein scale of the news where the events are reported. We connect these elements to build three different social networks and to extract new network metrics, which prove their value in extending the predictive power of models only based on the inclusion of other economic or demographic indices. We find that the number of news, their tone, the network constraint of nations and their betweenness centrality oscillations are important predictors of the Gross Domestic Product per Capita and of the Business and Consumer Confidence indices.

Keywords

Economic forecasting, GDELT, Business Confidence Index, Consumer Confidence Index, GDP, social networks

Date received: 7 February 2017; accepted: 6 June 2017

Introduction

Hamermesh,[1] in his study about the changing nature of economics research over the last six decades, found that in 1963 economic research was 50.7% theoretical and empirical research was negligible. In 2011, we saw an overturn, with empirical research reaching 64% and theoretical research dropping to 19%.

One of the main determinants of this shift is the larger quantity and variety of data that are now easily available, in connection to all different sorts of economic events. Data are now available in very large volumes, in both structured (organized and easily searchable) and unstructured formats.[2] To unveil the insights that can come from data analysis, many organizations recruited data scientists and established data analysis departments. Dutton[3] stated that we should move from the data driven organizations era towards the data driven culture. Similarly, Bizer et al.[4] showed that thanks to big data we have new approaches to solve old problems.

As regards news data, the collection process often comprised a deep search or manual extraction of paper documents.[5] Nowadays, computers and text mining algorithms[6] allow the extraction of many document-related variables. For example, how many times an article has been viewed, tweeted, liked, pinned or commented, the sentiment of its content, its appreciation, the number of readers per country and so on. Thus, we can use new variables, which probably carry new insights and useful information.

Simultaneously to the rise of data and web technologies, the study of network interactions – both within and outside

Department of Enterprise Engineering, University of Rome Tor Vergata, Rome, Italy

Corresponding Author:
Andrea Fronzetti Colladon, Department of Enterprise Engineering, University of Rome Tor Vergata, Via del Politecnico 1, 00133 Rome, Italy.
Email: fronzetti.colladon@dii.uniroma2.it

firms – proved its importance. Positions and interaction dynamics in communication and business networks proved to be connected, for instance, to employees' engagement and turnover intentions,[7] innovative capabilities of start-ups[8] or fraudulent activities.[9] An additional example is given by Shipilov and Li[10] who found that in corporate networks, higher centrality predicts higher market performance, while higher values of the network constraint are associated with a lower performance. The study of socio-economic networks is relevant to this research, which follows a social network analysis (SNA) approach[11] to determine which elements of these networks can contribute to the prediction of macroeconomic indicators. Socio-economic networks can offer a graphical representation of economic, political and social interactions. For example, they can be used to represent the way buyers and sellers interact in a market during the trade of goods and services, the diffusion of behaviours and opinions or the formation of commercial alliances. Recently, socio-economic networks evolved from just being able to illustrate patterns of connections among small individual groups, to represent large-scale, high-density statistical graphs, with millions of actors and links.[12]

Past research on forecasting macroeconomic variables[13–16] focused mainly on providing new mathematical and statistical forecasting techniques, examining small and large sets of data with varying degree of success. However, these forecasting models were mostly using other economic predictors – such as the gross domestic product (GDP) or inflation and interest rates – while ignoring network interactions dynamics.

In this context, we developed a novel approach based on the extraction of socio-economic networks from the content of news articles related to international business operations. We present an analysis of these networks with the purpose of demonstrating how SNA measures can help in forecasting the degree of optimism about the state of economy represented by the Consumer Confidence Index (CCI) and the Business Confidence Index (BCI) and the GDP per Capita (GDPCapita) of the 10 biggest economies in the Euro area. Our choice to predict those indexes has two main reasons: first, they are sensible indicators of the economic stability of a country, as they relate to the spending activity of consumers and to the investment activity of firms; second, journalism and news media can influence the opinion of both consumers and business managers.[17]

With this study, we offer useful insights for scholars and economic analysts, who are interested in improving existing forecasting models of macroeconomic indicators. The novelty of our contribution mainly lies in the metrics we use, in the network mapping techniques we present and in the use of less explored data sources, such as the Global Database of Events, Language, and Tone (GDELT) database. Our metrics – extracted from data sources which are freely available and accessible in real time – are mostly meant to be integrated in other existing models to improve their accuracy. Our findings can also be partially relevant to policymakers supporting, for instance, the importance for a country to be more central in the economic networks we describe.

Forecasts of macroeconomic variables

Macroeconomic variables prediction has been a subject of interest in the academia since decades.[1] Many scholars tried to develop different models, which mostly included economic variables, to predict macroeconomic indicators at the country level. Sims[18] was among the first authors to test and propose a linear vector autoregression (VAR) model to forecast the US macroeconomic variables related to the business cycle fluctuations. The model provided good results for the in-sample forecast but a poor output for sample forecast because of a problem of overfitting. In a later study, Litterman[19] proposed the Bayesian VAR technique which solved the over fitting problem by reducing the VAR's parameters. At the same time, King et al.[20] proposed a vector error correction model, which considered long run effects and stochastic trends as random walks, to constraint the VAR parameters. More recently, Gupta and Schaling[21] proposed an integrated model that combined the prementioned techniques in order to forecast the GDP and the interest rates of South Africa. Similarly, Kuzin et al.[22] forecasted the quarterly GDP growth for the Euro area.

Other models were developed to consider non-linear effects. Within this field, the self-exciting threshold autoregressive (SETAR) model[13] received a large attention. Cuaresma[23] used a two-regime (SETAR) process to estimate the quarterly GDP for 15 countries composing the European Union (EU). Feng and Liu[24] used the SETAR model in addition to an autoregressive integrated moving average model to forecast the Canadian GDP.

Despite the above-mentioned approaches showed good results, they were all efficient only for long-term forecasting, using long series of historical data. For the short-term forecasting, the academia got oriented towards the bridge equations (BEs). Several scholars[25–28] tested the BEs approach as a method which could combine real and forecasted data for the nowcasting of the GDP in the Euro area.

With the significant growth of computational potentials, which allowed the reconsideration of many abandoned information, other forecasting techniques arose based on machine learning and text mining algorithms. Stock and Watson[29] proposed a dynamic factor model to estimate macroeconomic indexes of the US economy. Giannone et al.[30] proposed a factor model to evaluate the current-quarter real measure of the GDP growth rate. Their factor model was applied to a large data set of about 200 macroeconomic, financial and surveys indicators. Results showed that the use of those large data sets increased the forecast precision monotonically with each inclusion of new data. Schumacher[31] tested the forecasting performance of two alternative factor models, static and dynamic, based on large pools of Germany quarterly time series data.

In this research, we use a simpler approach based on multilevel regression, with the inclusion of a new set of mostly unexplored variables. With this approach, we can easily stress the predictive power of these new variables, as this is the main contribution of our research, together with the presentation of new network mapping techniques. To put it in other words, our objective is not to present a new set of final forecasting models but to give evidence to the role played by the new variables we discuss in the article. We maintain this metrics could be useful for scholars, as they could integrate them in other existing models, to improve their accuracy.

Using network analytics to improve economic forecasting

SNA received big attention on the last decade. Initially, mainly used in the field of sociology, the SNA approach was then implemented in many other fields including anthropology, psychology, epidemiology, politics, physics and economics.[32]

In the economic field, one of the main reasons for this growth is the innovative support that SNA can offer in terms of producing new variables that help to improve existing models. Jackson[33] stated that recent interest in SNA mainly comes from the fact that the most classic economic analysis techniques have been pushed to the limit and they often fail in explaining the social circumstances observed during a certain economic phenomenon. Mizruchi and Stearns[34] applied SNA to investigate the management of uncertainty in evaluating potential transactions in a multinational leading commercial bank, particularly studying the means by which managers close transactions with corporate clients. They mapped the network of social relationships between bank employees at three locations and analysed how the network structure evolved under the conditions of uncertainty about the nature of a certain deal. Their findings are consistent with several studies[35–37] and proved that the formation of sparse networks, in which weak ties are more dominant, can promote the access to diverse and non-redundant ideas; by contrast, too dense networks can lead to redundant knowledge and communications schemes. Similarly, Granovetter[38] studied the impact of social structures on economic outcomes, investigating speed of flow, quality of the information, mechanisms of reward and punishment and trust relationship in networks. He gave evidence to the contribute of network analysis in studying price fluctuations, behaviours of business groups, firms' strategy and productivity and innovation in advanced economies.

Hidalgo et al.[39] investigated the network of relatedness between products, which they called 'product space', explaining one of the reasons why poor countries have more trouble to develop competitive exports and fail to reach the income levels of rich nations.

Garlaschelli et al.[40] constructed a directed weighted network representation of the world trade web (WTW) in order to capture the interplay between the dynamics of the GDP and the trade movements. They weighted each link (trade) according to the amount of wealth flowing from one actor (country) to the other and characterized each actor through its GDP value. Results showed that GDP values have a great influence on the evolution of WTW and vice versa, thus the probability of two countries being tightly connected is higher in the case of a similar level of GDP.

Our approach has some similarities with the above-mentioned studies with two main differences. First, we do not collect the available WTW data from online data banks: we use a novel approach to build interaction networks based on the analysis of the world's news articles and broadcasts included in the GDELT database. Second, we build predictive models to assess the informative power of new SNA and news-related metrics, while making macroeconomic predictions.

Case study

The GDELT is an open-source repository of news articles, which are continuously updated and made available to researchers. The dataset includes news organized in more than 300 different categories of events, which took place since 1979. Kwak and An[41] called it a tale of the world. It connects a massive amount of data regarding persons, organizations, locations and events across the planet. The database relies on tens of thousands of broadcasts, print and online news sources from every corner of the globe.[42] In order to crawl such a large amount of data, we used the Google BigQuery (https://cloud.google.com/bigquery/) service, that is a platform provided by Google which allows fast queries (written using the standard SQL programming language) on very big datasets. The GDELT database is already connected with Google BigQuery, thus making it easily analysable.

The news data were collected on a daily basis for the period between 1 January 2010 and 31 March 2016. We filtered the news referred to business events involving firms operating in each country of the world, and we recorded the countries involved. The resulting dataset contained more than one million and seven hundred thousand entries.

We organized our analysis with the intent of building a model for the prediction of the GDP, BCI and CCI indices, for the top 10 EU countries (Germany, France, Italy, Spain, Netherlands, Belgium, Austria, Greece, Finland and Portugal). To obtain the dependent variables, we consulted the data bank of the intergovernmental Organisation for Economic Cooperation and Development. We could export the GDPCapita on quarterly basis, represented in thousands of US dollars, and the CCI and BCI which were available with a monthly frequency. Both confidence indices are based on responses to survey questions about the business conditions of a country and the likely developments in the months

Table 1. Description of the variables extracted from GDELT.

Variable	Description
Actor country code	A three-character CAMEO code which identifies the countries involved in an event.
Action type code	A three-character CAMEO code for the type of action where a country is involved. This can be a specific action categorized as business, education, media and others. We limited our selection to business actions.
Host country code	Specifies the location in which the event takes place. Can be the same as the countries involved in the event, or a different country.

GDELT: Global Database of Events, Language, and Tone; CAMEO: Conflict and Meditation Event Observations.

ahead. In the case of CCI, surveys are mainly administered to households and consumers to evaluate their current economic situation, their evaluation about the labour market and their expectations for the near future. The BCI is similar, but for this index surveys are administered to business managers, assessing their overview about the state of the economy and their optimism about the future of their organizations. The BCI survey assesses the enterprises state about production, stocks, current value and expected value in the near future.[43,44]

Network extraction and graphing

On Google BigQuery, we searched for events included in the category of Conflict and Meditation Event Observations (CAMEO). Subsequently, we restricted the search criteria to business events in order to exclude other possible types of events that fall under the economic cooperation category, such as education, media, military and so on. In addition, we considered exclusively articles for which the following three variables could be clearly defined: (a) actor country code, (b) host country code and (c) action type code. These variables are described in Table 1.

In our case study, each interaction had three actors involved: two contracting countries that were the two parties involved in each event and a host country where the event took place, with the possibility, for all countries, to play a multiple role. For instance, a German company could be a contractor in one transaction and the host in another one.

We analysed the data at three different levels: firstly, only considering the connections between contractors; secondly, including the location of the agreement; thirdly, combining the first two networks. Network construction techniques are shown in Figure 1. In each graph, the different countries were represented as nodes and the interactions among them as graph edges (linking pairs of nodes who interacted according to the rules described in the following).

Interaction network. In this graph, we created links between the contracting countries, for each transaction, without considering the host country representation. For example, if a German firm makes an agreement with a Spanish one in Italy, the transaction will be represented by two nodes, Germany and Spain connected by a link, ignoring the role of Italy.

Location network. In this network, we stress the role of the host country, ignoring the link between the countries of the contractors. Using the same example, if a German firm makes an agreement with a Spanish one in Italy, the transaction will be represented with two ties: one connecting Italy to Germany and the other one connecting Italy to Spain. Accordingly, each transaction is represented by three nodes and two links. This network has a major role in the identification of relevant countries where business events take place.

Joint network. This final network is simply the combination of the previous two. Therefore, this network maps each transaction by three nodes and three links, connecting all parties involved in a transaction.

It is important to notice that even if our analysis is focused on 10 European countries, the above-mentioned networks comprise much more nodes (representing all the countries involved in economic events during the study period). This choice is vital to properly assess the role of the 10 analysed countries without losing information; accordingly, the social network metrics for the 10 nodes are calculated considering the full networks. The choice of separating the networks is not mandatory and we leave to the analyst the final decision. In our case, this choice is important for two reasons: it gives evidence to different network mapping techniques and to the flexibility that can be used while analysing relational data; it also allows to separate the contribution coming from different network metrics.

Independent and control variables

Network studies are often useful since the analysis of relational data can add informative power to the individual attributes of social units.[45] In this study, we considered well-known centrality measures, which are commonly used to assess the influence and positional power of nodes within networks.[45,46] In particular, centralities explain how strategic is the position of a certain actor in terms of connectivity and possibility of being in-between the patterns that keep together the other social actors.[47] Therefore, centralities express a tacit ranking of actors. Numerous centrality measures were developed and discussed in the previous literature.[48] In this study, we refer to some of the most common and widely accepted measure of degree centrality, betweenness centrality and closeness centrality.[47]

Degree centrality. The degree centrality of a node is the number of edges which are connected to that node.[49] The degree centrality is interpreted by some researches as a

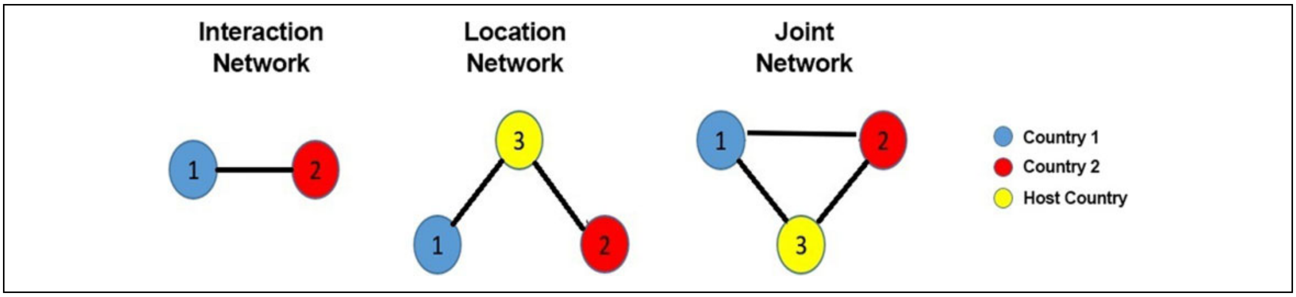

Figure 1. Network construction techniques.

density measure, because a high degree is representative of a high number of direct connections and often linked to higher activity levels.[50]

Closeness centrality. Closeness centrality is a measure which defines how close a node is to all others, in terms of network distance. It does not depend only on the direct connections, but it also considers the indirect connections. It is calculated as the inverse of the distance of a node from all others.[45]

Betweenness centrality. Similarly to closeness centrality, betweenness centrality is a path-dependent centrality measure. It measures how many times a node lies in the shortest paths that interconnect the other nodes (i.e. the paths that allow a connection between two generic nodes, passing through the minimum possible number of edges). The more in-between the node, the more relevant its social position.[45]

Betweenness centrality oscillations. Betweenness centrality oscillations counts the number of significant changes in betweenness centrality, for each node, over specific time frames. If a node betweenness centrality remains rather constant, we observe no variations; on the other hand, if it reaches a local maxima or minima, we count this changes in position.[51] This measure has also been called 'rotating leadership' and proved to be an important proxy for the assessment of the innovation capabilities of individuals and firms.[8,52]

Network constraint. The network constraint indicates the extent to which the direct links of a node are concentrated in groups of mutually connected peers or distributed to reach different groups of non-interconnected nodes.[53] Nodes with a more open ego-network have higher opportunities to mediate connections between peers, since they are less constrained by pre-existing links. Such positions proved to be related to positive performance evaluations, promotions and good ideas.[54] The value of network constraint is calculated as suggested by Burt.[55]

Multiple edges between two nodes are possible, if the corresponding countries were involved together in more than one economic event. Before calculating the above-mentioned centrality measures, we replaced multiple edges with single edges, weighted with the number of interactions between each pair of countries.

With regard to the following measures, we collected, from the GDELT database, their daily values. Subsequently, we calculated their monthly average or sum, over the months included in our study period – to match the frequency of our dependent variables.

Number of articles. A quantitative measure calculated for each country, which counts the number of news which mentioned that country, in the occasion of a business event.

Average tone. The average tone is a numerical score that indicates the degree of positivity/negativity delivered by each news, calculated by means of text mining techniques, and already included in the GDELT database.

Goldstein scale. First introduced by Goldstein,[56] it is a scaling technique that assigns a score, from -10 to $+10$, to each type of event, capturing the likely impact that an event can have on the stability of a country, based on its type (a riot, for instance).[42] Also this measure was already included in the GDELT database.

Finally, we considered a set of control variables, which were commonly used in previous research when trying to make macroeconomic predictions.[57–63] For the BCI and CCI, we used the nominal GDP, inflation rates and interest rates as control variables. For the GDPCapita, we used the inflation and interest rates and the country population.

We carried out an explorative study to test the relevance of the above-mentioned variables to make macroeconomic predictions. From previous studies, we got suggestions to think that social actors with more dynamic network positions (higher betweenness centrality oscillations) could better perform in business operations.[8] Moreover, one could also expect that activity and sentiment can influence the CCI[17] and BCI[64] and ultimately the GDPCapita.[65]

Results

Table 2 shows the correlation coefficients for each of the study variables for all the networks. These primary results fully support our idea that SNA can help explaining macroeconomic indicators.

Degree centrality shows a strong positive association for all the three macroeconomic variables in all the networks. This seems to suggest that more economic interactions with different parties are a healthy indicator for a national economy. Betweenness centrality shows a positive association with the GDPCapita in the interaction network and with the BCI and CCI in all the graphs. It seems that the more a country is in the economic paths that interconnect the other countries, the higher its GDP and the confidence levels expressed by businesses and consumers. By contrast, closeness centrality does not show any significant association with the dependent variables: this might be due to the relatively small size of these networks and to the possibility for most nodes to quickly reach the others, with short paths. As regards betweenness oscillations, this seems to be one of the most important predictors, being positively and significantly associated with all the indicators, in all networks. Lastly, the network constraint is negatively associated with our dependent variables in the interaction networks. This finding is consistent with the idea that a more open, less constrained, network position can be beneficial to business interactions.[55]

Looking at the other metrics extracted from GDELT, we see that the number of news articles referred to a specific country are positively associated with its GDP, BCI and CCI. It seems that this relationship still holds, regardless of the average tone used in the articles. In fact, we found no significant association of tone with GDPCapita and BCI; the only exception is for CCI, where we find a positive link of the tone with the consumers' optimism. Finally, the Goldstein scale showed a negative association with the GDPCapita and the BCI. Additionally, results showed significant correlation between the control variables and almost all the social network measures, implying that those variables could be considered as potential predictors even for the GDP, inflation and interest rates. To expand our findings, we implemented the multilevel regression models presented in Tables 3, 4 and 5. We used these models as they allow varying intercepts across countries, as well as to distinguish the proportion of variance which is attributable to the differences at a country level. We refer to the work of Nezlek[66] and to other well-known studies[67,68] for a better description of the potentialities of these models. We also tested the inclusion of random slopes, without obtaining better results. Additionally, we tried including lagged predictors in our models, with no improvements in this specific sample.

We built our models grouping repeated observations over time (level 1) by country (level 2). We tested our independent variables in blocks, avoiding to put together those metrics for which we identified collinearity problems. For each dependent variable, we started by running an empty model with no predictors, in order to measure the intraclass correlation coefficient (ICC).[66] Subsequently, we ran further models to assess the predictive power of each block on its own, then all the significant predictors were tested together in a final model, for the evaluation of the variance reductions. The ICC is 46.4%, 39.5% and 37.8% for the BCI, CCI and GDPCapita, respectively, indicating that at least one-third of each sample variance depends on differences at the country level. The models show that each macroeconomic variable can be better predicted by a specific set of independent variables. Specifically, betweenness oscillation in the location network and the average tone can help explaining the variance of BCI. CCI, on the other hand, seems to be more affected by the betweenness oscillations of countries in the interaction network, by the number of news and by the value of network constraint in the joint graph. Lastly, the GDPCapita can be predicted with the inclusion of a larger set of significant metrics, with consequently larger variance reductions: betweenness oscillations and network constraint in the interaction network, network constraint in the location network, number of articles and the Goldstein index.

Overall, the inclusion of the variables we proposed allowed a fair improvement of the predictive models, when compared with the empty models and with the models which just comprised the control variables.

Discussion and conclusions

In this study, we gave evidence to the informative power of novel metrics coming from SNA, while making economic predictions. In particular, we crawled the GDELT database, considering a 6-year period, to obtain useful data about news and world's broadcasts, related to economic events, which we converted into three network graphs. In order to build these graphs, we analysed the economic events which involved each nation in the world. As a second step, we built forecasting models to try predicting the GDPCapita, the CCI and the BCI, for the top 10 economies in the EU.

We propose a novel approach which combines network analysis and news-related variables (such as their volume and tone), offering new metrics that can be easily integrated in other existing predictive models, with the aim of improving their accuracy.

We find that the BCI is partially explained by the betweenness centrality oscillations of a nation and by the tone of the news referred to the economic events that involved that nation. The CCI, mostly dependent on consumers' perception, can be partially predicted considering a country betweenness oscillations, network constraint and number of mentions in economic news. Similarly, the GDPCapita seems to be affected by a country network constraint, oscillations in betweenness centrality, number of news articles and their Goldstein index. We also find additional evidence that the BCI and CCI are good predictors of the GDP, which is consistent with previous studies.[59,69,70]

The findings of our research are consistent with other previous studies which support the value of the information extracted from new online platforms, such as GDELT.[71] Our research is also partially linked to the literature focused on the relationship between network structures and innovation capabilities.[72] This study, mainly explorative, has several limitations, mostly related to the process of news data collection. The events in the GDELT database are classified according to the CAMEO data book: it might be the case that we missed some economic events, for example connected to the education sector, or sometimes not classified

Table 2. Correlation coefficients.

		Dependent variables			Control variables				Interactions network					Location network					Joint network					Non-network measures		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	GDPCapita	1																								
2	BCI	0.49*	1																							
3	CCI	0.50*	0.77*	1																						
4	GDP	0.35*	0.43*	0.28*	1																					
5	Inflation	0.57*	0.53*	0.60*	0.39*	1																				
6	Interest rates	−0.67*	−0.67*	−0.69*	−0.33*	−0.68*	1																			
7	Population	0.10	0.35*	0.21*	0.93*	0.27*	−0.23*	1																		
8	Betweenness centrality	0.18*	0.28*	0.15*	0.83*	0.25*	−0.19*	0.87*	1																	
9	Degree centrality	0.30*	0.40*	0.31*	0.92*	0.46*	−0.32*	0.88*	0.85*	1																
10	Betweenness oscillation	0.34*	0.41*	0.26*	0.89*	0.42*	−0.30*	0.80*	0.72*	0.86*	1															
11	Closeness centrality	0.01	−0.07	−0.08	0.07	−0.13*	0.07	0.08	0.09	−0.01	0.04	1														
12	Network constraint	−0.30*	−0.29*	−0.14*	−0.32*	−0.36*	0.23*	−0.28*	−0.30*	−0.29*	−0.27*	−0.01	1													
13	Betweenness centrality	0.11	0.24*	0.13*	0.80*	0.25*	−0.17*	0.86*	0.90*	0.84*	0.74*	0.09	−0.11	1												
14	Degree centrality	0.30*	0.38*	0.29*	0.90*	0.48*	−0.32*	0.86*	0.81*	0.98*	0.84*	−0.01	−0.13*	0.84*	1											
15	Betweenness oscillation	0.31*	0.42*	0.26*	0.90*	0.34*	−0.30*	0.84*	0.67*	0.84*	0.83*	0.04	−0.12	0.68*	0.81*	1										
16	Closeness centrality	0.01	−0.08	−0.10	0.08	−0.15*	0.09	0.09	0.07	−0.02	0.07	0.99*	−0.09	0.08	−0.01	0.07	1									
17	Network constraint	−0.12	−0.08	−0.09	−0.16*	−0.21*	−0.03	−0.13*	−0.24*	−0.28*	−0.31*	−0.01	0.40*	−0.02	−0.23*	−0.17*	−0.14*	1								
18	Betweenness centrality	0.10	0.24*	0.13*	0.72*	0.23*	−0.19*	0.78*	0.84*	0.72*	0.58*	0.08	−0.21*	0.95*	0.76*	0.61*	0.10	−0.21*	1							
19	Degree centrality	0.28*	0.39*	0.30*	0.91*	0.46*	−0.32*	0.88*	0.84*	0.99*	0.85*	−0.02	−0.28*	0.85*	0.99*	0.82*	−0.01	−0.17*	0.76*	1						
20	Betweenness oscillation	0.31*	0.42*	0.26*	0.88*	0.37*	−0.31*	0.82*	0.71*	0.83*	0.83*	0.04	−0.28*	0.67*	0.81*	0.95*	0.04	−0.13*	0.60*	0.82*	1					
21	Closeness centrality	0.01	−0.08	−0.09	0.06	−0.14*	0.08	0.07	0.07	−0.02	0.03	0.99*	0.00	0.07	−0.03	0.06	0.99*	−0.11	0.08	−0.03	0.03	1				
22	Network constraint	−0.13*	−0.10	−0.07	−0.17*	−0.23*	0.00	−0.15*	−0.13*	−0.14*	−0.13*	−0.06	0.49*	−0.24*	−0.19*	−0.16*	−0.09	0.98*	−0.22*	−0.19*	−0.14*	−0.08	1			
23	Number of articles	0.21*	0.33*	0.27*	0.80*	0.45*	−0.28*	0.80*	0.80*	0.95*	0.79*	−0.04	−0.24*	0.77*	0.94*	0.72*	−0.04	−0.09	0.68*	0.94*	0.72*	−0.05	−0.11	1		
24	Average tone	0.08	0.07	0.16*	0.01	−0.28*	0.26*	0.01	0.08	−0.12*	−0.10	0.29*	−0.05	0.02	−0.14*	0.01	0.32*	−0.06	0.04	−0.14*	−0.03	0.29*	−0.05	−0.19*	1	
25	Goldstein scale	−0.11*	−0.16*	−0.05	−0.24*	−0.11	0.18*	−0.25*	−0.25*	−0.24*	−0.24*	−0.06	0.15*	−0.23*	−0.24*	−0.24*	−0.05	0.06	−0.20*	−0.24*	−0.24*	−0.05	0.05	−0.22*	0.13*	1

GDP: gross domestic product; GDPCapita: GDP per Capita; BCI: Business Confidence Index; CCI: Consumer Confidence Index.
*$p < 0.05$.

Table 3. Predicting the BCI: Multilevel regression results.

		M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10
Control variables	GDP		0.010								0.018
	Inflation		0.002								0.004
	Interest rates		−0.162*								−0.189*
Interactions network	Betweenness centrality			−3.16							
	Degree centrality			0.007*							
	Betweenness oscillation				0.005*						
	Closeness centrality				−11.1E3						
	Network constraint				0.361						
Location network	Betweenness centrality					−3.913					
	Degree centrality					0.009*					
	Betweenness oscillation						0.005*				0.008*
	Closeness centrality						−13.4E3				
	Network constraint						1.149				
Joint network	Betweenness centrality							6.14			
	Degree centrality							0.008*			
	Betweenness oscillation								0.009*		
	Closeness centrality								−10.9E3		
	Network constraint								1.255		
Non-network measures	Number of articles									0.005	
	Average tone									0.046*	0.068*
	Goldstein scale									−0.065	
	Constant	99.7*	99.7*	99.5*	99.6*	99.5*	99.4*	99.4*	99.3*	100.2*	100.2*
	ICC	46.4%									
	Variance level 2	0.638	0.146	0.487	0.455	0.450	0.490	0.450	0.490	0.557	0.082
	Variance level 1	0.738	0.604	0.733	0.717	0.728	0.711	0.728	0.712	0.720	0.518
	Variance reduction level 2 (%)		77.1	23.7	28.7	29.5	23.2	29.5	23.2	12.7	87.1
	Variance reduction level 1 (%)		18.2	0.7	2.8	1.4	3.7	1.4	3.5	2.4	29.8
	AIC	670.7	614.8	670.4	666.6	667.9	665.3	668.2	665.4	669.8	603.3
	BIC	681.3	636.1	688.1	687.8	685.5	688.1	685.8	686.6	690.9	631.5

AIC: Akaike information criterion; BCI: Business Confidence Index; BIC: Bayesian information criterion; GDP: gross domestic product; ICC: intraclass correlation coefficient.
*$p < 0.05$.

Table 4. Predicting the CCI: Multilevel regression results.

		M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10
Control variables	GDP		0.006								0.004
	Inflation		0.016*								0.021*
	Interest rates		−0.201*								−0.195*
Interactions network	Betweenness centrality			11.73							
	Degree centrality			0.004							
	Betweenness oscillation				0.002*						0.01*
	Closeness centrality				−12.6E3						
	Network constraint				−2.48						
Location network	Betweenness centrality					7.62					
	Degree centrality					0.003					
	Betweenness oscillation						0.009*				
	Closeness centrality						−12.8E3				
	Network constraint						3.08*				
Joint network	Betweenness centrality							15.33			
	Degree centrality							0.006			
	Betweenness oscillation								0.004*		
	Closeness centrality								−12.6E3		
	Network constraint								−3.15*		−2.59*
Non-network measures	Number of articles									0.008	0.047*
	Average tone									−0.142*	
	Goldstein scale									0.032	
	Constant	99.4*	98.3*	99.4*	98.1*	99.2*	98.2*	99.4*	98.1*	99.7*	100.2*
	ICC	39.5%									
	Variance level 2	0.974	0.094	0.906	0.881	0.859	0.906	0.867	0.856	0.961	0.043
	Variance level 1	1.494	1.283	1.487	1.403	1.367	1.487	1.494	1.375	1.366	1.109
	Variance reduction level 2 (%)		90.3	7.0	9.5	11.8	7.0	11.0	12.1	1.3	95.6
	Variance reduction level 1 (%)		14.1	0.5	6.1	8.5	0.5	0.0	8.0	8.6	25.8
	AIC	844.3	763.2	846.5	834.2	827.8	846.5	847.2	829.1	828.7	758.0
	BIC	854.8	784.3	864.1	855.4	848.9	864.1	864.8	850.3	849.8	786.2

CCI: Consumer Confidence Index; GDP: gross domestic product; ICC: intraclass correlation coefficient.

*$p < 0.05$.

Table 5. Predicting the GDPCapita: Multilevel regression results.

		M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10
Control Variables	Inflation		0.024*								0.026*
	Interest rates		0.034								0.049*
	Population		0.002								0.004
Interactions network	Betweenness centrality			9.187							
	Degree centrality			0.008*							
	Betweenness oscillation				0.005*						0.008*
	Closeness centrality				5573.4						
	Network constraint				−3.28*						−1.91*
Location network	Betweenness centrality					18.4					
	Degree centrality					0.006					
	Betweenness oscillation						0.002*				
	Closeness centrality						−942.4				
	Network constraint						−1.75*				−1.16*
Joint network	Betweenness centrality							8.331			
	Degree centrality							0.004			
	Betweenness oscillation								0.007*		
	Closeness centrality								2159.8		
	Network constraint								−1.94*		
Non-network measures	Number of articles									0.002*	0.172*
	Average tone									0.090*	
	Goldstein scale									−.156*	−.151*
	Constant	36.02*	32.8*	35.5*	36.1*	35.6*	36.1*	35.8*	36.3*	35.7*	33.6*
	ICC	37.8%									
	Variance level 2	0.417	0.389	0.400	0.390	0.408	0.398	0.411	0.400	0.400	0.323
	Variance level 1	0.688	0.551	0.653	0.619	0.690	0.638	0.683	0.639	0.639	0.415
	Variance reduction level 2 (%)		6.7	4.1	6.5	2.2	4.6	1.4	4.1	4.1	22.5
	Variance reduction level 1 (%)		19.9	5.1	10.0	−0.3	7.3	0.7	7.1	7.1	39.7
	AIC	695.1	647.2	686.3	675.2	682.9	686.3	697.2	683.0	683.1	593.8
	BIC	705.6	668.4	704.0	696.3	703.5	703.9	714.8	704.1	704.1	632.6

GDP: gross domestic product; GDPCapita: GDP per Capita; ICC: intraclass correlation coefficient.
*p < 0.05.

as business events. In addition, some events might be discussed in a larger number of news whereas other neglected, depending on the strategies adopted by different newspapers or news agencies. It could also happen that a specific event is discussed in some news, after a long time from its first occurrence. Consistently, new data cleaning strategies could be tested to see if our findings can be further improved. We also suggest to test our method on different datasets (considering, e.g., the economies outside the EU). Lastly, we suggest to retest the effects of lagged predictors. This could help avoiding possible biases attributable to reverse causality. In this study, we neither aimed at proving causality nor at discussing the effects of predictors at various lags; we were mostly interested in giving evidence to the potentials of the new presented variables. An investigation of the time taken by economic news and agreements to impact the CCI, BCI and GDP could be the objective of future research. Accordingly, we advocate further studies based on more structured causal designs, which could be helpful for policymakers to understand both the impact of economic news and agreements on macroeconomic indicators and the advantages of maintaining central positions in economic networks.

Authors' note

Mohammed Elshendy and Andrea Fronzetti Colladon contributed equally to this work.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

1. Hamermesh DS. Six decades of top economics publishing: who and how? *J Econ Lit* 2013; 51: 162–172.
2. De Mauro A, Greco M and Grimaldi M. Library review a formal definition of big data based on its essential features. *Libr Rev J* 2016; 65: 122–135.
3. Dutton G. Five steps to building a data-driven culture [internet]. *Forbes*, http://www.forbes.com/sites/emc/2014/06/06/5-steps-to-a-data-driven-culture/#6a042f243d52 (2014, accessed 20 August 2016).
4. Bizer C, Boncz P, Brodie ML, et al. The meaningful use of big data: four perspectives – four challenges. *SIGMOD Rec* 2011; 40(4): 56–60.
5. Einav L and Levin J. The data revolution and economic analysis. *Innov Policy Econ* 2014; 14: 1–24.
6. Gupta V and Lehal GS. A survey of text mining techniques and applications. *Journal of Emerging Technologies in Web Intelligence* 2009; 1: 60–76.
7. Gloor PA, Fronzetti Colladon A, Grippa F, et al. Forecasting managerial turnover through e-mail based social network analysis. *Comput Hum Behav* 2017; 71: 343–352.
8. Allen TJ, Gloor PA, Fronzetti Colladon A, et al. The power of reciprocal knowledge sharing relationships for startup success. *J Small Bus Enterp Dev* 2016; 23(3): 636–651.
9. Fronzetti Colladon A and Remondi E. Using social network analysis to prevent money laundering. *Expert Syst Appl* 2017; 67: 49–58.
10. Shipilov AV and Li SX. Can you have your cake and eat it too? Structural holes' influence on status accumulation and market performance in collaborative networks. *Adm Sci Q* 2008; 53: 73–108.
11. Borgatti SP, Mehra A, Brass DJ, et al. Network analysis in the social sciences. *Science* 2009; 323: 892–896.
12. Newman MEJ. The structure and function of complex networks. *Siam Rev* 2003; 45: 167–256.
13. Clements MP and Smith J. The performance of alternative forecasting methods for SETAR models. *Int J Forecast* 1996; 13: 463–475.
14. Forni M, Hallin M, Lippi M, et al. Do financial variables help forecasting inflation and real activity in the euro area? *J Monet Econ* 2003; 50: 1243–1255.
15. Artis MJ, Banerjee A and Marcellino M. Factor forecasts for the UK. *J Forecasting* 2005; 24(4):279–298.
16. Gupta R and Kabundi A. Forecasting macroeconomic variables in a small open economy: a comparison between small-and large-scale models. *J Forecasting* 2010; 29(1-2): 168–185.
17. Doms ME and Morin NJ. Consumer sentiment, the economy, and the news media. *FRB San Fr Work Pap* 2004; 9: 1–70.
18. Sims CA. Macroeconomics and reality. *Econometrica* 1980; 48: 1–48.
19. Litterman RB. Forecasting with Bayesian vector autoregressions – five years of experience. *J Bus Econ Stat* 1986; 4: 25–38.
20. King R, Plosser CI, Stock JH, et al. Stochastic trends and economic fluctuations. *Am Econ Rev* 1991; 4: 819–840.
21. Gupta R and Schaling E. Forecasting the South African economy: a DSGE-VAR approach. *J Econ Lit* 2007; 37: 181–195.
22. Kuzin V, Marcellino M and Schumacher C. MIDAS vs. mixed-frequency VAR: nowcasting GDP in the Euro area. *Int J Forecast* 2011; 27: 529–542.
23. Cuaresma JC. Forecasting European GDP using self-exciting threshold autoregressive models: a warning. *IHS Economic Series* 2000; 79. http://europa.eu/rapid/press-release_MEMO-17-1339_en.htm.
24. Feng H and Liu J. A SETAR model for Canadian GDP: nonlinearities and forecast comparisons. *Appl Econ* 2003; 35(18): 1957–1964.
25. Rünstler G and Sédillot F. Short-term estimates of Euro area real GDP by means of monthly data. *ECB Work Pap* 2003; 276: 1–33.
26. Baffigi A, Golinelli R and Parigi G. Bridge models to forecast the euro area GDP. *Int J Forecast* 2004; 20: 447–460.
27. Diron M. Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data. *J Forecast* 2008; 27: 371–390.
28. Angelini E, Camba-Mendez G, Giannone D, et al. Short-term forecasts of Euro area GDP growth. *Econom J* 2011; 14: C25–C44.

29. Stock JH and Watson MW. Macroeconomic forecasting using diffusion indexes. *J Bus Econ Stat* 2002; 20: 147–162.
30. Giannone D, Reichlin L and Small D. Nowcasting: the real-time informational content of macroeconomic data. *J Monet Econ* 2008; 55: 665–676.
31. Schumacher C. Forecasting German GDP using alternative factor models based on large datasets. *J Forecast* 2007; 26: 271–302.
32. Borgatti SP, Everett MG and Johnson JC. *Analyzing social networks*. New York: Sage, 2013, p. 304.
33. Jackson MO. *Social and economic networks*. Princetown, NJ: Princetown University Press, 2008, p. 506.
34. Mizruchi MS and Stearns LB. Getting deals done: the use of social networks in bank decision-making. *Am Sociol Rev* 2001; 66: 647–671.
35. Levin DZ and Abrams LC. The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer. *Acad Manag Proc Membsh Dir* 2002; 8: D1–D6.
36. Kavanaugh AL, Reese DD, Carroll JM, et al. Weak ties in networked communities. *Inf Soc* 2005; 21: 119–131.
37. Hauser C, Tappeiner G and Walde J. The learning region: the impact of social capital and weak ties on innovation. *Reg Stud* 2007; 41: 75–88.
38. Granovetter MS. The impact of social structure on economic outcomes. *J Econ Perspect* 2005; 19: 33–50.
39. Hidalgo CA, Klinger B, Barabási AL, et al. The product space conditions the development of nations. *Science* 2007; 317(5837): 482–487.
40. Garlaschelli D, Di Matteo T, Aste T, et al. Interplay between topology and dynamics in the World Trade Web. *Eur Phys J B* 2007; 57: 159–164.
41. Kwak H and An J. Two tales of the World: Comparison of widely used world news datasets GDELT and event registry. *arXiv preprint arXiv*:1603.01979. 2016 Mar 7. https://arxiv.org/abs/1603.01979.
42. Leetaru K and Schrodt PA. GDELT: global data on events, location and tone, 1979-2012. Paper presented at the *Annual Meeting of the International Studies Association*, San Francisco, CA, 2013, pp. 1–49.
43. OECD. Consumer confidence index (CCI) (indicator). 2017. DOI: 10.1787/46434d78-en.
44. OECD. Business confidence index (BCI) (indicator). 2017. DOI: 10.1787/3092dc4f-en.
45. Wasserman S and Faust K. *Social network analysis: methods and applications*. New York: Cambridge University Press, 1994, p. 825.
46. Brandes U and Erlebach T. *Network analysis: methodological foundations*. Berlin, Germany: Springer-Verlag Berlin Heidelberg, 2005, p. 472.
47. Freeman LC. Centrality in social networks conceptual clarification. *Soc Networks* 1978; 1: 215–239.
48. Borgatti SP. Centrality and network flow. *Soc Networks* 2005; 27: 55–71.
49. Kay E, Bondy JA and Murty USR. *Graph theory with applications*. London: The Macmillan Press Ltd, 1976, p. 270.
50. Everett MG, Everett M and Borgatti SP. Ego network betweenness. *Soc Networks* 2005; 27: 31–38.
51. Kidane YH and Gloor PA. Correlating temporal communication patterns of the Eclipse open source community with performance and creativity. *Comput Math Organ Theory* 2007; 13: 17–27.
52. Davis JP and Eisenhardt KM. Rotating leadership and collaborative innovation: recombination processes in symbiotic relationships. *Adm Sci Q* 2011; 56(2): 159–201.
53. Burt RS. Social structure of competition. *Explorations in Economic Sociology* 1993; 19: 65–103.
54. Burt RS. Structural holes and good ideas. *Am J Sociol* 2004; 110(2): 349–399.
55. Burt RS. *Structural holes: the social structure of competition*. Cambridge: Harvard University Press, 1995, p. 313.
56. Goldstein JS. A conflict-cooperation scale for WEIS events data. *J Conflict Resolut* 1992; 36: 369–385.
57. Berardi A. Term structure, inflation, and real activity. *J Financ Quant Anal* 2009; 44(4): 987–1011.
58. Garner CA. Consumer confidence after September 11. *Econ Rev Reserv Bank Kansas City* 2002; 87: 5–26.
59. Cologni A, Manera M, Lavoro NDI, et al. Oil prices, inflation and interest rates in a structural cointegrated VAR model for the G-7 countries. *Energy Econ* 2005; 30: 856–888.
60. Taylor K and McNabb R. Business cycles and the role of confidence: evidence for Europe. *Oxf Bull Econ Stat* 2007; 69: 185–208.
61. Krznar I and Kunovac D. Impact of external shocks on domestic inflation and GDP. *CNB Occasional Pulications Working Papers* 2010; 26: 1–24.
62. Matsusaka JG and Sbordone AM. Consumer confidence and economic fluctuations. *Econ Inq* 1995; 33: 296–318.
63. Borio C. The financial cycle and macroeconomics: what have we learnt? *J Bank Financ* 2014; 45: 182–198.
64. Jaimovich N and Rebelo S. Can news about the future drive the business cycle? *Am Econ Rev* 2009; 99(4): 1097–1118.
65. Gleditsch KS. Expanded trade and GDP data. *J Conflict Resolut* 2002; 46(5): 712–724.
66. Nezlek JB. An introduction to multilevel modeling for social and personality psychology. *Soc Personal Psychol Compass* 2008; 2(2): 842–860.
67. Hoffman L and Rovine MJ. Multilevel models for the experimental psychologist: foundations and illustrative examples. *Behav Res Methods* 2007; 39(1): 101–117.
68. Snijders TAB and Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. London: Sage, 1999, p. 272.
69. Claveria O, Pons E and Ramos R. Business and consumer expectations and macroeconomic forecasts. *Int J Forecasting* 2007; 23(1): 47–69.
70. Mourougane A and Roma M. Can confidence indicators be useful to predict short term real GDP growth? *Appl Econ Lett* 2003; 10(8): 519–522.
71. Elshendy M, Fronzetti Colladon A, Battistoni E, et al. Using four different online media sources to forecast the crude oil price. *J Inf Sci* 2017; doi:10.1177/0165551517698298.
72. Ferraro G and Iovanella A. Revealing correlations between structure and innovation attitude in inter-organisational innovation networks. *Int J Comput Econ Econometrics* 2016; 6(1): 93–113.