The Age of Big Data: data mining and communication
9 November 2017

Speaker: Pierre Deville (Head of Analytics, Bisnode Group Analytics, Belgium)

Moderator: William Echikson (Associate Senior Fellow and Director of the Digital Forum, Centre for European Policy Studies, Belgium)

As data becomes more ubiquitous and easy to collect, including via social media, we are faced with a "tsunami" of potential data points. With the explosion of big data, the key is to take advantage of innovative opportunities to put raw data to use in actionable ways. This is where data mining, the automated extraction of hidden predictive information from large datasets, comes into play as a powerful tool for predicting trends and behaviours, allowing for knowledge-driven decisions on engaging audiences across channels. This session explored the potential of data mining for optimising communication and dug deeper into the many ways data mining can have an impact.

The moderator, **William Echikson**, started the session by noting the need for a more comprehensive policy and regulatory framework in Europe when dealing with big data. The explosion of communication possibilities and the high level of technological innovation have led to fundamental policy challenges in Europe and for regulators. We therefore need to ask how we can exploit big data and the exponential capabilities of information technologies, while minimising their potential risks. Mr Echikson raised the issue of free speech and whether policy-makers should be able to narrow it, for instance with the goal of protecting privacy or public order.

The speaker, **Pierre Deville**, then introduced the topic by asking what big data means, and where it comes from. By way of an answer, he quoted Dan Ariely, professor of psychology and behavioural economy at Duke University: Big data is like teenage sex; everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it. The question is what big data really means and how it can impact communication.

Mr Deville then introduced the big data phenomenon and its meaning. He first pointed out that there is an exponential increase in the technological capabilities for collecting (storage) and processing (computation) data, hence a "data explosion". For instance, at the beginning of 2000, the MIT Sloan Digital Sky Survey telescope managed to map 25% of the visible sky in 15 years, while today the same project can scan 100% of the sky in 3 days. Another example concerned social data: every daily activity produces around 3 exabytes of data, corresponding to the total amount of data produced until 2003. With this increasing amount of available data, there is a substantial need for new innovative approaches on how to process them.

Mr Deville then explained the value of big data for communicators and the various uses that could be made of big data. Big data analysis makes it possible to extract and derive knowledge through data mining. This activity refers to the design and building of efficient

algorithms, which are able to identify and extract patterns or make predictions on various topics. To put it simply, what matters is not the data itself but the knowledge that can be extracted through data mining. To illustrate this, Mr Deville used the example of AlphaGo, designed by Google and used to play the Chinese game of Go; Facebook ads, targeting advertisements according to the social data it collects from the social network, and IBM Watson, a computer system relying on semantic analysis and designed to answer questions through language processing and analysing.

Mr Deville then spoke about the new possibilities linked to this exponential increase in available data, and the challenges arising for communicators. He pointed first to the multiplication of massive decentralised democratic information sources. In other words, nowadays everyone can become a source of information through social media, blogs or comments, which can lead to four major applications of data mining for communication.

First of all, data mining would make it possible to identify key opinion leaders in various social networks and to further engage with them. Targeting those opinion leaders could then help to maximise the propagation of specific information. Mr Deville quoted the case of the recent US presidential elections where, by sending simple signals to the epicentre of a network, communicators would be able to influence not only one person's voting behaviour, but also that of the people around them.

Mr Deville also argued that data mining could make it possible to tailor communication strategies towards certain segments of the population more effectively. In that regard, communicators could rely on social data, available thanks to social media, but also on more "scientific information" such as demographic and socio-economic data.

A third application identified was the detection of threats and social phenomena through decentralised new sources. By monitoring social media, agencies or local authorities could pinpoint abnormal behaviour and prevent hazards such as terrorist attacks, and to some extent riots and civil unrest.

Finally, given this huge amount of data, Mr Deville identified a fourth application: when comparing the information available on a similar topic, it becomes possible to filter this information and decide what to address and what to ignore. Studies have been able to identify bias in media and thus spot fake news accurately.

However, beyond these applications, Mr Deville identified a few challenges associated with data manipulation. First of all, there is a new paradox between the amount of data that is available, and the veracity of these data. It is getting harder and harder to figure out which information is correct (as opposed to fake), which leads to problems of accuracy when designing your prediction models.

The second challenge identified is an ethical problem, hence the need for a consolidated legal framework. A banking organisation could use prediction models to take decisions on whether a person should receive money or not. This raises the question of how far should

this model go, and what data to use without leading to discrimination based on socio-economic factors or personal details.

Finally, the last challenge identified by Mr Deville was the question of education and qualification. It is increasingly challenging to find staff able to process big data efficiently. Consequently, the age of big data should open the door to more specialised education.

Mr Echikson then opened the floor, and participants asked about the accurate identification of fake news, and the dangers and opportunities of artificial intelligence. A first set of questions referred to the way data scientists could address fake news. Mr Deville argued first that we need to distinguish between detecting and addressing fake news. He then proposed a method of identification: some news spreads very quickly, but only within a specific bubble, while more "reliable" news would spread everywhere. According to Mr Echikson, this also raised the issue of the accountability and liability of platforms used to broadcast fake news or disturbing content, reinforcing the need for more detailed legislation.

This led to a question on addressing inaccuracy when people are not sharing honest information, and **Mr Deville** suggested that in such cases, we have to collaborate with various actors to try to collect as much information on a given topic as possible, in order to reduce the error margin.

Another set of questions touched on the potential challenges associated with the rise of artificial intelligence. According to both Mr Deville and Mr Echikson, AIs should be seen as carriers of opportunities, provided they are developed by the right people (such as universities), under an appropriate legal framework.

Finally, participants asked about ownership of large datasets by private companies and what role could be played by governments. Mr Deville argued that the current EU legal framework requires that companies disclose the type of data they are using when designing their marketing or communication strategies. Similarly, governments own a great deal of information, and data analysis often favours public institutions rather than private ones.