Predictors of organized crime and subversion: A machine learning approach[1]

Patricia Prüfer and Pradeep Kumar[2]

4 February 2018

FOR CONFERENCE PURPOSES ONLY – PLEASE DO NOT CITE OR DISTRIBUTE

Abstract: The share of offences that can be related to organized crime and subversive activities is increasing. Organized crime engages in systematic violation of the law with serious effects on society and is able to cover these violations in a very effective way. The same is true for subversion, which can be seen as "the paralysis of the (regional) society". *Notwithstanding the serious negative effects of these activities, it is difficult to combat them as they are performed illegally and* 'under the shadow of the law'. *What if one can disentangle patterns and methods of organized crime and subversion? We address the quest for indicators that can predict the occurrence of criminal and subversive activities by using a wide range of data sources and a broad combination of data science techniques. Based on patterns and best predictors, we develop a method with which policy makers can judge the level of criminal and subversive activities taking place in a certain area.*

Introduction

Although crime rates in general as well as the level of perceived crime are decreasing, the share of offences that can be related to organized crime and subversive activities is increasing in the Netherlands (Boerman et al. 2017). Organized crime is hereby related to "a form of criminal activity in which groups that focus primarily on illegal profits engage in systematic violation of the law with serious effects on society, and are able to cover these violations in a very effective way, especially by their willingness to use physical violence or corruption to deactivate people" (Fijnaut, 1996: 24/25). Closely related to organized crime is subversion or subversive crime with the following facets: infraction of authority of (regional) government and police; infiltration of social and welfare institutions, subtle societal acceptation of violations of the law (often via monetary transfers), impairment of market forces, and imputation of institutions focusing on legal activities (Tops and Van de Torre, 2014). In other words, subversive crime can be seen as "the paralysis of the (regional) society and economy by violating the law, intimidating, bribing, or blackmailing people, firms, politicians, and the civil service" (Studiecentrum voor Bedrijf en Overheid, 2017: 19).

[1] This research is funded by the Dutch Ministry of Justice and Security under the grant BD/WODC/EWB 2836.
[2] CentERdata, Tilburg University.

In the Netherlands, the venues where organized criminal activities concentrate and from where subversive activities spread can increasingly be found on industrial areas (Kolthoff and Khonraad, 2016). Therefore, it is important to get deeper insights on what is going on in industrials areas, and to uncover patterns and methods of organized crime operating from industrial areas. This is especially relevant as also the means of subversion used by (organized) criminal organizations to influence legal institutions and the society in general, have gained in subtleness and variety (Tops and Van der Torre, 2014).

Although the relevance of industrial areas for organized crime and subversion is undisputed, so far no instrument or method exists with which one can measure the level of criminal activities and subversion in a specific industrial area. The problem is that a lot of these activities are performed illegally and 'under the shadow of the law'. Besides direct hurts to victims, organized crime and subversion also have detrimental systemic effects, for example by eroding trust in a given society and/or economy. Moreover, they can lead to a neutralization of social control, which means that criminals engaged in these activities are very well capable of fencing off themselves and their business. In some cases, this can lead to an imagination of invulnerability of offenders and can make them a living proof of the incapability of the police and judiciary to take on organized crime and subversion. Ultimately, this image-forming can become a self-fulfilling prophecy (Kruisbergen et al., 2012).

What we do in this paper

In this paper, we address the quest for indicators that can predict the occurrence of criminal and subversive activities by using a wide range of data sources and a broad combination of data science techniques. By doing so, we can disentangle patterns of organized crime and subversion on Dutch industrial areas. Based on these patterns and the best predictors, we develop a method with which policy makers can judge the level of criminal and subversive activities taking place in a certain area. This method provides an instrument, think of a 'sliding scale' or an 'organized crime hot spot' providing insights in the state of a certain industrial area: (i) no organized crime and subversion are happening; (ii) a slight amount of organized crime and subversion are happening; or (iii) a high amount of organized crime and subversion are happening.

Our research is related to a couple of other examples of data science applications for data-driven policy making in the Netherlands. The national police has implemented a Dutch version of predictive policing (Criminal Anticipation System, *Criminaliteit Anticipatie Systeem*, CAS) where an algorithm is predicting the occurrence of high impact crime in a certain district based on more than 250 predictors. Also some municipalities are experimenting with data science approaches, for example to improve the environment or reduce burglary based on various sensor data.

In this paper, we relate to the ongoing digitalization and datafication in various ways. On the one hand, data are getting more and more important and available and also (governmental) institutions are starting to grasp the importance of working more data-driven and basing insights and political decisions more on empirical evidence. On the other hand, this also affects the nature of criminal activities that can occur, such as cybercrime, and the type of criminal activities that can be measured, for example by combing very different data sources or by harvesting the Darknet.

Besides the data that we scrape from the Internet (and eventually the Darknet), we use all types of available data and information related to different aspects of organized crime and subversion. For example (marketing) information from a suspicious person or firm; all sorts of files, for example on 'clients'; annual reports and balance sheets; tax information; and registries from the Chamber of Commerce. In theory, also social media data and the browse/search behavior of a suspicious person or firm can be analyzed with respect to interesting trends and patterns. A search engine usually registers when and where a certain query has taken place and use this information in the search results provided to this specific query. Therefore, the amount of queries being done for a certain search term and the ranking of search results presented for these queries can be very informative. To mine, cluster, and analyze these structured as well as unstructured data sources, we mainly use techniques related to natural language processing.[3]

We can access these 'big data'[4] sources in collaboration with the Dutch Ministry of Justice and Security, the Dutch national police, the national tax authority, and various other providers of relevant data. By collaborating with companies and managers of some large industrial areas, we can also analyze information from their wireless networks in a specific area, sensor data on movements and lightning (in an area as well as in buildings), GPS data, for example from trucks, smartphone and app data, and data from CCTV's. All available data are analyzed with machine and deep learning techniques. The principle of these techniques is to recognize and use existing patterns in (complex) data, and in our paper we are interested in finding patterns on the existence, level and migration of organized crime and subversion.

Machine and deep learning techniques cannot only be used for 'big data', but also for the traditional (administrative) data sources we use in our research (Prüfer and Prüfer, 2018). These recent techniques can also find new patterns in those kind of data, for example by analyzing answers to surveys that were conducted among victims of crime or to assess the level of security. Based on all the patterns found, we can derive the best predictors for organized crime and subversion on Dutch industrial areas. This provides a method and an instrument with which policy makers can judge the level of criminal and subversive activities that take place in a certain area.

[3] For example topic modelling, lemmatization, tagging, word2vec word-space alignment, parsing, term-frequency determination and ontological relationship extraction.

[4] For data to be classified as 'big', a couple of features have to apply. The most important ones are: volume (amount) of data that is being produced every second; velocity (speed) with which new data are generated, collected, and analyzed; variety (types) of data that are being used.

References

Boerman, F., Grapendaal, M., Nieuwenhuis, F. and Stoffers, E. (2017), *Nationaal dreigingsbeeld 2017: Georganiseerde criminaliteit*, Dienst Landelijke Informatieorganisatie, Politie, Driebergen.

Fijnaut, C. (1996), *Eindrapport Georganiseerde criminaliteit in Nederland*, Tweede Kamer, vergaderjaar 1995–1996, 24 072, nr. 16, The Hague.

Kolthoff, E., Khonraad, S. (2016), Ondermijnende aspecten van georganiseerde criminaliteit en de rol van de bovenwereld, *Tijdschrift voor Criminologie* 58(2),76-90.

Kruisbergen, E., Van de Bunt, H. and Kleemans, E. (2012), Georganiseerde criminaliteit in Nederland, *Vierde rapportage op basis van de Monitor Georganiseerde Criminaliteit*, WODC.

Prüfer, J., Prüfer, P. (2018), Data Science for Institutional and Organizational Economics, TILEC DP No. 2018-011, Tilburg University.

Tops, P., Van der Torre, E. (2014), *Wijkenaanpak en Ondermijnende Criminaliteit*, Tilburg University, Tilburg.

Studiecentrum voor Bedrijf en Overheid (2017), *Ondermijning & Georganiseerde Criminaliteit* 2017, e-book, Eindhoven.