
1. INTRODUCTION

A main objective of our research is to obtain a description of the socio-economic environment characterizing a trial leading to a criminal sentence, as well as the differences in the conduct of the trial between different jurisdictional administrations.

In this paper we study the juridical response to organized crime activities in Sicily, Italy, by analyzing a corpus of criminal sentences using computational techniques of information extraction and social network analysis. In particular, the analyzed criminal sentences were pronounced in the four courthouses of Sicily from 2000 through 2006, and were declared irrevocable for at least one defendant. A first element of originality of this paper is in the choice of the source of sentences: criminal sentences are tools of social regulation (although as outcome of a trial), and are used as a “magnifying glass” to study mechanisms and contexts from which one can deduce the

* D. De Felice and G. Giuffrida are assistant professors at the Department of Political and social sciences of the University of Catania (Italy); G. Giura, PhD in Sociology and methods of social sciences, works at the Department of European studies and international integration of the University of Palermo (Italy); V. Verendel has a PhD in the Complex systems group at the Chalmers University of Technology in Gothenburg (Sweden); C. Zarba, PhD, works for Neodata Intelligence S.r.l.

real balances of power and value hierarchies established by the institutional response to organized crime and mafia-related crimes. A second element of originality is the synergy, pursued since the phase of data retrieval, between computer scientists and social scientists when it comes to methodology and the approach of the research team.

By using direct (expert-based) knowledge of the semantic importance of the source of sentences, we made a preliminary coding and performed statistical exploration of a corpus of sentences. To the same corpus, we designed and applied automatic textual extraction algorithms trained using computational machine learning techniques. A result of using this approach, especially compared to the source of judiciary statistics, has the following advantages: a larger variety of the information material contained in the sentences, greater reliability of the classification of the crime categories dealt by the sentences, greater versatility of the information to the cognitive requirements of the researcher rather than that of the administrative monitoring requirements of the judiciary machine.

We use all of these advantages and report here on our results of information extraction and social network analysis between actors (pairs of actors that repeatedly co-occur in the text). An outcome of our result is the study of a social network with several familiar properties, which suggests that the combination of information extraction and social network analysis may provide novel insights on crime as a social phenomenon.

In this paper we describe the first example in Italy of a large digital archive of criminal sentences and demonstrate how a computational structural analysis can be performed on the corpus of data.

1.1. Criminal Sentences from a Sociological Perspective

In the analysis of the jurisprudence a sentence is the result of a reconstruction process of reality that depends on the purpose for which such reconstruction is made (that is to judge), and on the (juridic) technique used in order to arrive to the qualification of the facts.

The judge must evaluate using the current law, and therefore his action is limited to the possibilities contemplated by the law. This inevitably implies, on the one hand, that the qualificatory process influences the way that the reality is reconstructed in the sentence and, on the other hand, that the outcome of the sentence will depend upon what the judge will deem significant for the application of the relevant norm to the concrete case. Moreover,