This is the accepted version of:
Calderoni, Francesco, and Carlo Piccardi. 2014. "Uncovering the Structure of Criminal Organizations by Community Analysis: The Infinito Network." In 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems (SITIS), 301–8. Marrakech: IEEE Computer Society Press. doi:10.1109/SITIS.2014.20.
http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=7081563&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D7081563

Uncovering the structure of criminal organizations by community analysis: the Infinito network

Francesco Calderoni
Università Cattolica del Sacro Cuore
and Transcrime
Milan, Italy
e-mail francesco.calderoni@unicatt.it

Carlo Piccardi
DEIB, Politecnico di Milano
Milan, Italy
e-mail carlo.piccardi@polimi.it

***Abstract*—Criminal organizations tend to be clustered to reduce risks of detection and information leaks. Yet, the literature has so far neglected to explore the relevance of subgroups for their internal structure. The paper focuses on a case study drawing from a large law enforcement operation (”Operazione Infinito”). It applies methods of community analysis to explore the structure of a 'Ndrangheta (a mafia from Calabria, a southern Italian region) network representing the individuals' co-participation in meetings. The results show that the network is significantly clustered and that communities are partially associated with the internal organization of the 'Ndrangheta into different *locali* (similar to mafia families). The implications of these findings on the interpretation of the structure and functioning of the criminal network are discussed.**

I. Introduction

Academics and law enforcement agencies are increasingly applying network analysis to organized crime networks. Yet, the current applications mainly focus on the identification of the key criminals through centrality measures [1]. The analysis of the subgroups and their influence on the criminal activities received very limited attention so far.

Subgroups are a natural occurrence in criminal networks. Criminal organizations may structure themselves in functional, ethnic, or hierarchical units. Furthermore, the constraints of illegality limit information sharing to prevent leaks and detection, as criminal groups face a specific efficiency vs. security trade-off [2]. This makes criminal organizations naturally sparse, clustered networks, often showing both scale-free and small-world properties [3]. Also, the larger the criminal organization, the most likely and relevant is the presence of subgroups. These considerations suggest that the analysis of subgroups in criminal networks may provide insight on both the internal structure of large organized crime groups and on the best preventing and repressive strategies against them.

The mafias are a clear example of large organized crime groups, often comprising several families or clans with a specific hierarchy and a strong cohesion. These units may show different interactions among them, ranging from open conflict to pacific cooperation. Each mafia family is a subgroup within a larger criminal network, and inter-family dynamics are determinant for the activities of the mafias. Nevertheless, possibly due to the difficulties in gathering reliable data, the literature has so far neglected the role of the family in the structure and the activities of the mafias.

In the literature of network analysis (e.g., [4]–[6]), one of the most challenging areas of investigation in recent years is *community analysis*, which is aimed at revealing possible subnetworks (i.e., groups of nodes called communities, or clusters, or modules) characterized by comparatively large internal connectivity, namely whose nodes tend to connect much more with the other nodes of the group than with the rest of the network. A huge number of contributions have explored the theoretical aspects of community analysis, and proposed a broad set of algorithms for community detection [7]. Most notably, community analysis has revealed to be a powerful tool for deeply understanding the properties of a number of real-world complex systems in virtually any field of science, including biology [8], ecology [9], economics [10], information [11], [12] and social sciences [13], [14].

This paper aims to apply the methods of community analysis to mafia families in a network describing the co-participation in the meetings of a large criminal organization. The exercise aims to explore the relevance of subgroups in criminal networks in general. The case study draws data from a large law enforcement operation in Italy (”Operazione Infinito”), which arrested more than 150 people and concerned the establishment of several 'Ndrangheta (a mafia from Calabria, a southern Italian region) groups in the area around Milan, the capital city of the Lombardy region and Italy's ”economic capital” and second largest city. The exploration may have a double relevance. First, it may improve the understanding of the internal functioning of criminal organizations, demonstrating whether the Infinito network is clustered in subgroups, and whether the subgroups identified by community analysis overlap with the internal organization of the 'Ndrangheta. Second, it may contribute in the development of law enforcement intelligence capacities, providing tools for early identification of the internal structure of a criminal group.

The internal organization of the 'Ndrangheta provides an interesting opportunity to explore the relevance of subgroups in criminal networks. Indeed, this mafia revolves around the blood family [15], [16]. One or several 'Ndrangheta families, frequently connected by marriages, godfathering and similar social ties, form a *'ndrina*. The *'ndrine* from the same area may form a *locale*, which controls a specific territory [17]. The *locale* is the main structural unit of the 'Ndrangheta. Each *locale* has a number of formal charges, tasked with specific functions: the boss of the *locale* is the *capobastone* or *capolocale*, the *contabile* (accountant) is responsible for the common fund of the locale, the *crimine* (crime) oversees violent actions, and the *mastro di giornata* (literally "master of the day") takes care of the communication flows within the *locale*.

Since the organization in *locali* plays such an important role in the structure of the 'Ndrangheta, our investigation is specifically oriented to assess their significance in the sense of community analysis. Therefore, after illustrating some details on the network data (Sec. II), we first quantify the cohesiveness of each *locale* in the Infinito network (Sec. III), discovering a quite diversified picture where very cohesive *locali* coexist with others apparently not so significant. We then use two different approaches to the community analysis (Sec. IV): the results show that the Infinito network is significantly clustered, suggesting that subgroups play an important role in its internal organization. If we try and match the clusters obtained by community analysis with the *locali* composition, we interestingly discover that in most cases clusters correspond either to *locali*, or to subsets of them (with a precise hierarchical structure), or to unions of them. It must be acknowledged that the limitations of the data sources inevitably affect the quality of the results: for several *locali* the investigation identified only few members, and a number of individuals remained unidentified. In Sec. V we address this issue and replicate part of the analysis by excluding the smallest *locali* and those individuals with unknown affiliation. The results largely improve, denoting that higher accuracy in gathering data could significantly enhance the capability of describing the structure of the criminal organization with the tools of network analysis.

II. Data

A. *The Infinito network*

The most important source for this study was the pretrial detention order issued by the preliminary investigation judge ("Giudice per le indagini preliminari") of Milan upon request by the prosecution [18]. Most of the investigation, from background checks to wiretaps and surveillance activities, focused on describing the organizational structure of the 'Ndrangheta with a particular care in charting the Lombardy hierarchy and the different *locali* existing in the region.

The documentation of "Operazione Infinito" provides information on a large number of meetings among members of the 'Ndrangheta criminal organization in Lombardy. The meetings occurred in private (e.g., houses, cars) or public places (e.g. bars, restaurants or parks). The two sets of the meetings and of the participants define a standard bipartite (or two-mode) network [6]. The projection of the bipartite network onto the set of 256 participants leads to a (one-mode) weighted, undirected network, whose giant component – which we will denote hereafter as the *Infinito network* – has $N = 254$ nodes and $L = 2132$ links (the density is $\rho = 2L/((N(N-1))) = 0.066$). The weight w_{ij} is the number of meetings between participants i and j, and ranges from 1 to 115. However, the mean value of the (nonzero) weights is $\langle w_{ij} \rangle = 1.88$ and about 70% of them is 1, denoting that only very few pairs of individuals co-attended a large number of meetings. Similarly, the distributions of the nodes degree k_i and strength $s_i = \sum_j w_{ij}$ display a quite strong heterogeneity (see Fig. 1): indeed, their average values are, respectively, $\langle k_i \rangle = 16.8$ and $\langle s_i \rangle = 31.5$, but the most represented individual in the sample has both degree and strength equal to 1.

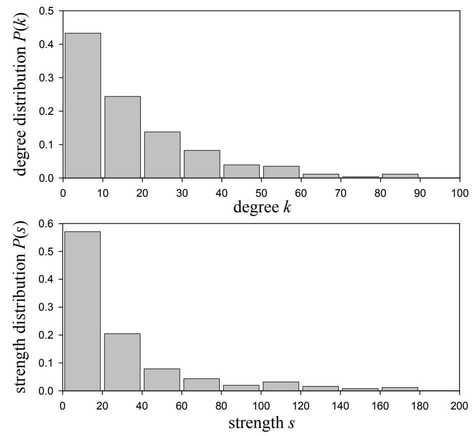

Fig. 1. Histograms of the degree distribution (above) and strength distribution (below) of the Infinito network.

B. *The* locali *partition*

As already pointed out, one of the main features of the 'Ndrangheta is its structural organization in *locali*, namely groups of individuals who control the criminal activities in a specific territory. The affiliation of an individual to a *locale* is formal, and each *locale* has a boss who is responsible of all the activities in front of the higher hierarchical levels (see [1] for further details).

"Operazione Infinito" was able to classify the *locale* membership of most of the individuals tracked during the investigation. Specifically, it was possible to associate 177 individuals (over 254) to one of the 17 *locali* identified in Milan area, the region under investigation. Of the remaining individuals who participate in the meetings, 35 were known to belong to *locali* based in Calabria (the region of Southern Italy where the 'Ndrangheta had origin), 3 came from a Lombardy *locale* not in the area of investigation (Brescia), and 8 were known to be non affiliated to 'Ndrangheta, whereas the correct classification of the remaining 31 remained undefined.

The network is displayed in Fig. 2. The figure, produced by Pajek [19], has a layout obtained by Kamada-Kawai algorithm [20], which implements an attraction/repulsion mechanism among nodes and tends to spontaneously highlight subgraphs which are significantly cohesive, as it will be discussed in the next section.

III. Testing the significance of the *Locali* partition

We firstly want to assess whether the partition of the individuals defined by their membership to a *locale* is significant in the sense of community analysis, namely whether the intensity of intra-*locale* meetings is significantly larger than that of the contacts among members of different *locali*. If so, this would confirm, on one hand, the modular structure of the crime organization; on the other hand, it would provide a

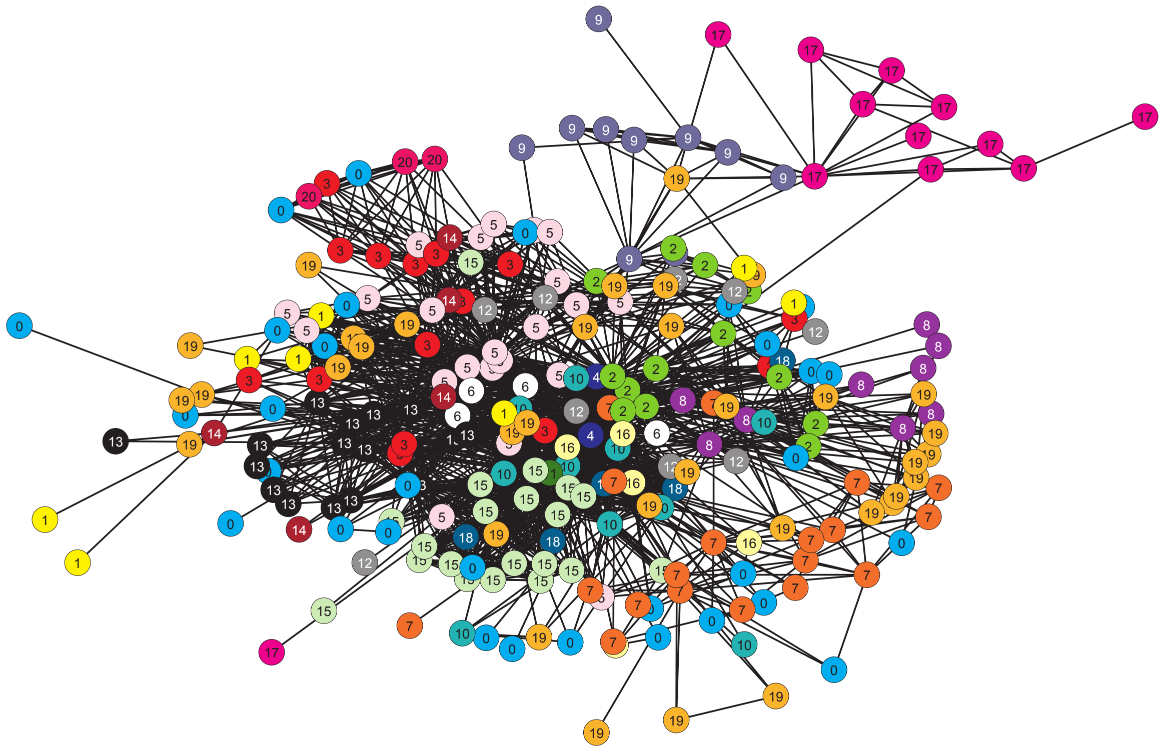

Fig. 2. The Infinito network. Labels (and colors) refer to the *locali* partition, see Table I.

tool for investigations, as the composition of the *locali* could endogenously be derived by mining meetings data.

Let us denote by C_k the subgraph induced by the nodes belonging to *locale* k. We quantify the cohesiveness of C_k by the *persistence probability* α_k, which proved to be an effective tool for the structural analysis of networks [21], [22]. In an undirected network it takes the form:

$$\alpha_k = \frac{\sum_{i \in C_k} \sum_{j \in C_k} w_{ij}}{\sum_{i \in C_k} \sum_{j \in \{1,2,\ldots,N\}} w_{ij}}. \tag{1}$$

The quantity α_k is the fraction of the total weight of the links connected to the nodes of C_k that remains within C_k. Radicchi et al. [23] defined *community* a subnetwork when $\alpha_k > 0.5$. Obviously, the larger α_k, the larger is the internal cohesiveness of C_k. It can be proved that α_k is the probability that a random walker, which is in one of the nodes of C_k, remains in C_k at the next step (indeed, the expected escape time from C_k is $(1-\alpha_k)^{-1}$): hence the name of persistence probability [21].

Since α_k tends to grow with the size N_k of C_k (trivially, $\alpha_k = 1$ for the entire network), large α_k values must be checked for their statistical significance. For that, from the given network we derive the empirical distribution of the persistence probabilities $\bar{\alpha}_k$ of the connected subgraphs of size N_k (we do that by randomly extracting 1000 such networks), and we quantify the significance of α_k by the *z-score*:

$$z_k = \frac{\alpha_k - \mu(\bar{\alpha}_k)}{\sigma(\bar{\alpha}_k)}. \tag{2}$$

To summarize, a large value of α_k (i.e., $\alpha_k > 0.5$) reveals the strong cohesiveness of the subgraph C_k, while a large value of z_k (i.e., $z_k > 3$) denotes that such a cohesiveness is not trivially due to the size of the subgraph, but it is anomalously large with respect to the subgraphs of the same size.

Computing α_k and z_k on the subgraphs corresponding to the *locali*, we obtain the results of Table I. Rows from L2 to L18 refer to the 17 *locali* under investigation, all based in Milan area (Milan itself plus 16 small-medium towns); L19 collapses the individuals, participating in some of the meetings, belonging to any of the Calabria *locali*, and L20 contains those affiliated to Brescia, not subject to investigation and whose members participated in the meetings only occasionally; L0 are the individuals with non specified affiliation, L1 those who are not affiliated. Four *locali* (highlighted in bold in the table) reveal strong – and statistically significant – cohesiveness, proving to actually behave as communities in the sense of network analysis. It cannot be claimed, however, that the *locali* partition as a whole forms a well defined clusterization.

IV. Community Analysis

A. Max-modularity

We now reverse our approach by putting aside for a while the a priori *locali* partition and performing a standard max-modularity community analysis. Obviously, we cannot expect the latter to be able to recover the full *locali* partition, as most of the *locali* proved not to be significant communities. On the other hand, we might perhaps be able to disclose actual communities not necessarily matching the *locali* partition, thus providing important information on the structure of the crime organization.

Given a K-subgraph partition $C_1, C_2, \ldots, C_K$ of the nodes of a weighted, undirected network, the *modularity* Q [24], [25]

TABLE I. TESTING THE *Locali* PARTITION

	locale	N_k	α_k	z_k
L0	*Not specified*	*31*	*0.08*	*-3.15*
L1	*Not affiliated*	*8*	*0.03*	*-0.84*
L2	Bollate	13	0.25	1.31
L3	Bresso	15	0.39	2.72
L4	Canzo	2	0.10	0.47
L5	Cormano	22	0.41	3.96
L6	Corsico	4	0.12	0.21
L7	**Desio**	**19**	**0.63**	**6.40**
L8	Erba	9	0.37	2.44
L9	**Giussano**	**10**	**0.63**	**5.26**
L10	Legnano	10	0.20	0.77
L11	Limbiate	1	0	
L12	Mariano Comense	9	0.27	1.40
L13	**Milano**	**16**	**0.62**	**5.78**
L14	Pavia	5	0.13	0.25
L15	Pioltello	20	0.43	3.83
L16	Rho	5	0.18	0.78
L17	**Seregno**	**12**	**0.93**	**8.73**
L18	Solaro	5	0.06	-0.42
L19	*Calabria locali*	*35*	*0.19*	*-0.97*
L20	*Brescia*	*3*	*0.17*	*0.98*

TABLE II. RESULTS OF MAX-MODULARITY COMMUNITY ANALYSIS

	N_k	α_k	z_k
C1	12	0.93	9.07
C2	18	0.72	7.79
C3	25	0.66	9.85
C4	25	0.63	9.11
C5	45	0.68	8.20
C6	62	0.78	8.30
C7	67	0.67	5.72

is given by

$$Q = \frac{1}{2s} \sum_{k=1,2,\ldots,K} \sum_{i,j \in C_k} \left(w_{ij} - \frac{s_i s_j}{2s} \right), \qquad (3)$$

where $s = \sum_i s_i/2$ is the total link weight of the network. Modularity Q is the (normalized) difference between the total weight of links internal to the subgraphs C_k, and the expected value of such a total weight in a randomized "null network model" suitably defined [24]. Community analysis seeks the partition with the largest Q: large values ($Q \to 1$) typically reveal a high network clusterization. Although the exact max-Q solution cannot be obtained because computationally unfeasible even for small-size networks [7], many reliable sub-optimal algorithms are available: here we use the so-called "Louvain method" [26].

The result is a partition with 7 clusters ($Q = 0.48$), whose persistence probabilities are then computed and reported in Table II. All clusters are strongly cohesive (α_k much larger than 0.5, with very good statistical significance). Overall, the Infinito network displays therefore strong clusterization, with community size from small (12) to medium-large (67, about 26% of the network size).

We can quantify the similarity between the partition defined by *locali* (Table I) and the one obtained by max-modularity analysis (Table II). Several indicators have been proposed for comparing two network partitions (see [7], [27] for surveys): the most popular are based on node pairs counting, or on set-matching criteria, or on information theoretical notions. The latter are nowadays considered the most reliable: two of them are often used by scholars in network analysis, namely the *variation of information* V and the *normalized mutual information* I (see, e.g., [7] for definitions and properties).

TABLE III. COMPARING THE *Locali* WITH THE MAX-MODULARITY COMMUNITIES ($p_{hk}/r_{hk}/f_{hk}$)

	locale	C1	C2	C3	C4	C5	C6	C7
L0	*Not specified*							
L1	*Not affiliated*					*.11/.62/.19*		
L2	Bollate							.12/.61/.20
L3	Bresso		.56/.67/.61					
L4	Canzo							.03/1/.06
L5	Cormano					.40/.81/.54		
L6	Corsico					.07/.75/.12		
L7	Desio							.25/.89/.40
L8	Erba							.10/.78/.18
L9	Giussano				.40/1/.57			
L10	Legnano						.15/.90/.25	
L11	Limbiate						.02/1/.03	
L12	Mariano C.							
L13	Milano			.64/1/.78				
L14	Pavia					.07/.60/.12		
L15	Pioltello						.31/.95/.46	
L16	Rho						.08/1/.15	
L17	Seregno	1/1/1						
L18	Solaro						.06/.80/.12	
L19	*Calabria locali*							
L20	*Brescia*		*.17/1/.29*					

Both range from 0 to 1, with $V = 0$ and, respectively, $I = 1$ if and only if the two partitions are coincident. In our case, we obtain $V = 0.43$ and $I = 0.47$, which are intermediate values denoting a mild correlation. Overall the above results show that the criminal organization is indeed compartmentalized in its structure and function, but not necessarily the compartments match with the *locali* – or, at least, not all of them.

Perhaps more interesting is the pairwise comparison among the *locali* $L0, L1, \ldots, L20$ (see Table I) and the communities $C1, C2, \ldots, C7$ obtained by max-modularity. Here we adopt two indicators typically used in information retrieval, namely *precision* and *recall* (e.g., [28]). Let m_{hk} be the number of nodes classified both in L_h and in C_k. Then the precision $p_{hk} = m_{hk}/|C_k|$ is the fraction of the nodes of C_k that belongs to L_h whereas, dually, the recall $r_{hk} = m_{hk}/|L_h|$ is the fraction of the nodes of L_h that belongs to C_k. If we interpret L_h as the "true" set and C_k as its "prediction", then the precision quantifies how many of the predicted nodes are true, and the recall how many of the true nodes are predicted. The two quantities can be combined in the so called *f-score* $f_{hk} = 2p_{hk}r_{hk}/(p_{hk} + r_{hk})$, which is 1 if and only if both precision and recall are 1, and 0 if at least one of them is 0.

Table III summarizes the results of this analysis (only those cases in which at least one among p_{hk}, r_{hk}, and f_{hk} is larger than or equal to 0.5 are displayed). We firstly note that *locale* L17 perfectly matches community C1. Moreover, *locali* L13 and L9 are approximately identified as C3 and C4, respectively, whereas C2 corresponds to a large extent to the union of L3 and L20. In this respect, notice that a large recall ($r_{hk} \to 1$) means that most of the nodes of L_h belong to C_k (if not all, when $r_{hk} = 1$), but that C_k includes some other nodes too. Therefore, that last three columns of Table III put clearly in evidence that C5, C6 and C7 actually behave, to a large extent, as unions of *locali*. Overall, the picture is that of a quite strongly compartmentalized network, where compartments coincide with single *locali* or unions of them.

B. Local community analysis

Max-modularity, as well as most methods for community analysis [7], seeks for a partition of the entire network (i.e., each node is assigned to one and only one cluster) by optimizing a global indicator (modularity). An alternative (local) approach is that of starting from a given node, growing a subgraph by including one node at a time selected among those neighboring the current subgraph, and terminating when a suitable quality indicator stops improving. The procedure is repeated starting from all nodes, and the result is a set of communities with the following features:

- They might partially overlap: a few nodes could belong to two or more communities.
- They might be organized hierarchically: a small community might be completely included in a larger one.
- They do not need to cover the entire network, which is natural when part of the network is not significantly clusterized.

A number of local methods for community analysis have been proposed in recent years (e.g., [29], [30]). Here we use an algorithm which is a slight variation of [31], based on maximizing the persistence probability. Here is a detailed description.

Searching for local communities. We start from a single node k, so that the initial current subgraph is $C_k = \{k\}$ and the persistence probability $\alpha_k = 0$ (see eq. (1)). At each step, we include into C_k the node, selected among those neighboring C_k, that attains the maximal increase of α_k. We stop when we get a local maximum for α_k, namely when any new node insertion would decrease α_k. More precisely, to filter out possible small fluctuations of α_k, we stop when α_k decreases of at least $r = 0.05$ if a new node is introduced (this value has been tuned by trial-and-error). The community C_k is the subgraph which attains the maximum of α_k: it is retained only if $\alpha_k > 0.5$, otherwise it is discarded.

The procedure is repeated for each starting node $k = 1, 2, \ldots, N$, yielding the set $\mathbf{C} = \{C_1, C_2, \ldots\}$ of communities. Notice that not necessarily a valid community C_k is found from any starting node k, because α_k might increase monotonically to 1 (i.e., no maxima) as C_k grows to the entire network, or the maximum could have $\alpha_k \leq 0.5$ denoting a community not sufficiently cohesive.

Pruning. The set $\mathbf{C} = \{C_1, C_2, \ldots\}$ might contain communities C_i, C_j partially or totally coincident. If a subset $\mathbf{C}' = \{C_i, C_j, \ldots\} \subset \mathbf{C}$ exists such that

$$\frac{|C_i \cap C_j|}{|C_i \cup C_j|} > \mu, \tag{4}$$

for any $C_i, C_j \in \mathbf{C}'$, then we remove all the entries of $\mathbf{C}'$ from $\mathbf{C}$, except the one with the largest persistence probability. We set $\mu = 0.9$ having checked that, in the case under scrutiny, this value yields an effective pruning of subgraphs that typically differ by one node only.

Applying the above procedure to the Infinito network, we obtain 15 local communities which in total include 141 nodes (58% of the network). The main features of these communities are reported in Table IV. If we compare them with the *locali* composition, we obtain the precision/recall/f-score values summarized in Table V.

TABLE IV. RESULTS OF LOCAL COMMUNITY ANALYSIS

	N_k	α_k	z_k
C1	25	0.78	10.2
C2	25	0.67	9.27
C3	35	0.79	11.6
C4	21	0.71	7.79
C5	20	0.72	7.35
C6	4	0.74	6.96
C7	4	0.53	4.76
C8	15	0.59	5.27
C9	12	0.76	6.58
C10	5	0.82	7.17
C11	36	0.70	9.50
C12	32	0.63	8.52
C13	41	0.68	8.45
C14	12	0.93	8.65
C15	30	0.66	9.65

The detailed – node by node – analysis of the composition of the 15 local communities reveals a hierarchical organization, as summarized by the set relationships among communities displayed in Fig. 3. We note that two nodes (white in the figure) are shared among communities not hierarchically related, which therefore overlap. The two nodes, however, do not seem to have a role of effectively ”bridging” different communities, as they display low betweenness ranking.

The joint analysis of Table V and Fig. 3 highlights a few important facts. *Locale* L17 is again perfectly detected by the local method (community C14), as it was by max-modularity. But the hierarchical structure reveals that it contains two well cohesive sub-communities (C6 and C10) and, in its turn, it is part of larger subgraphs: C3, which clusters *locali* L9 and L17, and – at a higher level – C13, which includes L12 too. Community C1 is mostly coincident with *locale* L7, whereas C5 is approximately the union of L8 (one of the Lombardy *locali*) with the Calabria *locali* L19. Finally, the four communities C9, C4, C2, C15, hierarchically included one into the other, are all correlated to the union of *locali* L3 and L20. Overall, this analysis confirms that several *locali* actually behave as communities in the sense of network analysis, namely they show a strong internal cohesiveness which proves they behave as organized groups. Furthermore, the hierarchical structure of communities seems to indicate, on one side, the existence of some sort of *sub-locali* where the connectivity is even tighter and, on the other side, that the *locali* can cooperate forming larger organized (yet still cohesive) sets.

V. FILTERING DATA: THE REDUCED INFINITO NETWORK

The previous analysis reveals that the Infinito network has, overall, a fair level of clusterization. Some of the *locali* are actually well cohesive sets of individuals, whereas some others seem not to display such a feature. It should be quite clear that these evidences relate to a network whose reliability could be questioned in many respects. On one hand, there could be non recorded meetings, so that the social structure of the criminal organization could be significantly different: obviously, no information is available in this sense. On the other hand, sources of uncertainty in the network definition come from the choice of including those individuals who are not affiliated to any *locale* or whose affiliation is unknown,

TABLE V. COMPARING THE *Locali* WITH THE LOCAL COMMUNITIES ($p_{hk}/r_{hk}/f_{hk}$)

		C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15
L0	*Not specified*															
L1	*Not affiliated*							*.50/.25/.30*								
L2	Bollate															
L3	Bresso		.48/.80/.60		.57/.80/.67					.58/.47/.52						
L4	Canzo															
L5	Cormano															
L6	Corsico															
L7	Desio	.68/.89/.77											.56/.95/.71			
L8	Erba					.35/.78/.48						.19/.78/.31				
L9	Giussano			.29/1/.44										.24/1/.39		
L10	Legnano															
L11	Limbiate															
L12	Mariano C.													.20/.89/.32		
L13	Milano															
L14	Pavia															
L15	Pioltello															
L16	Rho															
L17	Seregno			.34/1/.51			1/.33/.50				1/.42/.59			.29/1/.45	1/1/1	
L18	Solaro															
L19	*Calabria locali*					*.60/.34/.44*										
L20	*Brescia*		*.12/1/.21*		*.14/1/.25*					*.25/1/.40*						*.10/1/.18*

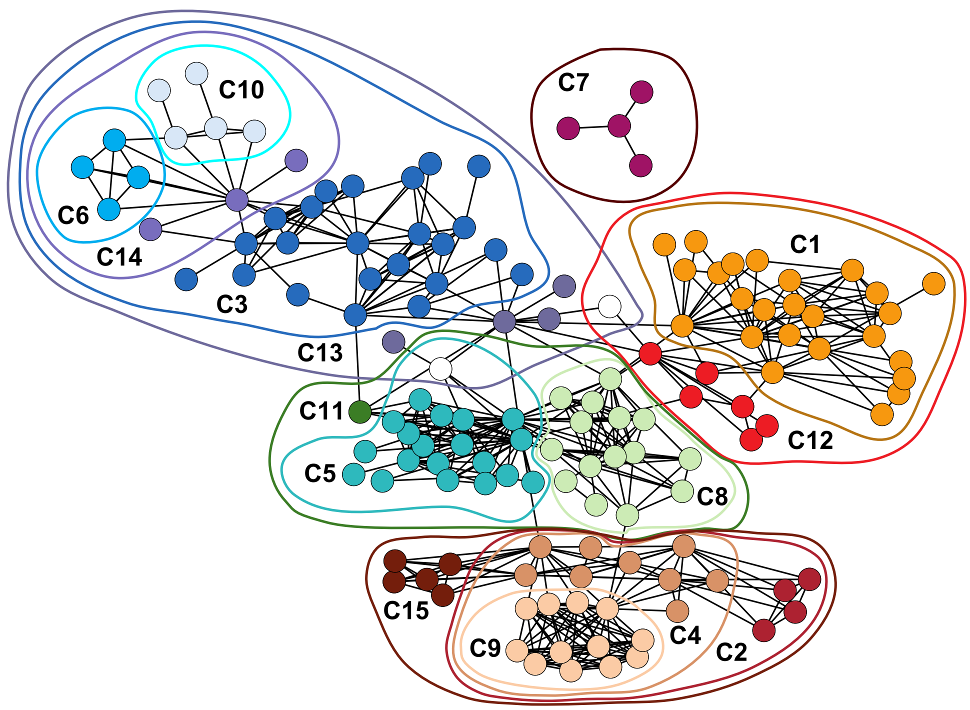

Fig. 3. The hierarchical organization of the local communities in the Infinito network. The figure displays the subgraph induced by the nodes belonging to the 15 identified local communities of Table IV.

as well as from the inclusion of very small *locali*, which are perhaps the result of an incomplete identification.

For all these motivations, and with the aim of understanding more deeply the structure of the criminal organization, we filter the network by excluding all nodes which are non classified, non affiliated, belonging to Calabria *locali* (who had only occasional meetings with the Lombardy 'Ndrangheta members), and belonging to Lombardy *locali* with size less than or equal to 5. We obtain what we call the *reduced Infinito network*, with $N = 155$ nodes (61% of the original network) classified in 11 *locali*, and $L = 1090$ links.

We first test the significance of the *locali* partition by computing the persistence probability α_k (and its statistical significance z_k) for each one of the 11 *locali*. The results are in Table VI, which should be compared to the analogous Table I that refers to the full Infinito network. In the reduced network, 8 *locali* over 11 (highlighted in bold in the table) show large cohesiveness ($\alpha_k > 0.5$), while they only were 4 over 17 in the full network.

If we perform a max-modularity community analysis on the reduced Infinito network, we obtain an optimal partition composed of 7 communities: all of them are significant, since their persistence probabilities range from $\alpha_k = 0.58$ to 0.90, with large z_k. Comparing the max-modularity partition with the 11-cluster *locali* partition, we obtain a variation of information $V = 0.26$ and a normalized mutual information $I = 0.69$, both definitely improved with respect to the full-network case

TABLE VI. TESTING THE *Locali* PARTITION OF THE REDUCED INFINITO NETWORK

	locale	N_k	α_k	z_k
L2	Bollate	13	0.35	1.22
L3	**Bresso**	**15**	**0.52**	**2.51**
L5	**Cormano**	**22**	**0.53**	**3.07**
L7	**Desio**	**19**	**0.73**	**4.52**
L8	**Erba**	**9**	**0.59**	**3.16**
L9	**Giussano**	**10**	**0.77**	**4.60**
L10	Legnano	10	0.29	0.84
L12	Mariano Comense	9	0.34	1.26
L13	**Milano**	**16**	**0.75**	**4.41**
L15	**Pioltello**	**20**	**0.56**	**3.10**
L17	**Seregno**	**12**	**0.95**	**6.21**

(they were $V = 0.43$ and $I = 0.47$). Overall, filtering the network has strengthened its clusterization and improved the detectability of the *locali* by means of community analysis. Indeed, a careful (node by node) comparison of the two partitions highlights that the max-modularity communities are good approximations of single *locali* or of unions of them.

VI. CONCLUSION

In this paper, methods of network analysis have been used to investigate the structure of a mafia organization. The participation of members to meetings has been adopted as a proxy for describing their network of relationships, and community analysis has been used to assess whether the criminal organization has a clusterized structure. Furthermore, since the formal membership to a *locale* (a local cellular group with a precise hierarchy and organization) was known for most of the tracked individuals, such information was used for benchmarking the ability of community analysis to recover the organizing partition of the mafia network.

The result is that the criminal network shows significant clustering, which supports the intuition that subgroups matter in this type of organizations. As expected, clusters often coincide with the *locali* – or with unions of them, apparently cooperating. Overall, these findings reinforce the idea that the tools of network analysis can be fruitfully adopted for enhancing the understanding of the structure and function of organized crime, albeit their use as a support for law enforcement intelligence still needs further exploration.

The research can be extended in many directions. First of all, a deeper structural analysis on a pool of criminal networks would be needed, aimed at assessing whether peculiar structural attributes turn out to be recurrent in such networks. Then, coming back to the problem of community detection, other methods might prove to be more effective – including those specifically devoted to bipartite networks, as it is our data structure before projection (see Sec. II). Finally, once the structure has been thoroughly understood, the challenge is clearly that of linking it with the *function* of the network, namely to fully understand how structural properties relate to criminal activities.

ACKNOWLEDGMENT

The authors would like to thank Giulia Berlusconi, Nicola Parolini and Marco Verani for many useful discussions.

REFERENCES

[1] F. Calderoni, “Identifying mafia bosses from meeting attendance,” in *Networks and Network Analysis for Defence and Security*, ser. Lecture Notes in Social Networks, Masys, A, Ed. Springer International Publishing Switzerland, 2014, pp. 27–48.

[2] C. Morselli, C. Giguere, and K. Petit, “The efficiency/security trade-off in criminal networks,” *Soc. Networks*, vol. 29, no. 1, pp. 143–153, 2007.

[3] A. Malm and G. Bichler, “Networks of Collaborating Criminals: Assessing the Structural Vulnerability of Drug Markets,” *J. Res. Crime Delinq.*, vol. 48, no. 2, pp. 271–297, 2011.

[4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. H. Hwang, “Complex networks: Structure and dynamics,” *Phys. Rep.*, vol. 424, no. 4-5, pp. 175–308, 2006.

[5] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[6] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[7] S. Fortunato, “Community detection in graphs,” *Phys. Rep.*, vol. 486, no. 3-5, pp. 75–174, 2010.

[8] P. Jonsson, T. Cavanna, D. Zicha, and P. Bates, “Cluster analysis of networks generated through homology: automatic identification of important protein communities involved in cancer metastasis,” *BMC Bioinformatics*, vol. 7, 2006.

[9] A. E. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. W. Taylor, “Compartments revealed in food-web structure,” *Nature*, vol. 426, no. 6964, pp. 282–285, 2003.

[10] C. Piccardi, L. Calatroni, and F. Bertoni, “Communities in italian corporate networks,” *Physica A*, vol. 389, pp. 5247–5258, 2010.

[11] G. Flake, S. Lawrence, C. Giles, and F. Coetzee, “Self-organization and identification of web communities,” *Computer*, vol. 35, no. 3, pp. 66–71, 2002.

[12] M. A. Fortuna, J. A. Bonachela, and S. A. Levin, “Evolution of a modular software network,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 108, no. 50, pp. 19 985–19 989, 2011.

[13] M. Girvan and M. Newman, “Community structure in social and biological networks,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 99, no. 12, pp. 7821–7826, 2002.

[14] A. Arenas, L. Danon, A. Diaz-Guilera, P. Gleiser, and R. Guimera, “Community analysis in social networks,” *Eur. Phys. J. B*, vol. 38, no. 2, pp. 373–380, 2004.

[15] L. Paoli, *Mafia brotherhoods: organized crime, Italian style*. New York, NY: Oxford University Press, 2003.

[16] F. Varese, “How mafias migrate: The case of the Ndrangheta in northern Italy,” *Law Soc. Rev.*, vol. 40, no. 2, pp. 411–444, 2006.

[17] L. Paoli, “Mafia and organised crime in Italy: The unacknowledged successes of law enforcement,” *West Eur. Polit.*, vol. 30, no. 4, pp. 854–880, 2007.

[18] Tribunale di Milano, *Ordinanza di applicazione di misura coercitiva con mandato di cattura - art. 292 c.p.p. (Operazione Infinito)*. Ufficio del giudice per le indagini preliminari, 2011.

[19] V. Batagelj and A. Mrvar, “Pajek - Analysis and visualization of large networks,” in *Graph Drawing Software*, ser. Mathematics and Visualization, Jünger, M., and Mutzel, P, Ed. Springer-Verlag Berlin, 2004, pp. 77–103.

[20] T. Kamada and S. Kawai, “An algorithm for drawing general undirected graphs,” *Inf. Process. Lett.*, vol. 31, no. 1, pp. 7–15, 1989.

[21] C. Piccardi, “Finding and testing network communities by lumped Markov chains,” *PLoS One*, vol. 6, no. 11, p. e27028, 2011.

[22] F. Della Rossa, F. Dercole, and C. Piccardi, “Profiling core-periphery network structure by random walkers,” *Sci. Rep.*, vol. 3, p. 1467, 2013.

[23] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, “Defining and identifying communities in networks,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 9, pp. 2658–2663, 2004.

[24] M. E. J. Newman, “Modularity and community structure in networks,” *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 23, pp. 8577–8582, 2006.

[25] A. Arenas, J. Duch, A. Fernandez, and S. Gomez, “Size reduction of complex networks preserving modularity,” *New J. Phys.*, vol. 9, p. 176, 2007.

[26] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *J. Stat. Mech.-Theory Exp.*, p. P10008, 2008.

[27] M. Meilă, “Comparing clusterings - an information based distance,” *J. Multivar. Anal.*, vol. 98, pp. 873–895, 2007.

[28] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Addison Wesley, 1999.

[29] J. Bagrow and E. Bollt, “Local method for detecting communities,” *Phys. Rev. E*, vol. 72, no. 4, Part 2, p. 046108, 2005.

[30] A. Clauset, “Finding local community structure in networks,” *Phys. Rev. E*, vol. 72, no. 2, Part 2, p. 026132, 2005.

[31] P. Landi and C. Piccardi, “Community analysis in directed networks: In-, out-, and pseudocommunities,” *Phys. Rev. E*, vol. 89, no. 1, p. 012814, 2014.