Granados and Vargas *EPJ Data Science* (2022) 11:6
https://doi.org/10.1140/epjds/s13688-022-00318-w

EPJ Data Science
a SpringerOpen Journal

REGULAR ARTICLE Open Access

The geometry of suspicious money laundering activities in financial networks

Oscar M. Granados[1,2*] and Andrés Vargas[2]

*Correspondence: oscarm.granadose@utadeo.edu.co
[1]Department of Economics and International Trade, Universidad Jorge Tadeo Lozano, Carrera 4 No. 22-61, Bogotá, Colombia
[2]Departamento de Matemáticas, Pontificia Universidad Javeriana, Carrera 7 No. 40-62, Bogotá, Colombia

Abstract

Corruption and organized crime are social problems that affect different communities around the world, involving public and private organizations in diverse sectors and activities. However, these problems are global phenomena that transcend economic, cultural, and social borders, especially, when corrupt individuals use the global financial system to protect their illegal money. This paper aims to evaluate the money laundering mechanism in financial networks, studying the structure of some suspicious money laundering groups, and how they could be detected by the use of topological and geometrical considerations that avoid the need of possibly non-available (or restricted) information.

Keywords: Financial services; Financial crime; Money laundering; Network topology; Network geometry

1 Introduction

The financial ecosystem is a complex network of institutions and transactions from different places and a variety of volumes, values, currencies, and instruments, where financial institutions have created different vehicles, structures, services, and products to take advantage of globalization and growing financial complexity. However, some clients and institutions use these platforms to develop money laundering activities, and the amount of money laundered by year is between 2% and 5% of the global Gross Domestic Product (GDP), according to the United Nations Office on Drugs and Crime (UNODC) [1]. Namely, economic crimes and criminal organizations use different mechanisms to protect their illegal money, and taking advantage of the global financial system is one of them. Criminals use this system to transfer or invest funds between several jurisdictions around the world to try to legalize their money.

Preliminary studies have investigated money laundering as a global economic cross phenomenon in a few cases [2–7] or as a social network analysis [8–10]. However, the integration of these phenomena could emerge as a complex systemic risk and propagate across different systems to a global scale [11–14]. The growth of a broad range of illegal activities and the volume of interactions in financial networks have created increasing complexity and, probably, constitute a context in which criminals are the beneficiaries because financial crimes have become more frequent and diverse. Thus, this phenomenon needs

© The Author(s) 2022. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

new approaches constantly because criminal organizations change rapidly their activity schemes like sophisticated structures of ownership (parent firms and subsidiaries) to outwit authorities [15], and even more when the financial transaction volume rises every day. It is our view in this paper that with a topological analysis of the network structure it is possible to identify suspicious groups, and how their members are separated from each other, in order to determine if some agents of these groups can develop financial crimes like money laundering. Our study suggests that the structure of these networks facilitates the emergence of certain patterns and relationships among financial and non-financial intermediaries and offshore entities that are characteristic of money laundering activities.

The aim of this work is to gain a better understanding of the money laundering mechanisms in financial networks by studying the topological structure of these suspicious money laundering groups. Indeed, we apply some tools from topology and geometry that are well-adapted to this setting to reveal and detect the most relevant groups of agents in the network that develop suspicious interactions that are frequently associated with money laundering activities. For this purpose, we employ as a basis the analysis of Caribbean financial networks that has appeared in [16], and introduce a proposal for the topological detection of characteristic money laundering interactions in a network. We will also employ as an aid in our study a discrete quantity called the Forman–Ricci curvature of a graph that in the case of simple undirected and unweighted networks provides information that can be obtained by degree centrality alone, but it has the advantage of having a geometric interpretation and is well-defined in settings where more information like weights associated to agents or interactions is present in the network. This is desirable for a more refined analysis of suspicious activities but it will not be attempted here since our purpose is to consider the case when only the minimum quantity of information is available.

This paper is divided as follows. In the Sect. 2, we present some basic network and financial terminology employed in the text and describe the datasets used in this study. The Sect. 3 is devoted to a brief description of the discrete topological and geometrical notions employed in this work, including a general definition of geometric simplicial complexes. In the subsequent Sect. 4, we present a simple set of rules associated to money laundering activities which, nevertheless, capture several of the most common characteristics of this type of fraudulent activity. Moreover, we propose a strategy to detect and study the emergence of suspicious groups in financial networks in order to identify with greater precision their possible members. The Sect. 5 presents the most relevant results obtained by the use of the proposed strategy to identify the largest suspicious money laundering activities in our financial networks, including observations that are derived from topological and geometrical considerations. We close the paper with Sect. 6, devoted to highlight important aspects of our approach and provide directions for future work.

2 Materials

2.1 Data collection

The financial networks used in this work were constructed from the Bahamas Leaks and the Panama Papers datasets based on the International Consortium of Investigative Journalists [17]. These datasets contain information about 0.7 million agents (intermediaries, offshore entities, and officers) and almost 1 million interactions among these agents. We have used the same data structure of the source where the data set is divided into different subsets with information of intermediaries, offshore entities, officers, and one additional

subset with the list of interactions. These interactions are given by their corresponding nodes, but they do not have specific information about dates. Namely, edges could correspond to permanent or occasional interactions, and it is impossible to know anything beyond that.

We compiled the data, but these datasets require some considerations that deserve mentioning. (1) The datasets are historical aggregate data for the 1980–2017 with unweighted nodes (also called agents in this setting) and undirected interactions (edges). (2) We did not use irregular information appearing in the data like repeated names, non-registered addresses, and non-classified agents. (3) We removed all data corresponding to intermediaries, offshore entities, and officers that appear without jurisdiction or other incomplete information, i.e., we used only complete registers available. (4) We used several information elements of the original datasets of Bahamas Leaks and Panama Papers except for the information sources list. (5) For legal concerns, we did not use names or other information that could be sensitive. Only those agents who have been legally being proven their participation in money laundering activities are identified (e.g., Mossack Fonseca).

Other considerations about the data are: the agents (nodes) are classified into three groups (see Table 1): intermediaries, offshore entities, and officers. Intermediaries are institutions that connect financial and business operations, and their activities include a broad range of financial, consulting, legal, accounting, and management services. Intermediaries have two subgroups financial and non-financial intermediaries. The financial intermediaries comprise all financial institutions like banks, private banks, trust companies, funds, or wealth managers with operations in the Caribbean jurisdictions. The non-financial intermediaries refer to consulting, legal, accounting, and management services for corporate and private entities. Offshore entities are companies incorporated in jurisdictions with low taxation rates (tax havens), and their activities are developed overseas, i.e., the company does not undertake business with persons resident in that jurisdiction. Additionally, its profits are not repatriated. Officers represent an intermediary or an offshore entity legally, habitually. They are employees of non-financial intermediaries or, in some cases, financial institutions.

Table 1 Type of Agents and Interactions

Agents/Nodes	Type	Sub-type
Intermediaries	Financial	Funds Banks Private Banks Trust Companies Wealth Managers
	Non-Financial	Legal Consulting Accounting Management
Offshore Entities	–	–
Officers	–	–
Edges Interactions	Type Financial Legal Consulting Accounting Management	

Agents are distinguished by their characteristics: size, activity, service, global network, to list a few, all of which can define the agent's interactions with others. For example, first, it may be that a law firm in several jurisdictions with a specific service is more likely to be visible (e.g., Mossack Fonseca). Second, it may be that the global banks are more active in a particular financial service. Third, several non-financial intermediaries with certain specialized services are more likely to attend to particular clients like global corporations, ultra-high wealth families, or public figures. Those examples are a sample that interactions are distinguished by the agents' characteristics and indicate how many agents that agent had contact in different jurisdictions.

Furthermore, the agents in the network (intermediaries, offshore entities, and officers) can interact between themselves through five types of edges: financial, business, legal, accounting, and management relations (see Table 1). This classification is a proxy because agents have interactions of any kind without being defined as some specific kind of transaction during the analyzed period. Additionally, some offshore entities and intermediaries have the date of incorporation, or establishment, respectively, but do not have specific data about transactions date. For this reason, interactions could be continuous or occasional, but it is not possible to know.

2.2 Data analysis

The data analysis using some network metrics identifies that networks have a low average node degree (see Table 2), which means that a big part of the agents is connected to at most another one. The average path length for the Bahamas Leaks and Panama Papers networks is 9.62 and 10.42, respectively, confirming the above.

In general, it is noteworthy that a moderate number of particular agents are highly connected to many others that only connect to it. Not all agents are connected between themselves. On the contrary, most agents are not connected to many others. Namely, the principal heterogeneity observed in the network has to do with two different types of interacting agents: a) Central nodes connected to many other nodes and b) nodes with low degrees connected to central nodes. Finally, the assortativity for the Bahamas network is –0.2404, and for the Panama Papers network, it is –0.0521. In this case, a network that is non-assortative, i.e., with negative assortativity, may comprise agents that are themselves highly assortative but others lowly assortative. A particularity of corrupt and money laundering networks is that some agents do not have the interest to connect with other high degree agents, only in some situations or through another agent [16].

Table 2 Some Network Metrics

Variable	Bahamas leaks	Panama papers
Total Nodes	219,856	559,433
Total Edges	246,291	657,488
Intermediaries	541	14,110
Offshore Entities	175,888	213,634
Officers	25,262	238,402
Average Node Degree	2.24	2.35
Avg. Shorthest Path Length	9.62	10.42
Assortativity	–0.2404	–0.0521
Connected Components	390	11,043
Detected Communities	536	11,569

The community analysis presented in that reference shows these networks' large-scale distribution and complexity, revealing a sophisticated internal organization and compartmentalized structure into components and communities. Although the networks are highly disconnected, both of them carry a very significant central component where the community detection algorithm still detects several smaller (but interconnected) groups. Thus, it becomes relevant to determine the possible role of these groups in the structure of money laundering. As it will become clear later, our topological approach shows that the community detection strategy is not precise enough to describe the fine structure of suspicious money laundering activities because they may comprise agents in different communities.

Furthermore, the metrics in Table 2 indicate that it is necessary to use other tools to single out suspicious money laundering groups. One of them is the negative assortativity of the networks because it is a representation of a characteristic that may differ for each node in the graph [18]. In the next section, using topological and geometrical considerations, we implemented a procedure to identify the most relevant candidate members of these kinds of groups.

3 Methods

To study the topological and geometrical structure of the financial networks under consideration, we employ mathematical tools from topology and discrete geometry adapted to this setting. We present here a brief overview without entering into technicalities, but relevant references with precise details are provided in the text.

We define a *graph* $G = (V, E)$ as a pair consisting of a set V of *nodes*, and a set E of pairs of nodes called *edges* and a subgraph $G' = (V', E')$ of G is a graph with $V' \subset V$ and $E' \subset E$, where E' contains all elements of E connecting vertices of V'. A *complete graph* is a graph with edges connecting any possible pair of its vertices [19]. Subsets of the collection of edges of a graph that are used in our analysis are:

- A *path* is a finite sequence of distinct edges joining a sequence of distinct vertices.
- A *circuit* is a non-empty path in which the first and last vertices are repeated.
- A *cycle* or *simple circuit* is a circuit in which the only repeated vertices are the first and the last vertices. If there are no external edges connecting any two of its vertices it is called a *chordless cycle* or *hole*.
- A *clique* is a set of vertices for which the corresponding subgraph is a complete graph. A *maximal clique* is a clique that is not properly contained in a larger one. A clique with k vertices will be called a *k-clique* (although there are other notions with the same name in the literature).

By the topological structure of a network, we refer to all the aspects related to its connections, i.e., the description of how the network nodes are interconnected between themselves giving rise to its characteristic distribution of paths and voids, without requiring any notion of length or size. In contrast, by its geometric structure, we refer mainly to those aspects directly associated with some notion of size, which can be given in terms of distances on the network (arising from the minimum number of edges connecting two nodes, or from any particular metric), in terms of a discrete notion of volume related to the degree of the nodes, or in terms of another similar quantity. Although, in general, the geometry of the network is not independent of its topology, it provides further insights into its structure.

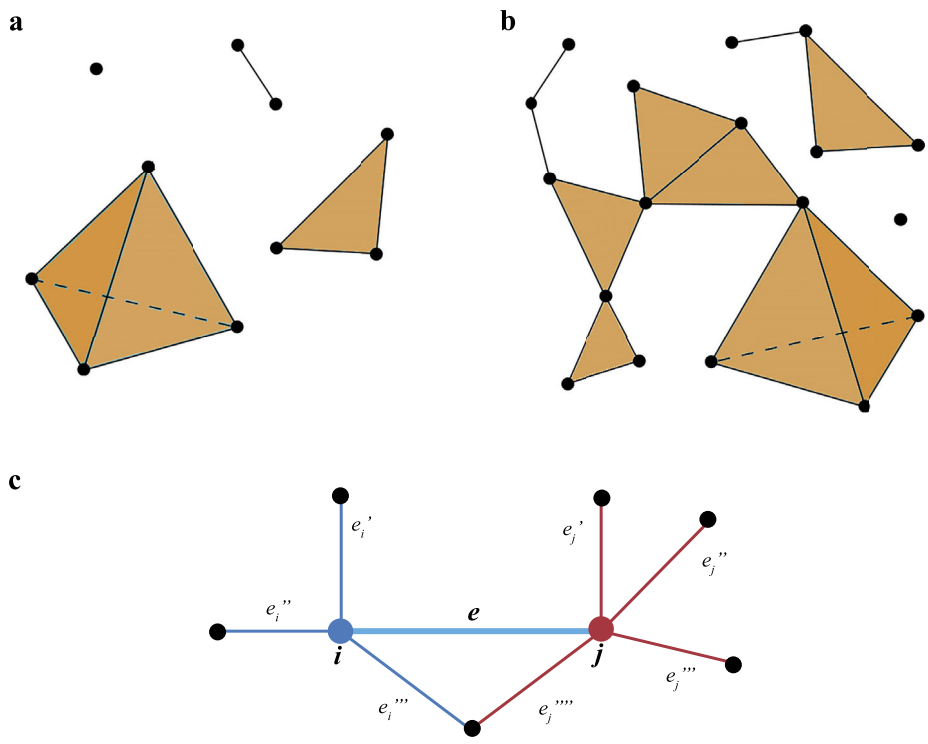

Figure 1 Simplices and geometric simplicial complex. (**a**) Example of geometric n-simplices for $n = 0$ (vertex), $n = 1$ (edge), $n = 2$ (3-clique including its 2D interior) and $n = 3$ (4-clique including its 3D interior, where the dashed line corresponds to an edge that is hidden by the solid 3-simplex body in 3D). (**b**) A geometric realization of a simplicial complex associated to a graph represented by black vertices and edges. The faces or 2-simplices, the 3-simplices and higher dimensional simplices are not usually considered for a proper (i.e., one-dimensional) network but can be considered in a higher dimensional approach. (**c**) Forman–Ricci curvature of an edge e on a one-dimensional approach to networks is calculated in terms of the number of other edges incident to its vertices i and j

For the topological characterization of networks, their simplicial structure is necessary to calculate invariants such as the Euler characteristic and their simplicial homology, although they are not explicitly needed for our purposes. These standard tools of algebraic topology can be described in different degrees of generality, but we adopt here a point of view based on the geometric realization of abstract simplicial complexes in the context of graphs (Fig. 1a and Fig. 1b). As far as the analysis presented here goes, we only need to provide some definitions and remark that the reduction procedure described in the forthcoming sections leaves the topological structure of our networks invariant. Therefore, it does not alter quantities like the Euler characteristic or their homology.

3.1 Topological notions

A geometric n-simplex of a graph G is a set of $(n + 1)$ completely connected nodes or vertices that define a complete subgraph, hence, they corresponds to an $(n + 1)$-clique. An orientation of an n-simplex is an ordering of its vertices (i.e., a directed $(n + 1)$-clique), but here we will deal only with undirected graphs and the orientation will not be required. In this way, a graph G gives rise to a geometric simplicial complex understood as the finite collection $\mathcal{S}(G)$ of all the possible simplices determined by its vertices and edges. By definition, $\mathcal{S}(G)$ is required to contain all the proper subsimplices, or faces, that belong to

any of its member simplices. Indeed, the defining property of a simplicial complex is to be closed under the operation of taking subsimplices of its elements. It is not difficult to show that the intersection of two simplices of $\mathcal{S}(G)$ neither of which contains the other is a (possibly empty) proper subsimplex of both, i.e., a common face.

Denoting by $\mathcal{S}_n$ the subset of all n-simplices of $\mathcal{S}(G)$ it is clear that we can write $\mathcal{S}(G) = \mathcal{S}_0 \cup \mathcal{S}_1 \cup \cdots \cup \mathcal{S}_N$ if $n = N$ is the highest order of any n-simplex of G. If $|\mathcal{S}_n|$ denotes the number of n-simplices of G, the Euler characteristic of G can be defined by

$$\chi(G) = \sum_{n=0}^{N} (-1)^n |\mathcal{S}_n|. \tag{1}$$

The Euler characteristic is a well-known topological invariant of a space and in this context it is a simple quantity that encodes global topological information of a network. In our context, its relevance lies only in the fact that this quantity remains invariant under the reduction procedure presented in the Money Laundering section but its particular value will not be needed in our analysis.

3.2 Geometry and curvature

In the last decade, there has been substantial interest in studying useful geometric notions like distances and curvatures (among others) that are naturally adapted to the network context. Some of these notions can be induced on a network by embedding it into an appropriate ambient space, and in this way, the relevance of hyperbolic geometry has been established in the literature [20] using embeddings into some of the models of this geometry. Although such an approach appears explicitly extrinsic, network models have been established for which hyperbolic geometry emerges naturally by specifying growing or attachment conditions for the nodes [21]. Other approaches to these notions come from the discretization of smooth geometric quantities using different considerations. Due to its purely combinatorial definition, which is intrinsic (i.e., independent of any embedding of the network into an ambient space) and its simplicity for calculations, we have chosen as a tool to describe the geometry of our networks the so-called Forman–Ricci curvature that was introduced in [22] in the general mathematical context of CW-complexes (Fig. 1c), and which has a very simple definition in our situation of undirected, unweighted networks. Nevertheless, it should be mentioned that some of the advantages of this notion lie precisely in providing a sensible definition when there are weighted nodes and/or edges in the network, and this situation will not be considered here. The precise mathematical definition is as follows.

For an unweighted graph $G = (V, E)$ the Forman–Ricci curvature $\mathrm{Ric}_F(e)$ of an edge $e \in E$ is defined as the number of its vertices, i.e. 2, plus the number of faces or true topological 2-simplices to which it belongs (that for proper graphs that carry only zero and one-dimensional simplices is always zero), minus the number of edges of G parallel to it (which are all their neighboring edges), see Fig. 1c. The total value is therefore equal to the sum of the degrees of its vertices minus 2. Explicitly, this can be written as

$$\mathrm{Ric}_F(e) = 2 - \left|\left\{e' \in E : e' \parallel e\right\}\right| = 4 - \sum_{v \sim e} \deg(v), \tag{2}$$

where the relation $\parallel$ denotes parallelism and $\sim$ denotes incidence.

From an intuitive point of view, the Forman–Ricci curvature of an edge quantifies the dispersion or divergence of paths at the two ends (i.e., vertices) of that edge and provides a geometric assessment of the importance of that edge in terms of how it is connected to its neighbors. Being an accurate discretization of a continuous notion, the Forman–Ricci curvature provides a direct and intrinsic geometric analog in the network setting of a well-established notion of Ricci curvature that is very important in applications and can be interpreted as a measure of how the "discrete volume" of connections changes around the node/edge under consideration. Although there are correlations between Forman–Ricci curvature and other more standard network metrics, they depend on the particular network, for example, for scale-free networks, there is a negative correlation with the degree of nodes, but the correlation is weaker for real networks than for network models. Additional details can be found, e.g., in [23, 24] and references therein, including more general definitions of this notion and its relation to other metrics in real and model networks.

4 Money laundering networks

In this section, we present a set of rules common to several instances of money laundering activities and explain how these rules can be manifested in the topology of a financial network to propose a useful strategy in determining suspicious interactions. The importance of these rules is that they can be used in the absence of any additional information apart from the fact that there is some interaction between the network agents.

4.1 Rules of suspicious money laundering activities

The Financial Action Task Force (FATF) defines Money Laundering as the process by which money generated through criminal activity appears to have come from a legitimate source. For evaluating suspicious interactions related to money laundering, like the movements to distance money from their source (geographically or financially) or their business activity, the local laundering in the jurisdiction in which the money was generated, or the non-local laundering in other jurisdictions, distance indicators are important [7]. Although the money might be moved through the trading of financial instruments or wired through different banking accounts anywhere around the world as tax havens (or jurisdictions with lax laws and low regulatory standards), it is expected that at least to some extent, this money will return to the money launderer agent after various transactions (few or a lot) determining a cycle. Thus, money laundering activities exhibit some characteristics that have been identified by diverse international organizations [1, 25], and the most relevant ones for our topological analysis can be summarized as a set of rules to detect groups of suspicious agents (nodes) that could be related to money laundering without the need of additional (and possibly confidential or unavailable) information concerning those interactions:

- *Rule 1.* The money launderer agent's interactions do not grow rapidly because it is preferable to maintain only few interactions with other agents and to keep them as anonymous or covert as possible.
- *Rule 2.* The money launderer agent does not impose restrictions on the geographical distance required for those interactions. This includes the transference of money between a non-haven jurisdiction and a tax haven jurisdiction.
- *Rule 3.* The money launderer agent does not care how many transactions are needed to clean the illegal money (cycle), resorting to deposits triangulation between the same agents.

- *Rule 4.* The money launderer agent employs the breaking up of large amounts of money into smaller amounts to avoid suspicions. The money is then deposited into one or more bank accounts or other financial instruments either by different persons or by a single person over a period of time.
- *Rule 5.* After a process of money laundering interactions is initiated, it is to be expected that at some point (at least part of) the involved money returns to the money launderer agent completing a cycle.

4.2 Topology and geometry of money laundering

Using the previous list, we present here a simple translation of those rules to the language of graphs in order to explain how those suspicious activities may be reflected in a financial network.

We are going to assume that the information used in the construction of the graph carries the bare minimum data necessary to determine all the nodes involved in the financial interactions and that edges only represent a known interaction of some unspecified type.

Then, when studying the graph of the financial network the characterization can be reflected in the following rules:

- *Rule 1.* Look for paths for which a proportionally important part of their nodes are not as highly connected as the rest.
- *Rule 2.* Do not restrict the location of those paths to be confined to a particular community (or other subset) of the graph.
- *Rule 3.* The length of the paths considered should not be restricted.
- *Rule 4.* Look for paths with several bifurcation points, and study their behavior from those points.
- *Rule 5.* Among the possible paths in the graph, closed paths or cycles are more relevant that simple paths.

Consequently, for the determination of suspicious activities/interactions in a financial network, we are led to concentrate as a fundamental task in determining those cycles in the network passing through bifurcation points (node s) that also belong to many other cycles. The larger cycles could be associated with a possibly stronger effort to obfuscate the money origins, so it is natural to start looking for cycles whose size varies from the largest to the shortest appearing in the network.

Nevertheless, since the number of cycles in a large network can be vast, the rules indicate that the presence of several crossing cycles should be much more suspicious than individual ones. We propose a simplified approach that can detect groups of agents where the previous rules hold simultaneously in the most concentrated possible way without specifically studying all paths or even all cycles in the network. The detection procedure can be summarized in the following.

Strategy Three practical steps in the determination of the most suspicious money laundering interactions in a financial network are:

- *Step 1. Reduction:* Replace each node of degree 2 (i.e., a bifurcation-free node) in the network and its incident edges by only one edge connecting the node's neighbors and repeat this procedure until the elimination of all those nodes. These operations will possibly leave pairs of nodes connected by two or more edges (multi-edges) and also nodes with one or several self-loops so that the reduced graph is a multigraph. See Fig. 2 for several instances of this reduction procedure.

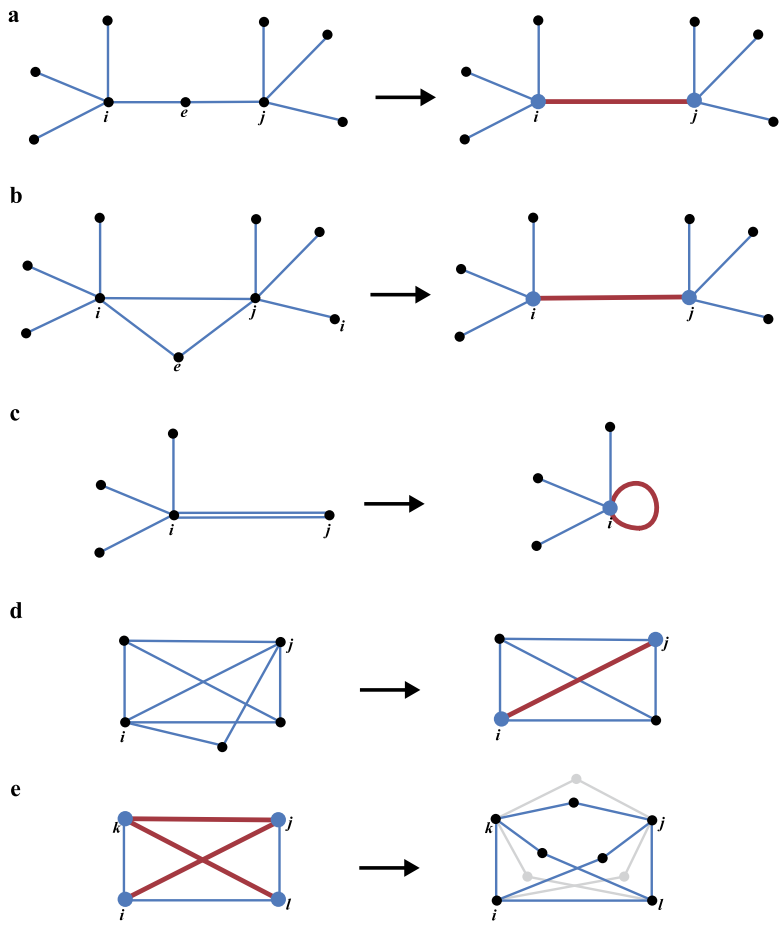

Figure 2 Explicit examples of the reduction procedure (Step 1). (**a**) Replacement of the 2-degree node *e* and its edges by a simple (red) edge. (**b**) After elimination of node *e*, the resulting multiedge is replaced by a simple (red) edge. (**c**) After the elimination of the 2-degree node *j*, a (red) self-loop appears that is also discarded. (**d**) After the reduction, the two paths joining *i* and *j* become a simple (red) edge. (**e**) After the detection of a true clique in the reduced network (called reduced clique), the complete cycles connecting their member nodes are determined in Step 3, restoring all the original members of the group. The right-hand side shows what is called an "embedded clique" in the unreduced network. It is not a true clique but a group of intersecting cycles, and hence a candidate focus of suspicious activities

- *Step 2. Clique search:* Once there are no bifurcation-free nodes in the network, proceed to determine all the maximal cliques in the reduced network, starting the search from the largest to the smallest ones. According to the rules and since the length of cycles plays no role after reduction, higher-order cliques should be more suspicious than lower-order ones. Furthermore, agents belonging to the intersection of several cliques could be pivotal centers for different suspicious activities, and their determination should also be highly prioritized.
- *Step 3. Embedding:* Find how the nodes of these cliques were embedded inside the original unreduced graph, restoring the possibly deleted vertices of 2-degree connecting their agents to obtain all the nodes involved in the cyclic interactions. Moreover, check some neighborhoods of those embedded cliques to study how they interact with other close members in the financial network and detect possible agents

connecting different cliques but which do not belong to any of them (and therefore could not have been detected before).

Although it is possible that strictly complete subgraphs, i.e., cliques, do not correctly appear in the unreduced financial network, the expected interconnection property of a corrupt community highly is a topological characteristic that is preserved by the reduction Step 1. From a practical point of view, this step only reduces path lengths but leaves the topological structure of the network invariant. Indeed, each application of this step actually removes one edge and one vertex of the original network, and therefore, preserves its Euler characteristic and all topological invariants (up to the removal of multiedges and self-loops). In other words, it preserves the connection structure of the network, which carries the money laundering characteristic activities. Furthermore, from this perspective, even the degree one nodes and their edges could also be deleted from the original network to reduce the network size when the dataset is too large.

Notice that the cliques of the reduced network (that we will call reduced cliques for short) do not necessarily correspond to true cliques in the original unreduced network. However, they indeed represent sets of maximally connected nodes that carry multiple cycles connecting groups of their members, and for this reason, they satisfy all the previously listed rules. Therefore, the relevance of this reduction is that any group of intersecting cycles of arbitrary lengths is converted to a clique or a set of intersecting cliques of different sizes but which are easier to detect and organize by size. When the nodes of the reduced k-cliques are viewed inside the original network by restoring their deleted nodes and connections, they will be called embedded k-cliques, even though they do not correspond to proper cliques in the unreduced original network. The reduction procedure does not alter the real cliques already present in that network.

After Step 1, some k-cliques of order $k \geq 3$ do necessarily appear in the reduced network, and the task is to find them all in Step 2. Starting from the set of highest order cliques to the lowest order ones, an analysis of the nodes appearing in these cliques is expected to provide an ample spectrum of corrupt candidate elements. However, since it is crucial to check their position in the original network, Step 3 must find the precise subgraph to which they belong, restoring possible previously deleted (and also suspicious) nodes.

Finally, the calculation of the Forman–Ricci curvature of the edges of all the embedded k-cliques as members of the original unreduced network helps to estimate the regions of the distribution of nodes that should be considered to be of interest for further investigation, but more importantly, it helps to corroborate that the embedded k-cliques found to carry some asymmetry in the volume of their connections, which is reflected in the fact that the curvature values lie around well-separated magnitudes.

Concerning the algorithmic implementation of these steps, we should point out that many of the standard software packages for the analysis of networks already provide functions to perform the reduction, the search for cliques (of particular sizes) and even for sets of paths connecting particular nodes.[1]

[1] for Steps 2 and 3, we have relied on standard network analysis functions like FindClique, FindCycle, and related ones available in Mathematica (v.11$^+$). For the reduction procedure of Step 1 on the function IGSmoothen implemented in the freely available Mathematica package IGraph/M.

5 Results

Applying the proposed strategy to find suspicious groups using topological and geometrical considerations introduced in the previous section, we obtained results in both financial networks. In this section, we present the obtained cliques from the Panama Papers and Bahamas Leaks networks as candidates for possible detection of money laundering of small groups. The usefulness of these results lies in catching agents performing interactions that generate suspicion because they have patterns commonly observed in money laundering activities. These agents go unnoticed with other methodologies due to the size of these financial networks.

Figure 3 shows the emergence of money laundering suspicious groups in the Panama Papers network. First, Fig. 3a shows the Panama Papers network and Fig. 3b shows the suspicious nodes in their communities represented in Fig. 3a. We find seven groups of 6-cliques that provide relevant information. One agent was one of the ten most essential

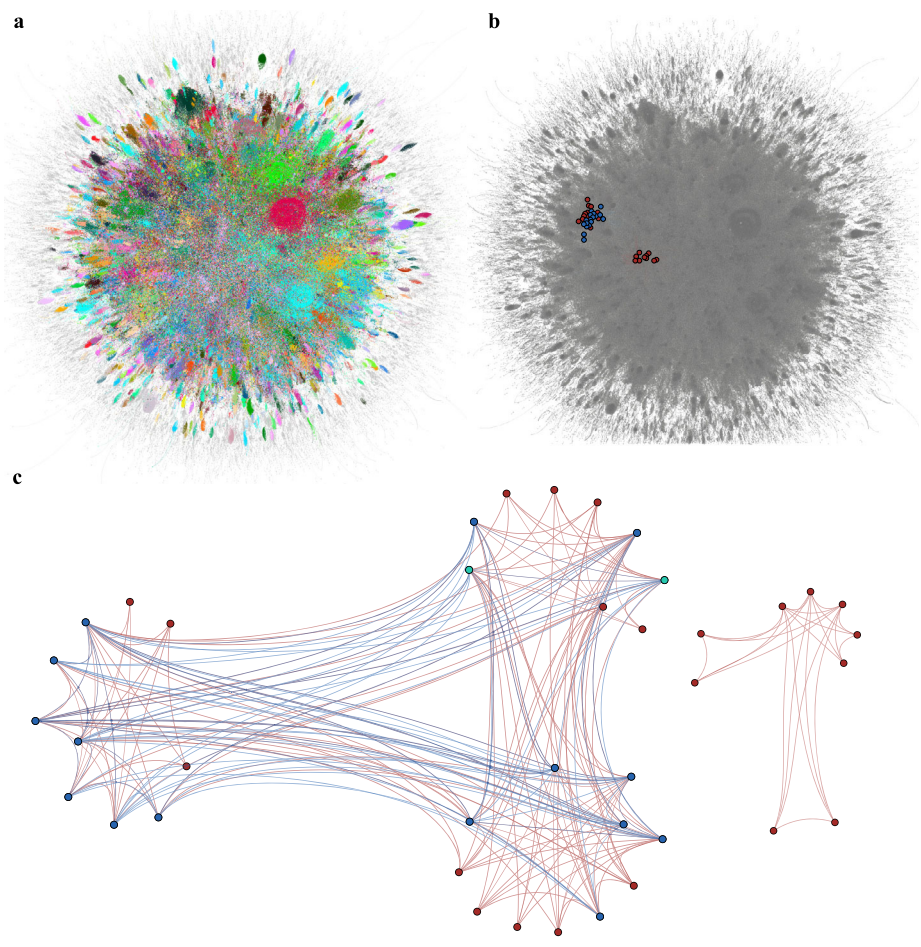

Figure 3 Reduced *k*-Cliques and Emerging Suspicious Groups in the Panama Papers Network.(**a**) Communities in the Panama Papers network. (**b**) Position of suspicious nodes in the network (zoom). (**c**) Simultaneous representation of all 5-cliques of the reduced network. Red nodes correspond to those that are also part of the reduced 6-cliques found in the network. Notice that here and in the accompanying figures, the 5 and 6-cliques are connected between themselves in the reduced network (i.e., they are not independent) and the figure shows all the connections between those nodes, hence there are more than 5 or 6-edges for many nodes

agents (central node) in the evolution analysis of Panama Papers communities (see [16]). For this reduced network of 5 and 6-cliques, this intermediate agent continues to be the most important one (Fig. 3c). The 5-cliques network includes agents of the 6-cliques one, but it was dominated by small agents (right network of Fig. 3c). We complemented this analysis with 2-neighborhood graphs of the embedded k-cliques. Figure 4a shows the 6-clique and how interactions between other agents and new groups appear with their features arising from the cliques' geometry and indicating possible money laundering patterns. Similarly, in the case of 5-cliques (Fig. 4b), new activities occur with small groups of agents.

We also obtained similar results on networks of other sizes, particularly the Bahamas network. Figure 5a and Fig. 5b show the suspicious nodes in their communities. Figure 5c shows the distribution of 4-clique agents in the network. These 4-cliques result from interactions between a few intermediaries and a group of offshore entities with a particularity: some of these have the same incorporate date and place (a money laundering pattern). In this case, one financial institution subsidiary interacts with one of these groups of cliques, but in other cases, intermediaries are non-financial intermediaries (Fig. 5c). It remains to check if some of these intermediaries have certain capital relations (shareholders) with financial institutions, although this requires further study and other methodologies.

Also, we identified these cliques in the Bahamas network and found new characteristics (Fig. 5a). First, some agents interacted with offshore entities created on the same date and place and a few times with financial and non-financial intermediaries, the cycle inside 4-clique (Fig. 5d). Additionally, we found interactions between nodes from different communities (appearing in the results of the previous section) with a particular characteristic: some of those communities were small, and they are not representative.

A summary of the number of maximal and total k-cliques found in all the reduced networks appears in Table 3. We conclude our analysis with some geometric considerations obtained from the study of the Forman–Ricci curvatures of the entire networks and the corresponding curvatures of the reduced cliques as they appear inside the original networks.

Figure 6 shows the distribution of Forman–Ricci curvature of the edges of the financial networks considered. It can be observed that most edges in these networks have negative curvature and broad distribution. By comparing distributions of edge curvatures in these networks, it is clear that the curvature distribution of the Panama network (Fig. 6b) decays very rapidly for negative values, similar to the behavior of a randomly constructed scale-free network of similar size whose Forman–Ricci curvature histogram is included for comparison in Fig. 6c. In contrast, the Bahamas network (Fig. 6a) has a curvature distribution that deviates more clearly from that basic scale-free model. A detailed analysis of the relation between deviation from scale-free behavior and the possible presence of fraudulent activities were considered in [16].

The Forman–Ricci curvature distributions for the embedded sets of 6-cliques and 5-cliques as they appear inside the original Panama Papers network appear in Fig. 7. Note that these Forman–Ricci curvature values are calculated for all the edges of those reduced cliques and not for their vertices, which explains why there are many more values than the number of vertices. From these results, the more important observation is that the edges accounted for in both histograms are divided into two well-separated groups, one whose

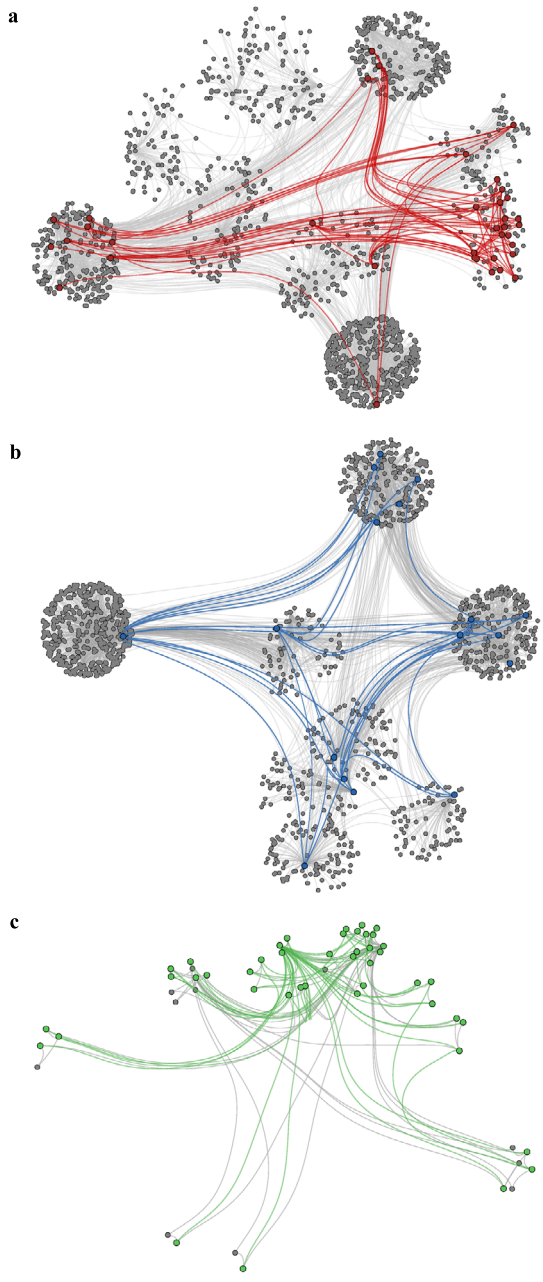

Figure 4 Reduced *k*-Cliques Embedded in the Unreduced Networks. (**a**) Simultaneous view of all reduced 5-cliques as they appear embedded into the original Panama Papers network after restoring their deleted 2-degree nodes and corresponding edges, thus they are not true cliques anymore but carry many connecting cycles. (**b**) Simultaneous view of all reduced 6-cliques as they appear embedded into the original Panama Papers network after restoring their deleted 2-degree nodes and corresponding edges. Although they are not true cliques in the unreduced network, they carry many connecting cycles. (**c**) Simultaneous view of all reduced 4-cliques as they appear embedded into the original Bahamas Leaks network after restoring their deleted 2-degree nodes and corresponding edges, in the original network they correspond to nodes that belong to many connecting cycles

edges have a negative curvature that is roughly half of the curvatures of the other group, showing a clear difference in the volume of interactions for each group.

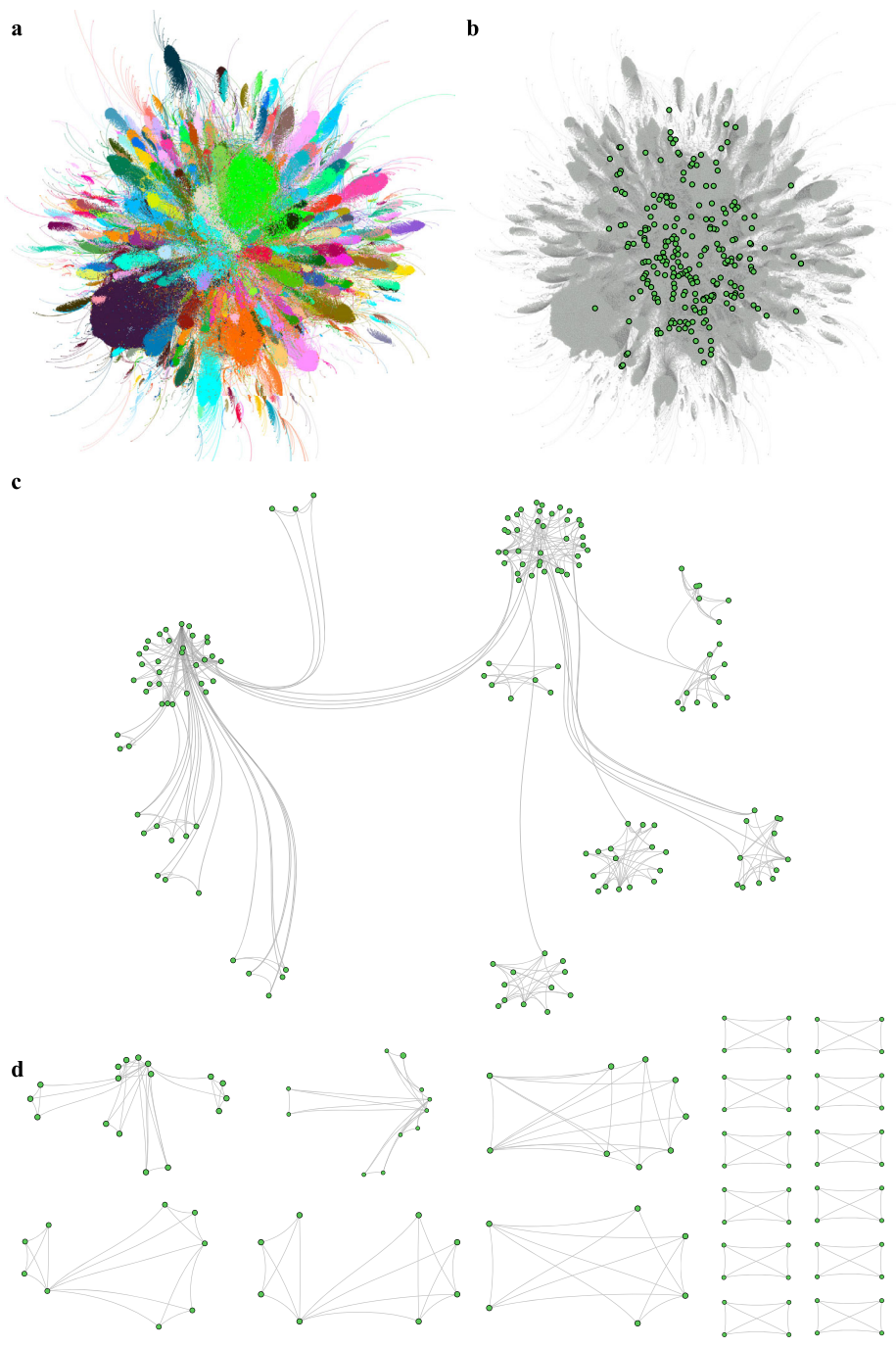

Figure 5 Reduced *k*-Cliques and Emerging Suspicious Groups in the Bahamas Leaks. (**a**) Communities in the Bahamas Leaks network. (**b**) Position of suspicious nodes in the Bahamas Leaks network (zoom). (**c**) Largest connected component of the set of all nodes belonging to 4-cliques of the reduced Bahamas Leaks network. (**d**) The remaining connected components of the set of nodes from 4-cliques of the reduced Bahamas Leaks network

This result shows that the groups of suspicious nodes under study are divided into two sets, one whose volume of connections to the rest of the network is higher than the other. This property could be considered a preliminary indication that these nodes satisfy the

Table 3 Number of maximal/total *k*-cliques in the reduced networks

Cliques sizes	Panama papers	Bahamas leaks
6-cliques	6/6	0/0
5-cliques	26/59	0/0
4-cliques	1100/1276	88/88
3-cliques	26,368/30,443	7577/7917

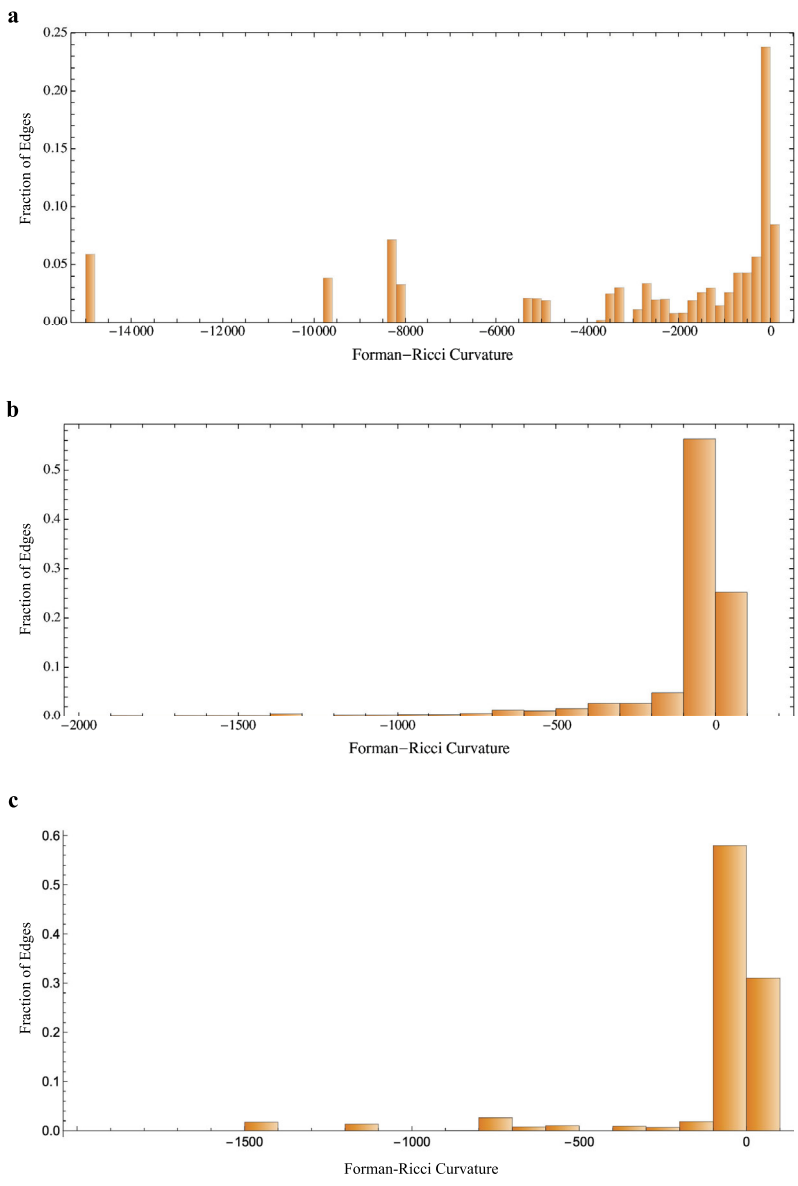

Figure 6 Forman–Ricci Curvature Histograms. The figures show histograms for the values of the Ricci–Forman curvature corresponding to the edges (i.e., interactions) of: (**a**) The Bahamas Leaks network. (**b**) The Panama Papers network. (**c**) A randomly generated scale-free network with comparable size

requirements established in the previously presented rules and particular Rule 1 of the Sect. 4.2 for lower-curvature nodes and Rule 4 for higher-curvature ones. The simultane-

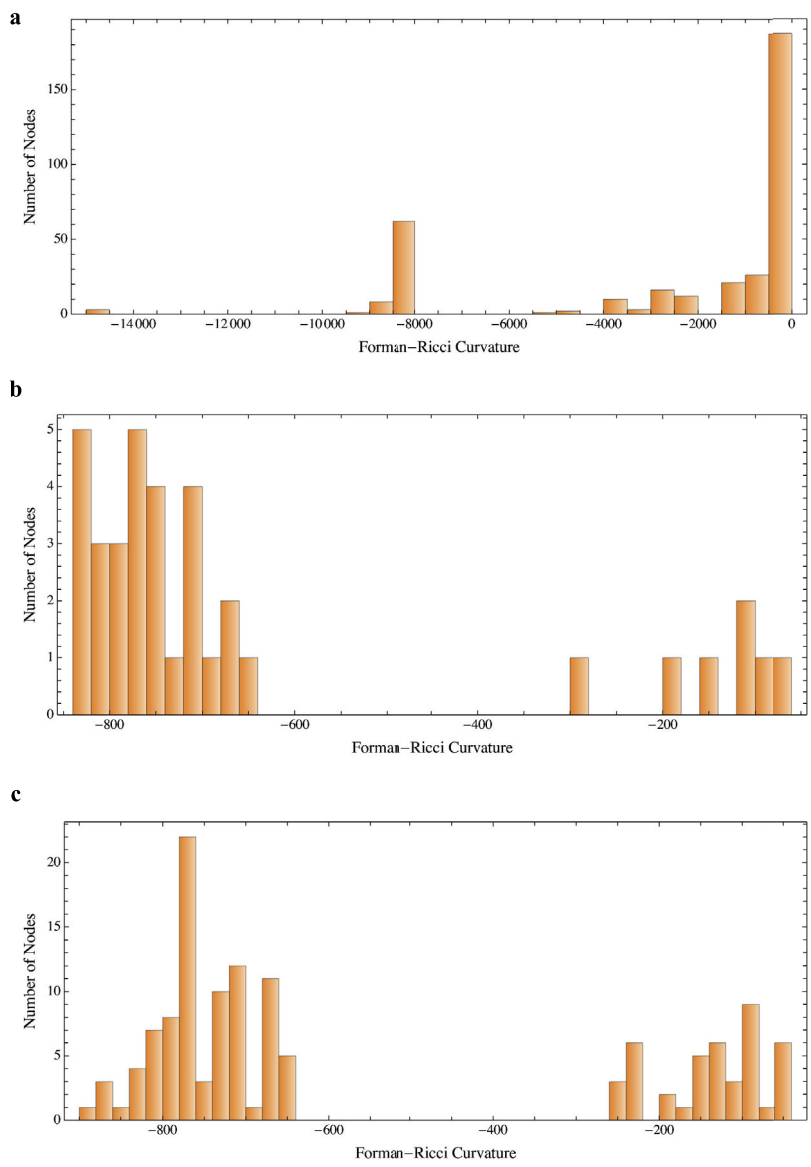

Figure 7 Forman–Ricci Curvature Histograms for Embedded Cliques in the Bahamas and the Panama Networks. The figures show histograms for the values of the Ricci–Forman curvature corresponding to the edges of the reduced *k*-cliques as they appear when viewed (embedded) inside the original network: (**a**) Values for the edges of all 6-cliques in Bahamas Leaks. (**b**) Values for the edges of all 6-cliques in Panama Papers. (**c**) Values for the edges of all 5-cliques in Panama Papers

ous presence of these two types of nodes reinforces the plausibility that those embedded cliques correspond to highly suspicious groups.

6 Conclusion

We analyzed suspicious interactions in financial network structures using information from the International Consortium of Investigative Journalists. As already observed in [16], very general aspects of this phenomenon in large-scale networks can be detected with several network science techniques. However, details of these crimes need specially adapted approaches because money launderers develop their activities through few in-

teractions and small groups that escape detection through standard methods like the determination of communities. As a result, many agents go unnoticed in these networks. The combination of topology and geometry tools can facilitate a new approach to analyzing money laundering in complex financial networks. Our proposed strategy allows us to identify common suspicious groups that lie at the root of global financial crime.

Although the complete characterization of money laundering activities through only topological considerations is not possible, we expect that Rules 1 to 5 provide a useful starting point to more refined developments. By its very definition, the strategy proposed here is in principle capable of detecting all groups of nodes that satisfy all the stated rules simultaneously. The difficulty lies in the large number of independent cliques that could appear, and their priority could be given by size. Nevertheless, it should be clear that no definite positives to money laundering can be given with our procedure, and particularly without further specific information on the involved interactions. The proposed strategy only allows determining suspicious groups that could be considered primary candidates for detailed scrutiny if more information is available.

For future work, a natural next step is to employ more refined topological and geometrical tools adjusted to this setting. More concretely, from the topological side, the use of homological techniques and invariants (e.g., Betti numbers, including the Euler invariant) can be used to measure and characterize the connectivity structure of the financial networks. At the same time, from the geometrical counterpart, further insights into the role of curvature in this context will provide a clearer picture of the suspicious interactions in these networks. A particular useful enhancement of the curvature considerations presented in this work is to extend the analysis with the use of augmented Forman–Ricci curvature by defining 2-simplices (or higher-order ones) in the network, taking advantage of the cyclic interactions that occur, or by using further information from the data or its observed behavior. In this way, the curvature values could encode more direct information concerning, for example, the suspicious interactions specified by Rule 5 of Sect. 4.2.

Even though a more detailed analysis is required, our findings provide a preliminary indication that the Forman–Ricci curvature can be used as a geometric tool to quantify how much a set of suspicious nodes satisfies some of the expected characteristics of money laundering interactions in these kinds of financial networks. Moreover, it is expected that its use can be highly enhanced once weights are given to agents and interactions using additional information from the data, which for the uniform treatment of the heterogeneous datasets employed was not viable in this study.

Acknowledgements
The authors would like to thank the anonymous referees for their valuable comments that helped to improve our manuscript.

Funding
This work was partially financed through the research project ID-20509 and ID-PROY 10439 of Vicerrectoría de Investigación, Pontificia Universidad Javeriana, Bogotá, Colombia.

Abbreviations
The article does not use abbreviations.

Availability of data and materials
The datasets analyzed during the current study come from the Offshore Leaks Database by the International Consortium of Investigative Journalists, available at: https://offshoreleaks.icij.org.

Declarations

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
OG designed the project. OG studied the financial structures. OG and AV studied the network structures and algorithms. AV studied the algorithm and applications of geometric and topological tools in the network setting. OG and AV wrote the manuscript. All authors read and approved the final manuscript.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 21 May 2021 Accepted: 17 January 2022 Published online: 08 February 2022

References
1. UNODC: Money-Laundering and Globalization (2018) https://www.unodc.org/unodc/en/money-laundering/globalization.html Accessed 2020-09-30
2. Walker J (1999) How big is global money laundering? J Money Laund Control 3(1):25–37. https://doi.org/10.1108/eb027208
3. Masciandaro D, Takáts E, Unger B (2007) Black finance. Edward Elgar Publishing, Cheltenham
4. Schneider F (2010) Turnover of organized crime and money laundering: some preliminary empirical findings. Public Choice 144(3/4):473–486
5. McCarthy KJ, van Santen P, Fiedler I (2015) Modeling the money launderer: microtheoretical arguments on anti-money laundering policy. Int Rev Law Econ 43:148–155. https://doi.org/10.1016/j.irle.2014.04.006
6. Imanpour M, Rosenkranz S, Westbrock B, Unger B, Ferwerda J (2019) A microeconomic foundation for optimal money laundering policies. Int Rev Law Econ 60:105856. https://doi.org/10.1016/j.irle.2019.105856
7. Walker J, Unger B (2009) Measuring global money laundering: "the Walker gravity model". Int Rev Law Econ 5
8. Dreżewski R, Sepielak J, Filipkowski W (2015) The application of social network analysis algorithms in a system supporting money laundering detection. Inf Sci 295:18–32. https://doi.org/10.1016/j.ins.2014.10.015
9. Colladon AF, Remondi E (2017) Using social network analysis to prevent money laundering. Expert Syst Appl 67:49–58. https://doi.org/10.1016/j.eswa.2016.09.029
10. Vaithilingam S, Nair M (2009) Mapping global money laundering trends: lessons from the pace setters. Res Int Bus Finance 23(1):18–30. https://doi.org/10.1016/j.ribaf.2008.03.003
11. Centeno MA, Nag M, Patterson TS, Shaver A, Windawi AJ (2015) The emergence of global systemic risk. Annu Rev Sociol 41(1):65–85. https://doi.org/10.1146/annurev-soc-073014-112317
12. Helbing D (2013) Globally networked risks and how to respond. Nature 497(7447):51–59. https://doi.org/10.1038/nature12047
13. Homer-Dixon T, Walker B, Biggs R, Crépin A-S, Folke C, Lambin EF, Peterson GD, Rockström J, Scheffer M, Steffen W, Troell M (2015) Synchronous failure: the emerging causal architecture of global crisis. Ecol Soc 20(3). https://doi.org/10.5751/ES-07681-200306
14. Andersen JJ, Johannesen N, Rijkers B (2020) Elite capture of foreign aid: Evidence from offshore bank accounts. World Bank. Working Paper 9150. http://hdl.handle.net/10986/33355
15. Kertész J, Wachs J (2020) Complexity science approach to economic crime. Nat Rev Phys. https://doi.org/10.1038/s42254-020-0238-9
16. Granados OM, Vargas A (2021) In: Granados OM, Nicolás-Carlock JR (eds) Financial networks and structure of global financial crime. Springer, Cham, pp 131–152. https://doi.org/10.1007/978-3-030-81484-7_8.
17. The International Consortium of Investigative Journalists (2017) Offshore Leaks Database. https://offshoreleaks.icij.org/ Accessed 2019-09-30
18. Noldus R, Van Mieghem P (2015) Assortativity in complex networks. J Complex Netw 3(4):507–542. https://doi.org/10.1093/comnet/cnv005
19. Newman M (2018) Networks. Oxford University Press, Oxford
20. Krioukov D, Papadopoulos F, Kitsak M, Vahdat A, Boguñá M (2010) Hyperbolic geometry of complex networks. Phys Rev E 82:036106. https://doi.org/10.1103/PhysRevE.82.036106
21. Wu Z, Menichetti G, Rahmede C, Bianconi G (2015) Emergent complex network geometry. Sci Rep 5:10073. https://doi.org/10.1038/srep10073
22. Forman (2003) Bochner's method for cell complexes and combinatorial Ricci curvature. Discrete Comput Geom 29(3):323–374. https://doi.org/10.1007/s00454-002-0743-x
23. Sreejith RP, Mohanraj K, Jost J, Saucan E, Samal A (2016) Forman curvature for complex networks. J Stat Mech Theory Exp 2016(6):063206. https://doi.org/10.1088/1742-5468/2016/06/063206
24. Weber M, Saucan E, Jost J (2017) Characterizing complex networks with Forman-Ricci curvature and associated geometric flows. J Complex Netw 5(4):527–550. https://doi.org/10.1093/comnet/cnw030
25. FATF: International Standards on Combating Money Laundering and the Financing of Terrorism & Proliferation (2021). www.fatf-gafi.org/recommendations.html Accessed 2021-01-30