Сергей В. Дуга[1], Виктория В. Ефимова[2], Андрей И. Труфанов[3]
[1]Судебно-экспертный центр Следственного комитета Российской Федерации,
Строителей ул., 8, корпус 2, Москва, 119313, Россия
[2]Следственное управление Следственного комитета Российской Федерации
по Иркутской области,
Володарского ул., 11, Иркутск, 664011, Россия
[3]Иркутский национальный исследовательский технический университет,
Лермонтова ул., 83, Иркутск, 664074, Россия
[1]e-mail: siber@list.ru, https://orcid.org/0000-0002-5894-9855
[2]e-mail: efimova.vika1977@mail.ru, https://orcid.org/0000-0003-3990-1917
[3]e-mail: troufan@gmail.com, https://orcid.org/0000-0002-6967-3495

АЛГОРИТМЫ СЕТЕВОГО АНАЛИЗА ДАННЫХ В РАСКРЫТИИ СХЕМЫ НАЛОГОВОГО ПРЕСТУПЛЕНИЯ

DOI: http://dx.doi.org/10.26583/bit.2022.1.08

Аннотация. В статье рассматривается возможность применения средств сетевого (графового) анализа при раскрытии схемы налогового преступления и формировании стратегии его расследования. Предложена модель данных, позволяющая построить сетевую топологию преступления основываясь как на непосредственных материалах о событии преступления, так и на дополнительной информации, полученной путем доступа к базам данных правоохранительных и контрольных органов. Рассмотрены алгоритмы сетевого анализа, способствующие выявлению новых сведений о преступной схеме. Данные сетевые алгоритмы позволяют обнаружить скрытые и неочевидные связи между фигурантами дела, выявить иерархическую структуру их отношений, что способствует установлению ключевых участников преступной схемы. На примерах различных уголовных дел налоговых преступлений показано, что применение сетевого анализа данных позволяет сформировать схему преступления, дать правильную криминалистическую характеристику преступления, определить круг его субъектов, предложить следователю криминалистическую методику расследования данного вида преступлений (алгоритм расследования) и типовые следственные версии, что в совокупности способствует организации расследования должным образом.

Ключевые слова: сетевой анализ, налоговые преступления, схема налогового преступления, алгоритмы сетевого анализа, расследование преступлений.

Для цитирования: ДУГА, Сергей В.; ЕФИМОВА, Виктория В.; ТРУФАНОВ, Андрей И. АЛГОРИТМЫ СЕТЕВОГО АНАЛИЗА ДАННЫХ В РАСКРЫТИИ СХЕМЫ НАЛОГОВОГО ПРЕСТУПЛЕНИЯ. Безопасность информационных технологий, [S.l.], т. 29, № 1, с. 82–93, 2022. ISSN 2074-7136. URL: https://bit.mephi.ru/index.php/bit/article/view/1409. DOI: http://dx.doi.org/10.26583/bit.2022.1.08.

Sergey V. Duga[1], Viktoriya V. Efimova[2], Andrey I. Trufanov[3]
[1]Forensic Expert Center of the Investigative Committee of the Russian Federation,
Stroitelej str., 8, k. 2, Moscow, 119313, Russia
[2]Investigative Committee of the Russian Federation Irkutsk Region,
Volodarskogo str., 11, Irkutsk, 664011, Russia
[3]Irkutsk National Research Technical University,
Lermontova str., 83, Irkutsk, 664074, Russia
[1]e-mail: siber@list.ru, https://orcid.org/0000-0002-5894-9855
[2]e-mail: efimova.vika1977@mail.ru, https://orcid.org/0000-0003-3990-1917
[3]e-mail: troufan@gmail.com, https://orcid.org/0000-0002-6967-3495

Algorithms of network data analysis in the disclosure of a tax crime scheme

DOI: http://dx.doi.org/10.26583/bit.2022.1.08

Abstract. The paper considers the possibility of using network (graph) analysis tools in the disclosure of a tax crime scheme and the formation of a strategy for its investigation. A data model is proposed that allows building a network topology of a crime based both on direct material about a crime event and on additional information obtained by accessing the databases of law enforcement and control agencies. The algorithms of network analysis that contribute to the identification of new information about the criminal scheme are also considered. These network algorithms make it possible to detect hidden and non-obvious connections between the defendants in the case, to identify the hierarchical structure of their relationships, which helps to identify the key participants in the criminal scheme. Using the examples of various criminal cases of tax crimes, it is shown that the use of the considered variants of network data analysis allows forming a crime scheme, providing the correct criminalistic characterization of the crime, determining the range of its subjects, offering the investigator a forensic methodology for investigating this type of crime (investigation algorithm) and standard investigative versions, that all together contribute to the proper organization of the investigation.

Keywords: mathematical modeling network analysis, tax crimes, tax crime scheme, network analysis algorithms, crime investigation.

For citation: DUGA, Sergey V.; EFIMOVA, Viktoriya V.; TRUFANOV, Andrey I. Algorithms of network data analysis in the disclosure of a tax crime scheme. IT Security (Russia), [S.l.], v. 29, n. 1, p. 82–93, 2022. ISSN 2074-7136. URL: https://bit.mephi.ru/index.php/bit/article/view/1409. DOI: http://dx.doi.org/10.26583/bit.2022.1.08.

Введение

Расследование преступления – сложный интеллектуальный процесс, в котором задействованы множество механизмов. Специалист в процессе расследования сталкивается со значительным числом трудностей, главный из которых – ограниченность временных ресурсов для установления всех обстоятельств преступления и завершения расследования. У каждого вида преступлений своя специфика и особенности при выборе стратегии и тактики расследования. Вместе с тем выделяется категория «интеллектуальных» преступлений, раскрытие которых напрямую зависит от правильности сбора первичного аналитического материала и способностей его анализа. В их числе налоговые преступления.

Предложенные в научной и практической литературе алгоритмы расследования налоговых преступлений не могут претендовать на всеохватность. Каждое преступление индивидуально, а применяемые криминальные схемы уникальны. Поэтому на первый план выходит способность подхода к сбору и анализу информации, формированию схемы преступления. От правильно построенной схемы налогового преступления зависит оперативность в расследовании и возможность применения превентивных мер к их совершению, обеспечение возмещения ущерба, причиненного государству.

В настоящее время на практике следователь является единственным лицом, который фиксирует ход следствия в установленном законом порядке, проводит аналитику собранных доказательств и обеспечивает выявление новых составов преступлений. Многие правоохранительные и контрольные органы имеют специализированные подразделения, включающие группы лиц, занимающихся отдельными направлениями процесса по выявлению фактов уклонения от уплаты налогов, и осуществляющих аналитику. Следователь такой поддержки лишен, так как в структуре Следственного комитета РФ отсутствуют собственные оперативные подразделения, а также аналитические службы, которые могли бы облегчить следователю процесс выявления преступления и формирования доказательств. При этом сбор информации для анализа, позволяющий отнести его результаты к доказательствам, осуществляется следователем в отсутствие каких-либо специфических программных комплексов, в так называемом «ручном» режиме, что существенно растягивает во времени процесс расследования.

Доказывание начинается со сбора информации для формирования схемы преступления. На данном этапе важнейшее значение приобретает оперативность получения

значимой информации из различных источников, формирование собственной базы данных, содержащей сведения об участии потенциальных субъектов схемы в преступной деятельности, связь таких субъектов с иными субъектами в схеме.

Традиционно, налоговые преступления отличаются длительностью их совершения, системным характером преступных действий и их проработанностью. Сбор аналитического материала для выявления и расследования налогового преступления осуществляется из огромного числа источников, наиболее значимые из них – базы данных налоговых, регистрирующих органов, коммерческих банков, таможенных органов, судов. Получение информации из таких источников в настоящее время занимает значительное время, осуществляется в формате «запрос-ответ» в бумажном виде, имеет ограниченный формат. Впоследствии полученные сведения следователь аккумулирует и обобщает, составляя схему преступления.

Применение средств сетевого анализа при раскрытии схемы налогового преступления и формировании стратегии и тактики расследования способно существенно облегчить процесс доказывания, оптимизировать работу следователя и сократить сроки расследования уголовных дел.

1. Обзор работ, близких по тематике

В настоящее время, со стороны исследователей всего мира, наблюдается рост интереса к применению новейших методов анализа данных для борьбы с преступностью в общем контексте [1–4], так и для расследования налоговых преступлений [5–7].

В [8] предложена аналитическая система для выявления влиятельных членов преступной организации. Подход состоит из последовательных этапов:

1) построение сети (сеть создается либо из данных мобильной связи, поддерживаемой преступной организацией, либо из отчетов о преступлениях, содержащих информацию о членах преступной организации),

2) назначение веса каждой связи в сети (вес связи представляет собой количество телефонных звонков/сообщений между двумя преступниками),

3) вычисление кратчайшего пути по степени посредничества (мера, которая отражает значимость узла (вершины) при передаче информации из одной части сети в другую),

4) присвоение оценки каждому узлу в сети на основе концепции зависимости существования. Преступники, представленные узлами (вершинами) высшего ранга, считаются влиятельными членами преступной организации.

В [9] представлена аналитическая система «ATTENet», предназначенная для обнаружения и объяснения подозрительных групп уклонения от уплаты налогов на основе аффилированных транзакций. Для решения задачи, во-первых, система создает сеть, которая включает данные о налогах и налогоплательщиках из официальной налоговой базы данных. Затем система объединяет основные характеристики и особенности структуры каждой группы в сети методом «Structure2Vec», после чего, с использованием алгоритма «Random Forest», обнаруживает подозрительные группы. Наконец, для изучения и объяснения результатов, система предоставляет визуализацию с интерактивными инструментами.

2. Используемые источники данных

Для наполнения информационной системы сведениями, нами используются различные источники. Кратко перечислим их:

– материалы уголовных дел. На текущий момент используется система «Pullenti» [10] для извлечения именованных сущностей и семантического анализа материалов уголовных дел, в частности протоколов допросов. На рис. 1 представлен пример такого анализа. Из первоначального текста извлекаются именованные сущности, а также, в результате семантического анализа, строится сетевая модель;

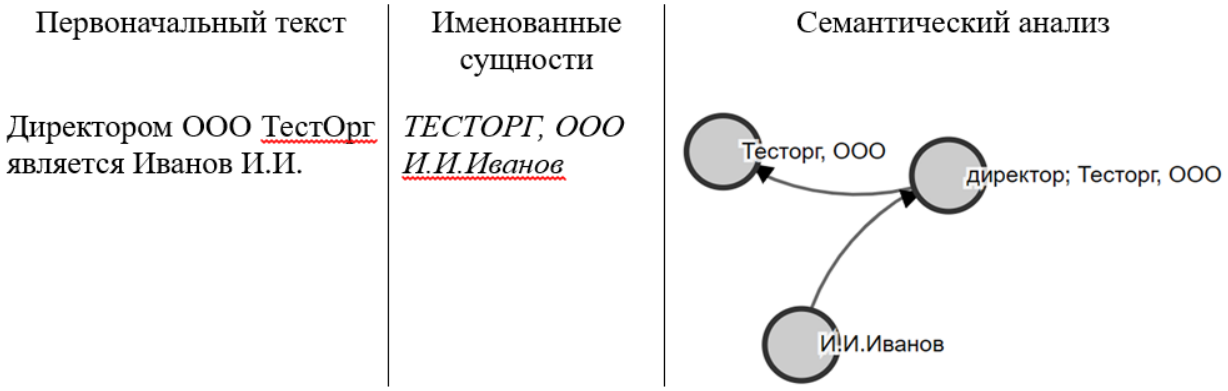

Рис. 1. Пример анализа текста с использованием системы «Pullenti»
Fig. 1. An example of text analysis using the "Pullenti" system

– данные из мобильных телефонов. По результатам осмотров мобильных телефонов фигурантов уголовных дел формируются отчеты, которые, в последующем, загружаются в систему. Использование средств коммуникации, таких как телефонные звонки и мессенджеры, оставляют цифровые следы, которые можно использовать для анализа. Это позволяет следователям лучше понимать внутреннюю иерархию преступных организаций, обнаруживая субъектов, которые играют центральную роль и/или обеспечивают связь между подгруппами;

– сведения от налоговых органов.

Также в систему загружаются сведения из программ учета финансово-хозяйственной деятельности, полученные в ходе производства осмотров электронных носителей информации, данные из единой информационной системы в сфере закупок, результаты арбитражных процессов.

3. Модель данных

Налоговое преступление, как правило, представляет собой сложную систему – совокупность объектов и субъектов, взаимодействующих друг с другом нетривиальным образом. При расследовании данного вида преступлений важно проводить анализ сети в целом, не сосредоточивая внимание на отдельных субъектах. Такой тип оценки может существенно выиграть от применения сетевого анализа, когда лица рассматриваются как акторы, которые связаны друг с другом в рамках взаимозависимой системы.

Чтобы изучить сложные системы в разных дисциплинах и областях, первым шагом является конкретное представление системы с использованием унифицированного математического языка. Кроме того, эти формализмы позволяют конструировать эффективные алгоритмы и могут использоваться для определения структуры, функции и динамики системы. Представляя исследуемые структуры в виде сетей, можно применять различные математические и сетевые методы для количественной оценки и выявления структурных особенностей.

В данном исследовании используется сетевая модель, которую можно представить в виде кортежа (*V, E, L(V), T(E), X*), представляющая собой неориентированный граф, вершины и ребра которого связаны с одной или несколькими метками, где:

V – множество вершин,

E – множество ребер,

L(V) – сюръективное отображение (v , l), которое связывает вершины с метками, такое, что каждой вершине $v \in V$, соответствует хотя бы одна метка $l \in L$,

T(E) – сюръективное отображение (e , t), которое связывает ребра с их типами, такое, что каждому ребру $e \in E$ соответствует хотя бы один тип $t \in T$.

Кроме того, каждая вершина связана с соответствующим вектором признаков. Здесь *X* является матрицей признаков для графа ($X \in \Re^{N \times D}$), так, что *i*-я строка *X* является вектором признаков для узла $v_i (i = 1,2 \ldots .|V|)$.

Для дальнейшего численного анализа определим матрицу смежности ($A^{N \times N}$) данного графа, такую что:

$$A_{ij} \begin{cases} 1, \text{если существует связь между вершинами } i \text{ и } j \\ 0, \quad \text{в ином случае} \end{cases}.$$

Кроме того, используются двудольные сети для разделения неоднородного графа на однородные подграфы, с целью последующего анализа. В двудольных графах вершины разбиты на два непересекающихся подмножества, так, что связи (ребра) могут возникать только в том случае, если вершины принадлежат разным множествам. Применение двудольных сетей (графов) может использоваться для представления связей субъект-организация, когда один тип представляет исходные узлы (например, люди), а другой представляет группы, к которым принадлежит первый тип узлов (например, организация). Двудольная сеть и ее граф определяется матрицей инцидентности ($A^{N \times G}$). Например, если имеется *n* субъектов и *g* организаций:

$$A_{ij} \begin{cases} 1, \text{если } j \text{ субъект принадлежит к организации } i \\ 0, \text{ в ином случае} \end{cases}.$$

Основываясь на предложенных в [11, 12] моделях, узлы используются для представления, таких сущностей как:

– Субъекты. Физические лица, имеющие отношение к расследуемому уголовному делу.

– Организации. Юридические лица, имеющие отношение к расследуемому уголовному делу.

– События – время, место, способ и другие обстоятельства совершения преступления, обстоятельства, способствовавшие совершению преступления (УПК РФ статья 73), а также иные обстоятельства, имеющие значение для уголовного дела (встречи людей, телефонные звонки, передача данных и пр.).

– Объектами могут быть любые предметы, которые служили орудиями, оборудованием или иными средствами совершения преступления, предметы и документы, которые могут служить средствами для обнаружения преступления и установления обстоятельств уголовного дела (УПК РФ статья 81).

– Места – место совершения преступления, домашний/рабочий адрес человека, адрес регистрации юридического лица и пр.

В зависимости от типов узлов используются различные связи между ними, например, связь между организацией и физическим лицом, а также между организациями, может быть установлена из программ учета финансово-хозяйственной деятельности

организаций, сведений, полученных из налогового органа или вручную. В свою очередь, связь между физическими лицами может быть охарактеризована различными типами: родственники, коллеги, частое общение, общаются редко. Тип связи задается вручную, или определяется автоматически на основе анализа телефонных соединений, обмена сообщениями и иных источников. Для иллюстрации изложенного на рис. 2 приведена часть концептуальной модели данных, сформированной в графовой базе данных «Neo4j».

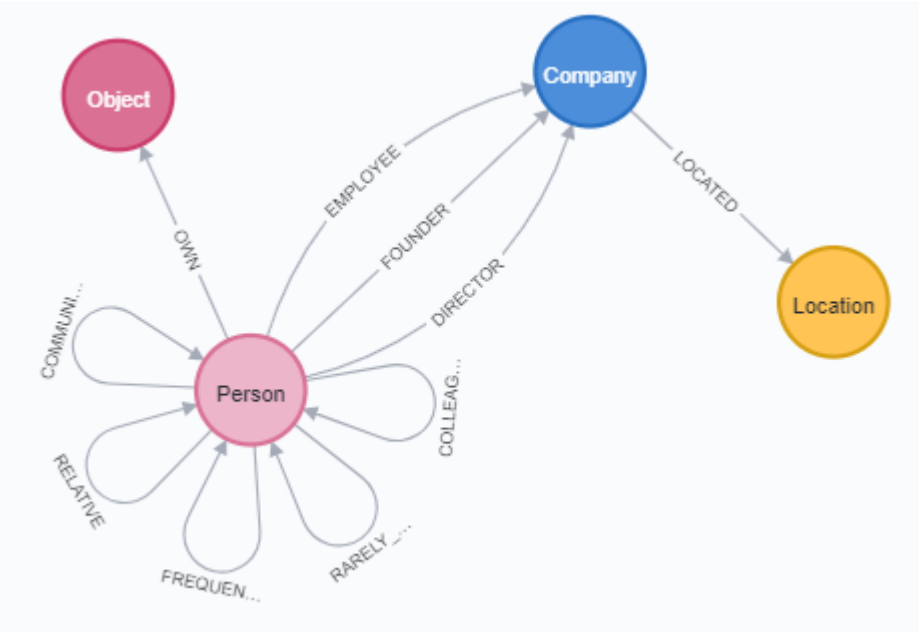

Рис. 2. Часть концептуальной модели данных
Fig. 2. Part of the conceptual data model

Сеть строится на основе данных, получаемых в рамках расследования уголовного дела, предполагая, как ручное внесение данных, так и автоматическое извлечение сведений из различных источников.

4. Используемые алгоритмы сетевого анализа

Анализ отношений между физическими и юридическими лицами представляет собой выявление и определение степени (важности) личных связей, имущественных прав. Такие связи могут быть как формальными, так и неформальными. Формальные связи легко проверить с помощью соответствующих документов и реестров. Неформальными связи можно определить на основе различных данных, полученных в ходе расследования преступления, например сведения о телефонных соединениях или финансовые операции. Анализ, основанный на различных источниках данных, также позволяет определить роль отдельных узлов во всей сети.

В ходе анализа связей необходимо, в частности, выявить следующие признаки аффилированности:

– участие одних и тех же лиц в создании, управлении организациями, в том числе через родственников или подконтрольных лиц;

– пересечение по работе в одной и той же организации, т.е. либо одновременно являются работниками организации, либо работали в ней в разное время;

– совпадение адресов государственной регистрации организаций.

5. Прогноз существования связи

Данную задачу можно сформулировать следующим образом:

Если имеется набор данных в виде сети $G = (V, E)$, где V – множество узлов, E – множество наблюдаемых связей, то задача заключается в том, чтобы предсказать, насколько вероятно существование ненаблюдаемой связи ($e_{ij} \notin E$) между произвольной парой узлов v_i, v_j.

Для решения подобных задач используются различные методы [13–15]. Одним из таких методов является метод «Общие соседи» (Common Neighbors) [16]. В основе данного метода лежит предположение, что два узла с большой вероятностью связаны между собой (они будут связаны в ближайшем будущем), если у них много общих соседей:

$$score(x, y) = |\Gamma(x) \cap \Gamma(y)|,$$

где $\Gamma(x)$ –набор узлов, смежных с узлом x, а $\Gamma(y)$ –набор узлов, смежных с узлом y.

Применение данного метода целесообразно ввиду того, что, зачастую, у следователя есть только частичная информация для анализа.

Несмотря на свою простоту, метод хорошо работает в большинстве реальных сетей и превосходит сложные подходы [17]. Например, при расследовании уголовного дела по факту хищения бюджетных средств путем незаконного возмещения НДС с деятельности ООО Б. в особо крупном размере решение о потенциальному круге фигурантов уголовного дела (Р., Ф., С.) сделан следователем с учетом данных о связях лица Р., заявленного в уставных документах в качестве руководителя организации, с иными лицами, хотя и не имеющими официального отношения к Обществу, но фактически являющимися его руководителями.

В ходе расследования были получены сведения от налогового органа о лицах, заявленных в качестве работников ООО Б., сведения из базы ЕГРЮ, сведения от кредитных организаций по месту открытия расчетных счетов общества. На основании полученных сведений была построена сетевая модель преступления и применен метод прогноза существования связи, который показал высокую вероятность наличия связи между ООО Б. и физическими лицами Ф. и С.

6. Использование метрики «центральность по собственному вектору»

Целью данной метрики является определение наиболее значимых фигурантов уголовного дела путем присваивания веса каждой вершине в сети посредствам использования алгоритмов центральности [18–21].

На основе предложенной модели вычисляются степени значимости (определение важности отдельных узлов в сети) как классическими методами сетевого анализа, так и методами, адаптированными для предметной области (предварительное следствие). Основная цель этой оценки – выявить те субъекты, которые с большей вероятностью будут причастны к деятельности по уклонению от уплаты налогов. Это особенно полезно, когда нет надежной информации, которая позволила бы идентифицировать соответствующих фигурантов. Выявление влиятельных членов преступной организации – одна из важнейших задач, которую берут на себя следователи по уголовным делам. Возможное решение – выявить наиболее значимых субъектов при помощи алгоритма центральности собственного вектора на основе матрицы смежности (Eigenvector Centrality) [22]. Оценка центральности (значимости) вершины пропорциональна сумме оценок всех вершин, соединенных с ней. Таким образом, если вершина соединена со многим значимыми вершинами, она также будет считаться значимой. Математически это можно выразить уравнением $Ax = \lambda x$, где A – матрица смежности с собственным значением λ (согласно теореме Перрона-Фробениуса

λ – наибольшее собственное значение матрицы смежности), x собственный вектор этой матрицы.

В рамках выше описанного примера, расследования уголовного дела по факту хищения бюджетных средств путем незаконного возмещения НДС с деятельности ООО Б. в особо крупном размере, решение об организаторе преступления и его исполнителях, сделано следователем с учетом информации о центральности указанных фигур в связи с большим количеством фактов хозяйственной деятельности ООО Б. по сравнению с иными потенциальными субъектами преступления: участие в создании ООО Б. и их аффилированных организаций, использованных в схеме преступления (из информации регистрационных дел налогового органа); представление интересов ООО Б. и аффилированных к нему организаций, участвующих в схеме преступления во взаимодействии с налоговыми органами (информация налогового органа по запросу следственного органа, анализ осмотра документов, изъятых в ходе обысков, выемок); участие в распоряжении денежными средствами ООО Б. и аффилированных к нему организаций, участвующих в схеме преступления (по данным юридических дел кредитных организаций); осуществление деятельности по адресу отправления налоговой отчетности по системе телекоммуникационной связи в налоговых орган (по данным налогового органа по IP-адресам отправки налоговой отчетности указанных организаций); наличие сведений о согласовании текущих вопросов деятельности ООО Б. и аффилированных к нему организаций, участвующих в схеме преступления (информация по результатам осмотров документов учета, изъятых в ходе обыска, выемок).

7. Поиск изоморфного подграфа

Многие реальные случаи уклонения от уплаты налогов реализуются посредством взаимодействия нескольких субъектов. Задача состоит в том, чтобы на основе сетевой онтологии, описывающей сетевые шаблоны подозрительной деятельности, извлекать топологическую информацию из комплексной сети, также известную как подграфы в анализе графов. Использование данного метода позволяет характеризовать, в частности, социальные связи, экономические взаимоотношения. Для примера приведём некоторые критерии аффилированности субъектов предпринимательской деятельности:

– участие одних и тех же лиц в создании, управлении организациями, в том числе через родственников или подконтрольных лиц;

– большая часть операций по расчетному счету приходится на взаимоотношения с аффилированной организацией;

– один адрес государственной регистрации.

Применения алгоритмов поиска сетевых шаблонов широко используется для выявления подозрительных групп при уклонении от уплаты налогов [23–27].

На основе используемой онтологии (рис. 3) сформулированы запросы на графовом языке запросов «Cypher». Используя данные запросы, можно, в частности, получить организации, связанные с интересующей организацией опосредованно. Данный алгоритм можно описать следующим образом:

1. указать интересующую организацию;
2. получить организации, связанные с интересующей организацией на удалении, например, до трех перемещений (прыжков);
3. для каждой организации, полученной на шаге 2, найти смежные вершины;
4. для каждого множества смежных вершин, полученных на шаге 3, вычислить коэффициент сходства Жаккарда с множеством смежных вершин интересующей организации.

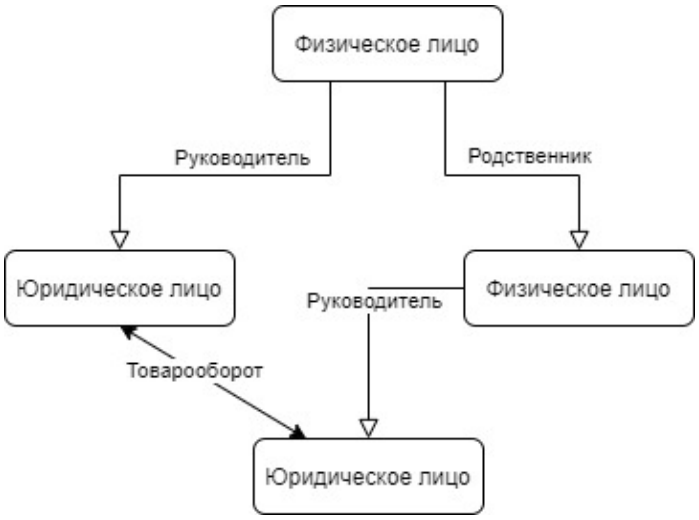

Рис. 3. Часть используемой онтологии
Fig. 3. Part of the ontology used

Коэффициент Жаккара (Jaccard Similarity) [28] в настоящее время является наиболее часто используемой мерой подобия в контексте анализа поведенческих связей, сходства между преступлениями. Привлекательность данного метода, отчасти, заключается в его простоте. Математически его можно выразить уравнением:

$$J = \frac{c}{a+b-c},$$

где a – множество смежных вершин интересующей организацией, b – множество смежных вершин организации (одной из), полученной на шаге 3 описанного выше алгоритма, c – пересечение этих множеств.

Применение данного алгоритма, в ходе расследования уголовного дела по факту уклонения от уплаты налогов с деятельности ООО Ж. в особо крупном размере путем необоснованного применения льготы по уплате НДС, позволило выявить признаки подконтрольности девяти организаций, формально не имеющих отношений к деятельности ООО Ж., однако, фактически с ним связанных. На основании информации, полученной из регистрирующего органа по результатам анализа документов регистрационных дел указанных организаций, установлено, что участниками указанных организаций являются лица, ранее являвшиеся административными работниками ООО Ж.; указанные организации также как и ООО Ж. имеют один адрес государственной регистрации; из информации, полученной из налогового органа установлено, что все организации имеют связь по месту осуществления бухгалтерского учета, формирования и направления налоговой отчетности; идентичными оказались и сведения о месте выхода в сеть по системе электронных расчетов Банк-Клиент всех указанных организаций, включая ООО Ж. Выявленные связи позволили сделать вывод о наличии подозрительной группы, состоящей из десяти субъектов (ООО Ж. и девять выше указанных организаций), использованных в преступной схеме.

Заключение

На примере расследования реальных уголовных дел продемонстрирована эффективность применения используемых алгоритмов в рассматриваемой предметной области. Показано, что использование алгоритмов сетевого анализа, таких как поиск изоморфного подграфа, предсказания связи, а также методов анализ социальных сетей, в частности определение наиболее значимых фигурантов, может способствовать значительному сокращению сроков расследования уголовного дела, выявить неочевидные/скрытые факты.

Предложенные алгоритмы позволяют анализировать собранную в рамках расследования информацию, включая сведения электронных баз данных учета налогоплательщика, данные электронных переписок, телефонных переговоров, соединений конкретных лиц, пр., позволяют при наличии межведомственного взаимодействия оперативно получать запрашиваемую информацию у различных источников (налоговые органы, органы внутренних дел, иные контрольные и правоохранительные органы), имеют существенное практическое значение для выявления и расследования налоговых преступлений, поскольку позволяют формировать схему налогового преступления, изменять ее с учетом новой криминалистической информации, обрабатываемой с применением информационных технологий, и на ее основе выдвинуть обоснованную и максимально достоверную в соответствии с установленным набором входных данных следственную версию события преступления.

СПИСОК ЛИТЕРАТУРЫ:

1. Bogahawatte K., Adikari S. 8th International Conference on Computer Science & Education. Intelligent criminal identification system, Colombo, Sri Lanka. 2013, p. 633–638. DOI: http://dx.doi.org/10.1109/ICCSE.2013.6553986.
2. Hamdy E., Adl A., Hassanien A.E., Hegazy O. and Kim T. -H. Criminal Act Detection and Identification Model, 2015 Seventh International Conference on Advanced Communication and Networking (ACN). 2015, p. 79–83. DOI: http://dx.doi.org/10.1109/ACN.2015.30.
3. Nath S.V. Crime Pattern Detection Using Data Mining, 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops. 2006, p. 41–44. DOI: http://dx.doi.org/10.1109/WI-IATW.2006.55.
4. Saravanan P., Selvaprabu J., Arun Raj L., Abdul Azeez Khan A., Javubar Sathick K. (2021) Survey on Crime Analysis and Prediction Using Data Mining and Machine Learning Techniques. In: Zhou N., Hemamalini S. (eds) Advances in Smart Grid Technology. Lecture Notes in Electrical Engineering, vol 688. Springer, Singapore. DOI: http://dx.doi.org/10.1007/978-981-15-7241-8_31.
5. Jihal H., Ounacer S., Ardchir S., Azouazi M. (2020) Clustering Model of False Positive Elimination in Moroccan Fiscal Fraud Detection. In: Ezziyyani M. (eds) Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). Advances in Intelligent Systems and Computing, vol 1104. Springer, Cham. DOI: http://dx.doi.org/10.1007/978-3-030-36671-1_12.
6. Wu Y., Dong B., Zheng Q., Wei R., Wang Z. and Li X. A Novel Tax Evasion Detection Framework via Fused Transaction Network Representation, 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). 2020, p. 235–244. DOI: http://dx.doi.org/10.1109/COMPSAC48688.2020.00039.
7. Didimo W., Grilli L., Liotta G., Menconi L., Montecchiani F. and Pagliuca D. Combining Network Visualization and Data Mining for Tax Risk Assessment, in IEEE Access. Vol. 8, p. 16073–16086, 2020. DOI: http://dx.doi.org/10.1109/ACCESS.2020.2967974.
8. Taha K., Yoo P.D. Using the Spanning Tree of a Criminal Network for Identifying Its Leaders, in IEEE Transactions on Information Forensics and Security. Vol. 12, no. 2, p. 445–453, 2017. DOI: http://dx.doi.org/10.1109/TIFS.2016.2622226.
9. Zheng Q. et al. ATTENet: Detecting and Explaining Suspicious Tax Evasion Groups. «Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence». P. 6584–6586, 2019. DOI: http://dx.doi.org/10.24963/ijcai.2019/964.
10. Золотарёв О.В. и др., Система PullEnti – извлечение информации из текстов естественного языка и автоматизированное построение информационных систем. Ситуационные центры и информационно-аналитические системы класса 4i для задач мониторинга и безопасности, тр. межд. научн. конф., т. 2, ИФТИ, Протвино. 2016, с. 28–35. URL: https://www.elibrary.ru/item.asp?id=28185224 (дата обращения: 30 03 2021).
11. Announcing the Neo4j Crime Investigation Sandbox. URL: https://medium.com/neo4j/announcing-the-neo4j-crime-investigation-sandbox-c0c3bd9e71b1 (дата обращения: 30 03 2021).
12. Neo4j and the Offshore Leaks: the Case of Azerbaijan. URL: https://neo4j.com/graphgist/neo4j-and-the-offshore-leaks-the-case-of-azerbaijan (дата обращения: 30 03 2021).
13. Adamic L.A. и Adar E., Friends and neighbors on the web Social networks. Vol. 25, no. 3, p. 211–230, 2003. DOI: http://dx.doi.org/10.1016/S0378-8733(03)00009-1.

14. Zhou T., Lü L., Zhang Y.C. Predicting missing links via local information. The European Physical Journal. Vol. 71, no. 4, p. 623–630, 2009. DOI: http://dx.doi.org/10.1140/epjb/e2009-00335-8.
15. Barabâsi A.L. et al. Evolution of the social network of scientific collaborations. Physica A: Statistical mechanics and its applications. Vol. 311, no. 3-4, p. 590–614, 2002. DOI: http://dx.doi.org/10.1016/S0378-4371(02)00736-7.
16. Newman M.E.J. Clustering and preferential attachment in growing networks. Physical review. Vol. 64, no. 2, 2001. DOI: http://dx.doi.org/10.1103/PhysRevE.64.025102.
17. Martínez V., Berzal F., Cubero J C. A survey of link prediction in complex networks. ACM computing surveys (CSUR). Vol. 49, no. 4, p. 1–33, 2016. DOI: http://dx.doi.org/10.1145/3012704.
18. Rungsawang A., Manaskasemsak B. An Efficient Partition-Based Parallel PageRank Algorithm. Proceedings of the 11th International Conference Parallel and Distributed Computing, 2004. DOI: http://dx.doi.org/10.1109/ICPADS.2005.85.
19. Brandes U. A faster algorithm for betweenness centrality. Journal of mathematical sociology. Vol. 25, no. 2, p. 163–177, 2001. DOI: http://dx.doi.org/10.1080/0022250X.2001.9990249.
20. Sariyüce A.E., Kaya K., Saule E and Çatalyiirek Ü.V. Incremental algorithms for closeness centrality, 2013 IEEE International Conference on Big Data. 2013, p. 487–492. DOI: http://dx.doi.org/10.1109/BigData.2013.6691611.
21. Bihari A. and Pandia M.K. Eigenvector centrality and its application in research professionals' relationship network, 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE). 2015, p. 510–514. DOI: http://dx.doi.org/10.1109/ABLAZE.2015.7154915.
22. Li X. et al. Identifying social influence in complex networks: A novel conductance eigenvector centrality model. Neurocomputing. Vol. 210, p. 141–154, 2016. DOI: http://dx.doi.org/10.1016/j.neucom.2015.11.123.
23. Didimo W. et al. Visual querying and analysis of temporal fiscal networks. Information Sciences. Vol. 505, p. 406–421, 2019. DOI: http://dx.doi.org/10.1016/j.ins.2019.07.097.
24. Ruan J. et al. Identifying suspicious groups of affiliated-transaction-based tax evasion in big data. Information Sciences. Vol. 477, p. 508–532, 2019. DOI: http://dx.doi.org/10.1016/j.ins.2018.11.008.
25. Jihal H., Ounacer S., Ardchir S., Azouazi M. (2020) Clustering Model of False Positive Elimination in Moroccan Fiscal Fraud Detection. In: Ezziyyani M. (eds) Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). Advances in Intelligent Systems and Computing, vol. 1104. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-36671-1_12.
26. Adamov A.Z. IEEE 13th International Conference on Application of Information and Communication Technologies (AICT). Machine Learning and Advanced Analytics in Tax Fraud Detection, 2019. DOI: http://dx.doi.org/10.1109/AICT47866.2019.8981758.
27. Stankevicius E., Leonas L. Hybrid approach model for prevention of tax evasion and fraud. Procedia-Social and Behavioral Sciences. Vol. 213, p. 383–389, 2015. DOI: http://dx.doi.org/10.1016/j.sbspro.2015.11.555.
28. Paul Jaccard. Etude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Socit Vaudoise des Sciences Naturelles, no. 37, p. 547–579. DOI: http://dx.doi.org/10.5169/seals-266450.

REFERENCES:

[1] Bogahawatte K., Adikari S. 8th International Conference on Computer Science & Education. Intelligent criminal identification system, Colombo, Sri Lanka. 2013, p. 633–638. DOI: http://dx.doi.org/10.1109/ICCSE.2013.6553986.
[2] Hamdy E., Adl A., Hassanien A.E., Hegazy O. and Kim T. -H. Criminal Act Detection and Identification Model, 2015 Seventh International Conference on Advanced Communication and Networking (ACN). 2015, p. 79–83. DOI: http://dx.doi.org/10.1109/ACN.2015.30.
[3] Nath S.V. Crime Pattern Detection Using Data Mining, 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops. 2006, p. 41–44. DOI: http://dx.doi.org/10.1109/WI-IATW.2006.55.
[4] Saravanan P., Selvaprabu J., Arun Raj L., Abdul Azeez Khan A., Javubar Sathick K. (2021) Survey on Crime Analysis and Prediction Using Data Mining and Machine Learning Techniques. In: Zhou N., Hemamalini S. (eds) Advances in Smart Grid Technology. Lecture Notes in Electrical Engineering, vol 688. Springer, Singapore. DOI: http://dx.doi.org/10.1007/978-981-15-7241-8_31.
[5] Jihal H., Ounacer S., Ardchir S., Azouazi M. (2020) Clustering Model of False Positive Elimination in Moroccan Fiscal Fraud Detection. In: Ezziyyani M. (eds) Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). Advances in Intelligent Systems and Computing, vol 1104. Springer, Cham. DOI: http://dx.doi.org/10.1007/978-3-030-36671-1_12.
[6] Wu Y., Dong B., Zheng Q., Wei R., Wang Z. and Li X. A Novel Tax Evasion Detection Framework via Fused Transaction Network Representation, 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). 2020, p. 235–244. DOI: http://dx.doi.org/10.1109/COMPSAC48688.2020.00039.

[7] Didimo W., Grilli L., Liotta G., Menconi L., Montecchiani F. and Pagliuca D. Combining Network Visualization and Data Mining for Tax Risk Assessment, in IEEE Access. Vol. 8, p. 16073–16086, 2020. DOI: http://dx.doi.org/10.1109/ACCESS.2020.2967974.
[8] Taha K., Yoo P.D. Using the Spanning Tree of a Criminal Network for Identifying Its Leaders, in IEEE Transactions on Information Forensics and Security. Vol. 12, no. 2, p. 445–453, 2017. DOI: http://dx.doi.org/10.1109/TIFS.2016.2622226.
[9] Zheng Q. et al. ATTENet: Detecting and Explaining Suspicious Tax Evasion Groups. «Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence». P. 6584–6586, 2019. DOI: http://dx.doi.org/10.24963/ijcai.2019/964.
[10] Zolotarev O.V et al., System PullEnti — information extraction from natural language texts and automatic construction of information systems. School-Seminar (International) on Situational Centers and Information-Analytical System 4i Class for Monitoring and Security Tasks Proceedings, v. 2, IFTI, Protvino, 2016, p. 28–35. URL: https://www.elibrary.ru/item.asp?id=28185224 (accessed: 30 03 2021) (in Russian).
[11] Announcing the Neo4j Crime Investigation Sandbox. URL: https://medium.com/neo4j/announcing-the-neo4j-crime-investigation-sandbox-c0c3bd9e71b1 (accessed: 30 03 2021).
[12] Neo4j and the Offshore Leaks: the Case of Azerbaijan. URL: https://neo4j.com/graphgist/neo4j-and-the-offshore-leaks-the-case-of-azerbaijan (accessed: 30 03 2021).
[13] Adamic L.A. и Adar E., Friends and neighbors on the web Social networks. Vol. 25, no. 3, p. 211–230, 2003. DOI: http://dx.doi.org/10.1016/S0378-8733(03)00009-1.
[14] Zhou T., Lü L., Zhang Y.C. Predicting missing links via local information. The European Physical Journal. Vol. 71, no. 4, p. 623–630, 2009. DOI: http://dx.doi.org/10.1140/epjb/e2009-00335-8.
[15] Barabâsi A.L. et al. Evolution of the social network of scientific collaborations. Physica A: Statistical mechanics and its applications. Vol. 311, no. 3-4, p. 590–614, 2002. DOI: http://dx.doi.org/10.1016/S0378-4371(02)00736-7.
[16] Newman M.E.J. Clustering and preferential attachment in growing networks. Physical review. Vol. 64, no. 2, 2001. DOI: http://dx.doi.org/10.1103/PhysRevE.64.025102.
[17] Martínez V., Berzal F., Cubero J C. A survey of link prediction in complex networks. ACM computing surveys (CSUR). Vol. 49, no. 4, p. 1–33, 2016. DOI: http://dx.doi.org/10.1145/3012704.
[18] Rungsawang A., Manaskasemsak B. An Efficient Partition-Based Parallel PageRank Algorithm. Proceedings of the 11th International Conference Parallel and Distributed Computing, 2004. DOI: http://dx.doi.org/10.1109/ICPADS.2005.85.
[19] Brandes U. A faster algorithm for betweenness centrality. Journal of mathematical sociology. Vol. 25, no. 2, p. 163–177, 2001. DOI: http://dx.doi.org/10.1080/0022250X.2001.9990249.
[20] Sariyüce A.E., Kaya K., Saule E and Çatalyiirek Ü.V. Incremental algorithms for closeness centrality, 2013 IEEE International Conference on Big Data. 2013, p. 487–492. DOI: http://dx.doi.org/10.1109/BigData.2013.6691611.
[21] Bihari A. and Pandia M.K. Eigenvector centrality and its application in research professionals' relationship network, 2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE). 2015, p. 510–514. DOI: http://dx.doi.org/10.1109/ABLAZE.2015.7154915.
[22] Li X. et al. Identifying social influence in complex networks: A novel conductance eigenvector centrality model. Neurocomputing. Vol. 210, p. 141–154, 2016. DOI: http://dx.doi.org/10.1016/j.neucom.2015.11.123.
[23] Didimo W. et al. Visual querying and analysis of temporal fiscal networks. Information Sciences. Vol. 505, p. 406–421, 2019. DOI: http://dx.doi.org/10.1016/j.ins.2019.07.097.
[24] Ruan J. et al. Identifying suspicious groups of affiliated-transaction-based tax evasion in big data. Information Sciences. Vol. 477, p. 508–532, 2019. DOI: http://dx.doi.org/10.1016/j.ins.2018.11.008.
[25] Jihal H., Ounacer S., Ardchir S., Azouazi M. (2020) Clustering Model of False Positive Elimination in Moroccan Fiscal Fraud Detection. In: Ezziyyani M. (eds) Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). Advances in Intelligent Systems and Computing, vol. 1104. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-36671-1_12.
[26] Adamov A.Z. IEEE 13th International Conference on Application of Information and Communication Technologies (AICT). Machine Learning and Advanced Analytics in Tax Fraud Detection, 2019. DOI: http://dx.doi.org/10.1109/AICT47866.2019.8981758.
[27] Stankevicius E., Leonas L. Hybrid approach model for prevention of tax evasion and fraud. Procedia-Social and Behavioral Sciences. Vol. 213, p. 383–389, 2015. DOI: http://dx.doi.org/10.1016/j.sbspro.2015.11.555.
[28] Paul Jaccard. Etude comparative de la distribution florale dans une portion des Alpes et des Jura. Bulletin del la Socit Vaudoise des Sciences Naturelles, no. 37, p. 547–579. DOI: http://dx.doi.org/10.5169/seals-266450.

Поступила в редакцию – 31 октября 2021 г. Окончательный вариант – 01 марта 2022 г.
Received – October 31, 2021. The final version – March 01, 2022.