

СИСТЕМА МЕТОДОВ КАРТОГРАФИРОВАНИЯ ЗАЩИЩАЕМОГО КИБЕРПРОСТРАНСТВА

А.Л. Сердечный

Настоящая работа структурирует методологию построения и анализа информационных карт. Предложено подробное описание системы методов, необходимых для построения и анализа информационных карт. В качестве основных групп методов рассматриваются: методы подготовки данных для информационной карты (моделирование данных на основе источников, веб-скрапинг и др.), методы визуализации информационной карты (методы укладки графов, построение тепловой карты и др.), методы актуализации информационной карты (метод совмещения графов, метод трансформации ландшафта), традиционные картографические методы (графический оверлей, разметка областей карты и др.), сетевые методы (методы расчёта метрик графа, методы кластеризации графа и др.). Предлагаемая система методов является достаточной для проведения всесторонних картографических исследований защищаемого киберпространства, так и других сетевых структур, моделирующих предметную область. Также допускается расширение данной системы дополнительными методами, направленными на повышение эффективности процессов построения и анализа информационных карт.

Ключевые слова: метод, информационная карта, киберпространство, граф.

Введение

Статья посвящена развитию картографического подхода исследования защищаемого киберпространства, изложенного в работах [1-3]. В ней изложены результаты систематизации методов построения и анализа информационных карт, использование которых позволяет повысить

эффективность экспертного анализа процессов и явлений, происходящих в киберпространстве за счёт использования возможностей человека по обработке визуальной информации.

Система методов, подробное описание которых представлено в настоящей работе, приведена на рис. 1.



Рис. 1. Система методов построения и анализа информационных карт

Методы подготовки данных для информационной карты

Подготовка исходных данных для построения информационных карт требует выполнение следующих этапов:

- моделирование данных информационной карты;
- сбор необходимых исходных данных;
- выбор из всего множества собранных данных тех, которые требуется использовать

при построении информационной карты (генерализация данных).

В зависимости от задач исследований, а также специфики исходных данных, на каждом этапе могут применяться собственные методы, рассматриваемые в настоящей работе.

Методы моделирования данных информационной карты

Методы моделирования данных имеют ключевое значение, так как с помощью них формируется основа для построения информационной карты. Методы моделирования данных направлены на определения облика необходимых исходных данных. С помощью данных методов не только определяется состав объектов киберпространства, формирующих ландшафт карты, но и осуществляется обоснование метрик близости между такими объектами. От того, насколько точно и полно будут описаны исходные данные, зависит качество и скорость построения информационной карты. Моделирование данных может осуществляться как на в результате анализа неструктурированных или частично структурированных информационных источников, так и на основании онтологий соответствующих предметных областей, если такие разработаны для решаемой с помощью информационных карт задачи.

Моделирование данных на основе источников

Метод моделирования данных на основе источников заключается в разработке модели представления исходных данных, необходимых для создания информационной карты. В ходе реализации метода осуществляется поиск наиболее крупных источников, полнота сведений в которых охватывает большую область фактов моделируемой предметной области (например, для анализа долговременных целенаправленных угроз воздействия на информационные системы в качестве таких источников могут быть выбраны основные базы данных, агрегирующие сведения из отчётов организаций, занимающихся исследованием вредоносного программного обеспечения).

В ходе реализации метода экспертом формируется представления о составе объектов (сведения о которых содержатся в рассматриваемых источниках) и их взаимосвязях. Как только экспертом построена адекватная модель данных, позволяющая описать сведения из источников, такая модель адаптируется под задачу построения информационной карты. В результате этого из модели выбираются только такие типы объектов и связей между ними, которые требуются для решения задачи построения и анализа соответствующей карты.

Некоторые объекты и связи могут быть преобразованы, если это позволяет упростить задачу исследования. Так, например, модель данных о Bitcoin-транзакциях для информационного ресурса Blockchain предполагает организацию передачи сведений о совершаемых операциях с использованием двух типов сущностей:

- «Транзакция», которая содержит сведения об отправителях и получателях криптовалюты в рамках совершённой операции;

- «Криптокошелёк», который связан с сущностью «Транзакция» как отправитель или как получатель определённой суммы (если «Криптокошелёк» выполняет роль отправителя, то по итогам транзакции с него списывается некоторая сумма, если отведена роль получателя – на него поступает криптовалюта).

Для некоторых задач (например, визуального анализа операций с криптовалютами) удобнее использовать более простую модель данных, где используется всего лишь одна сущность – «Криптокошелёк». В этом случае размер полученных или отправленных средств может быть записан в параметрах связей.

В ходе моделирования данных также должны быть рассмотрены возможности использования различных метрик близости. Должна существовать как минимум одна метрика. Если таких метрик несколько, то на основании разработанной модели данных можно будет построить различные проекции информационной карты. Метрики близости могут базироваться как на численных методиках расчёта (например, косинусное

сходство документов, представленных в векторном пространстве), так и на неформализованных алгоритмах (например, сходство объектов, определяемое рекомендательными системами). Во втором случае для обеспечения повторяемости процедуры построения информационной карты значения метрик должны быть внесены в состав набора данных.

В процессе анализа информационной карты могут быть выявлены противоречия в построенной модели данных или же найдены новые источники. Это может потребовать уточнение модели и изменения информационной карты, что приведёт к повышению качества исследований. На основе разработанной таким образом модели данных могут быть сформированы онтологии и базы знаний.

Моделирование данных на основе онтологий

Метод моделирования данных на основе онтологий аналогичен предыдущему методу. Его отличие заключается в том, что исследователь при разработке модели представления исходных данных, необходимых для построения информационной карты, работает уже с имеющейся онтологической моделью. С одной стороны, данный метод сокращает работу эксперта за счёт использования уже готовых сведений, которые не требуется получать в результате поиска и анализа дополнительных источников. Также использование готовых онтологий повышает обоснованность формируемой модели и даёт возможность её стандартизации.

Однако разработанные онтологии существуют не для всех задач. Для задач, которые не требуют сложных моделей данных (например, задачи связи источника информации и сведений об уязвимостях программного обеспечения), более простым вариантом оказывается использование метода моделирования данных на основе источник, нежели поиск и изучение сложных многокомпонентных онтологий, таких как STIX или CVRF (данные онтологии содержат в том числе связи между уязвимостями и источниками сведений о них).

В ходе анализа информационной карты онтологические модели также могут быть уточнены.

Методы сбора исходных данных информационной карты

Класс методов автоматизированного сбора исходных данных достаточно широк и связан с поиском, получением доступа и предварительной обработкой данных. На реализацию данных методов уходит большая часть времени построения информационных карт. В связи с этим важное значение приобретают информационные ресурсы, агрегирующие различные наборы данных (такие как [4, 5]). Такие ресурсы не только предоставляют удобный доступ к собранным и структурированным сведениям, но также содержат необходимую информацию о формате и истории происхождения соответствующих наборов данных.

Для некоторых типов данных существуют специализированные методы извлечения фактов. Так, например, для автоматического извлечения ключевых слов и терминов из текстовой информации существует отдельная задача в рамках компьютерной лингвистики. Анализ сцен в потоке видео также требует особых методов, связанных с анализом изображений и сопоставлением их содержимого с текстовой информацией.

Веб-скрапинг

Одним из основных методов получения исходных данных для информационных карт защищаемого киберпространства является сбор сведений из информационных ресурсов сети Интернет. Данный метод получил название «веб-скрапинг» (от слова *scrape* – очистка), так как для получения сведений с Интернет-сайтов требуется предварительная обработка веб-страницы и приведение полученных данных к нужному формату (в соответствии с разработанной моделью данных).

Сбор сведений осуществляется с помощью специальной программы, называемой парсером, для которой сформированы правила сбора и преобразования данных. Способы задания правил могут существенно отличаться в зависимости от используемого средства

автоматизации. Их состав варьируется от использования только лишь графического интерфейса до написания кода на каком-либо языке программирования с использованием специализированных библиотек. Основными техниками анализа содержимого веб-страницы является использование CSS-селекторов, разбор DOM-деревьев и применение регулярных выражений [6].

Обычно в ходе исследований требуется выполнение сбора данных с нескольких информационных ресурсов, которые используют собственные форматы и отличающиеся обозначения информационных объектов. В этом случае потребуются выполнение дополнительной работы по интеграции данных.

В ходе веб-скраппинга исследователю приходится преодолевать механизмы защиты от сбора данных. Некоторые механизмы могут быть преодолены лишь при помощи распределённой системы сбора данных, работающей в несколько тысяч потоков с различных IP-адресов для достижения требуемых скоростей получения исходных данных.

Задачи мониторинга обстановки с помощью информационной карты может потребовать выполнение периодического сбора сведений, что также усложняет архитектуру разрабатываемого парсера.

Зондирование информационного ресурса

Метод зондирования применяется для выборочного сбора сведений из информационных ресурсов с громадным динамически изменяемым объёмом связанных данных. Ярким примером таких источников являются социальные платформы, благодаря которым пользователи ежедневно генерируют гигабайты (для некоторых платформ терабайты) новых данных.

Впервые метод зондирования информационного пространства был реализован в поисковых системах, которые с помощью поисковых роботов осуществляют обход веб-страниц по гиперссылкам. В зарубежных источниках для метода зондирования веб-ресурсов используется термин Web Crawler. Аналогичный способ может быть применён к рекомендательным системам, когда осуществляется последовательный переход по ссылкам на

похожие объекты, которые, в свою очередь, связаны с другими похожими объектами, на которые также можно перейти.

Существуют различные стратегии зондирования информационных источников [7, 8]:

- масштабное зондирование (последовательный обход всех встречающихся ссылок);

- управляемое зондирование (исследователь задаёт направление обхода на основании информации, получаемой в ходе зондирования);

- зондирование на основании правил (ссылки для обхода генерируются на основании специальных правил, например, путём последовательного перебора идентификаторов);

- распределённое зондирование (имеются множество поисковых роботов, собирающих данные, для которых правила обхода задаются с помощью команд от сервера, который автоматически обрабатывает получаемые данные от поисковых роботов);

- непрерывное зондирование (ссылки могут повторно просматриваться с целью обнаружения изменений).

Эффективность стратегий и конкретных техник зависит от параметров источника, а также структурных особенностей связанных данных. Для её повышения в рамках реализуемых алгоритмов используются методы оптимизации, основанные на природных аналогиях (метод фейерверков, метод муравьиных колоний и др. [9, 10]). Оценка эффективности может быть представлена как решение задачи о канадском путешественнике (обобщение задачи о кратчайших путях на графе, которые частично видимы [11]).

Извлечение терминов, ключевых слов и именованных сущностей

Во многих задачах, связанных с защитой информации и обеспечением информационной безопасности, требуется определение сходства между информационными объектами на основании содержания его текстовой информации. Одним из распространённых подходов к решению этих задач является выявление терминов, ключевых фраз и других

информационных объектов. Так, например, именованные сущности могут извлекаться на основании стиля написания и использования словаря именованных сущностей.

Особое значение для информационной картографии является извлечение ключевых фраз. Термины и ключевые фразы могут использоваться как самостоятельные информационные объекты или же в качестве подписей объектов и слоёв информационной карты. Ввиду многозначности терминов, возможности их составления из нескольких слов, а также проблемы омонимии извлечение терминов является нетривиальной задачей.

Решение задачи извлечения терминов и ключевых фраз осуществляется в рамках компьютерной лингвистики. За последние 20 лет в данной области заметен прогресс, однако, точность результатов еще недостаточно высока даже для наиболее популярного среди исследователей английского языка [12]. Существуют различные алгоритмы извлечения терминов и ключевых слов [12, 13]:

1) построенные на базе размеченных наборов данных;

2) построенные без использования размеченных данных:

- основанные на статистике (TF-IDF, KP-Miner, YAKE и др.);

- основанные на графах (TextRank, SingleRank и др.);

- основанные на тематике (KeyClustering, TopicRank и др.);

- основанные на языковых моделях.

Наиболее распространёнными являются методы, основанные на графах. Однако статистические методы также привлекают внимание исследовательского сообщества.

Используя автоматические средства извлечения текстовой информации, необходимо иметь представление о полноте и точности извлекаемых данных, которые могут существенно различаться в зависимости от языка и тематической направленности текста.

Анализ и обобщение геопространственных данных

Важным источником данных для построения информационных карт являются геопространственные базы данных.

Благодаря совместным усилиям тысяч пользователей в них накоплены громадные объёмы данных, которые включают не только сведения о географических объектах, но и о связанных с ними информационных объектах, которые имеют отношение к защищаемому киберпространству. Так, например, такие сервисы, как Wikimapia [14] или «Народная карта» Яндекса [15] могут содержать отметки об атаках на элементы критической информационной инфраструктуры. Также благодаря геопространственным данным могут быть получены справочные сведения об организациях и объектах киберпространства (например, о расположении инфраструктуры операторов связи).

Анализ и обобщение геопространственных данных предполагает использование алгоритмов аналитической геометрии и математической статистики для определения характеристик объектов через их топологические, геометрические или географические свойства. К таким характеристикам относятся: положение объектов, расстояние между объектами, длина маршрута перемещения между объектами по дорожной сети и др.

В качестве основной операции работы с такими данными следует считать локализацию области размещения объекта на карте, после чего выполняются пространственные запросы, связанные с этой областью. Скорость выполнения этих запросов обеспечивается наличием пространственных индексов для различных слоёв, с которыми связаны данные. Для этой задачи разработаны специализированные базы данных, такие как PostGIS. Однако в некоторых случаях (например, для сбора данных о неизвестном маршруте в сложной сети дорог) могут дополнительно использоваться алгоритмы теории графов [16].

В качестве инструментов работы с пространственными данными используют геоинформационные системы или специализированные библиотеки для таких языков, как Python, C++, R [17].

Анализ сцены (анализ мультимедийных данных)

С развитием автономных киберфизических систем и активной

цифровизацией реальности популярность получило направление автоматического анализа сцены. В рамках данного направления осуществляется анализ разнородных сведений (представленные в виде фотоснимков, звуковых файлов, видеозаписей и текстовых сообщений) как единой системы, образующей сцену, в которой реализуется какой-либо процесс. Данное направление находится на начальном этапе развития и относится к области машинного зрения. В качестве аналитического аппарата обычно используются искусственные нейронные сети, а основными направлениями являются [18]:

- определение основных характеристик сцены;
- кодирование признаков сцены;
- распознавание шаблонов пространственной компоновки сцены;
- распознавание объектов сцены;
- анализ связей объектов сцены;
- анализ изменений сцены.

Методы анализа сцен имеют хорошие перспективы для извлечения исходных данных из потокового видео или набора изображений с целью обнаружения и картографирования деструктивного контента.

Очистка данных

Для согласования исходных данных, получаемых из разных источников требуется очистка данных. Очистка данных осуществляется автоматически (в соответствии с набором правил) или вручную. В рамках информационно-картографических исследований очистка данных может быть осуществлена несколько раз:

- выполнение очистки данных для построения первоначальной карты;
- выполнение очистки данных в результате обнаружения противоречий в ходе анализа информационной карты.

Первоначальной очистке данных должно быть уделено основное внимание. Если данные из различных источников не привести к единому формату или допустить существенные неточности в ходе их преобразования, то в результате анализа информационной карты могут быть сделаны ошибочные выводы в отношении состава источников или модели данных.

В ходе очистки данных выполняются следующие процедуры:

- форматирование исходных данных;
- разъединение строк по разделяющим символам или на основе регулярных выражений;
- согласование идентификаторов информационных объектов.

Согласование идентификаторов является самой трудоёмкой процедурой. Если в качестве идентификаторов информационных объектов используются их названия, то операции по очистке данных могут включать:

- удаление «зашумляющих» символов (пробелов, знаков препинания);
- приведение названий к единому регистру;
- внесение в название идентификатора уникального признака объекта (для различных объектов с одинаковыми названиями).

Удаление «зашумляющих» символов и приведение к единому регистру позволяют автоматически и быстро провести очистку данных. При этом в некоторых случаях такие операции могут стать причиной ложного объединения разных объектов. Так, например, приведение аббревиатуры к нижнему регистру может сделать её неотличимой от осмысленного названия совершенно другого объекта.

В случае существования для картографируемых информационных объектов онтологии с общепринятыми идентификаторами, следует выполнить процедуру отображения этих идентификаторов на соответствующие информационные объекты. В будущем это обеспечит согласованность информационной карты с новыми источниками и упростит процедуру картографирования.

Методы генерализации данных

Методы генерализации данных применяются для многослойных информационных карт и заключаются в отображении или скрытии её деталей в зависимости от необходимости фокусирования внимания аналитика на определённых объектах. Методы генерализации данных разделены по степени

автоматизации процесса на следующие методы:

- экспертная генерализация данных;
- автоматическая генерализация данных;
- автоматизированная генерализация

данных.

Экспертная генерализация данных осуществляется исследователем путём рассмотрения различных комбинаций отображения информационных карт на её слоях. Эксперт решает следующие вопросы:

- какой состав слоёв будет включать информационная карта;
- какие объекты будут образовывать ландшафты слоёв информационной карты;
- какие объекты и связи требуется отображать на каждом из слоёв.

Для ответа на эти вопросы эксперт руководствуется собственным опытом, а также представлением о важности и наглядности отображений информационных объектов.

Альтернативой экспертной генерализации данных выступает метод *автоматическая генерализация*. Для географической картографии алгоритмы генерализации данных достаточно развиты, что подтверждается существующими геоинформационными системами. Такие алгоритмы ориентированы на задачи, решаемые с помощью карт [19]. При этом необходимо отметить, что в традиционной картографии успех алгоритмов автоматической генерализации объясняется тем, что перечень задач и состав отображаемых объектов для традиционной картографии полностью известен и имеется богатый опыт практического использования географических карт. Для информационных карт эти условия не выполняются. Однако существуют предпосылки созданию подобных алгоритмов.

В качестве основы автоматической генерализации информационных карт являются метрики графа объектов, построенного на основании исходных данных. Например, в качестве возможных критериев выбора отображаемых объектов в зависимости от масштаба слоя могут быть метрики центральности соответствующих узлов.

Метод автоматической генерализации данных позволяет унифицировать процедуру

построения карты, а также сократить затраты на привлечение ручного труда высококвалифицированных экспертов. При этом необходимо отметить, что данный метод применим лишь для отработанных моделей данных и информационных объектов, для которых хорошо изучена специфика отображения и разработаны соответствующие алгоритмы представления данных.

Применения подобных алгоритмов к объектам с собственной спецификой в большинстве случаев приведёт к некорректным результатам. Это обусловлено недостатками автоматических алгоритмов анализа графов, о которых было сказано ранее.

Объединения преимуществ экспертного и автоматического методов возможно добиться в их последовательном применении. Для построения первоначальной информационной карты могут быть использованы различные алгоритмы автоматической генерализации данных, лучший из которых будет выбран в ходе предварительного анализа. Наилучший алгоритм адаптируется в соответствии с задачами исследования или принимается решение об использовании экспертного метода генерализации данных.

Таким образом, *автоматизированная генерализация* не ограничивается лишь этапом сбора исходных данных. Основная работа приходится на процедуру анализа первой версии информационной карты.

Методы визуализации информационной карты

Методы визуализации направлены на формирование ландшафта карты, а также отображение на нём информационных объектов. В основе методов визуализации лежат алгоритмы укладки графа.

Методы укладки графов

В зависимости от условий анализа, а также имеющихся технических средств возможны следующие варианты укладки графа:

- на плоскости;
- на поверхности сферы;
- в трёхмерном пространстве.

Укладка графа на плоскости

Применение информационной карты в печатных изданиях предполагает возможность использования лишь двумерного пространства. Кроме того, человеку на поверхности экрана удобнее воспринимать двумерное изображение. В связи с этим укладка графа на плоскости является наиболее распространённым вариантом. Однако данный метод обладает наибольшими искажениями из-за наибольшего влияния эффекта от снижения размерности.

При отображении графа в двумерном пространстве особое значение имеет фокус внимания, определяемый тем, какая часть графа будет расположена в центре. Также важное влияние имеют свойства графа, среди которых:

- наличие или отсутствие направлений рёбер (направленный или ненаправленный граф);
- наличие или отсутствие циклов (дерево или мультиграф);
- вес рёбер (взвешенный или невзвешенный граф);
- наличие вложенных элементов (гиперграф).

Для укладки графа на плоскости используются разнообразные алгоритмы. Наиболее распространёнными являются алгоритмы силовой укладки, основанные на

аналогиях с физическими силами, действующими на материальные объекты. Чаще всего используется алгоритм ForceAtlas2 [20], основанный на пружинной модели. Данный алгоритм реализован во многих программных продуктах визуализации графов, так как он достаточно прост в реализации, демонстрирует качественные результаты и высокую скорость. Это достигается за счёт выбора удачных оптимизаций, а также возможности распределённого вычисления. Укладка граф с помощью алгоритма ForceAtlas2 может быть распараллелена (в том числе с использованием высокопроизводительных видеокарт). За приемлемое время данный алгоритм способен на персональном компьютере визуализировать в двумерном пространстве граф размером до миллиона узлов.

Алгоритм ForceAtlas2 также может быть применён к взвешенным графам, причём имеется возможность регулирования влияния веса на моделируемые силы притяжения или отталкивания. Данная возможность позволяет визуализировать карты с учётом неоднородностей связей между информационными объектами.

В качестве недостатка алгоритма можно отметить локальную сходимость слабосвязанного графа из-за влияния оптимизирующих преобразований (рис. 2).

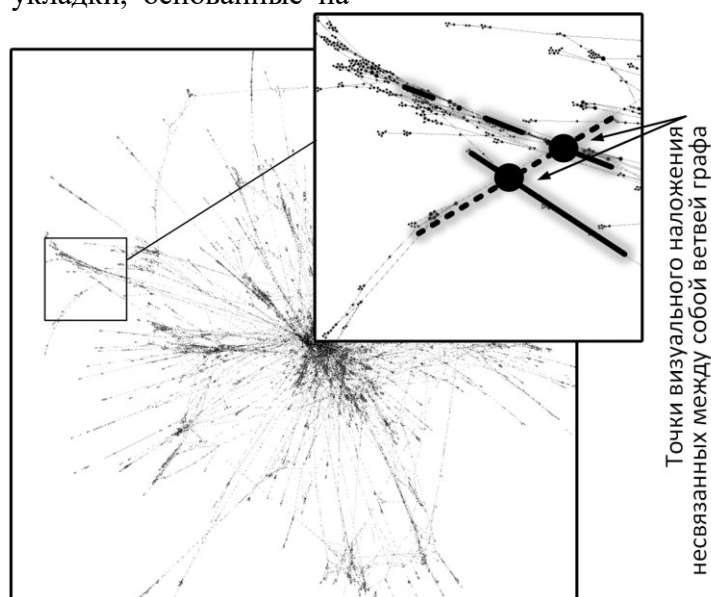


Рис. 2. Демонстрация состояния локального максимума после применения алгоритма силовой укладки ForceAtlas2 [20], при котором несвязанные между собой ветви графа имеют визуальные пересечения в двумерном пространстве

В качестве альтернативы алгоритму ForceAtlas2 в [21] предложен алгоритм BatchLayoutBH, который по заявлениям авторов обеспечивает более эффективный механизм распределённых вычислений при сохранении визуального качества укладки графа. Для графа с 1,5 млн. узлов и 114 млн. дуг расчёты были завершены за 224 секунды. Для графа с 74 тыс. узлами вычисления длились 10 секунд, что было в 14 раз быстрее аналогичных расчётов для алгоритма ForceAtlas2.

В качестве альтернативы силовой укладки могут быть использованы традиционные алгоритмы снижения размерности на основе матрицы признаков (в виде которой можно представить связи графа). Наиболее эффективными алгоритмами являются t-SNE и UMAP[22].

Несмотря на визуальное сходство результатов выполнения силовых алгоритмов укладки и алгоритмов, основанных на традиционных методах снижения размерности, в последних может нарушаться физическая аналогия. Близость объектов основана не на вычислении расстояний в

соответствии с законами действия сил, а на результатах минимизации значений функции попарного сравнения расстояний в пространствах исходной и сниженной размерности. Выбор этой функции может не предполагать минимизацию глобального значения функции стресса. Т.е. расположение элементов на информационной карте внутри кластеров будет удовлетворять требованию близости объектов, а позиции самих кластеров друг по отношению к другу – нет. Результаты анализа такой «карты» в отношении закономерностей расположения крупных структур могут быть ложно интерпретированы. Для алгоритмов UMAP и t-SNE это утверждение относится в меньшей степени, так как данные алгоритмы вместе с алгоритмом ForceAtlas2 относятся к классу алгоритмов, основанных на силе. В работе [23] было показано, что алгоритмы t-SNE [24], UMAP [22], ForceAtlas2 [20] и собственные карты Лапласа можно свести друг к другу путём изменения баланса сил притяжения/отталкивания в алгоритме t-SNE (рис. 3).

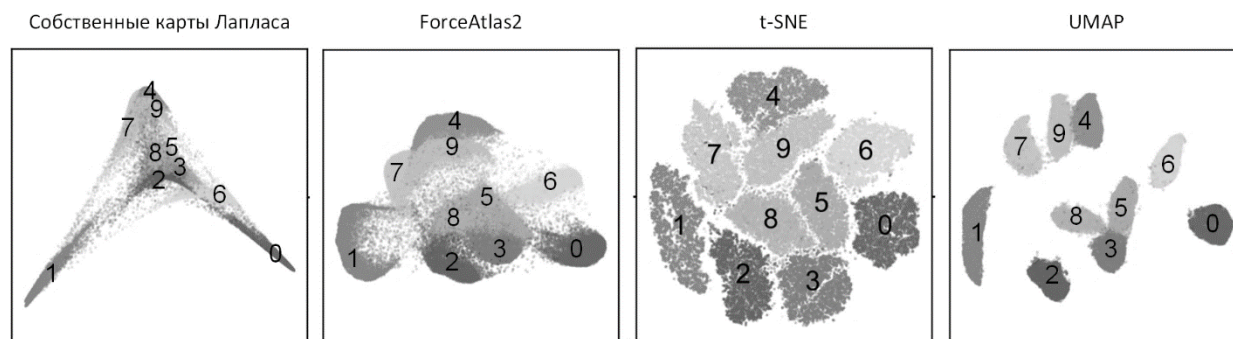


Рис. 3. Результаты сравнения различных алгоритмов укладки графа [23] (в качестве набора данных для классификации использована база данных 70 000 изображений рукописных цифр, кластеры которых обозначены соответствующей цифрой)

Ещё одной особенностью алгоритма ForceAtlas2 заключается в том, что данный алгоритм не учитывает направление рёбер графа. Если для построения информационной карты имеет значение направление и порядок следования вершин, то выбор алгоритма ForceAtlas2 не подходит. Одним из возможных решений здесь является использование идей Сугияма и Мисуэ, которые предложили расширение пружинной модели. В [25] предлагается «намагнитить» некоторые или все пружины и ввести

глобальное магнитное поле, которое действует на них. Магнитное поле может использоваться для управления ориентацией рёбер графа. Это позволяет учитывать направление связей, что привносит дополнительные возможности в анализ направленных графов.

Существует три основных типа алгоритма укладки магнитным полем [25]:

- параллельный (все магнитные силы действуют в одном направлении);

- радиальный (магнитные силы действуют радиально наружу от узла);
 - концентрический (магнитные силы действуют в области концентрических окружностей).

Говоря о методе укладки графа на плоскости, необходимо отметить его особенности применительно к деревьям. В виде деревьев часто представляются вложенные иерархические структуры, к которым также относятся слои масштабируемой информационной карты. Однако, как было показано в [26], использование алгоритмов, основанных на расчёте расстояний в евклидовых пространствах, не является оптимальным выбором для визуализации многоуровневых деревьев с разветвлёнными узлами. Авторы [26] предлагают использовать карты Пуанкаре для которых используются геометрия гиперболического пространства. Деревья, представленные в гиперболическом пространстве, могут быть отображены в двумерное пространство с сохранением парных расстояний между вершинами иерархии.

Другим способом представления деревьев является способ вложенных иерархий, когда узел, расположенный на более высоком уровне иерархии включает в себя все узлы потомки. Таким образом, вся поверхность холста распределяется между узлами верхнего уровня иерархии, которые ограничивают каждый свою часть пространства. Внутри этого пространства располагаются узлы потомки, которые распределяют эту часть между собой. Более подробно об укладке вложенных групп узлов графа сказано в работе [27].

Укладка графа на поверхности сферы

Ключевым свойством сферической поверхности для задач информационного картографирования является отсутствие границ (замкнутость поверхности). В случае, если информационная карта является генеральной, а перечень информационных объектов, образующих её пространство, является достаточно полным, причём их граф связей не имеет единого центра, то для изображения такой карты лучше всего использовать поверхность сферы. В качестве иллюстрации можно привести граф связей научных дисциплин (рис. 4) [28, 29, 30, 31, 32].

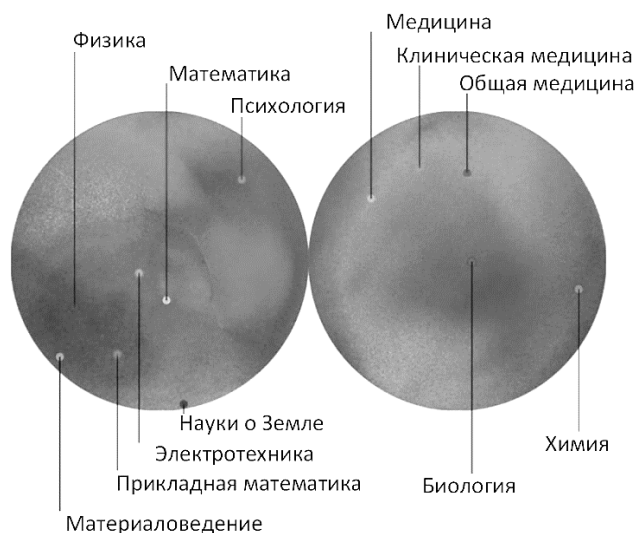


Рис. 4. Две полусферы, на котором изображён граф цитирования публикаций системы Web of Science, уложенный в сферическом пространстве и преобразованный к тепловой карте (точкам отмечены центры наиболее крупных кластеров, соответствующих 12-ти областям научного знания) [28]

Существует два основных способа визуализации графа на сфере [33]:

- проекционный способ;
 - сферическое многомерное шкалирование.

Первый способ заключается в проецировании результатов двумерной визуализации графа на сферическую поверхность. Такой способ не позволяет воспользоваться преимуществом отсутствия границ, а лишь имитирует сферическую

укладу, по сути, оставаясь укладкой графа на плоскости. Единственным преимуществом такого способа является фокусирование внимания на определённом участке графа за счёт эффекта «рыбьего глаза». В ближайшей к наблюдателю окрестности сферы соответствующий участок графа будет увеличен, в то время, как расстояния между элементами, оказавшимися на периферии будет уменьшено. Это позволяет акцентировать внимание на определённой области и лучше рассмотреть связи для узлов, попавших в эту область.

В работе [33] предложен способ сферического многомерного шкалирования, заключающийся в следующем:

- формирование исходной матрицы расстояний, в которой в качестве элементов выступают кратчайшие расстояния для каждой пары узлов графа, вычисленные с помощью алгоритма Дейкстры;

- расчёт матрицы расстояний между точками на сфере (каждая точка соответствует узлу исходного графа) таким образом, чтобы минимизировать функцию среднеквадратического отклонения расстояний для соответствующих пар точек на сфере и узлов исходного графа;

- визуализация точек в соответствии с рассчитанными значениями.

Для данного алгоритма остаётся открытым вопрос выбора радиуса сферы, названный автором [33] вопросом семантического масштабирования сферы.

В [1] предложен силовой алгоритм укладки графа на сфере, который заключается в следующем:

- положения узлов представляются в виде трехмерных векторов с длиной, равной единице (т.е. находятся на поверхности сферы радиусом 1, центрированной в начале координат);

- для вычисления начальных координат узлов используется силовой алгоритм компоновки Фрухтермана-Рейнгольда [34] (может быть использован любой алгоритм);

- на каждом шаге позиции узлов обновляются путем приложения сил притяжения и отталкивания к парам узлов.

Для больших графов попарное вычисление расстояния является слишком трудоёмкой процедурой, поэтому в [28] предлагается

оптимизация, основанная на аппроксимации, используемой для задачи моделирования n-тел (метод Барнса-Хата) [35].

На завершающем этапе выбирается фокусный узел, который помещается в центр видимой части сферы. Для узлов, расположенных близко от него вычисляется более точные расстояния, чтобы уменьшить ошибку из-за применения численных методов расчёта.

Укладка графа в трёхмерном пространстве

Метод укладки графа в трёхмерном пространстве направлен на преодоление негативных эффектов снижения размерности для двухмерного пространства. Этот эффект вызывает сложности с отображением некоторых классов реальных сетей.

Большие реальные сети с высоким коэффициентом кластеризации и небольшим средним значением кратчайшего пути трудно визуализировать в двухмерном пространстве, сохраняя принцип близости расстояний между схожими узлами [33]. В сильно связанных участках таких сетей размещение узлов будет носить случайный характер. Если же такие участки распространяются на достаточно крупную область, то в двухмерном пространстве возможно появление нескольких примерно равных значений глобальных максимумов. Для сетей с указанным выше свойством возможно передать лишь общую структуру крупных сообществ. При этом сети с малым значением средней длины кратчайших путей могут оказаться представленными в виде единого кластера, подобного «пучку волос». Ввиду данных свойств такие сети сложно использовать для визуального представления динамики процессов. Отсутствие явной структуры не позволяет «увидеть» причинно-следственные связи между узлами, участвующими в процессе. Поэтому для таких сетей в ряде источников [28, 8, 36] предлагается использовать метод визуализации, который в центр фокуса внимания помещает лишь один ключевой узел, сохраняя сетевые расстояния до всех соседних узлов. Это делается в ущерб сохранению сетевых расстояний между всеми парами узлов.

Также частично данную проблему решает переход к трёхмерному пространству, в котором имеется больше возможностей для расположения узлов графов. Силовая укладка в трёхмерном пространстве достаточно хорошо выполняется с помощью рассмотренного ранее алгоритма ForceAtlas2, для которого евклидово расстояние между узлами вычисляется не для двухмерного, а для трёхмерного пространства.

Важным ограничением трёхмерного пространства является более высокая сложность восприятия исследователем расположенных в нём объектов. Несмотря на возможность передачи перспективы, двухмерный экран ограничивает область пространства. Преодоление этого ограничения возможно благодаря дополнительным техническим средствам визуализации, таким как шлем виртуальной реальности. Для обеспечения работы в трёхмерной виртуальной среде требуется куда больше вычислительных ресурсов, чем для двухмерного пространства. При аналогичных технических возможностях в двухмерном пространстве может быть осуществлена визуализация графа гораздо большей размерности.

Методы формирования ландшафта

Методы формирования ландшафтов предназначены для визуализации контекста, который является иерархически объёмлющей по отношению к объекту исследований системой. Ландшафт представляет фон информационной карты. Попытка классификации способов формирования ландшафтов информационных карт была сделана в [37].

Одним из методов изображения ландшафта является метод, основанный на использовании исходного графа, уложенного в соответствующем пространстве. В этом случае ландшафт представляет собой изображение узлов и рёбер графа (к которым также могут быть применены специализированные алгоритмы укладки). Такой ландшафт назван *сетевым*.

Для графов, размер которых превышает критическое значение (обычно более 30 тыс. узлов), отображение всех его рёбер и узлов не

является оптимальным выбором, так как избыточная информация отвлекает исследователя и требует нерациональное потребление вычислительных ресурсов. Для изображения ландшафта информационных карт, построенных на основе таких графов, используются следующие методы [37]:

- метод формирования ландшафта на основе тепловой карты;

- метод формирования ландшафта на основе преобразования Вороного.

Метод формирования ландшафта на основе тепловой карты

Метод формирования ландшафта на основе тепловой карты заключается в интерполяции точек, соответствующих узлам исходного графа. Чем ближе точки расположены друг к другу, тем выше их плотность. Ландшафт, построенный с помощью данного метода, для краткости можно называть «тепловая карта».

Если ландшафт строится в двухмерном пространстве, то значение плотности задаёт цвет (тон) области вокруг точки (рис. 5.в). Для трёхмерного пространства скопление точек может быть представлено в виде рельефа при помощи эллиптических параболоидов.

Для построения информационного ландшафта вместо тепловой карты (heatmap) может использоваться карта горячих точек (hot spot map), которая по внешнему виду схожа с тепловой картой, но в ходе расчёта плотности точек учитывает их вес. Это может понадобиться для информационных объектов, обладающих разным влиянием. Например, для графа соавторов научных публикаций в качестве веса может использоваться метрика, отражающая его учёную степень и научное звание. В этом случае карта горячих точек будет акцентировать внимание на областях, в которых расположены более авторитетные члены научного сообщества.

Формирование ландшафта на основе тепловой карты включает выполнение следующих операций:

- выбор точечных данных, соответствующих информационным объектам;

- выполнение растровой интерполяции плотности точек;

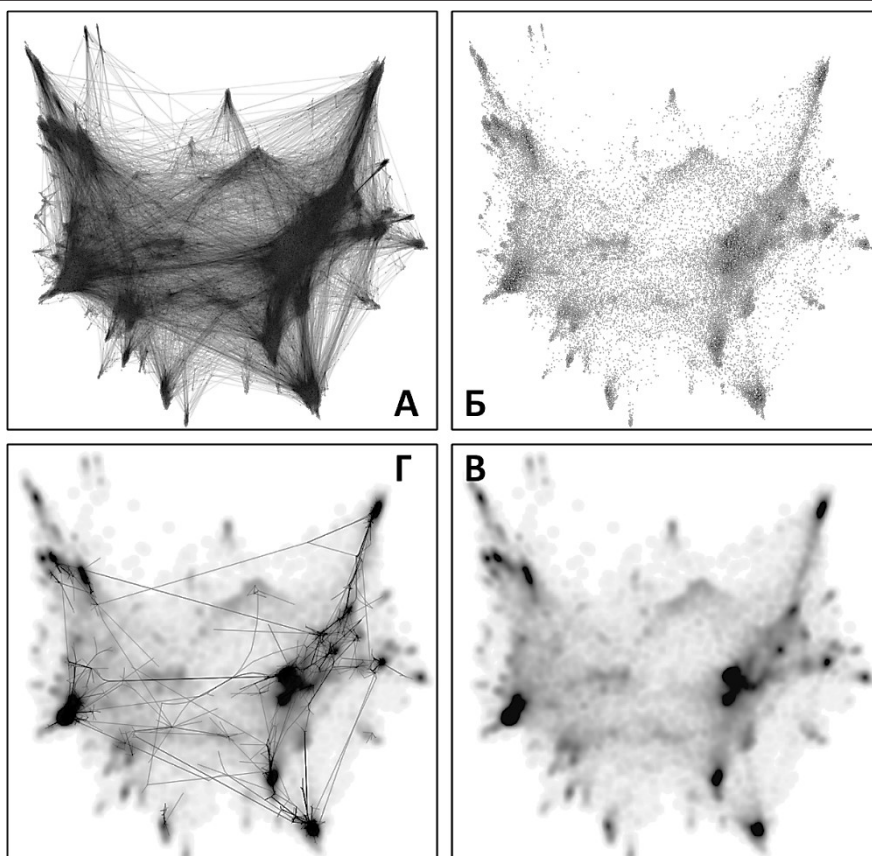


Рис. 5. Примеры ландшафтов, построенных на основе тепловых карт (А – исходный граф с 53 193 узлами и 188 640 рёбрами, Б – изображение только лишь узлов графа, В – тепловая карта, построенная на основе исходного графа, Г – изображение основных связей графа на тепловой карте)

- генерирование изолиний (линий, обозначающих участки одинаковой плотности) информационного ландшафта (при необходимости);

- создание полигонов на основе изолиний (при необходимости).

Выбор точечных данных осуществляется в соответствии с задачей исследований. В исходном графе могут присутствовать несколько типов узлов, каждый из которых влияет на общее расположения, но для задач исследований важны лишь некоторые типы. Например, требуется изобразить информационные объекты на основании тематического сходства новостных публикаций. В качестве узлов используются сами публикации, а связи между ними определяются на основании работы автоматического распознавания тем. Для корректировки итогового изображения в результате ошибок работы алгоритма автоматического обнаружения связей в исходный граф могут быть введены узлы,

представляющие собой метки тем, назначенные вручную. Такие узлы будут «притягивать» к себе связанные с ними узлы-новости, что положительно скажется на расположении соответствующих кластеров. Узлы-метки выполняют вспомогательную роль и для ландшафта карты не должны быть использованы.

Добавление изолиний к ландшафту позволяет очертить границы областей с одинаковой плотностью точек. Изолинии не только выполняют эстетическую функцию, но также могут играть роль координатной сетки в пределах соответствующего кластера. На основе изолиний могут быть построены полигоны для определения пространства кластеров. Как правило, для практических задач удобно строить два полигона. Полигон очерчивает область наиболее вероятного нахождения элементов кластера, а второй – область, где сконцентрированы элементы ядра кластера (рис. 6).

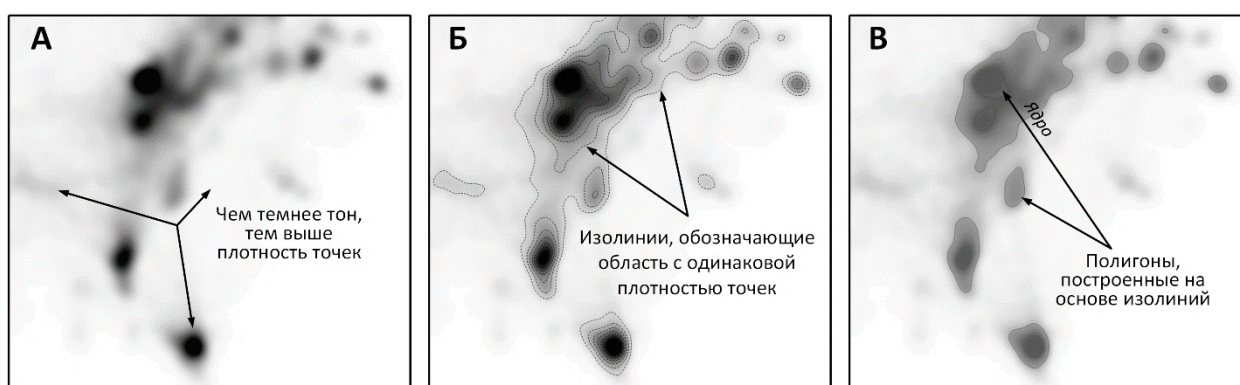


Рис. 6. Изолинии ландшафта информационной карты

Метод формирования ландшафта на основе преобразования Вороного

Диаграммы Вороного используются для разбиения плоскости на области в соответствии с некоторым множеством порождающих точек. Разбиение осуществляется таким образом, что внутри определённой зоны может находиться только одна порождающая точка, причём расстояние от неё до любой точки, находящейся внутри этой области меньше, чем расстояние от этих точек до любой другой порождающей точки.

Такое свойство позволяет формировать ландшафт информационной карты на основании расположения её особых объектов, ограничивающих зону влияния. Например, таким объектом может быть кластер, а соответствующая точка будет представлена

центроидом полигона этого кластера. Множество центроидов с помощью преобразования Вороного могут разделить пространства карты на области, каждая из которых соответствует определённому кластеру (рис. 7).

Из рис. 7 видно, что чем больше плотность точек, тем меньше площадь полигона. Одной из особенностей такого разбиения является то, что наиболее близкие между собой кластеры будут расположены рядом друг с другом (для кластеров сохраняется принцип близости). С другой стороны, такой способ не позволяет наглядно отобразить пересекающиеся кластеры. Чем больше таких кластеров, тем хуже выразительная способность метода формирования ландшафта.

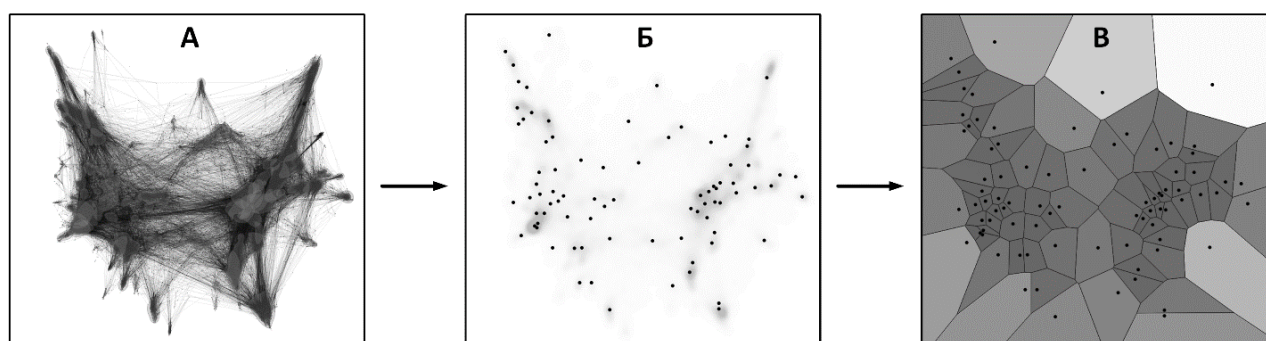


Рис. 7. Результат разбиения плоскости на участки, соответствующие кластерам исходного графа (А – исходный граф, Б – центроиды кластеров исходного графа, В – ландшафт информационной карты, построенной на основе диаграммы Вороного)

Другой подход к использованию диаграмм Вороного был продемонстрирован в работе [38] (рис. 8).



Рис. 8. Процедура формирования ландшафта с использованием диаграмм Вороного, реализованная в средстве R-Map [38]

Вначале, на основе исходного графа строится дерево (например, дерево результатов автоматической кластеризации). Затем дерево укладывается на плоскости с помощью пузырькового метода [39]. Для уложенного графа вычисляется охват, который ограничивает область холста. В этой области случайным образом генерируются точки, на основании которых строятся диаграммы Вороного. Данные полигоны используются для кодирования узлов в дереве репостов.

Методы визуализации объектов на информационной карте

На сформированном ландшафте можно проводить исследования объектов, которые должны быть представлены наглядным образом. Методы визуализации таких объектов зависят от их типа. Можно выделить следующие типы изображаемых объектов:

- точечные объекты;
- протяжённые объекты;
- вложенные объекты;
- динамические объекты;
- вспомогательные объекты (навигационные и справочные).

Визуализация точечных объектов реализуется с помощью:

- геометрических объектов разной формы, цвета и размера;
- пиктограмм;
- выносок с названием;
- анимированных объектов;
- тепловых карт.

Первый способ является одним из самых простых и действенных, особенно если точечных объектов достаточно много. Для данного способа требуется меньше всего вычислительных ресурсов, а восприятие точек интуитивно понятно. Если же точек слишком много, то для их визуализации

может быть применён метод создания тепловой карты, о котором было сказано ранее. Если визуализируемые объекты уникальны и, к тому же, имеют свое узнаваемое обозначение (например, логотип организации), то использование таких изображений будет лучшим вариантом. Выноска обычно используется, когда требуется указать название точечного объекта или же на карты нет свободного места. Анимированные объекты используются для демонстрационных целей, если требуется привлечь внимание наблюдателей.

Если требуется отображение читаемых меток для большого количества объектов, необходимо воспользоваться автоматическими алгоритмами укладки меток, классификация которых приведена в [40]. Открытым остаётся вопрос автоматического выбора ключевых слов для названия меток.

Визуализация протяжённых объектов, как правило, представляется в виде линий или полигонов. Как правило, это маршруты, границы территорий или зон распространения.

Для визуализации маршрутов могут применяться алгоритмы, позволяющие оптимизировать расположение линий, которыми они обозначены. Существуют алгоритмы силового связывания рёбер графа [41, 42], которые преобразуют прямые линии, соединяющие узлы графа, таким образом, что общие участки этих линий объединяются в пучки и не загромождают карту (рис. 9).

Основная идея силового связывания рёбер [42] заключается в следующем:

- линии разбиваются на участки, размер которых задаётся в качестве входного параметра алгоритма;

- итерационно осуществляется расчёт воздействия сил на такие участки (если два фрагмента ребра расположены близко друг к другу и выполняется ряд условий, таких как

угол между фрагментами меньше порогового значения, то такие фрагменты притягиваются).

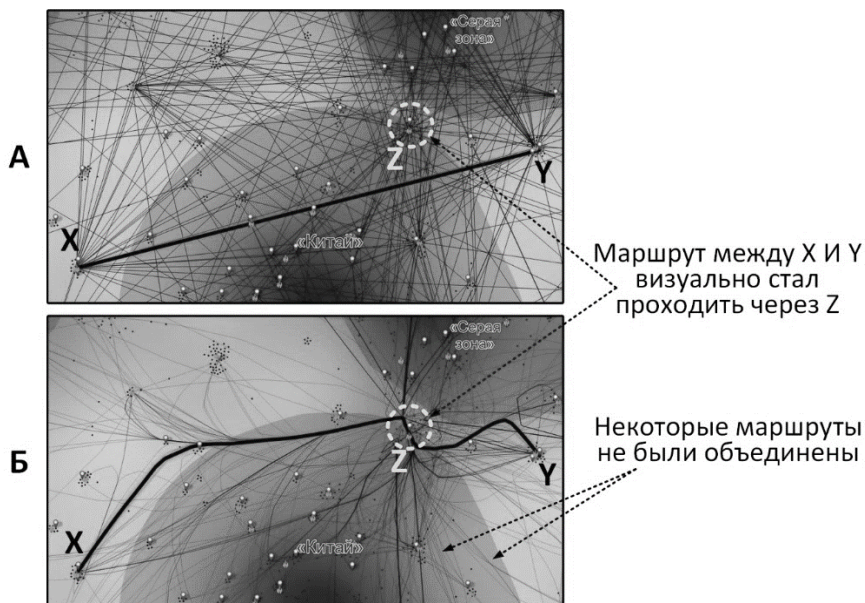


Рис. 9. Пример изображения точечных объектов и протяжённых объектов с использованием алгоритма силового связывания рёбер (А – исходный граф с прямыми рёбрами, Б – рёбра графа уложены в соответствии с алгоритмом силового связывания [42])

При этом необходимо отметить, что алгоритм имеет ряд оптимизаций из-за необходимости сокращения перебора большого числа комбинаций в ходе расчёта воздействия сил на фрагменты графа. Как правило, если в графе более 5 000 рёбер, то укладку рёбер достаточно сложно завершить за приемлемое время. Кроме того, как показано на рис. 9, некоторые рёбра могут остаться разделены, несмотря на их визуальную близость. Однако, что более существенно, появляются соединительные точки, через которые проходят новые маршруты. Наличие таких точек может

создать ложное впечатление о том, что в этом месте находится какой-либо узел графа. Если же такая точка попадает на место реально существующего узла, то исследователем могут быть сделаны ошибочные выводы в отношении проходящих маршрутов. Для прямых линии, если они проходят через узел, с которым не связаны, это можно заметить в отличии от изгибающихся линий.

Для графа, уложенного на сфере, проблема пересечения рёбер и наложение их на узлы решается за счёт использования третьего измерения (рис. 10).

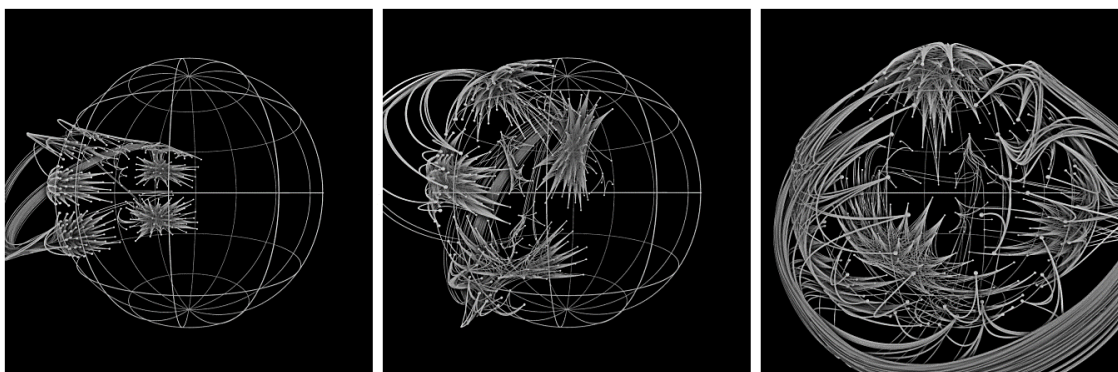


Рис. 10. Результаты применения алгоритма укладки рёбер на сфере [43]

В работах [27, 44] рассмотрены различные алгоритмы вложения графов. Их использование особенно важно при работе с картами мелкого масштаба, особенно при добавлении на холст вложенных объектов. Такая задача решается при разработке различных онтологических моделей, а также в ходе анализа признаков вредоносного программного обеспечения.

Отдельным направлением является визуализация процессов и динамических графов. В общем случае, процесс может быть изображён с помощью статического состояния карты (благодаря условным изображениям нескольких карт, соответствующих различным состояниям процесса), или же с помощью анимации изменяющихся структур ландшафта и/или объектов исследования. Существует два подхода к изображению изменяющихся структур [45, 46, 47, 48]:

- формирование ландшафта из всех объектов, участвующих в процессе;

- трансформация ландшафта каждый раз, когда происходит появление или исчезновение объектов.

В первом случае ландшафт информационной карты включает в себя полный спектр объектов, которые были активны как на ранних стадиях процесса, так и при его завершении. С одной стороны, данный способ удобен в реализации и не требует от исследователя каждый раз привыкать к новому ландшафту. При этом, если состав информационных объектов ландшафта и их взаимосвязи существенно отличается для различных стадий визуализируемого процесса, то контекст, изображаемый общим ландшафтом, не может адекватно передать особенности среды для соответствующей стадии.

Для интерактивных информационных карт появляется возможность быстрого перемещения (навигации) по карте, а также отображения дополнительной информации по представляющему интерес объекту или группе объектов. Возможности визуализации навигационных и справочных объектов карты зависят от средств ввода/вывода, а также наличия соответствующей вычислительной мощности. Способы традиционного интерактивного взаимодействия существенно

отличаются от взаимодействия с использованием средств виртуальной реальности. В первом случае средства и методы достаточно хорошо отработаны и заключаются в использовании манипулятора мыши и системы окон, которые можно открывать, закрывать и располагать в разных частях экрана. В таких окнах могут отображаться различные графики и диаграммы, «миникарты» (позволяющие быстро перемещаться в нужную точку информационной карты), окна с наборами фильтров и т.д. Данный способ лучше всего подходит для взаимодействия с картами в двухмерном пространстве.

Методы актуализации информационной карты

Одной из важнейших задач информационной картографии является возможность быстрой актуализации карт. Актуализация может потребоваться как на уровне сведений об объектах, исследуемых с помощью информационной карты, так и на уровне ландшафта. В отличие от географических карт, ландшафт информационной карты может меняться очень быстро, особенно в условиях динамически развивающихся информационных технологий. Кроме того, наличие различных проекций представления одних и тех же объектов киберпространства открывает возможность их совмещения. Методы актуализации информационной карты направлены на решение этих вопросов и включают:

- метод совмещения графов;
- метод трансформации ландшафта.

Совмещение графов

Две объединяемые информационные карты (будь то карты различных состояний одного и того же процесса, или различные проекции одинаковых объектов) могут быть объединены на уровне графов, лежащих в их основе.

Задача совмещения графов включает:

- оценку соответствия моделей данных, по которым формировались исходные данные для обоих графов;
- оценку соответствия наборов данных, используя которые были сформированы ландшафты информационных карт;

- оценку соответствия сведений об объектах, отображаемых на информационных картах.

Для каждой процедуры принимается одно из следующих решений:

- рассматриваемые данные идентичны для обеих карт и их корректировка не требуется;

- данные одной из карт имеют преимущества и должны быть взяты за основу при объединении;

- данные обеих карт обладают своими преимуществами и недостатками, поэтому требуется разработка нового решения, объединяющего положительные свойства обоих вариантов.

Чем больше различий в каждом из компонентов (модели данных, набор данных или сведения об объекте), тем проще решение задачи совмещения графов.

Пример использования метода совмещения графов был продемонстрирован в работе [29], где авторы провели анализ 20 существующих «карт» науки, с целью разработки единой карты науки. Несмотря на то, что полученная «карта» не удовлетворяет требованиям к информационной карте (не соблюдаются принципы измеримости сходства объектов, близости изображений схожих объектов и принцип изображения контекста), интересной представляется процедура формирования итогового графа связей научных дисциплин (рис. 11).

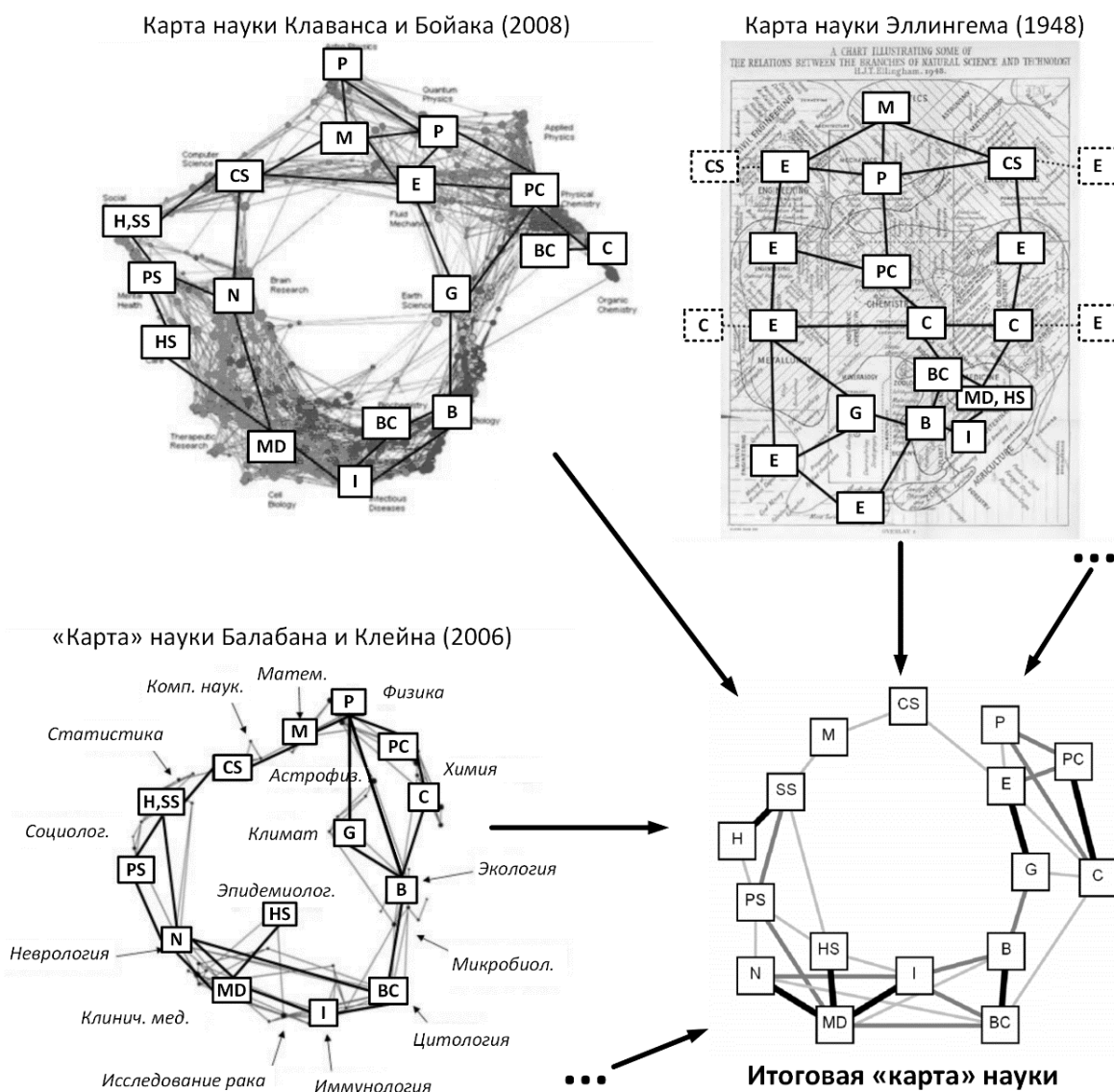


Рис. 11. Пример реализации метода совмещения графов [29]

Исследователями [29] в ходе анализа методов построения соответствующих карт и изображений науки было установлено, что в их основе используются различные модели данных (в некоторых случаях для построения карт использовались графы цитирования научных публикаций, где-то применялся граф связей научных журналов, другие карты строились на основании субъективного понимания конкретных исследователей).

Исходные данные для каждой карты отличались полностью. Кроме того, наблюдались существенные отличия в структуре графов и итоговой форме изображений. Для возможности совмещения результатов авторами предложен подход, основанный на выборе общих ключевых точек графов, присутствовавших в большинстве изображений и графов. Этими точками оказались крупные научные направления. В результате для каждой из 20 рассмотренных карт и изображений науки был сформирован граф связей таких ключевых точек, некоторые из которых представлены на рис. 11. На основании наиболее распространённых связей между ключевыми точками сформирована итоговая «карта» науки, включающая 16 научных направлений, расположенных в виде кольца, что подчёркивает наличие центральной науки.

При этом, необходимо отметить, что если анализируемых карт меньше, а каждая из карт обладает сходством как в модели, так и в исходных данных, то процедуру анализа проводить проще, а результирующая карта может обладать большими чертами сходства с каждой из анализируемых, чем карты, рассмотренные в [29].

В рассмотренном примере показан экспертный способ совмещения графов, однако, существуют возможности его автоматизации за счёт использования алгоритмов автоматического выравнивания графов (graph alignment, также используется термин «выравнивание сетей» – network alignment). Данные алгоритмы получили активное развитие благодаря прикладным задачам в области медицины (поиск данных в больших базах, сравнение белковых сетей, исследование структур мозга и др.) [49, 50, 51]. Алгоритмы выравнивания применяются

для сопоставления онтологий, баз знаний [52], а также других задач, связанных со сравнением сетевых структур.

В общем случае алгоритмы выравнивания делятся на глобальные и локальные. Первый класс позволяет сопоставить графы на основании их топологии, второй – определяет схожие фрагменты в структуре сравниваемых сетей. В [50] предложен алгоритм, объединяющий оба класса (сначала проводится сопоставление крупных структурных элементов, после чего осуществляется поиск соответствующих элементов). Обзор основных техник выравнивания и их практических реализаций представлен в [51], где техники разделены на два вида:

- спектральные (направлены на сопоставление двух сетей на основе их матриц смежности);
- сетевые методы обучения (учитываются дополнительные признаки узлов сетей).

Таким образом, если в результате совмещения графов, было принято решение об изменении основы информационной карты, то потребуются её полная перестройка, включая выполнение процедуры построения ландшафта.

Трансформация ландшафта

Метод трансформации ландшафта основан на изменении лишь формы информационной карты без перестройки графа, лежащего в её основе. Реализация метода связана только с геометрическими преобразованиями векторного или растрового изображения ландшафта. Применить данный метод легче, чем осуществить полный цикл операций совмещения графов, однако его результаты имеют низкую точность интеграции данных с совмещённой карты (трансформация ландшафта сохраняет связь между объектами только в рамках координат двухмерного или трёхмерного пространства). Данный метод полезен в качестве механизма быстрого сравнения двух карт. В этом случае одна карта используется как базовая, а вторая загружается в виде изображения, которое трансформируется для того, чтобы основные точки ландшафтов совпали друг с другом (рис. 12)

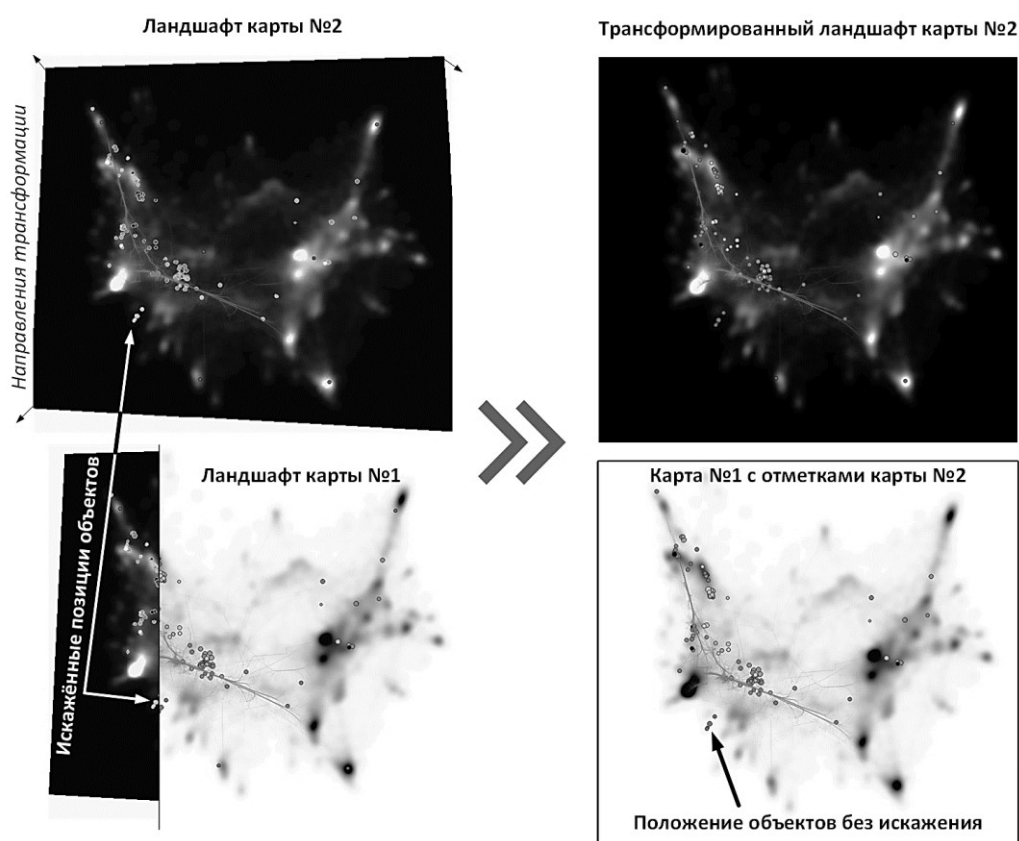


Рис. 12. Результаты совмещения двух информационных карт, для получения которых была выполнена трансформация ландшафта второй карты

Обычно для трансформации ландшафта используются алгоритмы, основанные на опорных точках. Предварительно выбираются несколько точек на ландшафтах обеих карт, между которыми имеется однозначное соответствие. Такие точки называются опорными. На основании этих точек определяются правила трансформации. К наиболее распространённым типам трансформации относятся [53]:

- линейная трансформация (равномерное масштабирование ландшафта без вращений и других видов преобразований);
- трансформация Гельмерта (допускает операцию вращения);
- полиномиальная трансформация (дополнительно позволяет осуществлять операции сдвига и искривления фрагментов ландшафта);
- трансформация TPS-RPM (трансформация сплайнами, позволяющая более тонко изменять ландшафт, требует больше опорных точек, чем предыдущие алгоритмы).

Из-за ошибок снижения размерности, трансформация ландшафта приводит к усилению искажений, особенно в тех

областях, где значение функции стресса наибольшее.

Картографические методы анализа карт защищаемого киберпространства

В информационной картографии могут применять следующие методы традиционной картографии:

- графический оверлей;
- разметка областей карты;
- навигация и интерактивное взаимодействие с картой.

Графический оверлей является картографическим методом, направленным на получение знаний в результате анализа совмещённых карт [54]. Для реализации метода необходимо приведение карт к единой проекции.

Применительно к информационной картографии данный метод может быть реализован в трёх вариантах (рис. 13):

- совмещение изображений различных процессов или явлений на неизменном ландшафте информационной карты (вариант 1 – наложение объектов);
- совмещение различных ландшафтов (выполненных в единых проекциях) при

неизменности изображаемого процесса или явления (вариант 2 – наложение ландшафтов);

- совмещение различных состояний одного процесса или явления для различных состояний ландшафта информационной карты (вариант 3 – наложение состояний).

Первый вариант является самым распространённым и используется для выявления сходства и различия между исследуемыми объектами в рамках единого контекста.

Второй и третий варианты предназначены для исследования динамических процессов. Как было сказано ранее, представление ландшафта для динамических процессов может меняться. Для того, чтобы отследить это изменение применяется второй вариант. Если же одновременно нужно понять, какие при этом произошли изменения с объектами исследования, то применяется третий вариант.

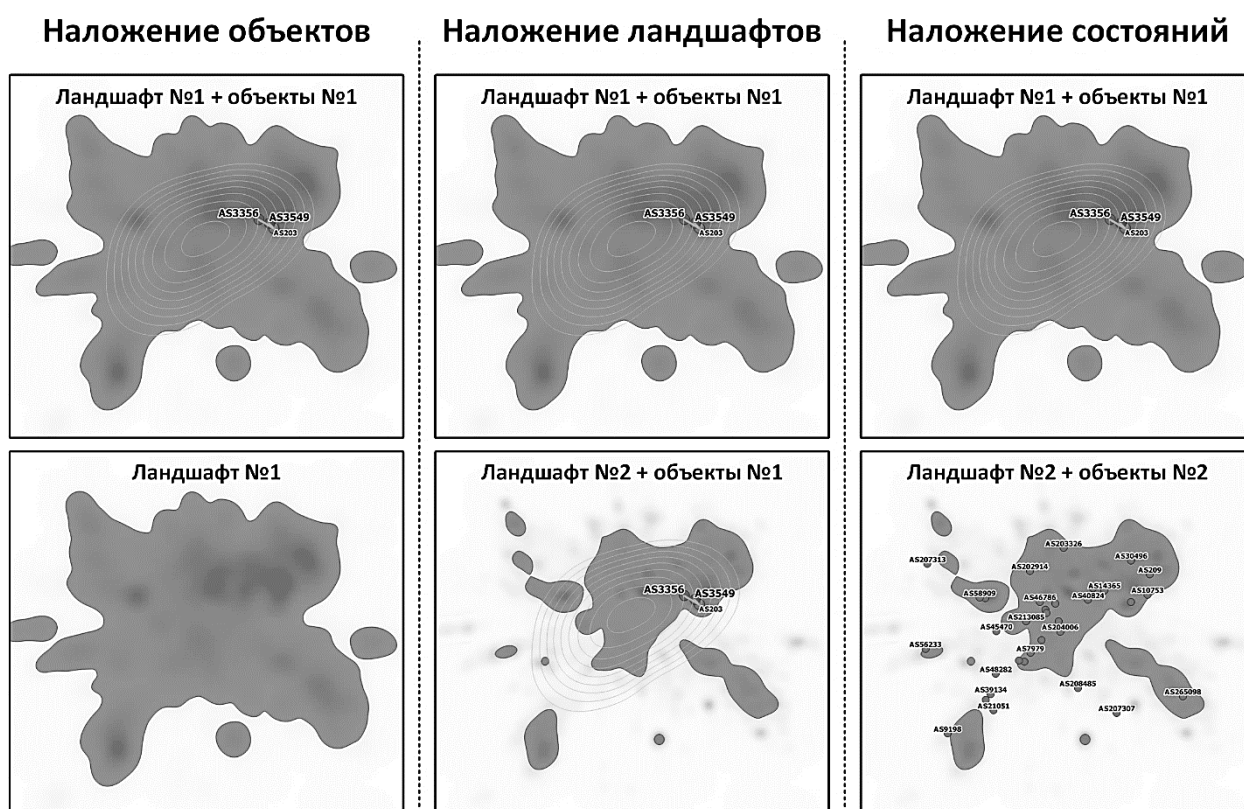


Рис. 13. Варианты реализации графического оверлея

Разметка областей карты, осуществляется применительно к ландшафту и может осуществляться как в ручном, так и в автоматическом режиме. Операции разметки включают:

- нанесение точечных объектов (определение объектов исследования);
- нанесение линейных объектов (формирование границ и маршрутов);
- нанесение полигональных объектов (формирование зон и территорий).

Автоматическое формирование точечных объектов может быть осуществлено на основании их координат в пространстве ландшафта информационной карты или по

отношениям с уже размещёнными на ней объектами.

Формирование линейных объектов осуществляется аналогичным образом. Кроме того, линейные объекты могут быть сформированы на основании точечных объектов, если указан их порядок следования.

Полигональные объекты могут быть сформированы различными способами, среди которых:

- задание координат вершин полигонов и кривизны их сторон;
- охват выбранных объектов информационной карты (зона, включающая все выбранные объекты);

- использование изолиний тепловой карты, сформированной на основании точечных объектов;

- трансформация замкнутых линейных объектов;

- применение диаграмм Вороного;

- арифметические операции с полигональными объектами (объединение, вычитание, исключение и др.).

Для точечных, линейных и полигональных объектов могут быть установлены названия и другие свойства. Наличие разметки позволяет автоматизировать процедуры анализа (например, рассчитывать центральные области кластеров, производить пространственный выбор объектов на основании их попадания в заданную область и др.).

Навигация по карте может осуществляться:

- на основании координат в рамках ландшафта информационной карты;

- путём использования маркеро-закладок, привязанных к определённым областям информационной карты;

- в результате перемещения к выделенному объекту;

- путём использования «миникарты»;

- в ручном режиме «масштабирования-перемещения».

Для перемещения по карте, представленной в трёхмерном пространстве, могут быть использованы специализированные устройства-манипуляторы.

В работах [55, 56] представлены классификации *интерактивных операций*, выполняемых исследователем в ходе взаимодействия с картой. На основании их анализа дополнительно (к операциям навигации) выявлены следующие операции:

а) операции включения:

- импорт данных;

- экспорт данных;

- возврат к историческим состояниям;

- редактирование карты (исходных данных, ландшафта, объектов);

- аннотирование карты;

б) операции изменения внешнего вида:

- переключение между несколькими состояниями;

- размещение нескольких карт;

- изменение стиля карты;

- фильтрация и выбор объектов.

Операции импорта позволяют загружать наборы данных или ранее созданную карту. Операции обычно применяются в начале сеанса работы с картой, но по мере необходимости могут загружаться дополнительные сведения. Также могут загружаться потоковые данные (данные, наносимые на карту или добавляемые в систему хранения по мере их поступления).

Операции экспорта позволяют сохранять информационную карту, результаты её анализа или собираемые исходные даты, лежащие в её основе, с целью использования их за пределами карты.

Операции возврата к историческим состояниям позволяют отменять или возвращаться к ранее выполненным операциям в зависимости от их результата. Они достаточно важны для интерактивного анализа, так как некоторые операции выполняются достаточно долго, но при выявлении ошибок в результате их завершения должен быть предусмотрен лёгкий способ отмены совершённых действий, позволяющий с наименьшими затратами скорректировать параметры таких расчётов или не выполнять их вовсе.

Операции редактирования исходных данных позволяют вносить изменения в исходные данные, чтобы повлиять на процедуру построения карты. Это крайне важно при обнаружении недостатков в исходных данных. Должен быть предусмотрен механизм повторной реализации операций построения карты на основе уточнённых данных, чтобы минимизировать усилия и ошибки исследователя.

Операции редактирования ландшафта карты предполагают возможность изменения формы ландшафта, сформированного на основе исходных данных.

Операции редактирования карты связаны с добавлением, созданием, удалением, присоединением и перемещением объектов, нанесённых на карту.

Операции аннотирования заключаются в добавлении текстовых меток или связывании друг с другом объектов карты, ландшафта и

исходных данных. В отличие от разметки карты, аннотирование относится к объектам, изображаемым на карте, а не к областям ландшафта. Аннотирование позволяет добавлять на карту вспомогательную информацию. Данная операция может выполняться совместно несколькими исследователями. При этом должны фиксироваться сведения об авторе аннотации и времени её добавления (изменения, удаления). Совместное аннотирование выполняет важную коммуникационную функцию, позволяя синхронизировать понимание карты между различными участниками исследования.

Операции переключения между несколькими состояниями информационной карты обычно используются для следующих задач:

- исследование динамических процессов;
- выявление различий между источниками информации;
- демонстрации результатов анализа.

Операции размещения в видимой области сразу нескольких карт обеспечивают возможность одновременного (в рамках единого визуального пространства) сравнения нескольких карт или их состояний.

Операции изменения стиля карты позволяют определять единые шаблоны для визуального представления всех объектов слоя. Эта функция позволяет быстро адаптировать внешний вид карты для задач использования информационной карты. Изменение стиля может относиться как ко всей карте в целом, так и к отдельным слоям или объектам карты. К операциям изменения стиля карты также относятся операции панорамного вида, трёхмерного представления карты, стилизации под географическую карту и др.

Фильтрация и выбор объектов карты могут осуществляться на основании правил или в ручном режиме. Фильтрация (выбор) на основании правил позволяет задавать характеристики информационных объектов на основании его свойств (например, значения каких-либо полей данных, связанных с объектами) или пространственных отношений (например, попадание в область, определяемую другими объектами). В ручном режиме с помощью

манипулятора типа «мышь» пользователь огораживает область пространства карты, в которую попадают нужные объекты. Далее с помощью контекстного меню осуществляются действия с выбранными объектами (например, отображение меток или вспомогательных данных).

Для выбранных объектов могут быть изменены их характеристики, такие как цвет, форма и др. Также эти объекты могут быть сохранены, скопированы или удалены с карты и из системы хранения данных.

Сетевые методы анализа карт защищаемого киберпространства

Сетевые методы анализа карты позволяют использовать автоматические алгоритмы расчёта разнообразных метрик графа, лежащего в основе информационной карты. Такие метрики предоставляют эксперту дополнительные возможности для визуальных и аналитических рассуждений. Сетевые методы подробно рассмотрены в монографии [57]. Результаты автоматической кластеризации, а также метрик узлов и рёбер позволяют дополнить визуальное представление информационной карты, расширить набор признаков её информационных объектов и более эффективно решать следующие задачи:

- выявление типовых структур графа;
- выявление особых узлов графа;
- выявление особых связей графа;
- анализ маршрутов графа;
- анализ изменений графа.

Выявление типовых структур графа

Под типовой структурой графа понимается его подграф, обладающий рядом выраженных свойств, удовлетворяющих критериям решаемой задачи.

Выявление типовой структуры связано с описанием фрагмента графа по характеристикам его узлов и рёбер. Такими фрагментами могут выступать:

- клики (группа узлов графа, в рамках которой все узлы связаны между собой);
- кластеры (группы узлов графа, которые обладают схожими признаками, например, наибольшей связанностью узлов внутри кластера);
- последовательности узлов с заданными параметрами.

Выявление таких структур может быть осуществлено как на основе формальных правил, так и в результате визуального анализа. Зачастую, для определения формальных правил требуется понимание того, какие структуры надо найти. Для этого можно использовать визуальный метод применительно к графу, уложенному в пространстве. Эффективность такого анализа во многом зависит от способа укладки и опыта исследователя.

Во многих задачах защиты информации и обеспечения информационной

безопасности требуется выявление типовых структур. Например, для анализа ущербов от компьютерных атак в общем потоке финансовых транзакций требуется выявление только тех транзакций, которые совершали киберпреступники. Т.е. для графа транзакций (рис. 14), узлами которого являются идентификаторы участников финансового обмена, а рёбрами – суммы перевода, требуется найти типовые структуры, характерные для финансовых транзакций киберпреступников.

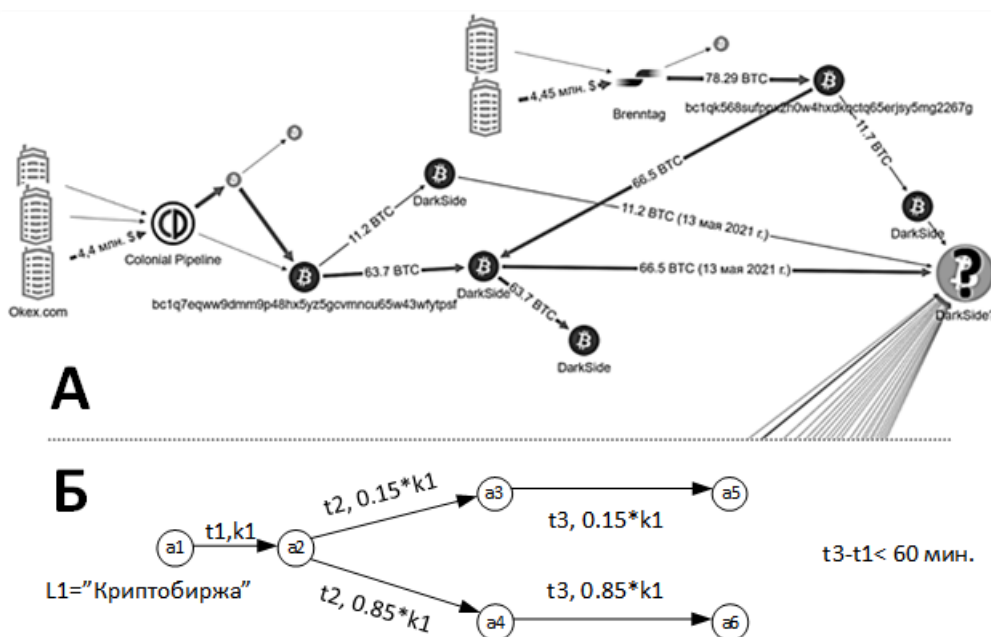


Рис. 14. Визуализация транзакций злоумышленника в ходе атак DarkSide (А) и её типовая структура (Б)

Для определения формы типовых структур требуется знание поведения злоумышленников. Одним из стратегий киберпреступников, к которым относятся программы-вымогатели, является разделение выкупа, полученного от жертвы, между партнёрами, участвовавшими в атаке. Например, после получения вознаграждения злоумышленники в течении часа разделяет полученную сумму на две части в отношении 15 % на 85 %, каждая из которых переводится на новый кошелёк (рис. 14, А).

То есть на основании типовой транзакции (рис. 14, Б) можно сформировать шаблон, при помощи которого будут найдены фрагменты общего графа транзакций, для которых выполняются соответствующие условия.

Важное значение для выявления типовых структур графа имеют алгоритмы кластеризации. Кластеризации графов посвящено много (в том числе [58, 59]). Существует проблема выбора подходящего алгоритма кластеризации и настройки его параметров (рис. 15).

Например, иерархическая кластеризация позволяет разбить исходный кластер на группу вложенных кластеров (рис. 15, Б). Такая кластеризация полезна при распределении структур по слоям информационной карты. Каждый кластер будет виден при соответствующем масштабе карты. Лейденский алгоритм в зависимости от разрешения позволяет разбить исходный граф на разное число непересекающихся кластеров. При низком разрешении будут выявлены крупные структуры графа, при

высоком – граф разбивается на больше число кластеров, при этом не всегда для разных частей графа удаётся подобрать нужное разрешение. Так при разрешении 4 можно

выявить кластеры 8 и 9 (рис. 15 Г), однако узлы 5 и 6 в другой части графа распознаются как отдельные кластеры, что является ошибкой.

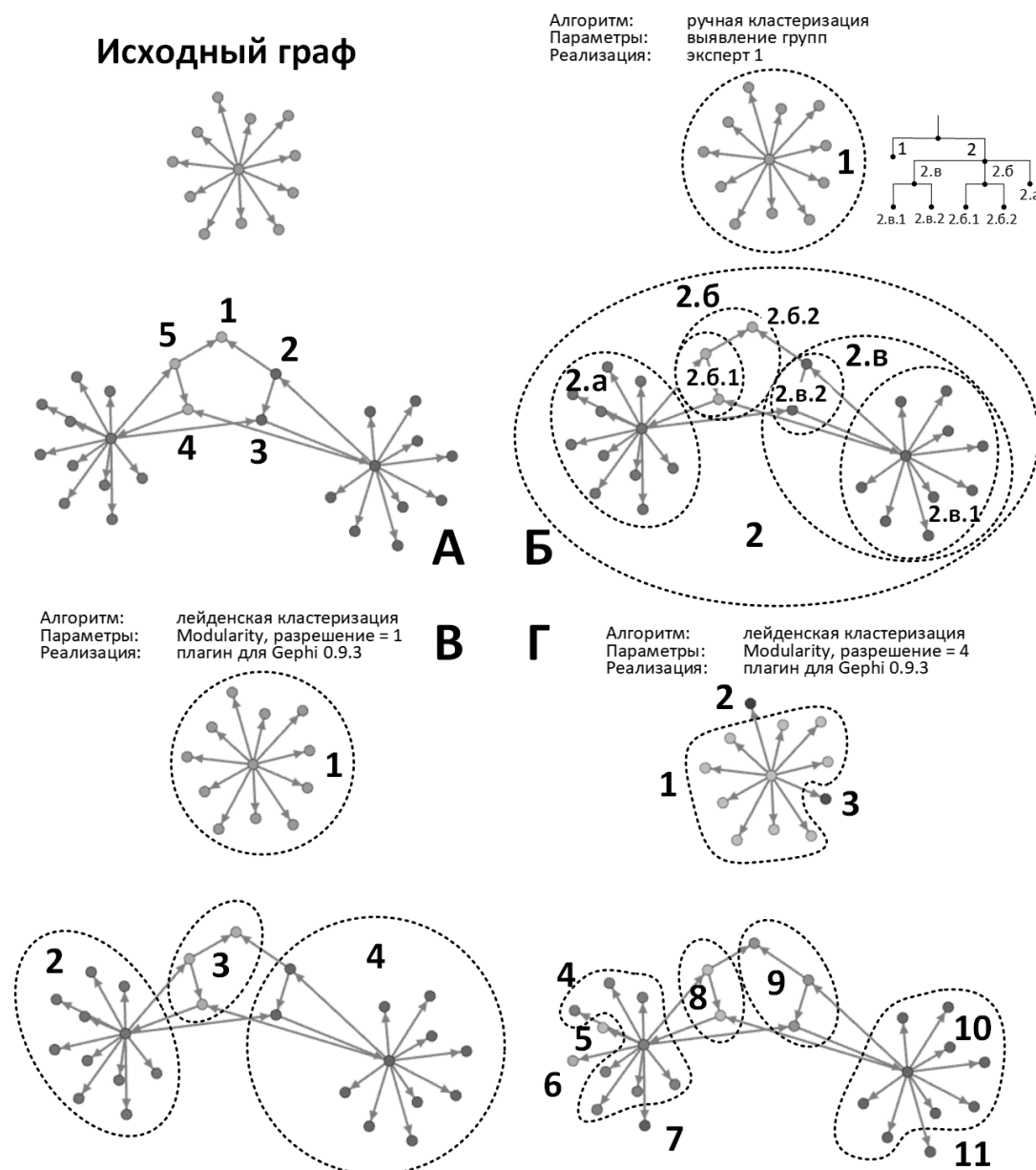


Рис. 15. Сравнение различных алгоритмов кластеризации (А – исходный граф, Б – иерархическая кластеризация, В и Г – лейденская кластеризация)

Выявление особых узлов графа

Метод выявления особых узлов в отличии от предыдущего метода направлен на поиск отдельного элемента, а не конкретного графа. Для этой задачи используются различные метрики центральности.

Различие метрик центральности обусловлено не только алгоритмом их расчёта, и выбором начальных параметров, но также существенно зависит от задачи

исследования и структуры данных. В одних случаях требуется поиск узлов, которые играют ключевую роль в своём кластере, другие задачи могут быть связаны с поиском узлов, объединяющих два кластера. Продемонстрируем особенности выявления особых узлов на примере графа цитирования научных публикаций.

Способы анализа информационной карты, показанной на рис. 16, основаны на значениях следующих метрик:

- веса вершины, подсчитанного для входящих рёбер (рис. 16, А);
- веса вершины, подсчитанного для исходящих рёбер (рис. 16, Б);
- метрики PageRank, подсчитанной для ориентированного графа (рис. 16, В);
- метрики PageRank, подсчитанной для неориентированного графа (рис. 16, Г);
- метрики Strongly-Connected (рис. 16, Д);
- коэффициента кластеризации (рис. 16, Е).

Граф цитирования научных публикаций уложен с помощью алгоритма ForceAtlas [20]. Размер узлов графа отражает его значение соответствующей метрики.

Количество входящих связей в графе цитирования научных публикаций показывает, насколько много источников цитирует соответствующая публикация. Если у публикации много входящих связей и все они относятся к вершинам одного кластера, то такая публикация скорее всего является обзорной. Если связанные узлы расположены в разных кластерах, то это междисциплинарная публикация, которая скорее всего является книгой. Анализируя визуальное представление рис. 16, Б можно быстро выявить такие публикации.

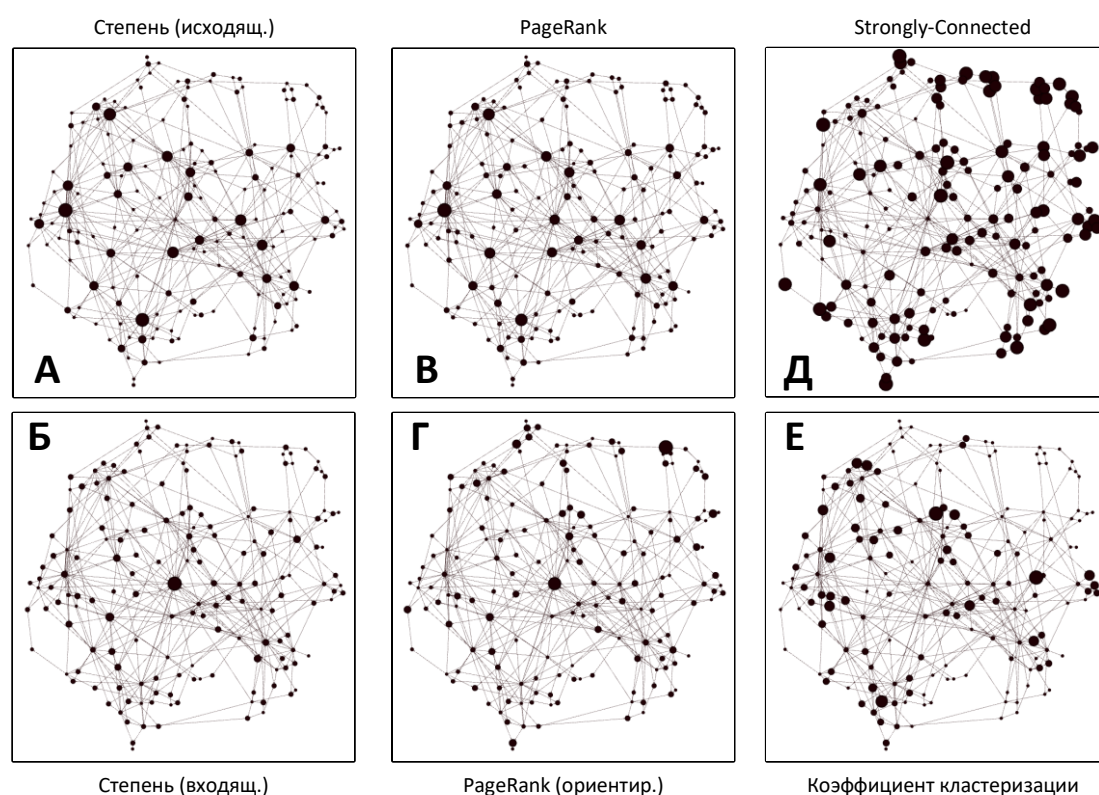


Рис. 16. Примеры визуализации метрик центральности, значения которых показаны в виде размеров узлов

С другой стороны, если публикацию цитируют много научных статей (т.е. степень вершины, измеренная по количеству исходящих связей, имеет высокое значение), то такая публикация является первоисточником популярной идеи. Обнаружить такие публикации визуально достаточно просто (рис. 16, А). Как видно из рисунка, таких публикаций больше, чем обзорных.

Публикации-источники идей и обзорные статьи расположены в центрах соответствующих кластеров, поэтому их

можно назвать кластерообразующими узлами. Лучше всего кластерообразующие узлы можно выявить с помощью метрики PageRank, вычисленной для неориентированного графа (рис. 16, В).

Для задачи анализа научных публикаций также интерес представляют узлы, которые связывают несколько кластеров и, при этом, не имеют большого количества исходящих рёбер. Как правило, это междисциплинарные научные статьи, содержащие новые идеи, которые еще не получили развития в других публикациях. Их поиск может быть

осуществлен с помощью метрик Strongly-Connected (рис. 16, Д) и коэффициента кластеризации (рис. 16, Е). Первая выделяет узлы-мостики, а вторая – узлы, относящиеся к минимальному количеству кластеров. Искомые узлы должны иметь большие значения по данным показателям.

Аналогичным образом осуществляется выявление особых узлов для информационных карт.

Выявление особых связей графа

Метод выявления особых связей графа аналогичен предыдущему методу, но в отношении метрик его рёбер. Выявление особых связей обычно связано с задачей анализа различных типов маршрутов графа. К таким задачам, к примеру, можно отнести:

- классическую задачу выявления кратчайших маршрутов (для задачи информационной картографии кратчайший маршрут может использоваться в качестве значения расстояния между информационными объектами защищаемого киберпространства в многомерном пространстве признаков, оно может существенно отличаться от расстояния между изображениями таких объектов на информационной карте для различных проекций) [60];

- задачу коммивояжёра (например, для задачи планирования операции может потребоваться изображение на информационной карте самого короткого маршрута атаки, проходящей через все ключевые точки атакуемой информационной системы) [60];

- задачу канадского путешественника (например, как и в предыдущем случае задачи планирования операции может потребоваться изображение на информационной карте самого короткого маршрута атаки, но, при этом, нет достоверных сведений о наличии связей между её компонентами, а есть информация только лишь о вероятностных характеристиках) [11].

В некоторых случаях для выявления особых связей могут использоваться метрики узлов (например, для отображения маршрутов единичной длины, проходящих через самые «влиятельные» узлы графа).

Для выявления особых связей графа обычно применяются взвешенные сети,

методы анализа которых были подробно рассмотрены в монографии [61].

Таким образом, методы сетевого анализа являются базовыми аналитическими методами информационной картографии и применяются как для решения задач построения ландшафтов информационных карт, так и для анализа объектов исследования, отображаемых на таких ландшафтах.

Список литературы

1. Сердечный А.Л. Концептуальные основы картографии защищаемого киберпространства. Часть 1 // Информация и безопасность. Т. 24. Вып. 3. 2021. С. 373-386.
2. Сердечный А.Л. Концептуальные основы картографии защищаемого киберпространства. Часть 2 // Информация и безопасность. Т. 24. Вып. 3. С. 413-424.
3. Сердечный А.Л. Информационно-картографические системы как инструментальная основа картографии защищаемого киберпространства / Системы управления и информационные технологии. 2021. № 4 (86). С. 41-46.
4. Kaggle Datasets. URL: <https://www.kaggle.com/datasets> (дата последнего обращения 10.09.2021).
5. Mitchell R. Web scraping with Python: Collecting more data from the modern web // O'Reilly Media. 2018. 255 с.
6. Gunawan R. Comparison of web scraping techniques: regular expression, HTML DOM and Xpath / R. Gunawan, A. Rahmatulloh и др. // International Conference on Industrial Enterprise and System Engineering (IcoIESE 2018) Comparison. 2019. Т. 2. С. 283-287.
7. Kumar M. A survey of Web crawlers for information retrieval / M. Kumar, R. Bhatia, D. Rattan // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2017. Т. 7. №. 6. С. e1218.
8. Kumar S. Review of Focused Crawling Schemes for Search Engine / S. Kumar, M.A. Gupta // Smart Trends in Computing and Communications: Proceedings of SmartCom 2020. 2021. С. 311-317.
9. Slowik A. Swarm Intelligence Algorithms (Two Volume Set) // CRC Press. 2021. 363 с.

10. Molina D. Comprehensive taxonomies of nature-and bio-inspired optimization: Inspiration versus algorithmic behavior, critical analysis recommendations / D. Molina, J. Poyatos, J. Del Ser [и др.] // *Cognitive Computation*. 2020. Т. 12. №. 5. С. 897-939.
11. van Ee M. Approximation and complexity of multi-target graph search and the Canadian traveler problem / M. van Ee, R. Sitters // *Theoretical Computer Science*. 2018. Т. 732. С. 14-25.
12. Sun C. A Review of Unsupervised Keyphrase Extraction Methods Using Within-Collection Resources / C. Sun, L. Hu, S. Li [и др.] // *Symmetry*. 2020. Т. 12. №. 11. С. 1864.
13. Papagiannopoulou E. A review of keyphrase extraction / E. Papagiannopoulou и др. // *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. 2020. Т. 10. №. 2. С. e1339.
14. Wikimapia. URL: <https://wikimapia.org/> (дата обращения 10.09.2021).
15. Народная карта Яндексa . – URL: <https://n.maps.yandex.ru/> (дата обращения 10.09.2021).
16. Déom P. Analyse des solutions de spatialisation dans les graphes NoSQL afin d'exploiter la topologie arc nœud des réseaux routiers. Quelle est actuellement la solution la plus performante pour réaliser des requêtes spatiales sur un réseau routier important?. 2019. 80 с.
17. Oyana T.J. Spatial Analysis with R: Statistics, Visualization, and Computational Methods // CRC press. 2020. 355 с.
18. Xie L. Scene recognition: A comprehensive survey / L. Xie, F. Lee, L. Liu и др. // *Pattern Recognition*. 2020. Т. 102. С. 107205.
19. Как и зачем мы внесли 22 тысячи изменений в дизайн Яндекс.Карт. URL: <https://habr.com/ru/company/yandex/blog/544940/> (дата обращения 10.09.2021).
20. Jacomy M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software / M. Jacomy, T. Venturini, S. Heymann и др. // *PloS one*. 2014. Т. 9. №. 6. С. e98679.
21. Rahman M.K. BatchLayout: A batch-parallel force-directed graph layout algorithm in shared memory / M. K. Rahman, M. H. Sujon, A. Azad // 2020 IEEE Pacific Visualization Symposium (PacificVis). IEEE, 2020. С. 16-25.
22. McInnes L. Umap: Uniform manifold approximation and projection for dimension reduction / L. McInnes, J. Healy, J. Melville // *arXiv preprint arXiv:1802.03426*. 2018. С. 1-51.
23. Böhm J.N. A unifying perspective on neighbor embeddings along the attraction-repulsion spectrum / J. N. Böhm, P. Berens, D. Kobak // *arXiv preprint arXiv:2007.08902*. 2020. С. 1-20.
24. Van der Maaten L. Visualizing data using t-SNE / L. Van der Maaten, G. Hinton // *Journal of machine learning research*. №. 11. Т. 9. 2008. С. 1-20.
25. Sugiyama K. Graph drawing by the magnetic spring model / K. Sugiyama, K. Misue // *Journal of Visual Languages & Computing*. 1995. Т. 6. №. 3. С. 217-231.
26. Klimovskaia A. Poincaré maps for analyzing complex hierarchies in single-cell data / A. Klimovskaia, D. Lopez-Paz, L. Bottou и др. // *Nature communications*. 2020. Т. 11. №. 1. С. 1-9.
27. Dogrusoz U. A layout algorithm for undirected compound graphs / U. Dogrusoz, E. Giral, A. Cetintas и др. // *Information Sciences*. 2009. Т. 179. №. 7. С. 980-994.
28. Schulz C. Visualizing spreading phenomena on complex networks / C. Schulz // *arXiv preprint arXiv:1807.01390*. 2018. С. 1-17.
29. Brunn S.D. What is where? The role of map representations and mapping practices in advancing scholarship / S. D. Brunn, M. Dodge // *Mapping Across Academia*. Springer, Dordrecht, 2017. С. 1-22.
30. A Chart Illustrating Some of the Relations between the Branches of Natural Science and Technology. URL: <https://scimaps.org/map/7/2> (дата обращения 10.09.2021).
31. Griffith B.C. The structure of scientific literatures II: Toward a macro-and microstructure for science / B.C. Griffith, H.G. Small, J.A. Stonehill // *Science studies*. 1974. Т. 4. №. 4. С. 339-365.
32. Small H. The geography of science: disciplinary and national mappings / H. Small, E. Garfield // *Journal of information science*. 1985. Т. 11. №. 4. С. 147-159.
33. Perry S. Drawing Graphs on the Sphere / S. Perry, M. S. Yin, K. Gray [и др.] //

- Proceedings of the International Conference on Advanced Visual Interfaces. 2020. С. 1-9.
34. Fruchterman T.M.J. Graph drawing by force-directed placement / T.M.J. Fruchterman, E.M. Reingold // *Software: Practice and experience*. 1991. Т. 21. №. 11. С. 1129-1164.
35. Gove R. A Random Sampling O (n) Force-calculation Algorithm for Graph Layouts / R. Gove // *Computer Graphics Forum*. 2019. Т. 38. №. 3. С. 739-751.
36. A Historical View of the AS Core. URL: http://www.caida.org/research/topology/as_core_network/historical.xml (дата обращения 10.09.2021).
37. Höggräfer M. The State of the Art in Map-Like Visualization / M. Höggräfer, M. Heitzler, H. J. Schulz // *Computer Graphics Forum*. 2020. Т. 39. №. 3. С. 647-674.
38. Chen S. R-Map: A Map Metaphor for Visualizing Information Reposting Process in Social Media // *IEEE Transactions on Visualization and Computer Graphics*. 2019. Т. 26. №. 1. С. 1204-1214.
39. Grivet S. Bubble tree drawing algorithm / S. Grivet D. Auber J. P. Domenger // *Computer vision and graphics*. Springer, Dordrecht, 2006. С. 633-641.
40. Bekos M.A. External labeling techniques: A taxonomy and survey / M. A. Bekos, B. Niedermann [etc.]. // *Computer Graphics Forum*. 2019. Т. 38. №. 3. С. 833-860.
41. Battista G.D. Graph drawing: algorithms for the visualization of graphs / G.D. Battista, P. Eades, R. Tamassia, [и др.] // Издательство Prentice Hall PTR. 1998. 400 p.
42. Graser A. Untangling origin-destination flows in geographic information systems / A. Graser, J Schmidt // *Information Visualization*. 2019. Т. 18. №. 1. С. 153-172.
43. Kwon O.H. A study of layout, rendering, and interaction methods for immersive graph visualization / O H. Kwon, C. Muelder, K. Lee [и др.] // *IEEE transactions on visualization and computer graphics*. 2016. Т. 22. №. 7. С. 1802-1815.
44. Vehlow C. The State of the Art in Visualizing Group Structures in Graphs / C. Vehlow, F. Beck, D. Weiskopf // *EuroVis (STARs)*. 2015. С. 21-40.
45. Meidiana A. New Quality Metrics for Dynamic Graph Drawing / A. Meidiana, S.H. Hong, P. Eades // arXiv preprint arXiv:2008.07764. 2020. 1-17 с.
46. Beck F. A taxonomy and survey of dynamic graph visualization / F. Beck, M. Burch, S. Dieh [и др.] // *Computer Graphics Forum*. 2017. Т. 36. №. 1. С. 133-159.
47. Bender S. The art and science of dynamic network visualization / S. Bender-deMoll, D. A. McFarland // *Journal of Social Structure*. 2006. Т. 7. №. 2. С. 1-38.
48. Chen X. A survey of multi-space techniques in spatio-temporal simulation data visualization / X. Chen, L. Shen, Z. Sha [и др.] // *Visual Informatics*. 2019. Т. 3. №. 3. С. 129-139.
49. Elmsallati A. Global alignment of protein-protein interaction networks: A survey / A. Elmsallati, C. Clark, J. Kalita // *IEEE/ACM transactions on computational biology and bioinformatics*. 2015. Т. 13. №. 4. С. 689-705.
50. Milano M. GLAlign: Using global graph alignment to improve local graph alignment / M. Milano, M. Cannataro [и др.] // *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2016. С. 1695-1702.
51. Trung H.T. A comparative study on network alignment techniques / H. T. Trung, N.T. Toan, T. Van Vinh [и др.] // *Expert Systems with Applications*. 2020. Т. 140. С. 112883.
52. Geng Q. Cross-domain ontology construction and alignment from online customer product reviews / Q. Geng, S. Deng, D. Jia [и др.] // *Information Sciences*. 2020. Т. 531. С. 47-67.
53. QGIS. Правила трансформации. URL: https://docs.qgis.org/testing/en/docs/user_manual/working_with_raster/georeferencer.html#defining-the-transformation-settings (дата обращения 10.09.2021).
54. Берлянт А.М. Картографический метод исследования // Изд-во Моск. ун-та. 1978. 257 с.
55. Roth R.E. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization / R. E. Roth // *IEEE transactions on visualization and computer graphics*. 2013. Т. 19. №. 12. С. 2356-2365.
56. Rossi R.A. Interactive visual graph mining and learning / R.A. Rossi, N.K. Ahmed, R. Zhou, H. Eldardiry // *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2018. Т. 9. №. 5. С. 1-25.

57. Остапенко А.Г. Сете-информационная эпидемиология / А.Г. Остапенко, Е.Б. Белов, А.О. Калашников [и др.]; [Под ред. чл.-корр. РАН Д.А. Новикова.] Серия Теория сетевых войн; Вып. 6.: М: Горячая линия – Телеком, 2021. 216 с.
58. Veerabhadrapa R. Compatibility evaluation of clustering algorithms for contemporary extracellular neural spike sorting / R. Veerabhadrapa, M. Ul. Hassan, J. Zhang и др. // *Frontiers in systems neuroscience*. 2020. Т. 14. 34 с.
59. Lu J. Clustering-based force-directed algorithms for 3D graph visualization / J Lu, Y. W. Si // *The Journal of Supercomputing*. 2020. Т. 76. №. 12. С. 9654-9715.
60. Карпов Д.В. Теория графов //СПб.: Санкт-Петербургское отделение Мат. института им. ВА Стеклова РАН. 2017.
61. Остапенко А.Г. Атакуемые взвешенные сети / А.Г. Остапенко, В.Г. Плотников, А.О. Калашников [и др.]; [Под ред. чл.-корр. РАН Д.А. Новикова.] Серия Теория сетевых войн; Вып. 2. М: Горячая линия – Телеком, 2017. 248 с.

Воронежский государственный технический университет
Voronezh State Technical University

Поступила в редакцию 12.09.2021

Информация об авторе

Сердечный Алексей Леонидович – канд. техн. наук, старший преподаватель, Воронежский государственный технический университет, e-mail: alex-voronezh@mail.ru

SYSTEM OF PROTECTED CYBERSPACE MAPPING METHODS

A.L. Serdechnyy

This paper structures a methodology for the construction and analysis of information maps. A detailed description of the system of methods required for the construction and analysis of information maps is offered. The main groups of methods considered are: data preparation methods for information map (source-based data modeling, web scraping, etc.), information map visualization methods (graph layout methods, heat map, etc.), information map updating methods (graph matching method, landscape transformation method), traditional cartographic methods (graphic overlay, map area markup, etc.), network methods (graph ranking, graph clustering etc.) The proposed system of methods is sufficient to conduct comprehensive mapping studies of the protected cyberspace, as well as other network structures modeling the subject area. It is also possible to extend this system with additional methods aimed at increasing the efficiency of information map construction and analysis processes.

Keywords: method, cyberspace, information map, graph.

Submitted 12.09.2021

Information about the author

Alexey L. Serdechnyy – Cand. Sc. (Technical), Senior Lecturer, Voronezh State Technical University, e-mail: alex-voronezh@mail.ru