

# Методы и средства визуализации сетей соавторства и сетей цитирования больших научных порталов

*Апанович З.В., Винокуров П.С.*

## Аннотация

Благодаря быстрому развитию направления Semantic Web и его новой ветви Linked Open Data, в Интернете становятся доступными большие объемы структурированной информации, размещенной на научных порталах, посвященных различным научным направлениям. Наиболее достоверным источником информации, посвященной любому научному направлению, являются собственно научные публикации, составляющие основное наполнение таких порталов. Эти данные нуждаются в средствах анализа, которые могли бы способствовать упрощению их понимания и оптимизации научного менеджмента. В данной работе описываются новые алгоритмы визуализации графов, реализованные в ИСИ СО РАН, и демонстрируется применение этих алгоритмов для визуализации сетей соавторства и сетей цитирования, извлеченных из научных порталов, входящих в облако Linked Open Data.

**Ключевые слова:** онтология, информационное наполнение, методы визуализации информации, силовой алгоритм, радиальный алгоритм, иерархические жгуты ребер, поуровневое изображение ориентированного графа, сети цитирования, Open Linked Data.

## Введение

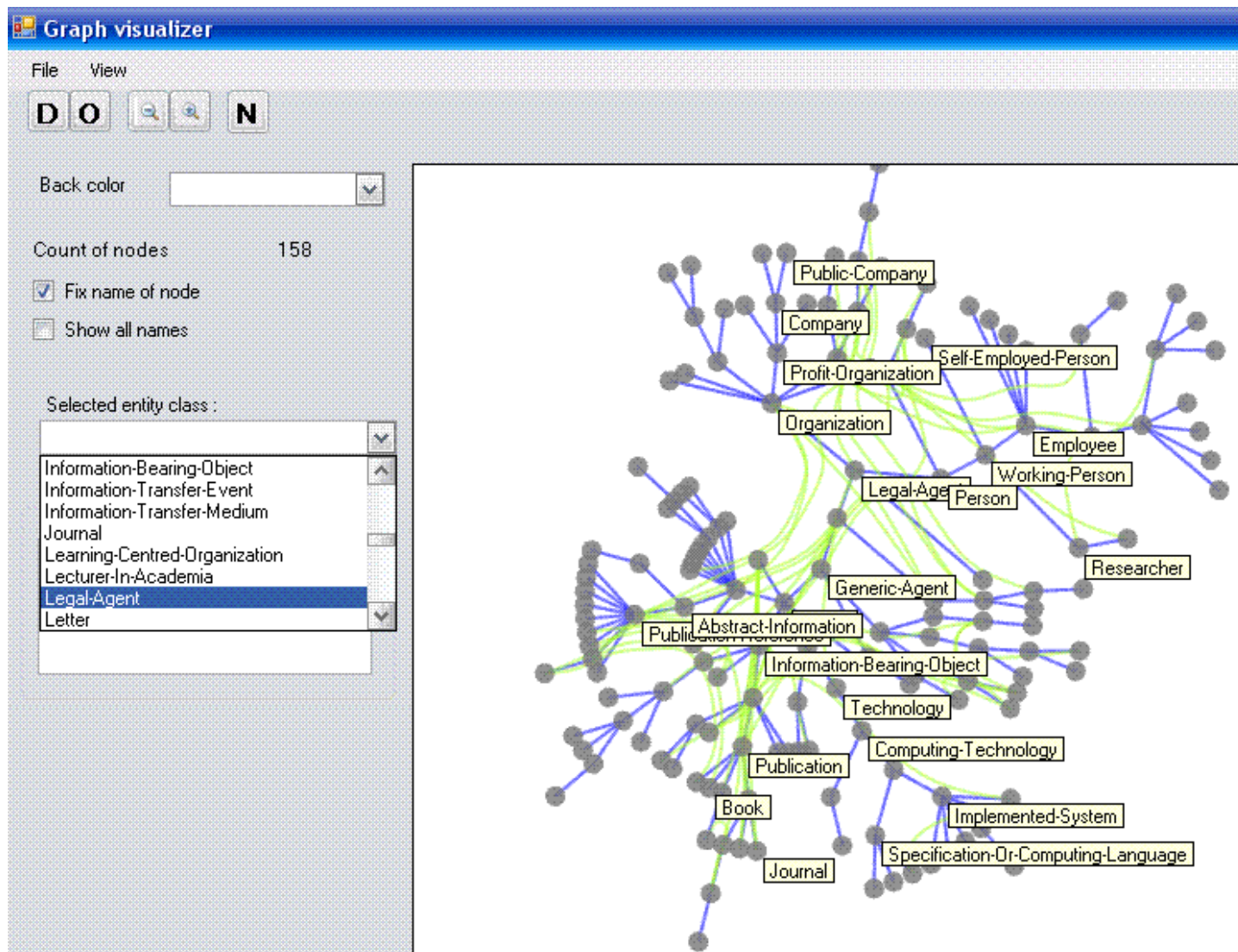
В связи с бурно развивающимся направлением Semantic Web и его новой ветвью Linked Open Data в Интернете становятся доступными большие объемы информации, посвященной различным научным направлениям. В число таких ресурсов входят информационные системы, цифровые библиотеки и специализированные порталы. Наиболее достоверным источником информации, посвященной любому научному направлению, являются собственно научные публикации, составляющие основное наполнение научных порталов и цифровых библиотек. Самые активные и влиятельные исследователи, организации, в которых они работают, и места, в которых расположены научные организации, – вся эта информация становится доступной в формате rdf/xml. Важно также отметить, что эта информация эволюционирует во времени и стремительно увеличивается в объеме. Исследование и анализ этих данных необходимы для оптимизации процессов управления научными исследованиями. Для обеспечения понимания этих стремительно расширяющихся данных нужны новые инструменты.

Одним из таких общепризнанных инструментов является визуализация информации с применением графовых моделей. Следует заметить, что осмысленные множества данных имеют разную структуру и требуют существенно различных стратегий при визуализации. Ранее нами были рассмотрены методы визуализации информации о научном сотрудничестве, представимой при помощи сетей соавторства, извлекаемых из небольших русскоязычных информационных порталов, посвященных таким научным направлениям как археология и компьютерная лингвистика [1, 2]. Но эти данные имели достаточно локальный характер и обладали сравнительно небольшим объемом. Для того, чтобы опробовать наши алгоритмы визуализации на общеизвестных данных большего объема, мы воспользовались общеизвестными данными порталов, входящих в облако Open Linked Data[3]. В процессе экспериментов с этими данными были реализованы новые алгоритмы визуализации, описанные ниже.

## 1. Построение сетей соавторства и сетей цитирования на основе Linked Open Data

Прежде чем решать проблему анализа библиографических данных, необходимо решить проблему их получения. Задача сбора данных весьма трудоемка и ресурсозатратна. В настоящее время функционирует большое количество электронных библиотек и информационных порталов. Эти библиотеки и порталы имеют различную структуру, основаны на разных онтологиях, что затрудняет доступ к ним. Последнее время наметились большие сдвиги в унификации доступа к библиографической информации, благодаря новому проекту сообщества Semantic Web, который называется Linked Open Data (LOD) [3]. В рамках этого проекта проделана большая работа по объединению разрозненных наборов данных в единое целое и обеспечению единого механизма доступа. В частности, в рамках этого проекта предоставляется доступ к большому объему структурированной библиографической информации. Множества структурированных данных, посвященных научным исследованиям, предоставлены такими известными порталами как DBLP, Citeseer, CORDIS, NSF, EPSRC, ACM, IEEE и др. Данные предоставляются в формате RDF и имеют весьма внушительные объемы. Например, RDF-данные, предоставленные порталом CiteSeer, содержат 8 146 852 троек RDF[12], данные портала ACM насчитывают 12,402,336 троек RDF[14], портал DBLP предоставил 28 384 790 троек RDF[13]. Пользователь может либо скачивать файлы в формате RDF, либо генерировать данные при помощи запросов sparql.

Важно также отметить, что за последнее время LOD-сообществом проделана огромная работа по переводу всех этих множеств данных на единую онтологию AKT Reference Ontology [9], представляющую собой объединение нескольких онтологий, таких как Support Ontology, Portal Ontology, Extensions Ontology и RDF Compatibility Ontology. Онтология Portal Ontology является основной среди этих онтологий, она описывает такие понятия как организации, персоны, проекты, публикации, географические данные и т. д. Онтология AKT представляет собой весьма глубокую иерархическую структуру, это видно на рис. 1.



**Рис. 1.** Основные классы онтологии АКТ.

На этом рисунке показано также одно из окон пользовательского интерфейса программы визуализации. Меню «File» позволяет выбрать файл для визуализации и указать формат входных данных, меню «View» - выбрать один из алгоритмов визуализации, имеющихся в системе. В данном случае для визуализации используется радиальный алгоритм. Также указано общее количество вершин изображаемой структуры. В данном случае в иерархии имеется 158 вершин.

Следует заметить, что размеры исследуемой структуры могут быть весьма значительными, и, хотя в системе визуализации имеется режим подписи всех вершин («Show all names»), полученное изображение будет слишком загроможденным. Поэтому нами реализовано много возможностей взаимодействия с изображением, позволяющих пользователю пошагово изучать исследуемую структуру. Хорошо себя зарекомендовал режим «Fix name of node», в котором все надписи являются всплывающими, но при выборе мышью интересующей пользователя вершины ее название фиксируется на экране. Этот режим позволяет пользователю в интерактивном режиме проставить надписи только интересующих его вершин. Также на панели «Selected entity class» пользователю выдается список всех вершин изображаемой структуры и при выборе нужной вершины на экране высвечиваются связи этой вершины. В связи с большим объемом изображаемых далее структур, авторы не имеют возможности вывести названия всех их вершин (изображение было бы черным от надписей), но пользователь системы имеет возможность в интерактивном режиме увидеть названия всех интересующих его вершин.

Что касается АКТ-онтологии, для описания публикаций имеется два корневых класса «Information-Bearing-Object» и «Abstract-Information». Подклассами класса «Information-Bearing-Object» являются также классы «Recorded-Audio», «Recorded-Video», «Publication», «Edited-Book», «Composite-Publication», «Serial-Publication», «Periodical-Publication», «Book». Все элементы этого класса имеют отношение «has-publication-reference», указывающее на объекты класса «Publication-Reference», который является подклассом класса «Abstract-Information». В свою очередь класс «Publication-Reference» имеет в качестве подклассов классы «Web-Reference», «Book-Reference», «Edited-Book-Reference», «Conference-Proceedings-Reference», «Workshop-Proceedings-Reference», «Book-Section-Reference», «Article-Reference», «Proceedings-Paper-Reference», «Thesis-Reference» и «Technical-Report-Reference». Эти объекты имеют такие отношения как: «has-date», «has-title», «has-place-of-publication», «cites-publication-reference» и др. Для описания организаций имеется класс «Organization», который является подклассом класса «Legal-Agent», а класс «Legal-Agent» является подклассом класса «Generic-Agent». Точно так же класс «Person» является подклассом класса «Generic-Agent».

Несмотря на то, что все хранилища библиографических данных облака LOD приведены к единому словарю, данные, доступные в этих хранилищах, очень разнородны и опираются на очень узкие подмножества большого словаря. Для описания реальных объектов используются, как правило, классы самого верхнего уровня иерархии.

Так, например, для описания публикаций самыми ходовыми классами являются «Publication-Reference» и «Article-Reference», при этом совсем не используются такие классы, как «Proceedings-Paper-Reference», что затрудняет извлечение из базы данных публикаций по одной заданной теме. Также многие поля, имеющиеся в этой богатой онтологии, остаются незаполненными при описании реальных данных. Тем не менее, единый механизм доступа открывает большие возможности для работы с этими данными. Достаточно просто

извлечь из любого репозитория облака LOD данные для построения сетей соавторства. Любая публикация, описанная в этих репозиториях, имеет название публикации (отношение "has-title") и авторов (отношение "has-author"). Поэтому сеть соавторства для любого из перечисленных выше порталов можно сгенерировать с помощью sparql-запроса следующего вида:

```
CONSTRUCT{?y :co_author ?z}

WHERE{

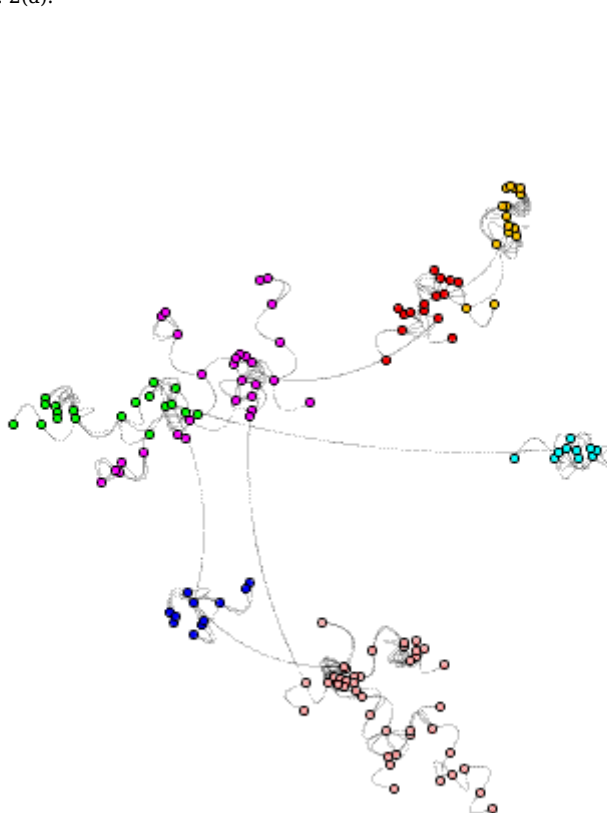
    ?x akt:has-author ?y ;

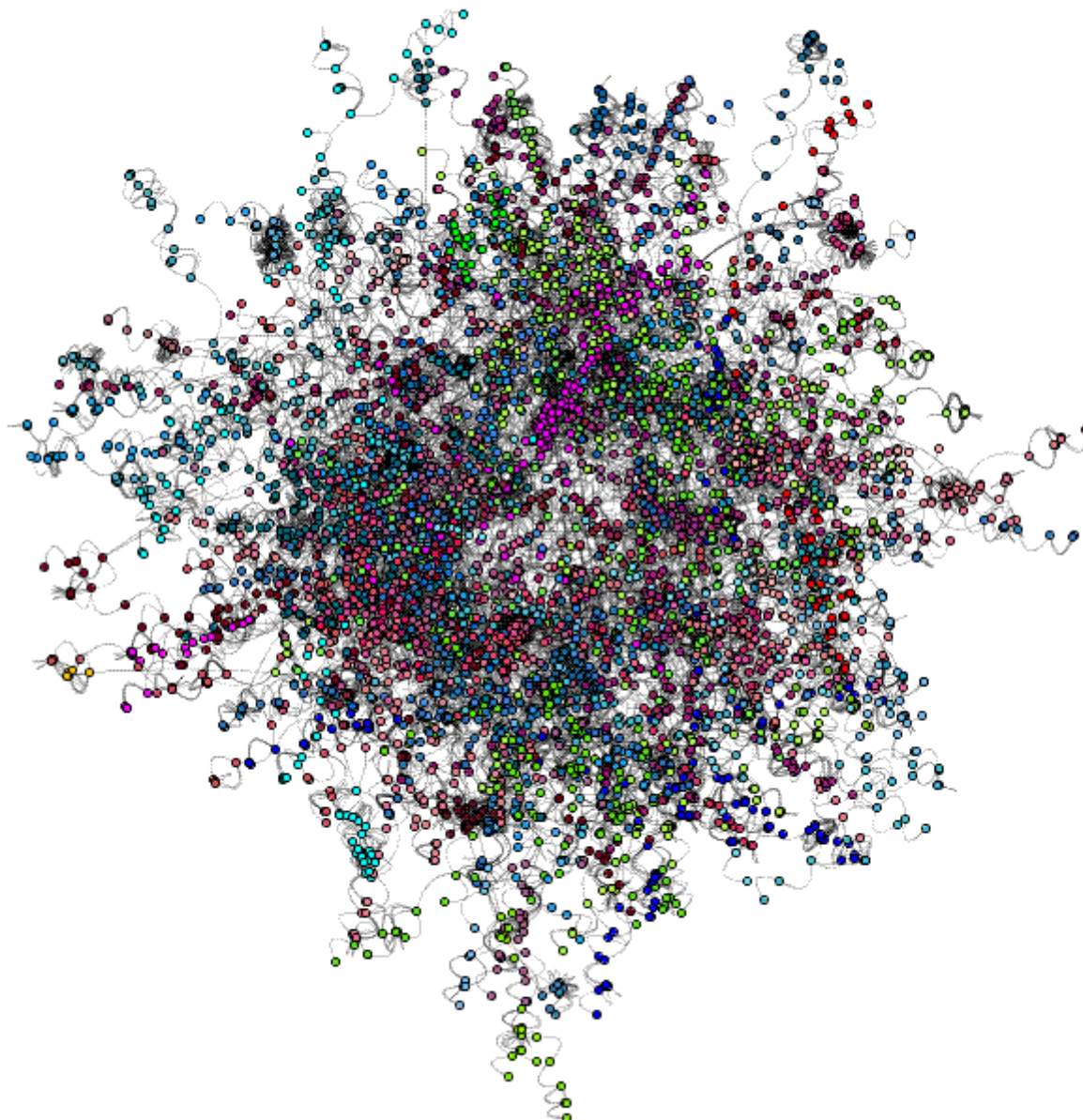
    akt:has-author ?z ;

    a ?type .

    FILTER(?y != ?z &&(?type = akt :Publication-Reference ) ).
```

Для выбора данных нужного объема используется модификатор запроса LIMIT N. В настоящее время мы сравнительно легко извлекаем сети соавторства объемом 20-30 тысяч вершин. Следует сказать, что при таком способе генерации сетей соавторства их связность и плотность напрямую связаны с объемом. Например, для портала DBLP при установке лимита на количество ребер в сети соавторства, равном десяти тысячам, наибольшая связная компонента этой сети имеет всего 140 вершин и 191 ребро. Изображение этой небольшой компоненты связности показано на рис. 2(a).





**Рис. 2.** Изображение связанных компонент сетей соавторства, сгенерированных по данным портала DBLP.

При возрастании лимита на объем сети до 50000 ребер наибольшая связанная компонента имеет уже 3001 вершину и 4983 ребра. Для анализа таких компонент связности необходимы специальные алгоритмы.

## 2. Кластеризация и визуализация больших сетей соавторства

В предыдущих работах [1, 2] нами был представлен алгоритм выделения сообществ (кластеризации) из сетей соавторства на основе принципа модулярности [8]. На рис. 2(а) показаны семь сообществ, выделенных прежним алгоритмом. Вершины, принадлежащие одному кластеру, раскрашены в один цвет, что позволяет визуально оценить качество кластеризации. Рисунок 2(б) показывает размещение большой компоненты связности сети соавторства, имеющей 3001 вершину и 4983 ребра, до и после работы старого алгоритма кластеризации. Визуально на нем не так просто выделить сообщества. В настоящий момент реализована многоуровневая версия алгоритма выделения сообществ, которая повышает качество кластеризации. Для описания алгоритма напомним определение модулярности:

Определим симметричную матрицу  $e$  размерности  $k \times k$ . Элемент  $e_{ij}$  этой матрицы равен отношению количества ребер, соединяющих два сообщества  $i$  и  $j$ , к общему количеству ребер в сети. Также можно определить суммы по столбцам (или по строкам)  $a_i = \sum_j e_{ij}$ , которые соответствуют отношению ребер, соединяющих вершины в сообществе  $i$  к общему количеству ребер. Модулярность (modularity) выражается через  $a_i$  и  $e_{ij}$ :

$$Q = \sum_i (e_{ii} - a_i^2)$$

Экспериментально показано [8], что значение модулярности, превышающее 0,3, является указателем на реальное наличие сообществ в сети.

Новая реализация алгоритма выделения сообществ состоит из грубой кластеризации и итеративного улучшения. На этапе грубой

кластеризации первоначально сеть соавторства разбивается на кластеры, каждый из которых состоит из одной вершины. Затем кластеры, дающие наилучшее увеличение модулярности, попарно объединяются в кластеры большего размера до тех пор, пока еще возможно увеличение значения модулярности. Промежуточные результаты попарного объединения кластеров хранятся в виде бинарного дерева. Заметим, что получившийся в результате первого шага набор кластеров не является оптимальным, вследствие того, что на начальных этапах работы алгоритма возможно объединение вершин из разных сильно связанных сообществ. Поэтому на втором этапе применяется алгоритм итеративного улучшения, идея которого заимствована у [7].

Определим величину  $\Delta_{Q_{v \rightarrow D}}$  как число, на которое увеличится модулярность, если переместить вершину  $v$  из ее текущего кластера в кластер  $D$ . Тогда алгоритм **Итеративного Улучшения кластеризации** работает следующим образом:

Шаг 1: Для каждой вершины  $v$  ищется кластер с максимальным значением  $\Delta_{Q_{v \rightarrow D}}$ . Если  $\Delta_{Q_{v \rightarrow D}} > 0$ , вершина  $v$  перемещается в кластер  $D$ .

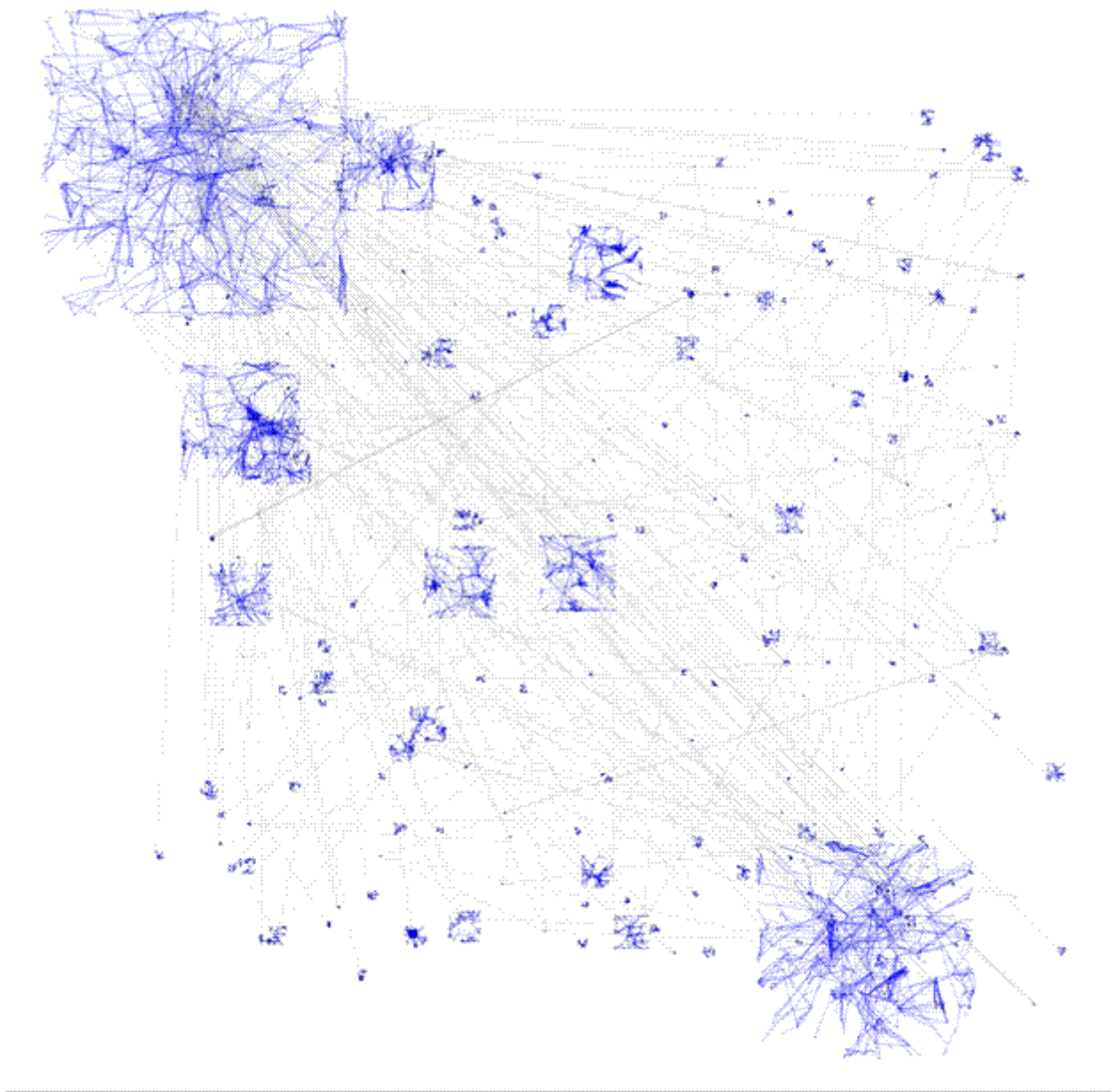
Шаг 2: Шаг 1 повторяется до тех пор, пока существует хотя бы одно перемещение, улучшающее значение модулярности.

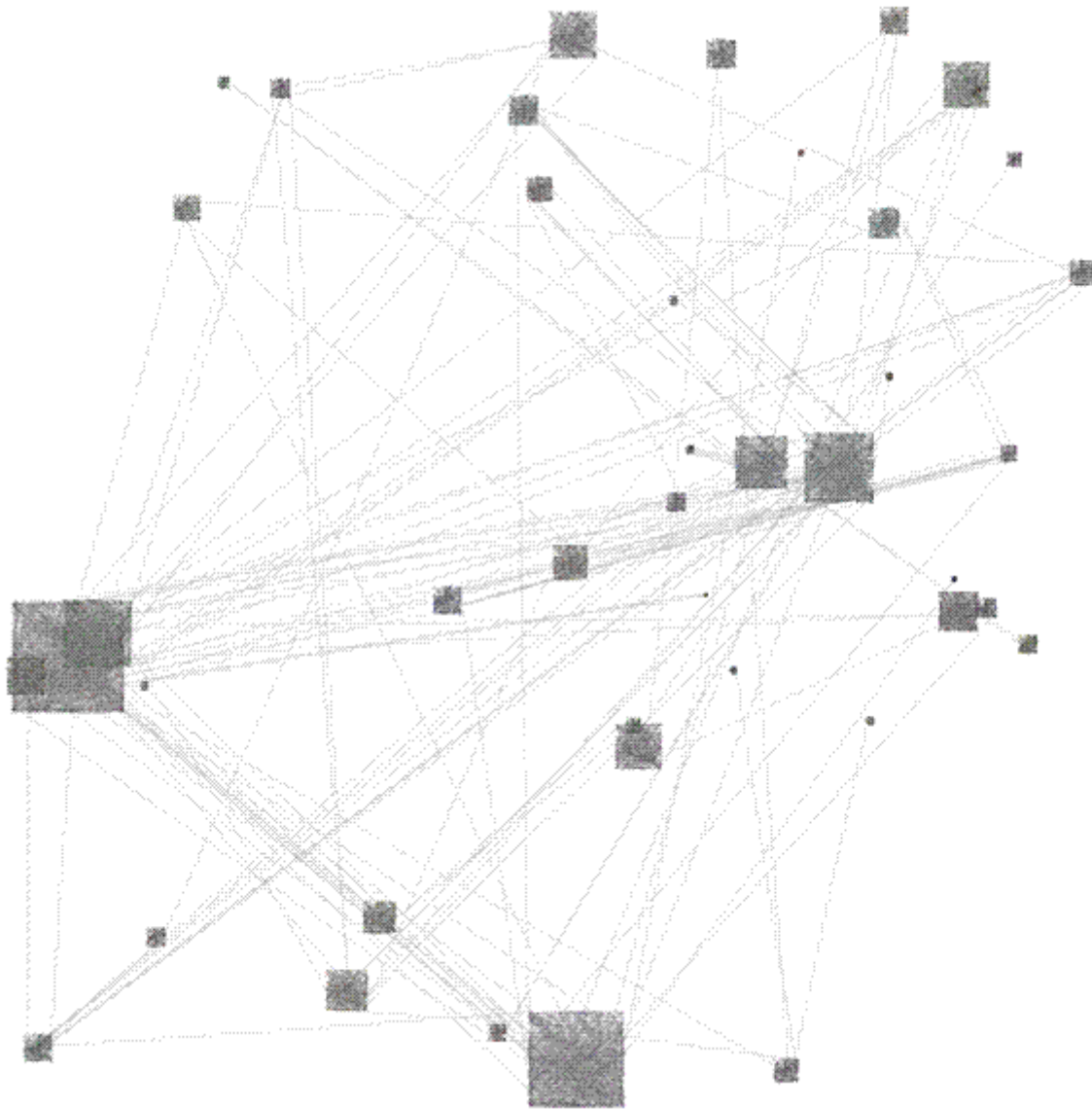
Этот алгоритм перемещает по одной вершине и не может переместить сразу группу сильно связанных вершин. Поэтому лучше всего этот алгоритм применять для улучшения промежуточных результатов грубой кластеризации. Для этого в бинарном дереве объединения кластеров выделяются уровни, между которыми количество кластеров сократилось в два раза. Для каждого такого уровня имеется набор текущих кластеров, а в качестве перемещаемых вершин используются кластеры, полученные на предыдущем уровне грубой кластеризации. Для каждого такого уровня запускается алгоритм **Итеративного Улучшения**, что позволяет еще немного улучшить значение модулярности, а вместе с ней и качество кластеризации.

После выделения научных сообществ необходимо построить изображение сети соавторства с найденными научными сообществами. Мы хотим построить такое изображение, чтобы в нем легко просматривались найденные сообщества, а также и связи между этими сообществами. Для построения такого изображения используется трехуровневый алгоритм размещения. Сначала осуществляется глобальное размещение графа, вершинами которого являются найденные компоненты. На этом этапе используется стандартный силовой алгоритм [5]. В процессе размещения считается, что идеальная длина ребра, соединяющего компоненты  $i$  и  $j$ , пропорциональна величине  $e_{ij}$ , количеству ребер между ними.

Детальное изображение каждого сообщества строится тоже при помощи силового алгоритма. Но на этом этапе все вершины одной группы располагаются примерно на одинаковом расстоянии друг от друга. Это идеальное расстояние существенно меньше того, что используется при глобальном размещении. Оно обратно пропорционально количеству вершин в каждом кластере. Наконец, детальное изображение каждой компоненты подставляется в глобальное размещение компонент и заново отрисовываются все межкомпонентные ребра.







**Рис. 3.** Пример разбиения на сообщества сети соавторства, имеющей 5625 вершин и 10103 ребра

На рис. 3(а) показан пример изображения сети соавторства, полученной прежним алгоритмом кластеризации (количество вершин 5625, количество ребер 10103, модулярность 0,922, 197 сообществ. На рис. 3(б) показано разбиение на сообщества той же самой сети многоуровневым алгоритмом (48 сообществ, модулярность 0,948). Заметим, что данные рисунки сильно уменьшены для того чтобы показать всю сеть соавторства, и поэтому названия вершин не проставлены на изображении, но в режиме интерактивной визуализации пользователь имеет возможность получить всю информацию о каждой вершине, входящей в тот или иной кластер.

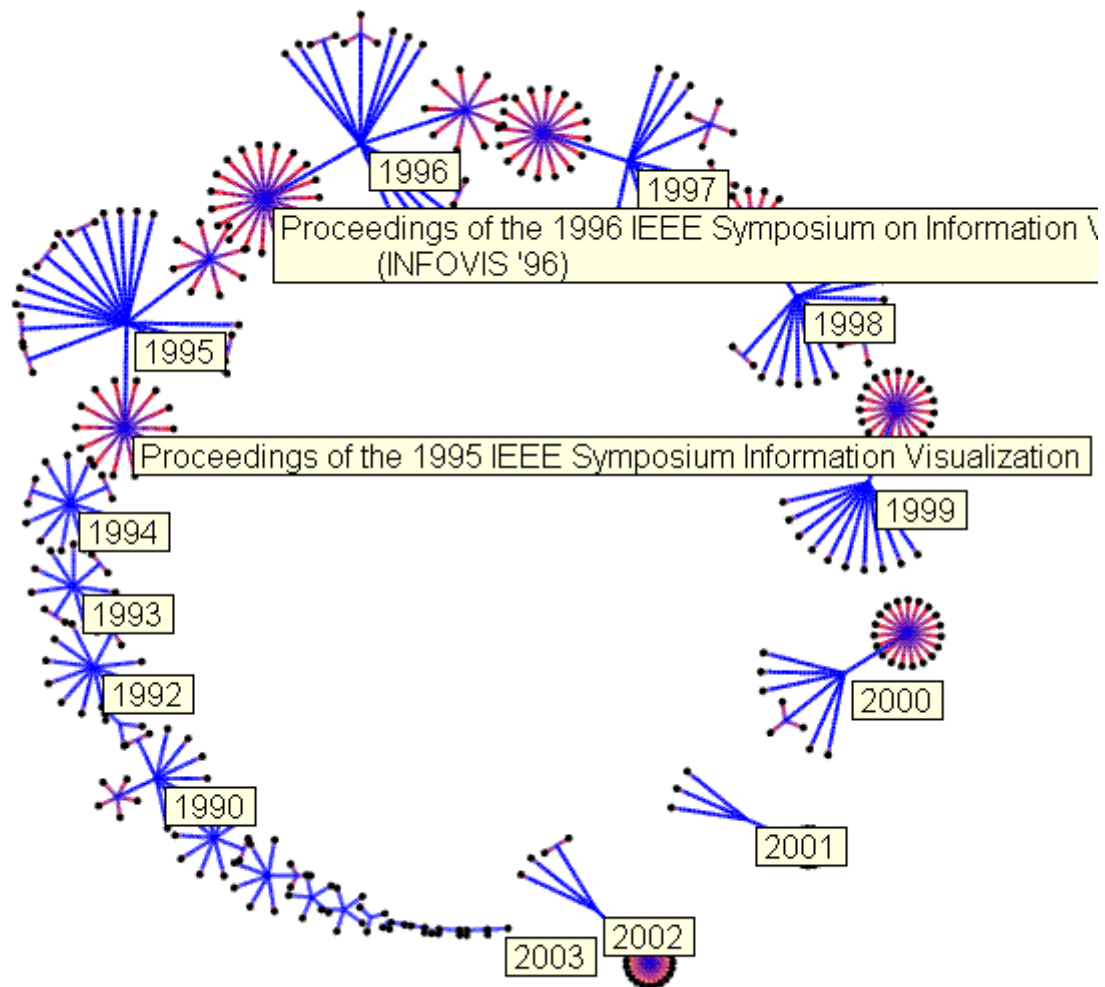
### 3. Методы визуализации сетей цитирования

Если для любого портала облака LOD не составляет большого труда сгенерировать сеть соавторства любого заданного объема, ситуация с сетями цитирования обстоит существенно сложнее. Во-первых, построение списков цитируемой литературы требует гораздо больше технических усилий, поэтому в открытом доступе эта информация предоставляется только небольшим количеством порталов. Среди порталов облака LOD такими порталами являются CiteSeer и ACM. Во-вторых, следует отметить, что методы, применяемые при визуализации сетей соавторства, оказались мало пригодными в случае сетей цитирования.

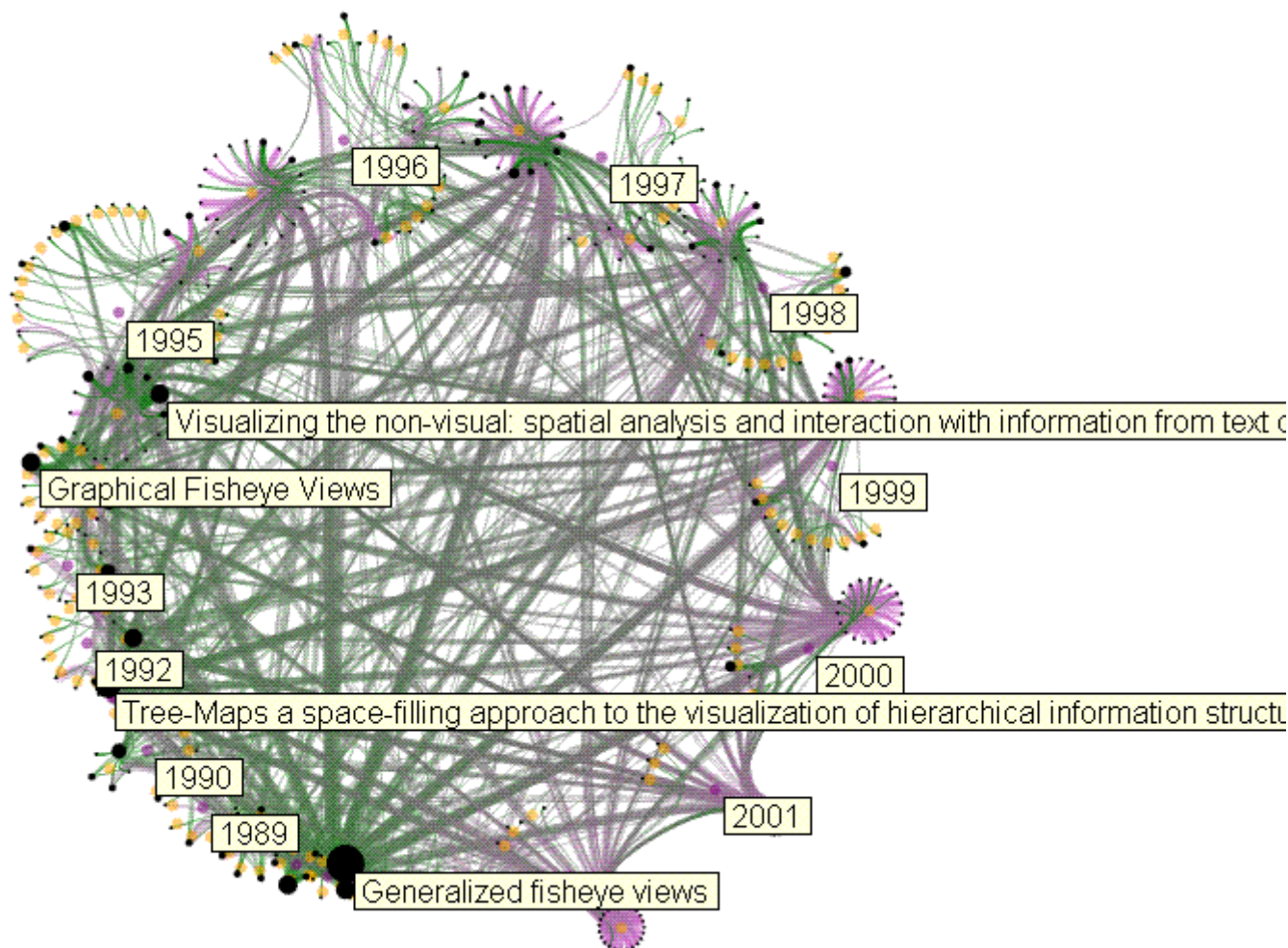
Мы начали наши эксперименты по визуализации сетей цитирования с попыток напрямую использовать реализованный ранее метод иерархических жгутов ребер, который хорошо себя зарекомендовал при визуализации сетей соавторства. Этот метод позволяет совместное изображение сети и онтологических отношений информационного портала, что позволяет визуально анализировать различные характеристики рассматриваемой сети. В частности, в рассматриваемых ранее примерах этот метод позволял анализировать распределение вершин сети по географическим местностям, разделам науки и т.д.

Для экспериментов была сгенерирована небольшая сеть цитирования, имеющая в качестве вершин 561 публикации по визуализации информации за 10 лет. На Рис. 4(а) показана трехуровневая иерархия, состоящая из годов публикации, научных конференций и собственно публикаций, изображенная при помощи кругового алгоритма. (Годы включают конференции, конференции включают публикации). Ребра этой иерархии раскрашены голубым цветом. Публикации изображены кружочками черного цвета, конференции - кружочками желтого цвета, а годы публикации - кружочками фиолетового цвета. На эту иерархию накладываются ребра сети цитирования, вершинами которой являются публикации. (Рис. 4 (б)). Поскольку сеть цитирования является ориентированным графом, направление ребер сети цитирования (от цитирующей публикации к цитируемой публикации) изображается интерполяцией цвета от фиолетового к зеленому. Хотя изображение позволяет достаточно легко увидеть, годы, на публикации которых много ссылок, и годы, от публикаций которых есть много ссылок, в целом изображение несет не очень много информации, поскольку трудно идентифицировать наиболее цитируемые публикации. Поэтому, прежде всего, мы модифицировали алгоритм визуализации, позволив размеру вершины меняться в зависимости от индекса цитирования. Поскольку мы не хотели, чтобы из-за увеличения размеров вершин увеличивалась бы

площадь изображения, мы разрешили наложение вершин, при этом введя функцию изменения видимости вершин в зависимости от их индекса цитирования. Аналогично, пользователю предоставлена возможность менять ширину и видимость связей цитирования.







**Рис. 4.** Иерархическая структура и сеть цитирования (а) Трехуровневая иерархия состоит из годов, конференций и публикаций. (б) Изображение сети цитирования методом иерархических жгутов ребер.

Возможные варианты функций могут быть такими.

[1]

$$y = (o_{\max} - o_{\min}) \frac{I - I_{\min}}{I_{\max} - I_{\min}} + o_{\min}$$

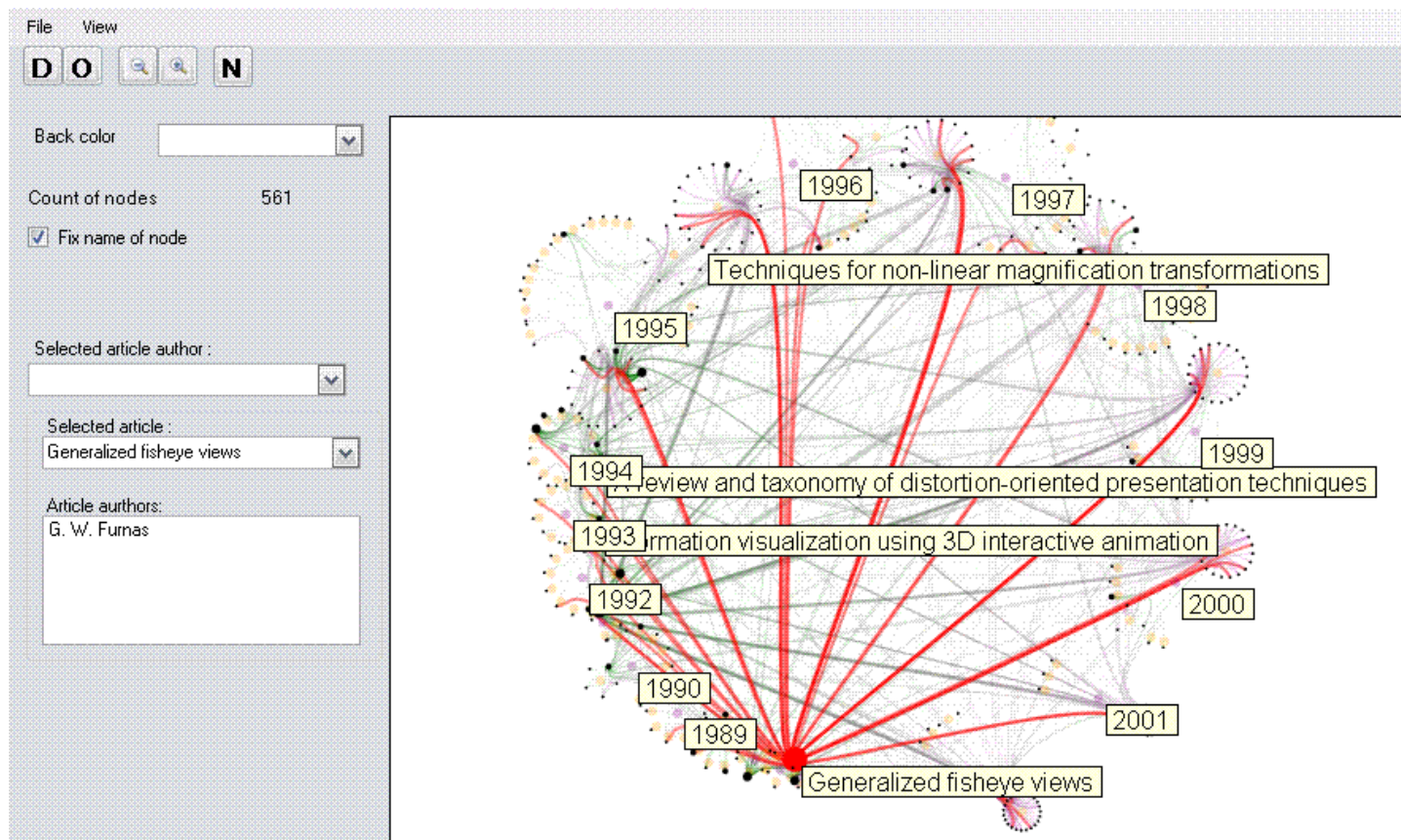
[2]

$$y = (o_{\max} - o_{\min}) \cdot \left( 1 - \sqrt{\frac{I_{\max} - I}{I_{\max} - I_{\min}}} \right)$$

Где  $I$  – индекс цитирования,  $I_{\max}$  and  $I_{\min}$  – наибольший и наименьший индекс цитирования в рассматриваемой сети,  $o_{\max}$  и  $o_{\min}$  – верхняя и нижняя граница значений для  $y$ .

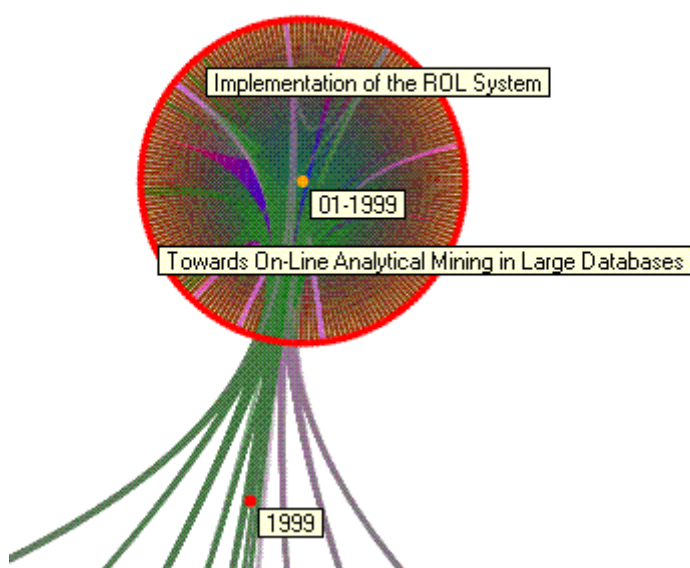
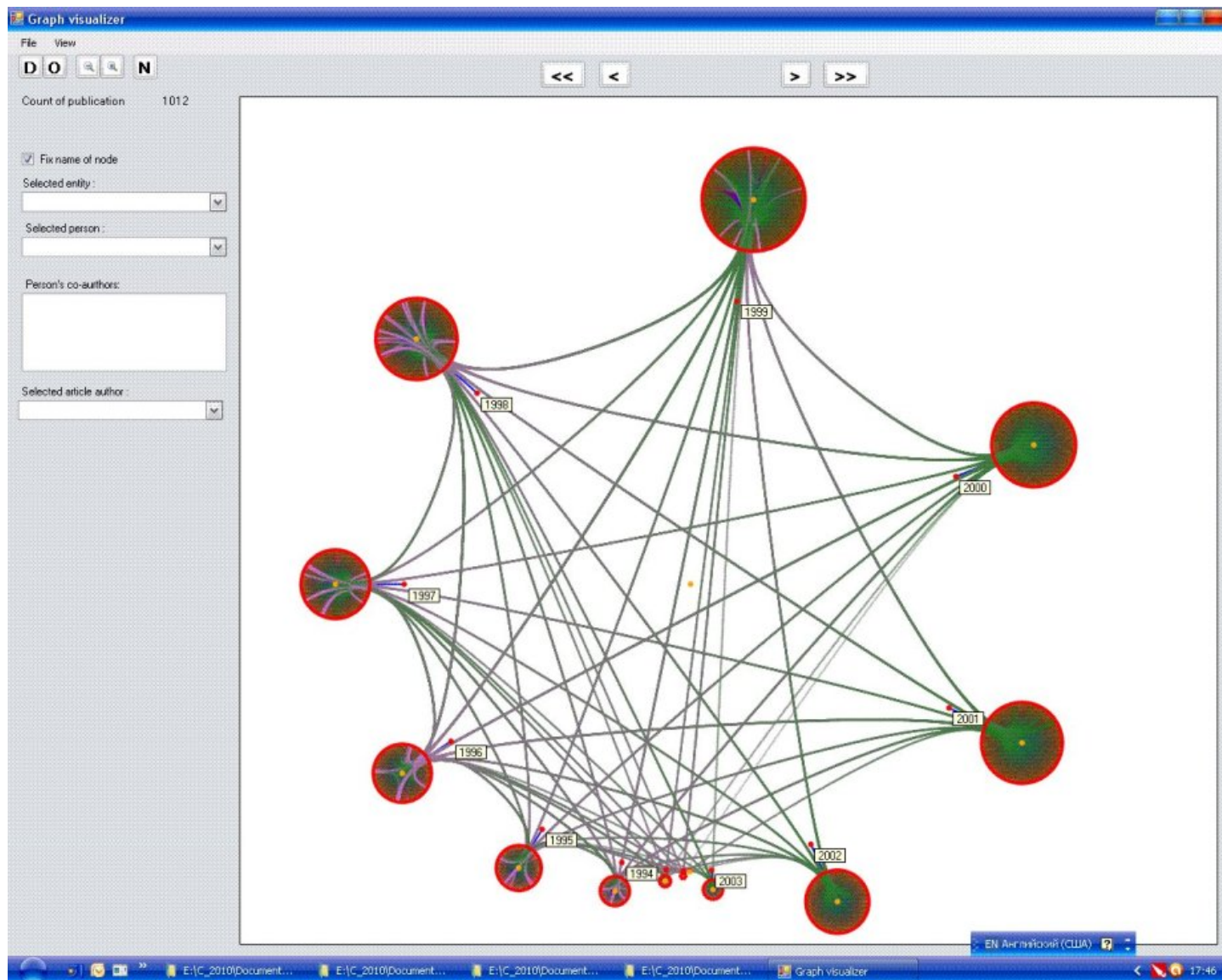
Формула (1) в силу своей линейности позволяет идентифицировать группы наиболее цитируемых публикаций, такие публикации имеют примерно одинаковый размер. Формула (2) позволяет выделить самую цитируемую публикацию, так как размер такой публикации существенно отличается от размера остальных публикаций.

Когда выявлены наиболее цитируемые публикации, пользователь может выбрать любую такую публикацию и просмотреть дополнительную информацию о ней: название, авторов, какие другие публикации на нее ссылаются, как это можно видеть на Рис. 5. Ребра-ссылки на выбранную публикацию со стороны других публикаций показаны красным.



**Рис. 5.** Наиболее цитируемая публикация и ссылки на нее от других публикаций.

При возрастании сети цитирования данный метод визуализации перестает быть полезным. Например, на рисунке 6 показана сеть цитирования, содержащая 20 000 вершин, извлеченная с портала Citeseer и покрывающая временной интервал с 1993 по 2003 год. Поскольку в имеющейся базе данных не оказалось информации о конференциях, мы сгенерировали только двухуровневую иерархию, состоящую из месяца публикации и года публикации. В результате получилось изображение, разреженное в центре и очень плотное на периферии. На изображении легко видеть, что больше всего публикаций в этом тестовом множестве приходится на 1998 and 1999 годы и не очень много публикаций – на 2003. Но более детальную информацию из этого изображения получить трудно. В центре изображения имеется полный граф, соответствующий факту, что от каждого года есть ссылки на публикации более ранних годов. А публикации по отдельным годам настолько многочисленны, что трудно выбрать отдельную вершину для дальнейшего исследования.



**Рис. 6.** Сеть цитирования из 20 000 вершин сгенерированная по данным портала from Citeseer. (а) Глобальный вид, (б) Публикации за один месяц 1998 года.

Вдобавок, изображение хронологически упорядоченной информации по кругу не всегда выглядит естественно. Поэтому для лучшей визуализации хронологической информации был реализован метод динамического поуровневого изображения сетей цитирования.



## 4. Динамическое поуровневое изображение сетей цитирования

Поскольку сеть цитирования является ориентированным графом, желательно изобразить граф так, чтобы направление ребер соответствовало хронологическому порядку появления публикаций. Поэтому за основу визуализации был взят поуровневый метод изображения ориентированного графа[9]. Суть данного метода состоит в том, что вершины-публикации разбиваются на слои, соответствующие различным годам.

На Рис. 7 показана сеть цитирования, изображенная при помощи поуровневого метода. Годы публикаций показаны прямоугольниками разного цвета в верхней части изображения. Все статьи, опубликованные в одном году, располагаются в вертикальных столбцах соответствующих году публикации. Ребра этой сети соответствуют отношению цитирования. Каждое ребро сети цитирования соответствует отношению akt:cites-publication-reference и ориентировано справа налево. Цвет каждого ребра совпадает с цветом года, в котором опубликована цитируемая публикация. Чем больше ссылок имеется на определенную публикацию, тем больше входных ребер имеет соответствующая вершина и тем больше ее радиус. В результате наиболее цитируемые публикации становятся заметны с первого взгляда. Также, в программе реализована возможность отслеживания динамики цитирования по годам. Для этого в верхней части экрана расположены кнопки, позволяющие перемещаться по изображению с заданными интервалами времени. В данный момент размер минимального интервала равен одному году. При нажатии кнопки «>>>» изображается вся имеющаяся сеть цитирования, а при нажатии «<

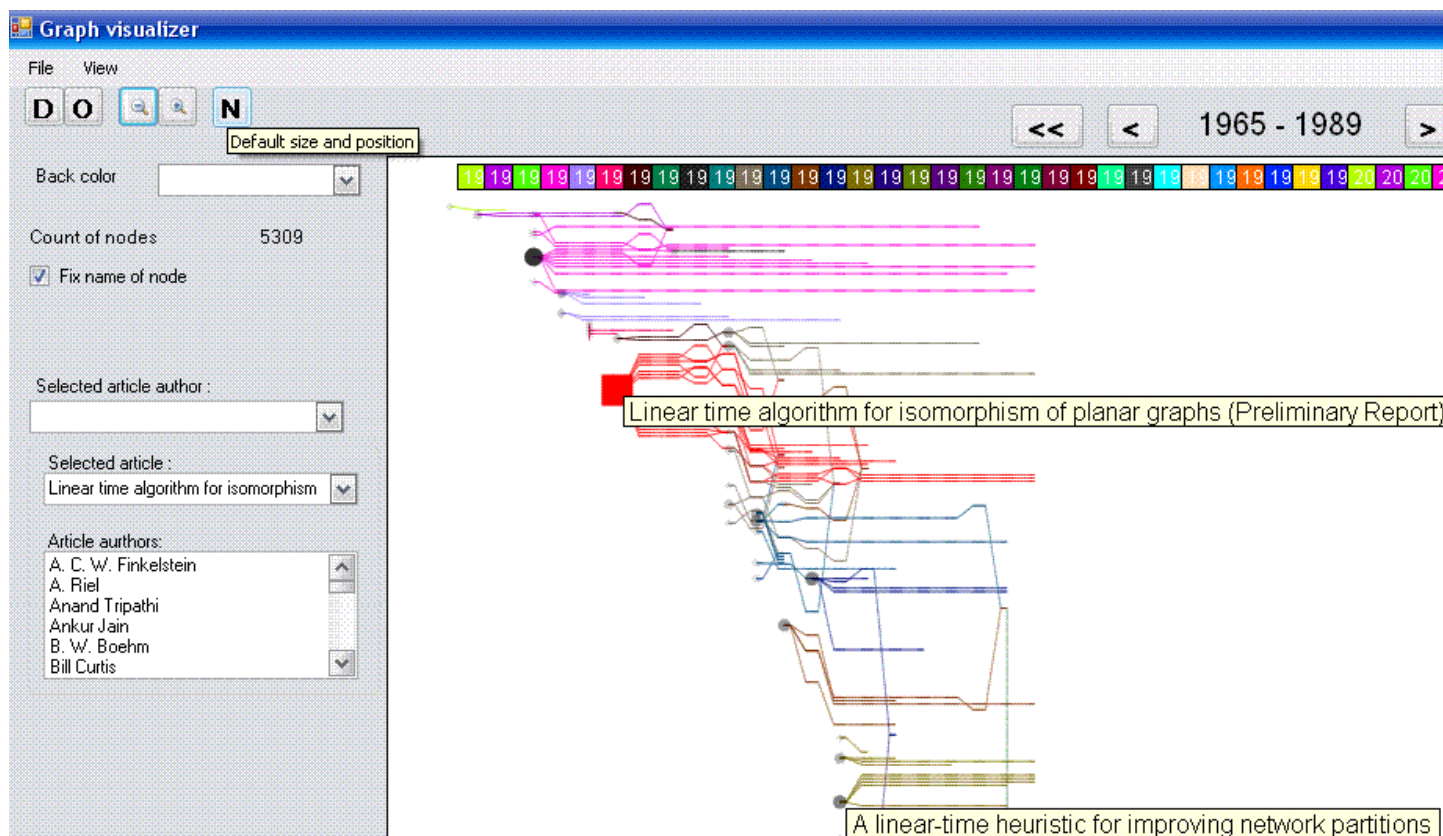
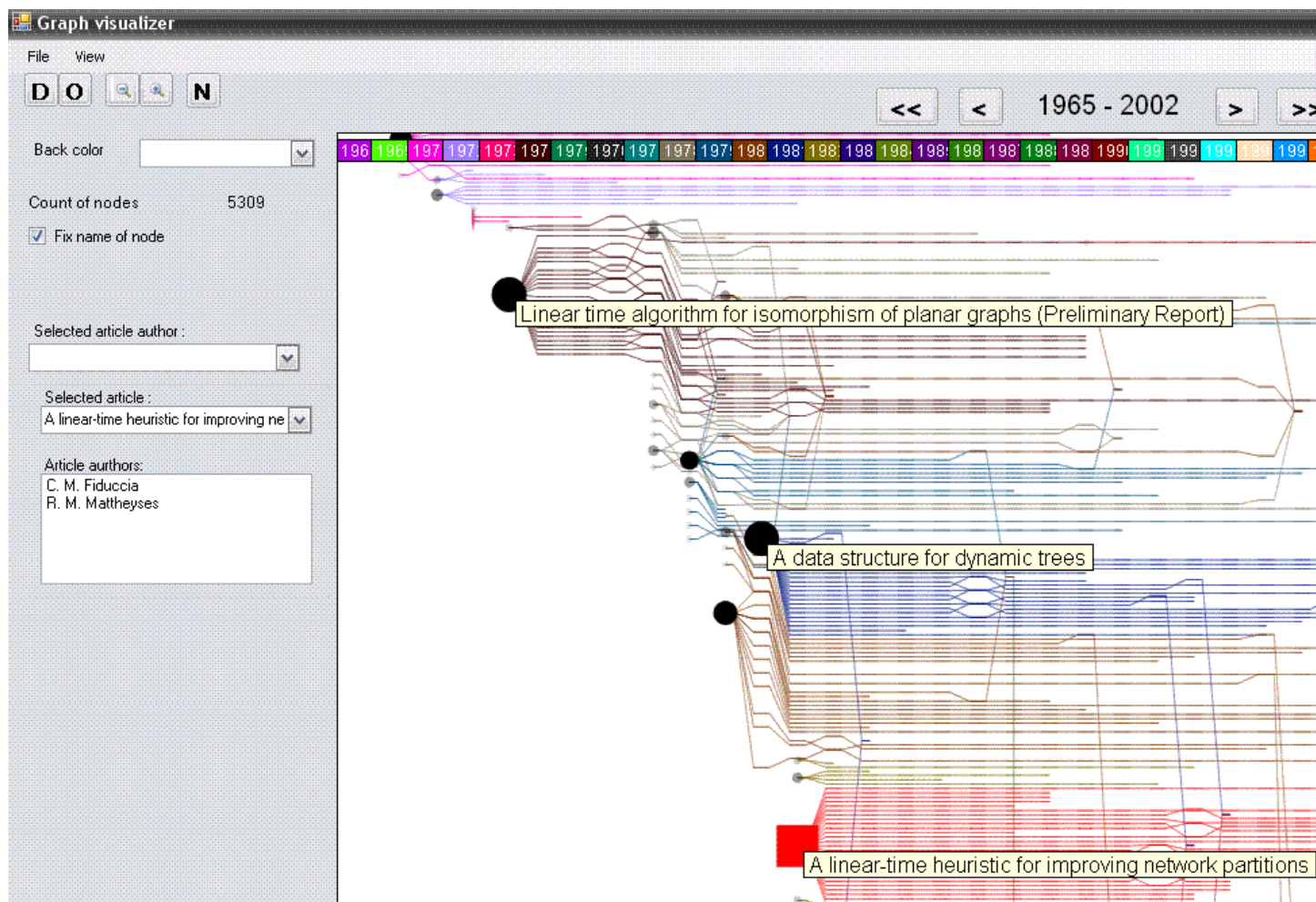


Рис. 7. Изображение сети цитирования поуровневым методом.

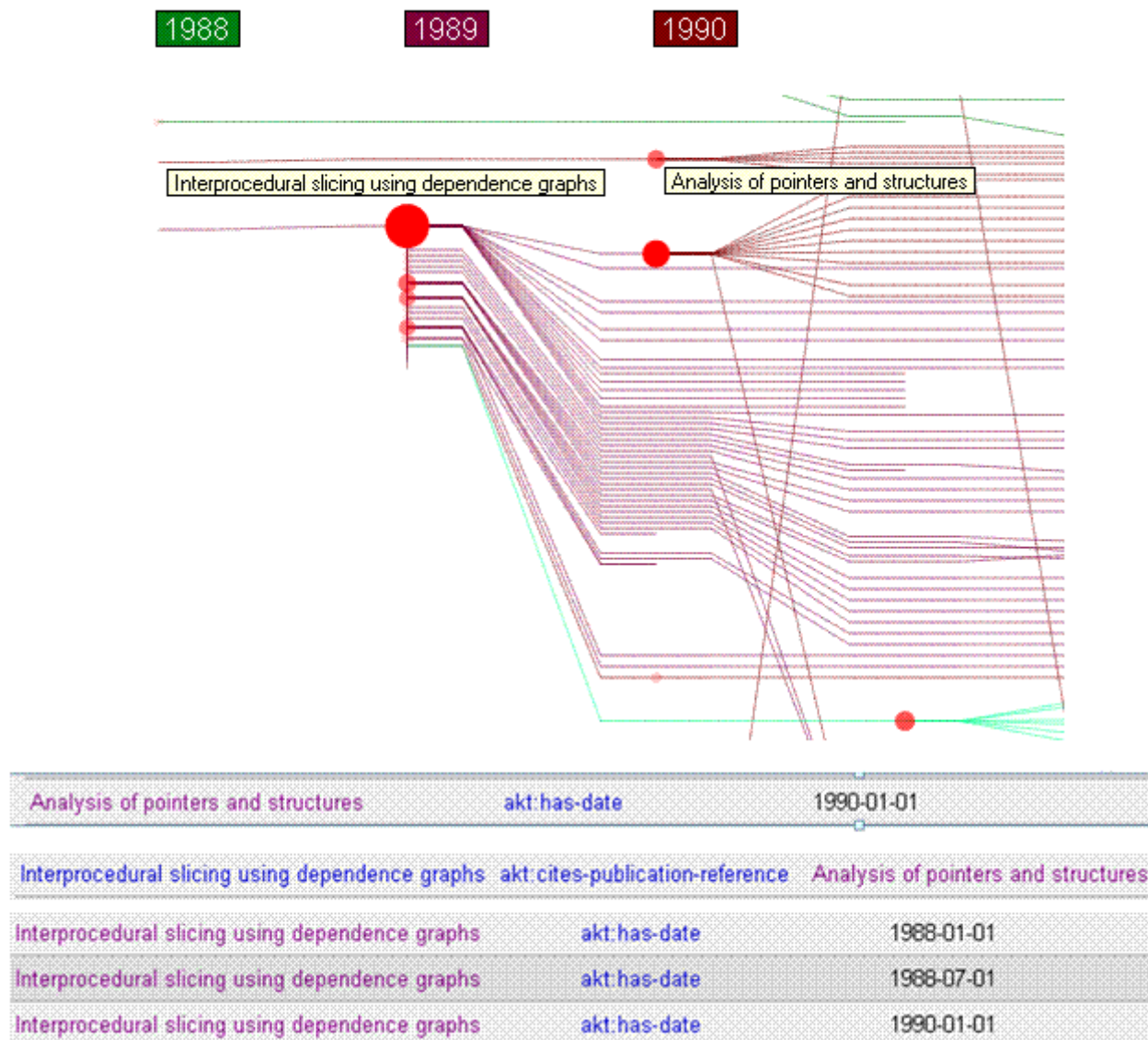
Перемещение по изображению осуществляется при помощи кнопок «<» и «>», позволяя наблюдать изменение сети цитирования во времени. Рисунок 7 показывает фрагмент сети цитирования для публикаций по теории графов в период с 1965 по 1989 и пользовательский интерфейс, позволяющий получать дополнительную информацию. Можно видеть, что в этой выборке среди публикаций по теории графов доминирует «Linear-time algorithm for isomorphism of planar graphs». На это указывает большой радиус и яркость соответствующей вершины. Дополнительную информацию о вершине можно получить благодаря пользовательскому интерфейсу. Весьма мало заметна в этот период публикация «A linear-time heuristic for improving network partition». Радиус этой вершины мал, потому что на нее имеется мало ссылок. Для сравнения на рисунке 8 показано, как выглядит эта же сеть цитирования к 2002 году. Размер вершины, соответствующей публикации «A linear-time heuristic for improving network partition» существенно возрастает, как и размер шлейфа публикаций, ссылающихся на нее.



**Рис. 8.** Изменение сети цитирования к 2002 году.

Помимо всего прочего, такой способ визуализации позволяет обнаруживать ошибки и неточности в библиографических данных. На Рис. 9(a) показан фрагмент сети цитирования, сгенерированной по базе данных ACM для временного интервала с 1988 по 1990 годы. Бордовое ребро соединяет вершину «Analysis of pointers and structures» paper опубликованную в 1990 году и вершину «Interprocedural slicing using dependence graphs» опубликованную в 1988 году. Поскольку цвет ребра совпадает с цветом прямоугольника 1990 года это должно означать, что статья, опубликованная в 1988 году, ссылается на статью, опубликованную в 1990 году. Просмотрев реальные данные, мы обнаружили, что статья «Interprocedural slicing using dependence graphs» имеет несколько дат публикации, из-за чего и появилось подобное изображение.





**Рис. 9.** Ориентация ребер и даты публикаций

Следует также отметить, что данный алгоритм изображения сетей цитирования находится в процессе развития. В настоящий момент реализуется модификация этого алгоритма, позволяющая осуществлять кластеризацию параллельно идущих ребер для их объединения в жгуты, как это показано на рис. 10. Новый алгоритм визуализации существенно уменьшает загруженность изображения и, предположительно, должен лучше демонстрировать тенденции развития научного направления.

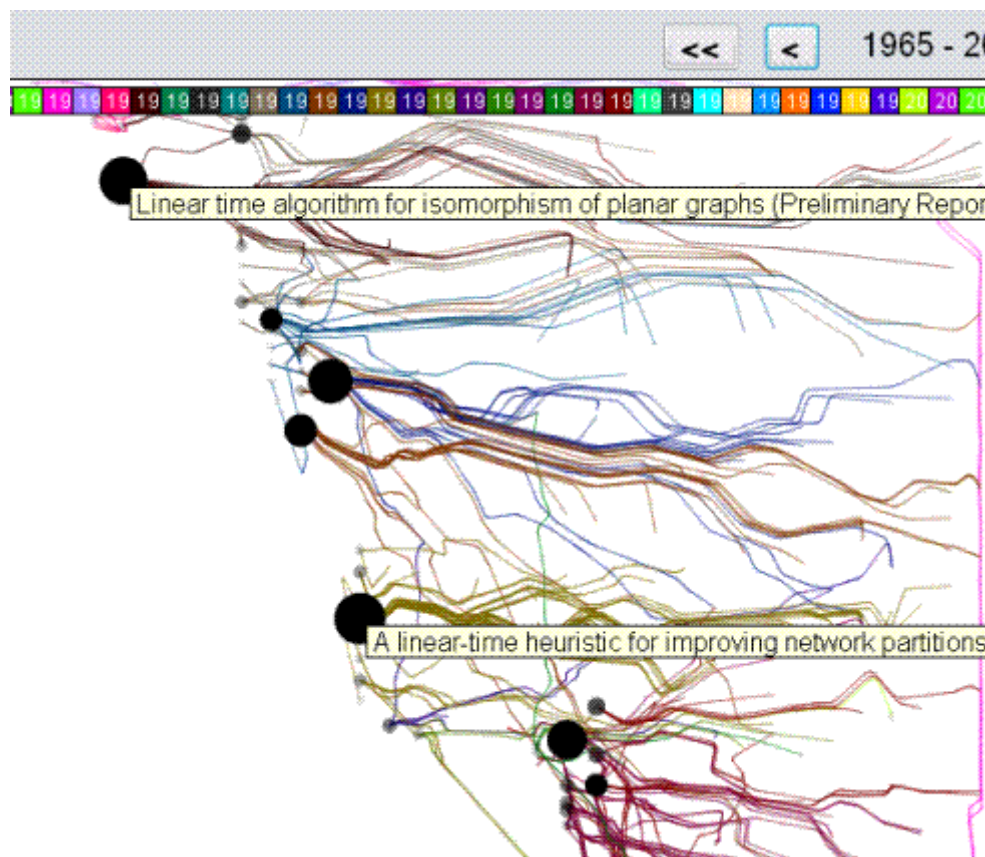


Рис. 10. Применение идеологии жгутов ребер к методу поуровневого размещения вершин.

## Заключение

В данной работе рассмотрены методы извлечения сетей соавторства и сетей цитирования большого объема из баз данных, созданных в рамках проекта Linked Open Data, а также продемонстрированы новый метод кластеризации для сетей соавторства и новый метод динамической визуализации сетей цитирования. Генерируемые при помощи нашего метода изображения наглядно представляют информацию по цитированию публикаций, позволяют анализировать и оценивать научный уровень работ, продуктивность исследователей и показатели значимости отдельных публикаций. В настоящий момент ведется разработка метода геометрического группирования ребер, благодаря которому удастся уменьшить визуальную перегруженность в графе и работать с большими объемами данных.

## Список литературы

1. Апанович З.В., Кислицына Т.А. Расширение подсистемы визуализации наполнения информационного портала средствами визуальной аналитики // Проблемы управления и моделирования в сложных системах: Труды XII Международной конференции (Самара, 21-23 июня 2010 г.). 2010. С. 518-525.
2. Apanovich Z. V., Vinokurov P. S. Ontology based portals and visual analysis of scientific communities//First Russia and Pacific Conference on Computer Technology and Applications, 6-9 September, 2010. Vladivostok, Russia. 2010. P. 7-11.
3. Bizer, C., Heath, T. and Berners-Lee, T. Linked Data - The Story So Far. //Int. J. Semantic Web Inf. Syst., 5 (3). 2009. P. 1-22.
4. W. Cui, H. Zhou, H. Qu, P. C. Wong, X. Li, "Geometry-Based Edge Clustering for Graph Visualization", IEEE Transactions on Visualization and Computer Graphics //Proceedings Visualization / Information Visualization 2008). — vol. 14, no. 6.— 2008.
5. Fruchterman T. M. J., Reingold E. M. Graph Drawing by Force-Directed Placement//Software -Practice and Experience, 1991, Vol. 21, N11, P. 1129-1164.
6. Holten D. , Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data// IEEE Transactions on Visualization and Computer Graphics.— v.12, n.5.— 2006. — pp. 741-748.
7. Lin, Shen; Kernighan, B. W. "An Effective Heuristic Algorithm for the Traveling-Salesman Problem". Operations Research, 1973. 21(2). P. 498-516.
8. Newman M. E. J., Girvan M. Finding and evaluating community structure in networks// Physical Review E, 69.26113. 2004.
9. Sugiyama K., Tagawa S., Toda M. Methods for Visual Understanding of Hierarchical System Structures, //IEEE Trans. Systems, Man, and Cybernetics. 1981. P. 109-125.
10. <http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets>.
11. Описание онтологии AKT: <http://www.aktors.org/ontology>.
12. Данные портала CiteSeer: <http://citeseer.rkbexplorer.com/>.
13. Данные портала DBLP: <http://dblp.rkbexplorer.com/>.

14. Данные портала ACM: <http://acm.rkbexplorer.com/>.

## **Об авторах**

1. Апанович Зинаида Владимировна - к.ф.-м.н, с.н.с. ИСИ СО РАН (Новосибирск), E-mail: [apanovich@iis.nsk.su](mailto:apanovich@iis.nsk.su);

[apanovich@iis.nsk.su](mailto:apanovich@iis.nsk.su)

2. Винокуров Павел Сергеевич - аспирант ИСИ СО РАН (Новосибирск), E-mail: [vinokurov.pasha@gmail.com](mailto:vinokurov.pasha@gmail.com)

[vinokurov.pasha@gmail.com](mailto:vinokurov.pasha@gmail.com)

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (гранты № 09-07-00400 и 11-07-00388) и проекта РАН 2/12 «Формальные языки и методы спецификации, анализа и синтеза информационных систем».