Средства визуального анализа информационного наполнения порталов, входящих в облако Linked Open Data♣

© З.В. Апанович[1], П.С. Винокуров[1], Т.А. Кислицина[2]
[1]Институт систем информатики СО РАН
[2]Новосибирский государственный университет
apanovich@iis.nsk.su

Аннотация

Благодаря быстрому развитию направления Semantic Web 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 демонстрируются средства визуализации сетей соавторства и сетей цитирования на примере данных, извлеченных из научных порталов, входящих в облако Linked Open Data.

1. Введение

В связи с бурно развивающимся направлением Semantic Web в Интернете становятся доступными большие объемы информации, посвященной различным научным направлениям. В число таких ресурсов входят информационные системы 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 rdf/xml формате. 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, 7]. 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[2,8]. В процессе экспериментов с этими данными были реализованы новые алгоритмы визуализации, описанные ниже.

2. 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 является весьма трудоемкой и ресурсозатратной. В настоящее время функционирует большое количество информационных порталов, имеющих различную структуру, основанных на разных онтологиях, что затрудняет доступ к ним. Последнее время наметились большие сдвиги в

Труды 13й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» - RCDL'2011, Воронеж, Россия, 2011.

унификации доступа к библиографической информации, благодаря новому проекту сообщества Semantic Web, который называется Linked Open Data (LOD) [2]. В рамках этого проекта проделана большая работа по объединению разрозненных наборов данных в единое целое и обеспечению единого механизма доступа [8]. В частности, в рамках этого проекта предоставляется доступ к большому объему структурированной библиографической информации [4-6]. 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, данные портала ACM насчитывают 12,402,336 троек RDF, портал DBLP предоставил 28 384 790 троек RDF. 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 [3], представляющую собой объединение нескольких онтологий, таких как Support Ontology, Portal Ontology, Extensions Ontology и RDF Compatibility Ontology. Онтология Portal Ontology (Рис.1) 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. Так, например, 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", akt:addresses-generic-area-of-interest” 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".

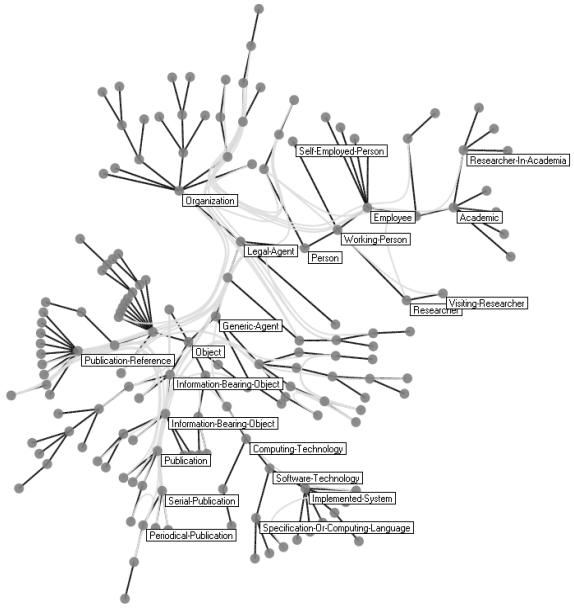


Рис. 1. Основные классы онтологии AKT.

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 простейшую сеть соавторства для любого из перечисленных выше порталов можно сгенерировать с помощью sprql -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.

(а) (б)

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

Следует сказать, что при таком способе генерации сетей соавторства их связность и плотность напрямую связаны с объемом. Например, для портала DBLP[6] при установке лимита на количество ребер в сети соавторства, равном 10000, наибольшая связная компонента этой сети имеет всего 140 вершин и 191 ребро. Для анализа такой сети достаточно обычного алгоритма размещения Фрюхтерман-Рейнгольда[10]. Изображение этой небольшой компоненты связности показано на рисунке 2(а).

При возрастании лимита на объем сети до 50000 ребер, наибольшая связная компонента имеет уже 3001 вершину и 4983 ребра. Для анализа таких компонент связности необходимы специальные алгоритмы. В предыдущих работах [1, 7] нами был представлен алгоритм кластеризации сетей соавторства на основе принципа модулярности [13]. На рисунке 2(а) показано изображение компоненты связности сети соавторства содержащей, 140 вершин и 191 ребро. После кластеризации нашим алгоритмом, получилось 7 кластеров. Вершины, принадлежащие одному кластеру, раскрашены в один цвет (в данном случае, один оттенок серого). Рисунок 2(б) показывает размещение большой компоненты связности сети соавторства, имеющей 3001 вершину и 4983 ребра, после работы нашего старого алгоритма кластеризации. Основным недостатком этого алгоритма является то, что принадлежность кластеру показана при помощи цвета вершин. Расстояние между вершинами практически не зависит от того, какому сообществу принадлежит та или иная вершина, поэтому при большом количестве вершин изображение становится нечитабельным: вершины, принадлежащие разным сообществам располагаются «вперемешку». Для улучшения визуализации нам нужен такой алгоритм размещения, который располагал бы вершины одного сообщества близко друг к другу, а вершины разных сообществ-далеко друг от друга. В настоящий момент реализована многоуровневая версия этого алгоритма, которая существенно повышает качество кластеризации.

3. Кластеризация и визуализация больших сетей соавторства

Для описания алгоритма напомним определение модулярности:

Определим симметричную матрицу $\boldsymbol{e}$ размерности $k \times k$. Элемент $e_{i\,j}$ 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) $a_i = \Sigma_j e_{ij}$, которые соответствуют отношению количества ребер, соединяющих вершины в сообществе i, к общему количеству ребер. Модулярность (modularity) выражается через a_i и e_{ij}:

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

.

Экспериментально показано [9], что значение модулярности, превышающее 0,3, является указателем на реальное наличие сообществ в сети.

Прежний алгоритм выделения сообществ применялся к каждой связной компоненте сгенерированной сети соавторства. Он осуществлялся при помощи удаления ребер, имеющих наибольшую реберную промежуточность. Для оценки реберной промежуточности подсчитывались все кратчайшие пути между всеми парами вершин, и определялось, сколько кратчайших путей проходит через каждое ребро. Затем выбиралось ребро с наибольшим значением промежуточности и удалялось из сети соавторства. Если в результате удаления очередного ребра происходило увеличение количества компонент связности, для нового разбиения подсчитывалась модулярность. При оценке модулярности учитывались все ребра исходного графа. Если новое найденное значение модулярности оказывалось выше, чем прежнее, то это состояние запоминалось, и процесс удаления ребер

продолжался до тех пор, пока разница между текущим значением модулярности и наилучшим значением не станет больше, чем *Параметр_останова*. В этот момент процесс кластеризации завершался и компоненты, соответствующие наилучшему найденному значению модулярности, выдавались в качестве результата кластеризации.

Новая версия этого алгоритма состоит из грубой кластеризации и итеративного улучшения. На этапе грубой кластеризации сеть соавторства разбивается на кластеры, состоящие из одной вершины, затем кластеры, дающие наилучшее увеличение модулярности, попарно объединяются в кластеры большего размера до тех пор, пока еще возможно увеличение значения модулярности. 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 Lin–Kernighan [12]. Алгоритм работает следующим образом.

Определим величину $\Delta Q_{v->D}$ как число, на которое изменится модулярность Q, если переместить вершину v из ее текущего кластера в кластер D.

Тогда алгоритм **Жадного Улучшения кластеризации** состоит из двух шагов:

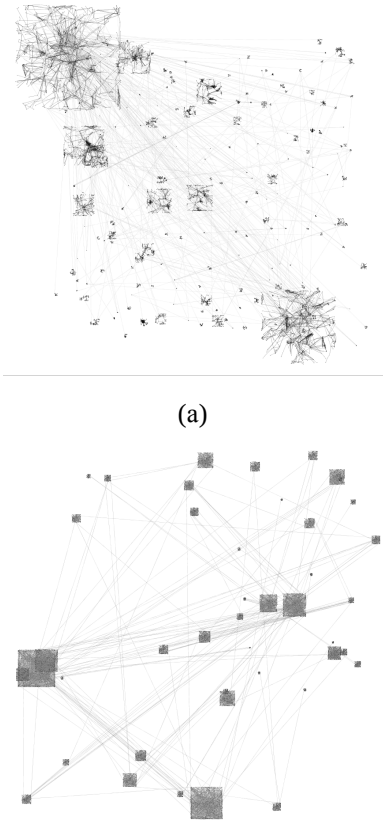
Шаг 1: Для каждой вершины v находится кластер D с максимальным значением $\Delta Q_{v->D}$.Если $\Delta Q_{v->D} > 0$ вершину v перемещается в кластер D.

Шаг 2: Повторяем Шаг 1 до тех пор, пока найдется хотя бы одно перемещение, улучшающее модулярность.

Этот алгоритм перемещает по одной вершине из одного кластера в другой, он 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 кластеров, полученном на этапе грубой кластеризации, выделим уровни, между которыми количество кластеров отличается в 2 раза. Для каждого такого уровня мы имеем набор текущих кластеров, а в качестве перемещаемых вершин используются кластеры, полученные на предыдущем уровне грубой кластеризации. Применяем алгоритм Жадного Улучшения Кластеризации. Этот шаг позволяет еще немного улучшить значение модулярности.

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 сообщества. На этом этапе используется стандартный силовой алгоритм [10]. В процессе размещения считается, что идеальная длина ребра, соединяющего сообщества 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 членов сообщества. Наконец, детальное изображение каждой компоненты подставляется в глобальное размещение компонент и заново отрисовываются все межкомпонентные ребра.



(а)

(б)

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) .

4. Визуализация 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 больших технических усилий, поэтому в открытом доступе эта информация предоставляется только небольшим количеством порталов. Среди порталов облака LOD такими порталами являются Citeseer и ACM[4, 5].

Во-вторых, для генерации информативных сетей цитирования нужны дополнительные усилия. В случае портала Citeseer нами применялась двухуровневая схема генерации сетей цитирования, а в случае портала ACM дополнительно использовалась собственная онтология этого портала, позволяющая выбирать публикации относящиеся к определенному разделу науки.

Наконец, следует отметить, что методы, применяемые при визуализации сетей соавторства, оказались мало пригодными в случае сетей цитирования. Прежде всего, сеть цитирования является ориентированным графом, поэтому для понятного изображения этой сети желательно, чтобы все ребра были направлены в одну сторону. Направление ребер может соответствовать хронологическому порядку публикаций. В принципе, метод изображения иерархических жгутов ребер [11], реализованный нами ранее, соответствует этому требованию. Но применение стандартного метода иерархических жгутов ребер затруднено тем фактом, что в такой базе данных как Citeseer нет достаточно глубокой иерархии, на которую можно было бы наложить сети цитирования. На основе информации о публикациях, мы в своих экспериментах строили иерархию дат публикаций, которая имела всего 2 уровня: год публикации - месяц публикации. В результате получалось изображение, достаточно разреженное в центре и сильно перегруженное на периферии, как это можно видеть на Рис. 4. На этом рисунке показано изображение сети цитирования из 20 000 вершин, извлеченной из базы данных портала Citeseer. Временной период этих публикаций с 1993 по 2003 год. Поскольку ребра в этой сети ориентированные, для облегчения задачи определения направления ребер их концы раскрашены в разные цвета. Входной конец ребра (инцидентный цитируемой вершине-публикации) раскрашен сиреневым цветом, а выходной конец ребра (инцидентный цитирующей публикации) раскрашен зеленым цветом. Можно заметить, что наибольшее количество публикаций в этом множестве приходится на 1998 и 1989 годы. При этом можно рассмотреть достаточно много ссылок на публикации этих лет (жгуты сиреневого цвета), а также заметить, что публикации 2003 года весьма немногочисленны, и от них идут жгуты зеленого цвета – ссылки на более ранние публикации. Для более детального изучения надо рассматривать это изображение фрагментами. При возрастании размеров сети цитирования, в особенности, при увеличении временного интервала, которому принадлежат публикации сети цитирования, эта задача становится весьма трудной для данного алгоритма изображения.

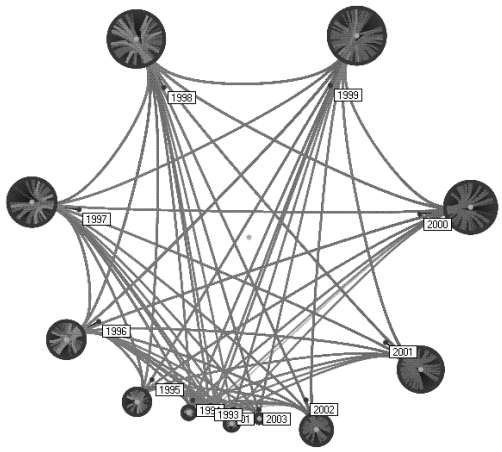
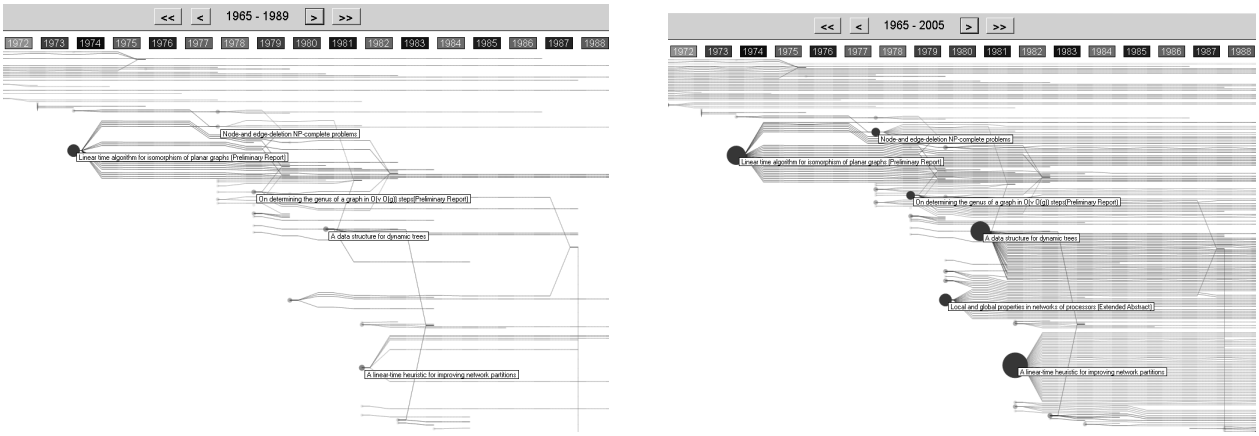


Рис. 4. Изображение сети цитирования, извлеченной из RDF-данных портала Citeseer и содержащей 20 000 вершин.

Для того чтобы сделать возможным просмотр и анализ изменения сетей соавторства на больших промежутках времени, нами был реализован метод поуровневого размещения ориентированного графа с минимизацией пересечений ребер [14]. Суть данного метода состоит в том, что вершины-публикации разбиваются на слои, соответствующие различным годам. Индекс цитирования публикации, т.е. ее значимость, отображается радиусом вершины и интенсивностью ее цвета, что позволяет сразу увидеть самые важные публикации за определенный интервал времени.

На рисунке 5 показано изображение сети цитирования, полученное при помощи поуровневого метода размещения. Вершины этой сети, соответствующие отдельным публикациям, упорядочены хронологически по годам публикаций. Годы публикаций показаны прямоугольниками разного цвета в верхней части изображения. Все публикации, появившиеся в одном году, располагаются в вертикальном столбце, соответствующем этому году. Ребра этой сети соответствуют отношению цитирования. Каждое ребро сети цитирования соответствует отношению akt:cites-publication-reference и ориентировано справа налево. Чем больше ссылок в сети цитирования имеется на некоторую публикацию, тем больше входных ребер имеет соответствующая вершина, и тем больше ее радиус. Цвет каждого ребра, соответствует цвету года цитирующей публикации. Для того чтобы легче было отследить количество ссылок на одну и ту же публикацию, используется процедура минимизации количества



(а) (б)

Рис. 5. Изменение значимости публикаций во времени.

пересечений ребер. В каждом вертикальном ряду осуществляется сортировка вершин, переставляющая каждую вершину в центр тяжести вершин, расположенных в ближайшем к ней ряду слева, с которыми она связана ребром цитирования. Для того чтобы такие перестановки были возможны, каждое длинное ребро цитирования разбивается фиктивными вершинами на короткие ребра. Длинными считаются ребра-ссылки на публикации, с момента появления которых до рассматриваемого момента прошло несколько лет. Каждое короткое ребро соединяет вершины, расположенные в соседних вертикальных рядах. Благодаря этой трансформации, ребра цитирования одной и той же публикации образуют хорошо различимые на рисунке жгуты. 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 « << » происходит очистка изображения.

Перемещение по изображению осуществляется при помощи кнопок « <» и « > », позволяя наблюдать изменение сети цитирования во времени. Технически, эта возможность реализована при помощи фильтрации вершин и ребер сети цитирования.

На рисунке 5 показана изменяющаяся во времени сеть цитирования для публикаций по теории графов. Два рисунка покрывают фрагменты временного интервала с 1965 по 2005 год. В период с 1965 по 1989 (Рис.5 (а)) среди публикаций по теории графов доминирует «Linear-time algorithm for isomorphism of planar graphs». Эта вершина имеет самый большой радиус и большой коричневый шлейф. А в 2005 году (Рис.5(б)) публикация «A linear-time heuristic for improving network partition» становится самой цитируемой.

Можно так же видеть, как появляется интерес к публикации «Node-and-edge-deletion NP-complete problems», причем она ссылается на ранее доминировавшую публикацию «Linear-time algorithm for isomorphism of planar graphs», т.е. образуется цепочка значимых связанных публикаций.

Помимо всего прочего, такой способ визуализации, позволяет обнаруживать ошибки и неточности в библиографических данных.

5. Геометрический метод построения жгутов ребер

Проблемой с применением обычного поуровневого метода изображения сетей цитирования является то, что очень быстро возникает перегруженность изображения, а применение фильтрации, удаляющей малозначимые публикации, искажает реальность: малозначимые публикации вносят основной вклад при определении значимости других публикаций. Поэтому возникла необходимость в алгоритме визуализации, который уменьшал бы визуальную загруженность изображения, формируя жгуты ребер на основе их собственной геометрии, а не привнесенной извне иерархии [9]. Общая схема алгоритма выглядит следующим образом:

- Сгенерировать прямоугольную сетку размера *N*x*N* и наложить ее изображение графа, построенное любым способом.
- Для каждой ячейки прямоугольной сетки вычислить основное направление ребер, пересекающих эту ячейку.
- Объединить соседние ячейки с направлениями, отличающимися не более чем на пороговое значение a, в зоны.
- Вычислить основное направление в каждой зоне, и перпендикуляр к основному направлению зоны.

- Построить отрезки, проходящие перпендикулярно направлению зоны до пересечения с границей зоны.
- Использовать полученные точки пересечения с границей каждой зоны для построения новой сетки при помощи триангуляции.
- Для каждого ребра построенной триангуляции найти точки пересечений с ребрами исходного изображения графа. Вычислить центр среди этих точек.
- Для каждого ребра графа *G* построить *b*-сплайн, проходящий через центральные точки ребер контрольной сетки, которые пересекает ребро графа *G*.

На Рис. 6. показано применение алгоритма геометрических жгутов ребер к изображению, полученному при помощи поуровневого изображения сети цитирования, показанной на Рис.5б.

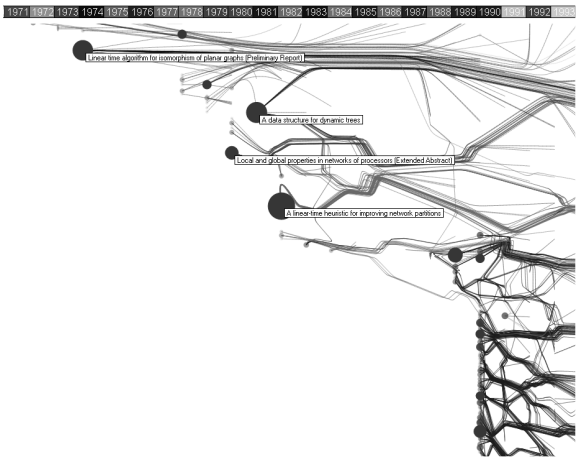


Рис. 6. Применение идеологии жгутов ребер к методу поуровневого размещения вершин.

На настоящем этапе, имеется гораздо больше вопросов, связанных с этим методом, чем ответов на них. Как наилучшим образом выбрать направление прямоугольной сетки? Как зависит направление жгутов ребер от размера сетки? Как выбрать наилучшее направление внутри каждой зоны? Тем не менее, даже в настоящий момент можно констатировать, что этот алгоритм существенно уменьшает загруженность изображения и, мы надеемся его развить до состояния, когда с его помощью можно будет диагностировать тенденции развития научного направления.

Заключение

В данной работе рассмотрены методы извлечения сетей соавторства и сетей цитирования большого объема на примере 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.

Литература

[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.), С. 518-525, 2010.

[2] Данные облака Linked Open Data: http://www.w3.org/wiki/TaskForces/CommunityProjects/LinkingOpenData/DataSets.

[3] Описание онтологии AKT: http://www.aktors.org/ontology.

[4] Данные портала ACM: http://acm.rkbexplorer.com/.

[5] Данные портала CiteSeer: http://citeseer.rkbexplorer.com/.

[6] Данные портала DBLP: http://dblp.rkbexplorer.com/.

[7] 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, pp. 7-11, 2010.

[8] Bizer, C., Heath, T. and Berners-Lee, T. Linked Data - The Story So Far. //Int. J. Semantic Web Inf. Syst., 5 (3), pp. 1-22, 2009.

[9] Cui W., Zhou H.,Qu H.,Wong P.C.,Li X. Geometry-Based Edge Clustering for Graph Visualization // IEEE Transactions on Visualization and Computer Graphics, vol.14 (6), pp.1277-1284, 2008.

[10] Fruchterman T. M. J., Reingold E. M.:”Graph Drawing by Force-Directed Placement” Software - Practice and Experience, Vol. 21, N11, pp. 1129-1164, 1991.

[11] Holten D., Hierarchical Edge Bundles: Visualization of Adjacency Relations in Hierarchical Data// IEEE Transactions on Visualization and Computer Graphics, v.12, n.5, pp. 741-748, 2006.

[12] Lin, Sh., Kernighan, B. W. "An Effective Heuristic Algorithm for the Traveling-Salesman Problem". Operations Research, 21(2). pp. 498–516, 1973.

[13] Newman M. E. J., Girvan M. Finding and evaluating community structure in networks// Physical Review E, 69.26113. 2004.
[14] Sugiyama K., Tagawa S., Toda M. Methods for Visual Understanding of Hierarchical System Structures, //IEEE Trans. Systems, Man, and Cybernetics, pp. 109-125, 1981.

Tools for Visual Analysis of Information Content of Portals Included in Linked Open Data Cloud

© Z.V. Apanovich, T.A. Kislicyna, P.S.Vinokurov

Due to the fast development of Semantic Web and its new branch of Linked Open Data, large amounts of structured information on various scientific areas become available. Digital libraries, information systems and portals based on ontologies are the most reliable sources of this information that need careful investigation in order to optimize management of science. A generally accepted way to facilitate understanding of such large and complex data sets is a graph visualization. This paper is devoted to visualization of citation networks extracted from information portals and digital libraries based on ontologies.

♣ 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».