

# ДИНАМИЧЕСКИЕ СРЕДСТВА ВИЗУАЛИЗАЦИИ СЕТЕЙ ЦИТИРОВАНИЯ, ИЗВЛЕЧЕННЫХ ИЗ НАПОЛНЕНИЯ НАУЧНЫХ ПОРТАЛОВ

Апанович З. В.

Институт систем информатики Сибирского отделения Российской академии наук  
г. Новосибирск, 630090, Россия  
тел.: 383-3309344, e-mail: [apanovich@iis.nsk.su](mailto:apanovich@iis.nsk.su)

**Аннотация** — Разработано два метода динамической визуализации сетей цитирования, извлекаемых из наполнения научных порталов. Генерируемые изображения наглядно представляют информацию по цитированию публикаций, позволяют анализировать и оценивать научный уровень работ, продуктивность исследователей и показатели значимости отдельных публикаций

## I. Введение

В связи с бурно развивающимся направлением Semantic Web и его новой ветвью Linked Open Data [1] 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 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. В данной работе описаны методы визуализации сетей цитирования на примере таких порталов, как Citeseer и ACM [4,5].

## II. Визуализация сетей цитирования

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.

Для того, чтобы сделать возможным просмотр и анализ изменения сетей соавторства на больших промежутках времени нами были реализованы два метода динамического поуровневого размещения ориентированного графа [8]. Суть первого метода состоит в том, что вершины-публикации разбиваются на слои, соответствующие различным годам. Индекс цитирования публикации, т.е. ее значимость, отображается площадью вершины и интенсивностью ее цвета, что позволяет сразу увидеть самые важные публикации за определенный интервал времени.

На Рис. 1 показано изображение сети цитирования, сгенерированной по RDF- данным портала ACM. Вершины этой сети соответствуют отдельным публикациям, упорядоченным по годам публикаций. Ка-

ждое ребро сети цитирования соответствует отношению акт:cites-publication-reference АКТ-онтологии и ориентировано справа налево. Годы публикаций показаны прямоугольниками разного цвета в верхней части изображения. Все публикации, появившиеся в одном году, располагаются в вертикальном столбце, соответствующем этому году. 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. 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.

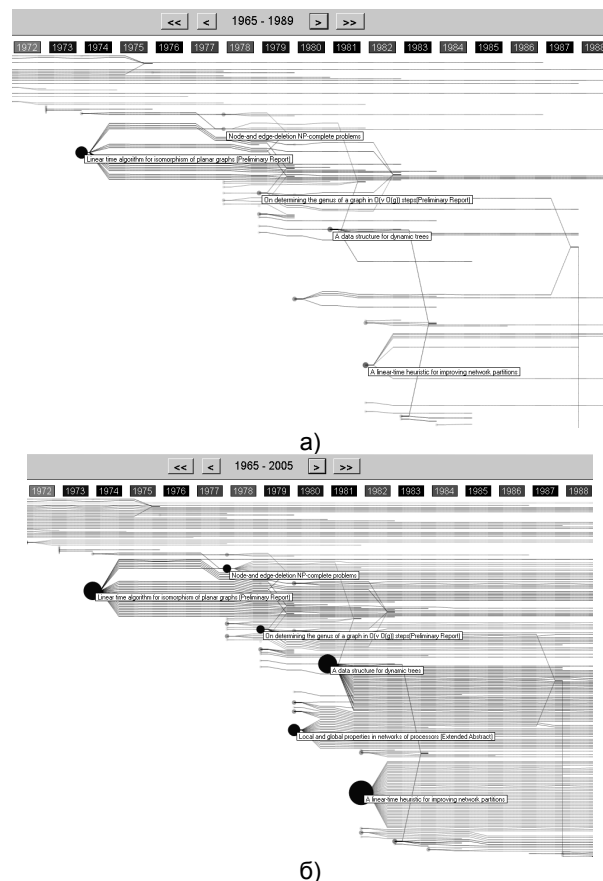


Рис. 1. Публикации по теории графов с 1965 года по 2005 год.

Fig. 1. Publications on Graph Theory from 1965 to 2005

Два рисунка покрывают разные фрагменты временного интервала с 1965 по 2005 год. В период с 1965 по 1989 (рис.1(а)) среди публикаций по теории графов доминирует «Linear-time algorithm for isomorphism of planar graphs», которая имеет самый

большой радиус и большой коричневый шлейф. В 2005 году (Рис.1(б)) публикация «A linear-time heuristic for improving network partition» является 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.

В настоящий момент реализована модификация этого алгоритма, позволяющая осуществлять кластеризацию параллельно идущих ребер для их объединения в жгуты, как это показано на Рис. 2.

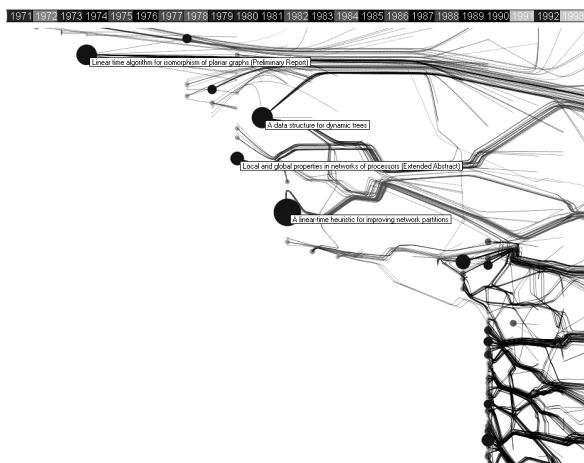


Рис. 2. Геометрические жгуты ребер.

Fig. 2. Geometrical edge bundles

Новый алгоритм визуализации находится в экспериментальной эксплуатации. Он существенно уменьшает загруженность изображения, и мы надеемся, что с его помощью будет легче выявлять тенденции развития научного направления.

### III. Заключение

В данной работе продемонстрированы два метода динамической визуализации сетей цитирования. Генерируемые при помощи наших методов изображения наглядно представляют информацию по цитированию публикаций, позволяют анализировать и оценивать научный уровень работ, продуктивность исследователей и показатели значимости отдельных публикаций.

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.

### V. Список литературы

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

- [2] 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.
- [3] Апанович З. В., Vinokurov P. S. Ontology based portals and visual analysis of scientific communities // First Russia and Pacific Conference on Computer Technology and Applications, (Vladivostok, September 6—9, 2010). Vladivostok, 2010. P. 7—11.
- [4] ACM. URL: <http://acm.rkbexplorer.com/> (дата обращения: 09.05.2011).
- [5] CiteSeer. URL: <http://citeseer.rkbexplorer.com/> (дата обращения: 09.05.2011).
- [6] Sugiyama K., Tagawa S., Toda M. Methods for Visual Understanding of Hierarchical System Structures // IEEE Trans. Systems, Man, and Cybernetics. 1981. P. 109—125.
- [7] Cui W., Zhou H., Wong P., Li X. Geometry-based edge clustering for graph visualization // IEEE Transactions on Visualization and Computer Graphics. 2008. Vol. 14. No. 6. P. 1277—1284.

## DYNAMIC TOOLS FOR VISUALIZATION OF CITATION NETWORKS EXTRACTED FROM THE CONTENT OF SCIENTIFIC PORTALS

Apanovich Z. V.

Institute of Informatics Systems SB RAS  
Novosibirsk, 630090, Russia

Ph.: 383-3309344, e-mail: [apanovich@iis.nsk.su](mailto:apanovich@iis.nsk.su)

**Abstract** — This paper describes new tools for dynamic visualization of citation networks generated from the content of large scientific portals. The first method is a dynamic version of the well-known layered algorithm. The second method generates edge bundles with respect to the obtained by the first algorithm drawing.

### I. Introduction

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 science management. A generally accepted way to facilitate understanding of such large and complex data sets is graph visualization.

### II. Main Part

Two dynamic algorithms for visualization of citation networks extracted from the content of scientific portals are demonstrated. The first method is a dynamic version of the well-known layered algorithm. The second method generates edge bundles with respect to the obtained by the layered algorithm drawing.

### III. Conclusion

It is shown that the developed methods can be quite useful for visual analysis of citation networks. They allow a user to recognize easily the most important publications and the most influential scientists. The second visualization method reduces visual clutter compared to the first one.