

ИССЛЕДОВАНИЕ СВЯЗНОСТИ ОФИЦИАЛЬНОГО НАУЧНО-ОБРАЗОВАТЕЛЬНОГО ВЕБА



Ю.В. Чуйко, А.В. Чирков

Институт прикладных математических исследований
Карельского научного центра РАН

постановка задачи

Исследуемые веб-подмножества:

- T_A – целевое множество официальных сайтов учреждений и организаций РАН
- T_U – целевое множество сайтов классических университетов РФ

Для них исследуются вопросы:

- определение структуры связей внутри и между веб-подмножествами
- оценка степени связанности подмножеств
- взаимное влияние подмножеств на связанность

переход к терминологии теории графов

Графовое представление веб-подмножества:

ориентированный граф $G=\langle V, E \rangle$, где $V=\{\text{множество сайтов}\}$ – вершины,
 $E=\{\text{уникальные ссылки между узлами } V\}$ – направленные дуги

Задачи:

- визуализация графов
- выделение компонент сильной связности
- определение мощностей и диаметров компонент сильной связности
- определение изменения структуры компонент связности при объединении подмножеств

методика решения задач

- экспертным образом формируются множества сайтов T_A , T_U
- путем сканирования по http собирается информация о ссылках между их узлами
- находятся компоненты сильной связности для множеств и их объединений, а также их диаметры
- результаты графически представляются в наглядном виде для их визуальной оценки
- делаются выводы о степени связанности подмножеств сайтов между собой и взаимного влияния их друг на друга

инструменты

сканирование сайтов:

разработано веб-приложение на PHP - сканер для сбора в базу данных информации о гиперссылках со страниц сайта

анализ связности веб-графов:

разработано приложение на языке Java SDK 2 для нахождения компонент сильной связности

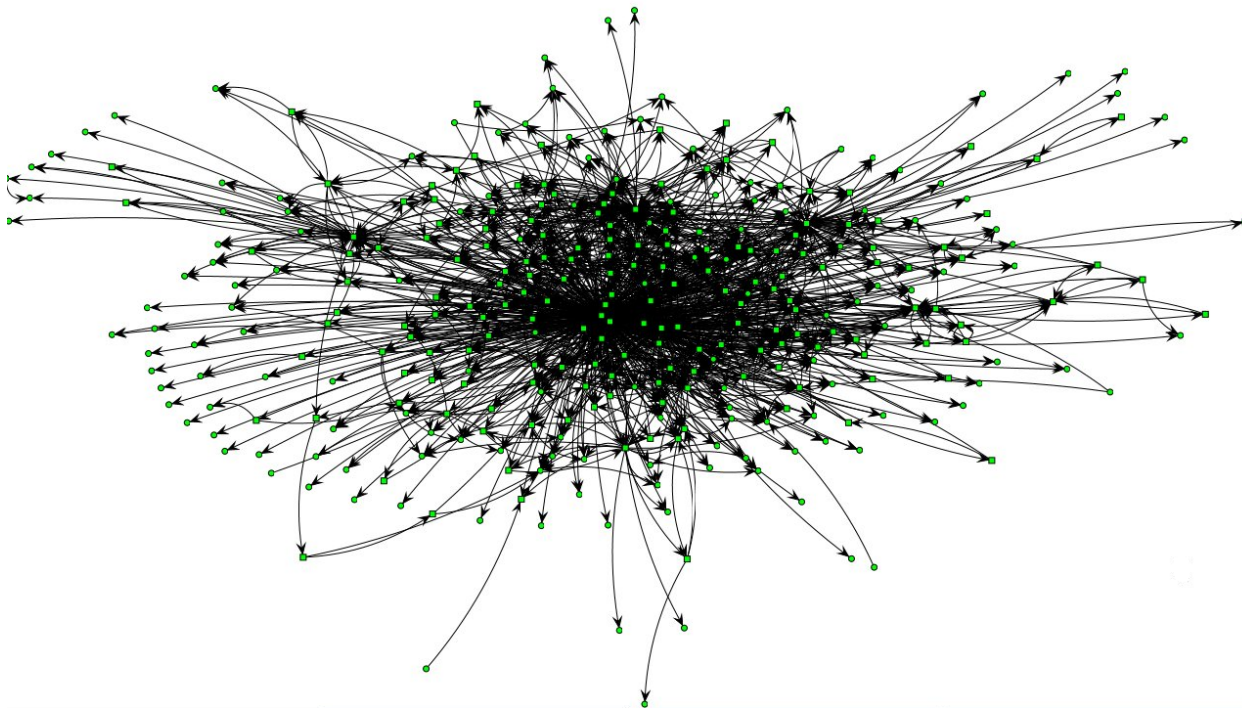
визуализация графов:

использована Java-библиотека JUNG 2 (<http://jung.sourceforge.net/>)

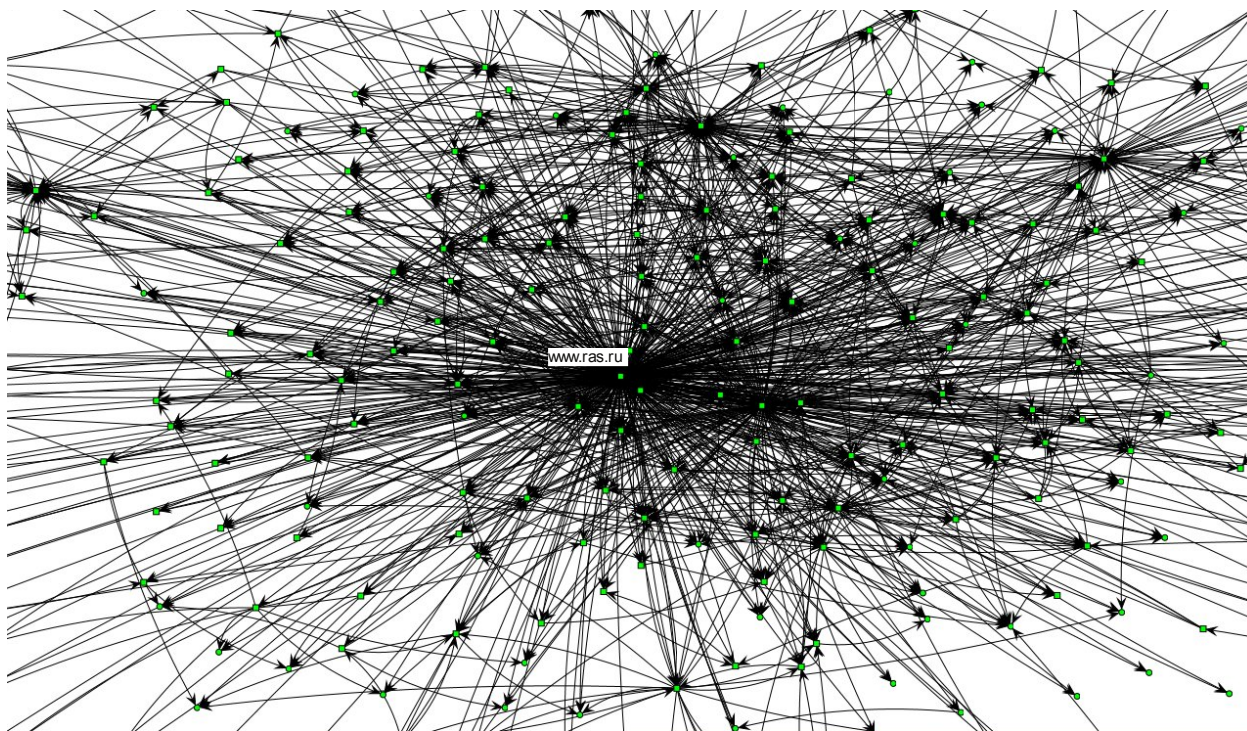
исследование связности академического веба

Множество T_A включает в себя 349 официальных сайтов учреждений и организаций РАН, из которых выделяется единственная нетривиальная КСС - компонента сильной связности (мощность - 180, диаметр - 8).

На рисунке представлена компонента слабой связности данного множества. Квадратами обозначены узлы, входящие в КСС.



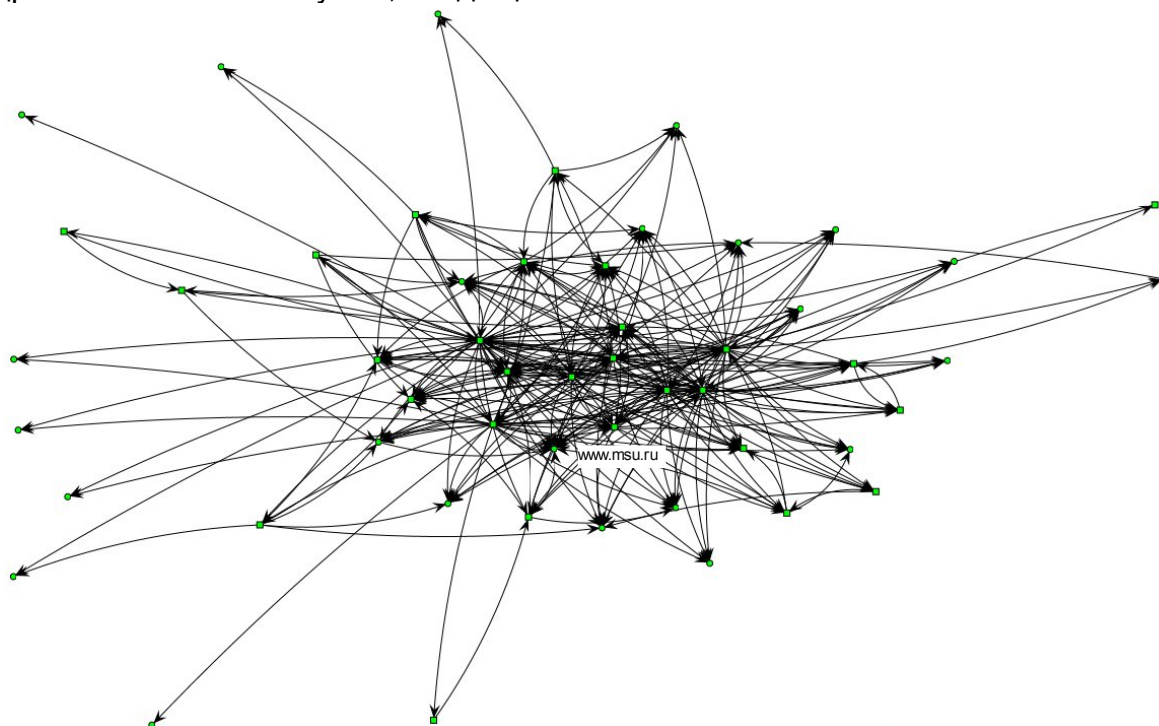
центральная часть рисунка в увеличенном масштабе



исследование связности университетского веба

Множество T_U включает в себя 56 сайтов классических университетов РФ, из которых выделяется единственная нетривиальная КСС - компонента сильной связности (мощность - 28, диаметр - 5).

На рисунке представлена компонента слабой связности данного множества. Квадратами обозначены узлы, входящие в КСС.



наблюдения и выводы

- В обоих случаях в целевом множестве выделяется единственная и достаточно большой мощности компонента сильной связности.
- Структура связности веб-подмножеств аналогична для обоих случаев: академического и университетского веба:
 - единственная компонента сильной связности мощностью > 1
 - примерно совпадают относительные мощности компонент связности в исследованных комбинациях подмножеств
- Для академического веба характерно вхождение в КСС сайтов наиболее значимых организаций, в то время как в университетском наблюдается их некоторая обособленность (в частности, www.msu.ru, www.spbu.ru не попадают в КСС).

Выделение единственной КСС большой мощности в подмножестве важно для входящих в нее узлов. Если узел не принадлежит большой КСС, вероятность его посещения пользователем значительно меньше по сравнению с узлами, входящими в КСС.

Следующая задача – **объединить два подмножества** с целью исследования структуры взаимной связности.

исследование взаимной связанности научного и образовательного веба

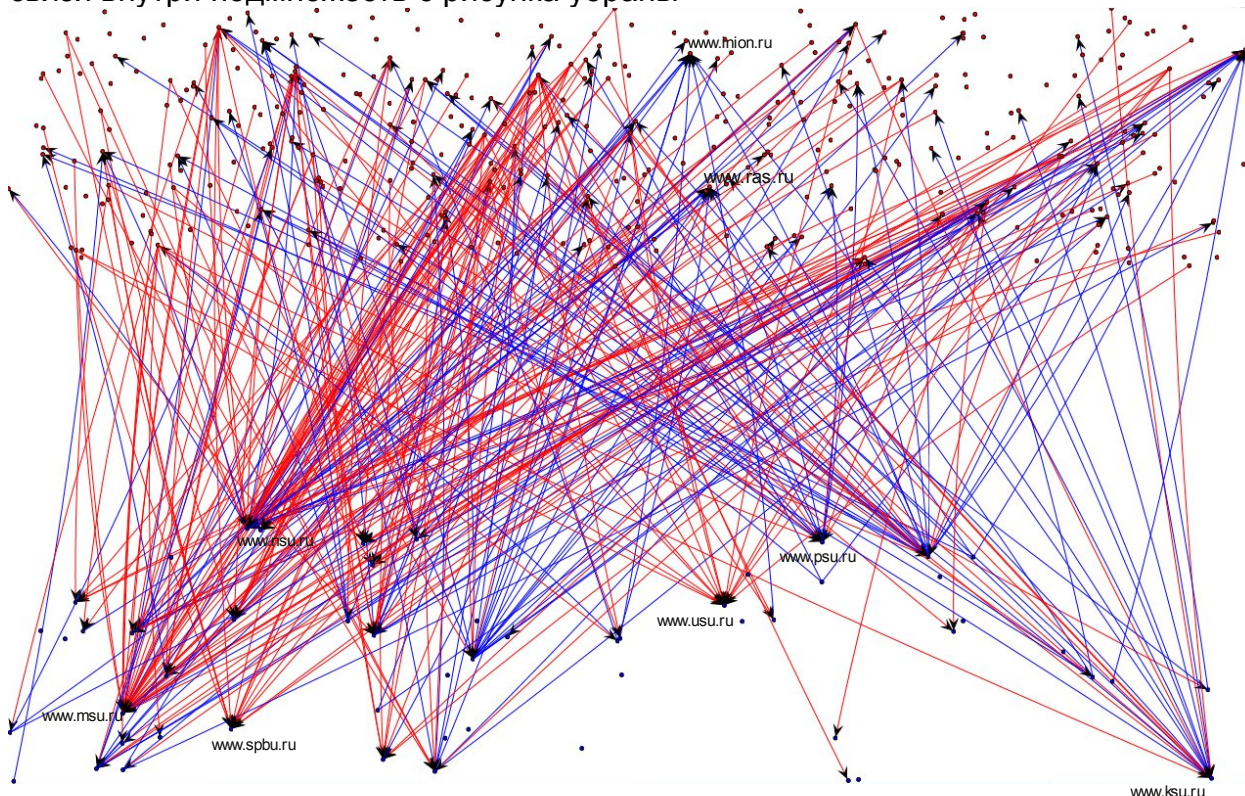
взаимные ссылки между подмножествами T_A и T_U

Сверху (красные) - вершины академического веба, снизу (синие) - университетского.

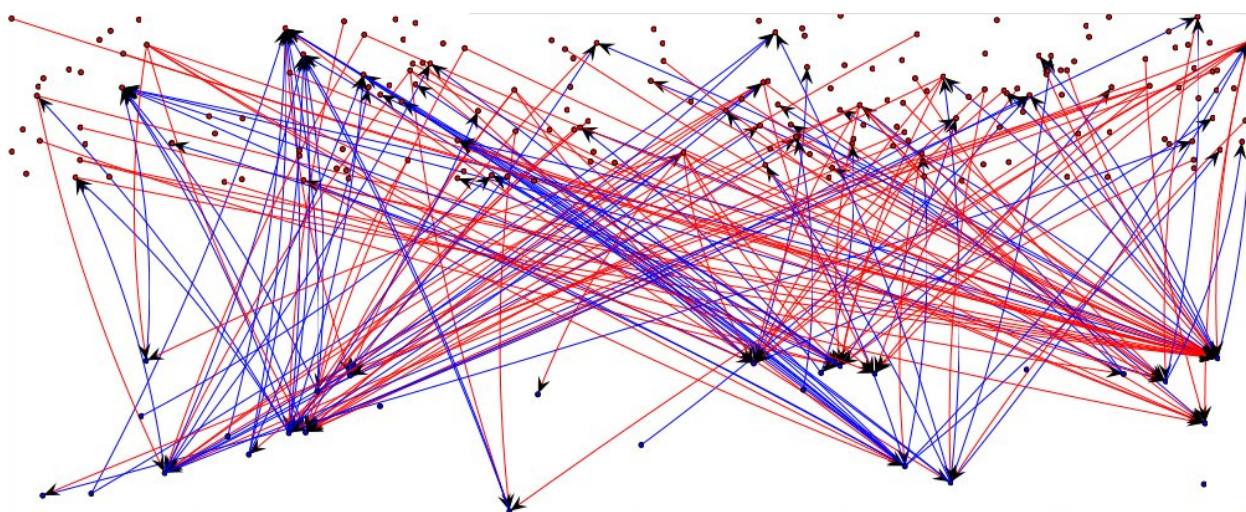
красные стрелки - ссылки с сайтов T_A на T_U

синие стрелки - ссылки с сайтов T_U на T_A

связи внутри подмножеств с рисунка убраны



Два подмножества T_A и T_U достаточно сильно связаны между собой.
То же можно сказать об их КСС:

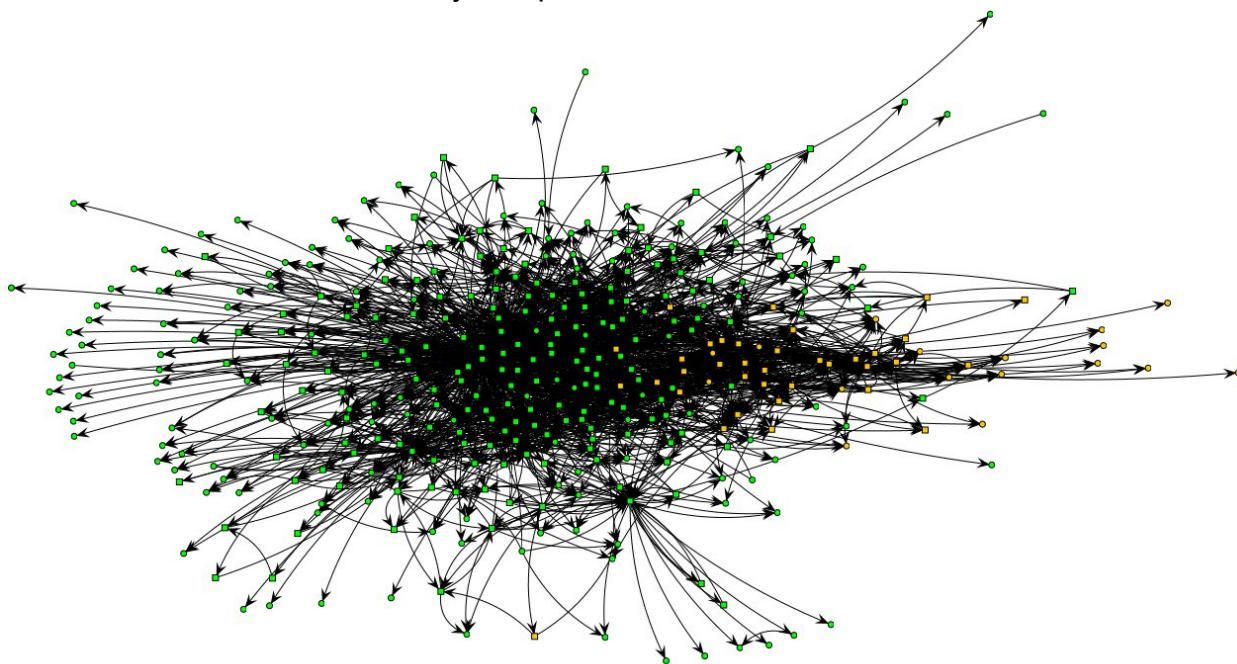


В КСС двух подмножеств выделяются большие группы узлов, ссылки между которыми связывают между собой обе КСС, что обеспечивает наличие единственной КСС в объединенном целевом множестве, включающей обе КСС.

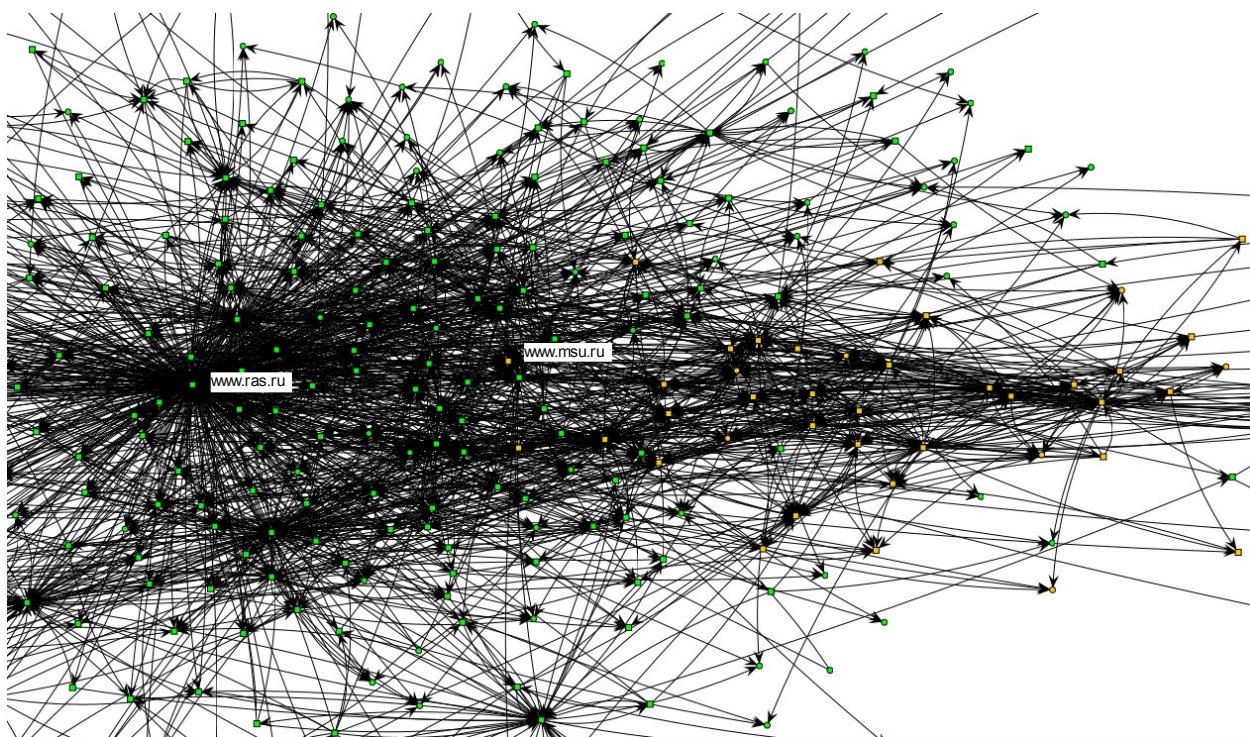
компонента сильной связности научно-образовательного веба

Множество T включает в себя 349 официальных сайтов учреждений и организаций РАН и 56 сайтов классических университетов РФ, из которых выделяется единственная нетривиальная КСС - компонента сильной связности (мощность - 220, из них 36 университетских и 184 академических сайтов, диаметр - 7)

На рисунке квадратами обозначены узлы, входящие в КСС
зеленые узлы - сайты учреждений и организаций РАН
желтые - сайты классических университетов РФ



центральная часть рисунка в увеличенном масштабе



наблюдения и выводы

- В случае объединения целевых множеств в результирующем множестве также выделяется единственная и достаточно большой мощности компонента сильной связности.
- Связность университетского веба значительно возрастает (вместо 28 в КСС входят 36), когда в качестве коммутаторов к нему добавляются академические сайты. Для академических сайтов такое изменение невелико: с 180 на 184.
- КСС научно-образовательного веба включает в себя все узлы КСС академического и университетского подмножеств, а также дополнительное множество узлов, вошедших в КСС, но не попадавших в КСС подмножеств (например: www.philosophy.nsc.ru, www.ipae.uran.ru, www.ieie.nsc.ru из академических, www.msu.ru, www.kamgu.ru, www.masu.ru из университетских сайтов).
- Наиболее значимые организации академического веба входят в КСС как академического, так и научно-образовательного веба.
- Часть значимых организаций университетского веба, не входившие в его КСС, вошли в КСС научно-образовательного (www.msu.ru), в то время как другая их часть осталась в компоненте слабой связности (www.spbu.ru).
- Диаметр КСС научно-образовательного веба меньше диаметра КСС академического веба, несмотря на ее большую мощность.

Из данных наблюдений можно сделать следующие выводы

- Академия наук и университеты сильно связаны между собой взаимным сотрудничеством.
- Связность университетского веба с академическим значительно более важна для университетов, чем для организаций РАН.
- Высокая степень связности научно-образовательного веба способствует увеличению посещаемости сайтов обоих целевых подмножеств.

ЛИТЕРАТУРА:

1. ВЕБОМЕТРИКА. Институт прикладных математических исследований КарНЦ РАН. [Электронный ресурс] – Режим доступа: <http://webometrics.krc.karelia.ru>.
2. А.А. Печников, Н.Б. Луговая, Ю.В. Чуйко, И.Э. Косинец "Разработка инструментов для вебметрических исследований гиперссылок научных сайтов" // Вычислительные технологии. 2009. Том 14. №5. С. 66-78.
3. Программа для ЭВМ "Информационная система для вебметрических исследований". Свидетельство о государственной регистрации программы для ЭВМ "Информационная система для вебметрических исследований" Федеральной службы по интеллектуальной собственности, патентам и товарным знакам РФ № 2010610941 от 29 января 2010 г.
4. Н. Кристофидес Теория графов: алгоритмический подход. - М.: Мир, 1978. - 429 с.