



Конференция «Философия. Язык. Культура»

Можно ли измерить науку: философия, язык и культура современной наукометрии*

И. Ефименко,
школа филологии
ФГН НИУ ВШЭ
iefimenko@hse.ru

* Презентация к пленарному докладу, 29.04.2015



План презентации

- История вопроса
- Наукометрия. Базовые понятия и индикаторы
- Библиометрические базы и сервисы
- Динамические модели наукометрии
- Наукометрия и Big Data
- Повестка современной наукометрии
- Примеры исследований и аналитических сервисов
- Гуманитарные науки в наукометрии



Начнем...

История вопроса

Наукометрия — дисциплина, изучающая эволюцию науки через многочисленные измерения и статистическую обработку научной информации (количество научных статей, опубликованных в данный период времени, цитируемость и т. д.).

Scientometrics is the study of measuring and analyzing science, technology and innovation. Major research issues include the measurement of impact, reference sets of articles to investigate the impact of journals and institutes, understanding of scientific citations, mapping scientific fields and the production of indicators for use in policy and management contexts.



История вопроса

1. История, философия, социология науки:
 - ✓ Область на стыке со сферой политики, как средство поддержки «принятия решений» (в т.ч. самими учеными)
 - ✓ Полезный исследовательский инструмент – например, для исследования эволюции науки и технологий

2. Герменевтика, семиотика:
 - ✓ «Знаковая система» в коммуникации ученых
 - ✓ Средства интерпретации «паттернов» наукометрических данных
 - «Наукометрический анализ наукометрии» – траектории исследователей: «физики и лирики», химики и философы

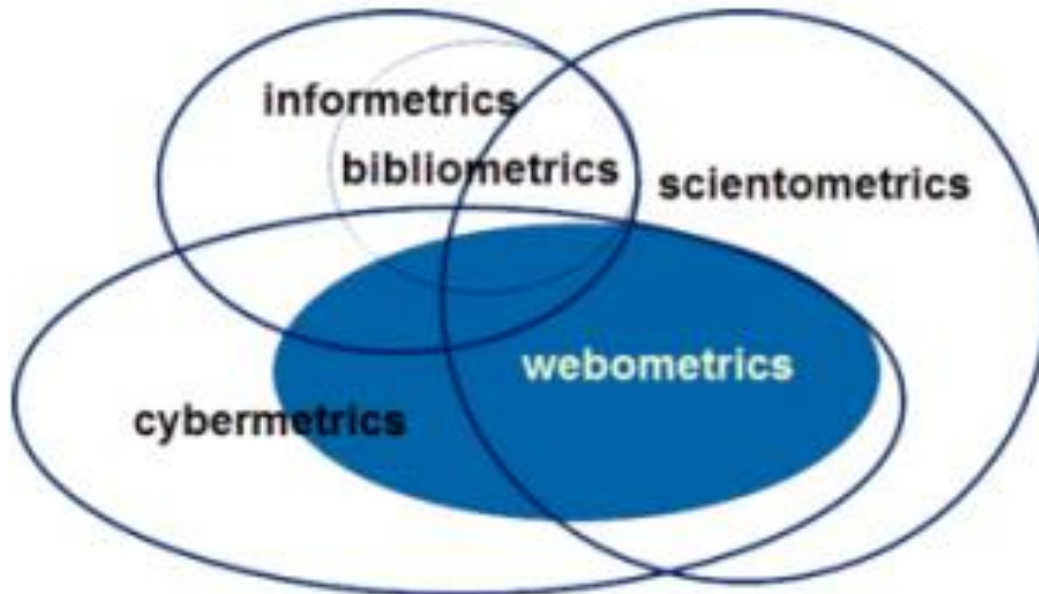
3. Наука о данных, искусственный интеллект



История вопроса

- (Лишь некоторые) проблемы:
 - ✓ Критерии результативности и прогресса?
 - ✓ Игнорирование важных нюансов (Модель vs. Действительность)
 - ✓ «Эффект Матфея» (Р. К. Мертон) или «Клуб богатых»

История вопроса



<http://blogs.ubc.ca/>



История вопроса

Webometrics also called **cybermetrics** is a relatively new information discipline that aims to quantitatively measure web phenomena. By web phenomena we mean the number of hyperlinks on a web site, their structure and patterns of use by web users, and their reciprocal links by other websites.

Generally speaking, this data can be mined by search engine websites although several experts in webometrics recommend the use of *data cleansing* heuristics to ensure their reliability.

<http://blogs.ubc.ca/>

История вопроса

Bibliometrics and **Scientometrics** are two closely-related fields that aim to measure scientific publications and science in general. A lot of the research that falls under this topic involves citation analysis, or examining how scholars cite one another in publications.

Author citation data can show a lot about scholar networks and scholarly communication, linkages between scholars, and the development of areas of knowledge over time.

<http://blogs.ubc.ca/>

- Василий Налимов
- Дерек Джон де Солла Прайс («Малая наука, большая наука», 1963)
- Юджин Гарфилд
“*The Russians are coming*” paper (*Red Hot Soviet Scientists*)

Ключевые методы. *Пример*: коцитирование (Г. Смолл, И. Маршакова, 1973)



История вопроса

Два крупных периода в истории науки – «малая наука» и «большая наука»:

- *«Малая наука»: разрозненные усилия учёных по наблюдению за окружающим миром, выведению закономерностей и постулатов, описывающих функционирование природы и человека.*
- *«Большая наука»: с возникновением научных обществ и научных учреждений начался новый период в истории науки, который положил начало «большой науке». Именно с этого периода, когда наука стала управляемым и профессиональным видом деятельности, следует отсчитывать настоящую историю науки.*

Различные точки зрения. «Технонаука» и др.



История вопроса

Altmetrics (*alternate or alternative metrics*) is a field of web-based metrics that accounts for total author influence. Quantifying this scholarly activity is new and differs from citation metrics. *Altmetrics*, or alternative citation metrics, provides researchers and scholars with new ways to track influence for emerging scholarly communication on the web (number of links, downloads, etc.).

<http://blogs.ubc.ca/>



История вопроса

- Наукометрия — дисциплина, изучающая эволюцию науки через многочисленные измерения и статистическую обработку информации.
- Наукометрию часто применяют как абсолютную основу оценки выполнения и финансирования различных организационных единиц (институтов, команд, индивидуумов, а также проектов и разработок).

Термин «наукометрия» был впервые введен в монографии В.В. Налимова, З. М. Мульченко «Наукометрия: изучение науки как информационного процесса» (1969). Затем появились различные переводы.

- Scientometrics is the study of measuring and analysing science, technologies and innovations.
- Major research issues include the measurement of impact, reference sets of articles to investigate the impact of journals and institutes, understanding of scientific citations, mapping scientific fields and the production of indicators for use in policy and management contexts.

Wikipedia

Наукометрия: базовые понятия и индикаторы

«Эффективность X». Простое vs. комплексное понимание





Базовые понятия и индикаторы

- Инфометрия – научная дисциплина, предметом которой являются количественные измерения хранимой и используемой информации
 - Библиометрия (Bibliometrics) – научная дисциплина, занимающаяся изучением документов на основе количественного анализа первичных и вторичных источников информации с помощью формализованных методов с целью получения данных об эффективности, динамике, структуре и закономерностях развития исследуемых областей
 - Наукометрия (Scientometrics) – научная дисциплина, занимающаяся изучением количественных методов развития науки как информационного процесса
 - Вебометрика (Webometrics) – раздел информатики, где исследуются количественные аспекты конструирования и использования информационных ресурсов, структур и технологий применительно к среде Интернет



Базовые понятия и индикаторы

- Индекс цитирования научных статей (ИЦ) — реферативная база данных научных публикаций, индексирующая ссылки, указанные в пристатейных списках этих публикаций, и предоставляющая количественные показатели этих ссылок.

В России распространена особая интерпретация понятия «Индекс цитирования»:

- ✓ **Показатель, указывающий на значимость данной статьи и вычисляющийся на основе последующих публикаций, ссылающихся на данную работу.**



Базовые понятия и индикаторы

- Первый индекс цитирования был связан с юридическими ссылками и датируется 1873 г. (Shepard's Citations). См. также Index Medicus (1879 г.). Последний просуществовал вплоть до 2004 г.
- В 1960 году Институт научной информации (ISI), основанный Юджином Гарфилдом, ввёл индекс цитирования для статей, опубликованных в научных журналах, положив начало Science Citation Index (SCI), а затем включив в него индексы цитирования по общественным наукам (Social Sciences Citation Index, SSCI) и искусствам (Arts and Humanities Citation Index, AHCI; в настоящий момент – еще ряд индексов).
- ИЦ представлен, например, в Web of Science.
- Начиная с 2006 г. появились другие источники подобных данных (например, Google Scholar).

«Фото из семейного архива».

Книга «Руководство по наукометрии»



Лирическое
отступление



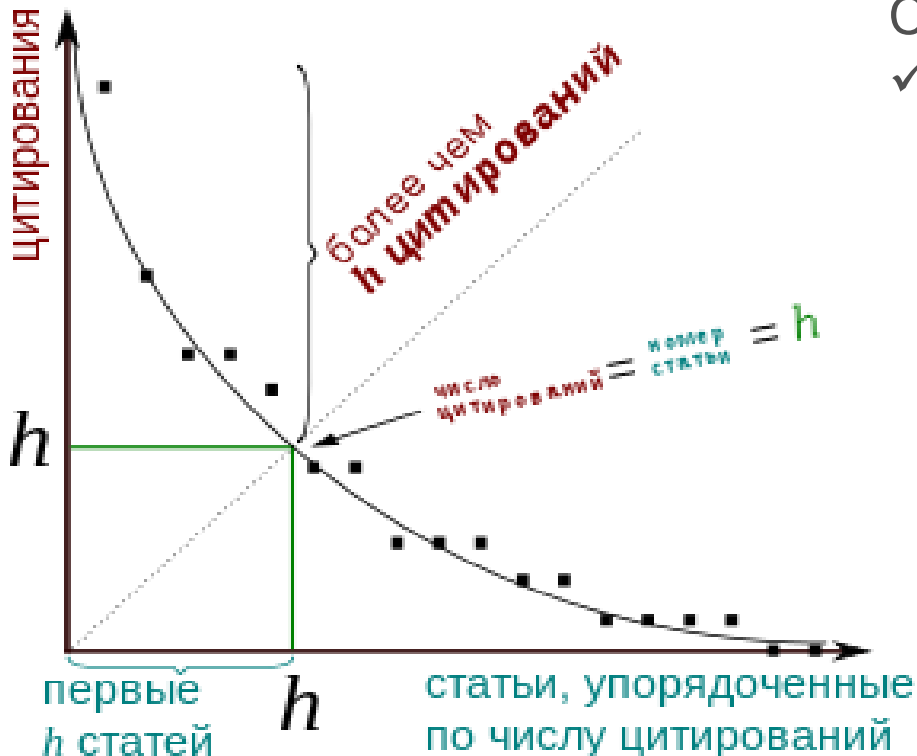
Базовые понятия и индикаторы

- h -индекс (индекс Хирша) – наукометрический показатель, предложенный в 2005 г. аргентино-американским физиком Хорхе Хиршем из Калифорнийского университета в Сан-Диего.

Индекс Хирша является количественной характеристикой продуктивности ученого, группы ученых, научной организации или страны в целом, основанной на количестве публикаций и количестве цитирований этих публикаций.

Базовые понятия и индикаторы

h -индекс вычисляется на основе распределения цитирований работ данного исследователя.



Согласно Хиршу:

- ✓ Учёный имеет индекс h , если h из его N_p статей цитируются как минимум h раз каждая, в то время как оставшиеся $(N_p - h)$ статей цитируются не более, чем h раз каждая.

Примеры вычисления h -индекса:

- Если ученый имеет индекс h , он опубликовал h статей, на каждую из которых сослались как минимум h раз. Так, если у данного исследователя опубликовано 100 статей, на каждую из которых имеется лишь одна ссылка, его h -индекс равен 1. Таким же будет h -индекс исследователя, опубликовавшего одну статью, на которую сослались 100 раз.
- В то же время, если среди публикаций исследователя имеется
 - 1 статья с 9 цитированиями,
 - 2 статьи (включая уже упомянутую статью с 9 цитированиями) с ≥ 8 цитированиями,
 - 3 статьи с ≥ 7 цитированиями, ...
 - 5 статей с ≥ 5 цитированиями, ...
 - 9 статей с ≥ 1 цитированием каждой из них,
- то его h -индекс равен 5 (так как на 5 его статей сослались как минимум по 5 раз).

Базовые понятия и индикаторы

- g-индекс – индекс для измерения научной продуктивности, рассчитываемый на основе библиометрических показателей (2006, Leo Eggh)

Индекс рассчитывается на основе распределения цитирований, полученных публикациями ученого, следующим образом:

Для данного множества статей, отсортированного в порядке убывания количества цитирований, которые получили эти статьи, g-индекс – это наибольшее число, такое что g самых цитируемых статей получили (суммарно) не менее g^2 цитирований.

(Борьба с проблемами h-индекса)



Базовые понятия и индикаторы

- i -индекс - индекс публикационной активности научной организации, рассчитываемый на основе библиометрических показателей. Предложен в 2006 г. независимо М. Космульским и Г. Пратхапом.

Индекс рассчитывается на основе распределения индекса Хирша ученых из данной научной организации:

- *Научная организация имеет индекс i , если не менее i ученых из этой организации имеют h -индекс не менее i .*

Критика и практическое использование h-индекса. «Логика» h-индекса



Не (только) методы разрушающего контроля

- Направление Indicator Engineering и профессиональный юмор как индикатор зрелости (*STI 2014 Conference*)

We encourage anyone with information to share to visit us there, and we wish you a very nice conference.

cons of this idea? Please contact the office in the foyer.



DATA INDICATOR BUBBLES

DID METRICS KILL THE CAT?

According to the famous measurement problem in quantum mechanics, one cannot do a measurement without severely influencing the measured system.

A somewhat similar problem can be identified in Scientometrics.

In his presentation, Paul Wouters warned that researchers in the scientometrics field should be aware that their measurements influence the scientific world. When asked whether this is a problem for

Лирическое отступление

Не (только) методы разрушающего контроля



THOMSON REUTERS
Web of Science



Лирическое
отступление





Не (только) методы разрушающего контроля

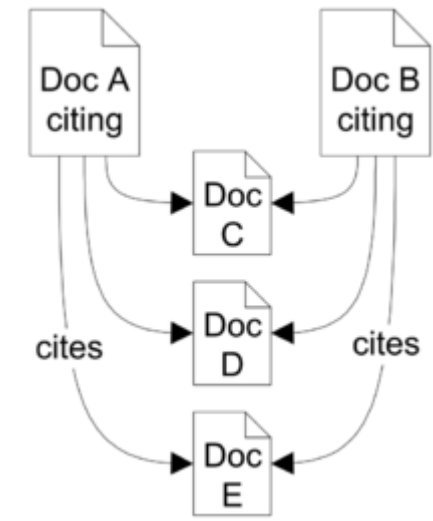
- Направление «О разумном и обоснованном применении индикаторов»
- «Типовые сценарии» использования рейтингов и многое другое

Базовые понятия и индикаторы: немного сложнее

- Исследовательские фронты

Метод библиографического связывания (ретроспективный метод)

Метод коцитирования (проспективный метод)





Базовые понятия и индикаторы: немного сложнее

Примеры исследовательских фронтов, вычисляемых WoS

Исследовательские фронты	Число публикаций, составляющих фронт	Число ссылок	Среднее число ссылок на статью	Год формирования
Unconventional superconductivity; superconductivity close; spin-singlet superconductivity; bulk superconductivity; iron-based layered superconductor	46	19041	413.9	2008
Graphene growth; uniform large area graphene; large-area graphene films; uniform graphene films; wafer scale homogeneous bilayer graphene films	46	11167	242.8	2008
Edge-oxidized zigzag graphene nanoribbons; nitrogen-doped zigzag graphene nanoribbons; edge-modified zigzag graphene nanoribbons; disordered graphene nanoribbons; form graphene nanoribbons	41	10282	250.8	2007



Базовые понятия и индикаторы: немного сложнее

- Импакт-фактор (ИФ, IF) – численный показатель важности научного журнала.

С 60-х годов прошлого века он ежегодно рассчитывается Институтом научной информации (ISI, который в 1992 г. был приобретен корпорацией Thomson и ныне называется Thomson Scientific) и публикуется в журнале Journal Citation Report.

В соответствии с ИФ (в основном, в других странах, но в последнее время и в России) оценивают уровень журналов, качество статей, опубликованных в них, дают финансовую поддержку исследователям и принимают сотрудников на работу.

Расчёт импакт-фактора основан на 3-х летнем периоде. Например, импакт-фактор журнала в 2011 году:

- $IF_{2011} = A/B$, где: A – число цитирований в течение 2011 г. в журналах, отслеживаемых ISI, для статей, опубликованных в данном журнале в 2009-2010 гг.; B – число статей, опубликованных в данном журнале в 2009-2010 гг.



Базовые понятия и индикаторы: немного сложнее

- В 1974 году в ВИНТИ были предприняты попытки создания отечественного указателя научного цитирования (УНЦ), который в технологическом плане должен был стать «аналогом» SCI.
- В 1987 году Китай запускает проект по созданию Китайского индекса научного цитирования Chinese Science Citation Index, а в 1988 году появляется его конкурент – China Scientific and Technical Papers and Citations. В 1997 году начинается разработка китайского индекса цитирования по общественным наукам Chinese Social Sciences Citation Index.
- В 1995 году Япония приступает к созданию национального индекса цитирования Citation Database for Japanese Papers, разработчиком которого становится Национальный институт информатики Японии.
- Разработки национальных индексов ведутся в Тайване (Taiwan Humanities Citation Index), а также в ряде европейских стран (Польша, Испания).



Базовые понятия и индикаторы: немного сложнее

- Национальные индексы и Web of Science
- Проект «1000 журналов»



Базовые понятия и индикаторы: немного сложнее

- Модель vs. Действительность. Неотъемлемое свойство любой модели
- «Не догма, а руководство к действию»
- Что такое индикатор? (Эвристическая) природа индикаторов
- Постоянная критика и постоянное усовершенствование индикаторов. Нарботка аппарата. Новый этап: Big Data – структурированная и неструктурированная информация
- «Сложная» математика и методология
- Положительная и отрицательная роль библиометрических БД. Доказательность. Подход «от данных»

Базовые понятия и индикаторы: немного сложнее

Системы поддержки принятия решений и «мечта всех ЛПР»



Большая красная кнопка

Цель – не (только) в том, чтобы дать ответ, а в том, чтобы предоставить мощный и гибкий инструментарий



Некоторые полезные цитаты

Картографирование науки. И.В. Маршакова-Шайкевич)

Д. Прайс: «Научная статья отнюдь не является единицей информации, которую публикуют, накапливают, находят и выдают по требованию. Она — меняющаяся часть **социальной ткани науки**, и она производится в одних условиях, а используется в других».

Дж. Бернал: «Обзоры и карты столь же необходимы в науке, как и в навигации. Они заслуживают того, чтобы им сегодня уделяли больше внимания. Основные области исследования требуют четкого указания **границ**, а список основных проблем данной отрасли должен **пересматриваться в короткие сроки**»

Д. Прайс : научные работы **«конгломерируются в континенты и государства»**



Некоторые полезные цитаты

Целью кластеризации (более 1 млн публикаций и более 10 млн ссылок) было получение **глобальных всеохватывающих карт науки пяти уровней**.

Самый высокий уровень кластеризации — уровень пятый (С5) — называется «макроуровень». Карта науки этого уровня представляет собой глобальную карту областей исследования в естественных и социальных науках.

Начиная с макроуровня, надо двигаться как бы вниз: выбрав какой-то узел (кластер), можно на следующем уровне К. н. видеть карту этого кластера, также состоящую из совокупности кластеров — подобластей или дисциплин и т. д. до второго уровня картографирования С2, где выделенные кластеры представляют собой **активные исследовательские фронты**, над которыми работает научное сообщество. Нижний уровень С1 включает **кластеры «ядерных» публикаций**, образующих фронты и являющихся, по сути, классическими работами в этих исследовательских направлениях.

Первые карты науки, сформированные в ISI

*к 1990 году, выявили **циклический характер** науки, **неоднородную** ее структуру, **маргинальный статус** отдельных областей, в которых, как правило, происходит **национальная специализация** научных направлений.*



Некоторые полезные цитаты

К. н. методом коцитирования показало, что социальные науки можно включать в одну и ту же структуру вместе с естественными и что существуют **значительные связи** между этими областями. Изучение библиометрических характеристик исследовательских фронтов позволяет выявлять **вклад стран** в становление и развитие новых идей и проблем в науке. Становится очевидным тот факт, что немногие области в науке полностью **изолированы**. Примерно четверть исследовательских направлений уже на третьем уровне агрегации карт являются **междисциплинарными**, что приводит к **пересмотру границ в науке**.

*Систематическое изучение карт науки разных лет дает возможность философам и социологам науки по-новому оценивать те концептуальные изменения, которые постоянно происходят в науке. Сравнение карт коцитирования публикаций и карт коцитирования авторов показало, что эти карты **не конгруэнтны**, т.е. **когнитивная структура** области не изоморфна ее **социальной структуре**; очень немного авторских коллективов соотносится с одним исследовательским фронтом первого уровня.*

*К. н. позволяет проследить **эволюцию** научного знания как на микроуровнях развития, так и на макроуровне, на котором фиксируются изменения, происходящие в крупных, **традиционных** отраслях знания.*



Библиометрические базы и сервисы

- **Web of Knowledge (Web of Science)** – поисковая платформа, объединяющая реферативные базы данных публикаций в научных журналах и патентов, в том числе базы, учитывающие взаимное цитирование публикаций, разрабатываемая и предоставляемая компанией Thomson Reuters. Web of Knowledge охватывает материалы по естественным, техническим, биологическим, общественным, гуманитарным наукам и искусству. Платформа обладает встроенными возможностями поиска, анализа и управления библиографической информацией.



Библиометрические базы и сервисы

- **SCOPUS** – библиографическая и реферативная база данных и инструмент для отслеживания цитируемости статей, опубликованных в научных изданиях. Индексирует 21 тыс.+ наименований научных изданий по техническим, медицинским и гуманитарным наукам ~5 тыс. издателей. База данных индексирует научные журналы, материалы конференций и серийные книжные издания. Разработчиком и владельцем Scopus является издательская корпорация Elsevier.
- В отличие от Web of Knowledge, в Scopus не используется понятие импакт-факторов (см. SJR), но широко применяется индекс Хирша.



Библиометрические базы и сервисы

- **Академия Google (*Google Scholar*)** – свободно доступная поисковая система, обеспечивающая полнотекстовый поиск научных публикаций всех форматов и дисциплин. Система работает с ноября 2004 года, первоначально в статусе бета-версии.
- Индекс *Google Scholar* включает в себя большинство рецензируемых онлайн журналов Европы и Америки крупнейших научных издательств.
- По функциям он похож на свободно доступные системы Scirus от Elsevier, CiteSeerX и getCITED (в настоящий момент поддерживается только CiteSeerX), а также на платные сервисы Scopus и Web of Science.



Библиометрические базы и сервисы

- **РИНЦ** - библиографическая база данных научных публикаций российских ученых.
- Для получения необходимых пользователю данных о публикациях и цитируемости статей на основе базы данных РИНЦ разработан аналитический инструментарий Science Index.
- Проект РИНЦ разрабатывается с 2005 года компанией «Научная электронная библиотека» (ELIBRARY.ru).

Библиометрические БД: Текущая ситуация и накопление информации за длительный период времени (статика и динамика)

Свет потухшей звезды

*Крабовидная
туманность (M 1, NGC
1952, Taurus A) —
газообразная туманность в
созвездии Тельца,
являющаяся остатком
сверхновой SN 1054
и плерионом*



Из презентации М. А. Акоев. *Прогнозирование и планирование ИТ-процесса*, 2015. <http://thomsonreuters.ru/>

Journal Citing Half-Life

- The citing half-life is the median age of articles cited by the journal in the JCR (*J. Cit. Report*) year. For example, in JCR 2003, the journal *Food Biotechnology* has a citing half-life of 9.0. That means that 50% of all articles cited by articles in *Food Biotechnology* in 2003 were published between 1995 and 2003 (inclusive).

Агрегированный показатель для ПО

<http://admin-apps.webofknowledge.com/>



Динамические модели наукометрии

Journal Cited Half-Life

- The median age of the articles that were cited in the JCR year. Half of a journal's cited articles were published more recently than the cited half-life. For example, in JCR 2001 the journal *Crystal Research and Technology* has a cited half-life of 7.0. That means that articles published in *Crystal Research and Technology* between 1995-2001 (inclusive) account for 50% of all citations to articles from that journal in 2001.
- A higher or lower cited half-life does not imply any particular value for a journal. Cited half-life figures may be useful to assist in collection management and archiving decisions.
- Dramatic changes in cited half-life over time may indicate a change in a journal's format. Studying the half-life data of the journals in a comparative study may indicate differences in format and publication history.

Агрегированный показатель для ПО

Динамические модели наукометрии

Sleeping beauty, «Синдром Менделя»



- *“Being ahead of one’s time” has always both fascinated and frightened scientists. Their dearest publication may become a “Sleeping Beauty”, an article that goes unnoticed (“sleeps”) for a long time and then, almost suddenly, attracts a lot of attention.*

Sleeping Beauties in science. A. Van Raan. Scientometrics, 2004

Параметры (индикаторы):

- Глубина сна
- Продолжительность сна
- «Интенсивность» пробуждения

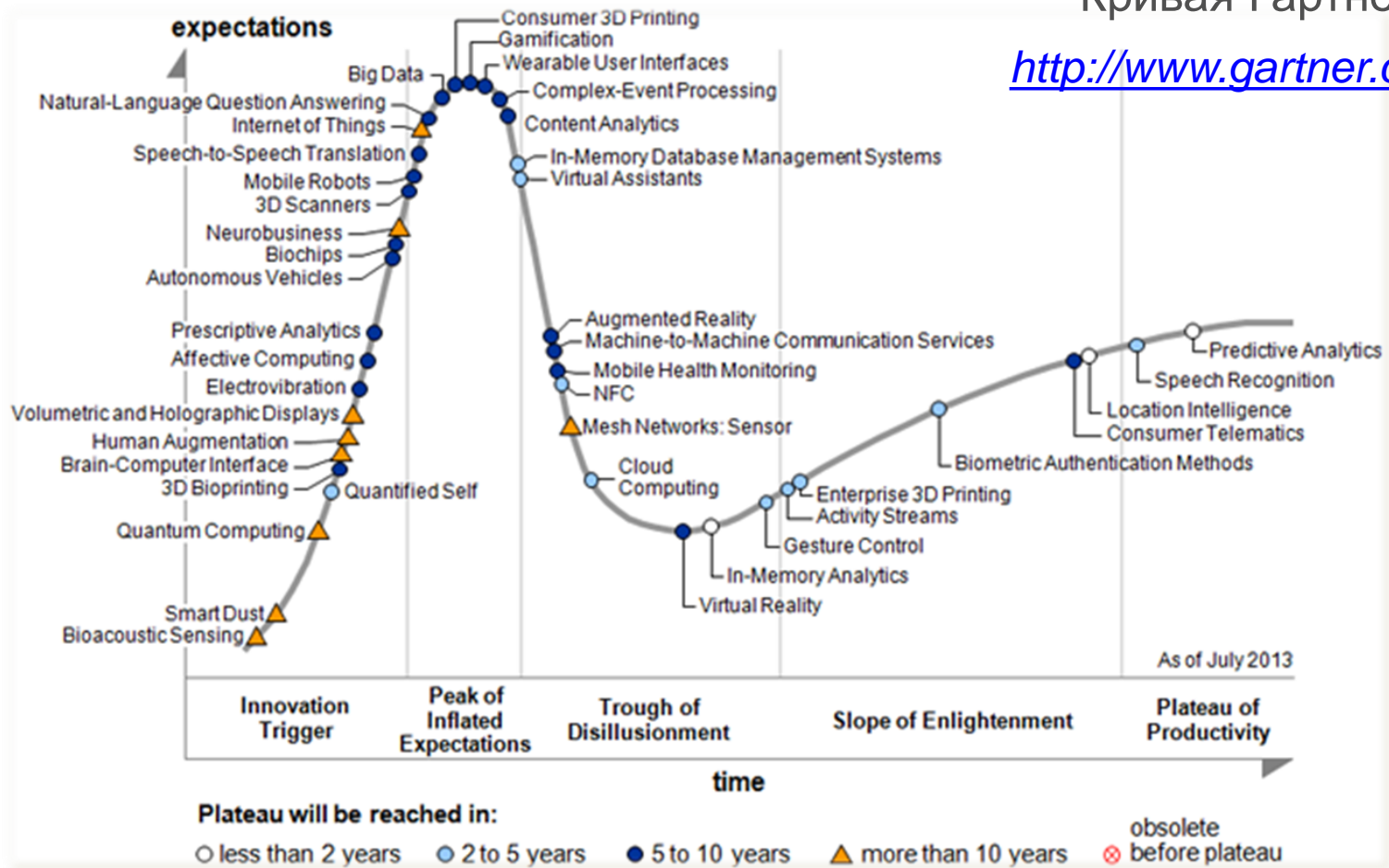
«Экстремальные» спящие красавицы (совокупность критериев, форма кривой)

L. J. ROMANS, Physics Letters B, 169 (1986) Massive $N = 2a$ supergravity in ten dimensions

Динамические модели наукометрии

Кривая Гартнера

<http://www.gartner.com/>





Динамические модели наукометрии

- Динамические модели – Tech Mining –
«Большие данные»...



Наукометрия и Big Data

- Парадигма «Больших данных»
- Аналитический паралич
- 80/20 → 90/10... Velocity. Volume vs. Value
- Обратный тренд: рост объемов «полезной», качественной, достоверной информации, включая

научный контент

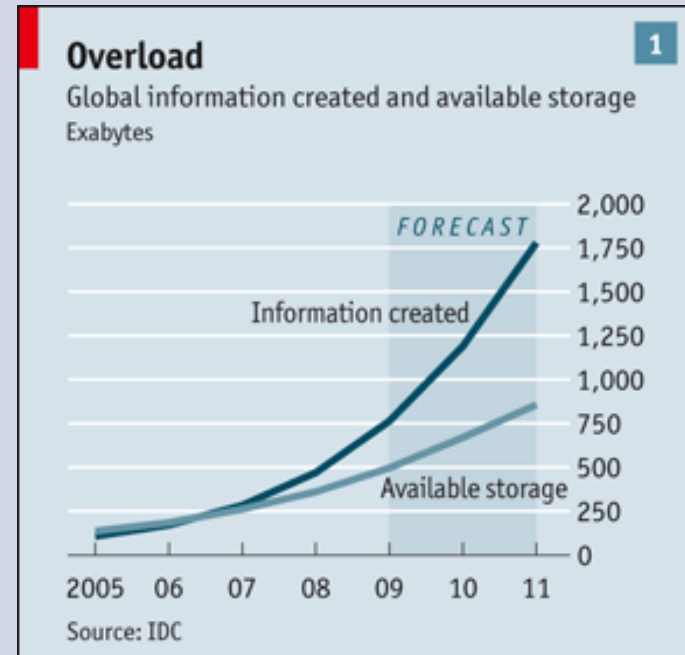
Big Data: Текущая ситуация (2010)

Примеры

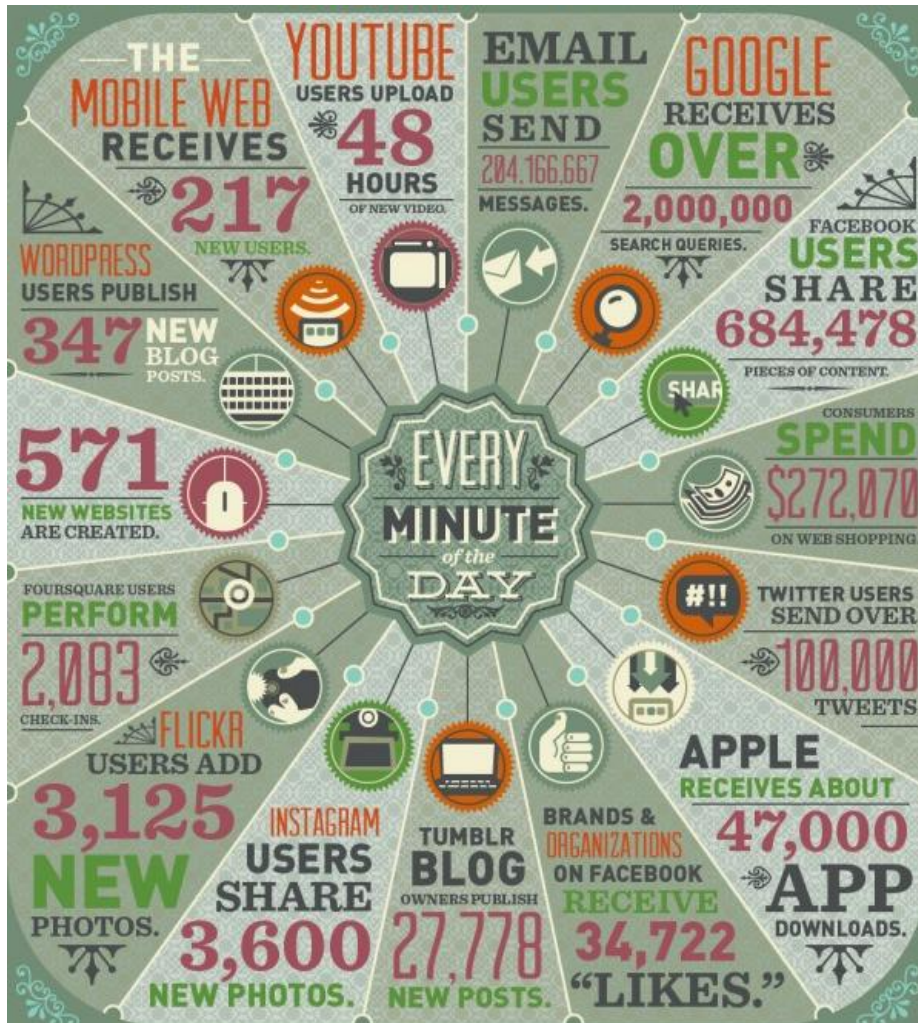
- Телескоп Sloan Digital Sky Survey в Нью-Мексико. За первые недели больше данных, чем за всю историю астрономии. За 10 лет – 140 ТБ
- Large Synoptic Survey Telescope, Чили, с 2016 года. 140 ТБ каждые 5 дней

<http://www.economist.com/node/15557443>

Мрачные прогнозы



Big Data: Текущая ситуация



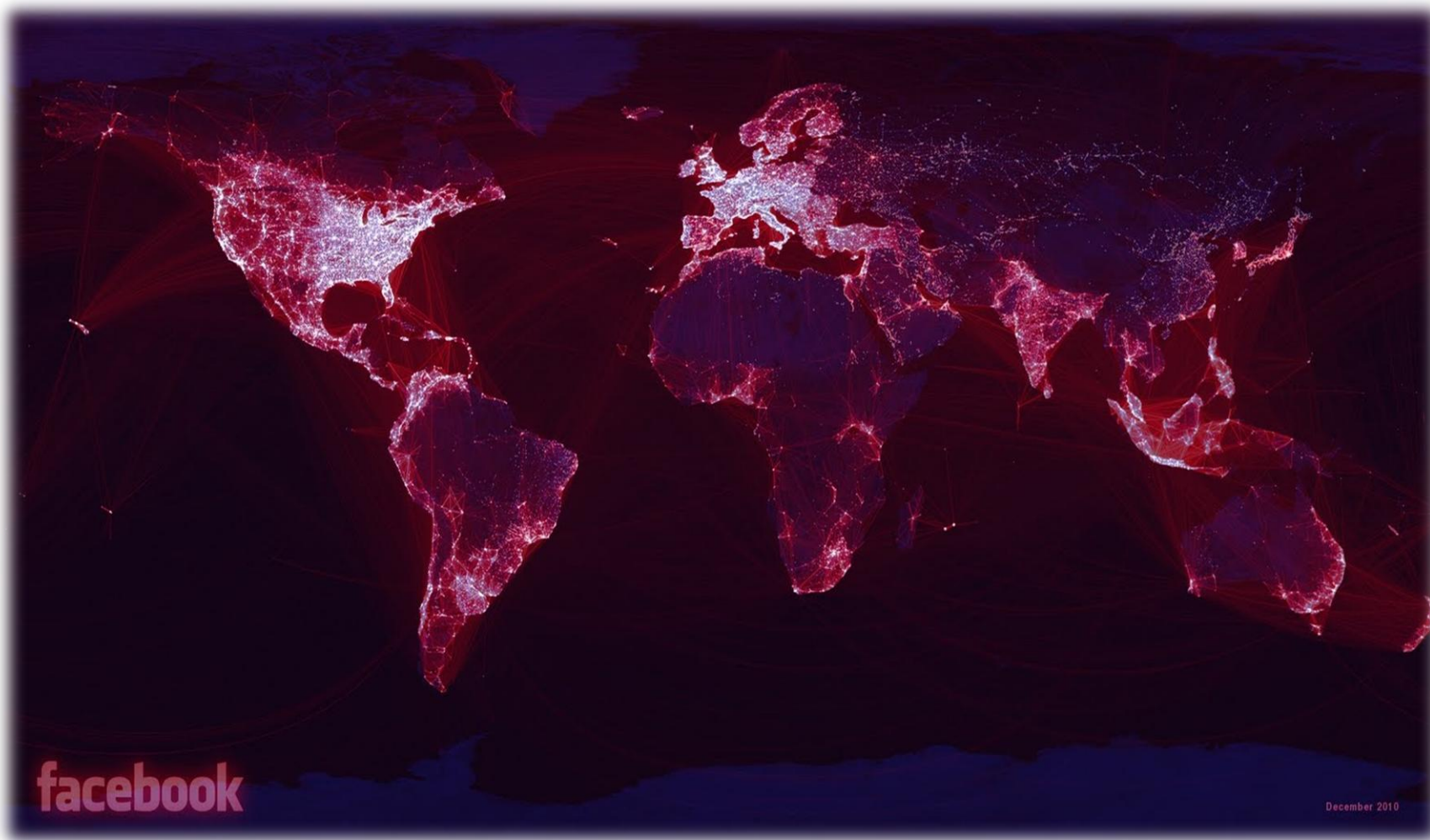
<http://www.visualnews.com/2012/06/19/how-much-data-created-every-minute/?view=infographic>

<http://www.sciencedaily.com/releases/2013/05/130522085217.htm>



Facebook vs. сети коцитирования

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ» 2014 г.



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

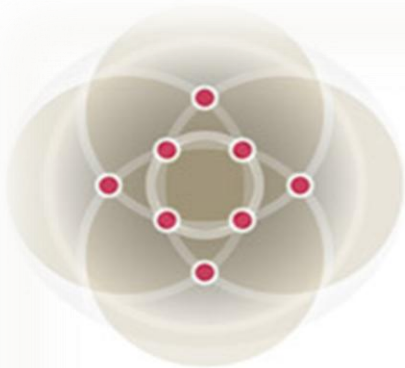
Вызов «традиционной» наукометрии

19th International Conference on Science and Technology Indicators

“Context Counts: Pathways to Master Big and Little Data”

3 - 5 September 2014

Leiden, The Netherlands



5th Annual

Global TechMining
Conference

2015.09.16 • Atlanta, Georgia

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ» 2014 г.



Вызов «традиционной» наукометрии

Текст, неструктурированная информация



«Квалитативный» аспект



Повестка наукометрии

По результатам анализа ведущих российских и международных конференций в области наукометрии за 2014 год

- Ключевой предмет исследования:
 - ✓ Эффективность («качество», производительность, видимость...)
- Дизайн:
 - ✓ Взаимовлияние характеристик объектов
 - ✓ Внешние факторы (Δ , финансирование)

Методы vs. задачи



Повестка наукометрии

- Indicator Engineering (новые индикаторы или модификация существующих):
 - ✓ «Классические индикаторы» в применении к национальным моделям (Δ , с учетом специфики национальных журналов или методов оценки). Национальные индикаторы и БД
 - ✓ Нормализация индикаторов в применении к традициям различных областей науки
 - ✓ «Вторичная аналитика»: анализ адекватности самих индикаторов
 - ✓ Анализ области применения различных индексов («жанровые» ресурсы, Δ , VCI, национальные ресурсы, специализированные по ПО ресурсы)

Повестка наукометрии

■ Влияние индикаторов на поведение ученых

We encourage anyone with information to share to visit us there, and we wish you a very nice conference.

cons of this idea? Please contact the office in the foyer.

DID METRICS KILL THE CAT?

According to the famous measurement problem in quantum mechanics, one cannot do a measurement without severely influencing the measured system.

A somewhat similar problem can be identified in Scientometrics.

In his presentation, Paul Wouters warned that researchers in the scientometrics field should be aware that their measurements influence the scientific world. When asked whether this is a problem for



DATA INDICATOR BUBBLES

Лирическое
отступление

Повестка наукометрии

- Углубленный анализ географической информации:
 - ✓ Использование в интересах решения технологических задач (Δ, идентификация авторов, организаций и других объектов)
 - до 5 или > вариантов транскрипции даже не самых замысловатых фамилий.
(K)horos(c)hevsky / yi / ii...
 - «научный сотрудник... того же института»
 - ✓ Персонализированные сервисы для пользователей (Δ, подбор партнеров с учетом географии)
 - ✓ География как фактор влияния (Δ, анализ влияния «географии» команды на уровень цитируемости)
 - ✓ «Глобальная» аналитика (Δ, межнациональные предпочтения в цитировании; страны / регионы-лидеры в определенной предметной области)
 - ✓ Анализ библиометрических баз данных (Δ, адекватность предоставления информации с учетом языков оригинальных версий статей, стран, регионов)

Повестка наукометрии

- Анализ цитирования, коцитирования, библиографического связывания (*метод*):
 - ✓ «Вторичная аналитика»: анализ индикаторов
 - ✓ Аналитика на сетях цитирования в разрезе различных факторов (Δ, межстрановое, междисциплинарное цитирование)
 - Различные бизнес-задачи (Δ, поиск win-win партнерств с учетом междисциплинарности и других факторов). *Примеры*
 - ✓ Жизненные циклы научных статей и «знаний» («история идей», «трансфер знаний»)
 - ✓ Особые виды цитирования (Δ, самоцитирование). «Плохое» и «хорошее» самоцитирование
 - ✓ Сложные сети и карты науки
 - ✓ Исследовательские фронты
 - ✓ Углубленный анализ цитирования с учетом семантики



Повестка наукометрии

- Технологические задачи:
 - ✓ Разрешение неоднозначностей, глобальная идентификация авторов, научных организаций и менее «классических» объектов (Δ, финансирующих организаций)
 - ✓ Сетевой анализ, методы оптимизации структуры сети
 - ✓ Выравнивание классификаторов

Повестка наукометрии

- Патентный анализ:
 - ✓ «Классические» задачи патентного анализа (зародившиеся за пределами области наукометрии: Δ , анализ областей монопольного владения, лакун и т.п.)
 - ✓ Межжанровое цитирование
 - ✓ Сопоставление данных по научным статьям и патентам
 - ✓ «Классические» задачи наукометрии (innovative women и т.п.)



Повестка наукометрии

- Альтметрика (Altmetrics), Open Access (рост интереса со стороны ключевых «вендоров»: WoS, Elsevier):
 - ✓ Новые методы и подходы к оценке научной эффективности (большое число методологических проблем)
 - ✓ Сравнительная аналитика (Δ , сопоставление числа скачиваний и цитирований в библиометрических БД)
 - ✓ «Классические» задачи науко- и библиометрии (Δ , оценка в страновом, институциональном, предметном разрезе)



Повестка наукометрии

- Научно-техническая политика (научные исследования как субъект и как объект):
 - ✓ Анализ эффективности взаимодействия бизнеса и государства на основе данных о цитировании
 - ✓ Факторы, влияющие на эффективность ученых, коллективов, университетов
 - ✓ Влияние научных исследований на общество
 - ✓ Влияние отдельных крупных проектов на публикационную активность стран
 - ✓ Анализ эффективности различных форм и источников финансирования (Δ , стоимость статьи или цитаты)
 - ✓ Invisible colleges



Повестка наукометрии

- Тренд – наука и бизнес:
 - ✓ Аналитика на результатах взаимодействия университетов и фирм
 - ✓ Композитные индикаторы, применимые к сфере бизнеса (оценка инноваций и др.)
 - ✓ Вклад различных жанров научно-технологической информации в принятие бизнес-решений



Повестка наукометрии

- Аномальные явления, типологии:
 - ✓ Поиск «звезд», top producers и top toppers (К. Бояк, Р. Клаванс, Э. Нойонс и др.)

...“top toppers”, i.e. scholars among the 25% most productive, the 25% most productive of highly cited publications, and also among the 25% scholars publishing in the best journals. A third typology of scholars are the “high impact” scholars, i.e., those belonging to the top 25% of the world in terms of publishing highly cited publications and impact of their journals, but belonging to the segment of producers between the median and percentile 25...



Повестка наукометрии

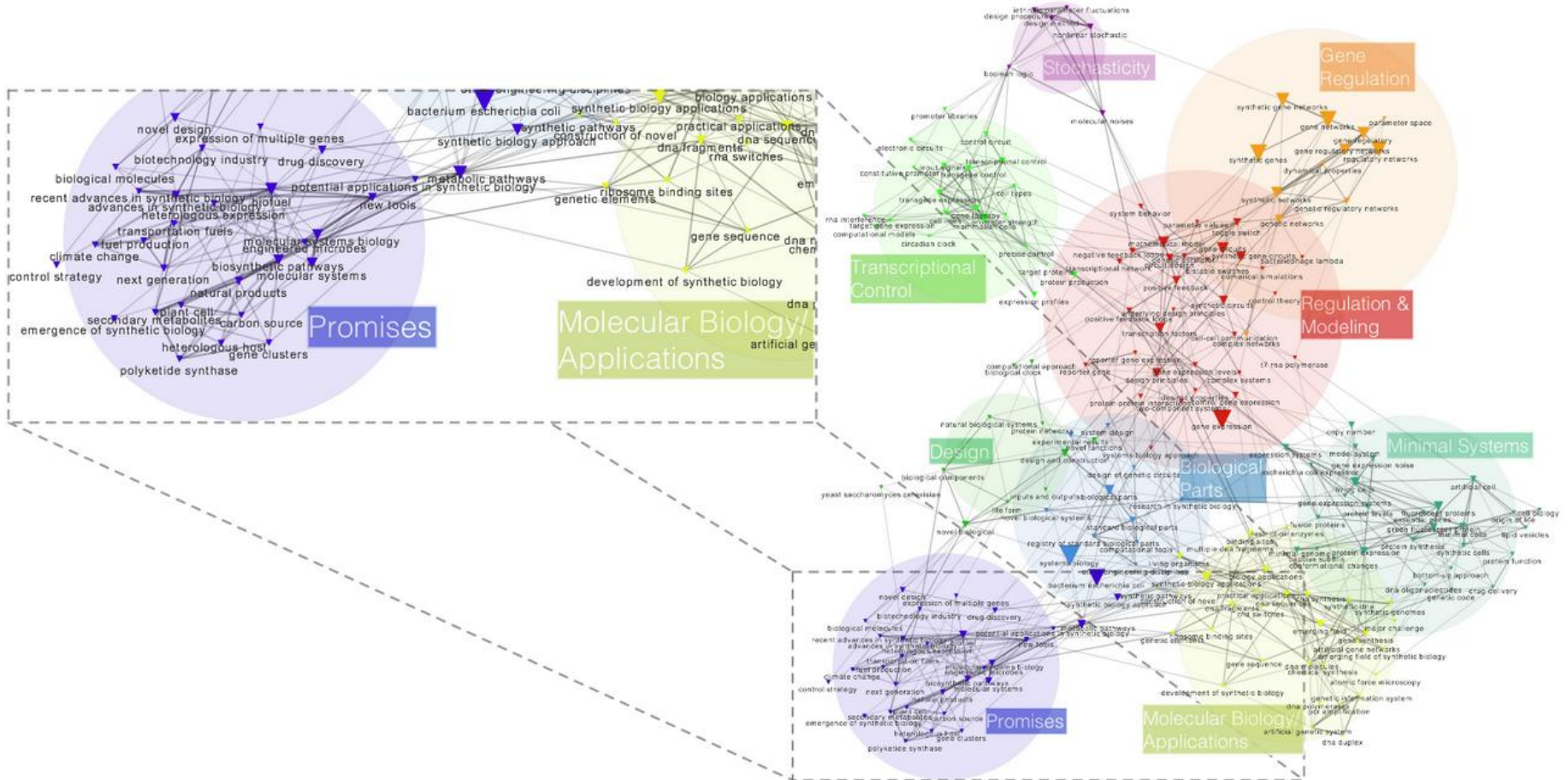
- Рейтинги (Лейден и др.). Методологические споры: можно ли считать рейтинг агрегирующим индикатором?
- «Другие» индикаторы. Эвристики. Hot Science, Emergence и др. Не значение параметра, а сложный паттерн данных
- Отсутствие информации как индикатор.
Примеры из атомной отрасли



Повестка наукометрии

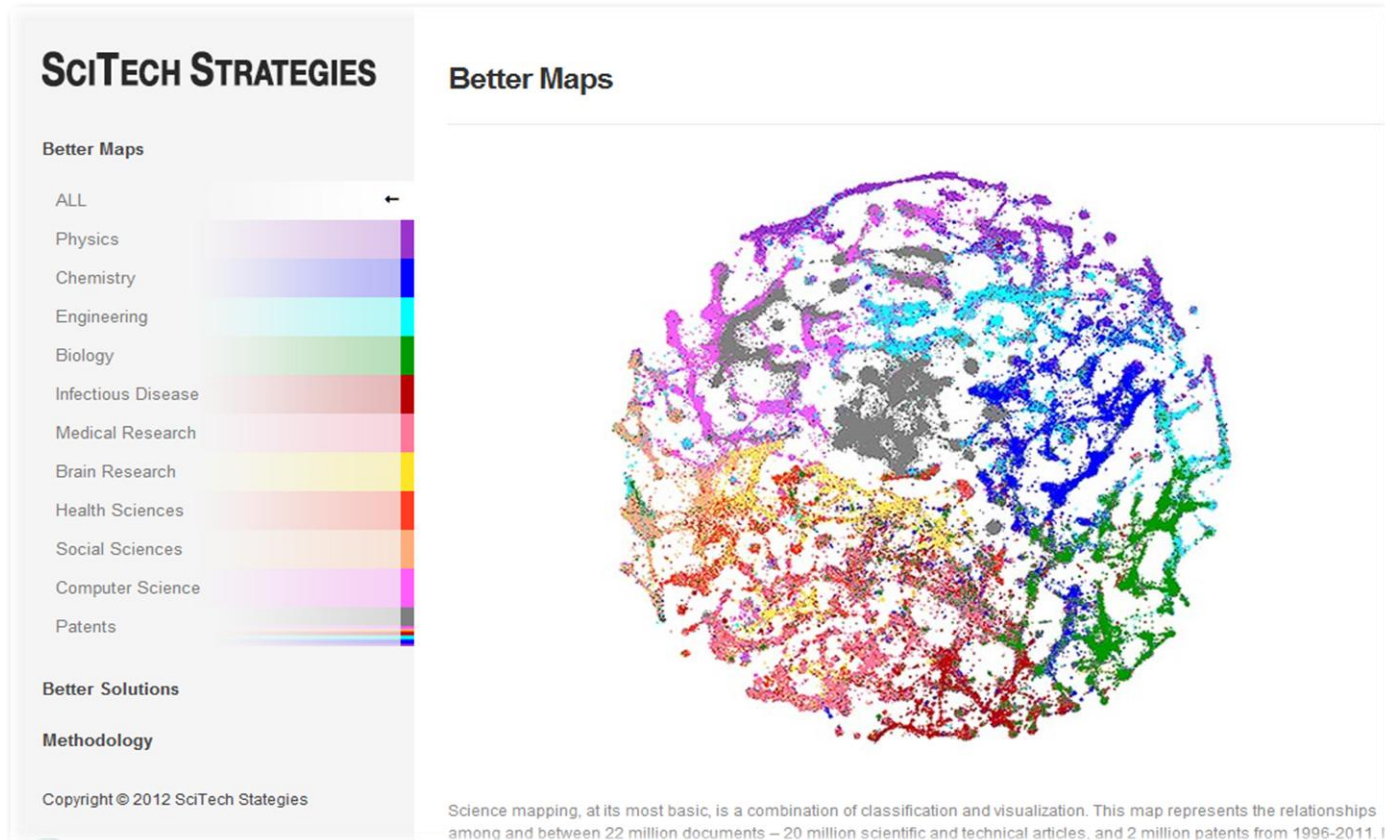
- Аналитика по типам источников (методы и сервисы):
 - ✓ Кросс-цитирование между журналами, книгами, конференциями, патентами, в т.ч. для выработки мер научно-технологической политики и создания новых индикаторов, поиска внедрений и др.
 - ✓ Новые библиометрические сервисы для библиотек, исследование роли библиотек как провайдеров таких сервисов. *Пример ВШЭ*

Некоторые примеры: сети цитирования, коцитирования и не только

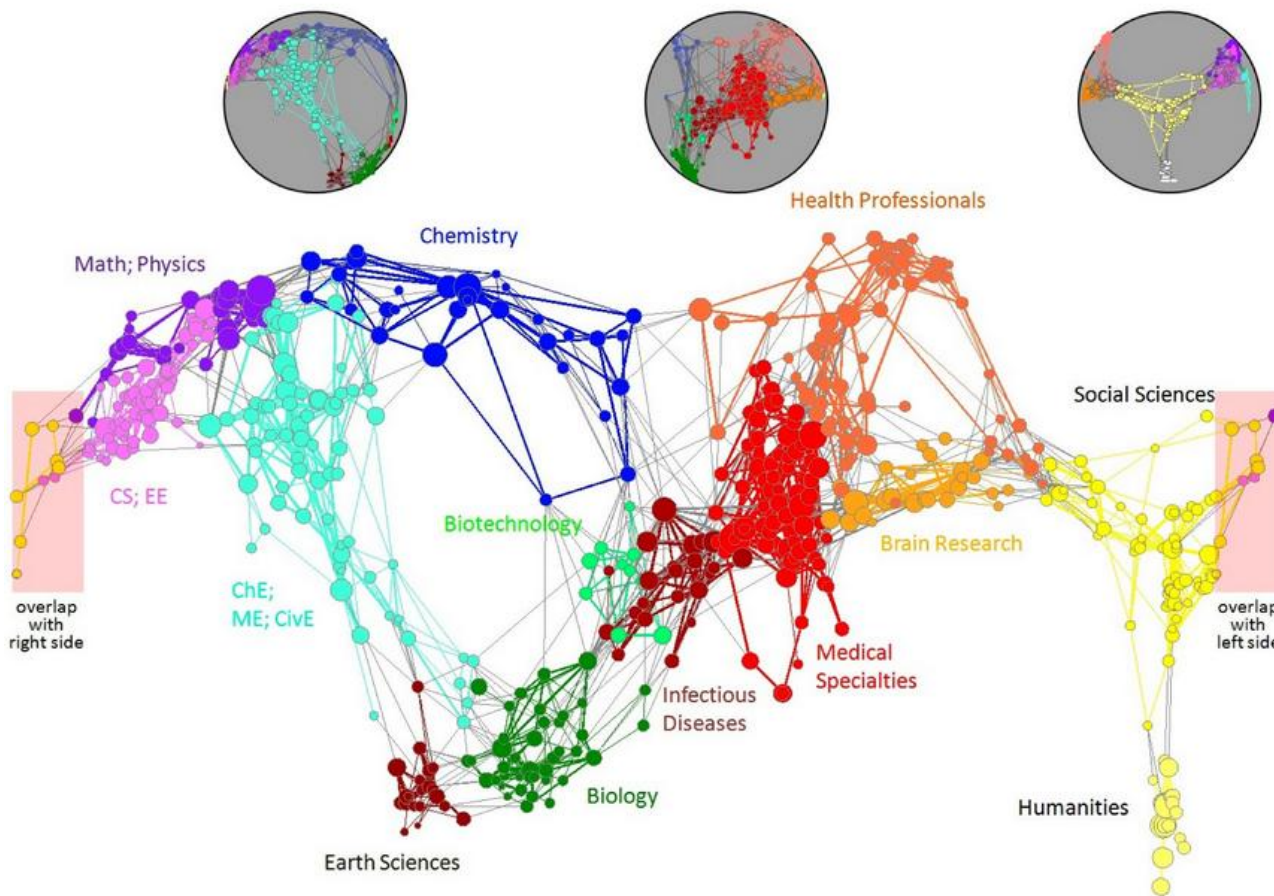


http://digitalmethods-seminar.org/wp-content/uploads/2014/03/2014_SyntheticBio.png

Некоторые примеры: картирование науки

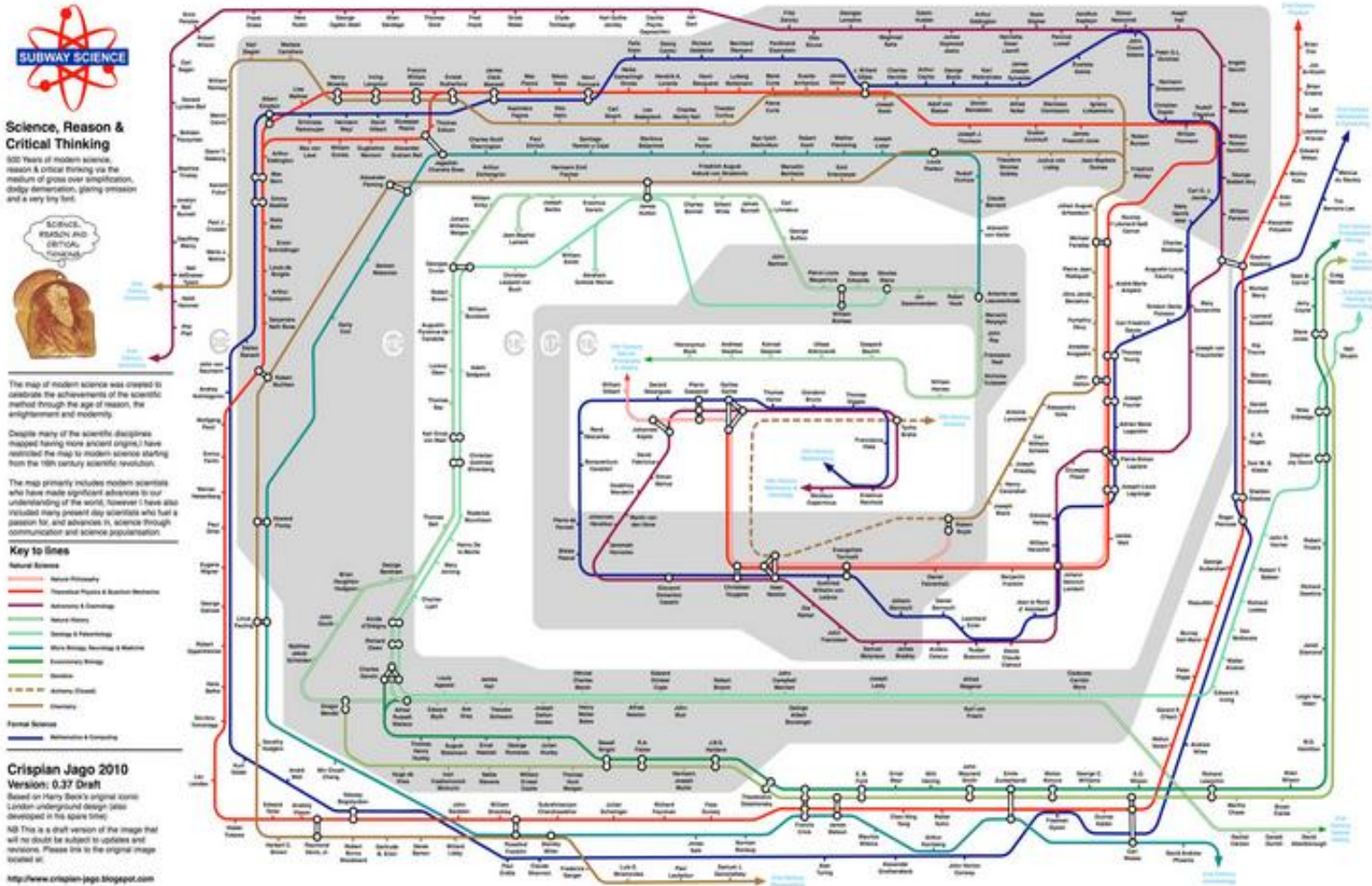


Некоторые примеры: картирование науки



<http://www.soic.indiana.edu/news/story.html?story=Global-Maps-Science>

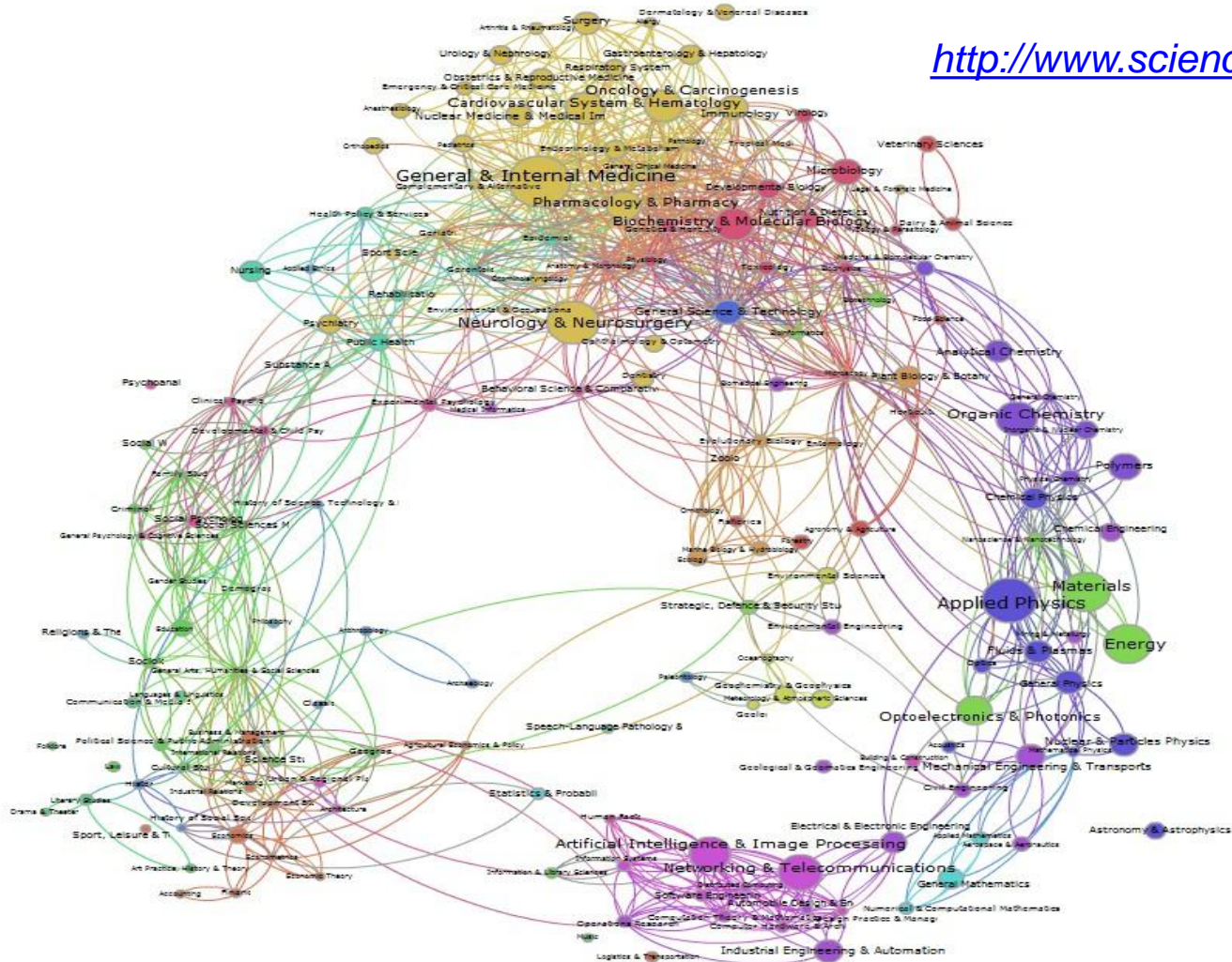
Некоторые примеры: карты связей направлений и ученых



<http://www.crispian.net/CrispiansScienceMap.html>

Некоторые примеры: онтологии и динамика развития

<http://www.science-matrix.com>



Некоторые примеры: РИНЦ

Аналитический инструментарий РИНЦ ScienceIndex

Название организации ?	Период	Публ.
■ Вычислительный центр им. А.А. Дородницына РАН (Москва)	1994-2014	8
■ Институт программных систем им. А.К. Айламазяна РАН (Переславль-Залесский)	1997-1999	2

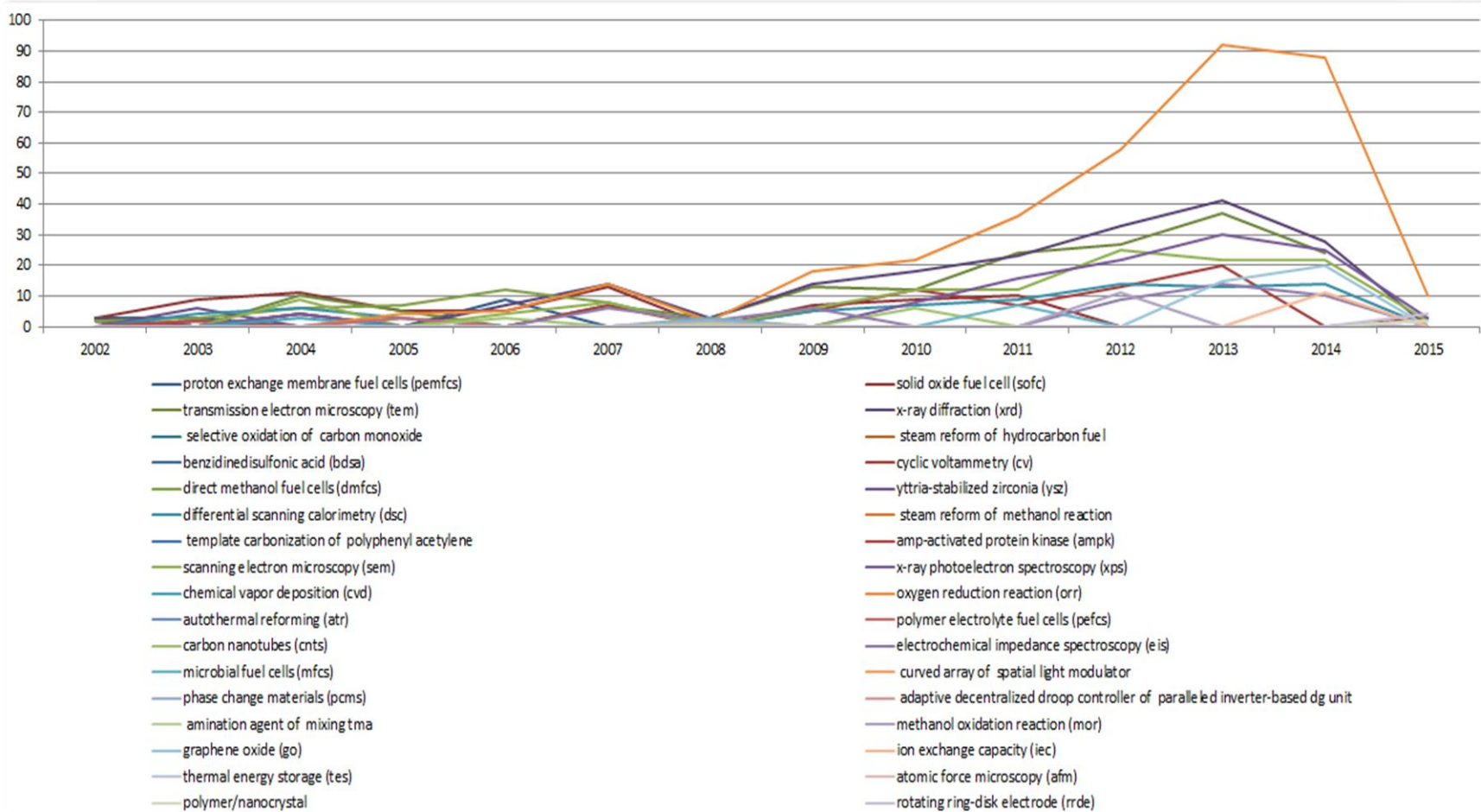
ОБЩИЕ ПОКАЗАТЕЛИ

Название показателя ?	Значение
■ Число публикаций автора в РИНЦ ?	16
■ Число публикаций автора с учетом статей, найденных в списках литературы ?	21
■ Год первой публикации ?	1979
■ Число цитирований публикаций автора в РИНЦ ?	34
■ Число цитирований публикаций автора с учетом статей, найденных в списках литературы ?	49
■ Суммарное число цитирований автора ?	894
■ Число публикаций, процитировавших работы автора ?	816
■ Число ссылок на самую цитируемую публикацию ?	21
■ Индекс Хирша ?	4
■ Индекс Хирша без учета самоцитирований ?	4

РАСПРЕДЕЛЕНИЕ ЦИТИРУЮЩИХ ПУБЛИКАЦИЙ ПО КЛЮЧЕВЫМ СЛОВАМ

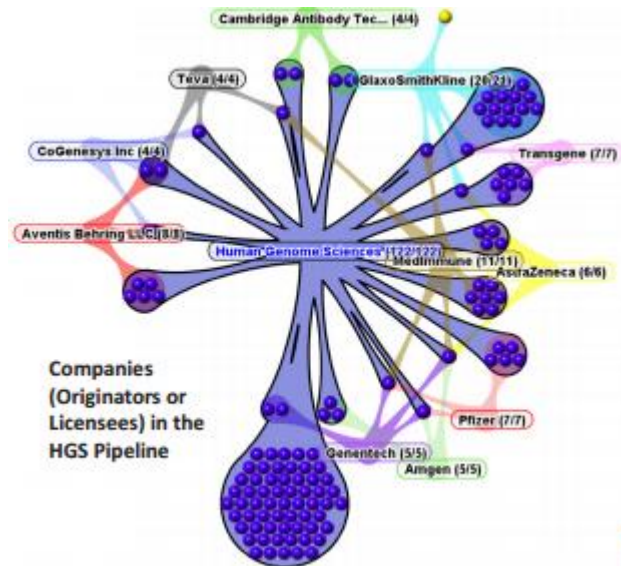
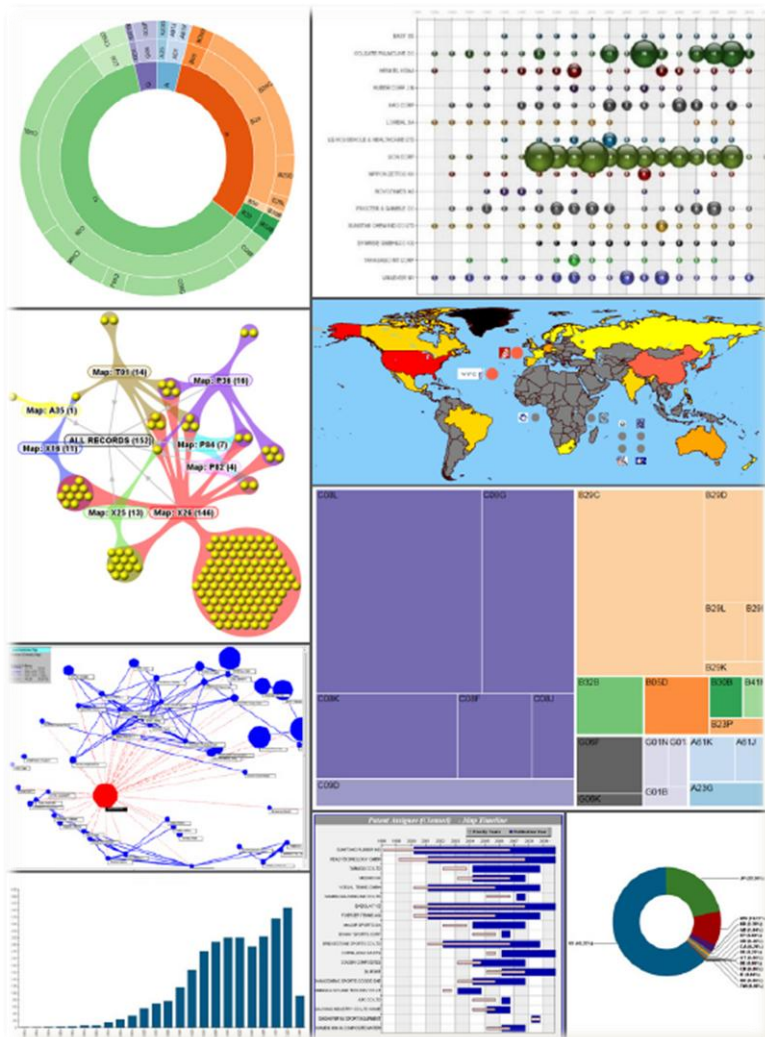
№	Ключевое слово	Публикаций
1.	БАЗА ЗНАНИЙ	64
2.	ОНТОЛОГИЯ	53
3.	KNOWLEDGE BASE	43
4.	ONTOLOGY	40
5.	ЭКСПЕРТНАЯ СИСТЕМА	38
6.	ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ	29
7.	ЭКСПЕРТНЫЕ СИСТЕМЫ	26
8.	EXPERT SYSTEM	23
9.	EXPERT SYSTEMS	19
10.	ЗНАНИЯ	19
11.	ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ	18
12.	ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА	16
13.	ARTIFICIAL INTELLIGENCE	14
14.	KNOWLEDGE	14
15.	БАЗЫ ЗНАНИЙ	14

Некоторые примеры: мониторинг трендов



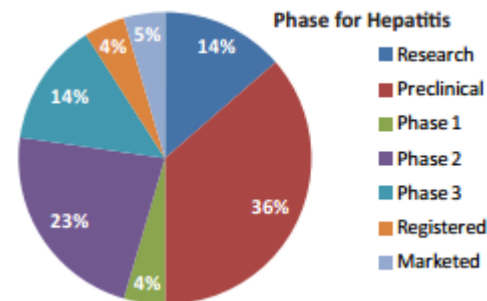
Efimenko, V. Khoroshevsky (в печати)

Некоторые примеры: инструменты BI



Компании

Ландшафт



<https://www.thevantagepoint.com/>



Некоторые примеры: зарождение новых областей

SciTech Stategies, Inc.

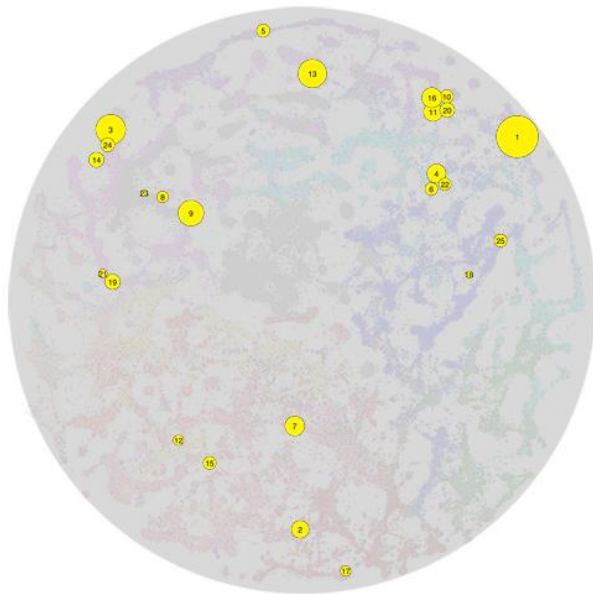
Emergence =
Novelty (or newness) + Growth

Сложный алгоритм, основанный на кластеризации с использованием двух моделей цитирования и коцитирования (*Identifying emerging topics in science and technology. H. Small, K. Boyack, R. Klavans. Research Policy, 2014*).

«Поощрение» и «штрафы»

Некоторые примеры: слабые сигналы и поиск команд

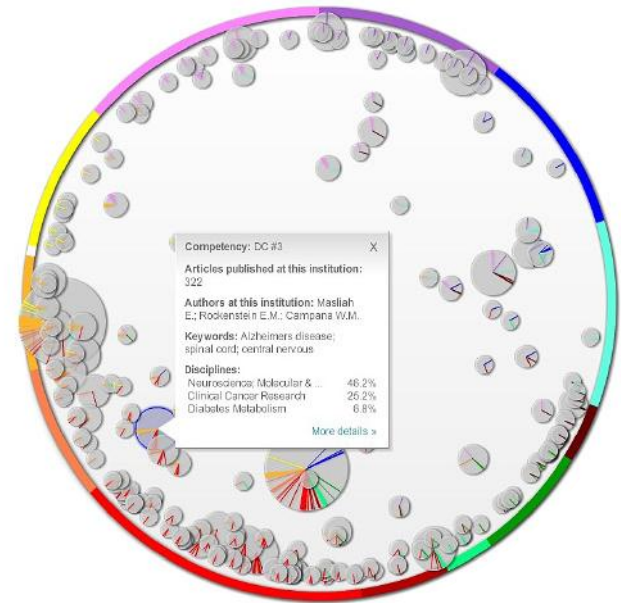
Динамические карты для областей с «прорывным» или «разрушающим» потенциалом



«Слабые сигналы». Джокеры и «Черные лебеди»

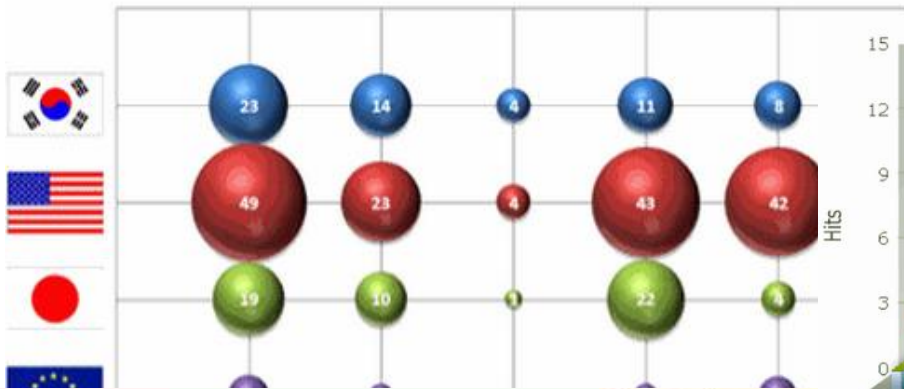


Поиск команд.
Взаимодополнение коллективов.
«Центры компетенции»

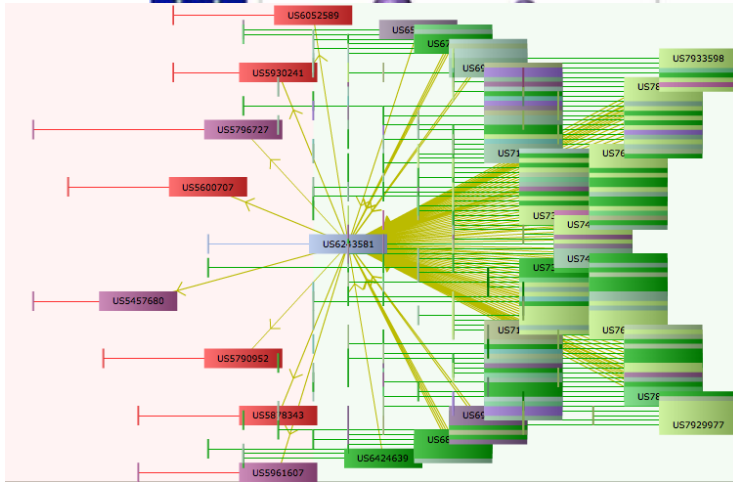
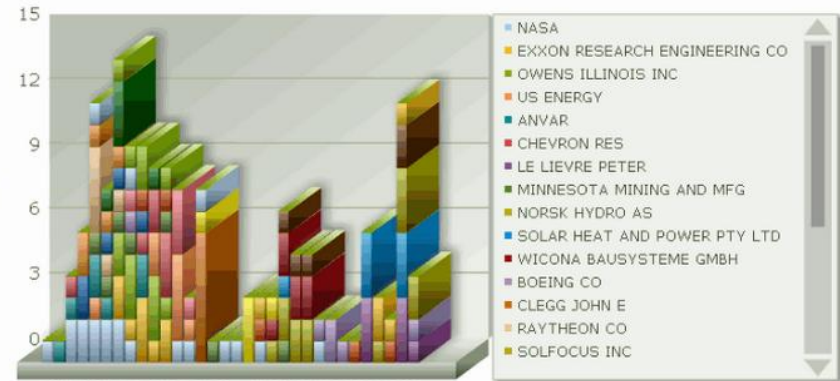


Некоторые примеры: патентная аналитика – широкий спектр возможностей

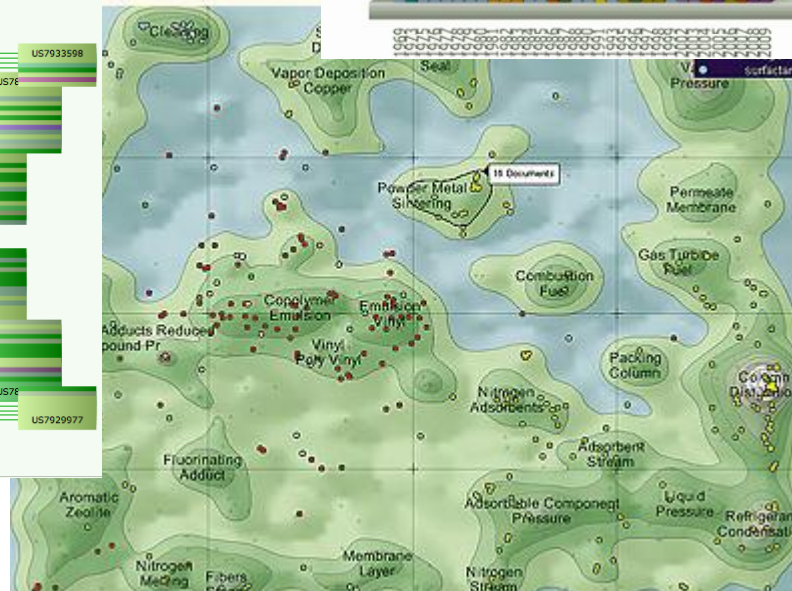
✓ Displaybank



✓ Intellogist



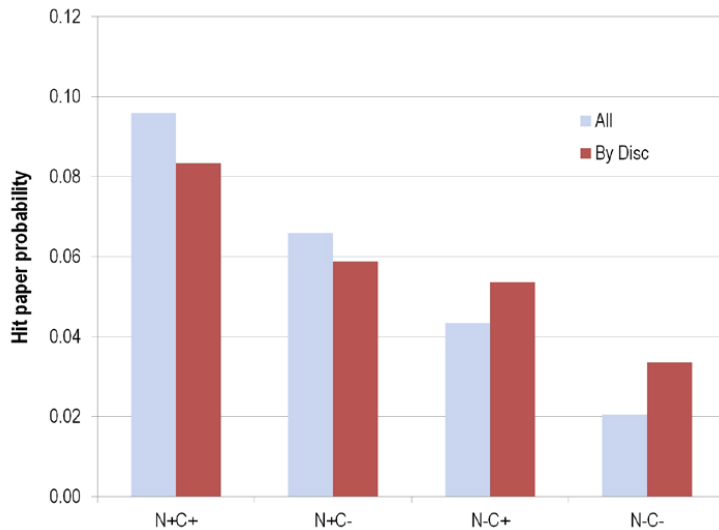
✓ Sumobrain



✓ ThemeScape:
Aureka
✓ WoS и др.

Некоторые примеры: конвенциональность и НОВИЗНА

	UMSJ (1990-2000)		This study (2001-2005)	
	% sample	Prob	% sample	Prob
High Novelty, High Convention (N+C+)	6.7%	0.0911	9.5%	0.0959
High Novelty, Low Convention (N+C-)	26%	0.0533	30.6%	0.0659
Low Novelty, High Convention (N-C+)	44%	0.0582	40.5%	0.0433
Low Novelty, Low Convention (N-C-)	23%	0.0205	19.4%	0.0205



Atypical combinations are confounded by disciplinary effects

Kevin W. Boyack* and Richard Klavans**

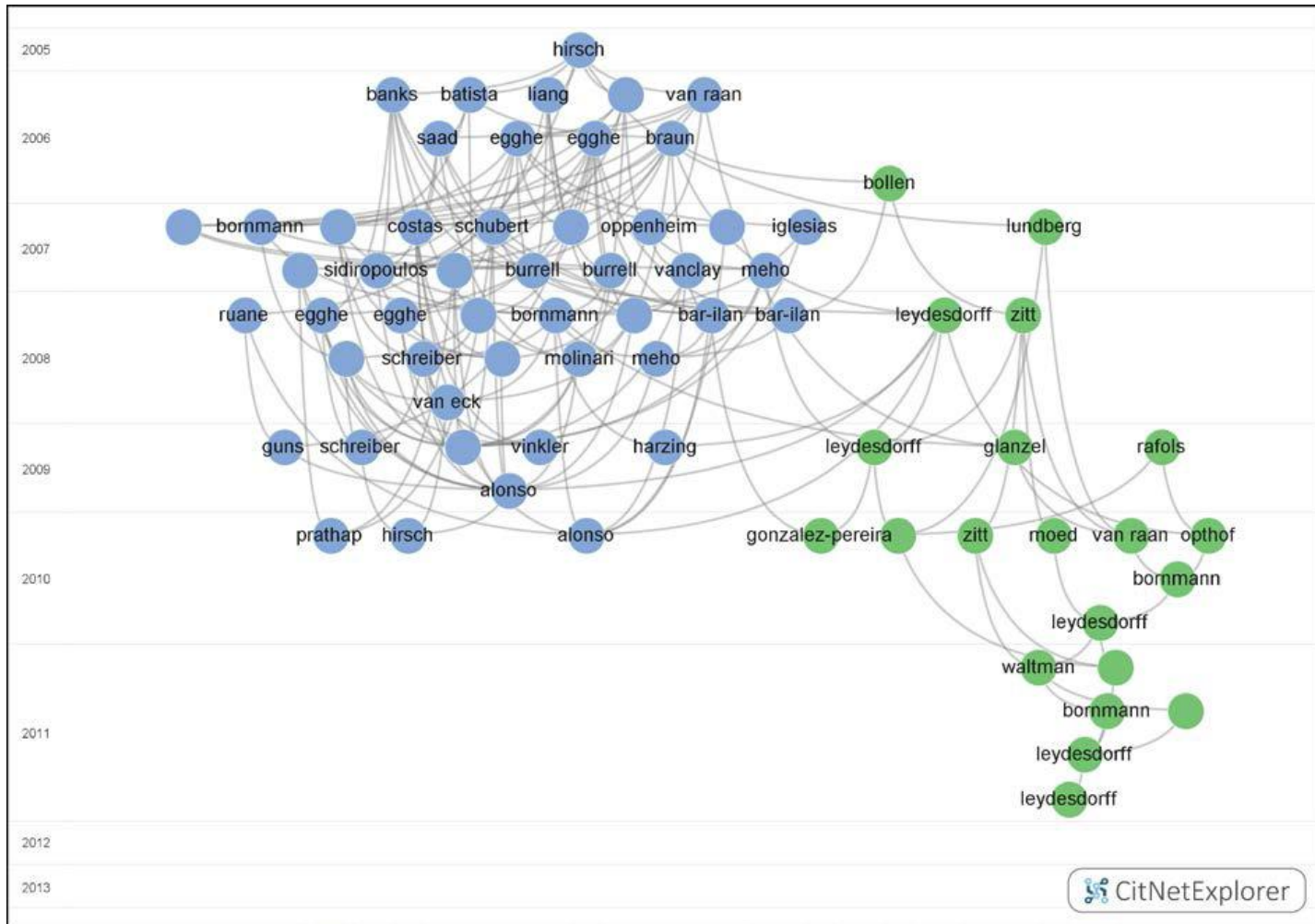
* kboyack@mapofscience.com

SciTech Strategies, Inc., Albuquerque, NM, 87122 (USA)

** rklavans@mapofscience.com

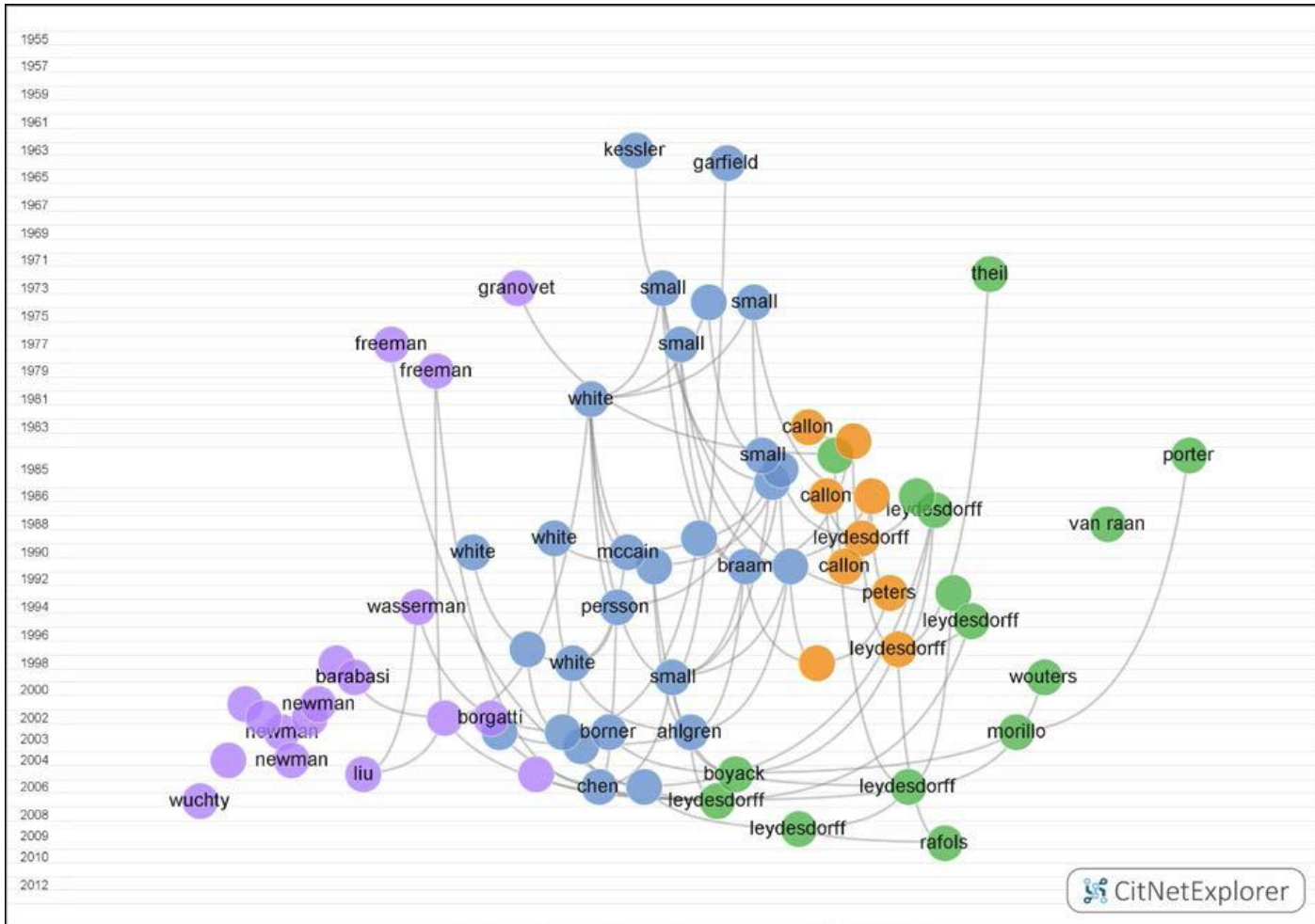
SciTech Strategies, Inc., Berwyn, PA, 19312 (USA)

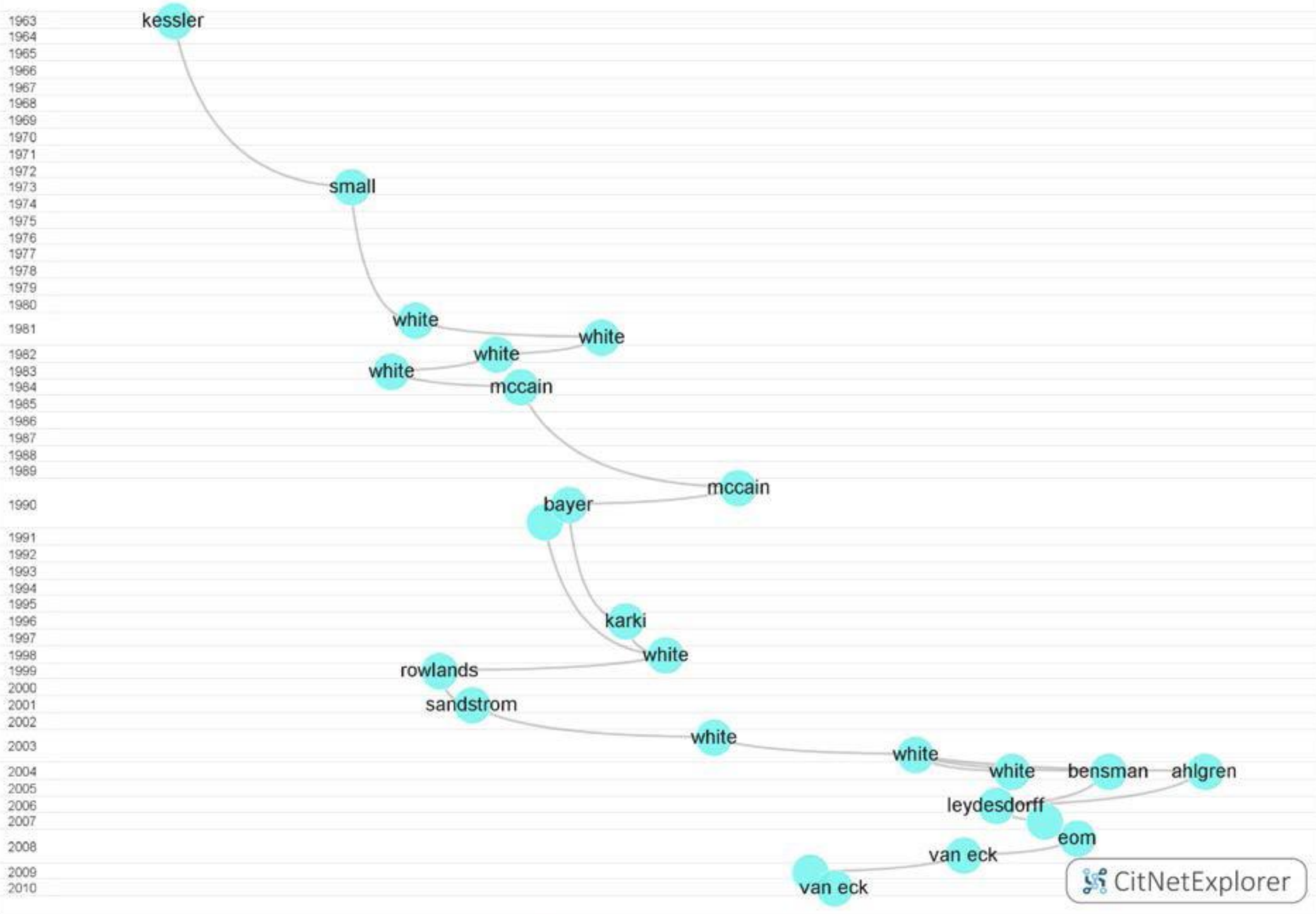
Некоторые примеры: эволюция науки и «личный вклад»



CitNetExplorer

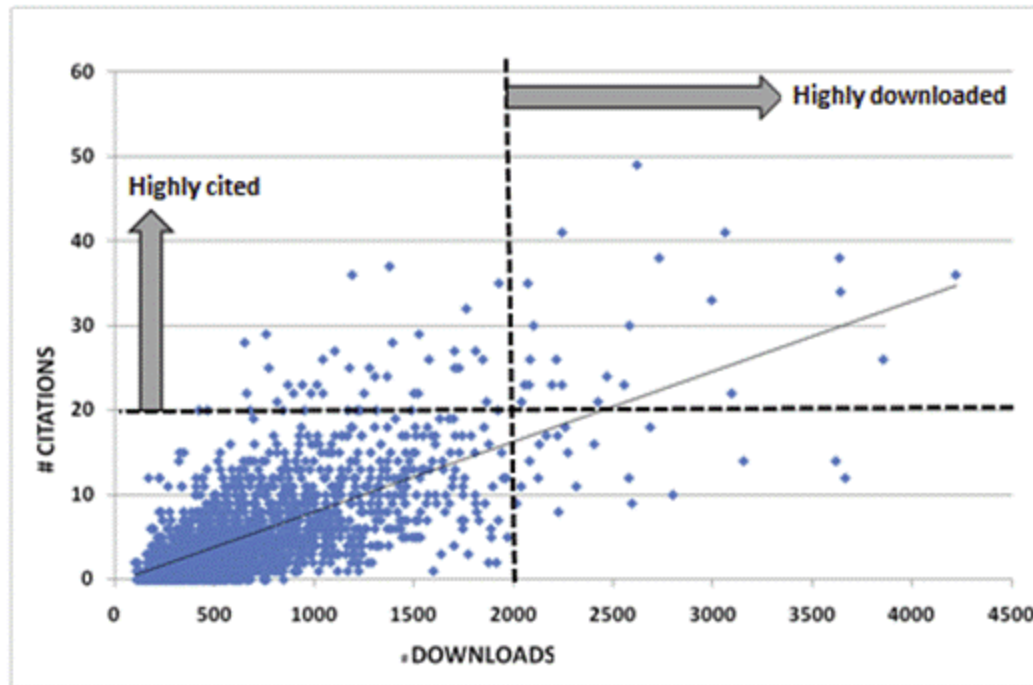
Некоторые примеры: «ядерные» публикации, кратчайший и самый длинный пути...





Некоторые примеры: альтметрика

Труды конференции STI 2014, Нидерланды, Лейден



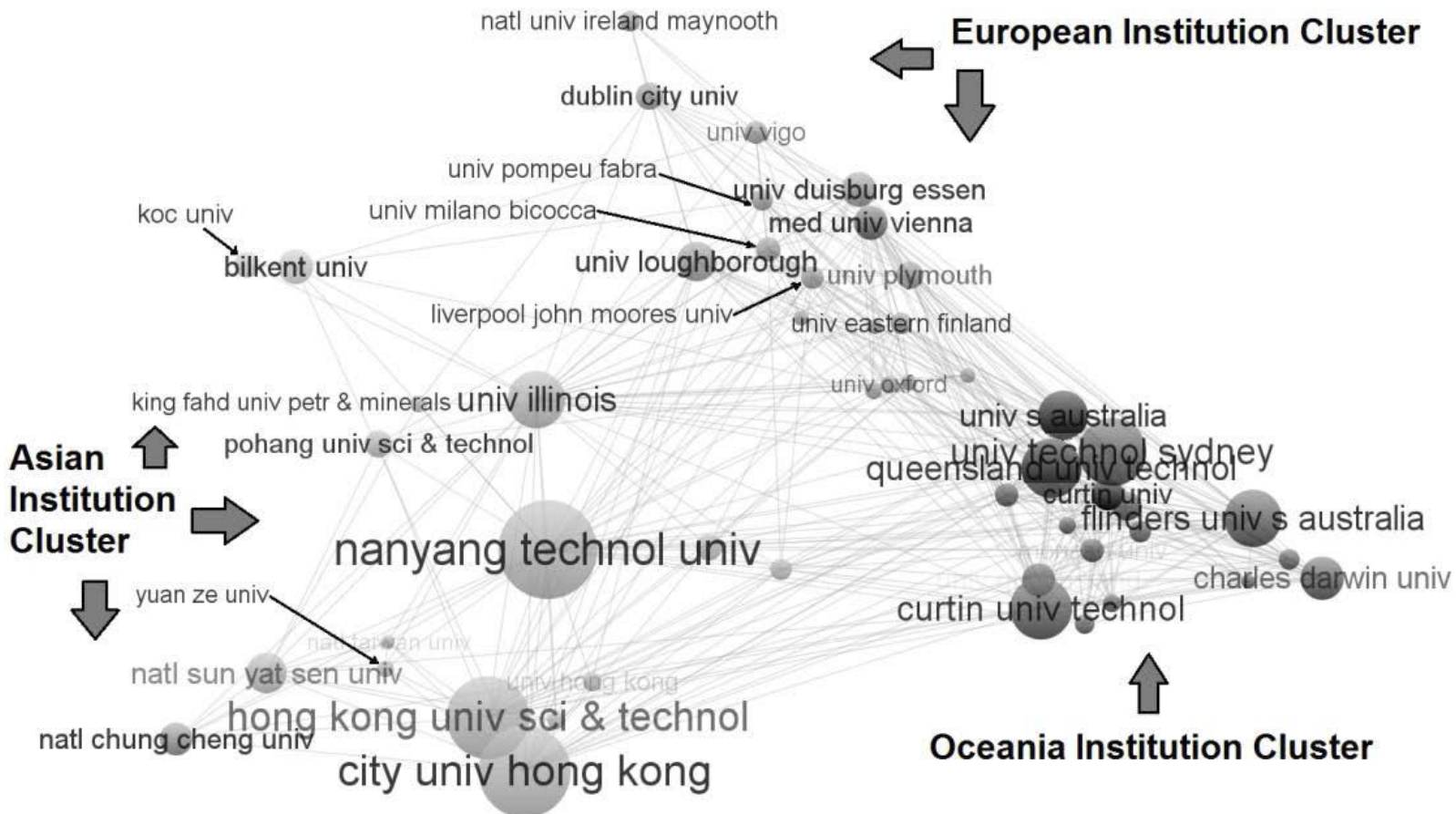
18

Usage patterns of scientific journals and their relationship with citations.

Gali Halevi, Henk F. Moed. 2014

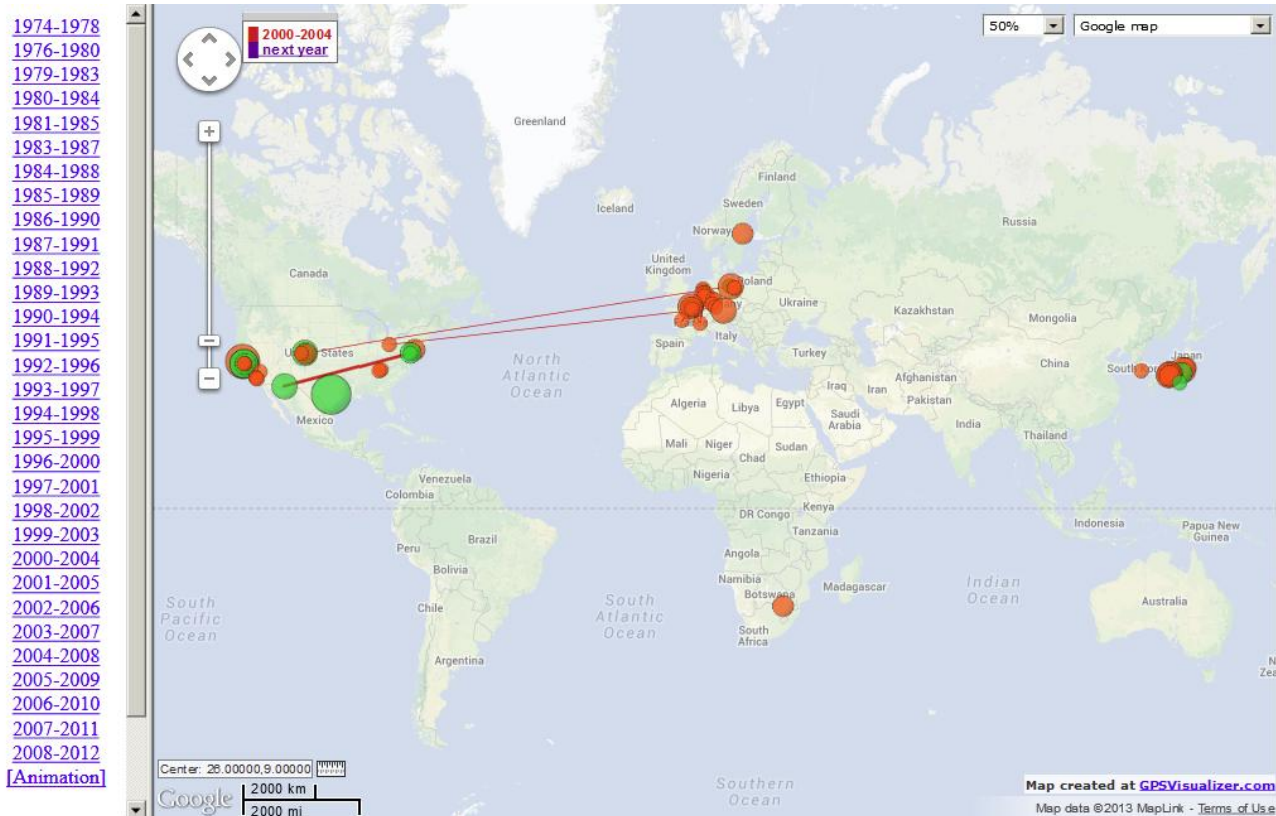
Некоторые примеры: молодые университеты и кооперация

How has International Collaboration been beneficial to Young Universities? A Case Study.
K. A. KHOR and L.-G. YU. 2014



Некоторые примеры: география инноваций

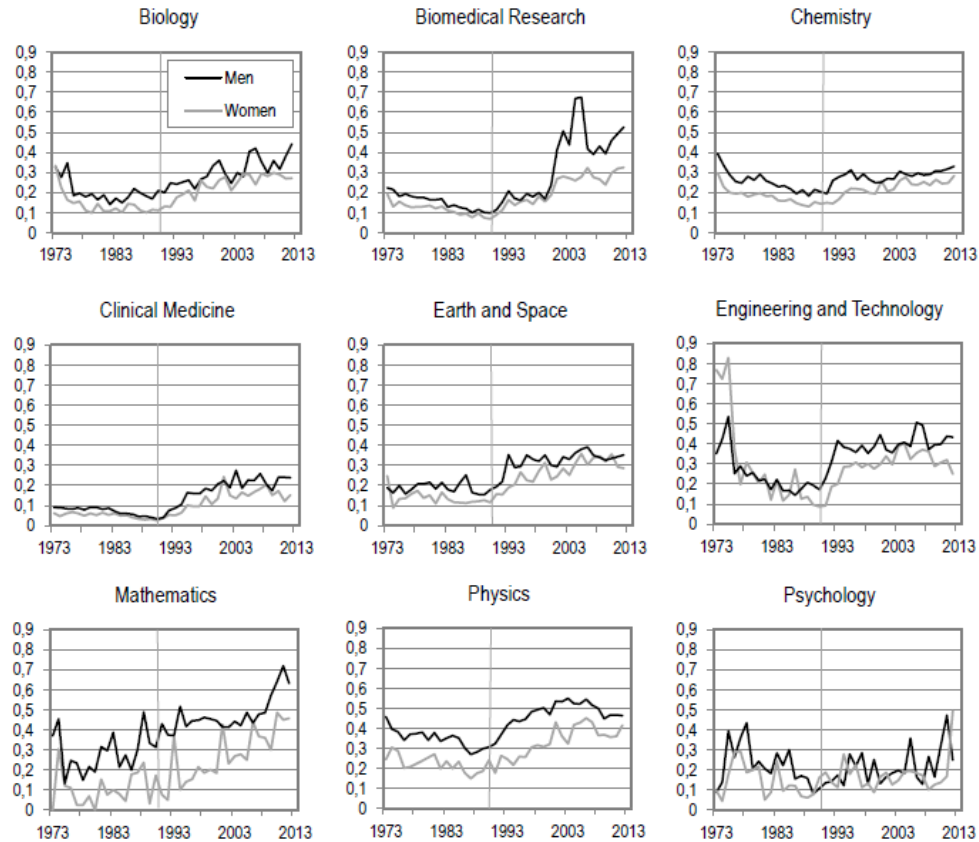
Patents as Instruments for Exploring Innovation Dynamics: Different Perspectives on “Photovoltaic Cells”. Loet Leydesdorff, Floortje Alkemade, Gaston Heimeriks. 2014



Некоторые примеры: женщины в науке

Women and science in Russia: a historical bibliometric analysis.

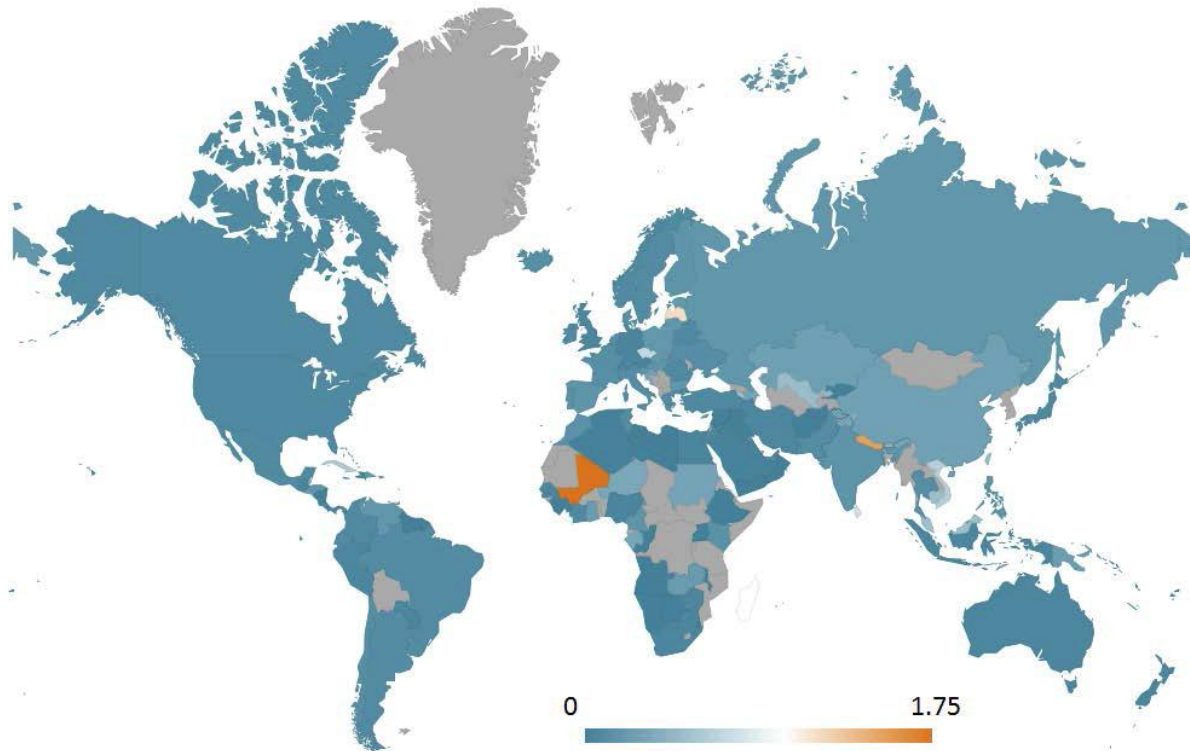
Adèle Paul-Hus,
Rébecca L. Bouvier,
Chaoqun Ni, Cassidy R
Sugimoto, Vladimir
Pislyakov and Vincent
Larivière. 2014



Некоторые примеры: ... и технике

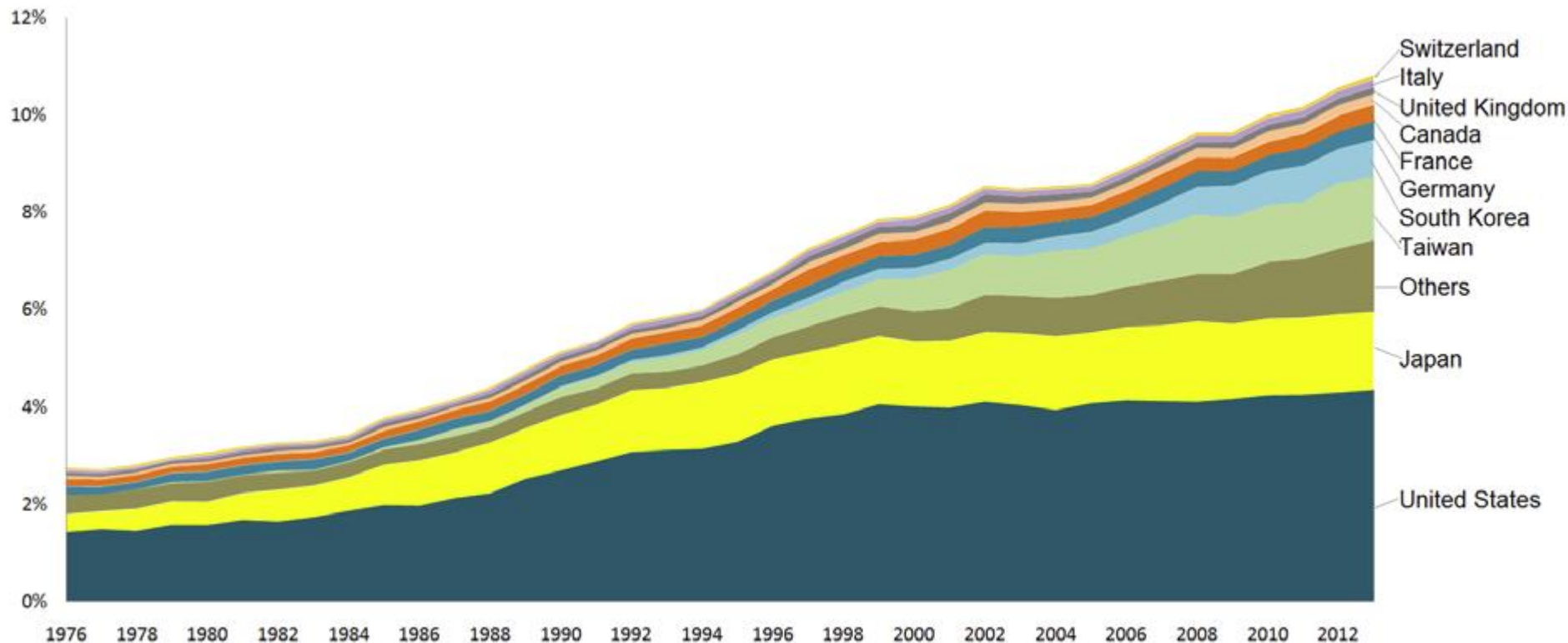
Five countries are female dominated; however, these all have fewer than 35 fractionalized patents (Mali, Nepal, Latvia, Madagascar, and Liberia).

Innovative women: an analysis of global gender disparities in patenting. Cassidy R. Sugimoto, Chaoqun Ni, Jevin D. West, and Vincent Larivière. 2014



Некоторые примеры: женщины в науке

Top 10 countries' (by number of patents) contribution to global female patenting

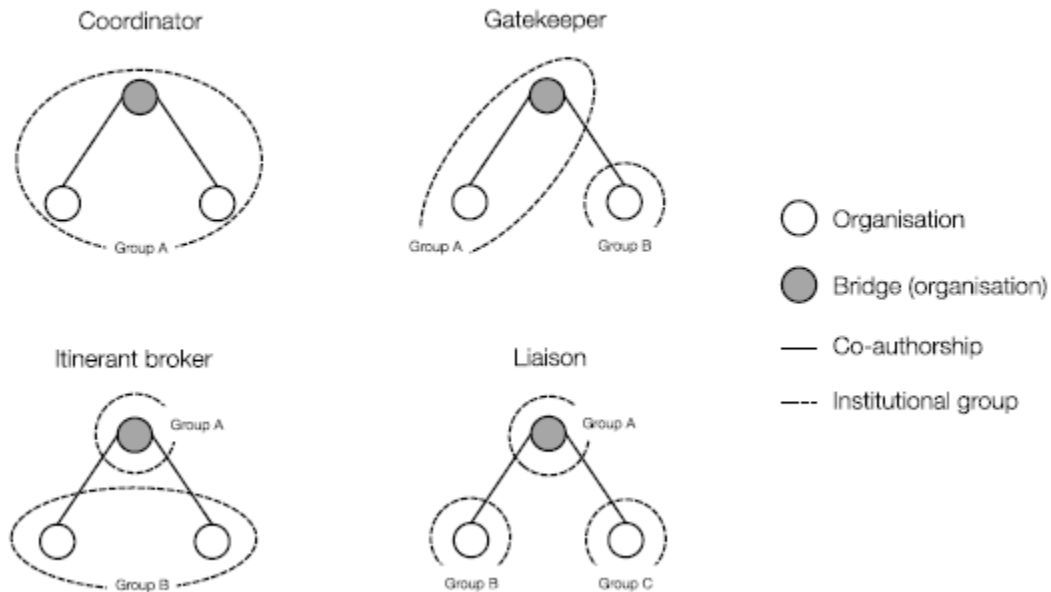


Некоторые примеры: роль институтов в развитии дисциплины

The Emergence of Molecular Biology in the Diagnosis of Cervical Cancer: A Network Perspective.

Daniele Rotolo, Michael Hopkins, Ismael Rafols and Stuart Hogarth. 2014

Figure 1: Institutional groups and bridging roles.





Некоторые примеры: семантический анализ исследовательских фронтов

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ



В. Ф. Хорошевский, И. В. Ефименко,
обработка данных РИНЦ

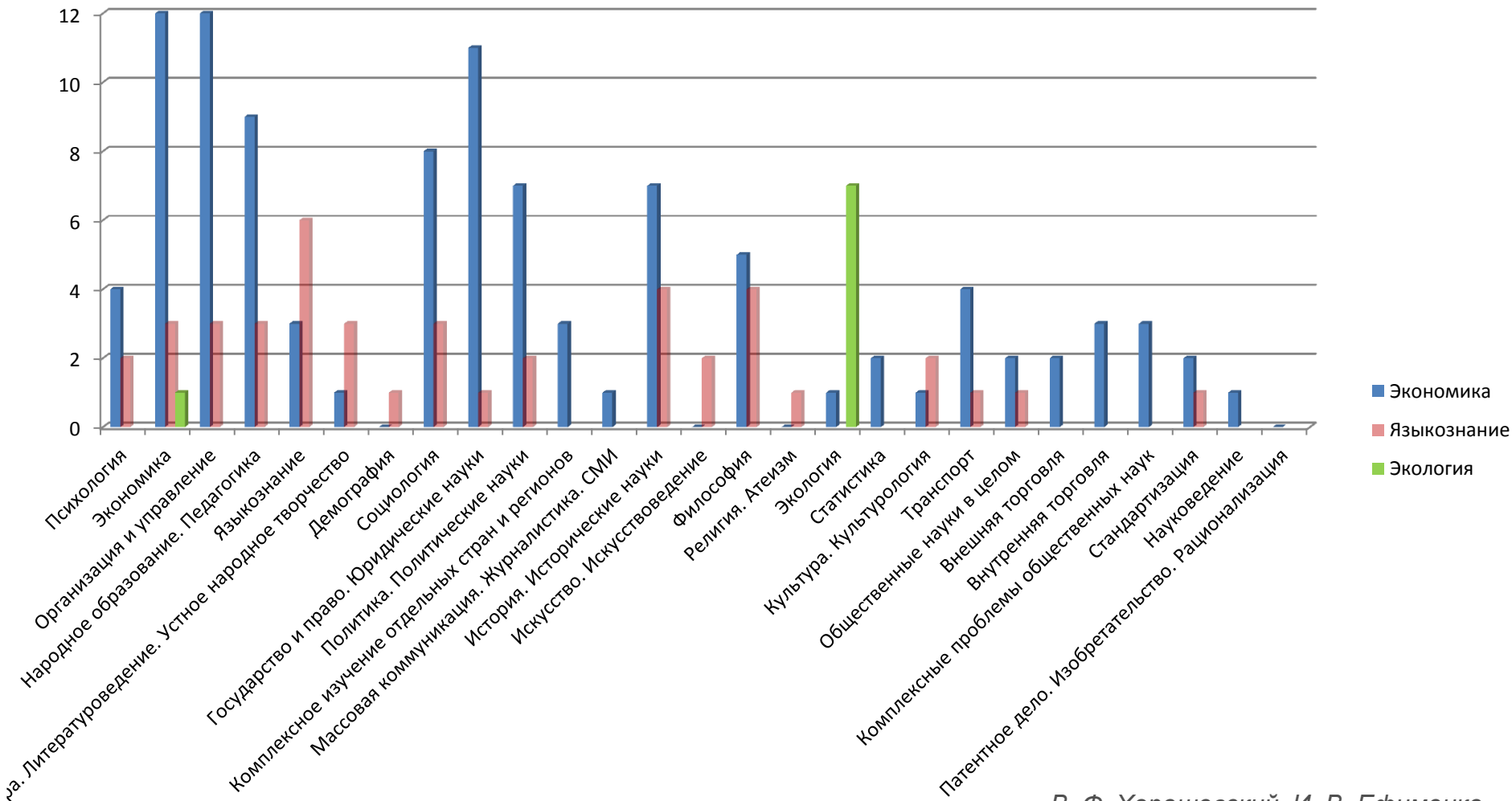
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ «ВЫСШАЯ ШКОЛА ЭКОНОМИКИ» 2014 г.

Некоторые примеры: семантический анализ исследовательских фронтов



В. Ф. Хорошевский, И. В. Ефименко,
обработка данных РИНЦ

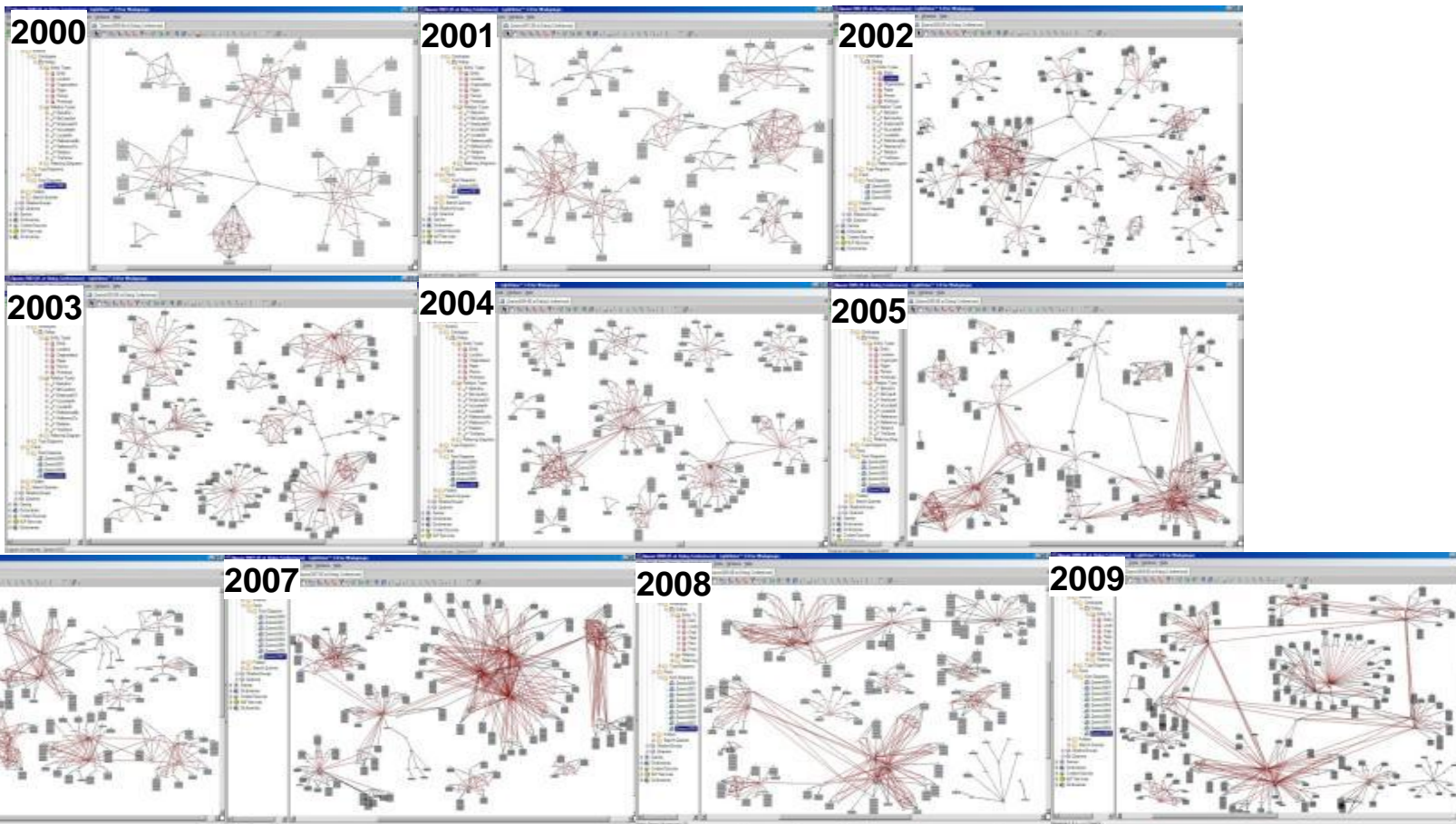
Некоторые примеры: междисциплинарные связи в рамках исследовательских фронтов



В. Ф. Хорошевский, И. В. Ефименко,
обработка данных РИНЦ

Некоторые примеры: «центры превосходства»

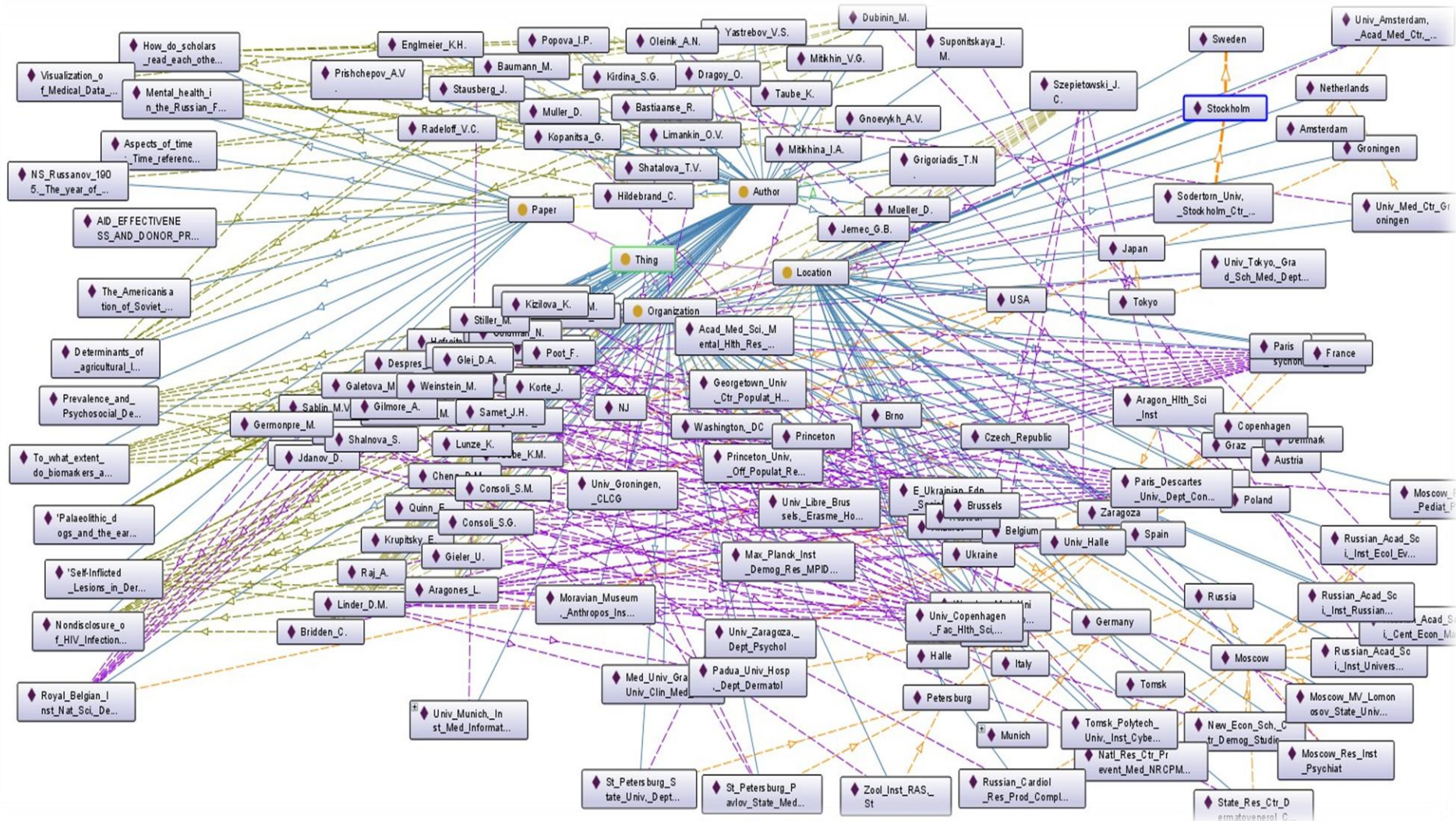
«Коврики Зенкина». Прикладная семиотика. Примеры: инкапсулированность направлений



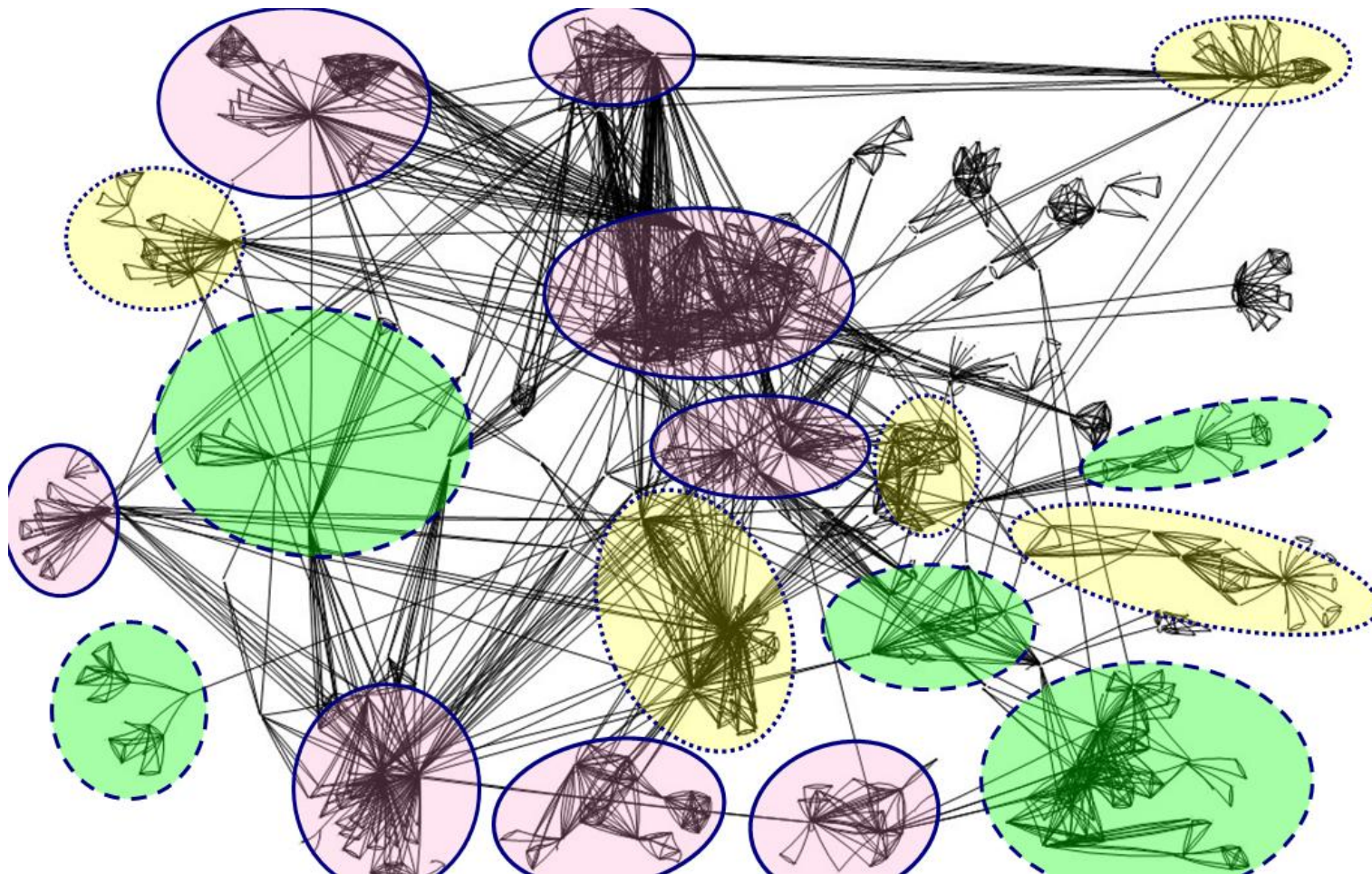
Извлечение информации из текстов на конференциях серии Диалог: взгляд соседа по лестничной клетке.

В. Ф. Хорошевский. Труды международной конференции «Диалог–2010». М.: 2010

Некоторые примеры: «скрытые коллективы»



Некоторые примеры: «центры превосходства»



Извлечение информации из текстов на конференциях серии Диалог: взгляд соседа по лестничной клетке.
В. Ф. Хорошевский. Труды международной конференции «Диалог–2010». М.: 2010



Гуманитарные науки в наукометрии

Статистика (данные РИНЦ):

- Россия, 2008-2012: 1 499 219
- Социальные и гуманитарный науки: 535 837
- процитировано не менее 1 раза: 117 457
- Цитирования: 293 303
- Цитирующие статьи: 171 761
- 3 907 самых высокоцитируемых



Гуманитарные науки в наукометрии

ГРНТИ	Название раздела рубрикатора	Статей	Нецит.	Нецит.,%
00	Общественные науки в целом	2329	1948	83,64%
02	Философия	18375	14761	80,33%
03	История. Исторические науки	35043	28412	81,08%
04	Социология	21586	16491	76,40%
05	Демография	2092	1759	84,08%
06	Экономика. Экономические науки	123135	90507	73,50%
10	Государство и право. Юридические науки	95796	74996	78,29%
11	Политика. Политические науки	18904	14315	75,72%
12	Науковедение	2501	2015	80,57%
13	Культура. Культурология	11855	9981	84,19%
14	Народное образование. Педагогика	69771	54029	77,44%
15	Психология	22555	17143	76,01%
16	Языкознание	31050	26438	85,15%



Гуманитарные науки в наукометрии

ГРНТИ	Название раздела рубрикатора	Цитирований	Самоцит.	Самоцит., %
00	Общественные науки в целом	777	364	46,85%
02	Философия	9565	2665	27,86%
03	История. Исторические науки	13118	5172	39,43%
04	Социология	14100	3048	21,62%
05	Демография	787	244	31,00%
06	Экономика. Экономические науки	90468	24948	27,58%
10	Государство и право. Юридические науки	50257	8526	16,96%
11	Политика. Политические науки	11891	2466	20,74%
12	Науковедение	1201	496	41,30%
13	Культура. Культурология	3697	1257	34,00%
14	Народное образование. Педагогика	38608	11807	30,58%
15	Психология	13141	4797	36,50%
16	Языкознание	9676	2941	30,39%

Гуманитарные науки в наукометрии



Гуманитарные науки в наукометрии

Тренд
(данные WoS)

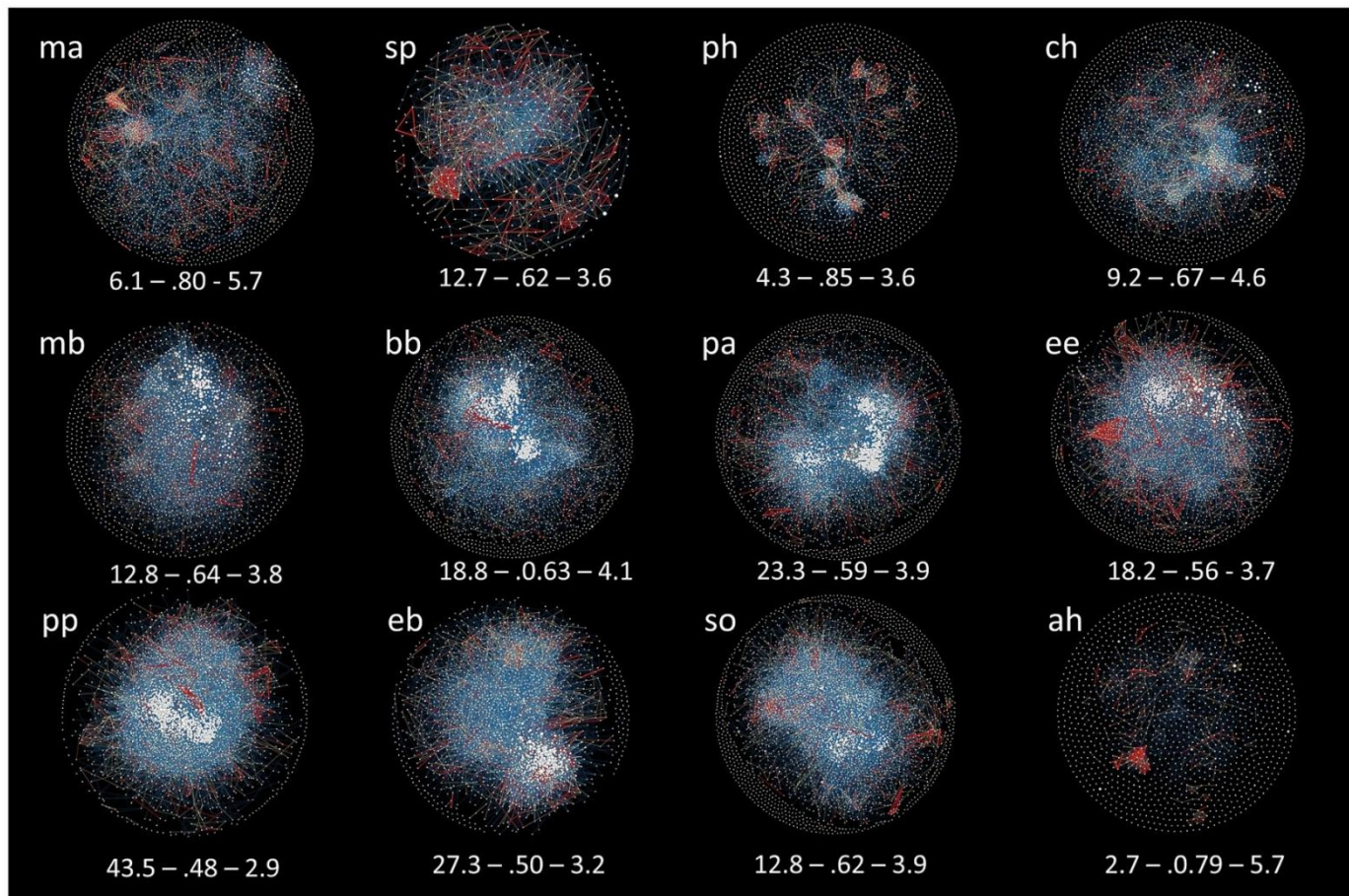
Field: Publication Years	Record Count	% of 211	Bar Chart
2014	32	15.166 %	■
2013	24	11.374 %	■
2011	23	10.900 %	■
2012	21	9.953 %	■
2010	16	7.583 %	■
2009	14	6.635 %	■
2015	13	6.161 %	■
2006	10	4.739 %	■
2008	8	3.791 %	■
2005	7	3.318 %	■
2000	5	2.370 %	■
2007	5	2.370 %	■
1995	4	1.896 %	■
2001	4	1.896 %	■
2004	4	1.896 %	■
1992	3	1.422 %	■
1993	3	1.422 %	■
1999	3	1.422 %	■
1988	2	0.948 %	■
1996	2	0.948 %	■
2002	2	0.948 %	■
2003	2	0.948 %	■
1989	1	0.474 %	■
1991	1	0.474 %	■
1997	1	0.474 %	■
1998	1	0.474 %	■

Гуманитарные науки в наукометрии

Bibliographic coupling networks. Fanelli, Daniele; Glänzel, Wolfgang.
2013

“Number of shared references between any two papers: blue=1; yellow= ≥ 2 ; red ≥ 5 ”.

Длина пути, количество ссылок и др.



Сложное устройство гуманитарных наук.

Примеры:

- *Филология (теоретическая лингвистика, литературоведение, компьютерная лингвистика, нейролингвистика...)*
- *Философия (формальные модели, логика...)*

Особые типы источников: книги, переводы и др.

Междисциплинарность и семантика цитирования

- **“Hopeful monsters” на картах науки**





Публикации автора по теме

- Irina Efimenko, Vladimir Khoroshevsky, Ed Noyons. (Map of Science)²: Fields of S&T in Scientometric and Tech Mining Papers. 2016
- Irina Efimenko, Vladimir Khoroshevsky, Ed Noyons. Anticipating Future Pathways of Science, Technologies & Innovations: (Map of Science)² Approach. 2016
- Vladimir Khoroshevsky, Irina Efimenko. From Mining to Meaning: Identification of Meaningful Patterns in Technology Trend Monitoring. 2016
- Irina V. Efimenko, Vladimir F. Khoroshevsky. Peaks, Slopes, Canyons, Plateaus: Identifying Technology Trends throughout the Life Cycle // International Journal of Innovation and Technology Management. 2015 (forthcoming)
- Efimenko I. V., Khoroshevsky V. F. New Technology Trends Watch: An Approach and Case Study // Lecture Notes in Computer Science. 2014. No. 8722. P. 170-177
- Ефименко И. В., Хорошевский В. Ф. Формальная модель лучших практик построения дорожных карт // В кн.: XIV Национальная конференция по искусственному интеллекту с международным участием КИИ-2014. Казань : Российская ассоциация искусственного интеллекта, 2014. С. 118-127
- Хорошевский В. Ф., Ефименко И. В. Искусственный интеллект: карта научного направления в трудах конференций РАИИ // В кн.: XIV Национальная конференция по искусственному интеллекту с международным участием КИИ-2014. Казань: Российская ассоциация искусственного интеллекта, 2014. С. 160-168
- Ефименко И. В. Гибридный подход к выявлению комплексных объектов в области научно-технического прогнозирования: принцип "черного ящика" // В кн.: Открытые семантические технологии проектирования интеллектуальных систем. Материалы III международной научно-технической конференции. Минск: БГУИР, 2013
- Ефименко И. В. Модели использования и интерпретации оценочной информации в прогнозировании: время, состояние, вероятность // В кн.: XIII Национальная конференция по искусственному интеллекту с международным участием КИИ-2012. Белгород : Российская ассоциация искусственного интеллекта, 2012. Том 1. С. 244-251



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ

And so...



Спасибо за терпение!

iefimenko@hse.ru
veassi@mail.ru