

Вычислительный Центр Российской Академии Наук

# **Анализ данных о поведении пользователей сети Интернет**

К. В. Воронцов, В. А. Лексин, К. В. Рудаков

# Анализ Клиентских Сред

- **Клиентская среда**
  - Клиенты (пользователи)
  - Услуги (ресурсы)
  - Протокол действий пользователей
- **Принцип сходства**
  - Клиенты схожи, если они пользуются схожим набором услуг
  - Услуги схожи, если ими пользуются схожие клиенты.

# Web Mining и Web Usage Mining

- ***Web Content Mining***
  - анализ контента (содержимого документов)
- ***Web Structure Mining***
  - анализ структуры гиперссылок и внутреннего представления документов
- ***Web Usage Mining (WUM)***
  - анализ поведения пользователей

# Исходные данные для WUM

- Поисковые машины
- Счетчики посещений
- Интернет-магазины
- Интернет-провайдеры и прокси-серверы
- Веблоги и форумы

100002171080304956

Валютное регулирование и валютный контроль в РФ 1110371715 281743 0

<http://tandem-forum.ru/seminars/info/seminar39.html> 1110371732

<http://tandem-law.ru/seminars/info/seminar283.html> 110371757

<http://www.nns.ru/krizis/kabinet/2104/krizis8.html>

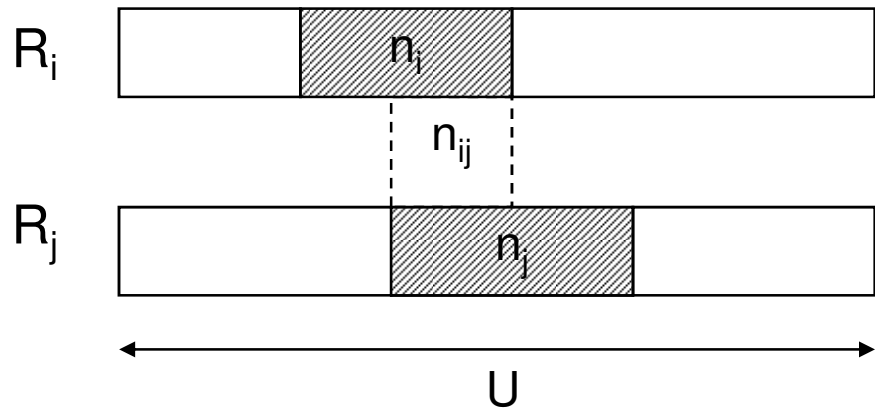
<http://dit.perm.ru/articles/management/data/021217.htm> 1110371805

<http://www.pomosch.com/article.php?sectionId=6&articleId=53>

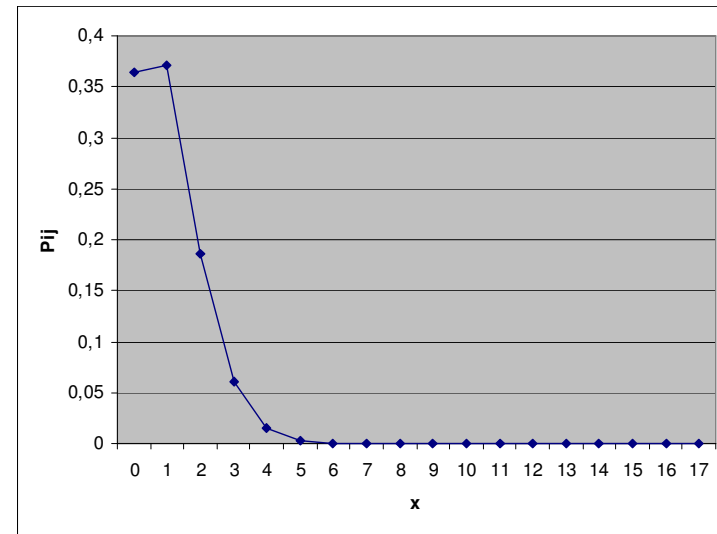
...



# Вычисление оценок сходства ресурсов



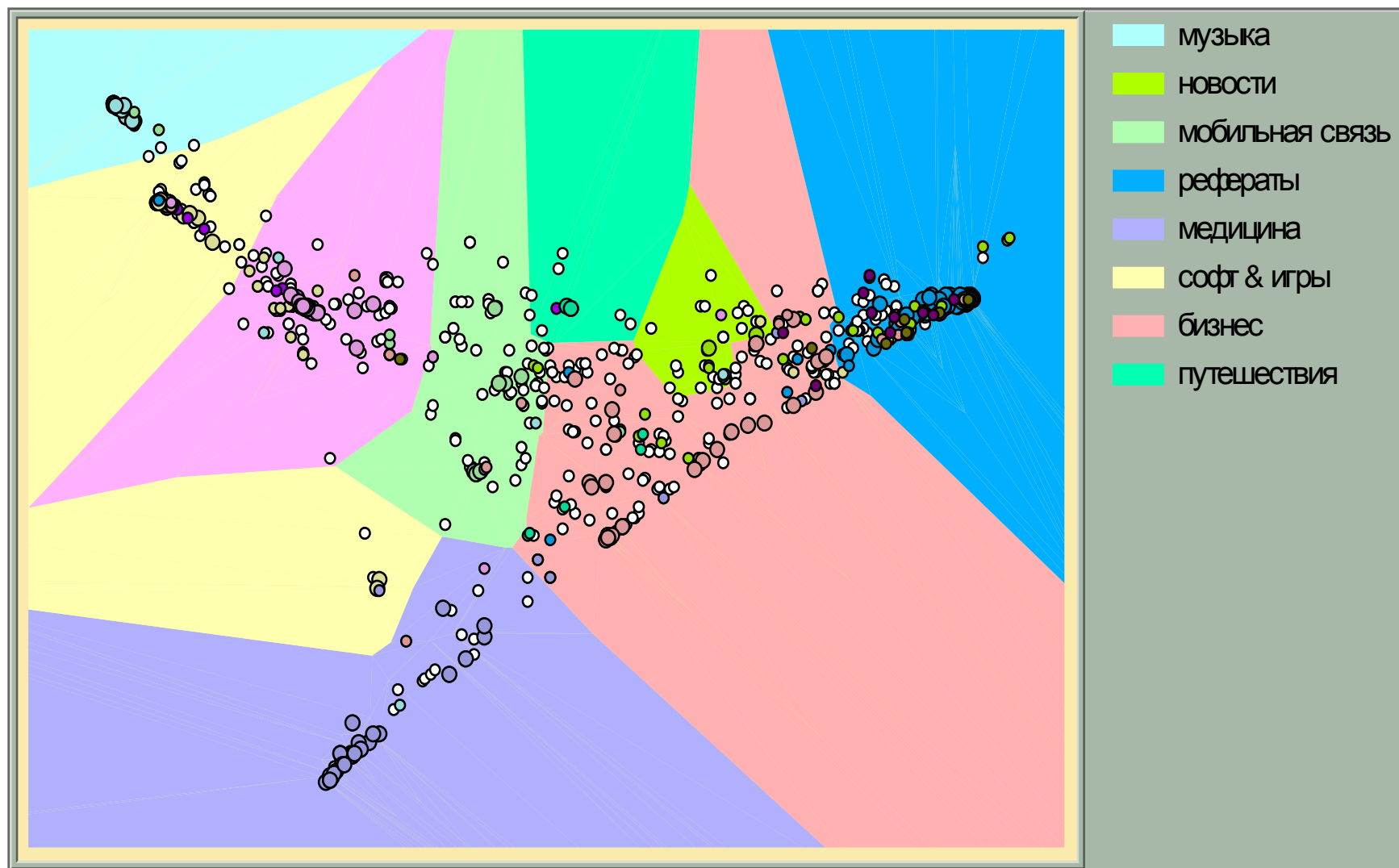
Гипергеометрическое распределение:



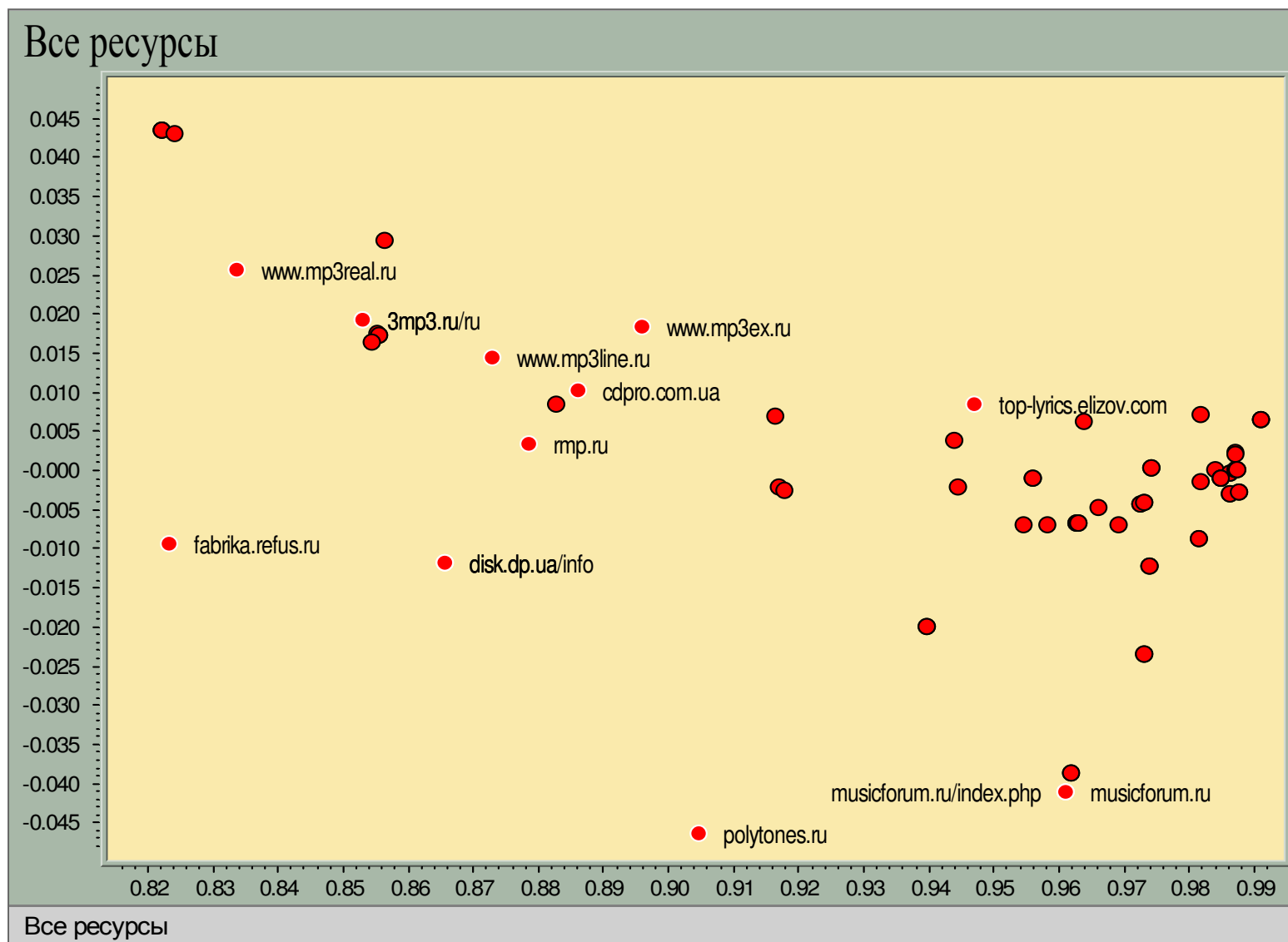
$$P_{ij} = P(n_{ij} = x) = \frac{C_{n_i}^x C_{U-n_i}^{n_j-x}}{C_U^{n_j}}$$

$$\rho(i, j) = \left( \frac{|\ln \alpha|}{|\ln P_{ij}|} \right)^3$$

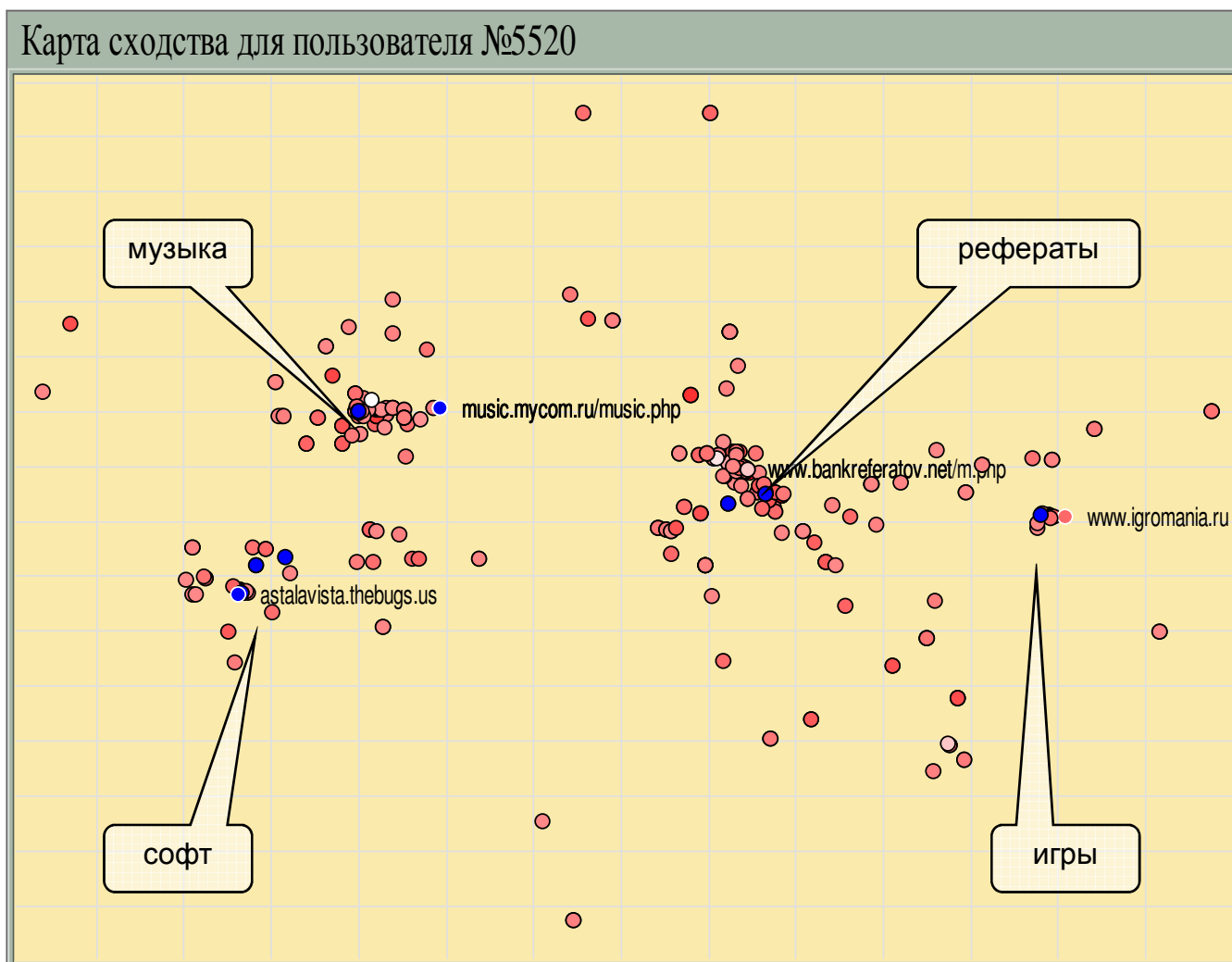
# Карта сходства ресурсов



# Фрагмент карты сходства



# Персональное предложение пользователю





## Практические применения АКС

- Карты сходства – новое средство навигации
  - Персонализация контента
  - Направленное предложение
  - Выявление web-сообществ
  - Маркетинговые исследования
- 
- Технология АКС универсальна:
    - Широкий класс прикладных областей
    - Широкий класс решаемых задач