Министерство науки и высшего образования Российской Федерации

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт прикладной математики и компьютерных наук

КУРСОВАЯ РАБОТА

Выявление факторов влияющих на формирование научных групп и команд ТГУ на основе библиометрического анализа.

Энгельке Сергей Алексеевич

01.03.02 - Прикладная математика и информатика

Математические методы в экономике

Руководитель работы ученая степень, звание

\_\_\_\_\_\_\_\_\_\_\_\_\_\_Т.В Кабанова

подпись

«\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_2024 г.

Автор работы

студент группы № \_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_С.А Энгельке

подпись

«\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_2024 г.

Томск – 2024

Оглавление

[КУРСОВАЯ РАБОТА 1](#_Toc158408048)

[Введение 3](#_Toc158408049)

[Глава 1 Теоретическая часть 3](#_Toc158408050)

[Научные группы и команды: определение, характеристики, особенности. 3](#_Toc158408051)

[Характеристики научных групп и команд: 3](#_Toc158408052)

[Особенности научных групп и команд: 4](#_Toc158408053)

[1.2 Библиометрический анализ: методы и подходы 4](#_Toc158408054)

[1.3 Факторы, влияющие на формирование и разрушение научного взаимодействия 4](#_Toc158408055)

[Глава 2 Методология исследования 5](#_Toc158408056)

[2.1 Объект исследования 5](#_Toc158408057)

[2.2 Методика библиометрического анализа 5](#_Toc158408058)

[2.3 Стохастическое акторно-ориентированное моделирование: суть метода, его применимость в исследовании научных групп и команд 5](#_Toc158408059)

[2.3.1 Допущения ислледования 6](#_Toc158408060)

[2.3.2 Спецификация Акторно ориентированной модели: 7](#_Toc158408061)

[2.3.3 Оценка параметров: 12](#_Toc158408062)

[Глава 3 Анализ результатов 14](#_Toc158408063)

[3.1 Формирование начальных социологических гипотез. 14](#_Toc158408064)

[3.2 Оценка сформированных гипотез. 14](#_Toc158408065)

[3.2.1 Интерпретация результатов 15](#_Toc158408066)

[Заключение 17](#_Toc158408067)

[Список литературы 17](#_Toc158408068)

# Введение

В наше время научная деятельность играет важную роль в развитии общества и науки. Однако, для достижения наилучших результатов, необходимо эффективное взаимодействие между научными группами и командами. В связи с этим, вопрос выявления научных групп и команд, а также факторов, влияющих на формирование и разрушение научного взаимодействия, становится все более актуальным.

Целью данной курсовой работы является выявление факторов влияющих на формирование научных групп и команд Томского Государственного Университета на основе библиометрического анализа, а также определение факторов, влияющих на формирование и разрушение научного взаимодействия. Для достижения поставленной цели используется стохастическое акторно-ориентированное моделирование (САОМ).

Стохастическое акторно-ориентированное моделирование используются для изучения динамики социальных сетей. Класс моделей САОМ объединяет теорию полезности и марковские процессы. САОМ применялся для анализа динамики социальных сетей в исследованиях образования [1], для анализа экономических взаимодействий между географическими акторами [2] и многих других исследованиях социальных сетей.

Было построено несколько временных снимков сети соавторства Томского Государственного Университета на основании данных с информационного портала ТГУ.Сотрудники. Похожее исследование динамики научного сотрудничества с с ипользованием САОМ проводилось для национальной научной системы Словении [3].Результаты такого рода исследования могут быть использованы для оптимизации научного взаимодействия в рамках Томского Государственного Университета и дальнейшего развития научной деятельности в целом.

# Глава 1 Теоретическая часть

## Научные группы и команды: определение, характеристики, особенности.

Научные группы и команды - это организованные сообщества ученых, сотрудничающих для достижения определенных научных целей. Научные группы и команды образуются на базе научных институтов, университетов, научно-исследовательских центров или лабораторий.

### Характеристики научных групп и команд:

* Члены научной группы или команды имеют общие научные интересы и цели;
* Члены научной группы или команды имеют определенные роли и задачи;
* Члены научной группы или команды сотрудничают для выполнения научных исследований, в том числе и научных публикаций;
* Научные группы и команды могут иметь различные размеры и структуры;
* Научные группы и команды могут работать в рамках одной научной области или пересекаться с другими научными областями.

### Особенности научных групп и команд:

* Результаты работы научных групп и команд могут быть использованы для разработки новых технологий и продуктов;
* Научные группы и команды могут привлекать внешних экспертов для решения конкретных задач;
* Научные группы и команды могут иметь различные формы финансирования, включая государственное и частное финансирование.

## 1.2 Библиометрический анализ: методы и подходы

Библиометрический анализ является методом исследования научной литературы и состоит в количественном анализе научных публикаций, цитирований и авторских коллективов с целью выявления закономерностей и тенденций в различных научных областях.

В рамках данной курсовой работы я будут рассматриваться такие методы и подходы библиотемтричского анализа:

* Анализ сетевых структур. Этот подход базируется на анализе сетевых структур, создаваемых научными сообществами в процессе обмена информацией и знаниями. Он позволяет выявить ключевых игроков и их роли в сообществе, а также группы, которые наиболее тесно связаны между собой.

## 1.3 Факторы, влияющие на формирование и разрушение научного взаимодействия

Научное взаимодействие представляет собой процесс сотрудничества между учеными с целью достижения общей научной цели. Однако, как и любой другой процесс, научное взаимодействие может изменяться под влиянием различных факторов, которые могут способствовать или, наоборот, препятствовать его формированию и развитию.

Основными факторами, влияющими на формирование и разрушение научного взаимодействия, которые можно выделить в рамках взаимодействия в Томском Государственном Университете:

* Общая научная тематика: Ученые, работающие в одной области науки, имеют больше возможностей для сотрудничества, чем те, кто занимаются разными научными направлениями.
* Финансирование: Наличие достаточных финансовых ресурсов может способствовать формированию научных групп и команд. Однако, в случае недостатка финансирования, ученые могут ощутить трудности в формировании и поддержании научных контактов.
* Личностные факторы: Личностные характеристики ученых, такие как общительность, коммуникабельность и т.д., могут оказывать существенное влияние на формирование и поддержание научных связей.
* Технологический прогресс: Технологический прогресс может предоставить новые возможности для научного сотрудничества.
* Наличие научной инфраструктуры: Наличие необходимой инфраструктуры, такой как лаборатории, библиотеки, научные журналы, тоже является важным фактором для формирования научных групп и команд.

# Глава 2 Методология исследования

## 2.1 Объект исследования

Объектом исследования данной курсовой работы являются научные группы и команды Томского Государственного Университета, а конкретно данные, собранные о сотрудниках ТГУ, их научных публикациях,подразделении в котором они работают и сетевых связях. Используя библиометрический анализ и стохастическое акторно-ориентированное моделирование, можно выявить факторы, влияющие на формирование и разрушение научного взаимодействия внутри этих групп и команд.

## 2.2 Методика библиометрического анализа

Для исслeдования использованы сетевые структуры и метрики. Данные выгруженны с веб сайта ТГУ.Сотрудники. Ипользуемая информация о авторах: Опыт работы, подразделение в ТГУ и список публикаций автора, образующий сеть соавторства. Исходный код парсера данных (программы по выгрузке и систематизации данных), а так же данные, которые были использованы для исследования доступны по ссылке.

## 2.3 Стохастическое акторно-ориентированное моделирование: суть метода, его применимость в исследовании научных групп и команд

Стохастическое акторно-ориентированное моделирование [4] (Stochastic Actor-Oriented Modeling, SAOM) - это метод социально-сетевого анализа, который позволяет моделировать и предсказывать эволюцию социальных сетей на основе поведения и взаимодействий индивидуальных акторов в них. SAOM используется для изучения процессов формирования и разрушения социальных связей в группах и командах, а также для оценки влияния различных факторов на эти процессы.

Применительно к исследованию научных групп и команд, SAOM может быть использован для моделирования процессов формирования научных связей между учеными в рамках группы или команды.

Используя SAOM возможно выявить такие факторы, влияющие на формирование или разрушение связей, как: влияние стажа на то, с какой частотой актор будет принимать решения относительно изменения сети, с какой частотой изменяется сеть, как влияет научное подразделение актора на динамику сети, и т.д.

Преимущества SAOM включают возможность моделирования процессов на уровне индивидуальных акторов, учет динамической природы социальных связей, а также возможность оценки влияния различных факторов на эти процессы. Однако SAOM также имеет некоторые ограничения, такие как необходимость большого количества данных для моделирования и ограниченная возможность предсказания долгосрочных эффектов.

Таким образом, SAOM является мощным инструментом для исследования научных групп и команд, который может помочь в выявлении факторов, влияющих на формирование научных связей и определении эффективных мер по их сохранению.

### Допущения ислледования

1. Необходимо учесть факт, что сеть соавторства является ненаправленной:

,

где является элементом графа и обозначает связь .

1. В силу ненаправленности сети формирование или разрушение связи между акторами происходит по одному из двух сценариев:

* 1.1 Одностороння инициатива: Выбирается один актор , который получает возможность произвести изменение.
* 1.2 Двустороння инициатива: Выбирается упорядоченная пара акторов , при и получает возможность принять новое решение о связи .
* Процесс выбора моделируется по одному из двух сценариев:
* D. Диктаторский(Dictatorial): Один актор может навязывать решение о связи другому актору.
* M. Взаимный(Mutal): Оба актора дают согласие на существование связи между ними.

1. Между временными наблюдениями происходят изменения, которые не наблюдаются.
2. Вероятность изменения состояния сети зависит от её текущего состояния, и не зависит от истории изменений.
3. В любой момент времени в сети происходит не более одного изменения. Этими изменениями могут быть: создание или уничтожение связи.

Допущения 2-4 позволяют рассматривать модель как неприрывную цепь Маркова.

### 2.3.2 Спецификация Акторно ориентированной модели:

Модель является акторной в том смысле, что изменение связи моделируется, как результат выбора сделанного актором. Модель изменения связи состоит из двух компонент: время и выбор. Время изменения определяется в терминах возможности изменения, а не факта изменения. В случае если актора устраивает его текущее состояние он не будет изменять свои связи. Рассмотрим текущий момент времени и обозначим текущее состояние сети . Частота принятия решений актором  зависит от параметра . Обозначим .

Моменты времени распределены в соответствии с экспоненциальным распределением.

Эта формула соответствует модели «first past the post».Все участники имеют стохастическое время ожидания. Первый получивший возможность произвести изменение делает свой выбор и все начинается с начала, но уже в новом состоянии.

Совокупность выборов  актора образуют собой поток Пуассона  с параметром .

Пример:

, c интенсивностью



Рисунок 1 – Поток выборов актора 1

, c интенсивностью



Рисунок 2 – Поток выборов актора 2

, c интенсивностью



Рисунок 3 – Поток выборов актора 3

, c интенсивностью



Рисунок 4 – Поток выборов актора n

Образуют собой совокупный поток N(t) с интенсивностью .



Рисунок 5 – Совокупный поток выборов всех акторов.

Сумма всех потоков – есть так же поток Пуассона с параметром , где времена между моментами скачков независимы и распределены экспоненциально.

Вероятность того, что актор актор в выбранный промежуток времени примет решение о изменении сети:

,

где  остаточный член высшего порядка.

Вероятность того, что какой то актор в выбранный промежуток времени совокупного потока , где  примет решение о изменении сети :

.

У каждого актора есть  вариантов выбора, и всего  вариантов изменения сети. Построить\разрушить связь с кем то из других вершин, или ничего не делать. Для выбранного актора строится функция полезности для оценки вероятности выбора каждого из  вариантов. Функция полезности определяет вероятности того, что при следующей смене связей данный актор перейдет из состояния  в .

Функция полезности:

,

где  – итератор актора;

 – текущее состояние сети, для графа и ;

 – граф, который идентичен по всем ребрам, кроме пары , для которой существование связи является переключателем , причем возможен случай когда  это значит, что актор не изменит текущее состояние сети.

 определена на множестве всех пар сети и , таких, что . В интерпретации полезности функция полезности может рассматриваться как чистая полезность, которую получает актор от перехода из состояния  в .

Предположим, что в момент времени , при текущей сети у актора появилась возможность изменения в состояние



Это распределение вероятностей выборов актора представляет собой модель множественного выбора при n вариантов решений у вершины. Подробнее про построение функции полезности: В рассматриваемой модели функция полезности – есть линейная комбинация параметров  и функции сети, где функция сети описывает то, как выглядит изменение сети из текущего состоянии в состояние новое с точки зрения актора .

Функция сети:

,

где  – индекс параметра модели. Существует несколько реализаций функции сети [6] [4]. Некоторые из реализаций:

* Базовый компонент исходящих степеней:

.

Он подобен постоянному члену в регрессионных моделях и всегда включается. Этот компонент балансирует между созданием и прерыванием  
связей. Учитывая предыдущее состояние , следующее состояние либо  
имеет на одну связь больше, либо на одну связь меньше, или они идентичны.  
Если имеет параметр то для создания связи имеет вклад в функцию  
полезности , для разрушения связи . Таким образом, роль эффекта степени исходящих связей в модели заключается в вкладе в пользу создания  
связей по сравнению с их прерыванием. Обычно сети разрежены, поэтому  
возможностей для создания связей гораздо больше, чем для их прерывания.

* Взаимный выбор:



Фундаментальный аспект почти всех направленных социальных сетей, потому что почти всегда существует некий обмен или другая  
взаимная зависимость. Это отражается в степени взаимности.

Функция полезности:

.

Функция полезности описывает полезность для  актора перехода из  
состояния в .

Вероятность изменения связи :

,

где  – матрица смежности с элементами представляющая сеть в момент времени .

Учитывая факт того, что в рамках этой работы рассматривается  
ненаправленная сеть то создание связи может быть как односторонней  
инициативой, так и совместной, в отличии от разрушения связи. Разрушение  
связи происходит в одностороннем порядке. Рассмотрим разные варианты  
моделирования возникновения связи.

 – есть вероятность возникновения связи в разных сценариях.

1. Диктаторский сценарий
   1. Односторонняя инициатива:  
       
   2. Двусторонняя инициатива:   
      
2. Взаимный сценарий:

,

где  позволяет учесть факт существования или отсутствия связи в рассматриваемый момент времени.

Скорость изменения модели с односторонней инициативой:

.

Скорость изменения модели с двусторонней инициативой:

, где  – скорость изменения пары.

Таким образом Акторно-ориентированная модель (САОМ) работает по  
следующему алгоритму:

1. Зададим .
2. Генерируем  в соответствии с экспоненциальным временем с параметром .
3. Если  установить  и остановиться.
4. Выбираем случайный  используя вероятности .
5. Выбираем случайным образом , используя вероятности .  - множество всех состояний в которое может измениться сеть  в следствии решения актора .
6. Задать .
7. Задать .
8. Вернуться к шагу (2).

На принятие решения влияет множество  
параметров. Дальнейшей задачей является оценка этих параметров.

### 2.3.3 Оценка параметров:

Для оценки параметров используется имитационное моделирование.  
Реальная система заменяется моделью, и производится множество  
повторений алгоритма. Когда в среднем модель будет хорошо  
согласовываться с данными процесс оценки будет остановлен, и мы будем  
считать, что параметры оценены.  
Оценка параметров может быть произведена несколькими способами. В  
рамках пакета RSiena, как следствие, и в моей курсовой работе используется  
метод моментов [5].

Обозначим оцениваемый параметр как

.

Для каждого эффекта существует статистика, чувствительная к этому параметру. Для , влияющего на общее количество изменений используется статистика Hamming distance,



Для параметра, обозначающего насколько сильно скорость изменения актора  зависит от  , существует статистика – ковариата или позиционаая характеристика актора, такая как исходящая степень . Имеется в виду что скорость изменения может зависеть от пола, возраста, и тд.

.

Для функции полезности, где  не зависит от , большие значения  будут стремиться к сетям  для который значение  больше для любого актора . Для оценки  статистика имеет вид

.

Комбинируя статистики и используя предположение о марковской цепи для наблюдаемых данных оценочные уравнения имеют вид.

,









,

где  и  номера элементов и .

Для решения этих уравнений используется стохастическая оптимизация на основе алгоритма Роббинса-Монро. Алгоритм применяет многомерную версию алгоритма Роббинса-Монро с улучшениями, предложенными Поляком и Руппертом. Техника "двойного усреднения" также применяется для улучшения результатов. Алгоритм реализован в пакете RSiena на языке R.

Он состоит из трех фаз:

1. Определение чувствительности ожидаемых статистик к параметрам.
2. Обновляет параметры с использованием симуляций динамики сети.
3. Используется для оценки приближения полученных уравнений и вычисления стандартных ошибок.

Для вычисления производных ожидаемых значений по отношению к параметрам используется метод функции оценки. Этот алгоритм является надежным, но затратным по времени.

# Глава 3 Анализ результатов

Формирование начальных гипотез, а так же формирование социологических выводов – это задача прикладных специалистов.

В рамках моей курсовой работы я использую некоторые предположения выдвинутые в подобном исследовании для национальной научной системы Словении [3].

## 3.1 Формирование начальных социологических гипотез.

Согласно исследованию Anuška Ferligoj, Luka Kronegger, Franc Mali, Tom A B Snijders, Patrick Doreian в журнале Scientometrics (2015, №104, с. 989–990):«В целом, мы предлагаем следующие механизмы влияния на научное соавторство: встраиваемость в сеть: соавторы соавторов стремятся стать соавторами соавторов; преимущественная привязанность: авторы предпочтительно ищут соавторов, у которых уже много соавторов; институциональная встроенность (принадлежность к одной исследовательской группе и одной научной дисциплине, возрастное сходство также может подпадать под эту категорию, поскольку означает принадлежность к общей когорте ученых, которые взаимодействуют друг с другом больше, чем представители разных когорт) и контрольные переменные, в частности стаж.»

## 3.2 Оценка сформированных гипотез.

Коэффицент сходимости модели достаточно низок ( Overall maximum convergence ratio: 0.0508 ) , для того, чтобы считать модель сходящейся[7].

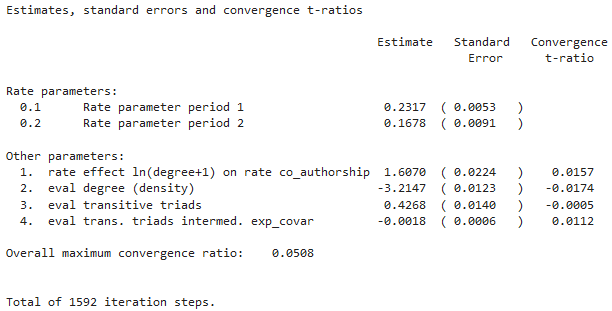


Рисунок 6 – Таблица оценок, стандартных ошибок и t-критерий Стьюдента

### 3.2.1 Интерпретация результатов

Необходимо учитывать, что функция эволюции сети задана формулой:

 ,

где 

,

где  не зависит от  причем может зависеть от  для другого актора .

Она отражает полезность для  актора. Используется для сравнения полезности различных изменений. С другой стороны когда  актор делает изменение из  в , и   два возможных результата шага то отношение вероятностей этих вариантов есть: . Следовательно сетевые параметры стоит расценивать как логарифмические.

Для оценки зависимых параметров необходимо учитывать, что возможные изменения в переменной , и оценки не должны выходить за пределы допустимого диапазона. Параметры целевой функции это вклады в логарифмические вероятности увеличения зависимой переменной на  
1 единицу при увеличении эффекта на 1 единицу. Где зависимая переменная либо , либо .

Пример:

Рассчитан эффект average alter он равен 1,1561. Это означает, что при сравнении двух акторов равных во всех отношениях, кроме того, что связи первого в среднем на 1 выше по шкале рассматриваемой ковариаты (Индекс Хирша, и тд), чем у второго, шанс увеличения значения ковариаты по сравнению с отсутствием изменений ( в рамках одного шага относительно ковариаты) выше в  раза, чем для второго.

Не любой эффект можно трактовать как в приведенном примере.

Пример:

Для эффекта out-degree (количество связей) для актора  . Пусть число связей, чьи значения  меньше, равны или больше значения  для , обозначим через a, b и c. Обозначим диапазон (максимальное минус минимальное значение) через . Тогда, в случае минимального шага в отношении поведения, вклады cуммарного эффекта сходства в лог-вероятности изменений  определяются следующим образом  и . Вклады для среднего  
эффекта сходства составляют  и . Это показывает, что влияние связей на эффекты сходства зависит только от того, имеют ли они большие или меньшие значения, чем у рассматриваемого актора, а не от того, насколько эти значения больше. Этот также показывает, что для эффектов сходства важна дисперсия ценностей связей, а не  
а не только их среднее значение, в то время как для эффекта среднего изменения имеет значение только среднее значение.

Не все эффекты могут быть описаны как изменение некоторой функции полезности. В таких случаях используются элементарные эффекты [8].

Элементарный эффект - это вклад в создание или поддержание связи, определенный напрямую, то есть без выражения его на основе изменения какой-либо функции оценки. Это означает, что элементарные эффекты более общие, чем эффекты оценки, и все эффекты могут быть представлены как элементарные эффекты. Однако для удобства интерпретации предпочтительно использовать формулировку функции оценки.

Элементарные эффекты могут применяться аналогично к созданию и поддержанию связи; или они могут применяться исключительно к созданию связи или исключительно к поддержанию связи.

Основной пример ситуации, когда невозможно описать как изменение функции полезности - тенденция к созданию связные троек (транзитивные замыкания). Если актор  является центральным, то связи, приводящие к замыканию, представлены как →  и  → . Однако первая связь подразумевает создание пути  →  → , в то время как вторая связь означает установление связи с актором . Актор , в свою очередь, делает тот же исходящий выбор для третьего актора , что указывает на структурную эквивалентность; таким образом, это различные процессы.

##### density

Оценка параметра этого эффекта указывает на то, что с увеличением исходящих степеней снижается вероятность появления новой связи.

##### outRateLog

Логарифмическая зависимость скорости (параметра  для актора ) от количества исходящих степеней. С увеличением числа связей актор увеличивает частоту принятия решений.

##### transitive triads

Оценка этого эффекта показывает количество транзитивных паттернов относительно актора . Для двух акторов, где каждый из них является соавтором с  индивидуально, вероятность стать соавторами увеличивается в  раза.

eval trans. triads intermed. exp\_covar

Значение оценки параметра этого эффекта означает, что с увеличением стажа ослабевает эффект transitive triads.

# Заключение

В рамках курсовой работы я ознакомился с дисциплиной сетевой анализ и ознакомился с современным методом анализа динамики социальных сетей - Стохастическое акторно ориентированное моделирование (Stohastic Actor-Oriented Models for Network Dynamics). Полученные результаты исследования могут оказать практическую значимость и служить основой для дальнейших исследований в данной области.

# Список литературы

1. Докука C.B., Валеева Д.Р. Статистические модели для анализа динамики социальных сетей в исследованиях образования // Вопросы образования. - 2015. - №1. - С. 201-213.
2. Tom Broekel, Pierre-Alexandre Balland, Martijn Burger, Frank Oort Modeling knowledge networks in economic geography: a discussion of four methods // The Annals of Regional Science. - 2014. - №53. - С. 423-452 .
3. Anuška Ferligoj, Luka Kronegger, Franc Mali, Tom A B Snijders, Patrick Doreian Scientific collaboration dynamics in a national scientific system // Scientometrics. - 2015. - №104. - С. 985–1012.
4. Snijders, T.A.B., & Pickup, M. (2016). Stochastic Actor-Oriented Models for Network Dynamics. Retrieved June 10, 2016.
5. Snijders, T. A. B. he Statistical Evaluation of Social Network Dynamics // Sociological Methodology. - 2001. - №31. - С. 361-395.
6. Siena algorithms // Department of Statistics - University of Oxford URL: <https://www.stats.ox.ac.uk/~snijders/siena/Siena_algorithms.pdf> (дата обращения: 06.01.2024).
7. Manual for RSiena / M. R. Ruth. — Текст : электронный // University of Oxford: Department of Statistics; Nuffield College University of Groningen: Department of Sociology : [сайт]. — URL: <https://www.stats.ox.ac.uk/~snijders/siena/RSiena_Manual.pdf> (дата обращения: 14.01.2024).
8. Snijders The statistical evaluation of social network dynamics / Snijders, A T. — Текст : непосредственный // Sociolog- ical Methodology. — 2001. — № 31. — С. 361-395.