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*

«\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_2023 г.

Автор работы

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

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

*подпись*

«\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_2023 г.

Томск – 2023

Оглавление.

Введение

1 ТЕОРИТИЧЕСКАЯ ЧАСТЬ

1.1 Научные группы и команды: определение, характеристики, особенности

1.2 Библиометрический анализ: методы и подходы

1.3 Факторы, влияющие на формирование и разрушение научного взаимодействия

2 Методология исследования

2.1 Объект исследования

2.2 Методика библиометрического анализа

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

3 Анализ результатов

3.1 Формирование начальных социологических гипотез.

3.2 Оценка сформированных гипотез.

3.3 Оценка факторов, влияющих на формирование и разрушение научного взаимодействия.

Заключение

Список литературы

# 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. Разработал САОМ Tom A.B. Snijders[[1]](#footnote-0). 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. Я построю несколько временных снимков сети соавторства Томского Государственного Университета используя данные с информационного портала ТГУ.Сотрудники[[2]](#footnote-1). Исследования с ипользованием САОМ проводилось для динамики научного сотрудничества в национальной научной системе Словении [3].

Результаты исследования могут быть использованы для оптимизации научного взаимодействия в рамках Томского Государственного Университета и дальнейшего развития научной деятельности в целом.

1 ГЛАВА ТЕОРИТИЧЕСКАЯ 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. 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 буду рассматривать такие мотоды 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.
* Кластерный анализ. Этот подход основан на группировке научных работ по подразадлению авторов. Он позволяет выявить кластеры научных работ, которые имеют схожую тематику, и определить их особенности и тенденции.
  1. 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 выделть в рамках взаимодейтсвия в Томском Государственном Унивеситете:

* 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.

2 ГЛАВА 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 исслодования использованы сетевые структуры и метрики. Выгруженные 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[[3]](#footnote-2).

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.

2.3.1 Допущения ислледования

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

,

где явлется элементом графа и обозначает связь

Мы предполагаем, что формирование или разрушение связи между акторами происходит по одному из двух сценариев:

* 1. Одностороння инициатива: Выбирается один актор i, который получает возможность произвести изменение.
  2. Двустороння инициатива: Выбирается упорядоченная пара акторов (i,j), 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.

2. Между временными наблюдениями происходят изменения, которые не наблюдаются.

3. Вероятность изменения состояния сети зависит от её текущего состояния, и не зависит от истории измененией.

4. В любой момент изменений сети происходит не более одного изменения. Этими изменениями могут быть: создание или уничтожение связи.

Допущения 2 и 3 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 i 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 i актора образуют собой поток Пуассона , где 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 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 изменения сети, и всего n вариантов выбора. 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 n 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 (i,j), для которой существование связи является переключателем

=,

причем возможен случай когда j=i 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 .

Для определения этой вероятности используем следующие обозначения: Предположим, что в момент времени , при текущей сети у актора появилась возможность изменения состояния сети в состояние задана

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 i.

Функция сети: , где – индекс параметра модели. Существует несколько реализаций функции сети [4]. В качестве примера приведу некоторые:

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.

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 ij:

,

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

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.

Диктаторский сценарий:

Односторонняя инициатива:

Двусторонняя инициатива, актор i делает бинарный выбор о том будет ли существовать связь ij :

Взаимный сценарий:

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 λ.
3. Если t + > установить и остановиться.
4. Выбираем случайный используя вероятности
5. Выбираем случайный 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 этапе моделирования принятия решения получается 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 существует статистика, чувствительная к этому параметру. Для , влияющего на общее количество изменений используется статистика Hamming distance,

.

Для параметра , обазначаюшего насколько сильно скорость изменения актора i 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 i. Для оценки статистика имеет вид

.

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

где номера элементов .

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: первая фаза используется для определения чувствительности ожидаемых статистик к параметрам, вторая фаза обновляет параметры с использованием симуляций динамики сети, а третья фаза 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.

3 ГЛАВА 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.

По итогам оценки параметров можно сказать, что (список эффектов влияющих) имеют значимый эффект в процессе принятия решения актором.

Некоторые из эффектов акторы стараются наоборот минимизировать (список).

Эти эффекты не наблюдадются вопреки теоретическому ожиданию (список)

3.3 Оценка факторов, влияющих на формирование и разрушение научного взаимодействия.

Социологическая интерпретация цифорок, которые мы получили.

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.

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. Broekel, T., Balland, P.A., Burger, M., & van Oort, F. (Year not provided). Modeling knowledge networks in economic geography: a discussion of four methods.
7. Mali, F., Kronegger, L., Doreian, P., & Ferligoj, A. (Year not provided). Dynamic Scientific Co-Authorship Networks.
8. Palla, G., Pollner, P., Barabási, A.L., & Vicsek, T. (Year not provided). Social Group Dynamics in Networks.
9. Snijders, T.A.B. (2017). Stochastic Actor-Oriented Models for Network Dynamics. In Annual Review of Statistics and Its Application, Volume 4. University of Groningen, University of Oxford.

1. Профессор кафедры статистики в социальных науках Наффилдского колледжа, Оксфорд [↑](#footnote-ref-0)
2. ТГУ.Сотрудники // URL: https://persona.tsu.ru/ [↑](#footnote-ref-1)
3. GitHub репозиторий // URL: https://github.com/Amatikay/Parsing-data-from-TSU.Employees [↑](#footnote-ref-2)