Оглавление

[Перечень условных обозначений 2](#_Toc167203315)

[Введение 2](#_Toc167203316)

[1. Обзор литературы 2](#_Toc167203317)

[2. Теоретическая часть САОМ 2](#_Toc167203318)

[2.1 Постановка задачи приводящей к САОМ 2](#_Toc167203319)

[2.2 Описание САОМ 2](#_Toc167203320)

[2.3 Оценка параметров 2](#_Toc167203321)

[3. Применение модели на реальных данных 2](#_Toc167203322)

[3.1 Объект исследования 2](#_Toc167203323)

[3.2 Применимость САОМ в исследовании научных групп 2](#_Toc167203324)

[3.3 Формирование гипотез о динамике рассматриваемой сети 2](#_Toc167203325)

[3.4 Описание модели 2](#_Toc167203326)

[3.5 Оценка сформированных гипотез 2](#_Toc167203327)

[3.5.1 Оцененные параметры при эффектах 2](#_Toc167203328)

[3.5.2 Интерпретация полученных оценок параметров 2](#_Toc167203329)

[4. Построение имитационной модели 2](#_Toc167203330)

[4.1 Описание модели 2](#_Toc167203331)

[4.2 Допущения 2](#_Toc167203332)

[4.3 Описание работы имитационной модели 2](#_Toc167203333)

[4.4 Оценка работы имитационной модели 2](#_Toc167203334)

[Заключение 2](#_Toc167203335)

[Список использованных источников и литературы 2](#_Toc167203336)

# Перечень условных обозначений

Социальная сеть - объединение социальных акторов и их связей.

Актор - действующий субъект, совершающий действие, направленное на других акторов.

Ковариата - частная характеристика актора.

САОМ, SAOM – класс стохастических акторно ориентированных моделей (Stochastic Actor-Oriented Models, САОМ, SAOM)

# Введение

В современном мире изучение социальных сетей приобретает всё большую значимость. Это связано с необходимостью анализа и оптимизации различных социальных процессов, будь то в науке, экономике или политике. С развитием сети интернет появилась возможность собирать и анализировать объективные данные о социальных сетях, и их структурах. Это вызвало развитие такой дисциплины, как сетевой анализ. В рамках сетевого анализа существует класс акторно-ориентированных моделей. Стохастическое акторно-ориентированное моделирование является мощным инструментом для анализа и предсказания сетевой динамики.

Объектом данного исследования является класс стохастических акторно-ориентированных моделей, а именно построение этих моделей, их применимость на реальных данных. Предметом исследования выступают факторы, влияющие на динамику сети соавторства в Томском государственном университете.

Целью данной работы является анализ факторов, влияющих на сеть соавторства Томского государственного университета, с использованием стохастического акторно-ориентированного моделирования. Построение и анализ имитационной модели стохастической акторно ориентированно модели.

# Обзор литературы

# Теоретическая часть САОМ

## Постановка задачи приводящей к САОМ

Рассмотрим некоторую социальную сеть. Изменение сети можно рассматривать как изменение актором своего «положения» в этой сети. Каждое изменение связей актора, а именно создание новой связи с другим актором, или разрушение существующей связи приводит к изменению всей сети.

Для разных сетей могут быть разные требования к этому изменению. Например: для сети дружбы между одногруппниками процесс создания или разрушения связи ( в данном случае связь – считает ли актор другого актора своим другом) происходит в одностороннем порядке, и не требует подтверждения принимающей стороны ( можно считать своим другом того, кто не считает своим другом). Однако в случае сети научного соавторства необходимо подтверждение обеих сторон для создания связи. Так же следует заметить, что в примере с сетью научного соавторства нельзя разрушить связь.

Как при выборе того, с кем актор хочет изменить сеть, так и при выборе какое изменение следует внести (создание или разрушение связи) актор руководствуется некоторыми критериями оптимальности.

При изучении реальных социальных сетей в большинстве случаев не возможно наблюдать за динамикой сети непрерывно. Пусть исследователю доступны лишь несколько снимков состояния сети.

Для анализа и предсказания динамики социальных сетей существует класс стохастических акторно-ориентированных моделей. Под потянем акторно-ориентированные подразумевается, что все изменения сети происходят из за принятий решений акторами этой сети.

## Описание САОМ

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. Обозначим , как состояние сети в момент времени .

В любой момент времени в сети происходит не более одного изменения. Этими изменениями могут быть: создание или уничтожение связи. Вероятность изменения состояния сети зависит от её текущего состояния, и не зависит от истории изменений. Таким образом – непрерывная цепь Маркова.

Частота принятия решений актором *i* зависит от параметра . Обозначим . .

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Моменты времени распределены в соответствии с экспоненциальным распределением.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

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 с параметром .

Пример:

, с интенсивностью



Рисунок 1 – Поток выборов актора 1

, с интенсивностью



Рисунок 2 – Поток выборов актора 2

, с интенсивностью



Рисунок 3 – Поток выборов актора 3

, с интенсивностью



Рисунок 4 – Поток выборов актора n

Образуют собой совокупный поток 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 *i* актор в выбранный промежуток времени примет решение о изменении сети:

|  |  |  |
| --- | --- | --- |
|  | , | (3) |

где остаточный член высшего порядка.

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

|  |  |  |
| --- | --- | --- |
|  | .. | (4) |

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 перейдёт из состояния в .

Функция полезности:

|  |  |
| --- | --- |
|  | (5) |

где *i* – итератор актора принимающего решение о связи с актором *j*;

– текущее состояние сети, для графа и ,

– граф, который идентичен по всем рёбрам, кроме пары , для которой существование связи является переключателем , причём 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 *i* от перехода из состояния . .

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

|  |  |
| --- | --- |
|  | (6) |

Это распределение вероятностей выборов актора представляет собой модель множественного выбора при 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:

|  |  |
| --- | --- |
| , | (7) |

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

* Базовый компонент исходящих степеней:

|  |  |
| --- | --- |
|  | (8) |
|  |  |

Он подобен постоянному члену в регрессионных моделях и всегда включается. Этот компонент балансирует между созданием и разрушением

связей. Учитывая предыдущее состояние x, следующее состояние либо

имеет на одну связь больше, либо на одну связь меньше, или они идентичны. Функция сети умножается на параметр , он показывает вклад функции полезности в вероятность создания новой связи, вклад в разрушение существующей связи . Таким образом, роль эффекта степени исходящих связей в модели заключается в вкладе в пользу создания

связей по сравнению с их разрушением. Обычно сети разрежены, поэтому

возможностей для создания связей гораздо больше, чем для их прерывания.

* Число взаимных связей:

|  |  |
| --- | --- |
|  | (9) |

Фундаментальный аспект почти всех направленных социальных сетей, потому что почти всегда существует некий обмен или другая взаимная зависимость.

Функция полезности:

|  |  |
| --- | --- |
|  | (10) |

Функция полезности описывает полезность перехода из  
состояния  в для *i* актора.

Вероятность изменения связи :

|  |  |
| --- | --- |
|  | (11) |
|  |  |

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

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:

|  |  |
| --- | --- |
|  | (12) |

|  |  |
| --- | --- |
| б) Двусторонняя инициатива: | (13) |

1. Взаимный сценарий:

,(14)

|  |  |
| --- | --- |
|  |  |

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

|  |  |
| --- | --- |
|  | (15) |

Скорость изменения модели с односторонней инициативой:

|  |  |
| --- | --- |
|  | (16) |

Скорость изменения модели с двусторонней инициативой:

|  |  |
| --- | --- |
|  | (17) |

– матрица интенсивностей переходов цепи маркова .

Таким образом Стохастические Акторно-ориентированные модели (САОМ) работают по  
следующему алгоритму:

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

На принятие решения может влияет множество  
эффектов, и параметров при этих эффектах. Дальнейшей задачей является оценка этих параметров.

## 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 метод моментов [5].

Обозначим оцениваемый параметр как

|  |  |
| --- | --- |
| . | (18) |

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

|  |  |
| --- | --- |
|  | (19) |

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

|  |  |
| --- | --- |
| . | (20) |

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

|  |  |
| --- | --- |
| . | (21) |

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

|  |  |
| --- | --- |
| ,, | (22) |
| ,, | (23) |
|  |  |
| ,  , | (24) |

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 Руппертомх[10]. Техника "двойного усреднения" также применяется для улучшения результатов. Алгоритм реализован в пакете RSiena языка R.

Он состоит из трёх 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 надёжным, но затратным по времени[10].

# Применение модели на реальных данных

## Объект исследования

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

Данные выгружены с веб сайта «ТГУ.Сотрудники». Используемая информация о авторах: Опыт работы и список публикаций автора, образующий сеть соавторства.

## Применимость САОМ в иследовании научных групп

Научное взаимодействие представляет собой процесс сотрудничества между учёными 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: Учёные, 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, учёные могут ощутить трудности в формировании и поддержании научных контактов.
* Личностные факторы: Личностные характеристики учёных, 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.

САОМ - это метод социально-сетевого анализа, который позволяет моделировать и предсказывать эволюцию социальных сетей на основе поведения и взаимодействий индивидуальных акторов в них. САОМ 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, САОМ может быть использован для моделирования процессов формирования научных связей между учёными в рамках группы или команды.

Используя САОМ 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 САОМ включают возможность моделирования процессов на уровне индивидуальных акторов, учёт динамической природы социальных связей, а также возможность оценки влияния различных факторов на эти процессы. Однако САОМ также имеет некоторые ограничения, такие как необходимость большой вычислительной мощности для моделирования и ограниченная возможность предсказания долгосрочных эффектов.

Таким образом, САОМ 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 а так же оптимизации научного взаимодейтсвия.

## Формирование гиптез о динамике рассмтриваемой сети

Формирование начальных гипотез, а так же формирование социологических выводов – это задача прикладных специалистов. Будут использованы некоторые предположения выдвинутые в подобном исследовании для национальной научной системы Словении [3].

**[ПЕРЕВОД]** 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 учёных, которые взаимодействуют друг с другом больше, чем представители разных когорт) и контрольные переменные, в частности стаж.»

## Описание модели

Необходимо учесть факт, что сеть соавторства является ненаправленной:

|  |  |  |
| --- | --- | --- |
|  | , | (25) |

где является элементом графа и обозначает связь .

В силу ненаправленности сети формирование связи между акторами происходит по одному из двух сценариев:

* 1. Односторонняя инициатива: Выбирается один актор *i*, который получает возможность произвести изменение.
* 2. Двусторонняя инициатива: Выбирается упорядоченная пара акторов , 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.

Согласно выдвинутым социологическим предположениям[3] формирование связи будет происходить согласно двусторонней инициативы, и оба актора будут давать согласие на создание связи.

Факторы влияющие на оценку полезности создания связи:

Изменить названия эффектов

* + 1. Количество исходящих связей
    2. Транзитивные тройки
    3. Влияние стажа на эффект транзитивных троек
    4. Влияние количества исходящих связей на частоту принятия решений

## 3.5 Оценка сформированных гипотез

### 3.5.1 Оцененные параметры при эффектах

Коэффициент сходимости модели достаточно низок 0.0508, для того, чтобы считать модель сходящейся[7]. Что означает, что выдвинутые предположения о виде модели корректно описывают представленные снимки сети. Так же абсолютное значение *Convergence t-ratio*[13] для каждого параметра меньше , что говорит о значимости каждого из эффектов.

|  |  |  |  |
| --- | --- | --- | --- |
| Эффект | Оценка | Стандартное отклонение | Convergence t-ratio |
| Интенсивность в 1 периоде | 0.2317 | 0.0053 |  |
| Интенсивность в 2 периоде | 0.1678 | 0.0091 |  |
| Количество исходящих связей | -3.2147 | 0.0224 | 0.0157 |
| Транзитивные тройки | 0.4268 | 0.0140 | -0.0005 |
| Влияние стажа на эффект транзитивных троек | -0.0018 | 0.0006 | 0.0112 |
| Влияние количества исходящих связей на частоту принятия решений | 1.6070 | 0.0224 | 0.0157 |

Табл 1 Оценённые значения эффектов

### 3.5.2 Интерперетация полцченных оценок параметров

Необходимо учитывать, что функция эволюции сети задана формулой:

|  |  |
| --- | --- |
| , | (26) |
| , | (27) |

где не зависит от причём может зависеть от для другого актора *j*.

Она отражает полезность для *i* актора. Используется для сравнения полезности различных изменений. С другой стороны когда *i* 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  
 единицу при увеличении эффекта на единицу. Где зависимая переменная либо , либо .

Пример:

Рассчитан эффект *среднее число альт*е*ра* [Ссылка, что альтер - термин] он равен . 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 приведённом примере.

Пример:

Для эффекта *количество исходящих связей* для актора . Пусть число связей, чьи значения меньше, равны или больше значения для *i*, обозначим через *a*, *b* и *c*. Обозначим диапазон (максимальное минус минимальное значение) через . Тогда, в случае минимального шага в отношении поведения, вклады суммарного 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. Это 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, определённый 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.

Пример ситуации, когда невозможно описать как изменение функции полезности - тенденция к созданию связных троек (транзитивные замыкания). Если актор *i* является центральным в тройке, 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 *h*. Актор *h*, в свою очередь, делает тот же исходящий выбор для третьего актора *j*, что указывает на структурную эквивалентность; таким образом, это различные процессы.

Таким образом полученные оценки можно интерпретировать так:

* *Количество исходящих связей* указывает на то, что с увеличением исходящих степеней снижается вероятность появления новой связи.
* *Зависимость скорости* (параметра для актора ) *от количества исходящих степеней*: c увеличением числа связей актор увеличивает частоту принятия решений.
* *Транзитивные тройки*: показывает количество транзитивных паттернов относительно актора . Для двух акторов, где каждый из них является соавтором с индивидуально, вероятность стать соавторами увеличивается раза.
* *Влияние стажа на эффект транзитивных троек* означает, что с увеличением стажа ослабевает эффект транзитивных троек.

# Построение имитационной модели

## Описание модели

## Допущения

## Описание работы имитационной модели

## 4.4 Оценка работы имитационной модели

# Заключение

# Список использованных источников и литературы