

УДК 001

# О ВЗАИМОСВЯЗИ СЕТИ СОАВТОРСТВА И ТЕМАТИЧЕСКОЙ БЛИЗОСТИ УЧЕНЫХ НА ПРИМЕРЕ ИПУ РАН

**Д.А. Губанов**

*Институт проблем управления им. В.А. Трапезникова РАН*  
Россия, 117997, Москва, Профсоюзная ул., 65  
E-mail: dmitry.a.g@gmail.com

**А.Г. Чхартишвили**

*Институт проблем управления им. В.А. Трапезникова РАН*  
Россия, 117997, Москва, Профсоюзная ул., 65  
E-mail: sandro\_ch@ipu.ru

**Ключевые слова:** сети научного соавторства, тематическое пространство, теория управления, профиль ученого.

**Аннотация:** В докладе исследуется взаимосвязь между сетью научного соавторства (в которой узлами являются ученые, а связью – наличие совместных публикаций) и показателями тематической близости ученых на примере ИПУ РАН. Под тематической близостью понимается близость в тематическом пространстве теории управления, вычисленная на основе методологии, реализованной в Информационной системе автоматизации научной деятельности (ИСАНД), которая разрабатывается в ИПУ РАН. Сформулированы и исследованы гипотезы относительно соотношения тематической близости и близости в сети соавторства.

## 1. Введение

Данный доклад посвящен исследованию взаимосвязи сети научного соавторства (в которой узлами являются ученые, а связью – наличие совместных публикаций) и показателей тематической близости ученых на примере ИПУ РАН. Под тематической близостью мы будем понимать близость в тематическом пространстве теории управления [1].

В настоящее время существует большое количество классификаторов, используемых для различных целей: УДК [2], международный классификатор OECD [3], классификатор Российского научного фонда [4], ГРНТИ [5] и др. Каждый из них является своего рода «системой координат», позволяющей выявлять и классифицировать относящиеся к тем или научным направлениям объекты (в первую очередь – публикации, но также и ученых, научные организации и т.д.). В данной работе мы опираемся на онтологию научного знания теории управления, разработанный экспертами ИПУ РАН для Информационной системы автоматизации научной деятельности (ИСАНД).

Далее в разделе 2 будут описаны исходные данные для анализа: профили ученых (раздел 2.1) и сеть соавторств (раздел 2.2). В разделе 3 изложены полученные результаты, в Заключение – их краткое обсуждение и перспективы развития.

## 2. Исходные данные

### 2.1. Формирование профилей ученых

Разработанная в ИПУ РАН онтология научного знания теории управления имеет четырехуровневую структуру, уровни которой (кроме нижнего) представляют собой дерево. Уровни нумеруются числами от 0 до 4.

Нулевой уровень содержит 3 фиксированные вершины «Математический аппарат», «Проблемная область», «Сфера применения». Предполагается, что при возможных расширениях онтологии этот уровень не изменяется. Он отражает различные аспекты научных исследований, относящихся к теории управления: используемый математический инструментарий (теорию графов, теорию игр, ...), проблемную область, т.е. некоторую прикладную теорию (мехатронику, теорию управления в организационных системах, ...) и конкретную сферу применения (транспорт, энергетику, ...). Вершины нулевого уровня называются *аспектами*.

Каждая из вершин нулевого уровня является корнем тематического поддерева, раскрывающего ее содержание. Например, поддерево «Математический аппарат» содержит вершину «Теория игр» (1-й уровень) и детализирующие ее вершины 2-го уровня: «Теория некооперативных игр» и др. Соответственно, поддерево «Проблемная область» среди прочих содержит вершину «Теория управления в организационных системах» и детализирующие ее вершины 2-го уровня, например, «Механизмы планирования», а поддерево «Сфера применения» содержит вершины «Энергетика» (1-й уровень) и «Атомная энергетика» (2-й уровень). Вершины первого уровня будем называть *темами*, вершины второго уровня – *подтемами*.

Третий уровень – это вершины-*термины*, характеризующие подтемы. На уровне терминов древовидность структуры онтологии может нарушаться, т.е. один термин может относиться (в редких случаях) к нескольким различным подтемам или даже темам.

Для каждой публикации можно рассчитать стохастические вектора тематических профилей (далее будем называть их просто *профилями*) по подтемам, темам и аспектам на основе частоты вхождения в ее текст терминов, характеризующих каждую из подтем, с применением агрегирования по подтемам по темам, а тем по аспектам. Получив профили каждой публикации, далее можно рассчитать профили авторов публикаций (подробнее см. [1]). В данной работе мы ограничимся исследованием профилей по темам.

Отметим, что можно определять расстояние между профилями, лежащее в промежутке от 0 до 1. Мы будем пользоваться расстоянием, основанном на манхэттенской метрике (подробнее см. [1]).

### 2.2. Формирование сети соавторства

Научная деятельность порождает немалое количество объектов и связей между ними. Например, в научных публикациях принято ссылаться на предыдущие исследования. Имея соответствующие данные, можно построить направленную сеть, в которой вершинами являются публикации, а дугами между вершинами – ссылки одной публикации на другую (см., напр., [6, 7]).

В данной работе мы рассматриваем сеть соавторства (см., напр., [8]). В ней узлами являются ученые, а ненаправленная дуга между двумя узлами означает наличие совместных публикаций.

Исходная сеть соавторства (на основе данных из внутренней базы публикаций ИПУ РАН) состоит из 6 502 вершин и 24 285 связей. Далее учитывались только актуальные сотрудники лабораторий ИПУ РАН, такой граф состоит из 430 вершин и

1 110 связей соавторства. Для сотрудников были рассчитаны расстояния между профилями.

### 3. Результаты исследования публикаций по теории управления сотрудников ИШУ РАН

В данной работе предлагается при помощи профилей ученых и связей соавторства рассмотреть следующие исследовательские вопросы (гипотезы).

- i) Сотрудники пишут совместную публикацию в двух случаях: а) если их тематические профили (научные интересы) близки, б) если они взаимодополняют друг друга, т.е. их профили далеки друг от друга.
- ii) Среднее расстояние между профилями двух авторов меньше в том случае, если они связаны в сети соавторства сильной связью.

Под *сильной связью* между двумя авторами мы будем понимать следующую ситуацию: существует третий автор, который имеет хотя бы одну публикацию с первым автором без второго и хотя бы одну публикацию со вторым без первого (т.е. связан отдельно с каждым из двух авторов).

Заметим, что среднее расстояние между профилями сотрудников является довольно большим и составляет 0,85. Это означает, что публикации сотрудников относятся, вообще говоря, к различным областям теории управления.

Для ответа на первый исследовательский вопрос рассмотрим распределение расстояний между профилями соавторов статей – см. рис. 1.

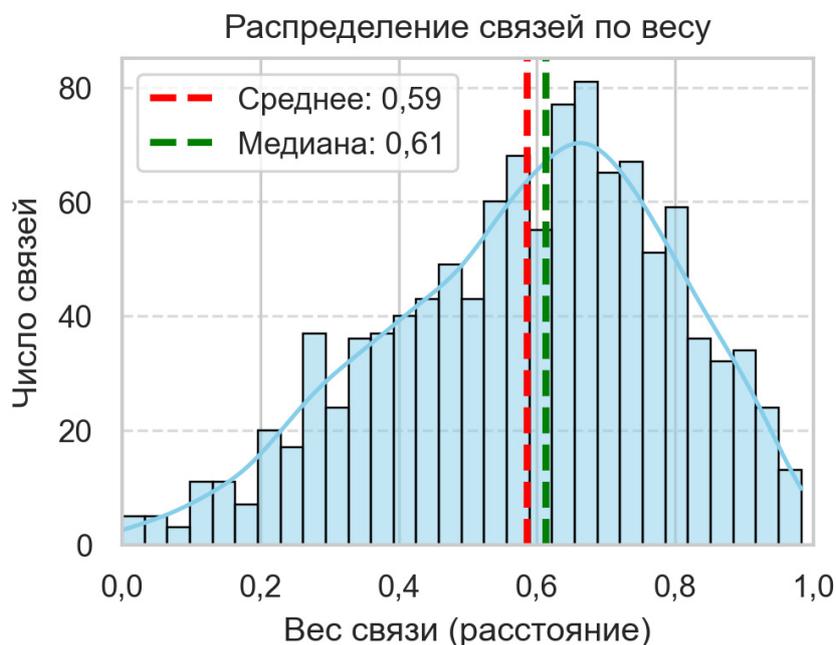


Рис. 1. Распределение для графа соавторств.

Исходя из распределения нельзя выделить (поскольку отсутствуют соответствующие пики или сгущения плотности) такие кластеры публикаций, в одном из которых интересы авторов почти совпадают, а в другом существенно различаются. Поэтому первую гипотезу в данном случае следует отвергнуть.

Для проверки второй гипотезы были составлены два набора значений:

- расстояния между профилями авторов, связанных в графе соавторства сильной связью;
- расстояния между профилями авторов, соединенных в графе соавторства, но не связанных сильной связью.

В результате проверки по статистическому критерию Вилкоксона-Манна-Уитни было выявлено, что профили авторов, соединенных сильной связью, в среднем ближе друг к другу ( $p$ -value =  $4,9 \times 10^{-6}$ , средние значения составляют 0,59 для сильной связи и 0,64 для связи, не являющейся сильной). Таким образом, вторая гипотеза подтверждается.

## 4. Заключение

В данном докладе были исследованы две гипотезы о взаимосвязи сети соавторства и близости профилей авторов публикаций.

Первая гипотеза состояла в том, что сотрудники пишут совместную публикацию либо с тематически близкими соавторами (т.е. авторы обладают схожими тематическими компетенциями), либо, наоборот, с тематически далекими соавторами (т.е. авторы обладают далекими компетенциями, дополняющими друг друга при написании публикации). На имеющихся данных гипотеза не подтвердилась.

Вторая гипотеза состояла в том, что у сильно связанных в сети соавторства ученых тематические профили ближе, чем у просто соединенных связью. Эта гипотеза подтвердилась.

Сеть соавторств научных публикаций представляет собой активную сетевую структуру, во многих аспектах аналогичную другим активным сетевым структурам, в частности – социальным сетям. Поэтому следует ожидать, что методов сетевого анализа может оказаться применимыми для анализа этой сети.

## Список литературы

1. Губанов Д.А., Кузнецов О.П., Суховеров В.С., Чхартишвили А.Г. О построении профилей в тематическом пространстве теории управления / Материалы 9-й Международной конференции «Знания-Онтологии-Теории» (ЗОНТ-2023, Новосибирск). Новосибирск: Институт математики им. С.Л. Соболева СО РАН, 2023. С. 89-94.
2. ГОСТ 7.90.2007. Система стандартов по информации, библиотечному и издательскому делу. Универсальная десятичная классификация. Структура, правила введения и индексирования: издание официальное.
3. Коды международной классификации (OECD). [https://storage.tusur.ru/files/134958/kody\\_OECD.pdf](https://storage.tusur.ru/files/134958/kody_OECD.pdf) (дата обращения: 15.09.2023).
4. Классификатор Российского научного фонда (РНФ). <https://rscf.ru/contests/classification/> (дата обращения: 15.09.2023).
5. Государственный рубрикатор научно-технической информации. <https://grnti.ru/> (дата обращения: 14.06.2023).
6. Печников А.А., Чебуков Д.Е. Структура графа цитирования журналов Math-Net.Ru // Научный сервис в сети Интернет: труды XXIII Всероссийской научной конференции (20-23 сентября 2021 г., онлайн). М.: ИПМ, 2021. С. 265-278.
7. Yang S., Wang F. Visualizing information science: Author direct citation analysis in China and around the world // Journal of Informetrics. 2015. Vol. 9, No. 1. P. 208-225.
8. Sarigöl E., Pfitzner R., Scholtes I. et al. Predicting scientific success based on coauthorship networks // EPJ Data Science. 2014. No. 3. P. 9.