

АКАДЕМИЧЕСКАЯ ГЕНЕАЛОГИЯ КАК МЕТОД АНАЛИЗА ЭВОЛЮЦИИ ИССЛЕДОВАТЕЛЬСКИХ НАПРАВЛЕНИЙ В СФЕРЕ ОБРАЗОВАНИЯ

С. Г. Григорьев ¹⁾, А. Х. Мариносян ²⁾

¹⁾ *Московский городской педагогический университет, Россия, Москва,
grigorsg@yandex.ru*

²⁾ *Московский городской педагогический университет, Россия, Москва,
a.marinosyan@yandex.ru*

В докладе представлена методология количественного сравнения тематических профилей научных школ, позволяющая определить принадлежность исследуемых школ к единому научно-исследовательскому направлению. Подход основан на семантическом анализе корпусов диссертационных работ с использованием векторизации, кластеризации и последующем сравнении тематических профилей школ во времени с помощью набора статистических и информационных метрик.

Ключевые слова: научная школа; академическая генеалогия; концептуальное сходство; сравнительный анализ; семантический анализ; кластеризация; статистические метрики; информационные метрики; диссертации; информатизация образования.

ACADEMIC GENEALOGY AS A METHOD FOR ANALYZING THE EVOLUTION OF RESEARCH DIRECTIONS IN EDUCATION

S. G. Grigoriev ¹⁾, A. K. Marinosyan ²⁾

¹⁾ *Moscow City University, Russia, Moscow, grigorsg@yandex.ru*

²⁾ *Moscow City University, Russia, Moscow, a.marinosyan@yandex.ru*

The report presents a methodology for quantitative comparison of thematic profiles of scientific schools, allowing to determine whether the studied schools belong to a unified research direction. The approach is based on semantic analysis of dissertation corpora using vectorization, clustering, and subsequent comparison of schools' thematic profiles over time using a set of statistical and information metrics.

Keywords: scientific school; academic genealogy; conceptual similarity; comparative analysis; semantic analysis; clustering; statistical metrics; information metrics; dissertations; informatization of education.

Введение

Современная научная деятельность характеризуется ростом объемов публикаций, расширением междисциплинарных исследований и усложнением структуры научного знания. В этих условиях актуальность приобретают подходы, позволяющие эффективно структурировать и анализировать научную информацию. Одним из таких подходов является академическая генеалогия – междисциплинарная область исследований, изучающая структуру, динамику и эволюцию науки в контексте отношений научного руководства и интеллектуальной преемственности [1].

Исторически предметом анализа в рамках академической генеалогий были формальные связи между научным руководителем и учеником, визуализируемые в виде графов научного руководства, где узлами выступают ученые, а ребрами – подтвержденные факты защиты диссертации под руководством. Успешными примерами являются проекты Mathematics Genealogy Project и Neurotree [2; 3]. Нами были проведена работа по построению деревьев научного руководства для школы академика РАО А.А. Кузнецова [4].

Однако формальные связи не отражают всю сложность интеллектуального взаимодействия, поскольку ученые испытывают влияние не только своих непосредственных руководителей, но и более широкого круга исследователей. Концептуальные связи, в отличие от формальных, основаны на содержательном анализе научных работ и позволяют выявить преемственность идей, методов и теоретических подходов независимо от наличия формальных отношений между авторами.

Проблема заключается в том, что научные школы, объединенные фигурой общего основателя, могут значительно различаться по тематике исследований, а школы с разными основателями могут демонстрировать концептуальное сходство. Это обуславливает необходимость разработки *методов, позволяющих идентифицировать и объединять концептуально близкие исследования в единые научно-исследовательские направления.*

Методология исследования

Разработанный метод имеет целью количественное сравнение тематической эволюции научных школ для выявления степени их сходства, различия и идентификации общих исследовательских направлений. Сравнение тематических профилей научных школ с использованием статистических методов и информационных метрик позволяет оценить их близость и проследить динамику их взаимоотношений во времени. Школы со статистически схожими и стабильно близкими распределениями

по тематике диссертационных работ могут рассматриваться как составляющие одного исследовательского направления.

Предлагаемая методология включает следующие шаги. На первом этапе идентифицируются научные школы, как правило, объединенные фигурой основателя или ведущего научного руководителя. На втором этапе собирается корпус авторефератов диссертационных работ, защищенных в рамках данных школ за определенный период. На третьем этапе проводится содержательный анализ и категоризация тематики диссертаций. Выделяются ключевые элементы, отражающие суть исследования (название, актуальность, новизна, предмет, объект и т. д.). После производится векторизация каждого элемента с помощью семантических эмбедингов и затем конкатенация этих векторов для получения общего вектора работы, который затем используется для кластеризации и присвоения категорий (один кластер рассматривается как одна категория). Названия категорий присваиваются экспертами. Важно отметить, что категории должны быть унифицированы для всех сравниваемых научных школ.

На четвертом этапе рассчитывается процентное распределение работ по выделенным категориям для каждой школы – как за весь период исследования, так и для отдельных временных интервалов (например, пятилетий), чтобы отследить динамику. Пятый, ключевой этап – сравнение полученных распределений между школами с использованием набора статистических методов и метрик. Для проверки гипотезы о сходстве распределений применяются критерии согласия, такие как хи-квадрат Пирсона и G-критерий отношения правдоподобия. Эти критерии позволяют оценить статистическую значимость различий между наблюдаемыми частотами категорий в разных школах. Для количественной оценки силы связи и степени сходства распределений используются коэффициент V Крамера и коэффициент ранговой корреляции Спирмена. Кроме того, для оценки «расстояния» между распределениями как вероятностными объектами применяются метрики из теории информации: расстояние Кульбака-Лейблера, дивергенция Йенсена-Шеннона и дистанция Хеллингера. Эти метрики дают интегральную оценку различий между тематическими профилями школ. Анализ этих показателей в динамике позволяет выявить периоды сближения или расхождения научных интересов школ. Дополнительно, для разделения влияния факторов «научная школа» и «временной период» может быть применен двухфакторный дисперсионный анализ (ANOVA).

Результаты и их обсуждение

Для иллюстрации предложенной методологии был проведен сравнительный анализ тематических профилей диссертационных исследований научных школ Александра Андреевича Кузнецова (102 работы) и Сергея Григорьевича Григорьева (151 работа) за период с 1995 по 2024 год. Было выделено 8 тематических категорий, а весь период разбит на 6 пятилетних подпериодов. Применение критериев согласия (хи-квадрат и G-критерий) показало, что при рассмотрении всего периода исследования (1995–2024 годы) статистически значимых различий в распределении категорий между школами не выявлено (p -значения > 0.05). Это подтверждается и высоким значением косинусного сходства (0.9307) за весь период, указывающим на значительную общность в тематической направленности. Однако анализ по пятилетним периодам выявил более сложную динамику. Единственный период, где наблюдались статистически значимые различия (по G-критерию, $p = 0.0357$), – это ранний этап 1995–1999 годы. В этот период также зафиксировано самое низкое косинусное сходство, отрицательная ранговая корреляция Спирмена и максимальные значения расстояний по информационным метрикам, что свидетельствует о существенном расхождении в структуре тематических приоритетов на начальном этапе развития школ. В последующие периоды наблюдались колебания степени сходства. Например, периоды 2010–2014 и 2020–2024 годы характеризовались высокой степенью сходства по большинству метрик (низкие информационные расстояния, высокие значения косинусного сходства и корреляции Спирмена).

Анализ корреляций между различными метриками показал высокую согласованность между информационными метриками ($r > 0.9$). Двухфакторный дисперсионный анализ подтвердил отсутствие статистически значимого влияния фактора «научная школа» или «временной период», а также их взаимодействия, на общую структуру распределения долей работ по категориям, если рассматривать все данные вместе. Это говорит о схожей (статистически неразличимой) общей динамике развития тематических профилей двух школ.

Таким образом, можно сделать вывод, что научные школы А.А. Кузнецова и С.Г. Григорьева демонстрируют значительную общность тематических интересов в области информатизации образования на протяжении почти трех десятилетий, что позволяет рассматривать их как части *единого исследовательского направления*. Тем не менее выявленные флуктуации сходства, особенно выраженное расхождение на раннем

этапе и некоторые различия в отдельные периоды, подчеркивают наличие собственной специфики и динамики развития у каждой школы.

Представленный подход, интегрирующий академическую генеалогию с количественным анализом содержания исследований, является перспективным инструментом для изучения тенденций развития научных направлений в образовании. Он позволяет не только констатировать формальные связи, но и оценивать реальную тематическую близость научных школ и ее изменение во времени. Дальнейшее развитие методологии может включать ее масштабирование для анализа большего числа школ.

Библиографические ссылки

1. *Лернер И. М., Мариносян А. Х., Григорьев С. Г., Юсупов А. Р., Аникьева М. А., Гарифуллина Г. А.* Подход к формированию интеллектуальной академической генеалогии с использованием больших языковых моделей // *Электромагнитные волны и электронные системы*. 2024. Т. 29. № 4. С. 108–120.
2. *Mulcahy C.* The Mathematics Genealogy Project comes of age at twenty-one // *Notices of the AMS*. 2017. Vol. 64. No. 5. P. 466–470.
3. *David S. V., Hayden B. Y.* Neurotree: A collaborative, graphical database of the academic genealogy of neuroscience // *PloS One*. 2012. Vol. 7. No. 10. e46608.
4. *Григорьев С. Г., Мариносян А. Х., Лернер И. М.* Цифровые методы в анализе исследований в области образования: научная школа А. А. Кузнецова по информатизации образования // *От информатики в школе к цифровой трансформации образования : материалы науч.-практ. конф., Москва, 25 окт. 2024 г. / под науч. ред. Л. Л. Босовой. М. : Рос. акад. образования, 2024. С. 183–188*