

Прикладная эконометрика, 2016, т. 44, с. 100–118.

Applied Econometrics, 2016, v. 44, pp. 100–118.

Н. Н. Матвеева, **О. В. Польдин**¹

Цитируемость ученых и сети научного соавторства: анализ данных Google Scholar

В работе анализируются корреляции между параметрами сети соавторства и библиометрическими характеристиками ученых в Google Scholar. Оценивание проводилось с помощью моделей счетных данных по выборке, состоящей из более чем 30 тысяч авторов с первым цитированием после 2007 года. Найдена положительная связь между цитируемостью ученого и числом соавторов, цитируемостью и центральностью автора в сети, цитируемостью ученого и средней цитируемостью его соавторов. Индекс Хирша и индекс $i10$ значимо связаны с числом соавторов и их средней цитируемостью.

Ключевые слова: наукометрия; сети соавторства; библиометрический анализ; Google Scholar; модели счетных данных.

JEL classification: A14; D83; Z13.

1. Введение

Одним из основных и актуальных вопросов, которые интересуют современную «науку о науке», является связь между библиометрическими показателями ученых и их вовлеченностью в сотрудничество с другими исследователями (Persson et al., 2004). Несмотря на то что строгие математические исследования в этом направлении начались сравнительно недавно, сами ученые уже давно заметили, что расширение контактов и сотрудничества повышает их «научный выход». Огромное число конференций проходит с целью не только обменяться информацией, но и интенсифицировать взаимодействия между научными группами.

Взаимодействие между учеными является неотъемлемой чертой научной деятельности. Одной из форм взаимодействия является работа над научной проблемой нескольких исследователей. В процессе сотрудничества между участниками происходит обмен идеями, обсуждение проблем, результатов, генерация новых идей. Такого рода взаимодействие способно повысить результативность последующих исследований ученых. Таким образом, социальные взаимодействия выступают важным ресурсом ученого и влияют на его положение в профессиональном сообществе.

¹ Матвеева Наталья Николаевна — Национальный исследовательский университет «Высшая школа экономики», Нижний Новгород; Nata_Plast@mail.ru.

Польдин Олег Викторович — Национальный исследовательский университет «Высшая школа экономики», Нижний Новгород; opoldin@hse.ru. Скончался 29.10.2016. Редакция журнала скорбит об утрате и выражает искренние соболезнования родным и близким Олега Викторовича.

Значимость и научный вклад публикаций принято оценивать таким параметром, как цитируемость — общее число ссылок на данную публикацию в других работах. В связи с этим, актуальным является вопрос, какие факторы влияют на показатели цитируемости ученого и как их можно оценить.

С появлением интернета, электронных баз данных научного цитирования и научных поисковых систем, таких как Web of Science, Scopus, Microsoft Academic Search, Google Scholar и других, задача сбора информации об авторах и их публикациях значительно упростилась, после чего исследования в этом направлении существенно интенсифицировались (Meho, Yang, 2007; Kulkarni et al., 2009; Abbasi et al., 2011; Li et al., 2013; Uddin et al., 2013; Orduña-Malea et al., 2015). Стоит отметить, что каждая электронная база данных имеет свою специфику, что необходимо учитывать при анализе используемых данных, в том числе, анализе сетей соавторства. Основные различия представленных данных связаны с количеством индексируемых изданий, охватом научных областей, способами анализа библиометрических характеристик, доступностью (Jacsó, 2005; Bakkalbasi et al., 2006; Yang, Meho, 2006; Kulkarni et al., 2009; Бедный, Сорокин, 2012; Bensman, 2013; Orduña-Malea et al., 2015; Ortega, 2015).

Представляемое исследование основано на анализе базы данных Google Scholar. В отличие от Web of Science и Scopus, Google Scholar имеет открытый доступ, индексирует большее число источников, содержит поддерживаемые авторами профили пользователей, однако не все индексируемые издания являются научными.

На сегодняшний день непосредственно на базе профилей сети Google Scholar проведено немного исследований. Возможными причинами этого являются ошибки идентификации авторов и их работ, наличие «фантомных» авторов, а также неправильная идентификация года публикации, что, безусловно, влияет на показатели цитируемости (Jacsó, 2008a). Несмотря на имеющиеся недостатки, стоит заметить, что библиометрические индексы, рассчитанные на основе Google Scholar, сопоставимы с аналогичными показателями, рассчитанными на основе Web of Science с традиционными импакт-факторами (Harzing, van der Wal, 2008; Wildgaard, 2015).

В данной работе исследуется связь сетевых характеристик ученого, таких как число соавторов (центральность по степени), среднее число соавторов и центральность по близости с его библиометрическими показателями: общим количеством цитирований, индексом Хирша, индексом *i10*. Используемые сетевые характеристики представляют собой метрики сети соавторства, построенной на основе анализа профилей ученых базы данных Google Scholar.

Сеть соавторства представляет собой граф, в котором узлами являются авторы, а связь между узлами — это научное сотрудничество, выраженное, как правило, в совместных статьях. Сети соавторства ученых позволяют не только визуализировать взаимодействия научного сообщества, но и определить тенденции его развития, ключевых участников, пути распространения информации.

В работах (Heffner, 1981; Kretschmer, 1994; Katz, Martin, 1997; Newman, 2001; Girvan, Newman, 2002; Glanzel, 2002; Newman, 2004) исследуются способы построения и виды сетей соавторства, а также представлены основные способы расчета метрик сети, которые используются в настоящей работе.

В рамках анализа сетей соавторства стоит уделять внимание поиску закономерностей и различий между метриками сети разных научных областей, поскольку параметры сети соавторства ученых могут различаться в зависимости от принадлежности авторов к той или иной научной области (Kretschmer, 1994).

Сетевой анализ соавторства позволяет выявить тенденции и законы взаимодействия участников сети, отражая тем самым формы социального взаимодействия между авторами. Социальное взаимодействие, в свою очередь, может выступать фактором, влияющим на результативность научной деятельности. Вопрос оценки эффективности научной деятельности в настоящее время весьма актуален, поскольку, если в долгосрочной перспективе важность тех или иных результатов научных исследований уже можно оценить, например, по их влиянию на экономическое развитие общества и рост уровня жизни, то в краткосрочной перспективе такое влияние может быть еще не очевидным. Основные способы оценки эффективности научной деятельности представлены в работах (Hirsch, 2005; Egghe, 2006; Bergstrom et al., 2008; Bornmann, Daniel, 2008; West et al., 2010; Kreiman, Maunsell, 2011; Fooladi et al., 2013; Писляков, 2010).

Количественная оценка значимости научных результатов, как правило, базируется на индексе цитируемости публикаций, в которых эти результаты опубликованы. В работах (Glänzel, Schubert, 2004; Newman, 2004; Abbasi et al., 2011; Méndez-Vásquez et al., 2012; Uddin et al., 2013) исследуется наличие зависимости между сетевыми характеристиками и эффективностью научной деятельности, выраженной, как правило, в цитируемости и зависимых от нее параметров (индекс Хирша, индекс $i10$).

Ограничения предыдущих исследований связи библиометрических и сетевых характеристик ученого связаны с использованием небольшого числа соавторов из разных областей науки (Zuckerman, 1967) или же данных об ученых одной научной сферы (Avkiran, 2013; Uddin et al., 2013; Puuska et al., 2014; Yu et al., 2014), что в обоих случаях влияет на достоверность выводов. Также на качество анализируемых данных влияет использование для построения сети соавторов информации о статьях, а не о самих авторах, поскольку возникает проблема поиска соответствия между авторами с одинаковыми фамилиями и их работами (Méndez-Vásquez et al., 2012). В настоящей статье удалось избежать данных ограничений. Специфика Google Scholar позволяет использовать данные об ученых и их соавторах, не прибегая для этого к анализу разрозненных записей об отдельных публикациях или к перекрестному анализу других библиографических баз данных. Непосредственный анализ профилей ученых позволяет избежать проблем при нахождении соответствий авторов и их работ, которые возникают в случае формирования базы данных по отдельным публикациям.

Научная новизна представленной работы заключается в использовании объема данных, значительно превосходящего объемы данных в аналогичных публикациях. Исследование впервые основано на анализе данных профилей Google Scholar, с учетом разделения ученых по научным специализациям и стажу. Кроме того, насколько известно авторам, впервые изучается наличие статистической связи между показателями цитируемости ученого и средней цитируемостью его соавторов.

2. Контекст исследования и данные

Google Scholar

В настоящей работе исследуется связь сетевых характеристик ученого и его библиометрических показателей (цитируемости, индекса Хирша, индекса $i10$) с учетом влияния стажа цитируемости. Исследование основано на анализе профилей ученых базы данных Google

Scholar. Для этой цели было использовано 110 тысяч профилей ученых из различных стран и научных областей, на базе которых была построена сеть соавторства.

Google Scholar (GS) является свободно доступной поисковой системой, которая обеспечивает полнотекстовый поиск научных публикаций всех форматов и по всем научным дисциплинам. Система работает с ноября 2004 года. Помимо классических библиометрических возможностей, Google Scholar содержит элементы социальной сети, т. к. позволяет пользователям регистрировать свои научные профили, оптимизируя индексацию собственных научных публикаций. Зарегистрированные пользователи имеют возможность определять круг своих соавторов в специальном разделе личной страницы.

В отличие от Web of Science и Scopus, Google Scholar является более удобным прототипом сети соавторства благодаря: 1) существующей системе организации структуры профилей участников; 2) индексации большего числа научных источников; 3) бесплатному доступу и сравнительной простоте извлечения информации с помощью программ-парсеров веб-страниц.

К недостаткам Google Scholar можно отнести индексирование ненаучных источников (статьи в нереферируемых журналах, блоги, презентации), ошибки в идентификации авторов и их работ и, как следствие, неправильные значения показателей цитируемости (Jacsó, 2008a, 2008b). Помимо этого, существуют некоторые ограничения в структуре профилей GS: поля должности, аффилиации и научных интересов не обязательны к заполнению; поле соавторов у некоторых профилей может быть не заполнено; ошибка индексации ученых с китайскими (и не только) фамилиями («перепутывание» таких профилей); нет отслеживания карьеры и смены аффилиации ученых. Однако стоит отметить, что и для Web of Science и для Scopus последние две проблемы также стоят довольно остро.

Несмотря на недостатки системы профилей Google Scholar, стоит отметить, что: 1) профили Google Scholar достаточно полно и автоматически отображают библиометрические показатели ученых; 2) для профилей с незаполненным списком соавтором существует возможность восстановить этот список по другим профилям участников сети.

Показатели цитируемости

Цитируемость статьи или ученого представляет собой показатель, равный числу ссылок на данную статью или ученого. В отличие от числа публикаций, цитируемость учитывает качественный аспект — считается, что более качественные работы цитируются чаще (Garfield, 1979). К недостаткам цитируемости как меры для сравнения ученых можно отнести зависимость данного показателя от области знаний и возраста публикации. Кроме того, упоминание чужой работы бывает и в критическом ключе, тогда это обусловлено не ее качеством, а, наоборот, недостатками (Toutkoushian, Webber, 2011). Показатель цитируемости также варьируется в зависимости от используемой библиографической базы (Bakkalbasi et al., 2006).

Индекс Хирша является популярной количественной характеристикой продуктивности ученого, т. к. он одновременно учитывает как число статей, так и их цитируемость. Ученый имеет индекс h , если h из его N статей цитируются как минимум h раз каждая, в то время как оставшиеся $N-h$ статей цитируются не более чем h раз каждая (Hirsch, 2005). Основным недостатком данного индекса является то, что он не отражает число высокоцитируемых, т. е. значимых, работ.

Помимо индекса Хирша, в Google Scholar используется индекс *i10*, который указывает число научных публикаций, имеющих как минимум 10 цитирований. Этот индекс введен в 2011 году специально для Google Scholar. К его достоинствам можно отнести более точное (в сравнении с индексом Хирша) отражение числа высокоцитируемых (имеющих как минимум десять цитирований) работ, а также простоту подсчета, а к недостаткам — локальность использования.

Для подсчета показателей цитируемости GS, в отличие от Web of Science и Scopus, учитывает не только журнальные статьи и англоязычные работы, но и большое число публикаций другого вида, что увеличивает значения показателей. Однако существует высокая корреляция между рейтингами цитируемости, составленными на основе этих баз данных (Franceschet, 2009; Delgado, Repiso, 2013; Wildgaard, 2015).

Профиль ученого в GS содержит следующую информацию: публикации, занимаемая позиция и аффилиация, научные интересы и домен, данные о числе цитирований, индексе Хирша и индексе *i10* за все время и за последние пять лет, список зарегистрированных в системе и указанных данной персоной соавторов.

Сетевые характеристики

Сетевые характеристики представляют собой метрики графа соавторства ученых, построенного на основе данных профилей ученых в Google Scholar.

Для расчета сетевых характеристик были использованы следующие параметры: число соавторов (degree centrality), среднее число связей для разных выборок, положение автора в сети относительно других участников, измеренное с помощью центральности по близости (closeness centrality). Центральность по близости измеряет, насколько близко конкретный участник расположен к остальным узлам сети и насколько легко его достичь².

В качестве стартового листа были использованы 12 профилей российских исследователей, затем была собрана информация об их соавторах, соавторах соавторов и так далее до 110 тысяч. На основе полученных данных была построена матрица смежности, определены кратчайшие пути и сетевые параметры: число соавторов, центральность по близости, средняя цитируемость всех соавторов. Также по ключевым словам научных направлений были выделены группы профилей, в которых обозначена одна из следующих специализаций автора: компьютерные науки (computer science), экономика и финансы (economics and finance), биология и медицина (biology and medicine), физика и химия (physics and chemistry), математика (mathematics), социально-гуманитарные (social sciences and humanities).

Оценивание эмпирических моделей проводилось для 34 тыс. профилей из сети в 110 тыс. профилей, но с учетом связей со всеми участниками сети. Ограничение вызвано тем, что в число объясняющих переменных включен стаж цитируемости ученого. Из-за особенностей Google Scholar информация о стаже может быть извлечена только для авторов с первым цитированием после 2007 года.

² Центральность по близости определяется формулой $C_i = (N - 1)^{-1} \left(\sum_j d_{ij}^{-1} \right)$, где C_i — центральность i -го участника; N — общее число участников; d_{ij} — кратчайшее расстояние в графе от участника i до участника j .

Описательные статистики

В таблице 1 представлены средние значения и стандартные отклонения переменных, подсчитанные как по полной выборке из 34701 авторов, так и по подвыборкам из ученых с однородной научной специализацией. Наибольшие показатели цитируемости, индексы Хирша и *i10* наблюдаются в отраслях биомедицины, физики и химии, наименьшие — в экономике и финансах. Среднее число соавторов для всех дисциплин — 6 человек, выше среднего данный показатель в естественных науках, ниже среднего — в экономике и финансах. На основе исследуемой выборки был подтвержден результат, полученный Ortega (2015), что в сети Google Scholar доминируют ученые, в той или иной степени занятые в области вычислительных наук.

Таблица 1. Описательные статистики: среднее значение (стандартное отклонение)

	Полная выборка	Специализация					
		BM	CS	EF	Math	PhCh	SH
Цитируемость	333.9 (1041.9)	542.8 (1529.2)	262.1 (873.4)	234.1 (414.0)	266.0 (504.3)	470.3 (1153.3)	307.6 (611.7)
Индекс Хирша	6.375 (4.942)	7.604 (5.624)	5.765 (4.299)	5.714 (3.834)	6.520 (4.264)	7.980 (5.832)	6.542 (4.755)
Индекс <i>i10</i>	5.855 (9.265)	7.773 (10.33)	4.828 (7.324)	4.748 (5.720)	5.880 (7.249)	8.432 (11.98)	6.113 (8.597)
Число соавторов	6.032 (6.619)	5.933 (6.343)	6.655 (7.597)	4.068 (3.754)	6.106 (5.668)	5.928 (5.394)	5.913 (6.521)
Средняя цитируемость соавторов	4285.2 (5728.4)	6355.0 (7630.2)	3705.9 (4498.8)	4209.0 (6239.2)	3984.9 (4600.6)	5060.1 (5720.8)	4019.3 (5194.9)
Публикационный стаж	4.928 (1.725)	5.111 (1.651)	4.758 (1.746)	5.221 (1.672)	5.088 (1.669)	5.041 (1.661)	5.081 (1.671)
Логарифм публикационного стажа	1.508 (0.466)	1.557 (0.431)	1.466 (0.482)	1.573 (0.454)	1.550 (0.439)	1.542 (0.429)	1.548 (0.438)
Центральность по близости	0.185 (0.0163)	0.185 (0.0141)	0.190 (0.0136)	0.165 (0.0188)	0.185 (0.0154)	0.178 (0.0134)	0.180 (0.0186)
Число наблюдений	34701	1878	6514	1004	1247	1353	2259

Примечание. BM — biology and medicine, CS — computer science, EF — economics and finance, Math — mathematics, PhCh — physics and chemistry, SH — social sciences and humanities.

Стоит отметить, что, в отличие от параметра цитируемости, такие параметры, как число соавторов или центральность, существенно зависят от того, какая часть сети соавторов оказывается за пределами полученных данных. В представленном исследовании для анализа использовались данные связанной сети соавторства, что позволило свести к минимуму смещение оценок сетевых характеристик.

Важной переменной, влияющей на показатели цитируемости, является исследовательский стаж автора. Поскольку присутствие в выборке авторов с разной продолжительностью карьеры влияет на их библиометрические показатели (Ductor et al., 2014), авторы разделили имеющуюся выборку по годам первого цитирования, чтобы избежать влияния временного

фактора на оценки связи сетевых и библиометрических параметров. Помимо этого, такой подход позволит проверить гипотезу о наличии влияния продуктивных авторов на библиометрические показатели их молодых коллег. К примеру, наличие в сети соавторства профессоров с высокой цитируемостью и большой сетью соавторов может влиять на показатели цитируемости и центральности их аспирантов.

3. Эмпирические модели

В данном исследовании зависимые переменные по своей природе измеряют число цитирований и являются неотрицательными целыми числами. Для регрессионного анализа подобных случаев часто применяются модели со счетными зависимыми переменными (count data models). Здесь используются варианты модели с отрицательным биномиальным распределением (negative binomial model), которая может рассматриваться как обобщение пуассоновской модели (Cameron, Trivedi, 2013). В частности, модель с отрицательным биномиальным распределением обладает большей гибкостью для описания дисперсии зависимой переменной и лучше подходит для моделирования ситуаций, в которых отдельные факты цитирования являются взаимосвязанными событиями (Ajiferuke, Famoye, 2015).

Модель с отрицательным биномиальным распределением

Переменная, распределенная по пуассоновскому закону

$$Y \sim \text{Poisson}(\mu), \quad (1)$$

имеет функцию вероятности

$$f(y|\mu) = \frac{e^{-\mu} \mu^y}{y!}, \quad y = 0, 1, 2, \dots, \quad (2)$$

так что ее среднее значение и дисперсия равны друг другу:

$$E(Y) = \mu, \quad \sigma^2(Y) = \mu. \quad (3)$$

В действительности дисперсия индексов цитирования превосходит их математическое ожидание, поэтому пуассоновская модель некорректна. Этот недостаток можно исправить, введя параметризацию дисперсии. Например, если имеются две случайные величины Y и v с распределениями

$$Y \sim \text{Poisson}(\mu v), \quad v \sim \text{Gamma}(1, \alpha), \quad (4)$$

где $\alpha > 0$, то переменная Y будет распределена по отрицательному биномиальному закону

$$Y \sim \text{Negative Binomial}(\mu, \alpha) \quad (5)$$

с математическим ожиданием и дисперсией

$$E(Y) = \mu, \quad \sigma^2(Y) = \mu + \alpha \mu^2. \quad (6)$$

Функция вероятности переменной с отрицательным биномиальным распределением имеет вид

$$f(y|\mu, \alpha) = \frac{\Gamma(\alpha^{-1} + y)}{\Gamma(\alpha^{-1})\Gamma(1 + y)} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu} \right)^{\frac{1}{\alpha}} \left(\frac{\mu}{\alpha^{-1} + \mu} \right)^y, \quad y = 0, 1, 2, \dots, \quad (7)$$

где $\Gamma(\cdot)$ — гамма-функция.

В регрессионной модели предполагается, что математическое ожидание и дисперсия зависимой переменной зависят от объясняющих переменных через экспоненциальную функцию

$$\mu = \exp(x'\beta), \quad \sigma^2(y) = \exp(x'\beta) + \alpha(\exp(x'\beta))^2. \quad (8)$$

Поскольку авторы с нулевой цитируемостью не рассматриваются, для случаев, когда зависимой переменной y является цитируемость, выборка оказывается усеченной, содержащей только наблюдения с $y \geq 1$. Условная функция вероятности с исключенными нулевыми значениями зависимой переменной имеет вид

$$f(y|y \geq 1) = \frac{f(y)}{1 - f(0)}, \quad y = 1, 2, \dots, \quad (9)$$

где $f(y)$ — функция вероятности с включенными нулевыми значениями. Данная корректировка учитывается в моделях с усеченными нулями (zero-truncated models).

Двухчастная модель

Индекс *i10* оценивает число публикаций автора, преодолевших порог в 10 цитирований. Поскольку не все авторы имеют такие публикации, для эмпирического моделирования данного индекса использовались модели с барьерами (hurdle models). Эти модели относятся к классу двухчастных моделей, т. е. состоящих из двух компонентов. Первый компонент описывает распределение нулевых значений зависимой переменной, второй — ненулевых. Совместное распределение имеет вид

$$f(y) = \begin{cases} f_1(0), & \text{если } y = 0, \\ \frac{1 - f_1(0)}{1 - f_2(0)} f_2(y), & \text{если } y \geq 1. \end{cases} \quad (10)$$

С точки зрения регрессионных оценок оцениваются два уравнения. В нашем случае первое уравнение представляет собой логистическую модель бинарного выбора для вероятности автора достичь ненулевого индекса *i10*. Второй компонент моделирует поведение индекса при ненулевых значениях с помощью отрицательной биномиальной модели с усеченными нулями.

4. Результаты

Для анализа корреляционных связей показателя цитируемости использовалась модель с отрицательным биномиальным распределением и исключенными нулевыми значениями (цитируемости). Результаты оценивания трех спецификаций модели для полной выборки приведены в табл. 2. Спецификации различаются набором объясняющих переменных. В спецификации 1 характеристикой положения автора в сети выступает число его соавторов, в спецификации 2 — центральность по близости, в спецификации 3 — оба сетевых параметра. Общими переменными во всех спецификациях являются средняя цитируемость соавторов и логарифм стажа цитируемости автора. Как следует из сопоставления результатов, центральность по близости играет второстепенную роль относительно числа соавторов. В полной спецификации (3) центральность по близости незначима, но становится значимой при исключении числа соавторов, принимая на себя роль переменной, характеризующей положение автора в сети. Оценки для спецификации без центральности по близости (1) и полной спецификации (3) очень близки. Средняя цитируемость соавторов и стаж значимы на 1%-ном уровне. Зависимость средней цитируемости от стажа примерно квадратическая, однако нужно учитывать, что в выборке присутствуют авторы со стажем не более 7 лет, поэтому данная зависимость является аппроксимацией для авторов на начальном этапе карьеры.

Таблица 2. Регрессионные оценки для показателя цитируемости

	Спецификации		
	(1)	(2)	(3)
Средняя цитируемость соавторов $\times 10^{-4}$	0.854*** (0.033)	0.817*** (0.035)	0.862*** (0.035)
Число соавторов	0.061*** (0.002)		0.063*** (0.003)
Центральность по близости		14.84*** (0.726)	-1.066 (1.115)
Стаж (логарифм)	2.109*** (0.040)	2.247*** (0.036)	2.108*** (0.040)
Константа	1.307*** (0.0634)	-1.212*** (0.153)	1.491*** (0.213)
$\ln(\alpha)$	0.160*** (0.021)	0.240*** (0.018)	0.160*** (0.022)
Число наблюдений	34 701	34 701	34 701
Псевдо- R^2	0.0543	0.0482	0.0543

Примечание. В скобках указаны стандартные погрешности оценок; *** — значимость на уровне 1%.

С точки зрения интерпретации полученных оценок удобнее рассматривать не сами коэффициенты при переменных, находящихся в показателе экспоненциальной функции в модели (4), а изменения зависимой переменной при приращении объясняющих факторов. Поскольку объясняющие переменные различаются по масштабу, рассматривались эффекты от дискретного приращения переменных от своих средних значений на одно стандартное отклонение, при этом остальные переменные были зафиксированы на уровне средних

значений. В таблице 3 приведены математические ожидания эффектов и границы 95%-ных доверительных интервалов как для полной выборки, так и отдельных специализаций.

Таблица 3. Изменение цитируемости при увеличении независимых переменных на 1 ст. отклонение

	Δy	Нижняя граница	Верхняя граница	$\Delta y/\sigma$
<i>Полная выборка</i>				
Средняя цитируемость соавторов	116.8	103.9	129.7	0.11
Число соавторов	92.4	86.1	98.8	0.09
Стаж (логарифм)	309.9	292.7	327.0	0.30
<i>Computer science</i>				
Средняя цитируемость соавторов	65.5	49.1	81.9	0.07
Число соавторов	105.8	91.6	120.1	0.12
Стаж (логарифм)	223.4	206.1	240.6	0.26
<i>Economics and finance</i>				
Средняя цитируемость соавторов	24.7	6.5	42.8	0.06
Число соавторов	71.6	55.3	88.0	0.17
Стаж (логарифм)	266.7	231.0	302.4	0.64
<i>Biology and medicine</i>				
Средняя цитируемость соавторов	176.2	123.5	228.8	0.12
Число соавторов	156.5	117.4	195.5	0.10
Стаж (логарифм)	483.4	416.6	550.2	0.32
<i>Physics and chemistry</i>				
Средняя цитируемость соавторов	124.4	85.6	163.2	0.11
Число соавторов	112.0	81.5	142.5	0.10
Стаж (логарифм)	420.8	355.3	486.4	0.36
<i>Mathematics</i>				
Средняя цитируемость соавторов	86.8	44.6	129.1	0.17
Число соавторов	80.1	61.5	98.7	0.16
Стаж (логарифм)	260.4	218.9	302.0	0.52
<i>Social sciences and humanities</i>				
Средняя цитируемость соавторов	93.4	46.5	140.2	0.15
Число соавторов	88.2	69.2	107.2	0.14
Стаж (логарифм)	223.1	137.1	309.0	0.36

В полной выборке увеличение средней цитируемости соавторов на стандартное отклонение, равное 5728, ассоциируется с увеличением цитируемости на 116.8 единиц, при росте числа соавторов с 6 до 13 человек цитируемость растет на 92.4. Из рассмотренных специализаций наибольший абсолютный эффект наблюдается для авторов, работающих в области биологии и медицины, а наименьший — в области экономики и финансов. Однако среднее значение и дисперсия цитируемости для биологии и медицины значительно выше таковых для экономики и финансов. В последнем столбце табл. 3 абсолютные эффекты нормированы на стандартные отклонения цитируемости в соответствующей выборке. Приведенные

цифры показывают, на сколько стандартных отклонений изменяется цитируемость при росте объясняющего фактора на одно стандартное отклонение. В этих терминах средняя цитируемость соавторов имеет наибольший эффект для специализации в области математики, число соавторов — для экономики и финансов. Количественно оценки эффекта числа соавторов варьируются от 0.09 до 0.17, эффект цитируемости соавторов — от 0.06 до 0.17.

В таблице 4 содержатся регрессионные оценки для индекса Хирша в полной выборке. Оценивалась обычная модель с отрицательным биномиальным распределением без исключения нулевых значений индекса. В качестве объясняющих переменных использовались средняя цитируемость соавторов, число соавторов и логарифм стажа цитируемости автора. Все переменные статистически значимы. Эффект числа соавторов примерно в два раза превосходит эффект цитируемости соавторов, в то время как для цитируемости по полной выборке (из табл. 3) величины этих эффектов сопоставимы. Другое отличие касается корреляции с публикационным стажем: для индекса Хирша связь практически линейная (для цитируемости — квадратичная), но в целом нормализованный эффект стажа велик (0.66) и в два раза превосходит таковой для цитируемости (0.30).

Таблица 4. Регрессионные оценки для индекса Хирша

	Коэффициент	Δy при изменении Δx на 1 ст. отклонение			
		Δy	Нижняя граница	Верхняя граница	Δy/σ
Средняя цитируемость соавторов × 10 ⁻⁴	0.173*** (0.010)	0.58	0.52	0.65	0.12
Число соавторов	0.031*** (0.001)	1.25	1.18	1.32	0.25
Стаж (логарифм)	0.985*** (0.008)	3.26	3.20	3.32	0.66
Константа	-0.021* (0.012)				
ln(α)	-2.246*** (0.035)				
Число наблюдений	34 701				
Псевдо-R ²	0.123				

Примечание. В скобках указаны стандартные погрешности оценок; *, *** — значимость на уровне 10 и 1% соответственно.

Оценки двухчастной модели с барьерами для индекса *i10* показаны в табл. 5. Первая часть представляет собой логистическую модель бинарного выбора, в которой зависимой переменной является вероятность того, что индекс *i10* превышает нулевое значение, т. е. у автора есть как минимум одна публикация, имеющая 10 и более цитирований. Увеличение средней цитируемости соавторов на одно стандартное отклонение связано с ростом вероятности преодолеть порог на 2.6%, а числа соавторов — на 6%. Во второй части зависимой переменной является само значение индекса *i10* при условии, что оно не нулевое. Оценивалась модель с отрицательным биномиальным распределением и исключенными нулевыми значениями. Как видно из последнего столбца, нормализованный эффект от дискретного изменения средней цитируемости соавторов (0.12) заметно слабее, чем от изменения числа соавторов (0.22).

Таблица 5. Регрессионные оценки для индекса $i10$

	Коэффициент	Δу при изменении Δх на 1 ст. отклонение			
		Δу	Нижняя граница	Верхняя граница	Δу/σ
Зависимая переменная — P(i10 > 0)					
Средняя цитируемость соавторов×10 ⁻⁴	0.730*** (0.060)	0.026	0.023	0.030	
Число соавторов	0.208*** (0.008)	0.060	0.058	0.062	
Стаж (логарифм)	2.804*** (0.040)	0.058	0.056	0.061	
Константа	−3.374*** (0.062)				
Число наблюдений	34 701				
Псевдо- <i>R</i> ²	0.342				
Зависимая переменная — i10 i10 > 0					
Средняя цитируемость соавторов × 10 ⁻⁴	0.365*** (0.019)	1.156	1.021	1.292	0.12
Число соавторов	0.050*** (0.001)	2.024	1.921	2.128	0.22
Стаж (логарифм)	1.861*** (0.030)	4.093	3.925	4.262	0.44
Константа	−1.909*** (0.051)				
ln(α)	−0.384*** (0.030)				
Число наблюдений	28 690				
Псевдо- <i>R</i> ²	0.0738				

Примечание. В скобках указаны стандартные погрешности оценок; *** — значимость на уровне 1%.

Таким образом, для всех рассмотренных показателей цитируемости — числа цитирования, индекса Хирша, индекса $i10$ — найдены значимые корреляционные связи с числом соавторов и средней цитируемостью соавторов.

5. Заключение

Работа основана на анализе данных 34 тысяч профилей ученых Google Scholar (с полным охватом сети в 110 тысяч профилей) с учетом даты первого цитирования. Отсечение по году первого цитирования обусловлено включением в число объясняющих переменных стажа цитируемости ученого, что из-за особенностей Google Scholar возможно не для всех профилей. Поэтому полученные результаты относятся к авторам, имеющим относительно небольшой публикационный опыт.

Было проверено наличие статистической связи между положением ученого в сети научного соавторства, показателями цитируемости соавторов и результативностью его науч-

ной деятельности, выраженной в показателях цитируемости. Научная новизна данного исследования заключается в использовании для построения и анализа сетевых характеристик данных непосредственно об ученых и их соавторах, без привлечения анализа разрозненных записей об отдельных публикациях или перекрестного анализа других библиографических баз данных, что позволяет избежать проблемы поиска соответствия между авторами с одинаковыми фамилиями и их работами. Данные охватывают ученых из различных научных областей: компьютерные науки, экономика и финансы, биология и медицина, физика и химия, математика, социально-гуманитарные науки.

Использование большого числа ученых из различных научных областей позволяет избежать смещений при оценке показателей цитируемости. Показано, что существует положительная связь между цитируемостью ученых и числом соавторов, между цитируемостью и центральностью по близости. Также была получена положительная связь между цитируемостью ученого и средней цитируемостью их соавторов. Для оценивания степени статистической связи между переменными в работе использовались нормированные коэффициенты, характеризующие изменение зависимой переменной (в абсолютных цифрах и стандартных отклонениях) при увеличении объясняющих факторов на одно стандартное отклонение. Полученные количественные характеристики варьируются в зависимости от научной специализации автора, но сопоставимы по порядку величин. Помимо общего числа цитируемых работ, в качестве характеристики ученого также использовались два других показателя — индекс Хирша и индекс *i10*. Данные показатели значимо коррелируют с числом соавторов и их средней цитируемостью, при этом нормированный эффект числа соавторов примерно в два раза превосходит эффект цитируемости соавторов.

Таким образом, результаты анализа данных профилей авторов научных публикаций в Google Scholar показывают, что ученые, которые нарабатывают больше контактов и активнее других взаимодействуют с другими учеными, имеют, в среднем, лучшие библиометрические показатели цитируемости. Данные результаты согласуются с другими исследованиями, например, (Glanzel, 2002; Persson et al., 2004; Uddin et al., 2011), в которых отмечается тенденция роста цитируемости публикации с увеличением числа написавших ее соавторов.

Важно отметить, что выявленные статистические связи между переменными не позволяют однозначно интерпретировать их как действие того или иного причинно-следственного механизма. Из-за особенностей представления информации в профилях Google Scholar невозможно было использовать динамические данные об анализируемых переменных. Более фундаментальная проблема кроется в эндогенности социальных связей вообще и связей соавторства в частности, обусловленной тем, что связи формируются на добровольной основе.

Полученные результаты могут быть объяснены несколькими гипотезами, нуждающимися в дополнительной проверке. С одной стороны, корреляцию цитируемости ученого с числом соавторов и средней цитируемостью соавторов можно объяснить влиянием со стороны количества (чем больше соавторов, тем больше цитирований может получить работа) и качества соавторов на цитируемость совместной публикации (и косвенно, на последующие труды ученого). С другой стороны, при формировании коллектива соавторов существует самоотбор: сложные научные проблемы требуют для своего решения соответствующего количества и качества исследователей, поэтому попадание ученого в компанию высокочитируемых соавторов объясняется востребованностью его собственных творческих способностей. Оба механизма варьируются в зависимости от исследовательского опыта соавторов и разли-

чий в этом опыте, и их сосуществование непротиворечиво. Более того, оба механизма допускают их сознательное использование для увеличения научной результативности ученого.

Для относительно молодых ученых, как в рассмотренной выборке, важным мотивом написания совместных публикаций является развитие собственных исследовательских навыков и получение новой информации через взаимодействие с коллегами. Поэтому для начинающих авторов важным является взаимодействие с более опытными коллегами, как происходит, например, в парах «аспирант—руководитель». Делая совместную работу в большем коллективе, молодые ученые способны достичь лучших результатов. В этих случаях в объяснении связи цитируемости, числа соавторов и средней цитируемости соавторов значителен механизм влияния. Для более зрелых ученых мотив получения новой информации от соавторов менее значим, чем возможность принять участие в разработке сложной проблемы. Поэтому состоявшие ученые могут активно завязывать новые контакты, чтобы дополнить собственные компетенции в решении сложной научной задачи. В этом случае механизм самоотбора преобладает над механизмом влияния.

Благодарности. Работа выполнена в рамках проекта «Институциональные факторы поведения агентов на различных рынках услуг» Программы фундаментальных исследований НИУ ВШЭ в 2016 г. и при поддержке Правительства РФ в рамках реализации «Дорожной карты» Программы 5/100 НИУ ВШЭ.

Авторы выражают признательность старшему научному сотруднику ИПФ РАН Льву Матвееву за помощь в создании программы-парсера и анонимным рецензентам за замечания, которые позволили улучшить статью.

Список литературы

Бедный Б., Сорокин Ю. М. (2012). О показателях научного цитирования и их применении. *Высшее образование в России*, 3, 17–28.

Писляков В. (2010). Соавторство российских ученых с зарубежными коллегами: публикации и их цитируемость. *Препринты Высшей школы экономики*. WP6/2010/01, Москва.

Abbasi A., Altmann J., Hossain L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, 5 (4), 594–607.

Ajiferuke I., Famoye F. (2015). Modelling count response variables in informetric studies: Comparison among count, linear, and lognormal regression models. *Journal of Informetrics*, 9 (3), 499–513.

Avkiran N. K. (2013). An empirical investigation of the influence of collaboration in finance on article impact. *Scientometrics*, 95 (3), 911–925.

Bakkalbasi N., Bauer K., Glover J., Wang L. (2006). Three options for citation tracking: Google scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3:7.

Bensman S. J. (2013). Eugene Garfield, Francis Narin, and Pagerank: The theoretical bases of the Google search engine. *ArXiv preprint*, arXiv:1312.3872.

Bergstrom C. T., West J. D., Wiseman M. A. (2008). The eigenfactor metrics. *The Journal of Neuroscience*, 28 (45), 11433–11434.

Bornmann L., Daniel H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64 (1), 45–80.

- Cameron A. C., Trivedi P. K. (2013). *Regression analysis of count data*. Cambridge University Press.
- Delgado E., Repiso R. (2013). The impact of scientific journals of communication: Comparing Google Scholar metrics. *Comunicar*, 21 (41), 45–52.
- Ductor L., Fafchamps M., Goyal S., van der Leij M. J. (2014). Social networks and research output. *Review of Economics and Statistics*, 96 (5), 936–948.
- Egghe L. (2006). Theory and practice of the g-index. *Scientometrics*, 69 (1), 131–152.
- Fooladi M., Salehi H., Yunus M. M., Farhadi M., Chadehagani, A., Farhadi H., Ebrahim N. (2013). Does criticisms overcome the praises of journal impact factor? *Asian Social Science*, 9 (5), 176–182.
- Franceschet M. (2009). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, 83 (1), 243–258.
- Garfield E. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*, New York: Wiley.
- Girvan M., Newman M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99 (12), 7821–7826.
- Glanzel W. (2002). Coauthorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library trends*, 50 (3), 461–475.
- Glänzel W., Schubert A. (2004). Analysing scientific networks through co-authorship. In: H. F. Moed, W. Glänzel and U. Schmoch (eds.). *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems*. Springer, 257–276.
- Harzing A.-W. K., van der Wal R. (2008). Google Scholar as a new source for citation analysis. *Ethics in science and environmental politics*, 8 (1), 61–73.
- Heffner A. (1981). Funded research, multiple authorship, and subauthorship collaboration in four disciplines. *Scientometrics*, 3 (1), 5–12.
- Hirsch J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (46), 16569–16572.
- Jacsó P. (2005). As we may search — comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89 (9), 1537–1547.
- Jacsó P. (2008a). The pros and cons of computing the h-index using Google Scholar. *Online information review*, 32 (3), 437–452.
- Jacsó P. (2008b). Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for FW Lancaster. *Library trends*, 56 (4), 784–815.
- Katz J. S., Martin B. R. (1997). What is research collaboration? *Research policy*, 26 (1), 1–18.
- Kreiman G., Maunsell J. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, 5 (48), 1–6.
- Kretschmer H. (1994). Coauthorship networks of invisible colleges and institutionalized communities. *Scientometrics*, 30 (1), 363–369.
- Kulkarni A. V., Aziz B., Shams I., Busse J. W. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *Jama*, 302 (10), 1092–1096.
- Li E. Y., Liao C. H., Yen H. R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy*, 42 (9), 1515–1530.
- Meho L. I., Yang K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58 (13), 2105–2125.

Méndez-Vásquez R. I., Suñén-Pinyol E., Cervelló R., Camí J. (2012). Identification and bibliometric characterization of research groups in the cardio-cerebrovascular field, Spain 1996–2004. *Revista Española de Cardiología (English Edition)*, 65 (7), 642–650.

Newman M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (2), 404–409.

Newman M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical review E*, 69, 066133.

Orduña-Malea E., Ayllón J. M., Martín-Martín A., López-Cózar E. D. (2015). Methods for estimating the size of Google Scholar. *Scientometrics*, 104 (3), 931–949.

Ortega J. L. (2015). How is an academic social site populated? A demographic study of Google Scholar citations population. *Scientometrics*, 104 (1), 1–18.

Persson O., Glänzel W., Danell R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60 (3), 421–432.

Puuska H.-M., Muhonen R., Leino Y. (2014). International and domestic co-publishing and their citation impact in different disciplines. *Scientometrics*, 98 (2), 823–839.

Toutkoushian R. K., Webber K. (2011). Measuring the research performance of postsecondary institutions. In: C. J. Shin, K. R. Toutkoushian and U. Teichler (eds.). *University rankings: Theoretical basis, methodology and impacts on global higher education*. Springer, 123–144.

Uddin S., Hossain L., Rasmussen K. (2013). Network effects on scientific collaborations. *PLoS ONE*, 8 (2), 1–12.

Uddin S., Hossain, L., Abbasi A., Rasmussen, K. (2011). Trend and efficiency analysis of co-authorship network. *Scientometrics*, 90 (2), 687–699.

West J., Bergstrom T., Bergstrom C. T. (2010). Big Macs and eigenfactor scores: Don't let correlation coefficients fool you. *Journal of the American Society for Information Science and Technology*, 61 (9), 1800–1807.

Wildgaard L. (2015). A comparison of 17 author-level bibliometric indicators for researchers in astronomy, environmental science, philosophy and public health in Web of Science and Google Scholar. *Scientometrics*, 104 (3), 873–906.

Yang K., Meho L. I. (2006). Citation analysis: A comparison of Google Scholar, Scopus, and Web of Science. *Proceedings of the American Society for Information Science and Technology*, 43 (1), 1–15.

Yu Q., Shao H., Long C., Duan Z. (2014). The relationship between research performance and international research collaboration in the C&C field. *Experimental and Clinical Cardiology*, 20 (6), 145–153.

Zuckerman H. (1967). Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *American Sociological Review*, 32 (3), 391–403.

Поступила в редакцию 11.06.2016;
принята в печать 26.09.2016.

Matveeva N., Poldin O. Citation of scholars in co-authorship network: analysis of Google Scholar data. *Applied Econometrics*, 2016, v. 44, pp. 100–118.

Nataliya Matveeva

National Research University Higher School of Economics, Nizhnii Novgorod, Russian Federation;
Nata_Plast@mail.ru

Oleg Poldin

National Research University Higher School of Economics, Nizhnii Novgorod, Russian Federation;
opoldin@hse.ru

Citation of scholars in co-authorship network: Analysis of Google Scholar data

In this study, we analyze correlations between the co-authorship network parameters and citation characteristics in Google Scholar. We estimate the count data regression model in a sample of more than 30 thousand authors with the first citation after 2007. There is a positive relationship between scholar's citation counts and number of co-authors, between citations and the author's centrality, and between scholar's citations and the average citation of co-authors. The h-index and i10 index are significantly associated with the number of co-authors and average citation of co-authors.

Keywords: scientometrics, co-authorship network; bibliometric analysis; Google Scholar; count data models.

JEL classification: A14; D83; Z13.

References

- Bednyj B., Sorokin Ju. M. (2012). O pokazateljah nauchnogo citirovaniya i ih primenenii. *Higher Education in Russia [Vyshee obrazovanie v Rossii]*, 3, 17–28 (in Russian).
- Pislyakov V. (2010). International coauthorship of Russian scientists: Papers and their citedness: Working paper WP6/2010/01. Moscow: State University — Higher School of Economics (in Russian).
- Abbasi A., Altmann J., Hossain L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, 5 (4), 594–607.
- Ajiferuke I., Famoye F. (2015). Modelling count response variables in informetric studies: Comparison among count, linear, and lognormal regression models. *Journal of Informetrics*, 9 (3), 499–513.
- Avkiran N. K. (2013). An empirical investigation of the influence of collaboration in finance on article impact. *Scientometrics*, 95 (3), 911–925.
- Bakkalbasi N., Bauer K., Glover J., Wang L. (2006). Three options for citation tracking: Google scholar, Scopus and Web of Science. *Biomedical Digital Libraries*, 3:7.
- Bensman S. J. (2013). Eugene Garfield, Francis Narin, and Pagerank: The theoretical bases of the Google search engine. *ArXiv preprint*, arXiv: 1312.3872.
- Bergstrom C. T., West J. D., Wiseman M. A. (2008). The eigenfactor metrics. *The Journal of Neuroscience*, 28 (45), 11433–11434.
- Bornmann L., Daniel H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64 (1), 45–80.
- Cameron A. C., Trivedi P. K. (2013). *Regression analysis of count data*. Cambridge University Press.

- Delgado E., Repiso R. (2013). The impact of scientific journals of communication: Comparing Google Scholar metrics. *Comunicar*, 21 (41), 45–52.
- Ductor L., Fafchamps M., Goyal S., van der Leij M. J. (2014). Social networks and research output. *Review of Economics and Statistics*, 96 (5), 936–948.
- Egghe L. (2006). Theory and practice of the g-index. *Scientometrics*, 69 (1), 131–152.
- Fooladi M., Salehi H., Yunus M. M., Farhadi M., Chadegani, A., Farhadi H., Ebrahim N. (2013). Does criticisms overcome the praises of journal impact factor? *Asian Social Science*, 9 (5), 176–182.
- Franceschet M. (2009). A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, 83 (1), 243–258.
- Garfield E. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*, New York. Wiley.
- Girvan M., Newman M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99 (12), 7821–7826.
- Glanzel W. (2002). Coauthorship patterns and trends in the sciences (1980–1998): A bibliometric study with implications for database indexing and search strategies. *Library trends*, 50 (3), 461–475.
- Glänzel W., Schubert A. (2004). Analysing scientific networks through co-authorship. In: H. F. Moed, W. Glänzel and U. Schmoch (eds.). *Handbook of quantitative science and technology research: The use of publication and patent statistics in studies of S&T systems*. Springer, 257–276.
- Harzing A.-W. K., van der Wal R. (2008). Google Scholar as a new source for citation analysis. *Ethics in science and environmental politics*, 8 (1), 61–73.
- Heffner A. (1981). Funded research, multiple authorship, and subauthorship collaboration in four disciplines. *Scientometrics*, 3 (1), 5–12.
- Hirsch J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102 (46), 16569–16572.
- Jacsó P. (2005). As we may search — comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases. *Current Science*, 89 (9), 1537–1547.
- Jacsó P. (2008a). The pros and cons of computing the h-index using Google Scholar. *Online information review*, 32 (3), 437–452.
- Jacsó P. (2008b). Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for FW Lancaster. *Library trends*, 56 (4), 784–815.
- Katz J. S., Martin B. R. (1997). What is research collaboration? *Research policy*, 26 (1), 1–18.
- Kreiman G., Maunsell J. (2011). Nine criteria for a measure of scientific output. *Frontiers in Computational Neuroscience*, 5 (48), 1–6.
- Kretschmer H. (1994). Coauthorship networks of invisible colleges and institutionalized communities. *Scientometrics*, 30 (1), 363–369.
- Kulkarni A. V., Aziz B., Shams I., Busse J. W. (2009). Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles published in general medical journals. *Jama*, 302 (10), 1092–1096.
- Li E. Y., Liao C. H., Yen H. R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy*, 42 (9), 1515–1530.
- Meho L. I., Yang K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, 58 (13), 2105–2125.

Méndez-Vásquez R. I., Suñén-Pinyol E., Cervelló R., Camí J. (2012). Identification and bibliometric characterization of research groups in the cardio-cerebrovascular field, Spain 1996–2004. *Revista Española de Cardiología (English Edition)*, 65 (7), 642–650.

Newman M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98 (2), 404–409.

Newman M. E. (2004). Fast algorithm for detecting community structure in networks. *Physical review E*, 69, 066133.

Orduña-Malea E., Ayllón J. M., Martín-Martín A., López-Cózar E. D. (2015). Methods for estimating the size of Google Scholar. *Scientometrics*, 104 (3), 931–949.

Ortega J. L. (2015). How is an academic social site populated? A demographic study of Google Scholar citations population. *Scientometrics*, 104 (1), 1–18.

Persson O., Glänzel W., Danell R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60 (3), 421–432.

Puuska H.-M., Muhonen R., Leino Y. (2014). International and domestic co-publishing and their citation impact in different disciplines. *Scientometrics*, 98 (2), 823–839.

Toutkoushian R. K., Webber K. (2011). Measuring the research performance of postsecondary institutions. In: C. J. Shin, K. R. Toutkoushian and U. Teichler (eds.). *University rankings: Theoretical basis, methodology and impacts on global higher education*. Springer, 123–144.

Uddin S., Hossain L., Rasmussen K. (2013). Network effects on scientific collaborations. *PLoS ONE*, 8 (2), 1–12.

Uddin S., Hossain, L., Abbasi A., Rasmussen, K. (2011). Trend and efficiency analysis of co-authorship network. *Scientometrics*, 90 (2), 687–699.

West J., Bergstrom T., Bergstrom C. T. (2010). Big Macs and eigenfactor scores: Don't let correlation coefficients fool you. *Journal of the American Society for Information Science and Technology*, 61 (9), 1800–1807.

Wildgaard L. (2015). A comparison of 17 author-level bibliometric indicators for researchers in astronomy, environmental science, philosophy and public health in Web of Science and Google Scholar. *Scientometrics*, 104 (3), 873–906.

Yang K., Meho L. I. (2006). Citation analysis: A comparison of Google Scholar, Scopus, and Web of Science. *Proceedings of the American Society for Information Science and Technology*, 43 (1), 1–15.

Yu Q., Shao H., Long C., Duan Z. (2014). The relationship between research performance and international research collaboration in the C&C field. *Experimental and Clinical Cardiology*, 20 (6), 145–153.

Zuckerman H. (1967). Nobel laureates in science: Patterns of productivity, collaboration, and authorship. *American Sociological Review*, 32 (3), 391–403.

Received 11.06.2016; accepted 26.09.2016.