



Александр БИРЮКОВ¹

Факультет информационных и компьютерных наук, Университет Тренто, Италия;
Европейский инновационный альянс, Гент, Бельгия

Джозеф Р. ВЕЙКЛИНГ

Европейский инновационный альянс, Гент, Бельгия

Клаудио БАРТОЛИНИ

Лаборатория сервисной автоматизации и интеграции, HP, Palo Alto, Калифорния, США

Фабио КАЗАТИ

Факультет информационных и компьютерных наук, Университет Тренто, Италия

Маурицио МАРКЕЗЕ

Факультет информационных и компьютерных наук, Университет Тренто, Италия

Кацярына МИРЫЛЕНКА

Факультет информационных и компьютерных наук, Университет Тренто, Италия

Нардин ОСМАН

Научно-исследовательский институт искусственного интеллекта (IIIA-CSIC), Барселона, Испания

Аддзурра РАГОНЭ

Факультет информационных и компьютерных наук, Университет Тренто, Италия; Exprivia SpA, Мольфетта, Италия

Карлос СИЕРРА

Научно-исследовательский институт искусственного интеллекта (IIIA-CSIC), Барселона, Испания

Аалам ВАССЕФ

Peerevaluation.org, Париж, Франция

АЛЬТЕРНАТИВНОЕ РЕЦЕНЗИРОВАНИЕ: НОВЫЕ ПОДХОДЫ к ЭКСПЕРТНОЙ ОЦЕНКЕ

В этой статье мы рассмотрим несколько новых подходов к оценке результатов исследований. Сначала вкратце опишем суть экспертной оценки и показатели, которые используются для оценки эффективности, затем – возникающие противоречия и качество такого типа оценки в целом. Мы остановимся на 5 подходах, которые включают и экспертизу, основанную на репутации (ее продвигает проект LiquidPub и команды ученых, которые сотрудничают с ним).

Эти подходы являются либо альтернативой рецензированию, либо дополняют традиционную экспертную оценку. Мы обсудим все плюсы и минусы предлагаемых подходов и поговорим о будущем экспертной оценки. Мы считаем, что ни одна из существующих систем в научном сообществе не может удовлетворить запросы всех заинтересованных сторон.

¹ Александр Бирюков сейчас работает в редакции компьютерных наук, Springer-Verlag GmbH, Гейдельберг, Германия

1. ВВЕДЕНИЕ

Внешняя экспертная оценка или ее аналоги были частью научно-публикационного процесса как минимум с XVIII века (Kronick, 1990). Практика экспертной оценки, как и ее



научное сообщество само вывело разнообразные методы для оценки научного знания

установленные нормы, изменялась в зависимости от дисциплины и исторического периода (Burnham, 1990; Spier, 2002), но ключевые моменты всегда оставались неизменными: забота о корректности научной работы; борьба с искажением научного знания путем распространением ложных утверждений; необходимое для авторов признание истинности их работы; поддержка репутации организации, издателя или редколлегии, ответственных за работу; и, в то же время, работа над тем, чтобы на пути новых ценных идей не возникало препятствий различного характера. В XX веке в связи с увеличением объема публикаций процесс рецензирования стал необходимостью. Из огромного числа материалов предстояло выбрать те, которые будут представлены на страницах самых авторитетных журналов (Ingelfinger, 1974), объем которых был ограничен. Одним из последствий соревнования за читательское внимание стало то, что рецензентов все чаще просили оценить не только техническую сторону статьи, но и написать по сути редакционную рецензию, которая включает перечисление возможности использования данных в статье, ее потенциальное влияние и ценность для дальнейших исследований (Lawrence, 2003).

Научное сообщество само вывело разнообразные методы для оценки научного знания, которые не ограничивались только односторонним слепым рецензированием (рецензент знает имя автора, но автор не знает о рецензенте), двусторонним слепым рецензированием (автор и рецензент не знают имен друг друга) и открытым рецензированием (и автор, и рецензент знают друг о друге). Редакторы журналов также играют роль как в первичной оценке (стоит ли вообще отправлять статью на рецензирование), так и в принятии решения о публикации на основании рекомендаций рецензента. Степень различия предпочтений редактора и рецензента в разных изданиях может очень сильно различаться (McCook, 2006). На сегодняшний день, несмотря на повсеместное использование рецензирования и всеобщее одобрение его основных преимуществ для исследовательской работы (Ware and Monkman, 2008; Sence About Science, 2009), есть много опасений по поводу такого способа оценки: противоречивость и предвзя-

тость, неэффективный поиск ошибок и фальсификаций, создание барьера для инновационных исследований и проектов.

В этой статье мы обсудим разные модели оценки, которые являются либо полной альтернативой традиционному рецензированию, либо дополняют его. В следующем разделе мы представим краткий обзор традиционного процесса экспертной оценки и его противоречий, включая обзор существующей литературы и анализ рецензирования. Также мы проанализируем поведение рецензента в разных дисциплинах при различных видах рецензирования. Затем мы дадим обзор ряда количественных показателей для оценки качества и эффективности экспертной оценки в целом. Эти показатели используются для проверки надежности и устойчивости процесса рецензирования, степени согласия и предвзятости рецензентов, а также для проверки способности рецензентов предсказать будущее влияние определенной научной работы.

Затем мы перейдем к рассмотрению экспериментальных методик в рецензировании и сравним быстрое ранжирование статей, тендерный метод для рассмотрения статей и традиционный способ оценки. Мы также обсудим два подхода к оценке научного труда, которые основываются на использовании явной или неявной обратной связи с научным сообществом: *OpinioNet* и *UCount*. В выводах мы поговорим о плюсах и минусах представленных подходов и нашем видении процесса рецензирования в будущем.

2. РЕЦЕНЗИРОВАНИЕ: ИСТОРИЯ И ПРОТИВОРЕЧИЯ

Процесс рецензирования в том или ином виде был частью научно-публикационного процесса, начиная с первых научных журналов, особенно *Philosophical Transactions of the Royal Society*. Первый задокументированный процесс экспертной оценки отмечают в журнале *Medical Essays and Observations*, который издавался Королевским обществом Эдинбурга (Kronick, 1990).

Хотя типы рецензирования сильно различались (Burnham, 1990), существенный рост научной литературы в двадцатом веке привел к тому, что рецензирование стало практически повсеместным и начало рассматриваться как ключевой механизм оценки научной работы (Ingelfinger, 1974; Ware and Monkman, 2008; Sence About Science, 2009).

Несмотря на повсеместную практику (хотя более правильным было бы сказать о

большом разнообразии практик, которые имеют одно название), экспертная оценка была слабо изучена до последних десятилетий. Результаты этих исследований, возможно, сильно удивят, поскольку они зачастую допускают двойное толкование и не отвечают на вопрос, действительно ли рецензирование играет роль «фильтра» для выбора качественной работы и исправления ошибок (Jefferson et al., 2007). В значительном числе работ указывалось, что экспертная оценка представляет собой процесс, эффективность которого «вопрос веры, а не доказательства» (Smith, 2006), то есть процесс «непроверенный» и «неопределенный» (Jefferson et al., 2002b), о реальных последствиях которого мы знаем очень мало, потому что ученые редко получают доступ к данным, имеющим отношение к этому вопросу.

Например, Lock (1994) утверждает, что экспертная оценка может только помочь обнаружить серьезные ошибки, не более того, а реальным критерием оценки документа может стать то, как часто его содержание используется и упоминается спустя несколько лет после публикации. Другие исследования ставят под сомнение способность рецензирования обнаружить важные ошибки в работе (Godlee et al., 1998). Однако до сих пор рецензирование считается процессом, у которого нет достойной альтернативы (Kassier and Campion, 1994; Smith, 2006).

Один из аспектов проблемы заключается в том, что цель рецензирования и его процесс может отличаться в зависимости от дисциплины и журнала. Исследования экспертной оценки различаются по типу и количеству информации и использования различных систем показателей для оценки ее эффективности. Действительно, наличие четкой цели для анализа – это один из ключевых и самых трудных моментов, так как зачастую неясно, как измерить эффективность экспертной оценки (Jefferson et al., 2002a). Однако в целом все показатели можно разделить на две группы: те, с помощью которых можно определить эффективность и правильность экспертной оценки (см. ниже), и те, которые направлены на понимание того, что автор подразумевает под «положительными» характеристиками рецензирования (этот вопрос мы обсуждаем в разделе «Количественный анализ рецензирования»).

Первую категорию исследования можно разделить на две большие группы: одна группа показывает способности рецензирования обнаруживать ошибки, с помощью другой

можно измерить возможность экспертной оценки предвидеть будущее влияние работы, которое обычно измеряется количеством цитирований.

Исследование по обнаружению ошибок было проведено в 1994 г. (Goodman et al.). Авторы изучили 111 работ, которые были присланы в *Annals of Internal Medicine* в период с марта 1992 года по март 1993. Они изучали документы до и после рецензирования, чтобы выяснить, смог ли рецензент обнаружить ошибки. Существенного различия в статьях до и после публикации обнаружено не было. В результате экспертной оценки были обнаружены только небольшие недостатки – неточности в рисунках, статистике и описании результатов. Интересное исследование проводил Godlee и соавторы (1998). Они намеренно внесли ошибки в статью, которая уже была принята в *British Medical Journal* (BMJ) [1] и разделили 420 рецензентов на 5 групп. 1 и 2 группы не знали о личности авторов, а группы 3 и 4 – знали. Первой и третьей группе было предложено подписать свои рецензии, а вторую и четвертую группу попросили прислать рецензию без подписи. Единственное различие между группами 4 и 5 было в том, что рецензенты из 5-й группы знали, что они принимали участие в исследовании. Godlee и др. (1998) сообщают, что в среднем существенных ошибок было обнаружено 2 из 8. 16% не нашли ошибок совсем, 33% рецензентов приняли статью, несмотря на ошибки. К сожалению, в исследовании нет данных о том, вместе ли рецензенты обнаруживали ошибки (эта информация очень помогла бы для исследований процессов в научном сообществе, которые будут рассматриваться далее) и были ли некоторые ошибки замечены чаще, чем другие.

Подсчет цитирований широко используется в качестве критерия оценки в исследованиях Bornmann и Daniel. Первое исследование проверяет эффективность рецензирования в выборе рецензентов с более высокими показателями цитирования и показывает, что такая взаимосвязь действительно существует (Bornmann and Daniel, 2005b). Более поздняя

цель рецензирования и его процесс могут отличаться в зависимости от дисциплины и журнала

Ключевые слова:
оценка научных исследований, рецензирование, показатели, тендерный метод, LiquidPub, UCount

Keywords:
research evaluation, peer review, metrics, bidding, opinions, LiquidPub, UCount





даже категорическое несогласие рецензентов друг с другом может оказаться полезным

работа авторов рассматривает первоначальную оценку рукописей, которую проводили редакторы крупного химического журнала, по сравнению с более поздними оценками внешних рецензентов (Bornmann and Daniel, 2010a). Хотя оценки редакторов коррелировали с цитированиями статьи в будущем, в 2/3 случаев редакторы были либо не способны, либо не рискнули оценить статью. Окончательные оценки после второго рецензирования коррелировали с числом цитирований гораздо сильнее, что говорит о положительном эффекте вне зависимости от того, согласились редакторы с первоначальным решением или нет. Эти результаты можно сравнить с теми, которые получили Orthof и соавторы в 2002 году в отношении рукописей, представленных в медицинский журнал, в котором изначальные оценки редакторов не коррелировали с более поздним подсчетом цитирований, в то время как оценки внешних рецензентов коррелировали сильнее, поскольку учитывалось больше участников. Лучшим прогностическим значением оказалась комбинация рейтингов редакторов и рецензентов, указывающая на то, что различия в прогностической ценности ослабевали, поскольку рецензенты и редакторы улавливали разные аспекты качества статьи.

3. КОЛИЧЕСТВЕННЫЙ АНАЛИЗ ЭКСПЕРТНОЙ ОЦЕНКИ

В этом разделе мы рассмотрим исследования, которые касаются количественного анализа экспертной оценки. Эффективность или правильность рецензирования может быть измерена с учетом различных показателей, которые включают способность предсказывать будущую позицию статьи в рейтингах цитирования, разногласия между рецензентами и необъективность рецензента.

Самым очевидным количественным показателем для измерения корреляции между оценками рукописей и их последующим влиянием на науку можно назвать цитаты. Как мы говорили в предыдущем разделе, результаты могут очень сильно зависеть от конкретного контекста. Например, Bornmann и Daniel (2010b) изучали данные 1899 представленных в *Angewandte Chemie International* рукописей

и обнаружили положительную корреляцию между рекомендациями рецензентов и последующим влиянием – самое интересное, что более сильная корреляция была там, где в процессе принятия статьи к публикации принимало участие меньшее число рецензентов [2].

С другой стороны, Ragone et al. (2011, 2013) изучали большой банк данных из 9000 рецензий, которые охватывали около 3000 статей, поданных на 10 конференций в области информатики. Исследователи наблюдали несколько статистически значимых корреляций, когда ранжирование документов согласно мнению рецензентов сопоставлялось с ранжированием по цитированиям [3].

Другим важным показателем экспертной оценки является согласие между рецензентами (Casati et al., 2010), который измеряет, насколько различаются оценки статьи, полученные от разных рецензентов. Смысл этого показателя состоит в том, что в то время как точки зрения рецензентов могут различаться в зависимости от их квалификации, областей знания и т.д., мы можем ожидать, что в их заключениях будет достигнут некий консенсус по ключевым преимуществам статьи или их отсутствию. С другой стороны, если оценка, данная рецензентами, сопоставима с оценками, данными случайным образом, то результат процесса экспертной оценки будет также случайно эффективен, что ставит достижение изначальной цели под угрозу. Есть несколько причин для того, чтобы сотрудничать с несколькими рецензентами: для оценки на основе консенсуса или мнения большинства и для того, чтобы иметь возможность провести многократную экспертизу (например, один рецензент будет оценивать методологию исследования, а два других – содержание статьи).

Более того, даже категорическое несогласие рецензентов друг с другом может оказаться полезным. Это значит, что сами привлеченные эксперты не могут оценить значение работы для науки. Этот показатель может быть полезен для улучшения качества процесса рецензирования, может помочь решить, нужны ли еще рецензенты для того, чтобы процесс экспертной оценки был действительно качественным.

Значительная часть исследований, касающихся рецензирования, сосредоточены на том, чтобы выявить факт предубежденности рецензента и понять, как это влияет на результат оценки. Действительно, объективность рецензента – это основополагающее качество процесса экспертной оценки, «идеальный рецензент», отмечает Ingelfinger (1974), «должен быть полностью объективным, другими словами, сверхъ-

естественным». Разные подходы к анализу предубежденности экспертной оценки определяют несколько типов предубежденности: аффилиационная зависимость, это означает, что исследователи из выдающихся учреждений находятся в более выгодном положении в глазах рецензента (Ceci and Peters, 1982); предубеждение, связанное с предпочтением исследователей из США (Link, 1998), дискриминация по половому признаку в отношении женщин-исследователей (Wenneras and Wold, 1997; Bornmann, 2007; Marsh et al., 2009; Ceci and Williams, 2011), «порядковое» предубеждение (Bornmann and Daniel, 2005a), это значит, что при рассмотрении в алфавитном порядке кандидатов на получение стипендий аспирантов и пост-доков (postdocs), в более выгодном положении оказываются те, чьи фамилии находятся в начале списка. Хотя не всегда легко отличить очевидное предубеждение от других факторов, как, например, разница в квалификации. Как минимум, существуют предубеждения, основанные на национальности рецензентов и авторов – и они остаются даже в том случае, если учитывается качественный аспект работы (Lee et al., 2006; Lynch et al., 2007). Другие типы предубеждения, основанные на предпочтении статистически значимых результатов (Olson et al., 2002; Lee et al., 2006) или дискриминации по половому признаку (Marsh et al., 2009; Ceci and Williams, 2011), становятся более редкими по сравнению с другими факторами, отличными от самого процесса рецензирования. Кроме того, возможно вычислить рейтинг необъективности, когда рецензент последовательно дает более высокие или более низкие оценки вне зависимости от вклада, который статьи вносят в науку. Это тоже является своего рода предубеждением, которое проявляется довольно часто. Но это легко обнаружить и можно скорректировать с помощью довольно простых процедур для улучшения точности экспертной оценки (Ragone et al., 2011).

Одним из способов определения предвзятости является сравнение результатов одностороннего и двустороннего слепого рецензирования. Одностороннее слепое рецензирования предусматривает анонимность рецензентов и используется для защиты рецензентов от ответных «нападков» авторов. Во многих областях исследований одностороннее слепое рецензирования является обычной практикой. Тем не менее, в других областях, таких как информационные системы или, по крайней мере, на конференции Association for Computing Machinery Special

Interest Group on Management of Data (ACM SIGMOD), где рецензент и автор не знают имен друг друга, это норма. Целью двустороннего слепого рецензирования является помощь рецензентам в оценке только научных аспектов статьи без учета других факторов и, как следствие, обеспечение беспристрастности.

Анализ преимуществ двустороннего слепого рецензирования привел к нескольким двусмысленным результатам. Ранние исследования, проведенные McNutt et al. (1990) and Fisher et al. (1994) на базе журналов с двусторонним слепым рецензированием показали, что такой вид рецензирования положительно влияет на качество рецензии по подсчетам редакторов, хотя на результаты последнего исследования, возможно, оказал влияние тот факт, что «слепые» рецензенты знали о том, что принимают участие в исследовании, в то время как «зрячие» эксперты не знали об этом. Позже появилось исследование Justice et al. (1998), в котором все рецензенты были в курсе того, что принимают участие в исследовании. Автор не выявил статистически значимых различий. Другое исследование, проведенное van Roooyen и соавторами (1999), включало и знающих и не знающих об исследовании рецензентов. Оно не выявило никаких статистически значимых различий ни от стиля рецензирования (одно- или двустороннее слепое), ни от знания, является это рецензирование исследованием или нет. Однако большое исследование материалов медицинских конференций, проведенных Ross et al. (2006), показало, что двустороннее слепое рецензирование успешно устранило множество предубеждений, связанных с полом, национальностью, престижем организации и другими факторами.

Одним из основных факторов, который может объяснить эти противоречивые результаты, является то, что скрытые данные автора не всегда могут скрыть информацию о нем, поскольку авторы часто ссылаются на свои предыдущие работы в статьях (Cho et al., 1998; Katz et al., 2002).

Вероятность такого случайного раскрытия информации повышается для объемных работ, таких как журнальная статья. Это усложняет двустороннее слепое рецензирование и сказывается на его результате, в отличие от неболь-

исследований по открытому рецензированию сейчас довольно мало



ших по объему работ: тезисов или кратких сообщений. Доля раскрытия данных широко колеблется в зависимости от журнала. Может быть и так, что объем материалов и размер самого пишущего научного сообщества также влияют на легкость определения личности автора (Ross et al., 2006). Также автор имеет возможность опознать рецензента по его комментариям. Поэтому позитивный результат от использования такого вида рецензирования может быть довольно трудно получить на практике.

Исследований по открытому рецензированию (когда имя рецензента известно автору) сейчас довольно мало. Первоначальные исследования показали, что открытые рецензии были сделаны более тщательно, они были более вежливыми, а рецензент обычно уделял им больше времени, чем обычно (Walsh et al., 2000). Примером открытого рецензирования, принятого на конференциях серии *PLoP [4], было «пастырство», где пастух (рецензент) работает вместе с овцами (авторами) для того, чтобы улучшить статью. Основной проблемой открытого рецензирования является борьба с нежеланием некоторых рецензентов раскрытия информации о них (Ware and Monkman, 2008), хотя журналы, которые используют такой вид рецензирования, сообщили о хороших результатах на практике (Godlee, 2002).

Исследования показывают, что для улучшения результатов рецензирования иногда достаточно обращать внимание на детали. Например, оценочная шкала может влиять на рецензентов и заставлять их использовать только определенные оценки вместо всего предложенного диапазона (Casati et al., 2010; Medo and Wakeling, 2010). Было показано, что в шкале с диапазоном от 1 до 5 с дистанцией в полшага рецензенты используют, как правило, только целые значения, а если предложить им диапазон от 1 до 10 (без полушагов), рецензенты используют всю шкалу, чтобы оценить статью (Casati et al., 2010). В шкале с диапазоном от 1 до 7 рецензенты имеют тенденцию концентрироваться в середине шкалы (Casati et al., 2010).

Одной из основных проблем анализа экспертной оценки является доступ к данным. Как правило, работы по изучению рецензирования ограничены анализом всего одной-двух конференций, заявок на гранты или стипендии. Например, Reinhart (2009) проанализировал 496 заявок на проект по базовому финансированию; Vornmann and Daniel (2005a) изучали процесс отбора 1954 докторских и 743 кандидатских заявок на получение сти-

пендии; Vornmann и соавторы (2008) анализировали 668 заявок на финансирование; Godlee и соавторы (1998) отбирали для своего эксперимента 420 рецензентов из базы данных журнала; Goodman и соавторы (1994) проанализировали 111 рукописей, принятых к публикации. Как мы уже упоминали выше, один из крупнейших банков данных был использован в работе Ragone и соавторов (2011), в которой они собрали данные 10 конференций – в общей сложности 9032 рецензий, предоставленных 2295 рецензентами на 2797 статей, отправленных на рецензию.

4. ЭКСПЕРИМЕНТЫ В РЕЦЕНЗИРОВАНИИ

В настоящее время ученые и редакторы изучают альтернативные подходы к решению некоторых распространенных проблем в традиционной экспертной оценке (Akst, 2010). Альтернативы включают: предоставление авторам возможности пересылать отзывы из одного журнала в другой (Akst, 2010), размещение комментария рецензента по аналогии с публикацией статьи [5], или запуск процесса традиционной экспертной оценки параллельно с публичным обсуждением (Akst, 2010). На конференции ACM SIGMOD также экспериментировали с вариациями классической модели рецензирования, в которой статьи оцениваются в два этапа. При такой схеме на первом этапе отфильтровываются статьи, которые вряд ли будут приняты – это позволяет сосредоточить усилия рецензентов на рассмотрении ограниченного числа работ. В исследовании Casati et al. (2010) авторы предложили модель многофазного обзора, который может улучшить процесс экспертной оценки с помощью уменьшения усилий для написания самой рецензии, необходимой для принятия решения по множеству представленных документов, сохраняя при этом результат такого же качества.

Далее мы сосредоточимся на трех экспериментальных подходах к экспертной оценке: просьбе к рецензентам ранжировать статьи вместо того, чтобы рассматривать их обычным способом; тендер на рецензирование статьи и открытое рецензирование.

4.1. Эксперимент. Сравнение ранжирования статей со стандартным способом рецензирования

Для воркшопа «Менеджмент в сфере информационных технологий, основанный на

потребностях бизнеса», проводимого при содействии Института инженеров электротехники и электроники (Business-Driven IT Management Workshop, IEEE BDIM) в 2010 году, члены Технического программного комитета (ТПК) экспериментировали с коллективным разумом при отборе статей. Цель эксперимента состояла в оценке жизнеспособности альтернативного механизма отбора, в котором рецензенты (некоторые из них) могли ранжировать документы на основе быстрого чтения, а не погружаться глубоко в текст для оценки качества.

Вот порядок, которому они следовали:

Члены ТПК попросили рецензентов разделить на примерно равные по количеству группы, одна из которых предпочитала коллективный разум в оценке статьи (а), а другая тяготела к традиционному подходу (б). Затем члены ТПК разделили на группы тех рецензентов, которые не ответили либо не высказали конкретного мнения [6].

Члены ТПК точно знали, в какой группе они находились, но не знали о том, где находятся другие рецензенты.

Группа (б) делала обычно 3-4 традиционных рецензии.

В конце процесса обзора отзывы группы (б) были приведены к среднему значению, результатом которого, как обычно, явился общий список всех представленных на рассмотрение статей.

Группе (а) был передан PDF-файл, содержащий все материалы (исключая те, в которых был конфликт интересов) без информации об авторе, таким образом, был запущен процесс двустороннего слепого рецензирования.

Группа (а) должна была представить общий список (или список большинства) документов, которые были переданы на рассмотрение, потратив не более 3-5 минут на каждую статью.

Члены ТПК объединили списки, равно оценивая значение двух способов рецензирования. Лучшие статьи были разделены на типы (расширенная презентация, обычная презентация, короткая презентация, стендовый доклад, отклоненная) в соответствии с согласованным упорядоченным списком. Члены ТПК устраивали дополнительные обсуждения там, где это было необходимо.

Авторы получили отзыв, который содержал:

- информацию о принятии или отклонении статьи;
- тип принятой статьи, если такое деление было применимо (расширенная, обычная, короткая статья (презентация) или стендовый доклад);

возможность улучшения статьи до публикации многие исследователи считают самой важной в рецензировании



- полную информацию о процессе рецензирования;
- по крайней мере три представленные рецензии;
- оценку статьи по результатам традиционного рецензирования и рейтинг по результатам рецензирования посредством «коллективного разума»;

Были получены следующие результаты:

- рецензенты разделились ровно на две равные группы – ровно половина предпочла традиционный подход, другая половина – метод «коллективного разума»;

• выборка проводилась следующим образом: каждую статью рецензировали три эксперта, поэтому члены ТПК считали вес количественного показателя результатов «коллективного разума» равным трем экспертам. Они трансформировали оценку путем усреднения оценок всех экспертов и представили в линейном виде средний рейтинг с диапазоном количественных показателей десятков традиционных рецензий.

- результаты были таковы, что в тройку лучших статей и в четверку худших попали одни и те же статьи по результатам двух типов рецензирования. Тем не менее, в случае с работами, оказавшимися в середине рейтинга, членам ТПК пришлось учитывать не только оценки рецензентов, но и содержание самой рецензии, придавая больший вес оценкам более опытных экспертов. Для работ, попавших в среднюю категорию, «коллективный разум» не помог, и, в основном, его результаты отклонялись от конечных результатов отбора.

Эксперимент показал, что «коллективный разум» можно применять для быстрого ранжирования статей и определения работ в лучшей и худшей категории. Однако для тех статей, которые находятся вне этих категорий, такой подход не годится.

4.2. e-Scripts. Тендерный метод рецензирования

Большинство исследователей отдают предпочтение рецензированию как ключевому механизму для оценки исследований (Ware

and Monkman, 2008; Sense About Science, 2009). Основным мотивирующим фактором здесь является возможность экспертной оценки не только для непосредственной оценки работы или фильтра, но и для возможности улучшения статьи до публикации (Goodman et al., 1994; Purcell et al., 1998; Sense About Science, 2009), и многие исследователи считают эту возможность самой важной в рецензировании (Sense About Science, 2009).

С другой стороны, некоторые из основных разочарований авторов (и редакторов), касающиеся рецензирования, связаны как раз с теми случаями, когда рецензент либо немотивирован, либо не знаком с предметом. На конференциях (например, EuroPLoP) с этой проблемой чаще всего борются, позволяя членам программного комитета выбирать статьи на рецензию посредством *тендера* на основании их названий и аннотаций. Таким образом, каждый член программного комитета мог надеяться на то, что получит статью для рецензирования, которая будет отвечать его профессиональным интересам и области знаний. Роль председателя программного комитета тоже существенно упросталась из-за меньшего количеством работы, связанного с назначением рецензентов для статьи.

Система управления электронными материалами *e-Scripts* [7] разработана Институтом компьютерных наук, социальной информатики и телекоммуникаций (ICST) и Европейским инновационным альянсом (EAI), которые попытались перенести принципы и преимущества в процесс рецензирования научных журналов. Названия и аннотации представленных статей размещаются в открытом доступе, и в течение примерно двух недель заинтересованные читатели могут голосовать за те статьи, которые пришли им по душе. В конце публичного тендерного периода редактор утверждает ранжированный список кандидатов в рецензенты путем выбора из проголосовавших кандидатов и рецензентов, предложенных редактором или авторами статьи. Приглашение на рецензирование автоматически рассылаются людям, находящимся в верхней части списка.

Цель здесь в том, чтобы использовать энтузиазм и желание помочь, которые движут хорошими рецензентами, но не слепо полагаться на эти качества: в отличие от некоторых неудачных попыток коллективных рецензий (Greaves et al., 2006), в *e-Scripts* редакторы все еще несут ответственность за назначение и поиск надежных рецензентов. В этом случае рецензенты, которые участвуют в голосова-

нии за статьи, включаются в список экспертов, а не заменяют его. Кроме того, система поддерживает режим конфиденциальности для еще не опубликованной работы, редактор журнала контролирует доступ к тексту.

Такой подход не только улучшает качество отдельных рецензий, но и может генерировать дополнительные данные для того, чтобы помочь редакции принять решение. Во-первых, поскольку по статистике просмотров можно предположить потенциальный уровень цитирования (Brody et al., 2006), мы можем предположить, что корреляция между количеством голосов и потенциальной важностью статьи будет довольно высока. Во-вторых, такая система может быть использована для построения профиля интересов рецензента, с помощью которого можно будет помочь автоматизировать процесс привлечения рецензента. Эта система является основой проекта UCount, который реализует EAI, наряду с другими средствами оценки и ранжирования потенциальных рецензентов. Мы поговорим о нем в разделе 5.2.

4.3. Peerevaluation.org: доверие ученых во времена социальной сети (social web)

Для поколения Y шэринг, анализ, распространение и получение немедленных отзывов стали не только обычными потребностями, на них возложены вполне определенные ожидания. Для почти миллиона пользователей Facebook такие потребности и ожидания включены в ежедневный поток потребления информации, коммуникации, развлечения, работы и доступа к знаниям.

4.3.1. Создание социальной репутации

В социальной сети все имеют право быть всеми сразу: авторами, рецензентами, распространителями и потребителями. С такой возможностью перетасовки ролей кажется вполне логичным, что появляются альтернативные механизмы *построения репутации*.

4.3.2. История Джона

Джон сочинил песню, загрузил ее на YouTube и поделился с друзьями. Песня стала хитом и стремительно распространилась по сети. У Джона теперь есть репутация композитора, он собрал вокруг себя сеть из 500 000 тысяч слушателей, фанатов и критиков. В истории Джона музыкальные издатели, распространители и журналисты не имели никакого значения. Джон полагался на неформальное распространение,

критику и *построение социальной репутации*. Затем ему предложила контракт звукозаписывающая компания, который он подписал, чтобы добиться большего признания и распространения своей музыки.

4.3.3. История Софи

Софи – это младшая сестра Джона, нейробиолог, которая только что защитила кандидатскую диссертацию. Софи владеет интернетом так же хорошо, как и Джон, и ждет, что ее карьера будет очень перспективной. Софи знает, что ее будущее в качестве исследователя будет зависеть от ее способности внести свой вклад в нейробиологию в виде оригинальных и подтвержденных методов и результатов, а также от достаточного финансирования.

Чтобы убедить фонды профинансировать ее исследования, Софи нужно найти способ, который подтвердит, что ее научно-исследовательские проекты имеют важное значение для нейробиологии, и что ее методы и результаты корректны.

Конечно, Софи в курсе, что может публиковать статьи в рецензируемых журналах, которым доверяют научные фонды, но она уже знает об опыте Джона и разочарована медлительностью процесса рецензирования, затратами на публикацию и сложными и непрозрачными механизмами подсчета научной репутации и степени влияния.

На самом деле Софи, как и Джон, ценит расширение прав и возможностей, актуальность, прозрачность и качественные показатели оценки ее работы, в отличие от автоматизированных и количественных показателей влияния ее исследований.

4.3.4. Мир Софи

Софи не нуждается в 500 000 тысячах просмотров и перепостов. У нее в смартфоне есть адреса электронной почты 20 коллег со всего мира, которые специализируются в ее области, 20 нейробиологов, которые могли бы отрецензировать ее работу. Всё, что ей нужно, это место, где она может продемонстрировать, что она уважает правила *научной репутации*, что ее методы и результаты действительно были просмотрены и оценены квалифицированными и объективными коллегами. В этом месте информация должна распространяться широко и быстро, чтобы ее могли обсуждать и рекомендовать в эксклюзивном сообществе коллег-специалистов.

Наконец, научные фонды, как правило, перегружены предложениями, поэтому, чтобы привлечь к себе внимание, Софи придется предложить им обобщенную и всеобъемлющую информацию, которая даст представление о том, насколько ее исследование действительно значимо, оригинально и одобрено коллегами, которые считают, что оно будет полезно для науки и для развития человечества в целом.

Именно эти вопросы решает платформа peerevaluation.org, осознающая, что платформа, поддерживающая открытую науку (Open Science), коллективное рецензирование и распространение не может добиться успеха без мощной материальной поддержки, инновационного управления правом собственности и, наконец, надежного представления научной достоверности, которые отвечают ожиданиям политиков и финансирующих организаций.

Peerevaluation.org стремится стать тем местом, в которое ученые приходят, чтобы быть уверенными в том, что они получают лучшее из возможного: повышение распространения, видимости, доступности; комментарии и обсуждения, плодотворное сотрудничество и, наконец, доказательство воздействия, влияния и повторного использования их результатов.

Базовая модель peerevaluation.org для распространения и удаленной пре- и постпубликационной экспертной оценки и комментариев реализуется следующим образом: (а) вы загружаете PDF-файл вашей последней статьи; (б) вы экспортируете аннотацию статьи и ссылку на PDF и размещаете их в вашем блоге, аккаунте системы Mendeley и в репозитории, например, CiteSeerX, (в) одновременно эта информация индексируется в Google Scholar и Microsoft Academic Search; (г) вне зависимости от того, где находится сам файл, люди могут комментировать, обсуждать и рекомендовать его, делиться им, иметь доступ к метрикам и показателям влияния ваших статей; (д) все эти удаленные социальные взаимодействия одновременно объединяются и отображаются в вашем аккаунте на peerevaluation.org, где их можете видеть вы и все те, кому они могут пригодиться.

5. ПОДХОДЫ К КОЛЛЕКТИВНОЙ ОЦЕНКЕ

Проблемы, связанные с рецензированием, и новые инструменты экспертной оценки, которые возникли после появления Web 2.0, стали спусковым крючком для новых направлений в оценке научных работ, выводя на первый план экспертной оценки доверие и репутацию (см., например, подход Peerevaluation.org). Репутация

отражает мнение сообщества о результатах работы того или иного исследователя по одному или нескольким критериям. В этом разделе мы рассмотрим два подхода к оценке исследований, которые по-новому используют явную и неявную обратную связь научного сообщества, а именно:

1. OpinioNet подсчитывает репутацию исследователей, основываясь на общественном мнении: отзывах или цитатах;
2. UCount использует специальные опросы, которые позволяют сообществу сделать вывод о результатах работы ученого как в качестве исследователя, так и в качестве рецензента.

5.1. OpinioNet: репутация научного исследования, основанная на большом количестве мнений

OpinioNet – это инструмент, который основан на большом количестве мнений, сведенных в структурные графы. В OpinioNet репутация поданной научной работы строится не только на основе отзывов, которые она получает, но и на том положении, которое она занимает в структурном графе публикаций. Например, конференция имеет хорошую репутацию, потому что принимает качественные статьи. Кроме того, люди обычно предполагают, что при отсутствии какой-либо информации о представленной статье, тот факт, что статью принял журнал с хорошей репутацией, означает, что и статья должна быть хорошего качества. Таким образом, существует понятие распространения мнений и, как следствие, репутации через отношения «является частью» в структурном графе.

На рис. 1 представлен пример структурного графа исследовательской работы. На этом рисунке есть конференция-серия CS , которая проводит несколько конференций $\{CP_1, \dots, CP_n\}$, проведение каждой конференции состоит из множества статей. Кроме того, существует журнал J , который имеет множество выпусков $\{V_1, \dots, V_n\}$, каждый из которых состоит из множества статей. Заметим, что если бы статьи были разделены на секции $\{S_1, \dots, S_n\}$, то некоторые секции, такие как «обзор литературы» и «исходные данные» могли бы быть включены в различные статьи.

Текущая количественная оценка репутации в публикационной сфере фокусировалась, в основном, на показателях на основе цитирования, таких, как индекс Хирша. Прямые рецензии (или мнениям) не уделяли должного внимания за пределами публикационного процесса из-за того, что в публикационной сфере такая информация, как правило, отсутствует. В то же время, эта информация очень распространена в моделях, основанных на электронной коммерции, таких, как Amazon или eBay. OpinioNet решает эту проблему, предоставляя средства, которые помогают одному исследователю составить свое мнение о научных статьях или об ученых на основе их собственных мнений о составных частях мирового структурного графа публикаций. Соответственно, репутация (или коллективное мнение) рассчитывается путем объединения мнения отдельных исследователей.

Кроме того, OpinioNet также может использовать косвенные мнения. При вычислении репутации исследователей и их научно-исследовательской работы мы говорим о том, что достаточно большое количество существующей информации может быть интерпрети-

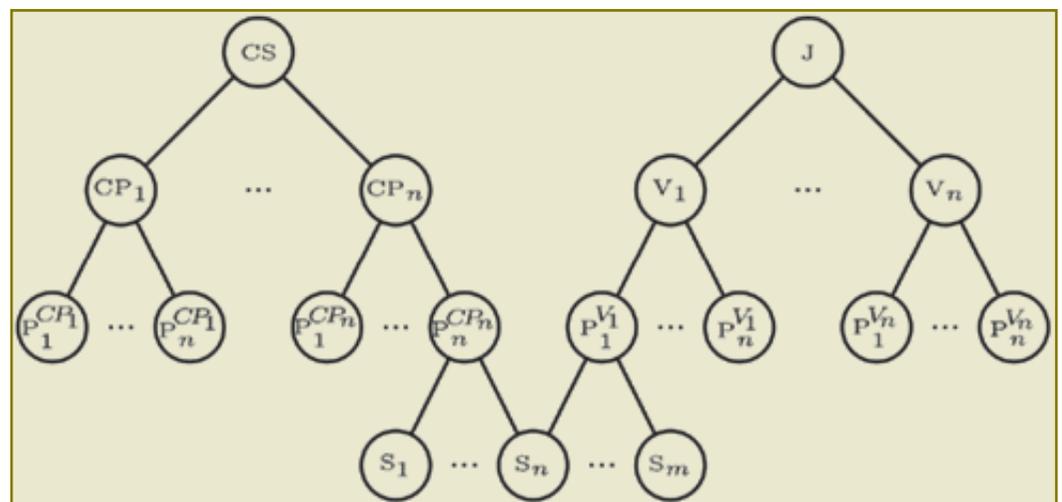


Рис. 1. Образец структурного графа в области публикаций

ровано как мнение о данном исследователе или исследовательской работе. Например, нынешняя система публикаций дает нам прямые (явные) мнения: количественные показатели. Кроме того, могут быть рассмотрены косвенные (неявные) мнения. Например, цитаты можно рассматривать как индикатор того, насколько хорошей является эта исследовательская работа, т.е. положительное мнение ссылающегося автора о цитируемой научно-исследовательской работе. Подписку на журнал можно рассматривать как индикатор того, насколько хорошо журнал принимают в обществе, т.е. положительное мнение читателя о журнале. Существуют большие объемы информации, которая может быть истолкована как мнение. Алгоритм OpinioNet (Osman et al., 2010b) использует эти мнения вне зависимости от того, были ли они прямыми или косвенными, выводит мнение исследователя о данной научной статье [8] и на основании таких мнений выводит общее мнение исследовательского сообщества о статье. Более того, OpinioNet может быть использован для любой комбинации информационных источников, хотя в разных областях исследования один показатель может иметь разный вес.

Таким образом, OpinioNet легко настраивается в соответствии с требованиями разных сообществ или дисциплин. Например, известно, что в зависимости от дисциплины подход к оценке исследований очень разнится. С OpinioNet можно выбрать источник и/или отдельные мнения, чтобы сосредоточиться на тех показателях, которые имеют больший вес. Например, можно настроить OpinioNet так, чтобы использовать только собственные мнения исследователя, прямые мнения сообщества целиком, показатели только на основе цитирования или на сочетании цитирований и прямых мнений. OpinioNet может также увеличить вес статей, принятых в журнал, а не на конференцию, или наоборот, и т.д.

Кроме того, у OpinioNet нет необходимости дополнительно стимулировать исследователей менять устоявшиеся практики. Конечно, наличие открытой системы, в которой люди читают и оценивают другие работы, могло бы быть очень выгодным. Но OpinioNet при этом работает с данными, которые доступны уже сейчас. Уже есть огромное число мнений, как прямых, так и косвенных, таких как рецензии, цитаты, факт принятия статьи в журнал или на конференцию, подписки на журналы, ссылки на нетра-

диционные источники (например, блоги) и т.д. Что действительно необходимо сейчас, так это системы, такие, как OpinioNet, которые могут получать доступ к такой информации, интерпретировать ее и вычислить репутацию научной работы. В настоящее время, как мы считаем, доступ к таким данным и их обработка представляют наибольшие трудности.

OpinioNet затрагивает проблему наличия потенциального предубеждения во мнении, принимая во внимание репутацию оценки источника мнения, чтобы убедиться в достоверности заключения. Например, если мы говорим о том, что человек очень хорошо разбирается в определенной области, он, как правило, будет справедливо оценивать других ученых, которые в этой же области работают.

Это догма. Примером существующей практики на основе этой догмы является выбор членом комитетов, консультативных советов и т.д. Помимо слепого доверия экспертам, можно использовать дополнительные методы для определения в OpinioNet предвзятости исследований или попыток обмануть систему. Например, изучая предыдущие отзывы человека, можно сказать, является ли он предвзятым по отношению к определенному полу, национальности, технике проведения исследования и т.д. Кроме того, при анализе предыдущих отзывов можно сказать, насколько сильно совпадало мнение этого человека с мнением его коллег. Прошлый опыт может использоваться и для оценки потенциальных попыток обмануть систему, сговора. Вся эта информация является взаимодополняющей и может быть использована OpinioNet, чтобы помочь определить степень достоверности суждений.

После рассказа об основных понятиях и целях OpinioNet, мы кратко расскажем о технической стороне вопроса. Для получения более подробной информации мы рекомендуем читателю обратиться к работам Osman и др. (2010b). И для получения информации об оценке OpinioNet и его влияния на поведение исследователей посредством моделирования, мы рекомендуем обратиться к другой работе Osman et al. (2011).

5.1.1. Репутация научной работы

Репутация научной работы основывается на распространении и агрегации мнений в структурном графе. Алгоритм распространения мнений в OpinioNet базируется на трех главных концепциях:

- *Влияние узла.* От того, что исследователи могут написать и разделить свою иссле-



репутация источника мнения используется для определения, насколько корректны мнения, им предоставленные

довательскую работу на разные «дочерние узлы» (например, на секции статей или различные статьи в сборнике материалов конференций), невозможно узнать точный вес для каждого дочернего узла при оценке его влияния на родительские узлы (и наоборот). В OpinioNet влияние конкретного узла n во время t основано на соотношении узлов, которые получили прямые отклики в структурном поддереве n . Другими словами, в оценке репутации узла OpinioNet опирается на внимание, которые узел получает (вне зависимости от того, положительное оно или отрицательное). Например, если одна статья из журнала получила огромное количество откликов (положительных или отрицательных), а другая не получила никакого внимания, то та статья, которая получила огромное количество откликов, будет иметь более сильное влияние на репутацию журнала, чем «пустая» статья.

- *Направление распространения.* Направление распространения в структурном графе имеет решающее значение. Каждое распространение имеет особое значение. «Нисходящее» распространение оценивается, чтобы определить мнение по умолчанию, например, когда статья наследует репутацию журнала, который ее опубликовал. Мнение по умолчанию представляет мнение об узле, которое наследуется от родителей и, как правило, используется, когда не хватает информации о дочерних узлах, которые помогают формировать общее мнение об узле, о котором идет речь. «Восходящее» направление определяет развитие мнений, как, например, мнение о конференции получается путем агрегирования репутации поданных статей. Затем каждый раз, когда новое мнение добавляется к узлу в графе, то репутация по умолчанию и динамическая репутация соседних узлов, соответственно, обновляются. Затем обновление значений одного узла инициирует обновление соседних узлов, в результате чего распространение волны происходит во всем структурном графе.

- *Снижение информационной ценности.* Мы говорим о том, что все теряет свою ценность со временем. Отзывы не являются исключением, и мнения о некоторых узлах n , сделанных во время t со временем теряют

свою ценность (очень) медленно, стремясь в сторону распределения вероятности (или репутации по умолчанию), следуя функции снижения, которая заставляет мнение об узле сходить к мнению по умолчанию.

Хотелось бы заметить, что OpinioNet распространяет мнения *одного* ученого о заданном атрибуте (например, качество исследования) в структурном графе. Однако мнения могут быть представлены с несколькими атрибутами, например, новизна, правильность исследования, и т.д. Мнения могут быть получены от более чем одного ученого. В этом случае различные методы агрегации могут быть использованы, чтобы получить окончательное мнение группы о заданной исследовательской работе. Работа Osman et al. (2010b) содержит несколько примеров того, как проводить агрегацию мнений для получения конечного значения репутации. Однако, как говорилось ранее, важно заметить, что репутация каждого источника мнения используется для определения, насколько корректны мнения, им предоставленные. Другими словами, репутация источников мнения (ученых) используется для взвешивания мнения при агрегации.

5.1.2. Репутация исследователей

У каждого узла структурного графа есть свой собственный автор или множество соавторов. Авторы разных частей статьи могут различаться, хотя могут быть и некоторые пересечения в множествах соавторов. Аналогично, авторы разных статей на конференции могут различаться. И так далее. В OpinioNet репутация автора в данный момент времени представляет собой совокупность репутаций всех его научных работ. Однако при подсчете учитывается количество соавторов, которое есть у каждого документа. Таким образом, при использовании совокупности репутаций (см. Osman et al., 2010a) получается, что чем больше соавторов у научно-исследовательской работы, тем меньшее влияние она оказывает на репутацию каждого из ее соавторов.

5.2. UCount: коллективная оценка исследований

Метод UCount [9] (Parra et al., 2011) предоставляет средства для коллективной оценки суммарного значения научных исследований и деятельности исследователей в качестве рецензентов. Оценка суммарного значения научных достижений производится путем опросов [10], которые показывают мнение научного сообщества о том,

1. Указал ли рецензент на важность исследуемого вопроса?				
1	2	3	4	5
совсем нет			вопрос всесторонне рассмотрен	
2. Оценил ли рецензент оригинальность статьи?				
1	2	3	4	5
совсем нет			вопрос всесторонне рассмотрен и приведены ссылки	
3. Отметил ли рецензент сильные и слабые стороны исследования (план исследования, совокупность данных, анализ данных)?				
1	2	3	4	5
совсем нет			проведен глубокий анализ	
4. Оставил ли рецензент полезные для автора комментарии по поводу языка, стиля, графических и табличных данных?				
1	2	3	4	5
совсем нет			дана исчерпывающая информация	
5. Прокомментировал ли рецензент авторскую интерпретацию результатов?				
1	2	3	4	5
совсем нет			вопрос всесторонне рассмотрен	
6. Как вы в целом оцениваете качество рецензии?				
1	2	3	4	5
очень плохо			очень хорошо	

Рис. 2. Шкала Review Quality Instrument (RQI), разработанного van Rooyen и соавторами (1999). Общий балл вычисляется как среднее арифметическое первых 7-ми пунктов, в то время как 8-й общий пункт обеспечивает дополнительный контроль качества

насколько ценный вклад определенный исследователь внес в науку. Результаты отражаются в виде рейтинга. В этом разделе мы опишем использование UCount для оценки рецензентов.

UCount для оценки рецензентов был специально разработан для получения данных об ученых на основе *их деятельности в качестве рецензентов*, в противовес другим критериям, например, библиометрическим: хороший исследователь не обязательно будет хорошим рецензентом (Black et al., 1998). UCount интегрирован в вышеупомянутый e-Scripts, систему рецензирования, которая используется для журналов серии ICST Transactions. Это позволяет авторам и редакторам получать обратную связь от рецензентов, которые используют Review Quality Instrument (RQI), разработанный для редакторов British Medical Journal (van Rooyen et al., 1999). Этот инструмент, основанный психометрией, используется в нескольких исследованиях экспертной оценки (Jefferson et al., 2007).

RQI представляет собой опросник, который состоит из шкалы из 8 пунктов (см. рис. 2), в которой каждый пункт оценивается по 5-балльной шкале Лайкерта (1=плохо; 5=отлично). Первые 7 пунктов позволяют узнать о разных аспектах рецензирования,

включая, в том числе, обсуждение важности и оригинальности работы, комментарии о сильных и слабых сторонах метода исследования и представления результатов, конструктивности комментариев и тому, насколько они подкреплены ссылками на статью. 8-й (последний) пункт позволяет оценить качество самого рецензирования и может быть рассчитан как среднее значение первых 7 пунктов.

На основе этих оценок каждые три месяца (привязанный к выпускам ICST Transactions) формируется рейтинг рецензентов. Рецензенты, представившие по меньшей мере три отзыва, будут ранжироваться по нескольким критериям: лучший общий результат, общее количество законченных отзывов, полезность и конструктивность обратной связи. Кроме того, в процессе выбора рецензентов для статьи редактор сможет видеть рейтинг рецензентов, составленный на основе их предыдущей работы. Рейтинг будет основан на обратной связи RQI.

- Первые места будут занимать кандидаты со средним значением RQI выше определенного порога (мы предполагаем, что начальное значение порога может равняться медиане, 3), упорядоченные по значению RQI.

- Далее следуют кандидаты без RQI, включая как новых рецензентов, так и рецензентов, которые завершили меньше 3 рецензий в последние

12 месяцев. Эти кандидаты будут упорядочены в стандартном порядке: потенциальный рецензент-автор-редактор.

- Последними будут кандидаты в рецензенты, у которых средний RQI ниже приемлемого порога, упорядоченные в порядке уменьшения рейтинга.

По возможности, RQI для кандидатов будет отображаться для уточнения рейтинга. У редактора будет возможность изменить порядок выбора рецензентов, как он считает нужным. Мы считаем, что это даст возможность выбирать лучших рецензентов, а также обеспечит им признание в научном сообществе, в отличие от нынешней ситуации в большинстве журналов, где только члены редколлегии получают бонусы, в то время как рецензенты остаются неизвестными.

UCount сейчас внедряется для работы со статьями в Европейском альянсе по инновациям (EAI) и Институте компьютерных наук, социальной информатики и телекоммуникаций (ICST).

6. ВЫВОДЫ И ОБСУЖДЕНИЯ

В этой статье мы представили ряд возможных дополнений и альтернатив для традиционного способа экспертной оценки. Разнообразие этих подходов отражает широкий спектр дополнительных факторов, которые могут использоваться при определении значимости научного вклада. Действительно, качество работы очень часто зависит от контекста: например, в некоторых случаях технически несовершенная, но творческая и эмоциональная работа может быть более ценной, чем тщательно выверенное исследование (Underwood, 2004), а в некоторых случаях – наоборот. Разнообразные потребности требуют разнообразных решений.

Выборка подходов к видам рецензирования далеко не полная, она отражает в первую очередь исследования, проведенные в рамках проекта LiquidPub его участниками и партнерами [12]. Существует много других подходов, которые могли бы быть дополнением к экспертной оценке, например, подобные методы реализуются на сайте Faculty of 1000 [13] или в персонализированных рекомендательных системах (Adomavicius and Tuzhilin, 2005; Zhou et al., 2010).

Далее мы обсудим спорные моменты подходов, которые рассмотрены в статье.

6.1. Тендер как индикатор важности

Учитывая известные результаты, отраженные в статистике скачиваний статьи (Brody

et al., 2006) и заключения по тендерному эксперименту, описанному в Части 1, можно ожидать, что система голосования будет довольно надежным (хотя и не без недостатков) средством, которое сможет определить величину будущего влияния научно-исследовательской работы. Проблема может быть в следующем: как и в случае с цитированиями, люди могут голосовать не только за те статьи, которые их интересуют, но и за те статьи, которые они готовы критиковать и отклонять.

Мы склоняемся к тому, что это не так часто случается, как можно подумать, по двум основным причинам. Во-первых, результаты онлайн рейтингов, таких как 5-звездная система рейтинга, используемая YouTube, показывают, что существует очень сильный уклон в сторону положительных оценок и что люди чаще всего просто безразлично относятся к тем вещам, которые им не нравятся, а не активно их критикуют (Hu et al., 2009). Можно предположить, что аналогичный принцип сработает и в случае с тендером: потенциальные рецензенты будут игнорировать плохие статьи, а не терять драгоценное время и не критиковать то, что по их мнению будет отклонено в любом случае. Во-вторых, если оставить в стороне плохие статьи, мы можем предположить высокую активность рецензентов при рассмотрении статей, с которыми они категорически не согласны. Это может создать проблему для редактора журнала, который должен контролировать прецеденты конфликта интересов, но это никак не связано с последующим влиянием статьи. Работы на провокационные темы будут цитироваться даже больше, а не меньше.

Дополнительный риск кроется в том, что тендеры проводятся на основе названия и аннотации, что может привести к переоценке статей, которые будут казаться значительно более ценными, чем есть на самом деле. Это, конечно, повсеместная проблема, затрагивающая не только тендерную систему: авторы пытаются рекламировать свои статьи, чтобы привлечь рецензентов и внимание читателя (Lawrence, 2003). Главный вопрос, который нужно будет решить в будущем, будет ли тендерная система искажать статистику сильнее, чем цитирования и статистика скачиваний уже это делают.

Есть и позитивные моменты. Участие в тендере согласуется с одним из ключевых мотивов для участия ученых в рецензировании: таким образом они могут улучшить и внести свой вклад в работу коллег (Goodman et al., 1994; Purcell et al., 1998; Sense About Science, 2009). Этот этический аспект профессиональ-

ного альтруизма помогает компенсировать риски, описанные выше, и объясняет еще одну причину, почему тендерная система имеет такую большую значимость и влияние – это более привлекательный вариант внести вклад в работу, которая по мнению рецензента будет значимой еще долгое время.

6.2. Peerevaluation.org против UCount

Сервисы Peerevaluation и UCount нацелены на более открытую и прозрачную экспертную оценку. Однако UCount направлен на постепенную модификацию традиционного рецензирования в журнале с введением обратной связи, в то время как Peerevaluation не предлагает каких-то заметных изменений, которые в этом случае сказались бы на работе редактора. Мы считаем, что эти два подхода могли бы быть объединены в будущем, например, выводы UCount могут быть использованы для выбора рецензентов в Peerevaluation, а история предыдущих рассмотрений пригодилась бы для UCount.

6.3. Использование коллективного мнения

Подходы OpinioNet и UCount используют общественное мнение для того чтобы оценить репутацию исследователя. Учитывая то, что большинство не всегда право, вес мнений в OpinioNet зависит от степени доверия к источнику этого мнения, например, опирается на предыдущие решения рецензента. UCount, однако, стремится просто учитывать мнение сообщества без каких-либо дополнительных настроек. UCount не ставит перед собой цель поиска ответа на вопрос: «Действительно ли этот человек является лучшим рецензентом (или исследователем)?», но может утверждать, что «сообщество X считает, что этот человек является лучшим рецензентом (или исследователем)». Оба подхода опираются на полученные от сообщества данные: OpinioNet использует данные, которые доступны на основе цитирования, информации о соавторстве и публикациях, а UCount просит авторов заполнить анкету, данные из которой могут быть использованы как прямые мнения в OpinioNet.

6.4. Стимулирование участников

Предоставление прямых мнений о рецензентах в UCount можно рассматривать как дополнительное усилие со стороны автора. Тем не менее, заполнение опросника –

это минимальное количество временных затрат по сравнению с написанием статьи или рецензии. Поэтому мы считаем, что если действительно хорошие начнут собирать мнения авторов о рецензентах (например, как это было предложено UCount), то люди будут принимать участие в таком проекте, а затем к такой практике присоединятся и другие журналы. Кроме того, и в UCount, и в Peerevaluation у рецензентов есть стимул для того, чтобы писать качественные рецензии, поскольку они либо изначально знают, что они будут оценены либо напрямую (UCount), либо косвенно, потому что отзывы являются открытыми (Peerevaluation). Кроме того, рецензенты таким образом делают себе хорошую рекламу. UCount стимулирует авторов участвовать в процессе, поскольку таким образом они помогают редакторам выбрать лучших рецензентов и, соответственно, получить впоследствии лучшие рецензии. Если в какой-то момент окажется, что хороших рецензентов не хватает, возможно, имеет смысл пересмотреть стимулирующую политику.

Среди спорных, но тем не менее возможных стимулирующих методов, есть плата рецензентам, возможность подачи статьи только после того, как автор отрецензировал три статьи, или снижение платы за регистрацию на конференции для ученых, которые рецензируют поступающие на конференцию статьи.

6.5. Роль Интернета

Давно известно, что существование интернета породило множество возможностей для изменения в области исследовательских публикаций и оценки (Harnad, 1990; Ginsparg, 1994; Swan, 2007). Если говорить о самых простых вещах, то публикация в электронном виде уменьшает расходы на хранение, распространение и пересылку информации до нуля, а также существенно облегчает создание и совместное использование документов (Odlyzko, 1995).

Публикация в электронном виде значительно облегчает поиск и индексирование документов, а скорость электронного обмена информацией позволило значительно снизить время для рассмотрения и публикации научной работы (Spier, 2002). Электронная публикация также позволяет распространение информации во множестве форматов и на разных носителях, помимо стандартного формата научной статьи. Такие новые форматы включают в том числе банки данных, про-

граммное обеспечение, видео и многие другие формы вспомогательных материалов.

Эти же факторы содействуют массовой экспертной оценке, которая описана в этой статье и которая широко используется в социальных сетях, видеохостингах и других онлайн сообществах. Человеку становится легко и дешево оценивать и комментировать электронные объекты, а массовое комментирование и построение рейтинга дает очень много ценной информации для анализа, что в свою очередь принесет пользу в дальнейшем при использовании коллективной оценки (Masum and Zhang, 2004).

Одна из проблем, которая связана с таким подходом, заключается в том, что несмотря на то, что электронная коммуникация служит для расширения доступа и доступности информации, практический эффект от поиска, репутации ученого и других рекомендательных инструментов состоит как раз в том, чтобы сузить этот круг (Evans, 2008). С одной стороны, это может быть связано с улучшением фильтрации плохих работ. Однако вполне возможно, что репозитории и электронные системы оценки повысят и без того известный принцип цитирования «богатый богатеет» (de Solla Price, 1976; Medo et al., 2011) и, возможно, усилят существующее неравенство. Одним из способов решения этого вопроса может быть стимулирование электронных систем оценки делать сильный акцент на разнообразии предоставляемой информации, как на необходимом сервисе (Zhou et al., 2010). Это, конечно, подчеркивает мысль, которую мы ранее высказывали в этой статье, о том, что требуется многообразие показателей для того, чтобы подтвердить, что много разных типов вклада в науку должным образом признаются и вознаграждаются.

Вторая проблема относится к доступности. Многие из инструментов и методов, которые описаны здесь, предполагают повсеместный доступ к интернету и легко доступны в более богатых странах, но эти условия все еще трудно обеспечить во многих странах мира (Best, 2004). Даже там, где доступ не является проблемой, неполадки могут быть, например, в распространении мультимедиа файлов. Тем не менее, электронные технологии и сообщества также служат для того, чтобы сократить географические и экономические неравенства, что в конечном итоге позволяет создавать документы аналогичного качества (Ginsparg, 1994) и допускать виртуальное общение и встречи, стоимость которых в реальном мире может быть очень серьезным препятствием

для исследователей (Gichora et al., 2010). Переход к онлайн сообществам как организаторам научной оценки, безусловно, должно сопровождаться сильным толчком для обеспечения доступа к работам.

6.6. Наше видение рецензирования в будущем

Один из выводов, к которому мы пришли, заключается в том, что ландшафт научных публикаций, несомненно, меняется, как меняются и способы оценки результатов исследования и исследователей. Как можно видеть в разделах 2, 3 и 4.2., цель рецензирования (для поиска ошибок или для улучшения статьи) по-разному истолковывается в разных сообществах. Мы прогнозируем рост разных инструментов для оценки исследовательской деятельности в ближайшие годы, в том числе и проектов с открытым исходным кодом и тех, которые работают с открытым API протоколом. Такие инструменты будут в первую очередь работать в интернете и включать в себя различные методы для оценки, так что председатели программных комитетов конференций и редакторы журналов (или даже люди, которые будут занимать должности, которых пока не существует) будут иметь возможность выбирать наиболее удобные для себя инструменты. Примеры инструментов с такими функциональными возможностями уже появляются (это, например, Mendeley, Peerevaluation.org, Interdisciplines), но пока не очень хорошо понятно, как эти инструменты будут связаны друг с другом и какие из них будут использоваться достаточно широко для того, чтобы стать нормой. Мы считаем, что разные инструменты и практики рецензирования будут приняты в разных научных сообществах и что не существует уникального подхода, который бы мог удовлетворить спрос всех ученых на планете. Более того, одному и тому же исследователю при работе в разных контекстах понадобятся разные инструменты, и эффективные системы оценки должны иметь возможность дать эту альтернативу, она должна быть уже встроена в дизайн [14]. Учитывая это, следует уделять меньше внимания проектированию «единой системы научной оценки завтрашнего дня» - просто потому, что, например, экспертная оценка будет развиваться по-разному при учете потребностей различных научных дисциплин и сообществ. Вместо этого внимание следует сосредоточить на обеспечении разнообразия и взаимодействия многих возможных инструментов, которыми можно будет воспользоваться для того, чтобы оценить результаты научной работы.

ЗАЯВЛЕНИЕ О КОНФЛИКТЕ ИНТЕРЕСОВ

Авторы утверждают, что исследование было проведено без влияния каких-либо коммерческих или финансовых отношений, которые могли бы быть истолкованы как потенциальный конфликт интересов

БЛАГОДАРНОСТИ

Эту работу поддерживал проект Liquid Publication EU ICT. Проект LiquidPub благодарит за финансовую поддержку программу Future and Emerging Technologies (FET) в рамках седьмой рамочной программы для исследований Европейской комиссии в соответствии с грантом FET-Open. Номер гранта: 213360.

Авторы благодарят коллектив журнала «Научная периодика: проблемы и решения» за перевод статьи на русский язык, выполненный с согласия всех авторов и при участии Александра Бирюкова и Кацярыны Мирыленка.

Сноски

1. «С согласия авторов в уже отрецензированной и принятой к публикации в *VMJ* статью было внесено 8 ошибок в оформление, анализ или интерпретацию данных» (Godlee et al., 1998).

2. Это показывает странное расхождение с результатами Orthof et al. (2002), в которых большее число рецензентов сделали более точный прогноз. Объяснить это можно тем, что в медицинских исследованиях может быть огромное число различных факторов, которые обязательно нужно учитывать при оценке статьи, следовательно, несколько рецензентов с разным опытом смогут обеспечить более высокое качество рецензирования.

3. Корреляция между рейтингами на основе рейтингов рецензентов и рейтингов на основе цитирований была измерена с помощью τ Кендалла для 5 различных конференций, среди которых были 2 слабые, но статистически значимые корреляции ($\tau = 0,392$, $p = 0,0001$ и $\tau = 0,310$, $p = 0,005$; две конференции имели, соответственно, 150 и 100 представленных статей). На других более крупных конференциях не было статистически значимых корреляций (Ragone et al., 2013).

4. *PloP расширяется как Pattern Languages of Programs и включает в себя несколько конференций: EuroPloP, PloP, VikingPloP, etc. По ссылке можно найти полный список <http://www.hillside.net/europlop/europlop2011/links.html>

5. <http://interdisciplines.org/> - веб-сайт, на котором междисциплинарные конференции создаются в формате диалога.

6. Следует отметить, что технически этот эксперимент ближе к квази-эксперименту, потому что рецензентам было разрешено выбирать тип процесса рецензирования. Если любая из групп содержала более опытных участников, это могло бы повлиять на результаты.

7. <http://escripts.icst.org/>

8. Как могут определяться косвенные мнения, смотрите в исследовании Osman et al. (2010a).

9. <http://icst.org/ucount/>

10. Примеры таких обзоров можно посмотреть по ссылке: <http://icst.org/UCount-Survey/>

11. <http://project.liquidpub.org/>

12. Полный обзор исследований, проведенный по этим темам, можно посмотреть по ссылке: <http://project.liquidpub.org/research-areas/research-evaluation>.

13. <http://f1000.com/>

14. Например, Confy, система рецензирования статей для конференций, которую использует EAI и ICST, позволяет выбирать вариативные модели для проведения рецензирования – с тендером или без, настраивает формы рецензии и использует другие настройки. Confy сейчас доступна по ссылке <http://cameraready.eai.eu/>.

Источники

1. Adomavicius, G., and Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* 17, 734–749.

2. Akst, J. (2010). I hate your paper. *Scientist* 24, 36.

3. Best, M. L. (2004). Can the internet be a human right? *Hum. Rights Hum. Welf.* 4, 23–31.

4. Black, N., van Rooyen, S., Godlee, F., Smith, R., and Evans, S. (1998). What makes a good reviewer and a good review for a general medical journal? *J. Am. Med. Assoc.* 280, 231–233.

5. Bornmann, L. (2007). Bias cut. women, it seems, often get a raw deal in science – so how can discrimination be tackled? *Nature* 445, 566.

6. Bornmann, L., and Daniel, H.-D. (2005a). Committee peer review at an international research foundation: predictive validity and fairness of selection decisions on post-graduate fellowship applications. *Res. Eval.* 14, 15–20.

7. Bornmann, L., and Daniel, H.-D. (2005b). Selection of research fellowship recipients by committee peer review. reliability, fairness and predictive validity of board of trustees' decisions. *Scientometrics* 63, 297–320.

8. Bornmann, L., and Daniel, H.-D. (2010a). The validity of staff editors' initial evaluations of

manuscripts: a case study of *Angewandte Chemie International Edition*. *Scientometrics* 85, 681–687.

9. Bornmann, L., and Daniel, H.-D. (2010b). The usefulness of peer review for selecting manuscripts for publication: a utility analysis taking as an example a high-impact journal. *PLoS ONE* 5, e11344. doi:10.1371/journal.pone.0011344

10. Bornmann, L., Wallon, G., and Ledin, A. (2008). Does the committee peer review select the best applicants for funding? An investigation of the selection process for two European molecular biology organization programmes. *PLoS ONE* 3, e3480. doi:10.1371/journal.pone.0003480

11. Brody, T., Harnad, S., and Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *JASIST* 58, 1060–1072.

12. Burnham, J. C. (1990). The evolution of editorial peer review. *J. Am. Med. Assoc.* 263, 1323–1329.

13. Casati, F., Marchese, M., Mirylenka, K., and Ragone, A. (2010). Reviewing Peer Review: A Quantitative Analysis of Peer Review. Technical Report 1813. University of Trento. Available at: <http://eprints.biblio.unitn.it/archive/00001813/>

14. Ceci, S. J., and Peters, D. P. (1982). Peer review: a study of reliability. *Change* 14, 44–48.

15. Ceci, S. J., and Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3157–3162.

16. Cho, M. K., Justice, A. C., Winker, M. A., Berlin, J. A., Waeckerle, J. F., Callahan, M. L., and Rennie, D. (1998). Masking author identity in peer review: what factors influence masking success? *PEER Investigators*. *JAMA* 280, 243–245.

17. de Solla Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inf. Sci.* 27, 292–306.

18. Evans, J. A. (2008). Electronic publication and the narrowing of science and scholarship. *Science* 321, 395–399.

19. Fisher, M., Friedman, S. B., and Strauss, B. (1994). The effects of blinding on acceptance of research papers by peer review. *J. Am. Med. Assoc.* 272, 143–146.

20. Gichora, N. N., Fatumo, S. A., Ngara, M. V., Chelbat, N., Ramdayal, K., Opap, K. B., Siwo, G. H., Adebisi, M. O., El Gonnouni, A., Zofou, D., Maurady, A. A. M., Adebisi, E. F., de Villiers, E. P., Masiga, D. K., Bizzaro, J. W., Suravajhala, P., Ommeh, S. C., and Hide, W. (2010). Ten simple rules for organizing a virtual conference – anywhere. *PLoS Comput. Biol.* 6, e1000650. doi:10.1371/journal.pcbi.1000650

21. Ginsparg, P. (1994). First steps towards electronic research communication. *Comput. Phys.* 8, 390–396.

22. Godlee, F. (2002). Making reviewers visible: openness, accountability, and credit. *JAMA* 287, 2762–2765.

23. Godlee, F., Gale, C. R., and Martyn, C. N. (1998). Effect on the quality of peer review of blinding reviewers and asking them to sign their reports a randomized controlled trial. *JAMA* 280, 237–240.

24. Goodman, S. N., Berlin, J., Fletcher, S. W., and Fletcher, R. H. (1994). Manuscript quality before and after peer review and editing at *annals of internal medicine*. *Ann. Intern. Med.* 121, 11–21.

25. Greaves, S., Scott, J., Clarke, M., Miller, L., Hannay, T., Thomas, A., and Campbell, P. (2006). Overview: Nature's peer review trial. *Nature*. doi:10.1038/nature05535.

26. Harnad, S. (1990). Scholarly skywriting and the prepublication continuum of scientific enquiry. *Psychol. Sci.* 1, 342–344.

27. Hu, N., Pavlou, P. A., and Zhang, J. (2009). Overcoming the J-shaped distribution of product reviews. *Commun. ACM* 52, 144–147.

28. Ingelfinger, F. J. (1974). Peer review in biomedical publication. *Am. J. Med.* 56, 686–692.

29. Jefferson, T., Rudin, M., Folse, S. B., and Davidoff, F. (2007). Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane* 41, MR000016.

30. Jefferson, T., Wager, E., and Davidoff, F. (2002a). Measuring the quality of editorial peer review. *JAMA* 287, 2786–2790.

31. Jefferson, T., Alderson, P., Wager, E., and Davidoff, F. (2002b). Effects of editorial peer review: a systematic review. *JAMA* 287, 2784–2786.

32. Justice, A. C., Cho, M. K., Winker, M. A., Berlin, J. A., Rennie, D., and PEER Investigators. (1998). Does masking author identity improve peer review quality? A randomized controlled trial. *JAMA* 280, 240–242.

33. Kassirer, J. P., and Campion, E. W. (1994). Peer review: crude and understudied, but indispensable. *J. Am. Med. Assoc.* 272, 96–97.

34. Katz, D. S., Proto, A. V., and Olmsted, W. W. (2002). Incidence and nature of unblinding by authors: our experience at two radiology journals with double-blinded peer review policies. *Am. J. Roentgenol.* 179, 1415–1417.

35. Kronick, D. A. (1990). Peer review in 18th-century scientific journalism. *JAMA* 263, 1321–1322.

36. Lawrence, P. A. (2003). The politics of publication. *Nature* 422, 259–261.

37. Lee, K., Boyd, E., Holroyd-Leduc, J., Bacchetti, P., and Bero, L. (2006). Predictors of publication: characteristics of submitted manuscripts associated with acceptance at major biomedical journals. *Med. J. Aust.* 184, 621.

38. Link, A. M. (1998). Us and non-US submissions: an analysis of reviewer bias. *JAMA* 280, 246–247.
39. Lock, S. (1994). Does editorial peer review work? *Ann. Intern. Med.* 121, 60–61.
40. Lynch, J. R., Cunningham, M. R., Warme, W. J., Schaad, D. C., Wolf, F. M., and Leopold, S. S. (2007). Commercially funded and united states-based research is more likely to be published; good-quality studies with negative outcomes are not. *J. Bone Joint Surg. Am.* 89, 1010–1018.
41. Marsh, H. W., Bornmann, L., Mutz, R., Daniel, H.-D., and O'Mara, A. (2009). Gender effects in the peer reviews of grant proposals: a comprehensive meta-analysis comparing traditional and multilevel approaches. *Rev. Educ. Res.* 79, 1290–1326.
42. Masum, H., and Zhang, Y.-C. (2004). Manifesto for the reputation society. *First Monday* 9 [Online].
43. McCook, A. (2006). Is peer review broken? *Scientist* 20, 26.
44. McNutt, R. A., Evans, A. T., Fletcher, R. H., and Fletcher, S. W. (1990). The effects of blinding on the quality of peer review: a randomized trial. *JAMA* 263, 1371–1376.
45. Medo, M., Cimini, G., and Gualdi, S. (2011). Temporal effects in the growth of networks. Available at: <http://arxiv.org/abs/1109.5560>
46. Medo, M., and Wakeling, J. R. (2010). The effect of discrete vs. continuous-valued ratings on reputation and ranking systems. *Europhys. Lett.* 91, 48004.
47. Odlyzko, A. M. (1995). Tragic loss or good riddance? The impending demise of traditional scholarly journals. *Int. J. Hum. Comput. Sci.* 42, 71–122.
48. Olson, C. M., Rennie, D., Cook, D., Dickersin, K., Flanagan, A., Hogan, J. W., Zhu, Q., Reiling, J., and Pace, B. (2002). Publication bias in editorial decision making. *JAMA* 287, 2825–2828.
49. Opthof, T., Coronel, R., and Janse, M. J. (2002). The significance of the peer review process against the background of bias: priority ratings of reviewers and editors and the prediction of citation, the role of geographical bias. *Cardiovasc. Res.* 56, 339–346.
50. Osman, N., Sabater-Mir, J., and Sierra, C. (2011). “Simulating research behaviour,” in 12th International Workshop on Multi-Agent-Based Simulation (MABS'11), Taipei.
51. Osman, N., Sabater-Mir, J., Sierra, C., de Pinninck Bas, A. P., Imran, M., Marchese, M., and Ragone, A. (2010a). Credit attribution for liquid publications. Deliverable D4.1, Liquid Publications Project. Available at: https://dev.liquidpub.org/svn/liquidpub/papers/deliverables/LP_D4.1.pdf
52. Osman, N., Sierra, C., and Sabater-Mir, J. (2010b). “Propagation of opinions in structural graphs,” in ECAI 2010: Proceedings of the 19th European Conference on Artificial Intelligence, Vol. 215 of *Frontiers in Artificial Intelligence and Applications*, eds H. Coelho, R. Studer, and M. Wooldridge (Lisbon: IOS Press), 595–600.
53. Parra, C., Birukou, A., Casati, F., Saint-Paul, R., Wakeling, J. R., and Chlamtac, I. (2011). “UCount: a community-driven approach for measuring scientific reputation,” in *Proceedings of Altmetrics11: Tracking Scholarly Impact on the Social Web*, Koblenz.
54. Purcell, G. P., Donovan, S. L., and Davidoff, F. (1998). Changes to manuscripts during the editorial process: characterizing the evolution of a clinical paper. *J. Am. Med. Assoc.* 280, 227–228.
55. Ragone, A., Mirylenka, K., Casati, F., and Marchese, M. (2011). “A quantitative analysis of peer review,” in 13th International Society of Scientometrics and Informetrics Conference, Durban.
56. Ragone, A., Mirylenka, K., Casati, F., and Marchese, M. (2013). On peer review in computer science: analysis of its effectiveness and suggestions for improvement. *Scientometrics* 97 (2), 317–356. doi: 10.1007/s11192-013-1002-z
57. Reinhart, M. (2009). Peer review of grant applications in biology and medicine. reliability, fairness, and validity. *Scientometrics* 81, 789–809.
58. Ross, J. S., Gross, C. P., Desai, M. M., Hong, Y., Grant, A. O., Daniels, S. R., Hachinski, V. C., Gibbons, R. J., Gardner, T. J., and Krumholz, H. M. (2006). Effect of blinded peer review on abstract acceptance. *J. Am. Med. Assoc.* 295, 1675–1680.
59. Sense About Science. (2009). Peer Review Survey: Preliminary Results. Available at: <http://www.senseaboutscience.org.uk/index.php/site/project/29/>
60. Smith, R. (2006). Peer review: a flawed process at the heart of science and journals. *J. R. Soc. Med.* 99, 178–182.
61. Spier, R. (2002). The history of the peer-review process. *Trends Biotechnol.* 20, 357–358.
62. Swan, A. (2007). Open access and the progress of science. *Am. Sci.* 95, 198–200.
63. Underwood, A. J. (2004). It would be better to create and maintain quality rather than worrying about its measurement. *Mar. Ecol. Prog. Ser.* 270, 283–286.
64. van Rooyen, S., Black, N., and Godlee, F. (1999). Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. *J. Clin. Epidemiol.* 52, 625–629.
65. Walsh, E., Rooney, M., Appleby, L., and Wilkinson, G. (2000). Open peer review: a randomised controlled trial. *Br. J. Psychiatry* 176, 47–51.
66. Ware, M., and Monkman, M. (2008). Peer Review in Scholarly Journals: Perspective

of the Scholarly Community – An International Study. Survey Commissioned by the Publishing Research Consortium. Available at: <http://www.publishingresearch.net/PeerReview.htm>

67. Wenneras, C., and Wold, A. (1997). Nepotism and sexism in peer-review. *Nature* 387, 341–343.

68. Zhou, T., Kuscsik, Z., Liu, J.-G., Medo, M., Wakeling, J. R., and Zhang, Y.-C. (2010). Solving the apparent diversity-accuracy dilemma of recommender systems. *Proc. Natl. Acad. Sci. U.S.A.* 107, 4511–4515.

Оригина́л публи́кации: Birukou A, Wakeling JR, Bartolini C, Casati F, Marchese M, Mirylenka K, Osman N, Ragone A, Sierra C and Wassef A (2011) Alternatives to peer review: novel approaches for research evaluation. *Front. Comput. Neurosci.* 5:56. doi: 10.3389/fncom.2011.00056

Aliksandr BIRUKOU

Department of Information Engineering and Computer Science, University of Trento, Trento, Italy, European Alliance for Innovation, Gent, Belgium

Joseph Rushton WAKELING

European Alliance for Innovation, Gent, Belgium

Claudio BARTOLINI

Service Automation and Integration Lab, HP Labs, Palo Alto, CA, USA

Fabio CASATI

Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

Maurizio MARCHESE

Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

Katsiaryna MIRYLENKA

Department of Information Engineering and Computer Science, University of Trento, Trento, Italy

Nardine OSMAN

Artificial Intelligence Research Institute (IIIA-CSIC), Barcelona, Catalonia, Spain

AZZURRA RAGONE

Department of Information Engineering and Computer Science, University of Trento, Trento, Italy, Exprivia SpA, Molfetta, Italy

Carles SIERRA

Artificial Intelligence Research Institute (IIIA-CSIC), Barcelona, Catalonia, Spain

Aalam WASSEF

Peerevaluation.org, Paris, France

Alternatives to Peer Review: Novel Approaches for Research Evaluation

In this paper we review several novel approaches for research evaluation. We start with a brief overview of the peer review, its controversies, and metrics for assessing efficiency and overall quality of the peer review. We then discuss five approaches, including reputation-based ones, that come out of the research carried out by the LiquidPub project and research groups collaborated with LiquidPub. Those approaches are alternative or complementary to traditional peer review. We discuss pros and cons of the proposed approaches and conclude with a vision for the future of the research evaluation, arguing that no single system can suit all stakeholders in various communities.