

ارزیابی کمی سیر تکامل ضریب تاثیر علمی شخصی

معرفی

در اغلب جنبه های عملکرد بشری، از ورزش تا مهندسی، مسیر پیشرفت خوب به منحنی یادگیری تند و تلاش طولانی نیازمند است. علم نیز از این قاعده مبرا نیست: اکتشافات برجسته اغلب بر نشریات با تاثیر کمتر مقدم بوده است. اما صرف نظر از تمایل رو به رشد برای تشخیص دانشمندان خوب تاکنون، الگوهای تخصصی جدیدی که مزیت علمی را مشخصه بندی نماید ناشناخته باقی مانده است.

انگیزه اساسی تحقیق

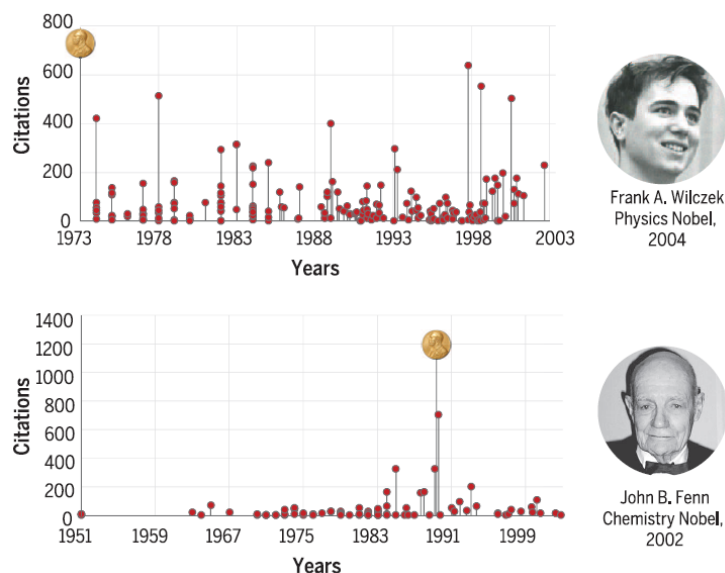
چگونه تاثیر و بهره وری در یک تخصص علمی تغییر خواهد کرد؟ آیا ضریب تاثیر که معمولا شایعترین معیار بازدهی علمی مربوطه است، از الگوهای پیش بینی شده ای پیروی خواهد کرد؟ آیا می توانیم زمان رسیدن یک دانشمند به موفقیتی برجسته را تخمین بزنیم؟ آیا می توانیم تخصص علمی را از لحاظ کمی و مولفه های تخمینی مدلسازی نماییم؟ با این سوالات ارائه شده، در اینجا به بررسی کمی سیر تکامل ضریب تاثیر و بهره وری در بین هزاران کار علمی خواهیم پرداخت. این کار را با بازبینی تاثیر بلندمدت روی جامعه علمی که براساس معیار تعداد نقل قول از مقالات مشخص شده است، در نشریات ثبت شده علمی از هفت رشته مختلف انجام خواهیم داد.

نتایج

به این نتیجه رسیدیم که کار با بالاترین تاثیر در دوره کاری یک دانشمند به طور تصادفی در بدنه کارهای او توزیع شده است. به این معنا که کار با بالاترین تاثیرگزاری با احتمال یکسان می تواند هرجایی از زنجیره مقالات منتشرشده توسط یک دانشمند باشد- یعنی اولین انتشار، در اواسط دوره کاری، یا آخرین انتشار او باشد. این قانون

تأثیر تصادفی، برای دانشمندان رشته های مختلف با دوره تخصصی متفاوت، در حال کار در دهه های مختلف، با انتشار انفرادی یا تیمی و با تایید اعتبار علمی یکنواخت یا متفاوت بین همکاران، صادق است.

قانون تأثیر تصادفی به ما این امکان را خواهد داد که یک مدل کمی را توسعه دهیم که به صورت متقارن قانون بهره وری و موفقیت در هر رشته علمی را روشن سازد. در این مدل فرض بر این است که هر دانشمند یک پروژه را با تأثیر تصادفی p انتخاب کرده و با یک ضریب Q_i آن را بهبود می بخشد و منتج به انتشاری با تأثیر گذاری Q_{ip} می شود. پارامتر Q_i قابلیت دانشمند i برای پیشرفت دانش در دسترس، به طریقی است که تأثیر احتمالی p یک مقاله را بهبود $(Q_i > 1)$ یا کاهش $(Q_i < 1)$ دهد. این مدل پیش بینی می کند که اکتشافات با تأثیر حقیقی بالا به ترکیبی از Q_i و p بالا نیازمندند و بهره وری زیاد به تنهایی نمی تواند شانس ایجاد یک کار با تأثیرگذاری خیلی زیاد را بهبود بخشد. همچنین نشان می دهیم که Q دانشمند که نشان دهنده توانایی شخصی برای انتشار مقالات با تأثیر بالاست، مستقل از مرحله تخصصی او می باشد. این مورد در تقابل با معیارهای کنونی، شامل تعداد کلی نقل قول ها تا مقیاس h - است که با زمان افزایش خواهند یافت. مدل Q توصیفی تجربی از این معیارهای تأثیر سنتی ارائه می کند، به ما امکان پیش بینی سیر تکاملی آینده هر دانشمند را می دهد و دستیابی به معیارهای رسمی نظیر جوایز نوبل را پیش بینی خواهد کرد.



توضیح تصویر بالا:

قانون تاثیرگذاری تصادفی: تاریخچه انتشارات دو برنده جایزه نوبل، فرانک ویلزچک (جایزه نوبل فیزیک، 2004) و جان بی فن (نوبل شیمی، 2002) نشان می دهد که کار با بالاترین تاثیرگذاری می تواند با احتمال یکسان هرجایی از زنجیره مقالات منتشره توسط یک دانشمند باشد. هر خط عمودی مربوط به یک مقاله تحقیقی است. ارتفاع هر خط نشان دهنده تاثیرگذاری مقاله است که براساس تعداد نقل قول ها از مقاله پس از 10 سال کمی سازی شده است. ویلزچک جایزه نوبل را برای اولین مقاله منتشره خود برد، در حالی که فن کار برنده جایزه نوبل خود را حتی پس از بازنشستگی اجباری از دانشگاه، در اواخر دوره کاری خود ارائه کرد.

صرف نظر از استفاده مکرر از شاخص های کمی زیادی برای سنجش تاثیر حرفه ای یک دانشمند، در مورد اینکه چگونه دانشمند در طول زمان پیشرفت و ظهور می کند، اطلاعات کمی وجود دارد. در اینجا، ما تغییرات تاثیرگذاری و بهره وری تخصصی در یک علم را از لحاظ کمی بررسی خواهیم کرد و این تاثیر را برحسب انتشارات موثر خواهیم سنجید که به طور تصادفی درون زنجیره انتشارات دانشمند نهفته است. این قانون تصادفی به ما امکان تنظیم یک مدل تصادفی را خواهد داد که تاثیرات بهره وری، قابلیت شخصی و شانس را از هم مجزا خواهد نمود و وجود الگوهای جهانی پیش برنده موفقیت دانشمندان را به طور واضح نشان خواهد داد. این مدل یک پارامتر شخصی واحد Q را به هر دانشمند نسبت می دهد که در طول کار تخصصی پایدار است، و به طور دقیق سیر تکاملی تاثیر یک دانشمند را با استفاده از مقیاس h برای نقل قول های تجمعی و تقدیرات مستقلی نظیر جوایز، پیش بینی می کند. بهره وری که به تعداد انتشارات تالیف شده توسط دانشمند در طول زمان اشاره دارد و تاثیرگذاری که اغلب با شمار نقل قول های انجام شده از یک انتشار تخمین زده می شود (1-4)، مکررا به عنوان معیارهایی برای سنجش بازدهی دانشمندان استفاده شده اند. صرف نظر از استفاده گسترده، ما از فقدان فهمی کمی از الگوهای این معیارها حین کار تخصصی یک دانشمند رنج می بریم (5). با تطابق رو به افزایش دانشمندان برای کار آکادمیک (4 و 11)، بکارگیری آنها به عنوان ورودی معیارهای متعددی مانند مقیاس h و متغیرهای مربوط به آن که مکررا برای مقایسه بازدهی اشخاص استفاده شده اند (12-14) جالب خواهد بود. با علاقه رو به افزایش برای پیش بینی مقدار این شاخص ها (5 و 15)، در اینجا این پرسش را مطرح می کنیم که: چگونه تاثیر و بهره وری در طول کار حرفه ای یک دانشمند

تغییر خواهد کرد؟ آیا تاثیرگزاری که معمولترین معیار سنجش بازدهی است از الگوهای قابل پیش بینی پیروی می کند؟ آیا می توانیم زمان رسیدن یک دانشمند به برجستگی علمی را پیش بینی کنیم؟ آیا می توانیم قاعده ای برای تاثیر، بهره وری و شانس در طول تخصص کاری یک دانشمند تعریف نماییم؟

برای پاسخ به این سوالات، پروفایل انتشارات دانشمندان را با استفاده از چند قاعده بازسازی می کنیم و هر کدام از انتشارات آنها را با یک ضریب تاثیر مرتبط خواهیم کرد که با نشان داده شده است که معرف تعداد نقل قول ها در 10 سال بعد از انتشار مقاله است (شکل 1الف؛ روش ها و بخش س1 را مشاهده نمایید).

با گرفتن ایده از کارهای با تاثیر بالای دانشمندانی (16 و 17)، نظیر رادیواکتیویته ماری کوری و مارپیچ دوگانه واتسون و کریک، برای هر دانشمند مقاله با بیشترین نقل قولش را با $C*10$ مشخص کردیم، که مقاله ای است که بیشترین تعداد نقل قول را 10 سال پس از انتشار داشته است. توزیع $P(C*10)$ برای دانشمندان مطالعه شده نشان می دهد که فقط 5٪ دارای $C*10$ بالاتر یا مساوی 200 بوده اند، بنابراین اکثر دانشمندان محقق تاثیر بیشینه محدود داشته اند. به منظور تمییز سیستماتیک بین متخصصین برپایه بیشترین تاثیرشان، هر دانشمند را به گروه های با تاثیر بیشینه بالا (5٪ بالا، $C*10$ بیشتر از 200)، تاثیر بیشینه کم (20٪ پایین، $C*10$ کمتر از 20) و تاثیر بیشینه متوسط (75٪ وسط، $C*10$ بین 20 و 200) گروه بندی می کنیم (شکل 1ب و بخش س2).

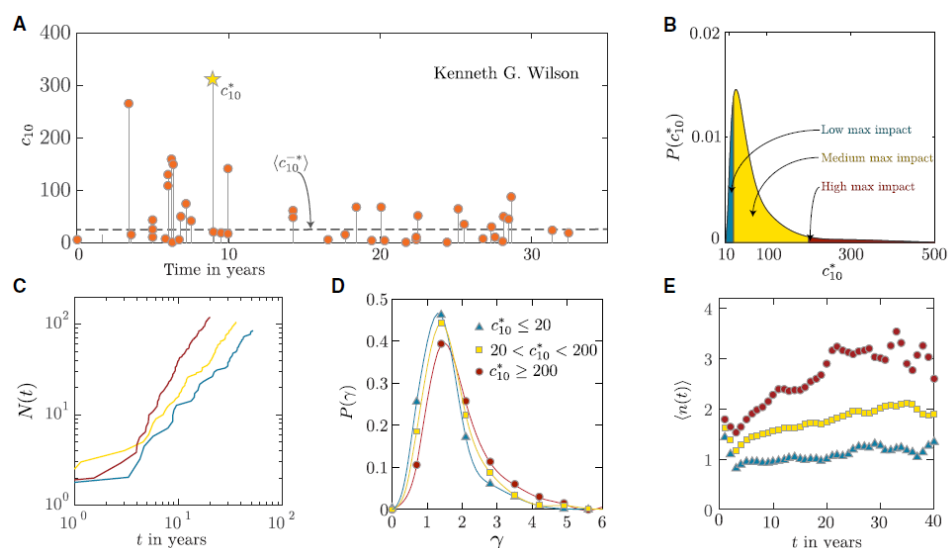
الگوهای تاثیر و بهره وری در تخصص های علمی

تعداد کلی مقالات دانشمند i منتشرشده تا زمان t پس از اولین انتشارش، $N_i(t)$ ، از منحنی مجانب $N_i(t) \sim t_i^\gamma$ پیروی می کند (شکل 1ث) (18). بنابراین، بهره وری سالیانه، $n_i(t)$ ، از مقیاس مشابه با توان $(\gamma_i - 1)$ پیروی می کند (شکل س5). اما توان معادله برای دانشمندان با تاثیر کم، متوسط و بالا متفاوت است (شکل 1ث). به این نتیجه رسیدیم که برای دانشمندان با تاثیر کم این توان برابر با 1/55 است که نشان می دهد به طور میانگین افزایش بهره وری پایداری دارند. برای محققین با تاثیر بالا، افزایش بیشتری وجود دارد که این ضریب 2/05 است (شکل 1د). این معیارها با تغییرات بهره وری سالیانه $n(t)$ نیز تایید می شود: برای دانشمندان با تاثیر بالا، بهره وری اغلب در طول دوره تخصصیشان افزایش می یابد، در حالی که این افزایش برای دانشمندان با تاثیر کم، متوسط خواهد بود (شکل

1ای). شکل 1(د و ای) نشان می دهد که بهره وری در طول تخصص کاری یک دانشمند تغییر خواهد کرد. بهر حال ما دریافتیم که این مقیاس توسط تاثیرگذاری تعدیل خواهد شد: رشد بهره وری برای دانشمندان با تاثیرگذاری بالا قابل ملاحظه است و برای دانشمندان با تاثیرگذاری پایین متوسط است (شکل 1، ث تا ای).

همانطور که شکل 2الف نشان می دهد، تاثیرگذاری از منحنی های مشابه بهره وری پیروی می کند (شکل 1ای). اگرچه c10 در طول تخصص کاری یک دانشمند با تاثیرگذاری بالا افزایش می یابد، افزایش این مقیاس برای اشخاص با تاثیرگذاری کم متوسط قابل ملاحظه تر است. اما اگر تاثیرگذاری در مجاورت t^* (زمان انتشار کار با بیشترین تاثیر C^*10) را بررسی نماییم به الگوی متفاوتتری می رسیم. با رسم c10 برای زنجیره مقالات قبل و بعد از مقاله با بیشترین تاثیر یک دانشمند C^*10 (شکل 2ب)، افزایش تدریجی در تاثیرگذاری قبل و بعد از این شاهکار مشاهده نمی شود. در عوض، الگوی مشاهده شده یک رفتار عجیب را نشان می دهد. این شگفتی می تواند در نتیجه تغییرات میانگین تاثیر تصادفی در تخصص کاری باشد. اما دریافتیم که در صورت استفاده از یک میانگین متحرک با در نظر گرفتن انتشارات با بیشترین تاثیر در یک پنجره چرخشی (بخش س 2. 1 و شکل س 6) نتایج محکمتری حاصل خواهد شد و با استفاده از تئوری تطابق $\hat{c}_{10}^i(t) \sim \alpha_i t$ که قبل و بعد از t^* (بخش س 2.2 و شکل س 7) بدون تغییر α_i است، اعتبارسنجی خواهد شد (19). همچنین، مقالات منتشرشده قبل و بعد از t^* تفاوت قابل تشخیصی از لحاظ میانگین نقل قول ها نشان نمی دهند (شکل 2ث). سرانجام، ما هر تخصص کاری را با بدون تغییر نگه داشتن تمام معیارهای بهره وری [تعداد کلی مقالات N_i و $lni(t)$ و ترکیب تاثیرگذاری تمام مقالات یک تخصص کاری، به طور تصادفی انتخاب کردیم (شکل 2ث). عدم وجود تفاوت بین تخصص های کاری تصادفی شده و اصلی این نتیجه را حاصل کرد: تغییرات تخصصی قبل و بعد از کار با بیشترین تاثیر یک دانشمند وجود ندارد. استحکام این نظریه را برای نمونه های مختلف دانشمندان (شکل های س 8 و س 9)، برای تعاریف مختلف تاثیرگذاری (بخش س 1. 6 و شکل س 10) و در مجموعه داده هایی که برای تاثیر مختلف هر نویسنده یک مقاله وجود داشت (بخش س 6 و شکل س 11) آزمایش کردیم که به نتیجه مشابه رسیدیم. اما به طور قطع نمی توانیم بگوییم که متغیرهای تاثیرگذاری یا مجموعه داده هایی یافت نشود که الگوهایی قبل و بعد از مقاله با بیشترین تاثیرگذاری را رد خود نداشته باشد.

برای فهمیدن زمانی که یک دانشمند کار با بیشترین تاثیرگزاری را منتشر می کند، احتمال $P(t^*)$ را سنجش کردیم که مقاله با بیشترین تاثیرگزاری در زمان t^* بعد از اولین انتشار مقاله منتشر شده است (شکل 2د). $P(t^*)$ بالا بین 0 تا 20 سال نشان می دهد که اغلب دانشمندان مقاله با بیشترین تاثیر خود را در اوایل یا اواسط دوره کاری خود منتشر می کنند. افت $P(t^*)$ پس از 20 سال نشان می دهد که احتمال وجود کار با بیشترین نقل قول یک دانشمند در اواخر دوره کار او باشد که نتیجه ای است که در ادبیات موضوع خلاقیت نیز به خوبی مستند شده است (به بخش س 3. 1 رجوع کنید) (20 و 21). به منظور فهم مبدا این الگو، c_{10} را با حفظ بهره وری وابسته به زمان دانشمند و تاثیر مقاله به مقاله، و انتخاب تصادفی مرتبه انتشارات او، در میان تمام مقالات یک دانشمند ترکیب کردیم. این حقیقت که $P(t^*)$ برای این کارهای ترکیبی از داده های اصلی غیر قابل تمییز است (شکل 2د) نشان می دهد که $P(t^*)$ به دلیل زنجیره های تاثیرگزاری خاص یا دیگر مشخصات نیست، بلکه با تغییرات سال به سال بهره وری در طول کار تخصصی دانشمند تشریح می شود (شکل س 12) (20 و 21).



شکل 1. الگوهای بهره وری طی دوره کاری یک دانشمند. توضیح شکل ها: الف) تاریخچه انتشارات کنت جی ویلسون (برنده جایزه نوبل فیزیک، 1982) محور افقی شمار سال های پس از اولین انتشار دانشمند را نشان می دهد و هر خط عمودی مربوط به یک مقاله تحقیقی است. ارتفاع هر خط نشانگر c_{10} است که شمار نقل قول ها از مقاله تا 10 سال پس از انتشار آن است (بخش های س 1. 3 و س 1. 6). مقاله با بیشترین تاثیر ویلسون، در سال 1974

منتشر شد که 9 سال بعد از اولین انتشار مقاله او می باشد و هفدمین مقاله از مجموع 48 مقاله اوست؛ بنابراین t^* $N^* = 17$ ، $N = 48$. ب) توزیع مقاله با بیشترین تاثیر $P(C^*10)$ برای تمام دانشمندان. در پایین 20٪ مساحت آبی شده است که مربوط به دانشمندان با تاثیر بیشینه کم است (C^*10 کمتر از 20)؛ ناحیه قرمز دانشمندان با تاثیر بیشینه بالا را نشان میدهد (5٪ بالا، C^*10 بیشتر از 200). ناحیه زرد مربوط به 75٪ باقیمانده با تاثیر بیشینه متوسط است (C^*10 بین 20 و 200). در صورت خارج کردن مقالات مروری از تحلیل، این الگوها تغییر نخواهد کرد (شکل های س 4 و س 36 را ببینید). ث) تعداد مقالات $N(t)$ منتشرشده تا زمان t برای سه دانشمند با تاثیر کم، متوسط و بالا که تعداد مقالات نهایی قابل مقایسه ای در کل دوره کاری خود دارند. د) توزیع توان بهره وری λ (18). بهره وری دانشمندان با ضریب تاثیر بالا از دانشمندان با ضریب تاثیر کم، رشد بیشتری دارد. ای) تحرک بهره وری که با شمار میانگین مقالات منتشرشده در هر سال برای دانشمندان با ضرایب تاثیر بالا، متوسط و پایین نشان داده شده است. $t = 0$ مربوط به سال اولین انتشار مقاله دانشمند می باشد.

این نتایج ما را وا داشت که موقعیت N^* مقاله با تاثیر بیشینه را در زنجیره انتشارات N یک دانشمند با سنجش $P(N^*/N)$ استخراج نماییم که مربوط به این احتمال است که کار با بیشترین نقل قول در اوایل (N^*/N کم) یا اواخر (N^*/N تقریباً برابر با 1) زنجیره مقالات منتشرشده دانشمند باشد. دریافتیم که $P(N^*/N)$ (شکل 2ای) یکنواخت است. همچنین به یک نتیجه عجیب دیگر نیز رسیدیم که نشان دهنده یافته های تجربیمان است: با وجود زمان انتشار یا ترتیب زنجیره انتشار مقالات، تاثیرگذاری به طور تصادفی درون کار یک دانشمند توزیع شده است. ما به این مورد قانون تاثیر تصادفی می گوئیم زیرا نشان می دهد که کار با بیشترین تاثیر می تواند با احتمال یکسان، هرجایی از زنجیره مقالات منتشرشده دانشمند باشد. دریافتیم که قانون تاثیر تصادفی در صورت طول دوره کاری مختلف، کار در دهه های مختلف و انتشار تنها یا گروهی مقالات یا تایید یا عدم تایید اعتبار دانشمند مسان همکاران، باز هم صادق است (بخش های س 1.4 و س 6.1) (22).

قانون تاثیر تصادفی می تواند تاثیرگذاری روبه رشد در طول دوره کاری دانشمند را تشریح نماید (شکل 2الف). برای مشاهده این مورد، ما بار دیگر تاثیر مقالات درون هر تخصص را با بدون تغییر نگه داشتن بهره وری شخصی ترکیب

می کنیم. تغییرات تاثیرگزاری دانشمندان تصادفی از داده اصلی برای اشخاص با تاثیر بالا و پایین، غیرقابل تمییز است (شکل 2الف). بنابراین، تاثیر میانگین رو به رشد استنادشده در شکل 2الف، نتیجه ترکیب بهره وری میانگین فزاینده () با ذات موثر توزیع نقل قول می باشد (6، 23-25). c_{10} پایدار نیست بلکه با شمار انتشارات زیاد می شود، که موجب افزایش تاثیرگزاری مشاهده شده خواهد شد (شکل 2الف). بنابراین، بهره وری رو به رشد، بیشتر از توانایی یا برتری می تواند باعث رشد تاثیرگزاری میانگین حین دوره تخصصی دانشمند شود.

در صورت تحلیل نمونه های مختلف دانشمندان در زمان بیش از 20 سال ثبت انتشارات، تعریف قانون بهره وری در زمان کار با بالاترین تاثیرگزاری علمی کارآمد خواهد بود. ما مجموعه های مختلفی از دانشمندان نظیر (1 گروه بندی آنان با طول کاری مختلف 2) گروه بندی آنان براساس دهه کار فعال (شکل های س 15 و س 16) 3 حذف مقالات با چند نویسنده (شکل س 17) 4 شامل نمودن مقالات منتشرشده در یک زیربخش فیزیک به تنهایی (شکل س 38) 5 عدم ایجاد فیلتر و شامل نمودن تمام دانشمندان (شکل های س 18 و س 19) 6 استفاده از تعریف های مختلف تاثیرگزاری (شکل های س 20، س 21 و س 22) یا 7 در نظر گرفتن شش قاعده مختلف در مجموعه داده حالت 2 (شکل های س 22 و س 23) را در نظر گرفتیم. در تمام این موارد، موقعیت بیشینه احتمال وقوع کار با بیشترین تاثیرگزاری تغییر می کند اما در کارهای تخصصی تصادفی شده هیچ اختلافی نمی بینیم. بنابراین، شکل خاص $P(t^*)$ فقط تابعی از انتخاب دانشمندان و الگوهای بهره وری زمانی آنان خواهد بود، در حالی که تاثیرگزاری همواره به طور تصادفی درون زنجیره انتشارات یک دانشمند توزیع شده است.

قانون مستند تاثیر تصادفی یک سوال را در بردارد: قانون توانایی شخصی یک محقق که تخصص عالی دارد، چگونه است؟ ما دو مدل کمی برای پاسخ به این سوال پیشنهاد خواهیم کرد.

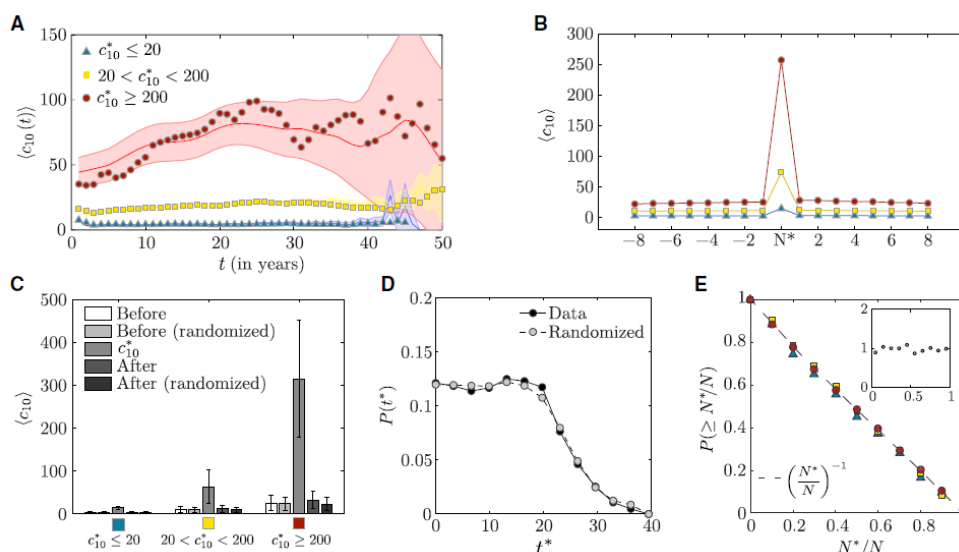
مدل تاثیر تصادفی (R-model)

می توانیم از قانون تاثیر تصادفی برای ساختن یک مدل خنثی از کارهای تخصصی دانشمندان بهره ببریم: فرض می کنیم که هر دانشمند زنجیره مقالاتی منتشر می کند که تاثیرشان به طور تصادفی از توزیع تاثیر یکسان ($P(c_{10})$) انتخاب شده است. در نتیجه، تنها تفاوت بین دو دانشمند بهره وری کلی N آنهاست. با داشتن توزیع $P(c_{10})$ و

$P(N)$ (شکل 3، الف و ب) به عنوان ورودی، مدل R به دست آمده (بخش س 4. 2) به طور دقیق تصادفی بودن زنجیره تاثیرگزاری $P(N^*/N)$ (شکل 2ای) را بازتولید خواهد کرد، اما دو تخمین را نیز ایجاد خواهد کرد که در تضاد با داده ها است.

الف) بهره وری به تنهایی باعث موفقیت می شود: اگر تاثیر هر مقاله به طور تصادفی از $P(c_{10})$ یکسان انتخاب شود، یک دانشمند بهره ور (با N بالا) تمایل بیشتری به داشتن امتیاز c^*_{10} بالا خواهد داشت (معادلات س 7 و س 8 را ببینید) (26 و 27). اما مدل R به طور صحیح رشد مشاهده شده به عنوان تابعی از $\langle c^*_{10} \rangle$ را بازتولید نمی کند (شکل 3ث).

ب) تاثیر واگرا: هرچه تاثیر میانگین انتشارات یک دانشمند بدون در نظر گرفتن انتشار با بیشترین نقل قول $\langle c^*_{10} \rangle$ (شکل 1الف) بالاتر باشد، تاثیر مقاله با بیشترین نقل c^*_{10} بالاتر خواهد بود (شکل 3د). بنابراین، مقالات با تاثیر بالای حقیقی توسط دانشمندانی منتشر می شوند که ثبت پایداری از تاثیر بالا داشته اند. مدل R نمی تواند این رفتار را پیش بینی کند، بلکه تخمین می زند که وقتی $\log(c^*_{10}) \rightarrow 1.97$ (شکل 3د) $\langle c^*_{10} \rangle$ واگرا خواهد شد که این رفتار ناشی از ذات طبیعی لگاریتم است (بخش س 4. 1 و شکل س 27).



شکل 1. الگوهای تاثیرگزاری طی تخصص کاری یک دانشمند. توضیح: الف) تحرک تاثیر که با تاثیر میانگین سالیانه مقالات $\langle c_{10}(t) \rangle$ برای دانشمندان با تاثیر بیشینه بالا، متوسط و کم نشان داده شده است، که $t = 0$ مربوط به

سال اولین انتشار دانشمند می باشد. نمادها مربوط به داده است، درحالی که ناحیه سایه دار به حد مطمئن 95٪ کارهای تخصصی اشاره دارد که تاثیر انتشارات به طور تصادفی درون هر تخصص کاری مخلوط شده است. (ب) تاثیر میانگین $\langle C_{10} \rangle$ مقالات منتشرشده قبل و بعد از مقاله با بیشترین تاثیر C^*_{10} برای دانشمندان با تاثیرگزاری بالا، متوسط و پایین. این نمودار خاطرنشان می کند که هیچ تغییرات قابل تشخیصی در تاثیرگزاری قبل و بعد از کار با بیشترین تاثیر دانشمند وجود ندارد. (ث) $\langle C^*_{10} \rangle$ و $\langle C_{10} \rangle$ قبل و بعد از مقاله با بیشترین تاثیر دانشمند. در هر گروه، ما تاثیر میانگین مقاله با بیشترین تاثیر $\langle C^*_{10} \rangle$ را محاسبه نموده و تاثیر میانگین تمام مقالات قبل و بعد از آن را نیز محاسبه می کنیم. همچنین معیارهای مشابه به دست آمده در زنجیره های انتشاراتی را نیز گزارش می کنیم که C^*_{10} ثابت است در حالی که تاثیر تمام مقالات دیگر به صورت تصادفی مخلوط می شود. (د) توزیع زمان انتشار t^* مقاله با بیشترین تاثیر برای کارهای دانشمندان (دایره های مشکی) و برای کارهای با تاثیر تصادفی (دایره های خاکستری). عدم وجود اختلاف بین دو منحنی ($P = 0.70$ برای تست U من-ویتنی بین دو توزیع) قانون تاثیر تصادفی را تایید می کند؛ بدان معنا که تاثیر دورن زنجیره انتشارات دانشمند تصادفی است. توجه نمایید که افت پس از 20 سال مشخصا به دلیل تمرکز ما روی کارهایی است که حداقل 20 سال پوشش دارند (شکل س 22 را ببینند). (ای) توزیع انباشته $P(\geq N^*/N)$ برای دانشمندان با $N \approx 50$ ، که N^*/N به مرتبه N مقاله با بیشترین تاثیر در طول کار دانشمند اشاره دارد که بین $1/N$ و 1 متغیر است. توزیع انباشته N^*/N یک خط مستقیم با شیب 1 است، که نشان می دهد N در هرجای زنجیره مقالات منتشره توسط یک دانشمند، احتمال وقوع یکسان دارد. همواری $P(N^*/N)$ (تمام دانشمندان، ضمیمه) این نتیجه را تایید می کند که احتمال زمانی مقاله با بیشترین تاثیر یکسان است. اختلافات کوچک بین سه نمودار ناشی از تعداد متفاوت مقالات N در سه گروه دانشمندان است [شکل س 24 را برای نمودار $P(\geq N^*/N)$ برای مقادیر دیگر N ببینید و شکل س 25 و س 26 را برای خودهمبستگی درون دوره کاری یک دانشمند ببینید].

نقص های الف و ب ما را برانگیخت تا این فرضیه را بگوییم که مقالات تحقیقی همگی از توزیع تاثیر مشابهی برداشته شده اند و بنابراین محققین هیچ تاثیر شخصی قابل تشخیصی ندارند و این مورد ما را مجبور کرد که رابطه بین بهره وری، تاثیر و شانس را به طور دقیقتر بررسی نماییم.

مدل Q

متاسفانه در مدل R، دانشمندان با بهره وری یکسان تاثیر غیر قابل تشخیصی دارند. در واقعیت، تاثیر به طور قابل ملاحظه بین دانشمندان فرق دارد (شکل 3ای)، که وجود یک پارامتر پنهان Q_i برای تعادل تاثیرگزاری را خاطرنشان می کند که برای هر دانشمند i مقدار واحدی دارد.

ذات طبیعی لگاریتم $P(c_{10})$ (شکل 3الف) (24) وجود فرایندهای افزایشده را خاطرنشان می کند، که ما را برمی انگیزد تا تاثیر $c_{10,ia}$ مقاله α منتشره توسط دانشمند i را به این صورت بنویسیم،

$$c_{10,ia} = Q_i p_{\alpha} \quad (1)$$

که p_{α} تاثیر پتانسیلی مقاله α در زنجیره مقالات منتشره توسط دانشمند i است.

پارامتر Q_i ، قابلیت دانشمند i برای رشد دانش موجود را بیان می دارد که در صورت پیشرفت ($Q_i > 1$) یا تنزل ($Q_i < 1$) خواهد بود. ما مقدار این پارامتر Q_i را طی دوره تخصصی یک دانشمند ثابت می گیریم که این تئوری بعداً اعتبارسنجی خواهد شد (شکل 5 و بخش س 4.9). مدل به دست آمده فرض می کند که هر دانشمند به طور تصادفی یک پروژه با پتانسیل p_{α} را انتخاب می کند و با فاکتور Q_i که برای هر دانشمند واحد است آن را بهبود می بخشد و منتج به مقاله ای با تاثیر معادله 1 خواهد شد. بنابراین، انتشارات با تاثیر بالای حقیقی، نتیجه دانشمند با Q_i بالا است که برحسب شانس یک پروژه با p_{α} بالا را انتخاب کرده است؛ هر دانشمند صرف نظر از پارامتر Q_i خودش، می تواند مقالات با تاثیر پایین را با انتخاب p_{α} پایین منتشر کند.

فرایند تصادفی پشت مدل معادله 1، با احتمال مشترک $P(p, Q, N)$ ، با روابط نامعین بین p ، Q و N تعیین می شود. ذات طبیعی لگاریتم (شکل 3الف) به ما این امکان را می دهد که را سنجش کنیم، و بفهمیم که با یک تابع لگاریتمی معمولی تطابق دارد (شکل 3ف). با فرض اینکه نیز لگاریتمی است (بعداً تایید خواهد شد)، به دست می

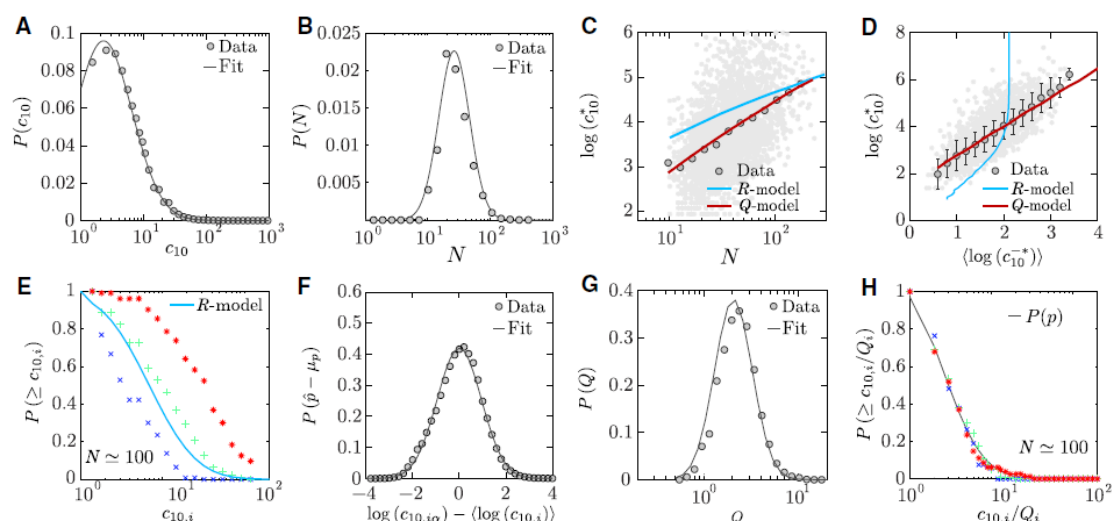
آوریم $\hat{p} = \log p$ ، $\hat{Q} = \log Q$ ، $\hat{N} = \log N$ ، که توزیع سه متغیره $P(\hat{p}, \hat{Q}, \hat{N}) = \mathcal{N}(\mu, \Sigma)$ به دست خواهد آمد. با استفاده از یک رویکرد احتمالی بیشینه (بخش س 4.4 را ببینید)، از داده ها میانگین $\mu \equiv (\mu_p, \mu_Q, \mu_N) = (0.92, 0.93, 3.34)$ و ماتریس کوواریانس زیر را تخمین می زنیم:

$$\Sigma \equiv \begin{pmatrix} \sigma_p^2 & \sigma_{p,Q} & \sigma_{p,N} \\ \sigma_{p,Q} & \sigma_Q^2 & \sigma_{Q,N} \\ \sigma_{p,N} & \sigma_{Q,N} & \sigma_N^2 \end{pmatrix} = \begin{pmatrix} 0.93 & 0.00 & 0.00 \\ 0.00 & 0.21 & 0.09 \\ 0.00 & 0.09 & 0.33 \end{pmatrix} \quad (2)$$

ماتریس معادله 2 دو تخمین کلیدی را در بر دارد: 1) نشان می دهد که تاثیر پتانسیلی مقاله $\rho\alpha$ مستقل از بهره وری دانشمند N_i و پارامتر پنهان او Q_i است. بنابراین، دانشمندان پتانسیل تاثیر هر مقاله را به طور تصادفی از توزیع $P(p)$ انتخاب می کنند که برای تمام اشخاص یکسان است و مستقل از Q و N خواهد بود.

2) غیرصفر نشان می دهد که پارامتر پنهان Q و بهره وری N به همدیگر وابسته هستند (بخش س 4.4)، اما مقدار کوچک آن نشان می دهد که Q بالا خیلی کم با بهره وری بالا مرتبط است.

عدم وجود رابطه بین p و (Q, N) به ما این امکان را می دهد که به صورت تجربی وابستگی مقاله با بالاترین تاثیر $\langle \log c_{10}^* \rangle$ به بهره وری N (بخش س 4.10) و تاثیر میانگین مقالات منتشره دیگر توسط یک دانشمند خاص $\langle c_{10}^* \rangle$ (بخش س 4.10 را ببینید) را محاسبه نماییم. تخمین این مدل برای $\langle \log c_{10}^* \rangle(N)$ و $\langle \log c_{10}^* \rangle(\langle c_{10}^* \rangle)$ توافقی عالی با داده ها دارد (شکل 3 ث و د و شکل س 30)، که نشان می دهد پارامتر پنهان Q و تغییرات مربوط به بهره وری N می تواند اختلافات تاثیر تجربی مشاهده شده بین دانشمندان را توصیف نماید و نقوص مدل R را تصحیح کند.



شکل 3. مدل Q. توضیح اشکال: الف) توزیع تاثیر مقالات c10 در راستای تمام مقالات موجود در مجموعه داده. خط خاکستری مربوط به تابع لگاریتمی با میانگین $\mu = 1.93$ و $\sigma^2 = 1.93$ ($R^2 = 0.98$) می باشد. ب) توزیع تعداد کلی مقالات منتشره توسط یک دانشمند (بهره وری). خط خاکستری یک لگاریتم معمولی با $\mu = 3.6$ و $\sigma^2 = 0.57$ می باشد [تست وزنی کولموگورف-اسمیرنوف، $P = 0.70$]. ث) نقل قول های مقاله با بیشترین تاثیر c10، در برابر تعداد انتشارات N طی دوره کاری یک دانشمند. هر نقطه خاکستری نمودار پراکندگی مربوط به یک دانشمند است. دایره ها منبع داده پراکندگی است. با فرض اینکه تاثیر هر مقاله به طور تصادفی از توزیع $P(c10)$ شکل 2 الف استخراج شده است، نمودار آبی تخمین مدل R را در بر دارد. منحنی قرمز مربوط به تخمین تجربی (معادله س 35 را ببینید) مدل Q است؛ ($R^2 = 0.98$ ؛ بخش س 4.6 و شکل س 29 را برای تناسب ببینید). د) $\log c10^*$ برابر $\langle \log c10^* \rangle$. هر نقطه خاکستری در نمودار پراکندگی مربوط به یک دانشمند است، که $\langle \log c10^* \rangle$ لگاریتم متوسط تاثیر مقاله او بدون در نظر گرفتن مقاله با بیشترین تاثیر c10* است. نمودار آبی مربوط به تخمین مدل R و قرمز تخمین تجربی (معادله س 36 را ببینید) مدل Q است ($R^2 = 0.99$ ؛ بخش س 4.6 و شکل س 29 را برای تناسب ببینید). ای) توزیع تاثیر انباشته تمام مقالات منتشر شده توسط سه دانشمند با بهره وری یکسان $N \approx 100$ و Q متفاوت. ف) توزیع $P(\hat{p})$ برای تمام انتشارات. برای هر مقاله α دانشمند i ، داریم $p_{\alpha} = \log c10_{i,\alpha} - \log Q_i$ که $\log Q_i = \langle c10_i \rangle - \langle \hat{p} \rangle$. بنابراین توزیع $\hat{p} = \log p$ ، بجز برای فاکتور تبدیل μ_p ، مربوط به توزیع

$\log c_{10,ia} - \langle \log c_{10,i} \rangle$ نیز می باشد که لگاریتم معمولی برای $\mu=0$ و $\sigma^2=0.95$ است (تست KS ، $p = 0.48$).
 ج) توزیع پارامتر Q ، $P(Q)$ ، برای تمام دانشمندان. خط خاکستری مربوط به تابع لگاریتم معمولی با $\mu=0.93$ و $\sigma^2=0.46$ می باشد (تست KS وزنی، $p = 0.48$). ه) توزیع انباشته تاثیر با مقیاس جدید $c_{10,ia}/Q_i$ برای سه دانشمند در شکل ای. خط مشکی مربوط به توزیع کلی $P(p)$ می باشد. سقوط آن با معادله 1 پیش بینی شده است.

به طور خلاصه، مدل Q با اختصاص یک پارامتر Q شخصی و یک بهره وری N به هر دانشمند که از توزیع $P(Q, N)$ استخراج خواهد شد، به ما این امکان را خواهد داد که زنجیره ای ترکیبی از انتشارات ایجاد نماییم. هر مقاله در زنجیره تاثیری مشخص به صورت خواهد داشت، که p به صورت تصادفی از توزیع $Q \times p$ خارج می شود و برای تمام دانشمندان یکسان است.

سنجش و دقت پارامتر پنهان Q

این مدل به ما این امکان را می دهد که پارامتر Q_i را از زنجیره انتشارات $c_{10,ia}$ هر دانشمند محاسبه نماییم (بخش 4.5)، که برای N_i بزرگ معادله زیر را داریم (معادله 28 را برای N_i محدود، و شکل 28 را برای رابطه بین دو تخمین پارامتر Q ببینید)،

$$Q_i = e^{\langle \log c_{10,i} \rangle - \mu p} \quad (3)$$

با فرض وابستگی Q به $c_{10,ia}$ ، Q با یک مقاله با تاثیر کم یا زیاد تغییر زیادی نخواهد، اما درعوض قابلیت دانشمند برای تغییر سیستماتیک پروژه خود به انتشار با تاثیر بالا (یا تاثیر پایین) را نشان می دهد. برای مثال اگرچه سه دانشمند شکل 3ای بهره وری یکسان 100 دارند، معادله 3 مقادیر Q بسیار متفاوتی را برای آنان پیش بینی می کند که برابر با 9/99، 3/31 و 1/49 است. این مقادیر به طور دقیق اختلافات قابل ملاحظه در زنجیره انتشارات نان را نشان می دهد: محقق با $Q = 9.99$ به صورت پایدار مقالاتی با تاثیر بالا را منتشر می کند در حالی که انتشارات

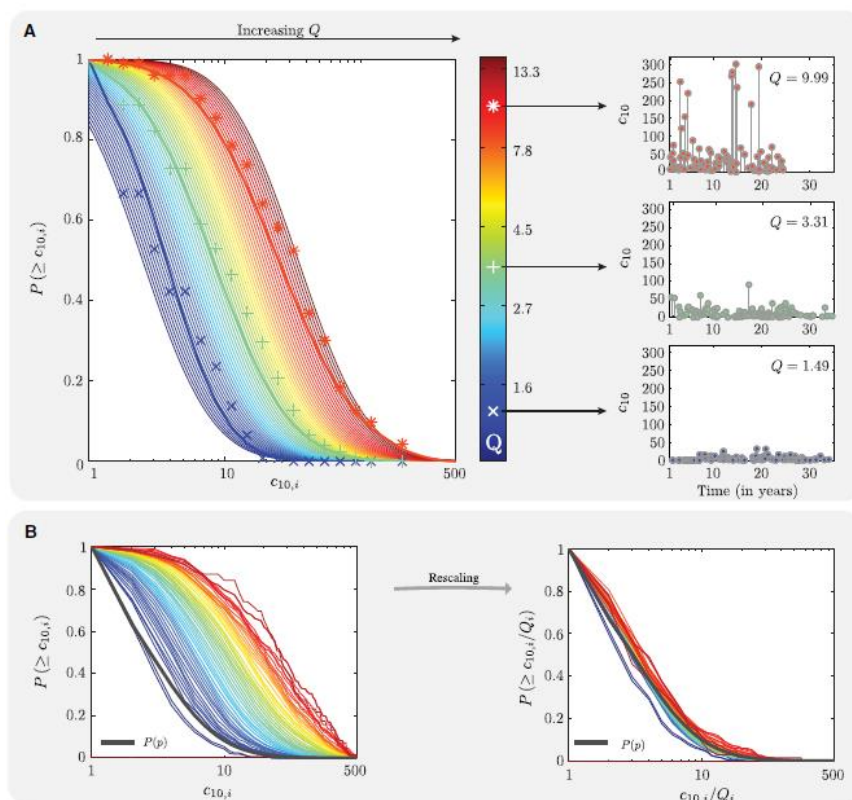
محقق $Q = 1.49$ به صورت پایدار با تاثیر محدود است. بنابراین، پارامتر Q قابلیت متفاوت دانشمندان برای برای گرفتن پروژه های تصادفی و تغییر سیستماتیک آنها به انتشارات با تاثیر بالا یا پایین را نشان می دهد.

مدل Q پیش بینی های عجیبی می کند که اگر از متغیر کاهشی $p_a = c_{10,ia}/\hat{Q}_i$ استفاده نماییم (28.29)، صرف نظر از اختلافات آشکار در مسیرهای کاری شخصی، اختلافات تاثیرات مقالات شخصی نیز باید ظاهر شود. اگرچه توزیعات $P(c_{10,ia})$ شخصی خیلی فرق دارد، توزیعات $P(c_{10,ia}/Q_i)$ برای تمام دانشمندان به یک منحنی کلی واحد $P(p)$ افت می کند (شکل 4ب)، که ذات فراگیر تاثیر تمام کارهای تخصصی را نشان می دهد (30). درانتها، لگاریتم معمولی $P(Q)$ (شکل 3ج) پایداری ذاتی ریاضی مدل را تایید می کند.

یک محدودیت اساسی تمام معیارهای استفاده شده در دانش، عدم سکون آنهاست: بهره وری، شمار انباشته نقل قول ها و مقیاس- h ، همگی با زمان رشد خواهند کرد که باعث خواهد شد تا مقایسه اشخاص در مراحل مختلف دوره تخصصیشان سخت شود. درمقابل، دریافتیم که پارامتر Q مستقل از مرحله کاری می باشد. برای اثبات این موضوع، یک پنجره مقاله با $\Delta N = 30$ را برای سنجش تغییرات Q در طول دوره کاری یک دانشمند استفاده نمودیم، که مشاهده شده شد پارامتر Q تغییرات کمی در دوره تخصصی دارد (شکل 5الف). مقدار این تغییرات با ذات تصادفی Q (بخش س 4.9)، فقط برای 75٪ دانشمندان توضیح داده شده است، زیرا پارامتر تخمینی Q درون محدوده نامعینی که توسط مدل فراهم شده است واقع خواهد شد. در 25٪ باقیمانده موارد، تغییرات کمی بیشتر از تغییرات پیش بینی شده با ذات تصادفی مدل است (شکل 5ب). اما مقدار این تغییرات اضافه هرگز از 15٪ بیشتر نخواهد شد و خطای نسبی میانگین همواره کمتر از 10٪ است (بخش س 4.9 و اشکال س 31 و س 32).

در انتها، برای آزمودن پایداری پارامتر Q در کل تخصص کاری و بدون تابعیت از بهره وری، دانشمندان با حداقل 50 مقاله را در نظر گرفته و پارامتر Q را برای اوایل و اواخر کار آنان (بترتیب Qearly و Qlate)، با استفاده از معادله 3 بترتیب در نیمه اول و دوم مقالاتشان محاسبه می کنیم (95.1٪، شکل 5ث). در این حالت، عدم شناسایی تصادفی اختلافات بین Qearly و Qlate برای اکثر دانشمندان را تشریح خواهد کرد. همچنین این سنجش ها

نشان می دهد که پارامتر Q به طور کلی در دوره تخصصی پایدار است، که امکان ارائه تخمین های کمی برای ارزیابی کار علمی را میسر خواهد کرد.



شکل 4. متخصصان و پارامتر Q مربوطه. الف) چپ) توزیع تاثیر انباشته تخمینی تجربی برای Q مختلف. این نمودار توزیع تاثیر سه دانشمند نشان داده شده در شکل 2 ای را نیز خاطرنشان می کند. ثبت جزییات انتشارات هر دانشمند در سمت راست گزارش شده است. ب) چپ: توزیع تاثیر انباشته شخصی $P(c_{10,i})$. با وجود شمار کمی از مقالات برای مشخصه بندی دانشمندان و غیرممکن بودن محاسبه آماری توزیعات بامعنا برای اغلب آنان، هر توزیع برای تمام مقالات تمام دانشمندان با مشابه محاسبه شده است. نمودار قرمز Q_i آنها را همانند شکل الف نشان می دهد. راست: توزیعات انباشته تاثیر با مقیاس جدید $c_{10,i}/Q_i$ برای دانشمندان، که نشان می دهد توزیعات اشخاص به توزیع کلی $P(p)$ افت می کند.

توان تخمینی پارامتر پنهان Q

مقدار حقیقی پارامتر Q به توان تخمینی آن بستگی دارد:

(1) پارامتر Q به ما این امکان را می دهد که تعداد مقالات مورد نیاز یک دانشمند برای رسیدن به مقاله با بیشترین تاثیر $c*10$ را تخمین بزنیم (شکل 6ب). دریافتیم که دانشمندان با Q پایین (تقریبا 1.2) باید حداقل 100 مقاله را بنویسند تا یک از آنها به طور میانگین 30 نقل قول بخورد. اما دانشمندی با بهره وری مشابه و $Q = 10$ انتظار مقاله با $c10 * = 250$ را خواهد داشت. دوبرابر شدن بهره وری فقط باعث افزایش هفت نقل قول برای مقاله با بیشترین تاثیر دانشمند با Q کم خواهد شد، درحالی که برای دانشمند با Q بالا $c*10$ افزایش 50 نقل قول را خواهد داشت. به طور کلی، شکل 6ب استناد می کند که برای دانشمندان با Q کم، افزایش بهره وری نمی تواند پایداری شانس انتشار یک کار با تاثیر بالا را تضمین کند؛ بنابراین شانس موفقیت آنان پایین خواهد بود.

(2) مقیاس h یک دانشمند، که خاطرنشان می کند مقالات با نقل قول h بیشتر آنان حداقل نقل قول h را دریافت خواهند کرد (12 و 15)، با پارامتر Q و بهره وری N مرتبط است (بخش س 4.11). این تخمین تجربی فقط مقیاس h مشاهده شده تمام دانشمندان را بازتولید نمی کند (شکل س 33ب)، بلکه سیر تکامل مقیاس h در دوره تخصصی دانشمندان را نیز نشان می دهد (شکل 6 ث و د و شکل س 34الف). معادلات مشابهی که تعداد انباشته نقل قول ها (شکل 6د و اشکال س 33د و ف و شکل س 34ب) و مقیاس g (بخش س 4.11) را دربردارد، نشان می دهد که معیارهای سنتی بازدهی به صورت واحد توسط Q تعیین خواهند شد. با دانستن اینکه Q برحسب زمان ثابت است، نتیجه می گیریم که بهره وری فقط می تواند معیاری برای تغییرات گستره تخصصی این سنجش ها باشد (شکل 6 ث و د).

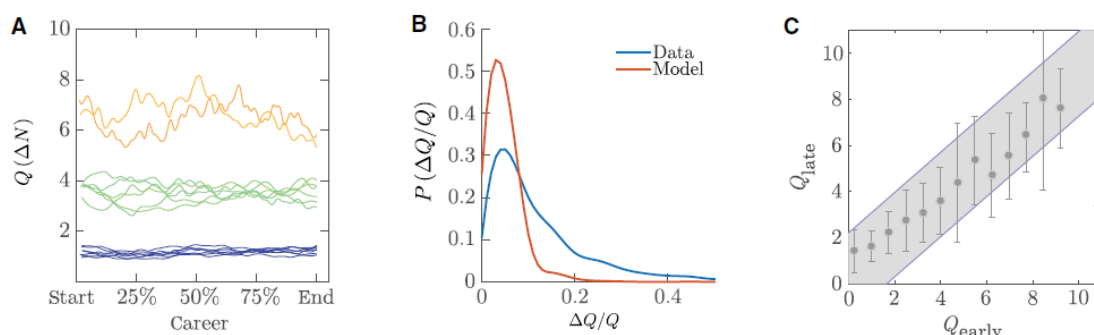
(3) با تعیین مقدار Q در طول مراحل اولیه تخصص کاری دانشمند، می توانیم از آن برای تخمین تاثیر آینده استفاده نماییم. تخمین خطای ΔQ برای Q ، برحسب تعداد مقالات منتشر شده N کاهش می یابد و پس از $N = 20$ کمتر از 10٪ خواهد شد (بخش س 4.12). بنابراین می توانیم Q را براساس N_0 اولین مقالات منتشره دانشمند در معادله 3 تخمین بزنیم و سپس از معادله تجربی مقیاس h و تعداد کلی نقل قول ها برای تخمین تاثیر آینده دانشمند استفاده نماییم (بخش س 4.12 و شکل س 35). با دانستن ذات تصادفی مدل، محدوده نامعین مقدار سنجش تاثیر مناسبتری را نشان خواهد داد. در شکل 6ای، پس از اینکه Q را از مقالات $N_0 = 20$ (بالا) و $N_0 = 50$

(پایین) تخمین زدیم، برای دو دانشمند تخمین مقیاس- h تا $N = 150$ را نشان داده ایم. اگرچه مقیاس- h آغازی برای دو دانشمند همپوشانی دارد، تاثیر بلندمدت آنان واگراست که این اختلاف به طور دقیق توسط مدل Q پیش بینی شده است. با تعمیم به نمونه بزرگتری از دانشمندان، رابطه ای بین مقیاس- h تخمینی و تجربی خواهیم یافت (شکل 6ف). برای کمی سازی دقت تخمین کلی مدل Q ، کسرهای زمانی را که درون محدوده مورد نظر، مقیاس- h افت می کند را برای دانشمندان با حداقل 100 مقاله سنجش کردیم. امتیاز Z_N برای هر دانشمند، SD های انحراف مقیاس- h واقعی از مقیاس- h با بیشترین احتمال پس از N مقاله را نشان می دهد. دریافتیم که 71٪ دانشمندان $2 \leq z_{40} \leq 50$ برپایه $N_0 = 50$ دارند، که برای $N_0 = 50$ و z_{70} به 81٪ بهبود می یابد (شکل 6ج). همچنین به این نتیجه رسیدیم که تخمین پارامتر Q در مراحل اولیه قابلیت پیش بینی تاثیر بلندمدت تخصصی را خواهد داشت.

4) برای آزمایش اینکه آیا Q_i با تاثیر رابطه دارد، دانشمندان را براساس Q , N , C_{tot} , c_{10} و مقیاس- h آنان دسته بندی کردیم. برای اعتبارسنجی این دسته بندی ها، از یک نمودار مشخصه اجرایی دریافتی (ROC) استفاده کردیم که کسر جوایز نوبل در بالای لیست دسته بندی شده را سنجش می کند (شکل 6الف). دریافتیم که دسته بندی برپایه Q ، کارهای برنده نوبل را خیلی دقیق تخمین می زند (بخش س 7 و شکل س 45). همچنین باید خاطرنشان کرد که توان های تخمین C_{tot} , c_{10}^* و مقیاس- h از یکدیگر غیرقابل تشخیص است و بهره وری خیلی کم قابل تخمین خواهد بود. با استفاده از Q_i برای بررسی برندگان دیراک-بولتزمن نیز نتایج مشابهی به دست آمد (اشکال س 46 و 47). Q اولیه کاری نیز دقت خوبی برای تخمین جوایز نوبل دارد (بخش س 7.1 و شکل س 48).

اکتشافات با تاثیر بالا اغلب ناشی از کار تیمی است (31-33). برای کشف تاثیر کار تیمی (34 و 35) از یک الگوریتم اعتبارسنجی (22 و 36) برای تعیین تاثیر نقش هر نویسنده استفاده کردیم. سپس تحلیل کلیمان را تکرار کردیم و دریافتیم که مدل Q با پارامترهای کمی بیشتر، می تواند نتایج را توضیح دهد (بخش س 6.1 و اشکال س 40 تا س 43). علاوه براین، دریافتیم که Q_i با حذف برخی همکاران قابلیت بهبود خواهد داشت (بخش س 6.2 و شکل

س 44). بنابراین اگرچه اثرات همکاری باعث تعادل موفقیت انتشارات خاصی خواهد شد، افراد بتنهایی تاثیر محدودی بر Q_i خواهند داشت.



شکل 5. پایداری پارامتر Q (الف) تغییر زمانی پارامتر Q در طول دوره تخصصی فرد. برای دانشمندان با حداقل 100 مقاله و Q برابر با 1.2، 3.8 و 6.5، $Q(\Delta N)$ را گزارش کرده ایم که در پنجره متحرکی با $\Delta N = 30$ مقالات سنجش شده است. برای 75٪ دانشمندان، تغییرات به این دلیل است که ما تعداد محدودی مقاله در پنجره متحرک داشتیم، و مقدار تغییرات با تخمین مدل Q ثابت قابل مقایسه است (بخش س 4.9). (ب) تغییرات پارامتر Q در مدل و داده ها. توزیع نامعین $Q(\Delta N) / \Delta N$ ، را در داده ها و کارهای ترکیبی با Q ثابت ($\Delta N = 5$) مطالعه کردیم. برای 74 درصد دانشمندان، تغییرات با مدل قابل مقایسه است. برای 26 درصد باقیمانده SD کمی بیشتر از تخمین مدل است. (ث) مقایسه بین پارامتر Q اولیه و پایانی. پارامتر Q را در مرحله اولیه و پایانی 823 دانشمند با حداقل 50 مقاله مقایسه کرده ایم. دو مقدار پارامترها را بترتیب با استفاده از نیمه اول و دوم مقالات منتشرشده سنجش کردیم. این سنجش ها را روی داده واقعی (دایره ها) و روی کارهای تخصصی تصادفی اجرا کردیم، که ترتیب مقالات ترکیب شده است (نواحی سایه دار خاکستری). بری غلب کارهای تخصصی، 95٪، تغییرات بین مراحل اولیه و پایانی نزدیک به تغییرات تخمینی مدل خنثی با ترتیب تصادفی مقالات است، که خاطرنشان می کند پارامتر Q در طول دوره کاری پایدار است. تغییرات مشاهده شده با تعداد محدودی مقالات در دوره کاری دانشمند تشریح شده است.

خلاصه و بحث

به طور خلاصه، گواهی تجربی را ارائه کرده ایم که تاثیرگذاری به طور تصادفی درون زنجیره انتشارات دانشمند توزیع شده است، و تغییرات موقت تاثیرگذاری در طول دوره کاری دانشمند می تواند با تغییرات بهره وری، شانس و پیشینه

توزیع تاثیر شخصی دانشمند توضیح داده شود. این یافته به ما این امکان را داد که قانون بهره وری، شانس و Q یک دانشمند را به طور سیستماتیک حل نماییم، که تخمین می زند مقالات با تاثیر بالای حقیقی به ترکیبی از Q و شانس نیازمندست و بهره وری بالا به تنهایی تاثیر محدودی روی احتمال انجام یک کار با تاثیر بالا در دوره کاری دانشمند خواهد داشت. پارامتر قابل سنجش نمایانگر قابلیت پایدار دانشمند برای انتشار مقالات با تاثیر بالا (یا پایین) است.

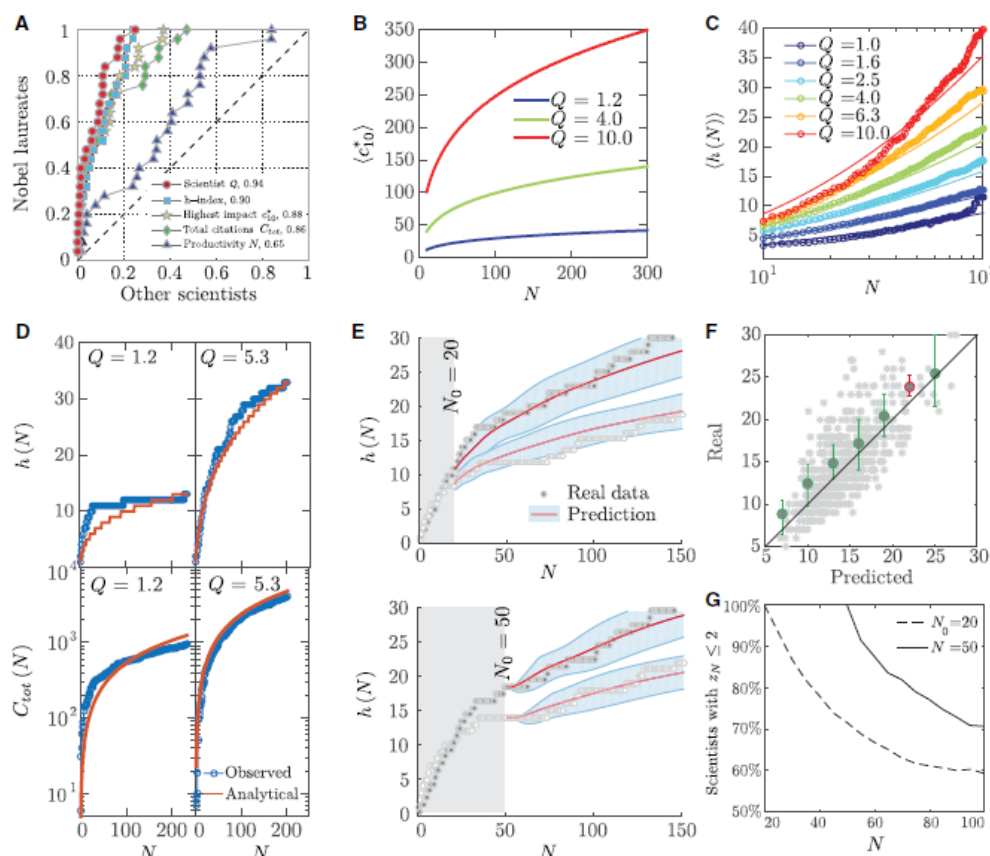
تمام این معیارهای استفاده شده بازدهی، در طول دوره کاری دانشمند تغییر می کند. در مقابل، Q در طول دوره کاری یک دانشمند برای اغلب دانشمندان (76٪) ثابت است، و توسط یک مقاله یا همکار خاص تغییر چندانی نخواهد کرد. ذات اساسی پارامتر Q با این حقیقت پشتیبانی می شود که معیارهای کنونی استفاده شده برای موفقیت، از مقیاس h گرفته تا نقل قول های انباشته، می توانند از این معیار Q محاسبه شوند. Q نه تنها مقدار این معیارها را سنجش می کند بلکه سیر تکامل زمانی آنها را نیز پیش بینی خواهد کرد (شکل 6 ث و ف).

تمام یافته های اراده شده بالا، براساس زیرمجموعه ای از 2887 فیزیکدان با دوره کاری حداقل 20 سال و مقالات کافی است. دانشمندان به واسطه یا پایان کار تخصصی خود رسیده اند بنابراین می توانند در صورت انتخاب درست در دانشگاه ها، موفق در نظر گرفته شوند. البته یافته های ما برای دانشمندان جوان که دانشگاه را زود ترک کرده و مقالات اندکی منتشر نموده اند صادق نیست.

در این کار، ما با تاثیر بلندمدت که با $c*10$ نشان داده شده است، به عنوان یک متغیر برونزا رفتار کرده ایم. معقول به نظر می رسد که بهره وری و تاثیرگذاری برهمدیگر اثر بگذارند. برای مثال از منظر کارکردی، برخی انتشارات اولیه دانشمند می تواند به جذب منابعی که باعث رشد بهره وری بالا شود کمک نماید. $\langle c10 \rangle$ برای 10 مقاله اول دانشمند با تداوم دوره کاری او مرتبط است، یعنی احتمال ماندن در دانشگاه کمی متأثر از ضریب تاثیر مقالات اولیه است (شکل س 49). همچنین مدل Q خاطرنشان می کند که تعداد کلی مقالات در دوره تخصصی ارتباط کمی با Q بالا دارد (معادله 2). اگرچه مدل Q و تخمین های فراهم شده در اینجا، از ارتباط بین تاثیر اولیه و بهره وری کلی

(بخش س 5) سخنی نمی گوید، این یافته های اولیه، نیازمند سنجش و مدلهای بیشتری است که بتواند سیر تکامل موازی تاثیرگزاری و بهره وری کوتاه مدت را نشان دهد (37).

اگرچه Q می تواند به طور دقیق تاثیر یک کار را تخمین بزند، وابستگی Q به فاکتورهای بیشتر نظیر کیفیت آموزش و همکاری (38 و 39)، اندازه انجمن تحقیقی (24 و 40)، جنسیت (41 و 42)، دینامیک زیررشته ها (43 و 44) یا عادت های انتشار مقالات ناشناخته باقی مانده است. از لحاظ ریاضی، در صورت بازتاب فاکتورهای دیگر مشخص کننده یک دانشمند توسط پارامتر Q ، تفاوت زیادی ایجاد نخواهد شد. بررسی های اعتبار مختلفی که برای کشف فاکتور موثر محتمل نظیر طول دوره تخصصی، دهه، اثرات تیمی و تحلیل قواهد مختلف و مجموعه داده های متنوع انجام دادیم، از ارائه تشریحی ساده و مستقیم برای علت مقادیر Q مختلف دانشمندان ناتوان بوده است. علاوه براین، پارامتر توسط فاکتورهای مختلفی تحت تاثیر قرار می گیرد و اطلاعات بیشتر حول طبیعت آن باید مطالعه شود. اما تفاوت تعیین کننده کلیدی Q از شانس باید مجزا باشد. Q از یک مقاله تنها برحسب شانس تعیین نشده است، بلکه از بازدهی بالای پایدار در طول دوره تخصصی دانشمند تعیین شده است. این مورد در مولفه $\log c10 <$ در معادله 3 بازتاب دارد، که نشان می دهد یک مقاله واحد با ضریب تاثیر بالا تنها تاثیر کمی روی Q دارد. یک دانشمند نیازمند مقالات چندگانه با $c10$ بالا است تا Q بالایی داشته باشد. عدم پوشش دلیل پارامتر یک هدف یشرو خواهد بود، که نه تنها می تواند درک بهتری از ضرورت و سیر تکامل تعالی دانشمند ارائه نماید، بلکه قادر به بهبود توانایی آموزش و پرورش دانشمندان با تاثیر بالا نیز خواهد بود.



شکل 6. رابطه بین Q و معیارهای تاثیر دیگر. الف) نمودار نمایشگر رتبه بندی دانشمندان براساس Q , C_{tot} , h - N index, c_{10}^* و N . هر منحنی نسبت جوایز نوبل در برابر نسبت دانشمندان دیگر برای حد رتبه بندی مشخص را نشان می دهد. قطر (خط بدون تفکیک) مربوط به رتبه بندی تصادفی است؛ ناحیه زیر هر منحنی دقت ما برای رتبه بندی جوایز نوبل را نشان می دهد. دقت رتبه بندی در توضیحات شکل آمده است. درمورد دقت به عنوان تابعی از رتبه بندی در بخش س 7 بحث شده است. ب) نقل قول های موردانتظار برای مقاله با بیشترین تاثیر برای یک دانشمند. این نمودار شانس بسیار پایین محقق با Q کم را برای انتشار مقاله با تاثیر بالا نشان می دهد. ث) رشد مشاهده شده در برابر رشد تخمینی مقیاس- h برای دانشمندان با Q متنوع. این نمودار توافق بین مقیاس- h تخمینی تجربی (معادله س 38، خط پیوسته) و مقدار مشاهده شده $\langle h(N) \rangle$ را تایید می کند که با میانگین گیری مقیاس- h برای دانشمندان با Q یکسان (دایره ها) به دست آمده است. د) بالا: رشد مقیاس- h برای دو دانشمند با حداقل 200 مقاله و Q متفاوت به عنوان تابعی از بهره وری N (دایره های آبی) در مقایسه با تخمین معادله س 38

(خط قرمز). پایین: برای دو دانشمند در پتل بالا، تعداد انباشته نقل قول را به عنوان تابعی از N سنجش کرده ایم $C_{tot}(N)$ ، و با تخمین معادله س39 مقایسه نمودیم. توافق نزدیک بین مشاهده و تخمین در شکل 3 و د نشان می دهد که Q مستقل از زمان یک مشخصه ذاتی دانشمند را نشان می دهد، و اینکه معیارهای دیگری نظیر مقیاس- h یا نقل قول های انباشته، به طور واحد توسط Q و بهره وری تعیین شده اند. ای) برای دو دانشمند، تخمین مقیاس- h را به عنوان تابعی از N ، با استفاده اطلاعات دوره کاری اولیه نشان داده ایم، که داریم $N_0 = 20$ (top) و $N_0 = 50$ (bottom) برای تخمین پارامتر Q . اگرچه مقیاس- h آغازی تا $N_0 = 20$ همپوشانی زیادی برای دو دانشمند دارد، تاثیر بلندمدت آنان و اگر است که این اختلاف به طور دقیق توسط مدل Q تخمین زده شده است. ف) نمودارهای پراکنش مقیاس- h تخمینی و واقعی در $N = 60$ براساس Q تخمینی در $N_0 = 20$. میله های خطا چارک های تخمینی (25 و 75٪) در هر نمونه را نشان می دهد و در صورت وقوع $y = x$ بین دو چارک در آن نمونه سبز هستند و درغیراینصورت قرمز هستند. دایره ها مربوط به مقیاس- h میانگین در آن نمونه هستند. ج) امتیاز z_N برای هر دانشمند، تعداد SD های انحراف مقیاس- h واقعی از مقیاس- h محتمل بعد از N انتشار را نشان می دهد. $z_N \leq 2$ نشان می دهد که داده واقعی درون محدوده تخمین قرار گرفته است.

روش های جمع آوری داده

ما دو نوع مجموعه داده را استخراج کردیم: 1) ثبت انتشارات 236884 فیزیکدان منتشرکننده مقالات در ژورنال Physical Review از 1983 تا 2010 [مجموعه داده انجمن فیزیک آمریکا (APS)، بخش س1.1 و شکل س1 و س2 را ببینید] و 2) ترکیبی از 24630 داده پروفایل کاری گوگل اسکولار با Web of Science (WoS) که 514896 مقاله حول زیست شناسی، شیمی، علوم انسانی، زمین شناسی اقتصادی و روانشناسی را پوشش داده است (مجموع داده WoS، در بخش س1.2 و شکل س3 توضیح داده شده است). نتایج نشان داده شده در این مقاله به 2887 دانشمند اشاره دارد، که حداقل طی 20 سال انتشار داشته اند و حداقل 10 مقاله را منتشر نموده و حداقل 5 سال یکبار انتشار مقاله داشته اند که از مجموعه داده APS استخراج شده اند (بخش س1.3 را ببینید).

توجه نمایید که مجموعه داده فقط شامل نقل قولهای درون APS می باشد (بخش س 1.1 را ببینید)؛ به همین دلیل تعداد نقل قولهای این مجموعه به نسبت داده های WoS کمتر است.

یافته های ما با تحلیل نمونه های مختلف دانشمندان در مجموعه داده APS که با استفاده از چند مقیاس مختلف (بخش س 1.4 را ببینید) انتخاب شده و با تحلیل قواعد دیگر مجموعه داده WoS که در الحاقات و مراجع مقاله ذکر شده، نیز تایید شده است.

سنجش نقل قول ها

سنجش های تاثیرگزاری براساس نقل قول تحت تاثیر سه مساله مهم هستند: 1) نقل قول ها برای مقالات مختلف دینامیک متفاوتی دارند (6و45)، 2) تعداد متوسط نقل قول ها با زمان تغییر می کند، 3) تعداد نقل قول بستگی به زیررشته مربوطه دارد. برای کنارآمدن با مساله 1، ما برای هر مقاله شمار انباشته نقل قولهای دریافتی مقاله تا 10 سال پس از انتشار c10، را به عنوان معیاری از تاثیر علمی آن استفاده نمودیم (6و45). با نرمالسازی c10 توسط میانگین <c10> مقالات منتشرشده در یک سال مشابه می توانیم مسائل 2 و 3 را تصحیح نماییم.

مدل Q

فرایند تصادفی پشت مدل Q با احتمال مشترک $P(p, Q, N)$ تعیین شده است. این مدل فرض می کند که دانشمند i بهره‌وری N_i و یک پارامتر Q_i نمونه گیری از توزیع $P(Q, N)$ دارد، و سپس N_i را از p توزیعات حالت $P(p, Q, N)$ خارج می کند. با این فرض که $P(p, Q, N)$ از توزیع لگاریتمی سه متغیره با پارمترهای $\sum$ و μ پیروی می کند، می توانیم احتمال تابع L_i را بنویسیم که یک دانشمند i با Q_i و N_i زنجیره مقالات $\{\alpha\}$ با تاثیر $\{Q_i p\alpha\}$ دارد (معادله 1 را ببینید). درنهایت با بهینه سازی عددی، بیشینه تابع احتمال لگاریتمی کلی $\mathcal{L} = \sum_i \mathcal{L}_i$ را تشخیص می دهیم، که تخمین عددی $\sum$ و μ گزارشده در معادله 2 را فراهم می کند (بخش های س 4.3 و س 4.4 را نیز ببینید). این رویکرد Q_i را نیز تخمین می زند که با بیشینه کردن تابع احتمال L_i برای هر دانشمند به دست آمده است. این بیشینه سازی تشریحی تجربی برای Q_i فراهم می کند که با بهره وری N_i بالا، به معادله 3 همگرا می شود (بخش س 4.5 را ببینید).

REFERENCES AND NOTES

1. I. Fuyuno, D. Cyranoski, Cash for papers: Putting a premium on publication. *Nature* **441**, 792 (2006). doi: [10.1038/441792b](https://doi.org/10.1038/441792b); pmid: [16778850](https://pubmed.ncbi.nlm.nih.gov/16778850/)
2. J. A. Evans, J. Reimer, Open access and global participation in science. *Science* **323**, 1025 (2009). doi: [10.1126/science.1154562](https://doi.org/10.1126/science.1154562); pmid: [19229029](https://pubmed.ncbi.nlm.nih.gov/19229029/)
3. P. Azoulay, Research efficiency: Turn the scientific method on ourselves. *Nature* **484**, 31–32 (2012). doi: [10.1038/484031a](https://doi.org/10.1038/484031a); pmid: [22481340](https://pubmed.ncbi.nlm.nih.gov/22481340/)
4. B. Owens, Research assessments: Judgement day. *Nature* **502**, 288–290 (2013). doi: [10.1038/502288a](https://doi.org/10.1038/502288a); pmid: [24132272](https://pubmed.ncbi.nlm.nih.gov/24132272/)
5. O. Penner, R. K. Pan, A. M. Petersen, K. Kaski, S. Fortunato, On the predictability of future impact in science. *Sci. Rep.* **3**, 3052 (2013). doi: [10.1038/srep03052](https://doi.org/10.1038/srep03052); pmid: [24165898](https://pubmed.ncbi.nlm.nih.gov/24165898/)
6. D. Wang, C. Song, A.-L. Barabási, Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013). doi: [10.1126/science.1237825](https://doi.org/10.1126/science.1237825); pmid: [24092745](https://pubmed.ncbi.nlm.nih.gov/24092745/)
7. J. A. Evans, Future science. *Science* **342**, 44–45 (2013). doi: [10.1126/science.1245218](https://doi.org/10.1126/science.1245218); pmid: [24092715](https://pubmed.ncbi.nlm.nih.gov/24092715/)
8. S. Lehmann, A. D. Jackson, B. E. Lautrup, Measures of measures. *Nature* **444**, 1003–1004 (2006). doi: [10.1038/4441003a](https://doi.org/10.1038/4441003a); pmid: [17183295](https://pubmed.ncbi.nlm.nih.gov/17183295/)
9. S. Lehmann, A. D. Jackson, B. E. Lautrup, A quantitative analysis of indicators of scientific performance. *Scientometrics* **76**, 369–390 (2008). doi: [10.1007/s11192-007-1868-8](https://doi.org/10.1007/s11192-007-1868-8)
10. F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, Diffusion of scientific credits and the ranking of scientists. *Phys. Rev. E* **80**, 056103 (2009). doi: [10.1103/PhysRevE.80.056103](https://doi.org/10.1103/PhysRevE.80.056103); pmid: [20365039](https://pubmed.ncbi.nlm.nih.gov/20365039/)
11. D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, I. Rafols, Bibliometrics: The Leiden Manifesto for research metrics. *Nature* **520**, 429–431 (2015). doi: [10.1038/520429a](https://doi.org/10.1038/520429a); pmid: [25903611](https://pubmed.ncbi.nlm.nih.gov/25903611/)
12. J. E. Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16569–16572 (2005). doi: [10.1073/pnas.0507655102](https://doi.org/10.1073/pnas.0507655102); pmid: [16275915](https://pubmed.ncbi.nlm.nih.gov/16275915/)
13. S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, F. Herrera, h-Index: A review focused in its variants, computation and standardization for different scientific fields. *J. Informet.* **3**, 273–289 (2009). doi: [10.1016/j.joi.2009.04.001](https://doi.org/10.1016/j.joi.2009.04.001)
14. L. Bornmann, R. Mutz, S. E. Hug, H.-D. Daniel, A multilevel meta-analysis of studies reporting correlations between the h index and 37 different h index variants. *J. Informet.* **5**, 346–359 (2011). doi: [10.1016/j.joi.2011.01.006](https://doi.org/10.1016/j.joi.2011.01.006)
15. D. E. Acuna, S. Allesina, K. P. Kording, Future impact: Predicting scientific success. *Nature* **489**, 201–202 (2012). doi: [10.1038/489201a](https://doi.org/10.1038/489201a); pmid: [22972278](https://pubmed.ncbi.nlm.nih.gov/22972278/)
30. T. Vicsek, A question of scale. *Nature* **411**, 421 (2001). doi: [10.1038/35078161](https://doi.org/10.1038/35078161); pmid: [11373653](https://pubmed.ncbi.nlm.nih.gov/11373653/)
31. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007). doi: [10.1126/science.1136099](https://doi.org/10.1126/science.1136099); pmid: [17431139](https://pubmed.ncbi.nlm.nih.gov/17431139/)
32. B. F. Jones, S. Wuchty, B. Uzzi, Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008). doi: [10.1126/science.1158357](https://doi.org/10.1126/science.1158357); pmid: [18845711](https://pubmed.ncbi.nlm.nih.gov/18845711/)
33. J. A. Evans, J. G. Foster, Metaknowledge. *Science* **331**, 721–725 (2011). doi: [10.1126/science.1201765](https://doi.org/10.1126/science.1201765); pmid: [21311014](https://pubmed.ncbi.nlm.nih.gov/21311014/)
34. J. Moody, D. R. White, Structural cohesion and embeddedness: A hierarchical concept of social groups. *Am. Soc. Rev.* **68**, 103–127 (2003).
35. R. D. Malmgren, J. M. Ottino, L. A. N. Amaral, The role of mentorship in protégé performance. *Nature* **465**, 622–626 (2010). doi: [10.1038/nature09040](https://doi.org/10.1038/nature09040); pmid: [20520715](https://pubmed.ncbi.nlm.nih.gov/20520715/)
36. C.-T. Zhang, A proposal for calculating weighted citations based on author rank. *EMBO Rep.* **10**, 416–417 (2009). doi: [10.1038/embor.2009.74](https://doi.org/10.1038/embor.2009.74)
37. K. Börner, J. T. Maru, R. L. Goldstone, The simultaneous evolution of author and paper networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5266 (2004). doi: [10.1073/pnas.0307625100](https://doi.org/10.1073/pnas.0307625100); pmid: [14976254](https://pubmed.ncbi.nlm.nih.gov/14976254/)
38. P. Deville et al., Career on the move: Geography, stratification, and scientific impact. *Sci. Rep.* **4**, 4770 (2014). doi: [10.1038/srep04770](https://doi.org/10.1038/srep04770)
39. A. Clauset, S. Arbesman, D. B. Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Sci. Adv.* **1**, e1400005 (2015). doi: [10.1126/sciadv.1400005](https://doi.org/10.1126/sciadv.1400005); pmid: [26601125](https://pubmed.ncbi.nlm.nih.gov/26601125/)
- and hierarchy in faculty hiring networks. *Sci. Adv.* **1**, e1400005 (2015). doi: [10.1126/sciadv.1400005](https://doi.org/10.1126/sciadv.1400005); pmid: [26601125](https://pubmed.ncbi.nlm.nih.gov/26601125/)
40. J. Kaur, F. Radicchi, F. Menczer, Universality of scholarly impact metrics. *J. Informet.* **7**, 924–932 (2013). doi: [10.1016/j.joi.2013.09.002](https://doi.org/10.1016/j.joi.2013.09.002)
41. V. Larivière, C. Ni, Y. Gingras, B. Cronin, C. R. Sugimoto, Bibliometrics: Global gender disparities in science. *Nature* **504**, 211–213 (2013). doi: [10.1038/504211a](https://doi.org/10.1038/504211a); pmid: [24350369](https://pubmed.ncbi.nlm.nih.gov/24350369/)
42. S. F. Way, D. B. Larremore, A. Clauset, *Proceedings of the 25th International Conference on World Wide Web*, Geneva, Switzerland, 11 to 15 April 2016.
43. K. K. Mane, K. Börner, Mapping topics and topic bursts in PNAS. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5287–5290 (2004). doi: [10.1073/pnas.0307626100](https://doi.org/10.1073/pnas.0307626100); pmid: [14978278](https://pubmed.ncbi.nlm.nih.gov/14978278/)

16. B. F. Jones, B. A. Weinberg, Age dynamics in scientific creativity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18910–18914 (2011). doi: [10.1073/pnas.1102895108](https://doi.org/10.1073/pnas.1102895108); pmid: 22065777
17. P. Azoulay, J. S. Graff Zivin, G. Manso, Incentives and creativity: Evidence from the academic life sciences. *RAND J. Econ.* **42**, 527–554 (2011). doi: [10.1111/rand.2011.42.issue-3](https://doi.org/10.1111/rand.2011.42.issue-3)
18. A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli, Persistence and uncertainty in the academic career. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5213–5218 (2012). doi: [10.1073/pnas.1121429109](https://doi.org/10.1073/pnas.1121429109)
19. C.-K. Peng *et al.*, Long-range correlations in nucleotide sequences. *Nature* **356**, 168–170 (1992). doi: [10.1038/356168a0](https://doi.org/10.1038/356168a0); pmid: 1301010
20. D. K. Simonton, Career landmarks in science: Individual differences and interdisciplinary contrasts. *Dev. Psychol.* **27**, 119–130 (1991). doi: [10.1037/0012-1649.27.1.119](https://doi.org/10.1037/0012-1649.27.1.119)
21. D. K. Simonton, Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychol. Rev.* **104**, 66–89 (1997). doi: [10.1037/0033-295X.104.1.66](https://doi.org/10.1037/0033-295X.104.1.66)
22. H.-W. Shen, A.-L. Barabási, Collective credit allocation in science. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12325–12330 (2014). doi: [10.1073/pnas.1401992111](https://doi.org/10.1073/pnas.1401992111); pmid: 25114238
23. D. de Solla Price, *Little Science, Big Science... and Beyond* (Columbia University, 1963).
24. F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: Toward an objective measure of scientific impact. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17268–17272 (2008). doi: [10.1073/pnas.0806977105](https://doi.org/10.1073/pnas.0806977105); pmid: 18978030
25. M. J. Stringer, M. Sales-Pardo, L. A. Nunes Amaral, Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *J. Am. Soc. Inf. Sci. Technol.* **61**, 1377–1385 (2010). doi: [10.1002/asi.v61:7](https://doi.org/10.1002/asi.v61:7); pmid: 21858251
26. E. J. Gumbel, *Statistics of Extremes* (Dover Publications, 1958).
27. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer-Verlag, 2001).
28. D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, H. A. Makse, Scaling laws of human interaction activity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12640–12645 (2009). doi: [10.1073/pnas.0902667106](https://doi.org/10.1073/pnas.0902667106); pmid: 19617555
29. D. Garlaschelli, G. Caldarelli, L. Pietronero, Universal scaling relations in food webs. *Nature* **423**, 165–168 (2003). doi: [10.1038/nature01604](https://doi.org/10.1038/nature01604); pmid: 12736684
44. R. Sinatra, P. Deville, M. Szell, D. Wang, A.-L. Barabási, A century of physics. *Nat. Phys.* **11**, 791–796 (2015). doi: [10.1038/nphys3494](https://doi.org/10.1038/nphys3494)
45. S. Redner, Citation statistics from 110 years of *Physical Review*. *Phys. Today* **58**, 49–54 (2005). doi: [10.1063/1.1996475](https://doi.org/10.1063/1.1996475)