

به کارگیری قانون بنفورد در شبکه‌های نشر دانشگاهی^۱

لکساندر توسی سی^۱، جرنج ویسیج^۲

۱. دانشکده ریاضیات، علوم طبیعی و فناوری اطلاعات دانشگاه پریمورسکا، کپر، اسلوونی

Izola ۶۳۱۰, ۶ InnoRenew CoE Livade

۲. مرکز تحقیقات اسلوونی آکادمی علوم و هنر اسلوونی، موسسه فران راموف، اسلوونی

کلمات کلیدی: قانون بنفورد، استناد شبکه کتابشناسی

ترجمه^۳ افروز همراهی

دکتری علم اطلاعات و دانش‌شناسی، پژوهشگاه علوم و فناوری اطلاعات ایران،

hamrahi@students.irandoc.ac.ir

چکیده

قانون بنفورد، که به عنوان قانون رقم اول نیز شناخته می‌شود، به طور گسترده‌ای برای آزمایش ناهنجاری در طیف مختلفی از داده‌های نظیر کشف تقلب در حسابداری، قیمت سهام و قیمت خانه گرفته تا قبوض برق، تعداد جمعیت و نرخ مرگ و میر مورد استفاده قرار گرفته است. با افزایش دسترسی به داده‌ها، نمودار همکاری‌های علمی نیز به طور گسترده مورد مطالعه قرار گرفته است. غالب این پژوهش‌ها به تجزیه و تحلیل الگوها، توپولوژی نمودارهای استنادی و نمودارهای هم‌نویسندگی تأکید کرده‌اند. با وجود اینکه جامعه علمی پیچیده و ناهمگون است بیشتر کشورها برای اندازه‌گیری دقیق بروندهای پژوهشی، انتشارات را در گروه‌های مختلف تقسیم‌بندی می‌کنند. این در حالی است که حوزه‌های علمی مختلف، ممکن است فرهنگ‌های متفاوت انتشاراتی داشته باشند که ساخت معیارهای واحد برای ارزیابی بروندهای پژوهش را مشکل‌ساز سازد. در نظام‌های پیچیده‌ای مانند نظام انتشارات، مهم است که ناهنجاری‌های بالقوه به طور منظم مشاهده و با دقت بیشتری بررسی شود تا مدل ارزیابی بهبود یابد یا نقاط ضعف و سوء استفاده‌های احتمالی شناسایی گردد. در این مطالعه، کاربرد بالقوه قانون بنفورد در «پایگاه داده تحقیقات رسمی اسلوونی»^۴ مورد بررسی قرار گرفته است. شواهدی ارائه شده است که نشان می‌دهد معیارهایی مانند تعداد پژوهشگران مقاله با توزیع بنفورد مطابقت دارد؛ اما تعداد نویسندگان در هر مقاله با توزیع مطابقت ندارد. همچنین، استدلال‌هایی بالقوه‌ای برای برخی از ناهنجاری‌های که در این نظام مشاهده می‌شود ارائه شده است.

کلمات کلیدی: قانون بنفورد، شبکه استنادی، کتابشناسی

۱. مقدمه

این مقاله به کاربرد قانون بنفورد در تحلیل شبکه همکاری پژوهش‌های علمی می‌پردازد. نمونه این مطالعه، انتشارات علمی و نویسندگان اسلوونی را دربرمی‌گیرد. این مطالعه می‌تواند مبانی علمی برای روش ارائه شده و دستورالعمل‌های ساده برای بکارگیری آن، فراهم نماید. مطالعه انجام شده بر روی شبکه همکاری تحقیقاتی اسلوونی نشان می‌دهد که قانون بنفورد در شبکه انتشارات دانشگاهی نیز صدق می‌کند. برای تعیین صحت این مسئله که آیا توزیع ارقام اول در نشر علمی از قانون بنفورد تبعیت می‌کند یا خیر؟ چندین حوزه علمی و چند رشته‌ای^۵ نیز مورد آزمایش قرار گرفت.

۱. Use of Benford's law on academic publishing networks

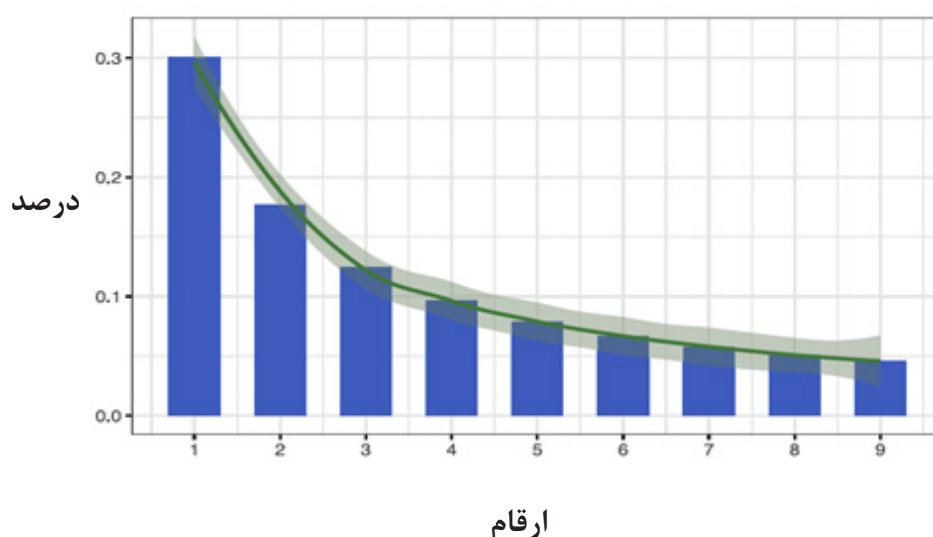
۲. Tošić, Aleksandar, and Jernej Vičič

۳. این مقاله ترجمه‌ای از Tošić, Aleksandar, and Jernej Vičič (2021). «Use of Benford's law on academic publishing networks.» Journal of Informetrics 15, no. 3: 101163.

۴. The Official Research Database of Slovenia

۵. چند رشته‌ای، مفهومی متفاوت از میان رشته‌ای است. در حوزه‌های چند رشته‌ای مانند درمان چندین رویکرد مانند پزشکی، تغذیه، روانشناسی می‌توانند در رابطه با یک مفهوم مورد بررسی قرار گیرند.

پدیده معروف «قانون بنفورد» (بنفورد^۶، ۱۹۳۸؛ سینگلتون^۷، ۲۰۱۱) که با نام «قانون رقم اول» نیز شناخته می‌شود، به مشاهده توزیع فراوانی ارقام اصلی در داده‌های عددی از مجموعه‌های واقعی می‌پردازد. به عبارت دیگر، توزیع ارقام در فرآیندهای طبیعی و اجتماعی را توصیف می‌کند. اعداد به شکل توزیع لگاریتمی هستند. با وجود اینکه، قانون بنفورد بسیار قدیمی است و به طور گسترده مورد بررسی قرار گرفته است، اما هنوز ابزار جالبی برای یافتن ناهنجاری‌ها در داده‌هاست. علاوه بر این، برای تشخیص ناهنجاری، به حجم زیادی از داده‌های تولید شده، به روش‌های ساده و مؤثر نیازمند است. در حالی که قانون بنفورد مخالف یک خواستگاه علمی است (برگر و هیل^۸، ۲۰۱۱)، اما به جای پیشینه نظری بر کاربرد آن تمرکز کرده است (نیگرینی^۹، ۱۹۹۶).



شکل ۱. توزیع ارقام مطابق با قانون بنفورد (سینگلتون، ۲۰۱۱)

این کاربردها، تشخیص تقلب در انتخابات، تشخیص دستکاری تصویر، تشخیص تقلب در حسابداری (دورتسچی، هلیسون و پاچینی^{۱۰}، ۲۰۰۴)، تشخیص تقلب علم (رانستام^{۱۱} و همکاران، ۲۰۰۰) و نظایر آن را در برمی‌گیرد. قانون بنفورد به طور جدی و مستمر در حسابداری مورد استفاده قرار گرفته است (باتاچاریا و کومار^{۱۲}، ۲۰۰۸، نیگرینی، ۲۰۱۲). این روش همچنین با موفقیت در تحلیل‌های دیگر مانند مدل‌سازی ویژگی‌های رفتاری کاربران شبکه‌های اجتماعی (گلبک^{۱۳}، ۲۰۱۵) و رویدادهای هواشناسی، مانند، مسافت‌های طی شده طوفان‌های استوایی از سال ۱۸۴۲ به کار گرفته شده است. هدف این مقاله ارزیابی استفاده بالقوه از قانون بنفورد در شبکه‌های تحقیقاتی است. از این روش می‌توان برای تشخیص اختلافات احتمالی در یک سطح کلان استفاده کرد. پس از شناسایی توزیعی که مشکوک شناخته می‌شود، این روش می‌تواند مشاهداتی ایجاد کند که منحرف (کج) شود (شناسایی چولگی در توزیع داده‌ها). این مشاهدات، یک مدخل عالی برای تجزیه و تحلیل بیشتر و دقیق‌تر برای توضیح و استدلال مشاهدات مشکوک است. این روش برای خود ارزیابی جوامع پژوهشی و همچنین پختگی حوزه‌های پژوهشی خاص مفید است. این قانون می‌تواند برای تنظیم‌گر مثل یک حلقه بازخورد، خدمت کند و در جهت بهبود نظام ارزش‌گذاری و پاسخ به تغییر جهت ناخواسته بالقوه، عمل کند. این روش روی داده‌های زمانی بویژه برای بررسی زمانی که اختلاف رخ داده است، نیز می‌تواند اعمال شود.

۶. Benford

۷. Singleton

۸. Berger & Hill

۹. Nigrini

۱۰. Durtschi, Hillison, & Pacini

۱۱. Ranstam

۱۲. Bhattacharya and Kumar

۱۳. Golbeck

در ادامه این مقاله: بخش ۲، به تشریح قانون بنفورد (سینگلتون، ۲۰۱۱) و کاربردهای آن پرداخته است و با فرض اینکه مقاله چند رشته‌ای است، وضعیت پیشرفت این قانون در حوزه‌های علمی مختلف بیان شده است. بخش ۳، نیز به عنوان مطالعه موردی برای توصیف روش معرفی شده، روش شناسی و نتایج مورد استفاده، به شبکه علمی/تحقیقاتی اسلوونیایی پرداخته است. در انتها نیز بحث و نتیجه‌گیری بیان شده است.

٢. قانون بنفورد

قانون بنفورد یا قانون رقم اول، مشاهده توزیع فراوانی ارقام اصلی است. این قانون به عنوان قانون نیوکمب بنفورد یا قانون بنفورد^{۱۴} نیز شناخته می‌شود. ظاهراً اولین بار توسط پلایمیت نیوکمب^{۱۵} کشف شد و در نیوکمب (۱۸۸۱) منتشر شد و بعداً توسط فیزیکدان اف. بنفورد کشف مجدد شد و در بنفورد (۱۹۳۸) ارائه شد. قانون بنفورد (سینگلتون، ۲۰۱۱) برای ارقام اصلی در هر نوع داده‌ی عددی با ویژگی‌های زیر، توزیع احتمالی ثابتی را تعریف می‌کند:

- داده‌ها با مقادیری که از طریق ترکیب ریاضی اعداد، از چندین توزیع تشکیل شده‌اند.
 - داده‌هایی که تنوع زیادی در تعداد ارقام دارند (به عنوان مثال داده‌هایی با مقادیر زیاد در صدها، هزاران، ده‌ها هزار و غیره)
 - مجموعه داده‌ها نسبتاً بزرگ که به عنوان یک قانون حداقل ۵۰ تا ۱۰۰ مشاهده دارد (کنی^{۱۶}، ۲۰۱۵).
 - داده‌هایی که دارای چولگی در سمت راست هستند (یعنی میانگین بزرگتر از میانه است)، و توزیع به جای متقارن بودن دارای بلندی در سمت راست توزیع است.
 - داده‌ها هیچ مقدار حداکثری یا حداقلی از پیش تعریف شده‌ای ندارند (به استثنای حداقل صفر).
 - توزیع ارقام در شکل ۱ ارائه شده است. رقم ۱ تقریباً در ۳۰ موارد رخ می‌دهد و ارقام دیگر از منحنی لگاریتمی پیروی می‌کنند.
- مطالعات نشان داده است، این نتیجه برای طیف گسترده‌ای از مجموعه داده‌ها صدق می‌شود (سینگلتون ۲۰۱۱)، از جمله، قبوض برق، آدرس خیابان، قیمت سهام، قیمت خانه، تعداد جمعیت، میزان مرگ و میر، و طول رودخانه‌ها. در فرمول زیر، معادله توزیع اولین ارقام داده‌های مشاهده‌شده، ارائه شده است.

$$P(d) = \log \lambda \cdot (d + 1) - \log \lambda \cdot (d) = \log (\lambda + 1) \lambda \cdot /d)$$

۳. شبکه تحقیقات اسلوانی،

در شبکه تحقیقاتی اسلوونی در اسلوونی، پژوهشگران با روشی که توسط آژانس تحقیقات اسلوونی^{۱۷} منتشر شده است، ارزیابی می‌شوند. این روش، مجموعه‌ای شفاف از قوانین است که در نظام ملی اطلاعات تحقیقاتی^{۱۸} نیز پیاده‌سازی شده است (کروسک^{۱۹}، ۲۰۱۴). مشارکت‌های علمی که به پایگاه داده اضافه می‌شوند به توپولوژی طبقه‌بندی می‌شوند و پس از تأیید، به آنها امتیاز اختصاص می‌یابد. تأکید ویژه بر مقالات علمی است که بر اساس داده‌های وب او ساینس^{۲۰}، با نمایه‌های مختلفی نظیر اس سی آی/ای/اس اس سی آی/ای اچ سی آی^{۲۱}؛ جی سی آر^{۲۲}، استنادات وب او ساینس و اسکوپوس^{۲۳}، اسنپ^{۲۴} ارزیابی می‌شوند (فالاگاس، پیتسونی، مالیتریس و پاپاس^{۲۵}، ۲۰۰۸). چنین نظام قوی، برای ارائه یک روشی جامع به دولت به منظور تأمین مالی پژوهشگران و طرح‌های داخلی و خارجی مورد نیاز است. این نظام از دکتري به بعد، برای ارتقای رتبه علمی با در نظر گرفتن حداقل الزامات نیز استفاده می‌شود.

مروری دقیق و نظام‌مند از شبکه تحقیقاتی اسلوونی توسط کرک^{۲۶} در سال ۲۰۱۹ ارائه شده است. مقالات مجلات با توپولوژی ۱،۰۱، ۱،۰۲ و ۱،۰۳ جایگذاری شده و به چهار دسته (چهارم)، مطابق با موقعیت مجله در حوزه علمی، رتبه‌بندی شده است تا برای محاسبه امتیاز، به کار رود؛ سپس به طور مساوی بین نویسندگان تقسیم شود.

۱۴. Newcomb Benford law or Benford's law

15. polymath Newcomb

16. Kenny

LY-ARRS

1.4. SICRI

19. Korosec

۲۰. Web of Science (WoS)

۲۱. SCIE/SSCI/AHCI

۲۲. Journal of Citation Reports impact factor) JCR(

۲۳. Scopus

۲۴. Source Normalized Impact per Paper impact (SNIP)

25. Falagas, Pitsouni, Malietzis, & Pappas

۲۶. Curk

امتیازهای اضافی بر اساس تعداد استناد به طور مساوی به هر نویسنده اضافه می‌شود. یکی دیگر از جنبه‌های مهم، حداقل الزامات پایان‌نامه‌های دکتری است؛ که داوطلبان را ملزم می‌کند حداقل یک مقاله در یک مجله نمایه شده اس سی آی^{۲۷} منتشر کنند. هنبیرگ^{۲۸} (۲۰۱۶) مشاهده کرد که شاخص‌های کتاب‌سنجی به طور فزاینده‌ای بر مشاغل، بودجه و شهرت تأثیر می‌گذارند و سبب شکل‌گیری نگرش افراطی جدیدی می‌شود که معتقد است «جایی که یک دانشمند منتشر کرده است، بسیار مهم‌تر از آن چیزی است که منتشر کرده است» که با مطالعات دیگر (هولوب، تاپینر و ابرهارتر^{۲۹} ۱۹۹۱؛ و شیبایاما و بابا^{۳۰} ۲۰۱۵). نیز همخوانی دارد. در حالی که به همان اندازه که برای ارزیابی مشارکت‌های پژوهشی، یک نظام جامع و قوی ضرورت دارد، به نظامی که بتواند مشخصات حوزه پژوهشی را هم توصیف کند، نیاز است. با توجه به عادات و فرهنگ‌های انتشاراتی بسیار متفاوت در برخی از حوزه‌های پژوهشی، ارائه چارچوب یکپارچه ارزیابی با قابلیت مقایسه و سادگی، دشوار است. لاریویر، آرشمبل، گینگراس و وینگنولا-گانه^{۳۱} (۲۰۰۶) به این نتیجه رسیدند که تفاوت‌های قابل توجهی در عادات انتشار بین علوم اجتماعی و علوم انسانی و علوم طبیعی و مهندسی^{۳۲} وجود دارد که مشکل خاصی را در حوزه ارزش‌گذاری کتاب‌سنجی ایجاد می‌کند.

اختلاف بین معیارهای ارزش‌گذاری و عادات انتشار می‌تواند انگیزه‌های ناخواسته‌ای ایجاد کند. این مسئله به همراه یک نظام تأمین مالی مبتنی بر نقطه سیستمی، می‌تواند انگیزه‌های نامطلوبی برای سوء استفاده از روش ارزش‌گذاری به نفع خود، بوجود آورد. علاوه بر این، با توجه به توپولوژی مشارکت، حوزه‌های پژوهشی مختلف، فرهنگ‌های انتشاراتی متفاوتی دارند. در سطح کلان، مشاهده و نظارت بر شبکه انتشارات به منظور شناسایی روندها و رفع اختلافات بالقوه اهمیت دارد. به طور کلی، هم برای دانشگاه به منظور شناسایی روندها و هم برای دولت به منظور تنظیم معیارهای ارزش‌گذاری و هدایت پژوهش به سمت اهداف از پیش تعریف شده، اهمیت زیادی دارد. علاوه بر این، لازمه شناسایی تغییرات غیرمنتظره در شبکه نشر در سطح کلان، تحلیل دقیق‌تر و جزئی‌تر در سطح خرد است.

۴. پیشرفت قانون

قانون بنفورد به طور کامل مورد بررسی قرار گرفته و مبانی نظری آن در بسیاری از مقالات علمی به اثبات رسیده است. این پدیده با جزئیات بیشتری در مطالعات برگر و هیل^{۳۳} (۲۰۱۱) مورد بحث قرار گرفته است. این مقاله، نسخه گسترش یافته و شواهد ساده شده‌ای از بسیاری از نتایج کلیدی را در پیشینه ارائه می‌دهد. بسیاری از پژوهشگران تأیید کرده‌اند که قانون به طور گسترده رعایت می‌شود، اما همچنین خاطرنشان کرده‌اند که توضیحات رایج کاملاً رضایت بخش نیستند (فیوستر^{۳۴} ۲۰۰۹). مطالعات کتاب‌سنجی و اطلاع‌سنجی به موضوع شبکه دانشگاهی و شبکه هم‌نویسندگی پرداخته است که به طور گسترده از سطح جهانی یا ملی (براون و گلنزل^{۳۵} ۱۹۹۶، لیدسدورف و واگنر^{۳۶} ۲۰۰۸) به سطح فردی (ملین^{۳۷} ۲۰۰۰، نیومن^{۳۸} ۲۰۰۴) می‌رود. آریل زو و چانگ^{۳۹} (۲۰۲۰) نشان می‌دهند که شبکه هم‌نویسندگی به خوبی به عملکرد دانشگاهی بستگی دارد. مقالات متعددی وجود دارد که رابطه مثبتی بین همکاری بین‌المللی و تأثیر پژوهش پرداخته است (نارین و ویتلو^{۴۰} ۱۹۹۰، کاتز و هیکس^{۴۱} ۱۹۹۷).

۲۷. SCI

۲۸. Heneberg

۲۹. Holub, Tappeiner, and Eberharter

۳۰. Shibayama and Baba

۳۱. Larivière, Archambault, Gingras, & Vignola-Gagné

۳۲. social sciences and humanities (SSH) and natural sciences and engineering (NSE)

۳۳. Berger and Hill

۳۴. Fewster

۳۵. Braun and Glaenzell

۳۶. Leydesdorff and Wagner

۳۷. Melin

۳۸. Newman

۳۹. Ariel Xu and Chang

۴۰. Narin and Whitlow

۴۱. Katz and Hicks

پیشینه وضعیت فعلی و روندهای شبکه‌های اجتماعی دانشگاهی توسط کنگ، شی، یو، لئو و شیا^{۴۲} (۲۰۱۹) مورد بررسی قرار گرفته و یافته‌ها ارائه شده است. مطالعه‌ای (بناونت پرز، گوریز، گامپنبرگر و دو مویا-آنگن^{۴۳} ۲۰۱۲) بر روی همکاری پژوهشی و مطالعه‌ای دیگر (بوردونز، گومز، فرناندز، زلوئتا و مندز^{۴۴} ۱۹۹۶) به خوبی مطالعه قبلی به تأثیر جغرافیای بین‌المللی و همکاری‌های پژوهشی بین‌المللی پرداخته است. اورتگا^{۴۵} (۲۰۱۴) تحلیلی از رابطه بین تأثیر پژوهش و ویژگی‌های ساختاری شبکه‌های هم‌نویسندگی ارائه کرده است. برای مشاهده بلوغ حوزه پژوهش، روش توصیف شده در این مقاله مناسب‌ترین روش است. در راستای این پژوهش مطالعاتی مانند کیتلی-هرینگت^{۴۶} و همکاران (۲۰۱۶) لاکو، رویز، سانز، تارانکون و کلمنته-گالاردو^{۴۷} (۲۰۲۱) انجام شده است، که تحلیلی از تکامل یک حوزه علمی هدفمند را ارائه می‌دهد. یکی از اهداف، مشاهده تغییرات در شبکه کتابشناختی در طول زمان است که به بلوغ شبکه منجر شده است. باتاگلج و مالتسوا^{۴۸} (۲۰۲۰) با استفاده از سال انتشار آثار، روشی را برای تبدیل شبکه‌های کتابشناختی به شبکه‌های زمانی متناظر، پیشنهاد می‌کند که براساس کمیت‌های زمانی، ویژگی‌های زمانی جالب گره‌ها، پیوندها و گروه‌های آنها تعریف می‌شود تا بینشی از تکامل شبکه‌های کتابشناختی فراهم نماید.

۴-۱. کاربردهای قانون بنفورد

رایج‌ترین کاربردها قانون بنفورد، در تشخیص تقلب حسابداران است و تحقیقات زیادی در این حوزه انجام شده است مانند دریک و نیگرینی^{۴۹} (۲۰۰۰) که دانش آموزان را با قانون بنفورد و تجزیه و تحلیل دیجیتال (تحلیل الگوهای رقمی و عددی مجموعه داده) آشنا می‌کند، که می‌تواند به عنوان یک روش تحلیلی و ابزار تشخیص تقلب مورد استفاده قرار گیرد. نیگرینی (۲۰۱۷) مروری بر پیشینه فعلی حوزه ارائه می‌دهد. دورتسچی^{۵۰} و همکاران (۲۰۰۴) قانون بنفورد را به عنوان ابزاری ساده و مؤثر برای تشخیص تقلب معرفی کرده‌اند. هدف این پژوهش، کمک به حساب‌رسان برای استفاده موثرتر از تجزیه و تحلیل دیجیتال بر اساس قانون بنفورد است تا مجموعه داده‌هایی که می‌توان انتظار داشت که از توزیع بنفورد پیروی کند شناسایی شود و انواع تقلب‌هایی که با چنین تجزیه و تحلیلی «کشف یا تشخیص داده نمی‌شوند»، ارائه گردد. کلیری و تیودو^{۵۱} (۲۰۰۵) به برخی مشکلات ذاتی اشاره می‌کنند که در استفاده از قانون بنفورد در فرآیند حسابرسی به طور بالقوه ایجاد می‌شود. این مقاله محاسن قانون بنفورد و روش آماری معمولی آزمون به آزمون^{۵۲} را مقایسه می‌کند. سادگی قانون بنفورد به عنوان یک ابزار، طیف وسیعی از کاربردها را امکان‌پذیر می‌سازد. هیکمن و رایس^{۵۳} (۲۰۱۰) آمار جرم و جنایت را در سطح ملی، ایالتی و محلی ایالات متحده مورد بررسی قرار داده‌اند تا مطابقت آن را با توزیع بنفورد آزمایش کنند. برک و کینکانن^{۵۴} (۱۹۹۱) توزیع ارقام اولیه ثابت‌های فیزیکی^{۵۵} را مشاهده کرده‌اند، گرچه نتایج آنها قطعی نیست. یکی از پژوهش‌های اخیر در حیطه قانون بنفورد، مطالعات ژانگ^{۵۶} (۲۰۲۰) هست، نویسندگان آزمایشی را پیشنهاد می‌کنند که بر اساس آن تعداد گزارش شده‌ی موارد ابتلا به ویروس کرونا در سال ۲۰۱۹ در چین با قانون بنفورد ارائه شود، نتایج نشان می‌دهد که تعداد گزارش‌شده‌ی افراد مبتلا از قانون بنفورد تبعیت می‌کنند.

۴۲. Kong, Shi, Yu, Liu, and Xia

۴۳. Benavent-Pérez, Gorraiz, Gumpenberger, and de Moya-Anegón

۴۴. Bordons, Gomez, Fernández, Zulueta, and Méndez

۴۵. Ortega

۴۶. Keathley-Herring et

۴۷. Pelacho, Ruiz, Sanz, Tarancón, and Clemente-Gallardo

۴۸. Batagelj & Maltseva

۴۹. Drake & Nigrini

۵۰. Durtschi

۵۱. Cleary and Thibodeau

۵۲. test-by-test

۵۳. Hickman and Rice

۵۴. Burke and Kincaid

۵۵. ثابت فیزیکی یک کمیت فیزیکی است که در تمام جهان ثابت است و در طول زمان تغییر نمی‌کند. مانند: سرعت نور

۵۶. Zhang

۲-۴. یکپارچگی تحقیقات و رفتار غیراخلاقی

قانون بنفورد با موفقیت در کشف تقلب حسابداری استفاده شده است و تعداد کمی از پروژه های تحقیقاتی مانند ژانگ (۲۰۲۰) با هدف استفاده از ابزار مشابه برای کشف تقلب در سایر حوزه ها نظیر تشخیص گزارش های جعلی کوید ۱۹ استفاده کرده اند. طبیعی است که از حوزه مالی الهام گرفته و آن را به حوزه جدید انتقال دهیم، اما هدف این نیست؛ استفاده از ابزار ارائه شده به عنوان معیار تشخیص تقلب، کماکان باید مورد بررسی قرار گیرد. مشکل یکپارچگی پژوهش و رفتار غیراخلاقی گسترش یافته است. در اشکال و ابعاد جدید که توسط بسیاری از محققان مانند مارتین^{۵۷} (۲۰۱۳) و بوهانون^{۵۸} (۲۰۱۳) مشاهده شده است. در حال حاضر مطالعات فزاینده ای (مانند مارتین^{۵۹} ۲۰۱۳ و پترسون^{۶۰} ۲۰۰۷) در مورد یکپارچگی علمی و تخلفات، و همچنین ارائه دستورالعمل ها و بهترین شیوه ها وجود دارد. با استفاده از کتاب های درسی (مانند ماکرینا^{۶۱} ۲۰۱۴ و اسپون هولز^{۶۲} ۲۰۰۰) دوره های کاملی در سطح دانشگاه به این موضوع اختصاص داده شده است. در سال های اخیر مقالات و نظرات پژوهشی بسیاری (مانند گودچارل، نمری و دیریکس^{۶۳} ۲۰۱۴ و برنشتاین^{۶۴} ۱۹۸۴) منتشر شده است که در خصوص صحت تحقیقات، ابزار نگرانی می کنند و به موردی از مسائل اخلاقی مطرح شده می پردازند. ادواردز و روی^{۶۵} (۲۰۱۷) در رابطه با این که، دانشمندان در رقابت برای تأمین مالی پژوهشی و توسعه معیارهای کمی برای اندازه گیری عملکرد، به طور فزاینده ای منحرف شده اند؛ بحث می کنند. در مطالعاتی مانند ون در هایدن، ون دو درکس ون و اپتوف^{۶۶} (۲۰۰۹)، اثربخشی نظام داوری همتا به عنوان ابزاری برای جلوگیری از سوء رفتارهای علمی به چالش کشیده شده است. نظرسنجی ها حاکی از تعداد فزاینده و افراطی رفتارهای نادرست است (جان، لوونشتاین و پرلک^{۶۷} ۲۰۱۲). این مطالعه، در پی آن نیست که موضوع ظریف صداقت علمی را از نو تعریف کند یا ابزاری ارائه نماید که بتواند رفتار نادرست یک نویسنده یا یک مشارکت پژوهشی ضعیف را تبیین کند؛ بلکه، می خواهد روشی را برای پیگیری تغییرات و ارزیابی بلوغ نظام پژوهشی پیشنهاد نماید.

۳-۴. پژوهش در شبکه تحقیقاتی اسلوونی

هر منبع کتابشناختی جدید از نظر مناسب بودن آن برای تحلیل-های کتابسنجی، باید مورد آزمایش قرار گیرد، مانند کاوش دانشگاهی میکروسافت (اورتگا، ۲۰۱۴). شبکه دانشگاهی/پژوهشی اسلوونی، از دیدگاه های مختلف از جمله رویکرد روش شناختی کمی و کیفی به همکاری علمی (مالی، پوستور، کوگماس و فرلیگوج^{۶۸} ۲۰۱۸)، مطالعه ساختارهای جامعه از طریق هم نویسنده گی علمی (کوگماس، فرلیگوج، و کرونگر^{۶۹} ۲۰۱۹) به طور گسترده مورد بررسی قرار گرفته است. پیسانسکی^{۷۰}، پیسانسکی و پیسانسکی (۲۰۲۰) برای سهولت مصورسازی شبکه های بزرگ مانند شبکه های کتابشناختی دو روش ارائه کرده اند و آن ها را در شبکه تحقیقاتی اسلوونی به نمایش گذاشته اند. فرلیگوج، کرونگر، مالی، اسنایدرز و دوریان^{۷۱} (۲۰۱۵) ساختارهای همکاری و پویایی شبکه هم نویسنده گی همه پژوهشگران اسلوونی را مورد بررسی قرار داده اند. هدف از این بررسی، شناسایی عوامل کلیدی ایجاد همکاری و تفاوت های اصلی در رفتار همکاری است. در مطالعات کارلوفوسک و ملادنچیک^{۷۲} (۲۰۱۵) مقیاس جدیدی برای حوزه های بین رشته ای پیشنهاد شده است که محتوا و ساختار گراف را در نظر می گیرد. مقیاس پیشنهادی جدید، در تجزیه و تحلیل اکتشافی جامعه تحقیقاتی اسلوونی مورد استفاده قرار می گیرد. در تفسیر این مقاله، همچنان که رودلا^{۷۳} (۲۰۱۶) از ارزش مقاله حمایت می کند، به دو کاستی آن نیز می پردازد. لوزار، لوناجیچ، پو و پرک^{۷۴} (۲۰۱۴) مطالعه ای در مورد پویایی علوم بین رشته ای در شبکه علمی اسلوونی ارائه کرده اند.

۵۷. Martin

۵۸. Bohannon

۵۹. Peterson

۶۰. Macrina

۶۱. Sponholz

۶۲. Godecharle, Nemery, and Dierickx

۶۳. Bernstein

۶۴. Edwards and Roy

۶۵. Van der Heyden, van de Derks Ven, and Opthof

۶۶. John, Loewenstein, & Prelec

۶۷. Mali, Puštorh, Cugmas, and Ferligoj

۶۸. Cugmas, Ferligoj, & Kronegger

۶۹. Pisanski

۷۰. Ferligoj, Kronegger, Mali, Snijders, and Doreian

۷۱. Karlovčec and Mladenić

۷۲. Rodela

۷۳. Lužar, Levnajić, Povh, and Perc

۵. روش شناسی

همانطور که در بخش ۱ ذکر شد، این مقاله روشی برای پیگیری تغییرات و ارزیابی بلوغ نظام پژوهش، پیشنهاد می‌کند. به این ترتیب، هدف ارائه مبانی علمی است که امکان، امکان‌سنجی و کاربرد روش را فراهم کند و همچنین مجموعه‌ای از دستورالعمل‌های استفاده و یک مورد کاربردی که در آن فرضیه‌ها تأیید شوند را پیشنهاد کند. همانطور که در بخش ۲ توضیح داده شد، مجموعه مشاهدات باید با تمام پیش نیازهای اساسی قانون بنفورد، مطابقت کند. در بخش ۳ اشاره شد که به عنوان نمونه‌ای از شبکه انتشارات علمی، شبکه تحقیقاتی اسلوونی مورد استفاده قرار گرفته است. روش‌ها و مکان‌ها به راحتی برای سایر کشورهای اروپایی و غیر اروپایی قابل استفاده است. داده‌های شناسایی شده در این مطالعه شامل موارد زیر است، اما به این موارد محدود نمی‌شود:

- تعداد انتشارات به ازای هر نویسنده
- تعداد نویسندگان مشترک به ازای هر نویسنده
- طبقه‌بندی نویسندگان معمولاً در یک یا چند حوزه علمی
- ارتباط نویسندگان با یک یا چند مؤسسه تحقیقاتی که می‌توانند موقعیت جغرافیایی بیشتری بوجود آورد.
- برچسب‌گذاری همه انتشارات با سال انتشار

مقایسه دیگری که معمولاً هنگام مشاهده داده‌های ارائه شده انجام می‌شود «تعداد نویسندگان در هر نشریه» است، با اینکه از نظر تئوری، این تعداد نامحدود است اما در عمل اینگونه نیست و تنها تعداد کمی از بی‌شمار انتشارات حذف می‌شوند و بیشتر اعداد ۱ یا ۲ رقمی مانده‌اند. فرض بر این است که پایایی در بخش ۷ تأیید می‌شود و سنج‌های زیر همگی مشمول قانون بنفورد هستند:

تعداد نویسندگان مشترک در ازای نویسنده

تعداد انتشارات به ازای هر نویسنده

بیشترین تعداد انتشارات برای هر نویسنده در مجموعه داده‌های این پژوهش ۸۴۸۳ بود که بیش از ۰/۵ آن‌ها بالاتر از ۱۰۰۰ بود، که این پیش نیاز دوم که در بخش ۲ بیان شده است (هرچند تا حدودی) را برآورده می‌کند. در مجموعه‌های انتخاب شده هر یک از مفروضات را می‌توان بیشتر توزیع کرد:

۱. کل شبکه علمی (اسلوونیایی)

۲. تقسیم به حوزه‌های علمی (به عنوان مثال، علوم طبیعی، علوم اجتماعی، و غیره)

۳. گروه‌بندی نشریات در دوره‌های زمانی گسسته

۴. گروه‌بندی نویسندگان در مناطق جغرافیایی

دلیل انتخاب این مجموعه‌ها به شرح زیر است: ۱. همه انتشارات، ۲. آیا حوزه علمی وجود دارد که در مراحل اولیه توسعه است، یا انتشارات و ارتباطاتی که با برخی ویژگی‌های دیگر منحرف شده‌اند، ۳. آیا دوره زمانی وجود داشت که مجموعه مشاهده شده، مطابق قانون بنفورد رفتار نمی‌کرد. ۴. مناطق جغرافیایی (معمولاً مربوط به مرزهای ملی - برداشت قومی یا نهادی از نشریات پژوهشی) وجود دارد که در آن، مجدداً، اعداد با قانون بنفورد مطابقت ندارند. آزمایش اینکه داده‌ها با توزیع بنفورد مطابقت دارند، با آزمون‌های برازش فراوان انجام شده است. آزمایش انطباق داده‌ها با توزیع بنفورد، با بسیاری از آزمون‌های نیکویی برازش انجام شده است که از آزمون‌های چي پیرسون^{۷۴}، آمار کولموگوروف-اسمیرنوف دی^{۷۵}، اصلاح فریدمن آمار یوتو واتسون^{۷۶}، آمار فاصله اقلیدسی^{۷۷} و بسیاری آزمون‌های دیگر را دربرمی‌گیرد.

با این حال، هیچ داده واقعی هرگز از توزیع دقیق پیروی نمی‌کند. از این رو، اکثر تجزیه و تحلیل‌ها، آزمون‌های آماری را با نمایش‌های گرافیکی تکمیل می‌کنند که با نشان دادن الگوهای مشکوک در داده‌ها به بررسی‌های بیشتر کمک می‌کند. علاوه بر این، آزمون‌های مختلف، واکنش‌های متفاوتی نسبت به اندازه نمونه دارند. آزمون مجذور خی از مشکل توان اضافی رنج می‌برد، زیرا وقتی تعداد مشاهدات زیاد می‌شود بیش از ۵۰۰۰ رکورد تخمین زده شده (نیگرینی ۲۰۱۲) نسبت به خوشه‌های ناچیز حساس‌تر می‌شود و به این نتیجه می‌رسد که داده‌ها مطابقت ندارند.

۷۴. Pear-son's Chi squared,

۷۵. Kolmogorov-Smirnov D statistics

۷۶. Freedman's modification of Watson U2 statistics

۷۷. Euclidean distance dstatistics

نیگرینی (۲۰۰۱) پیشنهاد کرد، برخی از آزمون‌های در صورت اعمال بر تعداد زیادی از مشاهدات می‌توانند نتایج گمراه کننده‌ای ارائه کنند. از سوی دیگر، درویکا، اوانسیا و والسان^{۷۸} (۲۰۱۸) به این نتیجه رسیدند که آزمون ماد(میانۀ قدر مطلق انحراف)^{۷۹} حداقل با ۲۰۰ مشاهده قابل اعتماد است.

الکساندر^{۸۰} (۲۰۰۹) آزمون قوس مانتیسا^{۸۱} را پیشنهاد کرد که یک آزمون هندسی بسیار جالب است. متأسفانه، انحراف کمی از توزیع‌های بنفورد را تحمل می‌کند. نیگرینی (۲۰۱۲) به این نتیجه رسید که بهترین آزمون، میانۀ قدر مطلق انحراف (ماد) است، همچنین مقدار عینی بحرانی برای انطباق (۰,۰۰۰)، انطباق قابل قبول (۰,۰۰۶)، انطباق تا حد قابل قبول (۰,۰۱۲) و عدم انطباق را عدد (۰,۰۱۵) تعیین می‌کند. برای اندازه گیری میانگین انحراف بین ارتفاع میله‌ها و خط بنفورد، ماد اقتباس شده استفاده می‌گردد. هرچه ماد بالاتر شود، میانگین تفاوت بین نسبت‌های واقعی و مورد انتظار بیشتر می‌شود. با توجه به اینکه در این مطالعه، اندازه‌های نمونه به خوبی در محدوده قابل قبول است، تمام آزمون‌های انطباق با استفاده از هر سه آزمونی که در این پاراگراف ذکر شد انجام می‌شود و با نمایش‌های گرافیکی آزمون‌های آماری تکمیل می‌گردد. نتایج در بخش ۷ ارائه شده است. دستور اولیه: یک مجموعه تقریباً بزرگ از داده‌های جمع شده را انتخاب کنید که با پیش نیازهای قانون بنفورد مطابقت دارد. داده‌ها را جمع‌آوری نمایید و مقادیر را شمارش کنید. مقادیر پیشرو را بشمارید و میانگین انحراف مطلق (ماد) را روی داده‌های جمع‌آوری شده اعمال کنید. برای هر رقم اول، نمودارهای ساده را با اعداد ترسیم کنید تا توزیع را مشاهده کنید. اگر داده‌ها با قانون بنفورد مطابقت نداشت، بیشتر بررسی کنید.

۵-۱. توصیف مجموعه داده‌ها و فرآیند جمع‌آوری داده‌ها (شبکه دانشگاهی اسلوونی)

شبکه دانشگاهی اسلوونی از طریق دو سرویس عمومی ارائه می‌شود: نظام اطلاعات پژوهشی فعلی اسلوونی^{۸۲} (کرک^{۸۳} ۲۰۱۹) که داده‌های مربوط به پژوهش‌های علمی مانند آموزش دانشمندان / پژوهشگران را ذخیره می‌کند. مدرک، رشته علمی، وابستگی، نوع اشتغال، و پروژه‌های تحقیقاتی با بودجه ملی؛ و نظام و خدمات کتابشناسی تعاونی^{۸۴} (سلجاک بوشنچک^{۸۵} ۲۰۰۶) که داده‌های مربوط به همه انتشارات (از جمله انتشارات پژوهشی که دغدغه اصلی این مقاله هست) را ذخیره می‌کند.

برای شناسایی منحصر به فرد پژوهشگران، یک طرح طبقه‌بندی پژوهشی استفاده می‌شود. کتابشناسی پژوهشگران در فرآیند فهرست‌نویسی مشترک ایجاد می‌شود. با این حال، استفاده از روش‌شناسی یکسان مدارک/آثار برای طبقه‌بندی اقلام کتابشناختی الزامی است. این دو سرویس به طور یکپارچه با هم ترکیب شده و از طریق یک رابط پرس‌وجو در دسترس عموم قرار می‌گیرند. اگرچه داده‌ها به صورت عمومی در دسترس هستند و می‌توان برای اهداف پژوهشی، آن‌ها را به دست آورد؛ اما در این پژوهش، گزینش داده‌های موجود ترجیح داده شده است. در مارس ۲۰۲۰، با استفاده از خدمات یک خزنده آرام، یک پایگاه داده محلی به منظور جلوگیری از مشکلات بارگذاری ساخته شد.

هر پژوهشگر با یک شناسه پژوهشگر منحصر به فرد^{۸۶} که توسط آژانس تحقیقاتی اسلوونی صادر شده بود، نمایش داده می‌شد. در این روش، نویسندگان همکار خارجی ابهام‌زدایی نشده‌اند. برای هر نویسنده مشترک که در مدخل کتابشناختی مشاهده می‌شود، یک موجودیت جدید ایجاد شده است. داده‌های جمع‌آوری‌شده در این فرآیند تا زمانی معتبر هستند که نویسندگان اسلوونیایی را مشاهده کنید و از نویسندگان همکار خارجی فقط به صورت تجمیع شده استفاده نمایید. ساختار پایگاه داده محلی در شکل ۲ ارائه شده است.

۷۸. Druica, Oancea, and Vâlsan

۷۹. MAD

۸۰. Alexander

۸۱. Mantissa Arc test

۸۲. Slovenian Current Research Information System (SICRIS)

۸۳. Curk

۸۴. Cooperative Bibliographic System and Services (COBISS)

۸۵. Seljak & Bošnjak

۸۶. MSTID

۶. استفاده نادرست از معیار پیشنهادی

این بخش نمونه‌ای را ارائه می‌کند که در آن معیار پیشنهادی نتایج گمراه کننده‌ای به همراه خواهد داشت. قانون بنفورد برای تعداد نویسندگان به ازای هر مورد مشارکت پژوهشی، قابل اجرا نیست زیرا محدودیت دوم ذکر شده در بخش ۲ را نقض می‌کند که به تنوع زیادی از تعداد ارقام، گستره‌ای از صدها و هزارها، نیاز است. حداکثر تعداد نویسندگان همکار با یک مشارکت در شبکه دانشگاهی اسلوونی ۴۰ نفر است و تنها ۹۲ مشارکت با ۳۰ همکار یا بیشتر، و فقط ۳۲۳ مشارکت با ۲۰ همکار یا بیشتر وجود دارد. شکل ۳، توزیع نامتقارن ارقام اول را نشان می‌دهد. ارقام اول با حجم نمونه ۷۸۸۴۱۰ مشاهده، تجزیه و تحلیل شد. انطباق ماد (نگرینی ۲۰۱۲) با تمام آزمون‌های آماری برابر عدم انطباق طبقه‌بندی شد و مقادیر پی^{۸۷} نزدیک به گزارش شد، زیرا تعداد نویسندگان در هر مقاله با قانون بنفورد مطابقت ندارد.

۷. بحث و نتیجه‌گیری

در این بخش نتایج روش‌های اعمال شده بر روی مجموعه‌های شناسایی شده ارائه می‌گردد (همه در بخش ۵ توضیح داده شده است). شکل ۴ نشان می‌دهد که توزیع ارقام اول برای تعداد مشارکت‌های هر نویسنده برای کل مجموعه داده‌ها (به عنوان مثال، کل شبکه علمی اسلوونی) مطابق با بنفورد است. قانون انطباق ماد براساس مطالعات نگرینی (۲۰۱۲) قابل قبول است. جدول ۲، نتایج آزمایش حوزه‌های پژوهشی فردی را، جداگانه بر روی مقالات کنفرانس‌ها و مقالات مجلات بطور خلاصه بیان می‌کند. آمار اصلی مورد استفاده برای تعیین اینکه آیا داده‌ها با قانون بنفورد مطابقت دارند یا خیر، میانگین انحراف مطلق یا همان انطباق ماد بود (نگرینی، ۲۰۱۲). مدل عامل اعوجاج^{۸۸}، نشان می‌دهد که آیا الگوهای رقمی بیش از حد یا کمتر از مقدار و میزان اعوجاج هست، همچنین میزان اعوجاج را نیز نشان می‌دهد. علاوه بر این، در مواردی که تا حدودی انطباق وجود دارد آزمون‌های کای دو پیرسون و مانیتسا آرک^{۸۹} برای وضوح و تأیید، ارائه شده است. بر اساس آزمون‌های فوق، مجموعه داده‌ها به صورت عدم انطباق، مرزی، قابل قبول یا انطباق دقیق طبقه‌بندی می‌شود. نتایج نشان می‌دهد، به استثنای تعداد مقالات منتشر شده در علوم اجتماعی و مقالات کنفرانس‌های منتشر شده در علوم انسانی و بیوتکنولوژی، بیشتر حوزه‌ها به شدت مطابقت دارند. برای مجموعه داده‌های واحد، تجزیه و تحلیل بیشتر از نمودارهای توزیع رقمی می‌توان ترسیم کرد. شکل‌های ۵ و ۶ بینش بیشتری در حوزه‌های پژوهشی فردی و توپولوژی مشارکت ارائه می‌دهد. اگر اختلاف مجذور بین توزیع رقم مشاهده شده و توزیع بنفورد بیشتر از ۴ شود، مشکوک‌ها (با رنگ قرمز) علامت‌گذاری می‌شوند. یافتن آستانه صحیح یک فرآیند تجربی است که در آن برای فراهم نمودن مجموعه‌ای بزرگتر از مشاهدات که بیشترین انحراف را دارند، لازم است مشاهدات مشکوک را از بالاترین آستانه (مشاهده با بالاترین مجذور اختلاف) شروع کرد و آن را در یک فرآیند گام به گام کاهش داد تا بدست آید. این مشاهدات می‌تواند بینش گسترده‌تری در مورد نتایج انطباق عینی ماد ارائه دهد. آستانه به صورت تجربی صرفاً برای نمایش گرافیکی، معنادار انتخاب شد. علوم اجتماعی دارای بیشترین انحراف در سه رقم اول است، که می‌توان به دلیل فرهنگ نشر متفاوت یا افزایش تعداد پژوهشگرانی که تنها به نشر یک مقاله برای مقاصد علمی یا پایان نامه دکتری نیاز دارند توضیح داده شوند. این پژوهش، ثابت می‌کند که، به طور کلی، تعداد مقالات منتشر شده به ازای هر نویسنده با توزیع بنفورد، با برخی از واریانس‌های بین حوزه‌های مختلف پژوهشی مطابقت دارد، که می‌توان به فرهنگ‌های مختلف انتشاراتی نسبت داد. با این حال، استفاده از قانون بنفورد اگر بدون در نظر گرفتن جنبه زمانی داده‌ها انجام شود، برای شناسایی احتمالی ناهنجاری‌ها در انتشارات بسیار مفید نیست. با این حال، استفاده از قانون بنفورد برای شناسایی احتمالی ناهنجاری‌ها در انتشار، اگر بدون در نظر گرفتن جنبه زمانی داده‌ها انجام شود، بسیار مفید نیست. این پژوهش ثابت می‌کند که، قانون بنفورد می‌تواند در بازه‌های زمانی کوتاه‌تر برای محدود کردن جستجو به میزان قابل توجهی استفاده شود.

۸۷. P-value

کمترین مقدار خطای نوع اول (آلفا) که منجر به رد شدن فرض صفر می‌شود.

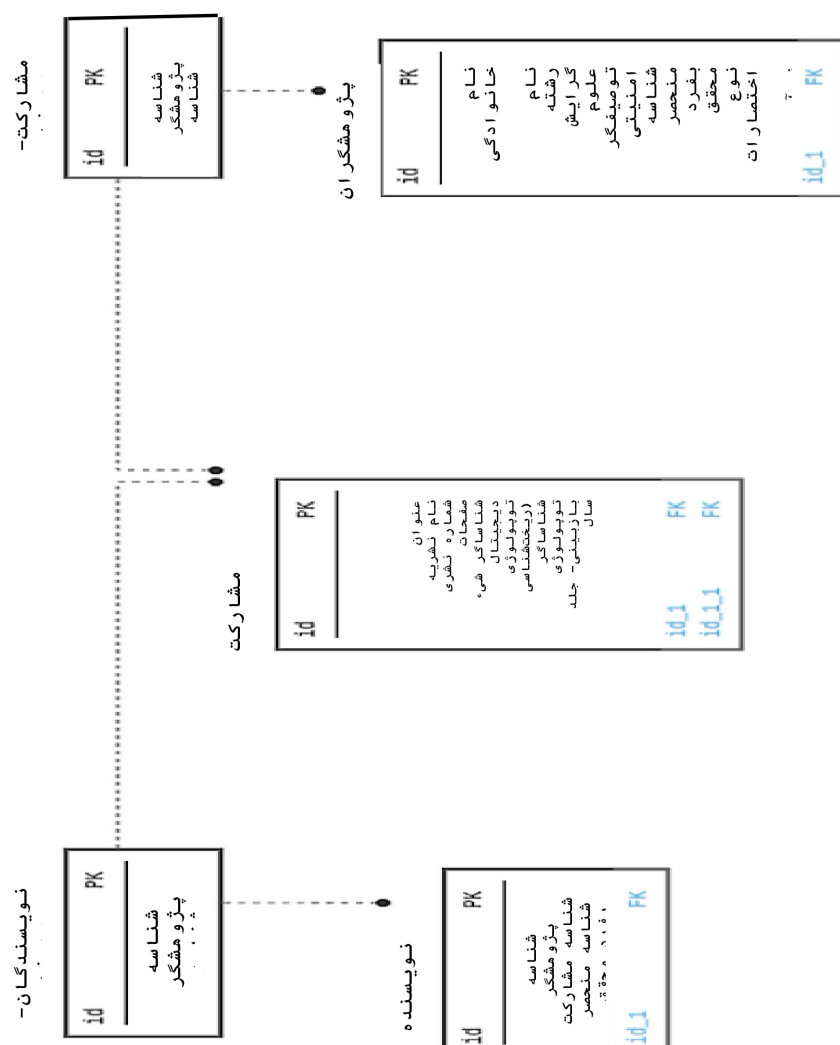
۸۸. به هرگونه تغییر نامطلوب در شکل موج یک سیگنال الکتریکی که از درون یک مدار که دارای دستگاه انتقال است، اعوجاج گفته می‌شود.

۸۹. Pearson's Chi-squared and Mantissa Arc

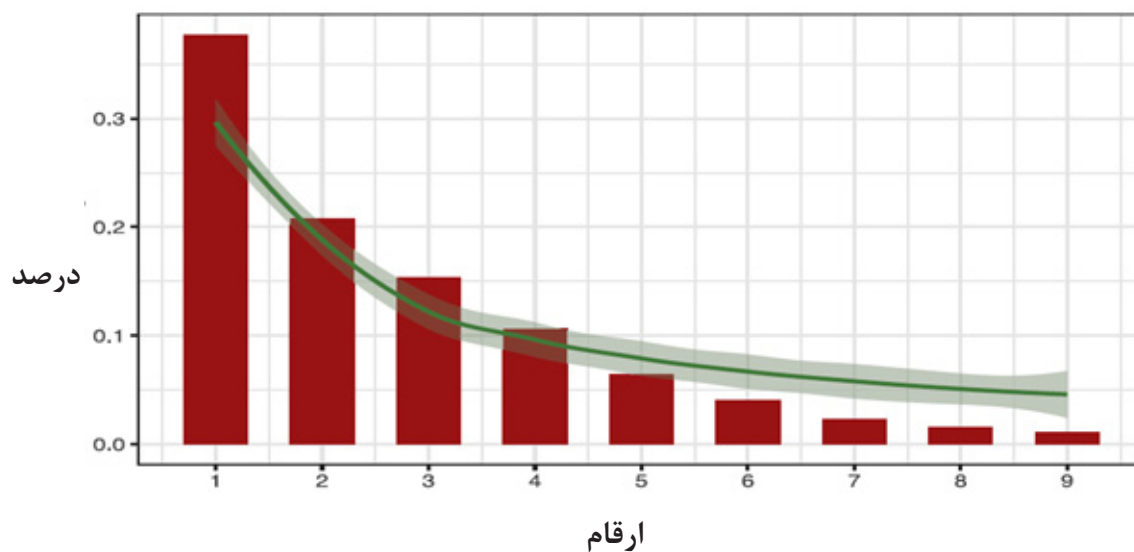
تجزیه و تحلیل داده‌های زمانی با استفاده از توزیع بنفورد قبل (سمبرج، تکالچیچ و جکسون^{۹۰} ۲۰۱۰) نشان داده شده است که در آن نمونه‌ها به فواصل زمانی تقسیم شده و به صورت جداگانه آزمایش شدند. در شکل ۷، توزیع رقمی داده‌های مشابه را به تفکیک دهه نشان داده شده است. مشاهده می‌کنید که از سال ۱۹۶۰ تا ۱۹۹۰ توزیع رقمی مطابقت ندارد با این حال، بعد از دهه ۱۹۹۰ ما شاهد یک ترکیب قوی هستیم. داده‌های ۱۹۶۰-۱۹۷۰ فقط ۳۸ مشاهده دارند که بسیار پایین‌تر از آستانه است. مشارکت‌های منتشر شده بین سال‌های ۱۹۷۰ تا ۱۹۸۰ فقط به ۴۹۳ مشاهده می‌رسد، که نشان می‌دهد حداقل حجم نمونه می‌تواند به طور قابل توجهی در این مورد خاص افزایش یابد.

جدول ۱، نتایج عددی برای آزمون‌های انطباق را نشان می‌دهد. درک کلیدی این است که در بازه‌های زمانی کمتر، تا زمانی که حجم نمونه به اندازه کافی بزرگ است، می‌توان از قانون بنفورد استفاده کرد. علاوه بر این، ما شاهد افزایش قابل توجهی در تعداد مقالات منتشر شده بین سال‌های ۱۹۹۰-۲۰۰۰ هستیم. این احتمالاً نتیجه افزایش دسترسی به اینترنت است که همکاری بین‌المللی را تقویت نموده و منجر به انتشار مقالات بیشتر شده است.

شکل ۲. ساختار مجموعه داده‌های جمع‌آوری شده

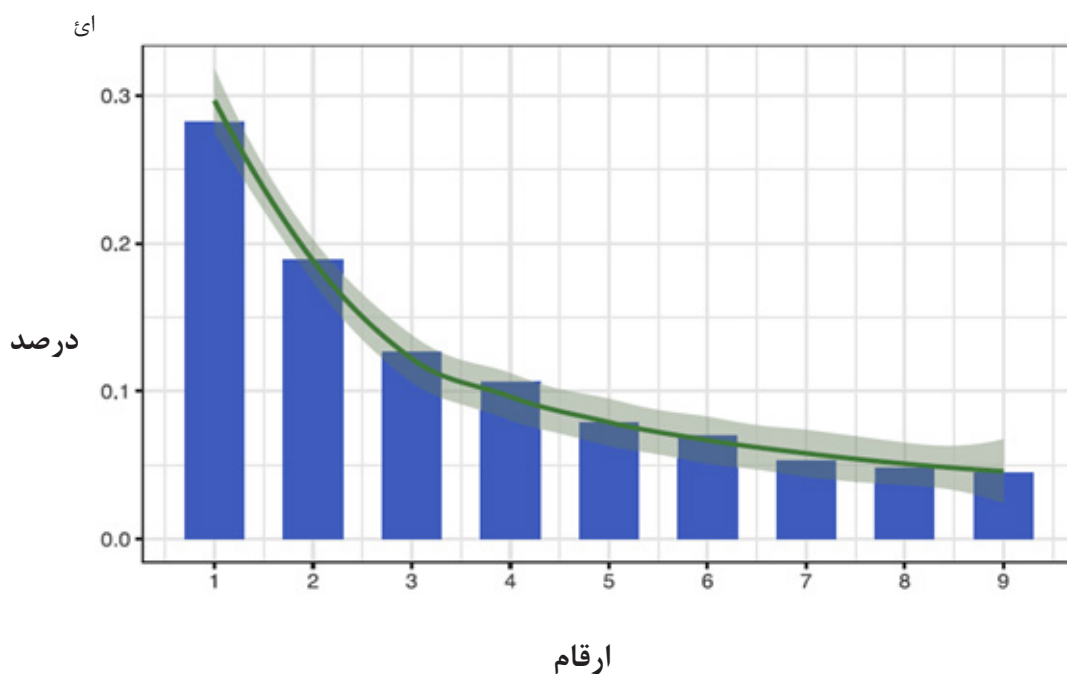


۹۰. Sambridge, Tkalčič, & Jackson

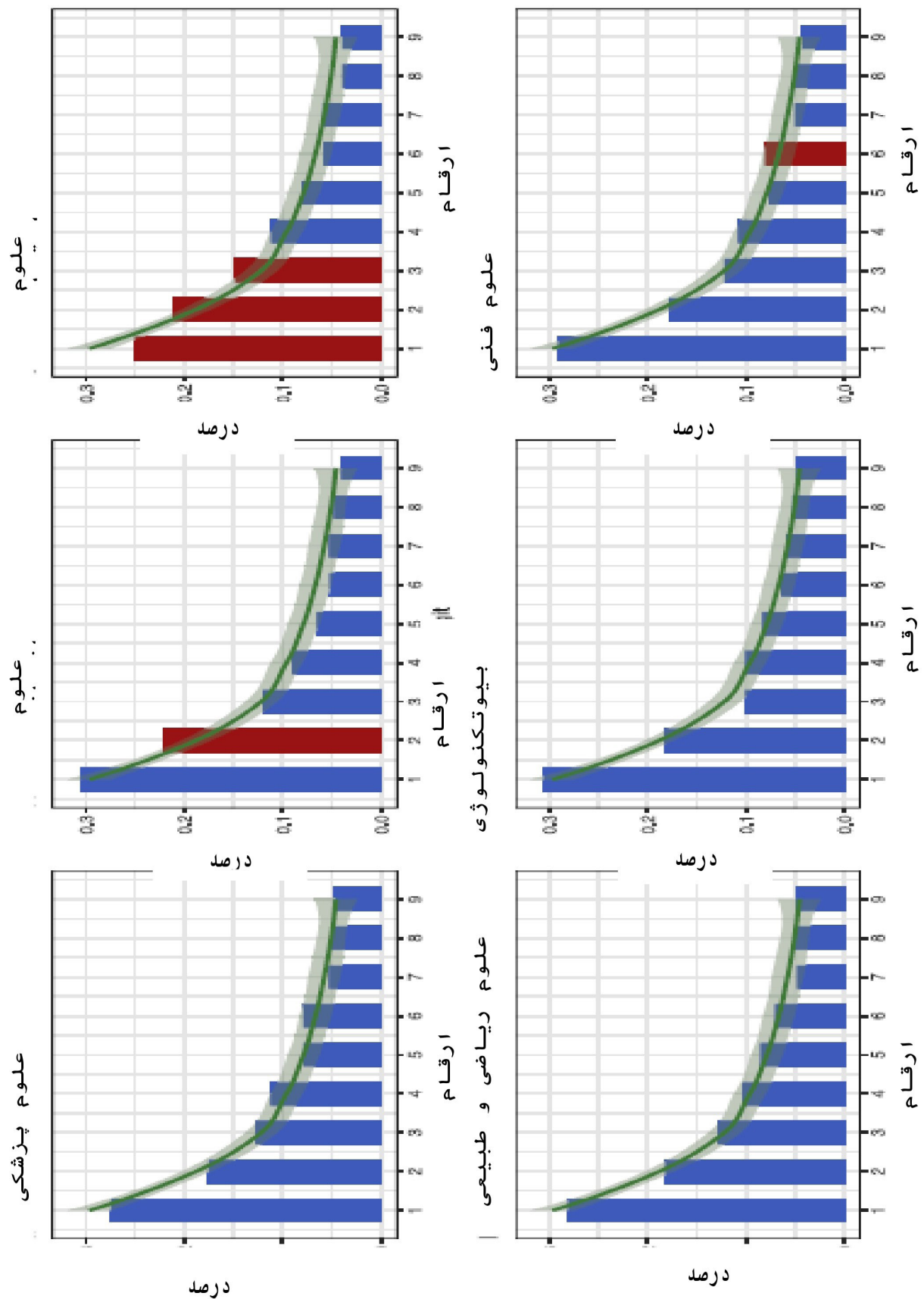


شکل ۳. نمونه‌ای از استفاده نادرست از معیار. نمودار نشان می‌دهد که داده‌های مشاهده شده با قانون بنفورد مطابقت ندارند. داده‌ها با قانون بنفورد مطابقت ندارند، زیرا تعداد نویسندگان در هر انتشار به ندرت از ۱۰ فراتر می‌رود. قانون بنفورد برای داده‌های ارائه شده قابل اجرا نیست.

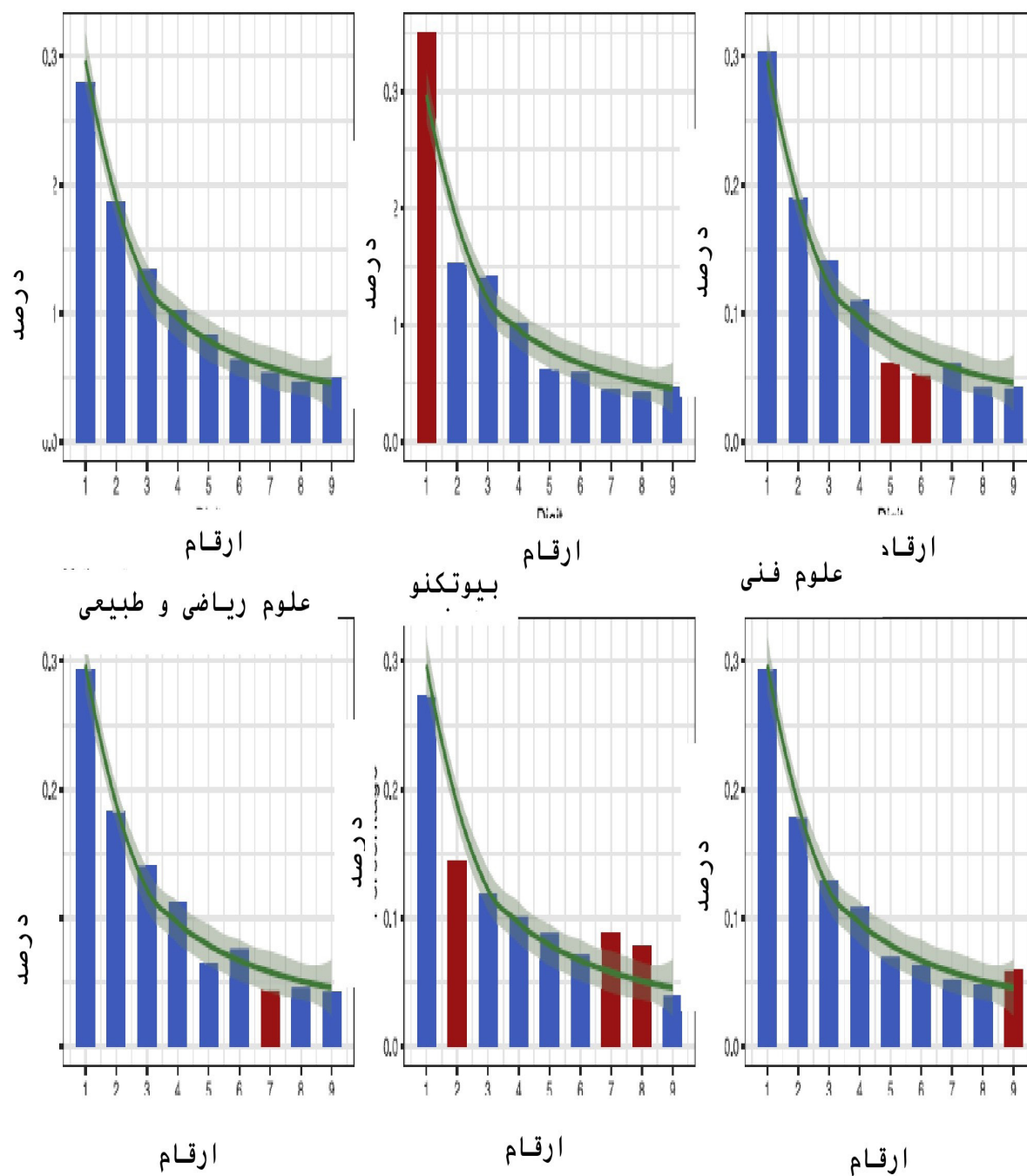
تمام پژوهش‌ها



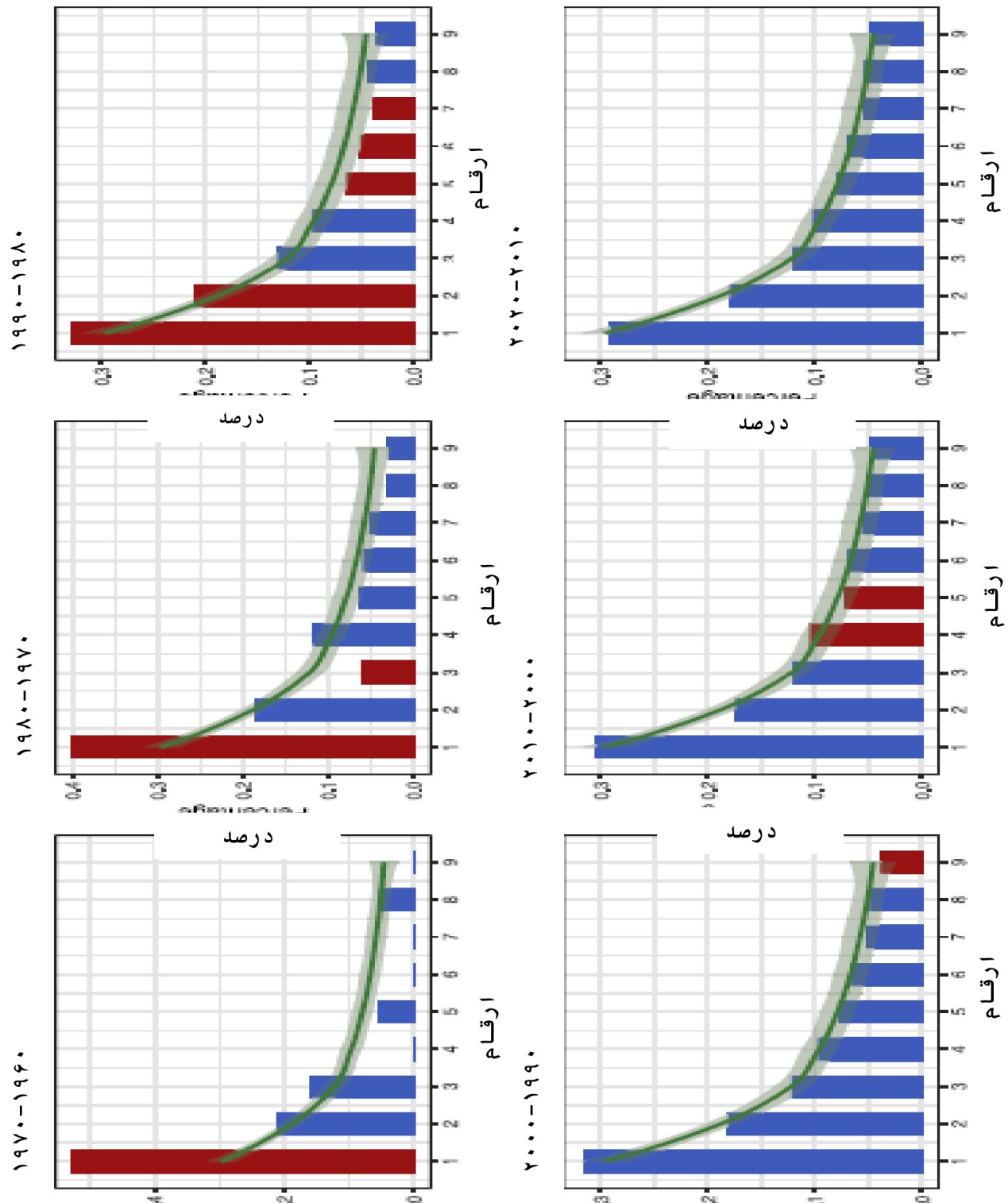
شکل ۴. توزیع ارقام اول برای تعداد مشارکت به ازای هر نویسنده برای کل شبکه علمی اسلوونی با قانون بنفورد مطابقت دارد. انطباق ماد قابل قبول است (نیگرینی، ۲۰۱۲).



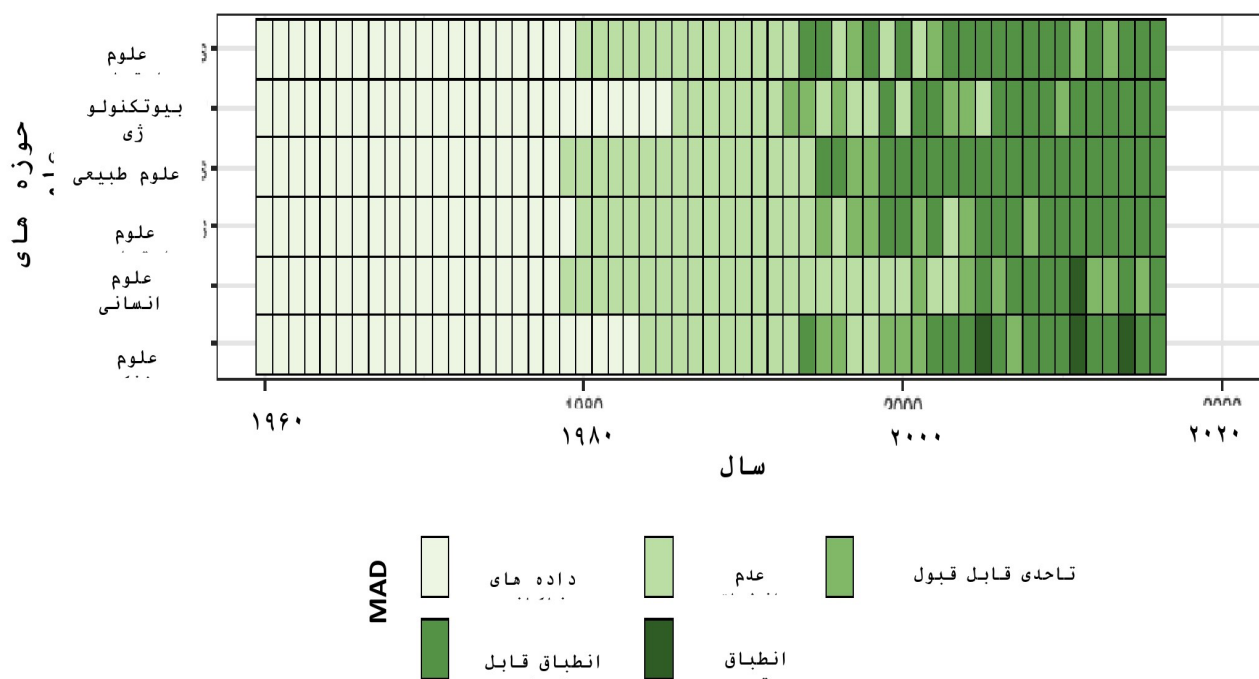
شکل ۵. توزیع اعداد پیشرو در مقالات مجلات (۱،۰۱)



شکل ۶. توزیع اعداد پیشرو در مقالات کنفرانس (۱،۰۸)



شکل ۷. تعداد مقالات به ازای هر نویسنده، گروه بندی شده بر اساس دهه



شکل ۸. انطباق بنفورد در هر حوزه پژوهشی به صورت سالانه

جدول ۱. آزمون‌های انطباق برای تعداد مشارکت به ازای هر نویسنده در هر دهه

مجموعه داده‌ها	میزان مشاهدات	آزمون‌های خی		تست قوس		MAD	انطباق ماد	ضریب
		X-Squared	P-Value	12	P-Value			
۱۹۷۰-۱۹۷۹	۳۸	۵۰۲	۰۲۵۲۹	۱۲۳۷۶	۰۰۹۰۷	۰۶۵۳	عدم	۵/۴۱۳۵۴
۱۹۸۰-۱۹۸۹	۴۰	۳۵/۳۴۴	۲/۳۱E-۰۸	۰۴۶۲۲	۸/۱۲E-	۰۲۸۴	عدم	۴/۲۷۵۴۸
۱۹۹۰-۱۹۹۹	۱۸۹	۴/۵۲۲	۵/۰۵E-۰۸	۰۰۴۷۱۵۴	۰۰۱۲۹	۰۱۴۹	تأیید قابل	۱/۱۹۱۹۷
۲۰۰۰-۲۰۰۹	۵۱۷۶	۱/۵۵۶	۰۹۴۱۱	۰۰۱۱۹۴۲	۰۰۲۰۶۸	۰۰۴۲	انطباق دقیق	۱/۵۰۲۴۲
۲۰۱۰-۲۰۱۹	۸۸۱۹	۱/۸۷۲	۰۸۵۱۵	۰۰۴۰۲۳۳	۰۲۸۷۸	۰۰۳۲	انطباق	۷/۷۳۵۰۷
۲۰۲۰-۲۰۲۹	۹۶۶۰	۱/۹۰۹	۰۸۴۱۸	۰۰۲۴۹۵	۰۸۹۷۸	۰۰۳۹	انطباق	۹/۴۹۴۶۶

جدول ۲. انطباق بنفورد بر اساس رشته و توپولوژی. اختصارات: H = علوم انسانی، SS = علوم اجتماعی، TS = علوم فنی، B = علوم بیوتکنیک، MNS = ریاضیات و علوم طبیعی، و MS = علوم پزشکی.

مجموعه داده‌ها	میزان مشاهدات	آزمون‌های خی		تست قوس		MAD	انطباق ماد	ضریب
		X-Squared	P-Value	12	P-Value			
H (1.01)	۹۷۷	۱۸/۸۴۳	۰/۰۱۵۷۲	/۰۰۲۷۱۱	۰/۰۷۰۷۴	۰/۱۱۳	انطباق	۲/۴۳۴۹۵
H (1.08)	۸۱۷	۱۸/۴۰۶	۰/۰۱۸۳۸	۰۰۷۹۶۶۹	۰۰۱۴۹	۰/۱۵۶	عدم	۴/۱۳۰۳۹
SS (1.01)	۱۴۰۱	۳۸/۳۷۴	۶/۴۲E-۰۶	۰۰۶۷۳۲۹	۸/۰۱E-	۰/۱۷۱	عدم	۱/۳۵۶۲۲
SS (1.08)	۱۳۴۶	۱۹/۸۰۷	۰/۰۱۱۰۹	۰۰۰۶۸۸۲	/۳۹۶	/۰۱۰۴	انطباق	۴/۱۲۵۵۴
TS (1.08)	۲۵۶۷	۱۹/۲۰۳	۰/۰۱۳۸۱	۰۰۰۰۱۲۴۸	۹۶۸۵	/۰۰۶	انطباق	۱/۴۰۰۵۸
BS (1.01)	۷۷۵	۴/۵۹۱۸	۰/۸۰۰۲	۰/۰۰۱۱۳۸۷	/۴۱۳۷	۰/۰۰۵	انطباق	۱/۰۷۰۵۹
BS (1.08)	۷۱۳	۲۹/۰۹۹	۰۰۰۳۰۴۶	/۰۱۲۰۹۷	۰۰۰۱۷۹	۰/۰۱۶	عدم انطباق	/۴۶۹۷۶۱
MNS (1.08)	۱۴۷۷	۱۹/۰۵۸	۰/۰۱۴۴۱	۰۰۰۲۰۸۶۹	۰/۷۳۴۷	۰/۰۱	انطباق قابل	۱/۸۳۴۴۴
MNS (1.01)	۱۹۱۹	۹/۲۶۸۵	۰/۰۳۲۰۲	/۰۰۰۶۴۱۵۹	۰/۲۹۱۹	/۰۰۶۸	انطباق قابل	۱/۲۴۰۶۵
MS (1.01)	۱۵۸۳	۱۲/۰۸۴	۰/۱۴۷۵	۰۰۱۵۵۸۳	/۰۸۴۸۵	/۰۰۷	انطباق قابل	۱/۵۵۷۵۷
MS (1.08)	۱۳۸۸	۶/۰۳۳	۰/۶۴۳۵	۰۰۰۱۵۴۳	۰/۸۰۷۲	۰/۰۰۷	انطباق قابل	۱/۸۹۶۷۲

۸. نتیجه‌گیری و آثار بیشتر

این مقاله روشی را برای پیگیری تغییرات و ارزیابی بلوغ نظام پژوهش با استفاده از قانون بنفورد پیشنهاد می‌کند. این مقاله نوع مجموعه داده‌هایی را که می‌توان برای این منظور استفاده کرد را مشخص می‌کند. روش ارائه شده در یک مورد آزمایشی در دنیای واقعی شبکه تحقیقاتی اسلوونی مورد ارزیابی قرار گرفته است. یافته‌های پژوهش نشان می‌دهد که این روش می‌تواند برای شناسایی موارد ایزوله نشده احتمالی انحرافات مورد استفاده قرار گیرد. این روش به شناسایی گروه‌ها می‌پردازد و روی موارد فردی متمرکز نیست. به منظور تجزیه و تحلیل مشاهدات مشکوک برای شناسایی موارد فردی، به بررسی‌های دقیق‌تری نیاز است. اما دو رویکرد یعنی تجزیه و تحلیل حوزه‌های پژوهش فردی و تجزیه و تحلیل زمانی^{۹۱} برای محدود کردن فضای جستجو در نتیجه، این روش می‌تواند به طور قابل اعتمادی در حوزه‌های پژوهش فردی مورد استفاده قرار گیرد و تفاوت‌هایی در علوم اجتماعی و علوم انسانی نشان دهد که می‌توان به فرهنگ انتشارات حوزه خاص نسبت داد که به روش‌های دقیق‌تری به منظور انطباق نیاز دارد. علاوه بر این، تجزیه و تحلیل زمانی تأیید می‌کند که این روش می‌تواند به طور قابل اعتماد در فواصل ۱۲ ماهه مورد استفاده قرار گیرد، مشروط بر اینکه نمونه بتواند تعداد آستانه مشاهده را برآورده نماید. تجزیه و تحلیل زمانی، به امکان استفاده از گزارش‌های پژوهشی سالانه در نظارت مستمر بر رفتار حوزه‌های پژوهش فردی اشاره دارد.

علاوه بر این، افزایش غیرمنتظره‌ای در تعداد مقالات نوشته‌شده بین سال‌های ۱۹۹۰-۲۰۰۰ مشاهده می‌شود. برای این سال‌ها دو فرضیه احتمالی شناسایی شده است:

- در سال ۱۹۹۱ اسلوونی اولین جمهوری شناخته می‌شد که از یوگسلاوی جدا شد. تغییر در ساختار سیاسی پیامدهای زیادی بر بودجه، ارتقاء و ساختار کلی شبکه علمی داشت.

- در دهه ۹۰، اینترنت به یک رسانه پرکاربرد برای تبادل اطلاعات تبدیل شد. اینترنت این امکان را داد که شکاف‌های فیزیکی قبلی پر شود و امکان همکاری بین‌المللی بیشتری فراهم گردد.

علاوه بر این، این مقاله مثالی ارائه می‌دهد که در آن، این روش بطور عمدی مورد سوء استفاده قرار گرفته است، از این رو، نتایج همراه‌کننده‌ای تولید می‌کند. اهمیت این مسئله، برای ممانعت از شکل‌گیری مثبت کاذب مورد تأکید قرار گرفته است.

با این حال، ما از کمبودهای احتمالی استفاده از نمونه داده‌های کوچک‌تر کاملاً آگاه هستیم و پیشنهاد می‌کنیم این روش در کشورهایی با مجموعه داده‌های بزرگ‌تر تأیید شود. همه داده‌های جمع‌آوری شده (و ناشناس) در زودود^{۹۲} در دسترس هستند: [Dataset] <https://doi.org/10.5281/zenodo.3935770>

این امکان را به پژوهشگران می‌دهد که آزمایش ما را تکرار کنند و از داده‌های ما برای تحقیقات بیشتر استفاده مجدد نمایند.

مایلیم روش پیشنهادی را در آینده، روی یک مورد آزمایشی جدید امتحان کنیم. منتخبی از نامزدهای احتمالی عبارتند از: مخزن ملی فرانسه اچ ای ال^{۹۳}، سیستم اطلاعات پژوهشی فعلی در نروژ (کریستین^{۹۴})، کتابشناسی علمی مجارستان، و سیستم اطلاعاتی پژوهش، توسعه تجربی^{۹۵} و تحقیقات نوآوری جمهوری چک.

نویسندگان همچنین این فرضیه را مطرح می‌کنند که این روش همانطور که پیش بینی شده است بر روی گروهی از نویسندگانی که فراتر از مرزهای ملی هستند، مانند نویسندگانی که در حوزه‌های علمی جهانی از روش‌شناسی مشابهی استفاده می‌کنند، کارایی دارد (زدرافسکی^{۹۶} و همکاران ۲۰۱۹).

مشارکت نویسندگان:

ایده‌پردازی و طراحی تجزیه و تحلیل: الکساندر توسی سی، جرنج ویسیچ

جمع‌آوری داده‌ها: الکساندر توسی سی

تجمیع داده‌ها یا ابزار تحلیل: الکساندر توسی سی، جرنج ویسیچ

انجام تجزیه و تحلیل: الکساندر توسی سی، جرنج ویسیچ

نوشتن مقاله: الکساندر توسی سی، جرنج ویسیچ

سپاسگزاری

نویسندگان از کمیسیون اروپا برای تأمین مالی پروژه اینورینو^{۹۷} (توافقتنامه کمک مالی شماره ۷۳۹۵۷۴) تحت برنامه گروهی گسترده افق ۲۰۲۰ و از جمهوری اسلوونی برای سرمایه‌گذاری، و صندوق توسعه منطقه ای اروپای اتحادیه اروپا قدرانی می‌کنند.

۹۲. Zenodo

۹۳. HAL

۹۴. Cristin

۹۶. Zdravevski

۹۷. InnoRenew

۹۵. توسعه تجربی کار نظام‌مندی است که بر اساس دانش به دست آمده از تحقیقات و یا تجربیات عملی به تولید مواد، محصولات یا دستگاه‌های جدید هدایت می‌شود.

منابع:

- Alexander, J. C. (۲۰۰۹). Remarks on the use of Benford's law. Available at SSRN ۱۵۰۵۱۴۷
- Ariel Xu, Q., & Chang, V. (۲۰۲۰). Co-authorship network and the correlation with academic performance. Internet of Things, ۱۰۰۳۰۷, ۱۲. <https://doi.org/۱۰.۱۰۱۶/j.iot.۲۰۲۰.۱۰۰۳۰۷> <https://www.sciencedirect.com/science/article/pii/S۲۵۴۲۶۶۰۵۲۰۳۰۱۳۹۶>
- Batagelj, V., & Maltseva, D. (۲۰۲۰). Temporal bibliographic networks. Journal of Informetrics, ۱۰۱۰۰۶, ۱۴. <https://doi.org/۱۰.۱۰۱۶/j.joi.۲۰۲۰.۱۰۱۰۰۶> <http://www.sciencedirect.com/science/article/pii/S۱۷۵۱۱۵۷۷۱۹۳۰۱۴۳۹>
- Benavent-Pérez, M., Gorraiz, J., Gumpenberger, C., & de Moya-Anegón, F. (۲۰۱۲). The different flavors of research collaboration: A case study of their influence on university excellence in four world regions. Scientometrics, ۵۸-۴۱, ۹۳.
- Benford, F. (۱۹۳۸). The law of anomalous numbers. Proceedings of the American Philosophical Society, ۵۷۲-۵۵۱.
- Berger, A., & Hill, T. P. (۲۰۱۱). A basic theory of Benford's Law. Probability Surveys, ۱۲۶-۱, ۸. <https://doi.org/۱۰.۱۰۱۱/۱۰۱۲۱۴PS۱۷۰> <http://projecteuclid.org/euclid.ps/۱۳۱۱۸۶۰۸۳۰>
- Bernstein, G. S. (۱۹۸۴). Scientific rigor, scientific integrity: A comment on sommer and sommer. American Psychologist, ۳۹, ۱۳۱۶. <https://doi.org/۱۰.۱۰۳۷X.۳۹.۱۱.۱۳۱۶.a>
- Bhattacharya, S., & Kumar, K. (۲۰۰۸). Forensic Accounting and Benford's Law [In the Spotlight]. IEEE Signal Processing Magazine, ۲۵ <https://doi.org/۱۰.۱۱۰۹/MSP.۱۵۰۰-۱۵۲.۲۰۰۷.۹۱۴۷۲۴>; <http://ieeexplore.ieee.org/document/۴۴۷۲۲۵۸/>
- Bohannon, J. (۲۰۱۳). Who's afraid of peer review? Science, ۶۵-۶۰, ۳۴۲. <https://doi.org/۱۰.۱۱۲۶/science.۳۴۲.۶۱۵۴.۶۰>
- Bordons, M., Gomez, I., Fernández, M., Zulueta, M., & Méndez, A. (۱۹۹۶). Local, domestic and international scientific collaboration in biomedical research. Scientometrics, ۲۹۵-۲۷۹, ۳۷.
- Braun, T., & Glaenzel, W. (۱۹۹۶). International collaboration: Will it be keeping alive east European research? Scientometrics, ۲۵۴-۲۴۷, ۳۶.
- Burke, J., & Kincanon, E. (۱۹۹۱). Benford's law and physical constants: The distribution of initial digits. American Journal of Physics, ۵۹, ۹۵۲. <https://doi.org/۱۰.۱۱۸۳۸/۱۰.۱۱۱۹>
- Cleary, R., & Thibodeau, J. C. (۲۰۰۵). Applying digital analysis using Benford's law to detect fraud: The dangers of type I errors. AUDITING: A Journal of Practice & Theory, ۲۴, ۷۷-۸۱. <https://doi.org/۱۰.۲۳۰۸/aud.۲۰۰۵.۲۴.۱.۷۷>
- Cugmas, M., Ferligoj, A., & Kronegger, L. (۲۰۱۹). Scientific co-authorship networks. Advances in Network Clustering and Blockmodeling, ۳۸۷-۳۶۳.
- Curk, L. (۲۰۱۹). Implementation of the evaluation of researchers' bibliographies in Slovenia. Procedia Computer Science, ۸۳-۷۲, ۱۴۶. <https://doi.org/۱۰.۱۰۱۶/j.procs.۱۴.۲۰۱۹.۰۱.۰۸۲> 14th International Conference on Current Research Information Systems, CRIS۲۰۱۸, FAIRness of Research Information. <http://www.sciencedirect.com/science/article/pii/S۱۸۷۷۰۵۰۹۱۹۳۰۰۸۷۰>
- Drake, P. D., & Nigrini, M. J. (۲۰۰۰). Computer assisted analytical procedures using Benford's law. Journal of Accounting Education, ۱۴۶-۱۲۷, ۱۸.
- Druica, E., Oancea, B., & Vâlsan, C. (۲۰۱۸). Benford's law and the limits of digit analysis. International Journal of Accounting Information Systems, ۸۲-۷۵, ۳۱. <https://doi.org/۱۰.۱۰۱۶/j.accinf.۲۰۱۸.۰۹.۰۰۴> <http://www.sciencedirect.com/science/article/pii/S۱۴۶۷۰۸۹۵۱۷۳۰۱۰۱X>
- Durtschi, C., Hillison, W., & Pacini, C. (۲۰۰۴). The effective use of Benford's law to assist in detecting fraud in accounting data. Journal of Forensic Accounting, ۳۴-۱۷, ۵.
- Edwards, M. A., & Roy, S. (۲۰۱۷). Academic research in the ۲۱st century: Maintaining scientific integrity in a climate of perverse incentives and hypercompetition. Environmental Engineering Science, ۶۱-۵۱, ۳۴.

- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (۲۰۰۸). Comparison of pubmed, scopus, web of science, and google scholar: Strengths and weaknesses. *The FASEB Journal*, ۳۴۲-۳۳۸, ۲۲.
- Fanelli, D. (۲۰۱۳). Redefine misconduct as distorted reporting. *Nature*, ۱۴۹-۱۴۹, ۴۹۴.
- Ferligoj, A., Kronegger, L., Mali, F., Snijders, T. A., & Doreian, P. (۲۰۱۵). Scientific collaboration dynamics in a national scientific system. *Scientometrics*, ۱۰۱۲-۹۸۵, ۱۰۴.
- Fewster, R. M. (۲۰۰۹). A simple explanation of Benford's law. *The American Statistician*, ۳۲-۲۶, ۶۳. <https://doi.org/10.1198/tast.2009.0005>
- Godecharle, S., Nemery, B., & Dierickx, K. (۲۰۱۴). Heterogeneity in european research integrity guidance: Relying on values or norms? *Journal of Empirical. Research on Human Research Ethics*, ۹۰-۷۹, ۹.
- Golbeck, J. (۲۰۱۵). Benford's law applies to online social networks. CoRR abs/۱۵۰۴.۰. arXiv:۱۵۰۴.۰۴۳۸۷
- Heneberg, P. (۲۰۱۶). From excessive journal self-cites to citation stacking: Analysis of journal self-citation kinetics in search for journals, which boost their scientometric indicators. *PLOS ONE*, ۱۱.
- A. Tošić and J. Vičić *Journal of Informetrics* ۱۰۱۱۶۳ (۲۰۲۱) ۱۵
- Van der Heyden, M. A., van de Derks Ven, T., & Ophhof, T. (۲۰۰۹). Fraud and misconduct in science: The stem cell seduction. *Netherlands Heart Journal*, ۲۹-۲۵, ۱۷.
- Hickman, M. J., & Rice, S. K. (۲۰۱۰). Digital analysis of crime statistics: Does crime conform to Benford's law? *Journal of Quantitative Criminology*, ۳۴۹-۳۳۳, ۲۶. <https://doi.org/10.1007/s116-9۰۹۴-۰۱۰-۱۰۹۴۰>
- Holub, H. W., Tappeiner, G., & Eberharter, V. (۱۹۹۱). The iron law of important articles. *Southern Economic Journal*, ۳۲۸-۳۱۷.
- John, L. K., Loewenstein, G., & Prelec, D. (۲۰۱۲). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, ۵۳۲-۵۲۴, ۲۳.
- Karlovćec, M., & Mladenović, D. (۲۰۱۵). Interdisciplinarity of scientific fields and its evolution based on graph of project collaboration and co-authoring. *Scientometrics*, ۹۵۴-۹۳۳, ۱۰۲.
- Katz, J., & Hicks, D. (۱۹۹۷). How much is a collaboration worth? A calibrated bibliometric model. *Scientometrics*, ۴۰۵۴-۵۴۱.
- Keathley-Herring, H., Van Aken, E., Gonzalez-Aleu, F., Deschamps, F., Letens, G., & Orlandini, P. C. (۲۰۱۶). Assessing the maturity of a research area: Bibliometric review and proposed framework. *Scientometrics*, ۹۵۱-۹۲۷, ۱۰۹.
- Kenny, D. A. (۲۰۱۵). Measuring model fit. <http://davidakenny.net/cm/fit.htm>
- Kong, X., Shi, Y., Yu, S., Liu, J., & Xia, F. (۲۰۱۹). Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications*, ۱۰۳-۸۶, ۱۳۲. <https://doi.org/10.1016/j.jnca.2019.01.029> <https://www.sciencedirect.com/science/article/pii/S104848۰۴۵۱۹۳۰۰۴۳۸>
- Korošec, A. (۲۰۱۴). Sicris, v3. Organizacija Znanja, ۲۶-۲۲, ۱۹. <http://www.dlib.si/?URN=URN:NBN:SI:DOC-KY3LXZHR>
- Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (۲۰۰۶). The place of serials in referencing practices: Comparing natural sciences and Engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, ۱۰۰۴-۹۹۷, ۵۷.
- Leydesdorff, L., & Wagner, C. S. (۲۰۰۸). International collaboration in science and the formation of a core group. *Journal of Informetrics*, ۳۲۵-۳۱۷, ۲.
- Lužar, B., Levnajić, Z., Povh, J., & Perc, M. (۲۰۱۴). Community structure and the evolution of interdisciplinarity in Slovenia's scientific collaboration. *Network. PLOS ONE*, ۹, e9۴۴۲۹.

- Shibayama, S., & Baba, Y. (۲۰۱۵). Impact-oriented science policies and scientific publication practices: The case of life sciences in Japan. *Research Policy*, ۹۵۰-۴۴, ۹۳۶.
- Singleton, T. W. (۲۰۱۱). IT audit basics: Understanding and applying Benford's law. *Isaca Journal*, ۶, ۳.
- Sponholz, G. (۲۰۰۰). Teaching scientific integrity and research ethics. *Forensic Science International*, ۵۱۴-۵۱۱, ۱۱۳.
- Zdravevski, E., Lameski, P., Trajkovik, V., Chorbev, I., Goleva, R., Pombo, N., et al. (۲۰۱۹). Automation in systematic, scoping and rapid reviews by an nlp toolkit: A case study in enhanced living environments. *Enhanced living environments*, ۱۸-۱.
- Zhang, J. (۲۰۲۰). Testing case number of coronavirus disease ۲۰۱۹ in china with Newcomb-Benford law. *ArXiv*: ۲۰۰۲, ۰۵۶۹۵