

## **Influential Article Review - Exploring Issues on Academic Systems**

**Monica Valdez**

**Jodi Ingram**

**Clarence Hawkins**

*This paper examines identity. We present insights from a highly influential paper. Here are the highlights from this paper: In biology, last names have been used as a proxy for genetic relatedness in pioneering studies of neutral theory and human migrations. More recently, analyzing the last name distribution of Italian academics has raised the suspicion of nepotism, with faculty hiring their relatives for academic posts. Here, we analyze three large datasets containing the last names of all academics in Italy, researchers from France, and those working at top public institutions in the United States. Through simple randomizations, we show that the US academic system is geographically well-mixed, whereas Italian academics tend to work in their native region. By contrasting maiden and married names, we can detect academic couples in France. Finally, we detect the signature of nepotism in the Italian system, with a declining trend. The claim that our tests detect nepotism as opposed to other effects is supported by the fact that we obtain different results for the researchers hired after 2010, when an anti nepotism law was in effect. For our overseas readers, we then present the insights from this paper in Spanish, French, Portuguese, and German.*

*Keywords: academic systems, isonomy, gender imbalance, nepotism*

### **SUMMARY**

- Here, we have taken an ostensibly meager source of data—a list of names of professors along with their field of research and geographic information—and used elementary randomizations to investigate differences in academic systems. Importantly, we produced a specific randomization for each angle that we wanted to probe, showing that even extremely simple methods can shed light on subtle patterns in the data.
- In Italy, names cluster by city, showing that professors tend to work where they were born. The American system, however, is geographically well-mixed. Even when accounting for geographical and field-specific distribution of last names, Italian academics display an excess of last name sharing within departments. The results of our additional analysis are consistent with the hypothesis of nepotism as testified by the fact that we can detect the effects of an anti nepotism law in effect for the period 2010–2015. Importantly, our analysis shows that nepotism is field- and region-specific and likely driven by a handful of departments. Moreover, after a large increase in the number of faculty between 2000 and 2005, the size of Italian academia has been steadily declining,

with a staggering 10% overall loss during the last decade. The numbers look even worse when examined at the level of regions, fields, or single institutions: Toscana and Liguria lost one-quarter of their faculty, and geology and the humanities have lost a large fraction of their professors. Solving the problem of nepotism by disbanding the university system would be throwing the baby out with the bathwater. Second, anti nepotism laws can have negative side effects, especially when targeting spousal hires.

## HIGHLY INFLUENTIAL ARTICLE

We used the following article as a basis of our evaluation:

Grilli, J., & Allesina, S. (2017). Last name analysis of mobility, gender imbalance, and nepotism across academic systems. *Proceedings of the National Academy of Sciences*, 114(29), 7600–7605.

This is the link to the publisher's website:

<https://www.pnas.org/content/114/29/7600/tab-figures-data>

## INTRODUCTION

Since its inception, science has been a worldwide endeavor, with scholarly publications and conferences connecting researchers across the globe. Despite the many similarities (for example, the organization of scholars into departments and the ubiquitous academic ranks), academic systems around the world are, however, quite distinct in their goals and practices. In many European countries, for example, professors are civil servants, and therefore, their hiring procedures are subject to special regulations. In contrast, American universities have more freedom in choosing their faculty. Salaries, duties, and resources also vary widely both within and between systems.

Here, we examine differences in academic systems using a very simple form of data: a list of names of professors working at a given institution along with their rank, field of study, and geographic location. These data are easy to obtain and can be used to unveil patterns in mobility and immigration (are researchers employed in the region where they were born and raised?), gender imbalance (are women underrepresented in certain fields?), and even nepotism (do professors hire their relatives for academic posts?).

The use of last names as a form of data has a long history in biology, starting with George Darwin (son of Charles), who used the distribution of last names in England to estimate the prevalence of marriages by first cousins (like his parents) (1). Soon dubbed the “poor’s man population genetics” (2), the study of isonymies (occurrences of people with the same name) provided a cheap source of (large) data, with the advantage that last names would well-approximate neutral alleles (2, 3), allowing for the study of human migrations (4). With the advent of modern molecular methods, last names have been associated with Y-chromosome haplotypes (5). More recently, the association of ethnic-specific first and last names has been shown to be predictive of occupational success (6). Closer to the spirit of this work, the distribution of last names in Italian academics has been used to test the hypothesis of nepotistic hires (7, 8): these studies have highlighted a significant scarcity of last names in certain fields and regions, raising the suspicion of nepotistic hires, in which professors recruit relatives for academic positions.

Here, we expand on these results by presenting an international comparison and by introducing specific randomizations that probe different aspects of each academic system. Although our focus is on academia, the same approach could be used in a variety of contexts [for example, in studies of social mobility (9) or health disparities (10)] and even to test whether longevity is related to inbreeding (11).

We analyze last names in three datasets of unprecedented quality and size: all Italian academics in four different years (2000, 2005, 2010, and 2015), researchers currently working at the CNRS in France, and academics working at research-intensive public institutions in the United States. These datasets allow us to track the evolution of last names in time (Italy) and the geographic variability both within and between

countries. Special features of the data allow us to detect the presence of academic couples in France and probe the effects of anti nepotism legislation in Italy.

Results show that the Italian academic system tends to attract researchers mostly at the local level—many researchers have last names that are typical of the region or even the city in which they work—whereas the American system is geographically well-mixed, with a strong influence of immigration. Moreover, in the United States, certain last names are typical of specific scientific fields—meaning that immigration and researchers of given ethnic/cultural backgrounds tend to target predominantly specific areas of research. Using the distribution of first names, we show strong gender imbalance in science, technology, engineering, and mathematics (STEM) disciplines in all systems. Finally, we show that nepotism is present (but declining) in Italy.

## CONCLUSION

Here, we have taken an ostensibly meager source of data—a list of names of professors along with their field of research and geographic information—and used elementary randomizations to investigate differences in academic systems. Importantly, we produced a specific randomization for each angle that we wanted to probe, showing that even extremely simple methods can shed light on subtle patterns in the data.

In Italy, names cluster by city (SI Appendix, Fig. S3), showing that professors tend to work where they were born. The American system, however, is geographically well-mixed (SI Appendix, Fig. S5). The strong signal of immigration is highlighted by the US randomizations, where, for example, physics and mathematics test significantly when randomizing by city but not when randomizing by field: certain names are associated with specific fields, consistent with field-specific immigration and the fact that American researchers of certain heritages tend to target preponderantly science and engineering.

The analysis of married vs. maiden names for the French system shows that our methods can detect the signal of family ties when they are present. Note that, in the Italian system, all women keep their maiden name, whereas in the United States, an unspecified fraction of married women takes their husbands' names—possibly explaining the excess of IPs in pedagogy and other fields. The analysis of first names highlights strong gender imbalance in STEM fields.

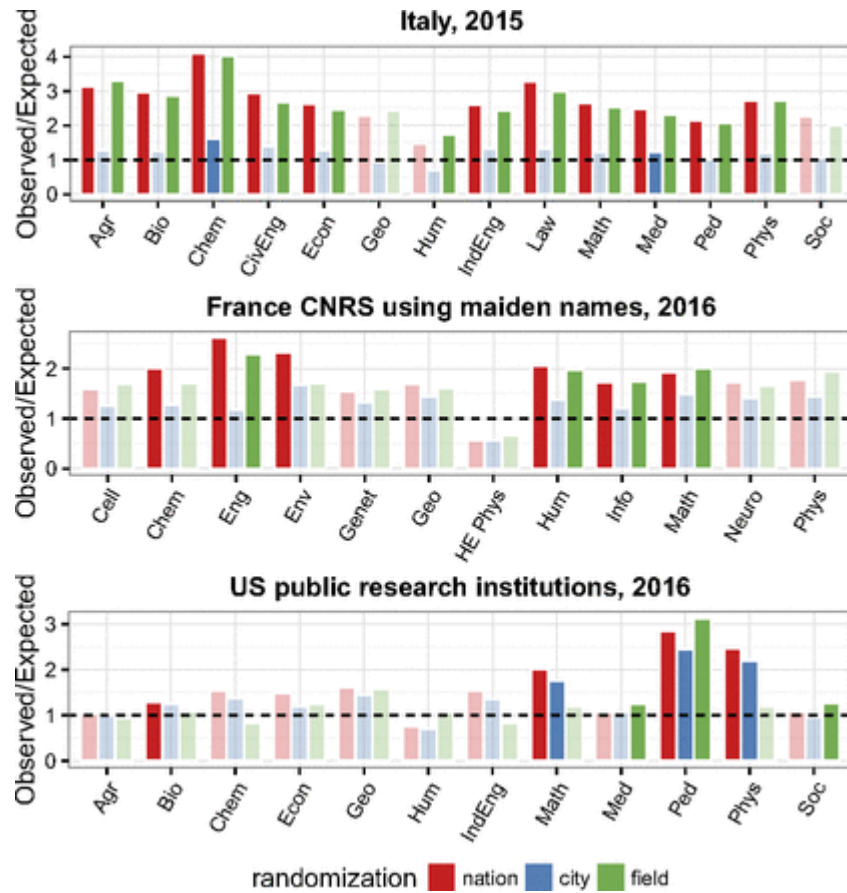
Even when accounting for geographical and field-specific distribution of last names, Italian academics display an excess of last name sharing within departments. The results of our additional analysis are consistent with the hypothesis of nepotism as testified by the fact that we can detect the effects of an anti nepotism law in effect for the period 2010–2015. Importantly, our analysis shows that nepotism is field- and region-specific and likely driven by a handful of departments. For example, when measuring  $\alpha^*$  for the hires in 2005, we found that 10% of departments had an  $\alpha^* \geq 0.1$  (we would expect 2.4% at random), whereas the vast majority of departments had  $\alpha^* \approx 0$ . Similarly, the randomizations in Figs. 1, 2 and 5 show that specific regions and fields drive the results.

For the Italian system, evidence of the efficacy of anti nepotism laws and the fact that the phenomenon seems to be declining should be greeted as good news, with two caveats. First, the decrease in IPs is largely because of retirements: we showed that retirees are more likely to share last names than new hires. Moreover, after a large increase in the number of faculty between 2000 and 2005, the size of Italian academia has been steadily declining, with a staggering 10% overall loss during the last decade. The numbers look even worse when examined at the level of regions, fields, or single institutions (SI Appendix): Toscana and Liguria lost one-quarter of their faculty (Siena, −30.2%; Florence, −29.3%; Genoa, −24.3%), and geology (−21.4%) and the humanities (−18.9%) have lost a large fraction of their professors. Solving the problem of nepotism by disbanding the university system would be throwing the baby out with the bathwater. Second, anti nepotism laws can have negative side effects, especially when targeting spousal hires. For example, in the first half of the 20th century, anti nepotism laws in the United States created the phenomenon of the “vanishing wives” (17): because spouses could not be hired in the same department as their husbands, many women worked as unpaid guests, slowing down the process leading to equal gender representation.

The examples of France, which has hiring procedures that are quite close to those of the Italian system, and the United States, where practices are, however, very different, show that one can build a fair academic system without the need for especially harsh measures. Indeed, many US institutions welcome couples (spousal hires; often extended to domestic partners), although anti nepotism provisions are in place, so that one partner cannot be responsible for the other partner's career advancements.

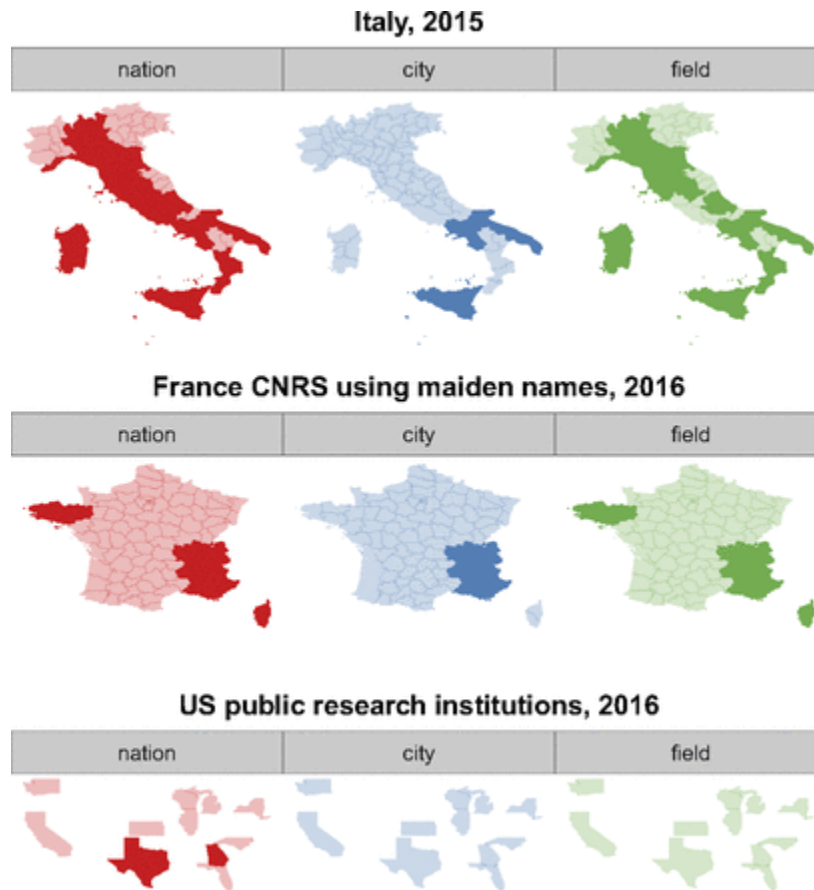
**APPENDIX**

**FIGURE 1**  
**RATIO BETWEEN OBSERVED AND EXPECTED NUMBERS OF IPS FOR EACH ACADEMIC SYSTEM AND FIELD**



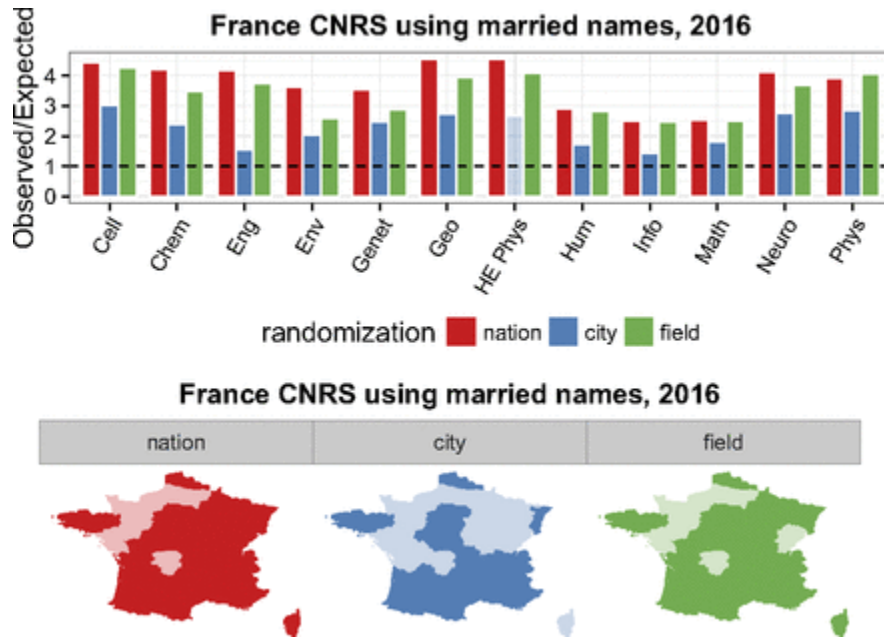
Different colors stand for three randomizations explained in the text; saturated colors mark fields in which the probability of finding a higher or equal number of IPs by chance is  $\leq 0.05$  per number of fields (i.e., significant after applying a Bonferroni correction for multiple hypothesis testing). Agr, agriculture; Bio, biological sciences; Cell, cell and molecular biology; Chem, chemistry and pharmaceutical sciences; CivEng, civil engineering and architecture; Econ, economics and statistics; Eng, engineering; Env, environmental sciences; Genet, genetics; Geo, geology and Earth sciences; HE Phys, high-energy physics; Hum, philology, literature, archeology; IndEng, industrial, electronic, and electronic engineering; Info, information and communications sciences; Law, law; Math, mathematics and computer science; Med, medical sciences; Neuro, neuroscience; Ped, pedagogy, psychology, history, philosophy; Phys, physics and astrophysics; Soc, social and political sciences.

**FIGURE 2**  
**RATIO BETWEEN OBSERVED AND EXPECTED NUMBERS OF IPS FOR EACH REGION**



Saturated colors stand for significantly higher numbers of IPs than expected at random (i.e., P value  $\leq 0.05$  per number of regions).

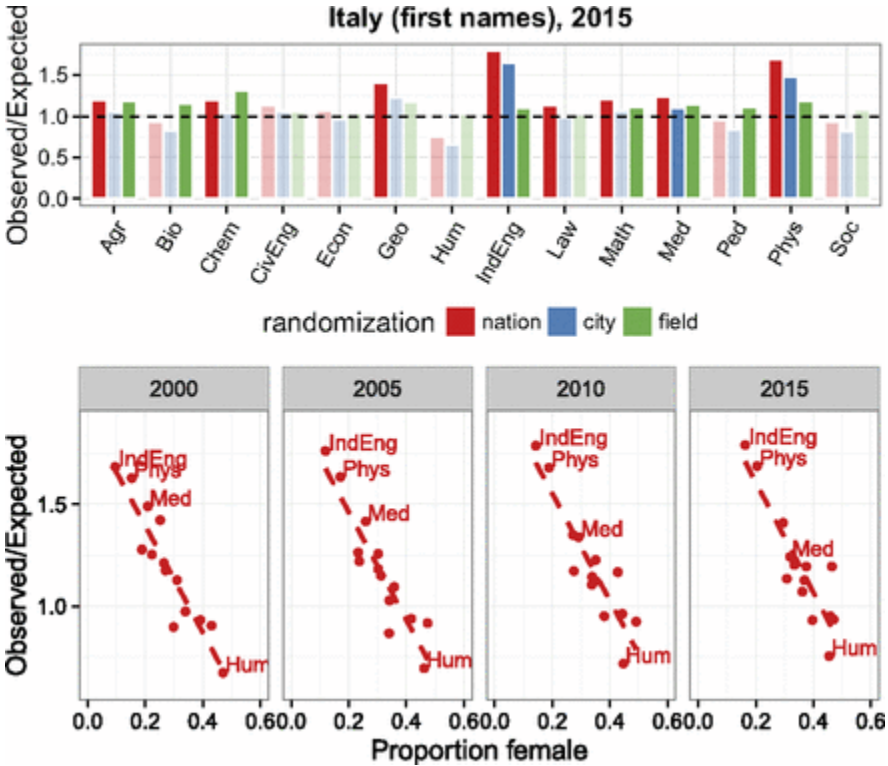
**FIGURE 3**  
**RATIO BETWEEN OBSERVED AND EXPECTED NUMBERS OF IPS USING MARRIED**  
**NAMES INSTEAD OF MAIDEN NAMES**



The large difference in the results is caused by married couples working in the same department. Saturated colors mark significant results once accounted for multiple hypothesis testing. Cell, cell and molecular biology; Chem, chemistry and pharmaceutical sciences; Eng, engineering; Env, environmental sciences; Genet, genetics; Geo, geology and Earth sciences; HE Phys, high-energy physics; Hum, philology, literature, archeology; Info, information and communications sciences; Math, mathematics and computer science; Neuro, neuroscience; Phys, physics and astrophysics.

**FIGURE 3**

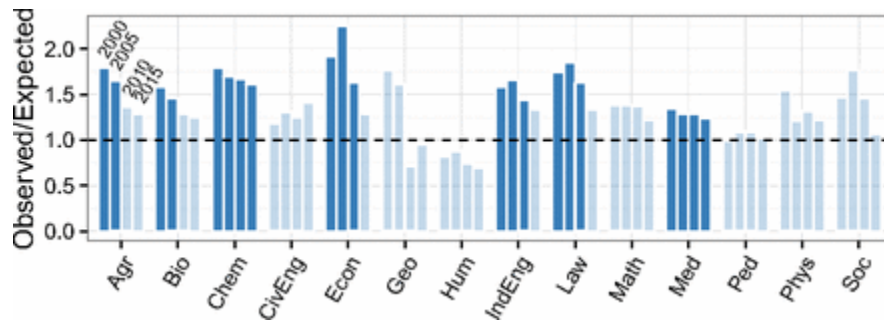
**RATIO BETWEEN OBSERVED AND EXPECTED NUMBERS OF IPS USING FIRST NAMES AND RATIO BETWEEN OBSERVED AND EXPECTED NUMBER OF IPS VS. PROPORTION OF WOMEN FOR ALL YEARS (NATIONAL RANDOMIZATION)**



(Upper) The same as in Fig. 1 but using first names instead of last names. (Lower) Ratio between observed and expected number of IPs vs. proportion of women for all years (national randomization). Some of the fields are highlighted for reference. Saturated colors mark significant results once accounted for multiple hypothesis testing. Agr, agriculture; Bio, biological sciences; Chem, chemistry and pharmaceutical sciences; CivEng, civil engineering and architecture; Econ, economics and statistics; Geo, geology and Earth sciences; Hum, philology, literature, archeology; IndEng, industrial, electronic, and electric engineering; Law, law; Math, mathematics and computer science; Med, medical sciences; Ped, pedagogy, psychology, history, philosophy; Phys, physics and astrophysics; Soc, social and political sciences.

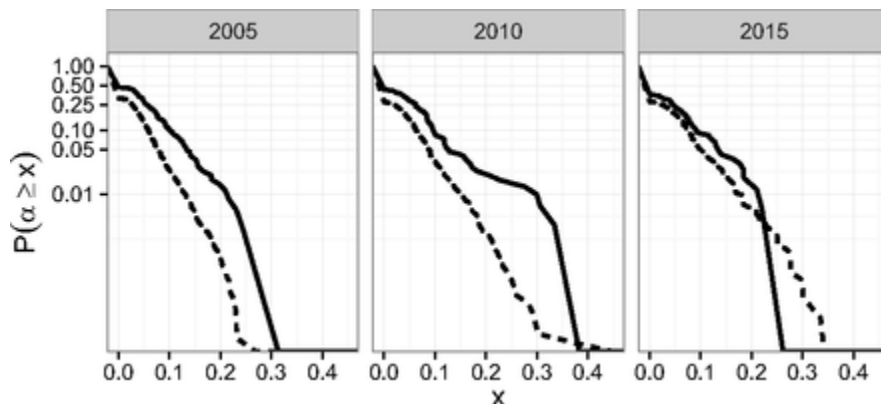
**FIGURE 5**

## EVOLUTION OF THE RATIO BETWEEN OBSERVED AND EXPECTED NUMBER OF IPS IN ITALY BETWEEN 2000 AND 2015



Saturated colors mark significant results once accounted for multiple hypothesis testing. Agr, agriculture; Bio, biological sciences; Chem, chemistry and pharmaceutical sciences; CivEng, civil engineering and architecture; Econ, economics and statistics; Geo, geology and Earth sciences; Hum, philology, literature, archeology; IndEng, industrial, electronic, and electronic engineering; Law, law; Math, mathematics and computer science; Med, medical sciences; Ped, pedagogy, psychology, history, philosophy; Phys, physics and astrophysics; Soc, social and political sciences.

**FIGURE 6**  
CUMULATIVE DISTRIBUTION OF THE MAXIMUM LIKELIHOOD ESTIMATES  $\alpha$



For each department,  $\alpha$  is the maximum likelihood estimate of the probability of sampling new hires from the names already present in the department as opposed to the rest of the city. The solid lines show the distribution of the data, whereas the dashed lines are obtained repeatedly by randomizing the last names of all new hires in a 5-year period. For example, for the hires between 2000 and 2005, we find that 10% of the departments yield a  $\alpha \geq 0.1$ , whereas in the randomizations, we find that only 2.4% of the departments should have such elevated values of  $\alpha$ .

### REFERENCES

- Allesina S (2011) Measuring nepotism through shared last names: The case of Italian academia. *PLoS One* 6: e21160.
- Allesina S (2012) Measuring nepotism through shared last names: Response to Ferlazzo and Sdoia. [arXiv:1208.5792](https://arxiv.org/abs/1208.5792).
- Clark G, Cummins N, Hao Y, Vidal DD (2015) Surnames: A new source for the history of social mobility. *Explor Econ Hist* 55:3–24.
- Crow JF (1983) Surnames as markers of inbreeding and migration. *Discuss Hum Biol* 55:383–397.
- Darwin GH (1875) Marriages between first cousins in England and their effects. *J Stat Soc Lond* 38:153–184.



- Durante R, Labartino G, Perotti R (2011) Academic Dynasties: Decentralization, Civic Capital and Familism in Italian Universities. Working paper 17572 (National Bureau of Economic Research, Cambridge, MA). Available at [www.nber.org/papers/w17572](http://www.nber.org/papers/w17572). Accessed May 22, 2017.
- Elliott MN, et al. (2009) Using the census bureau's surname list to improve estimates of race/ethnicity and associated disparities. *Health Serv Outcome Res Methodol* 9:69–83.
- Ferlazzo F, Sdoia S (2012) Measuring nepotism through shared last names: Are we really moving from opinions to facts? *PLoS One* 7: e43574.
- Goldin C, Shim M (2004) Making a name: Women's surnames at marriage and beyond. *J Econ Perspect* 18:143–160.
- Goldstein JR, Stecklov G (2016) From Patrick to John F.: Ethnic names and occupational success in the last Era of mass migration. *Am Sociol Rev* 81:85–106.
- Jobling MA (2001) In the name of the father: Surnames and genetics. *Trends Genet* 17:353–357.
- Kessler DA, Maruvka YE, Ouren J, Shnerb NM (2012) You name it—how memory and delay govern first name dynamics. *PLoS One* 7: e38790.
- Lan F, Hale K, Rivers E (2015) Immigrants' Growing Presence in the U.S. Science and Engineering Workforce: Education and Employment Characteristics in 2013 (NSF InfoBriefs, Arlington, VA), pp 15–328.
- Lykknes A, Opitz DL, Van Tiggelen B, eds (2012) For Better or for Worse? Collaborative Couples in the Sciences (Springer Science & Business Media, Basel), Vol 44.
- Montesanto A, Passarino G, Senatore A, Carotenuto L, De Benedictis G (2008) Spatial analysis and surname analysis: Complementary tools for shedding light on human longevity patterns. *Ann Hum Genet* 72:253–260.
- Piazza A, Rendine S, Zei G, Moroni A, Cavalli-Sforza LL (1987) Migration rates of human populations from surname distributions. *Nature* 329:714–716.
- Zei G, Guglielmino CR, Siri E, Moroni A, Cavalli-Sforza LL (1983) Surnames as neutral alleles: Observations in Sardinia. *Hum Biol* 55:357–365.

## **TRANSLATED VERSION: SPANISH**

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

## **VERSION TRADUCIDA: ESPAÑOL**

A continuación se muestra una traducción aproximada de las ideas presentadas anteriormente. Esto se hizo para dar una comprensión general de las ideas presentadas en el documento. Por favor, disculpe cualquier error gramatical y no responsabilite a los autores originales de estos errores.

## **INTRODUCCIÓN**

Desde sus inicios, la ciencia ha sido un esfuerzo mundial, con publicaciones académicas y conferencias que conectan a investigadores de todo el mundo. A pesar de las muchas similitudes (por ejemplo, la organización de académicos en departamentos y los rangos académicos omnipresentes), los sistemas académicos de todo el mundo son, sin embargo, bastante distintos en sus objetivos y prácticas. En muchos países europeos, por ejemplo, los profesores son funcionarios públicos y, por lo tanto, sus procedimientos de contratación están sujetos a regulaciones especiales. Por el contrario, las universidades estadounidenses tienen más libertad a la hora de elegir su facultad. Los salarios, deberes y recursos también varían ampliamente tanto dentro como entre los sistemas.

Aquí, examinamos las diferencias en los sistemas académicos utilizando una forma muy simple de datos: una lista de nombres de profesores que trabajan en una institución determinada junto con su rango,

campo de estudio y ubicación geográfica. Estos datos son fáciles de obtener y se pueden utilizar para revelar patrones en movilidad e inmigración (¿están empleados investigadores en la región donde nacieron y se criaron?), desequilibrio de género (¿están las mujeres subrepresentadas en ciertos campos?), e incluso el nepotismo (¿contratan profesores a sus familiares para puestos académicos?).

El uso de apellidos como forma de datos tiene una larga historia en biología, comenzando con George Darwin (hijo de Charles), quien utilizó la distribución de apellidos en Inglaterra para estimar la prevalencia de matrimonios por primos primos (como sus padres) (1). Pronto apodado la "genética de la población del hombre pobre" (2), el estudio de los isonymies (ocurrencias de personas con el mismo nombre) proporcionó una fuente barata de datos (grandes), con la ventaja de que los apellidos se aproximarían bien a los alelos neutros (2, 3), permitiendo el estudio de las migraciones humanas (4). Con el advenimiento de los métodos moleculares modernos, los apellidos se han asociado con los haplotipos del cromosoma Y (5). Más recientemente, se ha demostrado que la asociación de nombres y apellidos específicos de las etnias es predictiva del éxito profesional (6). Más cerca del espíritu de este trabajo, la distribución de apellidos en los académicos italianos se ha utilizado para probar la hipótesis de las contrataciones nepotistas (7, 8): estos estudios han puesto de relieve una escasez significativa de apellidos en ciertos campos y regiones, elevando la sospecha de contrataciones nepotistas, en las que los profesores reclutan familiares para puestos académicos.

Aquí, ampliamos estos resultados presentando una comparación internacional e introduciendo aleatorizaciones específicas que sondea diferentes aspectos de cada sistema académico. Aunque nos centramos en la academia, el mismo enfoque podría utilizarse en una variedad de contextos [por ejemplo, en estudios de movilidad social (9) o disparidades de salud (10)] e incluso para probar si la longevidad está relacionada con la endogamia (11).

Analizamos los apellidos en tres conjuntos de datos de calidad y tamaño sin precedentes: todos los académicos italianos en cuatro años diferentes (2000, 2005, 2010 y 2015), investigadores que actualmente trabajan en el CNRS en Francia, y académicos que trabajan en instituciones públicas intensivas en investigación en los Estados Unidos. Estos conjuntos de datos nos permiten realizar un seguimiento de la evolución de los apellidos en el tiempo (Italia) y la variabilidad geográfica tanto dentro como entre países. Las características especiales de los datos nos permiten detectar la presencia de parejas académicas en Francia y sondear los efectos de la legislación de antinepotismo en Italia.

Los resultados muestran que el sistema académico italiano tiende a atraer investigadores principalmente a nivel local —muchos investigadores tienen apellidos típicos de la región o incluso de la ciudad en la que trabajan—, mientras que el sistema estadounidense está geográficamente bien mezclado, con una fuerte influencia de la inmigración. Además, en los Estados Unidos, ciertos apellidos son típicos de campos científicos específicos, lo que significa que la inmigración y los investigadores de determinados orígenes étnicos/culturales tienden a dirigirse a áreas de investigación específicas previamente. Utilizando la distribución de nombres, mostramos un fuerte desequilibrio de género en las disciplinas de ciencia, tecnología, ingeniería y matemáticas (STEM) en todos los sistemas. Por último, demostramos que el nepotismo está presente (pero disminuyendo) en Italia.

## CONCLUSIÓN

Aquí, hemos tomado una fuente de datos ostensiblemente escasa (una lista de nombres de profesores junto con su campo de investigación e información geográfica) y hemos utilizado aleatorizaciones elementales para investigar las diferencias en los sistemas académicos. Es importante destacar que producimos una aleatorización específica para cada ángulo que queríamos sondear, mostrando que incluso los métodos extremadamente simples pueden arrojar luz sobre patrones sutiles en los datos.

En Italia, los nombres se agrupan por ciudad (Apéndice SI, Fig. S3), mostrando que los profesores tienden a trabajar donde nacieron. El sistema americano, sin embargo, está geográficamente bien mezclado (Apéndice SI, Fig. S5). La fuerte señal de inmigración es resaltada por las aleatorizaciones de los Estados Unidos, donde, por ejemplo, la física y las matemáticas prueban significativamente cuando se aleatoriza por ciudad, pero no al aleatorizar por campo: ciertos nombres están asociados con campos específicos,

consistentes con la inmigración específica del campo y el hecho de que los investigadores estadounidenses de ciertas herencias tienden a apuntar preponderantemente a la ciencia y la ingeniería.

El análisis de los nombres de casada frente a doncellas para el sistema francés muestra que nuestros métodos pueden detectar la señal de los lazos familiares cuando están presentes. Tenga en cuenta que, en el sistema italiano, todas las mujeres mantienen su nombre de soltera, mientras que en los Estados Unidos, una fracción no especificada de mujeres casadas toma los nombres de sus maridos, posiblemente explicando el exceso de ips en la pedagogía y otros campos. El análisis de los primeros nombres pone de relieve un fuerte desequilibrio de género en los campos STEM.

Incluso cuando se contabiliza la distribución geográfica y específica de los apellidos, los académicos italianos muestran un exceso de intercambio de apellidos dentro de los departamentos. Los resultados de nuestro análisis adicional son consistentes con la hipótesis del nepotismo como lo demuestra el hecho de que podemos detectar los efectos de una ley de antinepotismo vigente para el período 2010-2015. Es importante destacar que nuestro análisis muestra que el nepotismo es específico del campo y de la región y probablemente impulsado por un puñado de departamentos. Por ejemplo, al medir el valor de las contrataciones en 2005, encontramos que el 10% de los departamentos tenían un 0,1 euros (esperaríamos un 2,4% al azar), mientras que la gran mayoría de los departamentos tenían un 0% . Del mismo modo, las aleatorizaciones de las Figs. 1, 2 y 5 muestran que regiones y campos específicos impulsan los resultados.

Para el sistema italiano, la evidencia de la eficacia de las leyes de antinepotismo y el hecho de que el fenómeno parece estar disminuyendo debe ser recibida como una buena noticia, con dos advertencias. En primer lugar, la disminución de las direcciones IP se debe en gran medida a las jubilaciones: mostramos que los jubilados son más propensos a compartir apellidos que las nuevas contrataciones. Además, después de un gran aumento en el número de profesores entre 2000 y 2005, el tamaño de la academia italiana ha ido disminuyendo constantemente, con una asombrosa pérdida global del 10% durante la última década. Las cifras se ven aún peor cuando se examinan a nivel de regiones, campos o instituciones individuales (Apéndice SI): Toscana y Liguria perdieron una cuarta parte de su facultad (Siena, 30,2%; Florencia, 29,3%; Génova, 24,3%), y geología (21,4%) y las humanidades (18,9%) han perdido una gran fracción de sus profesores. Resolver el problema del nepotismo disolviendo el sistema universitario sería tirar al bebé con el agua del baño. En segundo lugar, las leyes de antinepotismo pueden tener efectos secundarios negativos, especialmente cuando se dirigen a las contrataciones conponsales. Por ejemplo, en la primera mitad del siglo XX, las leyes de antinepotismo en los Estados Unidos crearon el fenómeno de las "esposas desaparecidas" (17): debido a que los cónyuges no podían ser contratados en el mismo departamento que sus maridos, muchas mujeres trabajaban como invitados no remunerados, ralentizando el proceso que conducía a la misma representación de género.

Los ejemplos de Francia, que tiene procedimientos de contratación muy cercanos a los del sistema italiano, y los Estados Unidos, donde las prácticas son, sin embargo, muy diferentes, muestran que se puede construir un sistema académico justo sin la necesidad de medidas especialmente duras. De hecho, muchas instituciones estadounidenses dan la bienvenida a las parejas (contrataciones con cónyuges, a menudo extendidas a parejas domésticas), aunque existen disposiciones sobre antinepotismo, por lo que una pareja no puede ser responsable de los avances profesionales del otro socio.

## **TRANSLATED VERSION: FRENCH**

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

## **VERSION TRADUITE: FRANÇAIS**

Voici une traduction approximative des idées présentées ci-dessus. Cela a été fait pour donner une compréhension générale des idées présentées dans le document. Veuillez excuser toutes les erreurs grammaticales et ne pas tenir les auteurs originaux responsables de ces erreurs.

## INTRODUCTION

Depuis sa création, la science est une entreprise mondiale, avec des publications savantes et des conférences reliant des chercheurs à travers le monde. Malgré les nombreuses similitudes (par exemple, l'organisation des chercheurs dans les départements et les rangs universitaires omniprésents), les systèmes universitaires du monde entier sont toutefois très distincts dans leurs objectifs et leurs pratiques. Dans de nombreux pays européens, par exemple, les professeurs sont des fonctionnaires et, par conséquent, leurs procédures d'embauche sont soumises à des réglementations spéciales. En revanche, les universités américaines ont plus de liberté dans le choix de leur faculté. Les salaires, les tâches et les ressources varient également considérablement à l'intérieur et à l'extérieur des systèmes.

Ici, nous examinons les différences dans les systèmes académiques en utilisant une forme très simple de données: une liste de noms de professeurs travaillant dans un établissement donné avec leur rang, domaine d'études, et l'emplacement géographique. Ces données sont faciles à obtenir et peuvent être utilisées pour dévoiler les modèles de mobilité et d'immigration (les chercheurs sont-ils employés dans la région où ils sont nés et élevés ?), le déséquilibre entre les sexes (les femmes sont-elles sous-représentées dans certains domaines ?) Et même le népotisme (les professeurs embauchent-ils leurs proches pour des postes universitaires ?).

L'utilisation des noms de famille comme une forme de données a une longue histoire en biologie, à commencer par George Darwin (fils de Charles), qui a utilisé la distribution des noms de famille en Angleterre pour estimer la prévalence des mariages par des cousins germains (comme ses parents) (1). Bientôt surnommée la « génétique de la population humaine des pauvres » (2), l'étude des isonymies (occurrences de personnes du même nom) a fourni une source bon marché de (grandes) données, avec l'avantage que les noms de famille seraient bien approximativement allèles neutres (2, 3), permettant l'étude des migrations humaines (4). Avec l'avènement des méthodes moléculaires modernes, les noms de famille ont été associés aux haplotypes du chromosome Y (5). Plus récemment, il a été démontré que l'association des prénoms et des noms de famille spécifiques à l'ethnie était prédictive de la réussite professionnelle (6). Plus proche de l'esprit de ce travail, la répartition des noms de famille dans les universitaires italiens a été utilisée pour tester l'hypothèse des embauches népotistes (7, 8): ces études ont mis en évidence une pénurie significative de noms de famille dans certains domaines et régions, ce qui soulève la suspicion d'embauches népotistes, dans lesquelles les professeurs recrutent des parents pour des postes universitaires.

Ici, nous élargissons ces résultats en présentant une comparaison internationale et en introduisant des randomisations spécifiques qui sondent différents aspects de chaque système académique. Bien que nous nous concentrons sur le milieu universitaire, la même approche pourrait être utilisée dans divers contextes [par exemple, dans les études sur la mobilité sociale (9) ou les disparités en matière de santé (10)] et même pour vérifier si la longévité est liée à la consanguinité (11).

Nous analysons les noms de famille dans trois ensembles de données d'une qualité et d'une taille sans précédent : tous les universitaires italiens en quatre années différentes (2000, 2005, 2010 et 2015), les chercheurs travaillant actuellement au CNRS en France, et les universitaires travaillant dans des institutions publiques à forte intensité de recherche aux Etats-Unis. Ces ensembles de données nous permettent de suivre l'évolution des noms de famille dans le temps (Italie) et la variabilité géographique à l'intérieur et entre les pays. Les particularités des données nous permettent de détecter la présence de couples universitaires en France et de sonder les effets de la législation antinépotisme en Italie.

Les résultats montrent que le système universitaire italien tend à attirer des chercheurs principalement au niveau local - de nombreux chercheurs ont des noms de famille qui sont typiques de la région ou même la ville dans laquelle ils travaillent - alors que le système américain est géographiquement bien mélangé, avec une forte influence de l'immigration. De plus, aux États-Unis, certains noms de famille sont typiques de domaines scientifiques spécifiques, ce qui signifie que l'immigration et les chercheurs d'origines

ethniques et culturelles ont tendance à cibler des domaines de recherche prépondérants spécifiques. En utilisant la distribution des prénoms, nous montrons un fort déséquilibre entre les sexes dans les disciplines des sciences, de la technologie, de l'ingénierie et des mathématiques (STEM) dans tous les systèmes. Enfin, nous montrons que le népotisme est présent (mais déclinant) en Italie.

## CONCLUSION

Ici, nous avons pris une source apparemment maigre de données — une liste de noms de professeurs avec leur domaine de recherche et d'information géographique — et utilisé des randomisations élémentaires pour étudier les différences dans les systèmes universitaires. Fait important, nous avons produit une randomisation spécifique pour chaque angle que nous voulions sonder, montrant que même des méthodes extrêmement simples peuvent faire la lumière sur les modèles subtils dans les données.

En Italie, les noms groupe par ville (SI Annexe, Fig. S3), montrant que les professeurs ont tendance à travailler là où ils sont nés. Le système américain, cependant, est géographiquement bien mélangé (SI Annexe, Fig. S5). Le signal fort de l'immigration est mis en évidence par les randomisations américaines, où, par exemple, la physique et les mathématiques testent de manière significative lors de la randomisation par ville, mais pas lors du randomisation par domaine : certains noms sont associés à des domaines spécifiques, compatibles avec l'immigration spécifique au terrain et le fait que les chercheurs américains de certains héritages ont tendance à cibler la science et l'ingénierie de façon prépondérante.

L'analyse des noms de famille mariés ou de jeunes filles pour le système Français montre que nos méthodes peuvent détecter le signal des liens familiaux lorsqu'ils sont présents. Notez que, dans le système italien, toutes les femmes conservent leur nom de jeune fille, alors qu'aux États-Unis, une fraction non précisée des femmes mariées prend le nom de leur mari, ce qui explique peut-être l'excès de IP dans la pédagogie et d'autres domaines. L'analyse des prénoms met en évidence un fort déséquilibre entre les sexes dans les domaines stem.

Même lorsqu'ils expliquent la répartition géographique et spécifique des noms de famille, les universitaires italiens affichent un excès de partage de noms de famille au sein des départements. Les résultats de notre analyse supplémentaire sont compatibles avec l'hypothèse du népotisme, comme en témoigne le fait que nous pouvons détecter les effets d'une loi sur l'antinépotisme en vigueur pour la période 2010-2015. Fait important, notre analyse montre que le népotisme est spécifique au terrain et à la région et probablement dirigé par une poignée de ministères. Par exemple, lorsque nous mesurons  $\alpha^*$  pour les embauches en 2005, nous avons constaté que 10 % des départements avaient un  $\alpha^* \geq 0,1$  (nous nous attendions à 2,4 % au hasard), alors que la grande majorité des ministères avaient  $\alpha^* < 0$ . De même, les randomisations dans les figures 1, 2 et 5 montrent que des régions et des champs spécifiques sont à l'origine des résultats.

Pour le système italien, la preuve de l'efficacité des lois sur l'antinépotisme et le fait que le phénomène semble décliner devrait être accueillie comme une bonne nouvelle, avec deux mises en garde. Premièrement, la diminution des IP est en grande partie attribuable aux départs à la retraite : nous avons montré que les retraités sont plus susceptibles de partager leur nom de famille que les nouveaux employés. En outre, après une forte augmentation du nombre de professeurs entre 2000 et 2005, la taille du milieu universitaire italien n'a cessé de diminuer, avec une perte globale stupéfiante de 10 % au cours de la dernière décennie. Les chiffres semblent encore pires lorsqu'ils sont examinés au niveau des régions, des champs ou des institutions individuelles (Si Annexe): Toscana et la Ligurie ont perdu un quart de leur corps professoral (Sienne, -30,2%; Florence, -29,3 %; Gênes, -24,3 %) et géologie (-21,4 %) et les sciences humaines (-18,9%) ont perdu une grande partie de leurs professeurs. Résoudre le problème du népotisme en dissolvant le système universitaire serait jeter le bébé avec l'eau du bain. Deuxièmement, les lois sur l'antinépotisme peuvent avoir des effets secondaires négatifs, surtout lorsqu'elles ciblent les embauches de conjoints. Par exemple, dans la première moitié du xxe siècle, les lois sur l'antinépotisme aux États-Unis ont créé le phénomène des « épouses en voie de disparition » (17) : parce que les conjoints ne pouvaient pas être embauchés dans le même département que leurs maris, de nombreuses femmes travaillaient comme invitées non rémunérées, ralentissant ainsi le processus menant à une représentation égale entre les sexes.

Les exemples de la France, qui a des procédures d'embauche assez proches de celles du système italien, et des États-Unis, où les pratiques sont cependant très différentes, montrent que l'on peut construire un système académique équitable sans avoir besoin de mesures particulièrement sévères. En effet, de nombreuses institutions américaines accueillent les couples (embauches de conjoints ; souvent étendus à des partenaires domestiques), bien que des dispositions en matière d'antinépotisme soient en place, de sorte qu'un partenaire ne peut pas être responsable des avancements professionnels de l'autre partenaire.

## **TRANSLATED VERSION: GERMAN**

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

## **ÜBERSETZTE VERSION: DEUTSCH**

Hier ist eine ungefähre Übersetzung der oben vorgestellten Ideen. Dies wurde getan, um ein allgemeines Verständnis der in dem Dokument vorgestellten Ideen zu vermitteln. Bitte entschuldigen Sie alle grammatikalischen Fehler und machen Sie die ursprünglichen Autoren nicht für diese Fehler verantwortlich.

## **EINLEITUNG**

Seit ihrer Gründung ist die Wissenschaft ein weltweites Unterfangen, mit wissenschaftlichen Publikationen und Konferenzen, die Forscher auf der ganzen Welt verbinden. Trotz der vielen Ähnlichkeiten (z. B. Die Organisation von Gelehrten in Abteilungen und die allgegenwärtigen akademischen Ränge) sind akademische Systeme auf der ganzen Welt jedoch in ihren Zielen und Praktiken recht unterschiedlich. In vielen europäischen Ländern sind beispielsweise Professoren Beamte, und daher unterliegen ihre Einstellungsverfahren besonderen Vorschriften. Im Gegensatz dazu haben amerikanische Universitäten mehr Freiheit bei der Auswahl ihrer Fakultät. Auch die Gehälter, Zölle und Ressourcen sind sowohl innerhalb als auch zwischen den Systemen sehr unterschiedlich.

Hier untersuchen wir Unterschiede in akademischen Systemen anhand einer sehr einfachen Form von Daten: eine Liste von Namen von Professoren, die an einer bestimmten Institution arbeiten, zusammen mit ihrem Rang, Ihrem Studienbereich und ihrem geografischen Standort. Diese Daten sind leicht zu erhalten und können verwendet werden, um Muster in Mobilität und Einwanderung zu enthüllen (sind Forscher in der Region beschäftigt, in der sie geboren und aufgewachsen sind?), Geschlechterungleichgewicht (sind Frauen in bestimmten Bereichen unterrepräsentiert?) Und sogar Vetternwirtschafts (stellen Professoren ihre Verwandten für akademische Stellen ein?).

Die Verwendung von Nachnamen als eine Form von Daten hat eine lange Geschichte in der Biologie, beginnend mit George Darwin (Sohn von Charles), der die Verteilung von Nachnamen in England nutzte, um die Prävalenz von Ehen durch erste Cousins (wie seine Eltern) zu schätzen (1). Bald als "Arme-Mann-Bevölkerung Genetik" (2) bezeichnet, lieferte die Untersuchung von Isonymien (Vorkommen von Menschen mit dem gleichen Namen) eine billige Quelle von (großen) Daten, mit dem Vorteil, dass Nachnamen neutrale Allele gut annähern würden (2, 3), was die Untersuchung menschlicher Migrationen (4) ermöglicht. Mit dem Aufkommen moderner molekularer Methoden wurden Nachnamen mit Y-Chromosom-Haplotypen assoziiert (5). In jüngerer Zeit hat sich gezeigt, dass die Assoziation ethnisch-spezifischer Vor- und Nachnamen den beruflichen Erfolg vorhersagt (6). Näher am Geist dieser Arbeit, die Verteilung der Nachnamen in italienischen Akademikern wurde verwendet, um die Hypothese der nepotistischen Einstellungen zu testen (7, 8): Diese Studien haben eine erhebliche Knappheit an Nachnamen in bestimmten Bereichen und Regionen deutlich gemacht, was den Verdacht der nepotistischen Einstellungen erhöht, in denen Professoren Verwandte für akademische Positionen rekrutieren.

Hier erweitern wir diese Ergebnisse, indem wir einen internationalen Vergleich präsentieren und spezifische Randomisierungen einführen, die verschiedene Aspekte jedes akademischen Systems untersuchen. Obwohl unser Schwerpunkt auf der Wissenschaft liegt, könnte derselbe Ansatz in einer Vielzahl von Kontexten verwendet werden [z. B. In Studien über soziale Mobilität (9) oder gesundheitliche Unterschiede (10)] und sogar testen, ob Langlebigkeit mit Inzucht zusammenhängt (11).

Wir analysieren Nachnamen in drei Datensätzen von beispielloser Qualität und Größe: alle italienischen Akademiker in vier verschiedenen Jahren (2000, 2005, 2010 und 2015), Forscher, die derzeit am CNRS in Frankreich arbeiten, und Wissenschaftler, die an forschungsintensiven öffentlichen Einrichtungen in den Vereinigten Staaten arbeiten. Diese Datensätze ermöglichen es uns, die Entwicklung der Nachnamen in der Zeit (Italien) und die geographische Variabilität sowohl innerhalb als auch zwischen den Ländern zu verfolgen. Spezielle Merkmale der Daten ermöglichen es uns, die Anwesenheit akademischer Paare in Frankreich zu erkennen und die Auswirkungen der Antinepotismusgesetzgebung in Italien zu untersuchen.

Die Ergebnisse zeigen, dass das italienische akademische System dazu neigt, Forscher vor allem auf lokaler Ebene anzuziehen – viele Forscher haben Nachnamen, die typisch für die Region oder sogar die Stadt sind, in der sie arbeiten –, während das amerikanische System geografisch gut gemischt ist und einen starken Einfluss der Einwanderung hat. Darüber hinaus sind in den Vereinigten Staaten bestimmte Nachnamen typisch für bestimmte wissenschaftliche Bereiche – was bedeutet, dass Einwanderung und Forscher mit einem bestimmten ethnischen/kulturellen Hintergrund dazu neigen, überwiegend spezifische Forschungsbereiche ins Visier zu nehmen. Anhand der Verteilung von Vornamen zeigen wir in allen Systemen ein starkes Ungleichgewicht zwischen den Geschlechtern in den Disziplinen Wissenschaft, Technik, Ingenieurwesen und Mathematik (MINT). Schließlich zeigen wir, dass Vetternwerden in Italien vorhanden ist (aber rückläufig ist).

## **SCHLUSSFOLGERUNG**

Hier haben wir eine scheinbar dürftige Datenquelle – eine Liste von Namen von Professoren zusammen mit ihrem Forschungsgebiet und geographischen Informationen – genommen und elementare Randomisierungen verwendet, um Unterschiede in akademischen Systemen zu untersuchen. Wichtig ist, dass wir für jeden Winkel, den wir untersuchen wollten, eine spezifische Randomisierung erstellt haben, die zeigt, dass selbst extrem einfache Methoden auf subtile Muster in den Daten beleuchten können.

In Italien gruppieren sich die Namen nach Städten (SI-Anhang, Abb. S3), was zeigt, dass Professoren dazu neigen, dort zu arbeiten, wo sie geboren wurden. Das amerikanische System ist jedoch geografisch gut gemischt (SI Appendix, Abb. S5). Das starke Signal der Einwanderung wird durch die US-Randomisierungen hervorgehoben, wo zum Beispiel Physik und Mathematik deutlich testen, wenn sie nach Stadt randomisieren, aber nicht bei der Randomisierung nach Feld: bestimmte Namen sind mit bestimmten Feldern verbunden, im Einklang mit feldspezifischer Einwanderung und der Tatsache, dass amerikanische Forscher bestimmter Erbe dazu neigen, überwiegend Wissenschaft und Technik ins Visier zu nehmen.

Die Analyse von Ehe- und Mädchennamen für das französische System zeigt, dass unsere Methoden das Signal von familiären Bindungen erkennen können, wenn sie vorhanden sind. Beachten Sie, dass im italienischen System alle Frauen ihren Mädchennamen behalten, während in den Vereinigten Staaten ein nicht näher bezeichneter Teil der verheirateten Frauen die Namen ihrer Ehemänner annimmt – was möglicherweise den Überschuss an ips in der Pädagogik und anderen Bereichen erklärt. Die Analyse von Vornamen zeigt ein starkes Ungleichgewicht zwischen den Geschlechtern in den MINT-Fächern.

Selbst wenn man die geographische und feldspezifische Verteilung der Nachnamen berücksichtigt, weisen italienische Akademiker einen Überschuss an Nachnamen innerhalb der Abteilungen auf. Die Ergebnisse unserer zusätzlichen Analyse stimmen mit der Hypothese der Vetternwirtschaftskraft überein, die durch die Tatsache belegt wird, dass wir die Auswirkungen eines Antinepotismusgesetzes für den Zeitraum 2010-2015 erkennen können. Wichtig ist, dass unsere Analyse zeigt, dass Vetternwirtschaft feld- und regionsspezifisch ist und wahrscheinlich von einer Handvoll Abteilungen angetrieben wird. So stellten wir beispielsweise bei der Messung der Einstellungen im Jahr 2005 fest, dass 10 % der Abteilungen eine Anzahl von 0,1 % hatten (wir würden nach dem Zufallsprinzip 2,4 % erwarten), während die überwiegende

Mehrheit der Abteilungen 0 % hatte. In ähnlicher Weise zeigen die Randomisierungen in den Abb. 1, 2 und 5, dass bestimmte Regionen und Felder die Ergebnisse antreiben.

Für das italienische System sollten Beweise für die Wirksamkeit von Antiepotismusgesetzen und die Tatsache, dass das Phänomen abnimmt, als gute Nachricht mit zwei Vorbehalten begrüßt werden. Erstens ist der Rückgang der ips hauptsächlich auf die Pensionierungen zurück: Wir haben gezeigt, dass Rentner eher Nachnamen teilen als Neueinstellungen. Darüber hinaus ist die Zahl der italienischen Hochschulen nach einem starken Anstieg der Zahl der Fakultäten zwischen 2000 und 2005 stetig zurückgegangen, wobei in den letzten zehn Jahren ein gesamter Verlust von 10 % zu verzeichnen war. Die Zahlen sehen noch schlechter aus, wenn man sie auf der Ebene der Regionen, Felder oder einzelnen Institutionen untersucht (SI-Anhang): Toscana und Ligurien verloren ein Viertel ihrer Fakultät (Siena, 30,2 %; Florenz, 29,3 %; Genua, 24,3 %) und Geologie (21,4 %) und die Geisteswissenschaften (18,9 %) einen großen Teil ihrer Professoren verloren haben. Das Problem der Vetternwirtschaft durch Auflösung des Universitätssystems zu lösen, würde das Kind mit dem Bade ausschütten. Zweitens können Antiepodose-Gesetze negative Nebenwirkungen haben, vor allem, wenn sie auf Ehegatten-Mieten abzielen. In der ersten Hälfte des 20. Jahrhunderts beispielsweise schufen Antiepotismusgesetze in den Vereinigten Staaten das Phänomen der "verschwindenden Ehefrauen" (17): Weil Ehegatten nicht in derselben Abteilung wie ihre Ehemänner eingestellt werden konnten, arbeiteten viele Frauen als unbezahlte Gäste, was den Prozess verlangsamte, der zu einer gleichberechtigten Geschlechtervertretung führte.

Die Beispiele Frankreichs, das Einstellungsverfahren hat, die denen des italienischen Systems sehr nahe kommen, und die Vereinigten Staaten, wo die Praktiken jedoch sehr unterschiedlich sind, zeigen, dass man ein faires akademisches System aufbauen kann, ohne dass besonders harte Maßnahmen erforderlich sind. Tatsächlich begrüßen viele US-Institutionen Paare (Ehegattenmieten; oft auf inländische Partner ausgedehnt), obwohl Antiepotismus-Bestimmungen vorhanden sind, so dass ein Partner nicht für den beruflichen Aufstieg des anderen Partners verantwortlich sein kann.

## **TRANSLATED VERSION: PORTUGUESE**

Below is a rough translation of the insights presented above. This was done to give a general understanding of the ideas presented in the paper. Please excuse any grammatical mistakes and do not hold the original authors responsible for these mistakes.

## **VERSÃO TRADUZIDA: PORTUGUÊS**

Aqui está uma tradução aproximada das ideias acima apresentadas. Isto foi feito para dar uma compreensão geral das ideias apresentadas no documento. Por favor, desculpe todos os erros gramaticais e não responsabilize os autores originais responsáveis por estes erros.

## **INTRODUÇÃO**

Desde a sua criação, a ciência tem sido um esforço mundial, com publicações e conferências acadêmicas que ligam investigadores em todo o mundo. Apesar das muitas semelhanças (por exemplo, a organização de académicos em departamentos e as fileiras académicas ubíquas), os sistemas académicos em todo o mundo são, no entanto, bastante distintos nos seus objetivos e práticas. Em muitos países europeus, por exemplo, os professores são funcionários públicos e, por conseguinte, os seus procedimentos de contratação estão sujeitos a regulamentações especiais. Em contraste, as universidades americanas têm mais liberdade na escolha do seu corpo docente. Os salários, os deveres e os recursos também variam muito, tanto dentro como entre sistemas.

Aqui, examinamos as diferenças nos sistemas académicos utilizando uma forma muito simples de dados: uma lista de nomes de professores que trabalham numa determinada instituição juntamente com o seu posto, campo de estudo e localização geográfica. Estes dados são fáceis de obter e podem ser usados para desvendar padrões de mobilidade e imigração (são investigadores empregados na região onde



nasceram e cresceram?), desequilíbrio de gênero (são mulheres sub-representadas em determinadas áreas?), e até nepotismo (os professores contratam os seus familiares para cargos acadêmicos?).

O uso de apelidos como forma de dados tem uma longa história na biologia, a começar por George Darwin (filho de Carlos), que usou a distribuição de apelidos em Inglaterra para estimar a prevalência de casamentos por primos primos (como os seus pais) (1). Logo apelidado de "genética populacional do homem pobre" (2), o estudo das isonígenas (ocorrências de pessoas com o mesmo nome) forneceu uma fonte barata de dados (grandes), com a vantagem de que os apelidos aproximariam os alelos neutros (2, 3), permitindo o estudo das migrações humanas (4). Com o advento dos métodos moleculares modernos, os apelidos têm sido associados com haplótipos cromossomas Y (5). Mais recentemente, a associação de primeiros e últimos nomes étnicos tem-se revelado preditiva do sucesso profissional (6). Mais perto do espírito deste trabalho, a distribuição de apelidos em académicos italianos tem sido usada para testar a hipótese de contratações nepotísticas (7, 8): estes estudos têm destacado uma escassez significativa de apelidos em determinadas áreas e regiões, levantando a suspeita de contratações nepotísticas, em que os professores recrutam familiares para cargos académicos.

Aqui, expandimos estes resultados apresentando uma comparação internacional e introduzindo aleatoriedades específicas que sondam diferentes aspetos de cada sistema académico. Embora o nosso foco seja a academia, a mesma abordagem poderia ser usada numa variedade de contextos [por exemplo, em estudos de mobilidade social (9) ou disparidades de saúde (10)] e até para testar se a longevidade está relacionada com a consanguinidade (11).

Analizamos os apelidos em três conjuntos de dados de qualidade e dimensão sem precedentes: todos os académicos italianos em quatro anos diferentes (2000, 2005, 2010 e 2015), investigadores atualmente a trabalhar no CNRS em França, e académicos que trabalham em instituições públicas intensivas de investigação nos Estados Unidos. Estes conjuntos de dados permitem-nos acompanhar a evolução dos apelidos no tempo (Itália) e a variabilidade geográfica dentro e entre países. Características especiais dos dados permitem-nos detetar a presença de casais académicos em França e sondar os efeitos da legislação antinepotismo em Itália.

Os resultados mostram que o sistema académico italiano tende a atrair investigadores principalmente a nível local — muitos investigadores têm apelidos típicos da região ou mesmo da cidade em que trabalham — enquanto o sistema americano é geograficamente bem misturado, com uma forte influência da imigração. Além disso, nos Estados Unidos, certos apelidos são típicos de campos científicos específicos, o que significa que a imigração e os investigadores de determinadas origens étnicas/culturais tendem a visar áreas de investigação preponderantemente específicas. Utilizando a distribuição de primeiros nomes, mostramos um forte desequilíbrio de gênero nas disciplinas de ciência, tecnologia, engenharia e matemática (STEM) em todos os sistemas. Finalmente, mostramos que o nepotismo está presente (mas em declínio) em Itália.

## CONCLUSÃO

Aqui, temos tomado uma fonte ostensivamente escassa de dados - uma lista de nomes de professores juntamente com o seu campo de pesquisa e informação geográfica - e usamos aleatoriedades elementares para investigar diferenças nos sistemas académicos. Importante, produzimos uma aleatoriedade específica para cada ângulo que queríamos sondar, mostrando que mesmo métodos extremamente simples podem lançar luz sobre padrões subtis nos dados.

Em Itália, nomes de cluster por cidade (SI Apêndice, Fig. S3), mostrando que os professores tendem a trabalhar onde nasceram. O sistema americano, no entanto, é geograficamente bem misturado (SI Apêndice, Fig. S5). O forte sinal da imigração é destacado pelas aleatoriedades dos EUA, onde, por exemplo, a física e a matemática testam significativamente quando se aleatoriedade pela cidade, mas não quando aleatoriamente por campo: certos nomes estão associados a campos específicos, consistentes com a imigração específica do campo e o facto de os investigadores americanos de certos patrimónios tendem a visar preponderantemente a ciência e a engenharia.

A análise de nomes casados vs. Donzelas para o sistema francês mostra que os nossos métodos podem detetar o sinal de laços familiares quando estão presentes. Note-se que, no sistema italiano, todas as mulheres mantêm o seu nome de solteira, enquanto nos Estados Unidos, uma fração não especificada de mulheres casadas toma o nome dos seus maridos, possivelmente explicando o excesso de ips na pedagogia e em outros campos. A análise dos primeiros nomes destaca o forte desequilíbrio de género nos domínios STEM.

Mesmo quando contabilizam a distribuição geográfica e específica de apelidos, os académicos italianos exibem um excesso de partilha de apelidos dentro dos departamentos. Os resultados da nossa análise adicional são consistentes com a hipótese do nepotismo como testemunha pelo facto de podermos detetar os efeitos de uma lei antinepotismo em vigor para o período 2010-2015. Importante, a nossa análise mostra que o nepotismo é específico do campo e da região e provavelmente impulsionado por um punhado de departamentos. Por exemplo, ao medir  $\alpha^{\wedge}$  para as contratações em 2005, descobrimos que 10% dos departamentos tinham um  $\alpha^{\wedge} \geq 0.1$  (seria de esperar 2,4% aleatoriamente), enquanto a grande maioria dos departamentos tinha  $\alpha^{\wedge} \approx 0$ . Da mesma forma, as aleatoriedades nos Figs. 1, 2 e 5 mostram que regiões e campos específicos impulsionam os resultados.

Para o sistema italiano, a evidência da eficácia das leis antinepotismo e o facto de o fenómeno parecer estar em declínio devem ser saudados como boas notícias, com duas ressalvas. Em primeiro lugar, a diminuição dos ips deve-se, em grande parte, às reformas: mostrámos que os reformados têm mais probabilidades de partilhar apelidos do que novas contratações. Além disso, após um grande aumento do número de docentes entre 2000 e 2005, a dimensão da academia italiana tem vindo a diminuir de forma constante, com uma perda global de 10% na última década. Os números parecem ainda piores quando examinados ao nível das regiões, campos ou instituições individuais (apêndice SI): Toscana e Ligúria perderam um quarto do seu corpo docente (Siena, -30,2%; Florença, -29,3%; Génova, -24,3%), e geologia (-21,4%) e as humanidades (-18,9%) perderam uma grande fração dos seus professores. Resolver o problema do nepotismo ao dissolver o sistema universitário seria deitar o bebé fora com a água do banho. Em segundo lugar, as leis antinepotismo podem ter efeitos colaterais negativos, especialmente quando visam as contratações de cônjuges. Por exemplo, na primeira metade do século XX, as leis antinepotismo nos Estados Unidos criaram o fenómeno das "esposas desaparecidas" (17): porque os cônjuges não podiam ser contratados no mesmo departamento que os seus maridos, muitas mulheres trabalhavam como hóspedes não remunerados, abrandando o processo que conduzia à igualdade de representação de género.

Os exemplos da França, que tem procedimentos de contratação muito próximos dos do sistema italiano, e dos Estados Unidos, onde as práticas são, no entanto, muito diferentes, mostram que se pode construir um sistema académico justo sem a necessidade de medidas especialmente duras. Na verdade, muitas instituições norte-americanas acolhem casais (contratações conjugais; muitas vezes estendidas a parceiros domésticos), embora existam disposições antinepotismo, para que um parceiro não possa ser responsável pelas progressões na carreira do outro parceiro.