# SHIFTING WINDS: GENDERED STRUCTURES OF ACADEMIC MENTORSHIP

**Jiabao Li**
stephlijiabao@gmail.com

**Houjiang Liu**
liu.ho@utexas.edu

**Jilie Zeng**
zengjilie@gmail.com

**Di Wu**
wudi9901@utexas.edu

**Ying Ding**
ying.ding@ischool.utexas.edu

**Alec McGail**
am2873@cornell.edu

The University of Texas at Austin

**www.mentortrees.com**

**Abstract**
Every researcher alive today had their mentors, those who helped assimilate them into a life of scholarly work. And in turn they each had their mentors, and so on to the dawn of knowledge. In the same way, each researcher's mentees take their perspectives and methods to future mentees, and to their mentees, etc. These comprise the roots and branches, respectively, of the academic tree of a single researcher. If we let these ancestors' and descendants' genders affect these trees like a "wind," most curl nearly to the earth. We depict and describe the structure of these trees, and how this wind has changed over the decades. To set these trees growing upright again we visualize giving differential weight to male and female researchers.

**Authors Keywords**
science of science, gender in academia, visualization design

**Introduction**
Mentors pass the torch of scholarly practice to their mentees in the most intimate academic relationship available. Mentorship reproduces science from one generation to the next, as mentees take on the perspectives and methods of their advisors [4]. And mentees' academic careers are strongly impacted by their mentors' past grant success, as well as mentors' citations and prestige [7, 8]. Academics working today, of both genders, have incredibly gendered roots. Higher education has been exclusively male until the late 19th century. In the 1920s there was a boom, with 32.5% of professors, college presidents, and instructors being women by 1930 [1]. This declined after WWII and has steadily risen to present [1]. And although we have come a long way in recent decades towards bringing women into academia, in 2012 75% of full professors were still male [6] and in 00s in IEEE VIS only around 30% of authors were female [9]. Women still transition to principal investigator approximately 20% slower than men [5]. And this progression to more equality is heavily dependent on locale, as Japanese, Saudi Arabian, and Venezuelan women make up less than 30% of authors [2]. Homophily and academic resource concentration, most exaggerated at the top of the academic world, then make for a gendered imbalance of academic power [8]. This all while the youngest generation of female academics equal or outperform their male counterparts [10].
In this paper we depict the gendered academic tree of each individual researcher from two perspectives. The first tree, their descendants, is composed of a researcher, their mentees, their mentees' mentees, and so on until present. The second, their ancestors, is composed of a researcher, their mentors, their mentors' mentors, and so on back to the beginning of time. We present an online interactive narrative, visualizing the gendered "wind" which has bent most trees nearly to the ground. The upshot is a hope that we can detach ourselves from this past. If, however, history defines current academia, we should make sure women are well represented across all positions in academia, especially as mentors.

In the spirit of mingling spaces, this project has been the culmination of a collaborative space we have built joining the scientific and the visual. That is, between scientists and those depicting them. We are students of the social sciences, computer science, and of design and visualization, working together to draw the gendered structure of science. The work itself is of interest to practicing scientists, scholars of science, and students of scientific visualization. The trees we draw then depict the real mingling of disciplines, in the colors painted at the end of each branch.

VISAP'22, Pictorials and annotated portfolios.

## DRAWING THE TREE

A tree starts with a single scholar, sprouts with their mentees and has roots in their mentors. For current scholars, the tree has only just sprouted, with them, and the roots are the most interesting feature. For long-dead and influential scholars their descendant tree is most interesting. Descendant trees will be the focus of much of this paper. To draw the descendant tree starting at Isaac Newton, for instance, we draw Isaac Newton first, growing straight up. From that we draw sprouts for Roger Cotes, William Whiston, and John Flamsteed, the three scholars who are noted in Academic Family Tree as pupils of Newton. These all have recognizably male first names, so they are drawn equally spaced in a fan tilting to the left, in green. They were each quite prolific, and their mentees are plotted in the same way, tilting left or right if they are male or female (see Figure 4 for reference).

Concretely, male mentee number sprouts an angle -iM relative to their mentor, and female mentee number sprouts an angle iF, where M and F are positive tunable parameters. For all trees presented here M = F, but we let users tune these parameters independently on the website.

Mentees with unknown gender sprout in the same direction as their mentors. We color female mentees red, male mentees in green, and unknown mentees in grey. The outermost scholars, the canopy of the tree, are those who have no mentees. Each scholar's "sprout" in the tree has width proportional to the total number of descendants they have. Thus the base is the thickest sprout, supporting the rest of the tree, while the canopy, descendants with no mentees, are the thinnest.

notabilita.net [11] uses a similar paradigm of curling branches, depicting instead discussions to keep or delete Wikipedia articles. The sequence of votes to keep or delete an article bends the discussion's branch left or right as it grows. The two trees of interest in their analysis are articles which are eventually kept and those which are eventually deleted, which have markedly distinct shapes. In their trees each branch is of interest, whereas in ours it is the tree as a whole. Their trees span a few days, while ours span many generations. Their trees are made of a collection of independent shoots, while in ours each branch can sprout new shoots through further mentorship. The following sections describe ways to describe mentorship trees, and mentorship forests, giving a bird's eye view of thousands of researchers' roots.

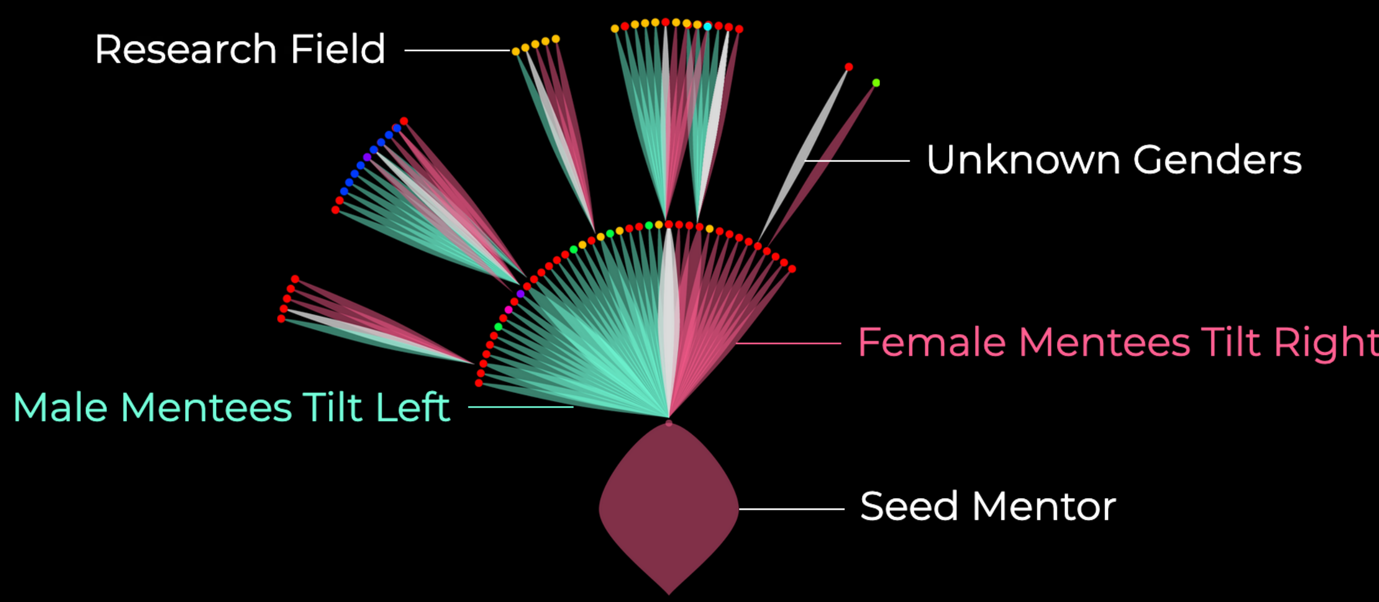



Figure 1. Mentor tree and legend of Jennifer A. Doudna

## DATA

The MENTORSHIP dataset [3] which we use in this paper is based on the Academic Family Tree, a crowdsourced effort to collect academic genealogy. Registered users can input, update, or confirm the information posted there. This includes relational data, keeping track of advisors and their advisees, collaborators, research assistants, and postdoctoral students. The dataset now holds 774,733 researchers in total across 112 fields. As a crowdsourced effort, this information is surely incomplete and nonrepresentative. Yet insofar as findings are dramatic and consistent, they certainly act as a window into the history, present, and future of mentorship in academia.

To narrow down the scope of visualizations and make sure that the mentorship data contains relatively complete and correct information for visualization, we investigate only the most represented research areas. These research areas are anthropology, cell biology, chemistry, computer science, engineering, history, linguistics, literature, math, microbiology, neuroscience, nursing, philosophy, physics, political science, and sociology.

For researchers' gender, we use the labels provided in [3]. They use a neural network based in BERT embeddings to predict gender, and train it on 1) popular male and female names as collected by the Social Security Administration (SSA) in the United States, as well as 2) an international dataset produced by the agreement of the Genni and SexMac tools for gender-tagging. Names which have very few occurrences, or which can be either male or female names, are tagged as Unknown. 82.5% of the scholars in this dataset are coded as either Man or Woman, while 17.5% of researchers in this dataset are labeled Unknown. This proportion of "Unknowns" has varied over time, from around 5% throughout the 1800s, trending up to 10% by 1940, and rising again starting in 1980 to its present peak of 21% (2010-2020). This increase could be explained by a rise in global academia. Some countries, like Vietnam and Japan, have many more androgynous names. In addition, the recent widespread revolution in societal gender norms may be affecting the usage of androgynous names overall. And because we cannot tag extremely unusual names, any influx of entirely novel names would cause an increase in the percent of Unknowns.

To be clear, we cannot study the full complexity of gender in this paper. This tool is trained to tag the typical legal sex of a first name, which has a varied correspondence to identified and enacted gender through time and space. The modern landscape of formal non-binary genders, and the fluidity of gender expression throughout the ages, are not considered in this paper. Instead, our study focuses on the binary imposed by the societies the researchers were embedded in, and how this binary has structured researchers' academic roots and branches. We are limited insofar as gender is not expressed in the first name, but where names are attached to a sex-category we can track the gendering work done by the academic institution as a whole. A smaller-scale qualitative analysis would be better suited to addressing the various other aspects of gender in science which are inaccessible to this methodology.

For 18.0% of mentor-mentee relationships there is a start date, while 82.4% have an end date. We assign a date of graduation for each student based on the first of their mentorships which have a date attached, priveledging end-dates over start-dates.

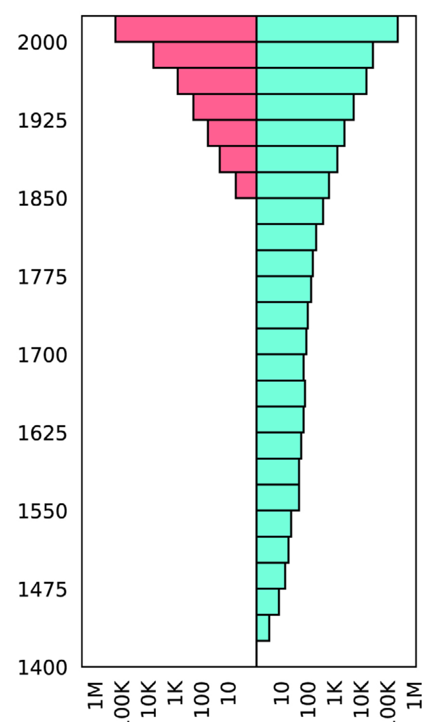



Figure 2: The change in number of mentees and gender distribution in mentees, over time. Bars are 25 years wide. Darker red bars are women and lighter green are men. Those with unknown gender were excluded.

VISAP'22, Pictorials and annotated portfolios.

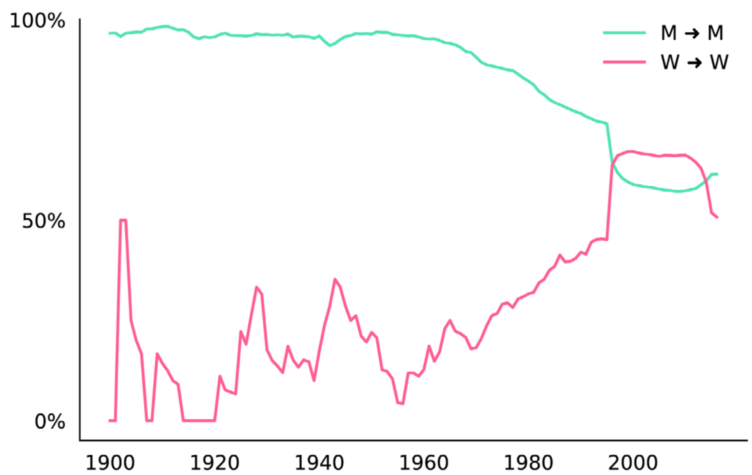



Figure 3. The percentage of men's mentees who were men, compared with the percentage of women's mentees who were women. Unknowns were excluded, and we present a 5-year rolling average beginning at the date on the x-axis.

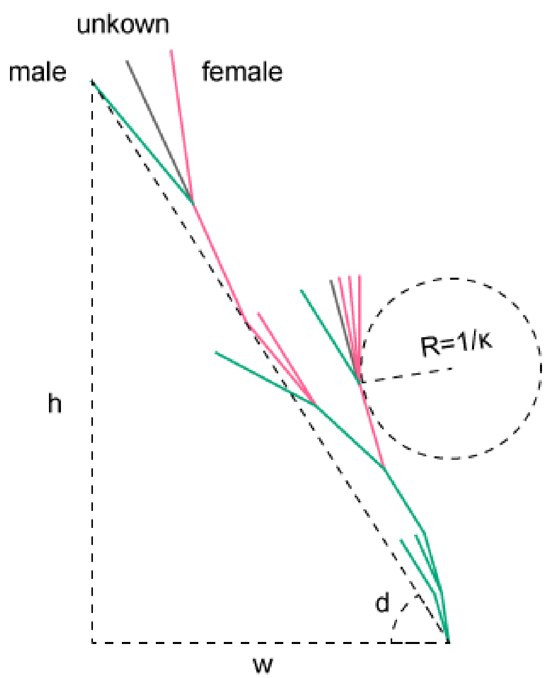



Figure 4: Visual metrics of a tree, w = width, h = height, d = angle in degrees, R = radius of curvature, k = curvature

VISAP'22, Pictorials and annotated portfolios.

This assigns graduation dates to 90.3% of mentees in the dataset, 568,630 scholars. Figure 2 displays the number of mentees in this dataset in each year. The x-axis is logarithmic, so the linear trend we see in number of females from 1850 to present represents exponential growth. In the years since 2000 there have been almost 8 times as many mentees as in all previous years combined. Women made up 5% of mentees in 1925-1949, 6% of mentees 1950-1974, 24% of mentees 1975- 1999, and 49% of mentees 2000-present. The years since 2000 has seen 4.4 times more male mentees than all previous years combined, and almost 23 times more female mentees than all previous years combined.

Looking now at the proportion of males' mentees who were male, and the proportion of females' mentees who were female (Figure 3) we see that females' mentees were not more than 50% female until 1996-2001. Perhaps more surprisingly, we can see the F-F mentorship rate plumitting around 2011, from over 65% female mentees amongst female mentors back to 50%. The decreasing F-F mentorship rate after 1944 matches the history of women being shouldered out of academia after WWII.

## CHARACTERISTICS OF TREES

To compare trees, we first identify the "seed" scholars in every field, individuals who have no mentors, no formal roots in their field. These are the ultimate bases of ancestor trees, and typically reside where further mentorship is not known or has not been entered into Academic Family Tree. We draw these trees with equal female and male curls, and record their width and height. We also record the radius and curvature of each mentee in the tree (see Figure 4).

The tallest and widest trees, then, correspond to the maximum height and width of these drawn trees across 16 research fields. By averaging the radius of curvature over all mentees in each tree, we find the most tilted trees, to the right and to the left. And by computing the standard deviation of all curvature for each datapoint, we find the most diverse tree, the tree with the most variation in curl amongst its mentees. We depict the widest and tallest tree, the most tilted trees, and the most diverse tree on the website. All code for generating these visualizations is available online: https://github.com/JialingJia/scimap-FA-2021

Figure 5a. Nursing, the most female-tilted field

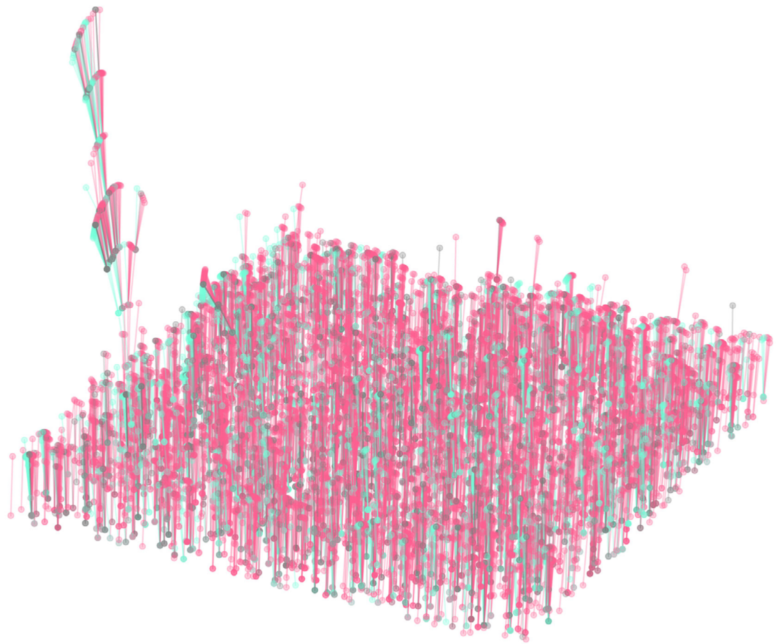
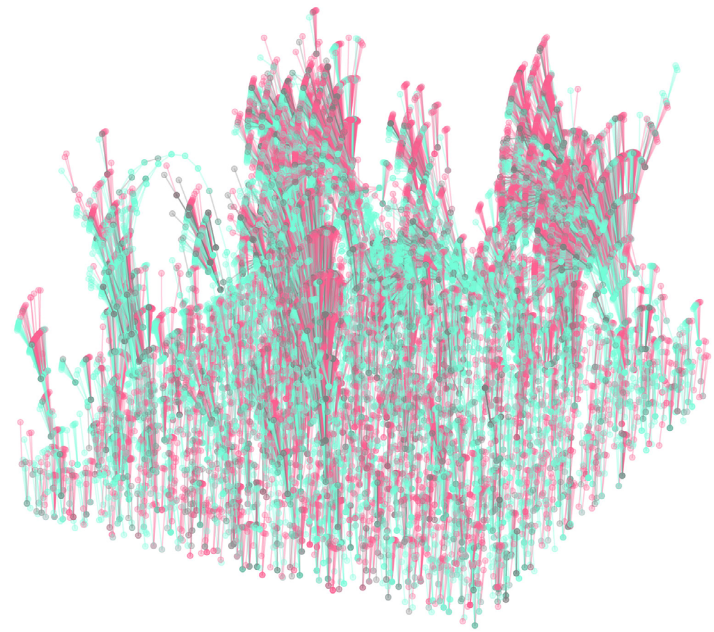


Figure 5b. History, the most male-tilted field

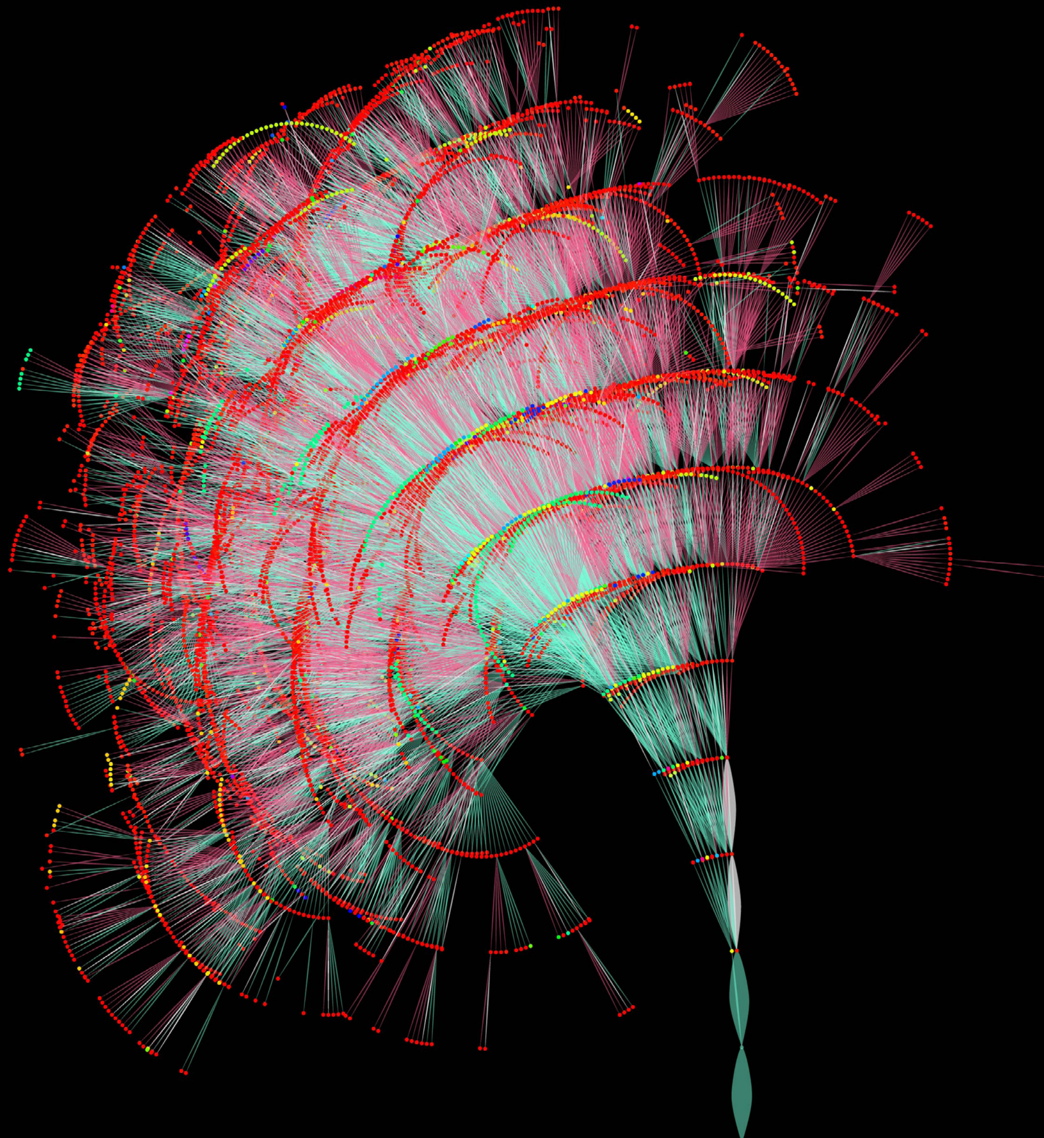


## CHARACTERISTICS OF FORESTS

In addition to exploring prototypical trees, it is also important to understand how the gender distribution is visually different across different research areas. We run a simple statistical analysis by adopting the previous visual metrics of a single tree and analyzing the visual metrics of the mentorship forests. We use the mean value of angles in degrees, width, height, and curvature to represent the visual metrics of forest.

By aggregating the mean of tree angles that lean towards both genders, we find that the most male-tilted research field is history, while the most female-titled research field is nursing. Nursing also has fewer generations of mentor-mentee relationships compared to other fields. Interestingly, history is also the field with the most variation in drift. That is, there are highly tilted trees leaning towards both genders in the field of history. The widest research field on average is neuroscience, while the tallest field on average is chemistry. Mentorship trees in neuroscience and chemistry also have relatively more mentees on average than in other fields.

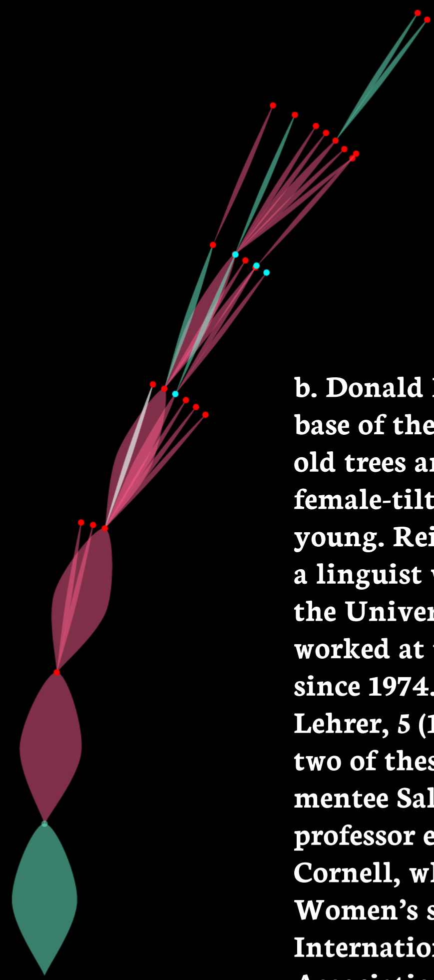
To visually display such statistical differences, we draw Nursing and History as a forest of mentorship trees in three dimensions (Figure 5a, 5b). Trees grow from a random spot in their terrain in the same way described above. The more crowded the forest, the more seed researchers as roots; the denser the canopy of the forest, the more researchers in total are in the trees.



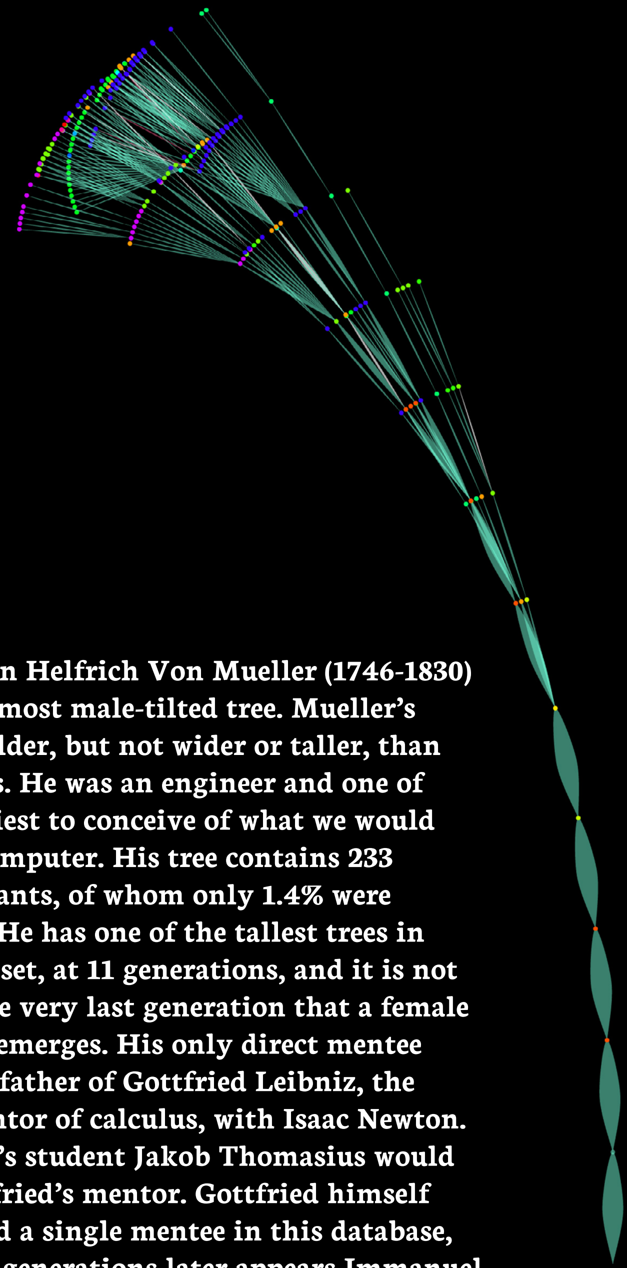
## SPECIFIC TREES

We have chosen some trees which are exceptional in their structure, and others which were of interest to the authors. We display these trees in the following pages, and the rest of this section describes these scholars and their mentorship trees.

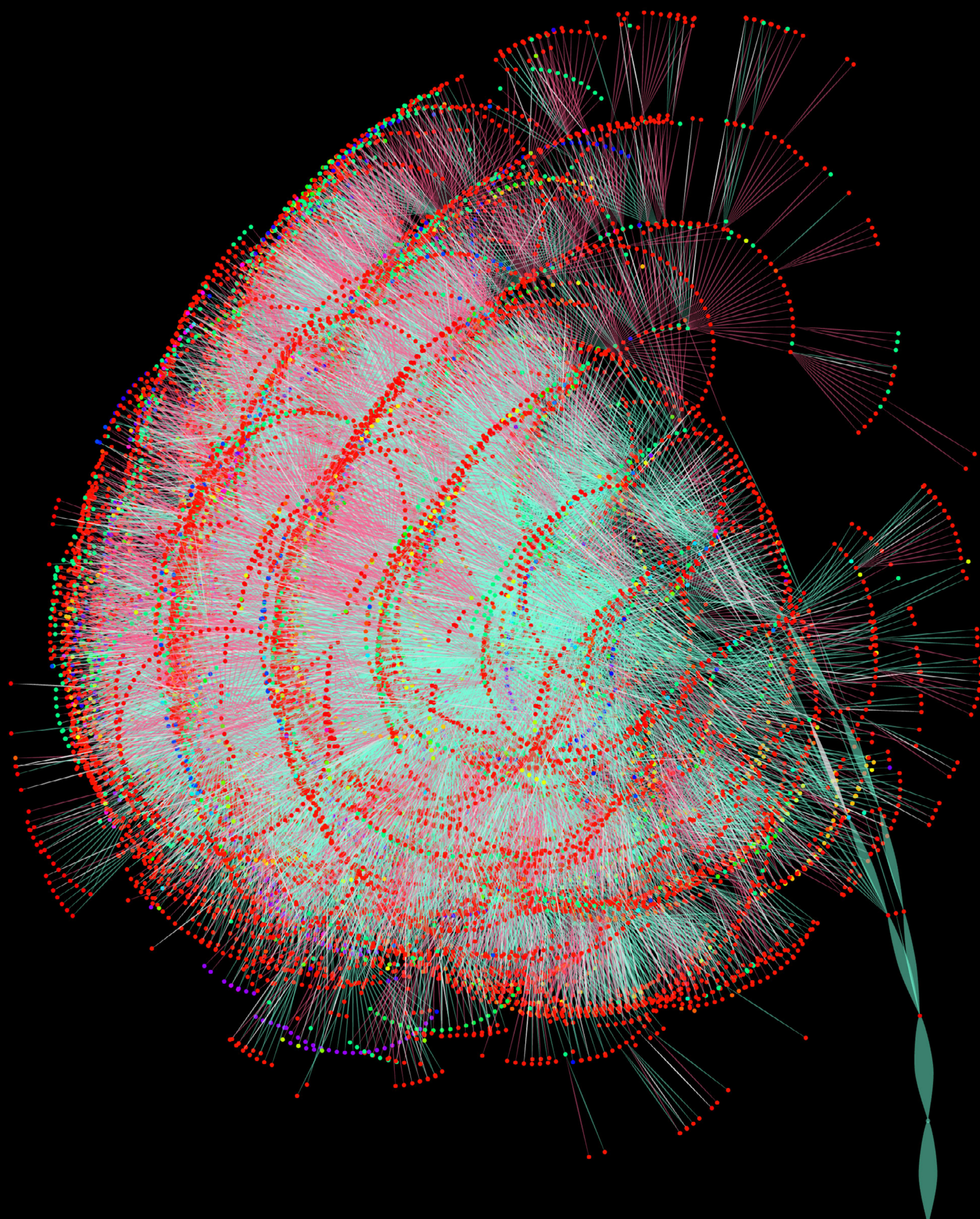
**a. Francis Galton (1822 - 1911), sits at the base of the widest and tallest tree in this dataset. For a tree to be wide or tall it must be old and enduring. Galton was an incredibly influential polymath, contributing to statistics, sociology, psychology, and anthropology, amongst many other things. He was also a proponent of social Darwinism, eugenics, and scientific racism. Galton's tree clearly bends toward male, and its curliest leaves are male. Yet, 51.1% of the scholars in Galton's tree are female! His mentee James Cattell, a psychologist, has a relatively balanced and massive tree, 52% female, while the influential statistician Karl Pearson's tree had 791 mentees, and only 33% were female. There are other "tallest" trees, which tower at 11 generations. These include Charles Sanders Pierce, William James, and Ludwig Bolzmann.**



b. Donald Reiff (1937 - present) sits at the base of the most female-tilted tree. Just as old trees are more male tilted, the most female-tilted trees are comparatively young. Reiff mentored Adrienne Lehrer, a linguist who got her PhD in 1968 at the University of Rochester. She has worked at the University of Arizona since 1974. Of the 26 descendants of Lehrer, 5 (19%) were male. All except two of these came through Lehrer's mentee Sally McConnell-Ginet, professor emerita of linguistics at Cornell, who served as director of Women's studies and president of the International Gender and Language Association, amongst other posts.



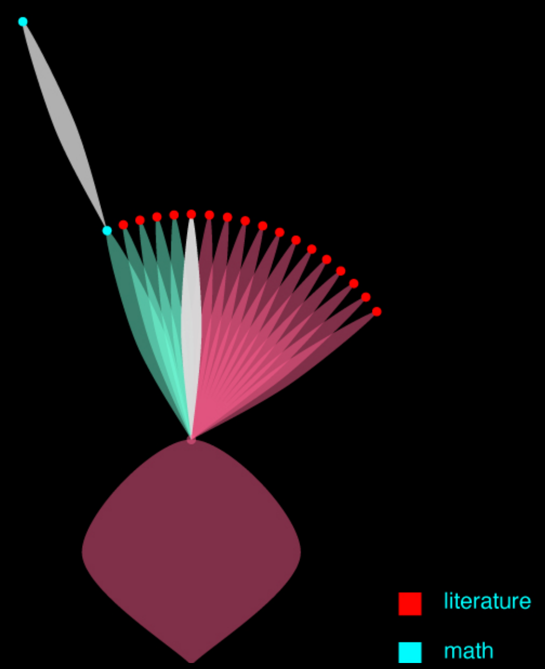
c. Johann Helfrich Von Mueller (1746-1830) has the most male-tilted tree. Mueller's tree is older, but not wider or taller, than Galton's. He was an engineer and one of the earliest to conceive of what we would call a computer. His tree contains 233 descendants, of whom only 1.4% were female. He has one of the tallest trees in the dataset, at 11 generations, and it is not until the very last generation that a female mentee emerges. His only direct mentee was the father of Gottfried Leibniz, the co-inventor of calculus, with Isaac Newton. Mueller's student Jakob Thomasius would be Gottfried's mentor. Gottfried himself only had a single mentee in this database, but two generations later appears Immanuel Kant, who has 140 in his descendant tree.

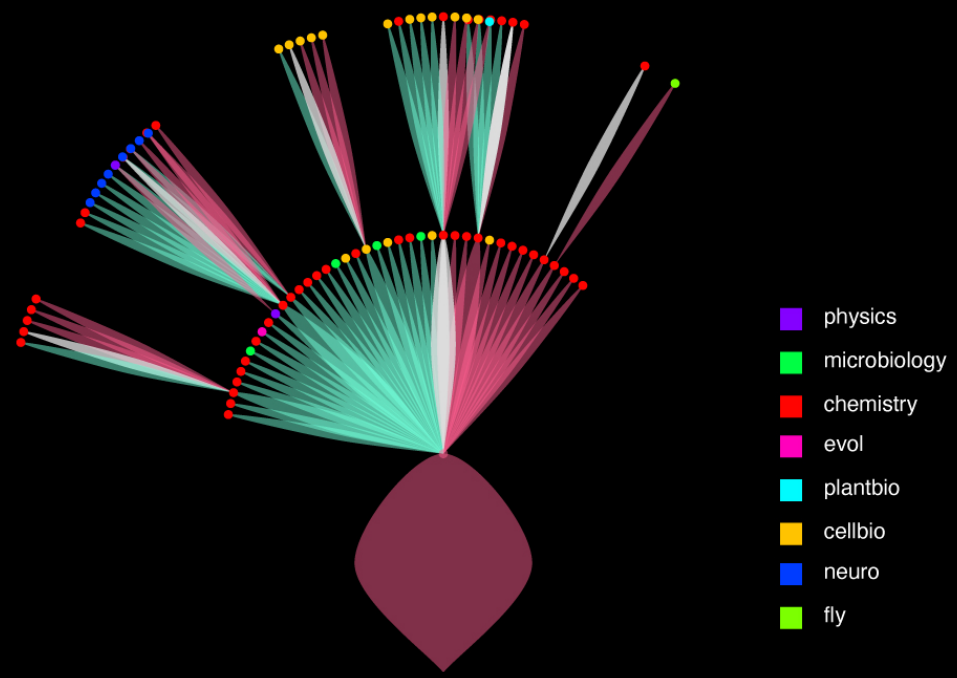


d. Christian Gottfried Daniel Nees Von Esenbeck (1766 - 1858) had the most diverse tree. That is, there is the most variety in how the branches in his tree curl. He was a botanist, physician, zoologist, and natural philosopher. He has 34,480 descendants and is also one of the tallest trees, at 11 generations. 2,393 of his descendants have 10 consecutive male ancestors, and 5,036 have 9. Yet his tree is 42% female, and includes 4,749 descendants with at least five females in their lineage back to Gottfried. The relatively high variation in curl in this tree gives it its distinct fullness.

e. Donna Haraway (1944 - present) made contributions to our thinking on human-machine and human-animal relations. She became the first tenured professor in feminist theory in the United States in 1980, and has received numerous accolades for her work. She has been teaching since 1974 and her tree is 72% female. Donna herself has had 12 female mentees, 5 male mentees, and one of unknown gender.

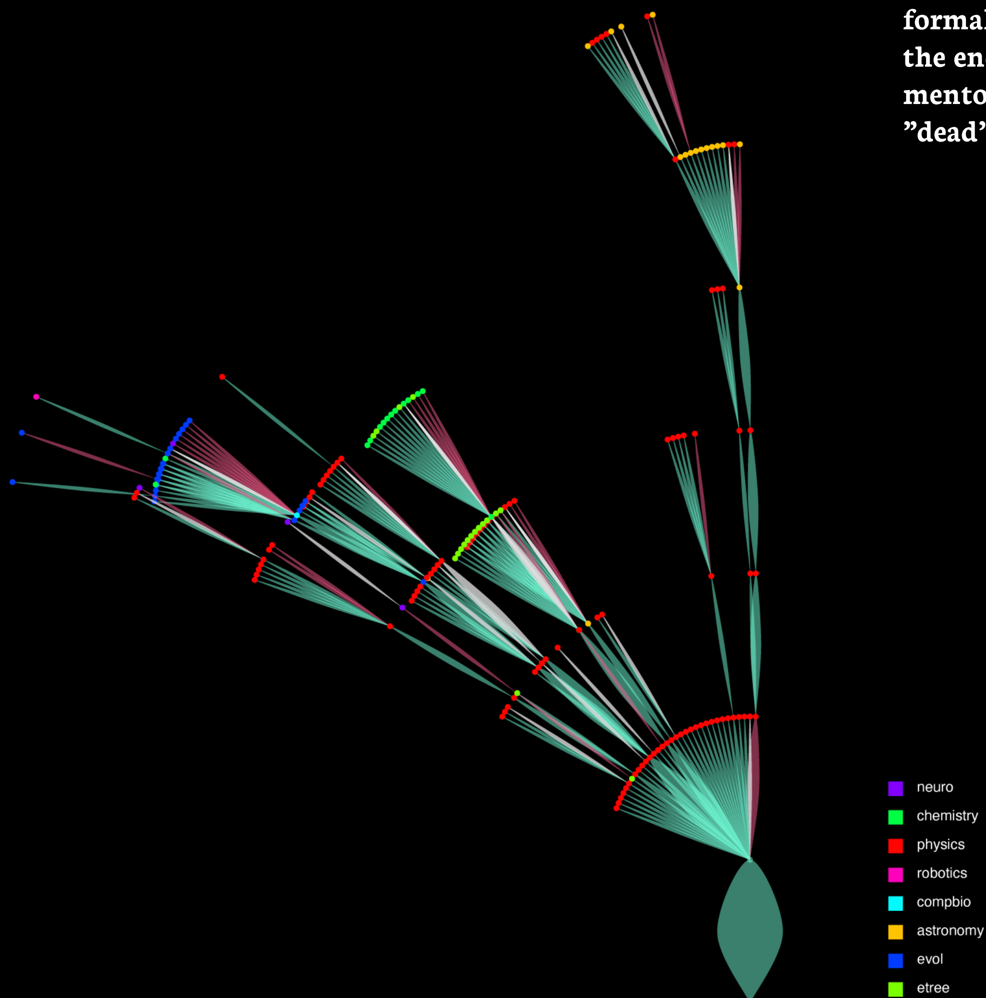
f. Jennifer A. Doudna (1964 - present) has made significant contributions to biochemistry and genetics, and is most famous for her work developing CRISPR, for which she received the Nobel Prize in Chemistry in 2020. Her tree is 40.5% women and contains 98 descendant scholars. This is a full tree, considering its age. Also because it is young and relatively balanced, it is almost upright.

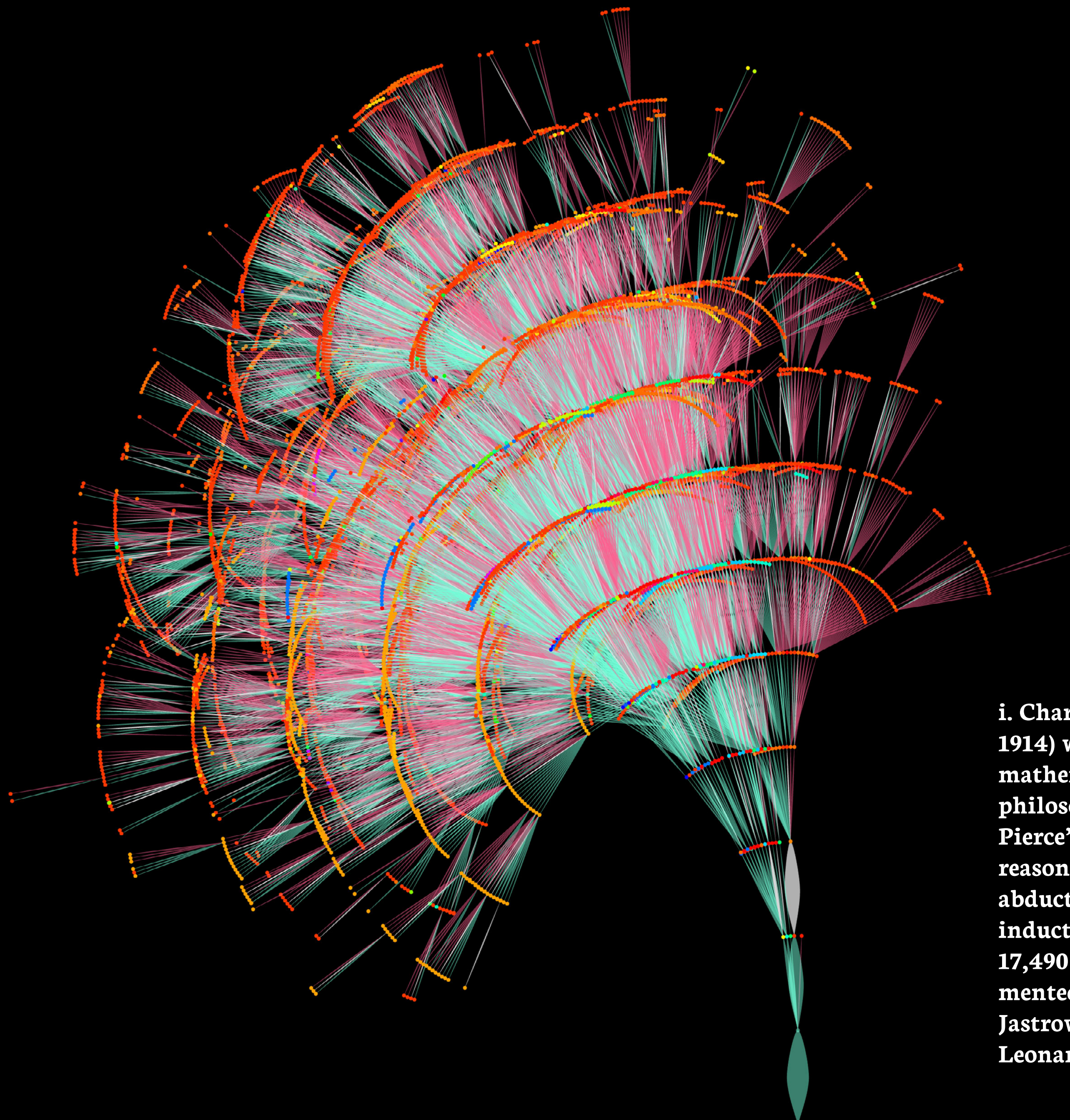


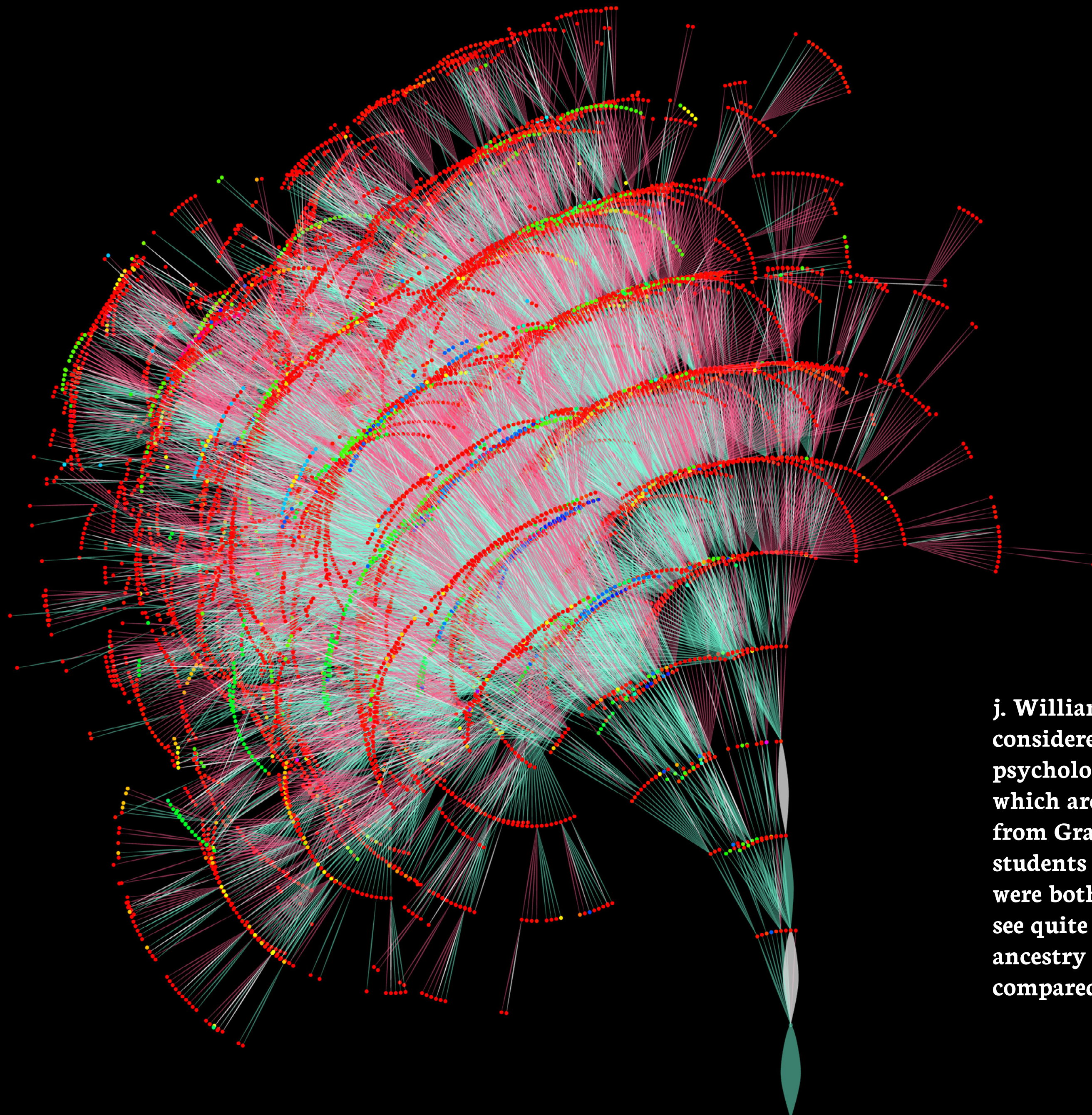
g. Richard Feynman (1911 - 1988) was a theoretical physicist, who introduced the path integral formulation of quantum mechanics. He was a prolific teacher, mentoring 32 males and 2 females. In all he had 230 descendants, 16% of which were female.

h. Jane Goodall (1934 - present) devoted her life to the study of chimpanzees in Tanzania. In 1960, she was the first person to see a chimpanzee use a twig to fish termites from the dirt. Through her study and then as head of the Jane Goodall Institute, she pioneered the systematic study of primates' social lives. She has written more than 15 books and is the subject of more than 40 films and now spends most of her year advocating for chimps, the environment, and our world. Incredibly, Jane Goodall had no formal academic mentees whatsoever. This is a stark reminder that formal mentorship as encoded in this dataset is not the end of the eminence of a scholar, and that a lack of mentorship at the leaves does not necessarily indicate a "dead" tree.

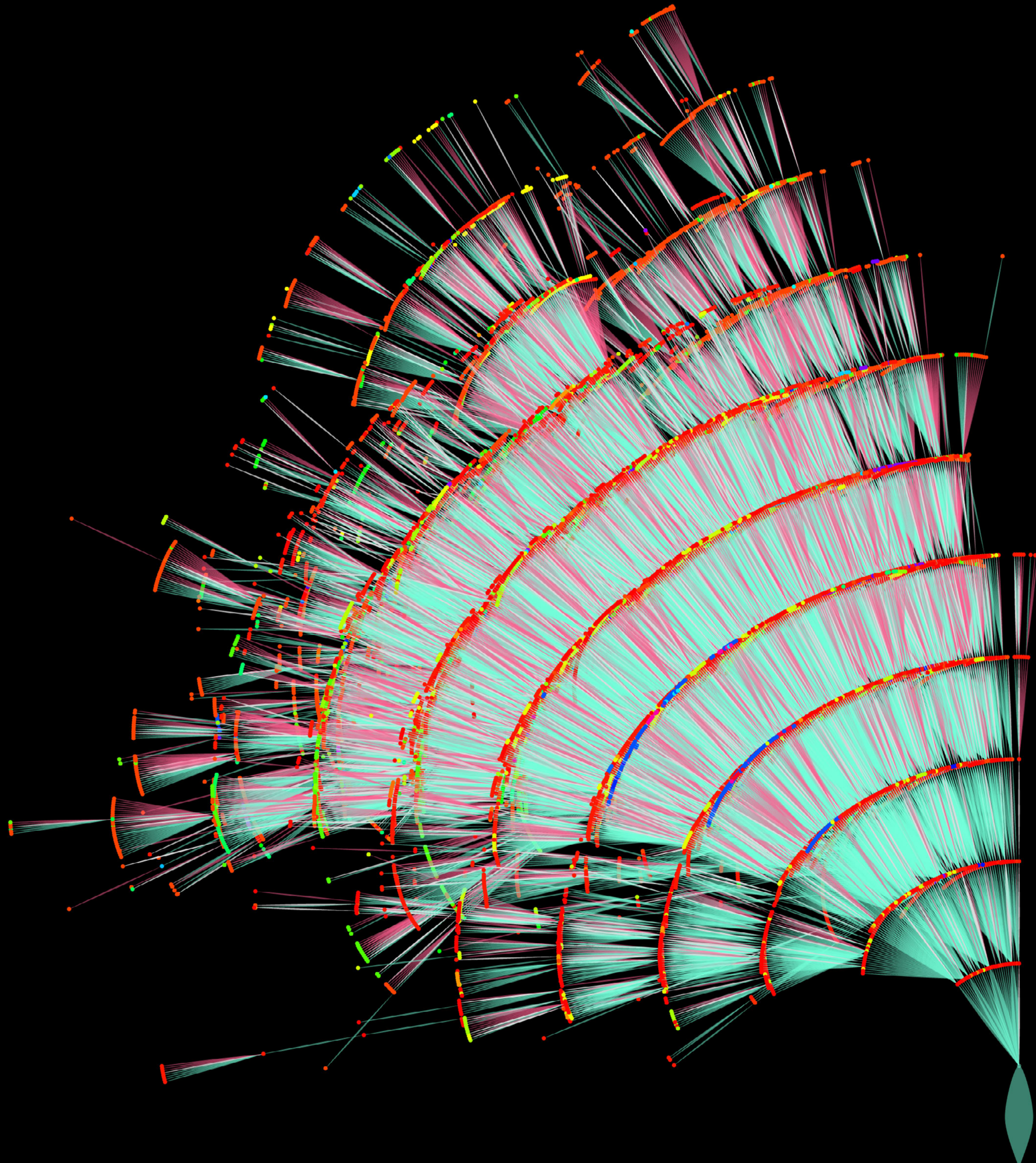


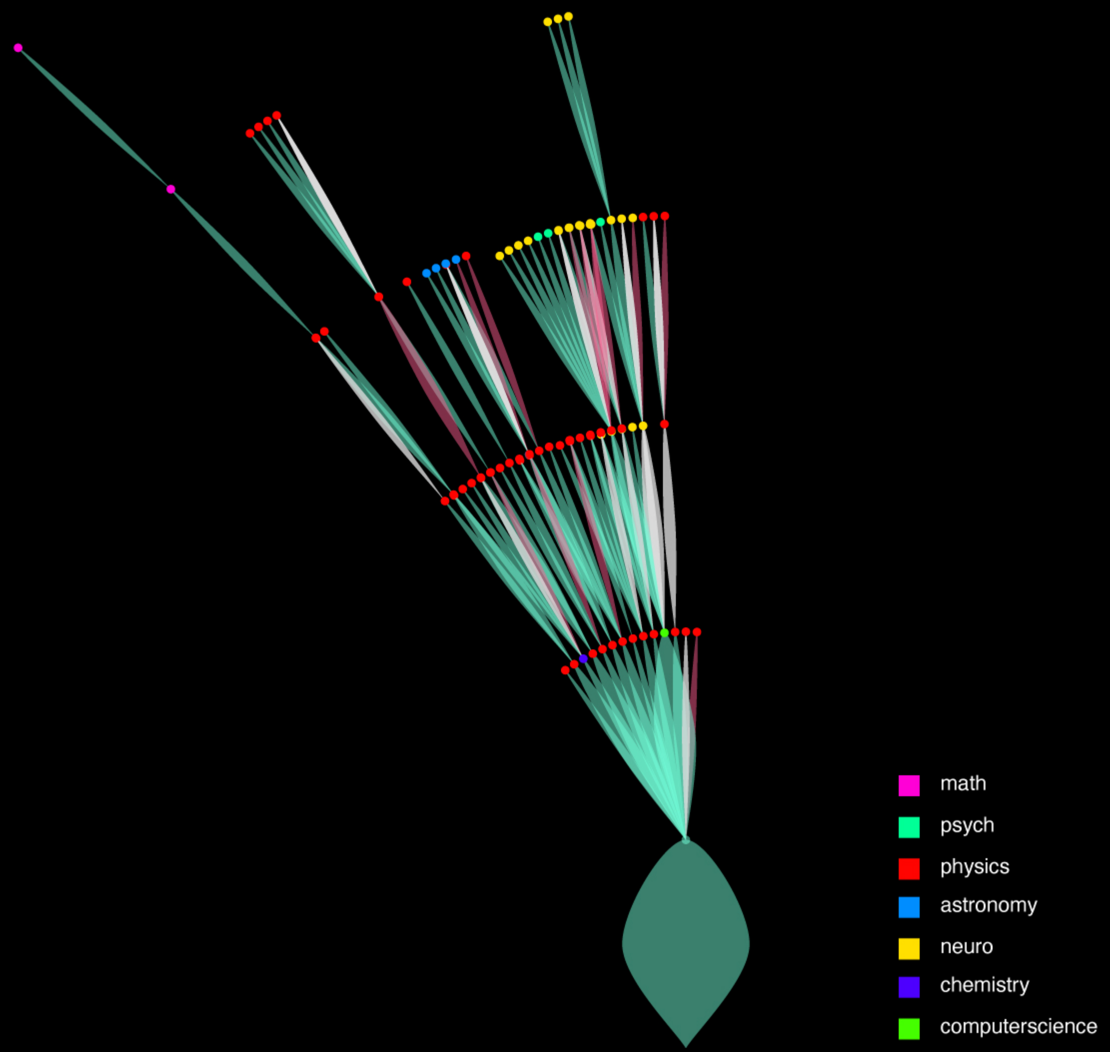
**i. Charles Sanders Peirce (1839 - 1914) was a philosopher and mathematician who pioneered the philosophy of science and reasoning. Pierce's first logical rule was that the reasoner must "wonder," and described abduction, the shuttling between induction and deduction. Peirce has 17,490 descendants mostly through his mentee Joseph Jastrow (11,742), because of Jastrow's quite reproductive mentee Clark Leonard Hull, a behavioral psychologist.**



j. William James (1842 - 1910) is considered the father of American psychology. He has 21,123 descendants, which are 51% female. Most of these grow from Granville Stanley Hall, and his students James Cattell and John Dewey were both incredibly influential. We can see quite a number of female-dominated ancestry towards the right of this tree, compared even to the curliest tree.



k. Niels Bohr (1885 - 1962) was a groundbreaking physicist, who revolutionized our understanding of the internal structure of atoms. He had 45,999 descendants, 32% of which were female. His mentees included the physicists Wolfgang Pauli and John Wheeler, and chemist Linus Pauling, along with 27 other male mentees. He had no female mentees.




1. Stephen Hawking (1942 - 2018) was a theoretical physicist who studied the properties of black holes and at the time of his death was head of the Centre for Theoretical Cosmology at Cambridge. His 101 descendants are just 16% female, sprouting rather equally in number and gender distribution from his 14 mentees. His most prolific mentee was Alan Yuille, professor in Computational Cognitive Science at John Hopkins, and had 31 descendants. He was the source for the neuro and CS students in the upper right of the tree.

## WEBSITE

**www.mentortrees.com**
We have produced a website which steps through the story of each of the prototypical trees we present here. Before introducing these trees we explain the analogy, the context, and how to read the visualizations. At the bottom of the website, users can use the sliders to adjust the tilting weight of female or male branches, the parameters M and F which we used in plotting the tree (see Drawing the Tree). The analogy is that of rebalancing the influences of our gendered academic past, to support and inform current researchers in a way that moves beyond that past.

The website also allows the user to generate the academic mentorship tree for any researcher. This supports displaying both leaves and roots, so that current practicing researchers can investigate their own gendered academic genealogy.

And finally, we put on display the straightening effect which results from chopping a descendant tree from its base. For instance, truncating Francis Galton's in this way starts with removing the root, Galton himself. We are left with two trees, one for each of his mentees, which tilt slightly less towards male. Chopping another generation leaves trees for each of their mentees, again slightly less male-tilted. And we can keep chopping to the edge of the canopy. This simulates the truncation of influence, the forgetting of founders, which results in significantly shorter trees which nevertheless stand virtually upright.

## LIMITATIONS

There are at least three limitations to the approach we are taking here. Firstly, we have opted to fan out mentees, such that some single-step mentorship relationships can result in a very high added curvature to the visualization. As a consequence the ordering of mentees at each level has an effect on the look of the tree, especially in earlier generations. This is apparent in looking at Richard Feynman's tree, where multiple first generation male mentees were tilted more than 45 degrees, and so their mentees are also extremely tilted, despite having no more male mentees than the branches which extend nearly vertical.

Secondly, this paper operationalizes gender based on the first name of the researcher. Although [3] claims the "Unknown" gender is an attempt to capture non-binary genders, the first name is only an indicator society's gendering through naming. In addition, Unknown captures androgenous names, including many Chinese and Korean names, leaving those authors' genders out of the analysis. Despite this, there is real meaning to the gender binary, in the perceptions and behaviors scientists have attached to it. Naming is only one aspect of how a society genders its inhabitants, but it has been an important one.

Finally, the data we use is crowd-sourced, drawn from registered users who input new researchers, mentors and mentees, and information. This selectivity is opaque, but we might surmise that people primarily add themselves, their colleagues, their mentors, or scholars of interest to them. There is also a process of maintenance by more involved members of the community, checking for correctness and completeness.

## CONCLUSION

Because of historical and social disadvantages, women and gender-diverse people are not represented in the roots of today's academic knowledge. The past of academia is deeply gendered, pushing the ancestry and descendant trees of nearly all researchers to the earth. The analogy begs us to visualize today's researchers' trees growing as if on the side of a cliff, liable to fall over and die. In this image, detaching from roots could mean death, or a new rebirth.

Notice how adding more weight to female researchers could help with balancing the tree. Treating men, women, and gender-diverse people equally (giving them the same tilting weight for the branches) might not actually be fair, and might even create further biases in current and future science. This is especially poignant to the extent that the methods and pursuits of a field are influenced directly or indirectly by gender. The scope and import of a science should also increase by virtue of further diversity in perspective. We hope these mentorship trees can stand straight again, built through attention to as many mothers of a discipline as fathers. That is, a more equal attention to founders, and foundational ideas, across the gender binary so long imposed by society at large.

The way in which gender interacts with the trees we have drawn makes visceral the gendered past of knowledge. It reminds us that intellectual influence, resources, and the privilege of leading labs, have only very recently been extended beyond just men. And that although institutions now look and feel more equal, our knowledge has deeper roots.

**ACKNOWLEDGMENTS**

The authors would like to thank Soojin Park, Yunting Chuang, Amy Li, and Raja Vijaya Sankar for their design, conceptualization, and prototyping work in this project.

**REFERENCES**

[1] P. A. Graham. Expansion and exclusion: A history of women in american higher education. Signs: Journal of Women in Culture and Society, 3:759–773, 7 1978. doi: 10.1086/493536
[2] L. Holman, D. Stuart Fox, and C. E. Hauser. The gender gap in science: How long until women are equally represented? PLOS Biology, 16:e2004956, 4 2018. doi: 10.1371/journal.pbio.2004956
[3] Q. Ke, L. Liang, Y. Ding, S. V. David, and D. E. Acuna. A dataset of mentorship in science with semantic and demographic estimations. arXiv preprint arXiv:2106.06487, 2021.
[4] E. Leahey. Transmitting tricks of the trade: Advisors and the development of research knowledge. Teaching Sociology, 34(2): 93–110, April 2006.
[5] M. J. Lerchenmueller and O. Sorenson. The gender gap in early career transitions in the life sciences. Research Policy, 47: 1007–1017, 7 2018. doi: 10.1016/ j.respol.2018.02.009
[6] A. A. of University Professors. Distribution of faculty by rank, gender, category, and affiliation, 2011-2012, 2012.
[7] C. M. Rawlings and D. A. McFarland. Influence flows in the academy: Using affiliation networks to assess peer effects among researchers. Social Science Research, 40:1001–1017, 5 2011. doi: 10.1016/j. ssresearch. 2010.10.002
[8] L. P. Schwartz, J. Lienard, and S. V. David. Impact of gender on the formation and outcome of mentoring relationships in academic research. 4 2021.
[9] N. Tovanich, P. Dragicevic, and P. Isenberg. Gender in 30 years of ieee visualization. IEEE Transactions on Visualization and Computer Graphics, 28:497–507, 1 2022. doi: 10.1109/TVCG.2021.3114787
[10] P. van Arensbergen, I. vander Weijden, and P. vanden Besselaar. Gender differences in scientific productivity: A persisting phenomenon? Scientometrics, 93:857–868, 12 2012. doi: 10.1007/s11192-012-0712-y
[11] M. Stefaner, D. Taraborelli, and G. L. Ciampaglia, Notabilia. [Online]. Available: https://notabilia.net/. [Accessed: 22-Aug-2022].