Academic Genealogy

Cassidy R. Sugimoto

Introduction

Academic genealogy is the quantitative study of intellectual heritage operationalized through chains of students and their advisors. In most cases, this is the doctoral student and advisor, although the concept of doctoral advisor has been liberally interpreted in some projects. For example, the Mathematics Genealogy project traces genealogy back as far as 1380—centuries before the establishment of the research dissertation or the doctor of philosophy degree (Clark, 2006). Although academic genealogy has occasionally been termed “intellectual genealogy” or “scientific genealogy,” it should not be confused with the former, which is a term used to describe any kind of intellectual influence (not necessarily a formal or institutional mentor) (Lubek et al., 1995), nor the latter, which is a term used to elevate the studies of family histories to the level of a science (Davenport, 1915).

Academic genealogy has been criticized for having low levels of generalizability and rigor, used only to honor eminent scholars or to satisfy a
scholar’s curiosity about their academic ancestors. Similar criticisms were made regarding early citation analyses, which, due to the painstaking manual data collection, were often no more than local case studies with limited generalizability or use. However, as with the rise of citation indexes, there are now sources and means for collecting large-scale data on academic genealogy that make possible rigorous studies in this area. Academic genealogy is no longer the exclusive purview of academics attempting to trace their own roots, but also of interest to those who study science—from historical, philosophical, sociological, and scientific perspectives.

The underlying assumption of academic genealogy is that disciplines are propagated through knowledge transfer activities (Abbott, 1999; Turner, 2000)—in this case, through doctoral mentoring. As noted by Kuhn (1962/1996) and others, disciplines are governed by norms or paradigms—those cultural rules that inform members of appropriate practices in the discipline. It is assumed that many of these practices are transferred both tacitly and explicitly during interactions with trusted mentors (Girves & Wemmerus, 1988). Academic genealogy provides a means to measure and analyze these interactions and to study the relationship between mentoring and “disciplining” (Abbott, 2001; Foucault, 1975/1995). As a proxy for studying the transfer of disciplinary knowledge, academic genealogy can also be used to examine interactions between and among disciplines as faculty members “disciplined” in one area move to another (Sugimoto, Ni, Russell, &
Bychowski, 2011). The migration of doctoral students in one discipline to faculty members in another has a direct impact on the knowledge landscape, potentially changing the topical trajectory of the next generation of acolytes.

As an evaluative metric, academic genealogy can serve to counteract the “neglect of silent evidence” (Taleb, 2010, p. 103), by making visible the contributions of mentors, particularly those with high mentorship fecundity: that is, those who produce a large number of protégés (Malmgren, Ottino, & Amaral, 2010). Metrics of academic genealogy (Russell & Sugimoto, 2009) serve to demonstrate and contextualize these contributions within the larger academic and disciplinary spheres. The rationale underlying the use of academic genealogy as an evaluative metric is the belief that the “most effective way for a scientist’s work to live beyond their time is for them to populate the next generation of academics with people that they have mentored…[in order that] their ideas, contributions, and views will continue to influence scientific thought” (Andraos, 2005, p. 1405). A scholar’s lifetime is finite, but his contribution is amplified, enhanced, and extended through successive generations of mentees.

The objectives of this chapter are to introduce the concept of academic genealogy, review previous studies of academic genealogy, provide a typology of academic genealogies with corresponding motivations and outcomes, outline appropriate methods for the conduct of academic genealogy, and discuss implications and future research directions.
State of existing research

Published studies of academic genealogy have examined Neuroscience (David & Hayden, 2012), Organic Chemistry (Andraos, 2005), Mathematics (Chang, 2010; Malmgren, Ottino, & Amaral, 2010), Physiology (Bennett & Lowe, 2005; Jackson, 2011), Energy Expenditure (Durnin, 1991), Primatology (Kelley & Sussman, 2007), Exercise and Sports Sciences (Mitchell, 1992; Montoye & Washburn, 1980), Pharmaceutical Sciences (Stella, 2001; Tyler & Tyler, 1992), Library and Information Science (LIS) (Marchionini, Solomon, Davis, & Russell, 2007; Russell & Sugimoto, 2009; Sugimoto, Ni, Russell, & Bychowski, 2011), and Psychology (Boring & Boring, 1948; Lubek et al., 1995; Newton, 1995; Robertson, 1994; Williams, 1993). Online projects range in size and coverage.

Bennett and Lowe (2005) include a link from their article to a companion website recording the academic descendants of a single individual. The University of Illinois at Urbana-Champaign hosts a database for Chemists,¹ and the Department of Computer Science at the University of Texas at Austin collects academic genealogy information for the Artificial Intelligence community.² The largest

---

¹ http://www.scs.illinois.edu/~mainzv/Web_Genealogy/index.php. At the time of writing, the database contained data from the faculty of UIUC and nine other chemistry departments. In total, the database included 1,070 advisors, 1,032 students, and 44 scientists (as detailed in a personal email communication from Vera Mainz.)

² http://aigp.eecs.umich.edu/about. At the time of writing, this site claimed to have 16,741 researchers, from 1,569 schools and 52 countries.
single-discipline database is the Mathematics Genealogy Project, hosted by the Department of Mathematics at North Dakota State University. As of September 1, 2012, the project contained data on 163,855 mathematicians.

Two projects that began as single-discipline databases have expanded to include other disciplines. The MPACT Project, hosted jointly by Indiana University Bloomington and the University of North Carolina at Chapel Hill, began as a study of LIS, but contains, at the time of writing, data on 9,037 individuals across 217 disciplines, 319 schools, and 156 countries. The Academic Family Tree, an extension of the neuroscience-focused Neurotree.org (for an analysis of which, see David and Hayden, 2012), is an interdisciplinary project currently listing 28 fields of study. The number of people in the database per area ranges from 37,387 for Neuroscience to 51 for Ingestive Behavior. The content is entirely user-supplied and, although the trees are networked between fields, each field has its own portal. The growth rate for Neuroscience is estimated at 150 new additions per week, and the web site maintains a page demonstrating these growth rates. However, this source, as with the previously mentioned ones, is heavily biased towards a single discipline.

**Types of academic genealogy**

---

3 [http://genealogy.math.ndsu.nodak.edu/](http://genealogy.math.ndsu.nodak.edu/)
5 [http://academictree.org/about.php](http://academictree.org/about.php)
Academic genealogy has experienced some of the same criticisms that have plagued other scientometric methods; namely, that these are mere navel-gazing exercises done for personal pleasure or to celebrate an individual. Ego and honor do play a part; however, there are a number of motivations for conducting this type of research and, more importantly, a variety of research questions that can be answered by the construction of an academic genealogy. Therefore, I propose a five-part typology for academic genealogies: honorific, egotistic, historic, paradigmatic, and analytic. Each is briefly described below. It should be noted, however, that these are not exclusive categories; most academic genealogies function in at least two of these ways.

Honorific

Many academic genealogies focus on the descendants of a single individual, in order to honor this individual and demonstrate the scholar’s impact upon generations of academics (e.g., Bennett & Lowe, 2005). At the conclusion of one such study, readers were “encouraged to explore their own academic legacies as a way of honoring those who prepared the way for us” (Jackson, 2011, p. 120). Honorific genealogies are typically prepared as part of a Festschrift, workshop, or conference aimed at honoring a canonical, eminent, or recently deceased member of a field. Those honoring canonical figures are likely to be historical (e.g., Zittoun, Perret-Clermont, & Barrelet, 2008), while those honoring eminent or
recently deceased figures tend to focus on one generation of descendants and contemporary accomplishments.

*Egotistic*

Egotistic academic genealogies take as their point of departure a single individual or small set of scholars and trace antecedents. Typically, the author of the study conducts a personalized search to understand his academic ancestors (e.g., Durnin, 1991; Jackson, 2011; Robertson, 1994; Williams, 1993), although it could also involve a department, school, or other group of individuals tracing their collective past. Curiosity and community building (Newton, 1995) can motivate egotistical genealogy. There is also a sense of establishing oneself or one’s unit by delineating connections with prestigious predecessors. As Lubek et al. (1995) noted regarding academic genealogies of psychologists:

> Amateur genealogists, in pursuing their family trees, often hope to uncover some illustrious ancestor who might add a measure of glamour to their lineage. Collective genealogies, like those of departments of psychology, may serve similar functions. The recognition that we have ‘descended,’ in a manner of speaking, from some celebrated figure in the history of psychology may infuse our circumscribed daily activities with some transcendental value, with the recognition that although we are ordinary workers in the fields of Kuhnian normal science, we are
nonetheless part of an historical endeavour led by persons of some renown. (p. 52)

Demonstrating that one descended from Wilhelm Wundt (Lubek et al., 1995; Robertson, 1994; Williams, 1993) is a form of credentialing. Demonstrating that one of your forbearers won the Nobel Prize (Tyler & Tyler, 1992) is a way to flaunt your academic genes.

**Historical**

The historical nature of an academic genealogy is perhaps unavoidable, regardless of the type of genealogy constructed. Explicit historical genealogies are those which use academic genealogy as a platform for a historical analysis. In practice, these often involve a historical narrative around academic influences in an honorific genealogy and potted histories of each advisor in an egotistic genealogy (e.g., Durnin, 1991; Jackson, 2011; Tyler & Tyler, 1992). The underlying motivation of a historic genealogy is not to honor or confer prestige on any individual, but rather to describe the development of a field or area of study through academic lineages.

“Founding Fathers” are often the centerpiece of such studies. In empirical studies, the goal is to identify founding fathers by tracing current members of a field back to a few key individuals (e.g., Boring & Boring, 1948). For example, Lubek et al. (1995) examined the academic genealogy of faculty in five Canadian
psychology departments and found that 75% of the faculty members could be traced to nine “pioneer figures”. In this way, one can empirically identify the canonical figures in a field. Alternatively, canonical scholars can serve as the starting point of the analysis and the development of the field can be demonstrated by tracing academic descendants (e.g., Tyler & Tyler, 1992).

Other studies revolve around contemporary elders (e.g., Mitchell, 1992); that is, current eminent scholars in the field. Tracing the ancestry of these individuals provides insights into the formation of the field. These studies can be used to show the growth of an intellectual area (e.g., Kelley & Sussman, 2007) and the networked nature of such studies can throw “into relief several facets of our collective history” (Lubek et al., 1995, p. 66). This collective history can have pedagogical value for neophytes. As Andraos (2005) commented: “[The] predictability of what is important and worth looking into as a viable research question or area markedly improves when one knows the chronology of ideas and associates themselves with the key players in shaping those ideas or at least their direct scientific descendants” (p. 1405). Presenting new doctoral students with this research or requiring them to search their own pedigrees can be a useful initiation into the discipline.

*Paradigmatic*
Paradigmatic genealogy can be used to not only define formal relationships between people, but also to study the extent to which knowledge and epistemic practices are transmitted through these relationships. It has been cautioned that institutional genealogy (i.e., tracing genealogy through formal lineage) should not be conflated with intellectual genealogy (i.e., tracing how ideas move through mentors, collaborators, and colleagues over time) (Lubek et al, 1995). However, as De Mey (1982) remarked on Zuckerman’s (1977) study of Nobel Prize laureates:

Though innovation might seem to be coupled to erratic and unpredictable movement, there are also impressive chains of continuity in science through master-apprentice relationships…Apprentices need not remain in the area of their master to exemplify this continuity…The conservation and transmission of a tradition of scientific leadership and innovativeness has again both sociological and cognitive aspects. Through their position in the social organization of science, masters can introduce their protégé more effectively into the leading circles and innovative areas…. [high quality mentoring] is probably the most effective way of installing cognitive structures which remain intact when the apprentice migrates to other fields or encounters revolutionary turmoil. (p. 146)

Studies have supported the claim of “impressive chains of continuity,” finding evidence of shared methods or epistemology across academic generations (e.g.,
Jackson, 2011; Kelley & Sussman, 2007; Robertson, 1994). However, Lubek et al. (1995, p. 65) warned that genealogical trees can “sometimes give a false picture of continuity” when they fail to take into account the “model of discontinuous discoveries in science” and urged for more “critical approaches…sensitized to the tension between continuity and discontinuity” to address this limitation.

Many studies of a paradigmatic nature have lacked rigor, merely exploring the “self-evident” replication of theories, methods, and practices in an academic lineage. However, the introduction of robust topic modeling techniques and large-scale dissertation data provide a new lens for exploring paradigmatic genealogies. For example, Sugimoto, Li, Russell, Finlay, and Ding (2011) analyzed the topics in LIS dissertations over an 80-year timespan by applying topic modeling techniques to dissertation titles and abstracts, revealing the main topic areas by decade. Given that each of the titles and abstracts is matched with a student who is matched with an advisor, one can apply network analysis to this to reveal underlying topic communities (Yan, Ding, Milojević, & Sugimoto, 2012) within and across academic family trees. Qualitative analyses might also be used to reveal the transmission of ideas and practices from one academic to another—interviews with a cohort of students advised by a single advisor might provide insight on the nature of knowledge, skills, and practices transferred through formal mentoring relationships.
Analytic

The growth of large-scale databases and rigorous statistical analyses has made possible a new type of genealogy. Analytic academic genealogies scientifically address explicit research questions and provide evaluative assessments of scholars and scholarship. Such studies are typically descriptive and evaluative in nature and, in some cases, predictive. Early analytic genealogies sought to establish sets of metrics for quantifying academic genealogy—calculating the number of times an individual served as an advisor and committee member and various weighted approaches to these calculations, as well as metrics for calculating the breadth and depth of academic lineage (Marchionini et al., 2007; Russell & Sugimoto, 2009). Subsequent studies have also sought to examine distances in relationships among individuals in the family tree (e.g., David & Hayden, 2012). Metrics of genealogy have been used to investigate the relationship between quantitative advising metrics and other metrics of academic productivity (Sugimoto, Russell, Meho, & Marchionini, 2008). The evidence suggested that the relationship was not linear, but rather individual and dependent upon the time in the academic lifecycle. Malmgren, Ottino, and Amaral (2010) similarly noted that mentorship fecundity (that is, the number of protégés a mentor trains) for mathematicians had different rates over the course of the academic lifecycle: fecundity was higher than expected for the first two-thirds of a high performing academic’s career and lower than expected in the last third of the academic’s mentoring career. That is, a
student mentored in a high-performing scholar’s early years would be more likely to have a large number of academic progeny than those produced later.

Many indicators of interdisciplinarity are static—describing the distribution of units (e.g., authors or documents) across a discipline at a given point in time. However, interdisciplinarity can also be measured using the backgrounds of advisors represented at various points in time, providing a diachronic indicator of the heterogeneity of a field. The use of academic genealogy to study interdisciplinary was introduced by Sugimoto, Ni, Russell, and Bychowski (2011), who investigated the discipline in which advisors and committee members on LIS dissertations had received their degrees. This work demonstrated that the degree of the advisor had a direct impact on the intellectual content of the advisee’s dissertation—when an advisor received their degree in one discipline and then moved to another, the subsequent advisee’s use of interdisciplinary research was significantly different from those mentored by “within-field” advisors. This provides evidence of the transmission of knowledge through advisee-advisor chains and the possible impact of interdisciplinarity in advisorship networks. The diachronic nature of academic genealogy studies provides the added benefit of being able to study not only the heterogeneity of a discipline, but also its permeability at various points across time (Klein, 1996). Depictions of the dynamic nature of the discipline may provide better descriptions of “life cycle of scientific specialties” (de Mey, 1992, p. 148).
Network science provides additional opportunities for analytic academic
genealogy. In some fields, it is common for people to do post-doctoral training
with their academic grandfather, that is, their advisor’s advisor (Andraos, 2005).
This form of academic inbreeding produces strong networks and can be
“particularly effective in securing consistent and credible recommendation letters
from people who know one another professionally as well as personally”
(Andraos, 2005, p. 1404). Studies have also shown that doctoral origins work to
create strong ‘power-cliques’ in scholarly communication. Yoels (1971) found a
disproportionate number of editors coming from a few doctoral programs and
showed that this distribution was not proportional to the rate of production of
doctoral graduate across all programs; furthermore, Yoels revealed that the
editorial boards’ doctoral origins tended to align with the doctoral origin of the
editor-in-chief. These studies suggest that the “Matthew effect” (Merton, 1973)
does not only result in cumulative advantages for the advisor, but also residual
benefits for the student; quite simply, “student[s] who have trained with any of the
key scientific contributors or their immediate descendants are highly sought after
as the next generation of academics” (Andraos, 2005, p. 1405). Many direct
correlations have been identified with respect to mentorship fecundity, including
publications, citations, and membership in prestigious associations (Malmgren,
Ottino, & Amaral, 2010; Sugimoto, Russell, Meho, & Marchionini, 2008). Future
studies of academic genealogy could identify more ways in which the
productivity and success of the doctoral student are directly related to the academic network in which the student was trained.

Method

Academic genealogies appear, on the surface, to be fairly simple: create links between individuals based on formal academic affiliations. However, several approaches can be taken and have various advantages and disadvantages. There is also the possibility of error and misinterpretation if variables are not explicitly operationalized. As Boring and Boring (1948) noted in their construction of an academic genealogy of psychologists: “Our experience with the living shows that we may easily have made some errors with the dead” (p. 527). The next section reviews the methods of academic genealogy, including initializing a search, operationalizing links, identifying data sources, and visualizing results.

Initializing

Academic genealogies typically begin with a purposive sample of one or more individuals. In the case of an honorific genealogy, a single individual is selected and all the descendants of this individual are identified (e.g., Bennett & Lowe, 2005). In the case of an egotistic genealogy, a contemporary scholar (typically the researcher of the study) identifies his own lineage (i.e., his advisors, his advisor’s advisor, and so on until no more information can be found) (e.g., Jackson, 2011;
Williams, 1993). The goal of a single-individual study is either to demonstrate the importance of a key scholar or to trace a personal genealogy.

Multiple individuals are usually selected when the motivation is to study the growth and development of a field or discipline. In similar fashion to single-individual studies, these can be done either from the perspective of descendants or antecedents. Certain canonical authors may be chosen and their collective genealogical descendants traced to outline the parameters of the discipline. Conversely, canonical authors could be identified by studying the genealogies of all current members of the discipline (e.g., Lubek et al., 1995). Contemporary scholars may be identified using prolific authors in a high impact venue (e.g., Mitchell, 1992; Montoye & Washburn, 1980), doctoral graduates from representative schools in the discipline (e.g., Sugimoto, Ni, Russell, & Bychowski, 2011), or current faculty members at representative schools or in representative associations.

Great care should be taken in selecting the individual(s) used to “seed” the genealogy. Given that the majority of academic genealogy studies employ purposive sampling, caution should be taken when generalizing the results (unless the goal is explicitly honorific or egotistic). Studies of an historical, paradigmatic, or analytic nature should always justify the sample and, when a census is not possible, explore possible sampling options.
Operationalizing links

It is critical that any academic genealogy study begin with a clear operationalization of the criteria necessary to establish a link between student and mentor. For the strictest studies, this implies a degree of formality in the relationship, typically as academic advisor, dissertation chair, major professor, or supervisor: that is, the person formally and primarily responsible for guiding the student through the process of completing a doctoral degree (e.g., Kelley & Sussman, 2007; Lubek et al., 1995; Mitchell, 1992; Robertson, 1994; Williams, 1993). However, this operationalization is often not followed in strict fashion, particularly when academic genealogies reach beyond the modern doctoral education system and the modern dissertation.

There is some criticism of equating an academic advisor with a mentor and with the exclusive use of doctoral advisors to denote links between students and mentors. It is argued that many individuals can serve as the primary mentor for an individual and that individuals are often influenced by a “mentoring constellation” (Sugimoto, 2012) rather than an individual. Despite this, general consensus and empirical evidence (Sugimoto, 2012) suggest that the doctoral advisor is the primary advisor in most cases and is an adequate proxy when examining large-scale genealogies. It may be useful in micro-level analyses, however, to look beyond the dissertation advisor for other forms for intellectual influence upon a scholar, be it postdoctoral advisors (e.g., Andraos, 2005; Bennett
& Lowe, 2005; Jackson, 2011; David & Hayden, 2012), collaborators, or other informal mentors (e.g., Zittoun, Perret-Clermont, & Barrelet, 2008), and various levels of students or “pupils” (Bennett & Lowe, 2005; Durnin, 1991).

The links in the genealogy end when no additional verified information can be located on an advisor/advisee. Other potential discontinuities in the family tree are the rare advisors who come from non-academic origins (‘‘self-starters,’’ as Lubek et al. [1995] label them), and students who bear no academic children themselves (largely as a result of going into a non-academic position). There is also a practice of maintaining an intra-disciplinary perspective and taking as a “point of disciplinary discontinuity” any branch in the tree where someone was trained by someone from outside the discipline (Lubek et al., 1995, p.53). The goal of these (typically historic or paradigmatic) genealogies is to create a disciplinarily homogeneous genealogy. This can be useful in describing the extent and breadth of disciplinary histories. However, it has been suggested that academic genealogies should not avoid these discontinuities; but rather use them to inform our understanding of the birth, maturation, and interaction of disciplines (Sugimoto, Ni, Russell, & Bychowski, 2011). Such interdisciplinary genealogies integrate homogeneous academic family trees into a “canopy of trees” (“The Academic Family Tree,” n.d.) that display the interconnectivity of doctoral education and, by extension, of knowledge.
Identifying data sources

The most reliable source for information on doctoral advisors and committee members is the actual dissertation. Ideally, names and roles will be typed underneath or next to a signature line on the cover page. If neither a cover page nor a typed name can be found, and if the signatures (assuming they are available) are indecipherable, the next best source is the acknowledgment, where the authors will often explicitly thank their advisors and committee members. However, acknowledgments can be ambiguous, as when the author thanks a number of faculty members without specifying their roles. Acknowledgments can also shed light on transitions: the author may thank multiple advisors who guided them during one point in their doctoral career, but who did not sign the final dissertation for various reasons (e.g., leaving the university, death, etc.). In some cases, the final advisor may be more of an administrative signatory, rather than a true intellectual counsel. However, operationalization must remain strict as the researcher cannot be responsible for identifying “true mentors” in each situation (see earlier section on operationalization).

The most comprehensive data source for dissertations is ProQuest’s Dissertations and Theses database, “covering 40% of all dissertations from major universities” (Andersen & Hammarfelt, 2011, p.374). For the years 1848 to 2009, this includes about 2.3 million dissertations conferred at 1,490 research institutions in 66 countries (Ni & Sugimoto, 2012). However, ProQuest is heavily
biased towards English-language degrees conferred at North American universities, introducing a bias for global studies. An additional limitation for historical, paradigmatic, and analytic studies is the lack of information on the school or department in which the dissertation was conferred. Although recent dissertations contain this information, most do not. Therefore, to gain a "disciplinary sample", a researcher must rely on ProQuest Subject Categories. While these may be useful for identifying some disciplines, they are poorly suited to identify highly interdisciplinary areas. Sugimoto, Russell, and Grant (2009) detail the substantial number of false positives generated when searching by Subject Categories and then validating by examining the cover page of the dissertation. There are also false negatives that can only be validated if one begins with an accurate sampling frame, generated either from the graduating school or a directory of recent graduates. In short, ProQuest Subject Categories are mere proxies for disciplinarity, and inferences based on them should acknowledge the limitations.

The ProQuest database can be useful in identifying a sample, but it has certain limitations for conducting academic genealogies: only half of the dissertations contain information on advisors, and even fewer contain information about committee members. In addition, the majority of dissertations containing advisorship information are from the last 20 years, making the study of multiple generations difficult. Nonetheless, ProQuest does have some advantages. For
many dissertations, it offers a “Preview” that provides the first 24 pages of the dissertation—including the cover page and acknowledgment section where information for an academic genealogy can be found. Unfortunately, this means manual data collection and entry—a tedious and time-consuming activity for a large genealogy project. In addition, “Previews” tend to be available for more recent dissertations. Therefore, conducting an academic genealogy before 1950 requires using the physical resources of a library—either from a local collection or via interlibrary loan. Researchers should keep in mind that many of these items will only be available on microfiche, so access to a microfiche reader will be necessary.

Scholars may also be interested in exploring the data provided by open access sources. For dissertations, the largest of these is the Networked Digital Library of Theses and Dissertations (NDLTD),7 which contains more than one million theses. This site also lists a number of other electronic thesis and dissertation (ETD) sources, many of which are country specific. Country-specific sources are often government-mandated and therefore fairly reliable and current.

In many cases, the dissertation itself does not provide the information needed for the genealogy; other sources must be consulted. Several studies have used direct communication with scholars listed in the genealogy (Kelley & Sussman, 2007; Lubek et al., 1995; Montoye & Washburn, 1980), standard

7 http://www.ndltd.org/find
bibliographic reference material (Andraos, 2005; Robertson, 1994; Williams, 1993), obituaries (Williams, 1993), and solicitations for people to contact the authors with corrections or additional information (Montoye & Washburn, 1980). Recent projects have utilized the Web to solicit and crowdsourc this information. For example, two large academic genealogies on the Web, the Mathematics Genealogy Project and Academic Family Tree, use crowdsourced content as their primary data source.

User-generated content, whether via the Web or personal contact with scholars has large potential for error. There is a tendency to “obscure or to disown inadvertently a non-prestigious institutional supervisor in favour of a highly visible intellectual mentor” leading to “mythical creation of a genealogical link” (Lubek et al., 1995, pp.59-60). In some cases, these are unintentional: previous studies (Boring & Boring, 1948) have based their analyses on vague questions such as, “Who was it who influenced you most in psychology up to the time you got the Ph.D?” or even the more explicitly phrased, “Whose student were you?” However, both questions tend to yield answers that are not precisely in keeping with the notion of doctoral advisors as understood in an academic genealogy. Biographies can often provide erroneous information based on historical myths, or provide ambiguous statements (e.g., “x was a student of y”) that do not identify precisely the relationship between the individuals. Therefore, validation studies (e.g., David & Hayden, 2012) are necessary to ensure the accuracy of
crowdsourced academic genealogy data. Lastly, the often highly emotional relationships between students and their advisors can cause problems—in cases where the relationship ended on poor terms, either the student or advisor may be unwilling to acknowledge the relationship. In such instances, the student often declares that another member of the dissertation committee was more of a “mentor” than the official chair. This brings up issues of ethics pertaining to genealogies, since anonymity is impossible. However, given the public nature of the data, studies typically do not require permission from institutional review boards (IRBs) or those named in the genealogy.

**Visualizing results**

Published visualizations of academic genealogies mimic the structure of genetic family trees: names are provided, beneath these names a place and date, vertical lines connect “parents” to “children,” and academic “siblings” are connected via horizontal lines. There is software available for the construction of academic genealogies (e.g., Kelley and Sussman [2007] used Microsoft Visio), although many scholars simply visualize by hand or use word processing software. Some academic genealogies contain historic birth and death dates (e.g., Stella, 2001; Tyler & Tyler, 1992); however, it is more common to list the date on and institution at which a doctoral degree was conferred (e.g., Andraos, 2005; Mitchell, 1992). There are alternative visualization approaches; for example:
listing a series of advisors in a table, without any vertical lines (e.g., Robertson, 1994), providing photographs of each advisor (e.g., Jackson, 2011), crafting a descending matrix in tabular form (e.g., Bennett & Lowe, 2005), or extending the tree metaphor by generating a tree with the listing of descendants in branches (e.g., Bennett & Lowe, 2005). However, there are limits to what can be done in print (or traditionally print-based) publications.

Online projects provide more flexibility in generating visualizations of academic genealogies. The MPACT project uses the DOT language (rendered by the open source software Graphviz) to provide the directed graphs that represent the academic genealogies. This allows automatic rendering of the representations whenever new information is added to the database. Furthermore, users can interact with the visualization by selecting any node in the graph and generating a new graph based on this node (where all descendants are visualized and each ancestor is provided). Neurotree visualizes the family tree through a set of PHP scripts and also includes a brief biographical information page for each individual (David & Hayden, 2012). However, there are limitations, as visualizations only provide links between advisors and not between committee members or additional intellectual influences. Other online projects have used modifications upon family genealogy software (e.g., Bennett & Lowe, 2005) or simple web programming language, using hyperlinks to navigate the genealogy (e.g., the Mathematics Genealogy Project). Many of these online projects offer personalized genealogy
Conclusion

Dissertations are an underutilized source of information on the growth and evolution of science (Andersen & Hammarfelt, 2011). What we know about the emergence and development of disciplines, the diffusion of knowledge, and the evolution of science has been predominantly charted through analysis of journal articles. Symbolic capital is calculated by ascertaining the number of publications, the impact factors of the venues in which one is published, and the number of citations one receives. However, as noted by many (e.g., Cronin & La Barre, 2004; Larivière, Archambault, Gingras, & Vignola Gagné, 2006; Sula, 2012), journal articles do not tell the entire story or, for some disciplines, even a significant part of the story. In addition, contributions to scholarship made by those who engage in substantial amounts of mentoring are largely overlooked. Academic genealogy provides a means to recognize and metricize these contributions and to provide a new lens for understanding knowledge diffusion.

This chapter provides the first typology of academic genealogies and the outcomes of each. Although of limited scholarly value, there is certainly a place for honorific and egotistic genealogies: these serve to highlight the importance of mentoring in an academic career, contextualize one’s work in the larger scholarly
sphere, and can be used as pedagogical tools. However, most scientific value is likely to be derived from historical, paradigmatic, and analytic genealogies, which provide opportunities to systematically depict disciplinary histories, interdisciplinarity, and knowledge diffusion processes.

Recent developments in network science (see West & Vilhena’s chapter) and big data initiatives should serve to enhance the academic genealogy research front. Many timely and important questions can be addressed by positioning academic genealogy research in the context of network science, such as examining bias and power-cliques in doctoral mentoring networks. For example, a recent report on National Institutes of Health (NIH) funding demonstrated that white applicants were awarded a disproportionate number of awards (Ginther et al., 2011). A social networking explanation was given by an *Economist* summary suggesting that advisors may unconsciously choose advisees of their own race, providing a perpetuating advantage in scholarship (“Racial Discrimination in Science,” 2011). This could be empirically tested by adding race and other social variables to an academic genealogy analysis, which is particularly important given the change in doctoral student demographics in recent decades (Thurgood, Golladay, & Hill, 2006). In addition, anecdotal evidence suggests that pedigree and academic genealogy create power-cliques in scholarly communication and the academic labor market, with the ability to constrict academic and disciplinary mobility. Analytic genealogy could be used to examine, explain, and hopefully
eradicate these and other violations of the universal norms of science (Merton, 1973).

Mentors and doctoral students leave traces of their scholarly interaction in acknowledgments (Cronin, 1991) and in patterns of co-authorship (Sugimoto & Cronin, 2012; Sugimoto, 2011). However, for many advisor-advisee relationships, there is no trace besides the cover page of the dissertation. Bringing these relationships to light could reveal underlying paths of knowledge diffusion, academic mobility, and disciplinary developments, particularly when academic genealogy is used in conjunction with other scientometric and social variables to provide a multi-dimensional view of the scholarly landscape.

Acknowledgment

Work on this chapter was enabled by the Faculty Research Support Program at Indiana University Bloomington.

Cited References


