University of California

Los Angeles

Between Genre and Paradigm: The Development of U.S. Social Sciences, 1900-1925

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Sociology

by

Brooks Ambrose

2016

Abstract of the Dissertation

Between Genre and Paradigm: The Development of U.S. Social Sciences, 1900-1925

by

Brooks Ambrose

Doctor of Philosophy in Sociology

University of California, Los Angeles, 2016

Professor Lynne G. Zucker, Chair

Lorem ipsum.

The dissertation of Brooks Ambrose is approved.

Gabriel Rossman

Barbara Lawrence

Jacob Foster

Lynne G. Zucker, Committee Chair

University of California, Los Angeles

2016

Lorem ipsum.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Lorem ipsum.

Vita

2006	B.A. (Sociology), Haverford College.
2011	M.A. (Sociology), UCLA.
2011–2012	Teaching Assistant, Sociology 210A-210B Intermediate Statistical Methods, Cameron Campbell and Gabriel Rossman, UCLA.
2012	Teaching Assistant, Sociology 180A: Sociology of Cheating, Rick Grannis, UCLA.
2012–2013	Teaching Associate, Sociology 101: Development of Sociological Theory, César Ayala and Stefan Bargheer, UCLA.
2012–2013	Research Assistant, Cambodia DSS, Patrick Heuveline, UCLA. Prepared and analyzed demographic surveillance system data from 2,000 households on a Cambodian island. Wrote programs in R and Stata to clean and format birth, death, and migration statistics.
2013	Research Assistant, Democracy at War, Stefan Bargheer, UCLA. Measured historical rates of attention to Alexis de Tocqueville versus Frederick Jackson Turner in 5 social science journals using JSTOR's Data for Research service. Wrote program in R to identify false positives.
2014–2015	Teaching Fellow, Sociology 1: Online Introduction to Sociology, William Roy, UCLA

2014–2015	Graduate Assistant, D-Lab, UC Berkeley. Coordinator of Computational Text Analysis Working Group. Consultant in social network analysis and text analysis. Instructor in R programming.
2015–present	Instructor, School of Information, Masters of Information and Data Science, UC Berkeley. Lead live online classes and develop course material for W201 Research Design and Applications for Data and Analysis under supervision of Steve Weber.

Publications

Lorem ipsum.

CHAPTER 1

Foundations

1.1 Introduction

This collection of studies seeks to accomplish a simplistic goal. I wish to enumerate schools of thought in the historical record of U.S. social science scholarship. While this is little more than a census-taking exercise, what will become interesting about it is what we must learn when taking operationalization seriously. By attempting to resolve difficulties that arise between ontology (what we assume is there) and observation (what we see when we actually look), I am forced to update my sociological imagination. While the studies below are empirical their purpose is theoretical in the imaginary sense. I do not seek an unbiased view of the world, but merely to move in an unbiased direction.

In Chapter 2 I begin by achieving clarity on what I think a school of thought is ontologically, a particular set of patterns among personalities, cultures, and social structures. From the perspective of this orienting theory in Chapter 3 I take a critical look at the work of three theorists of intellectual development, Jen Lena, Thomas Kuhn, Harry Collins. In Chapter 4 I motivate the selection of the U.S. social sciences as a case of intellectual development. In Chapter 5 I introduce the source material for the studies that follow. In Chapter 6 I execute five brief empirical studies that test different aspects of my expectations about intellectual communities and their development. In Chapters 7 and 8 I summarize findings

and discuss limitations of the study. In Chapter 9 I conclude by updating the theory building of Chapters 2 and 3 with the results of the studies.

1.1.1 Mechanisms

Struggles to establish disciplines as professions were hard fought, harder still because each was a contest on cultural, cognitive, and social fronts. While all human behavior can be analyzed as consisting in different ratios of all three components, the institutional development of professions procedes in a conditional order.

In sociology to declare something an institution is to ask how patterns of human behavior became regular and to excavate the hidden mechanisms that maintain that regularity within limits. Religious practice is regulated by the church, political practice by the state (**???**172). Such large scale organizations eclipse the cultures that provide the content around which they initially organized. They assure their own cultural inputs, and may be open or closed with respect to novel culture.

Paradigms reflect an advanced stage of institutional development in the history of disciplines. They presuppose professions, which provide the organizational resources necessary to enforce conventions. Though paradigms are often associated with ideas.

Culture preceeds cognition in the sense that practices develop tacitly before they are "recognized" explicitly, and indeed recognition is, while often transformative, not actually necessary. On the contrary a cultural thing, whether an object or a practice, must already exist for it to be recognized. Likewise cognition, especially classification, preceeds social control Cultural processes center on human interaction with meaningfully constituted objects. This far reaching concept is usefully characterized in the tradition of Geertz where culture exists at the

intersection of symbolically organized thought and concrete practice. Since Geertz priority has been placed This polarity between real and ideal may be adapted in a tr

A discipline must cohere culturally before it can professionalize, and this process occurs in four stages. First, a prototypical set of productions–articles, books and lectures–had either to be found or invented, and the patterns they established had to be reproduced without the benefits of consistent resources or conventions. Second, assembeled productions had to become recognizable; they had to be labeled and grouped together according to a consistent symbolism, and that symbolism had to be learned within a broader milieu. Third, after a symbolic index could be taken for granted, disciplines would be stillborn if they could not maintain productivity. Would-be disciples had to produce enough new material to support an audience intially of peers and then of larger publics. Fourth, to emerge as professions, disciples had to have a reasonable chance of being awarded scarce resources within the academy. The ability to attach the disciplinary label to departments and professorships marked the beginning of a viable adolescence. If a discipline could aqcuire the machinery of education it could control the presumption of its own legitimacy, at least among new generations of students.

Tabs/ch0/devstages.tex

This study is about the second stage of development, recognition. I assume that protoypes of disciplinary knowledge were readily available in the United States by the end of the 19th century, and that the challenge for disciples was to relabel what had already been accomplished in an effort to create an occupation out of what was formally a personal undertaking, an obsession or a pastime.

1.1.1.1 Cultural Sensemaking

In order to regulate their own creative activity, disciples curated prototypes into sets that drew a boundary, however roughly, around the tacit definition of what they thought they were doing. These acts of sorting allowed prototypes to be organized without requiring the organizer to explain the rules of their order.

Wherever personal sets overlapped the intersection would form a smaller set of higher status. This tendency to overlap one's personal cultural toolkit with that of others was a form of deference to peers as well as to common culture. This allowed disciples, without necessarily intending, to both claim membership in the discipline and indulge in their idiosyncratic variations on the essence of the discipline. Indeed idiosyncrasy would be tolerable only to the extent that a scholar could in the same breath genuflect to what was already understood.

While new references to the overlapping set reinforced its status they also winnowed its content. Paradoxically by begging to peers that they recognize some novel prototype scholars would have to also pay homage to the core. They could not then elevate their own interests above that of the growing stock of taken for granted knowledge, since introducing something could not garner attention without attaching it to something old.

The rejoinder to the claim that nothing can overtake the core is found in Kuhnian accounts of scientific revolution. Yet the theory of referential coherence could not be further from that of paradigmatic coherence. The great advantage of using references as the currency of disciplinary communication and exchange is that their meanings can be assumed, and the inevitable disagreements over interpretation can be easily ignored so long as conversation floats just above matters of substance. Paradigmatic agreement is not culturally, that is autonomously, possible unless one makes a very strong empirical convergence assumption. Exogenous reality must

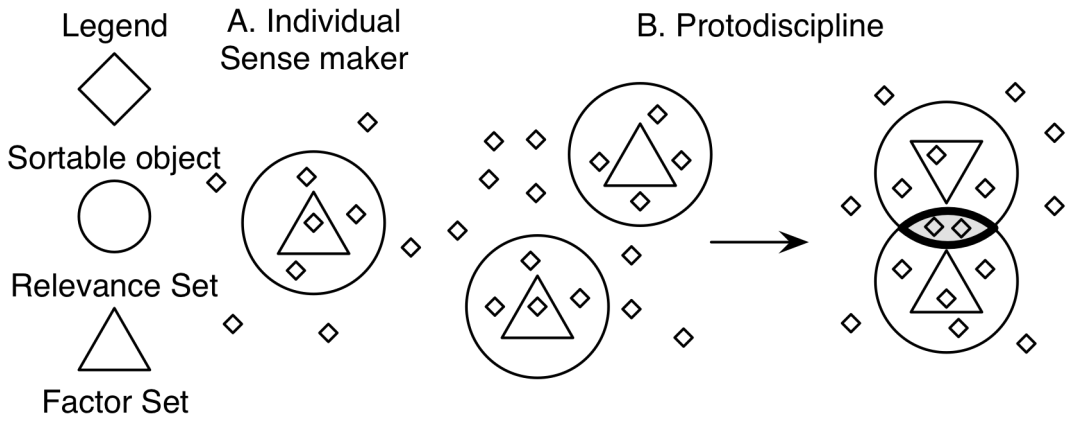

Figure 1.1: Sensemaking entails discriminating first between relevent and irrelevant material, and second between relevant and productive resources.

weigh so heavily on the mind that the concept of interpretation can be dismissed. If however interpretive variation is the norm then we can also posit that within limits the same larder satisfies many chefs so long as each is free to cook according to her particular tastes.

Such tastes are not merely metaphorical; an important feature of this conceptualization is that scholars produce work for their own consumption first, especially early in the development of a discipline. There must be a point of origin where cultural sensemaking is idiosyncratic to the creator. It is only with time that multiple idiosyncracies can be commensurated, and given such commensuration there can be no expectation that consensus develops to the point of uniformity in thought even in a solitary thinker not to mention a community of scholars.

If cultural sensemaking were the only mechanism of disciplinary coherence then it would draw people together but only weakly. Though the kernel of a core set of references may grow this does not imply a tendency for disciplines to become closed culturally. What truly bonds scholars is the opportunity to stand up and be seen through ceremonies around the core while being otherwise totally invisible when pursuing idiosyncratic interests.

1.1.1.2 Relevance

1.1.1.3 Cognition

Here I attempt to combine internalist and externalist theories of culture. "Art for art's sake" characterizes the internalist view; cultural expression is both motivated and regulated by a concept of the object and standards of quality that, no matter how they got there, are autonomously held in the mind of the creator. "Art for the artists's sake" characterizes the externalist view, where the interests of the creator may be multiply determined by a number of social pressures including fame, fortune, fraud, or force. Many couplets of opposing forces–knowledge and professions, culture and society, truth and power–merely replicate this basic distinction.

Yet the internal/external distinction is generic and restrictive. The underlying references concern how a person might orient herself to cultural production, and how such productive activity may be embedded in larger social structures and cultural milieus. I break this distinction into three simple processes that make creative work easier.

1.1.1.4 Social Structures

It is during the period where recognition develops that social control begins to collide with personal sensemaking. This control may at first go no further than a group of peers beginning to enforce conventions of practice or symbolism by asking or demanding that a concept be manifest in this way and not that way, or that it not be expressed at all. If the chief mode of personal sensemaking is sorting objects into sets of similarity or difference, the nascent form of social intrusion on cultural sensemaking is attaching labels to the sets of others. If a sensemaker is exposed

to sets with the same label and different contents then any attending pressure of contradiction would be most easily resolved by direct interference with someone else's process of personal sensemaking. Disengaging from the contradiction is certainly an option, but it is likely that the dopplegangers would find themselvs at odds again if they inhabit the same milieu.

If there is a weak social force drawing thinkers together at the overlap of their sensemaking, it will pale in comparison to exogenous sources of social control that can define cultural relevance by fiat, force, fame, or facilitation.

Even politically motivated ideologies require constant social control to combat the drift in personal sensemaking.

If a person seeks out social control over their own sensemaking then they are thoroughly socialized. This tautology may describe some people most of the time and everyone some of the time. Even if I stipulate that a sensemaker is delighted when someone else expresses a thought that makes sense to her, such delight must be predecated on the prior development of recognition, or else identical thoughts would pass each other unrecognized like ships in the night.

A cartographic metaphor will be useful. The indexing of the set occurs when a set is labeled, and this is analogous to the name of a contiguous territory on a map. The territory is defined only by the boundary that encloses it, and no knowledge of its contents is necessary to refer to the label. Any content discovered must be categorized de novo, and the boundaries provide the decision rule to accomplish this. Recognition of bounded content occurs when the content is given an address. The address need only lead a searcher to a point of contact with the content, for the address is only a reference and bears no knowledge of its own.

It is entirely possible that discourse around a set of references can be sustained without ever referring to underlying content. The great efficacy of organizing

cultural objects into sets of addresses is that it is much easier to agree on the address than it is to agree on the character of what is addressed.

In Parsons's terms the addresses are a form of influence. Influence is a currency bearing social status that allows communication to occur generically and at a pace faster than would be the case if the real underlying content had to be mobilized. Money is the archtypical form of generalized media, allowing exchanges can be defined prior to the actual mobilization of goods and services.

By organizing cultural history into categories, disciples solved two important problems. First, they sanctioned ignorance; lack of knowledge beyond the curated set would be no threat to disciplinary credentials. Second, they created a method of communicating tacit knowledge where formal codification or training was absent. If a scholar were able to reference the contents of a curated set in ongoing conversations he would be rewarded with membership in a self-organized community of scholars including anyone willing to uphold the legitimacy of the discipline. While variations in the quantity and quality of set references confer differences in status iternally, a reference no matter how controversial signaled deference to the discipline and would be treated as a credential for membership.

Such curation served as a critical factor to scholarly production in both of its forms, research and education. The cultural artifacts of curation left by research were the footnotes and bibliographies of books and articles, that left by education were lectures and course syllabi.

Professional activity routinizing the reproduction of the pattern these protypes established. They developed legitimacy for the low level of extant cultural production, low relative to the subsequent historical growth of disciplines.

Recognition, a process that amounts to the social organization of existing cultural material. This is the process of genre formation. Though the term is sometimes

use to refer to the entire developing art world or field of cultural production (Lena and Peterson 2008), *genre* is better reserved to refer to the development of the cognitive institution alone. Here genres represent the collision of social constraints with cultural production.[1] The initial symbols were indexical, allowing productions to be labeled according to a cogntive categories representing the disciplines. For the U.S. social sciences, by the end of the ninteenth century symbolism had already been developed around the . In this regard scholarly disciplines behave according to a logic of development that has been articulated in the context of artistic genre.

Tendencies leading to each of these thresholds followed different yet often intersecting developmental logics.

1.2 Theory

A cultural action is teleological, autonomously controlled, socially neutralized, facility constrained, and indeterminant. A social action is teleological, heteronomously controlled, culturally neutralized, constrained normatively by prestige, sanctions, and rightness or competitively by scarcity, acquisition, and attrition, and determinant. A sociocultural behavior is merely an admixture of the two.

- A. Cultural
- B. Personal
- C. Social
- D. Facilitory
- E. Developmental

[1] In Habermas's terms, genres represent a development of system against lifeworld. In Bourdieu's terms, genres are fields, specifically the definition of positions to be taken.

1.2.1 Typology of Relationships

1.3 Critical Review

- Paradigms, Structural Mechanisms, and Stage-sequential Development

1.3.1 Lena

In the sociology of music a very different developmental theory has been proposed centering on the concept of genre (2008). Genres differ significantly from paradigms in that they do not require a strong cultural consensus to be reproduced. Disparate cultural objects and practices may be united under a genre if similarities are successfully socially constructed. While genre classifications may become normative they are usually difficult to enforce; exogenous pressure is usually required to condemn something as violate. This is because genre inclusion tends to occur when a threshold of similarity is crossed by, so to speak, checking of enough boxes on a checklist of attributes, while remaining unclassified features of the object are ignorable if a preponderance evidence in favor of inclusion is established.

1.3.2 Kuhn

It is the contention of this study that this transition in scholarly practice between casual and informal citation behavior is both a cause and a consequence of a developmental transition in the social an cultural structures of scholarly fields.

A substantive theory of such structures was introduced by Kuhn in the *The Structure of Scientific Revolutions* (**???**). Rather than reproducing Kuhn's theory, my aim here is to generalize his model–which was intended for application to the

natural sciences especially prior to the maturation of the university system in the twentieth century–to scholarship in general, in the United States, and in the first half of the twentieth century. Though my theoretical scope aims to be inclusive all disciplines finding a support in the U.S. university system, I will restrict my empirical investigations to selected disciplines in the social sciences.

To generalize Kuhn I draw on more recent work in the sociologies of knowledge, of cultural production, and of professions. Whereas Kuhn focused intently on three relationships, of scientists to paradigmatic works, to their peer scientists in the narrowest sense, and to their own direct labor of scientists to

Paradigm has at least two referents, one concerning the process of cultural production which emphasizes the relationship of creators to their craft, and the other concerning the normative regimes within the community of craftspeople. The cultural reference is embedded in the eytemology of the term; a paradigm is a prototypical cultural product that serves as a model or basis of comparison for the reproduction and extension of knowledge, either that gleaned from the lesson of the object or knolwedge about the process of creation. Not all objects are good paradigms in this sense; a paradigm has staying power if it poses problems without answering them, and if the solutions are partially but not fully explicated.

The pattern of oscillation between normal and revolutionary science in Kuhnian theory can be thought of as recurring moments in the continuous development of a stable thing (professional science), or it can treated as a very simple ecological model with only two things (rival paradigms) drawing on the same scarce resource base (professional science).

Not so with paradigms; a novel object must be thoroughly vetted to be considered relevant to the cause of normal science. If it does not pass a sniff test it will be ignored, and if it insists on being recognized and remains incommensurate with

extant expectations then the familiar revolutionary crisis ensues.

1.3.3 Collins

1.3.4 Abbott

1.3.5 Bourdieu

To begin it will be useful to consider a theory that is diametrically opposed to the classification of groups. I find one in Bourdieu's essay, *The Social Space and the Genesis of Groups* {*Bourdieu:1985wh}. Bourdieu claims that Marxian social class categories, and their implied boundaries, are analytical constructs with no real referent in the world. The rules for placing an uncategorized actor into a class are well established in any number of class theory traditions. The rules relate to attributes of the actor which may be independently measured, e.g. does she sell labor or buy it? If she sells it, she is a proletarian, if she buys it she is a bourgeoise. Because these categorical rules are the analyst's, Bourdieu decries the reification of the boundary between classes so constituted. Rather than a multidimensional multinomial space, Bourdieu argues that people actually exist in the world in a multidimensional euclidean space, where, in capitalism at least, each dimension is a form of capital. Money, for instance, may be used to place people on a scale and the difference in their holding defines their distance along one dimension of that space. cultural capital (the same thing would be true, mutatis mutandis, of the economic game) determines the aggregate chances of profit in all the games in which cultural capital is effective, thereby helping to determine position in social space" (1985:724) Murkier is Bourdieu's notion of relationship. In the field theory relationships tend to be competitive bids for profit in a field constituted by a form of capital. (**???**)

It is only within these subdivisions–including the power elite as a small and local community of its own–that Bourdieu's fields of cultural production operate according to a logic of peer recognition. Yet these local communities do not form the larger share of a professor's quotidian reward. As a sanctioned officer of the academy professors enjoy influence not with their peers but with their laity–students and members of the disciplinary public. Horizontal stratification also allows professors to safely exchange recognition across disciplinary boundaries without fear of competition.

1.4 Historical Case: U.S. Social Sciences, 1900-1942

_ References and the Development of Scholarly Disciplines

Between 1900 and 1925 each American social science discipline distinguished itself as an autonomous profession. .

A key component of the establishment of each as a field was the development of a pan-disciplinary convention of citing references. We take for granted the use of citations as a currency of information flow and authorial recognition, but early in the century it was not a norm to provide precisely codified descriptions of consulted publications. Citations were often very casual, referencing an author by title and surname only, and referring to an idea and not any work in particular. . These proto-citations required a contemporary grasp of context to be intelligible, as do citations today, but they lacked the address-like codification that would allow the unknowing reader to actually locate the source in question.

Over several decades however the act of referencing became both more common and more consistently codified. Whereas in 1900 the average number of references in an article bibliography for the journal <> was <>, by 1925 it had grown to

<> and by 1940 to <>.

These references indicate the general technical upgrading of scholarly rigor. Scholars were, especially among peers, expected to be transparent about their influences, to be deferent to their predecessors and fair to their opponents, and to provide readers with the means to follow up on claims made about sources.

But what had scholars been doing previously if not scholarship? If disciplinary conventions were not solid, was their work undisciplined? The answers are not merely semantic; they involve the task of disentangling the relationship between two very different aspects of creative occupations. On one hand, scholars have always struggled to make sense in public about what they have struggled to make sense of in private. However we conceptualize the basis of thought and expression, internal and external representations are ontologically distinct processes and the mechanisms of translation between them are nontrivially constructed and maintained. What distinguishes early scholars from the inheritors of their craft is a difference of proportion between internal and external requirements; beset by a smaller and less powerful audience early scholars were more free to engage in idiosyncrasy, to commit their energy to an internal struggle of making sense of their own ideas and those cast in front of them. This did not mean that they worked in isolation; insofar as they were required to meet heteronomous standards these were content standards, either defined by peer pressure, by the personal relationships of rivalry and comraderie (c.f. restricted fields of cultural production [/@Bourdieu:1993vo/ :39]) or by the tastes of lay audiences who mattered for the scholar's livelihood.

The relationships defining these content standards were institutions like any other mechanisms restricting what scholars might otherwise do if left alone. However they were not disciplinary standards. Adherence to them did not signal membership in the profession of scholarship. Instead content standards signaled their intellectual

location directly; recognizing scholars required understanding the content of what they were speaking.

As scholarship grew in published work and personnel, references also became a sign of dependence, since one scholar could not be expected to master more than a component of her milieu.

References then constitute an important element of not only scholarly practice, but after a period of institutional development, scholarly orthopraxy. It is only in the narrowest sense of a subfield that normativity can be characterized as a form of orthodoxy, or correct thought. The normative power of a profession can at best be a form of correct practice that puts only very weak restrictions on correct cultural or intellectual content. But differently, two scholars on equal footing may consider each other's substantive work profane, but this mere fact of heterodoxy would not normally disqualify either of them from membership in the profession. The submission of an "unprofessional" manuscript, one that, for instance, did not adhere to citation conventions, would bar the work from publication.

Scholarship then is a cultural lifeworld steered by a professional system (**???**). Habermas's concepts provide a more precise forumulation of the familar embeddeness concept in sociology. The metaphorical content of embeddeness suggests close contact between the focal field of practice and the more or less hidden mechanisms that support it. However, these supporting mechanisms often stand at arms length from the practices of, for instance, scholarly discourse. If we presume that scholarly discourse is exemplifed by verbal exchanges, for instance, among students in a seminar or symposium, between a reader and the written words of a reference, or between a writer and her own consciousness, then by mechanisms of discourse we must mean the ways that proximate utterances shape their antecedeants in streams of discourse (**???**), (**???**), (**???**)]. Professions then are embedded much like pylons struck into the shifting sands of language. However this reverses the embeddeness

metaphor. Scholarly discourse is not supported by the unseen mechanisms of the profession, rather the profession is supported by the unseen mechanisms of the discourse. Without the churn of language and thought, editors and reviewers would have no content to split and bolster, and no audience to serve the cream to. We might say that system and lifeworld have an epigenetic relationship. However what we are uncovering here is not the hierarchy between professions and culture, but the lack thereof. Professions and culture are interpenetrating; the resources and chance opportunities determining one's control over the other turnover to such an extent that we might as well accept Sewell's anti theory of structural multiplicity (**???**)?:].

Abbott calls culture or lifeworld an interactional field. Cf abbots basketweave

Abbott (2001)]

Citation practices however did not necessarily represent better scholarly practice, in the sense of a capacity to communicate more clearly and with fewer errors of fact or interpretation. These different possible meanings of a reference made them polysemic in a way that belied a "disciplined" discourse. Rather than trying to "get Kuhn right", I will draw selectively to fill out the conceptual matrix outlined below.

A lacuna in Kuhn's model of scientific development concerns a clear conceptual model of what might be called non-cyclical or original/genetic development. The basic cyclical pattern is between revolutionary and normal science, where each is always a prerequisite for the other. This poses a chicken and egg question; presumably a different process explains an "original" revolution developing out of no prior paradigm. Kuhn is able to forestall this question through an historical approach where antecedent moments in historical (not analytical) development are always at hand for the researcher who has done her homework.

1.5 Empirical Foundations

This brief foray into the different sociocultural functions of citations may be demonstrated by the observation of the formative moments of U.S. social science. The impressions left by the earliest social scientists became a terrain of disciplines and invisible colleges that could be landscaped but not easily turned over by future generations. Paradigms or hegemonic cultures developed in the first half of the 20th century. These paradigms endured even during times of social upheaval such as the Great Depression and WWII. However, at the dawn of the 1960s, such monoliths were toppled in quick succession. For the first time in American history, the cultural heritage was treated in an a la carte fashion by new generations.

What was really different about the 1960s?

1.5.1 Study 1: Ordered Diffusion of Disciplinary Bases

Audience and contemporary may become obvious, and paths of influence may present themselves as stimulating creative opportunities.

Categories then are not the foundation of institutionalizations, but rather they serve as a hinge between cultural and social phases of institutional development.

In real historical situations an author cannot wait for her chosen obsession to become institutionalized, and indeed the likilihood of this happening for her is slim given the rate of cultural survival.

Nothing forbids creators from inventing knowledge about the world before it is clear how to label that knowledge.

On the contrary she will try to master whatever extant opportunities exist, then using the traditional production as a security base, delve into her idiosyncratic

persuit with whatever resources are left over.

The development of such social structures is necessarily endogenous to the recognition of particular forms of cultural production as belonging to a category.

The empirical trace of recognition is the appearance of flags toward which categorical thinking and boundary work may be directed.

The organization of one's work along disciplinary lines does however tend to make one more productive.

The stereotype of the artists who suffers for her craft and dies young and consumed by her obsession is the exception that proves the rule.

The term is vague and serves only to establish a beachhead on a set of objects that might otherwise be explained by any other ontology, such as from another intellectual discipline, a religion, popular culture, or common sense.

There may not be a formative moment for disciplines, but there are tendencies of development that when compared at relatively large times scales show transitions from separate, substantively heterogenous, and methodologically anisomorphic persuits, more simply personal obsessions, into institutionalized and then professionalized relationships.

These adjectives begin to describe the feature of an object as being explicable by the knowledge that the discipline may bring to bear.

Ultimately disciplines intersect with organizational resources, first as a market for tutorship and ultimately as a professorship (c.f. Weber's description of prebends (**???**343)).

We submit that recognition is empirically identifiable if there exists a flag around which categorizing coalesces.

What interests us presently is to trace the development of scholarly disciplines back to a point where they were weakly recognized even if they were already culturally prolific.

What is more likely is a shifting of resource deployment, first with an investment in the aqcuisition of an income generating activity and secondly with a shift of attention to the obsession for as long as neglect of the day job can be tolerated.

Without a rallying point a discipline cannot develop as a meta-category.

Yet, even before these disciplines exist all that someone can do to categorize the explanation is apply the modifier.

Ideally, I would analyze the entire stock of recorded publication material to give the best chance of observing when authors contravene institutional boundaries. Practically, I must take a sample, however sampled networks are not small versions of the population network (Handcock and Gile 2010). Sampling may have the effect of degrading the network cohesion on which community detection methods depend, such that a method will not detect the same boundaries in a sample as it would in the population. To avoid a sampling effect on network cohesion, I draw a full census of articles, reviews, and book reviews from each journal selected.

Sampling on journals creates another problem, which is to merely reproduce boundaries coextensive with the journals from which the articles are drawn. Even though journals market themselves at catering to particular disciplines and sub-fields, I should not assume that authors, editors, and reviewers always obey these distinctions. If a scholarly field exists with a grounding in two or more journals, the omission of one may also degrade its cohesion to the point of rendering its boundaries undetectable. As an indicator of affiliation among journals, I use Leydesdorff's (2010)

I should also expect to observe boundaries due to several other institutional levels higher than journals, like publishers, disciplines, or national and language groups. To provide ample opportunity to observe boundaries existing in the space between journals and between academic disciplines, I also take a large sample of journals

1.5.2 Study 2: Settlement of the Citation Landscape

- Poisson Permutation Test to Locate Transition from Extensive to Intensive Development

Growth in the number of scholars and the number of published scholarly works is attended by a qualitative transitions between extensive and intensive patterns of citation. When disciplines are very young scholars are almost always exploring new or at least unclaimed terrain with little interest in covering the same ground twice. As disciplines develop a transition invariably occurs; scholars become much more likely to retrace familiar ground. Much of the work of this study aims to understand the significance of this fact.

First the fact must be established. Intellectual terrain is often imagined as a space of meaning or a population meaningfully distinguishable ideas. Because ideas cannot be directly observed, several indicators of their presence have been used, citations chief among them. Empirically, then, we will start on the basis of the citations as a useful indicator of ideas. Later we will discuss the limitations of the ideational theory, and we will present an alternative interpretation of a citation space. Luckily the facts at issue will not change.

So, then, it will be useful to treat the terrain of scholarship as the accumulating stock of already cited references. The act of exploration, so to speak, of this space consists in the inclusion, in the reference list of a scholarly publication, of a particular set of citations and not another. A footprint in this space, left

by one publication, may be represented as a count of each citation pair in the list of references. This operationalization allows footprints to overlap completely, partially, or not at all. By enumerating citation pairs or co-citations instead of their individual counts, we also claim that the meaning of a reference may vary in combination with other references.

A more empiricist and less theory laden interpretation is to claim that we may identify how disparate acts of cultural production hang together, without knowing why they do so. Citations provide merely one kind of thread, but were we to trace out several more modes of relatedness then we might provide a fuller picture of the sociocultural structure underpinning scholarship. Such a task is beyond scope for the present study, but we can at least specify ignorance (**???**). Clearly there is much more to the content of a publication than its list of references. But even considering this narrow slice of its meaningful content, we are already at pains to generalize from the observation of a citation pattern to the cause of that pattern appearing in a particular time and place. It will be difficult, for instance, to posit a choice mechanism, for we cannot discern whether the inclusion of a reference was the choice of the author, the editor of a journal, the reviewers refereeing the manuscript, a colleague listed in the acknowledgements, an uncredited inspiration, etc. I therefore make no effort to identify an actor responsible for an included reference, but rather consider it the outcome of the local art world surrounding the production of that piece of scholarship [c.f. theories of authorship @ ;@]. What is a critical problem to solve for the intellectual historian may a fool's errand for the population researcher. It is a mistake to treat any particular citation, and especially to treat the entire reference list, as reflective of the choice of the author. Indeed this mystifies the production process behind scholarship.

An extensive pattern of citation then is one that both introduces never before cited references and one that favors those extant references that have been cited the

least by others. A Poisson distribution is a simple first approximation of a random search in this space, and observed citation counts with a mean below the random pattern (underdispersion) can be considered to represent the extensive pattern, while means above the same (overdispersion) may represent the intensive pattern.

This extensive pattern of development may be compared to the paradigmatic model described by Kuhn (**???**10). Once a paradigm, in the sense of a model to be extended, takes hold among a community of scholars, normal science ensues as a process of narrowing the range of possibilities opened by the paradigm. The specifically scientific pattern of development is to retrace familiar problems until they are solved, and then to relegate the solution to one or another form of black box, such as mathematical codification, textbook explication, or codification in technology. Familiar ground is in one moment intensely retraced, and in the next systematically forgotten. Indeed Kuhn aims to demonstrate that the ideology of cumulative development in the sciences is a consequence of black boxing, which serves to render subsequent generations of scientists ignorant of a history better described by a cyclical or sinusoidal trend.

While pre-history of disciplines are beyond Kuhn's scope, here they are paramount. This emphasis is based in a hunch that the mechanisms that govern the genesis of disciplines may be implicated in their ongoing development.

Not all of the things people "do" are behaviors in this sense. Teleological behaviors have a goal or end known to the subject

Whether knowledge is found or invented is less important than whether it is addressable; knowledge that is not recoverable might as well not exist. Addressing occurs when a source is recoverable via a portable reference to its location. Because using knowledge tends to reinforce rather than deplete it, knowledge is only paradoxically scarce. Knowledge is scarce to the extent that it is easily lost.

Such references went through a process of formalization wherein casual statements of credit were replaced with precise street address-like registrations of locations. When for whatever reason citations became precisely codified, especially in the journal space, it became possible to use references as a form of currency in the profession. Here the function of citations became something more than an aide to understanding the text; they became on one hand raw material for sustained cultural production and on the other a set of credentials for membership in social science disciplines and their specialties.

Citations now play several roles, some of which are largely decoupled from the underlying cultural content to which the citation ostensibly points. The role of citations at the center of this article concerns how they are used by researchers to jockey for legitimacy in a competitive professional field. Citations in this sense are neither forms of cultural capital, for they may be used without understanding their significance, nor forms of social capital, because they rarely connote personal ties among the authors. Instead I consider the claim that citations are adoptable traits which refer to the abstract social categories of inclusion and exclusion that form one of the institutional bases of the academic profession. Insofar as particular lists of references must be cited to request or signal membership in an extant professional club, citations become a currency of status exchange.

If a list of citations acts as a credential, there is variability in the length and content of the list, and this variability may become the basis of esoteric hierarchies within disciplines.

Rewards are stratified into two tiers, the first of which is socially dominant and the second culturally dominant. In the first tier the reward is membership in the professorial occupation. Given the first, the second reward is prority ranking. This is the formal recognition that accrues to legendary individuals. A horizontal stratification attends the segmentation of disciplines–subdivisions allowing more

space for local heroes.[2] (Gustin 1973)

1.5.3 Study 3: Subdiscipline Formation

- k-Clique Percolation Clustering of Co-reference Network

1.5.4 Study 4: Survival of Knowledge

- Cox Proportional Hazard Analysis of k-Cliques

1.5.5 Study 5: Genre Signaling

- Association between k-Clique and Topic Model Classifications

References

Abbott, Andrew. 2001. "Part 1 Self-Similarity in Social Science." In *Chaos of Disciplines*, 1–153. Chicago: University of Chicago Press.

Bourdieu, Pierre. 1985. "The Social Space and the Genesis of Groups." *Theory and Society* 14 (6): 723–44.

Gustin, Bernard H. 1973. "Charisma, Recognition, and the Motivation of Scientists." *The American Journal of Sociology* 78 (5): 1119–34.

Handcock, Mark S, and Krista J Gile. 2010. "Modeling social networks from

[2]Except for rare celebrities that a discipline would honor beyond its own boundaries, but again this is a cultural gesture whose tangible rewards are irrelevant to the recipient.

sampled data." *ArXiv* stat.AP (October).

Lena, Jennifer C, and Richard A Peterson. 2008. "Classification as Culture: Types and Trajectories of Music Genres." *American Sociological Review* 73 (5): 697–718.

Leydesdorff, Loet, Felix de Moya-Anegon, and Vicente P Guerrero-Bote. 2010. "Journal Maps on the Basis of Scopus Data: A Comparison with the Journal Citation Reports of the ISI." *Journal Of The American Society For Information Science And Technology* 61 (2): 352–69.

CHAPTER 2

Data

2.1 A lexical sample

To observe some of these high level institutions, I draw sets of journals from four social science disciplines–anthropology, sociology, economics, and political science–and I draw these in blocks from the same publisher. Journals were selected from the disciplinary affiliations signaled in their titles. From a JSTOR master list of archived materials, journals were selected if they contained any of the disciplinary prefixes anth-, soci-, econ-, and poli-. {{Though not all journals that are affiliated with a discipline signal this with a word containing the signature prefix, those that do are affiliated with a high degree of accuracy. Soci is an exception, and journals like the Royal Society of Statistics [madeup] are excluded.}} This list was cross referenced with the TR WOK database.

Following these trends in the use of prefixes, we develop a sample or journals that use them under the assumption that these signal domain relevance for the disciplines.

We reduce the sampling frame in several steps. First, we require that the journal publisher be located in the United States. Second, we require that the journal be included in the JSTOR database. Third, we require that the journal title contain, with some exceptions, at least one of the prefixes . Of the journals in the WOK

master list, these criteria limit the sample to titles or less than half of a percent of the original sampling frame.

The WOK master as of listed titles. It is not clear what sample of historical population of journals this represents, but it is a substantial substantive starting point.

The source data are observations on documents spanning years.

Each study depends on a database of records of the contents of journals. This database is compiled from two sources, JSTOR and the Thompson Reuters Web of Knowledge Social Science Citation Index (WOK).

2.2 Sources

2.2.1 Google Books

2.2.2 Thompson Reuters Web of Knowledge

2.2.3 JSTOR Data for Research

2.3 Research Databases

2.3.1 Sample Selection

2.3.2 Entity Recognition

2.3.3 Formats

2.3.3.1 Edgelist

2.3.3.2 Flat + k-Clique

2.3.3.3 Survival

2.4 Drawer

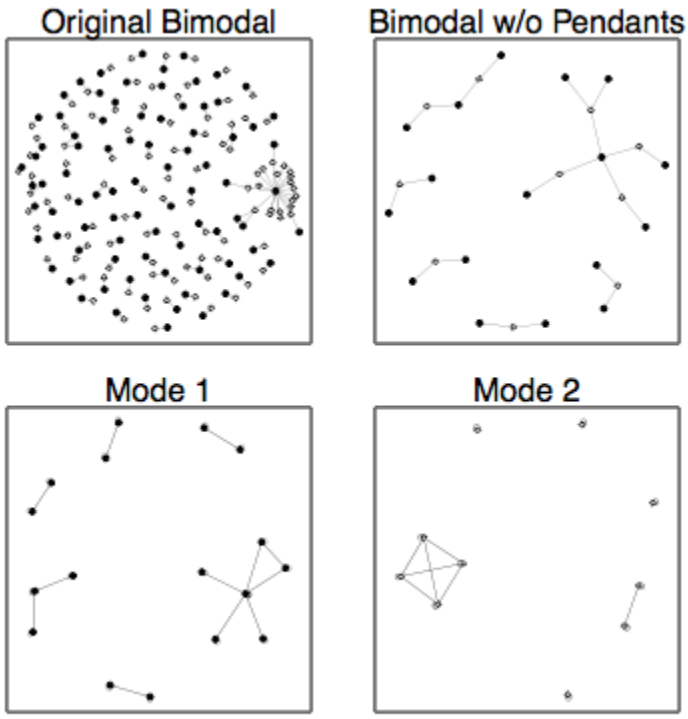

Figure 2.1: Mode Projections

CHAPTER 3

Ordered Diffusion of Disciplinary Bases: A Time Series Analysis

3.1 Introduction

A discipline is established as a categorization mechanism when the set of possible labels for the same cultural object shrinks to a short and ranked list.

The position at the top of that list will be occupied by what we call the disciplinary prefix, and below it variations on nouns used to capture different aspects of cultrual content from created work to the authors themselves.

Consider five social science disciplines–anthropology, economics, political science, psychology, and sociology.

In english the labels that like flags lay claim to disciplinary resources are the prefixes anth-, econ-, poli-, psyc-, and soci-.

In sociology the term "social problem" is an example of the flag being established as a claim to disciplinary relevance.

We contend that these prefixes will diffuse first as generic and weakly categorical terms that could modify and lay claim to any worldly object.

3.2 Temporal Sequencing Methods

Correlations between time series are difficult to tease out due to several dynamics that if not controlled for can lead to spurious correlations. Before we can attempt to test causal order we must decompose historical trends in terms into their systematic and residual components, such that we may test the residuals for patterns between two series.

ARIMA models have been criticized for their irrealism (Isaac and Griffin 1989:877). After establishing statistical considerations and laying bare our assumptions, we will discuss the historical and ontological limitations of the statistical approach.

3.2.1 Series

Table 3.1: Terms searched in the Google Books Ngrams Database

	soci	econ	anth	poli	psyc
Genre	social	economic	cultural	political	mental
Technique	sociological	economical	anthropological	political	psychologi
Ontology	society	economy	culture	polity	mind
Discipline	sociology	economics	anthropology	political science	psychology
Profession	sociologist	economist	anthropologist	political scientist	pscyhologi
Subdiscipline	sociology of	economics of	anthropology of	political science of	psychology

3.2.2 ARIMA model

ARIMA, or AutoRegressive Integrated Moving Average, models are effective in decomposing several categories of within-series correlations.

$$\mathrm{I} = \frac{\mathrm{MA}}{\mathrm{AR}} \tag{3.1}$$

This says that I, the change in our series, is a function of MA, a moving but systematic average (a line or higher order polynomial) and

$$(1-B)^d y_t = \frac{c + (1 + \theta_1 B + \cdots + \theta_q B^q) e_t}{(1 - \phi_1 B - \cdots - \phi_p B^p)} \tag{3.2}$$

Where c is a constant drift up or down,

3.2.3 Granger Causality

3.3 Results

As table **??** shows.

3.4 Which came first?

Granger tests can help determine which (Thurman and Fisher 1988; Granger 1969)

Clear secular trends and period effects surrounding WWII and the baby boom. To control:

- Model the trends. We could estimate the linear trend or splines and then subtract them.
- First differences. Subtract from each point the previous point.
- Link relatives. Divide each point from the point before it.

Box Cox doesn't mean

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} + e_t, \tag{8.1}$$

References

Granger, C W J. 1969. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." *Econometrica* 37 (3): 424.

Isaac, Larry W, and Larry J Griffin. 1989. "Ahistoricism in Time-Series Analyses of Historical Process: Critique, Redirection, and Illustrations from U.S. Labor History." *American Sociological Review* 54 (6): 873.

Thurman, Walter N, and Mark E Fisher. 1988. "Chickens, Eggs, and Causality, or Which Came First?" *American Journal of Agricultural Economics* 70 (2): 237.

CHAPTER 4

Settlement of Citation Landscape

- Poisson Permutation Test to Locate Transition from Extensive to Intensive Development

CHAPTER 5

Subdiscipline Formation

- k-Clique Percolation Clustering of Co-reference Network

5.1 Introduction

This appendix explains the research procedures used to produce tables and figures in *Productions of Culture: Knowledge Survival in Art and Science*.

5.2 Compiling Database from Source Records

5.2.1 Thomson Reuters Web of Knowledge

Web of Knowledge (WOK) data that are available through regular subscriptions may be reported in a long, field-tagged, plain text format. The `wok2dbl.f` function recursively searches a directory for plain text batches of records and quickly imports them into R in a long format that may be easily queried. By default letter case standardization and deduplication by record ID is performed. The function returns a data.table object and optionally saves the same to the hard drive. Use of `data.table` provides accomodation for very large databases that perform poorly when treated as a `data.frame`.

By default the data.table is keyed by its WOK id (the "UT" field), then by the

field. This makes querying easy. To see the authors of the first three articles, we might enter:

By default the value field is not keyed. While there are scenarios where this would be useful—e.g. for calling every record by a particular author, or every record in a particular year—keying also sorts the data.table, and the original sort order is important for fields like "TI" (title) which may be broken across several observations.

The `wok2dbl` object remains in the long format of the original data source. We can see this by simply calling the `wok2dbl` object itself or by using `expand.grid` to query multiple keys at once. Here was ask for the source journal, publication year, number of references, and total citations for each of the first three records.

To sort results by record instead of field:

It is convenient to keep the original source data in a long format and to reshape it as necessary for use in different methods. This will be discussed in the *Formatting* section below.

5.2.2 JSTOR Data for Research

Where WOK data are superior for the study of citations, the JSTOR Data for Research (JSTOR) service provides much of the bibliographic information available in WOK and sometimes more accurately. This makes it useful as a cross reference when assessing the quality of a WOK sample, or for augmenting fields such as authors' names.

In addition to the usual variables, JSTOR data also provide ngram frequencies. These data are very valuable and allow limited full-text analysis using "bag of words" methods. The `jstor2dbw.f` function imports dfr.jstor.org records directly

from the compressed files returned by queries to the service. Parallelization of the importing process is available and suitable for systems with fast disks. The function performs a standard set of text pre-processing procedures (e.g. stemming and stop word, punctuation, digit and idiosyncratic word removal) on the ngram frequency tables contained in zip archives that include them. These ngram frequency tables are returned in the indexed format expected by the `stm` package, and all other bibliographic data available are returned as a `data.table`. A character vector attribute called `vocab` is attached to which the indexes in the `jstor2dbw$bow` refer.

Inspecting the `jstor2dbw` object without bags of words ("bow") or abstracts reveals the standard information, and in the conventional wide, flat file format. The only complex value here is author, and multiple authors are listed with names separated by commas.

While the `bow` variable contains the indexed ngram frequency table, which indexes the `vocab` attribute of the jstor2dbw object.

5.3 Identity Resolution

Also known as named entity recognition, identity resolution is a data quality problem preventing the researcher from identifying the same thing with a unique label. This happens whenever variations of a label exist. As a consequence the researcher may fail to connect two events to the same thing. When correcting for low identity resolution, the opposite error may be introduced, where two different entities are erroneously treated as the same thing.

The approach to identity resolution involves supervised machine learning. Because this method is not fully automatic it is difficult to implement as a straightforward

routine. For now, the results of this analysis are exploited without a manual for conducting the resolution itself.

5.4 Formatting

Depending on the analysis or data manipulation to be performed, the `wok2dbl` and `jstor2dbw` objects may need to be converted to a different format, including network formats, which allow us to take advantage of records containing information on multiple units.

5.4.1 Flat File

The `reshape2` package makes it easy to return the wide or flat file format of a query of a `wok2dbl` object.

Many of the interesting fields in WOK records are complex, having multiple observations per record. Some are falsely complex, such as title (TI), which stores a single observation across several fields. Simple and falsely complex values are often trivial features of the document itself. Truely complex field usually store named entities to which the article is related. The most important complex fiels are author (AU and AF) and cited reference (CR). Source journal is an example of a named entity field that is always simple, because a document is only published in one source at a time, though it may have several authors and citations.

5.4.2 Network Formats

The simplest network data format to work with is an edgelist. An edgelist typically has two columns, the name of the node sending an edge in the first column and

that of the node receiving the edge in second column.

5.4.2.1 Bipartite Edgelist

When considering the different relationships among things that could be treated as a network, the `wok2dbl` object is naturally in the format of a bipartite edgelist. For instance we may treat the sender as the paper (UT record id) and the receiver is the citation (CR) to create a citation network.

Or we could treat the author as the reciever to create a bipartite co-authorship network.

However, because of several problems of identity resolution of the CR field in particular, we recommend using the `dbl2bel.f` utility, which normalizes citation codes through case transformation, removal of digital object identifiers, and deduplication. It also optionally allows for data reduction of citations by flagging citations referenced only once (pendants). A report of the results of pendant treatment is printed.

The `dbl2bel` object is appropriate for import into methods designed for bipartite graphs. Because of the nature of record keeping, each complex unit is relateable to others only indirectly by virtue of common inclusion in an article-level record. With a few lines of code we could merge an article to author data.table to an article to citation data.table to yield an author to citation edgelist.

5.4.2.2 Monopartite Edgelist

A more common operation however is to reduce a bipartite graph to a monopartite graph. This is called a reprojection of the graph, and involves a trivial loss of data. Because many network methods assume monopartite data, we include the

`bel2mel.f` utility. The function expects a two column matrix, so when choosing to drop pendants you must do so explicitly and leave off the pendant column.

Assuming that there is at least one 2-star (node of degree two or more) in the bipartite graph, `bel2mel.f` will by default return both monopartite projections. Each projection is the inverse of the other in the sense that what are nodes in the first projection are edges in the second, and vice versa.

5.4.3 Bag of Words

The `jstor2dbw` object contains a variable `bow` and associated attribute `vocab` which can be fed directly to the `stm` package for topic modeling. Usage will be described below.

5.4.4 Merging

5.4.5 Survival

5.5 Analytical Method

5.5.1 Clique Percolation

5.5.2 Topic Modelling

The `jstor2stm.f` function is a simple wrapper for the `stm` package for stuctural topic modeling.

5.5.3 Survival Analysis

5.6 Reporting

5.6.1 Study 1: The Transition from Extensive to Intensive Development

Growth in the number of scholars and the number of published scholarly works is attended by a qualitative transitions between extensive and intensive patterns of citation. When disciplines are very young scholars are almost always exploring new or at least unclaimed terrain with little interest in covering the same ground twice. As disciplines develop a transition invariably occurs; scholars become much more likely to retrace familiar ground. Much of the work of this study aims to understand the significance of this fact.

First the fact must be established. Intellectual terrain is often imagined as a space of meaning or a population meaningfully distinguishable ideas. Because ideas cannot be directly observed, several indicators of their presence have been used, citations chief among them. Empirically, then, we will start on the basis of the citations as a useful indicator of ideas. Later we will discuss the limitations of the ideational theory, and we will present an alternative interpretation of a citation space. Luckily the facts at issue will not change.

So, then, it will be useful to treat the terrain of scholarship as the accumulating stock of already cited references. The act of exploration, so to speak, of this space consists in the inclusion, in the reference list of a scholarly publication, of a particular set of citations and not another. A footprint in this space, left by one publication, may be represented as a count of each citation pair in the list of references. This operationalization allows footprints to overlap completely,

partially, or not at all. By enumerating citation pairs or co-citations instead of their individual counts, we also claim that the meaning of a reference may vary in combination with other references.

A more empiricist and less theory laden interpretation is to claim that we may identify how disparate acts of cultural production hang together, without knowing why they do so. Citations provide merely one kind of thread, but were we to trace out several more modes of relatedness then we might provide a fuller picture of the sociocultural structure underpinning scholarship. Such a task is beyond scope for the present study, but we can at least specify ignorance (**???**). Clearly there is much more to the content of a publication than its list of references. But even considering this narrow slice of its meaningful content, we are already at pains to generalize from the observation of a citation pattern to the cause of that pattern appearing in a particular time and place. It will be difficult, for instance, to posit a choice mechanism, for we cannot discern whether the inclusion of a reference was the choice of the author, the editor of a journal, the reviewers refereeing the manuscript, a colleague listed in the acknowledgements, an uncredited inspiration, etc. I therefore make no effort to identify an actor responsible for an included reference, but rather consider it the outcome of the local art world surrounding the production of that piece of scholarship [c.f. theories of authorship @ ;@]. What is a critical problem to solve for the intellectual historian may a fool's errand for the population researcher. It is a mistake to treat any particular citation, and especially to treat the entire reference list, as reflective of the choice of the author. Indeed this mystifies the production process behind scholarship.

An extensive pattern of citation then is one that both introduces never before cited references and one that favors those extant references that have been cited the least by others. A Poisson distribution is a simple first approximation of a random search in this space, and observed citation counts with a mean below the random

pattern (underdispersion) can be considered to represent the extensive pattern, while means above the same (overdispersion) may represent the intensive pattern.

This extensive pattern of development may be compared to the paradigmatic model described by Kuhn (**???**10). Once a paradigm, in the sense of a model to be extended, takes hold among a community of scholars, normal science ensues as a process of narrowing the range of possibilities opened by the paradigm. The specifically scientific pattern of development is to retrace familiar problems until they are solved, and then to relegate the solution to one or another form of black box, such as mathematical codification, textbook explication, or codification in technology. Familiar ground is in one moment intensely retraced, and in the next systematically forgotten. Indeed Kuhn aims to demonstrate that the ideology of cumulative development in the sciences is a consequence of black boxing, which serves to render subsequent generations of scientists ignorant of a history better described by a cyclical or sinusoidal trend.

While pre-history of disciplines are beyond Kuhn's scope, here they are paramount. This emphasis is based in a hunch that the mechanisms that govern the genesis of disciplines may be implicated in their ongoing development.

CHAPTER 6

Survival of Knowledge

- Cox Proportional Hazard Analysis of k-Cliques

CHAPTER 7

Genre Signaling

- Association between k-Clique and Topic Model Classifications

CHAPTER 8

Conclusion

8.1 Summary of Findings

8.2 Limitations

8.3 Conclusion

CHAPTER 9

Appendices

9.1 Plagiat

9.2 Reproducible Software Container