

LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Essays on Synergies from the Geographic Clustering of Literary Artists

A THESIS SUBMITTED TO THE UNIVERSITY OF DUBLIN, TRINITY COLLEGE

IN APPLICATION FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY

Sara Mitchell

Supervised by

Professor John O'Hagan



Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

2016



PhD
SOCIAL SCIENCES
& PHILOSOPHY

50734000
* 5 0 7 3 4 0 0 0 *

THESIS
11430

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

TRINITY LIBRARY
24 OCT 2017
DUBLIN
Thesis 11430

Summary

This thesis consists of five chapters on synergies from the geographic clustering of literary artists. Chapter 1 provides an introduction to the thesis.

Chapter 2 utilises a unique, purpose-built panel dataset on prominent authors in the UK and Ireland born 1700–1925 to estimate the productivity gains associated with agglomeration of an industry with few capital requirements and no apparent need to cluster geographically. I analyse the patterns of migration and clustering of authors, and I construct age-productivity profiles to determine the productivity gains (if any) associated with the geographic clustering of literary activity.

Individual fixed effects are utilised to control for spatial sorting by skill level, and the within estimator reveals the average author annual productivity is 10.68% per annum higher when the author resides in London, the only major literary cluster during the time period. There is no evidence of dynamic learning effects, and there is no evidence that the results are driven by a dynamic self-selection process.

In Chapter 3, I explore the hypothesis that physical proximity is a sufficient condition for peer effects to exist between authors. I construct age-productivity profiles to determine the productivity gains associated with the mean quality of peers, mean peer output, and peer group size. Individual fixed effects are used to control for spatial and group sorting.

The within estimator reveals that peer group size has a modest negative effect on productivity, and these results are robust to several alternate specifications. When concentration

of authors in London is cotrolled for, mean author quality while residing in London reduces productivity by 0.12 works per year. These results are primarily driven by older authors, although there is abstantial heterogeneity of returns by own-author quality.

Chapter 4 quantifies the gins that may be realised through the use of social network information. I combine a ew, unique dataset of the social connections with the existing dataset with location, biogaphical and peer data (in which peers are defined as authors living in the same city). I fit consider the predictive power of social network information in terms of annual productivit. Models with social network information provide an increase in predictive power above tht of an age-productivity profile, and this increase in predictive power is one to two orders f magnitude greater than the increase in predictive power of a model that only includes per effects variables.

I then consider two alternae measures of output: total works produced over an author's lifetime and author Impac Index score (a measure of contemporary critical attention). There is an improvement i predictive power with both measures of output. However, social network information as substantially more predictive power for the Impact Index score than lifetime works. This suggests that way output is defined is of critical importance, and a failure to account for his may result in incomplete or incorrect conclusions regarding the effect of social network nformations on productivity.

I also provide a detailed dcussion of the author network and OLS estimates of the relationship between social ntwork information and productivity. While I do not establish a causal relationship, I proide a discussion on the ways in which social network connectivity and centrality may ifluence productivity. I discuss the challenges associated with establishing a causal relatiosship when using social network data to provide some guiding intuition for future studies n authorship and other creative and cultural industries.

Chapter 5 provides a conclding discussion.

Acknowledgements

I would like to thank Prof. John O'Hagan for his guidance and support as my PhD supervisor throughout my study and research, and I thank the Grattan Scholar programme for funding my study. This PhD would not have been possible without the support of the Grattan donors. I express my appreciation to Vahagn Galstyan and Ronan Lyons for serving on my confirmation panel. Their comments and insightful were valued greatly.

The research assistance of Rahul Dewan and Rónán O'Connor on the data collection enabled this project to be possible. I thank them for their contributions. I also thank Lukas Kuld for his assistance with R, and I am ever grateful to Luís Serralheiro for his help in writing the javascript code used to check and correct the consitency of the data.

I would also like to thank Nathaniel Russell, Michael O'Grdy, Sanna Nivakoski, and Gavin Morrison for reading over various drafts of the thesis chpters. Their keen eyes and suggestions were vital to this dissertation. I also thank Bibi Banchi, Miriam Sorace, and Luís Serralheiro for their patience, support, and advice througout my study.

Finally, I thank my friends and family for their continous spport and encouragement throughout the PhD.

Contents

1 Introduction **1**

2 London Calling? Agglomeration Economies in Literature since 1700 . **7**
2.1 Introduction 8
2.2 Theoretical Framework 9
2.2.1 Geographic Proximity and Tacit Knowledge 9
2.2.2 Agglomeration Economies and Productivity 11
2.2.3 Agglomeration of the Creative Industries 13
2.3 A Brief History of Publishing in Britain 14
2.4 Data 17
2.4.1 Selection of Authors 17
2.4.2 Migration Trends and Summary Statistics 18
2.5 Identification Strategy 22
2.6 Results 24
2.6.1 Robustness Analysis 26
2.7 Conclusion 33

Appendices **35**
2.A Detailed Data Collection Methodology 36
2.A.1 Selection of Authors 36
2.A.2 Quality of Authors 39
2.B Tables and Figures 42
2.B.1 Figures 42
2.B.2 Tables 46

3 Peer Effects in Literary Production **58**
3.1 Introduction 59
3.2 Existing Research 61
3.2.1 The 'Black Box' of Agglomeration Economies 61
3.2.2 Social Interaction and Peer Effects 62
3.3 The Data 66
3.4 Identification of Peer Effects 69
3.5 Results 73
3.5.1 General Peer Effects 73
3.5.2 Peer Effects within London 75
3.6 Conclusion 81

Appendices **82**
3.A Tables and Figures 83

3.A.1 Figures 83
3.A.2 Tables 86

4 The Predictive Power of Social Networks in Literary Production . . . 94
4.1 Introduction 95
4.2 Theoretical Framework 97
4.2.1 Proximity, Social Interaction, and Creative Production 97
4.2.2 Power and Prestige in Social Networks 99
4.3 The Author Network 102
4.3.1 Network Connection and Distance 107
4.3.2 Measures of Network Centrality 111
4.4 Econometric Approach 114
4.5 Empirical Findings 118
4.5.1 The Predictive Power of Social Networks 118
4.5.2 Social Networks and Productivity 121
4.5.3 Sensitivity Analysis: Lifetime Accomplishments 123
4.6 Conclusion 129

Appendices 131
4.A Tables and Figures 132
4.A.1 Figures 132
4.A.2 Tables 137

5 Conclusion 146

List of Figures

2.1 Share of Authors in Major Cities 42
2.2 Total Authors and Output per Annum 43
2.3 Total Output per Annum by Location 43
2.4 Frequency of Output Observations 44
2.5 Impact of Static and Dynamic Effects on Age-Productivity Profiles 44
2.6 Estimated London Premium 45
2.7 Estimated Pre- and Post- Migration Trend 45

3.1 Distribution of Author Quality 83
3.2 Share of Authors in Major Cities 84
3.3 Distribution of Number of Peers 84
3.4 Distribution of Number of Peers (excluding obs. with zero peers) 85

4.1 Graph of Author Network 132
4.2 Degree Distribution 133
4.3 Sir Walter Scott – Local Network 133
4.4 Virginia Woolf – Local Network 134
4.5 William Butler Yeats – Local Network 134
4.6 An Example of Power in an Author Network 135
4.7 James Joyce – Local Network 136
4.8 Charles Dickens – Local Network 136

List of Tables

2.1 Top 30 Literary Artists by Ranking 47
2.2 Number of Authors Born from 1700–1925, by Birth Location 48
2.3 Most Populous Cities (1851 Census) vs Top Author Destinations 49
2.4 Summary Statistics by Gender 50
2.5 Main Results 51
2.6 Main Results with 2-Year Time Dummies 52
2.7 Main Results with 5-Year Time Dummies 53
2.8 Probability of Migrating to London 53
2.9 Probability of Migrating to London Based on Quality (High vs. Low) . . . 54
2.10 Probability of Migrating to London by Century of Birth 55
2.11 Effect of Minor Clusters 56
2.12 Duration Effects 57

3.1 Summary Statistics 86
3.2 Main Result 87
3.3 Minor Variations in Specification 88
3.4 Variations on Number of Peers 89
3.5 City-Peer Effects in London 90
3.6 Effect of Younger Peers vs Older Peers 91
3.7 Effect of Living in City with a "Superstar" 92
3.8 Heterogeneity of Returns by Own-Quality 93

4.1 Author Summary Statistics 137
4.2 Network Summary: Connection and Distance 138
4.3 Network Summary: Centrality 138
4.4 Top 10 Authors by Impact Index 138
4.5 Degree – Top 10 Authors 139
4.6 Betweenness – Top 10 Authors 139
4.7 Closeness – Top 10 Authors 139
4.8 Clustering – Top 10 Authors 140
4.9 Top 10 Authors by Lifetime Output 140
4.10 Predictive Power of Social Networks 141
4.11 Effect of Network Centrality on Productivity 142
4.12 Correlation Table of Network Variables 143
4.13 Predictive Power: Lifetime Accomplishments 143
4.14 Lifetime Accomplishment: Lifetime Output 144
4.15 Lifetime Accomplishment: Impact Index 144
4.16 Predictive Power: Citation Index as a Measure of Lifetime Accomplishment 145
4.17 Lifetime Accomplishment: Citation Index 145

Chapter 1

Introduction

By the early 17th century, the publication of literature by artistic patronage was in decline, and a commercial market primarily based in London began to flourish under a system of perpetual copyright (Encyclopaedia Britannica, 2014). The first copyright law was introduced when the British parliament passed the Copyright Act of 1709. While English publishers acted as a cartel to maintain their monopoly power and uphold the previous system of perpetual copyright, Scottish publishers were subject to the active enforcement of the copyright law.

This allowed for a dramatic expansion of the publishing and book manufacturing industry in Scotland, particularly in Edinburgh. Scottish printers able to produce books and sell printing equipment at lower costs than printers in England and Wales. They were also able to sell books in England, even as far south as Oxford and Cambridge, for prices substantially lower than the London cartel (St Clair, 2004).

Although there is infrastructure required in the production of books, the printing process did not require the author to be present. Transcripts could be mailed through the post, and many business dealings could be handled by a publisher or agent (St Clair, 2004). There were also no reputational advantages to living and working in a particular city, at least not in terms of developing a reputation with the reading public. Romantic poets and Victorian novelists were typically anonymous, and the reading public generally did not demand to know or appear to care about authors' identities (St Clair, 2004).

With large publishing industries in both Edinburgh and London, the ability to submit of writings by post, and no reputational advantages of being associated with a particular city, where would authors chose to live and work? Unlike a factory or science laboratory, writing does not require a centralised location of production. Therefore, why would authors tend to cluster geographically at all? If authors did co-locate in particular cities, did they benefit from this literary agglomeration? If so, did authors become more productive by simply living and working among other authors or do authors benefit more from certain types of relationships? Was the average author like the Brontë sisters, who are often styled as solitary geniuses – or more like the Lost Generation of Paris in the 1920s, a dense network of artists, musicians, writers, and friends?

While the agglomeration of industry and the mechanisms driving the resulting gains are evident for industries with traditional modes of production, less is known empirically about these dynamics for the creative and cultural industries. I constructed a unique panel dataset on 370 prominent authors in the UK and Ireland born 1700–1925 to examine these questions and provide to a greater understanding of the circumstances that promote creative and cultural production. I begin in Chapter 2 by providing greater historical context and discussing the methodology used to collect this dataset.

I then analyse the patterns of migration and clustering of authors. The spatial distribution of authors differs significantly from the spatial distribution of the total population, and this result is not driven by trends in birth location. London is the only major geographic cluster during the time period. At its peak, London was home to over 50% of all authors. While Dublin, Edinburgh, Oxford, and Cambridge emerge as minor clusters, there are consistently fewer than 5 authors in each city in a given year. Large cities, such as Manchester and Liverpool, did not see the co-location of authors.

I construct age-productivity profiles to determine the productivity gains (if any) associated with the geographic clustering of literary activity. Individual fixed effects are used to control for spatial sorting by skill level, and the within estimator reveals the average author annual productivity is 10.68% per annum higher when the author resides in London, the

only major literary cluster during the time period. There is no evidence of dynamic learning effects, and there is no evidence that the results are driven by a dynamic self-selection process. There is also no evidence that authors strategically build a portfolio of works before moving to London and publishing them in quick succession.

I argue that productivity gains from agglomeration effects occur via tacit knowledge transfer and that these forces operate as both a by-product and driver of geographic clustering of literary production, but I acknowledge that it is not possible to empirically confirm this mechanism due to the 'Marshallian equivalence' of agglomeration forces. These findings relate to a broader 'black box' of agglomeration economies: "the many excellent studies of productivity have told us about the existence of agglomeration economies and also about their scope across industries, locations, and time. They have not, however, had much to say about the sources of agglomeration economies" (p 2148, Rosenthal and Strange, 2004.)

I explore this 'black box' of agglomeration effects in literary production in Chapter 3 by determining how various peer effects impact own-productivity. However, from both theory and empirical evidence, it is not clear in which areas or fields peer effects should exist, what factors should matter in a given circumstance, the 'correct' level of aggregation in which peer effects may be identified, or what kinds of relationships in which peer effects are most likely to exist. Azoulay et al. (2010), for example, argues that spillovers between academic scientists are more likely to occur between individuals in 'idea space' than in 'physical space' or 'social space.'

While the hypothesis is likely true for highly specialised fields requiring a large knowledge base, it is not clear whether the same restrictions should hold for other creative and cultural industries. It is possible that the 'idea space' condition is met by authors writing in the same city at the same time. Therefore, I also examine the hypothesis that physical proximity is a sufficient condition for peer effects to exist between authors.

I construct age-productivity profiles to determine the productivity gains associated with the mean quality of peers, mean peer output, and peer group size. Individual fixed effects are used to control for spatial and group sorting. I also determine whether simply having

peers (binary indicator) or whether categories of peer group size better capture the effect of number of peers. I explore peer effects within London, the only location with agglomeration effects as determined in Chapter 2.

I also consider the possibility that certain types of peers may matter more. I investigate the effect of younger versus older peers: younger peers may represent up-and-coming generations, increased competition, while older peers may be mentors, the status-quo, or the 'giants' on whose reputation the author must build. I consider the possibility that living in the same city as a 'superstar' may matter more than average peer quality. Finally, I also look at effects by heterogeneity of own-quality, as own-quality may impact ability to form useful connections.

The within estimator reveals that peer group size has a modest negative effect on productivity, and these results are robust to several alternate specifications. When concentration of authors in London is controlled for, mean author quality while residing in London reduces productivity by 0.129 works per year. These results are primarily driven by older authors, although there is substantial heterogeneity of returns by own-author quality.

These results could indicate a strong competition effect. During much of the time period, the time and resources of publishers were extremely constrained due to technological limitations. To a large degree, publication was a zero-sum game: if a printer and publisher are nearing capacity from producing other authors' work, then own-ability to get work published is reduced. It is also possible that publishers were able to identify and give preferential treatment to high-quality authors or that high-quality authors are better at making high-quality social connections or better at navigating the publishing process. The decrease in productivity could also be due to learning from peers that results in increased *quality* of work with the trade-off of reduced *quantity* of work.

Though I are unable to confirm these mechanisms, the results suggest that the city-level is not the most appropriate level of agglomeration to detect spillovers between authors. Studies on peer effects within the creative and cultural industries should either incorporate better proxies for peer interaction or data on the interactions themselves. I note that

anecdotal evidence on collaborative groups, such as the "Lost Generation" of writers, artists and other creative workers in Paris in the 1920s, suggests that the spillovers occurred not broadly across creative industries or generally within a city but specifically within social networks.

Social network information may not only provide greater explanatory power compared to broadly-defined peer effects but also provide greater insight into the conditions in which innovation and creative production thrive. However, the cost of collecting social network information is often time- and resource-intensive, particularly when studying historical figures. Thus, the application of social networks in studies of creative and cultural production have been limited, as researchers opt for data that can be more efficiently collected (e.g. basic biographical information such as birth and death location, years of birth and death, total works produced, etc.)

Chapter 4 quantifies the gains that may be realised through the use of social network information. I combine a new, unique dataset of the social connections with the existing dataset with location, biographical and peer data. I first consider the predictive power of social network information in terms of annual productivity. I then consider two alternate measures of output: total works produced over an author's lifetime and author Impact Index score (a measure of contemporary critical attention).

Models with social network information provide an increase in explanatory power above that of an age-productivity profile, and that this increase in explanatory power is one to two orders of magnitude greater than the increase in explanatory power of a model that only include peer effects variables. There is also an improvement in predictive power with both measures of lifetime accomplishment. The results suggests that way output is defined is of critical importance, and a failure to account for this may result in incomplete or incorrect conclusions regarding the effect of social network informations on productivity.

I also provide a detailed discussion of the author network and OLS estimates of the relationship between social network information and productivity. While I do not establish

a causal relationship, I provide a discussion on the ways in which social network connectivity and centrality may influence productivity. I also discuss the challenges associated with establishing a causal relationship when using social network data in order to provide intuition for future studies on authorship and other creative and cultural industries.

The expanded use of social network information could greatly contribute to the empirical and theoretical literature on creative and cultural production beyond what is possible through more traditional peer effect measures. For example, the results also suggest that the gains from interacting with peers may not occur between those who are physically proximate but instead occur between certain types of social connections. Therefore, it may be useful to incorporate factors such as an author's associated literary groups (e.g. the Inklings), an author's primary genre of writing or the genre of each work, and dynamic network measures to better understand the kinds of social connections that are most important in the production process and to understand the boundaries between 'social space' and 'idea space' for authors.

Chapter 2

London Calling? Agglomeration Economies in Literature since 1700

2.1 Introduction

The potential for gains from agglomeration of industry is evident for industries with traditional modes of production. However, less is known empirically about these dynamics for industries with few capital requirements and no apparent need to cluster geographically, such as literary production. Unlike a factory or science laboratory, writing does not require a centralised location of production. There are few capital requirements for the writing of literature, and the printing process does not require the author to be present. Authors are able to send drafts and conduct most business with a publisher or literary agent through the post. Authors are also highly mobile. Given these unique characteristics, do authors tend to cluster geographically at all? If so, are there gains in productivity associated with the agglomeration of literary activity? In order to answer these questions, I determine the cities in which literary activity clusters, and I empirically estimate these returns to literary agglomeration.

Specifically, I utilise a unique, purpose-built dataset with information on the birth location and lifetime migration, productivity (in terms of number of publications), and demographic characteristics of 370 authors in the UK and Ireland since 1700. I begin by discussing the data collection methodology. I then analyse the patterns of migration and clustering of authors, and I construct age-productivity profiles to determine the productivity gains (if any) associated with the geographic clustering of literary activity. Individual fixed effects are utilised to control for spatial sorting by skill level, and the within estimator reveals the average author annual productivity is 10.68% per annum higher when the author resides in London, the only major literary cluster during the time period. There is no evidence of dynamic learning effects, and there is no evidence that the results are driven by a dynamic self-selection process. There is also no evidence that authors strategically build a portfolio of works before moving to London and publishing them in quick succession.

The remainder of this chapter is structured as follows. Section 2.2 outlines the theoretical framework by discussing the relationship between geographic proximity and tacit knowledge, agglomeration economies and recent empirical evidence. A brief history of publishing

is provided in Section 2.3. Next, I discuss the data collection methodology and descriptive statistics in Section 2.4. Section 2.5 outlines the identification strategy, and results are presented in Section 2.6. Section 2.7 concludes.

2.2 Theoretical Framework

2.2.1 Geographic Proximity and Tacit Knowledge

One of the earliest economic explanations of geographic clustering of industry was posited by Alfred Marshall:

> *"When an industry has thus chosen a locality for itself, it is likely to stay there long: so great are the advantages which people following the same skilled trade get from near neighbourhood to one another. The mysteries of the trade become no mysteries; but are as it were in the air, and children learn many of them unconsciously. Good work is rightly appreciated, inventions and improvements in machinery, in processes and the general organization of the business have their merits promptly discussed: if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas. And presently subsidiary trades grow up in the neighbourhood, supplying it with implements and materials, organizing its traffic, and in many ways conducing to the economy of its material."* (Marshall 1890, p. 225)

More formally, the geographic concentration of industry occurs because firms benefit from economies of scale via increased input-output linkages, labour market pooling, and technology spillovers (Krugman, 1991; Duranton and Puga, 2004; Combes and Gobillon, 2015).

Duranton and Puga (2004) discuss the micro-foundations of Marshallian agglomeration economies and divide them into broad types based on the mechanisms that drive their formation: sharing, matching, and learning. Duranton and Puga highlight several types of

learning mechanisms that are particularly relevant in the context of creative production: knowledge generation, diffusion and accumulation. Their discussion of these mechanisms alludes to two important concepts – that of 'space of places' (the importance of location for learning and innovation) and that of 'space of flows' (the importance of networks in the transfer and diffusion of knowledge) (Ter Wal and Boschma, 2009). The importance of these concepts in learning-based agglomeration economies becomes clear with a careful understanding of the difference between information and knowledge, particularly with regard to tacit knowledge.

Tacit knowledge is gained through an individual's perception of information through repeated interaction in a shared learning process; this contrasts with explicit knowledge, which is information that can be codified and stored in media such as an encyclopaedia or a textbook (Howells, 2012). Tacit knowledge is transmitted most efficiently between individuals in close proximity, so there are natural geographic boundaries to its flows and spillovers.[1] Thus, it is natural that "innovative activity should concentrate geographically in those industries where the direct knowledge-generating inputs are greatest and where knowledge spillovers are the most prevalent" (p 630, Audretsch and Feldman, 1996).[2]

However, it is argued that geographic proximity is a necessary but not a sufficient condition for the transfer of tacit knowledge (Boschma, 2005; Torre and Rallet, 2005). Torre and Rallet (2005) outline four additional types of proximity that, in combination with geographic proximity, provide a sufficient condition for tacit knowledge transfer: cognitive proximity (a shared knowledge base), social proximity (socially embedded relationships), institutional proximity ('common habits, routines, established practices, rules, or laws'), and organisational proximity ('the ability of an organisation to make its members interact').[3] Yet, these types of proximity are difficult to quantify, and data on the individual or firm level is largely unavailable. Because of these limitations, research on learning-based

[1] See Audretsch, 1998; Gertler, 2003; Von Hippel, 1994 .

[2] As Glaeser et al. (1992, p 1126) note, "After all, intellectual breakthroughs must cross hallways and streets more easily than oceans and continents."

[3] Also see Mokyr, 2005; D'Este et al., 2013; Torre and Rallet, 2005; Boschma, 2005; Hellmanzik, 2013; Audretsch, 1998; Rallet and Torre, 1998.

agglomeration economies and knowledge spillovers has primarily relied on geographic proximity as a proxy for latent tacit knowledge transfer.

Authors are anecdotally described as solitary and independent workers, so authorship could be an inherently individual process and thus authors would not benefit to the same degree (if at all) from agglomeration. However, authors are cognitively proximate by the nature of their profession, and it is likely a sample of authors will be socially or institutionally proximate if the sample is limited to a relatively small geographic region, as they are likely to share a common language and culture and interact with the same members of a wider literary-industrial complex (publishers, editors, critics, etc.) Thus, tacit knowledge transfer is possible in this context. Given these necessary conditions for tacit knowledge transfer, I hypothesise that productivity gains from geographic co-location can (and do) exist for literary production.

2.2.2 Agglomeration Economies and Productivity

In addition to driving the formation of agglomeration economies, location and proximity also play important roles in increasing productivity. Geographic proximity increases the potential for tacit knowledge transfer which then, in turn, increases firm or individual productivity (Combes and Gobillon, 2015). These gains attract new firms that relocate to exploit the gains, further increasing the geographic concentration of innovative activity. Chatterji et al. (2013) find decreasing returns to such spillovers, while Ter Wal and Boschma (2009) warn of cognitive lock-in – the risk of getting 'locked in' to established ways of thinking instead of continuing to develop new ideas increases as the density of a cluster increases (also see Boschma, 2005; Jones, 2009; Audretsch, 1998).

This cyclical relationship between location, proximity, and knowledge spillovers also makes identification of causal relationships difficult (Audretsch and Feldman, 2004). Porter (2011) discusses how competition and rivalry due to geographic proximity increase overall standards and the probability competing firms or individuals will be considered 'prominent.'

Kelly and O'Hagan (2007), however, warn that the direction of the clustering/prominence relationship is not clear and may even be cyclical: some prominent agents may choose to cluster to reap the benefits of doing so, while others become prominent because of the gains they received from working in a cluster.

More generally, an important on-going debate is on the role of localisation (an increase in the concentration of an industry) versus urbanization (an increase in the size of a city) in facilitating innovative activity (Rosenthal and Strange, 2004). The oldest account of localisation is Marshall (1890), and through the years, this theory was advanced and formalised into what is known today as the Marshall-Arrow-Romer (MAR) externality. The MAR externality suggests that the higher degree to which firms are identical, the lower transaction costs and communication costs are. This, in turn, leads to a higher probability of knowledge spillovers between firms and, thus, more innovative activity (Glaeser et al., 1992). Porter (2011) also suggests that concentration of an industry promotes innovative activity, although he argues that local competition (rather than a local monopoly) is the driving force.

While theories of localisation stress that spillovers occur between identical firms within a single industry, theories of urbanisation argue that spillovers occur between industries. An early and influential discussion of the role of urbanisation is Jacobs (1969), who focuses on the importance of urban diversity for innovation. The motivation behind theories of urbanisation is perhaps best illustrated by Duranton and Puga (2004) who note that "it is very difficult to conceive how interactions within an 'army of clones' could generate sufficient benefits to justify the existence of modern cities" (p 2110).[4]

The greatest challenge to detecting spillovers is the 'Marshallian equivalence' of agglomeration mechanisms – that many different factors can result in the same observable outcome (Duranton and Puga, 2004; Carlino and Kerr, 2015). While increased concentration of industry can increase the number of new ideas, Helsley and Strange (2002) argue that "it is possible that concentration encourages innovation by assisting in the realization of ideas

[4]Also see Audretsch and Feldman, 2004; Helsley and Strange, 2002; Rosenthal and Strange (2004); Duranton and Puga, 2004; among others for further discussion of the 'localisation vs urbanisation' debate.

rather than in (or perhaps in addition to) the generation of ideas" (p 26). Their model is based on the matching of entrepreneurs' ideas with local inputs to produce an innovation. In markets with dense networks of input suppliers, this process of input sharing reduces the cost of realising ideas. Carlino and Kerr (2015) also support the idea that innovative activity can be facilitated by thick local markets that enable more efficient sharing of inputs and provide a larger pool of specialised workers and services, but they note that "knowledge can flow rapidly through these linkages," which highlights the difficultly of empirically disentangling these effects.

2.2.3 Agglomeration of the Creative Industries

Furthermore, there have been attempts to estimate the gains from economic clustering across the creative industries and understand how these dynamics change by field. Perhaps the most prevalent method of studying the gains from geographic proximity is by "following the paper trail" in analysing the spatial clustering of patent registrations from R&D by private firms and universities. (See Carlino et al., 2007; Jaffe et al., 1993; Moser et al., 2014; Audretsch and Feldman, 2004; Azoulay et al., 2010; Audretsch and Feldman, 1996, among others.) Audretsch and Feldman (2004) note: "the basic results agree: patent citations are highly localized, indicating that location and proximity clearly matter in exploiting knowledge spillovers" (p 2723).

A more recent set of literature analyses agglomeration and knowledge spillovers by studying the individual scientist, often in a historical context. Waldinger (2012a) evaluates the productivity losses (in terms of academic publications rather than patents) associated with the breakdown of research clusters (i.e. university science departments), utilising the dismissal of mostly Jewish scientists and the destruction of facilities by Allied bombing in Nazi Germany as exogenous shocks. He finds significant decreases in departmental productivity, due to both human capital and physical capital shocks. Waldinger (2012b, 2010) conduct similar studies on university mathematics and science departments in Nazi Germany, finding that higher cluster density and higher quality of peers within the cluster

both increase the productivity of researchers and have a positive effect on the outcomes of the PhD students within their department.

Another new body of literature looks at the clustering of broadly defined creative activity. Borowiecki (2013) conducts a historical study examining the impact of geographic clustering on the productivity of classical composers. He finds that clustering has a positive and significant impact on composer productivity and finds heterogeneity of returns across clusters and composer quality. Borowiecki (2012) and Borowiecki and O'Hagan (2013) follow this by analysing the impact of war on clustering and life-cycle creativity. They find that the share of composers drops during periods of conflict, composer productivity is negatively impacted by the presence of a conflict, and the outward migration of composers significantly reduces a country's long-term creative potential. The gains from the clustering of prominent visual artists are analysed by Hellmanzik (2010), who finds that the artwork of prominent visual artists is valued higher when produced in a creative cluster. She also finds heterogeneity of returns by cluster location and artist quality. Giorcelli and Moser (2015) exploit the timing of Napoleon's military victories to determine the effect of copyright on the creativity of Italian opera composers. They find that copyrights lead to an increase in both the number of operas and the number of high-quality operas, and they find that cities with pre-existing industry-related infrastructure received greater gains.

2.3 A Brief History of Publishing in Britain

By the early 17th century, the publication of literature by artistic patronage was in decline, and a commercial market primarily based in London began to flourish under a system of perpetual copyright regulated and enforced by the Stationers' Company (a London-based guild that held monopoly power over the publishing of books in the United Kingdom).[5] The first copyright law was introduced in 1710 when the British parliament passed the *Statute of Saint Anne* (or *Copyright Act of 1709*). This Statute mandated that all books

[5]See Encyclopaedia Britannica (2014) sections on '*Copyright*' and the '*History of Publishing*' for more information.

be given a copyright of 14 years if published after April 10, 1710 or 21 years if published before.[6]

While this Statue legally applied in both Scotland and England, the English publishers acted as a cartel, maintaining their monopoly over English publishing and continuing to uphold the previous system of perpetual copyright. In Scotland, however, the courts actively enforced the copyright law. While a small book printing and publishing industry had existed in Scotland since the early 16th century, this divergence in enforcement of copyright allowed for a dramatic expansion of the publishing and book manufacturing industry in Scotland and particularly in Edinburgh.[7] Not only were the Scottish printers able to produce books at lower cost, they were also able to sell books in England – even as far south as Oxford and Cambridge – on credit for prices "three or more times more favourable than those offered by London" (St Clair 2004, p. 106-107). By the 1750s, the Scottish printing industry also offered lower prices on printing equipment.

From the date copyright was introduced, the London booksellers and printers fought the copyright law in court and appealed to parliament for a repeal, eventually engaging in a 30-year legal 'battle of the booksellers' that ended in 1774 with the *Donaldson v. Becket* judgment upholding statutory copyright limits (Rose, 1988). However, St Clair (2004) notes that "the London publishers continued to deal in legally unenforceable perpetual intellectual properties for at least forty years after 1774" (p. 111). While there were other changes in copyright law through the 19th and 20th centuries, none resulted in further major divergences between Scottish and English publishing industries, as the London cartel's monopoly power over copyright continued to weaken. Edinburgh and London remained

[6]See St Clair (2004) and *'History of Publishing'* in Encyclopaedia Britannica (2014).

[7]St Clair (2004) notes: "At the time Scottish judges made their first decisions, there was no substantial book industry in Scotland, nor any powerful constituency of intellectual property owners. The effect was therefore not only to create a difference in what was legally permitted in the two countries of the union, but to set up economic incentives for the development of a Scottish book industry, as a kind of offshore rival to England but without any of the usual legal, physical or tariff barriers...To reprint a title which had passed into the public domain under the terms of the 1710 Act was not illegal, nor subject to injunction, in Scotland, nor was it illegal to sell such books in England or in the British colonies in America and elsewhere...By the middle of the eighteenth century, a substantial Scottish book industry had grown up in Edinburgh, and others in Glasgow and Aberdeen, publishing a variety of newly written texts under statutory copyright, but mainly reprinting works which lay outside the limits of the 1710 Act. The Scottish publishers were able to do this profitably at prices which were between half and two-thirds of the levels of the London industry." (p. 105)

the main book manufacturing centers for the Romantic and Victorian periods and into the 20th century (St Clair, 2004).

Commercial borrowing was an established industry by the 17th century, and circulating lending libraries were the main source of reading through most of the 18th century.[8] However, subscriptions at that time were primarily purchased by aristocrats and wealthy professionals. "Authors who aspired to a larger readership could send short pieces to a newspaper or a literary review. Wordsworth, Southey, and Coleridge all started their literary careers this way" (St Clair 2004, p.159). By the end of this period, dozens of periodicals were publishing poetry, prose and short stories, and commercially successful authors could receive good payments for their submissions (St Clair, 2004).

Poetry, periodical essay, and drama were the most popular genres through most of the 1700s. While the novel began to develop as a literary genre in the 1700s, it did not become the leading literary genre until the Victorian era in the following century. The Triple-Decker, or three-volume novel, was developed by Edinburgh-based Archibald Constable in the 1820s as a more profitable way to publish novels (Sir Walter Scott's work, in particular), and this system of publishing novels became the standard form during the 19th century when borrowing from commercial circulating libraries that could be accessed via subscription was still a common form of book consumption.[9] The 'Railway Library' made its debut in Euston Station, London, in 1848, and similar bookstalls offering inexpensive reprints (for prices as low as one shilling) opened at railway stations across Britain soon after (Barnes and Barnes, 2004). While the difference in copyright enforcement stimulated the development of a competitive printing market leading to the production of inexpensive, mass-produced books, this development of a widespread and easily accessible bookselling market helped to expand book consumption and widened readership.

With large publishing industries in both Edinburgh and London, a growing readership across the British Isles and abroad, and the ability to submit writings by post, where

[8] "By 1801, there were believed to be about 1000 circulating libraries in Great Britain...and their numbers had grown to about 1500 in 1821, with many more established later" St Clair (2004), p. 237.

[9] See Hewitt (2004) and '*Library*' in Encyclopaedia Britannica (2014).

would authors chose to live and work? Is the average author like the Brontë sisters who are often styled as solitary geniuses – or more like the Lost Generation of Paris in the 1920s, a dense network of artists, musicians, writers, and friends? The next section describes the authors studied in this chapter and highlights the trends in migration and geographic clustering.

2.4 Data

2.4.1 Selection of Authors

This chapter utilises a unique dataset combining location and biographical information of prominent literary artists constructed for the purposes of this study. These dataset was developed by collecting every individual associated with British or Celtic literature born between 1700–1925 with an entry in Encyclopaedia Britannica Online.[10] To be defined as an author for the purposes of this study, an individual must have made at least one unique contribution to poetry or prose, which eliminated individuals whose contributions were strictly limited to translations, textbooks, manuals or guides, songwriting, literary criticism, or publishing.

Individual-level data was collected from three online encyclopaedias: Encyclopaedia Britannica (2014), The Literary Encyclopaedia (2014), and Literature Online (2014). This data include age, lifespan, age at first publication, number of publications per annum, lifetime publications, career duration, gender, and location for every year of the author's life. I rely on observable measures of productivity (output per annum) for the analysis. Output is measured in terms of number of known publications rather than number of works ever written. Publications include any type of unique literary publication: novel, collection of poetry or short stories, anthologies, contributions to literary magazines, plays,

[10]Due to posthumous publishing and potential biases that may be associated with an increased interest and analysis in an author immediately following his or her death, living or recently deceased literary artists were not included. Thus, all literary artists in this dataset passed away before the year 2000.

memoirs, etc. Individuals who only published non-fiction works (textbooks, dictionaries, journalism, etc.) and/or translations are not included. All publications were given equal weight. There is no measure of the quality of individual publications. Of the roster of 537 significant figures, only 370 have complete or near-complete lifetime location and publication information.[11] This sample of 370 authors is used in the econometric analysis in the following sections.

Following the methodology of Murray (2003) and Verboord (2003), I create a purpose-built index of author quality or impact, defined as the amount of contemporary critical attention in academic literary studies given to one author relative to her peers. This index is constructed from two components: an indexed measure of the total words designated to an author from the three biographical sources and an indexed measure of the total number of citations within literary criticism as listed in the Literature Online (2014) database. These measures carry equal weight in the index. A sample of the top 30 literary artists and their respective Impact Index value is listed in Table 2.1 in Appendix 2.B.2.

A more detailed description of the data collection process may be found in Appendix 2.A.1. A more detailed description of the creation of the Impact Index may be found in Appendix 2.A.2.

2.4.2 Migration Trends and Summary Statistics

The Greater London Area is the location with the most births of literary artists, as 79 (or 21.35%) of the artists were born within this region. (See Table 2.2 in Appendix 2.B.2.) I also evaluate spatio-temporal trends in lifetime movement to determine if authors tend to migrate and cluster in certain locations. As seen in Figure 2.1, clustering intensity is surprisingly high throughout the sample. London consistently emerges as the largest

[11]The term 'near-complete' is used specifically in regard to lifetime location because it was not unusual for individuals to have an unknown location for less than five years of their life. Unknown locations were particularly common during periods of conflict for males participating in military service. Such individuals were kept in the sample. With regard to publications, authors either had a comprehensive bibliography listed in Literature Online or a few select works listed across all three online dictionaries. Authors who do not have a comprehensive bibliographical entry were not included in the sample.

literary cluster. Until the end of the 19th century, over 30% of all authors in a given year reside in London. (See Appendix 2.A.1 for the definition of London over time.) A Herfindahl-Hirschman Index (HHI), defined as

$$HHI = \sum_{i=1}^{N} s_i^2$$

where s_i is the share of author-year observations of location i and N is the number of locations, can provide a more formal measure of author concentration. The HHI has a value of 0.177, indicating a moderate degree of concentration.[12]

At its peak, London was home to over 50% of all authors. Dublin, Edinburgh, Oxford, and Cambridge emerge as the only other cities that see minor clustering of authors at any point in the sample. Due to the small sample sizes in these minor clusters (consistently fewer than five authors clustering in each city per annum), I define London as the only literary cluster for the primary analyses in this study. Due to the small sample sizes, the minor clusters are only utilised in a robustness analysis to determine if locating in any one of the minor clusters results in an increase in productivity. (See Section 2.6.1.)

For comparison, Table 2.3 provides the ten most populous cities in England, Scotland and Wales [13] in the 1851 census and the top ten destinations for authors (in terms of author-year observations). Census data on city population is drawn from the Online Historical Population Reports (2016). While this only provides a single year observation for the most populous cities in England, there tends to be persistence in spatial distribution over time.[14] As can be observed in Table 2.3, authors did not tend to live in large British cities, with the exception of London. Manchester, for example, is the 9th most common author destination, yet 74 of the 100 author-year observations can be attributed to two authors (Samuel Bamford and Elizabeth Gaskell) who spent a substantial part of their lives in the city. All other authors lived in Manchester for less than 10 years. While there is some

[12]For comparison, a market with a similar HHI value may be subject to scrutiny under US competition laws. See US Department of Justice and the Federal Trade Commission (2010) for further information.

[13]Historical population data on Ireland are largely unavailable before the 1901 census, as only fragments remain of the 1821-1851 censuses. See National Archives of Ireland (2016) for more information.

[14]See Online Historical Population Reports (2016) for further additional census population reports.

overlap in the years Bamford and Gaskell spent in Manchester, there is little to no overlap in the times other authors lived in the city. Therefore, it is not possible to disentangle a 'large city' effect from an 'intellectually diverse city' effect in this context of this chapter.

It is notable that the clustering intensity of authors in London decreased from nearly 50% at the end of the 19th century to around 25% during the 20th century. (See Figure 2.1.) This migration of authors out of London during this period was not accompanied by an increase in the clustering intensity in the minor literary clusters or a similar decrease in their clustering intensity. The decline of London in the early 20th century could be due to authors participating in or fleeing from the two World Wars. It could also be due to the rise of Paris as a major destination for authors by the 1920s. This may be an interesting and useful context to investigate the decline of one cluster and the rise of another; although, this particular research question is beyond the scope of this chapter.

I also investigate the variation in movement in and out of London, as a key concern is that London arises as a cluster only due to its role as a major location of birth. Of the sample of 370 authors, 71 authors never lived in London and only 4 authors never left London, indicating that approximately 80% of the sample spent at least some time living in London. Thus, the clustering intensity of London is not driven by the fact that London is a key birth location and is instead due migration in and out of London.

The trends in birth versus lifetime location reveal a surprising degree of mobility, migration, and clustering. The degree of clustering is significant in and of itself. There is no infrastructure required in the writing of books. Although there is infrastructure required in the production of books, little (if any) of the printing process required the author to be present. Transcripts could be mailed through the post, and many business dealings could be handled by a publisher or agent. "Felicia Hemans, for example, who lived in Wales, built her literary reputation first with Murray in London and later with Blackwood in Edinburgh, without visiting either city" (St Clair 2004, p. 159).

While the importance of London as a literary cluster is not surprising, the fact that centres of learning and culture emerge as the only other city-level clusters rather than simply

large, urban areas is revealing. Gains from agglomeration of literary activity will not be due to economies of scale effects associated with general production in a large city. If the geographic concentration of authors was only due to availability of local publishing infrastructure, ease of social access to publishers, lower costs of production, etc., then one would expect more authors to be located in Edinburgh. Instead, it suggests that having a diverse intellectual elite may be a more important factor driving migration and clustering, supporting the Jacobs (1969) theory that diversity of industry facilitates knowledge exchange and increases innovative activity. However, detailed data on other cultural industries or intellectual individuals is not available, so I am not able to explicitly test this hypothesis in this study.

Summary statistics for the sample of 370 authors are presented in Table 2.4. Authors tended to be engaged in work-related activities for a great portion of their lives. On average, authors lived for 66 years, published their first work at age 25 and worked for just under 40 years. Over the course of their careers, authors produced an average of 31 publications, publishing an average of 0.88 works per year. Over their lifetimes, authors spent around 21 years in London. There are some distinct gender differences, particularly with regard to total lifetime works, with male authors publishing around 10 more works over the course of their lifetimes.

Figure 2.2 shows the total number of authors and total output (of all authors) per annum over time. There is no significant change in total production for the first 150 years, followed by a sharp and steady increase in total production from around 1850. It is notable that this is not accompanied by a corresponding dramatic increase in the total number of authors.[15] The dynamics of publishing are further detailed in Figure 2.3, which decomposes the total number of publications between those published by authors in London and those published by authors living in all other locations. Until the turn of the twentieth century, authors

[15]This shift in total production could be due to the invention and widespread use of industrial printing press during this time or due to the continual increase in the total number of authors led to an increased competitiveness that manifested itself in increased total output. Perhaps, as well, this is due to rise in literacy and the advent of the "penny dreadful" and "shilling shocker" in the mid to late 19th century. Most likely, it is some combination of all these factors (St Clair, 2004).

residing in London consistently produce at least as many publications as all authors in all other locations combined.

2.5 Identification Strategy

I begin by determining if living in a geographic cluster (London) results in an increase in productivity and estimating these returns. A standard approach is to estimate the following relationship by ordinary least squares using a pooled panel of workers:

$$Output_{it} = \delta Location_{it} + \beta \mathbf{x}_{it} + \upsilon_{it} \tag{2.1}$$

in which some measure of productivity (output per annum) is a function of the location in which worker i resides in time t, a vector of individual-specific characteristics, and an error term.

An OLS estimation of Equation 2.1 will yield unbiased and consistent results only if $\mathbf{x}_{it}$ is measured without error; however, some elements of $\mathbf{x}_{it}$ are unobservable. Specifically, Equation 2.1 suffers from potential source of bias, as it does not account for spatial sorting by worker skill. There is an endogenous relationship between productivity and location, as there is likely to be some unobserved individual-level characteristic, such as natural ability, that is linked to an individual's likelihood of migrating to a particular location. If individuals with higher unobserved ability are more likely to migrate, then the OLS estimate of $Location_{it}$ will be upwardly biased. The bias will be downwards in the opposite case.

A standard solution is to find an exogenous instrument for location choice and then estimate

the pooled OLS.[16,17] However, if panel data on individual workers is available (as is the case in this chapter), Glaeser and Maré (2001), Combes et al. (2008), de la Roca and Puga (2012) argue that the use of individual fixed-effects can adequately address the issue of workers sorting across locations on unobservables.[18]

I first provide OLS estimates for the pooled panel of authors as a baseline estimate of the correlation between location and productivity along the cross-section. I then address the identification issue discussed above by estimating the following equation:

$$output_{it} = \delta London_{it} + \beta_1 age_{it} + \beta_2 age_{it}^2 + \sigma_i + \mu_t + \epsilon_{it} \tag{2.2}$$

where $output_{it}$ measures the number of known works by author i published in time t.[19] The variable of interest, $London_{it}$, is a binary variable equal to 1 if author i was living in London in time t. The variables age_{it} and age_{it}^2 control for author-specific life-cycle ageing effects, with the quadratic term allowing for productivity to diminish as a author's age increases. μ_t is a vector of time dummies to account for yearly changes in productivity that impact all authors in the same way. I also include a full set of author fixed effects, given by σ_i. The combination of the linear component of the age profile, 1-year dummy variables and individual author fixed effects can lead to the "age-time-cohort" identification problem, so I repeat this estimation with 2-year and 5-year time dummies.

Furthermore, there are many years in which authors do not publish at all. In years in

[16] Ciccone and Hall (1996) suggest using long lags of population density to instrument for the size or density of local population. They argue that the spatial distribution of the population persists over time and that this differs substantially from the contemporary factors impacting productivity on the local level.

[17] Borowiecki (2013) uses birthplace-cluster distance to instrument for the spatial sorting of prominent Western classical music composers, as birthplace-to-cluster distance captures not only the costs associated with travel but also factors such as cultural and linguistic distance. Since most specifications define London as the only cluster, it is not possible to use a time-invariant instrument and individual fixed effects. Additional detailed individual-level information would be needed to control for individual heterogeneity in order to utilise birthplace-London distance as an instrument, which is often not available in the limited historical sources. Furthermore, given the limited geographic scope of this chapter, the travel distances are much smaller and the culture and language less varied. Particularly after the development of extensive rail lines by the mid-1800s, distance across the UK was not likely such a deterring factor in moving to London.

[18] Combes and Gobillon (2015) note that "as long as individual location decisions depend only on the explanatory terms introduced in the specification, which can go as far as the individual fixed effect, some time-varying individual characteristics such as age, and a location-time fixed effect, there is no endogeneity bias" (p 286).

[19] I rely on observable measures of productivity for the analysis. As mentioned in the previous section, productivity is measured in terms of number of known publications rather than number of works written. I have no measure of the quality of individual publications.

which they do publish, most authors publish only a single work per year, with increasingly fewer having two, three, etc. publications per year, as seen in Figure 2.4. Therefore, I limit the study to potential active years, which I define as age 16 and older. This represents the lower bound for age at first publication. Additionally, many authors in the sample began university at age 17 or 18. Thus, potentially the first migration decision the authors made themselves would have been made around age 16 or 17. Because of this, the sample used in the analysis begins in 1725 (the first year with multiple observations) and ends in 1999 (when the last authors died). I also supplement the analyses with a negative binomial model, which is a more appropriate specification for over-dispersed count data (Wooldridge, 2010).[20]

2.6 Results

The estimates for Equation 2.2 are presented in Table 2.5. Column (1) shows the estimates of the OLS relationship between locating in the London cluster and the number of publications in a given year using the pooled panel of authors, utilising cross-sectional variation to identify the effect of living in London. In Column (2), I introduce author fixed effects, drawing on temporal variation within individuals to identify the effect of living in London while controlling for sorting by skill level. Column (3) includes the coefficients for the negative binomial model with individual fixed effects, and Column (4) provides the respective incident rate ratios. In all four models, robust standard errors are clustered on the individual level to account for serial correlation within individuals.[21]

As seen in Column (1), much of the variation in production between individuals cannot be explained by location. However, when individual fixed effects are introduced in Column

[20] Although not reported, the conditional variance of the output variable is greater than the conditional mean; therefore, the data is over-dispersed.

[21] Clustering of standard errors on the individual level is utilised rather than multi-way clustering on the individual and location levels because many cities were inhabited by only one author. There are fewer than 42 locations with consistently more than 1 author, and multiple authors per location is required in order to separate the idiosyncratic variation in output across people from the variation across location. Therefore, it is possible that the bias introduced by clustering on too few clusters is greater than the potential reduction in bias from multi-way clustering of standard errors. (See Angrist and Pischke (2008).)

(2), an author residing in London experiences productivity gains of 0.094 additional works per annum compared to years of her life when residing elsewhere. Given that the mean output per annum is 0.88, this translates to a 10.68% increase in annual productivity.[22] The negative binomial incident rate ratios (NB IRR) reported in Column (4) indicate that an author sees a 24.3% increase in the probability of publishing in a given year while residing in London compared to years of her life when residing elsewhere. In all specifications, location is not correlated with productivity in the cross-section, but individual authors do experience productivity gains from locating in a geographic cluster throughout their life-cycles. If there existed a correlation between location and productivity in the cross-section, this would suggest that standard large-city economy of scale effects are a primary underlying factor. The fact that there is only a significant relationship within author observations suggests that there exists some sort individual-level process resulting in productivity gains.

While there is a positive link between locating in London and creative activity, the specific channels through which these gains occur have not been identified. It is unlikely that such gains occurred because of reputational effects, at least not because authors gained any London-specific reputation in the eyes of the reading public. Romantic poets and Victorian novelists were typically anonymous, particularly for an author's first book.[23] A few authors, such as Edgeworth, Godwin and Amelia Opie, wrote under their own name for later works, but the reading public generally did not demand to know or appear to care about authors' identities (St Clair, 2004). This is perhaps illustrated best by the fact that "throughout the Romantic period, all novels, whether anonymous or not, were normally classified in booksellers' lists and in circulating library catalogues, alphabetically by title, not under the name of the author" (St Clair 2004, p. 174).[24]

[22]The introduction of individual-level fixed effects increased the R^2 from 0.103 to 0.120. The fact that there was not a more substantial increase indicates that time invariant characteristics, such as gender, social class, or level of education (attained before age 15), does not explain a large share of the variation in literary production in a given year. It is possible that time variant factors beyond those included in the regression could have greater explanatory power, but this data is difficult to consistently collect for all authors (if it is available at all) due to the nature of historical data.

[23]For example, Erickson (2002) estimates that by the "1770s sixty percent of all poetry books were published anonymously, but by the period's final decade from 1826 to 1835 only twenty-five percent were" (p 247).

[24]The 'Author of Waverly' (Sir Walter Scott) is a notable exception as one of the few novelists that the reading public actively tried to identify. Jane Austen's identity, on the other hand, was only revealed years after her death.

It is also important to note that these results show the impact of residing in London relative to periods of an author's life when she resided elsewhere. If living in other large British cities (e.g. Liverpool or Manchester), one of the minor clusters (Cambridge, Edinburgh, Dublin or Oxford) or another large, culturally diverse city outside the British Isles (e.g. New York or Paris) also result in productivity gains, then the effect of living in London may be seen as a lower bound. I argue this increase in creative activity could be due to knowledge spillovers, but the 'Marshallian equivalence' of this and other mechanisms cannot be empirically disentangled.

In all specifications, age has a positive but diminishing effect, which is consistent with general findings in the literature on individual productivity.[25] The R-squared value of the fixed effects model is also notable: 12% of the variation in productivity can be explained by only four variables (age, London, and time and author fixed effects). Creative production is often anecdotally viewed as a highly ethereal process – one that is organic and impalpable, beyond quantification. Yet, a substantial amount of literary production can be explained by a simple life-cycle production process and geographic proximity, not so dissimilar to the production processes of R&D researchers. (See Levin and Stephan, 1991.)

Tables 2.6 and 2.7 provide the results with 2-year and 5-year time dummies, respectively. These results are consistent with the previous estimates, although the effect is more modest. With 2-year (5-year) dummies, an author living in London increases productivity by 0.08 (0.09) works per annum and her probability of publishing is increased by 22% (23.5%) while living in London compared to living elsewhere.

2.6.1 Robustness Analysis

Spatial Selection as a Dynamic Process

Empirically, establishing a causal link is a challenge due to the potential endogeneity of the propensity to cluster. This is controlled for by the inclusion of author fixed effects if the

[25] See, for example, Levin and Stephan (1991).

only source of endogeneity is spatial sorting by skill level; however, it is still possible that there is a dynamic feedback effect between past output and the probability of migrating to London. In this case, the strict exogeneity assumption is violated, and the estimates in Table 2.5 will be biased. The primary concern in this context is that authors' recent career success – who are not necessarily those with greater innate ability – systematically migrate to London at higher rates. In this case, the fixed effects estimates will be overestimates of the true effect. Similarly, if authors who experience recent career difficulty move to London at systematically higher rates (potentially believing this will improve their chances of experiencing a 'big break' or will provide them with more career alternatives if their pursuit of a literary career is unsuccessful), the fixed effects estimates will be a lower bound. I explicitly address this issue by using a logit regression to determine the role of past success (in terms of output) in the probability of migrating to London.[26]

In Table 2.8, I present logit regressions predicting the probability of migrating to London to examine the role of past success in determining migration. In Column (1), I use the previous year's output to examine if immediate past success increases the probability of migration. In Column (2), I use the total output to date to determine if cumulative success increases the probability of migration. In both specifications, previous success has no statistically significant impact on the probability of migrating to London. The effects of both output measures are negative, suggesting that greater past success decreases the probability of migrating. I repeat these specifications by quality level (low vs. high)[27] and by century of birth to determine if these results are being driven by a particular cohort. In both specifications, the results are consistent (see Tables 2.9 and 2.10 in the Appendix, respectively). Thus, I conclude that the results on the gains from the geographic clustering of literary activity in London are not driven by the systematic self-selection of recently successful authors.

[26] I use 25-year dummies rather than year dummies for the logit models to be parsimonious with the degrees of freedom.

[27] See Section 2.4.1 and Appendix 2.A.2 for further discussion of the quality variable.

Agglomeration Effects in Other Cities

It is possible that authors also receive productivity gains from locating in other cities. Authors may receive a 'large city' effect from locating in one of the most populous cities. Particularly for skilled workers, such cities would offer more alternatives if a literary career was unsuccessful or if the author needed to supplement their income from writing with other sources. Authors may also receive productivity gains from locating in Edinburgh, in particular, due to the well-developed local book publishing infrastructure and low costs of production. Authors may benefit from locating in smaller university cities, in which exposure to a diverse intellectual elite promotes an increase in creative activity. The primary difficulty in examining any of these potential factors is the low degree of concentration in non-London cities. As discussed in Section 2.4.2, authors did not tend to locate in the most populous cities in the United Kingdom. The only non-London cities that are consistently a location of residence for multiple authors in the same year are Cambridge, Dublin, Edinburgh and Oxford.

Because there are so few observations for the minor cluster cities (Cambridge, Dublin, Edinburgh and Oxford), it is not possible to evaluate the effect of each city separately within the framework utilised in this chapter. It is possible to determine the effect of an author locating in any one of the minor clusters by defining a general 'minor clusters' indicator: a binary variable equal to 1 if the author lived in Cambridge or Dublin or Edinburgh or Oxford in a given year and equal to 0 otherwise. Even at this level of aggregation, 48.58% of the sample of authors have a standard deviation of 0, indicating they either spent no time in any of the minor clusters or they spent their entire lives in the clusters. Given that so few authors were present in any of these cities in a given time (as discussed in Section 2.4.2), I can assume most of these authors never lived in any of the minor clusters.

I follow the same strategy described in Section 2.5 and present the results in Table 2.11. Columns 1 and 2 provide the coefficients for an OLS fixed effects estimation, and Columns 3 and 4 present negative binomial fixed effects incident rate ratios. Columns 1 and 3 only

include the 'Minor Clusters' indicator, while Columns 2 and 4 also include the indicator for residing in London. These results should be interpreted cautiously, as there is not much within-author variation for the 'Minor Clusters' variable. Locating in any one of the minor clusters has a negative and significant impact on productivity in the OLS fixed effects specification but does not have a significant effect when controlling for periods when residing in London. In the negative binomial models, locating in a minor cluster has no significant impact on productivity, although the coefficient also indicates a negative relationship.

In general, these results suggest that there is something unique about the conditions of London that is not present in other literary cities, including the rival publishing city of Edinburgh. Access to literary infrastructure alone is not enough for authors to receive an increase in productivity. Clustering intensity could be one factor. It could be that there is a minimum number of authors needed in a location for gains to realise. Productivity is likely to increase as the number of peers increases due, for example, to an increase in local competition, stronger local social networks, a reduction in how costly it is for a new innovation to be realised or increased knowledge spillovers. Unfortunately, there is not enough variation in cluster size to test this hypothesis. On the other hand, the diversity of industry or concentration of other creative industries in London could also be important features. Without detailed data on such industries, this cannot be empirically tested either.

Duration Effects

In the main analysis, the regression specifications have the explicit assumption that individuals receive an instantaneous productivity gain upon arrival in London with no change to the rate of productivity. However, Glaeser and Maré (2001) and de la Roca and Puga (2012) argue that this approach to agglomeration is too simplistic. Individuals receive a static, instantaneous gain from agglomeration (e.g. from immediate access to better local resources), but they may also receive a dynamic gain reflecting learning that accumulates over time. De la Roca and Puga (2012) find that the magnitude of this dynamic gain is

positively related to city size, and they argue that the presence of dynamic agglomeration effects is clear evidence of learning and tacit knowledge spillovers. Helsley and Strange (2002), however, argue that input sharing and knowledge spillovers both take time to realise, and a similar argument could be made for labour market pooling in the case of authors.

This theoretical distinction is illustrated in Figure 2.5. The solid black (lowest) line represents the age-productivity profile of an individual who does not live in London. The solid red (middle) line represents the age-productivity profile of an individual upon moving to London under the assumption that there is only a static gain. The dashed red (upper) line represents both the fixed gain and the dynamic gain. Mathematically, the static gain is represented by an increase in the intercept of the original age-productivity function, and the dynamic gain is represented by an increase in the slope of the original age-productivity function.

This graphical analysis is also useful in assessing another potential issue. It is possible that authors may build a large portfolio of writings before moving to London. Upon arrival, the author may make connections with local literary agents and publishers who then publish a number of the author's works in quick succession. In this case, the estimate for *London* will be an overestimate of learning effects as it would also capture the fact that authors strategically withhold writings before 'dumping' them on the market within the first years of arriving in London. Although the average London effect indicates an increase in productivity relative to the periods the author does not reside in London, the author would actually be receiving a sharp spike in productivity immediately after arrival in London before experiencing a drop to at or near pre-migration levels.

Following the methodology of de la Roca and Puga (2012), I consider the following extension in order to simultaneously estimate the dynamic and static agglomeration effect of living in London:

$$\begin{aligned} output_{it} =& \delta London_{it} + \gamma_1 Exper_{ict} + \gamma_2 Exper_{ict}^2 + \theta_1 (Exper * London)_{it} \\ &+ \theta_2 (Exper * London)_{it}^2 + \beta_1 age_{it} + \beta_2 age_{it}^2 + \sigma_i + \mu_t + \epsilon_{it} \end{aligned} \tag{2.3}$$

in which $Exper_{ict}$ is defined as the number of consecutive years individual i has spent in city c since the year of migration to that city, and $Exper_{ict}^2$ allows for diminishing returns to duration in a city. The terms $Exper * London_{it}$ and $Exper * London_{it}^2$ are the interaction between London and the experience effects, providing the effect of the number of consecutive years an author has spent in London.

The results are provided in Table 2.12. Column (1) provides the effect of *London* and *Experience* before the introduction of the interaction term. The net duration effect, which is found by estimating this model without *London*, is given in Column (2). The breakdown of this effect into static and dynamic effects is shown in Column (3): the estimate for *London* gives the fixed effect and the estimate for *Experience* gives the dynamic effect. Life-cycle aging effects added in Column (4). The corresponding negative binomial incident rate ratios for these models are given in Columns (5) to (8), respectively.

Plotting the point estimates allows for the visual detection of the existence of 'dumping' while also determining if gains from agglomeration accumulate. I re-estimate the full model specified by Equation 2.3 using experience dummies rather than an experience trend using OLS with author fixed effects.[28] Figure 2.6 provides an illustration of this post-migration premium by static, dynamic and net agglomeration effects. In this graph, the London premium in time 10 is interpreted as the productivity of individual i in the 10th consecutive year in London after migrating relative to the years she was not in London.

As can be seen in Table 2.12 and in Figure 2.6 (and in contrast to the findings of de la Roca and Puga (2012)), there is little evidence to support the hypothesis that the gains from agglomeration accumulate over time above life-cycle aging effects. This implies that

[28]These results are not reported but are consistent with those in Table 2.12.

individual i does not accumulate knowledge or skill in a way that increases annual productivity in terms of number of works produced. This exercise is also useful for visualising the net agglomeration gains by comparing the pre- and post-migration trends. To do this, I re-estimate Equation 2.3 with dummies for duration in London and for each of the 5 years before migrating to London. These estimates are illustrated in Figure 2.7. There is a distinct jump in productivity once the individual moves to London. Relative to the periods in which individual i is not in London, the gains from living in London are essentially the same whether she has lived in London for 2 consecutive years or 10 consecutive years.

There is no evidence of strategic 'dumping' within the first few years of arrival in London. Authors receive relatively constant gains for the duration of their stay in London. Given the types of contracts that were most common, this is not surprising. The most common contract was a lump sum payment made to the author in exchange for the exclusive copyright to the work. Under this type of contract, the publisher assumed all financial risks, but the author would not receive any additional payment if the publication became more successful than anticipated.[29] If a book was commercially successful, an author could demand higher payment for later works. Authors could also agree to sell the copyright for a set edition size and renegotiate payment for future editions, or authors could agree to share the net profits. Neither of these contract types were as common, as authors were (often rightly) suspicious that publishers were not being honest in their records of sales or costs. Authors could also chose to self-finance their publication or publish 'on commission' by raising funds through local donors.[30]

These types of contracts may have provided the incentive to withhold publication so as not to produce 'too many' books simultaneously: if an author's work became commercially

[29] St Clair (2004) notes that this type of contract was not always beneficial for the publisher or for the author: "When the Bristol publisher, Joseph Cottle, bought Lyrical Ballads from Wordsworth in 1798, he made a poor bargain, for when he transferred all his business assets to Longman not long after, the copyright was valued at nil. By contract when, in the same year, a Glasgow publisher agreed to print The Pleasures of Hope by the twenty-year old Thomas Campbell in return for the right to buy copies at the trade discount, fortune favoured the publisher. The poem was an instant and continuing bestseller, netting thousands of pounds over the following decades, but nothing more for Campbell except an occasional ex gratia payment which did more to enrage than to soothe his feelings. Campbell never overcame his sense of having been cheated." (p. 161)

[30] A few authors, such as Coleridge, Byron and Shelley, began successful literary careers this way (St Clair, 2004).

successful, she could then bargain for higher pay for the next book published. Perhaps living in London allowed authors to make better bargains with local publishers, allowing them to publish at a more steady rate and for higher pay – i.e. by "reducing the costs of realizing ideas" (Helsley and Strange, 2002). Although, presumably this would also have been possible in Edinburgh – perhaps even more so given the concentration of authors was much lower, providing Edinburgh-based authors easier access, less competition and, for a time, lower costs of production.

It could be that London-based authors had access to better local social networks (both within literary circles and across other creative industries), potentially providing both information on local publishers and literary critics and preferential access to inputs through social ties. It could be that authors in London simply needed to produce at a higher and more steady rate to cover the costs of living in London. More importantly for the learning hypothesis, it is also possible that the author accumulates skills that result in producing *higher quality work* rather than a slow increase in *quantity* of work. That output does not measure quality is an important fact and needs consideration, but capturing quality of output is a challenge that is not unique to this study and is one not easily resolved. In general, without detailed data on any of these factors, it is not possible to assess the underlying driving forces beyond speculation.

2.7 Conclusion

This study contributes to current research by empirically analysing the positive externalities associated with the agglomeration of literary activity in historical Britain and Ireland. Specifically, I utilise a unique dataset with information on the birth location and lifetime migration, productivity (in terms of number of publications), and demographic characteristics of 370 authors in the UK and Ireland since 1700 that I constructed for the purposes of this study. I analyse the migration and clustering of literary artists and found that London was the major literary cluster throughout the sample. I then construct age-productivity

profiles to estimate the productivity gains associated with the geographic clustering of literary activity.

I find that authors who reside in London experience productivity gains of 10.68% per annum compared to their peers living elsewhere. These results are not driven by self-selection of recently successful authors in migrating to London. This increase in productivity is primarily due to a static agglomeration effect and not due to a dynamic learning effect. There is also no evidence that authors strategically withhold writings until migrating to London and then 'dump' them on the market.

The geographic concentration of creative workers is certain to have played an important role in the transfer and diffusion of knowledge, as well as the generation of new knowledge and ideas. Authors in London likely had access to stronger and more advantageous social networks, in terms of increased connections with their peers (other authors), individuals with influence within the publishing industry (agents, publishers, critics, etc.), and those who are a part of the intellectual and cultural elite (artists, musicians, wealthy patrons). Authors in London also could have taken advantage of the related economic infrastructure and gained from the resulting economies of scale, allowing for a more efficient transformation of ideas into physical book-form. However, further research is needed to fully understand the channels through which authors receive these gains and the degree to which these mechanisms are geographically localised.

Due to the nature of the data used in this study, broad generalisations of these results are limited. However, the insights into the positive externalities associated with the agglomeration of an industry with few capital requirements are still of relevance to economic researchers, and the contributions to broader understanding of how to access and enhance knowledge spillovers are of interest to both firms and public policy-makers. For example, parallels may be drawn between the creative processes of the historical author and the contemporary software developer. As software development continues to shift from more tradition modes of production to production via remote work, it will be important to understand if (and why) agglomeration of IT firms persists and how individual productivity

will be impacted. In this, historical analyses of literary production and other creative industries may be particularly useful.

2.A Detailed Data Collection Methodology

2.A.1 Selection of Authors

I aim to identify the centres of co-location (or the geographic clustering of individuals) in order to examine the geography of literary production. In order to do so, I must first systematically determine which individuals to study. In recent years, there have been several attempts to identify significant figures across various fields. Perhaps the most comprehensive of these is the work of Murray (2003), in which Murray identifies and ranks over 4,000 of the most globally-significant individuals across the arts and sciences from 800 B.C. to 1950. Murray develops his inventories of figures by determining which individuals were consistently included across a large cross-cultural and multilingual collection of sources, including histories, encyclopaedias, and biographical dictionaries. Murray's work includes a roster of 835 significant figures in Western literature. However, Murray's data on individual authors is limited to name, year of birth, national origin, and an index score indicating relative significance. Therefore, additional biographical sources were need to generate the panel dataset needed for this analysis.

The cultural bias of encyclopaedias, biographical dictionaries, and similar resources has been well documented in works such as Murray (2003), O'Hagan and Kelly (2005), and O'Hagan and Borowiecki (2010). As Murray (2003) explains, this cultural bias is more prominent in literature than other artistic and scientific fields. Additionally, comprehensive literary biographies are quite limited in number. Reference works on literature tend to be specific to one period, movement, and/or language and often include more literary criticism than detailed biographical information. It would require a multi-linguistic effort to systematically and consistently collect biographical information, including information on lifetime location and migration, for all authors. Therefore, due to natural constraints, Murray's inventory of significant figures was not deemed usable for the purposes of this study.

Consequently, I begin by limiting the scope culturally, linguistically, geographically, and temporally to individuals associated with British or Celtic literature born between 1700–1925 with an entry in Encyclopaedia Britannica Online.[31] While there are other countries that are culturally and linguistically similar (e.g. the United States or Canada), transcontinental and transoceanic migration and clustering across such large distances was not common for most of the sample time period. For the purposes of this study, richness of data over time was preferred to richness of data across geographic reach. A key implication of this is that all results may be specific to the culture and social norms of the location and period, and thus it may not be possible to generalise the results to all authors globally and temporally.

I develop a unique inventory of literary artists by collecting every individual associated with British or Celtic literature born between 1700–1925 with an entry in Encyclopaedia Britannica Online. To be defined as a literary artist, an individual must have made at least one unique contribution to poetry or prose, which eliminated individuals whose contributions were strictly limited to translations, textbooks, song-writing, literary criticism, or publishing. Due to posthumous publishing and potential biases that may be associated with an increased interest and analysis in an author immediately following his or her death, living or recently deceased (i.e. post-1999) literary artists were not included. This gave an inventory of 537 literary artists.

Data on the location and migration of the sample of authors was collected from Encyclopaedia Britannica, Literature Online, and The Literary Encyclopaedia. Location information was collected on four levels within the British Isles: city, ceremonial county or lieutenancy area, a modified NUTS 1 regional definition, and country.[32] The modified NUTS 1 regional definition includes all NUTS 1 level statistical regions for England, Scotland, and Wales. England is composed of 9 NUTS 1 regions. Scotland and Wales each compose a single

[31] Celtic literature, in this particular usage, is defined as literature associated with Celtic nations within the British Isles (Ireland, Scotland, Isle of Man, Wales, and Cornwall) and does not strictly refer to literature written in the Celtic languages. 'British' literature is not used as an all-encompassing term due to the long and complicated history associated with the concept of 'British identity', the discussion of which lies far beyond the aims of this paper.

[32] NUTS 1 statistical regions are the *Nomenclature of Territorial Units for Statistics, Level 1* regional standards developed by the European Union.

NUTS 1 region. This definition is not used for the island of Ireland, as it lists Northern Ireland and the Republic of Ireland as separate NUTS 1 regions. Historically, these divisions are not so distinctive. Therefore, the traditional provinces of Ireland (Leinster, Ulster, Munster, and Connacht) are used to instead of the NUTS 1 regions. For locations outside the British Isles, information was collected on three levels: city/state, country, and a general category designating distance from British Isles (Rest of Europe, North America, and Rest of World.) In addition, London's city borders have changed dramatically over time, and sources often ambiguously report an individual as "moving to London" rather than specifying that person's exact location. Particularly in the latter part of the sample, the "moving to London" may refer to any part of the metropolitan area. Therefore, I define 'London' as the Greater London ceremonial county rather than the area within the official City of London limits or the London borough.

The roster was reduced further only based on availability of relevant biographical information. To be included in the final roster, authors must have 'near complete' lifetime location information. It was not unusual for individuals to have an unknown location for less than 5 years of their life. Unknown locations were particularly common during periods of conflict for males participating in military service. Such individuals were kept in the sample. With regard to publications, authors either had a comprehensive bibliography listed in Literature Online or a few select works listed across all three online dictionaries. Authors who do not have a comprehensive bibliographical entry were not included in the sample. These exclusions gave a final roster of 370 individuals.

Following the definitions used by Murray (2003), these 370 authors represent the most significant or prominent figures in British and Celtic literature during the period 1700-1999. As this set represents the upper-tail of the author distribution, the results of this study may not be generalisable to the entire author population, both globally and within the British Isles. For example, these prominent authors are likely to have advantageous relationships with other authors, publishing houses, critics, and other individuals associated with the creative industry. Therefore, these 370 authors may also be more likely to benefit from geographic clustering of artistic activity and from author social network connects. Thus,

the results should be interpreted as applying to the 'literary elite' rather than authors in general.

2.A.2 Quality of Authors

Murray (2003) uses an advanced column-inch methodology – which relies on the principle that the amount of space devoted to an individual within a source (i.e. total column inches), the more significant that person is – to construct a measure of the relative significance of each person within a field.

Most studies, however, concentrate on developing methods to rank individuals within a specific field. Galenson (2002) uses a textbook illustration method to identify the most significant French artists of the 19th century. This methodology is similar to that of the column-inch but is somewhat more labour-intensive: once the most important sources have been determined, one simply counts the number of illustrations reproduced in each book for the artist. Galenson defines a significant figure as having illustrations in no less than three of his five primary sources and then ranks their significance using the total number of illustrations from a larger sample of 33 art history textbooks. O'Hagan and Kelly (2005) compare the column-inch and textbook-illustration ranking methodologies for 35 visual artists and contrast these results to those of Galenson. They find that both methods are useful for broadly grouping artists by significance and that there is little evidence to suggest that more elaborate ranking methods produce better results.

Verboord (2003) focuses on the relative significance or 'literary prestige' of 502 authors born after 1880. Verboord's measure of significance is composed of a range of indicators within primary areas: number of literary prizes won, entries in literary encyclopaedias (measured by word count rather than column inches), number of academic studies mentioning the author, and the literary reputation of the author's publishing house.[33] Verboord's

[33]The literary reputation of the author's publishing house is determined by the total number of prestigious authors on the publisher's list. This definition assumes that the prestige of the publishing house is likely to influence the prestige of the individual author.

methodology includes diverse and comprehensive indicators, but it is not possible to use all of these indicators to estimate literary prestige for earlier time periods. Consistent and comprehensive information on authors' publishing houses is difficult to obtain, particularly information from before the mid-1800s. Furthermore, it is not possible to use the number of literary prizes won for earlier time periods, as literary prizes are quite a recent invention. Literary prizes were limited to schools and universities until the mid-nineteenth century, and the earliest national literary prize in Great Britain was not founded until 1919 (Shaffer, 2008).

It should be noted that while those in the literary and wider art world are often hesitant to attach ordinal rankings to artists or to designate one author as indisputably "better than" another, the previously discussed methodologies do not create an arbitrary cardinal measure of quality (e.g. the stars rating system generally associated with movie ratings.) Instead, such rankings simply reflect the amount of critical attention given to one author relative to her peers. Furthermore, previous studies have shown that systematic rankings based on measurable attributes can provide comparable judgement to experts in the field and, in some cases, may even outperform them. Throsby (1994), for example, found that the aesthetic judgements of art experts are not random but systematically based on observable characteristics such as artist's school of work, career stage, and past achievement.

The methodology used to determine significance in this study is similar to that of Verboord (2003); though, the indicators in this study are limited to entries in literary encyclopaedias and attention given in literary studies. Similar to the measures of significance used in other studies, the Impact Index simply reflects the amount of contemporary critical attention an individual author has received relative to her peers.

Sources of information on the biographies and critical reception of the authors was limited to three online encyclopaedias: Encyclopaedia Britannica (2014), The Literary Encyclopaedia (2014), and Literature Online (2014). In addition to providing a biographical entry for each author, Literature Online also contains a database of literary criticism and reference works, such as ABELL (Annual Bibliography of English Language and Literature 1920-),

MLA International Bibliography, an electronic library of 392 full-text journals of literary criticism, an indexed directory of 4,400 periodicals, and a number of other reference works. Each author's main page includes links to journal articles, citations, and reference articles about that author. Journals of literary criticism include peer-reviewed academic articles on new research, reviews or critiques of previous works, and book reviews. The criticism component of the Impact Index was created by recording the total number of entries under each author's respective criticism page.

For the word-count component of the Impact Index, total word count rather than column-inch method is used because differences in font and format made cross-resource column-inch comparisons infeasible. Additionally, the absolute word count is used rather than a measure weighted by the total length of the source. The length of an entry relative to the total length of each encyclopaedia could not be determined because there exists no simple method of determining the complete length of these online resources due to the format of the online systems. Furthermore, due to the nature of online hosting, publishers do not experience the same space limitations associated with physical encyclopaedias, and so the authors of each biographical entry has greater freedom to expand.

2.B Tables and Figures

2.B.1 Figures

Figure 2.1: Share of Authors in Major Cities

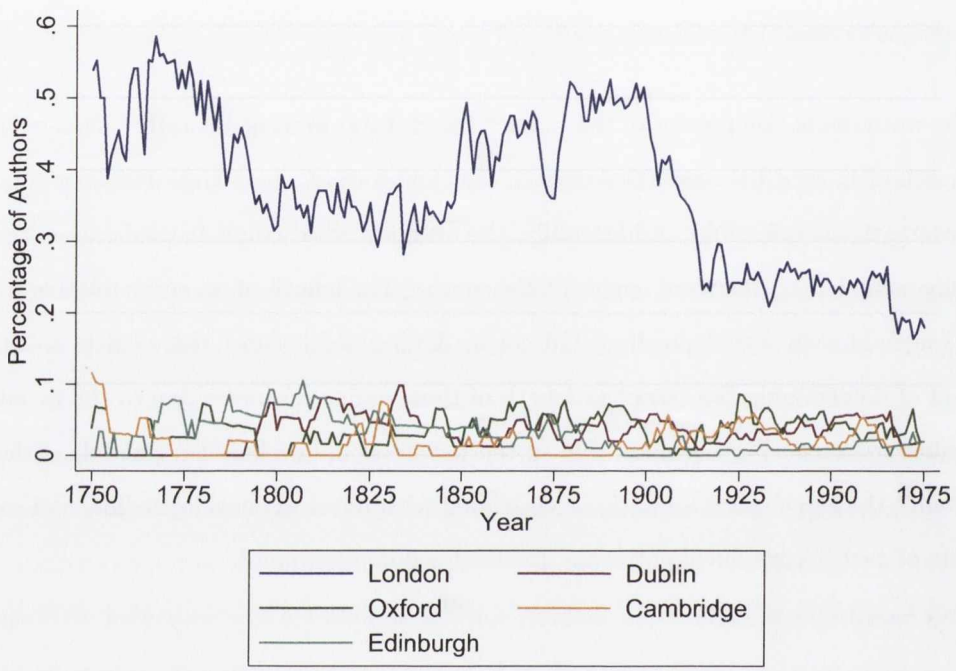


Figure 2.2: Total Authors and Output per Annum

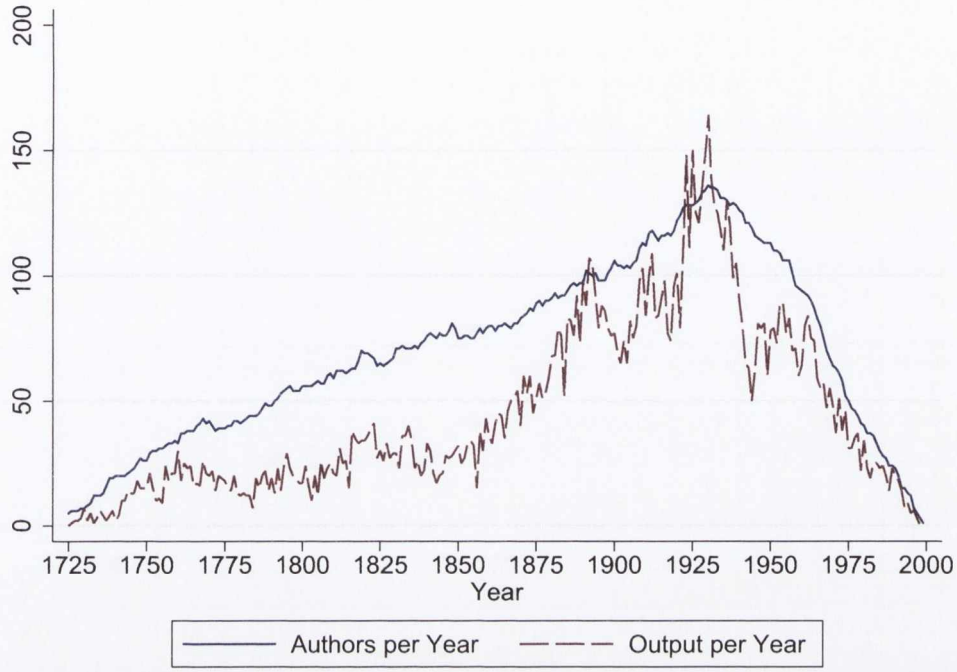


Figure 2.3: Total Output per Annum by Location

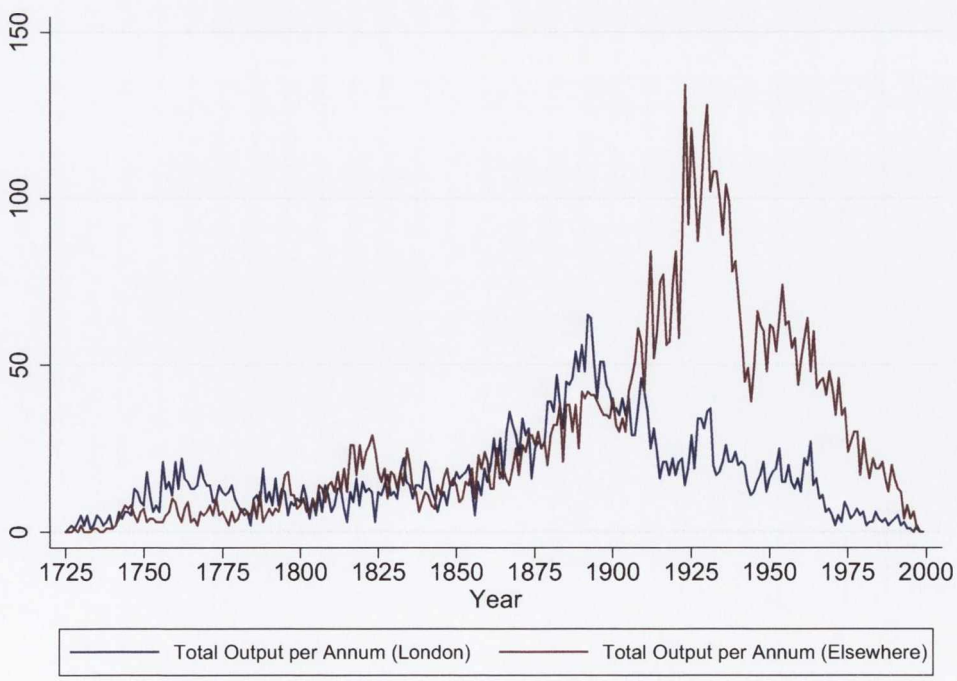


Figure 2.4: Frequency of Output Observations

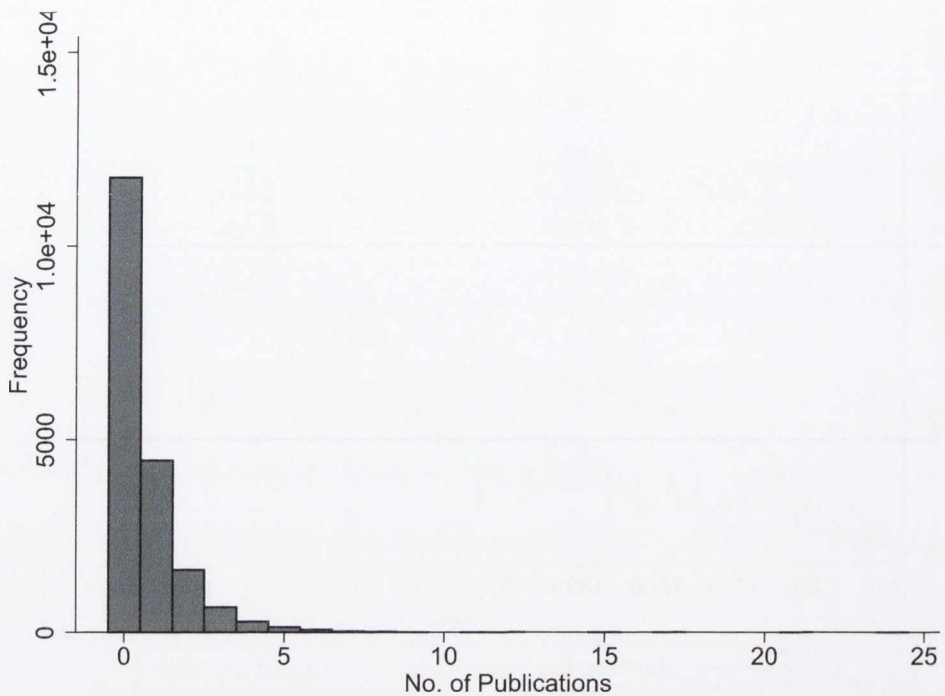


Figure 2.5: Impact of Static and Dynamic Effects on Age-Productivity Profiles

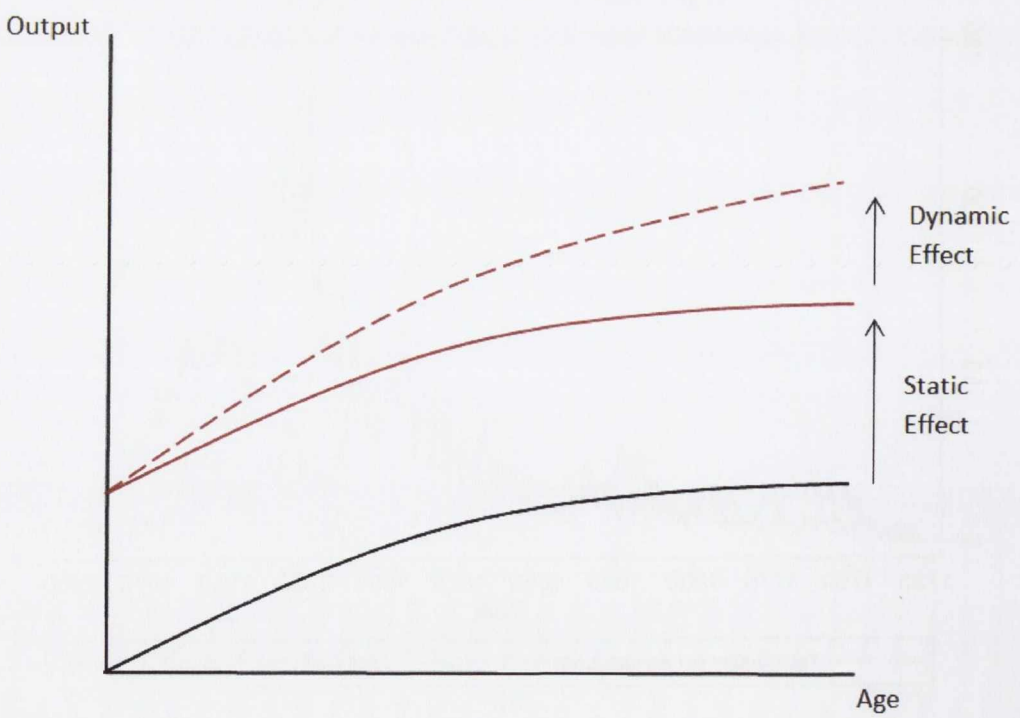


Figure 2.6: Estimated London Premium

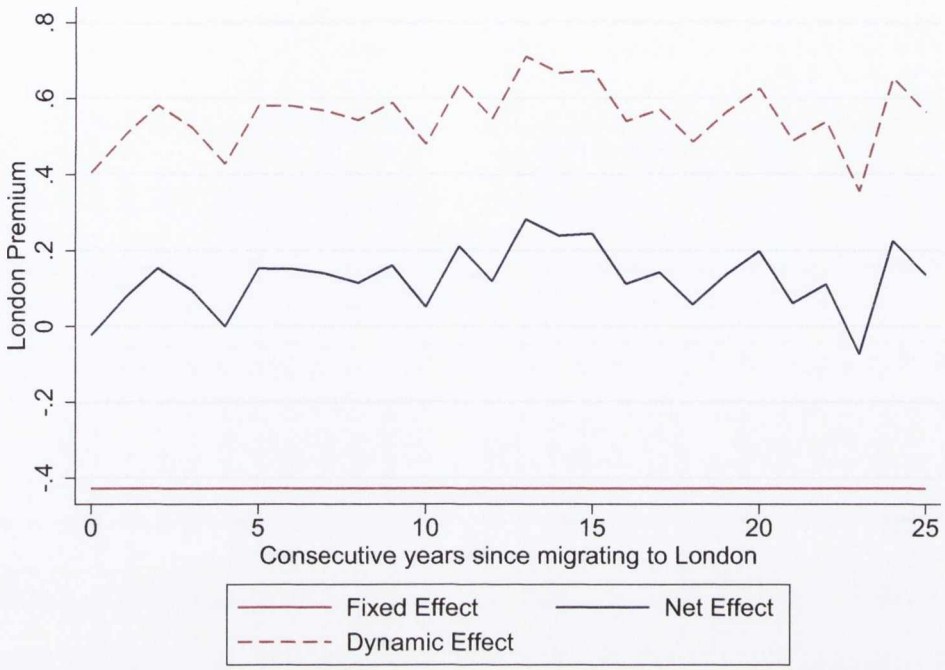
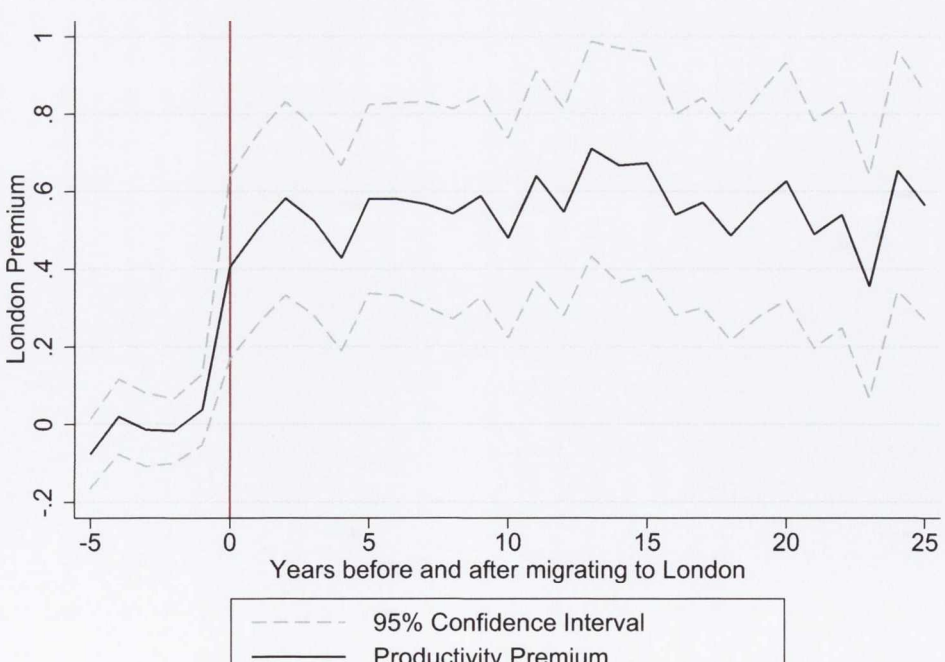


Figure 2.7: Estimated Pre- and Post- Migration Trend



2.B.2 Tables

Table 2.1: Top 30 Literary Artists by Ranking

Literary Artist	Word Count Index	Citation Index	Impact Index
Joyce, James	1	1	100
Dickens, Charles	0.75	0.76	75.47
Lawrence, D. H.	0.69	0.43	56.39
Conrad, Joseph	0.60	0.50	55.07
Woolf, Virginia	0.80	0.25	52.61
Blake, William	0.76	0.23	49.29
Wordsworth, William	0.56	0.42	48.86
Coleridge, Samuel Taylor	0.63	0.34	48.81
Austen, Jane	0.52	0.43	47.20
Eliot, T. S.	0.42	0.49	45.46
Yeats, William Butler	0.68	0.21	44.53
Byron, George Gordon	0.61	0.28	44.48
Beckett, Samuel	0.63	0.22	42.67
Johnson, Samuel	0.74	0.11	42.52
Hardy, Thomas	0.41	0.42	41.78
Auden, W. H.	0.66	0.13	39.33
Orwell, George	0.64	0.13	38.26
Shaw, George Bernard	0.50	0.24	37.02
Eliot, George	0.42	0.31	36.28
Keats, John	0.49	0.23	35.89
Kipling, Rudyard	0.55	0.14	34.15
Shelley, Percy Bysshe	0.46	0.22	34.00
Moore, George	0.64	0.03	33.81
Browning, Robert	0.45	0.22	33.75
Carlyle, Thomas	0.55	0.11	33.05
Wilde, Oscar	0.43	0.22	32.50
Stevenson, Robert Louis	0.56	0.09	32.29
Betjeman, Sir John	0.63	0.00	31.55
Meredith, George	0.57	0.06	31.25
Alfred, Lord Tennyson	0.41	0.21	31.16

A note on interpreting these values: the highest-ranked author is James Joyce. James Joyce has the most number of words in his biographical entries and the highest number of citations, so his Word Count Index (WCI) and Citation Index (CI) values are both equal to 1. The second-highest ranked author is Charles Dickens. Dickens has a WCI value of 0.75 (rounded to two decimal places), indicating that Dickens's biographical entries contained 75% the number of words as those of Joyce. Dickens has a CI value of 0.76, indicating he received 76% the number of citations as Joyce. Thus, Dickens' Impact Index value is 75.47 – the average of his WCI and CI values normalised to 100.

Table 2.2: Number of Authors Born from 1700–1925, by Birth Location

Region	1700–1724	1725–1749	1750–1774	1775–1799	1800–1824	1825–1849	1850–1874	1875–1899	1900–1925	Total Authors per Region
Connacht	0	0	0	1	0	0	3	0	1	5
East Midlands	1	2	1	1	4	3	0	3	1	16
East of England	2	1	4	0	6	5	0	3	3	24
Greater London	3	5	7	7	13	11	9	15	9	79
Leinster	0	2	1	6	2	1	4	4	4	24
Munster	1	0	0	1	1	1	0	0	2	6
North East England	1	0	0	0	1	0	0	1	0	3
North West England	0	0	1	3	3	1	3	2	3	16
Rest of Europe	0	2	0	1	0	1	5	1	1	11
Rest of World	0	0	0	0	1	2	4	6	5	18
Scotland	3	3	7	6	2	4	7	5	4	41
South East England	3	2	1	4	6	7	7	8	7	45
South West England	3	2	4	3	3	5	1	3	2	26
Ulster	0	0	0	0	0	0	1	4	2	7
Wales	0	1	0	0	0	1	3	0	5	10
West Midlands	2	1	0	0	2	1	3	6	6	21
Yorkshire and the Humber	0	0	0	0	3	1	1	4	5	14
Unknown	1	1	1	0	0	0	0	0	1	4
Total Authors per Period	20	22	27	33	47	44	51	65	61	370

Table 2.3: Most Populous Cities (1851 Census) vs Top Author Destinations

	England, Wales and Scotland*		Author Destination	
Rank	City	Population	City	Author-Year Obs.
1	Liverpool	375,955	London	6,620
2	Manchester	303,382	Oxford	777
3	Birmingham	232,841	Dublin	651
4	Leeds	172,270	Edinburgh	484
5	Glasgow	148,116	Cambridge	425
6	Bristol	137,328	Paris	265
7	Sheffield	135,310	New York	125
8	London^{+}	127,869	Bath	105
9	Bradford (York)	103,778	Manchester	100
10	Newcastle upon Tyne	87,784	Norwich	84

Source: Population data is taken from the 1851 Census of Great Britain (Online Historical Population Reports, 2016). Data on top author destinations comes from author's calculations.

* Historical population data on Ireland are largely unavailable before the 1901 census, as only fragments remain of the 1821-1851 censuses. See National Archives of Ireland (2016) for more information.

+ London here refers to the borough of London within the *City* of London, as noted in the footnote of the original census table. For further discussion on the difficulties of defining 'London' over time, see Section 2.4.1.

Table 2.4: Summary Statistics by Gender

		Median	Mean	Std. Dev.
Male	Lifespan	69.00	65.96	16.43
N = 300	Age at First Publication	26.00	26.46	7.27
	Years in London	15.00	20.83	20.80
	Lifetime Works	23.00	34.70	32.44
	Career Length	41.00	39.50	17.23
	Works per Annum	0.67	0.92	0.79
Female	Lifespan	70.50	68.56	15.58
N = 70	Age at First Publication	27.50	28.33	8.06
	Years in London	21.50	24.29	20.51
	Lifetime Works	18.50	27.00	30.84
	Career Length	43.00	40.23	17.43
	Works per Annum	0.55	0.68	0.56
Total	Lifespan	69.00	66.45	16.28
N = 370	Age at First Publication	26.00	26.81	7.45
	Years in London	16.00	21.48	20.76
	Lifetime Works	21.00	33.24	32.24
	Career Length	42.00	39.64	17.24
	Works per Annum	0.64	0.88	0.76

Table 2.5: Main Results

	(1) OLS	(2) Fixed Effects	(3) Negative Binomial	(4) NB IRR
Age	0.0766***	0.0870***	0.149***	1.160***
	(0.00511)	(0.00549)	(0.00468)	(0.00542)
Age-Squared	-0.000728***	-0.000738***	-0.00157***	0.9984***
	(0.0000510)	(0.0000492)	(0.0000426)	(0.0000425)
London	0.0331	0.0944***	0.218***	1.243***
	(0.0528)	(0.0364)	(0.0322)	(0.0400)
Constant	-1.232***	-1.562***	-2.143***	0.117***
	(0.134)	(0.158)	(0.266)	(0.031)
1-Year Dummies	Yes	Yes	Yes	Yes
Author FE	No	Yes	Yes	Yes
R^2	0.103	0.120		
No. Authors	370	370	370	370
Observations	19022	19022	19022	19022

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses. Coefficients for the negative binomial are presented in Column 3. The corresponding incident rate ratios are presented in Column 4.

Table 2.6: Main Results with 2-Year Time Dummies

	(1) OLS	(2) Fixed Effects	(3) Negative Binomial	(4) NB IRR
Age	0.0753***	0.0762***	0.160***	1.176***
	(0.00498)	(0.00474)	(0.00397)	(0.00466)
Age-Squared	-0.000711***	-0.000712***	-0.00149***	0.9985***
	(0.0000500)	(0.0000473)	(0.0000406)	(0.0000406)
London	0.0248	0.0806**	0.199***	1.220***
	(0.0523)	(0.0374)	(0.0320)	(0.0391)
Constant	-1.064***	-1.120***	-2.183***	0.1127***
	(0.0855)	(0.106)	(0.128)	(0.014)
2-Year Dummies	Yes	Yes	Yes	Yes
Author FE	No	Yes	Yes	Yes
R^2	0.079	0.096		
No. Authors	370	370	370	370
Observations	19022	19022	19022	19022

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses. Coefficients for the negative binomial are presented in Column 3. The corresponding incident rate ratios are presented in Column 4.

Table 2.7: Main Results with 5-Year Time Dummies

	(1) OLS	(2) Fixed Effects	(3) Negative Binomial	(4) NB IRR
Age	0.0764***	0.0757***	0.150***	1.161***
	(0.00506)	(0.00592)	(0.00450)	(0.00523)
Age-Squared	-0.000726***	-0.000735***	-0.00156***	0.9984***
	(0.0000505)	(0.0000487)	(0.0000428)	(0.0000427)
London	0.0335	0.0939***	0.211***	1.235***
	(0.0524)	(0.0362)	(0.0323)	(0.0398)
Constant	-0.783***	-0.799	1.002**	2.723**
	(0.211)	(0.542)	(0.483)	(1.316)
5-Year Dummies	Yes	Yes	Yes	Yes
Author FE	No	Yes	Yes	Yes
R^2	0.097	0.111		
No. Authors	370	370	370	370
Observations	19022	19022	19022	19022

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses. Coefficients for the negative binomial are presented in Column 3. The corresponding incident rate ratios are presented in Column 4.

Table 2.8: Probability of Migrating to London

	(1) Output Last Year	(2) Total Output to Date
Age	0.968*	0.967*
	(0.0167)	(0.0166)
Age-Squared	1.000	1.000
	(0.000187)	(0.000183)
Output	0.993	0.997
	(0.0552)	(0.00603)
Year Dummies	Yes	Yes
Author FE	Yes	Yes
No. Authors	265	265
Observations	14040	13769

Logit odds ratios reported. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Probability of Migrating to London Based on Quality (High vs. Low)

	Low Quality		High Quality	
	(1)	(2)	(3)	(4)
	Output Last Year	Output To-Date	Output Last Year	Output To-Date
Age	0.948**	0.949*	0.982	0.982
	(0.0256)	(0.0255)	(0.0225)	(0.0226)
Age-Squared	1.000	1.000	1.000	1.000
	(0.000296)	(0.000293)	(0.000248)	(0.000241)
Output	1.066	1.002	0.944	0.992
	(0.0874)	(0.0102)	(0.0671)	(0.00809)
Year Dummies	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes
No. Authors	82	82	134	134
Observations	6986	6873	7054	6896

Logit odds ratios reported. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.10: Probability of Migrating to London by Century of Birth

	18th Century		19th Century		20th Century	
	(1)	(2)	(3)	(4)	(5)	(6)
	Output Last Year	Output To-Date	Output Last Year	Output To-Date	Output Last Year	Output To-Date
Age	1.001	1.015	0.970	0.968	0.978	0.992
	(0.0371)	(0.0391)	(0.0238)	(0.0235)	(0.0416)	(0.0422)
Age-Squared	0.999	0.999	1.000	1.000	1.001	1.000
	(0.000426)	(0.000417)	(0.000275)	(0.000276)	(0.000455)	(0.000444)
Output	0.888	0.971*	1.013	1.005	1.144	0.992
	(0.105)	(0.0162)	(0.0704)	(0.00731)	(0.168)	(0.0162)
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes	Yes	Yes
No. Authors	75	75	150	149	82	82
Observations	3683	3683	8163	7892	2194	2194

Logit odds ratios reported. Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.11: Effect of Minor Clusters

	OLS Fixed Effects		Negative Binomial IRR	
	Minor Clusters	All Clusters	Minor Clusters	All Clusters
Age	0.0761***	0.0865***	1.161***	1.160***
	(0.00499)	(0.00546)	(0.00540)	(0.00544)
Age-Squared	-0.000725***	-0.000735***	0.998***	0.998***
	(0.0000501)	(0.0000487)	(0.0000427)	(0.0000427)
Minor Clusters	-0.0969*	-0.0433	0.925	0.985
	(0.0511)	(0.0420)	(0.0466)	(0.0507)
London		0.0881**		1.241***
		(0.0371)		(0.0407)
Constant	-1.195***	-1.546***	0.128***	0.118***
	(0.133)	(0.158)	(0.034)	(0.031)
Year Dummies	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes
R^2	0.103	0.120		
No. Authors	370	370	370	370
Observations	19022	19022	19022	19022

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses.

The heading 'Negative Binomial IRR' denotes that the incident rate ratios are reported for the negative binomial regressions.

The 'Minor Clusters' is a binary variable equal to 1 if the author lived in Cambridge or Dublin or Edinburgh or Oxford in a given year and equal to 0 otherwise.

Table 2.12: Duration Effects

	(1) FE 1	(2) FE 2	(3) FE 3	(4) FE 4	(5) NB 1	(6) NB 2	(7) NB 3	(8) NB 4
London	0.133***		0.0728	0.0558	1.305***		1.148***	1.168***
	(0.0391)		(0.0451)	(0.0419)	(0.0408)		(0.0525)	(0.0533)
Duration in City	0.0182***	0.0118***	0.0136***	-0.00308	1.036***	1.024***	1.023***	0.992**
	(0.00367)	(0.00432)	(0.00455)	(0.00405)	(0.00290)	(0.00334)	(0.0036)	(0.0035)
Duration in City^2	-0.000481***	-0.000342***	-0.000374***	0.0000641	0.9991***	0.9993***	0.9993***	1.0002***
	(0.0000830)	(0.000102)	(0.000107)	(0.0000911)	(0.000061)	(0.000076)	(0.000080)	(0.000081)
Duration in London		0.0185***	0.0123*	0.00917		1.034***	1.022***	1.015***
		(0.00592)	(0.00712)	(0.00679)		(0.00415)	(0.00574)	(0.00571)
Duration in London^2		-0.000394***	-0.000286*	-0.000243		0.9994***	0.9996***	0.9996***
		(0.000146)	(0.000163)	(0.000149)		(0.000103)	(0.000125)	(0.000124)
Age				0.0857***				1.161***
				(0.00537)				(0.00575)
Age^2				-0.000728***				0.9984***
				(0.0000481)				(0.000046)
Constant	-0.202*	-0.164	-0.190	-1.541***	1.5007*	1.6340**	1.564**	0.120***
	(0.120)	(0.119)	(0.121)	(0.157)	0.33303	0.36184	0.34733	0.03252
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.077	0.077	0.078	0.121				
No. Authors	370	370	370	370	370	370	370	370
Observations	19024	19024	19024	19022	19024	19024	19024	19022

Robust standard errors are clustered on the author level and are reported in parentheses.

Coefficients for OLS fixed effects are presented in Columns 1-4, and incident rate ratios for negative binomials are presented in Columns 5-8.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Peer Effects in Literary Production

3.1 Introduction

Chapter 2 finds positive agglomeration effects for prominent literary artists living in London relative to periods in which the author lives elsewhere, but the mechanisms underlying these gains remain undiscovered. These findings relate to a broader 'black box' of agglomeration economies: "the many excellent studies of productivity have told us about the existence of agglomeration economies and also about their scope across industries, locations, and time. They have not, however, had much to say about the sources of agglomeration economies" (p 2148, Rosenthal and Strange, 2004.) This chapter aims to explore this 'black box' of agglomeration effects in literary production and provide greater evidence of potential channels through which these gains occur by determining how various peer effects impact own-productivity.

However, empirical evidence of peer effects is quite mixed. While some studies find evidence of peer spillovers, others find no evidence in similar contexts. For example, Waldinger (2012b) finds no evidence of peer effects between academics within departments but does find evidence of peer effects among co-authors, while Waldinger (2010) find that average faculty quality does have an impact on the outcomes of PhD students. Azoulay et al. (2010) argue such patterns are observed because spillovers are more likely to occur between individuals in 'idea space' than in 'physical space' or 'social space.' While this likely to hold for highly specialised fields requiring a large knowledge base and technical jargon such as STEM fields (i.e. science, technology, engineering, and mathematics), it is not clear whether the same restrictions should hold for other creative and cultural industries. I argue that the 'idea space' condition is likely met simply by authors writing in the same city at the same time.

In this chapter, I explore the hypothesis that physical proximity is a sufficient condition for peer effects to exist between authors, individuals who work in an industry that is relatively unspecialised and does not require as large a knowledge base compared to academic research in STEM fields, for example. In doing so, I also determine if the city is an appropriate

level of agglomeration for author peer effects to occur and explore several mechanisms that may underlie the agglomeration gains observed in Chapter 2. Specifically, this chapter utilises a unique dataset with information on the birth location and lifetime migration, productivity (in terms of number of publications), quality, and demographic characteristics of 370 authors in the UK and Ireland since 1700 that was constructed for the purposes of this study. I construct age-productivity profiles to determine the productivity gains associated with the mean quality of peers, mean peer output, and peer group size.

I also determine whether simply having peers (binary indicator) or whether categories of peer group size better capture the effect of number of peers. I explore peer effects within London, the only location with agglomeration effects as determined in Chapter 2. I also consider the possibility that certain types of peers may matter more. I investigate the effect of younger versus older peers: younger peers may represent up-and-coming generations, increased competition, while older peers may be mentors, the status-quo, or the 'giants' on whose reputation the author must build. I consider the possibility that living in the same city as a 'superstar' may matter more than average peer quality. Finally, I also look at effects by heterogeneity of own-quality, as own-quality may impact ability to form useful connections.

Individual fixed effects are used to control for spatial and group sorting. The within estimator reveals that peer group size has a modest negative effect on productivity, and these results are robust to several alternate specifications. When concentration of authors in London is controlled for, mean author quality while residing in London reduces productivity by 0.129 works per year. These results are primarily driven by older peers, although there is substantial heterogeneity of returns by own-author quality.

The remainder of this chapter is structured as follows. In Section 3.2, I outline the theoretical framework by discussing social interactions between peers, how these interactions relate to agglomeration economies and recent empirical evidence. Next, I discuss the data and provide descriptive statistics in Section 3.3. Section 3.4 outlines the identification strategy. The results are presented in Section 3.5. Section 3.6 concludes.

3.2 Existing Research

3.2.1 The 'Black Box' of Agglomeration Economies

Chapter 2 argues that the agglomeration gains observed for London-based authors were likely due to tacit knowledge spillovers, although the underlying mechanisms could not be confirmed. These findings align with general conclusions of many agglomeration studies: "location and proximity clearly matter in exploiting knowledge spillovers...[but] there is no understanding of the way in which spillovers occur and are realized at the geographic level" (p 2726, Audretsch and Feldman, 2004). Thus, these underlying mechanisms remain the 'black box' of agglomeration economies.

On a broad level, there is no consensus on whether localisation or urbanisation plays a greater role in facilitating innovative activity – or if the effect is industry-specific (Rosenthal and Strange, 2004). Theories of localisation argue that creativity and innovation are enhanced by the increased concentration of a single industry. Audretsch and Feldman (2004) explain:

> "*...to the degree that individuals in the population are identical and engaged in identical types of activities, the costs of communication and transactions are minimized. Lower costs of transaction in communication result in a higher probability of knowledge spilling over across individuals within the population. An important assumption of the model is that knowledge externalities with respect to firms exist, but only for firms within the same industry. Thus, the relevant unit of observation is extended from the firm to the region...but the spillovers are limited to occur solely within the relevant industry.*" (p 2727)

This theory of local monopoly is referred to as the Marshall-Arrow-Romer (MAR) externality. Porter (2011), on the other hand, argues that local competition between firms in the

same industry is the underlying force. However, Jacobs (1969) argues that urbanisation – diversity of industry – is more important for creativity and innovation.[1]

In both cases, agglomeration gains are dependent upon localised interactions of firms or individuals. Determining which factors are most important in these interactions is complicated by the fact that many different factors can result in the same observable outcome – a feature known as "Marshallian equivalence" (Duranton and Puga, 2004; Carlino and Kerr, 2015). For example, thick local markets may enable more efficient sharing of inputs and provide a larger pool of specialised workers and services, which will foster innovation (Carlino and Kerr, 2015). It is also possible that input sharing reduces the cost of realising ideas, particularly for markets with dense networks of input suppliers (Helsley and Strange, 2002). Carlino and Kerr (2015) suggest that spillovers likely occur through professional and social networks.

However, I argue that it may be possible to disentangle effects associated with peer interactions and effects associated with the increased concentration of authors within a city with detailed panel data. Thus, the aim of this chapter is to understand the dynamics of the peer interactions that may underlie agglomeration gains for prominent authors.

3.2.2 Social Interaction and Peer Effects

Social interaction is a key feature of theories of both urbanisation and localisation. Durlauf and Ioannides (2010) define social interactions as "interdependencies among individuals in which the preferences, beliefs, and constraints faced by one person are directly influenced by the characteristics and choices of others" (p 452). These interactions may impact behaviour and outcomes through a variety of mechanisms – e.g. facilitating flow of information, influencing tastes and preferences, or providing risk-sharing devices (Topa and Zenou, 2015).

[1] Also see Audretsch and Feldman (2004); Helsley and Strange (2002); Rosenthal and Strange (2004); Duranton and Puga (2004); among others for further discussion of the 'localisation versus urbanisation' debate

Social interaction effects – or peer effects – can be broadly grouped into three categories: endogenous, exogenous (or contextual) and correlated effects. *Endogenous* peer effects reflect the tendency of groups to conform: if authors within a city are highly productive, then an individual author may feel more pressure to be highly productive to match the average productivity of her peers. In this way, each individual's expectation of the group outcome further contributes to the actual group outcome. *Exogenous* effects reflect the tendency for individuals to self-select into groups of similar individuals and thus have similar outcomes. *Correlated* effects capture the tendency of individuals to act similarly because they face the same environment. These effects could include culturally specific behaviours associated with a particular location or technological limitations associated with a particular time period. (See Manski, 1993).

Manski (1993) argues that only the *endogenous* effect represents a 'true' peer effect, as this effect operates as a social multiplier while *exogenous* and *correlated* effects do not. Through the *endogenous* effect, individuals adjust their own behaviour to match average peer behaviour. This, in turn, changes the group average, creating a cyclical relationship in which behavioural changes are aggregated – hence the term 'social multiplier.' Glaeser et al. (2002) describe the social multiplier as the "ratio of aggregate coefficients to individual coefficients" and state that "if an individual's outcome rises x percent as his neighbour's average outcome, then the social multiplier roughly equals $\frac{1}{1-x}$ for large enough groups." Thus, social multipliers tend not to be very large, unless the change in individual behaviour is greater than 33%.

Estimation of this social multiplier is complicated by sorting on unobservables or social interactions. If individuals sort on positive interactions, the estimated social multiplier will be an overestimate of the true effect and vice versa for sorting on negative interactions (Glaeser et al., 2002). Bramoullé et al. (2009) also note that a key assumption of the linear-in-means models and many peer effects models is that "individuals *interact in groups*. This means that the population is partitioned in groups, and that individuals are affected by all other in their group and by none outside of it" (p 42). (Also see Durlauf and Ioannides,

2010.) However, they mention that this is not likely to reflect typical interactions or relationships between individuals.

Yet, empirical evidence of peer effects has been found in a broad range of economic literature. Peers can affect health outcomes such as mental health, fertility and sexual activity, drug and alcohol use, etc.[2] Peers can also have an impact on human capital accumulation, both in terms of schooling (e.g. graduation rates, drop-out rates, years of schooling, test scores, and college attendance/post-secondary education)[3] and in terms of labour force participation and performance (e.g. hours worked, labour force participation, income/wages).[4]

There has also been a recent growth in interest in peer effects in creative industries. Azoulay et al. (2010) find evidence of positive spillovers between academic 'superstars' and their collaborators. Similarly, Kim et al. (2009) find positive spillovers between colleagues in economics and finance departments in elite universities. Agrawal et al. (2014) find that the department-level output of academic scientists increases after the hire of a 'star' and that much of the output gains are due to increase in department quality and not due to increase in department size. Although, they also find that hiring a 'star' results in an increase in productivity of related incumbents and a decrease in productivity of unrelated incumbents.

Waldinger (2010) and Waldinger (2012b) both utilise the dismissal of Jewish academics in Germany during World War II as an exogenous shock to investigate peer effects. Waldinger (2010) finds faculty quality has a positive effect on PhD student outcomes, while Waldinger (2012b) finds evidence of positive peer effects for co-authors. Hellmanzik (2016) also uses World War II as an exogenous shock to estimate the effect of peers on artistic production. She finds that artists who live and work dense high-quality peer group (i.e. in Paris) produce systematically higher priced paintings.

[2]See Aizer and Currie (2004), Turley (2003), Hogan and Kitagawa (1985), Plotnick and Hoffman (1999), Rivkin (2001), Ginther et al. (2000), Case and Katz (1991), Brewster (1994), Crane (1991), among others.

[3]See, for example Duflo et al. (2011), Sacerdote (2001), Carrell et al. (2009), Rivkin (2001), Ginther et al. (2000), Imberman et al. (2012), Zimmerman (2003), Plotnick and Hoffman (1999).

[4]See Ioannides and Datcher Loury (2004), Topa (2001), Corcoran et al. (1992), Case and Katz (1991), Datcher (1982), Plotnick and Hoffman (1999), Rivkin (2001), Weinberg et al. (2002), etc.

However, empirical evidence of peer effects is quite mixed. While some studies find evidence of peer spillovers, others find no evidence in similar contexts. For example, Waldinger (2012b) finds no evidence of peer effects between academics within departments but does find evidence of peer effects among co-authors, while Waldinger (2010) find that average faculty quality does have an impact on the outcomes of PhD students. Cohen-Cole and Fletcher (2008), on the other hand, find implausible peer effects within friendships: strong correlations in height, acne, and headaches.

From both theory and empirical evidence, it is not clear in which areas or fields peer effects should exist, what factors should matter in a given circumstance, the 'correct' level of aggregation in which peer effects may be identified, or what kinds of relationships in which peer effects are most likely to exist. Azoulay et al. (2010) explore peer effects in contemporary academic science co-author networks and conclude that spillovers are more likely to occur between individuals in 'idea space' than in 'physical space' or 'social space.' They suggest that physical proximity and social interaction are not sufficient conditions for knowledge spillovers between academic scientists and that spillovers can only occur between academics who work within the same boundaries within a scientific field.

In contrast, Agrawal et al. (2008) find that both spatial and social proximity increase the probability of knowledge flows between inventors. They also find that the marginal benefit of spatial proximity is greater for inventors who are not socially close, implying that spatial and social proximity are substitutes in terms of access to knowledge. However, they find greater marginal benefits for inventors working in the same technical area, which provides further evidence of 'idea space' as optimal for knowledge spillovers.

While the hypothesis that spillovers primarily occur within 'idea space' is likely to hold for highly specialised fields requiring a large knowledge base and technical jargon such as STEM fields or economics, it is not clear whether the same restrictions should hold for other creative and cultural industries. For example, authors of literature are, in many ways, far less specialized than academics. Is it reasonable to assume that the intellectual and knowledge boundaries between one Victorian novelist and another are so great that

spillovers will not exist in 'physical space' or 'social space'? It is likely that the 'idea space' condition is met by authors writing in the same city at the same time.

In this chapter, I explore the hypothesis that physical proximity is a sufficient condition for peer effects to exist between authors. In doing so, I contribute to literature on peer effects by exploring an industry that is relatively unspecialised and does not require as large a knowledge base compared to academic research in STEM fields. I also contribute to agglomeration literature by exploring the 'black box' of agglomeration effects for authors by determining if the city is an appropriate level of agglomeration for author peer effects to occur and exploring several mechanisms that may underlie the agglomeration gains observed in Chapter 2.

3.3 The Data

This chapter utilises a unique dataset combining location, biographical and peer data of 370 prominent literary artists born between 1700–1925 and associated with British or Celtic literature constructed for the purposes of this study.[5,6] These data include age, lifespan, age at first publication, number of publications per annum, lifetime publications, career duration, gender, author quality, and location for every year of the author's life. As in Chapter 2, I limit the study to potential active years, which I define as age 16 and older as this represents the lower bound for age at first publication. The sample used in the analysis begins in 1725 (the first year with multiple observations) and ends in 1999 (when the last authors died).

Author quality is defined as the amount of contemporary critical attention in academic literary studies given to one author relative to her peers. The amount of critical attention

[5]Celtic literature includes literature associated with Celtic nations within the British Isles (Ireland, Scotland, Isle of Man, Wales, and Cornwall) and does not strictly refer to literature written in the Celtic languages.

[6]Authors are defined as individuals who made at least one unique contribution to poetry or prose, which eliminated individuals whose contributions were strictly limited to translations, textbooks, manuals or guides, song-writing, literary criticism, or publishing.

has two components: an indexed measure of the total words designated to an author from the three biographical sources and an indexed measure of the total number of citations within literary criticism as listed in the Literature Online (2014) database. These measures carry equal weight in an overall index of author quality.[7]

The author who receives the most contemporary critical attention is given a measure of 100. An author with a quality measure of 50 receives 50% of the critical attention that the highest-ranked author receives. As seen in Figure 3.1, there is a sharp decline in the amount of attention given to the second and third ranked authors relative to the first. Most authors have a quality score of less than 20, and just over 20% of authors have a quality score of between 5 and 10. The mean quality score is 14.5.

Summary statistics are presented in Table 3.1. On average, authors lived for 66 years, published their first work at age 27, and worked for just under 40 years. Over the course of their careers, authors produced an average of 33 publications, publishing an average of 0.88 works per year. Output is defined as the number of known works published in a given year.[8]

Peers are defined as authors who live in the same city in the same year. Due to the nature of the data, it is not possible to define peers on a more localised basis.[9] Thus, peer effects in this context capture city-level interactions and spillovers between authors. Peer effects are intrinsically linked to the spatial distribution and geographic clustering of authors. Table 4.1 also provides the top 10 author destinations in terms of author-year observations. The most common destination is London, and authors spend an average 21 years in London over the course of their lifetimes. However, this does not effectively describe the geographic clustering of authors. For example, the 100 author-year observations for Manchester could be 20 authors spending 5 years each in Manchester or 4 authors spending 25 years each in

[7] A more detailed description of the data collection process and the creation of the Impact Index may be found in Chapter 2.

[8] Publications include any type of unique literary publication: novel, collection of poetry or short stories, anthologies, contributions to literary magazines, plays, memoirs, etc. Individuals who only published non-fiction works (textbooks, dictionaries, journalism, etc.) and/or translations are not included. All publications were given equal weight. There is no measure of the quality of individual publications.

[9] Historical biographies most commonly state that an author moved to a city or lived in a city in a given year without mentioning the specific neighbourhood.

Manchester. It also does not provide any information about co-location: authors living in the same location at the same time.

Figure 3.2 shows the share of authors in the top 5 cities of co-location. London is home to the highest share of authors for the entire period. Over 30% of all authors in a given year reside in London until the end of the 19th century, and over 50% of authors lived in London in the late 1700s and again in the late 1800s. Only Dublin, Edinburgh, Oxford, and Cambridge have consistent co-location of authors throughout the sample, although there tend to be less than 5 authors in each of these cities in any given year. The Herfindahl-Hirschman Index (HHI) of author concentration over the entire time period is 0.177, indicating a moderate degree of concentration.[10]

As seen in Figure 3.3, there is a substantial number of author-year observations in which authors have no peers. Figure 3.4 shows the distribution of the number of peers by author-year observations, excluding observations with no peers. There is a substantial number of author-year observations in which authors have 5 or fewer peers, including over 900 author-year observations in which authors only have one peer. This reflects the consistent co-location of authors in Dublin, Edinburgh, Oxford and Cambridge, as well as other cities such as Manchester, Bath and Paris that saw the co-location of very few authors for short periods. The variation in number of peers above this level likely reflects the variation in the number of people within London over the time period.

Table 3.1 includes summary statistics for peer variables. Authors have an average of 11 peers per year within their lifetimes. Mean peer quality per annum is 1.7, and mean peer output per annum is 0.06. The low mean peer output per annum reflects the many author-year observations in which no works are produced. All peer means are defined as leave-out means (i.e. they exclude own-author values.) For authors who visit more than one location in a single year, peer means are calculated using a weighted average with the number of peers in a location as the weight.

[10]The HHI is defined as $HHI = \sum_{i=1}^{N} s_i^2$ where s_i is the share of author-year observations of location i and N is the number of locations, and it can provide a more formal measure of author concentration.

3.4 Identification of Peer Effects

Studies on peer and social effects often utilise a linear-in-means model to empirically estimate the role of various social interactions in determining an individual's outcome. The standard linear-in-means model can be expressed as:

$$Y_{ic} = \alpha + \beta X_{ic} + \lambda \bar{Y}_c + \delta \bar{X}_c + \upsilon_{ic} \tag{3.1}$$

in which some outcome Y of individual i in location c is a function of personal characteristics X_{ic}, mean outcome $\bar{Y}_c$, and other mean peer characteristics $\bar{X}_c$. *Endogenous* peer effects are given by $\bar{Y}_c$, and *exogenous* peer effects are given by $\bar{X}_c$. Manski (1993) also warns that identification is further complicated by the presence of *correlated* effects – i.e. the tendency of individuals to act similarly because they face the same environment.

Manski refers to this as the *reflection problem.* As Bramoullé et al. (2009) note, "simultaneity in behaviour of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics" (p 41), making it difficult to separately identify the endogenous effect from the exogenous and correlated effects.[11] In the context of this chapter, does an author produce more because she has highly productive peers – or do her peers produce more because she is more productive? Do authors self-select into groups of similar individuals and thus display similar outcomes purely because they share similar characteristics? Is author productivity simply a reflection of local environment – location, time, literary trends, technological constraints, etc.?

The use of leave-out means (the group mean excluding one's own-observation) partially alleviates the *endogenous* reverse-causality issue but does not eliminate it. For example, the individual author plays a diminishing role in determining other authors' productivity as the size of the local population grows; however, the productivity of peers is likely to

[11] See Manski (2000), Bramoullé et al. (2009), Angrist (2014), Durlauf (2004), Durlauf and Ioannides (2010), Lee (2007), Topa and Zenou (2015), and Davezies et al. (2009) for further discussion of the identification of peer and neighbourhood effects.

increase as the local population grows. (See Bramoullé et al., 2009.) The use of lags will also address this issue of reverse causality: past peer output impacts current own-output, while current own-output cannot impact past peer output. Therefore, I utilise leave-out means for all peer variables and one-year lags for mean peer output. I control for correlated effects by using the combination of city and time fixed effects.

The validity of this approach hinges on the absence of serial correlation. Serial correlation would occur if publication in one year was likely to lead to publication in the following years. Serial correlation could arise because authors strategically time publication to have a smoother flow of income or because a publishing house is more likely to publish an author's work if they have already published previous works. However, authors would also need to have a steady stream of ideas and continued commercial success. It is possible that this is true for some authors while others produced at irregular intervals.[12]

Authors are likely to self-select into location and self-select into groups (although in this context these two self-selection issues are identical, as peer group is defined as location). (See Durlauf, 2004.) I utilise author fixed effects in order to control for individual self-selection that does not vary over time. The use of time and location fixed effects also mitigates the issue of selection associated with a fixed city characteristic (i.e. London being the most populous city or universities existing in Cambridge and Oxford) or a fixed time characteristic (i.e. the London Blitz in 1940 and 1941). I use 2-year time dummies to control for changes over time while also avoiding the age-cohort-time problem that will arise with the linear combination of age, year dummies and individual fixed effects.

In sum, I will estimate the following equation:

[12] A contemporary example of this is George R.R. Martin's struggle to complete the sixth book in the *A Song of Fire and Ice* series despite commercial success and pressure from publishers and fans.

$$Output_{ict} = \alpha + \boldsymbol{\beta}\mathbf{X}_{ict} + \lambda\left(\frac{1}{n_{ct-1}}\sum_{j=1,j\neq i}^{n_{ct-1}} Output_{jct-1}\right) + \delta_1\left(\frac{1}{n_{ct}}\sum_{j=1,j\neq i}^{n_{ct}} PeerQuality_{jct}\right)$$
$$+ \delta_2\left(\frac{1}{n_{ct}}\sum_{j=1,j\neq i}^{n_{ct}} PeerQuality_{jct}^2\right) + \delta_3 NumberPeers_{jct}$$
$$+ \delta_4 NumberPeers_{jct}^2 + \mu_i + \sigma_c + \gamma_t + \epsilon_{ict} \tag{3.2}$$

in which output of individual i in city c at time t is a function of a vector of observable individual characteristics, mean peer output in the previous period, and a vector of other mean peer characteristics where n_{ct} is the number of authors in city c at time t, excluding author i (i.e. all peer averages are defined as a leave-out means). Under this specification, λ captures the *endogenous effect*, δ_1 to δ_4 capture the *exogenous* (or *contextual*) *effects*, and the combination of σ_c and γ_t captures the *correlated effects* associated with time and location.

Regressors include own-age, one-year lagged peer output, average quality of peers and peer group size (i.e. number of peers). I allow for own-age, peer group size, and quality of peers to have non-linear effects by including quadratic terms. Mean peer output is defined as the mean peer output in the previous period, unless specified otherwise. If an author resided in more than one location in a given year, the peer means for that year are calculated using a weighted average with the number of peers as the weight, unless specified otherwise.

I expect age to have an increasing but diminishing effect on productivity, reflecting standard ageing effects. Peer group size is likely to have an increasing but diminishing effect on productivity, reflecting the increased generation and sharing of ideas and increasing competition for resources. Mean peer quality is likely have a positive but diminishing effect. High quality peers likely inspire authors and promote greater creativity, increasing the generation and diffusion of new ideas. However, it is also possible that peer quality may increase quality of output while not impacting the quantity of output, which is a limitation

of the measure of productivity. The potential effect of mean peer output is ambiguous. Authors could produce more to stay relevant (positive effect), authors compete for limited resources (negative effect), or authors could simply aim to conform to the productivity of their peers (positive or negative effect – i.e. when authors change locations, the effect depends on mean peer productivity of origin and destination cities.)

I also consider several alternative specifications as robustness checks. I repeat the estimation with a specification in which all peer variables are lagged by one year. I consider the possibility that mean peer output-to-date may be more important than mean peer output in the previous year. I also investigate whether the use of unweighted averages rather than weighted averages for peer variables has any significant impact on the results. I also determine whether simply having peers (binary indicator) or whether categories of peer group size better capture the effect of number of peers. I explore peer effects within London, the only location with agglomeration effects as determined in Chapter 2.

I also consider the possibility that certain types of peers may matter more. I investigate the effect of younger versus older peers: younger peers may represent an up-and-coming generation and increased competition, while older peers may be mentors, the status-quo, or the 'giants' on whose reputation the author must build. I consider the possibility that living in the same city as a 'superstar' may matter more than average peer quality. Finally, I also look at effects by heterogeneity of own-quality, as own-quality may impact ability to form useful connections. As Combes and Gobillon (2015) note, "some evidence suggests for instance that more productive workers are also the ones more able to reap the benefits from agglomeration" (p 261). (Also see Glaeser and Maré, 2001 and de la Roca and Puga, 2012.)

3.5 Results

3.5.1 General Peer Effects

The coefficients for the estimates of Equation 3.2 are presented Table 3.2. Columns (1)-(3) introduce each peer variable separately. Column (1) shows the coefficients for a standard age-productivity profile with peer quality only. The coefficients for a standard age-productivity profile with peer group size (i.e. number of peers) and with peer output are reported in Columns (2) and (3), respectively. Column (4) displays the coefficients for a specification with all three peer variables. In all four models, robust standard errors are clustered on the individual level to account for serial correlation within individuals.

When included separately in the age-productivity profile, peer quality and peer group size have a statistically significant impact on productivity. As seen in Column (1), peer quality has a positive and significant effect on productivity. There is no evidence to support the hypothesized positive but diminishing effect of peer quality on productivity. Instead, the relationship between peer quality and own-output appears to be linear. The magnitude of the effect is modest: an increase in peer quality by one standard deviation is associated with an increase of 0.015 additional works per year or just less than a 0.1% increase in output per annum. Peer group size has a negative and significant effect on productivity, as seen in Column (2). The quadratic term is also not significant, again suggesting the hypothesized effect was incorrect: the relationship between peer group size and own-output appears to be linear rather than quadratic. The magnitude of the effect is also modest: one additional peer reduces productivity by 0.017 per year or 10 additional peers reduces productivity by 1.7 works over 10 years. Peer output in the previous year does not have a statistically significant effect, as seen in Column (3).

When all peer variables are considered simultaneously, only peer group size has a significant effect. However, the magnitude of the effect is still quite modest: one additional peer within the life of an author reduces annual productivity by 0.0175 per year or 10 additional

peers reduces productivity by 1.75 works over 10 years. In the absence of random group assignment, I interpret these results cautiously, as there could be some remaining self-selection or omitted variable bias. Consistent with the findings of Chapter 2, age has a positive but diminishing effect in all four specifications. This is also consistent with general findings in the literature on individual productivity. (For example, see Levin and Stephan, 1991.) Overall, there is not much evidence supporting the existence of peer or social interaction effects within author groups at the city-level.

Specifications with Alternate Variable Definitions

I consider several additional specifications to determine if the results are being driven by the way in which the variables are defined. I allow for all peer variables to be lagged rather than only peer output. I also investigate whether mean peer output-to-date provides greater explanatory power than mean peer output in the previous year. Output-to-date is defined as the total number of works published by an author to date. Finally, I determine whether there are significant differences between using unweighted and weighted averages to calculate mean peer variables when authors live in more than one location in a given year.

The coefficients for these specifications with alternate variable definitions are presented in Table 3.3. Column (1) displays the results from Column (4) in Table 3.2 for ease of comparison. Column (2) includes the coefficients for a regression in which all peer variables are lagged. Column (3) shows the results when using mean peer output-to-date instead of mean peer output in the previous year. Column (4) includes the coefficients when using normal averages rather than weighted averages for peer variables.

When all peer variables are lagged, none have a statistically significant effect on productivity. When mean peer output-to-date is used instead of mean peer output in the previous year, peer group size has a negative and significant effect, while the magnitude of the effect is slightly smaller than that of peer output per year displayed (-0.0155 compared to

-0.0175). Peer group size has a negative and increasing impact on productivity when using unweighted averages rather than weighted averages for peer variables. The magnitude of the effect is only slightly greater than the original specification, and while this effect is increasing, it is increasing at a low rate.

Given the modest results in the original specification, the results with alternate variable definitions are fairly consistent in terms of both the magnitude and significance of the coefficients for corresponding regressors. In this respect, the results are not being driven by specific definitions.

Variations in Peer Group Size Effects

Given that much of the variation in peer group size is likely due to variation in author concentration within London over time, then it may be useful to investigate whether authors may benefit from simply having peers. I explore this by using a binary indicator of whether an author has peers. Authors may also receive different intensity of gains from different sized groups that are not captured by looking at one unit variations from the mean. Utilising categorical variables for group size with zero peers and positive numbers of peers in increments of, for example, 10 (i.e. no peers, having 1-10 peers, having 11-20 peers, etc.) may be a more effective way of detecting peer group size effects.

The results for estimations with variations in peer group size effects are presented in Table 3.4. Again, the results from Column (4) in Table 3.2 are presented in Column (1) for ease of comparison. Column (2) shows the coefficients for the regression with a binary indicator of whether an author has peers rather than a variable for peer groups size, and Column (3) presents the results with peer group size categories. Neither alternate method of capturing peer group size effects has a statistically significant effect on author productivity.

3.5.2 Peer Effects within London

In the chapter thus far, there has been weak evidence for general peer effects. There is only a small reduction in annual productivity associated with peer group size within (not between)

authors. Merely having peers is not linked to productivity. However, I return again to the observation that much of the variation in peer group size is likely due to variation in author concentration within London over time. Because very few authors co-locate in non-London cities, it is possible that peer effects may only occur within London. Chapter 2 found that agglomeration gains were London-specific, making this an interesting exercise because it provides a good context in which peer effects might be more easily observed and it allows for the examination of mechanisms through which the London-specific agglomeration gains might occur.

I investigate peer effects within London by interacting the peer regressors (peer quality, peer output, and peer group size) with London. These results are presented in Table 3.5. Column (1) displays the results from Column (4) Table 3.2 for ease of comparison. In this case, the coefficient for London is reported. It is notable that both the direction and magnitude of the London fixed effect differ from results found in Chapter 2. The within estimator reveals that during years in which an author lives in London, productivity is reduced by 1.19 works per annum. This contrasts with the findings of Chapter 2, in which an author's productivity increases by 0.09 works per year during the years in which the author lives in London. Column (2) presents estimates that include the peer-London interaction terms. The magnitude of the London effect is smaller but the direction is still negative. While size of peer groups does not have a statistically significant effect on productivity, peer quality in London does. A one unit increase in peer quality in London reduces productivity by 0.129 works per year, and productivity is reduced at an increasing rate.

Peer quality was not linked to productivity in more general specifications and is only influential when authors live in London and variation in concentration of authors within London is controlled for. There is no similar effect for authors living in cities other than London. This could be due to a competition effect. During much of the time period, the time and resources of publishers were extremely constrained due to technological limitations. To a large degree, publication was a zero-sum game: if a printer and publisher are

nearing capacity from producing other authors' work, then an author's own ability to get work published is reduced.

Perhaps publishers are able to identify high-quality authors and give them preferential treatment during the publishing process. It is also possible that high-quality authors are better at making high-quality social connections or better at navigating the publishing process, even if publishers could not so easily determine author quality. As the quality of peers increases, own-ability (or lack thereof) is more apparent. The reduction in productivity may reflect the advantage higher-quality peers receive and thus the disadvantaged position of the author herself. It would be useful to investigate the role of own-quality to better understand the dynamics of author quality that underlie this result.

There is also likely to be some trade-off between *quantity* of works and *quality* of works. For example, the quantity of works may decrease as an author learns to write higher quality works by living and working near higher-quality peers, particularly if the quality-quantity trade-off is quite high. Thus, it is also possible this decrease in productivity is due to learning from peers reflected in the *quality* of work rather than the *quantity* of work; however, I am unable to confirm this mechanism due to the way output is defined. The definition of output is an important limitation as mentioned previously in Chapter 2, and future studies on the productivity of creative workers should consider the potential importance of quality of output in their design.

Effect of Peer Age

I also consider the possibility that certain types of peers may have a greater influence on productivity. Younger peers (defined as own-age and younger) may represent an up-and-coming generation of authors, competitors and innovators, who build and innovate upon the stylistic standards developed by their predecessors. Older peers, on the other hand, may represent the status-quo, individuals of known quality and dependability with well-established social connections, who may also serve as mentors. The zero-sum nature

of publication could lead competition effects to dominate, with both older and younger peers reducing own-productivity. However, one or both groups may also promote greater exchange of ideas and thus increase productivity. Both effects could work through the peer quality channel or through direction competition via output.

I consider the effect of younger peers and older peers separately before estimating the effect of both peer types simultaneously. I also include London interactions, as I previously confirmed the existence of unique, London-based peer effects. These results are presented in Table 3.6. Column (1) includes the coefficients when only younger peers are included, and Column (2) includes the respective coefficients when only older peers are included. Column (3) reports the results for both older and younger peers. Column (4) adds peer-London interaction terms.

As seen in Column (1), neither the quality nor the output of younger peers has a statistically significant impact on own-productivity. Similar insignificant results are found when only including older peers (Column (2)) and when estimating younger and older peers simultaneously (Column (3)). The coefficients for peer group size and the London fixed effect for all three results are consistent with those found previously in Table 3.5. This result of no general peer effects is not unexpected, given that little evidence of general peer effects was found in previous specifications.

However, there is some evidence of peer influence when peer-London interactions are included (see Column (4)). As found in the previous section, peer group size is not significant and the London fixed effect is negative. The quality of older peers in London has a negative effect on productivity, reducing per annum productivity by 0.05 works. Older peers are more likely to have established reputations in terms of commercial success, quality, and dependability and may also have well-established social connections. As mentioned in the previous section, higher quality peers may also be given preferential treatment by publishers or may simply be better able to form advantageous social connections. On the other hand, by the quality-quantity trade-off hypothesis, learning from older peers (who

may serve as mentors or individuals who are already knowledgeable about the literary industry) may result in a reduction in productivity. In both cases, it is reasonable that older peers would have a negative effect on own-productivity through a peer quality mechanism.

Effect of Living in a City with a 'Superstar'

There is some evidence that the upper tail is more important than the average peer for innovation, knowledge spillovers, and human capital development. (See, for example, Rosen, 1981; Azoulay et al., 2010; Waldinger, 2010; and Squicciarini and Voigtländer, 2015.) I also consider the possibility that sharing the same location as a *superstar* may have a greater impact on productivity than average peer quality. I investigate three definitions of a *superstar* (i.e. someone in the upper tail of the quality distribution) to test the sensitivity of the definition. I define a *superstar* as (i) an author in the top 10%, (ii) in the top 25%, and (iii) in the top 50%. I also investigate general peer effects and peer effects within London.

The results of these estimations may be found in Table 3.7. Columns (1)-(3) include the results for general peer effects, while Column (4)-(6) include the results with peer-London interactions. Columns (1) and (4) show the coefficients for specifications in which a *superstar* is defined as an author in the top 10%. Columns (2) and (5) show the respective coefficients when a *superstar* is defined as an author in the top 25%. Column (3) and (6) display the respective coefficients when a *superstar* is defined as an author in the top 50%. In short, living in the same city as a superstar has no impact on productivity, regardless of how a superstar is defined.

This result highlights a key limitation of how peers are defined. In this study, peers are defined as authors living in the city in the same year. Due to the nature of biographical entries for historical figures, more localised data is largely unavailable. There is an implicit assumption that geographic proximity on a city level can serve as a proxy for unobservable social interactions. While many empirical studies on agglomeration utilise city-level data, it is likely that this level of aggregation is not the optimal level to study peer or interaction

effects. It is possible that *superstar* effects could be observed when using more detailed data on social interaction or social connections – known collaborators, friendship and kinship networks, attending the same university, etc. Similarly, it may simply be easier to detect competition effects related to peer quality at the city-level since publication itself is a zero-sum game.

Heterogeneity of Returns by Author Quality

Given the previous results suggesting that own-quality may play a role in productivity and in an author's ability to realise productivity gains through peer interactions, it may be valuable to investigate the heterogeneity of returns by author quality. Author quality may be an (imperfect) proxy for social capital, and Glaeser et al. (2002) note that a person's social capital or "a person's social characteristics – including social skills, charisma, and the size of his Rolodex...enables him to reap market and non-market returns from interactions with others" (p F438).

There is a growing body of empirical evidence of quality-related returns. For example, Rosenthal and Strange (2008) find sharper rates of attenuation of human capital spillovers for workers with a university education compared to those with less than a university education. Waldinger (2010) finds the greatest gains for PhD students at the top 10 universities, and similarly Borowiecki (2013) finds the top 10 composers receive the highest productivity gains from geographic clustering of Western classical composers.

In order to preserve identification via author fixed effects and investigate the impact of a fixed characteristic (quality), I divide the sample into quartiles based on own-quality and re-estimate an age-productivity profile with peer variables and peer-London interactions for each quartile. I also look at heterogeneity within the upper tail by estimating a similar regression for authors in top 10% in terms of quality. Due to the small subsample sizes, these results will provide suggestive, although not conclusive, evidence on the role of author quality in realising productivity gains from peer interactions.

The results from these regressions are presented in Table 3.8. Column (1) displays the coefficients for authors in the bottom quartile. Column (2) includes results for authors in the 2nd quartile. Coefficients for authors in the third quartile are presented in Column (3), and coefficients for authors in the top quartile are located in Column (4). Estimates for authors in the top 10% are presented in Column (5).

The most notable result is the heterogeneity between quartiles in terms of which peer variables influence productivity. Authors in the lowest quartile receive a substantial gain from the London fixed effect but have no statistically significant peer interaction effects, while authors in the second quartile receive no statistically significant effect from either. Peer interaction effects only arise for authors in the top 50%, suggesting there may be a link between quality and social capital.

For authors in the third quartile, peer output in London has a modest but positive effect on productivity. Peer group size (non-London) reduces the productivity of authors in the top 10%, but this effect does not hold when the group is expanded to the entire top quartile. Peer quality in London has a negative effect on productivity for authors in the top two quartiles, as well as those in the top 10%. This effect reduces productivity at an increasing rate for all three groups of authors. The magnitude of this effect increases with author quality, ranging from -0.26 for authors in the third quartile to -0.30 for the top 10% of authors.

These results provide evidence that more detailed data on social capital will likely shed greater light on productivity than the broadly defined peer effects used in this chapter. Peer groups based on location rely on the assumption that proximity will lead to greater probability of interaction; however, it is clear that the level of aggregation needs to be refined.

3.6 Conclusion

There is modest evidence of general peer effects through the channel of peer group size; however, there is stronger evidence of highly-localised peer effects. Mean quality of peers

in London has a negative effect on annual productivity. These effects could be due to competition – publication is a zero-sum game and higher quality authors are likely to have better connections or be given preferential treatment by publishers. This quality effect is due to older peers who are more likely to have an established reputation or serve as mentors to younger authors. There are also substantial heterogeneity of returns by own-quality, and this deserves further exploration in another framework in which the quality component of the author fixed effects are retained.

In general, the peer variables studied in this chapter do little to shine light into the 'black box' of agglomeration effects. However, an important observation is that the level of agglomeration is critical, a view supported by Glaeser et al. (2002). These results could indicate that cities are not optimal level of agglomeration for detecting social interaction effects, and more localised location information should be collected when possible. This may present a major challenge for similar studies, as historical information – biographies for historical figures in particular – can be somewhat vague in discussing location.

It is likely that geographic proximity is a necessary but not sufficient condition for peer effects to occur between authors, in line with the Azoulay et al. (2010) observation that spillovers are more likely to occur in 'idea space' than in 'physical space'. However, it is unclear how an 'idea space' should be defined for authors – and if there is an observable difference between 'idea space' and 'social space' for the creative industries in general. For example, anecdotal evidence on collaborative groups, such as the "Lost Generation" of writers, artists and other creative workers in Paris in the 1920s suggests that the spillovers occurred not only across creative industries but also within social networks. "Indeed, there is increasing recognition among economists in general of the role played by *social network* in structuring interactions among agents" (p 42, Bramoullé et al., 2009). Future studies on author peer effects should either incorporate better proxies for peer interaction or data on the interactions themselves.

3.A Tables and Figures

3.A.1 Figures

Figure 3.1: Distribution of Author Quality

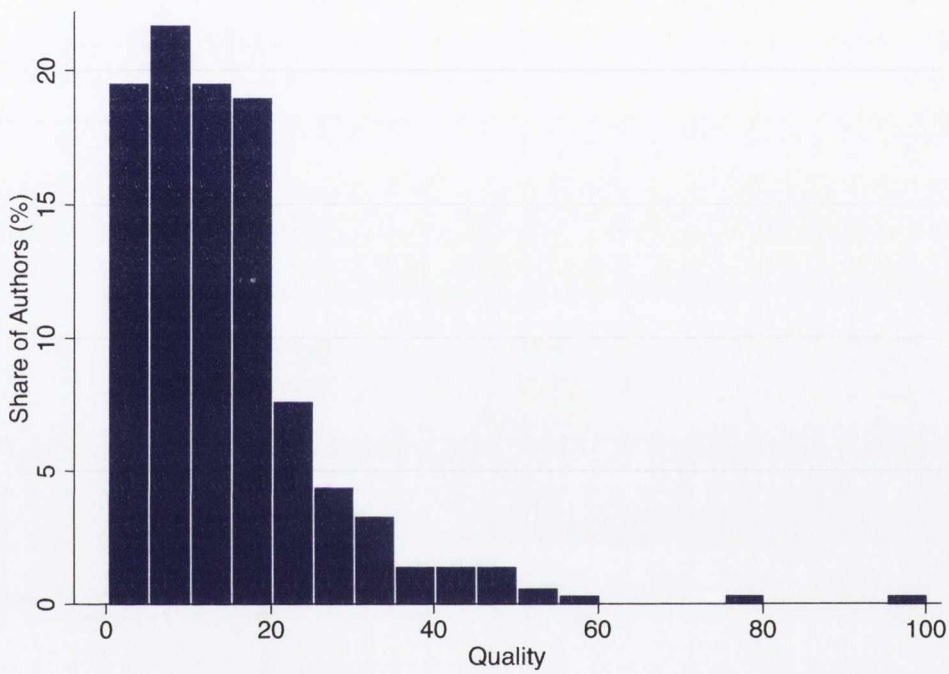


Figure 3.2: Share of Authors in Major Cities

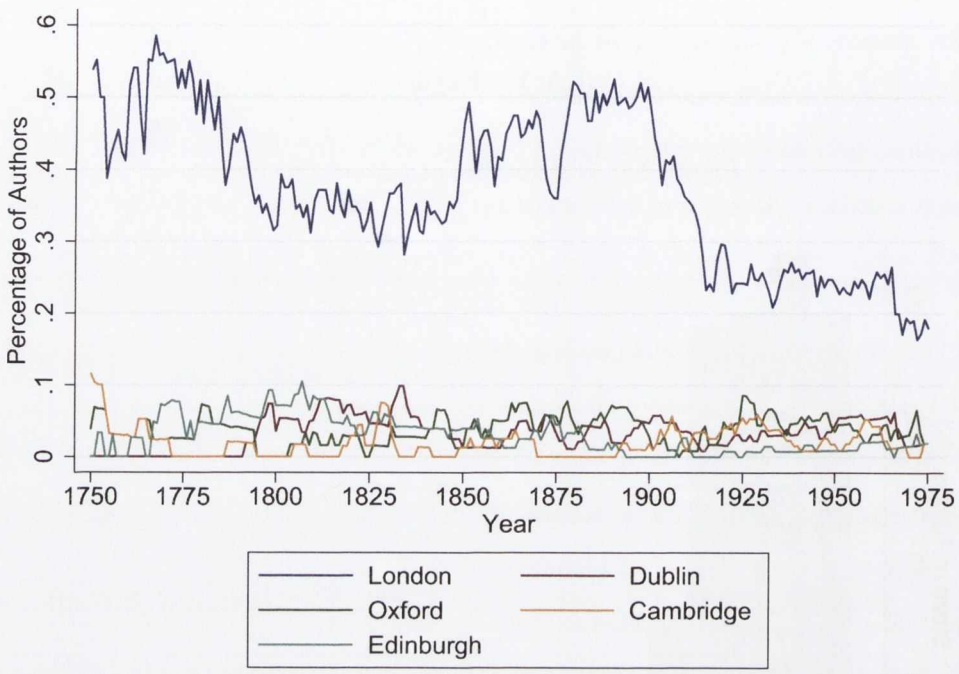


Figure 3.3: Distribution of Number of Peers

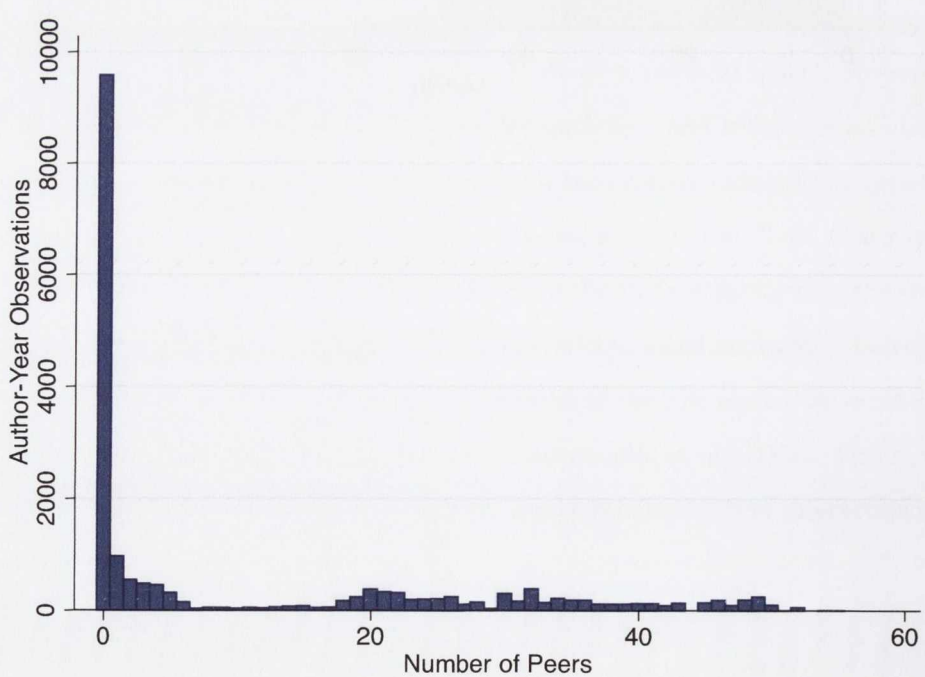
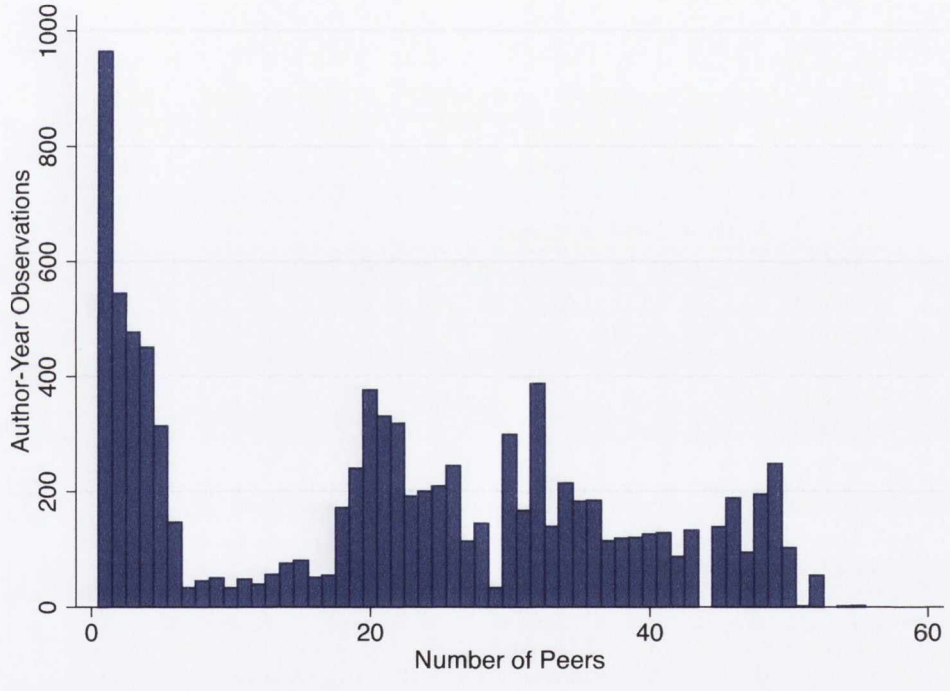


Figure 3.4: Distribution of Number of Peers (excluding obs. with zero peers)



3.A.2 Tables

Table 3.1: Summary Statistics

	Mean	St. Dev.	Min.	Max.
Individual Author Data				
Lifespan	66.45	16.28	18.00	97.00
Age at first publication	26.81	7.45	9.00	61.00
Total years in London	17.89	18.41	0.00	67.00
Total works over lifetime	33.21	32.23	1.00	186.00
Career length	39.64	17.24	1.00	81.00
Quality	14.48	11.93	0.23	100.00
Output per annum	0.87	0.76	0.02	5.67
Peer Data*				
No. of peers per annum	10.88	10.65	0.00	40.32
Mean peer quality per annum	1.70	2.34	0.00	16.01
Mean peer output per annum	0.06	0.09	0.00	0.65

Top 10 Author Destinations

City	*Author-Year Obs.*
London	6,620
Oxford	777
Dublin	651
Edinburgh	484
Cambridge	425
Paris	265
New York	125
Bath	105
Manchester	100
Norwich	84

*All peer means are defined as leave-out means (i.e. exclude own-author values.) For authors who visit multiple locations in a single year, peer means are calculated using a weighted average with the number of peers per location as the weight.

Table 3.2: Main Result

	(1) Peer Quality	(2) No. of Peers	(3) Peer Output	(4) Combined
Age	0.0765***	0.0771***	0.0773***	0.0779***
	(0.00494)	(0.00495)	(0.00515)	(0.00515)
Age^2	-0.000715***	-0.000722***	-0.000723***	-0.000729***
	(0.0000492)	(0.0000494)	(0.0000510)	(0.0000512)
Peer Quality	0.00633*			0.00562
	(0.00359)			(0.00356)
Peer $Quality^2$	-0.0000957			-0.0000849
	(0.0000648)			(0.0000668)
Number of Peers		-0.0172*		-0.0175*
		(0.00906)		(0.00910)
Number of $Peers^2$		0.000258		0.000267
		(0.000163)		(0.000164)
Peer Output			0.0167	0.00248
			(0.0287)	(0.0283)
Constant	0.406*	0.405*	0.387*	0.353
	(0.232)	(0.226)	(0.229)	(0.234)
Time Dummies	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes
R^2	0.149	0.150	0.141	0.142
No. Authors	370	370	370	370
Observations	19020	19020	18650	18650

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses.

Table 3.3: Minor Variations in Specification

	(1) Baseline	(2) All Lagged Peer Variables	(3) Peer Output-to-Date	(4) Unweighted Peer Variables
Age	0.0779***	0.0769***	0.0771***	0.0779***
	(0.00515)	(0.00516)	(0.00495)	(0.00516)
Age^2	-0.000729***	-0.000719***	-0.000722***	-0.000730***
	(0.0000512)	(0.0000511)	(0.0000494)	(0.0000512)
Peer Quality	0.00562	0.00519	0.00319	0.00171
	(0.00356)	(0.00331)	(0.00388)	(0.00379)
Peer $Quality^2$	-0.0000849	-0.0000943*	-0.0000633	-0.0000121
	(0.0000668)	(0.0000542)	(0.0000673)	(0.0000900)
Number of Peers	-0.0175*	0.00135	-0.0155*	-0.0190**
	(0.00910)	(0.00454)	(0.00897)	(0.00943)
Number of $Peers^2$	0.000267	0.00000525	0.000235	0.000285*
	(0.000164)	(0.000120)	(0.000161)	(0.000167)
Peer Output	0.00248	0.00372	0.00303	0.0130
	(0.0283)	(0.0309)	(0.00318)	(0.0242)
Constant	0.353	0.394*	0.377	0.362
	(0.234)	(0.229)	(0.243)	(0.231)
Time Dummies	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes
R^2	0.142	0.142	0.150	0.142
No. Authors	370	370	370	370
Observations	18650	18650	19020	18650

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Robust standard errors are clustered on the author level and are reported in parentheses.
Column (1) displays the same results as Column (4) in Table 3.2 for ease of comparison.

Table 3.4: Variations on Number of Peers

	(1) No. of Peers	(2) Author Has Peers	(3) Peer-Group Size Categories
Age	0.0779***	0.0774***	0.0776***
	(0.00515)	(0.00515)	(0.00516)
Age^2	-0.000729***	-0.000724***	-0.000726***
	(0.0000512)	(0.0000510)	(0.0000512)
Peer Quality	0.00562	0.00887	0.00811
	(0.00356)	(0.00581)	(0.00567)
Peer $Quality^2$	-0.0000849	-0.000129	-0.000117
	(0.0000668)	(0.0000835)	(0.0000826)
Peer Output	0.00248	0.00748	0.00674
	(0.0283)	(0.0291)	(0.0289)
Number of Peers	-0.0175*		
	(0.00910)		
Number of $Peers^2$	0.000267		
	(0.000164)		
Author has Peers		-0.0497	
		(0.0799)	
Author has 1-10 Peers			-0.0440
			(0.0777)
Author has 11-20 Peers			-0.111
			(0.168)
Author has 21-30 Peers			-0.165
			(0.158)
Author has 31-40 Peers			-0.103
			(0.159)
Author has 41-50 Peers			-0.109
			(0.178)
Author has over 50 Peers			-0.181
			(0.215)
Constant	0.353	0.363	0.361
	(0.234)	(0.235)	(0.235)
Time Dummies	Yes	Yes	Yes
Author FE	Yes	Yes	Yes
Location FE	Yes	Yes	Yes
R^2	0.142	0.142	0.142
No. Authors	370	370	370
Observations	18650	18650	18650

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses. Column (1) displays the same results as Column (4) in Table 3.2 for ease of comparison.

Table 3.5: City-Peer Effects in London

	(1) London FE	(2) London Interactions
Age	0.0779***	0.0776***
	(0.00515)	(0.00515)
Age^2	-0.000729***	-0.000727***
	(0.0000512)	(0.0000511)
Peer Quality	0.00562	0.00600
	(0.00356)	(0.00376)
Peer Quality^2	-0.0000849	-0.0000970
	(0.0000668)	(0.0000671)
Number of Peers	-0.0175*	-0.0111
	(0.00910)	(0.0188)
Number of Peers^2	0.000267	-0.00116**
	(0.000164)	(0.000584)
Peer Output	0.00248	-0.00329
	(0.0283)	(0.0284)
London	-1.189***	-0.896***
	(0.248)	(0.293)
Peer Quality in London		-0.129***
		(0.0459)
Peer Quality in London^2		0.00458***
		(0.00151)
No. of Peers in London		-0.0189
		(0.0234)
No. of Peers in London^2		0.00157**
		(0.000675)
Peer Output in London		0.299
		(0.196)
Constant	0.353	0.357
	(0.234)	(0.237)
Time Dummies	Yes	Yes
Author FE	Yes	Yes
Location FE	Yes	Yes
R^2	0.142	0.143
No. Authors	370	370
Observations	18650	18650

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses.

Column (1) above displays the same results as Column (4) in Table 3.2 for ease of comparison. The coefficient for London is reported above, while the results in Table 3.2 simply state that location fixed effects are included without reporting the coefficients.

Table 3.6: Effect of Younger Peers vs Older Peers

	(1) Younger Peers	(2) Older Peers	(3) Younger and Older Peers	(4) Young and Older Peers in London
Age	0.0778***	0.0779***	0.0778***	0.0778***
	(0.00516)	(0.00516)	(0.00516)	(0.00515)
Age^2	-0.000729***	-0.000729***	-0.000729***	-0.000730***
	(0.0000512)	(0.0000512)	(0.0000512)	(0.0000511)
Number of Peers	-0.0179*	-0.0183**	-0.0180**	-0.0112
	(0.00912)	(0.00916)	(0.00911)	(0.0189)
Number of Peers^2	0.000272*	0.000277*	0.000273*	-0.00107*
	(0.000164)	(0.000164)	(0.000164)	(0.000616)
Quality of Younger Peers	0.00197		0.00202	0.00192
	(0.00234)		(0.00233)	(0.00239)
Output of Younger Peers	0.00613		0.00580	-0.00442
	(0.0365)		(0.0365)	(0.0344)
Quality of Older Peers		0.000450	0.000738	0.00110
		(0.00298)	(0.00297)	(0.00306)
Output of Older Peers		-0.0212	-0.0213	-0.0257
		(0.0330)	(0.0331)	(0.0336)
London	-1.189***	-1.190***	-1.190***	-1.122***
	(0.243)	(0.244)	(0.244)	(0.268)
Quality of Younger Peers in London				-0.00218
				(0.0131)
Output of Younger Peers in London				0.336
				(0.215)
Quality of Older Peers in London				-0.0538**
				(0.0243)
Output of Older Peers in London				0.154
				(0.187)
No. of Peers in London				-0.00924
				(0.0228)
No. of Peers in London^2				0.00137**
				(0.000692)
Constant	0.366	0.373	0.368	0.374
	(0.229)	(0.230)	(0.230)	(0.232)
Time Dummies	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes
R^2	0.142	0.142	0.142	0.142
No. Authors	370	370	370	370
Observations	18650	18650	18650	18650

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses.

Table 3.7: Effect of Living in City with a "Superstar"

	(1) Peer in Top 10%	(2) Peer in Top 25%	(3) Peer in Top 50%	(4) Peer in Top 10% in London	(5) Peer in Top 25% in London	(6) Peer in Top 50% in London
Age	0.0779***	0.0779***	0.0778***	0.0779***	0.0779***	0.0779***
	(0.00515)	(0.00516)	(0.00516)	(0.00515)	(0.00515)	(0.00514)
Age^2	-0.000730***	-0.000730***	-0.000729***	-0.000731***	-0.000730***	-0.000730***
	(0.0000511)	(0.0000512)	(0.0000512)	(0.0000513)	(0.0000511)	(0.0000511)
Number of Peers	-0.0220**	-0.0186*	-0.0192**	-0.0214	-0.0146	-0.0170
	(0.0106)	(0.00951)	(0.00949)	(0.0217)	(0.0253)	(0.0258)
Number of Peers^2	0.000327*	0.000282*	0.000290*	-0.000818	-0.00116	-0.00119
	(0.000183)	(0.000169)	(0.000169)	(0.000650)	(0.000736)	(0.000813)
Peer Output	0.00635	0.00946	0.00839	-0.00186	0.00234	0.000860
	(0.0276)	(0.0278)	(0.0275)	(0.0274)	(0.0276)	(0.0274)
Peer in Top 10%	0.0605			0.106		
	(0.0570)			(0.0705)		
Peer in Top 25%		0.0142			0.0400	
		(0.0457)			(0.0652)	
Peer in Top 50%			0.0281			0.0540
			(0.0454)			(0.0637)
London	-1.174***	-1.188***	-1.207***	-1.202***	-0.881**	-0.720
	(0.244)	(0.244)	(0.245)	(0.251)	(0.348)	(0.525)
No. of Peers in London				0.00814	0.00137	0.00538
				(0.0253)	(0.0278)	(0.0281)
No. of Peers in London^2				0.00103	0.00137*	0.00138
				(0.000712)	(0.000784)	(0.000849)
Peer Output in London				0.206	0.173	0.208
				(0.190)	(0.203)	(0.180)
Peer in Top 10% in London				-0.157		
				(0.118)		
Peer in Top 25% in London					-0.408	
					(0.271)	
Peer in Top 50% in London						-0.639
						(0.493)
Constant	0.358	0.360	0.373	0.350	0.344	0.376
	(0.229)	(0.231)	(0.231)	(0.230)	(0.238)	(0.233)
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.142	0.142	0.142	0.142	0.142	0.142
No. Authors	370	370	370	370	370	370
Observations	18650	18650	18650	18650	18650	18650

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses.

Table 3.8: Heterogeneity of Returns by Own-Quality

	(1) Bottom Quartile	(2) 2nd Quartile	(3) 3rd Quartile	(4) Top Quartile	(5) Top 10%
Age	0.0459***	0.0855***	0.0739***	0.113***	0.0876***
	(0.00689)	(0.0121)	(0.0102)	(0.0113)	(0.0138)
Age^2	-0.000440***	-0.000827***	-0.000696***	-0.00103***	-0.000752***
	(0.0000696)	(0.000132)	(0.0000969)	(0.000111)	(0.000120)
Peer Quality	0.00386	-0.00469	0.00532	0.00480	0.0409*
	(0.00615)	(0.0106)	(0.0118)	(0.0114)	(0.0210)
Peer Quality^2	-0.0000570	-0.0000667	-0.0000219	0.000155	-0.00124
	(0.0000704)	(0.000275)	(0.000280)	(0.000309)	(0.000835)
No. of Peers	0.0546	-0.0453	0.121	-0.0170	-0.130*
	(0.0652)	(0.0871)	(0.0766)	(0.0334)	(0.0729)
Number of Peers^2	-0.0162	0.00349	-0.0112	-0.00103	0.00749
	(0.0121)	(0.00977)	(0.0125)	(0.00103)	(0.00824)
Peer Output	-0.00351	-0.0105	-0.00367	-0.0484	-0.0389
	(0.0500)	(0.0445)	(0.0449)	(0.0507)	(0.0808)
London	2.710***	0.0293	0.509	-0.0358	2.611***
	(0.871)	(0.290)	(0.431)	(0.562)	(0.628)
Peer Quality in London	-0.678	0.0408	-0.285***	-0.286*	-0.303***
	(0.480)	(0.0461)	(0.0998)	(0.147)	(0.0965)
Peer Quality in London^2	0.116	-0.00142	0.00905***	0.0156*	0.0125**
	(0.0756)	(0.00146)	(0.00285)	(0.00878)	(0.00525)
No. of Peers in London	-0.0974	0.0160	-0.128	-0.0510	-0.0237
	(0.0734)	(0.0901)	(0.0824)	(0.0462)	(0.0804)
No. of Peers in London^2	0.0168	-0.00295	0.0111	0.00196	-0.00545
	(0.0122)	(0.00980)	(0.0125)	(0.00126)	(0.00817)
Peer Output in London	0.465	0.313	1.012***	0.281	0.0603
	(0.962)	(0.233)	(0.347)	(0.171)	(0.112)
Constant	-2.225***	-0.991***	-0.960***	-0.369	-1.125***
	(0.175)	(0.265)	(0.274)	(0.352)	(0.399)
Time Dummies	Yes	Yes	Yes	Yes	Yes
Author FE	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes
R-squared	0.144	0.181	0.173	0.204	0.218
Number of author	91	94	93	92	37
Observations	4,592	4,773	4,645	4,638	1,782

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Robust standard errors are clustered on the author level and are reported in parentheses.

Chapter 4

The Predictive Power of Social Networks in Literary Production

4.1 Introduction

The circumstances in which creativity and innovation thrive must be understood in order to develop policy that encourages creativity and cultural production. For example, interactions between firms or individuals can increase innovative and creative activity through the knowledge and skills, learning, and the sharing of resources. Agglomeration and peer effects studies traditionally proxy social interaction with physical proximity under the assumption that greater proximity increases the probability of interaction; however, geographic proximity can be a poor proxy for social interaction, particularly as location data are aggregated.

Social networks provide more informative measures of interaction by describing the relational power an individual has in behaviours such as the exchange of knowledge and resources. Measures of social network connectivity and centrality describe the power each individual has to control over the flow of information and resources across the network and individual prestige (how an individual is positioned as a frame of reference for others). While social network information better describes interactions and relationships, the cost of collecting such data is often time and resource intensive, particularly when studying historical figures. Thus, the application of social networks in studies of creative and cultural production has been limited, as researchers opt for data that can be more efficiently collected (e.g. basic biographical information such as birth and death location, years of birth and death, total works produced, etc.)

In order to quantify the gains that may be realised through the use of social network information, I combine a new, unique dataset of the social connections of 370 prominent literary artists born between 1700–1925 and associated with British or Celtic literature with an existing dataset with location, biographical and peer data (in which peers are defined as authors living in the same city). Both datasets were constructed for the purpose of this study. I use a repeated random sub-sampling validation (Monte-Carlo cross-validation) procedure to determine the increase (if any) in predictive power of models containing

social network information beyond that contained in a standard age-productivity profile with peer effects. I also determine which of the network variables are informative.

I first consider the predictive power of social network information in terms of annual productivity. Models with social network information provide an increase in predictive power above and beyond that of an age-productivity profile, and this increase in predictive power is one to two orders of magnitude greater than the increase in predictive power of a model that includes a standard age-productivity profile and peer effects variables. Degree centrality (the number of social connections) provides the greatest predictive power.

I then consider two alternate measures of output: total works produced over an author's lifetime and author Impact Index score (a measure of contemporary critical attention). There is an improvement in predictive power with both measures of output. However, social network information has substantially more predictive power for the Impact Index score than lifetime works. Degree, closeness, and clustering have greater predictive power for measures of lifetime accomplishments than for annual productivity. This suggests that the way output is defined is of critical importance, and a failure to account for this may result in incomplete or incorrect conclusions regarding the effect of social network information on productivity.

I also discuss the author network and provide OLS estimates of the relationship between social network information and productivity in order to illustrate the ways in which social network connectivity and centrality may influence productivity and highlight the challenges associated with establishing a causal relationship when using social network data. This may provide some guiding intuition for future studies on authorship and other creative and cultural industries.

The remainder of this chapter is structured as follows. Section 4.2 outlines the theoretical framework by discussing the relationship between geographic proximity, social interaction, and creative production. Next, the data collection methodology, summary statistics, and the author network are described in Section 4.3. Section 4.4 outlines the econometric approach, and results are presented in Section 4.5. Section 4.6 concludes.

4.2 Theoretical Framework

4.2.1 Proximity, Social Interaction, and Creative Production

Whether through increased concentration of an industry (localisation) or increased size of a city (urbanisation), through urban diversity or local monopolies, geographic proximity is at the core of theories on the circumstances that promote innovative activity. Innovative activity can be facilitated through a variety of mechanisms. Geographically proximate firms and individuals can benefit from the pooling of shared resources, increased variety of resources, specialisation in production, better quality matches between workers, firms, inputs, and ideas, increased generation of new knowledge, and the transmission of knowledge and skills.[1]

These mechanisms can be broadly divided between those that are associated with physical proximity and those related to the interactions between firms or individuals. For example, Ter Wal and Boschma (2009) make the distinction between the importance of location for learning and innovation (i.e. 'space of places') and the importance of networks in the transfer and diffusion of knowledge (i.e. 'space of flows'). Azoulay et al. (2010) distinguish between 'physical space', 'social space', and 'idea space' and argue that spillovers are most likely to occur in 'idea space.'[2] Torre and Rallet (2005), on the other hand, describe four types of proximity in addition to geographic proximity: cognitive proximity (a shared knowledge base), social proximity (socially embedded relationships), institutional proximity ('common habits, routines, established practices, rules, or laws') and organisational proximity ('the ability of an organisation to make its members interact'). They argue that geographic proximity is a necessary but not sufficient condition for spillovers and that at least one additional type of proximity is required for spillovers to occur.

[1] See, for example, Audretsch and Feldman, 2004; Helsley and Strange, 2002; Rosenthal and Strange, 2004; Duranton and Puga, 2004; Jacobs, 1969; Porter, 2011; Glaeser et al., 1992.

[2] For example, academic department colleagues share physical and social space but only co-authors share idea space.

Interactions between firms or individuals are critical for innovation and creative production through mechanisms such as knowledge and skill transfer, learning, sharing of resources, etc. Yet, the effect of such relationships on innovation is rarely tested. "Direct data on information interconnections is typically unavailable to economists. Consequently, economic investigations of the process of social learning have typically made assumptions that relate observed relationships between individuals – such as geographic proximity – to unobserved flows of information" (p 36, Conley and Udry, 2010). Chapter 3, for example, used geographic proximity as a proxy for social interaction under the assumption that the geographic proximity of authors increases the probability of their interaction.

Granovetter (1973) provided one of the first formal arguments that the processes of interpersonal networks (i.e. interaction in small groups) provides more useful insights into macro-level patters. More recently, Jackson (2016) argues that accounting for patterns of interactions is necessary to understand many economic behaviours and that failing to account for such interactions can lead to an incomplete or incorrect understanding of an observed behaviour. For example, the structure of these social relationships will determine the patterns of interactions that can occur. Networks and behaviours 'co-evolve' (i.e. social connections influence behaviours, behaviours influence which social connections are established), and individuals or firms may interact in many different ways at once. Social network methods can, for example, describe assortative matching of group members, a reduction of transaction costs and increase in trust between group members, relative degrees of power and influence, and collaborative activity.

Understanding such relationships is critical to understanding social learning and diffusion processes. Jackson (2016) notes that: "incorporating network structure into the analysis gives us more precision in distinguishing various social forces, since things like diffusion of information, norms, and pressure from peers, all depend in different ways on network structure" (p 4). Thus, social network analysis can provide insight to many interactions described as 'peer effects' and the 'black box' of agglomeration economics. It can also provide a greater understanding of the circumstances in which creativity and cultural

production thrive. In the context of this chapter, information on author networks, such as relative power and influence, may prove useful in explaining variation in literary production.

4.2.2 Power and Prestige in Social Networks

Hanneman and Riddle (2005) and Jackson (2010) provide a comprehensive discussion of social network theory and various network summary statistics, and I summarise their discussions below. I highlight the application of social network theory and methods in describing how the relationships between individuals impact creative production, as studied in the economic geography, peer effects, and cultural economics literature discussed in previous chapters. I focus on how the measures of connection and centrality describe power (how an individual has control over the flow of information and resources across the network) and prestige (how an individual is positioned as a frame of reference for others) within a network.

In social network theory, a network expresses the relational power an individual has in relationships such as the exchange or sharing of knowledge and resources. The way an individual is positioned within a network can impose constraints or offer opportunities. Hanneman and Riddle (2005) note that: "Having a favored position means that an actor may extract better bargains in exchanges, have greater influence, and that the actor will be a focus for deference and attention from those in less favored positions."

The network not only describes the way individuals are connected and the amount of power one individual has relative to her immediate connections but also describes the amount of power within population as a whole. For example, if the network as a whole is not very interconnected, not much power can be exerted by any individual (Hanneman and Riddle, 2005).

I focus on four measures of connection and centrality within the network: degree, betweenness, closeness, and clustering, as these represent the most direct measures of centrality that also provide a comprehensive summary of the structure of an undirected network.

There are more complex measures of centrality based on the characteristics of a node's neighbours, which can capture other aspects of the network, particularly for directed networks. For simplicity, I chose to focus on the characteristics of the node rather than those of its neighbours for the purposes of this paper.

Degree: The degree coefficient is defined as the number links incident upon node i (i.e. the number of connections node i has). With degree centrality, the more connections an individual has, the more power she wields. Individuals with more connections have greater opportunities because they have more choices and are able to take part in more exchanges. They are less dependent on other individuals and have greater access to the resources of the network. The individual is not only able to have access to more resources but may also serve as a central point of reference for her connections.

An important limitation of this measure is that it only considers the immediate connections. Degree does not describe the number of indirect ties an individual has. Therefore, an individual might be directly connected to many people but still disconnected from the network as a whole. If an individual's direct ties are highly interconnected, then the person will only have power and influence within that local neighbourhood.

Betweenness: the betweenness centrality coefficient is defined as the number of shortest paths from all vertices to all others that pass through a given node. For node i, betweenness is defined as:

$$between_i = \Sigma \frac{\Sigma_{st(i)}}{\Sigma_{st}} * \frac{2}{(N-1)(N-2)}$$

where Σ_{st} is the total number of shortest paths from node s to node t and $\Sigma_{st(i)}$ is the number of those shortest paths that pass through node i. Assuming that resources and ideas are transferred along the shortest paths, an individual with a high betweenness coefficient has a large amount of power over the transfer of resources and ideas through the network. While a person with high betweenness can serve as a point of contact between others, she also has the power to isolate individuals or groups of individuals and prevent the flow of ideas and resources.

For example, consider a network composed of two distinct groups in which the individuals within each group are highly interconnected and only a few individuals serve as 'bridges' between the two groups. The individuals serving as 'bridges' will have high betweenness centrality. These individuals may not have power or influence in terms of having many connections but will wield great power for their ability to allow or prevent flow of resources and ideas between the groups. Such individuals may benefit in terms of heightened creativity, as information that flows between the groups must also pass through them. These individuals may also benefit from extracting high 'service charges' for their role in connecting the two groups. These 'service charges' can take the form of financial transactions (payment for services) or non-financial transactions ('owing a favour').

Closeness: The closeness coefficient describes how close a node is to all other nodes in the network. I first determine farness, defined as the sum of the distance from each actor to all others in the network: $farness_i = \frac{\Sigma(dist_{ik})}{k}$. I then calculate nearness, defined as the reciprocal of farness: $nearness_i = \frac{1}{farness_i}$. I define closeness as nearness normalised to 1:

$$closeness_i = nearness_i(n-1)$$

with $dist_{ik}$ being the length of the shortest path from node i to node k in a network with n nodes.

Individuals with high closeness centrality are able to reach others at shorter path lengths, but this does not necessarily indicate that they serve as important 'bridges' between groups. Through closeness, power comes from acting as a reference point by which others in the network judge themselves. Individuals with high closeness can be very influential voices in the network, as their position as the 'centre of attention' allows their views to be heard by a large number of others. However, information flows both ways. Individuals with high closeness coefficients are also more reachable. Thus, they may benefit from greater exposure to ideas and resources from diverse sources they are not immediately connected to.

Clustering: The local clustering coefficient of node i in the network graph g is defined as the share of network contacts of i ($N_i(g)$) who are directly connected among themselves:

$$cluster_i(g) = \frac{\#\{jk \in g | k \neq j, j \in N_i(g), k \in N_i(g)\}}{\#\{jk | k \neq j, j \in N_i(g), k \in N_i(g)\}}$$

The clustering coefficient is the ratio of existing links connecting an individual's neighbours to each other to the maximum possible number of such links. In other words, clustering describes the degree to which one's connections are also interconnected – or "all of my friends are friends." A high clustering coefficient is an indication of a 'small world.' As Hanneman and Riddle (2005) explain: "Most of the people we know may also know one another – seeming to locate us in a very narrow social world. Yet, at the same time, we can be at quite short distances to vast numbers of people that we don't know at all."

A high clustering coefficient can be beneficial, as it may indicate a high level of familiarity and trust within a group. For example, a high degree of clustering can reflect the presence of a collaborative circle, such as a group of academics who regularly co-author together or literary group who review and edit early-stage drafts. Ter Wal and Boschma (2009) also warn that such groups "carry the risk of getting locked in established ways of thinking and a lack of new knowledge coming in" (p 745), making it more difficult to innovate.

4.3 The Author Network

This chapter combines a new unique, purpose-build dataset of author social connections with an existing dataset with location, biographical and peer data of 370 prominent literary artists born between 1700–1925 and associated with British or Celtic literature.[3,4] The existing dataset includes age, lifespan, age at first publication, number of publications per annum, lifetime publications, career duration, author quality, and location for every

[3]Celtic literature is defined here as literature associated with Celtic nations within the British Isles (Ireland, Scotland, Isle of Man, Wales, and Cornwall) and does not strictly refer to literature written in the Celtic languages.

[4]Authors are defined as individuals who made at least one unique contribution to poetry or prose. The dataset does not include individuals whose contributions were strictly limited to translations, textbooks, manuals or guides, song-writing, literary criticism, or publishing.

year of the author's life. Output is measured in terms of number of known publications.[5] Author quality, given by the Impact Index, is defined as the amount of contemporary critical attention in academic literary studies given to one author relative to her peers.[6] For reference, the ten highest quality authors are listed in Table 4.4.

This dataset also includes data on number of peers, mean peer quality and mean peer output for each author in each year. Peers are defined as authors who live in the same city in the same year. Peer variable averages are calculated as leave-out means, i.e. own-observations are excluded from the calculation. The author-year observations are limited to potential active years defined as age 16 and older which represents the lower bound for age at first publication. The sample used in the analysis begins in 1725 (the first year with multiple observations) and ends in 1999 (when the last authors died).

Summary statistics of these biographical data are presented in Table 4.1. On average, authors lived for 66 years and spent 21.5 years in London, the only major literary cluster (see Chapter 2). The average author published her first work around age 27 and worked for just under 40 years. Authors produced an average of 33 publications over the course of their careers and published an average of 0.88 works per year. Of the 370 authors, 70 are female. England was the birthplace to the most authors (245), followed by Ireland (44), Scotland (41), and Wales (10). One author (Henry Luttrell) has an unknown birth location, 11 authors were born in other parts of Europe, and 18 authors were born elsewhere in the world.

There may be concern that authors born outside the UK and Ireland may be systematically disadvantaged in establishing social connections with other authors in the network. The

[5]Publications include any type of unique literary publication: novel, collection of poetry or short stories, anthologies, contributions to literary magazines, plays, memoirs, etc. Individuals who only published non-fiction works (textbooks, dictionaries, journalism, etc.) and/or translations are not included. All publications were given equal weight. There is no measure of the quality of individual publications.

[6]The amount of critical attention has two components: an indexed measure of the total words designated to an author from the three biographical sources and an indexed measure of the total number of citations within literary criticism as listed in the Literature Online (2014) database. These measures carry equal weight in an overall index of author quality. The author who receives the most contemporary critical attention is given a measure of 100. An author with a quality measure of 50 receives 50% of the critical attention the highest-ranked author receives. A more detailed description of the data collection process and the creation of the quality index may be found in Chapter 2.

locations elsewhere in the world primarily consist of parts of the British Empire: India (6), Australia (1), Burma – now Myanmar (1), New Zealand (2), South Africa (1), and Southern Rhodesia – now Zimbabwe (1). Others were born to British parents living abroad but outside the British Empire who often returned to the British Isles at relatively young ages. There are some notable exceptions. However, all such individuals successfully integrated into British society. There is no evidence to suggest that being born abroad or being an immigrant systematically disadvantaged any of the authors.

For example, Wyndham Lewis was born to a British father and an American mother. Though he was born on his father's yacht off the coast of Canada, his parents subsequently returned to England. W.H. Hudson's parents were Americans (of British and Irish origin). While Hudson was born in Buenos Aires, he moved to England at age 28. Hudson became a naturalised British citizen in 1900. T.S. Eliot was born in Missouri, USA to American parents. Eliot moved to England at age 25, became a naturalised British citizen, and renounced his US citizenship.[7]

Jennie Jerome Churchill was born in New York, USA to American parents. Churchill moved to Paris at age 13 and never lived in the US again. She married Lord Randolph Churchill at age 20 and became a well-known British socialite. Churchill is also famously the mother of Sir Winston Churchill (the former Prime Minister of the UK), who is also an author included in this dataset.[8,9]

Joseph Conrad is notable as the only author who was not a native English speaker. Joseph Conrad was born in Ukraine to Polish parents. Conrad first travelled to England in 1878 after serving on a British freighter. At the time, he had very limited knowledge of the English language (though his literary works were written almost exclusively in English). He eventually settled in England and was naturalised in 1886. Despite the language disadvantage, Conrad is recognised as "a writer of complex skill and striking insight, but above

[7]See Encyclopaedia Britannica (2014) entries on *Wyndham Lewis, W.H. Hudson, and T.S. Eliot* for more information.

[8]Although he is more famous as a politician and non-fiction writings, Sir Winston Churchill also wrote a novel, a short-story, and poetry.

[9]See Encyclopaedia Britannica (2014) entries on *Jennie Jerome Churchill and Winston Churchill* for more information.

all of an intensely personal vision, he has been increasingly regarded as one of the greatest English novelists" (*Joseph Conrad*, Encyclopaedia Britannica, 2014.)

The data on author social connections were collected from three online encyclopaedias: Encyclopaedia Britannica (2014), The Literary Encyclopaedia (2014), and Literature Online (2014). All names mentioned in an authors' biographical entries were recorded as possible social connections. For the purposes of this study, social connections are strictly limited to other authors in the dataset who have overlapping lives (i.e. both authors were alive simultaneously for any period of time). Social ties are assumed to be undirected: if author A is connected to author B, then author B is also connected to author A. The author network graph is presented in Figure 4.1. The size of the nodes is increasing with degree centrality.

The list of social connections is not a comprehensive list of all individuals an author met or interacted with. Rather, the list includes individuals such as friends, family, professional contacts, important rivals, etc. who played an important role in an author's personal life, professional life, or both. It is important to emphasise that authors likely had more actual social connections than those included in the dataset. The dataset only includes connections mentioned in the three biographical sources. As the individuals who wrote the biographies have expert knowledge on the respective author, I rely on their expertise to determine which connections are important enough to merit mention in a biography.

I argue that this data limitation is a reasonable one. Not all social connections would have an impact on an author's writing process or access to important literary actors, and those who were vital to an author would be worth mentioning. While other important figures were listed in the biographies, the network only includes other authors in the existing dataset. Therefore, I qualify the discussion of the network and the results by noting that I only consider the effect of other prominent authors associated with British or Celtic literature.

Sir Walter Scott, for example, was an important influence on reception and critical evaluation of Jane Austen's work, even though there is no indication in the biographies that the

two ever met in person. Jane Austen's early novels were well-reviewed and widely read. Despite this success, her experimentation with literary style was not fully appreciated, and she was largely regarded as a 'merely domestic' novelist. "During her lifetime there had been a solitary response in any way adequate to the nature of her achievement: Sir Walter Scott's review of *Emma* in the *Quarterly Review* for March 1816, where he hailed this 'nameless author' as a masterful exponent of 'the modern novel' in the new realist tradition...[This review] provided the foundation for serious criticism of Jane Austen" (*Jane Austen*, Encyclopaedia Britannica, 2014.)

J.R.R. Tolkien, on the other hand, is connected with fellow writers and collaborators, including C.S. Lewis and John Barrington Wain. Tolkien and Lewis were members of an informal collaborative group of writers (known as the Inklings) based in Oxford who met regularly through the 1930s and 40s. The group began as casual meetings between Tolkien, C.S. Lewis and Owen Barfield, during which they discussed politics and philosophy. The group slowly expanded to include other writers, and eventually – and perhaps most influentially – they began to discuss their own unfinished works. Encyclopaedia Britannica (2014) notes that:

> *"The group contributed significantly to its members' success through its criticism, support, and encouragement, an indebtedness evident in the acknowledgment pages and dedication pages of many of their works: Lewis's The Problem of Pain, Williams's The Forgiveness of Sins, and the first edition of Tolkien's The Lord of the Rings were dedicated to the Inklings. Lewis wrote of the Inklings, "What I owe to them all is incalculable," and Tolkien noted that "only by [Lewis's] support and friendship did I ever struggle to the end" of The Lord of the Rings."*[10]

John Barrington Wain is connected to the Inklings because C.S. Lewis was his tutor at Oxford, though he was never a member of that collaborative group. He was, however, a member of another literary group known as The Movement. His connections, Sir Kingsley Amis and Philip Larkin, were also associated with this group. The Movement was a group

[10]See Encyclopaedia Britannica (2014) entry on *The Inklings*

of writers and poets who shared ideas, criticised each other's works, and supported and encouraged one another, much like the Inklings. The Movement, however, "wrote understated and disciplined verse that avoided experimentation and grandiose themes" (*Kingsley Amis*, Encyclopaedia Britannica, 2014) in direct contrast to the complex, mythopoeic style of the Inklings' generation Farrell (2001). Thus, the social connection data contain links between generations of authors, as well as within a generation of authors.

4.3.1 Network Connection and Distance

A summary of measures of network connection and distance are provided in Table 4.2. The graph is relatively disconnected, since 70 of the 370 nodes are isolates – i.e. 70 authors are not connected to any other authors. A more informative property is whether any pair of nodes are reachable. That is, whether a path exists between any two pair of nodes. For example, Wasserman and Faust (1994) note that:

> *"If we consider a network of communications among people in which lines in a graph represent channels for transmission of messages between people, then if two actors are reachable, it is possible for a message to travel from one actor to the other by passing the message through intermediaries. If two actors are not reachable, then there is no path between them, and no way for a message to travel from one actor to the other."* (p 107)

Given a network with 370 nodes, the maximum number of undirected paths (i.e. connections between any two nodes) is $\frac{370*369}{2} = 68265$. The number of existing paths (i.e. connections between any two nodes in the largest connected component of the graph) is 44253. Therefore, 64.8% of the pairs of nodes are reachable.[11]

In comparison, Newman (2004) studied scientific co-authorship networks and found that 21% of biomedical research papers published from 1995 to 1999 were authored by a single

[11] Jackson (2010) provides a comprehensive discussion of social network theory and the various network summary statistics.

person. Bearman et al. (2004) studied the structure of adolescent romantic and sexual networks in the context of a broader study on STD transmission. In their study, just under 25% of the nodes were isolates. Ahuja (2000) used patents to measure the innovative collaborations between firms in the international chemical industry. The number of collaborations increased over the 10-year period from 1981 to 1991: 26% of firms were isolates in 1981, while only 14% of firms were isolates in 1991. In this study, 18.9% of authors are isolates. Given the very long time period and the historical context, the author network has a high level of connection.

The flow of knowledge or ideas across a network can also be described in terms of path length or the 'number of steps' it takes to get from one node to another. The average shortest path is based on the existing 44253 paths and finds the shortest path between every two pair of nodes in the network. This concept is known in popular culture as 'six degrees of separation' – the idea that any two people in the world can be connected through six or fewer 'friends of friends.' The average shortest path or 'degrees of separation' between the authors in this study is 4.91. In contrast, the diameter of the author network is 13, indicating that it takes 13 'steps' to get from the two most distant nodes in the largest connected component of the graph.

The average path length of the author network is shorter than those of some contemporary creative networks. Goyal et al. (2006) studied economic co-authorship networks and found that the average shortest path length between economists decreased from 12.86 in the 1970s to 9.47 in the 1990s. As this study focuses on social connections rather than professional collaborations, a more relevant comparison may be the Ugander et al. (2011) study of the network of Facebook users in May 2011. They found that the average shortest path length between users in the US was 4.3 and the average shortest path length between all users globally was 4.7. In more recent study, Edunov et al. (2016) found that the average degree of separation between any two of the 1.59 billion Facebook users in 2016 is 3.57. Indeed, the average shortest path length between authors is closer to that of contemporary social media users than scientific co-author networks.

Similarly, the diameter of the author network is relatively short compared to other creative networks. The Newman (2004) study of various scientific co-authorship networks found the diameter of the biologist network to be 24. The diameter of the mathematician network was 27, while the diameter of the physicist network was 20. In contrast, the Bearman et al. (2004) study of adolescent romantic and sexual networks found the student network had a diameter of 37.

While various measures of path length describe how far apart individuals in the network are, they do not indicate the distribution of connections or the structural pattern of the network. Rather intuitively, individuals tend not to be equally connected to all others in the network. Instead, social groups tend to form 'cliques' or local neighbourhoods in which all individuals are directly tied to one another. The degree of 'cliqueishness' of a network is given by the clustering coefficient. For an individual node, a clustering coefficient of 0.75 means that 75% of all possible ties in the neighbourhood are present, and a coefficient of 1 indicates that all of an individual's connections are connected to one another.

The average clustering coefficient of the author network is 0.164, indicating that the average author is not part of a well-defined and exclusive 'clique.' The global clustering coefficient is the ratio of the number of existing closed triplets (sets of three fully connected authors) to the number of possible triplets. The global clustering coefficient for the author network is 0.195, indicating that 19.5% of all triplets are closed. While there is some degree of group formation, the average author also interacted with others outside these groups. This is somewhat expected, given the long time period. Authors are connecting across generations as well as within generations. There will be some degree of overlap across generations, but there are natural limits to the number of triplets that can be closed.

In this respect, the global clustering coefficient is relatively high. There appears to be a strong legacy element to the author network. Younger generations of authors tend to arise with a connection to older generations. This may be due to the nature of authors themselves. Established authors may seek out rising stars to mentor or may serve as literary critics who review new writings. Given that this dataset is limited to prominent

authors, it is possible that an author is viewed as eminent because she is directly connected to other eminent authors. In this case, prominence is associated with status quo, even if the nature of the status quo with respect to genres and styles evolves over time. In this respect, social connections may signal reputation in addition to serving as a means for knowledge transmission.

The level of clustering is similar to other creative networks. For example, Goyal et al. (2006) found the clustering coefficient of the co-author network of economists had a clustering coefficient of 0.193 is the 1970s and a clustering coefficient of 0.157 in the 1990s. Newman (2004) found wide variation in the clustering coefficients across scientific fields. The network of biologists had a clustering coefficient of 0.066, while the network of mathematicians had a clustering coefficient of 0.15. In contrast, the physicist network had a clustering coefficient of 0.43. It may also be useful to compare the author network to social networks in addition to collaborative scientific networks. Ugander et al. (2011) found that for Facebook users with 100 friends, the average clustering coefficient was 0.14.

Network density indicates the level of cohesion in the network. Density is defined as: $density = \frac{NC_a}{NC_p}$ where NC_a is the number of actual connections and NC_p is the number of possible connections. The number of possible connections is defined as: $NC_p = \frac{n*(n-1)}{2}$ where n is the number of nodes. The network density is quite low: 1% of all possible connections actually exist. In comparison, Ahuja (2000) found that the network of firms in the international chemical industry exhibited density of 2.9% in 1981 and 5.6% in 1991. Degree centralisation describes the extent to which the network is organised around focal points. Degree centralisation of the author network is 0.055.

Given the low density and moderate clustering coefficient, combined with a low 'degree of separation' and small diameter, this suggests that there are a few individuals who serve as important connections or 'bridges' between 'cliques' between dense neighbourhoods of authors. This indicates that certain authors serve an important role in connecting groups of authors across generations and within generations. These authors would also serve a key role in the transmission of knowledge and ideas. They have the power to transmit information between groups – and the power to withhold it.

4.3.2 Measures of Network Centrality

Summary statistics for measures of network centrality are presented in Table 4.3. The degree coefficient is the number of ties an author has. The average author has degree of 3.8. The distribution of degree is displayed in Figure 4.2. The degree coefficient ranges from 0 to 24. 60% of authors have 3 or fewer ties, and 90% of authors have less than 10 ties. The higher degree an author has, the more power the author has. The author is less dependent on others and has access to greater resources and ideas. However, the author's power may be quite limited if her local network is disconnected from the greater network.

Table 4.5 presents the authors with the ten highest degree coefficients. The list includes authors from all three centuries, from a variety of birth locations, and across literary forms. The top authors are Sir Walter Scott (a Scottish playwright, poet, and novelist publishing in the late 18th and early 19th centuries) and William Butler Yeats (an 20th century Irish poet) who have 24 and 23 ties, respectively. The 18th century English playwright David Garrick and the 20th century English modernist writer Virginia Woolf are also among the most connected authors. The list also includes T.S. Eliot, a naturalised British citizen and a 20th century essayist, playwright, and poet.

The local networks for a sample of authors are listed in Table 4.5. Figure 4.3 displays the local network of Sir Walter Scott. Figure 4.4 shows the local network of Virginia Woolf, and Figure 4.5 presents the local network of William Butler Yeats. Sir Walter Scott's connections include Romantic poets and novelists such as Lord Byron, William Wordsworth, Robert Southey, and Amelia Opie. Virginia Woolf's connections include many prominent female authors such as Vita Sackville-West, Dame Edith Sitwell, Katherine Mansfield, Dorothy Richardson, and Alice Meynell. William Butler Yeats is primarily connected to prominent Irish writers such as Oscar Wilde, Edward Martyn, James Joyce, George Bernard Shaw, Oliver St. John Fogarty, Frank O'Connor, AE (George William Russell), and Lady Augusta Gregory.

The betweenness coefficient is equal to the number of shortest paths between all authors that pass through an individual, standardised to 1. The mean betweenness coefficient is

0.014, and betweenness coefficients range from 0 to 0.167. Table 4.6 displays the authors with the ten highest betweenness coefficients. Ford Madox Ford has the highest betweenness coefficient, indicating that he has the most power and influence over the transfer of knowledge and ideas through the network. Other authors with high betweenness who also had high degree, for example, William Wordsworth, Virginia Woolf, Sir Walter Scott, and William Butler Yeats. These authors have a large amount of local influence and access to greater local resources and ideas, and they also have power over the transfer of knowledge and ideas across the network.

Ford Madox Ford, for example, is an pivotal connection for Joseph Conrad. Joseph Conrad and Ford Madox Ford are famous for their collaborative relationship. Farrell (2001) notes that:

> *"Without that interaction, Ford might never have developed into a serious novelist. Careful analysis of his work during the collaboration, especially his successive revisions of the "Romance" manuscript, shows how he matured as a novelist as a direct consequence of working with Conrad...For Conrad, the collaboration contributed to his mastery of English, brought economy of expression to his sometimes flamboyant style, and suggested scenes, settings, and plots for his work. More important, the relationship stabilized Conrad emotionally and contributed to his ability to explore his own psychological conflicts through his writing."* (pp 132-133)

Their relationship was not only important for the development of their respective works but also because of the connections Ford had with prominent writers. Figure 4.6 illustrates the relative power Ford has compared to Conrad. Conrad has degree 7, but few of his contacts have a high degree coefficient (as displayed by the size of their respective nodes) Through Ford Madox Ford, however, Conrad is linked to prominent and well-connected authors such as William Butler Years, Thomas Hardy, James Joyce, Dante and Christina Rossetti, and D.H. Lawrence. Without these links, it is possible that neither of their works would have had the same degree of reach or exposure, since success as an author depends on ability to get works printed and gain an audience in addition to writing quality.

The closeness coefficient is a measure of how close an author is to all other authors in the network (normalised to 1). Closeness coefficients of the author network range from 0 to 0.35. Mean closeness is 0.21, indicating that the average author tends to be embedded in a local cluster rather than being highly connected to many others. Again, due to the long time period, this is somewhat expected. However, it does imply that individuals with relatively high closeness coefficients play an important role in the spread of ideas and knowledge through the network, as they have access to more authors at shorter path lengths and thus can more efficiently gain information from more people. In contrast to those with high betweenness, authors with high closeness do not necessarily have the ability to prevent or allow knowledge from transferring between groups.

The authors with the ten highest closeness coefficients are listed in Table 4.7. Ford Madox Ford also has the highest closeness coefficient. Ford was able to efficiently gather information from many sources (high closeness) and had a relatively high degree of power in determining whether that information gets transferred between groups (high betweenness). In fact, many of the authors with high betweenness coefficients listed in Table 4.7 also have high closeness coefficients listed in Table 4.6, with the exception of Robert Browning and Thomas Hardy. This supports the conclusion drawn in Section 4.3.1 that authors form loose groups and that a few individuals serve as important connections between the denser groups of authors.

The clustering coefficient provides a measure of whether all of author's local network is closed – i.e. whether all her contacts are interconnected. As mentioned in Section 4.3.1, the average clustering coefficient of the author network is 0.164. Thus, on average, 16.4% of an author's connections are also connected. As the clustering coefficient is defined as a share, the value is bound between 0 and 1. Table 4.8 lists ten authors with clustering coefficients of 1.

While high clustering coefficients can be an indication of 'small worlds' within the network, this can also simply be a sign that an author had very few interconnected contacts. As mentioned previously, 60% of authors have 3 or fewer ties, so the first ten authors listed

alphabetically are presented. For example, Anne Brontë's only connections are her sisters, Emily and Charlotte. Lascelles Abercrombie is connected only to fellow poets Rupert Brooke and John Drinkwater, and Arthur Hugh Clough's sole connections are his friends and fellow poets Matthew Arnold and Francis Turner Palgrave. However, it is important to re-emphasise that these individuals likely had more actual social connections than those listed. The dataset only includes connections mentioned in the three biographical sources who are also authors included in the dataset.

Furthermore, it is notable that neither of the two highest quality authors (Table 4.4) are present in any of the 'top ten' lists of centrality coefficients. The local network of James Joyce is presented in Figure 4.7, and the local network of Charles Dickens is shown in Figure 4.8. While neither author is among the listed authors with the highest degree, betweenness, or closeness coefficients, they have direct connections to multiple authors on those lists. Joyce is connected to Ford Madox Ford (betweenness and closeness) and William Butler Yeats (betweenness and degree). Dickens is connected to Leigh Hunt (degree, betweenness, and closeness), Thomas Carlyle (degree), and Alice Meynell (betweenness and closeness).

It is also notable that none of the highest producing authors are present in any of the 'top ten' lists. The top ten authors who produced the most over their lifetime are listed in Table 4.9. Of the local author networks presented thus far, only Robert Graves is present as a connection to Virginia Woolf. While being 'well connected' does appear to be significant in terms of the amount of contemporary literary attention an author receives, the link between connectedness and output in terms of the number of works produced is less apparent.

4.4 Econometric Approach

I aim to quantify the gains in predictive power that may be realised through the use of social network information. Following a methodology similar to that of Ductor et al. (2014), the data is randomly partitioned into a training sample and a test sample. I estimate the model

using the training sample. I then test the extent to which this training sample can predict the test sample. The data is then replaced, randomly split into training and test samples, and then trained and tested again. I perform 200 replications of this repeated random sub-sampling validation (i.e. Monte-Carlo cross-validation) procedure for each model, and the final result is the average of the results of the results for the 200 splits. I perform this procedure for four models. Model 1, the baseline model, is a standard age-productivity profile with with a quadratic age term and time dummies (γ_t):

$$\text{Model 1:} \qquad output_{it} = \alpha + \delta_1 age_{it} + \delta_2 age_{it}^2 + \gamma_t + \epsilon_{it}$$

Model 2 adds to Model 1 a vector of peer effects variables (i.e. those utilised in Chapter 3) and a London fixed effect:

$$\text{Model 2:} \qquad output_{it} = \alpha + \delta_1 age_{it} + \delta_2 age_{it}^2 + \phi London_{it} + \lambda \mathbf{PeerEffects}_{it} + \gamma_t + \epsilon_{it}$$

Model 3 adds to Model 2 one of the network variables:

$$\begin{aligned}\text{Model 3:} \qquad output_{it} =& \alpha + \delta_1 age_{it} + \delta_2 age_{it}^2 + \phi London_{it} + \lambda \mathbf{PeerEffects}_{it} \\ &+ \beta NetworkCentrality_i + \gamma_t + \epsilon_{it}\end{aligned}$$

There are four variants of Model 3, one for each of the network variables. Model 4 adds to Model 2 all of the networks variables simultaneously:

$$\begin{aligned}\text{Model 4:} \qquad output_{it} =& \alpha + \delta_1 age_{it} + \delta_2 age_{it}^2 + \phi London_{it} + \lambda \mathbf{PeerEffects}_{it} \\ &+ \beta_1 Degree_i + \beta_2 Betweenness_i + \beta_3 Closeness_i \\ &+ \beta_4 Clustering_i + \gamma_t + \epsilon_{it}\end{aligned}$$

Cross-validation provides insight into how a model will perform when applied to an independent dataset. Multiple rounds of a validation procedure are used to reduce variability, and the mean measures of fit are used to assess the model's performance. Repeated random sub-sampling validation, or Monte-Carlo cross-validation, provides several advantages over other validation methods. This method is asymptotically consistent, and it has a lower risk of overfitting the model compared to other methods, such as leave-one-out validation. This

method also has a larger probability of selecting a model with the best prediction ability. (See Xu et al., 2004.) It is possible for the same observation to be selected in the training set in multiple splits. Therefore, a disadvantage is that authors with longer lifespans are more likely to have an observation included in the training sample and, similarly, authors with very short lifespans may be under-represented in the procedure.

I present the R^2 of the regression on the training data for each model and the test RMSE for each model. I compare the RMSE of each model to the baseline model and show the coefficient of each regressor. With this procedure, it is not possible to cluster standard errors. The reported coefficients and significance levels for the forecasting procedure reflect this. The greater the RMSE differential, the more explanatory power a model has a beyond the baseline age-productivity profile. I expect the models that include social network variables to provide greater explanatory power than the baseline model. I also expect that the RMSE differential for the social network models will be greater than the model with peer network variables, in line with the hypothesis that spillovers are more likely to occur in "social" or "idea" space than "physical" space.

Though a causal relationship cannot be established, this validation procedure provides evidence of how social network centrality relates to production for those who work in the creative industries and whether these variables provide greater explanatory power than localised peer effect variables. Collection of social network data for historical figures can be a time- and resource-consuming process, as often this information is not easily available. Thus, it tends to be much easier to simply define peers based on proximity. If social network variables provide a substantial improvement in explanatory power compared to proximity-based relationships, then this may provide greater incentive for others to investigate the role of social networks for other cultural and creative industries.

Ideally, I would also establish a causal relationship between an author's position in the social network and her productivity. A simple and straightforward approach is to estimate the following relationship by ordinary least squares using a pooled panel of workers:

$$Output_{it} = \alpha + \beta_1 Degree_i + \beta_2 Betweenness_i + \beta_3 Closeness_i + \beta_4 Clustering_i + \boldsymbol{\delta}\mathbf{x}_{it} + v_{it} \tag{4.1}$$

where $Degree_i$, $Betweenness_i$, $Closeness_i$, and $Clustering_i$ are the degree, betweenness, closeness, and clustering centrality values for author i, respectively, and $\mathbf{x}_{it}$ is a vector of individual-level characteristics that vary over time. However, that the network centrality measures do not change over time presents a challenge for identification.

An instrumental variable approach can control for self-selection into location and social groups. For example, in order to account for spatial sorting, Ciccone and Hall (1996) suggest using long lags of population density to instrument for the size or density of local population based on the argument that the spatial distribution of the population persists over time and that this differs substantially from the contemporary factors impacting productivity at the local level. Borowiecki (2013) uses birthplace-cluster distance to instrument for the spatial sorting of prominent Western classical music composers, as birthplace-to-cluster distance captures not only the costs associated with travel but also factors such as cultural and linguistic distance.

However, detailed data on the population of cities in the UK and Ireland are not available for most of the time period covered in this analysis. Given the limited geographic scope of this study, the travel distances are much smaller and the culture and language less varied. Particularly after the development of extensive rail lines by the mid-1800s, distance between cities was not likely a deterring factor in migration within the UK. Additional detailed individual-level information would be needed to control for individual heterogeneity in order to utilise birthplace-city distance or lags of population density as an instrument, which is not available for all authors in the limited historical sources. Furthermore, neither approach addresses the problems associated with natural ability: productivity and social connectedness are both likely to be positively affected by natural ability.

It is possible to focus on a subset of authors and utilise an unanticipated, exogenous shock and a difference-in-differences approach to find a causal relationship. For example,

Waldinger (2010) and Waldinger (2012b) use the dismissal of mainly Jewish scientists and mathematicians from German universities during World War II as an exogenous shock to establish causality. As Waldinger explains, this was an unexpected political move and none of the academics could access to similar facilities in Germany after their dismissal. None of the academics continued to work as normal under the circumstances and many left the country, leading to a discernible loss of human capital within the academic departments.

In the UK and Ireland, there was no comparable exogenous shock, even during World War II. While men were drafted and women worked for the war service domestically, neither were arguably unexpected by the time the UK became involved in the war and by the time conscription became compulsory. While this shock would impact localised peer networks as defined in Chapter 3, this did not necessarily disrupt social networks, at least not in the same way the dismissal of academics disrupted collaborative research networks.

In the absence of a suitable instrument or exogenous shock, I can only provide descriptive evidence. However, such results are informative as they illustrate how various interactions may impact the productivity of authors, which may provide some guiding intuition for future work. I first estimate age-productivity profiles with a quadratic age term, time dummies, and one network variable using pooled OLS models. I repeat this estimation with all network variables included simultaneously. I add to that model a London fixed effect. I then add the peer variables utilised in Chapter 3. In all regressions, standard errors are clustered at the author level.

4.5 Empirical Findings

4.5.1 The Predictive Power of Social Networks

The results to the validation procedure are presented in Table 4.10. As mentioned in Section 4.4, Model 1 is an age-productivity profile with quadratic age term and time dummies. Model 2 adds to Model 1 peer variables and a London fixed effect. Model 3

adds to Model 2 one of the network variables. Model 4 adds to Model 2 all of the networks variables, simultaneously. Column (1) presents the R^2 of the regression on the test sample data for each model. Column (2) shows the training sample RMSE for each model. Column 3 compares the RMSE of each model to the benchmark model, Model 1. Column 4 shows the coefficient of each regressor. The greater the RMSE differential, the more explanatory power a model has a beyond the baseline age-productivity profile.

An age-productivity profile with time dummies (Model 1) explains approximately 11.5% of variation in annual output. In Model 2, the coefficients of London, mean peer quality, and peer group size are statistically significant, but this does not necessarily indicate that these variables are informative. The RMSE differential indicates that Model 2 improves predictive power by 0.009%. Thus, the peer effects variables and London fixed effect provide essentially no improvement in predictive power compared to a standard age-productivity profile.

In Model 3, each of the four network variables improves predictive power relative to Model 1, confirming that social network variables better explain variation in production than proximity-based peer variables. Of the social network variables, clustering has the least predictive power. Given the limitations of the clustering coefficient described in Section 4.3.2, this result is not surprising. While a high clustering coefficient can indicate a high degree of interconnectedness of an author's social connections, it can also reflect that an author has only a few connections who know one another. High interconnectedness may increase the ease of transfer of ideas and knowledge, but it may also increase the risk of getting 'locked-in' to established ways of thinking, particularly in the case of few but interconnected social connections. As the clustering coefficient likely captures multiple and opposing mechanisms, it is reasonable that clustering would have low predictive power.

Betweenness and closeness clustering coefficients improve predictive power relative to Model 1 by 0.92% and 0.86%, respectively. While this improvement is somewhat modest, it is two orders of magnitude greater than that of Model 2 with proximity-based peer effects. The degree coefficient has the greatest predictive power, with an RMSE differential

of 1.79%. When all social network variables are estimated simultaneously in Model 4, the RMSE differential is 1.65%, suggesting that a full model may perform no better than a more parsimonious model that only adds degree to Model 2.

While peer effects variables (in which peers are defined by proximity) have little to no improvement in predictive power relative to a standard age-productivity profile, models with social network variables yield a much larger improvement in predictive power. Thus, social network variables not only provide more intuitive explanations of the mechanisms underlying production, but they also provide greater predictive power relative to proximity-based peer effects.

This has important implications for the study of knowledge spillovers across 'idea space', 'social space', and 'physical space.' It is likely that physical proximity is not a sufficient condition for knowledge spillovers between authors, at least in terms of spillovers that affect annual productivity. It is also possible that physical proximity is not a necessary condition for spillovers once an initial social connection is established. Agrawal et al. (2006), for example, find evidence of enduring social capital for inventors – that social ties facilitate knowledge transfer that persists even when inventors are no longer co-located. However, dynamic network data would be required to empirically test this.

Furthermore, in fields that require a large knowledge base, such as STEM (science, technology, engineering, and mathematics), the proximal relationships that do (and do not) facilitate spillovers are distinct. For example, Azoulay et al. (2010) do not find evidence of spillovers between academic science department colleagues, who are presumably socially and physically proximate. They do find evidence of spillovers between academics working within the same boundaries within an academic field, i.e. working in the same 'idea space'. For these types of creative workers, there are clear boundaries between 'idea space' and 'social space.' While some colleagues may fall under both categories for a given academic (e.g. friends and co-authors), spillovers are limited to shared 'idea space.'

The distinction between 'idea space' and 'social space' is not so clear when it comes to authors. The results suggest that simply being an author living and writing in the same

time period may be sufficient to qualify as 'idea space.' Future studies determining a causal relationship between social network centrality and productivity could examine these distinctions by incorporating factors such as an author's associated literary groups (e.g. the Inklings), an author's primary genre of writing or the genre of each work, and dynamic network measures. For example, it may be possible to limit the study to a subsample of authors and track the exchange of letters to determine the direction and intensity of interaction.

4.5.2 Social Networks and Productivity

In the absence of a suitable instrumental variable, exogenous shock, or randomised experiment, the results discussed in this section only provide descriptive evidence. The coefficients for pooled OLS estimates of the effect of various measures of network centrality on annual productivity are provided in Table 4.11. Columns (1)–(4) provide the estimates of an age-productivity profile with betweenness, closeness, clustering, and degree, respectively. Column (5) includes all four centrality measures simultaneously. Column (6) adds a London fixed effect, and Column (7) adds peer group size, peer quality, and peer output. All regressions include time dummies, and standard errors are clustered at the author level.

Betweenness, closeness and degree have statistically significant and positive effects when included separately. Based on the summary statistics discussed in Section 4.3.2, it is likely that these variables are highly correlated. This is confirmed in Table 4.12: degree is highly correlated with betweenness and closeness. Thus, the estimates in Columns (5)–(7) likely suffer from multicollinearity. When all network centrality measures are estimated simultaneously, only degree has a statistically significant effect. This is symptomatic of multicollinearity, but it also suggests that degree has greater explanatory power than betweenness and closeness.

In the most parsimonious model in Column (4), a one unit increase in degree (i.e. one additional social connection) is associated with a 0.035 unit increase in per annum productivity. The mean works produced per annum is 0.88, indicating that one additional social

connection is associated with a 4% increase in annual productivity. However, I note that this relationship is not causal. The coefficients for degree are consistent across all four models in Columns (4)–(7): the values range from 0.035 to 0.036, thus the explanatory power of degree remains robust to variations in specification.

Consistent with the results of Chapter 3, living in a city with more people is associated with a reduction in productivity, yet having more social connections within the author network is associated with an increase in productivity. The selection process in choosing social ties is not necessarily the same as for choosing peer group (i.e. location), and it is likely that proximity-based peer effects and social network effects operate through different mechanisms. Defining peers by location implies that the only peers that are important are those based in the same location and that authors in other locations do not influence productivity. This is not necessarily a ideal assumption, but it is often made for practical purposes when more detailed data on interactions are not available. Perhaps more importantly, when peers are defined by location, it is not possible to disentangle the effects of learning, sharing ideas, and collaboration from the effects of economies of scale and competition in access to resources.

The peer effects findings in Chapter 3 likely capture competition effects associated with access to limited resources. Due to physical capacity constraints, publishers can only publish a certain number of author's works in a year. Therefore, in location with more authors, it may be more difficult for an author to get work published. Although, it could be argued that as the number of authors in a location increases, it become more difficult to innovate as authors becomes 'locked-in' to established methods and ways of thinking. Without more detailed data on publishing capacity, it is not possible to disentangle these effects.

In contrast, having more social connections (higher degree) more directly captures that an author has greater access to knowledge flowing through the network (e.g. learning about trends and style preferences). An author with high degree may be better able to call on social contacts and take advantage of resources of the network as a whole – and have a

greater chance of being introduced to an important publisher or literary critic. An author with high degree may have greater influence, prestige, and reputation within the network (making an author very valuable to publishers who may use her to actively recruit other authors to publish with the agency), or the author may have more connections with whom to collaborate or simply to read and edit drafts.

Degree is a measure of social connectedness over an author's entire lifetime, while the peer variables are determined on an annual basis. Authors are not necessarily residing in the same location as their social connections but, by definition, are residing in the same location as their peers. Therefore, higher degree does not necessarily result in an increase in competition for limited resources in a location. This may explain the contrasting results for peer group size and degree. Furthermore, it is possible that even a static measure of social connectedness may be a better proxy for social interaction than geographic proximity – and that studies of author productivity are highly sensitive to the types of proxies for social interaction. This is a key observation, as it suggests that proxying interaction with location may result in misleading or incorrect conclusions about the role of fellow authors in production.

4.5.3 Sensitivity Analysis: Lifetime Accomplishments

The above results may be sensitive to the definition of output, so I determine whether social network measures still display greater predictive power when other measures of output are used. Specifically, I look at two measures of lifetime accomplishments as alternative definitions of output: total number of works produced and Impact Index score. While the use of the total works produced in an author's lifetime is an intuitive measure of lifetime output, the use of an author's Impact Index score requires some explanation. Borowiecki (2013) argues that similarly-defined measure of 'quality' (i.e. Murray Index score) can be used as an alternative measure of output for Western classical music composers. Because both indices are based on the amount of contemporary critical attention an artist receives,

they reflect both lifetime output *and* accomplishment (in terms of quality, prestige, and lasting influence).

I also consider two measures of London: a binary variable equal to 1 if an author ever lived in London and zero otherwise and the total years an author spent in London. The results from Chapter 2 suggest that authors receive an immediate gain from moving to London that does not accumulate over time. However, the effect of living in London may impact the quality of work rather than the quantity of work, and such gains may accumulate over time. Thus, I consider the predictive power of both measures of a London effect.

I use the repeated random sub-sampling validation procedure used in Section 4.5.1 to understand the predictive power of models with social network information relative to more parsimonious models. I partition the data into samples. I leave one sample out and estimate the model. I then test the extent to which this training sample can predict the sample that was left out (i.e. the test sample.) I begin with Model 1 (a parsimonious baseline model), which regresses lifespan, a measure of time spent in London, birth region dummies, and birth cohort (25 year period) on a measure of lifetime output:

$$\begin{aligned} \text{Model 1:} \quad LifetimeOutput_i = &\alpha + \delta Lifespan_i + \phi London_i + \lambda BirthLocation_i \\ &+ \gamma BirthCohort_i + \epsilon_i \end{aligned}$$

I then repeat this procedure for Model 2, which adds to Model 1 all of the networks variables simultaneously:

$$\begin{aligned} \text{Model 2:} \quad LifetimeOutput_i = &\alpha + \delta Lifespan_i + \phi London_i + \lambda BirthLocation_i \\ &+ \gamma BirthCohort_i + \beta_1 Degree_i + \beta_2 Betweenness_i \\ &+ \beta_3 Closeness_i + \beta_4 Clustering_i + \epsilon_i \end{aligned}$$

I consider four variations of Model 2. First, I utilise total works produced over lifetime as the measure of output. I also use (separately) two definitions of London: (1) a dummy variable equal to one if an author ever lived in London and zero otherwise and (2) the total years spent in London. Then, I utilise the Impact Index score as the measure of output. Again, I also consider two definitions of London. As in Section 4.4, I perform

200 replications of the forecasting procedure for each model. In Table 4.13, Column (1) presents the R^2 of the regression on the training data for each model, and Column (2) displays the the test RMSE for each model. I compare the RMSE of each model to the baseline model, reported as the RMSE differential in Column (3).

There is very little difference between the models that uses a binary definition of London and the models that uses duration in London. The R^2 values of Model 1 are comparable for both London specifications for the regressions using lifetime works as the measure of lifetime accomplishment and for the regressions using the Impact Index score as the measure of lifetime accomplishment. With lifetime works, the R^2 of both baseline models is 0.27. With the Impact Index score, the R^2 of both baseline models is 0.15. The RMSE differentials are quite similar as well. With lifetime works, the RMSE differential is 6.8% with the binary London indicator, and the RMSE differential 7.0% with duration in London. With the Impact Index score, the RMSE differential is 18.2% with the binary London indicator, and the RMSE differential 18.8% with duration in London.

There is no substantial difference in the predictive power between models using the different definitions of London. The results from Chapter 2 suggest locating in London causally increases the productivity of an author, though there is no evidence of learning (that results in increased output) from living in London for a longer period of time. I might expect to see different results when quality or influence of work is included in the measure of output. I suggested that, by living in London, authors may learn how to write better quality works though not necessarily produce a higher quantity of works.

The London results suggest this is not necessarily the case. Perhaps authors, like inventors, also benefit from enduring social capital (Agrawal et al., 2006). It is also possible that authors benefited from economies of scale by living in London due to infrastructure that was not present in other cities, even in Edinburgh with its large and competitive publishing industry. London was a larger, more culturally and intellectually diverse city throughout the time period. It is possible that authors benefited from this urban diversity, in line with the Jacobs (1969) theory of urbanisation. Without detailed data on book publishers and printers, it is not possible to disentangle these effects.

On the other hand, the addition of social network information does improve the predictive power of the models, regardless of which definition of London and which definition of lifetime accomplishment is used. Approximately 27% of the variation in total works produced over an author's lifetime can be explained by lifespan, birth location, and birth cohort. The addition of social network information increases the predictive power of the model by 6.8–7%, depending on the specification. In contrast, only roughly 15% of the variation in the Impact Index scores can be explained by the information contained in the baseline model. When social network information is added, the model explains roughly 40% of the variation in the Impact Index scores, and there is a greater increase in predictive power of the model (18.2–18.8%, depending on the specification).

I also provide the OLS estimates of the effect of social network variables on lifetime accomplishment using the following specification:

$$\begin{aligned} LifetimeAccomplishment_i = &\alpha + \delta Lifespan_i + \phi London_i + \lambda BirthLocation_i \\ &+ \gamma BirthCohort_i + \beta_1 Degree_i + \beta_2 Betweenness_i \\ &+ \beta_3 Closeness_i + \beta_4 Clustering_i + \epsilon_i \end{aligned}$$

I provide OLS estimates of all four variations of the above equation. Table 4.14 provides the coefficients when output is defined as total works produced over lifetime. Table 4.15 provides results when output is defined as impact index score. In both tables, the regressions presented in Column (1) utilise a binary variable equal to 1 if an author ever lived in London and zero otherwise, and the specifications in Column (2) uses total years in London.

The results for total works produced over an author's lifetime are consistent with the findings in Section 4.5.2. Higher lifespan and degree are associated with higher production of works. One additional year of life is correlated to roughly 0.6 additional works, and one additional social connection is correlated to just under two additional works. These results are not highly sensitive to the London measure used, both in terms of magnitude and statistical significance of coefficients and in terms of explanatory power given by the R^2.

Neither measure of London has a statistically significant correlation with lifetime output. There is no statistically significant difference in lifetime production if an author spends any amount of time in London compared to those who spend no time in London, and the duration in London has no statistically significant impact on lifetime production either.

The results when an author's Impact Index score is used to capture lifetime accomplishment show a more complex picture. This suggests that measures of output should incorporate quality or influence of work and that a failure to do so will result in a loss of important information on the dynamics underlying creative production.

Lifespan does not have a statistically significant relationship with the Impact Index score, and the coefficients of lifespan are negative. While authors continue to produce new work as they get older, this does not necessarily lead to greater prestige and lasting influence. While some authors may produce their greatest works very late in life, it is also possible that a tragic or early death may increase intrigue over an author's life and works. Thus, the relationship between lifespan and Impact Index is not straightforward, and multiple and opposing factors can result in similar amounts of critical attention.

While neither measure of London is statistically significant, there are some key differences between the results. Clustering does not have a statistically significant relationship with the Impact Index score when the binary London measure is used. However, when duration in London is controlled for, a higher degree of clustering – or a higher degree to which one's connections are also interconnected – the lower the Impact Index score. Having a higher degree coefficient is also associated with a higher Impact Index score, as is having a higher closeness coefficient.

It is possible that the positive correlation between the Impact Index and degree exists by construction: the longer the encyclopaedia entry, the more social connections are mentioned. In order to confirm that the construction of the variable is not driving this result, I perform the same analysis with only the citation component of the Impact Index (the number of times an author is cited in contemporary literary criticism and academic journal

articles). The Monte Carlo cross-validation results are presented in Table 4.16, and Table 4.17 provides the results when output is defined as the citation index score.

When using the Citation Index as a definition of lifetime accomplishment, the RMSE differential is 13.24% when London is defined as a binary indicator equal to 1 if an author has ever lived in London and zero otherwise, and the RMSE differential is 13.96% when London is defined as the total years an author spent in London. Thus, the Citation Index has greater predictive power than lifetime output, but the composite Impact Index which also includes a measure of total word count of the encyclopaedia entries for each author has greater predictive power than both the Citation Index and lifetime output.

More importantly, the strong correlation between degree and lifetime impact does not disappear when the Citation Index is used instead of the composite Impact Index, indicating that the result is not being driving by potential correlation between length of encyclopaedia entry and number of social connections listed within an each entry. The strength of the relationship between degree and lifetime impact does decrease slightly when the Citation Index is used compared to Impact Index – 1.381 and 1.385 for the Impact Index for binary London measure and total year in London, respectively, versus 1.331 and 1.307 for the Citation Index, but there is greater difference in the significance and magnitude of the relationships between clustering and lifetime output and between closeness and lifetime output. Similar differences were observed when using lifetime output compared to the Impact Index, so it is likely the differences are a further reflection of how sensitive clustering and closeness are to different measures in lifetime accomplishment.

Further work needs to be done to collect the appropriate data needed to establish a causal relationship, though it is possible to hypothesise about how social network connectedness increases both quantity of output and quality and influence of output. Having many connections who are not too interconnected themselves and also being positioned to efficiently gain information from others results in better quality of output through the sharing of ideas and resources. Being part of a group that is 'too interconnected' may result in getting 'locked-in' to established ways of thinking, making it harder for all individuals in the group to produce innovative work.

The duration of the writing process may be reduced through efficiencies gained from collaborative relationships (i.e. authors reviewing and editing one another's drafts), which would increase the rate at which works could be produced. Having central or prestigious connections may also increase favourable access to publishers and critics, which might increase an author's ability to get works published and receive favourable reviews for new works. Being well positioned within the author network may also make it easier to establish more or better collaborative relationships. As noted by Farrell (2001), collaborative relationships like those of the Inklings or between Joseph Conrad and Ford Madox Ford resulted in better quality works.

However, it is also possible that greater contemporary attention is given to authors associated with the status quo. In this case, measures of author centrality reflect strength of ties to others of historic importance and prestige. Being linked to other important figures will influence the contemporary assessments of the historical importance and quality of an author. Individuals outside the 'status quo' will not receive the same degree of contemporary critical attention due to their 'fringe' status, even if their work is of equal quality. In addition, authors likely self-select into groups. Higher quality authors may be better able to have more contacts or better quality contacts or position themselves strategically within the network to control information flows. Thus, an author's position in the network is a function of own-quality, which is also captured by the Impact Index.

4.6 Conclusion

In this chapter, I use repeated random sub-sampling validation (i.e. Monte-Carlo cross-validation) procedures to quantify the increase in predictive power of models containing social network information beyond that contained in (1) a standard age-productivity profile and (2) an age-productivity profile with peer effects. I find that location-based peer effects provide a 0.009% increase in predictive power relative to a standard age-productivity profile when estimating annual productivity in terms of number of works published. The addition

of social network information results in a 0.86% to 1.79% increases in predictive power, depending on the model specification.

Greater gains are observed when looking at lifetime accomplishments rather than annual productivity. When using total works produced in an author's lifetime as the measure of lifetime accomplishment, the addition of social network information leads to 6.8% – 7.0% increases in predictive power, depending on the specification. When the Impact Index score (a measure based on the amount of contemporary critical attention an artist receives, reflecting both lifetime output and accomplishment in terms of quality, prestige, and lasting influence) is used as the measure of lifetime accomplishment, social network information leads to gains of 18.2% – 18.8%, depending on the specification.

I also determine which of the network variables are informative and provide OLS estimates of each model to illustrate the way social network information can better explain creative and cultural production. I also identify relationships that deserve more critical analysis in a causal framework. For example, the results also suggest that the gains from interacting with peers may not occur between those who are physically proximate but instead occur between certain types of social connections. Future studies determining a causal relationship between social network centrality and productivity should incorporate factors such as an author's associated literary groups (e.g. the Inklings), an author's primary genre of writing or the genre of each work, and dynamic network measures to better understand the kinds of social connections that are most important in the production process and to understand the boundaries between 'social space' and 'idea space' for authors.

Being connected to many people is likely to be beneficial, unless those people are highly interconnected. Being a part of a group that is 'too interconnected' can result in getting 'locked-in' to established ways of thinking, making it harder for all individuals in the group to produce innovative work. This should be tested explicitly in a causal framework. Furthermore, the results suggest that measures of output should incorporate quality or influence of work and that a failure to do so will result in a loss of important information on the dynamics underlying creative production.

These relationships deserve more critical analysis, although endogeneity presents a challenge to establishing a causal relationship. The results discussed above provide tentative evidence of the types of relationships that may exist and be of importance, and the results provide greater insight into the importance of how output is defined and the implications this may have for detecting spillovers from peer interactions. The results also show the potential of social network measures to reveal the mechanisms underlying creative production. The advantages of social network information – in terms of predictive power and intuitive explanation – are substantial, even in light of the fact that the collection of social network information is often more time and resource intensive compared to more general biographical information. The expanded use of social network information could greatly contribute to the empirical and theoretical literature on creative and cultural production beyond what is possible through more traditional peer effect measures.

4.A Tables and Figures

4.A.1 Figures

Figure 4.1: Graph of Author Network

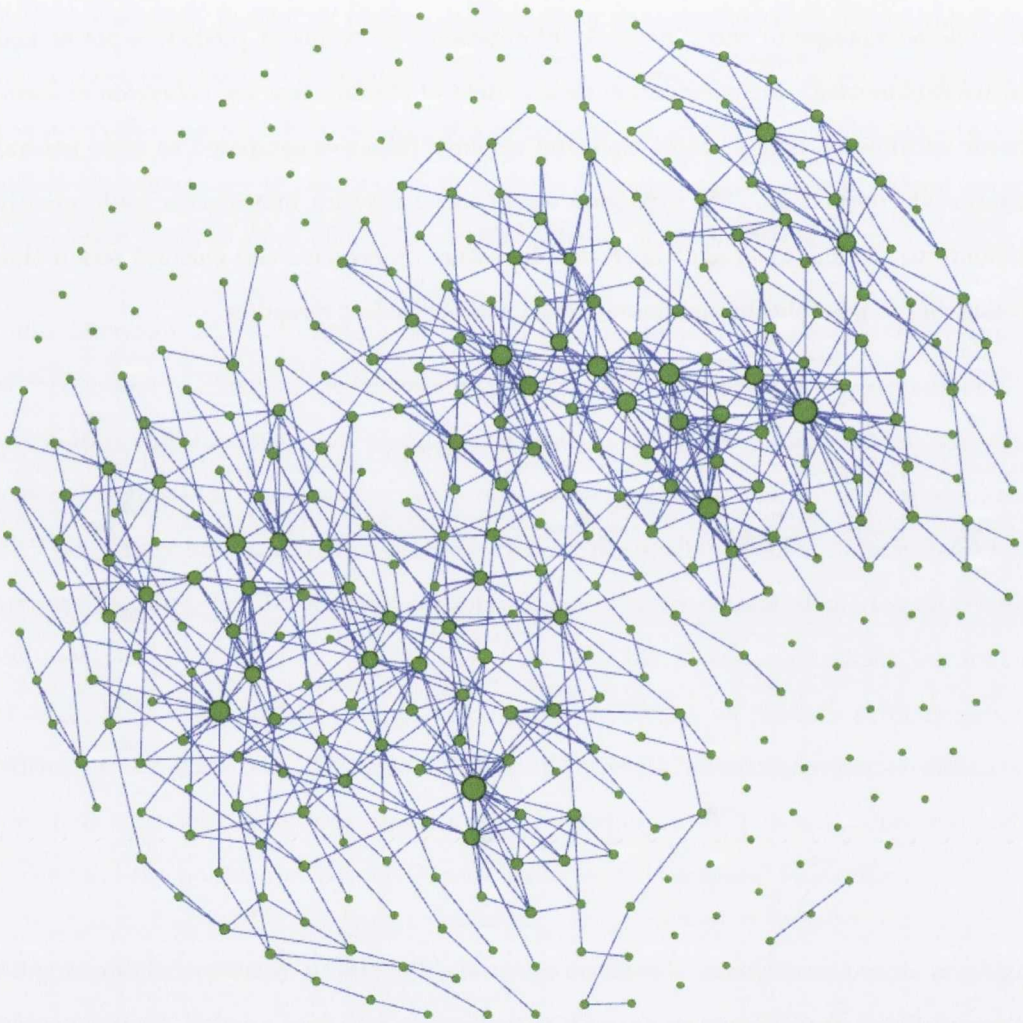


Figure 4.2: Degree Distribution

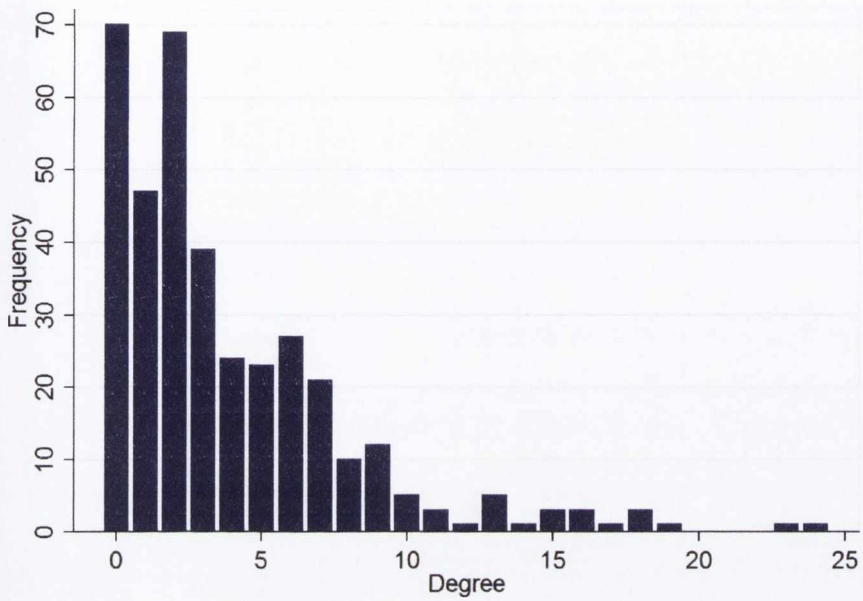


Figure 4.3: Sir Walter Scott – Local Network

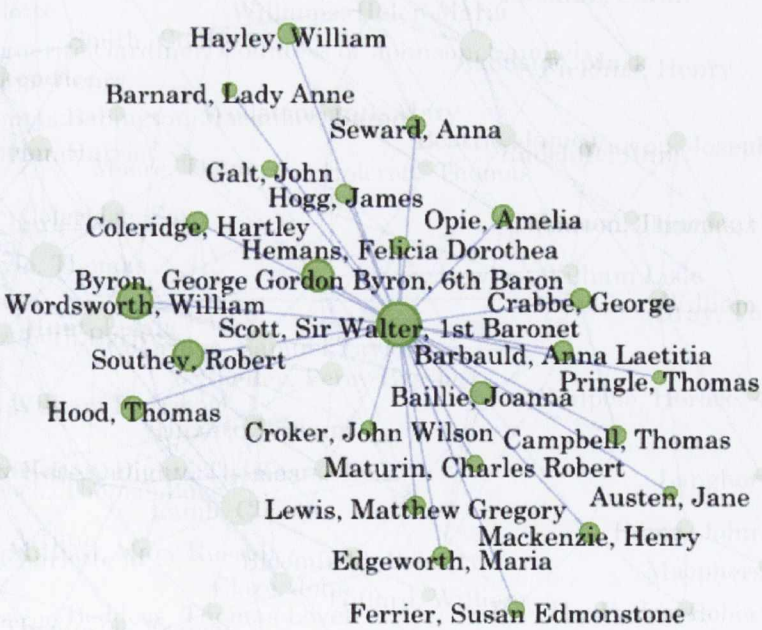


Figure 4.4: Virginia Woolf – Local Network

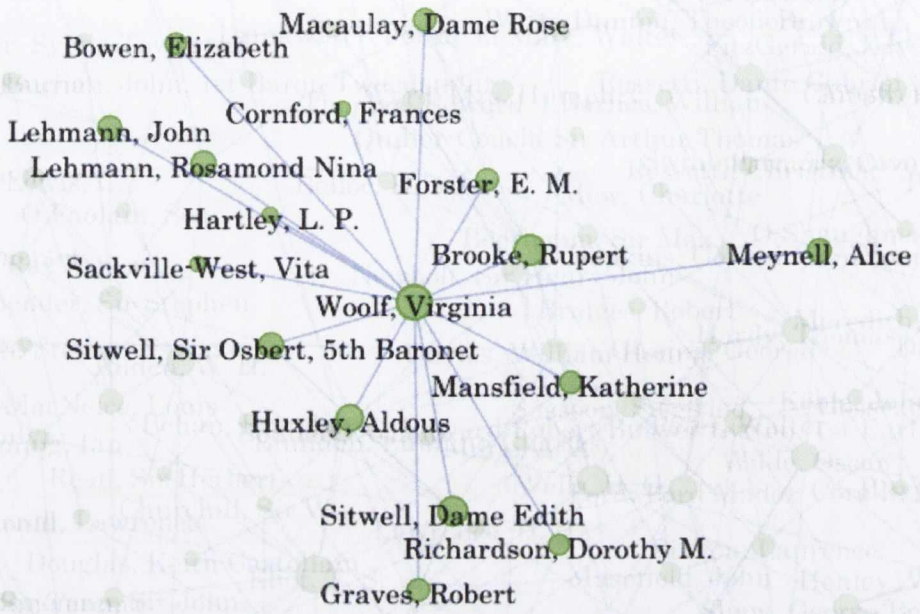


Figure 4.5: William Butler Yeats – Local Network

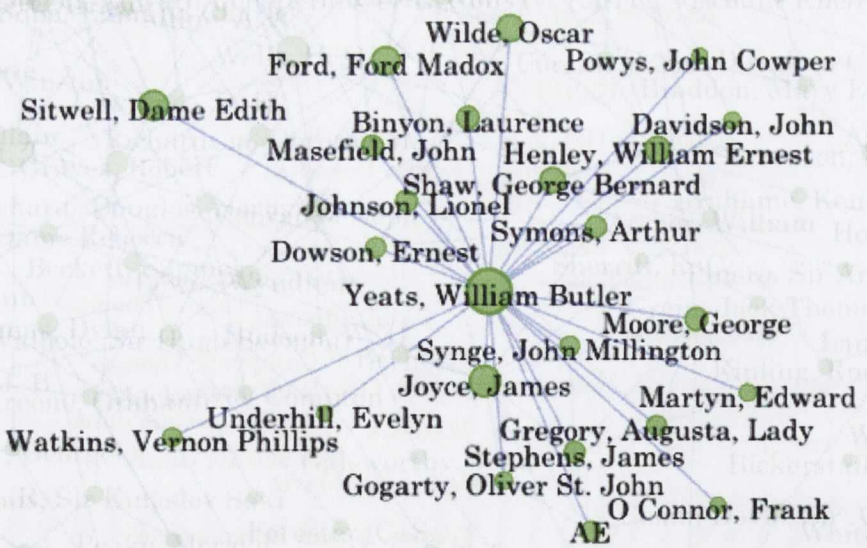
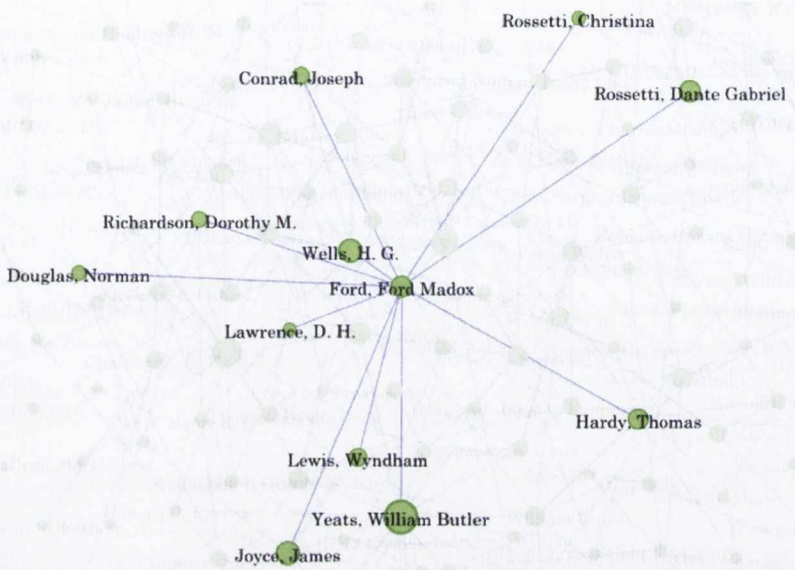


Figure 4.6: An Example of Power in an Author Network

(a) Ford Madox Ford – Local Network



(b) Joseph Conrad – Local Network

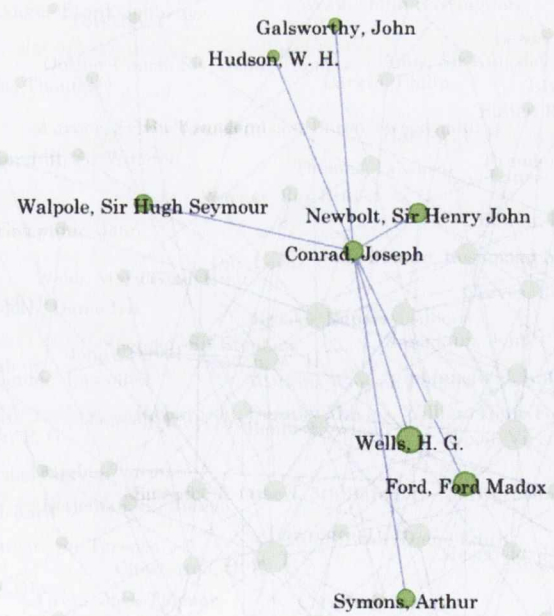


Figure 4.7: James Joyce – Local Network

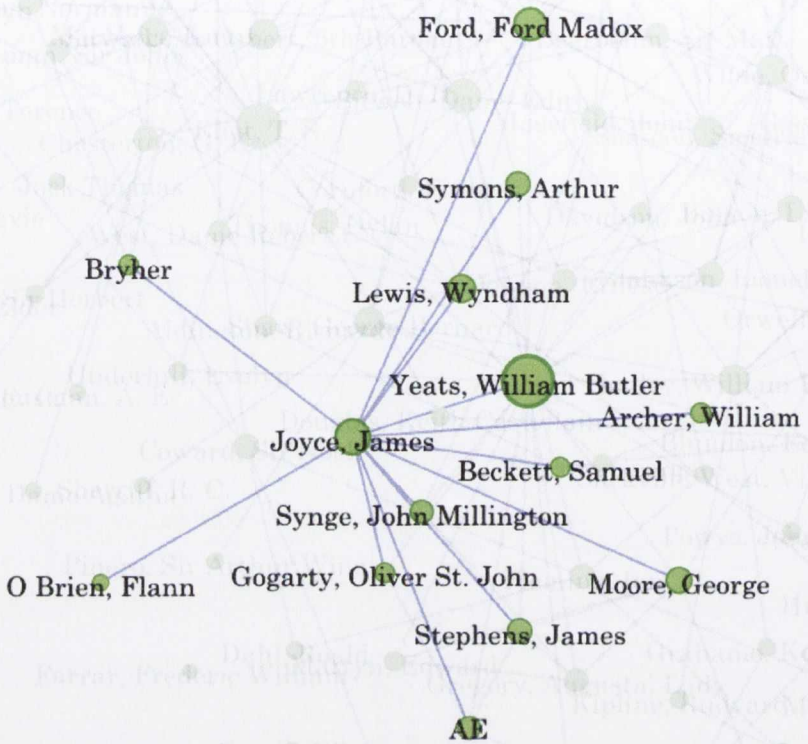
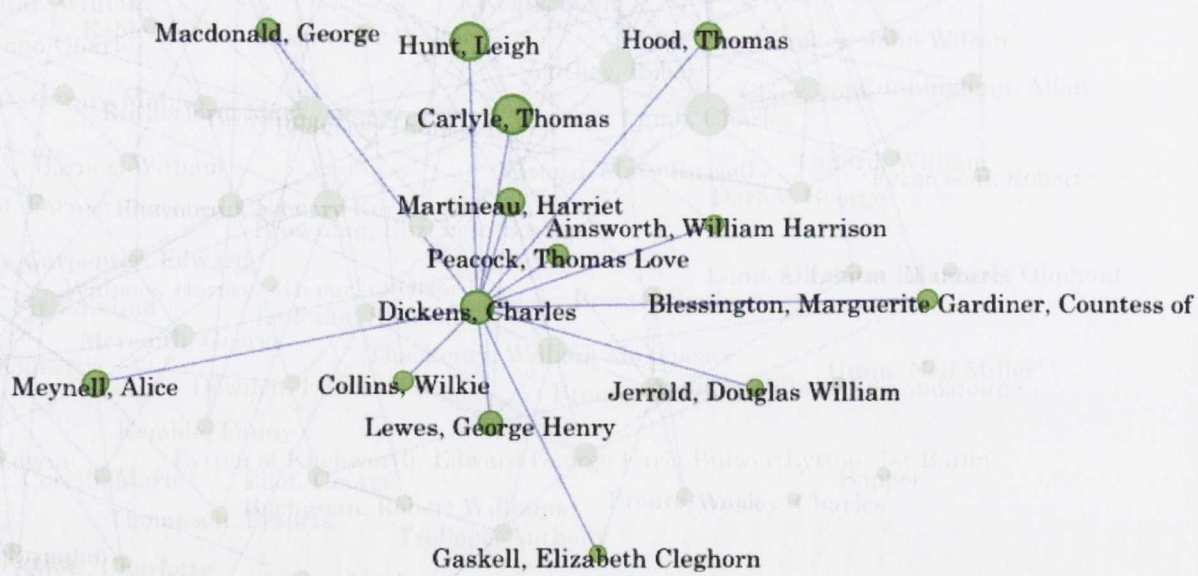


Figure 4.8: Charles Dickens – Local Network



4.A.2 Tables

Table 4.1: Author Summary Statistics

Gender		
Male	300	
Female	70	
Total	370	
Place of Birth		
England	245	
Ireland	44	
Scotland	41	
Wales	10	
Rest of Europe	11	
Rest of World	18	
Unknown	1	
Century of birth		
18^{th} Century	102	
19^{th} Century	207	
20^{th} Century	61	
Summary of Lifecycle	Mean	Std. Dev.
Lifespan	66.45	16.28
Years in London	21.48	20.76
Lifetime Works	33.24	32.24
Career Length	39.64	17.24
Works per Annum	0.88	0.76

Table 4.2: Network Summary: Connection and Distance

Nodes	370
Edges	707
Isolates	70
Paths	44253
Diameter	13
Average Shortest Path	4.9137
Density	0.0104
Degree Centralization	0.0550
Average Clustering Coefficient	0.1639
Overall Clustering Coefficient	0.1946

Table 4.3: Network Summary: Centrality

	Median	**Mean**	**Std. Dev.**	**Min**	**Max**
Degree	2	3.8216	4.0471	0	24
Betweenness	0.0016	0.0138	0.0270	0	0.1670
Closeness	0.2473	0.2095	0.1102	0	0.3518

Table 4.4: Top 10 Authors by Impact Index

Author	**Impact Index**
Joyce, James	100
Dickens, Charles	75.47
Lawrence, D. H.	56.39
Conrad, Joseph	55.07
Woolf, Virginia	52.61
Blake, William	49.29
Wordsworth, William	48.86
Coleridge, Samuel Taylor	48.81
Austen, Jane	47.20
Eliot, T. S.	45.46

Table 4.5: Degree – Top 10 Authors

Author	Degree
Scott, Sir Walter, 1st Baronet	24
Yeats, William Butler	23
Lamb, Charles	19
Alfred, Lord Tennyson	18
Eliot, T. S.	18
Wordsworth, William	18
Carlyle, Thomas	17
Garrick, David	16
Hunt, Leigh	16
Woolf, Virginia	16

Table 4.6: Betweenness – Top 10 Authors

Author	Betweenness
Ford, Ford Madox	0.1670
Rossetti, Dante Gabriel	0.1530
Hunt, Leigh	0.1440
Wordsworth, William	0.1311
Arnold, Matthew	0.1287
Woolf, Virginia	0.1272
Gosse, Sir Edmund	0.1269
Scott, Sir Walter, 1st Baronet	0.1226
Yeats, William Butler	0.1156
Meynell, Alice	0.1133

Table 4.7: Closeness – Top 10 Authors

Author	Closeness
Ford, Ford Madox	0.3528
Meynell, Alice	0.3518
Rossetti, Dante Gabriel	0.3508
Browning, Robert	0.3501
Patmore, Coventry Kersey Dighton	0.3488
Hunt, Leigh	0.3462
Gosse, Sir Edmund	0.3449
Woolf, Virginia	0.3398
Arnold, Matthew	0.3395
Hardy, Thomas	0.3376

Table 4.8: Clustering – Top 10 Authors

Author	Clustering
Abercrombie, Lascelles	1
Brontë, Anne	1
Clough, Arthur Hugh	1
Cornford, Frances	1
Croker, John Wilson	1
Dahl, Roald	1
Drinkwater, John	1
Fielding, Sarah	1
Galt, John	1
Gogarty, Oliver St. John	1

Table 4.9: Top 10 Authors by Lifetime Output

Author	No. Lifetime Works
Yonge, Charlotte M.	186
Kipling, Rudyard	165
Belloc, Hilaire	158
Wells, H. G.	145
Drinkwater, John	144
Auden, W. H.	138
Binyon, Laurence	135
Priestley, J. B.	131
Graves, Robert	131
Blunden, Edmund Charles	130

Table 4.10: Predictive Power of Social Networks

	R^2	RMSE	RMSE Differential	Coefficients
Model 1				
Age	0.1150	1.1610	–	0.077***
Age^2				-0.0007***
Model 2				
London	0.1172	1.1609	0.009%	0.425***
Peer Quality				0.006*
Peer $Quality^2$				-0.0001
No. Peers				-0.030***
No. Peers2				0.0005***
Peer Output				0.020
Model 3				
Betweenness	0.1259	1.1503	0.922%	3.871***
Closeness	0.1252	1.1510	0.861%	0.926***
Clustering	0.1172	1.1600	0.086%	0.0418
Degree	0.1330	1.1402	1.792%	0.035***
Model 4				
Betweenness	0.1338	1.1418	1.654%	-0.734
Closeness				0.252***
Clustering				-0.108***
Degree				0.036***

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Model 1 includes time dummies. Model 2 adds to model 1 peer variables and a London fixed effect. Model 3 adds to model 2 one of the network variables. Model 4 adds to model 2 all of the networks variables, simultaneously. We perform 200 replications of the forecasting procedure for each of the four models. The number of authors is 370. The number of observations is 19,023.

Table 4.11: Effect of Network Centrality on Productivity

	(1) Betweenness	(2) Closeness	(3) Clustering	(4) Degree	(5) Pooled OLS	(6) Pooled OLS with London FE	(7) Pooled OLS with Peer Effects
Age	0.0766***	0.0773***	0.0771***	0.0767***	0.0770***	0.0767***	0.0759***
	(0.00498)	(0.00503)	(0.00498)	(0.00496)	(0.00496)	(0.00504)	(0.00495)
Age^2	-0.000731***	-0.000738***	-0.000732***	-0.000734***	-0.000738***	-0.000735***	-0.000729***
	(0.0000499)	(0.0000505)	(0.0000499)	(0.0000497)	(0.0000496)	(0.0000501)	(0.0000493)
Betweeenness	3.880***				-0.734	-0.739	-0.734
	(0.934)				(1.411)	(1.412)	(1.414)
Closeness		0.940***			0.284	0.285	0.252
		(0.227)			(0.323)	(0.323)	(0.321)
Clustering			0.0341		-0.120	-0.117	-0.108
			(0.129)		(0.140)	(0.139)	(0.139)
Degree				0.0352***	0.0356**	0.0355**	0.0360***
				(0.00708)	(0.0138)	(0.0138)	(0.0137)
London						0.0183	0.395***
						(0.0497)	(0.151)
No. Peers							-0.0298***
							(0.0105)
$No. Peers^2$							0.000493***
							(0.000184)
Peer Quality							0.00264
							(0.00589)
$Peer Quality^2$							-0.0000780
							(0.000116)
Peer Output							0.0209
							(0.0541)
Constant	-1.234***	-1.332***	-1.177***	-1.333***	-1.358***	-1.365***	-1.553***
	(0.104)	(0.123)	(0.106)	(0.116)	(0.127)	(0.128)	(0.186)
Time Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.112	0.111	0.103	0.119	0.120	0.120	0.122
No. Authors	370	370	370	370	370	370	370
Observations	19023	19023	19023	19023	19023	19023	19023

Robust standard errors are clustered on the author level and are reported in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.12: Correlation Table of Network Variables

	Betweenness	Closeness	Clustering	Degree
Betweenness	1.0000			
Closeness	0.4444	1.0000		
Clustering	-0.0120	0.3280	1.0000	
Degree	0.7774	0.6304	0.2060	1.0000

Table 4.13: Predictive Power: Lifetime Accomplishments

	R^2	RMSE	RMSE Differential
Lifetime Output			
Ever lived in London			
Model 1	0.2745	944.9752	–
Model 2	0.3283	880.7693	6.79%
Total years in London			
Model 1	0.2748	947.1047	–
Model 2	0.3306	880.4398	7.04%
Impact Index			
Ever lived in London			
Model 1	0.1463	140.6225	–
Model 2	0.4025	115.0301	18.20%
Total years in London			
Model 1	0.1457	140.5945	–
Model 2	0.4045	114.1815	18.79%

Table 4.14: Lifetime Accomplishment: Lifetime Output

	(1) Ever Lived in London	(2) Total Years in London
Lifespan	0.591***	0.637***
	(0.0868)	(0.0973)
London	3.746	-0.110
	(3.347)	(0.0961)
Betweeenness	-31.30	-32.82
	(73.20)	(72.72)
Clustering	-3.417	-4.150
	(7.554)	(7.645)
Degree	1.838**	1.907***
	(0.739)	(0.725)
Closeness	8.025	7.121
	(18.25)	(18.28)
Constant	-29.20***	-25.41**
	(10.65)	(10.19)
Birth Cohort Dummies	Yes	Yes
Birth Region Dummies	Yes	Yes
R^2	0.260	0.261
No. Authors	370	370
Observations	370	370

Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.15: Lifetime Accomplishment: Impact Index

	(1) Ever Lived in London	(2) Total Years in London
Lifespan	-0.0329	-0.0206
	(0.0364)	(0.0377)
London	-0.478	-0.0443
	(1.839)	(0.0309)
Betweenness	-3.804	-3.174
	(36.94)	(36.33)
Clustering	-4.097	-4.196*
	(2.502)	(2.437)
Degree	1.381***	1.385***
	(0.380)	(0.371)
Closeness	13.80**	13.82**
	(6.392)	(6.364)
Constant	9.967*	10.02**
	(5.143)	(5.028)
Birth Cohort Dummies	Yes	Yes
Birth Region Dummies	Yes	Yes
R^2	0.337	0.340
No. Authors	370	370
Observations	370	370

Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.16: Predictive Power: Citation Index
as a Measure of Lifetime Accomplishment

	R^2	RMSE	RMSE Differential
Citation Index			
Ever lived in London			
Model 1	0.1369	93.5450	–
Model 2	0.3241	82.6079	13.24%
Total years in London			
Model 1	0.1307	93.5542	–
Model 2	0.3158	82.8235	13.96%

Table 4.17: Lifetime Accomplishment: Citation Index

	(1) Ever Lived in London	(2) Total Years in London
Lifespan	-0.0496*	-0.0476
	(0.0297)	(0.0307)
London	-2.604	-0.0331
	(1.797)	(0.0265)
Betweeenness	-37.17	-35.30
	(34.94)	(34.36)
Clustering	-5.192**	-5.043**
	(2.136)	(2.128)
Degree	1.331***	1.307***
	(0.420)	(0.412)
Closeness	-2.377	-1.934
	(5.725)	(5.596)
Constant	3.658	2.024
	(3.214)	(2.801)
Birth Cohort Dummies	Yes	Yes
Birth Region Dummies	Yes	Yes
R^2	0.257	0.251
No. Authors	370	370
Observations	370	370

Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 5

Conclusion

While the agglomeration of industry and the mechanisms driving the resulting gains are evident for industries with traditional modes of production, less is known about these dynamics for the creative and cultural industries. Although the focus of this dissertation is on literary artists, the broader aim is to provide a greater understanding of the circumstances that promote creative and cultural production. Literary artists are unique in that other types of artists, such as visual artists, composers, dancers, etc., have greater physical capital requirements and require more formal training. Therefore, the writing of literature presents a particularly interesting case.

Due to this unique nature of literary production, it is possible that authors do not behave like workers in other industries. In the introduction, I presented a series of questions that highlighted pertinent issues: where do authors chose to live and work? Do authors tend to cluster geographically? Do authors benefit from literary agglomeration? Do authors become more productive from proximity to other authors? Do authors become more productive from due to certain types of social interactions? In order to address these questions, I created a unique panel dataset on 370 prominent authors and combined this with theory and tools from economic history, urban economics, labour economics, and social network analysis.

Chapter 2 begins by providing historical context and discussing the methodology used to collect the dataset used in this analysis. I then analyse the patterns of migration and clustering of authors and found that the spatial distribution of authors differs significantly from the spatial distribution of the total population. London arises as the only major geographic cluster during the time period. Dublin, Edinburgh, Oxford, and Cambridge emerge as minor clusters but are home to consistently fewer than 5 authors in a given year. Large cities, such as Manchester and Liverpool, do not see the consistent co-location of authors.

I construct age-productivity profiles to estimate the productivity gains associated with the geographic clustering of literary activity in London. I find that authors who reside in London experience productivity gains of 10.68% per annum compared to when living elsewhere. These results are not driven by self-selection of recently successful authors in migrating to London. This increase in productivity is primarily due to a static agglomeration effect and not due to a dynamic learning effect. There is also no evidence that authors strategically withhold writings until migrating to London and then 'dump' them on the market.

In Chapter 3, I explore one possible channel through which these gains may occur – peer effects. I define peers as authors living in the same city in the same year, and I construct age-productivity profiles to determine the productivity gains (if any) associated with the mean quality of peers, mean peer output, and peer group size. Individual fixed effects are used to control for spatial and group sorting. I explore a number of alternate specifications, including estimating peer effects within London and determining whether certain types of peers (age of peers, living with a superstar) has a greater effect on productivity.

There is modest evidence of general peer effects through the channel of peer group size but stronger evidence of highly-localised peer effects. The within estimator reveals that peer group size has a modest, negative effect on productivity, and these results are robust to several alternate specifications. When concentration of authors in London is controlled

for, mean author quality while residing in London has a negative effect on annual productivity. These results are primarily driven by older authors, although there are substantial heterogeneity of returns by own-author quality.

The effect of peer group size is likely due to competition – publication is a zero-sum game and higher quality authors are likely to have better connections or be given preferential treatment by publishers. It is also possible that publishers were able to identify and give preferential treatment to high-quality authors or that high-quality authors are better at making high-quality social connections or better at navigating the publishing process. The decrease in productivity could also be due to learning from peers that results in increased *quality* of work with the trade-off of reduced *quantity* of work.

While these results did not clearly identify a channel through which agglomeration effects occur, they were still informative. These results indicate that cities are not optimal level of agglomeration for detecting social interaction effects, and more localised location information should be collected when possible. The results also suggest that that geographic proximity is not sufficient condition for peer effects to occur. Spillovers may be more likely to occur in 'idea space' than in 'physical space'. Though, it is unclear how an 'idea space' should be defined for authors – and if there is an observable difference between 'idea space' and 'social space' for the creative industries.

Anecdotal evidence suggests that spillovers between authors no only occur across creative industries or generally in a geographic area but also within social networks. I argue that social network information may provide greater explanatory power compared to broadly-defined peer effects and provides greater insight into the conditions in which innovation and creative production thrive. However, the cost of collecting social network information is often time and resource intensive, particularly when studying historical figures. Thus, the application of social networks in studies of creative and cultural production have been limited, as researchers opt for data that can be more efficiently collected (e.g. basic biographical information such as birth and death location, years of birth and death, total works produced, etc.).

Chapter 4 quantifies the gains that may be realised through the use of social network information. I combine a new, unique dataset of the social connections with the existing dataset with location, biographical and peer data. I use repeated random sub-sampling validation (i.e. Monte-Carlo cross-validation) procedures to quantify the increase in predictive power of models containing social network information beyond that contained in (1) a standard age-productivity profile and (2) an age-productivity profile with peer effects. I find that location-based peer effects provide a 0.009% increase in predictive power relative to a standard age-productivity profile when estimating annual productivity in terms of number of works published. The addition of social network information results in a 0.86% to 1.79% increases in predictive power, depending on the model specification.

Greater gains are observed when looking at lifetime accomplishments rather than annual productivity. When using total works produced in an author's lifetime as the measure of lifetime accomplishment, the addition of social network information leads to 6.8% – 7.0% increases in predictive power, depending on the specification. When the Impact Index score (a measure based on the amount of contemporary critical attention an artist receives, reflecting both lifetime output and accomplishment in terms of quality, prestige, and lasting influence) is used as the measure of lifetime accomplishment, social network information leads to gains of 18.2% – 18.8%, depending on the specification.

I also determine which of the network variables are informative and provided OLS estimates of each model to illustrate the way social network information can better explain creative and cultural production. I use to this to identify relationships that deserve more critical analysis in a causal framework. For example, the results suggest that the gains from interacting with peers may not occur between those who are physically proximate but instead occur between certain types of social connections. Future studies determining a causal relationship between social network centrality and productivity should incorporate factors such as an author's associated literary groups (e.g. the Inklings), an author's primary genre of writing or the genre of each work, and dynamic network measures to better understand the kinds of social connections that are most important in the production process and to understand the boundaries between 'social space' and 'idea space' for

authors.

The results also suggest that being connected to many people is beneficial, unless those people are highly interconnected. Being a part of a group that is 'too interconnected' can result in getting 'locked-in' to established ways of thinking, making it harder for all individuals in the group to produce innovative work. This should be tested explicitly in a causal framework. Furthermore, the results suggest that measures of output should incorporate quality or influence of work and that a failure to do so will result in a loss of important information on the dynamics underlying creative production.

The social interactions between creative workers in a geographically concentrated area likely played an important role in the transfer and diffusion of knowledge, as well as the generation of new knowledge and ideas. For example, authors in London may have had access to stronger and more advantageous social networks, in terms of increased connections with their peers (other authors), individuals with influence within the publishing industry (agents, publishers, critics), and those who are a part of the intellectual and cultural elite (artists, musicians, wealthy patrons). Authors in London also could have taken advantage of the related economic infrastructure and gained from the resulting economies of scale, allowing for a more efficient transformation of ideas into physical book-form.

However, further research is needed to fully understand channels through which authors receive productivity gains and the degree to which these mechanisms are geographically localised. The results also show the potential of social network measures to reveal the mechanisms underlying creative production. The advantages of social network information – in terms of predictive power and intuitive explanation – are substantial, even in light of the fact that the collection of social network information is often more time and resource intensive compared to more general biographical information. The expanded use of social network information could greatly contribute to the empirical and theoretical literature on creative and cultural production beyond what is possible through more traditional peer effect measures.

Due to the nature of the data used in this study, broad generalisations of these results are limited. However, the insights into the role of social interactions and the positive externalities associated with the agglomeration of an industry with few capital requirements are still of relevance to economic researchers, and the contributions to broader understanding of how to access and enhance knowledge spillovers are of interest to both firms and public policy-makers. For example, parallels may be drawn between the creative processes of the historical author and the contemporary software developer. As software development continues to shift from more tradition modes of production to production via remote work, it will be important to understand if (and why) agglomeration of IT firms persists and how individual productivity will be impacted. In this, historical analyses of literary production and other creative industries may be particularly useful.

Bibliography

Agrawal, A., Cockburn, I., and McHale, J. (2006). Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, 6(5):571–591.

Agrawal, A., Kapur, D., and McHale, J. (2008). How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics*, 64(2):258–269.

Agrawal, A. K., McHale, J., and Oettl, A. (2014). Why stars matter. Technical report, National Bureau of Economic Research.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3):425–455.

Aizer, A. and Currie, J. (2004). Networks or neighborhoods? Correlations in the use of publicly–funded maternity care in California. *Journal of Public Economics*, 88(12):2573–2585.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30:98–108.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Audretsch, B. (1998). Agglomeration and the location of innovative activity. *Oxford Review of Economic Policy*, 14(2):18–29.

Audretsch, D. B. and Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review*, 86(3):630–640.

Audretsch, D. B. and Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. In *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2713–2739. Elsevier.

Azoulay, P., Zivin, J. S., and Wang, J. (2010). Superstar extinction. *Quarterly Journal of Economics*, 125(2):549–589.

Barnes, J. J. and Barnes, P. P. (2004). Routledge, George (1812-1888). In *Oxford Dictionary of National Biography*. Oxford University Press.

Bearman, P. S., Moody, J., and Stovel, K. (2004). Chains of affection: The structure of adolescent romantic and sexual networks 1. *American Journal of Sociology*, 110(1):44–91.

Borowiecki, K. J. (2012). Are composers different? Historical evidence on conflict-induced migration (1816–1997). *European Review of Economic History*, 16(3):270–291.

Borowiecki, K. J. (2013). Geographic clustering and productivity: An instrumental variable approach for classical composers. *Journal of Urban Economics*, 73(1):94–110.

Borowiecki, K. J. and O'Hagan, J. W. (2013). Impact of war on individual life-cycle creativity: Tentative evidence in relation to composers. *Journal of Cultural Economics*, 37(3):347–358.

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1):61–74.

Bramoullé, Y., Djebbari, H., and Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55.

Brewster, K. L. (1994). Race differences in sexual activity among adolescent women: The role of neighborhood characteristics. *American Sociological Review*, pages 408–424.

Carlino, G. and Kerr, W. R. (2015). Agglomeration and innovation. In Gilles Duranton, J. V. H. and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 349–404. Elsevier.

Carlino, G. A., Chatterjee, S., and Hunt, R. M. (2007). Urban density and the rate of invention. *Journal of Urban Economics*, 61(3):389–419.

Carrell, S., Fullerton, R. L., and West, J. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.

Case, A. C. and Katz, L. F. (1991). The company you keep: The effects of family and neighborhood on disadvantaged youths. Technical report, National Bureau of Economic Research.

Chatterji, A., Glaeser, E. L., and Kerr, W. R. (2013). Clusters of entrepreneurship and innovation. Technical report, National Bureau of Economic Research.

Ciccone, A. and Hall, R. E. (1996). Productivity and the density of economic activity. *American Economic Review*, 86(1):54.

Cohen-Cole, E. and Fletcher, J. M. (2008). Detecting implausible social network effects in acne, height, and headaches: Longitudinal analysis. *BMJ: British Medical Journal*, 337.

Combes, P.-P., Duranton, G., and Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2):723–742.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Gilles Duranton, J. V. H. and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 247 – 348. Elsevier.

Conley, T. G. and Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *The American Economic Review*, 100(1):35–69.

Corcoran, M., Gordon, R., Laren, D., and Solon, G. (1992). The association between men's

economic status and their family and community origins. *Journal of Human Resources*, pages 575–601.

Crane, J. (1991). The epidemic theory of ghettos and neighborhood effects on dropping out and teenage childbearing. *American Journal of Sociology*, pages 1226–1259.

Datcher, L. (1982). Effects of community and family background on achievement. *The Review of Economics and Statistics*, pages 32–41.

Davezies, L., d'Haultfoeuille, X., and Fougère, D. (2009). Identification of peer effects using group size variation. *The Econometrics Journal*, 12(3):397–413.

de la Roca, J. and Puga, D. (2012). Learning by working in big cities. CEPR Discussion Papers 9243.

D'Este, P., Guy, F., and Iammarino, S. (2013). Shaping the formation of university – industry research collaborations: What type of proximity does really matter? *Journal of Economic Geography*, 13(4):537–558.

Ductor, L., Fafchamps, M., Goyal, S., and van der Leij, M. J. (2014). Social networks and research output. *Review of Economics and Statistics*, 96(5):936–948.

Duflo, E., Dupas, P., and Kremera, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *The American Economic Review*, 101(5):1739–1774.

Duranton, G. and Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2063–2117. Elsevier.

Durlauf, S. N. (2004). Neighborhood effects. *Handbook of Regional and Urban Economics*, 4:2173–2242.

Durlauf, S. N. and Ioannides, Y. M. (2010). Social interactions. *Annual Review of Economics*, 2(1):451–478.

Edunov, S., Diuk, C., Filiz, I. O., Bhagat, S., and Burke, M. (2016). Three and a half degrees of separation. `https://research.facebook.com/blog/three-and-a-half-degrees-of-separation`.

Encyclopaedia Britannica (2014). `www.britannica.com`.

Erickson, L. (2002). "Unboastful Bard": Originally anonymous English Romantic poetry book publication, 1770–1835. *New Literary History*, 33(2):247–278.

Farrell, M. P. (2001). *Collaborative Circles: Friendship dynamics and creative work*. The University of Chicago Press.

Galenson, D. W. (2002). Quantifying artistic success: Ranking French painters – and paintings – from Impressionism to Cubism. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 35(1):5–19.

Gertler, M. S. (2003). Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, 3(1):75–99.

Ginther, D., Haveman, R., and Wolfe, B. (2000). Neighborhood attributes as determinants of children's outcomes: How robust are the relationships? *Journal of Human Resources*, pages 603–642.

Giorcelli, M. and Moser, P. (2015). Copyright and creativity: Evidence from Italian operas. Working paper.

Glaeser, E. and Maré, D. (2001). Cities and skills. *Journal of Labor Economics*, 19(2):316–42.

Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., and Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6):1126–52.

Glaeser, E. L., Laibson, D., and Sacerdote, B. (2002). An economic approach to social capital. *The Economic Journal*, 112(483):F437–F458.

Goyal, S., Van Der Leij, M. J., and Moraga-González, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, 114(2):403–412.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1.

Hanneman, R. A. and Riddle, M. (2005). *Introduction to social network methods (e-book)*. University of California, Riverside.

Hellmanzik, C. (2010). Location matters: Estimating cluster premiums for prominent modern artists. *European Economic Review*, 54(2):199–218.

Hellmanzik, C. (2013). Democracy and economic outcomes: Evidence from the superstars of modern art. *European Journal of Political Economy*, 30:58–69.

Hellmanzik, C. (2016). Creative production and peer effects: Evidence from the exodus of superstar painters from Paris. *Working Paper*.

Helsley, R. W. and Strange, W. C. (2002). Innovation and input sharing. *Journal of Urban Economics*, 51(1):25–45.

Hewitt, D. (2004). Constable, Archibald (1774-1827). In *Oxford Dictionary of National Biography*. Oxford University Press.

Hogan, D. P. and Kitagawa, E. M. (1985). The impact of social status, family structure, and neighborhood on the fertility of black adolescents. *American Journal of Sociology*, pages 825–855.

Howells, J. (2012). The geography of knowledge: Never so close but never so far apart. *Journal of Economic Geography*, 12(5):1003–1020.

Imberman, S. A., Kugler, A. D., and Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *The American Economic Review*, 102(5):2048–2082.

Ioannides, Y. M. and Datcher Loury, L. (2004). Job information networks, neighborhood effects, and inequality. *Journal of economic literature*, 42(4):1056–1093.

Jackson, M. (2016). *The Oxford handbook of the economics of networks*, chapter The Past and Future of Network Analysis in Economics. Oxford University Press.

Jackson, M. O. (2010). *Social and Economic Networks.* Princeton University Press.

Jacobs, J. (1969). *The economy of cities.* Vintage international. Random House.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3):577–598.

Jones, B. F. (2009). The burden of knowledge and the "death of the Renaissance man": Is innovation getting harder? *Review of Economic Studies*, 76(1):283–317.

Kelly, E. and O'Hagan, J. (2007). Geographic clustering of economic activity: The case of prominent Western visual artists. *Journal of Cultural Economics*, 31(2):109–128.

Kim, E. H., Morse, A., and Zingales, L. (2009). Are elite universities losing their competitive edge? *Journal of Financial Economics*, 93(3):353 – 381.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499.

Lee, L.-f. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2):333–374.

Levin, S. G. and Stephan, P. E. (1991). Research productivity over the life cycle: Evidence for academic scientists. *American Economic Review*, pages 114–132.

Literature Online (2014). `www.lion.chadwyck.co.uk`.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, 14(3):115–136.

Marshall, A. (1890). *Principles of Economics.* Macmillan, London.

Mokyr, J. (2005). Mobility, creativity, and technological development: David Hume, Immanuel Kant and the economic development of Europe. *Northwestern University, unpublished manuscipt.*

Moser, P., Voena, A., and Waldinger, F. (2014). German Jewish émigrés and US invention. *American Economic Review*, 104(10):3222–3255.

Murray, C. A. (2003). *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BC to 1950*. HarperCollins New York.

National Archives of Ireland (2016). About pre-1901 census fragments. `http://www.census.nationalarchives.ie/help/pre1901.html`.

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5200–5205.

O'Hagan, J. and Borowiecki, K. J. (2010). Birth location, migration, and clustering of important composers: Historical patterns. *Historical Methods*, 43(2):81–90.

O'Hagan, J. and Kelly, E. (2005). Identifying the most important artists in a historical context: Methods used and initial results. *Historical Methods*, 38(3):118–125.

Online Historical Population Reports (2016). Census of Great Britain, 1851, Population tables, I. Number of the inhabitants in 1801, 1811, 1821, 1831, 1841 and 1851. Vol. I BPP 1852-53 LXXXV (1631), pages 68-69. `www.histpop.org`.

Plotnick, R. D. and Hoffman, S. D. (1999). The effect of neighborhood characteristics on young adult outcomes: Alternative estimates. *Social Science Quarterly*, pages 1–18.

Porter, M. E. (2011). *Competitive advantage of nations: Creating and sustaining superior performance*. Simon and Schuster.

Rallet, A. and Torre, A. (1998). On geography and technology: The case of proximity relations in localized innovation networks. In Steiner, M., editor, *Clusters and Regional Specialisation*. Pion Publication.

Rivkin, S. G. (2001). Tiebout sorting, aggregation and the estimation of peer group effects. *Economics of Education Review*, 20(3):201–209.

Rose, M. (1988). The author as proprietor: Donaldson v. Becket and the genealogy of modern authorship. *Representations*, (23):51–85.

Rosen, S. (1981). The economics of superstars. *American Economic Review*, pages 845–858.

Rosenthal, S. S. and Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In *Cities and Geography*, volume 4 of *Handbook of Regional and Urban Economics*, pages 2119 – 2171. Elsevier.

Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2):373–389.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics*, pages 681–704.

Shaffer, B. W. (2008). *A companion to the British and Irish novel 1945-2000*, volume 45. John Wiley & Sons.

Squicciarini, M. P. and Voigtländer, N. (2015). Human capital and industrialization: Evidence from the Age of Enlightenment. *The Quarterly Journal of Economics*, 130(4):1825–1883.

St Clair, W. (2004). *The Reading Nation in the Romantic Period*. Cambridge University Press.

Ter Wal, A. L. and Boschma, R. A. (2009). Applying social network analysis in economic geography: framing some key analytic issues. *The Annals of Regional Science*, 43(3):739–756.

The Literary Encyclopaedia (2014). `www.litencyc.com`.

Throsby, D. (1994). The production and consumption of the arts: A view of cultural economics. *Journal of Economic Literature*, 32(1):1–29.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *The Review of Economic Studies*, 68(2):261–295.

Topa, G. and Zenou, Y. (2015). Neighborhood and network effects. *Handbook of Regional and Urban Economics*, 5:561–624.

Torre, A. and Rallet, A. (2005). Proximity and localization. *Regional Studies*, 39(1):47–59.

Turley, R. N. L. (2003). When do neighborhoods matter? The role of race and neighborhood peers. *Social Science Research*, 32(1):61–79.

Ugander, J., Karrer, B., Backstrom, L., and Marlow, C. (2011). The anatomy of a facebook social graph. `https://arxiv.org/pdf/1111.4503.pdf`.

US Department of Justice and the Federal Trade Commission (2010). Horizontal merger guidelines. `https://www.justice.gov/atr/horizontal-merger-guidelines-08192010`.

Verboord, M. (2003). Classification of authors by literary prestige. *Poetics*, 31(3):259–281.

Von Hippel, E. (1994). "Sticky information" and the locus of problem solving: Implications for innovation. *Management Science*, 40(4):429–439.

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany. *Journal of Political Economy*, 118(4):787–831.

Waldinger, F. (2012a). Bombs, brains, and science: The role of human and physical capital for the creation of scientific knowledge. Technical report, Competitive Advantage in the Global Economy (CAGE).

Waldinger, F. (2012b). Peer effects in science: Evidence from the dismissal of scientists in Nazi Germany. *Review of Economic Studies*, 79(2):838–861.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press.

Weinberg, B., Reagan, P., and Yankow, J. (2002). Do neighborhoods affect hours worked? Evidence from longitudinal data. Technical report.

Wooldridge, J. M. (2010). Count, fractional, and other nonnegative responses. In *Econometric analysis of cross section and panel data*. MIT Press, 2 edition.

Xu, Q.-S., Liang, Y.-Z., and Du, Y.-P. (2004). Monte Carlo cross-validation for selecting a model and estimating the prediction error in multivariate calibration. *Journal of Chemometrics*, 18(2):112–120.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and statistics*, 85(1):9–23.