



## Durham E-Theses

---

# *Identities, Categories, and Clusters: A Study of Category Dynamics and Cluster Spanning in the Lebanese Newspaper Industry 1851-1974*

MOZAHM, NAJIB,ALI

### How to cite:

---

MOZAHM, NAJIB,ALI (2015) *Identities, Categories, and Clusters: A Study of Category Dynamics and Cluster Spanning in the Lebanese Newspaper Industry 1851-1974*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/11348/>

### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

Identities, Categories, and Clusters: A Study of  
Category Dynamics and Cluster Spanning in the  
Lebanese Newspaper Industry 1851-1974

Najib A. Mozahem

Durham Business School  
Durham University

2015

*Submitted in fulfillment of the degree of Doctorate in Business Administration*

# **Identities, Categories, and Clusters: A Study of Category Dynamics and Cluster Spanning in the Lebanese Newspaper Industry 1851-1974**

**Najib A. Mozahem**

## **Abstract**

Organizational ecologists have always argued for the need to take account of the whole industry while studying how certain characteristics might affect the mortality of organizations. They argued that concepts such as legitimacy and competition were the driving forces behind organizational dynamics. Recently, researchers have started to pay more attention to a usually ignored topic: audience perceptions. Legitimacy, for example, is no longer modeled as the number of organizations in a population. It is now thought to be dependent on how audience members perceive these organizations. By bringing audience members into the picture, the topic of identity has become of central importance. This thesis seeks to study how identities affect the life chances of organizations within the context of the Lebanese newspaper industry.

The first essay will study how the newspaper industry emerged. The essay will show that the newspaper industry in Lebanon started as a social movement. Social movements need to tell similar stories and justifications in order to be accepted by audience members. An examination of the stories told by the Lebanese newspapers during the period 1851 – 1879 will show that this was in fact the case.

The second essay studies the same period, but this time instead of looking at stories told by newspapers, this essay will build on the theoretic formulation of Hannan, Pólos, and Carroll (2007) and therefore will study that part of identity which is determined by the category spanning dynamics of organizations. The concept of cluster formation will also be introduced.

The third and final essay will apply the concept of cluster formation to the entire period under study (1851 – 1974). This essay will show that category spanning dynamics follow the pattern proposed by the Resource Partitioning theory, thereby questioning the legitimacy of the claim that category spanning is detrimental to organization survival. Instead, the essay will argue that cluster spanning increases the mortality hazard of organizations during all time periods.

# Table of Contents

<b>LIST OF FIGURES .....</b>	<b>III</b>
<b>LIST OF TABLES .....</b>	<b>IV</b>
<b>STATEMENT OF COPYRIGHT .....</b>	<b>V</b>
<b>ACKNOWLEDGMENTS.....</b>	<b>VI</b>
<b>1 CHAPTER ONE: INTRODUCTION .....</b>	<b>1</b>
1.1 INTRODUCTION.....	1
1.2 CATEGORY DYNAMICS .....	2
1.3 IDENTITY IN EMERGING MARKETS .....	5
1.4 EMPIRICAL SETTING .....	8
1.5 THESIS DIVISION AND THEORETICAL AIMS.....	9
1.6 CONCLUSION.....	11
<b>2 CHAPTER TWO: LITERATURE REVIEW AND DEVELOPMENT OF RESEARCH PROBLEMS .....</b>	<b>13</b>
2.1 INTRODUCTION.....	13
2.2 LITERATURE REVIEW – NICHE WIDTH .....	13
2.3 DEVELOPMENT OF RESEARCH PROBLEM – NICHE WIDTH.....	20
2.3.1 <i>Problems with the Distance-based Measures</i> .....	23
2.3.2 <i>Clusters: A Multi-level Approach</i> .....	29
2.3.3 <i>Revisiting Category Spanning: Formulation of Hypotheses</i> .....	36
2.4 LITERATURE REVIEW – INDUSTRY EMERGENCE.....	47
2.5 DEVELOPMENT OF RESEARCH PROBLEM – INDUSTRY EMERGENCE .....	50
2.6 CONCLUSION.....	56
<b>3 CHAPTER THREE: METHODOLOGY AND STATISTICAL MODELS .....</b>	<b>58</b>
3.1 INTRODUCTION.....	58
3.2 DATA COLLECTION AND CONTENT ANALYSIS.....	58
3.3 CLUSTER ANALYSIS.....	67
3.4 RANDOM SIMULATION EXPERIMENTS.....	71
3.5 MULTI-LEVEL LOGISTIC MODELS.....	72
3.6 SURVIVAL ANALYSIS .....	75
3.6.1 <i>Non-parametric Analysis</i> .....	76
3.6.2 <i>Semi-Parametric Analysis</i> .....	78
3.7 RESEARCH DESIGN.....	80
3.8 CONCLUSION.....	81
<b>4 CHAPTER FOUR: THE EMERGENCE OF THE INDUSTRY AS A SOCIAL MOVEMENT (1851-1879) .....</b>	<b>82</b>
4.1 INTRODUCTION.....	82
4.2 LITERATURE REVIEW – SOCIAL MOVEMENTS.....	84
4.3 EMPIRICAL SETTING: THE EMERGENCE OF LEBANESE NEWSPAPERS .....	90
4.3.1 <i>Micro-Level Grievances</i> .....	94
4.3.2 <i>Macro-Level Structural Changes</i> .....	103
4.3.3 <i>The Founding Fathers</i> .....	107
4.3.4 <i>The Framing Process</i> .....	113
4.3.5 <i>The Re-framing Process</i> .....	126
4.4 DISCUSSION AND CONCLUSION .....	132

<b>5</b>	<b>CHAPTER FIVE: CATEGORY DYNAMICS AND CLUSTER SPANNING DURING INDUSTRY EMERGENCE (1851-1879)</b> .....	<b>138</b>
5.1	INTRODUCTION .....	138
5.2	DATA COLLECTION.....	138
5.3	CLUSTER ANALYSIS.....	139
5.4	MULTILEVEL ANALYSIS.....	158
5.5	CONCLUSION.....	162
<b>6</b>	<b>CHAPTER SIX: CATEGORY DYNAMICS AND CLUSTER SPANNING THROUGHOUT THE LIFETIME OF THE INDUSTRY (1851 - 1974)</b> .....	<b>165</b>
6.1	INTRODUCTION .....	165
6.2	EMPIRICAL SETTING .....	165
6.3	DATA COLLECTION.....	169
6.4	CLUSTER ANALYSIS.....	171
6.5	MULTILEVEL ANALYSIS.....	185
6.6	SURVIVAL ANALYSIS .....	207
6.6.1	<i>Non-parametric Analysis</i> .....	207
6.6.2	<i>Semi-parametric Analysis</i> .....	209
6.6.3	<i>Other Aspects of Identity</i> .....	222
6.7	CATEGORIES OR FEATURES? .....	226
6.8	RESOURCE PARTITIONING: POSITIONING OR AUTHENTICITY?.....	230
6.9	CONCLUSION.....	232
<b>7</b>	<b>CHAPTER SEVEN: CONCLUSION</b> .....	<b>234</b>
7.1	INTRODUCTION .....	234
7.2	SUMMARY OF FINDINGS .....	234
7.3	RESEARCH IMPLICATIONS.....	237
7.3.1	<i>Discussion of Findings</i> .....	237
7.3.2	<i>Implications for Future Research</i> .....	239
7.4	GENERALIZABILITY AND LIMITATIONS .....	241
<b>8</b>	<b>REFERENCES</b> .....	<b>243</b>

## List of Figures

Figure 2.1 Cluster formation from categories .....	30
Figure 3.1 A newspaper issue on micro film at the Jafet library.....	61
Figure 3.2 Research design. Source: Krippendorff (2004) .....	81
Figure 4.1 Newspaper founding events in all Arabic countries up to 1879.....	92
Figure 4.2 Newspaper founding events in all Arabic countries up to 1879 excluding official newspapers.....	92
Figure 4.3 Newspaper founding events up to 1879 .....	127
Figure 4.4 Political content analysis excluding the scientific newspapers.....	130
Figure 5.1 The number of active newspapers in Lebanon 1851-1879.....	139
Figure 5.2 Dendrogram for all issues from 1851-1879.....	140
Figure 5.3 Dendrogram showing clusters forming after the 0.6 value.....	141
Figure 5.4 Simpsons index for newspapers between 1860 and 1879 .....	142
Figure 5.5 Monthly change in the standard deviation of the Simpson index .....	144
Figure 5.6 Dendrogram for the period 1851-1874.....	145
Figure 5.7 Dendrogram for the period 1875-1879.....	145
Figure 6.1 Number of founding events by Lebanese in their country and in other countries .....	166
Figure 6.2 Active newspapers 1851-1974 .....	170
Figure 6.3 Dynamics of the category Simpsons index over time .....	173
Figure 6.4 Cluster analysis 1880-1894 at the 0.6 level.....	175
Figure 6.5 Cluster analysis 1895-1913 at the 0.6 level.....	175
Figure 6.6 Cluster analysis 1919-1937 at the 0.6 level.....	176
Figure 6.7 Cluster analysis 1946-1949 at the 0.6 level.....	176
Figure 6.8 Cluster analysis 1950-1974 at the 0.6 level.....	177
Figure 6.9 Clusters active during every year 1895 – 1913 .....	184
Figure 6.10 Clusters active during every year 1950 - 1974.....	185
Figure 6.11 Effect of type and time on the dependent variable contraction .....	190
Figure 6.12 Effect of the period indicator and age on the dependent variable expansion .....	195
Figure 6.13 Effect of time, age and period of founding during the first period.....	203
Figure 6.14 Effect of age, period of founding and publication type during the fifth period .....	204
Figure 6.15 Survivor function from the Kaplan-Meier estimator .....	208
Figure 6.16 Smoothed hazard function from the Nelson-Aalen estimator .....	209
Figure 6.17 The graph of the hazard linear predictor .....	217
Figure 6.18 Newspaper founding events divided by religion of the founder .....	223
Figure 6.19 Non-parametric Survivor function based on the variable religion.....	224

## List of Tables

Table 3.1 Newspaper issues available for inspection and their life span. ....	62
Table 4.1 Newspaper issues inspected .....	114
Table 4.2 Names of the early newspapers.....	121
Table 4.3 Breakdown of the references to the government.....	129
Table 5.1 Distribution of newspaper titles among the groups 1851 – 1874.....	147
Table 5.2 Distribution of newspaper titles among the groups 1875 - 1879.....	147
Table 5.3 Simpson index for the newspaper titles before 1875 .....	149
Table 5.4 Simpson index for the newspaper titles after 1874 .....	149
Table 5.5 Average Simpson index for the cluster groups of 1851-1874 .....	152
Table 5.6 Average Simpson index for the cluster groups of 1875 - 1879.....	153
Table 5.7 Results for finding the optimal number of clusters 1875-1879.....	154
Table 5.8 The averages of the categories in the largest two clusters.....	155
Table 5.9 Results for finding the optimal number of clusters 1851-1874.....	156
Table 5.10 Multi-level logistic regression of the three variables Exiting, Entering, and Unsystematic movement.....	161
Table 6.1 Group analysis results for all periods .....	179
Table 6.2 Between group analysis.....	182
Table 6.3 Multilevel logistic regression for the dependent variable Exiting.....	187
Table 6.4 Multilevel logistic regression for the dependent variable Entering .....	192
Table 6.5 GEE regression of the Simpson index.....	199
Table 6.6 Proportion between Simpson index and category proportions .....	204
Table 6.7 Survival analysis results .....	211
Table 6.8 Linear predictors of the hazard function .....	216
Table 6.9 Results for finding the optimal number of clusters 1851-1974.....	219
Table 6.10 Survival analysis results including the environmental cluster spanning measure .....	221
Table 6.11 Average proportions dedicated to each category in each cluster 1950-1974 .....	229
Table 6.12 Correlations between categories 1950-1974 .....	229

## **Statement of Copyright**

The copyright of this thesis rests with the author. No quotation from it should be published without the author's prior consent and information derived from it should be acknowledged.



## **Acknowledgments**

First and foremost, I would like to thank my supervisors Professor László Pólos and Dr. Min Liu. I am grateful to them for agreeing to supervise me and was truly honored at having had the chance to pick their brains. Working on this thesis was one of the most fulfilling and enjoyable things that I had ever done. This is in no small part due to my supervisors' attitudes throughout the whole process. Anything good in what follows is a result of their comments and suggestions.

Next I would like to thank Durham Business School for giving me the chance to work with great researchers. More specifically I would like to thank everyone at the Doctoral Office for all their support and services. Thanks must also be given to Statistics.com and to all the contributors at Stata List for making the topic of statistical analysis more accessible to researchers.

Several people have helped me at one point or another, especially while I was just starting to collect the data. In no particular order, I would like to thank: Marwan el-Khatib, Professor As'ad abu-Khalil, Mohammed Taha, Hassan Mohanna, Mariam Mohanna, Rabie Barakat, Zeinab Salman, and Khoder Majed, Director of Research and Publications at the Lebanese Ministry of Information. I would also like to thank the staff of the Jafet library at the American University of Beirut and the staff of the Oriental library at Saint Joseph's University for being as helpful as possible while I spent the long hours collecting the data. I would also like to thank my friend Dr. Hussein Abdulsater for reading a first draft of the fourth chapter. His comments and suggestions were, as always, helpful. Special thanks must be given to Professor Joseph Illias. He was my primary source when it came to locating the data sources that I needed for this thesis. He was generous enough to take time out of his busy schedule, on several occasions, in order to assist a PhD student whom he had never heard of and with whom he shared no mutual friends. His generosity and knowledge will not be forgotten.

On a more personal note, I would like to thank my parents who have gone to great lengths to provide us with the chances that they never had. Their moral and financial support were crucial to the completion of this thesis. I can never repay them for all that they have given me. Also, all what I do, I do while thinking of my brother Mohammed. He will always be on my mind.

Finally I would like to thank my wife Alice for being in my life. She was the one that convinced me to continue my graduate studies. She has always been, and will always be, my friend, my partner, and my love. To her I dedicate this thesis and the following verses by Omar al-Khayam:

كَيْفَ يَحْوُمُ الْقَلْبُ يَوْمًا عَلَى  
غَيْرِكَ أَوْ يَبْغِي هَوَىَّ مَعَ هَوَاكَ  
إِنْ دُمُوعِي لَمْ تَدَعْ لِحِظَةَ  
عَيْنِي تَرْنُو لِحَبِيبِ سِوَاكَ

"يا كُمَيْل، إِنَّ هَذِهِ الْقُلُوبَ أَوْعِيَّةٌ، فَخَيْرُهَا أَوْعَاهَا، فَاحْفَظْ عَنِّي مَا أَقُولُ لَكَ:  
النَّاسُ ثَلَاثَةٌ: فَعَالِمٌ رَبَّانِيٌّ، وَ مُتَعَلِّمٌ عَلَى سَبِيلِ نَجَاةٍ، وَ هَمَّجٌ رِعَاعٌ أَتْبَاعُ كُلِّ نَاعِقٍ،  
يَمِيلُونَ مَعَ كُلِّ رِيحٍ، لَمْ يَسْتَضِيئُوا بِنُورِ الْعِلْمِ، وَ لَمْ يَلْجِئُوا إِلَى رُكْنٍ وَثِيقٍ"

علي

# **1 Chapter One: Introduction**

## **1.1 Introduction**

This work is based on recent theoretical advances that have made it possible for organizational ecologists to learn more about how industries emerge and about the effect that category spanning has on their well-being. Therefore, this thesis is best thought of as targeting two different questions that are nonetheless connected in that they are both concerned with identity. Categories have recently become a popular research topic in the field of organizational theory. This work hopes to contribute to this success by expanding on current work in the field.

During the twentieth century, organizational theorists took a huge leap forward when they started paying more attention to the environment that surrounds organizations. Prior to that, the organizations themselves were the sole focus of researchers, and this led to the development of what Burns and Stalker (1994) referred to as static theoretical models, of which the most famous was that of Weber (2009) and what he referred to as the “pure type”. These static models were soon to be replaced by dynamic ones that sought to study the effect that environments had on organizations while at the same time not neglecting the internal dynamics of organizations. This leap was largely initiated by March and Simon (1993) who argued that the social environment largely determined what alternatives the decision maker considers, and which ones he/she will ignore. Subsequently, in almost all the classical works of that period, one would be hard pressed to pick up a book about organizational theory that

did not refer to organizations as open systems (Blau and Scott 2003; Katz and Kahn 2007; Lawrence, Lorsch, and Garrison 1967; Thompson 2011). Eventually, this led to the rise of three different but related schools of thought: contingency theory, institutionalism, and organizational ecology. All three shared the premise that environments matter, and that it was not possible to study internal organizational dynamics without studying external environmental dynamics. However, they differed with regards to the level of analysis that they chose to concentrate on. Organizational ecologists believed that the key to developing dynamic models, which sought to explain the dynamics of organizations, was to study the population to which the organization belonged (Hannan and Freeman 1977). This led them to develop sophisticated quantitative methods that intended to study the forces of legitimation and competition in organizational populations. These models proved to be both popular and resilient, in that their results were replicated in many populations in diverse settings (Carroll and Hannan 2004). However, some researchers believed that these models, especially the variables that measure legitimation and competition, were too simplistic to be able to represent the complex dynamics that we witness in organizational populations. In seeking to address these criticisms Hannan, Pólos, and Carroll (2007) developed a completely new framework that utilized formal logic in order to develop a unified theory. At the heart of this new theory is the notion of categories, and subsequently, that of identities.

## **1.2 Category Dynamics**

One of the most enduring questions in organizational ecology is which organizations to include in the study of a specific population (Carroll and Hannan

2004). Without proper description and classification methods, the validity of the research will be highly questionable (McKelvey and Aldrich 1983). One of the earliest attempts at such a classification was undertaken by McKelvey (1982). Unfortunately it was never fully tested. Organizational ecologists have generally accepted that the traditional concept of density needs to be revised. Some ecologists speculated that weights need to be assigned to organizations based on their size (Barnett and Amburgey 1990), while others posited that geographical locations should be the key factor in establishing membership in the population under study (Lomi and Larsen 2001). Recently, in a series of publications, Pólos, Hannan and Carroll have presented a more radical position which calls for a constructionist view of populations (Hannan, Pólos, and Carroll 2007; Pólos and Hannan 2001; Pólos and Hannan 2002). In their view, audience members, be they regular customers or industry critics, tend to group similar organizations into categories to which they assign labels. Categories become forms once they achieve high levels of legitimation (Hannan, Pólos, and Carroll 2007).

The concept of categories has spawned a wide set of research papers which stress the importance of members abiding by the rules set forth by the categories of which they are members. This line of research points to the negative impact that spanning categories has on the life chances of the organizations (Hsu 2006; Hsu, Hannan, and Koçak 2009; Hsu, Negro, and Perretti 2008; Kovacs and Hannan 2009; Leung and Sharkey 2009; Negro, Hannan, and Rao 2010; Pontikes 2010; Zuckerman 2004). This need to abide by social codes is confronted with the need of organizations to differentiate themselves from the competitors (Deephouse 1999; Navis and Glynn 2010; Whetten and King 2008; Zuckerman 1999; Zuckerman 2004).

While the general findings of most of these papers are aligned, several findings seem to suggest that the consequences of category spanning are context-dependent (Hsu, Negro, and Perretti 2008). Pontikes (2008) developed the concept of “lenient” categories to explain why in some instances category spanning is not harmful, while Kovacs and Hannan (2009) and Hsu, Negro, and Perretti (2008) found that the effect of category spanning was contingent on the “fuzziness” of the category to which the organization belonged. Other researchers argued that time played a factor in the effect of category spanning (Ruef and Patterson 2009; Zuckerman, Kim, Ukanwa, and Rittmann 2003).

A promising line of work initiated by Kovacs and Hannan (2011) argued that the shape of the category space needs to be taken into account while studying the effects of category spanning. Their work led them to conclude that category spanning was most harmful when the distance between categories is large. By allowing for the effect of the category space, these researchers were able to successfully shift the debate from one that is concerned with internal characteristics of categories to one that is concerned with the relative position of categories in a wider resource space. By doing so, the researchers have managed to take into account the socially constructed nature of categories. However, their choice of distance measure is not without its problems. While their measure accounts for the distance between categories, it does not take into account the total category space. This thesis will argue that the category space is better accounted for by using a two level approach. While categories are the main units on the first level, clusters, which are formed from category spanning dynamics, are the units of interest on the second level. This formulation allows for two different dynamics: category spanning and cluster spanning. I will argue that the

effect of category spanning is time-dependent. More specifically, I will argue that the dynamics which follow from category spanning are closely linked to the resource-partitioning theory which posits that different forms have different survival chances in different times, and that it is possible for more than one form to co-exist together. I will then move on to argue that cluster spanning is hazardous to organizations' mortality rates at all times.

### **1.3 Identity in Emerging Markets**

Previous identity formulations (Albert and Whetten 1985) viewed identity as being determined solely by internal members (Hsu and Hannan 2005) or by internal enduring attributes (Wry, Lounsbury, and Glynn 2011) and features (Hannan, Baron, Hsu, and Koçak 2006). Recent theoretical advances made by Pólos and colleagues (Hannan, Pólos, and Carroll 2007; Pólos and Hannan 2001; Pólos and Hannan 2002) have argued that the identity of the organization consists of the social codes that audience members use for the categories of which it is a member of. Research has shown that identities play a central role in population dynamics (Carroll and Swaminathan 2000; Liu and Wezel 2013; Pozner and Rao 2006), in product success (Jensen 2010; Khessina and Carroll 2008), and ultimately the success of the firm as a whole (Carroll, Bigelow, Seidel, and Tsai 1996). Recently, thanks to the new theoretical foundation mentioned above, the attention of organizational ecologists has shifted to the period of industry formation. With this shift, increasing attention has been directed at the unresolved topic of legitimacy (Wry, Lounsbury, and Glynn 2011). According to Berger and Luckmann (1967), "Legitimation 'explains' the institutional order by ascribing cognitive validity to its objectivated



meanings. Legitimation justifies the institutional order by giving a normative dignity to its practical imperatives” (p. 93). Previously, it has been assumed that legitimacy was dependent on the number of organizations in the industry (Carroll and Hannan 2004). However, some research has shown that nascent small populations were able to gain sufficient legitimacy (Wry, Lounsbury, and Glynn 2011). The new theoretical formulation argued that while the number of organizations in an industry mattered, a new variable, which was termed the “contrast”, was also crucial. If a category had high contrast then the category would be legitimated, hence becoming a form (Hannan 2010). Categories are tied to identities in that the identity of an organization is formed, in part, from the categories of which it is a member of (Hsu and Hannan 2005; Wei 2005). Therefore, according to Hannan, Pólos, and Carroll (2007), organizations in an emerging industry needed to project a unified identity. Recent research has been able to show that identity played a central role in the emergence and legitimation of organizational forms (McKendrick and Carroll 2001; McKendrick, Jaffee, Carroll, and Khessina 2003; Perretti, Negro, and Lomi 2008). However, this research has produced diverging results as to the necessary conditions for successful legitimation. McKendrick and colleagues have argued that a unified identity is a facilitating condition for successful form emergence while Perretti, Negro, and Lomi (2008) have argued that the effect of identity is context dependent. They conclude that “...different domains offer alternative default settings, and the way a new organizational form emerges is in the context of interpretation and social classification of the identity of a domain as well as the identity of candidates” (p. 543). This position gained credence when King,

Clemens, and Fry (2011) showed that a unified identity was not a necessary condition for the emergence of new forms.

In order to move forward with the study of identities in emerging industries, it is important to understand that identity as a concept is multidimensional (Ashmore, Deaux, and McLaughlin-Volpe 2004) and that most previous research in the area of organizational ecology have assumed, either implicitly or explicitly, that it is one-dimensional. This thesis intends to look at identity from two angles. First, a close inspection of the early period of the newspaper industry in Lebanon shows that the industry lends itself well to the tools of social movement analysis. This thesis will show that the early newspapers formed a social movement that used “framing” (Benford and Snow 2000; Goffman 1986) to justify their newly found existence. It will become clear that in most cases the founders framed the problems facing the society in a unified way and identified similar goals to each of their newspapers. By describing these framing processes, this thesis will show that all participants projected the part of the identity of the newspapers that dealt with how the founders saw their movement in a very similar and coherent manner. The thesis will then move on to study the part of the identity which was reflected in the categories spanned by the topics of the newspapers, i.e., what the customers saw. The thesis will study whether the category spanning actions were systematic or random, and whether they moved in parallel among the different newspapers. By looking at these dynamics it will be possible to see if all, or most, newspapers projected that part of the identity coherently and unanimously.

## 1.4 Empirical Setting

The theories developed in this paper will be tested within the context of the Lebanese newspaper industry. The newspaper industry has generally been a popular empirical setting in the past. It has been used in studies of organizational founding (Delacroix and Carroll 1983), organizational mortality (Carroll and Delacroix 1982), resource distributions (Boone, Witteloostuijn, and Carroll 2002), market partitioning (Carroll 1985), management change (Carroll 1984), inter-organizational linkages (Miner, Amburgey, and Stearns 1990), and organizational change (Amburgey, Kelly, and Barnett 1993). The choice of industry was motivated by two reasons. First, there has been no study concerned with the organizational ecology model that has been tested within any Arab industry. Second, newspapers are well suited to the study of categories due to the nature of the product.

The Lebanese newspaper industry offers a very interesting context. This thesis will focus on the industry from 1851 - 1974. The year 1974 was chosen as the cut-off period because the Lebanese civil war started in 1975 and many of the newspapers had to shut down because they could no longer operate during the war. In addition, many of the issues of the time were lost. Throughout the years, the number of newspapers published in Lebanon has been disproportionately large when compared to those published in other Arabic countries. This affinity to newspaper publication was not unique to the Lebanese who resided in Lebanon. During the period 1851 – 1974, the Lebanese published 1056 newspapers in Lebanon and 703 newspapers outside of Lebanon<sup>1</sup>. Inside

---

<sup>1</sup> These numbers are obtained from the data that I collected. A detailed explanation of the data collection procedure is found in chapter 3.

Lebanon, starting newspapers was such a common activity that the government had to issue a decree in 1953 in order to limit the number of newspapers active at any one time in the country. The industry has been the subject of study of several previous publications. Most of these studies however studied the industry from a historical perspective without paying attention to the reason for the emergence of the industry. The primary intention of the authors was documenting general information about each newspaper title. In the case of one single work, that of al-Rifa'i (1967), we can see an attempt to study the actions that these newspapers undertook. Unfortunately, the author completely disregards the true emergence period, which I identify as 1851 – 1879, and instead tries to link the emergence of the industry to “stories” told in issues that were published after 1879. Ayalon (1995) offers a very brief account of the industry emergence in Lebanon due to the fact that his work is concerned with the entire Arab world and the author generally pays more attention to Egypt. This thesis is the first research to be conducted in this empirical setting from an organizational, instead of historical, perspective.

## **1.5 Thesis Division and Theoretical Aims**

The aim of this thesis is to better understand how identity forms and how it affects the life chances of organizations. Identity is multidimensional, therefore it is necessary to indicate which dimensions are being studied.

- Chapter 2 will serve as the theoretical foundation of the thesis. It will start with a discussion of recent theoretical developments in the field of organizational ecology, followed by a discussion of current points of contention relating to the topic of identity during industry emergence and

the concept of category spanning throughout the life history of the industry. The main aim of this chapter will be to highlight the areas that this thesis will try to contribute to, specifically the concept of clusters, and why clusters are well suited to represent the category space.

- Chapter 3 will provide an overview of the statistical methods that will be used to test the various hypotheses which will be developed in chapter 2. The chapter will offer an explanation as to why the statistical methods were chosen in addition to a description of these methods.
- Chapter 4 will study the collective identity of the newspapers during the founding period. Here, collective identity refers to a "shared sense of 'oneness' or 'we-ness' anchored in real or imagined shared attributes and experiences among those who comprise the collectivity and in relation or contrast to one or more actual or imagined sets of 'others'" (Snow and Soule 2010, p. 125). Therefore, chapter 4 will look at identity as it was being projected by the newspapers themselves. This will be done by examining the newspaper editorials. The chapter will show that the newspapers projected a unified collective identity in which their primary goal was to spread knowledge.
- Chapter 5 will also study the identity of the newspapers during the founding period, but this time instead of looking at the collective social identity, the chapter will study the individual identity of the newspapers as determined by their category spanning dynamics. This definition of identity builds on the work done by Hannan, Pólos, and Carroll (2007), in which organizational identity consists of social codes which specify the features than an organization is expected to possess. These social codes

are paired with labels in order to form codified categories. The chapter will show that, unlike the collective social identity, the category spanning dynamics were not uniform and took some time to stabilize. By the end of the founding period, the chapter will show that two different clusters emerged from the dynamics of category spanning.

- Chapter 6 will study the entire period (1851 - 1974). Like chapter 5, this chapter will also consider the part of identity that is determined by the category spanning dynamics of the newspapers. The main aim of this chapter is to show that category spanning is not hazardous throughout all stages of industry development. Specifically the chapter will show that category spanning is hazardous once the "resource-partitioning" phase has kicked in, i.e. once the industry witnesses the rise of the specialists and the consolidation of the generalists. Finally, chapter 6 will show that clusters, which form from category spanning dynamics, are better suited to explain the hazardous effects of a change in identity. Cluster spanning will be shown to be a hazardous act throughout the entire period of study.
- Chapter 7 will serve as the conclusion where the results will be discussed within a single context along with a description of what these results imply for future research.

## **1.6 Conclusion**

This chapter served as a brief introduction and motivation for the thesis. The next chapter will provide a more detailed overview of the current literature. The two main areas of interest in this thesis are the nature of the identity of organizations in nascent fields and the effects of category spanning on

organizational life chances. Each of these areas will be dealt with separately. Ultimately however, the two questions are tied together because both are concerned with the identity of organizations. Chapter 2 will conclude by providing a description of how the rest of the thesis is divided. This description is left to the end of the second chapter because by then it should be clear what this thesis aims to accomplish.

## **2 Chapter Two: Literature Review and Development of Research Problems**

### **2.1 Introduction**

In this chapter I will attempt to provide the theoretical foundation on which the thesis is built. There are two parts to the chapter. The first part will discuss the topic of niche width and how it applies to category spanning. A review of the literature will be presented, and it will be followed by the identification of the research problems that this thesis will try to target. The second part of this chapter deals with the issue of identity formation during the emergence period of industries. After reviewing past literature on the subject I will attempt to identify the areas that are of prime interest to this thesis.

### **2.2 Literature Review – Niche Width**

One of the earliest, and most important, theoretical developments of early organizational ecologists was the notion of niche width. According to Freeman and Hannan (1983), “niche width refers to a population’s tolerance for changing levels of resources, its ability to resist competitors, and its response to other factors that inhibit growth” (p. 1118). The concept of niche width is at the heart of the study of generalist and specialist organizations. In essence, the study of the distribution and the dynamics of these two types of organizations is the study of niche width (Hannan and Freeman 1977). Subsequent studies have compared the founding, disbanding, and survival rates of organizations that were classified into one of these two types. In his seminal paper, Carroll (1985) studied how the



dynamics of generalists in fact led to the proliferation of specialists. The fundamental point made by that paper was that the two types of organizations were in fact connected in the resource space, and that the activities of one group had a significant effect on the dynamics of the other. The findings of this research were replicated in a large number of studies in different markets (Barnett and Carroll 1987; Boone, Witteloostuijn, and Carroll 2002; Carroll and Swaminathan 1992; Lomi 1995; Swaminathan 1998).

The resource partitioning theory provides a very attractive theoretical framework for explaining real-world phenomenon that ran counter to dominant theories that predicted homogeneity in organizational forms within populations. Freeman and Hannan (1993) describe how three environmental variables dictate what the best form of organization is: environmental variability, grain and dissimilarity. They posit that environments that are coarse-grained and uncertain are favorable to generalists, while environments that are fine-grained and low on variability favor specialist organizations. In a similar manner, institutionalists posit that organizational fields are subjected to a strong force which pushes them in the direction of homogenization (DiMaggio and Powell 1983; March 1981). If organizations are forced to become homogeneous due to environmental pressures, then how can we explain the co-existence of both generalists and specialists in a diverse set of mature industries (Carroll and Hannan 2004)? The resource partitioning theory provided a very plausible explanation.

An important question that researchers were faced with was how to determine the niche width of organizations. In other words, how do we decide which organization was a specialist and which was a generalist? This question is

of fundamental importance because unless the researcher is able to systematically and consistently classify the organizations into the proper group, then no utility could be made of the results obtained in the statistical analysis. This question is very similar to another question which has long been on the mind of population ecologists: which organizations should we include in the study of certain populations (McKelvey 1982; McKelvey and Aldrich 1983)? Previously, both these questions were answered in a similar manner using a discrete mechanism. In their study of restaurants, Freeman and Hannan (1983) grouped 33 forms into three categories: generalists, fast-food establishment, and specialists. Carroll (1985) defined as generalists those organizations which published one or more general interest newspapers, while classifying the rest as specialists. Carroll and Swaminathan (1992) and Swaminathan and Carroll (1995) classify organizations in one of three forms: mass producers, microbreweries and brewpubs. Carroll and Swaminathan (2000) take extra care with hybrid firms in order not to “double count” any organization. Boone, Witteloostuijn, and Carroll (2002) considered the national newspapers which operated in every province to be the generalists, while local newspapers were taken as the specialists. Lomi (1995) divided the banks into core national banks and rural cooperative banks.

The studies mentioned thus far have followed a binary logic, in that an organization was classified into one of several groups with different niche widths. This has also been the same type of logic used in determining whether an organization belongs to a population or not, an issue which has been the target of several studies (Barnett and Amburgey 1990; Barron, West, and Hannan 1994; Baum and Mezias 1992; Carroll and Wade 1991; Lomi and Larsen 2001).

However, the issue of determining which organizations were generalists and which were specialists did not receive the same attention. Some studies attempted to develop more flexible measures to determine the niche width of organizations. Baum and Singh (1994) used a measure of niche width which was based on the age groups which centers catered to, McPherson (1983) utilized a measure which was based on the demographic information of the members of voluntary organizations, Dobrev, Kim, and Carroll (2002) utilized the size of the organization to create a continuum along which the larger the organization the more “generalist” it was, Dobrev, Kim, and Hannan (2001) measured the niche width of an automobile producer as the spread of engine capacity over all models that are produced by a firm, but they only use the maximum and minimum engine capacities, and Hsu (2006) calculated the niche width as the number genres that each film was classified under. With the exception of the use of size as an indicator for the level of generalism, which is in itself problematic (Blau and Scott 2003; Dobrev, Kim, and Hannan 2001), all these measures share a basic assumption, and that is that the organization is a full-fledged member of each area in the resource space which it targets.

In recent years, Pólos, Hannan and Carroll have developed a new theoretical framework which was based on the use of non-monotonic logic (Hannan, Pólos, and Carroll 2007; Pólos and Hannan 2001; Pólos and Hannan 2002). A primary departure from previous theoretical models is that “...instead of treating categories as analytic, as products of researchers distinctions, we assume that the members of audience segments create categories and forms” (Hannan, Pólos, and Carroll 2007, p. 33). In their view, audience members, be they regular customers or industry critics, tend to group similar organizations

into categories to which they assign labels. Organizations are no longer thought of as being full-fledged members of the category or not being members at all. Instead, each producer is assigned a grade-of-membership which can take any value between, and including, zero and one.

Using this new formulation, Hsu, Hannan, and Koçak (2009) revisited the topic of niche width. They proposed a measure which allowed for examining “the degrees of generalism, rather than a binary comparison of specialists versus non-specialists” (p. 158). The measure proposed is Simpson’s index of diversity (Simpson 1949), which is represented by the following formula:

$$1 - \sum_{l \in l(m)} \mu_{i(l)}^2(x),$$

where  $\mu_{i(l)}(x)$  represents the grade-of-membership of the producer in the category  $l$ . The grade-of-membership was determined by the following formula:

$$\mu_{i(l)} = \frac{l_i(x)}{l_x},$$

where  $l_x$  is the total number of labels applied to the producer  $x$ . The numerator was set to one when a certain label was applied to the producer and to zero otherwise. The Simpson’s index can take on any value in the interval  $[0, 1]$ , with zero indicating that the producer spans only a single category in which his grade-of-membership is one. The higher the index the higher the degree of generalism. This new conceptualization of the niche width was used in several studies, thus replacing the old binary classification scheme (Kovacs and Hannan 2009; Pontikes 2008).

Kovacs and Hannan (2011) took the next logical step by proposing that the niche width of organizations needs to account for, at least to some degree, the structure of the category space. According to them, the measure of niche

width needs to include the “distance” between the categories that an organization spans. They argued that spanning categories that were close to one another should increase the niche width of the organization far less than spanning categories that had a large distance between them. The grade-of-membership of a certain producer was now defined as such:

$$\mu_i(x) = \frac{l_i(x)}{1 + \sum_{j \in L} l_j(x) d_{ij}}$$

where  $d_{ij}$  is the distance from label  $i$  to label  $j$ . Just like before, the addition of new labels lowered the grade-of-membership of the producer, but this time the magnitude of the decline in the grade-of-membership depended on the distance between the labels. Next the authors tackle the question of how to determine the distance between categories. This leads them to adopt the following formula based on the work of Shepard (1987):

$$s(i, j) = \exp(-\gamma d_{ij}), \gamma > 0,$$

where  $s(i, j)$  is the similarity between the two categories. As we see, similarity is inversely related to distance. So the question of finding the distance now becomes one of finding the similarity. To find the similarity between two categories, the authors modified the measure proposed by Jaccard (1901) so that it allowed for asymmetry. Their measure is represented by the following formula:

$$Sim(i, j) = \frac{|i \cap j|}{|i \cap j| + \alpha|i - j| + \beta|j - i|}, \alpha, \beta > 0$$

With these new definitions of grade-of-membership, distance and similarity, the authors also proposed a new formula for the niche width. The reason for this was that if we used Simpson’s index together with the new grade-of-membership

formula, then the result can be negative, and this does not make sense (Kovacs and Hannan 2011). To develop the new measure the authors first calculated the total pairwise distance among the labels assigned and its average as follows:

$$D_x = \sum_{i \in L} \sum_{j \in L} l_i(x) l_j(x) d_{ij}$$

and

$$\bar{d}_x = \frac{D_x}{l_x(l_x - 1)}$$

From these two measures, the authors proposed two different methods for calculating the niche width as follows:

$$W_C(\mu_x) = \begin{cases} 0 & \text{if } l_x = 1; \\ l_x \bar{d}_x & \text{otherwise.} \end{cases}$$

and

$$W_N(\mu_x) = \begin{cases} 0 & \text{if } l_x = 1; \\ 1 - (1 + l_x \bar{d}_x)^{-1} & \text{otherwise.} \end{cases}$$

These measures simplify to the following:

$$W_C(\mu_x) = D_x(l_x - 1)^{-1}$$

$$W_N(\mu_x) = 1 - \left(\frac{D_x}{l_x - 1}\right)^{-1}$$

The main difference between both measures is that while with the first measure the “effect of increasing total pairwise distance does not depend on the level of the niche width”, the second measure “increases more with a given increment of distance at low levels of total distance spanned than at higher levels” (Kovacs and Hannan 2011, p. 13-14). With these two measures, it now became possible for an organization that spanned many, but close, categories to have a lower niche width than another one which spanned a fewer number of distant categories. Pontikes and Hannan (2012) adopted the above definitions of

distance and category spanning into their study of the software industry after they had previously studied the industry without taking into consideration the category structure (Pontikes 2008).

### **2.3 Development of Research Problem – Niche Width**

The topic of niche width is inexplicably linked to the concept of categories. Organizations that have a large niche width span categories while those that have a small niche width tend to concentrate their efforts on specific locations in the resource space (Hsu, Hannan, and Koçak 2009). The work of Kovacs and Hannan (2011), Hsu (2006), Hsu, Hannan, and Koçak (2009), and Hannan, Pólos, and Carroll (2007) all point in the same direction: spanning categories is a hazardous act. The concept of categories has spawned a wide set of research papers which stress the importance of members abiding by the rules set forth by the categories of which they are members. Zuckerman (1999) and Zuckerman (2000) have shown that critics and analysts penalize products that do not fit in certain categories due to confusion over the product's identity, while Zuckerman and Kim (2003) have also shown that market boundaries severely restrict the range of identities that economic actors can presume. Such results have been corroborated by a large number of researchers (Hsu 2006; Hsu, Hannan, and Koçak 2009; Hsu, Negro, and Perretti 2008; Kovacs and Hannan 2009; Leung and Sharkey 2009; Negro, Hannan, and Rao 2010; Pontikes 2010; Zuckerman 2004). Several reasons have been proposed for such a downgrade: being a member of several categories means abiding by different social codes at the same time (Hsu and Hannan 2005) which may lead to clashes of category codes (Kovacs and Hannan 2009); there is an inherent trade-off between niche

width and the capacity for performance (Hsu 2006); organizations that do not abide by social codes lose legitimacy and hence appeal (Whetten and King 2008); audience members find difficulty in evaluating the offerings of such organizations (Leung and Sharkey 2009), or they simply do not view them favorably (Leung 2010). Finally, Kovacs and Johnson (2012) found that a number of the above factors might be involved. Therefore, it seems that organizations are subjected to two forces. The first is a need to project a common image as that of its competitors so that they will be visible to the audience, and the second is to differentiate themselves from their competitors (Deephouse 1999; Navis and Glynn 2010; Whetten and King 2008; Zuckerman 1999; Zuckerman 2004).

While the general findings of most of these papers are aligned, several findings seem to suggest that the consequences of category spanning are context-dependent (Hsu, Negro, and Perretti 2008). Pontikes (2008), for example, found that members of “lenient” categories were able to span boundaries much more easily than members of categories which were not lenient. A similar result was obtained by Kovacs and Hannan (2009) and Hsu, Negro, and Perretti (2008), who found that the effect of category spanning was contingent on the “fuzziness” of the category to which the organization belonged. Kovacs and Hannan (2011) found that once the structure of the space of categories is accounted for, quantitative results would not only differ by magnitude, but by sign as well. Time has also appeared to play an important role. Zuckerman, Kim, Ukanwa, and Rittmann (2003) found that film stars that have many years of experience are able to straddle categories by acting in different genres, while less experienced actors were heavily penalized for the same action. On the other hand Ruef and Patterson (2009) found that in the original stages of



classification systems there were low risks of devaluation because category heuristics were not well developed. With time, however, these heuristics developed into strict codes that members had to abide by. Rao, Monin, and Durand (2005) highlighted the dynamic nature of categories when they noted that the strength of category boundary was eroded as more and more producers crossed them. Recently an interesting finding has suggested that in some cases, it might not be the properties of the categories themselves that determine the effect of category spanning, but factors exogenous to the categories. Pontikes (2010) found that different types of audiences react differently to category spanners while Damon and Zuckerman (2001) found that it was the position of the producer inside the category that decided whether conforming was good or bad. Smith (2011) on the other hand has found that the identity of organizations determines whether border-crossing activities are viewed positively or negatively.

I believe that the most promising work done to date to address the implications of category spanning is that of Kovacs and Hannan (2011). The answer to the question “when is category spanning hazardous and when is it not” lies in the category space. Unlike the work by Pontikes (2008), which labeled categories as lenient, I believe that the effect of category spanning is related to the structure of the space and the location of the category within that space rather than to inherent properties of certain categories. This is why I propose to continue the line of work started by Hsu, Hannan, and Koçak (2009) and extended by Kovacs and Hannan (2011). However, while I believe that the category space needs to be accounted for, I believe that the formulation posited by Kovacs and Hannan (2011) could be further improved. This can be done in

two ways. First, I believe that the entire structure of the category space needs to be taken into consideration, not just the distance between two specific categories. Second, I believe that the concept of distance needs to be accounted for not on the level of categories, but on the level of clusters that are formed from the dynamics of categories. First I will elaborate on the shortcomings of the distance measure, and then I will introduce the concept of clusters that, I believe, will alleviate the aforementioned problems.

### **2.3.1 Problems with the Distance-based Measures**

One issue with the distance-based measure is that the sum of the grade-of-memberships of a certain producer does not necessarily add up to one. Theoretically, this is not really a problem because the grade-of-membership is not a probabilistic function and therefore there is no requirement that the sum be equal to one (Hannan 2010). However, by constraining the grades-of-membership to be equal to one, or at least equal to each other, we would be in alignment with what Freeman and Hannan (1983) referred to as the principle of allocation:

*“Levins (1968), whose lead we follow, proposed a principle of allocation: each population has a constant sum of fitness that may be allocated across strategies for playing the evolutionary game. Specialist populations follow the strategy of betting all of their fitness chips on specific outcomes; generalists hedge their bets”*(p. 1119).

This same point was re-iterated by Hannan (2010) when he stated that “the level of total category engagement is normally the same for all producers in a market” (p. 168), and by Hannan, Pólos, and Carroll (2007) who stated that “In particular,

we impose the “constant-sum” constraint that increased breadth of a niche comes at the expense of lowered appeal at some positions” (p. 180). By constraining the sum of the grade-of-membership to be equal across all producers we would be in harmony with previous theoretical statements.

Another problem with the distance-based measure, which is also present in some papers that utilized the non-distance based approach, is that the grade-of-membership function is neither affected by features of the product or producer nor by the producer’s engagement in the category. Hannan, Pólos, and Carroll (2007) state that “A natural step in seeking to comprehend the bases of membership in a label involves relating the feature values of the producers/products to their grades of membership in a label” (p. 60), and “...the grade of membership in a position is the share of (producer) x’s engagement devoted to the audience at that position” (p. 180). If the grade-of-membership function reflects the appeal of an offering (Hannan, Pólos, and Carroll 2007), and engagement increases the appeal of the producer (Kocak, Hannan, and Hsu 2009), then surely engagement affects the grade-of-membership of that producer (Hsu, Hannan, and Koçak 2009). Most research so far has used a binary function such that if a producer spans a category then the function is set to one and if not then it is set to zero (Carroll, Feng, Le Mens, and McKendrick 2010; Hsu, Hannan, and Koçak 2009; Hsu, Negro, and Perretti 2008; Kovacs and Hannan 2009; Pontikes 2008; Pontikes 2010). Therefore, if a producer spanned three categories then his grade-of-membership will be  $1/3$  in each. The grade-of-membership function which is used in the distance based approach utilizes the same logic only this time it takes distance into consideration. I believe that the

grade-of-membership should distinguish between the levels of commitment that the producer applies in each category.

The problem of binary thinking is also present in the formulation of the similarity function. The Jaccard measure is after all a binary similarity measure (Everitt, Landau, Leese, and Stahl 2011) in that the similarity between two categories relies on a producer being mentioned in both. An example would help illustrate the problem with this. Imagine a category space in which two categories are spanned by 30% of the organizations. Imagine also that all of these organizations dedicate 70% of their resources to the first category, 10% to the other, and the final 20% are distributed between other categories. Now imagine that there are two other categories which are also spanned by 30% of the organizations, but that this time these organizations dedicate 50% of their resources to the first and 50% to the second. According to the Jaccard similarity measure, both pairs of categories have equal similarity, and hence are separated by the same distance. It is my belief that the distance between the second pair of categories should be smaller than the distance between the first pair.

The next issue with the distance-based niche width is more important than all of the above. Previous studies have shown that in many industries there comes a time when generalists take over the market. This is usually manifested it an increase in the market concentration ratio (Dobrev, Kim, and Carroll 2002; Lomi 1995; Marquis and Lounsbury 2007). The increase in the number of generalists means that the proportion of times that categories are mentioned together will also increase. This, in turn, will lead to a decrease in the distance between the categories since the distance measure is based on the similarity measure, which is based on a frequentist approach. The decrease in the distance

between the categories will lead to a decrease in the niche-width of the generalist organization. So by increasing the number of generalists, these generalists are no longer generalists because their niche width decreases. An extreme example would help illustrate the direction in which this kind of thinking leads to. Imagine for example a market in which many producers are active, with some of them being specialists and others being generalists. Now imagine that this particular environment favors generalist organizations and that with time the specialists are eliminated and generalist organizations proliferate. In order to become generalists, the organizations start expanding into new categories. It might be the case that all these generalist organizations end up spanning the same group of categories. This would mean that the distance between the categories is zero, thus making the niche width of the individual organization also zero. Although this is an extreme example, it illustrates how an environment that favors generalists will witness the rise of generalists, and that this rise will itself lead to a decrease in the niche width which theoretically means that these generalists become specialists even though they continue to span a wide set of categories.

One of the proposed reasons that spanning categories is hazardous to organizations is that audience members penalize non-conformity. Imagine now that there exists a category space in which two certain categories always co-occur together at a certain point in time. Now imagine that a new organization comes along and decides that it wants to span just one of these categories. This would cause a decrease in the grade of membership of all exiting organizations, while the new organization will have a zero niche width because it only spans one category. Since a smaller niche width is more favorable than a larger niche

width, the original organizations end up being punished while the new organization is not penalized for ignoring the second category even though it is the new organization that has not conformed to the market dynamics. Instead, what is happening is that all the organizations that used to exist and that did not change anything are the ones being penalized.

The final problem with the proposed distance-based measure is that it only takes into account the actions of organizations that span the two focal categories when it calculates the distance between these categories. This might seem to be the logical thing at first, but it is important here to take note of the nature of categories. It is generally accepted that categories are socially constructed entities. However, the method of their construction is an area of disagreement. White (1981) and White (2000) have argued that categories form via the actions that producers undertake while watching each other from behind a one-way mirror that prevents them from “seeing” the customers. Kennedy (2005) and Kennedy (2008) claimed that the media mediated the exchanges that took place on either side of the mirror. Rosa, Porac, Runser-Spanjol, and Saxon (1999), on the other hand, have attempted to bring the consumers into the heart of the category making process. According to them, categories are “...theoretical constructs, developed and agreed to by market actors to make sense of producer and consumer behaviors” (p. 64). In other words, the distance between categories is not an absolute measure that is independent of the surroundings. The distance between two categories is relative due to their socially constructed nature, and hence the context in which they exist must be taken into account. Again I will use an example to illustrate this point. Imagine a market that is made up of 20 organizations. The category space of this market is made up of eight

categories. Now imagine that two of these categories are each mentioned ten times, and that they co-occur five times. This means that the similarity measure is  $5/(10+10-5) = 1/3$ . Imagine that the total number of category membership claims is 80. Next imagine that after a certain period of time the market contains 100 organizations, that the number of categories has doubled to 16, and that the total number of category membership claims has subsequently increased to 960. Imagine also that the two categories in which we are interested are still mentioned ten times, and also still co-occur five times. In such a case the distance between the two categories at both time intervals is still the same,  $1/3$ . This is despite the fact that the number of organizations increased five folds, the number of categories doubled, and the number of category membership claims increased tenfold also. It should also be noted that initially on average each organization claimed to be a member of 4 categories while at the next time interval the average was 9.6, thus indicating an increase in the average niche width. However, according to the distance-based measure, all these changes do not affect the distance between the two categories as long as the numerator and the denominator of the similarity measure remain the same. Previously researchers have posited that the actions and resources of one population affects other populations (Freeman and Hannan 1983), that competition is socially constructed (Porac, Thomas, Wilson, Paton, and Kanfer 1995) and that it is largely a function of relative positions in the resource space (Baum and Haveman 1997), and that the creation of reality in organizations is an ongoing process (Weick 1995). Therefore, if in a socially constructed world (Berger and Luckmann 1967), actions that take place in one organizational population can affect organizations in another population, it seems natural to argue that actions

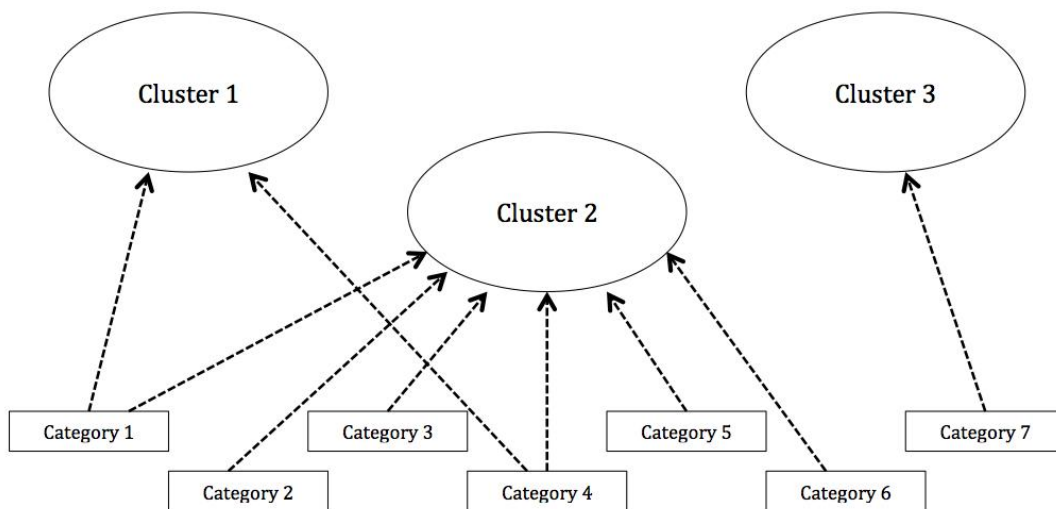
that take place, within the same population, in one category have an effect on other categories. In addition, if organizational reality and competition are themselves socially constructed, then significant shifts in the competition landscape (i.e., number of competitors and their average niche width) should affect how audience members evaluate all competitors. Dobrev, Kim, and Hannan (2001) argued that “an organization can experience a change in its relative position without making any changes in its own product offerings” (p. 1306). I believe that the same applies for categories.

### **2.3.2 Clusters: A Multi-level Approach**

How do we incorporate the category structure in the study of categories while overcoming the above issues? I believe that the solution lies in dividing the analysis into two parts. The distance-based measures proposed by Kovacs and Hannan (2011) incorporate the concept of distance at the category level thus affecting the definition of grade-of-membership and niche width. My solution leaves the definition of niche width as proposed by Hsu, Hannan, and Koçak (2009) and incorporates the category space on a second level which I refer to as clusters. It is important to note that the meaning of clusters depends on the observations of the first level. Cluster analysis is used to study the similarities and differences among the observations. Therefore, clusters derive their meaning from the observations that we are interested in studying. In this thesis, I am concerned with the clusters that are formed by the dynamics of organizations as they span different categories. Figure 2.1 below illustrates the formation of clusters from categories. The figure shows that there are seven categories in total and that three clusters form from the dynamics of the organizations’



category spanning dynamics. Cluster 1 contains organizations that span categories one and four, cluster 2 contains generalist organizations that span 6 categories while cluster 3 contains specialist organizations that span category seven only. It is important to note here that when we say an organization belongs to cluster 1 for example, it does not mean that all organizations that span category 1 belong to cluster 1. As we can see from the figure, generalists that span all categories, including category 1, belong to the second cluster. Also important is that cluster formation depends on the extent of the category spanning activity.



**Figure 2.1 Cluster formation from categories**

If we take newspapers as an example, a cluster may form from newspapers that dedicate 80% of their content to politics and the other 20% to economics, and another cluster might form from newspapers that dedicate 50% of their content to politics and the remaining 50% to economics. In other words, cluster formation depends on category spanning and organizational engagement within each category.

It is important here to note that clusters, as formulated in this thesis, are different from what Rosch et al. (1976) referred to as superordinate categories. The delineation between subordinate, basic, and superordinate categories is a question of finding the “natural” level of categorization:

*“The aim of the present research is to show that the world does contain ‘intrinsically separate things.’ The world is structured because real-world attributes do not occur independently of each other...That is, combinations of attributes of real objects do not occur uniformly”(p. 383).*

And,

*“A working assumption of the present research is that in the real world information-rich bundles of perceptual and functional attributes occur that form natural discontinuities and that basic cuts in categorization are made at these discontinuities” (p. 384-385).*

Clusters, which are formed from category spanning activity, are on the other hand not natural. Their formation is highly dependent on the category spanning dynamics of organizations, and therefore they are non-natural objects that are dynamically formed. This difference between superordinate categories and clusters is reflected in the fact that the question of the level of categorization is related to the taxonomy:

*“A taxonomy is a system by which categories are related to another by means of class inclusion. The greater the inclusiveness of a category within a taxonomy, the higher the level of abstraction...Each category within a taxonomy is entirely included within one other category (unless it is the highest level category) but is not exhaustive of that more inclusive category.*

*Thus, the term level of abstraction within a taxonomy refers to a particular level of inclusiveness”(p. 383).*

The same cannot be said about clusters. As discussed above, a cluster does not necessarily include all of a category and a certain category can be included within more than one cluster. A cluster is not just another level in the taxonomy. It is a representation of the category-space which helps us understand how organizations span, and dedicate their resources to, different categories, hence the relationship between clusters and niche width as discussed above.

I determine the grade-of-membership  $\mu$  of an issue in each category as the proportion of that issue that is dedicated to the category. So, for example, if a newspaper consisted of eight pages in which two were dedicated to sports, then the grade of membership of that issue in the category sports would be  $2/8$  or  $0.25$ . The main advantage of this method is that this measure takes into account the engagement of the producer in each category. The greater the space dedicated to a certain topic the higher the grade-of-membership. We can also see that this method allows for the sum of the grades-of-memberships to be equal to one for all newspapers. This means that we remain faithful to previous theoretical conceptualizations of the fitness curve having an equal area, and that an increase of engagement in one category will lead to a decrease in another. Now that we have a grade of membership for each issue in each category, we follow the lead of Hannan, Pólos, and Carroll (2007) and Hsu, Hannan, and Koçak (2009) and define the niche width using Simpson's index of dissimilarity as

$$1 - \sum_{i=1}^I \mu^2$$

This measure takes values between zero and one with a zero indicating that the entire issue is dedicated to a single category. Increasing values indicate an increase in the diversity of the newspaper and hence its niche width. This niche width does not make any assumptions above the size of organizations. It is possible for a large organization to publish a newspaper that is dedicated only to economics for example, and for another small organization to publish a newspaper that deals with several different topics. What determines the niche width is the number of categories and the engagement in each of these categories. A newspaper that has a grade of membership of 0.9 in one category and 0.1 in another has a smaller niche width than one that has a grade of membership of 0.6 and 0.4 in the same categories.

Now that we have the grade-of-membership and the niche width of the issue, we need to determine to which cluster the issue belongs. To do this I use the statistical method of Cluster Analysis (Everitt, Landau, Leese, and Stahl 2011). According to Everitt, Landau, Leese, and Stahl (2011):

*“Cluster analysis techniques are concerned with exploring data sets to assess whether or not they can be summarized meaningfully in terms of a relatively small number of groups or clusters of objects or individuals which resemble each other and which are different in some respects from individuals in other clusters”(p. 13).*

In the cluster analysis presented in this thesis, I used as a measure of proximity the Pearson correlation measure and used the weighted-average linkage as the linkage method. The reason for these two choices is as follows: first with regards to the proximity measure, Everitt, Landau, Leese, and Stahl (2011) note that the nature of the data should determine the choice of this measure. In this data set, I

am using cluster analysis on variables that refer to proportions. The variables include the proportion of the issue that is dedicated to politics, the economy, sports, and so on. These variables are all measured on the same scale. They lie between zero and one. In such a case it makes more sense to talk about the correlation between the data than to talk about the distance. The variables provide an indication of the “relative profile” of the observation. The correlation between observations  $i$  and  $j$  is found using the following formula:

$$\phi_{ij} = \frac{\sum_{a=1}^p (x_{ia} - \bar{x}_i)(x_{ja} - \bar{x}_j)}{\{\sum_{a=1}^p (x_{ia} - \bar{x}_i)^2 \sum_{b=1}^p (x_{jb} - \bar{x}_j)^2\}^{1/2}}$$

Where  $\bar{x}_i = \sum_{a=1}^p x_{ia} / p$  and  $p$  is the number of variables (Stata Corporation 2011b, p. 506). Since the correlation takes on values in the interval  $[-1,1]$ , it is converted into a distance measure which takes on values in the interval  $[0,1]$  using the formula  $\delta_{ij} = (1 - \phi_{ij})/2$ . Unlike the Jaccard index mentioned previously, this measure takes into account the engagement of each producer in each category. The difference between an issue that dedicates all of its space to politics and a second issue is different than the difference between one that dedicates 40% of its space to politics and the same second issue. With regards to the choice of the linkage method, the weighted-average method “gives equal weights to each cluster regardless of how many observations it contains. Such methods consequently tend to work better for detecting clusters of unequal size” (Hamilton 2009, p. 358). Everitt, Landau, Leese, and Stahl (2011) state that such a linkage method can be used with both distance and similarity proximity measures.

The above formulation has several desirable properties that overcome the limitations of including a distance-based measure on the category level. First, the grade-of-membership is dependent on the producers' engagement and the area under the fitness curve is constant and equal for all producers. Second, the proximity measure utilized also depends on the proportion of resources dedicated to categories. It is not enough to state that a certain organization spans a certain category in order to calculate its distance from other categories. Third, the determination of the best-fit number of clusters (which will be described in detail in the following chapter) depends on all the information present, not just on information pertaining to certain categories. Changes in categories and organizations that take place outside the focal categories affect the best-fit number of clusters, and hence affects the assignments of all producer to the clusters. The determination of the best-fit number of clusters takes into account the differences between the category dynamics as well as the similarities (Everitt, Landau, Leese, and Stahl 2011), thus all the information is utilized to determine the number of clusters. This method stresses the fact that distance is a relational concept and that it should not be determined in vacuum. Fourth, we no longer have the problem that a rise in generalists would lead to a decrease in the niche width of the generalists since we have uncoupled the concept of niche width from the concept of distance. Fifth, if an organization belongs to a cluster which is made up of generalists and then decides to exist a certain category, it is no longer the case that the other organizations are considered as the non-conformists and thus are penalized. Instead, what will happen is that the organization which exits the category, and therefore is the real non-conformist, ends up in another cluster.

### **2.3.3 Revisiting Category Spanning: Formulation of Hypotheses**

Given the above theoretical formulation, we can revisit the topic of category spanning in order to have a clearer understanding of why organizations span categories, and whether the act is hazardous, and if so, under what conditions. The most important concept presented above is that niche width and the category space must be separated. By splitting these two concepts, it will be possible to study two different dynamics: category spanning and cluster spanning. In this section I will try to formulate hypotheses regarding these two dynamics.

Theoretically speaking, category spanning is linked to the concept of niche width. The larger the number of categories spanned, the greater the niche width of the organization. If spanning more categories leads to a decrease in the engagement that the producer dedicates to each category, why do producers seek to span more and more categories? Past research has shown that after a population has become legitimated, more and more pressure is placed on producers in order to distinguish their offerings from those of their competitors (Navis and Glynn 2010). The result is that producers expand to new categories. In the case where the market is susceptible to economies of scale, this pattern is further enhanced because in addition to distinguishing their offers, category spanning will also drive down costs since the more categories a producer spans, the larger the number of audience members it attracts. This pattern will eventually lead to the rise of generalists and the market will become dominated by organizations that span as many categories as possible. In the words of Carroll (1985), "In newspaper publishing, the dominance of generalists in the market represents the equilibrium state of a long, competitive process dictated

by economies of scale” (p. 1268). Therefore, we would expect to see that organizations expand their offerings by entering new categories during the period of the “rise of the generalists”. This rise of the generalists will take place after the industry has successfully formed, since it is after the process of legitimation that organizations find themselves under more and more pressure to distinguish their offerings (Navis and Glynn 2010).

***Hypothesis 1:*** *After the industry has successfully formed, the rise of generalists will be caused by incumbent organizations entering into new categories that they previously did not span.*

According to the Resource Partitioning theory, this increase in the number of generalist organizations will lead to a highly competitive landscape in which only a few generalists can exist. By expanding their offerings in order to benefit from economies of scale and in order to attract as many audience members as possible, these organizations will find themselves competing with other large organizations in the same resource space. Carroll (1985) has shown how such dynamics will lead to a portion of the resource space being abandoned by the generalists, who concentrate all their efforts in the center of the resource space. The abandoned edges of the resource space will attract new organizations that tailor their offerings in such a way that fills a small portion of the resource space. By doing so, these new specialist organizations will avoid competition with the large and established generalists, while at the same time they would manage to service audience members that are currently being neglected by the generalists. Thus, the Resource Partitioning model predicts that increased



concentration among generalists enhances the life chances of specialized newspapers (Carroll 1985). Boone, Witteloostuijn, and Carroll (2002) have shown that concentrated markets prove to be fertile areas for the operation of specialist newspapers. Although the process of resource partitioning is seen in many industries, the mechanism through which it takes place might differ<sup>2</sup>. Carroll's model refers to what is called "resource release", where generalists forgo certain areas in the resource space in order to concentrate on a larger central area. Swaminathan (1998), on the other hand, has shown that resource partitioning took place due to socially constructed images of producer authenticity. He showed that the greater the degree of concentration in the brewing industry, the higher the founding rate of microbreweries and brewpubs. I would expect that, after the generalists have dominated the market for a while, specialist organizations would start entering the population. According to Carroll (1984), "...Hawley's model predicts that the losing competitors are transformed; the resource partitioning model predicts that those units die and that the differentiated units come from new sources" (1278-79). Therefore, these specialists organizations are mostly made up of new organizations that seek to occupy a narrow niche that they have noticed is not being properly served. I do not expect that incumbent generalists decrease their category spanning activities in order to become specialists and escape from the intense competition. It is not only organizational ecologists that have presented an explanation for the segregation of the market between generalists and specialists. The sunk cost theory of Industrial Organization (Sutton 1991) highlights the importance of economies of scope in market segregation (Boone and Witteloostuijn 2004).

---

<sup>2</sup> I would like to thank Dr. Gabor Peli for brining this point to my attention.

While the two schools of thought offer different explanations, they both acknowledge and seek to explain the same phenomenon.

***Hypothesis 2:*** *Once the industry has matured, generalists will continue to be generalists while newly founded organizations will be founded as specialists.*

The above two hypotheses serve to explain when, and why, do organizations increase or decrease their category spanning dynamics. Based on the above discussion, I do not believe that we can say, "Category spanning is hazardous". It is well documented that category spanning is alive and well in many industries. Smith (2011) has argued that, "Nevertheless, atypical organizations not only persist in many markets, they sometimes elicit significant attention and thrive" (p. 61). Category spanning helps organizations distinguish themselves, and helps them benefit from economies of scale. The above two hypotheses try to explain that if we follow the reasoning of the Resource Partitioning theory, then we would expect that the dynamics of category spanning be related to the age of the industry. The effect of environments on organizations has long been a topic of interest and has received much attention from organizational theorists (Aldrich 2008; Aldrich and Reuf 2006; Lawrence, Lorsch, and Garrison 1967; Pfeffer and Salancik 2003). Both exogenous and endogenous environmental factors can cause a shift in the effect of the niche width of organizations (Swaminathan 1998). One of the greatest examples of a change in exogenous factors that favored generalism as opposed to specialism was when General Motors overtook Ford to become the largest automobile company. This was made possible by several changes in the environment such as

the newly created availability of easy credit and a shift in consumer preferences (Sloan 1964). Resource-partitioning theory describes how endogenous changes can cause a significant shift in the effect of the niche width. Hannan and Freeman (1984) have argued that certain environments clearly favor generalist organizations. Many factors can help favor one form over the other, such as economies of scale (Dobrev, Kim, and Carroll 2002) and network embeddedness (Echols and Tsai 2005). More importantly, the timing in the industry evolution has proved to be critical in determining the effect of niche width (Dobrev, Kim, and Hannan 2001). Research has shown that with time, the effect of niche width on organizational mortality and on the founding rates of different organizational forms is reversed. Peterson and Berger (1975) and Swaminathan (1998) show how founding rates of specialist organizations increase as the concentration of the industry increases, while Dobrev, Kim, and Hannan (2001) and Dobrev, Kim, and Carroll (2002) show that as market concentration increases the effects of niche width on the hazard rate changes from negative to positive. Therefore, I believe that the effect of category spanning, which determines the niche width, is not uniform throughout the lifetime of the industry. I believe that, as the Resource Partitioning theory predicts, at certain stages in the industry lifetime, the environment would favor generalists who are trying to distinguish themselves, or who are benefiting from economies of scale. During these periods, I do not believe that spanning categories will be hazardous to organizations.

***Hypothesis 3: The effect of category spanning is time dependent.***

However, once the industry has matured, the severe competition between generalists in the middle of the resource space will make it increasingly harder for generalists to survive. It is during this period that specialist organizations start forming. By not spanning many categories, these specialists will avoid direct competition with the generalists. If, however, an organization enters as a generalist, then it will be faced with severe competition from incumbent generalists, and therefore suffer from a high mortality rate. In addition, the intense competition for the center of the resource space will result in the death of some generalists since only a limited number of generalists can survive in the market (Carroll 1985). So overall, I predict that during this time period, we will see an increase in the hazard of mortality for organizations that attempt to span more and more categories.

***Hypothesis 4: Category spanning increases the hazard of mortality or organizations once the industry matures and specialists start to emerge.***

The above four hypotheses link the dynamics of category spanning to the Resource Partitioning theory. However, this thesis has cited many papers that have found category spanning to be hazardous in some cases, and not harmful in other cases during the same time period (Kovacs and Hannan 2009; Pontikes 2008). The above hypotheses do not offer an explanation to this phenomenon. It is here where the concept of clusters becomes useful. As explained above, cluster formation depends on the category dynamics in all the resource space, and not just between two focal categories or producers. It is important to note here that cluster spanning and category spanning are two separate things that can, in

some instances, involve the same act. It is possible for an organization to expand its contents by spanning a new category. The result may be that this organization ends up in a new cluster. However, this is not always the case. For example, in a category space in which only two clusters exist, it might be the case that increasing the number of categories spanned by a certain organization does not warrant an increase in the number of clusters or even a shift to the other cluster. In other words, it is possible that category spanning will not have an effect on cluster spanning. In other cases, it is possible for an organization to span a new cluster by decreasing the number of categories spanned. For example, an organization that is a generalist and decides to become a specialist will end up in a new cluster in the category space. Finally, it is possible for an organization to span a new cluster even if the number of categories that it spans remains the same. Specialist organizations have considerable freedom in their choice of “location” (Dobrev, Kim, and Carroll 2002; Dobrev, Kim, and Hannan 2001). A specialist organization that spans only a single category might choose to remain a specialist organization but in a different category (switching from a political publication to an economical one). In such a case the number of categories is still one, but the organization would belong to a new cluster.

Among the reasons proposed for the negative effect that category spanning has is that it confuses audience members. I propose that audience members are confused when producers span clusters. I believe that the act of moving from cluster to cluster is hazardous to the organizations health because audience members will have difficulty in determining to which group the organization belongs and, therefore, they will view it negatively. Clusters are a representation of the category space. When a producer moves to a new cluster,

then it will occupy a new position in this category space. This movement is determined by many factors, not just by whether the producer spanned more categories or not. Because the number of clusters is determined by using data from all producers, it might be the case that a producer ends up in a different cluster simply because the majority of organizations changed their category spanning dynamics. In such a case, the entire category space shifts and, even though the focal producer did not change its category spanning dynamics, it ends up in a new position. Since audience members view the position of the producer in relation to the position of all other producers, the end result will be that the producer's position has changed. Whether this change was due to its actions or the actions of other organizations does not matter. Another reason that is given to explain the hazardous effect of category spanning is that the organization that attempts to span a new category is not conforming to the "rules of the game". An organization that crosses boundaries must be penalized, especially in a market in which boundaries are clearly delineated. However, as discussed in the section before this, it might be the case that the norm in the industry is to span many categories. Many studies have talked about a certain industry being crowded by generalist organizations. In such cases, it is the specialist organization that does not conform to the norms. Therefore, it is the organization that does not span categories that must be penalized for being the nonconformist. Here again, clusters help capture these dynamics. If an organization, by following certain category dynamics such as becoming a specialist, ends up exiting its current cluster and creating its own cluster while all other organizations remain in the original generalist cluster, then this act of moving from cluster to cluster is going to be hazardous because it signals nonconformity. Finally, the work of Kovacs

and Hannan (2011) explains that spanning distant categories is hazardous. While I follow their lead in taking into account the resource space, I do not mix the concepts of niche width and category distance together. I account for distance by using clusters that represent the category space. Since distance is determined by the correlation between observations in this thesis, and the best fit number of clusters is determined by the correlation that all observations have with all other observations, two different clusters represent two different sets of correlations and therefore the distance between a producer in one set and another producer in another set is greater than the distance between two producers in the same set. This means that spanning clusters also captures the concept of distance. Instead of saying, "Spanning distant categories is hazardous", my model says, "changing your category dynamics relative to the dynamics of other producers in such a way that warrants a shift from cluster to cluster is hazardous". This is similar to the argument made by Ruef and Patterson (2009), which states that rare combinations of components would increase the hazard faced by organizations. Here, the word "rare" indicates a relative concept. In my model, spanning a combination of categories that is "rare" might lead the focal organization to end up in a new cluster of its own creation. Therefore, given all of the above, I believe that cluster spanning dynamics capture the essence of several proposed explanations for the detrimental effect of category spanning. In addition, unlike the issue of category spanning, the effect of cluster spanning is not time dependent because we are not confounding two different issues: niche width and "confusion".

***Hypothesis 5:*** *Organizations that span clusters will have a greater hazard of failure than those that do not span clusters. This negative effect is not time dependent.*

This thesis has described how category spanning is different than cluster spanning, even though in some cases they may be caused by the same action. I have argued that the effect of category spanning is time dependent, while the effect of cluster spanning is uniform because it is confusing to audience members and because it might signal nonconformity. To test the above hypotheses, the models will need to control for both dynamics at the same time. This thesis will attempt to show that category spanning is hazardous in two cases. As proposed by hypothesis 4, category spanning is hazardous when the industry has matured and generalists have taken over. This results in the founding of specialists. In addition to this, category spanning is hazardous when it results in cluster spanning. As described above, not all cluster spanning is a result of category spanning, but in those cases in which this is actually the case, category spanning will be detrimental to the mortality of the organization. This helps explain why previous research has found that category spanning is not always hazardous. Unless generalists dominate the industry, and unless the category spanning results in cluster spanning, category spanning will not be hazardous. In fact, if the environment favors generalists then category spanning will have a positive effect on the organizational life chances.

The above hypothesis deals with the effect of category spanning on the organization that engaged in that act. One reason that was proposed for this negative effect was that audience members would be confused when an



organization spans a category (Leung and Sharkey 2009). I have attempted to show that confusion arises out of cluster spanning, since category spanning is a phenomenon that is related to the issue of niche width and therefore is more or less determined, and affected, by environmental variables such as the age of the industry. If an organization is penalized for spanning clusters due to the fact that it confuses audience members, what can we say about the total level of cluster spanning in the population during a certain time frame? I have argued above that the dynamics of organizations that belong to a certain population have an effect on other organizations. Therefore, is a high level of cluster spanning activity within a population detrimental to the health of all organizations that are active within that population, disregarding whether these organizations have actually spanned clusters themselves or not? I propose that a very high level of cluster spanning within a population would confuse the audience members about the entire population. We can determine the level of cluster spanning activity during a specific period and thus study the effect that periods with high cluster spanning dynamics have on organizations. Some research has argued that confusion during periods of industry emergence will lead to a failure in legitimation and hence a failure in emergence (McKenrick, Jaffee, Carroll, and Khessina 2003). If audience members, who are looking at an industry and trying to make sense of the activities of the producers within that industry are met with severe shifts in cluster formation and cluster dynamics, then this will confuse the audience. If the level of confusion is very high due to a high level of cluster spanning activities, I believe that all organizations within the population will suffer from this confusion. As mentioned above, clusters are determined by the total activity, and if the cluster dynamics around an organization keep changing,

then, due to the socially constructed nature of clusters, the position of the organization in the category space will also appear to be shifting, even if that certain organizations is not changing its behavior. I believe that organizations that are operational during periods that involve a large degree of cluster spanning will be exposed to a higher hazard of failure than organizations that are active during periods with a lesser degree of cluster spanning due to the fact that severe shifts in cluster formation will confuse audience members with regards to the industry as a whole.

***Hypothesis 6:** Organizations that are active during periods that witness a high level of cluster spanning will be subjected to a higher hazard of failure.*

## **2.4 Literature Review – Industry Emergence**

Organizational ecologists have managed to develop sophisticated tools which they have used to explore the dynamics of a wide variety of populations (Aldrich and Reuf 2006). However these tools have been mostly applied to stable, well-defined populations (Astley 1985). But what about populations that are not well-defined and/or stable? More specifically, can organizational ecologists study the emergence of new organizational forms which are characterized by instability (Navis and Glynn 2010) and incoherence (Santos and Eisenhardt 2009)? Previously, the emergence of organizational forms was seen through the lens of institutional theory (Hayagreeva 2004), evolutionary theorists (Aldrich and Fiol 1994), and community ecology (Ruef 2000). The theoretic advances made by Hannan, Pólos, and Carroll (2007) enable us to

examine the emergence of organizational populations using the concepts of categories, social code, membership, and identity.

The new theoretic formulation presents two significant departures from previous theories. First, legitimation is no longer defined purely as the number of organizations in a specific category. Instead, legitimation is achieved when members of a category abide by codes that the audience uses to make sense of the actions of organizations. Categories may therefore have a high number of members that nonetheless the audience has difficulty in making sense of, therefore causing the category to have what is called “low contrast”. Categories in which members follow similar rules manage to distinguish themselves from their surroundings and therefore achieve “high contrast” (Hannan 2010). Second, unlike previous formulations (e.g. Albert and Whetten (1985)) identity is no longer perceived as being determined solely by internal members (Hsu and Hannan 2005) or by internal enduring attributes (Wry, Lounsbury, and Glynn 2011) and features (Hannan, Baron, Hsu, and Koçak 2006). Ultimately, the identity of the organization will consist of the social codes which audience members use for the categories of which the organization is a member of (Negro, Kocak, and Hsu 2010; Pólos, Hannan, and Carroll 2002). Previous studies have shown that identity plays an important role in many dynamics. Recent research has shown that identity, not quality, is what drives resource partitioning in established markets (Carroll and Swaminathan 2000; Liu and Wezel 2013), and that the identity of the organization plays a pivotal role in the success of the product (Jensen 2010; Khessina and Carroll 2008), and inevitably the firm as a whole (Carroll, Bigelow, Seidel, and Tsai 1996). More importantly, research has also shown that the emergence of new organizational forms is heavily dependent

on the identity of the entrants, not on their total count (McKendrick and Carroll 2001; McKendrick, Jaffee, Carroll, and Khessina 2003).

Producers, especially in nascent markets (Santos and Eisenhardt 2009), attempt to redefine categories such that their product is considered as the prototype (Navis and Glynn 2010), to identify their rivals (Kocak, Hannan, and Hsu 2009), and to keep the status quo in their favor (Lounsbury and Rao 2004); while consumers, especially gatekeepers (Hannan, Pólos, and Carroll 2007) or activists (Kocak, Hannan, and Hsu 2009), tend to classify products into different categories because this classification is the basis on which audience members make choices (Beck, Swaminathan, Wade, and Wezel 2009; Biggart and Beamish 2003; Khaire and Wadhvani 2010; Lounsbury and Rao 2004; Wei 2005). Categories are especially important when audience members do not have full information regarding certain producers. By typecasting, audience members are able to apply defaults to these producers thereby filling in the missing information (Hsu, Hannan, and Pólos 2009).

Work by McKendrick and Carroll (2001) and McKendrick, Jaffee, Carroll, and Khessina (2003) highlighted the fact that the period of emergence of industries was a fertile ground for research. Since the identity of organizations is formed, in part, from categories of which they are a member (Hsu and Hannan 2005; King, Clemens, and Fry 2011; Wei 2005), this means that organizations in emerging categories need to display a unified identity (Hsu, Hannan, and Pólos 2009; Lamont and Molnár 2002). Failure to do so would confuse the audience (Hannan, Pólos, and Carroll 2007). With time, as audience engagement increased, organizations would have to differentiate themselves from their

competitors, thus there will be an increase in the subtypes of the new category (Kocak, Hannan, and Hsu 2009).

## **2.5 Development of Research Problem – Industry Emergence**

Two problems seem to arise from the above line of reasoning. First, in the work of McKendrick and his colleagues, identity is measured on a single dimension, which in this case is the origin of the company. However, identity, especially collective identity, is best viewed as multidimensional (Ashmore, Deaux, and McLaughlin-Volpe 2004). Second, looking at identities alone may prove to be too restrictive. Perretti, Negro, and Lomi (2008) found that the effect of identity is context dependent. Their research led them to conclude that “...different domains offer alternative default settings, and the way a new organizational form emerges is in the context of interpretation and social classification of the identity of a domain as well as the identity of candidates” (p. 543). Not surprisingly therefore, recent research has argued that a unified identity was not a necessary condition for the emergence of new categories (King, Clemens, and Fry 2011). In fact, Weber, Heinze, and DeSoucey (2008) have criticized the fact that initial variation in organizational forms have been under examined. This raises two questions: First, how do organizational identities in an emerging industry form? And second, what are their dynamics and their effects? A discussion of these questions must be preceded by an elaboration of how organizational identities are determined. Smith (2011) used a conceptualization of identities where he categorized organizations into conformists and non-conformists by measuring the distance of the organization from the “mean”,

Powell and Sandholtz (2011) highlighted firm-level attributes which they believed to reflect distinct domains of origin and function, while King, Clemens, and Fry (2011) determined identity by identifying elements which related to theme, resources and services, and target population. Other studies looked at the “stories” told by organizations in order to identify their identity (Wry, Lounsbury, and Glynn 2011). A multidimensional view of identity is yet to be incorporated into the study of population level dynamics.

I propose that in order to truly understand the dynamics and effect of identity, the researcher will have to take a multi-dimensional view of the concept. The researcher will have to study the context, the stories, and the category spanning dynamics. Therefore, this paper will study the identity of the Lebanese newspapers during the period 1851 – 1879 using two tools. The first tool is that of social movement analysis, and the second tool is that of categories and clusters. First, a close inspection of the early period of the newspaper industry in Lebanon shows that the industry lends itself well to the tools of social movement analysis. This thesis will show that the early newspapers formed a social movement which, unlike previous formulations of social movements, was dependent on both resource mobilization and audience grievances. The thesis will also show that grievances were an important part of the “framing” process (Benford and Snow 2000; Goffman 1986), and that the founders of the newspaper industry in Lebanon used grievance framing as a powerful tool. In order for social movements to be successful, they will have to develop collective action frames (Benford and Snow 2000) that will help in bridging the social distance between producers and consumers (Lounsbury, Ventresca, and Hirsch 2003; Weber, Heinze, and DeSoucey 2008). It will become clear that in the

Lebanese newspaper industry, in most cases, the founders framed the problems facing the society in a unified way and identified similar goals to each of their newspapers. By describing these framing processes, this paper will hopefully show that all participants projected the part of the identity of the newspapers that dealt with how the founders saw their movement in a very similar and coherent manner.

The qualitative study of the emergence of the newspaper population will provide us with an understanding of the context in which the industry emerged. It will also allow us to study the “stories” told by these newspapers in order to legitimate their coming to life. By studying these stories, we will be looking at that part of identity that these newspapers wanted to project to the audience members. This thesis will show that as a social movement, the newspapers in Lebanon told stories that fit the historical context in which they were situated. In other words, the newspapers had to justify their existence by claiming to accomplish a goal that was related to the problems facing the people during that time. If the organizations tried to project different goals and different reasons for their emergence then the emergence process would be jeopardized.

Perretti, Negro, and Lomi (2008) have shown that with regards to de alio and de novo entrants, the context plays an important role in determining what the best identity of new entrants is. In the same spirit, I will show that the identity of the founders of the early newspaper organizations is in-line with the context present at the time of the emergence of the Lebanese newspapers industry. Since the context was an important factor in determining the stories told by the early newspapers, I believe that the background of the founders was consistent with the stories told by their newspapers.

The second tool that will be used is that of category spanning and the subsequent cluster formation. As mentioned above, current theoretical formulations claim that the identity of organizations is partially determined by the categories that they span. Therefore, by studying category spanning and cluster formation, we will be looking at another dimension of identity. While I intend to show that the identity that was related to the stories and the context had to be unified, this is not necessary the case with regards to the identity that is determined from the categories spanned. King, Clemens, and Fry (2011) have shown that organizations were best grouped into more than one form, while McKendrick, Jaffee, Carroll, and Khessina (2003) argued that a unified identity was a necessity of successful emergence. This thesis intends to contribute to this debate by applying the tools of cluster analysis in order to study how the early newspapers spanned categories. Previous research has used the organizations' background as a determinant of identity or the attributes which the organizations claim to possess. I, on the other hand, will use category dynamics in order to study how the partial identities have formed during emergence.

***Hypothesis 7A:*** *The founding newspapers formed a single cluster.*

***Hypothesis 7B:*** *The founding newspapers formed more than a single cluster.*

Ruef and Patterson (2009) show that category heuristics were less developed during the initial industry emergence period. Aldrich and Reuf (2006) state that the conditions faced by the founding fathers are less secure than the conditions faced by their predecessors. Based on these two lines of thought, I



believe that the first newspapers spanned categories in a more haphazard manner than later ones. The fact that these early newspapers had no precedent to build upon and that category heuristics were yet to become well developed lend support to this belief. Although I expect that newspapers told the same stories and used a unified framing process, this does not mean that they were in agreement with regards to category spanning dynamics. While the problems and goals were uniformly identified, I believe that the content of the newspapers was not a matter of agreement. In fact, I do not even believe that issues belonging to the same newspaper spanned categories uniformly because I do not believe that the individual newspapers had a clear formulation of the type of content that they intended to include. With time, I believe that the newspapers were able to concentrate their efforts on certain categories that they believed reflected their identity and helped them reach their goals. This “maturity” spread among organizations due to the fact that organizations monitor one another (White 1981).

***Hypothesis 8:*** *During the emergence period, newspapers tended to enter new categories and exit old ones in a more random way than later newspapers.*

Because cluster formation depends on the category spanning dynamics, I believe that during the emergence period, the early newspapers created more clusters than the later newspapers due to their haphazard category spanning activities. Severe shifts in category dynamics will result in significant changes in the category space. As time goes by, newspapers should develop more systematic dynamics due an increase in their knowledge of the market and also due to an

increase in scrutiny by external audience members. Therefore, I believe that the number of best-fit clusters during the later period is less than the number during the early period. In addition to forming a larger number of clusters, I believe that the early newspapers were more likely to span clusters than later newspapers. Again, this is due to their random acts of entering and exiting categories that were mentioned in Hypothesis 8. With time, as individual newspapers developed a better understanding of what type of content they wanted to include, successive newspaper issues tended to end up in the same cluster. This can be thought of as a reflection of expansion in knowledge and maturity on the part of newspapers.

***Hypothesis 9:*** *With time, newspapers started forming less clusters during the emergence period.*

***Hypothesis 10:*** *Early newspapers were engaged in more cluster spanning dynamics than latter ones during the emergence period.*

Note that hypotheses 7 to 10 deal with the part of identity that is related to category spanning activities, while the qualitative section talked about the “stories” told by newspapers. By projecting identity on two dimensions, we can separately study whether organizations behaved in the same manner on these two dimensions. While I expect that organizations behaved similarly with regards to the stories told because, as mentioned above, and as will be explained in more detail in Chapter 4, social movements need to uniformly frame their actions. Any deviation would be detrimental to their collective fate during the

period of emergence. However, this is not necessarily the case on the second dimension. Past research has produced conflicting results. It should be noted however, that past research did not study identity using the category spanning activities. Therefore, this thesis will try to contribute to this debate by presenting a new dimension in the study of identities that is based on the work of Hannan, Pólos, and Carroll (2007). As can be deduced from hypotheses 7 to 10, I expect that, with time, newspapers started to behave more systematically with regards to category spanning dynamics. The question remains though whether by the end of the period of emergence the category space was made up of a single unified cluster, or more than one.

## **2.6 Conclusion**

This chapter started with a discussion of recent theoretical developments in the field of organizational ecology. This was followed by a discussion of current points of contention relating to the topic of identity during industry emergence and the concept of category spanning throughout the life history of the industry. The main aim of this chapter was to highlight the areas that this thesis will try to contribute to. The rest of the thesis is organized as follows: Chapter 3 will provide an overview of the statistical methods that will be used to test the various hypotheses which were developed in chapter 2. The chapter will offer an explanation as to why the statistical methods were chosen in addition to a description of these methods. Finally, the chapter will show how these various methods are interlinked together. After that, and in order to present the material in its proper chronological order, the thesis will start with examining the hypotheses relating to the period of the emergence of the industry 1851-

1879. Chapter 4 will provide a review of the social movement literature and will then move on to show that the Lebanese newspaper industry emerged as a social movement. In the process, the chapter will also show how the goals of the newspapers were developed within the historical, economic and cultural context of the era. Chapter 5 will move on to study the dynamics of category spanning and cluster formation during that same period. Both chapters 4 and 5 study the identity of the newspapers during the period of emergence but from different perspectives, or on different dimensions. After that, chapter 6 will deal with the issue of niche width and cluster spanning during the entire period of 1851 – 1974. Finally, chapter 7 will serve as the conclusion where the results will be discussed within a single context along with a description of what these results imply for future research.

## **3 Chapter Three: Methodology and Statistical Models**

### **3.1 Introduction**

Chapter 2 provided the theoretical underpinnings of this thesis and presented the hypotheses that are to be tested. Chapters 5 and 6 will test these hypotheses. In order to do so, several statistical techniques and models will be used, especially in chapters 5 and 6 since chapter 4 will be more concerned with qualitative reasoning and analyses. In this chapter, I will introduce the quantitative models that will be used to test the hypotheses of chapters 5 and 6 and will explain the reasoning that led to them being chosen. This chapter will describe how the data was collected, the statistical methods of Content Analysis, Cluster Analysis, simulation experiments, multi-level logistic models, and Survival Analysis (a.k.a. Event History Analysis). Finally, the chapter will show how all these models are interrelated.

### **3.2 Data Collection and Content Analysis**

This thesis studies how organizations span categories and the effect that this spanning activity has on the mortality of the organizations. In order to do so, I had to gather information about both the life history data and of the category spanning dynamics. Therefore the data collection process was divided into two parts. In the first stage I collected the life history data of all the Lebanese newspapers from 1851 up to and including 1974. Unfortunately there was no

single source that contained such information so I had to rely on several sources. The primary sources for this stage were Tarazi (1933), Dagher (1978), Khoury (1990b), and Illias (1997). I then used Mruwi (1961), al-Majlis al Thaqafi al-Lubnani(1996), Khoury (1999), and Taleb (2011) to fill in some missing information. Another important source of information was the Lebanese Official Journal that is published by the Lebanese government and contains new rules and regulations including those that pertain to the newspaper industry such as the issuing of new licenses or the modification of the license to allow a specific journal to include political content. The journal also contains rules that penalized publications. This information was useful in cases where the ending dates were missing because they allowed me to censor the entries at certain dates since the rules meant that they were in publication at that time. Other important sources of censored data are Naqabat Muhariri al-Sahafa al-Lubnaniya(1943) and Wizarat al-Anba'(1950). Both of these sources are lists of active publications in 1943 and 1950 respectively. In very few cases I managed to secure sources that gave detailed information about a single newspaper. For example, Ghaleb (1988) provides us with a detailed history about the longest living Lebanese newspaper Lisan al-hal while Khoury (1990a) has made available to us an extensive collection of al-Jam'iyah al-'ilmiyah al-Suriyah. Finally, I conducted several interviews with industry historian Joseph Illiyas in order to fill in some of the missing information that I was not able to find in any of the sources. By the end of this stage I was able to collect information about the starting date, ending date, format (newspaper/magazine) and frequency of publication. A publication was considered a magazine if there was a table of contents at the beginning of the issue, or if the page numbering was continuous, i.e. if the first page in a magazine

started off where the last page of the previous issue stopped. I was also able to record dates of when the ownership changed hands, or when the journal changed its format. One issue that presented itself is how to treat multiple death events. If a newspaper stopped operating for ten years and suddenly appeared again, should I consider it the same newspaper? In order to systematically treat all such events in a uniform manner I decided that if a newspaper disbanded at some point and re-emerged later, I would treat it as the same newspaper only if it retained the same format, title and owner. If however, for example, the newspaper was re-issued by a new owner after a period of death then I considered it to be a different organization. Therefore, unlike previous studies of organizational survival studies, my datasets allows for multiple occurrences of disbanding events. This makes sense in the present context because during the First World War for example many newspapers stopped operating for several years before resuming after the war. It does not make sense to consider the issues published after the war to be originating from a different organization than the issues before the war, and neither does it make sense to consider that the newspapers were operational for the entire period of time while they were clearly not operational for four or five years.

By the end of this stage I had collected the following information: there is a total of 1759 newspapers, with 1056 being published in Lebanon and 703 outside Lebanon. Around 81% of the publications in Lebanon were published in Beirut while around 35% of publications outside Lebanon were published in Egypt. With regards to the publications in Lebanon, there are 59 cases of missing founding dates and 427 cases of missing disbanding dates. Out of these 427 cases, there are 164 cases of censoring before the end of the study. This means

that while I did not find the ending dates of these publications, I managed to secure dates at which I knew that they were still in circulation. Finally, there are 119 cases of censoring due to the study ending at the end of 1974.

Chapters 5 and 6 cover different time periods, but with regards to the data collection both chapters relied on the same technique and the same source. All the newspaper issues which were examined for this thesis are available in one of two locations: the Jafet Memorial Library at the American University of Beirut and the Oriental Library at the Saint Joseph University, both of which are located in Beirut. These two sources have overlapping data, and I chose to use the Jafet library as the main source. This choice was facilitated by the fact that I am an alumni member of the American University of Beirut and therefore need no special permission to access the library records, in addition to the fact that the newspaper collection in Jafet library has a more advanced online search system which makes searching for certain issues easier. Most of the issues at the Jafet library are stored on microfilms like the one shown in Figure 3.1 below. A small proportion of the issues were only available as hard copies. In contrast, all issues examined at the Oriental library were available only as hard copies.



Figure 3.1 A newspaper issue on micro film at the Jafet library



Chapters 4 and 5 study the period 1851 – 1879. All the data collected was obtained from the Jafet library since the entire collection of the Oriental library of that period is a subset of the collection at Jafet library. The table below shows the issues that were available for inspection.

<b>Newspaper</b>	<b>Life Span (censored at December 1879)</b>	<b>Issues Inspected</b>
Al-Jam'iyah al-'Ilmiyah al-Suriyah	January 1868 – May 1869	January 1868 – May 1869
Hadiqat al-Akhbar	January 1858 – December 1879	January 1858 – April 1868
Majmoo' al-Fawa'id	January 1851 – January 1856	January 1851–January 1856
Nafeer Suriyah	September 1860 – April 1861	September 1860–April 1861
al-Basheer	September 1870 – December 1879	September 1870–December 1879
al-Taqadom	January 1874 – 1878	January 1874 – May 1875
Thamarat al-Funun	April 1875 – December 1879	April 1875–December 1879
al-Jinan	January 1870 – August 1879	January 1870 – August 1879
al-Zahra	January 1870 – December 1870	January 1870–December 1870
al-Nahla	May 1870 – December 1870	May 1870 – December 1870
al-Najah	January 1871 – December 1874	January 1872–December 1874
al-Mishkat	April 1878 – July 1878	April 1878 – July 1878
Lisan al-Hal	October 1877 – December 1879	October 1877–December 1879
al-Janna	June 1870 – December 1879	July 1879 – August 1879
al-Moktataf	June 1876 – December 1879	June 1876 – December 1879
al-nashra al-Osbo'iyah	January 1871 – July 1877	January 1872 – July 1877
al-Tabeeb	January 1874 – December 1876	January 1874–December 1876
al-Sharaka al-Shahriyah	January 1866 – September 1866	January 1866–September 1866
Al-Jareeda al-Rasmiyah	1860 - unkown	None
Akhbar 'an intishar al-Injeel fee al 'alam	March 1863 – June 1868	None
al-Junayna	February 1871 - 1875	None
Kawkab al-Sobh al-Muneer	January 1871 – December 1879	None
Lubnan	June 1868 – 1869	None
Al-Mujama' al-Vaticany	January 1870 – August 1870	None

**Table 3.1 Newspaper issues available for inspection and their life span.**

Unfortunately, newspaper collections are not a random sample (Riffe, Lacy, and Fico 1998). However, this does not mean that we cannot, or should not, study such collections (Riffe, Lacy, and Fico 1998). Given the historical nature of the data and the fact that it is being collected from primary and not secondary sources, it is acceptable, and expected, that the entire collections could not be

located. As can be seen from Table 3.1, the size of the collection at the AUB is significant: the entire collection of 14 of the 26 newspapers is available, with a considerable proportion of another four newspapers also available.

Chapter 4 will study the collective identity of the newspapers during the emergence period. As mentioned in Chapter 2, Chapter 4 will study the stories told by newspapers in order to determine what their collective identity was. The stories refer to how the newspapers identified themselves in their editorials. These stories are specifically linked to the framing process. I went through the articles of the newspapers listed in Table 3.1 looking at how the newspapers identified the problems that were faced by society, and how the newspapers believed that they could alleviate or solve these problems. In the terms of social movement analysis, I was able to identify how the newspapers framed the problems that were being faced by society at that time.

The next step was to use Content Analysis in order to study the category spanning dynamics. This meant that I first had to identify the categories. The justification for the choice of categories is a question which must be addressed. In the case of this thesis, the justification is of a more practical nature than it is theoretical. While going through the newspapers, I noted the nature of the news stories and started classifying them. Items which were of a political nature or an economic nature were well expected in advance, but some other items came as a surprise. For example, I was surprised to see that many newspapers dedicated a significant portion of their space to literary issues that related to the Arabic language. This led me to decide to create a separate category for such items since their frequency meant that it would not be wise to disregard them by adding them to another more general category. In addition, the category "Art" was not

initially on my list of categories, especially since the early newspapers rarely mentioned anything that could fall in that category. However, later newspapers started dedicating more and more of their space to this category, and this led me to create a separate category. During the process of going through the newspapers, I was able to identify nine different categories. Since the identification of these categories is not based on theoretical reasoning, another researcher doing the same exercise to the newspaper industry in another country might identify different categories, depending on the sort of news items that were appearing in that country. The nine identified categories are :

Politics: This category consisted of news items that dealt with political issues such as the Sultan, ruling government, foreign representatives, and war. Also, if a news item was concerned with the death of a public official then this was also considered to fall under this category.

Economics: News items about the state of the economy, the well being of industries, and the prices of commodities fell under this category.

Social issues: This category included items that were concerned with social development issues such as the state of education, immigration and the need for better roads.

Knowledge: Items that fell into this category included pieces that were concerned with the spread of education and knowledge. Examples include articles about biology, math, philosophy and book reviews.

Literature: I included in this category all news items that dealt with Arabic grammar, poetry and novels. Although I could have included these items in the above category (knowledge), I thought that it was best to create a separate

category due to the importance of language in the emergence of the newspaper industry in Lebanon as will be shown in chapter 4.

Sport: All news items that were concerned with sports were included in this category.

Art: News about movies, the theatre, paintings and sculptures were included in this category.

Advertising: This category consisted of all items that advertised products created by people or companies.

Other: any item that could not be classified into any of the above groups was added to this group.

I sampled one article from each month for each newspaper title and went through it determining the space dedicated to each of the above nine categories. I read through the entire issue and used a ruler to determine the area of the page, since different publications had different page sizes. I multiplied the area of the page by the number of pages to get the total area of the publication. I then measured the area of the news items that belonged to each of the above categories. I then divided the area occupied by each category by the total area of the publication to get the ratio of the publication dedicated to each category. Photos were calculated as part of the area dedicated to the news item to which they belonged. So for example if a certain article contained a large picture and few lines of text I calculated the area occupied by the picture and the text as a ratio of the total area. Most of the issues were stored on microfilm as described earlier. After going through the microfilm collections, I coded the issues that were available in hardcopy. The final dataset consisted of 659 entries, with each entry representing one issue. One important issue here is the why the religion of

the founder of the newspaper was not one of the variables that were coded above. The reason is that while I was expecting religion to play a significant role in the period of emergence of the industry, my research uncovered that the newspapers made a conscious decision of not talking about religion. This decision was made because the founders of the newspapers wanted to create a unifying identity between them and the larger Muslim population. This was achieved by concentrating on newspapers as vehicles of knowledge and by concentrating on the Arabic language which everyone shared. This issue will become more visible in Chapter 4. At the same time, in Chapter 6, I perform some analysis to uncover any differences between publications made by Christians and those made by Muslims but the results indicate that there is no statistically significant difference. This result also supports my choice of not coding religion.

Chapter 6 utilized the same method to collect the data for the period 1880 – 1974. Coding the entire issues for the period from 1851 - 1974 took me around 11 months. In these 11 months, for six days every week, I did nothing but go through newspaper issues from the morning until the afternoon calculating the area occupied by each category. There was however one important point which had to be taken into consideration when studying such a long period of time. The content of the newspaper is probably dependent on the date in which it was published. For example, if an issue was published directly after a major political event, like the start or end of a war, then the newspaper will probably dedicate more of its space to politics. In order to minimize variances in the topics due to different events occurring at different dates I decided to choose certain dates for the issues to be analyzed. These dates were chosen based on the anniversary of the creation of whatever government was ruling Lebanon. The Mutasarifiya of

Mount Lebanon was created on 9 June 1861, the French mandate started on 1 September 1920, and finally the date of Lebanese independence is 22 November 1943. This means that for newspapers published before 1920 I selected the issue published on 9 June. For newspapers published during the French mandate, the date chosen was 1 September and finally, for issues published after independence the date 22 November 1943 was chosen. If the issue at the selected date was not found in the collections, I analyzed the issue that was closest to the date. The final dataset consisted of 2,478 entries, with belonged to 293 different newspapers. I used the same nine categories that I identified above and went through each issue determining the space that was dedicated to each category. Again, it is important to note that the data collection does not represent a random sample. I went through the entire collections of the Jafet Library of the American University of Beirut and the Oriental Library of Saint Joseph University. As I was going through these collections, I added and modified to the information that I collected in the first stage of the data collection process. Therefore, the dataset that I use here is the most complete single source of information about the Lebanese newspapers because it contains information from all major sources and also contains modifications based on my fieldwork. In the end, I had coded 2358 issues from the AUB and 120 from the USJ for a total of 2,478 issues<sup>3</sup>.

### **3.3 Cluster Analysis**

The next step after collecting the data and calculating the proportion of the space dedicated to each category is to determine how these observations (i.e.

---

<sup>3</sup> The USJ collection is much larger than what the number given suggests but there was considerable overlap between the contents of the two libraries. I started collecting the data at the AUB and did not consider the part of the USJ library that overlapped with what I had already collected.

the newspaper issues) group together. In order to do so I used Cluster analysis. In chapter 2 I presented the reasoning behind the adoption of the concept of clusters along with a description of the choice of proximity measure. In this section I will describe in more detail how the best-fit number of clusters is determined.

As described in chapter 2, Cluster Analysis is concerned with studying whether observations in a data set can be summarized in terms of groups or clusters. There are several methods that can be used in order to see how the data cluster together. At the most basic level, the researcher can use two or three-dimensional graphical displays in order to visually assess whether certain groups emerge from the data set. One of the problems with such methods is that they can only be used when the number of dimensions is small. In order to accommodate a larger number of dimensions, the researcher can use Principal Component Analysis (Pett, Lackey, and Sullivan 2003) in order to represent the data on fewer dimensions. While these techniques are useful during the preliminary stages of analysis, there exist more rigorous methods that do not rely on the researcher looking at a diagram and determining the clusters. More specifically, hierarchical clustering techniques allow the researcher to study data sets that are multi-dimensional.

According to Everitt, Landau, Leese, and Stahl (2011), “In a hierarchical classification the data are not partitioned into a particular number of classes or clusters at a single step. Instead the classification consists of a series of partitions, which may run from a single cluster containing all individuals, to  $n$  clusters each containing a single individual” (p. 71). Due to the step-wise nature of the process, hierarchical techniques can be summarized visually using a

dendrogram. The dendrogram shows how the observations are clustering at each stage and at what value of the proximity measure. There are two types of hierarchical methods: agglomerative and divisive methods. As their names suggest, agglomerative methods take a bottom-up approach where observations are successively fused together, while in divisive methods the observations are separated into finer groups at each successive step. Of the two methods, agglomerative methods are the most widely used (Everitt, Landau, Leese, and Stahl 2011). In fact, as of version 12, Stata, the statistical analysis software that I use, has no command for divisive methods. Their manual notes that with regards to divisive methods, “There are relatively few mentioned in the literature, and they tend to be particularly time consuming to compute” (Stata Corporation 2011b, p. 88).

Based on the above brief discussion of Cluster Analysis techniques, in this thesis I use hierarchical agglomerative methods. There are nine different categories in my data set and therefore nine dimensions. Simple graphical visualization will not be of any use. Also, I chose agglomerative methods because they are more widely used than divisive methods and Stata supports them. The next step after identifying the type of cluster analysis method to is determine both the proximity measure and the linkage method. The choice of these two measures has already been discussed in chapter 2 in light of the theoretical underpinnings of the research. All cluster analysis presented in this thesis is performed using the Pearson correlation measure as the proximity measure and the weighted-average as the linkage method.

The next step after determining the proximity measure and the linkage method is to determine the best fitting total number of clusters. Many measures



have been proposed. Everitt, Landau, Leese, and Stahl (2011) conclude that it is best not to depend on a single measure, and this is in fact the advise followed in this thesis. Milligan and Cooper (1985) examined 30 rules with the conclusion that while there was no single best rule for all situations, two seemed to work most of the time. These are the Calinski and Harabasz pseudo-F index and the Duda-Hart index. I used both rules to determine the optimal number of clusters for the datasets. The first measure was proposed by Caliński and Harabasz (1974) and is calculated using the following formula:

$$\frac{[\text{trace } \mathbf{B} / k - 1]}{[\text{trace } \mathbf{W} / n - k]}$$

where  $n$  is the total number of items,  $k$  is the number of clusters in the solution, and  $B$  and  $W$  are the between and pooled within cluster sum of squares and cross products matrices (Milligan and Cooper 1985). This index is calculated for each possible cluster solution, where the largest index indicates the best clustering of the data. Trace  $W$  will start at a large value, but as  $k$  approaches the optimal number of clusters, trace  $W$  will significantly decrease due to an increase in the compactness of each cluster. Trace  $B$  will behave in the opposite direction. The second measure was proposed by Duda and Hart (1973) and it consists of a hypothesis test in order to decide if a cluster should be subdivided into two sub-clusters. The measure is a ratio where the numerator is sum of squared errors within cluster when the data are partitioned into two clusters, and the denominator is the squared errors when only one cluster is present (Milligan and Cooper 1985). The ratio is then compared to a critical value in order to determine if the hypothesis is to be rejected or not. Once we have determined

the optimal number of clusters we can assign each producer/product to the appropriate cluster based on the results of the cluster analysis.

### **3.4 Random Simulation Experiments**

In addition to determining the best-fit number of clusters, this thesis is also interested in studying how different observations group together. In order to answer questions like “do issues from the same newspaper have a tendency to group together?” and “does this tendency, if it exists, change with time?” we can use random simulation experiments (Simon 1997). These experiments act as a useful tool in order to study whether a certain observed characteristic is due to randomness or whether what we observed is actually a very improbable event. To do that, I will identify a certain statistic and then calculate it for the observed group distributions which we find using the cluster analysis. After that, I reshuffle the observations among the various groups and re-calculate that same statistic. This same experiment is repeated 10,000 times. By the end of the experiment I would calculate the confidence interval of the statistic that I have chosen. I would then see where in this interval does the observed value of the statistic lie. If the observed value lied within the 98% confidence interval, then I would not be able to reject the null hypothesis that the value of the statistic was due to randomness and therefore no special meaning can be derived from it. If, on the other hand, I find the value of the statistic to lie outside the 98% confidence interval then I would be able to reject the null hypothesis and therefore search for a more meaningful explanation as to why the data behaves the way it does. These experiments are simple, yet they provide us insight to the data that might otherwise be lost to us. The results from these experiments will

allow me to build more meaningful models when I start using multi-level logistic regression and survival analysis. These types of experiments are used in chapters 5 and 6 where their results will be explained in detail.

### **3.5 Multi-level Logistic Models**

The act of spanning a category is recorded as a variable that takes on the value of zero when no such activity takes place and a value of one when a newspaper spans a category. Several dependent variables are coded in such a way in chapters 5 and 6. Since these variables are recorded as a series of ones and zeros then it is not possible to use linear regression models with them. The reason for that is that a primary assumption of linear regression models is that the error term be normally distributed. However, if the value of the dependent variable can only take on two values then the error term can also only take on two values and thus cannot be normally distributed (Ryan 2008). Several models have been proposed to handle dependent variables which are binary, the most popular of which is the logistic regression model (Hilbe 2009). The reason for its popularity is that it is the only model that allows for the estimation of odds ratio for the predictors (Hilbe 2009) and it is an extremely flexible function (Hosmer and Lemeshow 2004). The form of the logistic model is:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

In the above equation  $\pi(x)$  is the conditional mean of the dependent variable given  $x$ . A simple transformation of the above equation will yield the following formula:

$$\ln \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x$$

Due to the above transformation, we now have a model that resembles a linear regression model where the left hand side of the equation is referred to as the logit function.

The above model, like linear regression models, makes one important assumption about the data, and that is that each observation is independent from the other. In our data set, we have seen that several issues may belong to the same newspaper. In such a case, the assumption of independence between observations will most likely not be valid. To relax this assumption, I use the Random-intercept logistic regression model. In such a model I include a newspaper-specific random-intercept  $\zeta_j$  in order to relax the assumption of independence of observations that pertain to the same newspaper. The equation transforms to the following (Rabe-Hesketh and Skrondal 2012):

$$\text{logit}\{\Pr(y_{ij} = 1|x_{ij}, \zeta_j)\} = \beta_1 + \beta_2 x_{2j} + \beta_3 x_{3j} + \zeta_j$$

where the subscript  $j$  refers to the newspaper while the subscript  $i$  refers to the individual observation within each panel. Generally speaking, the individual value of the random-intercepts is not of great importance. What is of importance in most research is their population variances and covariances (Singer and Willett 2003). There are several covariance structures that the researcher can choose from. Not all structures are suitable for all conditions. In this thesis I use an identity covariance structure because, as will be seen in chapters 5 and 6, the best fit models will include only a random-intercept, and in such a case the identity covariance structure is the only possible structure (Stata Corporation 2011a). This implies that the variance-covariance matrix of the random effects is a multiple of the identity matrix and all random effects have equal variances and zero covariances.

The above model is used to model spanning categories, an event which is recorded as a dichotomous variable. The next step after that is to model Simpson's index of the issues (the use of Simpson's index was explained in Chapter 2). However, the above model is not suitable for modeling Simpson's index because the nature of this independent variable puts significant restraints on the type of model that we could use. Simpson's index can take on any value from zero to one inclusive. In such a case there are three possibilities. The first is to use the Stata *betafit* command which assumes that the proportions follow a beta distribution (Buis, Cox, and Jenkins 2012). However, this command assumes a distribution that does not include the values zero and one and therefore is not suitable for this case since the Simpson index can, and does, take on these values. A second option is to use a tobit regression. Again this is not suitable in our case because the observed data in our case is not censored from above by one and from below by zero (Baum 2008). The Simpson index cannot possibly take on values outside this range. The third, and only option, available is to use a fractional logit model. This is done in Stata by using the *glm* command along with the *family(binomial) link(logit)* options. Since our dataset is longitudinal in nature we can use the *xtgee* command instead of the *glm* command with the same options to account for the, probable, lack of independence in the observations within panels<sup>4</sup>. The *xtgee* command runs a generalized estimating equation (GEE) model instead of a random-effect model (Hardin and Hilbe 2013). The difference between GEE models and random effects models is that

---

<sup>4</sup> The use of the fractional logit in this case was confirmed to me by Dr. Joseph Hilbe during a discussion we had as I took one of his courses on the website statistics.com, while the generalization to the *xtgee* command was confirmed to me by Dr. Nick Cox's answer to my question on the Stata list. He was also careful to stress that I should relax the standard error assumptions, which is why I use the *robust* option in the models.

instead of focusing on the distribution of the random effects as the source of non-independence, we consider the marginal expectation of the outcome (integrated over the distribution) (Hardin and Hilbe 2013). So the equation becomes:

$$\text{logit}\{\Pr(y_{ij} = 1|x_{ij})\} = \beta_0^{PA} + x\beta^{PA}$$

where the subscripts PA stand for population-averaged. Note that the above equation no longer includes a subject-specific random effect. This means that the coefficients have a different meaning than in the case of random-effects models. In the random-effects models discussed above the coefficients were subject specific in that they referred to the change in the probability for a certain newspaper if the variable changed for that newspaper. In GEE, the coefficients are population-average estimators not subject-specific estimators. This means that the coefficients refer to the change in the average newspaper.

### **3.6 Survival Analysis**

The next step in the analysis is to study the survival times of the publications. In order to do that I will use survival analysis, otherwise known as event history analysis (Hosmer, Lemeshow, and May 2011). The reason that the time cannot be analyzed using regular regression techniques (OLS) is that the error component of the models does not follow a normal distribution (Cleves, Gould, Gutierrez, and Marchenko 2010). The time to an event in many cases is not symmetric, and in some cases can even be constant. Therefore, linear regression is not useful in such cases. Carroll and Delacroix (1982) also show that time series models are not adequate in the case of death events because they aggregate the data, and this would lead to the loss of birth date information.

Typically, the research data will not be complete in the sense that not all units would have experienced the event under study by the end of the observation period. For example, if the units under observation were followed for a period of 20 years to observe their time of death, at the end of this time, there is nothing that guarantees that all the units would be dead. In the case of the present research, our observation period ends in the end of 1974. However, there were many publications that were still a going concern after that date. Such data is referred to as being right censored, which is the most common type of censoring. A second type of censoring, referred to as left censoring, describes the situation where the event is known to have occurred to some unit before the start of observation. This is not an issue in our research because our period of observation starts with the birth of the very first newspaper and obviously no newspaper was subject to the event of death at that time. Finally, interval censoring refers to situations when the event is actually witnessed during the period of observation, but the information available only points to a certain interval of time instead of the exact time (Lee and Wang 2003). The primary function of interest is the hazard function, which is the instantaneous rate of failure. It represents the probability that the event of interest occurs conditional on the subject having survived up to the present time:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t + \Delta t > T | T > t)}{\Delta t}$$

$T$  denotes the nonnegative time to event while  $t$  denotes time.

### **3.6.1 Non-parametric Analysis**

Nonparametric methods place no restriction on the data. This is very useful when the research is new in that no similar data had been previously

collected and analyzed. In such a case it does not make sense to assume that the shape of the data is known. There are two main nonparametric methods used to analyze event history data: the Kaplan-Meier (KM) estimator and the Nelson-Aalen (NA) estimator.

The KM estimator is used to calculate the survivor function. The estimator is calculated using the following formula:

$$\prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$$

where  $n_i$  is the number of individuals at risk at time  $t_i$  and  $d_i$  is the number of failures at time  $t_i$ . We can see from the above a very important feature of the KM estimator, and of all nonparametric methods, which is that data is only processed at the times when events occur. This means that all changes in data that take place between two consecutive events are ignored. If the data contains time-varying covariates, this can cause inefficiencies because some changes in the covariate might go unnoticed. Of course, such inefficiencies are to be expected when using such flexible methods.

A consequence of the above equation is that once the estimator reaches zero, it stays there (any multiple of zero is zero). This can be problematic when at some point in time all observed units die, and after that some new units enter the experiment. To remedy this, another nonparametric tool can be used, but this time to calculate the cumulative hazard function. The tool is the Nelson-Aalen(NA) estimator which is calculated using the following formula:

$$\sum_{t_i \leq t} \frac{d_i}{n_i}$$



where  $n_i$  is the number at risk at time  $t_i$  and  $d_i$  is the number of failures at time  $t_i$ .

### 3.6.2 Semi-Parametric Analysis

The hazard function has two parts. The first part is called the baseline hazard and it represents the hazard faced by all individuals (or whatever the unit of observation may be) in the population. This part characterizes the inherent aging process (Hosmer, Lemeshow, and May 2011). The second part is a function of the covariates which are included in the model. If all the covariates are set to zero, then the hazard function will be equal to the baseline hazard. The most used formulation of the hazard function is

$$h(t) = h_o(t)e^{(\beta_o + x_j\beta_x)}$$

where  $h_o(t)$  is the baseline hazard and  $x_j$  is a matrix of covariates and  $\beta_x$  the corresponding coefficients. This is, in fact, a semi-parametric model that is commonly referred to as the Cox model. While many other formulations of the hazard function would suffice, a special property of the above equation is that the hazards are proportional in that the hazard faced by each unit is multiplicatively proportional to the baseline hazard (Cleves, Gould, Gutierrez, and Marchenko 2010). A considerable advantage of such a method is that we do not need to worry about the shape of the baseline hazard function, given that the two units are being compared at equal times, because it will be cancelled out. However, this can cause problems in settings where the research is concerned with the values of the hazard functions themselves, not just how they compare with each other. This is especially true in medical cases. In social settings however, the research topics are mainly concerned with the effects of the

covariates on the hazard, therefore this formulation is considered ideal. The values of the coefficients (the  $\beta$  in the above equation) are found using the method of maximum likelihood (ML). Therefore, we can see that the model is parametric because it allows us to condition on a set of covariates, while it is "semi-" because, just as in nonparametric models, no restriction is placed on the data to fit a certain predefined hazard function since the baseline hazard function remains unspecified. An assumption made by this model is that the effect of the covariates remains constant, hence the name proportional hazards model. The main advantage of semi-parametric models over nonparametric models is that they allows us to use regression modeling, as we do in linear regression for example. However, to calculate the values of the coefficients, we replace the usual likelihood function by a partial likelihood function because it has been shown that the likelihood function of the above parameterization cannot be used. Fortunately, the asymptotic properties of the partial likelihood function are similar to those of the regular likelihood function (Lee and Wang 2003).

Once we have modeled the above equation, we can calculate the survivor function by first integrating the hazard function to find the cumulative hazard function

$$\begin{aligned} H(t|x) &= \int_0^t h_{0(u)} e^{(x\beta_x)} du \\ &= H_0(t) e^{(x\beta_x)} \end{aligned}$$

Note that in the above equations, we have not included the coefficient of the intercept ( $\beta_0$ ). This is due to the fact that this coefficient gets absorbed into the

baseline hazard and hence becomes part of it. Now that we have the cumulative hazard function, we can calculate the survivor function using the relation

$$S(t) = e^{-H(t)} = e^{-H_0(t)e^{(x\beta x)}} = S_0(t)e^{x\beta x}$$

### **3.7 Research Design**

As can be seen from the above description of statistical models, this thesis will use a variety of tools to test the various hypotheses. Figure 3.2 helps illustrate how the various models are inter-linked. Content Analysis will be used to collect information from the primary data sources, which are the newspaper issues. Some of this collected information will then be used in order to examine the “stories” told by newspapers in chapter 4 regarding the emergence of the industry, while other collected information will be used to determine how the newspaper issues spanned categories. This information will then be used in Cluster Analysis and in the Random Simulation Experiments. In addition, I will collect other data from the “World of Others”, which, in our case, are the bibliographical sources which contained the life history data. This data will be used to perform Survival Analysis and Multi-Level logistic regression. All four statistical models will be used to test the hypothesis mentioned in chapter 2.

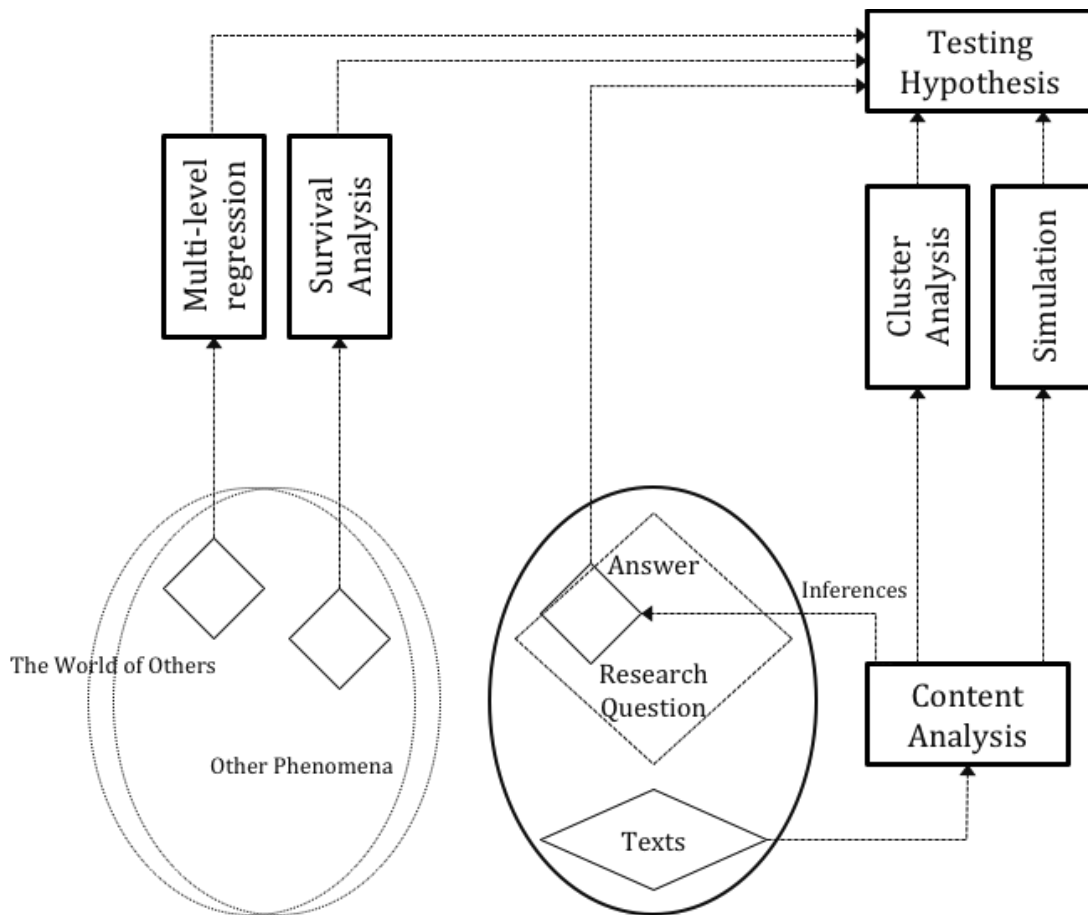


Figure 3.2 Research design. Source: Krippendorff (2004)

### 3.8 Conclusion

This chapter described the data collection procedure and the various statistical tools that will be used to analyse the data and test for the hypotheses which were developed in chapter 2. Most of these methods will be utilized in chapters 5 and 6. The next chapter, chapter 4, is more concerned with qualitative reasoning in order to study the reasons that led to the emergence of the newspaper industry in Lebanon.

## **4 Chapter Four: The Emergence of the Industry as a Social Movement (1851-1879)**

### **4.1 Introduction**

Organizational studies and social movement analysis are among the most creative fields in the social sciences (McAdam and Scott 2005). Within organizational studies, organizational ecology has made great strides in the study of patterns of whole industries. Organizational ecologists have been able to develop several tools with which they studied the dynamics of many different types of industries. While these tools were apt at tackling mature industries, i.e. industries which had successfully emerged and for which the boundaries had stabilized, the fertile area of industry emergence was barely touched (Astley 1985). The result was that the two fields, organizational ecology and social movement analysis, rarely crossed paths. This was because while organizational ecologists were busy studying stable industries, social movement analysts were concerned with the study of how movements challenge and change established systems. In the few cases where the two fields met it was at the hands of social movements theorists who tried to incorporate some of the ideas of organizational ecology into social movement analysis by studying the effect that organizations have on social movements (e.g. Minkoff (1997), Minkoff (1999), Olzak and West (1991), Olzak and Ryo (2007), and Olzak and Uhrig (2001)). Recently, armed with a new array of theoretical tools, organizational ecologists have turned their attention to one of the most important questions in the study

of any industry: why, and how, do certain industries emerge? Since emerging industries are characterized by instability (Navis and Glynn 2010) and incoherence (Santos and Eisenhardt 2009) this meant that a shift in perspective was inevitable for scholars of organizational ecology.

This chapter will look at the reasons that led to the emergence of the newspaper industry in Lebanon by looking at how these newspapers “framed” their goals. These goals surely constitute one element of their identity. A close inspection of the early issues of the founding newspapers in Lebanon will show that there were two dynamics at play. These two dynamics, macro-level structural changes and micro-level grievances, have been usually studied separately in the social movement literature. This chapter will show that it is not possible to study the emergence of the Lebanese newspaper industry without taking into consideration both perspectives. More so, this chapter will also show that neither macro-level structural changes nor micro-level grievances were secondary in importance. The emergence of the newspaper population is a product of both these dynamics, and the result would not have been achieved if one of them were absent or indeed artificial. In addition to the above, unlike previous studies of industries, this chapter will analyze the background of the founding fathers of the newspaper industry. It will be seen that most of these entrepreneurs came from the same background and that this background is a key element in understanding the framing process. Finally, the chapter will argue against the view of al-Rifa'i (1967) that claims that the first newspapers were revolutionary in that they considered the Ottoman Sultan an enemy and that they were proponents of a pan-Arab political governance structure. In fact, the founding newspapers were evolutionary and not revolutionary in that, at least at

first, they considered themselves obedient subjects of the Ottoman Sultanate. Most importantly I will show that the original framing process did not infringe on the political sphere, but instead it set as its target the educational sphere. This is in line with the view of Ayalon (1995) who argued that the first Lebanese newspapers were concerned with “disseminating” knowledge, via their newspapers. Ayalon (1995), however, did not offer a wider perspective in the sense of describing the cultural, political and economic factors that led to the birth of the industry. He just, rightly, stated that the first newspapers were not political but educational. This chapter intends to place the emergence of the industry in its proper context and show how it acted as a social movement.

## **4.2 Literature Review – Social Movements**

Recent studies in organizational theory have attempted to use the concept of social movements in order to explain certain dynamics in markets (Lounsbury, Ventresca, and Hirsch 2003; Weber, Heinze, and DeSoucey 2008). One of these dynamics which social movements can contribute to is the legitimation process in nascent industries (Weber, Heinze, and DeSoucey 2008). However, most of these studies have been applied to markets which were in the post-legitimation stage (Weber, Heinze, and DeSoucey 2008). This is unfortunate because the use of social movement analysis in emerging markets allows the researcher to incorporate agency into the process of legitimation (Benford and Snow 2000). In order to have a clear understanding of how these movements contribute to the emergence of markets or industries, it is imperative that we have a clear definition of the term. A flexible definition was provided by Rao, Morrill, and Zald (2000) who state that “Social movements may be defined as organized collective

endeavors to solve social problems” (p. 244). However, Weber, Heinze, and DeSoucey (2008) contend that “there is in fact limited consensus in the literature on collective behavior about what constitutes a social movement” (p. 531). A more rigorous definition was given by Touraine (1985) who had criticized studies of social movements for being too naïve:

*“There is an almost general agreement that social movements should be conceived as a special type of social conflict...A conflict presupposes a clear definition of opponents or competing actors and of the resources they are fighting for or negotiating to take control of...The type of conflict I will from now on call a “social movement” is defined by a clear interrelation between conflicting actors and the stakes of their conflict. These three components, which I identified long ago as the definition of the identity (i) of the actor, the definition of the opponent (o), and the stakes, that is the cultural totality (t) which defines the field of conflict, belong to the same universe; they express the central conflict of a societal type”(p. 750).*

Touraine (1985) identified eight different types of conflict, and he chose to refer to only one of them as "social movements":

*“...exists a different type of social conflict, whose stake is the social control of the main cultural patterns, that is, of the patterns through which our relationships with the environment are normatively organized” (p. 754-755).*

He distinguished this conflict from political and institutional conflicts. This thesis adopts the above definition because, as will be seen in this chapter, the conflict in which the newspapers were involved falls neatly into the above conflict type. The founders of the newspapers argued for a change in cultural patterns and in



the relationship with the environment. They believed that the decline of their homeland was a result of an unhealthy culture in which religion separated people. Therefore, as will be seen, they argued that science can help transform society and they concentrated on the Arabic language as a unifying force.

Extant literature in social movements has diverged along two clear paths: the resource-mobilization perspective and the psychofunctional perspective (Snow, Rochford Jr, Worden, and Benford 1986). Scholars adopting the first perspective concentrate on strategic issues which are found in organizational studies such as organizational structures and processes (McAdam and Scott 2005). Unlike the psychofunctional perspective, resource-mobilization rejects the notion that emotions are the driving factor behind these movements (Cohen 1985). It is therefore no surprise that this school of thought assumes high levels of rationality in the agents. While some resource-mobilization theorists do not deny the presence of grievances, they argue that such grievances are a natural product of all power plays and hence are present in all environments (Cohen 1985). Therefore, the real driving force behind the formation of social movements is to be found in the changing opportunities (Jenkins 1983; McCarthy and Zald 1977). Others, most notably Zald and McCarthy (1987), have taken an even more extreme view arguing that grievances are in some cases manufactured by the entrepreneurs in order to increase the likelihood of the success of the social movement. In general, all resource-mobilization theorists hold that the focus on micro-level factors such as the psychological state of those involved has pushed back important macro-level processes (McCarthy and Zald 1977).

Proponents of the psychofunctional perspective on the other hand stress the importance of grievances which are the result of “preexisting social arrangements” (McAdam and Scott 2005), while criticizing the assumption of excess rationalism on which the resource-mobilization perspective was built (Snow, Rochford Jr, Worden, and Benford 1986). In addition, just as they were criticized for ignoring macro-level processes, researchers who adopted the psychofunctional perspective criticized the proponents of resource-mobilization for ignoring micro-level processes (Jenkins 1983). The grievances, as argued by these researchers, were what provided the movements with the necessary social capital (McAdam and Scott 2005). However, it was necessary that these grievances be shared and understood by most agents alike. The heightening of these shared grievances would lead to the formation of social movements. Snow and Soule (2010) distinguished between individual-level grievances and mobilizing grievances and argued that while individual-level grievances were ubiquitous, mobilizing grievances were not. Resource-mobilization scholars, by failing to make this distinction, have wrongly assumed that all grievances were ubiquitous.

Although the above differences are substantial, there is a consensus among social movement scholars with regards to the importance of what Goffman (1986) referred to as “framing”. Goffman (1986) posited that frames helped people make meaning of the world. In that way, they are similar to schemata except that frames, unlike schemata which were predefined perceptions, are outcomes of negotiated meanings (Benford and Snow 2000). Social movement scholars, most notably Benford and Snow, used frames to

develop what they termed collective action frames. In the words of Benford and Snow (2000) framing,

*“denotes an active, processual phenomenon that implies agency and contention at the level of reality construction. It is active in the sense that something is being done, and processual in the sense of a dynamic evolving process. It entails agency in the sense that what is evolving is the work of social movement organizations or movement activists. And it is contentious in the sense that it involves the generation of interpretive frames that not only differ from existing ones but that may also challenge them. The resultant products of this framing capacity are referred to as ‘collective action frames’”(p. 614).*

These collective action frames are crucial in the formation of new markets, or industries, in the case of entrepreneurs who operate within social movements (Rao, Morrill, and Zald 2000). The founding fathers of any industry are faced with unique challenges that latter generations do not face (Aldrich and Ruef 2006; Aldrich and Fiol 1994; Alvarez and Barney 2005). In order to successfully create new industries, the entrepreneurs will have to bridge the social distance between producers and consumers (Lounsbury, Ventresca, and Hirsch 2003; Weber, Heinze, and DeSoucey 2008). This is where the importance of grievances comes in. For the scholars of resource-mobilization, the social movement will have to enhance, or even create, in such a way that they resonate with the audience. For scholars of the psychofunctional perspective, collective action frames must be preceded by an increase in the intensity of the grievances. Either way, the social movements will have to diagnose the problem, propose a solution, and provide the motivation for action (Evans 1997; Snow and Benford

1988; Weber, Heinze, and DeSoucey 2008). This implies that 1) agency has a central role in the formation of new industries (Benford and Snow 2000), 2) collective action frames are dynamic in their nature (Benford and Snow 2000; Snow and Soule 2010), 3) meaning creation and interpretation play an important role in the process (Benford 1997; Benford 2005; Benford and Hunt 1992; Snow, Rochford Jr, Worden, and Benford 1986). The ultimate question, therefore, “is not what’s going on here, but under what conditions do people believe in a particular version of reality?” (Benford 1997, p. 412).

In order for social movements to legitimate their activities, they have to strategically align their collective action frames with those of the audience (Benford and Snow 2000; Lounsbury, Ventresca, and Hirsch 2003; Snow, Rochford Jr, Worden, and Benford 1986). Frame alignment is concerned with linking the social movement’s orientations to those of the individual audience members (Snow, Rochford Jr, Worden, and Benford 1986). The presence of grievances by itself is not a sufficient condition for successful mobilization. What truly matters is that the social movements identify and give meaning to these grievances in a way that resonates with the intended audience (Benford 2005; Snow, Rochford Jr, Worden, and Benford 1986). Since collective action frames “encode” experiences (Snow and Benford 1992), social movements must make sure that this encoding is both intelligible and not counter-intuitive to audience members. Part of the alignment process is the creation of a vocabulary and a set of “stories” that help shape the audience’s perceptions (Wry, Lounsbury, and Glynn 2011). In doing so, social movements can help situate their goals and intentions within the wider context of the cultural and political environment. These linguistic tactics however also play a central role in the creation of a

unified collective identity, which is a necessary condition for successful emergence (Swaminathan and Wade 2001; Weber, Heinze, and DeSoucey 2008). According to Benford and Snow (2000), “an understanding of identity processes, and particularly collective identity, is fundamental to understanding the dynamics of social movements” (p. 631). This thesis holds that collective identity is socially constructed (Cerulo 1997; Cohen 1985) through the promulgation of a new vocabulary and the spread of stories which help establish a new social and cultural order (Eisenstadt 1998).

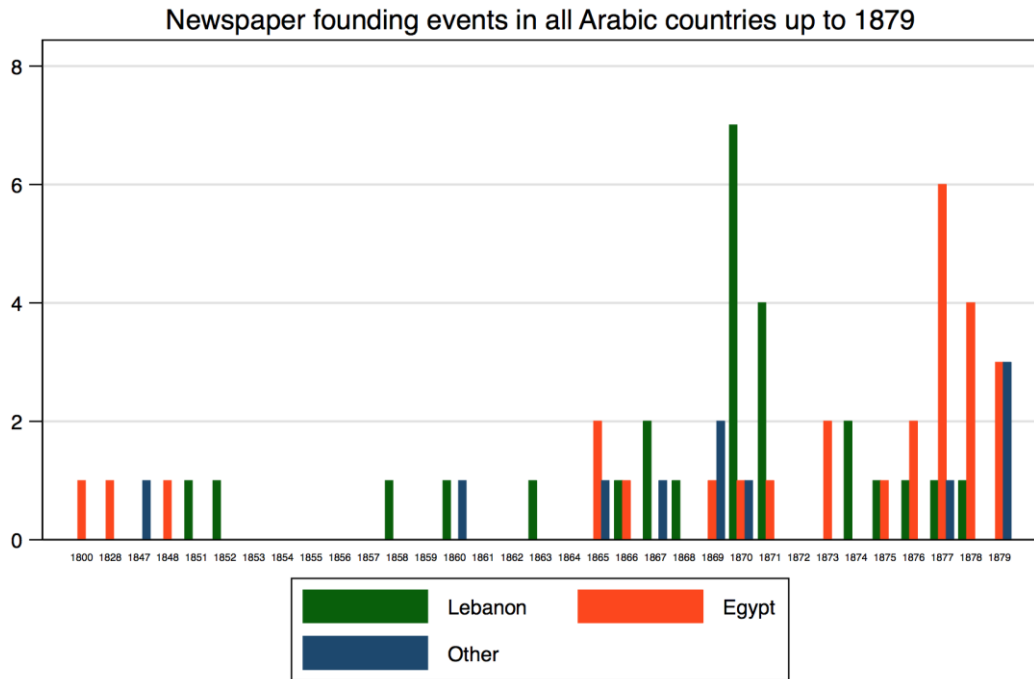
### **4.3 Empirical Setting: The Emergence of Lebanese Newspapers**

Lebanese industry historians consider the Lebanese newspaper *Hadiqat al-Akhbar*<sup>5</sup> to be the first Arabic newspaper (Illias 1997; Tarazi 1933). In order for them to make that claim they redefine what they mean by “first newspaper” as such: the first privately owned Arabic newspaper published in an Arabic country. This is because the first newspaper to be printed in the Arabic language was published on the orders of Napoleon after he conquered Egypt in 1799. The newspaper, *al-Hawadeth al-Yawmiya*, had a single job, and that was to facilitate communication between the conqueror and the conquered. It was discontinued in 1801 when the French army left Egypt. Nineteen years later, Mohammed Ali established the first printing press in Egypt. On the 20<sup>th</sup> of November 1828 the government owned *al-Waqa’i’ al-Masriya* was born. On 15<sup>th</sup> of September 1847 the French once again printed an Arabic newspaper in Algeria, which was named *al-Mubashir*. Again, the newspaper was used as a tool to facilitate

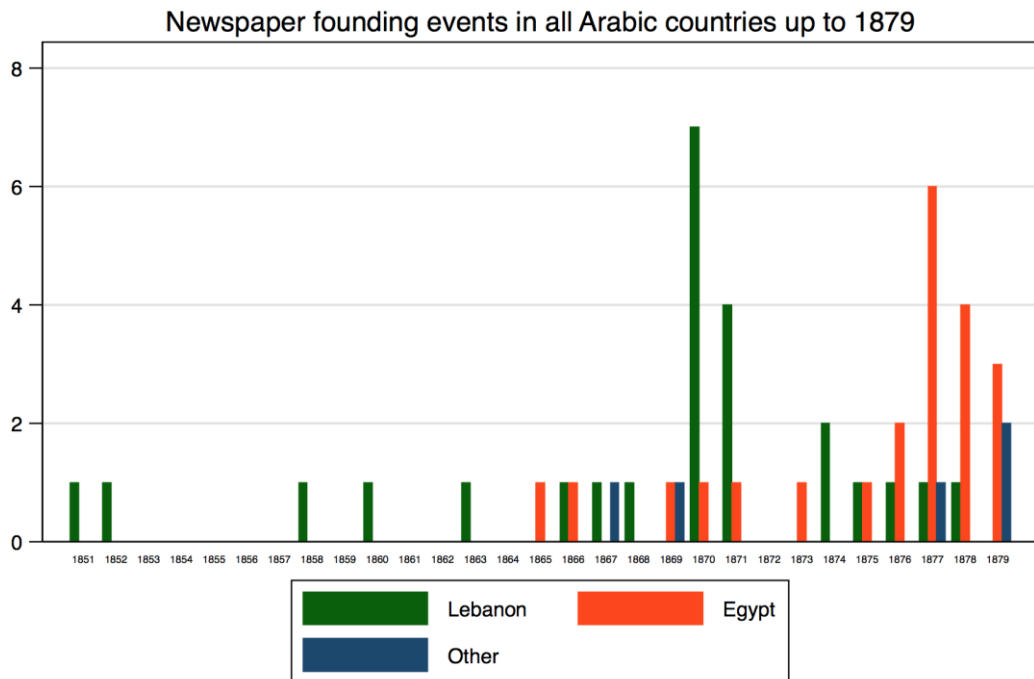
---

<sup>5</sup> With regards to the romanization of the Arabic names, I used the character ‘ to represent the Arabic character ء, and the character ’ to represent the Arabic letter ع.

communications between the French and their colonized subjects. Then in 1851 the American Protestant Missionaries in Beirut published *Majmoo' al-Fawa'id* which became the first Arabic language magazine. The magazine was concerned with theological, scientific, historical and geographic studies. In the following year, a group of Arabic and Western scholars in Beirut collaborated to produce a newspaper, *A'mal al-Jam'iyah al-Suriyah*, which was intended to spread scientific knowledge in the Arabic world. All of the above newspapers were published either by governments or by a group of individuals as an outlet for scientific news. In 1855 an Arabic Syrian by the name of Rizq Allah Hasoun published *Mir'at al-Ahwal*, the first privately owned Arabic newspaper, in Turkey. Some years later the owner of the newspaper escaped to Russia to flee the Ottoman government which he had managed to upset. Finally, the "first privately owned Arabic newspaper published in an Arabic country" was printed when on the first of January 1858 Khalil Khoury published the first issue of *Hadiqat al-Akhbar*. This newspaper was followed by a series of newspapers which were published by Lebanese, mostly Christians. Up until the end of 1879, the Lebanese had published 26 newspapers in Lebanon, out of which 25 were published in Beirut. The one newspaper not published in Beirut was a government owned newspaper and it was published instead in Beit al-Deen where the governor resided. Out of the first 26 newspapers, one was published by the government, two by groups of scholars, six by Christian missionary groups, sixteen by Christian individuals and one by a Muslim. Figure 4.1 below shows the founding events in Lebanon, Egypt, and all the remaining Arabic countries up to 1879.



**Figure 4.1 Newspaper founding events in all Arabic countries up to 1879**



**Figure 4.2 Newspaper founding events in all Arabic countries up to 1879 excluding official newspapers**

Although hailing from one of the smallest countries, the Lebanese were the leaders when it came to the emergence of the newspaper industry in the Arab world. Egypt is clearly a close second. However, Lebanon is the only Arab country, out of those that had published newspaper prior to 1880, where the government did not start the first newspaper. In all other countries, the industry started at the behest of the government. If we disregard the newspapers started by the government we end up with Figure 4.2.

Financially speaking, the conditions at the time of the emergence of the industry were not favorable. Illiteracy was very high in Lebanon, even in the culturally refined area of Beirut. Therefore, circulation numbers were very low. In 1858 for example, *Hadiqat al-Akhbar* had only 400 subscribers (Ayalon 1995). Many of the first newspapers had severe financial difficulties, and they constantly complained about them to the public. Almost all of the inspected newspapers (see Table 4.1 in the section about the framing process) continuously reminded their subscribers to pay for their subscription as soon as possible. Some newspapers tried to enlarge their newspaper from the usual 4 pages per issue, or to increase the frequency, but later backed down on their decision and stated that the reason was that they were not receiving enough funding or that the readers were not ready to pay a higher fee. Some newspapers also implored their readers not to share issues but to pay separately for each issue. The fact that the newspaper founders insisted on publishing their papers even at great personal cost indicates that money was not the driving force behind their actions. The situation in Lebanon would gradually turn to the better during the latter years of the second half of the nineteenth century as will be shown in the section dealing with macro-level structural changes.



### 4.3.1 Micro-Level Grievances

There was a clear change in the political and cultural mood in Lebanon in the nineteenth century. The political changes in the Ottoman Empire, especially the reformations, the cultural interaction between the Christians of Lebanon and the missionary groups, and the continual increase in the dependence of the Ottoman economy on Europe all had lasting effects on the people.

In 1516, the Ottomans conquered Greater Syria, the area that is today known as Syria, Lebanon, Transjordan, and Palestine. Initially, the Ottomans tried to control Mount Lebanon, but the Druze that inhabited the area proved to be too much trouble. After a series of battles that started in 1523, the Ottomans finally managed to subdue the Druze in 1586. Once the Druze accepted that the Ottomans were now their masters, the Ottomans chose one of them as emir. The Druze chief Fakhr al-Din Maan was allocated the *sanjaks* (administrative divisions) of Beirut and Sidon in 1590. The Emir was free to rule the land and tax the people, as long as he paid the Ottoman authorities the required respect, and taxes. At that time the land was solely owned by the government, and the peasant population was allowed to work on it in exchange for the payment of taxes (Owen 1993). Previously, the main feature of Middle Eastern commerce was the transit of Asian peppers, spices and silk. Most of these products were destined to Venice (Owen 1993). Now, the Ottomans allowed other European companies, most notably English, French and Dutch, access to the area. The result was an increase in the demand for silk which had previously been produced on a small scale for local use only (Salibi 2003). So in order to increase production of silk, the Druze landlords needed access to a larger labor force, and that is exactly what they found in the Maronites of the north. At around the same

time, the district of Bsharri, which is located in the north of Lebanon, witnessed a deterioration that increased the rate of immigration of the Maronite inhabitants to the south. The first of these migrations took place in sixteenth century, just as the demand for silk started increasing.

Eventually Fakhr al-Din rebelled against his masters and in 1633 he was captured and taken as a prisoner to Istanbul where he was executed (Salibi 2003). The rule of the mountain passed to the Shihabs, who were Sunnites from the Wadi al-Taym and not from Mount Lebanon. The Shihabs, who understood all too well that the Druze only followed their Druze overlords, sought the alliance of the Maronites in order to gain a foot holding in their new territory (Salibi 2003). Eventually the Shihabs started converting to Christianity. They were led by Emir Bashir Shihab II.

Throughout all this time, the Christians had been increasing in number at a far greater rate than their Druze counterparts. The relationship between Bashir Shihab and the Ottomans would soon turn sour, as Muhammad Ali Pasha would rebel against the Empire. He sent his son, Ibrahim Pasha, to invade Greater Syria in 1832. Bashir Shihab chose to align himself with the Egyptians. Unlike the Ottomans, the Egyptians decided to rule all of Greater Syria directly, and they proved to be much more efficient at it (Makdisi 2000). The area of the cultivated land in the Bekaa was increased, silk export more than tripled during the years 1833-1836 (Buheiry and Conrad 1989), and the cotton area in Syria more than doubled (Owen 1993). The Egyptians managed to extract much greater taxes from the area than the Ottomans ever did. In addition, they also set up strict military conscription laws and they paid special attention to making sure that these laws were enforced. The Druze chiefs were the first to rebel against these

two measures in 1838. The revolt soon spread to many areas, and in order to counter it, Ibrahim Pasha distributed arms to the Christians who stuck by the side of Bashir Shihab (Tarābulṣī 2007). This was to be the first time in the history of Mount Lebanon that the policies of the rulers were based on the sects of the people (Khalaf 1982). According to Makdisi (2000), 1841 was the birth of sectarianism in Lebanon. Eventually though, the Egyptians were far more efficient in tax collection and conscription than anyone could handle, and most inhabitants of the mountain joined with the Druze in their rebellion. At the same time, the British, in their never-ending competition with the French, backed the Ottomans both politically and militarily. In October 1840, the Egyptians were forced to retreat to Acre, Bashir Shihab was arrested and was exiled to Malta, and Greater Syria was once again part of the Ottoman Empire (Tarābulṣī 2007).

The Ottomans returned to a different Mount Lebanon than the one they had previously ruled. The old social hierarchy was starting to disintegrate and in its place a new system, which was based on sectarian divisions, was starting to form (Khalaf 1982; Makdisi 2000). As the Druze chiefs returned from their exile back to their villages and tried to reclaim the lands that they had previously controlled, they were faced with a hostile Christian peasantry. The peasants were backed by the new emir of Mount Lebanon, and the Druze were quick to retaliate. The fighting soon spread to other regions and hundreds of Druze and Christians were killed. To make matters even more complicated, the Maronites sought the protection of the French, who had started tidying up the mess that they created when they backed the Egyptian invasion, and the Druze looked upon the British for help (Makdisi 2000). Both sects understood that the Europeans now yielded far more power than they ever had and that it might be

possible to use this power to their advantage. The Ottomans stepped in and declared the end of the old system in which the Ottomans delegated their powers to a local emir. This time though, the Europeans joined them in the decision making process. The Europeans argued that the mountain should be divided along sectarian lines (Makdisi 2000), but the Ottomans objected because, as they argued, there were areas in which the population was too mixed. On December 7, the Ottomans conceded, but they clearly stated that they believed this to be a bad idea (Makdisi 2000).

Mount Lebanon, for the first time ever, was split along sectarian lines into two *qaimmaqamiyya*. The first was located in the north of Mount Lebanon, where the peasants and the chiefs were Maronite. To the south another *qaimmaqamiyya* was established. Here, however, things were more complicated because there were both Maronite and Druze peasants, with the former far exceeding the latter, while the chiefs were entirely Druze. These reforms did not last long. The Maronite peasants in the north refused to be slaves to the Khazin and Hubaysh families, the two Maronite *muqataji* families, while the Druze in the south, led by Said Jumblat, refused to recognize the authority of the newly appointed *qaimmaqam* (Tarābulṣī 2007). To make matters worse, the Ottomans recruited Maronites to help them fight the Druze rebellion in the south, thus further increasing sectarian tensions.

The missionary groups also contributed to the rising sectarian tensions. Makdisi (2000) described the arrival of missionary groups as a gentle crusade. Previously, contact between the Lebanese and Christian groups such as the Jesuits was confined to the Roman Maronite College in Rome and to some advisors for the Maronite Patriarch in Lebanon. In the nineteenth century this

field of contact was significantly expanded to include Mount Lebanon and coastal cities like Beirut. The arrival of the missionary groups, at a time when Bashir Shiab II was becoming more and more aware of his Christianity and the Egyptians and European powers were redefining the landscape along sectarian lines, would speed the process of 'sectarianization' in the mountain (Makdisi 2000). The Jesuits played an important role in alienating the Protestants because they believed that all areas inhabited by Catholics should be under their religious jurisdiction. The Protestants, naturally, sought other places, and as it happened, the British were on excellent terms with the Druze. In the end they found their sanctuary in the areas that were controlled by the Druze. Interestingly, the missionary groups were divided along the same lines which the mountain would be politically segregated into in the near future. When they first arrived, the missionaries, especially the Jesuits, were shocked to find that the Christian population was "corrupted by their neighborliness with Muslim and Druze villages" (Makdisi 2000, p. 91). A few years after their arrival, the situation had changed.

The years of 1856-1858 were economically one of the hardest on the peasants. By that time France had become the main recipient of the silk produced in Mount Lebanon. The dependence of the mountain on France was so large that when, during that period, Lyon suffered an economic recession, the production of silk in Mount Lebanon was halved (Tarābulsī 2007). To make matters worse, there was an economic downturn which began in 1856 (Buheiry and Conrad 1989), and which was followed by an especially severe winter in 1857, and a dry season in 1858. All these factors led to the commoners rebelling. This rebellion was soon to spread like a wild fire and a commoner by the name of

Tanyus Shahin emerged as the leader of the peasants in December 1858. Shahin complicated matters further yet when he mixed the social with the sectarian. Not only did he claim that the rebellion was directed against the unjust *muqatajis*, but he also insisted that the rebellion should spread to the south so that they could free their Maronite brothers from the Druze *muqatajis*. Once again, a social problem was to be given a sectarian spin, and the Christianity of the rebellion became its most defining characteristic (Makdisi 2000).

The Druze of the south were quick to act. The crumbling of the *muqataji* system was seen as a direct attack on them. In addition to that, their lands were populated by a large number of Maronites who were clearly sympathetic to the cause of their brothers from the north. The fighting spread from Mount Lebanon to the Bekaa and the Anti-Lebanon until it finally reached Damascus. By the end of the war 5,000 people were killed in Mount Lebanon alone with 200 villages being burnt (Tarābulṣī 2007).

The European powers were shocked by the events of 1860, while the Ottomans viewed the events as proof that the system had failed and that what was needed was direct rule by the Ottomans themselves. The European powers, most notably France and Britain had a strong say in things. The main reason was the ever increasing debt which the Ottoman Empire owed European banks (Owen 1993). The old system which was based on the *qaimmaqamiyya* was replaced by combining both into a single *mutasarrifiya* (Tarābulṣī 2007). This system proved to be successful and Mount Lebanon would not be the scene of any civil strife from then until the fall of the Ottoman Empire. However, Mount Lebanon was slowly to lose its prominence. The civil war caused many people,

mostly Christians, to immigrate. While Mount Lebanon was losing its importance, Beirut, on the other hand, was beginning to rise.

At the beginning of the nineteenth century, Beirut had a population of 6,000 and there was little indication that it would become the most important city in Lebanon (Buheiry and Conrad 1989). By the end of the century it had a population of 120,000 (Fawaz 1983). During the same period, it also grew from a small seaport that was overshadowed by Alexandria and Sidon to become the most important seaport in the Mediterranean. This led many European consulates to establish themselves in the city in the 1830s (Fawaz 1983). Unlike Mount Lebanon, the original inhabitants of Beirut were Sunnite Muslims and Greek Orthodox. These demographics would change with the civil wars of Mount Lebanon. Due to their defeat, many Maronites left the mountain and went to Beirut, thus ending the previous equality between Muslims and Christians in the city (Fawaz 1983). Many factors led to the rise of the city. One of the most important factors was the expansion of trade with Europe and the industrial revolution in Britain, which increased production and exports. Beirut's close location to the large cities of the Syrian interior also gave it an upper hand on its competitors, a position which was greatly enhanced with the building of the Beirut-Damascus road in 1859-1863 (Fawaz 1983), and the establishment of a public works department in 1867 which built small roads connecting the towns of the mountain to Beirut (Owen 1993). It was in the 1850s that new European financial institutions were created to channel savings into investments abroad (Owen 1993). One of the most important changes that came along with this financial system was the fact that the peasants were no longer dependent on the

wealthy landowning families like in the past. Instead, the peasants would borrow money themselves in order to finance their growing businesses.

Unfortunately, the increase in the material wealth of the city was not equally distributed among its inhabitants. The number of Europeans who spoke Arabic or Turkish was very small, and so they needed the assistance of local people who spoke those languages and who also understood Middle Eastern commercial practices (Owen 1993). The new middle class that connected European capital and Mount Lebanon was largely composed of Christians, with very few Sunnites (Issawi 1982). As early as the 1840s, only three of the twenty-nine merchant houses that traded directly with the Europeans were Muslim (Fawaz 1983). In addition to the financial gains, the Christians were the recipients of important legal benefits thanks to capitulations extended to them by the Europeans.

The above discussion clearly shows a significant change in the history of Lebanon. This change had considerable impact on the lives of the people of Mount Lebanon and Beirut. Prior to all these changes, there was a strict social system in which the poor lived to serve their feudalist lords. The people understood well where their loyalties should be and no one dared question the status quo. The emir was the political leader and he was subordinate only to the Ottomans. The Ottomans did not consider Mount Lebanon and Beirut to be important. Their attention was directed towards the Syrian interior, mostly to cities like Damascus and Aleppo. Also, while the people were aware of their religious identities, these identities did not conflict with the social hierarchy to which they belonged. Christians toiled night and day for the well being of their Druze lords. By mid-nineteenth century all of this had changed. The Druze and



the Maronites of Mount Lebanon became two distinct communities. The Sunnis and the Greek Orthodox of Beirut no longer lived together peacefully in Beirut due to sectarian tensions that spilled over into the city from the Mountain. To complicate matters further, the power of the Ottomans had declined tremendously, a decline that was equaled by the rise of the power of Europe. The people no longer had one master. The Christians sought the protection of the French while the Druze sought the protection of the British. The change in the economy also assisted in the destruction of the old order. Commoners had access to capital through channels other than their feudal lords, and the growing of mulberry trees and silkworms was within their reach. The Lebanese economy became interlinked with that of Europe more than any other place in the Ottoman Empire. All this meant that the people of Lebanon enjoyed an interaction with the Europeans, especially missionary groups, rivaled by no other people. The presence of a huge number of Christians within their ranks speeded up the process. The more they interacted with the Europeans, the more they realized how uneducated they were. They came to realize that there was a different world out there far more advanced than the Ottoman world. Most importantly, they were introduced to secularism and the idea that all subjects of the empire could be equal. Below when we discuss the framing process it will become clear how all these grievances manifested themselves in the early newspapers. Unlike what the resource-mobilization discourse suggests, these grievances were not blown out of proportion nor were they invented. These grievances were real and new, and they had a strong impact on the emergence of the newspaper industry.

### **4.3.2 Macro-Level Structural Changes**

The first printing press in the Arab world was established in 1610 Der Qazhiya in Tripoli, which is located in the north of Lebanon (Sabat 1958). This printing press did not last long though because, according to Sabat (1958), the environment was not ready for it. It would be more than one hundred years before another printing press was established in Lebanon. The first Arabic printing press in Lebanon was established in 1733 by the Deacon Abdullah Zakher. The printing press did not produce many books in its first one hundred years, again a fact which reflects the lack of education in Lebanon at that time (Sabat 1958).

The Muslims, on the other hand, at first refused to indulge in such technology because they refused to use European technology without the consent of their political and religious leaders (al-Rifa'i 1967). In 1716 the religious leader of the Ottoman Empire gave his consent to use the printing press as long as it did not print religious books (Sabat 1958). However, the people waited until an official decree explicitly allowed the setting up of a printing press. Therefore, they had to wait until 1727 when on the 2<sup>nd</sup> of July an imperial decree stipulated that books in the Ottoman Empire were to be printed. The first set of rules that explicitly dealt with publication issues were issued on the 6<sup>th</sup> of January 1857 in the era of Sultan 'abd al-Majid (al-Rifa'i 1967). We can see that there was a gap of more than 100 years between the adoption of the printing press by Christians and the adoption by Muslims.

Sabat (1958) identifies the years 1834-1869 as the years where the printing press took hold in Lebanon, especially in Beirut. The first printing press to be owned by an individual was the Syrian Printing Press which was founded

by Khalil Khoury the founder of *Hadiqat al-Akhbar* the first newspaper. In 1858 Ibrahim al-Najjar founded the Eastern Printing Press. In 1861 Youssef al-Shalfoun founded the General Printing Press. In 1865 two other printing presses were founded by two groups of Christian priests. Also in 1865 Jirjis Shahin founded the National Printing Press. Then in 1868 Khalil Sarkis founded the Knowledge Printing Press. There were several printing presses that were established outside Beirut, but unlike their counterparts in the city they were not very productive and in general did not survive long.

The missionary groups that landed on the shores of Lebanon had an immense impact on the cultural scene in Lebanon. The most important was the educational sphere. The first Christian group to come to Lebanon was the Jesuit fathers who had originally arrived in 1625. They faced great difficulties but nonetheless continued their mission as best as possible. They established a few schools and distributed books to people. However, their activities were severely restricted geographically due to sectarian sensitivities. In 1773 their organization was disbanded and they had to leave Lebanon. In 1831 a group of American Protestants came to Lebanon in an effort to convert Catholics to their sect. This group was part of a larger organization that had first chosen Malta as its base of operations. In 1832 the Egyptians invaded Lebanon and established a rule that was more tolerant and advanced than that of the Ottomans. The Egyptians founded many new schools in Aleppo and Damascus and provided the students with free accommodation. The Christian missionary groups benefited from this new atmosphere and began expanding their operations. In 1834 the American missionaries relocated their printing press from Malta to Beirut so that they could easily supply their schools with books. The Jesuits established their

first printing press in 1847. The Catholic missionaries and the Protestant missionaries soon started competing with each other and this led to the creation of many teaching institutions. In 1834 the Jesuit fathers reopened the 'ayntoura School that they had closed when they left in 1773.

The middle of the nineteenth century saw the birth of many small-scale schools. This is clearly reflected in a series of reports submitted to the Colonial and Continental Committee of the Free Church (Lowthian, Saleebey, and Saleebey 1856). According to the reports, a school was established in 1849 in the small town of Howarah by a Lebanese priest who was a scholar at the American Missionary School in Ibay. The school was located in his house. On the 26<sup>th</sup> of July 1853 a school was built in Btater. On the 3<sup>rd</sup> of November a school was built in Aramoon with an attendance of 62. On the 10<sup>th</sup> of November another school was opened in Bataloon. At the end of 1852 a school was opened in Allay and had an attendance of 62. In June 1855 a school was opened in Sook. The report also indicates that the attendance numbers were increasing each year. By 1862 the report indicates that there were a total of 18 schools in 15 different villages. A total of 700 students were recorded. The 1863 report states that there were 1500 pupils in attendance at all schools. In 1864 the report talks about 21 schools in 18 villages. The American missionaries on the other hand reported that by 1860 they had opened 33 schools which had a total of around 1,000 students (al-Rifa'i 1967). By 1862, the number of schools had increased to 41 (Tarābulṣī 2007). The Jesuits on the other hand had opened schools in Beirut in 1839, in Zahle in 1844, and in Damascus and Aleppo in 1873.

Soon, the schools that were being built were taking on a much larger scale. In 1863, the Lebanese scholar Butrus al-Bustani founded the National

School. In 1865 the Roman Catholic Patriarch Gregory the first also founded *al-Madrasah al-Batriyarkiya*. In 1866 the American Protestant missionaries relocated a school that they had previously started in Ibay to Beirut and called it the Syrian Protestant College. This was to become later the American University of Beirut. The Jesuit fathers followed their suite when in 1875 they moved a school which they had originally opened in Ghazir to Beirut and renamed it Saint Joseph College. That same year also saw the founding of *al-Hikmah* by Joseph al-Dibs, and in the following year the Muslim scholar al-Shaykh Ahmed Abas al-Azhari founded his own school. By 1869 there were 75 schools in Beirut with almost 6 percent of the population enrolled in them (Issawi 1988), and by 1878 the number had increased to 92 (al-Mishkat 1878). Official Ottoman registers indicate that by 1894 the percentage of illiterate persons above the age of ten in the administrative district of Beirut had dropped to 34.72%, giving Beirut a rank of 23/36 in terms of the highest illiterate percentage in all Ottoman districts (Karpas 1985). Other Arabic countries were not doing so good. Egypt, for example, had an illiteracy rate of 93 percent by as late as 1907 (Issawi 2006). Around the same time that literacy was increasing in Lebanon, several newspapers were talking about an increase in the demand for their product and mentioned plans to enlarge their newspapers. The 1/11/1870 issue of *al-Jinan* talks about a great demand for the newspaper. In the 15/4/1877 issue of the same newspaper the author says that this was the age of great demand for newspapers. In the 15/3/1878 issue the author mentions an increase in competition and promises to keep improving his paper. In May 1872 the publisher of *al-Najah* increased its frequency to three times per week and states that this was the result of an increase in both demand and praise for his

paper. According to him, people were more concerned with receiving news on time than ever before. Another signal that more and more people were reading newspapers was that the governor of Beirut placed an announcement in the 12/8/1878 issue of *Lisan al-Hal* forbidding shops to place their items on the street. Obviously he expected the notice to reach many of the shop owners this way. In its first issue dated 1/6/1876, *al-Moktataf* stated that one of the reasons for starting the paper was that many people had asked for a new scientific newspaper. Almost two years later the publisher declares that he will be adding four pages to the medical section because of their popularity.

### **4.3.3 The Founding Fathers<sup>6</sup>**

An examination of the histories of the founding fathers of the newspapers reveals that most of them came from a common background. They were men of science that had close contact with the government but were formally excluded from the highest positions. While the Muslims held government posts, their Christian counterparts, who could not hold such positions in an Islamic state, worked as interpreters for local governments and consulates (Hourani 1983).

The most famous of the founding fathers of the newspaper industry is also the most prominent intellectual figure in the history of Lebanon. Subsequent generations refer to him as “the teacher”. Butrus al-Bustani studied in the school of ‘ayn Waraqa. He worked as a translator for the American consul in Beirut and also worked with the American missionaries in their printing press and helped translate the Torah to Arabic. He authored many books that dealt with Arabic grammar, in addition to books that were concerned with history and science. In

---

<sup>6</sup> Most of the biographies of the founding fathers were taken from Tarazi, Felipe. 1933. *Tareekh al-Sahafa al-'arabiya*. Beirut: Dar Sader.

1863 he founded the National School, which was to become one of the most prominent schools in the history of the country. Butrus al-Bustani founded two newspapers: *Nafir Suriya* and *al-Jinan*. His son, Saleem Bustrus al-Bustani, founded two newspapers, *al-Janna* and *al-Junayna*, and worked with his father on *al-Jinan*. He studied Turkish, French and the English languages and, just like his father, worked as a translator for the United States consul. He taught English in his father's school while also helping him in the management of the school.

Khalil Khoury studied in the Roman Orthodox School where he learned the Arabic language and subsequently wrote poems which were praised by the famous poet Naseef al-Yazaji. He then learned the Turkish and French languages at the hands of specialized tutors. He published his poems in 6 books. He also wrote novels and wrote a book about the history of Egypt. He also translated books and articles from the Turkish and French languages to the Arabic language. In 1865 he was appointed as the manager of the government printing press in Syria. In 1870 he was appointed as inspector of the non-Islamic books.

Youssef al-Shalfoun studied the Arabic language and some other languages in the schools of Lebanon. He then worked in the printing press of Khalil Khoury. He was called upon by Fuad Pasha to edit the government letters to the foreign consuls in the events of 1860. He then founded a printing press that was responsible for printing more than sixty books that dealt with philosophy, religion, poetry and history among other topics. In 1867 he was called upon by Daoud Pasha to run the government printing press in Beit al-Din.

Louis al-Sabounji entered al-Shurfa School in Kisrawan. In 1854 he was sent to Rome where he earned his doctorate in philosophy. He later became the head of the Syriac Church in Beirut. He founded a printing press that printed

books in Arabic, Turkish and the Syriac language. He also founded a school that was to become one of the most famous in Lebanon. He also taught Latin, Turkish and Italian in other schools. While in England he invented a photographic machine that he earned a patent for from the government. He translated books from Italian and Latin to Arabic. He also authored books on history, philosophy and an English-Arabic dictionary.

Abd al-Qader Qabani first studied in the Islamic schools and then in the National School. He was a member of the al-Funun Association and the manager of its printing press. Even after the organization ceased to exist, he continued publishing their paper, *Thamarat al-Funun*, under his name. In 1880 he became a member of the administrative council of Beirut and was appointed as a judge in the court of first instance (Nashabi 1981).

Yakoub al-Sarouf studied at the Syrian Protestant College and was among the first class to graduate from it in 1870. He taught Arabic to American missionaries in Sidon for two years. He also became the principal of a school in Tripoli that was founded by the missionaries. He then taught mathematics at the Syrian Protestant College.

Khalil Sarkis studied at the American School in Beirut, which was close to the printing press in which he learned the craft. In 1868 he founded a printing press. He was elected a member of the Knowledge Council of the Vilayet and a member of the commission for al-Sana'i' office. He authored novels, history books and many books that were intended to be taught at school.

Faris Nimr studied at the English school in Beirut. Then he went to Jerusalem where he studied in the English Zionist School. After he graduated in 1874 he was appointed as an assistant to van Dyke at the Beirut observatory, of



which he became president in 1883, and also taught Algebra at the Syrian protestant College. He also taught English at the Patriarchic School of the Roman Orthodox.

Ibrahim Sarkis was the editor of *al-Nashra al-Shahriya*. He studied in the Protestant school in Ibay. When he moved to Beirut he was assigned by the American missionaries to proofread the translated edition of The Bible. He was then assigned as the manager of the printing press, a job that he held until his death.

As mentioned above, not all the newspapers were published by individuals. In the cases where the newspaper was founded by an organization, or even the government as in the case of the newspaper titled Lebanon, the people responsible for these newspapers also came from a similar background to the founders mentioned above. Hanna Bek Abu Sa'b, for example, who was the editor of Lebanon, was considered a genius since a young age. By the age of 14 he was appointed by the son of Bashir Shihab as his chief scribe. He learned how to write poems from Butrus Karamah. Later, he travelled to Malta where he learned the Italian, French and Turkish languages. He studied Arabic literature, grammar, mathematics, logic and astronomy. He was granted several medals by the Sultan 'abd al-Majid. He worked as a scribe for several government officials. After that he founded a printing press in 1852. When Daoud Pasha came to power after the events of 1860 he was appointed as the chief of the Arabic center, a position that he held until his death. He wrote many books that remained unpublished. These books dealt with logic, astronomy and animal science. Hussein Bayham, who was the head of the Syrian Scientific Society when it published its newspaper, was a poet and had held several government posts.

Some of the newspapers were founded by organizations, but most were founded by individuals. All of the entrepreneurs except one were Christians. Why so? Social movement theorists have argued that social movements are formed by groups of people who are excluded from formal channels (Rao, Morrill, and Zald 2000)<sup>7</sup>. A review of the literature concerning the Christians in the Ottoman Empire clearly shows that this was the case for them, even though, theoretically at least, their situation in the nineteenth century got better. Originally, religious identity was the basis of social hierarchy in the Ottoman Empire:

*“In the ottoman system the population of the Empire was organized upon a confessional basis, not upon a territorial or linguistic one. It was composed of religious communities each of which had its own internal organization and was controlled by a religious hierarchy. Socially and culturally each community formed a separate entity, each kept apart from the other. There was no attempt to create uniformity”* (Abu-Manneh 1980, p. 287).

Muslim supremacy was clear in almost all aspects of the empire (Hourani 1957). Imperial decrees from the eighteenth century clearly state that “Christians and Jews should have lower buildings than Muslims”, and that they should even dress differently (Göçek 1996). The main requirement for joining the ranks of the rulers was religious affiliation (Göçek 1996; Masters 2004) as the following passage shows:

*“The inability of the newly emerging institutions to incorporate minorities was evident throughout the social system. For instance, during the period 1859-79, all of the 162 graduates in the school for civil servants were Ottoman Muslims. Also, after 1879, even though the number of graduates increased fivefold, the small*

---

<sup>7</sup> I am thankful to Dr. Min Liu for bringing this point to my attention.

*number of minorities that had started attending hardly increased...Similarly, in the first and only Ottoman school founded abroad in Paris, in 1857, by the Ottoman sultan, even though minorities were actively recruited as students, of the thirty-five minority graduates with identified occupations, only four joined the Ottoman administrative service” (Göçek 1996, p. 84).*

Although the Ottoman reforms of 1839 and 1856, which were referred to collectively as the *Tanzimat*, declared that all subjects were equal regardless of their religious affiliation, the stark reality continued to remain different (Masters 2004). In fact, the period in which the reforms were declared happened to witness a rise in sectarian tensions between Muslims and Christians, especially in Lebanon (Khalaf 1982; Maoz 1982). The Muslims were increasingly hostile to the Christians due to the fact that the new reforms had made them politically their equal while the Christians retained their economic supremacy thanks to their ties with the Europeans whose influence was constantly increasing.

The above biographies of the early founders reflect what has been noted by historians (e.g. Hourani (1957)), and that is that the Christians, through their education, managed to secure jobs as scribes for the governments or translators for European consulates. However, none was able to rise to high places in the government bureaucracy. With rising hostilities and an inability to change the facts on the grounds through government participation and action, many Christians in the Ottoman Empire opted for a new way out: ethnic nationalism. This was most prominent among the population in the Balkans. The Christians of the Arab-speaking Ottoman world responded quite differently:

*“Although sectarian unrest occurred in Egypt and the Fertile Crescent, Arabic-speaking Christian intellectuals and community leaders eventually were able to*

*articulate several options with which to configure their political community as the empire collapsed under the weight of myriad ethnic antagonisms. Their choices were usually very different from those explored by their coreligionists elsewhere in the empire. This was due, in part, to the very crucial fact that Christian Arabs shared a common language and culture with their Muslim neighbors” (Masters 2004, p. 8).*

The Christians of Lebanon did not argue for separation from the Ottoman Empire, nor did they argue for “Christian rights”. Instead, they argued for something that was in line with the Ottoman reforms, and that was Ottoman citizenship with no regards to religion, and they used newspapers to spread their ideas. One of the main advantages of newspapers was that they were written in the Arabic language, which both the Christians and Muslims of Greater Syria shared. In this case, the language of the medium was in itself a message. Butrus al-Bustani argued that “Syria must not become a Babel of languages...as it is a Babel of religions and sects” (Abu-Manneh 1980, p. 291). Through newspapers, the Christians argued for education, because it defied religious boundaries. The next section will show that during the emergence of the newspapers, both the Arabic language and education would be constantly featured in articles because they were the two forces that transcended all religious and sectarian boundaries. Equally important is the identification of education as the ultimate goal of newspapers.

#### **4.3.4 The Framing Process**

The above section showed that the entrepreneurs that started the newspaper industry were comprised of men of knowledge that had close contact

with the government. The two sections before that showed the changes that had taken place in the environment in terms of the resources and also in terms of the cultural and political mood of the people of Lebanon. This section will examine how the newspapers, prior to 1880, identified the problem, proposed the solution, and motivated the people. It will be seen that knowledge, and not political achievements, acted as the driver to the social movement represented by the newspapers. This was an ideal solution because the founding fathers were first and foremost respected men of knowledge and by providing a framework within a field that they were masters of, they managed to convince their audience of their intentions.

<b>Newspaper</b>	<b>Issues Inspected</b>
Al-Jam'iyah al-'Ilmiyah al-Suriyah	January 1868 – May 1869
Hadiqat al-Akhbar	January 1858 – April 1868
Majmoo' al-Fawa'id	January 1851 – January 1856
Nafeer Suriyah	September 1860 – April 1861
al-Basheer	September 1870 – December 1879
al-Taqadom	January 1874 – May 1875
Thamarat al-Funun	April 1875 – December 1879
al-Jinan	January 1870 – August 1879
al-Zahra	January 1870 – December 1870
al-Nahla	May 1870 – December 1870
al-Najah	January 1872 – December 1874
al-Mishkat	April 1878 – July 1878
Lisan al-Hal	October 1877 – December 1879
al-Janna	July 1879 – August 1879
al-Moktataf	June 1876 – December 1879
al-nashra al-Osbo'iyah	January 1872 – July 1877
al-Tabeeb	January 1874 – December 1876
al-Sharaka al-Shahriyah	January 1866 – September 1866

**Table 4.1 Newspaper issues inspected**

From a total of 26 newspapers, I was able to inspect issues relating to 18 of them. Table 4.1 displays the issues that were inspected for this study. All these issues were accessed at the Jaffet Memorial Library at the American University of Beirut. All of the newspapers seemed to agree on the source of the problem that was facing the nation as a whole, and that was lack of education among the

people. In the first issue of *Hadiqat al-Akhbar* (the first newspaper), the author thanks God for providing the people with a great Sultan such as the one that is sitting on the throne now and describes the age as the age of knowledge and science. He says that the sciences have been renewed in this age at the hands of the sultan. The author then asks the people to support this great government that wants to spread knowledge, the arts and civilization. The author says that the greatest aspect of this generation is the science that aims at making life easier. The article goes on to analyze the importance of science in civilization. As it turns out, the word civilization appears in most of the early newspapers. The founders of the newspapers make it clear that the goal of the nation should be to strive for civilization because only on that path can it achieve true progress. In the 22/4/1861 issue of *Nafeer Souriya* the author talks about civilization and its importance for society. This takes up half of the issue<sup>8</sup>,

*“From the previous definition we can see that true civilization does not achieve its full potential and goal unless it was the cause of growth, both physically and intellectually. This is so because if one was only concerned with the physical aspects then he would be nothing but an animal.”*

The author then talks about how the countries of the east were the center of civilization. The author then criticizes the fake civilization in which people merely imitate the habits and clothes of the foreigners. A similar article appears in the 3/4/1870 issue of *al-Zahra*. In the 15/6/1875 issue of *Thamarat al-Funun* the author asks,

*“What was it that caused the human race to break the shackles of ignorance and to become civilized and enlightened other than science?”*

---

<sup>8</sup> All translated pieces from newspapers are my own.

The first issue of the *al-Jam'iya al-'Ilmiya al-Souriya* states that:

*"There is no doubt that if we were to contemplate what would cause us to rise after our fall, we find nothing more suited to the job than building schools and libraries followed by attending scientific organizations."*

The 1/1/1870 issue (first issue) of *al-Zahra* also states that its will to increase the spread of knowledge had led to its publication. In the 1/4/1878 issue (first issue) of *al-Mishkat* the author starts by praising the administrative unit of Syria, and mentions that the governor of Syria told him that newspapers are one of the best methods to spread education.

We can clearly see that education is mentioned as the sole solution to the problem facing the nation. Other newspapers identified the problems of the nation and linked them back to the fact that there was no education in their homeland. In the 26/6/1879 issue of *al-Moktataf*, for example, there is a letter that states that religious intolerance is the source of our problems. The solution to this, according to the first issue of the *al-Jam'iya al-'Ilmiya al-Souriya*, is science because it brings people together no matter their national or religious differences. This same argument is made in the 1/11/1870 issue of *al-Jinan* where the author says that knowledge will lead to a decrease in religious intolerance. It should be noted that religion itself was not viewed as a problem, but what was the problem was the mixing of the religious and political realms. This is explicitly mentioned in the 19/11/1860 issue of *Nafeer Suriya* where the author criticizes mixing the civil and the religious, but also criticizes the people for not being really religious. In the 22/2/1861 issue the author states that,

*"The civilized countries have experienced the damages caused by the mixing of the religious and political spheres and so have created a divide between the two."*

The secular identity of the education that the newspapers sought was clearly illustrated in a debate between *al-Bashir*, the newspaper published by the Jesuits and the scientific newspaper *al-Moktataf*. *Al-Bashir* had launched an attack on *al-Moktataf* because the later had printed an article about the Catholic faith which *al-Bashir* considered to be a bad article. In its response, *al-Moktataf* criticizes the “Jesuits sectarian newspaper” for launching an attack on “our scientific newspaper”. Another example about how the newspapers linked most problems to the absence of science can be found in the 8/1/1869 issue of the *al-Jam’iya al-’Ilmiya al-Souriya* where the author talks about how the French used scientific methods to advance the agriculture industry, which he considers to be the best of all industries. In the 4/4/1869 issue the author criticizes the current state of manufacturing in the Arab world and again argues that education provides the solution.

What role do newspapers play when it comes to education? As it turns out, the newspapers are seen, by their authors, as tools of education which are as important as schools and printing presses. According to the 22/4/1861 issue of *Nafeer Suriya*, the proper tools of civilization are: proper religion, political governance, and education (in which he includes newspapers), and trade. This is not the only place where we see that newspapers are included among educational institutes. In the 1/1/1871 issue of *al-Jinan* the author mentions newspapers along with schools when he is talking about the spread of education: “Ignorance has been lifted from upon us in 1870, and schools and the sciences have increased, and the newspapers have spread everywhere.”



Again, in the 2/4/1875 (first issue) of *Thamarat al-Funun* the author starts by thanking God and praising the Sultan. He then mentions the newspapers along with schools and printing presses as proof that education is spreading:

*“The clearest proof of the generosity of the government is the founding of schools and printing presses, which are the source of knowledge, and the spreading of the newspapers that contain what has happened in the morning and night.”*

In the 1/10/1870 issue of *al-Jinan* there is a letter from the ruler of Mount Lebanon to *al-Jinan* in which he subscribes to the newspaper for a school so that the children will read it in order for them to gain knowledge and to love the government. In the 6/10/1877 issue (first issue) of *Lisan al-Hal* the author starts by thanking God and the Sultan, and then states that newspapers are among the best tools for obtaining education. In fact, during the emergence period, the authors explicitly mention education as the goal of newspapers and not the reporting or analyzing of news items. The first issue of *Thamarat al-Funun* states that:

*“It is not hidden that the newspapers of this age are the cause of progress...because they spread the good deeds of the good people and the bad deeds of the bad people...and it presents to you feasts of useful information.”*

In the 11/5/1870 issue (first issue) of *al-Nahla* the author starts by thanking God. He then mentions the other newspapers favorably and states that they have managed to defeat the armies of ignorance and fix what has been ruined. In the 1/1/1870 issue of *al-Jinan* (first issue) the author also states that the main purpose of starting the newspaper is the spread of knowledge and scientific facts. In the 1/6/1876 issue (first issue) of *al-Moktataf*, the author says that he has started the newspaper as a service to his country because it will spread

useful knowledge. Again he also mentions that knowledge is the solution to the various problems that face the homeland.

The main motivation method used by the newspapers was to remind the people of the glorious past of their nation and promise them that through education, more specifically through newspapers, they can reclaim what they have lost. In the first issue of *Hadiqat al-Akhbar*, the author reminds the people of the past glorious days of this part of the world and argues that they could repeat their previous success with the aid of education. In the 22/4/1861 issue of *Nafeer Suriya* the author reminds the people of the glorious past and argues that they can reclaim their place among advanced nations. In the first issue of *al-Jam'iya al-'Ilmiya al-Suriya* the author also praises the past history of the Arabs and asks why the Arabs don't use science today to emulate their predecessors. The issue talks favorably about the past and critically about the present situation of the Arabs. The 10/4/1868 issue of the same newspaper also glorifies the history of the Arabs at a time when the Europeans were fighting barbarically. The same issue has a poem that says that Beirut has been dark because of the absence of science, but that it is now a shining planet thanks to the efforts of Sultan 'abd al-'aziz. In the 4/4/1869 issue there is an article that considers the history of Great Syria to be a glorious one and argues that the people of the West had taken their culture from Syria when they occupied it in the past. These same ideas are reflected upon in most of the other newspapers. In the 1/1/1870 issue of *al-Jinan* (first issue) the author states that the rays of knowledge were first emitted by the East towards the West, which has used this knowledge well. In the 1/3/1870 issue of *Hadiqat al-Akhbar* the author asks if the Arabs will one

day return to their glory and answers that it is inevitable that time will take us back to what politics has ruined. In the 15/2/1876 issue the author says:

*“If we were to ask history about our past we would see the beauty which was...we have now stripped our nation from this beauty...I doubt that there is anyone among us who does not acknowledge the fact that we are in retreat and that we have let go of what causes our progress...The progress of our homeland is dependent on love, science, and industry.”*

In the 15/11/1878 issue the author says:

*“Science is the best thing that we have, and the scientists are the heirs of the prophets. We have not done a tenth of what any civilized nation should do in order to preserve the sciences which our ancestors have nourished and delivered to the Europeans who we have now made our idols.”*

In the 12/8/1878 issue of *Lisan al-Hal* the author argues that racial differences do not matter when it comes to education. What matters are the tools that each nation has. In one case, in the 6/3/1862 issue of *Hadiqat al-Akhbar*, the author talks about some financial troubles that *al-Jawa'ib* was going through. He then blames the people for not buying the newspaper in large quantities and criticizes them for not supporting something that is a sign of civilization.

This emphasis on education and progress is reflected in the names of the newspapers. Table 4.2 shows the names of these newspapers in Arabic and their English translation<sup>9</sup>. Only one newspaper had a title that was the name of a geographic area. A recurring theme in the names of the newspapers is the idea of a paradise of knowledge. The word heaven appears in three titles.

---

<sup>9</sup> Some of the translations were taken from Ayalon, Ami. 1995. *The Press in the Arab Middle East: A History: A History*: Oxford University Press, USA.

Name (Arabic)	Name (English)	Name (Arabic)	Name (English)
حل المسألتين الشرقية و الغربية	The solution to the eastern and western question	اخبار طبية	Medical news
أخبار عن انتشار الانجيل في العالم	News about the spread of the Bible around the world	كوكب الصبح المنير	The planet of the shining morning
النشرة الاسبوعية	The weekly bulletin	لبنان	Lebanon
البشير	The Herald	لسان الحال	Voice of the present
التجارة	Trade	المجمع الفاتيكانى	The Vatican composite
التقدم	Progress	مجموع الفوائد	Collected useful lessons
ثمرات الفنون	Fruits of knowledge	اعمال الجمعية العلمية السورية	The works of the scientific Syrian society
الجنان	Gardens	المشكاة	The niche
الجنة	Paradise	المقتطف	Selections
الجنيانة	Little garden	المهماز	The spur
حديقة الأخبار	The garden of news	النجاح	Success
الزهرة	The flower	النحلة	The bee
الشركة الشهرية	The monthly corporation	نفير سورية	Syrian clarion

**Table 4.2 Names of the early newspapers**

We see that one newspaper is named “The garden of the news”. The founder of this newspaper had originally intended to name it “The Shining Dawn”. Another newspaper is named “the Flower” while another is named “the Bee”. In the first issue of “the Bee”, the author talks about how he intends for his newspaper to act like a bee and fly from flower to flower (in reference to other newspapers) picking the best from each. We can also see a newspaper named “Fruits of Knowledge”. In the 15/11/1870 issue of *al-Jinan*, there is a poem that praises the newspaper by comparing it to a paradise that contains the most delicious fruits and a river of knowledge. In the 15/7/1876 issue of the same newspaper there is a letter written to the newspaper that says:

*“...and the gardener would open up his heavens and we would enter into them while enjoying the breeze and the taste of their fruits and drink from the water of*

*literature which bursts from the fountains of wisdom and acumen so that we shall never go thirsty again."*

There are even instances in which the Lebanese newspapers refer to foreign newspapers in the same way. In the 13/1/1876 issue of *Thamarat al-Funun* the author mentions that a copy of the Iraq based newspaper *al-Zawra'* has been sent to him and that he was happy to "explored its rich garden".

We can see that newspapers were thought of as a garden that contained fruits. These fruits represented the knowledge contained within the newspaper. I inspected the names of all newspaper published between 1880 and 1890. There were a total of 14 newspapers founded in Lebanon during that period. None of the names of these newspapers referred to a garden or a paradise. As I argue later, the reason for this is that the newspaper industry matured during the end of the 1870s and the founders re-framed their movements and re-defined their goals.

While identifying education as their goal, the early newspapers constantly stressed the importance of the Arabic language. As mentioned in the section that dealt with the founding fathers, language was the crucial factor in the adoption of a secular ottoman citizenship by the Christians of Lebanon. By stressing the importance of language, the founders of the newspapers, who were mostly Christians, were strengthening the bonds between them and their Muslim compatriots. Language, just like education, transcended the religious boundaries. In the first issue of *Nafeer Suriya*, Butrus al-Bustani, in asking the people to unite and to abandon animosities, states that a single language binds the people. It was common for newspapers to include translated novels in their issues, either on the lower half of the first page or on the last page. The son of Butrus al-Bustani,

Salim, constantly translated European novels to the Arabic language and printed them in *A'mal al-Jam'iya al-'Ilmiya al-Suriya* and in his father's magazine *al-Jinan*. In the 5/12/1868 issue of *A'mal al-Jam'iya al-'Ilmiya al-Suriya*, one of the reasons he gave for this was that he wanted the people to remain interested in the Arabic language. Newspapers also contained many poems and in some instances these poems appeared on the first page of the newspaper. Half of the 3/5/1868 issue of *A'mal al-Jam'iya al-'Ilmiya al-Suriya* is a poem. Articles that explained some vague points in Arabic grammar also constantly appeared in the newspapers. The discussions regarding the Arabic language turned into a heated debate when the newspaper *al-Jawa'ib*, which was published by a Lebanese in Turkey, decided to write a critique of some of the literary works of the famous Arabic poet Naseef al-Yazaji. Naseef's son, Ibrahim, responded to these critiques on the pages of *al-Jinan*. Given all of the above, we would expect that a considerable proportion of the newspapers to be dedicated to issues that dealt with the Arab language. Almost one quarter of the 4/3/1868 issue of the scientific magazine *A'mal al-Jam'iya al-'Ilmiya al-Suriya* was occupied by poems, translated novels or technical articles dealing with Arabic grammar. Almost one full page out of a total of four in the 4/12/1858 issue of *Hadiqat al-Akhbar* was also taken up with such subjects as well. Of course, there were instances in which there was nothing that dealt with the Arabic language, like the 3/5/1860 issue of *Hadiqat al-Akhbar*, but in general the proportion dedicated to such issues was considerable.

According to Swaminathan and Wade (2001), "cooperation rather than competition is the norm as founders attempt to mobilize resources through collective action..." (p. 286-287). This was certainly true in the case of the Lebanese newspaper industry during its emergence period. The newspapers

constantly quoted from one another and when any new newspaper was published it was welcomed by the ones already in the industry. *Al-Bashir* quotes articles from *Hadiqat al-Akhbar* in the following issues: 6/5/1871, 3/6/1871, 7/10/1871, 12/4/1878, 10/5/1878, 9/8/1878, 7/3/1879 and 8/8/1879. The 4/5/1872, 7/9/1872, 1/2/1873, 5/4/1873 and 7/2/1879 issues quote from *al-Janna*. The 12/7/1878 and 12/9/1879 issues quote *Lisan al-Hal*. *Thamarat al-Funun*, in the 16/3/1876 issue, quotes *al-Bashir*. In the 16/11/1876, 13/4/1876 and 14/6/1877 issues the author quotes *al-Janna*. The 14/2/1878, 14/3/1878, 16/5/1878, 14/6/1878, 11/7/1878, 8/8/1878, 12/12/1878 and 14/4/1879 issues quote *Hadiqat al-Akhbar*. The 14/10/1879 issue quotes *al-Janna*. In the 13/10/1877 issue the author talks about the new newspaper *Lisan al-Hal* and wishes it the best. *Al-Jinan* on the other hand quotes *al-Bashir* in the 15/10/1872 and 15/11/1879 issues and *Hadiqat al-Akhbar* in the 1/4/1872 issue. It also mentions the scientific newspaper *al-Tabeeb* favorably in the 1/6/1875 issue. In its turn, *al-Tabeeb*, in its 1/6/1876 issue mentions the newly created *al-Moktataf* calling it useful and wishing it the best. *Al-Nahla*, in its 1/7/1870 issue quotes *al-Janna* and in its 6/4/1872 issue quotes *al-Bashir*. *Al-Janna* quotes *Hadiqat al-Akhbar* in the 17/8/1879 issue. *Lisan al-Hal* also quotes *Hadiqat al-Akhbar* in its 6/10/1877, 8/11/1877 and 6/12/1877 issues. In its 6/2/1879 and 6/11/1879 issues it quotes *al-Janna*. The 1/10/1877 and 1/1/1877 issues of *al-Moktataf* quote *Lisan al-Hal*. The 1/10/1878 issue quotes *Thamarat al-Funun* while the 1/12/1878 issue quotes *al-Taquadom*. In the 1/6/1879 issue the author mentions that *al-Nashra al-Osbooiya* is published again. *Al-Nashra al-Osbooiya* quotes *al-Junaynah* and *al-Bashir* in the 8/4/1873 issue, *al-Jinan* in the 10/2/1874 issue

and *al-Moktataf* in the 5/6/1877 issue. It also praises *al-Moktataf* in the same issue and asks the people to help make it successful.

The above is just a small portion of collaboration that was going on between the newspapers. However, there were also instances of discord. In some cases some newspapers criticized certain news items appearing in other publications as being not true or at least inaccurate. However, one of the most extreme, and most famous, cases of enmity was that between the founder of *al-Jinan* and the founder of *al-Nahla*. Interestingly, this dispute served to strengthen the collective identity that the newspapers themselves were trying to project, and that was discussed in the preceding section, i.e. newspapers as educational tools. The debate started when *al-Nahla* criticized *al-Jinan* for some items that it published. Soon the debate degenerated into a personal attack between the two founders. Interestingly the criticism of each party revolved around the education of the other person. In the 10/12/1870 issue of *al-Nahla* the author criticizes al-Bustani, the founder of *al-Jinan*, for studying “Voltaire, the master of the hypocrites”. It will be remembered from the section about the founding fathers that the founder of *al-Nahla* was a priest. Al-Bustani retaliates by stating that al-Sabounji was in fact not a learned man, to which al-Sabounji responds by ridiculing the education of al-Bustani. This debate clearly shows that by showing that the newspaper founder was not a learned man, you would be actually showing that the newspaper itself was not a valid educational tool. Such an attack strengthens the educational aspect of the newspaper and highlights what is truly important in these organizations.



#### 4.3.5 The Re-framing Process

Around the middle of the 1870s, the framing process utilized by the newspapers started shifting from talking about the need for education and started talking about the need for political reform. In fact, the newspapers redefined their goal as being one of delivering and analyzing news, in place of the older educational role. In the 1/9/1875 issue of *al-Jinan* we see one of the first acts of this re-framing process. The author talks about the importance of newspapers in analyzing events so that the average person would be aware of what is going on. By 1877 we start seeing more of this new goal orientation in more than one paper. In the 15/4/1877 issue of the same paper the author says:

*“The printed material in general, and the political, scientific, reformative, industrial and medical newspapers in particular are the cause of the spread of knowledge. This knowledge is the foundation upon which progress and growth form. And they are among the greatest methods with which a nation is to be aware of its rights and wrongs, and to project its needs and to preserve its rights from the trespassing of the officials. The freedom of the press is proof of the advance of any nation.”*

In the 15/6/1879 issue of the same newspaper there is an article about freedom of the press. Newspapers, according to the author, scrutinize the working of the government officials and so officials have to be careful and alert:

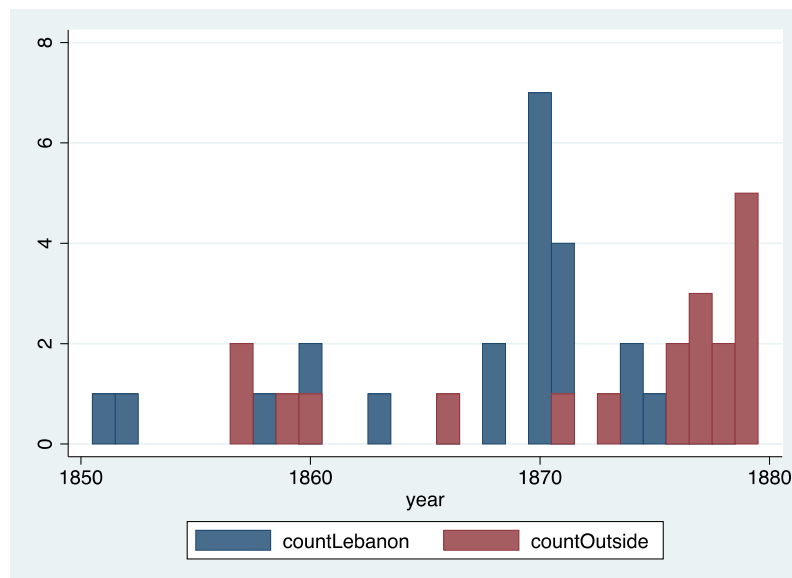
*“The newspapers seek to inform the public about the intentions of the government, while also informing the government of the needs of the people...One of the greatest and most critical of issues is that which relates to the freedom of the press.”*

Again, in the 17/11/1879 issue of *Thamarat al-Funun* the author criticizes the suspension of a newspaper in Aleppo and declares that this is tantamount to a declaration of tyranny. Such language was absent previously although many

newspapers were suspended for different reasons. In the 15/1/1879 issue the author says that reforms should be concentrated on spreading knowledge and differing viewpoints and questioning officials through freedom of speech and newspapers. This same idea is reflected in the 15/12/1879 issue of *Thamarat al-Funun* where the author defends the role that his newspaper is playing by stating that the job of newspapers is to inform the public about government actions and critically observing their implementation:

*“The job of the newspapers is to inform the people of the sources of such actions and to state the available resources and to highlight what good has been done.”*

If we look at Figure 4.3 we can see that around the time that the newspapers were going through this re-framing process, the number of founding events in Lebanon decreased while the number of founding events outside Lebanon by Lebanese founders increased.



**Figure 4.3 Newspaper founding events up to 1879**

Industry historians (e.g. al-Rifa'i (1967), Tarazi (1933), Mruwi (1961), Ayalon (1995)) have unanimously argued that the ascension of Sultan 'abd al-Hamid to the throne resulted in the persecution of the newspapers. In fact, the total

number of newspapers in the Ottoman empire declined after the arrival of the new sultan (Göçek 1996). While there is no doubt that the new sultan was less tolerant than his predecessors, the re-framing process itself also caused some of the damage that was to come. In order to investigate this, I sampled one issue per month of each newspaper from 1851 until the end of 1879. The total issues sampled were 659 with a total of 9826 pages. Of course, these pages differed in size. Some were A4 sizes, others were close to the sizes found in daily newspapers today. I then used content analysis in order to code the data (Riffe, Lacy, and Fico 1998). Specifically, I went through the issues to see any references to the authorities and coded a variable as “Supportive” if all references in the issue were positive, “Critical” if all references were negative, “Both” if the same issue contained positive and negative references, and finally “None” if there was no references whatsoever or if the references were neutral in that they did not criticize nor support the government. Table 4.3 below shows the result with the cells representing percentages. We notice that up to 1870 there were no critical references to the government in any newspaper. In addition, there were much more positive references than no references at all. In the lower bottom of the table we see that some issues both criticize and support the government while a small number only criticize the government. Also noticeable is the fact that the number of issues that do not mention the government is around the same as the number of positive references. The table shows us clearly that the newspapers have started operating differently than before. This happens around the same time as the re-framing process mentioned above. Therefore, it seems that the newspapers started looking at the government in a more critical way. Was this manifested in a way other than the mentioning of the government?

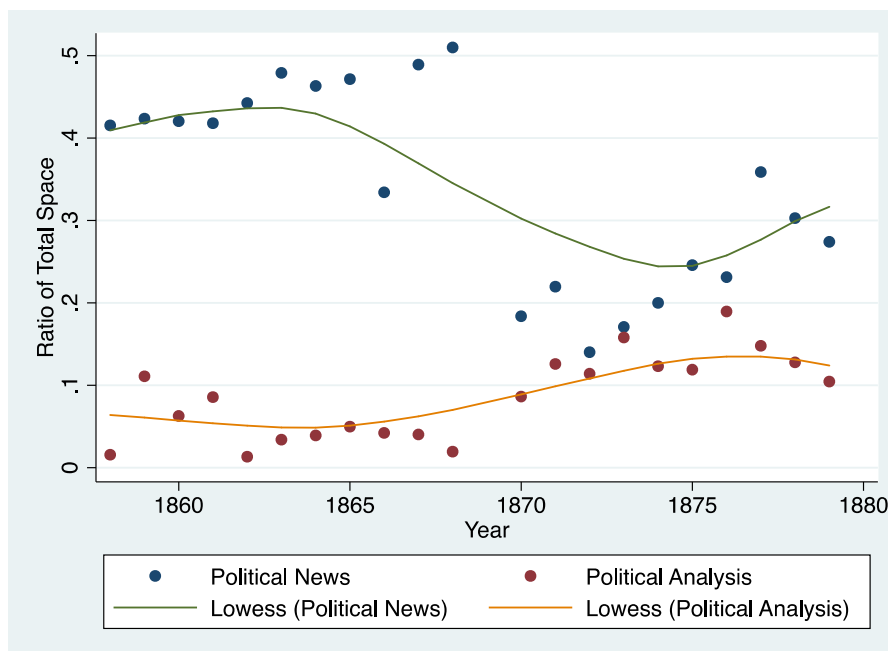
Year	Both	Supportive	Critical	None
1851	0.00	0.00	0.00	100.00
1852	0.00	0.00	0.00	100.00
1853	0.00	0.00	0.00	100.00
1854	0.00	0.00	0.00	100.00
1855	0.00	0.00	0.00	100.00
1856	0.00	0.00	0.00	100.00
1858	0.00	83.33	0.00	16.67
1859	0.00	41.67	0.00	58.33
1860	0.00	50.00	0.00	50.00
1861	0.00	86.67	0.00	13.33
1862	0.00	90.91	0.00	9.09
1863	0.00	83.33	0.00	16.67
1864	0.00	92.31	0.00	7.69
1865	0.00	87.50	0.00	12.50
1866	0.00	45.00	0.00	55.00
1867	0.00	83.33	0.00	16.67
1868	0.00	40.00	0.00	60.00
1869	0.00	40.00	0.00	60.00
1870	0.00	50.00	0.00	50.00
1871	0.00	25.00	4.17	70.83
1872	0.00	26.19	0.00	73.81
1873	0.00	31.91	0.00	68.09
1874	6.35	23.81	1.59	68.25
1875	0.00	29.31	0.00	70.69
1876	0.00	36.36	0.00	63.64
1877	0.00	51.72	0.00	48.28
1878	1.56	28.12	1.56	68.75
1879	3.23	32.26	1.61	62.90

**Table 4.3 Breakdown of the references to the government**

In addition to coding the above variable, I coded two more variables. The first reflected the amount of space taken up by political news and the second reflected the amount of space given to political analysis. The variable *political news* represented factual information while the variable *political analysis* represented pieces in which the author gave his (there were no females in the industry at that time) opinion regarding what had happened or what was about to happen. Based on the re-framing process described above, we would expect the amount of analysis to increase over time. If so, then this would indicate that the population was certainly maturing in both their new goals and in the methods implemented to reach these goals.

Figure 4.4 shows the scatter plot of the total amount of space dedicated to each category over the years 1851-1879. The figure also shows the loess graph

of both variables. I did not include the purely scientific newspapers in this graph because we are interested in the change between the two political variables and the scientific newspapers would not contribute to this change. This way we can see how the newspapers that dealt with political issues changed their content structure. This graph clearly shows that the political analysis sections of the newspapers took space away from the political news. Since space is a zero sum game, if one increases then the other decreased. The two loess graphs appear to move in harmony. In addition, we can see that at the far end of the graph the dynamic starts to reverse. The persecution of the newspapers led to many newspapers closing down, changing their content, or in some cases, moving to Egypt.



**Figure 4.4 Political content analysis excluding the scientific newspapers**

We can conclude from the above analysis that the re-framing process, along with the rise of ‘abd al-Hamid, had severe consequences. Previous studies only discussed the effect that the new Sultan had on the wellbeing of the organizations in the industry. The evidence shown here gives us a new

perspective on the issue. In addition to a sultan with an oppressive nature, the newspapers suffered because they had matured. They no longer had a view of themselves as being a paradise to which the reader was invited. They now demanded a more serious role, one that enabled them to question the government's actions and intentions. This was reflected in their definitions of the newspapers' goals as well as in the type of content that they published. As long as they were regarded as educational tools the government did not suppress them, but once they matured the government's response changed. The mere act of criticizing the government does not in itself constitute a re-framing process for newspapers. A change in the nature of the regime or a change in the economic or cultural aspects of society might entail a change in the way that newspapers approach their subject matter. If so, then such a change does not warrant to be labeled as a "re-framing process". The situation highlighted by the above discussion is quite different though. First of all Table 4.3 and Figure 4.4 above show that there was a change just after the beginning of the 1870s. Newspapers started analyzing events more than before and favorable mentions of the government started to decrease. Second, we have seen that around the same period newspapers started defining their goals in a completely different way. Newspapers were less regarded as tools of education and more regarded as tools to monitor government functioning. This was also reflected in the names of the newspapers that no longer reflected the image of a garden or paradise. These changes in the industry were not a result of a change in the environment. No major events took place around that period that would cause the newspapers to be less pleased with the government than before. There were also no significant economic or cultural shifts. The Ottoman government was in a state of steady

decline, but this had been the case for tens of years as shown in the history section of this paper. Finally, while the era of ‘abd al-Hamid was to proved to be severely restricting, this was not the case from the beginning of his ascension. ‘abd al-Hamid had been put on the throne because the bureaucrats believed that his predecessor was not responding properly to challenges. ‘abd al-Hamid had pledged to modernize the empire as a precondition for his becoming the new sultan. In fact, one of his first acts was the declaration of the constitution and parliament. This is why this thesis argues that the re-definition of the goals and the re-structuring of the contents of the newspapers were a result of a re-framing process undertaken by the newspapers themselves, and not the result of macro-events. According to Hannan and Freeman (1984) core changes cause an increase in organizational mortality. It has been shown by Minkoff (1999) that these core changes have a significant, and negative, impact on social movements. The re-framing process discussed above is clearly a case of a change in a “core feature” since it modifies the current identity (Hannan, Baron, Hsu, and Koçak 2006) and it resulted in a significant negative impact on the newspaper industry as can be seen from Figure 4.3.

#### **4.4 Discussion and Conclusion**

Organizational ecologists have recently directed their attention at the question of how do industries emerge. More specifically, they have attempted to answer the question whether a unified identity was a necessary condition for successful emergence. The understanding of this dynamic requires an understanding of collective identity. Social movement scholars have studied the framing process in which social movements alert their audience to problems,

propose a solution, and provide the necessary motivation. Both fields provide promising tools that will help shape our understanding of the emergence process. Some organizational theorists have used the concept of social movements in order to explain the emergence of either new market niches or entirely new industries. However, the use of this concept has been relatively lax. Touraine (1985) has argued for a more systematic approach to the question of what constitutes a social movement. In addition to this identification problem, another problem that has plagued social movement studies is the determination of the social forces that act as the trigger to the social movements. Two groups have emerged as a result of this question, where one argued that the intensification micro-level grievances were the trigger while the other group argued that macro-level resource-mobilization considerations were the central actor, with grievances being used merely as tools.

This chapter has attempted to address both concerns. With regards to the identification problem this chapter has showed that the newspaper industry in Lebanon, in its early formation days, constituted what is regarded as a social movement. Industry members stated that the nation faces the problem of falling behind when it comes to civilization economically, politically and culturally. They proposed that the only solution to this problem was the spreading of education and argued that newspapers were one of the most important tools in doing so. Finally the newspapers motivated the people by arguing that the previous glory days of the East can be repeated through this strive for education. Most of the newspapers have ascribed to themselves the role of educational tools and most of the founding fathers were themselves men of knowledge. In this regard, this



chapter has shown that the newspapers have projected a unified identity to their audience members in that they clearly had the same tools and the same goals.

Olzak and Uhrig (2001) state that the historical and social context of social movements matter and that any study of social movements needs to take them into consideration. This is certainly true in the case of the Lebanese newspapers industry. While Olzak and West (1991) found that an increase in ethnic conflict led to the founding of ethnic newspapers in the U.S., this chapter has shown that an increase in religious conflict led to the emergence of newspapers which argued against religious divisions and instead opted for a secular ottoman identity. In addition, the goal of the Lebanese newspapers during this emergence period differs from the goals attributed to the newspapers of other nations. In 1785 the founder of The Daily Universal Register, which later became The Times of London, stated the objectives of his newspaper as follows:

*“Such, it is intended, shall be the UNIVERSAL REGISTER, the great objects of which will be to facilitate the commercial intercourse between the different parts of the community, through the channels of Advertisement; to record the principal occurrences of the times; and to abridge the account of debates during the sitting of Parliament”* (Ayalon 1995, p. 6).

In addition, this chapter has contributed to the micro-macro debate by showing that in the case of the Lebanese newspaper industry neither of the competing explanations provides an adequate explanation on its own. Prior to the formation of the industry the Lebanese landscape witnessed severe changes in both the grievances faced by the people and the resources made available to them. The arrival of the missionary groups and the subsequent introduction of

the printing press and modern educational institutions were accompanied by the dissolving of the old stable social hierarchy and the creation of a new one in which sectarian identities were brought to the forefront. This new social system was the source of much of the population grievances. The place in which these dynamics mixed the most was in the coastal city of Beirut. Beirut was gradually emerging as the new political, economical and cultural center of Lebanon due to the introduction of European capital, and the demographic change that was caused by Christians escaping their persecution at the hands of the more powerful Druze landlords. The Protestant missionaries found a hard time permeating the Maronite areas on Mount Lebanon due to the influence of the Catholic Jesuits and so ended up founding most of their schools in Beirut or the areas around it. Most of the founding fathers of the newspapers were themselves students of these schools and eventually chose to live in the city due its many promises. In addition to that, almost all of the founding fathers were Christians. This chapter has shown that Christians were excluded from the formal channels because of the religious nature of the Ottoman Empire. Even though the Ottomans had introduced reforms in order to make all citizens equal regardless of their religion, the hard facts on the ground hardly changed. It was around that time that the Christians of Lebanon launched their newspapers in the language that was common between them and the Muslims. These newspapers stressed the importance of the Arabic language. It was clearly stated by the most prominent of all Lebanese intellectual figures, and a founder of several newspapers, Butrus al-Bustani, that it was not acceptable to allow the spread of more than the Arabic language across the Syrian landscape. The two main concerns of the newspapers, education and the Arabic language, were the two

forces which these Christian founders used to create a unity which was more inclusive to them than the Islamic identity of the empire.

All of the above were necessary ingredients in the founding of the newspaper industry. Contrary to what some industry historians had argued, the early newspapers identified themselves as educational tools and were not concerned with battling the Ottoman authorities. In fact, praise of the Ottoman Sultan was a recurring theme in most newspaper issues. Social movements can benefit from having proponents within the ranks of the ruling authorities (Olzak and Ryo 2007; Olzak and Uhrig 2001), a view also shared by institutional theorists with regards to organizational forms (Scott 2000). This was surely the case in the emergence period of the Lebanese newspaper industry. The main goal of this social movement, the spread of education, which transcended religious boundaries, was inline with the goals of the *Tanzimat*, which was the spread of a single ottoman citizenship with no regards to religion. Instead of being battle arenas, the newspapers saw themselves as paradises of knowledge. With time, the newspapers saw themselves in a different light and perhaps they realized that it was not enough to spread knowledge in a scientific way, but that what was needed was a more analytical and critical view of the government's function. This, coupled with 'abd al-Hamid's reign, led to strong persecution of the newspapers.

While this chapter has attempted to show qualitatively how the newspaper industry emerged, there remains to be seen quantitatively how its identity was formed. I have argued above that the newspapers themselves argued along similar lines and were concerned with similar matters, but, as mentioned previously, this is only one dimension of identity. The other

dimension is related to the category spanning activities of the newspapers. More specifically, did the newspaper talk about similar topics or was each concerned with its own specialty of topics? In the words of organizational ecologists, did the early newspapers span the same categories? In addition, was the category spanning, if the dynamic actually existed, systematic or random? Finally, how did clusters form from the dynamics of category spanning? The answers to these questions will help complete the picture of collective identity of the newspapers.

## **5 Chapter Five: Category Dynamics and Cluster Spanning During Industry Emergence (1851-1879)**

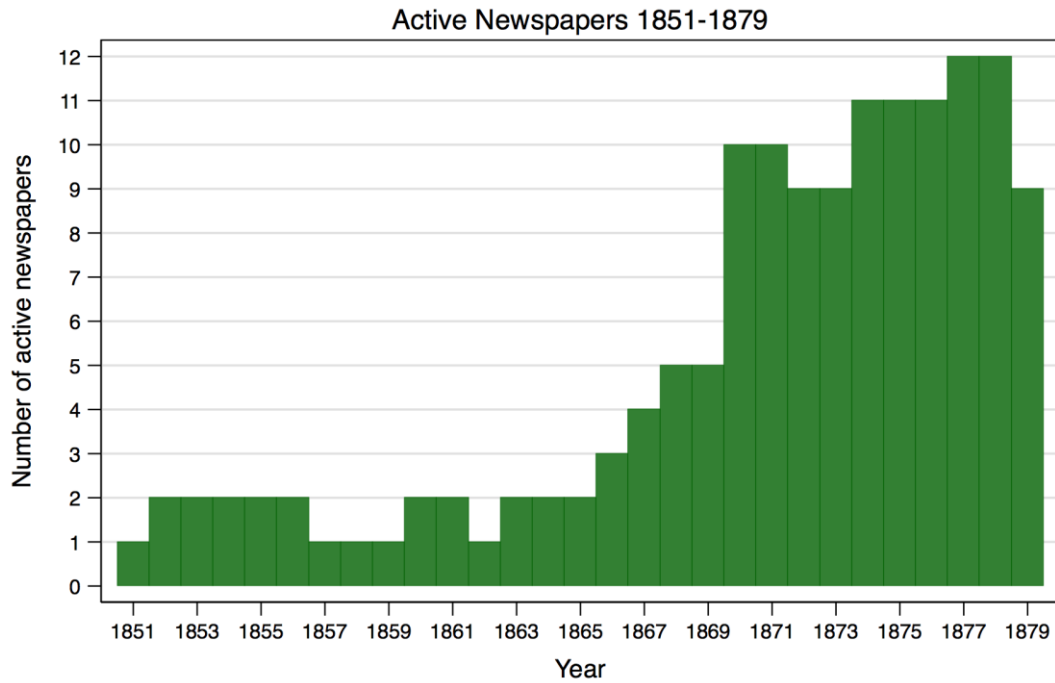
### **5.1 Introduction**

Chapter 4 showed that the newspaper industry emerged as a social movement seeking to spread education and language in order to unify the people. The chapter also showed that the part of identity that is related to the founding fathers and the stories told by the newspapers was uniform across the newspapers. The newspapers managed to do so using a framing process, thus projecting a unified identity. In this chapter I will study the category spanning dynamics and the cluster formation of these early newspapers. This will enable us to look at another dimension of identity. In doing so we will see that while along some dimensions uniformity is necessary, the same is not true with regards to other dimensions.

### **5.2 Data Collection**

The previous chapter provides the empirical setting of the analysis that follows. Figure 5.1 shows the number of active newspapers in the period 1851-1879. In the previous chapter, I studied the collective identity of the newspapers as projected by the founders themselves through the pieces they wrote for their papers. In this chapter, I study that part of the identity that is reflected in the different categories and groups that the individual newspapers span. In order to

do that, I went through the newspaper collections at the Jafet Memorial Library at the American University of Beirut. From the 26 newspapers, I was able to inspect issues relating to 18 of them.



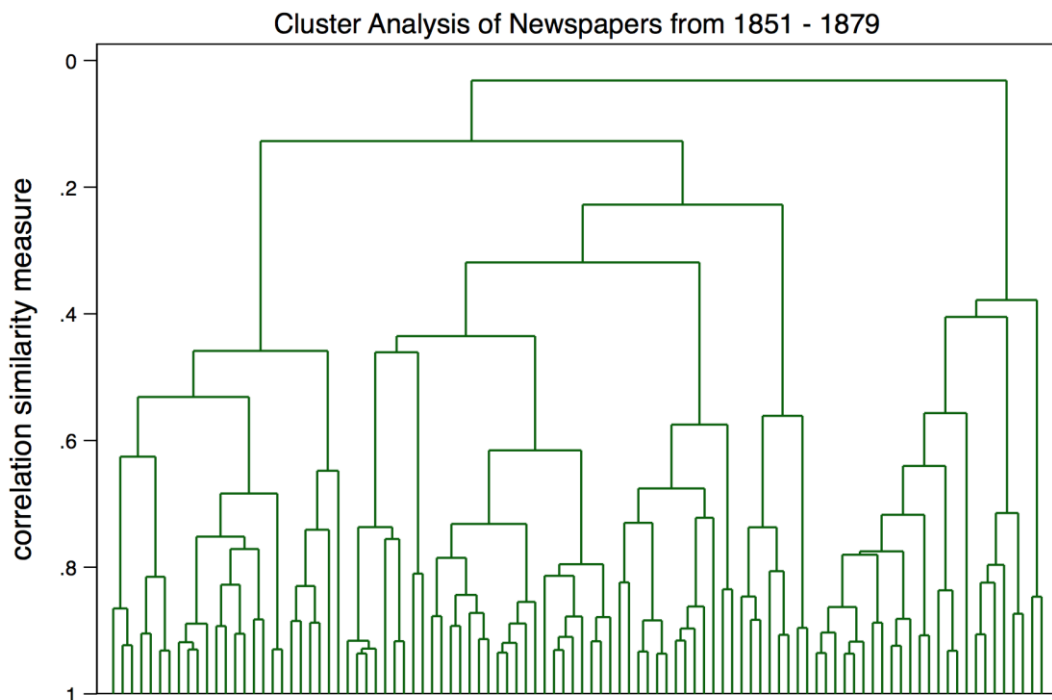
**Figure 5.1** The number of active newspapers in Lebanon 1851-1879

The data collection methods were discussed in detail in chapter 3. Nine different categories were identified and they are: politics, economics, social issues, knowledge, literature, sport, art, advertising, and “other”. I sampled one article from each month for each newspaper title and went through it determining the space dedicated to each of the above nine categories. The final dataset consisted of 659 entries.

### 5.3 Cluster Analysis

I used Stata version 12 in order to conduct the following analysis. Since my intention is this paper is to study whether the individual newspaper issues in the time period 1851-1879 formed one or more identities, I have chosen the

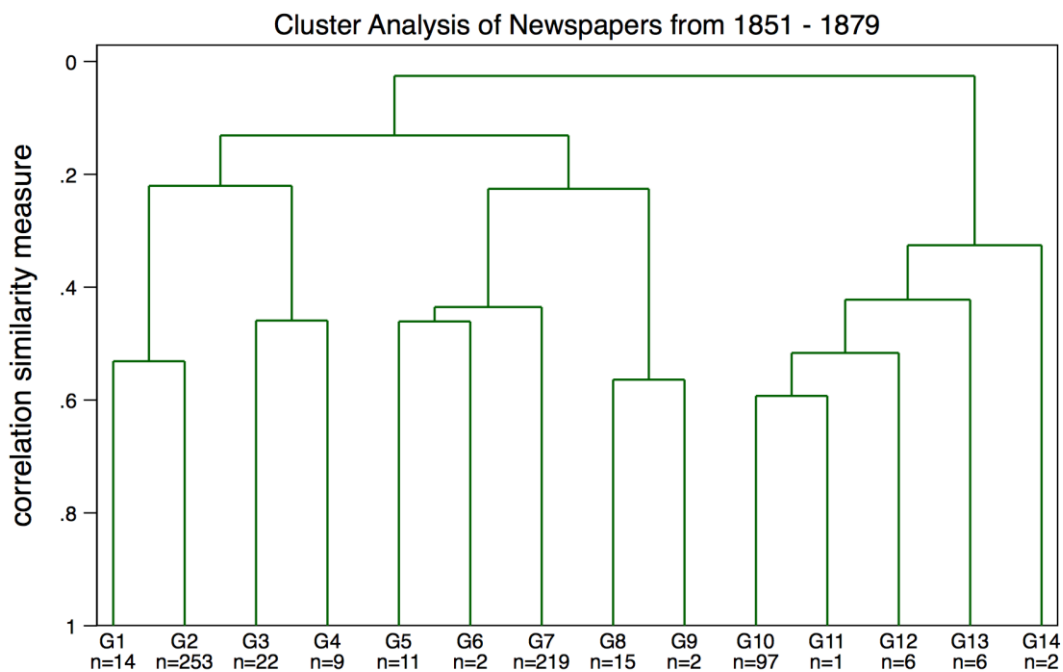
technique of Cluster Analysis as a starting point in the analysis. As a first step in the cluster analysis, I pooled all the issues in the period 1851 – 1879 together and performed a cluster analysis using as a measure of proximity the Pearson correlation measure and using the weighted-average as the linkage method. Figure 5.2 shows the dendrogram for this analysis.



**Figure 5.2 Dendrogram for all issues from 1851-1879**

The dendrogram starts at the bottom with 100 groups after some fusion have taken place because the number of observations is too great and therefore it is not possible to display the figure before any clustering has happened. The y-axis represents the correlation at which two clusters are fused together. Fusions at higher values of the correlation measure indicate greater similarity between the clusters. We can see that many fusions take place at values between 1 and 0.8. This means that many issues are actually similar with regards to their content.

The figure generally shows 5 distinct clusters forming at values which are higher than 0.4. It appears that the newspapers during the period under study did not span categories identically and therefore formed different clusters with regards to their content. The fact that several clusters formed at high values of correlations indicate that some kind of group formation was underway. In order to be more conservative, I look at the groups which have already been created at correlation values which are greater than 0.6. Correlation values greater than this number are generally considered to indicate significant correlations. Figure 5.3 is a dendrogram that only shows the clusterings that take place at values smaller than 0.6. At the bottom of the figure we see that by the time we have fused groups with correlations between 1 and 0.6 we end up having 14 groups.

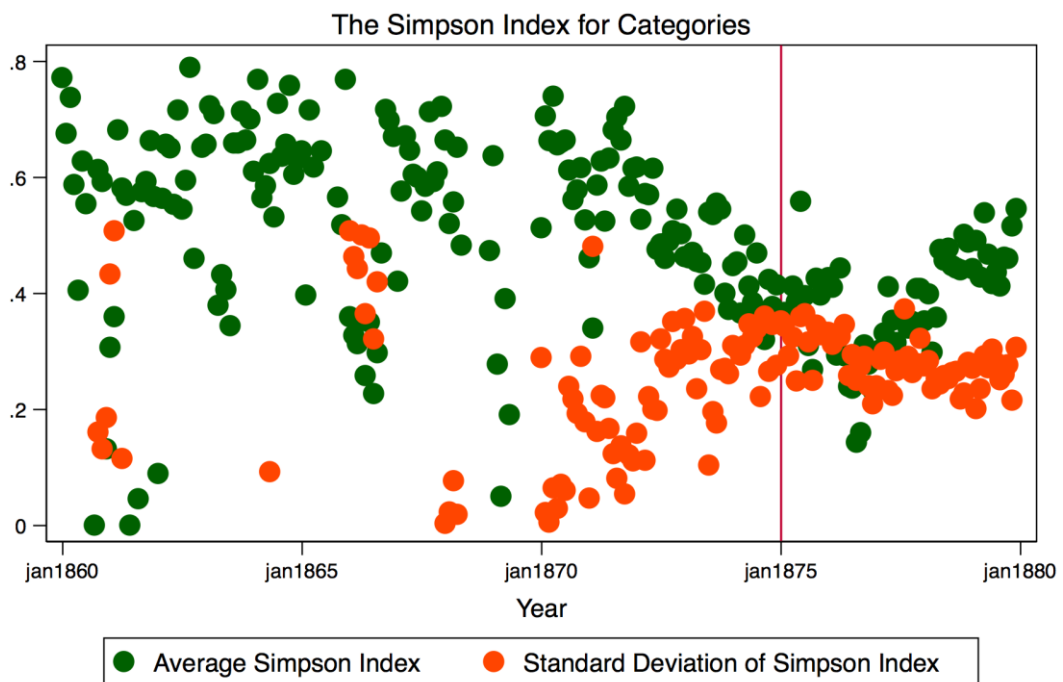


**Figure 5.3 Dendrogram showing clusters forming after the 0.6 value**



The numbers below the group labels represent the number of newspaper issues in each group. The largest group contains 253 issues, the one after that 219 issues, then after that 97 issues followed by 22, 15, 14 and 11 issues. All other groups contain less than 10 newspapers.

In the previous chapter of this thesis we saw that somewhere in the 1870s the newspapers modified their goals. That chapter argued that the newspapers matured during that period. Therefore, it would be interesting to see if the contents of the newspapers also took on a different trajectory during that same period. To look into this issue more deeply I divided the above analysis into two different periods. According to the previous chapter, the change in the content of the newspapers took shape somewhere around 1875. Therefore, I produced Figure 5.4 which shows two scatter plots.



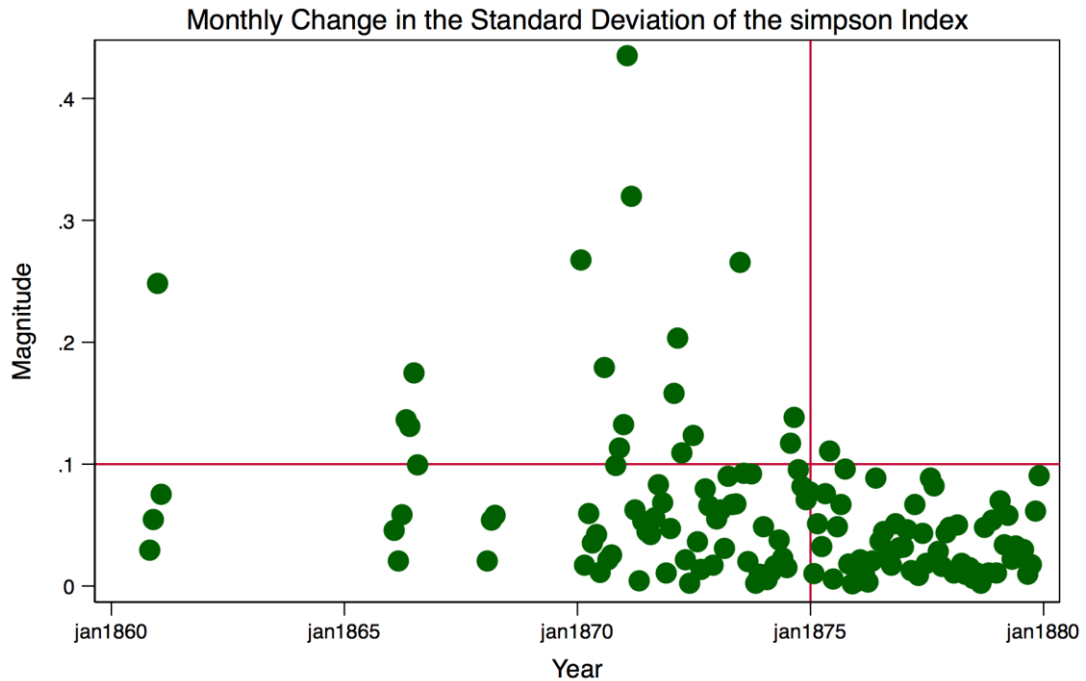
**Figure 5.4 Simpsons index for newspapers between 1860 and 1879**

Both of these plots refer to Simpson's index of the issues. The first step in this analysis is to study how the newspaper issues spanned the different categories over the duration of the study period. In such a case, it would be useful if we were able to summarize the category spanning activity using a single number. Building on the work of Hannan, Pólos, and Carroll (2007), we can think of the grade of membership  $\mu$  in a category as the proportion of that issue which is dedicated to the category. So, for example, if a newspaper consisted of eight pages in which two were dedicated to sports, then the grade of membership of that issue in the category sports would be  $2/8$  or  $0.25$ . Now that we have a grade of membership for each issue in each category, we follow the lead of Hannan, Pólos, and Carroll (2007) and Hsu, Hannan, and Koçak (2009) and define the niche width using the Simpson index of dissimilarity as:

$$1 - \sum_{i=1}^I \mu^2$$

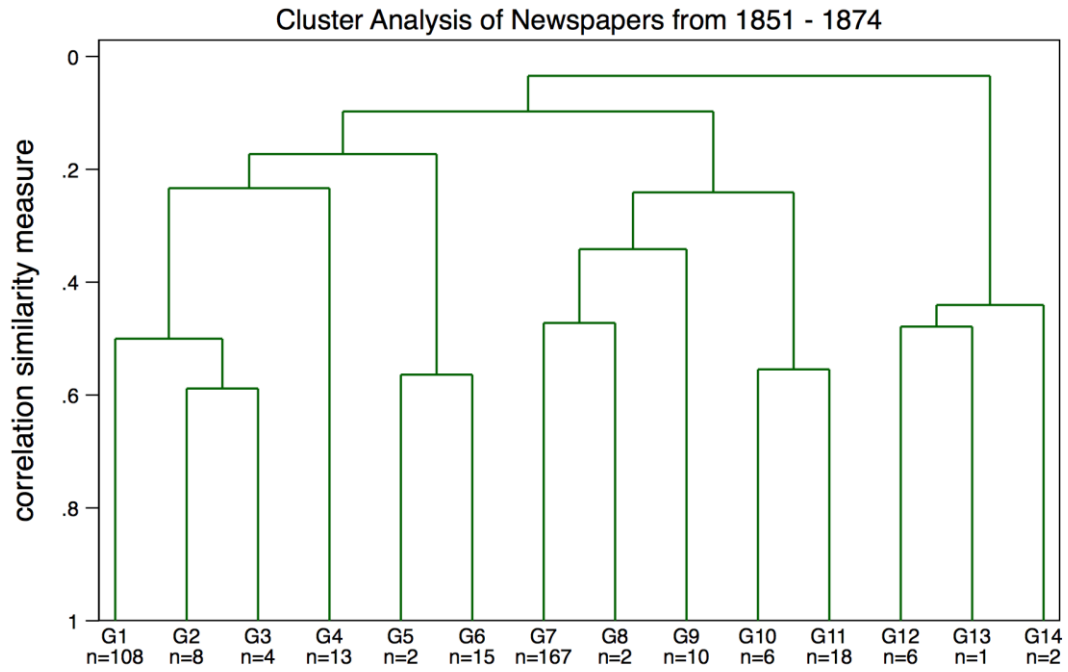
This measure takes values between zero and one with a zero indicating that the entire issue is dedicated to a single category. Increasing values indicate an increase in the diversity of the newspaper and hence its niche width.

To produce the figure I grouped the issues published in the same month together and calculated the average Simpson index and its standard deviation for each month. Both of these are plotted on the figure. I also drew a vertical line at January 1875 because it is the time that I believe when we can see a difference in the graphs. To the left of the line we see that there is significant variability in both the average and the standard deviation of the Simpson index. With time, this variability decreases. To the right of the line we see that the monthly variability is visibly much less than to the left of the line.

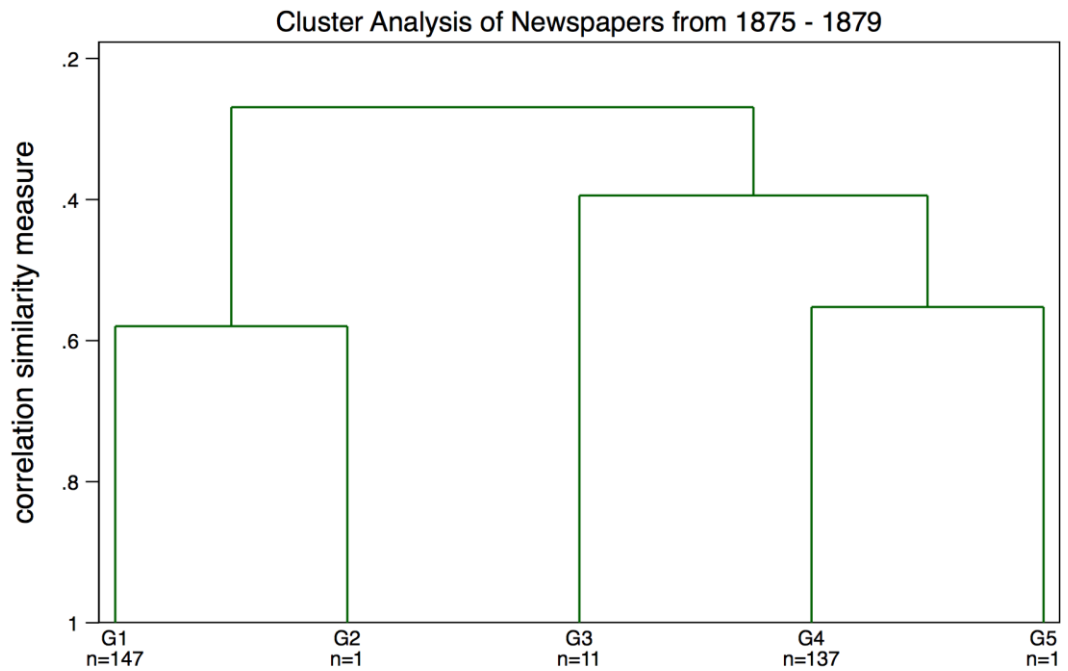


**Figure 5.5 Monthly change in the standard deviation of the Simpson index**

To verify the above observation I produced Figure 5.5. This figure shows the magnitude of the difference between the consecutive monthly standard deviations. The vertical line marks the beginning of 1875 while the horizontal line helps us see that starting in 1875 almost all of the points are less than 0.1 while before 1875 we see that there were much higher differences in consecutive months. These figures suggest that the year 1875 would make a good choice for a cut-off period, therefore I split the data into two time periods, one spanning the years 1851-1874, and consisting of 362 issues, while the second spans the period 1875-1879, and consisting of 297 issues. Figure 5.6 and Figure 5.7 present the dendrograms for the analysis. Both dendrograms only show the clustering taking place at values lower than 0.6.



**Figure 5.6 Dendrogram for the period 1851-1874**



**Figure 5.7 Dendrogram for the period 1875-1879**

A general look at the above figures shows that the clustering effect is much stronger in the latter period. There are 14 groups in the first period at values greater than 0.6 while there are only 5 groups in the second figure. More importantly though, in Figure 5.7 we can see that the smallest three groups have only 13 issues in total in them. This means that almost all of the newspaper issues have become part of one of the two largest groups. In addition, we can see that groups 1 and 4 are almost equal in size. What these numbers tell us is the following: in the first part of the emergence period the newspapers were similar enough to be clustered together at high correlation values but they formed several distinct groups each with a considerable number of members. Around the time of the “re-framing” process, the newspapers managed to form a smaller number of more equally sized coherent groups with very few exceptions.

The next question of interest is the following: did the issues of the same newspaper end up in the same group? In other words, was the content of each newspaper consistent from issue to issue? Table 5.1 and Table 5.2 below show the distribution of the issues of each title among the different groups for both time periods. We can see in Table 5.1 that in general there was a tendency for the issues of the same newspapers to end up in the same group, but this tendency had its exceptions. *Hadiqat al-Akhbar* for example is spread over 10 groups with three of them having a large number of issues. *Al-Nashra al-Osboo'iya* on the other hand had almost all of its issues ending up in the same group. In Table 5.2 we see that most of the issues seemed to be homogeneous except for *al-Jinan*.

	Group														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
<i>Amal alJamiya</i>	5	5	1	0	0	0	0	0	0	0	0	0	0	0	<b>11</b>
<i>Basheer</i>	0	0	0	3	0	0	49	0	0	0	0	0	0	0	<b>52</b>
<i>Hadiqat al-Akhbar</i>	2	0	0	0	1	13	71	2	8	1	18	0	1	1	<b>118</b>
<i>Jinan</i>	39	3	0	0	1	0	15	0	0	1	0	1	0	0	<b>60</b>
<i>MajmooFawaid</i>	4	0	2	0	0	0	0	0	0	0	0	0	0	0	<b>6</b>
<i>Nafeer</i>	0	0	1	0	0	0	5	0	0	0	0	0	0	1	<b>7</b>
<i>Nahla</i>	5	0	0	1	0	0	2	0	0	0	0	0	0	0	<b>8</b>
<i>Najah</i>	1	0	0	0	0	2	19	0	2	0	0	0	0	0	<b>24</b>
<i>Nashra</i>	35	0	0	1	0	0	0	0	0	0	0	0	0	0	<b>36</b>
<i>Sharaka</i>	4	0	0	0	0	0	0	0	0	4	0	0	0	0	<b>8</b>
<i>Tabeeb</i>	12	0	0	0	0	0	0	0	0	0	0	0	0	0	<b>12</b>
<i>Taqadom</i>	0	0	0	2	0	0	6	0	0	0	0	0	0	0	<b>8</b>
<i>Zahra</i>	1	0	0	6	0	0	0	0	0	0	0	5	0	0	<b>12</b>
	<b>108</b>	<b>8</b>	<b>4</b>	<b>13</b>	<b>2</b>	<b>15</b>	<b>167</b>	<b>2</b>	<b>10</b>	<b>6</b>	<b>18</b>	<b>6</b>	<b>1</b>	<b>2</b>	<b>362</b>

**Table 5.1 Distribution of newspaper titles among the groups 1851 – 1874**

	Group					
	1	2	3	4	5	
<i>Basheer</i>	2	0	7	38	0	<b>47</b>
<i>Janna</i>	0	0	0	2	0	<b>2</b>
<i>Jinan</i>	43	1	0	16	0	<b>60</b>
<i>Lisan</i>	0	0	1	25	1	<b>27</b>
<i>Mishkat</i>	4	0	0	0	0	<b>4</b>
<i>Moktataf</i>	43	0	0	0	0	<b>43</b>
<i>Nashra</i>	31	0	0	0	0	<b>31</b>
<i>Tabeeb</i>	24	0	0	0	0	<b>24</b>
<i>Taqadom</i>	0	0	0	3	0	<b>3</b>
<i>Thamarat</i>	0	0	3	53	0	<b>56</b>
	<b>147</b>	<b>1</b>	<b>11</b>	<b>137</b>	<b>1</b>	<b>297</b>

**Table 5.2 Distribution of newspaper titles among the groups 1875 - 1879**

It would be useful though if we could summarize the tables with a single number since this would enable us to compare the two time periods. Again we turn to Simpson's index, but this time instead of studying category spanning we look into cluster spanning. So, for example, if a newspaper had a total of 50 issues in our dataset and these issues were divided into two groups by the cluster analysis as such: thirty in one group and twenty in another, then the grade of membership of the newspaper in the first group is 30/50 and in the second 20/50. Again we use the formula:

$$1 - \sum_{i=1}^I \mu_i^2$$

This measure takes values between zero and one with a zero indicating that all issues of a certain newspaper lie in the same group. Increasing values indicate an increase in the diversity of the newspaper. I calculated these values and the results are presented in Table 5.3 and Table 5.4. The calculated numbers during the period prior to 1875 clearly vary from one newspaper to another with the smallest being 0 and the largest being 0.5970985. The average value indicates that, in general, the newspapers somewhat belonged to few groups. How do we check the statistical significance of this result? In order to do so I ran a simulation experiment in which I pooled the newspaper issues prior to 1875 together and randomly assigned them to 14 groups. The only condition I introduced was to keep the group sizes equal to the sizes of the groups obtained via the cluster analysis. I then calculated the Simpson index for each newspaper and the overall average. I repeated this trail 10,000 times, each time returning the average of the Simpson indexes obtained.

This experiment allowed me to obtain the expected values of the Simpson index if the assignment of the issues were in fact random. The result was that the 98% confidence interval of the returned value was [0.5761655, 0.685548]. Table 5.3 shows a value of 0.3905212, which is considerably out of the obtained interval. What does this mean? Since 0.3905212 is less than the lower limit of the 98% confidence interval it means that our obtained average from the Lebanese industry is in fact not due to randomness. The newspapers, in this early period, were in fact producing issues that, to some extent, seemed to fall into a small number of groups.

Newspaper	Simpson index
<i>Amal alJamiya</i>	0.5785124
<i>Basheer</i>	0.1442308
<i>Hadiqat</i>	0.5970985
<i>Jinan</i>	0.5383333
<i>MajmooFawaid</i>	0.4444444
<i>Nafeer</i>	0.4489796
<i>Nahla</i>	0.53125
<i>Najah</i>	0.3576389
<i>Nashra</i>	0.0540124
<i>Sharaka</i>	.5
<i>Tabeeb</i>	0
<i>Taqadom</i>	0.375
<i>Zahra</i>	0.5694444
Average	0.3905212

**Table 5.3 Simpson index for the newspaper titles before 1875**

Newspaper	Simpson index
<i>Basheer</i>	0.3223178
<i>Janna</i>	0.00
<i>Jinan</i>	0.415
<i>Lisan</i>	0.1399177
<i>Mishkat</i>	0.00
<i>Moktataf</i>	0.00
<i>Nashra</i>	0.00
<i>Tabeeb</i>	0.00
<i>Taqadom</i>	0.00
<i>Thamarat</i>	0.1014031
Average	0.0978638

**Table 5.4 Simpson index for the newspaper titles after 1874**



With regards to the years 1875-1879, Table 5.4 clearly shows that there was a decrease in the number of groups spanned by each newspaper. In fact, the most frequent occurring number in the table is zero. This indicates that, mostly, all issues of each newspaper fell into a single group in our cluster analysis. Again, in order to study the statistical significance of the number obtained I ran a similar simulation to the one described above, but this time on the period of 1875-1879. The 98% confidence interval of the average of the Simpson index I obtained was [0.3946162, 0.5367033]. Had the issues been randomly assigned to the five groups then we would have expected that the average of the Simpson index to lie within that interval. However, the value we obtained was 0.0978638 and this is much less than the lower limit of the 98% confidence interval. This tells us that the issues belonging to the same newspapers during that period mostly fell in the same group and this was not due to randomness. In addition, the number obtained is much less than that obtained for the period 1851 – 1874 so it seems that latter newspaper issues of each title were more likely to be in a smaller number of groups than those published before 1875. Again the question arises as to the statistical significance of this decrease in the average of the Simpson indexes. What if this decrease was due to the fact that we were dividing the issues into a fewer number of groups (5 as opposed to the previous 14). Surely this alone would result in a smaller Simpson index. Once again I used a simulation experiment, however, this time I pooled all issues from 1851 – 1879 together and then randomly assigned them into one of two groups. The only constraint imposed was that the two groups be equal in size to the groups in my original dataset. After assigning the issues to the two groups I re-ran the simulations described above on both groups, but this time I was interested in the

difference between the two averages of the Simpson index. The 98% confidence interval of the difference obtained was [0.0160145, 0.160887]. The difference in the original dataset is  $0.3905212 - 0.0978638 = 0.2926574$ , which is larger than the upper bound with a p-value that is significant at the 0.001 level. The fact that it is larger than the upper limit of the confidence interval suggests that there is a larger than expected decrease in Simpson's index. This means that, in fact, not only did the issues after 1874 form fewer groups, but also the issues relating to the individual newspapers were more likely to end up in the same group than the 1851 – 1879 period. This lends support to hypothesis 10.

The above analysis studies how the issues span the groups found in the cluster analysis. A question remains regarding the spanning of the different categories (politics, literature, economy, ...etc.). If a newspaper title has all its issues end up in a single group in the cluster analysis, then this does not mean that the newspaper's contents dealt with just one category. The group formed in the cluster analysis may be composed of newspapers that span the same categories in the same manner. Above, we found that the newspapers in the period 1875-1879 spanned fewer groups, but do they span fewer categories? To study this question, I again calculated the Simpson index, but this time I based the calculations on the ratio of the content of each category and not on the number of issues in each group. Instead of calculating the index for a newspaper title by dividing the total number of issues in a certain group by the total number of newspapers, I now calculate the index for each individual index by dividing the total amount of space dedicated to politics for example by the total space available in the issue. This way the index would tell us the niche of the issue with regards to the nine categories. After calculating the indexes, I calculated the

average of all the issues in each group of the cluster analysis section. This way I could see if a group was made up of specialist newspapers or more general newspapers. Table 5.5 and Table 5.6 show the results obtained for both periods. If we concentrate on the largest two groups in both tables, we notice that the lowest average in the period 1875-1879 is smaller than the lowest average in the 1851-1874 period. We also notice that the reverse is true for the higher of the two averages in both tables, that is, it is greater in the period 1851 – 1874 than the period 1875-1879. The tables seem to suggest that the issues of the period 1875-1879 were more specialized in their content than the issues of the 1851-1879 period. If this was really the case then not only did newspapers span a smaller number of group clusters in the latter period, but they also spanned a smaller number of categories. This would suggest that with time, newspapers became more alike and more focused in their content.

Cluster group	Number of issues in the group	Average of the Simpson index
1	108	0.378203
2	8	0.6112977
3	4	0.5486086
4	13	0.5460258
5	2	0.5330223
6	15	0.5907483
7	167	0.5635558
8	2	0.7077688
9	10	0.6905794
10	6	0.2221736
11	18	0.6715903
12	6	0.6861979
13	1	0.6715813
14	2	0.331775

**Table 5.5 Average Simpson index for the cluster groups of 1851-1874**

Cluster group	Number of issues in the group	Average of the Simpson index
1	147	0.2687042
2	1	0.7628535
3	137	0.478074
4	1	0.6837125
5	11	0.6876193

**Table 5.6 Average Simpson index for the cluster groups of 1875 - 1879**

Our final analysis concerns the number of clusters that best fit the newspapers published after 1874. Does the data suggest that all, or most, of these issues can be included in one single cluster thereby giving support to the argument that organizations need to project a unified identity (McKendrick and Carroll 2001; McKendrick, Jaffee, Carroll, and Khessina 2003), or do the newspapers cluster into more than one group, hence supporting the finding that more than a single identity is projected by organizations in an emerging industry (King, Clemens, and Fry 2011)? The question of the number of clusters that best fit datasets is extremely tricky. Many measures have been proposed. Everitt, Landau, Leese, and Stahl (2011) conclude that it is best not to depend on a single measure, and this is in fact the advise followed in this paper. Milligan and Cooper (1985) examined 30 rules with the conclusion that while there was no single best rule for all situations, two seemed to work most of the time. These are the Calinski and Harabasz pseudo-F index and the Duda-Hart index. I will use both rules to determine the optimal number of clusters for the given dataset. Table 5.7 shows the results of both measures calculated for 1 up to 15 clusters. With regards to the Calinski and Harabasz measure, the higher the value the better the fit. Unfortunately this measure gives us no information with regards to having a single cluster, but our second measure does. With regards to the Duda-Hart index, the best combination is the one which has as high a value as possible for

the  $Je(2)/Je(1)$  term and as low a value as possible for the pseudo T-squared. The best values for both measures are indicated in bold in Table 5.7.

Our first measure indicates that the number of clusters that best fit the data is two while the second measure indicates that the number is three. If the data is divided into two groups, based on the findings of the first statistic, then the distribution of the issues in the three groups will be as such: one group with 149 issues and a second group with 148 issues. If, on the other hand, the data is divided into 3 groups as the second measure suggests, then the distribution of the issues in the four groups will be as such: one group with 148 issues, a second group with 138 issues, and a third group with 11 issues. In both cases we see the same pattern, which is two large and equal sized groups and a very small number of residual issues that do not seem to fit into either group. This result suggests that no single cluster emerges from the dataset. Instead, the issues seem to form two distinct groups of equal size, thus supporting hypothesis 7B.

Number of clusters	Calinski and Harabasz index	Duda-Hart index	
		$Je(2)/Je(1)$	Pseudo T-squared
1	-	0.3013	684.04
2	<b>684.04</b>	0.8128	33.85
3	380.97	<b>0.9825</b>	<b>2.43</b>
4	255.24	0.9752	3.72
5	195.08	0.8810	18.23
6	164.66	0.1794	663.28
7	361.59	0.5896	6.26
8	313.83	0.5778	70.89
9	295.96	0.9137	4.35
10	265.19	0.6732	62.62
11	325.47	0.9136	1.99
12	297.36	0.7230	17.24
13	284.07	0.6211	2.44
14	263.90	0.7984	26.76
15	282.12	0.6938	8.83

**Table 5.7 Results for finding the optimal number of clusters 1875-1879**

Next we take a closer look at the contents of these two clusters in order to see the extent to which each cluster spans categories. Table 5.8 shows the averages of the ratios for each category in both clusters. The general finding is that we have two types of newspapers. The first has more than half its content dedicated to politics and the second has most of its content dedicated to knowledge. So by the end of the 1870s there were two clear identities: political and scientific newspapers.

	Cluster 1	Cluster 2
Politics	0.1051542	0.6378778
Economics	0.0060576	0.0313665
Social	0.0113373	0.0493177
Literature	0.0639691	0.0511335
Advertisement	0.0023351	0.0211913
Knowledge	0.7814186	0.0801641
Art	0.0003097	0.000906
Sport	0	0.0000168
Other	0.0252021	0.1283396

**Table 5.8 The averages of the categories in the largest two clusters**

I conducted the same analysis on the issues published before 1875. In the first part of this paper we found that issues published before 1875 were less similar to each other than issues published after 1874. Based on this, we would expect that these early issues would have a more difficult time clustering together. In other words, we would expect to see more than just two clusters forming. To test whether this was actually the case I calculated the same statistics for the period 1851 – 1874 that were described above. Table 5.9 shows the results obtained. Here we clearly see that both measures differ significantly with regards to the optimal number of clusters that fit the data. This is not a strange finding in cluster analysis. Although both measures give different numbers of clusters, what concerns us here is that in both cases the statistics

point to the fact that more clusters are formed by these issues than the number formed by issues published between 1875 and 1879<sup>10</sup>. We can see that the scores of the Calinski and Harabasz index for three, four, five, and six groups are close to each other and are much larger than the score for two clusters. In fact, the score for two clusters is the lowest of all scores, clearly indicating that the issues do not group into just two clusters. The Duda-Hart index goes even further and indicates than the issues are best grouped into eight clusters, lending support to hypothesis 9. This finding, taken with the findings from the first cluster analysis in this chapter, shows that issues published before 1875 form a larger number (hypothesis 9) of less coherent clusters (hypothesis 10) than the issues published after 1874.

Number of clusters	Calinski and Harabasz index	Duda-Hart index	
		Je(2)/Je(1)	Pseudo T-squared
1	-	0.9707	10.87
2	10.87	0.6012	232.85
3	<b>125.68</b>	0.7117	59.96
4	122.61	0.6960	57.23
5	120.23	0.8005	50.08
6	117.80	0.9065	18.25
7	104.75	0.5389	5.99
8	91.64	<b>0.9829</b>	<b>2.91</b>
9	80.85	0.6471	2.73
10	72.14	0.8749	16.87
11	69.99	0.4210	30.26
12	68.52	0.7077	6.20
13	63.77	0.3416	19.28
14	60.79	0.5534	8.88
15	57.92	0.8821	22.04

**Table 5.9 Results for finding the optimal number of clusters 1851-1874**

<sup>10</sup> In order to test the robustness of the results I repeated the analysis using *median linkage* instead of *weighted average linkage*. The results indicated that the period 1875-1879 produced fewer clusters than the period 1851-1874.

At this point, it would be interesting to ask whether the newspapers published after 1875 were aware of the fact that two clusters had formed, one political and the other scientific. In other words, do newspapers published after 1875 identify themselves with one of the clusters while those published prior to that date do not make the same effort due to the fact that the cluster formation was not well developed? In order to answer this question we need to look at how the newspapers identified themselves at the time of founding. With regards to the newspapers founded prior to 1875 (1851-1874) a review of the masthead of each newspaper revealed the following: *Hadiqat al-Akhbar* identified itself as a civil, scientific, trading and historic journal, *Nafeer Souriya* did not identify itself in any way, and neither did *al-Basheer* and *al-Jinan*, *al-Taqadom* identified itself as being a general journal, and *al-Zahra* and *al-Najah* identified themselves as weekly publication. We can see from this that the newspapers made no effort to “label” themselves as a certain type of publication. A different picture emerges when we look at the newspapers published after 1874. *Thamarat al-Funun* and *Lisan al-Hal* identify themselves as political and economic newspapers while *al-Moktataf* and *al-Mishkat* identify themselves as scientific publications. As can be seen from the above, the newspapers were aware that there now existed two dominant types of newspapers and they wanted the reader to know to which type each newspaper belonged. This was not the case in the earlier period where the cluster formation process was not as well developed as during the later period.



## 5.4 Multilevel Analysis

The next step in our analysis is to try to explain the change in the category spanning dynamics, and see whether the data support hypothesis 8. In our case, I thought of three events of interest. The first dependent variable, called *Exiting*, would record an event in which a current newspaper issue does not span a category that it used to span in the previous issue. The second dependent variable, called *Entering*, would record the event of spanning a category that it had not spanned in the previous issue. Finally, I created the third dependent variable by using the logical OR operator. This third variable takes on the value of one when either of the previous two dependent variables are one. The reason for the creation of this variable is that while each of the first two dependent variables tracked a dynamic that had a certain direction (either exiting or entering), the third variable has no direction. In other words, this third variable measures the unsystematic dynamic because no matter what direction the newspaper takes, this variable will record a one with any change. An important issue here is the following: when do we consider that a newspaper has spanned a category? In the dataset there are instances in which one newspaper issue does not dedicate any space to literature, but in the next issue the same newspaper dedicates just 5% to it. Should this be considered as a case of spanning? Probably not. So what is the lower limit that should be accepted? Since there are nine categories in total, I took 0.11 (1/9) as the accepted limit. Randomly speaking, an issue would dedicate 0.11 of its space to every single category. Therefore, when a newspaper dedicates more than that amount to a category (which would cause the space dedicated to another category decrease to below this limit) I considered that it spanned that category. These three variables are recorded as a

series of ones and zeros and are hence modeled using logistic regression (Hilbe 2009). In this type of model the logit link function is used to make the relationship between the dependent and independent variables linear. Because the data is of longitudinal nature, and observations belonging to the same newspaper will probably not be independent, I use a random effects model. More specifically, I include a newspaper-specific random-intercept  $\zeta_j$  in order to relax the assumption of independence of observations that pertain to the same newspaper as such (Rabe-Hesketh and Skrondal 2012):

$$\text{logit}\{\Pr(y_{ij} = 1|x_{ij}, \zeta_j)\} = \beta_0 + x\beta + \zeta_j$$

where the subscript  $j$  refers to the newspaper while the subscript  $i$  refers to the individual observation within each panel. I use an identity covariance structure because the best fit model will include only a random-intercept and the identity covariance structure is the only possible structure in such a case (Stata Corporation 2011a).

There are four control variables in the base model. The first is the type of publication (newspaper or magazine). This is to control for the fact that a certain type of publication might be more inclined to span categories and an increase in the number of such a publication in a certain period might be the cause of the increase in the probability of spanning categories. The second variable is the frequency of publication of the issues. Perhaps newspapers with a higher frequency would have a higher chance of spanning categories. The third variable is the number of pages of the issue. More pages might mean a higher probability of category spanning. The final variable is the total number of newspapers active during the time of publication. More titles might also cause an increase in category spanning.

The results of this analysis are shown in Table 5.10. First we start with the dependent variable *Exiting*. We see from the base model that issues from bi-weekly newspapers are statistically more likely to exit categories than other publications with different frequencies. Only the coefficient of bi-weekly is statistically significant. The only other variable which is also statistically significant is the number of active newspapers which has a negative coefficient. I next created a period indicator variable which was set to zero prior to 1875 and to one for issues published in the second period. This variable is used to test whether the issues of one period exited categories more often than the issues of the other period. The results (under Model 1) show that the coefficient is negative but not statistically significant. In Model 2 I replaced the period indicator with a continuous variable that recorded the time since emergence of the population (the date of the very first newspaper). Again the coefficient is negative but not statistically significant. Finally in Model 3 I replaced the time variable with the age of the newspaper. Since the time, both continuous and binary, was not significant, then perhaps the age of the individual newspapers was a better explanatory variable. We see that the coefficient of the variable age is negative and statistically significant. This means that the older the newspaper got, the less likely it was to exit a category. Next I followed the same strategy with regards to the second and third dependent variables. With regards to the variable *Entering*, the three added time variables were negative and statistically not significant. So it seems that while newspapers tended to decrease their category exiting activities with age, they retained their category entering activities.

	Exiting categories				Entering categories				Unsystematic category movement			
	Base Coef.	Model 1 Coef.	Model 2 Coef.	Model 3 Coef.	Base Coef.	Model 1 Coef.	Model 2 Coef.	Model 3 Coef.	Base Coef.	Model 1 Coef.	Model 2 Coef.	Model 3 Coef.
Frequency (base is bi-monthly)												
<i>Bi-weekly</i>	1.701343*	1.562821*	1.604934	1.263632	1.156006	1.095462	1.103122	.9025941	1.952803	1.730702	1.773637	1.402561
<i>Irregular</i>	.3156745	.3280339	-.0951772	-.2296452	-.2684769	-.2993611	-.4575896	-.5349552	-.6924957	-.7481581	-.1602802	-.1151953
<i>Monthly</i>	-.0344205	.0967091	.0308954	-.310834	-.0769286	-.0404292	-.0516032	-.2525044	-.2590298	-.1124904	-.184569	-.643778
<i>Weekly</i>	.9717399	.851204	.7944905	.7178168	.6234923	.5662044	.5382138	.4804657	1.046017	.868288	.7073139	.8029872
<i>Yearly</i>	-17.80174	-18.61402	-17.8258	-18.24017	-1.195504	-1.138448	-1.475437	-1.231831	-2.713508	-2.471926	-4.060184*	-2.671821
Type (base is magazine)												0
<i>newspaper</i>	.34832	.6267084	.3637016	.3420479	.238569	.3086556	.2479894	.2645217	-.1805581	.0558052	-.1231385	-.0826729
Pages	.0047411	.0054431	-.0002743	-.0014483	-.0112505	-.0118452	-.0134791	-.0136428	-.0170725	-.0190181	-.0248971	-.019571
Active newspapers	-.1327501*	-.0773368	-.0676776	-.1304601**	-.1443052**	-.1291721*	-.117652	-.1351247**	-.2261044***	-.1687694*	-.1253563	-.1928283**
Period		-.418046										
Time			-.0555299				-.140387			-.4900159		
Age				-.0068259*				-.0043614			-.1033537*	-.0102867**
Constant	-.4032355	-.805512	.3946393	.1727559	.3297689	.2608948	.6810376	.5688876	2.092221	1.795075	3.736369*	2.420477
Insig2u												
Constant	-2.340204	-3.195689	-2.112843	-2.361705*	-1.494508	-1.566694	-1.440816	-1.590217	-.4350034	-.5545673	-.3505714	-.5528825
AIC	770.0822	768.8263	770.0582	767.3396	775.6501	776.4511	777.2981	775.7795	722.227	721.5815	718.6239	715.9555
BIC	814.7125	817.9196	819.1515	816.4329	820.2804	825.5444	826.3914	824.8728	766.8573	770.6748	767.7173	765.0489
likelihood	-375.0411	-373.4131	-374.0291	-372.6698	-377.8251	-377.2256	-377.649	-376.8898	-351.1135	-349.7908	-348.312	-346.9778
N. of cases	641	641	641	641	641	641	641	641	641	641	641	641

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Table 5.10 Multi-level logistic regression of the three variables Exiting, Entering, and Unsystematic movement**

Finally, with regards to the last dependent variable, we see that both time and age are negative and statistically significant, but the model with age produces the lower AIC and BIC statistics thus indicating that age was the better predictor.

What do these results tell us? We have seen from Figure 5.4 that during the initial years there was a large variation in the average monthly Simpson's index. This variation decreased considerably in later years. The regression results shown above inform us that as newspapers aged, they were less likely to exit categories and less likely to follow a random pattern of entering and exiting. With regards to entering categories, we saw that none of the time variables were statistically significant. Together, these results show that the founding newspapers moved unsystematically between categories and, with age, they ceased exiting categories but continued their entering dynamics, thus lending support to hypothesis 8. In fact, we can see from Figure 5.4 that during the second period (after 1874) the average Simpson index started rising. This can be attributed to the fact that the newspapers stopped the exiting dynamic and continued the entering dynamic. It will be shown in the next chapter that this strategy was actually continued for several years to come thus giving rise to generalists.

## **5.5 Conclusion**

This chapter has used the logical formulations introduced by Hannan, Pólos, and Carroll (2007) in order to take an in-depth look at both group-spanning and category-spanning dynamics of the Lebanese newspapers at the time of their emergence. Unlike previous research that determined the categories which organizations spanned by referring to third parties like critics

(Zuckerman 2000), or to official organizational announcements and publications (Pontikes 2008), this research has extracted the information from the actual product itself. This information was used to study two levels of spanning. The first was the spanning which took place on the level of the categories themselves, while the second was the spanning which took place on the level of the clusters that were formed by the data. In both cases it was found that during later years in the formation period the newspapers became much more consistent in their data coverage in their respective issues, and also became more similar to each other. This result was supported by both cluster analysis and multi-level logistic regression. Another important result was the fact that the issues of the first period (1851-1874) did not produce haphazard content by jumping from category to category. Ever since the inception of the industry, the individual issues of the newspapers showed a tendency to display similar content, a tendency that increased with time. Another important finding was that in the middle of the 1870s there was a change in the content of the newspapers. Interestingly, this change happened around the same time that a change in the goals and motivation of the newspapers took place, as shown chapter 4. The observed change was best described as a process of maturation with age. The older the newspapers became the more focused their content became. Finally, this chapter found that two clearly separate identities had formed by the end of 1870s. One represented political newspapers and the other represented scientific newspapers. It took a significant amount of time for these two identities to be clearly separated. Therefore, this chapter has shown that, at least with regards to that part identity which is determined by the categories spanned

by the organization, industries can, and do, emerge by projecting more than a single identity.

This chapter has used the concept of clusters that was introduced in chapter 3 while discussing the theoretical aspects of this thesis. However, the chapter did not study the effect that spanning clusters has on the mortality of organizations. The reason is that in order to study the effect of cluster spanning we must study a longer period. This is left to the next chapter.

## **6 Chapter Six: Category Dynamics and Cluster Spanning Throughout the Lifetime of the Industry (1851 – 1974)**

### **6.1 Introduction**

The two previous chapters studied the emergence of the newspaper industry in Lebanon. While chapter 4 used qualitative methods, chapter 5 used quantitative methods to study the dynamics of category spanning and cluster formation. In this chapter I will use similar quantitative methods to study the dynamics of category spanning and cluster formation throughout the entire period of study that stretches from 1851 until 1974. In addition, this chapter will also look at the effect that spanning clusters has on organizational mortality rates.

### **6.2 Empirical Setting<sup>11</sup>**

Initially the Lebanese newspapers were in relative harmony with the Ottoman government. The newspapers identified education, not politics, as their primary concern. This, however, changed in the middle of the 1870s when the newspapers started asking for political and economic reform. This, along with the ascension of ‘abd al-Hamid to the throne in 1876, resulted in a crackdown on the newspapers that criticized the government<sup>12</sup>. On the 2<sup>nd</sup> of February 1877 the Sultan declared in a letter that no newspapers should be allowed to question

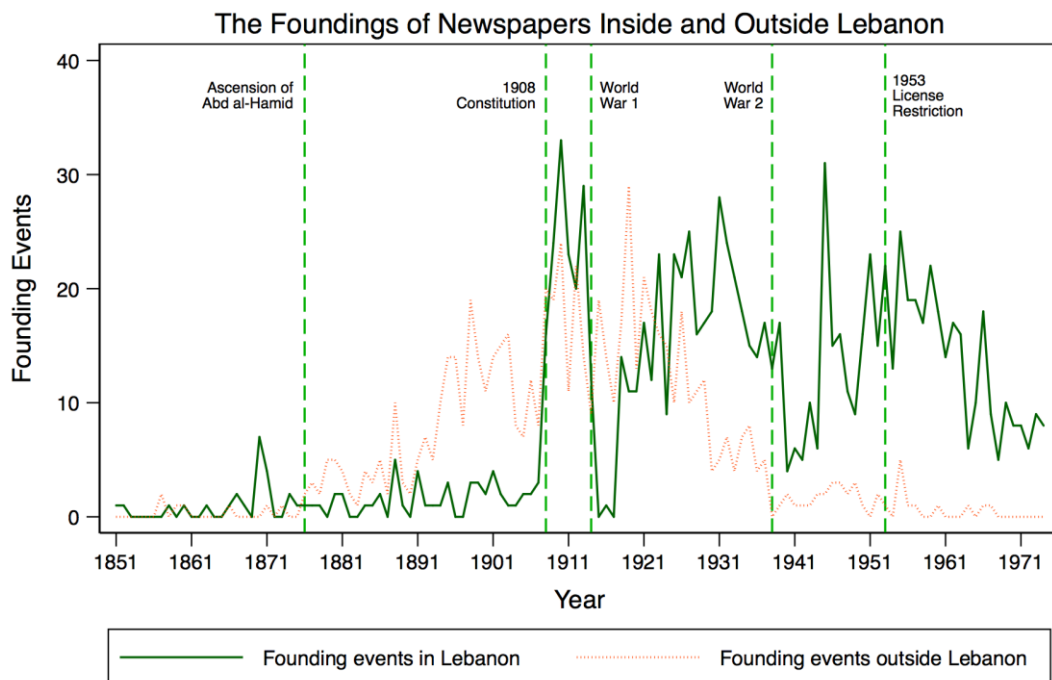
---

<sup>11</sup> This section starts of where the previous “Empirical setting” ended.

<sup>12</sup> Prior to ‘abd al-Hamid’s reign newspapers were suspended in many instances, but the frequency of such events increased after he became the ruler.



his actions (al-Rifa'i 1967). The number of newly founded newspapers, while initially on the rise, starting declining as more and more Lebanese left their country to start newspapers in areas with a little more freedom. This tendency is clearly shown in Figure 6.1 below.



**Figure 6.1 Number of founding events by Lebanese in their country and in other countries**

The primary destination of these Lebanese was Egypt whose newspaper industry benefitted significantly from the influx of experienced writers and editors. It should be noted that up until now the founders of Lebanese newspapers had considered themselves as Ottoman subjects and no one entertained the idea of separation. However, the newspapers started taking a more nationalistic stance as the persecution of the sultan increased. The increase in political pressure would result in the 1908 constitution that, on paper at least, granted the people more freedom. The number of newspapers in Lebanon increased dramatically following the constitution. Things soon reverted back to

the old ways, especially when, on the 6<sup>th</sup> of August 1916, the Ottomans executed 31 opposition members, among them 16 journalists.

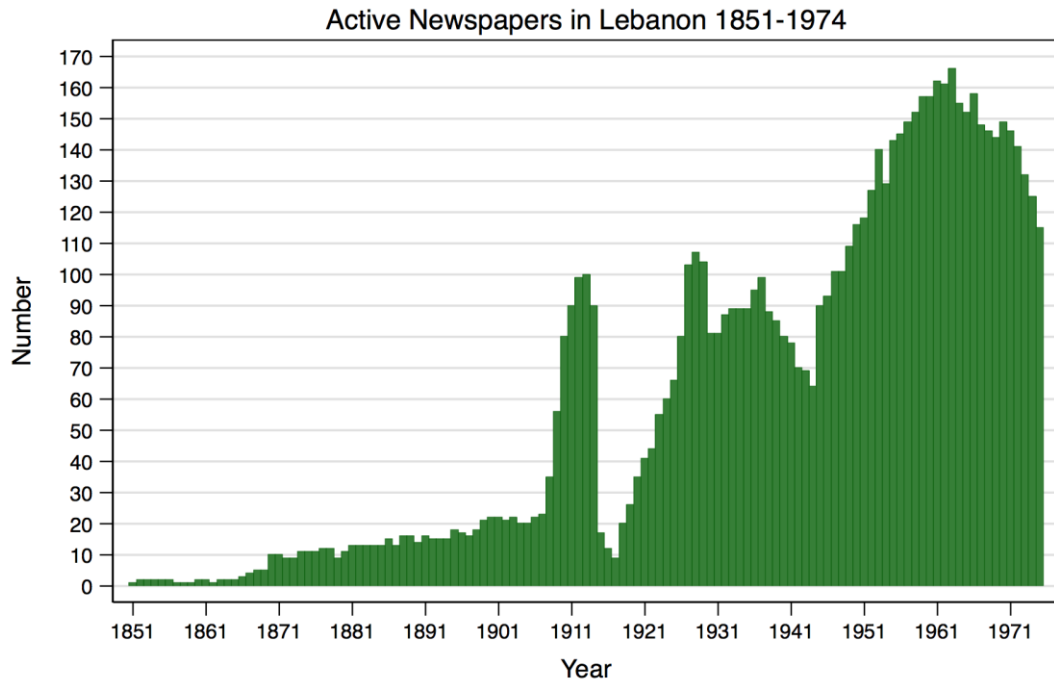
The First World War had a devastating effect on the press in Lebanon. Most newspapers had to stop operations because of lack of paper. Very few newspapers managed to do business as usual. After the war, however, all the major newspapers started operating once again. When the smoke of the First World War settled, the Ottoman rule of the Arabic countries was replaced by Western colonialism. The French were not more welcoming of criticism than the Ottomans, and would eventually persecute newspapers under the dubious title of “administrative reasons”. However, one major step forward for the Arabic press had started to be seen during that time. The spread of schools that began in the middle of the nineteenth century had started to bear its fruit much more than before. Newspapers however still relied on their owners and had their fates tied to them. Unlike in Europe and North America, no newspapers in Lebanon were owned, or even partly owned, by large corporations. This, in large part, explains why many of the newspapers of the time disappeared with the death of their founders. In some cases, the newspapers even suffered when the founder, who also often happened to be the editor, fell sick. Another problem that faced the Arabic press was that the Arabic countries did not manufacture paper, ink, or machines. Hence, the Arab world was highly dependent on Western countries for supplying them with what they needed. Although the percent increase in the number of people able to read newspapers increased substantively, it is important to note that the base of the increase was very small. For example, Ayalon (1995) reports that the famous newspaper *Hadiqat al-Akhbar* had 400 subscribers in 1858, while Mruwi (1961) stated that in the 1920s, if the

circulation of a newspaper could reach 6,000 then it would be considered a huge success. Although this is an increase of 1,400%, the number is still very low.

After independence in 1943, things still did not look better for the press in Lebanon. Although the French had left, their rules stayed, and the Lebanese officials who took the reins of power went to great lengths to insure that these rules were strictly adhered to. This led to a confrontation between the government and the press. The government managed to appease some journalists by issuing them licenses for paper trading. At that time paper was an important commodity that was in short supply. Some journalists became very rich due to trading in paper. The press once again fell on the bad side of the government when in 25 April 1947 they exposed the corruption and blatant falsification that accompanied the election of the members of parliament. Finally, on the 22<sup>nd</sup> of November 1952 the newspaper industry was to receive a huge boost with the issuing on a new rule that gave the press unprecedented freedoms. As a consequence, the number of newspapers and magazines in Lebanon exploded to the extent that there were more than 50 daily newspapers in Beirut alone. The government stepped in and, in April 1953, issued a decree that aimed at limiting the number of newspapers and magazines. This decree stipulated that no new licenses would be given to any publication until there are only 25 political daily newspapers and 20 political magazines. Also, in the case that two current publications ceased their operations simultaneously, then their owner shall be granted a new license for a single publication. Once the above figures would be reached, then one license will be given to the owner of a current publication that ceases operation. This study ends in the end of December 1974, while the Lebanese civil war started in 1975.

### **6.3 Data Collection**

The data collection process was explained in detail in chapter 3. The process was made up of two stages. In the first I collected data pertaining to the life history of all newspapers that existed in the time period 1851 – 1974. By the end of this stage I had collected the following information: there is a total of 1759 newspapers, with 1056 being published in Lebanon and 703 outside Lebanon. Around 81% of the publications in Lebanon were published in Beirut while around 35% of publications outside Lebanon were published in Egypt. With regards to the publications in Lebanon, there are 59 cases of missing founding dates and 427 cases of missing disbanding dates. Out of these 427 cases, there are 164 cases of censoring before the end of the study. This means that while I did not find the ending dates of these publications, I managed to secure dates at which I knew that they were still in circulation. Finally, there are 119 cases of censoring due to the study ending at the end of 1974. Figure 6.2 below shows the number of active newspapers during each year. It is important to note that newspapers with a missing disbanding date are not included in the figure. Newspapers with censored dates were included up to their date of censoring. So in other words censored dates were treated like disbanding dates in the figure. The figure clearly shows that the 1908 constitution was followed by a burst of newspaper founding events and how the effects of both world wars, especially the first one, was significantly negative on organizational density.



**Figure 6.2 Active newspapers 1851-1974**

One important point should be mentioned here. In his study about the political process in Lebanon, Hudson (1985) mentions that in 1961 there were a total of 255 Arabic publications in Lebanon. The figure above clearly shows that the number is less than 170 in my data. In fact the number is 162. Why this discrepancy? The reason is that Hudson (1985) used as his source the List of Registered Publications which was published by the Lebanese Press Syndicate in 1962. As the name suggests, this is the list of the registered publications, and it is very important to note that not all registered publications saw the light. For example, an individual with the name of Majid Toufiq al-Hamwi secured a license for a non-political weekly newspaper on 31/8/1955. No issue of this newspaper was published. On 26/3/1956 the owner changed the name of the publication. Then on 29/2/1960 he changed the name yet again and still not a single issue of this newspaper was published. Again on 9/8/1962 the owner yet again changed

the name of the newspaper and after that we finally know that the newspaper was published. In my database, I recorded a single founding event for the last title of which we know that at least a single issue was published.

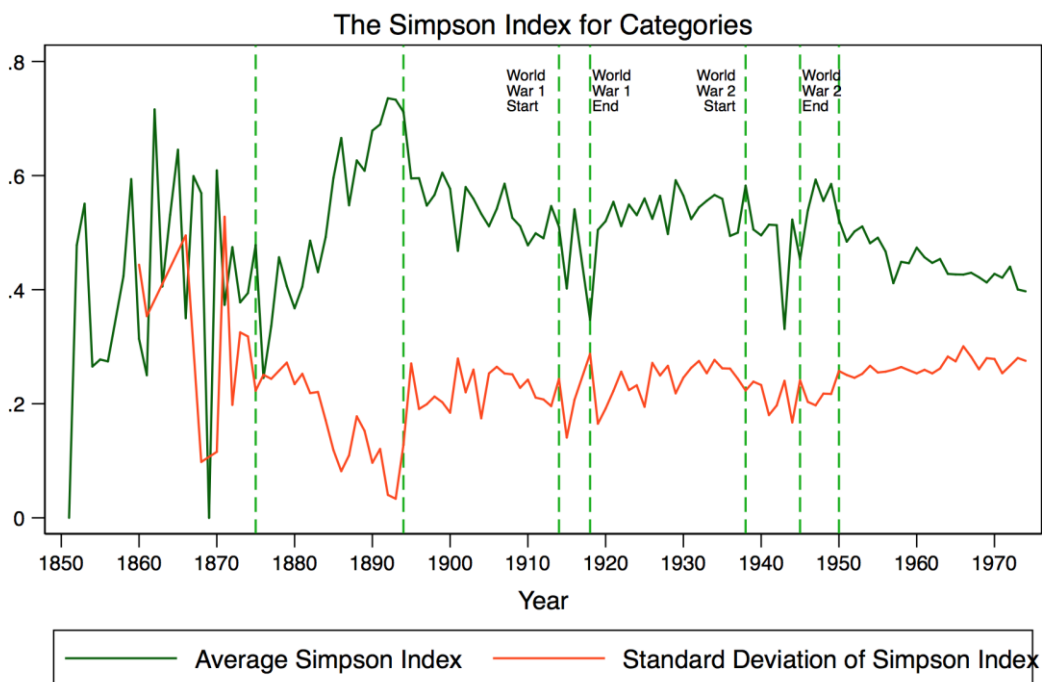
The second stage of the data collection process involved coding the contents of the newspapers. I identified the same categories as those described in chapter 3. In order to minimize variances in the topics due to different events occurring at different dates I decided to choose certain dates for the issues to be analyzed. These dates were chosen based on the anniversary of the creation of whatever government was ruling Lebanon. The Mutasarifiya of Mount Lebanon was created on 9 June 1861, the French mandate started on 1 September 1920, and finally the date of Lebanese independence is 22 November 1943. This means that for newspapers published before 1920 I selected the issue published on 9 June. For newspapers published during the French mandate, the date chosen was 1 September and finally, for issues published after independence the date 22 November 1943 was chosen. In the cases where the issue at the selected date was not found in the collections, I analyzed the issue that was closest to the date. The final dataset consisted of 2,478 entries, with belonged to 293 different newspapers.

## **6.4 Cluster Analysis**

The first step in the analysis is to study how the newspaper issues spanned the different categories over the duration of the study period. To do that I once again use Simpson's index. Increasing values indicate an increase in the diversity of the newspaper and hence its niche width. I calculated this index for every single newspaper issue in the dataset. I then grouped the issues according

to their year of publication and calculated the average of the index and its standard deviation for every year. The results are presented in Figure 6.3 below. The vertical lines in the graph have been inserted at the points in which the graph shows that there was some change in the dynamics. Prior to 1875 (the first vertical line) the graph clearly shows that both the average and the standard deviation of Simpsons index were shifting with no clear pattern. This is partly due to the fact that before 1865 there was a maximum of 2 issues per year, but in chapter 5 we showed that the variation in fact was not merely due to the small number of observations. After 1875 we see a completely different dynamic in that the average of Simpsons index kept increasing while the standard deviation was decreasing. The only explanation to this dynamic is that most of the newspapers were broadening their coverage. In other words, this period witnessed the rise of generalists. The fact that the standard deviation kept decreasing tells us that most of the newspapers followed this strategy. However, this dynamic was reversed after 1894 (the second line) where we can see that the average of Simpsons index started decreasing while the standard deviation increased. A plausible explanation to this is that some newspapers were becoming more specialized, concentrating on one, or at least on a few, categories. However, this was not a universal strategy as illustrated by the standard deviation. Therefore, we would expect that this period saw the rise of some specialists with some newspapers remaining generalists. The third and fourth lines in the graph show the start and end of the First World War. We can see that the category spanning becomes more erratic during this period and represents a discontinuity in the general dynamics of Simpsons index. The same can be seen during the Second World War (between the fifth and sixth lines). In between the

two world wars we see that neither the average of Simpsons index neither the standard deviation exhibit any changes. The seventh (and last) line represents the point when the average of Simpsons index starts decreasing while the standard deviation starts increasing. The dynamic during this period is similar to the one described above between 1894 and the start of the First World War in 1914. However, we can see that the slopes of the lines become more consistent and have less variation in the very last period. Again, the only explanation with regards to the last period is that some newspapers were more specialized while generalists continued to exist. This explanation is consistent with resource partitioning theory of Carroll (1985).



**Figure 6.3 Dynamics of the category Simpsons index over time**

In the previous chapter, it was shown that during the emergence period of 1851-1879, with time, the newspapers tended to span fewer categories. However, it was also shown that different issues relating to different newspapers



tended to span less groups, or clusters. That chapter distinguished between category spanning and group spanning. The logic behind this was that it was possible for certain newspapers that spanned the exact same categories to belong to the same group. These groups were determined using Cluster Analysis (Everitt, Landau, Leese, and Stahl 2011). Do the dynamics of category spanning and cluster spanning follow one another as found during the period of emergence? Or do they follow different paths and have different effects on the industry as a whole? To answer these questions, I once again used Cluster Analysis to determine what groups were formed. I used Figure 6.3 in order to determine how to split up the period of study into separate time periods. The figure indicates that certain dates proved to be critical points at which the industry dynamics changed. I do not study here the period prior to 1880 because it had been studied extensively in the previous chapter and also because the selection method of one issue per year leads to few observations during that period due to the fact that the number of newspapers that were active was relatively low. Therefore, the cluster analysis in this chapter will study the following periods separately: 1880 – 1894, 1895 – 1913, 1919 – 1937, 1946 – 1949, and 1950 – 1974. It should be noted that the years of the two world wars are not included in the analysis. During the First World War the Lebanese newspaper industry was almost completely eliminated because of the shortage in material and also because of the hardships that resulted from the famine. During the Second World War all newspapers had to radically decrease their number of pages and most of the content dealt with the ongoing war. It can be seen in Figure 6.3 above that these periods represent a discontinuity in the general trend of the two graphs.

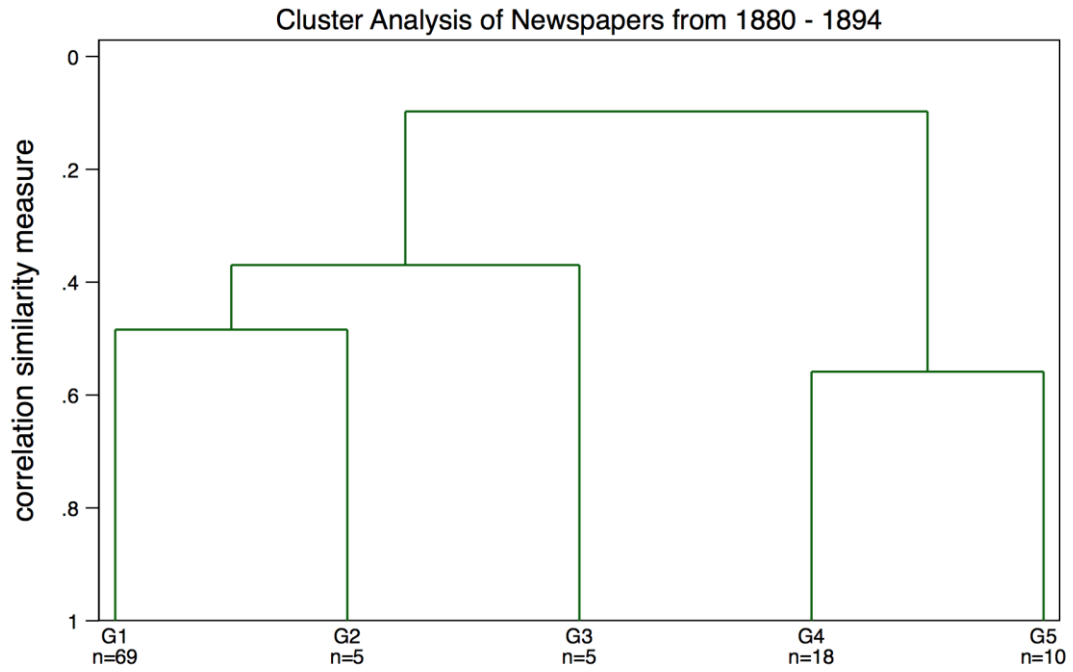


Figure 6.4 Cluster analysis 1880-1894 at the 0.6 level

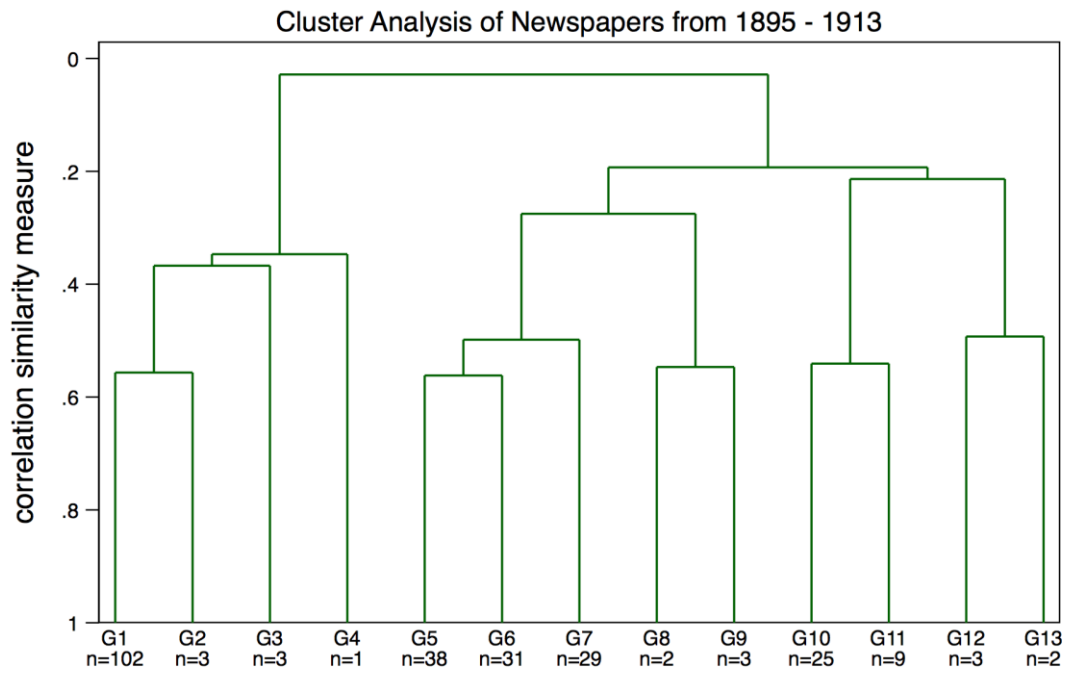
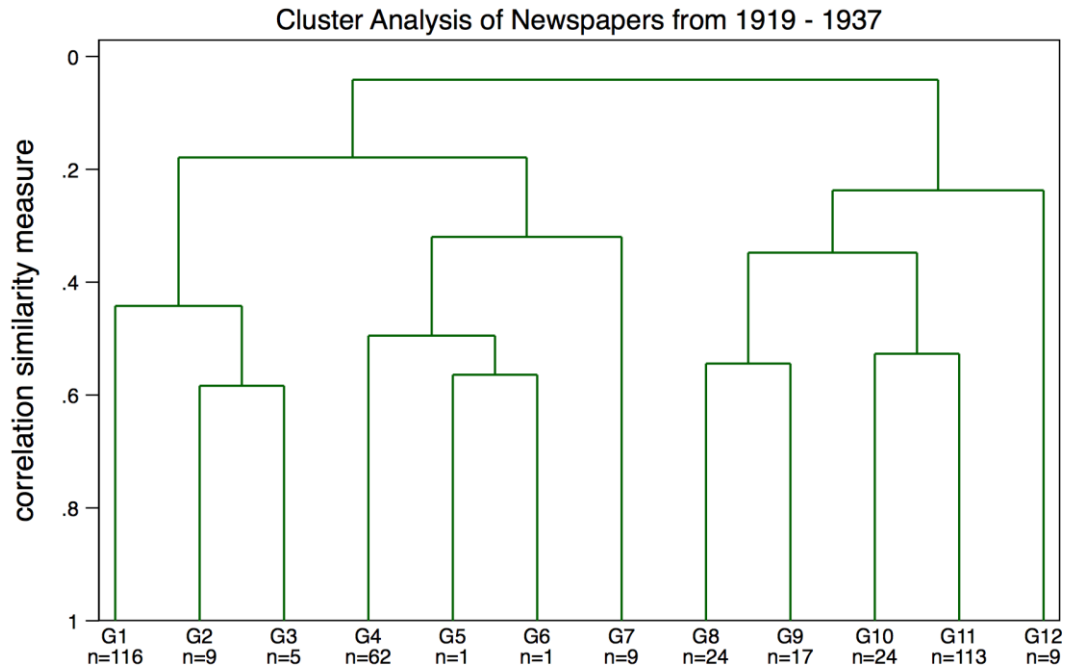
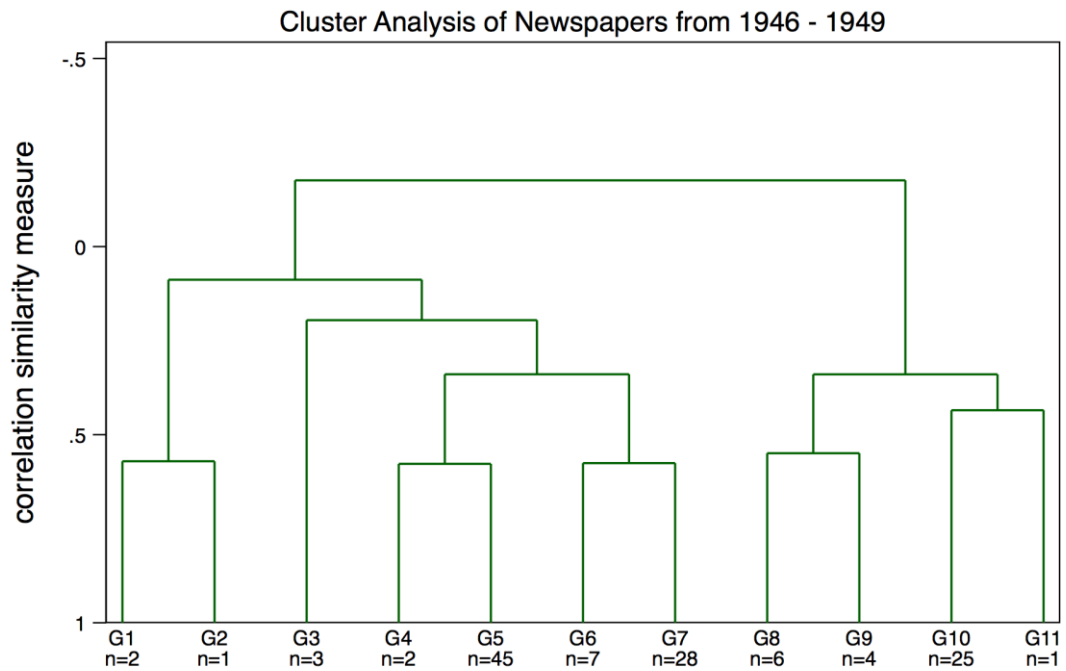


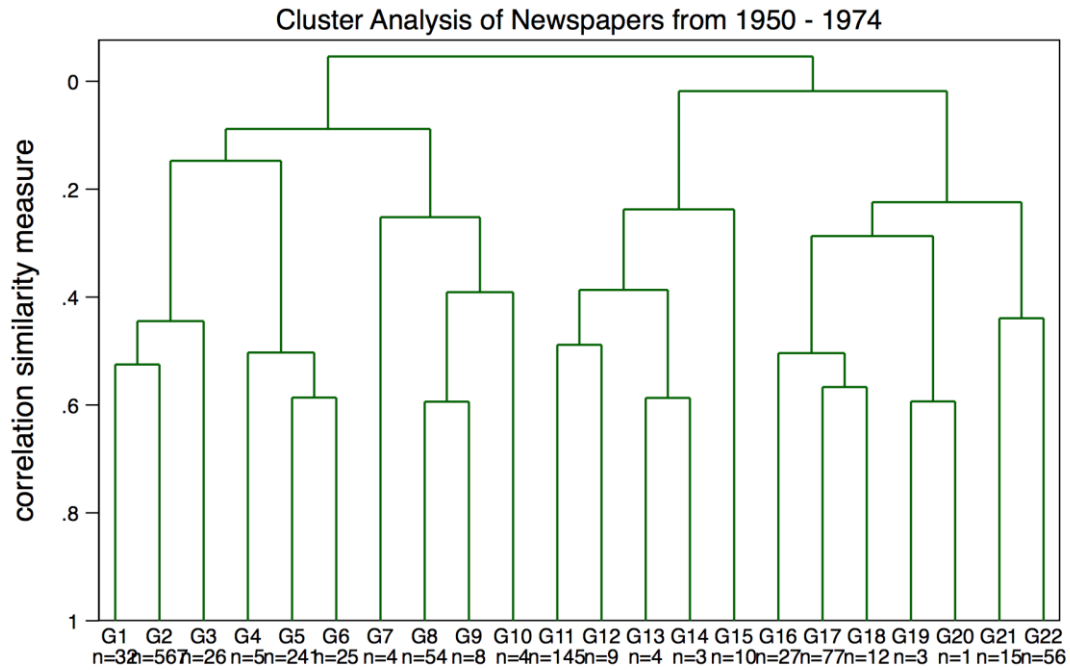
Figure 6.5 Cluster analysis 1895-1913 at the 0.6 level



**Figure 6.6 Cluster analysis 1919-1937 at the 0.6 level**



**Figure 6.7 Cluster analysis 1946-1949 at the 0.6 level**



**Figure 6.8 Cluster analysis 1950-1974 at the 0.6 level**

I used cluster analysis on each of the periods separately. Once again, I used as a measure of proximity the Pearson correlation measure and the weighted-average linkage as the linkage method. Figures 6.4 to 6.8 show the dendrogram for each cluster analysis. All of the figures show the cluster formation at the 0.6 level. It appears that issues were most alike during the period 1875 - 1894 and least alike during the period 1950 - 1974 (The results are summarized in Table 6.1 below along with other results which will be described soon). This is in fact consistent with our discussion about the average category spanning Simpson index. In the first period, there was a rise in generalists and so the newspaper population was relatively homogeneous. In the last period, there was a rise in specialists and so the population became more heterogeneous. I complemented this analysis with an attempt to determine the

number of clusters that best fit the dataset. I once again use the Calinski and Harabasz pseudo-F index and the Duda-Hart index. The results are shown in Table 6.1. Both measures agree on that the last period saw the largest number of clusters. However, the pseudo-F index indicates that the number of clusters was stable in the first four periods while the Duda-Hart index indicates that the number of clusters increased slightly in the second period before decreasing again. I checked the pseudo-F index for all cluster sizes in the second period, and while the best score was obtained by two clusters, the index for three, four, and five clusters was also very high, unlike the results for the first period in which all results were low except for the two cluster option. So it seems that there is some evidence that the number of clusters increased during the second period, but after that it dropped down back to two clusters. It seems that there was an attempt at diversification but that this attempt was discontinued.

Above, I calculated the average Simpson index of the category spanning dynamics. Here, I perform the same calculation, but this time with respect to the clusters that the cluster analysis has identified. Previously I asked if issues spanned multiple categories, now I am concerned with studying if different issues of the same newspaper belonged to different clusters. This way we can reflect on the tendency of newspapers to produce consistent products using a single number. To do this, I divided the analysis into the five periods specified above. Next I ran cluster analysis on all the issues within each period. The issues were assigned to one of the groups that had already been formed at the 0.6 level. I then calculated the Simpson index as follows: if a newspaper had a total of 20 issues in our dataset during a certain period and these issues were divided into two groups by the cluster analysis as such: fifteen in one group and five in

another, then the grade of membership of the newspaper in the first group is 15/20 and in the second 5/20. The Simpson index was created using the same formula used in the category spanning case above. I excluded the newspapers that had only a single issue because this way it would appear as if all the issues of this newspaper belonged to a single group and this would artificially deflate Simpson's index. Finally, I calculated the average of the Simpson index for all newspapers within a single period. The results are presented in Table 6.1.

Period	No. of clusters at 0.6 level	Calinski and Harabasz	Duda-Hart	Average Simpson index	Simulated experiments
1880 - 1894	5	2	3	0.1942309	[0.3373556, 0.4880385]
1895 - 1913	13	2	5	0.3942486	[0.5549536, 0.6485382]
1919 - 1937	12	2	3	0.2367467	[0.4827211, 0.565532]
1946 - 1949	11	2	2	0.2079248	[0.4332108, 0.5420752]
1950 - 1974	22	6	8	0.1825324	[0.5861586, 0.6299704]

**Table 6.1 Group analysis results for all periods**

What does the average Simpson index for each period tell us? Here, there are two aspects to the numbers. The first aspect is how high, or how low, each number is. In statistical terms, are the numbers simply random, or is there a dynamic involved? If the average Simpson index of a period was very low, much lower than we would expect under randomness, then that would mean that newspapers issues tended to end up in the same cluster more than expected. If the number was very high then that would indicate the opposite, i.e. that there was more difference among the issues of the same newspapers than expected and that it was highly unlikely for randomness to be the cause. In order to look into this aspect, I conducted simulation experiments on each period individually. In these experiments, I created the same number of groups as those observed in the data and randomly assigned all the issues to each group. The only condition I

introduced was to keep the group sizes equal to the sizes of the groups obtained via the cluster analysis. I repeated this 10,000 times, each time returning the average Simpson index. I then display the 98% confidence interval of the returned statistic (the upper 99<sup>th</sup> percentile and the lower 1<sup>st</sup> percentile). Finally, I inspected whether the observed average Simpson index was within the confidence interval obtained or not. The results are shown in the 6<sup>th</sup> column of Table 6.1. First, we see that in all periods the observed average Simpson index does not lie within the 98% confidence interval of the simulated statistic. This means that in all periods newspapers tended to produce issues that ended up in a few number of clusters. However, this does not tell us anything about this tendency relative to other periods. In other words, did this tendency change over time, and if so, how?

The second aspect of the numbers presented above is concerned with the change observed from period to period. In other words, after looking at the statistical significance of the numbers within each cluster, I now look at the statistical significance of the change observed between categories. It is important to take into account the difference in the number of clusters when we compare two periods. The calculation of the average Simpson index above is highly dependent on the number of clusters at the 0.6 level. For example, the first period had 5 clusters at the 0.6 level with an average Simpson value of 0.1942309. The period directly after that had 13 clusters with an average Simpson value of 0.3942486. How do we know that the increase in the average Simpson index is not just merely an artifact of dividing the issues into a larger number of groups? If we double the number of groups then we would expect that the average Simpson index would increase because there was a much higher

chance for issues that belong to the same newspaper to end up in different groups. In other words, what is the statistical significance of the results? To answer that question I again used random simulation experiments, but this time I pooled the issues from two consecutive periods together because we are mostly interested in the change between two consecutive periods. Lets take as an example the first two periods. There are five groups in the first period, with a total of 107 issues, and 13 in the second with a total of 238 issues (this excludes the newspapers which had only one issue as mentioned above). I pooled all these issues together and then randomly assigned them into one of the eighteen groups (five plus thirteen). Just like before, the only condition I introduced was to keep the group sizes equal to the sizes of the groups obtained via the cluster analysis. I then calculated the difference between the average Simpson index of both groups. I repeated this trial 10,000 times, each time returning the difference of the two average Simpson indexes obtained. This experiment allowed me to obtain the expected values of the difference of the average Simpson indexes if the assignment of the issues was in fact random. This time however, we are also interested in the direction of the change. Previously, we used a two-tailed test by looking at both ends of the confidence interval because we were interested to see if the observed Simpson index was lower or higher than what we would randomly get. Now however, we want to see if the increase in the Simpson index from period 1 to period 2 is significant, and if the decrease in the Simpson index between all other periods is significant. Therefore I use a one-tailed test here. After I retrieve the 10,000 calculated differences I am able to calculate the z-score of the observed difference and from that I calculate the p-value. The



results of these simulation experiments (the significance levels) are presented in Table 6.2.

Periods compared	Observed difference	Simulated experiments p-value
Period1 and Period2	0.2000177	p = 0.17828797
Period2 and Period3	-0.1575019	p = 0.00066941
Period3 and Period4	-0.0288219	p = 0.42527792
Period4 and Period5	-0.0253924	p = 0.03518125

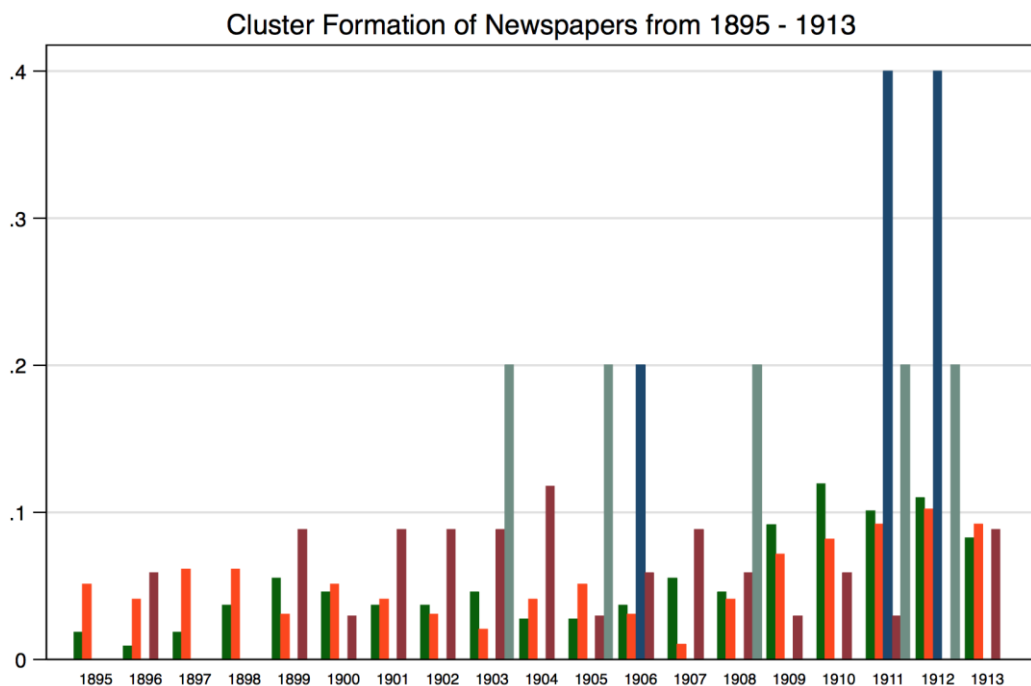
**Table 6.2 Between group analysis**

With regards to the difference in the Simpson index for the first two periods, the results show that the increase in the average Simpson index between period 1 and period 2 was not statistically significant. We then see that there is a decrease in the Simpson index from when we move from period 2 to period 3, and that this decrease is statistically significant. In periods 3 and 4 there is also a decrease but this decrease was not statistically significant. Finally, we see that in period 5 the drop in the Simpson index was statistically significant at the 0.05 level. We see that the newspapers in the fifth period were the least likely to span clusters, even though the number of clusters in that period was much more than any other period.

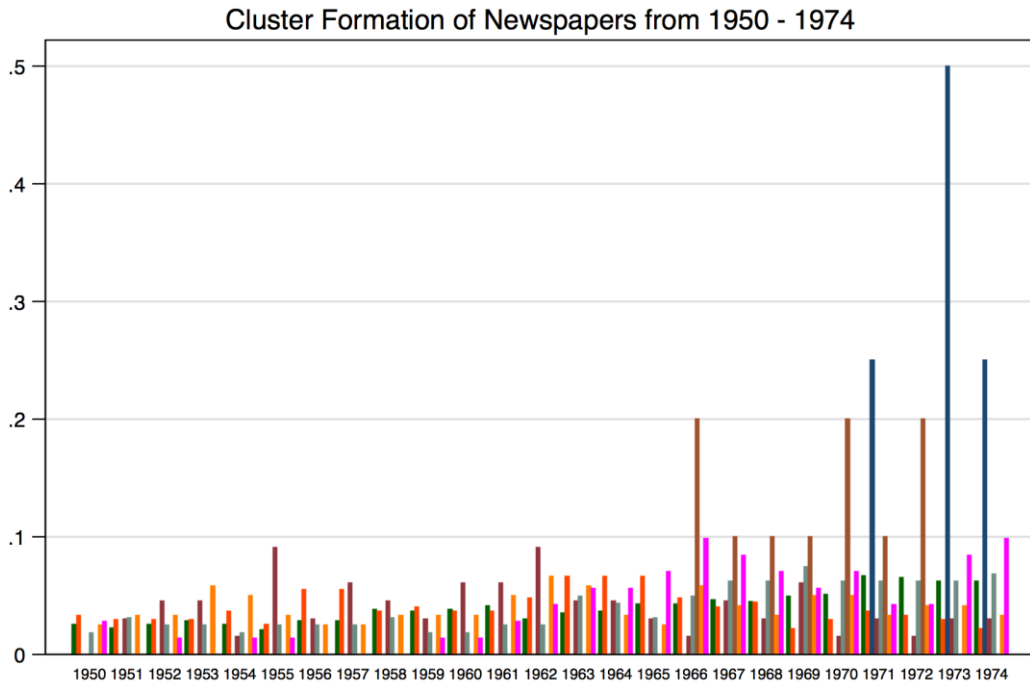
An interesting dynamic that can be discerned from the above analysis of category spanning and group spanning is the following: during the first period, 1880-1894, there was a rise in generalists that tended to produce homogeneous issues and thus the issues created few clusters. The issues of the same newspapers tended to end up in the same cluster. During the second period, 1895 – 1913, the newspapers started specializing with respect to category spanning and this caused a drop in the average Simpson index, but at the same time the newspapers altered their cluster spanning activities in that issues that belonged to the same newspaper were more likely to end up in different clusters when compared to the previous period. This diversification stage seems to have

failed because during the third period, 1919 – 1937, we saw that the best fit number of clusters decreased, but interestingly during this same period the Simpson index for cluster spanning decreased even though the number of clusters at the 0.6 level stayed almost the same. Table 6.2 above shows that the issues of the same newspaper in period 3 were significantly less likely to span different clusters when compared to those in period 2. Finally, during the last period under study (1950 – 1974) the industry seemed to have successfully matured into one where both specialists and generalists co-exist given the fact that most of these papers seem to have radically decreased their cluster spanning activities. Although the fifth period has the most number of clusters, it has the smallest average Simpson index indicating that the newspapers were highly unlikely to produce issues which ended up in different clusters. From the above, we see that there were two periods in which the industry saw a rise in specialists: the second period and the fifth period. However, individual newspapers in the second period spanned several clusters while those in the fifth period spanned very few clusters. In order to visually compare the diversification of the industry during these two periods I created Figure 6.9 and Figure 6.10 that show the clusters active during each year of every period. Different colors refer to different clusters. These figures display the best fitting number of clusters based on the calculations of the Duda-Hart index. Figure 6.9 contains five clusters while Figure 6.10 contains eight clusters. The y-axis of both figures shows the fraction of the total number of issues that belonged to that specific cluster in the entire period. We can see that in Figure 6.9 three of the clusters were present in every year and most of their bars are below 0.1 indicating that the issues that belonged to these clusters were spread over the

time axis evenly. However, the other two clusters had considerable gaps between consecutive appearances and by the end of the period we see that that these two clusters do not appear to be present anymore. This figure seems to confirm our previous assumption that the diversification process was a failure, an assumption that becomes even more visible when we compare this figure to Figure 6.10. This second figure reveals that all clusters survived for several consecutive years and that their distribution was more uniform than their counterparts in the period 1895 – 1914. Some of the clusters were active since the start of the period while others where only formed during later periods. By the end of the period we see that seven of the eight clusters were still active.



**Figure 6.9 Clusters active during every year 1895 – 1913**



**Figure 6.10 Clusters active during every year 1950 - 1974**

## 6.5 Multilevel Analysis

The resource partitioning theory of Carroll (1985) has found support in a wide range of industries (Barnett and Carroll 1987; Carroll and Swaminathan 2000; Lomi 1995). In the following I will attempt to tackle the same issue from a different perspective, that of categories. The above analysis has provided us some clues and hints as to the category spanning dynamics of the newspaper industry in Lebanon. I will now use regression models in order to try to determine the precise dynamics of category spanning during different periods of time.

Here I use two of the dependent variables that we used in the chapter 5: *Entering* and *Exiting*. *Entering* refers to a newspaper covering a category that it previously did not cover. *Exiting* refers to a newspaper exiting a category that it previously covered. The models are analyzed using the same multi-level random-

coefficient logistic regression techniques that we used in the previous chapter. When modeling both independent variables I include four control variables in the base model. The first is the type of publication (newspaper or magazine). This is to control for the fact that a certain type of publication might be more inclined to span categories and an increase in the number of such a publication in a certain period might be the cause of the increase in the probability of spanning categories. The second variable is the frequency of publication of the issues. Perhaps with a higher frequency newspapers would have a higher chance of spanning categories. The third variable is the number of pages of the issue. More pages might mean a higher probability of category spanning. The final variable is the total number of newspapers active during the time of publication. Table 6.3 below shows the results of all the fitted models for the dependent variable *Exiting*.

The Base model shows that the only statistically significant variables are the type of publication and the number of active newspapers. It appears that newspapers were more likely to exit a category than magazines. In Chapter 3 it was mentioned that magazines either had a table of contents, or had consecutive page numberings. Therefore, it seems that this continuity led the magazines to, more or less, not change their categories. We will see that the same finding was true for the dependent variable *Entering*. I keep all the variables in the following models to control for their effects. In longitudinal models, the variable of most interest usually measures the passage of time. In my dataset, each newspaper title was observed at date intervals separated by one year, I therefore included the *age* variable next in the model.

	Base	Model 1	Model 2	Model 3	Model 4	Model 5
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
Type of publication(base is magazine)						
<i>newspaper</i>	.735431**	.7326585**	.734268**	.707846**	.7069444**	.7067959**
Frequency (base is annually)						
<i>bi-monthly</i>	15.61916	15.93012	16.11841	15.73031	15.36317	15.72402
<i>bi-weekly</i>	15.16555	15.47381	15.66643	15.35039	14.97883	15.3405
<i>daily</i>	15.05302	15.35718	15.54655	15.09597	14.714	15.08046
<i>monthly</i>	15.34935	15.659	15.84904	15.37348	15.00718	15.36669
<i>quarterly</i>	15.84566	16.15157	16.33954	15.85014	15.48202	15.84632
<i>weekly</i>	15.39707	15.70383	15.89539	15.50745	15.13791	15.49874
Pages	-.0025454	-.0025729	-.0025518	-.0021924	-.0021881	-.0022565
Active newspapers	-.0030563*	-.0031477*	-.0031668*	-.0031698	-.0026865	-.0021827
Age		.0007073	-.0006008		.0003885	-.0006568
Age <sup>2</sup>			.0000208			
Time				.0450584**	.0467281**	.040696*
Time <sup>2</sup>				-.0002696**	-.0002872**	-.0002384*
Founded in period 5					-.0804666	
1.p5#c.age					.0132772	
Published in period 5						-.1881095
1.period5#c.age						.0022142
Constant	-15.58667	-15.88957	-16.07017	-17.31903	-17.00145	-17.20566
lns1_1_1						
Constant	-.3216664*	-.319796*	-.3214768*	-.3247867*	-.3163999*	-.3170394*
AIC	2715.429	2717.402	2719.374	2709.991	2715.484	2715.514
BIC	2777.865	2785.514	2793.162	2783.779	2806.3	2806.33
likelihood	-1346.715	-1346.701	-1346.687	-1341.996	-1341.742	-1341.757
N. of cases	2156	2156	2156	2156	2156	2156

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Table 6.3 Multilevel logistic regression for the dependent variable Exiting**

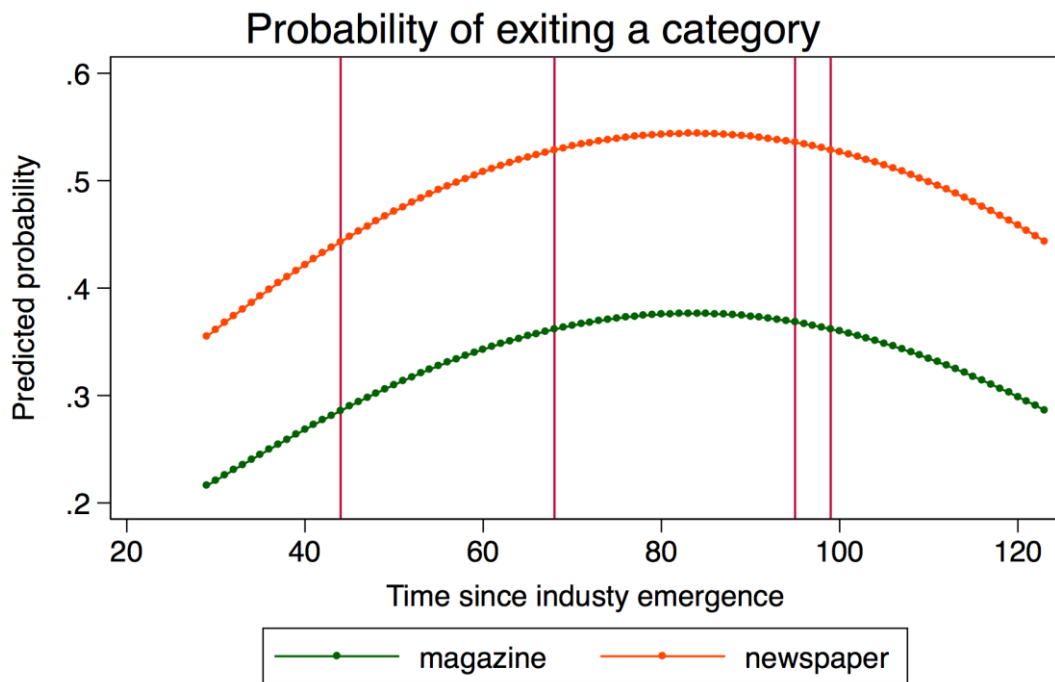
We can see in Model 1 that with age, the tendency of newspapers to exit categories increases slightly, but this result is not statistically significant. In a model not shown in the table, I added a random coefficient of *age*, but a likelihood ratio test revealed that this addition was not statistically significant so I retained the random intercept model in all later models. This result means that while the starting position was different for each newspaper, the changes that occurred over time were not statistically different. So in other words the newspapers, with respect to the independent variable exiting, followed similar dynamics. In Model 2 I added the square of *age* in order to see if perhaps the effect of age was linear or not. The result was also not statistically significant. The next variable of interest is labeled *time* and it tracks the time, in years, since the emergence of the newspaper industry in Lebanon (1/1/1851). I include both the variable and its square in Model 3. Here we see that both are statistically significant and they describe an inverted-U curvilinear.

Previously, I identified five different periods in the history of the industry as shown by Figure 6.3 (p1 – p5). I created five dummy variables to reflect these five different periods. The variable *p1* was set to one for newspapers founded after 1874 but before 1895, *p2* was set to newspapers founded after 1894 up to 1918 and so on. I next included each of these variables separately in a model (not shown in the table) but none of them was statistically significant. Model 4 shows the result of adding just the variable *p5* which indicates which issues belonged to newspapers which were founded in the fifth period and its interaction with age. The reason for this inclusion is that the fifth period was the period in which Figure 6.3 indicates that resource partitioning took place. Therefore, I was most interested in seeing if the newspapers founded in this period were founded as

specialized newspapers or if they were founded as generalists but adapted by becoming specialized by exiting categories. A negative coefficient for  $p5$ , as indicated by the results, means that the newspapers founded in the fifth period were actually less likely to exit categories, but this result is statistically insignificant. The positive coefficient of the interaction with age indicates that as the newspapers became older, they started to be more likely to exit categories, but again, the result is statistically insignificant. This led me to suspect that the newspapers founded in this period were founded mostly as specialists because specialists would not have a tendency to exit categories since they already span very few categories. This suspicion will be tested below when I model the average Simpson index. Finally, since being founded in the fifth period did not have a significant effect, I decided to test whether being published in the fifth period had a significant effect. In other words, while  $p5$  indicated if the issue belonged to a newspaper which was founded in period five, the variable  $period5$  indicates whether the issue itself was published in period 5. Again, I also include an interaction of this variable and  $age$ . If, for example,  $period5$  had a statistically significant positive coefficient while, as we saw,  $p5$  did not have a significant coefficient, then this would indicate that issues for all newspapers, those founded in period five or not, were more likely to exit categories. This would indicate that the decrease in the Simpson index observed in this period would have been partly due to newspapers, both old and new, specializing their content. The results, which are shown under the column Model 5, show that both of the variables are not statistically significant. Out of all these model the AIC, BIC, and, where the two models are nested in each other, the likelihood ratio test indicate



that Model 3 is the best fitting one. The model shows that time has an inverted-U effect on the dependent variable *Exiting*.



**Figure 6.11 Effect of type and time on the dependent variable contraction**

Figure 6.11 shows a visual representation of the results from the best-fit model. We clearly see the inverted-U effect of time on the probability of exiting a category. I also display the trajectories for both types of publications. The vertical lines indicate the different periods with the first line indicating the start of period 2 and the last line indicating the start of period 5. Period 1 had the lowest probability of exiting a category. This was the period in which the Simpson index increased steadily thus indicating that there was a rise in generalists, and generalists do not tend to exit categories. We see that the probability starts decreasing at the start of period 5, which we suspect is the period of the rise of the specialists. In that period we would expect both generalists and specialists to survive side by side. Again the graph provides

further evidence that newspapers founded in that period were founded as specialists instead of being generalists that became specialists by exiting categories. Had generalists exited categories during that period then we would not have seen a decrease in the probability of exiting. This issue will be further inspected later.

I next use the same reasoning as above to model the dependent variable *Entering*. The results obtained for all the models are shown in Table 6.4 below. The Base model includes the same control variables as those discussed above. Once again, newspapers had a statistically significant higher probability of entering new categories while other control variables are not significant. In Model 1 I added the variable *age* and in Model 2 I added the square of that variable. Both of these additions were not statistically significant. I also tried adding a random-coefficient for the variable *age* (not shown in the table) but once again the result was not statistically significant. In Model 3 I added the variable *time* and its square. Both of them turned out to be statistically significant with an inverted-U shape. I retain both of the variables in the next models. I next created indicator variables similar to the ones created when modeling contraction. I included the period indicator variables that indicated the period of founding individually (not shown in Table 6.4). None of them turned out to be significant.

	Base	Model 1	Model 2	Model 3	Model 4	Model 5
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
Type of publication(base is magazine)						
<i>newspaper</i>	.7455189**	.7440392**	.7415202**	.7026582**	.7019905**	.6554004**
Frequency (base is annually)						
<i>bi-monthly</i>	17.55893	17.50127	17.56229	17.47129	17.60463	18.60238
<i>bi-weekly</i>	17.30768	17.24807	17.30044	17.29239	17.45493	18.61769
<i>daily</i>	16.75349	16.6915	16.74958	16.66435	16.80013	17.86233
<i>monthly</i>	16.87719	16.81836	16.87501	16.73908	16.88442	17.89709
<i>quarterly</i>	17.33292	17.27206	17.3327	17.1773	17.31878	18.32101
<i>weekly</i>	17.14665	17.0862	17.14035	17.09219	17.23419	18.21761
pages	-.001255	-.0012704	-.0013046	-.000863	-.0008891	-.0008681
Active newspapers	-.0028629	-.0029158	-.0028793	-.0013382	-.0013778	-.0010841
Age		.0003997	.0027393		.001448	.00023
Age <sup>2</sup>			-.0000373			
Time				.0316617*	.0292357	.0272052
Time <sup>2</sup>				-.0002122*	-.0002031*	-.0001891
Founded in period 1					-.1568388	
1.p1#c.age					.0017087	
Published in period 1						-.9065902
1.period1#c.age						.0908417*
Constant	-17.41859	-17.35587	-17.43022	-18.45057	-18.45285	-19.42116
lns1_1_1						
Constant	-.2763829*	-.2758371*	-.2734871*	-.2805438*	-.2752747*	-.2899568*
AIC	2679.277	2681.269	2683.181	2677.154	2682.942	2676.313
BIC	2741.714	2749.381	2756.969	2750.942	2773.758	2767.129
likelihood	-1328.639	-1328.635	-1328.591	-1325.577	-1325.471	-1322.156
N. of cases	2156	2156	2156	2156	2156	2156

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Table 6.4 Multilevel logistic regression for the dependent variable Entering**

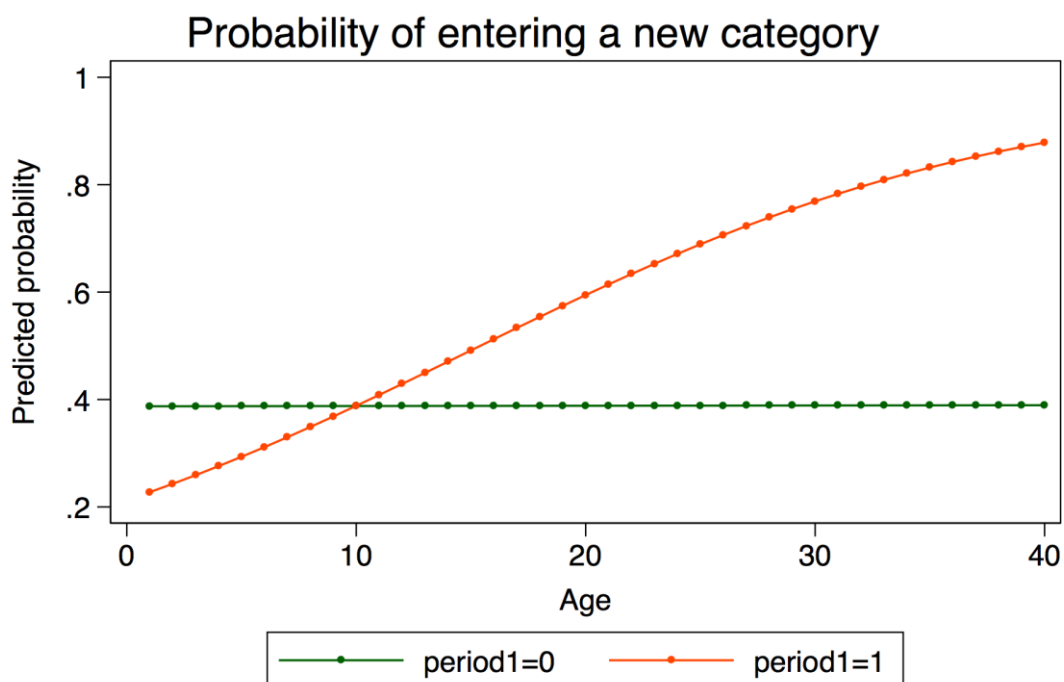
Next, Model 4 includes the indicator variable of being founded in the first period and its interaction with age. The logic here is the same as the logic used in modeling *Exiting* when the indicator of the fifth period was introduced. Figure 6.3 shows that the average of the Simpson index increased during the first period. Therefore, by including the indicator variable *p1* in the model, I was interested in seeing if newspapers founded during this period tended to enter new categories more than other newspapers. If so, this would mean that the newspapers founded in this period were not founded as generalists but became so with time. Because the coefficient of the variable *p1* and the coefficient of its interaction with *age* is not statistically significant, then we can not deduce that newspapers founded in period 1 were more likely than others to expand to new categories. This leads us to suspect that the newspapers founded in this period were founded as generalists and hence had no need of entering new categories since they already spanned most of them. This suspicion will be tested below when we model the variable Simpsons index. Finally, in Model 5 I included the variable *period1* and its interaction with *age*. This variable indicates whether a particular issue was published in the first period or in another period. We can see that while the main term is not statistically significant, the interaction term is statistically significant and so we keep both of them in the model (Hosmer and Lemeshow 2004). The negative coefficient of *period1* indicates that issues published in this period were less likely to span new categories, but this is only half the story, and not the more important half. The statistically significant positive coefficient of the interaction with age indicates that the older the newspaper the more likely it was to span new categories. This indicates that the

increase in the average Simpson index that was seen in Figure 6.3 was partially due to old newspapers, presumably those founded before period 1, expanding their contents, or in other words, old newspapers becoming generalists. This means that some of the generalists in this period were old newspapers that gradually spanned more and more categories. If we take both the main effect and the interaction together, we see that for very young newspapers published in the first period, the probability of entering new categories was small. This is inline with what we said about the newspapers which were founded during this period being founded as generalists and thus not needing to expand their contents while older newspapers had to expand their categories in order to become generalists. This will be tested below when I model the Simpson index. In order to get a clearer picture of whether it was age or time that had an effect, I ran a model in which I interacted the variable *period1* with time (the results are not shown) but the result was not statistically significant. Therefore, I deduce that it was the age of the newspapers that mattered with respect to newspapers published during the first period, and not the fact whether they were published at the beginning, in the middle, or at the end of the first period.

Now compare this with the results obtained when we modeled the dependent variable *Exiting*. In that case, we had found that neither being founded in period five nor being published in period 5 was statistically significant. This led us to suspect that the specialized newspapers of the period were founded as specialized instead of becoming specialized. We also saw that in later periods in the industry history exiting a category became less likely, presumably because generalists did not exit categories while specialists were founded as such and therefore had no need to exit. In modeling *Entering*, we saw

that issues that belonged to old newspapers during period 1 were more likely to expand their contents. I also suspect that newspapers founded in period 1 were founded as generalists.

While the time since the emergence of the industry had an inverted-U effect on the probability of exiting a category, the same is not true when it comes to entering categories. Here we see that a period indicator that indicates whether the issue was published during period 1 suffices. In addition, we see that there is an effect of age when the issue is published in the first period. Figure 6.12 below shows a visual representation of the effects of these variables on the probability of entering a new category.



**Figure 6.12 Effect of the period indicator and age on the dependent variable expansion**

The graph holds the variable time constant along with the other control variables. We can clearly see from the figure that the probability of entering a new category increases significantly with age for the issues that were published

in the first period. For very young newspapers, the probability of entering a new category is actually less than that of newspapers not published in the first period. However, as we discussed above, the older newspapers were more likely to enter new categories and thus developed into generalist organizations.

The findings of all the models discussed above indicates that the rise of the generalists in the first period was actually due to incumbents expanding their contents while the rise of specialists in the fifth period was not due to newspapers gradually exiting categories, but due to newly founded newspapers being founded as specialists while the older newspapers retained their generalists status. To test these findings, I decided to model the Simpson index of each issue in order to see what variables had an effect on it. This time however, the nature of the independent variable put significant restraints on the type of model that I could use. The Simpson index, as described above, can take on any value from 0 to 1. As discussed in detail in chapter 3, in such a case there are three possibilities: the Stata *betafit* command which assumes that the proportions follow a beta distribution (Buis, Cox, and Jenkins 2012), tobit regression, and the fractional logit model. It was explained in chapter 3 why the fractional logit model is the most suited for our case. This model is implemented in Stata by using the *glm* command along with the *family(binomial) link(logit)* options. Since our dataset is longitudinal in nature we can use the *xtgee* command instead of the *glm* command with the same options to account for the, probable, lack of independence in the observations within panels. The *xtgee* command runs a generalized estimating equation (GEE) model instead of a random-effect model (Hardin and Hilbe 2013). Using this model the equation becomes:

$$\text{logit}\{\Pr(y_{ij} = 1|x_{ij})\} = \beta_0^{PA} + x\beta^{PA}$$

where the subscripts PA stand for population-averaged. Note that the above equation no longer includes a subject-specific random effect. This means that the coefficients have a different meaning than in the case of random-effects models. In the random-effects models discussed above the coefficients were subject specific in that they referred to the change in the probability for a certain newspaper if the variable changed for that newspaper. In GEE, the coefficients are population-average estimators not subject-specific estimators. This means that the coefficients refer to the change in the average newspaper.

The results of modeling the Simpson index using the above method are shown in Table 6.5. The Base model includes the same control variables used in the models above. This time however we see that all of the control variables except for the number of pages turned out to be statistically significant. On average, newspapers had a higher Simpson index than magazines. Also bi-monthly, bi-weekly, daily, monthly, quarterly and weekly publications had a larger Simpson index than annual publications. Apparently annual publications were specialized in their content. We can also see that the higher the number of active publications the higher the Simpson index, but the coefficient, although statistically significant, is very small. In Model 1 I included the variable *time* and its square value. Only the squared term is significant, but I retain both variables in the model because it makes more sense to include lower degree terms in case of polynomial functions (Rabe-Hesketh and Skrondal 2012). The implied inverted-U relationship indicates the rise and then fall of the Simpson index that we observe in Figure 6.3. I next decided to include period indicators that indicate



during which period a newspaper was founded because I expect that dividing time into sections would prove instructive.

In Model 2 I include the indicator variable  $p5$ , which indicates whether a newspaper was founded in the fifth period or not. The negative coefficient indicates that issues that belong to newspapers that were founded in the fifth period had a smaller Simpson index, which meant that they were more specialized, thus supporting hypothesis 2. This seems to support our suspicion that we made above when we modeled the dependent variable *Exiting*, but we will leave the final discussion until we find the best-fit model. Notice that once we include the period indicator both time and its square become statistically non-significant. I therefore remove them from the next models. I next decided to include an interaction between  $p5$  and  $age$  in order to see if there was a tendency of these specialized newspapers to become more specialized or to expand their content (Model 3). We can see that both are statistically non-significant. However, I retain both in the next models because, as we will see, they will become statistically significant as we add more variables. In Model 4 I included the indicator variable  $period5$ , which as I explained previously, indicates whether the issue was published in the fifth period or not, and its interaction with  $age$ . This inclusion was made in order to see if newspapers that were founded prior to the fifth period had a tendency to increase or decrease their Simpson index. The result is statistically insignificant.

	Base	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
Type (base is magazine)										
<i>newspaper</i>	.2582813*	.2111539	.1634689	.2026042	.1986515	.1733945	.1556654	.1546833	.1683818	.2362185*
Frequency (base is annually)										
<i>bi-monthly</i>	2.398554**	2.365229**	2.224912**	2.239428**	2.259014**	2.190469**	2.228332**	2.229078**	2.258607**	2.251704**
<i>bi-weekly</i>	2.365615**	2.376614**	2.22789**	2.193638**	2.186689**	2.308343**	2.314247**	2.32337**	2.319087***	2.261557**
<i>daily</i>	2.2954**	2.274955**	2.09685**	2.098054**	2.078888**	2.101554**	2.130759**	2.135013**	2.161243**	2.125385**
<i>irregular</i>	1.146675	1.515016	1.301541	.9141122	.9607181	.9042177	.9935463	.9849303	.9872023	.9032072
<i>monthly</i>	2.068611**	2.016769*	1.861627*	1.896314**	1.891504**	1.874846**	1.900053**	1.903582**	1.928717**	1.890484**
<i>quarterly</i>	1.814805*	1.758468*	1.584529*	1.645607*	1.65087*	1.646715*	1.669061*	1.674428*	1.702716*	1.649295*
<i>weekly</i>	2.366546**	2.364498**	2.215749**	2.200588**	2.201825**	2.157054**	2.178784**	2.181758**	2.218458**	2.15215**
pages	-.0007752	-.0005654	-.0005988	-.0007715	-.0007394	-.0007012	-.000718	-.0007104	-.0007153	-.0008396
Active newspapers	.0020592**	.0026012**	.002789**	.0030136***	.0036916***	.0032696***	.003646***	.0036689***	.0037024***	.0036425***
Time		.0141493	.0126971					-.0002174	-.000515	-.0004417
Time <sup>2</sup>		-.0000958*	-.0000726							
Age				.0007354	.0022854	-.001486	-.0021081	-.0020717	-.0018159	.0004293
Founded in period 5			-.4918426***	-.5491953***	-.5305194***	-.6020847***	-.60581***	-.6006249***	-.5908505***	-.5712934***
1.p5#c.age				.0085722	.0092956	.0110378	.0117603	.0119336	.0119594*	.0151442*
Published in period 5					-.0727163					
1.period5#c.age					-.0018141					
Published in period 1						-.6153965**	-.7027714***	-.697448***	-.1882335***	-.1864043***
1.period1#c.age						.0492197***	.0533356***	.0531265***	.0329934***	.0347112***
Founded in period 1							.5038926*	.5139123	.39835	.4135586*
1.p1#c.age							-.0013565			
1.p1#c.time								-.0008225		
1.period1#c.time									.042004**	.0403956**
0.p5#2.type#c.age										-.0038356
1.p5#2.type#c.age										-.0164652*
Constant	-2.506911**	-2.951024**	-2.678733**	-2.220648**	-2.264819**	-2.181942**	-2.254204**	-2.24267**	-2.256**	-2.240156**
N. of cases	2478	2478	2478	2478	2478	2478	2478	2478	2478	2478

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Table 6.5 GEE regression of the Simpson index**

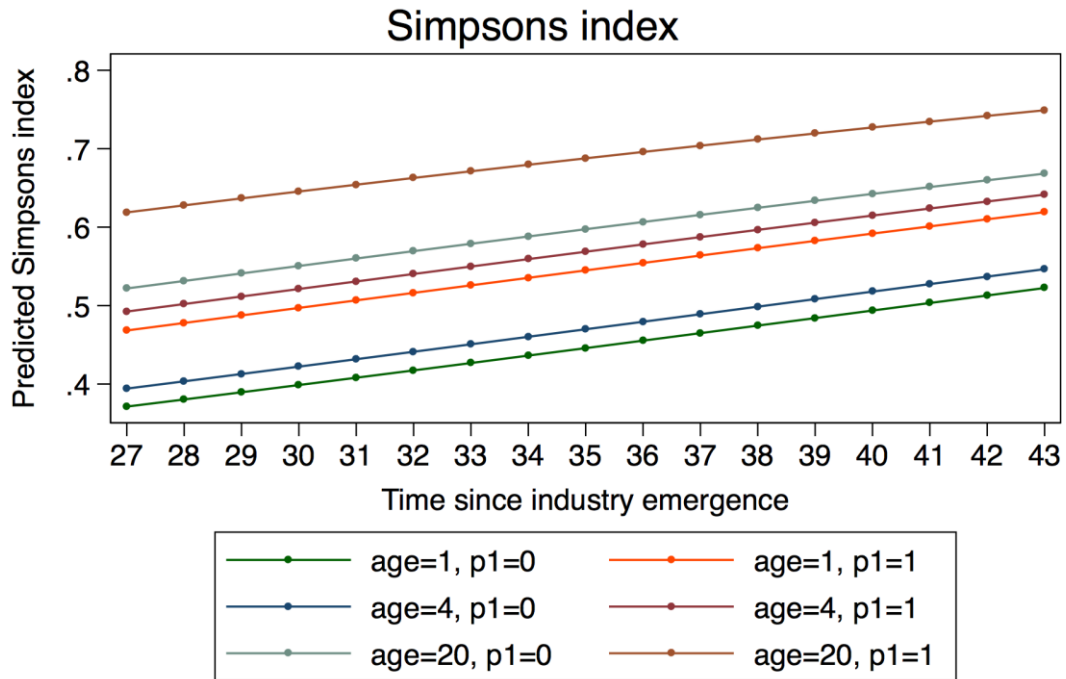
In Model 5 I decided to test the effects of being published in the first period because, as we saw above, we suspect that old newspapers that were founded prior to period 1 tended to become generalists while newspapers founded in this period were founded as generalists. Again I interact the variable *period1* with *age*. The results show that both coefficients are statistically significant with the main term being negative and the interaction being positive. The positive interaction supports the findings of modeling the *Entering* dependent variable because it shows that older newspapers published in that period tended to have a higher Simpson index, and hence were generalists. The negative main term suggests that newspapers started off with a low Simpson index. Again this supports our suspicion that the newspapers became generalists over time, thus supporting hypothesis 1.

In Model 6 I include the indicator variable *p1*, which indicates whether the issue belongs to a newspaper that was founded in the first period, and its interaction with *age*. Only the main term is statistically significant with a positive coefficient meaning that newspapers founded in the first period tended to have a higher Simpson index. When I modeled the variable *Entering*, I mentioned that I had a suspicion that the newspapers founded in this period were founded as generalists since they did not significantly expand their contents but the Simpson index of the period was high. Now we see that the suspicion was in fact correct. Newspapers founded in the first period tended to have a higher Simpson index and hence were founded as generalists. In Model 7 I interacted the indicator variable *p1* with *time* in order to see if it made a difference being founded in the beginning or the end of the period, but the result is statistically insignificant. In Model 8 I included the interaction between the indicator variable

*period1* and time. The reason I did that is that I wanted to see if time had an effect on the Simpson index like age. The result is both positive and statistically significant indicating that by the end of period 1, even the young newspapers were having a high Simpson index. This shows that both age and time had a positive effect on the Simpson index. This was a period in which both old newspapers and newer ones were having a high Simpson index.

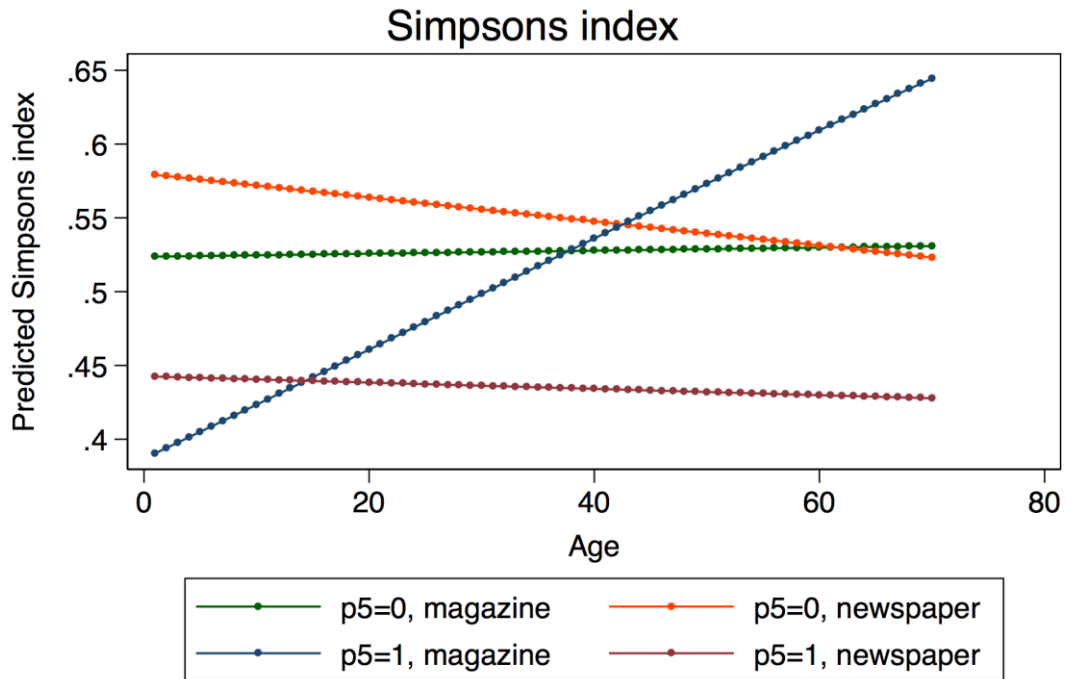
At this point, notice that the interaction between *p5* and *age* has become significant. According to Model 8, publications founded in the fifth period tended to span more categories with age, and this finding raised a flag for me. Hence I decided to investigate the matter further. Perhaps a possible explanation would be that newspapers and magazines behaved differently at this point. To look into this I included in Model 9 an interaction between the three variables *age*, *p5* and *type*. This allowed me to investigate whether all publications founded in the fifth period behaved in the same with age or whether newspapers and magazines behaved differently. Both of the new coefficients are negative, but only one is statistically significant, which is the one that belongs to publications founded in the fifth period. This result tells us that there was no statistically significant difference between newspapers and magazines that were founded prior to the fifth period with regards to the variable age, but that for those publications founded in the fifth period the two types behaved differently in that newspapers tended to have the smaller Simpson index, and hence tended to remain more specialized as they aged. For newspapers founded in period 5 we can see that positive interaction between *age* and *p5* is almost cancelled out by the negative interaction between *p5*, *age* and *type*. This is best understood by visualizing the results, which is what I do later.

The above results provide strong support to the resource partitioning theory, but it goes one step further in showing how the resource partitioning dynamics takes place. This is done by looking through the lens of categories. In the period when the generalists take over, we have seen that old newspapers tend to increase their expansion activities and hence their Simpson index increases (hypothesis 1). We also saw that newly founded newspapers also tended to be founded as generalists and so had no need to expand into new categories. This, however, is not the same dynamic that we see when the specialists start to appear. In that era, we saw that old newspapers tended to maintain their generalist status while newly founded newspapers were founded as specialists with no need to exit categories since they already did not span too many of those (hypothesis 2). I illustrate these results in Figure 6.13 and Figure 6.14 below. In the first figure, we see the effects that the variables age, time and period of publication have on Simpson's index during the first period. The graph clearly shows that both publications that were founded during the first period or prior to that period increased their Simpson index with both time and age. The graph also shows that the publications founded during the first period had a higher Simpson index and hence were more generalists than their precursors. However, this was not true for publications that were founded before the first period by a very long time and hence had a large age. We can see that a publication which was founded before the first period and was twenty years old when the first period began had a higher Simpson index than a publication which was founded at the very start of the first period.



**Figure 6.13 Effect of time, age and period of founding during the first period**

Figure 6.14 on the other hand compares effects of age and type of publication on the Simpson index between the newspapers founded in the fifth period and those founded prior to that. There are two things to note in this graph. As discussed previously, all types of publications founded in the fifth period started with a much lower Simpson index than those founded prior to the fifth period. What is significant here is that the period of founding is the significant variable and not the period of publication, as was the case for the first period. Second, we see that magazines founded in the fifth period develop a different trajectory in that they tend to become generalists as they age. This, as we mentioned, is contrary to our expectations.



**Figure 6.14 Effect of age, period of founding and publication type during the fifth period**

Why is it that magazines founded in the fifth period tended to become generalists as they aged? In order to investigate this, I re-examined the records that belonged to magazines founded after 1949. I calculated the Simpson index for the first issue and kept only those magazines that had a Simpson index less than 0.4 so as to eliminate the generalist magazines. I then calculated the correlation between the Simpson index and each of the variables which reflect the proportion of space dedicated to each category. The results are presented in Table 6.6 below.

Variable	Correlation with the Simpson index
Proportion dedicated to politics	-0.1406
Proportion dedicated to economy	-0.0114
Proportion dedicated to social issues	0.0070
Proportion dedicated to literature	0.0326
Proportion dedicated to advertisements	0.4328
Proportion dedicated to knowledge	0.0389
Proportion dedicated to art	0.2756
Proportion dedicated to sport	0.0018
Proportion dedicated to other news	0.0567

**Table 6.6 Proportion between Simpson index and category proportions**

From the above table we can see that the correlation between the proportion of space dedicated to advertisements and the Simpson index is much larger than all other correlations. In fact, it is the only correlation which is not classified as “weak” (Pett, Lackey, and Sullivan 2003). This seems to suggest that the increase in the Simpson index was due to an increase in the proportion of space dedicated to advertisements. At this point, the source of this “anomaly” occurred to me. As I was coding the contents of the publications, I noticed a significant increase in the amount of space dedicated to advertisements in specialized magazines founded after 1949. An example will illustrate this more clearly. The magazine *al-iktisad al-Lubnani wal ‘arabi* (The Lebanese and Arab economy) was founded in 1951. The 1951 issue had a total of 39 pages with 38 dedicated to economics, and one page of advertisements. The 1960 issue of the same magazine had 138 pages of which 128 were dedicated to economics and the other 10 were advertisements. The 1971 issue had 176 pages, where 117 dealt with economics and 59 dealing with advertisements. The Simpson indexes of these three issues respectively are: 0.04996713, 0.13442554, and 0.4456999. There is a huge increase in the Simpson index as a result of including these advertisements in the category advertisements. However, this increase was due to an increase in advertisement space in these specialized magazines. In order to verify this, I refit Model 9 in Table 6.6 (results are not shown) but this time I replaced the interaction between *p5* and *age* with an interaction between *p5* and the proportion of space dedicated to advertisements<sup>13</sup>. This interaction proves to be significant at the

---

<sup>13</sup> It should be noted here that my prime interest in this research in general is not to study the effects that different categories had on the Simpson index of each period. However, here we include such a covariate only to explain the seemingly strange result that I obtained. In the end, I am interested in Model 9 as presented in Table 6.6 even if including the variables which indicate the total proportion dedicated to each



0.001 level. The new model produced a much lower  $QIC_U$  statistic indicating that it was a better fit (Hardin and Hilbe 2013). This means that the increase in the Simpson index of specialist magazines founded in the fifth period was better explained by the fact that they gradually increased the amount of space dedicated to advertisements than by an aging process. It is important to note that this problem has no implications on the previous periods because such specialized issues with increased advertisements appeared only during the fifth period. I included similar interactions between *p5* and the other categories but none had a statistically significant positive result that explained the observed increase in the Simpson index. But what if the increase in the Simpson index that we observed in the first period for all publications, which we explained as the rise of generalists, was actually just caused by a similar increase in advertising space in all publications? To test this, I fit a model in which I included an interaction between the proportion of space dedicated to advertisements and *p1* (being founded in the first period). The results show that the coefficient was actually negative, but statistically insignificant. So the increase in the Simpson index during that period cannot be attributed to a mere spanning of the advertisements category by newspapers founded in that period. I next included an interaction between the proportion of space dedicated to advertisements and *period1* (being published in the first period), and the result was again statistically insignificant. This confirms that the increase in the Simpsons index was not due to newspapers solely increasing their advertisements.

---

category produces a better fit model. Based on our research interest here, we do not care what categories cause the change in the Simpson index.

## 6.6 Survival Analysis

The next step in the analysis is to study the survival times of the publications. This is performed on the entire period 1851 – 1974. In order to do that I will use survival analysis (Hosmer, Lemeshow, and May 2011). This statistical method was explained in chapter 3. The primary function of interest is the hazard function, which is the instantaneous rate of failure. It represents the probability that the event of interest occurs conditional on the subject having survived up to the present time:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t + \Delta t > T | T > t)}{\Delta t}$$

$T$  denotes the nonnegative time to event while  $t$  denotes time.

### 6.6.1 Non-parametric Analysis

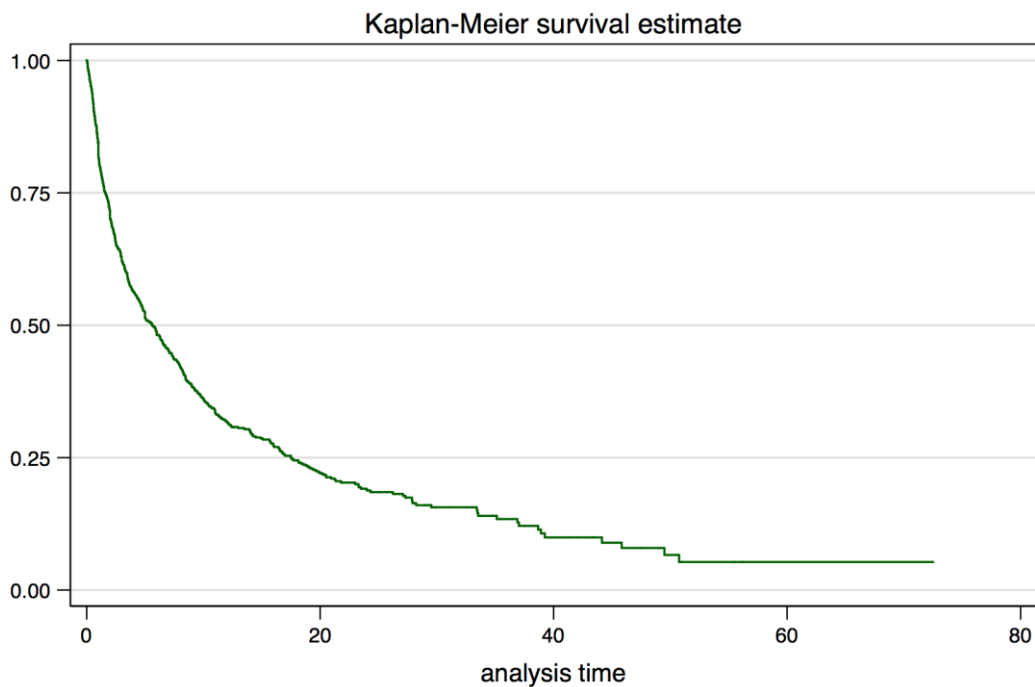
Figure 6.15 shows the survivor function that was produced from the KM estimator. We can clearly see from the graph that the probability of survival drops very quickly during the first years after founding before the rate starts to decrease during later years. Many studies, including ones conducted on newspaper populations (Carroll and Delacroix 1982), have provided empirical support for the liability of newness theory (Stinchcombe 1965) even when other factors such as size were controlled for (Pfeffer 1997). A possible explanation for the drop seen in the survival function in this case is that organizations usually start with a certain amount of capital, referred to as endowments. This capital is used to help sustain the company through the first period of operation, until the company can generate its own revenues<sup>14</sup>. With time, these endowments start to sun out and the effects of the liability of newness start to be seen. Figure 6.16

---

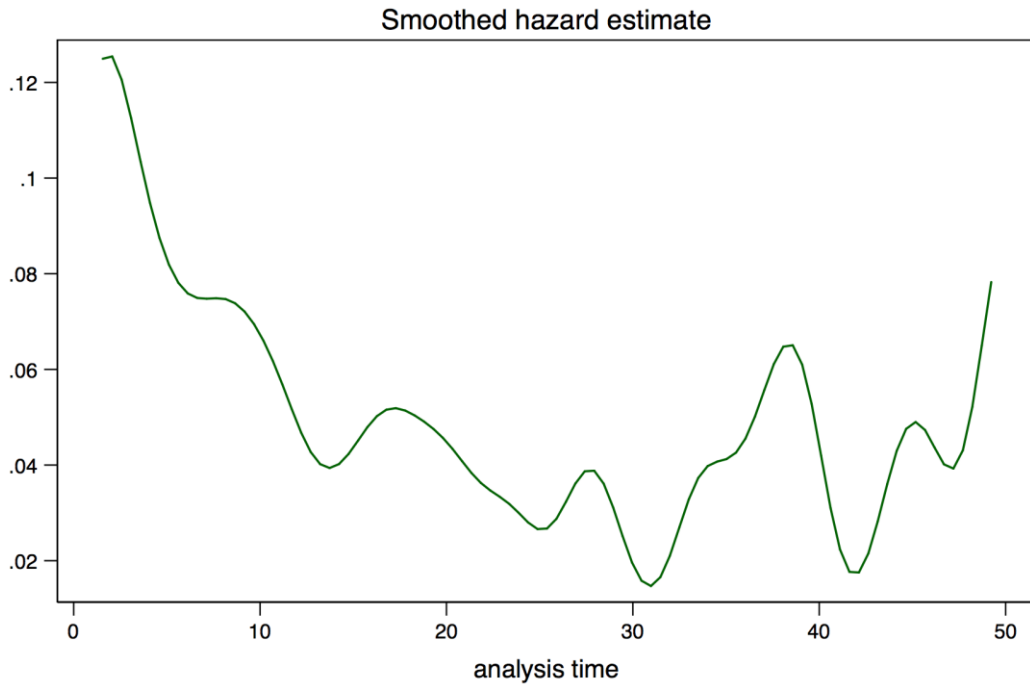
<sup>14</sup> I am grateful to Dr. Gabor Peli for raising this point

shows the smoothed hazard function that was obtained by using a Gaussian kernel function on the NA estimator. This graph also shows that young organizations face an extremely large hazard, and that this hazard decreased over time. However, we see that beyond a certain age (32 years) the hazard rate faced by organizations starts to increase considerably. This phenomenon has been explained by organizational theorists as being caused by either a liability of obsolescence or a liability of senescence (Barron, West, and Hannan 1994).

Next I ran a log-rank test in order to see if there was any difference in the survivor functions of political and non-political publications. The test indicated that the difference was not statistically significant. I also calculated the median survival time and found it to be 5.621 years.



**Figure 6.15 Survivor function from the Kaplan-Meier estimator**



**Figure 6.16 Smoothed hazard function from the Nelson-Aalen estimator**

### 6.6.2 Semi-parametric Analysis

I used the proportional hazards Cox model that was described in chapter 3. The results of the models are presented in Table 6.7 below. The models were successfully tested for the proportional hazards assumption. It should be noted here that all standard errors reported are robust because I used the *cluster* option since my data allows for multiple deaths for each newspaper. The first model in the table is the Base Model and it includes several control variables. The first four variables are period control variables. The first variable indicates the start of the reign of Sultan ‘abd al-Hamid up to the constitution of 1908. The next two variables were period indicators for the two world wars. The fourth variable indicates the period after the implementation of the 1953 law which restricted the number of new licenses. The only statistically significant period indicator was the World War 1 variable. We have seen in Figure 6.2 how the newspaper

industry was almost wiped out during that period. I next included some industry specific variables. The first two are the total number of active newspapers and its square. Past research in organizational ecology has uncovered a U-shaped relationship between mortality and organizational density (Carroll and Hannan 2004). However, our results indicate that there is a very narrow inverted-U shape, contrary to almost universal findings in previous research, but only the main term is significant. I also controlled for the total population of Lebanon. Numbers for the total population are severely lacking because the last census performed by the government was in 1932 and, due to religious and political grievances, the results of that consensus are highly questionable (Hudson 1985). I collected the available data from Faour (1991), Harris (1997), and Tarābulṣī (2007) and used linear interpolation to fill in the missing values. Although the coefficient is statistically significant, its value is extremely small. The last two variables I controlled for are the one-year lagged number of founding and disbanding events. Some periods saw an increase in the founding events, for example, after the 1908 constitution, and I wanted to see if an increase in founding events led to an increase in mortality due to increased competition. With regards to disbanding events I wanted to see if newspaper disbanding events tended to happen at the same time, i.e. that an increase in disbanding events in one year was following with more disbanding events in the next. The results show that both coefficients are significant. It seems that a high number of disbanding events in one year predicted higher mortality rates for those that survive in the next, while a high number of founding events predicted a lower mortality rate in the following year.

	Base Coef.	Model 1 Coef.	Model 2 Coef.	Model 3 Coef.	Model 4 Coef.	Model 5 Coef.	Model 6 Coef.	Model 7 Coef.
Reign of Sultan Abd al-Hamid?	.0194378	-.352521	.0170357	.0373108	-.002177	-.0211428	-.0337777	-.2152099
World War 1?	2.054869***	2.033788***	1.859282***	2.154345***	2.13591***	2.13289***	2.185482***	2.542394***
World War 2?	.1748053	-.0107065	-.0625812	-.0316291	.0005003	-.0031233	.0438507	-.0208161
After the 1953 license restriction law?	.3169871	-.2535052	.1206256	.4498086	.4953478	.4345334	.4876319	1.190341
Active newspapers	.0280945***	.0117969	.0312914*	.0254508	.0251011	.0252103	.0259798	.032773*
Square of active newspapers	-.0000805	-.0000423	-.0001021	-.0000835	-.0000826	-.0000818	-.0000871	-.0001089
Population of Lebanon	-1.29e-06***	-3.52e-07	2.67e-07	3.88e-07	3.87e-07	3.47e-07	3.21e-07	2.43e-07
Founding events (lagged)	-.0355948***	-.0195279	-.018732	-.0137723	-.0140624	-.0149797	-.0146178	-.027281
Disbanding events (lagged)	.0159225***	.0213356*	.0245825*	.0239093	.0235811	.0233124	.0218822	.0218015
Political?		-.3795896	-.412884	-.4254709	-.4256818	-.4084603	-.4117265	-.4859954
Number of pages		-.0035894	-.0036099	-.0036727	-.0037952	-.0035087	-.0036046	-.0035344
Type (base is magazine)								
<i>printed newspaper</i>		.7394218**	.5907409*	.7325638*	.6896002*	.7682437*	.707638*	.7747485*
<i>newspaper</i>		-.469057	-.412282	-.5545362	-.5800047	-.5190524	-.5352341	-.393172
Period (base is period 5)								
<i>Period 0</i>			3.583473***	3.360323**	3.547854**	3.262243**	3.42489**	4.490499**
<i>Period 1</i>			2.567032**	2.710461*	2.354251	2.690373*	2.352647*	3.048706*
<i>Period 2</i>			1.998999**	1.998924*	2.259908*	1.995211*	2.163153*	2.835326*
<i>Period 3</i>			1.449351*	1.744422*	1.971533*	1.706631*	1.857294*	2.720513**
<i>Period 4</i>			1.109103	1.460677	1.69893*	1.446339	1.598864*	2.174126*
Span (new category)				.3220182	.0618415			
Span*period1					1.160789			
Span*period5					.5348182			
Increase catnum						.0898936	-.3246206	-.44085
Catnum*period1							1.594216	2.055054*
Catnum*period5							.8278008*	.9200409*
Cluster spanning								.6985152*
AIC	6613.107	1462.433	1456.509	1242.215	1242.885	1245.229	1243.466	1146.941
BIC	6675.639	1538.484	1561.811	1351.526	1363.702	1354.539	1364.283	1273.302
likelihood	-3297.553	-718.2163	-710.2544	-602.1077	-600.4424	-603.6144	-600.7329	-551.4703
N. of cases	7693	2566	2566	2329	2329	2329	2329	2307

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Table 6.7 Survival analysis results**

Next in Model 1 I included some organization level control variables. I added a variable which indicated whether a publication dealt with politics or not, the number of pages, and the type of publication. Here it should be noted that other than newspapers and magazines, some journals were referred to as “matboo‘a” (publication) in the sources. There were only 27 cases such as this. The model results show that these “publications” had a statistically significant higher hazard rate when compared to magazines, but that the difference between magazines and newspapers was not significant. We see that after we added these organization level variables, the number of active newspapers and the lagged founding events are no longer significant. By looking at the AIC and BIC statistics we can clearly see that Model 1 is a much better fit than the Base Model. I next added indicator variables that indicated in which period the newspaper was founded (results are not shown in the table) but none of them was significant and the resulting AIC and BIC statistics were higher. So it seems that the period of founding was not significant.

In Model 2, I added indicator variables to indicate the period in which the issue was published. There are 6 period indicators here, with *period1* – *period5* referring to the same periods we used in the previous analysis in this chapter and *period0* referring to the years 1851-1874 inclusive, which I had identified in chapter 4 as the period of emergence. I used the fifth period (1950-1974) as the reference category. The results show that being published in any other period was more hazardous than being published in the fifth period. All coefficients, except the period 4 coefficient, are statistically significant. It is also interesting to see that the magnitude of the coefficients decreases with each consecutive period.

Next, in Model 3, I added the variable *span*. This variable is the same variable that we modeled as the dependent variable in Table 6.4. It records when a newspaper enters a new category. The reason for including this variable is that I wanted to see if entering new categories was statistically significant or not. It is important to note that this variable does not capture the number of categories spanned. It only captures if the newspaper enters a category that it had previously not included. The result indicates that spanning new categories increases the mortality rate, but the result is not statistically significant. Next I wanted to see if the variable *entering* was period dependent. Above, we have identified period 1 as the period in which generalists were formed and period 5 as the period in which specialists were formed. Perhaps spanning a new category is hazardous in the fifth period but not so in the first period. The result shows that spanning is hazardous in both periods. In fact, it is more hazardous in the first period than in the fifth period, but the result is not significant. These results indicate that entering a new category, without taking into account if the number of categories spanned by the newspaper decreases, stays the same, or increases, is not significant. Next I decided to see if increasing the number of categories spanned increases mortality. To do so I created a variable that I named *catnum*. This variable is set to one when the number of categories spanned by a certain issue is more than the number of categories spanned by the previous issue. Otherwise it is set to zero. I used the same logic that I used in previous sections in that I considered that an issue spanned a category when at least 11% of the issue was dedicated to the category. The result of adding this variable is shown under Model 5 in Table 6.7. We see that the coefficient is positive but that, once again, it is not statistically significant. Next I decided to see if the variable *catnum*



interacted with any certain periods. More specifically, I once again included the interaction of this variable with the period indicators for period 1 and period 5. Perhaps spanning a greater number of categories was not hazardous in period 1 but was hazardous in period 5. Model 6 shows that the coefficient of the variable has now become negative while both interactions are positive. However, only the interaction with period 5 is significant. These results indicate that spanning more categories is not always hazardous to the health of the organization, thus supporting hypothesis 3. Statistically speaking the model shows that spanning more categories is only hazardous in the fifth period, which is the period of specialization. With regards to which model is a better fit, Model 5 or Model 6, the AIC and BIC statistic give contradicting results with the AIC favoring Model 6 while the BIC favoring Model 5.

I have argued in this thesis that cluster spanning, more than category spanning, is what is most hazardous to organizational health. So far the results indicate that spanning more categories increases mortality only in the fifth period. But what about cluster spanning? In order to see the effect of cluster spanning on mortality rate I created a new variable that records whether an issue spans a different cluster than the previous issue. To do that I split each period into the number of clusters which were indicated by the Calinski and Harabasz index shown in Table 6.1<sup>15</sup>. Next, using cluster analysis, I assigned each issue in each period to one of the groups. An issue spanned a new cluster when it was not in the same cluster as the previous issue. So for example, if an issue of a newspaper was in the second cluster while the next issue was in the first cluster, I set the cluster spanning variable to 1. If they were in the same cluster then the

---

<sup>15</sup> I performed the same analysis using the groups as determined by the Duda-Hart index and got the same results.

variable was set to 0. Model 7 shows the result of including this variable in addition to the variables of Model 6. The results indicate that *catnum* is negative and not significant, but that its interactions with both the period 1 and period 5 indicators are positive. This means that spanning more categories increased mortality rates in both periods 1 and 5. The fact that spanning categories increases the hazard of mortality in period 5 lends support to hypothesis 4. What is even more important for us is that the coefficient of the binary variable which records cluster spanning is both positive and statistically significant indicating that spanning clusters increases mortality rates, thus supporting hypothesis 5. The AIC and BIC statistics clearly favor Model 7 above all other models. It is important to note here that a newspaper can span a different cluster without increasing the number of categories that it spans. For example, if an issue of a newspaper dedicates half its space to science and the other half to art, and then the following issue dedicates half its contents to politics and the other half to economics, then the number of categories spanned by both issues is the same, but both will probably belong to different clusters. What Model 7 tells us is that in such a case when *catnum* is zero, cluster spanning is hazardous to the organization's mortality. In addition, if an issue increases the number of categories that it spans in period 1 or period 5 and, by spanning more categories, is placed in a cluster different from the previous issue, then there are two positive coefficients that will cause an increase in the mortality rate. In other words, category spanning captures only one hazardous dynamic while cluster spanning captures another. Finally, in order to see if the effect of spanning clusters was period dependent, as was the case with category spanning, I interacted the spanning variable with the different time periods (results not

shown in the table). All of the interactions proved to be statistically insignificant and the resulting model was a worse fit than Model 7. This is in line with the statement in hypothesis 5, and that is that cluster spanning appears to be period independent.

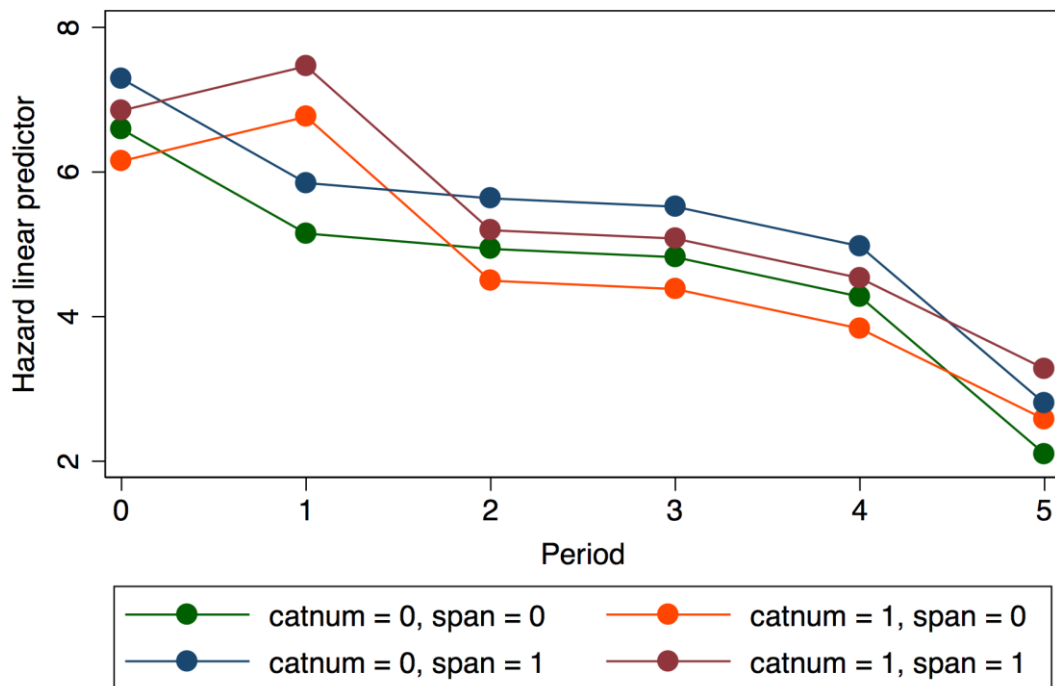
Table 6.8 below shows the value of the linear predictor for the hazard function for some combinations of the covariates at each time period. The first row shows a newspaper that is neither spanning a category nor spanning a cluster. The second row shows a newspaper that is spanning a category but without spanning a new cluster. The third row shows a newspaper that is spanning a new cluster but is not spanning extra categories. Finally, the last row shows a newspaper that spans both a new category and a different cluster. We see from the table that, in general, there is a decrease in the hazard as we move through time. The other covariates were taken at their mean values. Figure 6.17 plots these results in order for us to visualize the dynamics.

	Period 0	Period 1	Period 2	Period 3	Period 4	Period 5
No cluster spanning, no category spanning	6.5920	5.1502	4.9368	4.8220	4.2756	2.1015
Category spanning, no cluster spanning	6.1512	6.7644	4.4960	4.3812	3.8348	2.5807
No category spanning, cluster spanning	7.2905	5.8487	5.6354	5.5205	4.9742	2.8000
Category spanning, cluster spanning	6.8497	7.4629	5.1945	5.0797	4.5333	3.2792

**Table 6.8 Linear predictors of the hazard function**

We notice than in periods 0, 2, 3, and 4, the combination which results in the highest hazard function is the presence of cluster spanning and no category spanning. In periods 1 and 5 on the other hand, the presence of both cluster and category spanning produce the highest hazard function. This shows that, in all

periods, cluster spanning increases the rate of mortality, while this is not the case for category spanning. The effect of category spanning is period dependent in that during the first period (the rise of generalists) and the last period (the rise of specialists) it increases the mortality rate, while in other periods it has a negative (but statistically insignificant) effect on the mortality rate.



**Figure 6.17** The graph of the hazard linear predictor

The final analysis concerns hypothesis 6. Figure 6.9 and Figure 6.10 displayed the number and size of the active clusters during periods 2 and 5. We saw that the clusters formed during the fifth period were longer lasting and were more consistent than those formed during the second period. Both of these periods were described as periods of diversification because the average Simpson index decreased and I argued that this was because specialist newspapers were beginning to surface. However, we saw that the attempt made

during period 2 was not successful. This thesis has so far argued that cluster spanning is hazardous to organizations which attempt such actions. Hypothesis 6 posits that even organizations that do not span clusters face a higher risk if they are active within a period that is characterized with a lot of cluster spanning activity. Table 6.1 shows that the second period has the highest average Simpson index with respect to cluster spanning while the fifth period had the lowest even though it had the greatest number of clusters. Table 6.2 shows that the average Simpson index of the second period is statistically greater than both the period which preceded it and the period which followed it. According to hypothesis 6, the high amount of cluster spanning should be responsible for the failure of the diversification process. Diversification was successful during the fifth period because the organizations managed to form more clusters but with minimum cluster spanning activity. This was not the case for the second period. In order to test whether a high level of cluster spanning activity is detrimental to the survival of organizations I needed to come up with a measure that would adequately capture the phenomena. Because there is one issue of each newspaper during every year, it is not possible to calculate the measure on a yearly basis since one issue is not enough to see if the newspaper tends to span more clusters. The only way to tackle the issue was to use the concept of a window that has a width that is greater than one year. The problem here is in determining how large the window has to be. If I used a window of five years, then issues of newspapers that have been active for more than five years would be ignored while issues of newspapers that have been dead for three years would be included. This does not seem like a good way to approach the issue. Therefore, what I decided to do was to use a moving window which was not

bounded on the left but which only included the organization if it was still active during the year of investigation. So if I were calculating the level of spanning activity during year 1952, I would include in the calculation all of the previous issues of all newspapers that are still active during the year 1952 as well as new newspapers that were founded in that year. This measure is not affected by newspapers that have ceased to exist, and it takes into account all previous issues of existing newspapers. The next issue here is to decide on the number of clusters that the data should be divided to. The moving window might include newspapers from different time periods, so the previous division of the data into periods 0 to 5 will not suffice here. To solve this problem I ran a cluster analysis on all of the data in the dataset in order to find the optimal number of clusters that all the issues can be grouped to. The results are presented in Table 6.9.

Number of clusters	Calinski and Harabasz index	Duda-Hart index	
		Je(2)/Je(1)	Pseudo T-squared
1	-	0.8649	386.76
2	386.76	0.5937	1577.01
3	1132.78	0.8367	334.22
4	1008.22	0.7957	389.72
5	1010.86	0.6324	829.36
6	<b>1251.45</b>	0.9428	35.76
7	1059.95	0.8264	40.55
8	928.12	<b>0.9735</b>	<b>5.74</b>
9	814.85	0.9055	60.56
10	743.25	0.8166	272.63
11	762.95	0.8265	35.69
12	707.36	0.7357	74.01
13	674.82	0.9430	2.66
14	623.34	0.7753	25.80
15	587.62	0.6266	320.01

**Table 6.9 Results for finding the optimal number of clusters 1851-1974**

According to the Calinski-Harabasz index the optimal number is six while according to the Duda-Hart index the optimal number is eight. According to the Calinski-Harabasz index, eight groups has the sixth most optimal number of clusters, while according to the Duda-Hart index six groups is the third most optimal number. Therefore I chose to divide the data into six groups since it is the best number of groups if we take into consideration the two measures that we have used.

I next ran a program which calculates the spanning activity using the Simpson index that I have used throughout this thesis. At each year, this program would only keep the issues of newspapers that are currently active and separate them into six different clusters using cluster analysis. After that I calculated the Simpson index for each newspaper title by looking at the number of issues in each group with respect to the total number of issues. Finally, the program took the average of the Simpson index of all newspapers. I add this variable, which I call *environment*, to Model 7 and obtained the results that are presented in Table 6.10 under Model 8. We see from the results that the coefficient of the new variable is positive but that it is not statistically significant. At this point I decided to study whether the effect of this variable is binary rather than continuous. What if there was a certain ceiling above which cluster spanning activity in the environment becomes hazardous to all active organizations. To find whether this was true or not, I took as the ceiling the 90<sup>th</sup> percentile value of variable *environment* and coded a new variable, called *environment2*, as one if the measure during that particular year was greater than that ceiling and as a zero if it was less. The results of including this variable are shown under the

column of Model 9 in the same table. We can see that the coefficient is both positive and statistically significant thereby lending support to hypothesis 6.

	Model 8 Coef.	Model 9 Coef.
Reign of Sultan Abd al-Hamid?	-.2680784	-.0219979
World War 1?	2.381954***	2.611259***
World War 2?	-.1363554	.0032857
After the 1953 license restriction law?	1.082831	1.168867
Active newspapers	.0331561*	.0376986*
Square of active newspapers	-.0001098	-.0001281
Population of Lebanon	4.95e-07	2.86e-07
Founding events (lagged)	-.0199196	-.0243813
Disbanding events (lagged)	.0216259	.0249468
Political?	-.4722823	-.4557047
Number of pages	-.0036241	-.0036512
Type (base is magazine)		
<i>printed</i>	.8002299*	.745107*
<i>newspaper</i>	-.4320644	-.4378093
Period (base is period 5)		
<i>Period 0</i>	3.896204**	4.362457**
<i>Period 1</i>	3.19632*	3.212988*
<i>Period 2</i>	2.866524*	2.791085*
<i>Period 3</i>	2.698466**	2.763968**
<i>Period 4</i>	1.928286	2.182516*
Increase catnum	-.4472828	-.4751279
Catnum*period1	2.101795*	2.097342*
Catnum*period5	.9146806*	.9612848*
Cluster spanning environment	.6489488*	.6718665*
environment2	3.335832	1.434091*
AIC	1145.059	1145.799
BIC	1277.134	1277.875
likelihood	-549.5293	-549.8997
N. of cases	2304	2304

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

**Table 6.10 Survival analysis results including the environmental cluster spanning measure**



### 6.6.3 Other Aspects of Identity

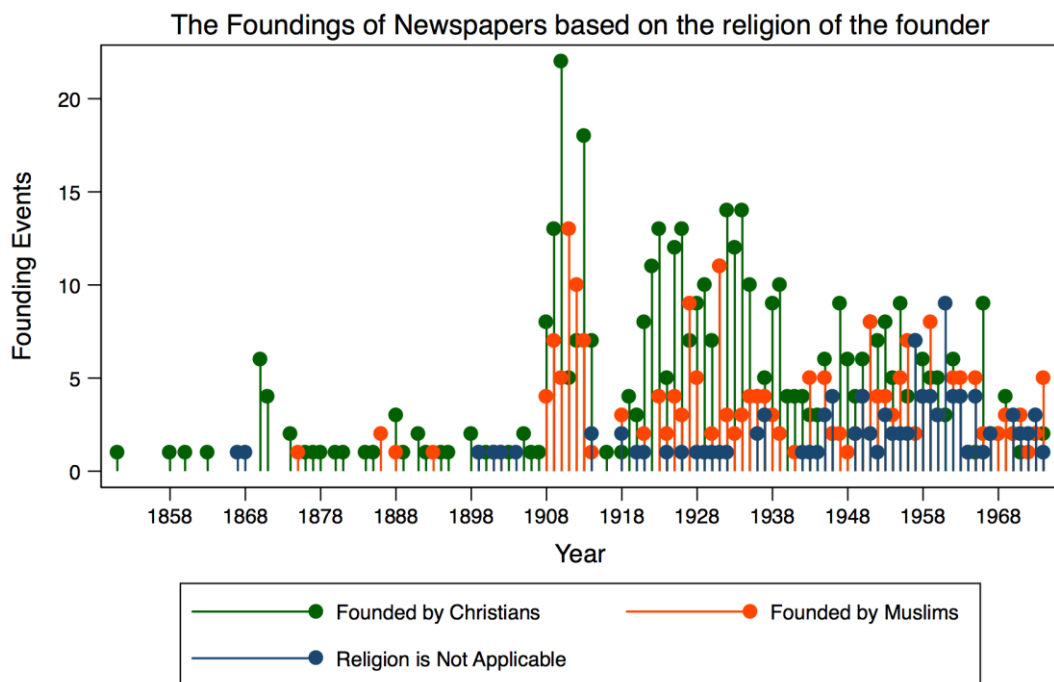
The results obtained in the present chapter indicate that spanning clusters is hazardous. The clusters were formed by the category spanning activities of the organizations. However, the identity of organizations is only partially captured by the combination of categories to which an organization belongs (Hannan, Pólos, and Carroll 2007). Chapters 4 and 5 of this thesis studied two different dimensions of identity during the period of emergence of the industry, and we saw that the results obtained depended on the dimension studied. Therefore, it would be interesting to see if other aspects of identity had a negative or positive effect on the mortality of the newspapers throughout the entire period of study (1851 – 1974)<sup>16</sup>. One of the most defining factors of Lebanon is the fact that the political system is based on power being shared among several religious sects, none of which has a clear majority. Of course, this “sharing” of power has been very problematic especially since most of the political and economic power was in the hands of Maronite Christians starting from independence in 1943 up to the end of this study in 1974 (Hudson 1985). Prior to that, religion was also an important factor that played an integral part in the shaping of loyalties during both the times of the Ottomans and the French. Therefore, the next step in this chapter is to study whether religion had any effect on the newspapers’ mortality rates. In order to do so I created a variable that I called *religion*<sup>17</sup>. This variable recorded the religion of the founder of the newspaper: Muslim, Christian, “Both” and “Not Applicable”. “Both” indicates that the newspaper had more than one founder and not all of them belonged to the same religion while “Not Applicable” indicates that the newspaper was founded by a non-religious organization, like

---

<sup>16</sup> I am thankful to Professor László Pólos and Dr Mehmet Asutay for raising this point.

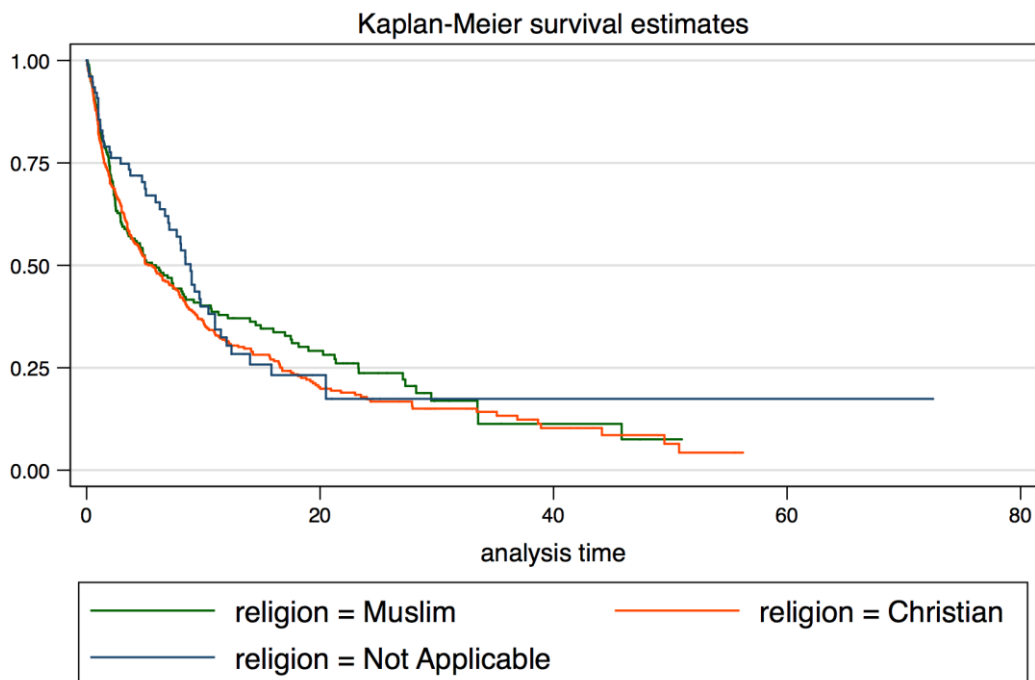
<sup>17</sup> I am thankful to Dr Hussein Abdulsater for helping me in coding this variable.

the Communist party for example. Based on my dataset, this variable indicated the following: out of a total of 1,056 newspapers 486 were founded by Christians, 244 by Muslims, 4 by “both”, and 110 by “Not applicable”. There were 212 instances of a missing value. Since very few newspapers fell under the heading “both”, I eliminated it from the following analysis. Figure 6.18 plots the founding events each year. We can see from the figure that throughout most of the period of study founding events by Christians far exceeded those by the other groups. We can also see that the magnitude of the difference decreased with time until eventually the founding events by Muslims equaled, and in some instances exceeded, those by Christians. We can also see that the founding events by organizations that had no religious affiliations increased significantly after independence.



**Figure 6.18 Newspaper founding events divided by religion of the founder**

The next step is to perform non-parametric survival analysis. Figure 6.19 below shows the Kaplan-Meier estimate of the survivor function for the three groups. We can see from the graph that the behavior of the three graphs is similar to one another. Newspapers founded by non-religious organizations start off with a higher survival rate but this does not continue for long. Newspapers founded by Muslims and those founded by Christians have similar trajectories initially, but eventually the survival rate of those founded by Muslims becomes greater before eventually dropping to the same level as that of the newspapers founded by Christians.



**Figure 6.19 Non-parametric Survivor function based on the variable religion**

Next I calculated the mean survival time of each type of newspaper. Those founded by Muslims had a mean survival time of 5.94 years, those founded by Christians had a mean survival time of 5.27 years, and the “Not Applicable” group had a mean survival time of 8.91 years. In order to study the statistical

significance of these results I performed a log-rank test. The results were not statistically significant. I then eliminated the “Not Applicable” group in order to see if the difference between newspapers founded by Muslims and Christians was statistically significant. The results were again not statistically significant.

Finally, I added this variable (*religion*) to the Cox proportional hazards model that we found to fit the data best. The results indicated that newspapers founded by Christians had a lowered hazard of mortality than those founded by Muslims, but the result was not statistically significant. Newspapers that fell under the “Not Applicable” group had a higher hazard but again the result was not statistically significant. The addition of this variable did not have any major effect on the results obtained previously in that no coefficient that was statistically significant lost its significance and no coefficient that was found to be not significant became significant. These findings indicate that the observed differences in survival rates between newspapers founded by Muslims and those founded by Christians were not statistically significant.

Chapter 4 studied the framing process during the industry emergence period. This framing process was studied by looking at the “stories” told by newspapers. Specifically, I looked at the problems identified by newspapers and how the newspapers believed that they could play a role in solving the problem. A similar approach might have been taken to analyze the framing process during the period after 1879. It would have been interesting to see if there were other shifts in the framing process, just like the shift that took place after 1875. The reason that I did not look into this matter was related to time constraints. Analyzing the stories told by newspapers, in connection to the social, economic, and political environment, as I did in Chapter 4 for the period of 1851 to 1879,

would have required a considerable amount of time. Such an analysis would require a separate thesis due to the time frame and the breadth of the topic. As mentioned in this thesis several times, identity is multidimensional, and the importance of this topic cannot be denied, but I had to make a deliberate choice on what I was interested in studying. In the end, I decided to study the identity of newspapers as projected by their category spanning dynamics throughout the whole period, while at the same time studying the framing process during the period of emergence. The framing process that was taking place during other time periods is an important and interesting topic and could be tackled in a whole different research. For the sake of this thesis, I found it useful to briefly look at the religious identity by accounting for the religion of the founder of the newspaper. This analysis did not reveal any statistically significant results. Other aspects of identity are not less interesting than the ones studied in this thesis, but they require, and deserve to be, more than to be a part of this thesis.

## **6.7 Categories or Features?**

This chapter has argued that clusters are formed from the combination of categories that organizations choose to span, and that spanning these clusters is hazardous. The theory formed in Chapter 2 and tested in the present chapter presented a critique of the view that spanning categories is hazardous at all times. However, Hannan, Pólos, and Carroll (2007) “define a type as a coupling of a label and a schemata that articulates a view about what pattern of feature values determine the applicability of the label” (p. 60). Later they state, “We can now define a category as a type for which an audience segment achieves a high level of extensional and intensional consensus” (p. 69). What if the “categories”

that were identified by this thesis were nothing more than features? This would imply that the “clusters” which were formed were actually categories, and thus the findings of this thesis would be exactly in line with present research on the effects of categories<sup>18</sup>. In order to show that this was not the case I would need to show that the categories identified in Chapter 3 (news, sports, social issues, economics ... etc.) were actual categories and not mere features.

I have shown in this chapter that the category spanning dynamics of the newspapers were in line with the theoretical reasoning of Carroll (1985). We have seen that Simpson’s index, which was calculated based on the categories that I have identified, increased during the early periods, indicating a rise in generalists and later decreased. We have also seen that the rise of generalists was caused by old newspapers expanding into new categories while the rise of specialists was caused by new newspapers being founded as specialists. If the categories that I identified were features and not categories then this thesis should not have achieved such results. There is nothing in the literature that states that the combination of features should mimic the dynamics of the rise of generalists and specialists. All the results obtained in this chapter depended on the identification of the categories and the calculation of Simpson’s index was also highly dependent on the category identification. In order to illustrate this more clearly, I wrote a program that displayed the amount of space dedicated to each category during the last period (1950 – 1974) of this study. This is the period that saw the rise of specialists. Specialists tend to span few categories, but they do not necessarily have few features. Therefore, if my identification of categories were correct, I would expect to see many of the clusters formed

---

<sup>18</sup> I am once again thankful to Professor László Pólos for bringing up this point.

during this last period to be populated by newspapers that concentrate their content on a certain category. If on the other hand most of the clusters contained newspapers that spanned more than one category equally then this would imply that what I have called categories throughout this thesis are actually features.

In order to move forward with the above I had to determine the best fit number of clusters. As we saw in this chapter the Calinski and Harabasz measure indicated that six clusters were the best while the Duda-Hart index indicated that eight clusters was a better fit. According to the Calinski and Harabasz measure, eight clusters is a very good fit, but not better than six, while according to the Duda-Hart index six clusters is not a good fit. Therefore I choose to divide the data into eight clusters because both measures agree that it fits the data well. After dividing the newspaper issues into the eight best-fit clusters I calculated the average proportion dedicated to each category in each cluster. The results are presented in Table 6.11. In this thesis I have used the value 0.11 as the cut-off value to determine whether a newspaper issue spanned a category or not. We can see that, except for the category “sports”, every single category was spanned by at least one cluster. We can also see that there was always a single cluster which dedicated more than half of its content to that category: politics (0.6181659), economics (0.7894179), social issues (0.782898), literature (0.5988283), advertisements (0.5226051), knowledge (0.7181807), and other (0.5385878), while art was a bit less than half (0.4231835). These results show that most of the clusters tended to dedicate the vast majority of their contents to a single category. They also show that each category was the center of attention of exactly one cluster.

Cluster	Politics	Economic	Social issues	Literature	Adv.	Knowledge	Art	Sports	Other
1	<b>.6181659</b>	.039677	.0555541	.0215158	.114132	.0378301	.0157275	.0122072	.0851905
2	.0194056	.0111833	.0373943	.119724	.0240238	<b>.7181807</b>	.0171571	.0003391	.052592
3	.0741246	.0624213	.0376506	.1558374	.0377099	.1997369	<b>.4231835</b>	0	.0093357
4	.0377098	.0049874	.0373405	<b>.5988283</b>	.0616321	.1406928	.061735	.0003157	.0567585
5	.0285686	<b>.7894179</b>	.0516549	.004174	.0635814	.0161905	.0055126	.0004871	.0404129
6	.0355	.3783652	.024375	.00175	<b>.5226051</b>	.0038881	0	.001875	.0316417
7	.0630539	.0147198	.0748345	.0589789	.0626158	.1442594	.0397717	.0031781	<b>.5385878</b>
8	.0109805	.0105513	<b>.782898</b>	.0279226	.0187935	.0271007	.0080899	.0010915	.112572

Table 6.11 Average proportions dedicated to each category in each cluster 1950-1974

	Politics	Economic	Social issues	Literature	Adv.	Knowledge	Art	Sports	Other
Politics	1.0000								
Economic	-0.2852	1.0000							
Social issues	-0.2172	-0.1160	1.0000						
Literature	-0.2824	-0.1880	-0.0919	1.0000					
Adv.	0.1052	-0.0307	-0.1274	-0.1044	1.0000				
Knowledge	-0.4673	-0.2431	-0.1668	0.0613	-0.3121	1.0000			
Art	-0.1092	-0.1254	-0.0539	0.1952	0.0250	-0.0364	1.0000		
Sports	0.2156	-0.1178	-0.0812	-0.1303	0.2517	-0.1830	-0.0158	1.0000	
Other	-0.1985	-0.1942	-0.0421	-0.0716	0.0015	-0.1404	0.0309	0.0542	1.0000

Table 6.12 Correlations between categories 1950-1974



Next I display the correlation matrix during that period for all categories. If these categories were in fact features then we would expect to see a high correlation between some of them, if in fact organizations do form categories by combining features. In the words of Rosch et al. (1976), “Subordinate categories have lower total cue validity than do basic because they also share most attributes with contrasting subordinate categories” (p. 385). If, however, my choice of categories was correct, then we should not see strong correlations because this was a period of specialization and that means that categories do not tend to be spanned together by a large number of organizations. The correlation matrix is displayed in Table 6.12. The highest positive correlation is 0.2517, which is very weak. Again, this result shows that the categories that I identified are not features. Another important thing to notice in the correlation matrix is that out of the 36 correlations, 27 are negative while only 9 are positive. Again this vindicates my choice of categories because it shows that if one category is spanned then other categories tend not to be spanned. Taken together, all of the above results show that what this thesis identified as categories were in fact categories and not features.

## **6.8 Resource Partitioning: Positioning or Authenticity?<sup>19</sup>**

It was mentioned in Chapter 2 that while a wide array of research has indicated that resource partitioning does take place, more than one reason was proposed for this phenomenon. One reason pointed to the fact that as generalists move to the center of the resource space, they forgo the periphery regions in order to concentrate on serving the needs of the larger audience. Other research

---

<sup>19</sup> I am grateful to Dr. Gabor Peli for raising this point.

has indicated that concepts such as authenticity are what drive resource partitioning. This raises the question of whether the resource partitioning that we have seen in this chapter was caused by specialists that wanted to serve the audience members whom were located at the peripheries, or whether resource partitioning was caused by authenticity.

Since newspapers have a limited space, category spanning is a zero sum game, in that if the content available to one category expands, there must be a decrease in the space dedicated to another category. This would lead me to suspect that the resource partitioning which has taken place in the Lebanese newspaper industry between the periods of 1959 and 1974 was caused by specialists targeting audience members that were being under served. If a newspaper wanted to be a generalist, it had to cover more and more categories. Hence, it cannot dedicate a large amount of space to a single category. In such a case audience members that are interested in a certain category would find that the generalist is no longer providing their focal category with the necessary amount of space. In order to see if the above is actually what happened in the case of the Lebanese newspaper industry, we can take a look at Table 6.11. As mentioned above, we can see from the table that every single cluster dedicated more than half of its content to a certain category, and that, with the exception of the category “sports”, this applied to every single category. This supports the suspicion that the specialists targeted a category that was being underserved due to space constraints. For example, some newspapers concentrated on economical issues and thus underserved the remaining categories, while other newspapers took advantage of this and dedicated more than half of their content to some other category that was not being served properly. Therefore, Table 6.11

indicates that different clusters chose to serve different customers and that they partitioned the resource space amongst themselves as such.

## **6.9 Conclusion**

This chapter studied the category and cluster dynamics of the newspapers throughout the period 1851 – 1974. I argued that category spanning was another aspect of the resource-partitioning theory. The resource-partitioning theory argued that the hazards faced by generalists and specialists are not uniform and that they were time dependent. I tested this same idea but instead of relying on count models that studied the number of generalists and specialists, and instead of classifying organizations as generalists or specialists, I relied on the Simpson index to do the work for me. Next, I argued that if it were the case that category spanning is an aspect of the resource-partitioning theory, then we would expect that the effect of category spanning was in fact not uniform throughout the life period of the industry. The survival models show that category spanning is period dependent. From there I argued that instead of thinking of category spanning we should be studying what I referred to as “cluster spanning”. Clusters are formed from the dynamics of category spanning. I argued that the act of spanning clusters is hazardous to organizations no matter in which time period they were active. Some category spanning activities lead to cluster spanning activities, but there exists some category spanning activities that do not necessarily result in spanning a new cluster. Even more interesting is the fact that it is possible to span a cluster even when there is no category spanning activity. It should be noted, however, that even when we take into account cluster spanning in the above models, the effects of category spanning

are still hazardous in certain periods. Ultimately, in some time periods, the highest hazard is faced by organizations that span both categories and cluster at the same time.

The chapter then studied the effect that other aspects of identity have on organizational mortality, namely the religion of the founder of the organization. This choice of variable is justified within the Lebanese context because of the political and social nature of the country. The results indicate that while there are some differences when we take into account the religion of the founder, these differences are not statistically significant. The most interesting finding with regards to this matter is the fact that founding events by Christians dominated the industry up to a certain point after which the Muslims caught up with them. It was also interesting to see that newspapers published by organizations that transcended religious boundaries started to proliferate during later years.

Finally the chapter ended by defending the choice of categories. The distinction between categories and features is central to the findings of this thesis. Feature dynamics do not necessarily mimic the findings of the resource partitioning theory, while category dynamics should lead to the same results. In addition, the fact that the categories identified by this thesis were at the heart of the division of the clusters after 1949 justifies my choice. By identifying the categories that organizations spanned, this thesis was able to identify the clusters that emerged from their dynamics. The distinction between cluster spanning and category spanning has important implications that will be discussed in the last chapter, among other things.

## **7 Chapter Seven: Conclusion**

### **7.1 Introduction**

This thesis started by giving a review of the current literature and identifying two areas that were fertile grounds for research: identity projection during industry emergence and category spanning effects on the survival chances of organizations. Later chapters provided both the theoretical underpinnings and the empirical findings related to the main research questions. This concluding chapter aims to summarize all of the findings while highlighting the main contributions of this thesis as well as the limitations that the nature of the data imposes on the findings.

### **7.2 Summary of Findings**

This thesis set out to study two different but related questions. The first was concerned with the collective identity of newspapers during the emergence of the industry. Taking into account that this is a multi-dimension issue, this thesis studied the identity from two different perspectives. The first perspective sought to explain the emergence of the industry as a social movement that utilized the technique of “framing” to justify its existence. Social movement scholars have studied the framing process in which social movements alert their audience to problems, propose a solution, and provide the necessary motivation. Two groups have emerged as a result of this question: one argued that the intensification micro-level grievances were the trigger and the other argued that macro-level resource-mobilization considerations were the central actor, with

grievances being used merely as tools. With regards to the identification problem this thesis has showed that the newspaper industry in Lebanon, during its formation days, constituted what is regarded as a social movement. Industry members have stated that the nation faces the problem of falling behind when it comes to civilization economically, politically and culturally. They proposed that the only solution to this problem was the spreading of education and argued that newspapers were one of the most important tools in doing so. Finally the newspapers motivated the people by arguing that the previous glory days of the East can be repeated through this strive for education. Most of the newspapers have ascribed to themselves the role of educational tools and most of the founding fathers were themselves men of knowledge. In this regard, this thesis has shown that the newspapers have projected a unified identity to their audience members in that they clearly had the same tools and the same goals.

The second perspective sought to study whether the category spanning dynamics of the newspapers were similar. Chapter 5 used the logical formulations introduced by Hannan, Pólos, and Carroll (2007) in order to take an in-depth look at both category-spanning and cluster-spanning dynamics of the Lebanese newspapers at the time of their emergence. The first was the spanning which took place on the level of the categories themselves, while the second was the spanning which took place on the level of cluster that were formed by the data. In both cases it was found that during later years in the formation period the newspapers became much more consistent in their data coverage in their respective issues, and also became more similar to each other. Another important finding was that in the middle of the 1870s there was a change in the content of the newspapers. Interestingly, this change happened around the same

time that a change in the goals and motivation of the newspapers took place, as shown in chapter 4. The observed change was best described as a process of maturation with age. The older the newspapers became the more focused their content became. Finally, this thesis found that two clearly separate identities had formed by the end of 1870s. One represented political newspapers and the other represented scientific newspapers. It took a significant amount of time for these two identities to be clearly separated. Therefore, this thesis has shown that, at least with regards to that part identity which is determined by the categories spanned by the organization, industries can, and do, emerge by projecting more than a single identity.

The thesis also argued that the clusters that emerge from the category spanning dynamics represent the category space. By understanding this category space it would be possible to understand how category spanning affects organizations. Chapter 6 studied the category dynamics of the newspapers throughout the period 1851 – 1974. The chapter showed that category spanning follows the pattern that is predicted by the resource-partitioning theory. Next, I argued that the effect of category spanning was in fact not uniform throughout the life period of the industry. The survival models show that category spanning is period dependent. From there I argued that instead of thinking of category spanning we should be studying what I refer to as “cluster spanning”. The thesis showed that the act of spanning clusters is hazardous to organizations no matter in which time period they were active. Some category spanning activities lead to cluster spanning activities, but there exists some category spanning activities that do not necessarily result in spanning a new cluster. Chapter 6 also showed that organizations that existed during time periods that were characterized by a

high level of cluster spanning faced a higher hazard of mortality than those that existed in periods with fewer cluster-spanning activities. The chapter then moved on to study whether the religion of the founder of the newspaper had any impact on organizational mortality rates. This was done because in Lebanon religion plays an important part in both the political and economic spheres. The results show that, at least with regards to organizational mortality, the variable *religion* was not statistically significant, but the chapter did show that the dynamics of founding events changed with time, with Muslim founding event equaling those initiated by Christians near the end of the study period. Finally, chapter 6 concluded by showing that my choice of categories was justified. This was important because otherwise the results obtained in chapter 6 would have been similar to previously obtained results.

## **7.3 Research Implications**

### **7.3.1 Discussion of Findings**

This thesis has produced three main findings that can be categorized in two time periods: the period of emergence (1851 - 1879) and a larger time frame that started with industry emergence and was censored in 1974. With regards to the period of emergence, the results presented in this thesis show that collective identity was a crucial factor in the successful legitimation of the newspaper industry in Lebanon. The first main finding that relates to the period of emergence was that the newspaper industry in Lebanon was founded as a social movement. As a social movement, it was crucial that the founding newspapers project a unified identity. The thesis has shown that all newspapers identified



the same problem and proposed the same solution: an increase in education. This is why the newspapers presented themselves as educational tools. Interest in using social movement analysis tools in the study of industry emergence has been increasing, and this thesis has contributed to these efforts. In addition, this thesis has shown that the same tools used to study social movements in Europe and America can be used to study similar movements in the Middle East.

Some research in organizational ecology has indicated that a unified identity was a facilitating condition for successful industry emergence. While it was expected that, as a social movement, newspapers project a unified collective identity through storytelling, the second finding of this thesis sought to contribute to the debate of whether, on another dimension, a unified identity was also to be expected. This thesis found that when identity was determined by the category spanning dynamics of organizations, a unified identity did not emerge. In fact, cluster analysis was used to show that two separate and distinct groups of organizations were active. It was also seen that the newspapers themselves were aware of this distinction, and this was reflected in their respective masthead. This is an important finding because it highlights the multidimensionality of identity, and it shows that while a unified identity needs to be projected on some dimensions, this was not the case on other dimensions. This allows for organizations who are seeking legitimation during the crucial period of emergence to both group themselves with other organizations while at the same time to differentiate themselves.

The third finding of the study relates to the effect that category spanning has on organizations. This effect has received considerable attention over the past few years. Unfortunately, with the exception of the work of Pontikes and

Hannan (2012), the concept of category space that was developed by Kovacs and Hannan (2011) is yet to be tested within diverse contexts. The third finding of this thesis is, in my opinion, the most significant finding. The reason for that is that the theory of cluster formation creates a much needed link between traditional research in organizational ecology and more recent research. Tradition research has shown that more than one form of organization can exist at the same time. Specifically, it is well documented that specialists and generalists can, and do, co-exist. Recent findings on the other hand argue that specialists always have the advantage and that generalists are, in most cases, penalized for spreading themselves too thin. By showing that category spanning dynamics follow the predictions of the resource partitioning theory, and by using clusters as a representation of the category space, this thesis has provided an explanation for the two seemingly contradictory findings. The thesis has shown that category spanning is hazardous when the environment favors specialists, or when spanning categories leads to cluster spanning. The thesis has also managed to build on previous work in order to show how the category space can be taken into account without affecting our conceptualization of the Simpson index, and without violating the principle of allocation.

### **7.3.2 Implications for Future Research**

The cluster-spanning theory that is developed in this thesis can be tested within the context of populations other than that of the Lebanese newspaper industry. By studying clusters, we would not only be studying the effects of category spanning, but we will also be studying how organizations group together during different periods throughout the lifetime of the industry. As

mentioned above, this thesis did not use a random sample, and therefore it would be interesting to know if the results obtained can be duplicated within other contexts. I believe that cluster analysis is a very promising direction which organizational ecologists should direct their attention to. Researchers in the field have always had their differences with regards to how to identify forms. This has led to huge efforts that were mostly directed at the study of a single organizational population. In a world in which boundaries between organizational populations are quickly eroding due to the rise of huge corporations, it would be useful if organizational ecologists were able to take into account more than one population at a time. Since cluster analysis is a multi-level approach, I believe that it allows researchers to do exactly that. While this thesis studied how clusters formed from category dynamics, there is no reason why the same approach cannot be used to study other levels<sup>20</sup>. For example, the researcher can replace categories with industries. This would allow him/her to study how clusters form from the activity of organizations in different industries. Organizational ecologists need no longer concentrate their efforts on single industries. On the other hand, the researcher might choose to move downwards instead of upwards, in that instead of moving from categories to industries, he/she can move from categories to features. Categories are formed from the combination of features. A valid research agenda is to study whether the combination of features changes over time and how do categories emerge from these combination. Again this can be done using the cluster analysis approach utilized in this thesis due to the fact that this approach allows for two separate but related levels. The researcher need not collect data that is continuous since

---

<sup>20</sup> I am once again thankful to Professor László Pólos for directing my attention to this point.

various binary measures exist in the cluster analysis literature. Therefore, the nature of data collected might differ from that used in this thesis, but that does not imply that cluster analysis cannot be used. The method used in this thesis can be applied to other settings with different sets of data, thus enabling researchers to test the theories developed in this thesis in a large, and diverse, number of settings.

#### **7.4 Generalizability and Limitations**

It is always important to determine whether the results obtained can be generalized, or if in fact they only make meaning within the context in which they were obtained. Chapter 4 showed that the industry started as a social movement. This, of course, is a context dependent result and has no implication for other populations. However, the fact that the Lebanese newspaper industry started as a social movement highlights the fact that industries can be created for non-economic reasons. This is not a new finding, but it is interesting to see that non-economic reasons were the reason for the creation of this industry, especially since newspapers founded in other countries were created for different reasons.

With regards to the results obtained using the quantitative tools, there are two things to consider. First, the development of the theory of clusters that was included in chapter 2 and the statistical tools used to analyze these theoretical constructs that were included in chapter 3 are generalizable. The theory that clusters form from category dynamics and that these clusters can be analyzed using cluster analysis and then included in survival models does not depend on the context of the study. These theories and models build on previous

theoretical formulations and hope to make a general contribution. The critique presented of the work of Kovacs and Hannan (2011) does not take into account any empirical setting. In addition, the work of Hannan, Pólos, and Carroll (2007), on which this thesis heavily relies, is intended to produce results that are generalizable and that address any previous conflicts.

When we consider the data used to test the theories that were developed in chapter 2, we run into a problem with regards to the generalizability of the results. The reason for that is that the data used does not represent a random sample due to the fact that newspaper collections are not random. This represents the greatest limitation of this thesis. The fact that more than one source were used to collect the data ameliorates the problem to some extent, but it does not completely address this concern. Therefore, this thesis cannot claim that the results obtained apply to other populations, even though the theory on which they were developed offers a sound basis from which to start. This brings us to the final point: implications for future research.

## 8 References

- Abu-Manneh, Butrus. 1980. "The Christians between Ottomanism and Syrian Nationalism: The Ideas of Butrus al-Bustani." *International Journal of Middle East Studies* 11:287-304.
- al-Majlis al Thaqafi al-Lubnani. 1996. *Sahafat Trablus wa al-Shamal fee mi'at Am 1893-1995*. Tripoli: Gros Press.
- al-Mishkat. 1878. "Beirut." Pp. 59-60 in *al-Mishkat*. Beirut.
- al-Rifa'i, Shams al-Din. 1967. *Tareekh al-Sahafa al-Souriya wal Lubnaniya*. Paris: Asmar.
- Albert, Stuart and David A. Whetten. 1985. "Research in organizational behavior." Pp. 263-295, edited by L. L. Cummings and B. M. Staw. Greenwich: JAI Press.
- Aldrich, Howard. 2008. *Organizations and Environments*: Stanford University Press.
- Aldrich, Howard and Martin Ruef. 2006. *Organizations Evolving*. California: Sage.
- Aldrich, Howard E. and C. Marlene Fiol. 1994. "Fools Rush in? The Institutional Context of Industry Creation." *The Academy of Management Review* 19:645-670.
- Alvarez, Sharon A. and Jay B. Barney. 2005. "How Do Entrepreneurs Organize Firms Under Conditions of Uncertainty?" *Journal of Management* 31:776-793.
- Amburgey, Terry L., Dawn Kelly, and William P. Barnett. 1993. "Resetting the Clock: The Dynamics of Organizational Change and Failure." *Administrative Science Quarterly* 38:51-73.
- Ashmore, Richard D., Kay Deaux, and Tracy McLaughlin-Volpe. 2004. "An Organizing Framework for Collective Identity: Articulation and Significance of Multidimensionality." *Psychological bulletin* 130:80-114.
- Astley, W. Graham. 1985. "The Two Ecologies: Population and Community Perspectives on Organizational Evolution." *Administrative Science Quarterly* 30:224-241.

- Ayalon, Ami. 1995. *The Press in the Arab Middle East: A History: A History*: Oxford University Press, USA.
- Barnett, William and Terry Amburgey. 1990. "Do Larger Organizations Generate Stronger Competition?" Pp. 78-102 in *Organizational Evolution: New Directions*, edited by J. Singh. California: Sage.
- Barnett, William P and Glenn R Carroll. 1987. "Competition and Mutualism Among Early Telephone Companies." *Administrative Science Quarterly*:400-421.
- Barron, David N, Elizabeth West, and Michael T Hannan. 1994. "A Time to Grow and a Time to Die: Growth and Mortality of Credit Unions in New York City, 1914-1990." *American Journal of Sociology*:381-421.
- Baum, Christopher F. 2008. "Stata Tip 63: Modeling Proportions." *Stata Journal* 8:299.
- Baum, Joel AC and Heather A Haveman. 1997. "Love Thy Neighbor? Differentiation and Agglomeration in the Manhattan Hotel Industry, 1898-1990." *Administrative Science Quarterly*:304-338.
- Baum, Joel AC and Stephen J Mezias. 1992. "Localized Competition and Organizational Failure in the Manhattan Hotel Industry, 1898-1990." *Administrative Science Quarterly*:580-604.
- Baum, Joel AC and Jitendra V Singh. 1994. "Organizational Niches and the Dynamics of Organizational Mortality." *American Journal of Sociology*:346-380.
- Beck, Nikolaus, Anand Swaminathan, James B. Wade, and Filippo Carlo Wezel. 2009. "The Impact of Organizational Prototypes and Industry Clusters on the Fates of Northern Bavarian (Franconian) Breweries."
- Benford, Robert D. 1997. "An Insider's Critique of the Social Movement Framing Perspective." *Sociological Inquiry* 67:409-430.
- . 2005. "You Could be the Hundredth Monkey." *The Sociological Quarterly* 34:195-216.
- Benford, Robert D. and Scott A. Hunt. 1992. "Dramaturgy and Social Movements: The Social Construction and Communication of Power." *Sociological inquiry* 62:36-55.

- Benford, Robert D. and David A. Snow. 2000. "Framing Processes and Social Movements: An Overview and Assessment." *Annual review of sociology*:611-639.
- Berger, Peter L. and Thomas Luckmann. 1967. *The Social Construction of Reality: A Treatise in the Sociology of Knowledge*. New York: Anchor Books.
- Biggart, Nicole Woolsey and Thomas D. Beamish. 2003. "The Economic Sociology of Conventions: Habit, Custom, Practice, and Routine in Market Order." *Annual Review of Sociology* 29:443-464.
- Blau, Peter M. and W.Richard Scott. 2003. *Formal Organizations: A Comparative Approach*. California: Stanford Business Books.
- Boone, Christophe, Arjen van Witteloostuijn, and Glenn R. Carroll. 2002. "Resource Distributions and Market Partitioning: Dutch Daily Newspapers, 1968 to 1994." *American Sociological Review* 67:408-431.
- Bone, Christophe and Arjen van Witteloostuijn. 2004. "A Unified Theory of Market Partitioning: An Integration of Resource-Partitioning and Sunk Cost Theories." *Industrial and Corporate Change* 13:701-725.
- Buheiry, Marwan R. and Lawrence I. Conrad. 1989. *The Formation and Perception of the Modern Arab World*: Darwin Press.
- Buis, Maarten L, Nicholas J Cox, and Stephen P Jenkins. 2012. "BETAFIT: Stata Module to Fit a Two-Parameter Beta Distribution." *Statistical Software Components*.
- Burns, Tom and G.M. Stalker. 1994. *The Management of Innovation*: Oxford University Press.
- Caliński, Tadeusz and Jerzy Harabasz. 1974. "A Dendrite Method for Cluster Analysis." *Communications in Statistics-theory and Methods* 3:1-27.
- Carroll, Glenn R, Mi Feng, Gaël Le Mens, and David G McKendrick. 2010. *Organizational Evolution with Fuzzy Technological Formats: Tape Drive Producers in the World Market, 1951–1998*, vol. 31: Emerald Group Publishing Limited.
- Carroll, Glenn R and James Wade. 1991. "Density Dependence in the Organizational Evolution of the American Brewing Industry Across Different Levels of Analysis." *Social Science Research* 20:271-302.



- Carroll, Glenn R. 1984. "Dynamics of Publisher Succession in Newspaper Organizations." *Administrative Science Quarterly* 29:93-113.
- . 1985. "Concentration and Specialization: Dynamics of Niche Width in Populations of Organizations." *American Journal of Sociology* 90:1262-1283.
- Carroll, Glenn R., Lyda S. Bigelow, Marc-David L. Seidel, and Lucia B. Tsai. 1996. "The Fates of De Novo and De Alio Producers in the American Automobile Industry 1885–1981." *Strategic Management Journal* 17:117-137.
- Carroll, Glenn R. and Jacques Delacroix. 1982. "Organizational Mortality in the Newspaper Industries of Argentina and Ireland: An Ecological Approach." *Administrative Science Quarterly* 27:169-198.
- Carroll, Glenn R. and Michael T. Hannan. 2004. *The Demography of Corporations and Industries*: Princeton University Press.
- Carroll, Glenn R. and Anand Swaminathan. 1992. "The Organizational Ecology of Strategic Groups in the American Brewing Industry from 1975 to 1990." *Industrial and Corporate Change* 1:65-97.
- . 2000. "Why the Microbrewery Movement? Organizational Dynamics of Resource Partitioning in the U.S. Brewing Industry." *American Journal of Sociology* 106:715-762.
- Cerulo, Karen A. 1997. "Identity Construction: New Issues, New Directions." *Annual review of Sociology*:385-409.
- Cleves, Mario, William Gould, Roberto G. Gutierrez, and Yulia V. Marchenko. 2010. *An Introduction to Survival Analysis Using Stata*: Stata Press.
- Cohen, Jean L. 1985. "Strategy or Identity: New Theoretical Paradigms and Contemporary Social Movements." *Social research*:663-716.
- Dagher, Yousouf. 1978. *Kamoos al-Sahafa al-Lubnaniya*. Beirut: Al-Maktabah al-Sharqiyah.
- Damon, J. Phillips and Ezra W. Zuckerman. 2001. "Middle - Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets." *American Journal of Sociology* 107:379-429.
- Deephouse, David L. 1999. "To Be Different, or to be the Same? It's a Question (and Theory) of Strategic Balance." *Strategic Management Journal* 20:147-166.

- Delacroix, Jacques and Glenn R. Carroll. 1983. "Organizational Foundings: An Ecological Study of the Newspaper Industries of Argentina and Ireland." *Administrative Science Quarterly* 28:274-291.
- DiMaggio, Paul J. and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48:147-160.
- Dobrev, Stanislav D, Tai-Young Kim, and Glenn R Carroll. 2002. "The Evolution of Organizational Niches: US Automobile Manufacturers, 1885–1981." *Administrative Science Quarterly* 47:233-264.
- Dobrev, Stanislav D, Tai - Young Kim, and Michael T Hannan. 2001. "Dynamics of Niche Width and Resource Partitioning." *American Journal of Sociology* 106:1299-1337.
- Duda, Richard O and Peter E Hart. 1973. *Pattern Classification and Scene Analysis*, vol. 3: Wiley New York.
- Echols, Ann and Wenpin Tsai. 2005. "Niche and Performance: The Moderating Role of Network Embeddedness." *Strategic Management Journal* 26:219-238.
- Eisenstadt, Shmuel N. 1998. "The Construction of Collective Identities Some Analytical and Comparative Indications." *European Journal of Social Theory* 1:229-254.
- Evans, John H. 1997. "Multi - Organizational Fields and Social Movement Organization Frame Content: The Religious Pro - Choice Movement." *Sociological Inquiry* 67:451-469.
- Everitt, Brian S., Sabine Landau, Morven Leese, and Daniel Stahl. 2011. *Cluster Analysis*: Wiley.
- Faour, Muhammad. 1991. "The Demography of Lebanon: A Reappraisal." *Middle Eastern Studies* 27:631-641.
- Fawaz, Leila T. 1983. *Merchants and Migrants in Nineteenth-Century Beirut*: Harvard University Press.
- . 1994. *An Occasion for War: Civil Conflict in Lebanon and Damascus in 1860*: University of California Press.
- Freeman, John and Michael T Hannan. 1983. "Niche Width and the Dynamics of Organizational Populations." *American Journal of Sociology*:1116-1145.

- Ghaleb, Abd al-Rahim. 1988. *Ma'at 'am min Tareekh al-Sahafa - Lisanul Hal*. Beirut: Gros Press.
- Göçek, Fatma M. 1996. *Rise of the Bourgeoisie, Demise of Empire: Ottoman Westernization and Social Change*: Oxford University Press.
- Goffman, Erving. 1986. *Frame Analysis: An Essay on the Organization of Experience*: Northeastern University Press.
- Hamilton, Lawrence C. 2009. *Statistics with Stata: Updated for Version 10*: Brooks/Cole.
- Hannan, Michael T and John Freeman. 1977. "The Population Ecology of Organizations." *American journal of sociology*:929-964.
- . 1984. "Structural Inertia and Organizational Change." *American sociological review*:149-164.
- . 1993. *Organizational Ecology*: Harvard University Press.
- Hannan, Michael T. 2010. "Partiality of Memberships in Categories and Audiences." *Annual Review of Sociology* 36:159-181.
- Hannan, Michael T., James N. Baron, Greta Hsu, and Özgecan Koçak. 2006. "Organizational Identities and the Hazard of Change." *Industrial and Corporate Change* 15:755-784.
- Hannan, Michael T., László Pólos, and Glenn R Carroll. 2007. *Logics of Organization Theory: Audiences, Codes, and Ecologies*: Princeton University Press.
- Hardin, James W. and Joseph M. Hilbe. 2013. *Generalized Estimating Equations*: CRC Press.
- Harris, William W. 1997. *Faces of Lebanon: Sects, Wars, and Global Extensions*: Markus Wiener Publishers.
- Hayagreeva, Rao. 2004. "Institutional Activism in the Early American Automobile Industry." *Journal of Business Venturing* 19:359-384.
- Hilbe, Joseph M. 2009. *Logistic Regression Models*: CRC Press.
- Hosmer, David W. and Stanley Lemeshow. 2004. *Applied Logistic Regression*: Wiley.
- Hosmer, David W., Stanley Lemeshow, and Susanne May. 2011. *Applied Survival Analysis: Regression Modeling of Time to Event Data*: John Wiley & Sons.

- Hourani, Albert. 1957. "The Changing Face of the Fertile Crescent in the XVIIIth Century." *Studia islamica*:89-122.
- . 1983. *Arabic Thought in the Liberal Age 1798-1939*: Cambridge University Press.
- Hsu, Greta. 2006. "Jacks of All Trades and Masters of None: Audiences' Reactions to Spanning Genres in Feature Film Production." *Administrative Science Quarterly* 51:420-450.
- Hsu, Greta and Michael T. Hannan. 2005. "Identities, Genres, and Organizational Forms." *Organization Science* 16:474-490.
- Hsu, Greta, Michael T. Hannan, and Özgecan Koçak. 2009. "Multiple Category Memberships in Markets: An Integrative Theory and Two Empirical Tests." *American Sociological Review* 74:150-169.
- Hsu, Greta, Michael T. Hannan, and László Pólos. 2009. "Typecasting and Legitimation: A Formal Theory." Stanford Graduate School of Business.
- Hsu, Greta, Giacomo Negro, and Fabrizio Perretti. 2008. "Hybrids in Hollywood: A Study of Genre Spanning in the U.S. Film Industry." Davis: University of California.
- Hudson, Michael C. 1985. *The Precarious Republic: Political Modernization in Lebanon*: Westview Press.
- Illias, Joseph. 1997. *al-Sahafa al-Lubnaniya: al-Kamoos al-Musawar 1858-1958*. Beirut: Dar Al-Nidal.
- Issawi, Charles. 1982. "The Transformation of the Economic Position of the Millets in the Nineteenth Century." Pp. 261-285 in *Christians and Jews in the Ottoman Empire: The Central Lands*, vol. 1, edited by B. Braude and B. Lewis. New York: Holmes & Meier Publishers.
- . 1988. *The Fertile Crescent, 1800-1914*: Oxford University Press, USA.
- . 2006. *An Economic History of the Middle East and North Africa*: Routledge.
- Jaccard, Paul. 1901. *Etude Comparative de la Distribution Florale dans une Portion des Alpes et du Jura*: Impr. Corbaz.
- Jenkins, J. Craig. 1983. "Resource Mobilization Theory and the Study of Social Movements." *Annual review of sociology*:527-553.

- Jensen, Michael. 2010. "Legitimizing Illegitimacy: How Creating Market Identity Legitimizes Illegitimate Products." Pp. 39-80 in *Categories in Markets: Origins and Evolution, Research in the Sociology of Organizations*, edited by G. Hsu, G. Negro, and O. Kocak. UK: Emerald Group Publishing Limited.
- Karpat, Kemal H. 1985. *Ottoman Population, 1830-1914: Demographic and Social Characteristics*: University of Wisconsin Press.
- Katz, Daniel and Robert L. Kahn. 2007. *The Social Psychology of Organizations*: John Wiley & Sons, Limited.
- Kennedy, Mark Thomas. 2005. "Behind the One-way Mirror: Refraction in the Construction of Product Market Categories." *Poetics* 33:201-226.
- . 2008. "Getting Counted: Markets, Media, and Reality." *American Sociological Review* 73:270-295.
- Khaire, Mukti and R. Daniel Wadhvani. 2010. "Changing Landscapes: The Construction of Meaning and Value in a New Market Category - Modern Indian Art." *Academy of Management Journal* 53:1281-1304.
- Khalaf, Samir. 1982. "Communal Conflict in Nineteenth-Century Lebanon." Pp. 107-134 in *Christians and Jews in the Ottoman Empire: The Arabic-speaking Lands*, vol. 2, edited by B. Braude and B. Lewis. New York: Holmes & Meier Publishers.
- Khessina, Olga M. and Glenn R. Carroll. 2008. "Product Demography of De Novo and De Alio Firms in the Optical Disk Drive Industry, 1983–1999." *Organization Science* 19:25-38.
- Khoury, Yousouf. 1999. *Fahras al-Sahafa al-Arabiya*. Kuwait: Dar Suad al-Sabah.
- . 1990a. *A'mal al-Jam'iyah al-'ilmiyah al-Suriyah 1868-1869*. Beirut: Dar al-Hamra'.
- . 1990b. *Mudawinat al-Sahafa al-'Arabiyyah*. Beirut: Centre of Arabic Development.
- King, Brayden G, Elisabeth S. Clemens, and Melissa Fry. 2011. "Identity Realization and Organizational Forms: Differentiation and Consolidation of Identities Among Arizona's Charter Schools." *Organization Science* 22:554-572.
- Kocak, Ozgecan, Michael T. Hannan, and Greta Hsu. 2009. "Audience Structure and Category Systems in Markets."

- Kovacs, Balazs and Michael T. Hannan. 2009. "Category Contrast, Spanning, and Appeal: Food-Service Organizations." Stanford Graduate School of Business.
- . 2011. "Category Spanning, Distance, and Appeal." Stanford Graduate School of Business.
- Kovacs, Balazs and Rebeka Johnson. 2012. "Contrasting Alternative Explanations for the Negative Effects of Category Spanning: A Study of Restaurant Reviews and Menus in San Francisco." University of Lugano.
- Lamont, Michèle and Virág Molnár. 2002. "The Study of Boundaries in the Social Sciences." *Annual Review of Sociology* 28:167-195.
- Lawrence, Paul R., Jay W. Lorsch, and James S. Garrison. 1967. *Organization and Environment: Managing Differentiation and Integration*: Division of Research, Graduate School of Business Administration, Harvard University Boston, MA.
- Lee, Elisa T. and John W. Wang. 2003. *Statistical Methods for Survival Data Analysis*: Wiley.
- Leung, Ming D. 2010. "Dilettante or Renaissance Man? How the Sequence of Category Memberships Affects Credibility in an Online Market for Services." PhD Thesis, Graduate School of Business, Stanford University.
- Leung, Ming D. and Amanda J. Sharkey. 2009. "Out of Sight, Out of Mind? The Mere Labeling Effect of Multiple-Category Membership in Markets." Stanford Graduate School of Business.
- Liu, Min and Filippo C. Wezel. 2013. "Small is Beautiful? Organizational Identity and Growth Rates in a Partitioned Market." Working Paper. Durham Research Online (DRO), Durham.
- Lomi, Alessandro. 1995. "The Population Ecology of Organizational Founding: Location Dependence and Unobserved Heterogeneity." *Administrative Science Quarterly*:111-144.
- Lomi, Alessandro and Erik Larsen. 2001. "Failure as a Structural Concept: A Computational Perspective on Age Dependence in Organizational Mortality Rates." Pp. 269-303 in *Dynamics of Organizations*:

- Computational Modeling and Organizational Theories*, edited by A. Lomi and E. Larsen. Massachusetts: MIT Press.
- Lounsbury, Michael and Hayagreeva Rao. 2004. "Sources of Durability and Change in Market Classifications: A Study of the Reconstitution of Product Categories in the American Mutual Fund Industry, 1944–1985." *Social Forces* 82:969-999.
- Lounsbury, Michael, Marc Ventresca, and Paul M. Hirsch. 2003. "Social Movements, Field Frames and Industry Emergence: A Cultural–Political Perspective on US Recycling." *Socio-Economic Review* 1:71-104.
- Lowthian, John, Elijah G. Saleebey, and Soloman Saleebey. 1856. *Narrative and Report Regarding Lebanon Schools, Superintended by John Lowthian, Esq., of Carlton House, Carlisle ; Mr. Elijah George Saleebey, Mount Lebanon ; and His Brother Mr. Soloman Saleebey.*
- Makdisi, Ussama S. 2000. *The Culture of Sectarianism: Community, History, and Violence in Nineteenth-Century Ottoman Lebanon*: University of California Press.
- Maoz, Moshe. 1982. "Communal Conflicts in Ottoman Syria during the Reform Era: The Role of Political and Economic Factors." Pp. 91-105 in *Christians and Jews in the Ottoman Empire: The Arabic-speaking Lands*, vol. 2, edited by B. Braude and B. Lewis. New York: Holmes & Meier Publishers.
- March, James G. 1981. "Footnotes to Organizational Change." *Administrative Science Quarterly* 26:563-577.
- March, James G. and Herbert A. Simon. 1993. *Organizations*: Wiley.
- Marquis, Christopher and Michael Lounsbury. 2007. "Vive la résistance: Competing Logics and the Consolidation of US Community Banking." *Academy of Management Journal* 50:799-820.
- Masters, Bruce. 2004. *Christians and Jews in the Ottoman Arab World: The Roots of Sectarianism*: Cambridge University Press.
- McAdam, Doug and W.Richard Scott. 2005. "Organizations and Movements." Pp. 4-40 in *Social Movements and Organization Theory*.
- McCarthy, John D. and Mayer N. Zald. 1977. "Resource Mobilization and Social Movements: A Partial Theory." *American journal of sociology*:1212-1241.

- McKelvey, Bill. 1982. *Organizational Systematics: Taxonomy, Evolution, Classification*. Berkeley: University of California Press.
- McKelvey, Bill and Howard Aldrich. 1983. "Populations, Natural Selection, and Applied Organizational Science." *Administrative Science Quarterly* 28:101-128.
- McKendrick, David G. and Glenn R. Carroll. 2001. "On the Genesis of Organizational Forms: Evidence from the Market for Disk Arrays." *Organization Science* 12:661-682.
- McKendrick, David G., Jonathan Jaffee, Glenn R. Carroll, and Olga M. Khessina. 2003. "In the Bud? Disk Array Producers as a (Possibly) Emergent Organizational Form." *Administrative Science Quarterly* 48:60-93.
- McPherson, Miller. 1983. "An Ecology of Affiliation." *American Sociological Review*:519-532.
- Milligan, Glenn W. and Martha C. Cooper. 1985. "An Examination of Procedures for Determining the Number of Clusters in a Data Set." *Psychometrika* 50:159-179.
- Miner, Anne S., Terry L. Amburgey, and Timothy M. Stearns. 1990. "Interorganizational Linkages and Population Dynamics: Buffering and Transformational Shields." *Administrative Science Quarterly* 35:689-713.
- Minkoff, Debra C. 1997. "The Sequencing of Social Movements." *American Sociological Review*:779-799.
- . 1999. "Bending with the Wind: Strategic Change and Adaptation by Women's and Racial Minority Organizations 1." *American Journal of Sociology* 104:1666-1703.
- Mruwi, Adeeb. 1961. *al-Sahafa al-'arabiya: Nash'atuha wa Tatawuruha*. Beirut: Dar Maktabat Al- Hayat.
- Naqabat Muhariri al-Sahafa al-Lubnaniya. 1943. *Daleel al-Sahafa al-Lubnaniya*. Beirut.
- Nashabi, Hisham. 1981. "Sheykh 'Abd al-Qadir al-Qabbani and Thamarat al-Funun." Pp. 84-91 in *Intellectual Life in the Arab East, 1890-1939*, edited by M. R. Buheiry. Beirut: Center for Arab and Middle East Studies, American University of Beirut.



- Navis, Chad and Mary Ann Glynn. 2010. "How New Market Categories Emerge: Temporal Dynamics of Legitimacy, Identity, and Entrepreneurship in Satellite Radio, 1990–2005." *Administrative Science Quarterly* 55:439-471.
- Negro, Giacomo, Michael T. Hannan, and Hayagreeva Rao. 2010. "Categorical Contrast and Audience Appeal: Niche Width and Critical Success in Winemaking." *Industrial and Corporate Change* 19:1397-1425.
- Negro, Giacomo, Ozgecan Kocak, and Greta Hsu. 2010. "Research on Categories in the Sociology of Organizations." Pp. 3-38 in *Categories in Markets: Origins and Evolution, Research in the Sociology of Organizations*, edited by G. Hsu, O. Kocak, and G. Negro. UK: Emerald Group Publishing Limited.
- Olzak, Susan and Emily Ryo. 2007. "Organizational Diversity, Vitality and Outcomes in the Civil Rights Movement." *Social Forces* 85:1561-1591.
- Olzak, Susan and S.C.Noah Uhrig. 2001. "The Ecology of Tactical Overlap." *American Sociological Review*:694-717.
- Olzak, Susan and Elizabeth West. 1991. "Ethnic Conflict and the Rise and Fall of Ethnic Newspapers." *American Sociological Review*:458-474.
- Owen, Roger. 1993. *The Middle East in the World Economy, 1800-1914*: I.B. Tauris.
- . 2004. "The Silk-reeling Industry of Mount Lebanon, 1840-1914: A Study of the Possibilities and Limitations of Factory Production in the Periphery." Pp. 271-283 in *The Ottoman Empire and the World-Economy*, edited by H. Islamoglu-Inan. Cambridge: Cambridge University Press.
- Perretti, Fabrizio, Giacomo Negro, and Alessandro Lomi. 2008. "E Pluribus Unum: Framing, Matching, and Form Emergence in U.S. Television Broadcasting, 1940–1960." *Organization Science* 19:533-547.
- Peterson, Richard A and David G Berger. 1975. "Cycles in Symbol Production: The Case of Popular Music." *American Sociological Review*:158-173.
- Pett, Marjorie A., Nancy R. Lackey, and John J. Sullivan. 2003. *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research*: SAGE Publications.
- Pfeffer, Jeffrey. 1997. *New Directions for Organization Theory: Problems and Prospects*. USA: Oxford University Press.

- Pfeffer, Jeffrey and Gerald R. Salancik. 2003. *The External Control of Organizations: A Resource Dependence Perspective*. California: Stanford University Press.
- Pólos, László and Michael T. Hannan. 2001. "Nonmonotonicity in Theory Building with Applications to Organizational Mortality." Pp. 405-437 in *Dynamics of Organizations: Computational Modeling and Organizational Theories*, edited by A. Lomi and E. Larsen. Massachusetts: MIT Press.
- . 2002. "Reasoning With Partial Knowledge." *Sociological Methodology* 32:133-181.
- Pólos, László, Michael T. Hannan, and Glenn R. Carroll. 2002. "Foundations of a Theory of Social Forms." *Industrial and Corporate Change* 11:85-115.
- Pontikes, Elizabeth G. 2008. "Fitting in or Starting New? An Analysis of Invention, Constraint, and the Emergence of New Categories in the Software Industry." PhD Thesis Thesis, Stanford Graduate School of Business, Stanford, CA.
- . 2010. "Two Sides of the Same Coin: How Category Ambiguity Affects Multiple Audience Evaluations " in *Working Paper*: University of Chicago Booth School of Business
- Pontikes, Elizabeth G. and Michael T. Hannan. 2012. "Toward an Ecology of Market Categories." in *Research Paper Series*: Stanford Graduate School of Business.
- Porac, Joseph F, Howard Thomas, Fiona Wilson, Douglas Paton, and Alaina Kanfer. 1995. "Rivalry and the Industry Model of Scottish Knitwear Producers." *Administrative Science Quarterly*:203-227.
- Powell, Walter W. and Kurt W. Sandholtz. 2011. "Amphibious Entrepreneurs and the Emergence of New Organizational Forms." *Strategic Entrepreneurship Journal* 6:94-115.
- Pozner, Jo Ellen and Hayagreeva Rao. 2006. "Fighting a Common Foe: Enmity, Identity and Collective Strategy." *Advances in Strategic Management* 23:445-479.

- Rabe-Hesketh, Sophia and Andres Skrondal. 2012. *Multilevel and Longitudinal Modeling Using Stata Vol 2: Categorical Responses, Counts, and Survival*: Stata Press.
- Rao, Hayagreeva, Philippe Monin, and Rodolphe Durand. 2005. "Border Crossing: Bricolage and the Erosion of Categorical Boundaries in French Gastronomy." *American Sociological Review* 70:968-991.
- Rao, Hayagreeva, Calvin Morrill, and Mayer N. Zald. 2000. "Power Plays: How Social Movements and Collective Action Create New Organizational Forms." *Research in Organizational Behavior* 22:237-282.
- Riffe, Daniel, Stephen Lacy, and Frederick Fico. 1998. *Analyzing Media Messages: Using Quantitative Content Analysis in Research*: L. Erlbaum.
- Rosa, José Antonio, Joseph F. Porac, Jelena Runser-Spanjol, and Michael S. Saxon. 1999. "Sociocognitive Dynamics in a Product Market." *The Journal of Marketing* 63:64-77.
- Rosch, Elizabeth, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. 1976. "Basic Objects in Natural Categories." *Cognitive Psychology* 8:382-439.
- Ruef, Martin. 2000. "The Emergence of Organizational Forms: A Community Ecology Approach." *American Journal of Sociology* 106:658-714.
- Ruef, Martin and Kelly Patterson. 2009. "Credit and Classification: The Impact of Industry Boundaries in Nineteenth-Century America." *Administrative Science Quarterly* 54:486-520.
- Ryan, Thomas P. 2008. *Modern Regression Methods*: John Wiley & Sons.
- Sabat, Khalil. 1958. *Tareekh al-Tiba'a fee al-Sharq al-'arabi*. Egypt: Dar al-Ma'aref.
- Salibi, Kamal S. 2003. *A House of Many Mansions: The History of Lebanon Reconsidered*: I.B. Tauris.
- Santos, Filipe and Kathleen Eisenhardt. 2009. "Constructing Markets and Shaping Boundaries: Entrepreneurial Power in Nascent Fields." *The Academy of Management Journal* 52:643-671.
- Scott, W. Richard. 2000. *Institutions and Organizations*: SAGE Publications.
- Shepard, Roger N. 1987. "Toward a Universal Law of Generalization for Psychological Science." *Science* 237:1317-1323.
- Simon, Julian L. 1997. *Resampling: The New Statistics*: Resampling Stats.

- Simpson, Edward H. 1949. "Measurement of Diversity." *Nature* 163.
- Singer, Judith D. and John B. Willett. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*: Oxford University Press, USA.
- Sloan, Alfred. 1964. *My Years with General Motors*: Random House Digital, Inc.
- Smith, Edward Bishop. 2011. "Identities as Lenses: How Organizational Identity Affects Audiences' Evaluation of Organizational Performance." *Administrative Science Quarterly* 56:61-94.
- Snow, David A. and Robert D. Benford. 1988. "Ideology, Frame Resonance, and Participant Mobilization." *International Social Movement Research* 1:197-217.
- . 1992. "Master Frames and Cycles of Protest." *Frontiers in Social Movement Theory*:133-155.
- Snow, David A., E. Burke Rochford Jr, Steven K. Worden, and Robert D. Benford. 1986. "Frame Alignment Processes, Micromobilization, and Movement Participation." *American Sociological Review*:464-481.
- Snow, David A. and Sarah A. Soule. 2010. *A Primer on Social Movements*: WW Norton & Company.
- Snow, David A., Sarah A. Soule, and Hanspeter Kriesi. 2004. "The Blackwell Companion to Social Movements." Pp. 3-16: John Wiley & Sons.
- Stata Corporation. 2011a. *Stata Longitudinal-data/panel-data Reference Manual: Release 12*: Stata Press.
- . 2011b. *Stata Multivariate Statistics Reference Manual-Release 12*: Stata Press.
- Stinchcombe, Arthur L. 1965. "Social Structure and Organizations." Pp. 142-193 in *The Handbook of Organizations*, edited by J. G. March.
- Sutton, John. 1991. *Sunk Costs and Market Structure: Price Competition, Advertising and the Evolution of Concentration*. Cambridge: MIT Press.
- Swaminathan, Anand. 1998. "Entry Into New Market Segments in Mature Industries: Endogenous and Exogenous Segmentation in the US Brewing Industry." *Strategic Management Journal* 19:389-404.
- Swaminathan, Anand and Glenn R. Carroll. 1995. "Beer Brewers." Pp. 223-243 in *Organizations in Industry: Strategy, Structure and Selection*, edited by G. R. Carroll and M. T. Hannan. New York: Oxford University Press.

- Swaminathan, Anand and James B. Wade. 2001. "Social Movement Theory and the Evolution of New Organizational Forms." Pp. 286-313 in *The Entrepreneurship Dynamic*, edited by C. B. Schoonhoven and E. Romanelli. California: Stanford University Press.
- Taleb, Ahmed. 2011. *al-Dawriyat al-Filastiniya al-Sadira fi Lubnan*. Beirut.
- Tarābulṣī, Fawaz. 2007. *A History of Modern Lebanon*. New York: Pluto.
- Tarazi, Felipe. 1933. *Tareekh al-Sahafa al-'arabiya*. Beirut: Dar Sader.
- Thompson, James D. 2011. *Organizations in Action: Social Science Bases of Administrative Theory*. New Jersey: Transaction Publishers.
- Touraine, Alain. 1985. "An Introduction to the Study of Social Movements." *Social Research* 52:749-787.
- Weber, Klaus, Kathryn L Heinze, and Michaela DeSoucey. 2008. "Forage for Thought: Mobilizing Codes in the Movement for Grass-fed Meat and Dairy Products." *Administrative Science Quarterly* 53:529-567.
- Weber, Max. 2009. *The Theory Of Social And Economic Organization*. New York: Free Press.
- Wei, Zhao. 2005. "Understanding Classifications: Empirical evidence from the American and French Wine Industries." *Poetics* 33:179-200.
- Weick, Karl E. 1995. *Sensemaking in Organizations*, vol. 3. California: Sage.
- Whetten, David A. and Brayden G. King. 2008. "Rethinking the Relationship Between Reputation and Legitimacy: A Social Actor Conceptualization." *Corporate Reputation Review* 11:192-207.
- White, Harrison C. 1981. "Where Do Markets Come From?" *American Journal of Sociology* 87:517-547.
- . 2000. "Modeling Discourse In and Around Markets." *Poetics* 27:117-133.
- Wizarat al-Anba'. 1950. *Al-Jadwal al-Naqabi li Naqabat al-Sahafa al-Lubnaniya*. Beirut.
- Wry, Tyler, Michael Lounsbury, and Mary Ann Glynn. 2011. "Legitimizing Nascent Collective Identities: Coordinating Cultural Entrepreneurship." *Organization Science* 22:449-463.
- Zald, Mayer N. and John D. McCarthy. 1987. "Social Movements in an Organizational Society." New Jersey: Transaction Pub.

- Zuckerman, Ezra W, Tai - Young Kim, Kalinda Ukanwa, and James Von Rittmann. 2003. "Robust Identities or Nonentities? Typecasting in the Feature - Film Labor Market." *American Journal of Sociology* 108:1018-1073.
- Zuckerman, Ezra W. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount." *American Journal of Sociology* 104:1398-1438.
- . 2000. "Focusing the Corporate Product: Securities Analysts and De-diversification." *Administrative Science Quarterly* 45:591-619.
- . 2004. "Structural Incoherence and Stock Market Activity." *American Sociological Review* 69:405-432.
- Zuckerman, Ezra W. and Tai - Young Kim. 2003. "The Critical Trade - off: Identity Assignment and Box - office Success in the Feature Film Industry." *Industrial and Corporate Change* 12:27-67.