Methodos Series 12

Emmanuel Lazega
Tom A.B. Snijders *Editors*

Multilevel Network Analysis for the Social Sciences

Theory, Methods and Applications

Springer

Methodos Series

Methodological Prospects in the Social Sciences

Volume 12

This Book Series is devoted to examining and solving the major methodological problems social sciences are facing. Take for example the gap between empirical and theoretical research, the explanatory power of models, the relevance of multilevel analysis, the weakness of cumulative knowledge, the role of ordinary knowledge in the research process, or the place which should be reserved to “time, change and history” when explaining social facts. These problems are well known and yet they are seldom treated in depth in scientific literature because of their general nature.

So that these problems may be examined and solutions found, the series prompts and fosters the settingup of international multidisciplinary research teams, and it is work by these teams that appears in the Book Series. The series can also host books produced by a single author which follow the same objectives. Proposals for manuscripts and plans for collective books will be carefully examined.

The epistemological scope of these methodological problems is obvious and resorting to Philosophy of Science becomes a necessity. The main objective of the Series remains however the methodological solutions that can be applied to the problems in hand. Therefore the books of the Series are closely connected to the research practices.

More information about this series at http://www.springer.com/series/6279

Emmanuel Lazega • Tom A.B. Snijders
Editors

Multilevel Network Analysis for the Social Sciences

Theory, Methods and Applications

Springer

Editors
Emmanuel Lazega
Institut d'Etudes Politiques de Paris, SPC
CSO-CNRS, Paris, France

Tom A.B. Snijders
Department of Sociology
University of Groningen
Groningen, The Netherlands

Nuffield College
University of Oxford
Oxford, UK

Methodos Series
ISBN 978-3-319-24518-8 ISBN 978-3-319-24520-1 (eBook)
DOI 10.1007/978-3-319-24520-1

Library of Congress Control Number: 2015957593

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Acknowledgments

The authors and editors of this volume thank Daniel Courgeau and Robert Franck, editors of the *Methodos* series, for their encouragement, advice, and patience. We are grateful to Julien Brailly, Guillaume Favre, Anuška Ferligoj, and Malka Older for advice in the preparation of this volume. The book benefitted from intellectual developments for which Mark Tranmer's Leverhulme group was very important. Finally, we thank the European Sciences Foundation, ECRP Program, as well as the Program 'Dynamique des réseaux multiniveaux' (DYREM) of Sorbonne Paris-Cité, for financial support for the meetings that led to this publication. Many thanks to Bernadette Deelen-Mans, Stéphanie Dubois, Edith Martine, and Céline Ottenwelter for providing administrative support for these meetings.

Contents

Chapter 1
Introduction

Emmanuel Lazega and Tom A.B. Snijders

Theoretical developments and the emergence of new epistemological insights are based on interactions between old problems and new methodologies (Courgeau 2003). At least two methodologies have helped social scientists of the past two generations in overcoming the traditional divide between individualistic and holistic approaches in the social sciences: multilevel analysis and social network analysis. The purpose of this book is to provide an exploration of the diverse ways in which these two methodologies can be brought together in statistical approaches to multilevel network analysis, specifically their combination in the development of three areas: theory, techniques, and empirical applications in the social sciences. The combination of approaches opens up new avenues of research and improves the necessary management of so-called 'ecological fallacies' (Robinson 1950; Courgeau 1999, 2002, 2004, 2007) in complex systems of inequalities: for example, when looking at problems as different as school performance of pupils or career development in labor markets.

With respect to theory, this book describes the development of multilevel network reasoning by showing how it can explain behavior by insisting on two different ways of contextualizing it. The first method consists of identifying levels of influence on behavior and identifying in sophisticated ways different aggregations of actors and behaviors as well as complex interactions between levels and therefore between context and behavior. The levels in multilevel analysis refer to the different units of

E. Lazega (✉)
Institut d'Etudes Politiques de Paris, SPC, CSO-CNRS, 19 rue Amélie, 75007 Paris, France
e-mail: emmanuel.lazega@sciencespo.fr

T.A.B. Snijders
Department of Sociology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands

Nuffield College, University of Oxford, Oxford, UK
e-mail: tom.snijders@nuffield.ox.ac.uk

E. Lazega, T.A.B. Snijders (eds.), *Multilevel Network Analysis for the Social Sciences*, Methodos Series 12, DOI 10.1007/978-3-319-24520-1_1

analysis. Each level of analysis corresponds to a population, so multilevel studies will refer to several populations (Bryk and Raudenbush 1992; Goldstein 1995; Bressoux et al. 1997; Snijders and Bosker 2012). For example, Kenny and LaVoie (1985) developed a Social Relations Model for dyadic dependent variables in which groups, individuals, and dyads are the relevant units of analysis. They propose a model in which level 1 is the individual, level 2 the dyad, and level 3 the group. Similarly, for the p_2 models which were proposed for binary dyadic dependent variables, "the multilevel p_2 model can be regarded as a three-level random effects model where Level 1 is formed by the tie observations, cross-nested in the actors (Level 2), who are nested in the networks (Level 3)" (Zijlstra et al. 2006, p. 3). In the same spirit, but with a dynamic perspective, Snijders and Baerveldt (2003) developed a multilevel model for friendship networks between pupils in several classes within the same school in order to understand the respective influence of each level on deviant behavior. In these network data structures the traditional approach of multilevel analysis based on hierarchical nesting cannot be followed exactly, because the levels of dyads and actors are not nested; but non-nested structures are also accommodated in multilevel analysis more generally (Courgeau 2003; Snijders and Bosker 2012).

A second, more recent method of contextualization, consists of identifying different systems of collective agency as distinct levels of analysis, differentiating for example among levels of collective action with different goals; specific resource interdependencies between members; and specific social processes that help members manage dilemmas of collective action at each level. Individuals today are members of an organizational society (Coleman 1990; Perrow 1991) because they act in organized, if not highly regulated and bureaucratized, social and economic contexts (companies, associations, families, etc.) that influence their behavior and that they in turn can try to shape. Individuals interact with each other, but are also embedded in (or construct) groups and organizations that interact with each other. Such superposed levels of agency can be examined separately as well as jointly, since they are linked by the affiliation of members of one level to collective actors at the higher level. Affiliations can be considered as indicators of deeper processes characterizing the "duality" of individuals and groups (Breiger 1974; Brass et al. 2004; Rousseau 1985), and thus the co-constitution of these levels as the expression of their vertical interdependencies and complexity. Their superposition is not static (Courgeau and Baccaïni 1997; Lazega 2012): through actors' efforts to endogenize context at each level, they influence each other's evolution. This raises issues of synchronization in these complex dynamics, and brings up the question of how the hidden social costs of this synchronization are shared, spread, or dumped (Lazega, this volume).

Another purpose of this book is to offer new case studies and datasets that explore new avenues of theorizing and modeling, as well as new applications of this methodology. As also shown in Rozenblat and Melançon (2013), an increasing number of datasets is being made available to test the value of theoretical ideas and the efficiency of methods. Although heterogeneous with respect to units of analysis and methods, models of multilevel network analysis presented in this volume tend to take into account a variety of structural dependencies, both within and between levels. The conclusion extends theoretical, methodological and empirical results of this new epistemology by speculating on the insights provided for our knowledge of societies that have become "organizational" societies, i.e. rationalized, managerialized, and marketized.

This book thus identifies a plurality of levels, assumes that actors operate across more than one of them, and provides a bouquet of models for multilevel network datasets to account for vertical and horizontal interdependencies in social life. It shows how concepts applied to analyze single-level networks can be extended to a multilevel perspective, and in turn be extended by it. In this way, it opens and explores new avenues of research for the emerging stream of multilevel network analyses. The volume ends with a general conclusion outlining the importance, limits and perspectives of these current methodologies.

The following outline summarizes the content of the book in terms of theory, methods and applications by suggesting the way in which each chapter contributes to the exploration of structure in multilevel network analysis, from descriptive and inductive techniques to stochastic models (from network autocorrelation models to p_2 models to ERGMs), accounting for both horizontal and vertical interdependencies.

Theory

Part I of the book provides the theoretical foundation for this combined approach. In Chap. 2, Tom Snijders describes the complementarity between these approaches from a methodological perspective. By providing a sketch of multilevel models, statistical models for social network analysis, multilevel network models, and models for multilevel networks, this chapter offers a background to the methods of analysis used in this book. Multilevel analysis, in which individuals' actions, beliefs and performances within groups are analyzed taking into account their nested collective memberships (Snijders and Bosker 2012; Multilevel Network Modeling Group 2012) does not take into account the dyadic interdependencies between individuals based on their relationships or links between groups. It is not plausible that such groups lack an internal structure, nor that they lack links among each other. Network analyses help in introducing more realistic approximations of the internal structure of these groups and of their interdependencies into the modeling of human and social action. This chapter summarizes the 'multilevel' perspective in network analysis. The basis for this is the presence in networks of units of various different, interconnected kinds: individuals, ties, subgroup structures, groups, and perhaps more. These kinds of units represent populations, which can be modeled as having random variability. The fundamental idea of multilevel analysis, to explain dependent variables by models containing multiple sources of random variation and including explanatory variables defined as aggregates over higher-order units, is fruitfully applied here to network models. This approach opens room for the simultaneous study of the contributions of several levels of social phenomena through the ‘multilevel analysis of networks’. The second method of contextualization mentioned above is expressed by the ‘analysis of multilevel networks’, which considers several interconnected system of agency. Following Wasserman and Iacobucci (1991), for cross-sectional data this can be expressed by the multilevel exponential random graph modeling (ERGM) approach of Wang et al. (2013). Each ‘level’ here is a set of actors, or agents, and the levels are interdependent with respect to the conditions for action and/or outcomes. A hierarchical nesting relation between the levels, which is the traditional basis of statistical multilevel analysis, is not required for the data structure of multilevel networks.

Multilevel network analysis means analyzing separately, then jointly, several levels of collective agency. In Chap. 3, Emmanuel Lazega argues that finding structure in society is a complex task if one is to take the meso-level of society seriously. His chapter explores the sociological meaning of introducing dynamics into the study of different and superposed systems of interdependencies and collective agency. In particular, he looks at the issue of “synchronization costs” between the temporalities that characterize the different levels. These specific social costs are related to carrying out collective action in the organizational society, i.e., a society in which multilevel structures, defined as superposed levels of collective agency, make cross-level social processes increasingly visible. These processes are modeled using network analysis. Synchronization costs are associated with building and maintaining specific social forms, in particular, social status and social niches, as intermediary relational infrastructure that helps individuals and groups manage their complex multilevel interdependencies and the dilemmas of their multilevel collective action. This helps them create new corporate entities that they can try to use as “tools with a life of their own” (Selznick 1949). It is suggested that the energy for creating and managing this relational infrastructure comes from catching-up dynamics between levels, where collective actors operate in different temporalities while under pressure to coordinate and stabilize this synchronization. Catching-up dynamics are associated with organized mobility of actors and relational turnover (OMRT) in their respective networks, a perspective combining White’s (1970), Snijders’ (1996), and Snijders et al. (2013) approaches. In this context, specific dimensions of social inequalities also become visible since actors who manage these social forms are in a position to benefit from their investments in synchronization costs as they become productive –in particular in terms of reshaping their meso-level opportunity structure – whereas others are likely to see their own investments in synchronization be lost, providing no return.

Methods

This new domain of interest brings together very different innovative methods, new theorizing, and applications to a wide diversity of problems. Part II of the book presents a series of different statistical frameworks and methods articulating social network analysis and multilevel analysis.

In Chap. 4 Filip Agneessens and Johan Koskinen use multilevel network analysis to look at the impact of network position and team structure on individual outcomes. They model individual outcomes using what they call a *Multilevel Social Influence* (*MSI*) model. This model explains individual differences in behavior and attitudes by considering the (individual level) *network position*, while simultaneously looking at the influence of the (group level) *network structure*. Such an approach requires a multilevel method, where both levels are explicitly modeled. However, while the network nature of the data offers the possibility of simultaneous investigation of the impact of the network level and the individual level position, the complex network interdependence within a single network make classical multilevel modeling unsuitable. The complex interdependence of social networks makes the models more complicated, as there is a need to control for both levels as well as for social contagion and autocorrelation. Their application considers an organizational setting focusing on the importance of trust relations for employee job satisfaction. They simultaneously consider how individual differences in being trusted by colleagues (within a team) impact a person's satisfaction, while at the same time also examining how the structure of the group (density and centralization) might impact the job satisfaction of all members of the group. The multilevel network nature of the data offers the possibility of simultaneous investigation of graph-level, positional and dyadic explanations. This introduces non-standard dependencies as the networks among level 1 units imply both contextual effects different from standard multilevel effects (such as team-level means) as well as direct network dependencies, the latter called level 1½.

In Chap. 5 Mark Tranmer and Emmanuel Lazega consider models for multilevel network dependencies, where one or more attributes of the level 1 network nodes varies across the levels of the multilevel network in which they are embedded. They apply Tranmer's multilevel model called Multiple Membership Multiple Classification (MMMC) model and explains how it can be used to estimate the relative share of variation in the different components of a multilevel network. They outline the ways in which this modeling approach differs from other models that are currently used for network dependencies. They also explain how the MMMC model can be used with statistical software. The approach is illustrated with an analysis of Lazega et al. (2008)'s multilevel network data on French cancer researchers, focusing on variations in research impact scores for the workers as the motivating and illustrative example. This approach can also be applied in the context of more traditional groups such as schools (Tranmer et al. 2014).

In Chap. 6, Peng Wang, Garry Robins and Petr Matous provide a summary presentation of Multilevel Network Analysis using ERGMs and their extensions. Through the integration of vertical dependencies, exponential random graph models (ERGMs) represent network structure as endogenous based on the assumption that network ties are conditionally dependent, that is, that the existence of a network tie depends on the existence of other network ties conditioning the rest of the network (Frank and Strauss 1986; Lusher et al. 2013; Snijders et al. 2006; Robins et al. 2007). In multilevel network contexts, ERGMs offer a statistical framework that captures complicated multilevel structure through some simple structural signatures or network configurations based on these tie dependence assumptions. But for multilevel network models, network ties are interdependent not only within levels but also across levels. The interpretation of ERGM parameters makes hypothesis testing about multilevel network structure possible.

Wang et al. (2013) pioneered ERGM specifications for multilevel networks, and demonstrated the features of multilevel ERGMs with simulation studies and modeling examples. Combining multilevel network structure and nodal attributes, Wang et al. (2015) proposed Social Selection Models (SSMs) where the existence of multilevel network ties are conditionally dependent on not only the existence of other network ties but also on nodal attributes. They demonstrated that nodal attributes may affect network structures both within and across levels. After reviewing the multilevel network data structure, multilevel ERGM and SSM specifications as proposed in Wang et al. (2013, 2015), the authors apply these models to a dataset collected among 265 farmers and their communication network in a rural community in Ethiopia. The resulting model provides an informative description of this farming community. There are similarities as well as clear distinctions between the entrepreneurial farmers and the rest. Without considering the meso- and cross-level effects, we might argue that the two types of farmers have similar network behavior, i.e., both are active within their religion and region; both have flat degree distribution, and both tend to form network closures. The meso- and cross-level effects, however, show that the network is segmented by the farmer types, where popular meso-level nodes tend not to communicate within levels, but popular within-level nodes tend to communicate across levels through the meso-level network. The example highlights the features of these models and their theoretical importance, i.e. within-level network structures are interdependent with network structures of other levels; and within level nodal attributes can affect multilevel network structures.

In Chap. 7, Mengxiao Zhu, Valentina Kuskova, Stanley Wasserman, and Noshir Contractor propose a correspondence analysis of multilevel networks. The past decade has seen considerable progress in the development of p^* (also known as exponential random graph) models. Ideally, social science theory should guide the identification of parameters that map on to specific hypotheses. However, in the preponderance of cases, extant theories are not sufficiently nuanced to narrow down the selection of specific parameters. Hence there is a need for some exploratory techniques to help guide the specification of theoretically sound hypotheses. They take the example of individuals being members of work teams. Modern technologies enable individuals to self-assemble and participate in more than one team. Teams often share one or more members with other teams and hence, are not independent of each other. In addition, the assemblage of these teams is embedded in prior communication and collaboration networks. The case becomes more complicated when considering relations at both the individual level and at a combination of individual and team levels.

In order to address these issues, they propose the use of correspondence analysis, incorporating multiple relations and attributes at both individual and team levels. The descriptive analysis preempts concerns about independence assumptions. Correspondence analysis can be used as an exploratory tool to examine the features of the dataset and the relationships among variables of interest, and the results can be presented visually using a graph that shows those relationships as well as observed raw data. They present the theory for this approach, and illustrate with an example focusing on combat teams from a fantasy-based online game. The results offer important multilevel insights and show how this approach serves as a stepping stone for more focused analysis using techniques such as multilevel p*/ERGMs.

In Chap. 8, Aleš Žiberna and Emmanuel Lazega present an application of Žiberna's (2014) method of blockmodeling multilevel network data and an application of this method. The chapter presents a blockmodeling analysis of multilevel (inter-individual and inter-organizational) networks. Several approaches are presented, and used to blockmodel such networks. Each blockmodel represents a system of roles (White et al. 1976) and therefore a form of division of work that is likely to change over time in fields of organized collective action. Using a case study, they show that while the systems of roles are quite similar at both levels (structures divided into core and periphery with bridging cores interpreted in terms of division of work between actors' and organizations' specialties, location, status, etc.), the roles are performed at different levels by units with different characteristics. The added value of this true multilevel analysis is to show how groups at different levels are connected. In the empirical case analyzed in the chapter, the division of work at the level of individuals and the division of work at the level of laboratories can complement and strengthen each other in the case of some segments of the population, while this reinforcement does not occur for other segments. For the same roles, the mix of specialties at one level is different from the mix of specialties at the other level, notably because the two levels do not manage the same resources. Thus, this analysis tracks the meeting of top down and bottom up pressures towards structural alignment between levels.

Applications

Although the differentiation between the 'methods' and 'applications' sections is not clear-cut, the following chapters contain examples of applications of the different methods described in the previous part. Several social areas are covered in these rich and original analyses: multilevel networks are analyzed in scientific fields and in various industries, markets and organizations. While several authors use traditional multilevel models applied to social networks, others use the neo-structural framework with separate levels of agency expressed by analysis of multilevel networks, depending on the kind of data that are available to them.

In Chap. 9, Bellotti et al. use a multilevel approach to compare scientific fields. They model the multilevel structure of scientific work, looking at social networks of collaborations between scientists, and how these networks are embedded in disciplinary and organizational levels. The dependent variable is the success of individual scholars in Italian academia. They adopt the structural approach of

Lazega et al. (2008) and analyze the local system of public funding to academic disciplines in Italy using bipartite networks. They observe the variability of structural effects across disciplinary areas that they expect to be organized in different but comparable ways. They find an overarching importance of academic rank and of brokerage roles in obtaining research funding, together with some other interesting results, like the less impactful but still significant importance of working with an established group of long-term colleagues, and differences between sub-disciplines. The importance of adopting a multilevel perspective is indicated by the relevance of the meso-categories, which combine individual network data with organizational properties. Despite the lack of impact of macro categories (university and sub-disciplinary affiliations), results show the necessity of controlling for these various nested levels, which the analysis of individual characteristics would not be able to account for. They show that in order to be successfully funded what counts more than being a big fish (a scientist with a lot of connections) working in a big pond (a large university) is being in a brokerage position interacting over the years with different research groups.

In Chap. 10, Julien Brailly, Guillaume Favre, Josiane Chatellet and Emmanuel Lazega revisit the notion of embeddedness by looking at networks of contracts as inter-organizational networks modelled jointly with social, interpersonal networks. Economic sociology has established the interdependencies between economic and social structures using the notion of the embeddedness of the economic in the social. Since Granovetter's (1985) and White's (1981, 1988) work about the interactions between economics and social relations, economic sociologists have shown that it is important to know the social structure of a specific milieu to understand its economic structure. For example, globalized markets require long distance partnerships between companies, "global pipelines" as Bathelt and Schuldt (2008) call them. But what kind of relationships do these partnerships represent? Behind each partnership between firms there are always inter-individual ties (Gulati 1995), with their own particular history. The authors use a multilevel framework to jointly analyze the economic networks between firms and the informal networks between their members in order to reframe the embeddedness hypothesis. Based on a network study of a trade fair for television programs in Eastern Europe they show that while each level has its own specific processes they are also partly nested. Beyond this result, they observe that these levels of agency emerge in different contexts and that they are diachronically related. They show that in order to understand performance in a market one needs to look at this dual positioning of individuals and organizations.

In order to explore the complex interactions between these embedded spheres, they provide a multilevel (individual and organizational) reading of an economic market by modeling its underlying social 'meta-system'. To illustrate, they reconstruct a multilevel network in the given market. They consider two levels of action: the first approximated by an advice network between individual actors; the second measured by the contract network between the organizations to which the individuals belong. The issue is to model the global structure generated by these two levels of agency that are in part nested. To investigate this meta-system, the formalization used is that of Wang et al. (2013) developed for multilevel ERGMs.

The multilevel ERGM represents the feedback between the inter-individual social relations and the inter-organizational economic relations (structural vertical dependence hypothesis between the levels). A traditional ERGM at each level shows differences in structuration and temporality between the levels. To manage these different temporalities, organizations develop specific mechanisms of learning and knowledge transmission (represented here by affiliation links). At the same time, recent contracts and current inter-organizational negotiations constitute a specific context for the inter-individual relations (inter-organizational links). The authors show that the cross-level effects and especially the multilevel tetradic substructure (Lazega et al. 2013; Brailly and Lazega 2012) are helpful in investigating the articulation of this meta-system.

In Chap. 11, Julia Brennecke and Olaf Rank examine the relationship between organizations' embeddedness in networks of research and development (R&D) collaborations, and their managers' and researchers' interpersonal knowledge networks in the context of high-tech clusters. Complex cross-level processes are assumed to characterize the networking activities of individuals at the micro-level and their organizations at the macro-level, leading to systematic interdependencies between knowledge networks at the two levels. They apply exponential random graph models (ERGMs) for multilevel networks to data collected in two German high-tech clusters and find that micro- and macro-level knowledge networks are highly interdependent. Specifically, organizations' tendency to maintain formal R&D collaborations interacts positively with their managers' popularity as providers of knowledge but negatively with their activity of seeking knowledge from colleagues. Moreover, managers and researchers exchange knowledge at the micro level if their organizations formally collaborate and vice versa. Their findings contribute to research on the determinants of formal and informal knowledge sharing in the context of institutionalized high-tech clusters.

In Chap. 12, Guillaume Favre, Julien Brailly, Josiane Chatellet and Emmanuel Lazega look at the same process of multilevel embeddedness as that in the chapter by Brailly et al. While a social exchange may involve two persons in the two firms, a transaction involves the two companies as entities at a different level. They therefore propose to use a multilevel framework to look at these networks at different levels of agency. In particular, they study the influence of inter-organizational relationships on the formation of inter-individual relationships in a context of a trade fair. Through a multilevel analysis of a trade fair for TV programs distribution in sub-Saharan Africa they study the influence of a deal network between companies on informal information exchanges among their members. While the inter-individual relationships which exist prior to the event are strongly influenced by the organizational structure, the relationships which are created during the event do not follow that logic. A process of synchronization is observed between levels, but not in the direct context of the trade fair. They argue that trade fairs could be conceived as temporary intermediary organizations in which individuals can break free from the influence of the organization to which they are affiliated and create ties without taking into account the organizational structure. Exponential random graph models are used at each level to measure and model this mutual influence between levels.

In Chap. 13 James Hollway and Johan Koskinen provide an application to international relations. They look at why and when some states establish multilateral treaties instead of bilateral treaties. This is a consequential question for vital issues such as international fisheries management. While multilateral treaties tend to represent attempts at establishing collective fisheries management and conservation policies, bilateral treaties tend to be more geared towards gaining access to coastal fisheries resources. The nature of the ties differs, which is in line with the concept of multilevel networks, and the authors argue that there are essential dependencies between the several networks. The first, meso-level network consists of a cross-level affiliation network of state ratifications to multilateral fisheries treaties. The second, micro-level network consists of states' dyadic bilateral treaty commitments with each other. Finally, these treaties succeed each other and deal with partially overlapping issues and regions, and such treaty references express additional higher-level dependencies and give a third one-mode, macro-level network. To adequately interrogate the resulting complex, interdependent multilevel system, they argue that it is necessary to address the multiple active levels simultaneously. For this, they draw on the conceptual tool of multilevel networks (Lazega et al., 2008; see also Breiger 1974). They apply recent advances in analyzing multilevel networks using exponential random graph models (Wang et al. 2013; see also Chap. 6 by Wang et al. in this volume). They find that a relatively parsimonious model that takes the multilevel dependencies into account explains the overall structure better than one that ignores these dependencies, combining parameters estimated for each network independently. Furthermore, the structural dependencies best describing 'big fish' (high bilateral or multilateral degree states) differ from those for the 'small fish' in both 'big ponds' (multilateral treaties) and 'small ponds' (bilateral treaties). While there is a geography effect, small fish sharing a bilateral treaty has little effect on whether they also share multilateral treaties. This shows that the interaction between bilateralism and multilateralism can be fruitfully analyzed using the multilevel network paradigm. Finally, they seek to explain what drives state choice of multilateral and bilateral treaties by incorporating and modeling the relational dynamics around several nodal attributes.

In Chap. 14, Paola Zappa and Alessandro Lomi provide an application of multilevel network analysis to the process of knowledge sharing in organizations. Their research question is about the effect of mandated hierarchical relations between organizational subunits on the presence of informal network ties connecting organizational members across those subunits. They argue that the failure of prior studies to address this multilevel question leaves uncertainty about the actual role that social networks play in organizations, and, more specifically, that informal network ties connecting organizational members across the formal boundaries of organizational subunits may not be independent from the relationship of hierarchical coordination linking the subunits. They focus on boundary-crossing ties because extant research has demonstrated their direct association with a wide variety of desirable organizational outcomes. They adopt the multilevel exponential random graph models of Wang et al. (2013) to examine how formal relations among organizational subunits affect the presence of interpersonal communication and exchange of advice among members of the top management team in a multiunit organization. They show that informal interpersonal ties are sustained and shaped by the hierarchical relations linking subunits in which organizational participants are located. In particular, ties across subunits are more likely to be observed between managers working in units that are themselves connected by mandated hierarchical relations. They also show that the dependence of interpersonal relations on formal hierarchical relations is partly moderated by the tendency of interpersonal interaction to weaken or reverse the direction of hierarchical relations. Finally, they suggest that the effect of formal structure is contingent on the specific relationship that under consideration.

Acknowledgements We are grateful to Sorbonne Paris-Cite (Dyrem Programme, 2013–2016) and to the European Science Foundation for supporting the preparation of this volume.

References

Bathelt, H., & Schuldt, N. (2008). Temporary face-to-face contact and the ecologies of global and virtual buzz. *SPACES Online,* 6, 1–23.

Brailly, J., & Lazega, E. (2012). Diversité des approches de la modélisation multiniveaux en analyses de réseaux sociaux et organisationnels. *Mathématiques et Sciences Sociales, 198*, 5–32.

Brass, D. J., Galaskiewicz, J., Greve, H. R., & Tsui, W. (2004). Taking stock of networks and organizations: A multilevel perspective. *Academy of Management Journal, 47*, 795–819.

Breiger, R. L. (1974). The duality of persons and groups. *Social Forces, 53*, 181–190.

Bressoux, P., Coustère, P., & Leroy-Audouin, C. (1997). Les modèles multiniveau dans l'analyse écologique: le cas de la recherche en éducation. *Revue Française de Sociologie, 38*, 67–96.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park: Sage.

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge: Harvard University Press.

Courgeau, D. (1999). De l'intérêt des analyses multi-niveaux pour l'explication en démographie. *Population, 54*, 93–116.

Courgeau, D. (2002). Evolutions ou révolutions dans la pensée démographique ? *Mathématiques et Sciences Humaines, 160*, 49–76.

Courgeau, D. (Ed.). (2003). *Methodology and epistemology of multilevel analysis* (Collection Methodos). Dordrecht: Kluwer Academic Publishers.

Courgeau, D. (2004). *Du groupe à l'Individu: Synthèse multiniveau*. Paris: Éditions de l'INED.

Courgeau, D. (2007). *Multilevel synthesis: From the group to the individual*. Dordrecht: Springer.

Courgeau, D., & Baccaïni, B. (1997). L'Analyse multiniveaux en sciences sociales. *Population, 4*, 831–864.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association, 81*, 832–842.

Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.

Granovetter, M. S. (1985). Economic action and social structure: The problem of embeddedness. *American Sociological Review, 91*, 481–510.

Gulati, R. (1995). Does familiarity breed trust? The implications of repeated ties for contractual choices in alliances. *Academy of Management Journal, 38*(1), 85–112.

Kenny, D. A., & La Voie, L. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology, 48*, 339–348.

Lazega, E. (2012). Sociologie néo-structurale. In R. Keucheyan & G. Bronner (Eds.), *Introduction à la théorie sociale contemporaine*. Paris: Presses Universitaires de France.

Lazega, E., Jourda, M.-T., Mounier, L., & Stofer, R. (2008). Catching up with big fish in the big pond? Multi-level network analysis through linked design. *Social Networks, 30*, 157–176.

Lazega, E., Jourda, M.-T., & Mounier, L. (2013). Network lift from dual alters: Extended opportunity structures from a multilevel and structural perspective. *European Sociological Review, 29*, 1226–1238.

Lusher, D., Koskinen, J., & Robins, G. (Eds.). (2013). *Exponential random graph models for social networks: Theory, methods, and applications* (Structural analysis in the social sciences series). New York: Cambridge University Press.

Multilevel Network Modeling Group. (2012). *What are multilevel networks*. University of Manchester. Available at: http://mnmg.co.uk/Multilevel%20Networks.pdf

Perrow, C. (1991). A society of organizations. *Theory and Society, 20*, 725–762.

Robins, G. L., Snijders, T. A. B., Wang, P., Handcock, M., & Pattison, P. E. (2007). Recent developments in exponential random graph (p*) models for social networks. *Social Networks, 29*, 192–215.

Robinson, W. S. (1950). Ecological correlations and the behaviour of individuals. *American Sociological Review, 15*, 351–357.

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior, 7*(1), 1–37.

Rozenblat, C., & Melançon, G. (Eds.). (2013). *Methods for multilevel analysis and visualisation of geographical networks* (D. Courgeau, & R. Franck, Methodos series, Vol. 11). Dordrecht: Springer.

Selznick, P. (1949). *Leadership in administration*. Evanston: Row, Peterson & Co.

Snijders, T. A. B. (1996). Stochastic actor-oriented models for network change. *Journal of Mathematical Sociology, 21*, 149–172.

Snijders, T. A. B., & Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology, 27*, 123–151.

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). London: Sage.

Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. (2006). New specifications for exponential random graph models. *Sociological Methodology, 36*, 99–153.

Snijders, T. A. B., Lomi, A., & Torlò, V. (2013). A model for the multiplex dynamics of two-mode and one-mode networks, with an application to employment preference, friendship, and advice. *Social Networks, 35*, 265–276.

Tranmer, M., Steel, D., & Browne, W. J. (2014). Multiple-membership multiple-classification models for social network and group dependences. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 177*(Part 2), 439–455.

Wang, P., Robins, G. L., Pattison, P. E., & Lazega, E. (2013). Exponential random graph models for multilevel networks. *Social Networks, 35*, 96–115.

Wang, P., Robins, G., Pattison, P., & Lazega, E. (2015). Social selection models for multilevel networks, *Social Networks*, http://dx.doi.org/10.1016/j.socnet.2014.12.003

Wasserman, S., & Iacobucci, D. (1991). Statistical modeling of one-mode and two-mode networks: Simultaneous analysis of graphs and bipartite graphs. *British Journal of Mathematical and Statistical Psychology, 44*, 13–44.

White, H. C. (1970). *Chains of opportunity: System models of mobility in organizations*. Cambridge: Harvard University Press.

White, H. (1981). Where do markets come from? *American Journal of Sociology, 87*, 517–547.

White, H. (1988). Varieties of markets. In B. Wellman & S. D. Berkowitz (Eds.), *Social structures: A network approach*. Cambridge: Cambridge University Press.

White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social structure from multiple networks. I. Blockmodels of roles and positions. *American Journal of Sociology, 81*, 730–780.

Žiberna, A. (2014). Blockmodeling of multilevel networks. *Social Networks, 39*, 46–61.

Zijlstra, B., Van Duijn, M., & Snijders, T. A. B. (2006). The multilevel p2 model a random effects model for the analysis of multiple social networks. *Methodology, 2*, 42–47.

Part I
Theory

Chapter 2
The Multiple Flavours of Multilevel Issues for Networks

Tom A.B. Snijders

Away from Atomistic Approaches

It is strange that the assumption that data obtained from human respondents represent independent replications has been so pervasive in statistical models used in sociological research. Sociology, after all, is about the interdependence among individuals, and about the ways in which individuals make up larger wholes such as families, tribes, organizations, and societies. Of course we know some of the reasons for this: statistical models founded on independence assumptions are convenient and have properties that can be mathematically ascertained; surveys are a major means of getting social information and ideally are obtained from probability samples containing a lot of independent operations in obtaining respondents; and, indeed, independence assumptions may yield good first-order approximations for statistical modeling. However, as early as 1959 Coleman (1959, p. 36) made an eloquent plea for taking social structure into account in methods of data collection and analysis. Coleman writes: "Survey methods have often led to the neglect of social structure and of the relations among individuals. (...) But (...) one fact remained, a very disturbing one to the student of social organization. The *individual* remained the unit of analysis. (...) Now, very recently, this focus on the individual has shown signs of changing, with a shift to groups as the units of analysis, or to networks of relations among individuals". He goes on to discuss methods for survey data collection and for data analysis that reflect this change in perspective, away from the focus on atomistic individuals. The analysis methods he discusses

T.A.B. Snijders (✉)
Department of Sociology, University of Groningen, Grote Rozenstraat 31, 9712 TG Groningen, The Netherlands

Nuffield College, University of Oxford, Oxford, UK
e-mail: tom.snijders@nuffield.ox.ac.uk

E. Lazega, T.A.B. Snijders (eds.), *Multilevel Network Analysis for the Social Sciences*, Methodos Series 12, DOI 10.1007/978-3-319-24520-1_2

include contextual analysis, the precursor of present-day multilevel analysis, and the study of subgroups and cliques, still now of crucial importance in social network analysis. He concludes by saying that these methods "will probably represent only the initial halting steps in the development of a kind of structural research which will represent a truly sociological methodology", and mentions the promise of electronic computers.

In the past half century, since Coleman wrote these words, great advances have been made in methodologies for analyzing groups, or collectives, along with individuals; or, more generally, for simultaneously analyzing variables defined on different domains. The name 'multilevel analysis' has replaced[1] 'contextual analysis'. Great strides also have been taken in the study of relations among individuals, known now as social network analysis. Network analysis likewise treats variables defined in various different domains, such as sets of nodes and sets of node pairs, and it is concerned with groups, but by and large multilevel analysis and social network analysis have developed separately, meeting each other only incidentally. Recently, however, developments in social network analysis have led to combinations of these two strands of methodology. We are still in an early phase of the junction of multilevel analysis and social network analysis, and we may echo Coleman in saying that this book presents some 'initial halting steps' of this junction. This chapter gives an overview of some concepts and techniques that now can be seen as playing important roles in the combination of multilevel and network analysis.

Multilevel Analysis

To be able to discuss multilevel network analysis, we need to present a sketch about 'regular' multilevel analysis.

Origins

Multilevel analysis, as a collection of methods, was born from the confluence of two streams. On the one hand, sociological methodologists had been developing quite some conceptual precision for inference relating individuals to collectives, for which variables need to be combined that are defined in several different domains. On the other hand, statisticians had already extended analysis of variance and regression analysis, the general linear model, to linear models combining fixed with randomly varying coefficients.

Let me first sketch some highlights on the sociological methodology side. Lazarsfeld and Menzel (1961), in their paper *On the relation between individual*

[1] Albeit with a shift of meaning.

and collective properties—written in 1956, reprinted as Lazarsfeld and Menzel (1993)—distinguish variables according to the set of units to which scientific propositions are meant to apply. For propositions about individual and collective properties, they state that there need to be sets of units both at the individual and at the collective level. Here 'individual' may refer to individual humans, but also, e.g., individual organizations or other groupings; 'collective' refers to sets of 'individuals'. Lazarsfeld and Menzel go on to define three types of properties defined for collectives. Analytical properties are obtained by a mathematical operation performed on each member, for example the mean of an individual variable, or the correlation between two variables. Structural properties are obtained by a mathematical operation performed on the relations of each member to some or all of the other members, for example the 'cliquishness' of a network. Global properties, finally, are properties of collectives that cannot be directly deduced from properties of individual members, e.g., the type of government of a city.

As for properties of individuals, Lazarsfeld and Menzel discuss that the correlation between individual variables may be considered as a correlation between the individuals but also between the collectives, pointing to the *ecological fallacy* presented in Robinson (1950): the mistake of regarding associations between variables at one level of aggregation as evidence for associations at a different aggregation level; an extensive review was given by Alker (1969). Researchers became aware of the importance of the different levels, or sets of units, in which variables are defined, and as suggested here the focus was on nested levels, representing individuals and collectives.

During the 1970s, methods for contextual analysis were developed taking into account these levels of analysis, and trying to avoid ecological fallacies. This was called 'contextual analysis' mainly by sociologists (Blalock 1984), and 'multilevel analysis' by educational researchers (Burstein 1980).

Statisticians had a few decades earlier developed models that waited to be discovered by these social scientists. In the analysis of variance, precursor and paradigmatic example of the general linear model, models had been developed where coefficients could themselves be random variables, allowing for the investigation of multiple sources of random variation in, e.g., agricultural and industrial production. Models with only fixed, fixed as well as random, or only random coefficients were called fixed, mixed, and random models, respectively (Wilk and Kempthorne 1955; Scheffé 1959).

In the early 1980s contextual analysis and linear mixed (or generalized linear mixed) models were brought together by several statisticians and methodologists: Mason et al. (1983), Goldstein (1986), Aitkin and Longford (1986), and Raudenbush and Bryk (1986). These researchers also developed estimation algorithms and implemented them in multilevel software packages, making use of the nested structure of the random coefficients to achieve efficiency in the numerical algorithms. The scientific gains from the combination of contextual analysis and random coefficient models are also discussed by Courgeau (2003). A more extensive history of these developments is given in Kreft and de Leeuw (1998).

Hierarchical Linear Model

The prototypical statistical model used in multilevel analysis is the *Hierarchical Linear Model*, which is a mixed effects linear model for nested designs (Raudenbush and Bryk 2002; Goldstein 2011; Snijders and Bosker 2012). This generalizes the well-known linear regression model. It is meant for data structures that are hierarchically nested, such as individuals in collectives, where each individual belongs to exactly one collective. The most detailed level (individuals) is called the lowest level, or level one. The Hierarchical Linear Model is for the analysis of dependent variables at the lowest level. The basic idea is that studying the simultaneous effects of variables defined at the individual level, as well as of other variables defined at the level of collectives, on an individual-level dependent variable requires the use of regression-type models that include error terms for each of those levels separately; the Hierarchical Linear Model is a linear mixed model that has this property.

In the two-level situation—let us say, individuals in groups—it can be expressed as follows. Highlighting the distinction with regular regression models, the terminology speaks of *units* rather than cases, and there are specific types of unit at each level. We denote the level-1 units, individuals, by i and the level-2 units, groups, by j. Level-1 units are nested in level-2 units (each individual is a member of exactly one group) and the data structure is allowed to be unbalanced, such that j runs from 1 to N while i runs, for a given j, from 1 to n_j. The basic two-level hierarchical linear model can be expressed as

$$Y_{ij} \;=\; \beta_0 \;+\; \sum_{h=1}^{r} \beta_h\, x_{hij} \;+\; U_{0j} \;+\; \sum_{h=1}^{p} U_{hj}\, z_{hij} \;+\; R_{ij}\;. \tag{2.1}$$

Here Y_{ij} is the dependent variable, defined for level-1 unit i within level-2 unit j; the variables x_{hij} and z_{hij} are the explanatory variables. Some or all of them may be defined at the group level, rendering superfluous the index i for such variables. Variables R_{ij} are residual terms, or error terms, at level 1, while U_{hj} for $h = 0, \ldots, p$ are residual terms, or error terms, at level 2. In the case $p = 0$ this is called a *random intercept model*, for $p \geq 1$ it is called a *random slope model*. The usual assumption is that all R_{ij} and all vectors $U_j = (U_{0j}, \ldots, U_{pj})$ are independent, R_{ij} having a normal $\mathscr{N}(0, \sigma^2)$ and U_j having a multivariate normal $\mathscr{N}_{p+1}(\mathbf{0}, \mathbf{T})$ distribution. Parameters β_h are regression coefficients (fixed effects), while the U_{hj} are random effects. The presence of both of these makes (2.1) into a linear mixed model. Similar models can be defined for nesting structures with more than two levels, e.g., employees in departments in firms.

In most practical cases, the variables with random effects are a subset of the variables with fixed effects ($x_{hij} = z_{hij}$ for $h \leq p$; $p \leq r$). The Hierarchical Linear Model can then be expressed in the appealing form

$$Y_{ij} \;=\; (\beta_0 + U_{0j}) \;+\; \sum_{h=1}^{p} (\beta_h + U_{hj})\, x_{hij} \;+\; \sum_{h=p+1}^{r} \beta_h\, x_{hij} \;+\; R_{ij}\;, \tag{2.2a}$$

which shows that it can be regarded as a regression model defined for the groups separately, with group-specific intercept

$$(\beta_0 + U_{0j}) \tag{2.2b}$$

and group-specific regression coefficients

$$(\beta_h + U_{hj}) \tag{2.2c}$$

for $h = 1, \ldots, p$; variables X_h for $p + 1 \le h \le r$ have regression coefficients that are constant across groups. This pictures the Hierarchical Linear Model as a linear regression model defined by the same model for all groups, but with regression coefficients that differ randomly between groups.

Going back to the teachings of Lazarsfeld and Menzel, it can be concluded that multilevel analysis elaborates the inference about individual and collective properties as a system of nested samples drawn from nested populations: a population of individuals nested in a population of groups (or collectives). The fact that, in practice, groups will be finite, whereas the populations are mathematically considered as if they were infinite, is usually glossed over in research aiming to generalize to social mechanisms or processes (as distinct from descriptive survey research about concrete groups, without the aim of generalization to other groups) (see Cox 1990; Sterba 2009).

Non-nested Data Structures

It soon transpired that the relevant data structures are not always nested, because social structures often are not. A basic example in studies of school effectiveness is that neighborhoods may also be an important factor for student achievement, but schools will have students coming from diverse neighborhoods while neighborhoods will have students attending different schools. This leads to a data set where students are nested in schools and also nested in neighborhoods, but schools and neighborhoods are not nested in each other; the term used for non-nested category systems is ‘crossed’, so that this would be called a cross-nested data structure. To present an extension of model (2.1) for such a cross-nested data structure, consider again a data structure with individuals i nested in groups j but now also nested in aggregates k of a different kind (in the example of the previous sentence, neighbourhoods). Denote by $k(i,j)$ the aggregate k to which individual i in group j belongs. In the simplest extension there is only a random intercept V_k associated with k, leading to the equation

$$Y_{ij} = \beta_0 + \sum_{h=1}^{r} \beta_h x_{hij} + U_{0j} + \sum_{h=1}^{p} U_{hj} z_{hij} + V_{k(i,j)} + R_{ij} \,. \tag{2.3}$$

The default assumption for the V_k is that again they are independent and normally distributed with mean 0 and constant variance, and independent of the U and R variables. A further extension is to mixed-membership models (Browne et al. 2001), in which individuals may be partial members of more than one group.

Frequentist and Bayesian Estimation

Multilevel models such as (2.2), in which parameters vary randomly between groups, provide a natural bridge between the frequentist paradigm in statistics, which treats parameters as fixed quantities which are unknown, 'out there', and the Bayesian paradigm, which treats parameters as random variables; in both paradigms, of course, the observations are the material that helps us get a grip on the values of the parameters. In the multilevel case, the random variation of parameters can be linked to a frequency distribution of parameters in the population of groups, which may be estimated from empirical data. Accordingly, this bridging ground is often called *empirical Bayes* (see, e.g., Raudenbush and Bryk 2002, and Chapter 5 of Gelman et al. 2014). Bayesian estimators[2] for the parameters such as (2.2a) and (2.2b), using the sample of groups to get information about the corresponding population, are called *empirical Bayes estimators*. For the parameters β, σ^2, and $\mathbf{T}$ in (2.1), frequentist as well as Bayesian estimators have been developed.

Especially for non-nested data structures, Bayesian estimators may have algorithmic advantages, and Bayesian Markov chain Monte Carlo ('MCMC') algorithms are often employed (Draper 2008; Rasbash and Browne 2008) for such more complex models. These are algorithms which use computer simulations, very flexible but also much more time-consuming than traditional algorithms. Today, Bayesian methods for multilevel analysis are often proposed and used without much attention paid to the distinct philosophical underpinnings. This lack of attention does not, however, take away the differences. The Bayesian approach can be a useful way to account for prior knowledge; this is discussed for the special case of multilevel analysis by Greenland (2000), and elaborated more practically in Chapter 5 of Gelman et al. (2014). Using this approach requires, however, that one pays attention to the sensitivity of the results to the choice of the prior distribution. In addition there are interpretational differences, but these may be less important because of the convergence between frequentist and Bayesian approaches discussed in Gelman et al. (2014, Chapter 4).

[2]In frequentist terminology these are not called estimators but predictors, because they refer to statistics that have the purpose to approximate random variables.

What Is a Level?

The various extensions of the basic multilevel model have made even more pressing the question '*What is a level?*' which has harrowed quite a few researchers even in the case of the more basic nested models. The mathematical answer is that, for applications of linear mixed or generalized linear mixed models, a level is a system of categories for which it is reasonable to assume random effects. More elaborately, this means that we assume that the categories j on which the variables U_j are defined (which are latent variables in model (2.1)) may be regarded has having been sampled randomly from some universe or population $\mathscr{G}$, making the U_j into independent and identically distributed random variables, and our aim is to say something about the properties of the population $\mathscr{G}$ rather than about the individual values U_j of the units in our sample. In the case that the U_j are one-dimensional quantities, the property of interest concerning population $\mathscr{G}$ could be, e.g., the variance of U_j. In practical statistical modeling, the assumption that the units in the data were randomly sampled from the population is usually taken with a grain of salt (again cf. Cox 1990; Sterba 2009). The essential assumption is *residual exchangeability*, which can be described as follows. The random effects, R_{ij} and U_j in (2.1) and also V_k in (2.3), are residuals given that the explanatory variables x_{hij} are accounted for; these residuals are assumed to be exchangeable across i and j (or k) in the sense that they are random and as far as we know we have no *a priori* information to distinguish them for different units in the data. Any R_{ij} could be high or low just as well as any $R_{i'j}$ in the same group j or any $R_{i'j'}$ in a different group j'; any U_{0j} could be high or low just as well as any other $U_{0j'}$; etc.

In this sense, multilevel analysis is a methodology for research questions and data structures that involve several sources of unexplained variation, contrasting with regression analysis which considers only one source of unexplained variation. Employing the Hierarchical Linear Model, as in (2.1) or its variants with additional levels, gives the possibility of studying contextual effects on the individual units. But also in more complex structures where nesting is incomplete, random effects will reflect multiple sources of unexplained variation. In social science applications this can be fruitfully applied to research questions in which different types of *actor* and *context* are involved; e.g., patients, doctors, hospitals, and insurance companies in health-related research; or students, teachers, schools, and neighborhoods in educational research. The word 'level' then is used for a type of unit, or a category system, for which a random effect is assumed. The basic phenomenon we are studying will be at the most detailed level (patients or students, respectively), and the other levels may contribute to the variation in this phenomenon, e.g., as contexts or other actors.

Lazarsfeld and Menzel (1961, first page) mentioned that, to be specific about the intended meaning of variables, we should 'examine (them) in the context of the propositions in which they are used'. This focus on propositions also sheds light on the question about what can be meaningfully considered as a '*level*' in multilevel analysis. We have to distinguish between the individual level, which is the level of the phenomena we wish to explain, the population of units for which the dependent variable is defined; and higher, collective levels, which do not need to be mutually nested, but in which the individuals are nested. To be a level requires, in the first place, that the category system is a population—a meaningfully delimited set of units with a basic similarity and for which several properties may be considered, such as a well-defined set of schools, of companies, of meetings. A category system then is a meaningful higher level if it is a population that we wish to use to explain[3] some of the variability in our phenomenon and also, potentially or actually, we may be interested in finding out which properties of the categories/units explain the variability associated with this category system.

To illustrate this, suppose we are interested in the phenomenon of juvenile delinquency as our dependent variable, and we consider neighborhoods as collectives. The individual level is, e.g., a set of adolescents living in a certain area at a certain time point; the dependent variable is their delinquency as measured by some instrument. We may observe that neighborhoods differ in average juvenile delinquency, and we then may wonder about the properties of neighborhoods—perhaps neighborhood disorder, of which a measurement may be available—that are relevant in this respect. This step, entertaining the possibility that there might be specific properties of neighborhoods associated with their influence on juvenile delinquency, and analyzing this statistically, is the step that makes the neighborhood a meaningful 'level' in the sense of multilevel analysis. In the paradigm of multilevel analysis we will then further assume that in addition to the effect of disorder there may be other neighborhood effects, but conditional on the extent of disorder and perhaps other neighborhood properties that we take into account, the neighborhoods are exchangeable (as far as we know) in their further, residual, effects.

The fact that we are interested in statistically analyzing the effect of the categories on the dependent variable also implies that for a level to be meaningful in a practical investigation, the total number of its units in the data set should be sufficiently large: a statistical analysis based on a sample of, say, less than 10 units usually makes no sense.

Dependent Variables at Any Level

The Hierarchical Linear Model is considered a model for dependent variables at the lowest level of the nesting hierarchy. However, it is so amazingly flexible that it can just as well be used for complex configurations of multiple dependent variables defined for several different levels. This was proposed, quite casually, already by

[3] 'Explaining' is meant here in the simple statistical sense, without considering deeper questions of causality.