Innovative Approaches to Investigating Social Determinants of Health - Social Networks, Environmental Effects and Intersectionality

Innovative Approaches to Investigating Social Determinants of Health – Social Networks, Environmental Effects and Intersectionality

Clare Rosenfeld Evans

A Dissertation Submitted to the Faculty of
The Harvard T.H. Chan School of Public Health
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Science
in the Department of Social and Behavioral Sciences

Harvard University
Boston, Massachusetts

November 2015

For David Evans,
With love.

Dissertation Advisor: Dr. S.V. Subramanian Clare Rosenfeld Evans

Innovative Approaches to Investigating Social Determinants of Health – Social Networks, Environmental Effects and Intersectionality

ABSTRACT

Contexts are important social determinants of individual health trajectories and population level patterns of health disparities. This dissertation examines three types of contexts—social networks, physical environments, and social positions—using innovative quantitative approaches. Chapter 1 examines the intersectional social positions created by interlocking social identities—race/ethnicity, sex, income, education, and age—and their relationship to health disparities in the obesity epidemic. We outline an innovative analytic approach to evaluating intersectionality using multilevel models. After adjustment for the contributions of the main effects, a large intersectional effect remains. While clear social patterning emerges, interactions are not necessarily patterned according to 'multiple jeopardy' and 'multiplicative benefit' as might have been expected. These findings reveal the complex social patterning of the obesity epidemic, and challenge us to consider possible refinements to intersectionality theory.

Chapter 2 evaluates whether U.S. adolescent social networks are segregated by family income level. Network segregation or integration may affect adolescent health trajectories through a variety of pathways, yet the extent to which networks are socioeconomically segregated is poorly understood. We approach the evaluation of income segregation through a novel lens by explicitly considering three scales of analysis within social networks: the network community level, the dyadic level, and a level in between. We find evidence of income

segregation at all three levels, though this segregation is neither extreme nor universal. Family income appears to be a socially salient factor in the structure of adolescent social networks.

In Chapter 3, three contexts of relevance to the adolescent obesity epidemic—schools, neighborhoods, and social networks—are examined simultaneously. Using a novel combination of social network community detection and cross-classified multilevel modeling, we compare the contributions of each of these contexts to the total variation in adolescent body mass index. After adjusting for relevant covariates, we find that the school-level and neighborhood-level contributions to the variance are modest compared with the network community-level. These results are robust to multiple sensitivity tests. This study highlights the salience of adolescent social networks and indicates that they may be a promising context to address in the design of health promotion programs.

TABLE OF CONTENTS

	TOPIC	PAGE
I.	Dissertation Introduction	1
II.	Chapter 1: Obesity and Body Mass Index Disparities at the Intersection of Multiple Social Categories: Evaluating the Embodiment of Intersectionality and Interactions	10
	Abstract	11
	Introduction	12
	Data and Methods	16
	Results	22
	Discussion	37
	References	42
III.	Chapter 2: Social Network Segregation by Family Income Level in U.S. Adolescent Friendship Networks	49
	Abstract	51
	Introduction	52
	Methods	61
	Analysis	63
	Results	69
	Discussion	80
	References	88
	Appendix A: Chapter 2 Supplemental Tables	96
IV.	Chapter 3: Multiple Contexts and Adolescent Body Mass Index: Schools, Neighborhoods, and Social Networks	108
	Abstract	110
	Introduction	111
	Methods	116
	Analysis	120
	Results	125
	Discussion	132
	References	138
	Appendix B: Chapter 3 Supplemental Tables	148

FIGURES WITH CAPTIONS

FIGURE NUMBER	FIGURE TITLE	PAGE
1.1	Strata-Level Residuals from Models of BMI, and the Location of Particular Strata Groups in the Distributions in Wave 2	32
1.2	Strata-Level Residuals from Models of Obesity, and the Location of Particular Strata Groups in the Distributions in Wave 2	33
1.3	Strata-Level Predictions of Mean BMI obtained using Multilevel approach versus OLS approach	35
1.4	Strata-Level Predictions of Proportion Obese obtained using Multilevel approach versus OLS approach	36
2.1	Cartoon Schematics of Levels Evaluated in Network Analysis	55
2.2	Comparison of Family Income Distributions Across Saturated Sample Schools	71
2.3	Visualizations of Detected Communities in One School Network Using Two Methods – K-clique Percolation and Modularity Maximization	72
3.1	Schematics Illustrating Data Structures in Full and Saturated Samples	123
3.2	Visualizations of detected communities in social networks using two algorithms – Modularity Maximization and K-clique Percolation	127

TABLES WITH CAPTIONS

TABLE NUMBER	TABLE TITLE	PAGE
1.1	Descriptive Characteristics of Sample	23
1.2	Number and Percent of Strata of a Given Sample Size (out of 384 possible strata)	24
1.3	MCMC Parameter Estimates for the Two-Level Hierarchical Bayesian Linear Regression Model of Body Mass Index (kg/m^2) in Wave 2	26
1.4	MCMC Parameter Estimates for the Two-Level Hierarchical Bayesian Logistic Regression Model of Obesity (BMI $\geq$ 30) in Wave 2	27
1.5	Strata that occupy the tails of the BMI residual distribution	29
1.6	Strata that occupy the tails of the Obesity residual distribution	30
2.1	Demographic Profile of Sample	70
2.2	Random Effects from Multilevel Models of Students nested in Network Communities	73
2.3	Mean Difference in Income for Dyads Within versus Dyads Between Network Communities – Modularity Maximization	76
2.4	Mean Difference in Income for Dyads Within versus Dyads Between Network Communities – K-clique Percolation	77
2.5	Dyadic Level Income Segregation – Logit Models of Difference in Income Between Dyad Pairs and 'Same Race' Indicator Predicting Probability of a Tie Existing	79
2.6	Summary of Findings	81
3.1	Sample Demographics	126
3.2	Multilevel Data Structure of Full and Saturated Samples	126
3.3	Full Sample – Random Effects Results from Single-Level and Two-Level Multilevel Models, Modularity Maximization Detection	128
3.4	Full Sample – Random Effects Results from Three-Level and Cross-Classified Multilevel Models, Modularity Maximization Detection	129

ACKNOWLEDGEMENTS

This dissertation was made possible because of the collaboration and support of many people. In particular, I would like to thank my advisor, S.V. Subramanian, for the insightful, thought-provoking questions he posed and the encouragement he offered throughout the development of my dissertation. I also owe a special debt of gratitude to the members of my dissertation committee, David Williams and Jukka-Pekka Onnela. Chapter 1 of this dissertation was inspired by a challenge posed by David Williams to his class one day early in my doctoral studies. His encouragement and shared expertise were invaluable to my professional development. Chapter 2 was only made possible through the tireless efforts of Jukka-Pekka Onnela to bring the latest in network science to the Harvard community. His instruction, mentorship, and enthusiasm encouraged me to always delve deeper and consider research questions from many angles.

A special thanks is also due to the entire community at the Harvard T.H. Chan School of Public Health, for providing me with so many remarkable opportunities. In particular, I would like to thank Cassandra Okechukwu, Erin Dunn, Tracy Richmond, and Adam Lippert for their collaboration, mentorship and encouragement throughout my doctoral studies. My friends and colleagues, Ying Chen, Mischa Haider, and Carly Milliren were always there for me when I needed support and a friendly audience for thinking through new ideas. Thanks also to the innumerable professors, colleagues, and cohort-members who enriched my time here and who taught me so much.

Finally, I would like to thank my family for their support and understanding while I moved across the country to pursue my dreams. In particular, I am grateful to my husband David, whose love and generosity lightened the load of doctoral studies and dissertation writing considerably.

DISSERTATION INTRODUCTION

Contexts come in many, varied forms, and the term can refer to *environmental contexts*, such as schools and neighborhoods, to *social contexts* such as social networks, and to *social positions*, which I define as locations within a larger social, political, and economic context. All three categories of context are socially defined, constructed, and perpetuated, and all three fall within the scope of what social epidemiologists and medical sociologists call the *social determinants of health*. Social determinants of health are often referred to as the *causes of causes* of both individual health trajectories and the social patterning of population health disparities (Berkman and Kawachi 2000). This is because social determinants—particularly social contexts—shape the health-promoting and health-harming behavior choices of individuals and populations, both by constraining the choices individuals can make, and by influencing their attitudes, beliefs, and preferences towards certain behaviors. Social epidemiology encourages us to look upstream in the causal pathways that result in poor health and health disparities, to ask: "*what is the root cause?*" Understanding how social contexts are structured and how population health varies across different contexts is of central concern to the field. This dissertation addresses contexts in all three forms, and seeks to contribute substantively to our understanding of them through the application of novel analytic approaches.

In its topical focus and the theoretical traditions from which it draws, this dissertation is explicitly situated within both social epidemiology and medical sociology. Chapter 1 draws on intersectionality theory, which posits that inequalities are generated by numerous interlocking systems of oppression such as racism, classism, and sexism (Collins 1990, Crenshaw 1989, McCall 2005, Nash 2008), to approach the evaluation of the relationships between social positions and health disparities in the obesity epidemic. While social disparities in the obesity epidemic have been widely documented along racial/ethnic, socioeconomic, gender, and age lines (Clarke et al. 2009, McLaren 2007, Ogden et al. 2012, Sobal and Stunkard 1989), the extent to which these multiple dimensions interact with each other to produce patterns of health disparities is still poorly understood. Are there, for instance, particular intersections of social identity, such as among black females of low socioeconomic status, where individuals experience multiple jeopardy to their health, or a burden above and beyond what might be expected based on the additive effects of each dimension of social identity? Conversely, are there multiply advantaged populations, such as white males of high socioeconomic status, who enjoy a multiplicative benefit, or additional protections to their health above and beyond what might have been expected? Resolving this issue is critical to understanding the true social patterning of the obesity epidemic across advantaged and disadvantaged social positions in society.

The limitations of currently available quantitative methods, including model inefficiency and lack of interpretability, have led intersectionality researchers to call for innovations in current methods (Bowleg 2012, Dubrow 2008, Hum and Simpson 2003, McCall 2005, Veenstra 2011, Veenstra 2013). In addition to contributing substantively to our understanding of the intersectionality of the obesity epidemic, Chapter 1 outlines a novel quantitative analysis approach to studying intersectionality. Briefly, this approach involves applying a two-level, hierarchical multilevel model (Raudenbush and Bryk 2002) that nests individuals (level 1) within their social positions (level 2), as defined by the intersectional identities of their race/ethnicity, sex, income, education and age. This study broadens our vision for the potential uses of multilevel models across a range of research questions in public health and the social sciences.

Chapter 2 addresses the structuring of adolescent social networks by socioeconomic status. Understanding adolescent network segregation by socioeconomic status is vital due to the implications of segregation (or integration) for the health, educational trajectories, and well being of adolescents. Social network segregation or integration by family income level could affect adolescent health outcomes through numerous pathways, yet the extent to which adolescent social networks are segregated is largely an unknown. This study therefore seeks to contribute substantively to our ability to characterize adolescent network segregation by family income level.

Care is taken to consider the multiple scales at which segregation may exist within a social network—ranging from the social network community (or social clique) level to the dyadic (or pair) level, and at points in-between. Scales other than the dyadic level are rarely considered explicitly in network segregation research (for one exception see: González et al. 2007), and often findings from the dyadic level are assumed to represent findings at other levels as well (Burgess, Sanderson and Umana-Aponte 2011, Cohen 1979, Maharaj and Connolly 1994, Mouw and Entwisle 2006). Yet this assumption may not be a valid one. For instance, some individuals may enjoy diversity in their general social group (i.e., the network community or social clique to which they belong), but will preferentially nominate individuals who are more similar to them when identifying their closest friends at the dyadic level. Therefore there is no guarantee that segregation at the dyadic level implies segregation at the network community level as well. This Chapter explicitly deals with the issue of scale, and applies relatively new methods from the field of network science—algorithms for network community detection (Fortunato 2010, Porter, Onnela and Mucha 2009)—in order to more fully characterize clustering by family income level within adolescent networks.

Chapter 3 addresses the issue of adolescents living and interacting within multiple contexts *simultaneously*. Social networks, schools and neighborhoods have all been implicated as potentially relevant contexts in shaping the adolescent obesity epidemic (Fletcher, Bonell and Sorhaindo 2011, Richmond and Subramanian 2008, Richmond et al. 2015, Townsend, Rutter and Foster 2012). Yet these contexts are very often studied in isolation. Resent methodological innovations in cross-classified multilevel modeling (CCMM) (Rasbash and Goldstein 1994) have enabled researchers to begin to compare the *relative* and *simultaneous* contributions of schools and neighborhoods to variation in health behaviors and outcomes.

The missing piece thus far has been social networks. Addressing this gap in our current knowledge is critical for two reasons. First, omitting potentially relevant contexts from analyses, particularly those using CCMM, may result in *omitted context bias*, or the attribution of variance associated with the omitted level to the included level or levels (Dunn et al. 2015, Meyers and Beretvas 2006). Second, understanding the relative contributions of each context to clustering of obesity status would perhaps enable researchers and policy makers to more effectively target interventions and policies to address heath inequalities. In this Chapter, network community detection algorithms and cross-classified multilevel modeling are combined in a novel way in order to compare the contributions of each of three contexts—schools, neighborhoods, and social networks—to the total variation in adolescent body mass index (BMI). In addition to contributing to our substantive understanding of the contributions of these three contexts in shaping the adolescent obesity epidemic, Chapter 3 outlines this novel analytic approach.

While each Chapter of this dissertation stands independently, they are unified by two themes. The first is a substantive focus on *contexts*—how they are structured and how health disparities are patterned according to context. The second theme is *methodological innovation* in social determinants research. Studying the social determinants of health from a quantitative perspective requires sophisticated methods, yet often the limitations of existing methods pose considerable challenges to researchers and to the advancement of the field. With this dissertation I seek to add to our methodological repertoire for the study of the social determinants of health.

REFERENCES

Berkman, LF and I Kawachi. 2000. "A Historical Framework for Social Epidemiology." in *Social Epidemiology*, edited by L. Berkman and I. Kawachi. New York, NY: Oxford University Press, Inc.

Bowleg, L. 2012. "The Problem with the Phrase Women and Minorities: Intersectionality— an Important Theoretical Framework for Public Health." *American Journal of Public Health* 102(7):1267-73.

Burgess, S, E Sanderson and M Umana-Aponte. 2011. “School Ties: An Analysis of Homophily in an Adolescent Friendship Network. Working Paper No. 11/267.” The Centre for Market and Public Organisation, Bristol Institute of Public Affairs, University of Bristol.

Clarke, P, PM O'Malley, LD Johnston and JE Schulenberg. 2009. "Social Disparities in Bmi Trajectories across Adulthood by Gender, Race/Ethnicity and Lifetime Socio-Economic Position: 1986-2004." *International Journal of Epidemiology* 38(2):499-509.

Cohen, J. 1979. "Socio-Economic Status and High-School Friendship Choice: Elmtown's Youth Revisited." *Social Networks* 2(1):65-74.

Collins, P. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

Crenshaw, K. 1989. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics." *University of Chicago Legal Forum*.

Dubrow, JK. 2008. "How Can We Account for Intersectionality in Quantitative Analysis of Survey Data? Empirical Illustration for Central and Eastern Europe." *ASK: Society, Research, Methods* 17:85-100.

Dunn, EC, TK Richmond, CE Milliren and SV Subramanian. 2015. "Using Cross-Classified Multilevel Models to Disentangle School and Neighborhood Effects: An Example Focusing on Smoking Behaviors among Adolescents in the United States." *Health and Place* 31:224-32.

Fletcher, A, C Bonell and A Sorhaindo. 2011. "You Are What Your Friends Eat: Systematic Review of Social Network Analyses of Young People's Eating Behaviors and Bodyweight." *Journal of Epidemiology and Community Health* 2011(65):6.

Fortunato, S. 2010. "Community Detection in Graphs." *Physics Reports* 486:75-174.

González, MC, HJ Herrmann, J Kertész and T Vicsek. 2007. "Community Structure and Ethnic Preferences in School Friendship Networks." *Physica A* 379(1):307-16.

Hum, D and W Simpson. 2003. "Labour Market Training of New Canadians and Limitations to the Intersectionality Framework." *Canadian Ethnic Studies/Etudes Ethniques au Canada* 35(3):56-69.

Maharaj, SI and JA Connolly. 1994. "Peer Network Composition of Acculturated and Ethnoculturally-Affiliated Adolescents in a Multicultural Setting." *Journal of Adolescent Research* 9(2):218-40.

McCall, L. 2005. "The Complexity of Intersectionality." *Signs* 30(3):1771-800.

McLaren, L. 2007. "Socioeconomic Status and Obesity." *Epidemiologic Reviews* 29(1):29-48.

Meyers, JL and SN Beretvas. 2006. "The Impact of Inappropriate Modeling of Cross-Classified Data Structures." *Multivariate Behavioral Research* 41(4):473-97.

Mouw, T and B Entwisle. 2006. "Residential Segregation and Interracial Friendships in Schools." *American Journal of Sociology* 112(2):394-441.

Nash, JC. 2008. "Re-Thinking Intersectionality." *Feminist Review* 89:1-15.

Ogden, CL, MD Carroll, BK Kit and KM Flegal. 2012. "Prevalence of Obesity in the United States, 2009-2010." *NCHS Data Brief 2012*:1-8.

Porter, MA, JP Onnela and PJ Mucha. 2009. "Communities in Networks." *Notices of the American Mathematical Society* 56(9):1082-97.

Rasbash, J and H Goldstein. 1994. "Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model." *Journal of Educational and Behavioral Statistics* 19(4):337-50.

Raudenbush, SW and AS Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.

Richmond, TK and SV Subramanian. 2008. "School Level Contextual Factors Are Associated with Weight Status of Adolescent Males and Females." *Obesity (Silver Spring)* 16:1324-30.

Richmond, TK, EC Dunn, CE Milliren, CR Evans and SV Subramanian. 2015. “Disentangling Overlapping Influences of Neighborhoods and Schools on Adolescent Body Mass Index.” Manuscript submitted for publication.

Sobal, J and AJ Stunkard. 1989. "Socioeconomic Status and Obesity: A Review of the Literature." *Psychological Bulletin* 105(2):260-75.

Townsend, N., H. Rutter and C. Foster. 2012. "Age Differences in the Association of Childhood Obesity with Area-Level and School-Level Deprivation: Cross-Classified Multilevel Analysis of Cross-Sectional Data." *International Journal of Obesity* 36(1):45-52. doi: 10.1038/ijo.2011.191.

Veenstra, G. 2011. "Race, Gender, Class, and Sexual Orientation: Intersecting Axes of Inequality and Self-Rated Health in Canada." *International Journal for Equity in Health* 10(1):3.

Veenstra, G. 2013. "Race, Gender, Class, Sexuality (Rgcs) and Hypertension." *Social Science and Medicine* 89:16-24.

Obesity and Body Mass Index Disparities at the Intersection of Multiple Social Categories: Evaluating the Embodiment of Intersectionality and Interactions

Clare Rosenfeld Evans[1]

David R. Williams[1]

Jukka-Pekka Onnela[2]

SV Subramanian[1]

[1] Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, MA

[2] Department of Biostatistics, Harvard School of Public Health, Boston, MA

Acknowledgements and Credits: We wish to thank the reference librarians of the Harvard University Countway Library of Medicine for their assistance in performing a systematic review of the literature.

Funding: No direct funding supported this study.

Keywords: Intersectionality, inequality, multilevel model, social determinants, social position, obesity, body mass index

ABSTRACT

Social disparities in the obesity epidemic along racial/ethnic, socioeconomic, gender, and age lines reflect a social patterning generated by numerous interlocking systems of oppression, including racism, classism, and sexism. However, the extent to which social categories interact is poorly understood. We outline an innovative analytic approach to evaluating intersectionality using multilevel models, and apply it to examine patterns of body mass index and obesity in the adult U.S. non-incarcerated population using Wave 2 of the National Epidemiologic Survey on Alcohol and Related Conditions ($N = 32{,}788$). After adjustment for the contributions of the main effects, 35% of between-strata variability for BMI, and 22% for obesity, remains unaccounted for, indicating a potentially large interaction effect. Many strata (22.8%) experience an interaction effect comparable to being in the highest income level relative to the lowest. While clear social patterning emerges, interactions are not necessarily patterned according to 'multiple jeopardy' and 'multiplicative benefit.'

Chapter 1: Obesity and Body Mass Index Disparities at the Intersection of Multiple Social Categories: Evaluating the Embodiment of Intersectionality and Interactions

Introduction

The obesity epidemic is a significant concern for medical sociologists and epidemiologists due both to obesity's numerous comorbidities (Ferraro and Kelley-Moore 2003, National Institute of Health 1998), and to the increased likelihood that overweight and obese individuals will experience weight-based discrimination and stigma (Carr and Friedman 2005, Carr and Friedman 2006, Hebl and Xu 2001). Social disparities in the obesity epidemic have been documented along racial/ethnic, socioeconomic, gender, and age lines (Clarke et al. 2009, McLaren 2007, Ogden et al. 2012, Sobal and Stunkard 1989) reflecting the expected social patterning of health status across society (Ben-Shlomo and Kuh 2002, Berkman and Kawachi 2000, Boardman et al. 2001, Boardman et al. 2005, Krieger 2011, Link and Phelan 1995, Mirowsky and Ross 2003, Schulz et al. 2000, Seng et al. 2012). Yet the extent to which these dimensions of social identity, status and position intersect and interact is poorly understood. Simply presenting the prevalence of obesity at each point of intersection, as in Flegal et al (2010), illustrates the social patterning of obesity in absolute terms, but it is unclear whether particular intersections (for instance, white females of age 60 years or more) are exhibiting the prevalence of obesity we might *expect* based on the additive effects of their sex, race and age, or whether they are more burdened or resilient than might have been expected. While these interactions can be, and often are (e.g., Ailshire and House 2011, Veenstra 2013), accounted for in regression models through the use of multiple interaction terms, such models are inefficient and the results are difficult to interpret when numerous dimensions of social identity, status and position are evaluated simultaneously. For instance, inclusion of five widely evaluated dimensions—sex, race/ethnicity, education, income, and age—can require the inclusion of hundreds of first-order, second-order, and higher-order interaction terms. This can create substantial interpretation problems. Additionally, insufficient samples at each point of intersection may result in unstable estimates. Yet failing to simultaneously address these numerous dimensions may cause us to miss important aspects of obesity disparities (Bowleg 2012, Chang and Lauderdale 2005, Davey Smith 2000, Farmer and Ferraro 2005, Schnittker 2004, Schulz and Mullings 2006, Warner and Brown 2011). We propose an innovative approach to evaluating the intersection of multiple social positions—multilevel (hierarchical) modeling of individuals nested within intersectional social identities—and apply it to examine patterns of body mass index (BMI) and obesity in the United States.

This study takes an explicitly intersectional approach to the study of health disparities. Intersectionality theorists argue that inequalities are generated by numerous interlocking systems of oppression such as racism, classism, and sexism (Collins 1990, Crenshaw 1989, McCall 2005, Nash 2008), and push back against the "additive approach" (Bowleg 2008), which treats the advantages or disadvantages conferred through simultaneous occupation of multiple social positions as simply accumulated. Instead, proponents argue that the intersection of multiple social dimensions creates unique social positions—which we refer to as *social strata*—with their own sets of societal expectations, stereotypes, opportunities, disadvantages, and sources of resilience (Kang and Bodenhausen 2014). From a complex interplay between structure and action (Bourdieu 1984), these unique social experiences become embodied through a multitude of intermediary pathways (Krieger 1994, Krieger 2011), and result in the intersectional patterning of health statuses and behaviors (Brown, O'Rand and Adkins 2012, Rosenfield 2012, Schulz and Mullings 2006, Seng et al. 2012, Veenstra 2013, Warner and Brown 2011). This intersectionality is detectable in a quantitative, fixed effects framework using interaction terms in regression models. The observed social patterning of obesity disparities along socioeconomic, racial/ethnic, sex and gender, and age dimensions (Clarke et al. 2009, Flegal et al. 2010, McLaren 2007, Ogden et al. 2012, Sobal and Stunkard 1989) indicates that taking an explicitly intersectional approach will likely deepen our understanding of obesity disparities.

The disadvantages of the fixed effects (FE) approach, however, are clear and often discussed in the literature (Bowleg 2012, Dubrow 2008, Hum and Simpson 2003, McCall 2005, Veenstra 2011, Veenstra 2013). First, as the number of main effect predictors increases, the number of interaction terms required increases geometrically, and most studies are forced to limit the scope of their analyses due to insufficient sample size in many strata. Second, partitioning the total intersectional effect for a particular stratum into many interaction parameters will generally reduce the magnitude of each parameter—thus reducing the likelihood of detecting an interaction. Third, even in cases where fitting such unwieldy models is possible, identifying and interpreting overarching patterns in the results limits the usefulness of such analyses. In the words of one prominent intersectionality scholar: "it is nearly impossible to publish grandly intersectional studies…using the categorical approach [wherein multiple categories are considered and compared]: the size and complexity of such a project is too great to contain in a single article" (McCall 2005). We tend to limit the visualizations we use when communicating the multidimensional nature of inequalities to two or three dimensions (e.g., sex *and* race *and* education) (Clarke et al. 2009, Ogden et al. 2012, Pamuk et al. 1998) for precisely this reason—because illustrations with more dimensions than that are usually unintelligible.

In this study we propose an innovative approach to quantitative intersectionality research—the application of multilevel models to evaluate the intersectionality of social strata.

While this approach has its own limitations and should certainly not be considered as a "free lunch" that addresses all limitations of currently available methods, we will argue that it does provide a more efficient and stable estimating procedure, and more importantly that it enables significantly improved interpretability of results.

Generally speaking, multilevel (or hierarchical) models partition the residual variation in a model into within-group and between-group variation (Raudenbush and Bryk 2002). Typically multilevel models are used when we wish to model the clustering of subjects by some observable clustering unit (e.g., children clustered by neighborhood) or when we wish to compensate for artifacts of the data collection processes (e.g., cluster-based sampling). However, multilevel models are capable of handling a more abstract type of clustering. Statistically relevant clustering occurs when the clustered units share something that creates similarity between them and ignoring this clustering would violate the regression assumption of independence. While the clustered individuals may share something concrete—like a neighborhood—they may also share something abstract, like a common set of social exposures. In other words, they occupy the same social stratum as defined by their sex, race, education, and so on.

The major objectives of this study are as follows. First, to evaluate the extent to which variation between social strata with respect to body mass index and obesity is attributable to intersectionality—above and beyond the additive effects. Second, to identify those strata that exhibit the greatest degree of intersectionality (i.e., deviate the most from what is predicted based on the additive effects). The notion of "multiple jeopardy" has been at the heart of intersectionality since the beginning, with its emphasis on the experiences of black women of low socioeconomic status (Collins 1990, McCall 2005, Nash 2008). Yet the question of whether white males of high socioeconomic status, for instance, enjoy a "multiplicative benefit" that counterpoints multiple jeopardy has yet to be resolved. We define "multiplicative benefit" as a beneficial effect that goes above and beyond the additive benefits of occupying multiple advantaged positions. This "unresolved theoretical dispute makes it unclear whether intersectionality is a theory of marginalized subjectivity or a generalized theory of identity" (Nash 2008:10). Therefore, the third objective of this study is to determine whether multiply disadvantaged strata exhibit signs of "multiple jeopardy" while multiply advantaged strata enjoy a "multiplicative benefit."

DATA AND METHODS

Data

The National Epidemiologic Survey on Alcohol and Related Conditions (NESARC) is a longitudinal study that was launched in 2001 by the National Institute on Alcohol Abuse and Alcoholism (NIAAA). It was designed to include a representative sample of the United States non-incarcerated civilian population, including citizens and non-citizens, aged 18 years and older who are residing in the United States and the District of Columbia. Its purpose was to determine the magnitude of population-level trends in alcohol-related disorders and to estimate the magnitude of health disparities. In this study, only data from Wave 2 of NESARC is used. For details about NESARC Wave 2, refer to Grant and Kaplan (2005). Wave 2 NESARC data was collected between 2004 and 2005, and the data is currently de-identified and publicly available.

The large sample size and intentional oversampling of young adults, Hispanics, and non-Hispanic blacks, by NESARC means that the data is sufficiently diverse to provide reasonable sample size within most evaluated social strata. In a smaller data set, or one that was less diverse, it would be difficult—despite the advantages of the multilevel modeling approach—to estimate strata effects.

Outcomes: Body Mass Index and Obesity

Body mass index (BMI), calculated as body weight in kilograms divided by height in meters squared, is commonly used to classify individuals as underweight, normal weight, overweight, and obese. In this study both BMI and obesity (BMI $\geq$ 30 kg/m^2) are evaluated; BMI is treated as continuous and obesity as dichotomous [1 = yes, 0 = no]. In Wave 2 of NESARC, the respondents' height and weight were elicited through self-report.

Social Dimensions and Strata

The social dimensions used to construct both the main effect predictors and the social strata were selected on the basis of being commonly used social dimensions in intersectionality and obesity research—*sex, race/ethnicity, education, income*, and *age*. All of these demographic variables were self-reported. Missing demographic data were imputed by the Census Bureau, and are included in the publicly available data set. For a detailed description of imputation methods used, refer to Grant et al (2003). Whenever possible, missing values were assigned using information provided by the respondent elsewhere in the surveys. Values were also imputed using information derived from other responding households with a variety of similar or matching characteristics.

It is vital to acknowledge from the outset that the categorizations used in this study undoubtedly overlook a significant amount of intracategorical and intercategorical complexity (McCall 2005, Warner 2008). This is particularly true with respect to gender, sexuality and race/ethnicity (Campbell and Troyer 2007, Frank, Akresh and Lu 2010, Jordan-Young 2010, Monk 2014a, Monk 2014b, Pfeffer 2014, Villarreal 2010, Worthen 2013). However, this data and our reliance on these admittedly rough and sometimes arbitrary classifications does enable us to examine in the social patterning and intersectionality of the obesity epidemic across a large and nationally representative sample of the U.S. population.

Sex. Interviewers who conducted the survey were instructed to ask the respondent what their sex was, if the sex of the respondent was "*not apparent*." Respondents were given the option of Male or Female.

Race/Ethnicity. Respondents were asked, "*Are you of Hispanic or Latino origin*?" and were instructed to "*select 1 or more categories to describe your race*" with the options: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, and White. According to the NESARC Data Notes, where more than one race or ethnicity was reported, a Census Bureau algorithm to code a single racial category was used. When more than one classification was selected by the individual, a single racial/ethnic category was selected from all those chosen in the following order of preference: (1) Hispanic or Latino, all races, (2) black or African American, (3) American Indian and Alaska Native, (4) Native Hawaiian and Other Pacific Islander, (5) Asian, (6) white. Only three racial/ethnic categories are included in this analysis—(1) white, not Hispanic or Latino, (2) black, not Hispanic or Latino, and (3) Hispanic or Latino—because other categories had insufficient sample size for this analysis. For simplicity we refer to these categories using the shorthand white, black and Hispanic.

Education. Education was measured in Wave 1 using self-report with a single item: "*What is the highest grade or year of school that you completed?*" and this information was updated (if necessary) with information collected in Wave 2. Fourteen response categories were provided by the survey; for the purposes of this analysis these categories were condensed into four: (1) *less than high school* (11 years or less), (2) *completed high school* (12 years plus degree or equivalency such as GED), (3) *some college no degree*, (4) *college degree or more* (including bachelor's degree, completed associate's or other technical 2-year degree, or graduate or professional studies). These categories were selected for two reasons: first, they reflect socially significant distinctions in educational achievement, and second, it partitioned the full sample into roughly equivalent groups.

Income. In Wave 2 of NESARC, annual family income was measured using the following item: "*During the last 12 months, what was YOUR TOTAL COMBINED FAMILY income received from jobs, businesses, and ALL OTHER SOURCES WE JUST TALKED ABOUT? Include ONLY immediate family members living in this household and report income before taxes and other deductions or net income after business expenses for self-employed family members. Include any tips, bonuses, overtime pay or commissions*." Income sources included food stamps. If an answer was not provided, a second question designed to provide income ranges was asked: "*Can you tell me which category on this card best represents YOUR TOTAL COMBINED FAMILY income in the last 12 months?*" Twenty-one response categories ranging from 'Less than $5,000' to '$200,000 or more' were provided as options.

In this study, annual family income was standardized by the number of related persons living in the household (including the respondent and children). For simplicity, we refer to this henceforth simply as *income*. The twenty-one income categories were condensed into four categories based on percent of the poverty threshold in 2000 (U.S. Census Bureau 2012)—low-income (below 100%), low-middle-income (100% to 199%), high-middle-income (200% to 399%), and high-income (400% or more). Cutoff values were derived based on estimates for a single person under the age of 65.

Age. Age was self-reported and recorded as ranging from 18 to 90 or older. Though measured continuously, we partition age into four categories—(1) 18 to 29 years, (2) 30 to 44 years, (3) 45 to 59 years, (4) 60 years and older. These categories were determined based on two factors—a need to distinguish socially meaningful age categories balanced against the benefits of distributing the sample roughly equally between the four categories.

Analysis

The proposed model is a two-level hierarchical model with individuals (level 1) clustered within social strata (level 2). Each social stratum is assigned a unique identifying number and a separate stratum is defined for every combination of social categories considered.

In order to control for the 'additive' aspects of the social categories considered, main effect predictors such as 'female sex' are included in the model as fixed effects. It is critical to note that, contrary to most multilevel models where social categories such as race and income are individual-level covariates, here such covariates are properties of the strata level. *Critically, no interaction terms between these predictors are included*. The stratum-level residual for each stratum therefore encompass the entirety of the interaction effect, and it is therefore possible to determine for a given stratum how much the prevalence of obesity, for instance, differs from what was 'expected' based on the contributions of the main effects alone. The strata-level residuals are assumed to be normally distributed, and it is the strata-level random effect (RE) variance term that characterizes the extent of intersectionality across all social strata. A large between-strata variance would imply that—for most strata—the main effects do an inadequate job of capturing the embodiment of the outcome of interest. On the other hand, a small between-strata variance would imply that main effects do a relatively good job and that few (if any) strata demonstrate intersectionality.

In its most general form, the linear model is specified as:

$$Y_{ij} = \beta\gamma_j + \mu_{0j} + e_{0ij}$$

$$\text{Level 2: } [\mu_{0j}] \sim N(0, \sigma^2_{strata})$$

$$\text{Level 1: } [e_{0ij}] \sim N(0, \sigma^2_{e0})$$

where Y_{ij} is the value of the outcome for respondent i in stratum j. γ_j is a vector of the main effect predictors for respondent i and β is a row vector of associated parameter values. The difference between the average value of the outcome in stratum j and the expected value of Y based on the main effects is given by the strata-level residual μ_{0j}, which is normally distributed with mean 0 and variance σ^2_{strata}. The difference between the value of the outcome for respondent i in stratum j and his or her stratum's average is given by e_{0ij}, which is also normally distributed with mean 0 and variance σ^2_{e0}. The two key parameters estimated by this model are the random effects parameters—σ^2_{strata} (the between-strata variance) and σ^2_{e0} (the within-strata variance). Note that a logistic version of this model, as required for the obesity outcome, does not estimate the within-strata variance.

In the empirical analysis, two multilevel models are fit for each outcome—a null model and a main effects model (which includes the main effect predictors). From these two models are derived the key estimates of interest. In the null model the between-strata variance term (σ^2_{strata}) represents the *total* amount of variability between strata, including that which is contributed by the main effects. In the main effects model, the same parameter is what remains after the main effects are adjusted for—the amount of between-strata variability potentially attributable to intersectionality. Thus dividing the strata-level variance term from the main effects model by the strata-level variance term from the null model provides an estimate of the *proportion of the between-strata variability that is potentially attributable to intersectionality* (assuming all relevant main effect parameters have been accounted for) for each outcome. Linear models are fit for the continuous BMI outcome, while logistic models are fit for the dichotomous obesity outcome.

In this analysis, the five social dimensions considered have the following number of categories each: sex (2), race/ethnicity (3), education (4), income (4), and age (4). This means that—treating all of the FE predictors as categorical—there are 12 FE predictors included in the model. Note that all predictors are dummy variables [0 = no, 1 = yes]. The individuals at level 1 are nested within the strata (N = 384) at level 2.

For the purposes of comparison, a more traditional model using ordinary least squares (OLS) estimation is also fit for each outcome. These OLS models include a full compliment of fixed effect dummy variables—one for each stratum. The OLS model and main effects multilevel model were compared based on Bayesian Information Criterion (BIC), a standard criterion measure which takes into account both model fit and parsimony, to empirically evaluate which approach is more efficient. The strata-specific values of mean BMI and percent obese were obtained from both models and graphically compared in order to determine whether the multilevel model provides more stable estimates (i.e., the OLS approach does not adequately address estimation for strata with very small sample size).

RESULTS

From the initial sample size of N = 34,653 in Wave 2 of NESARC, 330 respondents were excluded on the basis of missing responses to the height and weight items necessary for calculating BMI and obesity status, and a further 1,535 were excluded because they were not classified in the three race/ethnicity categories considered. The final sample size analyzed includes 32,788 respondents. The demographic profile of the sample is provided in Table 1.1. The sample is predominantly white non-Hispanic (61%) and female (58%), and is fairly well distributed across the various education, income, and age categories.

Table 1.1. Descriptive Characteristics of Sample

	Frequency *N*	**BMI** Mean (Std)	**% Obese** BMI ≥ 30
TOTAL	32788	27.90 (6.04)	29.49
Sex			
Male	13840	27.96 (5.11)	27.16
Female	18948	27.85 (6.64)	31.20
Race / Ethnicity			
White Non-Hispanic	19955	27.26 (5.78)	25.49
Black Non-Hispanic	6526	29.50 (6.67)	40.58
Hispanic / Latino	6307	28.25 (5.83)	30.70
Education			
Less than high school	5243	28.40 (6.35)	33.09
Completed high school	9038	28.41 (6.29)	33.07
Some college no degree	7043	27.93 (6.07)	29.99
College degree or more	11464	27.24 (5.61)	24.72
Income (% Poverty Threshold in 2000)			
Low income (Below 100%)	7666	28.57 (6.89)	33.89
Low-middle income (100% to 199%)	9144	28.01 (6.06)	31.20
High-middle income (200% to 399%)	9548	27.71 (5.67)	27.72
High income (400% or more)	6430	27.20 (5.35)	24.45
Age			
18 to 29 years	4628	26.96 (6.02)	24.68
30 to 44 years	9975	28.06 (6.18)	30.45
45 to 59 years	9148	28.63 (6.20)	33.52
60+ years	9037	27.44 (5.63)	26.83
US Region			
South	12485	27.92 (6.08)	29.80
Northeast	5757	27.90 (6.09)	29.29
Midwest	6190	28.05 (6.01)	30.58
West	8356	27.74 (5.97)	28.37

Note: Estimates are unadjusted for disproportionate sampling and therefore represent sample estimates (unadjusted sample proportions), not population estimates.

The raw values for mean BMI and percent obese by the various demographics (Table 1.1) are as expected. Higher education and income levels experience lower raw mean BMI scores and a smaller percentage of obese individuals; Hispanics and non-Hispanic blacks are more likely to be obese than non-Hispanic whites, and females are more likely to be obese than males. Age effects are also discernible, though not as clearly patterned.

The number of observations per stratum is important to consider in this analysis. While the multilevel approach accounts for the number of observations in strata by adjusting the estimates for strata means (estimates for strata with small sample size are pulled towards the global mean, and thus the residuals are pulled towards 0), it is still necessary to ensure that a sufficient proportion of strata have a reasonable number of observations. In this case, the sample size is sufficiently large and well distributed across the strata to prevent significant underestimation of the between-strata variability. Of the 384 strata considered in this analysis, 382 have at least one observation, and most have reasonable sample sizes (Table 1.2); 81% of strata have 20 or more respondents. Predictably, those strata with the fewest observations represent particularly unlikely combinations, such as those with both low education and high income.

Table 1.2. Number and Percent of Strata of a Given Sample Size (out of 384 possible strata)

Sample Size Per Strata	Frequency of Strata *N* (%)
100 or More	106 (27.60%)
50 or More	200 (52.08%)
30 or More	265 (69.01%)
20 or More	311 (80.99%)
10 or More	347 (90.36%)

All multilevel analyses were conducted in MLwiN version 2.26 (Rasbash et al. 2012) using Bayesian Markov Chain Monte Carlo (MCMC) estimation procedures (Browne 2009). The regression models were first fit using IGLS estimation to provide the Bayesian MCMC procedure with initialization values; Non-informative priors were used in all analyses. The main effect estimates did not differ significantly between the IGLS and MCMC procedures. The OLS version of the model was fit using IGLS.

The results of all multilevel models are presented in Table 1.3 (BMI) and Table 1.4 (obesity). It is first important to note that BMI is patterned as expected in the population. Comparing the highest to the lowest, increased education and income are associated with lower BMI. Compared with 18 to 29 year olds, older age groups experience higher BMI on average, but particularly those between 30 and 59 years. Minorities experience higher expected BMI scores than whites and females experience higher scores than males. The logistic models of obesity, which predict the probability of being obese, follow the same general patterns. It is critical to note that for logistic multilevel models, however, all higher-level predictors (such as those included in the present model) are not interpretable as Odds Ratios (ORs); instead they ought to be converted to median ORs (Larsen and Merlo 2005).

In the linear BMI null model, we see that the total between-strata RE (1.823) is significant, and after adjusting for the main effects of sex, race, education, income and age there remains considerable unexplained between-strata variability (0.643). Approximately 35% of between-strata variability is unexplained by the main effects. A similar result is obtained from the logistic models of obesity; The between-strata RE (0.179) from the null model is reduced when the main effects are included (0.039), but nearly 22% of the between-strata variability remains unaccounted for. The magnitude of the strata-level residuals in many cases is considerable. For instance, the effect on predicted BMI of belonging to a strata that is of the highest income level (compared with the lowest) is -0.584, and 22.8% of the 382 strata evaluated experience a strata-level residual of this magnitude or larger.

Table 1.3. MCMC Parameter Estimates for the Two-Level Hierarchical Bayesian Linear Regression Model of Body Mass Index (kg/m^2) in Wave 2

	Null Model Estimate (95% CI)	**Main Effects Model** Estimate (95% CI)
Fixed Effects		
Intercept	28.126 (27.965 , 28.293)	26.858 (26.433 , 27.288)
Sex		
Male (reference)		—
Female		0.081 (-0.149 , 0.316)
Race / Ethnicity		
White Non-Hispanic (reference)		—
Black Non-Hispanic		1.791 (1.511 , 2.066)
Hispanic / Latino		0.659 (0.383 , 0.941)
Education		
Less than high school (reference)		—
Completed high school		0.087 (-0.255 , 0.433)
Some college no degree		-0.240 (-0.591 , 0.123)
College degree or more		-0.813 (-1.167 , -0.460)
Income (% Poverty Threshold in 2000)		
Low income (Below 100%) (reference)		—
Low-middle income (100% to 199%)		-0.066 (-0.370 , 0.245)
High-middle income (200% to 399%)		-0.258 (-0.574 , 0.060)
High income (400% or more)		-0.584 (-0.953 , -0.210)
Age		
18 to 29 years (reference)		—
30 to 44 years		1.282 (0.944 , 1.624)
45 to 59 years		1.814 (1.477 , 2.152)
60+ years		0.523 (0.184 , 0.862)
Random Effects		
Strata	1.823 (1.503 , 2.196)	0.643 (0.488 , 0.826)
Individual	34.506 (33.984 , 35.035)	34.511 (33.977 , 35.047)
Percent of Between-Strata Variation Attributable to Intersectionality		**35.272 %**

Note: 95% Credible Intervals in parentheses. *P*-values are associated with frequentist approaches and are not available for Bayesian estimations.

Table 1.4. MCMC Parameter Estimates for the Two-Level Hierarchical Bayesian Logistic Regression Model of Obesity (BMI ≥ 30) in Wave 2

	Null Model	**Main Effects Model**
	Estimate (95% CI)	Estimate (95% CI)
Fixed Effects		
Intercept	-0.841 (-0.896 , -0.788)	-1.265 (-1.407 , -1.138)
Sex		
Male (reference)		—
Female		0.193 (0.122 , 0.264)
Race / Ethnicity		
White Non-Hispanic (reference)		—
Black Non-Hispanic		0.578 (0.495 , 0.662)
Hispanic / Latino		0.151 (0.064 , 0.238)
Education		
Less than high school (reference)		—
Completed high school		0.012 (-0.096 , 0.118)
Some college no degree		-0.104 (-0.217 , 0.007)
College degree or more		-0.322 (-0.435 , -0.213)
Income (% Poverty Threshold in 2000)		
Low income (Below 100%) (reference)		—
Low-middle income (100% to 199%)		0.037 (-0.055 , 0.134)
High-middle income (200% to 399%)		-0.075 (-0.173 , 0.025)
High income (400% or more)		-0.189 (-0.306 , -0.071)
Age		
18 to 29 years (reference)		—
30 to 44 years		0.359 (0.257 , 0.466)
45 to 59 years		0.514 (0.412 , 0.622)
60+ years		0.130 (0.023 , 0.240)
Random Effects		
Strata	0.179 (0.144 , 0.221)	0.039 (0.025 , 0.057)
Individual	—	—
Percent of Between-Strata Variation Attributable to Intersectionality		**21.788 %**

Note: 95% Credible Intervals in parentheses. *P*-values are associated with frequentist approaches and are not available for Bayesian estimations. Individual-level random effects are not estimated in logistic multilevel models.

We proceed now to consider those strata that are furthest from what is expected based on the contributions of the main effects. Table 1.5 identifies the ten strata that occupy the most extreme positions in each tail of the BMI residual distribution, while Table 1.6 identifies the ten strata in each of the extreme tails of the obesity residual distribution. While there is considerable overlap, there are some differences in rankings between the two outcomes. The top ten list for those experiencing lower BMI than expected is dominated primarily by white females and black males. On the other hand, six of the top ten experiencing higher BMI than expected are black female strata. Similar patterns are observed for obesity, however it is now white males (often of higher SES) that have a significantly higher proportion obese than expected. While these top ten lists are interesting for identifying the most extreme cases, it is more interesting to evaluate the general patterning that occurs across strata in the residual distribution.

Table 1.5. Strata that occupy the tails of the BMI residual distribution

Rank	Residual	Sex	Race	Education	Income	Age	*N* in Stratum
Lower BMI Than Predicted by Main Effects *							
1	-1.9588	Female	Hispanic or Latino	College degree or more	High	30 to 44	110
2	-1.8894	Female	White Non-Hispanic	College degree or more	High	30 to 44	442
3	-1.8040	Male	Black Non-Hispanic	Completed high school	Low	45 to 59	58
4	-1.7404	Male	Black Non-Hispanic	Less than high school	Low	60 +	112
5	-1.3322	Male	Black Non-Hispanic	Some college no degree	Low	45 to 59	23
6	-1.2813	Female	White Non-Hispanic	Some college no degree	High-middle	30 to 44	237
7	-1.2310	Female	White Non-Hispanic	College degree or more	High	18 to 29	114
8	-1.2304	Female	White Non-Hispanic	College degree or more	High-middle	30 to 44	598
9	-1.1814	Male	Black Non-Hispanic	Completed high school	Low-middle	45 to 59	70
10	-1.1795	Female	Hispanic or Latino	Some college no degree	Low	18 to 29	88
Higher BMI Than Predicted by Main Effects *							
373	1.1007	Male	White Non-Hispanic	Less than high school	High-middle	45 to 59	59
374	1.1164	Female	Hispanic or Latino	Less than high school	Low	60 +	217
375	1.1177	Female	Black Non-Hispanic	Some college no degree	Low	45 to 59	78
376	1.2115	Female	Black Non-Hispanic	Less than high school	Low	18 to 29	54
377	1.2138	Male	Hispanic or Latino	Completed high school	High-middle	30 to 44	53
378	1.2807	Female	Black Non-Hispanic	College degree or more	Low-middle	45 to 59	81
379	1.3067	Female	White Non-Hispanic	Less than high school	Low	45 to 59	68
380	1.3748	Female	Black Non-Hispanic	Completed high school	Low	30 to 44	223
381	1.5266	Female	Black Non-Hispanic	College degree or more	Low	30 to 44	119
382	2.2711	Female	Black Non-Hispanic	Completed high school	Low	45 to 59	149

* The top ten strata from each of the tails of the residual distribution (i.e., those with residuals of the greatest magnitude, positive and negative). Those strata furthest from what is predicted by the main effects are those that experience significantly higher or lower BMI on average than predicted by the main effects of the two-level multilevel model. Rank is out of 382 strata, with 1 representing the most negative residual, and 382 the most positive.

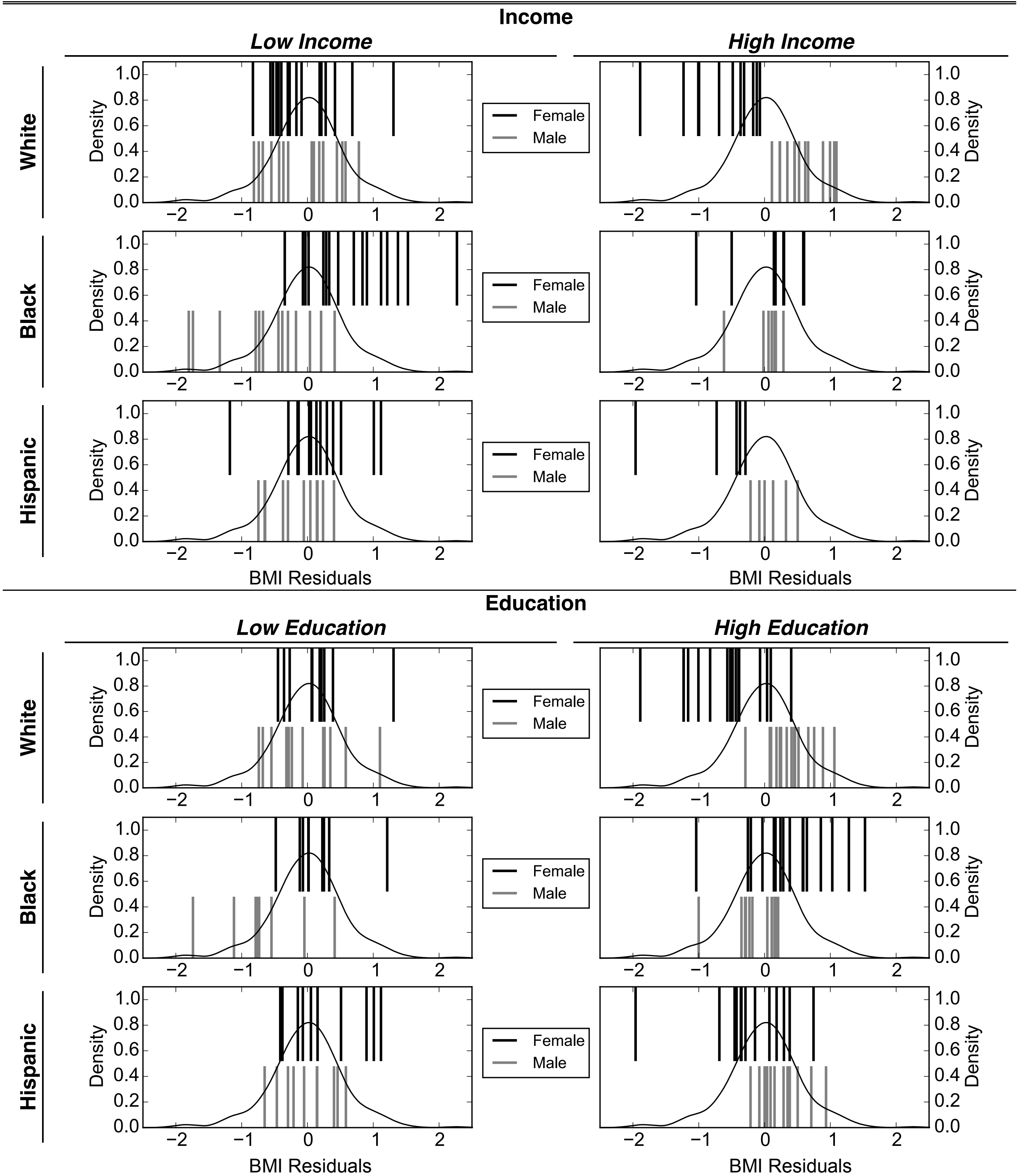
Table 1.6. Strata that occupy the tails of the Obesity residual distribution

Rank	Residual	Sex	Race	Education	Income	Age	*N* in Stratum
*Lower Proportion Obese Than Predicted by Main Effects **							
1	-0.37461	Female	White Non-Hispanic	College degree or more	High	30 to 44	442
2	-0.37243	Male	Black Non-Hispanic	Completed high school	Low	45 to 59	58
3	-0.31582	Female	White Non-Hispanic	College degree or more	High	18 to 29	114
4	-0.29841	Female	White Non-Hispanic	College degree or more	High-middle	30 to 44	598
5	-0.29263	Male	Black Non-Hispanic	Less than high school	Low	60 +	112
6	-0.29195	Female	Hispanic or Latino	College degree or more	High	30 to 44	110
7	-0.28635	Male	Black Non-Hispanic	Less than high school	Low-middle	60 +	98
8	-0.27924	Male	White Non-Hispanic	Completed high school	Low	60 +	122
9	-0.27769	Female	White Non-Hispanic	Completed high school	High-middle	30 to 44	194
10	-0.23670	Male	Black Non-Hispanic	Completed high school	High-middle	45 to 59	57
*Higher Proportion Obese Than Predicted by Main Effects **							
373	0.23700	Female	Black Non-Hispanic	Some college no degree	Low	30 to 44	136
374	0.24158	Male	White Non-Hispanic	Completed high school	Low	30 to 44	110
375	0.24529	Male	White Non-Hispanic	Completed high school	High	45 to 59	145
376	0.26318	Male	Hispanic or Latino	College degree or more	Low-middle	18 to 29	26
377	0.27351	Male	White Non-Hispanic	College degree or more	Low-middle	30 to 44	230
378	0.27523	Female	Hispanic or Latino	Less than high school	Low	18 to 29	116
379	0.28179	Male	White Non-Hispanic	College degree or more	High	60 +	425
380	0.28242	Female	Hispanic or Latino	Less than high school	Low	60 +	217
381	0.29163	Female	Black Non-Hispanic	Completed high school	Low	45 to 59	149
382	0.33833	Male	Hispanic or Latino	Completed high school	High-middle	30 to 44	53

* The top ten strata from each of the tails of the residual distribution (i.e., those with residuals of the greatest magnitude, positive and negative). Those strata furthest from what is predicted by the main effects are those that experience significantly more or less obesity on average than predicted by the main effects of the two-level multilevel model. Rank is out of 382 strata, with 1 representing the most negative residual, and 382 the most positive.

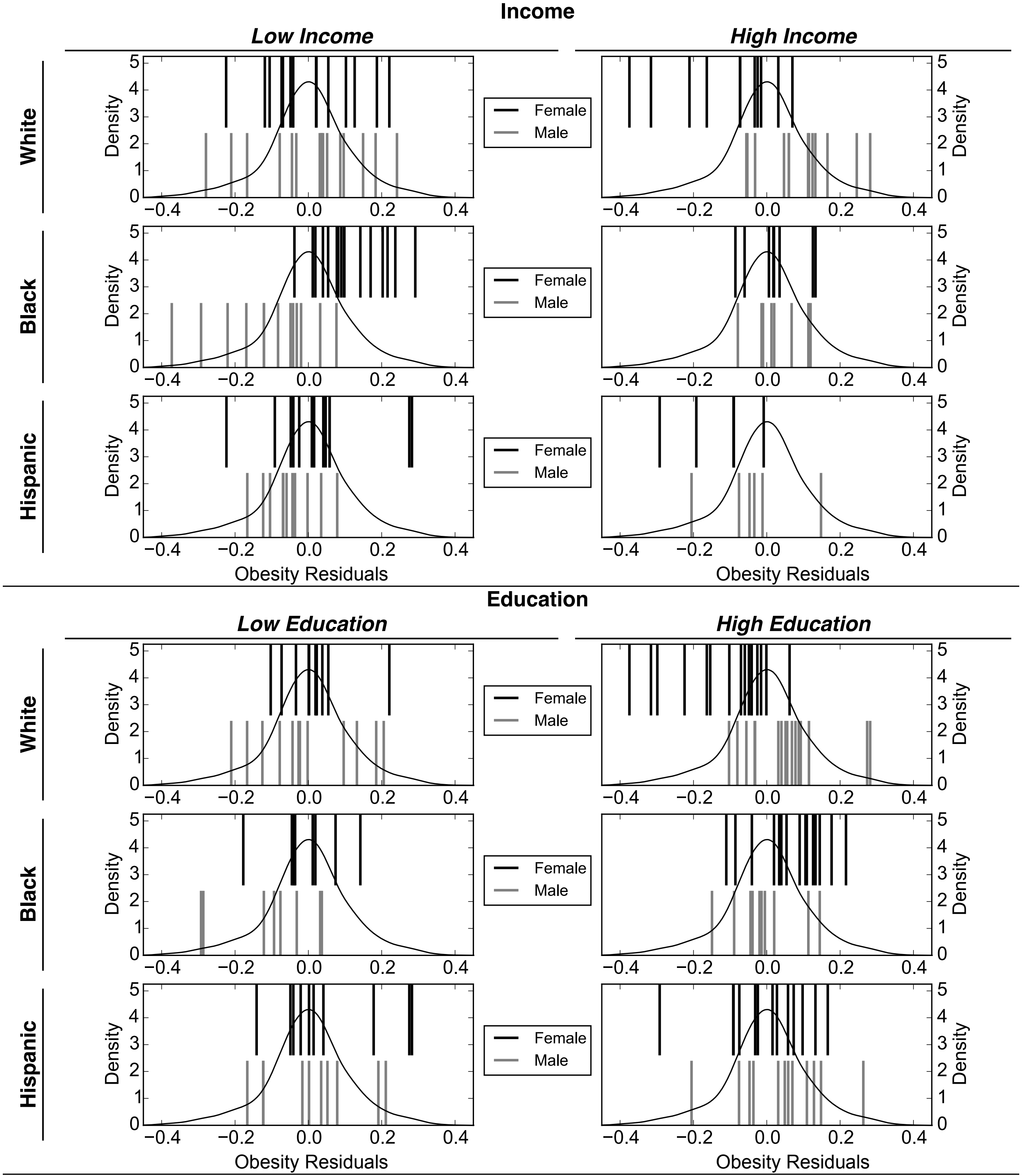
We turn now to Figures 1.1 and 1.2. Figure 1.1 superimposes on the strata-level residual distribution from the BMI model, markers indicating the location within the distribution of various particular strata. The classic triad of sex, race/ethnicity and socioeconomic status (SES) are considered as super-categories. For instance, numerous strata at varying levels of education and age belong to the grouping ‘white females of high income’ and therefore numerous strata markers are shown for each grouping. These figures enable us to visualize where strata belonging to particular super-categories fall within the strata-level residual distribution. If most of the strata markers for a particular super-category fall, for instance, on the *left-hand* side of the distribution, as is the case for white females of high income, then this indicates that generally those strata experience a *lower* mean BMI than might have been expected based on the main effects.

Figure 1.1. Strata-Level Residuals from Models of BMI, and the Location of Particular Strata Groups in the Distributions in Wave 2



Note: Each line represents one stratum's position in the residual distribution. Negative values indicate lower BMI 'than expected' while positive values indicate higher BMI 'than expected' for that stratum. Only strata with 20 or more observations are shown.

Figure 1.2. Strata-Level Residuals from Models of Obesity, and the Location of Particular Strata Groups in the Distributions in Wave 2



Note: Each line represents one stratum's position in the residual distribution. Negative values indicate lower proportion Obese 'than expected' while positive values indicate higher proportion Obese 'than expected' for that stratum. Only strata with 20 or more observations are shown.

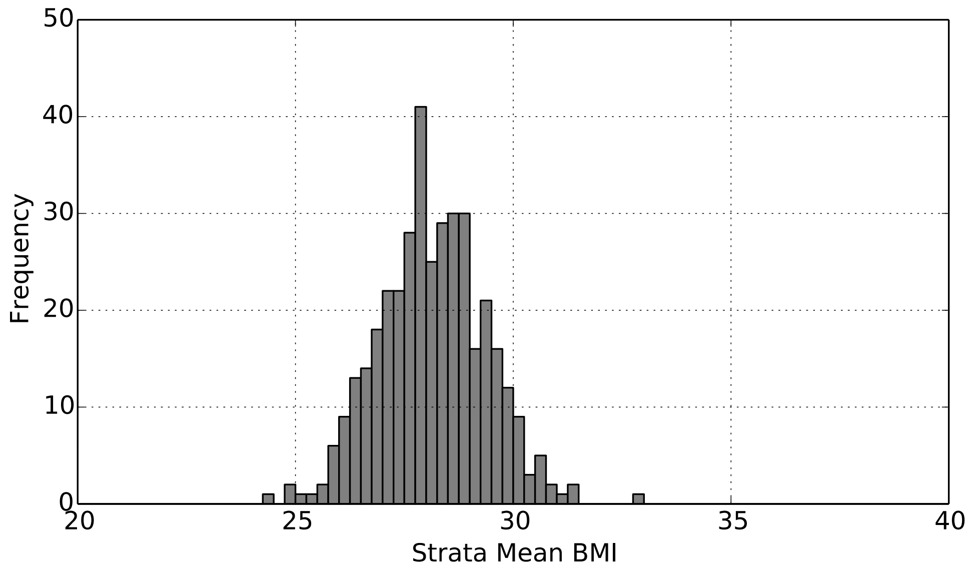
The intriguing patterns that emerge are that some groups, Hispanic males of high and low income, for instance, have strata clustered near the middle of the residual distribution, indicating that while they typically have non-zero strata-level residuals they are nevertheless accounted for relatively well by the main effects. On the other hand, strata for white females of high SES, Hispanic females of high income, and black males of low SES tend to cluster in the tail of the residual distribution indicating lower BMI than expected. Conversely, strata for white males of high SES and black females of low SES tend to cluster in the other tail of the residual distribution, indicating higher BMI than expected. Similar patterns emerge in Figure 1.2, which provides the same visualizations for the obesity strata-level residual distribution.

One point clearly illustrated in both Figures is that there is tremendous variability between strata belonging to the same "sex by race by SES" super-categories. An analysis that disregarded this further parsing and considered, for example, black females of low socioeconomic status to be a homogeneous group, would miss the variability within this group.

Both the multilevel models and the OLS models provide estimates for each stratum for the mean BMI and percent obese. These estimates were plotted as histograms for each model in Figures 1.3 and 1.4, enabling us to compare the distribution of estimates obtained using the two approaches. As is clear in both figures, the OLS models have fatter tails, indicating an over-estimation of between-strata variability. This is a direct result of the OLS model failing to adjust for the sample size of strata (i.e., strata with small sample size have biased and unreliable estimates). The BIC value for the multilevel model (209,533.2) was smaller than the BIC value for the OLS model (212,787.7)—a difference (3,254.5) that was likely due to model parsimony. The estimate distributions and difference in BIC support our claim that the multilevel approach is more efficient and provides more stable estimates.

Figure 1.3. Strata-Level Predictions of Mean BMI obtained using Multilevel approach versus OLS approach

Multilevel Model Estimates



OLS Model Estimates

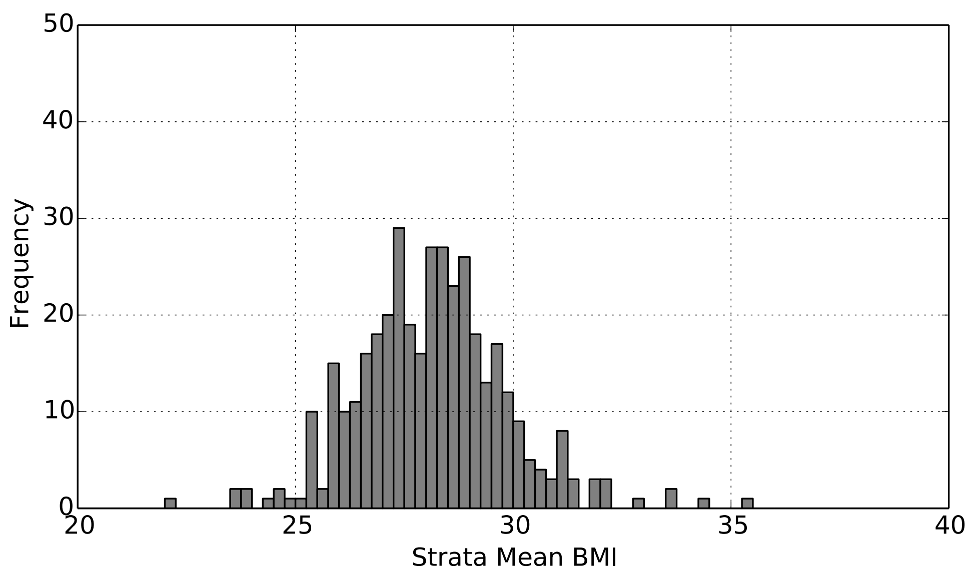
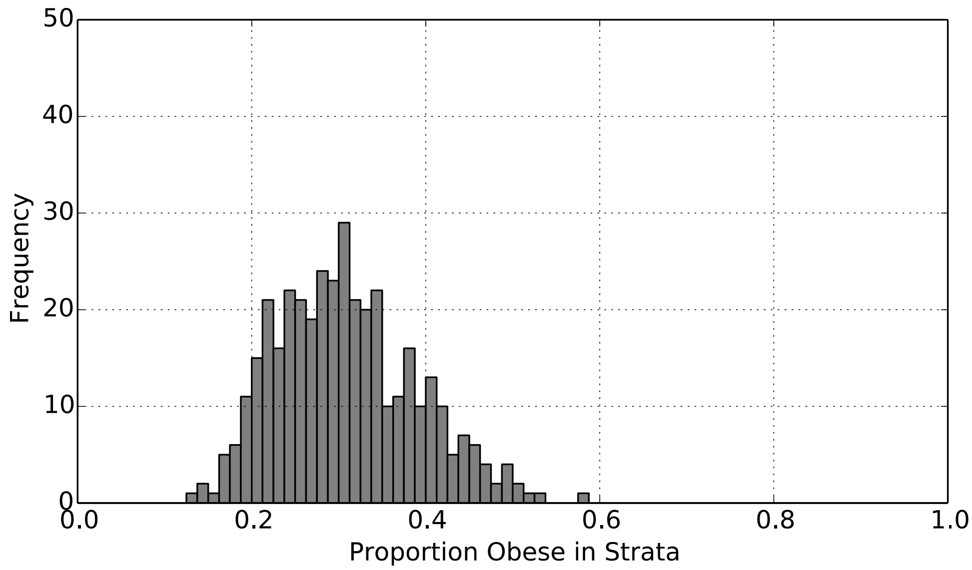
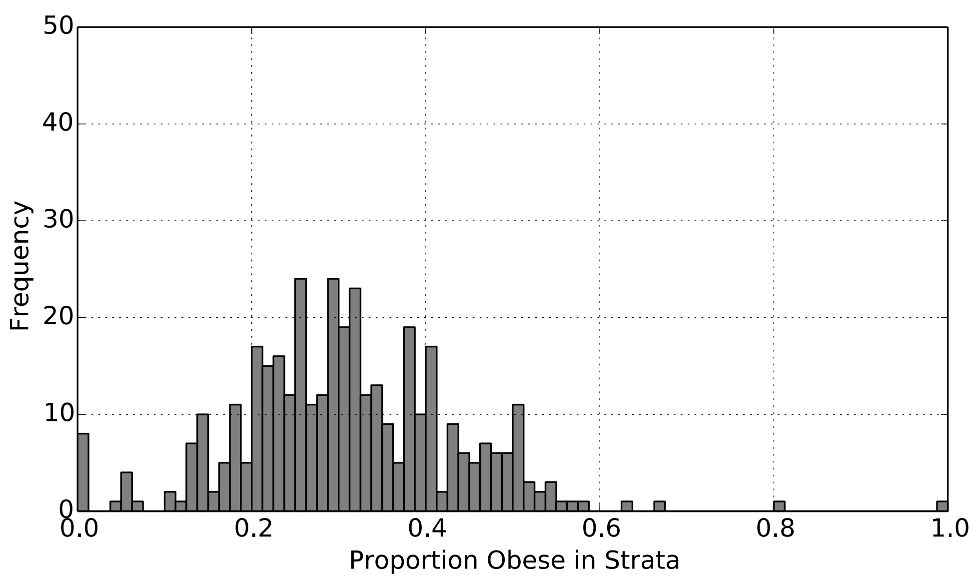


Figure 1.4. Strata-Level Predictions of Proportion Obese obtained using Multilevel approach versus OLS approach

Multilevel Model Estimates



OLS Model Estimates



DISCUSSION

Medical sociologists and social epidemiologists have previously established the strong links between social disadvantage, marginalization and an increased burden with respect to the obesity epidemic; these links are confirmed by this analysis. Minorities, females and those of the lowest SES (relative to the highest) experience an elevated burden of high BMI and obesity. The primary question of interest to this study, however, is whether existing at the intersection of multiple disadvantageous (or advantageous) social positions carries with it an impact on BMI and obesity above and beyond what is expected by the aggregation of the independent effects of these social positions.

This study has demonstrated a novel approach to evaluating the interactions between numerous social categories. It is essential to note, however, that it would be inadvisable to attribute all of the unexplained between-strata variability to intersectionality. Consideration of additional main effects that are predictive of BMI and obesity, currently not incorporated in the model, may reduce the estimate for the residual between-strata variability; one should therefore not attribute the observed effect to intersectionality alone, as omitted variable bias may be artificially inflating our estimate of intersectionality. Acknowledging this, these analyses do simultaneously account for five of the most commonly evaluated demographic characteristics when considering social disparities in the obesity epidemic.

The results of this analysis suggest that while there appears to be considerable intersectionality occurring, it is not always occurring as expected. A significant 35% of the residual between-strata variability for BMI, and 22% for obesity, remain after accounting for the contributions of the main effects, and many strata experience an impact on predicted BMI and obesity levels comparable to (or greater than) the impact of being of the highest income level relative to the lowest. Even if some of the effect were accounted for by omitted variables, these analyses provide strong evidence that some strata are demonstrating particular resilience or burden—above and beyond what we might have predicted for them—with respect to the obesity epidemic.

A second, major conclusion of this study is that while intersectionality appears to be occurring, it is not following the clear-cut story of "multiplicative benefit" and "multiple jeopardy." These results suggest that more complex interactions are occurring with respect to BMI and obesity than predicted, implying a need for an expansion of current theory. For instance, the multiply advantaged white males of high SES and the multiply disadvantaged black females of low SES both seem to experience a *greater* obesity burden than might have been expected. While this provides some support for the "multiple jeopardy" hypothesis with respect to black females of low SES, the opposite does not appear to be true—white males of high SES are not (generally) experiencing a multiplicative benefit. It is critical to recall, however, that in absolute terms the multiply disadvantaged black females of low SES are more burdened by obesity than the advantaged white males of high SES, due to the main effects contributions. On the other hand, black males of low SES, who also experience multiple marginalized and disadvantaged social positions, generally experience a lower obesity burden than might have been expected. From a theoretical perspective these results partially contradict our expectations, and indicate a clear need for further research.

Finally, these analyses provide a partial answer to the question of whether all strata experience intersectionality. Clearly not all do, as some are accounted for relatively well by the main effects, however many strata do, including those who experience multiple advantages. From the perspective of testing theory, it is essential to reiterate that these results will almost certainly vary by population, context, time, and outcome considered.

Limitations

The limitations of this study largely relate to limitations of the model itself. First, the model still requires certain assumptions about the functional form of the data, such as the assumption that observations are independent and identically distributed (iid) within strata. Second, this approach requires the use of very large data sets with sufficient diversity in the sample for reliable estimation of effects. Third, it still relies on the rough categorizations of individuals into sometimes arbitrarily defined groups, a practice that some intersectionality researchers have rightly criticized on the grounds that it fails to capture the complexity and diversity of such identities. Additionally, not all potentially relevant social dimensions were considered in this demonstration. Fourth, there remains the possibility of misclassification of respondents into strata due to the demographic variable imputations and the difficulty of parsing the sample into income categories.

Finally, and perhaps obviously, it is vital to acknowledge that this approach does not address all of the questions posed by intersectionality researchers. For instance, this approach does not identify those factors that *contribute* to the intersectionality of a given stratum. This approach ought, therefore, to be used as a complement to other approaches—particularly qualitative ones—and not as a substitute. So long as these limitations are acknowledged from the outset, this approach provides a new and valuable tool for the exploration of intersectionality from a quantitative perspective.

Significance and Innovation

This study makes several major substantive contributions. First, it has enabled us to quantitatively assess the intersectionality of social strata, providing a more complete picture of inequalities between socially defined groups than has been readily available before. Further research is needed to consider these results, and the results of similar analyses. There is the potential that some of the strata that are discovered to deviate significantly from what is predicted based on the main effect contributions of sociodemographic variables will not have been sufficiently studied in detail using qualitative methods, and this analytic approach may therefore highlight gaps in our understanding and point the way to promising avenues of future research. Second, the results of this study do not completely conform to our current understanding, and thus challenge us to expand our thinking in future theory-based research. While quantitative investigations such as this are capable of identifying social strata that demonstrate particular burden or resilience, we particularly encourage the use of qualitative and mixed-methods approaches to investigate further the uniqueness of the *identity, meaning* and *lived experiences* associated with occupying those social positions, as such investigations will enable the further development of theory.

This study has also contributed to our methodological repertoire by broadening our vision of the potential applications of multilevel models. Multilevel modeling has already demonstrated its value in the public health and social science literature. Yet as far as we are aware, a random effects approach has never been used to address clustering from so abstract a source as social position. Additionally, this study has demonstrated a novel and innovative methodology for studying intersectionality and envisioning disparities in society.

While we find that the multilevel approach is more efficient and provides more stable strata-level estimates than the standard OLS approach, we believe that the main contribution of this new approach is in improving the interpretability of results. The multilevel framework enables us to more explicitly parse variation in a population into between-strata and within-strata levels, to quantify the amount of intersectionality occurring above and beyond the mere additive, and to better envision the social patterning of health inequalities across social categories of interest.

REFERENCES

Ailshire, JA and JS House. 2011. "The Unequal Burden of Weight Gain: An Intersectional Approach to Understanding Social Disparities in Bmi Trajectories from 1986 to 2001/2002." *Social Forces* 90(2):397-423.

Ben-Shlomo, Y and D Kuh. 2002. "A Life Course Approach to Chronic Disease Epidemiology: Conceptual Models, Empirical Challenges and Interdisciplinary Perspectives." *International Journal of Epidemiology* 31:285-93.

Berkman, LF and I Kawachi, eds. 2000. *Social Epidemiology*. Oxford: Oxford University Press.

Boardman, JD, BK Finch, CG Ellison, DR Williams and JS Jackson. 2001. "Neighborhood Disadvantage, Stress, and Drug Use among Adults." *Journal of Health and Social Behavior* 42(2):151-65.

Boardman, JD, JM Saint Onge, RG Rogers and JT Denney. 2005. "Race Differentials in Obesity: The Impact of Place." *Journal of Health and Social Behavior* 46(3):229-43.

Bourdieu, P. 1984. *Distinction: A Social Critique of the Judgment of Taste*. Translated by R. Nice. Cambridge, MA: Harvard University Press.

Bowleg, L. 2008. "When Black Plus Lesbian Plus Woman Not Equal Black Lesbian Woman: The Methodological Challenges of Qualitative and Quantitative Intersectionality Research." *Sex Roles* 59(5-6):312-25.

Bowleg, L. 2012. "The Problem with the Phrase Women and Minorities: Intersectionality— an Important Theoretical Framework for Public Health." *American Journal of Public Health* 102(7):1267-73.

Brown, TH, AM O'Rand and DE Adkins. 2012. "Race-Ethnicity and Health Trajectories: Tests of Three Hypotheses across Multiple Groups and Health Outcomes." *Journal of Health and Social Behavior* 53(3):359-77.

Browne, WJ. 2009. *Mcmc Estimation in Mlwin V2.1*: Centre for Multilevel Modelling, University of Bristol.

Campbell, ME and L Troyer. 2007. "The Implications of Racial Misclassification by Observers." *American Sociological Review* 72(5):750-65.

Carr, D and MA Friedman. 2005. "Is Obesity Stigmatizing? Body Weight, Perceived Discrimination, and Psychological Well-Being in the United States." *Journal of Health and Social Behavior* 46(3):244-59.

Carr, D and MA Friedman. 2006. "Body Weight and the Quality of Interpersonal Relationships." *Social Psychology Quarterly* 69(2):127-49.

Chang, VW and DS Lauderdale. 2005. "Income Disparities in Body Mass Index and Obesity in the United States, 1971-2002." *Archives of Internal Medicine* 165(18):2122-28.

Clarke, P, PM O'Malley, LD Johnston and JE Schulenberg. 2009. "Social Disparities in Bmi Trajectories across Adulthood by Gender, Race/Ethnicity and Lifetime Socio-Economic Position: 1986-2004." *International Journal of Epidemiology* 38(2):499-509.

Collins, P. 1990. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. New York: Routledge.

Crenshaw, K. 1989. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics." *University of Chicago Legal Forum*.

Davey Smith, G. 2000. "Learning to Live with Complexity: Ethnicity, Socioeconomic Position, and Health in Britain and the United States." *American Journal of Public Health* 90(11):1694-98.

Dubrow, JK. 2008. "How Can We Account for Intersectionality in Quantitative Analysis of Survey Data? Empirical Illustration for Central and Eastern Europe." *ASK: Society, Research, Methods* 17:85-100.

Farmer, MM and K Ferraro. 2005. "Are Racial Disparities in Health Conditional on Socioeconomic Status?". *Social Science and Medicine* 60(1):191-204.

Ferraro, K and JA Kelley-Moore. 2003. "Cumulative Disadvantage and Health: Long-Term Consequences of Obesity?". *American Sociological Review* 68(5):707-29.

Flegal, KM, MD Carroll, CL Ogden and LR Curtin. 2010. "Prevalence and Trends in Obesity among US Adults, 1999-2008." *The Journal of the American Medical Association* 303(3):235-41.

Frank, R, IR Akresh and B Lu. 2010. "Latino Immigrants and the U.S. Racial Order: How and Where Do They Fit In?". *American Sociological Review* 75(3):378-401.

Grant, BF, K Kaplan, J Shepard and T Moore. 2003. "Source and Accuracy Statement for Wave 1 of the 2001-2002 National Epidemiologic Survey on Alcohol and Related Conditions.". *National Institute on Alcohol Abuse and Alcoholism: Bethesda, MD*.

Grant, BF and KD Kaplan. 2005. *Source and Accuracy Statement for the Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC)*. Rockville, Maryland: National Institute on Alcohol Abuse and Alcoholism.

Hebl, MR and J Xu. 2001. "Weighing the Care: Physician's Reactions to the Size of a Patient." *International Journal of Obesity* 25(8):1246-52.

Hum, D and W Simpson. 2003. "Labour Market Training of New Canadians and Limitations to the Intersectionality Framework." *Canadian Ethnic Studies/Etudes Ethniques au Canada* 35(3):56-69.

Jordan-Young, RM. 2010. *Brain Storm: The Flaws in the Science of Sex Differences*. Cambridge, MA: Harvard University Press.

Kang, SK and GV Bodenhausen. 2014. "Multiple Identities in Social Perception and Interaction: Challenges and Opportunities." *Annual Review of Psychology* 66.

Krieger, N. 1994. "Epidemiology and the Web of Causation: Has Anyone Seen the Spider?". *Social Science and Medicine* 39(7):887-903.

Krieger, N. 2011. *Epidemiology and the People's Health - Theory and Context*. New York, NY: Oxford University Press Inc.

Larsen, K and J Merlo. 2005. "Appropriate Assessments of Neighborhood Effects on Individual Health: Integrating Random and Fixed Effects in Multilevel Logistic Regression." *American Journal of Epidemiology* 161(1):81-88.

Link, BG and J Phelan. 1995. "Social Conditions as Fundamental Causes of Disease." *Journal of Health and Social Behavior* 35(Extra Issue):80-94.

McCall, L. 2005. "The Complexity of Intersectionality." *Signs* 30(3):1771-800.

McLaren, L. 2007. "Socioeconomic Status and Obesity." *Epidemiologic Reviews* 29(1):29-48.

Mirowsky, J and CE Ross. 2003. *Social Causes of Psychological Distress*. New York: Aldine de Gruyter.

Monk, EP. 2014a. "Skin Tone Stratification among Black Americans, 2001-2003." *Social Forces* 92(4):1313-37.

Monk, EP. 2014b. "The Cost of Color: Skin Tone, Discrimination, and Health among African Americans." Paper presented at the American Sociological Association, San Francisco, CA.

Nash, JC. 2008. "Re-Thinking Intersectionality." *Feminist Review* 89:1-15.

National Institute of Health. 1998. *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults—the Evidence Report*. Obes Res 6(Suppl 2): 51S-209S.

Ogden, CL, MD Carroll, BK Kit and KM Flegal. 2012. "Prevalence of Obesity in the United States, 2009-2010." *NCHS Data Brief 2012*:1-8.

Pamuk, E, D Makuc, K Heck, C Reuben and K Lochner. 1998. *Socioeconomic Status and Health Chartbook*. *Health, United States, 1998*. Hyattsville, MD: National Center for Health Statistics.

Pfeffer, CA. 2014. ""I Don't Like Passing as a Straight Woman": Queer Negotiations of Identity and Social Group Membership." *American Journal of Sociology* 120(1):1-44.

Rasbash, J, WJ Browne, M Healy, B Cameron and C Charlton. 2012. *Mlwin Version 2.26*: Centre for Multilevel Modelling, University of Bristol.

Raudenbush, SW and AS Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage Publications.

Rosenfield, S. 2012. "Triple Jeopardy? Mental Health at the Intersection of Gender, Race, and Class." *Social Science and Medicine* 74(11):1791-801. doi: 10.1016/j.socscimed.2011.11.010.

Schnittker, J. 2004. "Education and the Changing Shape of the Income Gradient in Health." *Journal of Health and Social Behavior* 45(3):286-305.

Schulz, AJ, DR Williams, B Israel, A Becker, E Parker, SA James and JS Jackson. 2000. "Unfair Treatment, Neighborhood Effects, and Mental Health in the Detroit Metropolitan Area." *Journal of Health and Social Behavior* 41(3):314-32.

Schulz, AJ and L Mullings, eds. 2006. *Gender, Race, Class, and Health: Intersectional Approaches*. San Francisco, CA: Jossey-Bass.

Seng, JS, WD Lopez, M Sperlich, L Hamama and CD Reed Meldrum. 2012. "Marginalized Identities, Discrimination Burden, and Mental Health: Empirical Exploration of an Interpersonal-Level Approach to Modeling Intersectionality." *Social Science and Medicine* 75(12):2437-45.

Sobal, J and AJ Stunkard. 1989. "Socioeconomic Status and Obesity: A Review of the Literature." *Psychological Bulletin* 105(2):260-75.

U.S. Census Bureau. 2012, "Poverty Thresholds 2000": Social, Economic, and Housing Statistics Division: Poverty. Retrieved 1 April 2014, (http://www.census.gov/hhes/www/poverty/data/threshld/thresh00.html).

Veenstra, G. 2011. "Race, Gender, Class, and Sexual Orientation: Intersecting Axes of Inequality and Self-Rated Health in Canada." *International Journal for Equity in Health* 10(1):3.

Veenstra, G. 2013. "Race, Gender, Class, Sexuality (Rgcs) and Hypertension." *Social Science and Medicine* 89:16-24.

Villarreal, A. 2010. "Stratification by Skin Color in Contemporary Mexico." *American Sociological Review* 75(5):652-78.

Warner, DF and TH Brown. 2011. "Understanding How Race/Ethnicity and Gender Define Age-Trajectories of Disability: An Intersectionality Approach." *Social Science and Medicine* 72(8):1236-48.

Warner, LR. 2008. "A Best Practices Guide to Intersectional Approaches in Psychological Research." *Sex Roles* 59:454-63.

Worthen, MGF 2013. "An Argument for Separate Analyses of Attitudes toward Lesbian, Gay, Bisexual Men, Bisexual Women, MtF and FtM Transgender Individuals." *Sex Roles* 68(11-12):703-23.

Social Network Segregation by Family Income Level in U.S. Adolescent Friendship Networks

Clare Rosenfeld Evans[1]

S.V. Subramanian[1]

David R. Williams[1]

Jukka-Pekka Onnela[2]

[1] Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA

[2] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

Acknowledgements and Credits: We wish to thank the reference librarians of the Harvard University Countway Library of Medicine for their assistance in performing a systematic review of the literature.

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Funding: SVS is partially supported through the Robert Wood Johnson Investigator Award in Health Policy Research.

Keywords: Social networks, segregation, adolescents, socioeconomic status, family income, race and ethnicity, schools, social stratification

ABSTRACT

This study evaluates whether adolescent social networks are segregated by socioeconomic status. In Wave 1 of the Add Health study, we consider 16 schools for which social network and self-reported family income data exist for most students (*N*=2,584). We approach the evaluation of income segregation through a novel lens by explicitly considering three scales of analysis within social networks: the network community (i.e., social clique) level, the dyadic level, and a level in between. We find evidence of income segregation at all three levels of adolescent social networks, though this segregation is neither extreme nor universal. Logistic models at the dyadic level reveal a modest, though consistent tendency for adolescents in five sample schools to select close friends who are more similar to them in family income level. In half of sample schools (46% to 50%, depending on methodology), the mean income gap between pairs of adolescents in the same network community is smaller than the mean income gap between pairs of adolescents in different network communities. Multilevel analyses reveal network community-level clustering of family income after adjustment for race/ethnicity. Family income appears to be a socially salient factor in the structure of adolescent social networks.

CHAPTER 2: SOCIAL NETWORK SEGREGATION BY FAMILY INCOME LEVEL IN U.S. ADOLESCENT FRIENDSHIP NETWORKS

INTRODUCTION

Segregation has long been acknowledged as a substantial concern to society, not only for its salience as a civil rights issue, but also because racial, ethnic, and socioeconomic segregation are associated with elevated disparities in health and constrained educational aspirations and achievement (Acevedo-Garcia and Lochner 2003, Aneshensel and Sucoff 1996, Crosnoe 2009, Massey 2004, Willms 1986, Wilson 1959). Historically, our conceptualization of segregation has tended to focus both on racial/ethnic segregation and on structural or spatial aspects of segregation, such as between schools, neighborhoods, or work places (Farley and Taeuber 1974, Fiel 2013, James 1989, Logan, Oakley and Stowell 2008, Massey and Denton 1993, Tomaskovic-Devey et al. 2006). Research on socioeconomic segregation has similarly emphasized the sorting of different income or occupational classes across residential areas and schools (Erbe 1975, Reardon and Bischoff 2011). These conceptualizations, while capturing vital aspects of segregation, are incomplete because they tend to overlook the difference between integration of an environment and the experiences of individuals who live or interact within those environments. Moody (2001) highlighted this distinction by demonstrating that even within racially integrated schools, the social networks of adolescents can be highly segregated along racial/ethnic lines. This implies that racial integration of a school—or neighborhood—does not necessarily indicate that the *lived experiences* of the individuals who interact within those environments are integrated ones. After conducting a systematic review of the literature, we found that the extent to which adolescent social networks are segregated by socioeconomic status has been seriously neglected. Additionally, the level at which segregation operates in social networks—be it between social network groups, or at the friendship-pair level—has not been explicitly addressed. In this study we seek to address these gaps in our current knowledge.

Using data from 16 schools surveyed in Wave 1 of the National Longitudinal Study of Adolescent to Adult Health (Add Health), for which we have high quality adolescent social network and family income data for the majority of students, we evaluate the extent to which adolescent school-based social networks are segregated by family income. Networks are evaluated at the network community (i.e., social clique) level, the dyad (i.e., pair) level, and at a point in between (dyads organized into "within" and "between" network communities), in order to determine to what extent segregation exists at different scales of analysis.

Background

Scales of Segregation and Adolescent Social Networks

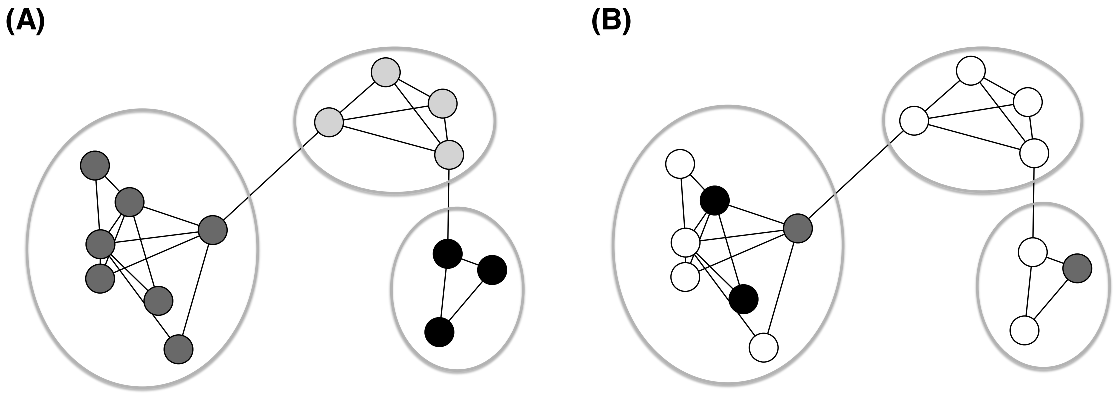
It is commonly accepted that adult social networks are highly segregated by socioeconomic status (Ajrouch, Blandon and Antonucci 2005, Bian et al. 2005, Blau, Ruan and Ardelt 1991, DiPrete et al. 2011, King 1961, Marques 2012, McPherson, Smith-Lovin and Cook 2001, Smith, McPherson and Smith-Lovin 2014) largely because adults live and work in segregated environments (due to a combination of institutional discrimination and selection) that limit opportunities for forming ties across socioeconomic levels. To the extent that schools are also segregated by socioeconomic factors (Crosnoe 2009, Willms 1986, Wilson 1959), child and adolescent social networks are likely similarly constrained in terms of presenting opportunities for cross-class friendships to form. However, it is unclear whether—within school environments that theoretically should provide opportunities for mixing between socioeconomic levels—we find socioeconomic network segregation among adolescents.

Put another way, there are four levels at which we can envision adolescent social networks as being segregated. The *first*, and perhaps most obvious, is between-schools or between-residential neighborhoods. If these environments are relatively homogeneous within and heterogeneous between, then we might assume this will impact the diversity of adolescent networks. In network analysis terms, residing in different neighborhoods or attending different schools limits opportunities for friendships to form across such environments, and this may result in disconnected (or mostly disconnected) network components. If schools or neighborhoods are socioeconomically segregated then this may result in socioeconomic *network* segregation as well.

The *second* level is within these environments. If we envision peering inside of a school and mapping the entire social network of students, then we might detect social groups or 'cliques'—referred to by network scientists as *network communities* (Fortunato 2010, Porter, Onnela and Mucha 2009), which are broadly defined as sets of individuals who are more densely connected to each other than to others elsewhere in the network. A cartoon visualization of network communities is provided in Figure 2.1a. These network communities would be expected to approximate social groups that regularly socialize together. Network community level segregation might be evaluated by considering whether network communities are heterogeneous with respect to family income. In other words, in a multilevel framework, does clustering occur by family income at the network community level?

Third, we can consider whether pairs of adolescents (dyads) within the same network community have a smaller mean income gap than pairs in different network communities situated in the same school (Figure 2.1b). And *fourth*, we can consider the dyadic level itself, entirely disregarding the borders of network communities (Figure 2.1c). At the dyadic level we might evaluate the extent to which friend pairs are homogeneous with respect to family income.

Figure 2.1. Cartoon Schematics of Levels Evaluated in Network Analysis



Note: Nodes (adolescents) are represented as circles, and the lines between them represent social connections, or links.

Panel A: *Network Community Level*. Nodes are nested within the network community to which they belong. Network community membership is indicated by color and network communities are encircled for illustration.

Panel B: *Dyads Within vs. Dyads Between Network Communities Level*. All pairs (dyads) of nodes, regardless of whether they are linked or not, are sorted into “Within” and “Between” categories. At this analysis level we address dyads but also consider the network community level. “Within” dyads (black) reside in the same network community. “Between” dyads (grey) reside in different network communities.

Panel C: *Dyadic Level*. All pairs (dyads) of nodes are considered without reference to network communities. Whether dyads are linked (yes = black) or not linked (no = grey) is the dependent variable in the logistic models.

These three "within school" levels of network segregation are the focus of this study. Each level captures qualitatively different aspects of adolescent social relationships. For instance, prior studies have demonstrated that friendship nominations, particularly when limited to a set number of ties, tend to capture particularly strong social connections (Moody and White 2003, Quillian and Campbell 2003, Schofield and Whitley Jr. 1983) while missing friendly acquaintances. On the other hand, a network community may include adolescents who regularly interact socially but who would not nominate each other as particularly close connections. Since some individuals might enjoy diversity in their general social group, but preferentially nominate individuals who are more similar to them when identifying their closest friends, there is no guarantee that segregation at the dyadic level implies segregation at the network community level as well.

Despite the qualitative differences between levels, network communities are rarely treated as the unit of analysis in network segregation studies, largely due to the dearth of sociocentric data (i.e., maps of the network) relative to egocentric data (i.e., derived from surveys of individuals that ask about their immediate social connections), and to the fact that community detection is a relatively new field. One example we have identified is a study by González et al (2007) which utilized network community detection to explore racial and ethnic preferences in adolescent social networks.

At the dyadic level, *homophily* refers to the tendency for people to form social relationships with others who have similar attributes as themselves, such as gender, education level, and race/ethnicity. While evaluations of adolescent network homophily at the dyadic level are more prevalent (though certainly not common), particularly with respect to race/ethnicity, sex and age (Currarini, Jackson and Pin 2010, McPherson, Smith-Lovin and Cook 2001, Moody

2001, Mouw and Entwisle 2006, Neal 2010, Wilson and Rodkin 2011), few studies have explored socioeconomic homophily. We will now briefly consider several papers that form the basis of our current knowledge about socioeconomic segregation in adolescent social networks.

The first, published by Jere Cohen (1979), includes a secondary analysis of data collected in 1958 (Coleman 1961) on the high school friendship and dating patterns in Elmtown High, as well as a comparison of the results with previously published data on the same school (Hollingshead 1949). Cohen concluded that parents' socioeconomic status (SES) was less important as a criterion of choice in 1958 than it was in 1942, but that it was still clearly important. The second paper, published more recently by Maharaj and Connolly (1994), addressed primarily racial/ethnic homophily among 896 Canadian adolescents enrolled in Grades 9-12 in a middle-class suburban high school near Toronto. In their analysis, adolescent SES (derived from parents' highest level of education) was used along with other demographic variables to account for "ethnocultural" homophily. In other words, it was not SES-based segregation per se that was evaluated, but the extent to which SES predicted racial/ethnic homophily in friendship networks.

The third paper, published by Burgess et al (2011) was an analysis of an adolescent friendship network (6,961 friendship linkages between 2,396 adolescents) in the West of England. Measures of SES evaluated included household income, parents' occupational class, and parents' highest educational attainment. Between friends, homophily based on parental education was found, but homophily based on household income and parent occupational class was not.

Fourth, in an analysis focusing on interracial friendships in U.S. schools by Mouw and Entwisle (2006), the authors considered the possibility that some of the racial/ethnic homophily detected in adolescent friendship ties could be due to an underlying preference for friends of the same SES, and due to race-based differences in SES it may be that apparent preference for same-race friendships is really due to preference for same-SES. While they did not consider SES-based homophily in isolation, they did include the difference between adolescents' family incomes and the difference between adolescents' parent education in analyses at the dyadic level. They concluded that increasing difference in parents' income or years of education decreased the probability of two individuals being friends. Finally, an analysis by Bearman, Moody and Stovel (2004) of a predominantly-white high school in the Add Health sample found that adolescents tended to select romantic and sexual partners of similar socioeconomic levels.

To summarize, studies addressing socioeconomic segregation in adolescent friendship networks—particularly recently in the United States—are rare, and none of these studies specifically address the issue of the scale, or level of analysis, in their consideration of network segregation. Dyadic level homophily is the default scale considered, and this scale is uncritically assumed to represent segregation at other network scales. The extent to which adolescent social networks are segregated by socioeconomic status, and the scale at which this segregation operates, is therefore unclear.

Implications of Adolescent Network Segregation

Understanding adolescent network segregation by socioeconomic status is vital due to the implications of segregation or integration for the health, educational trajectories, and well being of adolescents. Social network segregation or integration by family income level could affect adolescent health outcomes through numerous pathways. Network *integration* may benefit the health of adolescents from lower income families. For instance, friendships with students from advantaged backgrounds may potentially bring low-income adolescents into contact with professionally successful or highly educated adults (the parents of their friends). This represents a potential *resource* for these adolescents, who could benefit from such things as informational support about college and scholarship application processes, and from the social norms, expectations and role models that are likely to exist in households with highly educated parents, such as those pertaining to educational achievement and professional development. If these resources increase the likelihood of low-income students pursuing higher education, this will also provide health dividends throughout the rest of the life course, as higher education is strongly associated with improved health behaviors and outcomes (Ross and Wu 1995, Winkleby et al. 1992). Additionally, it is well established that health-related behaviors (e.g., healthy diet, consistent physical activity, not smoking) are socially patterned, with higher socioeconomic status adults (Emmons 2000) and their adolescent children (Hanson and Chen 2007) being more likely to engage in healthy behaviors. Network integration could therefore benefit low-income adolescents because they are likely to be exposed to healthier behaviors, both among their friends and their friends' parents, and this exposure could result in adoption of healthier behaviors through social influence (Ali and Dwyer 2010, Christakis and Fowler 2007, Christakis and Fowler 2008).

On the other hand, network *segregation* by income among adolescents does not just represent the loss of the previously mentioned opportunities, it also indicates that—at a young age—individuals may be highly conscious of socioeconomic status, and may attach social significance (such as peer status) to the possession of certain material markers of family income. There is evidence that, through numerous complex pathways, parental socioeconomic status (such as income and education) can translate into the social status of their adolescent children, and further that this status may translate into the adolescents' socioeconomic status as they transition into the adult world (Bjarnason 2000). It is possible, therefore, that adolescent social status may translate into health outcomes in later life via socioeconomic status mediating pathways. Additionally, if network segregation is occurring because of some form of socioeconomic-consciousness and the status-significance attached to family income, then this could adversely affect the developing adolescents. For instance, it might promote adverse behaviors (such as discrimination or bullying) and the formation and perpetuation of unflattering class-based stereotypes. Bullying—for both the perpetrator and victim—has been shown to have short-term (Forero et al. 1999, Nansel et al. 2001) and long-term (Sourander et al. 2007) implications for mental and psychosomatic health. Similarly, discrimination is stress inducing, and chronic activation of stress response pathways has been linked to various adverse health outcomes (Cohen, Janicki-Deverts and Miller 2007). Exposure to income-based discrimination may result in similar patterns of inequalities between social groups as is created by race-based discrimination (Fuller-Rowell, Evans and Ong 2012, Williams and Mohammed 2009).

It is, therefore, important to characterize the extent to which adolescent social networks are segregated or integrated, and the levels at which this occurs, in order to further our understanding of the social world that shapes educational and health trajectories of adolescents. Our first research question is: To what extent are adolescent social networks segregated by family income at each of the three levels of analysis—the network community level, the dyads *within* network communities versus dyad *between* network communities level, and the dyadic level? Our second research question is whether there is a correspondence between the levels in terms of the segregation (or integration) observed.

METHODS

Data: The National Longitudinal Study of Adolescent to Adult Health

Add Health is a longitudinal study of a nationally representative sample of U.S. adolescents who were in grades 7-12 in the first wave of interviews (1994-1995) (Harris et al. 2009). The primary sampling frame was derived from the Quality Education Database (QED) and was used to select a stratified sample of 80 high schools with probability proportional to size, as well as 52 middle schools that were paired to the high schools as feeders. Schools were stratified based on region, urbanicity, school type (public, private, parochial), ethnic mix and size. An important aspect of the Add Health data is that students were asked to nominate up to 10 of their closest friends (5 male and 5 female), and since many of the nominated students were also interviewed it is possible to construct a fairly complete social network for each school.

In Wave I, an in-home questionnaire was also administered to a sub-sample of randomly selected students from each school called the 'core sample,' and interviews with a parent of each child in the core sample were attempted. It is this parent interview that provides the family income data of interest in this study. Because the parent interview is only available for a sub-set of students from many schools in the Add Health sample, it is necessary to restrict the present analysis to 16 schools. In these 16 schools—2 large high schools plus 14 smaller middle schools and high schools—an in-home interview with all students and their parents was attempted. These 16 schools therefore represent a *saturated sample* for which social network and self-reported family income data are available for most students.

Study Population

The saturated sample of schools represents a variety of contexts. The two large high schools, for example, are significantly different—one is a predominantly White (94%) rural public high school (grades 9 to 12, N = 1,024) in the Midwest, while the other is a racially diverse (Hispanic = 39.3%, Black = 23.2%, Asian = 31.6%, White = 4.5%) suburban public high school (grades 10 to 12, N = 2,104) in the West. The other 14 schools range in size from 47 to 193 students and represent public (n = 9), private (n = 4), and Catholic (n = 1) schools in the West (n = 3), South (n = 4), Northeast (n = 3), and Midwest (n = 4) regions.

Measures

Family Income

Family income was measured through a single item on the Parent In-Home Questionnaire administered during Wave I of Add Health: "*About how much total income, before taxes, did your family receive in 1994? Include your own income, the income of everyone else in your household, and income from welfare benefits, dividends, and all other sources*." In an effort to boost the response rate, respondents were again reassured about confidentiality just prior to this question being asked. Family income was measured as a continuous outcome and response values between $0k and $999k were permitted. For this analysis, $0k was interpreted as "less than $1,000" and was assigned the value $500.

Race / Ethnicity

In order to address the potential mapping of income-based segregation onto racial/ethnic segregation, student race/ethnicity is also considered. The standard procedure for constructing a single race/ethnicity variable in Add Health Wave I was used (UNC Carolina Population Center project). Race and ethnicity were self-reported. Coding of race/ethnicity was determined based on the first category marked in the following list: Hispanic all races, Black or African American, Asian, Native American, Other, and White. Therefore, an individual who indicated both "Hispanic or Latino" ethnicity and "Black or African" race would be coded as "Hispanic, All Races." Alternatively, an individual who identified as both Asian and White would be coded as "Asian, Non-Hispanic."

ANALYSIS

Network Community Detection and Segregation Evaluation

In order to evaluate segregation at the network community level, we first apply network community detection algorithms. Network community detection is an active area of research within the field of network science, and so there are many different algorithms that have been developed to identify meaningful network communities. In this analysis, two separate community detection algorithms—*modularity maximization* and *k-clique percolation*—were applied because different algorithms yield subtly different membership lists for each detected network community and we wished to test the robustness of the results to varying definitions of 'network community.'

The concept of a network community is a fairly intuitive one—it represents a collection or group of individuals who are densely connected to each other but who are relatively sparsely connected to others outside the group. In essence, community detection entails identifying such clusters by determining which adolescents reside within each network community, but also determining the number of such clusters contained in the network. Unfortunately, while the concept of network community may be intuitive, identifying communities in practice is difficult because it is necessary to apply a strict definition of what a 'dense cluster' is and what level of 'bridging' between clusters is deemed sparse enough to count as a partition between two communities. In addition to the difficulty associated with defining the concept of network community precisely, there are also computational issues with most definitions, meaning that one needs to resort to different heuristics when implementing these methods in practice. In this sense, the choice of a community detection method is analogous to stating how one would like to ascertain communities, whereas the choice of a particular heuristic for the given method determines how communities are actually ascertained.

Modularity maximization (Newman 2006a, Newman 2006b) is one of the most popular detection methods because it has been shown to partition a variety of networks into what appear to be meaningful network communities (Porter, Onnela and Mucha 2009). Modularity refers to a quality function or a score given to a particular network partition. A common null model used in the formulation of modularity considers the fraction of edges (i.e., friendships) that run between nodes (i.e., adolescents) belonging to the same community and subtracts from it the fraction of such edges we would *expect* to find if the edges where positioned at random in the network while preserving the number of connections of each node. In principle we would like to find the partition that results in the maximum value for modularity by iterating through all possible network partitions until the modularity score is maximized. In practice, however, this is impossible for all but the smallest of networks given the computational complexity of such a procedure. Instead, modularity maximization algorithms utilize different heuristics to find an optimal partitioning of the network into communities. The particular algorithm we used in this study to maximize modularity was developed by Blondel et al (2008). Using this algorithm, each node is assigned to exactly one network community and no nodes that have social connections are "left over" (i.e., socially marginalized from network communities).

K-clique percolation is a network community detection approach that was pioneered by Palla, Derényi, Farkas and Vicsek (2005) and is becoming increasingly popular in evaluations of social networks (Fortunato 2010). It is a deterministic algorithm that begins with a specification for the size of a clique (*k*), or small group of nodes who are fully connected. A clique of size k = 3, therefore, indicates a cluster of three adolescents each of whom shares a friendship link with the other two. Note that although the user needs to fix the value of the clique, the number of communities for the given clique number is determined automatically by the algorithm. The previously mentioned study by González et al (2007) in Add Health found that cliques of size k = 3 were optimal for community detection in Add Health networks. A network community is defined as a set of cliques that are adjacent to one another, where clique adjacency is defined as two cliques sharing $k-1$ nodes. The advantage of this definition is that, unlike modularity, it allows a given adolescent to belong simultaneously to two or more network communities, which likely better reflects the complex reality of social network structures. K-clique percolation also, by definition, will exclude nodes from network communities if they are not members of cliques. These "marginalized" nodes may have very few friends or may have a number of friends spread sparsely across various cliques, but lack sufficient ties to a single clique to belong to it according to the clique definition.

Once the network communities within each of the 16 schools were identified, two segregation analyses were conducted. The primary analysis was a multilevel hierarchical model nesting adolescents (level 1) in their network communities (level 2). The multilevel models took the following forms: Models 1 and 2 controlled for school attended using dichotomous fixed effect covariates, and Model 2 also adjusted for adolescent race/ethnicity to account for correlation between adolescent family income and race/ethnicity.

To test the sensitivity of our results to community specification, both models were fit separately using both modularity maximization and k-clique network communities. Since the k-clique algorithm does not assign "marginalized" nodes (i.e., those who are not members of cliques of size *k*) to network communities, marginalized adolescents are not included in the k-clique version of multilevel models. Additionally, k-clique percolation allows nodes to simultaneously belong to multiple network communities; For the purpose of this analysis, these nodes were assigned to the network community to which they had the most links, and in the case of ties they were randomly assigned to one of the tied network communities.

In the primary analysis the goal is to determine the extent to which family income is clustered at the network community level using a multilevel modeling framework. In this case the emphasis is on assessing differences *between* network communities. In a secondary analysis we focus on the *within* network community homogeneity of moderate and large sized network communities (i.e., 10 or more nodes). Segregation in this context is defined in relative terms as network communities being *more* homogeneous with respect to family income than we might have expected based on the background diversity of income levels within the school. In other words, within community homogeneity implies between community heterogeneity (segregation). If, for instance, a network community of 10 students was identified within a given school, we first calculated the variance of family income within this subset of students. We then determined whether the empirical (observed) variance was smaller than we might have expected by chance (indicating it is relatively homogeneous, and therefore segregated) by repeatedly resampling from the entire school network samples of 10 students and determining for each sample the variance in family income. For each detected community, this process was repeated 1,000 times to obtain a distribution of variance. If the detected network community's observed variance in income was significantly smaller than what would be deemed likely by this distribution (at $p < 0.05$), then this would indicate that the network community was more homogeneous than expected by chance, therefore segregated from others by income level.

Dyads Within versus Dyads Between Network Communities

All pairs of adolescents (dyads) within a school, regardless of whether a friendship link exists between them or not, were sorted into two groups—those where both adolescents were nested within the same network community (*within* community dyads) and those nested in different network communities (*between* community dyads). For each dyad, the difference in family income was calculated. Using two-sided t-tests, we compared the mean difference in income for the *between* community dyads and the *within* community dyads. This analysis was completed for each of the sample schools using both network community detection algorithms. If the mean income difference for dyads within the same network community were smaller than the income difference for dyads in different network communities, then this would imply that adolescents in the same network community—who likely socialize with each other in a group regardless of whether they are particularly close friends—will be likely to have more similar income levels than adolescents in different network communities.

Analyses were repeated for each school and both network community detection algorithms. In k-clique percolation, three versions of models where fit to ensure that results were robust to the treatment of dyads with marginalized adolescents (those who were not assigned to any network community): (1) dyads where one or both were marginalized counted as *between* community dyads, (2) dyads where both were marginalized were dropped from analysis and dyads with one node marginalized counted as *between* community dyads, and (3) dyads where one or both nodes were marginalized were dropped from analysis.

Dyadic Level Income Homophily

Evaluation of income segregation at the dyadic level involves a more standard analytic approach. All possible dyads of students within a school can be grouped into two categories—the pairs that have a friendship tie and the pairs that do not. The absolute difference in family income ($\text{Difference\_Income}_{\_ijk}$) between all pairs of adolescents (i,j) was used to predict the probability of friendship ties existing between them. The model takes the following form:

$$\text{logit}\,[Y_{\_ijk}] = \beta_{0k} + \beta_{1k}\,(\text{Difference\_Income}_{\_ijk})$$

In this model, the outcome, $Y_{\_ijk}$, represents the existence of a friendship tie between individuals i and j (1 = yes, 0 = no) in school k. Logistic models were fit separately for each school. The intercept term (β_{0k}) captures the school-specific effect for school k, which may include the general density of ties within a school (since some schools may encourage the formation of more social ties, whereas other schools may not) and variation in school size. The parameter of interest in this analysis is β_{1k}, which captures the change in probability of a friendship tie existing between individuals i and j in school k for each unit increase in the difference between the family incomes of individuals i and j. A statistically significant and negative value for β_{1k} would indicate that as the differences in income between two adolescents shrinks, the probability of them being friends increases.

A second version of this model was also fit, with a dummy variable indicating whether adolescents i and j are of the "same race/ethnicity," and it was included as a covariate in the model. This version of the model accounts for the possibility of confounding of income homophily by racial/ethnic homophily.

RESULTS

Analyses were conducted in a variety of softwares: SAS 9.3 (SAS Institute Inc 2011), MLwiN 2.32 (Rasbash et al. 2015) using Markov Chain Monte Carlo estimation (Browne 2009), and Python 2.7 (Anaconda by Continuum Analytics 2015). Multiple Python libraries were used, including Python Pandas (McKinney 2011), NumPy (NumPy Developers 2005) , NetworkX (Hagberg, Schult and Swart 2008), Python-IGraph (Csardi and Nepusz 2006), and matplotlib (Hunter 2007).

The 16 schools in the saturated sample yielded a total of 3,702 respondents, of whom 2,584 (70%) had a parent report family income. A further 793 students across the 16 schools were identified through social network nominations, but for reasons such as absence or refusal they are not included in the Add Health sample. All students, including those missing family income responses, are included in the community detection analysis because considering them improves our understanding of the school's network structure. Basic demographic profiles of the sample are provided in Table 2.1 and more detailed profiles of the full sample and particular school samples are provided in Supplemental Tables 2.1, 2.2 and 2.3 of the Appendix.

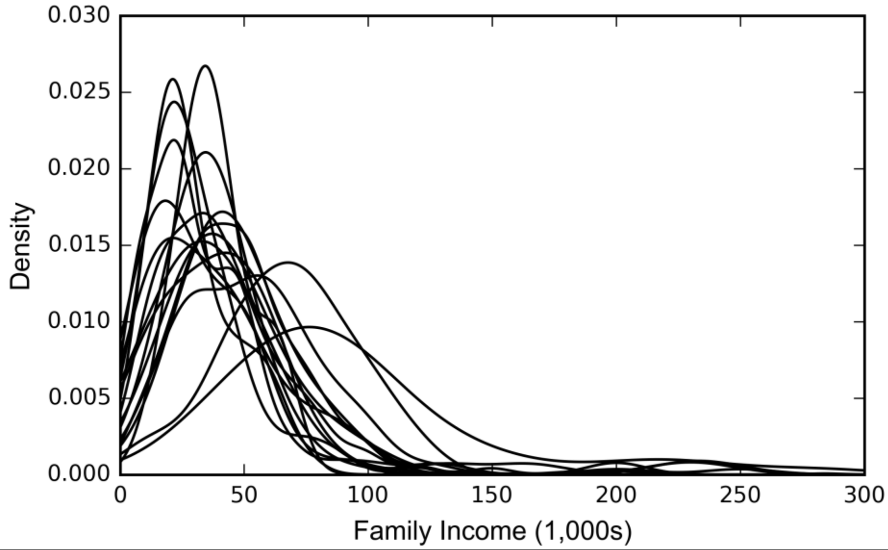
While between-school comparisons of family income are not the focus of the present analysis, it is helpful to understand the different socioeconomic profiles of the sample schools in order to characterize the sample and better contextualize subsequent results. Figure 2.2 shows the overlaid distributions of family income for each school. While most of the schools include students from very low-income backgrounds, there is substantial variation between schools on the mean and standard deviation of income. School "N," for instance, has a mean family income of $93,220 per year (SD = $52,900), while school "K" has a mean family income of $28,660 per year (SD = $17,470). The schools also vary significantly in their racial/ethnic composition, ranging from school "L," which is 100% white non-Hispanic, to school "K," which is nearly 98% black non-Hispanic. The two largest schools in the sample are the predominantly white non-Hispanic school "G" (mean income $49,030 per year (SD = $27,890)) and the racially diverse school "H" (mean income $40,190 per year (SD = $25,900)).

Table 2.1. Demographic Profile of Sample

	Frequency N	(%)
TOTAL SAMPLE [a]	**3702**	**(100.00%)**
Main Sample – Provided Income	2584	(69.80%)
Refused or Missing Income	1118	(30.20%)
MAIN SAMPLE	**2584**	**(100.00%)**
Sex of Adolescent		
Female	1232	(47.68%)
Male	1352	(52.32%)
Race / Ethnicity of Adolescent		
Missing	3	(0.12%)
Hispanic or Latino	437	(16.91%)
Black, Non-Hispanic	385	(14.90%)
Asian, Non-Hispanic	300	(11.61%)
Native American, Non-Hispanic	44	(1.70%)
Other, Non-Hispanic	15	(0.58%)
White, Non-Hispanic	1400	(54.18%)
Highest Education Level of Parents		
Less than high school	270	(10.45%)
Completed high school	675	(26.12%)
Some college no degree	830	(32.12%)
College degree or more	809	(31.31%)

[a] An additional 793 students were identified through friendship nomination, however they were not surveyed (perhaps due to absence on the day the survey was administered) and therefore little is known about them beyond their gender. These individuals are included in the network structure aspects of analysis and are coded as 'missing' most data.

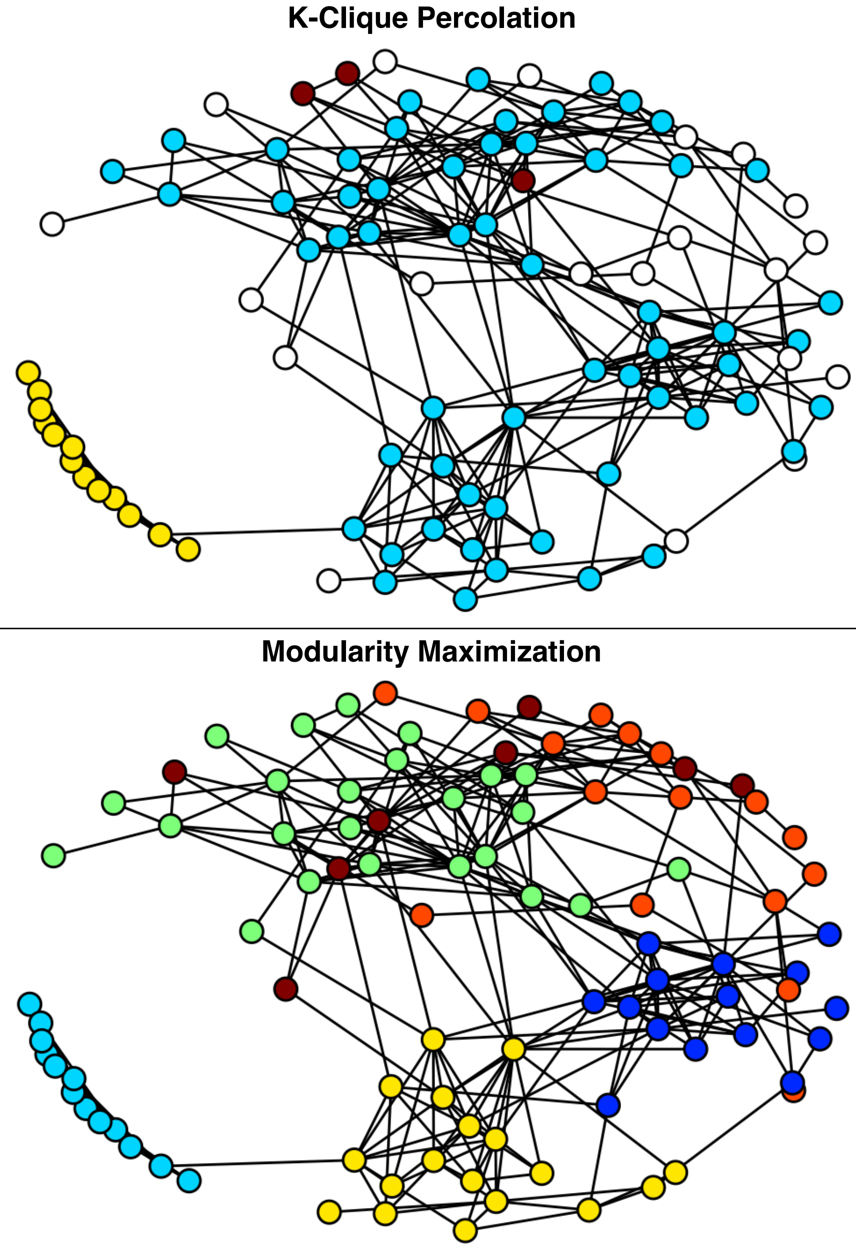
Figure 2.2. Comparison of Family Income Distributions Across Saturated Sample Schools



Network Community Detection and Segregation Evaluation

As expected, the two network community detection algorithms provided very different partitions of the school networks. A particularly striking example of this for one school is shown in Figure 2.3. In these illustrations, one for k-clique percolation and the other for modularity maximization, each adolescent is represented as a node (or dot) and is color coded according to their network community membership, except for the k-clique percolation graph, in which white color indicates no network community membership (i.e., marginalization). Friendship ties are denoted as lines linking nodes. In the full sample, 195 network communities were detected using modularity maximization, and 383 were detected using k-clique percolation (using a clique size of k=3). A school-by-school breakdown of network communities is provided in Supplemental Table 2.4.

Figure 2.3. Visualizations of Detected Communities in One School Network Using Two Methods – K-clique Percolation and Modularity Maximization



Note: Nodes of the same color within a single visualization of the network indicates membership in the same network community. In the k-clique percolation visual, white nodes are not members of any community.

Despite the differences between k-clique percolation and modularity maximization in the network communities identified, the results of the segregation analyses conducted at the network community level yielded similar results. In the primary analysis using multilevel models (Table

2.2) we found consistent clustering at the network community level. The intraclass correlation coefficient (ICC)—calculated by dividing the network community-level random effect variance by the total variance and converting the result to a percentage—was a significant 4.4% for modularity maximization and 7.7% k-clique percolation after adjustment for school and race/ethnicity. This indicates two things. First, family income does appear to cluster at the network community level—some network communities tend to have wealthier members, and some network communities have less affluent members. Second, however, there is still tremendous variation of family income *within* network communities, indicating that while there is some evidence of clustering at the network community level, it is in no way an exclusionary type of income segregation with perfect assortativity (e.g., where rich only associates with rich, and poor with poor).

Table 2.2. Random Effects from Multilevel Models of Students nested in Network Communities

	Model 1		Model 2	
	Estimate	95% Credible Interval	Estimate	95% Credible Interval
Modularity Maximization				
Between	49.66	[26.87 , 80.93]	34.66	[15.92 , 60.50]
Within	726.52	[719.17 , 808.62]	757.44	[714.07 , 803.06]
DIC	*22453.64*		*22409.52*	
ICC	**6.40 %**		**4.38 %**	
K-clique Percolation [a]				
Between	76.12	[30.82 , 131.96]	60.67	[14.22 , 116.15]
Within	734.96	[679.37 , 795.00]	734.31	[678.31 , 794.55]
DIC	*14636.46*		*14625.76*	
ICC	**9.39 %**		**7.73 %**	

Notes: DIC = Deviance Information Criterion; smaller value indicates better model fit. ICC = intraclass correlation coefficient; indicates the percent of total variance attributable to the "between" level.
Model 1: adjusted for school ID.
Model 2: adjusted for school ID and race/ethnicity.
[a] Analyzed sample included 15 of the 16 schools because school "A" had no k-cliques.

This second result is consistent with our findings from the secondary analyses of network communities, in which we focused on network communities of moderate and large size and evaluated whether the within group homogeneity of income was less than expected based on the background level of variability. Using this definition of segregation, no segregation by family income at the network community scale was detected (Supplemental Tables 2.5 and 2.6) because network communities were sufficiently heterogeneous with respect to income.

Because of the possibility that students in larger network communities might somehow differ in expected income variation from students not included in larger communities, we conducted sensitivity analyses whereby the repeated samples were drawn from various subsets of students in the two largest high schools. In the k-clique analysis we sampled in three additional ways: (1) only from students in larger cliques (10 or more students), (2) only from students belonging to cliques (of any size), and (3) only from students with two or more friends (who theoretically could have belonged to a clique of size 3). In modularity maximization, only the "larger cliques" alternative was used, because no students were considered marginalized as the method assigns all network nodes to communities. The result of the secondary analysis, that no segregation was detected at the network community level when defining segregation as within group homogeneity, was robust to all specifications.

Dyads Within versus Dyads Between Network Communities

Under the modularity maximization specification of network communities, in 8 of 16 schools the mean income difference was smaller for dyads in the same network community (dyads *within* communities) than for dyads where adolescents were in different network communities (dyads *between* communities) (Table 2.3). Similar results were found using k-clique network communities (Table 2.4). Of the 13 schools with sufficient numbers of *between* and *within* community dyads for meaningful comparisons, 6 schools had smaller income gaps for the dyads within the same k-clique network communities than in different network communities. Results were robust to all three alternative specifications of the k-clique analysis (Supplemental Tables 2.7 and 2.8).

With some exceptions, the difference in the income gap for *within* versus *between* tended to be modest though statistically significant. This indicates that, in many of the sample schools, adolescents were likely to encounter somewhat smaller income gaps between themselves and their fellow network community members than if they interacted with a randomly chosen individual from outside their network community.

Table 2.3. Mean Difference in Income for Dyads Within versus Dyads Between Network Communities – Modularity Maximization

	Within			Between			Different?		
School	N	Mean (X_1)	SD	N	Mean (X_2)	SD	($X_1 - X_2$)	t-test	p-value
A	25	**25.72**	16.84	965	27.94	21.33	-2.22	-0.6459	0.5240
B	298	38.86	51.52	1532	**34.89**	46.86	3.97	1.2346	0.2177
C	240	**30.00**	47.38	4320	30.83	36.06	-0.83	-0.2671	0.7896
D	1265	**20.47**	16.43	8746	22.86	18.27	-2.39	-4.7652	**<0.0001**
E	533	**20.80**	18.28	4123	21.22	19.85	-0.42	-0.4941	0.6214
F	1305	**30.00**	25.40	8425	32.18	27.39	-2.18	-2.8541	**0.0044**
G	19212	**29.09**	26.45	247603	30.29	27.14	-1.20	-6.0463	**<0.0001**
H	26483	**28.90**	27.51	717107	29.51	27.20	-0.61	-3.5450	**0.0004**
I	420	**21.14**	29.70	4431	24.20	32.23	-3.06	-2.0027	**0.0457**
J	514	**32.44**	45.51	2889	33.97	49.32	-1.53	-0.6932	0.4884
K	303	**18.35**	14.04	1975	20.16	15.79	-1.81	-2.0536	**0.0406**
L	28	**15.11**	9.48	92	17.12	12.80	-2.01	-0.8998	0.3719
M	225	**24.44**	25.21	1206	25.63	23.15	-1.19	-0.6582	0.5109
N	266	**46.05**	45.61	959	52.63	52.87	-6.58	-2.0083	**0.0452**
O	95	**28.00**	31.35	766	37.12	39.38	-9.12	-2.5930	**0.0106**
P	91	34.42	24.63	1184	**33.73**	25.57	0.69	0.2568	0.7978

Note: The difference in income (in thousands of dollars) is calculated for all possible dyads, regardless of whether they are linked or not. Dyads are sorted by whether both nodes in the pair lie within the same network community (Within) or in different network communities (Between). The mean for each group is calculated and compared using a two-sided t-test with significance 0.05. Bolded mean value indicates it is the smaller of the means for that school. Bolded p-values indicate statistical significance at 0.05.

Table 2.4. Mean Difference in Income for Dyads Within versus Dyads Between Network Communities – K-clique Percolation [a]

	Within			Between			Different?		
School	N	Mean (X_1)	SD	N	Mean (X_2)	SD	($X_1 - X_2$)	t-test	p-value
A [b]	0	—	—	990	27.88	21.22	—	—	—
B	235	**33.86**	47.54	1595	35.78	47.69	-1.92	-0.5778	0.5638
C	40	**25.84**	17.21	4520	30.83	36.86	-4.99	-1.7977	0.0794
D	979	**20.99**	16.10	9032	22.73	18.26	-1.74	-3.1679	**0.0016**
E	402	**18.31**	19.54	4254	21.44	19.66	-3.13	-3.0683	**0.0023**
F	603	31.89	26.80	9127	31.89	27.17	0.00	< 0.0001	> 0.9999
G	30467	**29.59**	26.36	236348	30.28	27.19	-0.69	-4.2846	**< 0.0001**
H	3716	**27.71**	27.04	739874	29.50	27.21	-1.79	-4.0252	**0.0001**
I	303	**16.68**	23.36	4548	24.42	32.47	-7.74	-5.4287	**< 0.0001**
J	1107	**22.93**	19.87	2296	38.95	57.01	-16.02	-12.0338	**< 0.0001**
K	116	21.78	14.86	2162	**19.82**	15.61	1.96	1.3803	0.1699
L	55	**16.07**	11.73	65	17.14	12.47	-1.07	-0.4837	0.6295
M	256	26.11	25.68	1175	**25.29**	22.98	0.82	0.4714	0.6376
N	299	52.00	50.06	926	**50.94**	51.89	1.06	0.3155	0.7525
O [b]	3	—	—	858	36.20	38.71	—	—	—
P [b]	3	—	—	1272	33.79	25.52	—	—	—

Note: The difference in income (in thousands of dollars) is calculated for all possible dyads, regardless of whether they are linked or not. Dyads are sorted by whether both nodes in the pair lie within the same network community (Within) or in different network communities (Between). The mean for each group is calculated and compared using a two-sided t-test with significance 0.05. Bolded mean value indicates it is the smaller of the means for that school. Bolded p-values indicate statistical significance at 0.05.

[a] In this specification, dyads where one or both of the pair is “marginalized” (i.e., not a member of any network community) are categorized as “between.”

[b] School contained no k-clique network communities or insufficient k-clique communities for a meaningful comparison.

Dyadic Level Income Homophily

Table 2.5 presents the results of the dyadic level analysis in which we model difference of family income level between pairs of adolescents to predict the likelihood of friendship, adjusted for dyads being of the 'same race/ethnicity.' Results from the unadjusted model are provided in Supplemental Table 2.9. Income homophily was detected in 5 of the 16 schools after adjusting for racial/ethnic homophily. Half of the 16 schools showed dyadic level homophily of some form—three by race/ethnicity, three by income, and two by both income and race/ethnicity. Most of the schools that showed income homophily in the unadjusted model continued to do so after controlling for race, with only one exception—school "H," the large and racially diverse high school. Initially, school "H" appeared to have homophily by income level at the dyadic level, but this effect disappeared when racial/ethnic homophily was controlled for.

Table 2.5. Dyadic Level Income Segregation – Logit Models of Difference in Income Between Dyad Pairs and 'Same Race' Indicator Predicting Probability of a Tie Existing

School ID	Parameter	Estimate	SE	*Z*-score	*P*-value
A	Intercept	-4.4407	0.5257	-8.4472	<.0001
	Income Difference	0.0034	0.0133	0.2556	0.7969
	Same Race	-0.1646	0.6161	-0.2672	0.7893
B	Intercept	-3.9665	0.5850	-6.7803	<.0001
	Income Difference	-0.0033	0.0025	-1.3200	0.1991
	Same Race	1.2985	0.5922	2.1927	0.0283 *
C	Intercept	-4.1678	0.5117	-8.1450	<.0001
	Income Difference	-0.0028	0.0038	-0.7368	0.4584
	Same Race	0.1092	0.5187	0.2105	0.8332
D	Intercept	-3.9125	0.5127	-7.6312	<.0001
	Income Difference	-0.0127	0.0040	-3.1750	0.0013 **
	Same Race	0.5391	0.5082	1.0608	0.2888
E	Intercept	-3.4648	0.1923	-18.0177	<.0001
	Income Difference	-0.0001	0.0043	-0.0233	0.9773
	Same Race	0.0394	0.1994	0.1976	0.8434
F	Intercept	-4.4213	0.1549	-28.5429	<.0001
	Income Difference	-0.0098	0.0028	-3.5000	0.0005 ***
	Same Race	1.8991	0.1540	12.3318	<.0001 ***
G	Intercept	-4.7527	0.0684	-69.4839	<.0001
	Income Difference	-0.0025	0.0008	-3.1250	0.0029 **
	Same Race	0.0597	0.0685	0.8715	0.3837
H	Intercept	-7.2220	0.0591	-122.1997	<.0001
	Income Difference	-0.0006	0.0009	-0.6667	0.5111
	Same Race	2.1272	0.0589	36.1154	<.0001 ***
I	Intercept	-3.9351	0.3410	-11.5399	<.0001
	Income Difference	-0.0008	0.0028	-0.2857	0.7713
	Same Race	0.4608	0.3484	1.3226	0.1859
J	Intercept	-3.7519	0.4536	-8.2714	<.0001
	Income Difference	-0.0044	0.0019	-2.3158	0.0188 *
	Same Race	1.2428	0.4578	2.7147	0.0066 **
K	Intercept	-2.7168	0.3797	-7.1551	<.0001
	Income Difference	-0.0017	0.0074	-0.2297	0.8214
	Same Race	-0.5820	0.3857	-1.5089	0.1313
L [a]	Intercept	-1.4470	0.4052	-3.5711	0.0004
	Income Difference	-0.0063	0.0203	-0.3103	0.7551
	Same Race	0	.	.	.
M	Intercept	-2.8986	0.1858	-15.6006	<.0001
	Income Difference	0.0004	0.0045	0.0889	0.9328
	Same Race	1.0957	0.2004	5.4676	<.0001 ***
N	Intercept	-1.9688	0.1708	-11.5269	<.0001
	Income Difference	-0.0044	0.0021	-2.0952	0.0408 *
	Same Race	0.0431	0.1894	0.2276	0.8200
O	Intercept	-2.3547	0.6123	-3.8457	0.0001
	Income Difference	-0.0083	0.0060	-1.3833	0.1649
	Same Race	-0.5865	0.6285	-0.9332	0.3507
P	Intercept	-3.8748	0.7476	-5.1830	<.0001
	Income Difference	-0.0113	0.0074	-1.5270	0.1258
	Same Race	0.8298	0.7321	1.1335	0.2570

Evidence of homophily (one tail test) at: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

[a] School was 100% white, making estimates of the 'same race' parameter infeasible.

Discussion

In this study we approached the evaluation of adolescent social network segregation by family income level in a novel way—through explicit consideration of the scale at which segregation occurs and the varying definitions of segregation which might be used. A summary of all analyses conducted and the results at each level is presented in Table 2.6.

Social network communities, which may serve as reasonable proxies for groups that socialize together, exhibited sizeable within-group family income heterogeneity. This was observed in both the multilevel analysis and the analysis comparing the empirical variation within network communities with the variation expected for network communities of comparable size. The analysis of within-group variation employed a far stricter definition of segregation—in order for a network community to appear segregated under this definition, it would have to have been highly homogeneous and sequestered from the variation in income observed in the rest of the school. This extreme form of segregation was not seen, however the multilevel analysis revealed that network communities differ by income level. A modest, though meaningful percent of total income variation in the sample occurs at the between-network community level (ICCs adjusted for race/ethnicity: modularity maximization=4.4%; k-clique=7.7%).

Table 2.6. Summary of Findings

Level	Analysis	Results
School Level	Distribution (mean and SD) of family income across the 16 sample schools was plotted.	There is clear overlap between the distributions, however the means and SD vary substantially. Since this is a small and non-representative sample the school level is addressed only to understand the sample, and not to evaluate between-school segregation. *See Figure 2.1 and Supplemental Table 2.2.*
Network Community Level	(1) Multilevel analysis: adolescents are nested within network communities, detected using modularity maximization and k-clique percolation algorithms.	(1) Variation between network communities was detected after controlling for school and race/ethnicity, supporting the claim that network communities are segregated. However clustering of family income status at the network community level is not *exclusionary* (i.e., rich associates only with rich, and poor with poor). Instead, there is substantial income heterogeneity within network communities. *See Table 2.2.*
	(2) Large and moderate sized network communities: variation of income within network communities is compared with variation expected in a network community of that size.	(2) No evidence of significantly lower variance of income within network communities than might have been expected if network communities were composed of randomly selected individuals in the network. *See Supplemental Tables 2.5 and 2.6.*
Dyads Within vs. Dyads Between Network Communities Level	Mean difference in income was calculated for two groups: dyads where both nodes are in the same network community (*within*), and dyads where nodes are in different network communities (*between*). Means are compared using two-sided t-tests.	In many—though not all—schools, the mean income difference was smaller for nodes in the same network communities (*within*) than for nodes in different network communities (*between*). The reverse was never true. *See Tables 2.3 and 2.4, and Supplemental Tables 2.7 and 2.8.*
Dyadic Level	Logistic models of *difference in income* between adolescents in dyads predicting likelihood of a tie existing between them. Models were fit separately for each school and control for homophilous race/ethnicity.	In some—though not all—schools, a smaller difference in income between any given pair of adolescents increases the likelihood of a friendship tie existing between them. Result is generally robust to adjustment for pairs being homophilous by race/ethnicity. *See Table 2.5 and Supplemental Table 2.9.*

The analysis of dyads sorted into dyads *within* network communities and dyads *between* network communities supports this finding of modest income level similarity between adolescents in the same network communities. Zooming in on the network further and disregarding the borders of network communities entirely, we found that income homophily exists at the dyadic level in 5 of the 16 schools, after the effects of racial/ethnic homophily were controlled for. Interestingly, of the 16 schools in the sample, income homophily appeared to be just as prevalent as racial homophily. Networks in three schools showed racial/ethnic homophily, three showed income homophily, and two showed both types of homophily. While the sample of 16 schools is insufficient to generalize any conclusions to a population-level, these results do indicate that income may be—in some schools—an alternative or additional consideration to race and ethnicity in the selection of close friends. It is intriguing, for example, that in the large high school ("G") with very little racial diversity we see income homophily, while in the large high school ("H") with significant racial diversity, we see racial homophily but not income homophily.

Taken together, these results indicate that segregation by family income level—defined in a variety of ways—does appear to exist in many adolescent school-based social networks. Furthermore, we find evidence that income segregation exists on all three levels of analysis. The segregation that appears to be occurring is not, however, extreme in the sense of adolescents forming social groups of homophilous income levels and then socially sequestering themselves. There is, in fact, income diversity within network communities and individual-level friend choices. Instead, we see a consistent, though modest preference at the dyadic level for friends with similar income levels, and also a tendency for network communities to bring together adolescents of more similar income levels than we might expect to see if we selected two adolescents at random from different network communities.

Each level analyzed represented a qualitatively different aspect of the social world. Evaluating homophily at the dyadic level implies that selection of particularly close social relationships is of upmost concern. This level is what is typically addressed in studies of segregation in social networks and it is the only level about which a prior literature on socioeconomic network segregation among adolescents existed. The best understood aspect of socioeconomic salience in adolescent social networks was with respect to romantic and sexual partners. In both older (Cohen 1979, Hollingshead 1949) and more recent data (Bearman, Moody and Stovel 2004) researchers have found evidence that U.S. adolescents are more likely to select partners with similar socioeconomic backgrounds. Some sparse literature (Maharaj and Connolly 1994, Mouw and Entwisle 2006) has suggested that homophily may also exist at the dyadic level among close friends, but these results are not consistent across all data sets and countries (Burgess, Sanderson and Umana-Aponte 2011). This study confirms the findings of Mouw and Entwisle, who also analyzed the Add Health sample, however this study adds to our understanding of network segregation by family income level because we place income homophily front and center in the analyzes, as opposed to treating it as a covariate of racial/ethnic homophily. Based on these findings, we encourage future researchers to consider socioeconomic homophily in evaluations of social networks, both independently from and in conjunction with racial/ethnic homophily.

We also encourage researchers to critically evaluate what types of relationships are captured in dyadic level data (and, indeed, in sociocentric data analyzed at higher levels as well). For instance, social networks can be constructed to show only romantic relationships, familial relationships, or friendships. Even the number of nominations survey respondents are allowed to make can affect what is captured. When respondents are encouraged to provide nominations of an unlimited number of social contacts, these nominations may include friendly acquaintances, whereas enforcing particularly strict limits on the number of nominations may restrict respondents to nominating only the closest connections. While the Add Health survey's limit of 10 nominations (5 male and 5 female) is not particularly strict, it did place an upper bound on the number of social contacts adolescents could nominate. Adolescents were therefore encouraged to be more selective in their friend nominations than if they had been allowed to nominate freely. This survey choice helps to create the qualitative differences between the dyadic level and the higher levels.

While the dyadic level has been previously evaluated, albeit sparsely, no studies of which we are aware have analyzed other scales at which socioeconomic segregation may exist in adolescent networks. As mentioned, addressing these scales implies that a fundamentally different aspect of social networks is of interest. Social network communities can, and often do, bring together individuals in a dense social fabric who may or may not be particularly close. Yet because they are bound tightly by the network community within which they are both embedded, these individuals will be more likely to interact with each other—even if only in a group setting— more regularly than if they belonged to different network communities. Put another way, individuals experience a social world shaped by people other than those they would list as being among their closest friends. An exclusive focus in the literature on the dyadic level will therefore neglect these other dimensions of the social world.

While evidence of income segregation was found at all three levels, there was no *a priori* reason to expect this concurrence across levels of analysis. By differentiating the levels at which segregation in adolescent social networks is occurring we have gained a richer understanding for how network structure reflects socioeconomic patterning. Based on this, we encourage researchers to explicitly consider multiple scales of analysis when evaluating segregation in social networks.

Limitations

One limitation of the present study concerns missing data. It is common for income-related items in survey research to have low response rates, and therefore the 70% response rate on the income item among parents surveyed was acceptable. A comparison of the demographic profiles of respondents versus refusers is provided in Supplemental Table 2.1. Another limitation concerns the completeness of the adolescent social networks. Students were restricted to nominating 5 male and 5 female friends which may not reflect the true size or diversity of students' relevant social groups, and the data did not include information about romantic relationship ties. The data was also unable to capture the diversity of networks that adolescents may have in other settings—such as beyond the reach of the Add Health sample (e.g., friendships formed during recreational activities outside of school, or with others from their neighborhood). Finally, only 16 schools were included in the saturated sample, limiting the generalizability of our findings. Despite the inherent limitations of the data, the saturated school Add Health sample is one of the few data sets with both rich network and family income data among adolescents, enabling a rare look into the structuring of adolescent social networks by socioeconomic status.

Significance

This study has addressed a major substantive gap in our understanding of income segregation in adolescent social networks. Furthermore it has demonstrated that when evaluating segregation in networks, it is critical to consider what scale the segregation may be occurring at, and to critically evaluate what definitions of segregation are being employed, as conclusions may differ widely depending on what choices are made. One key insight of this study is that while income segregation at the dyadic level is occurring among adolescents in some schools, school environments with sufficient socioeconomic diversity appear to foster friendship formation across socioeconomic levels, at both the dyadic and network community levels.

A second key insight is that family income does appear to have some social salience in the structure of adolescent social networks. The social processes that give rise to network structures with apparent income segregation are unclear. For instance, the simplest explanation is that adolescents are socially conscious of the material markers of family income and that they attach social significance, such as status, to these makers and select friends accordingly. Alternatively, adolescents may select friends based on preference for homophily by some other factor. By controlling for racial/ethnic homophily at both the dyadic level and the network community level, we find that preference for same-race friendships may partially explain the income homophily and segregation observed, yet it does not entirely account for what is observed. It is possible, however, that there are other unmeasured factors that could account for income segregation in adolescent networks. For instance, family income may be associated with adolescents' financial ability to participate in extracurricular activities (Snellman et al. 2015), or alternatively family income may be associated with adolescents' preferences for particular extracurricular activities (Bourdieu 1984). Adolescent networks may therefore become socioeconomically structured as a result of adolescents befriending others who share similar activities and interests.

In future studies of the social processes involved in creating segregation, and the health and educational implications of segregation, family income and other measures of socioeconomic status are worthy of consideration alongside race and ethnicity.

REFERENCES

Acevedo-Garcia, Dolores and Kimberly A. Lochner. 2003. "Residential Segregation and Health." in *Neighborhoods and Health*, edited by I. Kawachi and L. F. Berkman. New York, NY: Oxford University Press.

Ajrouch, Kristine J., Alysia Y. Blandon and Toni C. Antonucci. 2005. "Social Networks among Men and Women: The Effects of Age and Socioeconomic Status." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 60(6):S311-S17.

Ali, Mir M. and Debra S. Dwyer. 2010. "Social Network Effects in Alcohol Consumption among Adolescents." *Addictive Behaviors* 35(4):337-42.

Anaconda by Continuum Analytics. 2015. "Python 2.7." http://continuum.io/.

Aneshensel, Carol S. and Clea A. Sucoff. 1996. "The Neighborhood Context of Adolescent Mental Health." *Journal of Health and Social Behavior* 37(4):293-310.

Bearman, Peter S., James Moody and Kathrine Stovel. 2004. "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks." *American Journal of Sociology* 110(1):44-91.

Bian, Yanijie, Ronald Breiger, Deborah Davis and Joseph Galaskiewicz. 2005. "Occupation, Class, and Social Networks in Urban China." *Social Forces* 83(4):1443-68.

Bjarnason, Thoroddur. 2000. "Grooming for Success? The Impact of Adolescent Society on Early Intergenerational Social Mobility." *Journal of Family and Economic Issues* 21(4):319-42.

Blau, Peter M., Danching Ruan and Monika Ardelt. 1991. "Interpersonal Choice and Networks in China." *Social Forces* 69(4):1037-62.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics* P10008.

Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgment of Taste*. Translated by R. Nice. Cambridge, MA: Harvard University Press.

Browne, WJ. 2009. *Mcmc Estimation in Mlwin V2.1*: Centre for Multilevel Modelling, University of Bristol.

Burgess, Simon, Eleanor Sanderson and Marcela Umana-Aponte. 2011. "School Ties: An Analysis of Homophily in an Adolescent Friendship Network. Working Paper No. 11/267." The Centre for Market and Public Organisation, Bristol Institute of Public Affairs, University of Bristol.

Christakis, Nicholas A. and James H. Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *New England Journal of Medicine* 357(4):370-79.

Christakis, Nicholas A. and James H. Fowler. 2008. "The Collective Dynamics of Smoking in a Large Social Network." *New England Journal of Medicine* 358(21):2249-58.

Cohen, Jere. 1979. "Socio-Economic Status and High-School Friendship Choice: Elmtown's Youth Revisited." *Social Networks* 2(1):65-74.

Cohen, Sheldon, Denise Janicki-Deverts and Gregory E. Miller. 2007. "Psychological Stress and Disease.". *The Journal of the American Medical Association* 298(14):1685-87.

Coleman, James S. 1961. *The Adolescent Society*. Glencoe, Illinois: Free Press.

Crosnoe, Robert. 2009. "Low-Income Students and the Socioeconomic Composition of Public High Schools." *American Sociological Review* 74(5):709-30.

Csardi, Gabor and Tamas Nepusz. 2006. "The Igraph Software Package for Complex Network Research." InterJournal, Complex Systems 1695, http://igraph.org.

Currarini, Sergio, Matthew O. Jackson and Paolo Pin. 2010. "Identifying the Roles of Race-Based Choice and Chance in High School Friendship Network Formation." *Proceedings of the National Academy of Sciences USA* 107(11):4857-61.

DiPrete, Thomas A., Andrew Gelman, Tyler McCormick, Julien Teitler and Tian Zheng. 2011. "Segregation in Social Networks Based on Acquaintanceship and Trust." *American Journal of Sociology* 116(4):1234-83.

Emmons, Karen M. 2000. "Health Behaviors in a Social Context." in *Social Epidemiology*, edited by L. F. Berkman and I. Kawachi. New York, NY: Oxford University Press, Inc.

Erbe, Brigitte M. 1975. "Race and Socioeconomic Segregation." *American Sociological Review* 40(6):801-12.

Farley, Reynolds and Alma F. Taeuber. 1974. "Racial Segregation in the Public Schools." *American Journal of Sociology* 79(4):888-905.

Fiel, Jeremy E. 2013. "Decomposing School Resegregation: Social Closure, Racial Imbalance, and Racial Isolation." *American Sociological Review* 78(5):828-48.

Forero, Roberto, Lyndall McLellan, Chris Rissel and Adrian Bauman. 1999. "Bullying Behavior and Psychosocial Health among School Students in New South Wales, Astralia: Cross Sectional Survey." *BMJ* 319:344-48.

Fortunato, Santo. 2010. "Community Detection in Graphs." *Physics Reports* 486:75-174.

Fuller-Rowell, Thomas E., Gary W. Evans and Anthony D. Ong. 2012. "Poverty and Health: The Mediating Role of Perceived Discrimination." *Psychological Science* 23(7):734-39.

González, MC, HJ Herrmann, J Kertész and T Vicsek. 2007. "Community Structure and Ethnic Preferences in School Friendship Networks." *Physica A* 379(1):307-16.

Hagberg, AA, DA Schult and PJ Swart. 2008. "Exploring Network Structure, Dynamics, and Function Using Networkx." Paper presented at the Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA.

Hanson, Margaret D. and Edith Chen. 2007. "Socioeconomic Status and Health Behaviors in Adolescence: A Review of the Literature." *Journal of Behavioral Medicine* 30(3):263-85.

Harris, KM, CT Halpern, E Whitsel, J Hussey, J Tabor, P Entzel and JR Udry. 2009, "The National Longitudinal Study of Adolescent Health: Research Design [WWW Document]". (URL: http://www.cpc.unc.edu/projects/addhealth/design.).

Hollingshead, August B. 1949. *Elmtown's Youth: The Impact of Social Classes on Adolescents*. New York: John Wiley.

Hunter, JD. 2007. "Matplotlib: A 2D Graphics Environment." *Computing In Science and Engineering* 9(3):90-95.

James, David R. 1989. "City Limits on Racial Equality: The Effects of City-Suburb Boundaries on Public-School Desegregation, 1968-1976." *American Sociological Review* 54(6):963-85.

King, Morton B., Jr. 1961. "Socioeconomic Status and Sociometric Choice." *Social Forces* 39(3):199-206.

Logan, John R., Deirdre Oakley and Jacob Stowell. 2008. "School Segregation in Metropolitan Regions, 1970-2000: The Impacts of Policy Choices on Public Education." *American Journal of Sociology* 113(6):1611-44.

Maharaj, Sherry I. and Jennifer A. Connolly. 1994. "Peer Network Composition of Acculturated and Ethnoculturally-Affiliated Adolescents in a Multicultural Setting." *Journal of Adolescent Research* 9(2):218-40.

Marques, Eduardo. 2012. "Social Networks, Segregation and Poverty in Sao Paulo." *International Journal of Urban and Regional Research* 36(5):958-79.

Massey, Douglas S. and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Massey, Douglas S. 2004. "Segregation and Stratification: A Biosocial Perspective." *Du Bois Review* 1(1):7-25.

McKinney, W. 2011. "Python Pandas." AQR Capital Management, Lambda Foundry Inc, http://pandas.pydata.org/.

McPherson, M, L Smith-Lovin and J Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27(1):415-44.

Moody, James. 2001. "Race, School Integration, and Friendship Segregation in America." *American Journal of Sociology* 107(3):679-716.

Moody, James and Douglas R. White. 2003. "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups." *American Sociological Review* 68(1):103-27.

Mouw, T and B Entwisle. 2006. "Residential Segregation and Interracial Friendships in Schools." *American Journal of Sociology* 112(2):394-441.

Nansel, Tonja R., Mary Overpeck, Ramani S. Pilla, W. June Ruan, Bruce Simons-Morton and Peter Scheidt. 2001. "Bullying Behaviors among Us Youth: Prevalence and Association with Psychosocial Adjustment." *The Journal of the American Medical Association* 285(16):2094-100.

Neal, Jennifer W. 2010. "Hanging Out: Features of Urban Children's Peer Social Networks." *Journal of Social and Personal Relationships* 27(7):982-1000.

Newman, MEJ. 2006a. "Finding Community Structure in Networks Using the Eigenvectors of Matrices." *Physical Review E* 74(036104).

Newman, MEJ. 2006b. "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences USA* 103:8577-82.

NumPy Developers. 2005. "Numpy User Guide." http://www.numpy.org/.

Palla, Gergely, Imre Derényi, Illés Farkas and Tamás Vicsek. 2005. "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society." *Nature* 435(814-818).

Porter, Mason A., Jukka-Pekka Onnela and Peter J. Mucha. 2009. "Communities in Networks." *Notices of the American Mathematical Society* 56(9):1082-97.

Quillian, Lincoln and Mary E. Campbell. 2003. "Beyond Black and White: The Present and Future of Mutliracial Friendship Segregation." *American Sociological Review* 68(4):540-66.

Rasbash, J, WJ Browne, M Healy, B Cameron and C Charlton. 2015. *MLwiN Version 2.32*: Centre for Multilevel Modelling, University of Bristol.

Reardon, Sean F. and Kendra Bischoff. 2011. "Income Inequality and Income Segregation." *American Journal of Sociology* 116(4):1092-153.

Ross, Catherine E. and Chia-ling Wu. 1995. "The Links between Education and Health." *American Sociological Review* 60(5):719-45.

SAS Institute Inc. 2011. "SAS/AF ® 9.3 Procedure Guide." Cary, NC: SAS Institute Inc.

Schofield, Janet W. and Bernard E. Whitley Jr. 1983. "Peer Nomination Vs. Rating Scale Measurement of Children's Peer Preferences." *Social Psychology Quarterly* 46(3):242-51.

Smith, Jeffrey A., Miller McPherson and Lynn Smith-Lovin. 2014. "Social Distance in the United States: Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004." *American Sociological Review* 79(3):432-56.

Snellman, Kaisa, Jennifer M. Silva, Carl B. Frederick and Robert D. Putnam. 2015. "The Engagement Gap: Social Mobility and Extracurricular Participation among American Youth ". *The ANNALS of the American Academy of Political and Social Science* 657(1):194-207.

Sourander, Andre, Peter Jensen, John A. Rönning, Solja Niemelä, Hans Helenius, Lauri SIllanmäki, Kirsti Kumpulainen, Jorma Piha, Tuula Tamminen, Irma Moilanen and Fredrik Almqvist. 2007. "What Is the Early Adulthood Outcome of Boys Who Bully or Are Bullied in Childhood? The Finnish "from a Boy to a Man" Study." *Pediatrics* 120(2):397-404.

Tomaskovic-Devey, Donald, Catherine Zimmer, Kevin Stainback, Corre Robinson, Tiffany Taylor and Tricia McTague. 2006. "Documenting Desgregation: Segregation in American Workplaces by Race, Ethnicity, and Sex, 1966-2003." *American Sociological Review* 71(4):565-88.

UNC Carolina Population Center project. "Program Code for Race". Retrieved 09 Feb 2014, (http://www.cpc.unc.edu/projects/addhealth/data/code/race).

Williams, David R. and Selina A. Mohammed. 2009. "Discrimination and Racial Disparities in Health: Evidence and Needed Research." *Journal of Behavioral Medicine* 32(1):20-47.

Willms, J. Douglas. 1986. "Social Class Segregation and Its Relationship to Pupils' Examination Results in Scotland." *American Sociological Review* 51(2):224-41.

Wilson, Alan B. 1959. "Residential Segregation of Social Classes and Aspirations of High School Boys." *American Sociological Review* 24(6):836-45.

Wilson, Travis and Philip C. Rodkin. 2011. "African American and European American Children in Diverse Elementary Classrooms: Social Integration, Social Status, and Social Behavior." *Child Dev* 82(5):1454-69.

Winkleby, Marilyn A., Darius E. Jatulis, Erica Frank and Stephen P. Fortmann. 1992. "Socioeconomic Status and Health: How Education, Income, and Occupation Contribute to Risk Factors for Cardiovascular Disease." *American Journal of Public Health* 82(6):816-20.

APPENDIX A: CHAPTER 2 SUPPLEMENTAL TABLES

Supplemental Table 2.1. Comparison of Demographic Profiles Across Sample Subsets

	Main Sample		**Surveyed**		**Nominated in Network**	
	Provided Income		***Refused or Missing***		***Not Surveyed***	
	Frequency *N* (%)		Frequency *N* (%)		Frequency *N* (%)	
TOTAL	2584	(100.00%)	1118	(100.00%)	793	(100.00%)
Sex of Adolescent						
Female	1232	(47.68%)	579	(51.79%)	399	(50.32%)
Male	1352	(52.32%)	539	(48.21%)	394	(49.68%)
Race / Ethnicity of Adolescent						
Missing	3	(0.12%)	1	(0.09%)	—	
Hispanic or Latino	437	(16.91%)	317	(28.38%)	—	
Black, Non-Hispanic	385	(14.90%)	185	(16.55%)	—	
Asian, Non-Hispanic	300	(11.61%)	259	(23.17%)	—	
Native American, Non-Hispanic	44	(1.70%)	11	(0.98%)	—	
Other, Non-Hispanic	15	(0.58%)	11	(0.98%)	—	
White, Non-Hispanic	1400	(54.18%)	334	(29.87%)	—	
Highest Education Level of Parents						
Missing	0	(0.00%)	32	(2.86%)	—	
Less than high school	270	(10.45%)	199	(17.80%)	—	
Completed high school	675	(26.12%)	312	(27.91%)	—	
Some college no degree	830	(32.12%)	234	(20.93%)	—	
College degree or more	809	(31.31%)	341	(30.50%)	—	

Supplemental Table 2.2. Socioeconomic Profile of Saturated Sample Schools—Family Income and the Education Level of the Parent or Guardian with Highest Education

		Family Income [a]		Highest Education Level of Parent or Guardian			
School	**Frequency *N***	**Mean**	**SD**	**Less than high school**	**High school degree**	**Some college**	**College or more**
TOTAL	3702	43.86	29.63	12.67 %	26.66 %	28.74 %	31.06 %
A	45	36.70	24.65	20.00 %	40.00 %	26.67 %	13.33 %
B	61	50.23	39.91	0.00 %	9.84 %	40.98 %	49.18 %
C	123	51.40	34.06	2.44 %	34.96 %	33.33 %	28.46 %
D	161	29.89	19.04	13.04 %	55.28 %	22.36 %	9.32 %
E	111	31.14	20.35	7.21 %	56.76 %	21.62 %	13.51 %
F	152	53.55	29.45	3.29 %	16.45 %	32.89 %	47.37 %
G	832	49.03	27.89	4.09 %	32.81 %	33.53 %	28.85 %
H	1721	40.19	25.90	21.21 %	21.03 %	25.62 %	30.80 %
I	99	40.61	29.05	1.01 %	10.10 %	31.31 %	57.58 %
J	91	43.11	42.97	12.09 %	49.45 %	27.47 %	9.89 %
K	83	28.66	17.47	3.61 %	12.05 %	25.30 %	59.04 %
L	20	38.50	14.53	0.00 %	30.00 %	50.00 %	20.00 %
M	55	32.08	24.91	9.09 %	40.00 %	32.73 %	18.18 %
N	52	93.22	52.90	0.00 %	3.85 %	42.31 %	53.85 %
O	43	43.20	30.94	6.98 %	13.95 %	37.21 %	41.86 %
P	53	67.77	26.47	1.89 %	13.21 %	24.53 %	60.38 %

Note: This table profiles only the Add Health sample, and does not include students who were nominated as members of the network but were not surveyed.

[a] In thousands of dollars.

Supplemental Table 2.3. Demographic Profile of Saturated Sample Schools—Sex and Race/Ethnicity

		Sex	Race / Ethnicity [a]					
School	***N***	**% Female**	**% White**	**% Black**	**% Hispanic**	**% Asian**	**% Native American**	**% Other**
TOTAL	3702	48.92%	46.84%	15.40%	20.37%	15.10%	1.49%	0.70%
A	45	48.89%	48.89%	13.33%	35.56%	2.22%	0.00%	0.00%
B	61	52.46%	95.08%	0.00%	4.92%	0.00%	0.00%	0.00%
C	123	40.65%	97.56%	0.00%	1.63%	0.00%	0.81%	0.00%
D	161	43.48%	98.14%	1.24%	0.62%	0.00%	0.00%	0.00%
E	111	50.45%	87.39%	1.80%	0.00%	0.90%	9.01%	0.90%
F	152	58.55%	42.11%	45.39%	10.53%	0.66%	0.66%	0.66%
G	832	47.48%	93.87%	0.24%	1.08%	0.96%	3.37%	0.48%
H	1721	48.17%	4.53%	23.18%	39.22%	31.49%	0.46%	0.87%
I	99	53.54%	94.95%	0.00%	1.01%	1.01%	3.03%	0.00%
J	91	52.75%	96.70%	0.00%	0.00%	0.00%	3.30%	0.00%
K	83	53.01%	0.00%	97.59%	2.41%	0.00%	0.00%	0.00%
L	20	55.00%	100.00%	0.00%	0.00%	0.00%	0.00%	0.00%
M	55	49.09%	41.82%	14.55%	30.91%	5.45%	1.82%	5.45%
N	52	53.85%	75.00%	1.92%	15.38%	3.85%	0.00%	3.85%
O	43	51.16%	97.67%	0.00%	2.33%	0.00%	0.00%	0.00%
P	53	66.04%	94.34%	0.00%	5.66%	0.00%	0.00%	0.00%

Note: This table profiles only the Add Health sample, and does not include students who were nominated as members of the network but were not surveyed.

[a] Totals to less than 100% per school due to missing responses to race/ethnicity questions.

Supplemental Table 2.4. Number of Network Communities by School and Detection Algorithm

School	Modularity Maximization	K-Clique Percolation
TOTAL	195	383
A	8	0
B	5	3
C	18	11
D	10	15
E	9	14
F	8	23
G	19	89
H	48	200
I	13	6
J	6	3
K	10	6
L	4	1
M	8	3
N	7	7
O	11	1
P	11	1

Note: The full sample was utilized in order to identify network communities. After loss of some observations due to missingness, a total of 181 modularity maximization network communities and 350 k-clique network communities were used in the analyses.

Supplemental Table 2.5. K-clique (k = 3) Community Detection Analysis of Income Segregation in Network Communities of 10 or more nodes using various samples to estimate p-values.

School ID	Clique ID	Clique Size *N* nodes	P-value: Sample = All Students in School Network	P-value: Sample = Students in Large Cliques (≥10) [a]	P-value: Sample = Students in Any Size Clique [a]	P-value: Sample = Students with Degree ≥ 2 [a]
B	1	10	0.255			
	2	20	0.457			
	3	20	0.457			
D	1	10	0.368			
	2	46	0.457			
E	1	26	0.419			
	2	34	0.436			
F	1	10	0.336			
	2	16	0.293			
	3	18	0.316			
	4	27	0.352			
G	1	10	0.306	0.263	0.296	0.283
	2	11	0.297	0.272	0.276	0.302
	3	11	0.297	0.272	0.276	0.302
	4	13	0.291	0.245	0.275	0.308
	5	14	0.292	0.298	0.269	0.312
	6	14	0.292	0.298	0.269	0.312
	7	15	0.285	0.279	0.287	0.292
	8	16	0.301	0.271	0.293	0.286
	9	16	0.301	0.271	0.293	0.286
	10	29	0.358	0.369	0.357	0.325
	11	33	0.343	0.395	0.337	0.364
	12	38	0.384	0.400	0.383	0.372
	13	56	0.403	0.447	0.395	0.388
	14	69	0.401	0.469	0.463	0.436
	15	276	0.483	0.489	0.454	0.458
H	1	10	0.296	0.192	0.308	0.296
	2	10	0.296	0.192	0.308	0.296
	3	10	0.296	0.192	0.308	0.296
	4	10	0.296	0.192	0.308	0.296
	5	11	0.308	0.174	0.265	0.321
	6	11	0.308	0.174	0.265	0.321
	7	13	0.334	0.167	0.275	0.291
	8	14	0.324	0.166	0.262	0.317
	9	17	0.322	0.162	0.266	0.317
	10	18	0.330	0.133	0.279	0.288
	11	19	0.317	0.145	0.256	0.311
	12	23	0.345	0.117	0.254	0.334
	13	37	0.349	0.131	0.249	0.310
	14	38	0.321	0.140	0.248	0.288
	15	39	0.346	0.136	0.244	0.303
	16	77	0.342	0.216	0.274	0.292

Continued

Supplemental Table 2.5 Continued

School ID	Clique ID	Clique Size *N* nodes	P-value: Sample = All Students in School Network	Sample = Students in Large Cliques (≥10) [a]	Sample = Students in Any Size Clique [a]	Sample = Students with Degree ≥ 2 [a]
J	1	13	0.344			
	2	58	0.628			
K	1	13	0.477			
	2	15	0.453			
L	1	16	0.501			
M	1	23	0.436			
N	1	19	0.441			
	2	21	0.452			

Note: Four schools (A, C, O and P) did not have any large cliques (size 10 nodes or more) and so were not included in this aspect of the analysis.

P-values were estimated using repeated sampling from the designated sample (1,000 draws) and then comparing the empirical variance to the expected variance distribution.

[a] Alternative sub-samples from the school were evaluated as a means of sensitivity testing. Because repeated sampling from students belonging to cliques can only be done in schools with many cliques, only schools G and H are included in these sensitivity tests.

Supplemental Table 2.6. Modularity Maximization Community Detection Analysis of Income Segregation in Network Communities of 10 or more nodes using various samples to estimate p-values.

School ID	Community ID	Community Size *N* nodes	P-value: Sample = All Students in School Network	P-value: Sample = Students in Large Communities (≥10) [a]
B	1	14	0.299	
	2	14	0.299	
	3	15	0.357	
	4	24	0.510	
C	1	10	0.152	
	2	10	0.152	
	3	10	0.152	
	4	13	0.179	
	5	14	0.176	
	6	15	0.174	
	7	18	0.239	
D	1	11	0.346	
	2	19	0.392	
	3	20	0.391	
	4	27	0.408	
	5	28	0.431	
	6	30	0.455	
	7	33	0.445	
E	1	17	0.388	
	2	21	0.381	
	3	21	0.381	
	4	24	0.422	
	5	25	0.424	
F	1	11	0.319	
	2	17	0.325	
	3	23	0.307	
	4	28	0.320	
	5	39	0.337	
	6	40	0.329	
G	1	32	0.345	0.365
	2	37	0.363	0.330
	3	44	0.364	0.367
	4	45	0.392	0.394
	5	57	0.398	0.362
	6	58	0.436	0.394
	7	64	0.419	0.412
	8	70	0.404	0.430
	9	75	0.414	0.426
	10	80	0.448	0.457
	11	81	0.456	0.462
	12	87	0.469	0.429
	13	89	0.469	0.412
	14	116	0.482	0.460
	15	116	0.482	0.460

Continued

Supplemental Table 2.6 Continued

School ID	Community ID	Community Size *N* nodes	P-value: Sample = All Students in Network	P-value: Sample = Students in Large Communities (≥10) [a]
H	1	35	0.338	0.352
	2	37	0.349	0.306
	3	38	0.321	0.337
	4	38	0.321	0.337
	5	41	0.353	0.337
	6	43	0.341	0.308
	7	51	0.340	0.311
	8	51	0.340	0.311
	9	53	0.344	0.328
	10	55	0.334	0.321
	11	58	0.324	0.320
	12	60	0.317	0.295
	13	66	0.341	0.317
	14	72	0.353	0.320
	15	77	0.342	0.332
	16	82	0.343	0.328
	17	84	0.318	0.320
	18	94	0.331	0.290
	19	96	0.354	0.290
	20	97	0.341	0.344
	21	98	0.317	0.313
	22	108	0.349	0.319
	23	110	0.330	0.353
	24	116	0.335	0.324
	25	122	0.347	0.336
I	1	14	0.207	
	2	15	0.234	
	3	16	0.262	
	4	17	0.270	
J	1	13	0.344	
	2	15	0.387	
	3	16	0.403	
	4	18	0.473	
	5	24	0.561	
K	1	10	0.422	
	2	16	0.462	
	3	23	0.491	
	4	33	0.494	
M	1	10	0.329	
	2	12	0.375	
	3	14	0.430	
	4	14	0.430	
N	1	11	0.418	
	2	11	0.418	
	3	18	0.428	
	4	22	0.441	
O	1	10	0.368	

Note: P-values were estimated using repeated sampling from the designated sample (1,000 draws) and then comparing the empirical variance to the expected variance distribution. Two schools (A and P) did not have any large communities (10 nodes or more) and so were not included.
[a] These analyses were conducted in schools G and H as a sensitivity test.

Supplemental Table 2.7. Mean Difference in Income for Dyads Within versus Dyads Between Network Communities – K-clique Percolation, Alternative Specification 1 [a]

	Within			Between			Different?		
School	*N*	Mean (X_1)	SD	*N*	Mean (X_2)	SD	($X_1 - X_2$)	t-test	p-value
A [b]	0	—	—	0	—	—	—	—	—
B	235	**33.86**	47.54	1319	37.10	49.31	-3.24	-0.9571	0.3392
C	40	**25.84**	17.21	2242	27.59	31.24	-1.75	-0.6250	0.5352
D	979	**20.99**	16.10	7379	22.69	17.67	-1.70	-3.0678	**0.0022**
E	402	**18.31**	19.54	3551	21.39	19.70	-3.08	-2.9929	**0.0029**
F	603	31.89	26.80	8532	**31.83**	27.34	0.06	0.0531	0.9577
G	30467	**29.59**	26.36	226895	30.39	27.25	-0.80	-4.9538	**< 0.0001**
H	3716	**27.71**	27.04	579413	29.86	28.18	-2.15	-4.8302	**< 0.0001**
I	303	**16.68**	23.36	2718	24.52	33.92	-7.84	-5.2568	**< 0.0001**
J	1107	**22.93**	19.87	2065	36.25	52.66	-13.32	-10.2173	**< 0.0001**
K	116	21.78	14.86	1301	**20.77**	15.97	1.01	0.6970	0.4869
L	55	**16.07**	11.73	55	16.47	12.52	-0.40	-0.1729	0.8630
M	256	26.11	25.68	944	**26.00**	23.86	0.11	0.0617	0.9508
N	299	52.00	50.06	905	**51.35**	52.30	0.65	0.1925	0.8474
O [b]	3	—	—	117	35.61	36.37	—	—	—
P [b]	3	—	—	144	29.16	22.06	—	—	—

Note: The difference in income (in thousands of dollars) is calculated for all possible dyads, regardless of whether they are linked or not. Dyads are sorted by whether both nodes in the pair lie within the same network community (Within) or in different network communities (Between). The mean for each group is calculated and compared using a two-sided t-test with significance 0.05. Bolded mean value indicates it is the smaller of the means for that school. Bolded p-values indicate statistical significance at 0.05.

[a] In this specification, dyads where both nodes are "marginalized" (i.e., not a member of any network community) are dropped from analysis. Pairs where one of the nodes is "marginalized" are categorized as "between."

[b] School contained no k-clique network communities or insufficient k-clique communities for a meaningful comparison.

Supplemental Table 2.8. Mean Difference in Income for Dyads Within versus Dyads Between Network Communities – K-clique Percolation, Alternative Specification 2 [a]

	Within			Between			Different?		
School	N	Mean (X_1)	SD	N	Mean (X_2)	SD	($X_1 - X_2$)	t-test	p-value
A [b]	0	—	—	0	—	—	—	—	—
B	235	**33.86**	47.54	431	40.86	55.04	-7.00	-1.7157	**0.0868**
C	40	25.84	17.21	338	**22.63**	16.84	3.21	1.1180	0.2691
D	979	**20.99**	16.10	2507	21.72	16.58	-0.73	-1.1930	0.2330
E	402	**18.31**	19.54	1309	21.71	20.12	-3.40	-3.0301	**0.0025**
F	603	31.89	26.80	4857	**31.51**	28.09	0.38	0.3266	0.7440
G	30467	**29.59**	26.36	145061	30.66	27.48	-1.07	-6.3931	**< 0.0001**
H	3716	**27.71**	27.04	209162	30.67	30.14	-2.96	-6.6006	**< 0.0001**
I	303	**16.68**	23.36	400	28.38	41.18	-11.70	-4.7605	**< 0.0001**
J	1107	22.93	19.87	723	**21.17**	17.26	1.76	2.0074	**0.0449**
K	116	21.78	14.86	209	**21.69**	16.75	0.09	0.0500	0.9602
L [b]	55	16.07	11.73	0	—	—	—	—	—
M	256	**26.11**	25.68	240	27.34	25.32	-1.23	-0.5370	0.5915
N	299	**52.00**	50.06	604	55.47	56.08	-3.47	-0.9413	0.3469
O [b]	3	—	—	0	—	—	—	—	—
P [b]	3	—	—	0	—	—	—	—	—

Note: The difference in income (in thousands of dollars) is calculated for all possible dyads, regardless of whether they are linked or not. Dyads are sorted by whether both nodes in the pair lie within the same network community (Within) or in different network communities (Between). The mean for each group is calculated and compared using a two-sided t-test with significance 0.05. Bolded mean value indicates it is the smaller of the means for that school. Bolded p-values indicate statistical significance at 0.05.

[a] In this specification, dyads where one or both nodes are "marginalized" (i.e., not a member of any network community) are dropped from analysis.

[b] School contained no k-clique network communities or insufficient k-clique communities for a meaningful comparison.

Supplemental Table 2.9. Dyadic Level Income Segregation – Logit Models of Difference in Income Between Dyad Pairs Predicting Probability of a Tie Existing

School ID	Parameter	Estimate	SE	*Z*-score	*P*-value
A	Intercept	-4.4966	0.4873	-9.2276	<.0001
	Income Difference	0.0034	0.0133	0.2556	0.8012
B	Intercept	-2.7567	0.1280	-21.5367	<.0001
	Income Difference	-0.0029	0.0025	-1.1600	0.2531
C	Intercept	-4.0657	0.1588	-25.6026	<.0001
	Income Difference	-0.0028	0.0038	-0.7368	0.4619
D	Intercept	-3.3804	0.0986	-34.2840	<.0001
	Income Difference	-0.0129	0.0040	-3.2250	0.0011 **
E	Intercept	-3.4361	0.1239	-27.7328	<.0001
	Income Difference	-0.0001	0.0043	-0.0233	0.9872
F	Intercept	-3.2982	0.0961	-34.3205	<.0001
	Income Difference	-0.0094	0.0028	-3.3571	0.0007 ***
G	Intercept	-4.7002	0.0319	-147.3417	<.0001
	Income Difference	-0.0025	0.0008	-3.1250	0.0031 **
H	Intercept	-5.9751	0.0356	-167.8399	<.0001
	Income Difference	-0.0027	0.0010	-2.7000	0.0049 **
I	Intercept	-3.5149	0.1081	-32.5153	<.0001
	Income Difference	-0.0007	0.0028	-0.2500	0.8094
J	Intercept	-2.5714	0.0865	-29.7272	<.0001
	Income Difference	-0.0042	0.0019	-2.2105	0.0258 *
K	Intercept	-3.2361	0.1810	-17.8790	<.0001
	Income Difference	-0.0026	0.0073	-0.3562	0.7202
L	Intercept	-1.4470	0.4052	-3.5711	0.0004
	Income Difference	-0.0063	0.0203	-0.3103	0.7551
M	Intercept	-2.4160	0.1450	-16.6621	<.0001
	Income Difference	-0.0020	0.0043	-0.4651	0.6442
N	Intercept	-1.9440	0.1309	-14.8510	<.0001
	Income Difference	-0.0044	0.0021	-2.0952	0.0406 *
O	Intercept	-2.8939	0.2360	-12.2623	<.0001
	Income Difference	-0.0086	0.0060	-1.4333	0.1494
P	Intercept	-3.0978	0.2604	-11.8963	<.0001
	Income Difference	-0.0117	0.0074	-1.5811	0.1118

Note: The parameter estimates for income difference is almost always negative (though not always statistically significant), which indicates a consistent effect of increased income difference reducing the probability of a tie existing.
Evidence of homophily (one tail test) at: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$

Multiple Contexts and Adolescent Body Mass Index:

Schools, Neighborhoods, and Social Networks

Clare Rosenfeld Evans[1]

Jukka-Pekka Onnela[2]

David R. Williams[1]

SV Subramanian[1]

[1] Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA

[2] Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

Acknowledgements and Credits:

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

Funding: SVS is partially supported through the Robert Wood Johnson Investigator Award in Health Policy Research.

Keywords: Body mass index, adolescents, context, schools, neighborhoods, social networks

ABSTRACT

Adolescent health and behaviors are influenced by multiple contexts, including schools, neighborhoods, and social networks, yet these contexts are rarely considered simultaneously. In this study we combine social network community detection analysis and cross-classified multilevel modeling in order to compare the contributions of each of these three contexts to the total variation in adolescent body mass index (BMI). Wave 1 of the National Longitudinal Study of Adolescent to Adult Health is used, and for robustness we conduct the analysis in both the full sample available (122 schools; *N*=14,144) and a sub-set of the sample (16 schools; *N*=3,335), known as the saturated sample due to its completeness of neighborhood data. After adjusting for relevant covariates, we find that the school-level and neighborhood-level contributions to the variance are modest compared with the network community-level (σ^2_{school}=0.069, $\sigma^2_{neighborhood}$=0.144, $\sigma^2_{network}$=0.463). These results are robust to two alternative algorithms for specifying network communities, and to analysis in the saturated sample. While this study does not determine whether network effects are attributable to social influence or selection, it does highlight the salience of adolescent social networks and indicates that they may be a promising context to address in the design of health promotion programs.

CHAPTER 3: MULTIPLE CONTEXTS AND ADOLESCENT BODY MASS INDEX: SCHOOLS, NEIGHBORHOODS, AND SOCIAL NETWORKS

INTRODUCTION

Multiple contexts are relevant in shaping individual and population-level health and health behaviors. These include both *physically or spatially defined environments*, such as neighborhoods, schools, and workplaces, and *socially defined environments*, such as the social networks within which individuals are embedded. Historically these contexts have often been studied individually, likely due to the recentness of the availability of methods capable of addressing them simultaneously (Dunn et al. 2015b, Rasbash and Goldstein 1994), such as cross-classified multilevel modeling (CCMM). Since the development of CCMM, researchers have used them most frequently to study the *simultaneous* and *relative* contributions of schools and neighborhoods (Aminzadeh et al. 2013, Dunn et al. 2015a, Dunn et al. 2015b, Oberwittler 2007, Teitler and Weiss 2000, Townsend, Rutter and Foster 2012, Utter et al. 2011, West, Sweeting and Leyland 2004), and workplaces and neighborhoods (Moore et al. 2013, Muntaner et al. 2004, Muntaner et al. 2011, Muntaner et al. 2006, Virtanen et al. 2010) to variation in health behaviors and outcomes. However, studies have rarely bridged the domains of social networks and physical environments, and never within a CCMM framework. This gap in current knowledge is critical to address for two major reasons. First, there is tremendous value in ascertaining the *relative* contributions made by these contexts to the distribution of particular health behaviors and outcomes, as this would enable researchers and policy makers to more effectively target interventions and policies to address heath inequalities. Second, omitting potentially relevant contexts from analyses—particularly those using CCMM—may result in *omitted context bias*, or the attribution of variance associated with the omitted level to the included level or levels (Dunn et al. 2015b, Meyers and Beretvas 2006).

In this study we apply a novel combination of social network community detection analysis and cross-classified multilevel modeling to address this knowledge gap by directly and explicitly comparing the contributions of each of three contexts—schools, neighborhoods, and social networks—to the total variation in adolescent body mass index (BMI). The analysis is conducted using data from wave 1 of the National Longitudinal Study of Adolescent to Adult Health (Add Health). Adolescent body mass index (BMI) is the focus of this study for two main reasons. First, all three contexts have been implicated in prior research as highly relevant to shaping individual-level and population-level distributions of adolescent BMI. Second, the child and adolescent obesity epidemic in the United States represents a major public health challenge due both to its scope (Ogden et al. 2012) and numerous comorbidities (Ferraro and Kelley-Moore 2003, National Institute of Health 1998). Disentangling the contributions of relevant contexts that shape this epidemic will be key to addressing it.

Schools

The clustering of child and adolescent weight status by school-level has been found in a variety of data sets and populations (Procter et al. 2008, Richmond and Subramanian 2008, Richmond et al. 2015, Townsend, Rutter and Foster 2012, Utter et al. 2011). In particular, school-level factors that have been linked to student BMI, physical activity levels, and healthiness of diets, include: socioeconomic status (Miyazaki and Stack 2015, Richmond et al. 2006, Richmond and Subramanian 2008), the prevalence of school food practices (*e.g.*, using food as rewards and incentives) (Kubik, Lytle and Story 2005), aspects of the school built environment such as rural locality, school size and setting, and playground area (Gomes et al. 2014, Miyazaki and Stack 2015), and aspects of the school curriculum, such as frequency and duration of physical education classes, the qualification of physical education teachers, and the presence of school-based nutrition programs (Gomes et al. 2014, Veugelers and Fitzgerald 2005). These findings have situated schools in the policy limelight as both potential shapers of child and adolescent diet and physical activity, and as potential locales for the implementation of health promotion programs.

Neighborhoods

Neighborhoods have similarly been identified as salient to the clustering of child and adolescent BMI (Richmond et al. 2015, Townsend, Rutter and Foster 2012). Aspects of neighborhood *built environments*, such as proximity and access to parks, physical activity establishments, grocery stores, and fast food providers (Carroll-Scott et al. 2013, Schwartz et al. 2011), aspects of neighborhood *socioeconomic environments*, particularly area deprivation (Carroll-Scott et al. 2013, Grow et al. 2010, Rossen 2014, Schwartz et al. 2011, Townsend, Rutter and Foster 2012, Voorhees et al. 2009), and aspects of neighborhood *social environments*, including neighborhood crime, safety, and social connectivity (Carroll-Scott et al. 2013, Molnar et al. 2004, Utter et al. 2011), have been linked to child and adolescent BMI, healthy and unhealthy eating behaviors, physical activity levels, and hours of sedentary screen time.

Social Networks

The structuring of social networks by health status has become an intriguing new area of research. Among both adolescents (Trogdon, Nonnemaker and Pais 2008, Valente et al. 2009) and adults (Christakis and Fowler 2007), a tendency for overweight and obese individuals to cluster, or in other words, for friends to be similar to each other in terms of weight status, has been found. A recent review (Fletcher, Bonell and Sorhaindo 2011) of social network analyses evaluating the eating behaviors and bodyweight of young people found consistent evidence that school friends are clustered according to BMI, and that the frequency of fast food consumption clusters within groups of boys, whereas body image concerns, dieting, and eating disorders cluster among girls. Additionally, overweight and obese youth are less likely to be popular and more likely to be socially isolated.

It is not the purpose of this study to disentangle the roles of *selection* (the tendency for individuals to preferentially select friends who are similar to them in weight status, or other characteristics that are correlated with weight status) and *social influence* (the social contagion of behaviors with relevance to weight status, such as diet and exercise) in generating clustering of weight status in social networks. Instead we address another primary concern (Cohen-Cole and Fletcher 2008b, Fowler and Christakis 2008)—the disentangling of the roles of shared environments such as schools and neighborhoods from network effects.

Simultaneous Contexts

The substantive goal of this study is to determine the *relative* contributions of schools, neighborhoods of residence, and adolescent school-based peer networks to the variance of BMI observed. Studies addressing the simultaneous roles of schools and neighborhoods have consistently determined that both contexts contribute significantly to the variance in adolescent BMI and physical activity (Richmond et al. 2015, Townsend, Rutter and Foster 2012, Utter et al. 2011), yet such studies are still rare, and none that we are aware of have included adolescent peer networks as well.

Studies that have addressed the roles of both social networks (broadly defined) and environments to health outcomes of any kind are uncommon. In a recent review we conducted, these studies fell into three categories. In Category 1, network analyses involved the use of friend (or "alter") attributes to predict attributes of individuals (or "egos") of interest, while school environments were controlled for as fixed effects (Ali, Amialchuk and Dwyer 2011, Ali, Amialchuk and Renna 2011, Ali and Dwyer 2011, Ali, Dwyer and Rizzo 2011, Ali and Dwyer 2010, Cohen-Cole and Fletcher 2008a, Cohen-Cole and Fletcher 2009, Trogdon, Nonnemaker and Pais 2008). Variants on this theme include studies where the effect of alters on egos was evaluated based on geographic distance to determine whether the effect degraded as distance increased (Christakis and Fowler 2007, Christakis and Fowler 2008). The hallmark of studies belonging to this category is that environment is treated as a confounder to be adjusted for, rather than as a separate contributor to the variance that is of substantive interest.

Studies in Category 2 included both network covariates (such as rate of cholera in a social community) and environment covariates (such as rate of cholera in a spatial community) as fixed effect predictors in regression models (Emch et al. 2012, Giebultowicz et al. 2011a, Giebultowicz et al. 2011b). Variants of studies in this category would include covariates for constructs related to social networks, such as social capital (Richmond et al. 2014), though we did not specifically review that literature. While studies such as these enable comparisons of particular aspects of networks or environments that may be of interest, this approach does not enable an evaluation of the holistic contributions made by networks and environments.

Category 3 included only one study, which was recently published by Perez-Heydrich et al. (2013). In this study, networks were represented using fixed effects and neighborhood elements were included as spatial autoregression coefficients in order to correct for spatial dependence. This innovative approach to understanding both social and spatial processes is worthy of further exploration. However, for our current purposes this approach does not enable a direct comparison of the relative influence of networks and environment contexts.

This review highlights two points. First, BMI and obesity were addressed in a networks context but only in Category 1, where environment was not usually of substantive interest.

Second, an innovative approach is required in order to directly compare and better understand the simultaneity of multiple contexts. In this study we present a novel analytic approach, combining social network analysis and multilevel modeling, to disentangle and compare school, neighborhood, and social network contexts.

METHODS

Data

The National Longitudinal Study of Adolescent to Adult Health (Add Health) is a longitudinal study of a nationally representative sample of US adolescents who were in grades 7-12 in the first wave of interviews (1994-1995) (Harris et al. 2009). The primary sampling frame was derived from the Quality Education Database (QED) and was used to select a stratified sample of 80 high schools with probability proportional to size, as well as 52 middle schools that were paired to the high schools as feeders. Schools were stratified based on region, urbanicity, school type (public, private, parochial), ethnic mix and size. A unique aspect of the Add Health data is that students were asked to nominate up to 10 of their closest friends (5 male and 5 female), and therefore it is possible to construct sociocentric social network for each school. In wave 1, social network questions were administered in both the in-school questionnaire, which surveyed the full sample of participating students, and in an in-home questionnaire, which was administered to a sub-sample of students from each school (referred to as the "core sample"). The core sample was selected through a combination of stratified random sampling (to ensure a mix of students across grades and ages) and oversampling of racial and ethnic minorities. The in-home questionnaire also captured greater detail of student health and identifiers of neighborhood of residence.

In two large high schools and fourteen smaller middle schools and high schools, in-home interviews were attempted with *all* students in wave 1. These 16 schools represent a "saturated sample" for which neighborhood identifiers and BMI data are available for most students.

Add Health is an ideal data set in which to study the joint contributions of networks and environment, since few other data sets contain both excellent social network data and environment data.

Sample

In this study, primary analyses are conducted in the core sample (*N*=20,745) among adolescents for whom we have matching in-school questionnaires (referred to subsequently as the "full sample"). Initially the entire in-school sample's social network data (*N*=90,118) was used to construct relatively complete social networks for each school or pair of middle schools and high schools. These networks were used in the network community detection analyses to identify the social network communities, or social groups, to which individual students belonged. The in-school survey was very limited in scope and therefore using the in-school sample for the entire analysis is not possible. However, beginning with the in-school sample enables us to determine network community membership with improved validity, and the core sample that was selected from the in-school sample through (predominantly) random processes provides us with a sufficient representation of the entire school population. Respondents to the in-school questionnaire who completed in-home interviews were then taken as the full sample. Some schools were dropped from the sample due to insufficient sample size after the in-school and in-home matching processes and reductions based on missing the BMI outcome, leaving a sample of 122 schools and *N*=14,144 students (68% of the original core sample).

To test the robustness of our findings, analyses were also conducted in the "saturated sample" of 16 schools, for which we have nearly complete social network, BMI, and neighborhood data (*N*=3,335). The in-home questionnaire network data was used for the saturated sample. While not a random sample, the saturated sample schools are diverse. The two large schools were selected purposely by the Add Health researchers—one because it was a predominantly white rural public high school, and the other because it was racially and ethnically diverse (predominately Hispanic, black and Asian) and located in a major metropolitan area. The other 14 schools represent public (n = 9), private (n = 4), and Catholic (n = 1) schools in the West (n = 3), South (n = 4), Northeast (n = 3), and Midwest (n = 4) regions.

Outcome: Body Mass Index

Body mass index (BMI) was constructed using self-reported height and weight. Goodman, Hinden and Khandelwal (2000) have evaluated the validity of the Add Health self-report of height and weight by comparing the wave 2 self-report with actual measures of height and weight taken at wave 2 (similar measures were not taken in wave 1). They report that in wave 2 the correlation between self-report and measured BMI was very strong (r = 0.92) and self-report correctly classified 96% of adolescents as obese. Girls were no more likely than boys to be misclassified as obese. This is consistent with the larger literature, which generally holds that self-report is a valid measure of height and weight (Spencer et al. 2002, Stewart 1982).

Exposures

Neighborhoods and Schools

Unique IDs for each school and neighborhood of residence (census tract) are available in the Add Health sample, enabling a clear nesting of individual students within the schools they attend and neighborhoods they reside in.

Social Networks

Respondent nominations of friends were used to construct social networks. The nomination items took the following form: *"List your closest [*male/female*] friends. List your best [*male/female*] friend first, then your next best friend, and so on."* Students were provided with a roster of students attending their school and paired sister-school, enabling them to nominate friends using ID codes specific to each individual. All friend nominations are assumed to be reciprocated relationships (i.e., the networks are undirected) because this expands the variation in the numbers of connections observed (i.e., some individuals may have more than 5 male or 5 female friends but were unable to nominate them all due to the limitations imposed by the survey). Friends who were nominated but not surveyed, due to refusal or absence from school when the survey was administered, are included in network structures for network community detection purposes but are excluded from the full sample for missing BMI.

Covariates

Standard demographic covariates in BMI models, including sex, race/ethnicity, parent education, and age, are adjusted for in order to distinguish between composition and context-level variation. Models in the full sample also adjusted for US region (West, Midwest, South, and Northeast).

Sex and race/ethnicity were self-reported or self-confirmed by respondents. Sex was available only as a dichotomous variable. Race/ethnicity categories include: (1) Hispanic or Latino all races, (2) black or African American, (3) Asian, (4) Native American, (5) white, and (6) other. Age was calculated based on the interview date and the self-reported birthday of the respondent.

Parent education was defined as the highest completed level of education attained by either parent or parent-figure in the respondent's household. Since parent education information was requested multiple times in the wave 1 interviews, the order of preference for determination of education levels was: (1) provided by the parent or parent's spouse/partner in the parent in-home questionnaire, (2) provided by the student in the in-home questionnaire, (3) provided by the student in the in-school questionnaire.

ANALYSIS

Network Community Detection

The goal of network community detection is to identify clusters of individuals in networks that are relatively densely connected to each other and sparsely connected to others outside their group. Ideally, network community detection in *social* networks identifies groups that socialize regularly and potentially have their own social norms. These groups may bring together individuals who, while not necessarily nominating each other as friends, are at least closely woven together in their surrounding social fabric. Individuals within the same network community therefore may exert influence on each other through social processes and generally contribute to the shared social environment. This treatment of the networks expands our conceptualization of the relevant social context beyond the immediate social connections of each individual to consider the larger social environment, which includes relatively short indirect connections.

While network communities are conceptually fairly intuitive, identifying communities in practice requires application of strict definitions for what counts as a sufficiently dense cluster to warrant labeling a set of individuals as members of the same network community.

Modularity maximization network community detection algorithms have become widely used because they partition a variety of network graphs into apparently meaningful network communities (Porter, Onnela and Mucha 2009). The modularity maximization algorithm used in this analysis was developed by Blondel et al (2008). There are many different ways a given network can be partitioned, and modularity refers to the quality of a particular set of network partitions. Quality of a set of network partitions in this case is evaluated based on the number of ties that run between nodes (e.g., adolescents) in the same network communities relative to the number of such ties we would expect if ties were created between nodes at random (while holding constant the number of ties of each node) (Newman 2006a, Newman 2006b). Modularity maximization algorithms use a variety of heuristics to optimize the modularity score and partition networks into network communities.

Since different detection algorithms may yield different network community membership lists and different total numbers of network communities for the same network, we evaluated the robustness of our findings to network community specification by applying a second algorithm—*k-clique percolation*—which is becoming particularly popular in analyses of *social* networks (Fortunato 2010). K-clique percolation is a deterministic algorithm that begins with specifying the minimum size of a clique (k), or group where all members are friends with every other member (Palla et al. 2005). For instance, a clique of size $k = 3$ is a group of three adolescents, all of whom are friends with each other. Previous research using k-clique percolation in the Add Health data set found that cliques of size $k = 3$ were optimal (González et al. 2007). In k-clique percolation, all cliques within a network are identified, and then any cliques that share at least $k-1$ members will be defined as members of the same network community.

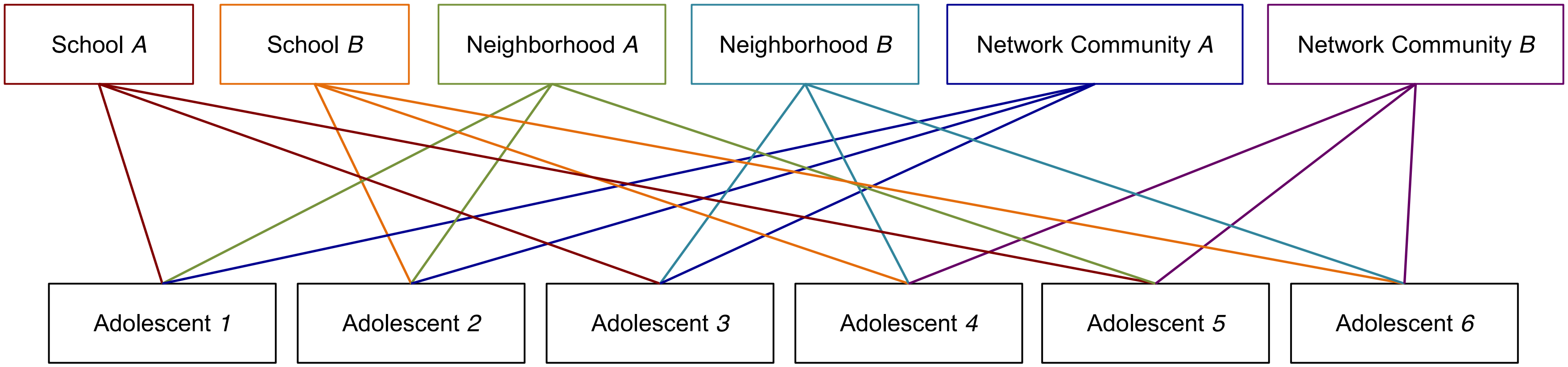
Within this definition, individuals are allowed to simultaneously belong to two or more network communities, whereas in modularity maximization individuals are nested within only one network community. In this analysis, in order to nest individuals within a single network community, if an adolescent belongs to multiple network communities he or she will be included as a member of the community to which they have the most friendship links. In the event of a tie, they are randomly assigned to one of the network communities to which they have the most friendship links. Additionally, some socially marginalized individuals who are not members of cliques are unable to be associated with particular network communities, and therefore are excluded in this analysis from models addressing networks.

Cross-Classified Data

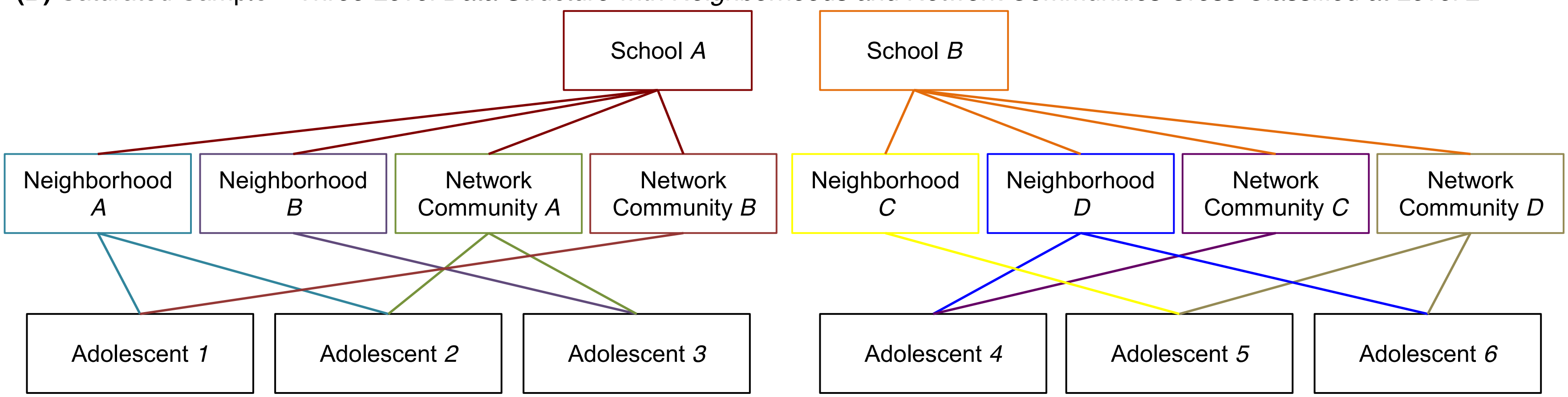
In the full sample, students from the same neighborhood may attend different schools, and students in the same school reside in different neighborhoods. Furthermore, because friend nominations can link individuals across paired schools, there is no clear hierarchical nesting of social network communities within schools. Students within a single network community can also reside in multiple neighborhoods, and students from the same neighborhood can participate in different network communities. In the full sample, therefore, the three contexts—schools, neighborhoods, and social network communities—are cross-classified, with no clear hierarchy (Figure 3.1A).

Figure 3.1. Schematics Illustrating Data Structures in Full and Saturated Samples

(A) Full Sample – *Cross-Classified, Non-Hierarchical Data Structure*



(B) Saturated Sample – *Three Level Data Structure with Neighborhoods and Network Communities Cross-Classified at Level 2*



Note:

Panel A: Because many schools in the full sample come in pairs and are located in the same geographic communities, network communities can bridge schools. There is no clear hierarchical nesting between schools, neighborhoods and network communities, and therefore the data structure is cross-classified.

Panel B: In the saturated sample, all 16 schools are situated in different geographic communities, and therefore neighborhoods and network communities are hierarchically nested within schools but cross-classified with each other.

In the saturated sample, none of the schools are paired and therefore there is a clear hierarchy—network communities are nested within schools and neighborhoods are nested within schools. However, because students from the same neighborhood can still participate in different network communities, and network communities are composed of students from multiple neighborhoods, the network community and neighborhood levels are still cross-classified (Figure 3.1B).

Models

A series of eight models are fit in both the full sample and saturated sample in order to iterate through all combinations of contexts. In the full sample, Model 1 is a single-level linear model of BMI where adolescents are not nested within any context. Models 2, 3 and 4 are two-level hierarchical models—Model 2 nests adolescents in neighborhoods, Model 3 nests adolescents in schools, and Model 4 nests adolescents in their social network communities. Models 5, 6, and 7 iterate through paired combinations of contexts using cross-classified multilevel models (CCMM). Model 5 nests adolescents (level 1) simultaneously in both network communities (level 2) and neighborhoods (level 2), Model 6 nests adolescents (level 1) in neighborhoods (level 2) and schools (level 2), and Model 7 nests adolescents (level 1) in network communities (level 2) and schools (level 2). The final model, Model 8, is a CCMM that nests adolescents (level 1) in schools (level 2), neighborhoods (level 2), and network communities (level 2). Because of the partially hierarchical structure in the saturated sample, a combination of CCMM and three-level hierarchical models are used as appropriate.

In the full sample, each model was fit four times: as a null model, a model adjusted for US region, a model adjusted for demographic covariates, and as a model adjusted for both demographic covariates and US region. In the saturated sample, school dummy variables were included when appropriate (in models not also treating school as a level).

RESULTS

Network community detection was performed in Python 2.7 (Anaconda by Continuum Analytics 2015). A variety of Python libraries were utilized, including Python Pandas (McKinney 2011), NumPy (NumPy Developers 2005) , NetworkX (Hagberg, Schult and Swart 2008), Python-IGraph (Csardi and Nepusz 2006), and matplotlib (Hunter 2007). The k-clique percolation algorithm *k_clique_communities* from the NetworkX library and modularity maximization algorithm *community_multilevel* from the IGraph library were used in network community detection. All multilevel analyses were conducted in MLwiN version 2.32 (Rasbash et al. 2015) using Bayesian Markov Chain Monte Carlo (MCMC) estimation procedures (Browne 2009). The regression models were first fit using Iterative Generalized Least Squares (IGLS) estimation to provide the Bayesian MCMC procedure with initialization values; Non-informative priors and burn-in of 500 iterations were used in all analyses. MCMC estimation was run in all models for a minimum of 150,000 iterations, though most models achieved convergence significantly before that point.

Descriptive statistics for the sample are provided in Table 3.1. The full sample (N=14,144) was predominantly white, black and Hispanic, with a mean age of 15.6 and mean BMI of 22.5 kg/m^2. Further information about the multilevel structure of the data is provided in Table 3.2. In the full sample, adolescents are distributed across 1931 neighborhoods, 122 schools, 930 modularity maximization network communities, and 2733 k-clique network communities (where clique size was k = 3). More detailed descriptions of the 16 school saturated sample are available in the online Supplemental Table 3.1.

Table 3.1. Sample Demographics

	Full Sample		Saturated Sample			
	N	**(%)**	***N***	**(%)**		
TOTAL	14144	(100.00 %)	3335	(100.00 %)		
Sex						
Female	7262	(51.34 %)	1630	(48.88%)		
Male	6882	(48.66 %)	1705	(51.12%)		
Race / Ethnicity						
Hispanic	2233	(15.79 %)	668	(20.03 %)		
White	7244	(51.22 %)	1596	(47.86 %)		
Black	3191	(22.56 %)	477	(14.30 %)		
Asian	1071	(7.57 %)	520	(15.59 %)		
Native American	248	(1.75 %)	49	(1.47 %)		
Other	153	(1.08 %)	23	(0.69 %)		
Missing	4	(0.03 %)	2	(0.06 %)		
Parent Education						
Less than High School	1506	(10.65 %)	413	(12.38 %)		
Completed High School	3592	(25.40 %)	883	(26.48 %)		
Some College	4034	(28.52 %)	955	(28.64 %)		
Completed College or More	4924	(34.81 %)	1056	(31.66 %)		
Missing	88	(0.62 %)	28	(0.84 %)		
	N	***N* missing**	**Mean**	**(SD)**	**Min**	**Max**
Full Sample						
Age in Wave 1 (years)	14141	3	15.61	(1.70)	11	21
BMI in Wave 1 (kg/m^2)	14144	0	22.54	(4.43)	11.22	63.56
Saturated Sample						
Age in Wave 1 (years)	3335	0	16.09	(1.58)	12	20
BMI in Wave 1 (kg/m^2)	3335	0	22.93	(4.51)	11.76	46.24

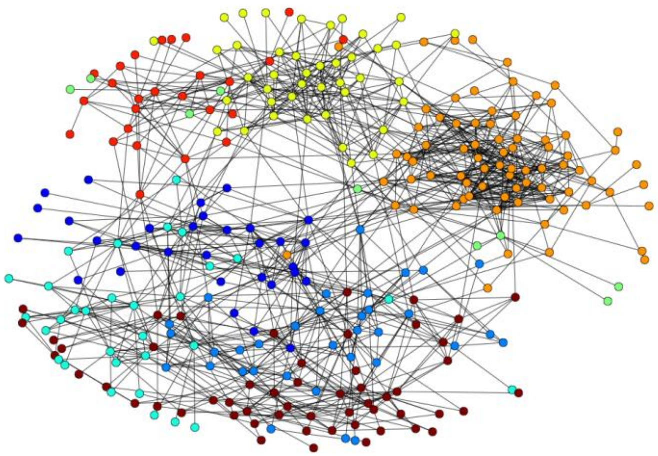
Table 3.2. Multilevel Data Structure of Full and Saturated Samples

	Full Sample		Saturated Sample	
	N	Mean Number of Students	*N*	Mean Number of Students
Students	14144	—	3335	—
Neighborhoods	1931	7.32	335	9.96
Schools	122	115.93	16	208.44
Network Communities				
Modularity Maximization	930	15.21	194	17.19
K-Clique Percolation	2733	5.18	394	8.46

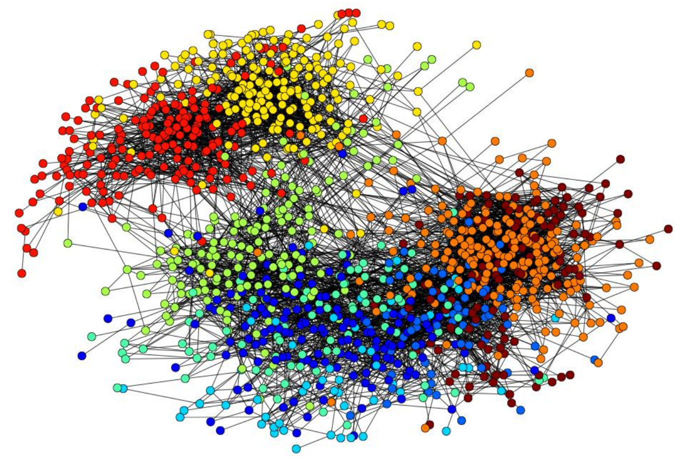
Figure 3.2. Visualizations of detected communities in social networks using two algorithms – Modularity Maximization and K-clique Percolation

(A) Modularity Maximization

Network 1

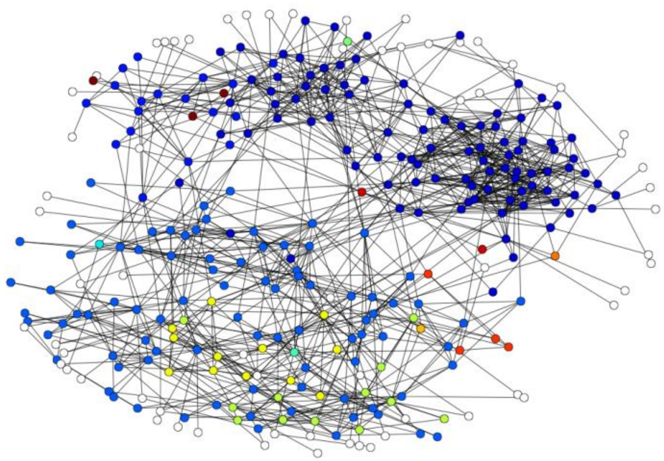


Network 2

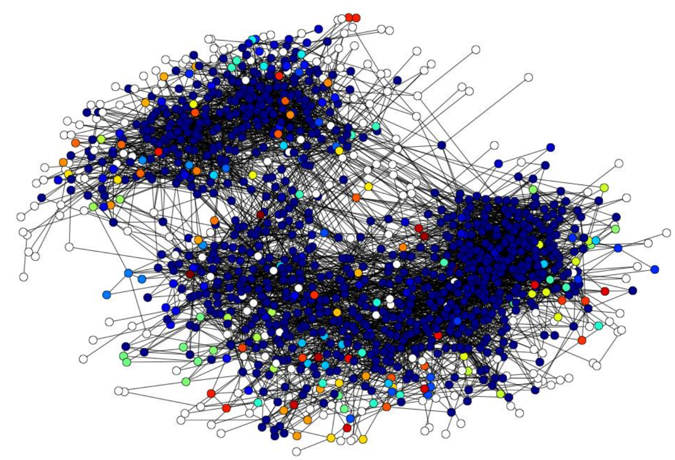


(B) K-Clique Percolation

Network 1



Network 2



Note: Nodes of the same color within a single visualization of a network indicates membership in the same network community. In panel B, the k-clique percolation visuals, white nodes are not members of any network community (i.e., they are marginalized).

Table 3.3. Full Sample – Random Effects Results from Single-Level and Two-Level Multilevel Models, Modularity Maximization Detection

	Null Model	**Adjusted Model – Region ***	**Adjusted Model – Demographics §**	**Adjusted Model – Demographics & Region *§**
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 1: *Single Level Model*				
Individual	19.655	19.616	18.625	18.623
	[19.203 , 20.117]	[19.165 , 20.078]	[18.197 , 19.066]	[18.192 , 19.063]
DIC	**82263.75**	**82235.32**	**80954.95**	**80952.77**
Model 2: *Two Level Model – Students nested in Neighborhoods*				
Neighborhood	0.603 **(3.06%)**	0.586 **(2.98%)**	0.277 **(1.49%)**	0.284 **(1.52%)**
	[0.426 , 0.806]	[0.414 , 0.787]	[0.154 , 0.421]	[0.162 , 0.430]
Individual	19.076	19.056	18.354	18.343
	[18.624 , 19.538]	[18.605 , 19.518]	[17.919 , 18.799]	[17.908 , 18.786]
DIC	**81238.32**	**81223.67**	**80155.93**	**80148.77**
Model 3: *Two Level Model – Students nested in Schools*				
School	0.860 **(4.35%)**	0.817 **(4.14%)**	0.205 **(1.10%)**	0.204 **(1.09%)**
	[0.616 , 1.177]	[0.578 , 1.127]	[0.110 , 0.327]	[0.111 , 0.326]
Individual	18.915	18.915	18.451	18.448
	[18.478 , 19.361]	[18.476 , 19.363]	[18.022 , 18.890]	[18.021 , 18.884]
DIC	**81720.14**	**81720.21**	**80822.88**	**80821.21**
Model 4: *Two Level Model – Students nested in Network Communities*				
Network	1.297 **(6.61%)**	1.255 **(6.41%)**	0.553 **(2.98%)**	0.551 **(2.96%)**
	[1.059 , 1.560]	[1.021 , 1.515]	[0.408 , 0.719]	[0.405 , 0.718]
Individual	18.313	18.318	18.035	18.035
	[17.879 , 18.755]	[17.885 , 18.761]	[17.609 , 18.471]	[17.608 , 18.470]
DIC	**81264.07**	**81266.72**	**80501.38**	**80502.60**

Notes: All models are fit using Markov Chain Monte Carlo (MCMC) estimation procedures. Models including school dummy variables are not fit when school is a level.

DIC = Deviance Information Criterion. ICC = Intraclass Correlation Coefficient.

* Demographic covariates: sex, race/ethnicity, parent education, and age.

§ U.S. Regions: West, Midwest, South, and Northeast.

Table 3.4. Full Sample – Random Effects Results from Three-Level and Cross-Classified Multilevel Models, Modularity Maximization Detection

	Null Model	Adjusted Model – Region *	Adjusted Model – Demographics §	Adjusted Model – Demographics & Region *§
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 5: *Cross-Classified Model – Students nested in Network Communities and Neighborhoods*				
Neighborhood	0.214 **(1.09%)**	0.219 **(1.12%)**	0.169 **(0.91%)**	0.186 **(1.00%)**
	[0.079 , 0.373]	[0.081 , 0.380]	[0.057 , 0.303]	[0.068 , 0.323]
Network	1.186 **(6.06%)**	1.145 **(5.86%)**	0.499 **(2.69%)**	0.490 **(2.64%)**
	[0.953 , 1.449]	[0.918 , 1.402]	[0.359 , 0.660]	[0.351 , 0.652]
Individual	18.175	18.174	17.908	17.901
	[17.737 , 18.625]	[17.734 , 18.621]	[17.475 , 18.348]	[17.469 , 18.341]
DIC	**80559.36**	**80558.44**	**79813.35**	**79808.04**
Model 6: *Cross-Classified Model – Students nested in Neighborhoods and Schools*				
School	0.833 **(4.21%)**	0.784 **(3.97%)**	0.136 **(0.73%)**	0.133 **(0.71%)**
	[0.585 , 1.150]	[0.545 , 1.095]	[0.047 , 0.258]	[0.043 , 0.251]
Neighborhood	0.143 **(0.72%)**	0.154 **(0.78%)**	0.146 **(0.78%)**	0.156 **(0.84%)**
	[0.016 , 0.300]	[0.033 , 0.312]	[0.011 , 0.290]	[0.025 , 0.309]
Individual	18.809	18.800	18.352	18.342
	[18.364 , 19.268]	[18.354 , 19.256]	[17.915 , 18.800]	[17.906 , 18.787]
DIC	**81039.86**	**81034.70**	**80155.38**	**80148.41**
Model 7: *Cross-Classified Model – Students nested in Network Communities and Schools*				
School	0.701 **(3.56%)**	0.677 **(3.44%)**	0.125 **(0.67%)**	0.136 **(0.73%)**
	[0.457 , 1.015]	[0.437 , 0.988]	[0.024 , 0.251]	[0.037 , 0.262]
Network	0.711 **(3.61%)**	0.706 **(3.59%)**	0.470 **(2.53%)**	0.461 **(2.48%)**
	[0.529 , 0.920]	[0.526 , 0.910]	[0.328 , 0.633]	[0.322 , 0.621]
Individual	18.283	18.286	18.013	18.014
	[17.852 , 18.726]	[17.855 , 18.727]	[17.586 , 18.450]	[17.585 , 18.448]
DIC	**81240.50**	**81242.31**	**80485.00**	**80485.62**

Continued

Table 3.4 Continued

	Null Model	**Adjusted Model – Region ***	**Adjusted Model – Demographics §**	**Adjusted Model – Demographics & Region *§**
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 8: *Cross-Classified Model – Students nested in Network Communities and Neighborhoods and Schools*				
School	0.672 **(3.41%)**	0.643 **(3.27%)**	0.061 **(0.33%)**	0.069 **(0.37%)**
	[0.431 , 0.985]	[0.405 , 0.952]	[0.001 , 0.180]	[0.002 , 0.191]
Neighborhood	0.101 **(0.51%)**	0.104 **(0.53%)**	0.135 **(0.73%)**	0.144 **(0.77%)**
	[0.007 , 0.243]	[0.001 , 0.249]	[0.006 , 0.274]	[0.018 , 0.297]
Network	0.717 **(3.64%)**	0.709 **(3.61%)**	0.472 **(2.54%)**	0.463 **(2.49%)**
	[0.534 , 0.929]	[0.527 , 0.920]	[0.332 , 0.635]	[0.322 , 0.623]
Individual	18.203	18.204	17.919	17.915
	[17.763 , 18.653]	[17.765 , 18.656]	[17.486 , 18.360]	[17.484 , 18.357]
DIC	**80580.92**	**80582.23**	**79821.49**	**79818.38**

Notes: All models are fit using Markov Chain Monte Carlo (MCMC) estimation procedures.
DIC = Deviance Information Criterion. ICC = Intraclass Correlation Coefficient.
* Demographic covariates: sex, race/ethnicity, parent education, and age.
§ U.S. Regions: West, Midwest, South, and Northeast.

Figure 3.2 illustrates an example from the data of how one social network is partitioned subtly differently depending on the algorithm used. In general, the Blondel et al. (2008) modularity maximization algorithm partitioned the network graphs into network communities of more equal size, while k-clique percolation tended to split networks into a small number of large network communities and a large number of small communities.

Tables 3.3 and 3.4 provide results for all models fit in the full sample using modularity maximization network community detection. Results for models fit using k-clique detection are available in Supplemental Tables 3.2 and 3.3, while results for models fit in the saturated sample (using modularity maximization) are available in Supplemental Tables 3.4 and 3.5.

In the full sample and using modularity maximization, results from the two-level models (Models 2-4) find that all three contexts, when considered individually, contribute substantially to the total variance in BMI (Intraclass Correlation Coefficients (ICC) in fully adjusted models: neighborhoods = 1.52%, schools = 1.09%, network communities = 2.96%). In Models 6, which cross-classifies neighborhoods and schools, both environmental contexts are attenuated slightly (ICCs: neighborhoods = 0.84%, schools = 0.71%), a result which is consistent with the literature and supports the claim that variance associated with omitted contexts will be attributed to contexts that are considered.

In the presence of cross-classification by network communities, both neighborhoods (Model 5: ICC = 1%) and schools (Model 7: ICC = 0.73%) are similarly attenuated. Surprisingly, the large network effect detected in the two-level model largely remains after cross-classification by the other contexts (ICC for network communities: Model 5 = 2.64%, Model 7 = 2.48%). These results indicate both that network communities contribute substantially to the variance—more so than either neighborhoods or schools—and that their inclusion in models attenuates the variance attributable to neighborhoods and schools.

Surprisingly, in the final model (Model 8) where adolescents are nested in all three contexts simultaneously, the school-level contribution to the variance is significantly reduced (ICC = 0.37%). The neighborhood-level continues to make a larger, though still modest contribution to the variance (ICC = 0.77%). The network community-level contributes to the variance (ICC = 2.49%) more than twice what neighborhoods and schools contribute combined. According to the Deviance Information Criterion (DIC), smaller values of which indicate better model fit, Model 8 (DIC = 79818.38) is modestly outperformed by Model 5 (DIC = 79808.04), indicating that the school-level does not substantially improve the model fit above and beyond what is achieved by cross-classifying neighborhoods and network communities.

These results are robust both to the specification of network communities using an alternative algorithm—k-clique percolation—and when the models are fit in the saturated sample. Interestingly, the magnitude of the network community-level contribution to the variance does depend on the detection method employed, however it is found to be substantial in both cases. In the fully adjusted Model 8, using k-clique network communities, the school-level and neighborhood-level contributions to the variance remain comparable to each other and modest relative to the network-level (ICCs: schools=0.58%, neighborhoods=0.53%, network=11.19%). In the saturated sample, the model is optimized when all three contexts are accounted for, though network communities clearly emerge as contributing more to the total variance.

DISCUSSION

In this study we present a novel analytic approach for determining the relative contributions to the variance of BMI among adolescents made by three contexts—schools, neighborhoods, and social network communities. Through a combination of network community detection and cross-classified multilevel modeling, we find that the network community-level contributes far more to the total variance than either neighborhoods and schools, and that there is some evidence (using modularity maximization) that the salience of schools recedes in the presence of network communities and neighborhoods. These surprising findings support the claim that omitting potentially salient contexts from analysis may result in the misattribution of variance to the contexts that are considered.

There are several critical points to make regarding how these results should be considered. First and foremost, as stated previously, there is no way to disentangle in these models the roles of selection and influence. While this is true for both the school and neighborhood-levels, this is a particular concern for the social network-level. There is a large literature indicating that individuals select friends who share similarities to them across a range of traits (McPherson, Smith-Lovin and Cook 2001, Shalizi and Thomas 2011), and these might include either weight status or characteristics and behaviors associated with weight status (de la Haye et al. 2011, Fletcher, Bonell and Sorhaindo 2011). There is a very real possibility, therefore, that a significant portion of the clustering effect at the network community-level is the result of selection, not social influence of behaviors that affect weight status.

However, we argue that regardless of the causal pathways that lead to the state of clustering, these findings are significant because they indicate the salience of the social environment. Whether adolescents are aware of their weight status and choose friends accordingly, or whether their friends influence their behaviors, the social environment will tend to present them with others who share their weight status and/or behaviors. This clustering will naturally result in the formation of local social norms, particular to social groups, that normalize or reinforce behaviors and attitudes about healthy diet, exercise, and weight. While this may serve a protective role in some cases, particularly given high levels of stigmatization of overweight individuals, these results also indicate that social networks may be ideal to recruit in health promotion activities. Interventions that frame weight-related behaviors as something that *individuals* can and should choose to address may run the risk of increasing individual-level anxiety and perceived stigma. On the other hand, interventions that approach the issue of health promotion as a group activity—something that can be recast as a mutually supportive and positive social experience—may improve the likelihood that participants will engage. Furthermore, if health promotion programs that recruit social groups succeed in shifting group-level norms then the groups themselves may succeed in perpetuating the new behavioral norms among their members, even after the conclusion of participation in the program.

The second critical point to be made is with respect to what is actually being captured in the "school-level" and "neighborhood-level" of these models. Speaking broadly of neighborhood-level and school-level clustering of body mass index is insufficient to characterize the multiple domains of influence within these environments. Three common domains addressed in the literature (Carroll-Scott et al. 2013, Sampson 2003) are the *built environment* (those physical structures and design elements that characterize the physical space of the neighborhood or school), the *socioeconomic environment* (which encompasses the socioeconomic composition of neighborhoods and schools), and the *social environment* (which is often broadly defined as referring to the social networks, social capital, social support, social norms, and/or social control that may operate within the physical purview of the neighborhood or school environment). Each of these domains may influence health and behaviors of individuals and populations through a range of mechanisms (Berkman and Kawachi 2000, Kawachi and Berkman 2003, Kawachi, Subramanian and Kim 2008, Link and Phelan 1995, Smith and Christakis 2008). It was not the purpose of this paper to address all of these domains, nor to determine the relative importance of each domain within each environment. It is essential, however, to recognize that in this study the social networks domain of the school environment was treated separately from the built and socioeconomic environment of the school. The social networks were, by and large, situated within school environments, and therefore the "social network-level" measured here is actually characterizing the social environment endogenous to the school. The neighborhood-level, on the other hand, may reflect any and all of these domains of influence, as no neighborhood social environment equivalent was included. Neighborhood-level social networks, if included, may well have attenuated the neighborhood-level effect, as neighborhood social environments have been found to influence adolescent BMI (Veitch et al. 2012).

Theoretically, a portion of the school and neighborhood-level socioeconomic environments is adjusted for using covariates, though it is possible that some of that domain remains unaccounted for. Largely, however, we may assume that after adjusting for demographics and region, the residual school-level variance is attributable to the built environment. Interestingly, in the null (unadjusted) version of Model 8, schools and network communities contribute approximately the same to the total variance (ICC: school = 3.41%, network community = 3.64%). Taken together, these findings indicate that the majority of the school-level variance is attributable to elements in school socioeconomic and social domains. In other words, socioeconomic and social domains within schools are the big players in terms of shaping distributions of BMI. To the extent that the school built environment and socioeconomic environment shapes the social networks within school walls—perhaps differentially for members of different social groups—schools may exert influences on BMI that is picked up by the social-level, rather than school-level in these models. In other words, it is possible that schools exert an influence on individual outcomes by operating through social network-level mediating pathways. We therefore strongly caution against interpreting these results as being unsupportive of the importance of schools. These results merely challenge us to consider more deeply which domains of a school environment are particularly salient in shaping adolescent BMI, and perhaps what elements of the school shape school-located social networks.

Finally, as mentioned with respect to the network-level, the lack of a causal relationship between a context and clustering of an outcome does not invalidate the potential salience of that context as a location for effective health promotion activities. Schools, for instance, can provide nutrition programs (Veugelers and Fitzgerald 2005) and physical education opportunities (Gomes et al. 2014) to improve student health, while curtailing programs that reinforce unhealthy behaviors (Kubik, Lytle and Story 2005).

One important limitation of this study is that the wave 1 data from Add Health, collected in the mid-1990s, is now somewhat out of date. Unfortunately, no new data source has arisen to allow for more contemporaneous analyses. Add Health remains one of the few studies to evaluate both the environments and social networks of a large and representative sample of the U.S. adolescent population. The fact remains however that in the 20 years since, the structuring and function of adolescent social networks may have changed. The relative influence of school-based social network communities, schools and neighborhoods may have shifted. An updated version of this data is required in order to evaluate the social contexts shaping the lives of adolescents in the present.

Despite limitations of the data, we have presented a novel approach to evaluating the simultaneous contributions of social and physical contexts to the variation of an outcome. This method holds promise for understanding the roles of multiple contexts in shaping a range of outcomes—including health outcomes, criminal behavior, and academic or work performance. It furthermore highlights the importance of evaluating multiple contexts in order to avoid misunderstanding the salience of certain contexts relative to others.

REFERENCES

Ali, MM, A Amialchuk and DS Dwyer. 2011. "Social Network Effects in Contraceptive Behavior among Adolescents." *Journal of Developmental and Behavioral Pediatrics* 32(8):563-71.

Ali, MM, A Amialchuk and F Renna. 2011. "Social Network and Weight Misperception among Adolescents." *Southern Economic Journal* 77(4):827-42.

Ali, MM and DS Dwyer. 2011. "Estimating Peer Effects in Sexual Behavior among Adolescents." *Journal of Adolescence* 34(1):183-90.

Ali, MM, DS Dwyer and JA Rizzo. 2011. "The Social Contagion Effect of Suicidal Behavior in Adolescents: Does It Really Exist?". *Journal of Mental Health Policy and Economics* 14(1):3-12.

Ali, MM and DS Dwyer. 2010. "Social Network Effects in Alcohol Consumption among Adolescents." *Addictive Behaviors* 35(4):337-42.

Aminzadeh, K, S Denny, J Utter, TL Milfont, S Ameratunga, T Teevale and T Clark. 2013. "Neighbourhood Social Capital and Adolescent Self-Reported Wellbeing in New Zealand: A Multilevel Analysis." *Social Science and Medicine* 84:13-21. doi: 10.1016/j.socscimed.2013.02.012.

Anaconda by Continuum Analytics. 2015. "Python 2.7." http://continuum.io/.

Berkman, LF and I Kawachi, eds. 2000. *Social Epidemiology*. New York, NY: Oxford University Press.

Blondel, VD, JL Guillaume, R Lambiotte and E Lefebvre. 2008. "Fast Unfolding of Communities in Large Networks." *Journal of Statistical Mechanics* P10008.

Browne, WJ. 2009. *Mcmc Estimation in Mlwin V2.1*: Centre for Multilevel Modelling, University of Bristol.

Carroll-Scott, A, K Gilstad-Hayden, L Rosenthal, SM Peters, C McCaslin, R Joyce and JR Ickovics. 2013. "Disentangling Neighborhood Contextual Associations with Child Body Mass Index, Diet, and Physical Activity: The Role of Built, Socioeconomic, and Social Environments." *Social Science and Medicine* 95:106-14.

Christakis, NA and JH Fowler. 2007. "The Spread of Obesity in a Large Social Network over 32 Years." *New England Journal of Medicine* 357(4):370-79.

Christakis, NA and JH Fowler. 2008. "The Collective Dynamics of Smoking in a Large Social Network." *New England Journal of Medicine* 358(21):2249-58.

Cohen-Cole, E and JM Fletcher. 2008a. "Is Obesity Contagious? Social Networks Vs. Environmental Factors in the Obesity Epidemic." *Journal of Health Economics* 27(5):1382-87.

Cohen-Cole, E and JM Fletcher. 2009. "Detecting Implausible Social Network Effects in Acne, Height, and Headaches: Longitudinal Analysis." *BMJ (Online)* 338(7685):28-30.

Cohen-Cole, E and JM Fletcher. 2008b. "Is Obesity Contagious? Social Networks Vs. Environmental Factors in the Obesity Epidemic." *Journal of Health Economics* 27:1382-87.

Csardi, G and T Nepusz. 2006. "The Igraph Software Package for Complex Network Research." InterJournal, Complex Systems 1695, http://igraph.org.

de la Haye, K, G Robins, P Mohr and C Wilson. 2011. "Homophily and Contagion as Explanations for Weight Similarities among Adoelscent Friends." *Journal of Adolescent Health* 49(4):421-27.

Dunn, EC, CE Milliren, CR Evans, SV Subramanian and TK Richmond. 2015a. "Disentangling the Relative Influence of Schools and Neighborhoods on Adolescents' Risk for Depressive Symptoms." *American Journal of Public Health* 105(4):732-40.

Dunn, EC, TK Richmond, CE Milliren and SV Subramanian. 2015b. "Using Cross-Classified Multilevel Models to Disentangle School and Neighborhood Effects: An Example Focusing on Smoking Behaviors among Adolescents in the United States." *Health and Place* 31:224-32.

Emch, M, ED Root, S Giebultowicz, M Ali, C Perez-Heydrich and M Yunus. 2012. "Integration of Spatial and Social Network Analysis in Disease Transmission Studies." *Annals of the Association of American Geographers* 102(5):1004-15.

Ferraro, K and JA Kelley-Moore. 2003. "Cumulative Disadvantage and Health: Long-Term Consequences of Obesity?". *American Sociological Review* 68(5):707-29.

Fletcher, A, C Bonell and A Sorhaindo. 2011. "You Are What Your Friends Eat: Systematic Review of Social Network Analyses of Young People's Eating Behaviors and Bodyweight." *Journal of Epidemiology and Community Health* 2011(65):6.

Fortunato, S. 2010. "Community Detection in Graphs." *Physics Reports* 486:75-174.

Fowler, JH and NA Christakis. 2008. "Estimating Peer Effects on Health in Social Networks: A Response to Cohen-Cole and Fletcher; and Trogdon, Nonnemaker, and Pais." *Journal of Health Economics* 27(5):1400-05.

Giebultowicz, S, M Ali, M Yunus and M Emch. 2011a. "The Simultaneous Effects of Spatial and Social Networks on Cholera Transmission." *Interdisciplinary Perspectives on Infectious Diseases* 2011.

Giebultowicz, S, M Ali, M Yunus and M Emch. 2011b. "A Comparison of Spatial and Social Clustering of Cholera in Matlab, Bangladesh." *Health and Place* 17(2):490-97.

Gomes, TN, FK dos Santos, W Zhu, J Eisenmann and JA Maia. 2014. "Multilevel Analyses of School and Children's Characteristics Associated with Physical Activity." *Journal of School Health* 84(10):668-76.

González, MC, HJ Herrmann, J Kertész and T Vicsek. 2007. "Community Structure and Ethnic Preferences in School Friendship Networks." *Physica A* 379(1):307-16.

Goodman, E, BR Hinden and S Khandelwal. 2000. "Accuracy of Teen and Parental Reports of Obesity and Body Mass Index." *Pediatrics* 106(1):52-58.

Grow, HMG, AJ Cook, DE Arterburn, BE Saelens, A Drewnowski and P Lozano. 2010. "Child Obesity Associated with Social Disadvantage of Children's Neighborhoods." *Social Science and Medicine* 71(3):584-91.

Hagberg, AA, DA Schult and PJ Swart. 2008. "Exploring Network Structure, Dynamics, and Function Using Networkx." Paper presented at the Proceedings of the 7th Python in Science Conference (SciPy2008), Pasadena, CA.

Harris, KM, CT Halpern, E Whitsel, J Hussey, J Tabor, P Entzel and JR Udry. 2009, "The National Longitudinal Study of Adolescent Health: Research Design [WWW Document]". (URL: http://www.cpc.unc.edu/projects/addhealth/design.).

Hunter, JD. 2007. "Matplotlib: A 2d Graphics Environment." *Computing In Science and Engineering* 9(3):90-95.

Kawachi, I and LF Berkman, eds. 2003. *Neighborhoods and Health*. New York, NY: Oxford University Press.

Kawachi, I, SV Subramanian and D Kim, eds. 2008. *Social Capital and Health*. New York, NY: Springer Science.

Kubik, MY, LA Lytle and M Story. 2005. "Schoolwide Food Practices Are Associated with Body Mass Index in Middle School Students." *Archives of Pediatrics and Adolescent Medicine* 159(12):1111-14.

Link, BG and J Phelan. 1995. "Social Conditions as Fundamental Causes of Disease." *Journal of Health and Social Behavior* 35:80-94.

McKinney, W. 2011. "Python Pandas." AQR Capital Management, Lambda Foundry Inc, http://pandas.pydata.org/.

McPherson, M, L Smith-Lovin and J Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27(1):415-44.

Meyers, JL and SN Beretvas. 2006. "The Impact of Inappropriate Modeling of Cross-Classified Data Structures." *Multivariate Behavioral Research* 41(4):473-97.

Miyazaki, Y and M Stack. 2015. "Examining Individual and School Characteristics Associated with Child Obesity Using a Multilevel Growth Model." *Social Science and Medicine* 128:56-66.

Molnar, BE, SL Gortmaker, FC Bull and SL Buka. 2004. "Unsafe to Play? Neighborhood Disorder and Lack of Safety Predict Reduced Physical Activity among Urban Children and Adolescents." *American Journal of Health Promotion* 18(5):378-86.

Moore, K, AV Diez Roux, A Auchincloss, KR Evenson, J Kaufman, M Mujahid and K Williams. 2013. "Home and Work Neighbourhood Environments in Relation to Body Mass Index: The Multi-Ethnic Study of Atherosclerosis (Mesa)." *Journal of Epidemiology and Community Health* 67(10):846-53. doi: 10.1136/jech-2013-202682.

Muntaner, C, Y Li, X Xue, P O'Campo, HJ Chung and WW Eaton. 2004. "Work Organization, Area Labor-Market Characteristics, and Depression among U.S. Nursing Home Workers: A Cross-Classified Multilevel Analysis." *International Journal of Occupational and Environmental Health* 10(4):392-400.

Muntaner, C, Y Li, E Ng, J Benach and H Chung. 2011. "Work or Place? Assessing the Concurrent Effects of Workplace Exploitation and Area-of-Residence Economic Inequality on Individual Health." *International Journal of Health Services* 41(1):27-50. doi: 10.2190/HS.41.1.c.

Muntaner, C, Y Li, X Xue, T Thompson, H Chung and P O'Campo. 2006. "County and Organizational Predictors of Depression Symptoms among Low-Income Nursing Assistants in the USA." *Social Science and Medicine* 63(6):1454-65. doi: 10.1016/j.socscimed.2006.03.042.

National Institute of Health. 1998. *Clinical Guidelines on the Identification, Evaluation, and Treatment of Overweight and Obesity in Adults—the Evidence Report*. Obes Res 6(Suppl 2): 51S-209S.

Newman, MEJ. 2006a. "Finding Community Structure in Networks Using the Eigenvectors of Matrices." *Physical Review E* 74(036104).

Newman, MEJ. 2006b. "Modularity and Community Structure in Networks." *Proceedings of the National Academy of Sciences USA* 103:8577-82.

NumPy Developers. 2005. "Numpy User Guide." http://www.numpy.org/.

Oberwittler, D. 2007. "The Effects of Neighbourhood Poverty on Adolescent Problem Behaviours: A Multi-Level Analysis Differentiated by Gender and Ethnicity." *Housing Studies* 22(5):781-803. doi: 10.1080/02673030701474727.

Ogden, CL, MD Carroll, BK Kit and KM Flegal. 2012. "Prevalence of Obesity in the United States, 2009-2010." *NCHS Data Brief 2012*:1-8.

Palla, G, I Derényi, I Farkas and T Vicsek. 2005. "Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society." *Nature* 435(814-818).

Perez-Heydrich, C, JM Furgurson, S Giebultowicz, JJ Winston, M Yunus, PK Streatfield and M Emch. 2013. "Social and Spatial Processes Associated with Childhood Diarrheal Disease in Matlab, Bangladesh." *Health and Place* 19(1):45-52.

Porter, MA, JP Onnela and PJ Mucha. 2009. "Communities in Networks." *Notices of the American Mathematical Society* 56(9):1082-97.

Procter, KL, MC Rudolf, RG Feltbower, R Levine, A Connor, M Robinson and GP Clarke. 2008. "Measuring the School Impact on Child Obesity." *Social Science and Medicine* 67(2):341-49.

Rasbash, J, WJ Browne, M Healy, B Cameron and C Charlton. 2015. *MLwiN Version 2.32*: Centre for Multilevel Modelling, University of Bristol.

Rasbash, J and H Goldstein. 1994. "Efficient Analysis of Mixed Hierarchical and Cross-Classified Random Structures Using a Multilevel Model." *Journal of Educational and Behavioral Statistics* 19(4):337-50.

Richmond, TK, RA Hayward, S Gahagan, AE Field and M Heisler. 2006. "Can School Income and Racial/Ethnic Composition Explain the Racial/Ethnic Disparity in Adolescent Physical Activity Participation?". *Pediatrics* 117(6):2158-66.

Richmond, TK and SV Subramanian. 2008. "School Level Contextual Factors Are Associated with Weight Status of Adolescent Males and Females." *Obesity (Silver Spring)* 16:1324-30.

Richmond, TK, CE Milliren, CE Walls and I Kawachi. 2014. "School Social Capital and Body Mass Index in the National Longitudinal Study of Adolescent Health." *Journal of School Health* 84(12):759-68.

Richmond, TK, EC Dunn, CE Milliren, CR Evans and SV Subramanian. 2015. "Disentangling Overlapping Influences of Neighborhoods and Schools on Adolescent Body Mass Index." Manuscript submitted for publication.

Rossen, LM. 2014. "Neighbourhood Economic Deprivation Explains Racial/Ethnic Disparities in Overweight and Obesity among Children and Adolescents in the U.S.A.". *Journal of Epidemiology and Community Health* 68(2):123-29.

Sampson, RJ. 2003. "Neighborhood-Level Context and Health: Lessons from Sociology." in *Neighborhoods and Health*, edited by I. Kawachi and L. Berkman. New York, NY: Oxford University Press.

Schwartz, BS, WF Stewart, S Godby, J Pollak, J DeWalle, S Larson, DG Mercer and TA Glass. 2011. "Body Mass Index and the Built and Social Environments in Children and Adolescents Using Electronic Health Records." *American Journal of Preventative Medicine* 41(4):e17-e28.

Shalizi, CR and AC Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods and Research* 40(2):211-39.

Smith, KP and NA Christakis. 2008. "Social Networks and Health." *Annual Review of Sociology* 34:405-18.

Spencer, EA, PN Appleby, GK Davey and TJ Key. 2002. "Validity of Self-Reported Height and Weight in 4808 Epic-Oxford Participants." *Public Health Nutrition* 5(4):561-65.

Stewart, AL. 1982. "The Reliability and Validity of Self-Reported Weight and Height." *Journal of Chronic Diseases* 35(4):295-309.

Teitler, JO and CC Weiss. 2000. "Effects of Neighborhood and School Environments on Transitions to First Sexual Intercourse." *Sociology of Education* 73(2):112-32.

Townsend, N, H Rutter and C Foster. 2012. "Age Differences in the Association of Childhood Obesity with Area-Level and School-Level Deprivation: Cross-Classified Multilevel Analysis of Cross-Sectional Data." *International Journal of Obesity* 36(1):45-52. doi: 10.1038/ijo.2011.191.

Trogdon, JG, J Nonnemaker and J Pais. 2008. "Peer Effects in Adolescent Overweight." *Journal of Health Economics* 27(5):1388-99.

Utter, J, S Denny, E Robinson, S Ameratunga and TL Milfont. 2011. "Social and Physical Contexts of Schools and Neighborhoods: Associations with Physical Activity among Young People in New Zealand." *American Journal of Public Health* 101(9):1690-95. doi: 10.2105/ajph.2011.300171.

Valente, TW, K Fujimoto, CP Chou and D Spruijt-Metz. 2009. "Adolescent Affiliations and Adiposity: A Social Network Analysis of Friendships and Obesity." *Journal of Adolescent Health* 45(2):202-04.

Veitch, J, MM van Stralen, MJM Chinapaw, SJ te Velde, D Crawford, J Salmon and A Timperio. 2012. "The Neighborhood Social Environment and Body Mass Index among Youth: A Mediation Analysis." *International Journal of Behavioral Nutrition and Physical Activity*:31.

Veugelers, PJ and AL Fitzgerald. 2005. "Effectiveness of School Programs in Preventing Childhood Obesity: A Multilevel Comparison." *American Journal of Public Health* 95(3):432-35.

Virtanen, M, M Kivimaki, J Pentti, T Oksanen, K Ahola, A Linna, A Kouvonen, P Salo and J Vahtera. 2010. "School Neighborhood Disadvantage as a Predictor of Long-Term Sick Leave among Teachers: Prospective Cohort Study." *American Journal of Epidemiology* 171(7):785-92. doi: 10.1093/aje/kwp459.

Voorhees, CC, DJ Catellier, JS Ashwood, DA Cohen, A Rung, L Lytle, TL Conway and M Dowda. 2009. "Neighborhood Socioeconomic Status and Non School Physical Activity and Body Mass Index in Adolescent Girls." *Journal of Physical Activity & Health* 6(6):731-40.

West, P, H Sweeting and A Leyland. 2004. "School Effects on Pupils' Health Behaviours: Evidence in Support of the Health Promoting School." *Research Papers in Education* 19(3):261-91. doi: 10.1080/0267152042000247972.

APPENDIX B: CHAPTER 3 SUPPLEMENTAL TABLES

Supplemental Table 3.1. Data Structure of Saturated Sample – Neighborhoods and Modularity Maximization Network Communities by School

School ID	*N* Students	*N* Neighborhoods	*N* Network Communities
TOTAL	3335	335	194
A	22	14	8
B	53	11	5
C	101	17	18
D	148	6	10
E	100	8	9
F	136	60	8
G	812	24	19
H	1521	96	48
I	86	37	13
J	82	8	6
K	71	17	10
L	19	3	4
M	48	15	8
N	49	12	7
O	40	3	10
P	47	4	11

Supplemental Table 3.2. Full Sample – Random Effects Results from Single-Level and Two-Level Multilevel Models, K-Clique Detection

	Null Model	**Adjusted Model – Region ***	**Adjusted Model – Demographics §**	**Adjusted Model – Demographics & Region *§**
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 1: *Single Level Model*				
Individual	19.655	19.616	18.625	18.623
	[19.203 , 20.117]	[19.165 , 20.078]	[18.197 , 19.066]	[18.192 , 19.063]
DIC	**82263.75**	**82235.32**	**80954.95**	**80952.77**
Model 2: *Two Level Model – Students nested in Neighborhoods*				
Neighborhood	0.603 **(3.06%)**	0.586 **(2.98%)**	0.277 **(1.49%)**	0.284 **(1.52%)**
	[0.426 , 0.806]	[0.414 , 0.787]	[0.154 , 0.421]	[0.162 , 0.430]
Individual	19.076	19.056	18.354	18.343
	[18.624 , 19.538]	[18.605 , 19.518]	[17.919 , 18.799]	[17.908 , 18.786]
DIC	**81238.32**	**81223.67**	**80155.93**	**80148.77**
Model 3: *Two Level Model – Students nested in Schools*				
School	0.860 **(4.35%)**	0.817 **(4.14%)**	0.205 **(1.10%)**	0.204 **(1.09%)**
	[0.616 , 1.177]	[0.578 , 1.127]	[0.110 , 0.327]	[0.111 , 0.326]
Individual	18.915	18.915	18.451	18.448
	[18.478 , 19.361]	[18.476 , 19.363]	[18.022 , 18.890]	[18.021 , 18.884]
DIC	**81720.14**	**81720.21**	**80822.88**	**80821.21**
Model 4: *Two Level Model – Students nested in Network Communities*				
Network	3.377 **(16.57%)**	3.307 **(16.28%)**	2.322 **(12.25%)**	2.340 **(12.34%)**
	[2.656 , 4.163]	[2.577 , 4.107]	[1.643 , 3.078]	[1.657 , 3.078]
Individual	17.008	17.012	16.626	16.618
	[16.512 , 17.521]	[16.512 , 17.523]	[16.126 , 17.139]	[16.123 , 17.129]
DIC	**65075.08**	**65076.60**	**64475.20**	**64469.12**

Notes: All models are fit using Markov Chain Monte Carlo (MCMC) estimation procedures. Models including school dummy variables are not fit when school is a level.

DIC = Deviance Information Criterion. ICC = Intraclass Correlation Coefficient.

* Demographic covariates: sex, race/ethnicity, parent education, and age.

§ U.S. Regions: West, Midwest, South, and Northeast.

Supplemental Table 3.3. Full Sample – Random Effects Results from Three-Level and Cross-Classified Multilevel Models, K-Clique Detection

	Null Model	**Adjusted Model – Region ***	**Adjusted Model – Demographics §**	**Adjusted Model – Demographics & Region *§**
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 5: *Cross-Classified Model – Students nested in Network Communities and Neighborhoods*				
Neighborhood	0.256 **(1.26%)**	0.259 **(1.28%)**	0.145 **(0.77%)**	0.154 **(0.81%)**
	[0.109 , 0.447]	[0.101 , 0.447]	[0.006 , 0.317]	[0.016 , 0.326]
Network	3.160 **(15.55%)**	3.087 **(15.24%)**	2.230 **(11.77%)**	2.214 **(11.69%)**
	[2.426 , 3.968]	[2.338 , 3.898]	[1.522 , 2.997]	[1.517 , 2.984]
Individual	16.911	16.914	16.570	16.567
	[16.409 , 17.432]	[16.405 , 17.434]	[16.065 , 17.091]	[16.060 , 17.086]
DIC	**64543.69**	**64545.29**	**63973.06**	**63970.49**
Model 6: *Cross-Classified Model – Students nested in Neighborhoods and Schools*				
School	0.833 **(4.21%)**	0.784 **(3.97%)**	0.136 **(0.73%)**	0.133 **(0.71%)**
	[0.585 , 1.150]	[0.545 , 1.095]	[0.047 , 0.258]	[0.043 , 0.251]
Neighborhood	0.143 **(0.72%)**	0.154 **(0.78%)**	0.146 **(0.78%)**	0.156 **(0.84%)**
	[0.016 , 0.300]	[0.033 , 0.312]	[0.011 , 0.290]	[0.025 , 0.309]
Individual	18.809	18.800	18.352	18.342
	[18.364 , 19.268]	[18.354 , 19.256]	[17.915 , 18.800]	[17.906 , 18.787]
DIC	**81039.86**	**81034.70**	**80155.38**	**80148.41**
Model 7: *Cross-Classified Model – Students nested in Network Communities and Schools*				
School	0.851 **(4.19%)**	0.766 **(3.79%)**	0.106 **(0.56%)**	0.128 **(0.68%)**
	[0.535 , 1.264]	[0.476 , 1.149]	[0.006 , 0.265]	[0.019 , 0.285]
Network	2.512 **(12.37%)**	2.495 **(12.35%)**	2.140 **(11.33%)**	2.123 **(11.24%)**
	[1.829 , 3.264]	[1.816 , 3.252]	[1.410 , 2.924]	[1.424 , 2.888]
Individual	16.945	16.948	16.647	16.639
	[16.447 , 17.459]	[16.450 , 17.461]	[16.143 , 17.168]	[16.139 , 17.154]
DIC	**65031.58**	**65033.43**	**64489.33**	**64484.30**

Continued

	Null Model	**Adjusted Model – Region ***	**Adjusted Model – Demographics §**	**Adjusted Model – Demographics & Region *§**
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 8: *Cross-Classified Model – Students nested in Network Communities and Neighborhoods and Schools*				
School	0.849 **(4.17%)**	0.772 **(3.81%)**	0.090 **(0.48%)**	0.110 **(0.58%)**
	[0.535 , 1.264]	[0.478 , 1.164]	[0.003 , 0.244]	[0.005 , 0.267]
Neighborhood	0.136 **(0.67%)**	0.123 **(0.61%)**	0.123 **(0.65%)**	0.101 **(0.53%)**
	[0.007 , 0.299]	[0.001 , 0.289]	[0.003 , 0.286]	[0.001 , 0.259]
Network	2.485 **(12.21%)**	2.489 **(12.28%)**	2.104 **(11.13%)**	2.118 **(11.19%)**
	[1.795 , 3.248]	[1.791 , 3.250]	[1.418 , 2.876]	[1.428 , 2.880]
Individual	16.879	16.883	16.594	16.593
	[16.370 , 17.399]	[16.373 , 17.407]	[16.085 , 17.120]	[16.086 , 17.116]
DIC	**64521.45**	**64524.89**	**63989.39**	**63988.70**

Notes: All models are fit using Markov Chain Monte Carlo (MCMC) estimation procedures.
DIC = Deviance Information Criterion. ICC = Intraclass Correlation Coefficient.
* Demographic covariates: sex, race/ethnicity, parent education, and age.
§ U.S. Regions: West, Midwest, South, and Northeast.

Supplemental Table 3.4. Saturated Sample – Random Effects Results from Single-Level and Two-Level Multilevel Models, Modularity Maximization Detection

	Null Model	**Adjusted Model – Schools**	**Adjusted Model – Demographics ***	**Adjusted Model – Demographics & Schools ***
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 1: *Single Level Model*				
Individual	20.313	19.675	19.158	19.063
	[19.359 , 21.309]	[18.748 , 20.651]	[18.259 , 20.103]	[18.163 , 20.011]
DIC	**19505.99**	**19399.12**	**19136.89**	**19120.42**
Model 2: *Two Level Model – Students nested in Neighborhoods*				
Neighborhood	0.968 **(4.71%)**	0.161 **(0.82%)**	0.163 **(0.85%)**	0.160 **(0.84%)**
	[0.436 , 1.727]	[0.004 , 0.461]	[0.004 , 0.451]	[0.012 , 0.442]
Individual	19.573	19.516	18.998	18.914
	[18.629 , 20.564]	[18.589 , 20.491]	[18.088 , 19.948]	[18.008 , 19.861]
DIC	**19300.37**	**19291.12**	**19028.08**	**19013.73**
Model 3: *Two Level Model – Students nested in Schools*				
School	1.068 **(5.15%)**	—	0.249 **(1.29%)**	—
	[0.382 , 2.489]		[0.003 , 0.855]	
Individual	19.674	—	19.080	—
	[18.748 , 20.643]		[18.178 , 20.028]	
DIC	**19399.13**		**19123.48**	
Model 4: *Two Level Model – Students nested in Network Communities*				
Network	2.466 **(11.93%)**	1.656 **(8.35%)**	1.494 **(7.76%)**	1.471 **(7.66%)**
	[1.639 , 3.539]	[1.035 , 2.490]	[0.944 , 2.215]	[0.911 , 2.235]
Individual	18.197	18.187	17.749	17.730
	[17.322 , 19.112]	[17.313 , 19.099]	[16.898 , 18.645]	[16.881 , 18.622]
DIC	**19138.96**	**19136.93**	**18884.43**	**18880.91**

Notes: All models are fit using Markov Chain Monte Carlo (MCMC) estimation procedures. Models including school dummy variables are not fit when school is a level.

DIC = Deviance Information Criterion. ICC = Intraclass Correlation Coefficient.

* Demographic covariates: sex, race/ethnicity, parent education, and age.

Supplemental Table 3.5. Saturated Sample – Random Effects Results from Three-Level and Cross-Classified Multilevel Models, Modularity Maximization Detection

	Null Model	Adjusted Model – Schools	Adjusted Model – Demographics *	Adjusted Model – Demographics & Schools *
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 5: *Cross-Classified Model – Students nested in Network Communities and Neighborhoods*				
Neighborhood	0.165 **(0.80%)**	0.119 **(0.60%)**	0.115 **(0.60%)**	0.138 **(0.72%)**
	[0.004 , 0.493]	[0.002 , 0.359]	[0.002 , 0.343]	[0.004 , 0.405]
Network	2.335 **(11.35%)**	1.646 **(8.30%)**	1.477 **(7.68%)**	1.461 **(7.61%)**
	[1.519 , 3.400]	[1.028 , 2.474]	[0.930 , 2.205]	[0.906 , 2.214]
Individual	18.069	18.064	17.639	17.603
	[17.185 , 18.991]	[17.195 , 18.979]	[16.780 , 18.541]	[16.748 , 18.505]
DIC	**19035.20**	**19034.48**	**18784.00**	**18777.20**
Model 6: *Three Level Model – Students nested in Neighborhoods nested in Schools*				
School	1.088 **(5.24%)**	—	0.238 **(1.23%)**	—
	[0.369 , 2.595]		[0.002 , 0.883]	
Neighborhood	0.141 **(0.68%)**	—	0.138 **(0.71%)**	—
	[0.002 , 0.407]		[0.004 , 0.405]	
Individual	19.528	—	18.947	—
	[18.604 , 20.506]		[18.037 , 19.896]	
DIC	**19293.28**	—	**19019.72**	—
Model 7: *Three Level Model – Students nested in Network Communities nested in Schools*				
School	0.916 **(4.41%)**	—	0.108 **(0.56%)**	—
	[0.237 , 2.346]		[0.001 , 0.607]	
Network	1.647 **(7.94%)**	—	1.466 **(7.59%)**	—
	[1.027 , 2.474]		[0.917 , 2.198]	
Individual	18.188	—	17.745	—
	[17.314 , 19.109]		[16.895 , 18.643]	
DIC	**19137.27**	—	**18883.17**	—

Continued

Supplemental Table 3.5 Continued

	Null Model	**Adjusted Model – Schools**	**Adjusted Model – Demographics ***	**Adjusted Model – Demographics & Schools ***
	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]	Estimate **(ICC)** [95% Credible Interval]
Model 8: *Three Level Cross-Classified Model – Students nested in Network Communities and Neighborhoods nested in Schools*				
School	0.954 **(4.59%)**	—	0.127 **(0.66%)**	—
	[0.241 , 2.468]		[0.001 , 0.691]	
Neighborhood	0.095 **(0.46%)**	—	0.116 **(0.60%)**	—
	[0.001 , 0.326]		[0.004 , 0.337]	
Network	1.638 **(7.89%)**	—	1.453 **(7.52%)**	—
	[1.024 , 2.457]		[0.909 , 2.188]	
Individual	18.081	—	17.629	—
	[17.213 , 18.994]		[16.772 , 18.525]	
DIC	**19037.77**	—	**18782.01**	—

Notes: All models are fit using Markov Chain Monte Carlo (MCMC) estimation procedures. Models including school dummy variables are not fit when school is a level.

DIC = Deviance Information Criterion. ICC = Intraclass Correlation Coefficient.

* Demographic covariates: sex, race/ethnicity, parent education, and age.