

HashtagHealth: A Social Media Big Data Resource for Neighborhood Effects Research

The University of Utah

PI: Quynh Nguyen

Grant Number: 1-K01S25433-01

My goal in seeking a Mentored Research Career Development Award is to acquire the necessary training, practical experience, and knowledge to become a leading independent investigator who harnesses biomedical Big Data Science for the investigation of multilevel influences on health. To continue my progress towards this goal, I am proposing to build the infrastructure to establish a neighborhood data repository, HashtagHealth, for public health researchers and policy makers. I am a highly trained researcher in social epidemiology and quantitative analyses, particularly large health surveys. Before coming to Utah, I worked as a full-time statistical programmer/data analyst on a NIH-funded project to evaluate the health effects of a large neighborhood relocation policy experiment on low-income families in five cities. Our study results suggested that moving from high- to lower poverty neighborhood is related to reductions in obesity and diabetes and improved mental health. Other extant research has provided evidence on associations between the neighborhood environment and mortality and morbidity-even after adjusting for individual characteristics. Poor access to healthy food, fast food chains, the lack of recreational facilities, and higher crime rates all correlate with higher obesity rates. Nonetheless, the dearth of neighborhood data, especially measures of neighborhood quality that are consistent across geographic areas, limits neighborhood effects research. Moreover, neighborhoods are not only defined by their resources, but also by the social interactions and activities of people who live there. The widespread usage of the internet and open recording of many transactions has led to the availability of massive amounts of data that permits capture of previously hidden micro-level interactions. We will build the data algorithms and infrastructure to harness relatively untapped, cost efficient, and pervasive social media data to develop neighborhood indicators such as food themes, healthiness of food mentions, frequency of exercise/recreation mentions, metabolic intensity of physical activities, and happiness levels. The creation of HashtagHealth requires the use and refinement of Big Data methods to perform data mining, processing and storing of heterogeneous, unstructured data. We will build a testable version of HashtagHealth for the state of Utah and then apply the data resource to the examination of neighborhood effects on young adult obesity. My rigorous training and previous research experiences in social determinants of health, causal inference, and data analyses uniquely prepare me to make significant contributions to the field of Big Data, particularly at the intersection of public health and social sciences. My Specific Aims are: 1) to develop a neighborhood data resource, HashtagHealth, for public health researchers, 2) to develop Big Data techniques to produce novel neighborhood quality indicators (e.g., healthiness of food mentions, frequency and type of exercise/recreation and happiness levels), and 3) to utilize HashtagHealth and individual-level data from the Utah Population Database to investigate neighborhood influences on obesity among young adults. My mentorship team includes experts in biomedical research (Drs. Ken Smith, Jim VanDerslice), computer science (Dr. Feifei Li), and statistics (Dr. Ming Wen). My team has the breadth of expertise to help me obtain critical multidisciplinary skills and successfully implement my research aims. In addition to my research aims, my Specific Career Development Aims include the following: 1) to develop expertise in data mining and database systems, 2) to acquire training in natural language processing and machine learning, 3) to further gain knowledge of geographic information systems (GIS), 4) to develop expertise in study design and analysis of neighborhood effects, and 5) and to develop grant writing and research management skills to lead future projects. The knowledge and experience gained from this proposal will allow me to successfully compete for R01 funding to create a national neighborhood data repository and to investigate national patterns of neighborhood effects on obesity. This proposal makes significant, relevant contributions to the field because 1) neighborhood environments are increasingly linked to important health outcomes, and 2) this project addresses the limits to research resulting from the lack of neighborhood data by providing new, cost-efficient data resources and methods for characterizing neighborhoods. PUBLIC HEALTH RELEVANCE: The epidemic rise in obesity and related chronic diseases in recent decades signal the importance of structural forces and social processes, but the dearth of data on contextual factors limits the investigation of multilevel effects on health. My overarching career goal is to contribute to the improvement of population health by conducting research in social determinants of health and developing new methodologies to address multidimensional research questions. The development of the Hashtag Health data repository will be a significant benefit to neighborhood effects researchers, harnessing the largely untapped potential of social media data to capture social and cultural processes with potential impact on health.