

Who and Where: A socio-spatial integrated approach for community-based health research

Jeanne-Marie R. Stacciarini¹, Raffaele Vacca², Liang Mao^{3*}

¹University of Florida, College of Nursing

²University of Florida, Department of Sociology and Criminology & Law

³University of Florida, Department of Geography

*Correspondence: liangmao@ufl.edu

Abstract

Social and spatial characteristics of a population often interact to influence health outcomes, suggesting a need to jointly analyze both to offer useful insights in community health. However, researchers have used either social or spatial analyses to examine community-based health issues and inform intervention programs. We propose a combined socio-spatial analytic approach to develop a *social network with spatial weights* and a *spatial statistic with social weights*, and apply them to an ongoing study of mental and physical well-being of rural Latino immigrants in North Florida, USA. We demonstrate how this approach can be used to calculate measures, such as social network centrality, support contact dyads, and spatial kernel density based on a health survey data. Findings reveal that the integrated approach accurately reflected interactions between social and spatial elements, and identified community members (*who*) and locations (*where*) that should be prioritized for community-based health interventions.

Keywords

Community health; Rural Latino immigrants; Mental well-being; Network analysis; Spatial statistics; Intervention

Introduction

In the last ten years, community-based health researchers have increasingly used either social or spatial analytics to examine health issues and inform intervention programs. While this approach has enabled researchers to reveal social or geographic connections within a population, intervention programs based on this approach have encountered two primary problems: targeted individuals can be misidentified due to incomplete social network information, and a targeted intervention site may be visited by very few people from the population of interest. Problems of this type can largely reduce the effectiveness of community-based health interventions. In real life, social and spatial dimensions in a community tend to interact and should not be treated as mutually exclusive in analysis [1]. For instance, social connections may change with people's spatial locations, and in turn people's spatial activities are shaped by their social connections [2,3]. As a result of the interaction between these factors, in the past twenty years, social scientists have increasingly called for more integration between social and spatial information (Goodchild and Janelle 2004), including in public health studies and interventions [4]. However, to date, little attention has been paid to integrating social and spatial analytics in order to understand community health and inform "who" and "where" an intervention should target.

Thus, in order to reflect ongoing interactions between social and spatial elements in the same community, this methodological article presents a unique socio-spatial approach that aims to study data using: 1) a *social network with spatial weights*, and 2) a *spatial statistic with social weights*. The data under study was derived from an exploratory research conducted among Latino immigrants in rural North Florida, an area of interest for interventions to promote mental and physical well-being. Given that rural populations tend to be less mobile than their urban living counterparts and have limited places to visit in their communities, the proposed integrative

approach could well fit these hard-to-reach populations and better inform health intervention design in rural communities.

A review of social and spatial analyses in community-based health research

Social analysis in community-based health research

Because social relationships can dramatically impact the health outcomes of a community [5,6], social analytics have been used frequently in community-based health research. One type of social analytics—social network analysis—uses a finite number of *nodes* to represent individuals and their characteristics such as age, gender, and occupation. *Links* are assigned between nodes to represent either social relationships or interactions, and *link weights* are used to measure characteristics of these connections, such as the strength of social relationships or the frequency of interaction. Researchers collect data about participants from community-wide social/health surveys that obtain information about social contacts, such as characteristics of respondents' friends, sources of professional help, and frequency of interaction with other. Using these data, researchers describe the network structure by calculating variables such as the number of links attached to each node (a measure known as degree centrality) or identifying cohesive subgroups of individuals who are strongly connected to each other. Next, statistical models, such as logistic regression and multilevel modeling, are used to explore associations between network measures and health outcomes. Findings may also be used to rank individuals and groups according to health status or network characteristics, and identify those who can benefit most from, or can be most helpful for, interventions [7]. Most social network studies of community health are aspatial: they locate people in a topological space of social relationships, where spatial arrangements of individuals are not considered.

Spatial analysis in community-based health research

Spatial analysis aims to highlight a relationship among participants' residential location, movement between locations, and health outcomes, such as depression, obesity, and access to healthcare [8-10]. Data collection may include searching hospital records to discover how participants access health care, asking participants to complete a diary of their travels or a survey, or examining their location data from social networking websites or GPS loggers. Next, a range of spatial statistics are used to summarize participants' spatial activities [11], such as a two dimensional ellipse that envelopes all locations visited by each participant, and a spatial density map that shows clusters of locations visited by all participants. Findings often reveal areas associated with negative health outcomes, and help researchers identify locations that may benefit from public health interventions. In most of these spatial studies of community health, the spatial relationships between individuals are emphasized (for instance their geographic proximity), but the social attributes or connections between them are neglected.

Although a large body of research exists that uses either social or spatial analysis to examine health outcomes, the findings are limited because social and spatial dimensions have ongoing, dynamic relationships that are better examined in tandem[12]. The following section elaborates the proposed socio-spatial integrated approach: 1) how to integrate spatial weights into social network analysis; and 2) how to assimilate social weights into spatial statistics.

Methodology

Integrating spatial weights with social network

Based on evidence that suggests a strong correlation between geographic proximity and social ties [3,13], we constructed a spatially-weighted social network that uses nodes to represent individuals as well as their demographic, social, and health-related characteristics. The links between nodes represent social connections inferred from spatial relationships between

individuals. For example, if significant overlap exists in the locations visited by two individuals (also known as *co-location*), then we can infer that those individuals are likely to be socially linked. The weights on links indicate the degree of co-location between nodes (or the likelihood of social connection), and are estimated by following steps. Each individual i is assigned a $1 \times L$ vector V_i , where L is the total number of locations in a study region. Each element in V_i records the probability of individual i visiting a location, and hereinafter the vector V_i is referred to as the *location vector* for i . Using the location vectors of two individuals i and j , the degree of co-location w_{ij} between them is measured as the cosine of angle between V_i and V_j in a L -dimensional location space, commonly known as cosine similarity:

$$w_{ij} = \cos\theta_{ij} = \frac{V_i \times V_j}{\|V_i\| \|V_j\|} = \frac{\sum_{k=1}^L V_{ik} V_{kj}}{\sqrt{\sum_k V_{ik}^2} \sqrt{\sum_k V_{jk}^2}} \in [0,1] \quad (Eq. 1)$$

If the two location vectors are similar, the angle between them is close to zero and the cosine value is near to 1, indicating a high degree of co-location and a great probability of social connection.

Once a spatially-weighted social network is built, network measures can be derived to identify key individuals or groups for intervention purposes. For example, a key lay-community leader in an intervention should be highly connected to other members in the network. The individual connectivity can be measured as a weighted degree centrality of each node, the sum of all weights on the links associated with the node:

$$DC_i = \sum_{j=1}^n w_{ij} = \sum_{j=1}^n \cos\theta_{ij} \quad (Eq. 2)$$

A high value of DC_i indicates that individual i has a high chance of being socially connected with other individuals, and thus can potentially be successful in influencing others' health

behaviors. This would suggest prioritizing individuals with high values of DC_i when recruiting community leaders to work for intervention programs. In addition, the community leaders' well-being can positively influence others. Thus, community leaders can be identified by evaluating both degree centrality and well-being measure for each node. For example, intervention designers may recruit the most degree-central individuals in the top decile of the well-being distribution in the community.

A similar combination of network and health measures can be leveraged to foster cohesion and social support in a community. Social support has long been known as a major determinant of health [14] while social isolation and perceived loneliness have a demonstrated deleterious impact on health [15]. Reducing perceived isolation and loneliness has been a major aim of health interventions in the past several years [16]. Thus, identifying who can help whom and exchange support in a community could be a primary goal for an intervention. Given a community network, this goal can be achieved by deriving possible contact dyads between people with poor and good well-being scores. Such contact dyads should satisfy two criteria: 1) a high likelihood of contact between two nodes; and 2) a large difference between their well-being measures. Accordingly, we define a 'support score' between a pair of nodes i and j as:

$$SS_{ij} = \sum_{j=1}^n \cos \theta_{ij} \times |HS_i - HS_j| \quad (Eq. 3)$$

where HS_i and HS_j can be any health or well-being measure on nodes i and j . The support score can be used to identify the best support provider for a specific person with a poor health measure: given person i with poor health, the best support provider j for this person is the node that maximizes SS_{ij} .

Integrating social weights into spatial statistics

Kernel density estimation (KDE) is a widely used spatial statistic to estimate the intensity of a point-based process, i.e., the number of observed homicides per unit area around a given location [17]. Typically, the kernel density estimation is carried out on a grid of locations covering the study area. A kernel-function defined window is centered on each location, and estimates the point intensity around the location (Figure 1). After the kernel visits all locations on the grid, a surface is generated to describe the variation of point intensity (e.g., homicides) over the study area.

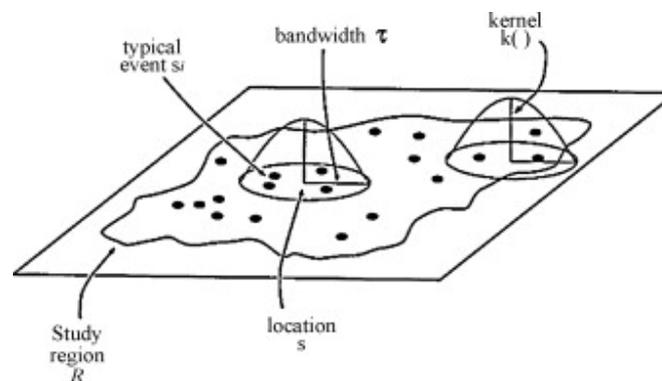


Figure 1. Diagram of how the quadratic kernel density estimation method works [18].

The first step of KDE is to center a kernel at a location s on a grid; each observed point event p ($=1, 2, 3, \dots, n$) is then assigned a weight according to the kernel function $k(d_{sp})$, which is a function of Euclidean distance from location s to each point p . In addition, the value measured at the point p , M_p (e.g., the number of homicides at p), can be included as a part of the weight. The density estimate at location s is the sum of the individual weight made from each observed point p (Eq. 4).

$$\hat{\lambda}_s = \sum_{p=1}^n M_p \times k(d_{sp}) \quad (\text{Eq. 4})$$

$$\text{where: } k(d_{sp}) = \frac{3}{\pi\tau^2} \left(1 - \frac{d_{sp}^2}{\tau^2}\right)^2; \quad d_{sp} < \tau \quad [19]$$

In the context of community health research, we view the point event p as a visit by an individual i to point p . The value measured at the point event p (M_p in Eq. 4) is now calculated as the product of the probability of individual i visiting point p (V_{ip} from the location vector) and the health measure of that individual (HS_i):

$$M_p = \sum_{i=1}^n HS_i \times V_{ip}$$

Thus, the kernel density estimation (Eq. 4) can be modified as:

$$\hat{\lambda}_s = \sum_{i=1}^n \sum_{p=1}^L HS_i \times V_{ip} \times k(d_{sp}) \quad (Eq. 5)$$

This KDE analysis integrates individuals' social/health characteristics with their spatial distribution and produces a density map that shows hotspots of visits, for example, where people with problems of social isolation (if HS_i is a measure of social isolation) or mental illness (if HS_i is an inverse measure of mental health) are highly likely to visit and cluster.

To exemplify our socio-spatial approach, we applied this method to data collected as part of a larger community-based research on rural Latino immigrants in three North Florida counties, with an overall goal of promoting well-being.

Case study of rural Latino immigrants in North Florida

According to the 2010 census, Latinos in Florida comprise 14.7% of rural and small town population in the state, surpassing African Americans (12.9%) as the largest minority group in rural and small town areas [20]. In the last 10 years, Florida has received a large number of Mexicans settling in rural areas that lack the significant social and health resources offered in urban areas such as Miami [21,22]. These rural Latinos tend to be young, poor, with low level of education and working primarily in farming or construction [22]. Because they often experience discrimination and feel excluded from local communities, rural Latino immigrants may avoid

certain social and spatial environments, which places them at risk for social isolation and lower levels of well-being [23-27]. Thus, in this methodological case study, we integrated social and spatial analyses to help inform the development of an intervention that targets social isolation and well-being in rural Latino immigrants in North Florida. Specifically, we aimed to: 1) identify community members who could be recruited as leaders for a well-being intervention; 2) identify dyads of community members who could provide support to each other to promote well-being; 3) map the intensity of participants' visits to non-home locations, particularly for those who felt socially isolated.

Data collection and geocoding

Data used in this case study included three rural communities in North Florida with vast farmland and scattered residences (Appendix Figure A1). Participants were rural Latino immigrants at least 25 years old, who had been living in the United States for up to 20 years, resided in one of three rural counties and spoke Spanish. The final sample for this case study was 60 participants (30 couple dyads). The study was approved by the university Institutional Review Board prior to study commencement.

The well-being of participants was assessed using measures of self-rated mental well-being, physical well-being, and perceived social isolation. Self-rated mental well-being of participants was gauged by the mental component score (MCS) of the SF-12v2™ Health Survey, which measures vitality, social functioning, effects of emotional problems on daily activities, and mental health. Self-rated physical well-being was determined by the physical component score (PCS) in the SF-12v2™ survey, which measures the effects of physical problems on daily activities, bodily pain, and general health. Proprietary scoring procedures were used to

standardize the mental and physical well-being scores between 0 and 100, with a mean of 50 and a Standard Deviation of 10. Higher scores reflected better perception of well-being.

The perceived social isolation was evaluated through a short version of the PROMIS Health Organization Social Isolation scale, which measures perceptions of being avoided, excluded, detached, disconnected from or unknown by others. The social isolation was also standardized as a score between 0 and 100. A score of 50 is the average for the general population in the United States (Std. Dev. =10), and higher scores indicate greater feelings of social isolation.

A questionnaire-based survey was delivered to obtain information regarding participants' spatial activities. Each participant was asked about routine activities in the past month, including the following questions: *Which ZIP code do you live in? Where have you traveled (e.g. specific places or ZIP codes) in the last month? What are the names of places you have visited in the last month? How long did it take you to get to each place? How often have you been there in the last month?* According to their reports, spatial activities of each participant were geocoded by following three steps:

- 1) *Home location.* To protect identity, the home of each participant was registered to a randomly selected location near to the population weighted mean center of the reported ZIP code (2010 census block populations were used as weights).

- 2) *Other activity places.* To limit respondent burden and to protect respondent privacy, participants were not asked to report specific names or addresses for their activity locations, but only the type of location (e.g., supermarket, church, and park) and their travel time to that location. This information was used to identify potential activity locations, as follows. First, the reported travel time was used to generate a travel ring including all areas that are reachable by

car from the participant's home location within the specified travel time (for example, in 0-10, 11-20, 21-30, or 31-40 minutes). The road network data of Florida and the ArcGIS network analyst were utilized to create these travel rings (Appendix Figure A2). Second, the travel ring was superimposed onto Google Earth to identify potential destinations of the type indicated by the respondent. For example, if '15 minutes to a church' was reported for a trip, a travel ring of 11-20 minutes was generated. All churches shown on the Google Earth and within the 11-20 minute travel ring were selected as potential places for this trip. Their latitude and longitude were extracted from Google Earth for subsequent mapping and analysis. Although the workplace is a critical place of daily activities, participants seldom reported this information, perhaps to protect their privacy. Hence, workplace locations were not mapped in this study.

3) *Visiting probability*. Each potential place was also associated with a daily probability of visit by a participant, computed from the reported frequency of visits by the participant during the last month. Specifically, the daily visit probability was set as 1 if the participant reported visiting the place 'every day' (e.g., the home), as 1/7 if the reported frequency was 'once a week', or as 1/30 if 'once a month'. If there were multiple potential places for an activity (as identified in step 2), the daily probability was prorated to each place by dividing it by the number of potential places. The visiting probabilities were used to construct the location vector V_i for each participant, used in Eq. 1.

Spatially weighted social network

Following the methodology described above, we built a 'potential contact' network composed of 60 participants as nodes. For each node, the scores for mental and physical well-being and social isolation were assigned as attributes. Links between participants were created and weighted based on Eq. 1, indicating the likelihood of social connections.

Using this network, we calculated the weighted degree centrality of each node (Eq. 2) to identify potential community leaders, who should have high degree centrality as well as high well-being scores. To find out who can help those with poor well-being, we calculated the ‘support score’ between every pair of nodes (Eq. 3), in terms of mental well-being and social isolation, respectively. We then identified potential supporters to help participants with low mental well-being scores or high social-isolation scores.

Socially weighted kernel density estimation

Here, we considered each visit to a non-home place by a participant i as a point event p in Eq. 5. We used the social isolation score as the health measure (HS_i) in Eq. 5, and thus the kernel density estimation detected hotspots of visits by socially isolated participants. A greater kernel density around a location indicates more socially isolated people visiting this location. Therefore, outreach and intervention activities at locations with higher density would be more likely to reach disadvantaged pockets of the community.

Results

Demographics

Participants were 26-45 years old, primarily from Mexico, who had been living in the United States for more than ten years and had 3-4 children living with them. Women had slightly more education than men, yet the employment rate was higher for men. More information about participants’ demographics can be found in Appendix Table 1. The social isolation mean score (39.7) was lower than the national average (50), with a few participants ($n=13$) scoring above the national average; but all scores fell within one standard deviation. The averages of mental and physical well-being scores were 54.2 and 54.4, respectively, both higher than the national average (50), indicating an overall better perception of well-being. There were a few participants

with relatively low scores, shown as outliers (Figure 2a). For spatial activities (Figure 2b), ‘shopping’ was the most frequently reported activity by participants, followed by ‘eating out’, ‘visiting friends’ and ‘going to church’. On average, women visited more places (6.4) in one month than the men visited in the same period (5.2).

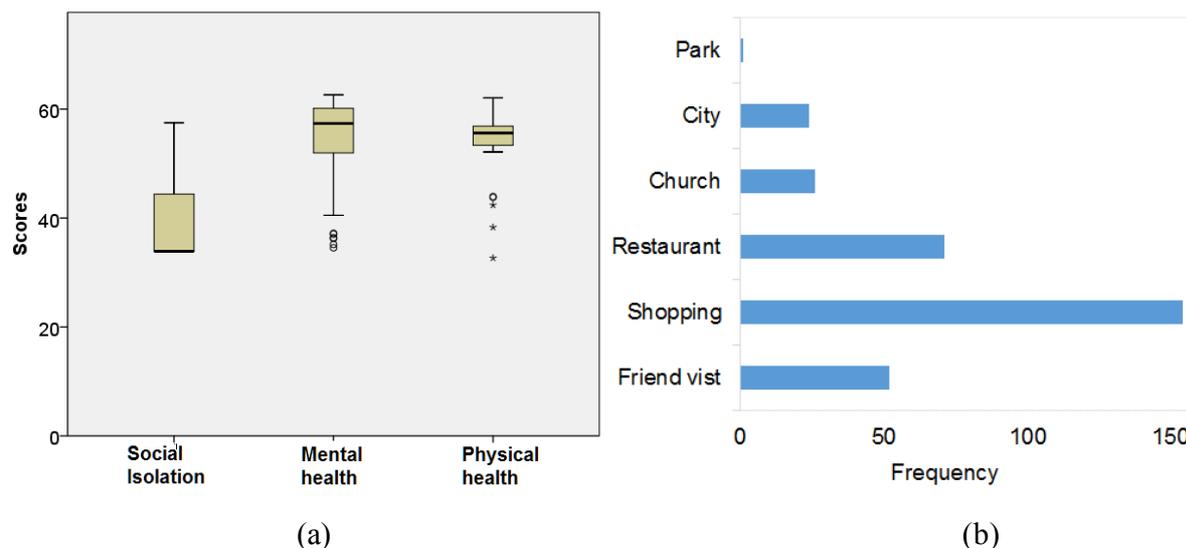


Figure 2. Statistical summary of surveyed data: a) Boxplot for three health scores of participants; b) Reported frequency of visiting different types of locations.

Spatially weighted social network analysis

The ‘potential contact’ network derived from this study was composed of 60 nodes (30 couples) and had 1394 links among them. A full structure of this network is visualized in Appendix Figure A3. For clearer visualization and analysis, we subset this large network into three sub-networks by ZIP code (community) (Figure 3), which helps identify potential community leaders with high degree centrality with good well-being scores.

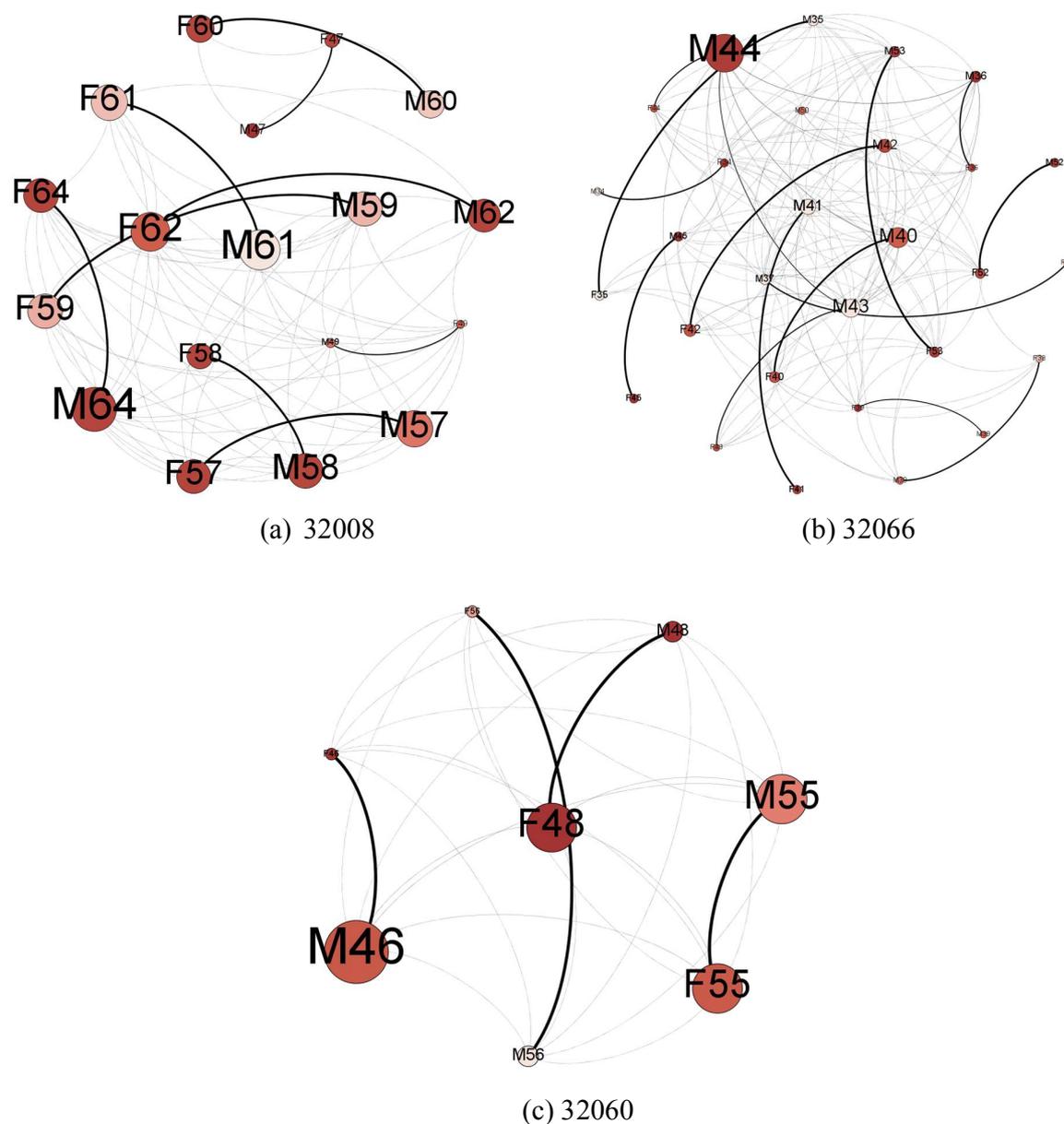


Figure 3. The inferred social network among Latino participants in each of the three communities (ZIP codes). The link width was scaled by the likelihood of social connection. The label ‘F’ stands for ‘Father’ and ‘M’ for ‘Mother’. The size of nodes indicates their weighted degree centrality, and the tone of color shows their mental well-being scores (the darker the color, the greater the mental well-being).

Figure 4 depicts possible contact dyads between participants with poor and good well-being scores. Here, ‘poor’ well-being refers to the social isolation score being above 50 or the mental well-being score being under 50. Those nodes with poor well-being were circled in Figure 4 and sized proportionally to their scores of social isolation (Figure 4a) and mental well-being (Figure 4b), indicating those who need more support from a potential intervention. Nodes i and j are connected by a link whose weight indicates the support score SS_{ij} (Eq. 3). To obtain a clearer network structure, the (i, j) link is only displayed if its support score is sufficiently large ($SS_{ij} > 1$). For each circled node, the optimal provider of support can be determined by comparing the width of links attached to it.

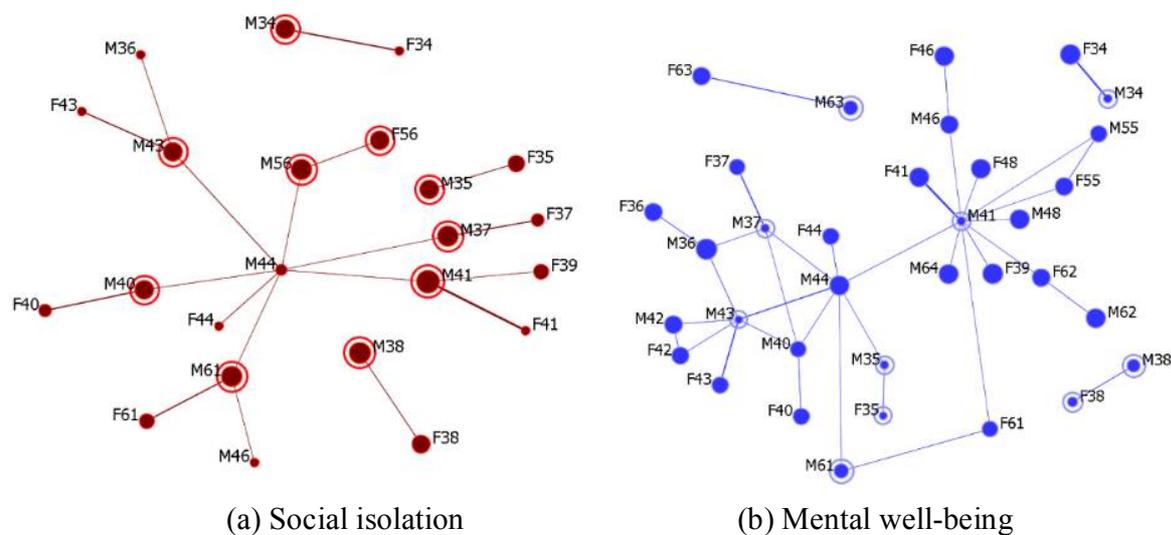


Figure 4. The potential dyads to identify supporters. Node size indicates the social isolation and mental well-being scores, while the links' width is proportional to the support score. For graph clarity, the links are only shown if their weights (the support scores) are sufficiently large (above 1.0).

Socially weighted kernel density estimation

Based on the survey of spatial activities, we identified 110 churches, 101 restaurants, 47 friend visits, and 33 shops (e.g., supermarkets and grocery stores). Generated from the kernel density function, Figure 5 shows four areas with the highest density (dark areas with labels), where visits of participants with high social isolation scores tended to cluster. Area 1, 2, and 3 (in the north and northwest) are completely or partially in participants' home communities (ZIP code areas). Area 4 in the northeast (a city) is about 30 to 45-minute driving distance to the three studied communities where the survey took place.

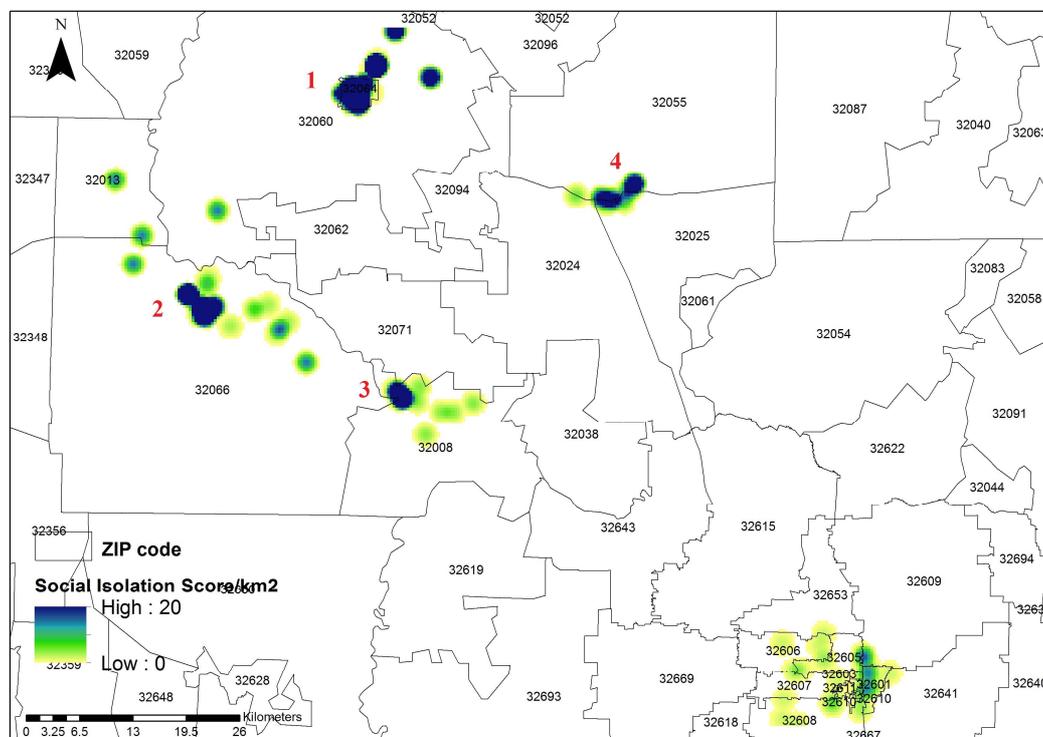


Figure 5. Kernel density map of visiting intensity of socially isolated participants. Dark colored areas indicate locations where participants with high social isolation scores were more likely to visit.

Discussion

Inferring social networks from spatial data

In community-based health research, most social network analyses rely on self-reports from study participants about their social ties. Collection of these data, however, might be excessively costly, time-consuming, or burdensome for respondents, particularly in certain disadvantaged populations (e.g., immigrant farmworkers with limited free time or people with disabilities that prevent them from taking long surveys). In addition, self-reports on social interactions might suffer from cognitive biases, errors of perception, and ambiguities [28]. In certain extreme settings, social network data collection is not feasible or is more strongly affected by respondent accuracy problems. For example, that can be the case if the community of interest is very large and social ties are too many to collect; or if the population under study is stigmatized (e.g., IV drug users) and respondents are reluctant to name other people in the population. In these cases, researchers can resort to the method described here and infer networks of potential interactions using several geospatial data sources, such as self-reports on activity locations, geo-tagged photos from Flickr, geo-tagged posts on Facebook, or mobile phone tracking data. Particularly in a rural community, the population is smaller in size, tend to be less mobile, and has limited places to visit, as compared to its urban counterpart. The co-location of two individuals is more likely to be a real social tie, and thus our inferential approach can produce better prediction of real social network.

In addition to the degree centrality used here (Eq. 2), betweenness centrality and bridging measures can also be calculated in the spatially weighted social network to identify a key lay-community leader. These two measures index the extent to which a node falls in between all other nodes, bridging otherwise unconnected areas of the network [29,30]. Relevant to our case study, participants with high betweenness or bridging values are people who visited a diverse set of locations, which are normally visited by very different groups of people. In other words, high-

betweenness individuals have a high likelihood of co-location and potential interaction with different and distant segments of the community, which are otherwise separate or spatially segregated from each other. Thus, in addition to community members with high degree centrality, individuals in bridging positions can also be very effective intervention champions, particularly in more fragmented communities with disconnected subgroups.

Who to intervene with: Identifying opinion leaders and social supporters

Figure 3 identifies the top two candidate leaders for each community who have high degree centrality and good well-being scores (Table 1). For example, Respondent M44 in ZIP code 32066, F55 in 32060, and M64 in 32008 all have high mental and physical well-being scores and low social isolation scores. These participants, with their wide connections to other people and positive influence, could assist researchers in the next stage of this study to deliver an intervention and promote the well-being of other community members.

Table 1. Top 2 potential community leaders by ZIP code area and their social-health measures

Node ID ^a	ZIP Code area	Degree centrality	Mental well-being score	Physical well-being score	Social isolation score
M44	32066	1.69	60.16	55.85	39.1
M42	32066	1.25	51.96	57.17	43.1
F55	32060	1.23	57.38	57.12	33.9
M46	32060	1.24	57.53	54.64	33.9
M64	32008	1.29	60.32	53.37	33.9
F62	32008	1.18	57.53	54.64	39.1

^aM stands for mother, while F stands for father.

Although spouses could often be the first candidate for social support, the network graphs in Figure 4 also suggest potential supporters outside individuals' nuclear families. For example, in Figure 4a, participant M44, the potential community leader of ZIP code 32066 (Table 1), may be an effective support provider for M40, M41, M56, and M37, who all reported high feelings of social isolation. Likewise, M46, as a potential leader in ZIP code 32060, could support M61. The supporters do not necessarily have to be potential community leaders listed in Table 1. For

instance, M36 could be a good supporter for M43, and F39 could be a good supporter for M41. Similarly, those who perceived poor mental well-being can find their helpers from Figure 4b.

Where to intervene: Prioritizing locations for outreach and intervention efforts

The kernel density estimation has been widely used to study people's spatial behavior, but often with no consideration for people's social or health attributes. In these applications, the identified hotspots only show where people are more likely to visit and cluster in general, but do not differentiate subgroups of interest, such as those who are socially isolated, at risk of mental illness, or with poor well-being. Interventions deployed in these hotspots may be effective at reaching people in general, but may be less likely to reach the subgroups of interest, and thus are less cost-effective. In contrast, by incorporating social or health-related weights into kernel density estimation, hotspots of interest can be more accurately identified, where specific, higher-priority subgroups of people tend to cluster. In our application, the resulting kernel density map (Figure 5) shows that the activities of most socially isolated participants were concentrated in the vicinity of their home locations, but many of them also traveled long distances to medium-size cities (two medium high-density areas in the south and southeast). In those hotspots, we identified three places where socially isolated participants were likely to gather, namely the Store A (in ZIP code 32008), Store B (32060), and Store C (32066). These are key places where community interventions can be deployed to more efficiently reach higher numbers of individuals in priority populations.

Although we focused on social isolation, similar analyses can be performed for mental and physical well-being, which are not illustrated here. Differently, the health score HS_i in Eq. 5 should be an inverse measure of health or well-being, such that participants with lower well-being have higher HS_i scores and higher weights (resulting in higher kernel density). For

example, if MCS_i is the mental component score of the SF-12v2™ Health Survey, then we could define $HS_i = 100 - MCS_i$ for KDE analysis

Limitations

Three limitations of this study should be noted. First, part of the analysis was based on a social network inferred from co-locations of participants' activities. This should be interpreted as a network of *potential* face-to-face encounters and interactions among participants, due to the spatial proximity of their daily locations; however, this network does not necessarily reflect *actual* face-to-face contacts among individuals. Participants who appear as connected in the network do not necessarily know or trust each other. The actual network of acquaintances, friendships or communications among participants could be different. This limitation should be kept in mind when interpreting the meaning of centrality measures, community leaders, or support scores derived from the network. Nevertheless, the spatial inferential approach has been increasingly adopted by network-related health studies, and is considered reliable and valuable when actual social network surveys are not feasible [13,31,32].

Second, the identification and geocoding of participants' activity locations was at times problematic. Many participants reported vague names for places they had visited, and were reluctant to provide their workplace information, which made their complete activity space difficult to map. To address this problem, our method assigned all possible locations of the specified type and within the specified travel ring to each participant's activity space. Although this ensures that individual privacy is protected, the incomplete information can introduce uncertainty in the structure and analysis of the resulting spatially-weighted network and kernel density estimation. This is, however, a data problem that can be addressed using GPS equipped

devices or Google Map during the survey, and does not affect the validity of our proposed approach.

Third, the sample size is relatively small in the socio-spatial analysis (30 women and men living together) and not random, which is typical of hidden or hard-to-reach populations such as immigrants in rural areas [33]. Our approach, however, can handle a larger sample and produce results that are more reliable. While all these limitations suggest that a more sophisticated and larger-scale survey would be beneficial in community health applications, they do not diminish the value of the integrated socio-spatial analysis approach. Such an approach seems a promising methodological tool for studying health disparities and designing health interventions in underrepresented and isolated communities.

Conclusions

Geographic space and social space are closely intertwined, and research should attempt to analyze them together to identify the “who” and “where” of health intervention endeavors. Previous studies have mostly focused on one of the two spaces, overlooking their interaction and the additional insights that can be gained from an integration. This suggests that there is much room for improving intervention programs by leveraging results from joint socio-spatial analyses. We offered one proposal for a socio-spatial analytic approach that integrates spatial co-location into social network analysis, and social/health measures into spatial kernel density estimation. Other ways to combine analyses of social and spatial data are certainly possible, and future research in community, public and rural health would greatly benefit from further investigating such integration. We believe that this line of work would enrich the methodological body of community health research, and bring real team-based science contributions, particularly when working with hard-to-reach rural populations, such as: 1) optimizing recruitment, retention and

eliminating barriers to participation of minority and rural populations in community interventions; 2) facilitating community partnerships with individuals and organizations; 3) assisting in negotiating intervention sustainment, and minimizing socio-spatial inequities for community-based interventions engagement.

References

1. Goodman, R.A.; Bunnell, R.; Posner, S.F. What is “community health”? Examining the meaning of an evolving field in public health. *Preventive medicine* **2014**, *67*, S58-S61.
2. Yin, L.; Shaw, S.-L. Exploring space–time paths in physical and social closeness spaces: A space–time gis approach. *International Journal of Geographical Information Science* **2015**, *29*, 742-761.
3. Toole, J.L.; Herrera-Yaqué, C.; Schneider, C.M.; González, M.C. Coupling human mobility and social ties. *Journal of The Royal Society Interface* **2015**, *12*, 20141128.
4. Gatrell, A.C.; Rigby, J.E. Spatial perspectives in public health. *MF a. J. Goodchild, DG (Ed.), Spatially Integrated Social Science: Examples in Best Practice* **2004**, 366-380.
5. Smith, K.P.; Christakis, N.A. Social networks and health. *Annu. Rev. Sociol* **2008**, *34*, 405-429.
6. Berkman, L.F.; Kawachi, I.; Glymour, M.M. *Social epidemiology*. Oxford University Press: 2014.
7. Valente, T.W. Network interventions. *Science* **2012**, *337*, 49-53.
8. McCray, T.; Brais, N. Exploring the role of transportation in fostering social exclusion: The use of gis to support qualitative data. *Networks and Spatial Economics* **2007**, *7*, 397-412.

9. de Castro, J.M.; King, G.A.; Duarte-Gardea, M.; Gonzalez-Ayala, S.; Kooshian, C.H. Overweight and obese humans overeat away from home. *Appetite* **2012**, *59*, 204-211.
10. Grubestic, T.H.; Durbin, K.M. Community rates of breastfeeding initiation: A geospatial analysis of kentucky. *Journal of Human Lactation* **2016**, *32*, 601-610.
11. Schönfelder, S.; Axhausen, K.W. Activity spaces: Measures of social exclusion? *Transport policy* **2003**, *10*, 273-286.
12. Adams, J.; Faust, K.; Lovasi, G.S. Capturing context: Integrating spatial and social network analyses. *Social Networks* **2012**, *34*, 1-5.
13. Onnela, J.-P.; Arbesman, S.; González, M.C.; Barabási, A.-L.; Christakis, N.A. Geographic constraints on social network groups. *PLoS one* **2011**, *6*, e16939.
14. House, J.S.; Umberson, D.; Landis, K.R. Structures and processes of social support. *Annual review of sociology* **1988**, *14*, 293-318.
15. Cacioppo, J.T.; Cacioppo, S. Social relationships and health: The toxic effects of perceived social isolation. *Social and personality psychology compass* **2014**, *8*, 58-72.
16. Masi, C.M.; Chen, H.-Y.; Hawkey, L.C.; Cacioppo, J.T. A meta-analysis of interventions to reduce loneliness. *Personality and Social Psychology Review* **2011**, *15*, 219-266.
17. Rogerson, P.; Yamada, I. *Statistical detection and surveillance of geographic clusters*. CRC Press: 2008.
18. Bailey, T.C.; Gatrell, A.C. *Interactive spatial data analysis*. Longman Scientific & Technical Essex: 1995; Vol. 413.
19. Diggle, P. A kernel method for smoothing point process data. *Applied statistics* **1985**, 138-147.

20. HAC. Race & ethnicity in rural america. In *Rural Research Briefs, April*, Housing Assistant Council: Washington DC, 2012; pp 1283-1288.
21. Terrazas, A. Mexican immigrants in the united states. February: 2010.
22. Douglas, K.M.; Saenz, R. No phone, no vehicle, no english, and no citizenship: The vulnerability of mexican immigrants in the united states. *Globalization and America: race, human rights, and inequality* **2008**, 161-180.
23. Williams, D.R.; Mohammed, S.A. Discrimination and racial disparities in health: Evidence and needed research. *Journal of behavioral medicine* **2009**, *32*, 20-47.
24. Stacciarini, J.-M.R.; Smith, R.; Garvan, C.W.; Wiens, B.; Cottler, L.B. Rural latinos' mental wellbeing: A mixed-methods pilot study of family, environment and social isolation factors. *Community mental health journal* **2015**, *51*, 404-413.
25. Stacciarini, J.-M.R.; Smith, R.F.; Wiens, B.; Pérez, A.; Locke, B.; LaFlam, M. I didn't ask to come to this country... i was a child: The mental health implications of growing up undocumented. *Journal of Immigrant and Minority Health* **2015**, *17*, 1225-1230.
26. Mora, D.C.; Grzywacz, J.G.; Anderson, A.M.; Chen, H.; Arcury, T.A.; Marín, A.J.; Quandt, S.A. Social isolation among latino workers in rural north carolina: Exposure and health implications. *Journal of Immigrant and Minority Health* **2014**, *16*, 822-830.
27. Ko, L.K.; Perreira, K.M. "It turned my world upside down": Latino youths' perspectives on immigration. *Journal of adolescent research* **2010**, *25*, 465-493.
28. Bernard, H.R.; Killworth, P.D.; Sailer, L. Informant accuracy in social-network data v. An experimental attempt to predict actual communication from recall data. *Social Science Research* **1982**, *11*, 30-66.

29. Everett, M.G.; Valente, T.W. Bridging, brokerage and betweenness. *Social networks* **2016**, *44*, 202-208.
30. Valente, T.W.; Fujimoto, K. Bridging: Locating critical connectors in a network. *Social networks* **2010**, *32*, 212-220.
31. Eagle, N.; Pentland, A.S.; Lazer, D. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences* **2009**, *106*, 15274-15278.
32. Scellato, S.; Noulas, A.; Mascolo, C. In *Exploiting place features in link prediction on location-based social networks*, Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, 2011; ACM: pp 1046-1054.
33. Spreen, M. Rare populations, hidden populations, and link-tracing designs: What and why? *Bulletin of Sociological Methodology/Bulletin de Methodologie Sociologique* **1992**, *36*, 34-58.
34. Mao, L.; Stacciarini, J.-M.R.; Smith, R.; Wiens, B. An individual-based rurality measure and its health application: A case study of latino immigrants in north florida, USA. *Social science & medicine* **2015**, *147*, 300-308.

Appendix

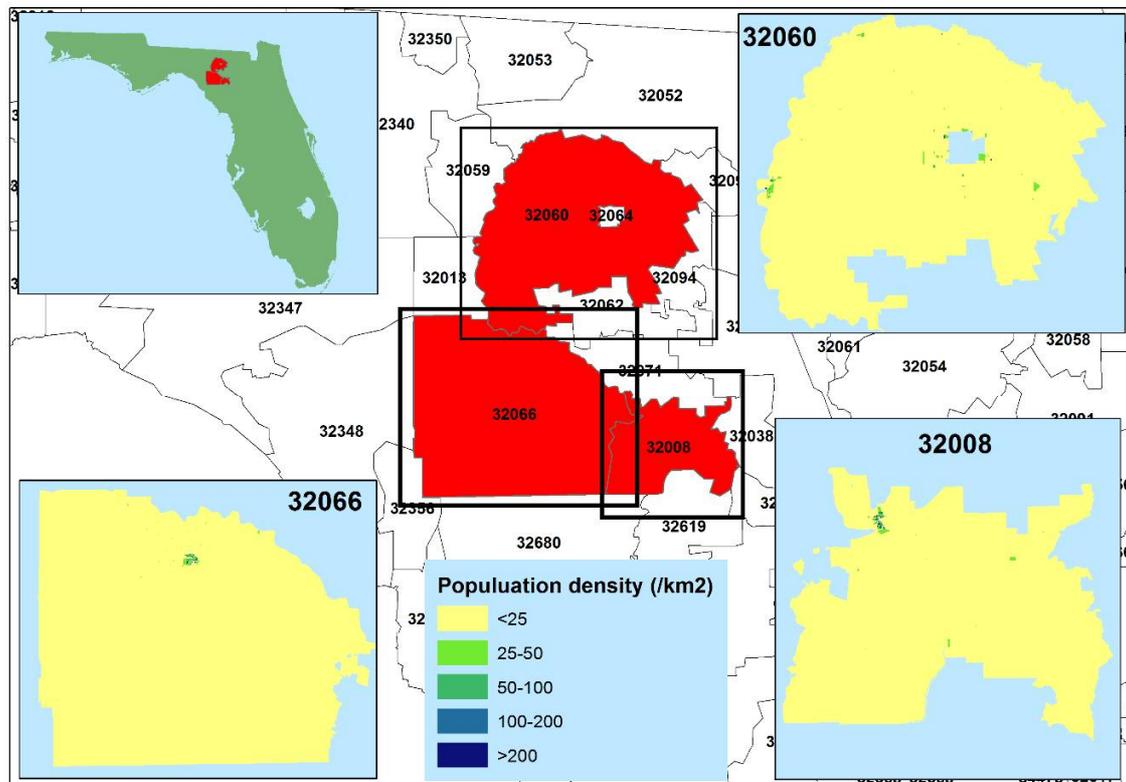


Figure A1. The central map shows the spatial extent of three surveyed communities (ZIP code areas). The upper left map indicates the geographic location of surveyed areas in Florida. Maps at other three corners depict the density of general population within three communities.

Table A1. Demographic summary of participants^a

	Women (n=30)	Men (n=30)
Country of Origin		
El Salvador	6.7%	6.7%
Mexico	90.0%	93.3%
Guatemala	3.3%	0%
Language (Spanish)	100%	100%
Years in US		
10 or less	13.3%	16.6%
More than 10	86.7%	83.4%
Age		
26-35	33.3%	33.3%
36-45	53.4%	50%
46-65	13.3%	16.7%
Number of children*		
1-2	24.1%	26.7%
3-4	65.5%	66.7%
5-6	10.4%	6.6%
Children in country of origin**		
No	92.9%	89.7%
Yes	7.1%	10.3%
Employed		
No	33.3%	3.3%
Yes	66.7%	96.7%
Years of education		
0	0%	6.7%
1-4	20%	13.4%
5-8	36.7%	46.6%
9-14	43.3%	33.3%
Family/friends nearby		
No	13.3%	6.7%
Yes	86.7%	93.3%
Mean number of places visited in last month***	6.4	5.2
Mean number of cities visited in last month*	2.9	2.5

Note: ^a This table is adopted from a previous publication[34]

* 1 missing, valid percentages used, **2 missing, valid percentages used ***>2 missing

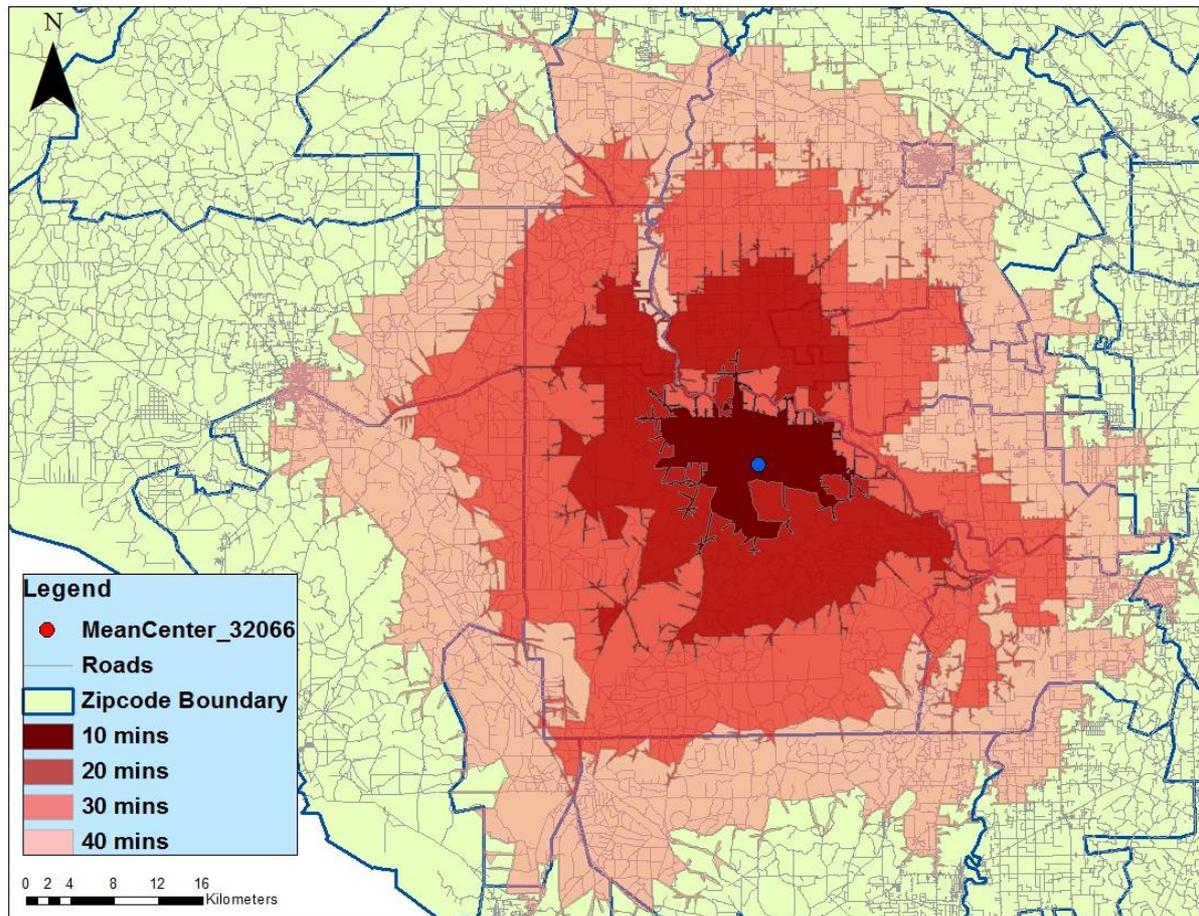


Figure A2. Travel rings of 0-10, 11-20, 21-30, 31-40 minutes around the population mean center of a ZIP code area 32066, created based on a real road network.

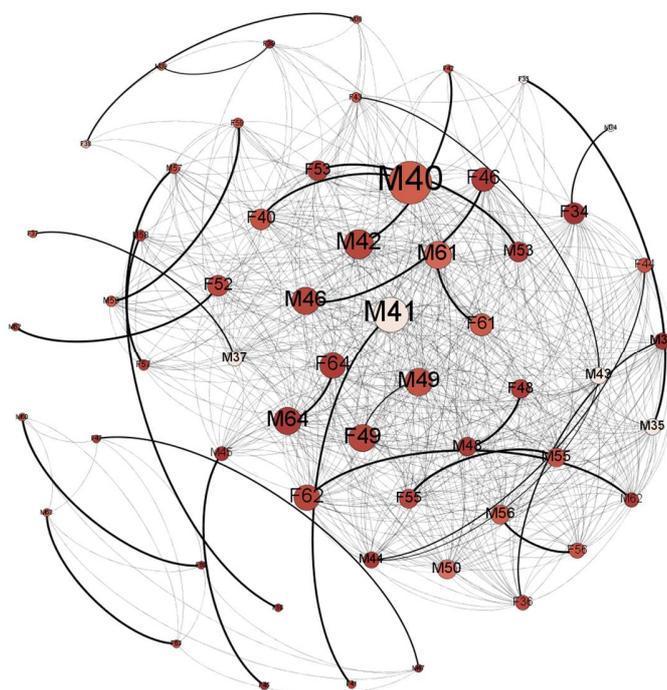


Figure A3. The inferred social network among Latino participants from all three communities combined. The link width was scaled by the likelihood of social connection. The label ‘F’ stands for ‘Father’ and ‘M’ for ‘Mother’. The size of nodes indicates their weighted degree centrality, and the tone of color shows their mental well-being scores (the darker the color, the greater the mental well-being).