Appendix 2: Data Collection and Attribute description

Data collection and cleaning is, as usual, an arduous task. Organizations were defined in the survey as "a group as having two or more members. This survey is intended for groups and organizations, not for individuals working alone, as sole proprietors, or as independent contractors. If you are a member of a group with multiple programs, please answer for your entire organization. If you are a respondent from a national organization, please answer for your local chapter." Cleaning consisted of getting an accurate list of organizations when referenced differently and combining organizations that were actually sub-organizations of a larger organization. Our quality assurance and data cleaning process dealt with inaccuracies by excluding individuals and by assigning branches/teams/programs to associated organization. Therefore these data are all standardized and cleaned before conducting SNA analysis. This process was the same for both Philadelphia and New York City.

Total respondents for the STEW-MAP survey in each city (New York City, n = 506; Philadelphia, n = 195) differs from respondents retained in each city's network model (New York City, n = 335; Philadelphia, n = 151), as not all respondents named network partners. We compared non-respondents to responding organizations by home office location and the type of organization (business, civic, governmental, and school). We saw no clear spatial pattern to the non-responses in either city. For organization type, the numbers of organizations not responding of each category was not significantly different than the organizational distribution of respondents. This was significantly different in New York, however, due to a large number of non-responding civic organizations. From looking at the names, in general these were small community groups. We speculate that this difference was due more to getting additional access to lists of small organizations in New York City than in Philadelphia, and that the actual networks of respondents are still comparable. Future research is planned to take advantage of social media and internet hyperlink data to get at changes in composition over time without relying on respondents. However, our data collection is still an almost an order of magnitude greater than most environmental social network studies (Groce et al. 2018).

There were many instances where respondents also named partners who were not part of the initial survey sample. Attribute information for these organizations was supplemented by internet searches. This additional data collection increased the number of non-respondents in each city (New York City n=440; Philadelphia, n=431). Where many network designs would remove alters that were not included in the survey (Robins et al. 2012), we maintain that including this data is critical for a better understanding of the networks in each city. For collaborating organizations that did not also respond to the survey (i.e., alters), a subset of the same organizational characteristics (501(c)(3) status, staff size, budget, mission, group focus), along with office location addresses, were identified through web searches and phone calls to verify our information. For organizations where home offices were identified, we could then also categorize their home office location according to US Census and land use data. See Appendix 1 for more information on handling the missing data.

Additional variables incorporated into network ERG models were collected from the STEW-MAP datasets and US Census American Community Survey socioeconomic data (2005 – 2009 5-year ACS data for New York City, US Census Bureau 2018a; 2011-2015 5-year ACS for Philadelphia, US Census Bureau 2018b). Variables derived from the STEW-

MAP survey include measures of both the organizational characteristics and neighborhood context. These include attributes of the organizations themselves such as primary focus area, measures of organizational formalization (i.e., staff, membership, volunteers, 501(c)(3) status). We did collect information on the age of the organization, but were unable to get models to converge when this variable was included. This might be due to the high correlations - see Figure A2.1 for correlations between vertex attributes. Paid staff, membership, and volunteers each were recoded as a presence/absence dichotomous variable. To assess neighborhood context, we classified the predominant land use of the census block where the home office was located, and we also used American Community Survey (ACS) data to characterize the census block group. The measures included represent the population density of the block, the median household income, and the median year individuals moved to the block (for residential areas). These variables were also dichotomized by calculating the median of the total observations within each city and coding each block as above (1) or below (0) this median. Finally, spatial measures were included in three ways: (1) distance between office locations, (2) overlap of the stewardship turf where each group has activities, and (3) the geographic extent of each group's stewardship turf (i.e., sub-neighborhood - small; across 2-5 neighborhoods - medium; and larger than 5 neighborhoods - large).

Figure A2.1 shows the correlations between the different vertex attributes (anything pertaining to a specific organization - this contrasts to edge or tie attributes like the distance between home offices). Only the significant correlations, as judged by a $p<0.05$ from a Pearson Correlation test, are displayed. From the correlation of the variables, we see that Age is negatively correlated with many other organizational attributes in New York City but not Philadelphia. We also see a much tighter correlation in New York City between the Census data for the home office block; in New York City, those with lower than median year moved are more likely to have higher than median household incomes and larger than median population density. These relationships are not seen in Philadelphia. We see many significant correlations between many of the different issue foci (note that Religion is not present in the New York City map as there were too few organizations with that focus to provide meaningful results). We see in both cities a correlation between Environment and Education foci as well as Youth and Seniors. None of the correlations are too high to worry about multicollinearity, and this is confirmed by our VIF tests with no score above 15 (Duxbury 2018).

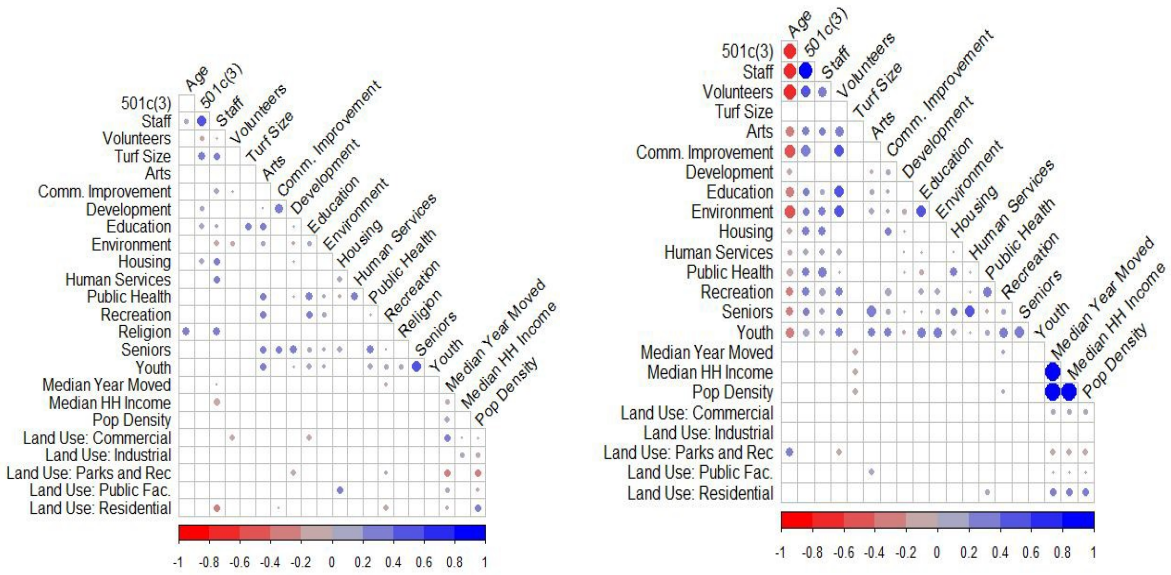

Figure A2.1: Correlation plots for vertex attributes. Philadelphia (left) and New York City (right)

Literature Cited:

Duxbury, Scott W. 2018. "Diagnosing multicollinearity in exponential random graph models." Sociological Methods & Research: 0049124118782543.

Groce, J. E., M. A. Farrelly, B. S. Jorgensen, and C. N. Cook. 2018. Using social-network research to improve outcomes in natural resource management. Conservation Biology.

U.S. Census Bureau. 2018*a*. *2005-2009 5-year American community survey*. U.S. Census Bureau, Washington, D.C., USA. [online] URL: https://www.census.gov/programs-surveys/acs/data/summary-file.2009.html

U.S. Census Bureau. 2018b. *2011-2015 5-year American community survey*. U.S. Census Bureau, Washington, D.C., USA. [online] URL: https://www.census.gov/programs-surveys/acs/data/summary-file.2015.html