A Spatial and Regression Analysis of Social Media in the United States Counties

Completed Research Paper

James B. Pick, Avijit Sarkar, and Jessica Rosales
School of Business
University of Redlands, Redlands, CA, USA
Contact: james_pick@redlands.edu

Abstract

The locational distribution and socio-economic determinants of social media are analyzed for the United States counties in 2012. A theory of determinants is presented that is modified from the Spatially Aware Technology Utilization Model (SATUM). Socio-Economic factors including demography, economy, education, innovation, and social capital are posited to influence social media factors, while spatial analysis is conducted including exploratory analysis of geographic distribution and confirmatory screening for spatial randomness. The determinants are identified through OLS regression analysis. Findings for the nation indicate that the major determinants are demographic factors, service occupations, ethnicities, and urban location. Further subsample analysis is conducted for the U.S. metropolitan, micropolitan, and rural subsamples. The subsamples differ most evidently in effects of ethnicities and construction occupations, and there are inverse effects of social capital at the micropolitan and rural levels.

The regression findings are discussed in terms of the literature mostly of larger geographic units, and the few nationwide studies at the county level. The exploratory spatial analysis generally indicates similar national geographic patterns of use. Among the results is that although Twitter users are more heavily concentrated in southern California and have strong presence in the lower Mississippi region, Facebook users are highly concentrated in Colorado, Utah and adjacent Rocky Mountain States. Social media usage is lowest in the Great Plains, lower Midwest, and South with the exceptions of Florida and the major southern cities such as Atlanta. The overall extent of spatial agglomeration is very high and is examined in detail for the nation and subsamples. The paper concludes by discussing the policy implications of the analysis at the county as well as the national levels.

Keywords. Social media, spatial analysis, United States, counties, regression, spatial autocorrelation, policies

Introduction and Literature Review

In this paper, we examine the demographic, socio-economic, governmental, and societal openness influences on information and communications technology (ICT) adoption and utilization in the counties of the United States. As the world's largest economy, and a leading developed nation, the United States has become one of the world's leading information societies. In 2012, the US was ranked seventeenth among 157 nations by the International Telecommunication Union in terms of the ICT Development Index, a composite index comprised of eleven indicators of ICT access, utilization, and skills. While the US lags advanced societies such as the Nordic nations, Singapore, United Kingdom, smaller European nations such as Netherlands and Switzerland, and countries such as Australia and New Zealand in terms of ICT development, the US is among the most populous nations worldwide and is also diverse in terms of factors that have been traditionally acknowledged to influence the digital divide within and between nations.

The digital divide has been defined as "the gap between individuals, households, businesses and geographic areas at different socio-economic levels with regard both to their opportunities to access information and communication technologies and to their use of the Internet for a wide variety of activities (OECD, 2011). Data from the 2010 Current Population Survey (CPS) of the US Census Bureau indicate that as of October 2010, 68 percent of households accessed broadband internet service, an increase of 64 percent since 2000. During the same time period, landline telephone subscription unsurprisingly grew by only 2 percent. On the other hand, at 310 million mobile cellular telephone subscribers, the US ranked third worldwide behind China and India in terms of this indicator in 2012. The CPS also found that more than three quarters of households owned a computer, the primary means by which households accessed the internet. Given that the United States is a large and increasingly diverse nation in terms of social, political, economic, and demographic attributes, it is important to analyze their impacts on ICT adoption and utilization, and on the digital divide within the US.

Overall, our research questions are the following:

(1) what are the factors that impact social media use in US counties,

(2) how does social media use vary geographically across US counties, and

(3) what county policies are implied from the findings of this research?

Conceptual Theory

The paper's conceptual model is drawn from the Spatially Aware Technology Utilization Model (SATUM), which is appropriate for research on composite influences of various social, economic, and political determinants on ICTs, which can include social media variables. SATUM is based on a large literature of studies, mostly of nations, but also of states, provinces, prefectures, and EU economic units (Nishida, Pick, and Sarkar, 2015; Pick, Sarkar, and Johnson, 2015; Pick and Sarkar, 2015; Vicente and Lopez, 2011). SATUM also has spatial relationships as components, so a research study can assess the extent of spatial bias in multivariate statistical analysis and can perform spatial cluster analysis and other spatial techniques (Farkas et al.,, 2016, in press). The present study's SATUM-based model is given in Figure 1.

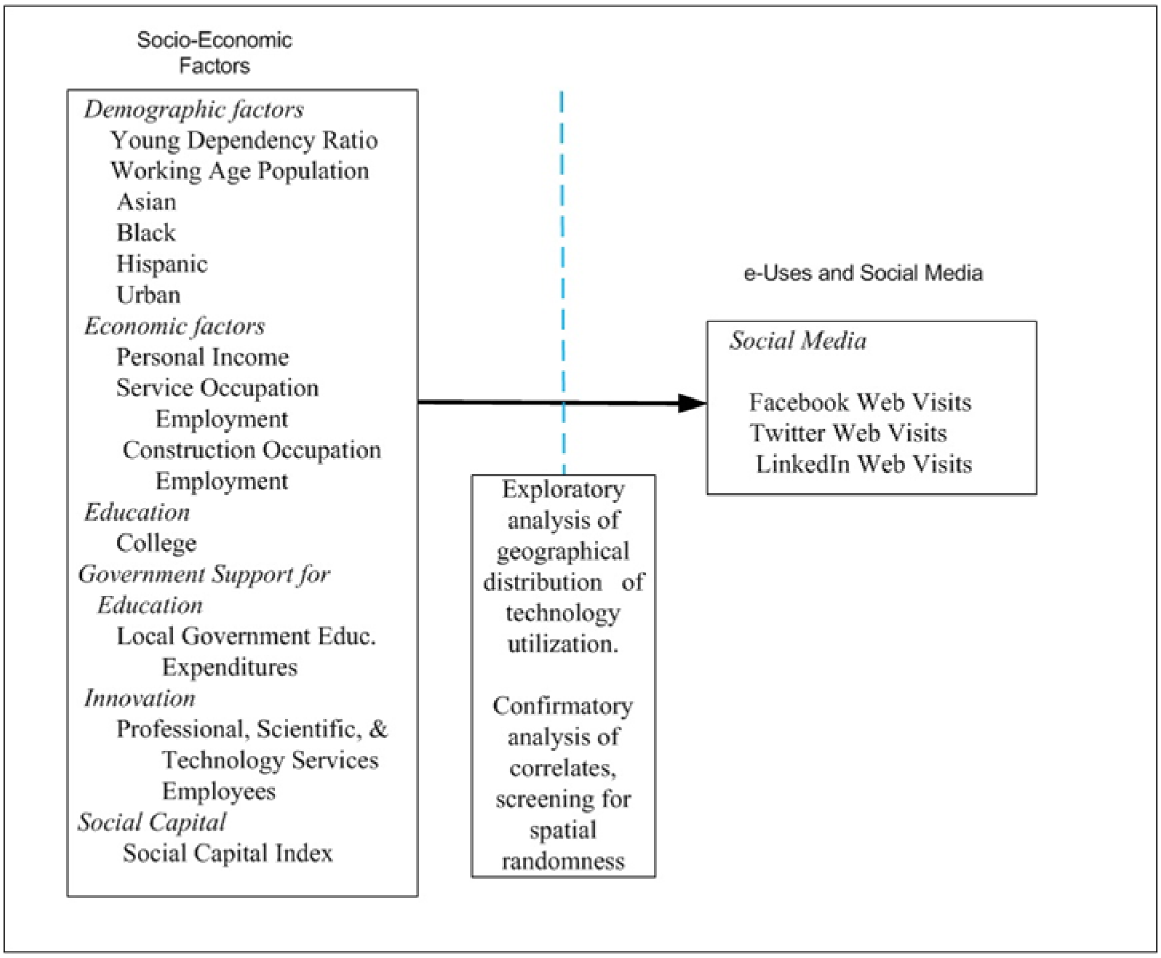

Figure 1. Conceptual Model for U.S. County Study

Demographic influences. Prior studies have documented an urban-rural differential for ICT utilization. NTIA provided evidence of such a differential for households using broadband in the home; rural households were found to trail their urban counterparts by 10 percent in 2010 (NTIA, 2011). It has been reasoned that more populous, wealthy urban regions in metropolitan areas with higher levels of educational attainment and higher median home values are more likely to have greater demand for broadband services. We posit that percent of urban population is associated social media utilization in US counties. Evidence of significant disparities in broadband usage at home by different racial/ethnic groups has also been documented; in 2010, Whites and Asians led Black and Hispanic households in broadband use by at least 18 percent (NTIA, 2011). Due to the disparity in internet access due to race and ethnicity we propose that race and ethnicity are associated with technology utilization in US counties; for Asians the association is posited as positive, and for Blacks and Hispanics the association is posited as negative.

Economic influences. Economic factors such as income, integration in the global economy, international trade openness, R&D, degree of economic openness, and GDP have been widely cited to influence ICT diffusion and utilization. Unemployed Americans, aged 16 years or older lag their employed counterparts in broadband usage by more than 10 percent in 2010 (NTIA, 2011). Income-based adoption disparity among individuals in the US for broadband has been documented: in households with annual income of $75,000 or more, 97 percent of individuals used the Internet, compared to 72 percent in households with $30,000 or less income (Perrin and Duggan, 2015). Hence, we posit that personal income per capita is associated with social media use in US counties.

Construction costs, specifically installation and maintenance costs were found to be associated with the probability of cell tower location for a spatial econometric study of cell phone coverage in sub-Saharan Africa. We argue that in the US, the construction sector will impact the physical and infrastructural aspects of ICT development. Therefore, we posit that services sector employment, and construction sector employment are associated with technology utilization in US counties.

Education influences. There is widespread evidence of the influence of education on ICT access and utilization. While adults with at least a Bachelor's degree are very likely to use broadband internet at home (almost 85 percent in 2010), the level of utilization drops rapidly to slightly more than 50 percent for individuals with a high school diploma (NTIA, 2011). In 2013, 93 percent of college graduates used the Internet, versus 50 percent for less than high school education (Perrin and Duggan, 2015). Other education correlates such as level of educational attainment, enrollments in tertiary education, and expenditures per student have been found to be significantly associated with internet use, broadband access, and personal computer use. Accordingly, we posit that college graduation is associated social media use. Government spending on education has been found to be associated with PCs, ICT expenditure, and ICT infrastructure quality (Author, 2008). Therefore, we also posit that local government education expense is associated with social media use.

Innovation influences. Innovation has impacted ICT utilization in some prior research. It can lead to more appealing and productive technology devices and services that increase usage. For instance, as social media capabilities and ease of use have improved through innovation, more users have tended to adopt it. In one study of the uses of several types of technologies for a worldwide sample of nations, innovation measured by scientific articles per capita was the most important independent variable (Author, 2008). Its importance has also been shown for in Asian nations (Quibria et al. 2003). Professional, scientific, and technical services workforce has been found to be significantly associated with receipts and payroll in three technology sectors (Author, 2005). The rationale was such a services sector is comprised of scientists, engineers, medical and health professionals, university researchers, lawyers and attorneys, who are naturally inclined towards innovative technology and are likely to be consumers of ICTs for research and development purposes. Accordingly, we propose that professional, technical, and scientific services employees per capita, as a proxy for innovation, is associated with social media use.

Social capital influences. Social capital is the scope of ties and linkages in a population through physical and communication means and well as by organizations that foster human collaboration or by bonding between people who possess resources (Putnam, 2001; Rupasingha, Goetz, and Freshwater, 2006). It was a key factor in a study of the digital divide for a survey sample of individuals in the US (Chen, 2003), as well as in a two-stage nationwide study of the decisions by persons to go online, subject to peer influences (Agarwal, Animest, and Prasad, 2009). Thus, we propose that social capital is related to usage of social media.

The dependent variables are three social media usage attributes of number of Facebook, Twitter, and LinkedIn website visits over a 30 day period. Facebook and Twitter subscription have been included in previous digital divide studies (Pick, Sarkar, and Nishida, 2014; Pick Sarkar, and Johnson, 2015).

Study Design and Methodology

Data

Data on various dependent and independent variables for this study were collected from multiple sources. Data on independent variables such as population, ethnicity, employment, income, governmental support for education, and sales revenue generated by publishers were obtained from various Census sources such as the US Decennial Census 2010, surveys conducted by the US Census Bureau such as the American Community Survey (ACS) of 2012, Local Education Agency (School District) Finance Survey of the National Center for Education Statistics of 2009, and US Economic Census of 2007. Data on US county-level social capital come from Pennsylvania State University's Northeast Regional Center for Rural Development (Rupasingha and Goetz, 2008). While data for several independent variables are obtained for the period 2010-12, social capital data are from the year 2009 which is the latest year for which such data are available. For dependent variables, data on the three social media dependent variables, usage of Facebook, Twitter, and LinkedIn, were extracted from Esri's Business Analyst software (Esri, 2014). Esri computes estimates of these dependent variables by combining 2013 geodemographic segmentation data with Doublebase® 2012 data from GfK MRI. Doublebase® 2012 data consist of an integration of information from four consumer surveys. The relevant data collection and estimation methodology statement for dependent variable data extracted from Esri's Business Analyst can be found at Esri (2015).

Usage of such data for US digital divide studies is unprecedented in the literature. Since all dependent variables, except for three, were from the three-year period 2010-12, time simultaneity has been assured for them. The fact that data for three of the independent variables lags by 3-5 years is justified as follows. Professional, scientific, and technical services employees are from the Economic Census of 2007, which was the latest date at which the variables were available. The 2009 variables of local government educational expenditures and social capital derive from specialized sources that do not have recent release dates. We feel the one-year lag for the period 2010-2012 is acceptable for these two variables, since they are not known to change rapidly year-to-year. All variables were converted to per capita whenever possible; variable definitions, sources, and descriptive statistics (N = 3,109 counties) of the dependent and independent variables are in Table 1.

Abbreviation	***Name of Dependent Variable***	***Source***	***Year of Data***	***Definition***	Min	Max	Mean	SD
FACEBOOKR	Visited website in last 30 days: facebook.com	BA	2010-2012	Facebook per capita	0.20	0.69	0.40	0.06
LINKEDINR	Visited website in last 30 days: LinkedIn.com	BA	2010-2012	LinkedIn per capita	0.01	0.19	0.03	0.02
TWITTERR	Visited website in last 30 days: twitter.com	BA	2010-2012	Twitter per capita	0.02	0.14	0.04	0.02
Variable	***Name of Independent Variable***	***Source***	***Year of Data***	***Definition***				
YOUNGDEPR	Young Dependency Ratio	CENDEC10_DP01	2010	**POP0-19/POP20-64**	0.17	0.88	0.45	0.07
COLLEGER	College Graduates or Higher, Age 18+	CENACS10_DP02	2008-2012	Population Estimate of College Graduates or Higher, Age 18+	0.03	0.54	0.13	0.06
PINCPC12	Personal Income Per Capita	BEA_CA1-3	2012	Personal Income Per Capita	17264.00	116843.00	36523.76	9182.78
WKAGEPOP	Working Age Pop (Pop. 20-64)/Total Population	CENDEC10_DP01	NA	**POP20-64/TOTPOP**	0.46	0.74	0.58	0.03
SERVICER	Service Occupations	CENACS10_DP03	2008-2012	Persons 16+ in Service Occupations, per capita	0.03	0.21	0.10	0.02
CONSTRUCTR	Construction Industry, Employed Persons 16+	CENACS10_DP03	2008-2012	Construction Industry, Employed Persons 16+, per capita	0.00	0.16	0.04	0.01
ASIANR	Asian Population	CENDEC10_DP01	2010	Asian Population per capita	0.00	0.36	0.01	0.02
BLACKR	Black Population	CENDEC10_DP01	2010	Black Population per capita	0.00	0.86	0.10	0.15
HISPANICR	Hispanic/Latino Population	CENDEC10_DP01	2010	Hispanic/Latino Population per capita	0.00	0.96	0.08	0.13
URBAN	Urban Population	CENDEC10_DP01	2010	Urban Population per capita	0.00	100.00	41.48	31.44
LGOVEDEXPR	Local Govt Education Expenditures	NCES_F33	2009	Total Expenditures (see variable definition document for more information)	69.57	36664.34	1985.97	1144.67
PSTSVCEMPR	Professional, Scientific, and Tech Services Employees	CENECON07	2007	Number of Professional, Scientific, and Technical Services Employees	0.00	0.18	0.01	0.01
FACTORAVE	Transformed Social Capital	Rupasingha and Goetz, 2008	2009	US County Social Capital	0.00	0.47	0.17	0.04
	SOURCES	Abbrev	Detailed Reference					
	US Census Bureau, DEC 2010, Table DP-01	CENDEC10_DP01	US Census Bureau, Decennial Census 2010, Summary File 1					
	US Census Bureau, ACS 2012, Table DP-02	CENACS12_DP02	US Census Bureau, American Community Survey 2012, 5-year estimates, Table DP-02					
	US Census Bureau, ACS 2012, Table DP-03	CENACS12_DP03	US Census Bureau, American Community Survey 2012, 5-year estimates, Table DP-03					
	US Census Bureau, ACS 2012, Table DP-04	CENACS12_DP04	US Census Bureau, American Community Survey 2012, 5-year estimates, Table DP-04					
	US Census Bureau, ACS 2012, Table DP-05	CENACS12_DP05	US Census Bureau, American Community Survey 2012, 5-year estimates, Table DP-05					
	Federal Communications Commission, Form 477	FCC_477	Federal Communications Commission, Form 477, Local Telephone Competition and Broadband Deployment					
	Bureau of Economic Analysis, CA1-3	BEA_CA1-3	US Dept of Commerce, Bureau of Economic Analysis, CA1-3, Personal Income, Per Capita Income					
	Bureau of Economic Analysis, CA04	BEA_CA04	US Dept of Commerce, Bureau of Economic Analysis, CA04, Personal Income Summary					
	US Census Bureau, Economic Census 2007	CENECON07	US Census Bureau, Economic Census 2007					
	US Department of Commerce, NTIA, State Broadband Initiative, Analyze Table	NTIA_SBI_Analyze	US Dept of Commerce, National Telecommunications and Information Administration, State Broadband Initiative (CSV format December 31, 2012).					
	National Center for Education Statistics, F-33	NCES_F33	US Census Bureau, Governments Division, Local Education Agency (School District) Finance Survey (F-33), National Center for Education Statistics, Common Core of Data					
	esri Business Analyst Data	BA	esri Business Analyst Data, GfK MRI DoubleBase Survey 2012					
	Rupasingha and Goetz, 2008	---	Pennsylvania State University's Northeast Regional Center for Rural Development					

Table 1. Definitions and Descriptive Statistics of Dependent and Independent Variables

Methodology

The methodology comprised techniques of descriptive statistics, correlation analysis, mapping of variables using a geographic information systems (GIS), testing spatial autocorrelation for the social media dependent variables, and ordinary least squares regression (OLS). For descriptive statistics, mean and standard deviations were computed for dependent and independent variables, in order to gauge the averages and extent of variation for all U.S. counties (Field, 2009). Correlation analysis was applied in order to screen the independent variables for multi-collinearity (Field, 2009, Hair et al., 2010).

Geographic Information Systems (GIS) can be utilized as an exploratory method to spatially describe and understand patterns of distribution of variables (Longley et al., 2011). In this investigation, GIS mapping is done to observe outlying values for the variables, with focus on the social media dependent variables. Given our sample of 3,109 counties, map displays reveal key trends, without overloading the display.

A key question is this study is whether social media usage in US counties shows statistically significant patterns of agglomeration of high and low values, or is social media usage in US counties spatially randomly distributed. We diagnose spatial autocorrelation – a common problem plaguing many technology diffusion datasets using Moran's I test statistic. The Moran's I test is inferential; the null hypothesis is that the values of a variable are randomly distributed spatially. Its interpretation is done by the p value for statistical significance (if p is not significant, the variable is randomly distributed spatially). Further, if the Z score is positive, the values of a variable are more geographically agglomerated (high values located near high ones and low values near low ones). If it is negative, the spatial pattern resembles a "checkerboard" pattern, in which high values are surrounded by low ones and vice versa (Moran, 1950; Openshaw, 1984).

Following the mapping analysis, ordinary least squares (OLS) regressions were performed for each social media usage dependent variable, in stepwise order, allowing in only those of the independent variables with significance levels of equal or less than 0.05. As an additional test of multi-collinearity, the variance inflation factor (VIF) was computed for each independent variable. We utilized the common cut-off of 5 or greater for VIF to be of concern (Myers, 1990) and no multicollinearity problems were detected.

Three diagnostic tests were administered to ensure that regression assumptions were met. Joint Wald Statistic is a test of the joint significance of several coefficients of individual independent variables (Wald, 1943). The Koenker (BP) Statistic Test is a test for heteroscedasticity, i.e. the variance of the residuals is not constant (Lyon and Tsai, 1996). The Jarque-Bera Statistic is a goodness-of-fit test of whether sample data, in this case regression residuals, have skewness and kurtosis that correspond to a normal distribution (Jarque and Bera, 1980). Additionally, regression residuals were tested for the presence of spatial bias using Moran's I test statistic. Model relationships which result in spatially random errors are regarded as valid. If errors in the model fit are spatially autocorrelated, it implies that the geographic forces are exogenous to the conceptual model. In case Moran's I testing indicates that regression residuals are not spatially randomly distributed, regressions results have to be treated with caution.

Findings

The regression findings for the entire country indicate that the most important determinants for the three social media variables are college graduation, young dependency ratio, proportion working age population, and percent urban. As seen in Table 2, the most important determinant is college education. This is in concert with other studies of U.S. counties (Author, 2005), household surveys of the U.S. (Chen, 2013), sample of U.S. states (Author, 2015), and international samples of nations (Baliamoune-Lutz, 2003; Yates et al., 2011). The mechanism may be that college educated people tend to be more conversant with social media through greater exposure to ICTs and the web during their years of education. That LinkedIn has the strongest education effect may be due to the enhanced presence of college educated users of LinkedIn compared to Facebook and Twitter.

	(Lower 48) Country			Metropolitan		
Independent Variable	FACEBOOK	TWITTER	LINKEDIN	FACEBOOK	TWITTER	LINKEDIN
Young Dependency Ratio	0.352***	0.357***	0.195***	0.493***	0.331***	0.225***
College Graduates or Higher, Age 18+	0.443***	0.353***	0.591***	0.627***	0.575***	0.761***
Personal Income Per Capita						
Working Age Pop (Pop. 20-64)/Total Population	0.350***	0.348***	0.267***	0.481***	0.414***	0.232***
Persons 16+ in Service Occupations	0.087***	0.092***		0.168***		
Persons 16+ in Construction Occupations					-0.209***	
Asian		0.135***	0.169***			0.173***
Black	-0.261***	0.181***		-0.178***	0.169***	
Hispanic	-0.249***			-0.154***	0.176***	
Urban	0.365***	0.272***	0.183***	N/A	N/A	N/A
Local Govt Education Expenditures						
Professional, Scientific, and Tech Services Employees						
Social Capital						
# of vars entered	7	7	5	6	6	4
Adjusted R-squared	0.760***	0.741***	0.820***	0.742***	0.755***	0.885***
Sample Size	3109	3109	3109	1161	1161	1161
OLS Regression Tests						
Joint Wald Statistic	7592.998***	6197.461***	7132.232***	1938.830***	2916.017***	5654.608***
Koenker (BP)	268.437***	228.120***	362.019***	33.406***	28.579***	90.485***
Jarque-Bera	616.561***	15713.923***	18055.782***	260.933***	1860.032***	121.843***
Spatial autocorrelation of dependent variable						
Moran's I	0.421***	0.363***	0.674***	0.434***	0.709***	1.168***
Spatial autocorrelation of regression residuals						
Moran's I	0.433***	0.030	0.107	0.286***	0.163**	0.252***
	* Signif. at 0.05, ** signif. at 0.01, *** signif at 0.001.					

Table 2. OLS Regression Findings for Socio-Economic Determinants of Social Media Variables, 2010-2012, Country and Metropolitan Samples

Demographic influences on ICT and social media have been reported in studies of Japan (Author, 2014) and the U.S. (Author, 2015). In the present study, both young dependency ratio and urban are positive; while for Japan, farm population is associated with reduced ICT usage including Facebook and Twitter, which implies that percent urban would increase them. On the other hand, young dependency ratio reduced some ICT variables and Twitter. This somewhat surprising finding is ascribed to young families being located in Japan mostly in rural areas, which would tend to have lower ICT and social media use.

Working age population has a strong positive correlation for all three social media indicators. This finding differs from the unimportance of proportion employed civilian workforce in a regression study of the U.S. states (Author, 2015). Service occupation is significantly associated with Facebook and Twitter use, although less strongly that for overall workforce. The findings on service occupations correspond to the importance of professional and service occupations for payroll and receipts in most technology sectors for U.S. counties in 1997-2000 (Author, 2005) and to the significant relationship of employment in services to technology level for 164 European Union sub-national regions in 27 European nations (Vicente and Lopez, 2011).

For the U.S. county sample, urban location is related to usage for all three social media variables, although its strength of association is highest for Facebook, followed by Twitter and LinkedIn. This may be the result of Facebook, versus Twitter and LinkedIn, having more areas of low use in rural regions including Appalachia, the lower South, and the southeast border area of Texas. The finding contrasts with lack of urban correlation for Facebook and Twitter for the U.S. states in 2010 (Author, 2015). However, it is similar to the finding for Japanese prefectures of an inverse relationship of proportion farm population with Facebook and Twitter users per capita in 2009-2010 (Author, 2015). The lack of effect for U.S. states might be due to the larger unit of analysis, i.e. state versus county, not being fine-grained enough to register the differences in and around the numerous American metropolitan areas. However, the finding corresponds to a survey study for the U.S. in 2010, in which 78 percent of urban adults used the Internet, versus 70 percent for rural dwellers (Perrin and Duggan, 2015, page 9).

The findings for influences of ethnicities varied among the three social media variables. For Facebook there is inverse association with Blacks and Hispanics, while for Twitter there is association with Asian and Blacks, and for LinkedIn association is present with Asian. We reason that Asian has generally opposite effect from Blacks and Hispanics, with Twitter being an exception for Blacks. This is based on other studies with similar findings (Perrin and Duggan, 2015; Author, 2015). For American adults in 2010, percent that used the Internet varied from 90 percent for Asians to 71 percent for Hispanics and 68 percent for Blacks (Perrin and Duggan, 2015, page 7). For U.S. states in 2009-2010, findings for contemporary technologies indicate positive association of Asians to a variety of ICT variables, while Hispanics associations are inverse. This included a positive Asian association and inverse Hispanic association with Facebook users, although no effects were evident for Twitter users (Author, 2015).

The results for the metropolitan, micropolitan, and rural subsamples largely correspond to those for the nation as a whole (Tables 2 and 3). Accordingly, only major differences from country-wide findings are noted here for the subsamples.

	Micropolitan			Rural		
Independent Variable	FACEBOOK	TWITTER	LINKEDIN	FACEBOOK	TWITTER	LINKEDIN
Young Dependency Ratio	0.545***	0.584***	0.338***	0.318***	0.619***	0.180***
College Graduates or Higher, Age 18+	0.385***	0.516***	0.725***	0.355***	0.187***	0.355***
Personal Income Per Capita						
Working Age Pop (Pop. 20-64)/Total Population	0.457***	0.539***	0.435***	0.202***	0.414***	0.222***
Persons 16+ in Service Occupations	0.175***	0.198***	0.160***	0.121***	0.161***	0.261***
Persons 16+ in Construction Occupations		-0.177***	-0.108***			0.089***
Asian	0.217***	0.130***		0.234***	0.125***	0.172***
Black	-0.291***	0.251***		-0.397***	0.258***	
Hispanic	-0.343***			-0.244***		
Urban	N/A	N/A	N/A	N/A	N/A	N/A
Local Govt Education Expenditures	-0.125***					-0.118***
Professional, Scientific, and Tech Services Employees			0.058**			0.086***
Social Capital		-0.087**	-0.111***			-0.152***
# of vars entered	8	8	7	7	6	9
Adjusted R-squared	0.694***	0.697***	0.771***	0.505***	0.361***	0.372***
Sample Size	637	637	637	1311	1311	1311
OLS Regression Tests						
Joint Wald Statistic	894.903***	779.111***	424.397***	1004.646***	213.198***	149.722***
Koenker (BP)	72.313***	112.968***	127.845***	244.571***	190.195***	256.122***
Jarque-Bera	116.315***	420.707***	1504.491***	89.059***	18472.807***	70571.308***
Spatial autocorrelation of dependent variable						
Moran's I	0.352	-0.178	-0.022	0.335***	-0.160*	0.105
Spatial autocorrelation of regression residuals						
Moran's I	0.853	0.128	0.770	0.077	-0.269***	-0.020
	* Signif. at 0.05, ** signif. at 0.01, *** signif at 0.001.					

Table 3. OLS Regression Findings for Socio-Economic Determinants of Social Media Variables, 2010-2012, Micropolitan and Rural Samples

Metropolitan subsample. The main difference from the national results is an inverse relationship of construction occupations with Twitter Use. This finding was not noted at the state level in 2009-2010 for Twitter or Facebook (Author, 2015). A possible indirect explanation is that areas having a high proportion of construction workers tend to be less educated and of lower income, which are known to decrease social media use. The direct effect related to the impacts of construction occupations is unexplained.

Micropolitan subsample. For the micropolitan subsample, there is a stronger association for the service occupations on all three social media variables than for the country or metropolitan samples, although correspondence to rural sample results. It may be that in the smaller cities and towns of micropolitan or rural America, there are proportionately more service workers, so with more presence, their influence on social media is greater. Another explanation is that social media may be a more important form of services communication between service workers and spread-out or even isolated service customers.

Social capital unexpectedly reduces Twitter and LinkedIn use. This is contrary to prior reported positive effects, at the U.S. state level, of social capital on desktop, internet, and positive effect broadband use and of the social-capital proxy variable of immigrant population on Facebook and Twitter (Author, 2015). We reason the influences for LinkedIn and Twitter are attributed to a substitution effect, in which those counties having strong social capital have lower average need for citizens to access LinkedIn, since those with a strong physical social network, usually mostly local or regional in extent, have less need for the worldwide professional networking of LinkedIn or social networking of Twitter. Facebook may be more valuable to users than Twitter or LinkedIn on a local basis, since it can contribute more to supplementing (rather than substituting for) existing physical networking.

The positive effect of professional/scientific/technical workforce, limited to only LinkedIn, is due to LinkedIn's market emphasis on business and professional people. The inverse effect on Facebook of local government education expenditures is unexplained.

Rural subsample. This subsample's inverse association of social capital with LinkedIn corresponds to the explanation just given for micropolitan. Likewise, there is the positive effect of professional/scientific/technical workforce on social media corresponds to the nation. The inverse effect on LinkedIn of local government education expenditures is unexplained.

Spatial Findings

In this section, the descriptive maps for the three social media variables are examined and interpreted, as well as the spatial autocorrelation findings. Since the maps show 3,109 counties, it is beyond the scope of this paper to examine their many hundreds of descriptive features, relationships, and differences, and accordingly this paper points only to the most prominent features nationally.

For Facebook (Figure 2), the findings show considerable range of percentage Facebook use, ranging from low use (20-35 percent) to high use (45 to 69 percent). Low levels are most evident in the southwest Texas border area, rural parts of the Great Plains stretching north-south from the Dakotas to Texas, the mid to lower-central South and in Appalachia and rural, inland parts of the Carolinas and Georgia. High areas of use are seen in the Boston to Washington megalopolis (see cutout on Figure 2), Atlanta, Chicago-Milwaukee, Minneapolis, Denver and Salt Lake City metropolitan areas, Seattle and Portland metro areas, San Francisco-San Jose, and parts of Southern California (i.e. San Luis Obispo, Santa Barbara, Orange, and San Diego Counties), but excluding Los Angeles County. This finer geographic detail was missing in a prior study's state map of Facebook users (Author, Figure 2), which for instance did not show the prominent features evident at the county level for Boston-Washington megalopolis, southwest Texas or Appalachia. Overall, these findings reflect the positive influence of location in large and creative metropolitan areas and reducing effect of poor and remote rural counties. The high Facebook levels in Denver, southern Wyoming, Salt Lake City, and surrounding Rocky Mountain areas have not been previously reported.

The concept of megapolis originated through Jean Gottman's study of the massive combined metropolitan areas of the Northeast coastal region stretching from Boston to Washington (Gottman, 1961). In 2000, the Northeast megalopolis was estimated to have 49.6 million population or a sixth of the U.S. population (Short, 2007). The high Facebook adoption by a half to two thirds of the population in this vast, dense megalopolitan region magnifies the social media impact, since the tens of millions of users are physically within several hours drive time of each other. This region also reflects the key positive determinants of Facebook use that were already discussed including college education, urban, working age population, and presence of significant Asian population in its large cities.

The spatial distribution of Twitter users (Figure 3) corresponds generally to the pattern to Facebook, with however the following differences: (1) the levels of Twitter use on the southwest border of Texas are much reduced compared to Facebook use, (2) usage in Southern California is substantially higher and includes all the counties in metropolitan Los Angeles and San Diego, (3) Denver and surrounding Rocky Mountain region is high in usage, but not as extremely high as Facebook, and (4) Appalachia and the mid-central South are moderate rather than very low for Facebook. The high Twitter use in the entertainment industry cluster of southern California might stem from Twitter's greater entertainment aspect than for Facebook or LinkedIn, while Twitter's relatively higher use in the lower-central South and Appalachia might be due to Twitter's simplicity and low storage needs, which are more suitable to the less affluent and educated population in those areas. LinkedIn's spatial distribution (Figure 4) resembles Twitter's, but has a north-south band of low use that extends from North Dakota to the Texas Panhandle and is unexplained.

The spatial autocorrelation analysis of the dependent variables (see Tables 2 and 3) reveals that results for Moran's I are highly significant for the country, metropolitan samples, inconsistent for the rural samples and not significant for the micropolitan sample. This reflects that the social media use is highly agglomerated in its metropolitan portion, which also influences the agglomeration level of the country as a whole, while micropolitan counties are less influential on their neighbors, resulting in lack of agglomeration of counties. For rural areas, agglomeration is significant for Facebook, for unknown reasons, while Twitter has significant inverse Moran's I value, implying that there is a trend towards lack of spatial autocorrelation, while there is a random spatial distribution for LinkedIn.

The spatial analysis went further and computed the extent of randomness of spatial autocorrelation, as measured by Moran's I, for the residuals of the OLS regressions. The results indicate a low spatial autocorrelation, compared to the high spatial autocorrelations values of the original variable, for the country and metropolitan samples, while the residual spatial autocorrelation level for the Micropolitan and Rural samples is random, except for an inverse Moran's I value for Twitter. Hence, in general, the OLS regressions have either reduced substantially or eliminated the spatial autocorrelation present in the original dependent variable, which means that the model and its independent variables are able to account quite fully for the very high autocorrelation for the observed dependent variables. Nevertheless some agglomeration remains, in particular for Facebook for the country sample and for all three social media variables in the metropolitan subsample, so those findings must be viewed cautiously. Figure 5 is a map of the standardized residuals for Facebook at the country level, and it is evident that the Facebook spatial autocorrelation persists in the residuals and is similar in pattern to the spatial autocorrelation pattern the Facebook raw variable (i.e. compare Figure 5 to Figure 2).

Discussion

This study has confirmed some well-known determinants from the digital divide literature. In particular among model variables, very prominent ones confirmed to have influence on social media, are age structure (Author, 2015), college graduation (Baliamoune-Lutz, 2003; Author, 2005; Author, 2008; Vicente and Lopez, 2011; Yates, Gulati, and Weiss, 2011; Chen, 2013), urban location (Arai and Naganuma, 2010; Fong, 2009, Chen, 2013), and ethnicity (Author, 2005; Perrin and Duggan, 2015; Florida, 2012). Among these are a pair of variables, urban location, and college graduation, which have commonly been closely correlated (Chen, 2013). Although missing from this research due to multi-collinearity, per capita income is often also correlated with this pair, so that it should be considered for inclusion in future research, perhaps to replace one or two in the triad.

Although professional, scientific, technical service occupation had a strong influence on ICT in a prior U.S. county study (Author, 2005), it was unimportant in the findings, except for LinkedIn for the micropolitan and rural subsamples. The difference is ascribed to a shift in the dependent variables, which, for the earlier study, were revenues and business receipts and payroll for the IS-Data Processing industry, broadcasting-telecommunication industry, and motion picture-sound industry. While for those complex sectors, the scientific/professional workforce would clearly have an impact, that impact is much less likely for the present countywide consumers of social media, who tend to be younger, educated individuals, but not necessarily technically inclined. Accordingly, for social media use, county policymakers need not concentrate on workforce development, by they but it should do to boost productivity in complex, digital industries.

Social capital is an unexpectedly weak determinant country-wide. For the subsamples, its effect was slight and only present as an inverse determinants in the micropolitan and rural samples for Twitter and LinkedIn. We ascribe the inverse findings to the substitution effect for physical social capital that was discussed earlier.

Construction workforce had a slight and mostly opposite, inverse effect. This is unexplained and points to further research to try to determine the mechanism.

The Spatially Aware Technology Utilization Model (SATUM) is applicable to this investigation. Nearly all the independent variables have at least some empirical association with the social media. The only exception is personal income, which was eliminated in the present study due to multi-collinearity. The present SATUM model has the potential to be applied to small geographic samples of counties, with sample sizes as small as 50. For example, two areas identified descriptively to have high usage of social media, western coastal counties stretching from Washington State to southern California and the Northeast megalopolis, might be studied as subsamples to determine their determinants for social media. Likewise, the social media determinants in the Great Plains and middle South counties could be analyzed by regression to determine their determinants for low social media usage.

For large samples, with varied environments, geographically weighted regression constitutes an alternative way to aggregate results over distinctive geographies. However, for the full set of U.S. counties, it is problematical due to the irregularity of the county polygons across the nation, although it could be applied in sub-national regions such as Midwest with more consistently shaped boundary shapes.

Spatial autocorrelation in the present investigation is very useful in assessing extent of agglomeration for raw variables and for regression residuals. For such a large sample, it has the advantage to summarize systematically the extent of agglomeration, which the eyeball would miss.

Given the very large sample sizes, more complex models could be constructed with Structural Equation Modeling (SEM), path analysis, or econometrics. For SEM, the set of independent variables could be enlarged by selecting additional ones from the U.S. Population and Economic Censuses, American Community Survey using 5-year range of data, or other robust county samples. Although a theoretical model has already been tested by SEM for a worldwide sample of nations (Author, 2011), a more complex factor model could be formulated for U.S. counties, and relevant theory could be applied from that study. However, it is likely other relationships would need to be induced. A challenge would be to take geography into account, in applying SEM.

Policy Implications and Limitations

County governments can set ICT policies, as can metropolitan governments which are defined as area based on one or more county (U.S. Census, 2015). These governments can influence ICT policy, by providing their own public Internet services, fostering or supporting ICT training for citizenry, encouraging the hiring of local ICT graduates (Kvasny and Keil, 2006), encouraging and helping with incentives for service industries make greater use of IT including by small businesses. Regarding IT training, the governments need citizens to go beyond just completing courses or certificates, but to leverage the training for the “next step,” which might be further education or a closely-related job (Kvasny and Keil, 2006).

It is more difficult to develop polices based in demographic determinants of social media, since the population processes are not under the control of the county government. Nonetheless, the counties could focus their training options on demographic categories associated with more ICT usage, such as young, urban, affluent people. However, at the same time, social equity would push county policymakers in the opposite direction, to favor training for technologically underprivileged age groups in rural areas.

Although our study suggests that county policies not emphasize formation of social capital, an earlier study at the state level indicated strong impact from social capital, and it pointed to the need for state policymakers to emphasize it (Pick, Sarkar, and Johnson, 2015).

Limitations. One limitation is that the dependent variables were collected in large-scale industry surveys, which are subject to a survey sampling error. However, when a newer survey becomes available, longitudinal comparisons can be made for error checking.

This also points to the limitation that the study is cross-sectional, so cannot recognize the magnitude changes of variables and varying influences over time. However, this limitation can be overcome by a future study that includes the present data for 2010-2012, along with data for a later period.

A limitation already mentioned is that the regression analysis does not model complex and/or bi-directional relationships. This can be addressed by using SEM, path analysis, or other techniques for analyzing complex models. A challenge for such studies will be to formulate a more complex theoretical model than SATUM, which could be derived from limited prior complex models, induction, and reasoning. Another methodology that could be applied to parts of the nation with fairly consistently shaped and sized polygons is geographically-weighted regression (GWR). This is a not a present limitation because the entire nation has polygons which are frequently irregularly sized and shaped, so GWR could not be applied.

The study is limited by use of data that are from 2010-2012. More recent data on the dependent variables is available; however, many of the dependent variables are not updated, so it might be several years before a data set with good time simultaneity will be available.

Finally a limitation is that case studies of counties are not available that could illustrate the SATUM theory and support or contradict the present empirical findings. If this need is met by developing case studies, interview questions would need to be designed to add higher levels of understanding of geographical aspects of county technology use.

Conclusion

This research has developed a conceptual model in which socio-economic variables influence social media use, with accompanying spatial analysis. The model is tested empirically by OLS regression, revealing that the key correlates of social media use are age structure, college education, young dependency ratio, working age population, and percent urban. The sample of U.S. counties is divided into subsamples of metropolitan, micropolitan, and rural counties, and analysis of those OLS findings allow comparisons to be made with the country-wide sample.

The spatial analysis consists of descriptive mapping and spatial autocorrelation analysis both of the dependent variables and of regression residuals. This latter analysis reveals the U.S. to be highly agglomerated in social media for counties. At the extreme of high valued agglomerated counties is the Northeast megalopolis and coastal southern California stretching up to San Francisco and eventually to Seattle. There also is an area of high social media use that centers on the Rocky Mountain region from Denver to Salt Lake City. Area of low social media use include the Great Plains, Appalachia, and the mid lower South.

Policy implications of the research are applied to formulation of county policies. Examples would be to encourage education, innovation, and training for service workers including for small businesses. This research points to further projects that could illuminate the county social media patterns, the forces that are related to social media use, and the geographic texture of usage in America.

References

Agarwal, R., Animesh, A., and Prasad, K. 2009. Social interactions and the digital divide: explaining variations in Internet use. *Information Systems Research* (20:2),pp. 277-294.

Arai,Y, and Naganuma, S. 2010. The geographical digital divide in broadband access and governmental policies in Japan: three case studies. *Networks and Communication Studies, NETCOM*, (24:1-2),pp.7-26.

Azari, R., and Pick, J.B. 2005. Socioeconomic influences on technological sectors for United States counties. *International Journal of Information Management* (25:1),pp. 25-37.

Baliamoune-Lutz, M. 2003. An analysis of the determinants and effects of ICT diffusion in developing countries. *Information Technology for Development* (10:3),pp.151-169.

Chen, W. 2013. The implications of social capital for the digital divides in America. *The Information Society* (29),pp.13-25.

Gottman, J. 1961. *Megalopolis: The Urbanized Northeastern Seaboard of the United States*. New York: The Twentieth Century Fund.

Esri 2014. Business Analyst. Retrieved from http://www.esri.com/software/businessanalyst, May 1.

Esri 2015. 2013 Esri US market potential data methodology statement. Retrieved from http://support.esri.com/en/knowledgebase/whitepapers/view/productid/153/metaid/2037.

Everitt, B.S., Landau, S., Leese, M., and Stahl, D. 2011. *Cluster Analysis*. 5th Edition. NewYork: John Wiley and Sons.

Farkas, Dan, Brian Hilton, James Pick, Hindupur Ramakrishna, Avijit Sarkar, and Namchul Shin. 2015. A Tutorial on Geographic information Systems: A Ten-year Update. *Communications of the Association for Information Systems*, in press.

Florida, Richard. 2012. *The Rise of the Created Class*. Revised Edition. New York: Basic Books.

Field, A. (2009). *Discovering Statistics Using SPSS*. 3rd Ed. Thousand Oaks, California: SAGE Publications Ltd.

Fong, M.W.I. 2009. Digital divide between urban and rural areas in China. *Electronic Journal of Information Systems in Developing Countries* (36:6),pp.1-12.

Griffith, D.A. 1987. *Spatial autocorrelation: a primer*. Washington, D.C.: Association of American Geographers.

Grubesic, T.H. 2004. The geodemographic correlates of broadband access and availability in the United States. *Telematics and Informatics* (21),pp.335-258.

Grubesic, T.H. 2006. A spatial taxonomy of broadband regions in the United States. *Information Economics and Policy* (18),pp.423-448.

Hair, J.F. Jr., Black, W.C., Babin, B.J., and Anderson, R.E. (2010). *Multivariate Data Analysis*. Upper Saddle River, NJ: Prentice Hall.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity, and serial independence of regression residuals. *Economics Letters, (6*:3),pp. 255-259.

Khatiwada, L.K., and Pigg, K.E. 2010. Internet service provision in the U.S. counties: is spatial pattern a function of demand? *American Behavioral Scientist* (53:9),pp.1326-1343.

Kvasny, L., and Keil, M. 2006. The challenges of redressing the digital divide: a tale of two U.S. Cities. *Information Systems Journal* (16:1),pp.22-53.

Longley P.A., Goodchild, M.F., Maguire, D.J., & Rhind, D.W. (2011). *Geographic Information Systems & Science*. 3rd Edition. New York, NY: John Wiley and Sons.

Lyon, J. D., & Tsai, C-L. (1996). A comparison of tests for heteroscedasticity. *The Statistician, (45*:3), pp. 337-349.

Malecki, E.J. 2003. Digital development in rural areas: potentials and pitfalls. *Journal of Rural Studies* (19),pp.201-214.

Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* (37:1),pp.17-23.

Myers, R. (1990). *Classical and Modern Regression with Applications*. 2nd Edition. Boston, MA, Duxbury.

Nishida, T., Pick, J.B., and Sarkar, A. 2014. Broadband Utilization in the Indian States: Socio-Economic Correlates and Geographic Aspects, in Jyoti Choudrie and Catherine Middleton (eds.), *Management of Broadband Technology Innovation*, Oxford, England: Routledge, pp. 269-296.

NTIA.2011. *Exploring the digital nation: Computer and internet usage at home*. Washington, DC: National Telecommunications and Information Administration, U.S. Department of Commerce.

OECD. 2011. *Understanding the digital divide*. Paris, France: Organization for Economic Co-operation and Development.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Geobooks, Norwich, England.

Perrin, A., and Duggan, M. 2015. Americans' Internet access: 2000-2015. Report in Series on Numbers, Facts and Trends Shaping the World. Washington, D.C.: Pew Research Center.

Pick, J.B., and Azari, R. 2008. Global digital divide: Influence of socioeconomic, governmental, and accessibility factors on information technology. Information Technology for Development (14:2),pp.91-115.

Pick, J.B., and Azari, R. 2011. Global model of utilization of technology based on governmental, social, economic, and business investment factors. Journal of Management Information Systems (28:1),pp.51-85.

Pick, James B., Avijit Sarkar, and Jeremy Johnson. 2015. United States Digital Divide: State Level Analysis of Spatial Clustering and Multivariate Determinants of ICT Utilization. *Socio-Economic Planning Sciences*, (49),pp.16-32.

Pick, James B., and Avijit Sarkar. 2015. The Global Digital Divides: Explaining Change. Heidelberg: Springer.

Perrin, A., and Duggan, M. 2015. Americans' Internet access: 2000-2015. Research Report. Washington, D.C.: Pew Research Center.

Putnam, R.D. 2001. Social capital. Measurement and consequences. *Canadian Journal of Policy Research* 2:41-51.

Putnam, R.D. 2007. E Pluribus Unum: diversity and community in the twenty-first century. The 2006 Johan Skytte Prize lecture. *Scandinavian Political Studies* (30:2),pp.137-174.

Quibria, M.G., Ahmed, S.N., Tschang, T., and Reyes-Macasaquit, M.-L. 2003. Digital divide: determinants and policies with special reference to Asia. *Journal of Asian Economics* (13),pp.188-825.

Rupasingha, A., Goetz, S. J., and Freshwater, D. 2006. The production of social capital in U.S. counties. *The Journal of Socio-Economics* 35:83-101.

Sarkar, Avijit, James B. Pick, and Jessica Rosales. 2016. ICT-enabled E-Entertainment Services in United States Counties: Socio-economic Determinants and Geographic Patterns. 2015. Manuscript in review for chapter in *Innovative ICT-enable Services and Social Inclusion*. J. Choudrie, S. Kurnia, and P. Tsatsou (Eds.), London: Routledge. In second review for publication.

Short, J.R. 2007. *Liquid City: Megalopolis and the Contemporary Northeast*. Washington, D.C.: Resources for the Future.

Takhteyev, Y., Gruzd, A., and Wellman, B. 2012. Geography of Twitter networks. *Social Networks* (34)pp.73-18.

Vicente, M.R., and Lopez, A.J. 2011. Assessing the regional divide across the European Union-27. *Telecommunications Policy* (13),pp.220-237.

Wald, A. (1943). Test of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society,* (54),pp.426-482.

Yates, D., Gulati, G.J., and Weiss, J.W. 2011. Different paths to broadband access: the impact of governance and polity on broadband diffusion in the developed and developing worlds. In *Proceedings of 44th Hawaiian International Conference on System Sciences*, IEEE.

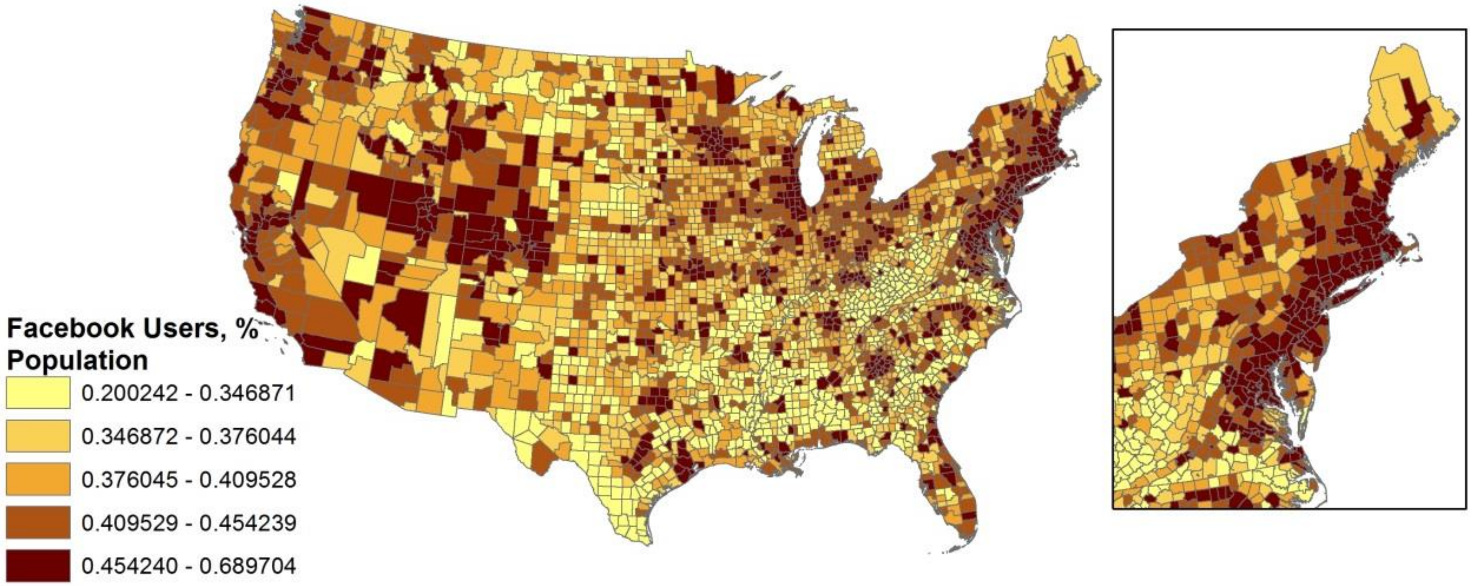

Figure 2. Distribution of Percent of Facebook Users, U.S. Counties, 2012

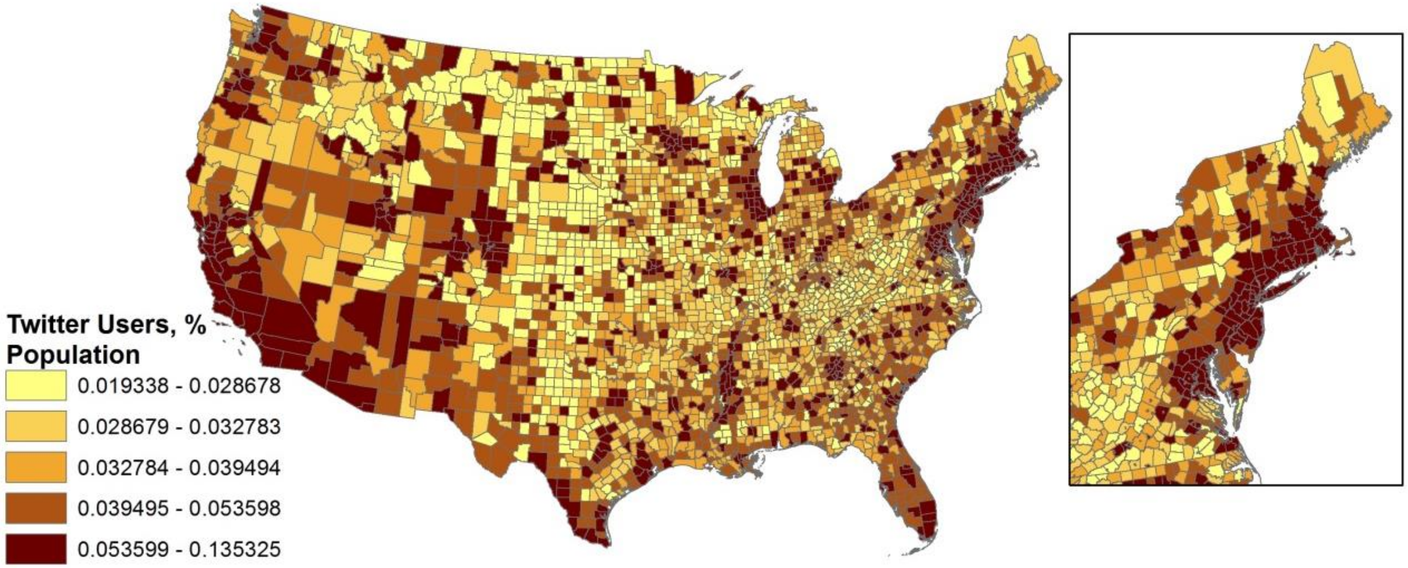

Figure 3. Distribution of Percent of Twitter Users, U.S. Counties, 2012

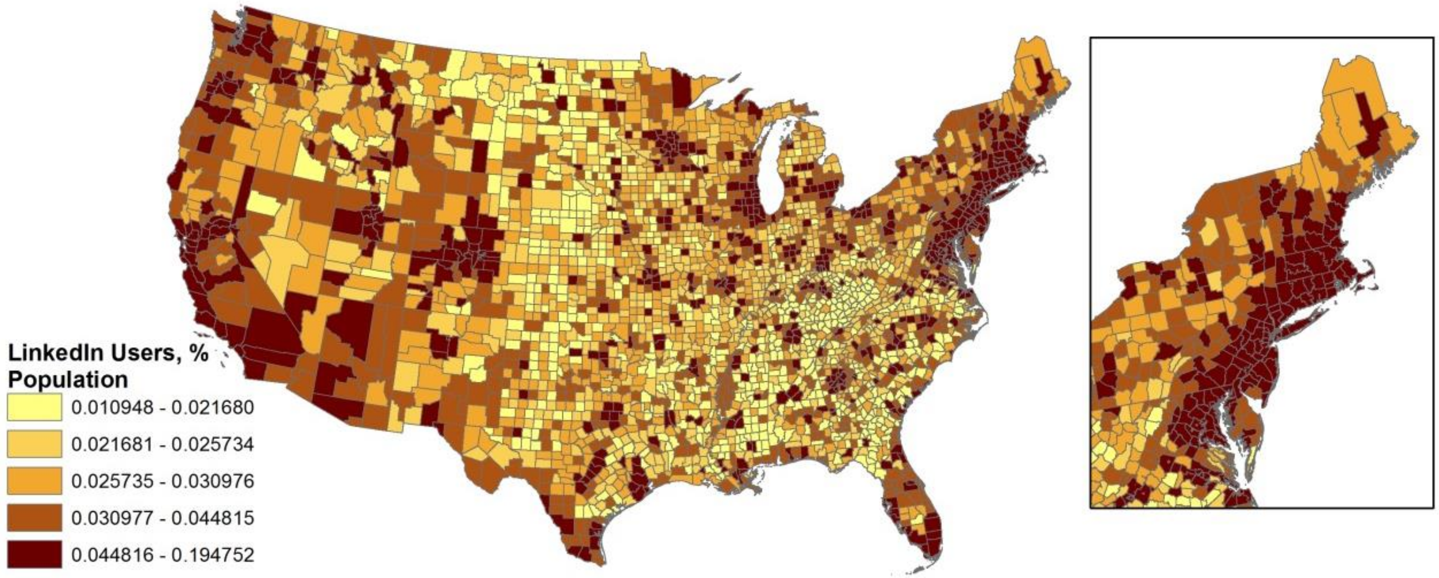

Figure 4. Distribution of Percent of LinkedIn Users, U.S. Counties, 2012

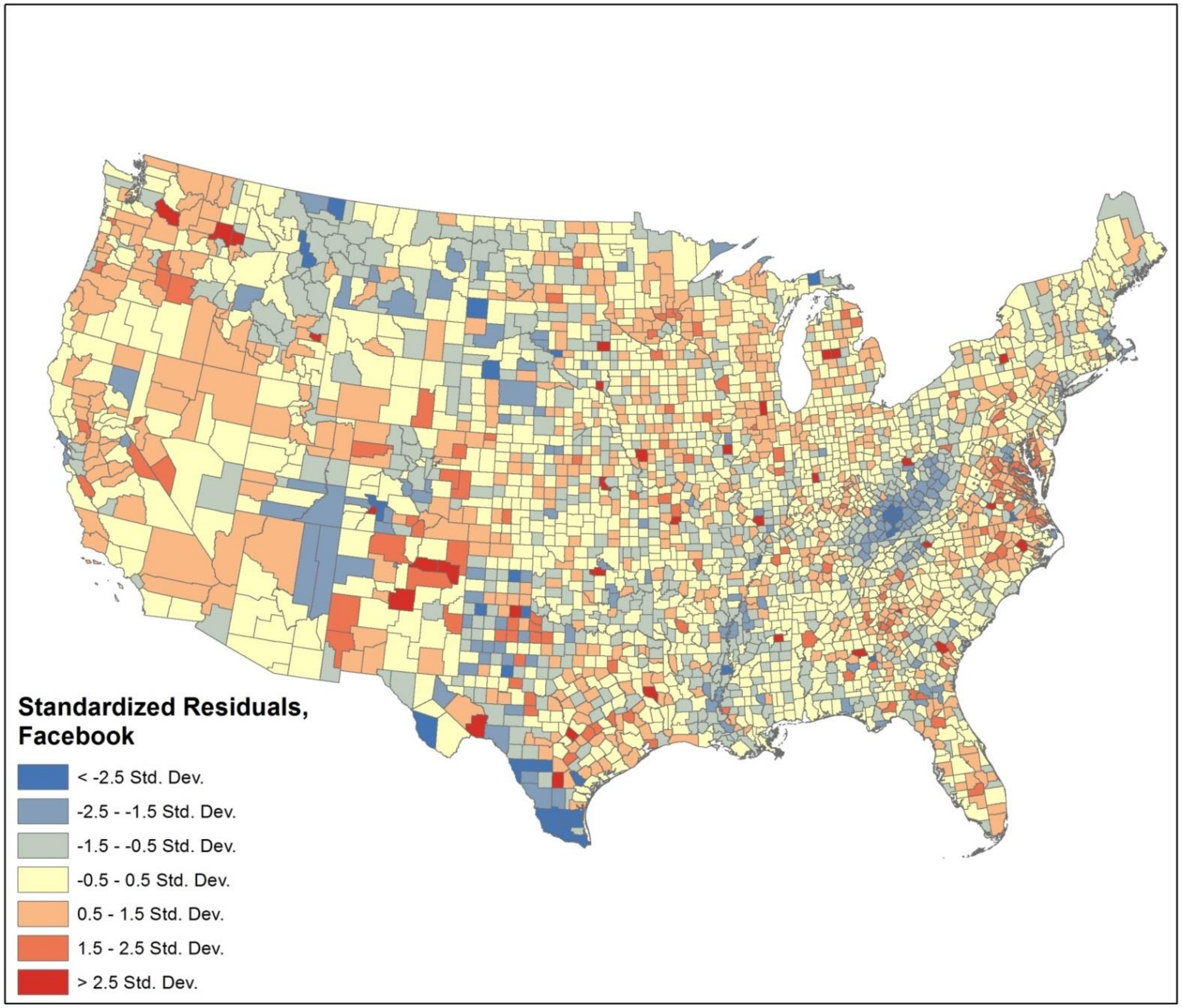

Figure 5. Distribution of OLS Regression Residuals (Standard Deviations) for Percent Facebook, U.S. Counties, 2012