Mapping the Distribution and Spread of Social Ties Over Time: A Case Study Using Facebook Friends

Clio Andris,1* Sara E. Cavallo,2 Elizabeth A. Dzwonczyk,3 Laura Clemente-Harding,2,4 Carolynne Hultquist2 and Marie Ozanne5

1School of City and Regional Planning and School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA.
2Department of Geography, The Pennsylvania State University, University Park, PA, USA.
3Department of Geography and Environmental Engineering, United States Military Academy, West Point, NY, USA.
4Geospatial Research Laboratory (GRL), Engineer Research and Development Center (ERDC), Alexandria, VA, USA.
5SC Johnson College of Business, The Hotel School, Cornell University, Ithaca, NY, USA.

*E-mail: clio@gatech.edu

Abstract

Relational geography asserts that social networks provide geographic benefits, and geographies are transmitted through the sharing of local knowledge and experience. To articulate the spatial expanse and geographic benefits of an individual’s social network, researchers require better social-spatial geographic information system models illustrating how contacts are dispersed, and how many distinct places they inhabit.

In this work, the authors conduct a case study to map social network ties in geographic space. The authors retrieve social network matrices for 20 volunteers (egos) via Facebook.com, amounting to over 8,500 friends (alters). Each ego listed the alter’s hometown city at two time periods: at relationship inception and at the time of the study. The authors measure specific tie locations, tie expanse, deviation from a gravity model prediction, and expansion of alter groups (family, clubs, neighbors, etc.) over time.

The authors find that social networks geographically spread over time, on average, from 2,679 km (standard distance) to 3,258 km (standard distance), and that the average ego had alters in 21 unique locations when they met, and 38 locations at the time of the study. Regarding friend groups, the authors discover that high school friends and friends from non-residential gatherings (ex. conferences) dispersed the most (over 1,900 km), and cultural groups (churches, sports teams) and family dispersed the least (less than 800 km) over time. Our results lead to a discussion of how mapping and measuring the distribution of social connections can uncover changing dynamics of social interaction, and one’s ability to access and engage with places through social ties.

Keywords

Spatial social networks, Geographic information systems, Friend mapping, Facebook, Social ties.

Individuals belong to geographically based social networks that include an individual’s ties (family, friends, professional contacts, etc.) in nearby and distant places (Acedo et al., 2017). A set of mapped ties comprises a spatial distribution, i.e., a unique “fingerprint” of geolocated social contacts (e.g., two friends in Rome, six family members in New York, a co-worker in Milan, etc.). This distribution is a natural part of social life, as humans have been “traveling, wandering and friending friends and kin seemingly forever” (Hampton and Wellman, 2003, p. 284).

While studies have estimated social network size (ex. Hill and Dunbar, 2003) and structure (examples abound), relatively less is known about the spatial distribution of an individual’s social ties. For example: in how many places (e.g., cities and towns) does the average person have a social tie? Does this number expand or contract for the individual over time? What is
the average distance at which an individual’s ties reside? In addition, longitudinal social network studies have examined contact strength, network benefits, individuals comprising the networks, and networks through life events such as marriage, divorce, retirement, and widowhood (Ertel et al., 2009). Yet, the shift in network geography over time is not as clear.

The answers to these questions can tell us more about an individual’s level of access to social and spatial resources, as having trusted nearby contacts or trusted contacts in particular locations-of-interest leads to the acquisition of a rich variety of resources. In short, it is not just where you live and what you can access nearby, but also what you can access through your social network ties. We situate these questions within a larger framework of the theory of relational geography, which affirms that social networks (and social capital) are spatialized, and that relationships are embedded within a larger spatial system that includes institutions and economic practices (Bathelt and Glückler, 2003). Relational geography theory also asserts that economic actors acquire distant resources across regions and countries in order to reap the benefits of innovation, diffusion, collaboration, collective learning processes, and knowledge generation (Bathelt and Glückler, 2005). Whether through new tie acquisition, inter-tie relationship changes, or changes in the spatial configuration of a network, time often changes the variety, diversity, and robustness of an individual’s personal Rolodex of people and places. As an individual moves, their spatial social network will add this new cost of visiting or maintaining communication, seek local alternatives for social life, or restructure in another way.

These life events and strategies to cope with the change of moving and changing geographic social networks are common, but the formal concepts are still abstract and challenging to measure. This serves as a motivation for quantitatively discretizing the spatial distribution of many individuals’ contacts. Mapping these ties over time can help codify complex structures in order to discern the cause and effect of social network life changes and to articulate commonalities of social network geographies across the population. Making this implicit, personalized knowledge more explicit and our analysis more procedural can show how humans may have a common experience. For instance, mapping contacts provides a way to find the typical individual’s balance of nearby and distant ties, and distinguishing this balance is important because nearby and distant ties offer different types of support. Nearby ties can provide help in emergency situations and yield increased power in local governance (Aldrich and Meyer, 2014) while distant friends are helpful when an agent plans a visit or seeks advice (Stafford, 2005).

To further motivate the mapping of geographic social networks, there are also practical travel and health benefits to enumerating contacts’ distances. Mapping tie locales can reveal how far people must travel to see their ties and to calculate the cost, or cost-prohibitive nature, of accessing social support (Carrasco et al., 2009). It can help measure the extent to which an individual is in jeopardy of isolation from known social ties, which can lead to loneliness and adverse health effects (Cacioppo and Patrick, 2008). Maps of contacts can also be used to estimate whether the average person is privy to information about many places (Dabbs et al., 1998; Golledge, 2002) and has opportunities to visit an array of cultural locales (Pultar and Raubal, 2009). Geographic theory supports this notion, through an understanding that contacts share their “insider” knowledge of places.

Engaging with the questions of geographic social network configuration is becoming increasingly facile with digital data sources. In particular, social media applications that allow individuals to track the geographic locations of their contacts provide instant information about the spatial expanse of a user’s contacts (where they move to, their vacation locations, etc.). Most commonly, social media acts as a repository in which users can look up locations from friends’ profiles (on platforms such as Facebook). However, more recently, the locations of large sets of friends has been mapped in real time by apps such as Snapchat’s Snap Maps, allowing users (most often teens and young adults) to peruse a changing heat map annotated with cartoon images of their friends (Juhász and Hochmair, 2018). This mass location sharing creates a “landscape” of many contacts mapped at one time (Næss, 2018). Users most often opt to share location with friends and family, as well as nearby contacts in order to ease the logistics of meeting up (Consolvo et al., 2005). The digital collection of these data encourages better studies not only on of social network size and structure, but also on the distribution of social tie locations.

Motivated by a conceptual systems approach of relational geography and enhanced offerings of online social network data, we perform a case study to measure the geographic distribution of social ties. The objective of this work is to take advantage of digital social network data, and present a proof-of-concept study that illustrates how geographic information systems (GIS) mapping and modeling can help analyze the underlying structure and organization of an individual’s relational geography. In our approach, we map the Facebook friends of a group of 20 volunteer egos. We use two locations for each alter: the location where the alter lived when they met the ego (pre-period)
and the alter’s location at the time of the study (post-period). We perform a descriptive analysis of this data set to answer the following research questions:

RQ1. In how many cities and countries do individuals have ties?
RQ2. Do individuals have ties in more locations and at farther distances over time?
RQ3. Do tie locations deviate from the expected distances at which an individual is likely to have ties?
RQ4. Do different types of alter groups (e.g., family, school friends, neighbors, etc.) tend to spread (disperse) over time?

This study illustrates our argument that an individual’s set of places they can access through ties will change over time, as will their ability to access geographic resources through ties. Instead of thinking about the individual’s geographic network as highly personalized and difficult to codify, the case study shows the technology needed to create a working model of geolocated social networks, what measurements and statistics are pertinent, and what visualizations are helpful. As a result, the network provides new insights about how social ties are spread across geographic space. This case study is a non-conventional approach to social network data analysis, as it combines the GIS analysis and measures of spatial network expanse at the individual level. By replicating the procedure described in our case study, others can implement their own networks and collect their own data for future studies.

In the following section, we review major concepts mobilized in this work, and briefly describe our case study approach. We next describe the Facebook data collection process, network measurements, GIS methods and summary statistics. We then describe the results of the 20 networks’ spatial distributions and their change over time. Finally, we contextualize our findings in a wider body of literature, discuss study limitations and conclude.

Background: concepts revisited in this work

This section explores previous research describing contacts’ expected and actual distribution over geographic space.

Estimating connectivity

The gravity model is a classic method for estimating a “baseline” of spatial interaction across locales. Interaction can represent number of phone calls, migrants, commuters, number of relationships or other flows that connect two locales. This model computes an expected value of interaction, $I_{ij}$, as the product of a source ($i$) and target ($j$) city’s respective populations, divided by the Euclidian distance between $i$ and $j$ raised to an exponent $\beta$ (called the coefficient of friction, most frequently parameterized as 2) (Dodd, 1950; Reilly, 1953) (Eq. (1)). Distance can also be redefined as travel time or another cost factor (de Smith, 2004). This value is multiplied by a constant ($K$) to better reflect the actual magnitudes of the interaction values.

$$I_{ij} = K \frac{\text{City}_i \times \text{City}_j}{\text{Distance}^{\beta}}$$

This equation has been used to estimate inter-city travel (Ben-Akiva and Lerman, 1985) and retail patronage (Huff, 1963) and has been reparametrized by adding demographic data such as income at a destination city (Greenwood, 1985). The gravity model provides an expectation from which to measure whether real-world inter-city connections deviate from this expected value. Recently, actual interaction data have been used to re-calibrate the coefficient of friction from the default value of two (Krings et al., 2009), and determine where sets of cities are over- or under-connecting in comparison to the gravity estimates (Dugundji et al., 2011; Takhteyev et al., 2012).

Next, survey data and large data sets describing tie distribution in geographic space have shown that distance affects the likelihood of relationships.

Relationships and distance

Distance plays a role in the likelihood of relationships and relationship maintenance. The likelihood of interaction (or friendship) tends to decrease with each increment of distance between individuals, a longstanding economic geography concept known as distance decay. Recently, large data sets from GPS traces, Location-based social networks (LBSNs), and call data records have confirmed that the probability of having a social tie in a certain location decreases exponentially with distance to that location (Blondel et al., 2008; Lambiotte et al., 2008; Leskovec and Horvitz, 2008; Liben-Nowell et al., 2005; Onnela et al., 2011; Preciado et al., 2012; Scellato et al., 2011), although rates can vary depending on data source (Spiro et al., 2016). Each different function reveals the likelihood of an individual having a contact at certain distances. For instance, in Belgium, this probability drops off significantly at 40km from the individual’s locale (Lambiotte et al., 2008). In studies of Twitter data,
it was shown that 34% of Twitter friends who follow one another live within 25 miles while only 18% of users who do not follow each other, but “mentioned” one another live within that radius (McGee et al., 2011), and that Twitter friendships are more likely to follow flight patterns than Euclidean distance (Takhteyev et al., 2012). Moreover, the average distance of an individual’s contacts has been shown to vary by hometown: a recent Facebook study showed that residents in isolated US areas (such as Eastern Kentucky) had up to 82% of Facebook friends living within 50 miles, whereas residents of the Western US had fewer than 43% of friends living within the same radius (Bailey et al., 2018).

In addition to big data harnessed from online social networks, studies using standard survey approaches confirm that the accessibility of social support is an important descriptor of a community. For example, in his study of the “spatial dimensions of personal relations”, Fischer (1982) finds that 26% of semi-rural residents’ relatives lived within a 5 minutes’ drive, but only 15% of urbanites’ relatives lived within a 5 minutes’ drive. The General Social Survey reports that 27% of respondents live within a 15-minute drive of their mother, and 12% live 12+ hours away (Smith et al. n.d.). Similar studies of long distance relationships find that 66% of elderly parents live within 30 minutes of an adult child (Lye, 1996; Stafford, 2005). These types of studies are particularly helpful because they distinguish different types of ties.

Survey data and distance decay functions provide rules-of-thumb for where relationship probability declines. Yet, they should be paired with map-based approaches that can communicate a richer portrait of relationships. Our case study aims to dig deeper into personal distributions for individuals by listing individual cities that egos interact with and describing how these dynamics change over time.

Case study, data and methods

Case study

In total, 20 volunteers downloaded their friendship networks (totaling 8,549 friends) from www.facebook.com. This platform was chosen because Facebook ties have been shown to mimic real-life acquaintances (Mayer and Puller, 2008) and because the platform has been widely used in human social behavior research in messaging patterns (Golder et al., 2007), social connections (Ellison et al., 2007), and cultural preferences (Gross and Acquisti, 2005; Lampe et al., 2006; Pempek et al., 2009).

Each volunteer (ego) annotated each of their friends (alters) with an accompanying home location at two time periods (the location from when the ego and alter met, and a current location). Individuals’ ties were mapped in geographic space and analyzed both individually and as a combined group. This study is equipped to respond to our four research questions because it specifically maps individuals’ social ties, calculates distance between ties, associates ties with different cities, and measures tie dispersion over time.

Participants

Volunteers were enrolled at The Pennsylvania State University (Penn State) as graduate or undergraduate students at the time of the study. All volunteers resided in State College, PA, a small college town in the northeastern USA, and were recruited to participate in the study through a seminar course project. Participants were required to have a Facebook account with associated “friends” (i.e., alters) and be at least 18 years old. The group ranged in age between 20–40 years old, and included 10 women and 10 men. It was comprised of 17 white and three Asian respondents, and four participants were non-native English speakers. Further information about egos was concealed for privacy reasons.

Social network acquisition

Within Facebook, each volunteer used the NameGenWeb application (as described in Hogan, 2011) to download their social network as a graph of their friends and friend inter-connections (i.e., if the ego’s friends are friends with one another). NameGenWeb has been used to show that agents in dense networks influence one another’s emotional status (Lin and Qiu, 2012), to find structurally- and semantically-related groups of nodes (Cruz et al., 2013), and to identify social groups and clusters (Brooks et al., 2014).

NameGenWeb created an undirected network of friends, where an ego with $k$ friends can render a list of $(k \times (k - 1))/2$ possible connections (Jackson, 2008). Each data set was downloaded as an edgelist of mutual friendships and later converted to other network data structures (i.e., Graph Markup Language) based on software input requirements. Data were stored within the Neo4j graph database structure. The NameGenWeb application became unavailable toward the end of 2014, and a similar application, NetVizz (Rieder, 2013) was used for four volunteers. As of January 1, 2015, the ability to download a friendship network was no longer a feature of Facebook or its applications. Yet, independent researchers still...
provide these functionalities. For example, the Lost Circles team of University of Konstanz in Germany offers a free plug-in for Facebook network visualization and download (https://lostcircles.com/).

**Network metrics**

We calculated standard network descriptors, including diameter, density, and average clustering coefficient using the “igraph” library (Csárdi and Nepusz, 2006) in the R statistical computing environment.

**Geolocation**

Next, volunteers geocoded the locations of each of their friends at two time steps (using recollection and assistance from social media, such as Facebook profiles): the alter’s home location when the relationship began (pre-city) and the alter’s home location at the time of the study (post-city). A few alters met in “cyberspace” or had unknown locations and were removed because, although virtual spaces are geographic (Chen et al., 2013), we required metrics of distance change, i.e., two distinct geographic points, over time. Volunteers assigned longitude and latitude coordinates to alters by geolocating cities in Google Maps, Mapquest or Esri’s global gazetteer.

Each ego and alter were assigned an ID number to preserve anonymity. After anonymizing the data, the geographic coordinates were mapped within the ArcMap GIS software environment. These coordinates were spatially joined to existing shapefiles (i.e., spatial data files) of core-based statistical areas (CBSA) (metropolitan areas) or, if in a rural area, counties. Coordinates outside the USA were assigned to global administrative areas (GADM). As a result, each alter was assigned to a standardized urban center or a US county.

**Spatial calculations**

We then calculated the Euclidean distance between the ego’s current location and the locations of each alter’s pre-city and post-city, using the Great Circle method from the R “geosphere” package based on the WGS84 ellipsoid (Hijmans et al., 2017). To examine how social networks spread over time, we found the standard distance (i.e., the standard deviation of the longest linear axis of a point pattern) of each ego’s alters for both time steps. We determined the mean centers, that is, the average center of friend distributions, and reported the extent of the shift in mean centers over time in the ArcMap environment.

**Ground truthing**

The gravity model (as described earlier) is used in this study to predict the places where the egos are likely to have ties. We compared the locations of the alters’ cities to the gravity model for US cities only. First, we found the product of the population of State College and each alter’s locale (using 2000 US census population counts). We next computed the Euclidean distance from the town’s centroid to all other cities’ centroids.

**Community detection**

Facebook friendship networks tend to naturally cluster around different areas of a person’s life (Hogan et al., 2007) with significant clustering within universities (Lewis et al., 2008). To find clusters, or modules, within each network, we used the Louvain method for community detection (i.e., modularity calculation) (Blondel et al., 2008) within the Gephi environment (Bastian et al., 2009). To avoid double-counting alters, we removed 44 “overlapping” alters, i.e., nodes that are friends with more than one ego.

For each modularity group, egos identified an institution that the modularity group represents and the year of the group’s inception. We provided a crowdsourced set of labels, including university, secondary school, place-based cultural groups, and non-residential gatherings, such as conferences and vacation travel. Volunteers were invited to label groups using the aforementioned examples but were also able to choose their own labels. A modularity group required three members to be considered, and 90% of egos annotated their groups with a classification. We assume that each unique cluster within a Facebook network represents a specific social context (Brooks et al., 2014).

**Results**

**Network description**

All 20 egos had a total of 8,549 alters ranging from 123 to 772 per ego (Table 1, Figure 1). The number of edges connecting the nodes ranged from 784 to 12,667. The average degree of friends ranged from 6.9 to 37.7, indicating that some alters had many shared friends in the same network while others did not. Density is defined as the number of edges divided by the total possible edges in a network and is computed as the number of actual edges (e) that exist over the number of possible edges (k × (k − 1))/2. Average density ranged from 0.024 to 0.172 (Table 1, Figure 1). Diameter, defined as the...
Table 1. Summary statistics and geographic distribution of alters for each ego’s network.

<table>
<thead>
<tr>
<th>Agent</th>
<th>Nodes</th>
<th>Edges</th>
<th>Average degree</th>
<th>Diameter</th>
<th>Density</th>
<th>Clustering coefficient</th>
<th>Modules; modularity</th>
<th>Average distance to friends (post-period)</th>
<th>Standard distance pre-period</th>
<th>Standard distance post-period</th>
<th>Change in standard distance</th>
<th>Change of mean center</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>376</td>
<td>2679</td>
<td>12.3</td>
<td>10</td>
<td>0.038</td>
<td>0.36</td>
<td>4; 0.61</td>
<td>431</td>
<td>1468</td>
<td>2392</td>
<td>924</td>
<td>192</td>
</tr>
<tr>
<td>B</td>
<td>336</td>
<td>2533</td>
<td>14.5</td>
<td>9</td>
<td>0.045</td>
<td>0.53</td>
<td>13; 0.67</td>
<td>7827</td>
<td>5570</td>
<td>5691</td>
<td>121</td>
<td>667</td>
</tr>
<tr>
<td>C</td>
<td>601</td>
<td>7392</td>
<td>23.5</td>
<td>11</td>
<td>0.041</td>
<td>0.49</td>
<td>15; 0.62</td>
<td>277</td>
<td>1592</td>
<td>2395</td>
<td>803</td>
<td>383</td>
</tr>
<tr>
<td>D</td>
<td>370</td>
<td>3686</td>
<td>19.4</td>
<td>9</td>
<td>0.054</td>
<td>0.57</td>
<td>11; 0.71</td>
<td>2567</td>
<td>4377</td>
<td>3686</td>
<td>-691</td>
<td>597</td>
</tr>
<tr>
<td>E</td>
<td>573</td>
<td>9833</td>
<td>33.8</td>
<td>10</td>
<td>0.06</td>
<td>0.47</td>
<td>12; 0.51</td>
<td>1409</td>
<td>2754</td>
<td>3727</td>
<td>973</td>
<td>1873</td>
</tr>
<tr>
<td>F</td>
<td>232</td>
<td>2706</td>
<td>22.3</td>
<td>5</td>
<td>0.10</td>
<td>0.53</td>
<td>12; 0.37</td>
<td>1174</td>
<td>2917</td>
<td>3903</td>
<td>986</td>
<td>455</td>
</tr>
<tr>
<td>G</td>
<td>361</td>
<td>2989</td>
<td>14.8</td>
<td>12</td>
<td>0.04</td>
<td>0.54</td>
<td>12; 0.71</td>
<td>215</td>
<td>887</td>
<td>2283</td>
<td>1396</td>
<td>368</td>
</tr>
<tr>
<td>H</td>
<td>437</td>
<td>6002</td>
<td>27.2</td>
<td>8</td>
<td>0.06</td>
<td>0.46</td>
<td>10; 0.51</td>
<td>1185</td>
<td>627</td>
<td>1267</td>
<td>640</td>
<td>188</td>
</tr>
<tr>
<td>I</td>
<td>403</td>
<td>5022</td>
<td>24.7</td>
<td>8</td>
<td>0.062</td>
<td>0.43</td>
<td>9; 0.52</td>
<td>687</td>
<td>3087</td>
<td>2383</td>
<td>-704</td>
<td>518</td>
</tr>
<tr>
<td>J</td>
<td>437</td>
<td>2763</td>
<td>11.5</td>
<td>10</td>
<td>0.029</td>
<td>0.52</td>
<td>13; 0.78</td>
<td>1619</td>
<td>3726</td>
<td>4157</td>
<td>431</td>
<td>409</td>
</tr>
<tr>
<td>K</td>
<td>639</td>
<td>9784</td>
<td>30.5</td>
<td>7</td>
<td>0.048</td>
<td>0.33</td>
<td>10; 0.46</td>
<td>712</td>
<td>2534</td>
<td>2893</td>
<td>359</td>
<td>575</td>
</tr>
<tr>
<td>L</td>
<td>188</td>
<td>1002</td>
<td>10.4</td>
<td>10</td>
<td>0.057</td>
<td>0.64</td>
<td>6; 0.74</td>
<td>697</td>
<td>3047</td>
<td>2605</td>
<td>-442</td>
<td>129</td>
</tr>
<tr>
<td>M</td>
<td>201</td>
<td>784</td>
<td>6.9</td>
<td>10</td>
<td>0.039</td>
<td>0.16</td>
<td>16; 0.46</td>
<td>8522</td>
<td>5727</td>
<td>5610</td>
<td>-117</td>
<td>325</td>
</tr>
<tr>
<td>N</td>
<td>772</td>
<td>7143</td>
<td>18.2</td>
<td>11</td>
<td>0.024</td>
<td>0.44</td>
<td>16; 0.76</td>
<td>529</td>
<td>3274</td>
<td>3357</td>
<td>83</td>
<td>292</td>
</tr>
<tr>
<td>O</td>
<td>123</td>
<td>1028</td>
<td>15.8</td>
<td>9</td>
<td>0.137</td>
<td>0.74</td>
<td>9; 0.40</td>
<td>6701</td>
<td>3588</td>
<td>5564</td>
<td>1976</td>
<td>448</td>
</tr>
<tr>
<td>P</td>
<td>659</td>
<td>11,708</td>
<td>34.9</td>
<td>8</td>
<td>0.054</td>
<td>0.50</td>
<td>14; 0.52</td>
<td>1508</td>
<td>3726</td>
<td>3552</td>
<td>-174</td>
<td>1150</td>
</tr>
<tr>
<td>Q</td>
<td>727</td>
<td>12,667</td>
<td>34.5</td>
<td>9</td>
<td>0.048</td>
<td>0.53</td>
<td>10; 0.69</td>
<td>492</td>
<td>1792</td>
<td>2951</td>
<td>1159</td>
<td>520</td>
</tr>
<tr>
<td>R</td>
<td>227</td>
<td>4412</td>
<td>37.7</td>
<td>6</td>
<td>0.172</td>
<td>0.55</td>
<td>10; 0.23</td>
<td>176</td>
<td>49</td>
<td>1658</td>
<td>1609</td>
<td>196</td>
</tr>
<tr>
<td>S</td>
<td>408</td>
<td>7639</td>
<td>37.0</td>
<td>7</td>
<td>0.092</td>
<td>0.52</td>
<td>9; 0.52</td>
<td>626</td>
<td>1693</td>
<td>2929</td>
<td>1236</td>
<td>441</td>
</tr>
<tr>
<td>T</td>
<td>566</td>
<td>4157</td>
<td>14.3</td>
<td>14</td>
<td>0.026</td>
<td>0.53</td>
<td>15; 0.75</td>
<td>1148</td>
<td>1156</td>
<td>2167</td>
<td>1011</td>
<td>560</td>
</tr>
</tbody>
</table>

Note: All distances measured in kilometers.
longest shortest-path that connects two nodes, ranged from 6 to 14. However, not all nodes were reachable, due to some isolates and disconnected components. Lower diameter values imply more “friends of friends” who are also friends (e.g., triads), whereas higher values imply more “chains” of friends who do not have common friends (Jackson, 2008). The average clustering coefficient, ranging from 0.16 to 0.74, represents the probability of a node being part of a network triangle. Dense networks with high clustering coefficients and low diameters can be considered tight-knit.

Geographic distribution and spread of social ties

Egos provided a total of 1207 recorded locations: 430 individual places logged in the pre-period and 777 in the post-period. Alters spread over time, producing more places for egos to visit, telecommunicate with, think about, and in which to claim they have social capital (i.e., the benefits that relationships can yield). Accounting for duplicates across the two time periods, there were 942 unique locations (and 265 repeats), which represent the collective places where egos are likely to interact at a personal level. At the time of the study, three egos could see friends at, on average, less than 300 km distance (C, G, R), whereas others (such as M) had ties, on average, 8,522 km away (Table 1). Some egos had fewer than five locations during friend inception, and others had over 50 locations. Some egos (such as R) met their alters within a small radius (49 km), while others (such as B or M) met alters over a wide geographic radius (over 5000 km) (Table 1). These differences are an interesting way to describe one’s “life experience” (i.e., gaining ties from very few or a variety of places), as each respective scenario may be associated with local or global experiences (Fischer, 1982). For instance, some egos have extensive international experience, while others were rooted in the region.
Friends have generally dispersed over time (Table 1), at distributions ranging from 83 km (V) to 1976 km (O) (Table 1). However, in some cases, friends moved closer to one another over time: the alters of five egos (D, I, L, M, P) decreased in standard distance ranging from −703 to −116 km. The collective distance decay distribution (Figure 2) illustrates that in the pre-period (blue line), more alters lived nearby, and in the post-period (red line), more alters live between 1,500–3,000 km from the ego. To illustrate spreading, a single ego’s example alter distribution shows ~35 locations when the ego met each alter, and nearly twice as many locations at the time of the study (Figure 3). As a result, the ego can access more locations over time, given the same set of friends. This map reveals the evolving potential for friends to meet, and the difficulties involved therein.

Along with the change in standard distribution, the mean center of each ego’s friends (i.e., the center point of the distribution) has shifted as well. The mean center of each ego’s set of alters changed as much as 1,873 km (E) and as little as 129 km (L) (Table 1). A small shift in mean center indicates that one’s friendship “center of gravity”, or most accessible place to all alters, does not change significantly over time, whereas a large shift means that friends’ locations may have changed.

Due to privacy concerns associated with our small sample size and requests from participants, we did not disclose the distinct countries where international friends are located. Indeed, volunteers were initially hesitant to provide international locational information. For example, a participant with many friends (~100) in a single foreign country was reluctant to publish the name of this country, although many countries had more than 100 alters (often furnished by one ego). When discussing privacy, we sometimes tend to insinuate personal privacy, yet the participants’ primary consideration was their friends’ privacy, especially in non-urban locations. Egos were also uncomfortable listing how many alters they had in a single city if data from no other alter was furnished. We gathered this information on privacy preferences by asking volunteers about their preferences during the data collection process.

Internationally, alters lived in 64 countries (including the USA) in the pre-period and 90 countries in the post-period. 7472 (87%) friends lived in the US during the pre-period, and there was a net gain of 145 alters abroad over time. Excluding the US, five countries had at least 50 alters in the pre-period and eight countries in the post-period, meaning certain countries attracted many individuals. Similarly, ten countries had between 10 and 49 alters in the pre-period and 12 had 10–49 alters in the post-period. None of the top five countries for alters have English as a primary language, but in the following ten, four have English as a primary language. Time yielded fewer South American and Asian-residing alters, and more alters in Australia, North America and Europe.

**Top cities for interaction in the USA**

In the US, participants met alters in 248 cities or counties during the pre-period, and had alters in 426 cities or counties during the post-period. Locations were mostly concentrated in the Eastern Atlantic United States (Table 2). A number of egos had alters in the US outside CBSAs, living in rural areas in Iowa, New York, Georgia, and Missouri.

A “heat map” result shows the loss and gain of alters between the two time periods for the collective group (Figure 4). This figure does not show the raw frequency of alters, but is normalized based on a quadratic kernel function (Silverman, 2018) (akin to a probability density function) that effectively smooths the data within a spatial radius. Alters have
moved from Pennsylvania, Washington, DC and the Midwest and now inhabit geographic pockets, notably in the East Coast, North Carolina, Denver, CO, and West Coast cities (Figure 4). This finding is in line with research using 2000 census data that heralds large metropolitan areas, and the USA South and West, as key places for young, college-educated adults (Franklin, 2003). In addition, the places that draw alters, such as Los Angeles, Seattle, San Francisco, and North Carolina’s Research Triangle are information technology hubs (He and Fallah, 2011) with advanced research and job opportunities in the tech sector, a field of expertise for a number of the egos. Incidentally, the seminar course that anchored this study was in the field of information science.

A general gravity model from State College predicts high interaction with following cities (in order): New York, NY, Philadelphia, PA, Washington DC, Harrisburg, PA, Pittsburgh, PA, Baltimore, MD, Lewistown, PA, Altoona, PA, and DuBois, PA. Of the 50 top cities for predicted interaction, 10 of the 20 pre- and post-cities are represented (Table 2).

In general, results followed the gravity model to a small extent. This value correlated with actual interaction at rates of 0.45 ($R^2$ value) in the pre-period, and 0.56 in the post-period. The lower predictability in the pre-period is likely due to the fact that the egos themselves may not have lived in the region when they met their alters. If the model was centered around the ego’s previous residence(s), their individual gravity models may have been highly predictive of friend locations. The post-period’s higher fit was driven by preferences toward high-population cities, such as Atlanta, Seattle and Los Angeles, regardless of distance. This result matches findings from life course migration theory, specifically that younger adults flock to large metropolitan areas not only in the US.
Table 2. Top 20 most popular domestic alter cities and the number of egos with alters in these locations in the pre-period and post-period.

<table>
<thead>
<tr>
<th>Pre-city</th>
<th>Alters</th>
<th>Egos</th>
<th>Post-city</th>
<th>Alters</th>
<th>Egos</th>
</tr>
</thead>
<tbody>
<tr>
<td>State College, PA</td>
<td>1384</td>
<td>17</td>
<td>State College, PA</td>
<td>797</td>
<td>18</td>
</tr>
<tr>
<td>Knoxville, TN</td>
<td>222</td>
<td>5</td>
<td>Chicago–Naperville–Joliet, IL–IN–WI</td>
<td>224</td>
<td>16</td>
</tr>
<tr>
<td>Atlanta–Sandy Springs–Marietta, GA</td>
<td>196</td>
<td>7</td>
<td>San Francisco–Oakland–Fremont, CA</td>
<td>183</td>
<td>14</td>
</tr>
<tr>
<td>San Francisco–Oakland–Fremont, CA</td>
<td>186</td>
<td>8</td>
<td>Los Angeles–Long Beach–Santa Ana, CA</td>
<td>133</td>
<td>17</td>
</tr>
<tr>
<td>Williamsport, PA</td>
<td>180</td>
<td>3</td>
<td>Dallas–Fort Worth–Arlington, TX</td>
<td>132</td>
<td>15</td>
</tr>
<tr>
<td>Austin–Round Rock, TX</td>
<td>178</td>
<td>5</td>
<td>Albany–Schenectady–Troy, NY</td>
<td>129</td>
<td>6</td>
</tr>
<tr>
<td>Dallas–Fort Worth–Arlington, TX</td>
<td>159</td>
<td>6</td>
<td>Austin–Round Rock, TX</td>
<td>120</td>
<td>15</td>
</tr>
<tr>
<td>Appleton, WI</td>
<td>125</td>
<td>2</td>
<td>Williamsport, PA</td>
<td>107</td>
<td>4</td>
</tr>
<tr>
<td>Riverside–San Bernardino–Ontario, CA</td>
<td>104</td>
<td>5</td>
<td>Cedar Rapids, IA</td>
<td>92</td>
<td>2</td>
</tr>
<tr>
<td>Harrisburg–Carlisle, PA</td>
<td>61</td>
<td>3</td>
<td>Augusta–Richmond County, GA–SC</td>
<td>82</td>
<td>5</td>
</tr>
<tr>
<td>San Diego–Carlsbad–San Marcos, CA</td>
<td>45</td>
<td>7</td>
<td>Riverside–San Bernardino–Ontario, CA</td>
<td>76</td>
<td>7</td>
</tr>
</tbody>
</table>

Note: Six pre-cities were removed for privacy because only one ego supplied alters from the city.

(Glaeser, 1999), but internationally as well (McCormick and Wahba, 2005). The pull to large cities suggests that the alters, like the egos, may be young, educated and highly mobile agents.

The differences between actual and predicted ranks of each city further support the proclivity toward large cities. Washington, DC’s pre-city and post-city rank is 2 and gravity rank is 3 out of all cities in the network. Therefore, Washington, DC could be described by the absolute value difference (δ) between its gravity model rank and our pre- and post-city ranks (δ Pre: 1, δ Post: 1). Nearby metropolitan areas, Philadelphia (Pre: 11, Post: 5) and especially New York (Pre: 16, Post:1), synchronize with the gravity model in the post-period. (The frequency of contacts in each locale at both time steps vis-à-vis gravity model predictions can be explored in Figures A1 and A2).
Figure 4. Individuals have fewer alters in red regions and more alters in blue regions over time, illustrating a shift toward coastal locales, North Carolina, and large cities in the West. Hot and cold spots are created using a kernel density function.

**Modules and geographic spreading**

In total, 143 modularity groups were identified. We found that, generally, social groups spread over time (Table 3). Groups associated with stronger ties to place and long-term membership (i.e., family, cultural groups) spread less than those associated with transitory institutions (such as universities, co-workers and secondary education). Educational groups, including university and high school, and non-residential gatherings, such as conferences and summer camps, spread over 1500 km over time. These also tend to be the “youngest” groups. Although they had more time to spread, older groups (those with longer average time since formation) spread the least since their inception. This finding matches social network research discoveries that some groups are more stable than others, particularly that family ties endure over extended friendship networks (Ertel et al., 2009). Recreational groups, including place-based cultural groups, did not incur significant spreading, and a review of social group longevity revealed that outdoor, cultural, or sports groups tend to dissolve time (especially for men), while religious groups tend to persist over time (Hyypä et al., 2008).

Neighbors were largely left out of these distinctions – very few egos labeled their modules as “neighbors”. This may be due to a lack of traditional “neighbors” being on Facebook, or that college neighbors (e.g., dormmates) are commonly regarded as friends. Nevertheless, neighbor groups do not tend to persist through time for highly mobile individuals. As Wellman et al. (1997) find, over a decade, 20% of neighbors maintained ties after moving.

**Discussion**

Within the framework of relational geography, we conducted a social network mapping case study. In response to our research questions, we found that egos had five to 50+ locations where they have met alters...
Table 3. Table of group types and changes in standard distances.

<table>
<thead>
<tr>
<th>Group type</th>
<th>Modules</th>
<th>Avg. standard distance before; after (km)</th>
<th>Difference in standard distance (km)</th>
<th>Avg. years since inception</th>
</tr>
</thead>
<tbody>
<tr>
<td>University</td>
<td>36</td>
<td>2298; 3810</td>
<td>1512</td>
<td>8.5</td>
</tr>
<tr>
<td>Professional</td>
<td>21</td>
<td>2468; 3927</td>
<td>1460</td>
<td>6.6</td>
</tr>
<tr>
<td>Secondary education</td>
<td>20</td>
<td>527; 2441</td>
<td>1914</td>
<td>14.2</td>
</tr>
<tr>
<td>Place-based cultural group</td>
<td>16</td>
<td>1437; 2224</td>
<td>788</td>
<td>10.7</td>
</tr>
<tr>
<td>Family</td>
<td>15</td>
<td>1678; 2459</td>
<td>781</td>
<td>22.2</td>
</tr>
<tr>
<td>Non-residential gathering</td>
<td>14</td>
<td>2898; 4865</td>
<td>1967</td>
<td>6.5</td>
</tr>
<tr>
<td>Non-place-based cultural group</td>
<td>8</td>
<td>2463; 4267</td>
<td>1805</td>
<td>9.5</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>4429; 5410</td>
<td>981</td>
<td>5.5</td>
</tr>
</tbody>
</table>

and the average tie distance for individual egos ranged from 83 to 1,976 km. These findings resemble Bailey et al.’s (2018) map of average tie distance in that some egos’ distance corresponds to a typical person in Penn State’s Appalachian region (where most friends live less than 100 km away), while other have much higher distances (typical of large cities and West coast cities). In terms of the individual cities that attracted alters, this geographic shadow points to alters also being young and educated, despite opportunities for Facebook friends to include more than these student volunteers’ peers (e.g., aunts and uncles).

Over time, contacts dispersed geographically. On average, egos had alters in 21 individual cities, counties and countries (depending on how they reported locales) in the pre-period and 38 in the post-period. Egos’ alters lived in a combined 64 countries in the pre-period and in 90 countries in the post-period – 145 alters moved abroad. Given the demographics of the egos, the mechanisms behind the spreading of alters are assumed to be job prospects and opportunities, or returning to a home country, rather than forced migration (political, economic, or otherwise). The notion of growing distances over time aligns with prior research, namely, findings that those who move farther maintain more dispersed networks, and those who move locally keep proximate social networks (Magdol, 2000). A number of egos in this study moved to State College for graduate school, some from international locations.

In the pre-period, about 45% of friend distribution locations were predicted by the gravity model, and this value rose to 56% in the post-period. This group had a distinct lack of social ties around State College, in nearby Pennsylvania cities Lewisburg, Altoona and DuBois, which may be explained by egos’ preference for large cities. The locations of potential interaction may be perpetuated by the network of universities and other institutions as opposed to nearness. We find that the predictive power of the gravity model is slightly less than previous research on the gravity model’s ability to predict inter-city US migration, which has been measured at about 60% (Andris et al., 2011). The increase in predictive power over time is expected, and is attributed to two factors: friends’ migration to larger cities, and that in the pre-period, the egos may not have lived in State College – providing a false center from which to measure interaction distance.

Groups of contacts spread at different rates, but each type of group experienced geographical dispersion. The spreading of education-based groups can be explained by migration research that cites higher-education cohorts’ above-average movement rates where the young and educated produce the highest rates of migration in the USA (Frey, 2005). It also reflects findings that that living far from home, a history of residential movement, and high education levels equate to geographically distributed networks (Viry, 2012).

Accordingly, education level may be the most distinguishing predictor variable of this cohort. As such, this cohort may want to pay special attention to university and professional contacts, and their respective locations in the future, as individuals tend to stay in touch with higher-education ties, especially those that provide career support (Feld et al., 2007), and more education leads to a greater capacity to maintain distant ties (Viry, 2012). In one study, between 25–33% of
classmates, co-workers and friends persisted across a decade, and educational attainment predicted the persistence of school/professional ties (Suitor and Keeton, 1997). Collectively, these findings suggest that the groups created from professional and university institutions may play a substantial supporting role for the egos in the future – and, thus, their locations matter.

The research questions in this work were largely conceptual, and the study has a set of limitations. Our sample size was relatively small. This is a highly specialized group of 20 egos that only serves as a proof-of-concept for mapping the social ties for this subset of the population. Recruiting volunteers was difficult due to the newness of geolocated social networks and the lack of previously successful studies to which we could refer potential contributors. As a result, we limit the generalizability of these particular results to the wider population, since our sample is a group of young, mobile university students, some of whom are not native to the region. In the future, examining diverse networks of the elderly, those with different levels of educational achievement, or migrant populations may yield different patterns of spreading behavior – perhaps flows to retirement communities, less spreading in general, or movement within and between ethnic enclaves, respectively.

There were also challenges associated with data creation. With each ego geolocating upwards of 500 highly mobile contacts, there was likely to be human error in tagging locations and relying on Facebook profiles to report the alter’s most current location. In a few cases, egos could not recall where they met or how they knew an alter, and thus, the alter was removed from the data set. However, since egos filtered their own data and validated their modularity groups with their raison d’être (in the form of “club”, “family”, etc.), digital footprint data quality issues (Lewis, 2015) such as lack of context and over-inclusive sample, were partially remedied. Regarding analysis, although we used modularity to detect social groups, a social group is not always synonymous with a module. Modularity algorithms may omit social group members or may capture individuals who were not in the social group but were grouped together mathematically due to mutual ties with the group’s members.

New research studies could be developed based on the social-spatial perspective proposed here. Future studies could also use data from Facebook or other social networks to create a larger and more diverse set of egos. With a larger data set, we can more reliably measure spatial network regularities and create benchmarks for cross-study comparisons. These measures should include how proximal an individual’s social support is and the variety of places in which an individual has social support, in order to determine whether there is a common spatial rhythm to the spatial distribution of human relationships.

An ideal study would map alters at multiple time steps, perhaps in real time, with technologies that resemble the Snap Maps described above. Egos should be interviewed or surveyed about how access to their own dynamic, visual address book of contacts affects decisions to travel, telecommunicate, maintain friendships, and perceive places, as one may develop different migration or travel plans after reexamining their tie locations. If the social tie has a hidden byproduct of helping the ego explore new places, it may be that a link between two people also indicates a link between two places, and we can explore the validity of this theory under various definitions and conditions. This exercise would also help characterize social networks. For example, although an ego may have a strong, vibrant social network, this network may be limited in terms of the places he or she can be exposed to through these ties. On the other hand, an ego may have a relatively sparse social network, but can access a variety of attractive places through these ties.

Despite the study’s drawbacks, this case study yielded new insights about how ties are distributed and how time affects social network geography. We hope this exploratory research can lead to future studies at larger scales to create new findings about these personalized distributions.

Acknowledgments

The authors thank Brian Swedberg, Timmy Huynh, and the anonymous volunteers who participated in the study. The authors also thank the anonymous reviewers who provided helpful suggestions.

References


Mapping Facebook Friends


Mapping Facebook Friends


Appendix. Maps of alter locations vis-à-vis gravity model predictions.

Figure A1. Top cities for ego interaction at the time of friendship inception (as contours and white hot spots) are compared with the top cities for gravity model interaction (as hot and cold colors). Each number represents the rank of the city by its likelihood for interaction, given the gravity model’s prediction. White outlines represent popular cities for friends and yellow outlines represent cities with high theoretical interaction but few friends. Unpopular cities for friendships (marked in yellow) are peripheral to the location of the university.
Figure A2. Top cities for ego interaction at the time of the study vis-à-vis top cities for gravity model interaction. Unpopular cities for friendships are still peripheral to the location of the university.