Computers, Environment and Urban Systems xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

Challenges for social flows

Clio Andris[a,b,*], Xi Liu[a,b], Joseph Ferreira Jr[a,b]

[a] Department of Geography, The Pennsylvania State University, 302 Walker Building, University Park, PA 16802, USA
[b] Department of Urban Studies and Planning, Massachusetts Institute of Technology, 77 Massachusetts Avenue Room 9-209, Cambridge, MA 02139, USA

ARTICLE INFO

Keywords:
GIS
Social networks
Spatial data
Origin/destination data
Spatial interaction

ABSTRACT

Social and interpersonal connections are attached to the built environment: people require physical infrastructure to meet and telecommunicate, and then populate these infrastructures with movement and information dynamics. In GIS analysis, actions are often represented as a unit of spatial information called the social flow–a linear geographic feature that evidences an individual's decision to connect places through travel, telecommunications and/or declaring personal relationships. These flows differ from traditional spatial networks (roads, etc.) because they are often non-planar, and unlike networks in operations systems (such as flight networks) provide evidence of personal intentionality to interact with the built environment and/or to perpetuate relationships with others. En masse, these flows sum to illustrate how humans, information and thoughts spread between and within places.

Amid a growing abundance and usage of social flow data, we extend formal definitions of this data type, create new typologies, address new problems, and redefine social distance as the manifestation of social flows. Next, we outline challenges to fully leveraging these data with commercial GISystems by providing examples and potential solutions for representing, visualizing, manipulating, statistically analyzing and ascribing meaning to social flows. The goal of this discussion is to improve the dexterity of social flow data for geographic, environmental and social research questions.

1. Introduction

Conceptually and cognitively, humans are rarely in just one place. They interact with other places through experiences of travel, moving to new locations, commuting, vacationing, telecommunicating with others or interacting with mass media (Adams, 2005). These human-centric activities are often called social flows. More specifically, social flows can be defined as *agent*-based decisions to connect places through movement, telecommunications or stated relationships (as in Andris, 2016). Social flows are units of analysis within part of a larger family of flow/interaction/spatial network data (as in Haggett & Chorley, 1969; Hagerstrand, 1968; Clarke, 1986). Historically, social flows have included household newspaper orders (Green, 1955), taxi trips (Goddard, 1970), migration (ex. Greenwood & Sweetland, 1972), postal volumes (Pred, 1973), and telephone calls (Rietveld & Janssen, 1990) have guided research on interaction for many decades.

However, their granularity has changed significantly from aggregate supply and demand values to GPS traces and geolocated social media (Dodge, Weibel, & Lautenschütz, 2008). Accordingly, social flow research is burgeoning due to exciting new datasets of taxi traces, phone calls, friendships, migrants, tourists and economic flows, that have spurned new analysis methods for detecting connection patterns between and within places (ex. Larsen, Axhausen, & Urry, 2006, Arentze & Timmermans, 2008, Carrasco & Miller, 2009, Carrasco et al., 2008; Ahas, Silm, Järv, Saluveer, & Tiru, 2010, Frei & Axhausen, 2011, Dugundji et al., 2011, Arentze, van den Berg, & Timmermans, 2012). Tracking these related studies is difficult as social flow keywords are diverse, including: networks, traces, interaction, connectivity, social topology, graph theory, social physics, distance decay, origin-destination matrices, and flow dynamics, and vary according to the researcher's domain.

Social flows are challenging to adopt in GIS for a number of reasons. Social variables in GISystems are often points (ex. people's locations) and polygons (ex. census data), but social flows are linear, largely non-planar spatial data types, this characteristic alone promises integration issues with GISystems (Thill, 2000). Next, their agent-based nature assumes the accumulation of spatial experience and exposure (Dodge et al., 2008), but also, the potential for social relationships. Unlike other linear structures in GISystems, such as roads and fences, social flows are marked with intentionality, cognition, information transfer and the philosophical underpinnings of a personal tie.

These data have been used abundantly, but are overlooked as part of

* Corresponding author at: Department of Geography, The Pennsylvania State University, 302 Walker Building, University Park, PA 16802, USA.
E-mail address: clio@psu.edu (C. Andris).

https://doi.org/10.1016/j.compenvurbsys.2018.03.008
Received 5 September 2016; Received in revised form 25 October 2017; Accepted 22 March 2018

Please cite this article as: Andris, C., Computers, Environment and Urban Systems (2018), https://doi.org/10.1016/j.compenvurbsys.2018.03.008

a family of GIS data types and are lacking a needs assessment for effective integration with GIScience (Torrens, 2010). Our contribution in this article is to explain the nature of these data and the challenges to their use. We extend previous network/GIS studies by applying methods for transportation and utility networks (Curtin, 2007) and paradigms of network-discrete entity fusion to data (Goodchild, 1998) to non-planar flows of agent-based decisions.

Social flow manipulation requires creative solutions that are learned through loose-coupling and tacit knowledge. Here, we compile and articulate these issues, and in response, suggest both new solutions and retrofit prior solutions to social flow challenges. We find that the facilitation of social flow usage in GISystems and GIScience requires a mixture of modern tools and statistical techniques from network science (ex. Chi, Thill, Tong, Shi, & Liu, 2016), a re-thinking of adjacency and distance in geographic space, and theoretically, a reflection on what it means for an agent to connect places through his or her actions. We argue that these data are valuable to geographers and spatial modelers, and that challenges to integration are surmountable. We specifically outline how they can be better integrated with other spatial layers.

In this article, we group social flow challenges into four categories (a) How to characterize and manage components, (b) What digital infrastructure and categories are needed represent and combine components, (c) How to analyze and apply computational methods to social flows, (d) How to contribute to knowledge in a meaningful way and derive meaning from these flows. We first discuss conceptual advancements in social flows within GISystems, including

- A formalization of the concept of social flow causes, collection methods and types
- A reinvigoration of the traditional concept of social distance to reflect 21st century data types
- A discussion of key problems of flow embeddedness; flow duality, and combining planar and network variable distributions
- A typology for social flows within a spectrum of flow/network type

We then revisit key technological issues for social flows within GISystems and offer potential solutions. These topics include difficulty visualizing flows, difficulty manipulating flows in a GISystem (spatial joins and interactive selection), a lack of spatial analysis methods, a lack of exploratory spatial data analysis (ESDA) tools, and unclear sampling methods for flows. Finally, we present a theoretical discussion on ascribing meaning to social flows.

This article is written primarily for an urban or environmental planner who is interested in learning more about social flow datasets and using these data within GISystems and in GIScience to augment geographical models or derive new place-based statistics.

1.1. Social flows

Social flows and can be divided into transportation and human travel (including migration), telecommunications (including mass media), and stated relationships that are collected through passive sensors and tabulated (e.g. administrative) records alike (Table 1: Collection Methods) producing data that connect two or more places (Table 1: Evidence).

GIS models of urban form, growth and patterning used for scenario modeling, forecasting, prediction and visualization (Batty, 2005; White, 2015) use traditional socio-economic and demographic data, such as unemployment rates, race, income and education level, to illustrate the human "layer". Social flows of commuter and leisure trips (Limtanakool, Schwanen, & Dijst, 2009) or spatio-temporal data of events, crowds, and a daytime/nighttime population (such as McKenzie, Janowicz, Gao, & Gong, 2015) can enhance the spatial and temporal resolution of population patterns and inferred activities, linguistics, topics of interest—i.e. what locals might be tweeting, posting or texting.

Through these connections, humans fold together non-adjacent places and sew deep patterns of movement, information transfer, and settlement in the world. This philosophical idea has been described as the contortion of Euclidean space—the pinching and expanding of spatial features to better reflect one's perception of distance (Golledge & Hubert, 1982) or the costs of travel (L'Hostis, 2009). Space is also conceptually contorted due to high volumes of behavioral activity that reflect flow volume (i.e. magnitudes) between places.

1.2. Social distance

When social flows are summed (i.e. aggregated) by common origin and destination pairs, they comprise a useful network that describes how humans, information and thoughts spread between and within places (as in Plane 1984, via Thill, 2011). This "social distance" can be defined as the dispersion between two or more places as measured by the inverse of the magnitude of social flows between the places. An increase of social flows will result in the "shortening" social distance between places (Fig. 1). ***Social distance*** described by Regional Scientist Walter Isard (Isard, 1966, Deutsch and Isard, 1961), using the example of Hollywood and New York to illustrate short social distance (as defined by heavy communications flows) despite a sizable physical distance between the pair, as a result of a lack of intervening opportunities in industry (Stouffer, 1940) between Hollywood and New York (Isard, 1960, p. 542).

This paradigm has also been called ***metaphorical or functional distance*** measured by "social contacts between places or individuals (frequency of shopping trips, telephone calls, cultural exchanges, journal subscriptions)" (Müller, 1982, p. 190); ***effective distance*** measured by following flows on the infrastructure network (Brockmann & Helbing, 2013); and ***logical distance*** which measures "dispersion between towns in terms of labor- or study-led activities" (De Montis et al., 2010, p. 46). These versions of social distance are pragmatic answers to calls to fuse networks with geography in the Digital Era (Gastner & Newman, 2006, Xu & Sui, 2007, Xu & Harriss, 2008, Limtanakool et al., 2009, Torrens, 2010, Zook, 2010).

Different from Euclidean and cost distance, which are commonly used in GIS and planning, social distance can reveal desire lines, i.e. the "hidden" routes where information, people, and artifacts (ex. disease or goods) are transmitted (Cowan, 2005). Social distance challenges the hypothesis that nearby places are similar (Tobler 1970), as two nearby cities may be disconnected in terms of language (Expert, Evans, Blondel, & Lambiotte, 2011), information transfer (Ratti et al., 2010, Chi et al. 2016), or genealogy (Cheshire, Longley, & Singleton, 2010). However, when the term "nearby" is translated into network distance, non-adjacent place similarity (or vice versa) can be explained. That is, planners and geographers can use social distance to explain that nearby places in Euclidean and cost distance may not be well-connected, or that faraway places are well-connected. Social distance can be used to predict how cities may change with the influx of particular connection patterns, and not simply by local activity and demographics. It can also be used to quantitatively symptomize social and economic heterogeneities between nearby places.

1.2.1. An example

We use an example of social flows between Austin, Texas, U.S.A. and other U.S. metropolitan statistical areas (MSAs) to show that different types of social flows exhibit different geographical ranges (Fig. 2). Austin has a reputation for technological advancements and progressive politics within a conservative surrounding[1] and may leapfrog (Seto et al., 2012) over its neighboring cities to connect with like-

[1] For instance, ethnographer Barry Shank (1994) writes of "cultural distinctions between the relatively liberal town of Austin and the remainder of the highly conservative state of Texas." (p. 16).