Position paper: **Tracking and understanding information spread through advanced space-time analysis**

Li An, lan@mail.sdsu.edu; Sarah Wandersee, wanderse@rohan.sdsu.edu

Department of Geography, San Diego State University

Events such as radical concepts or epidemic outbreaks affect our life in many significant ways. Understanding the diffusion patterns as well as potential drivers behind such patterns will significantly empower scientists, professionals in various fields, and the general public for better intervention and possible prevention purposes. Substantial advancements in web-based data collection (see position papers from Tsou, Gupta, Gawron, and Spitzberg, PIs of the NSF-CDI project "Mapping Ideas from Cyberspace to Realspace"; Award #1028177) have made data related to such events more obtainable over both space and time. Traditional separation between spatial analysis and time series analysis, however, has hindered the discovery and understanding of the mechanisms and relationships behind such data. In this context, space-time analysis of web search results aims to bridge this gap and reveal patterns and potential driving forces behind the associated idea (or disease) outspread.

Specifically, we will first employ a few metrics to quantify spatial patterns at discrete times, including those used in landscape ecology (e.g., network connectivity, dominance, proximity), map comparison (e.g., Pontius et al. 2004), and spatial statistics (global and local Moran's I, univariate and bivariate LISA, Geary's G; Anselin et al. 2006). In calculating dominance of a certain type of influence, the major type (e.g., with highest frequency) will be assigned to the city if a city has been substantially influenced by multiple types of events. By proximity we mean to measure how a focal city under a certain type of event is isolated from other cities (or clusters of cities) of the same type using the proximity index (Turner et al. 2001, p.114). Such metrics offer snapshot measures of the associated patterns at discrete time points.

Second, we will employ space-time trajectory measures that characterize the temporal change of spatial events, e.g., LISA time paths (Anselin et al. 2006; Ye 2009), space–time composites (Langran and Chrisman (1988), spatiotemporal objects (Worboys 1992), and space-time prisms (Hägerstrand 1970). Particularly, we will use the "hazard" (a term quite often used in sociology, demography, and epidemiology) concept to depict and quantify the risk of an area being dominated or substantially influenced by certain events (e.g., ideas). The concept hazard is calculated as a function of survival time T, or the time within which a city is free from a certain event (Allison 1995, p.63-66). As the time frame of our data collection and analysis is not too long, it is reasonable to assume that once the event happens for the first time at a certain place, then it can be labeled as "influenced" regardless of whether the same type of events would happen again or not. This will avoid working on hazards related to repeated events (e.g., a city may incur the same/similar topic multiple times), which is more complicated but still can be handled in survival analysis (Allison 1995, p.236-247; An and Brown 2008).

Third and last, we will employ survival analysis to link hazards with a set of independent variables (with the link to real world data such as census data), exploring what factors may affect the spread of ideas over space and time. These independent variables may take constant or changing values over time. The latter ones, termed time dependent variables in survival analysis, may include population density, unemployment rate, religion, ethnicity, and particularly the number and proportion of people who are at ages of 15-24 that are indicative of the existence of a "youth bulge" (e.g., Urdal 2006). These data will be assembled from a variety of sources, including census data from each country (if analysis is conducted within USA), Demographic and Health Survey data produced by Macro International, Multiple Indicator Cluster Surveys produced by United Nations Children's Fund, and the World Bank Living Standards Measurement Surveys.

Depending on how precisely we measure the survival times, we can choose different hazard models, such as the Cox proportional hazard model (Allison 1995, p. 113-114) or the three-step strategy proposed by An and Brown (2008) and An et al. (2011): (1) Generate a piecewise dataset where each of the city-period (or county-period depending on the spatial unit we use) observations incorporates all types of censoring and time-dependent variables; (2) treat all the types that are not being considered in a particular model as right-censored observations, which allows for competing risks; and (3) model the relationships between the hazards and the explanatory variables using the piecewise exponential model (the lifereg procedure in SAS). We will also use other types of models (e.g., logistic regression, generalized linear models) to build links between the idea diffusion patterns and various socioeconomic /demographic variables.

References

Allison, P. D. 1995. Survival analysis using SAS ┌R : A practical guide. Cary, NC: SAS Institute.

An, L., and D. G. Brown. 2008. Survival analysis in land-change science: integrating with GIScience to address temporal complexities. Annals of Association of American Geographers, 98(2): 323-344.

An, L., D. G. Brown, J. Nassauer, and B. Low. 2011. Variations in Development of Exurban Residential Landscapes: Timing, Location, and Driving Forces. Journal of Land Use Science, 6 (1): 13–32.

Anselin, L., Syabri, I., and Kho, Y. 2006. GeoDa: An introduction to spatial data analysis. Geographical Analysis, 38(1):5–22.

Hägerstrand, T. 1970. What about people in regional science? Papers of the Regional Science Association, 24, 7–21.

Langran, G., and N. R. Chrisman. 1988. A framework for temporal geographic information. Cartographica 25:1–14.

Turner, M. G., R. H. Gardner, and R. V. O'Neill. 2001. Landscape Ecology in Theory and Practice: Pattern and Process. New York: Springer.

Urdal, H. 2006. A Clash of generations? Youth bulges and political violence. International Studies Quarterly 50: 607–629.

Pontius, R.G., Jr., E. Shusas, and M. McEachern. 2004. Detecting important categorical land changes while accounting for persistence. Agriculture, Ecosystems and Environment, 101 (2004): 251–268

Worboys, M. F. 1992. A model for spatio-temporal information. In Proceedings: The 5th International Symposium in Spatial Data Handling, 2, ed. P. Bresnahan, E. Corwin, and D. Cowen, 602–11. San Jose, CA: American 1600 Congress of Surveying and Mapping.

Ye, X. 2009. Comparative space time dynamics (dissertation). Joint doctoral program between San Diego State University and University of California, Santa Barbara.

ACKNOWLEDGEMENT

This analysis is built upon the work and output from Tsou (GIS and cartography), Gupta (political science and terrorism research), Spitzberg (Communication), and Marc Gawron (Linguistics). SDSU Geography doctoral student Sarah Wandersee has played a critical role in performing space-time analysis as mentioned above.