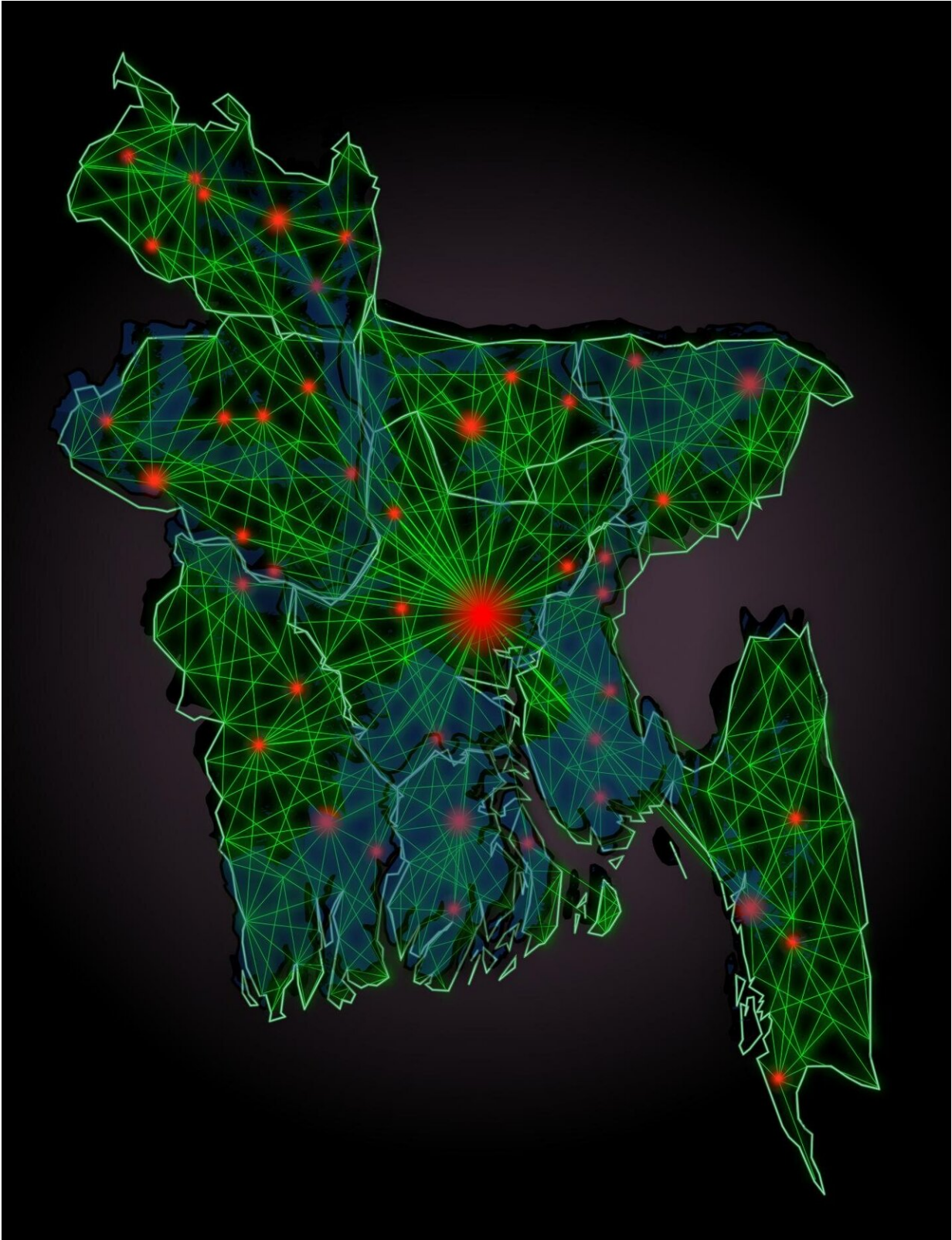


Novel method for measuring spatial dependencies turns less data into more data

October 21 2020



Depiction of interdependencies in human migration away from flood zones to

inland cities in Bangladesh. Credit: Maurizio Porfiri, Ph.D.

The identification of human migration driven by climate change, the spread of COVID-19, agricultural trends, and socioeconomic problems in neighboring regions depends on data—the more complex the model, the more data is required to understand such spatially distributed phenomena. However, reliable data is often expensive and difficult to obtain, or too sparse to allow for accurate predictions.

Maurizio Porfiri, Institute Professor of mechanical and aerospace, biomedical, and civil and urban engineering and a member of the Center for Urban Science and Progress (CUSP) at the NYU Tandon School of Engineering, devised a novel solution based on network and information theory that makes "little data" act big through, the application of mathematical techniques normally used for time-series, to spatial processes.

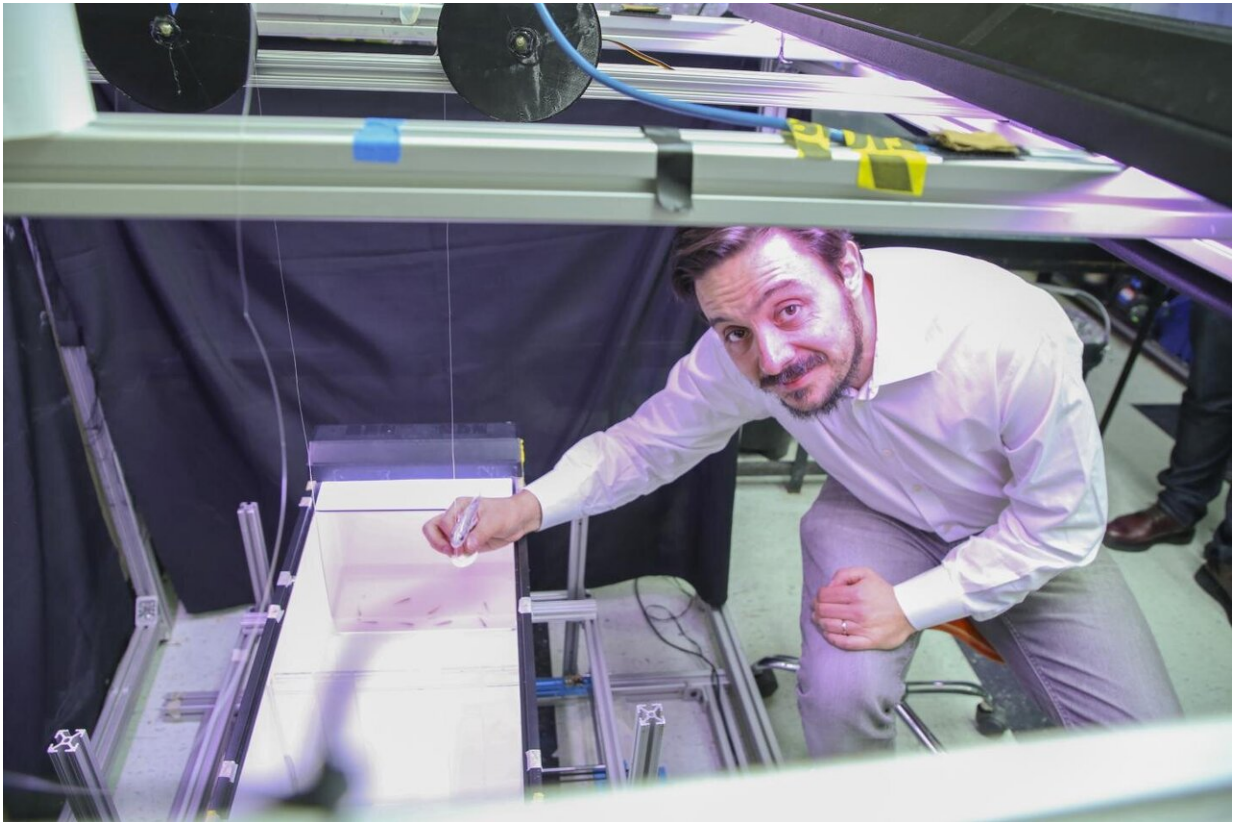
The study, "An information-theoretic approach to study spatial dependencies in small datasets," featured on the cover of *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, describes how, from a small sample of attributes in a limited number of locations, observers can make robust inferences of influences, including interpolations to intermediate areas or even distant regions that share similar key attributes.

"Most of the time the [data sets](#) are poor," Porfiri explained. "Therefore, we took a very basic approach, applying information theory to explore whether influence in the temporal sense could be extended to space, which allows us to work with a very small data set, between 25 and 50 observations," he said. "We are taking one snapshot of the data and drawing connections—not based on cause-and-effect, but on interaction

between the individual points—to see if there is some form of underlying, collective response in the system."

The method, developed by Porfiri and collaborator Manuel Ruiz Marín of the Department of Quantitative Methods, Law and Modern Languages, Technical University of Cartagena, Spain, involved:

- Consolidating a given data set into a small range of admissible symbols, similar to the way a machine learning system can identify a face with limited pixel data: a chin, cheekbones, forehead, etc.
- Applying an [information-theory](#) principle to create a test that is non-parametric (one that assumes no underlying model for the interaction between locations) to draw associations between events and to discover whether uncertainty at a particular location is reduced if one has knowledge about the uncertainty in another location.



Professor Maurizio Porfiri, at work in his lab at the NYU Tandon School of Engineering. Credit: NYU Tandon School of Engineering

Porfiri explained that since a non-parametric approach posits no underlying structure for the influences between nodes, it confers flexibility in how nodes can be associated, or even how the concept of a neighbor is defined.

"Because we abstract this concept of a neighbor, we can define it in the context of any quality that you like, for example, ideology. Ideologically, California can be a neighbor of New York, though they are not geographically co-located. They may share similar values."

The team validated the system against two [case studies](#): population

migrations in Bangladesh due to sea level rise and motor vehicle deaths in the U.S., to derive a statistically principled insight into the mechanisms of important socioeconomic problems.

"In the first case, we wanted to see if migration between locations could be predicted by geographic distance or the severity of the inundation of that particular district—whether knowledge of which district is close to another district or knowledge of the level of flooding will help predict the size of migration," said Ruiz Marín .

For the second case, they looked at the spatial distribution of alcohol-related automobile accidents in 1980, 1994, and 2009, comparing states with a high degree of such accidents to adjacent states and to states with similar legislative ideologies about drinking and driving.

"We discovered a stronger relationship between states sharing borders than between states sharing legislative ideologies pertaining to alcohol consumption and driving."

Next, Porfiri and Ruiz Marín are planning to extend their method to the analysis of spatio-temporal processes, such as gun violence in the U.S.—a major research project recently funded by the National Science Foundation's LEAP HI program—or epileptic seizures in the brain. Their work could help understand when and where gun violence can happen or seizures may initiate.

More information: Maurizio Porfiri et al, An information-theoretic approach to study spatial dependencies in small datasets, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* (2020). [DOI: 10.1098/rspa.2020.0113](https://doi.org/10.1098/rspa.2020.0113)

Provided by NYU Tandon School of Engineering

Citation: Novel method for measuring spatial dependencies turns less data into more data (2020, October 21) retrieved 2 May 2024 from <https://phys.org/news/2020-10-method-spatial.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.