UC San Diego News Center

April 30, 2014 | By Susan Brown

Phone Calls Reveal Population Patterns that Censuses Miss

It's no surprise that official tallies of the inhabitants of a city miss many residents. Not everyone wishes to be counted, particularly not those who lack legal permission to be there, and censuses are infrequent, if they happen at all.

Yet a measure of human density and distribution could help urban planners, whose work would also benefit from knowledge of population on scales far finer than the size of a census tract — the size of a city block, for example.



A high resolution map of population in Milan in near real time, assembled using call data. Each square is 235 meters per side. Official population data is aggregated into larger districts, outlined in black. Click on image for larger view.

So when Telecom Italia, a major mobile phone service in Italy, made available records for 600

million calls placed in metropolitan Milan as part of a big data competition, a multidisciplinary team led by David Meyer, professor of mathematics at UC San Diego and research scholar with the University of California's Institute for Global Conflict and Cooperation stepped up.

"We wanted to paint a picture of immigrant populations in Milan," said team member Megha Ram, a research associate with IGCC. "Some calls to numbers outside the country are for business, of course," Ram notes, "But one source may be immigrants calling home."

Call data provided by Telecom Italia contained a wealth of information calls from Milan, which Meyer's team extracted to find patterns of human connection in a project they call (Dis)assembling Milan. Their entry made it into the top 10 of 652 submissions in the competition earlier this month.

The company made available a data set containing information such as the country code of numbers called localized to 10,000 "grid squares," 235 meters per side, about the size of a city block. Two months of call data was aggregated into 10-minute bins. This made for fine

resolution in time and space, and also for a large amount of information to sort through.

"These data sets can be structured strangely," said Rex Douglass, a post-doctoral researcher with a background in political science and a quantitative bent. "It takes some imagination and creativity to come up with good ways to extract information."

The team's first step was to associate shear numbers of calls with the size of the population. "We wanted to construct a real-time population census at the city block level," Douglass said.

They aggregated call data for the block-sized grids into 68 of Milan's registration districts and looked for measures that corresponded with recorded population numbers, a strategy called machine learning.

"The idea is to find patterns in complex data even when we don't have good theoretical expectations to guide us," Douglass said.

The number of calls placed between 10 and 11 each morning to locations outside of Milan best predicted population in each census tract, they found, and that correlation held at larger scales as well, lending confidence to the measure when mapped to the far finer scale of 10,000 block-sized grids.

Large numbers of Egyptians live in Milan, but their numbers are uncertain. Estimates by the Egyptian consulate are about twice the numbers recorded by Italian officials, a discrepancy that likely results from many immigrants who arrived through unofficial channels, a status Italians call "irregular."

Meyer's team used a similar strategy to locate and count this subpopulation by analyzing patterns of calls from Milan to Egypt. They found that calls during dinnertime in Egypt were the best predictor of this population.

Using big data in this way could raise concern, but benefits in this case could outweight potential harm, the team said. Italy has a history of integrating immigrant populations through periodic amnesty programs that allow people to switch their immigration status to "regular." And organizations that provide social services and health care, and work to protect the human rights would like to identify neighborhoods where their help is most needed.

Calls from one grid to another provide a measure of human connections within Milan, "the practical geometry of a city," Meyer said. They found Milanese community to be more closely tied along the radial streets than around ring roads, a community structure better represented

by a map projected onto a hyperbolic manifold than a flat plane.

Meyer's group joins experts in data and analysis with researchers who specialize in social and political issues to address these kinds of questions, creating a unique sort of synergy.

"It allows us to explore questions in a completely different way," said Ram, who holds a degree in international studies with a concentration in economics. "I've gained a new perspective on quantitative methods in the field of political science that is expanding my thoughts about what's possible."

David Rideout, a project scientist in mathematics and Dongjin Song, a graduate student in electrical and computer engineering, completed the team for this competition.

MEDIA CONTACT

Susan Brown, 858-246-0161, sdbrown@ucsd.edu

UC San Diego's <u>Studio Ten 300</u> offers radio and television connections for media interviews with our faculty, which can be coordinated via <u>studio@ucsd.edu</u>. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <u>https://ucsdnews.ucsd.edu/media-resources/faculty-experts</u>.