Money on the Move: Big Data of Bank Card Transactions as the New Proxy for Human Mobility Patterns and Regional Delineation. The Case of Residents and Foreign Visitors in Spain

Stanislav Sobolevsky
SENSEable City Lab, MIT
Cambridge, MA, USA
Email: stanly@mit.edu

Izabela Sitko
Department of Geoinformatics - Z_GIS,
University of Salzburg
Salzburg, Austria
Email: izabela.sitko@sbg.ac.at

Remi Tachet des Combes
SENSEable City Lab, MIT
Cambridge, MA, USA
Email: rtachet@mit.edu

Bartosz Hawelka
Department of Geoinformatics - Z_GIS,
University of Salzburg
Salzburg, Austria
Email: hawelka.bartosz@sbg.ac.at

Juan Murillo Arias
New Technologies, BBVA
Madrid, Spain
Email: juan.murillo.arias@bbva.com

Carlo Ratti
SENSEable City Lab, MIT
Cambridge, MA, USA
Email: ratti@mit.edu

***Abstract*—Increasing availability of big data, which documents human activity in space and time, offers new solutions to well-known operational problems. Recent studies have demonstrated how topological community detection in large-scale networks of human interactions and mobility can produce geographically cohesive regions, which are meaningful for the regional division of countries. So far, those networks have mainly been built based on the country-wide datasets of telephone calls, typically available for residents of a country. However, it is natural to expect that foreign visitors explore a country in a different way, with patterns that vary depending on nationality. Understanding those differences can be of a great importance for the touristic industry and transportation planning. In this study, we demonstrate the potential of a new type of extensive data, namely bank card transactions executed in a variety of businesses by the domestic and foreign customers of a Spanish bank. We confirm applicability of this data to the regional delineation inline with other datasets and reveal new opportunities related to the distinction of customers by their origin. We point out the important differences between the optimal regional structure derived from the mobility of residents and of foreign visitors. The definition of the mobility network appears to be a crucial component of the methodology, and is potentially sensitive to the dataset being used. While a reasonable comparison of results obtained based on different data of course requires the consistency of such definition. We propose a novel, consistent way of constructing mobility networks using transactional data; a way transposable to a variety of other datasets. Finally, we perform a quantitative study of the impact of tourists' nationality on their mobility behavior. We find a surprisingly consistent trend between the distance from a given country to Spain, and the mobility characteristics of visitors coming from this country, i.e. the parameters of the gravity model estimation.**

***Keywords*-big data; bank card transactions; regional delineation; human mobility; network analysis**

I. Introduction

Nowadays, more and more aspects of human activity include some sort of interaction with digital technologies. Those interactions result in the increasing production of data, which can be interpreted and explored as digital traces of human behavior. Discovery of patterns occurring in such data creates tremendous new opportunities in the fields of geography, urban planning, as well as other areas of social science, which are experiencing flourishing developments due to the availability of large-scale datasets reflecting human interactions, communications, and mobility.

Many patterns in human behavior were discovered through mobile call records [1]–[5], social media posts [6]–[9], vehicle GPS traces [10], [11] or smart cards usage [12], [13]. For instance, it has been shown that community detection based on a country-wide network of cell phone communications creates spatially continuous and cohesive regions, similar in shape and number to the existing administrative divisions of the country [14]–[18]. This methodology was additionnally proven to be valid for human mobility networks extracted by means of call records [19], geo-referenced tweets [9], and transportation systems data [10] (such as taxi GPS routes).

In this paper, we explore a new type of data, giving insight on human economic activity: individual transactions performed via bank cards. This country-wide dataset, provided by the second largest Spanish bank BBVA, contains anonymized records of all bank card transactions executed by its customers (in any Spanish business) and of all transactions performed on BBVA card terminals, including by visitors from other countries. Unlike phone calls, this data does not allow to measure direct human-to-human

978-1-4799-5057-7/14 $31.00 © 2014 IEEE
DOI 10.1109/BigData.Congress.2014.28

interactions over space. However, a network can still be built between different locations that were co-visited by the same consumers bringing their money from one location to another, in a similar manner as communication flows or origin-destinations of trips were used before to connect locations across the country. This network can be also analyzed from the perspective of customer citizenship - distinguishing Spanish and foreign customers reveals specific patterns of country exploration by different nationalities.

Similarly to the recent work [18], community detection is performed within the general framework of modularity optimization approach [20], using the particularly efficient method introduced in [21]. We first validate that the partitioning created with the newly-defined networks shares the same general properties as the partitioning based on other types of data. The novel approach of constructing mobility networks from transactional data is presented in a consistent way for different subsets of the data, such as transactions performed by domestic or foreign customers. Next, the optimal regional structures obtained with the mobility networks of domestic customers and foreign visitors are compared and discussed. Finally, we present how the mobility of customers can be approximated with a gravity model [22], [23]. We discuss the differences in exponents noted for visitors of different nationalities, as well as comment the values for all visitors and Spanish residents, trying to understand their actual meaning and context.

II. The dataset

Our study relies on the complete set of bank card transactions recorded by Banco Bilbao Vizcaya Argentaria (BBVA) during 2011, all over Spain. Transactions were performed by two groups of card users. The first one consists of direct customers, residents of Spain, who hold a debit or credit card issued by BBVA. In 2011, the total number of such active customers was around 4.5 M, altogether they executed more than 178 M transactions in over 1.2 M points of sale, which resulted in a cumulative spending exceeding 10 billion euros. The second group of card users includes over 34.6M customers of other banks who used one of the approximately 300 thousand BBVA card terminals (in all types of Spanish businesses). Those customers come from both Spain and abroad - representing visitors from 177 countries. In total, they executed another 166 M transactions worth over 9 billion euros, while over 10.9M transactions were made by foreign customers.

Due to the sensitive nature of the subject, our dataset was anonymized by BBVA prior to sharing, in accordance to all local privacy protection laws and regulations. As a result, customers are identified by randomly generated IDs, connected with an indication of their residence at the spatially aggregated level - one of 368 *comarca* units (see description below) for direct customers of BBVA and the country of residence for all others. Of course such a coarse-grained level of spatial aggregation does not reveal any specific details about customer residence that might threaten his or her anonymity. Each transaction is characterized with its value, a time stamp, and additionally the retail location where it was performed.

The raw dataset is protected by the appropriate non-disclosure agreement and is not publicly available. However, the researches may share certain aggregated data upon request and for the purpose of paper findings validation.

From the spatial perspective, the study refers to three different levels of Spanish regional units. The smallest one is *comarca*, a unit that groups municipalities around one central town that offers shared services, such as justice courts, for a total population of around 125,000 inhabitants at least. The 368 comarcas serve as the nodes for all network definitions proposed in this paper. From the technical stand point, they offer an optimal balance between fine spatial granularity and a possible sparseness of the data used to construct the network. Two other spatial levels - 17 autonomous communities and 52 provinces that reflect the official territorial division of Spain, are used for comparative purpose. All relevant boundaries were obtained from the Global Administrative Areas spatial database, version 2[1]. Additionally, we utilized country-level statistics from the World Bank database[2], which were considered as factors that can potentially affect the mobility of different nationalities visiting Spain.

III. The partitioning method

From the methodological perspective, this study takes advantage of a standard community detection approach based on modularity optimization [20], [24]. The modularity measure relies on the scores of all edges of the network, calculated according to their relative strength and with respect to weights of the nodes that they connect. In particular, it aggregates the scores associated to the links between nodes that belong to the same communities. In this way, modularity optimization procedure fosters the connections of nodes with positive score of associated edges, and avoids connections for edges with negative score.

The particular optimization algorithm [21] is a novel high-performance technique based on an iterative combination of splits, merges and multi-node shifts from one community to another. The suggested algorithm was found to outperform the known state-of-the-art modularity optimization algorithms, in most cases producing the best modularity score on a variety of benchmark networks. For this reason, we use it here to ensure that we obtain an optimal partitioning as far as the modularity quality function is concerned.

[1] www.gadm.org
[2] data.worldbank.org

IV. Partitioning of the country-wide network of money flows for Spanish residents

We start by exploring the expenses of the Spanish BBVA customers. The network is defined in a way similar to the approach described in [14], where a link between each pair of locations across the country is created every time a person from one location called someone from the other, while the strength of the link depended on the call duration. Analogously in our case, each time a customer uses his or her bank card to make a transaction we create an edge between the customer's residence location (one of the 368 comarcas) and the region where the transaction was executed. The weight of the link is equal to the amount of the purchase. This approach allows to build a country-wide network of money flow, reflecting how money is circulating between customers and businesses locations.

Figure 1 shows the communities obtained via partitioning of the aforementioned network. The picture generally confirms validity of the approach. All the elements appear as spatially cohesive regions, similar in shape and size to the official autonomous communities of Spain. The number of communities we got - 19 - is also very close to the official 17 units. Three regions, Galicia, Canary and Balearic Islands, keep the exact boundaries of corresponding autonomous communities. This can be interpreted through their geographically peripheral location that is reflected by human economic behavior. There are also other regions, such as Cantabria or Extremadura, which are not delineated in perfect agreement, yet very closely with the official borders. A good alignment of the transactions-based communities can be observed at the provincial level, where almost all administrative borderlines are followed, with only few exceptions.

At the same time however, certain discrepancies can be recognized and explained by different factors. For instance, historical links between particular regions resulted in increased commercial ties between Asturias and the province of León, and similarly between Murcia and Alicante. On the contrary, historical barriers within currently united regions have been preserved in the human utilization of space, resulting in the split of Andalusia or Catalonia into separate communities. Interestingly, regions such as Western and Eastern Andalusia are still recognizable in the territorial structures of private organizations from this part of Spain. Regardless of the internal split, the outer borders of those large regions are preserved. Another important factor reflected in the spatial alignment of the communities is the influence of the main population center. This is particularly clear for Madrid - the detected community appears much bigger than its official spatial extent, which seems a logical effect of a capital city attracting more people and from farther away. In a similar way, we can understand the connection of Navarra, La Rioja, and the eastern part of Cantabria to the more prosperous region of Basque Country. Finally, we also consider physical factors such as proximity and orography, which influence the accessibility of certain service nodes. Communication issues may explain the separation of Dénia and Jávea from Alicante, their administrative capital, to the advantage of the link with Valencia. Location within the same valley reinforces the bond between the western part of the Toledo province and Cáceres. In a similar way the province of Huesca gravitates toward West Catalonia, as its main urban node (Frada) is located much closer to the city of Lérida (around 30km), than to the administrative capital of Zaragoza (around 100km).

A general observation stemming from the partitioning of the money flow network is the fact that human activity, as it was already proven with the cell phone data, tends to respect the existing official borders, however is also able to reveal important patterns where deviations appear. Most of them can be logically explained by either historical ties between regions, or the convenience of travels fostered by physical parameters of accessibility.

V. Partitioning of the country-wide mobility networks for Spanish residents and foreign visitors

Let us now compare the optimal regional structure discovered through the activity of domestic customers with the one created by foreign visitors. The case of foreigners is intricate, because we do not know their exact home location, which in any case lays outside of the country. Here for the sake of consistency, for both domestic and foreign customers, we consider an alternative definition of the mobility network independent from the customer residence location. The link between two locations a and b is defined as the probability that for a randomly selected pair of transactions made by the same customer, first transaction was made in location a and second in location b. More specifically, if the number of transactions made by customer c in location a is denoted by $t(c,a)$, then the link between locations a and b is defined as

$$link(a,b) = \sum_c \frac{t(c,a)\,(t(c,b) - \delta(a,b))}{T\,(t(c)-1)}, \qquad (1)$$

where $t(c) = \sum_a t(c,a)$, $T = \sum_c t(c)$ and $\delta(a,b)$ is the Kronecker delta being 1 if $a = b$ and 0 otherwise.

This approach creates a normalized network (the total weight of all links is 1) consistently defined for any spatial resolution. In the following, we again consider 368 comarcas as the set of network nodes.

Partitioning of the mobility networks for both domestic and foreign customers is shown in figure 2. As one can see, the result obtained for domestic customers is pretty similar to the communities built from money flows on figure 1. Certain aspects of regional borders are even captured in a more consistent way, e.g. borders of Extremadura and Aragon are now identical with the administrative units.

Figure 1. Optimal partitioning of money flow network based on the spending of domestic customers. The 19 communities are differentiated by color, while solid lines represent boundaries of the 17 official autonomous communities of Spain, dashed lines indicate provinces.

When it comes to the mobility network of foreign visitors however, the outcome is substantially different. First of all, the whole north-west region of Spain, less attractive from the purely touristic perspective, is merged into one big community. Communities detected in the remaining part of the country align according to the visitors' interests. For instance, we observe coastline-connected communities in southern Catalonia or Andalusia, as well as the merged region of Murcia and Valencia. Balearic Islands are, interestingly, split into two communities, which may reflect differences between the tourism taking place in Menorca and the one taking place in Mallorca, or perhaps simply accessibility of one island from the other. Finally, big cities, such as Barcelona or Sevilla, constitute separate communities, and the region of Madrid, attractive from the global perspective as a capital, becomes even bigger than the one delineated based on the mobility of Spaniards.

Following the approach of [18], it is also relevant to consider the optimal bi-partitioning of the above networks using the same algorithm but with additional limitation on the resulting number of communities. Interestingly, the bi-partitioning for domestic customer mobility reproduces the north-south division of the country (figure 3), well-known from the pattern of many socioeconomic parameters such as unemployment, mortality or educational indicators as measured by the National Statistics Institute of Spain[3]. The only exception is Madrid, being statistically connected to the more prosperous north.

Mobility of foreign visitors indicates a completely different division of Spain, where Mediterranean east, including Barcelona and Balearic Islands, is separated from the rest of the country. This picture can be interpreted through different travel motivations. For the Mediterranean arch, the usual driver is sun and beach leisure offers, whereas the inner country attracts different profile of visitors pursuing more diverse objectives. The affiliation of Badajoz (autonomous community of Extremadura) and Navarre with the eastern touristic region requires a more thorough investigation. One may suppose, that it reflects the frontier location of both areas, and may possibly result from the activity of visitors from Portugal and France.

[3]www.ine.es

VI. Modeling the mobility networks for Spanish residents and foreign visitors

It is well known that human activity over space, such as mobility or communication intensity, decays with distance. The average dependence is often approximated by a gravity law [22], [23], [25], [26]. Hence, it is natural to assume that mobility extracted from individual economic transactions is not an exception. Demonstrating that gravity models can be applied to our framework would lead to an additional confirmation of the consistency of the network definition.

In this analysis, we only consider the mainland of the country, as distance across sea cannot be expected to have a consistent impact with distance over land. For each pair of locations $a \neq b$, we model it with the following gravity law, in its most general form:

$$link(a,b) = C\frac{(w(a))^p\,(w(b))^q}{(dist(a,b))^d}, \qquad (2)$$

where $dist(a,b)$ is the geographical distance between two nodes and w denotes the weights of different locations. Those can be estimated based on the population of the place or alternatively on its total economic activity. We choose the latter solution since the networks we are going to study are not directly related to residential population. This way we define $w(a)$ as the total number of transactions performed by all the customers in the retail points of location a. The model

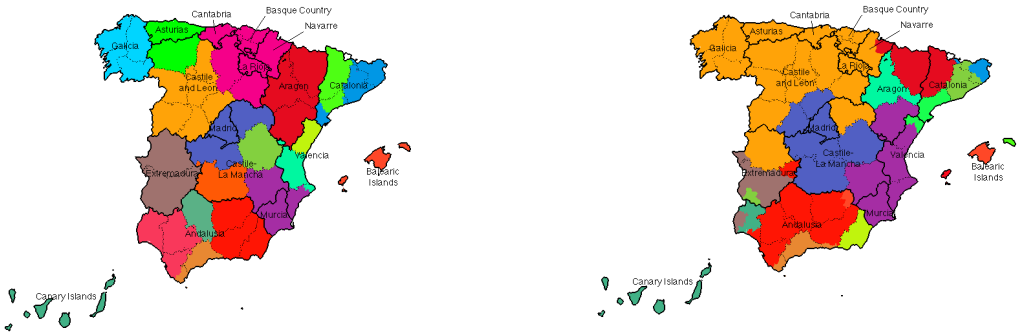

Figure 2. Optimal partitioning of mobility networks of domestic customers (on the left, 20 communities) and foreign visitors (on the right, 17 communities).

Figure 3. Optimal bi-partitioning of mobility networks based on domestic customers (on the left) and foreign visitors (on the right).

predicts all links of the network based on the weights of origin and destination and the distance between the two, with the exception of loop edges that reflect the probability for a customer to make a transaction within the same location.

However, before applying the model, one needs to define the exponents p, q, d as well as the normalization constant C. In the present analysis, those parameters are fitted from the actual networks constructed with our data, and used to characterize the customer mobility. For both networks - reflecting mobility of domestic and foreign customers - the model demonstrates a decent fit (figure 4) with the optimal values of the exponents estimated as $p = q = 0.9778$ and $d = 0.7640$ for foreigners and $p = q = 0.8022$, $d = 1.2111$ for domestic customers. Those values suggest pretty fast decay of the number of economic connections with distance. Additionally, the size of both origin and destination has an underlinear impact on the strength of the link.

Keeping those basic remarks aside, values for foreign and domestic customers exhibit substantial differences. First of all, the distance exponent for domestic customers is much higher than the one for foreign visitors. This can be interpreted as foreign visitors undertaking more frequent long travels within the country, while the mobility of Spaniards is more local. Since foreign visitors are already coming from far away, and do not have any local ties within the country, they are more inclined to visiting a variety of distant places across Spain. On the contrary, domestic customers spend most of their time within their primary local area. Another interesting pattern is revealed by the exponents attached to the origin and destination weights. The ones obtained for foreign visitors are higher, which means that they are relatively more attracted by major destinations in the country (perhaps they just do no know about other options), while domestic people mobility is more spread between small and large destinations.

VII. Impact of country of residence on foreign visitor mobility

In the previous sections, a substantial difference in mobility of domestic and foreign customers was found. Let us now explore more thoroughly the impact of the country of origin, in particular verifying if inhabitants of closer countries exhibit behavioral patterns more similar to those of the Spanish residents. On figure 5, one can see that it is indeed the case. Although the optimal gravity model parameters for the mobility of visitors change a lot from one

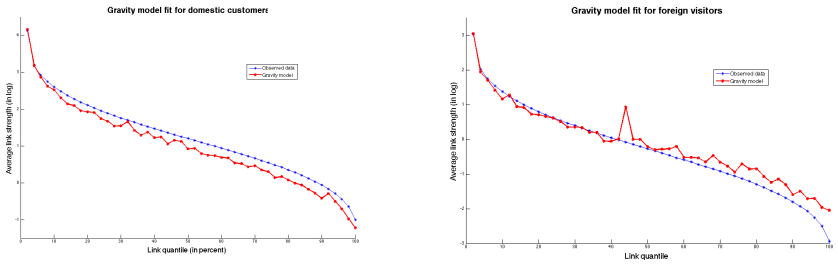

Figure 4. Fit of the gravity model estimation for domestic customers (on the left) and foreign visitors (on the right)

Table I
STATISTICAL RELIABILITY OF THE TRENDS

	Distance exponent			
	Slope	p-value	slope 95% CI	R^2
Distance to Spain	-0.494	1.6E-10%	[-0.608, -0.380]	52.25%
Population	-0.139	0.024%	[-0.211, -0.067]	18.35%
GDP	0.145	3.15%	[0.013, 0.277]	6.82%

	Weight exponent			
	Slope	p-value	slope 95% CI	R^2
Distance to Spain	0.45	2.9E-14%	[0.366, 0.534]	62.84%
Population	0.119	0.016%	[0.059, 0.178]	19.24%
GDP	-0.229	0.0016%	[-0.328, -0.131]	24,80%

country to another, the general tendency is well recognizable and similar to a linear trend for both distance and weight exponents versus the logarithm of distance from the country of origin to Spain. Linear regression for both types of exponents with the logarithm of distance indicates a sharp slope for both trends (table I). The corresponding p-values, obtained with the standard F-test, are far below 1%, which is generally accepted as a very strong statistical evidence of a trend [27]. Values of the coefficient of determination R^2 show that over 50% of the exponents variability can be attributed just to the distance to the country of origin.

Overall, we observe that the distance exponents substantially decay with the distance from origin, while weight exponents increase. The average distance and weight exponent values for foreign visitors from nearby are indeed close to the mobility parameters of local inhabitants. Those observations might be interpreted as follows: the farther away people come from, the larger the scale of their mobility within the country is and the more attracted to larger destinations they are. On the contrary, people from nearby countries explore smaller destinations and remain more often confined to smaller areas.

On figure 6, we also analyze the dependence of gravity model exponents on other important parameters of the country of origin such as size, reflected by its population, and economic prosperity, measured by its Gross Domestic Product per capita (GDP). In general, the bigger the country of origin, the smaller the distance exponent and the higher the weight exponent. This means that people coming from bigger countries also tend to travel more globally across Spain and are more attracted to larger destinations. As far as a GDP is concerned, it seems like the wealthier the country of origin is, the more local the mobility of visitors becomes (exponent is higher) and the more attractive smaller destinations are. However, while interpreting the country size and GDP impact, one should keep in mind that those two parameters show medium-level inverse linear correlation (with a correlation coefficient of -0.3932). Therefore the visible impact of one of the two might simply be an indirect consequence of the impact of the other.

As one can see from table I, the trends for country size and GDP are also statistically significant, as most of them are characterized by a $p-value$ below 1% , with the only exception of the relation between distance exponent and the logarithm of GDP with a $p-value$ of 3.15%, which is still generally accepted as strong evidence. However, the coefficient of determination R^2 is already much lower compared to the case of distance to the country of origin. Thus, both wealth and population size of the country might also matter in touristic mobility patterns, but remoteness seems to have the strongest impact.

VIII. CONCLUSIONS

Introducing the new dataset of digital traces of human individual economic activity through bank cards transactions in Spain, we first of all validated that this dataset possesses a significant potential for regional delineation, in a way similar to the mobile phone call records or transportation traces. More specifically, we analyzed two ways of defining the network between different pairs of locations across the country: a first, standard one, based on money flows and a second, novel, probabilistic approach. The community detection performed on such networks, relying exclusively on their topological properties without any pre-imposed conditions or spatial constraints, resulted in geographically cohesive and spatially connected communities of locations, similar in shape and size to the existing regions of the official autonomous communities.

In this way, our findings confirm that the predominantly local character of the economic activity of domestic population is well aligned with the existing administrative divisions. However, certain discrepancies exist and reveal

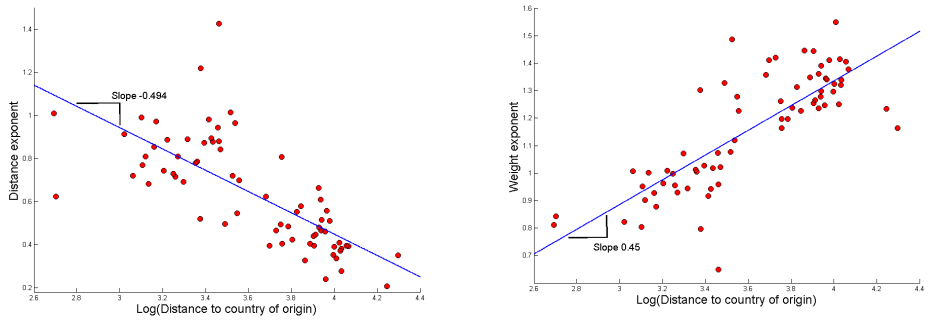

Figure 5. Parameters of the gravity model estimated for foreign visitors against distance from the country of origin

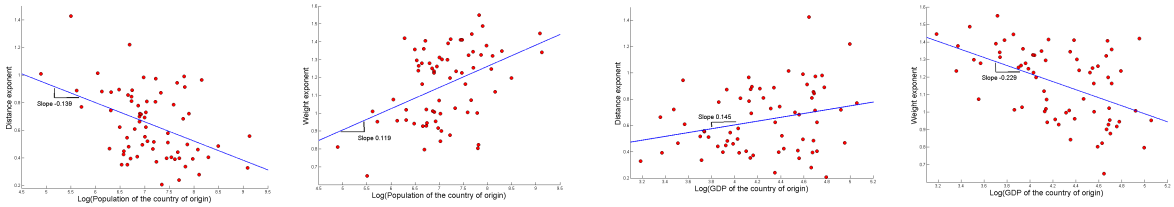

Figure 6. Parameters of the gravity model estimated for foreign visitors depending on the size (first and second plots from the left) and on the GDP per capita (third and fourth plots) of their country of origin.

important patterns. Those deviations are easily attributable to the additional factors affecting purchasing behavior such as: historical links or barriers among regions, which were discarded officially, yet often maintained in the collective perception; influence of important major urban centers, attracting inhabitants regardless of the administrative divisions; physical determinants of accessibility, such as proximity and orography. Importantly, the discovered regionalization remains mostly stable across both types of network definition.

Furthermore, as the new dataset allows to distinguish economic activity of domestic consumers from the one of foreign visitors, we constructed a separate mobility network for those visitors and compared regional patterns obtained through the partitioning procedure. Discovered differences are meaningful on both scales: the fine-grained partitioning into the optimal number of communities and the optimal bi-partitioning of the country. In the latter case, we see the Spaniards being strongly influenced by the north-south socioeconomic distinction, while mobility of foreign visitors is highly affected by their touristic profile, dividing the country into the Mediterranean arch and the interior.

Finally, we zoomed into the mobility of foreign visitors considering the impact of the country of origin. Being characterized by the optimal gravity model exponents, mobility of the foreigners demonstrates a surprisingly strong dependence on the distance to their origin. This trend indicates that the farther away visitors come from, the more large-scale their mobility inside the visited country is, and the more attracted to larger destinations they are. On the contrary, people from nearby countries show more local mobility patterns, exploring smaller destinations more easily. From that standpoint, their behavior is similar to the behavior of locals.

Overall, findings of the paper suggest substantial value of the new big dataset of bank card transactions for studies of various aspects of domestic and foreign people behavior inside the country. We can also conclude that the data possesses a unique potential to support policy decisions at the regional level, with its direct applicability to be further investigated and verified in detail.

ACKNOWLEDGMENT

The authors would like to thank the Banco Bilbao Vizcaya Argentaria (BBVA) for providing the dataset for this research. Special thanks to Assaf Biderman, Marco Bressan, Elena Alfaro Martinez and María Hernández Rubio for organizational support of the project and stimulating discussions. Furthermore, we thank the BBVA, Ericsson, MIT SMART Program, Center for Complex Engineering Systems (CCES) at KACST and MIT, the National Science Foundation,the MIT Portugal Program, the AT&T Foundation, Audi Volkswagen, The Coca Cola Company, Expo 2015, Ferrovial, The Regional Municipality of Wood Buffalo and all the members of the MIT Senseable City Lab Consortium for supporting the research. This research was also supported by the Austrian Science Fund (FWF) through the Doctoral College GIScience (DK W 1237-N23).

REFERENCES

[1] C. Ratti, S. Williams, D. Frenchman, and R. Pulselli, "Mobile landscapes: Using location data from cell phones for urban analysis," *Environment and planning B*, vol. 33, no. 5, p. 727, 2006.

[2] F. Calabrese and C. Ratti, "Real Time Rome," *Networks and Communication studies*, vol. 20, no. 3-4, pp. 247–258, 2006.

[3] F. Girardin, F. Calabrese, F. D. Fiore, C. Ratti, and J. Blat, "Digital footprinting: Uncovering tourists with user-generated content," *Pervasive Computing, IEEE*, vol. 7, no. 4, pp. 36–43, 2008.

[4] M. González, C. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns," *Nature*, vol. 453, pp. 779–782, 2008.

[5] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft, "Recommending social events from mobile phone location data," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, 2010, pp. 971 –976.

[6] A. Java, X. Song, T. Finin, and B. Tseng, "Why weTwitter: understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 2007, pp. 56–65.

[7] M. Szell, S. Grauwin, and C. Ratti, "Contraction of online response to major events," *arXiv preprint arXiv:1308.5190*, 2013.

[8] M. R. Frank, L. Mitchell, P. S. Dodds, and C. M. Danforth, "Happiness and the patterns of life: A study of geolocated tweets," *Scientific Reports*, vol. 3, 2013.

[9] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, "Geo-located Twitter as the proxy for global mobility patterns," *Cartography and Geographic Information Science*, 2014.

[10] C. Kang, S. Sobolevsky, Y. Liu, and C. Ratti, "Exploring human movements in Singapore: a comparative analysis based on mobile phone and taxicab usages," in *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. ACM, 2013, p. 1.

[11] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. Strogatz, and C. Ratti, "Taxi pooling in New York City: a network-based approach to social sharing problems," *arXiv preprint arXiv:1310.2963*, 2013.

[12] M. Bagchi and P. White, "The potential of public transport smart card data," *Transport Policy*, vol. 12, no. 5, pp. 464–474, 2005.

[13] N. Lathia, D. Quercia, and J. Crowcroft, "The hidden image of the city: Sensing community well-being from urban mobility," in *Pervasive Computing*, ser. Lecture Notes in Computer Science, J. Kay, P. Lukowicz, H. Tokuda, P. Olivier, and A. Krüger, Eds., 2012, vol. 7319, pp. 91–98.

[14] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, and S. H. Strogatz, "Redrawing the map of Great Britain from a network of human interactions," *PLoS One*, vol. 5, no. 12, p. e14248, 2010.

[15] V. Blondel, G. Krings, and I. Thomas, "Regions and borders of mobile telephony in Belgium and in the Brussels metropolitan zone," *Brussels Studies*, vol. 42, no. 4, 2010.

[16] V. Blondel, P. Deville, F. Morlot, Z. Smoreda, P. Van Dooren, and C. Ziemlicki. (2011, 11) Voice on the border: Do cellphones redraw the maps?

[17] F. Calabrese, D. Dahlem, A. Gerber, D. Paul, X. Chen, J. Rowland, C. Rath, and C. Ratti, "The connected states of America: Quantifying social radii of influence," in *Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on Social Computing (SocialCom)*. IEEE, 2011, pp. 223–230.

[18] S. Sobolevsky, M. Szell, R. Campari, T. Couronné, Z. Smoreda, and C. Ratti, "Delineating geographical regions with networks of human interactions in an extensive set of countries," *PloS ONE*, vol. 8, no. 12, p. e81707, 2013.

[19] A. Amini, C. Kang, K. Kung, S. Sobolevsky, and C. Ratti, "The impact of social segregation on human mobility in developing and urbanized regions," *arXiv preprint arXiv:1401.5743*, 2014.

[20] M. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[21] S. Sobolevsky, R. Campari, A. Belyi, and C. Ratti, "A general optimization technique for high quality community detection in complex networks," *arXiv preprint arXiv:1308.3508*, 2013.

[22] G. Krings, F. Calabrese, C. Ratti, and V. D. Blondel, "Urban gravity: a model for inter-city telecommunication flows," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 07, p. L07003, 2009.

[23] W.-S. Jung, F. Wang, and H. E. Stanley, "Gravity model in the Korean highway," *EPL (Europhysics Letters)*, vol. 81, no. 4, p. 48005, 2008.

[24] S. Fortunato, "Community detection in graphs," *Physics Report*, vol. 486, pp. 75–174, 2010.

[25] R. Lambiotte, V. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren, "Geographical dispersal of mobile communication networks," *Physica A*, vol. 387, no. 21, pp. 5317–5325, 2008.

[26] P. Expert, T. Evans, V. Blondel, and R. Lambiotte, "Uncovering space-independent communities in spatial networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 19, p. 7663, 2011.

[27] L. Wasserman, *All of statistics: a concise course in statistical inference*. Springer, 2004.