Power Centrality as a Relational measure of Urban Hierarchy.
Testing the Splintering Urbanism Theory with Social Media data from Santiago de Chile.

By

Francisco J. Humeres M.

Architect and Master in Urban Design
Pontificia Universidad Católica de Chile
Santiago, Chile (2011)

Submitted to the Department of Urban Studies and Planning
in partial fulfillment of the requirements for the degree of

Master in City Planning

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2014

Author____
Department of Urban Studies and Planning
October 9, 2013

Certified by ____
Associate Professor P. Christopher Zegras
Department of Urban Studies and Planning
Thesis Supervisor

Accepted by____
Associate Professor P. Christopher Zegras
Chair, MCP Committee
Department of Urban Studies and Planning

Power Centrality as a Relational measure of Urban Hierarchy. Testing the Splintering Urbanism Theory with Social Media data from Santiago de Chile.

by

Francisco J. Humeres M.

Submitted to the Department of Urban Studies and Planning on October 3, 2013 in Partial Fulfilment of the Requirements for the Degree of Master in City Planning

ABSTRACT

Power Centrality, a measure of node importance within a network, is borrowed from the field of Social Network Analysis and applied to the assessment of Urban Hierarchy. Based on the overlaps of human activity between places, Power Centrality is tried as a method for measuring a particular feedback property: How well connected are places to other well connected places.

In this research Power Centrality is used to assess a recent model of Urban Structure: The Splintering Urbanism Theory of Graham and Marvin (2001). This theory posits that the contemporary city is a fragmented agglomeration of isolated urban pieces, where distant but valuable fragments are highly connected between them, bypassing their less valuable surroundings. The causal explanation provided by Graham and Marvin is centered on their concept of premium networks: Networks customized for valuable (users in terms of income or power). The reach of this theory is assessed by studying the case of a mass transit system in a developing country: The Metro or subway of Santiago de Chile.

The spatial hypothesis of Graham and Marvin is tested empirically through the use of the Power Centrality Measure, applied to a dataset of 242.000 twitter statuses generated by Metro users, while the causal explanation is evaluated by comparing the results with an unbiased sample of 110.000 statuses. Power Centrality allowed the identification of central locations that by standard measures of spatial concentration would have remained undetected. Furthermore, the results evidenced how Metro could be acting as a mass public bypass that connects these emergent centralities, challenging the concept of premium networks posited by Graham and Marvin.

Thesis Advisor:

P. Christopher Zegras
Associate Professor of Transportation and Urban Planning
Department of Urban Studies and Planning

Thesis Reader:

Michael Szell
Post-Doctoral Fellow
Senseable City Lab

ACKNOWLEDGMENTS

That I am running out of time is not a mere metaphor. Unfortunate events that took place in my hometown thousands of kilometers away, as well as inconveniences here in Cambridge, have not allowed me to devote as I desired to this thesis. Therefore, I only have time for mentioning those who directly helped me to carry on with this task during one of the toughest months in my life. Time will judge if this thesis was finished in a good or bad shape, notwithstanding, I prefer to highlight the solidarity provided by the people mentioned here, that allowed me to finish this work.

For their hospitality here in Cambridge (and helping me to bear the stress of being a thesis wanderer): Flavio Sciaraffia, Juan Pablo Ugarte, Cyntia Barzelatto, Francisco Quintana, My Lam and Alexander Marks.

For their emotional and financial support: My aunt Tonia, and her husband Polo.

For helping me to deal with issues that only an International Student can understand: Faizan Jawed, Chia Yang-Weng, and Daniel Rosenberg.

For literally saving my life: My hard drive, and implicitly, my brother Hector, who gave it to me. His 90’s hacks inspired me to become an amateur programmer, skills that allowed me to access big amounts of data by my own.

I would also like to thanks professors Pablo Allard and Luis Valenzuela, they made me believe in something that seemed impossible: Get into this place. This includes the professional and personal support of the Garcia family, and the amateur therapists Juan Pablo and Andres, that helped me to overcome my fears. Thanks to Cesar A. Hidalgo for giving me a second chance here in Cambridge, and to DUSP faculty Lawrence Susskind and Dennis Frenchman for being supportive, open-minded and encouraging.

This work is dedicated to my buddies from Craighouse school, the MPUR pack, and my parents and siblings. My thoughts to my little niece, I hope one day she will cross the Atlantic and meet the Beaver.

Dedicated to the Memory of Teresa de Jesus Silva, widow of Marfan.
“Tata Tere”
R.I.P. (August 8, 2013)

INDEX

INTRODUCTION.

Power Centrality: A Relational Measure of Urban Hierarchy.

At the beginning of the 20th century, the Scottish biologist Patrick Geddes noticed how several belts of contiguous built areas were stretching through the industrial north of England. He coined a particular term to make reference to this emerging urban typology: Conurbations *(Geddes, 1997)*. These emerging conurbations complicated the characterization of urban organization, as markets, businesses, industries and housing were no more condensed within the limits of the urban core; Like never before, people had to travel during the day from home to work in the city *(Soja, 2000)*.

By that time it was becoming evident that despite their apparently organic growth, these emerging Metropolitan Areas did not have a random organization. Since the beginning of the 20th century, several models of Urban Structure have attempted to link observed land-use patterns with economic, social or cultural drivers. Traditionally, the focus has been on the concentration of employment or population (Figures 1.1-1.4), in order to establish an urban hierarchy according to the predominance (or absence) of these types. Thus, the key measure on which these models are based is the spatial concentration of uses or activities *(Anas, Arnott and Small, 1998)*.

The emergence of the contemporary polycentric Metropolis seems to pose conceptual challenges similar in nature to those addressed by Geddes. However, the measures employed by conventional Urban Structure models, such as land-value, retail, employment or population density, seem inapplicable in the contemporary urban context, where places combine diverse uses and demographics in a

Figure 1.1. Bid-rent theory of land use, derived by William Alonso (1964) from Von Thunnen's model of crop location (1820).

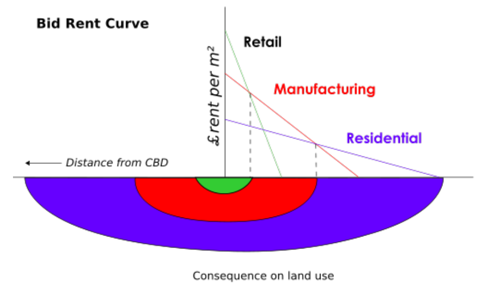


Figure 1.2. Walter Christaller's central place theory compared with the urban structure of southern Germany (1933).

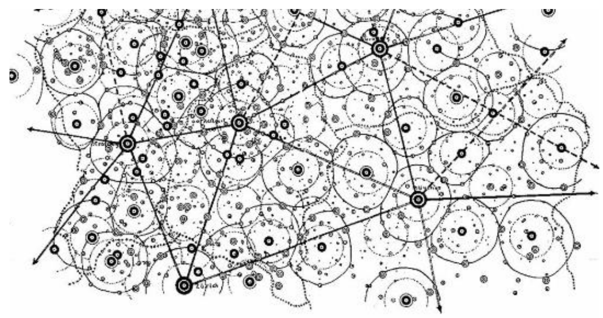


Figure 1.3. Sector model by Homer Hoyt.

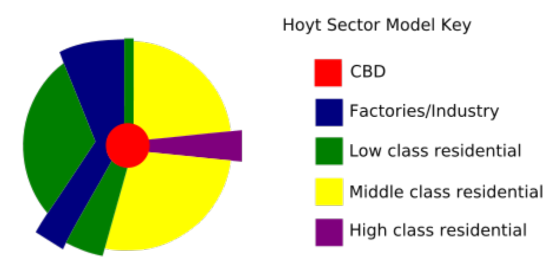
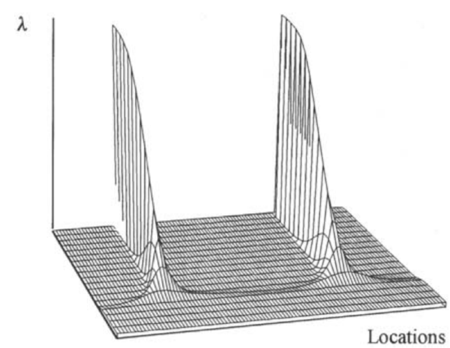


Figure 1.4. Krugman Edge Cities Model. In a continuos line that represents space, several sub-centers progressively increase their share of manufacturing space, converging to an equidistant set of central locations.



highly inarticulate fashion, phenomenon termed by the french sociologist Marc Auge as *"non-places" (Graham, Marvin, 2001)*.

Here, in contrast to measures based on spatial concentration, a relational approach is used for the identification of urban hierarchies, owing to the concept of *power-geometries* introduced by the neo-marxist geographer Doreen Massey. She defines space as "*the product of relations (including the absence of relations)",* which implies that rather than by what it contains, space *"is produced through the establishment or refusal of relations."(Massey, 2009)*

In order to translate the concept of *power-geometries* into an operational measure, two notions from the field of Social Network Analysis are applied to the characterization of Urban Structure: *"Social Proximity"* and *"Power Centrality"*. While *Social Proximity* makes reference to the connectedness of two places according to their degree of human interaction, *Power Centrality* expresses the relative importance of places based on the latter. The distinct feature of *Power Centrality* rests in the feedback quality it measures: How well connected are the elements of a network to other well connected elements (Bonacich, 1987).

In this research, the *Power Centrality* measure is used to test the underlying postulates of a recent model of Urban Structure: The *Splintering Urbanism* theory of Graham and Marvin (Graham, Marvin, 2001). The aim is to assess empirically a theory rich in metaphors and analogies, but that lacks the support of a measurable characterization of the contemporary city. Here the translation of metaphors into a formal measure is not considered a trivial exercise. As Paul Krugman showed when he developed his *"Edge Cities"* model based on the literary work of Joel Garreau (Figure 1.4), metaphors can prove an effective tool for abstracting complex and intractable urban phenomena.

The Splintering Urbanism Theory.

The *"Splintering Urbanism"* theory proposes that the contemporary city is namely a fragmented agglomeration of isolated urban pieces, where distant but valuable fragments are highly connected between them, by-passing their less valuable surroundings (Figure 1.7).

This theory is highly indebted to contemporary neo-marxist interpretations, that range from sociology and geography (Edward Soja, Manuel Castells, Mike Davis, David Harvey) to the arts (Figure 1.9). These interpretations posit that networks are collapsing the spatial constraints for a highly connected elite while increasing the poverty and isolation of groups that do not have access to this means of transport and communications. According to Graham and Marvin, the main drivers behind this fragmentation of urban space are networks, which allow determined groups of the population (the global or local *elites*) to bypass neighbouring areas considered of low-value, producing a *"tunnel effect"* (Figure 1.8).

In terms of Urban Structure, the

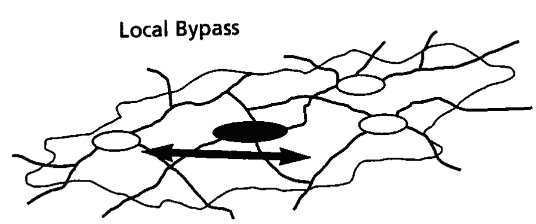


Figure 1.5. Local bypass as defined by Graham and Marvin (2001).

main argument of Graham and Marvin can be dissected into **three composing hypotheses:**

- **The Paradox of (Dis)Connection (sic):** With this literary device, Graham and Marvin make reference to how contemporary networks have segmented space and transformed the interactions on it. *"Valued"* users and places are *"effectively connected"* to each other, while *"by-passing non-valued"* locations.

- **Proximity ≠ Meaningful Relations! (sic):** Highly indebted to the concept of *"time-space compression*", introduced by David Harvey. As its name indicates, this concept expresses the relativeness of spatial constraints due to the expansion of communication and transport networks. According to Graham and Marvin, there is *"a sense of local disconnection"* in *"physically close, but socially and economically distant places and people" (Graham, Marvin, 2001)*.

- **Premium Networks:** This part of the argument can be logically derived from the two other hypotheses. Having *"valued"* users, in a context where places and people are *"socially and economically distant"*, implies the presence of a privileged segment of the population that benefits from contemporary networks. Graham and Marvin stress this view by referring to these networks as *"premium networks" (Graham, Marvin, 2001)*.

Evidently, in terms of location analysis *"high-value"* can be understood in multiple ways. The definition for a *"valued"* place in the *Splintering Urbanism*

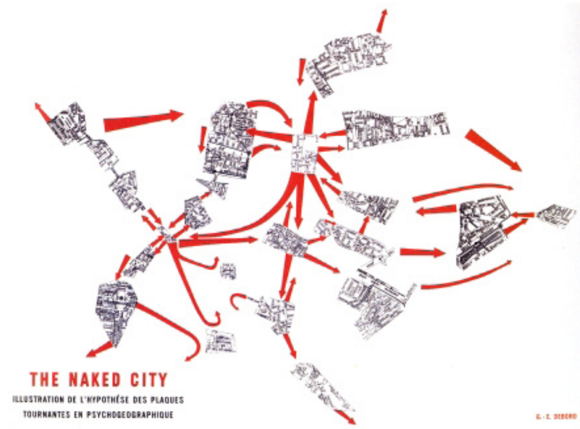


Figure 1.6. "The Naked City" (1957), by french situationist intellectual Guy Debord. Debord rearranged a tourist map of Paris by slicing its neighbourhoods and connecting them randomly. These itineraries expressed his concept of "drift" or "derive", where human trajectories are determined by chance.

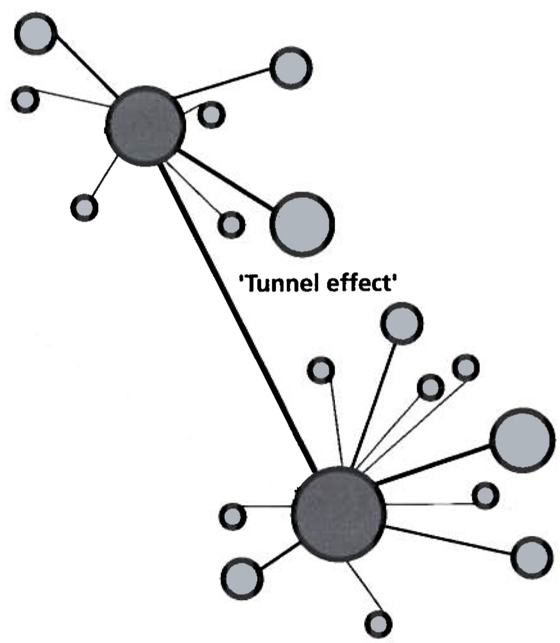


Figure 1.7. Illustration of the "tunnel effect", basis of the Splintering Urbanism theory, by Graham and Marvin. Areas of high interest for determined sectors of the population are highly connected between them, bypassing neighbouring areas considered of low value.

theory is context dependant: It can be an airport, a financial district, a shopping centre, a gated community, or the seat of political power. Given the ambiguity of the definition, that cannot be expressed in terms of fixed spatial indicators, it is the *"effectively"* connected keyword what provides the clue for its measurement.

- The **first hypothesis** to be tested is stated in terms of flows. If Graham and Marvin prove to be correct, then in a given urban system, this implies that the density of exchanges is going to be concentrated in a reduced set of privileged locations. As measured by the *Power Centrality* measure under the lens of *Social Proximity*, this makes reference to the exchanges of people between locations (or the overlap of human activity).

The second hypothesis of Graham and Marvin has a strict meaning in terms of space organization. For them, most of the models of Urban Structure produced during the 20th century belong to a *"modern ideal"*, expressed in the *"Uniplex Metropolis"*[1] *(Graham, Marvin, 2001)*. Graham and Marvin make a strong argument against the *"core-periphery"* models, that according to them underlie the modern ideal. Though this observation may seem trivial[2], it allows however a straightforward definition for comparison of Urban Structure. Already laws that explain polycentric urban structure have been formulated (Yang, et. al., 2012), however few of them in a relational measure.[3]

- The **second hypothesis** to be tested relates *Power Centrality* with dispersion. According to Graham and Marvin, in the *"Uniplex"* model of organization the central places of an urban system are restricted to the city core. In terms of *Power Centrality* this is expressed by comparing the areas of influence of the central places against the entire urban area, an index of urban eccentricity introduced by Yang et. al. (2012). If the ratio is relatively low, then the city presents a centralized rather than polycentric structure.

The third main aspect of the Splintering Urbanism theory relates directly with the neo-marxist interpretations that sustain it. For Graham and Marvin, the neo-liberal policies applied during the post-soviet era have accelerated the segregative effect of networks. They call this process the *"unbundling"* of infrastructures, where former state controlled monopolies are transferred to the private sector and opened to market competition. Their conclusion is that this former public goods are now customized for specific *"usually powerful"* users (Graham, Marvin, 2001). Despite being a compelling argument, on their excessive focus on the elites Graham and Marvin do not provide counterexamples that could aid in confirming their spatial hypothesis. Public transportation as well as subsidies and regulations targeted to the poor are barely mentioned[4]. Considering the lack of analysis of the new ways

1 A concept introduced by Patsy Haley, that groups most of the models of Urban Structure of the 20th century that assume the existence of a CBD (Central Business District) or predominant urban core. According to Haley, the "Uniplex" model also involves a "vision" of the city, or an ideal on which to organize cities in this fashion.

2 Since the 70's polycentric urban structures have been recognized and studied. See Anas, Arnott and Small (1998) for examples.

3 Among them, can be included the work of Batty with fine-grained commuting data. See Arcaute, Batty, et. al. (2013) for an example.

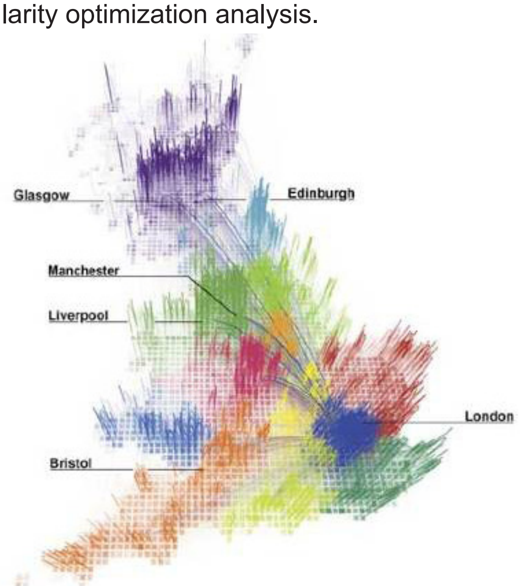
4 No assumption is made about the positive or negative effects of these policies for the disadvantaged populations.

of state intervention in the post-soviet era, the third hypothesis is based on the assessment of a counterexample to be presented in the case study.

- The **third hypothesis** of this research establishes that although the scenario depicted by the Splintering Urbanism theory may prove to be fairly representative of the contemporary Urban Structure, it is however incomplete. There indeed exist examples of societies that have undertaken deep privatization reforms where still direct state intervention has produced local bypasses that do not necessarily serve the elites.

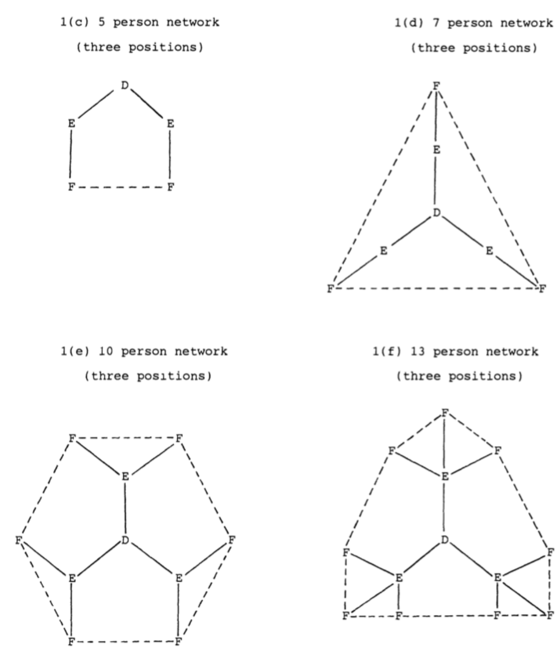
Social Proximity and Power Centrality.

Figure 1.8. Partitions derived from the social strength between places in Britain (Ratti et. al., 2010). The figure shows the strongest 80% of the links according the total talk time between areas. Each colour represents areas identified by a modularity optimization analysis.



The criterion used for assessing the connectivity between places is commonly referred to in Social Networks Analysis as *"Social Proximity"* or *"Social Affinity" (Borgatti and Halgin, 2011)*. These concepts make reference to the *"distance"* (or its inverse, the proximity) between *"actors"* (usually people or organizations) measured by relationships such as friendship, kinship or financial flows. In the methodological approach of this work, places are considered as the *actors*, while the distance between them is determined by their overlap of human activity. This approach has already been explored with massive datasets in a large geographical scale, such as with Senseable Citie's *"Borderline"* project (Ratti et. al., 2010), where millions of mobile calls were used to derive the social interaction strength between places (Figure 1.10).

Figure 1.9. Bargaining networks by Cook et. al. (1983). In all the networks, positions E rather than C (apparently more central), obtained the larger profits (measured empirically).



From the network of social overlaps between places the measure of *Power Centrality* is derived. In this case, the specific definition introduced by Phillip Bonacich in 1972, also known as *"Eigenvector Centrality" (Bonacich, 1987),* is applied. The underlying intuition behind this measure is that actors are more central not only by their sole amount of connections, but also by their connectivity to other central actors. The formal definition of Bonacich's centrality includes a factor that amplifies the strength of connectivity according to the *centrality* of the other *actors* to which an actor is connected. This factor, as Bonacich demonstrated, is equivalent to the output produced through a technique similar to Principal Component Anlysis (PCA)[5].

The effect of this procedure is the amplification of the *centrality* for *actors* that are well connected to other central *actors*. This measure, widely used in Network Analysis since the 70's[6], recently has been introduced for the study of patterns of human activity in the space, with examples in neighbourhood detection *(Cranshaw, Schwartz, Hong, Sadeh, 2012)*, place identification *(Calabrese, Reades, Ratti, 2010)*, mobility characterization *(Jiang, Ferreira and Gonzalez 2012)*, routine inference *(Eagle, Pentland, 2009)*, and centrality analysis *(Reades, Calabrese, Ratti, 2009)*.

Non-Convential Data Sources

The Splintering Urbanism model will be tested against data coming from Non-Conventional Data (NCD). In this work NCD makes reference to digital datasets that are readily accessible (usually public), frequently updated and obtained in the form of *metadata*[7].

This specific type of *metadata*, derived from activity registered on digital networks, offers extensive amounts of information about human interactions, but unlike conventional surveys, usually is not collected through a standard sampling method. The statistical analysis of this type of data, termed *"Reality Mining" (Pentland, 2009)*, combines a wide array of data mining[8] techniques for inferring meaningful

5 A technique for statistical dimension reduction.
6 Indeed it is the basis of the Page-Rank algorithm used by the Google search engine.

7 Data that relates to other forms of data.
8 The field that encompasses statistical techniques dedicated to the inference of patterns from large datasets.

Figure 1.10. Twitter and Instagram activity in London, by Eric Fischer.

Figure 1.11. Foursquare activity in Jakarta.

patterns from this somewhat unspecific sources of information. As millions of persons are acquiring these pervasive tracking systems, at an especially fast rate in the developing world *(Pentland, 2009)*[9], these sources of data emerge as cheap (in terms of access and time) proxies[10] for urban measures across a wide range of cultural, social and economic contexts.

The Case: Santiago Metro.

The case to be studied in this thesis corresponds to the Santiago Transportation system, and in specific, its grade-separated[11] railway network, Metro. The urban area of Santiago, the capital of Chile, is composed of roughly 36 municipalities or *comunas*, with no centralized authority with the exception of the *Intendente*, head of the broader Metropolitan Region (an administrative division of the country). This authority, appointed directly by the president of the country, performs namely administrative functions, leaving the key decisions of infrastructure development and regulation in the hands of the corresponding governmental agencies (such as the Ministry of Housing or the Ministry of Public Works)[12]. It is estimated that the urban area within the Metropolitan Area[13] comprises 975 square kilometres, with a population slightly above the 5.8 million

Figure 1.12. LANDSAT Image of Santiago, 2011.

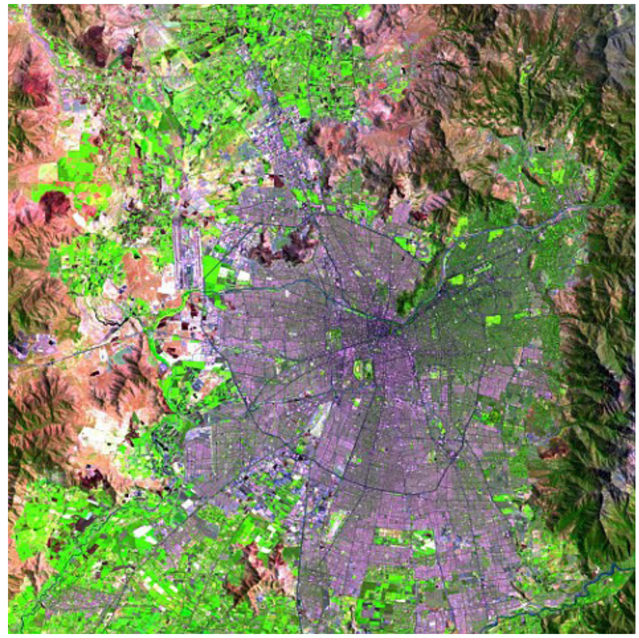
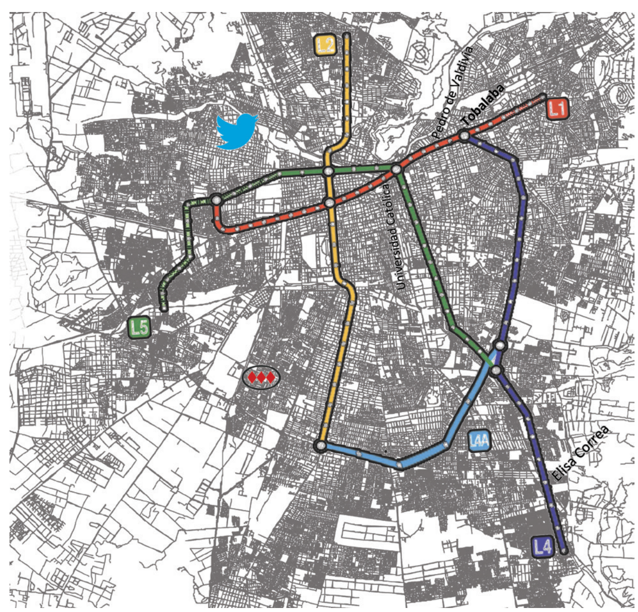


Figure 1.13. Road layout of Santiago de Chile and the Metro lines.



9 According to the International Data Corporation (IDC), Smartphones sales outpaced for the first time in history the sales of conventional mobile phones in 2013.
10 Variables that can be used to approximate the value of other variables of interest.
11 A mode of transport that is independent or not interrupted by other modes of transport.
12 And other investments of local extent in the hands of the comunas, that are governed by elected authorities.
13 The urban area of Santiago, not to be confused with the administrative Metropolitan Region.

people (ALC-BRT CoE, 2012). Overall, the Metropolitan Region, contains more than 6 million inhabitants. Although it varies according to the definition employed, satellite cities detached from the contiguous urban agglomeration are regarded as part of the Metropolitan Area, including Lampa and Colina to the north (which includes middle and high income large scale real estate projects)[14], Talagante to the south, and on some occasions, the small city of Melipilla to the west.

Since its opening in 1975, Metro was regarded until recently as an independent transportation system, without integration with other transit services. This situation changed in 2007 with the implementation of Transantiago, a BRT[15] inspired bus system that replaced the old system run by thousands of independent operators. The new transportation scheme introduced the fare integration of Metro with the buses, affecting severely the commuting patterns in both systems.

Transantiago has been arguably one of the most controversial public policies of the last decades, bringing high political costs for authorities and officials. One of its most debated aspects was the integration with Metro *(Quijada, Tirachini, Henriques, Hurtubia, 2007)*, which shifted dramatically the modal choice[16]. Before Transantiago, Metro concentrated a mere 1,6% of the daily trips in the city *(Steer Davies and Gleave, 2012)*. Five years after the implementation of this transportation reform, estimations indicate that Metro concentrates 31% of the transit trips, and in combination with bus transfers, adds up to other 66% of the trips (Figure 1.14). Consequently, Metro has become the de facto backbone of the Santiago's public transportation system, having incidence on almost two thirds of the city's transit daily trips *(ALC-BRT CoE, 2012)*.

14 This projects, known as "Conditioned Zones" (ZODUC and PDUC) are granted with development rights outside the legal urban limit conditioned to the fulfilment of infrastructure requirements imposed by the Metropolitan masterplan.
15 Accronym for Bus Rapid Transit. According to the BRT standard manual (2013, http://www.itdp.org/uploads/BRT_Standard_2013_ENG.pdf), a BRT corridor is defined as "A section of a road or contiguous roads served by a bus route or multiple bus routes that have dedicated lanes with a minimum length of 4 kilometres."

Metro as a counterexample of Premium Network.

The Metro of Santiago is offered as a counterexample to the *"premium networks"* concept of Graham and Marvin. The transportation sociologist Olivier Coutard has been one of the few to point the relativity of this concept. He engaged in a scholar debate with Simon Graham,

16 The distribution of trips within the Metropolitan Area between transport modes (car, bus, walk, Metro, etc.)

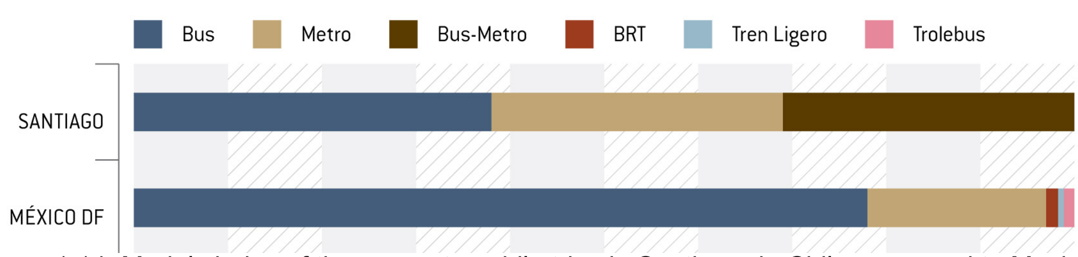


Figure 1.14. Modal choice of the non-automobile trips in Santiago de Chile compared to Mexico City, by BRT Across Latitudes (2013).

arguing that water infrastructures in developing countries have indeed targeted the poor after the privatization reforms *(Graham, 2002)*. With the example of Metro, this research intends to extend that critique to the domain of transportation infrastructures, profusely exposed in Graham and Marvin's book as prime examples of elite-tailored networks. In the case of Santiago, public transportation still concentrates 52% of the daily trips *(MTT, 2013)*[17], a modal choice that the government intends to maintain with the investment package planned in the 2025 Santiago Transportation Master Plan.

The shock that the implementation of Transantiago meant for Metro caused a chaotic situation during the first days of integration between both systems. Estimations set the saturation within the trains at record levels of 7 persons per square meter[18]. Through heavy investments, the government has increased the central role of Metro, transforming it into one of the most extensive Metro systems in the world[19] (Figure 1.15). The development strategy for the next 13 years deepens that trend, with a projected investment in Metro related infrastructure amounting to 41% of the budget[20].

Despite the massive reorganization of the travel routines induced by Transantiago, that have installed a general negative view of the transit system[21],

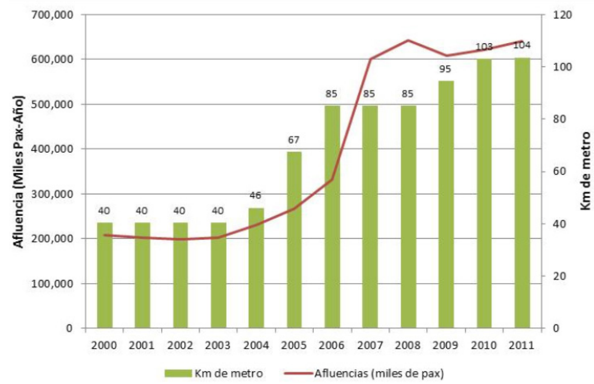


Figure 1.15. Santiago Metro Service: Number of passengers carried per year (red line, in thousands), and extension of the system (green bars, in kilometers). By Steer Davis and Gleeave (2012).

indicators show that the difference of travel time in public transportation compared to private mobility is not dismal[22]. Albeit the saturation of the Metro network, in terms of delivery[23], infrastructure, and rail technology, the service is internationally recognized among the best in the world.[24] The effect of the Metro network in the public transportation system is noticeable, making it the preferred mode of transportation for long distance travel, and increasing the average speed of the transit system,

17 It must noted however that other sources such as BRT Center of Excellence drop this figure to a much lower level. Specifically they indicate that the public transportation share is 41%.

18 Reported by the national TV station, TVN: http://www.youtube.com/watch?v=zPJNmV5C4KY

19 BBC Online: http://www.bbc.com/travel/blog/20130314-subway-systems-by-the-numbers

20 Budget hat also includes private initiatives regulated by the government.

21 Different polls show consistently a high level of disapproval of the system. According to Gfk-Collect, this year reached a record level of 61%: http://radio.uchile.cl/2013/01/28/transantiago-enfrenta-la-peor-evaluacion-ciudadana-desde-su-inauguracion

22 Considering that the public transportation figures include about 5 minutes of "access" time, that is the walk to a stop.

23 As reported by the local press (based on statistics provided by Metro), Santiago Metro ranks as the 7th most regular Metro system in the world, and 75% of the trains are on time with the schedule: http://www.latercera.com/noticia/nacional/2013/06/680-527912-9-metro-de-santiago-en-cifras-transporta-al-ano-casi-la-mitad-de-la-poblacion-de.shtml

24 Metro Santiago has been conferred international design and service awards (including the best Metro of the Americas): http://archpeace.blogspot.com/2012/04/metro-of-santiago-more-than-mere-public_14.html.

placing it over cities with well developed BRT systems such as Bogota and Curitiba, *(ALC-BRT CoE, 2013)*. Considering the share of the population served, the quality of the service, and the heavy state investment, certainly Metro constitutes an example of *"premium network"* that does not target the elites.

Data

The specific dataset analysed in this work comes from *twitter*, a *"micro-blogging"* service. *Microblogging* sites allow the exchange of short messages between their subscribers. On *twitter*, the limit of the messages (called *statuses* or *tweets*) is set to 140 characters. What defines *twitter* is its public character: Unless the subscribers explicitly block their content, this is available online not only to the rest of users in the network (which can obtain constant *"feeds"* from desired accounts by *"following"* them), but also to any other person on Internet. Likewise, their datasets are available online for software developers that connect to their API (Application Programming Interface), a tool commonly provided by most of the Social Media services (though with varying degrees of access). Through this means the data for this work was collected.

An extremely useful trait of *Social Media* services is their *metadata*. For the case of this work, of particular interest is the geo-located information included on Twitter. Approximately 1% of Twitter users have enabled the publication of geo-located content, that indicates where the *tweets* were generated. Other useful feature of *twitter* is the integration with other *Social Media* services. *Tweets* including data from a popular location-based social network, *foursquare*, are used as means

Figure 1.16. Twitter usage by country (2012), by Semiocast.

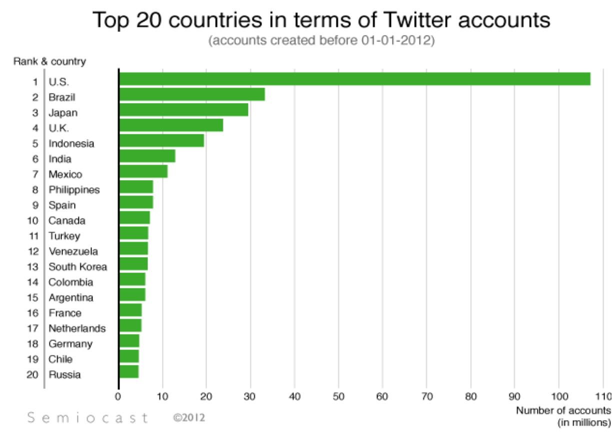
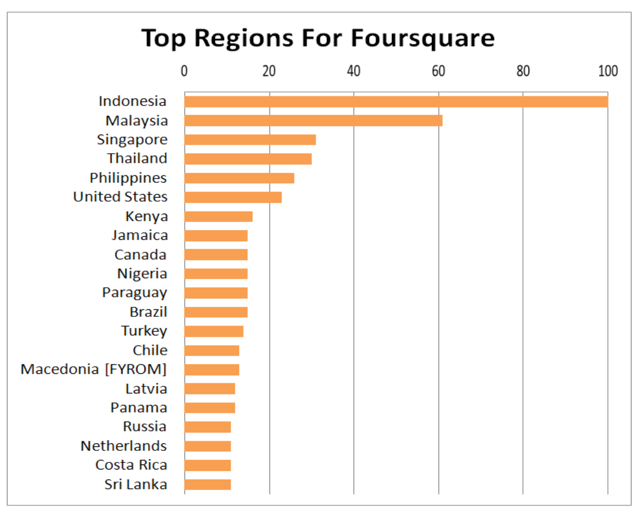


Figure 1.17. Foursquare usage by country (% of the population, 2012), by Ignite Social Media.



for tracing passengers of the Metro system. *Foursquare* allows to *“check-in”* or report the presence at specific *“venues”* or places contained in the database of the system. This allows analysts to search for specific locations that otherwise would be represented by highly scattered datapoints.

Twitter data (and its associated *foursquare* content) is considered of interest given the intensive use of *Social Media* services in Chile, among the highest in the world. By sheer number of accounts, Chile is amongst the top-20 countries in terms of *twitter* subscription (Figure 1.16), and in proportional terms its foursquare usage is among the highest in the world (Figure 1.17).

The final dataset used in this investigation was composed by 242.000 geo-located tweets, belonging to 1.532 users and representing a time frame that spanned for 400 days (from late 2010 to mid-2013). Of them, approximately 20.000 *tweets* contained *foursquare metadata*.

Structure of the Thesis

The following chapter introduces the intellectual background from Urban Theory considered relevant in terms of spatial relational analysis. Two seminal works, *“The Image of the City”* of Kevin Lynch, and *“A City is Not a Tree”* of Christopher Alexander are discussed as analytical approaches that stress the importance of human activity overlaps for defining the character of urban places. Linking with that intellectual background, a brief introduction to Social Networks concepts employed in this research is provided, including an overview of the analytical techniques used. The concept of *Power Centrality* is explained in formal terms.

The methods section describes in detail the analysis of the Santiago Metro *twitter* dataset, including characteristics of the dataset, specifications on the algorithm employed (including formal mathematical definitions), and results. The spatial hypotheses of the Splintering Urbanism theory are tested empirically, while the premium infrastructures concept is assessed qualitatively.

The conclusions section contains a brief discussion on the main topics addressed by this thesis in the light of the case studied: The effectiveness of the *Power Centrality* measure as a relational proxy for urban hierarchy, the reach of the *premium networks* concept of Graham and Marvin, and closing remarks on the shifting paradigms regarding digital networks and their impact on the human activity in space.

Figure Sources

Figure 1.1. Source: Wikimedia Commons. http://en.wikipedia.org/wiki/File:Bid_rent1.svg. Retrieved August 31, 2013.

Figure 1.2. Baskin, Carlisle W. 1966. "Central places in southern Germany." Trans, of Christaller (1933). Englewood Cliffs, NJ: Prentice-Hall.

Figure 1.3. Author: Suzanne K.N. Source: Wikimedia Commons. http://en.wikipedia.org/wiki/File:Hoyt_model.svg. Retrieved August 31, 201

•**Figure 1 .4.** Fujita, Masahisa, Paul R. Krugman, and Anthony Venables. 1999. The Spatial Economy : Cities, Regions and International Trade. Cambridge, Mass: MIT Press.

•**Figure 1 .5.** Graham, Stephen, and Simon Marvin. 2001. Splintering urbanism : networked infrastructures, technological mobilities and the urban condition / Stephen Graham and Simon Marvin: London ; New York : Routledge, 2001.

•**Figure 1 .6.** Ibidem.
(Graham, Marvin, 2001) Graham, Stephen, and Simon Marvin. 2001. Splintering urbanism : networked infrastructures, technological mobilities and the urban condition / Stephen Graham and Simon Marvin: London ; New York : Routledge, 2001.

•**Figure 1 .7.** Author: Guy Debord. Source: Wikimedia Commons. http://it.wikipedia.org/wiki/File:The_Naked_City.jpg. Retrieved August 31, 2013.

•**Figure 1 .8.** Ratti, Carlo, Stanislav Sobolevsky, Francesco Calabrese, Clio Andris, Jonathan Reades, Mauro Martino, Rob Claxton, and Steven H. Strogatz. 2010. "Redrawing the Map of Great Britain from a Network of Human Interactions." PLoS ONE no. 5 (12):1-6. doi: 10.1371/journal.pone.0014248.

•**Figure 1 .9.** Author: Cook, Emerson, Gilmore and Yamagishi Source: Bonacich, 1987

•**Figure 1 .10.** Author: Eric Fischer. Source: http://www.flickr.com/photos/walkingsf/5925800427/in/set-72157627140310742. Retrieved August 31, 2013.

•**Figure 1 .11.** Author: Foursquare.com Source: https://foursquare.com/infographics/500million. Retrieved August 31, 2013.

•**Figure 1 .12.** Author: USGS (2010). Source: http://landsat.usgs.gov/gallery_view.php?category=nocategory&thesort=mainTitle. Retrieved August 31, 2013.

•**Figure 1 .13.** Own elababoration from basemap (road network) of Observatorio Ciudades Universidad Catolica (OCUC). Retrieved August 31, 2013.

•**Figure 1 .14.** Author: BRT Across Latitudes, Benchmark Report (2013). Source: http://www.brt.cl/wp-content/uploads/2013/03/Asesor%C3%ADa-Experta-para-la-Ejecución-de-un-Estudio-Comparativo_0320.pdf. Retrieved August 31, 2013.

•**Figure 1 .15.** Steer Davies Gleave. 2012. "Estrategia de Desarrollo Regional de la Region Metropolitana de Santiago". Document prepared for the Regional Government (GORE) and the United Nations Development Programme (PNUD).

•**Figure 1 .16.** Authors: Semiocast (2012). Source: http://semiocast.com/en/publications/2012_01_31_Brazil_becomes_2nd_country_on_Twitter_superseds_Japan. Retrieved **August 31, 2013.**

•**Figure 1 .17.** Authors: Ignite Social Media (2012). Source: http://www.ignitesocialmedia.com/social-media-stats/2012-social-network-analysis-report/. Retrieved August 31, 2013.

POWER CENTRALITY: DEFINING URBAN HIERARCHIES FROM SOCIAL OVERLAPS.

Image of the City and A City is not a Tree: Recovering the social fabric.

The early 60's witnessed the appearance of seminal works in Urban theory that scathingly criticized modernist Planning. According to this works, planners during the first half of the century produced inhumane spaces, due to the pursue of their own ideological agenda, which led them to forget, or plainly deny, the social fabric that ultimate sustains urban development.

Though most of the Urban intellectuals during the period emphasized the relevance of social interactions as drivers of urban development -including landmark works such as *"City in History"* (Mumford, 1961) and *"Death and Life of Great American Cities"* (Jacobs, 1961)-, were Kevin Lynch and Christopher Alexander who explicitly situated human activity as the defining component of cities. Two of their iconic works, *"The Image of the City"* (Lynch, 1964), and the short essay *"A City is not a Tree"* (Alexander, 1965), directly attacked the way in which planners, according to them, where imposing deductive models of the city to its inhabitants. The core of their argument relied on the way in which cities are organized in social terms, which for them simply did not match the assumptions hold as true by planners.

The city as a collection of human overlaps.

In the case of Lynch, the critique targeted the universalist principles underlying the Planning practice at the time. Though the explicit goal of his study was the definition of a coherent *"city image"* from the aggregated experience of individuals, his findings showed a rather fragmentary perception of the urban environment. Lynch and his team at MIT interviewed a small sample of 60 persons in Boston, Jersey City and Los Angeles, producing *"mental maps"* that resembled abstracted networks of the built environment.

Lynch indeed borrowed terms from topology to name the elements described by the mental maps (*"paths", "edges"* and *"nodes"*). One of his main discoveries was the existence of several overlaps on the mental maps (Figures 2.1, 2.2). Regarding this analytical approach, Lynch's work constitutes one of the first challenges to the physical determinism imbued in the Planning practice at the time.

While for Lynch human activity overlaps where one amongst many of the elements that helped to define the image of city, for Alexander they were the defining characteristic of cities *par excellence*[1]. The main hypothesis underlying his essay *"A City is Not a Tree"* is that planners sought the segregation of the city into autonomous units, in open contradiction with the overlapping nature of human interactions. To exemplify, Alexander described two opposing urban typologies. The first one, designed by planners and usually built in a single run, were the *"Artificial Cities"* (Brasilia, Chandigarh, Milton Keynes) while the other typology, the *"Natural Cities"* corresponded to those grown spontaneously (or mainly in that fashion) over several generations.

Alexander -a Cambridge mathematician- resorted like Lynch to graph theory, though in a rather literal sense. According to his interpretation, artificial cities were organized as *"trees"*, following a hierarchical structure where spatially segregated units would follow in functional terms a

1 At least during his first works, produced during the second half of the 60's.

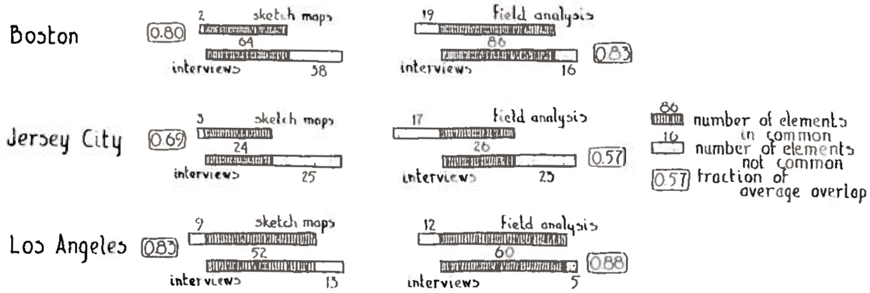


Figure 2.1. Summarized results for the surveys carried by Lynch and his team at MIT for the "Image of the City", indicating the number of common elements identified by city inhabitants.

nested organization. This kind of organization could encompass every aspect of the urban domain; Convenience stores would serve small neighbourhoods, medium retailers would serve districts, and shopping centres would serve a town or city; Roads would feed avenues, avenues would feed arteries, and arteries would feed highways, and so on. This hierarchy would not present overlaps, isolating units of similar equivalence -either in formal or functional terms- from each other.

Alexander's thesis was that people preferred natural cities because they where organized as *"Semi-Lattices"* (Figure 2.3), defined by him as a collection of sets that overlapped. Alexander's mathematical diagrams showed that rather than thinking cities as aggregates of neatly segregated units, they and their constituent parts should be defined by more subtle grouping constraints.

Principles of Social Affinity Analysis.

In *"A City is not a Tree"*, the use of social networks diagrams for comparing tree-like structures against *Semi-Lattices* was not casual. Just the year before Alexander presented his Phd. thesis based on the application of graph theory to Design problems, while working in computer science at MIT. Among the sources cited by Alexander, is the work of Alex Bavelas, one of the pioneers in the field of *Social Network Analysis*. The main challenge that research on this field faces is

Figure 2.2. Overlapping boundaries of neighboorhoods in central Boston, according to the results of Lynch's survey for the "Image of the City" .

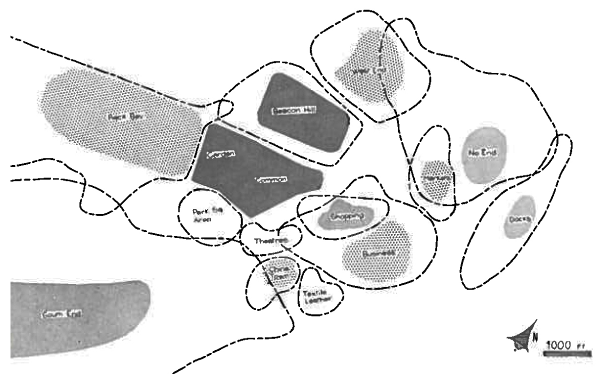
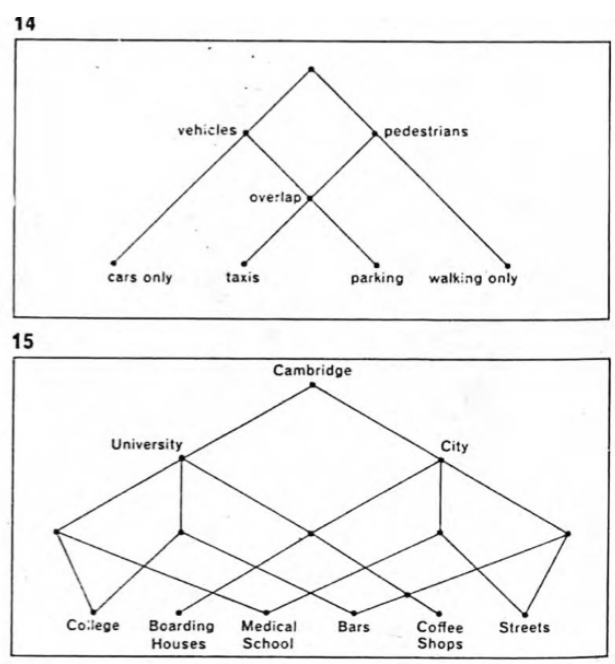


Figure 2.3. Examples of urban Semi-lattices structures, provided by Alexander, corresponding to a traffic network and a university-based city (Alexander, 1965).



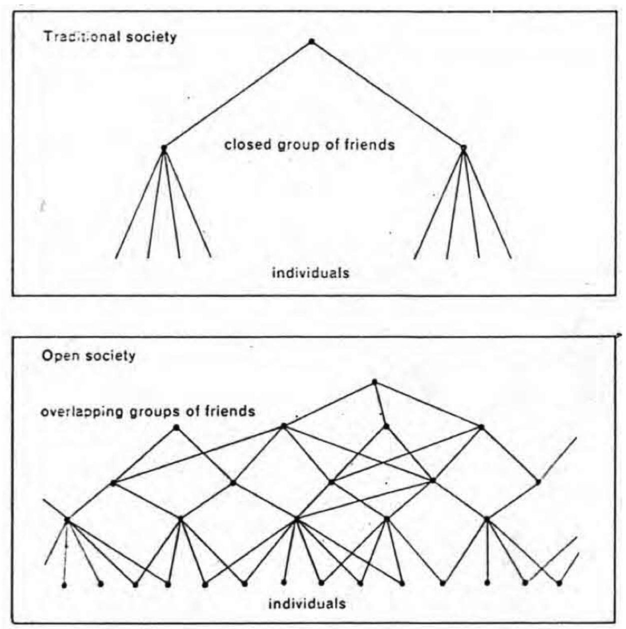
the identification of patterns in structures, such as social groups, that are complex by essence. This problem was well described by Alexander in his description of contemporary Social Networks:

"If we ask a man to name his friends and then ask them in turn to name their friends, they will all name different people, very likely unknown to the first person; these people would again name others, and so on outwards. There are virtually no closed groups of people in modern society. The reality of today's social structure is thick with overlap" (Alexander, 1965). (Figure 2.4)

In order to undertake this somewhat paradoxical task -how to identify groups in structures that present no apparent divisions-, *Social Networks Analysis* has resorted, since its birth around the 1930's *(Freeman, 2011)*, to the abstraction of social structures as graphs composed by *actors* (people, organizations, etc.) and links (relationships such as friendship, kinship, or financial flows). Especially since the early 70's, several algorithms of increasing sophistication have been developed for the empirical detection of subgroups within Social Networks. These algorithms termed interchangeably[2] as *"Community Detection Algorithms"*, *"Partioning Algorithms"* or *"Modularity routines"*, aim to the detection of optimal partitions according to a given criteria.

Despite the explosive expansion of research in the field during the last decades (specially since the increased interest of physicists and computer scientists on the study of networks)[3], the underlying intuitive notion has remained practically intact: A group is defined when its members interact more often between each other (an area of *"thick spots")* than with outsiders (being the interactions between groups areas of *"weak ties"*). The clustering approach used in this work relates to the analysis of *"Affiliation Networks"*, term that refers to the relationships between two set of actors (for example people and organizations) *(Borgatti and Halgin, 2011)*. For the case of this research the analogy is straightforward; locations are related with people.

Figure 2.4. Diagram of two opposing models of social networks, by Alexander (1965).



The analysis of *Affiliation* data is commonly based on three key concepts that define a regular procedure. The first corresponds to the definition of the net-

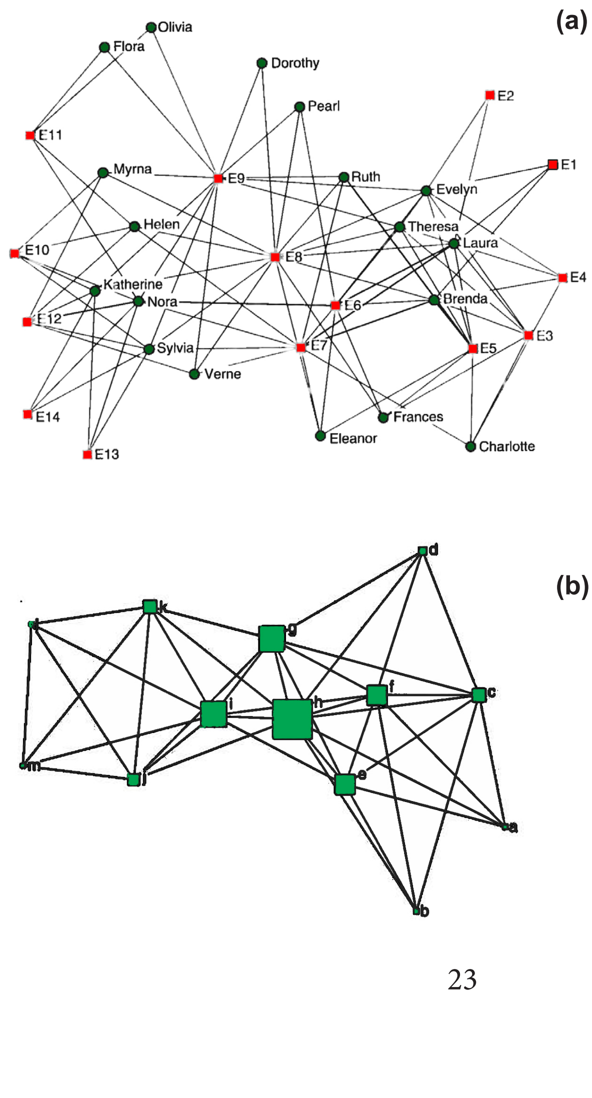
2 Though, depending on each case, these terms may not refer to exactly the same concept (but usually they are closely related).

3 Since the late 90's physicists have introduced a "revolution" in the study of social networks (see Freeman, 2011), based on the introduction of Network Models (in comparison to the conventional analytical approach of Social Networks Analysis). Despite their relevance, the discussion of these models are beyond the scope of this work. Among the most relevant models of this "revolution" are the Watts-Strogatz (1989), Barabasi-Albert (1998) and Newman's cut models.

work itself: Actors are separated into *“modes”* or different sets, which represent a distinct attribute in terms of the relationships being studied (for example, employees and corporations, or people and events as shown in Figure 2.5). The second concept relates with the measure of the distance between the actors, usually focusing on the actors of one of the modes. In order to define this proximity, the data is *“projected”* into *one-mode* data, containing only weighted relationships between one set of actors, according to their *“structural equivalence”*. *Structural equivalence* (here used in its broad definition) makes reference to how similar are the links connecting two given actors compared to the rest of the network (Figure 2.5). For this research, the original bipartite network (i.e. composed of two modes) relates a set of locations with a set of *twitter* users.

The third concept relates with the proper clustering of the *Affiliation Network*. Here, for the analysis of Santiago and its Metro system, a measure of centrality is used for grouping places: *Power Centrality*. This measure, defined by Phillip Bonachich in 1972 (Bonachich, 1987), resorts to a factor that amplifies the *“Degree Centrality”* of an actor -the sum of connections to others— according to the centrality of the actors to which it is connected. In it is simplest form, the factor is equivalent to the first *Eigenvector*[4] (Figure 3.8) of the *Adjacency Matrix* (a symmetric matrix where each cell indicates the connections between pairs of actors, see Figure 2.7). The score of each node or *actor* in the first *Eigenvector* is an indication not only of its aggregated centrality (how many links it has to others), but how well connected is to other powerful nodes. As an essentially feedback measure, the score of a small set of powerful and well connected nodes is significantly higher than the rest of the network. Several grouping procedures can be applied based on this measure. Notwithstanding, here powerful places will be defined according to thresholds of *Eigenvector* scores.

Figure 2.5. Deep south dataset by A. Davis (1941). (a) Two-mode network where actors (women, in green), are related to events (social events, in red). (b) Projection of the deep-south dataset into one-mode composed solely by the set of women. The size of the nodes indicates their centrality (in terms of number of social clubs they belong to).



4 Eigenvector makes reference to the output of Principal Component Anlysis (PCA), a data dimension reduction technique. The aim is the identification of a reduced set of vectors that account for much of the variability in the data. See figure 3.8 for an illustration of the technique.

	Gr1	Gr 2	Gr 3	Gr4	Gr 5
I1	1	0	0	0	0
I3	0	0	0	0	0
W1	1	1	1	0	0
W2	1	1	0	0	0
W3	1	1	1	0	0
W4	1	1	1	0	0
W5	0	0	1	0	0
W6	0	0	0	1	0
W7	0	0	0	1	1
W8	0	0	0	1	1
W9	0	0	0	1	1
S1	0	1	1	0	0
S2	0	0	0	0	0
S4	0	0	0	0	1

Figure 2.6. Two-Mode or Bipartite Matrix, indicating the relationships between women and events for Deep South dataset (Davis, 1941).

	EVE	LAU	THE	BRE	CHA	FRA	ELE	PEA	RUT	VER	MYR	KAT	SYL	NOR	HEL	DOR	OLI	FLO
EVELYN	8	6	7	6	3	4	3	3	3	2	2	2	2	2	1	2	1	1
LAURA	6	7	6	6	3	4	4	2	3	2	1	1	2	2	2	1	0	0
THERESA	7	6	8	6	4	4	4	3	4	3	2	2	3	3	2	2	1	1
BRENDA	6	6	6	7	4	4	4	2	3	2	1	1	2	2	2	1	0	0
CHARLOTTE	3	3	4	4	4	2	2	0	2	1	0	0	1	1	1	0	0	0
FRANCES	4	4	4	4	2	4	3	2	2	1	1	1	1	1	1	1	0	0
ELEANOR	3	4	4	4	2	3	4	2	3	2	1	1	2	2	2	1	0	0
PEARL	3	2	3	2	0	2	2	3	2	2	2	2	2	2	1	2	1	1
RUTH	3	3	4	3	2	2	3	2	4	3	2	2	3	2	2	2	1	1
VERNE	2	2	3	2	1	1	2	2	3	4	3	3	4	3	3	2	1	1
MYRNA	2	1	2	1	0	1	1	2	2	3	4	4	4	3	3	2	1	1
KATHERINE	2	1	2	1	0	1	1	2	2	3	4	6	6	5	3	2	1	1
SYLVIA	2	2	3	2	1	1	2	2	3	4	4	6	7	6	4	2	1	1
NORA	2	2	3	2	1	1	2	2	2	3	3	5	6	8	4	1	2	2
HELEN	1	2	2	2	1	1	2	1	2	3	3	3	4	4	5	1	1	1
DOROTHY	2	1	2	1	0	1	1	2	2	2	2	2	2	1	1	2	1	1
OLIVIA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2
FLORA	1	0	1	0	0	0	0	1	1	1	1	1	1	2	1	1	2	2

Figure 2.7. Adjacency Matrix, indicating the number of interactions in common for the Deep South dataset (Davis, 1941).

Figure 2.8. Illustration of Principal Component Analysis, by Markus Ringer (2008). a) Plot of two variables. b) Projection of two principal components (PC1 and PC2) on the plane formed by XBP1 and GATA3. These are the first two components (or eigenvectors) that capture most of the variability of the data (they are orthogonal to each other). c) Projection of the data onto PC1 (the orthogonal projection of datapoints onto this vector). d,e,f) More auxiliary plots showing the distribution of the data, and the projection of the original variables onto the new vector space.

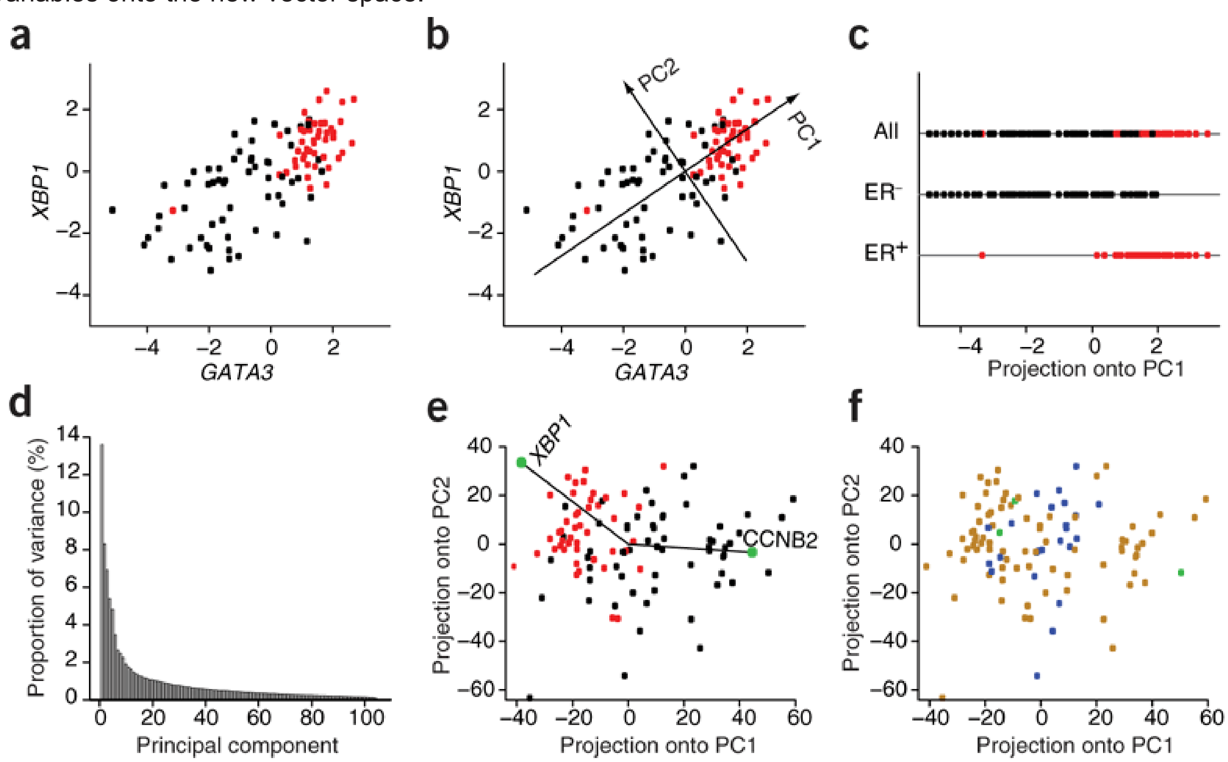


Figure References

Figure 2.1. Lynch, Kevin. 1964. The image of the city / Kevin Lynch, Publication of the Joint Center for Urban Studies: Cambridge, Mass. : M.I.T. Press, 1964. 1st M.I.T. Press pbk. ed.

Figure 2.2. Ibidem.

Figure 2.3. Alexander, Christopher, and Lewis Mumford. "A city is not a tree."

Figure 2.4. Ibidem.

Figure 2.5. Borgatti, Stephen P, and Daniel S Halgin. 2011. "Analyzing affiliation networks." The Sage handbook of social network analysis:417-433. (Borgatti and Halgin, 2011) Borgatti, Stephen P, and Daniel S Halgin. 2011. "Analyzing affiliation networks." The Sage handbook of social network analysis:417-433.

Figure 2.6. Ibidem.

Figure 2.7. Ibidem.

Figure 2.8. (Cha, 2007) Cha, Sung-Hyuk. 2007. "Comprehensive survey on distance/similarity measures between probability density functions." City no. 1 (2):1.
Ringnér, Markus. 2008. "What is principal component analysis?" Nature biotechnology no. 26 (3):303-304.

<This page intentionally left blank>

METHODS AND RESULTS.

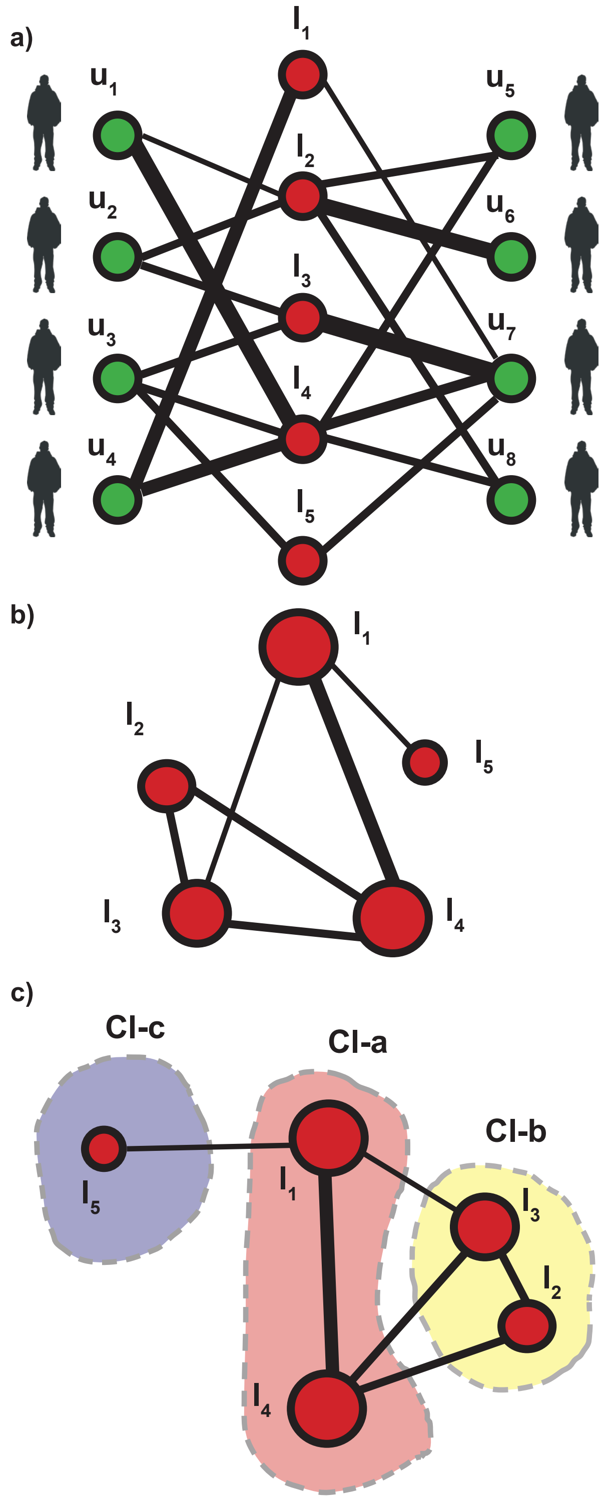
Socio-Spatial Power Centrality: Definition of the problem.

The method employed can be defined in relation to the problem composed by the following steps:

- Given a *two-mode* or *bi-partite* network, composed by a set of persons (users, *U*) and a set of locations (*L*), calculate the strength of the links between the actors of the two sets (how many times each user *u* visited a location *l*).
- Assess the social proximity between the set of locations by projecting the data into one mode. The new weighted links define the strength of relationships between places (according to the overlaps of activity).
- Compute the *centrality* of each node (location), defined by its own *centrality* as well as the *centrality* of the nodes to which they are connected. *Centrality* is derived from the social strength of the ties.

The goal is to amplify the *centrality* score of the small set of locations situated at the top rank, that supposedly will have a high strength of connections to other well connected locations. This set of locations should represent the group of places that are densely connected to each other and thus are more *central* in relation to the rest of the network, based on the strength of their flows.

Figure 3.1. Definition of the Problem. a) Split the network into two-modes, locations (L) and users (U) and calculate the weight of the ties (number of visits). b) Project the information into a one-mode network of locations, and calculate the centrality of the nodes. c) Identify groups of nodes with similar centrality (compared to the rest of the network).



Data Collection

The dataset is composed by *twitter* statuses (better know as *tweets*), which contain the following information for each observation (or *tweet*): user ID, a timestamp, latitude and longitude, and the source of the *tweet*. The source field indicates the origin of the status, that could be *twitter* itself, or other *Social Media* applications where the user enabled connection with *twitter*. When connection is enabled, any content generated on the other platforms is automatically reported in the form of *twitter statuses*.

For this work the most popular Location-Based Social Network (LBSN), *foursquare*[1], is used as means for identifying *twitter* users that have been at Metro stations. *Foursquare* works as an online game where users *"check-in"* or report their presence at unique places stored in the database of the system (with the monicker of *"venues"*). The application offers a set of nearby *venues* from which users can choose to *check-in*, according to the geographical coordinates reported by the smartphone. The basic principle for filtering data in the initial step is the search of *twitter statuses* that contain *foursquare check-ins* at Metro stations[2]. A list of 1.532 users that have *checked-in* at Santiago Metro venues is composed by filtering data with *twitter's "streaming"* and *"followers"* API tools[3]. Later, through *twitter's timeline* API tool , the last 3.200 *statuses* are retrieved for each one of these users[4].

Working with this data implied two main assumptions. As the *tweeting* frequency varies widely among its users (Figure 3.26)[5], it was assumed that the distortion produced by the overrepresented *twitter* users would not bias the sample in favour of determined locations. For the specific case of *foursquare check-ins*, it was assumed that users mostly tend to *check-in* at *venues* where they actually are or were recently, as there is no special incentive to cheat the system (i.e. report their presence at other locations). See the related *"Limitations and Discussion"* section for a more detailed description of these biases.

Data Preprocessing and Aggregation

As it can be observed in the time series graph (Figure 3.4), the amount of observations per day increased significantly since the first day of the observed

1 See Noulas et. al. (2011).

2 According to some estimations, 20% to 25% of Foursquare users have enabled the connection with Twitter. See Scellato et. al. (2010).

3 The first tool ("streaming") allows the filtering of twitter statuses according to combinations of desired kewyords to search: https://dev.twitter.com/docs/api/1.1/post/statuses/filter. This method is limited to the retrieval of tweets not older than one week approximately. By accumulating tweets periodically from the last week of 2012 to the first week of February 2013 a list of roughly 700 users was composed. The second tools retrieves lists of "followers" or users that automatically receive the tweets of a specific account: https://dev.twitter.com/docs/api/1.1/get/followers/list. A set of approximately 800 users that follow the official Santiago Metro account was composed from this list. As the followers list contains more than 300.000 users, only a small fraction of the list could be parsed (roughly 10%, which order was determined by twitter's sorting from newest to oldest followers). The combination of both lists, obtained independently through these two methods ("streaming" and "followers"), compose the final list of over 1.500 twitter users. These users register in their twitter accounts at least one check-in produced on foursquare at Metro stations.

4 https://dev.twitter.com/docs/api/1.1/get/statuses/user_timeline

5 Discussions have centered around on wether the distribution of messages within the network follows a proper power-law. For an example see http://techliberation.com/2009/08/17/twitter-power-laws/

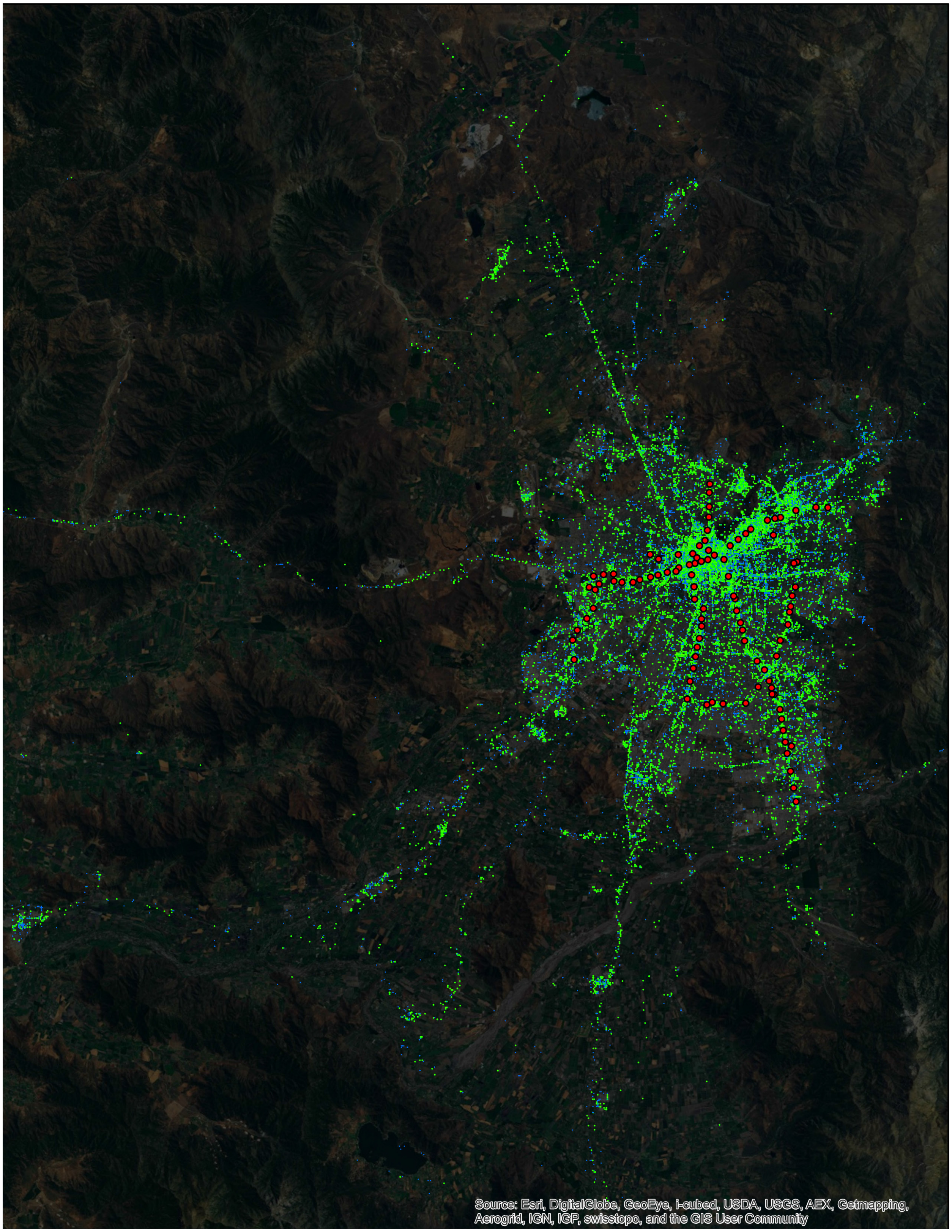


Figure 3.2. Distribution of observations for the initial dataset in the Santiago Metropolitan Area. In blue, foursquare locations. In green, twitter geo-located content (excepting Foursquare). In red, the Santiago Metro Stations reported through Foursquare.

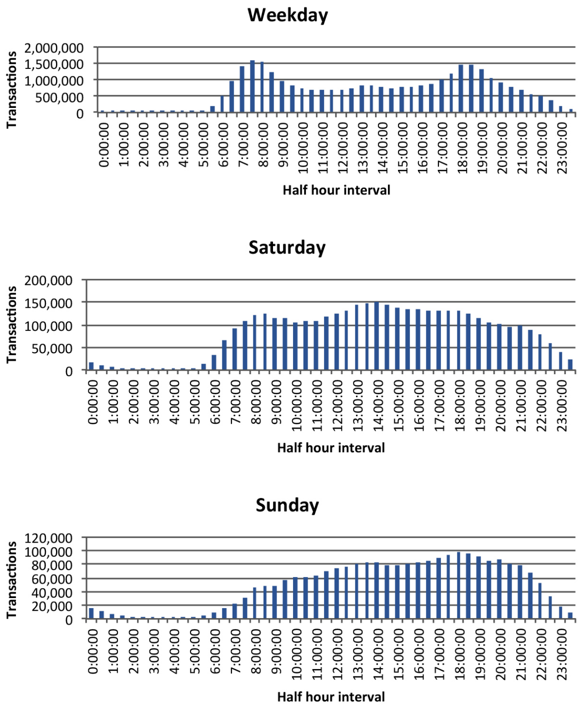


Figure 3.4. Average distribution of smartcards transactions by hour in Transantiago, during the week and week-ends, Munizaga et. al. (2013).

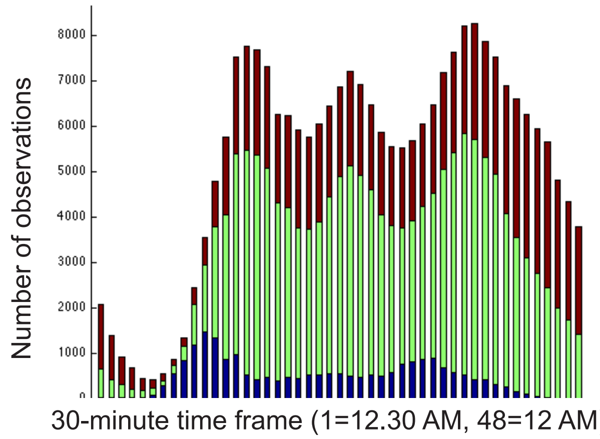
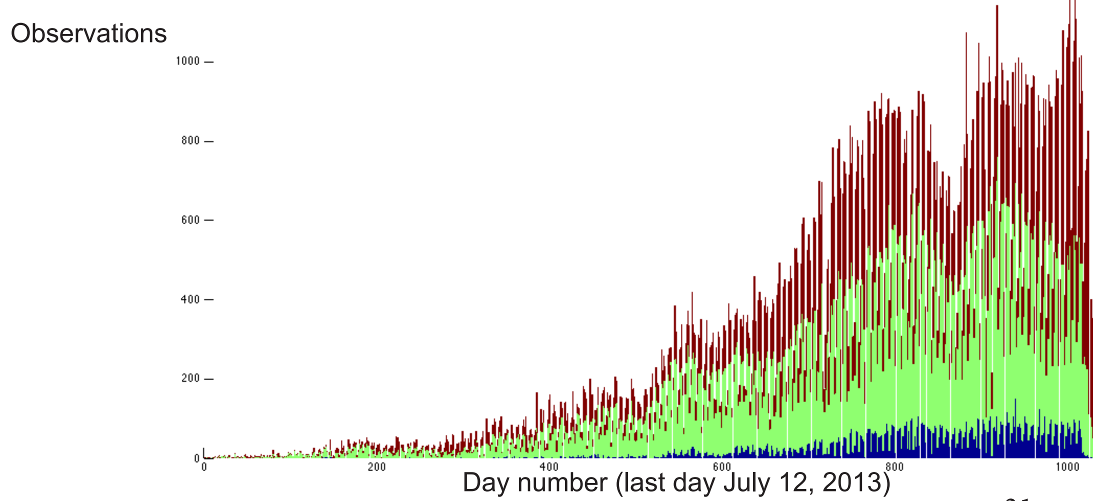


Figure 3.5. Number of observations from the initial dataset registered at each 30-minute time frame.

Colour Code (Figures 3.5-3.6).

twitter observations	foursquare observations	Metro observations

Figure 3.6. Time-series for the initial sample.



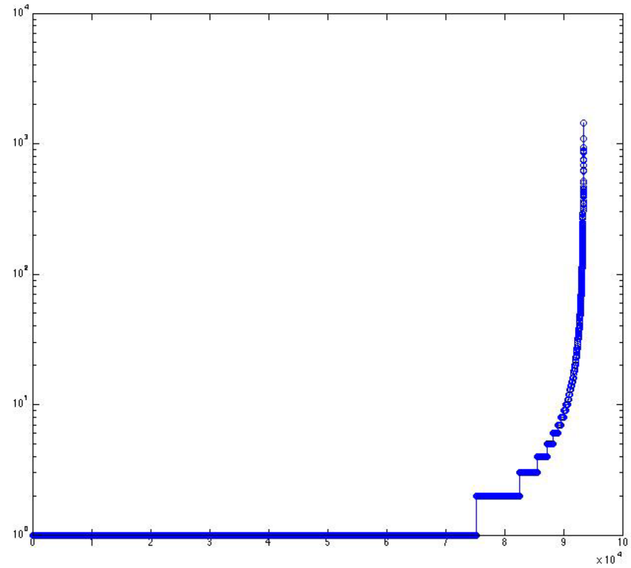
time frame (July 12, 2013). As PCA involves matrix operations with intensive computations, the time frame is reduced to the last 400 days since the most recent observation. A further preprocessing step was undertaken in order to strengthen the activity patterns that could be inferred from the data. Previous evidence from studies based on smartcard data (Figure 3.4), suggested that weekends would present different patterns in comparison to routine activity during the week. After the deletion of the weekend data, the set of observations is reduced to 242.000.

The comparison of time distribution (Figure 3.5) with the mentioned smartcard data shows patterns of activity consistent with the daily travel routines in Santiago, with similar peaks around 7.30 in the morning and 6.30 in the afternoon (and a mid-day peak around 1 PM). Though the assessment of *twitter* data as valid source for inferring commuting patterns is beyond the scope of this work, this comparison suggests promising opportunities for further research on the subject.

The data aggregates shown in Figures 3.5-3.6, include information regarding the source of the *tweet*, indicating wether it was generated from a Metro *foursquare venue*, other f*oursquare venues*, or a *twitter status* without any *foursquare* content. The probability among unique locations (Figure 3.7) shows a long-tail distribution. This is expected considering that highly precise twitter geo-located content[6] comprises roughly one-third of the data (and therefore locations usually with one observation), while the remaining two-thirds are comprised by unique *foursquare venues* that can concentrate

6 Up to eight decimal places in decimal degrees measure (latitude-longitude), see https://dev.twitter.com/docs/platform-objects/tweets#obj-coordinates.

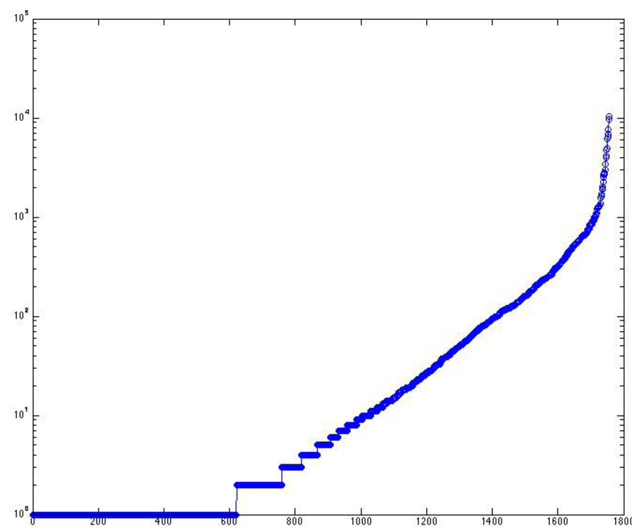
Number of Observations



Location rank by number of Obs. (Ascending).

Figure 3.7. Number of observations by location before the aggregation by spatial pixels process.

Number of Observations



Location rank by number of Obs. (Ascending).

Figure 3.8. Number of observations by 1000 x 1000 meter pixel after the aggregation process.

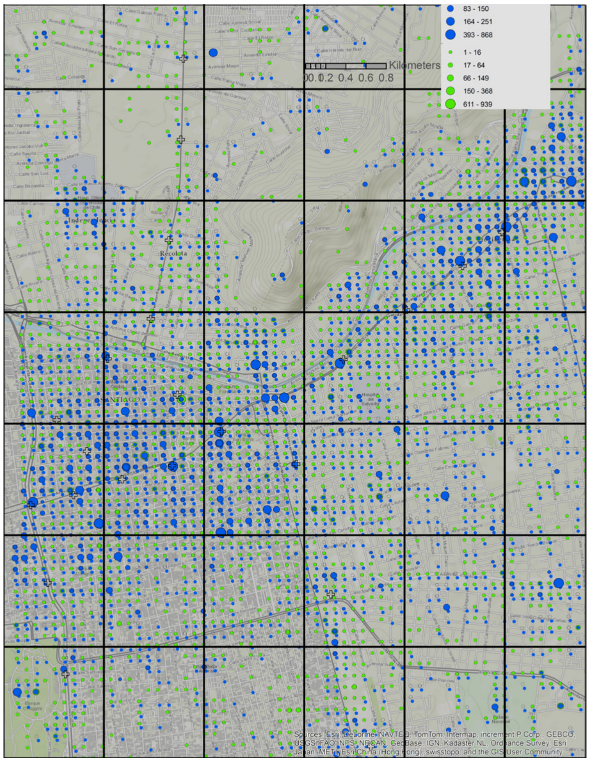


Figure 3.9. Downtown of Santiago with the imposed aggregation lattice, with number of observations by location (sources: blue=foursquare (f), green=tweeter (t)).

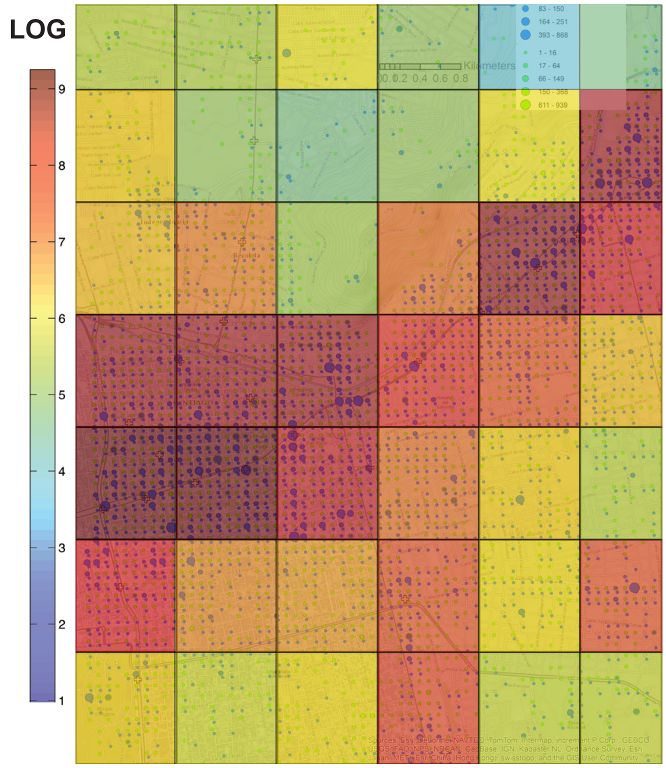


Figure 3.10. Activity a by pixel p, corresponding to the overall sum observations (tweets) within each grid cell.

Figure 3.11. Map of the Metropolitan Area of Santiago with the aggregation lattice.

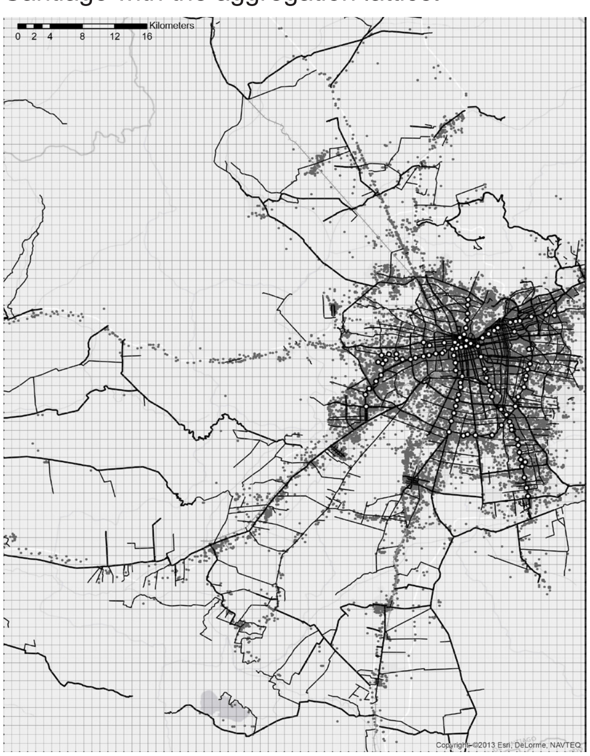
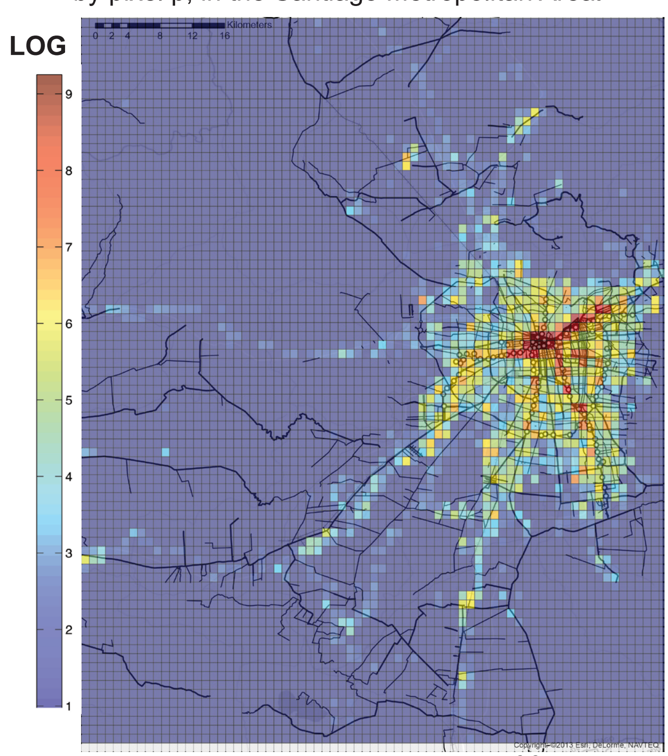


Figure 3.12. Activity (aggregate number of tweets) by pixel p, in the Santiago Metropolitan Area.



a high amount of observations (especially the Metro venues). The presence of several locations with low activity produces extremely large and sparse matrices, that exponentially increase the time spent on complex computations of the nature of PCA. In order to reduce the matrix of locations to be compared, a 1000 x 1000 meter lattice was imposed as level of aggregation[7] (the limitations of this level of aggregation are discussed at the end of this chapter). From this aggregation criterion, the following definitions are provided:

- A pixel p_i corresponds to a unique area contained by a square of 1000 x 1000 meters, defined by the projection of a regular lattice on the Metropolitan Area of Santiago. The centroid of a pixel is located in the intersection of the arithmetic mean for the latitude and longitude values of all the observations contained within it.

- The activity of a pixel a_i is defined as the number of observations contained within the area of a unique pixel.

Eigenvector Centrality Algorithm

The main goal of the *Eigenvector Centrality Algorithm* is to identify locations regarded as central due to their high level of social interaction with other locations that are central by the same token. The methodology applied in this work follows recent work on the study of connectivity between places according to measures of social interaction. Of particular relevance is considered the approach developed by the Urban Livehoods project *(Cranshaw, Schwartz, Hong, Sadeh, 2012)*, for the study of foursquare data in Pittsburgh[8]. According to their definition, the social proximity between a given pair of places is determined by their activity histogram similarity. For the specific model applied here, the affinity between places is determined by the sole overlap of activity; that is, the number of times two given locations were visited by the same users. Afterwards, PCA is performed on the *LxL* locations matrix, and the first *Eigenvector* is used as indicator of the *Power Centrality* score for each pixel.

The starting point is the construction of the aggregated *Activity Matrix* by users. From this activity matrix a *Social Affinity Matrix* between places was calculated, using the intersection similarity measure for comparing vectors of activity distribution by location. Both matrices are defined as follows:

- From the set of 242.000 observations *M*, we have a set of unique *twitter* users identified by their unique ID. The total *Activity Matrix T* is defined as the bipartite matrix for which each element $t_{i,j}$ indicates the number of observations for a user u_i at the pixel p_j. In this case, the total activity matrix *T* contains observations for 1532 users and 1756 pixels.

7 With latitude the orthogonal projections vary in magnitude. As reference were considered the conversions for 31 degrees of Latitude. As Santiago is Located at 33 degres south, these measures varied slightly, setting the actual measure of the pixels slightly below the 500 meter divison. For effects of this analysis, this difference (of less than three meters) was considered as negligible. For the conversion from decimal degree to meters, the following source is consulted: http://www.tceq.texas.gov/gis/geocoord.html

8 Now extended to several American and Canadian cities available on their website www.livehoods.org.

	p_1	p_2	p_3	.	p_{1576}
u_1	$T_{1,1}$	$T_{1,2}$	$T_{1,3}$	.	$T_{1,3818}$
u_2	$T_{2,1}$	.	.	.	.
u_3	$T_{3,1}$	.	.	.	.
.	.	.	.	.	.
u_{1532}	$T_{868,1}$	.	.	.	.

- The 1756 x 1756 *Social Affinity Matrix A* corresponds to the intersection of vectors of the total *Activity Matrix T.* For each pair of pixels $p_{i,j}$, their corresponding intersection similarity $A_{i,j}$ is defined as follows:

$$A_{i,j} = \sum_{n=1}^{1532} min(T_{ni}, T_{nj})$$

*definition of intersection similarity from *(Cha, 2007).*

After setting the diagonal of the *Social Affinity Matrix* to zero (in order to prevent double count of activity)[9] and normalizing the matrix by its highest value, the first *Eigenvector* is computed. From the application of the Eigenvector Centrality model, two measures of centrality are obtained for each pixel:

- The *Degree Centrality* for each pixel, D_i, corresponds to the sum of its intersection similarity vector (or the overall sum of activity overlaps with other pixels) :

$$D_i = \sum_{J=1}^{1756} T_{i,j}$$

- The *Eigenvector Centrality* for each pixel, E_i, corresponds to their corresponding score in the first *Eigenvector* of the normalized (by its highest value) *Social Affinity Matrix A*.

Results

For their comparison, these measures are normalized (by their maximum value) and then converted to a logarithmic scale (based on the natural logarithm). The effect of these transformations can be seen in Figure 3.13.

Figure 3.14 shows the output of the *Eigenvector Centrality* computation in the form of a gridded map (representing each pixel), while Figure 3.15 shows the interpolation map (using Natural Neighbour method) derived from the latter, using the centroids of the pixels with their corresponding *Eigenvector Centrality*. The histogram of the interpolated map approximates a normal distribution (Figure 3.16). The blue area, defined close to the mean of the distribution, detaches an area that

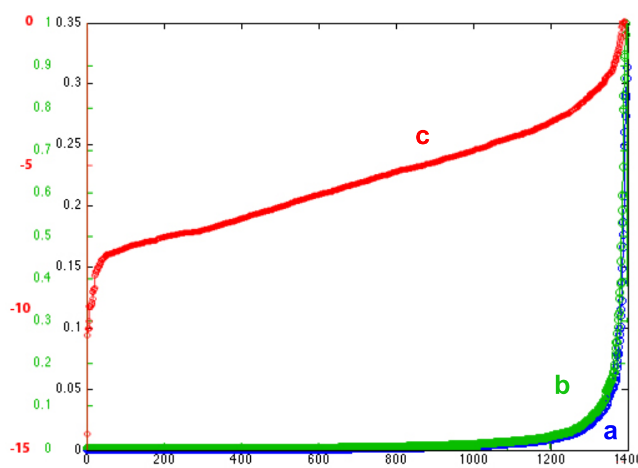


Figure 3.13. Eigenvector scores for the set of 1756 locations. a) Raw score. b) Normalized Score. c) LOG of Normalized Score.

9 This is a conventional practice in this type of analysis, as Eigenvector Centrality focuses on the importance of a node as determined by its connections. In the case were flows are not considered to work within themselves, the "weight" of a node (the sum of connections to other nodes) is considered a redundant measure that distorts the analysis. For an example, consult the Trade Network analysed by Wambeke et. al. (2012).

loosely resembles the urban area of Santiago, while the purple tone areas contain roughly 2,5 % of the interpolated pixels, defining a threshold for what could be regarded as the set of most central locations.

Considerable research has been dedicated to the unique properties of *Eigenvector Centrality (Ruhnau, 2000)*, including Bonacich's demonstrations[10] *(Bonacich, 1987)*. However, and as Bonacich acknowledged, *Eigenvector Centrality* could tend to be correlated with standard *centrality* measures such as *Degree Centrality*, fact that does not imply a strict equivalence (Bonacich, 2007).

The results obtained here are in line with Bonacich's observations. A simple regression analysis could lead to conclude -given the strong correlation- that these measures are equivalent (the most central places are mostly those with higher aggregate activity). However, Figure 3.17, that ranks the pixels according to *Degree Centrality*, shows that the variability of the other measures (specially *Eigenvector Centrality*), can be wide for certain ranges of the scale. This is confirmed by the residuals plot of the regression (Figure 3.18), that evidences a general penalization of the *Eigenvector* score for the pixels with low *Degree Centrality*, a wide variability in the middle range, and a relative high score for the pixels in the upper range. Figures 3.20-3.22, that compare the measures of *centrality* in a set of interpolated maps (using a common normalization, scale transformation and colorbar[11]), evidences how *Eigenvector Centrality* alters the *centrality* balance in comparison to the standard measures.

10 Based on idealized network structures, not observed social networks..

11 However with not the same colouring scheme of Figure 3.16. This is only a graphical change, that did not involve data transformations.

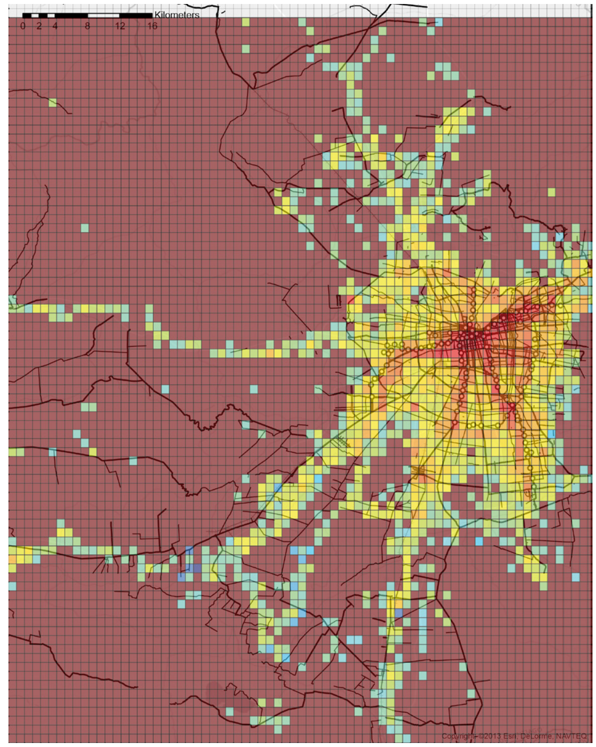


Figure 3.14. Eigenvector Centrality Map of Santiago (LOG scale).

The combined analysis of the *centrality* scores and the interpolated maps, shows that there is indeed a reduced set of central locations, for which *centrality* is increased by the *Eigenvector* measure. For the case of those locations in the mid-range of *Degree Centrality*, that tend to be situated on rather peripheral areas, this suggests the presence of the *"tunnel effect"* hypothesized by Graham and Marvin. The details of the most central locations is included in Appendix A.

For testing the second spatial hypothesis, that distance does not constrain the centrality dispersion across the urban area, the urban eccentricity measure of Yang et. al. (2013) is used. This measure is the ratio of the area of the city core against the entire urban area, measured with directional distributional ellipses. While Yang et. al. used population density for defining the borders of these ar-

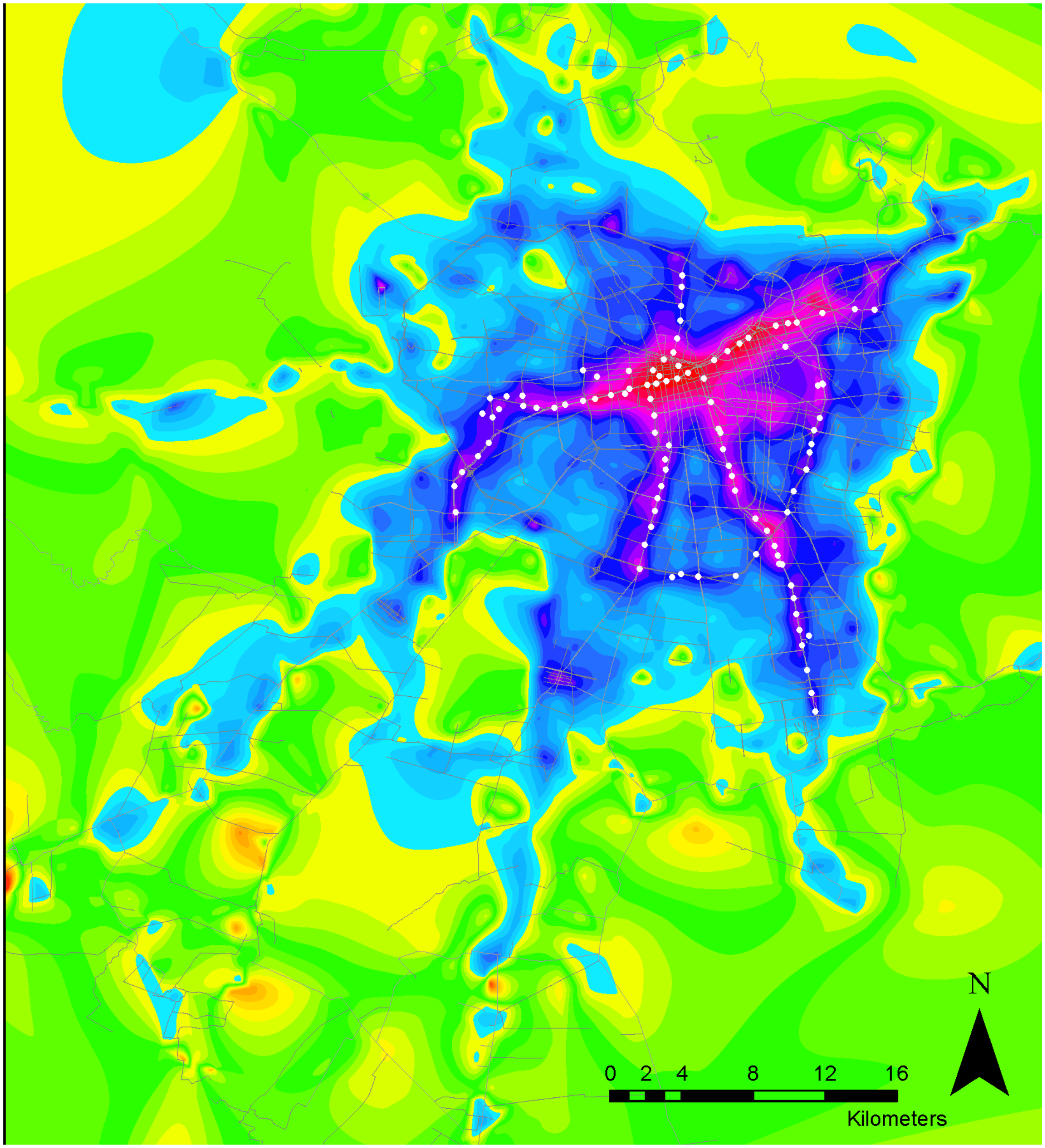


Figure 3.15. Eigenvector Centrality Map of Santiago (normalized-LOG score),produced through Nearest Neighbor interpolation (from the aggregation pixels centroids).

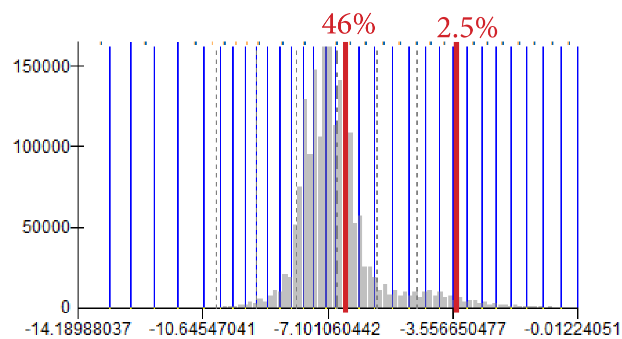


Figure 3.16. Frequency distribution for the Nearest Neighbour interpolation applied to the Eigenvector Centrality Map of Santiago. The blue lines represent the intervals for the colour categories.

Eigenvector and Degree Centrality Score
(LOG of the normalized measure).

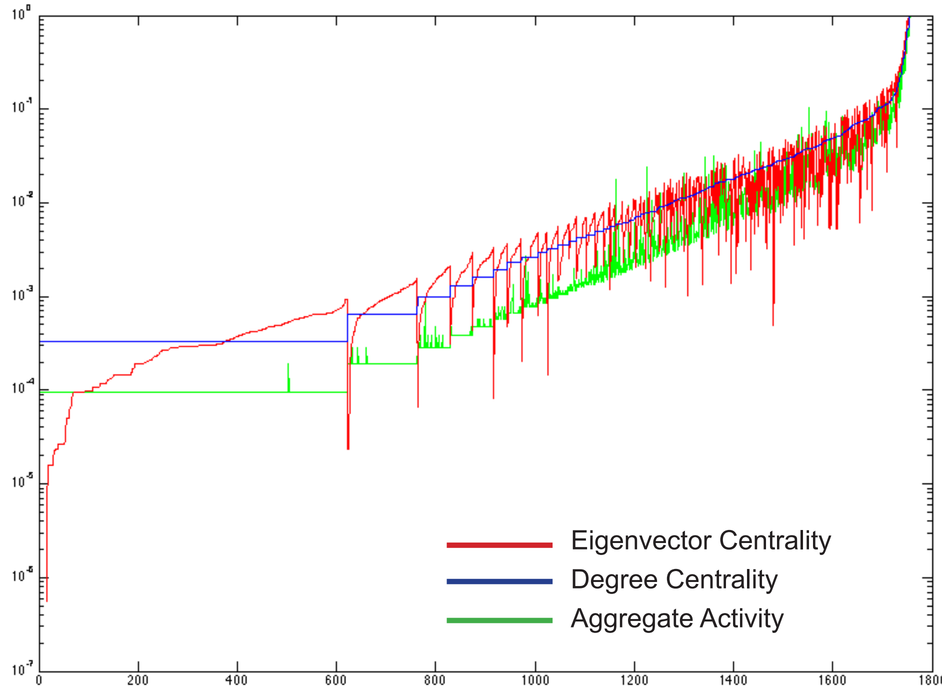


Figure 3.17. Comparison of Degree Centrality (aggregated activity) with Eigevenctor Centrality score by pixel.

Eigenvector Centrality Score
(LOG of the normalized measure).

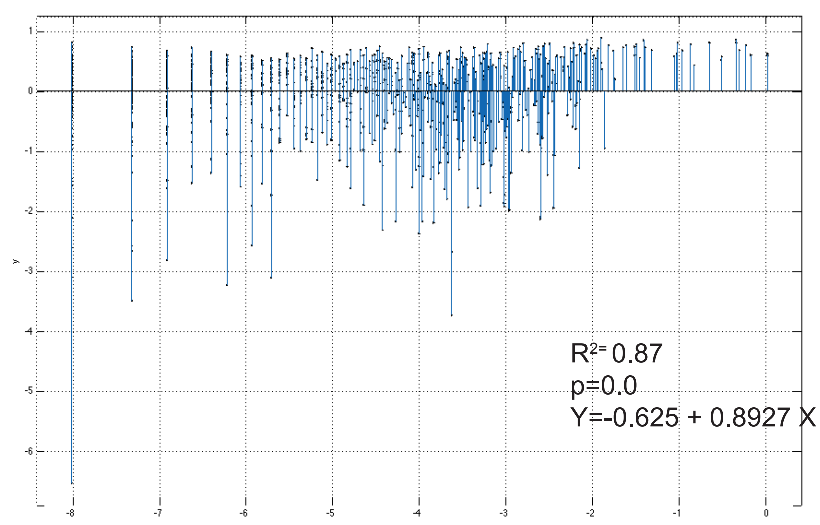


Figure 3.18. Residuals of the linear fit of Degree Centrality against Eigenvector Centrality.

Figure 3.19. Distance from the most central (Eigenvector score) pixel, obtained through Nearest Neighbour (NN) interpolation. Points represent the centroids of the pixels.

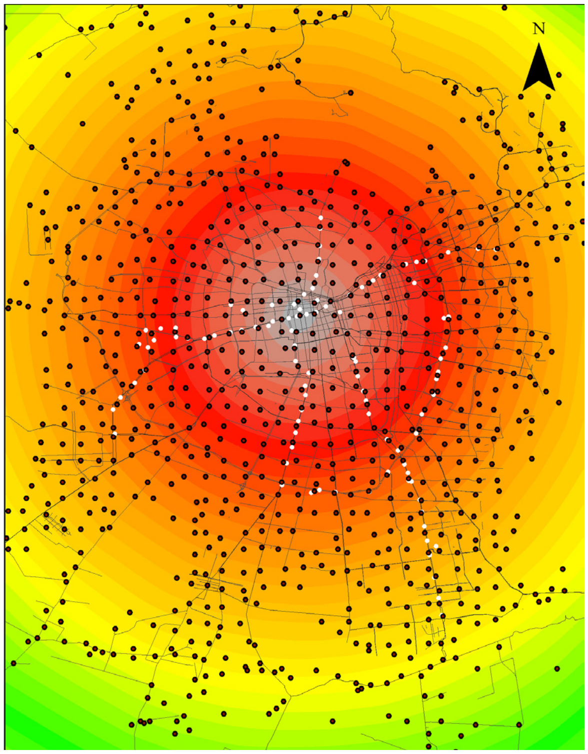


Figure 3.20. NN interpolation for Aggregated Activity by pixel (normalized-LOG score).

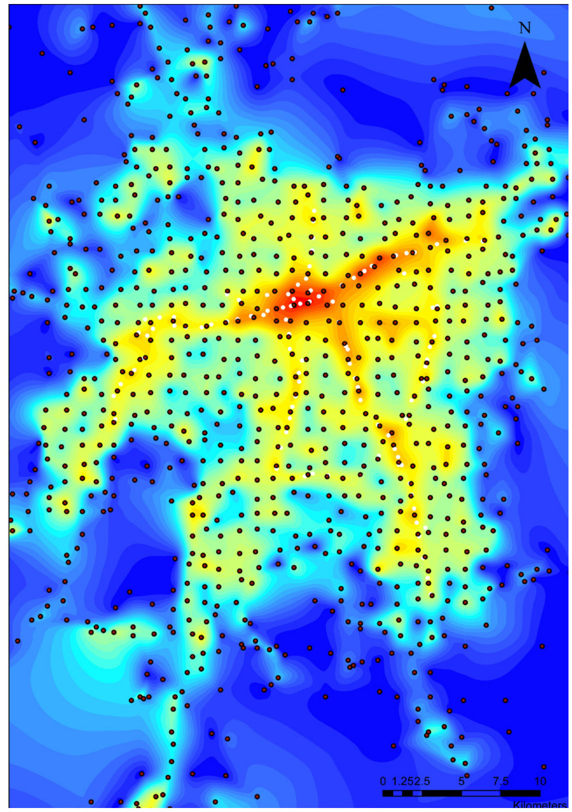


Figure 3.21. NN interpolation for Degree Centrality by pixel (normalized-LOG score).

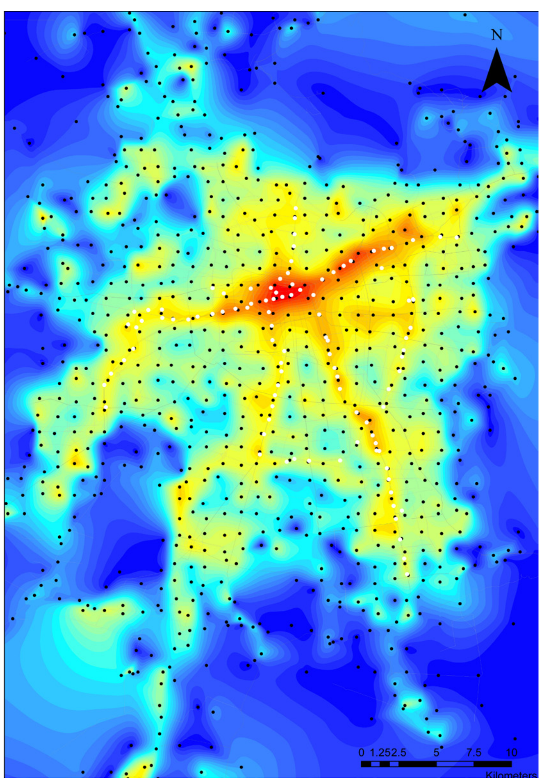
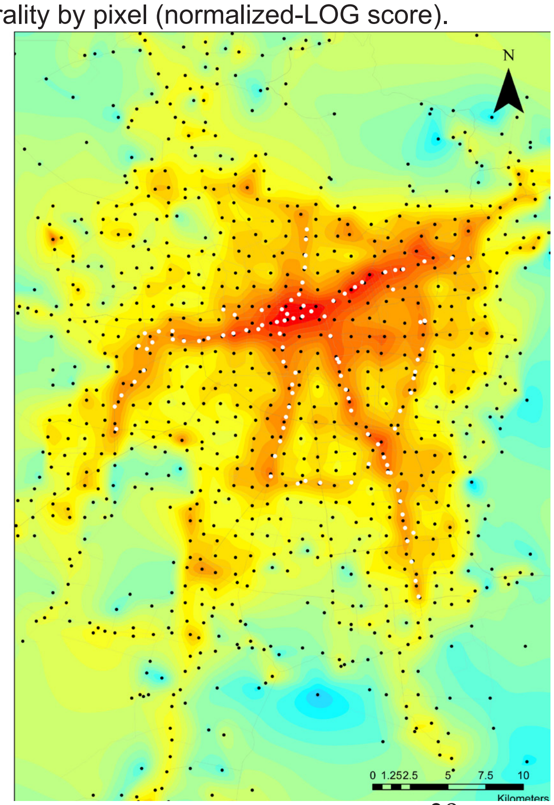


Figure 3.22. NN interpolation for Eigenvector Centrality by pixel (normalized-LOG score).



eas, here their procedure is replicated using the *Eigenvector centrality* score. The threshold values that determined the *"core"* and *"urban"* centroids are the same as those identified in the interpolation map (Figure 3.16). The corresponding ellipsoids, weighted by the *Eigenvector centrality* score and restricted to two standard deviations, are shown in Figure 3.23. The ratio of both ellipsoids is 18%. In comparison to the calculations of Yang et. al. -based as indicated, on population density-, Santiago is situated above cities such as Boston, Chicago or Atlanta (that register values around 15%), but well below other polycentric conurbations such as Los Angeles and San Francisco[12] (that score above 25%).

These results show that in Santiago there is indeed a reduced set of powerful well connected places, and that although the city does not present a completely polycentric structure, clearly distance does not constrain the dispersion

12 This study was carried for U.S. cities.

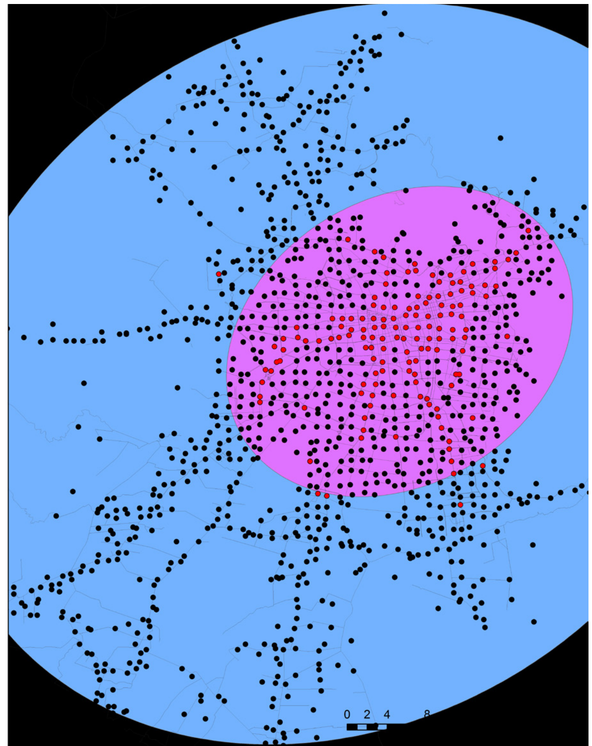


Figure 3.23. Directional Distributional Ellipsoids for central pixels (purple) and urban area pixels (blue), weighted by the Eigenvector Centrality Measure (ellipsoids bounded by 2 standard deviations). In red the centroids belonging to the central pixels (according to the thresholds shown in Figure 3.16).

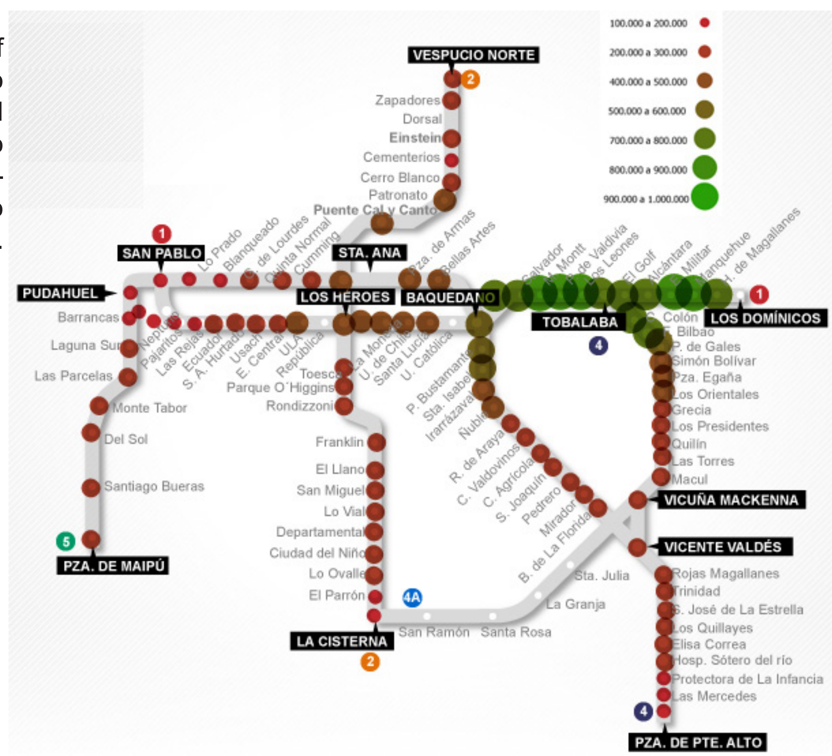


Figure 3.24. Average Income of residential blocks around Metro Stations (according to a weighted function that relates the station to the centroid of a comuna and its average income, as 2006). By Centro de Estudios Publicos (CEP), 2013.

of these *centralities*. When compared to the average income present around the Metro stations, the concept of *"premium networks"*, key to the Splintering Urbanism Theory, turns contradictory (Figure 3.24). The high *centralities* identified by the *Eigenvector* measure show that not only the *"valued"* places (in terms of income) are being connected to each other.

Certainly, this sole comparison is not exhaustive and conclusive by itself. It can be argued that many other factors are affecting the *centrality* scores, and thus their apparent mismatch with the areas of high income. This situation applies for areas where in general the residential density is low compared to the amount of employment they host (such as the historical downtown of the city). The travels between these areas and high-income areas would apparently confirm the connections between *"valued"* areas to the service of *"valued"* users. However, as it can be seen in Appendix C, the distribution of services, population density and other spatial indicators are far from being articulated in a coherent pattern across the Metropolitan Area. Areas such as Puente Alto or Maipu , densely populated mid to low-income areas, register a relative a high *centrality*, suggesting a strong connection with the other central areas of the city. Other places, such as *"Paradero 14"* (B. de La Florida station), presenting similar residential patterns, have conformed new centralities by concentrating retail and services space.

Naturally, the main concern with the validity of the result is related with the excessive bias that could be present in favour of the Metro network. It can be argued that this analysis left aside a large proportion of the population that does not use Metro (including the elites), that are effectively served by other types of by-passes.[13]

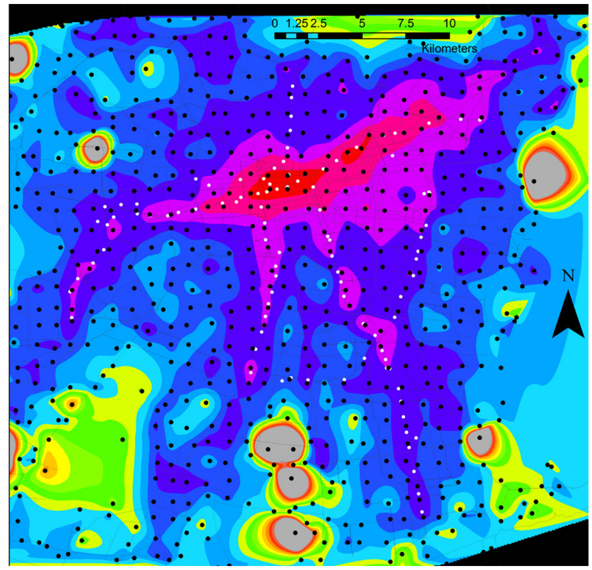


Figure 3.25. Results of the Power Centrality algorithm applied to a different dataset, corresponding to 110.000 tweets produced during the first week of April 2011.

A comparison with an alternative dataset shows however that the *centrality* patterns remain basically unaltered. This dataset, composed by 110.000 tweets obtained during a week of early 2011[14], without any special filtering procedure (Figure 3.25), was processed using the same *Eigenvector* algorithm, including the normalization and the logarithmic transformations. No significant change can be observed in high-income areas (that could be expected to increase their score in this unbiased dataset), while slight decreases in the centrality of peripheral locations can be noticed. Overall, it can be inferred from this type of data, that either the imprint of the Metro network in the overall mobility patterns is considerable, or Metro has

---

13 Such as the extensive Urban Highways system that were built in Santiago during the first half of the past decade. They are run by private operators with a free-flow toll system.

14 By the San Francisco based spatial analyst Eric Fischer. More details about the dataset and his work are included in Appendix B.

been efficient in following the urban development patterns during the last years. More details about the results of the algorithm performed on this alternative dataset can be found in Appendix B.

Limitations and Discussion

- In a research based on data coming from non-convential sources, and specifically Social Media, the sample biases are a significant issue to consider. The *"Digital Divide"* hypothesis probably constitutes the most widely discussed topic regarding this issue. This hypothesis states that a big share of the population would not have access to expensive and sophisticated electronic devices that provide the connection to these networks. This theory became extremely popular at the beginning of the 90's when the sociologist Manuell Castells coined the term *"Fourth World"* to make reference to populations that were being excluded from the global communications and exchange networks.

  This issue is specially sensitive in Latin America, the region in the world with the most skewed income distribution *(Martin, 2009)*. Chile, despite leading the region in income and development indicators, is not excluded from these concerns, considering that its wealth distribution ranks amongst the worst in the world.[15] Quantitative analysis have demonstrated that in Latin America, income is the most important factor that constrain digital access *(Martin, 2009)*. However, the explosive adoption of mobile technologies in developing countries (see Introduction) suggest that today income may not be as determinant as only few years ago. Appendix B contains a short regression analysis based on the alternative *twitter* dataset (Figure 3.25) and census data (year 2002). The results suggest that this type of data represent activity rather than demographic patterns. Though it cannot be asserted that high income users are not over-represented (as their income information is not available), the main concern was that places with predominantly high income population would be concentrating most of the *twitter* activity. According to the correlation analysis, the likelihood of that specific bias is low.
- The second main source of bias is the "*tweeting"* behaviour of the users themselves. From the sample used in this research, it can be observed that the distribution of tweets among users approximates a power law (Figure 3.26). This research worked under the assumption that this small set of highly active users where not going to bias the sample towards specific locations. By selecting users that were at Metro stations (reported via *foursquare*) some sources of bias are avoided. It is unlikely that *bots* representing commercial services where included among the users (as the development of applications of such kind is extremely difficult for *foursquare*). Regarding the veracity of the locations reported by the *foursquare* users, the risk of bias associated with this specific issue has yet not been ascertained. Recent research *(Noulas et. al., 2011)* has shown how the transitions between *venues* do follow logical patterns (for example, from train to train station, or

15 According to the World Bank, chile tops the region in GDP per capita, but its income inequality (measured by the GINI index) is comparable to African economies such as Zambia. See World Bank Dat: http://data.worldbank.org/indicator/

from one airport to other one).

As this research did not focus on written content of the messages, the precautions needed for the semantic analysis[16] were not considered necessary. See Szell, Grauwin and Ratti (2013) for a more detailed discussion on this sources of bias.

- A critical aspect to be improved is the unit of aggregation. For analytical simplicity, the lattice was chosen in comparison to clustering methods or existing transportation divisions[17]. Spatial clustering is not a trivial task, as it requires a deep understanding of the purpose sought for the areas the be defined[18]. Though the data analysed is spatial, the correspondence with fixed spatial indicators (such as population density), is low (see Appendix B). On the other hand, it was considered a risk to associate the dataset to existing transportation divisions given that the data has not been yet validated with transportation data.

The election of the size of the grid (1000 meters pixels) was based on the experience of previous analysis based on Smartcard data (Fischer Montt, 2010). According to their calibrations, the walking distance in Santiago is of one kilometer[19]. Naturally the lattice as aggregation method has serious caveats, being probably the most important the inherent properties of the data that could be being overlooked.

16 The topics being discussed by the users.

17 Transportation analysis performed by the Chilean government usually are carried on the ESTRAUS divisions, developed by local transportation analysts.

18 During the last years the research dedicated to the so-called spatio-temporal clustering models has grown explosively. Usually these techniques resort to advanced clustering algorithms such as OPTICS or DBSCAN. For a comprehensive survey on spatio-temporal clustering techniques consult: Maimon, Oded Z., and Lior Rokach. 2010. Data mining and knowledge discovery handbook [electronic resource] / edited by Oded Maimon and Lior Rokach: New York ; London : Springer, 2010.

19 The distance after which a person would either access a transit top or choose other mode of transportation (car, taxi, etc.)

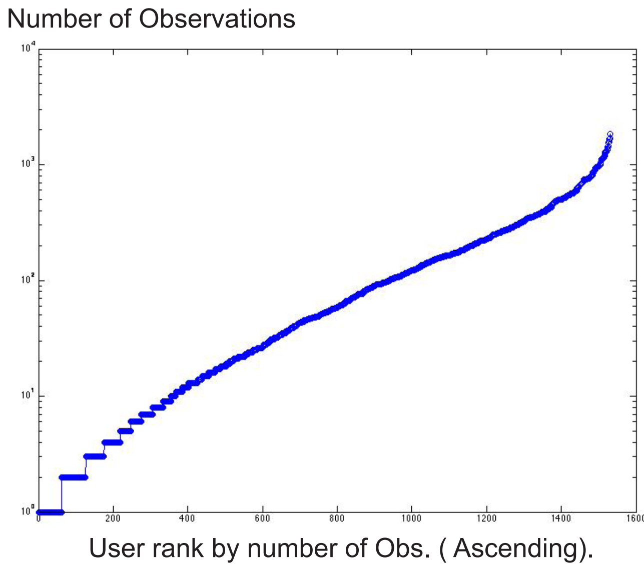


Figure 3.26. Semi-Log plot of observations by user in final dataset.

Figure References

Figure 3.4. Average distribution of smartcards transactions by hour in Transantiago, during the week and week-ends, Munizaga et. al. (2013).
(Munizaga et. al., 2013) Munizaga, Marcela, Devilaine, Flavio, Navarrete, Claudio, and Diego Silva. "Validating travel behavior estimated from smartcard data." Paper presented at the International Choice Modelling Conference, Sydney, Australia, July 3–5, 2013.

Figure 3.25. *Echenique, Juan, and Urzua, Segio. 2013. "Desigualdad, Segregacion y Resultados Educacionales. Evidencia desde el Metro de Santiago." Centro de Estudios Publicos (CEP) Puntos de Referencia. no. 359.*

CONCLUSIONS.

Power Centrality as a Synthetic Relational Measure

In this work *Power Centrality* was introduced with the intention of providing a synthetic measure for characterizing the hierarchy of places within Metropolitan Areas. The hope is that in the future, this or other relational measures could provide new insights for capturing the complex structure of contemporary cities. The case-study, Santiago, indeed shows that given the spatial distribution of population, services, retail or other proxies of place *centrality*, hardly any tractable regularities can be inferred in terms of Urban Structure (Appendix C).

The relational approach undertaken here, evidenced some *centralities* that could not have been noticed through conventional measures. As it can be seen in Appendix C, mid to low-income suburbs, satellite cities, and isolated shopping centers (outside the high income areas), surprisingly show a relative high *centrality*. Following the definition of Massey, it could be asked: What means for poor, peripheral or apparently disconnected locations to be highly related to *"powerful"* places? Are they indeed relevant places within the urban system that have been overlooked by conventional indicators of *centrality*, or just *"in transit"* areas without further impact in the urban hierarchy? Here the inclination is for the first explanation, however, it is acknowledged that further conceptual is analysis is required, as *Power Centrality* has been offered as measure for characterizing urban hierarchy, not as an explanatory model.

Limitations of the Splintering Urbanism Theory: The Premium Infrastructures

Through a three-legged process it was shown how for a city like Santiago, the spatial hypothesis of Graham and Marvin could prove to be true, but incomplete in its causal explanations. Santiago indeed showed an Urban Structure -in both a biased and a random sample-[1] far from the mono-centric ideal Graham and Marvin argue against. This is in general terms could be considered an expected result, given the current trend of Metropolitan Areas around the world to evolve towards polycentric patterns of organization, a phenomenon which as indicated in the introduction, has been noticed and studied since the early 70's.

The main contribution of this research has thus been the analysis of that phenomenon from the viewpoint of a relational *centrality* measure, and most importantly, to challenge part of the causal explanations of Graham and Marvin. In their theoretical framework, services such as a Metro do not play a relevant role -or plainly do not play any role at all- in the configuration of post-soviet cities. However, Metro in Santiago showed how a public infrastructure system, in a country that went trough a deep process of *"network unbundling"*[2], still has a large impact in the overall mobility structure of the city. Indeed, the notion of *"valued users"* has been subverted, as now areas with low-in-

1 By "biased" meaning the sample that was obtained by filtering through Metro foursquare users, and by "random" Fischer's sample obtained from twitter without any special filtering process beyond the biases induced by the users preferences or the own technical constraints of the service.

2 Since the late 80's the country undertook deep privatizations reforms, that included telecommunications, sanitation, energy and transportation sectors. Most of them were regarded as pioneering reforms (for good or bad) in Latin America.

come or non-powerful users are being effectively served by the bypasses theorized by Graham and Marvin. There are multiple causes that explain the emergence of this counterexample of *premium infrastructure*, among them the public will to remedy its own planning policies carried just years before[3], or the conjuncture of Transantiago, that dramatically changed the role of Metro[4] in the transportation system. Beyond this contex-specific explanations, the concept of *"premium networks"* introduced by Graham and Marvin deserves a deeper discussion, as the infrastructure *"unbundling"* they described certainly is not necessarily a linear process where to privatization follows the market segmentation and then the exclusion of non-valued users[5]. State intervention, private-public partnership, and the detection of market niches in former unvalued sectors[6], are among the intricate processes that explain the generation of urban bypasses that serve a wide array of users.

3 Before the period of greatest expansion of Metro during the 2000's, a massive social housing policy based on combination of public subsidies and private real-estate investment, relocated thousands of Santiaguinos from relatively central locations to new complexes on the outskirts of the city. According to estimations, one in six santiaguinos where relocated through this policy. For more information consult the work of Alfredo Rodriguez and Ana Sugranyes: http://www.scielo.cl/scielo.php?pid=S0250-71612004009100004&script=sci_arttext

4 Also it is worth to mention that Metro is technically a private company, but owned in its majority by the state. The main decisions on infrastructure investments are not determined by profits (as the capital is usually contributed by the state).

5 As they acknowledge, though briefly, in the concluding section of their book.

6 And also in an historical perspective, path-dependency: What would have been of Metro without Transantiago in the way it was implemented?

From the Space-Time Compression to Mankind's Central Nervous System

At the time *"Splintering Urbanism"* was published, Internet, and digital technologies in general, seemed as tools that would radically alter the patterns of human behaviour, including the relationship with space. This not only included techno-utopian interpretations (such as those of Francis Fukyama or Nicholas Negroponte), but also dismal scenarios where the main concern was the increasing gap between a digital elite and those excluded from these networks. During the last decade, and specially since the advent of Social Media and smartphones, digital devices and Internet have become accessible to former excluded segments at exponential rates *(Pentland, 2009)*.

The paradigm is swiftly changing from the *"Space-Time compression"* (with all the positive and negative effects associated) to the *"Mankind's Central Nervous System"* coined by Pentland (Pentland, 2009). This is not to say that human activity on space has not been significantly transformed during the last decades, but probably these transformations still are driven mostly by physical networks than by their virtual counterparts. The underlying theory of this work is that the major revolution brought by digital networks is not the large scale reshaping of human activity on space, but the enhanced capacity for observing our own behavioural patterns on it. How this will affect the planning decisions, and also the individual behaviour (as this information is becoming available to wider audiences) is one of the mayor questions for the following years.

<This page intentionally left blank>

APPENDICES.

Appendix A:
Images of Selected Locations

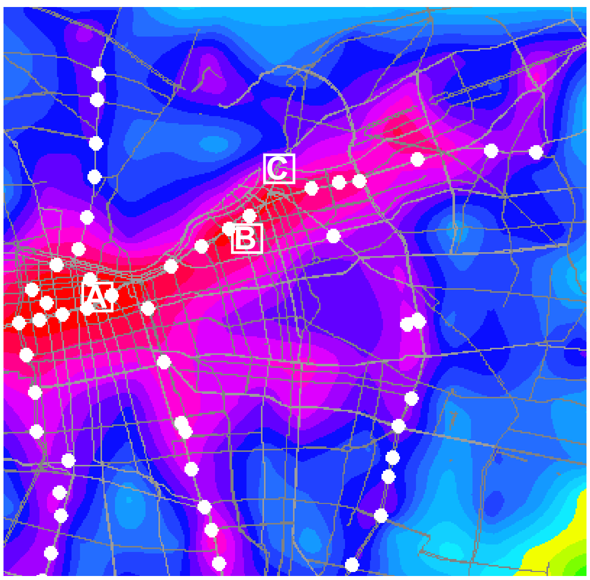


Image 1.3 Avenida Providencia (to the right of the Park, parallel to the river). On the background can be seen Costanera Center and Titanium towers.

Image 1.1 Stock-Exchange in the historical downtown of Santiago.

Image 1.2 Costanera Center, tallest skyscraper in South America, flanked by the second tallest tower in Santiago (Titanium). The area conforms the finance hub of the country.

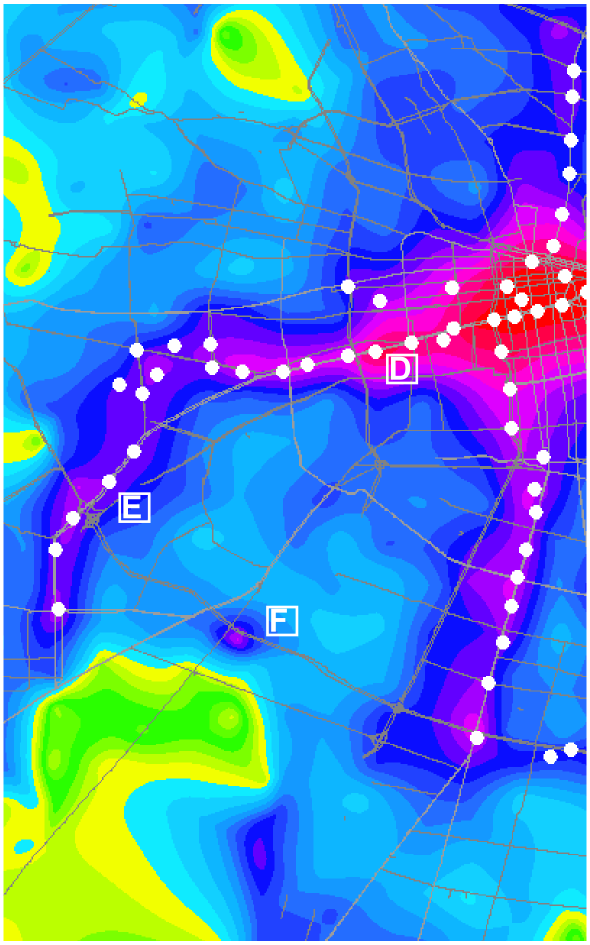


Image 1.4   Estacion Central, oldest train station in Santiago still in service. Nearby are located the main inter-urban bus hubs of the city.

Image 1.5   Metro tracks and stations across Maipu. On the background the “Templo Votivo”, one of the most important churches in the city and landmark of this comuna.

Image 1.6   Plaza Oeste shopping center.

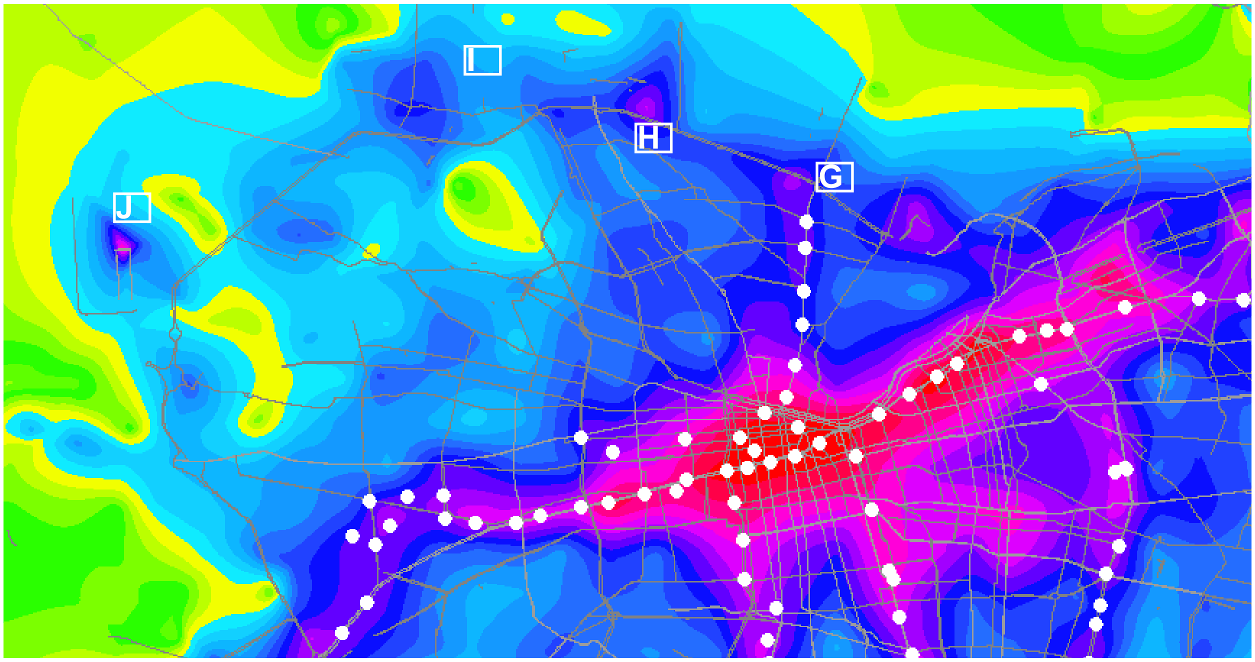


Image 1.7 Ciudad Empresarial ("Business Town"), new office project located on the outskirts of the city. Following the model of business parks, most of its locations are rented by creative and IT industries.

Image 1.8 Plaza Norte shopping center, close to the terminal station of Metro Line 2.

Image 1.9 Airport (AMB-SCL). During 2012 the total passenger traffic reached more than 14.000.000 (Source: Airport Administration, SCL: http://www.aeropuertosantiago.cl/)

Image 1.10 Business Park on the intersection of Santiago's outer ring (Americo Vespucio Highway) and one of the main north-south highways (Norte-Sur).

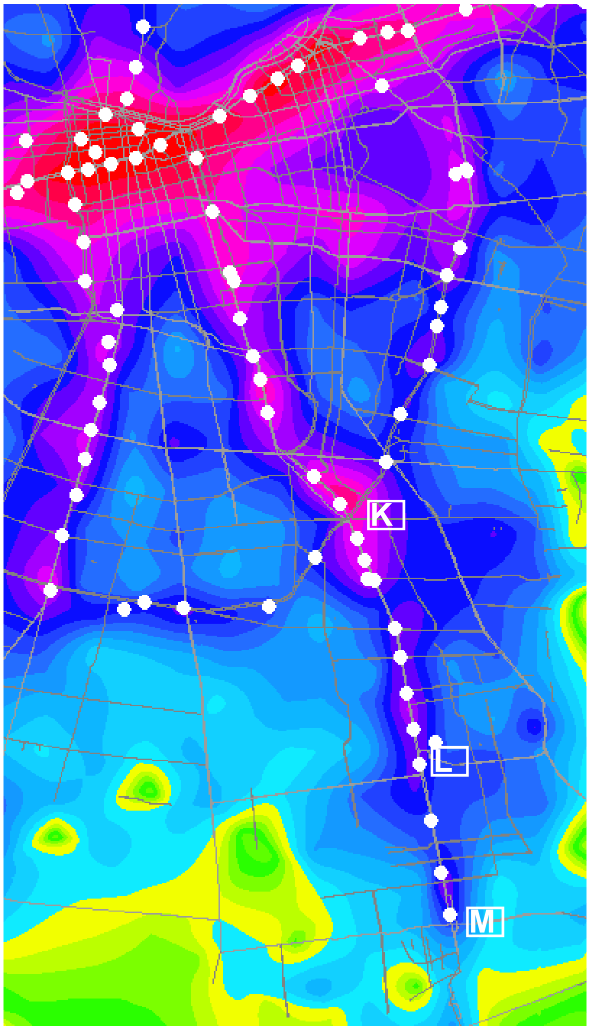
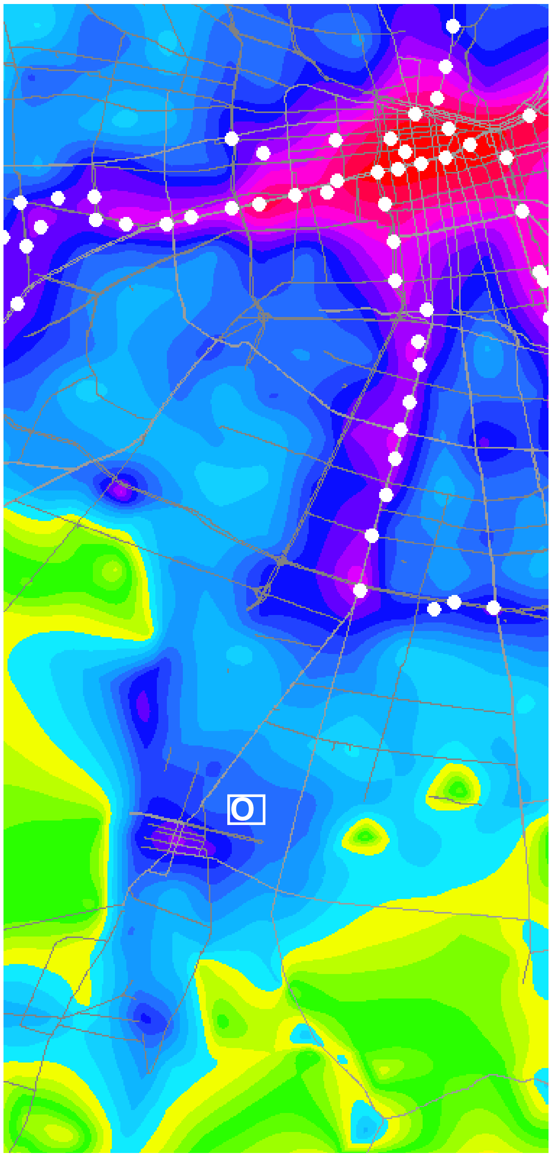


Image 1.11 Metro workshops in Puente Alto, southern suburb of Santiago. On the foreground, surrounding the workshops, can be seen typical social housing blocks from the 90's. On the back-ground, middle-class housing.

Image 1.12 Metro track across Puente Alto, surrounded by single-family social housing (right), and middle-class real estate projects (left).

Image 1.13 Plaza Vespucio Shopping Center, in La Florida, the largest of its type in the country. Located on the intersection of Santiago's outer ring (America Vespucio Highway) with Vicuna Mackenna, the main north-south avenue of the city (area known colloquially as "Paradero 14").



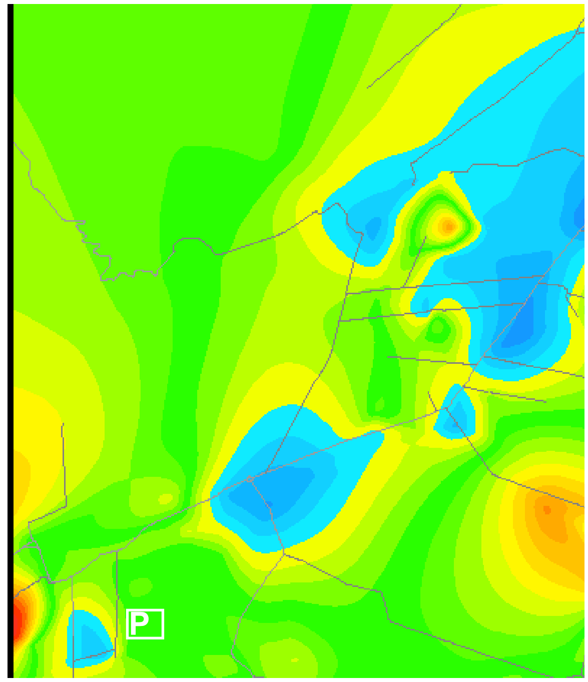


Image 1.15 Aerial view of San Bernardo, looking to the east.

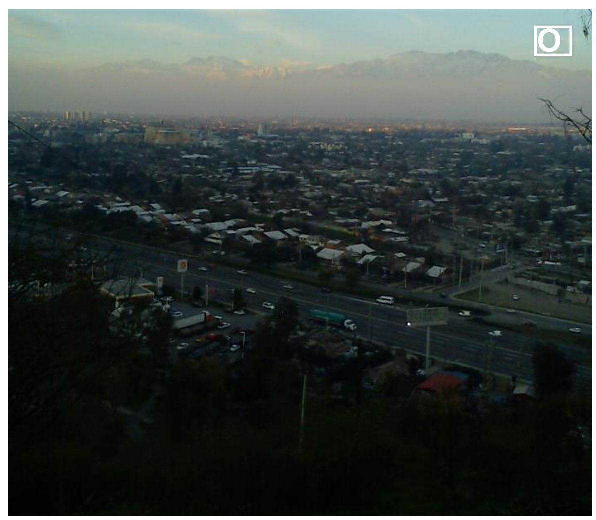


Image 1.16 Panoramic view of Melipilla, 100.000 people city located 20 kilometers south-west of Santiago.

Image Sources

•Image 1 .1. Source: Flickr ID 5379601956. http://www.geolocation.ws/v/L/5379601956/pedestrian-street/en (2010). Retrieved from source on 9 Jul 2013 2:41:53 -04:00.

•Image 1 .2. Own work.

•Image 1 .3. Author: Victor Roman (2010). Source: Wikimedia Commons. http://commons.wikimedia.org/wiki/File:Avenida_Providencia.jpg. Retrieved August 31, 2013.

•Image 1 .4. Author: Jorge Barrios (2007). Source: Wikimedia Commons. http://commons.wikimedia.org/wiki/File:Frontis_Estaci%C3%B3n_Central.jpg. Retrieved August 31, 2013.

•Image 1 .5. Author: Municipalidad de Maipu (2011). Source: Flickr. http://www.flickr.com/photos/maipu/5413596229/. Retrieved August 31, 2013.

•Image 1 .6. Author: Jorge Barrios (2010). Source: Wikimedia Commons. http://commons.wikimedia.org/wiki/File:Mall_Plaza_Oeste_2009.JPG. Retrieved August 31, 2013.

•Image 1 .7. Author: Christian Van Der Henst S (2012). Source: Flickr. http://www.flickr.com/photos/tecnovm64/6775151838/. Retrieved August 31, 2013.

•Image 1 .8. Source: ComunicaExtend(beta). http://www.extend.cl/comunica/2010/02/10/nuevas-ampliaciones-en-tres-centros-comerciales-mall-plaza/ August 31, 2013.

•Image 1 .9. Author: Phillip Capper (2010). Source: Wikimedia Commons. http://commons.wikimedia.org/wiki/File:Terminal_Aeropuerto_Pudahuel.jpg. Retrieved August 31, 2013.

•Image 1 .10. Author: Eduardo Mieres (2008). Source: Panoramio. http://www.panoramio.com/photo/13812736. Retrieved August 31, 2013.

•Image 1 .11. Author: Carlos Rosas (2009). Source: Flickr. http://www.flickr.com/photos/scalibur001/395488749/. Retrieved August 31, 2013.

•Image 1 .12. Author: Jorge Barrios (2007). Source: Wikimedia Commons. http://commons.wikimedia.org/wiki/File:Frontis_Estaci%C3%B3n_Central.jpg. Retrieved August 31, 2013.

•Image 1 .13. Author: Tomas Grau. Source: Panoramio. http://www.panoramio.com/photo/2662250. Retrieved August 31, 2013.

•Image 1 .14. Source: Wikimedia Commons (2012). http://commons.wikimedia.org/wiki/File:Panoramica_de_snbdo.jpg. Retrieved August 31, 2013.

•Image 1 .15. Author: Marcelo Rodriguez (2009). Source: Wikimedia Commons. http://commons.wikimedia.org/wiki/File:Melipilla_Panoramica.jpg. Retrieved August 31, 2013.

Appendix B: Alternative Dataset

The results of the application of the *Power Centrality* algorithm to a different dataset are shown here. This corresponds to more than 110.000 tweets provided by Eric Fischer, the San Francisco based spatial analyst who pioneered the representation of Social Media data on maps.[1] This dataset contains five days of activity during the first week of April 2011. Though it may include observation overlaps with the principal dataset used in this research, a general check revealed that it did not contain any observation produced at *foursquare* locations in the Santiago Metro.

In the Methods and Results section already is shown the interpolation map product of the *Eigenvector* scores calculated on this particular dataset (Figure 3.26). Figure B.1 shows the overlay of the observations from Fischer's dataset and the dataset used in this work, while Figure B.2 offers a comparison the diverse centrality scores by pixel (in the same fashion as shown in the results section). Given that the sample of centroids is not equivalent to the one used in the main body of this research (and thus the normalization applied to the centralities scores differ), a correlation analysis was considered potentially misleading.

Regarding the possible biases produced from predominant residential income at the locations, a regression analysis is presented here, comparing Fischer's dataset against census data from 2002. All the variables were transformed logarithmically (base e). As Figure B.3 shows, there is a negative correlation between Average Income by census block, and population density, being the relationship of relative high strength (R^2=0.35). When population density and Average Income[2]

1 His work is available on his Flickr site, http://www.flickr.com/photos/walkingsf/.

2 The census information contained the number of families belonging to an income classification commonly used in local marketing studies. This information was derived (by OCUC-UC) from the goods declared by the families in the census. The average income per-block was calculated as the weighted average, according to the socio-economic classification assigned to each family (the income for each socio-economic group was

Figure B.1: Dataset used in this work (red) overlaid by the datapoints of Fischer's dataset (green).

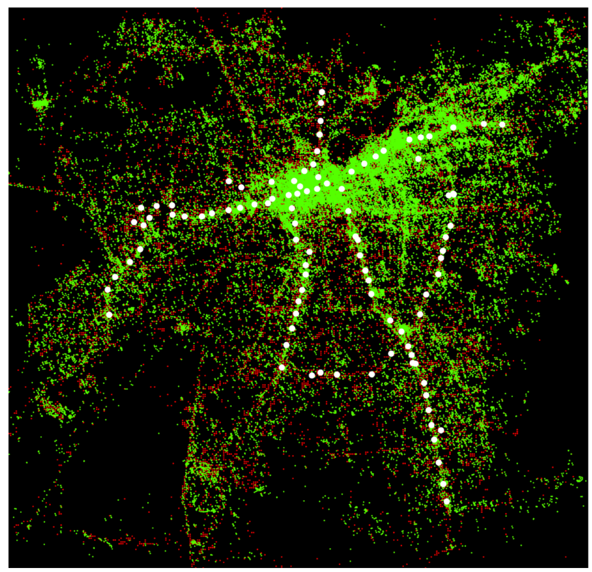
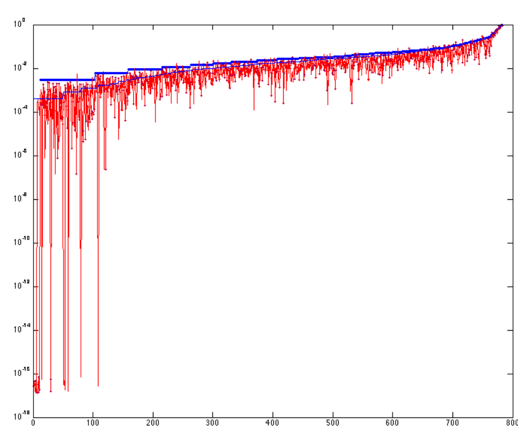


Figure B.2: Comparison of Degree Centrality with Eigenvector Centrality, from Fischer's dataset. All measures are normalized and converted to LOG scale.

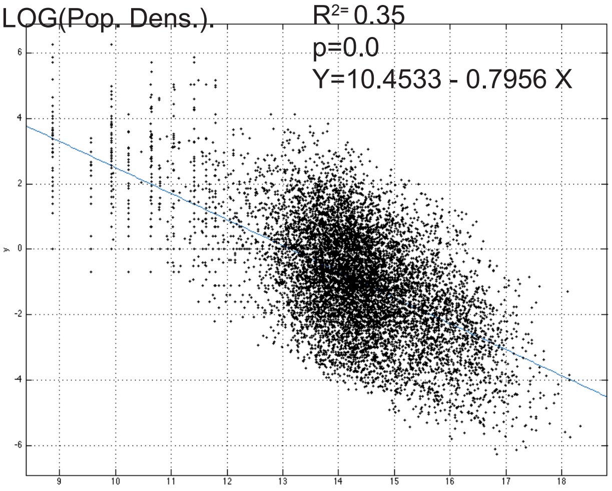


are tested as predictors of *tweeting* density by census block, the relationships, though significant, are far more weaker (almost negligible for the case of income). The final model (Figure B.6) corresponds to a multiple regression that combines Average Income and Population Density per census block, in order to explain the *tweeting* density. Again, despite being significant, the relationship is extremely weak (the model does not explain more than 4% of the variability).

These results suggest that the tweeter data tends to represent mobility rather than demographic patterns (or better said, activity rather than residence of people). Notwithstanding, it must be noted that no direct information regarding the income of the users can be inferred. This means that although this type of data might not be biased towards places with population of predominant high income, it does not necessarily constitute a representative sample of all the income groups.

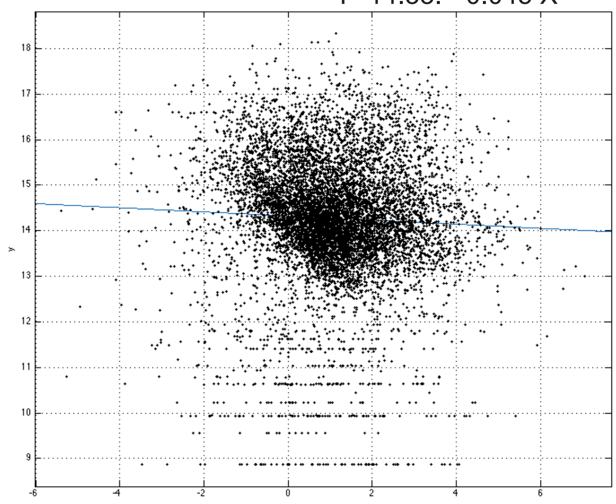
obtained from the consultancy Adimark, www. adimark.cl).

Figure B.3
LOG(Pop. Dens.).
R^2= 0.35
p=0.0
Y=10.4533 - 0.7956 X



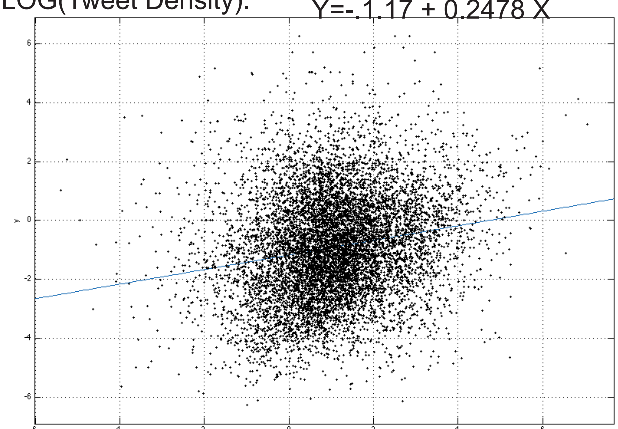
LOG(Avg. Income Census Block).

Figure B.4
LOG(Tweet Density).
R^2= 0.0022
p=0.0
Y=14.33. - 0.045 X



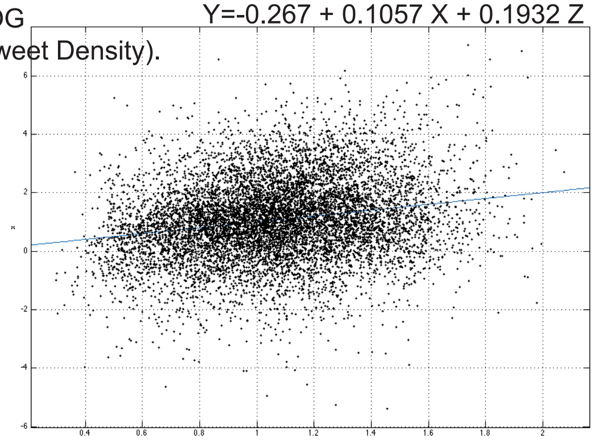
LOG(Avg. Income. Census Block).

Figure B.5
LOG(Tweet Density).
R^2= 0.04
p=0.0
Y=-.1.17 + 0.2478 X



LOG(Population Density Census Block).

Figure B.6
LOG
(Tweet Density).
R^2= 0.04
p1 & p2=0.0
Y=-0.267 + 0.1057 X + 0.1932 Z

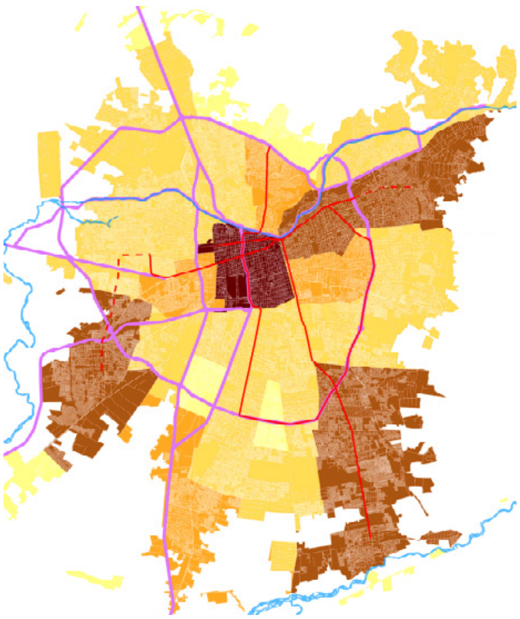


X=LOG(Avg.Incom) Z=Log(Pop. Dens.).

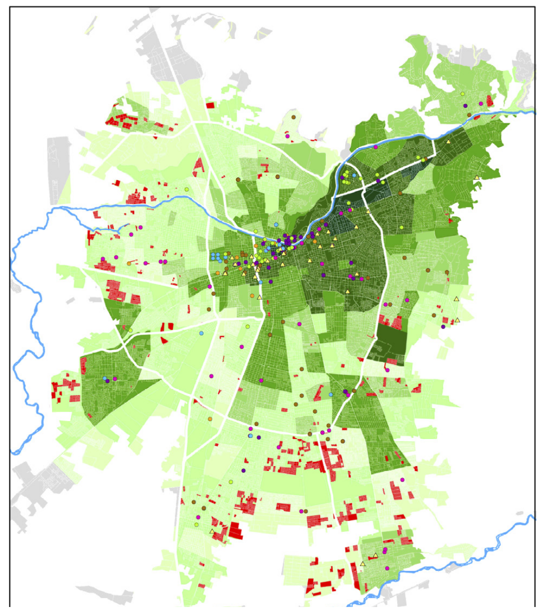
Appendix C:
Spatial Concentration Indicators

The following maps of Santiago were downloaded from the site of Observatorio de Ciudades Universidad Catolica (OCUC, www.ocuc.cl), with permission granted by Ricardo Trufello (author).

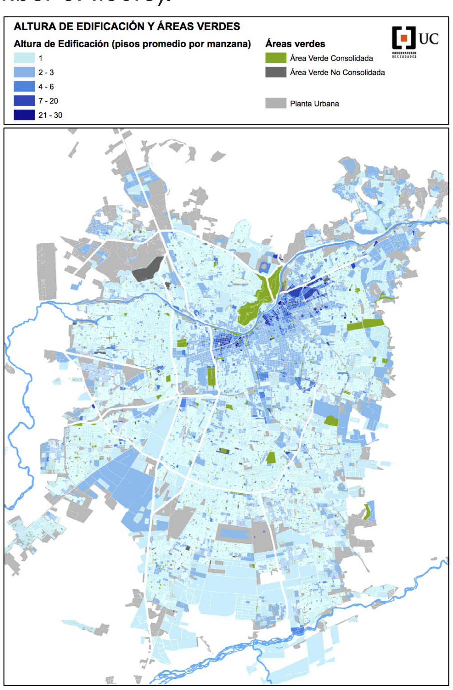
Map 1: Origin-Destination Index (darker==more trips generated or attracted), according to Origin-Destination Survey of 2001 (Chilean Government, MTT).



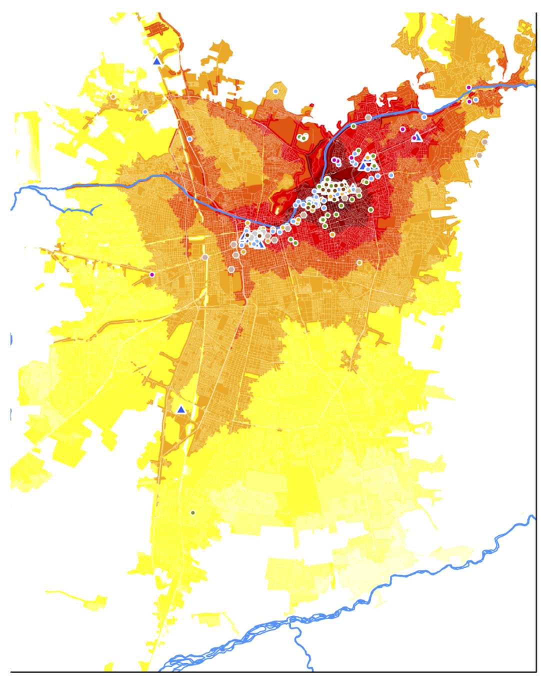
Map 2: Land Value (darker==higher land value). In red the complexes of social housing built during the period 1982-2000.

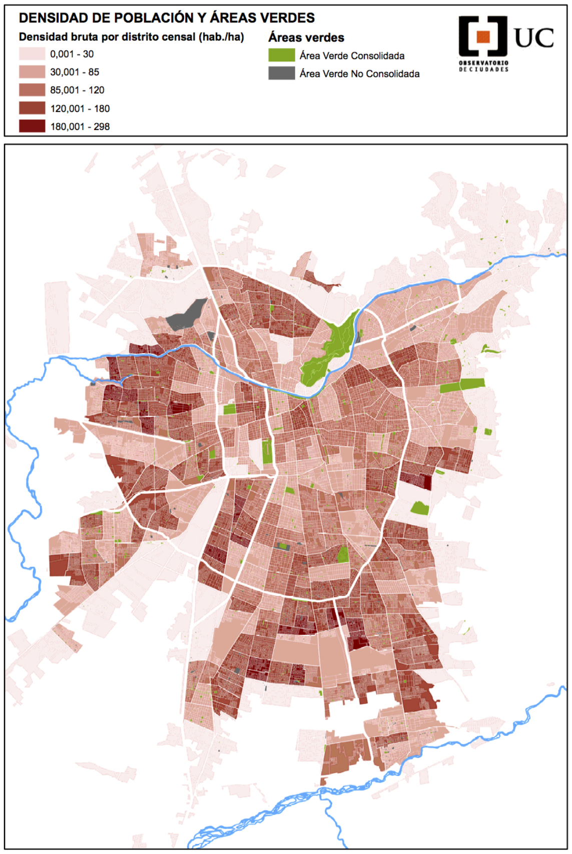


Map 3: Building Height (symbology indicates number of floors).

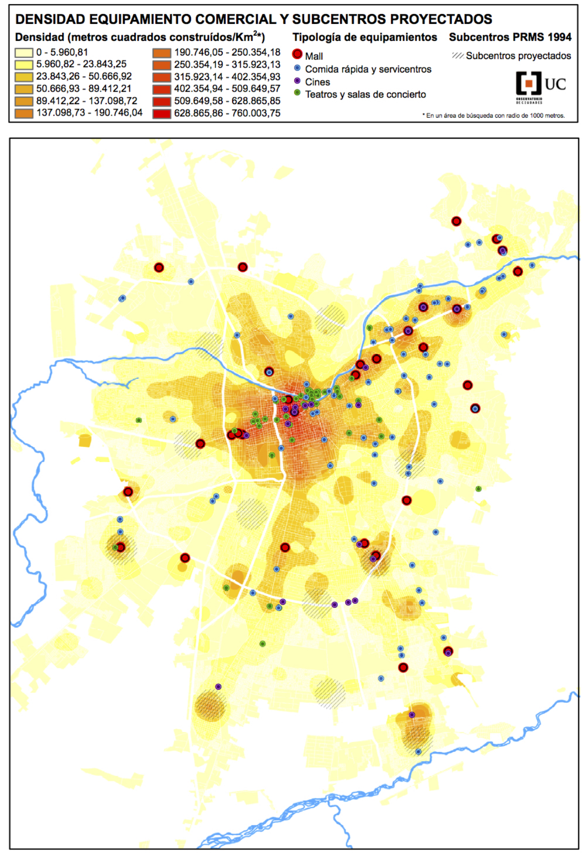


Map 4: Interpolation based on the location of major finance institutions, services and corporate headquarters (darker==higher centrality).

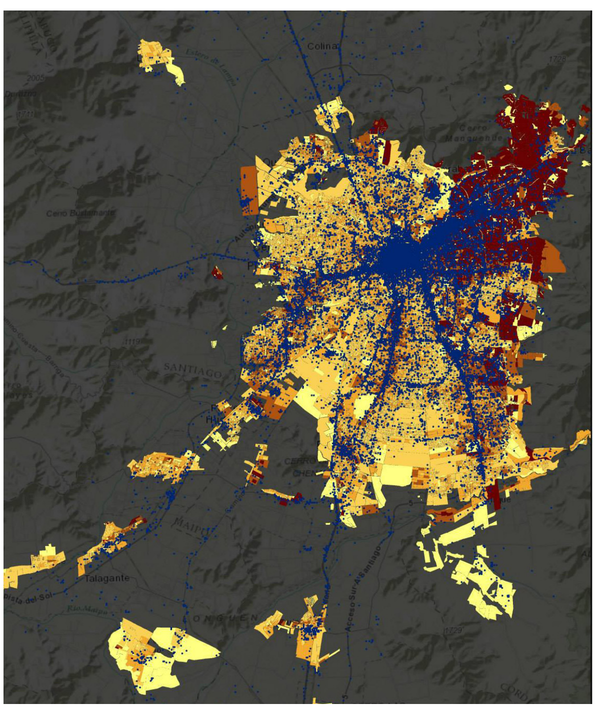




Map 5: Population Density (intervals on the top indicate number of persons per hectare).



Map 6: Centrality interpolation based on retail density (intervals on the top indicate built are per square kilometer). Red dots indicate shopping centres.



Map 7: Predominant income group by census block (red==higher income), overlaid by the observations from the dataset used in this research (own elaboration with data from OCUC).

REFERENCES

- Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998b. “Urban spatial structure.” Journal of Economic Literature no. 36 (3):1426.
- Alexander, Christopher, and Lewis Mumford. “A city is not a tree.”
- Arcaute, Elsa, Erez Hatna, Peter Ferguson, Hyejin Youn, Anders Johansson, and Michael Batty. 2013. “City boundaries and the universality of scaling laws.”
- Bus Rapid Transit Centre of Excellence (ALC-BRT CoE). 2012. “Asesoría Experta para la Ejecución de un Estudio Comparativo. Indicadores de Ciudades Latinoamericanas.”
- Bonacich, Phillip. 1987. “Some unique properties of eigenvector centrality.” Social Networks no. 29:555-564. doi: 10.1016/j.socnet.2007.04.002.
- Bonacich, Phillip. 1987a. “Power and centrality: a family of measures.” American Journal of Sociology no. 92:1170-1182. doi: 10.1086/228631.
- Borgatti, Stephen P, and Daniel S Halgin. 2011. “Analyzing affiliation networks.” The Sage handbook of social network analysis:417-433.
- Calabrese, Francesco, Jonathan Reades, and Carlo Ratti. 2010. “Eigenplaces: Segmenting Space through Digital Signatures.” IEEE Pervasive Computing no. 9 (1):78.
- Cha, Sung-Hyuk. 2007. “Comprehensive survey on distance/similarity measures between probability density functions.” City no. 1 (2):1.
- Cranshaw, Justin, Raz Schwartz, Jason I Hong, and Norman M Sadeh. 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. Paper read at ICWSM.
- Eagle, N., and A. S. Pentland. 2009. “Eigenbehaviors: identifying structure in routine.” BEHAVIORAL ECOLOGY AND SOCIOBIOLOGY no. 63 (7):1057-1066.
- Echenique, Juan, and Urzua, Segio. 2013. “Desigualdad, Segregacion y Resultados Educacionales. Evidencia desde el Metro de Santiago.” Centro de Estudios Publicos (CEP) Puntos de Referencia. no. 359.
- Fischer Montt, Daniel Andrés. 2010. “Obtención de Información para la Gestión del Transantiago a Través Datos Operacionales.”
- Freeman, Linton C. 2011. “The Development of Social Network Analysis–with an Emphasis on Recent Events.” The SAGE Handbook of Social Network Analysis:26-39.
- Freeman, Linton C, and Cynthia M Webster. 1994. “Interpersonal proximity in social and cognitive space.” Social Cognition no. 12 (3):223-247.
- Fujita, Masahisa, Paul R. Krugman, and Anthony Venables. 1999. The Spatial Economy : Cities, Regions and International Trade. Cambridge, Mass: MIT Press.
- Geddes, Patrick Sir. 1997. Cities in evolution : an introduction to the town planning movement and to the study of civics / Patrick Geddes, Social theories of the city: v. 5: London : Routledge/Thoemmes, 1997.
- Graham, Stephen. 2002. “On technology, infrastructure and the contemporary urban condition: a response to Coutard.” International Journal of Urban and Regional Research no. 26 (1):175-182.
- Graham, Stephen, and Simon Marvin. 2001. Splintering urbanism : networked infrastructures, technological mobilities and the urban condition / Stephen Graham and Simon Marvin: London ; New York : Routledge, 2001.
- Jiang, Shan, Joseph Ferreira, and Marta C González. 2012. “Clustering daily patterns of human activities in the city.” Data Mining and Knowledge Discovery no. 25 (3):478-510.
- Lynch, Kevin. 1964. The image of the city / Kevin Lynch, Publication of the Joint Center for Urban Studies: Cambridge, Mass. : M.I.T. Press, 1964. 1st M.I.T. Press pbk. ed.
- Martin, Hilbert. “When is Cheap, Cheap Enough to Bridge the Digital Divide? Modeling In-

come Related Structural Challenges of Technology Diffusion in Latin America." World Development no. 38:756-770. doi: 10.1016/j.worlddev.2009.11.019.

- Massey, Doreen. 2009. "Concepts of space and power in theory and in political practice." Documents d'Anàlisi Geogràfica (55):15-26.
- Munizaga, Marcela, Devilaine, Flavio, Navarrete, Claudio, and Diego Silva. "Validating travel behavior estimated from smartcard data." Paper presented at the International Choice Modelling Conference, Sydney, Australia, July 3–5, 2013.
- Ministry of Transportation and Telecommunications (MTT), Chilean Government. 2013. "Plan Maestro de Transporte Santiago 2025".
- Noulas, Anastasios, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. 2011. "An Empirical Study of Geographic User Activity Patterns in Foursquare." ICWSM no. 11:70-573.
- Pentland, Alex. 2009. "Reality mining of mobile communications: Toward a new deal on data." The Global Information Technology Report 2008–2009:1981.
- Quijada, R, A Tirachini, R Henríquez, and R Hurtubia. 2007. "Investigación al Transantiago: Sistematización de Declaraciones hechas ante la Comisión Investigadora, Resumen de Contenidos de los Principales Informes Técnicos, Información de Documentos Públicos Adicionales y Comentarios Críticos." Santiago: Departamento de Ingeniería de Transporte, Universidad de Chile.
- Ratti, Carlo, Stanislav Sobolevsky, Francesco Calabrese, Clio Andris, Jonathan Reades, Mauro Martino, Rob Claxton, and Steven H. Strogatz. 2010. "Redrawing the Map of Great Britain from a Network of Human Interactions." PLoS ONE no. 5 (12):1-6. doi: 10.1371/journal.pone.0014248.
- Reades, J., F. Calabrese, and C. Ratti. 2009. "Eigenplaces: analysing cities using the space-time structure of the mobile phone network." ENVIRONMENT AND PLANNING B-PLANNING & DESIGN no. 36 (5):824-836.
- Ringnér, Markus. 2008. "What is principal component analysis?" Nature biotechnology no. 26 (3):303-304.
- Ruhnau, B. 2000. "Eigenvector-centrality - a node-centrality?" SOCIAL NETWORKS no. 22 (4):357-365.
- Soja, Edward W. 2000. Postmetropolis : critical studies of cities and regions / Edward W. Soja: Oxford ; Malden, Mass. : Blackwell Publishers, 2000.
- Steer Davies Gleave. 2012. "Estrategia de Desarrollo Regional de la Region Metropolitana de Santiago". Document prepared for the Regional Government (GORE) and the United Nations Development Programme (PNUD).
- Szell, Michael, Sebastian Grauwin, and Carlo Ratti. 2013b. "Contraction of online response to major events." arXiv preprint arXiv:1308.5190.
- Yang, Jiawen, Steven French, James Holt, and Xingyou Zhang. 2012. "Measuring the Structure of U.S. Metropolitan Areas, 1970–2000." Journal of the American Planning Association no. 78 (2):197-209. doi: 10.1080/01944363.2012.677382.
- Wambeke, Brad W., Min Liu, and Simon M. Hsiang. 2012. "Using Pajek and Centrality Analysis to Identify a Social Network of Construction Trades." Journal of Construction Engineering & Management no. 138 (10):1192-1201. doi: 10.1061/(ASCE)CO.1943-7862.0000524.

- *(Anas, Arnott and Small, 1998) Anas, Alex, Richard Arnott, and Kenneth A. Small. 1998b. "Urban spatial structure." Journal of Economic Literature no. 36 (3):1426.*
- *(Alexander, 1965) Alexander, Christopher, and Lewis Mumford. "A city is not a tree."*
- *(Arcaute, Batty, et. al., 2013) Arcaute, Elsa, Erez Hatna, Peter Ferguson, Hyejin Youn, Anders Johansson, and Michael Batty. 2013. "City boundaries and the universality of scaling laws."*
- *(ALC-BRT CoE, 2012) Bus Rapid Transit Centre of Excellence (ALC-BRT CoE). 2012. "Asesoría Experta para la Ejecución de un Estudio Comparativo. Indicadores de Ciudades Latinoamericanas."*
- *(Bonacich, 2007) Bonacich, Phillip. 1987. "Some unique properties of eigenvector centrality." Social Networks no. 29:555-564. doi: 10.1016/j.socnet.2007.04.002.*
- *(Bonacich, 1987) Bonacich, Phillip. 1987a. "Power and centrality: a family of measures." American Journal of Sociology no. 92:1170-1182. doi: 10.1086/228631.*
- *(Borgatti and Halgin, 2011) Borgatti, Stephen P, and Daniel S Halgin. 2011. "Analyzing affiliation networks." The Sage handbook of social network analysis:417-433.*
- *(Calabrese, Reades, Ratti, 2010) Calabrese, Francesco, Jonathan Reades, and Carlo Ratti. 2010. "Eigenplaces: Segmenting Space through Digital Signatures." IEEE Pervasive Computing no. 9 (1):78.*
- *(Cha, 2007) Cha, Sung-Hyuk. 2007. "Comprehensive survey on distance/similarity measures between probability density functions." City no. 1 (2):1.*
- *(Cranshaw, Schwartz, Hong, Sadeh, 2012) Cranshaw, Justin, Raz Schwartz, Jason I Hong, and Norman M Sadeh. 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. Paper read at ICWSM.*
- *(Eagle, Pentland, 2009) Eagle, N., and A. S. Pentland. 2009. "Eigenbehaviors: identifying structure in routine." BEHAVIORAL ECOLOGY AND SOCIOBIOLOGY no. 63 (7):1057-1066.*
- *(Echenique and Urzua, 2013) Echenique, Juan, and Urzua, Segio. 2013. "Desigualdad, Segregacion y Resultados Educacionales. Evidencia desde el Metro de Santiago." Centro de Estudios Publicos (CEP) Puntos de Referencia. no. 359.*
- *(Fischer Montt, 2010) Fischer Montt, Daniel Andrés. 2010. "Obtención de Información para la Gestión del Transantiago a Través Datos Operacionales."*
- *(Freeman, 2011) Freeman, Linton C. 2011. "The Development of Social Network Analysis–with an Emphasis on Recent Events." The SAGE Handbook of Social Network Analysis:26-39.*
- *(Freeman, Webster, 1994) Freeman, Linton C, and Cynthia M Webster. 1994. "Interpersonal proximity in social and cognitive space." Social Cognition no. 12 (3):223-247.*
- *(Fujita, Krugman, and Venables, 1999) Fujita, Masahisa, Paul R. Krugman, and Anthony Venables. 1999. The Spatial Economy : Cities, Regions and International Trade. Cambridge, Mass: MIT Press.*
- *(Geddes, 1997) Geddes, Patrick Sir. 1997. Cities in evolution : an introduction to the town planning movement and to the study of civics / Patrick Geddes, Social theories of the city: v. 5: London : Routledge/Thoemmes, 1997.*
- *(Graham, 2002) Graham, Stephen. 2002. "On technology, infrastructure and the contemporary urban condition: a response to Coutard." International Journal of Urban and Regional Research no. 26 (1):175-182.*
- *(Graham, Marvin, 2001) Graham, Stephen, and Simon Marvin. 2001. Splintering urbanism : networked infrastructures, technological mobilities and the urban condition / Stephen Graham and Simon Marvin: London ; New York : Routledge, 2001.*
- *(Jiang, Ferreira and Gonzalez 2012) Jiang, Shan, Joseph Ferreira, and Marta C González. 2012. "Clustering daily patterns of human activities in the city." Data Mining and*

Knowledge Discovery no. 25 (3):478-510.

- (Lynch, 1964) Lynch, Kevin. 1964. The image of the city / Kevin Lynch, Publication of the Joint Center for Urban Studies: Cambridge, Mass. : M.I.T. Press, 1964. 1st M.I.T. Press pbk. ed.
- (Martin, 2009) Martin, Hilbert. "When is Cheap, Cheap Enough to Bridge the Digital Divide? Modeling Income Related Structural Challenges of Technology Diffusion in Latin America." World Development no. 38:756-770. doi: 10.1016/j.worlddev.2009.11.019.
- (Massey, 2009) Massey, Doreen. 2009. "Concepts of space and power in theory and in political practice." Documents d'Anàlisi Geogràfica (55):15-26.
- (Munizaga et. al., 2013) Munizaga, Marcela, Devilaine, Flavio, Navarrete, Claudio, and Diego Silva. "Validating travel behavior estimated from smartcard data." Paper presented at the International Choice Modelling Conference, Sydney, Australia, July 3–5, 2013.
- (MTT, 2013) Ministry of Transportation and Telecommunications (MTT), Chilean Government. 2013. "Plan Maestro de Transporte Santiago 2025".
- (Pentland, 2009) Pentland, Alex. 2009. "Reality mining of mobile communications: Toward a new deal on data." The Global Information Technology Report 2008–2009:1981.
- (Quijada, Tirachini, Henriques, Hurtubia, 2007) Quijada, R, A Tirachini, R Henríquez, and R Hurtubia. 2007. "Investigación al Transantiago: Sistematización de Declaraciones hechas ante la Comisión Investigadora, Resumen de Contenidos de los Principales Informes Técnicos, Información de Documentos Públicos Adicionales y Comentarios Críticos." Santiago: Departamento de Ingeniería de Transporte, Universidad de Chile.
- (Ratti et. al., 2010) Ratti, Carlo, Stanislav Sobolevsky, Francesco Calabrese, Clio Andris, Jonathan Reades, Mauro Martino, Rob Claxton, and Steven H. Strogatz. 2010. "Redrawing the Map of Great Britain from a Network of Human Interactions." PLoS ONE no. 5 (12):1-6. doi: 10.1371/journal.pone.0014248.
- (Reades, Calabrese, Ratti, 2009) Reades, J., F. Calabrese, and C. Ratti. 2009. "Eigenplaces: analysing cities using the space-time structure of the mobile phone network." ENVIRONMENT AND PLANNING B-PLANNING & DESIGN no. 36 (5):824-836.
- (Ringnér, 2008) Ringnér, Markus. 2008. "What is principal component analysis?" Nature biotechnology no. 26 (3):303-304.
- (Ruhnau, 2000) Ruhnau, B. 2000. "Eigenvector-centrality - a node-centrality?" SOCIAL NETWORKS no. 22 (4):357-365.
- (Soja, 2000) Soja, Edward W. 2000. Postmetropolis : critical studies of cities and regions / Edward W. Soja: Oxford ; Malden, Mass. : Blackwell Publishers, 2000.
- (Steer Davies and Gleave, 2012) Steer Davies Gleave. 2012. "Estrategia de Desarrollo Regional de la Region Metropolitana de Santiago". Document prepared for the Regional Government (GORE) and the United Nations Development Programme (PNUD).
- (Szell, Grauwin and Ratti, 2013) Szell, Michael, Sebastian Grauwin, and Carlo Ratti. 2013b. "Contraction of online response to major events." arXiv preprint arXiv:1308.5190.
- (Yang, et. al., 2012) Yang, Jiawen, Steven French, James Holt, and Xingyou Zhang. 2012. "Measuring the Structure of U.S. Metropolitan Areas, 1970–2000." Journal of the American Planning Association no. 78 (2):197-209. doi: 10.1080/01944363.2012.677382.