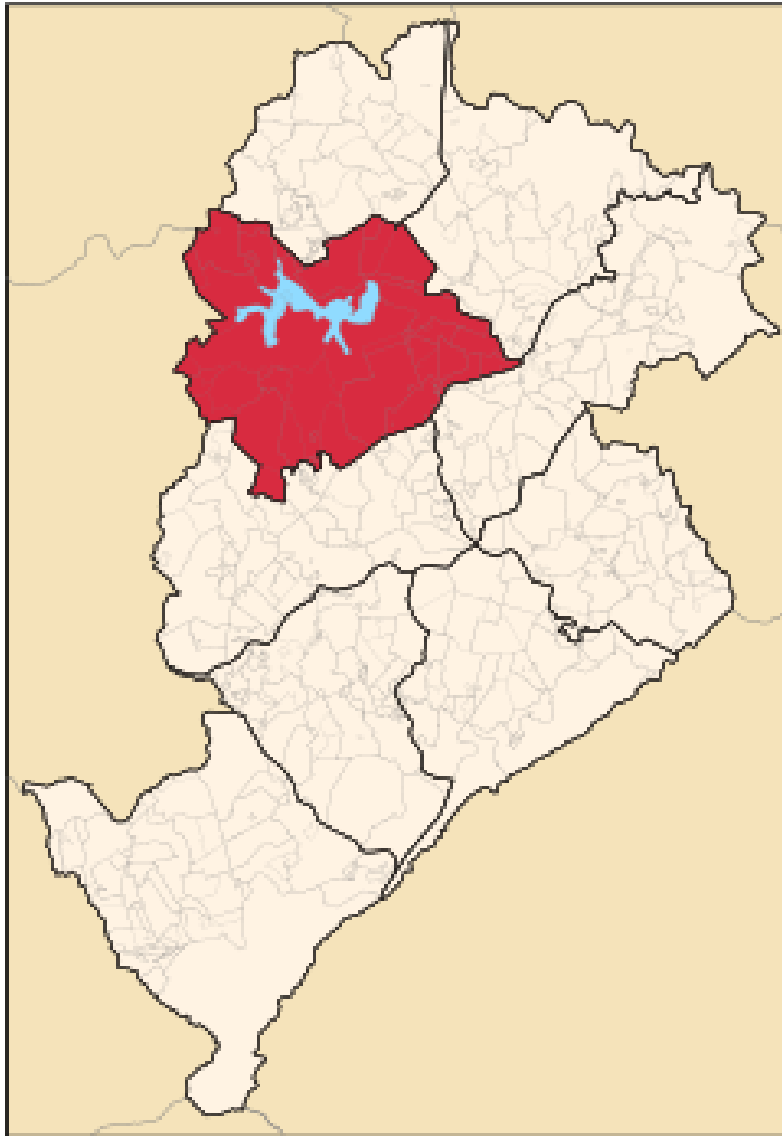

Traditional VS Data-driven Approaches: A Map Comparison for the Evaluation of Usability of Twitter Data

Stefano Pensa, Elena Masala

DAD - Department of Architecture and Design, Politecnico di Torino, Torino, Italy - stefano.pensa@polito.it

SiTI - Higher Institute on Territorial Systems for Innovation, Torino, Italy – elena.masala@polito.it

Case Study – Pampulha, Belo Horizonte



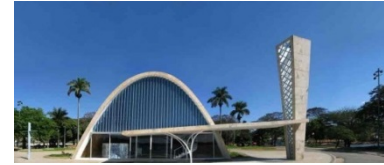
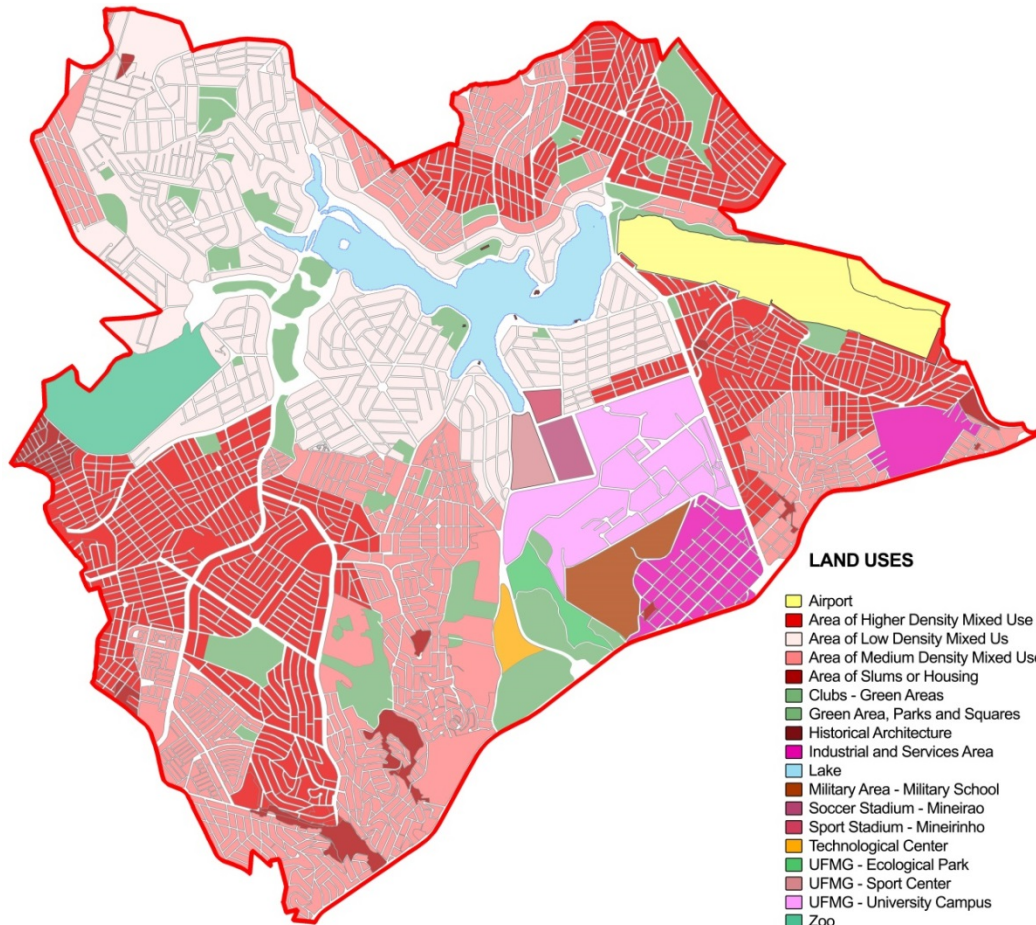
Belo Horizonte is the capital of the state of **Minas Gerais, Brazil**

The municipality has a population of **2.500.000** inhabitants (up to 5.000.000 in the Metropolitan Area)

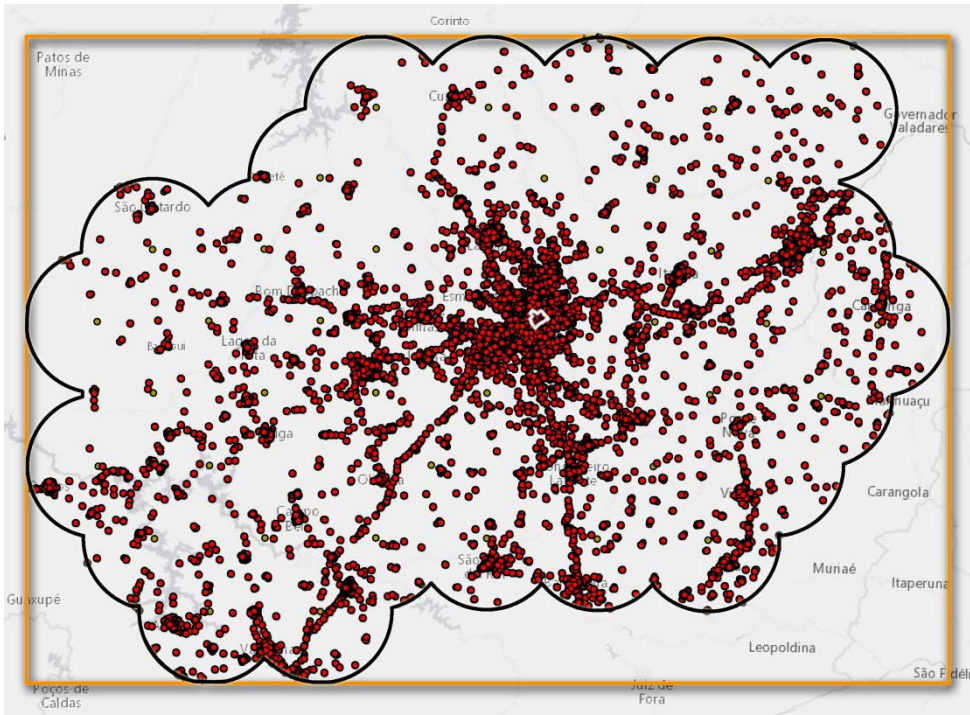
The Case Study is localised in **Pampulha** neighborhood.

Pampulha has a population of 145.000 inhabitants and a density of 3.082 inh/km²

Case Study – Pampulha, Belo Horizonte

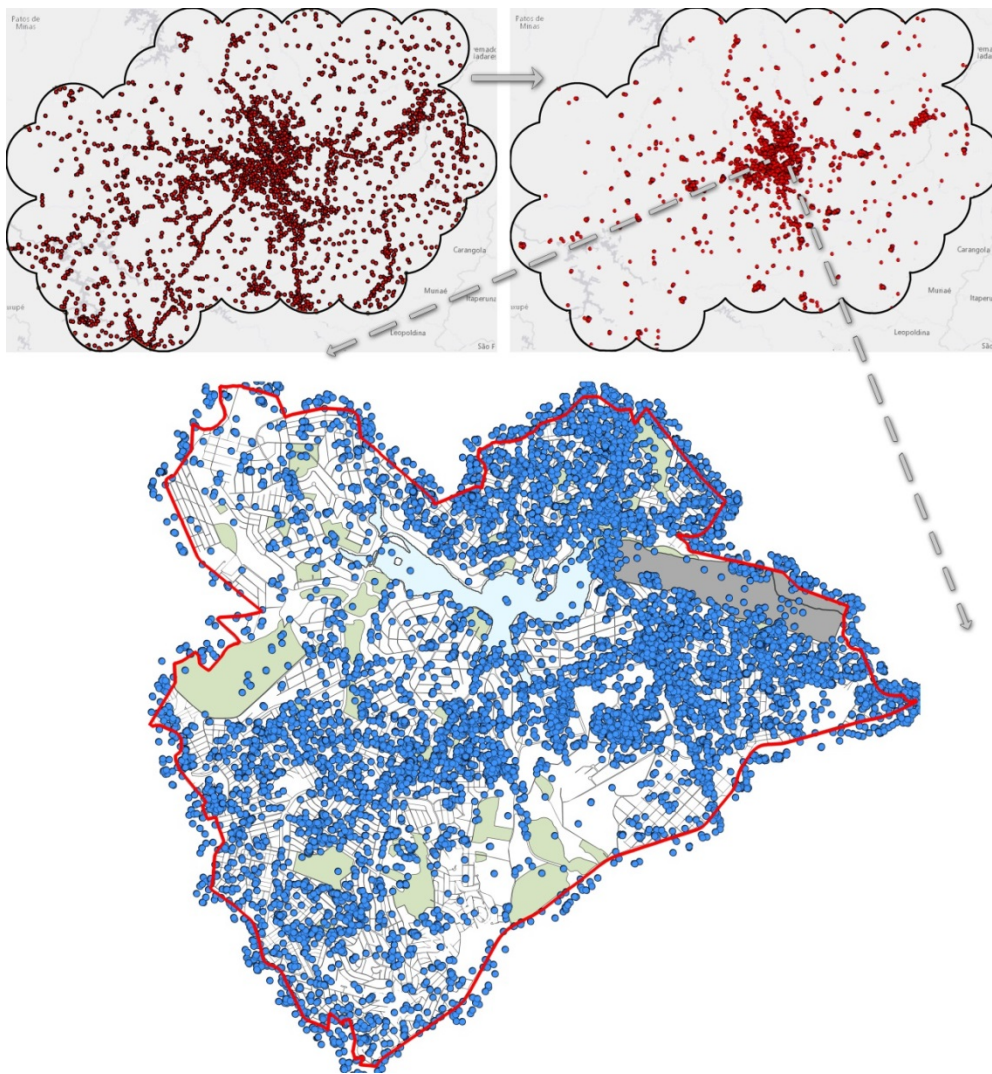


Data Collection



- Tweets collected through the **Streaming APIs** provided by the Twitter company itself
- Period between and **April and June 2015**
- Whole **Metropolitan Area** of Belo Horizonte.
N-S -18.418° , -21.698°
E-W -41.734° -46.715° W
- Tweets without a **geo-reference** are self-excluded
- **929.281** of tweets were collected (without inconsistencies), from **41.317** users

Data Collection



- All the tweets sent by the users who sent **at least one tweet from Pampulha**:
 - 42,991 tweets
 - 5,173 users.
- **4,6%** of the total amount of tweets have been sent from the area of Pampulha.
- **12,52%** of total amount of Twitter users tweeted from Pampulha
- **63%** of tweets from the outside of Pampulha

Data Collection

Tweets_pampulha_W84

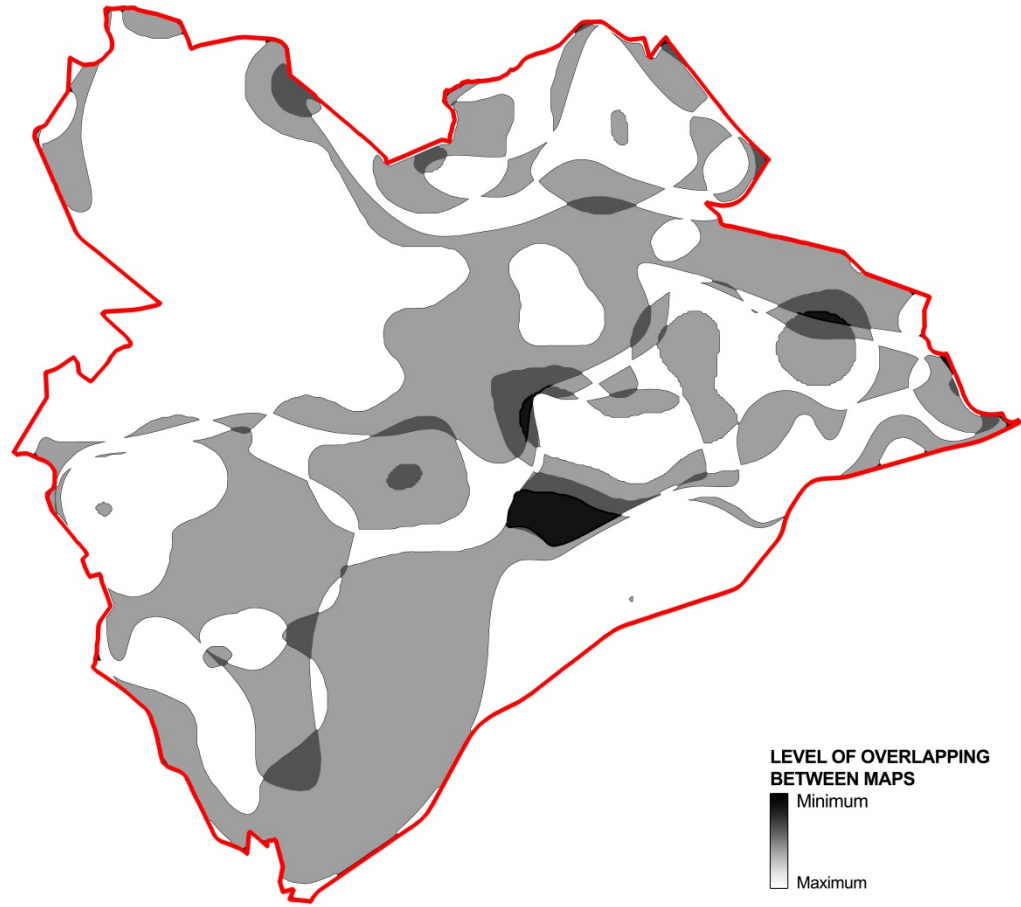
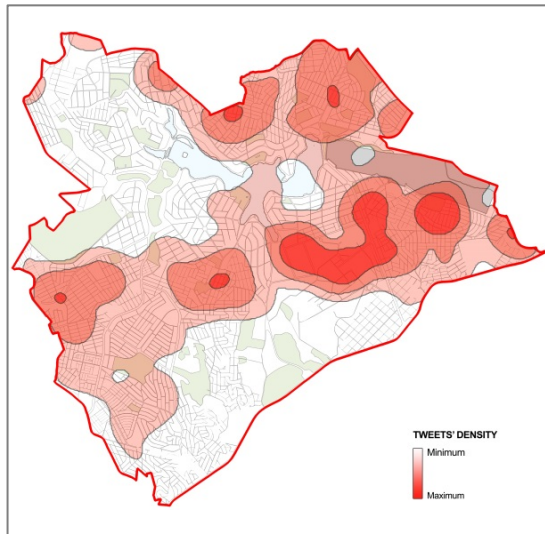
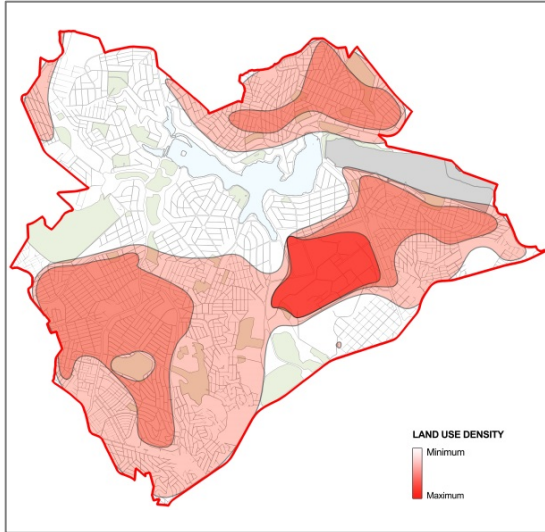
FID	Shape *	ID	MESSAGE	USER FROM	USER TO	DATE 01	LAT	LON	DAY W	MONTH	DAY MONTH	YEAR	DATA C
0	Point	210	Gosto de liberar [os alunos] ... \$ 9:00 para chegarmos mais ced	MatthSousa		Thu Mar 26 11:51:44 +0000 2015	-19,86591	-43,95981	Thu	Mar	26	2015	26/03/2015
1	Point	976	Agora fiquei com muito medo de estudar ... noite D:	MatthSousa		Thu Mar 26 11:47:45 +0000 2015	-19,86568	-43,95996	Thu	Mar	26	2015	26/03/2015
2	Point	987	To vendo a hora que essa Vera vai me perguntar alguma coisa	LeandroResLac		Thu Mar 26 11:45:38 +0000 2015	-19,86604	-43,9597	Thu	Mar	26	2015	26/03/2015
3	Point	988	Bom dia Espírito Santo o que vamos fazer juntos hoje?	claudiomnovo		Thu Mar 26 11:45:25 +0000 2015	-19,86912	-43,95008	Thu	Mar	26	2015	26/03/2015
4	Point	1004	Como j previa , o dia j começou com stress 'L'	LauraTeixeira22		Thu Mar 26 11:43:11 +0000 2015	-19,83163	-43,97163	Thu	Mar	26	2015	26/03/2015
5	Point	1007	@Silva_C dps de ontem n'Éo vai lembra msm	GuiAalvees	Silva_C	Thu Mar 26 11:42:43 +0000 2015	-19,84104	-43,96421	Thu	Mar	26	2015	26/03/2015
6	Point	1008	Os meninos t'Éo com um bong doido dmsss	GuiAalvees		Thu Mar 26 11:42:04 +0000 2015	-19,8409	-43,96417	Thu	Mar	26	2015	26/03/2015
7	Point	1009	No CIAAR. I just finished running 5.04 km in 26m:39s with #Endo	acantus94		Thu Mar 26 11:41:51 +0000 2015	-19,84957	-43,96029	Thu	Mar	26	2015	26/03/2015
8	Point	1021	@Silva_C I na praça, n'Éo paga de planta n'Éo	GuiAalvees	Silva_C	Thu Mar 26 11:39:54 +0000 2015	-19,84114	-43,96422	Thu	Mar	26	2015	26/03/2015
9	Point	1027	Semana passou rapidao ???????	PalomaRamalho17		Thu Mar 26 11:37:42 +0000 2015	-19,83613	-43,96052	Thu	Mar	26	2015	26/03/2015
10	Point	1029	T' ficando irritado mt f cil, tenho que parar com isso	GuiAalvees		Thu Mar 26 11:36:52 +0000 2015	-19,84109	-43,96416	Thu	Mar	26	2015	26/03/2015
11	Point	1030	Bm Dia, hoje ja , quinta ??????	PalomaRamalho17		Thu Mar 26 11:36:36 +0000 2015	-19,83613	-43,96052	Thu	Mar	26	2015	26/03/2015
12	Point	1032	@leorossatto , a única coisa que o faz mineiro. Mas isso pode s	HelenaNoce	leorossatto	Thu Mar 26 11:47:52 +0000 2015	-19,87396	-43,9802	Thu	Mar	26	2015	26/03/2015
13	Point	1033	Esporro por estar no celular, sempre rola ??????	_peguetedoian		Thu Mar 26 11:38:04 +0000 2015	-19,87183	-43,96252	Thu	Mar	26	2015	26/03/2015
14	Point	1034	@loversdefenders bom dia ??	itsmctly4ever	loversdefenders	Thu Mar 26 11:37:52 +0000 2015	-19,8721	-44,00528	Thu	Mar	26	2015	26/03/2015
15	Point	1035	Meu deus eu tenho a melhor professora de matem tica do mund	_peguetedoian		Thu Mar 26 11:35:48 +0000 2015	-19,87183	-43,96252	Thu	Mar	26	2015	26/03/2015
16	Point	1037	Atitudes durante a redaç'Éo da tese: Durante a redaç'Éo da tes	pathayde		Thu Mar 26 11:33:24 +0000 2015	-19,86909	-43,96638	Thu	Mar	26	2015	26/03/2015
17	Point	2273	I'm at Universidade Federal de Minas Gerais (UFMG) - @ufmgbr	ramonpepe		Thu Mar 26 11:58:54 +0000 2015	-19,86757	-43,96436	Thu	Mar	26	2015	26/03/2015
18	Point	2292	Essa professora parece tia de pr-escola... Para a aula pra cham	LeandroResLac		Thu Mar 26 11:58:06 +0000 2015	-19,86608	-43,9597	Thu	Mar	26	2015	26/03/2015
19	Point	2317	Muito importante a melhor que eu sei que eu sei... Éo que eu sei	LeandroResLac		Thu Mar 26 11:55:21 +0000 2015	-19,86608	-43,9597	Thu	Mar	26	2015	26/03/2015

Message
User from
User to

Time: dd/mm/yy hh:mm:ss

Geo Location: Latitude and Longitude

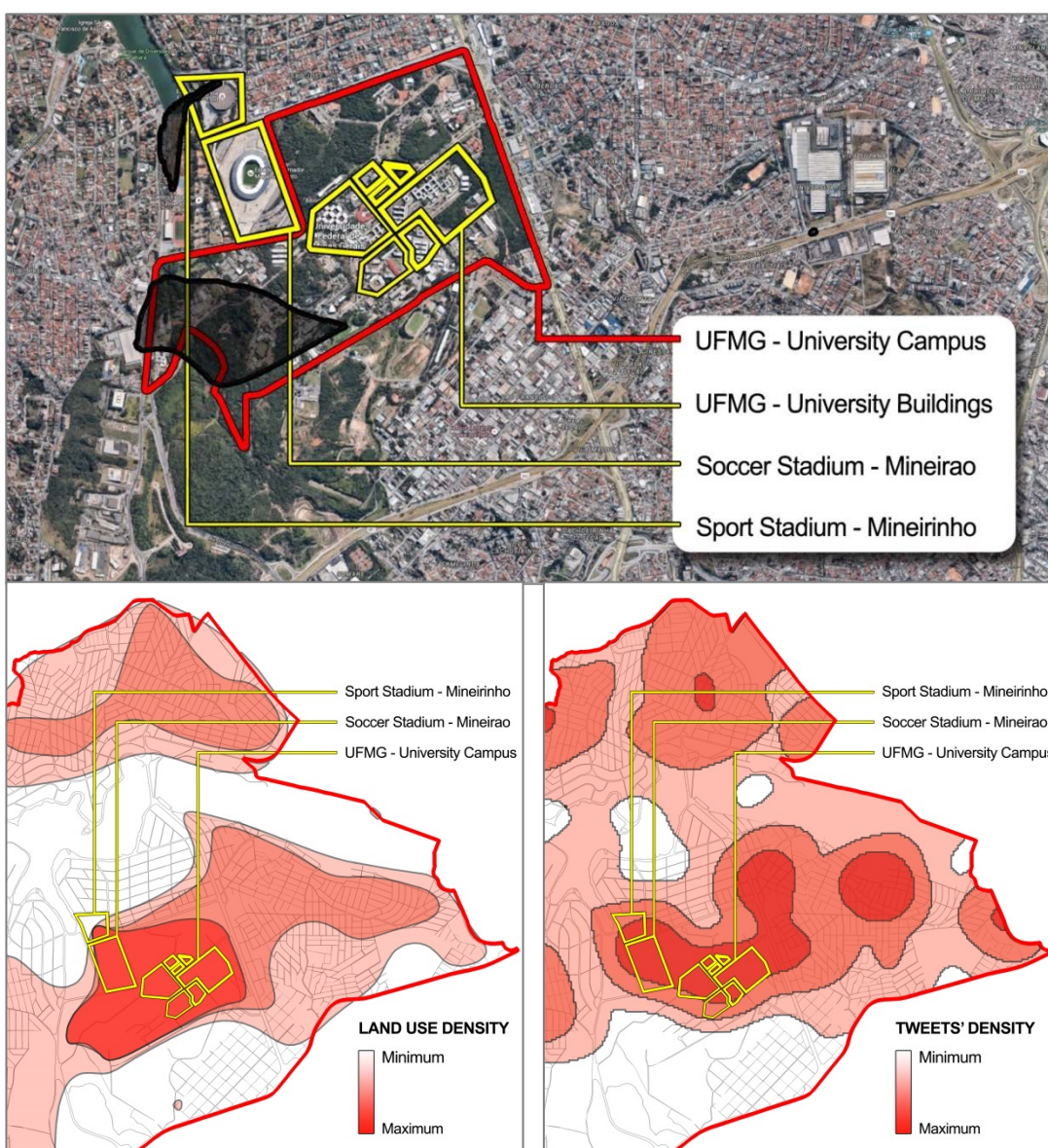
Methodology - Density Maps



Overlapping of land-use and tweet density maps (main)

Density map based on land-use vs Quartic Kernel function of tweets

Methodology - Density Maps

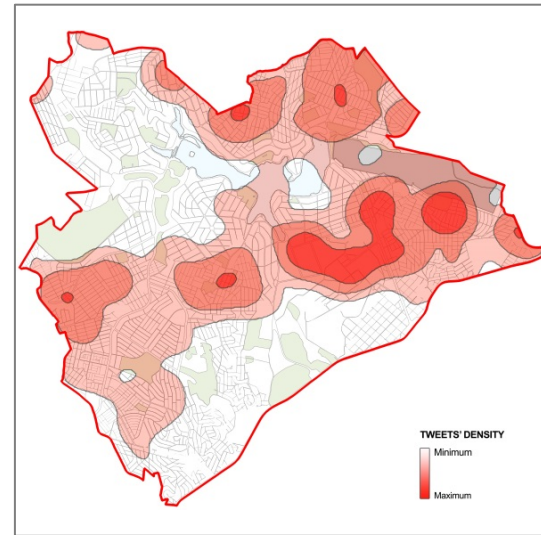
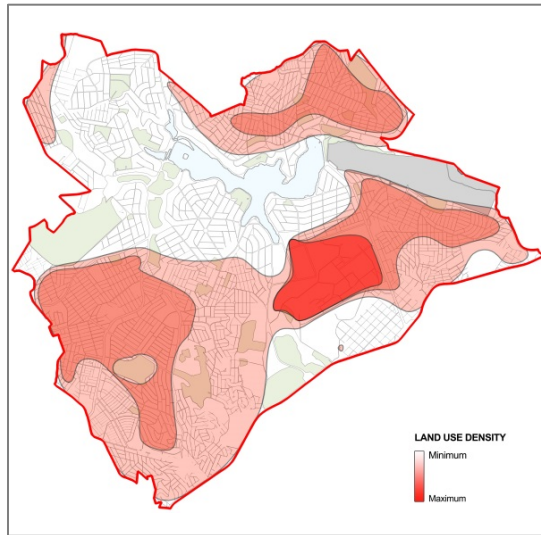


Zoom on the area with heterogeneous information:

the map on the top highlights the black areas (overlapping inconsistencies) and the shapes of plots and building in the University campus of UFMG and in the sport areas.

On the bottom, the images highlights the differences in the official land-use density map (on the left) and tweet density map (on the right).

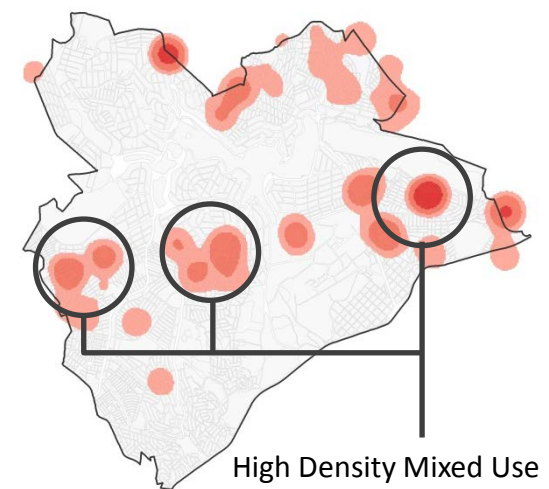
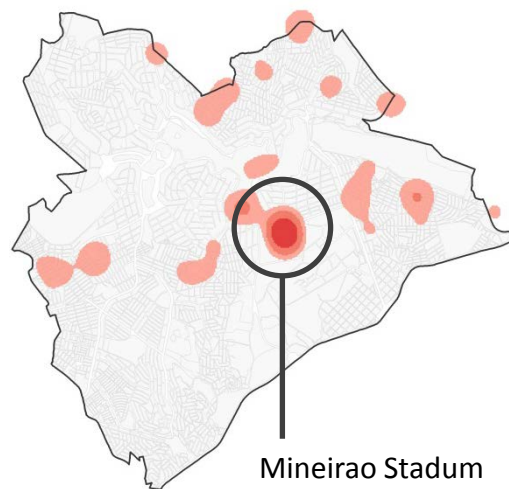
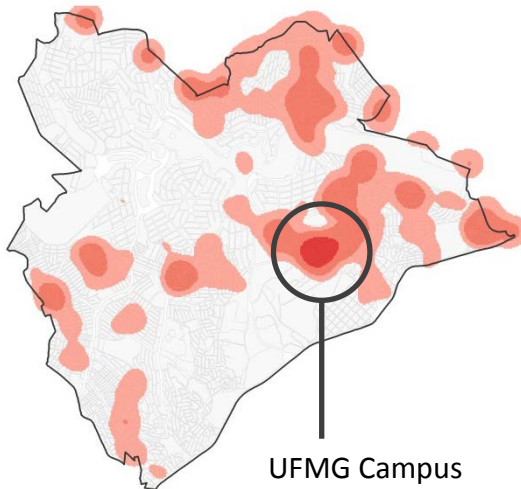
Methodology - Density Maps



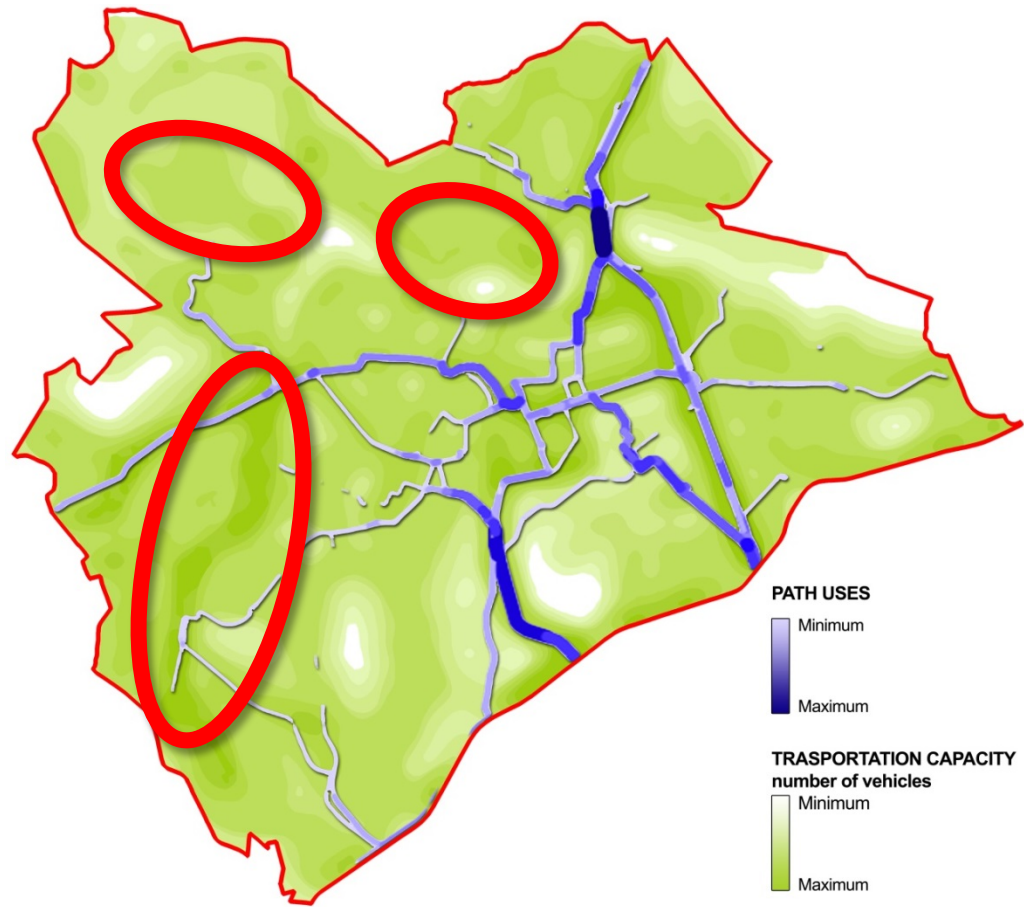
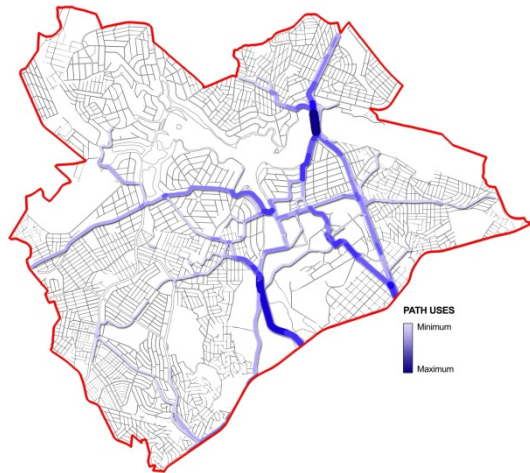
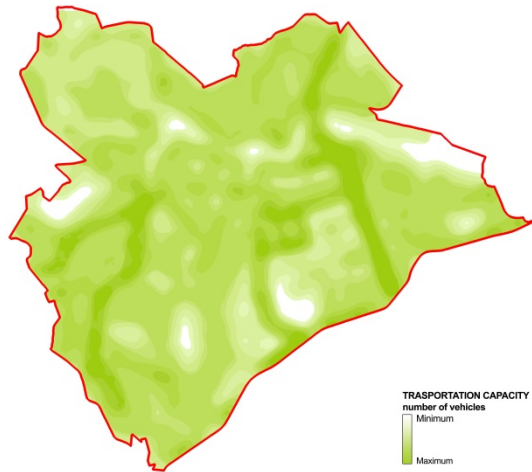
KF – 8.00 – 13.00 working day

KF – 14.00 – 19.00 Sunday

KF – 22.00 – 8.00 working day



Methodology – Shortest Paths



Capacity of roads in number of vehicles from 0 to 5400 (top left), shortest path algorithm to the couples of subsequent tweets* (bottom left) overlapping of both maps (main)

Conclusions

- Tweets can provide the reading of a place in a **particular period**. They can not be used to provide a general overview but they should be included within a specific context.
- They could be used for future simulations, but a **specific elaboration** should be done **to convert** them **in usable data**.
- The “Big” number of data provide a large number of possible correlations which could inspire different reasoning. However, only through the **exploration and analysis** of those data, the **cause-effect relationships can be distinguished from correlations** (Andrienko & Dykes, 2011; Andrienko, et al., 2007; Andrienko, Andrienko, Keim, MacEachren, & Wrobel, 2011; Jiang & Li, 2005; Jiang, Huang, & Vasek, 2003; Keim, Andrienko, Fekete, Görg, KohlhammeXr, & Melancon, 2008; Masala & Pensa, 2016).

Conclusions

- Maps derived from **different sources provide different information**. Whereas information matches, it could be justifiable taking those data as reliable. Otherwise, different information should be taken as a warning for dedicating some deeper investigation.
- It is evident that data need an interpretation because **data provides an information which is not unique**. This implies a stronger effort for planners and decision-makers in interpreting the information included in data.
- Data-driven approaches should provide data to the actors involved in the decision-making, but they should also let the **actors play with those data** in order to explore different combinations and discover correlations, cause-effects relationships and the best strategies for converting those dynamics into working elements for a spatial system.
- It is evident that tweets are sent by a self-selected population which is not a sample. Thus, they are not representative of the whole population, but can be the **mirror for comparing official data**.