Geo-social media as a proxy for meteorological data for flood monitoring

Camilo Restrepo-Estrada

camilo.restrepo@udea.edu.co

Phd Candidate University of São Paulo

São Carlos, SP (November 23th 2015)

Floods

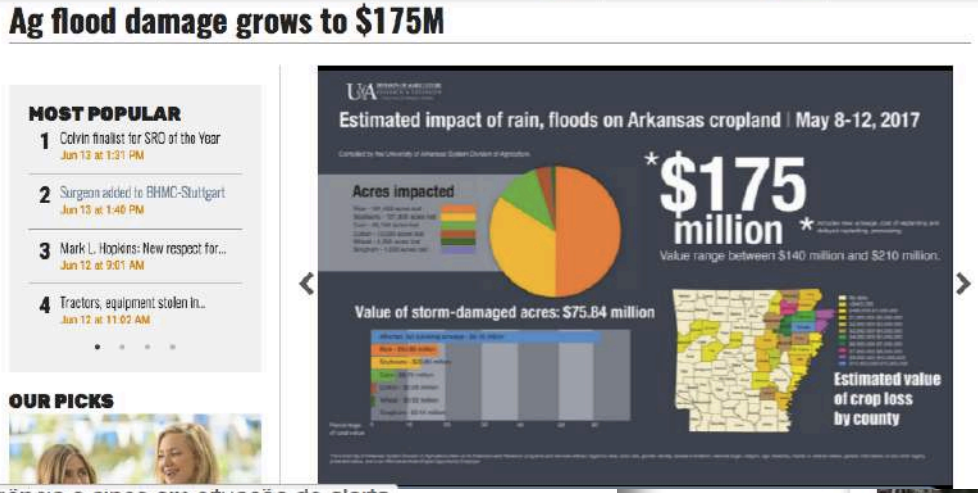
BBC
Entrar
Menu
BRASIL
Notícias | Brasil | Internacional | Economia | Saúde | Ciência
Enchentes podem prejudicar brasileiros até 2030, diz ONG
5 março 2015
Mudanças climáticas e o crescimento urbano desorde
mais de 400 mil brasileiros expostos a enchentes fluv
um estudo divulgado hoje pela ONG World Resources

theguardian
UK world sport football opinion culture business lifestyle fashion environment tech travel
browse all sections
home › environment › climate change
wildlife energy pollution
Flooding

EL TIEMPO
Fuertes lluvias en Medellín causaron varias inundaciones
inundaciones
Las emergencias se presentaron principalmente en Envigado, Itagüí y El Poblado.
Comentar
Facebook 2
Twitter
Google+
Linkedin
Enviar

globo.com g1 globoesporte gshow famosos & etc vídeos
MENU
JORNAL NACIONAL
Estudo diz que enchentes provocam prejuízos de R$ 9 bilhões para o Brasil
Segundo um estudo da organização WRI, o número de pessoas afetadas por alagamentos em todo o mundo pode triplicar nos próximos 15 anos.
Jornal Nacional
g1 primeira página

Stuttgart Daily Leader
YOU CAN NOW READ 7 ARTICLES IN A MONTH FOR FREE
READ AS MUCH AS YOU WANT ANYWHERE AND ANY TIME FOR JUST 99¢
RISE: Trump to release Cuba policy; troops to Afghanistan; Scalise improving
Authorities say escaped Georgia inmates captured
Pence hires outside legal counsel
THE DRIVE: DC police seek arrest of Turks; Aussie PM parodies...
Judge to deadlocked Cosby jury: Keep trying to reach...


Ag flood damage grows to $175M
MOST POPULAR
1 Colvin finalist for SRO of the Year
Jun 13 at 1:31 PM
2 Surgeon added to BHMC-Stuttgart
Jun 13 at 1:40 PM
3 Mark L. Hopkins: New respect for...
Jun 12 at 9:01 AM
4 Tractors, equipment stolen in...
Jun 12 at 11:02 AM
OUR PICKS
Estimated impact of rain, floods on Arkansas cropland | May 8-12, 2017
Acres impacted
*$175 million *
Value range between $140 million and $210 million.
Value of storm-damaged acres: $75.84 million
Estimated value of crop loss by county

DEPORTES
ENTRETENIMIENTO
BIENESTAR
TECNOLOGÍA
ESPECIALES

O Dia » Notícia » Bras
Recomendar
27/04/2015 17:39:47
Enchent
pessoas
Há um muni
situação de emergência e cinco em situação de alerta
AGÊNCIA BRASIL
Amazonas - Balanço divulgado nesta segunda-feira pela Defesa Civil do Amazonas informa que as enchentes no estado já afetaram 110.610 pessoas, o que corresponde a 22.116 famílias. Neste momento há um município em estado de calamidade pública: Boca do Acre, localizado na microrregião de Purus; 18 em situação de emergência e cinco em situação de alerta. Segundo o governo do estado 363 toneladas de alimentos foram enviados a 19 municípios atingidos pelas cheias.

MEDIO AMBIENTE
Hay que repensar la gest

RCN Contigo
País
Política
Economía
Bogotá
Regiones
Gente

MAURICIO ARANGO

mergencia en el centro de
n por inundaciones
¿Cree
más fu
bloque
Sí
No

What is that?

Alternative sources

Social Media Information

Waze - Similar concept

waze LIVE MAP MAJOR EVENTS SUPPORT BLOG ABOUT Login

Nothing can beat real people working together

Imagine millions of drivers out on the roads, working together towards a common goal: to outsmart traffic and get everyone the best route to work and back, every day.

Traffic is more than just red lines on the map

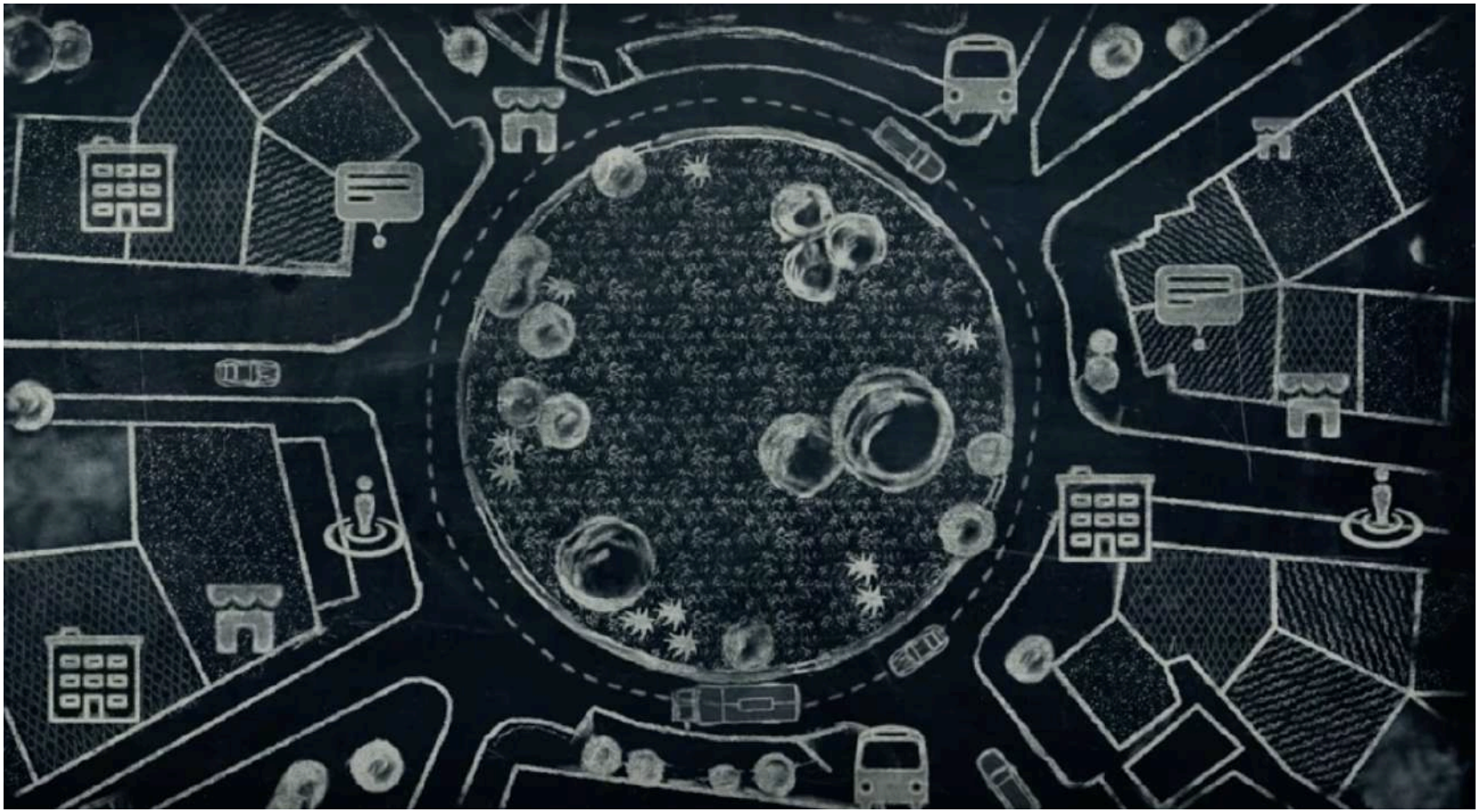
Get alerted before you approach police, accidents, road hazards or traffic jams, all shared by other drivers in real-time. It's like a personal heads-up from a few million of your friends on the road.

BIG DATA

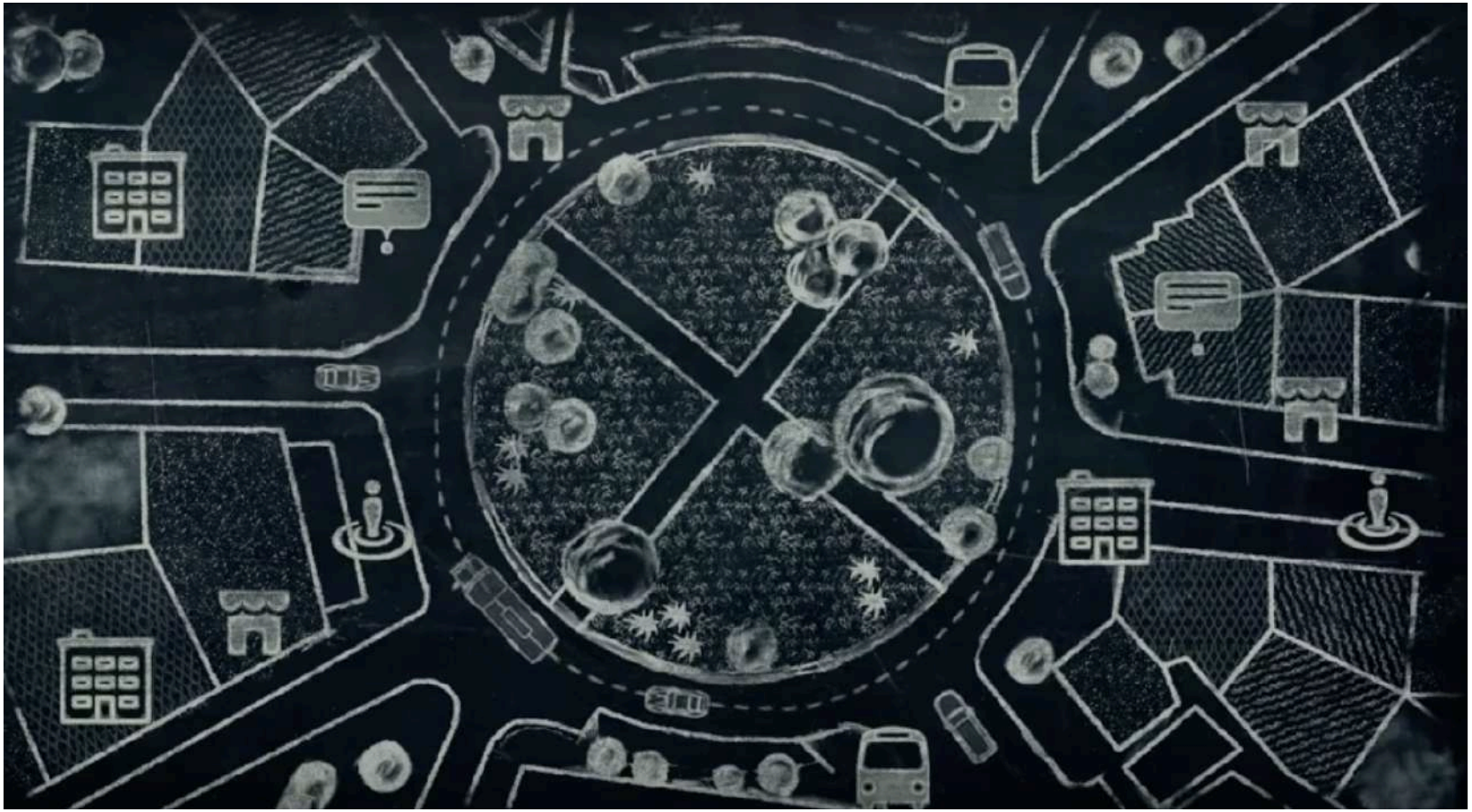
“Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big Data philosophy encompasses unstructured, semi-structured and structured data, however the main focus is on unstructured data.” source wikipedia

According to IBM: “Big Data is being generated at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it. Much of this data is coming to us in an unstructured form, making it difficult to put into structured tables with rows and columns. To extract insights from this complex data, Big Data projects often rely on cutting edge analytics involving data science and machine learning. Computers running sophisticated algorithms can help enhance the veracity of information by sifting through the noise created by Big Data's massive volume, variety, and velocity.”

BIG DATA

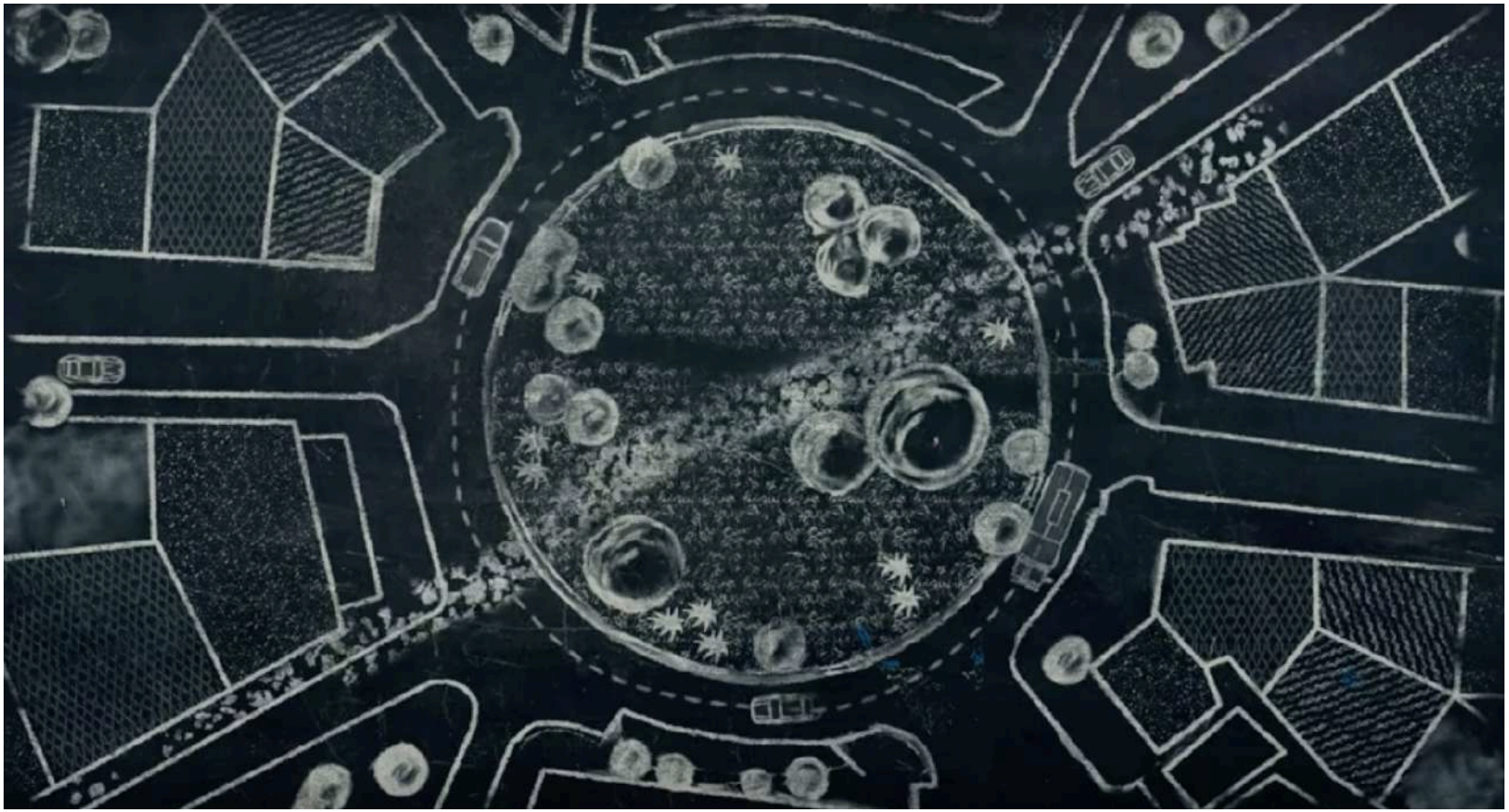


BIG DATA

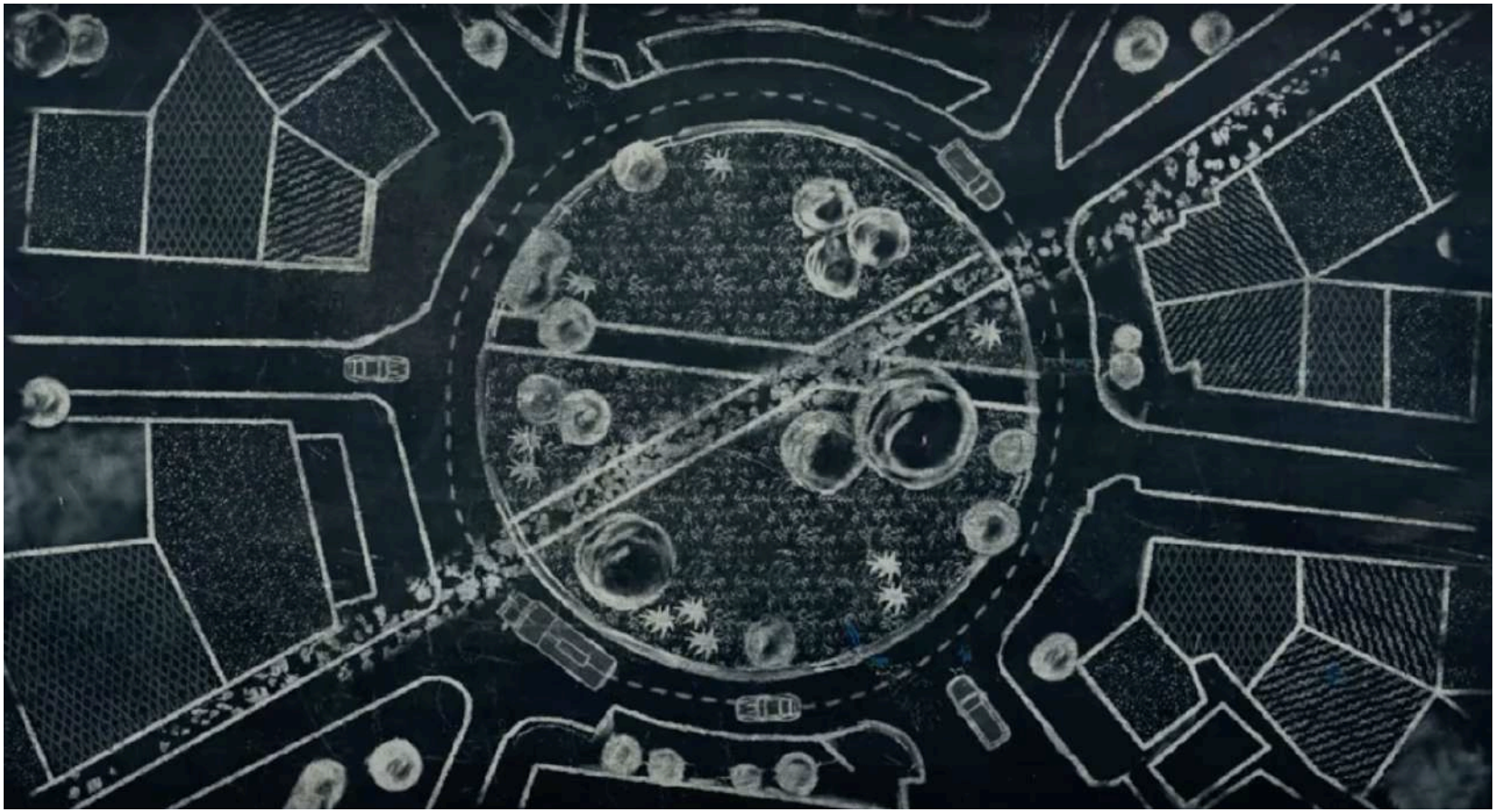


BIG DATA

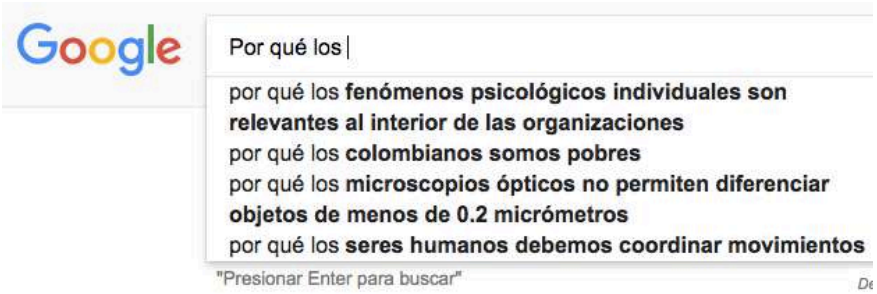
BIG DATA

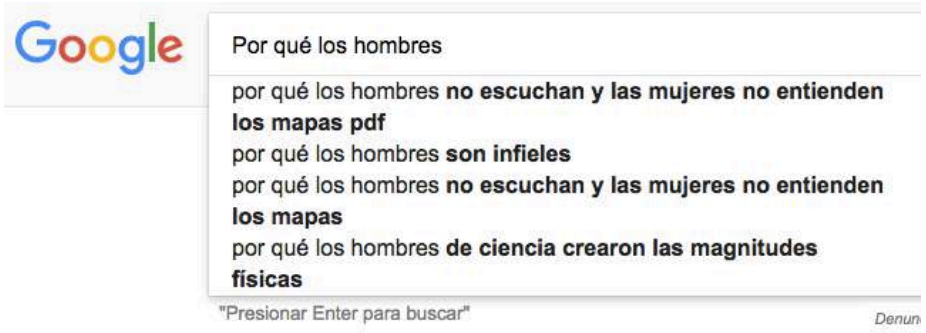


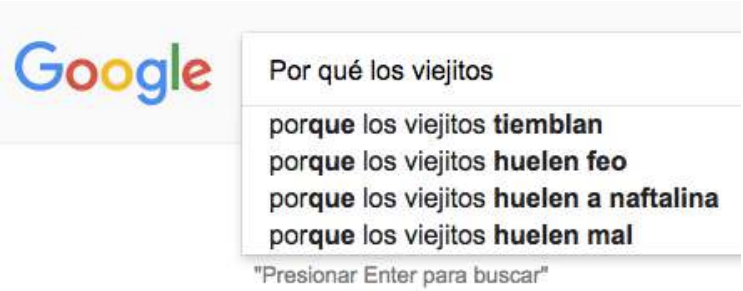
BIG DATA



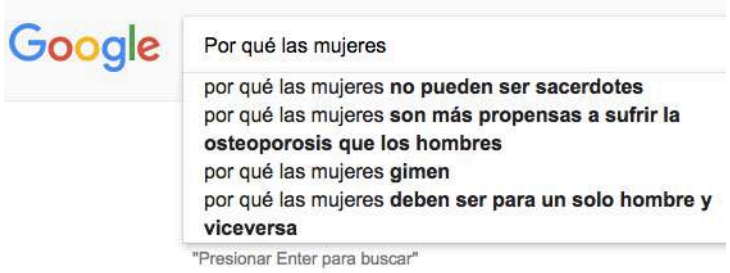
BIG DATA

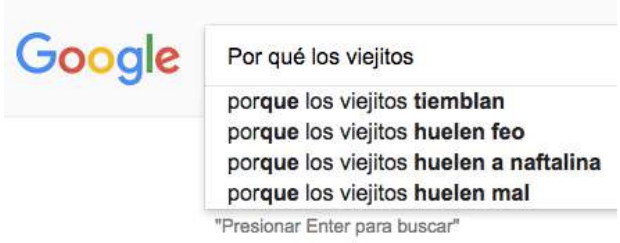

Google
Por qué los
por qué los **fenómenos psicológicos individuales son relevantes al interior de las organizaciones**
por qué los **colombianos somos pobres**
por qué los **microscopios ópticos no permiten diferenciar objetos de menos de 0.2 micrómetros**
por qué los **seres humanos debemos coordinar movimientos**
"Presionar Enter para buscar"


Google
Por qué los hombres
por qué los hombres **no escuchan y las mujeres no entienden los mapas pdf**
por qué los hombres **son infieles**
por qué los hombres **no escuchan y las mujeres no entienden los mapas**
por qué los hombres **de ciencia crearon las magnitudes físicas**
"Presionar Enter para buscar"


Google
Por qué los viejitos
porque los viejitos **tiemblan**
porque los viejitos **huelen feo**
porque los viejitos **huelen a naftalina**
porque los viejitos **huelen mal**
"Presionar Enter para buscar"

Google
Por qué los profesores
por qué los profesores **se resisten a planear sus cursos**
porque los profesores **estan en paro**
porque **protestan** los profesores
"Presionar Enter para buscar"


Google
Por qué las mujeres
por qué las mujeres **no pueden ser sacerdotes**
por qué las mujeres **son más propensas a sufrir la osteoporosis que los hombres**
por qué las mujeres **gimen**
por qué las mujeres **deben ser para un solo hombre y viceversa**
"Presionar Enter para buscar"


Google
Por qué los viejitos
porque los viejitos **tiemblan**
porque los viejitos **huelen feo**
porque los viejitos **huelen a naftalina**
porque los viejitos **huelen mal**
"Presionar Enter para buscar"

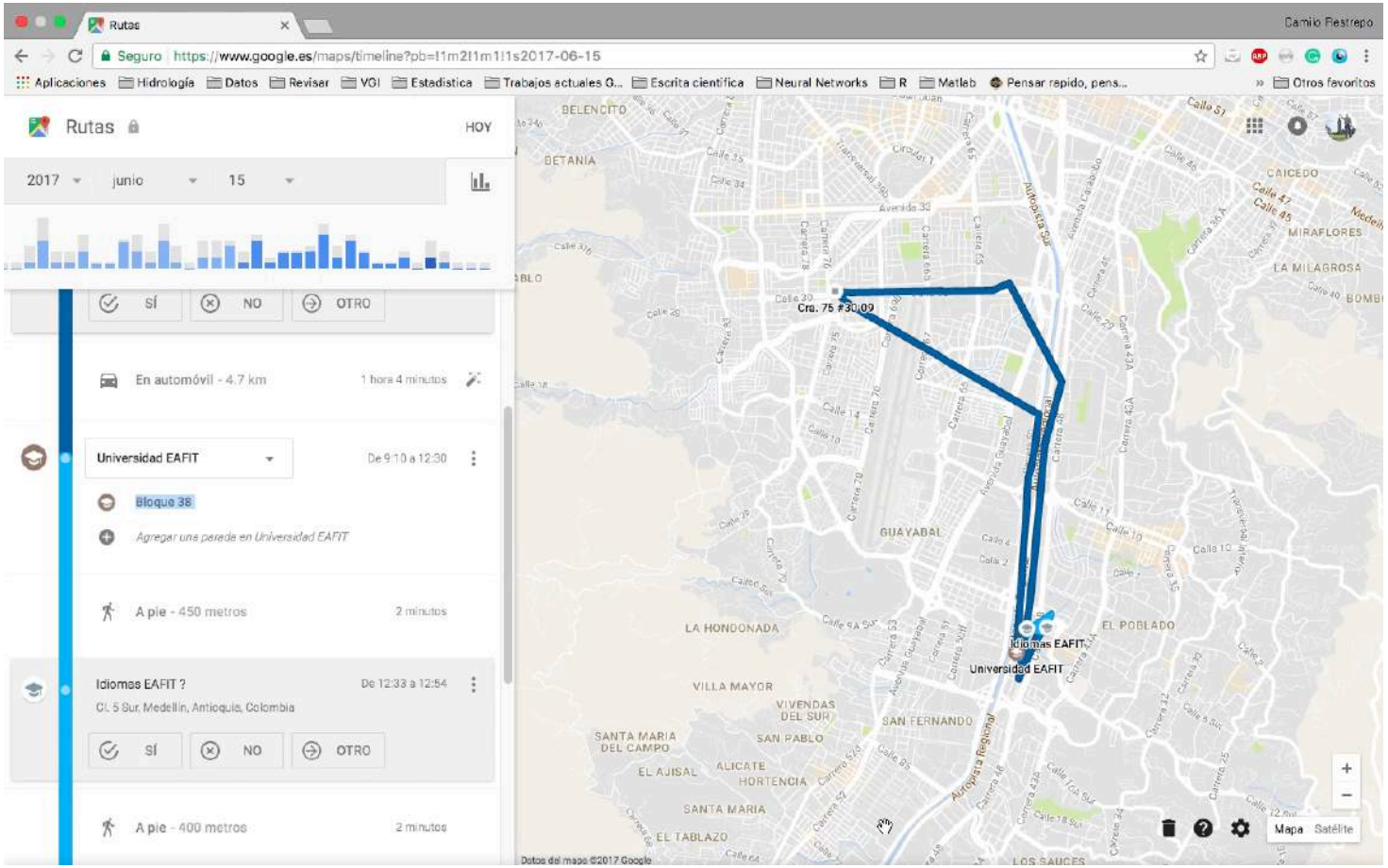
BIG DATA

ESTE ESTUDO FOI REALIZADO COM 65 TIPOS DE CEREAIS EM 10 SUPERMERCADOS
PERSONAGENS ADULTOS OLHAM PARA FRENTE
CEREAIS DE ADULTOS SÃO COLOCADOS NO DOBRO DA ALTURA DOS CEREAIS INFANTIS
PERSONAGENS INFANTIS OLHAM PARA BAIXO EM UM ÂNGULO DE 9,6°
ESSA TROCA DE OLHARES ENTRE O CONSUMIDOR E OS PERSONAGENS AUMENTA EM 16% A CONFIANÇA NA MARCA
FONTE:
CORNELL FOOD & BRAND LAB

TARGET
PARABÉNS, SUA FILHA ESTÁ GRÁVIDA!
The New York Times
Magazine
Times JOURNEYS
BOOK
How Companies Learn Your Secrets

How Google Flu Trends Works
O GOOGLE FLU TRENDS DETECTA EPIDEMIAS ANTES PORQUE ESTIMA A ATIVIDADE DA GRIPE EM TEMPO REAL

BIG DATA


Rutas
Damio Restrepo
Seguro https://www.google.es/maps/timeline?pb=!1m2!1m1!1s2017-06-15
Aplicaciones Hidrología Datos Revisar VGI Estadística Trabajos actuales G... Escrita científica Neural Networks R Matlab Pensar rapido, pens... Otros favoritos
Rutas
HOY
2017 junio 15
SÍ NO OTRO
En automóvil - 4.7 km 1 hora 4 minutos
Universidad EAFIT De 9:10 a 12:30
Bloque 38
Agregar una parada en Universidad EAFIT
A pie - 450 metros 2 minutos
Idiomas EAFIT ? De 12:33 a 12:54
Cl. 5 Sur, Medellín, Antioquia, Colombia
SÍ NO OTRO
A pie - 400 metros 2 minutos
Cra. 75 # 30-09
Idiomas EAFIT
Universidad EAFIT
BELENCITO
BETANIA
CAICEDO
MIRAFLORES
LA MILAGROSA
GUAYABAL
LA HONDONADA
EL POBLADO
VILLA MAYOR
VIVENDAS DEL SUR
SAN FERNANDO
SAN PABLO
SANTA MARIA DEL CAMPO
EL AJISAL
ALICANTE
HORTENCIA
SANTA MARIA
EL TABLAZO
LOS SAUCES
Mapa Satélite
Datos del mapa ©2017 Google

Social Media

Twitter Terms of Service

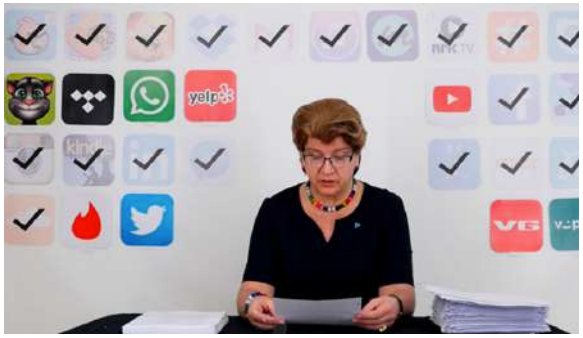
(if you live in the United States)

These Terms of Service ("**Terms**") govern your access to and use of our services, including our various websites, SMS, APIs, email notifications, applications, buttons, widgets, ads, commerce services, and our other covered services (https://support.twitter.com/articles/20172501) that link to these Terms (collectively, the "**Services**"), and any information, text, links, graphics, photos, videos, or other materials or arrangements of materials uploaded, downloaded or appearing on the Services (collectively referred to as "**Content**"). By using the Services you agree to be bound by these Terms.

1. Who May Use the Services
2. Privacy
3. Content on the Services
4. Using the Services
5. Disclaimers and Limitations of Liability
6. General

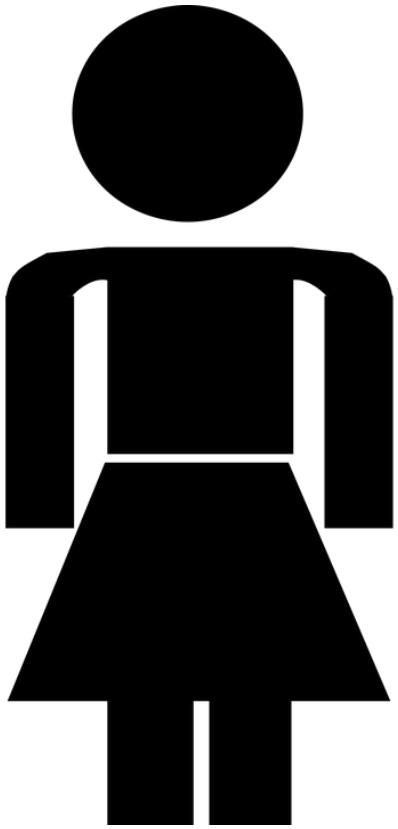
1. Who May Use the Services

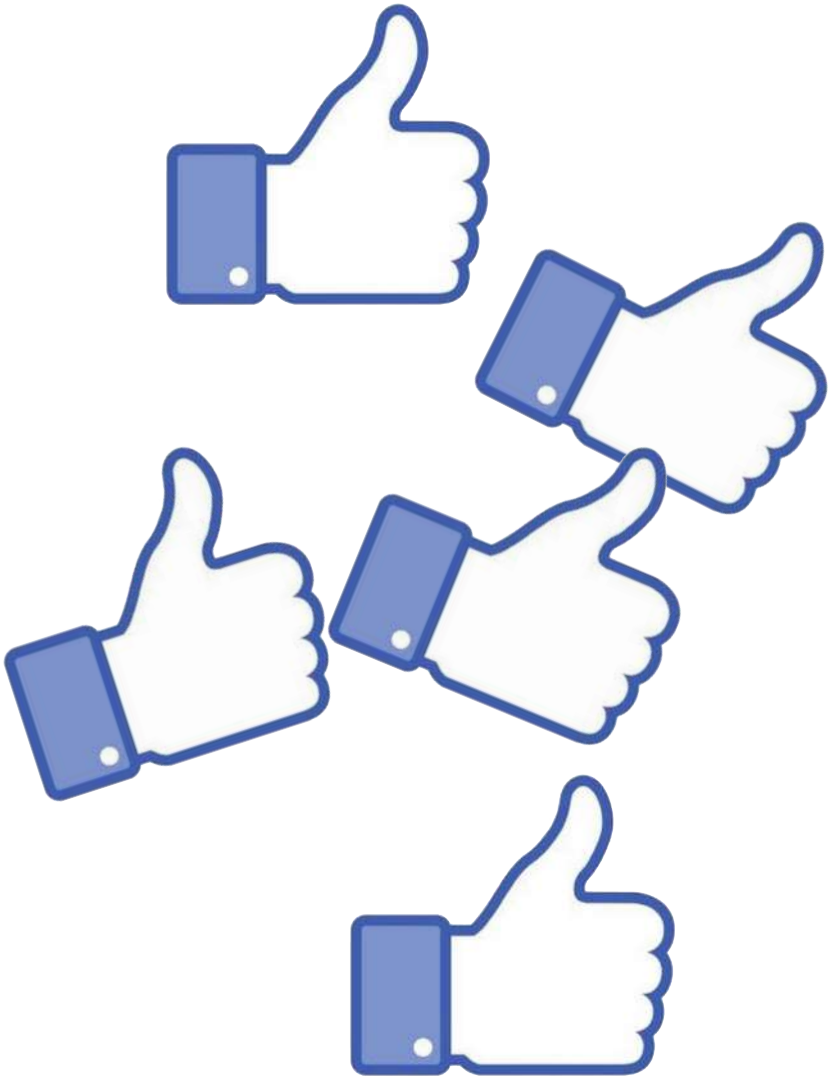
You may use the Services only if you agree to form a binding contract with Twitter and are not a person barred from receiving services under the laws of the applicable jurisdiction. In any case, you must be at least 13 years old to use the Services. If you are accepting these Terms and using the Services on behalf of a company, organization, government, or other legal entity, you represent and warrant that you are authorized to do so.



Social Media

Personality

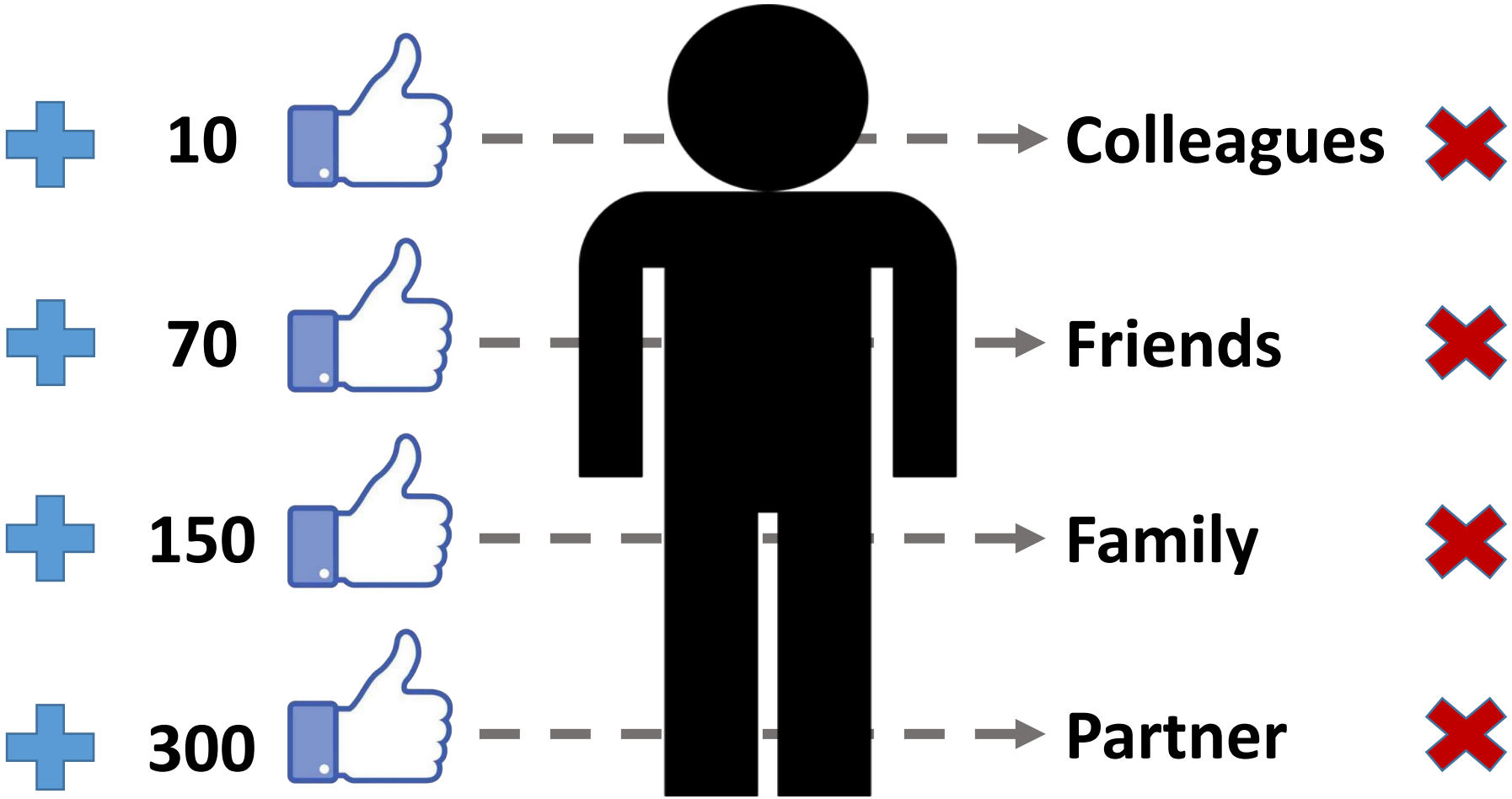




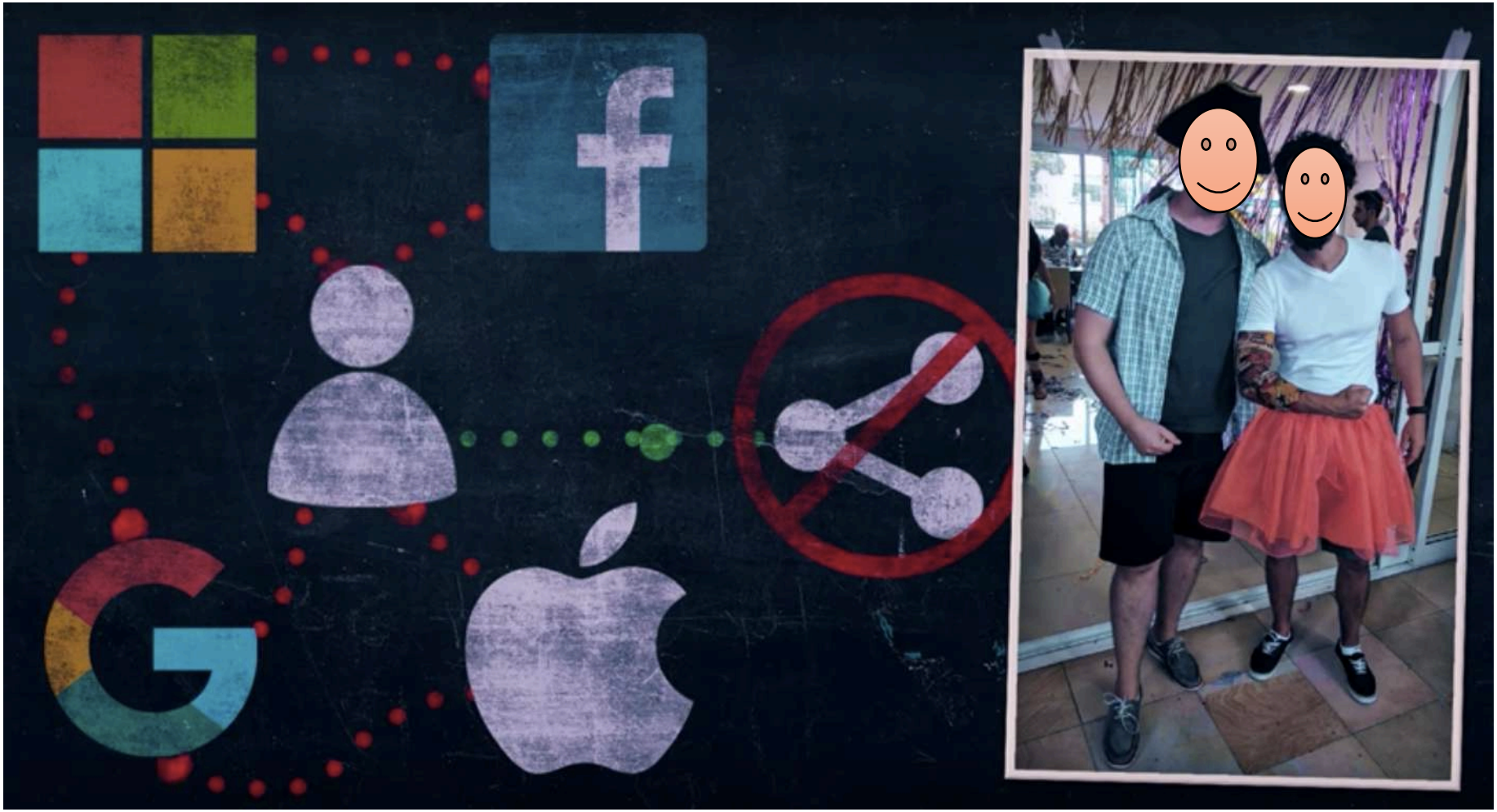
- Neuroticism
- Extroversion
- Kindness
- Politics filiation

Social Media

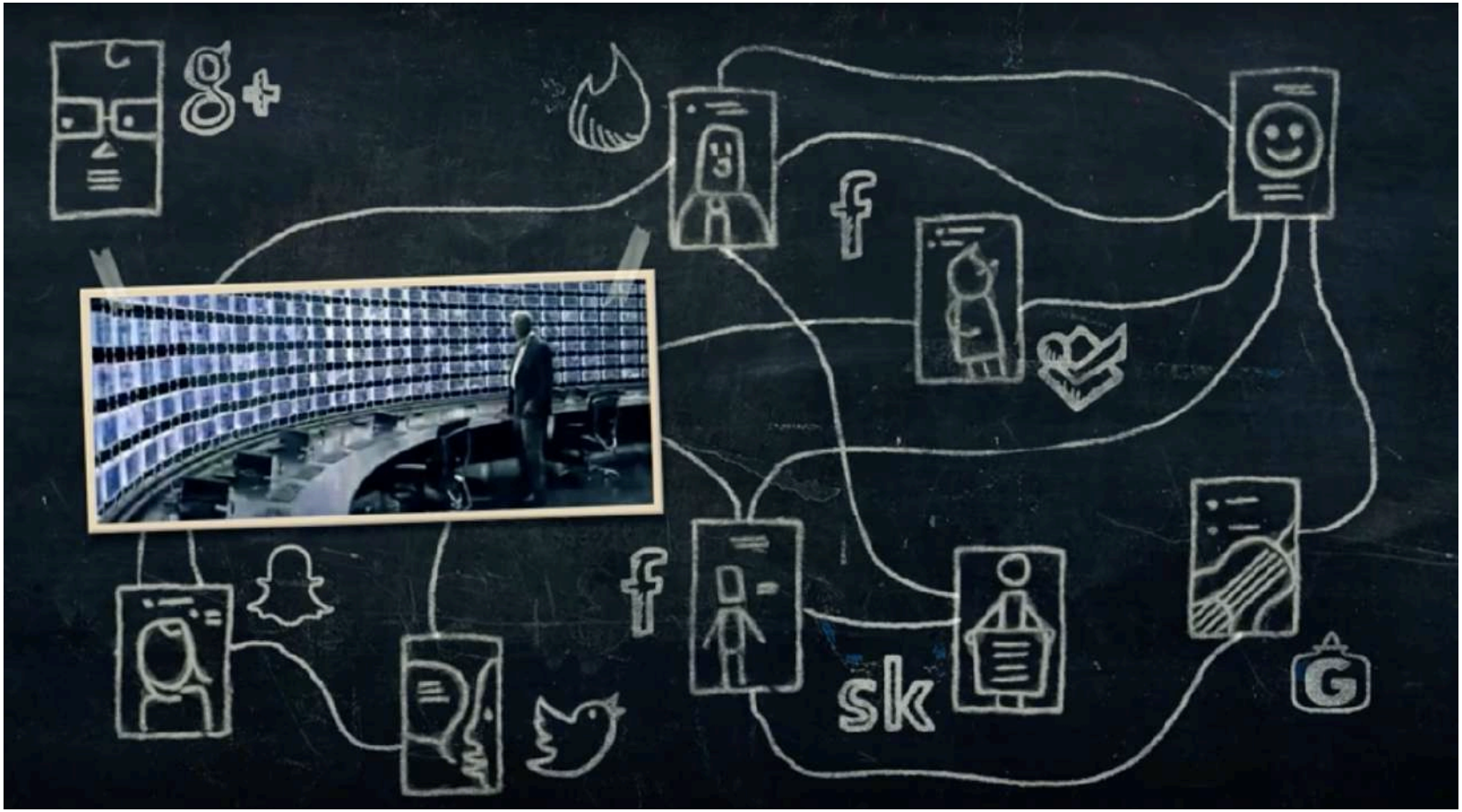
Personality



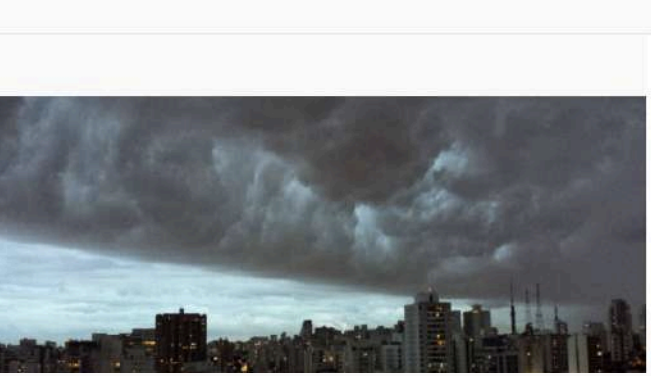
Social Media



Social Media



Social Media and hydrology



jheinerosa
Condominio Edificio B...

jheinerosa "Eu não sou senhor mas eu sei que vai chover..." #s #rain #vejasp #estadao #tvmin #splovers #saopaulocity #011

alex_minho WTF!

jheinerosa Tenso @alex_minho já começou a chuva

alex_minho Aqui também 😢

ericanacif Uauuu que foto

williamoleriano
Santa Cecília
Seguir

williamoleriano está chovendo e ao mesmo tempo nevando

lareism Nossaaaaaa...

williamoleriano Pois é muito gelo

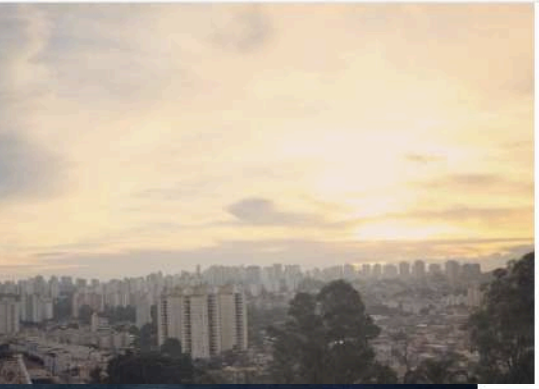
cahszsophia Nossa cunhado onde é isso?

williamoleriano Aqui na porta do meu serviço na Santa Cecília kkkkk, nevou e choveu ao mesmo tempo @cahszsophia

cahszsophia Meu Deus

115 visualizações

22 DE FEVEREIRO



vaniioberg
Condominio Vida Bela
Seguir

vaniioberg Faça sol ou chuva um lindo dia vai nascer

deborahalves_1
Esporte Clube Pinheir...
Seguir

deborahalves_1 Acho que vai chover um pouquinho por aqui! 😕#medo #chuva #terradagaroa #SP #pinhei #personalideh #sãopaulo #

carlotablotto Será???? 😂

castellanphotography Go

jacare_ems
Guarulhos

jacare_ems 02.02.2017 - PI ABORTADO POR CHUVA *Hoje não deu....vamos tent amanhã...rs www.strava.com/activities/l connect.garmin.com/moder 55321755 #bicycle #bike #bikemap #t #ciclismo #cycle #cycling #cyclingphotos #cyclist #g #garminconnect #instabike #mountainbike #mountainb #roadcycling #road #sport #stravacycling #stravaphot #vidadecicl ista #praquempe #sport #stravabrasil

reginaldomoura Impressior

wellington_original
Condominio Conjunto...
Seguir

wellington_original Será que tá chovendo???

vanbsm Rásca! 😍

Wellington Silva @wellsilva · 24 feb.
Será que tá chovendo??? @ Condominio Conjunto Morumbi
instagram.com/p/BQ6Hk1SjgaWI...

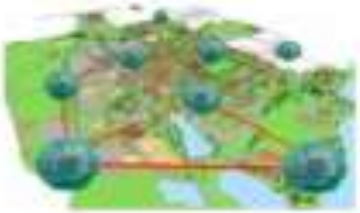
Traduzir del português

guistavares, hildetejesus65, pri_villanova, jessica_simplesassim, vanbsm, c3trio, gilvania_oliveira, flomueller87 e eulucke_blog curtiram isso

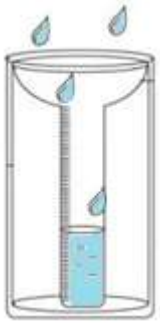
24 DE FEVEREIRO

Siga wellington_original para curtir ou

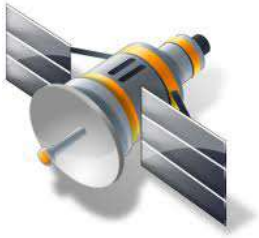
Context



Rede de sensores

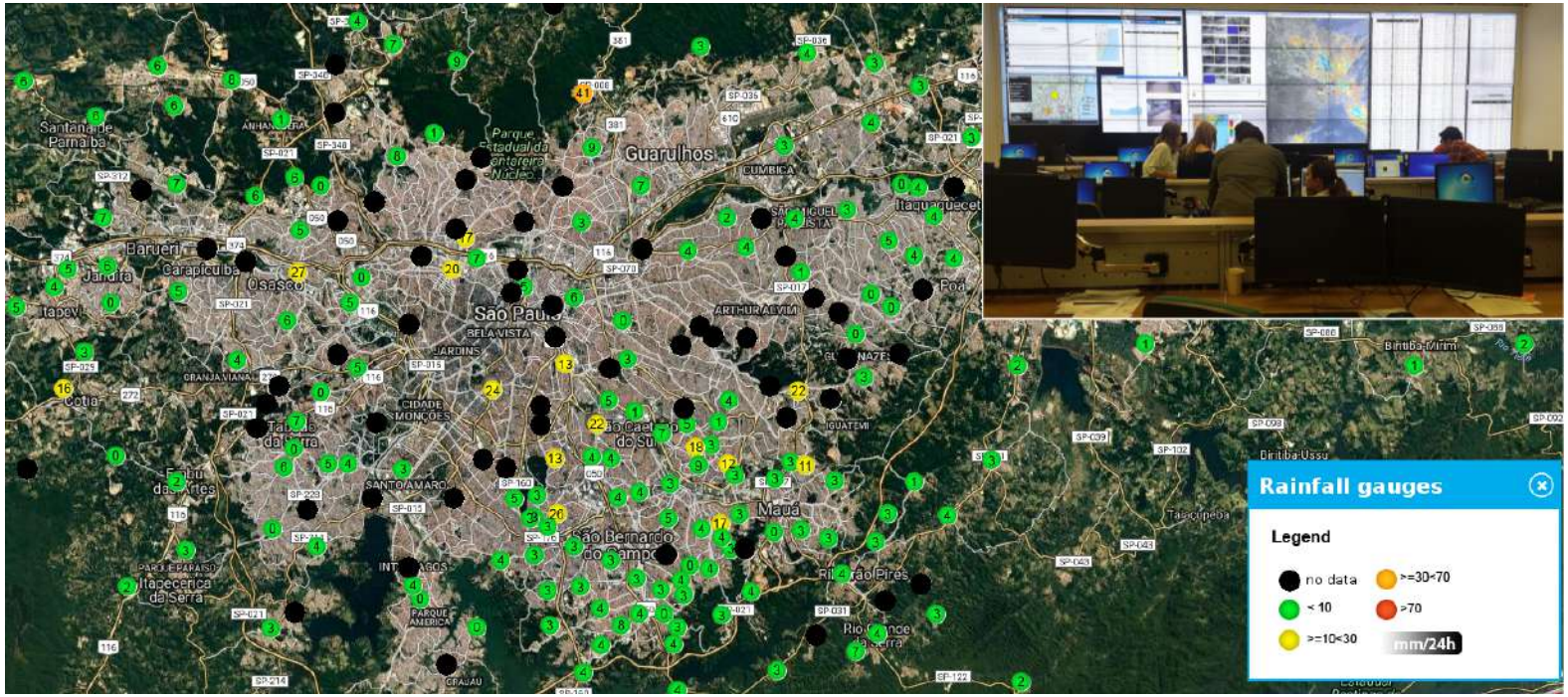


Pluviômetros

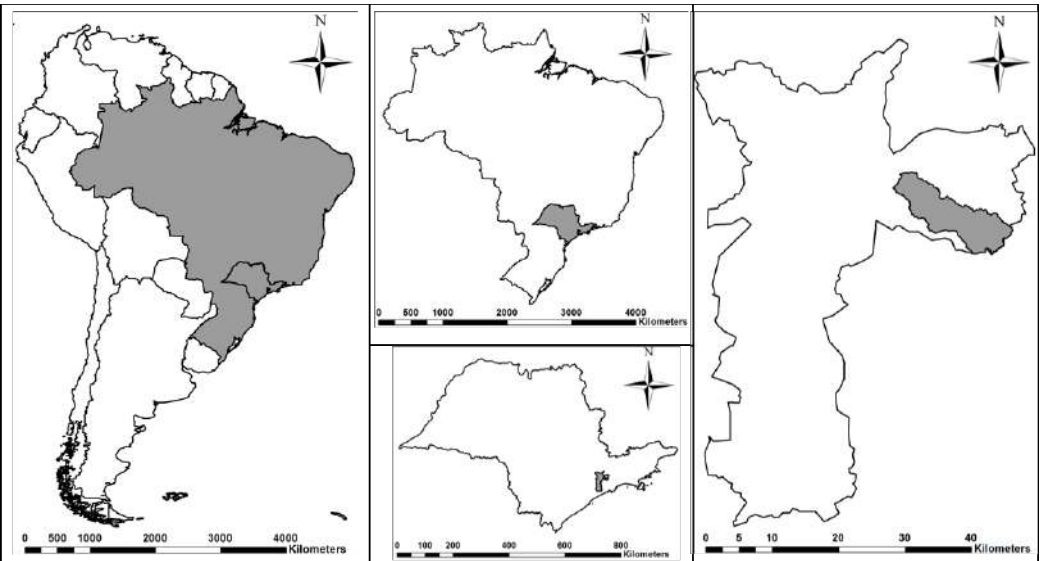


Satélites

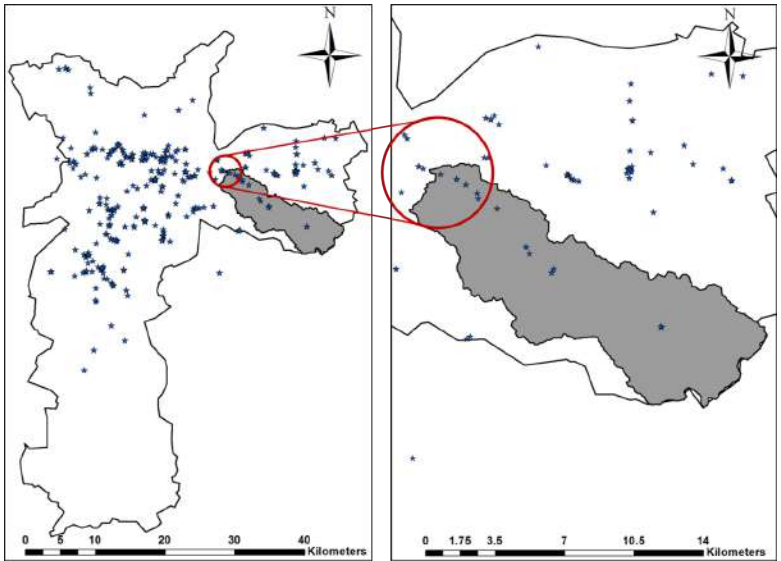
Problem statement



Aricanduva catchment

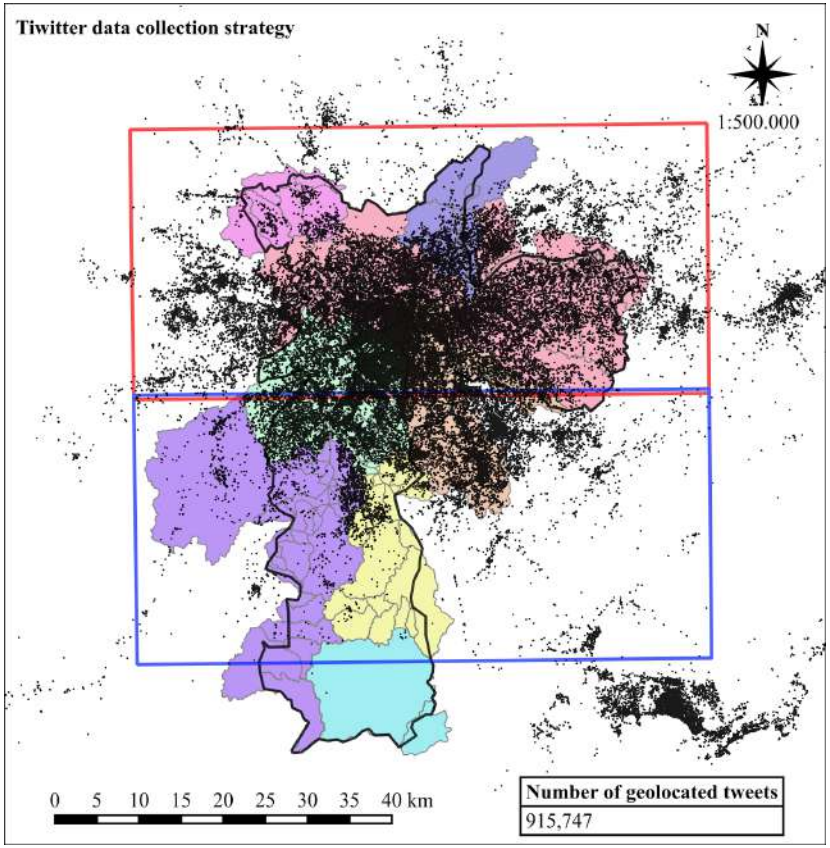


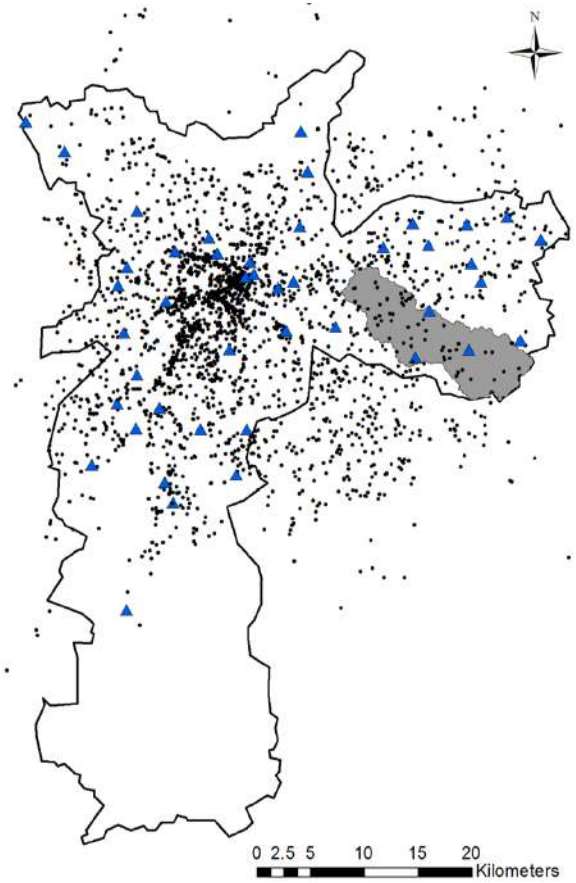
SAISP floods

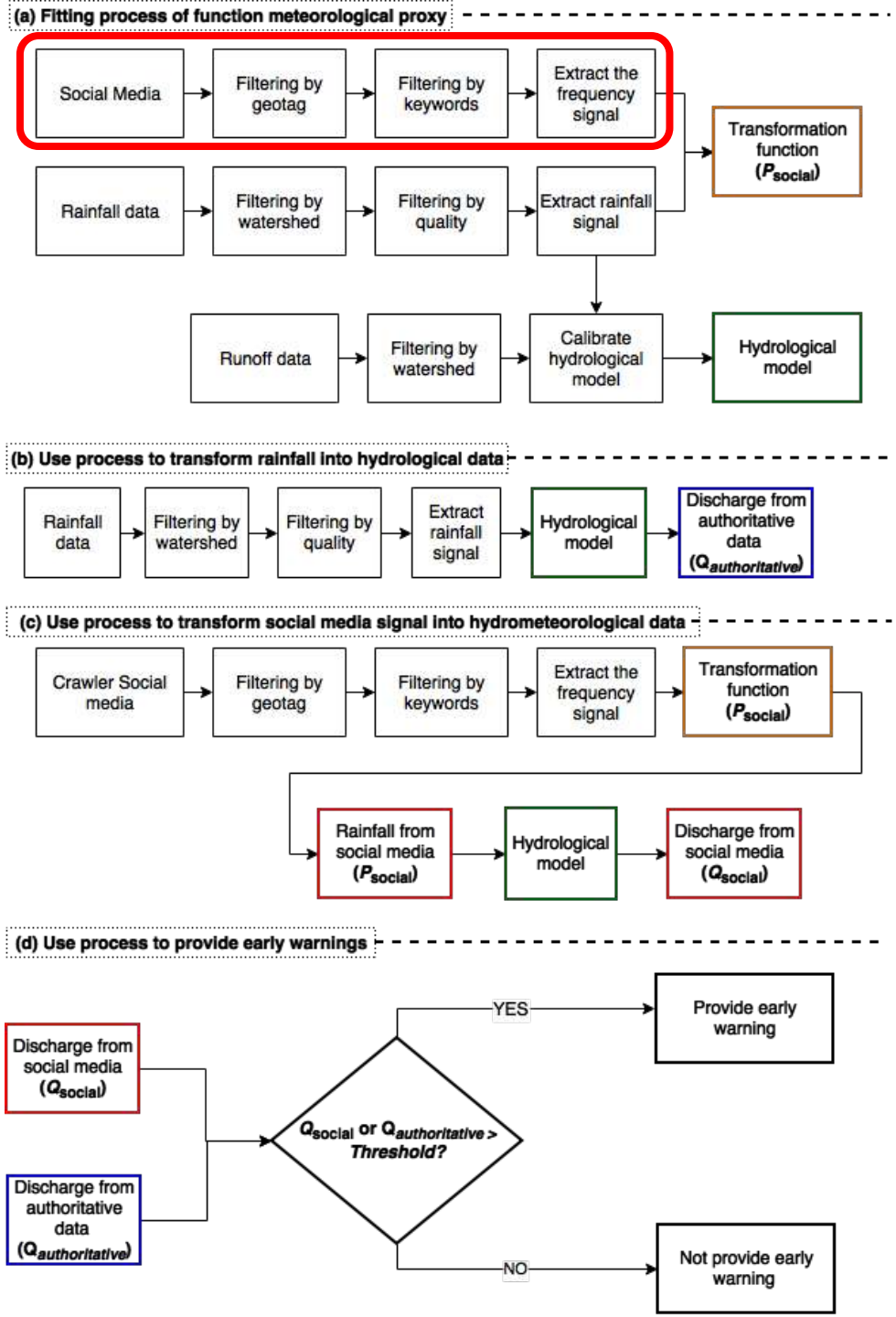


Related geolocated tweets

City of São Paulo with related tweets as black points, rainfall gauges as blue triangles and the Aricanduva catchment shaded gray.

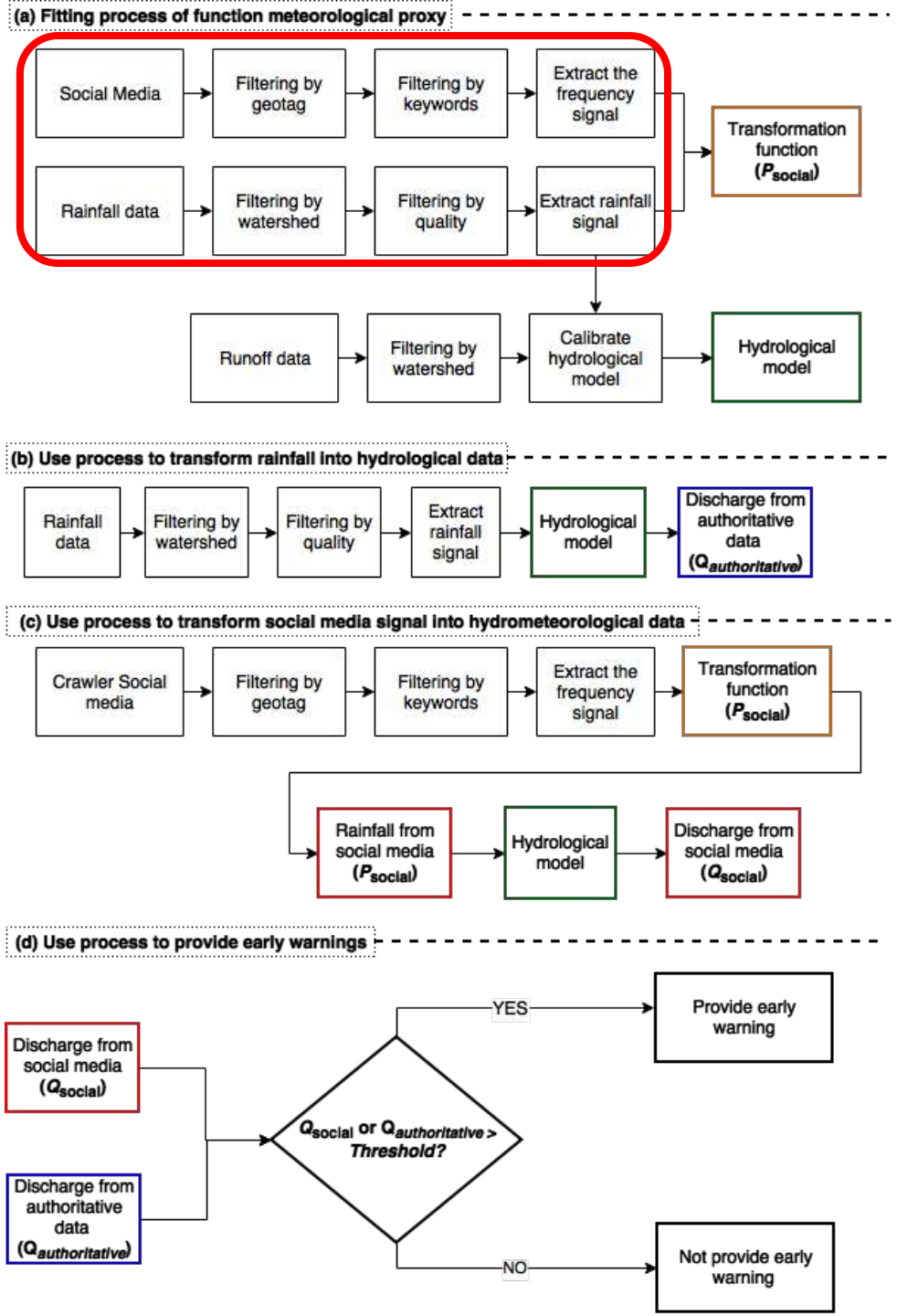




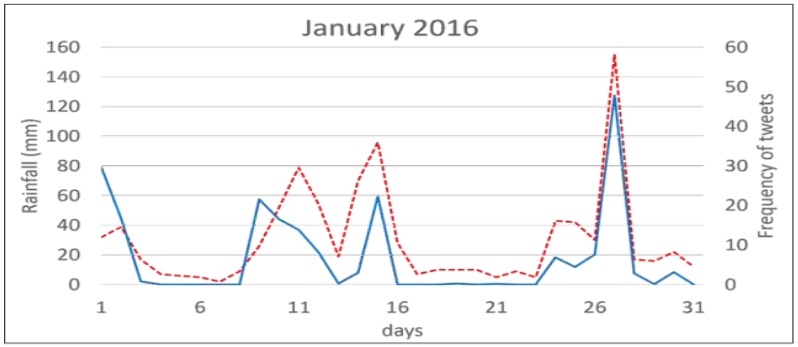

(a) Fitting process of function meteorological proxy
Social Media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Runoff data
Filtering by watershed
Calibrate hydrological model
Hydrological model
(b) Use process to transform rainfall into hydrological data
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Hydrological model
Discharge from authoritative data ($Q_{authoritative}$)
(c) Use process to transform social media signal into hydrometeorological data
Crawler Social media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall from social media (P_{social})
Hydrological model
Discharge from social media (Q_{social})
(d) Use process to provide early warnings
Discharge from social media (Q_{social})
Discharge from authoritative data ($Q_{authoritative}$)
Q_{social} or $Q_{authoritative}$ > Threshold?
YES
Provide early warning
NO
Not provide early warning

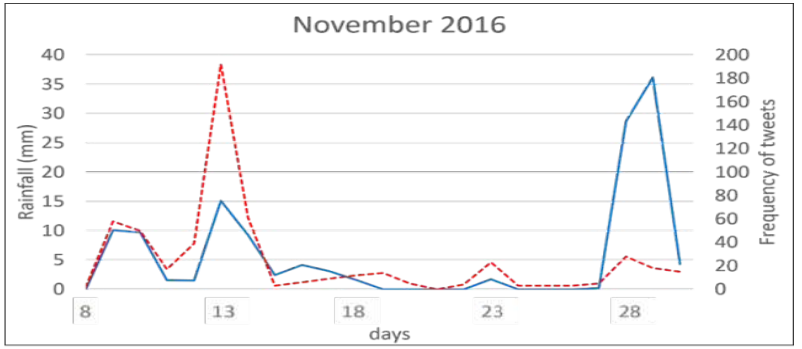
Key words and wrong key words

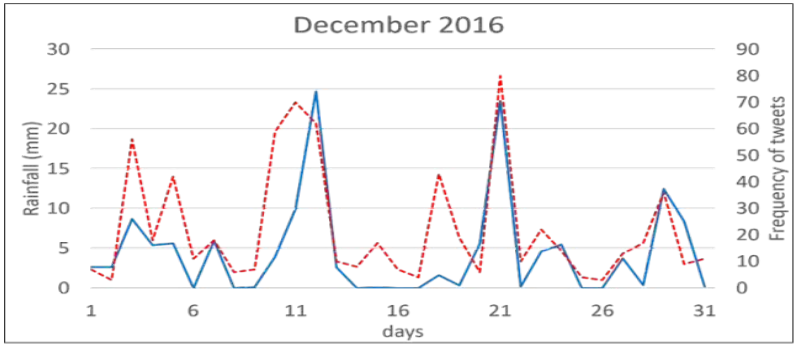
Variable	Words in Portuguese	Words in English
Keywords	%chuv% (chuva, chuvinha, chuvisco) %chov% (chove, chovendo, chover, choveu)	rain, rainy, drizzle It rains, to rain, rained
Wrong keywords	%bolinho de chuva%, %chuveiro%, %chuva negra%, %chuva de foto%, %chuva de benção%, %chuva de graça%, %chuva de confete%, %chuva de arroz%, %chuva de glitter%, %chuva de espuma%, %chuva de amor%, %chuva de mec%, %chuva de abacate%, %chuva de sol%	sweet snack, shower, rap group calls chuva negra, photo rain, rain of blessing, rain of grace, confetti rain, rice rain, glitter rain, foam rain, love rain, mec rain, avocado rain, sun rain.
Reinforcement	%chove forte%, %chove muito%, %muita chuva%, %chovendo muito%, %chove muito forte%	it rains hard, It rains a lot, lots of rain, It's raining a lot, It's raining very strong
Weakening	%arco íris%	rainbow

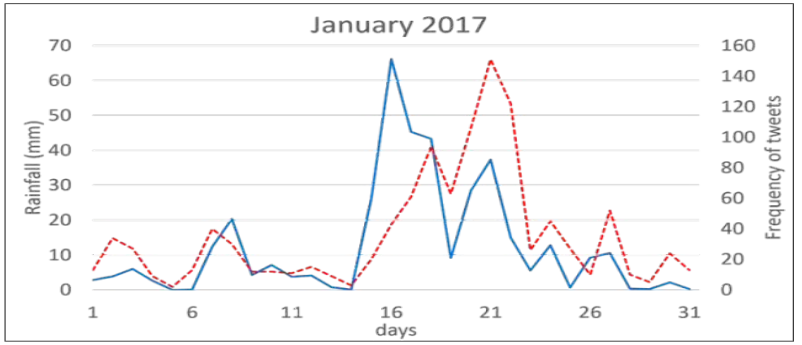

(a) Fitting process of function meteorological proxy
Social Media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Transformation function (P_{social})
Runoff data
Filtering by watershed
Calibrate hydrological model
Hydrological model
(b) Use process to transform rainfall into hydrological data
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Hydrological model
Discharge from authoritative data ($Q_{authoritative}$)
(c) Use process to transform social media signal into hydrometeorological data
Crawler Social media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall from social media (P_{social})
Hydrological model
Discharge from social media (Q_{social})
(d) Use process to provide early warnings
Discharge from social media (Q_{social})
Discharge from authoritative data ($Q_{authoritative}$)
Q_{social} or $Q_{authoritative}$ > Threshold?
YES
Provide early warning
NO
Not provide early warning

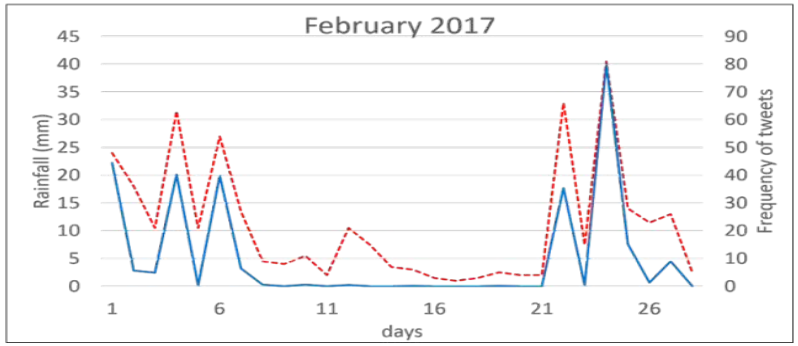
Preliminary results

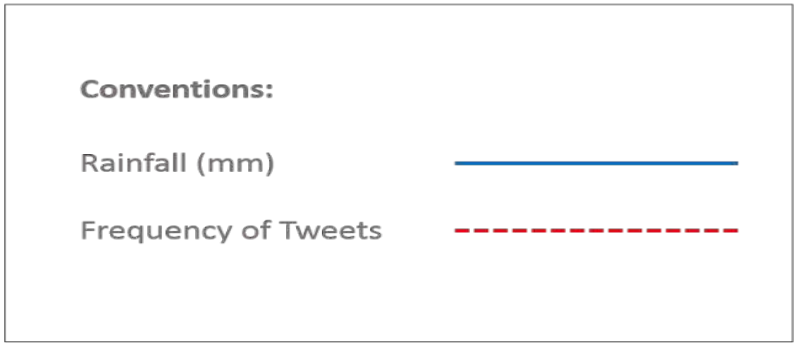


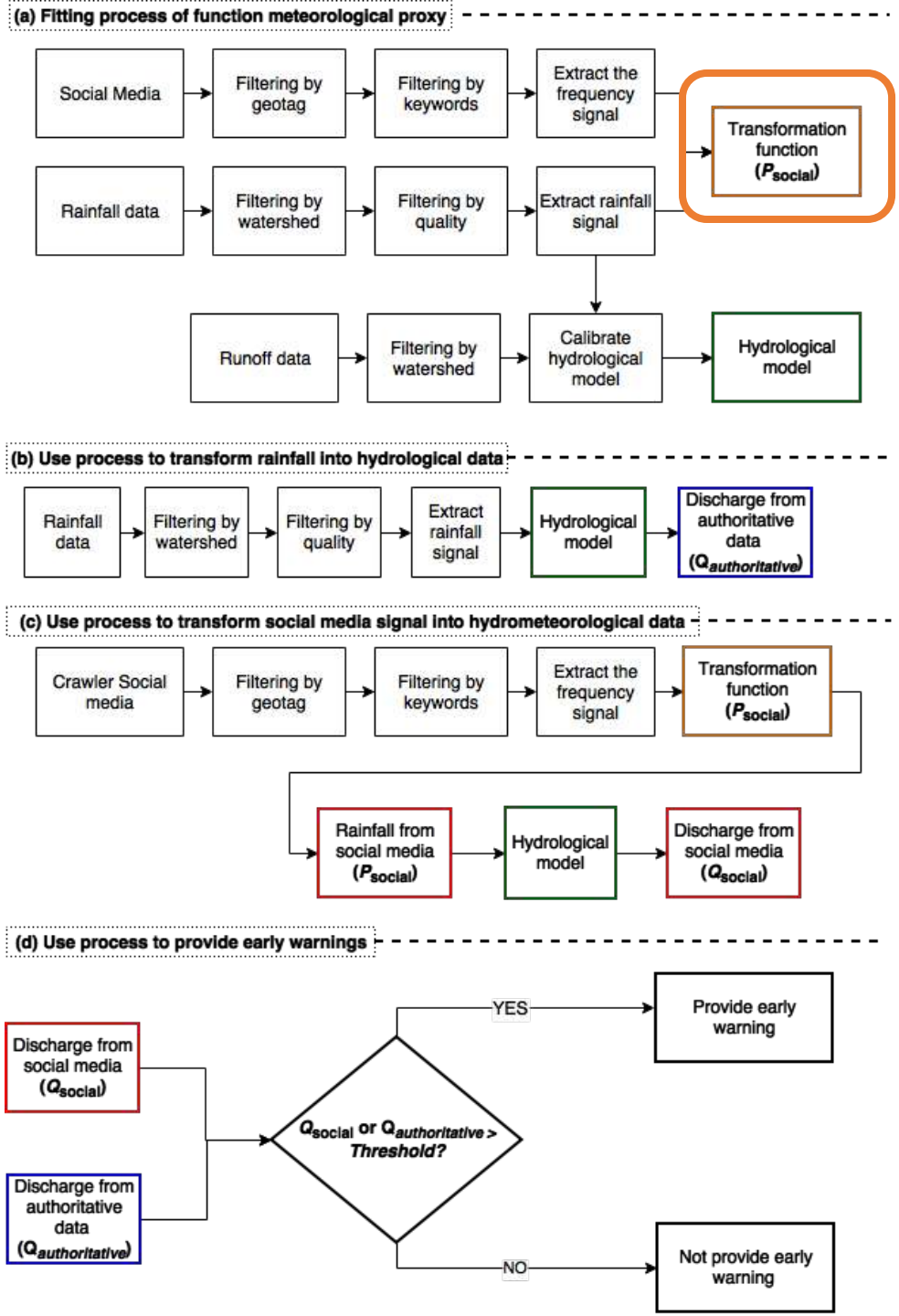






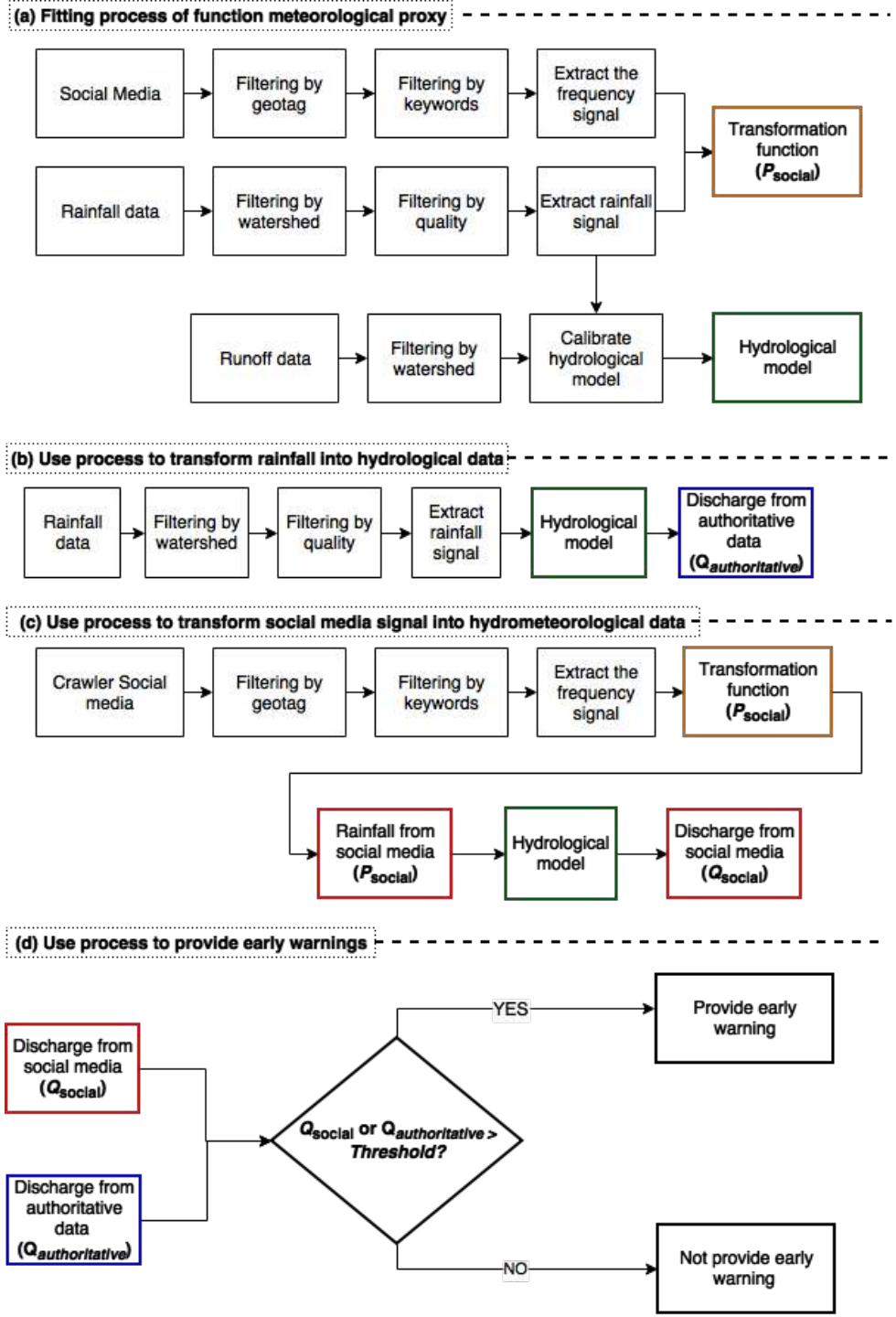


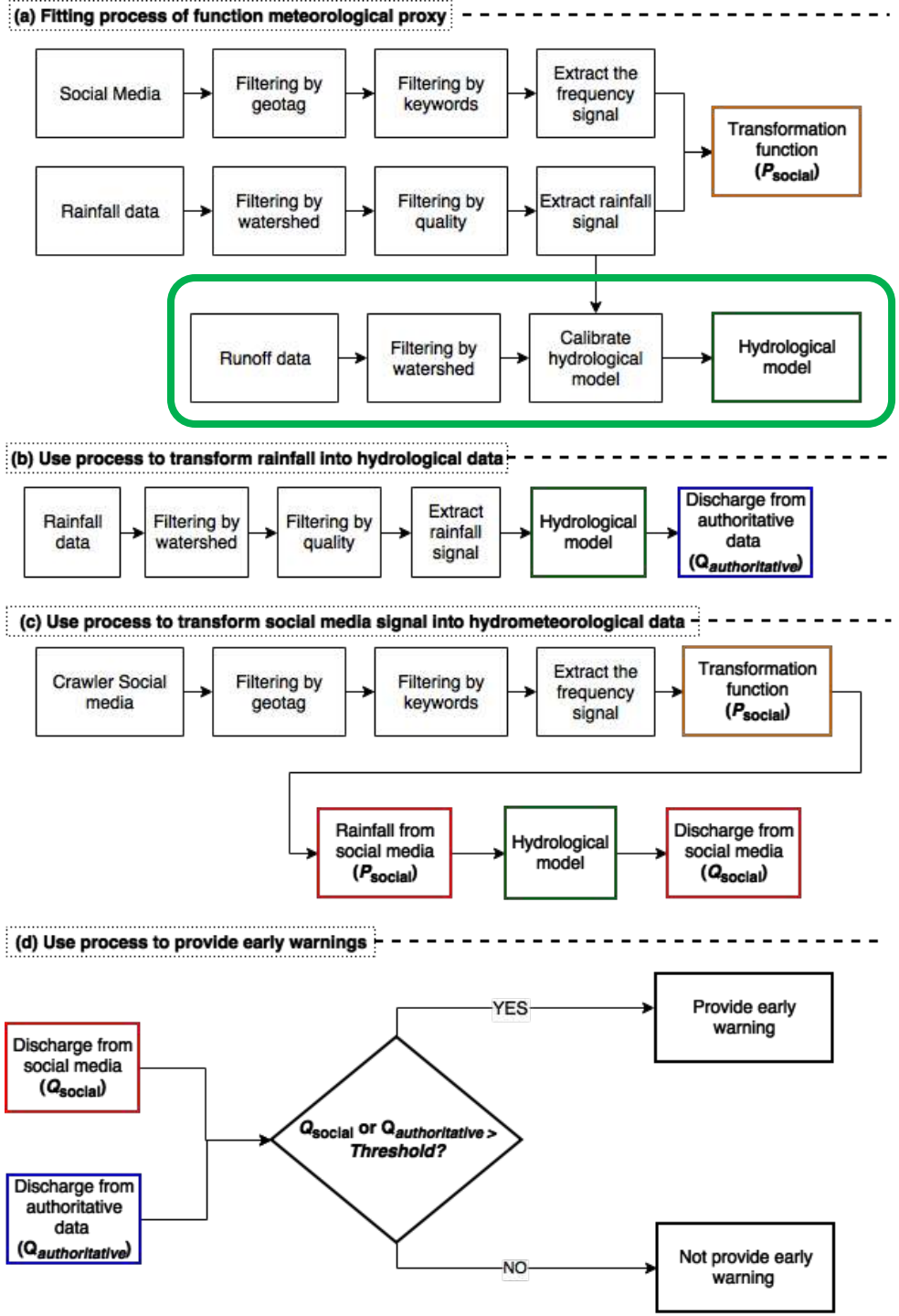



(a) Fitting process of function meteorological proxy
Social Media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Runoff data
Filtering by watershed
Calibrate hydrological model
Hydrological model
(b) Use process to transform rainfall into hydrological data
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Hydrological model
Discharge from authoritative data ($Q_{authoritative}$)
(c) Use process to transform social media signal into hydrometeorological data
Crawler Social media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall from social media (P_{social})
Hydrological model
Discharge from social media (Q_{social})
(d) Use process to provide early warnings
Discharge from social media (Q_{social})
Discharge from authoritative data ($Q_{authoritative}$)
Q_{social} or $Q_{authoritative}$ > Threshold?
YES
Provide early warning
NO
Not provide early warning

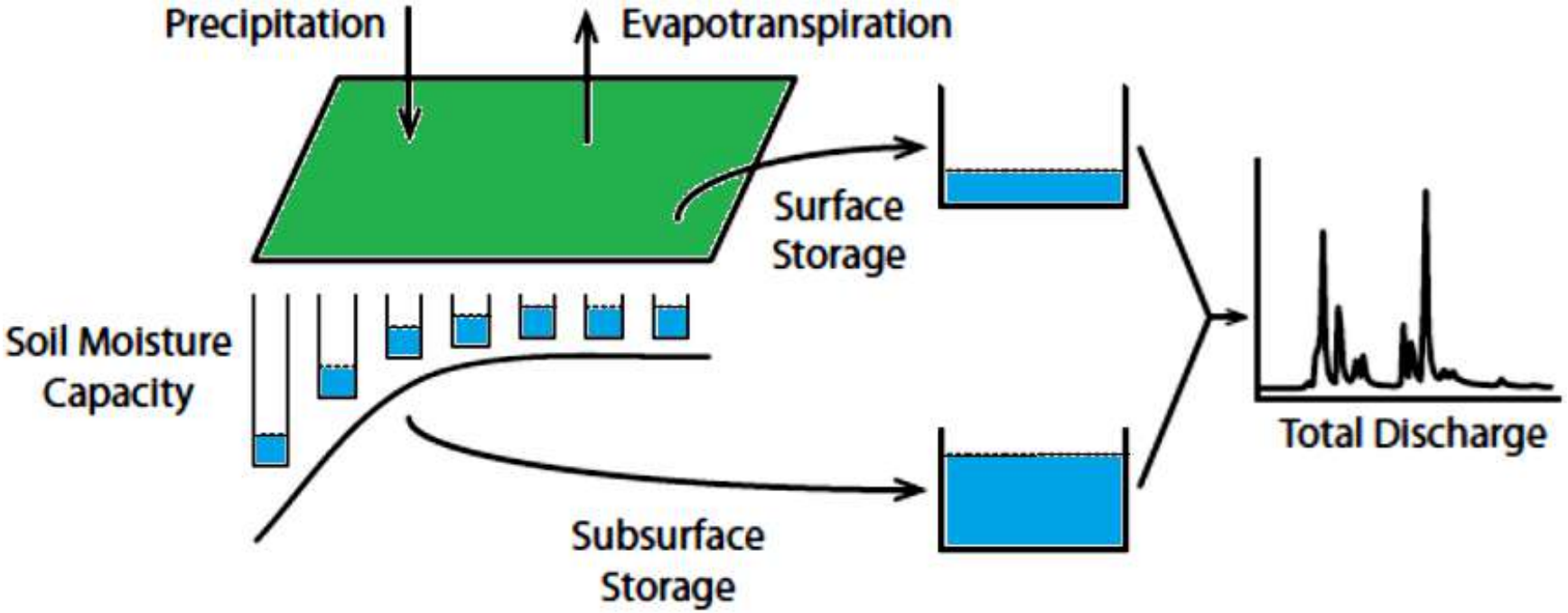
Transformation function

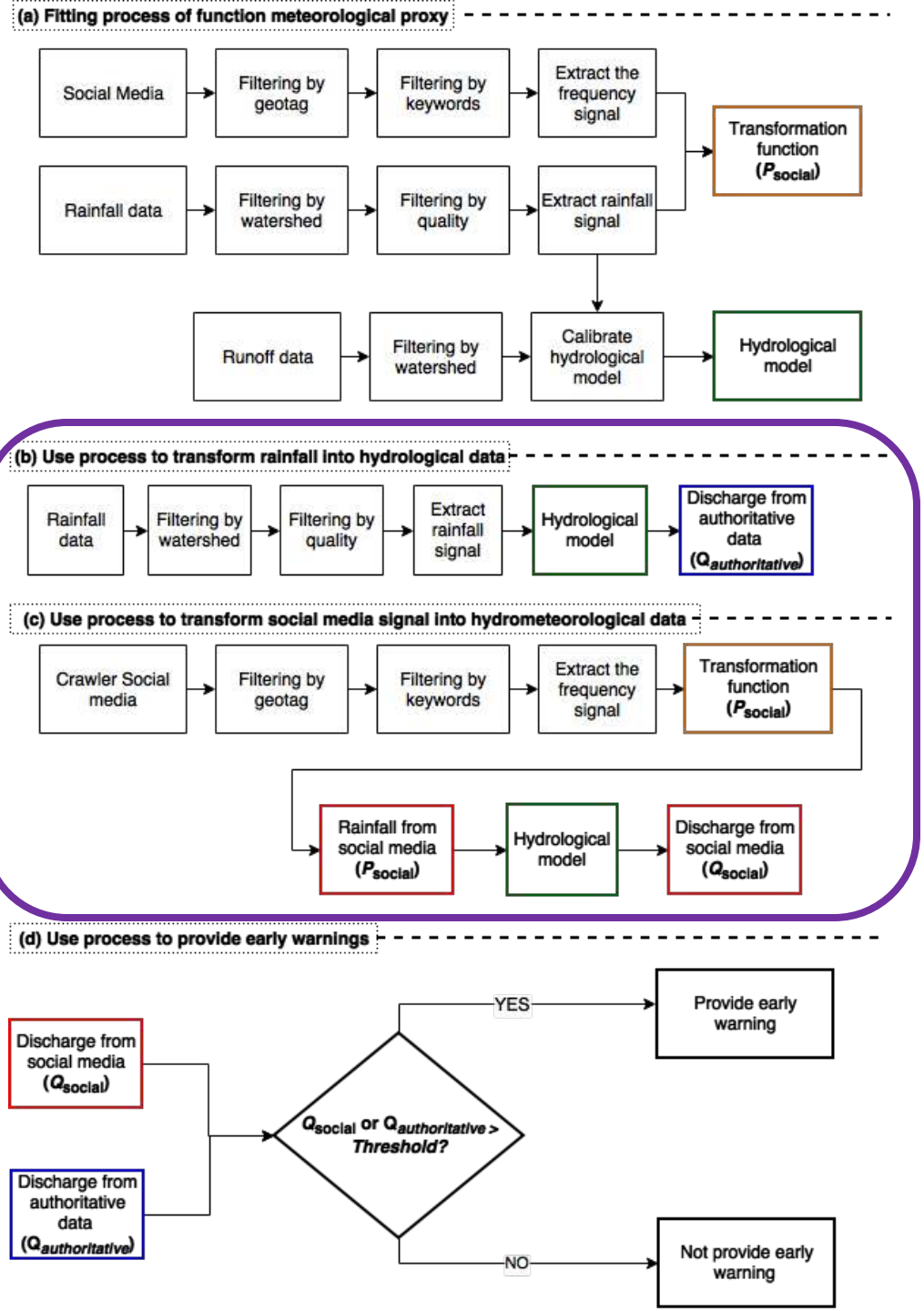
$$p_{social} = \alpha(1 + \eta_{strong} + \eta_{soft}) \frac{f_{kw}}{A_{interest}} + \sum_{i=2}^{n} \beta_i \frac{F_{kw(i)}}{A_{interest}}$$


(a) Fitting process of function meteorological proxy
Social Media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Runoff data
Filtering by watershed
Calibrate hydrological model
Hydrological model
(b) Use process to transform rainfall into hydrological data
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Hydrological model
Discharge from authoritative data ($Q_{authoritative}$)
(c) Use process to transform social media signal into hydrometeorological data
Crawler Social media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall from social media (P_{social})
Hydrological model
Discharge from social media (Q_{social})
(d) Use process to provide early warnings
Discharge from social media (Q_{social})
Discharge from authoritative data ($Q_{authoritative}$)
Q_{social} or $Q_{authoritative}$ > Threshold?
YES
Provide early warning
NO
Not provide early warning

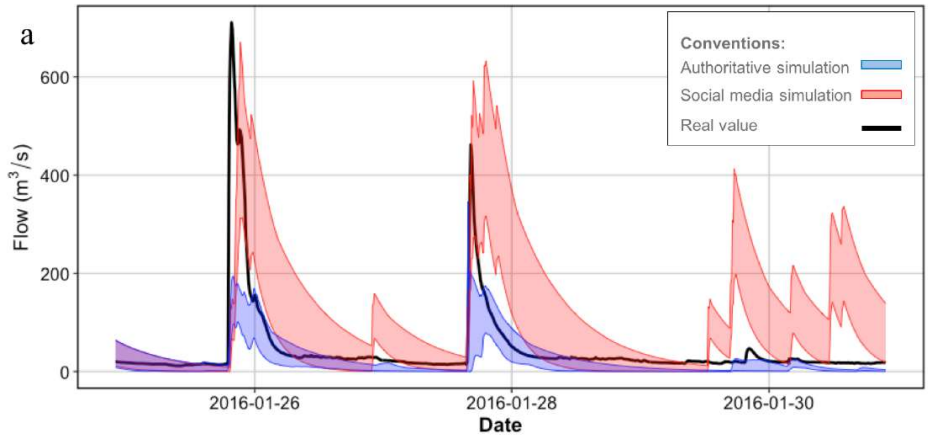

(a) Fitting process of function meteorological proxy
Social Media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Transformation function (P_{social})
Runoff data
Filtering by watershed
Calibrate hydrological model
Hydrological model
(b) Use process to transform rainfall into hydrological data
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Hydrological model
Discharge from authoritative data ($Q_{authoritative}$)
(c) Use process to transform social media signal into hydrometeorological data
Crawler Social media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall from social media (P_{social})
Hydrological model
Discharge from social media (Q_{social})
(d) Use process to provide early warnings
Discharge from social media (Q_{social})
Discharge from authoritative data ($Q_{authoritative}$)
Q_{social} or $Q_{authoritative}$ > Threshold?
YES
Provide early warning
NO
Not provide early warning

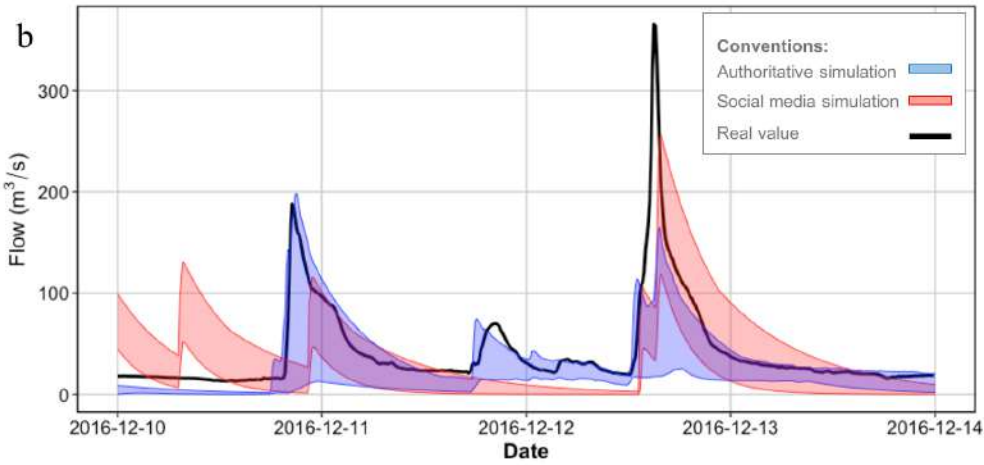
PROBABILITY DISTRIBUTED MODEL (PDM)

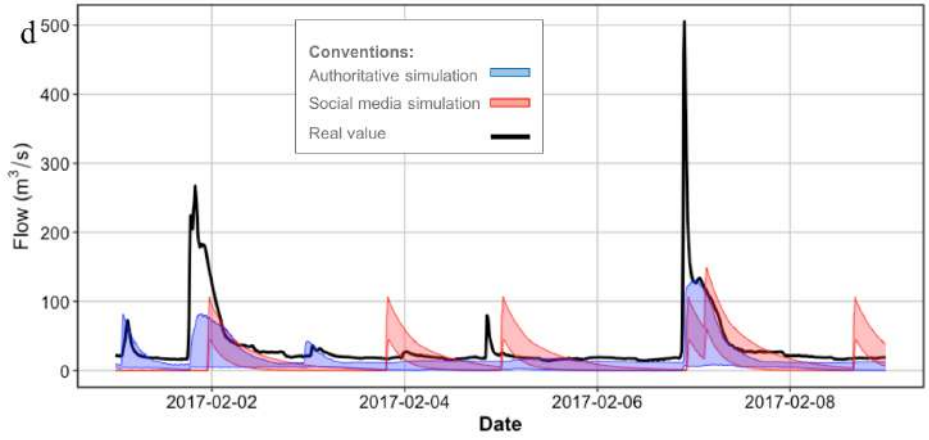


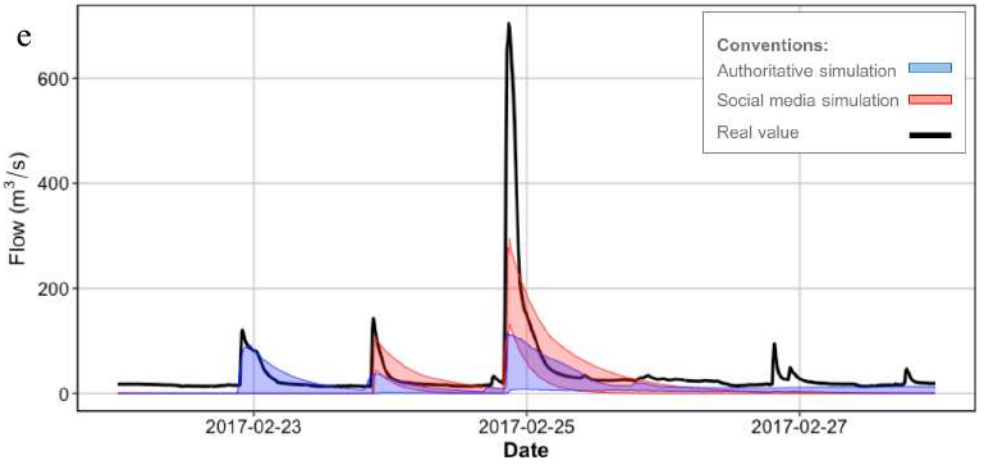

(a) Fitting process of function meteorological proxy
Social Media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Runoff data
Filtering by watershed
Calibrate hydrological model
Hydrological model
(b) Use process to transform rainfall into hydrological data
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Hydrological model
Discharge from authoritative data ($Q_{authoritative}$)
(c) Use process to transform social media signal into hydrometeorological data
Crawler Social media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall from social media (P_{social})
Hydrological model
Discharge from social media (Q_{social})
(d) Use process to provide early warnings
Discharge from social media (Q_{social})
Discharge from authoritative data ($Q_{authoritative}$)
Q_{social} or $Q_{authoritative}$ > Threshold?
YES
Provide early warning
NO
Not provide early warning

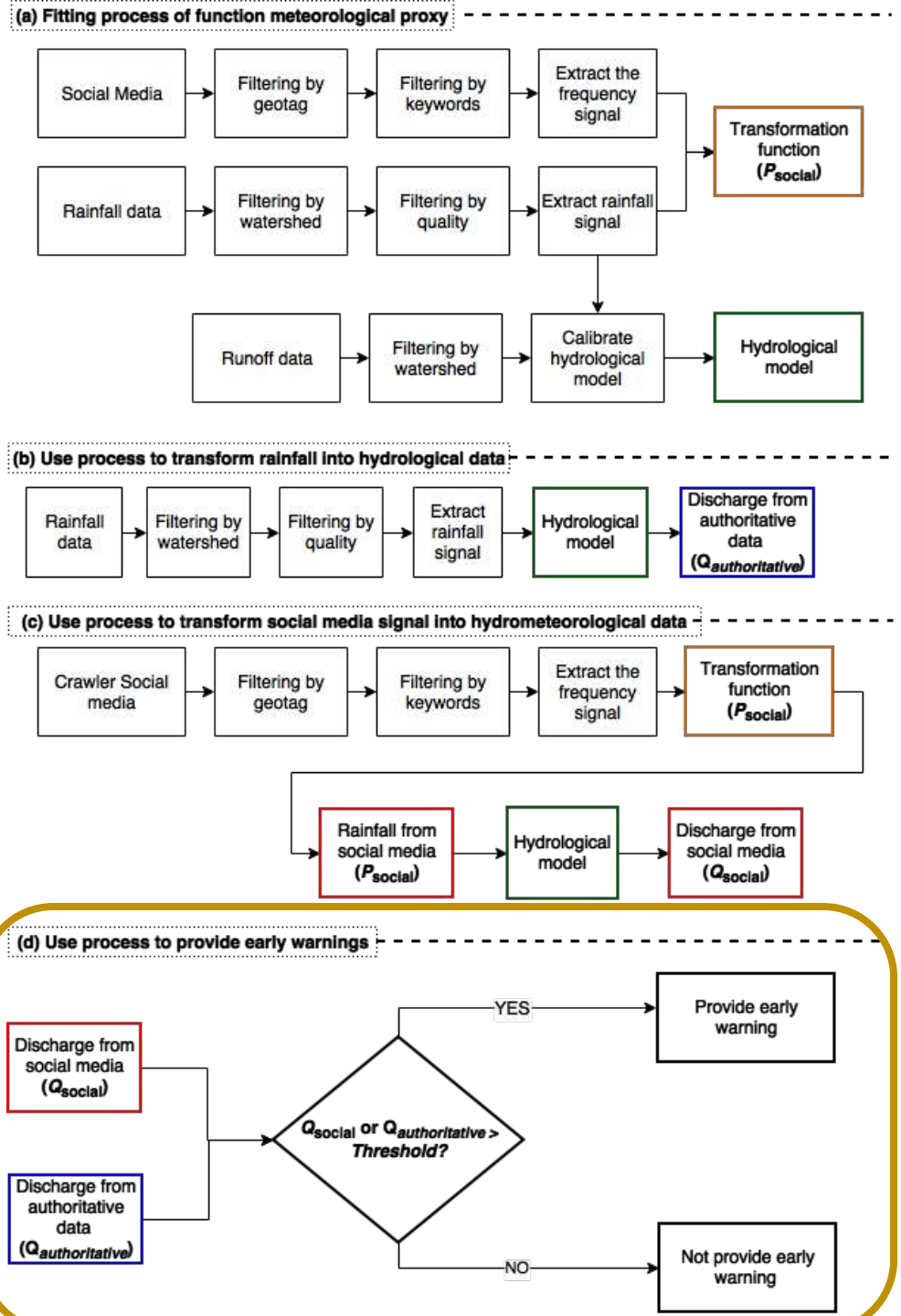
Results


a
Conventions:
Authoritative simulation
Social media simulation
Real value
Flow (m3/s)
0
200
400
600
2016-01-26
2016-01-28
2016-01-30
Date

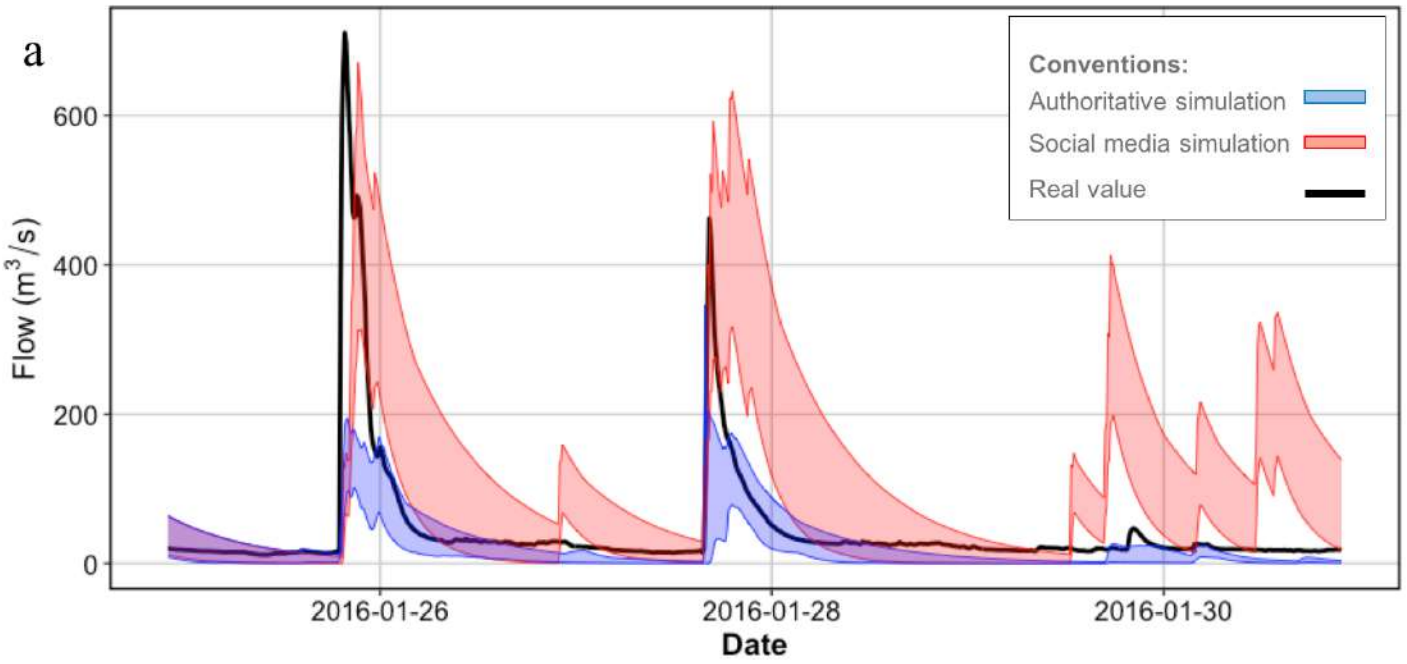

b
Conventions:
Authoritative simulation
Social media simulation
Real value
Flow (m3/s)
0
100
200
300
2016-12-10
2016-12-11
2016-12-12
2016-12-13
2016-12-14
Date

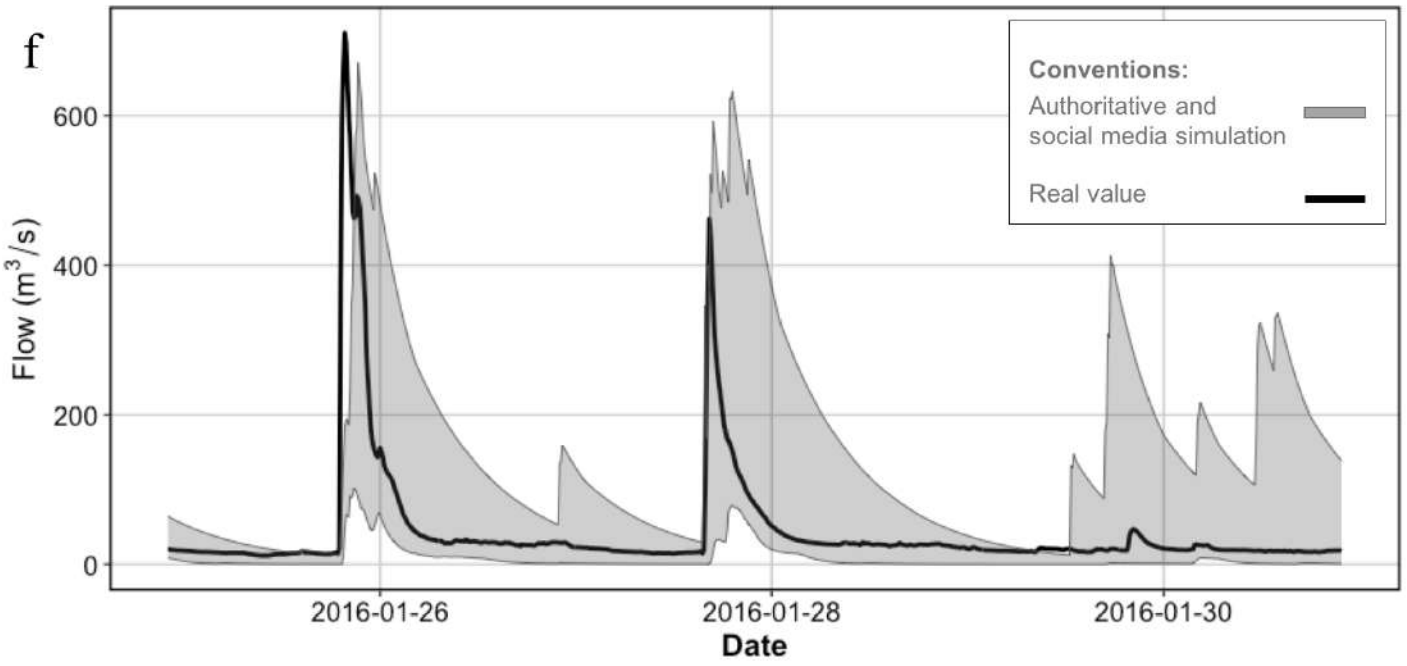

d
Conventions:
Authoritative simulation
Social media simulation
Real value
Flow (m3/s)
0
100
200
300
400
500
2017-02-02
2017-02-04
2017-02-06
2017-02-08
Date


e
Conventions:
Authoritative simulation
Social media simulation
Real value
Flow (m3/s)
0
200
400
600
2017-02-23
2017-02-25
2017-02-27
Date


(a) Fitting process of function meteorological proxy
Social Media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Runoff data
Filtering by watershed
Calibrate hydrological model
Hydrological model
(b) Use process to transform rainfall into hydrological data
Rainfall data
Filtering by watershed
Filtering by quality
Extract rainfall signal
Hydrological model
Discharge from authoritative data ($Q_{authoritative}$)
(c) Use process to transform social media signal into hydrometeorological data
Crawler Social media
Filtering by geotag
Filtering by keywords
Extract the frequency signal
Transformation function (P_{social})
Rainfall from social media (P_{social})
Hydrological model
Discharge from social media (Q_{social})
(d) Use process to provide early warnings
Discharge from social media (Q_{social})
Discharge from authoritative data ($Q_{authoritative}$)
Q_{social} or $Q_{authoritative}$ > Threshold?
YES
Provide early warning
NO
Not provide early warning

Results


a
Conventions:
Authoritative simulation
Social media simulation
Real value
Flow (m3/s)
600
400
200
0
2016-01-26
2016-01-28
2016-01-30
Date


f
Conventions:
Authoritative and
social media simulation
Real value
Flow (m3/s)
600
400
200
0
2016-01-26
2016-01-28
2016-01-30
Date

References

- Andrade, Sidgley Camargo de, Camilo Restrepo-Estrada, Alexandre C B Delbem, Eduardo Mario Mendiondo, and João Porto De Albuquerque. 2017. “Societal Geo-Innovation,” 19–37. doi:10.1007/978-3-319-56759-4.
- Restrepo-Estrada, Camilo; Andrade, Sidgley Camargo de; Albuquerque; João Porto de; Abe, Narumi; Fava, Maria Clara; and Mendiondo, Eduardo Mario. "Geo-social media as a proxy for meteorological data for flood monitoring". Computers & Geosciences. (Submitted 2017 - Ref: CAGEO_2017_513).
- BIG DATA | Nerdologia 40. https://www.youtube.com/watch?v=hEFFCKxYbKM
- Dilúvio de dados | Nerdologia 131. https://youtu.be/pNZq8BCS-Bc
- Privacidade na Internet | Nerdologia Tech 07 https://youtu.be/wZvDZfO13mA

Thanks for your attention

Camilo Restrepo
Email: camilo.restrepo@udea.edu.co

Acknowledgements:

Authors thank research funding grants from: CAPES-ProEx, CAPES-ProAlertas # 88887.091743/2014-01 (CEPED-"USP ALERTA").