Human Dynamics and Big Data 2016 Summer Specialist Meeting

Detecting location spoofing in social media: A Bayesian time geography approach













San Diego, August 1-2, 2016



Location

IncorporateApps

5. Fake GPS

New Horizon Apps

**** FREE

**** FREE





2. Location Spoofer

**** FREE

6. Mock GPS

**** FREE

3. Fake GPS location Lexa

4. Fake gps -

fake location

Andev *****

ŵ

**** FREE



7. Virtual Location maxxgreen

8. Location Spoofer - Fake App and Mobile ***** FREE **** FREE



Pro

FansBoy



10. Fake Your 9. Fake Location Location Free MJHDev Polli Apps **** FREE **** FREE

11. GPS Route 12. Location Simulator Spoofer Pro Fernando F. Galleg. LSDroid **** FREE **** \$1.67

Fake Location Apps (From Google Play Store)







GPS spoofers (From Google Shopping)

Location spoofing via a mobile GPS spoofer

Why Study Location Spoofing?



multiple web resources.

D ≁ Ξ Our Picks Popular Sections The Daily Dot Debug How to cheat at Pokémon Gotta catch 'em all, and you'll probably need to cheat to do it. Go and catch any Pokémon you want without leaving your couch

Find a Pokémon you want on Pokevision







Use PokeVision to get the coordinates of said Pokemon



Input your desired GPS location into the FakeLocation app



Google Play services com.google.android.gms

GPS Mocking (?)

The location information may be cached. If the location is not undated place wait a







20:00 to 21:59 EST

Methodology – Crawling and Harvesting Geospatial Big Data



Time Geography (Hägerstrand, 1970)

- Theoretical basis: Human constraints in space-time.
 - Capability constraints describe limitations on the activity of individuals since their physical/biological structure or available resources/vehicles.
 - Coupling constraints define the spatiotemporal limitation an individual has to be with other individuals or count on tools/materials to support producing, consuming and transacting.
 - Authority constraints point out the restricted space-time an individual (in)capable of accessing.
- Binary model: Space-time cube, prism, path, and ppa.
- Frequentist model: Time-geography Density Estimation (TGDE) (Downs et al, 2011)



A candidate and its consecutive geo-tagged tweet

A space-time cone (Hägerstrand, 1970)





Individual Activity Probability P(A)

$$P_{l}(A) = \frac{1}{(N-1)[t(1, N) \cdot v(u)]^{2}} \sum_{i=1}^{N-1} G\left(\frac{d(i, l) + d(l, i+1)}{t(i, j) \cdot v(u)}\right)$$

TGDE (Downs et al, 2011)

$$P_{l}(H) = \frac{1}{M[t(1, N) \cdot v(u)]^{2}} \sum_{j=1}^{M} G\left(\frac{d(j, l)}{rt(j, l) \cdot v(u)}\right)$$



Cill Bike Station 08:14 PM

Midtown Manhattan 03:04 PM

$P(H|A) = \begin{cases} P(H) & D > p \\ P(A) & 0 < D \le p_k \end{cases}$

Visiting probability

$$P(V) = P(A|H) = P(A) \cdot \frac{P(H|A)}{P(H)}$$

THE PARTY



Results

Fake geo-tags #

Source	# fake	# total	% Fake in total
Twitter for Android	1456	935859	0.16%
Instagram	1199	88292	1.36%
Foursquare	1140	70877	1.61%
Twitter for iPhone	990	1056878	0.09%
Total	4785	2151906	0.22%



Global distributions of fake geo-tagged tweets



Frequencies of posting geo-tagged tweets.





Fake geo-tagged tweets in New York city

Results - Nearest Neighbor

	Min	Max	Observed	Expected	Nearest	Z-	P-value
	Dist.	Dist.	Mean	Mean	Neighbor	score	
			Dist.	Dist.	Ratio		
Fake	6.19	3104680	84105	376413	0.223438	-44.57	0.000000
iPhone							
Fake	0.73	2131851	86534		0.229891	-44.20	
Android							
Fake	1.80	2176028	90139		0.239469	-43.65	
Instagram							
Fake	8.13	4679496	63738		0.169330	-47.67	
Foursquare							
Fake	1.60	2577011	99499		0.264334	-42.22	
Cleaned	8.13	2410705	90495		0.240415	-43.59	
Original	7.74	4462020	96788		0.286515	-42.63	

Nearest neighbor index (NNI) analysis results



Distributions of Z-score

(from ArcGIS online)





Cumulative Frequency of nearest neighbor distances

Cumulative Frequency of different sources



For Fun





For recording visited places

Discussion –

Identified Location Spoofers



For attracting audiences from elsewhere

For marketing

Discussion - Location Spoofers (Social Media Providers)



A subtle business advertising strategy (Foursquare)

Location customization (Instagram)

Discussion - Location Spoofers (Government/Military Agencies)



Controversial geo-tagged tweets sent by Taliban spokesman



Cuban Twitter "ZunZuneo", endorsed by USAID, was designed to stir remote unrest by organizing "smart mobs" (source: Google search)

Conclusions - Summary

Definition

An intentional act of masquerading one's location as somewhere else other than its true location

An information transmission process with several constituent elements

Spoofing motivations and locational inconsistency

Approach

A time geographic approach

(Ab)used by different types of location spoofers (e.g., social media providers) for multiple purposes

Implication

Multi-faceted impacts

Raising the awareness

Future Research

• Integrating with other detecting approaches





- Detecting Spoofing with cluster computing
- Further dealing with the multi-faceted nature of location spoofing

Human Dynamics and Big Data 2016 Summer Specialist Meeting

Thanks for your attention!

San Diego, August 1-2, 2016