

Geo-Social Targeting for Privacy-friendly Mobile Advertising

David Martens – University of Antwerp

Foster Provost – New York University, Coriolis Labs

Alan Murray – Coriolis Labs

In collaboration with
EveryScreen Media

Outline

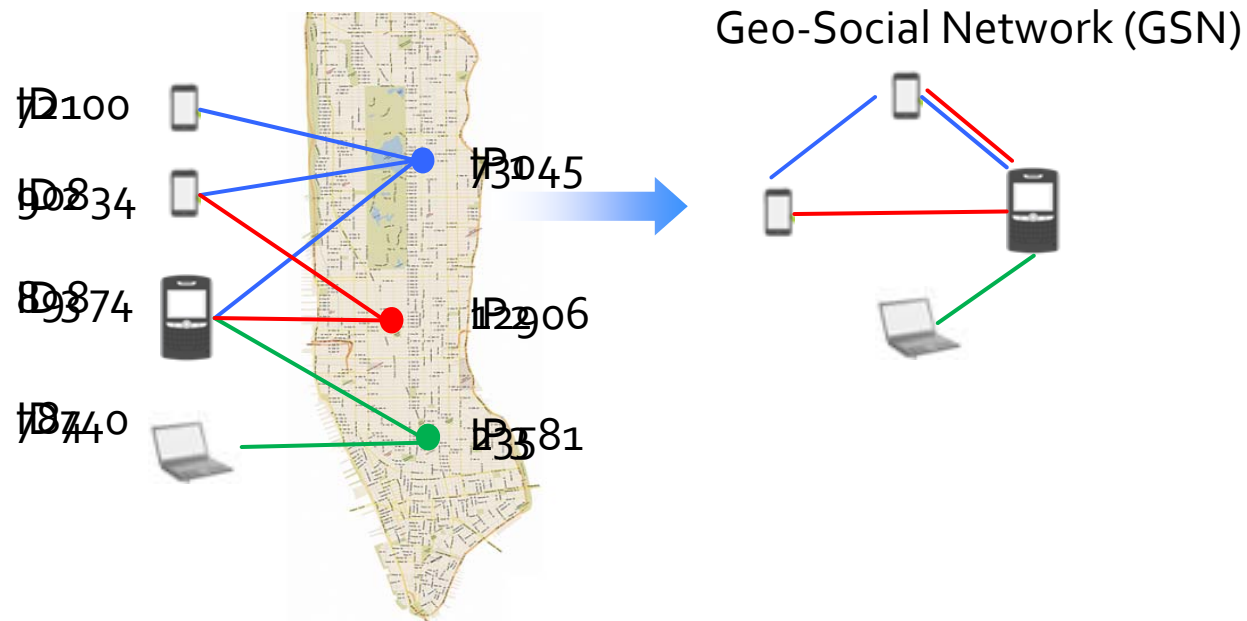
1. **Geo-social network design**
2. Empirical results
 - RTB data
 - Identifying the same user on multiple devices
 - Similarity in browsing behavior
3. Next steps

Geo-Social Network

- Combining advantages of **social** and **local** in **mobile** setting
 - Social: very effective
 - Local: hyper-local targeting
 - Mobile: huge growth
- Need for effective and privacy-friendly design

Geo-Social Network

- Based on similarity in location visits
 - Link **strength**: shared locations, location popularity
 - **Double de-identified**: privacy by design



Link strength metrics

- Link strength metrics
 - boolean
 - count
 - cosine
 - Jaccard
 - weighted by inverse location frequency
- Work in progress
 - Supervised location scores
 - Work on (huge) device by location matrix

GSN for targeting

- Why GSN for targeting
 - Intuitively, more shared locations likely same person
 - Fine-grained data contains latent information
 - Influence on purchase decision depends more on co-presence than on who we consider to be close friends (Crandall et al., 2011)

Outline

1. Geo-social network design
2. **Empirical results**
 - RTB data
 - Identifying the same user on multiple devices
 - Similarity in browsing behavior
3. Next steps

Data properties

- Using mobile checkins from RTB environments
 - “The **next revolution in advertising** for the benefits it will provide for all players in the ecosystem – the publisher, the user and the advertiser” – Pubmatic (2010)
- Data properties (sample!)
 - 1 week of RTB data
 - 24 websites
 - 100 million records/day

Identifying the same user

- Finding user “on every screen”

- To what extent are we able to find the same user within the GSN?

A. **Randomly** split up transactions

B. **Temporal** split up of transactions

- Before and after middle of the week

C. **TemporalDynamic** split up of transactions

- Before and after middle of all days the user is seen

D. Different users with **email** in url

E. Choosing the **personal emails** only

- e.g. not breakingnews@cnn.com

	IP1	IP2	IP3
John	1	5	0
John1	0	3	0
John2	1	2	0

Identifying the same user

- To what extent is the same user *connected* to itself?

	Percentage X1 and X2 connected
Random	81%
Temporal	43%
TemporalDynamic	51%
Email	38%
PersonalEmail	52%

- To what extent is the same user *strongly connected* to itself?

	AUC	Rank
Random	64%	39%
Temporal	66%	43%
TemporalDynamic	64%	43%
Email	58%	43%
PersonalEmail	58%	44%

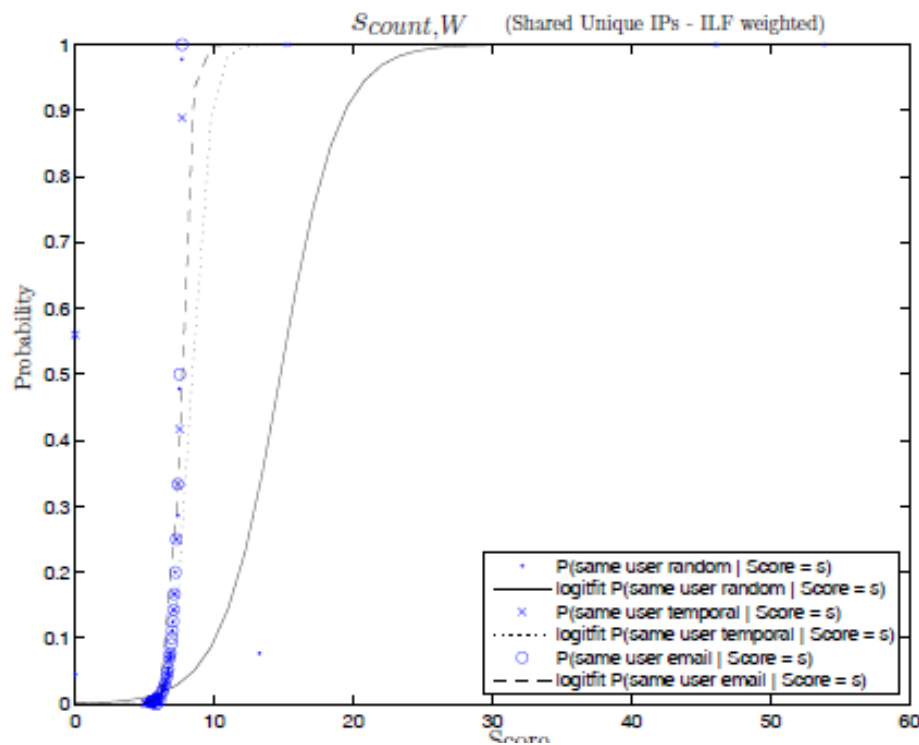
NB: Rank among neighbors is good as long it remains below 50%

E.g. if 2 neighbors and ranked first: rank = 50%

E.g. if 10 neighbors but all have same score, average taken: rank = 50%

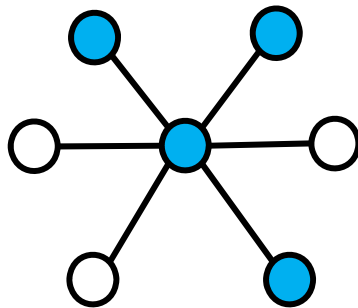
Identifying the same user

- Towards probability model of being same user



Do GSN neighbor visit similar websites?

- Website visit as proxy for brand affinity
 - Device that visited website X (e.g. cnn.com)
 - Build neighborhood
 - 50% of neighbors of brand actor also visit website X
 - 10% of all users visit the website X
- ➔ Lift of 5 (50/10) - Leverage of 40% (50% - 10%)



Do GSN neighbor visit similar websites?

- *Sample of 100 neighborhoods around known "seed" visitors*
 - Percentage of users visiting website
 - D: in entire population (baseline)
 - N(BA): in neighborhood of seed visitors
 - Lift^{10N}: only considering top 10 "closest" neighbors

	D	N(BA)	Lev.	Lift	Lift ^{10N}
Local interest websites					
A	32%	68%	36%	2.1	2.1
B	20%	33%	13%	1.7	2.5
C	13%	49%	36%	3.7	4.1
D	12%	35%	23%	2.9	4.2
E	11%	57%	46%	5.0	6.3
F	1%	21%	20%	14.8	30.3
Non-local websites					
G	3%	28%	25%	9.9	13.7
H	2%	20%	18%	12.3	27.1
I	1%	6%	5%	11.0	50.4
Social network websites					
J	7%	77%	70%	11.0	13.4
K	3%	21%	18%	6.3	10.6
L	1%	18%	17%	27.5	84.0



Do GSN neighbor visit similar websites?

- GSN Neighbors are similar
- GSN Neighbors *with high score* even more similar
 - Reasons for good performance:
 - **Soul mates**: similar users visiting the same IPs
 - **Many screens**: same user on different devices
 - **Multiple identities**: same user on same device with different IDs

Outline

1. Geo-social network design
2. Empirical results
 - RTB data
 - Identifying the same user on multiple devices
 - Similarity in browsing behavior
3. **Next steps**

GSN Use

- Compute GSN similarities for
 - Targeting
 - Entity resolution
 - Cross-channel campaign evaluation
 - Hyper-local targeting

Challenges and Next Steps

- Next steps
 - Validate results on full data
 - True brand action data
 - Cross-channel evaluation
- Challenges
 - Big data
 - Modeling influence of neighbors of high degree

Conclusion

- Geo-Social Network
 - Allows to find users “**on every screen**”
 - Neighbors are **similar** in browsing behavior
 - Can be built in **privacy-friendly** manner
 - Uses beyond targeting

Contact details

- Prof. dr. ir. David Martens
Faculty of Applied Economics
University of Antwerp
E: david.martens@ua.ac.be
W: www.ua.ac.be/david.martens
- www.everyscreenmedia.com

