

Spatio-Temporal Dynamics of Online Memes: A Study of Geo-Tagged Tweets



Today we will talk about

- Global footprint of hashtags
- Three spatial properties of hashtag propagation:
focus, entropy, spread
- Two methods for Characterizing locations based on hashtag spatial analytics



Global footprint of hashtags: Tweets Distribution



Input Data:

- List of tuples: <hashtag, timestamp, latitude, longitude>
- Geo-tagged tweets only

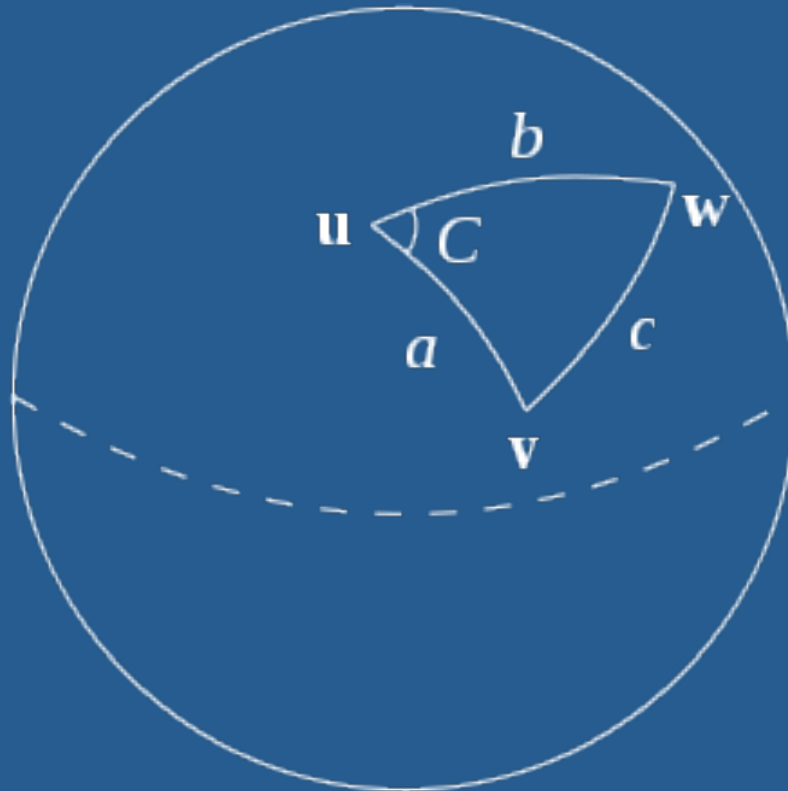


Global footprint of hashtags:

- Hashtag Sharing vs *Distance*
- Hashtag adoption Lag vs *Distance*

Haversine *Distance*

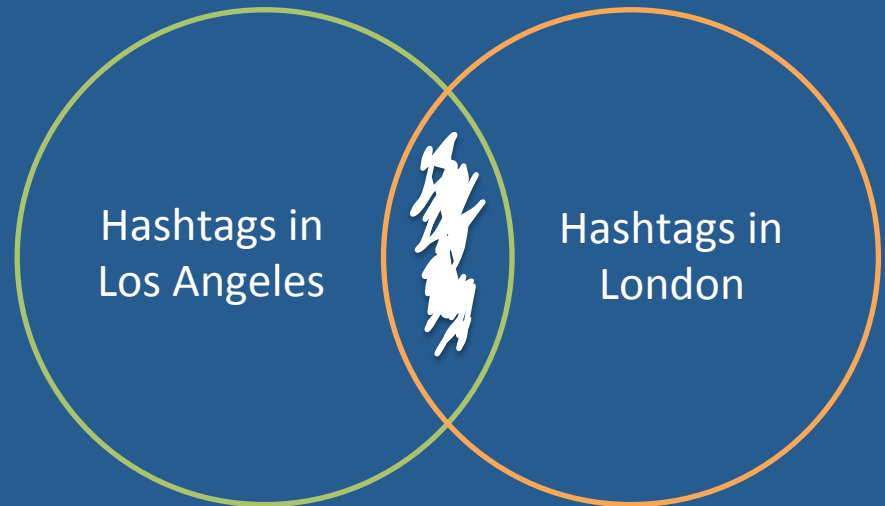
$D: \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$



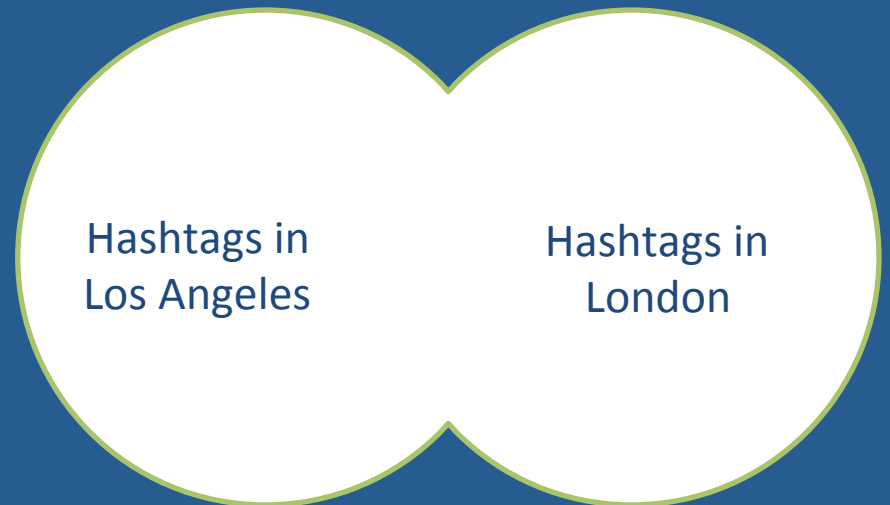
Hashtag Sharing vs *Distance*

To what degree does distance impact whether a hashtag is shared between two locations ?

- Jaccard coefficient



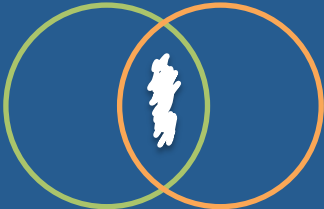
HashtagSimilarity(Los Angeles, London) =



Hashtag Adoption Lag vs *Distance*

Locations that are near, are more likely to share hashtags;

Are they more more likely to adopt hashtags at the same time?

$$\text{AdoptionLag}(\text{Los Angeles}, \text{London}) = \frac{1}{\text{Venn Diagram}} \times$$


$$\left(\begin{array}{l} \text{For each hashtag that belongs to} \\ \text{Venn Diagram} : \\ | \text{time hashtag } h \text{ was observed in LA} - \text{time hashtag } h \text{ was observed in London} | \end{array} \right)$$




Three spatial properties of hashtag propagation: *focus, entropy, spread*

Probability of observing hashtag # in London = $\frac{\text{All occurrences of \# in London}}{\text{For each } \underline{\text{location}} \text{ in Locations: } \{\text{occurrences of \# in } \underline{\text{location}}\}}$

Focus

Focus for hashtag # = Maximum (Probabilities of observing hashtag #)

- At a single location (e.g. London)



Three spatial properties of hashtag propagation: *focus, entropy, spread*

Entropy

- measures randomness in spatial distribution of hashtag
- determines minimum number of **bits** required to represent the spread
- entropy of zero for a hashtag indicates that it was posted from one 2^0 location only

Entropy for hashtag # = $-\left(\text{for each location in Locations :}\right.$
 $\left.\text{Probability of observing hashtag \# in location} \times \log_2(\text{Probability of observing hashtag \# in location})\right)$

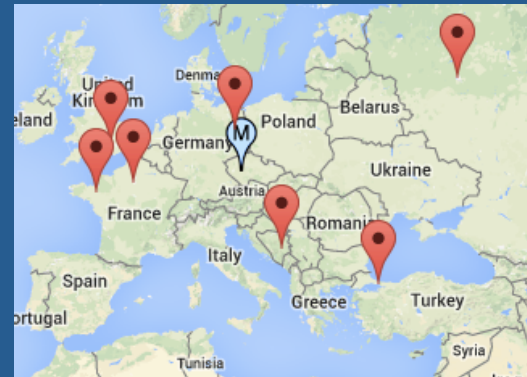


Three spatial properties of hashtag propagation: *focus, entropy, spread*

Spread

- mean distance for all occurrences of a hashtag from its Geographical midpoint

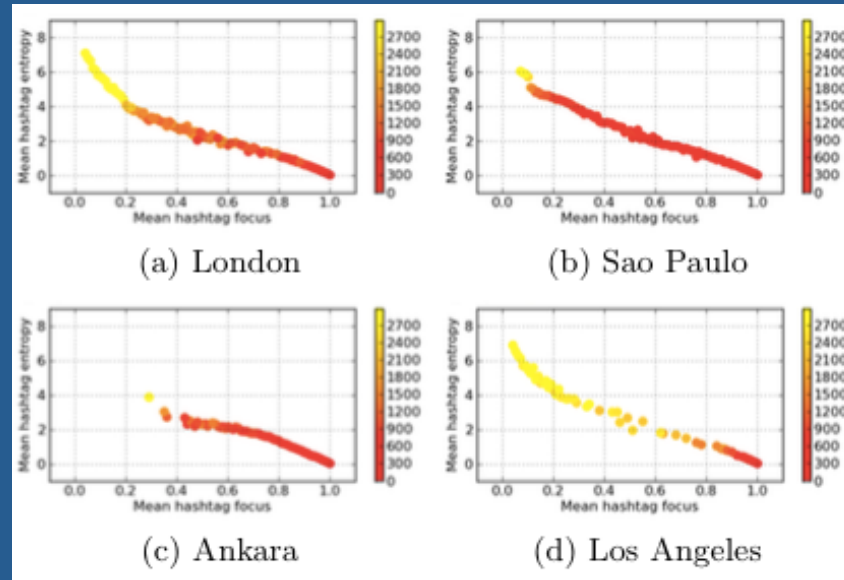
$$\text{Spread of hashtag \#} = \frac{1}{|\text{occurrences of \#}|} \times \left(\begin{array}{l} \text{For each occurrence of \# :} \\ \text{meanDistance(occurrence, Geo. Midpoint(all \# occurrences))} \end{array} \right)$$



Two methods for Characterizing locations based on hashtag spatial analytics

1. Location-based entropy-focus-spread plots

- For each location find *focused* hashtags;
 - For each hashtag focused on location L: Plot entropy vs focus;
 - For every focus-entropy pair: Calculate mean spread;

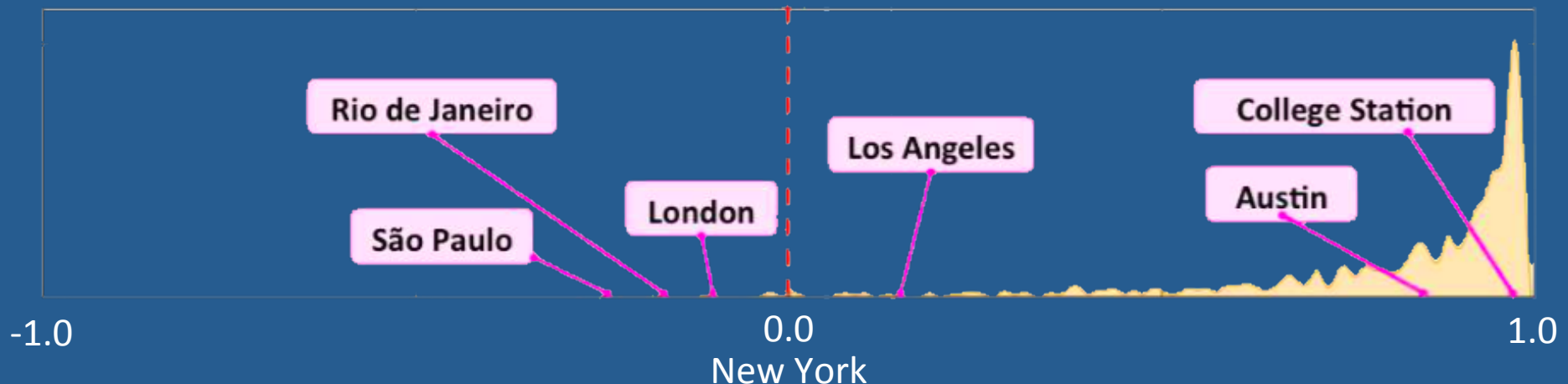


Two methods for Characterizing locations based on hashtag spatial analytics

2. Method for evaluating the spatial impact of locations

- Evaluates the impact a location has on other locations by measuring the hashtag-based spatial impact

$$\mathcal{I}_{l_i \rightarrow l_j}^h = \begin{cases} \frac{|O_{l_i}^h \prec O_{l_j}^h| - |O_{l_i}^h \succ O_{l_j}^h|}{|O_{l_i}^h \times O_{l_j}^h|} & \text{if } h \in H_{l_i} \text{ and } h \in H_{l_j} \\ 1 & \text{if } h \in H_{l_i} \text{ only} \\ -1 & \text{if } h \in H_{l_j} \text{ only} \end{cases}$$



What we are going to implement in our project

- Global footprint of hashtags
- Three spatial properties of hashtag propagation:
focus, entropy, spread
- Two methods for Characterizing locations based on hashtag spatial analytics

