

# Distance Matters: Geo-social Metrics for Online Social Networks

---

**Salvatore Scellato**

Computer Laboratory, University of Cambridge

Joint work with: Cecilia Mascolo, Mirco Musolesi, Vito Latora

**3rd Workshop on Online Social Networks**

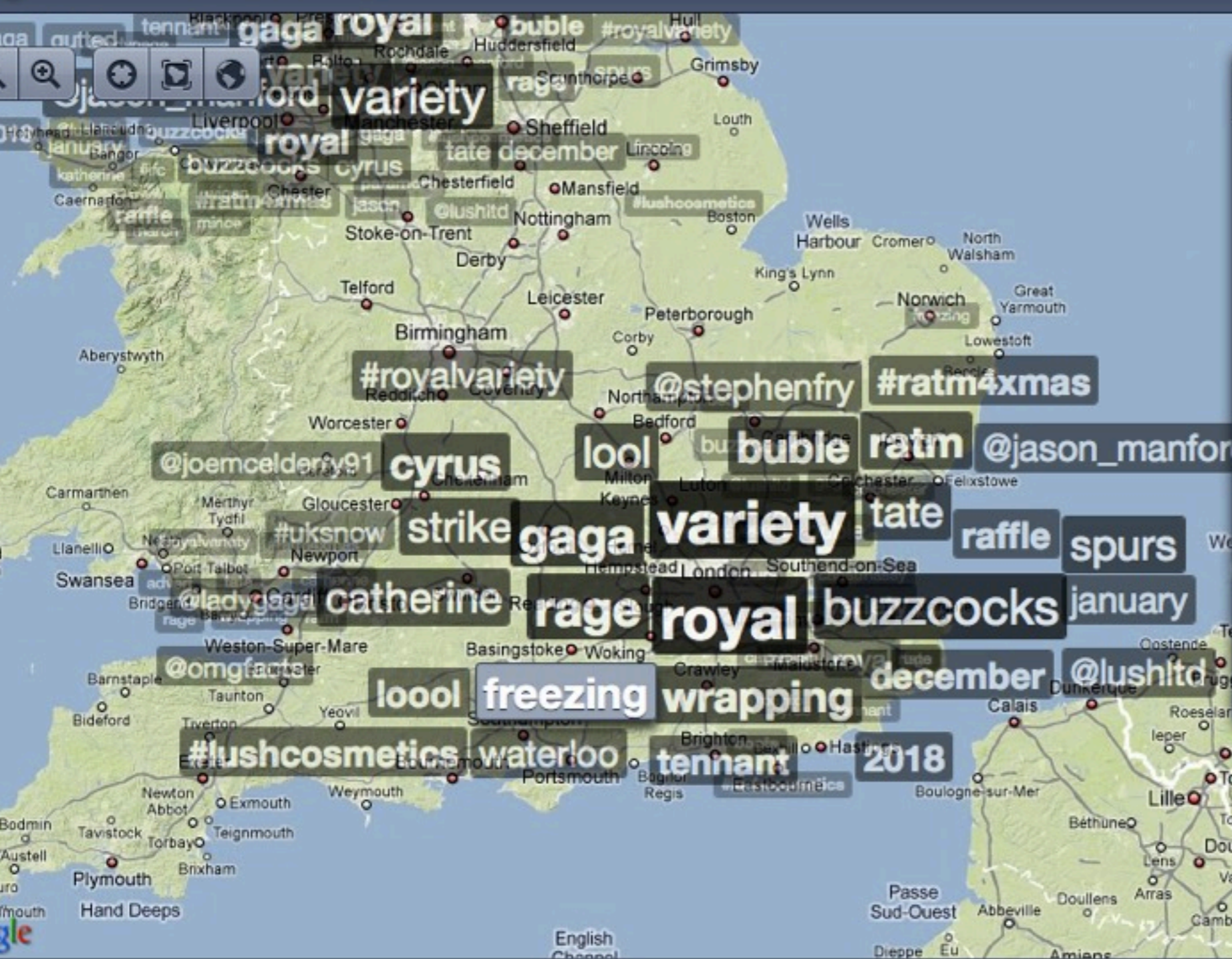
Boston, 22 June 2010





Location, location, location.  
And social networks.

Plethora of new services:  
increasingly important,  
excitingly new.



freezing in London

Forget the Big Chill, get freezing!

Status | Facebook

West Ham's owners benefit from freeze on debt...

Why doesn't my radiator convect any fucking heat round my room? I'm bloody freezing!  
London | 24 minutes ago

@eatlikeagirl it's supposed to be another freezing cold day tomorrow so a spicy tortilla soup oughtta warm ya up!  
London | 27 minutes ago

my @TheSouthButt was freezing after having to go on all those kiddies rides at gullivers!!  
London | 27 minutes ago

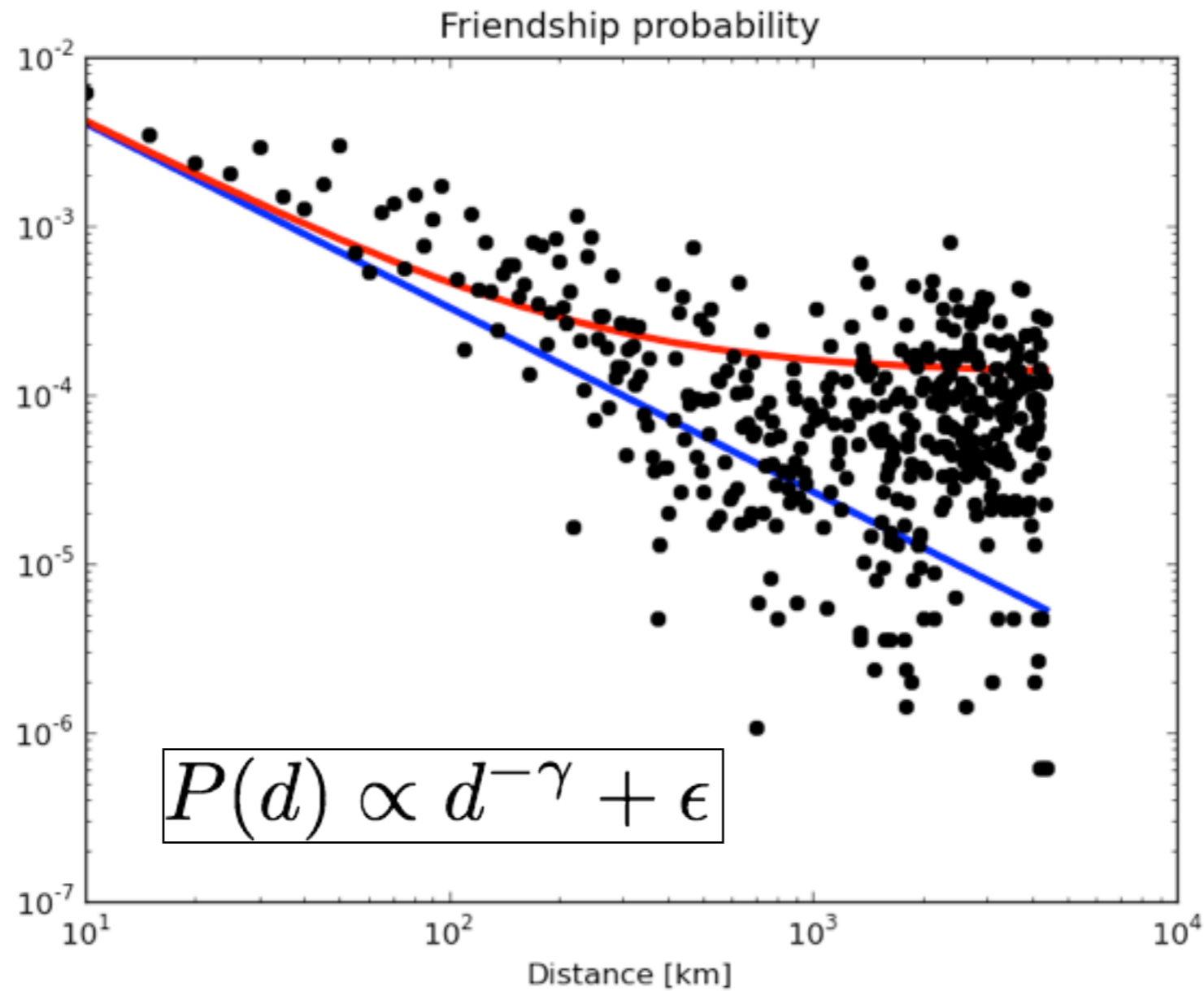
Information, social structure and space.

Geography may shape social structures and affect information flows.



Put people on a map and  
social ties across space.

We need new tools to model  
these networks.



Distance matters.

Probability of friendship decreases with distance.

# Interesting questions...

---

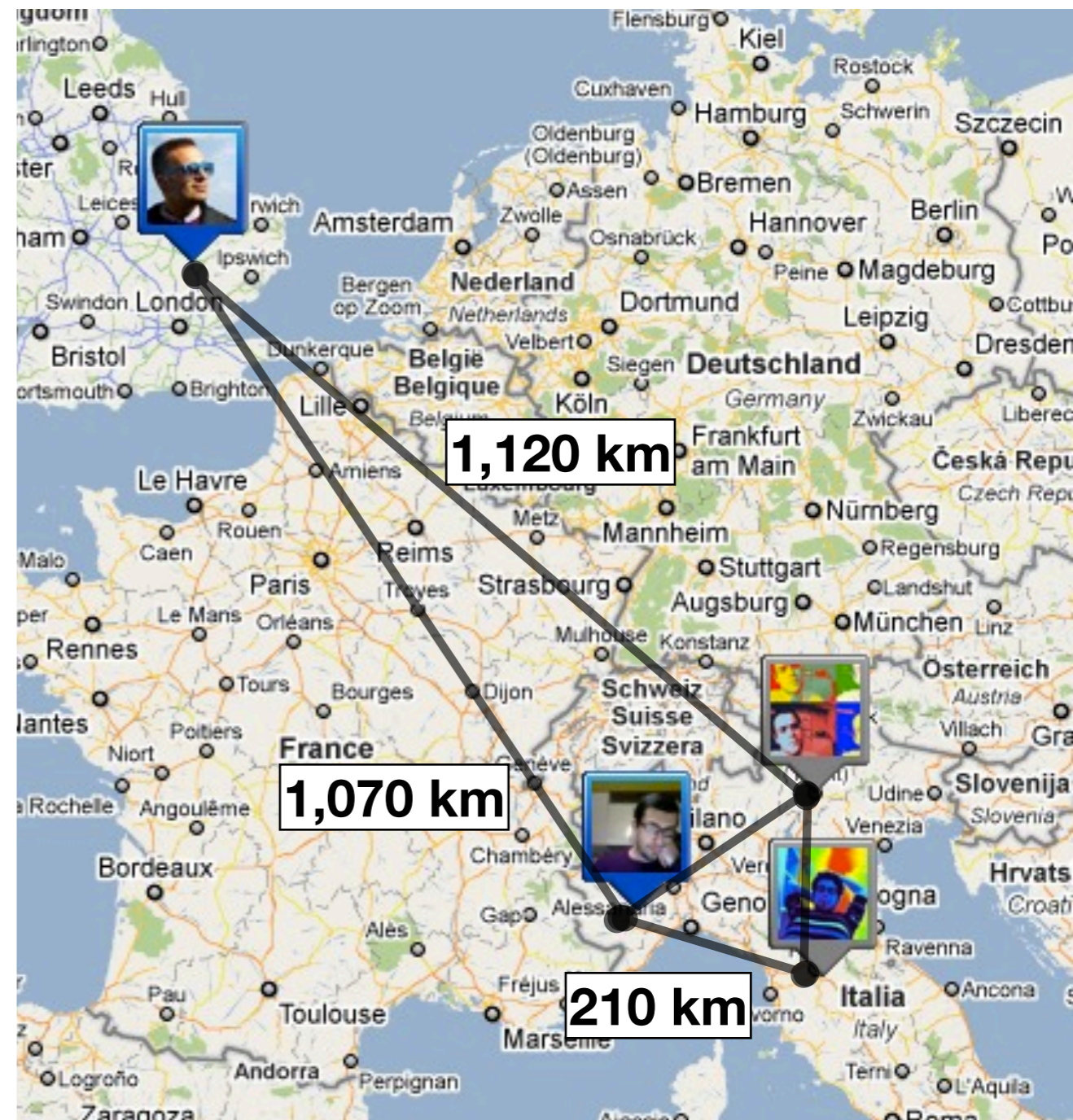
- Can we **discriminate between users** according to their attitude towards long-range ties?
- How geographically close are **clusters of friends**?
- How is **information spreading** across space over social links?
- Can we **improve real systems** exploiting geographic information in social networks?



# Geographic Social Network

Given a graph  $G=(N,K)$  and the **geographic location** of the nodes:

- Place all nodes in a 2D **metric space** adopting great-circle distance on the Earth.
- Assign a **weight** to each edge equal to the **geographic distance** between the two nodes.



# Geo-social metrics

How close are the neighbors of a given node to the node itself?

Node locality

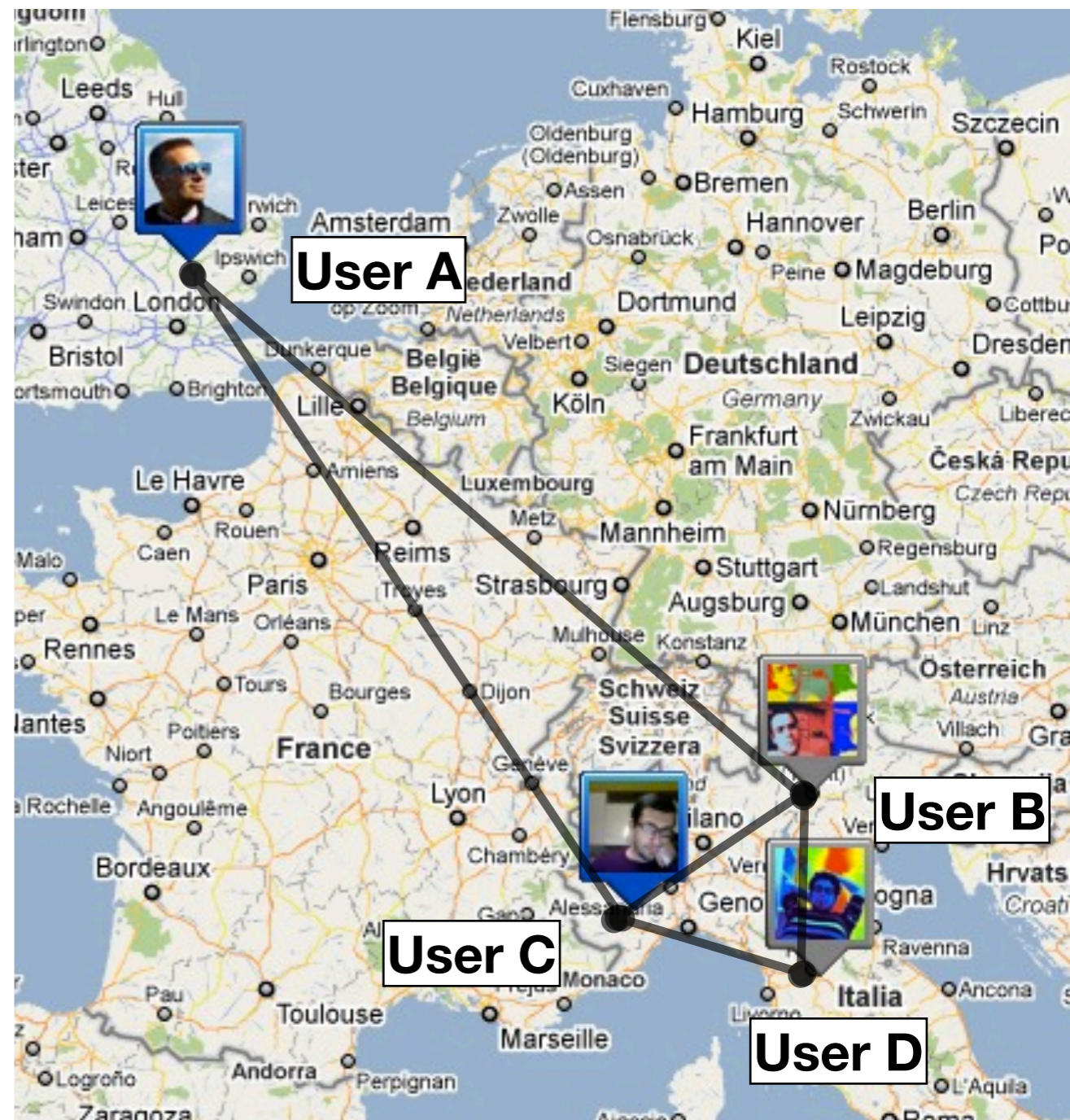
$$NL_i = \frac{1}{k_i} \sum_{j \in \Gamma_i} e^{-l_{ij}/\beta}$$

How spatially inter-connected are the neighbors of a given node?

Geographic clustering coefficient

$$\Delta_{ijk} = \max(l_{ij}, l_{ik}, l_{kj})$$

$$GC_i = \frac{1}{k_i(k_i - 1)} \sum_{j,k \in \Gamma_i} e^{-\Delta_{ijk}/\beta}$$





# Node locality

---

How close are the neighbors of a given node to the node itself?

Our aim is to:

- Highlight only extremely **short-range** social connections.
- Normalize this measure for nodes with **various degrees**.
- Allow networks at **different geographic scales** to be compared.

$$NL_i = \frac{1}{k_i} \sum_{j \in \Gamma_i} e^{-\frac{l_{ij}}{\beta}}$$

Node degree

Node neighborhood

Link length

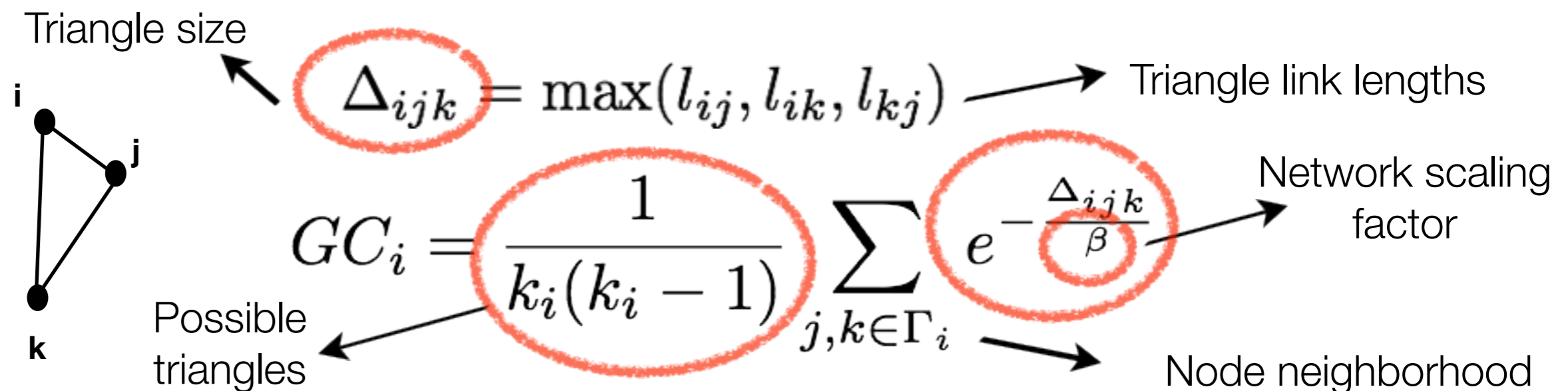
Network scaling factor

# Geographic clustering coefficient

How spatially inter-connected are the node's neighbours?

Our aim is to:

- Generalise the standard **clustering coefficient**.
- Highlight only extremely **short-range** social triangles.
- Allow networks at **different geographic scales** to be compared.

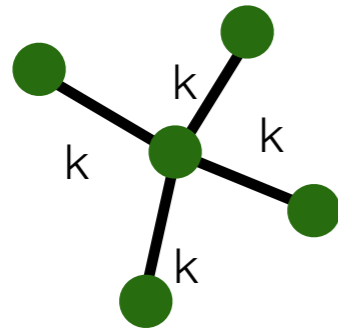


# Scaling factor

---

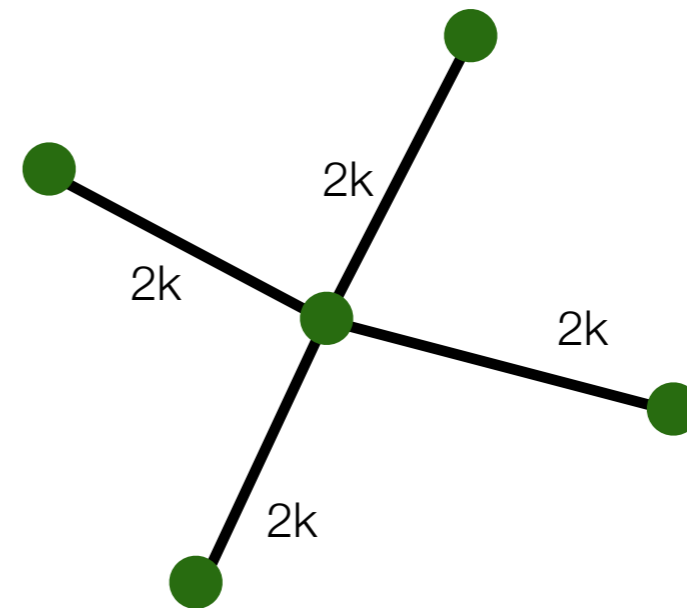
The scaling factor  $\beta$  allows us to compare geo-social metrics across networks with different scales. For example, by choosing  $\beta$  so that **if all lengths are rescaled,  $\beta$  is also rescaled**, geo-social metrics are not affected.

Graph 1



$$e^{-\frac{k}{\beta_1}}$$





Graph 2



$$e^{-\frac{2k}{\beta_2}} = e^{-\frac{2k}{2\beta_1}} = e^{-\frac{k}{\beta_1}}$$

# Dataset collection

---

Online Social Network	Collection method	Sampling	Location information
	Public API	Complete	GPS
	Public API	Snowball crawling	GPS
	Public API + HTML scraping	Snowball crawling	Text-based
	Public API	Snowball crawling	GPS or text-based

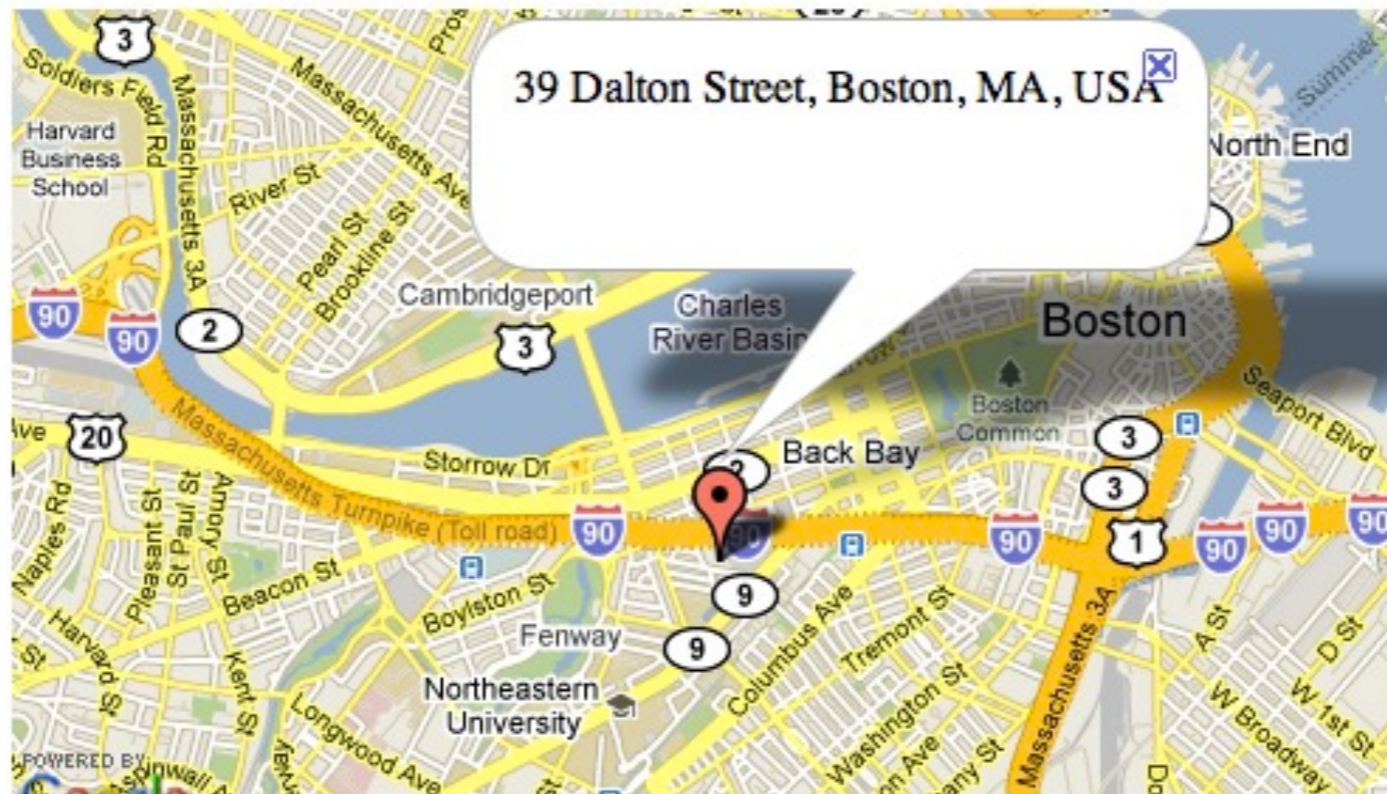
# Yahoo Geocoding API



Placemaker

39 Dalton Street, Boston, MA, USA

Go!



# Problems with geocoding

---



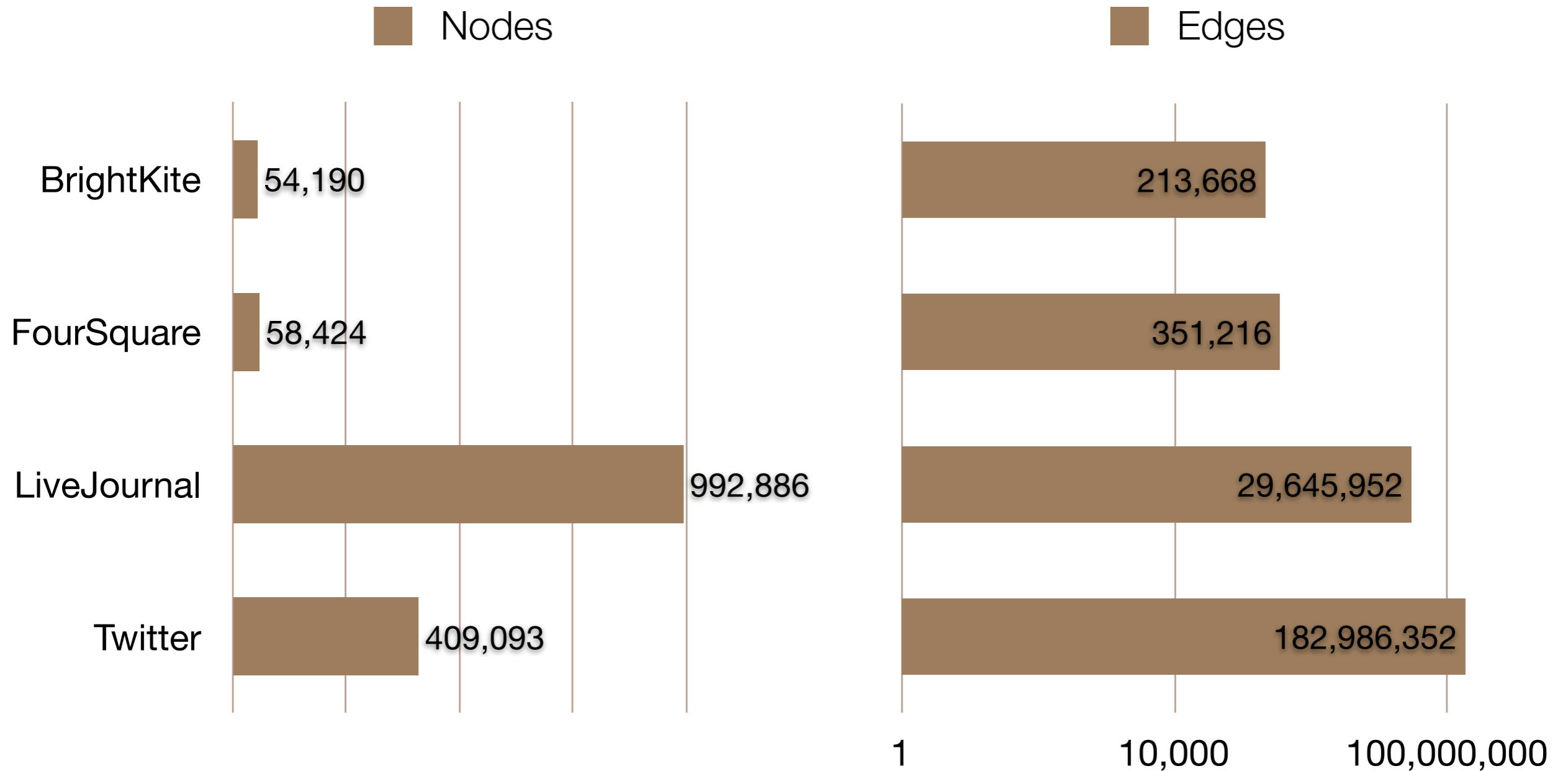
Hilton Paris



Paris Hilton

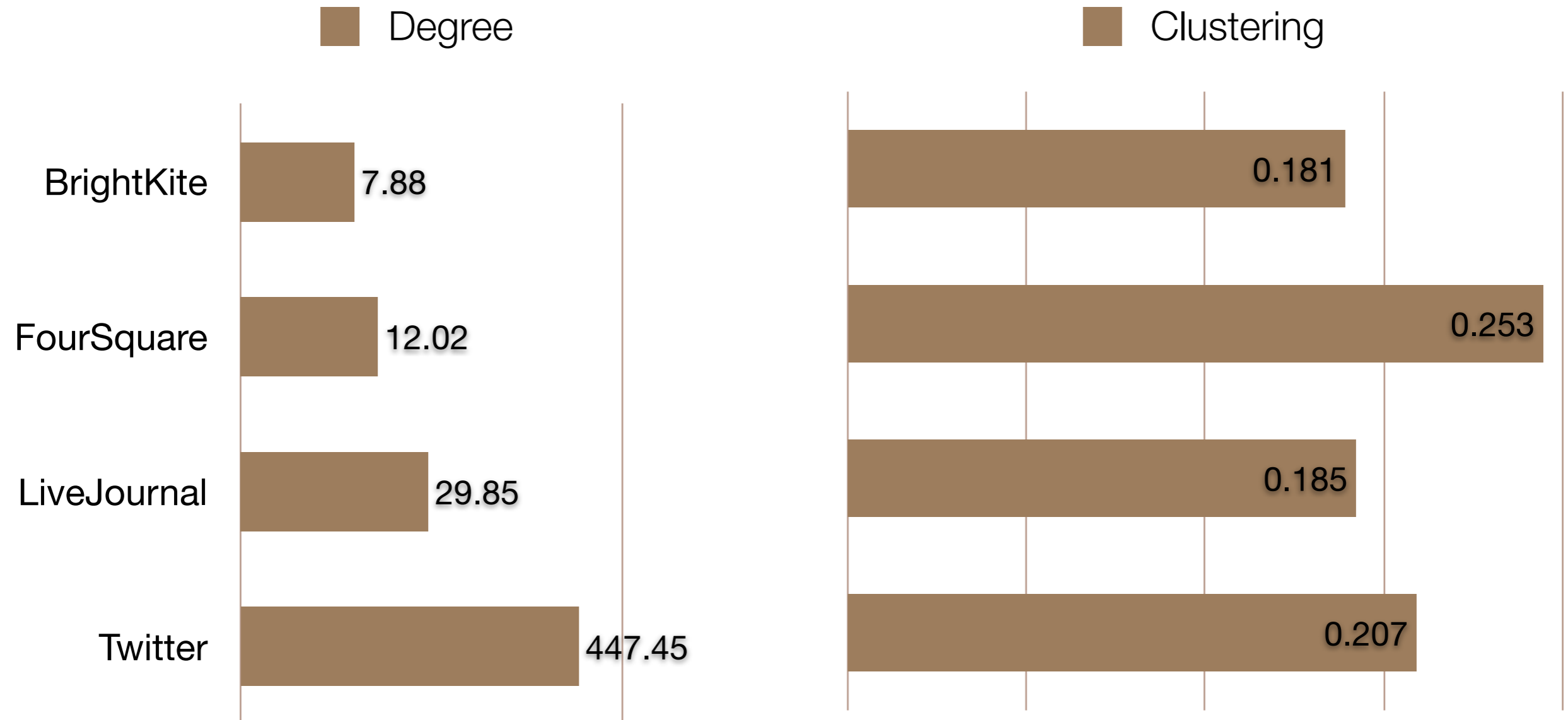
Keep only city-level accurate results

# Dataset properties



# Social Metrics

---

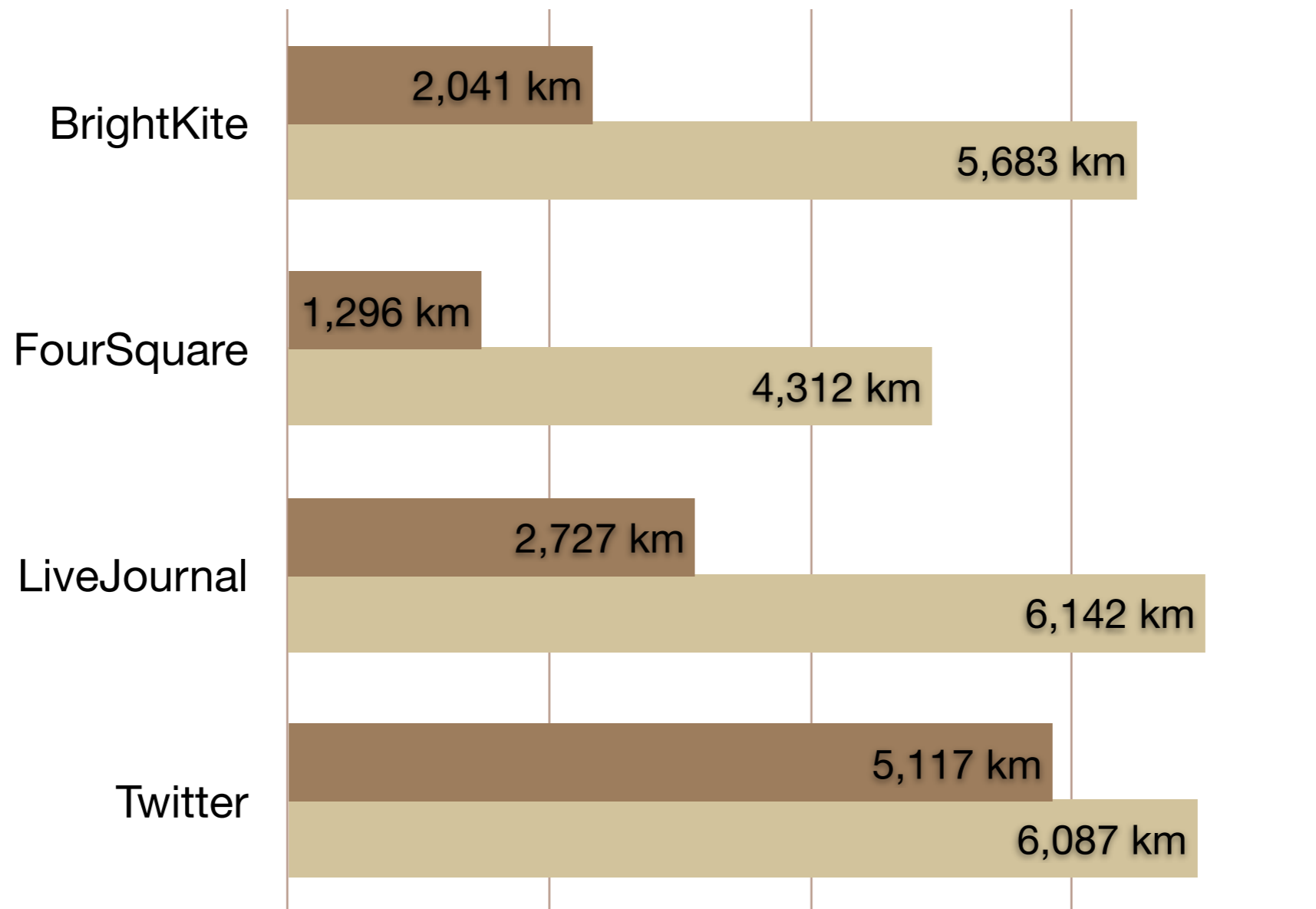




# Geographic Properties

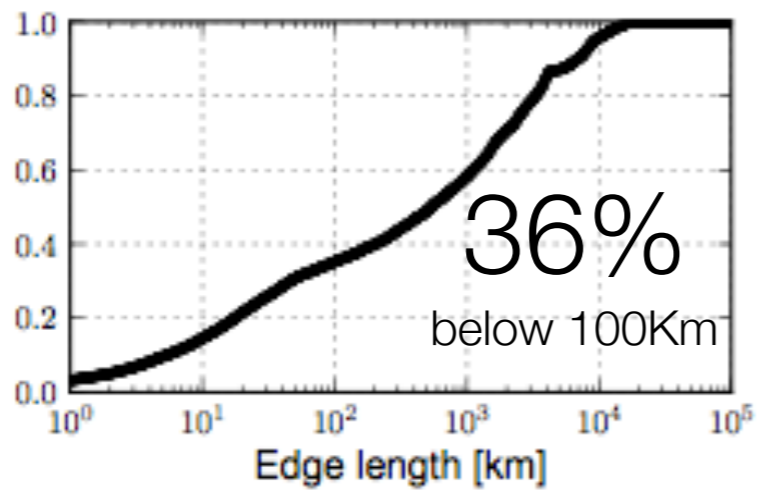
---

- Average link length
- Average user distance

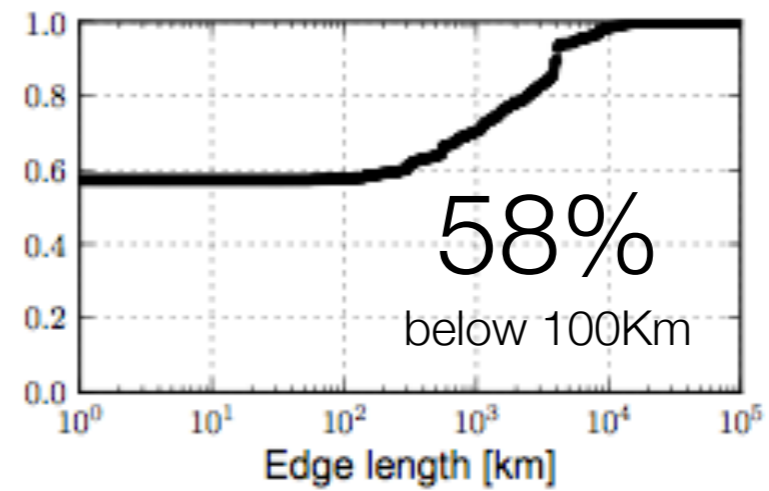


# Social Link Geographic Distance

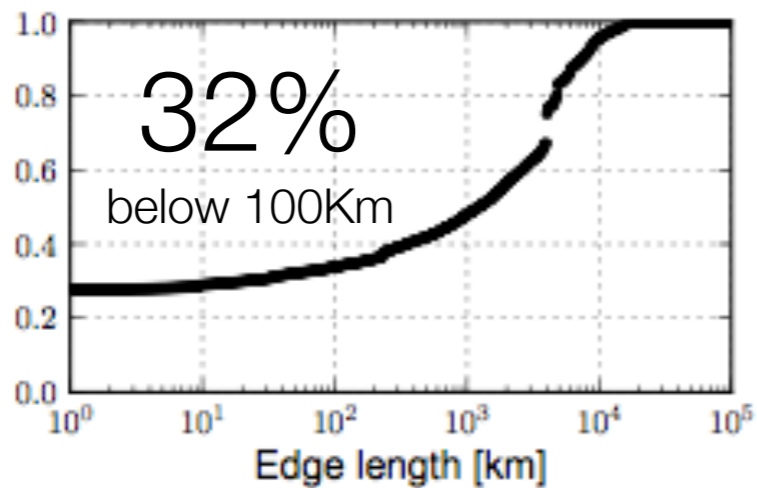
## BrightKite



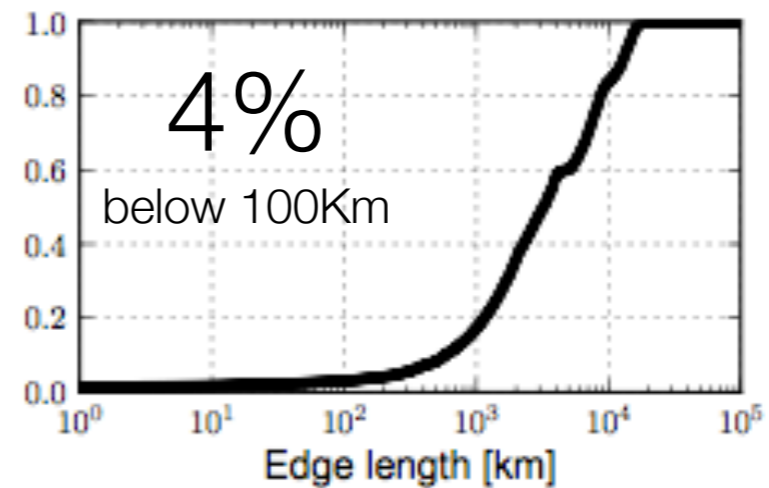
## FourSquare



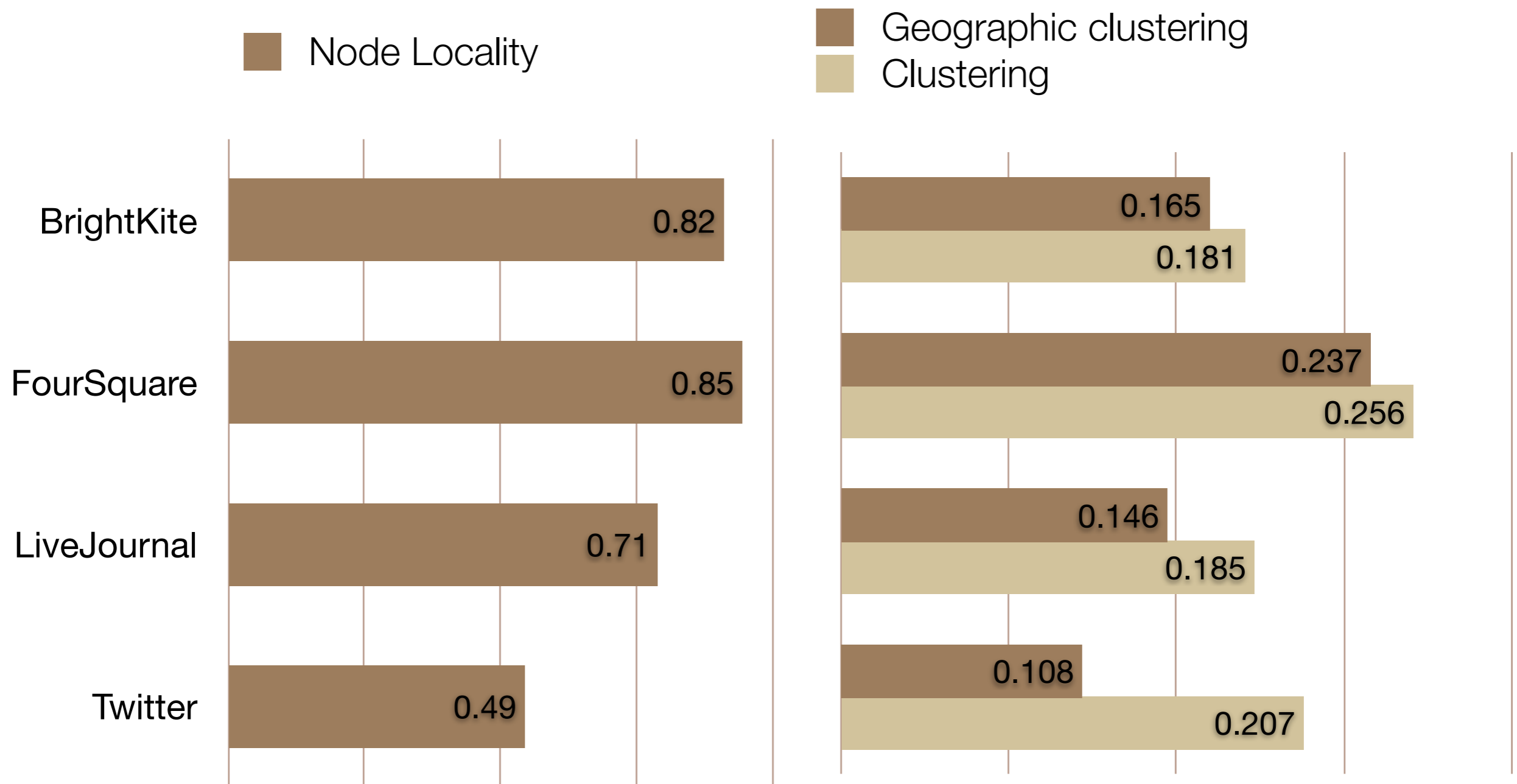
## LiveJournal



## Twitter

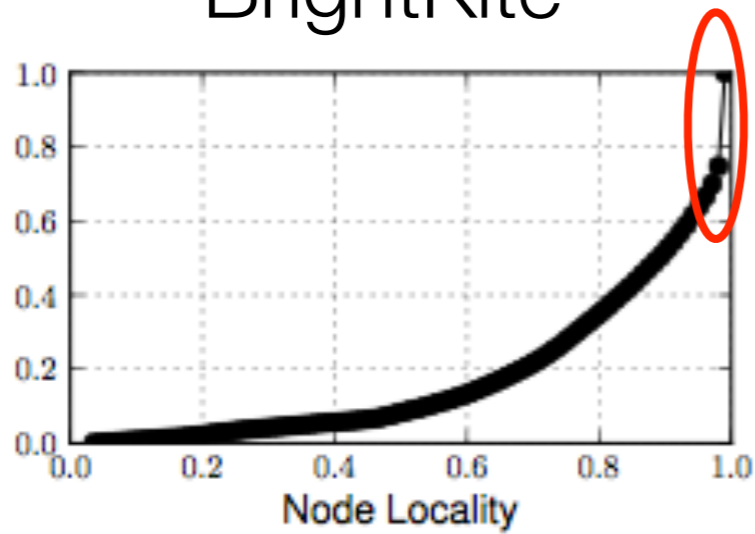


# Geo-social Metrics

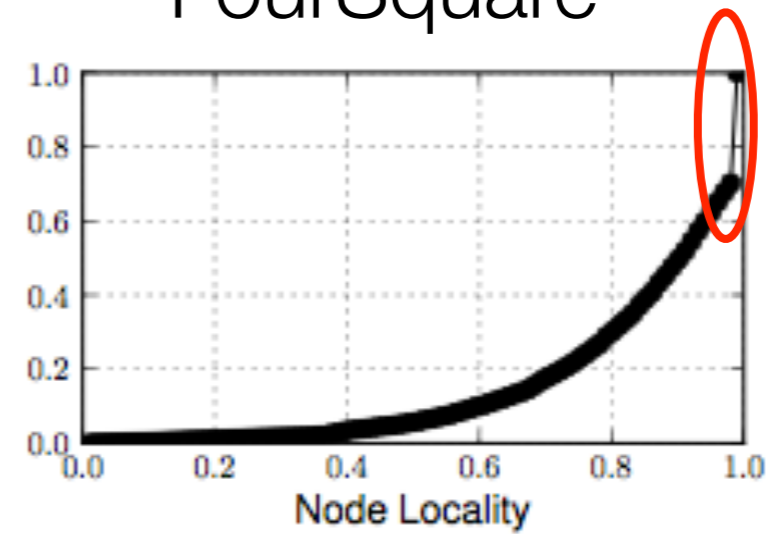


# Node Locality Distributions

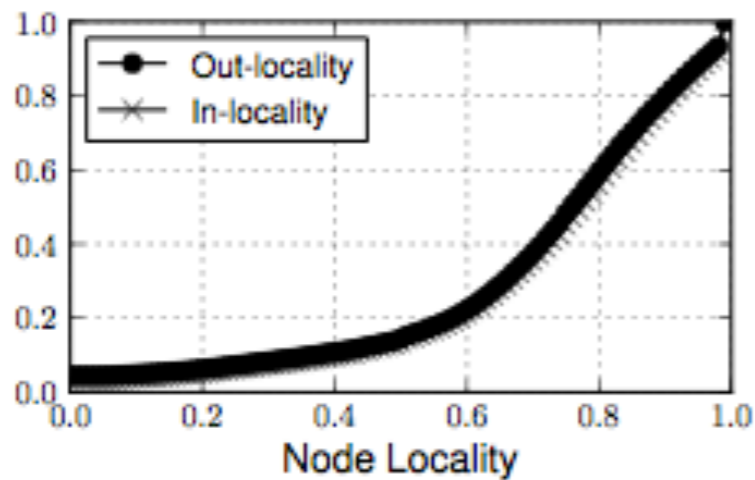
BrightKite



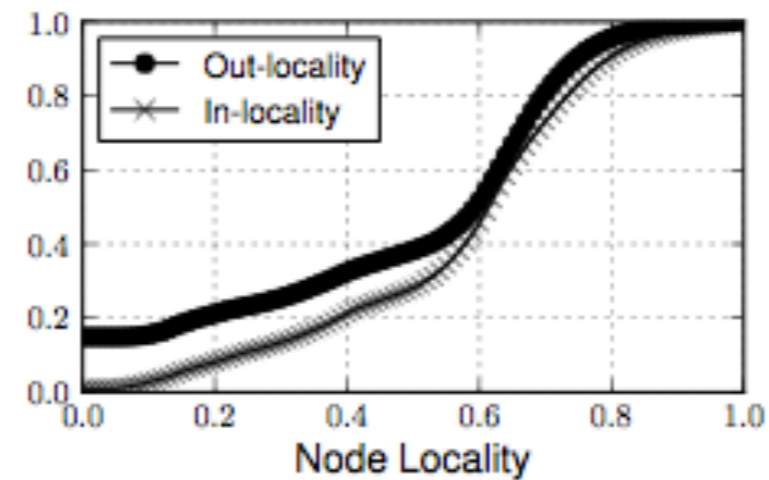
FourSquare



LiveJournal

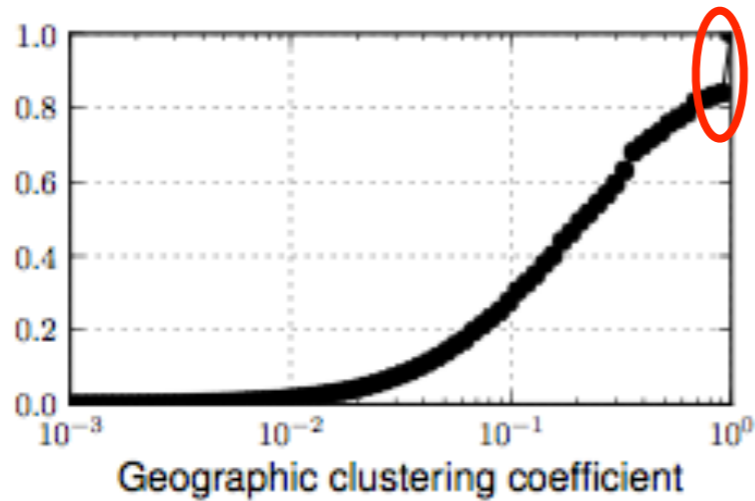


Twitter

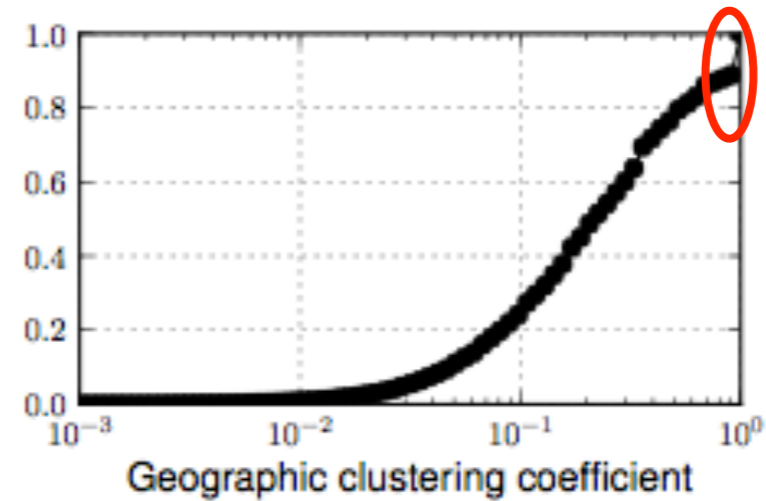


# Geographic Clustering Distributions

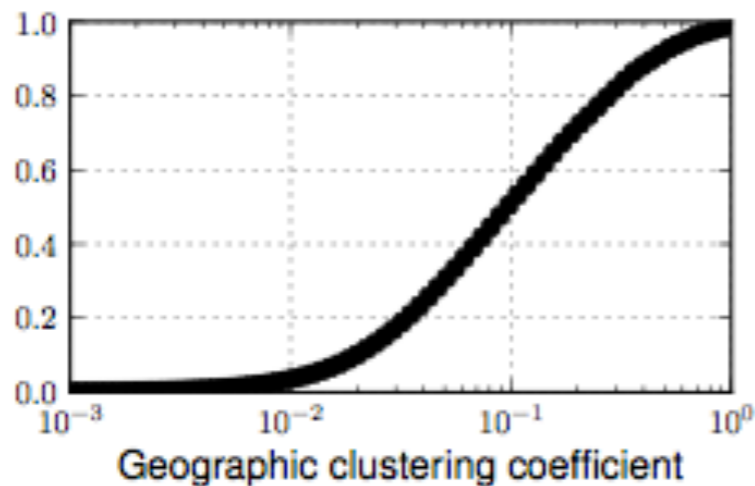
BrightKite



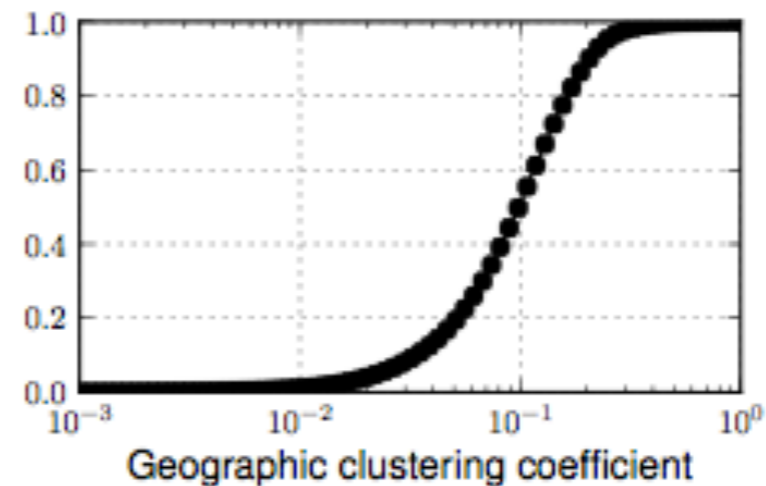
FourSquare



LiveJournal



Twitter



# Findings

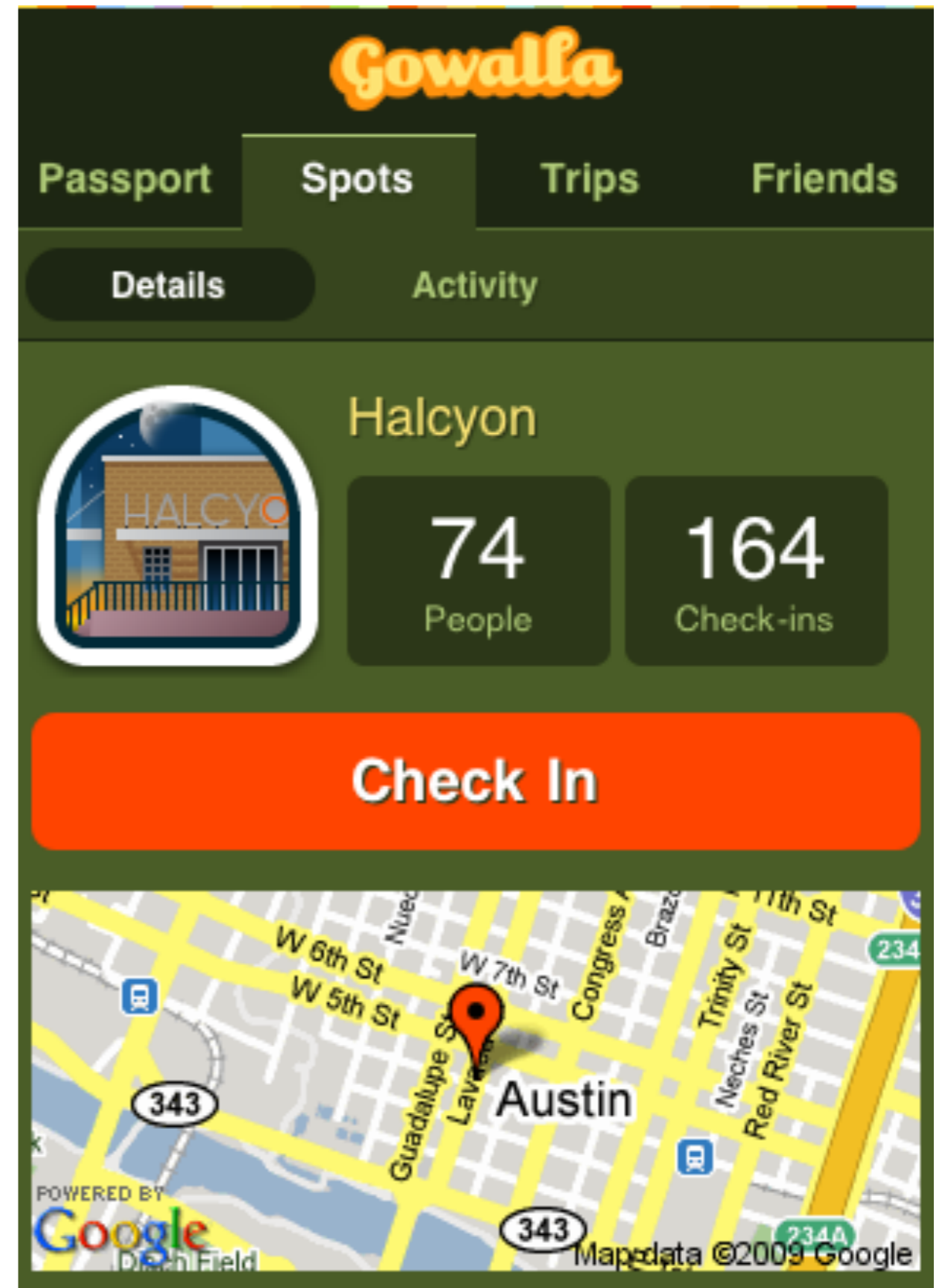
---

Location-based services (LBSs) **foster user interaction on shorter distance.**

LBSs have many users with **predominance of local ties** and local triangles.

Twitter does not exhibit this **'hyperlocal'** behaviour.

In general, **users with higher degrees appear more global**, (with the exception of Twitter).



# Conclusions and future works

---

We have shown how social networks with **geographic information** can be studied and represented.

We have defined two new **geo-social metrics** which take into account both social connections and geographic distance: **node locality** and **geographic clustering coefficient**.

We have collected **4 large-scale online datasets** and applied our metrics to their structure, highlighting **differences** between purely location-based social network services and other online social communities.

In future: **information propagation over space** on Twitter, combining **user mobility** with geo-social metrics, general **geographic generative model** for OSNs.

Thanks!

Questions?

---

**Salvatore Scellato**

**Email:** [salvatore.scellato@cl.cam.ac.uk](mailto:salvatore.scellato@cl.cam.ac.uk)

**Web:** <http://www.cl.cam.ac.uk/~ss824/>

**Twitter:** [www.twitter.com/thetarro](http://www.twitter.com/thetarro)

