Spatial Dissemination Metrics for Location-Based Social Networks

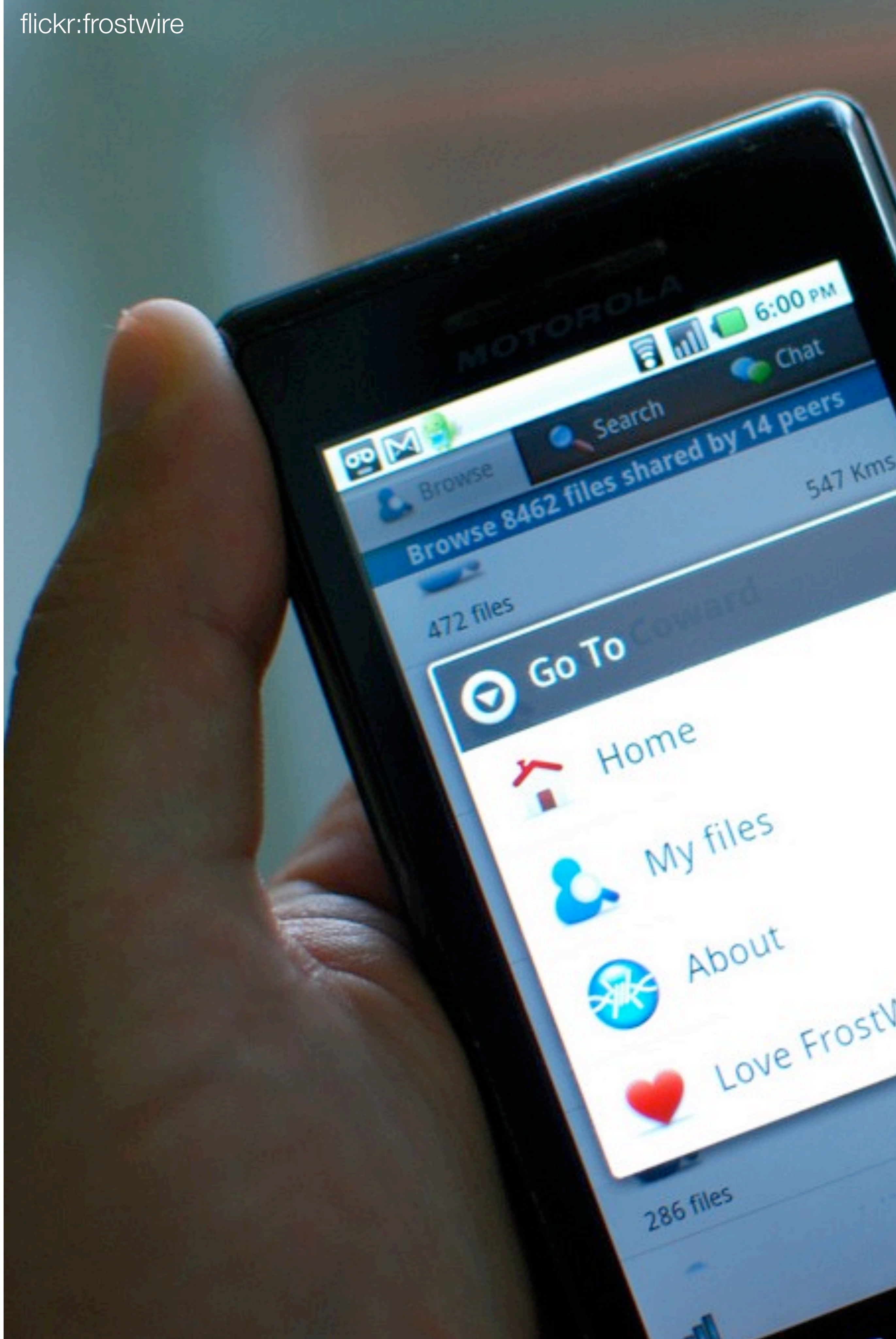
Antonio Lima
Joint work with Mirco Musolesi

LBSN 2012
Pittsburgh, PA - September 8, 2012

UNIVERSITY OF BIRMINGHAM

Information dissemination

- Mobile access to online services and social networks is increasingly common.
- Realtime information dissemination through these channels is important and in some contexts predominant.
- Who are the most important people in the network?

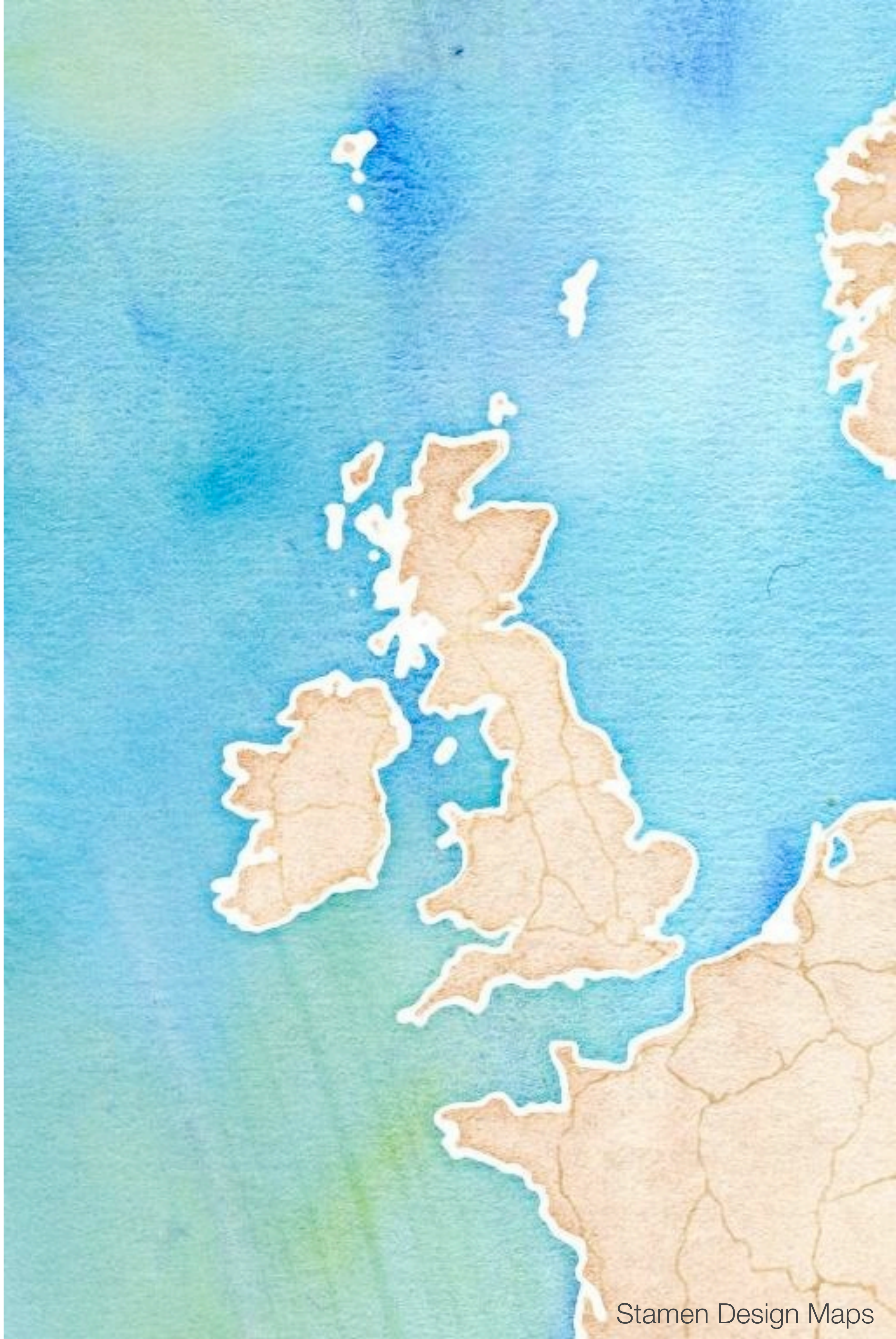


Influence is not global

Even the most influential people are influential in their field of action and in selected regions.

Research Question

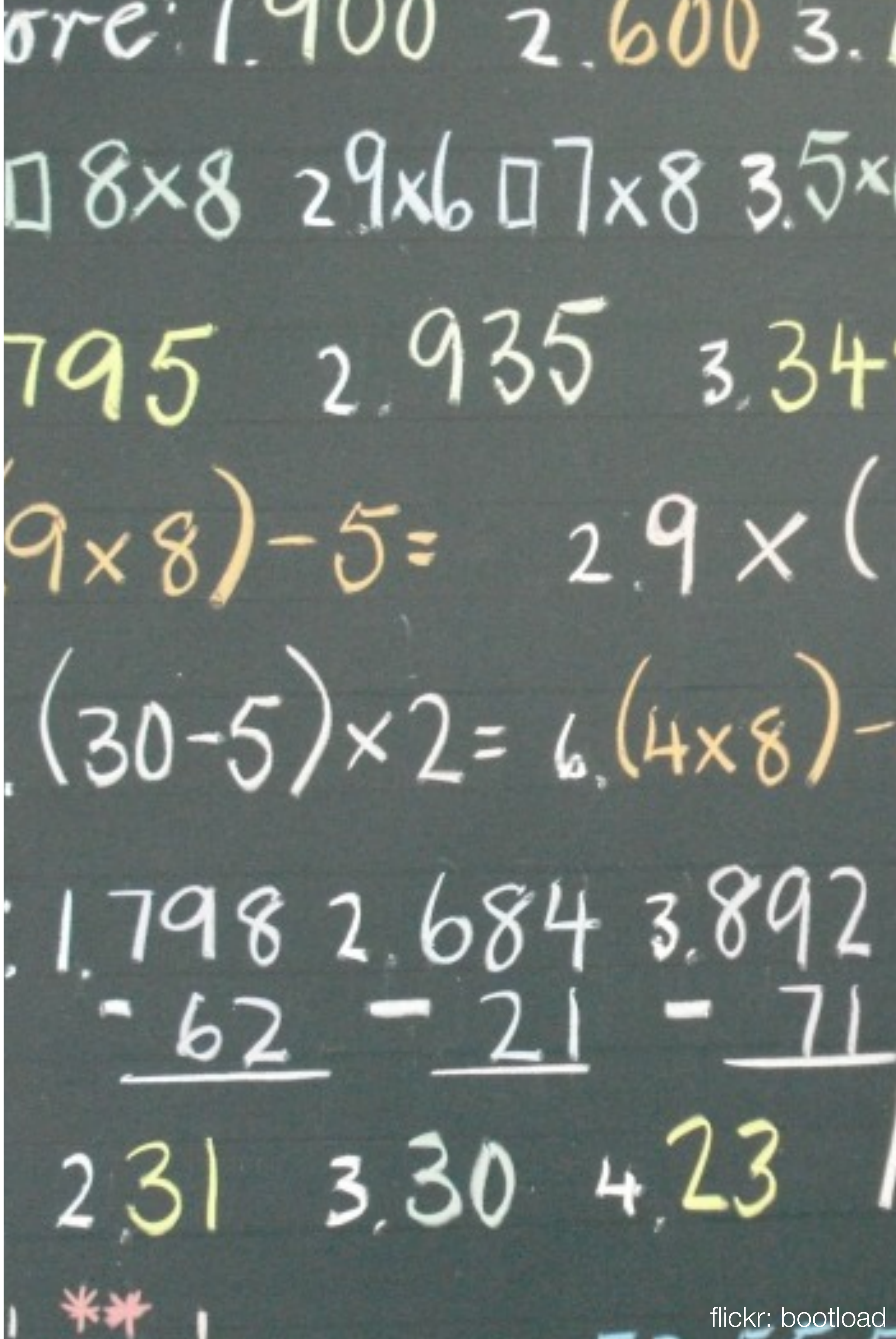
- Complex networks metrics allow us to find most central users in a social network.
- How to find people that are most central to a certain *geographic* region?
- Potential applications in a targeted information spreading and in building models of cultural influence.



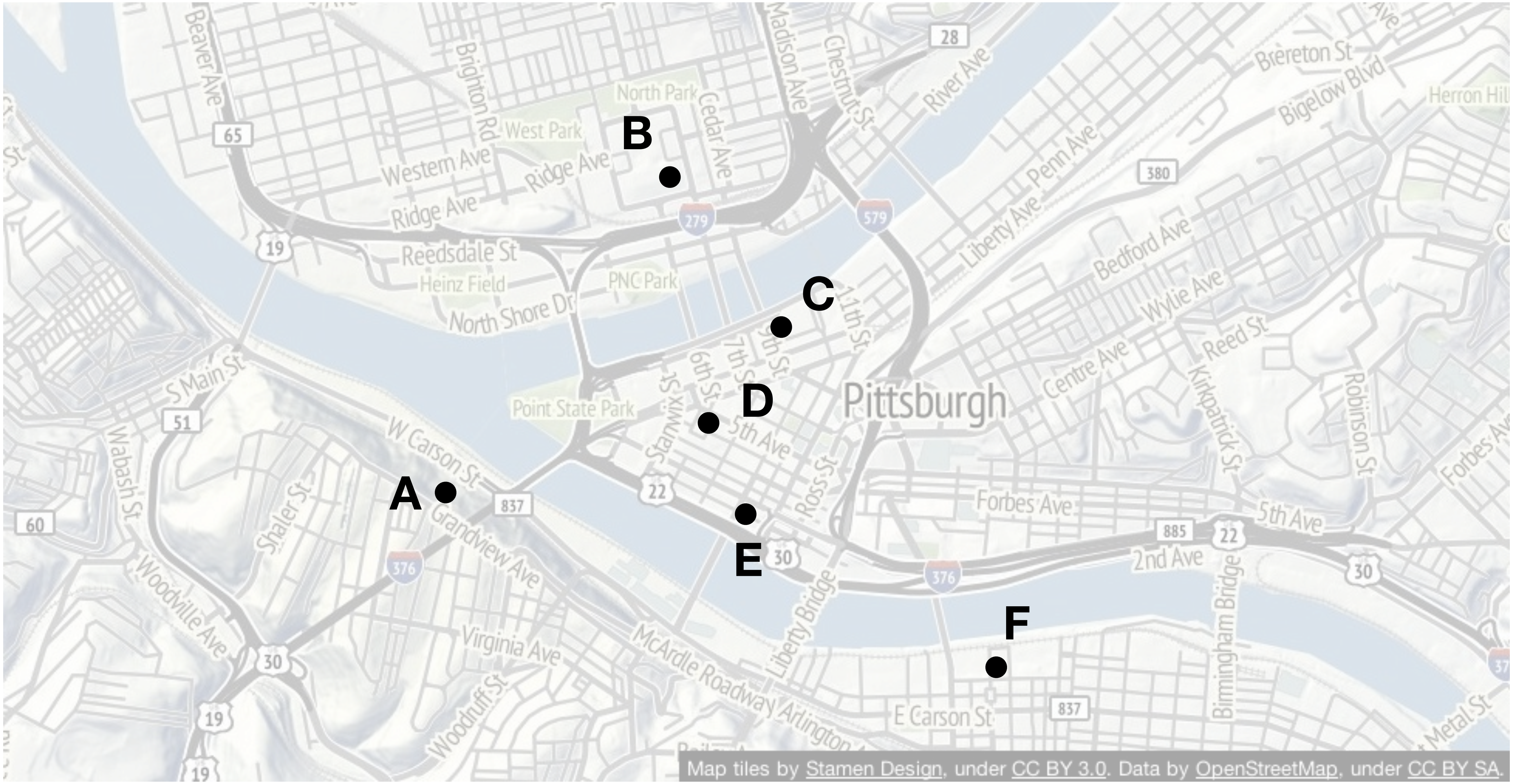
UNIVERSITY OF BIRMINGHAM

Our Approach

- A geo-social network model.
- Geographic extension of centrality measures defined for complex (social) networks. These are *structural* metrics.
- Analysis of real-world scenarios on two major social networks websites, Twitter and Foursquare.



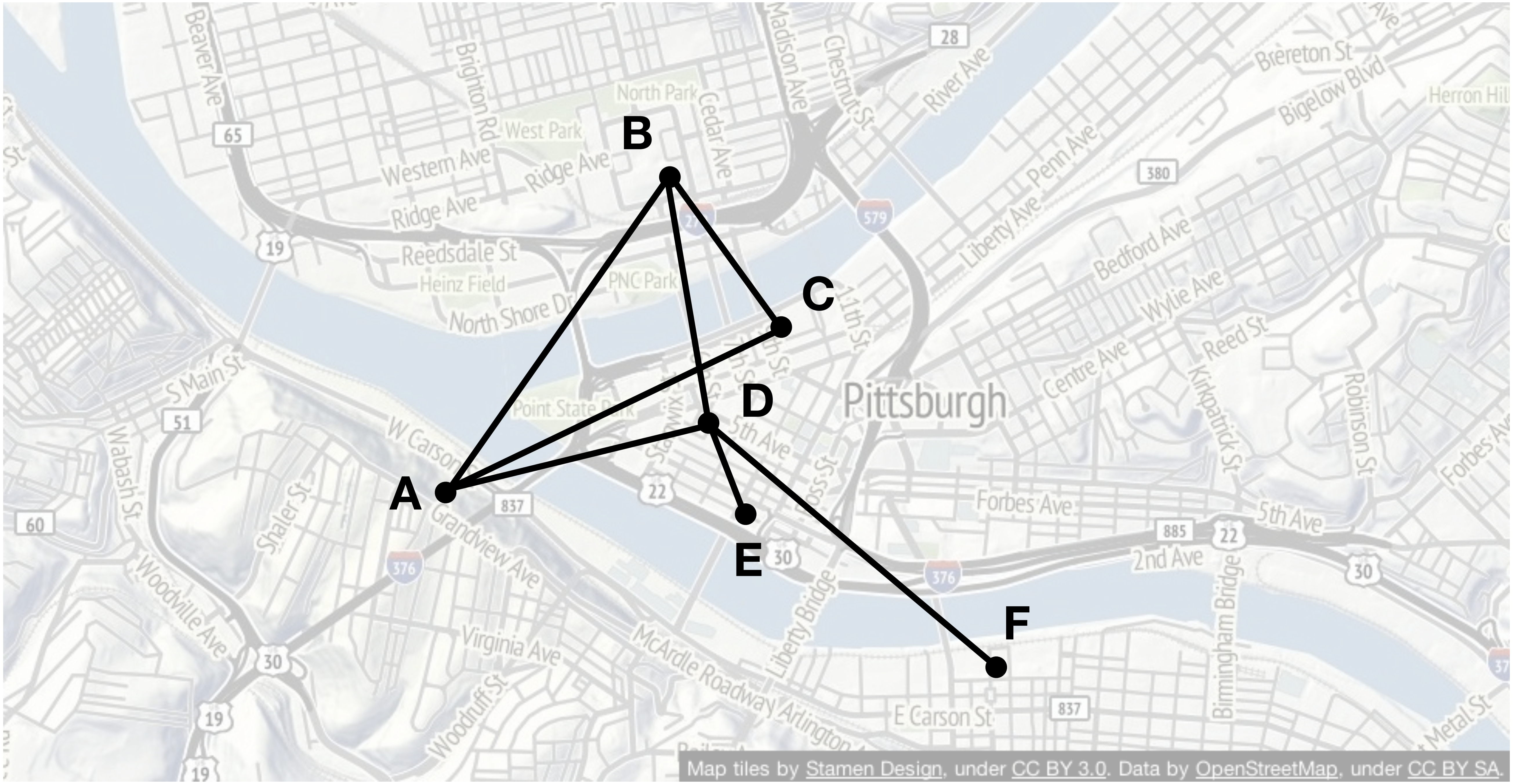
flickr: bootload



Geosocial Network Model

Every user is associated one or more significant point in a geographic space (home, office, favorite café, ...)

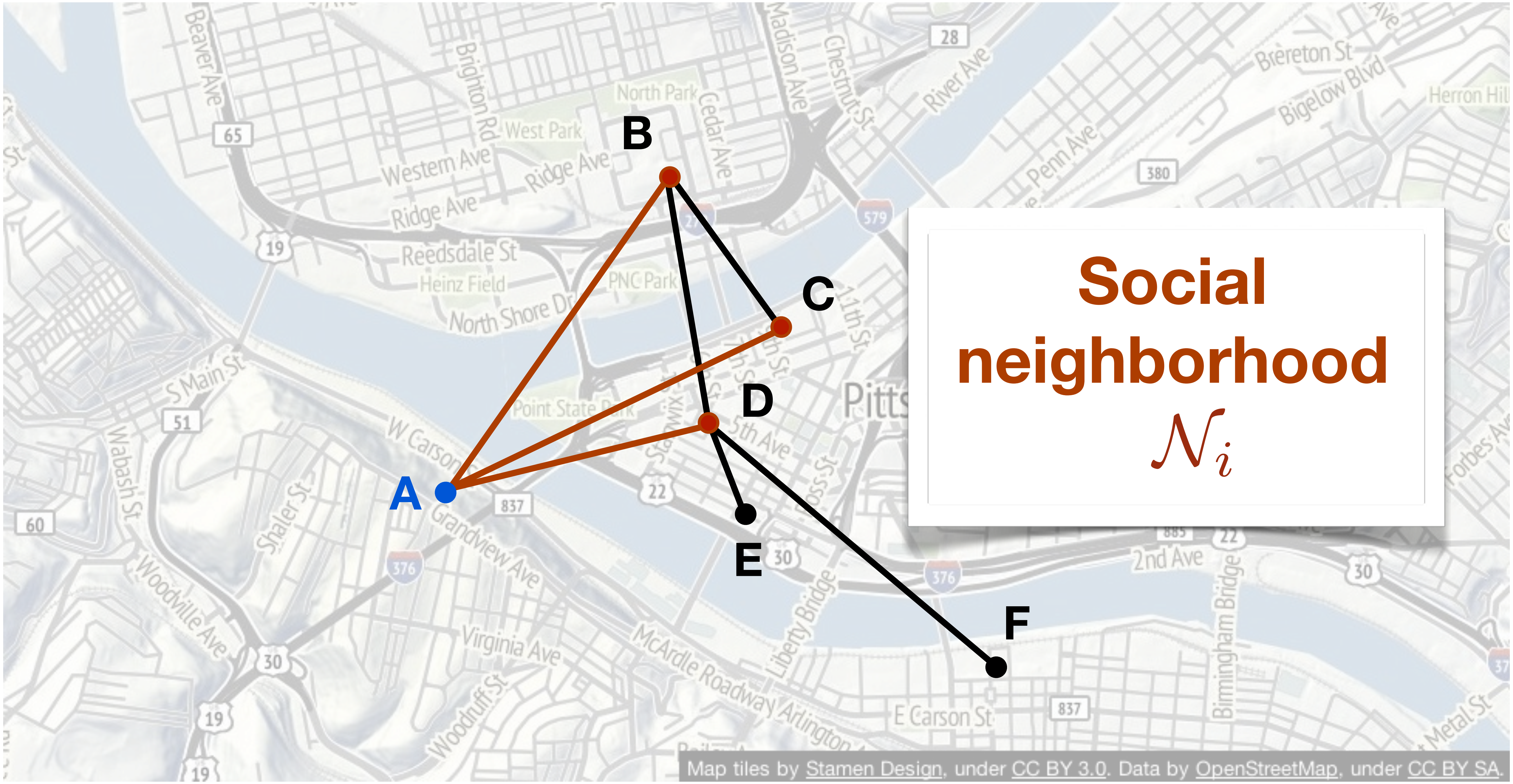
UNIVERSITY OF BIRMINGHAM



Geosocial Network Model

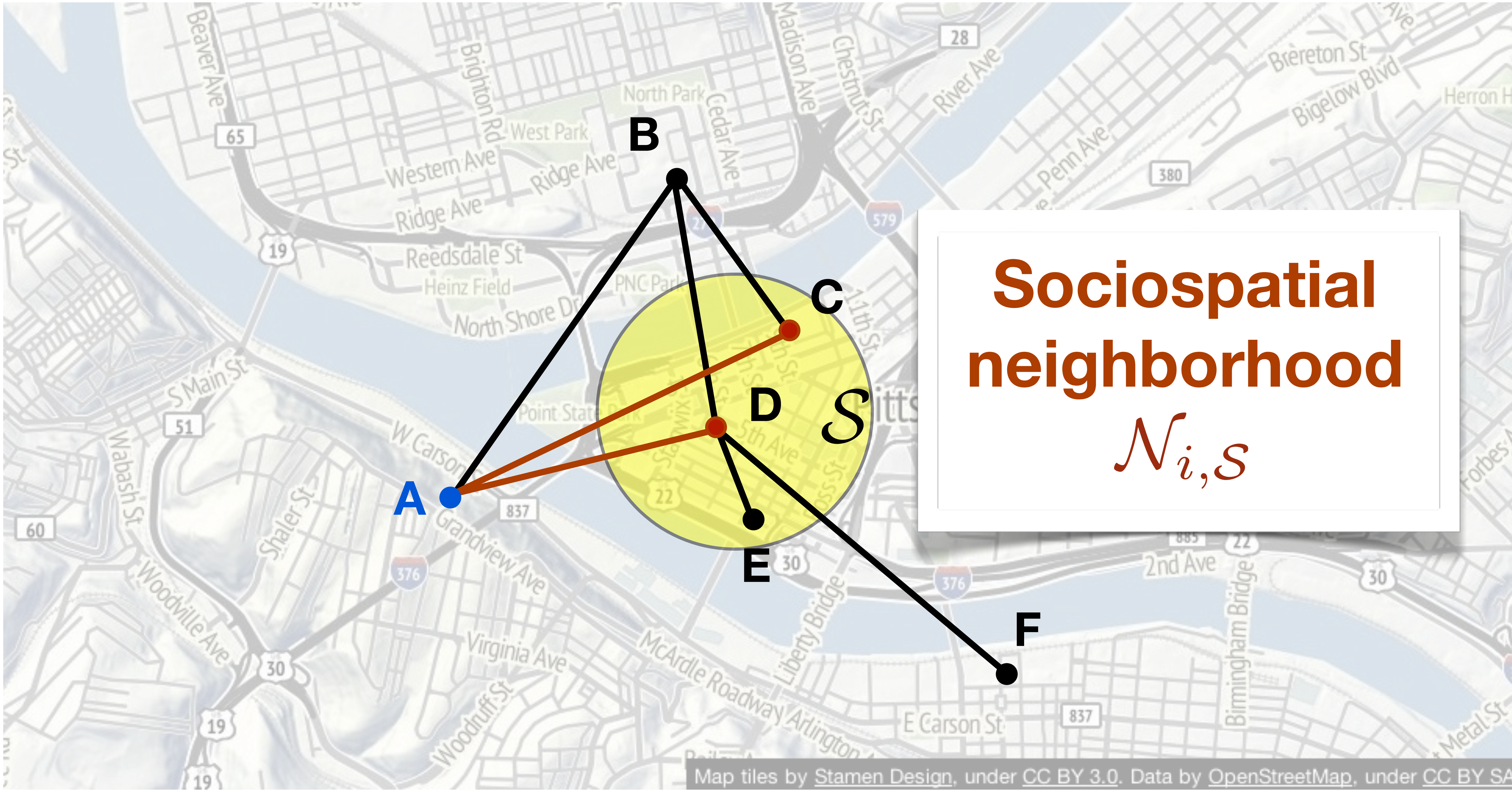
We also know the social network of this group of people.

UNIVERSITY OF BIRMINGHAM



Geosocial Network Model

Social neighborhood of a node. It is defined only on the social graph (no geographic info).



Geosocial Network Model

Spatio-Social neighborhood of a node w.r.t. to the yellow region.

UNIVERSITY OF BIRMINGHAM

Twitter Dataset

- Snowball sampling.

- 1375 seed users in San Francisco, CA and London, UK. 657K users (1375 seeds) and their social links.

- User significant point specified in their profile (location field).

- Location was geocoded using Google Geocoding API.

Foursquare Dataset

- Mayor of a venue: user with the highest number of check-ins in the last 60 days.

- Random crawling of venues in selected urban areas, their mayors' profile and friends.

- 177K users and their social links.

- Mayorships describe users significant points.

Spatial Degree Centrality and Spatial Degree Ratio

$$C_{i,\mathcal{S}} = |\mathcal{N}_{i,\mathcal{S}}|$$

Quantifies how many neighbors of *i* have significant points inside the region *S*.

Spatial Degree Centrality and Spatial Degree Ratio

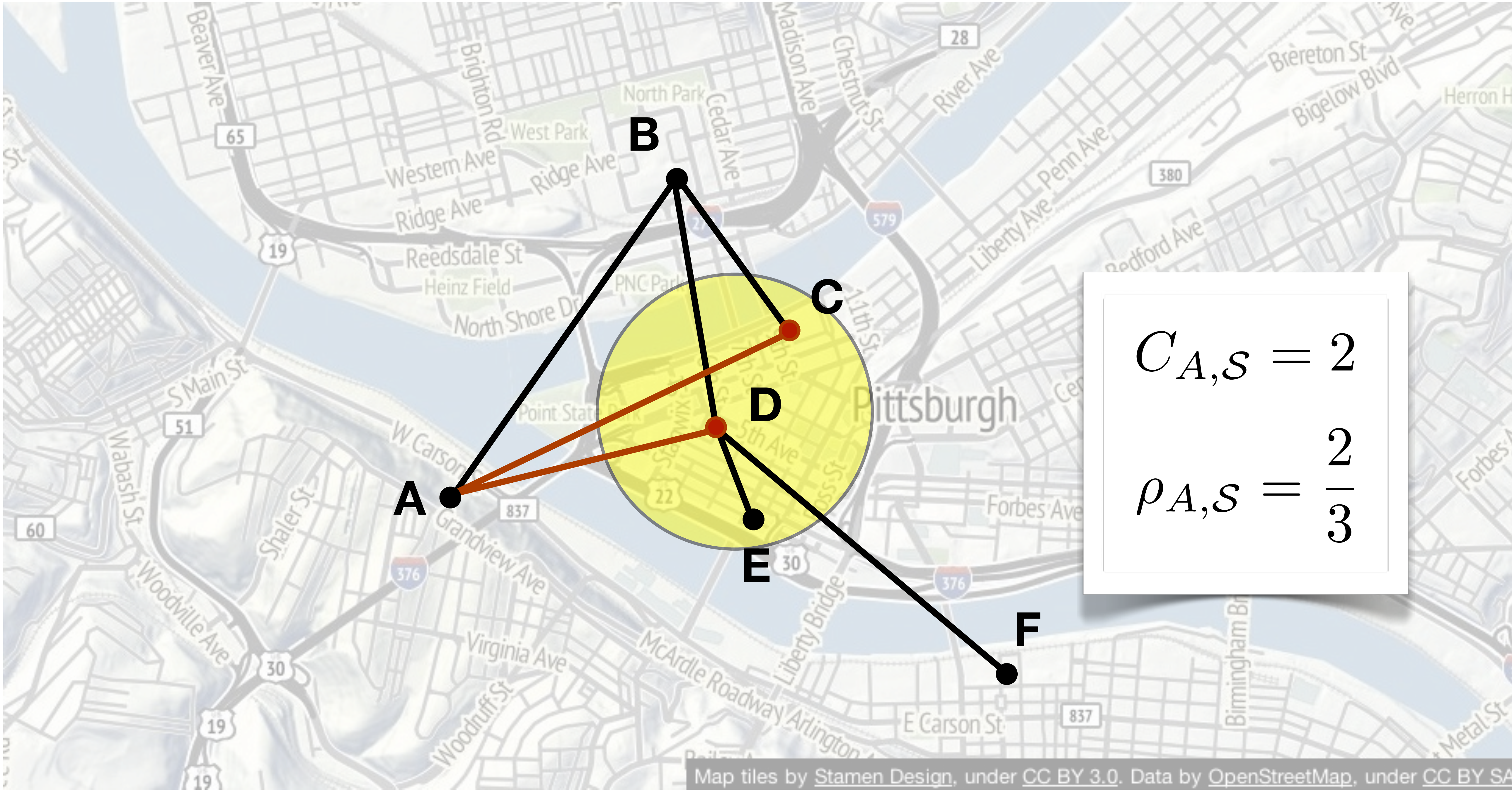
$$C_{i,\mathcal{S}} = |\mathcal{N}_{i,\mathcal{S}}|$$

Quantifies how many neighbors of *i* have significant points inside the region *S*.

Normalize

$$\rho_{i,\mathcal{S}} = \frac{|\mathcal{N}_{i,\mathcal{S}}|}{|\mathcal{N}_i|}$$

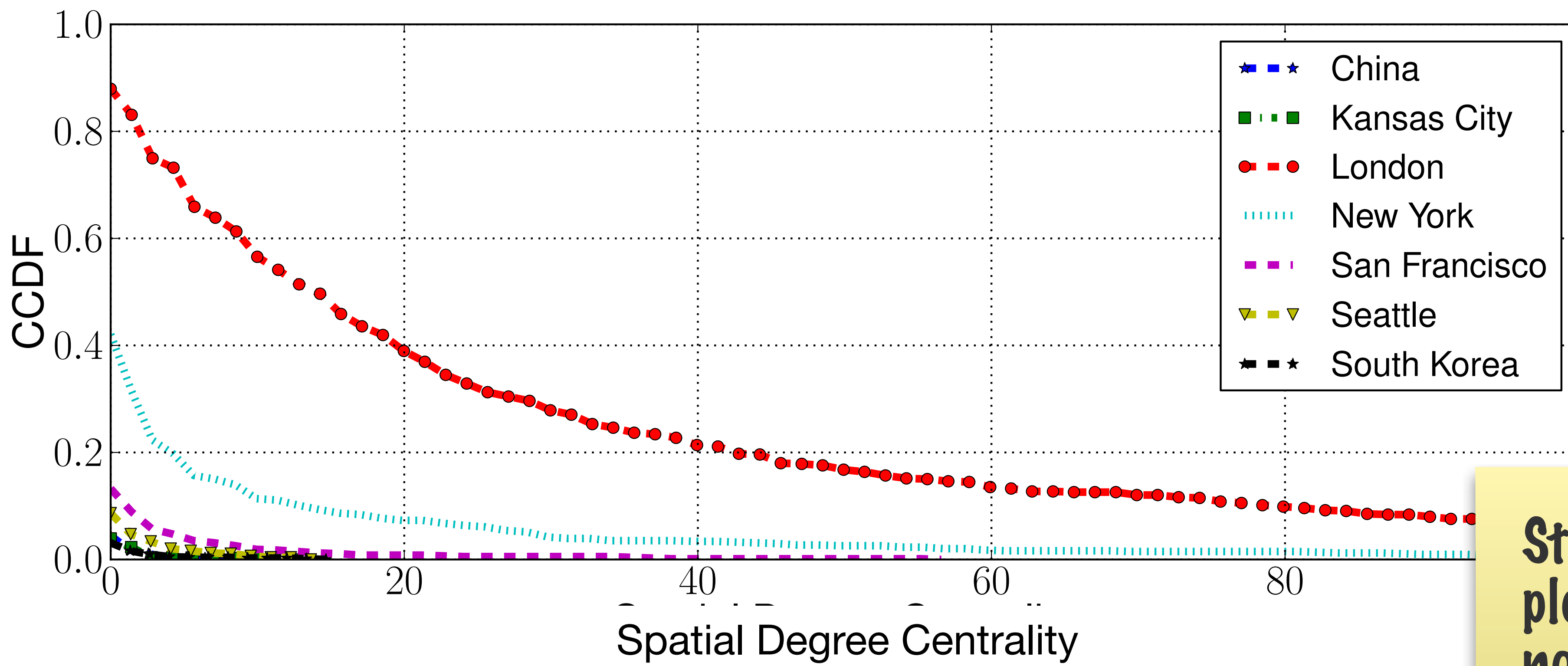
Quantifies the fraction of neighbors of *i* have significant points inside the region *S*.



$$C_{A,\mathcal{S}} = 2$$

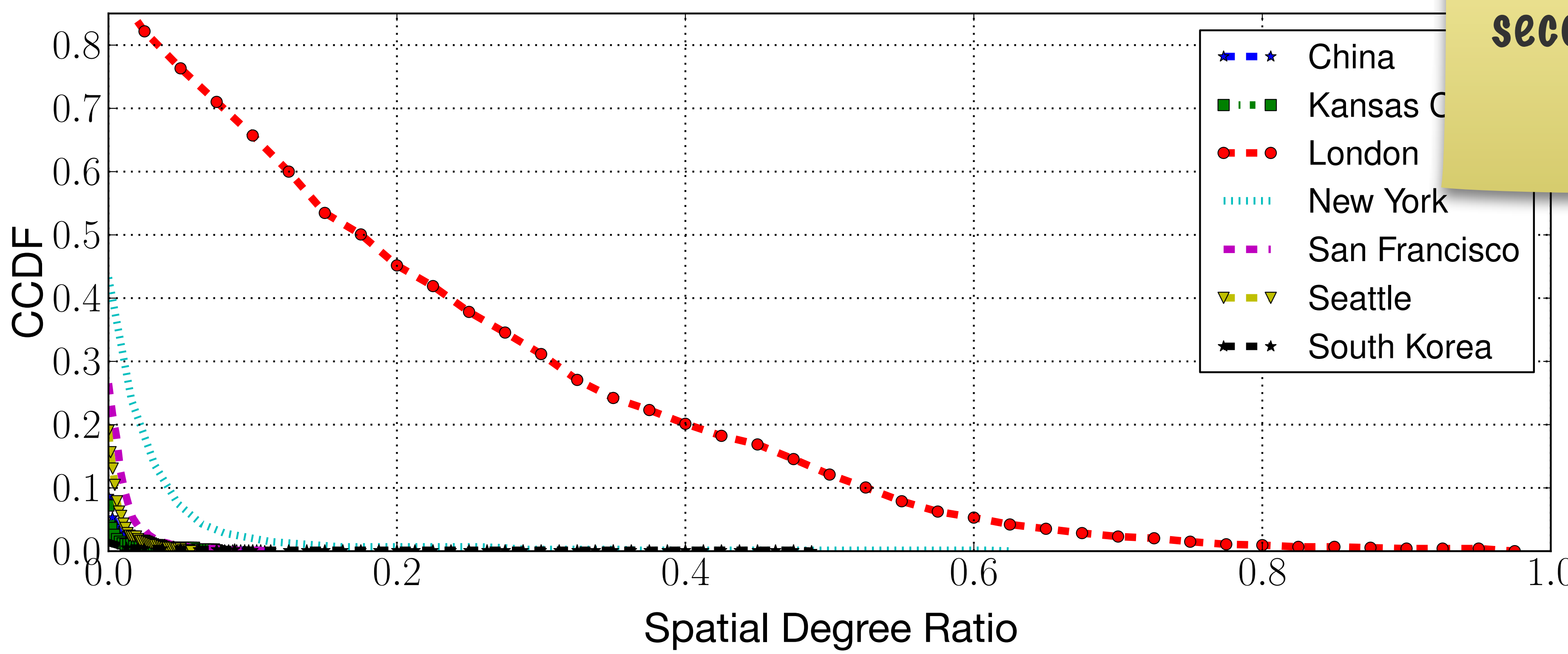
$$\rho_{A,\mathcal{S}} = \frac{2}{3}$$

$C_{i,\mathcal{S}}$



Stress that first plot is not normalized, second is

$\rho_{i,\mathcal{S}}$

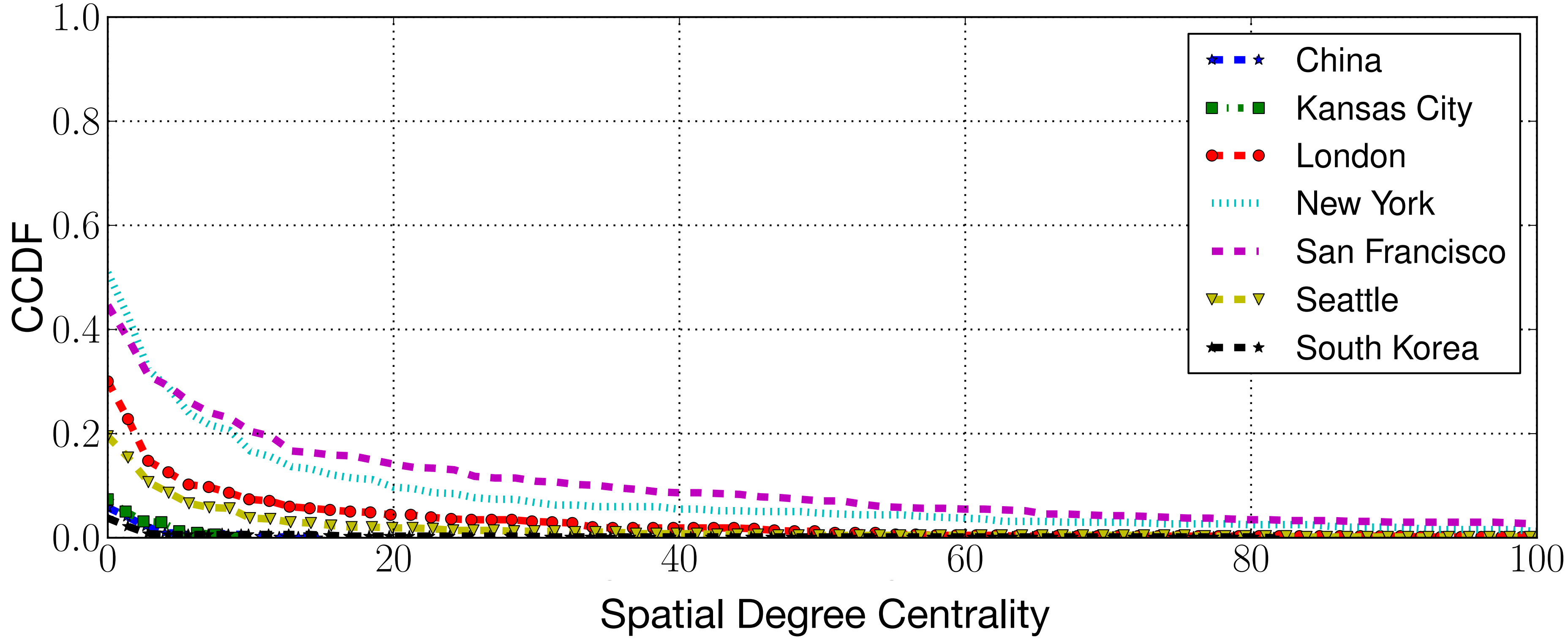


Users in London (Twitter)

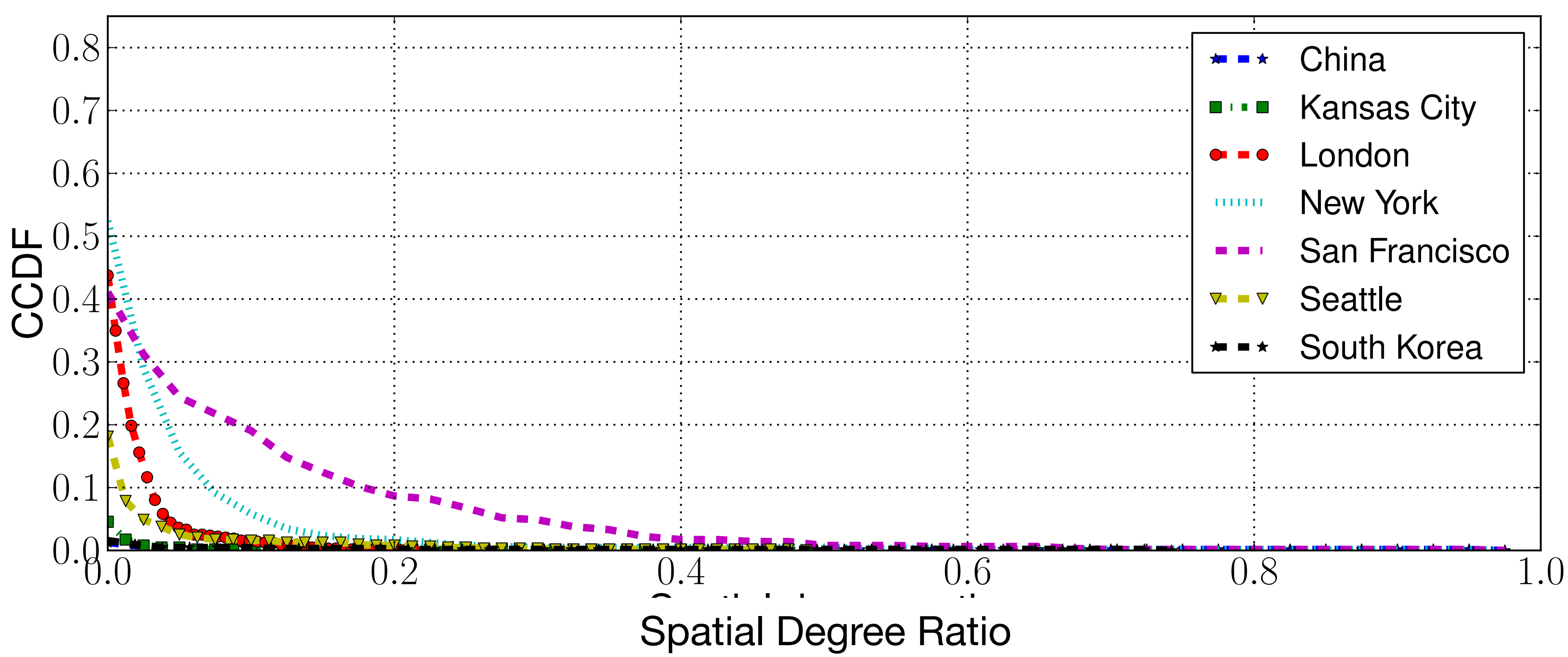
Londoners are mostly central towards London and somewhat central to New York as well.

UNIVERSITY OF BIRMINGHAM

$C_{i,\mathcal{S}}$



$\rho_{i,\mathcal{S}}$

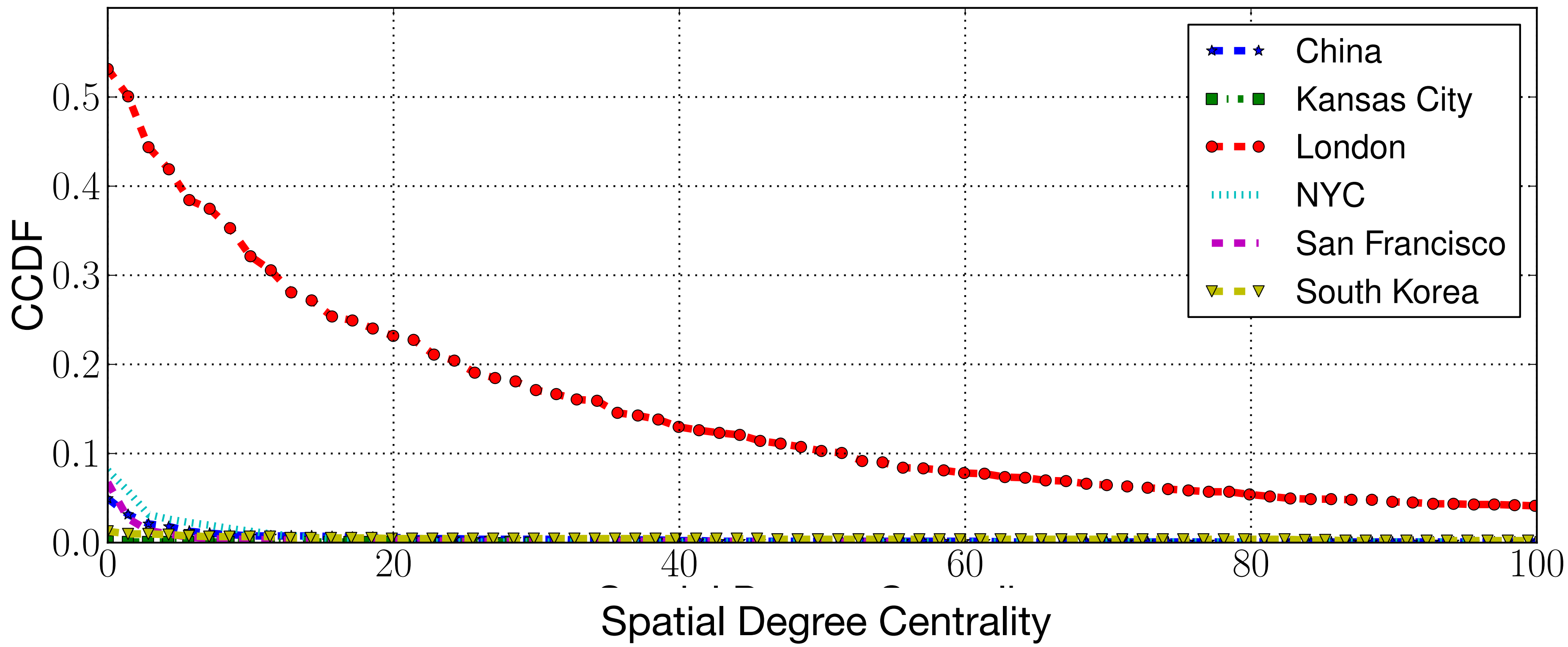


Users in San Francisco (Twitter)

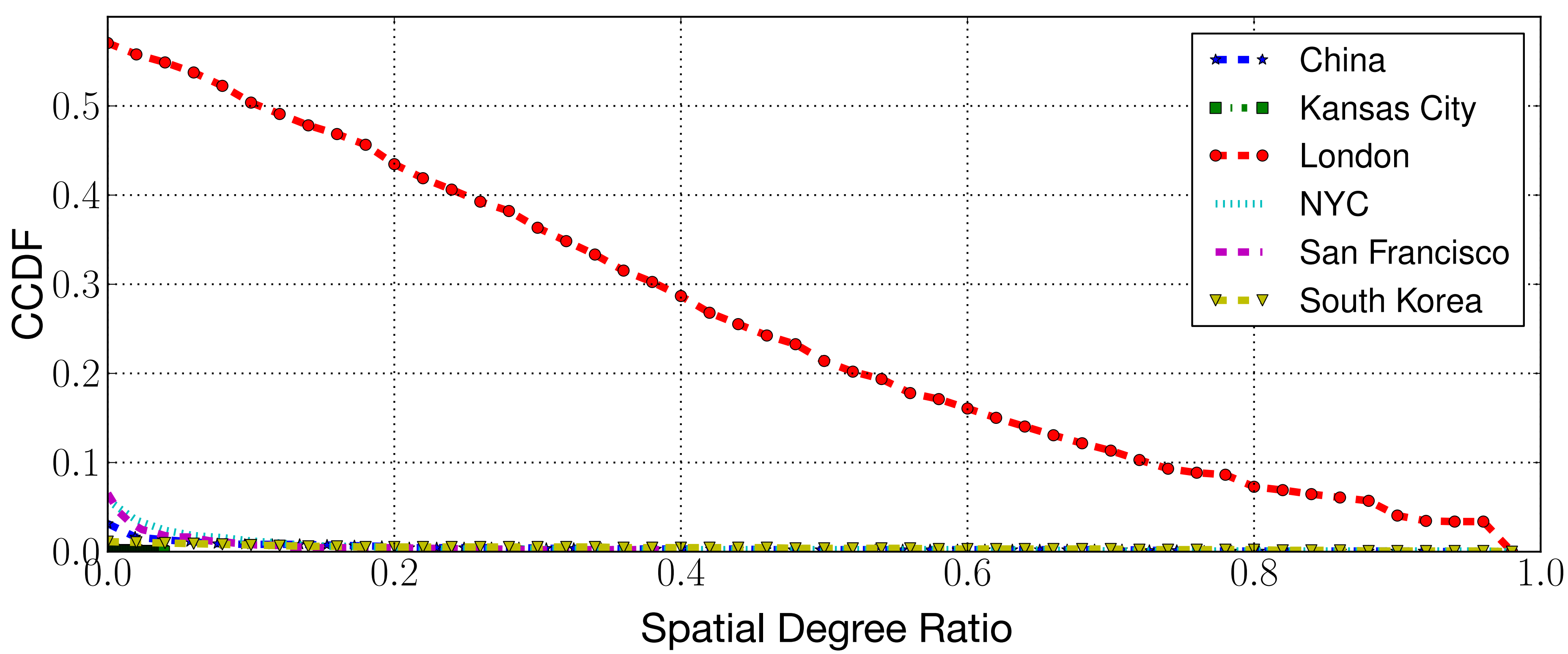
Interestingly, San Franciscan users have a similar distribution of centrality w.r.t. New York and San Francisco.

UNIVERSITY OF BIRMINGHAM

$C_{i,\mathcal{S}}$



$\rho_{i,\mathcal{S}}$

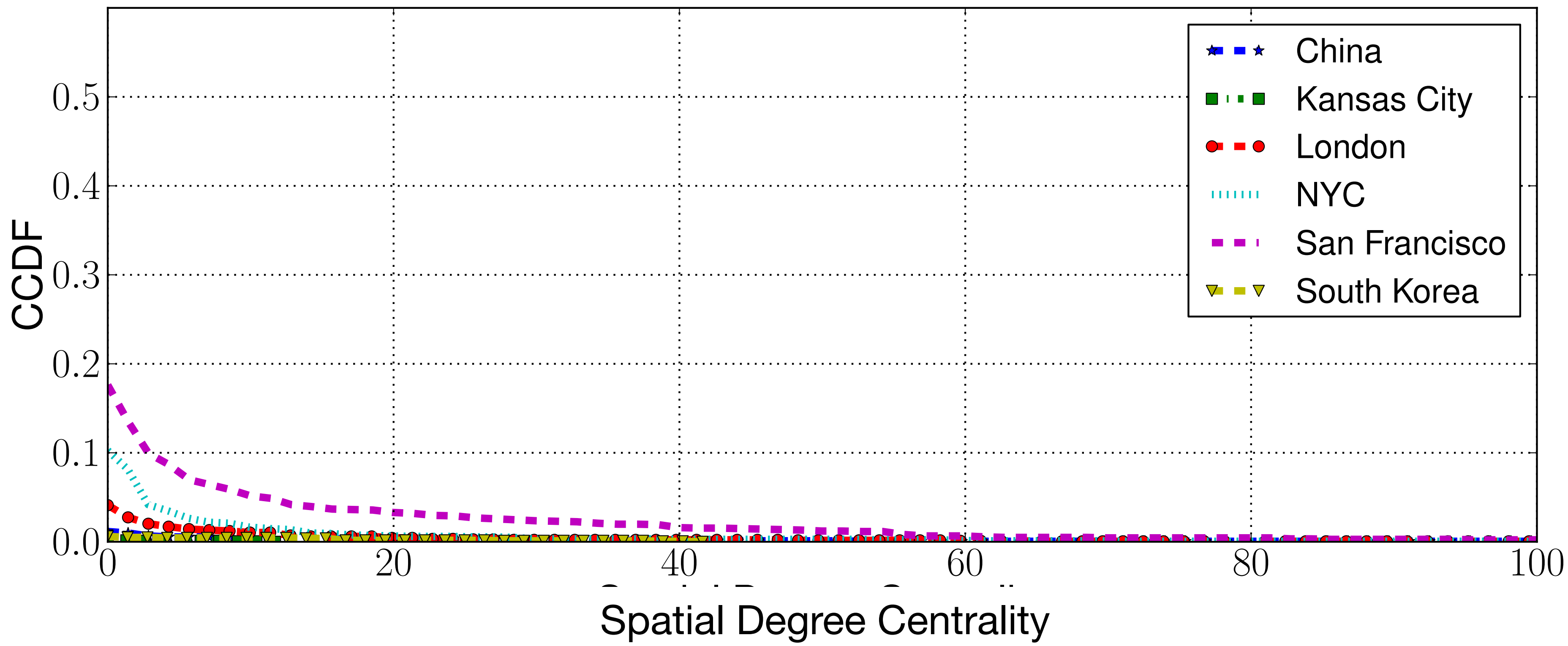


Users in London (Foursquare)

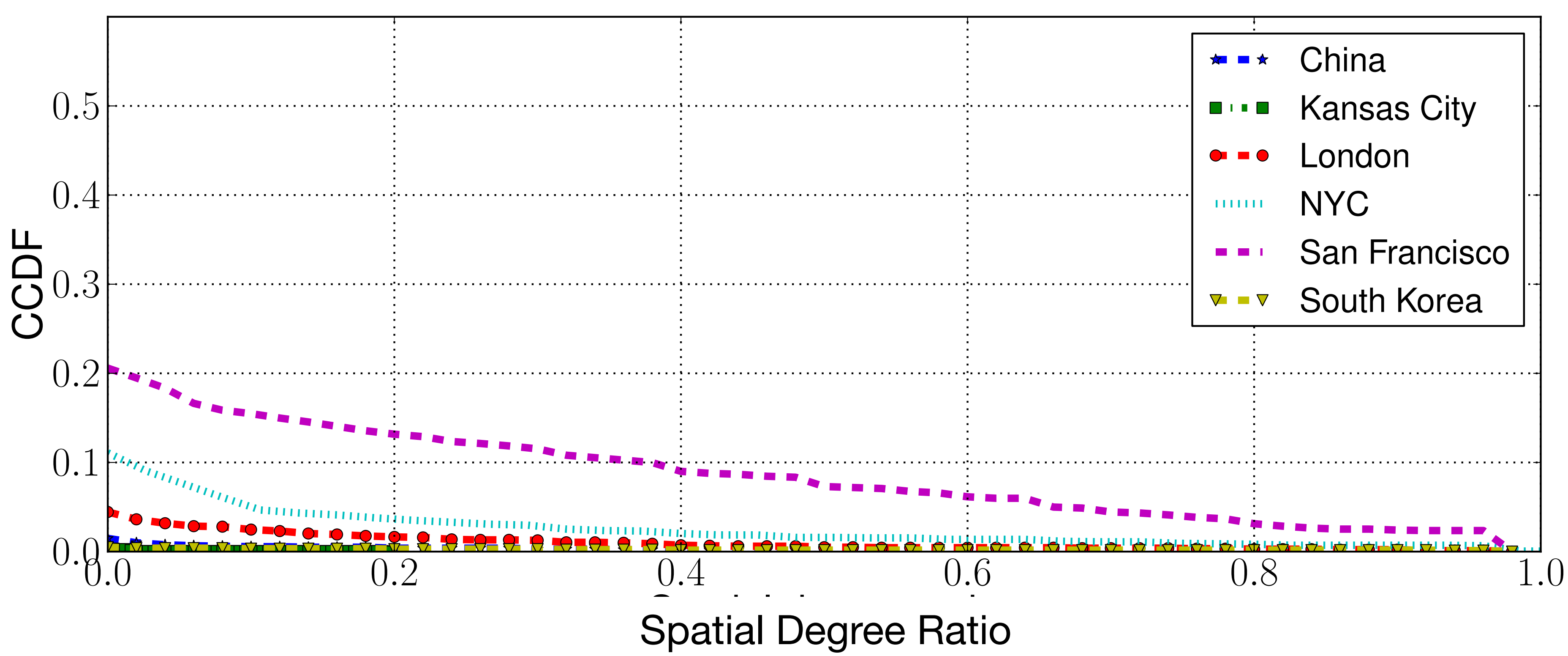
Foursquare exhibits lower avg degree (due to lower penetration rate). Results are in accordance with those observed for Twitter.

UNIVERSITY OF BIRMINGHAM

$C_{i,\mathcal{S}}$



$\rho_{i,\mathcal{S}}$



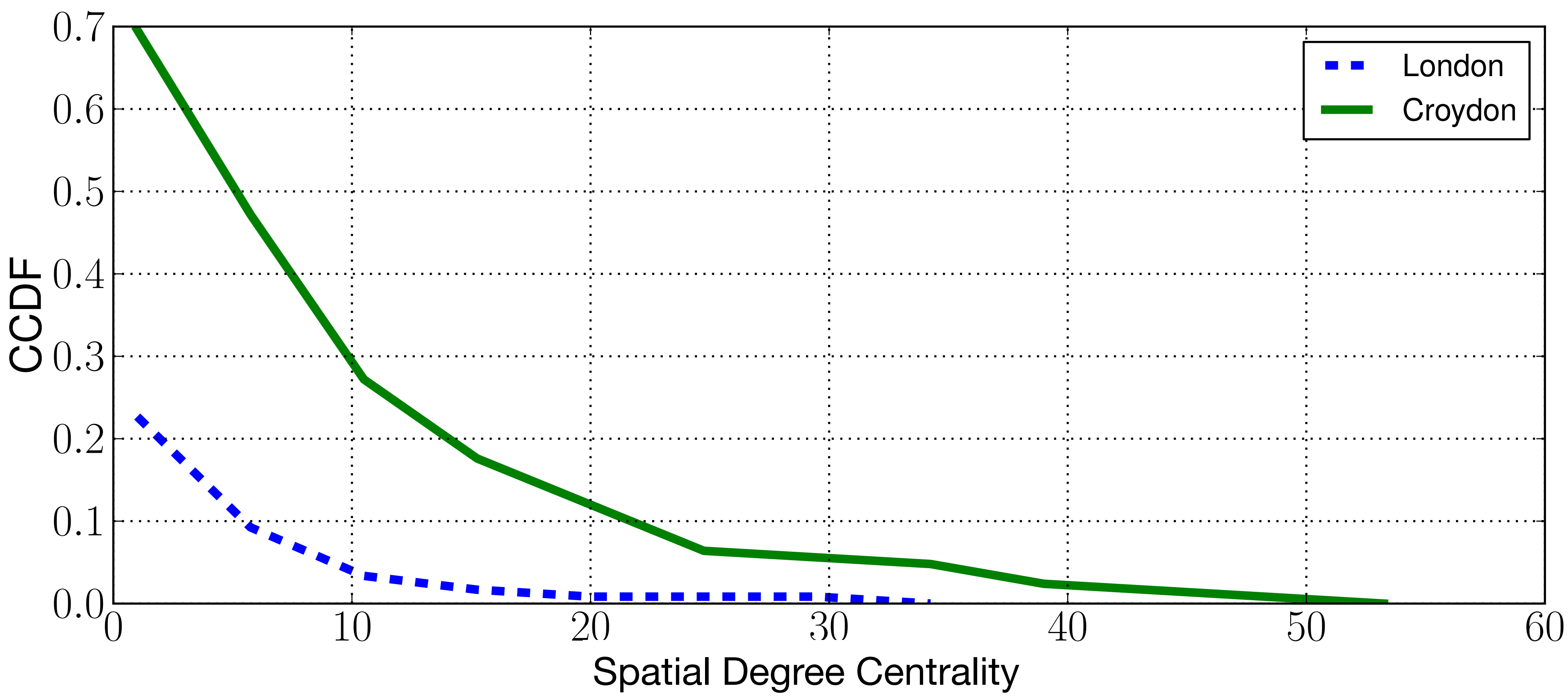
Users in San Francisco (Foursquare)

While London had no "influence" over other areas, San Francisco still has some influence on New York, as seen for Twitter.

Spatial Degree Centrality for Foursquare U
Croydon and London w.r.t. Croydon

The intra-city analysis cannot be carried out on the Twitter dataset.

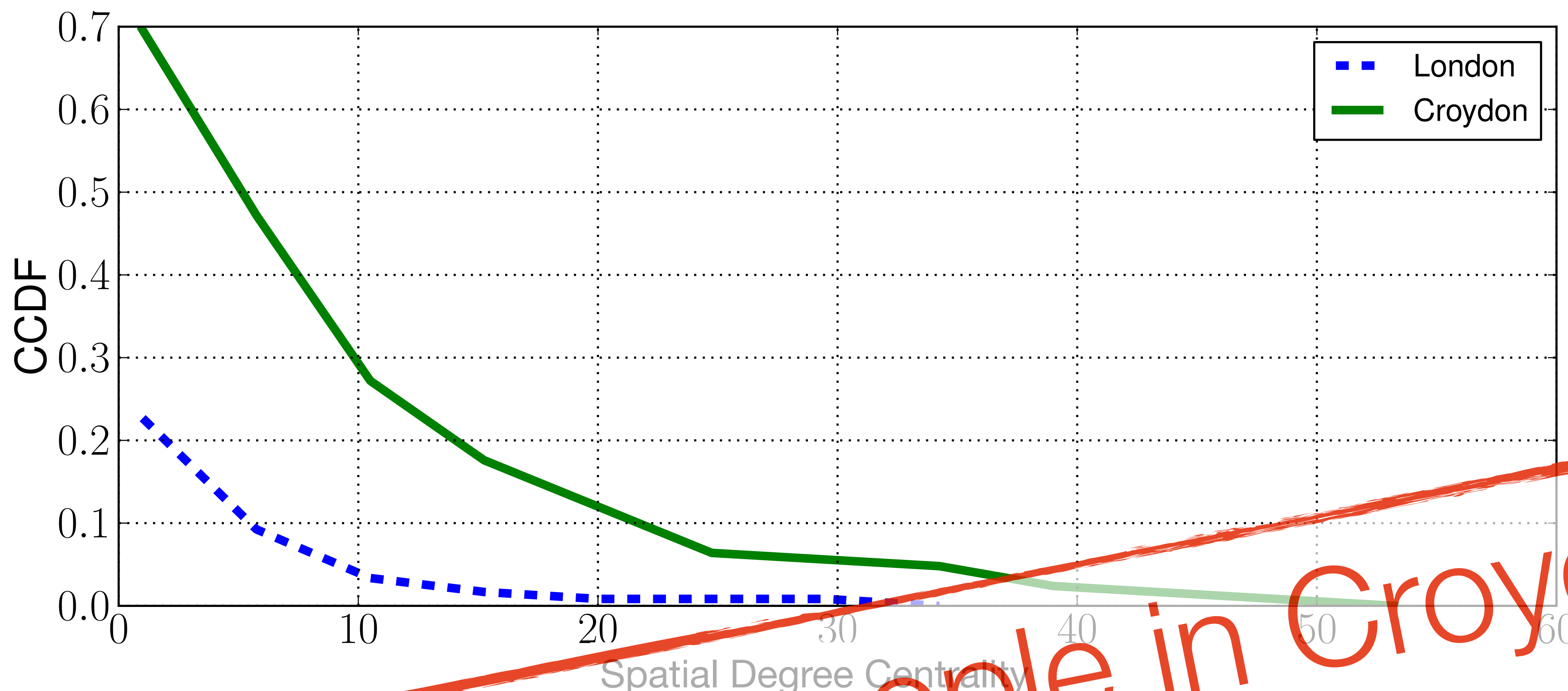
$$C_{i,\mathcal{S}}$$



Spatial Degree Centrality for Foursquare U
Croydon and London w.r.t. Croydon

The intra-city analysis cannot be carried out on the Twitter dataset.

$$C_{i,\mathcal{S}}$$



Targeting people in Croydon might give an advantage

UNIVERSITY OF BIRMINGHAM

Spatial Degree Centrality for Foursquare Users in SF Chinatown and SF w.r.t. San Francisco

- Avg centrality of Chinatown and San Francisco users w.r.t. Chinatown are comparable (3.20 vs 3.06).

- Avg centrality of Chinatown users w.r.t. to China is three times bigger than the centrality from San Franciscans (32.24 vs 11.87).

Potential cultural influence

Spatial Degree Centrality for Foursquare Users in SF Chinatown and SF w.r.t. San Francisco

- Avg centrality of Chinatown and San Francisco users w.r.t. Chinatown are comparable (3.20 vs 3.06).
- Avg centrality of Chinatown users w.r.t. to China is three times bigger than the centrality from San Franciscans (32.24 vs 11.87).

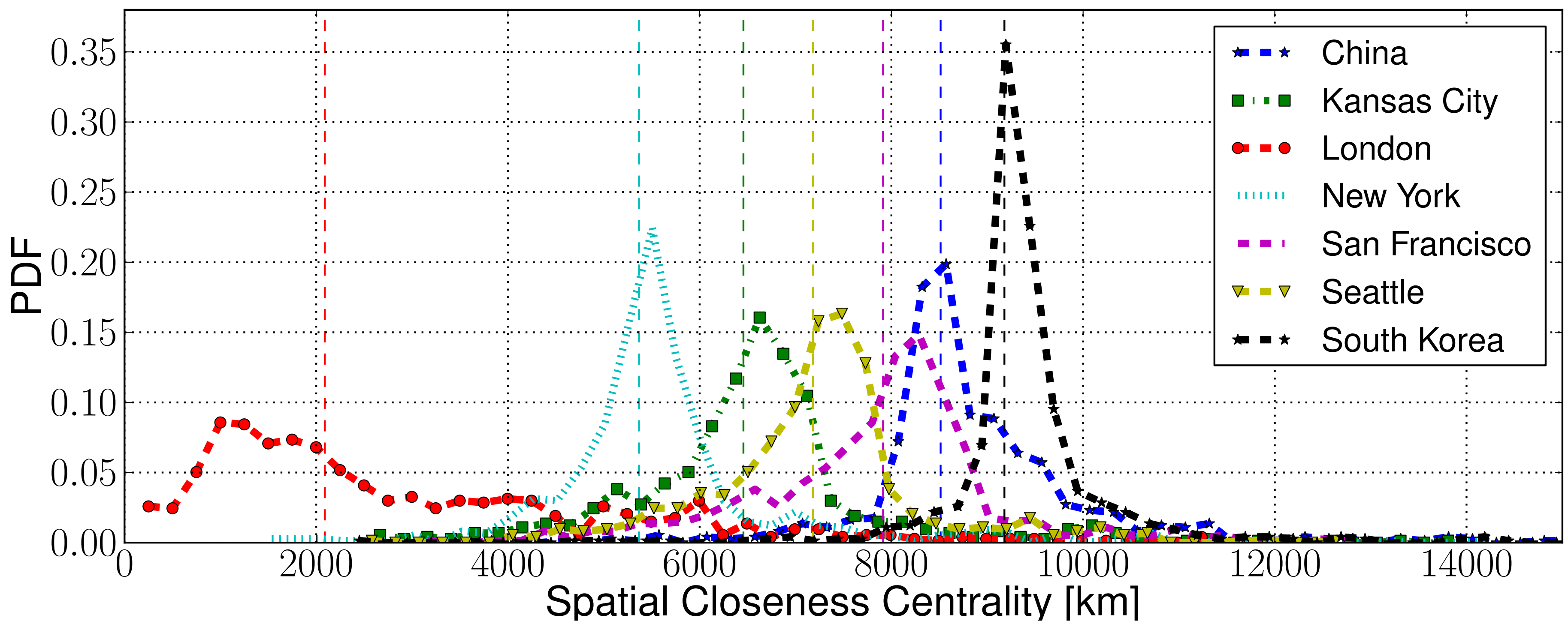
Advantage towards China, not SF Chinatown

Spatial Closeness Centrality

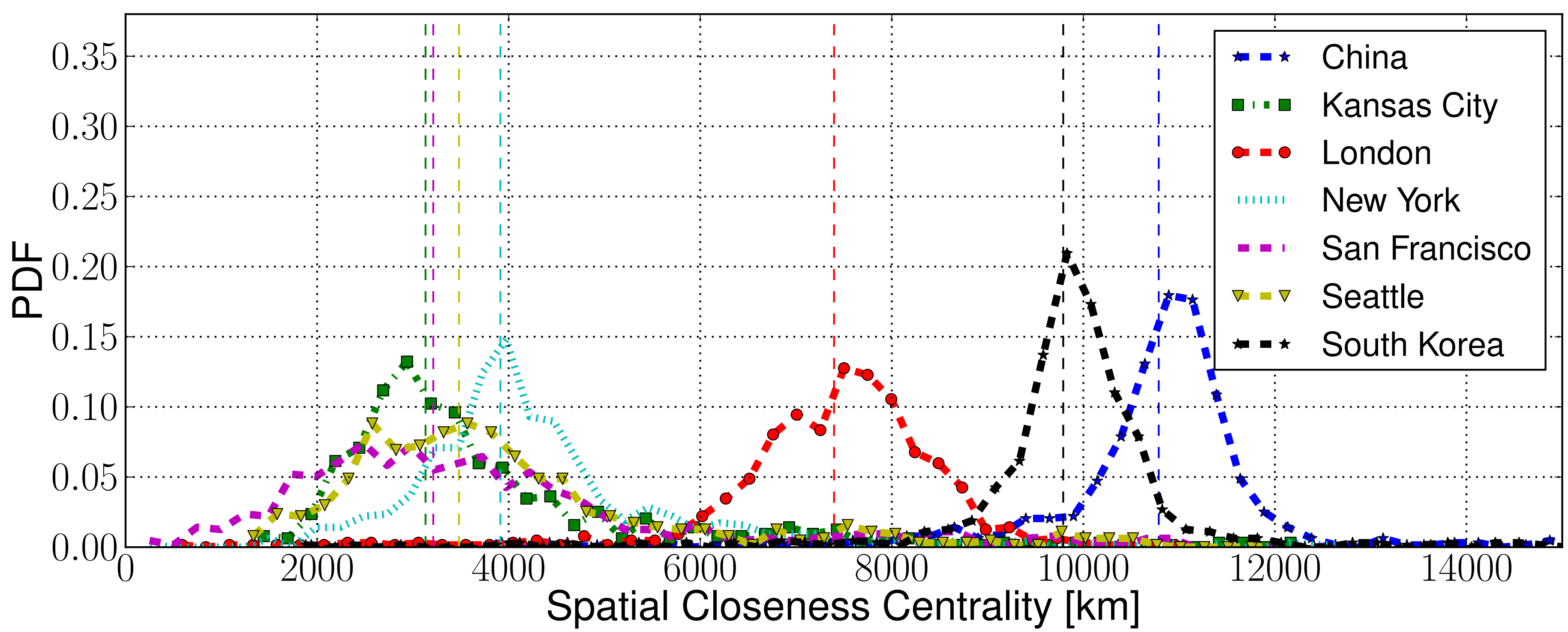
$$C^{C}_{i,p^{\star}} = \frac{1}{|\mathcal{N}_i|} \sum_{j \in |\mathcal{N}_i|} d_G(p_j, p^{\star})$$

It is the average geographic distance of all neighbors' significant places from a specific geographic point.

It is an indicator of how the influenced audience of a user is geographically close to a certain location.



from London



from San Francisco

Stress that it is London vs other cities and SF vs other cities

Spatial Closeness Centrality (Twitter)

Peak/median very close to the distance between considered locations.

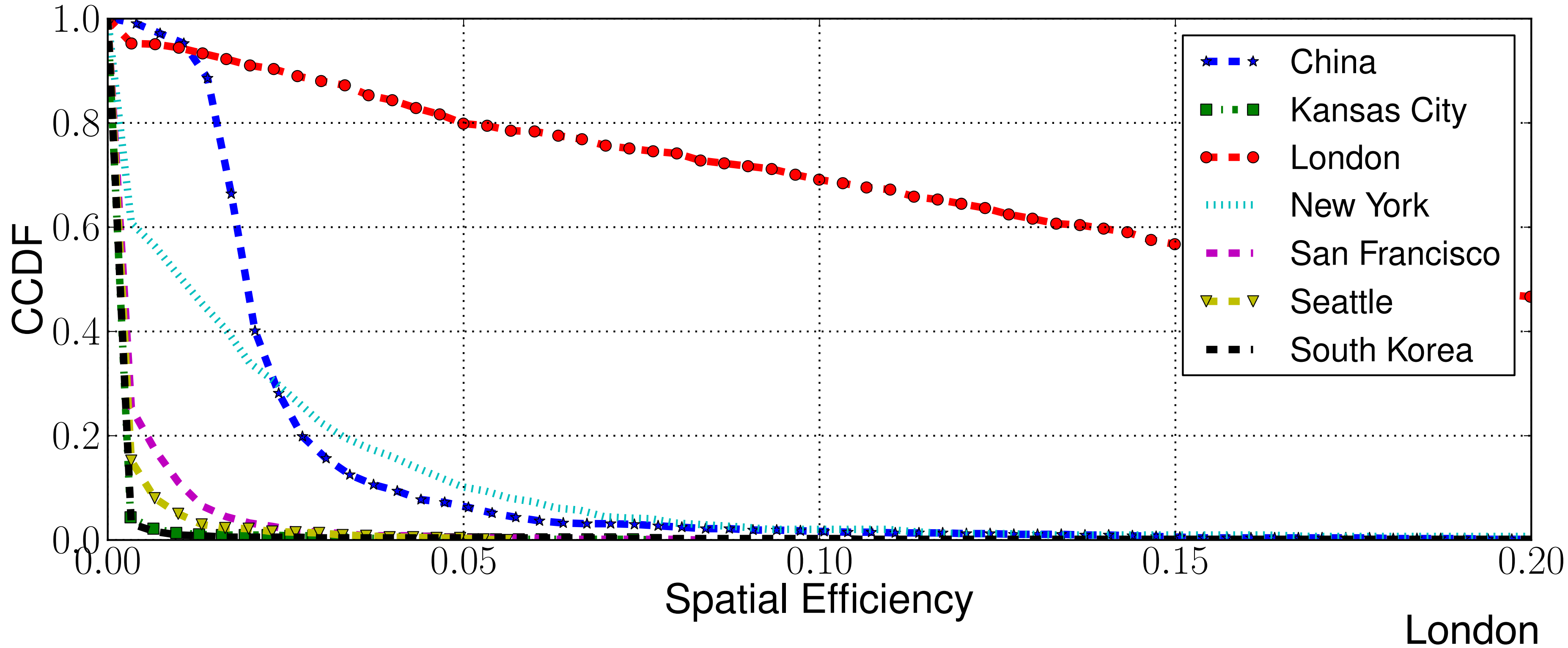
Spatial Efficiency Centrality

$$C^E_{i,p^\star} = \frac{1}{k_i} \sum_{j \in \mathcal{N}_i} \frac{1}{d_G(p_j, p^\star)}$$

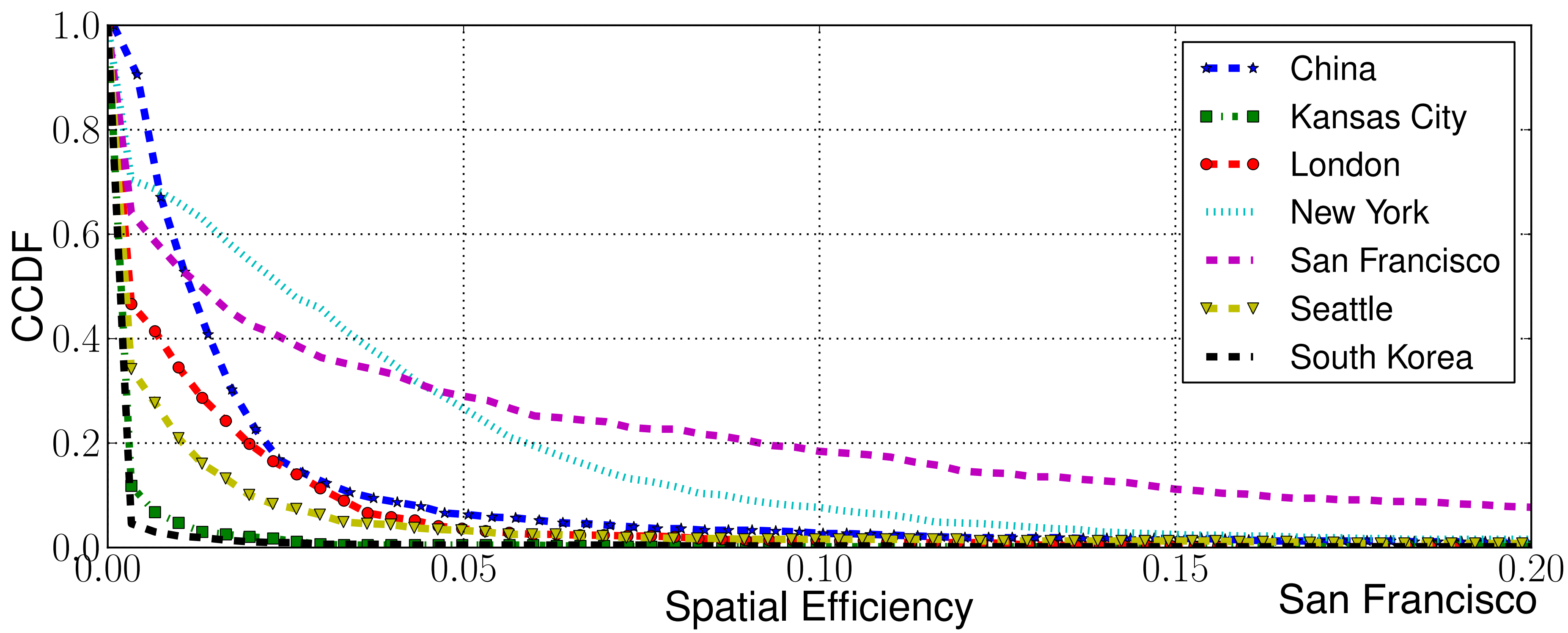
It can be thought of as a spatial extension of efficiency of traditional graphs.

Not defined if are coinciding!

$$C^E_{i,p^\star} = \frac{1}{k_i} \sum_{j \in \mathcal{N}_i} e^{-d_G(p_j, p^\star)/\gamma)}$$



from London



from San Francisco

Spatial Efficiency Centrality (Twitter)

High values of self-efficiency for Londoners. Distributions for SF more uniform.

UNIVERSITY OF BIRMINGHAM

Local Spatial Clustering Coefficient

$$C_{i,\mathcal{S}} = \frac{|\{e_{jk} \in E : j,k \in \mathcal{N}_{i,\mathcal{S}}\}|}{k_{i,\mathcal{S}}(k_{i,\mathcal{S}} - 1)}$$

It represents the fraction of users of *i* which for social triangles in the considered region *S.*

Nodes scoring high values are part of social circles in the region, making them potentially very influential.

Complexity

- All the defined metrics are *local*: no need to explore the whole graph.

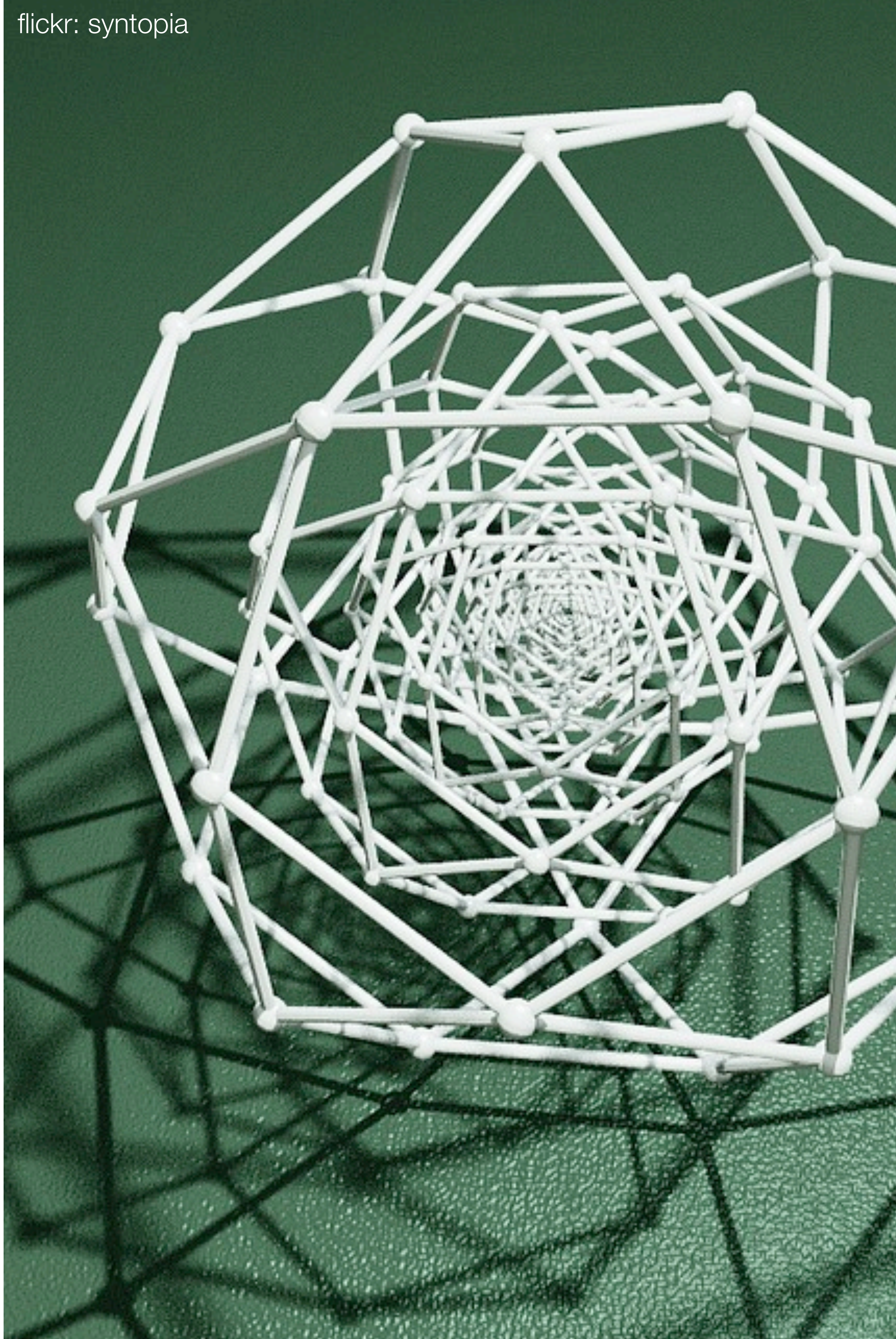
- Spatial degree/ratio/closeness centrality and spatial efficiency scale as

$$\mathcal{O}(nkt)$$

Local spatial clustering coefficient scales as

$$\mathcal{O}(nk^2t^2)$$

UNIVERSITY OF BIRMINGHAM



Complexity

- All the defined metrics are *local*: no need to explore the whole graph.

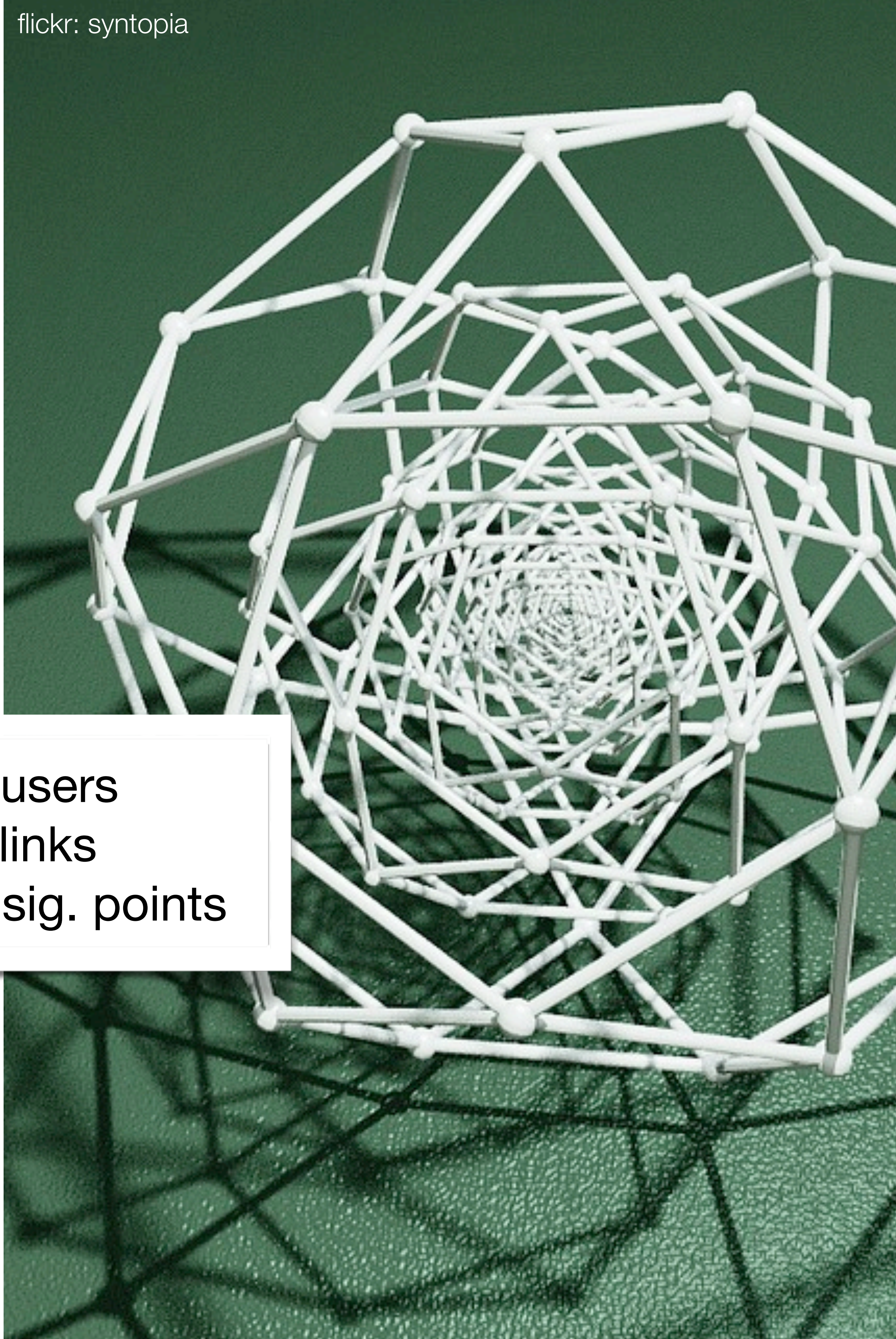
- Spatial degree/ratio/closeness centrality and spatial efficiency scale as

$$\mathcal{O}(nkt)$$

Local spatial clustering coefficient scales as

$$\mathcal{O}(nk^2t^2)$$

n	# users
k	# links
t	# sig. points



UNIVERSITY OF BIRMINGHAM

Future work

- We analysed *structural* properties, not processes dynamics (e.g. information diffusion).

- We plan to analyse processes happening on a network (e.g. retweets, mentions) and quantify the impact of spatial structure over these processes.

- We plan to explore real-time computation aspects.

UNIVERSITY OF BIRMINGHAM

Take-away Messages

Centrality metrics can be extended to measure spatio-social centrality.

Such metrics can be used to rank users according to their importance.

The presented metrics are local and scale well.

Thanks!

Questions?

Antonio Lima

a.lima@cs.bham.ac.uk
http://cs.bham.ac.uk/axl162
@themiurgo

UNIVERSITY OF
BIRMINGHAM