Structural and geographic properties of online social interactions

Yana Volkovich

Barcelona Media - Innovation Center

in collaboration with
A. Kaltenbrunner, D. Laniado, C. Mascolo, and S. Scellato

References

- Y. Volkovich, S. Scellato, D. Laniado, C. Mascolo, and A. Kaltenbrunner;
  “The length of bridge ties: structural and geographic properties of online social interactions”
  ICWSM-12 (International AAAI Conference on Weblogs and Social Media)
- A. Kaltenbrunner, S. Scellato, Y. Volkovich, D. Laniado, D. Currie, E. J. Jutemar, and C. Mascolo;
  “Far from the eyes, close on the Web: impact of geographic distance on online social interactions”;
  WOSN ’12 (ACM SIGCOMM Workshop on Online Social Networks)

online social connections:

- *explicit* (articulated)
  e.g. friendship connections
- *implicit* (behavioural)
  e.g. interactions

social graph: nodes and edges

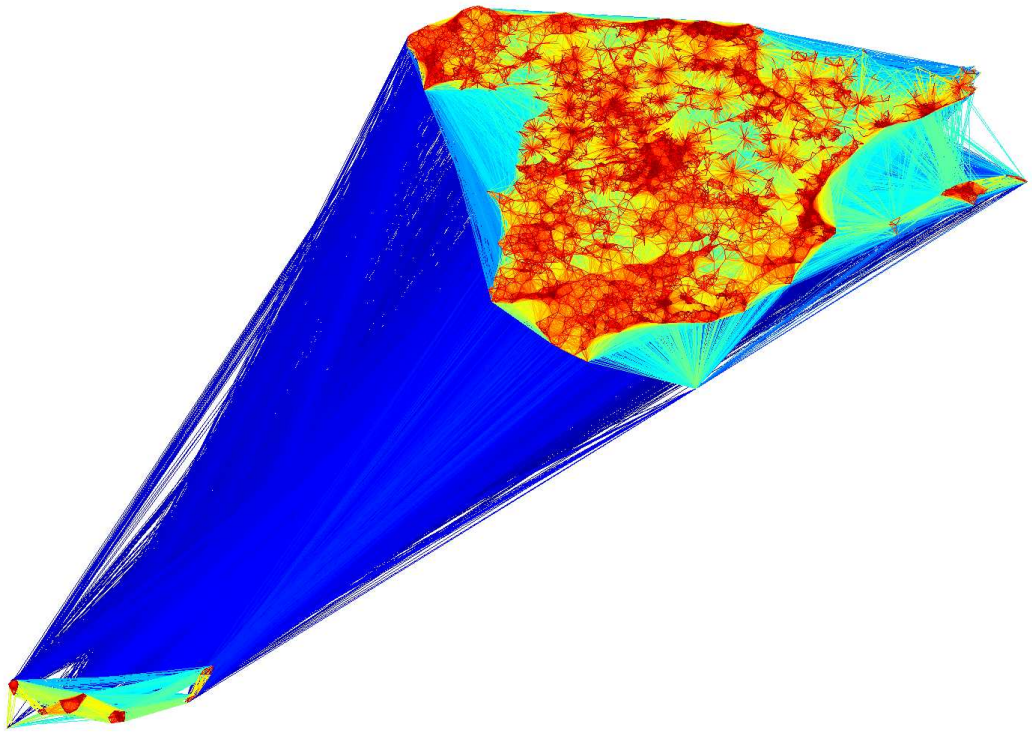
- connections could be more informative than nodes
- different types of social connections
- different ways to characterize social connections

different ways to characterize social connections

- interaction strength
- spatial distance
- structural position in a social graph

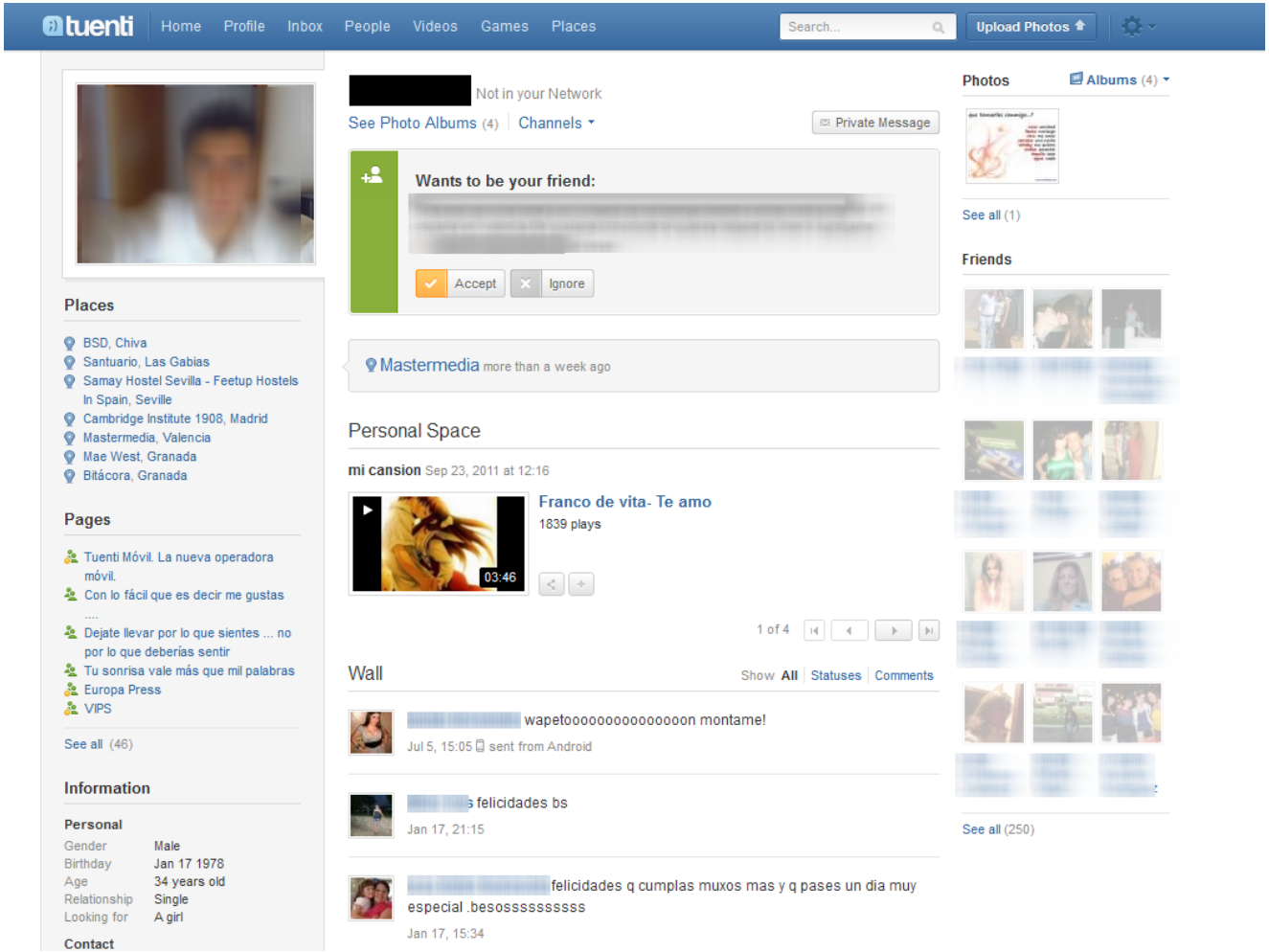
Dataset

Tuenti is the “Spanish Facebook”
a Spain-based, invitation-only social networking website



Tuenti

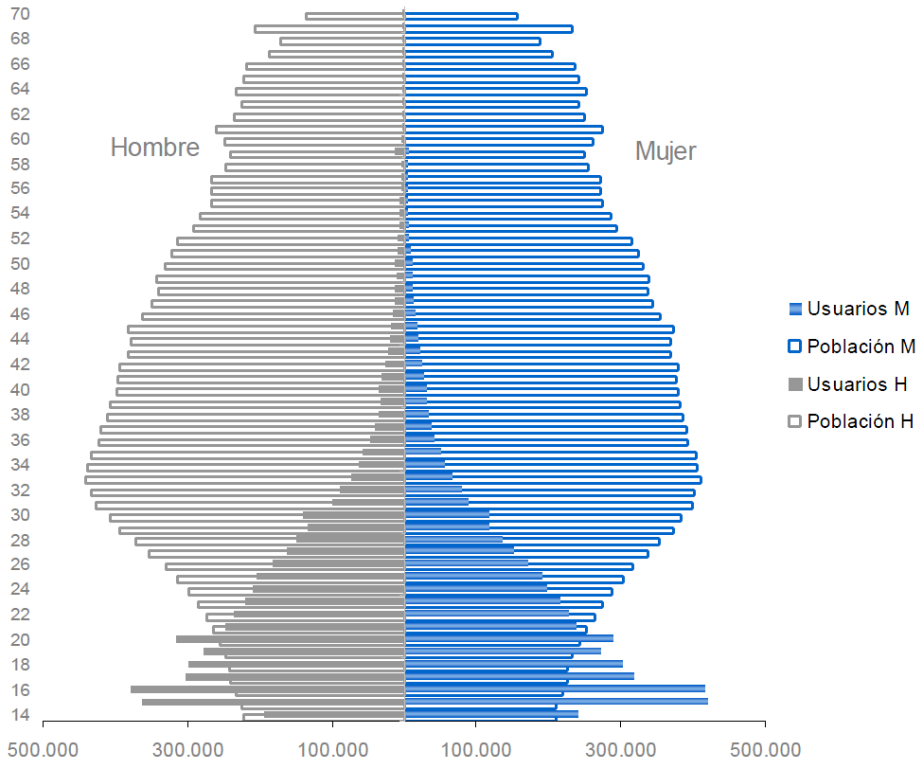
Tuenti website



Tuenti dataset:

- by Dec. 11, 2010;
- 9.88 million registered users (anonymous profiles);
- more than 1 174 million friendship links;
- 500 million messages exchanged during 3 months;

- age pyramid


70
68
66
64
62
60
58
56
54
52
50
48
46
44
42
40
38
36
34
32
30
28
26
24
22
20
18
16
14
Hombre
Mujer
500.000
300.000
100.000
100.000
300.000
500.000
Usuarios M
Población M
Usuarios H
Población H

by gender

- 50.6% female;
- 49.4% male.

by age (average)

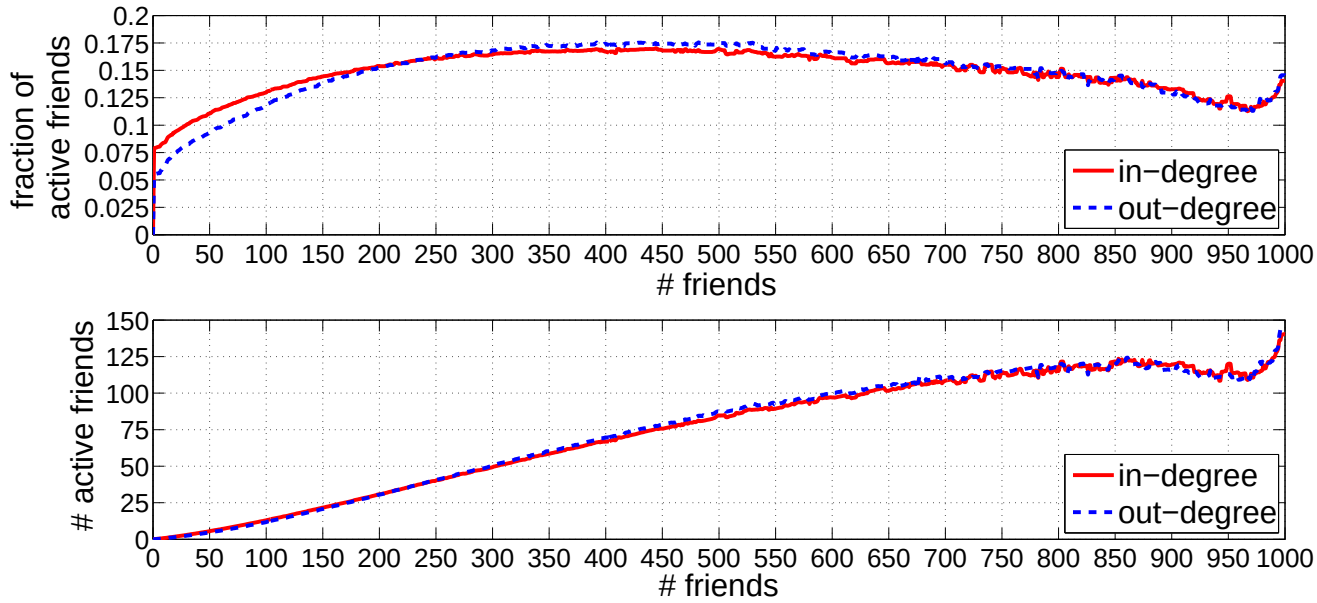
- female: 22 years;
- male: 28 years.

Tuenti users are very young

- 45% of users are between 14 and 20 years;
- 37.5% of users are between 21 and 30 years.
- 1.35 more teenagers than official population (due to Tuenti signing requirements).

implicit vs. explicit social connections

- Dunbar's number: an alleged theoretical cognitive limit to the number of people with whom one can maintain stable social relationship
- average fraction of friends and the average absolute number of friends a user interacts with as a function of the number of friends



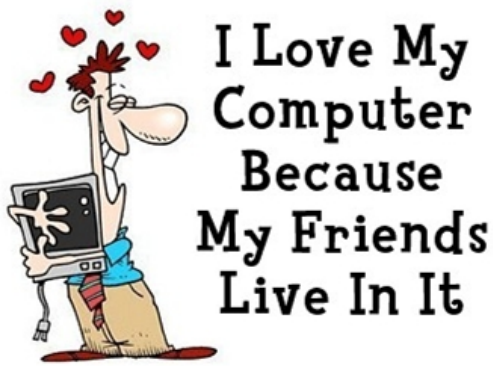
Characteristics for social connections

social ties and spatial distances:

- individuals try to minimize the efforts to maintain a friendship by interacting more with their spatial neighbors
- probability of a social interaction quickly decays as an inverse power of the relative geographic distance (Stewart [1941])

Social connections

spatial distance, related work



- online tools and long-distance travel might result in the ‘death of distance’
- probability of social connection between two individuals on online social networking services still decreases with their geographic distance (Backstrom et al. [2010], Liben-Nowell et al. [2005]).

spatial distance

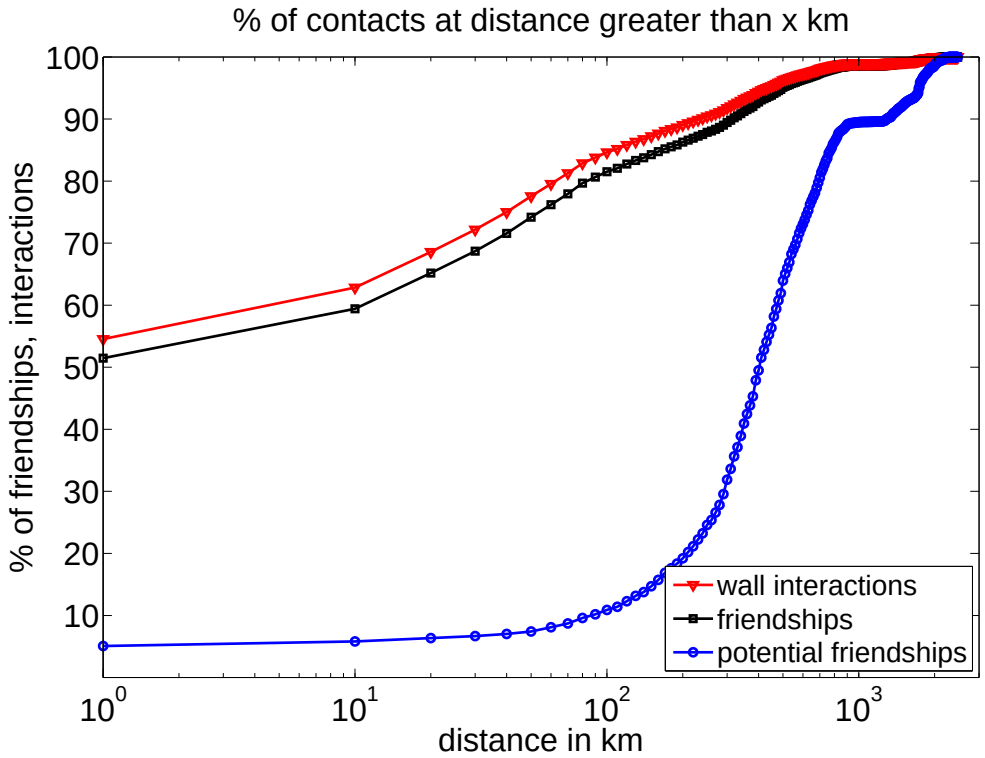
- $d_{i,j}$ is the geographic distance between the cities of residence of user i and user j;
  $d_{i,j} = 0$ if users report the same city of residence
- average geographic distances between users $< D >$ is about one order of magnitude larger than the average geographic distance between friends $< l >$

average geographic distance between nodes, km	531.2
average link length, km	79.9

Social connections

spatial distance

- spatially closer users are much more likely to engage in a social connection (e.g. become friends)
- about 50% of social links between users at a distance of 10 km or less



interaction strength

- close friends or just acquaintances
- quantitative estimation of a how much an online connection binds two users together

interaction strength

- $w_{i,j}$ is the number of messages user *i* posted on the wall of user *j*;
- $w_{i,j} = 0$ if user *i* has never left a message on user *j*'s wall;

balanced interaction weight:

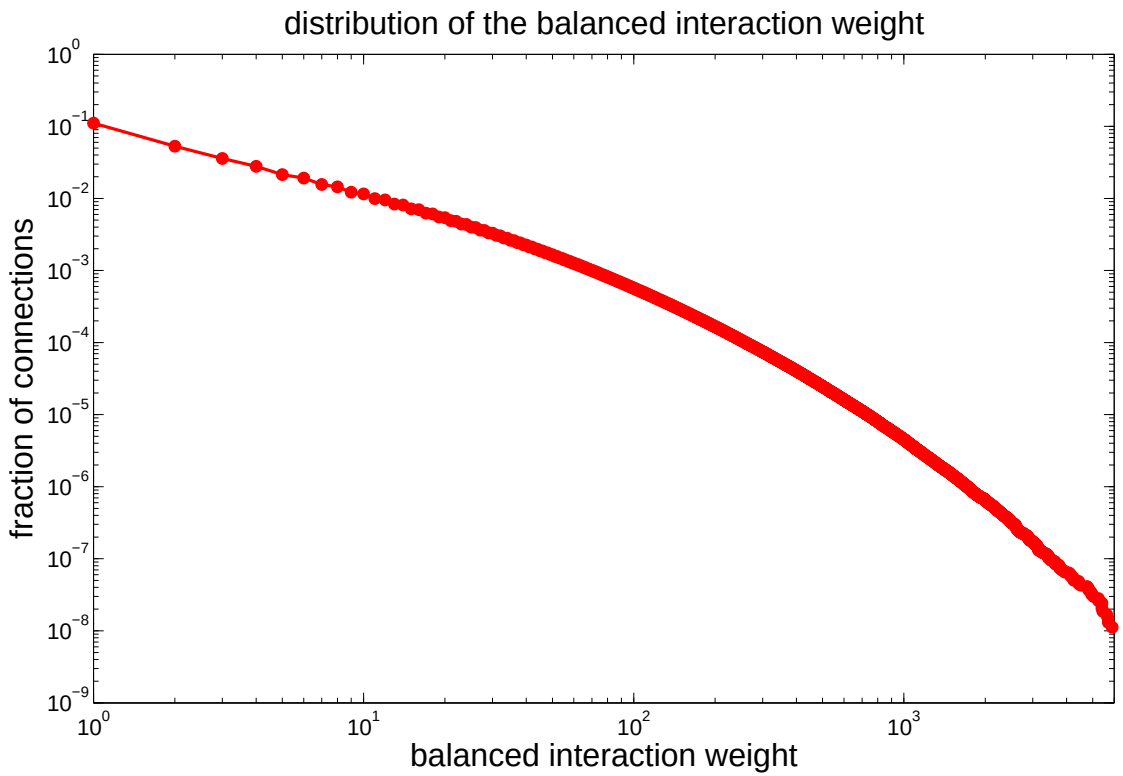
$$\bar{w}_{i,j} = \min(w_{i,j}, w_{j,i}) + (1 - \delta_{w_{i,j},w_{j,i}})/2,$$

Social connections

Interaction strength (log-log)

since non-reciprocated interactions may indicate spam:

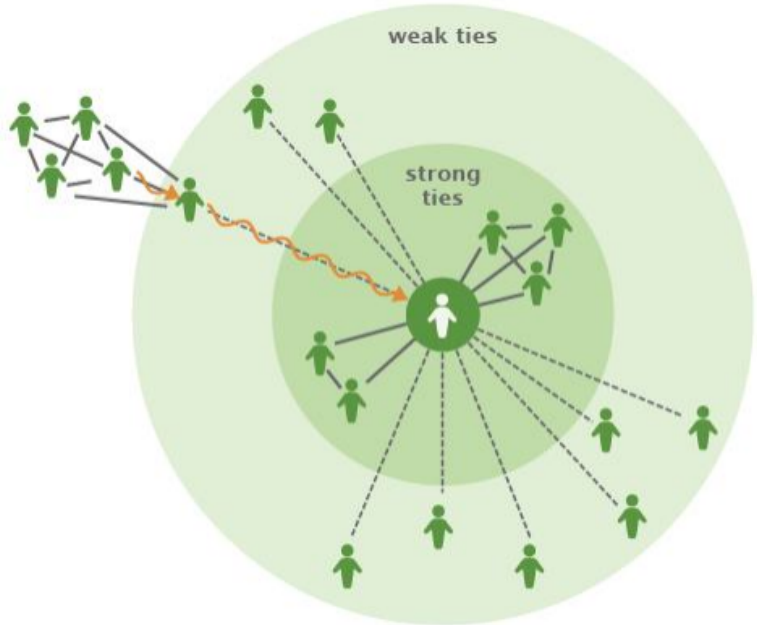
- the minimum of the interaction weights to emphasize reciprocated interactions;
- for the non-reciprocated interactions we only add 1/2 no matter the difference in the numbers of messages exchanged.



Social connections

structural properties

- weak ties are more likely to connect together otherwise separated portions of a network, playing an important role in information diffusion and resilience to network damage (Granovetter [1973])
- some social ties closing “structural holes” can be more powerful or more innovative (Burt [1992])



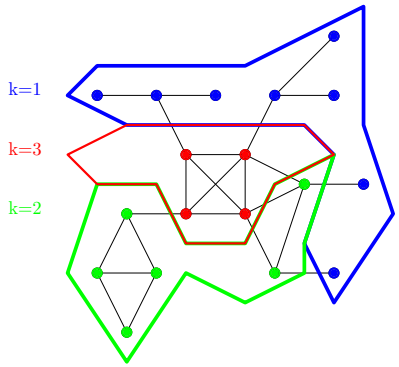
Bakshy [2012]

structural properties:

- **local position:** social overlap;
- *social overlap* of an edge $e_{i,j}$ as $o_{i,j} = |\Gamma_i \cap \Gamma_j|$, where Γ_i is the set of users connected to user i

structural properties:

- **global position:** k-index;
- *k*-core is the maximal subgraph in which each node is connected to at least *k* other nodes of the subgraph
- *k*-index of a node is *v* if it belongs to the *v*-core but not to the $(v+1)$-core
- *k*-index has been found to be an indicator of influential nodes within a social network (Kitsak et al. [2010])

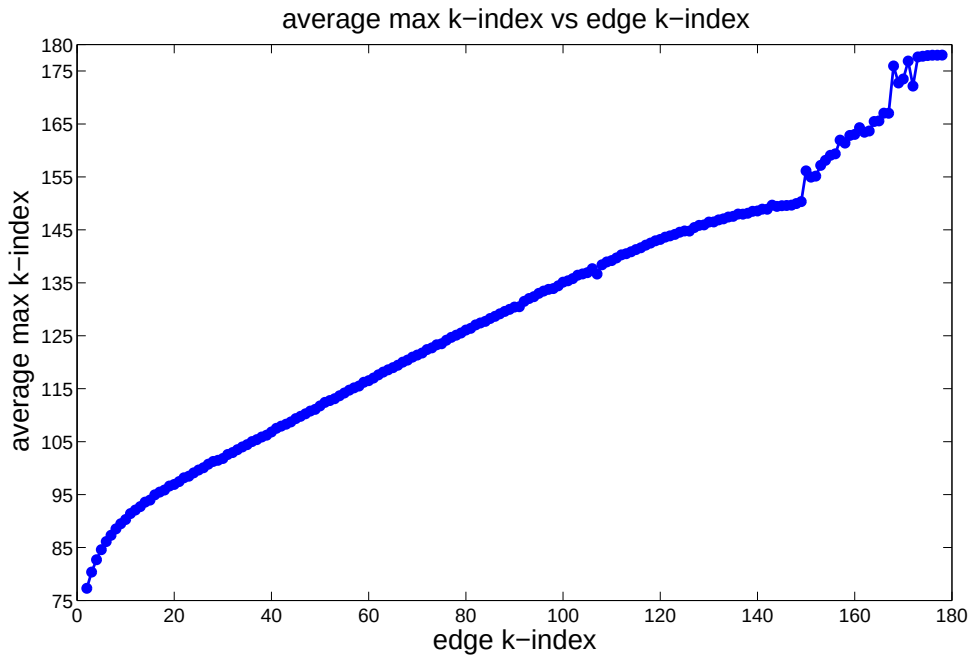


- central core/ smaller core in between/ periphery

Social connections

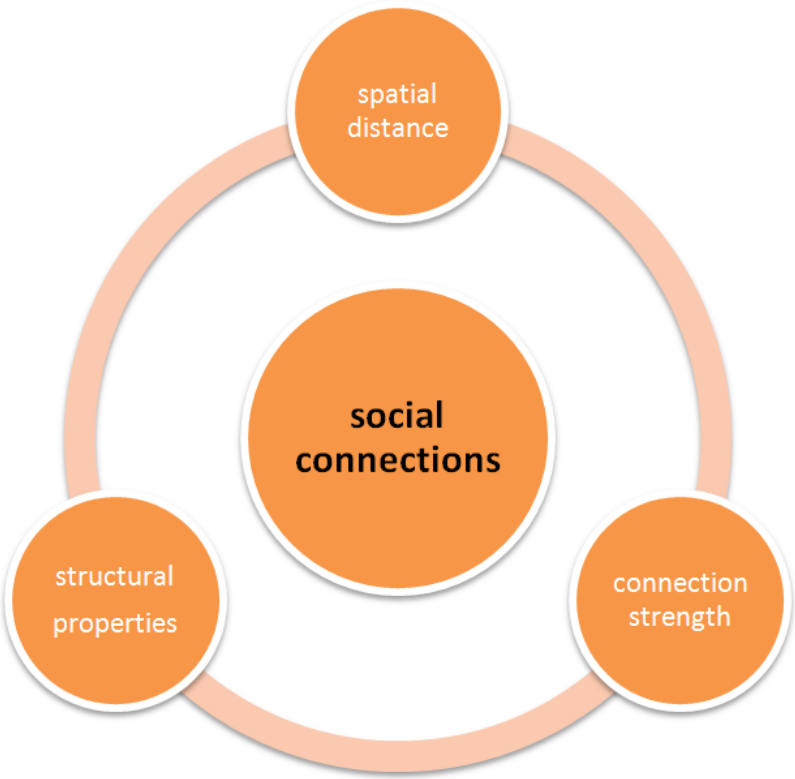
Structural properties:k-index of an edge

- *k-index* k_{ij} of an edge is the minimum of the *k*-indexes of two endpoints
- we distinguish if an edge connects nodes inside a network core or links to a node in the periphery



Combined analysis of social connections

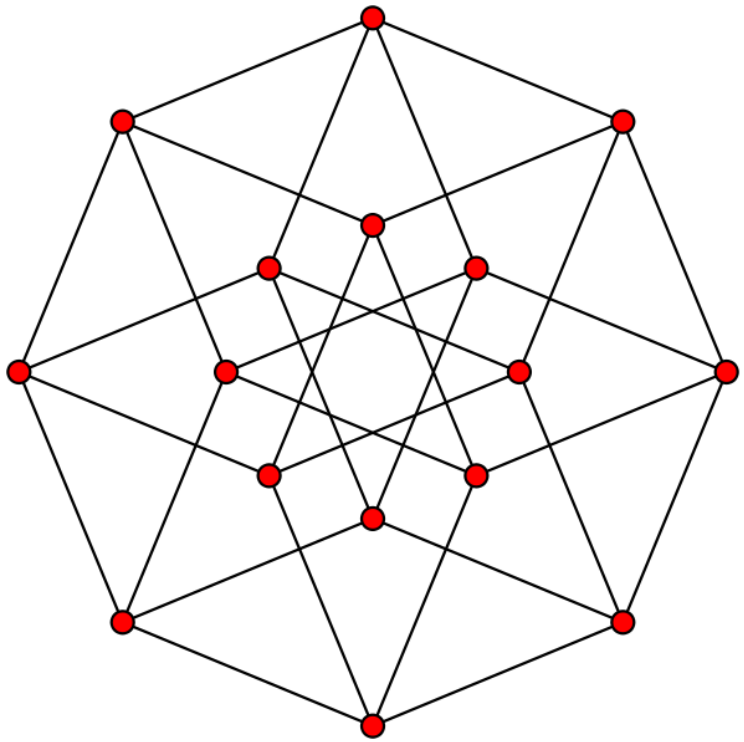
social connections



Combined analysis

Social overlap vs. k-index

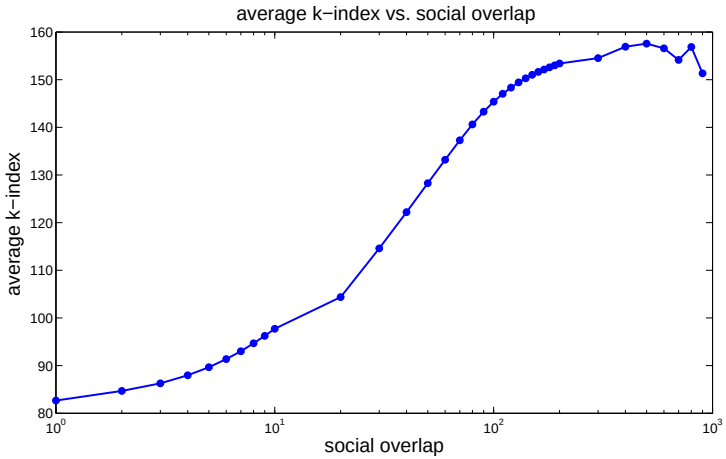
social overlap and *k*-index allow network scenarios where links may have high *k*-index and low overlap, or the other way round

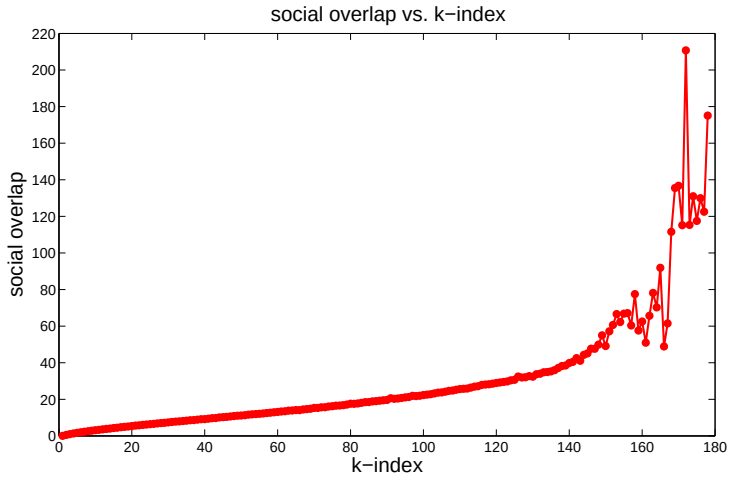


Combined analysis

Social overlap vs. k-index

- social overlap $\uparrow \Rightarrow$ *k*-index grows quickly
- *k*-index $\uparrow \Rightarrow$ the average social overlap grows slowly
- *there are inner cores where users are tightly connected to each other*
- *other parts of the network include more isolated users that tend to not belong to any community*

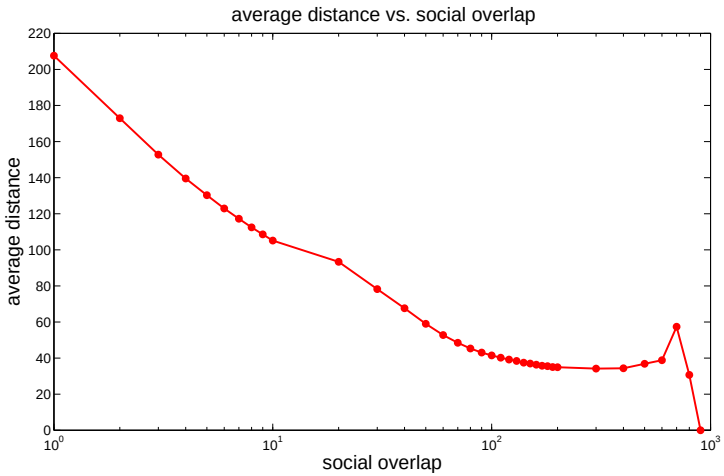


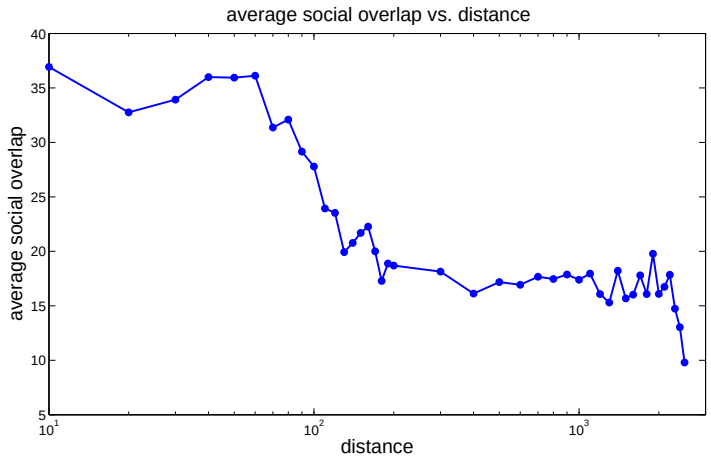


Combined analysis

Distance vs. social overlap

- the geographic distance between two connected users decreases as they share more and more friends
- social connections which span less than 60-80 km exhibit higher values of social overlap

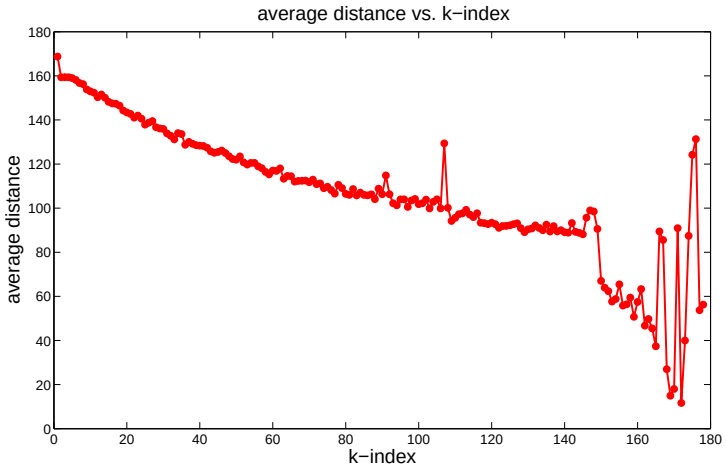


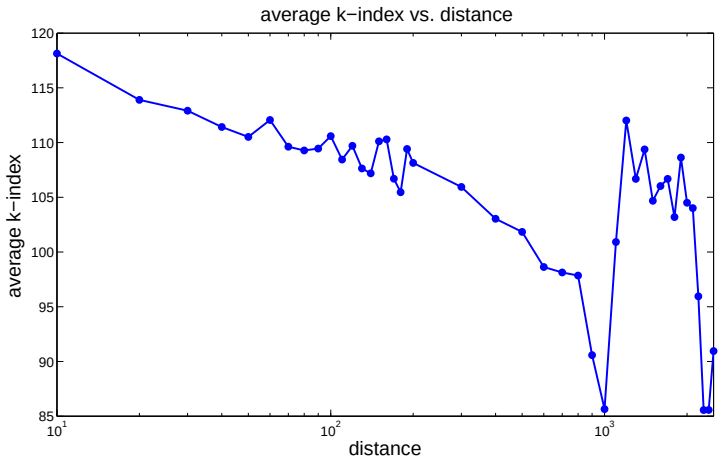


Combined analysis

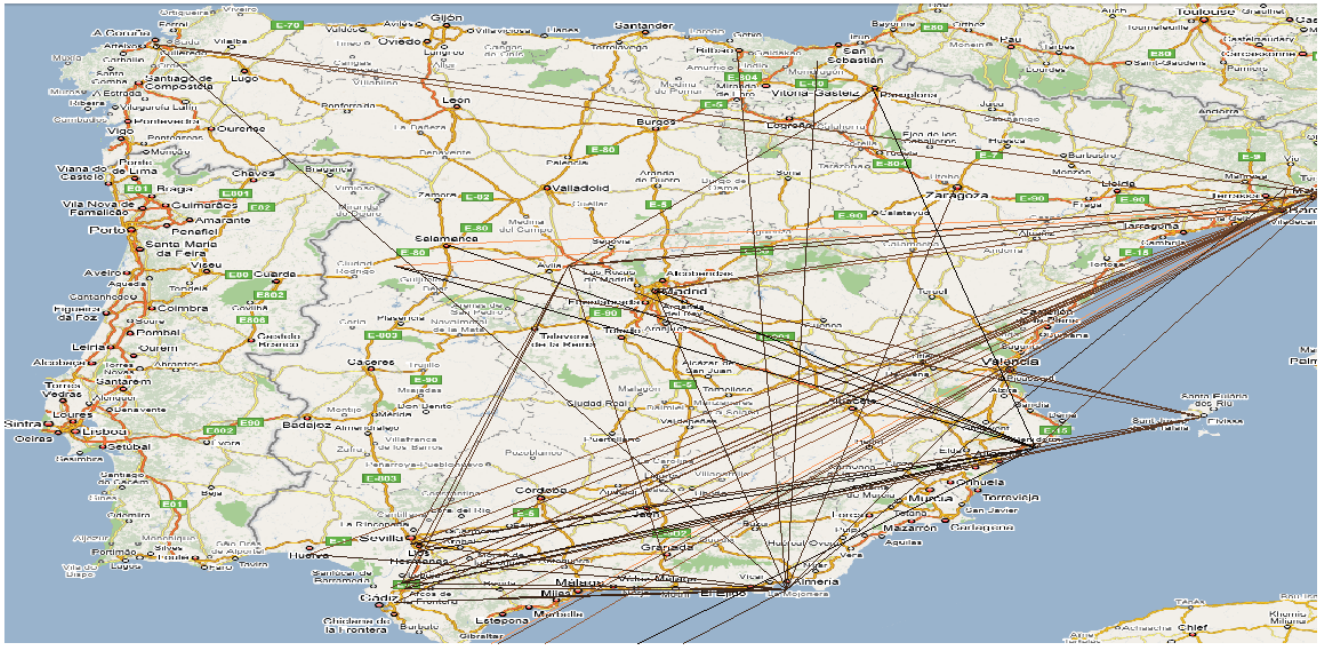
Distance vs. *k*-index

- the average spatial length of social links decreases as their *k*-index increases
- social links inside the core tend to be shorter than the ones reaching the periphery of the social network





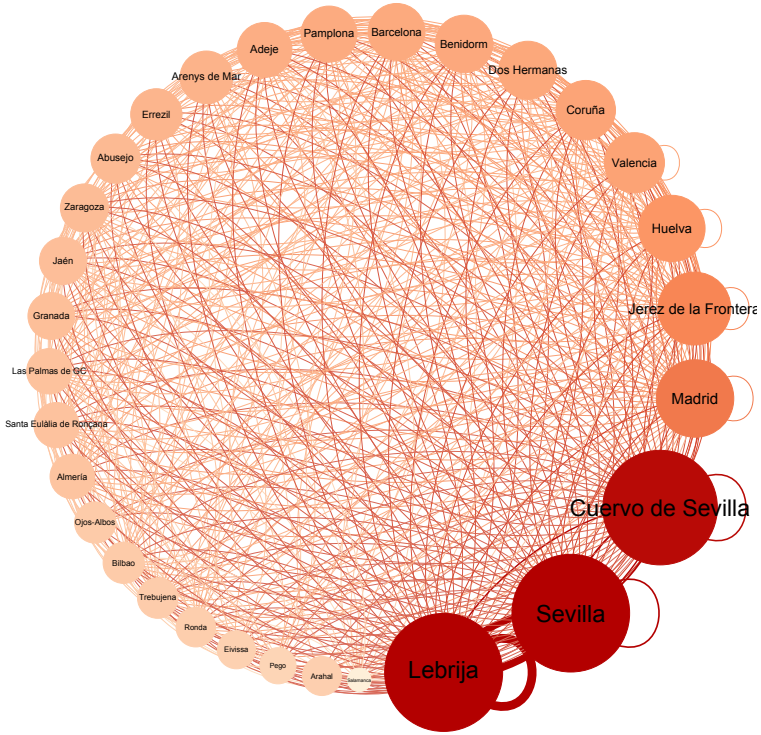
k_{max}-core


Gijón
Santander
Oviedo
Lugo
León
Burgos
Valladolid
Salamanca
Madrid
Zaragoza
Cáceres
Badajoz
Córdoba
Sevilla
Granada
Málaga
Almería
Murcia
Valencia
Toulouse

Combined analysis

Distance vs. *k*-index

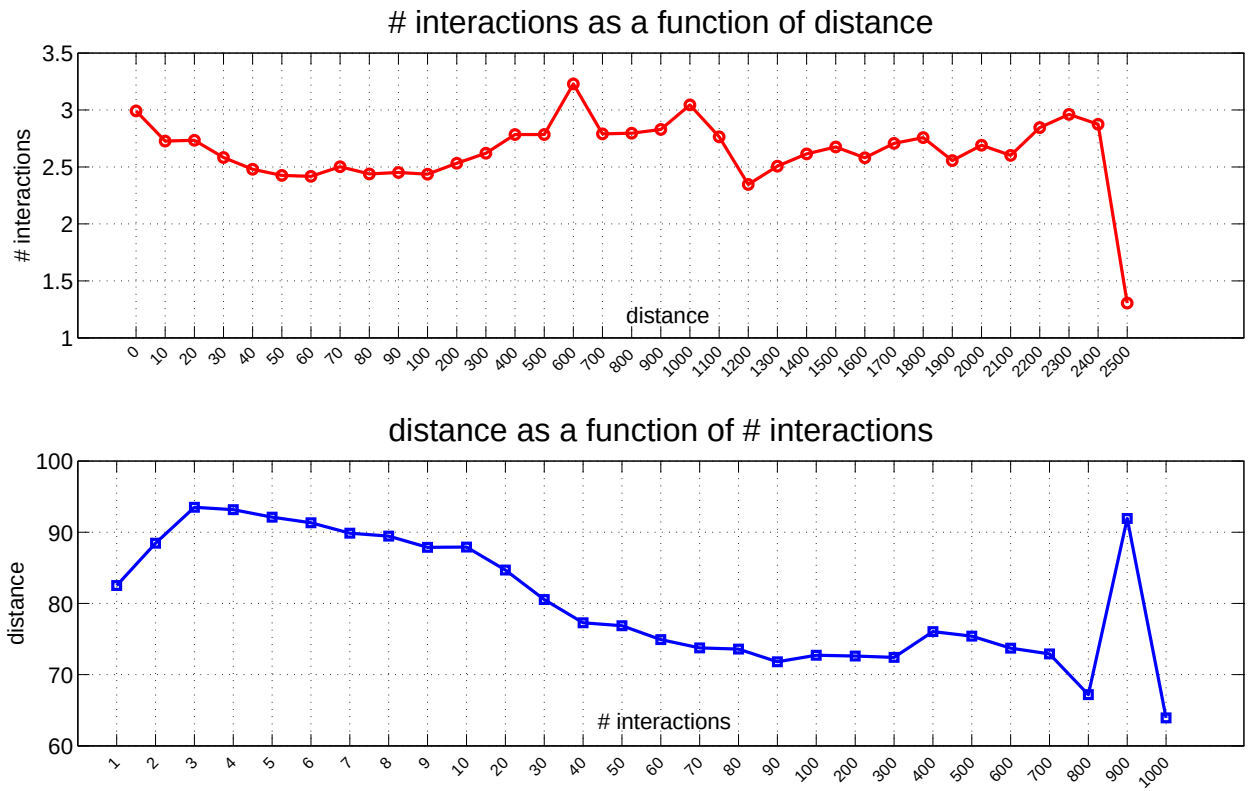
k_{max}-core



Combined analysis

Distance vs. interaction weight

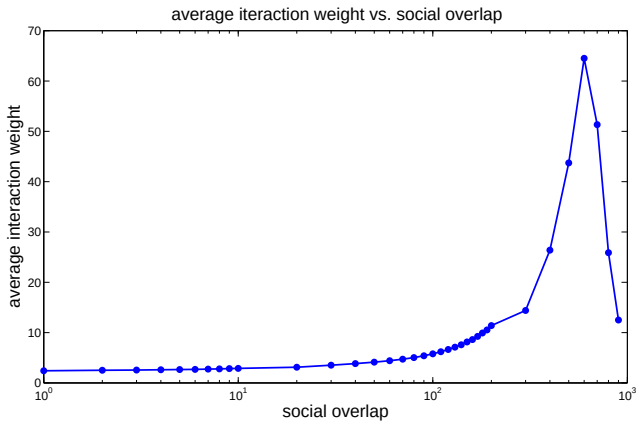
- *the amount of interaction is uncorrelated to spatial distance*
- note that the likelihood that two individuals are connected is heavily dependent on distance

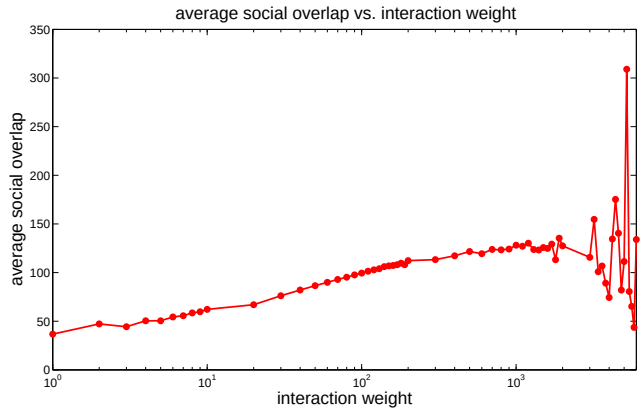


Combined analysis

Social overlap vs. interaction

- *the impact of social overlap remains fairly constant*
- the interaction weight only slowly increases the social overlap grows
- the extremely high levels of interaction mainly take place between users with several shared friends, which are likely to be in the network core

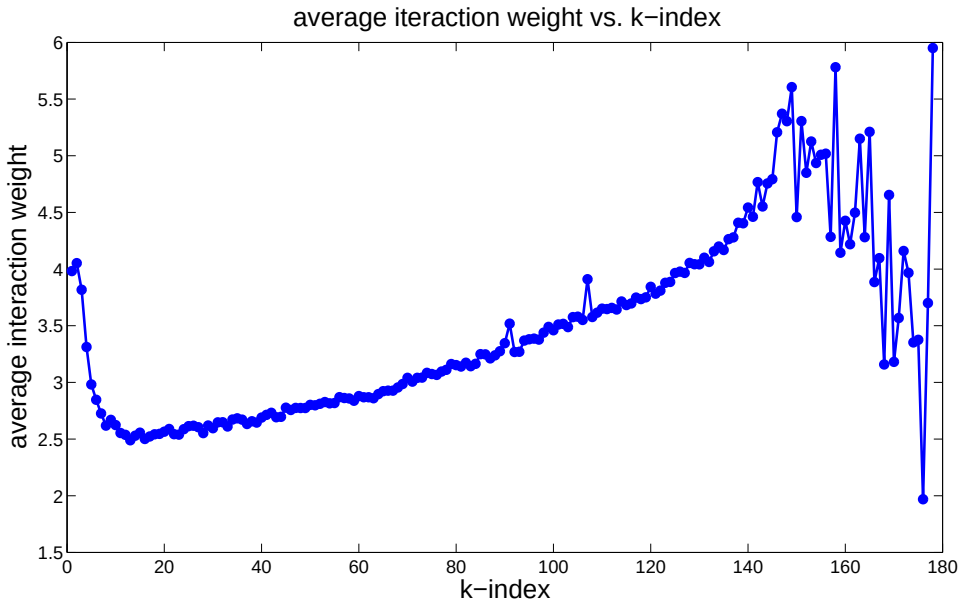




Combined analysis

k-index vs. interaction weight

- ties in the inner cores have the highest levels of interaction
- interaction weights are almost equally high for social ties with low *k*-index
- social ties with intermediate *k*-index, likely to bridge together different portions of the network, experience the lowest interaction levels



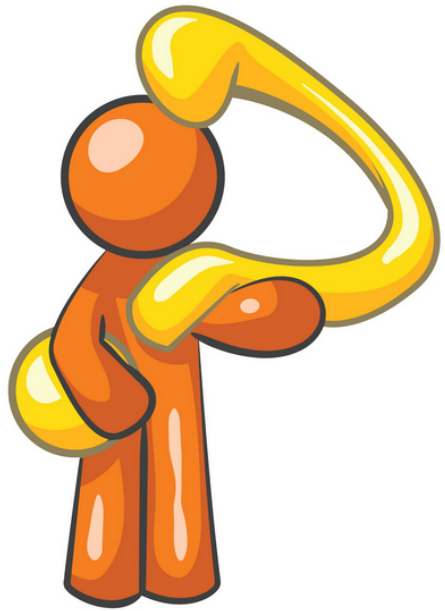
Conclusions

Conclusions

- social connections between users inside the core tend to have shorter geographic spans than connections stretching outside the core
- social ties outside the core tend to be much longer than the other links: the length of these bridge ties is thus creating not only network shortcuts, but also spatial shortcuts
- the amount of interactions appears independent of spatial distance
- interaction levels appear higher inside well-connected cores and on links connecting to the fringe of the network
- **edges could be more informative than nodes**

Questions

Questions



Bibliography I

L. Backstrom, E. Sun, and C. Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of WWW 2010*, Raleigh, North Carolina, USA, 2010.

E. Bakshy. Rethinking information diversity in networks, 2012. URL `www.facebook.com/notes/facebook-data-team/rethinking-information-diversity-in-networks/10150503499618859`.

R. S. Burt. *Structural holes: The social structure of competition*. Harvard University Press, Cambridge, MA, 1992.

M. S. Granovetter. The strength of weak ties. *The American Journal of Sociology*, 78(6):1360–1380, 1973. doi: 10.2307/2776392.

M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, Nov. 2010. URL `http://dx.doi.org/10.1038/nphys1746`.

Bibliography II

D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *PNAS*, 102(33): 11623–11628, Aug. 2005.

J. Q. Stewart. An inverse distance variation for certain social influences. 93(2404):89–90, 1941.