Consiglio Nazionale delle Ricerche

FROM EGO NETWORKS TO ONLINE SOCIAL NETWORKS

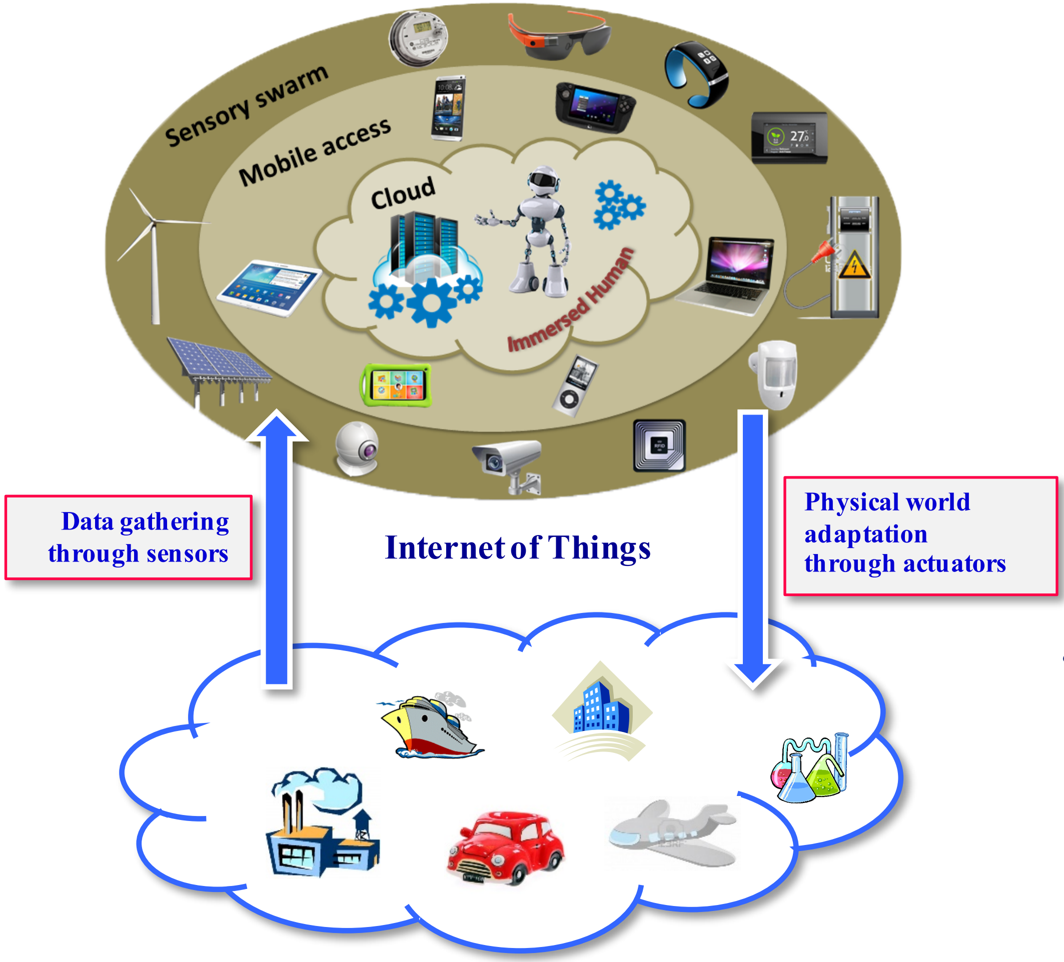
Marco Conti

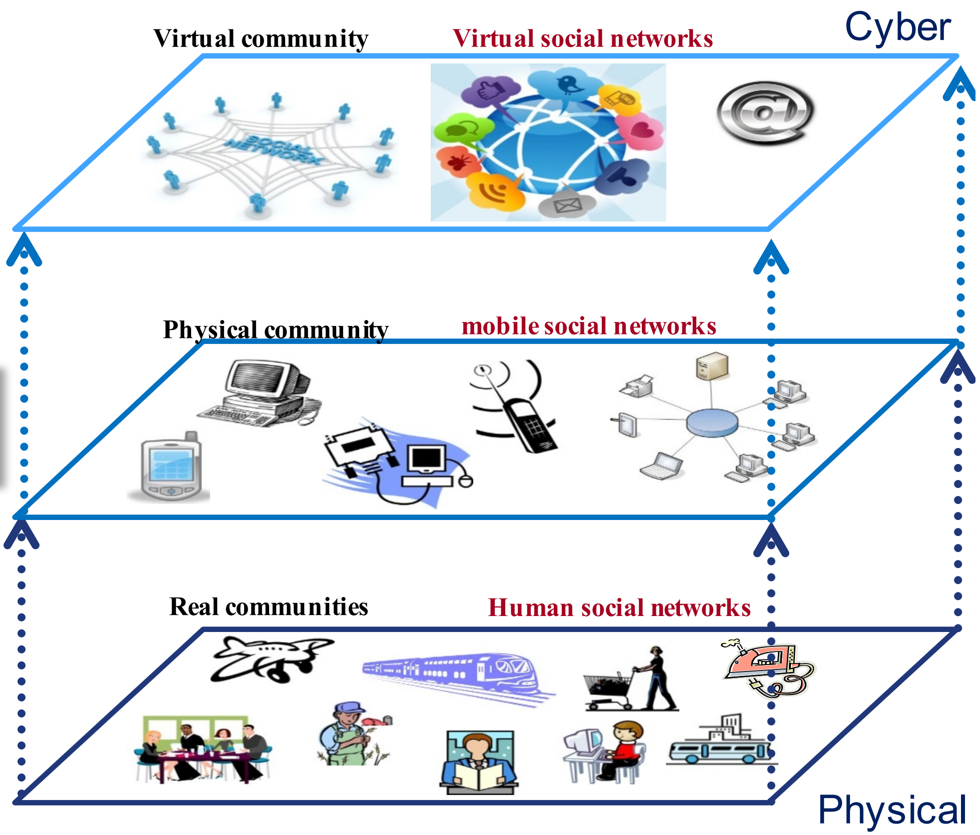
marco.conti@cnr.it

Istituto di Informatica e Telematica

Ubiquitous Internet Lab

CYBER-PHYSICAL CONVERGENCE





Smart City

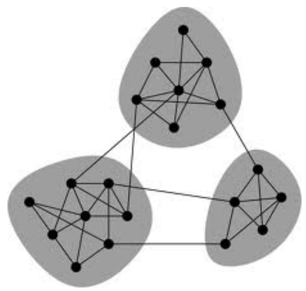
Google Self Driving Car

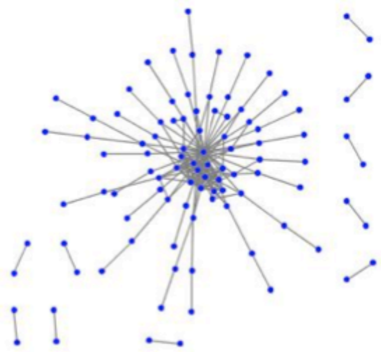
Industry 4.0

Social Networks: the Structural Properties of the Social Graph

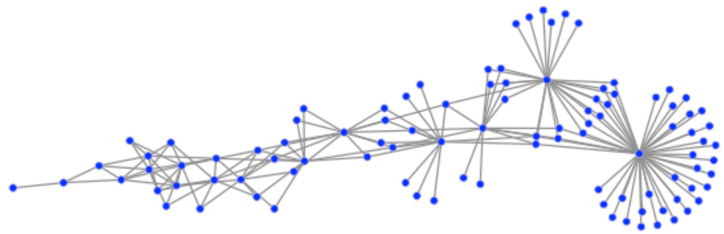
- Not about the analysis of content exchanged over the OSN

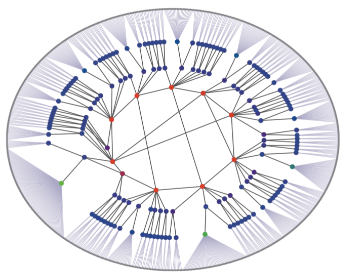
- Just about the structures in the graph
  - Presence of communities
  - Presence of hubs
  - "Connected core" of nodes
  - …

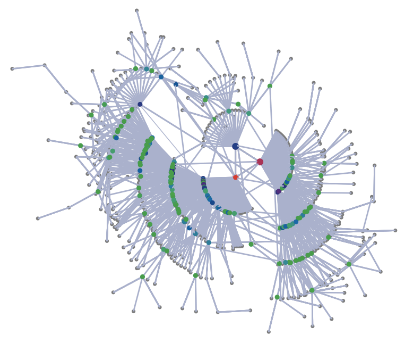




- Many phenomena occurring in OSN are determined by these structures
  - Information diffusion
  - Vulnerability to attacks and viruses
  - Transitivity of trust
  - Influence over people
  - Opinion dynamics
  - Recommendations effectiveness
  - ...

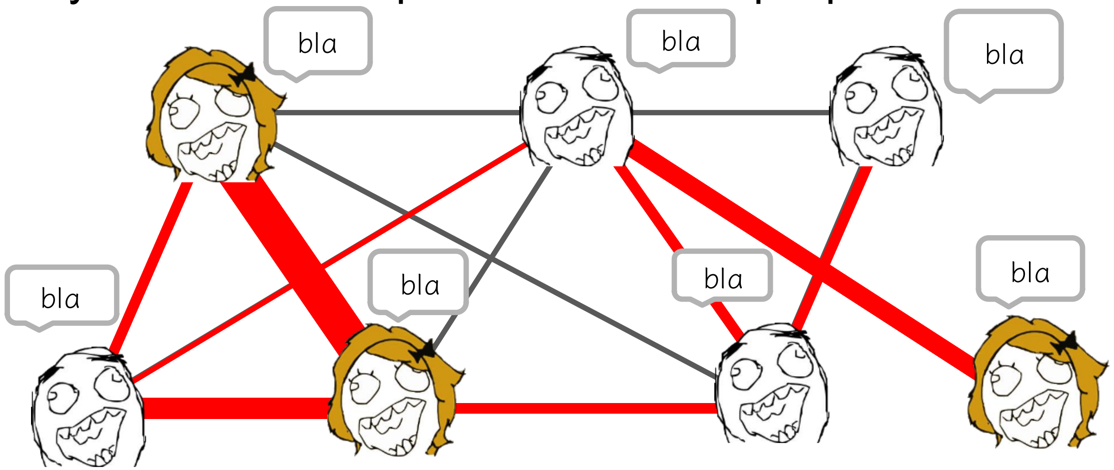






Social Graph vs Interaction Graph

- Interaction graph is about classifying the importance of different social links
- The mere existence of links is not sufficient to say whether that is a "strong" or a "weak" tie
  - We need to analyse interactions patterns between people



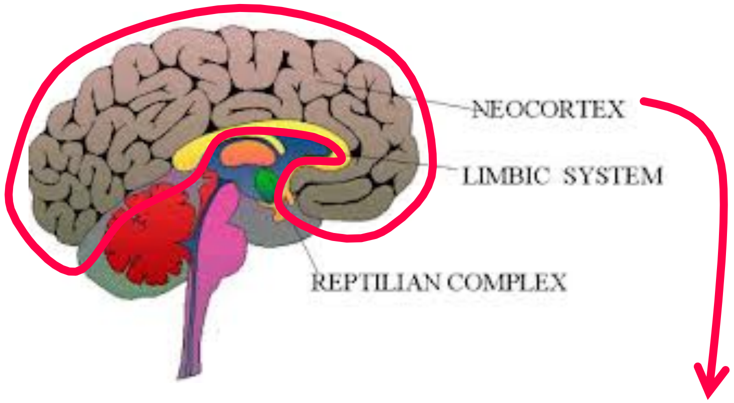
- Key impact on various aspects:
  - Spread of information (information drop for social distant friends)
    - Actual message spread might be less "small world" than expected
  - Trustworthiness of information
    - Trust in information received is a function of the path taken in the OSN

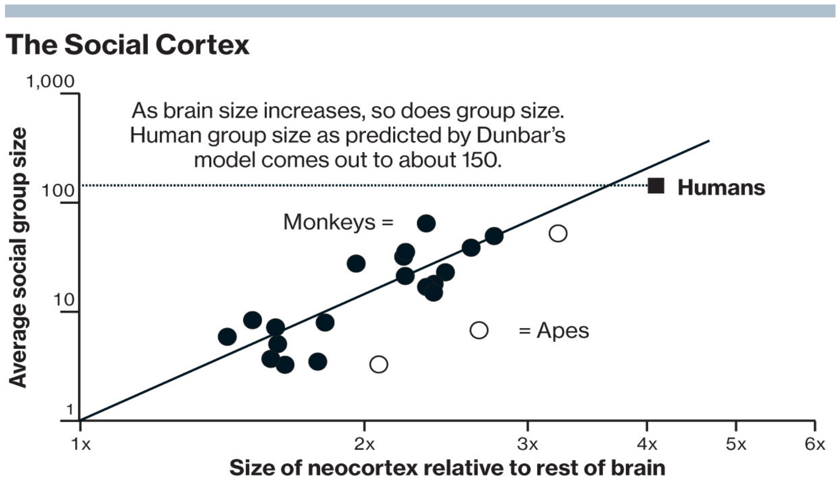
Macroscopic vs microscopic perspectives

THE SOCIAL BRAIN HYPOTHESIS

Our social capacity is bounded by a combination of the size of the human brain and of the limited time that can be allocated to the social relationships

The need to establish larger and larger social networks required more resources and thus bigger brains





- The structure of human ego networks is the result of cognitive and time constraints on our social capacity
- Mammals with bigger brains (neocortex) live in larger groups and thus have larger ego networks
- Humans, as predicted by Dunbar's Number, have an average of 150 active social relationships

EGO NETWORKS (ANTROPOLOGY): DUNBAR'S MODEL

The strength of relationships (or emotional closeness) is not evenly distributed, but forms a hierarchy of social "circles"

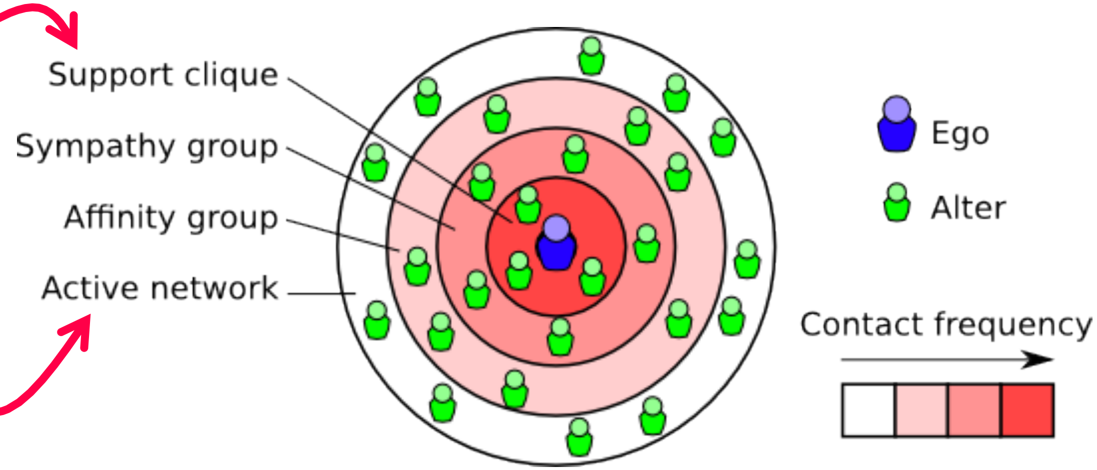
- One person (ego), and their social relationships
- Dunbar's model

sizes
~5
~15
~45
$\boxed{\sim 150}$

intimacy

intimate friends

acquaintances




WIKIPEDIA
The Free Encyclopedia

Main page
Contents

Article Talk

Dunbar's number

From Wikipedia, the free encyclopedia

See also: Dunbar (disambiguation)

Dunbar's number is a suggested cognitive limit

Measured sizes (Scaling ratio ~ 3)

Support clique	Sympathy group	Affinity group	Active network
4.6	14.3	42.6	132.5

OSN: Social Network in The Cyber World

- Online Social Networks are the largest-scale social interaction tool we ever had

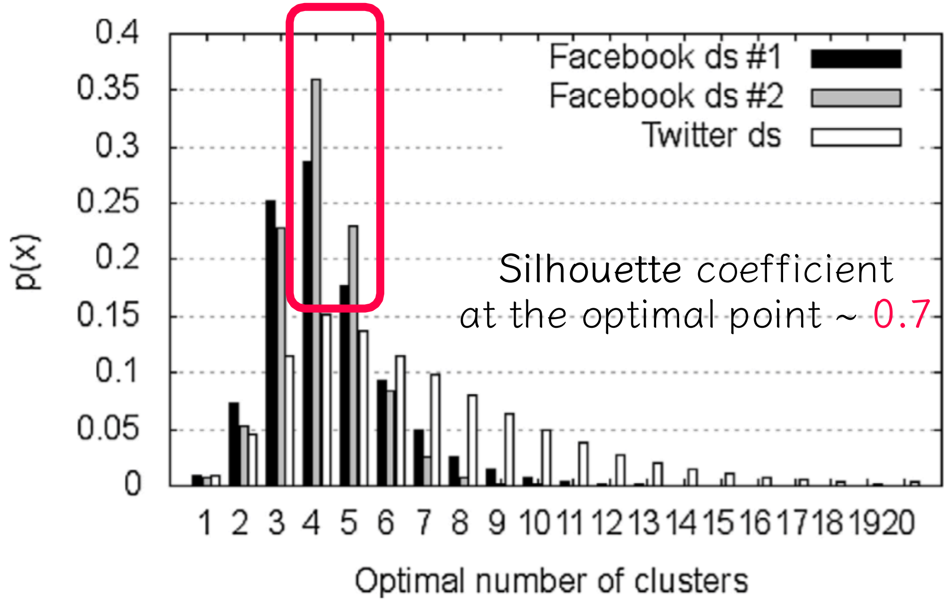
- Do they change the structure of ego networks?
  - Hypotheses
    - Yes, they do! because they offer an unprecedented rich tool for communication
    - No, they don't! because they are just another means of social interaction
- What impact do OSN ego networks have on properties (such as information diffusion) in OSN?
- Data Science can help us answering these questions

Ego Network Structure in OSN: a Data Study

- Mining Facebook and Twitter (through public API) to get the data
- Datasets
  - Facebook #1: 3M nodes, 23.5M edges
  - Facebook #2: 90K nodes, 3.6M edges (regional network of New Orleans)
  - Twitter: 300K nodes, 7.6M edges
    - With interaction events (post, likes, ...)
- Method
  - Evaluate contact frequency between egos and their alters
    - Contact frequency = standard proxy for intimacy and tie strength
  - Cluster contact frequencies between each ego and their alters
    - to see if there are groups
  - Average the size of the groups across egos

Ego Network Structure: Number of Groups

- Do groups exist at all?
  - Optimal number of clusters peaks at 4 or 5 (using the Akaike Information Criterion)
  - High Silhouette coefficient means clustering at the optimal point is not artificial



There is a "natural" grouping in OSN ego networks,
and the number of groups is similar to the Dunbar model

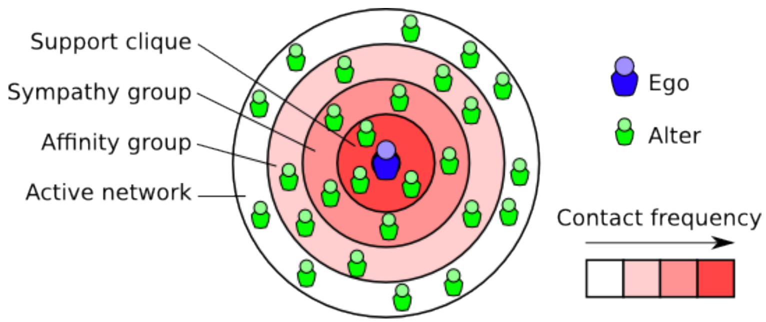
EGO NETWORK STRUCTURE: SIZE OF THE GROUPS

- Facebook #1, optimal clustering @ 4 groups

LAYER	0	1	2	3	4
AVG. SIZE	1.68	5.28	14.92	40.93	---
Dunbar's Model	---	4.6	14.3	42.6	132.5

- Twitter, optimal clustering @ 4 or 5 groups (below with 5)

LAYER	0	1	2	3	4
AVG. SIZE	1.55	4.52	14.92	28.28	88.31
Dunbar's Model	---	4.6	14.3	42.6	132.5



EGO NETWORK STRUCTURE: SIZE OF THE GROUPS

Facebook

LAYER	0	1	2	3	4
AVG. SIZE	1.68	5.28	14.92	40.93	---
Dunbar's Model	---	4.6	14.3	42.6	132.5

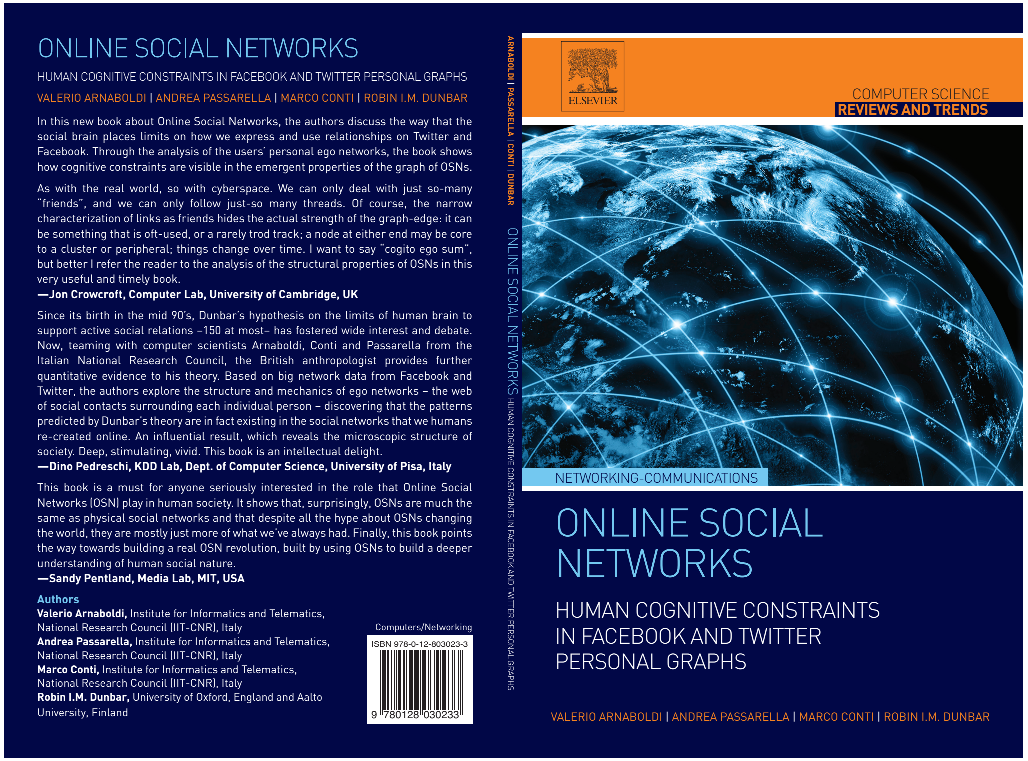
Twitter

LAYER	0	1	2	3	4
AVG. SIZE	1.55	4.52	14.92	28.28	88.31
Dunbar's Model	---	4.6	14.3	42.6	132.5

- Size of the layers is quite similar for internal layers
- External layers are less evident
  - Facebook: dataset is from 2009
  - Twitter: mixed use of external layers
- Layer 0
  - a "newcomer": not present in the "standard" Dunbar's model
  - Long-lasting hypothesis in antropology
    - "*there should be a best-friend layer inside the support clique, of size approximately equal to 2*"
    - Never measured in literature, due to the limited size of available datasets
  - Using OSN, we have possibly provided for the first time empirical evidence about this sociologically hypothesis
    - **OSN as a microscope for studying human social relationships**

ONLINE SOCIAL NETWORK PROPERTIES

- Do OSN change the structure of ego networks?
  - Hypotheses
    - Yes, they do! because they offer an unprecedented rich tool for communication
    - No, they don't! because they are just another means of social interaction
- Can we find other properties of ego networks thanks to availability of Big Data?
  - "Best-friend" layer measured thanks to large-scale OSN datasets



V. Arnaboldi, A. Passarella, M. Conti, R.I.M. Dunbar Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs Elsevier, October 2015

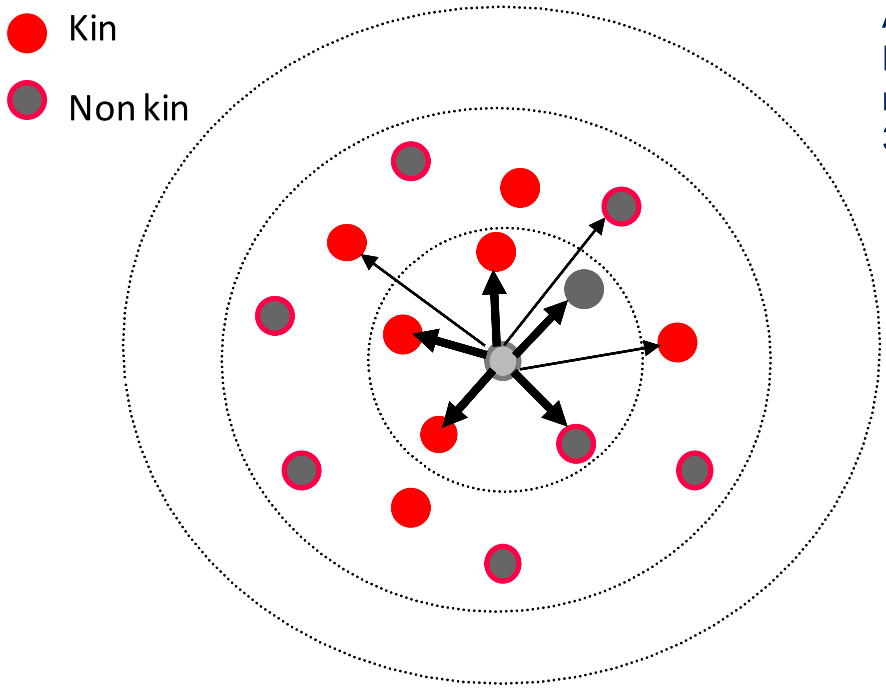
EGO NETWORK APPLICATIONS

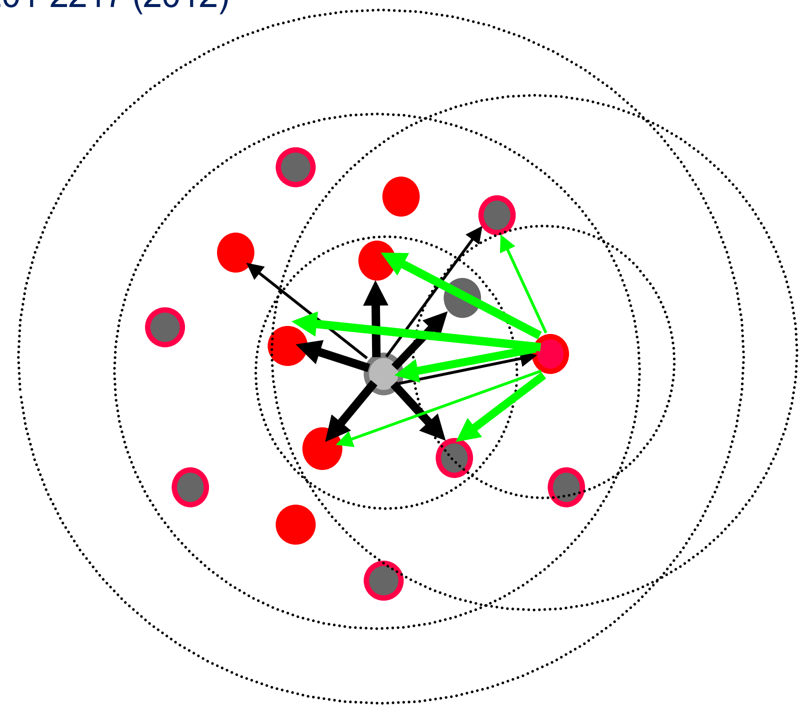
- Design of Decentralised OSN (DOSN)
- Develop effective policies for data forwarding and data dissemination in Mobile Social Networks (MSN)
- Generative models for OSN social and interaction graphs
- Information dissemination
  - Both in OSN and in MSN
- ...

FROM EGO NETWORKS TO OSN MODELS

- Step #1: A tool for synthetic generation of ego networks

Andrea Passarella, Robin I. M. Dunbar, Marco Conti, Fabio Pezzoni: Ego network models for Future Internet social networking environments. Computer Communications 35(18): 2201-2217 (2012)





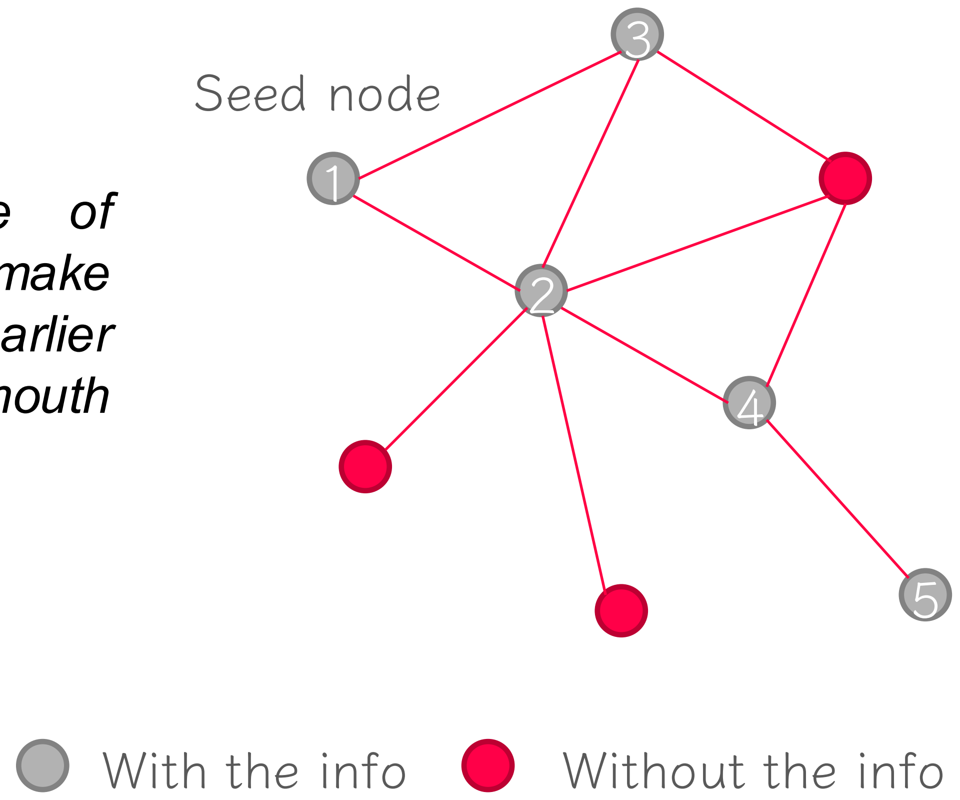
- Step #2: A generative model for social networks graphs based on the ego-network models. It reproduces both the macroscopic structure (e.g., its diameter and clustering coefficient), and the microscopic structure (e.g., the properties of the tie strength of individual social links) of human social networks

Marco Conti, Andrea Passarella, Fabio Pezzoni:
A Model to Represent Human Social Relationships in Social Network Graphs. SocInfo 2012: 174-187

From Ego Networks to Information Diffusion

- Why is it so important to study information diffusion?
  - Characterization of human behavior
  - Efficient targeted marketing campaigns
  - Control the diffusion process
  - Design Distributed and Mobile Social Networks

- Information Cascade: *a sequence of information adoptions for which users make decisions from inferences based on earlier actions from their contacts (word-of-mouth or infection)*

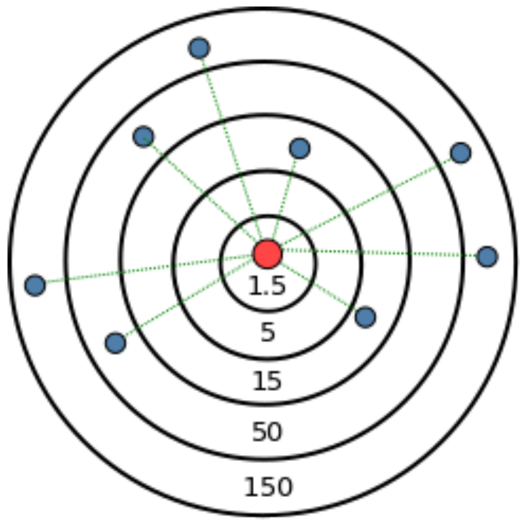


Ego Networks in OSN: Trust-Based Diffusions

- Trust impact on information diffusion
  - Not all messages are equal
    - General info without associated trust
    - General info, more or less reliable depending on from whom it arrives
    - ...
    - Private info

Level of trust

- We investigated a trust-based information diffusion
- We assumed that trust is positively correlated with tie strength
- The ego network model defines discrete levels of trust
- What happens to diffusions if we consider only contacts with a given trust level?



EGO NETWORKS IN OSN: IMPACT ON INFORMATION DISSEMINATION

- What happens if information diffusion stops at a certain layer of the ego networks?
  - E.g., trust sensitive information that one accepts only if it comes from most intimate friends
  - Information may spread very little
    - E.g., it reaches only 2.8% of users if only best friends accept it

Table 1 Percentage of nodes of the original graph covered by the largest component for the different thresholds. Thresholds are expressed in msg/month

Intimacy

Threshold	Percentage of nodes in the largest component					
	No insert	High freq	Low freq	Prob	Inv. prob	Rand
1/12 (act. cont.)	0.966	0.994	0.994	0.994	0.994	0.994
8/12 (friends)	0.297	0.714	0.705	0.726	0.722	0.725
1 (close fr.)	0.191	0.642	0.634	0.661	0.657	0.661
4 (v. intimate fr.)	0.028	0.386	0.385	0.453	0.444	0.456

EGO NETWORKS IN OSN: WHAT FOR?

- How can we improve information spreading?
  - Should we “convince” all people in the next layer to “accept” the information we propagate?
- Actually, not
  - Adding only 1 alter in the “next” layer results in coverage increase between ~2x and 10x

Policies to select the alter to be added

Table 1 Percentage of nodes of the original graph covered by the largest component for the different thresholds. Thresholds are expressed in msg/month

Intimacy

Threshold	**Percentage of nodes in the largest component**					
	No insert	**High freq**	**Low freq**	**Prob**	**Inv. prob**	**Rand**
1/12 (act. cont.)	0.966	0.994	0.994	0.994	0.994	0.994
8/12 (friends)	0.297	0.714	0.705	0.726	0.722	0.725
1 (close fr.)	0.191	0.642	0.634	0.661	0.657	0.661
4 (v. intimate fr.)	0.028	0.386	0.385	0.453	0.444	0.456

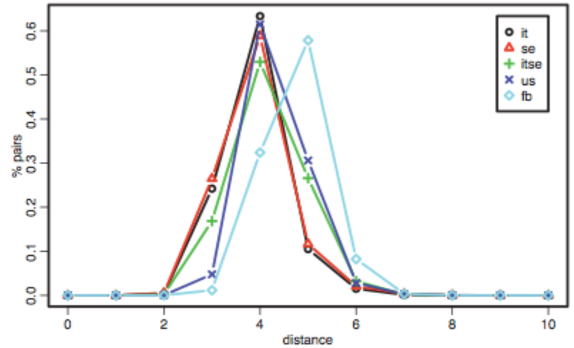
EGO NETWORKS IN OSN: WHAT FOR?

- Information diffusion: are we really @ 6 (or 4) degrees of separation?

Four Degrees of Separation

Lars Backstrom, Facebook, lars@fb.com
Paolo Boldi, Univ. degli Studi di Milano, boldi@dsi.unimi.it
Marco Rosa, Univ. degli Studi di Milano, marco.rosa@unimi.it
Johan Ugander, Facebook, jugander@fb.com
Sebastiano Vigna*, Univ. degli Studi di Milano, vigna@acm.org

WebSci 2012, June 22–24, 2012, Evanston, Illinois, USA.



Percentage of pairs at a certain distance

- Weighted shortest path
  - Weight of a link = tie strength of the link
  - Interpretation for maximal information spreading
    - weaker links are "more risky", and therefore have a higher cost in a path

- Whatever threshold we consider
  - Even using the entire ego network
- Whatever insertion policy
- The average path length is way longer than 4 (or 6)!

Table 6 Average length (# of nodes) of the weighted shortest paths in the largest component for the different thresholds representing the minimum contact frequency (msg/month) in the network

	Threshold - min. contact frequency			
Strategy	1/12 (active cont.)	8/12 (friends)	1 (close fr.)	4 (v. intimate fr.)
No insert	11.67	10.81	10.51	11.07
High freq	11.72	11.75	11.95	13.74
Low freq	11.68	11.93	12.19	16.11
Prob	11.71	11.95	12.21	16.16
Inv prob	11.71	11.97	12.30	17.42
Rand	11.74	11.95	12.28	17.15

V.Arnaboldi, M.La Gala, A.Passarella, and M.Conti, "Information diffusion in distributed OSN: The impact of trusted relationships", *Springer Peer-to-Peer Networking and Applications*, Aug 2015

DYNAMIC EGO-NETWORKS

Valerio Arnaboldi, Marco Conti, Andrea Passarella, Robin I.M. Dunbar, Structure of Ego-Alter Relationships of Politicians in Twitter, Journal of Computer-Mediated Communication, Volume 22, Issue 5, September 2017, Pages 231–247, https://doi.org/10.1111/jcc4.12193

TWITTER DATASETS

- 320 politicians
  - Italian cabinet members + recognised EU leaders
- ~14000 generic users
- REST API
  - Last 3200 tweets
    - Entire tweet history for several of them
- Active for at least 6 months
- Generating >3 tweets/month in at least 50% of the observation time window
- Tweet frequency compatible with human-like usage (to remove accounts that are handled by a pool of people)

Name	*Tweet Frequency*
Matteo Renzi	2.42
David Cameron	1.89
Enda Kenny	.74
Erna Solberg	2.16
Miro Cerar	3.06
Jean-Claude Juncker	1.88

STATIC EGO NETWORKS

- Tie strength: contact frequency over the entire observation window
  - Using direct communications only (reply, retweet, mention)
- Average sizes quite close to the conventional Dunbar's model
- Scaling ratios even closer to the model (expected value=3)
  - Smaller relative size of the confidence intervals
  - Stronger structural indication with respect to sizes

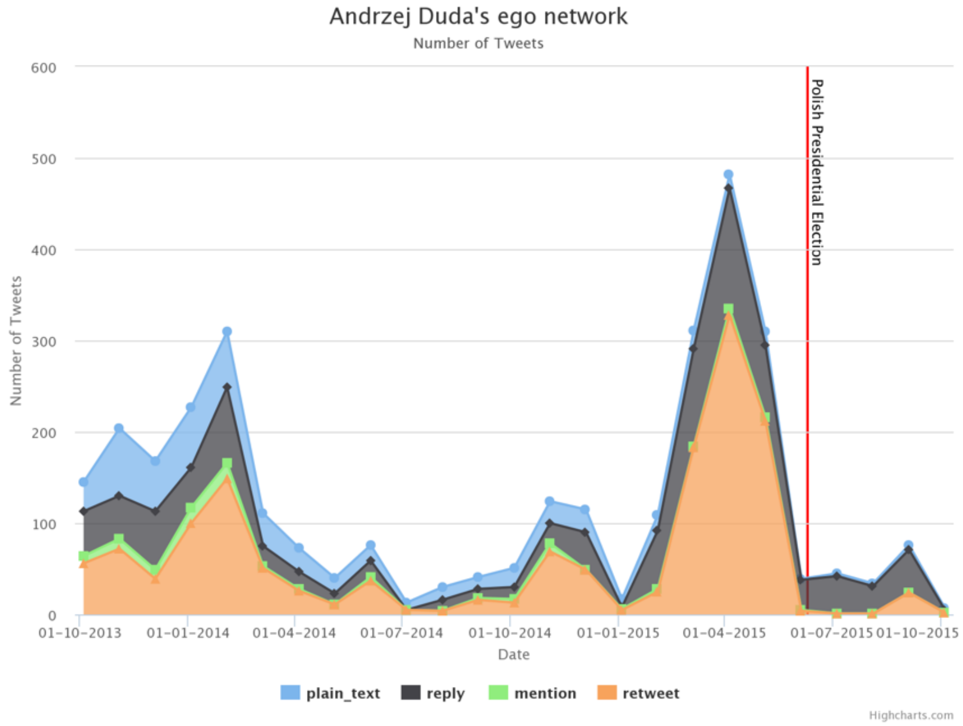
	Measure	*C1*		*C2*		*C3*		*C4*		*C5*
Italian MPs who mostly use replies	Size	2.07±.89		7.5±3.1		20.1±7.7		52±18.41		141±46.19
	Sc. ratio		3.92±.95		2.87±0.5		2.84±.49		2.85±.48	
	Measure	*C1*		*C2*		*C3*		*C4*		*C5*
Italian MPs who mostly use retweets	Size	1.63±.28		5.13±.92		14.23±2.41		39.81±6.3		134.9±22.8
	Sc. ratio		3.36±.42		2.88±.26		3.02±.27		3.52±.36	

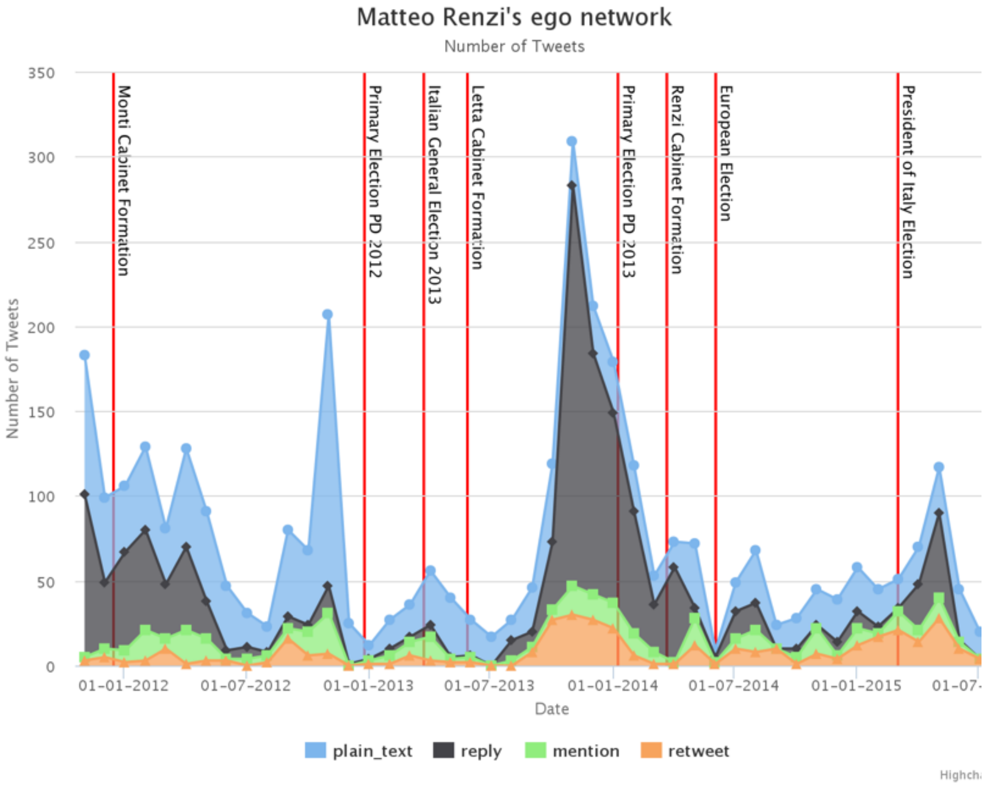
HOW MUCH INDICATIVE?

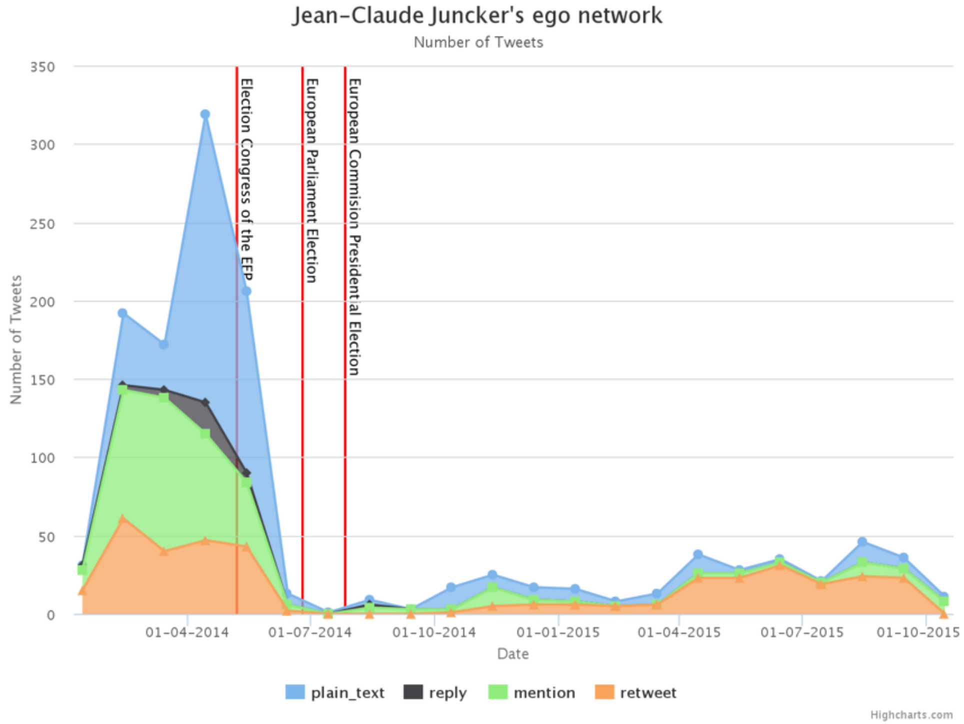
- Ego network over shifted 1-year time windows
  - 1 interaction/year threshold for being active

- Tweeting much variable over time
  - Spikes before key political events

- Need for a dynamic analysis of ego networks







DYNAMIC ANALYSIS

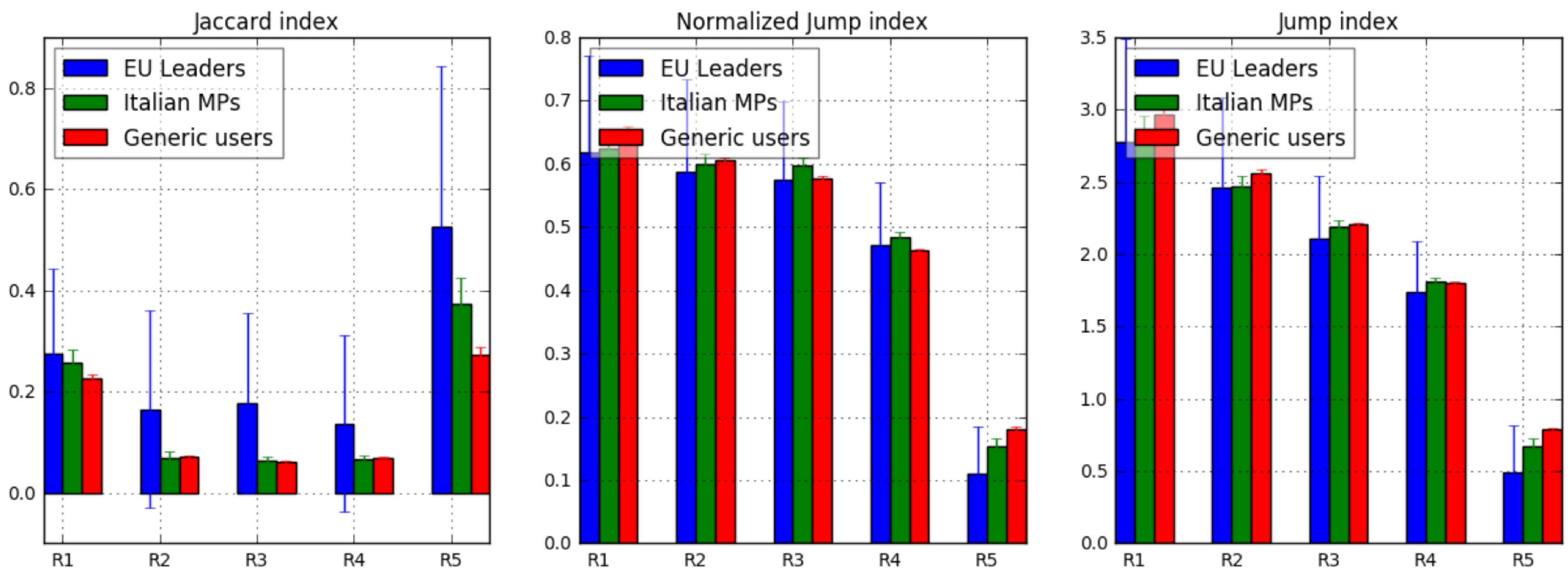
- Ego network over (not overlapped) 1-year time windows
  - 1 interaction/year threshold for being active
- Dynamic indices
  - Jaccard of membership layer-by-layer
  - Jump index (Conditioned to moving to a different ring)
    - #of layers jumped between consecutive time windows
    - Also a normalized version of the index wrt to the maximum jumps from its ring
  - Correlation of static and dynamic positioning
    - For each static layer, distribution over layers in the dynamic windows

JACCARD + JUMPS

- Inner layers more unstable than in human social networks
  - Different with respect to non-Twitter ego networks
  - Inner layers corresponds to short “strong social interactions”

- Jumps much longer from/to inner layers

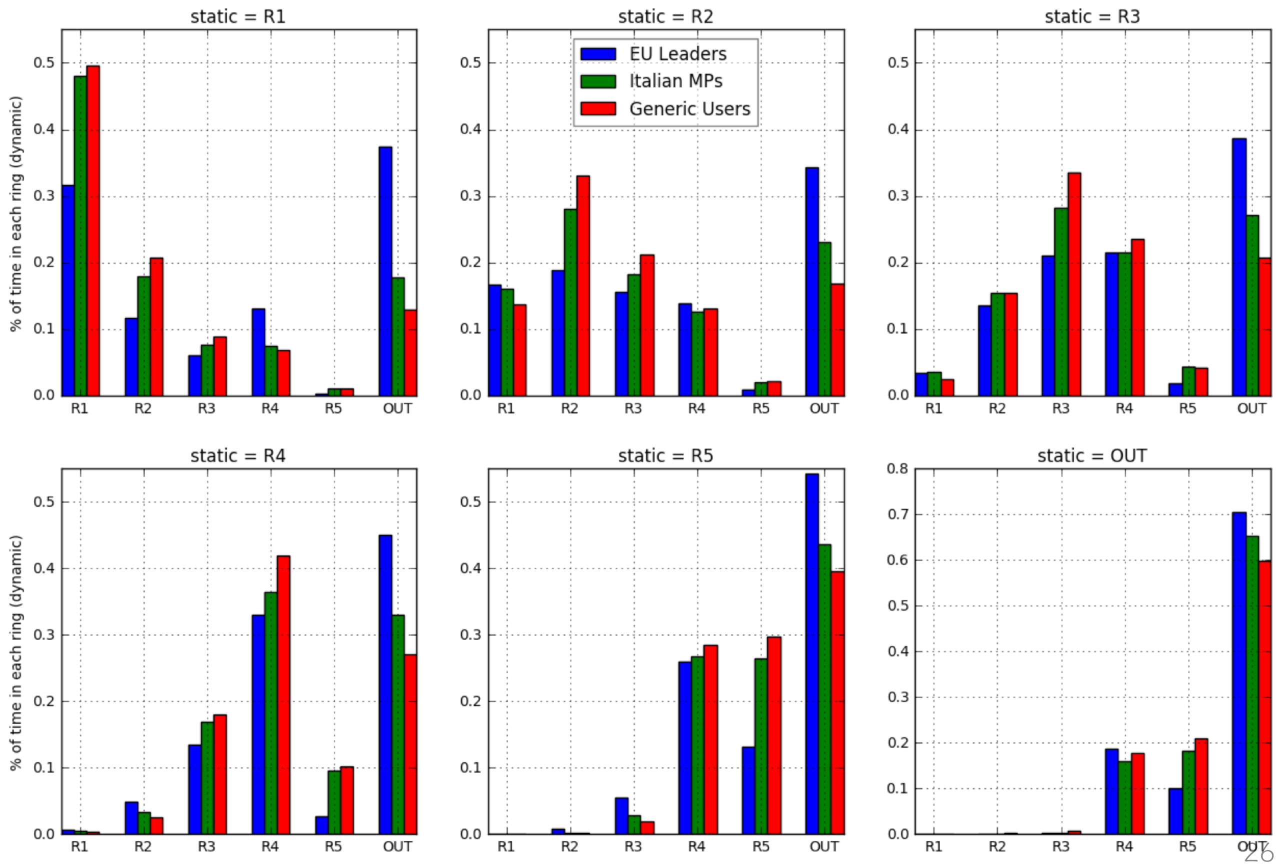
- Signs of sporadic, very active interactions in inner layers
  - Rather than strong social relationships, stable over time



CORRELATION STATIC/DYNAMIC

- Distribution over layers in dynamic windows, for relationships ending up in layer Rx in the static network
- For generic users there is a correspondence among rings in the static and dynamic case, but there is always a significant presence in OUT
- Peak in the corresponding layer of static for generic users and MPs; for EU leaders the peak is in OUT

- Very significant fraction spent outside of the ego network
- Non negligible presence in all layers

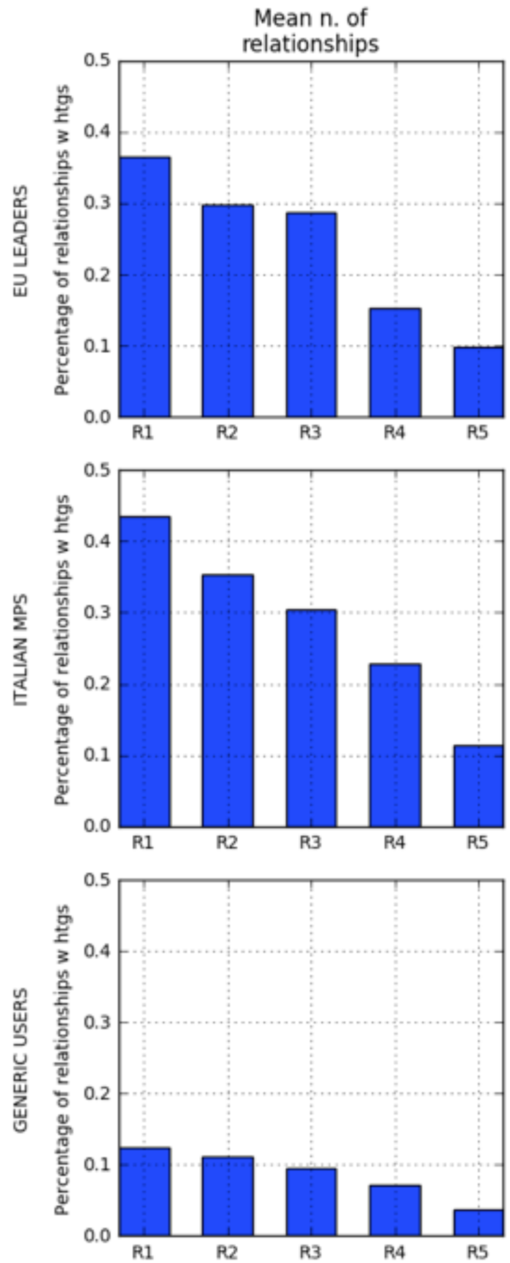


Twitter Ego Network Analysis Take Home

- Observed over long time period
  - Quite similar to the "conventional" Dunbar's model
- Observed over shorted time periods
  - Much more dynamic than expected
- Key features
  - More dynamic in the core than in outer layers
  - Long jumps and presence in all layers
    - "Come and go" into inner layers
- Common to both politicians and generic users
- *Is Twitter an information or a social network?*
  - *Is this different for politicians and generic users?*
- *We conjecture that especially for politicians Twitter is a tool for promoting discussions and visibility*

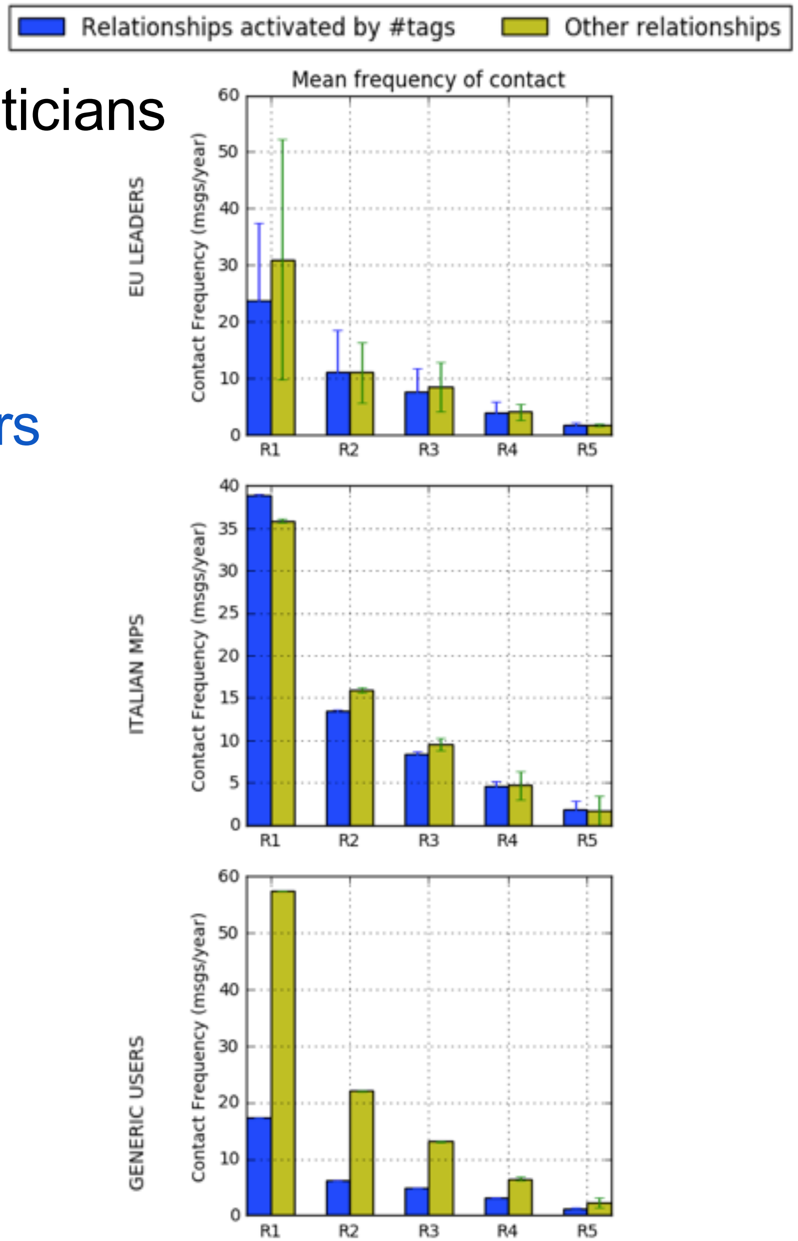
TWITTER AS AN INFORMATION NETWORK

- We analyze the impact of hashtags in activating ego-alter relationships
- Social links "activated" by hashtags
  - With a hashtag in the first tweet
- Fraction of such relationships at all layers
- Much higher for politicians at all layers



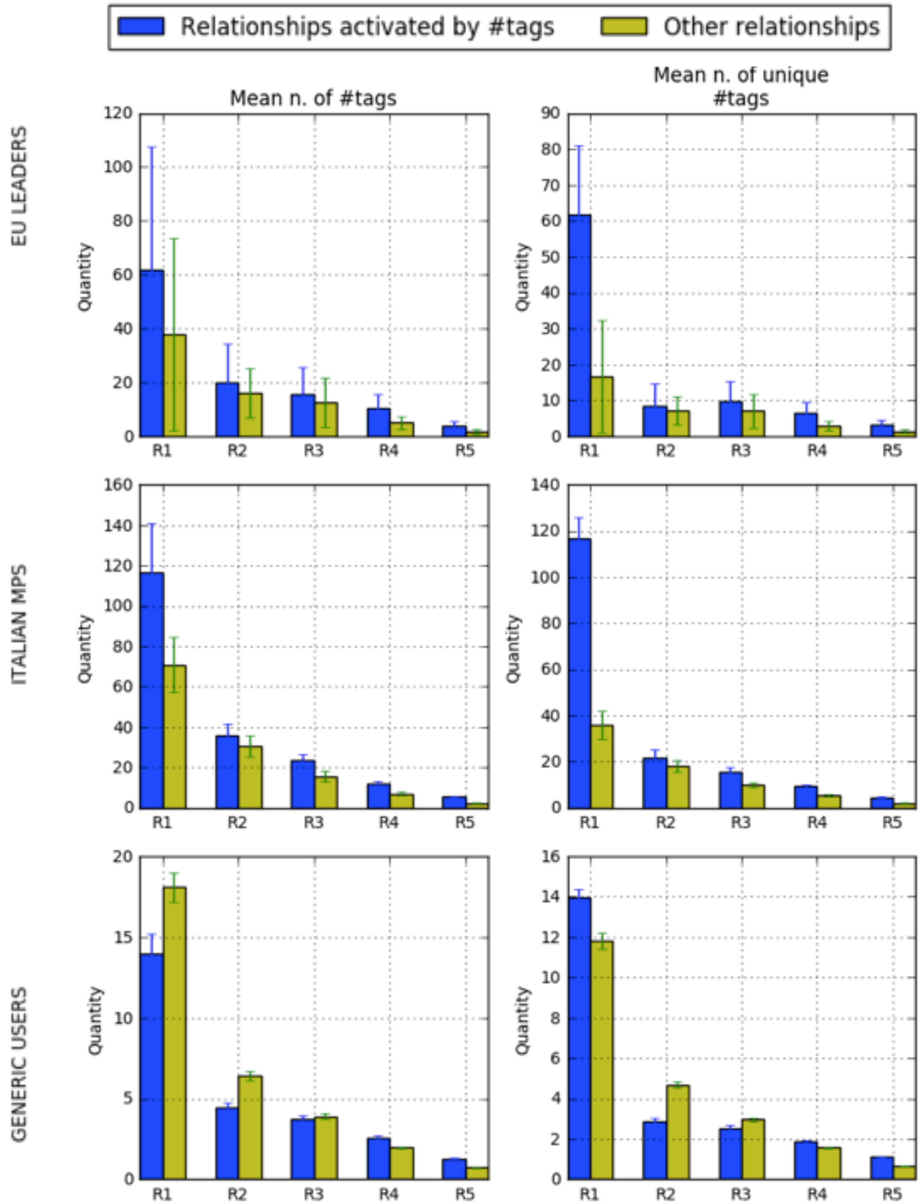
CONTACT FREQUENCY: POLITICIANS VS GENERIC

- Much higher contact frequency over relationships activated by hashtags for politicians
- Frequency similar to the one over "non information" social links
- This occurs primarily in the innermost layers



OF USED TAGS: POLITICIANS VS GENERIC

- Much more diversified used of tags
- Primarily for hashtag-activated relationships
- Primarily for inner-most layers
- Politicians tend to use much more the hashtags than generic users
- These results confirm that for politicians Twitter is used to expose public positions about specific topics through the interaction with a specific alter (e.g., journalists)



CONCLUSIONS

- Analysed politicians ego networks with respect to "generic" users
- Static analysis: ego networks similar to Dunbar's model
  - Same constraints shaping all kind of human social networks apply also here
- Dynamic analysis
  - More dynamic ego networks than expected
    - Both for politicians and generic users
  - Primarily so in the inner-most layers
- Inner layers of politicians
  - Very unstable relationships
  - Created for information diffusion/debates positioning
    - Lasting only as long as the time span of a "heated" discussion
- For politicians Twitter is mainly used as an information network for societal debates
  - Most social links are driven by tags (not by alters)
  - Primarily so in inner-layers
    - Which are also the most dynamic
  - Ego-networks shaped by the same cognitive constraints of offline/online SN or generic users

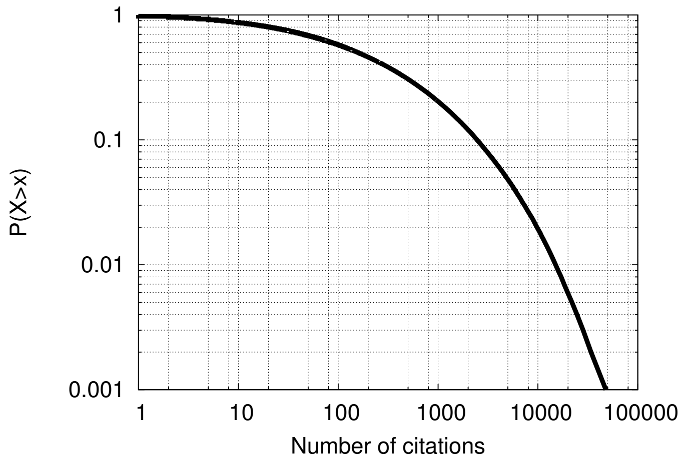
ANALYSIS OF CO-AUTHORSHIP EGO NETWORKS

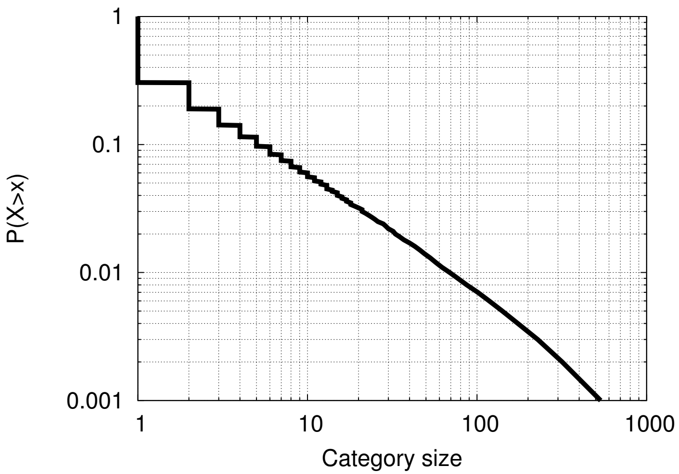
Valerio Arnaboldi, Robin I. M. Dunbar, Andrea Passarella, Marco Conti, "Analysis of Co-authorship Ego Networks", Proc. NetSci-X 2016, pp. 82-96

Co-Authorship Ego Networks

- We analyzed the structure of a number of co-authorship ego networks to see whether they show the same hierarchical structure found in social networks

- 312,207 Google Scholar profiles
- 188,657 categories (manually assigned by authors)
- 19,420,220 publications



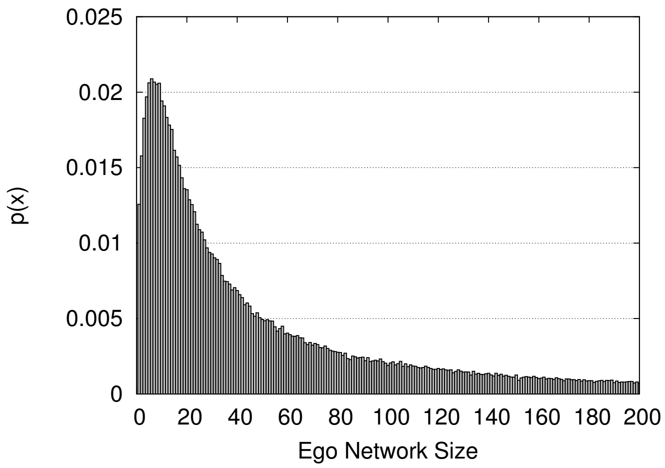


- We applied a community detection algorithm on the bipartite author-category graph to cluster authors into research areas and analyse possible differences between them

Categories of Authors in Google Scholar

Main Category	First 6 sub-categories (by size)	N. of Authors	N. of sub-categories
Machine learning	Artificial intelligence, computer vision, data mining, image processing, robotics, software engineering	81,409	23,615
physics	Nanotechnology, optimization, biochemistry, biophysics, chemistry, material science	80,419	27,666
neuroscience	Economics, psychology, education innovation, cognitive neuroscience, enterpreneurship	59,609	24,219
bioinformatics	Computational biology, genomics, molecular biology, evolution, genetics, conservation biology	59,428	19,802
ecology	Climate change, remote sensing, gis, gis&t, hydrology, geology	47,519	21,965
Molecular biology	Microbiology, medicine, hiv aids, biotechnology, immunology, epidemiology	20,431	19,738

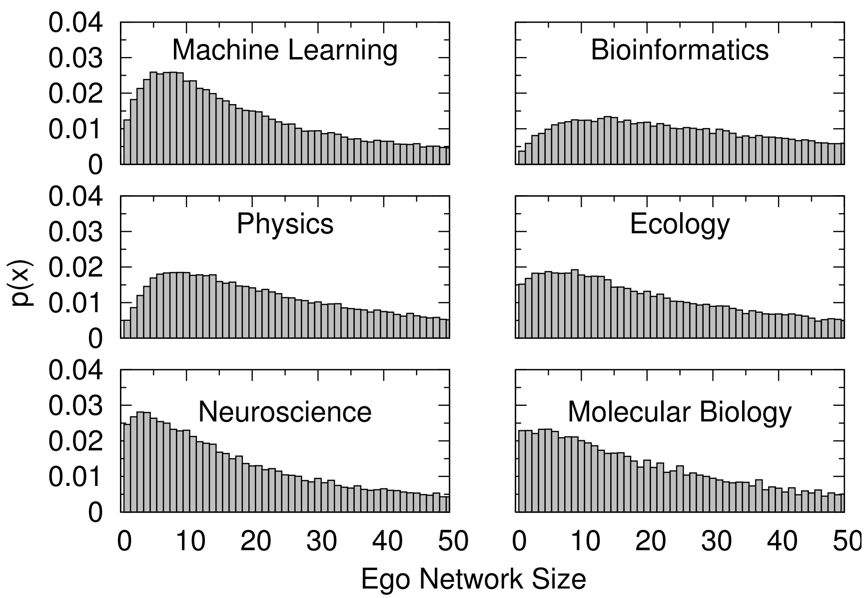
The Size of Co-Authorship Ego Networks



$$ts_{i,j} = \frac{1}{d_{i,j}} \sum_{p \in \{P(i) \cap P(j)\}} \frac{1}{k(p) - 1}$$

- Tie strength is the frequency of collaboration normalized by the number of co-authors in each paper
- The higher the number of co-authors in a paper, the lower the tie strength

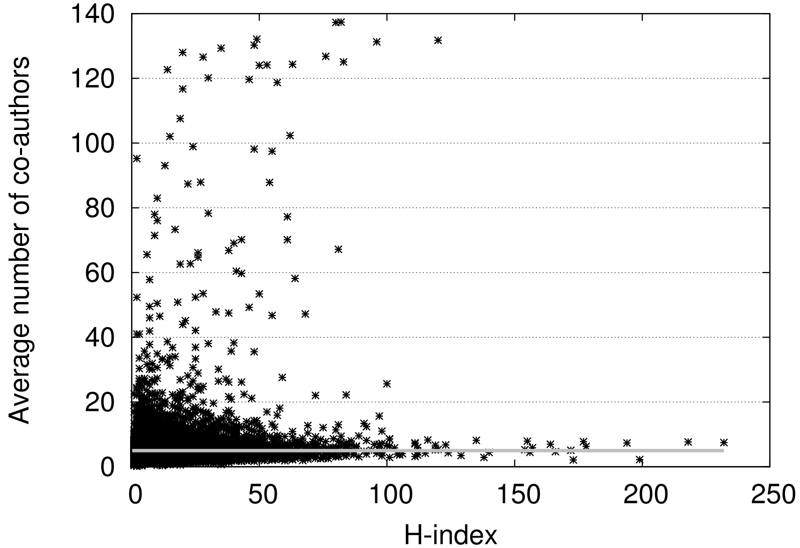
- We selected only authors (egos) with more than 3 years of publication activity
- We constructed ego networks by matching names of co-authors and counting the number of paper written with the ego



The Structure of Co-Authorship Ego Networks

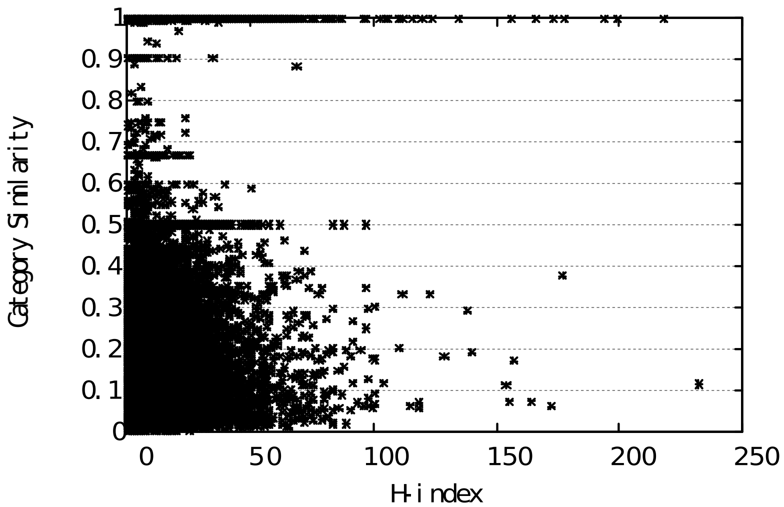
	C1	C2	C3	C4	C5
All authors	2.0	6.3	15.8	37.9	116.8
Machine learning	1.9	5.7	14.2	33.5	102.6
physics	2.0	6.4	16.1	38.6	119.5
neuroscience	1.8	5.5	13.3	31.1	100.1
bioinformatics	2.2	7.5	19.6	49.2	150.9
ecology	2.0	6.2	15.4	36.3	105.5
Molecular biology	2.1	6.2	15.0	34.3	97.0

Preliminary Results on Collaboration Strategies for Performances



- The average number of co-authors is distributed around 5, but with some exceptions due to some categories like physics
- Authors with very high h-index always have a number of co-authors around 5

- The similarity between categories is the average overlap between the categories of each author
- Authors with high similarity work in one field, authors with low similarity work in almost completely separated fields
- Successful authors seem to choose between working in one field or in completely separated ones



Co-authorship summary

- Humans organize their social relationships following a specific hierarchy
- This is true also in co-authorship networks, and the results confirm that there is a natural hierarchical grouping of everyday social structures
- Preliminary results indicates that 5 is the best group size for productivity – the size of the support clique
- Most productive authors seem to adopt two distinct strategies – work in a single field or in almost completely separated ones

Thanks!

References

- **Ego-networks in Facebook and Twitter**
  - Valerio Arnaboldi, Andrea Guazzini, Andrea Passarella: Egocentric online social networks: Analysis of key features and prediction of tie strength in Facebook. Computer Communications 36(10-11): 1130-1144 (2013)
  - Valerio Arnaboldi Andrea Passarella Marco Conti, Robin I. M. Dunbar: Online Social Networks - Human Cognitive Constraints in Facebook and Twitter Personal Graphs. Elsevier , 2015
  - Robin I. M. Dunbar, Valerio Arnaboldi, Marco Conti, Andrea Passarella: The structure of online social networks mirrors those in the offline world. Social Networks 43: 39-47 (2015)

- **Generative models**
  - Andrea Passarella, Robin I. M. Dunbar, Marco Conti, Fabio Pezzoni: Ego network models for Future Internet social networking environments. Computer Communications 35(18): 2201-2217 (2012)
  - Marco Conti, Andrea Passarella, Fabio Pezzon: A Model to Represent Human Social Relationships in Social Network Graphs. SocInfo 2012: 174-187

- **Dynamic studies on Twitter**
  - Valerio Arnaboldi, Marco Conti, Andrea Passarella, Robin Dunbar: Dynamics of personal social relationships in online social networks: a study on twitter. ACM COSN 2013: 15-2
  - Valerio Arnaboldi, Andrea Passarella, Marco Conti, Robin Dunbar: Structure of Ego-Alter Relationships of Politicians in Twitter. J. Computer-Mediated Communication 22(5): 231-247 (2017)

- **Data dissemination**
  - Valerio Arnaboldi, Marco Conti, Massimiliano La Gala, Andrea Passarella, Fabio Pezzoni: Ego network structure in online social networks and its impact on information diffusion. Computer Communications 76: 26-41 (2016)
  - Valerio Arnaboldi, Massimiliano La Gala, Andrea Passarella, Marco Conti: Information diffusion in distributed OSN: The impact of trusted relationships. Peer-to-Peer Networking and Applications 9(6): 1195-1208 (2016)

- **Other social networks**
  - Valerio Arnaboldi, Robin I. M. Dunbar, Andrea Passarella, Marco Conti: Analysis of Co-authorship Ego Networks. NetSci-X 2016: 82-96

- **Survey**
  - Valerio Arnaboldi, Marco Conti, Andrea Passarella, Robin I. M. Dunbar: Online Social Networks and information diffusion: The role of ego networks, Online Social Networks and Media 1, 44-55