

Psychological aspects of social communities

R. Lambiotte
Department of Mathematics
University of Namur, Belgium

Multi-relational Organization of Large-scale Social Networks in an Online World, M. Szell, R. Lambiotte and S. Thurner, PNAS, 107 13636-13641 (2010)

The Personality of Popular Facebook Users, D. Quercia, R. Lambiotte, M. Kosinski, D. Stillwell and J. Crowcroft, ACM CSCW (2012)

Psychological Aspects of Social Communities, A. Friggeri, R. Lambiotte, M. Kosinski and E. Fleury, SocialCom 2012

Computational social science

Small-scale questionnaire-based approaches



Fingerprints of individuals in electronic media (offline: mobile phone, or
online: email, Facebook, etc.)
+ Large-scale experiments in online media

Possibility to analyse the dynamics and organisation of large-scale social systems

D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, M. Van Alstyne, Science 323, 721-724 (2009).

Computational social science

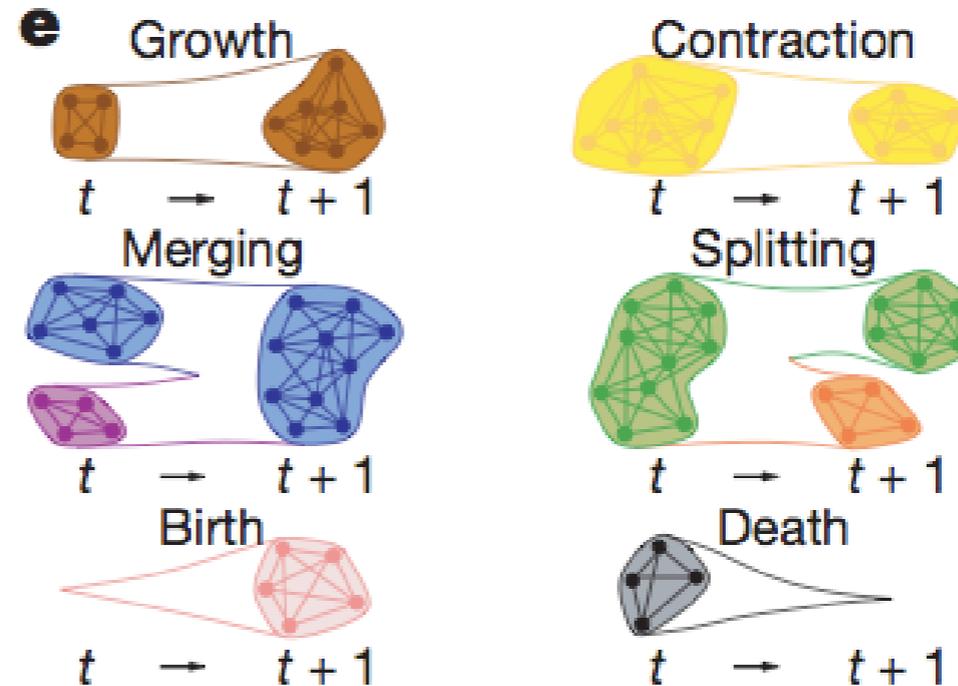
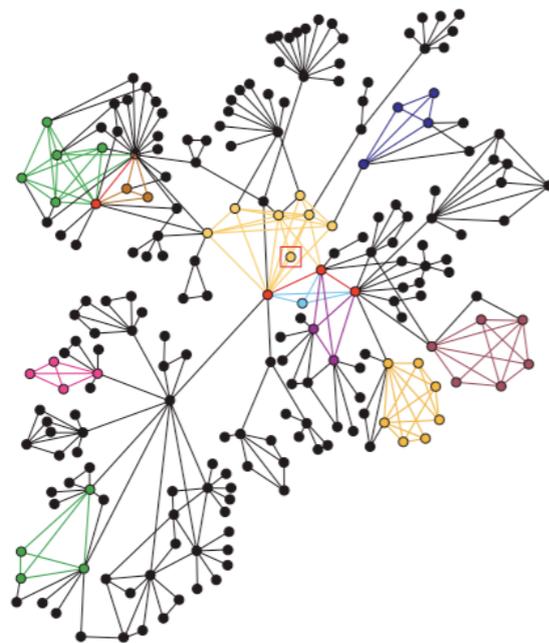
Small-scale questionnaire-based approaches



Fingerprints of individuals in electronic media (mobile phone, email, Facebook, etc.)

Longitudinal data-sets: quantification of the time evolution of social interactions

b Phone call



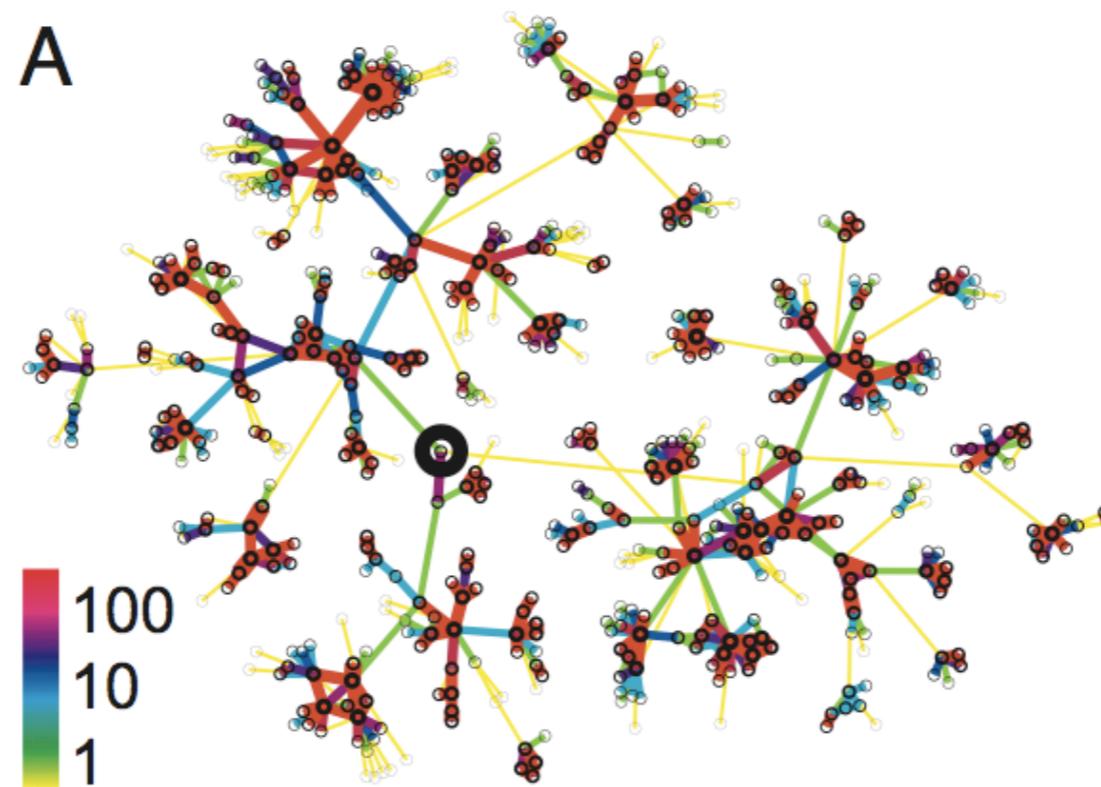
Computational social science

Small-scale questionnaire-based approaches



Fingerprints of individuals in electronic media (mobile phone, email, Facebook, etc.)

Weighted networks as a measure of the intensity of a tie



J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, A.-L. Barabási, PNAS 104 (18), 7332-7336 (2007).

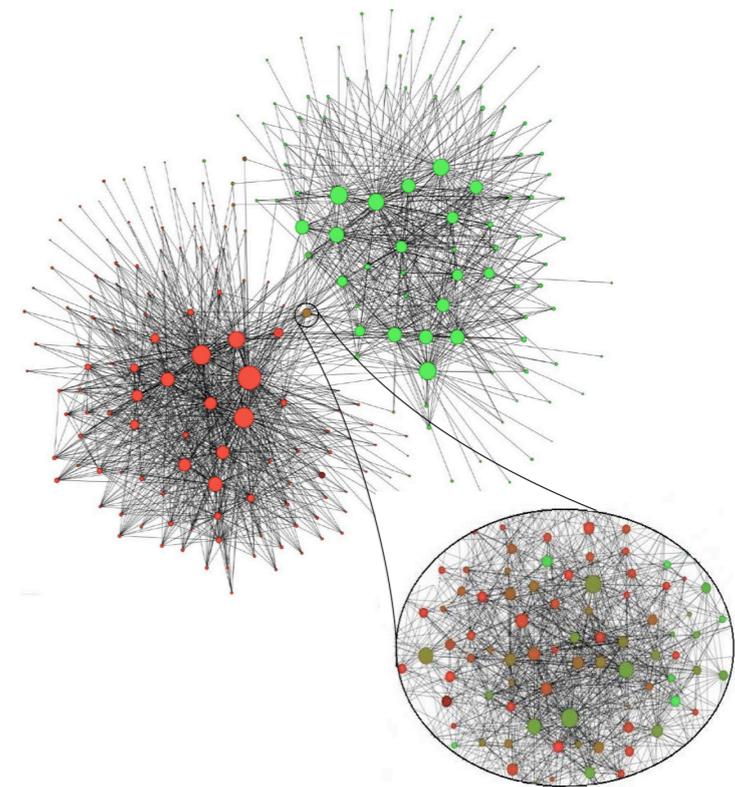
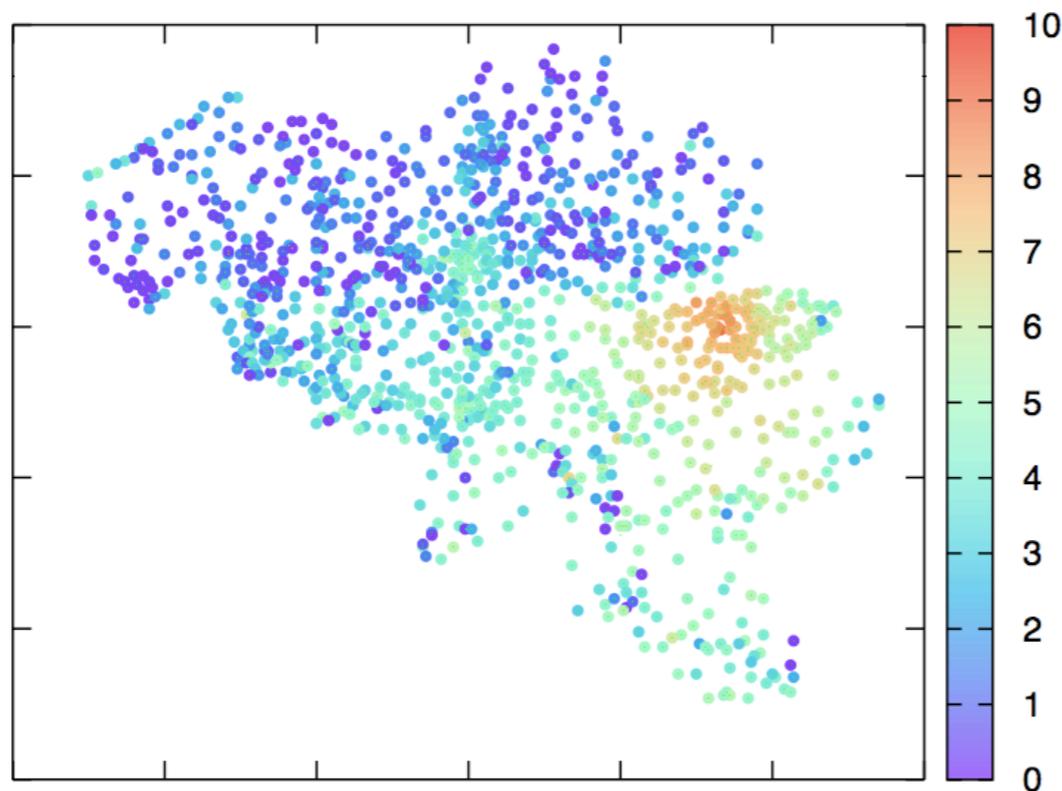
Computational social science

Small-scale questionnaire-based approaches



Fingerprints of individuals in electronic media (mobile phone, email, Facebook, etc.)

Node attributes: homophily and focus constraint



Computational social science

Small-scale questionnaire-based approaches

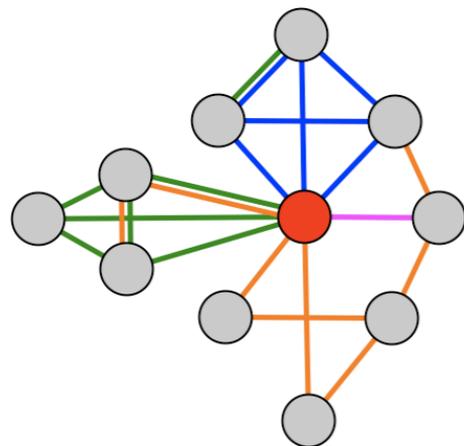


Fingerprints of individuals in electronic media (offline: mobile phone, or online: email, Facebook, etc.)
+ Large-scale experiments in online media

Possibility to analyse the dynamics and organisation of large-scale social systems

... but still a caricature of real-life social systems

Nature of the social relations?



Poor understanding of the nature of the agents

Despite our knowledge of some attributes, what about their personality?
Possibility to capture their intrinsic differences? Their tendency to react differently to the same cause?

Plethora of new services - new opportunities

Interweaving the social and the physical world

Offering more and more refined data about individuals and their interactions



Computational social science

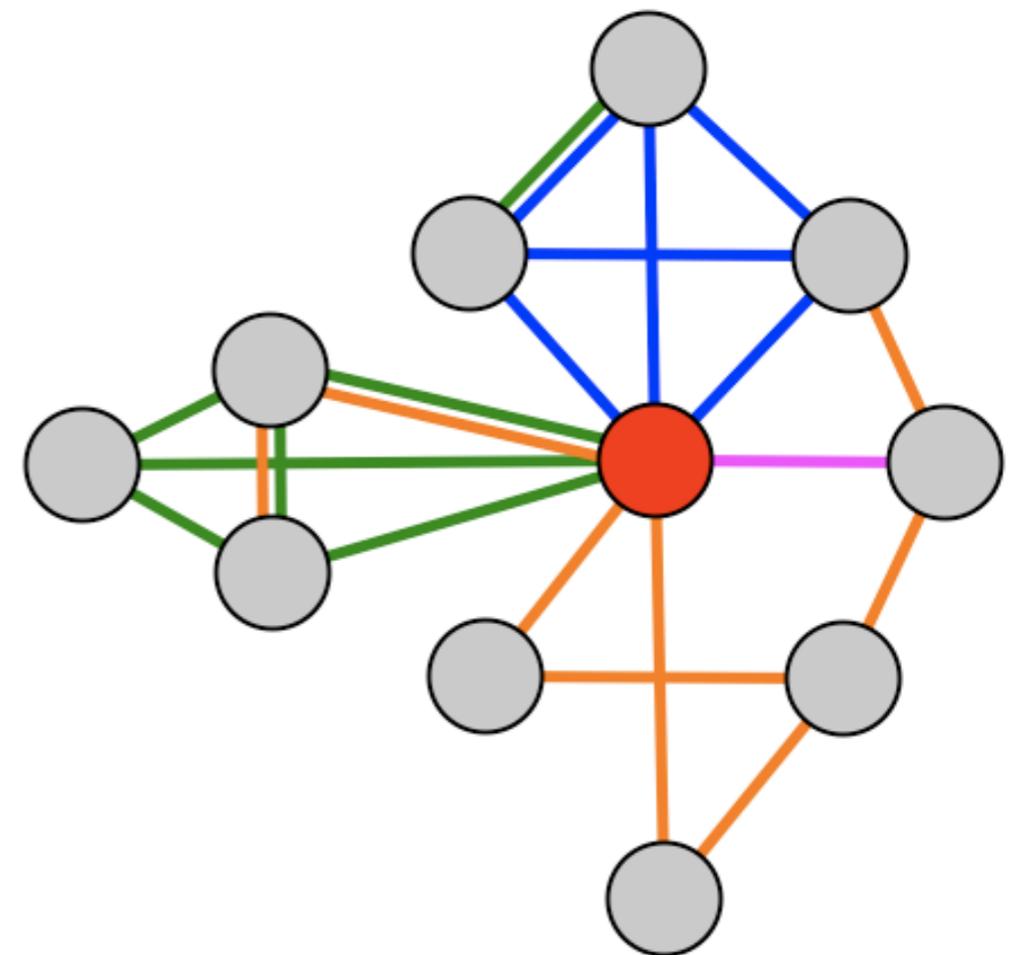
... but is blind to the existence of several types of social interactions between individuals

Relational ties are highly diverse and can represent a feeling, communication, exchange of goods (trade) or behavioural interactions
BUT electronic logs typically capture one channel of communication

Contrary to nodes (characterised by their age, sex, location, etc), the nature of interaction (family or work?) is usually unavailable in electronic data-sets

A society is characterised by the superposition of its constitutive socio-economic networks, all defined on the same set of nodes (multiplex networks)

A systemic understanding of a whole society can only be achieved by understanding these individual networks and how they influence and co-construct each other.



— Girlfriend
— Colleagues
— Family
— Friends

M. McPherson, L. Smith-Lovin and J.M. Cook (2001) *Annu. Rev. Sociol.* 27, 415.

K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis (2008) *Social Networks* 30, pp. 330-342.

S Wuchty, *PNAS* 2009 106 (36) 15099-15100

N Eagle, A Pentland and D Lazer, *PNAS* 2009 106 (36) 15274-15278

Computational social science

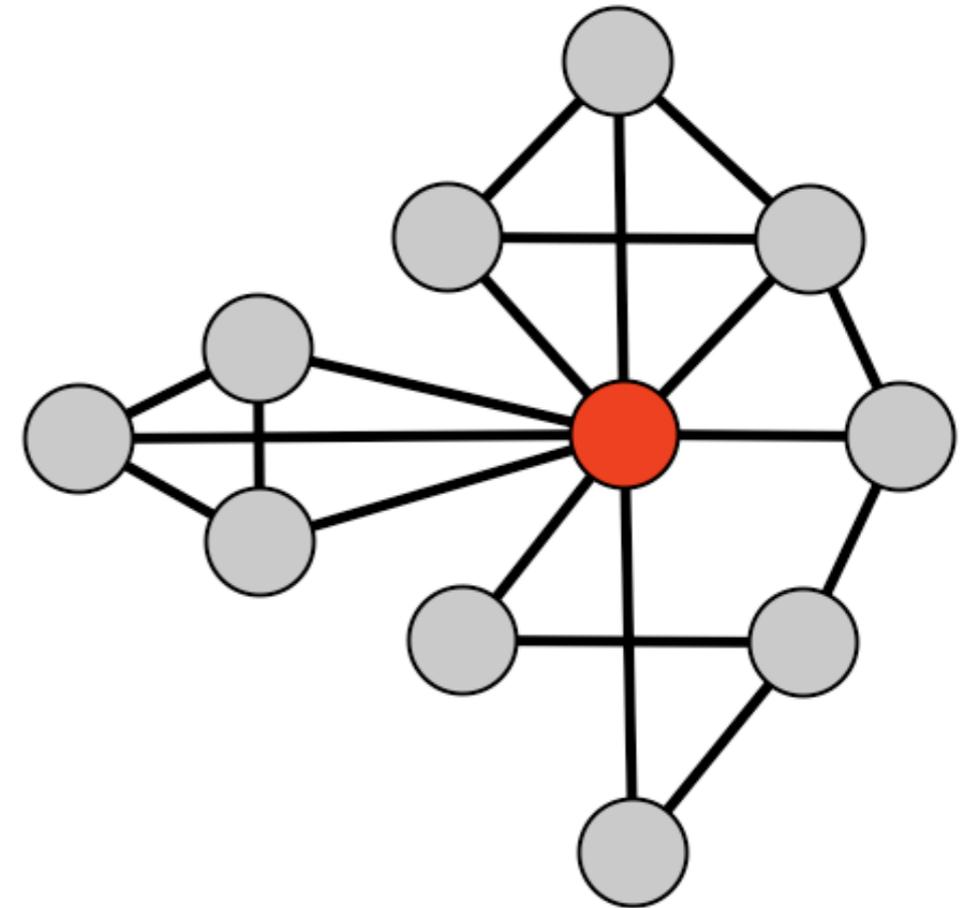
... but is blind to the existence of several types of social interactions between individuals

Relational ties are highly diverse and can represent a feeling, communication, exchange of goods (trade) or behavioural interactions
BUT electronic logs typically capture one channel of communication

Contrary to nodes (characterised by their age, sex, location, etc), the nature of interaction (family or work?) is usually unavailable in electronic data-sets

A society is characterised by the superposition of its constitutive socio-economic networks, all defined on the same set of nodes (multiplex networks)

A systemic understanding of a whole society can only be achieved by understanding these individual networks and how they influence and co-construct each other.



M. McPherson, L. Smith-Lovin and J.M. Cook (2001) *Annu. Rev. Sociol.* 27, 415.

K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis (2008) *Social Networks* 30, pp. 330-342.

S Wuchty, *PNAS* 2009 106 (36) 15099-15100

N Eagle, A Pentland and D Lazer, *PNAS* 2009 106 (36) 15274-15278

Computational social science

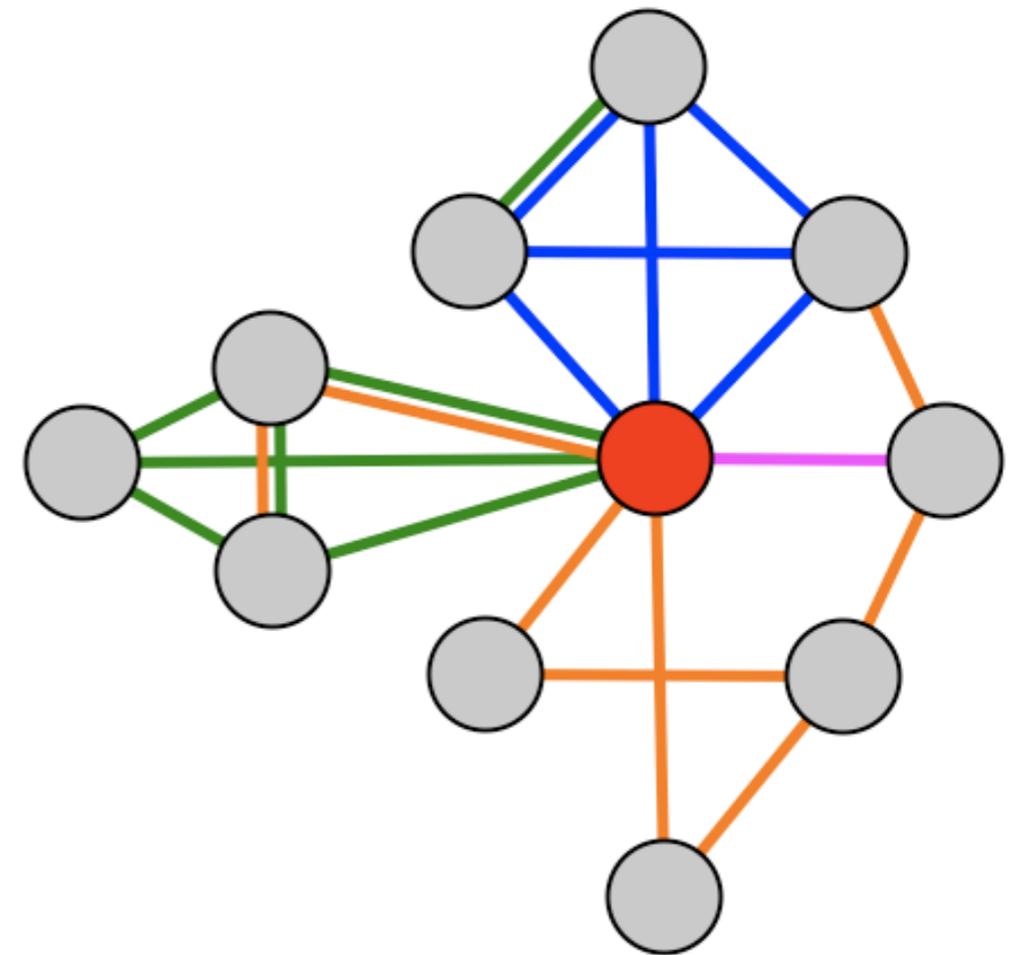
... but is blind to the existence of several types of social interactions between individuals

Relational ties are highly diverse and can represent a feeling, communication, exchange of goods (trade) or behavioural interactions
BUT electronic logs typically capture one channel of communication

Contrary to nodes (characterised by their age, sex, location, etc), the nature of interaction (family or work?) is usually unavailable in electronic data-sets

A society is characterised by the superposition of its constitutive socio-economic networks, all defined on the same set of nodes (multiplex networks)

A systemic understanding of a whole society can only be achieved by understanding these individual networks and how they influence and co-construct each other.



— Girlfriend
— Colleagues
— Family
— Friends

M. McPherson, L. Smith-Lovin and J.M. Cook (2001) Annu. Rev. Sociol. 27, 415.

K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis (2008) Social Networks 30, pp. 330-342.

S Wuchty, PNAS 2009 106 (36) 15099-15100

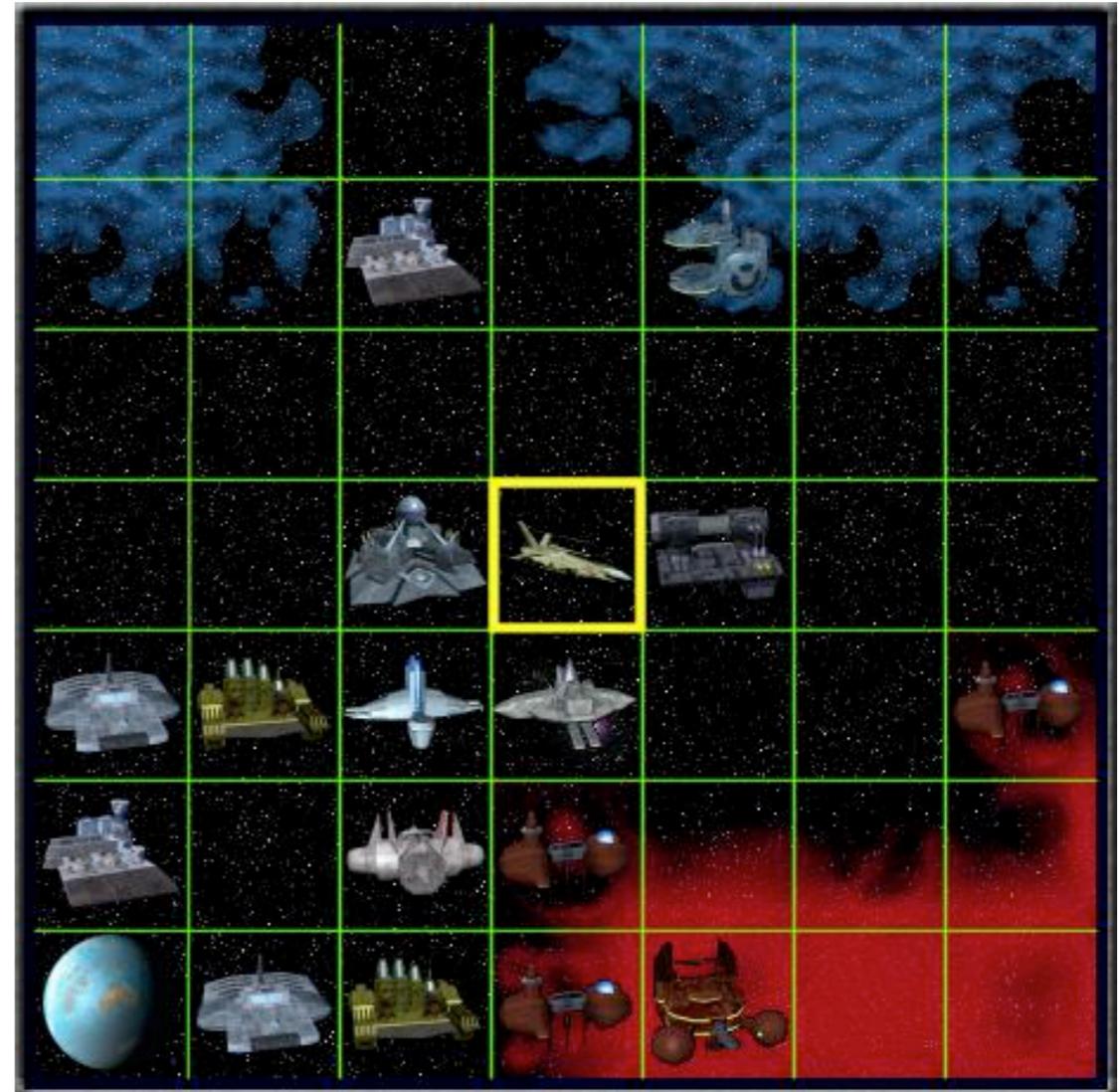
N Eagle, A Pentland and D Lazer, PNAS 2009 106 (36) 15274-15278

Nature of social relations: Massive Online Games

Players are immersed in a virtual world where they experience an *alternative* life with a variety of possible social interactions among players.

Motivation: establish friendships, gain respect and status in the virtual community.

All information about all actions taken by all players is stored in log-files



Massive Online Games

Pardus.at: Massive multiplayer browser game

330,000 registered, 13,000 active players

Played since 2004 (Free, optional 5\$/month)

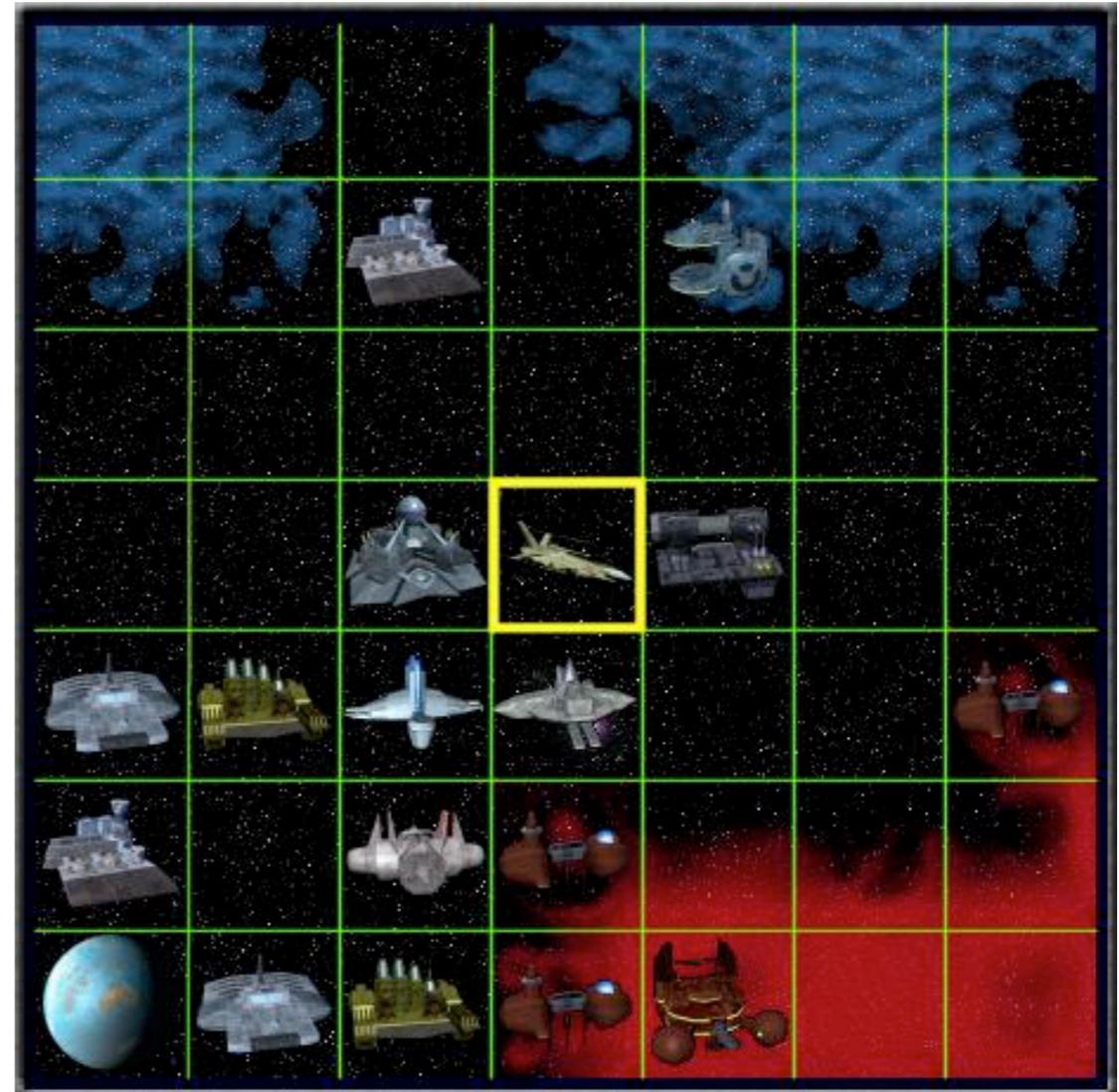
Open-ended game (no winner)

Players self-organise within groups and subgroups, claim territories, decide to go to war, etc., completely on their own account.

Economic life: Trade, production

Social life: Chats, forums, private messages

Exploratory life: explore of an unknown universe



Massive Online Games

Multiplexity: 6 types of directed, one-to-one interactions

Communication network: personal messages (similar to email)

Trade network: exchange of money for commodity

Friendship network: players can mark others as friends. Only the marker and the marked player know this information

Attack network: attacks performed by one player on the spaceship of another player

Bounty network: money promised for the destruction of a certain player

Enmity network: players can mark others as enemies. Only the marker and the marked player know this information

Massive Online Games

Multiplexity: 6 types of directed, one-to-one interactions

Communication network: personal messages (similar to email)

Trade network: exchange of money for commodity

Friendship network: players can mark others as friends. Only the marker and the marked player know this information

positive
interactions



Attack network: attacks performed by one player on the spaceship of another player

Bounty network: money promised for the destruction of a certain player

Enmity network: players can mark others as enemies. Only the marker and the marked player know this information

antagonistic
interactions



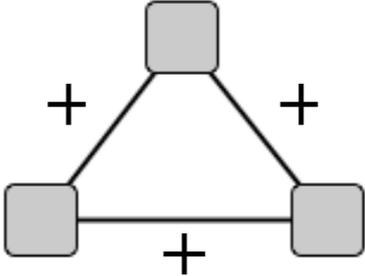
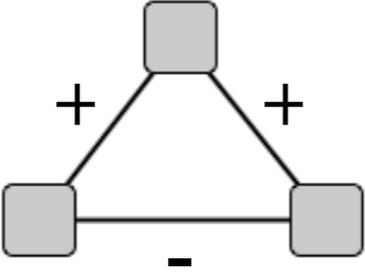
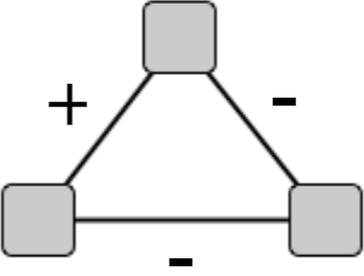
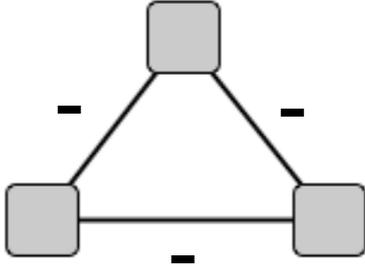
Static networks: Friendship and enmity networks are taken as snapshots at the last available day. All other networks are aggregated over time. For simplicity, we use unweighted, directed networks.

Undirected networks are also constructed: a link exist between i and j if there exists at least one directional edge between those nodes

Empirical verification of structural balance

Some configurations of signed motifs are socially and psychologically more likely than others

Unbalanced triads are sources of stress and therefore tend to be avoided by actors when they adapt their personal relationships

| |  |  |  |  | |
|-------------------------------|---|--|---|---|---------------------------|
| Strong formulation of balance | B | U | B | U | Cartwright (after Heider) |
| Weak formulation of balance | B | U | B | B | Davis |
| N_{Δ} | 26,329 | 4,428 | 39,519 | 8,032 | |
| $N_{\Delta,r}$ | 10,608 | 30,145 | 28,545 | 9,009 | |
| \mathcal{Z} | 71 | -112 | 47 | -5 | |

Predicting positive and negative links in online social networks, J. Leskovec, D. Huttenlocher, J. Kleinberg, ACM WWW Int Conf on World Wide Web (2010)

Multi-relational Organization of Large-scale Social Networks in an Online World, M. Szell, R. Lambiotte and S. Thurner, PNAS, 107 13636-13641 (2010)

Computational social science

... is often blind to the diversity in the nature/personality of individuals

In data such as facebook, individuals are characterized by variables such as their age, sex, physical location, etc.

First step toward a characterization fo the heterogeneity across users.

Is it possible to use facebook to measure the personality of users and to understand its influence on their (social) behaviour?

Measures of personality

The five-factor model of personality, or the big five, is the most comprehensive, reliable and useful set of personality concepts. The idea is that an individual can be associated with 5 scores that correspond to 5 main personality traits.

Personality traits predict a number of real-world behaviors. They, for example, are strong predictors of how marriages turn out: if one of the partner is high in Neuroticism, then divorce is more likely.

| Dimension | High scorers | Low scorers |
|--------------------------|------------------------------|-----------------------|
| Openness | Imaginative | Conventional |
| Conscientiousness | Organized | Spontaneous |
| Extraversion | Outgoing | Solitary |
| Agreeableness | Trusting | Competitive |
| Neuroticism | Prone to stress and worry | Emotionally stable |

Table 1: The big five personality dimensions.

Influence of OCEAN on sociometric popularity

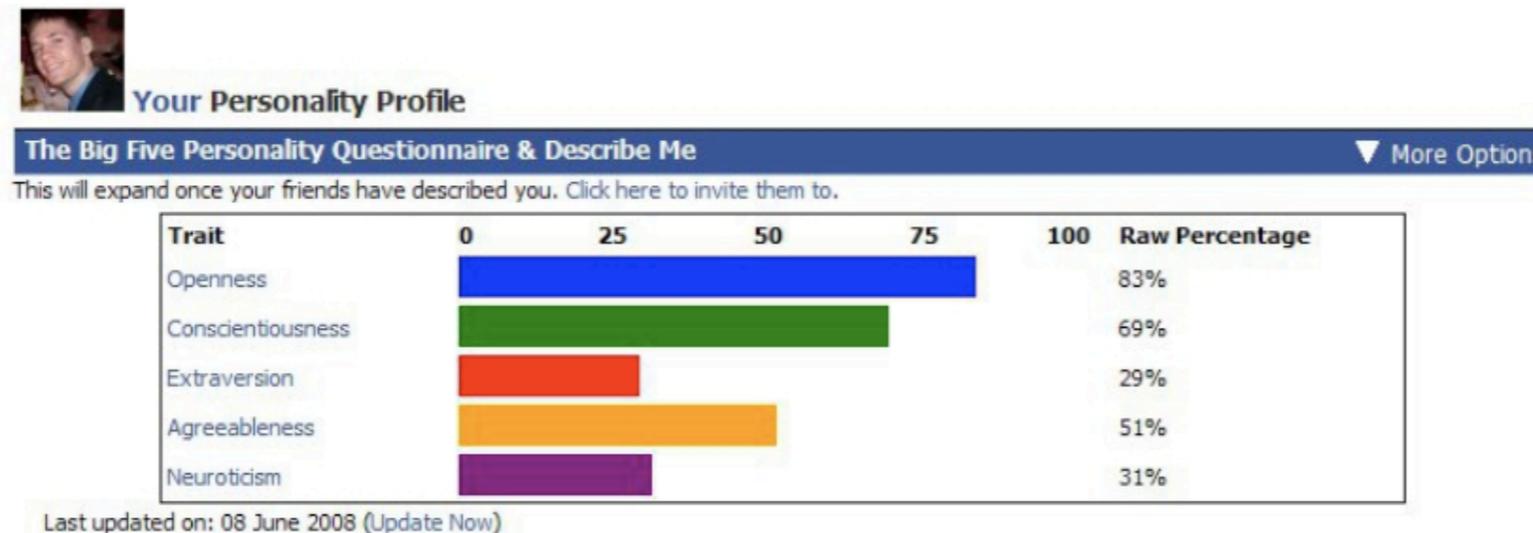
Numerous studies:

- The trait of Extraversion is the strongest predictor for the number of real-world friends. Extraversion has been shown to be consistently associated with structural measures of social support, including network size and contact with network members.
- Neuroticism has been associated with negative social interactions.
- Findings for the traits of Agreeableness, Conscientiousness and Openness are inconsistent.

But usual limitations of questionnaire-based studies: very small (poor statistics), biased samples (WEIRD, White, Educated, Industrialized, Rich, and Democratic)

Measuring OCEAN at a large-scale: myPersonality

- Facebook application: 5.5 million users
- Users can opt in and give their consent to share their profile information (40%)
- Right incentives: subjects are not paid nor receive college credits. myPersonality users are solely motivated by the prospect of receiving reliable feedback and test results that accurately describe their personalities.
- Unreliable results are removed. Numerous validity tests



- myPersonality is able to obtain test results that are more reliable than those in pen-and-paper studies.
- myPersonality users are far less biased than those studies' subjects for gender, age, and geography.
- VERY large scale data

Influence of OCEAN on sociometric popularity

200k users with between 30 and 1000 Facebook friends and between 18 and 54.

Significant factors for predicting $\log(k)$: linear regression

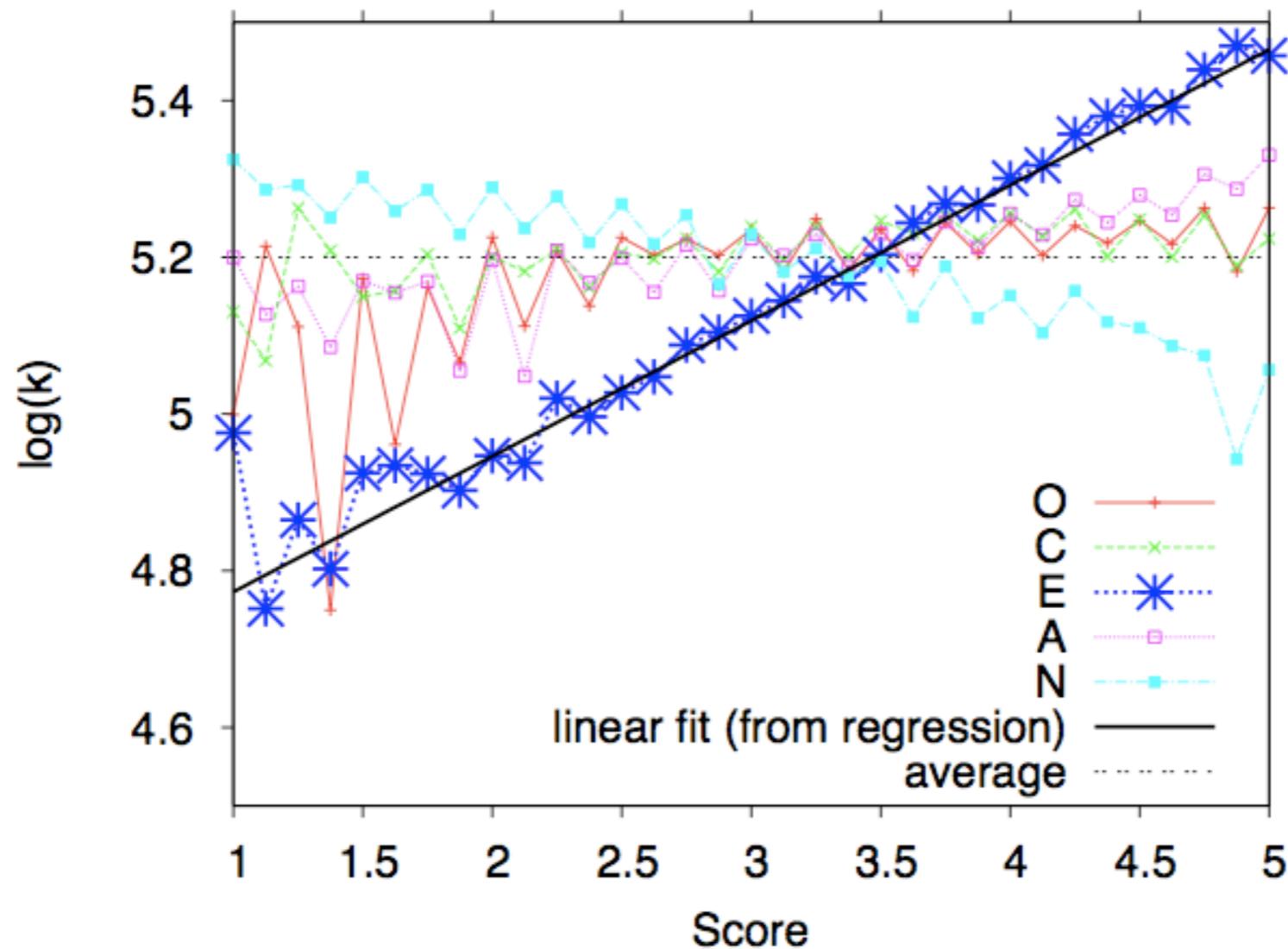
$$\log(k_i) = \alpha + \sum_{l=1}^5 \beta_l R_{l;i} + \beta_S S_i + \beta_A A_i + \epsilon_i$$

| Variables | β | t-test | β | t-test |
|-------------------|---------|---------------|---------|---------------|
| Openness | -0.031 | -12.6 | - | - |
| Conscientiousness | 0.004 | 1.5 | - | - |
| Extraversion | 0.174 | 79.7 | 0.171 | 79.8 |
| Agreeableness | 0.013 | 5.1 | - | - |
| Neuroticism | 0.005 | 2 | 0 | 0 |
| Sex | 0.005 | 1.4 | - | - |
| Age | -0.015 | -64.5 | -0.014 | -64.7 |

Influence of OCEAN on sociometric popularity

200k users with between 30 and 1000 Facebook friends and between 18 and 54.

Significant factors for predicting $\log(k)$: linear regression



Psychology and ego-networks

Ego-network of person *John*: friends of *John* and connections between them (*John* does not belong to his ego-network).

Number of friends = size of the ego-network

=> Extraversion affects the size of ego-networks, but what about its shape?

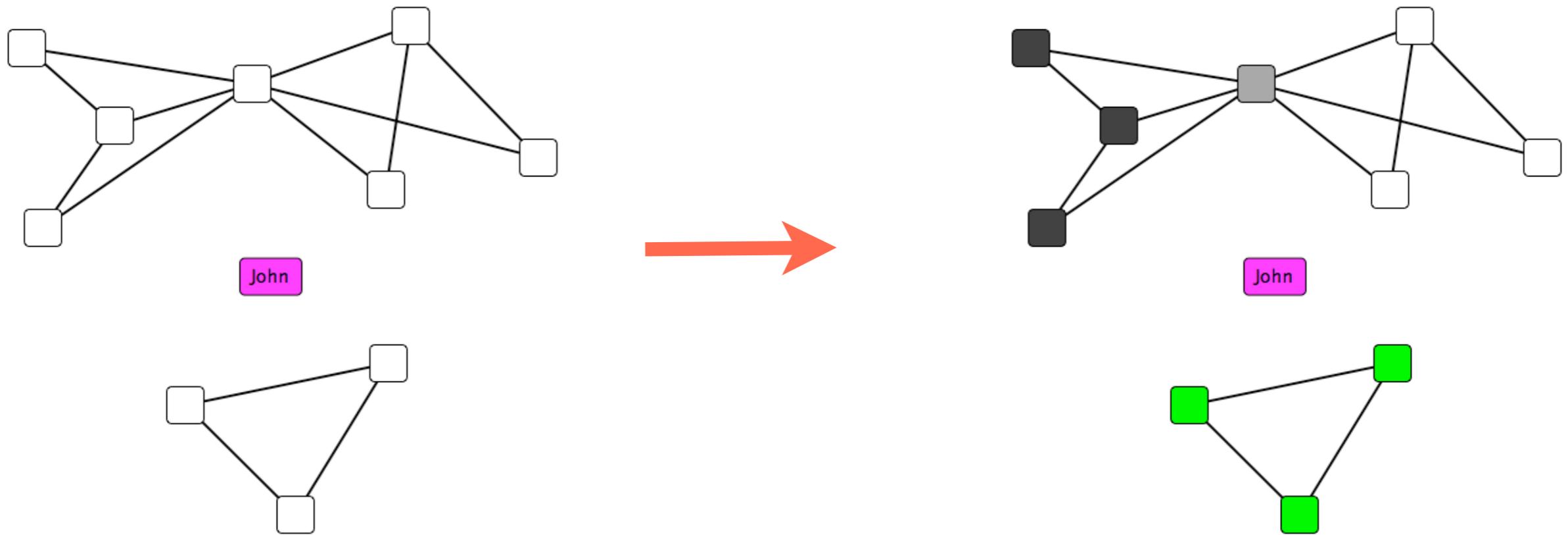
In the social science, long tradition in analysing and theorising ego-networks, e.g. their connection to social capital: dense ego-networks favor trust and facilitate information flow // open ego-networks indicate bridging capital as individuals bridge structural holes between disconnected others.

~ 50k users with number of friends comprised between 50 and 2000

PS: Local network analysis because of our very incomplete knowledge of the whole Facebook network (thousands vs billions)

From friends to communities

Organisation fo the ego-network into overlapping communities



From friends to communities

Hundreds of methods to uncover communities, overlapping or not

We applied the so-called C^3 method where the quality of a community is defined by its *social cohesion*

$$C(S) = \underbrace{\frac{\triangle(S)}{\binom{|S|}{3}}}_{\text{transitivity}} \times \underbrace{\frac{\triangle(S)}{\triangle(S) + \triangleleft(S)}}_{\text{isolation}}$$

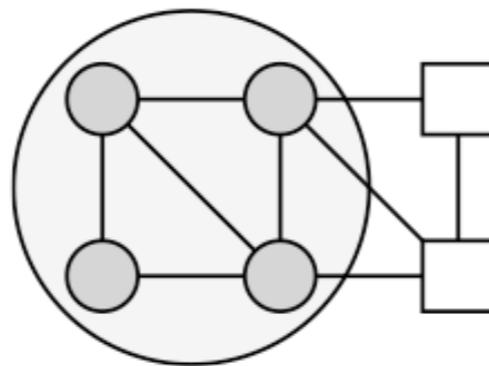


Fig. 3. In this example, the set of circle nodes contains 4 nodes, features 2 inbound triangles and only 1 outbound triangles, leading to a cohesion $C = \frac{1}{3}$.

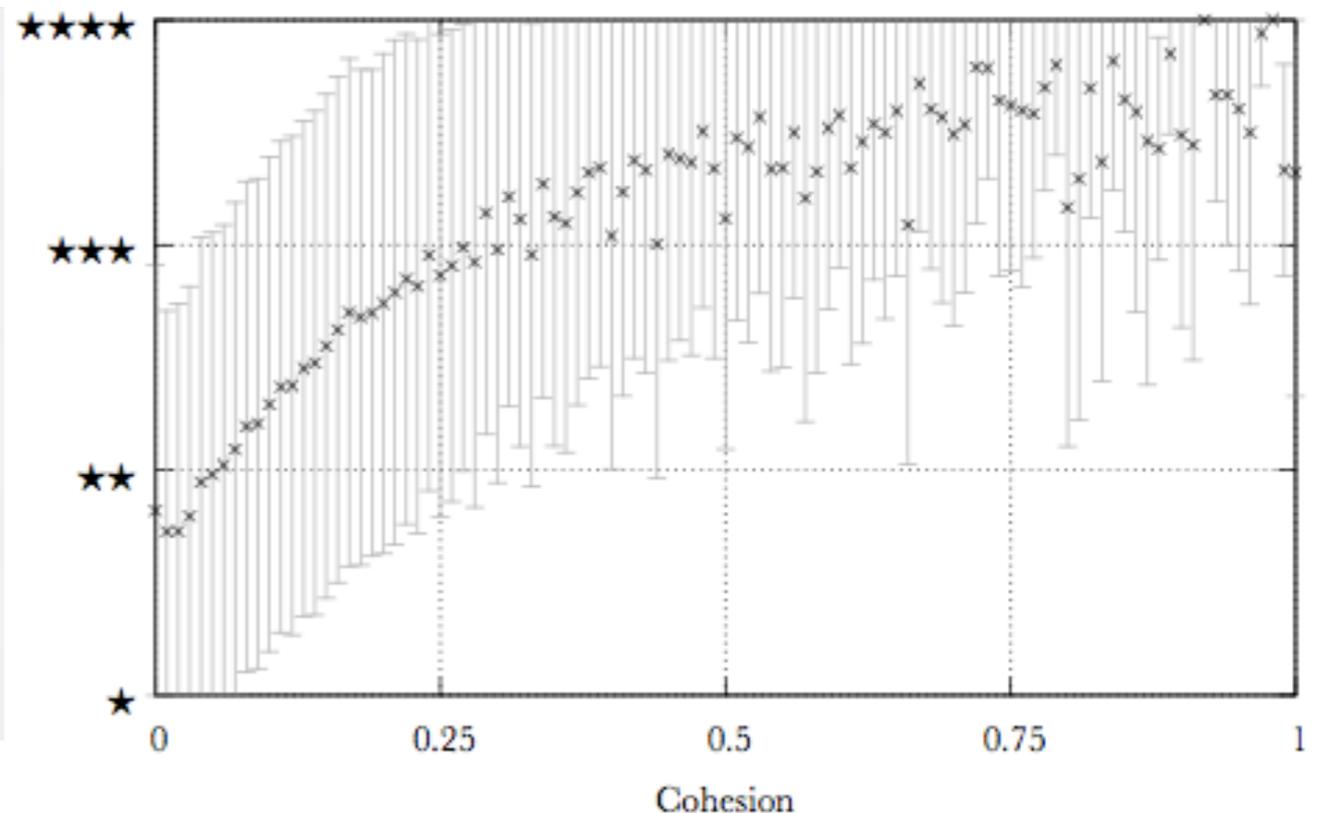
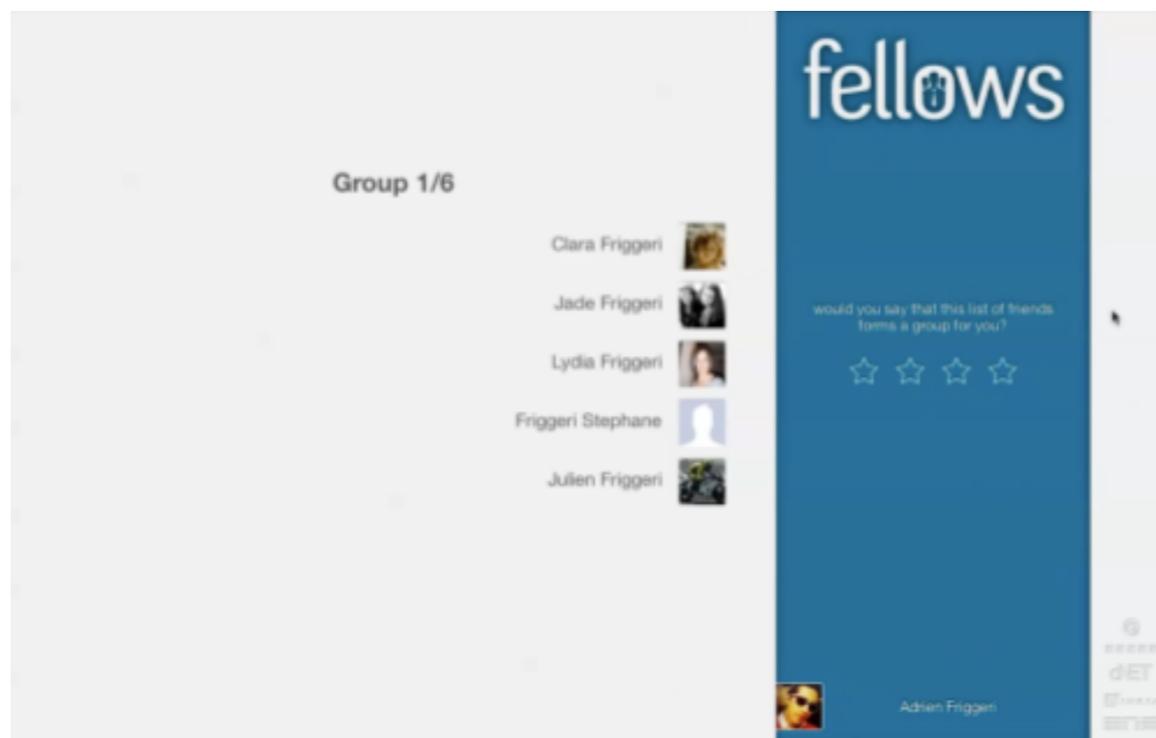
From friends to communities

Hundreds of methods to uncover communities, overlapping or not

We applied the so-called C^3 method where the quality of a community is defined by its *social cohesion*

Dedicated to, and tested on ego-networks.

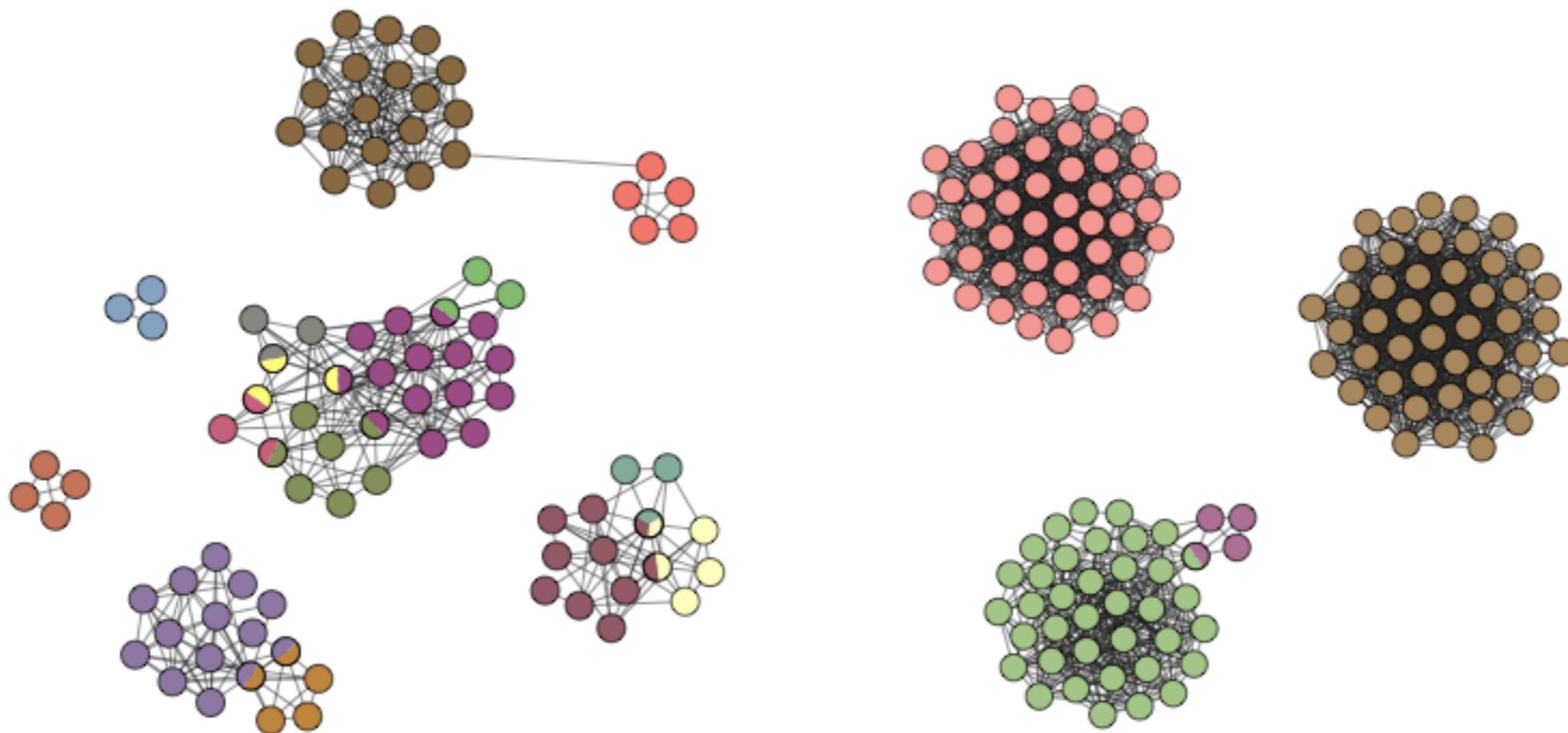
Validated on a large-scale experiment on Facebook (*Fellows*): algorithmic communities strongly correlated to the users' perception of the quality of social communities.



Extraversion and the organisation of ego-networks

Introverts tend to have less, larger communities: they *hide* into large communities.

Extroverts exhibit a higher overlap of the communities: they act as *bridges* between communities



(a) User A, 26 years old: high extraversion ($\text{ext} = 1.33$), 101 friends of which 91 are split across 15 communities of size varying between 3 and 19, and average cohesion $\bar{C} = 0.46$.

(b) User B, 19 years old: low extraversion ($\text{ext} = -1.21$), 145 friends of which 136 are split across 4 communities of size 4, 37, 48 and 48, and average cohesion $\bar{C} = 0.31$.

Fig. 8. Examples of two ego-networks of subjects with different psychological traits and structural features.

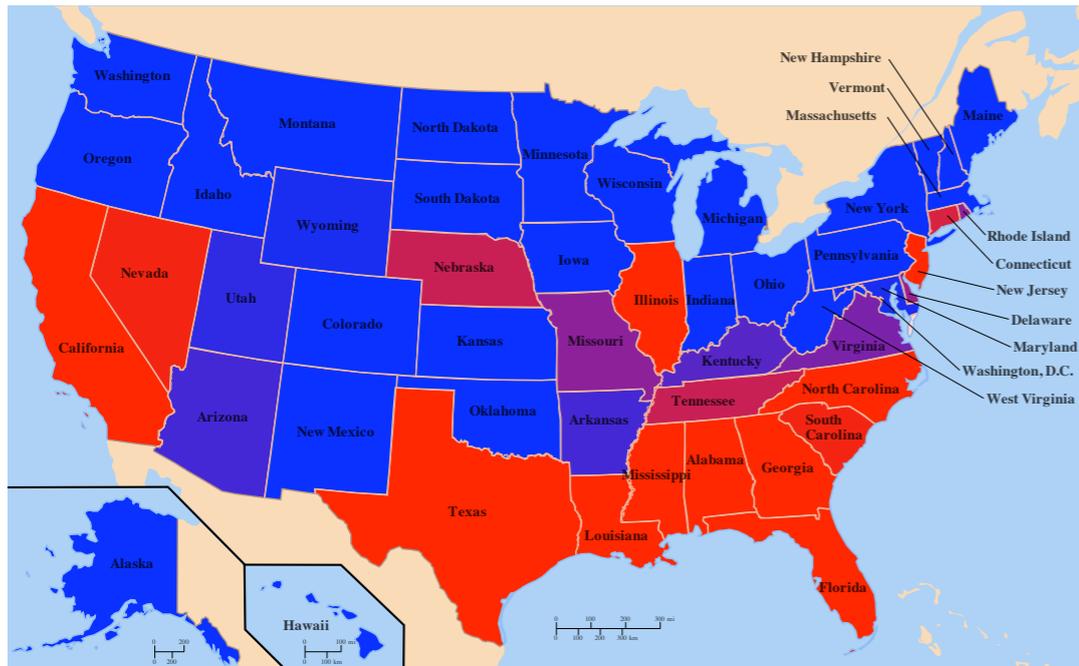
Conclusion

Beyond binary: richer and richer datasets unravel the complex nature of individuals and their interactions:

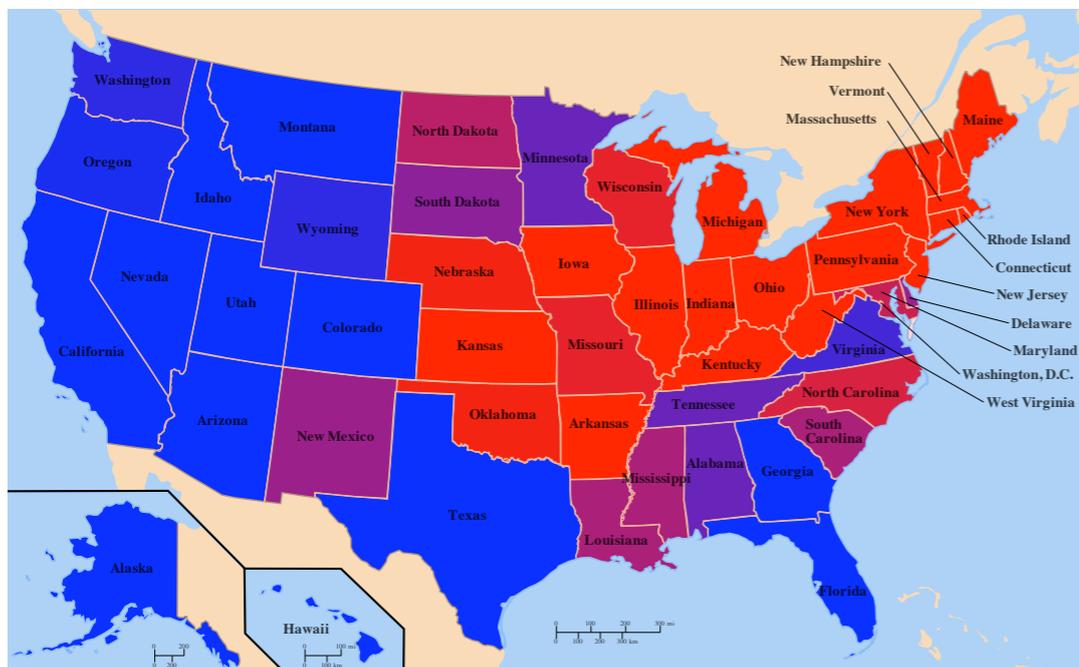
- Online games to reveal the nature of social relations and conflict dynamics
- Online experiment to survey a very large number of Facebook users with peer-reviewed personality tests, and to test the effect of personality scores on network topology

Conclusion

Relation between geography and psychology



“p-values” Extraversion



“p-values” Neuroticism

Fingerprints in online media reveal the organization of social systems, Renaud Lambiotte, Awareness (2012)

Relation between likes and personality traits

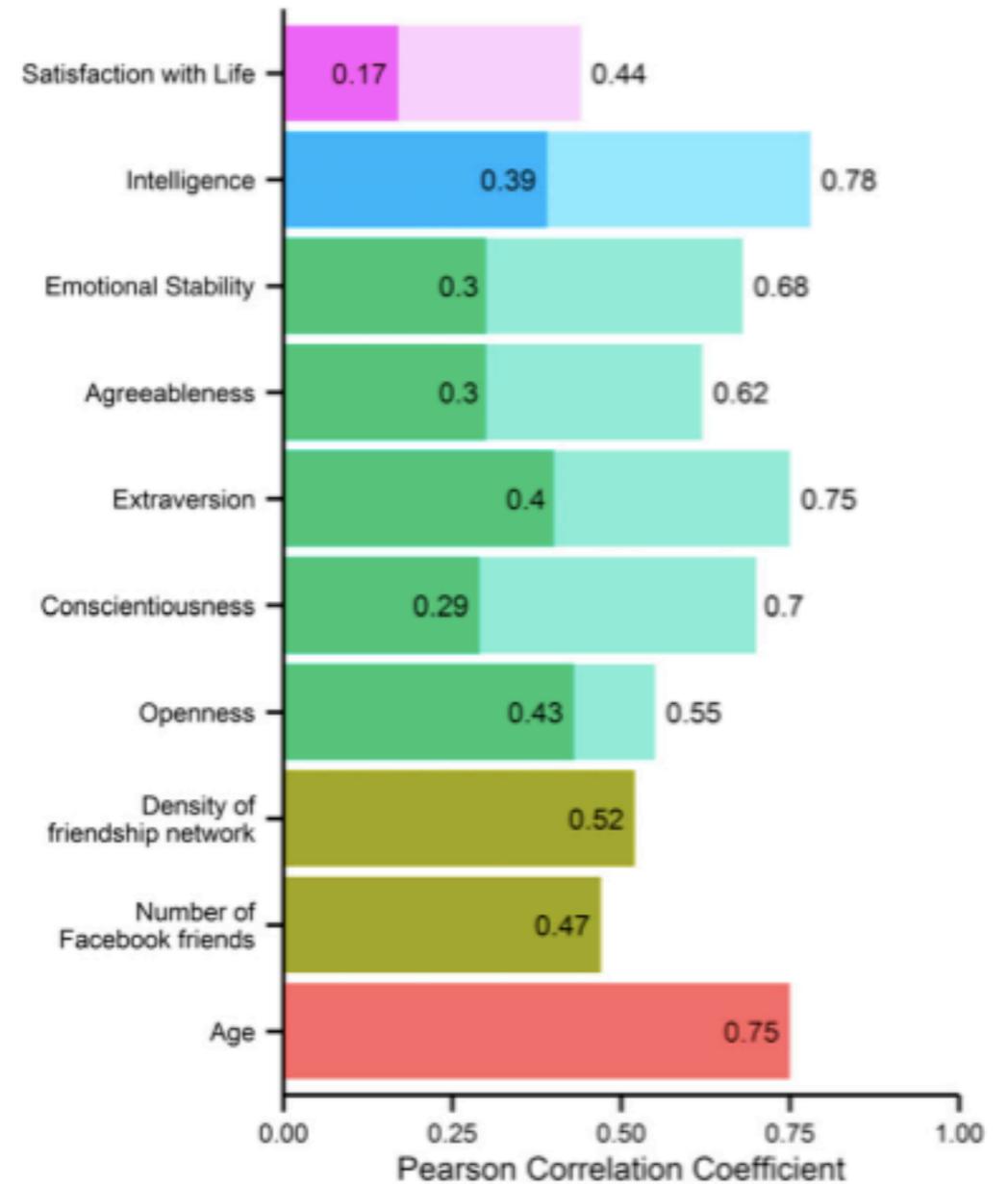


Fig. 3. Prediction accuracy of regression for numeric attributes and traits expressed by the Pearson correlation coefficient between predicted and actual attribute values; all correlations are significant at the $P < 0.001$ level. The transparent bars indicate the questionnaire's baseline accuracy, expressed in terms of test-retest reliability.

Private traits and attributes are predictable from digital records of human behavior. Michal Kosinski, David Stillwell, and Thore Graepel, PNAS 2013