Structure and tie strengths in a large-scale social network

J.-P. Onnela[1,2], J. Saramäki[2], J. Hyvönen[2], G. Szabó[3,4]
D. Lazer[5], K. Kaski[2], J. Kertész[6], A.-L. Barabási[3,4]

[1] Physics Department & Saïd Business School, University of Oxford, UK
[2] Laboratory of Computational Engineering, Helsinki University of Technology, Finland
[3] Center for Complex Networks Research, University of Notre Dame, USA
[4] Dana-Farber Cancer Institute, Harvard University, USA
[5] John F. Kennedy School of Government, Harvard University, USA
[6] Dept. of Theoretical Physics, Budapest University of Technology and Economics, Hungary

Overview

Introduction to social networks

- Social network theory since 1930s; Physicists involved since 1990s
- Nodes _ individuals, links _ social interaction (of some type / form)
- Classics: six-degrees of separation (1960's) & strength of weak ties (1973)

- Social network paradigm in the social sciences: Social life consists of the **flow** and exchange of norms, values, ideas, and other social and cultural resources **channelled through the social network**
- Key question: Microscopic interactions _ macroscopic social systems

- Physics perspective can contribute: Network flow & statistical mechanics

- Uncovering structure and function of social networks has been severely constrained by the difficulty of mapping social relations (interactions)

Introduction to social networks

- Traditional approach:
  - Data from questionnaires; $N \approx 10^2$
  - Scope of social interactions wide
  - Strength based on recollection

- New approach:
  - Electronic records of interactions; $N \approx 10^6$
  - Scope of social interactions narrower
  - Strength based on measurement

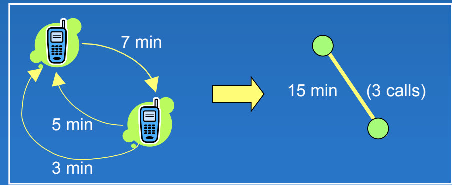
COMPLEMENTARY APPROACHES

- Mobile phone call network as a proxy for the underlying social network; Empirical study of a large scale **one-to-one weighted** social interaction network
  - Understanding of large-scale social networks (media multiplexity)
  - Foundation for social network models
  - Potential applications: epidemiology, collective behaviour

1. Introduction to social networks
2. Constructing the network
3. Network statistics
4. Local structure
5. Global structure
6. Diffusion of information
7. Conclusion

Constructing the network

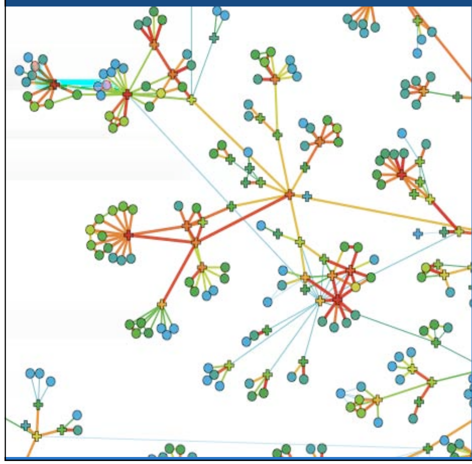
- Data from one operator in one country, covers ≈20% of population (*p=0.2*)
- Over 90% of country's population has mobile phone subscription
- Data aggregated from a period of 18 weeks (126 days)
- Records cover over 7 million private mobile phone subscriptions
- Focus on voice calls within the operator
- Require reciprocity of calls (**X→Y AND Y→X**) for a link
- Quantify tie strength (link weight) using (1) Aggregate call duration w_{ij}^D
  (2) Total number of calls w_{ij}^N



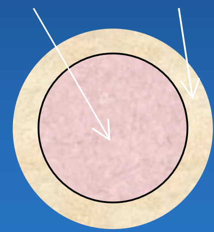
1. Introduction to social networks
2. Constructing the network
3. Network statistics
4. Local structure
5. Global structure
6. Diffusion of information
7. Conclusion

Network statistics

- Take a look at it!

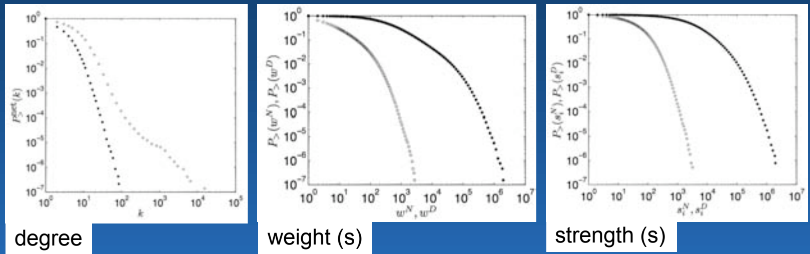


- Snowball sampling
  - Pick a node at random
  - Include all nodes within a given distance in the sample
- Bulk nodes & surface nodes



- Majority (~95%) are surface nodes
- Partial vs. full visibility wrt. neighbours

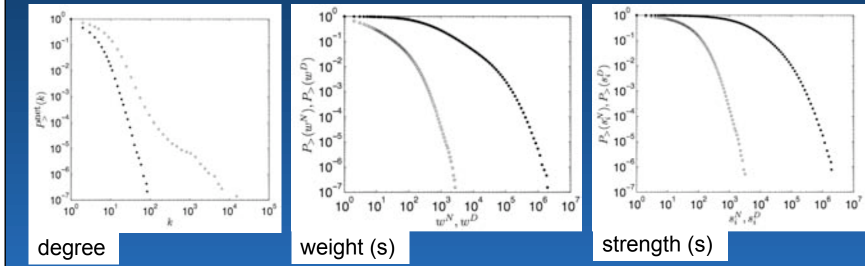
Network statistics



ALL PRIVATE !

	mean	std	max
degree k	3.3	2.5	144 (!)
weight w^N	15.4	37.3	3610
weight w^D	40.5 min	205.6 min	662.9 h
strength s^N	51.1	74.8	3644
strength s^D	134.6 min	385.9 min	690.2h

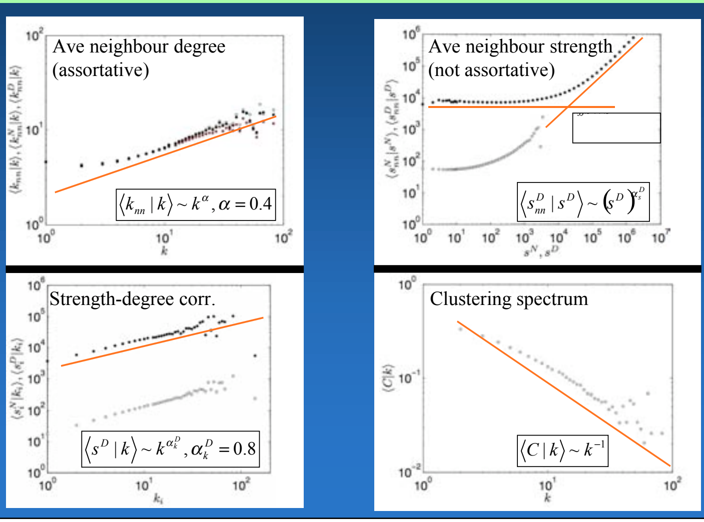
Network statistics



- Skewed degree distribution: most communicate with just a few others
- Broad link weight (tie strength) distribution
- Fat tailed tie strength distributions mainly in global transport network (fluxes in metabolic nets & packets on the Internet; local conservation)
- **No local conservation** to constrain or drive tie strengths here
- What is the relation between interaction (tie) strengths and topology?

$y = a(x+x_0)^{-\gamma}\exp(-x/x_c)$; $\{k_0 = 10.9, \gamma = 8.4, k_c = \infty\}$ $\{w_0 = 280, \beta = 1.9, w_x = 3.5\times10^5\}$

Network statistics

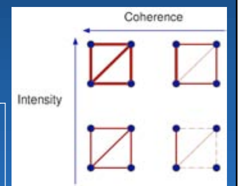


Network statistics

- Weighted measures for subgraphs (triangles, motifs, cliques):

Subgraph intensity

Subgraph coherence

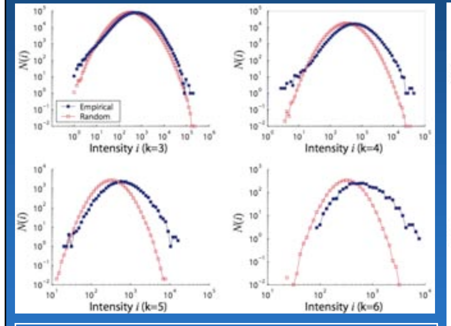


Weighted clustering coefficient

* J.-P. Onnela, J. Saramäki, J. Kertész, and K. Kaski, Intensity and coherence of motifs in weighted complex networks, *Phys. Rev. E* 71, 065103 (2005)

Network statistics

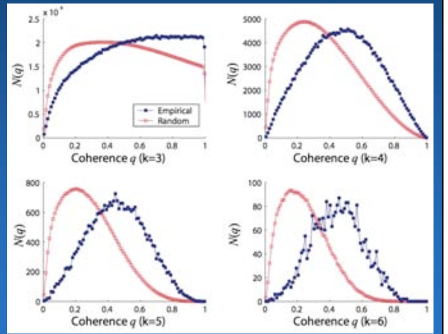
• Cliques are high in **intensity**

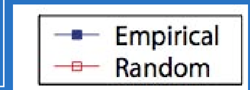


• Reference: shuffle weights & keep topology (*weight permuted reference*; H_0)

•Framework: compare distributions

• Cliques are high in **coherence**





Network statistics

• And many more in (2)

network characteristic	notation	"scaling"	see p.
degree distribution (cumulative)	$P_>(k)$	NA	6
link weight distribution (cumulative)	$P_>(w^N)$, $P_>(w^D)$	NA	7,8
node strength distribution (cumulative)	$P_>(s^N)$, $P_>(s^D)$	NA	7,8
degree–degree correlation	$\langle k_{nn} \vert k \rangle$	$\sim k^{\alpha}$, $\alpha \approx 0.4$	8,9
degree–degree correlation, w_{ij}^N–weighted	$\langle k_{nn}^N \vert k \rangle$		9
degree–degree correlation, w_{ij}^D-weighted	$\langle k_{nn}^D \vert k \rangle$		9
strength–strength correlation	$\langle s_{nn}^N \vert s^N \rangle$	$\sim (s^N)^{\alpha^N}$, see text	10
	$\langle s_{nn}^D \vert s^D \rangle$	$\sim (s^D)^{\alpha^D}$, see text	10
clustering spectrum	$\langle C \vert k \rangle$	$\sim k^{-1}$	10
weight–degree correlation	$\langle s^N \vert k \rangle$	$\sim k^{\alpha^N}$, $\alpha^N \approx 0.9$	10,11
	$\langle s^D \vert k \rangle$	$\sim k^{\alpha^D}$, $\alpha^D \approx 0.8$	10,11
strength product–degree product correlation	$\langle s_j^N s_j^N \vert k_i k_j \rangle$	$\sim (k_i k_j)^{\beta^N}$, $\beta^N \approx 0.7$	11
	$\langle s_j^D s_j^D \vert k_i k_j \rangle$	$\sim (k_i k_j)^{\beta^D}$, $\beta^D \approx 0.4$	11
weight–degree product correlation	$\langle w_{ij}^N \vert k_i k_j \rangle$	$\sim (k_i k_j)^{\gamma^N}$, $\gamma^D \approx -0.2$	11,12
	$\langle w_{ij}^D \vert k_i k_j \rangle$	$\sim (k_i k_j)^{\gamma^D}$, $\gamma^N \approx -0.1$	11,12
weight–strength product correlation	$\langle w_{ij}^N \vert s_i^N s_j^N \rangle$	$\sim (s_i^N s_j^N)^{\delta^N}$, $\delta^N \approx 0.5$	11,12
	$\langle w_{ij}^D \vert s_i^D s_j^D \rangle$	$\sim (s_i^D s_j^D)^{\delta^D}$, $\delta^D \approx 0.5$	11,12
average triangle intensity–strength correlation	$\langle \bar{i}(\Delta)^N \vert s^N \rangle$	$\sim (s^N)^{\epsilon^N}$, $\epsilon^N \approx 0.5$	15,16
	$\langle \bar{i}(\Delta)^D \vert s^D \rangle$	$\sim (s^D)^{\epsilon^D}$, $\epsilon^D \approx 0.7$	15,16
average triangle coherence–strength correlation	$\langle \bar{q}(\Delta)^N \vert s^N \rangle$	NA	15,16
	$\langle \bar{q}(\Delta)^D \vert s^D \rangle$	NA	15,16
weighted clustering–strength correlation	$\langle \tilde{C} \vert s^D \rangle$	$\sim (s^D)^{\zeta^D}$, $\zeta^D \approx 0.8$	10,16
	$\langle \tilde{C} \vert s^N \rangle$	NA	10,16
overlap–weight correlation	$\langle O \vert w^N \rangle$	NA	16-18
	$\langle O \vert w^D \rangle$	NA	16-18
overlap–cumulative weight correlation	$\langle O \vert P_c(w^N) \rangle$	NA	16-18
	$\langle O \vert P_c(w^D) \rangle$	NA	16-18
overlap–betweenness centrality correlation	$\langle O \vert b \rangle$	NA	19,20

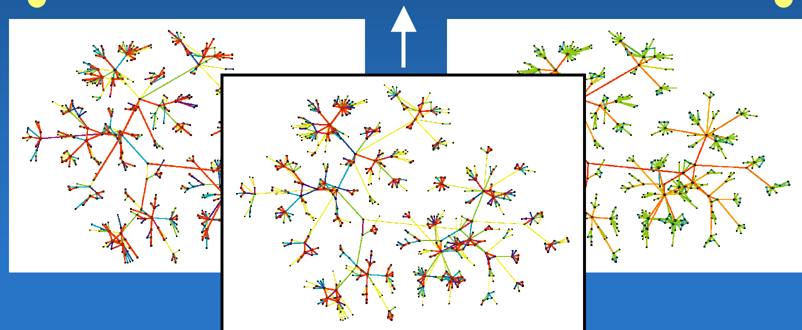
(2) J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, ***New Journal of Physics*** (2007). Pre-print physics/0702158.

Local structure

- What determines the strength of ties?
- Alternatives: (1) Dyadic hypothesis
  (2) Strength of weak ties hypothesis
  (3) Global efficiency hypothesis

Local $w_{ij} = w_{ij}$

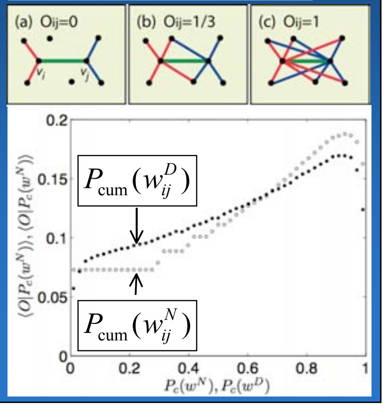
Global $w_{ij} \propto b_{ij}$



Local structure

- Weak ties hypothesis*: Relative overlap of two individual's friendship networks varies with the strength of their tie to one another (tie strength := time + reciprocity)
- Define ***overlap*** O_{ij} of edge (i,j) as the fraction of common neighbours
- Result: Average overlap increases as a function of (cumulative) link weights
- Local coupling between tie strengths and network topology, i.e. tie strength is in part driven by the structure surrounding the tie

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}}$$



* M. Granovetter, The strength of weak ties, *American Journal of Sociology* **78**, 1360 (1973).

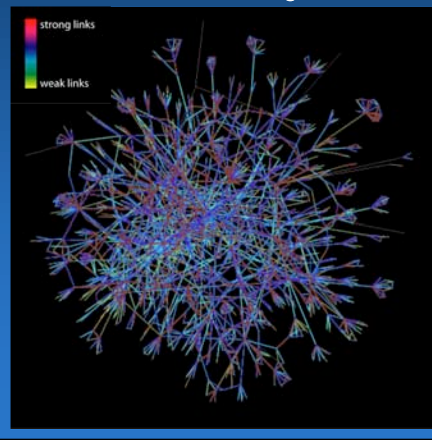
Global structure

- Probe the global role of links of different weight (w_{ij}) and local topology (O_{ij})
- Physicists' (and children's) approach: Break to learn!
- Thresholding (percolation):
  - Order links by weight (w_{ij})
  - Remove links, one by one, based on their order
- Control parameter f is the fraction of removed links
- We can move, in either direction, between the initial connected network ($f=0$) and the set of isolated nodes ($f=1$)

Global structure

Initial connected network ($f=0$), small sample

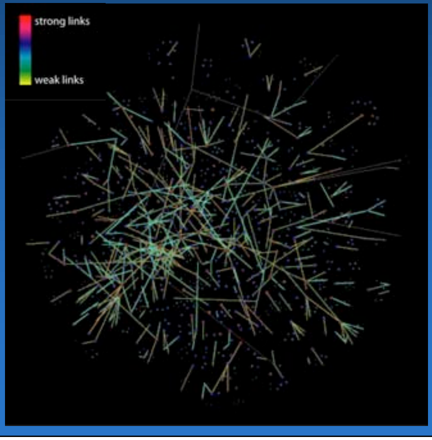
⇒ All links are intact, i.e. the network is in its initial stage



Global structure

Decreasing weight thresholded network (f=0.8)

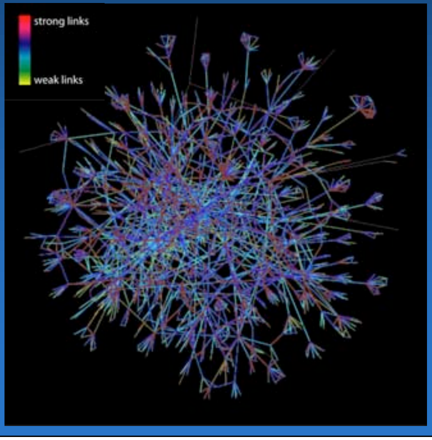
$\Rightarrow$ 80% of the strongest links removed, weakest 20% remain



Global structure

Initial connected network (f=0), small sample

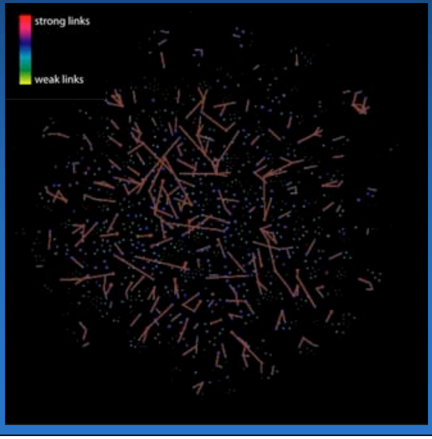
$\Rightarrow$ All links are intact, i.e. the network is in its initial stage



Global structure

Increasing weight thresholded network (f=0.8)

$\Rightarrow$ 80% of the weakest links removed, strongest 20% remain

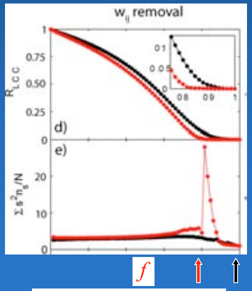


Global structure

- **Qualitative** difference in the **global** role of weak and strong links (FSS):
  - Phase transition when weak ties are removed first $f_c(\infty) \neq 1$
  - No phase transition when strong ties are removed first $f_c(\infty) = 1$
- Technological and biological networks: strong ties are structurally important
- Suggests a point of division between weak and strong links (f_c)

Existence of PT:
- Global role
- Comparison (N)
- Weak vs. strong

$w_c = P^{-1}_{cum}(0.80) \approx 27\,\text{min}$



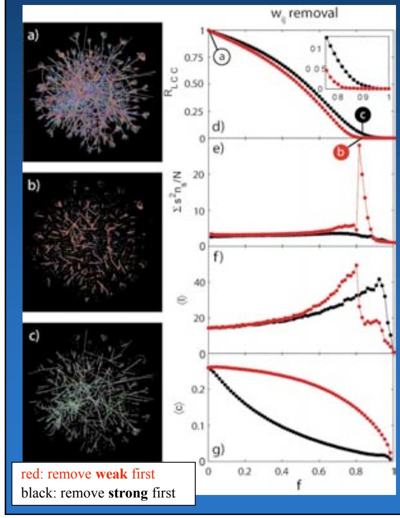
$f_c^w(\infty) = 0.80 \pm 0.04$

Order parameter R_{LCC}
- Def: fraction of nodes in LCC
- LCC is robust, collapses when $f \geq 0.80$

Susceptibility S
- Def: average cluster size (excl. LCC)
- Divergence? $\Rightarrow$ Global role of links?

red: remove **weak** first
black: remove **strong** first

Global structure



Order parameter R_{LCC}
- Def: fraction of nodes in LCC
- LCC is robust, collapses when $f \geq 0.80$

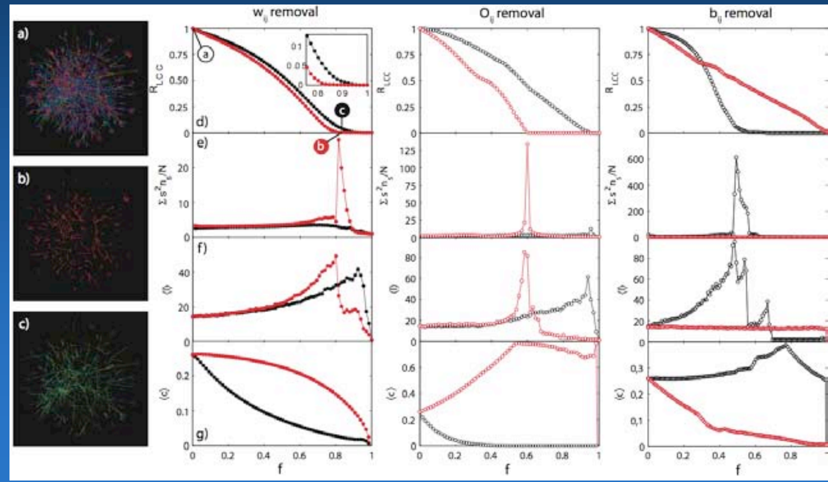
Susceptibility S
- Def: average cluster size (excl. LCC)
- Divergence? ⇒ Global role of links?

Average shortest path length $\langle l_{LCC} \rangle$
- Def: number of links along shortest path
- Remove weak links _ longer paths (!)

Average clustering coefficient $\langle C_{LCC} \rangle$
- Def: fraction of interconnected neighbours
- Removing strong links decreases $\langle C \rangle$
- WL removal 'invisible' locally; cmp. R_{LCC}

$$\left[S = \sum_{s<s_{max}} n_s s^2 / \sum_{s<s_{max}} n_s s;\ \tilde{S} = \sum_{s<s_{max}} n_s s^2 / N;\ C_i = t_i / 2k_i(k_i-1) \right]$$

Percolation properties



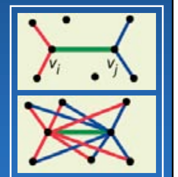
1. Introduction to social networks
2. Constructing the network
3. Network statistics
4. Local structure
5. Global structure
6. Diffusion of information
7. Conclusion

Diffusion of information

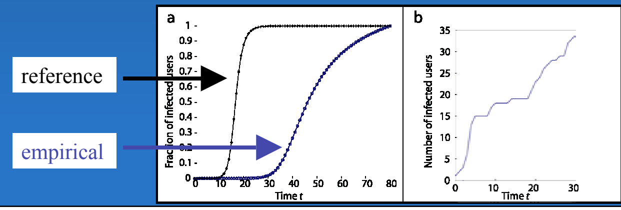
- Knowledge of information diffusion based on unweighted networks
- Use the present network to study diffusion on a weighted network: Does the local relationship between topology and tie strength have an effect?
- Spreading simulation: infect one node as in SI-model in epidemiology

(1) Empirical: $p_{ij} = a w_{ij}$

(2) Reference: $p_{ij} = a \overline{w} = \text{const.}$

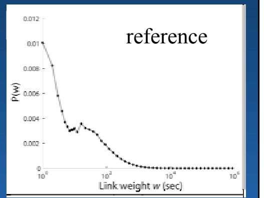


- Spreading significantly faster on the reference network
- Information gets trapped in communities in the real network

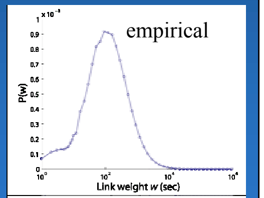


Diffusion of information

• Where do individuals get their information from?



Reference
- Transmission probability independent of link weight
- First transmissions through weak ties (Granovetter)



Empirical
- Transmission probability depends on link weight
- First transmissions through intermediate ties
- Weak ties: • access to new information ✓
  • low information transmission rate ✗
- Strong ties: • high information transmission rate ✓
  • rarely access to new information ✗

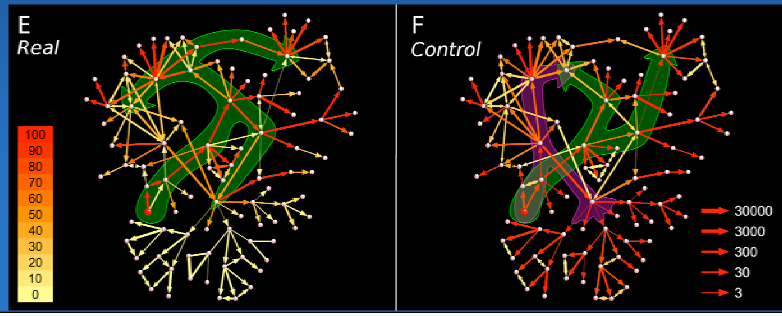
$\Rightarrow$ "weakness of weak and strong ties"

• Successful searches conducted primarily through intermediate to weak ties

P. S. Dodds, R. Muhamad, D. J. Watts, *Science* **301**, 827 (2003)

Diffusion of information

Impact on overall infromation flow in the network?
- Start spreading 100 times (large red node)
- Information flows differently due to the local organizational principle
  - (1) Empirical: information flows along a strong tie backbone (clean regions)
  - (2) Reference: information mainly flows along the shortest paths



Conclusion

• Local coupling between network topology and tie strengths
  $\Rightarrow$ Large-scale empirical verification of weak ties hypothesis

• Global role of ties
  Weak ties: maintain global integrity (phase transition)
  Strong ties: maintain local communities (no phase transition)

• First-time information diffusion occurs through ties of intermediate strength
  $\Rightarrow$ "weakness of weak and strong ties"

• Implications for modelling social networks and processes unfolding on them

(1) J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, ***PNAS* 104**, 7332 (2007).
(2) J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. de Menezes, K. Kaski, A.-L. Barabási and J. Kertész, ***New Journal of Physics*** (2007). Pre-print physics/0702158.
• See ***Science* 314**, 914 (2006).
• See http://www.physics.ox.ac.uk/users/Onnela/