Redes Complexas:
teoria, algoritmos e aplicações em computação

-Bloco #8-

``Flow of Information in networks”

Virgílio A. F. Almeida

Outubro de 2009

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais

factors influencing diffusion

- network structure (unweighted)
  - density
  - degree distribution
  - clustering
  - connected components
  - community structure

- strength of ties (weighted)
  - frequency of communication
  - strength of influence

factors influencing diffusion: proximity

Scalable Proximity Estimation and Link Prediction in Online Social Networks –ACM-IMC 2009

- A central concept in the computational analysis of social networks is *proximity measure, which quantifies the closeness or similarity* between nodes in a social network. Proximity measures form the basis for a wide range of important applications in social and natural sciences (*e.g., modeling complex networks [6, 13, 25, 42]),* business (*e.g., viral marketing [23], fraud detection [11]), information* technology (*e.g., improving Internet search [35], collaborative*filtering [7]), computer networks (*e.g., constructing overlay networks* [45]), and cyber security (*e.g., mitigating email spams [22],* defending against Sybil attacks [56]).

Studies influencing diffusion

1. The Strength of Weak Ties MS, Granovetter,1973

2. The Structure of Information Pathways in a Social Communication Network. Kossinets et al, 2008

Artigo 1:

The Strength of Weak Ties

MS **Granovetter** - American Journal of Sociology, 1973 - UChicago Press

www.si.umich.edu/~rfrost/courses/SI110/readings/In_Out_and_Beyond/Granovetter.pdf

Artigo 2:

G. Kossinets, J. Kleinberg, D. Watts.

The Structure of Information Pathways in a Social Communication Network.

Proc. 14th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, 2008.

http://www.cs.cornell.edu/home/kleinber/chrono.html

The Strength of Weak Ties

Mark S. Granovetter

The American Journal of Sociology,
1973

Introduction

- "One of the most influential sociology papers ever written" (Barabasi)
  - One of the most cited (*Current Contents, 1986)*

- Interviewed people and asked:
  "How did you find your job?"
  - Kept getting the the same answer:
    "through an acquaintance, not a friend"

how does strength of a tie influence diffusion?

- M. S. Granovetter: *The Strength of Weak Ties*, AJS, 1973:

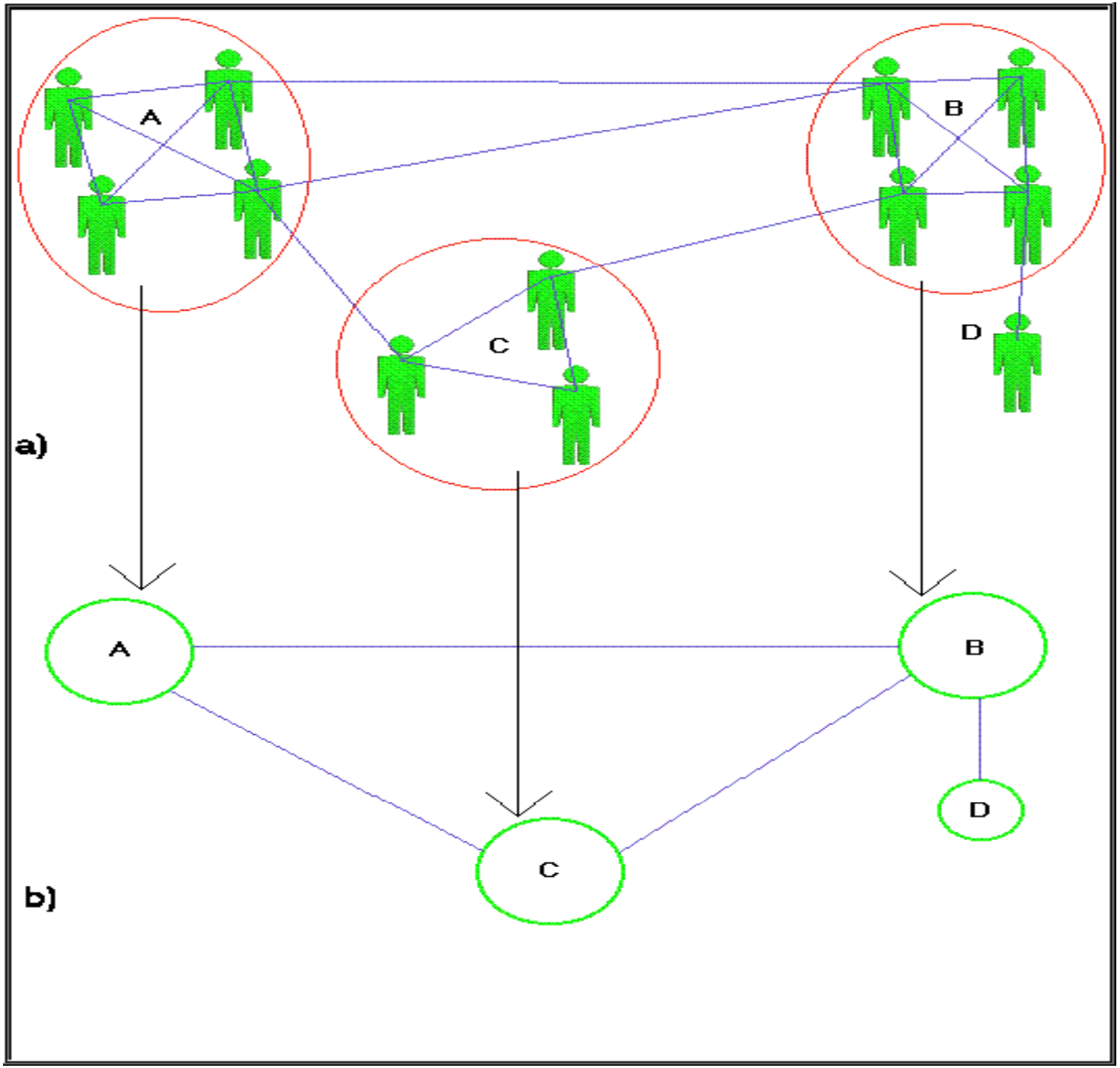
- finding a job through a contact that one saw
  - frequently (2+ times/week) 16.7%
  - occasionally (more than once a year but < 2x week) 55.6%
  - rarely 27.8%

- but… length of path is short
  - contact directly works for/is the employer
  - or is connected directly to employer

Context

- Lots of studies of macro patterns
  - Social mobility, community organization
- Data and studies for micro behavior
  - Interactions within small groups

- Limited understanding of how micro behavior translates into macro patterns


A
B
C
D
a)
A
B
C
D
b)

Network Analysis

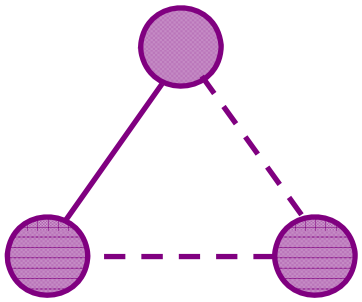
- Analysis of the interaction network
  - bridge the gap between micro and macro
- Interaction network
  - *Nodes:* People
  - *Edges:* Between people with a social relationship
    - *Weight:* strength of connection
      Quantize to either “weak” or “strong”

Basic argument

- Classify interpersonal relations as “strong”, “weak”, or “absent”
  - Strength is (vaguely) defined as “a (probably linear) combination of…
    - the amount of time,
    - the emotional intensity,
    - the intimacy (mutual confiding),
    - and the reciprocal services which characterize the tie
  - Negative and/or asymmetric ties (e.g. enemies or relations with power imbalance) are brushed aside for now

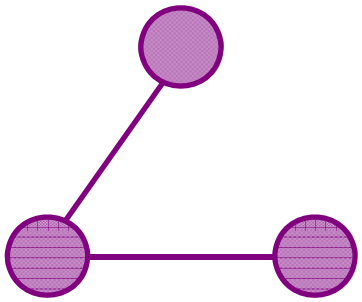
Basic argument (cont.)

- The stronger the tie between two individuals, the larger the proportion of people to which they are *both* tied (weakly or strongly)
  - In the extreme case, two people that are *always* together will be tied to the same individuals



Forbidden triad

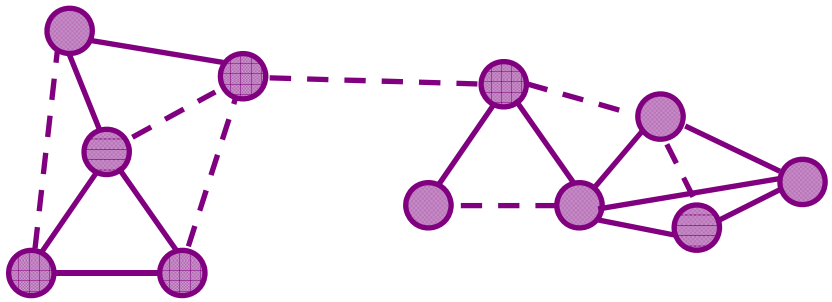
- If person A has a **strong** tie to both B and C, then it is unlikely for B and C not to share a tie.
  - Granovetter (admittedly) exaggerates and supposes such a triad *never* occurs



Bridges

- A bridge is “a line in a network which provides the *only* path between two points”

- Therefore, if the previous triad is in fact absent, *no strong tie is a bridge*
  - In other words, *all bridges are weak ties!*
  - (realistically, bridges can be *local* rather than *global*, but still weak)



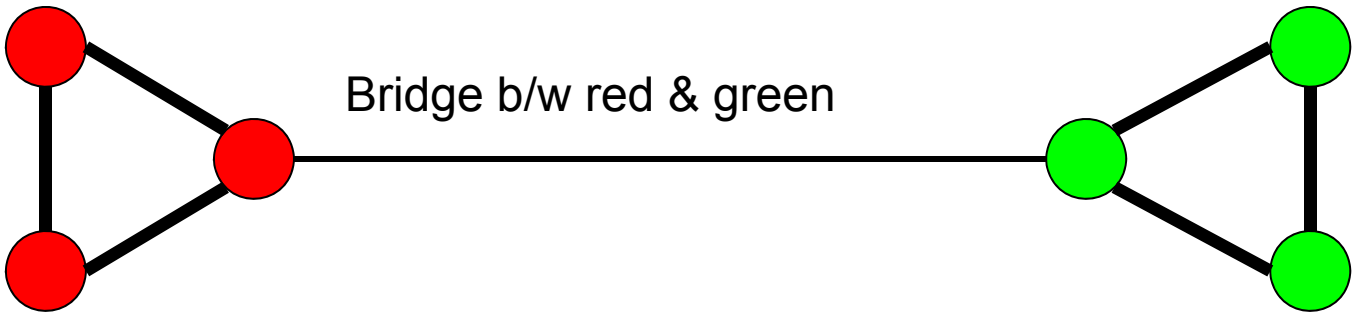
Strength of weak ties

- "Intuitively speaking, this means that whatever is to be diffused can reach a larger number of people, and traverse greater social distance (i.e., path length), when passed through weak ties rather than strong."

- Consequences
  - Diffusion of information (rumours, innovations, getting a job!)
    - Homophily
  - Group cohesion and trust
  - Traversal of networks and node coverage

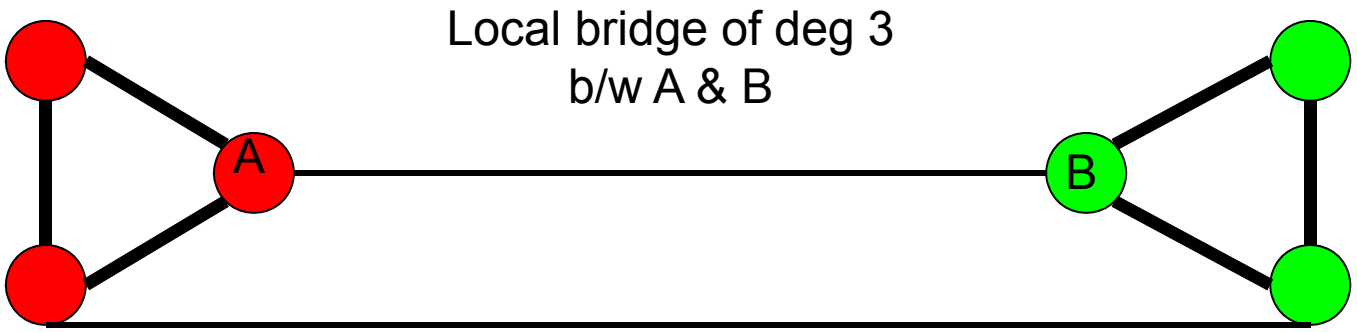
Bridges

- *Bridge:* An edge that is part of every path between two nodes



Local Bridges

- *Local Bridge of degree N:* An edge that is part of every path of length less than N
- Generalization of a bridge



Bridges

- Bridges allow diffusion of information between otherwise disconnected communities.
- Local bridges bring otherwise distant communities together

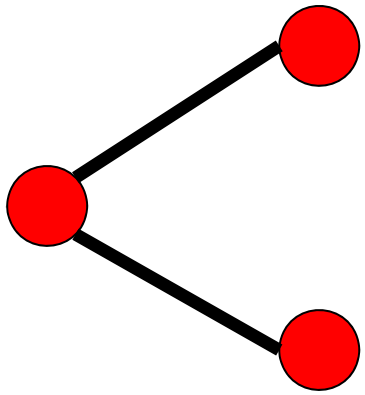
- “Bridge” concept provides an important piece of the micro => macro puzzle
  - What sort of relationships act as bridges?

Granovetter Transitivity

- The stronger the tie between A and B, the larger the overlap in their relationship circles
- Strong tie =>
  - lots of time together => lots of opportunity for B to meet the A's friends
  - similarity => greater chance that B will be "compatible" with A's friends
  - physiological need for congruence =>
    B will have a natural affinity for A's friends, based on A's opinion of them

Forbidden Triad

- This triad will resolve to a fully connected triad
  - New edge need not be strong
- Alternate: Any time strong tie A-B exists, then all of A's strong ties will be at least weakly connected to B
- Supported by evidence



All Bridges are Weak Ties!

- Proof:
  - If A-B and A-C are strong, then forbidden triad implies that B-C is at least weak
  - If A-B is deleted, then A can still reach B via A-C-B
  - Small corner case: if both nodes have only a strong edge to each other, and no other strong edges, than it is a bridge
    - Unlikely in reality
- All local bridges are also weak ties
  - Proof is identical

Implications

- Removal of weak ties raises path lengths more than removal of strong ties
- Assume: probability of info passing successfully between two nodes
  - is proportional to the number of paths connecting the two nodes
  - is inversely proportional to length of those paths
- Conclusion: Removal of a weak edge damages the connectivity more than the removal of a strong edge

Evidence

- Junior High Experiment:
  (Rapoport and Horvath, 1961)
  - Student writes down an ordered list of 8 friend
  - Pick a random starting student
  - Breadth first search on 1st and 2nd friends
  - Count number of students seen after each cycle
  - Repeat using 3/4th, 5/6th, 7/8th
- Largest number of people reached by using 7/8th, smallest using 1/2nd

Community Effects

Community Co-ordination

- Imagine a community organizing to defeat a common threat
- Requires organization and leadership
  - Leadership requires trust in the leaders
- Trust is difficult without a connection

Community Co-ordination

- Without weak links, community exists as a set of strongly connected, but disjoint cliques
  - No one suitable to act as a leader for all
- Example: Boston West End
  - Connections were mainly family-based
  - Few ways for weak links to be formed

Individual Effects

Access to Resources

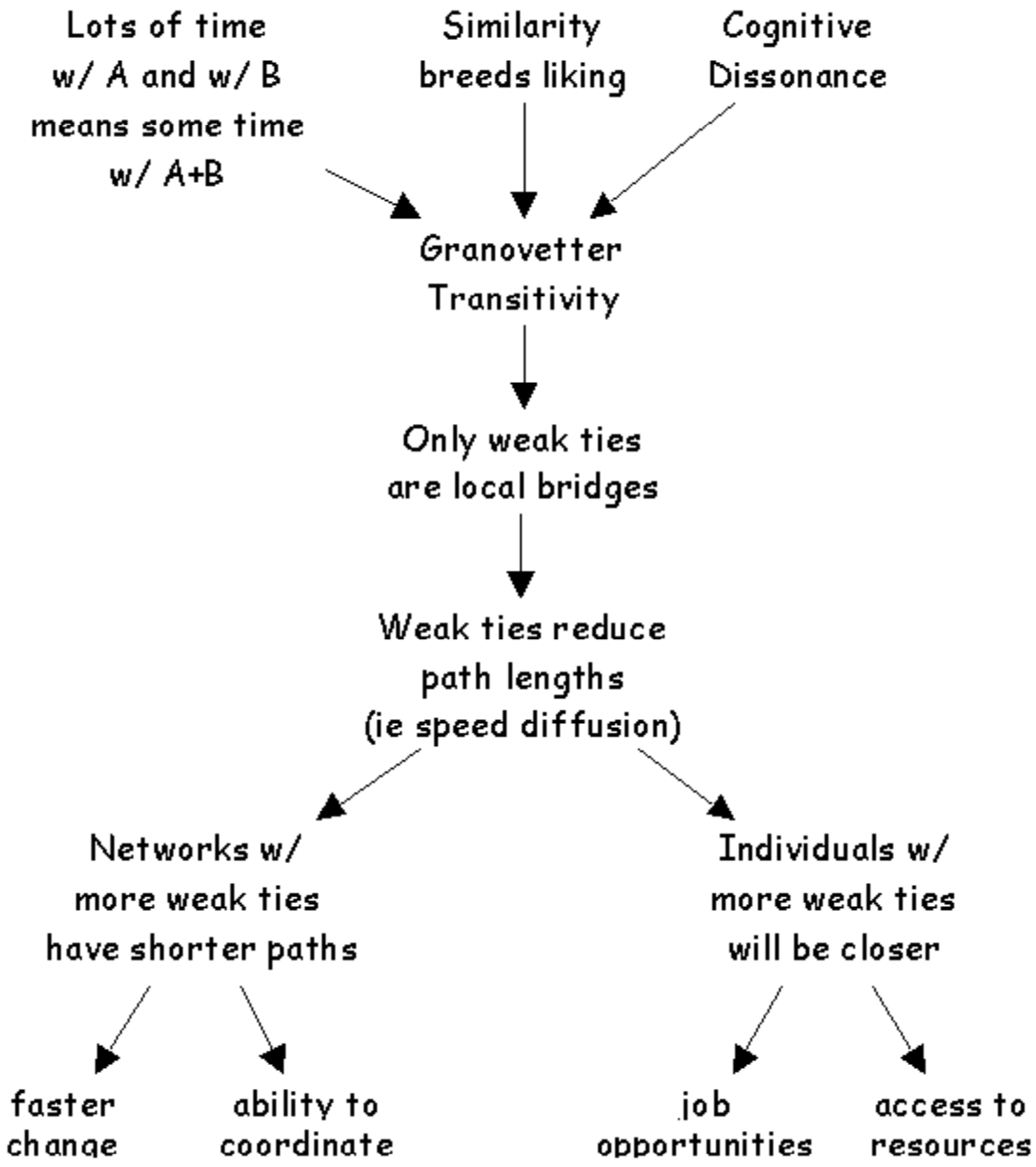
- Our weak ties are with people whose ties are with those socially distant to us.
  - Weak ties bring us knowledge of our community not available through friends
- Many weak ties => more access to wider community's ideas, resources, etc.
- Few weak ties => little information of outside world

Access to Resources

- Example: Academic Hiring
  - School's reluctance to hire your own PhD's
  - Want to prevent "intellectual inbreeding"

Finding a Job

- Do leads for new jobs come through strong or weak contacts?
  - Strong: More motivation to help you, since they know you better
  - Weak: Likely less overlap with leads you can easily get elsewhere
- Study by author shows that weak wins
  - Most job referrals come through those who we see rarely: old school friends, former co-workers, etc.



http://www.analytictech.com/networks/weakties.htm

Applications

- What is in it for **Online** Social Networks?

- "Weak ties are strong" is a valuable insight for…
  - Information diffusion
  - Threat edges
  - Network crawling
  - …

- But the micro-macro level bridge has deeper consequences
  - We take it for granted now, but it's a key assumption behind several of the papers we've read recently

Conclusions

- Personal relationships (micro) bound to large-scale social structure (macro)
- Opposite to what you might expect:
  - Weak personal relationships bind communities together
  - Exclusively strong ties lead to global fragmentation

Maintained Relationships on Facebook by cameron 2009

Measuring Networks on Facebook

To try and answer questions about network size on Facebook, we looked at the communications of a random sample of users over the course of 30 days. We defined networks in 4 different ways:

- **All Friends:** the largest representation of a person's network is the set of all people they have verified as friends.
- **Reciprocal Communication:** as a measure of a sort of core network, we counted the number of people with whom a person had had reciprocal communications, or an active exchange of information between two parties.
- **One-way Communication:** the total set of people with whom a person has communicated.
- **Maintained Relationships:** to measure engagement, we took the set of people for whom a user had clicked on a News Feed story or visited their profile more than twice.

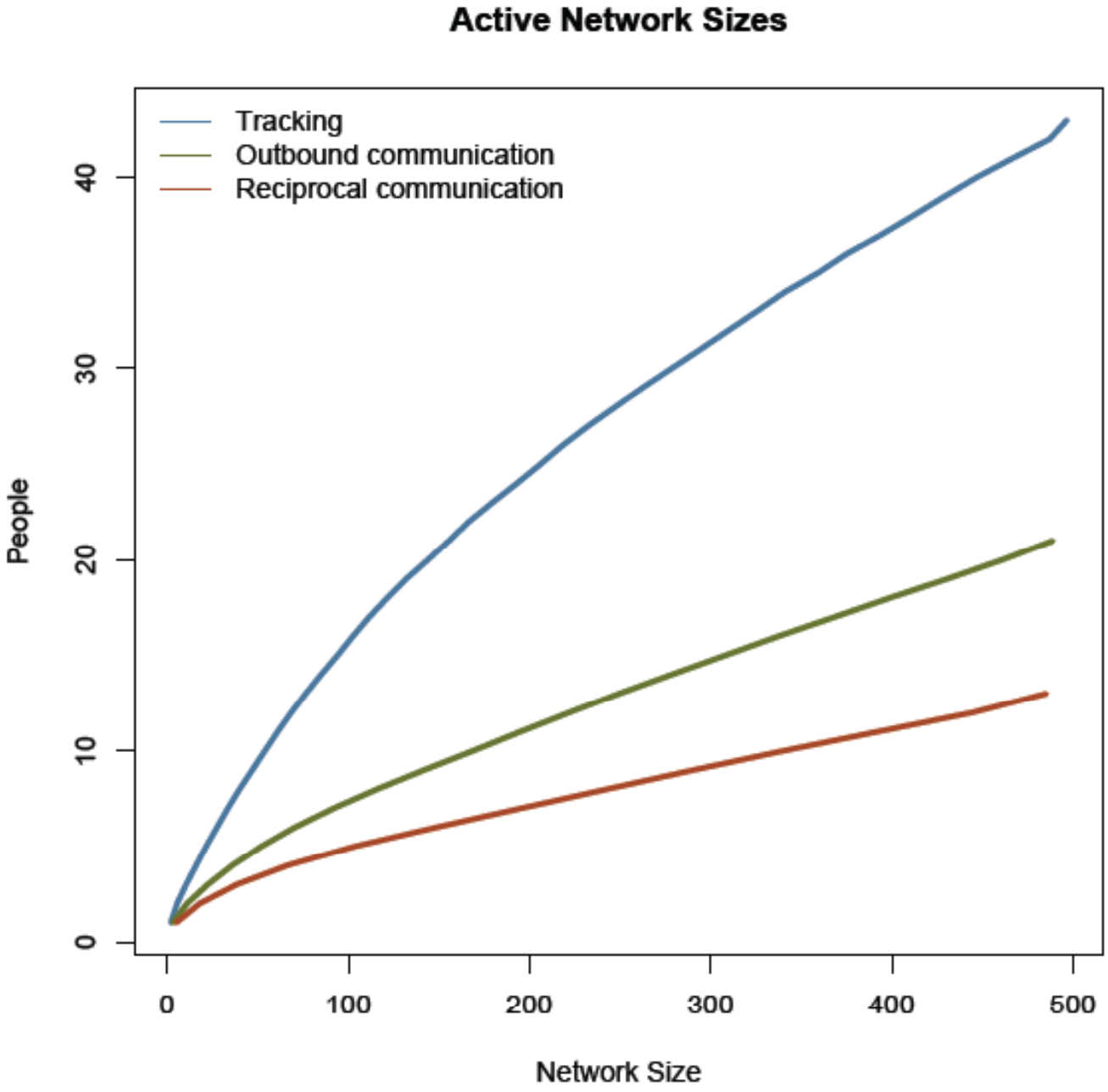
For each users we calculated the size of their reciprocal network, one-way network and network of maintained relationships, and plotted this as a function of the number of friends a user has. As Andreas mentions in his blog post about the article, the visualization (shown below) did not make it into the article, but presents a pretty clear picture of the relationship between these types of communication.

Tie Strength and Social Media:

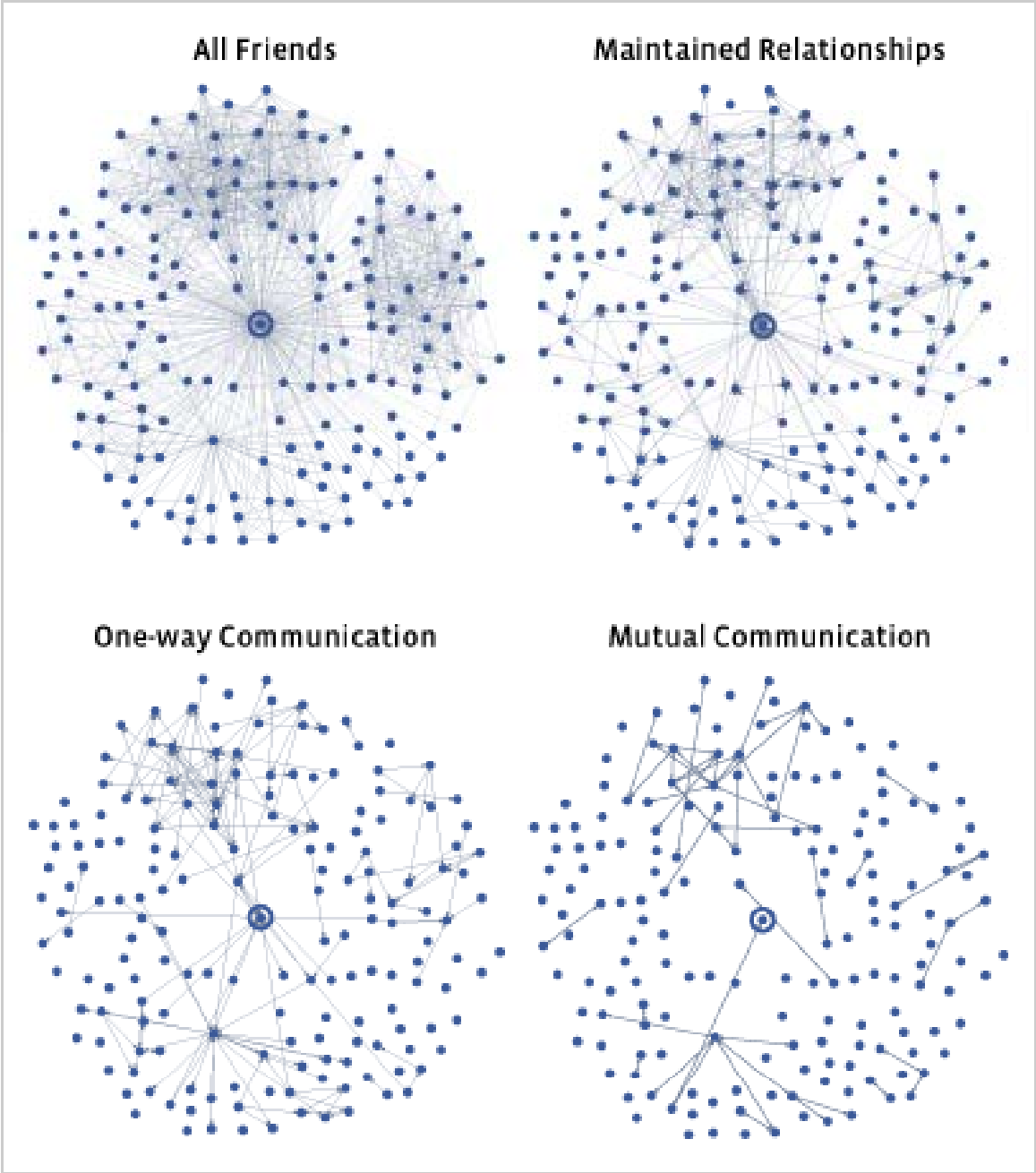
facebook and twitter

PEOPLE MAINTIAN HUNDREDS OF FRIENDSHIP LINKS: HOW MANY OF THESE CORRESPOND TO STRONG TIES THAT INVOLVE FREQUENT CONTACT, AND HOW MANY OF THESE CORRESPOND TO WEAK TIES THAT RAE ACTIVATED RELATIVELY RARELY**?**

Maintained Relationships o Facebook by cameron



Maintained Relationships on Facebook by cameron



Twitter under the microscope, Huberman 2009

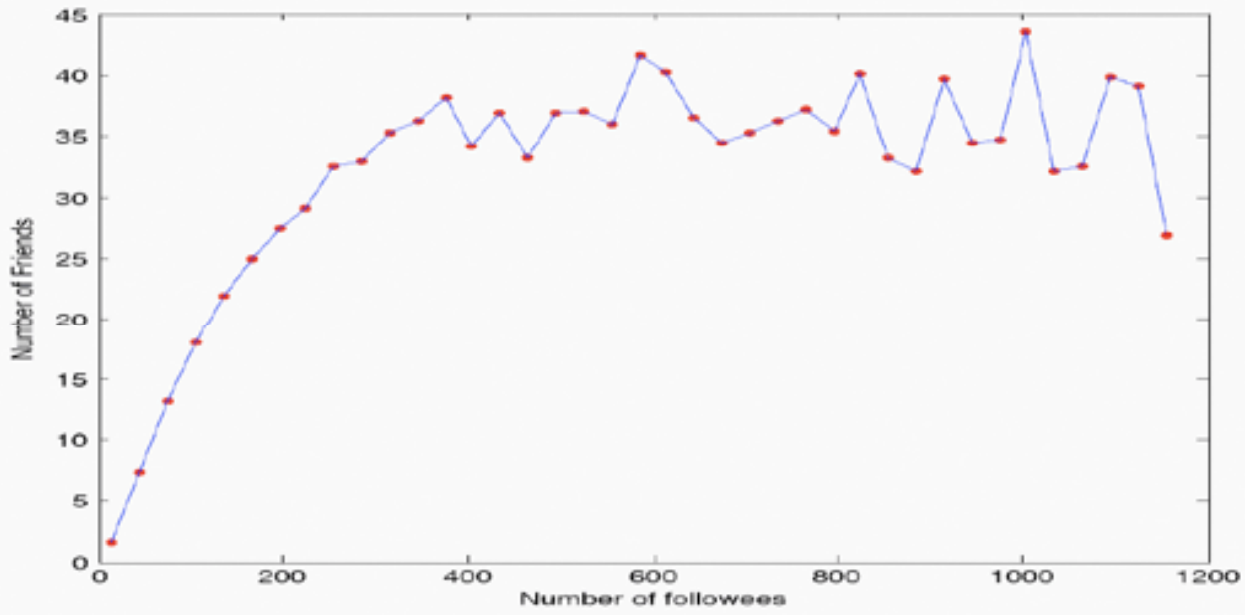
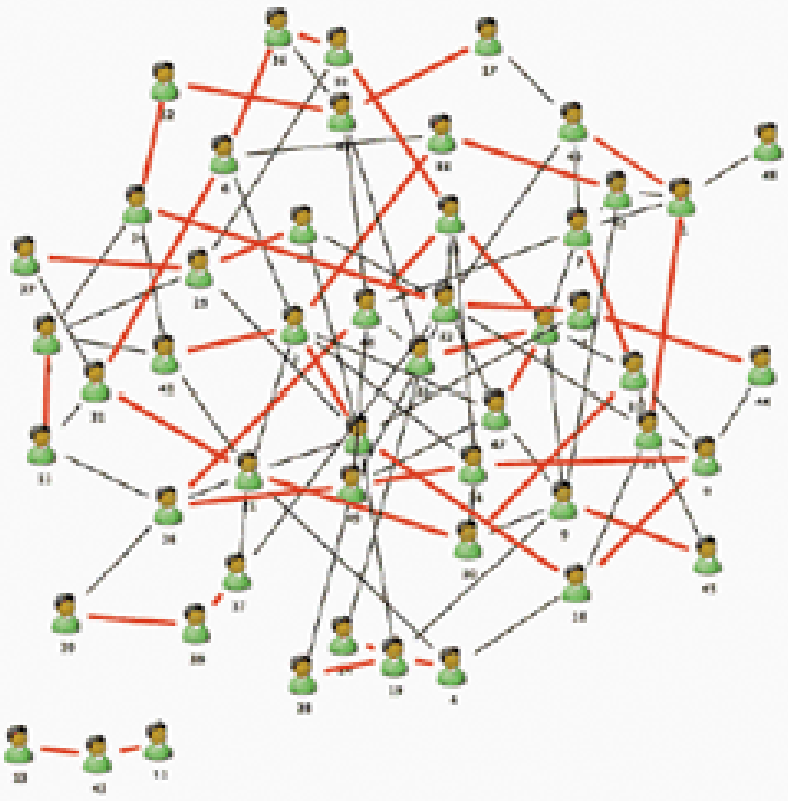
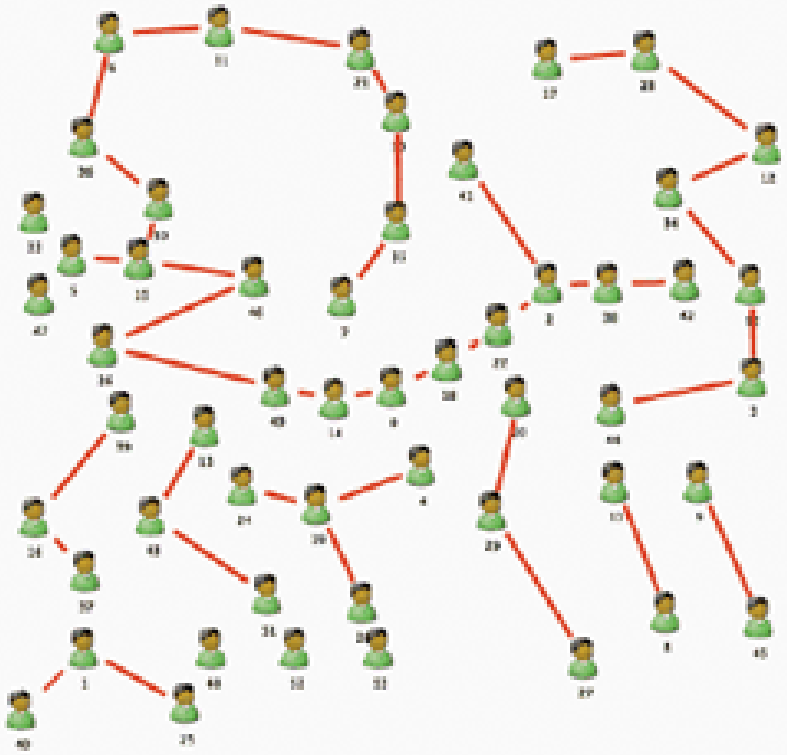


Figure 4: Number of friends as a function of the number of followees. The total number of friends saturates while the number of followees keeps growing due to the minimal effort required to add a followee.

Twitter under the microscope, Huberman 2009



(a) All links are declared followees and the red links are actual friends.



(b) After removing the black links and reorganizing the network look simpler than before. This is the hidden network that matters the most.

Structure of Information Pathways in a Social Communication Network

Gueorgi Kossinets Jon Kleinberg
Duncan Watts

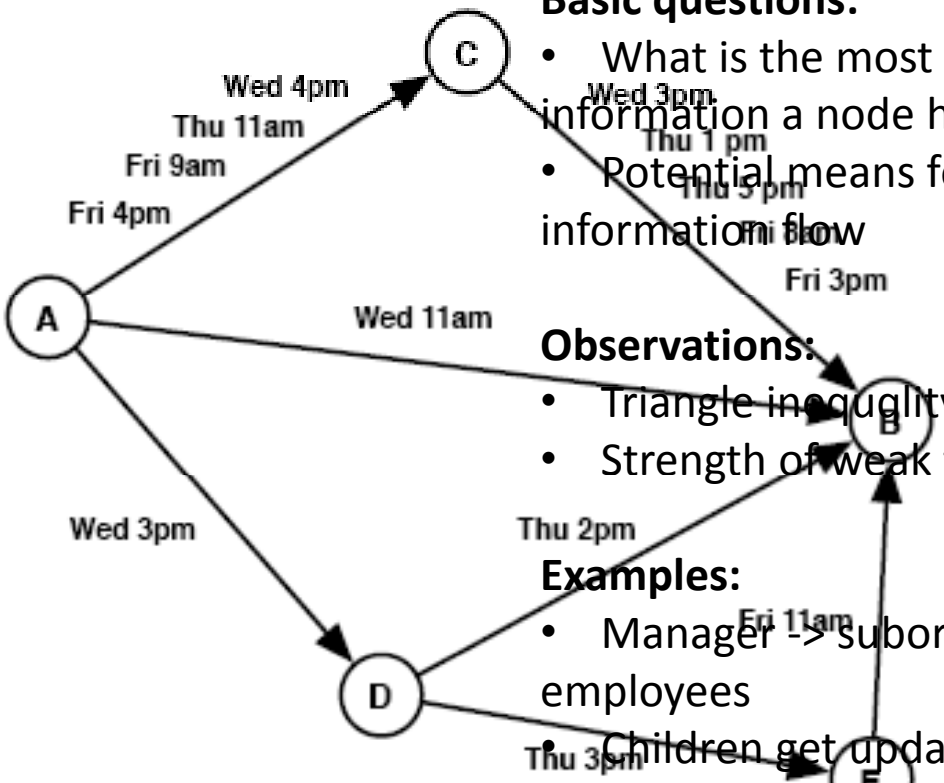
Goals

- Study temporal dynamics of communication
- Infer structural measures about the network based on the communication patterns
- Illustrate temporal notion of “distance” based on information flow in social networks
- Out-of-date and indirect paths

Communication Patterns

- Social networks consists of discrete communication acts
  - Rather than links tying nodes together
  - People at intermittent times communicate & exchange information
- Event-driven:
  - Triggered by a particular action
  - Information flows like a cascade
- Systemic:
  - Circulates information continuously through the network
  - Rhythm of everyday communication in a community
  - Governs the rate at which people communicate and remain up-to-date about each other
  - Allows information to piggy-back

Example



Basic questions:

- What is the most up-to-date information a node has about others
- Potential means for fastest information flow

Observations:

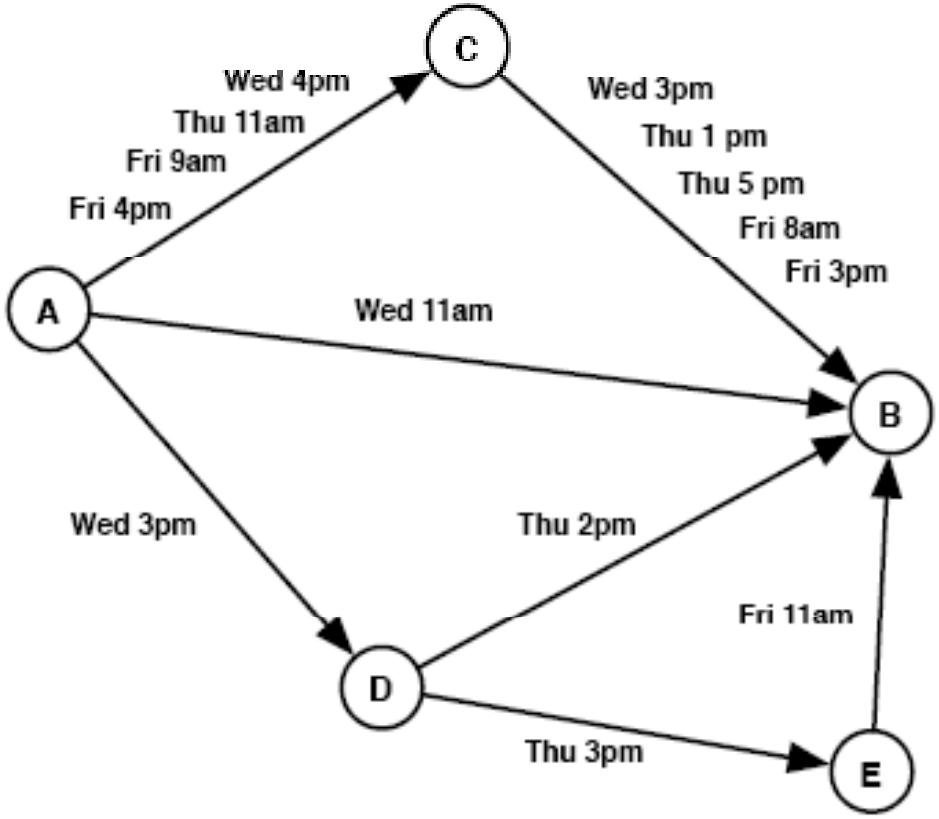
- Triangle inequality
- Strength of weak ties

Examples:

- Manager -> subordinates -> employees
- Children get updates from their parents about each other

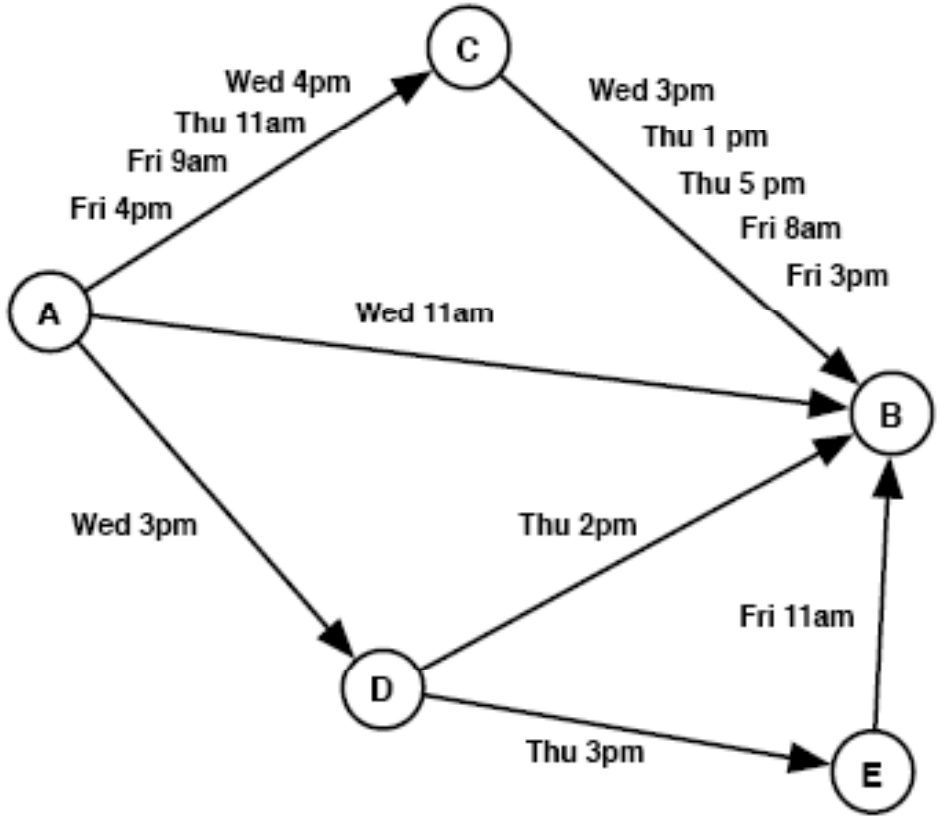
Out-of-date Views

- *v*'s view of *u* at time *t* [$\phi_{v,t}(u)$]: largest *t'* such that information originating at u at time *t'* reached *v* by *t*
- Vector clock: $\phi_{v,t} = \{\phi_{v,t}(u):u \in V\}$
  - $\phi_{B,Fri5pm} =$
- Information Latency: $t - \phi_{v,t}(u)$



Out-of-date Views

- *v*'s view of *u* at time *t* [$\phi_{v,t}(u)$]: largest *t'* such that information originating at u at time *t'* reached *v* by *t*
  - E.g., at Fri 5pm, *B*'s view of *A* is Fri 9 am,
- Vector clock: $\phi_{v,t} = \{\phi_{v,t}(u):u \in V\}$
  - $\phi_{B,Fri5pm}$ =Fri 9am, Fri 3pm, Thu 3pm, Fri 11am
- Latency: t- $\phi_{v,t}(u)$
- 8 hours
- 2 hours
- 26 hours
- 6 hours



Goals

- Out-of-date views & latency

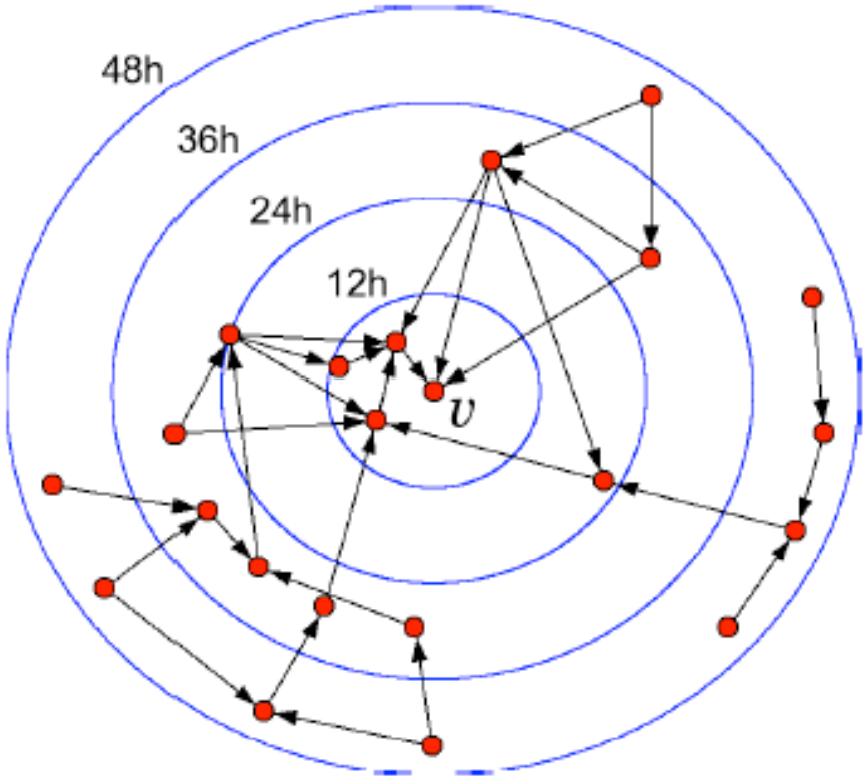
- Sparse set of pathways along which the information flows the quickest; *Backbone*

- Email dataset of 8000 faculty, staff, and students over 2 years
  - Chosen people with high activity: 1 email per hour
  - *{u, v, t}*

- Dataset Preprocessing – practical issues
  - q-fraction of highest volume email users (q=0.20)
  - Elimination of message with large recipient lists (> 5)

Latency

How far out-of-date the rest of the world is with respect to v



Ball Size = number of nodes in that ball

Aplicações:

Blogs influentes?

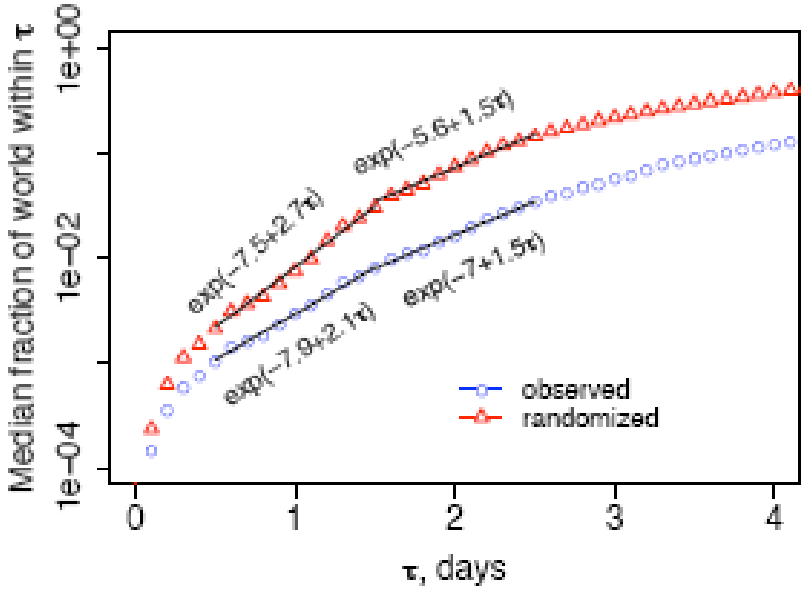
Controle industrial?

Bolsa-investimento?

Campanha politica?

Vacinação contra virus????

Latency



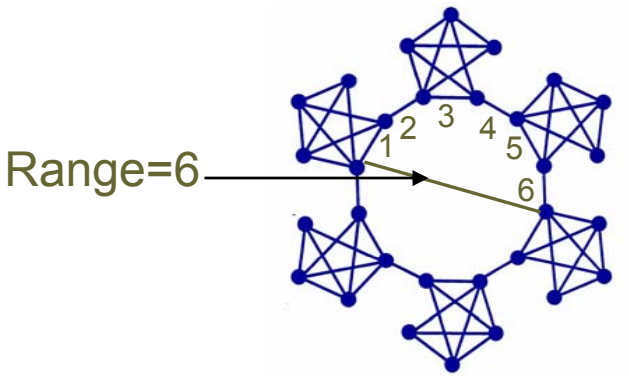
- General observation
  - Ball sizes grow piecewise exponentially
  - Choice of nodes: randomized selection of recipient – difference lies in the initial few hours
    - In randomized approach 50 people fall into 36 hr ball vs. 12 in the normal approach
    - Which means information travels much faster in a randomized epidemic fashion
- Very different view from “degrees of separation”
  - Average degree of separation: 3
  - To cover half of the network it takes about 7.5 days ball radius

Granovetter: Strength of Weak Ties

- Social ties vary in strength
  - Frequency of interaction
  - Trust, commitment, attachment
- Strong ties form short cycles
- Strength of weak ties is their range

Does Tie Strength Decline with Range?*

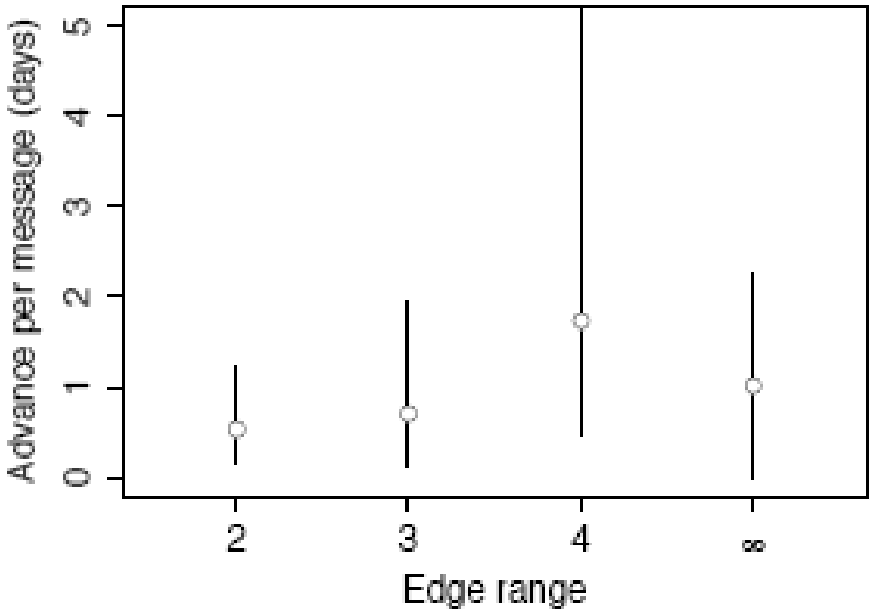
- Range is the path length traversed



- Tie strength
  - Number of emails, phone calls
  - Email lag times
  - Duration of phone calls

Strength of Weak Ties

- Important information travel through long-range edges [Granovetter 1973]
- **Edge Range:** length of shortest alternate path between two nodes
- Edge range > 2 – weak ties
- **Advance in clock:** sum of difference between vector clock of a node before and after a message from other node is received



Backbone Structures

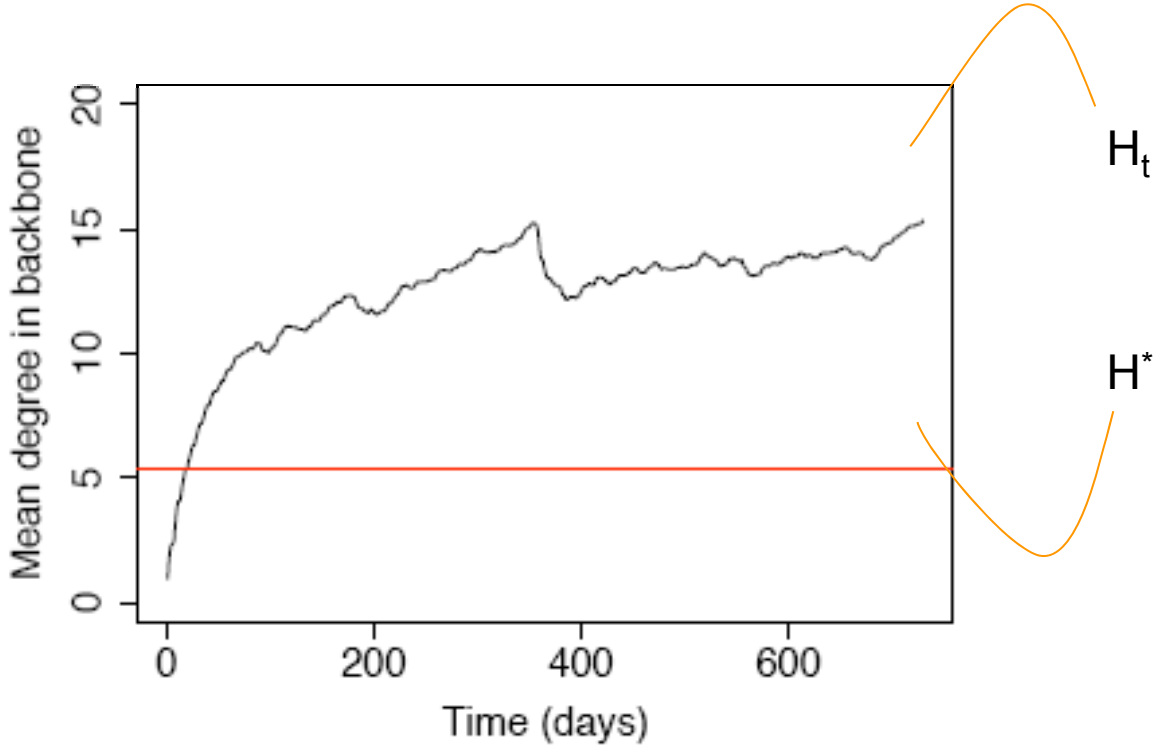
- Network backbone: the subset of edges in the social network that are not bypassed by a faster alternate path
- Essential edge: an edge is essential if *w*'s most up to date view from *v* comes from direct edge instead of sequence of updates along an indirect paths
  - E.g. (C,B) and (E,B) are essential at Fri 5pm; (A,B) and (D,B) are not
- Union of all essential edges → backbone at time t → instantaneous backbone (H_t)

Backbone Structures

- Aggregate Backbone (H^*)
  - v sent ρ messages to w in [0,T] time interval
  - Delay of edge (v,w) $\delta = T/\rho$
  - Path of minimum delay between v and w → fastest information could flow between v and w
  - Essential edge: no faster indirect path

Sparse Degree Distribution

- Communication skeleton is a dense graph whereas the aggregate backbone is sparse
- Average degree of the email network: 50



- From the point of information flow, significant majority of edges are bypassed with faster indirect paths

Sparse Degree Distribution

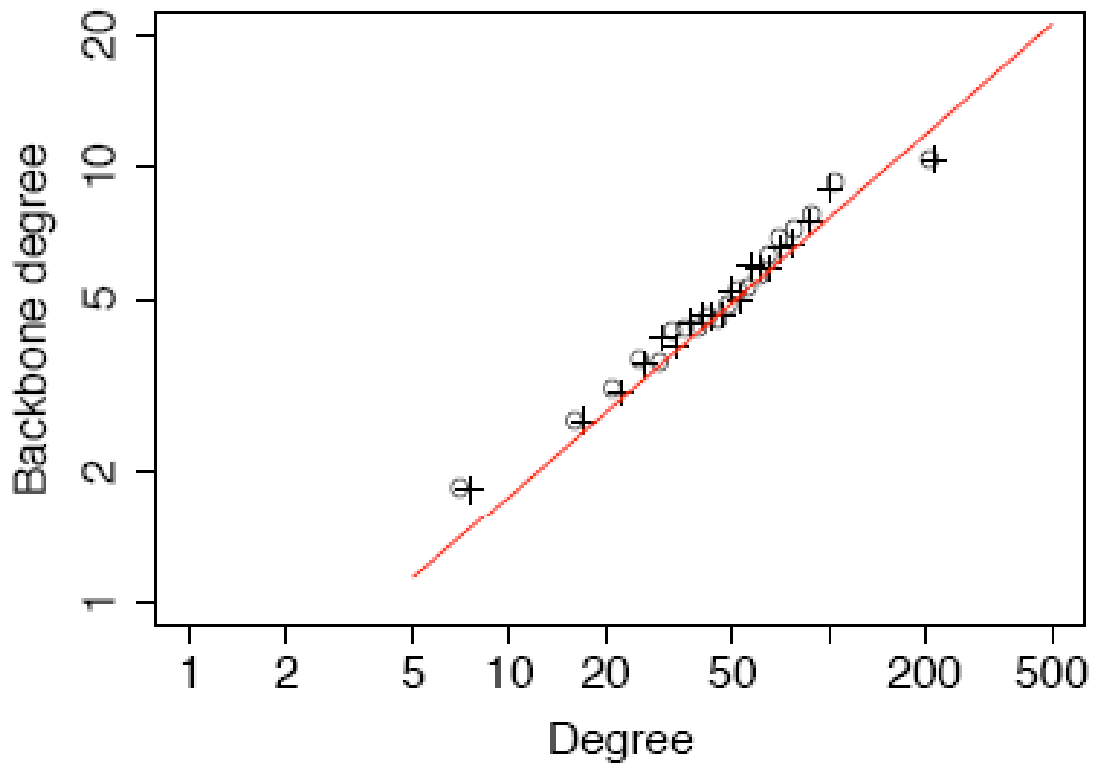


Figure 7: In-degree (circles) and out-degree (crosses) in the aggregate backbone as a function of degree in the full communication skeleton $\mathcal{G}$. The sublinear growth indicates that the backbone eliminates edges from high-degree nodes at a greater rate.

Communication Strategies

- Keeping the recipients fixed – are there ways to reduce the shortest path delays by changing the individual rates?

- Should one talk more actively to most frequent contacts – *load concentrating*

- Balance things by increasing communication with less frequent contacts – *load leveling*

Communication Strategies

- $\rho_{v,w} \rightarrow \rho_{v,w}^{\gamma}$
- $\gamma < 1$ corresponds to load-leveling
- $\gamma > 1$ corresponds to load-concentration