

# The structure of an Internet dating community (Introducing bipartivity)

January 17, 2003

**Petter Holme**

Department of Physics,  
Umeå University

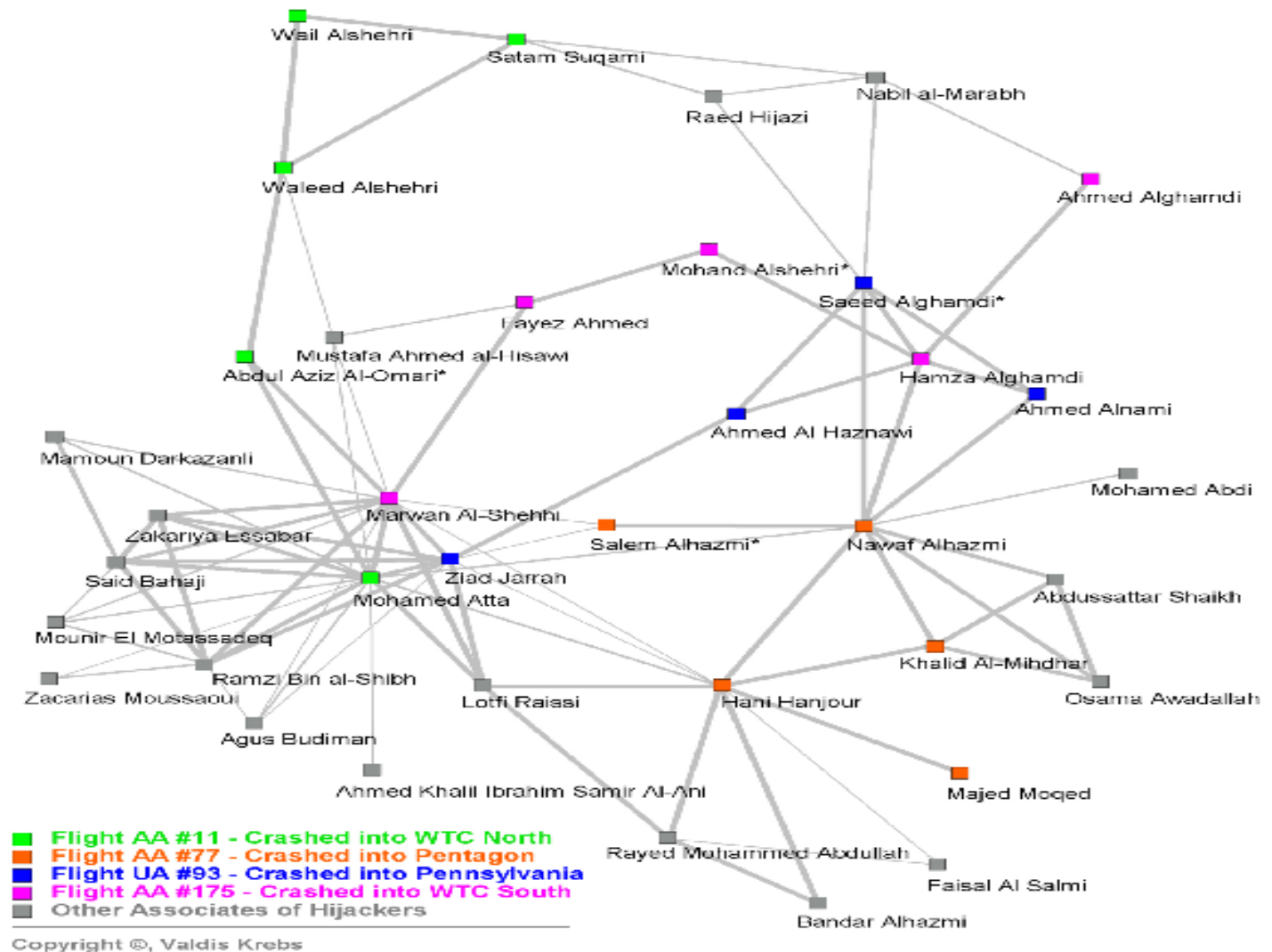
**Fredrik Liljeros &  
Christofer Edling**

Swedish Institute for  
Infectious Disease Control  
Department of Sociology,  
Stockholm University



# SOCIAL NETWORKS

- from group psychology . . .
- . . . to graph theory
- most intangible of networks
- epidemiology
- anthropology
- psychology
- defense



**Figure 4. Hijacker's Network Neighborhood**

## WHY INTERNET COMMUNITIES

---

- large size (nice for statistical mechanics style approach)
- precise timing resolution
- formation more close to real acquaintance network than other methods
- interesting *per se*

about pussokram.com:

- ~ 130 000 users
- ~ 70% female
- average age 21 yrs (?)

Hey you Friday, July 5, 2002 | Newest user: User B P16

» [Community](#) / [user A F20](#)

**User A F20**      *City*

You are logged in as: [user Z P20](#)

You have one new message

all in one place

[Message box](#)

[Secret corner](#)

[Community](#)

[Your homepage](#)

[Your questbook](#)

[Your diary](#)

[Write in your diary](#)

[Your friends](#)

[Search member](#)

[Forum](#)

[Member of the week](#)

[Ask Dr. Love](#)

[Cell phone](#)

[Competitions](#)

[Your settings](#)

[About pussokram](#)

[Log out](#)

number of online members  
**29**

Here User B has space to write about herself . . .

**No picture uploaded**

[Email me](#)

[Be my friend](#)

[Read my diary](#)

[Write in my guestbook](#)

**Member since:**  
April 16, 2000

**Last time online:**  
13:02, June 20

**Personal information:**

E-mail: [someone@somewhere.se](#)

ICQ: . . .

**Most recent visitors:**

[User D P23](#)

[User E P33](#)

[User F P19](#)

[User G F12](#)

[User H P30](#)

**Fast facts:**

Residence:	Apartment	Civil status	Single
Economy:	Rich	Favourite color:	Red
Movie taste:	Horror	Likes:	The World
Hair color:	Dyed	Hobbies:	Watch TV
Sense of humour:	Normal	Style of clothes	Streetwise
Length:	Taller than most	Favourite food:	Various
Music taste:	Everything	Eye color:	Blue
Personality:	Serious	Occupation:	Working

If you like User A you'd also like . . .

<a href="#">User D F25</a>	<a href="#">User H F25</a>	<a href="#">User L F23</a>
<a href="#">User E F21</a>	<a href="#">User I F25</a>	<a href="#">User M F22</a>
<a href="#">User F F28</a>	<a href="#">User J F22</a>	<a href="#">User N F27</a>
<a href="#">User G F20</a>	<a href="#">User K F29</a>	<a href="#">User O F26</a>

friends online

None of your Pussokram friends are currently online

sökofindern™

name of member

dr. love's answers

Here starts the text of a mail to "Dr. Love" . . .

[Read more . . .](#)

most recent diary

[User C F18:](#)

[Here goes the diary . . .](#)

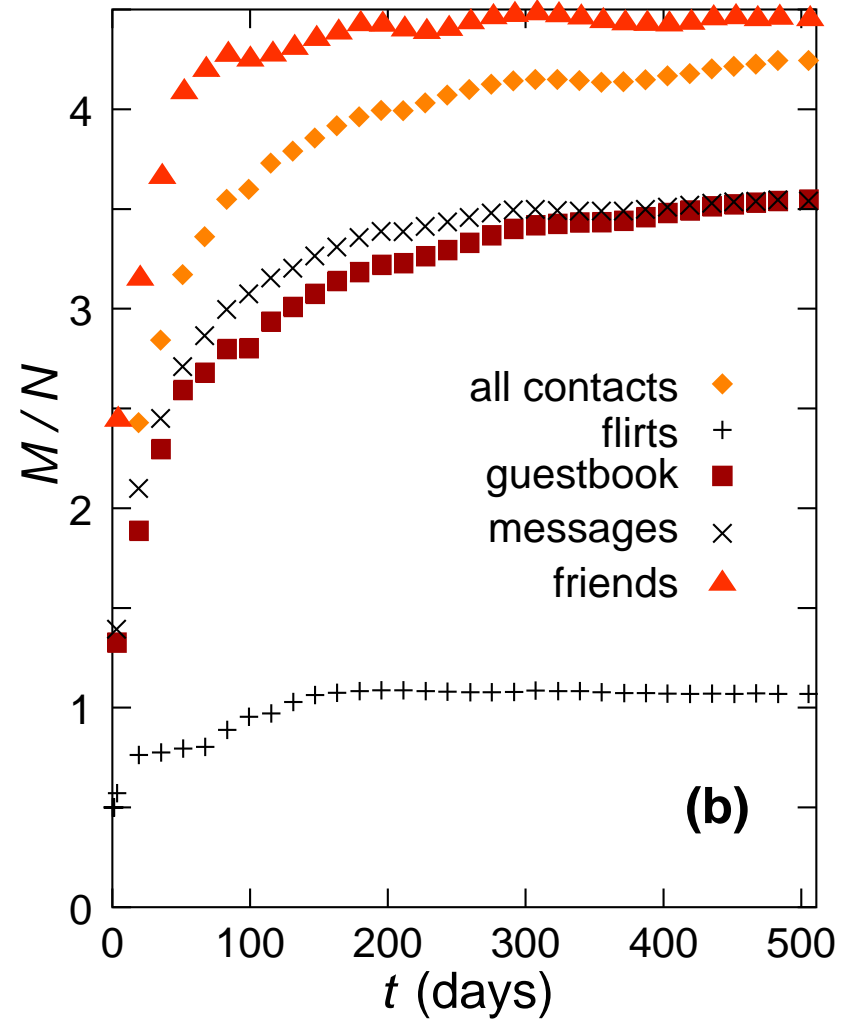
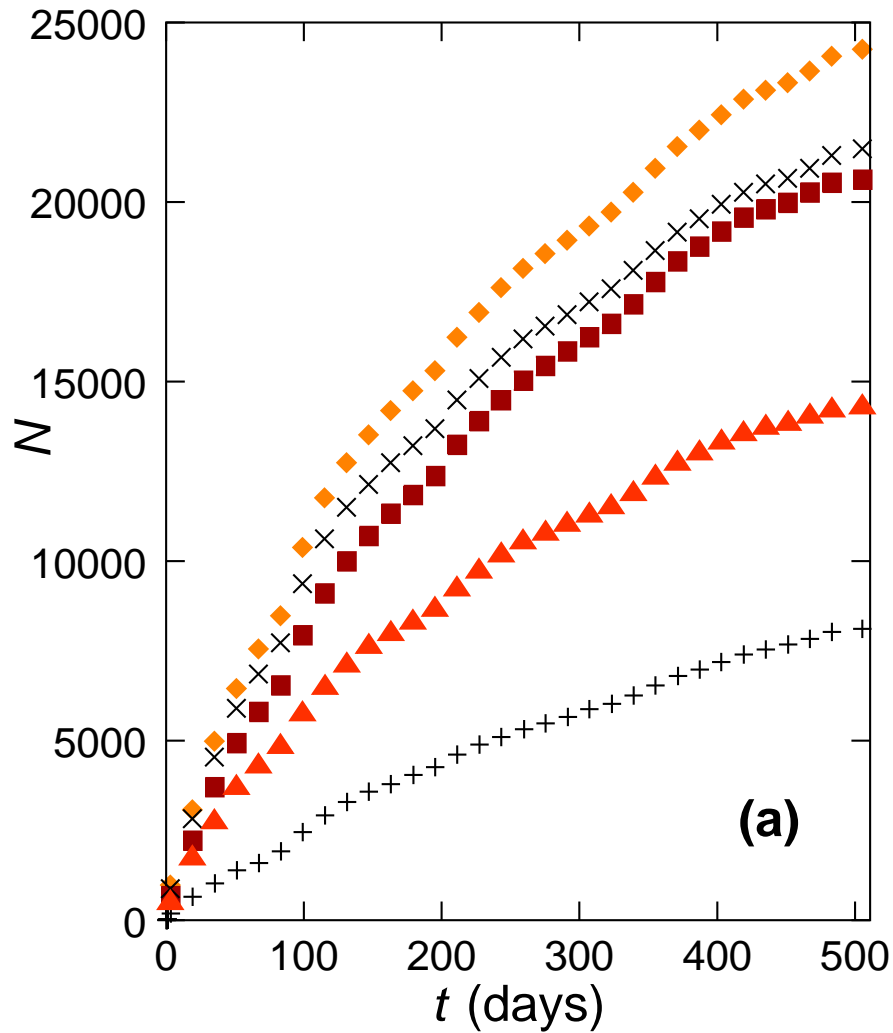
» [More diaries](#)

recently visited members

Here you can see the five most recently visited homepages

© 1999–2002 Manickel AB | [Information about pussokram.com](#) | [Press releases](#) | [Contact us](#) | [Disclaimer](#) | [Abuse](#)

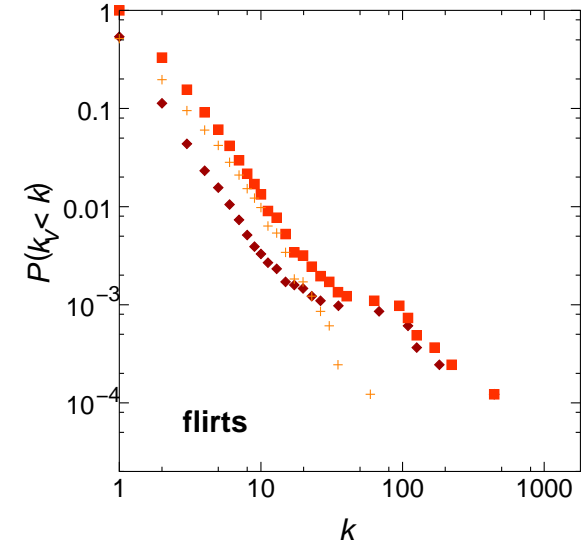
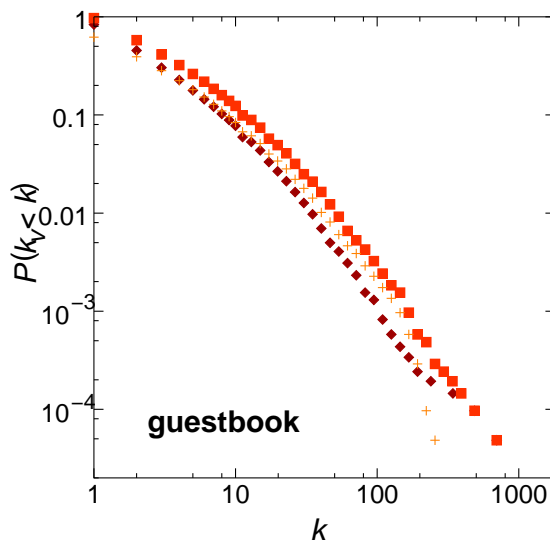
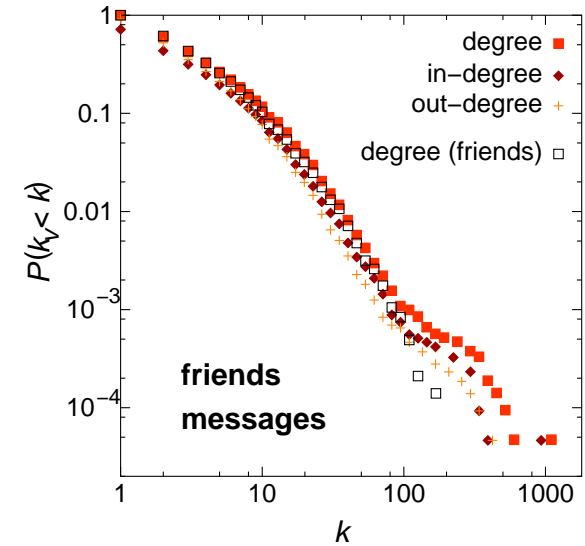
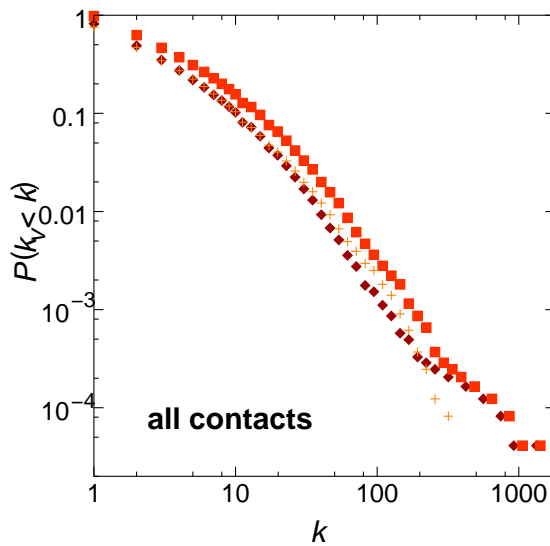
# TIME DEVELOPMENT



# DEGREE DISTRIBUTION

... quite scale-free, as seen in e.g. web of sexual contacts.

(There has to be a cut-off but its nowhere to be seen)

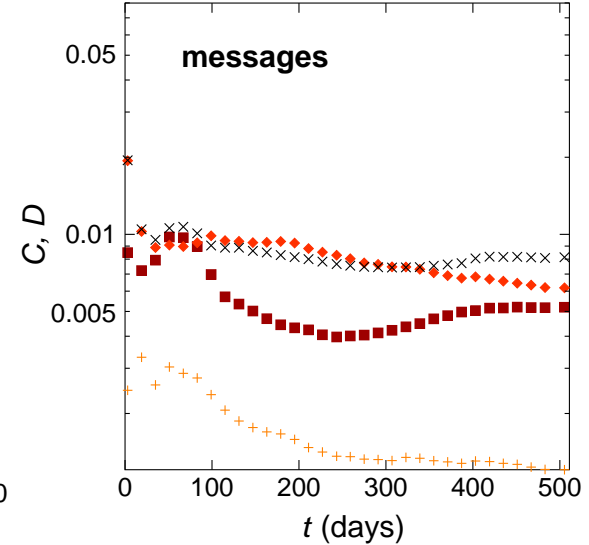
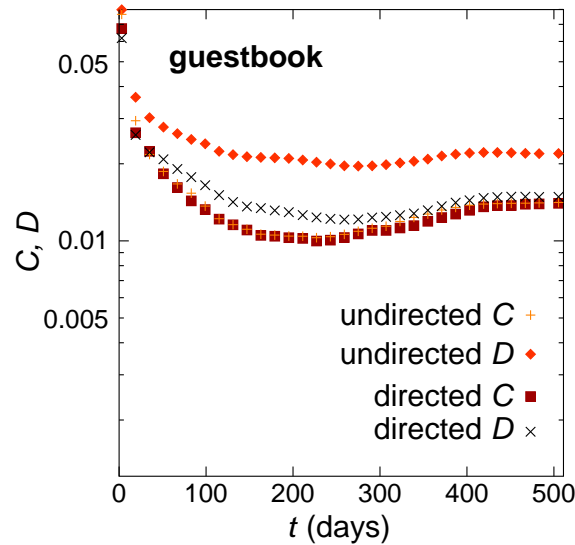
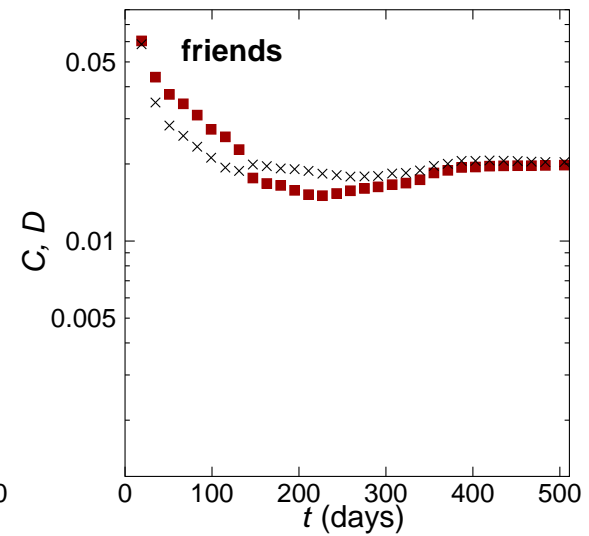
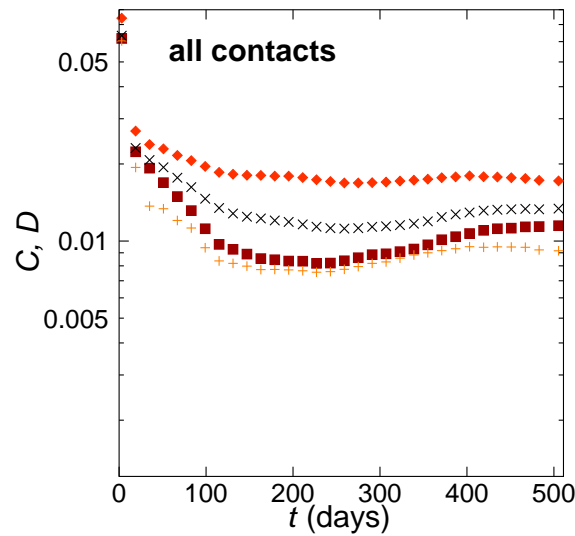


# DENSITY OF SHORT CIRCUITS

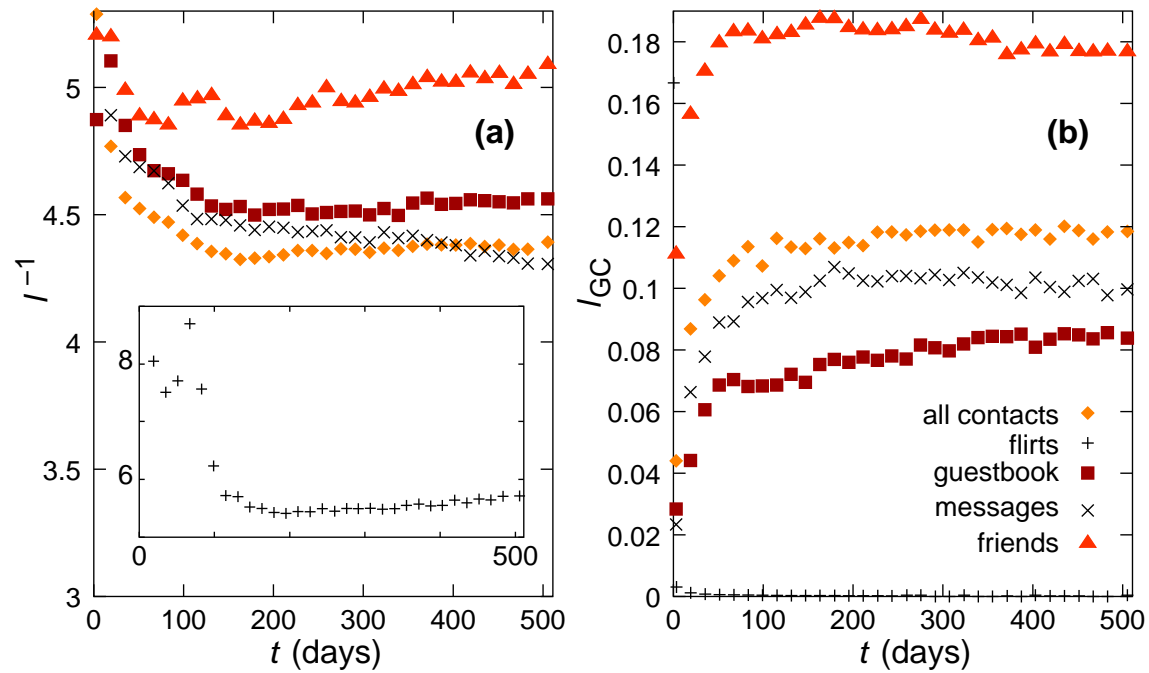
$C$ —density of triangles,  
 $D$ —density of squares.

High  $C$  in acquaintance networks.

$C > D$  in other empirical network data. Here  $D > C$  since most contacts are romantic & heterosexual.



# CHARACTERISTIC PATHLENGTH



$$l_{GSCC} = \frac{1}{|A_{GSCC}|} \sum_{(u,v) \in A_{GSCC}} d(u,v)$$

$$l^{-1} = \frac{1}{M} \sum_{(u,v) \in A} \frac{1}{d(u,v)}$$



## BIPARTIVITY . . .

---

In romantic networks there's another structure from heterophilous interaction, bipartivity—the network is closer to bipartite than a random network. How can this structure be measured in a general graph?

**Bipartivity measures** Supposing a strong heterophilous interaction  $\Rightarrow$

bipartivity = the fraction of unfrustrated edges in the ground state configuration of the antiferromagnetic Ising model: or

$$H = \sum_{(u,v) \in E} S_u S_v$$

where  $S_x \in \{-1, +1\}$  is a number assigned to every vertex.

Disadvantages with this measure:

- NP-complete optimization problem

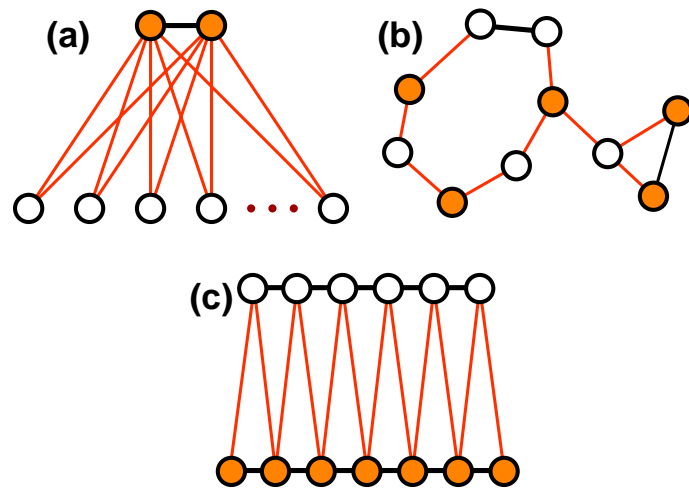
Solution: Calculate approximate value by simulated annealing.

Construct a heuristic measure calculable in polynomial time . . .

## . . . FROM HEURISTICS

Base the measure on odd circuit count . . .

To take into account . . .



(a) A highly bipartite graph can have many odd circuits.

(b) Not only triangles contribute.

(c) If many circuits are present, short ones are more important.

- Let  $C_n$  be the set of odd circuits of length  $\leq n$ .
- Let  $\Sigma(C_n)$  be the accumulated length of the circuits of  $C_n$  (so, for example  $\Sigma(C_3) = 3$  in (b)).
- Assign the cut-off  $3M$  to  $\Sigma(C_n)$ , and let  $\hat{n}$  be the smallest  $n$  such that  $\Sigma(C_n) \geq 3M$ .

This can be done in  $O(MC)$  ( $C$  is the total number of circuits) by Johnson algorithm. If the bound on  $n$  is realized the effective running time is  $O(M^2)$ .

Let  $\nu(a)$  denote the number of cycles in  $C_{\hat{n}}$  passing through the arc  $a$ .

1. Start with  $C = C_{\hat{n}}$ .
2. Sort the edges in order of  $\nu$ .
3. Repeat the following while  $C \neq \emptyset$ :

- (a) Mark the edge  $e$  with highest  $\nu$ .
- (b) Remove all cycles in  $C$  containing  $e$ .
- (c) Recalculate  $\nu$  for each edge.

Then the number of iterations  $m'$  is the assessment of  $M_{\text{fr}}$ , and we define our bipartivity measure as

$$b_2 = 1 - \frac{m'}{M}. \quad (1)$$

## . . . FROM OPTIMIZATION BY XMC

---

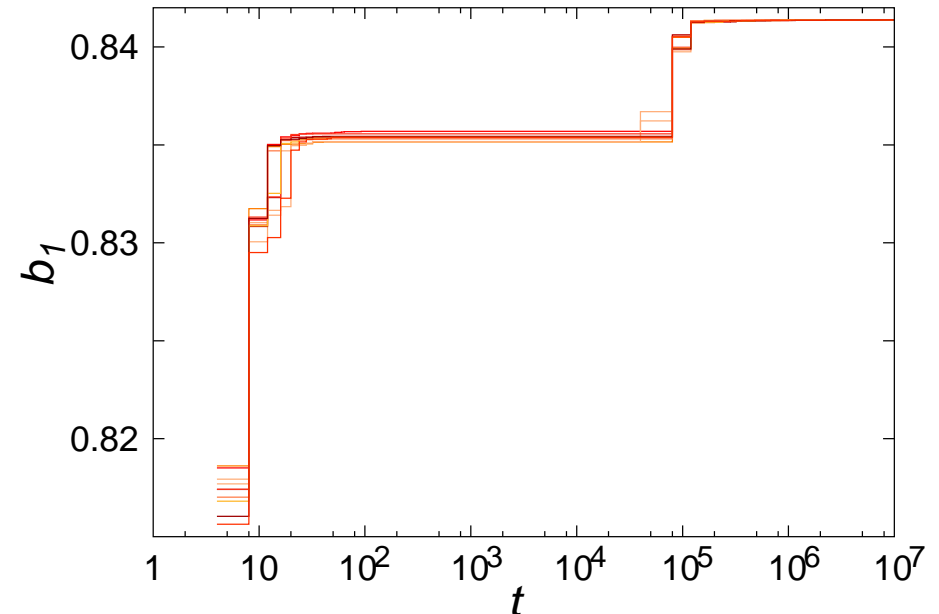
Simulated annealing: Run Monte Carlo at high temperature, and thus avoids getting stuck in a local minimum. Then slowly decrease the temperature, and save the lowest energy configuration attained.

To make this effective we use two tricks:

- Use Exchange Monte Carlo (XMC, to avoid getting stuck in local minima).
- Quench the system (flip spin as long as the energy is lowered) from time to time. (This is done to get the really low energies.)

In XMC one runs the replicas of the system at several temperatures simultaneously. At certain intervals replicas at adjacent temperatures are compared and swapped if a certain condition holds.

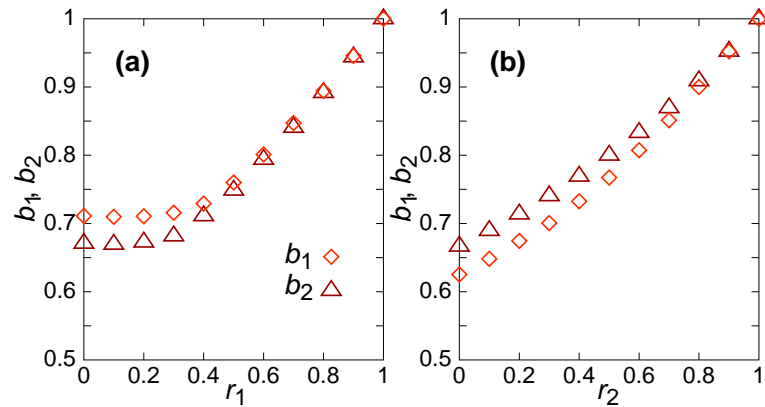
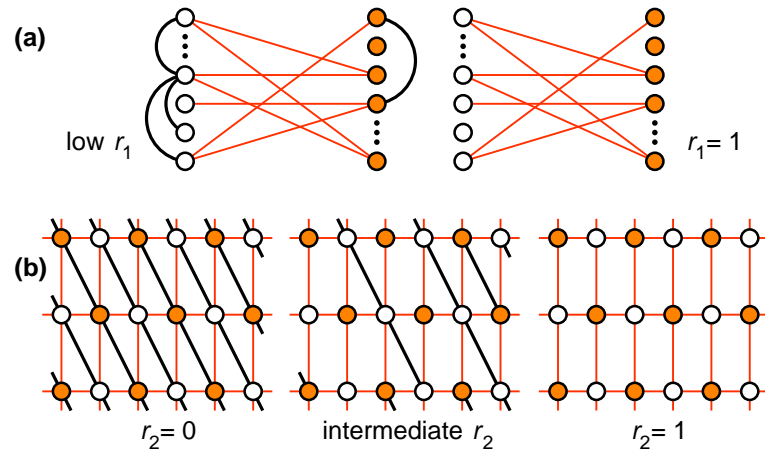
So different replicas can travel up and down



Quenches at times separated by 40 MC sweeps (giving the first jump). Exchange trials at times separated by  $2 \times 10^4$  MC sweeps (giving the second jump). So running XMC makes sense.

# RESULTS

## Model networks



## Real World networks

network	$N$	$M$	$b_1^{\text{dir}}$	$b_1$	$b_2^{\text{dir}}$	$b_2$
all contacts	29 341	174 662	—	0.860	0.948	0.928
messages	20 691	73 346	—	0.892	0.984	0.964
guestbook	21 545	76 257	0.863	0.889	0.943	0.965
nioki.com	50 259	405 742	0.842	0.855	0.956	0.975
arxiv.org	52 909	490 600	×	0.63	×	0.623
directors	7 475	48 899	×	0.549	×	0.507