Reputation in Self-Organized Communication Systems and Beyond

Jochen Mundinger (with Jean-Yves Le Boudec)

jochen.mundinger@epfl.ch

EPFL - IC - LCA

Lausanne, Switzerland

www.mundinger.org

Introduction

- Reputation
  - in (self-organized) communication systems
    - free-riding, malicious attacks and random failures
    - almost separate evolution in various communities: Peer-to-Peer (P2P), Mobile Ad-Hoc (MANET), Artificial Intelligence (AI)
  - in the social sciences
  - in other disciplines
    - economics, psychology
    - management science, marketing
- Define as *estimate about a person's actual quality* (following Oxford English Dictionary)
  - *person*: user, node, peer
  - *quality*: context dependent

Outline of talk

- Motivation, self-organized communication systems
- State-of-the art in computer science
  - historically oriented
- Our modelling results
  - impact of liars on their peers' reputation about a subject
- Reputation in the social sciences
- Fundamental directions for future research
  - coherent terminology and classification
  - coherent approach
  - fundamental questions as well as specific implementations
  - analytical models as well as simulation, measurement, testing

Self-organized communication systems

Self-organized communication systems

- Increasing attention in deployment and research
- Peer-to-peer (P2P) organization principle
  - equal participants in terms of capabilities, responsibilities
- Internet-based P2P systems
  - Napster, Kazaa, BitTorrent...
  - significant proportion of total Internet traffic
- Mobile Ad-Hoc Networks
- Wikipedia (but element of centralization)
- Similar systems in nature
  - biology
  - social networks

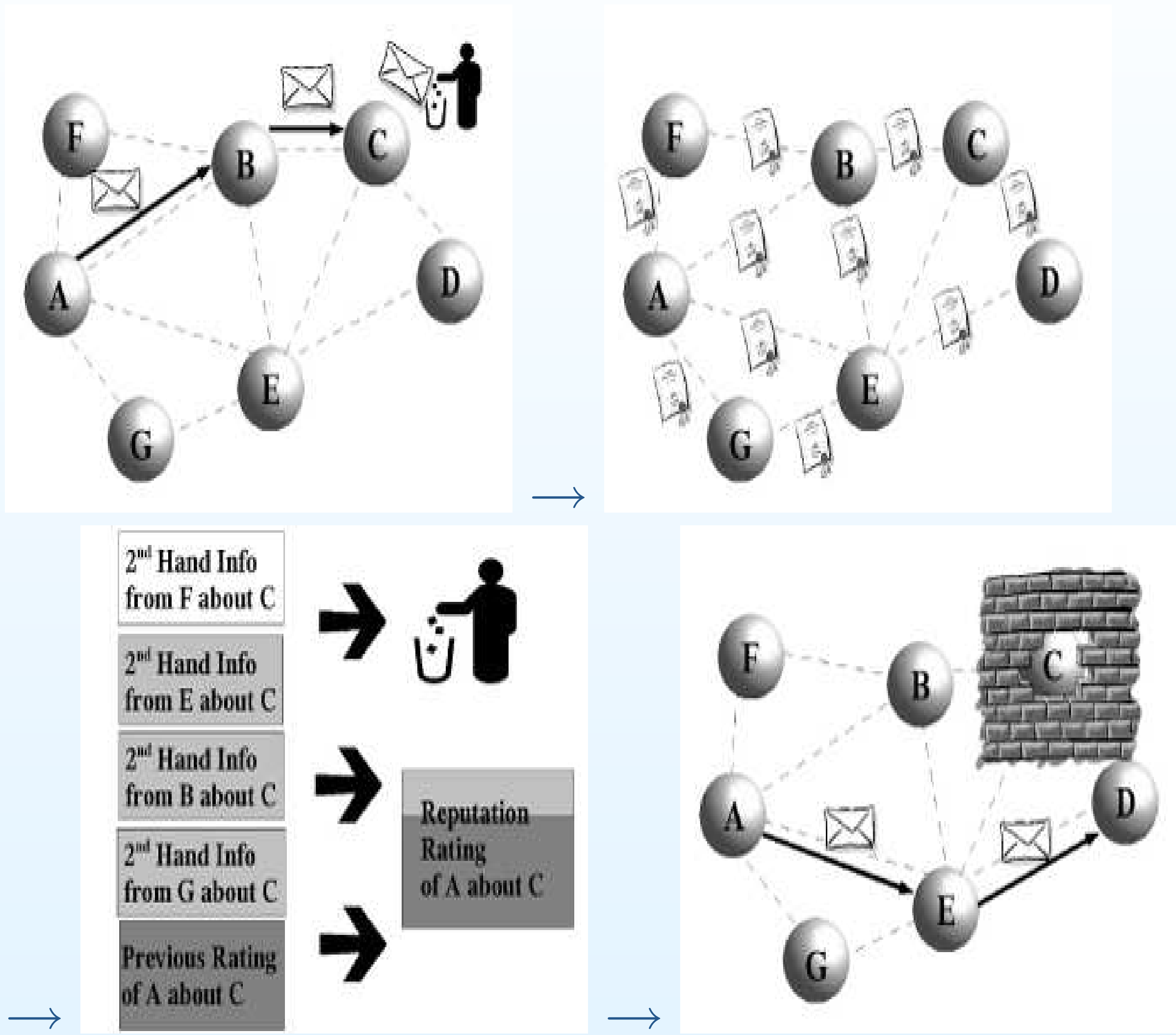
Problems and solutions approaches

- Advantages such as scalability
- Problems such as free-riding
  - self-interest (bandwidth, energy)
  - measurement studies: detrimental effects
  - *price of anarchy* (delays in Internet flows)
  - economics, public goods: services not provided at all or without sufficient quality of service
  - everyday life: air pollution, traffic jams
- Altruism?
- Solution approaches
  - incentive mechanisms: pricing, rules
  - reputation systems
  - artificial immune systems

Reputation systems

- Idea
  - users observe their peers' behaviour
  - exchange this information with their peers to compute a reputation value about each peer
  - use both first and second-hand information
    - can obtain accurate estimate faster
    - even without ever having interacted directly
    - but vulnerable to liars
  - users with a good reputation value are favoured
- Address malicious attacks and random failures as well as free-riding
- Have proven useful and popular in online systems (eBay, Amazon)
- But need to be fully distributed

Illustration for a Mobile Ad-Hoc Network



State-of-the-art in computer science

State-of-the-art in computer science

- Historically oriented overview
  - artificial intelligence community
  - Internet-based P2P community
  - Mobile Ad-Hoc Network community
- Incoherent definition, representation, terminology, classification
- Sometimes links are rare, despite similar ideas being used (*reinventing the wheel*)
- Often more links to social sciences than to other computer science communities

Sample classifications (1/2)

- (A1) conceptual model
- (A2) information sources (direct experiences, witness information, sociological information)
- (A3) prejudice (conclusions drawn from group membership)
- (A4) visibility types
- (A5) granularity
- (A6) agent behaviour assumption (honest, lie partially, lie and mechanisms)
- (A7) type of exchanged information (discrete, continuous)
- (A8) reliability measure

Sample classifications (2/2)

- (B1) social networks (by which they mean graph theoretic models that consider transitivity of trust along the edges)
- (B2) probabilistic estimation
- (B3) game theoretic models (both classical and evolutionary game theory)

- (C1) representation of information and classification
- (C2) use of second-hand information
- (C3) trust
- (C4) redemption and secondary response

- All relevant in all communities?

Conclusions and directions for further work

- Highlighted role of reputation for self-organized communication systems
- Overview and pointers to the various communities

- Definition and representation of reputation vary widely
- Even within computer science
- More coherent terminology certainly desirable
- More coherent classification

- Bring together different strands of research
- Within computer science as well as between disciplines

Our modelling results

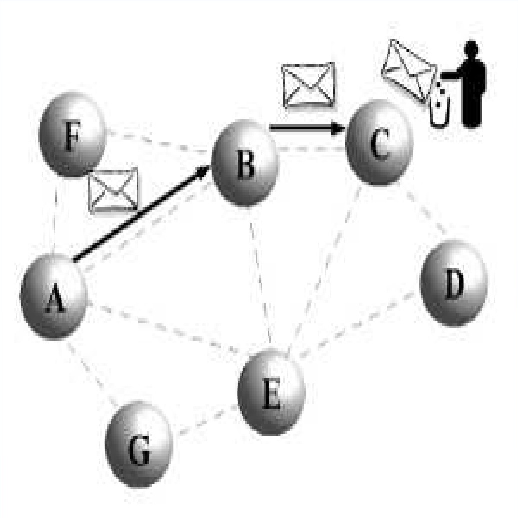
Modelling the liars' impact

- Tradeoff between speed and accuracy
- What is liars' impact on their peers' reputation about a subject? (Accuracy)
- Subject can be a peer, but also external entity
- Believe second-hand information if and only if it does not differ too much (deviation test)
  - simple, fully distributed, effective?
- Consider abstract model (w/o observation, reaction)
- Appropriate for range of communication networks and social networks, although different points-of-view
- Related work
  - typically based on simulation

Reputation vs. *Direct Observation*

- Address fundamental question, independently of implementation details
  - *Reputation* vs. *Direct Observation*
- *Reputation*
  - current estimate (based on both first and second-hand information) is passed on
- *Direct Observation*
  - only first-hand information is passed on
- What difference does this make for the reputation system?

Modelling assumptions



- Subject (here C) behaving positively w.p. θ
  - Case of many subjects can be decomposed
- N_h honest users and N_l liars
- Each user i has counters (x_n^i, y_n^i)
- Reputation values are obtained as $R_n^i = x_n^i/(x_n^i + y_n^i)$ in $[0, 1]$

Reputation system assumptions

- Second-hand information
  - deviation test
    - second-hand information is accepted only if it does not differ by more than Δ from current reputation value
  - if accepted, it is still weighted by a factor ω_{weight}

- Discounting
  - With each new observation, discount old observations by a factor ρ

Setting assumptions

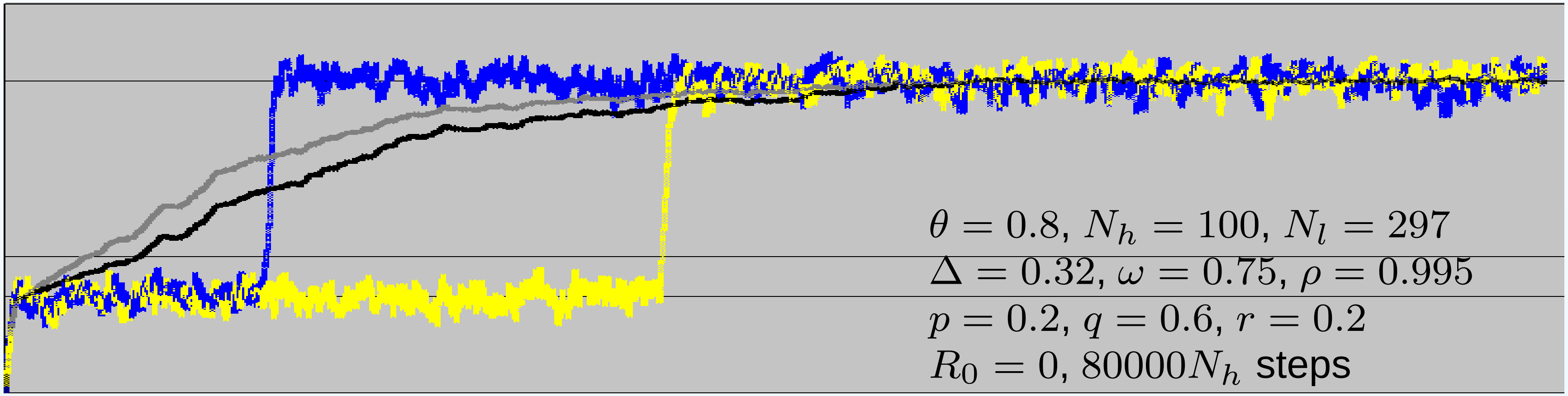
- Liars
  - extremely negative (o/w similar by symmetry)

- Interactions
  - symmetric interactions
  - Poisson processes
  - given interaction is direct/lie/honest w.p. $p/q/r$

Methodology

- Stochastic process formulation
  - based on assumptions
  - can be viewed as a generalization of the voter model in statistical physics
- Mean ordinary differential equation (ODE)
  - averaging the dynamics
  - *fast-time scaling* limit
  - considering average reputation
- Solve mean ODE and study fixed points
- Verification
  - simulation
  - direct computation

Typical sample path (*Direct Observation*)



- Reputation vs. time (steps/observations/interactions)
- Lines at 1, $R^*_{true} = 0.8$, R^*_{inter}, $R^*_{false} = 0.2$ and 0
- Two individual reputation values in blue and yellow
- Two average reputation values are plotted in black and grey
- Reputation values settle down at R^*_{false} and eventually R^*_{true}
- Confirms existence and values of R^*_{false} and R^*_{true}

Main Results

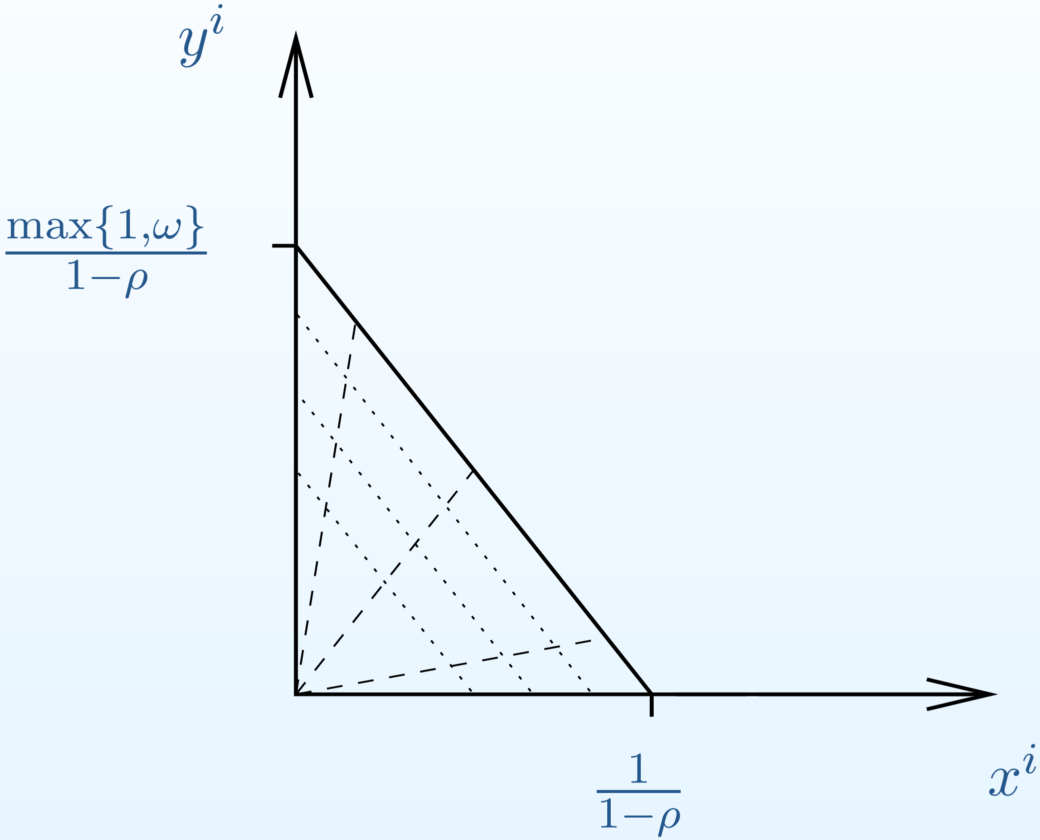
- Performance evaluation
  - Phase transition behaviour
    - no impact unless the proportion of liars exceeds a certain threshold
  - Alternatively, phrase in terms of Δ
    - no impact unless Δ exceeds a certain threshold
  - Precise formulae, quantify impact
- Performance optimization
  - For maximal speed subject to accuracy choose Δ equal to first critical value
- *Reputation* vs. *Direct Observation*
  - coincide if and only if $\theta > 2\Delta$
  - *Reputation*: second-hand information does not improve accuracy

Stochastic process (*Reputation*)

$$(x^i_{T^i_{n+1}}, y^i_{T^i_{n+1}}) = \rho(x^i_{T^i_n}, y^i_{T^i_n}) \; +$$

- $(1, 0)$ w.p. $p\theta$
- $(0, 1)$ w.p. $p(1-\theta)$
- $(0, \omega)1_{\left\{\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} \le \Delta\right\}}$ w.p. q
- $\omega(1-\rho)(x^j_{T^j_n}, y^j_{T^j_n})\mathbf{1}\left\{\left|\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} - \frac{x^j_{T^j_n}}{x^j_{T^j_n}+y^j_{T^j_n}}\right| \le \Delta\right\}$ w.p. r

State space



- For each user i, state space is subset of triangular area
- Dashed lines of constant reputation, increasing from bottom
- Dotted lines of constant certainty, increasing from left

Specific Results (*Reputation*) (1/2)

Theorem:

If $\Delta < \Delta_{c_1} = (p\theta)/(p + q\omega)$,

$$(x, y) = \frac{1}{(1 - r\omega)(1 - \rho)} (p\theta, p(1 - \theta)) \tag{1}$$

is the unique fixed point of the mean ODE.

It is asymptotically stable and all trajectories are attracted to it.

The corresponding reputation value is $R^*_{true} = \theta$.

Specific Results (*Reputation*) (2/2)

If $\Delta_{c_1} \leq \Delta < \Delta_{c_4} = \theta$, there is a second, false fixed point

$$(x, y) = \frac{1}{(1 - r\omega)(1 - \rho)} (p\theta, p(1 - \theta) + q\omega) \, . \qquad (2)$$

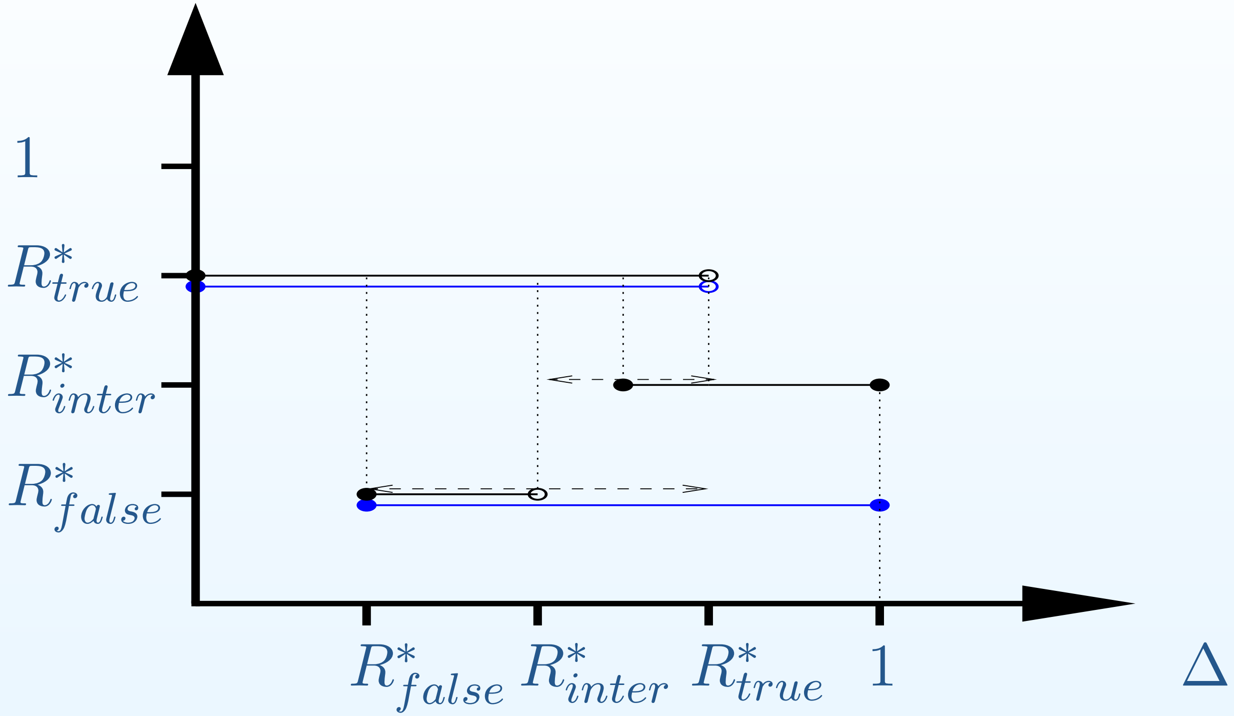
Both are asymptotically stable, attracting trajectories from $x(t)/(x(t) + y(t)) > \Delta$ and $x(t)/(x(t) + y(t)) \leq \Delta$ respectively. The corresponding reputation value is $R^*_{false} = \theta p/(p + \omega q)$.

If $\Delta_{c_4} \leq \Delta$, then only the latter, false fixed point is asymptotically stable and all trajectories are attracted to it.

Specific Results (*Direct Observation*)

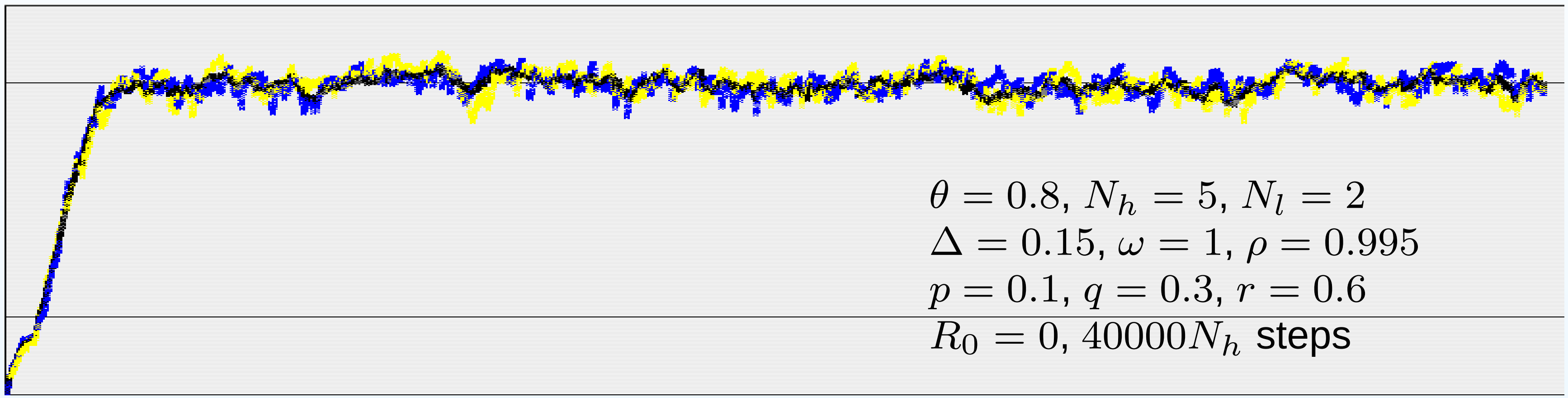
- Similar result as for *Reputation*
- But more complex
  - up to 3 possible fixed points
  - up to 4 critical values

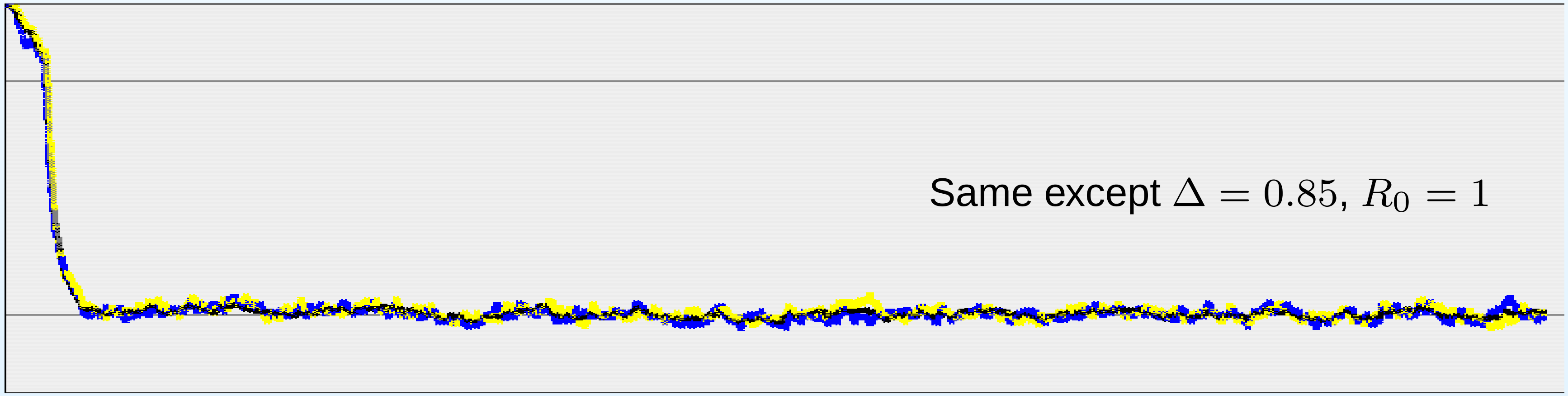
Reputation vs. *Direct Observation*



- Bifurcation plot in terms of Δ
- *Reputation* is shown in blue, *Direct Observation* in black
- *Reputation*: up to 2 fixed points, 2 critical values
- *Direct Observation*: up to 3 fixed points, 4 critical values

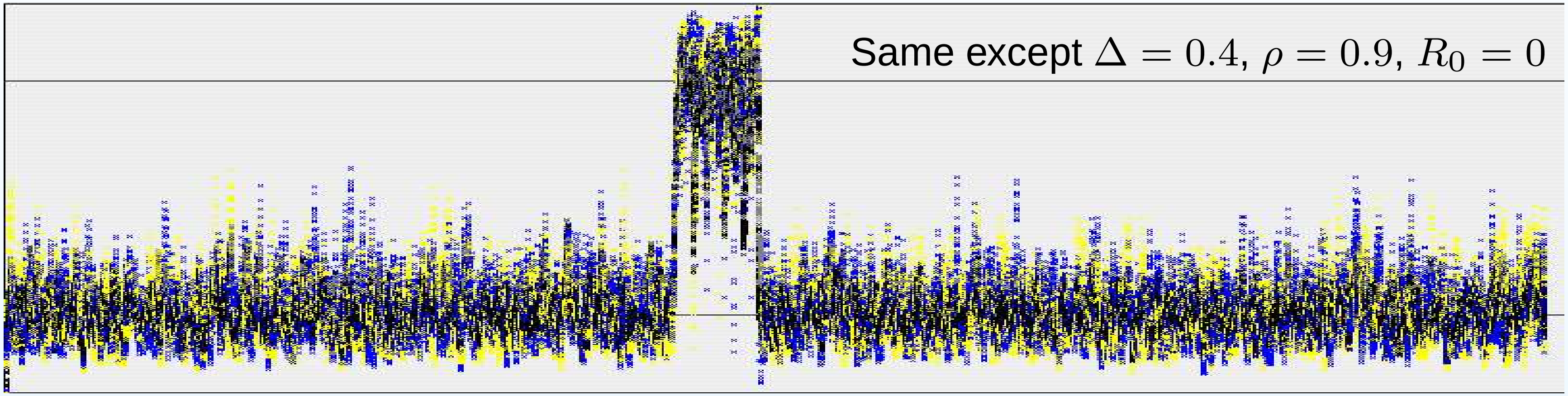
Simulation (*Reputation*) (1/3)

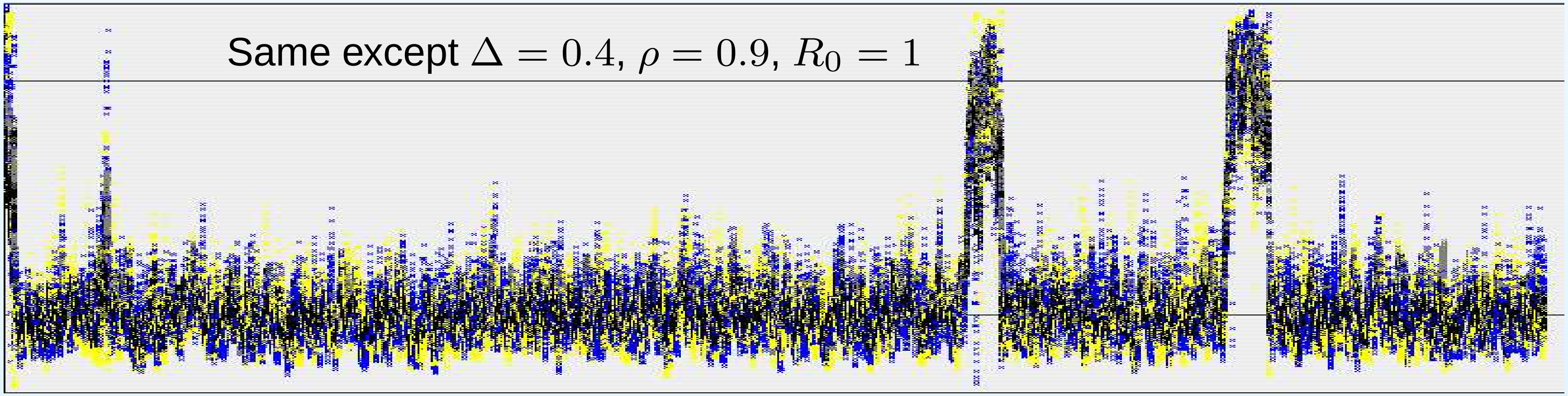




- $R^*_{true} = 0.8$, $R^*_{false} = 0.2$, $\Delta_{c1} = 0.2$, $\Delta_{c4} = 0.8$

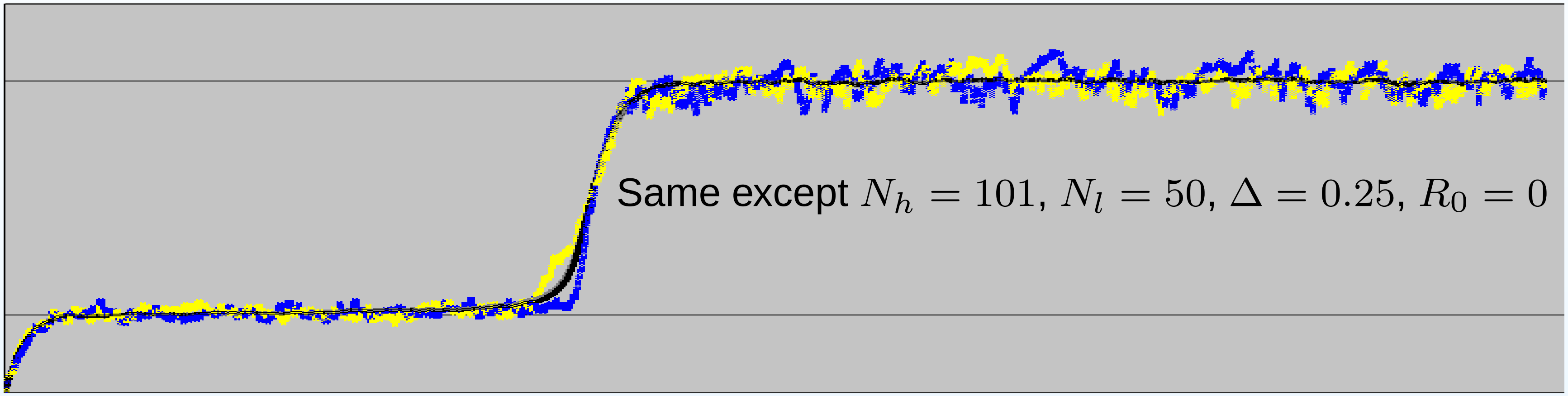
Simulation (*Reputation*) (2/3)

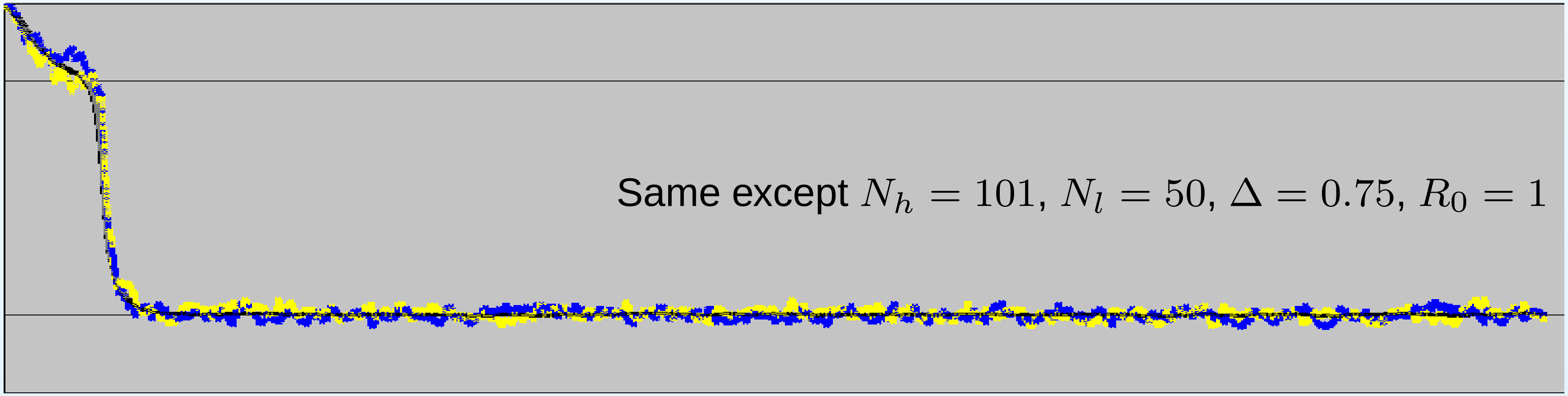




- $R^*_{true} = 0.8$, $R^*_{false} = 0.2$, $\Delta_{c1} = 0.2$, $\Delta_{c4} = 0.8$

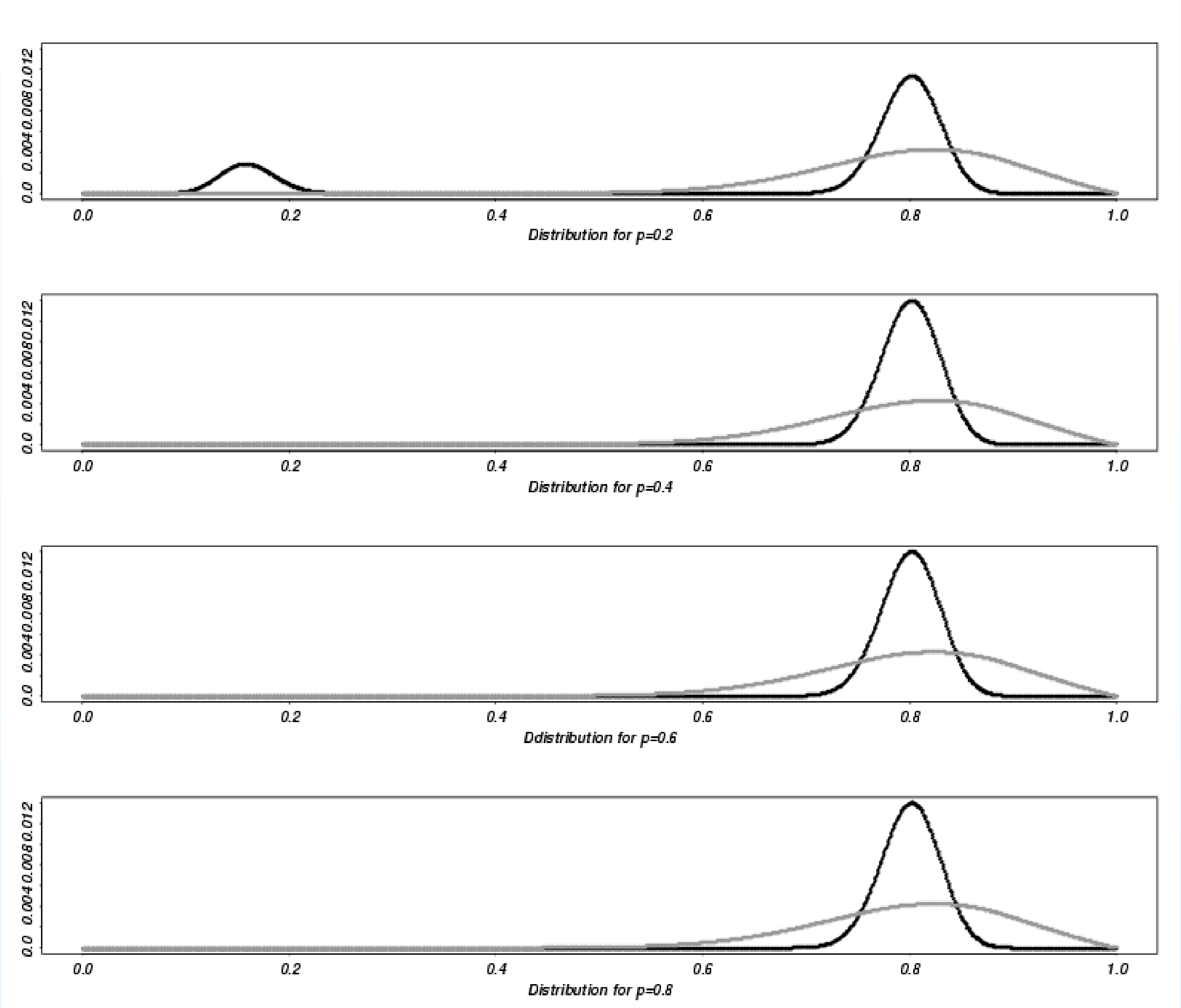
Simulation (*Reputation*) (3/3)



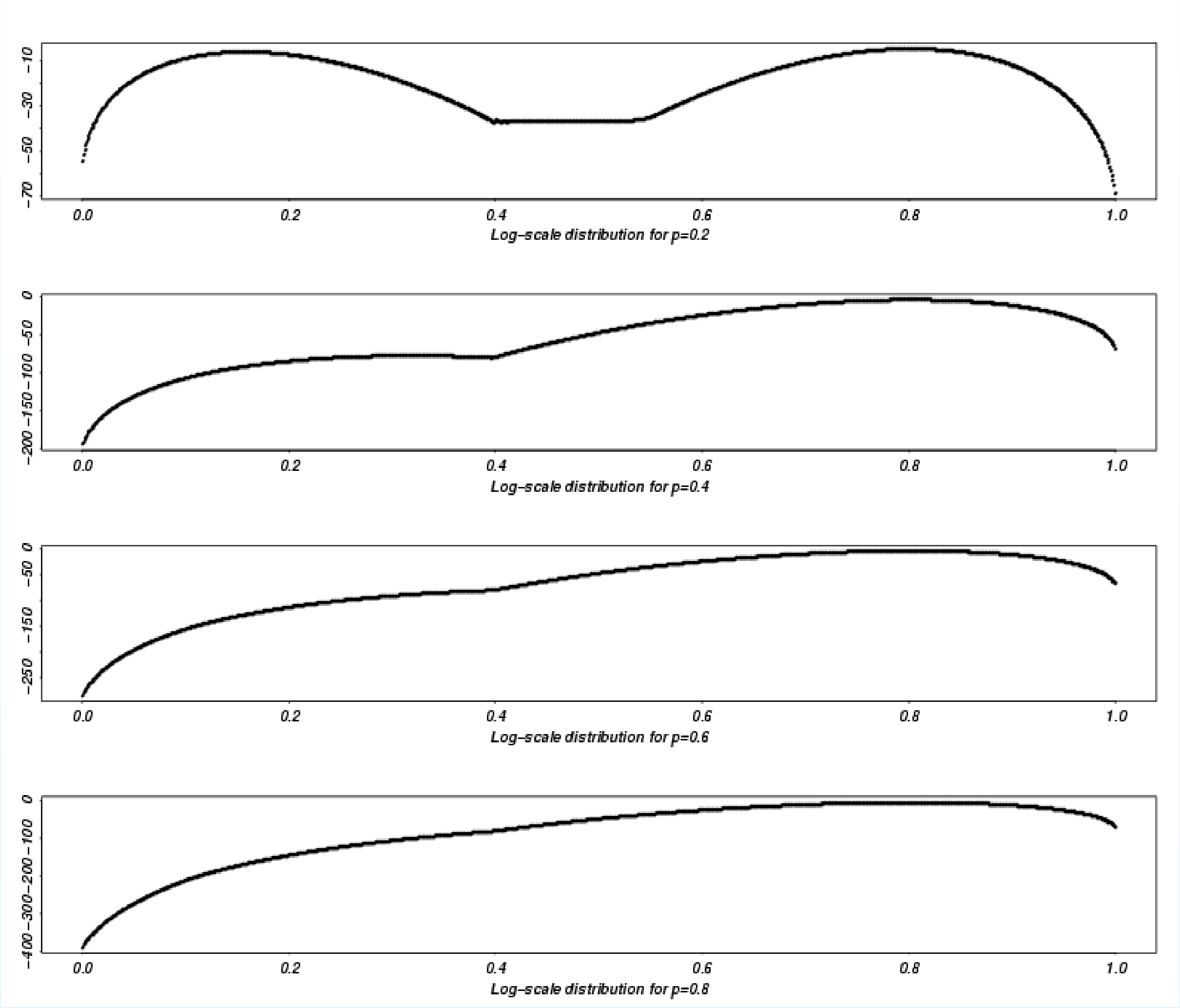


- $R^*_{true} = 0.8$, $R^*_{false} = 0.2$, $\Delta_{c1} = 0.2$, $\Delta_{c4} = 0.8$

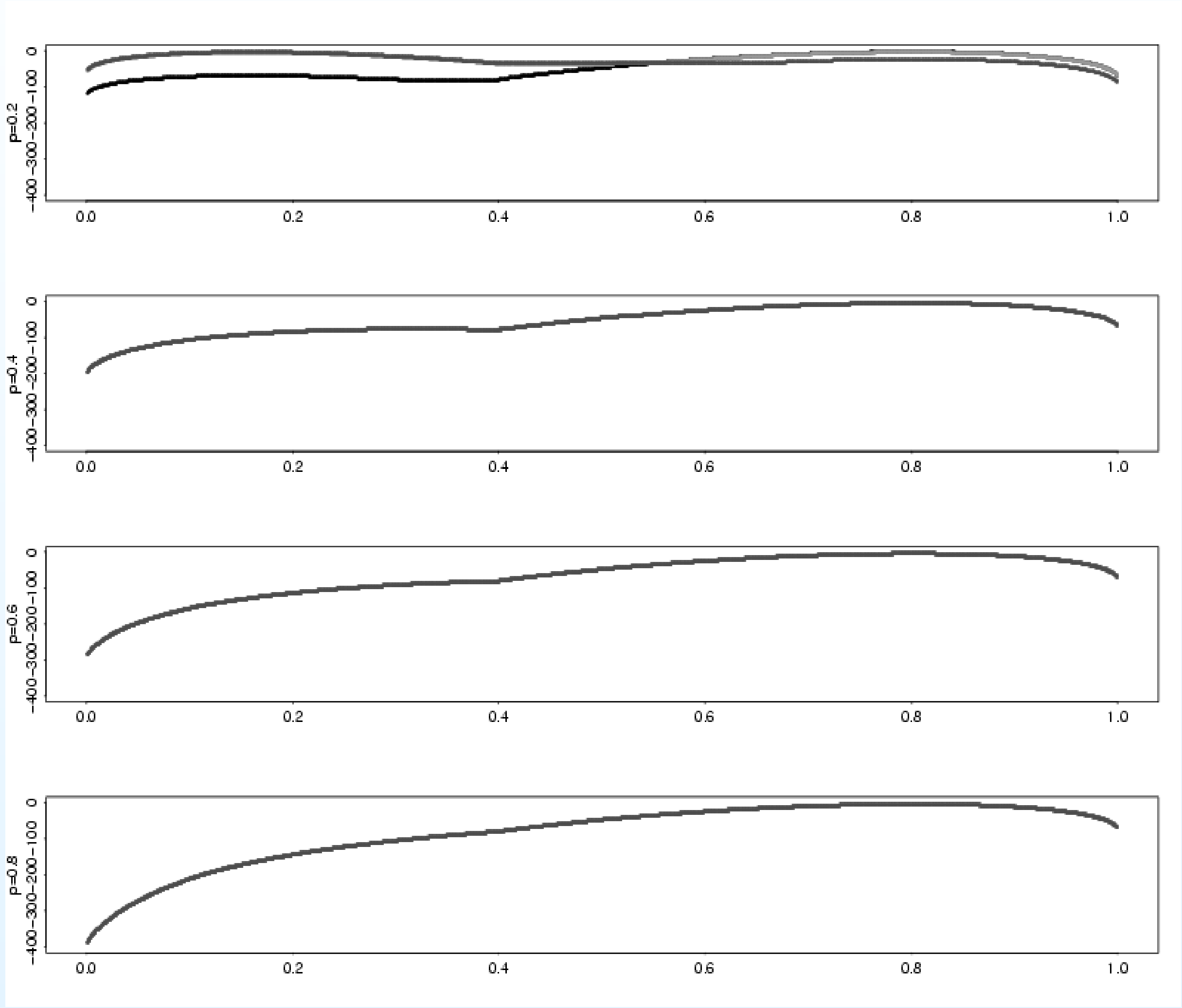
Computation of distribution (2 users)



Log of distribution (2 users)



Log of distribution, varying initial conditions (2 users)



Conclusions and further work

- Performance evaluation
  - Phase transition behaviour (proportion of liars, Δ)
  - Precise formulae, quantify impact
- Performance optimization (choice of Δ)
- *Reputation* vs. *Direct Observation*

- Correlated subject behaviour
- Strategic lying
- Asymmetric setting

- Address fundamental questions
- Use of analytical models

Reputation in the social sciences

Reputation in the social sciences

- Social network analysis
  - relationships of individuals in a society
  - natural, everyday life
- Liars in social networks?
  - dense, closed social networks
  - interest in behaviour of some subject (e.g. person, restaurant)
  - conversations (second-hand information, interactions)
  - reputation about the subject
  - liars in the absence of trust
  - confirmation bias (deviation test)
  - trust more what I have seen myself (weighting)
  - forgetful (discounting)

Modelling results from a social sciences viewpoint

- As before
  - phase transition behaviour
    - no impact unless the proportion of liars exceeds a certain threshold
  - precise formulae, quantify impact
  - *Reputation* and *Direct Observation*
    - *Reputation*: second-hand information does not improve accuracy
    - *Direct Observation*: second-hand information does improve accuracy
- No design issue here
- Explicatory nature
- Different approach for social scientists (data collection and interpretation)

Related work in the social sciences

- Foes deteriorate a person's labour market position
- Social capital, dependence on structure of network
  - *bandwidth hypothesis*: network closure enhances information flow
  - *echo hypothesis*: closure models merely create an echo that reinforces predispositions and leads to ignorant certainty
  - fundamental choice for theoretical models of trust
  - evidence: echo over bandwidth
  - cf. our lying, passing on reputation
- Herd behaviour

Fundamental Directions for Research

Fundamental Directions for Research

- Motivated by state-of-the-art
  - more coherent terminology and classification
  - bring together different strands of research
- Motivated by modelling results
  - fundamental questions as well as specific ones
  - analytical models as well as simulation, measurement, testing
    - often valid (at least qualitatively) even when assumptions violated

- Acknowledgements
  - Thanks to the organizers for their kind invitation
  - Thanks to Sonja Buchegger, Radu Jurca and Wojciech Galuba for helpful discussions