# Modelling Knowledge

## Bielefeld/Berlin-networks-group:

Andreas Krüger
Dima Volchenkov
Philippe Blanchard
Rainer Siegmund-Schultze
Sascha Delitzscher
Tyll Krüger

Vienna 14.11.2007

NEMO-Workshop

- Part 1:
  Generalized Epidemic Process (GEP)
  for Knowledge Diffusion („Spreading“)

- Part 2: for further modelling:
  Knowledge about *Knowledge*…

- Part 3:
  Shared/Complementary Knowledge Model
  for Partner Choice

# Part 1: Diffusion of knowledge

Generalized Epidemic Process (GEP):

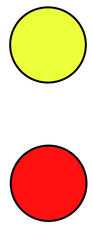
- classical epidemics
- threshold epidemics
- mean-field effect
- forgetting, i.e. active / passive knowledge
- Initial infection:
  „seed group“ of interconnected nodes

Bimodal Network with $N = N_{orgs} + N_{projs}$ nodes

$x \sim y \iff$ **edge** between $x$ and $y$

$$
\begin{aligned}
d(x) &= \text{degree of x} \\
&= \text{number of projects in which } \textbf{organisation } x \text{ participates} \\
&\phantom{=} \textit{or} \text{ number of participating organisations in } \textbf{project } x
\end{aligned}
$$

$$\omega(x) = \omega_t(x) = \begin{cases} 0 & \textit{unaware} \\ 1 & \textit{knowing} \end{cases} \quad \text{at time } t$$



first model:
*one* type of knowledge

project $x$ can be unaware/knowing and
organisation $x$ can be unaware/knowing

first model:
*no* distinction between
organisations & projects

## global observable

$$b = b_t = \frac{1}{N}\sum_{x=1}^{N}\omega(x) \quad \text{total knowledge prevalence (at time } t\text{)}$$

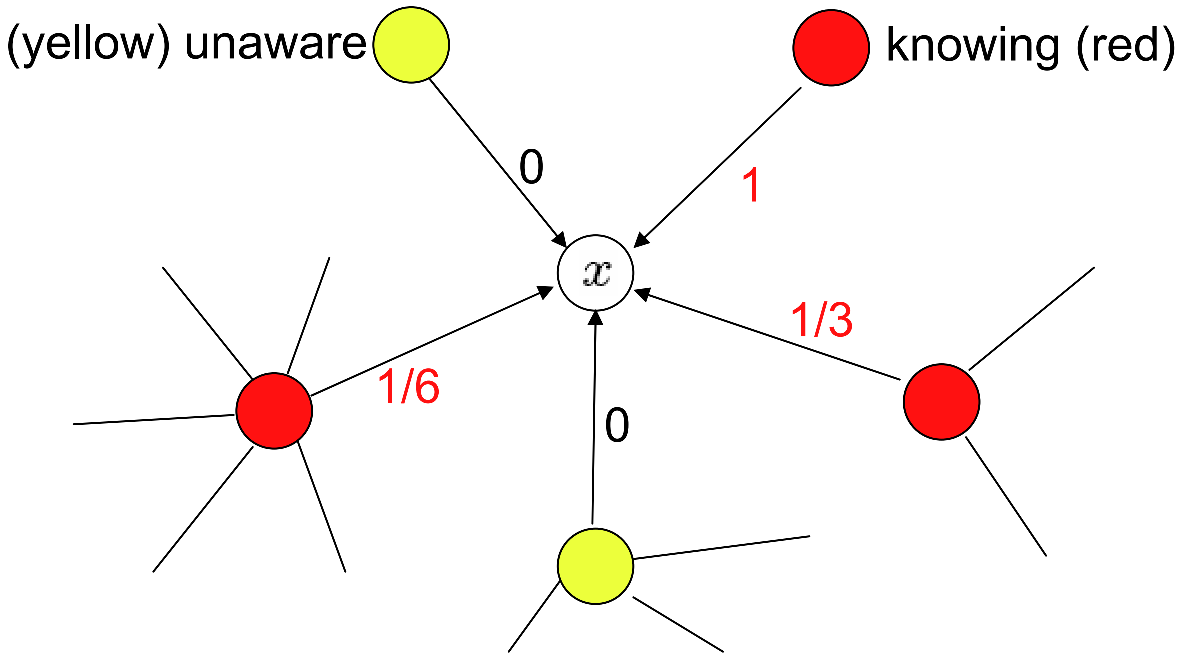
## local observables

$$\Omega_t(x) = \sum_{x\sim y}\omega(y) \qquad \text{number of knowing neighbours of } x$$ **= 3**

$$\Phi_t(x) = \sum_{x\sim y}\frac{1}{d(y)}\omega(y) \qquad \text{local knowledge inflow}$$ **=1+1/3+1/6 =1.5**

**Inner structure of projects is *not* FullGraph, but now we account for that:**

**1/degree weighing of the knowing neighbours**



# (ε) epsilon-process ~classical epidemics

- Local infection by knowing neighbours
- The epsilon-process has a very low probability ε, but:
- The more neighbours knowing, the higher the probability to get knowing:

$$\mathbf{P}_{0\to 1} = \epsilon \cdot \Phi(x)$$

$$\Phi_t(x) = \sum_{x\sim y} \frac{1}{d(y)} \omega(y)$$

But this rather weak epsilon process only happens below a threshold … $1 \leq \Omega(x) < \Delta$

# (α) alpha-process:
# delta-threshold infection

If the *number of knowing neighbours* exceeds a threshold Δ

$$\Omega(x) \geq \Delta$$

suddenly there is a *higher* probability α to get knowing

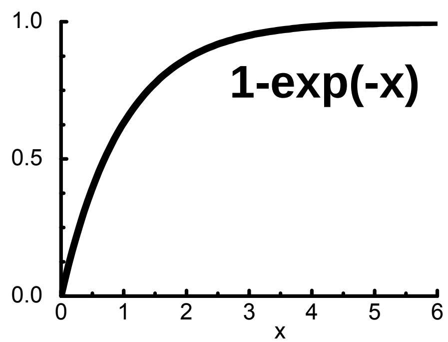
α >> ε

$$\mathbf{P}_{0\rightarrow 1} = \alpha \cdot \left(1 - e^{-\Phi(x)}\right)$$

Degree weighed inflow

$$\Phi_t(x) = \sum_{x\sim y} \frac{1}{d(y)} \omega(y) \quad \text{shift into } [0...1]:$$

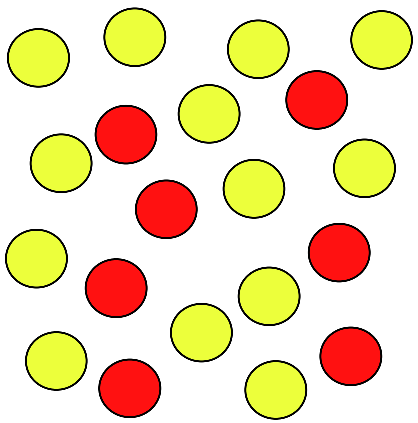
$$\left(1 - e^{-\Phi_t(x)}\right) = \begin{cases} \sim 0 & \text{for } \Phi_t(x) \text{ small} \\ \sim 1 & \text{for } \Phi_t(x) \text{ large} \end{cases}$$



# (β) beta-process: mean-field influence infects

- i.e. mass media, intuition about the state of the whole system, journals, …= „mean-field“.
- Proportional to square of relative prevalence

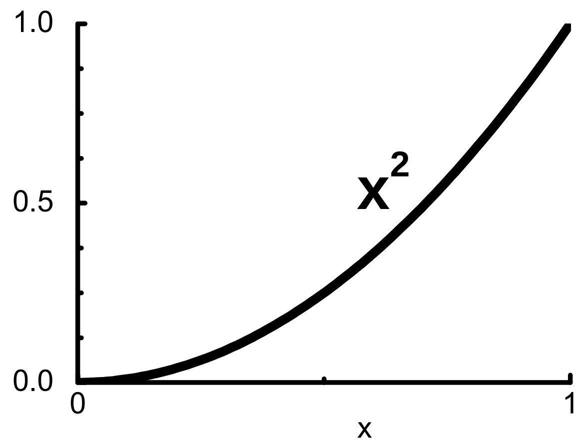
$$\mathbf{P}_{0\to 1} = \beta(b_t)^2$$



$$b_t = \frac{1}{N}\sum_{x=1}^{N}\omega(x)$$

total knowledge prevalence

7 / 20 = 0.35



# (γ) gamma-process: forgetting passive knowledge

- The less-knowing my neighbours, the higher my γ-process-forgetting

Ratio of **unaware** neighbours

$$\mathbf{P}_{1\to 0} = \gamma\left(1 - \frac{\Omega_t(x)}{d(x)}\right)$$

Ratio of **knowing** neighbours

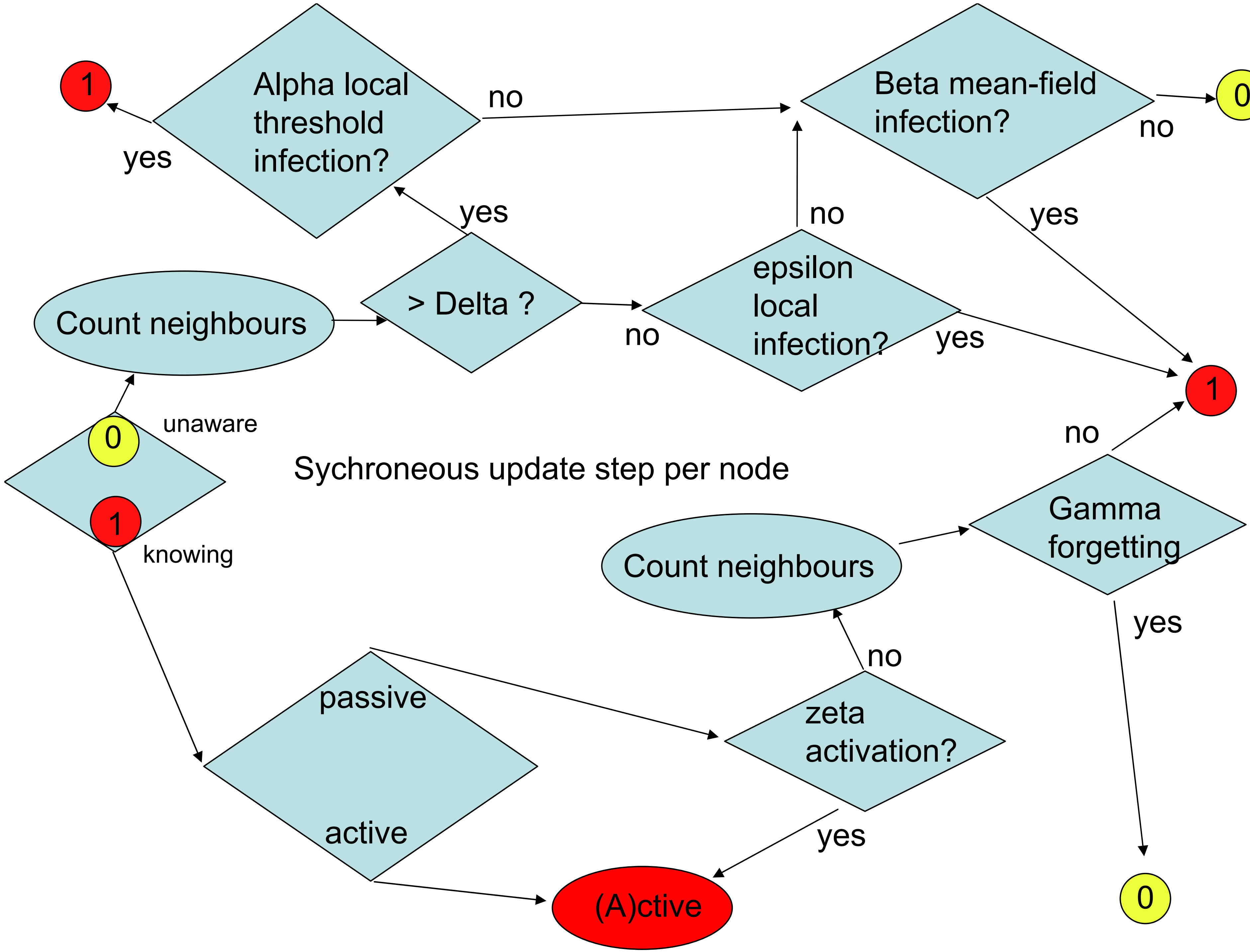
But I can only forget PASSIVE knowledge. ACTIVE knowledge stays with me…

# (ζ) zeta-process: activation of passive knowledge

- Each time step there is a (constant) probability ζ to get from „passive“ to „active“ knowledge

$$P_{1 \to A} = \zeta$$

- Only passive knowledge can be forgotten. Once activated, the node stays knowing forever.
- Possible extension: Active knowledge „counts“ more than passive knowledge (e.g. A=3)

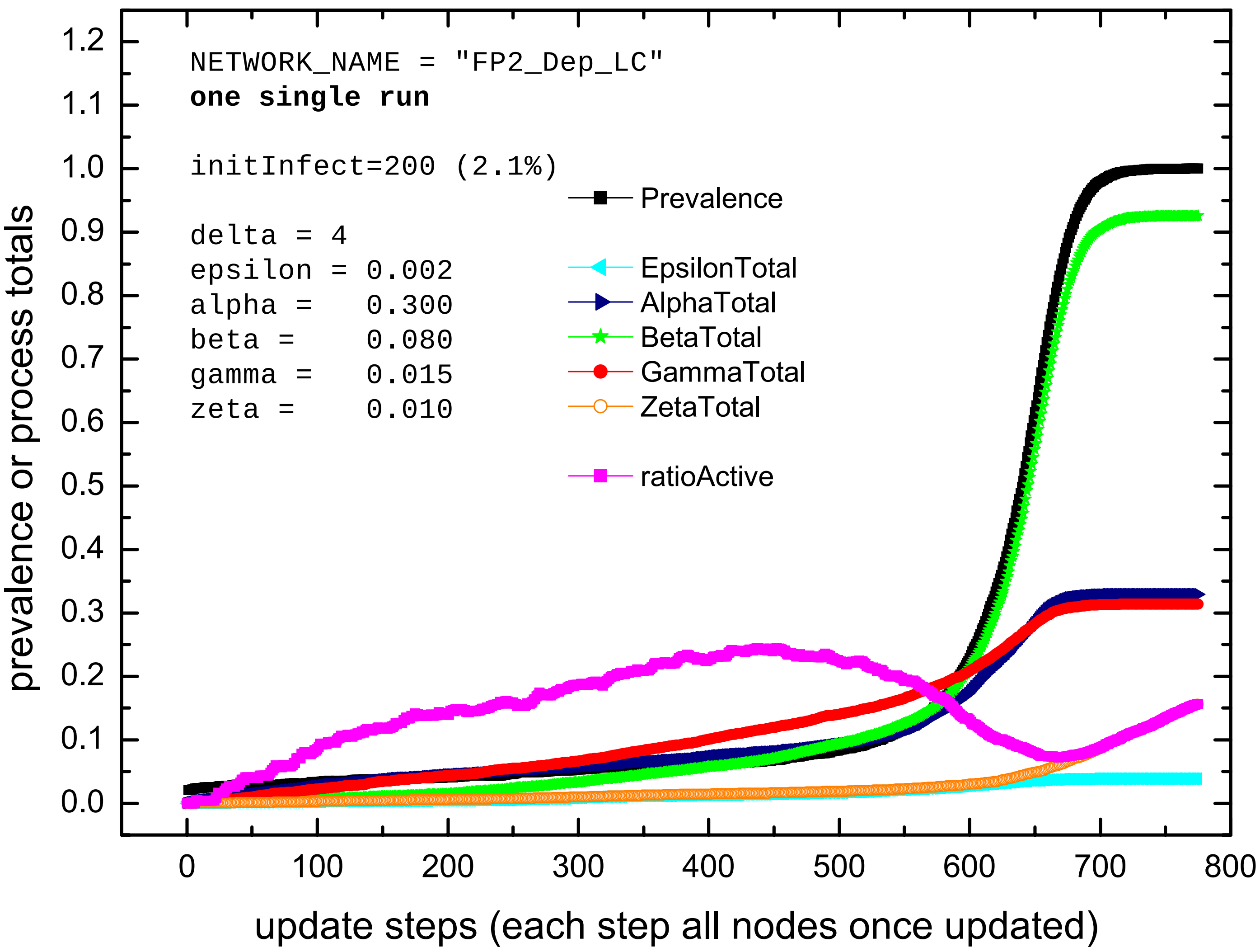


Sychroneous update step per node
0
unaware
1
knowing
Count neighbours
> Delta ?
yes
no
Alpha local threshold infection?
yes
1
no
Beta mean-field infection?
no
0
yes
epsilon local infection?
no
yes
1
passive
active
zeta activation?
no
yes
(A)ctive
Count neighbours
Gamma forgetting
no
yes
0

# GEP runs

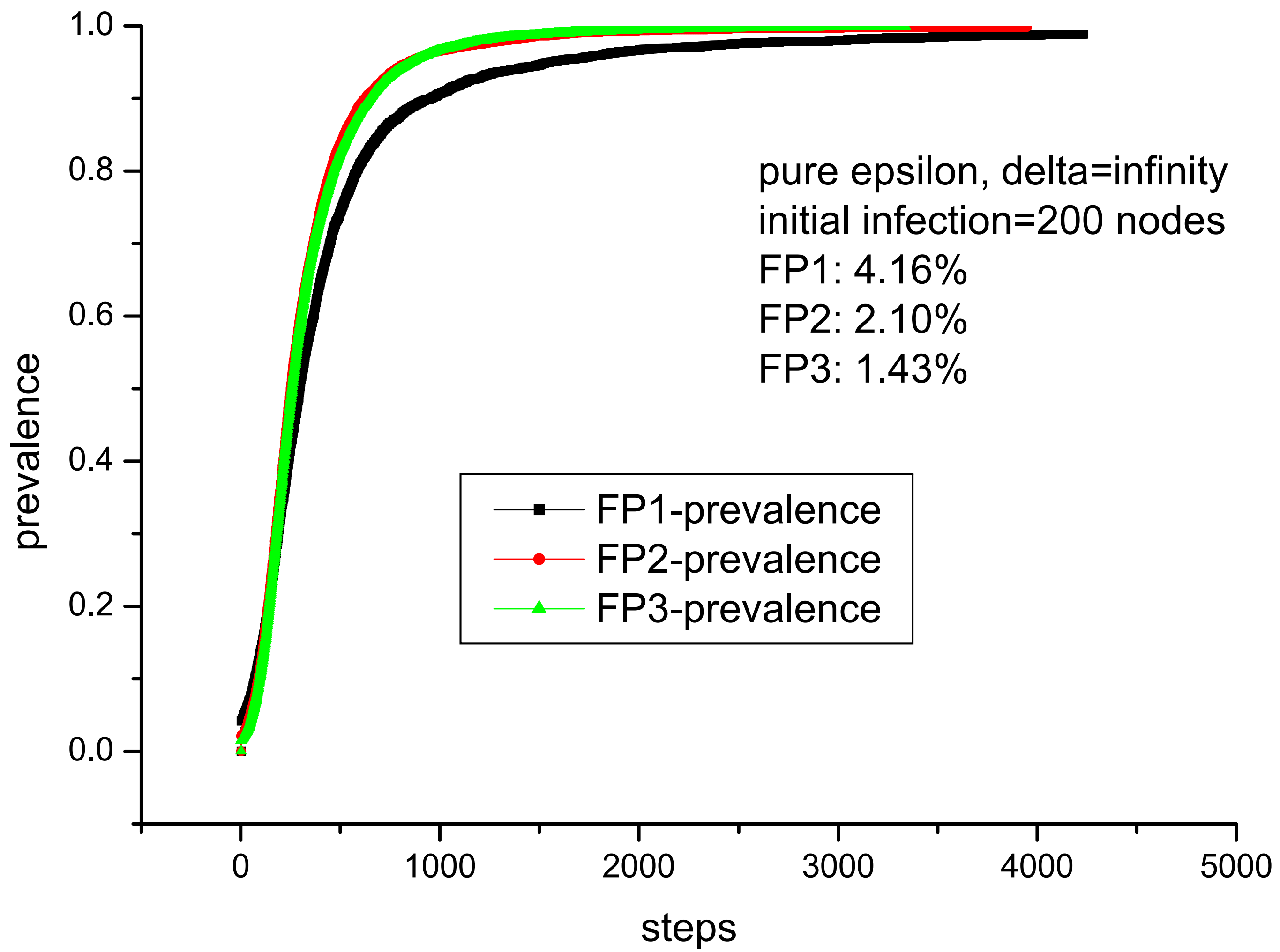
One infection run, on FP2

initial infection: 200 nodes



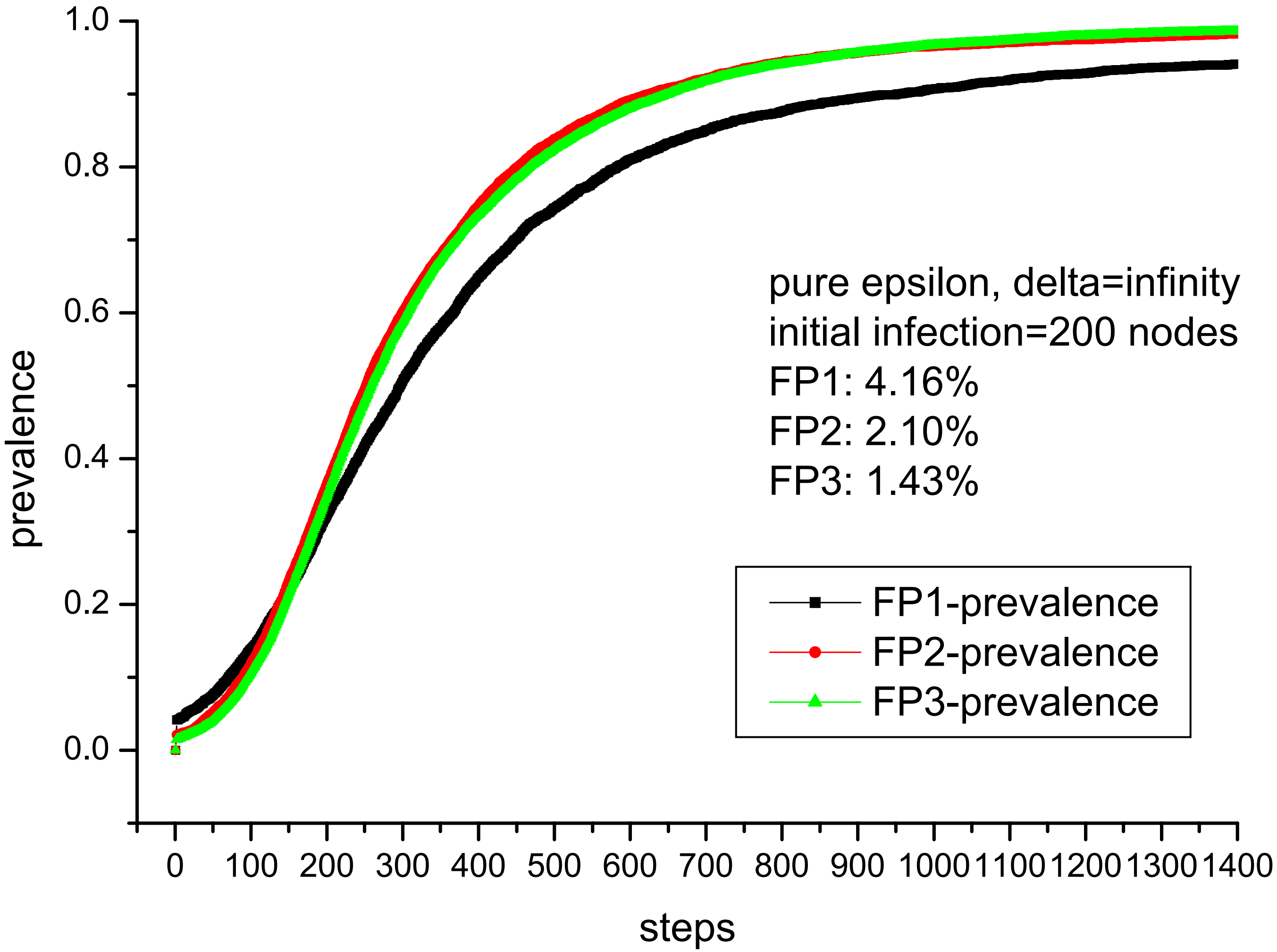
# Pure Epsilon process epsilon=0.04

alpha=0 beta=0 gamma=0 zeta=0 delta=infinity

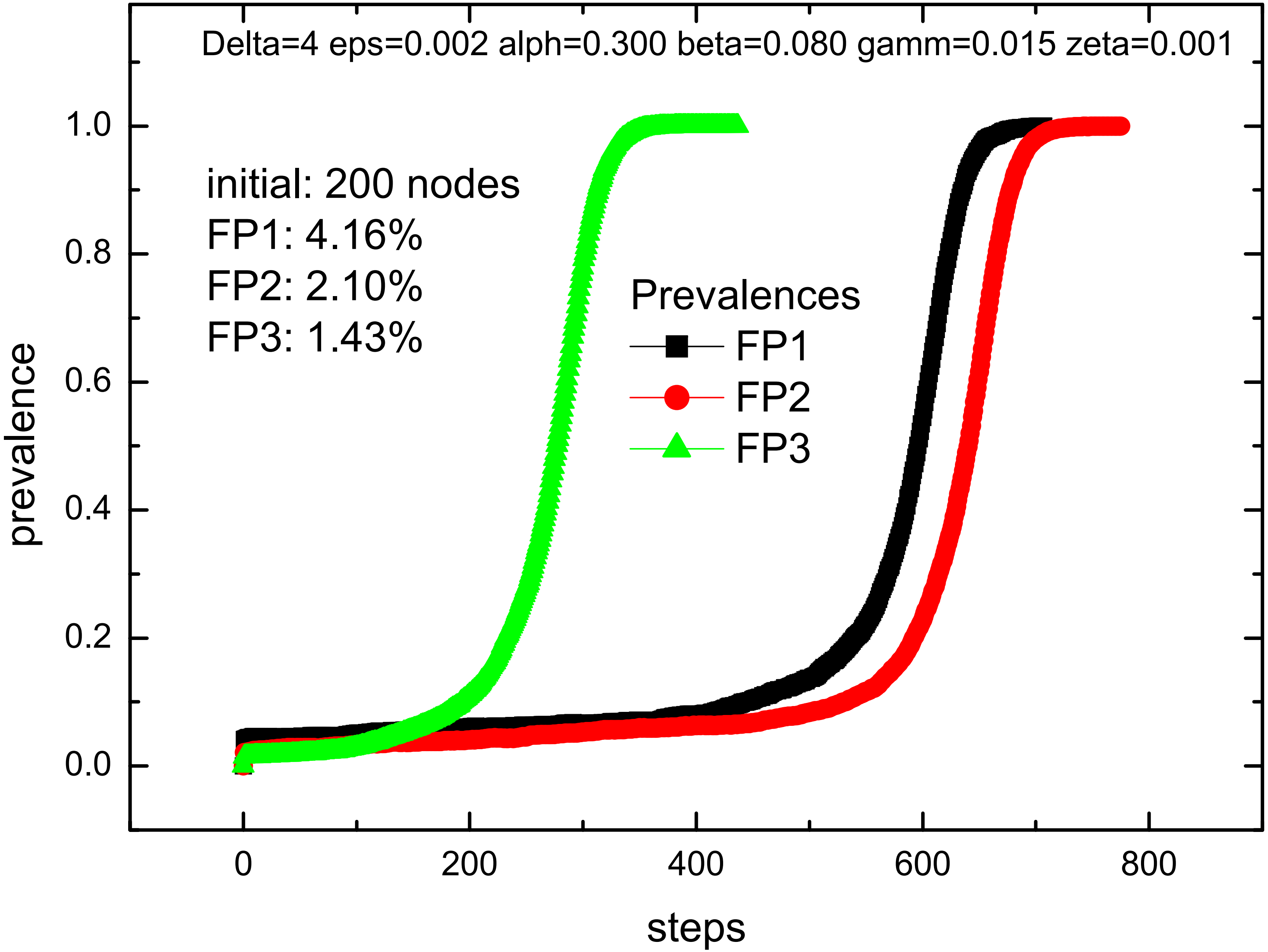


# Pure Epsilon process epsilon=0.04

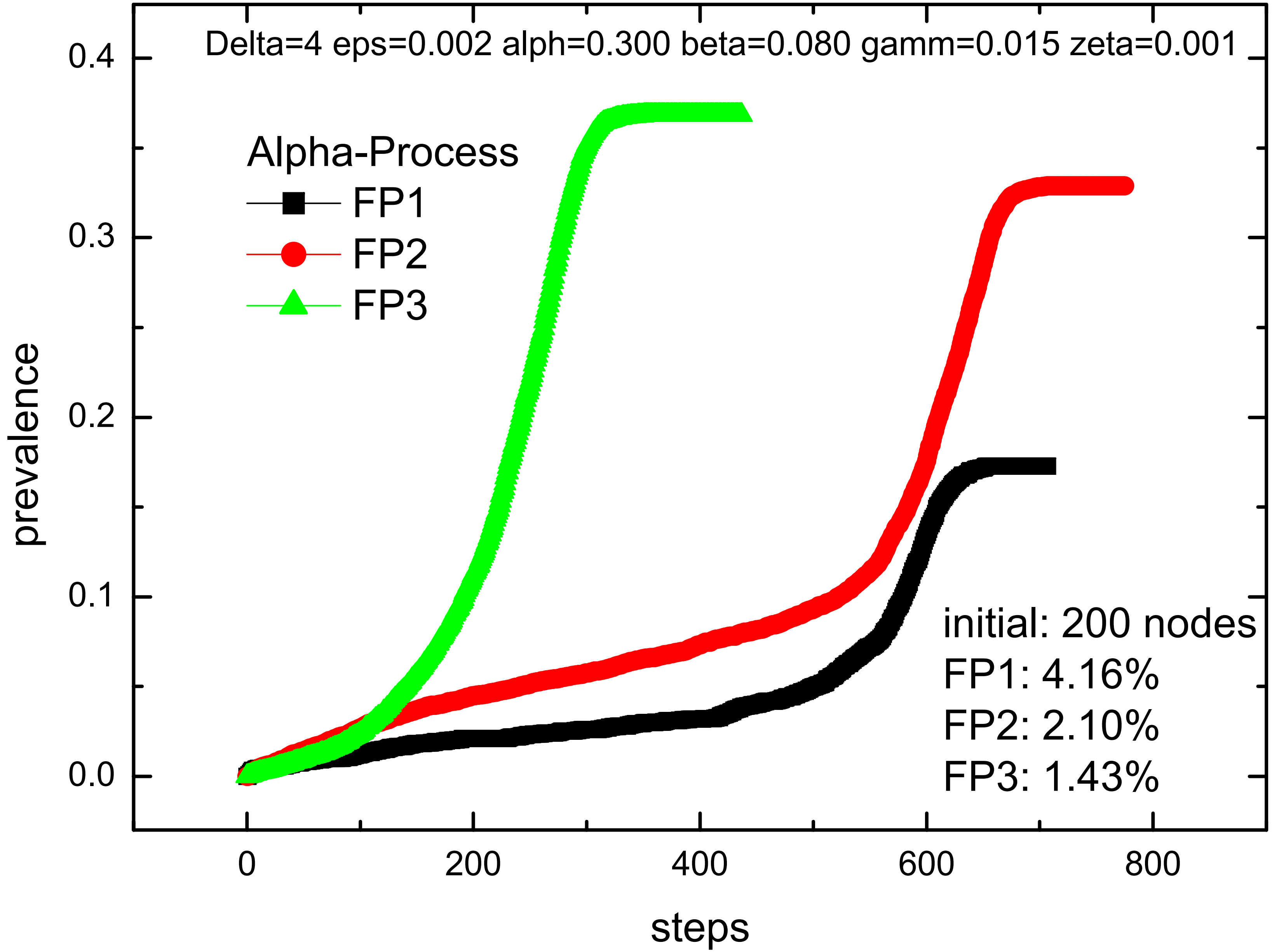
alpha=0 beta=0 gamma=0 zeta=0 delta=infinity



One infection run: **Prevalence** FP1 FP2 FP3 - comparison



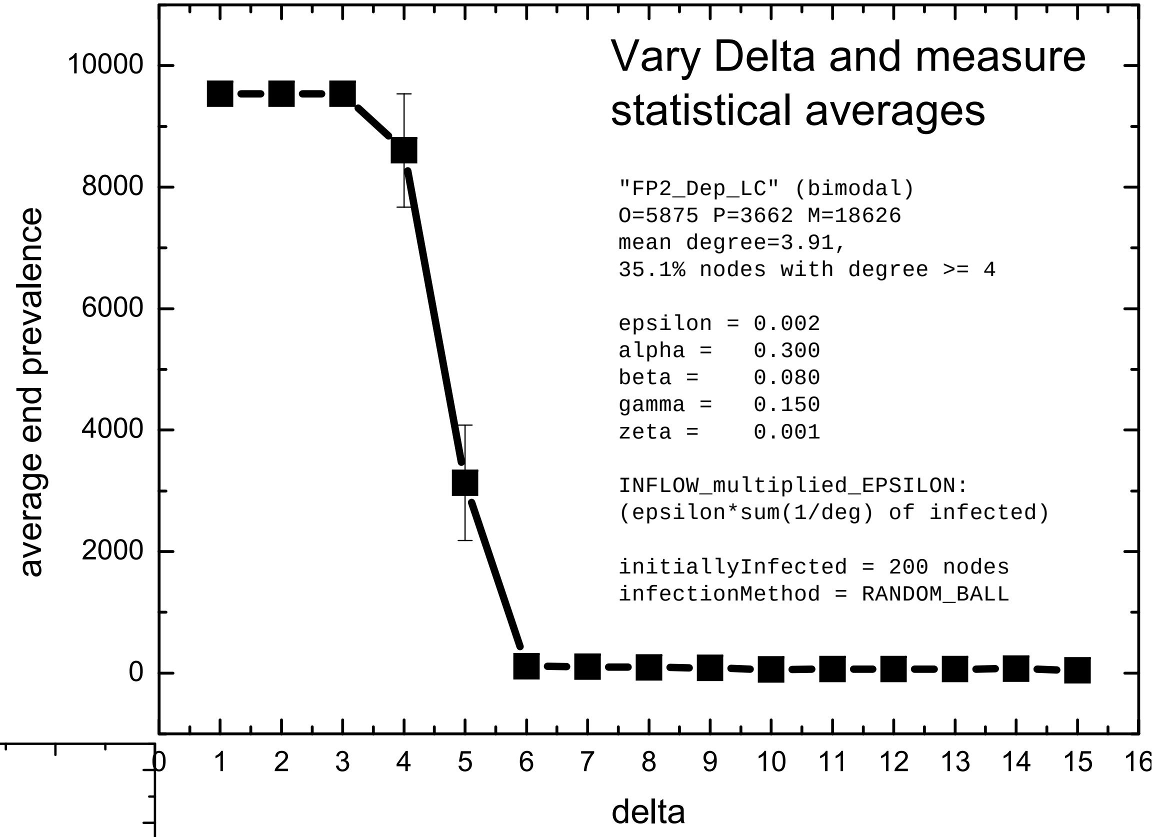
One infection run: **Alpha-Process** FP1 FP2 FP3 - comparison

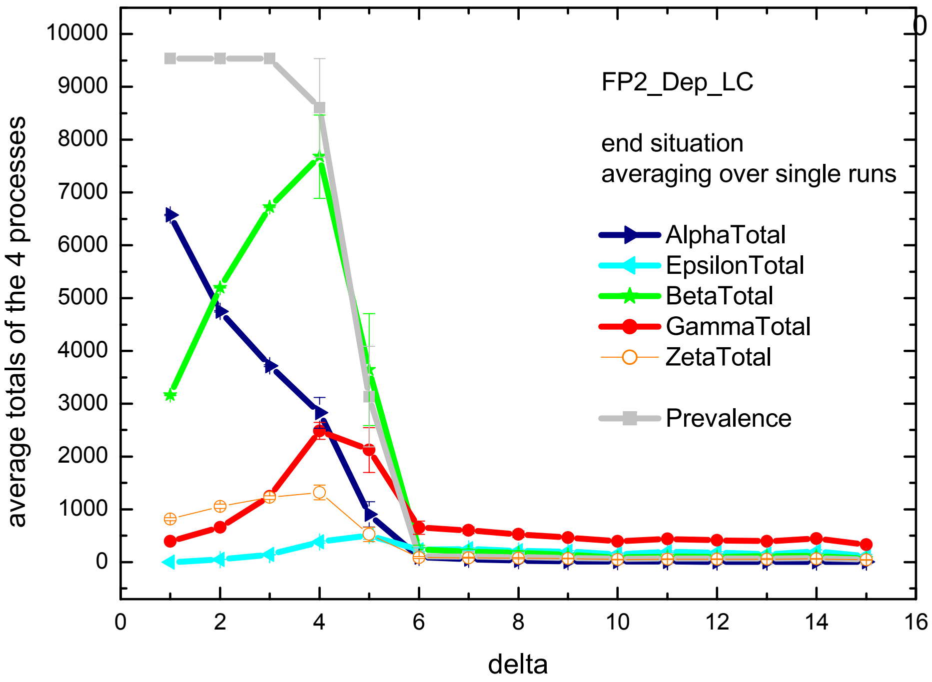


# FP2

Variation of delta

Many runs,
averages of **end** results

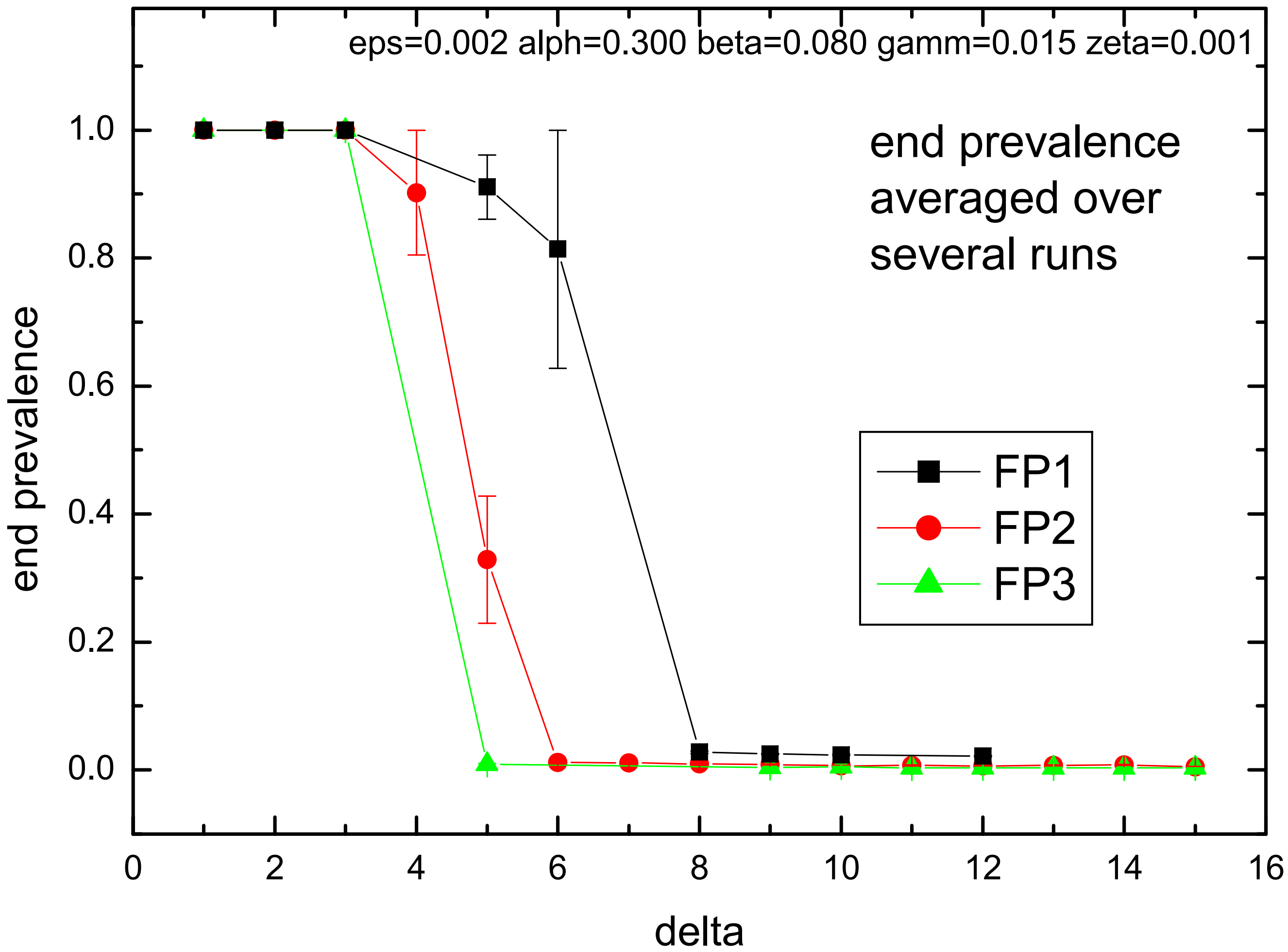


# FP1, 2, 3

# Variation of delta-threshold

Many runs, averages of **end** results



Planned *next* extensions:

- infectious time is only short after infection
- competing knowledge types:
  - first steps into high-dimensional knowledge representation
  - no *active* knowledge of all types possible
  - Majority rules for local and mean-field processes

# Part 2: stylized facts about knowledge..

NEMO
NETWORK MODELS, GOVERNANCE AND R&D COLLABORATION NETWORKS

Project no. NEST-2006-028875

NEMO

Network Models, Governance and R&D Collaboration Networks

Instrument: Specific Targeted Research Project (STREP)
Thematic Priority: NEST-Adventure

**Deliverable D1.1**
**Conceptual and empirical foundations of R&D network dynamics**


Manfred Paier
Petra Ahrweiler
Michael Barber
Andreas Brandes
Barbara Heller-Schuh
Nicolas Jonard
Magnus Kutz
Andreas Pyka
Bulat Sanditov
Petra Wagner-Luptacik
Matthias Weber



Submission date: 2007-09-16

Start date of project: 2006-09-01 Duration: 36 months

Organisation name of lead contractor for this deliverable: Austrian Research Centers GmbH – ARC


| Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006) | | |
|---|---|---|
| **Dissemination Level** | | |
| PU | Public | |
| PP | Restricted to other programme participants (including the Commission Services) | |
| RE | Restricted to a group specified by the consortium (including the Commission Services) | |
| CO | Confidential, only for members of the consortium (including the Commission Services) | X |

# Knowledge: 3 fields of application in NEMO

- **<u>knowledge diffusion</u>** = i.e. "spreading" of innovation
  - in an *existing* network - **how does knowledge spread**?
  - example: the term "***chaos***" (for nonlinear systems) has spread all over science
  - our GEP model adresses knowledge spreading (Part 1)
- projects: **<u>partner choice</u>** by their knowledge tuple
  - for project goal *and* for self-interest
  - choice of best match among network partners – locally …
  - … and global search
- **<u>network variation</u>**
  - after a (successful) project
  - strengthen and weaken some bonds
  - change the network accordingly

# Knowledge Types

- active versus passive knowledge
  - no forgetting of active knowledge
  - usage activates knowledge
  - active knowledge likely to have stronger infection effect

- Knowledge $K_i$ is a tuple in high-dimensional *metric* space

# Projects: *choice of partners*

- for mutual understanding
  **shared** knowledge necessary
- incentive for cooperation is
  **complementary** knowledge
- *project interest* versus all *self-interests*
- secret and non-transferable knowledge, patents
- trust ~ history
  - relational embeddedness: (positive) past experiences
  - structural embeddedness: indirect ties
- project-coordinators are
  more likely to be chosen again

# *Choice of partners*: Heterogeneity!

- local ties vs. cluster-spanning
  - banks experienced with *cluster-spanning ties* are more likely to to establish cluster-spanning ties in the future
  - … versus *within-clique-ties*
- more vs selective
  - organisations with *few* contacts tend to *add more* partners
  - organisations with *many* contacts are likely to be *selective*

# Project

- project <u>knowledge stock</u>
  - "sum" of participants
  - sum over knowledge-*sub*space only!
- during <u>cooperation</u>
  - generation of completely new knowledge
  - activation of passive knowledge
  - knowledge exchange, by imitation
  - directed versus mutual learning
- project <u>goal</u>
  - project *goal*: certain knowledge tuple
  - project *outcome*: certain knowledge tuple
  - *products* can be produced
    once the necessary knowledge tuple exists

# Project

- project proposal
  - Project goal is evaluated
  - Project *goal in proposal* might be different …
  - … from project *outcome that <u>can</u> be produced* from knowledge stock!
  - proposal generation phase is knowledge exchange already (even if project is not contracted)
- inner structure of project
  - 4-10 work packages
    - cooperation is mainly done in workpackages
    - intra project structure 11%-99% density of FullGraph
  - overlay of management/coordination network: star network or star-of-cliques

# Knowledge outside projects

- organisations bring *initial knowledge*
- working outside cooperation *also* increases knowledge
- mobility of knowledge workers across firms

# Limits of the individual (organisation)

- absorptive capacities:
  per timestep limited learning
- heterogeneity of individuals "intelligence„
- no one knows everything
- misunderstanding
- forgetting

# „subjectively meaningful“

- context-embedded knowledge
- Individual knowledge stock determines what *can* be learnt
- received knowledge less, different or other than sended knowledge
- from the outside, systems cannot be "informed" with a certain and sure knowledge transfer, but rather activated to learn themselves from given offers

# Part 3:

# multidimensional knowledge & partner choice

$x=(x_1,\ldots, x_n)$

$y=(y_1,\ldots, y_n)$

$|x_i - y_i| \geq \delta_{comp}$

$|x_i - y_i| \leq \delta_{shared}$

**Complementary Knowledge Subspace**

$D_{comp} = \{i : |x_i - y_i| \geq \delta_{comp}\}$

$d_{comp} = \#D_{comp}$

complementary knowledge number

**Shared Knowledge Subspace**

$D_{shared} = \{i : |x_i - y_i| \leq \delta_{shared}\}$

$d_{shared} = \#D_{shared}$

shared knowledge number

The probability $P(x \sim y)$ that $x$ cooperates with $y$ and produces new knowledge is a monotone rising function of the size of common speech, measured by $d_{shared}$:

$$P(x \sim y) \propto d_{shared}$$

The value of this new knowledge component is a monotone rising function of the number of topics in which the cooperators complement each other. This number is measured by $d_{comp}$:

$$x_{n+1} \propto d_{comp}$$

Thank you.

# Additional Slides

# Knowledge representation

As this is an interdisciplinary workshop, example first ☺

Given 4 knowledge tupels $K_1, K_2, K_3, K_4$

$$K_1 = (1, 1, 0...0)$$
$$K_2 = (1, 1, 0...0)$$
$$K_3 = (0, 1, 0...0)$$
$$K_4 = (0, 0, 1, 0..0)$$

Let us define a *distance function* $d$ with this behaviour:

- $d(K_1, K_2) = 0$
  knowledge $K_1$ and $K_2$ are *identical*
- $d(K_1, K_3) < d(K_1, K_4)$ both compared to $K_1$,
  knowledge $K_3$ is *more similiar* than $K_4$

# Knowledge representation: Metric space

A *metric space* $(M, d)$ are a set $M$ and a *distance* function $d : M \times M \rightarrow \mathbb{R}$.

The function $d$ can **compare** elements of $M$ and fulfills 3 conditions:

| | | |
|---|---|---|
| (1) | *identity of indiscernibles* | $d(x, y) = 0$ only if $x = y$ |
| (2) | *symmetry* | $d(x, y) = d(y, x)$ |
| (3) | *triangle inequality* | $d(x, y) + d(y, z) \geq d(x, z)$ |

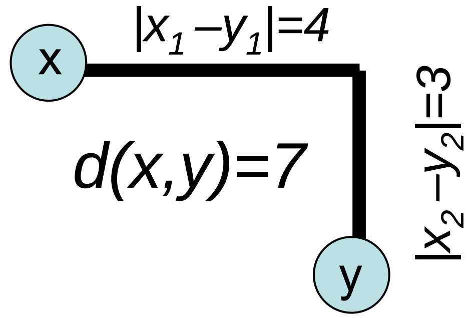
these three conditions combine to the property of

| | | |
|---|---|---|
| (4) | *non-negativity* | $d(x, y) \geq 0$ |

Let the *knowledge space* $\mathfrak{K} = \mathbb{R}^n$ be an $n-$dimensional metric space, over the real numbers $\mathbb{R}$. A knowledge tuple $K \in \mathfrak{K}$ can then be written as real coefficients $K = (k_{1,} k_{2,} ..., k_n)$ with $k_i \in \mathbb{R}$

Any function with conditions 1-4 (e.g. the Euclidian distance) can serve as a metric. Example: the **Manhattan distance** sums differences in all dimensions:

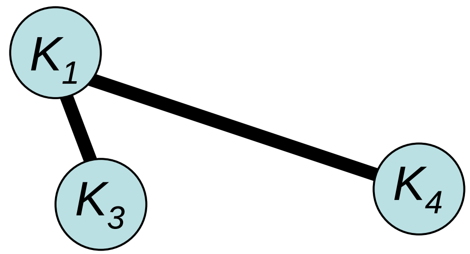
$$d(x,y) = \sum_{i=1}^{n} |x_i - y_i|$$
$$= |x_1 - y_1| + |x_2 - y_2| + ... + |x_n - y_n|$$




---

Given 4 knowledge tupels $K_1, K_2, K_3, K_4$

$$K_1 = (1, 1, 0...0)$$
$$K_2 = (1, 1, 0...0)$$
$$K_3 = (0, 1, 0...0)$$
$$K_4 = (0, 0, 1, 0..0)$$

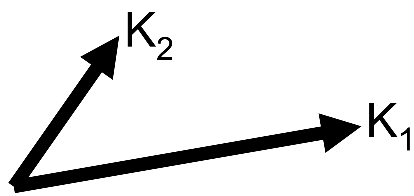



Metric spaces are about DISTANCES

- $d(K_1, K_2) = 0 = |1 - 1| + |1 - 1| + |0 - 0| + ...$
- $d(K_1, K_3) = 1 = |1 - 0| + |1 - 1| + |0 - 0| + ...$
- $d(K_1, K_4) = 3 = |1 - 0| + |1 - 0| + |0 - 1| + |0 - 0| + ...$

in <u>contrast</u> to *metric* space: *Vector space*

"within NEMO, an agreement has been made to represent knowledge pragmatically as **<u>vectors</u>** in multidimensional space"
(Deliverable D1.1 p. 43) → ???




## <u>Some properties of *Vector* Spaces</u>

Vector addition

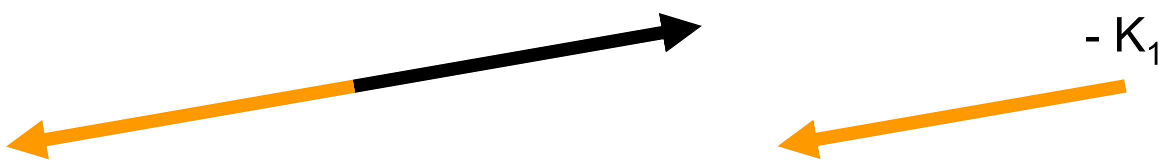
(parallelogram rule)


---

Vector multiplied by field element („scaling“)


---

Vector addition has inverse elements




For vector v exists vector w so that v+w=0 → ???

# *All* simulation results up to now

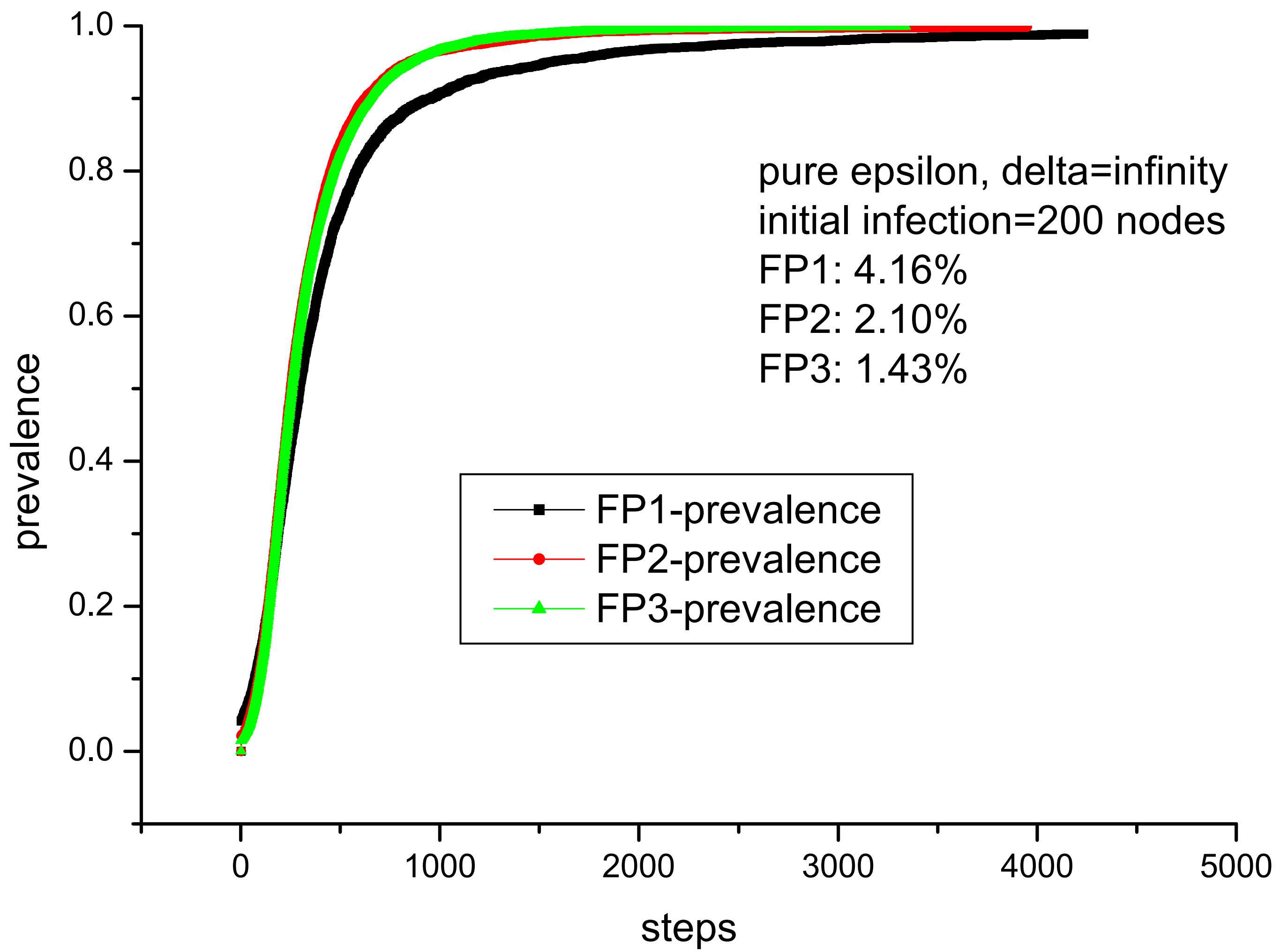
(before was a selection only)

# Pure epsilon process

classical epidemics + 1/degree-effect

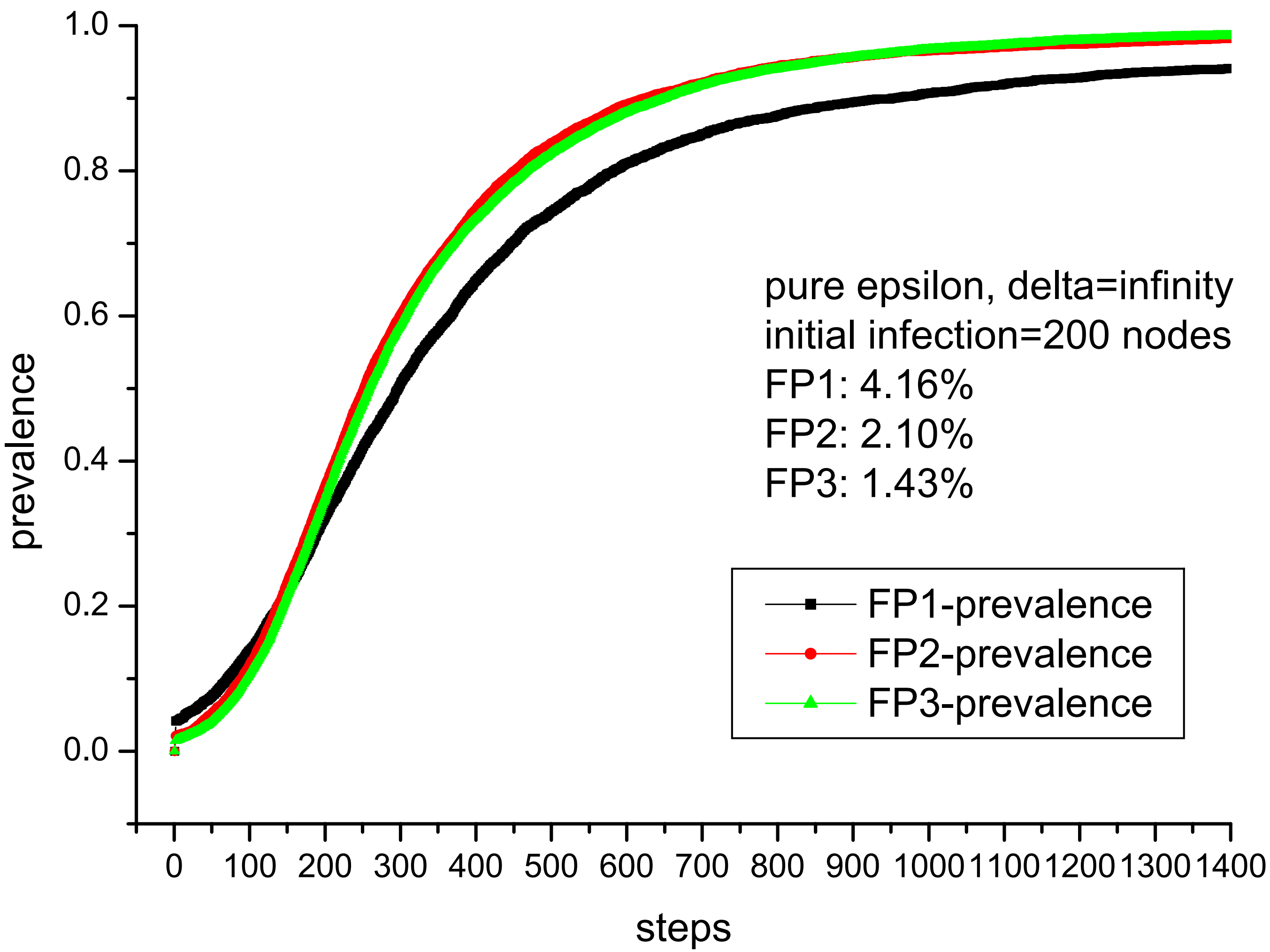
# Pure Epsilon process epsilon=0.04

alpha=0 beta=0 gamma=0 zeta=0 delta=infinity



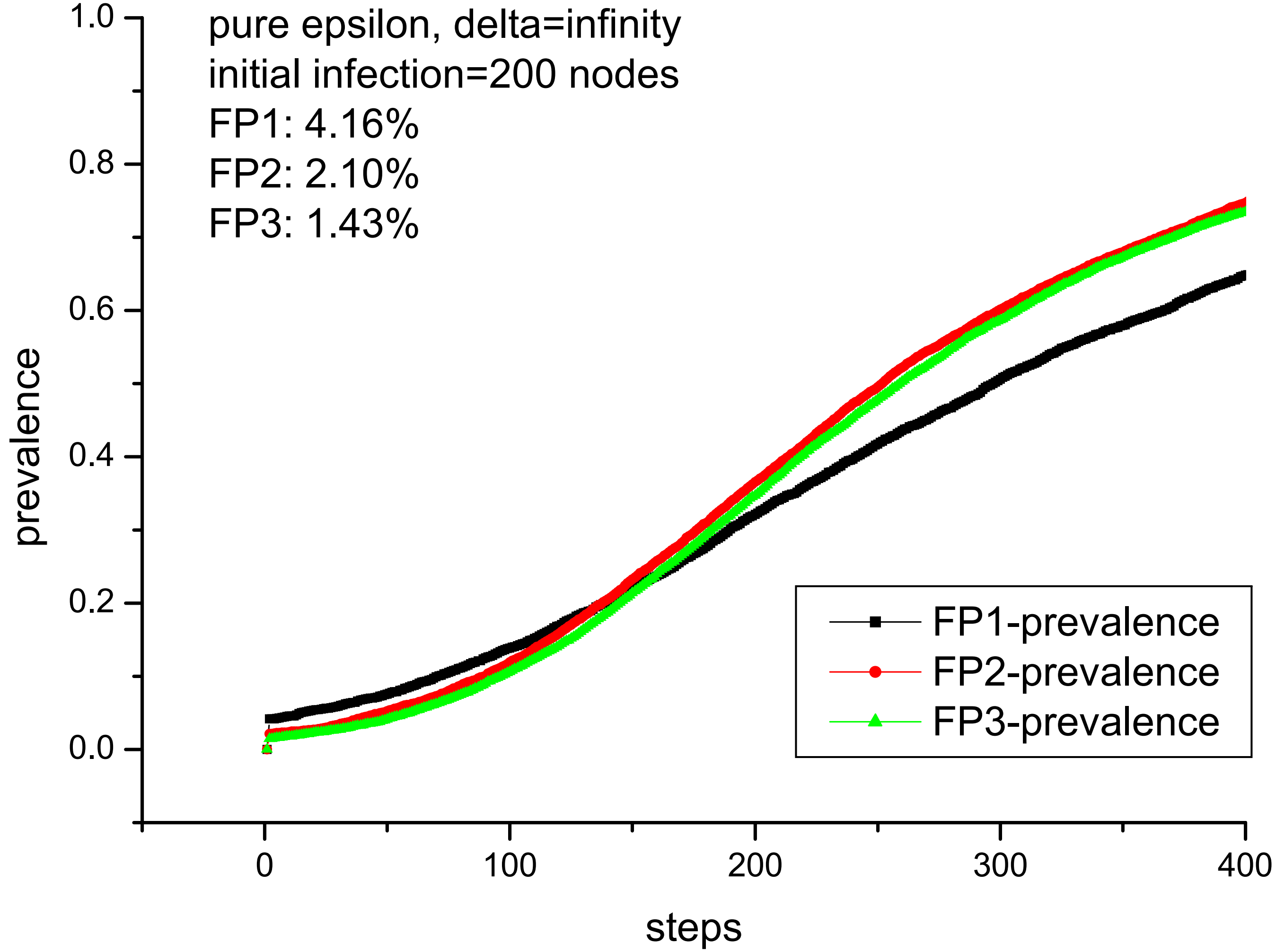
# Pure Epsilon process  epsilon=0.04

alpha=0 beta=0 gamma=0 zeta=0 delta=infinity



# Pure Epsilon process  epsilon=0.04

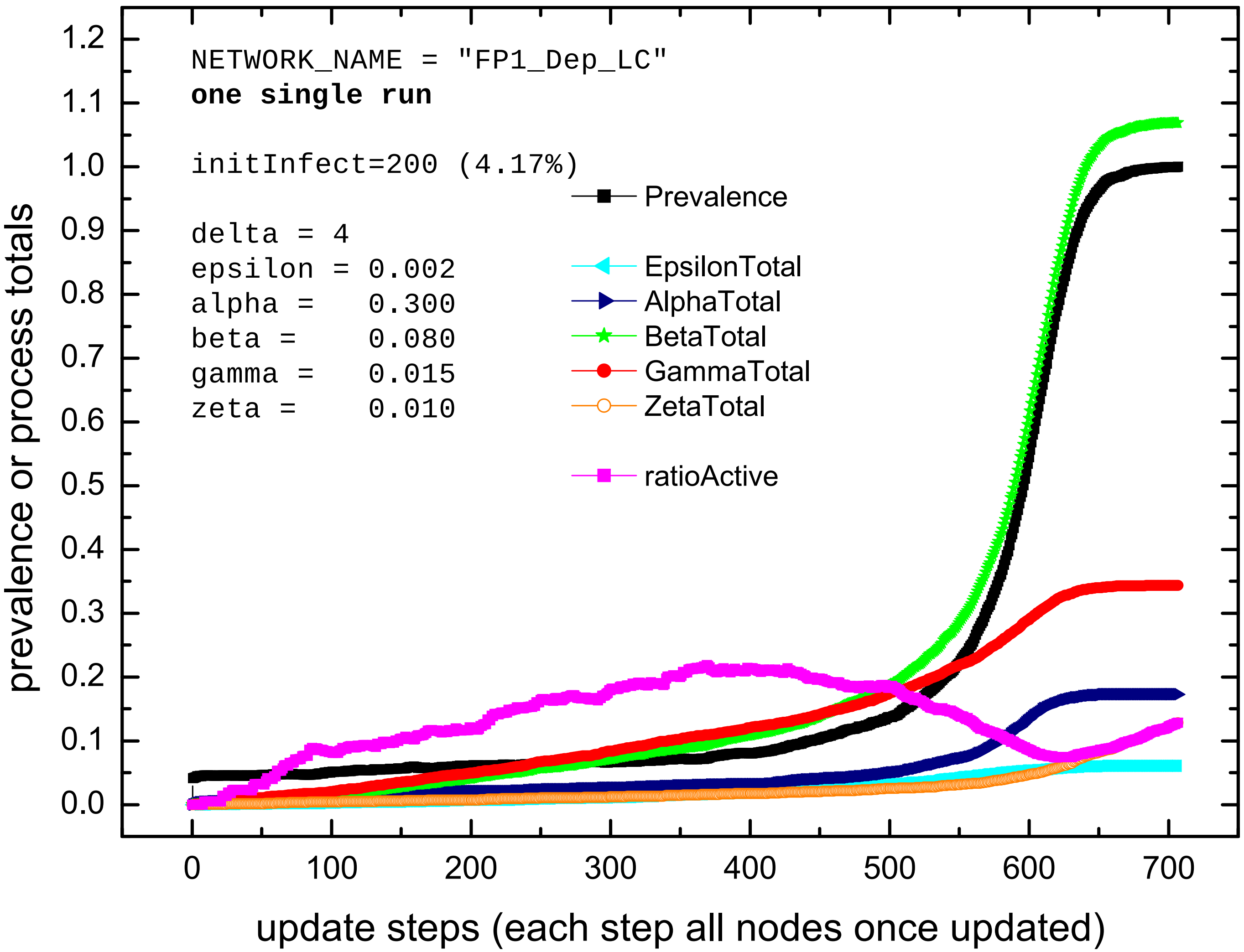
alpha=0 beta=0 gamma=0 zeta=0 delta=infinity



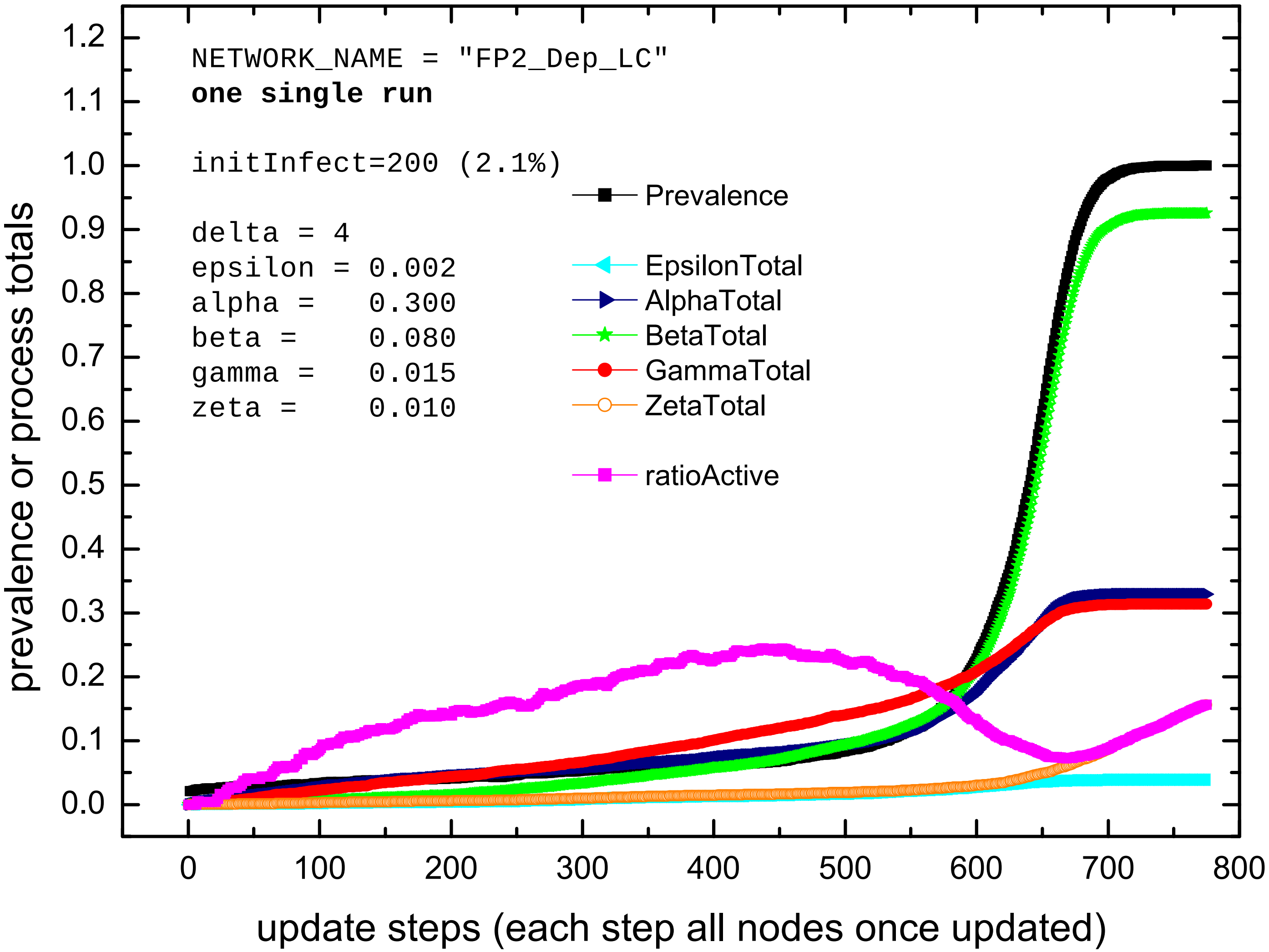
all processes switched on

One infection run, on FP1

initial infection: 200 nodes

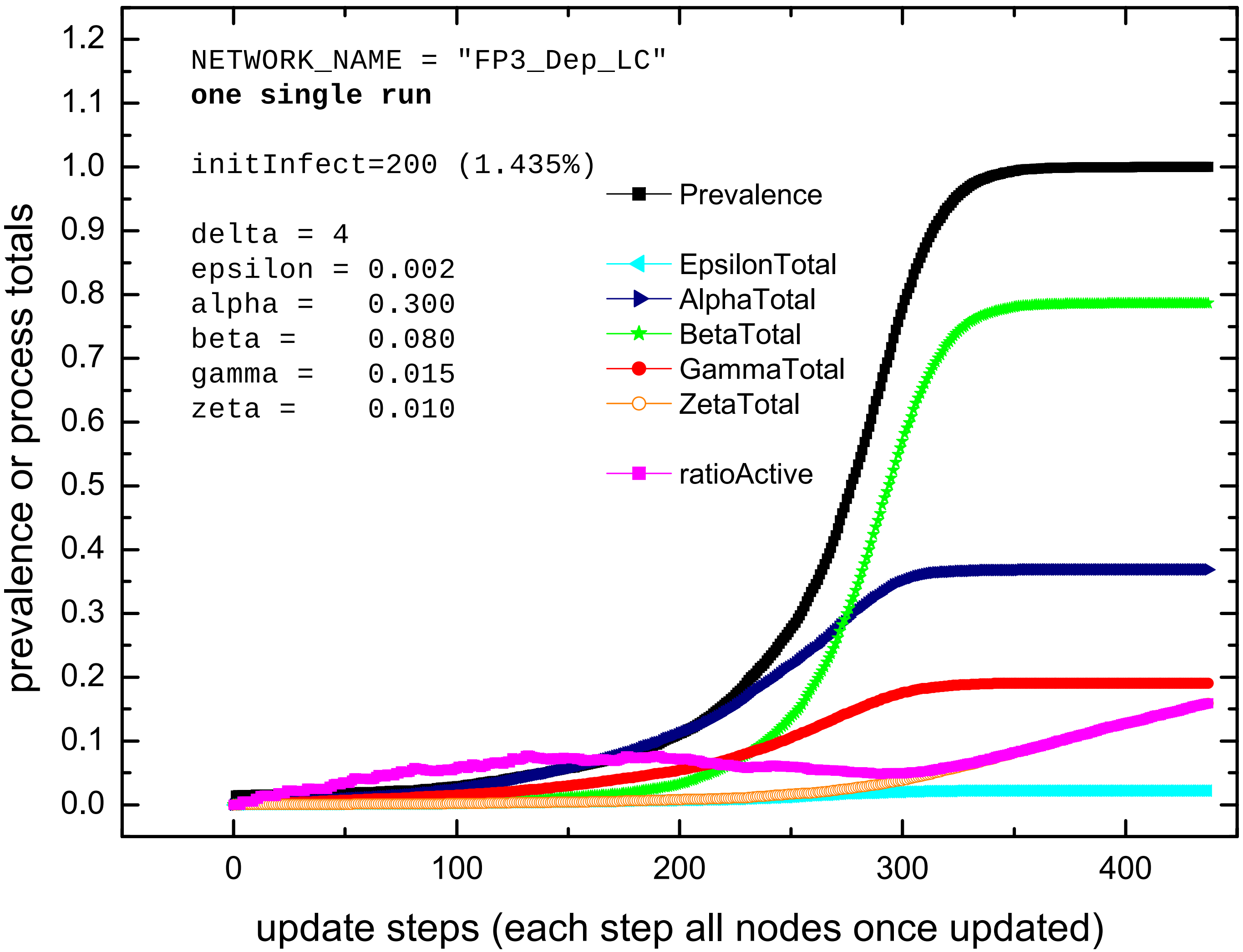


One infection run, on FP2 initial infection: 200 nodes



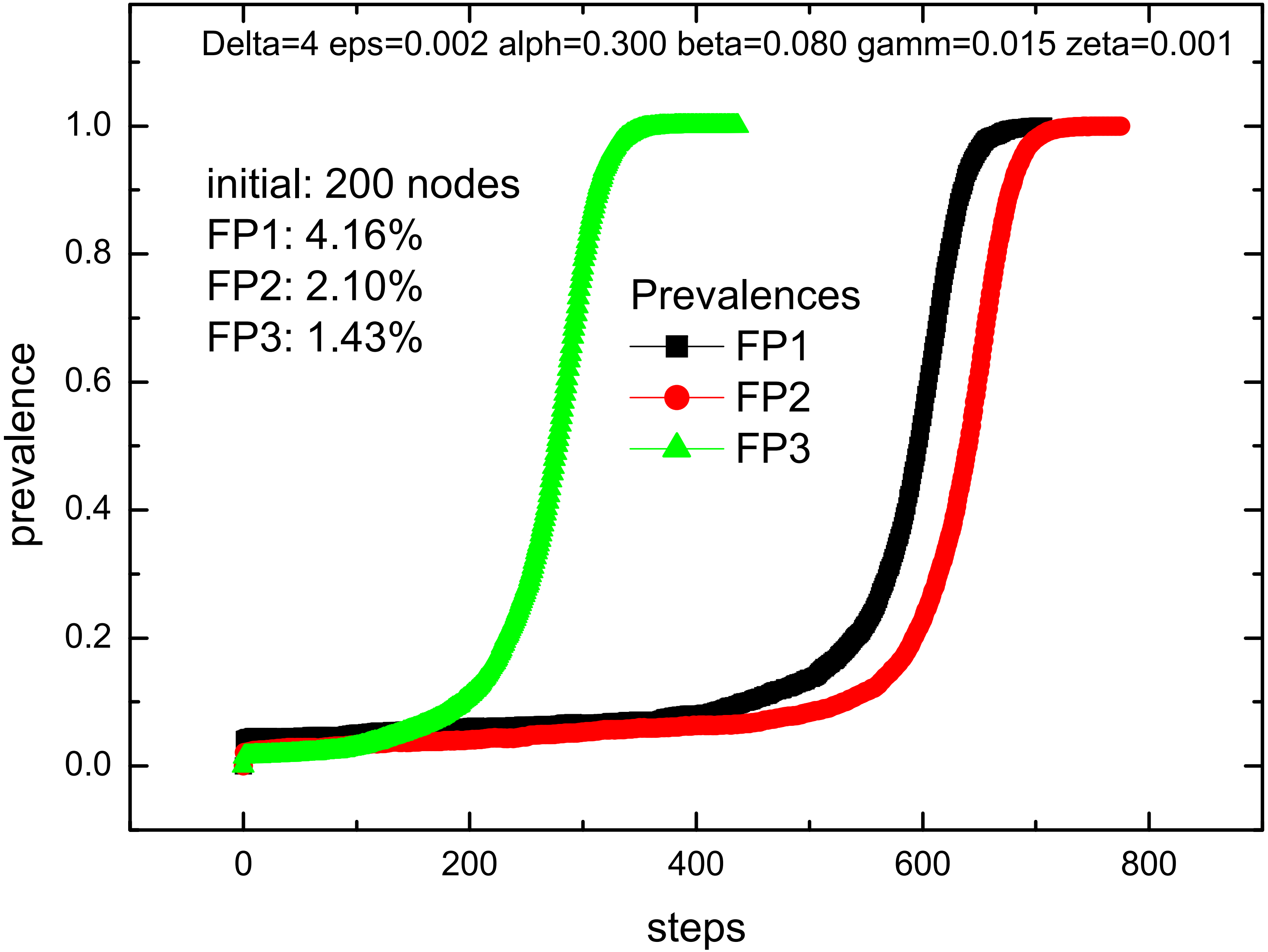
One infection run, on FP3

initial infection: 200 nodes

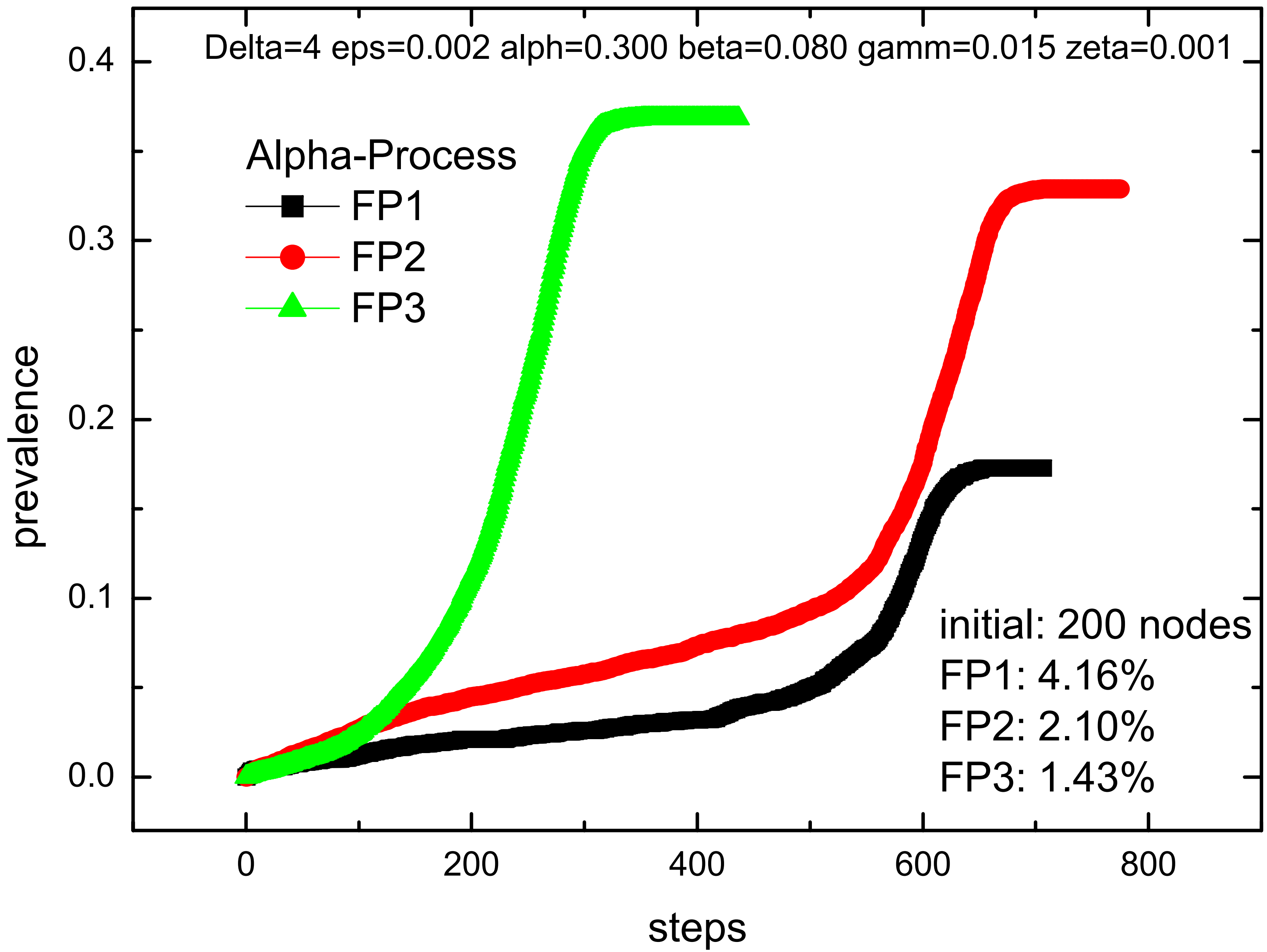


# Comparison of FP 1 2 3

One infection run: **Prevalence** FP1 FP2 FP3 - comparison



One infection run: **Alpha-Process** FP1 FP2 FP3 - comparison



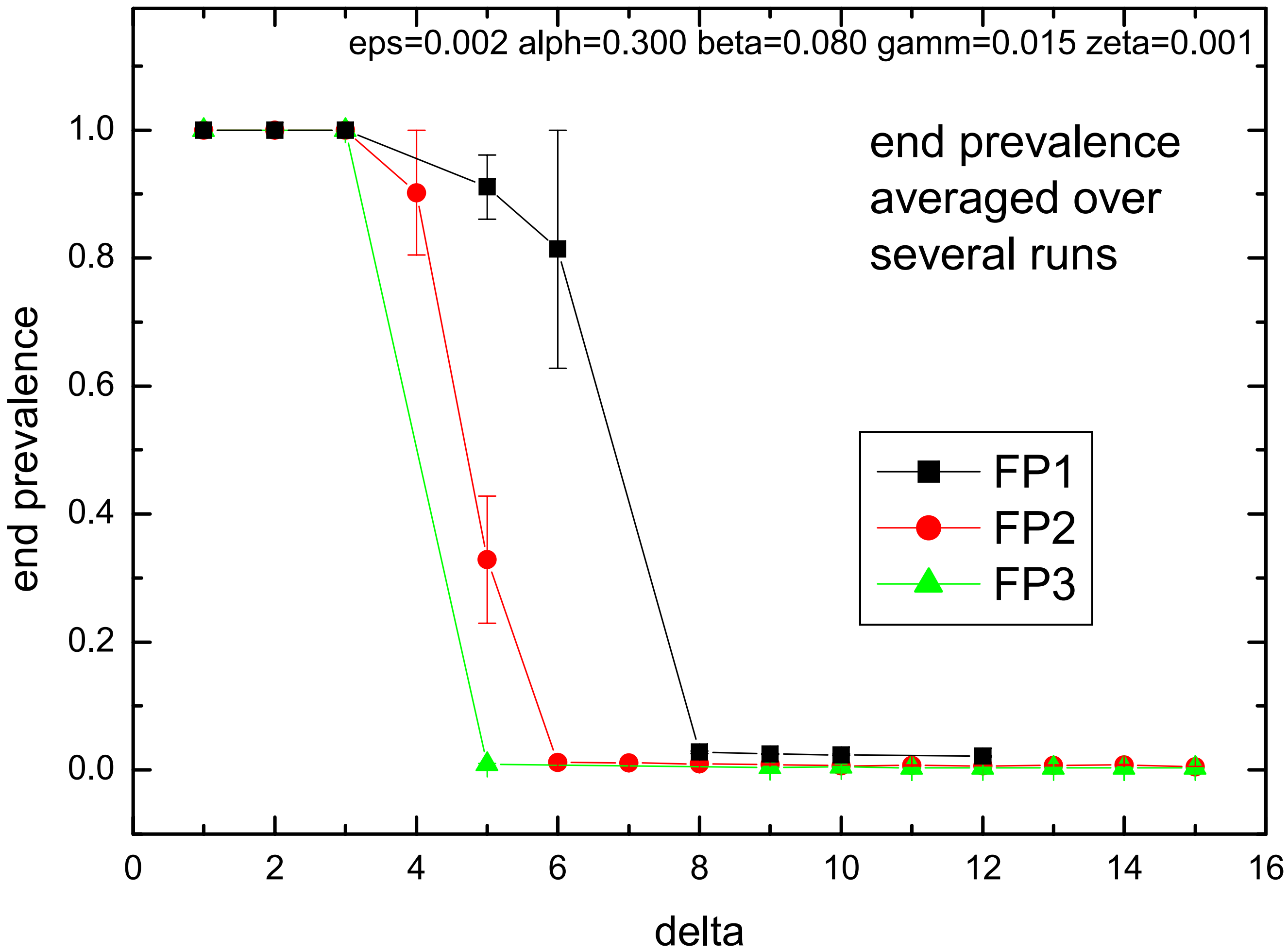
# delta sweep

- Variation of delta
  - Small delta = „big news“
  - Big delta: I need to hear it many times

# FP1, 2, 3

# Variation of delta-threshold

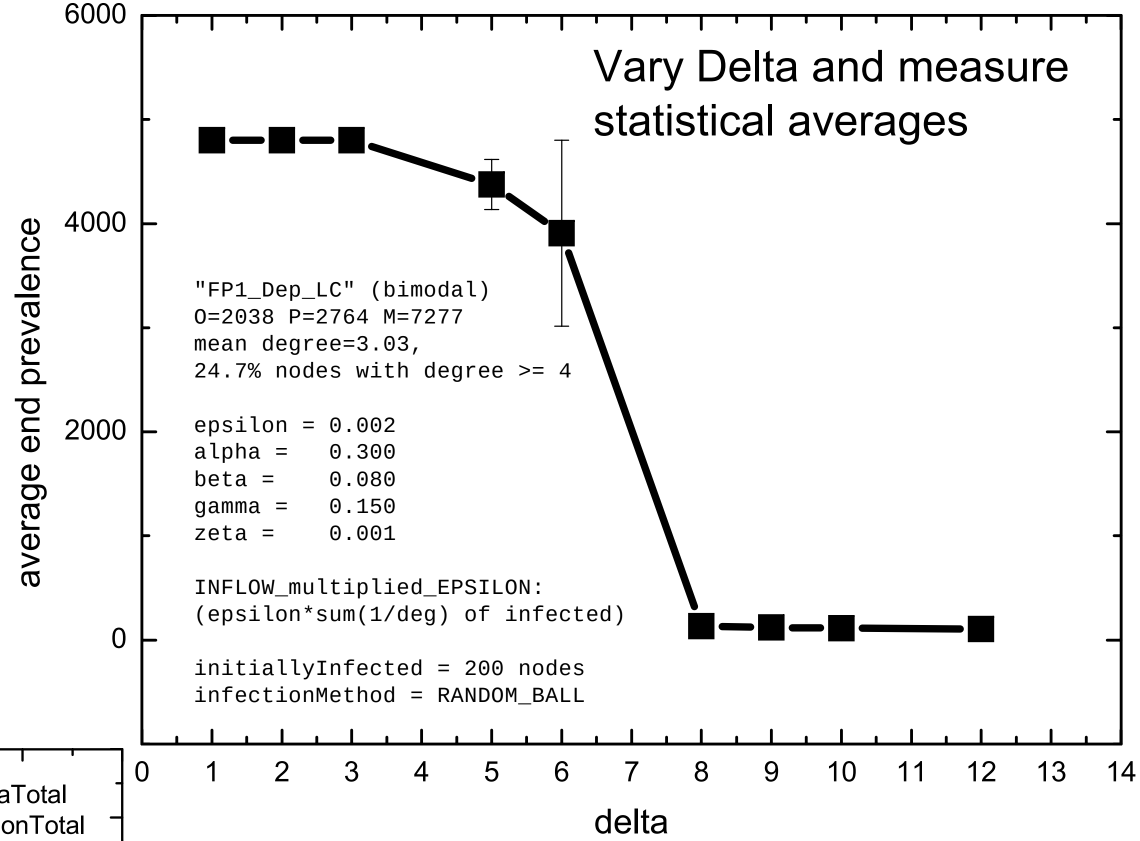
Many runs, averages of **end** results

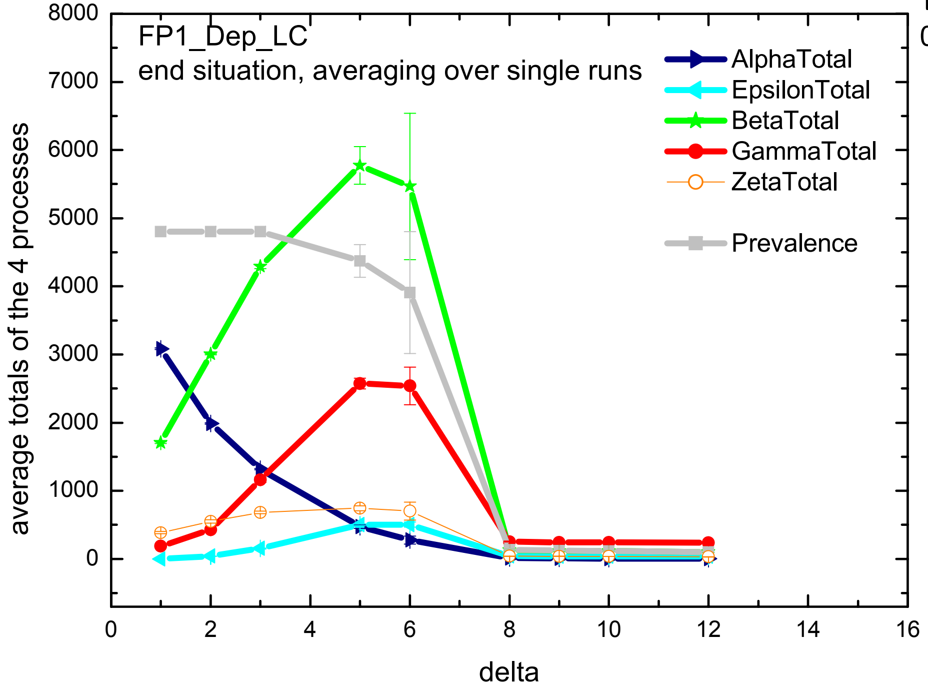


# FP1

Variation of delta

Many runs,
averages of **end** results

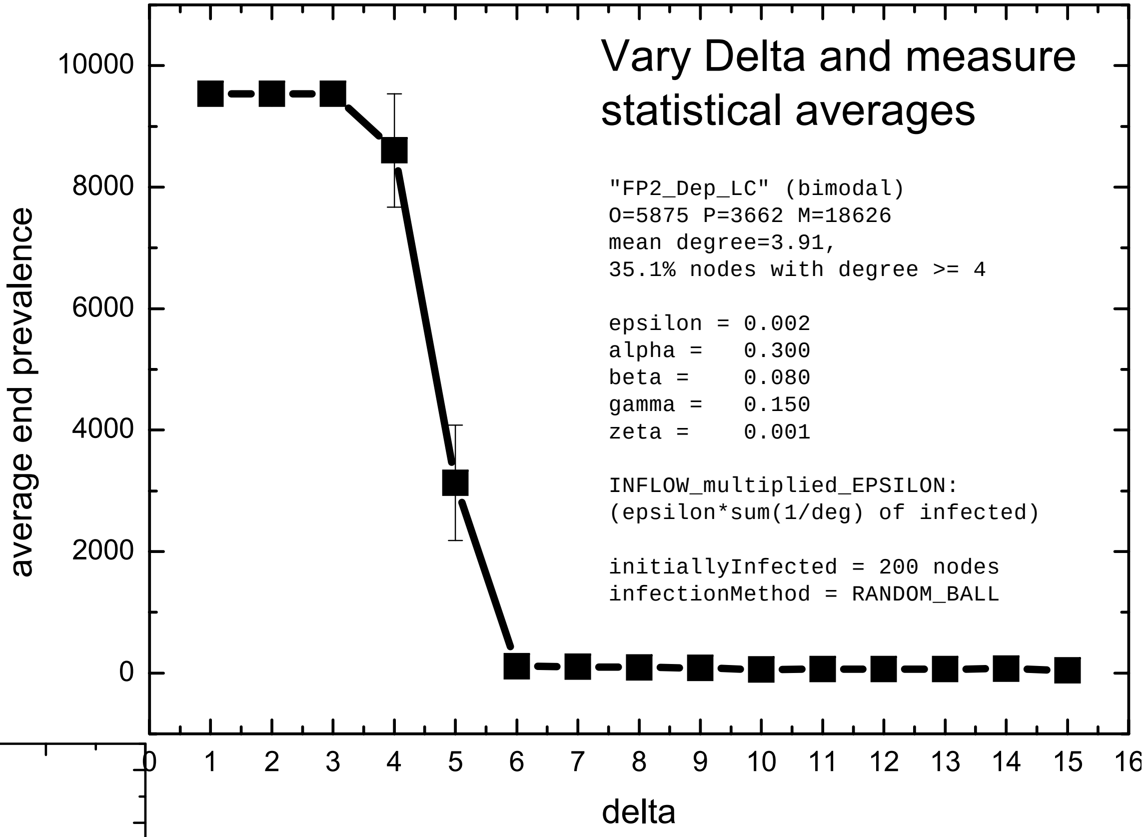


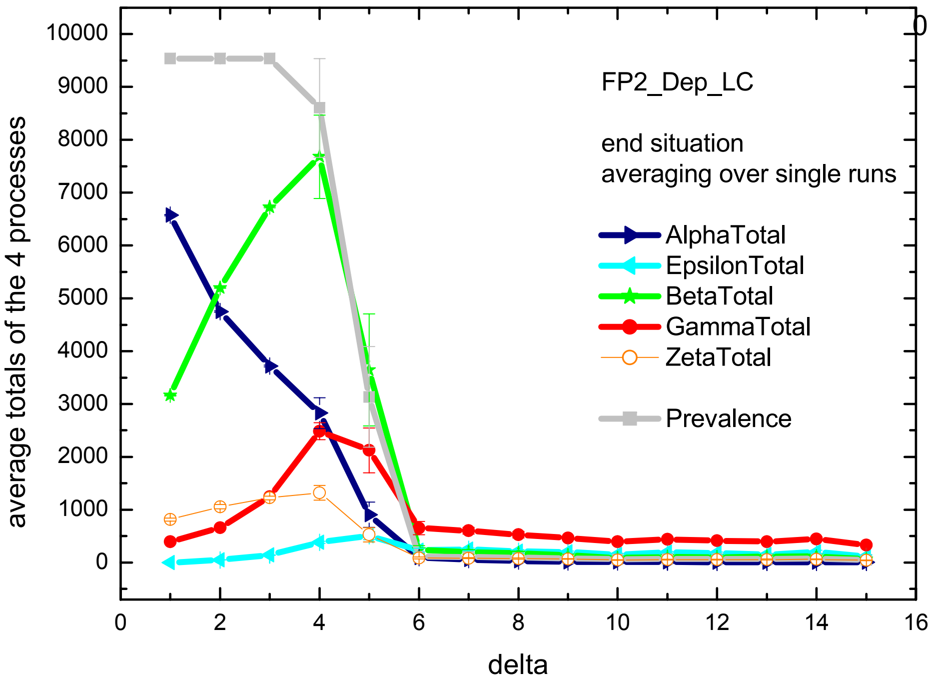



# FP2

Variation of delta

Many runs,
averages of **end** results

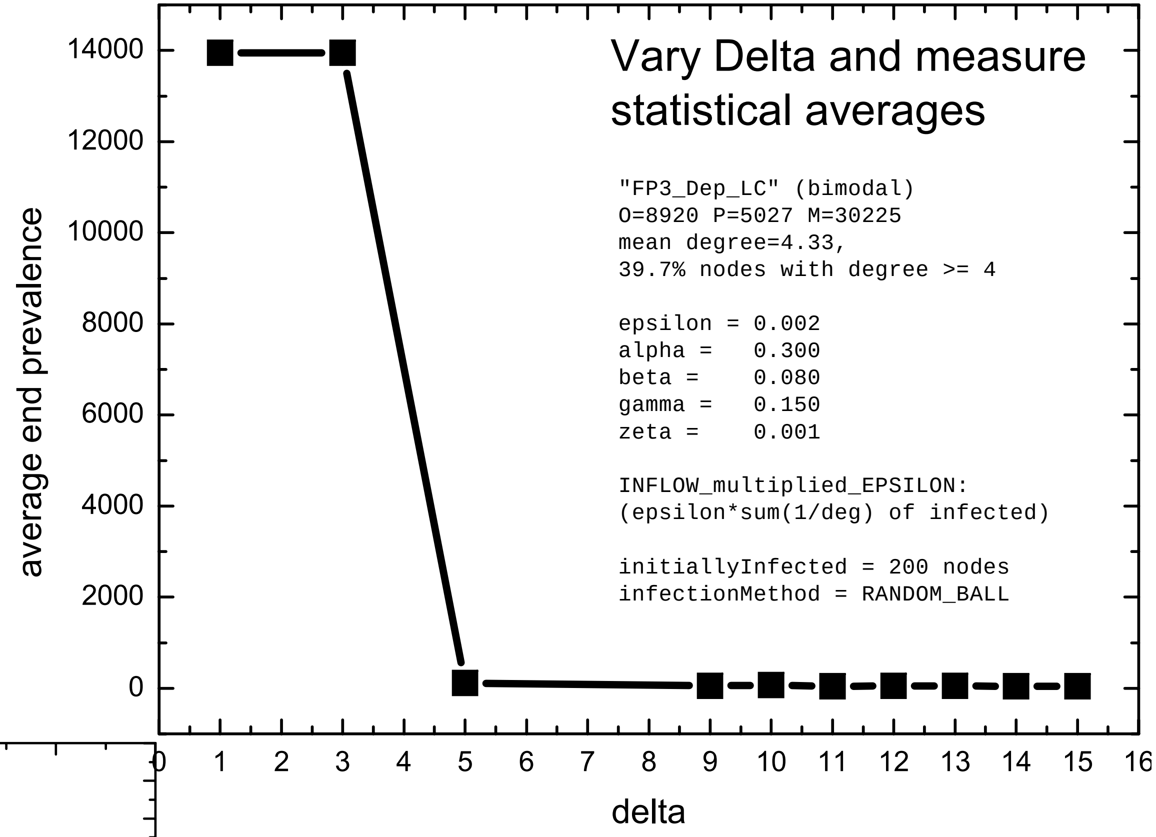


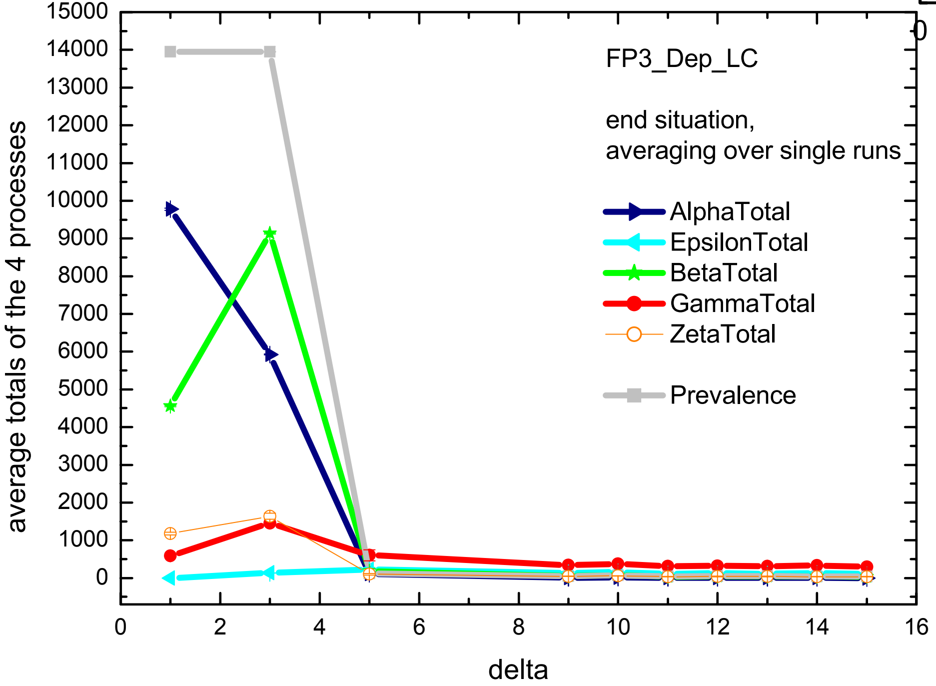



# FP3

Variation of delta

Many runs,
averages of **end** results

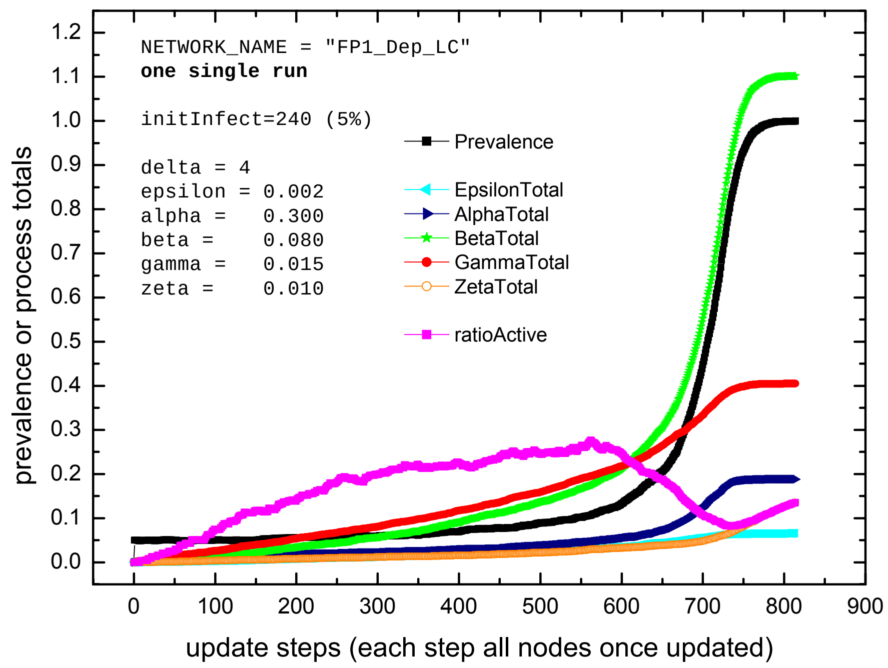


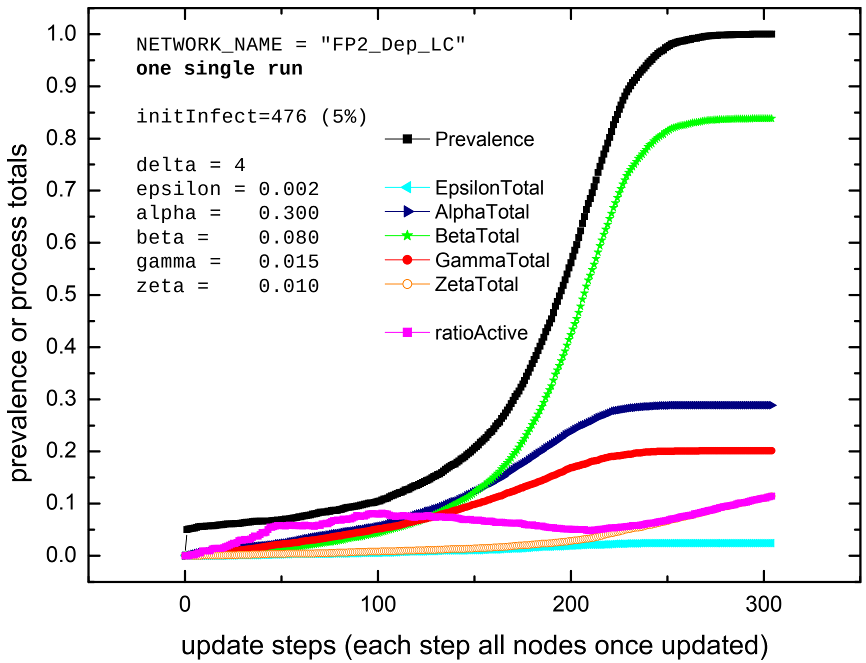



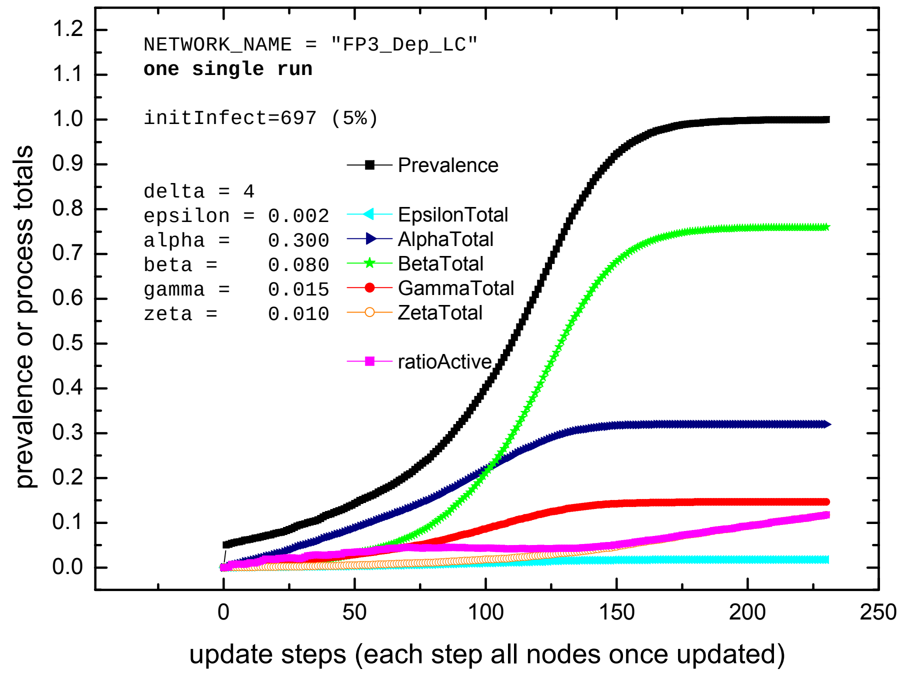
# Testweise nicht 200 Anfangsinfizierte, sondern

## 5% Anfangsinfektion

240 Knoten, 476 Knoten, 697 Knoten
je nach Gesamtsystemgröße



# Please…

If you read, use or cite this presentation, please notify me:

Networks (at) AndreasKrueger (dot) de

Thank you.