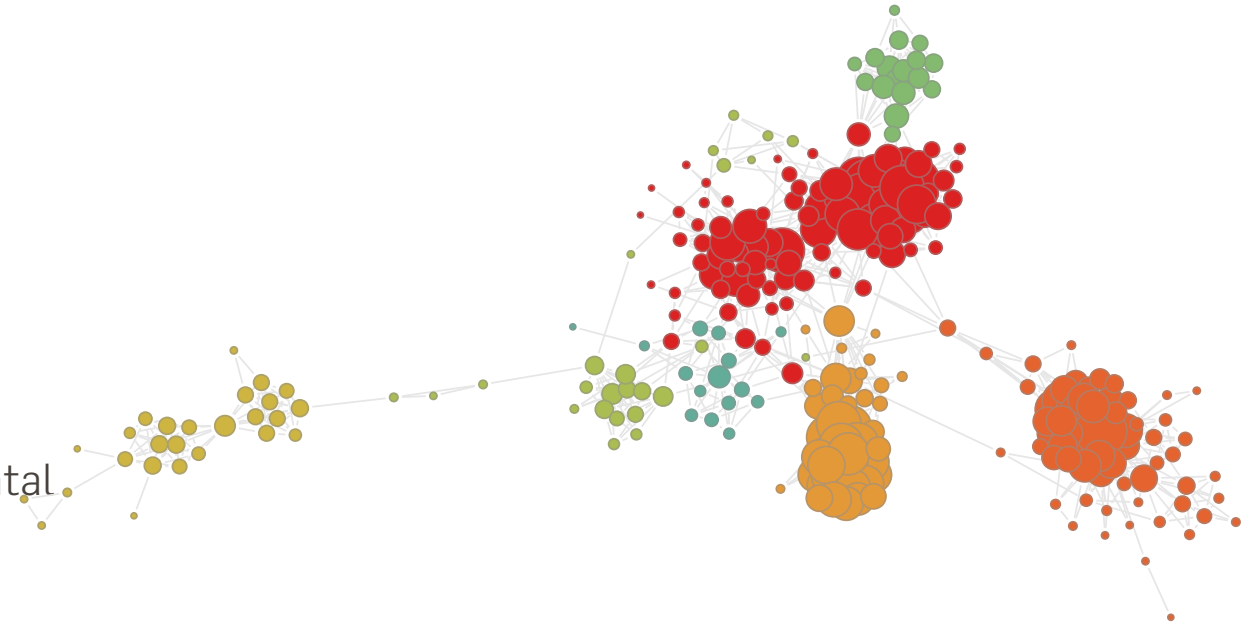
UNIVERSITAT DE BARCELONA

iSocial
Decentralized online
social networks

THE TOPOLOGICAL EVOLUTION OF ONLINE SOCIAL NETWORKS

February 4, 2014
iSocial Meeting
Stockholm



Kaj Kolja Kleineberg
Marián Boguñá

Departament de Física Fonamental
Universitat de Barcelona

AGENDA

1. Motivation & introduction
2. Empirical study
3. Basic model
4. Extended model
5. Summary & discussion

MOTIVATION & INTRODUCTION

DECENTRALIZED ONLINE SOCIAL NETWORKS

» The vision of a **decentralized** Ubiquitous Social Networking Layer in the **absence of central management and control.**«[1]

[1] http://isocial-itn.eu

DECENTRALIZATION IN MONETARY SECTOR

THE BENEFIT FROM MODELS

- Overlay Infrastructure for Decentralized Online Social Networking Services
- Data storage & distribution
- Security, privacy & trust
- **Modelling** and Simulation

THE BENEFIT FROM MODELS

- Overlay Infrastructure for Decentralized Online Social Networking Services
- Data storage & distribution
- Security, privacy & trust
- **Modelling** and Simulation

- **Quantitative** models can identify **mechanisms** for:
  - + Evolution of **topological structure** of the network
  - + **Competition** between different online social networks

THE BENEFIT FROM MODELS

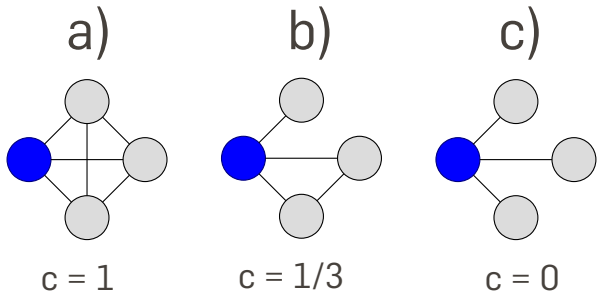
- Overlay Infrastructure for Decentralized Online Social Networking Services
- Data storage & distribution
- Security, privacy & trust
- **Modelling** and Simulation

- **Quantitative** models can identify **mechanisms** for:
  - + Evolution of **topological structure** of the network
  - + **Competition** between different online social networks

Identified **mechanisms** are helpful for designers of **next generation OSN service architectures.**

BRIEF INTRODUCTION

- **Local clustering coefficent:** c_i number of connected pairs of neighbors divided by total number of pairs of neighbors

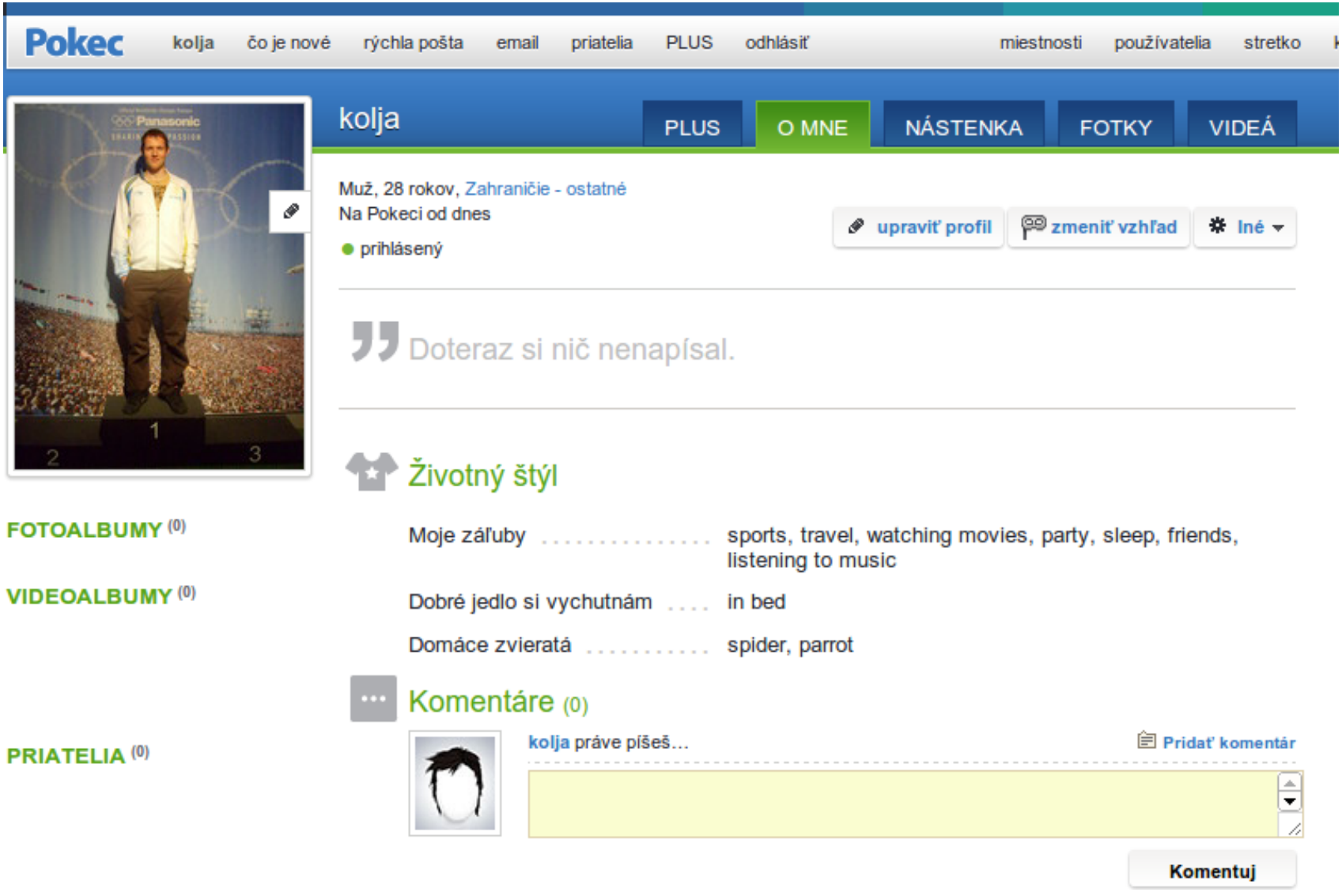


- **Mean local clustering:** average of c_i
- **Giant connected component (GCC):** Size of the largest subgraph of a network in which each pair of nodes is connected by a path.
- **Average shortest path length:** Average shortest distance (d_{ij}) between each pair of nodes in the GCC.
- **Network diameter:** The network diameter (within the GCC) is defined as the maximum of d_{ij}

$$l = \max_{ij} \{d_{ij}\} \ .$$

EMPIRICAL STUDY

CASE STUDY: »POKEC«



POKEC SLOVAKIAN ONLINE SOCIAL NETWORK

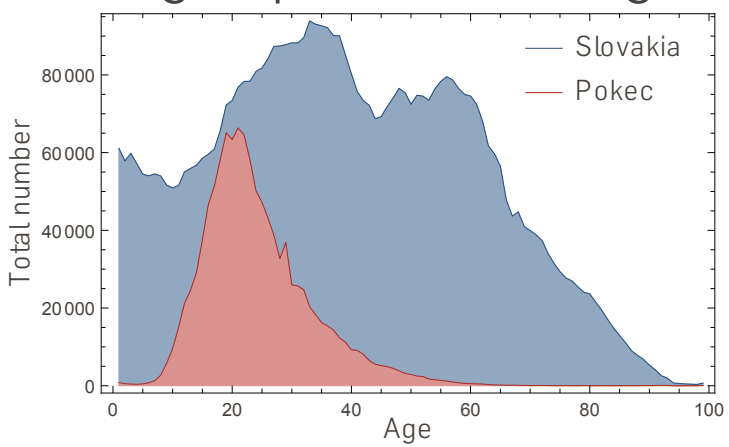
Slovakia

Population
5,410,836

Pokec empirical network

Users
1,632,803

Age-specific coverage



POKEC SLOVAKIAN ONLINE SOCIAL NETWORK

Slovakia

Population
5,410,836

Pokec empirical network

Users
1,632,803

- Isolated network
- Very high population coverage
- Evolution 1999-2012
- 1.2 million nodes in bi-directional subgraph

POKEC SLOVAKIAN ONLINE SOCIAL NETWORK

Slovakia

Population
5,410,836

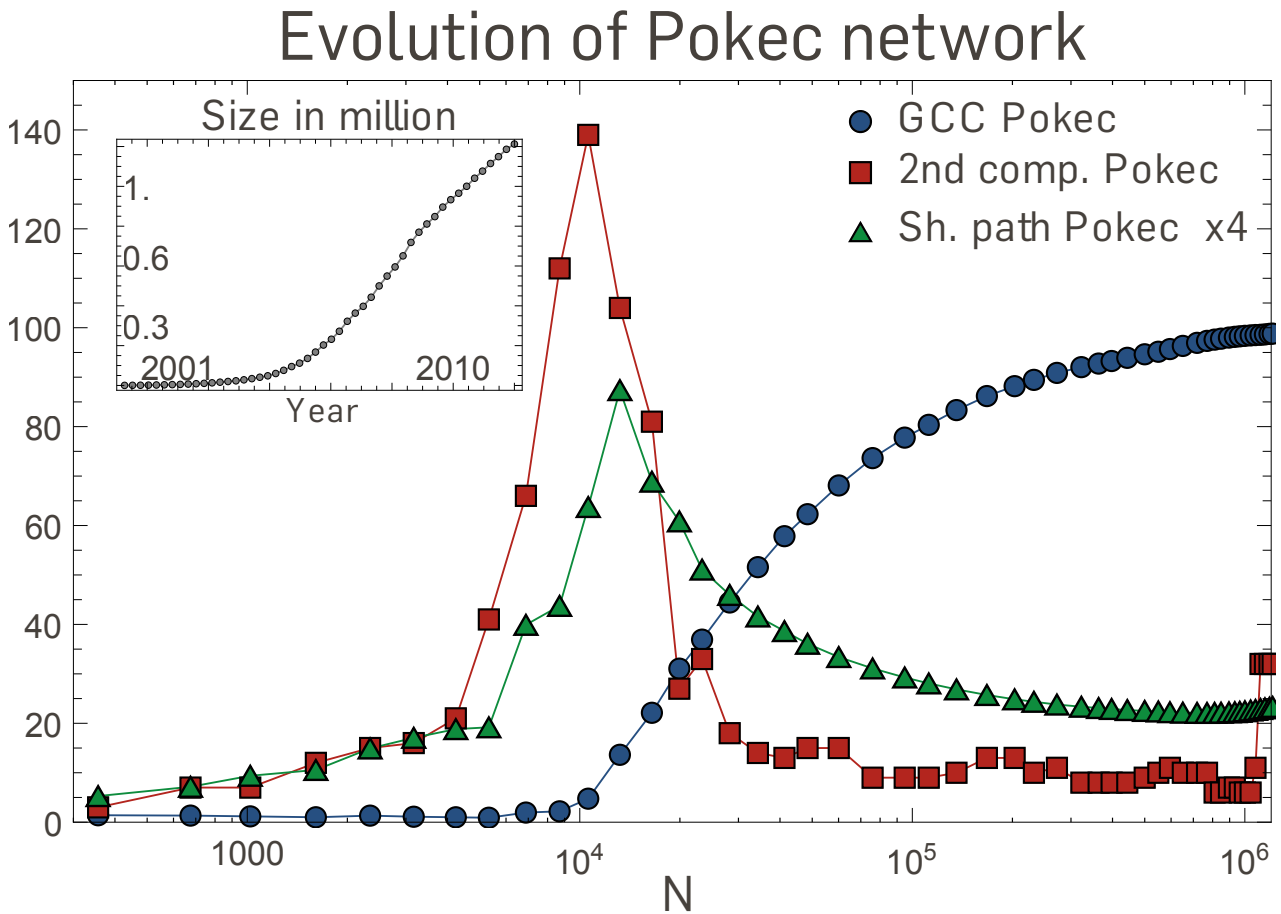
Pokec empirical network

Users
1,632,803

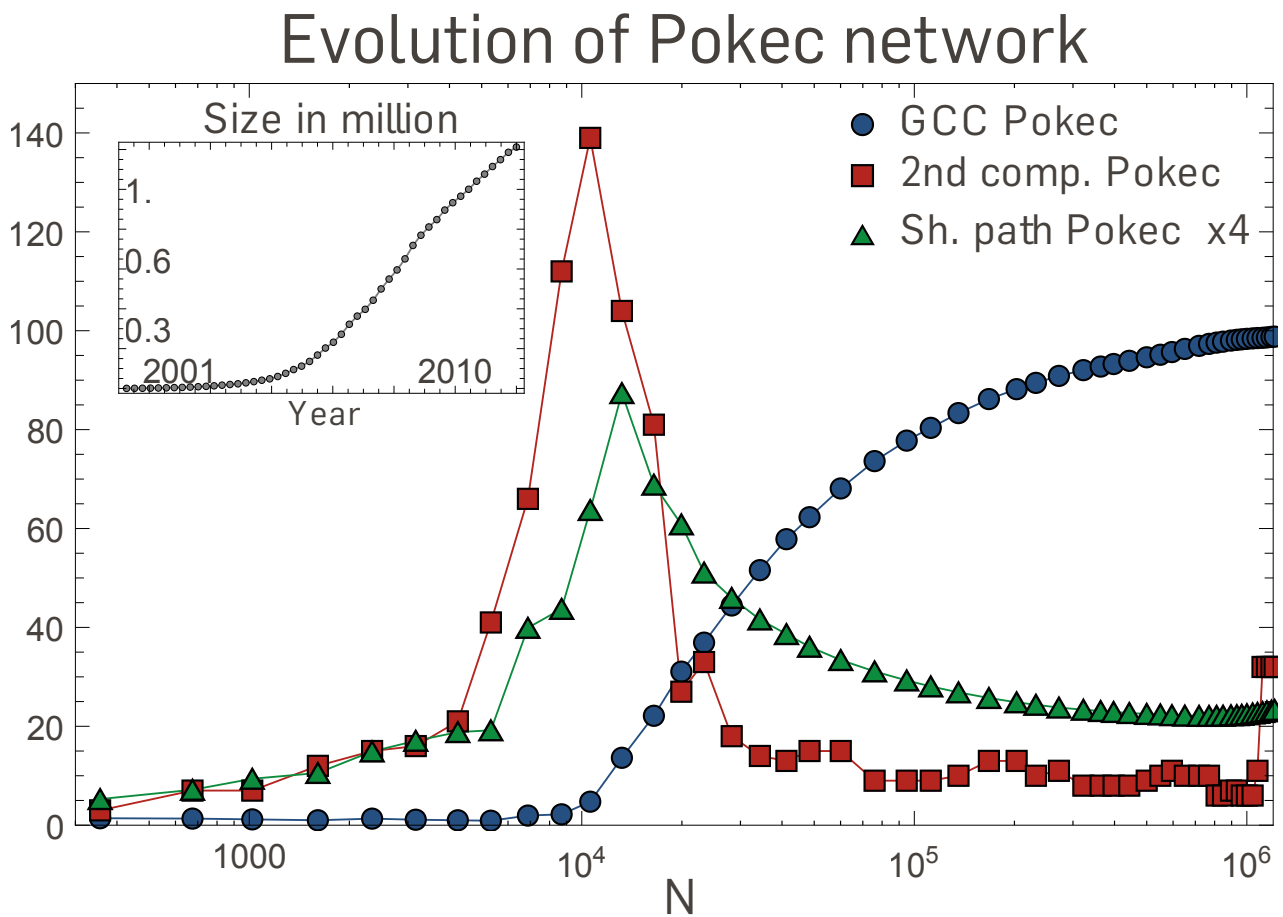
- Isolated network
- Very high population coverage
- Evolution 1999-2012
- 1.2 million nodes in bi-directional subgraph

Network is **isolated**, **active** & contains a **large fraction** of a priori susceptible individuals.

PHASE TRANSITION



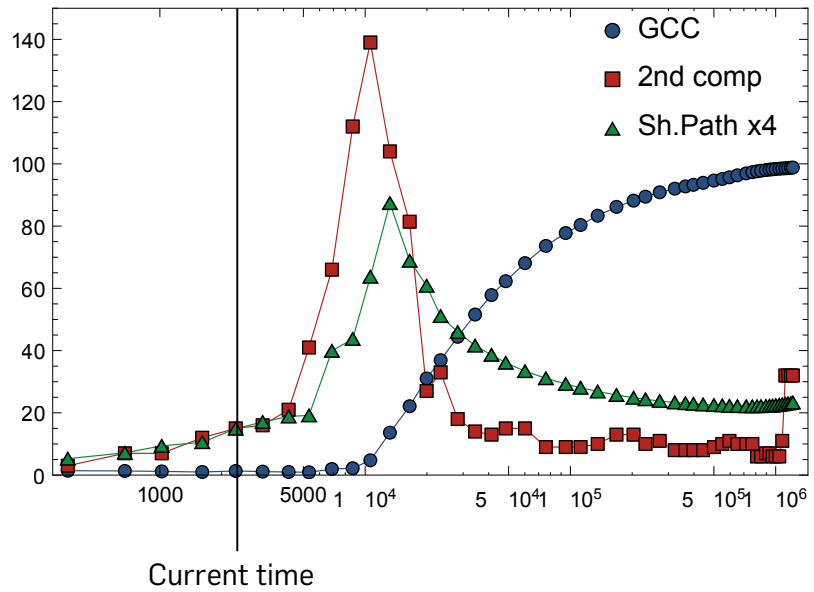
PHASE TRANSITION



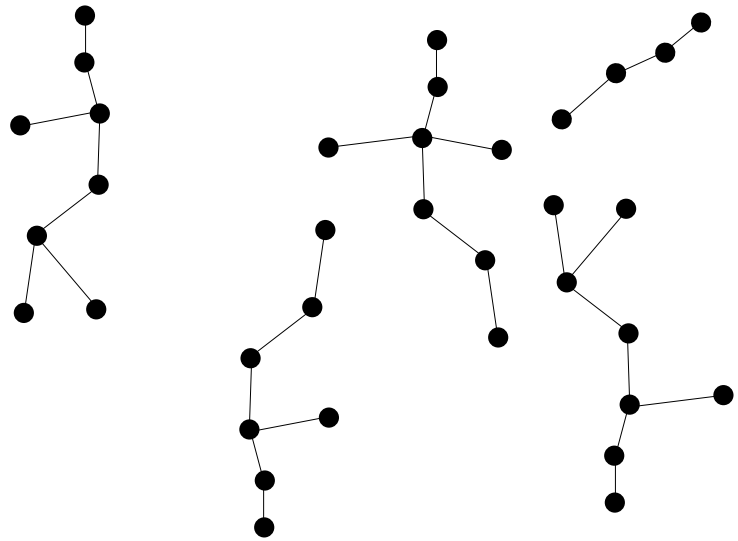
Network exhibits **dynamical phase transition** between **disconnected** and **connected** phase.

STRUCTURE

Connectivity

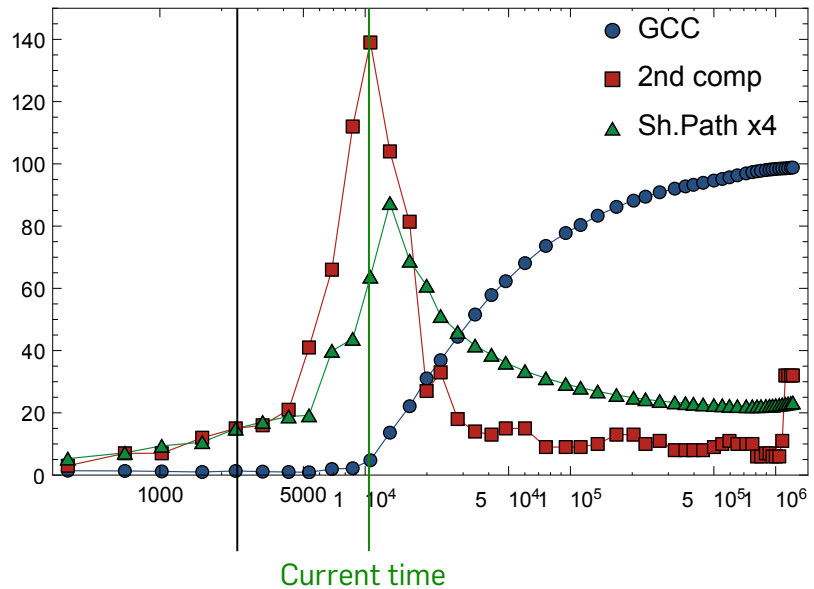


Structure

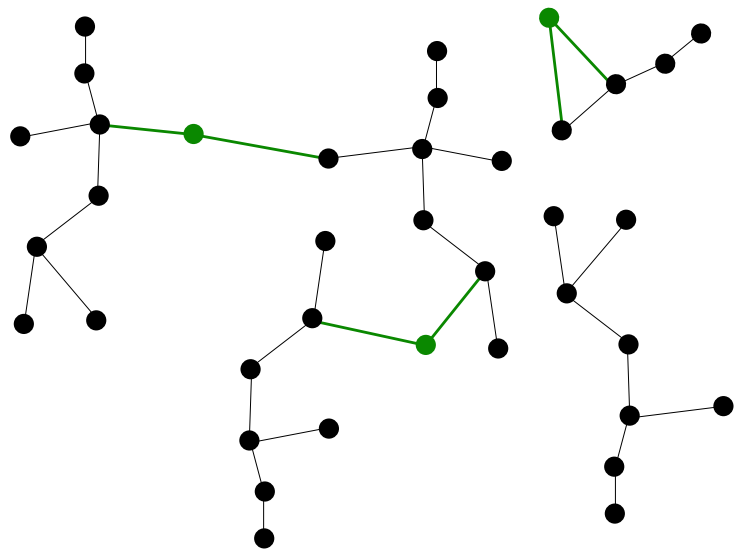


STRUCTURE

Connectivity

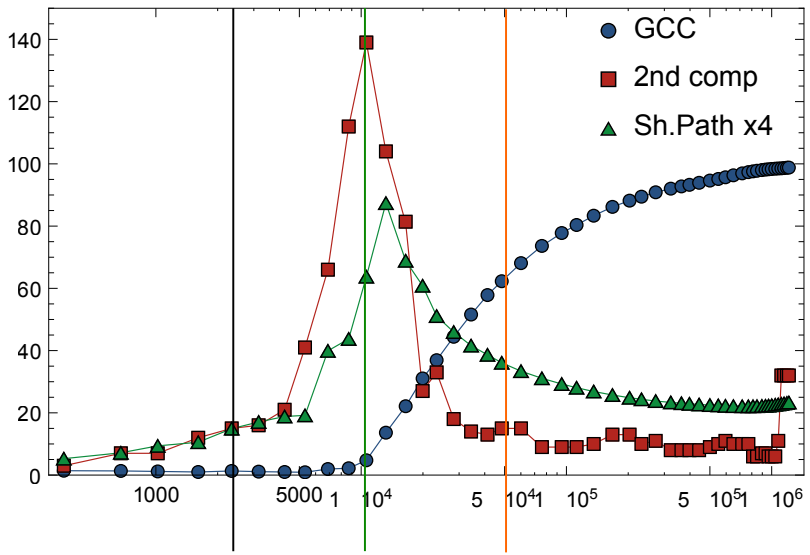


Structure

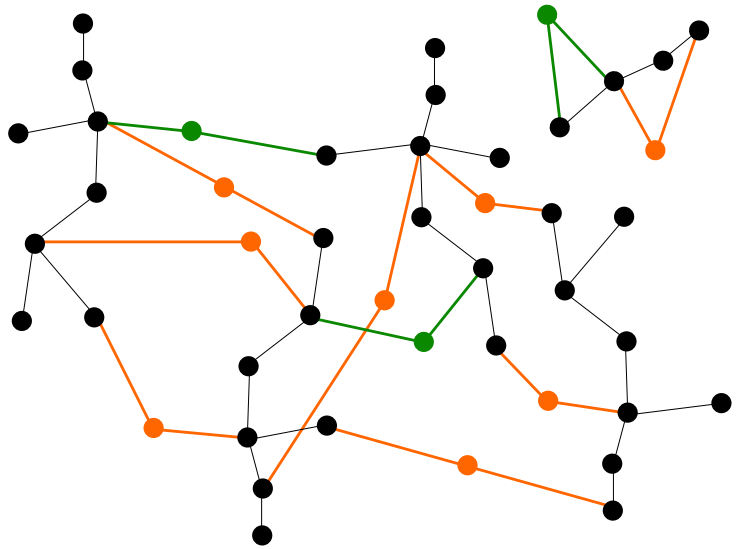


STRUCTURE

Connectivity

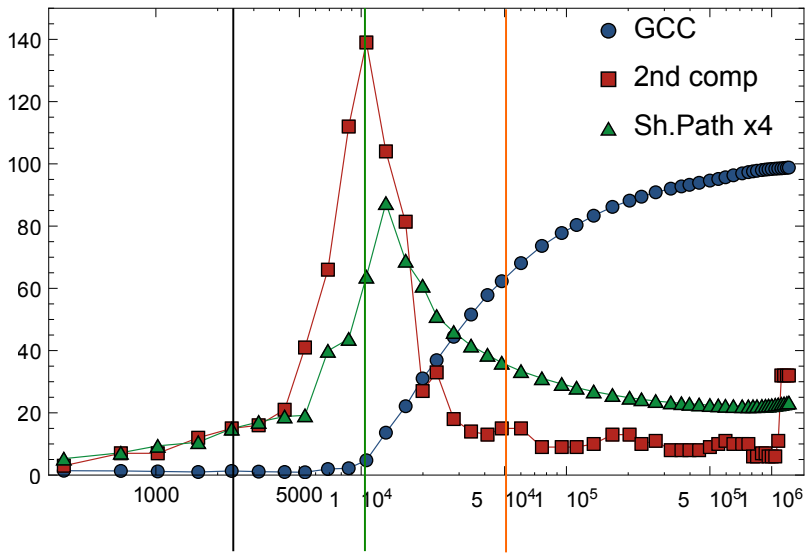


Structure

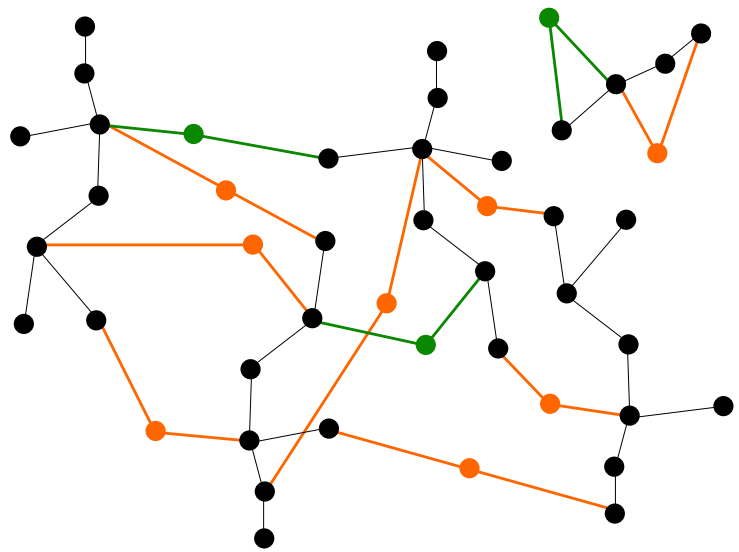


STRUCTURE

Connectivity



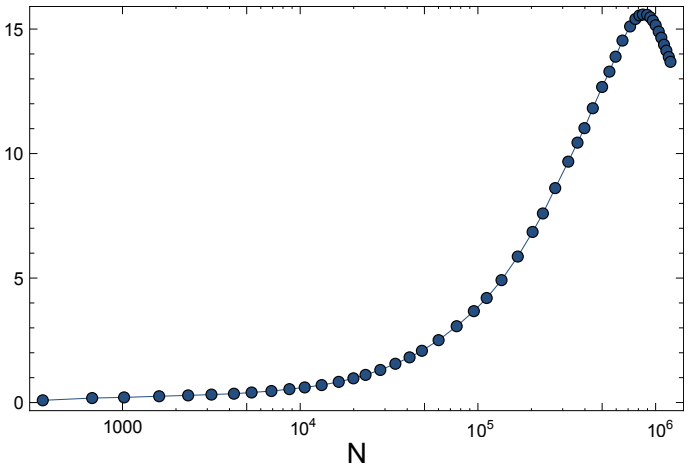
Structure



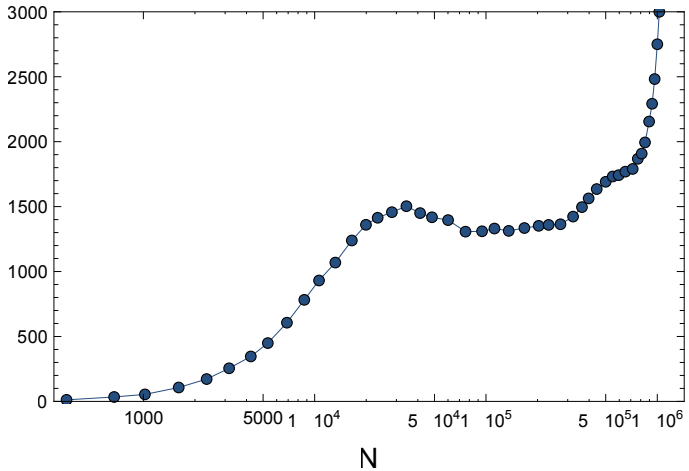
Multiple components grow to **chain-like** structures until they become **connected** and form **shortcuts**.

TOPOLOGICAL EVOLUTION OF POKEC

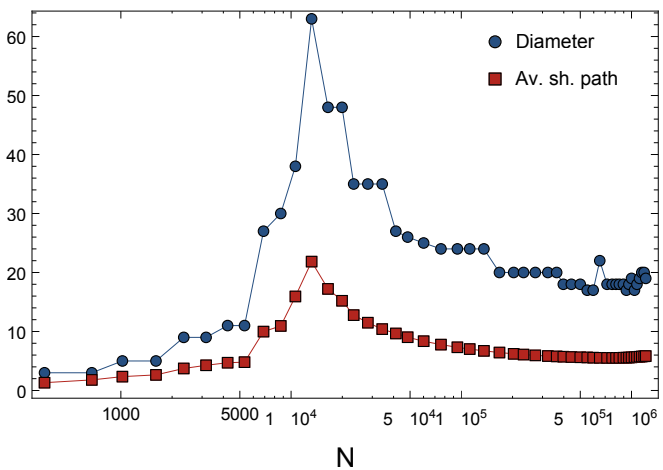
Growing **mean degree**



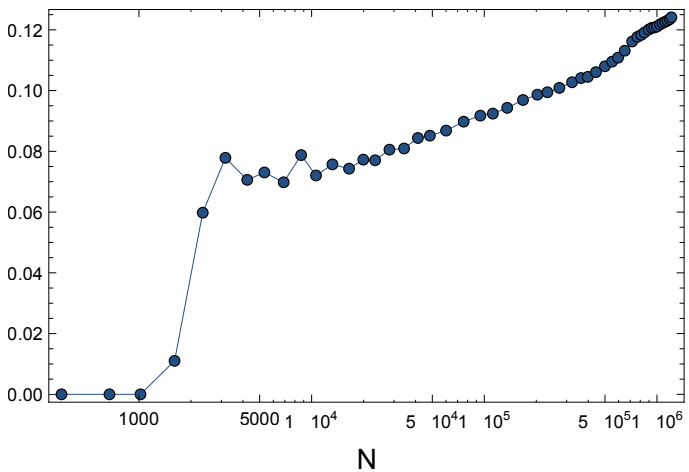
Number of **components**



Shrinking **pathlength**



Increasing **clustering**

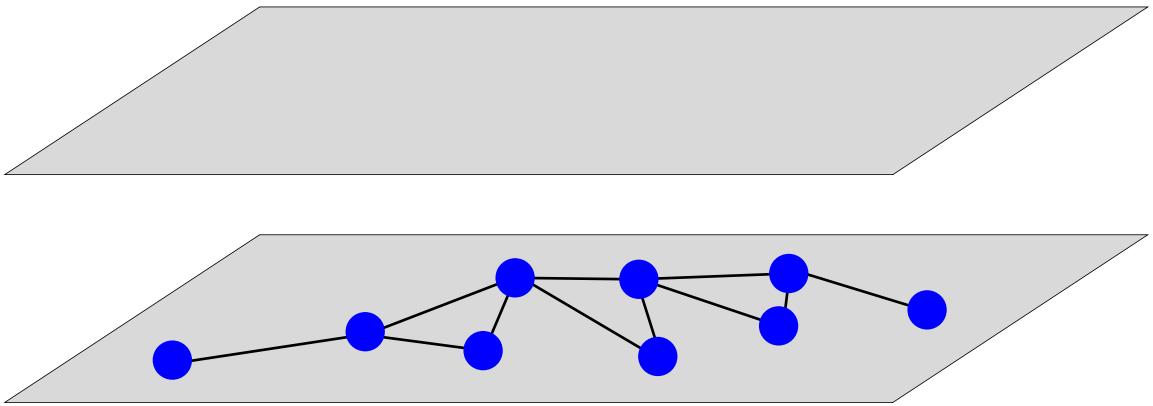


SUMMARY: EMPIRICAL STUDY

- Study evolution of **large, active & isolated** online social network
- Dynamical **phase transition** from disconnected to connected
- **Shrinking diameter** & **pathlength**
- **Increasing clustering** coefficient

BASIC MODEL

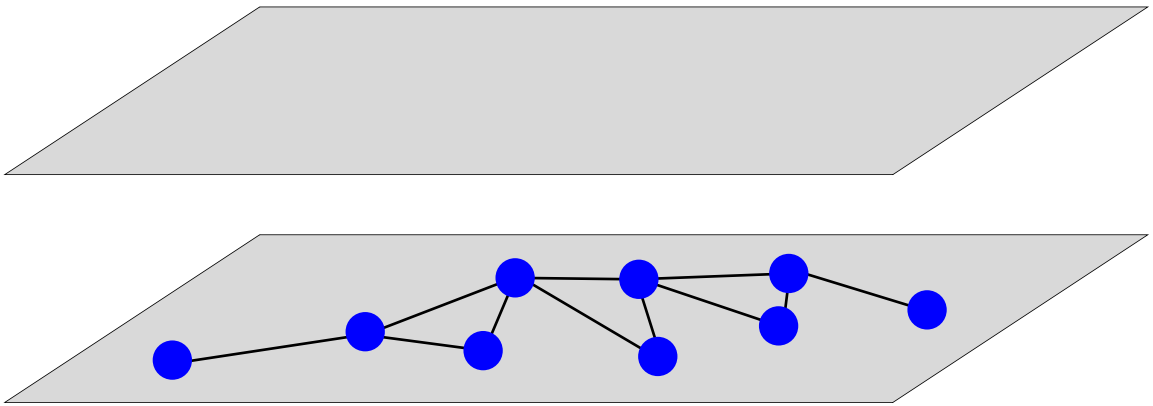
TWO-LAYER DESIGN: THE SOCIAL BACKBONE



Online social network layer

Traditional contact network layer

TWO-LAYER DESIGN: THE SOCIAL BACKBONE

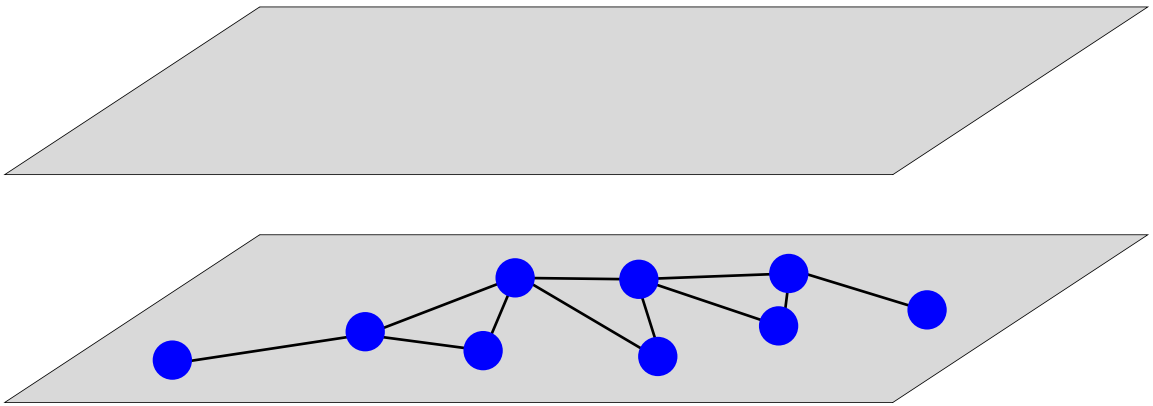


Online social network layer

Traditional contact network layer

- Underlying subgraph of **a priori susceptible** nodes from **aggregation** of all contact networks except OSN

TWO-LAYER DESIGN: THE SOCIAL BACKBONE

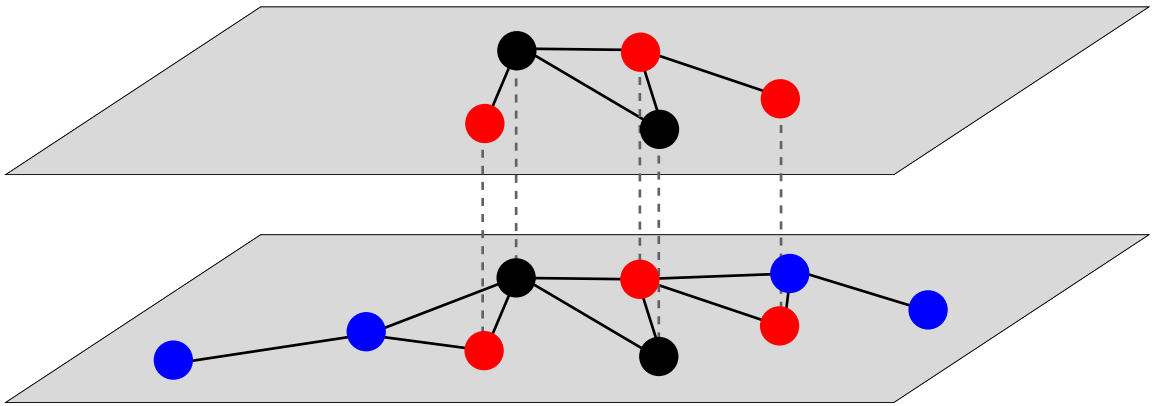


Online social network layer

Traditional contact network layer

- Underlying subgraph of **a priori susceptible** nodes from **aggregation** of all contact networks except OSN
- Start with **empty online social network** layer

TWO-LAYER DESIGN: THE SOCIAL BACKBONE

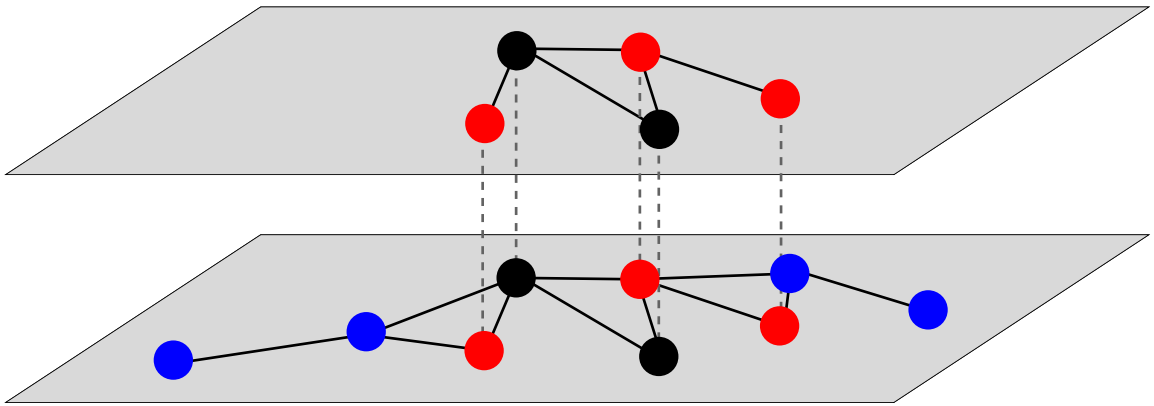


Online social network layer

Traditional contact network layer

- Underlying subgraph of **a priori susceptible** nodes from **aggregation** of all contact networks except OSN
- Start with **empty online social network** layer
- Nodes in OSN make **same connections** as in underlying

TWO-LAYER DESIGN: THE SOCIAL BACKBONE



Online social network layer

Traditional contact network layer

- Underlying subgraph of **a priori susceptible** nodes from **aggregation** of all contact networks except OSN
- Start with **empty online social network** layer
- Nodes in OSN make **same connections** as in underlying

Question: What do we use as **underlying contact network** layer?

UNDERLYING NETWORK: THE VALIDATION CONCEPT

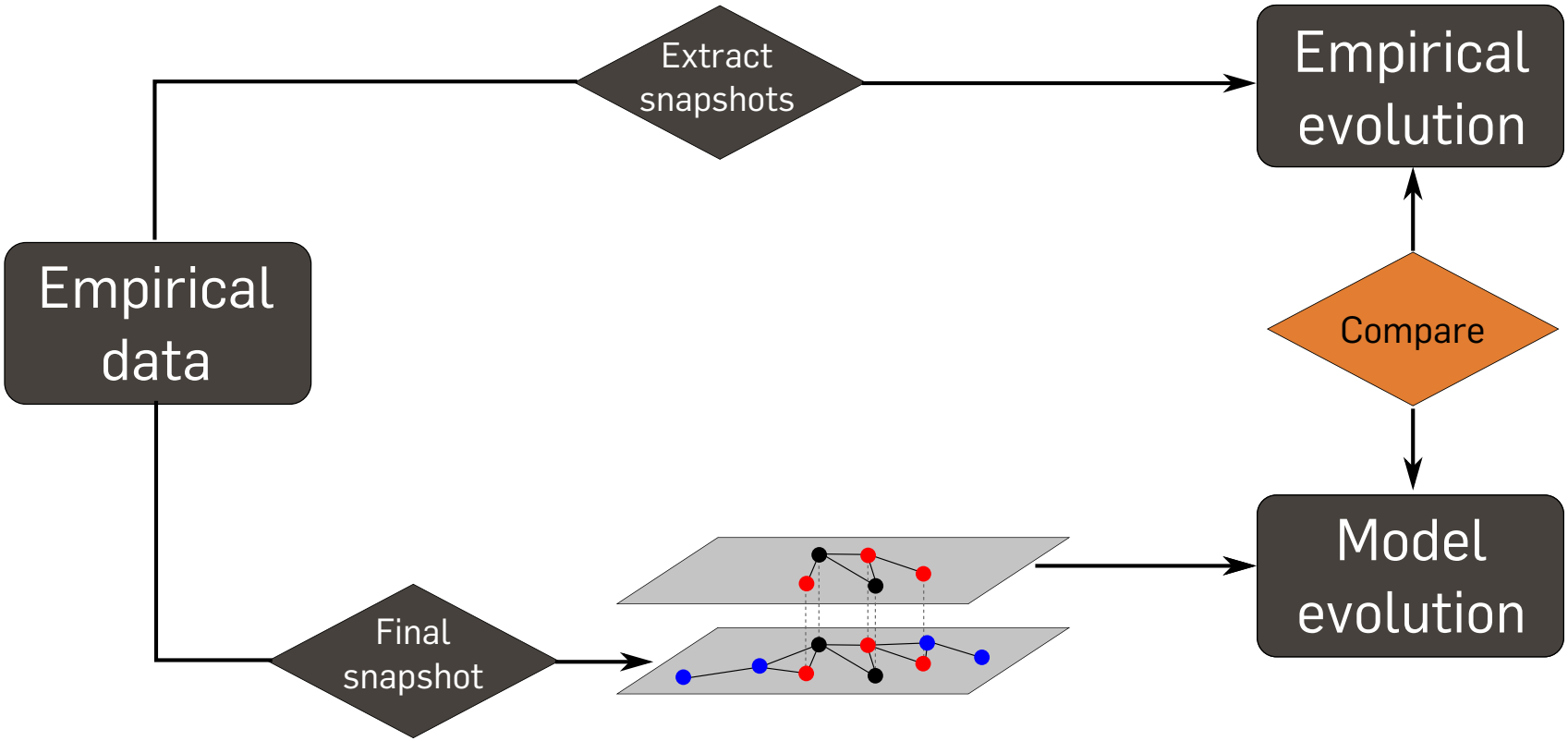
Synthetic network		Empiric network	
For instance S1-model		Final snapshot of empiric network	
Control of degrees, clustering and sizes	✓	Control of degrees, clustering and sizes	✗
Model validation	✗	Model validation	✓

UNDERLYING NETWORK: THE VALIDATION CONCEPT

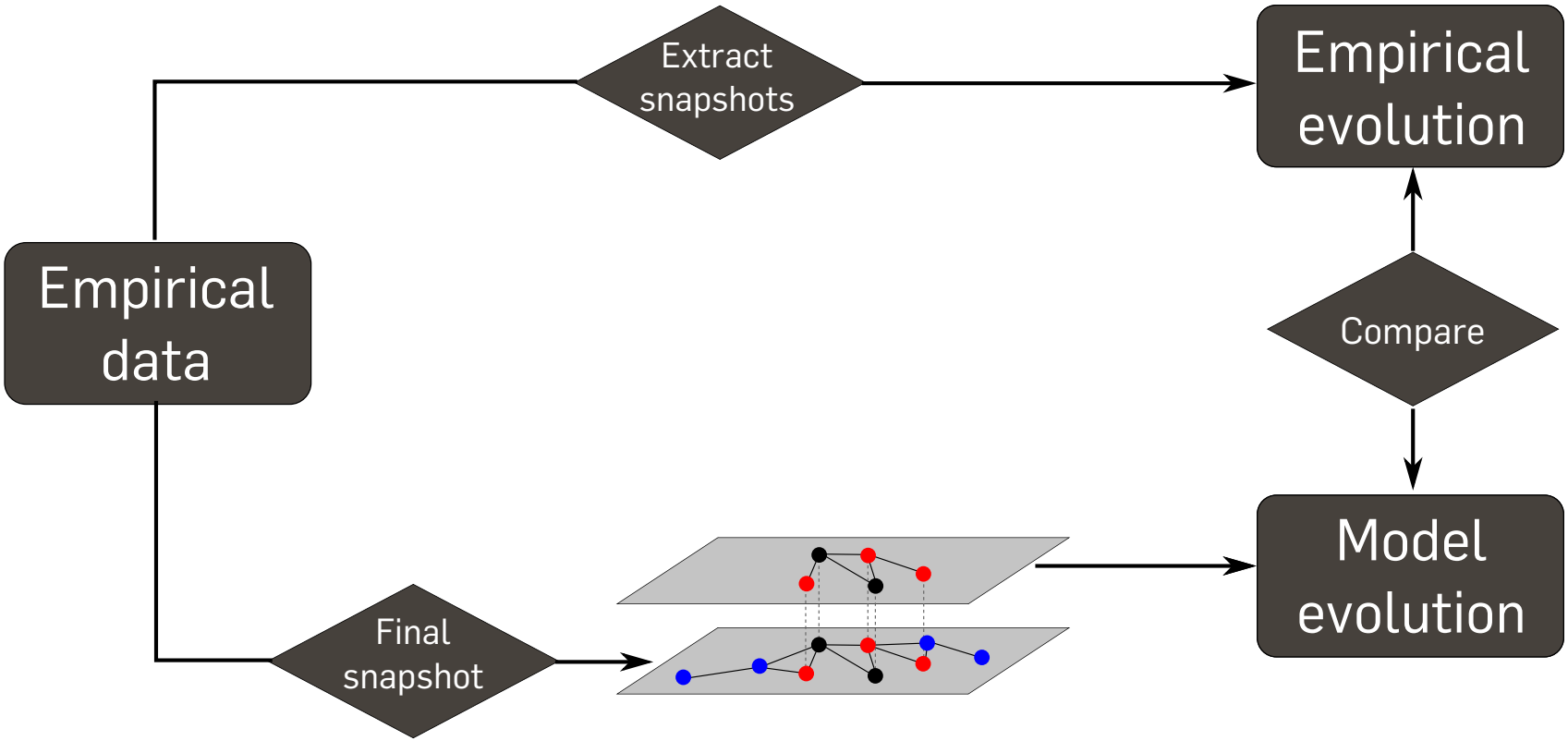
Synthetic network		Empiric network	
For instance S1-model		Final snapshot of empiric network	
Control of degrees, clustering and sizes	✓	Control of degrees, clustering and sizes	✗
Model validation	✗	Model validation	✓

Only **empiric network** as underlying layer allows **rigorous validation** of the model.

UNDERLYING NETWORK: THE VALIDATION CONCEPT



UNDERLYING NETWORK: THE VALIDATION CONCEPT



Question: Which are the **mechanisms** that determine subscriptions and allow to **reproduce the evolution**?

VIRALITY AND MASS MEDIA INFLUENCE

Mass media

»Sources of information that reach and **influence large numbers** of people.«

Model: Random subscriptions (affects everybody equally).

Impact: Creates many disconnected components.

Virality

»The tendency of information to be circulated rapidly and widely **from one to another.**«

Model: Like the spread of a virus in SIS model.

Impact: Increases the size of connected components.

VIRALITY AND MASS MEDIA INFLUENCE

Mass media

»Sources of information that reach and **influence large numbers** of people.«

Model: Random subscriptions (affects everybody equally).

Impact: Creates many disconnected components.

Virality

»The tendency of information to be circulated rapidly and widely **from one to another.**«

Model: Like the spread of a virus in SIS model.

Impact: Increases the size of connected components.

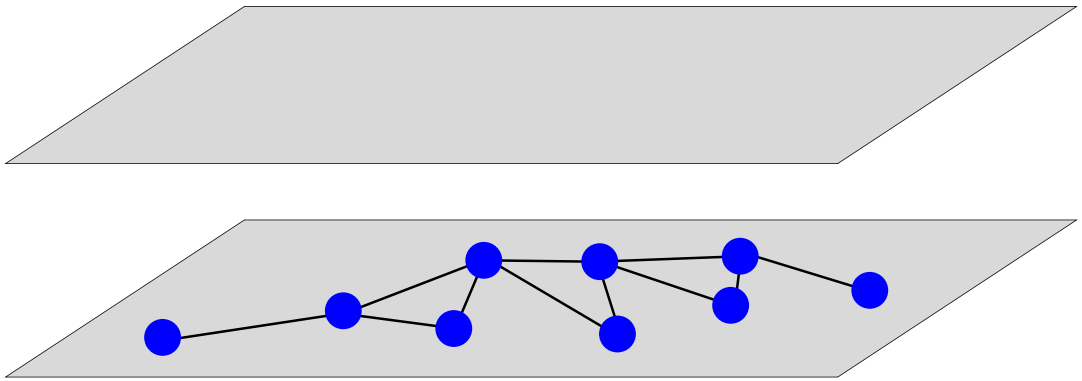
Virality and **mass media** effect play **complementary roles** for the evolution of the network topology.

MODEL MECHANISM

S: Susceptible
only offline
can be activated

P: Passive
online and offline
do not activate

A: Active
online and offline
do activate



Online social network layer

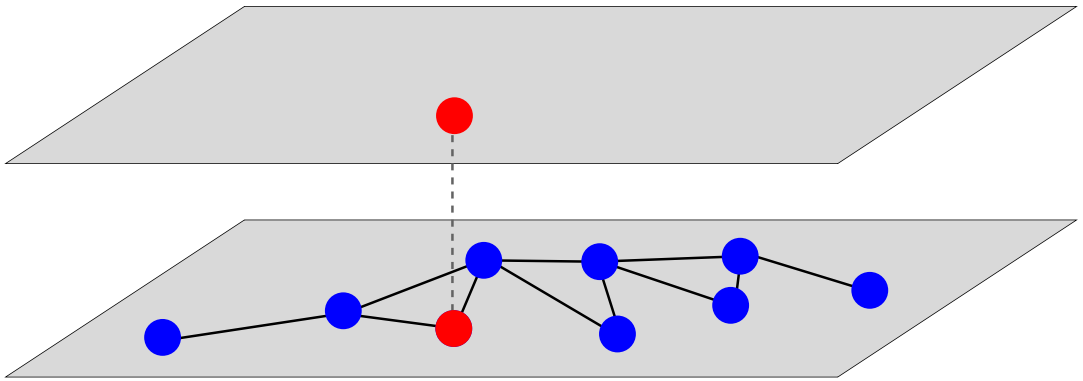
Underlying contact network layer

MODEL MECHANISM

S: Susceptible
only offline
can be activated

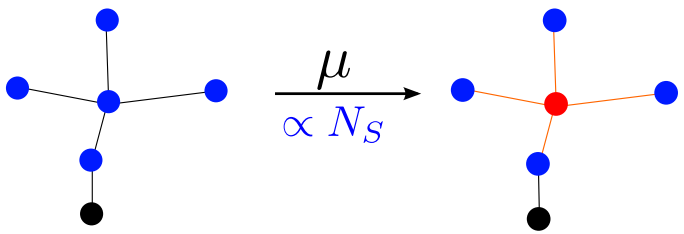
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate



Online social network layer

Underlying contact network layer



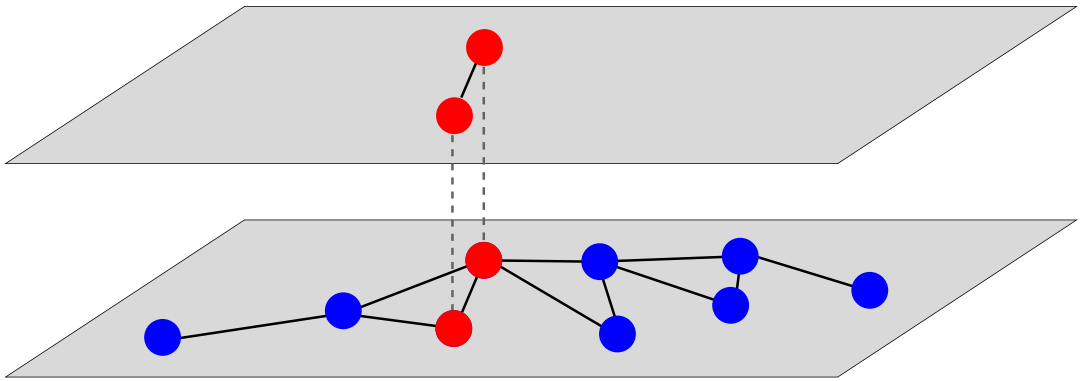
Mass media effect

MODEL MECHANISM

S: Susceptible
only offline
can be activated

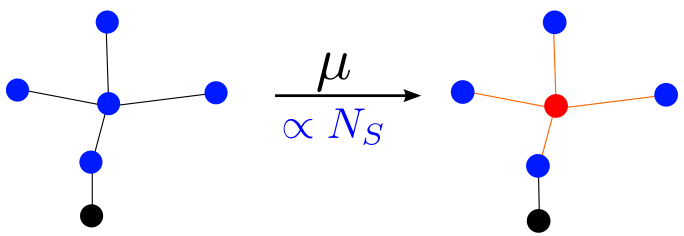
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate

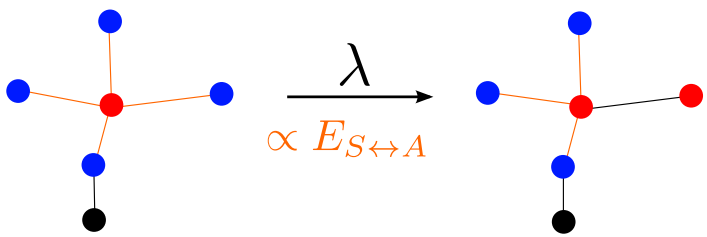


Online social network layer

Underlying contact network layer



Mass media effect



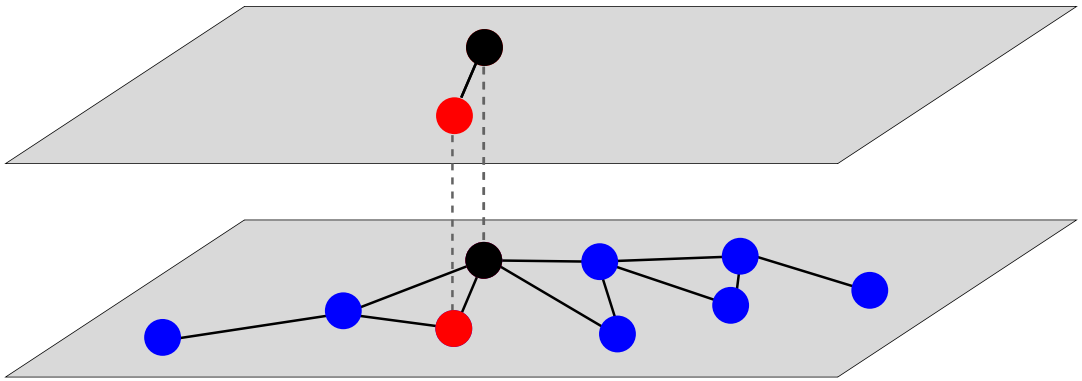
Viral activation

MODEL MECHANISM

S: Susceptible
only offline
can be activated

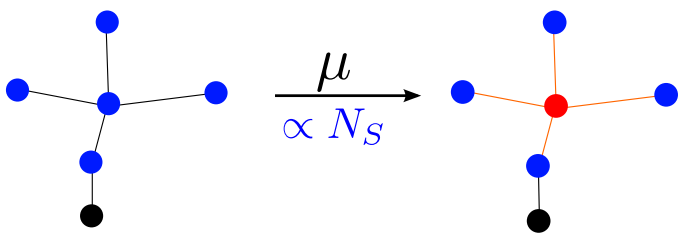
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate

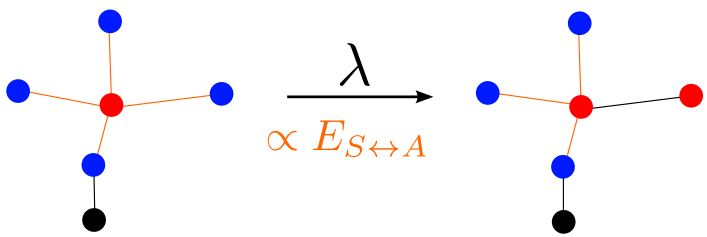


Online social network layer

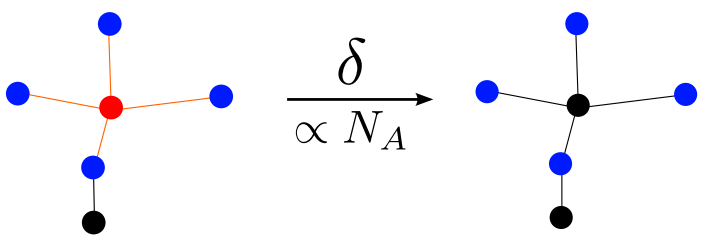
Underlying contact network layer



Mass media effect



Viral activation



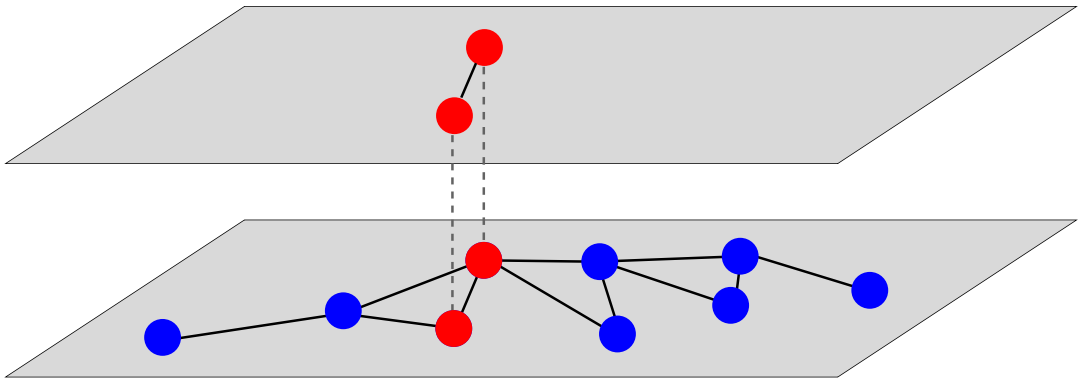
Deactivation

MODEL MECHANISM

S: Susceptible
only offline
can be activated

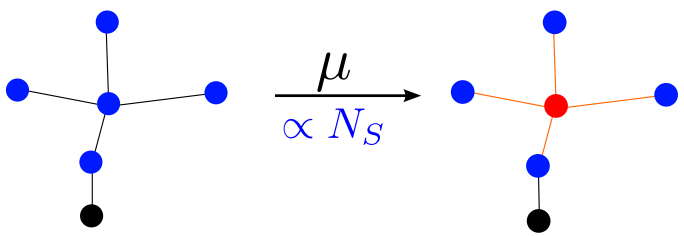
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate

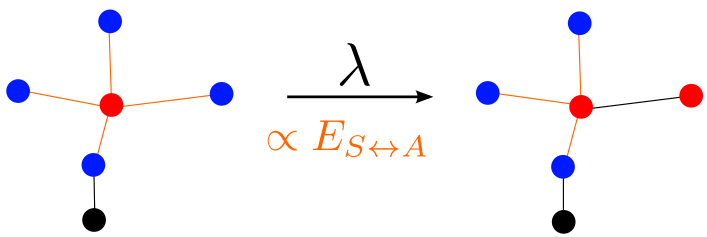


Online social network layer

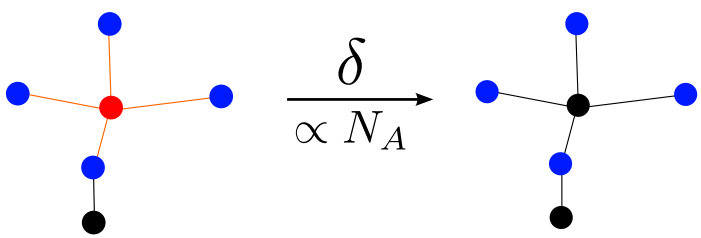
Underlying contact network layer



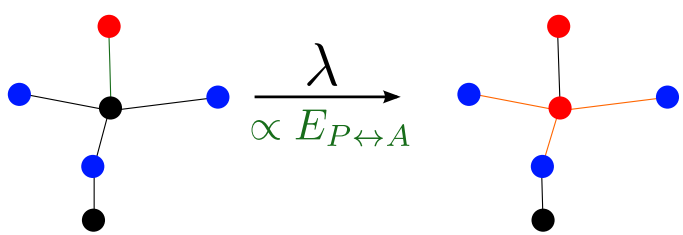
Mass media effect



Viral activation



Deactivation



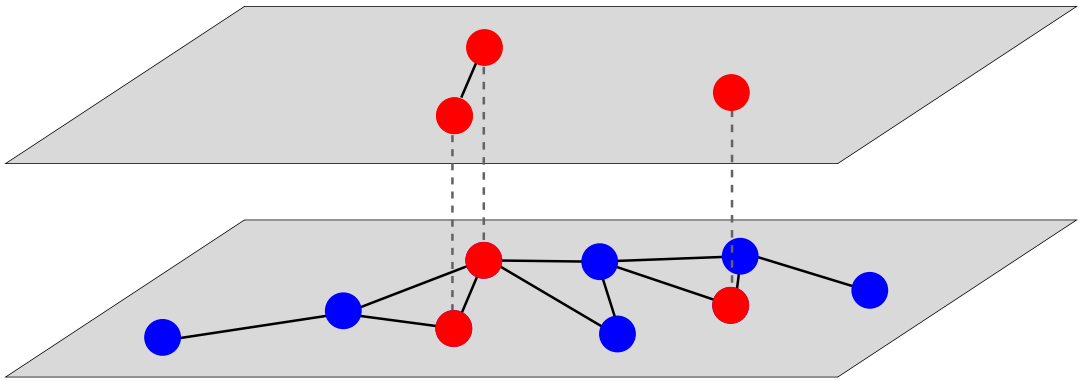
Viral reactivation

MODEL MECHANISM

S: Susceptible
only offline
can be activated

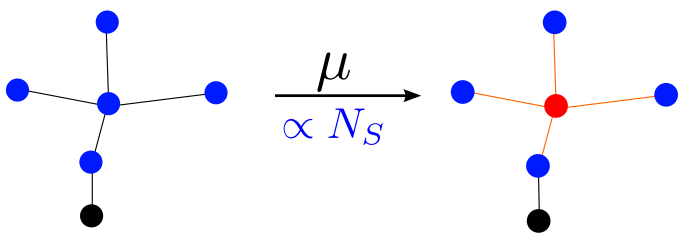
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate

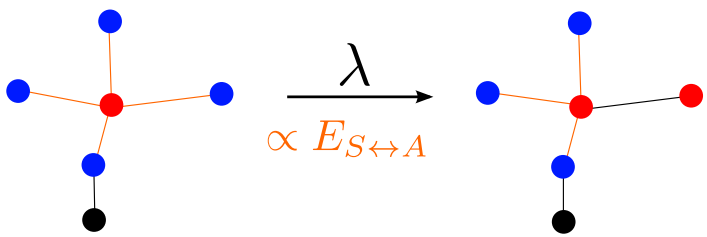


Online social network layer

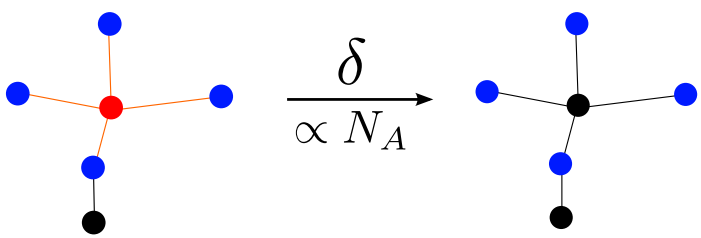
Underlying contact network layer



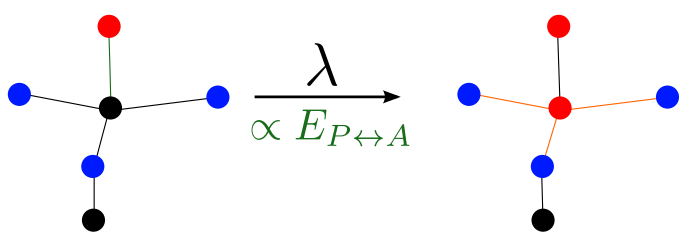
Mass media effect



Viral activation



Deactivation



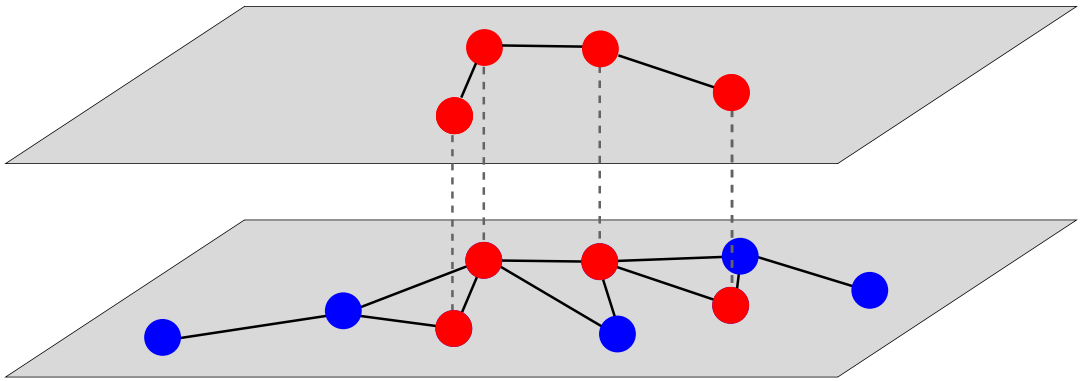
Viral reactivation

MODEL MECHANISM

S: Susceptible
only offline
can be activated

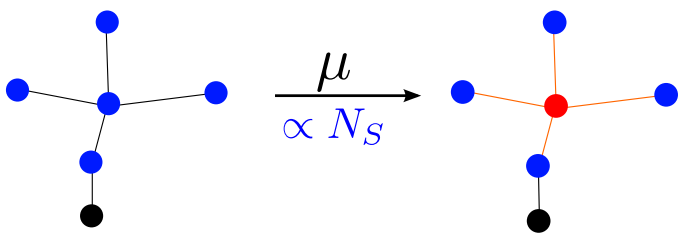
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate

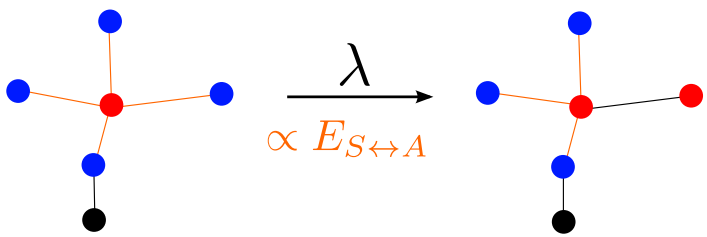


Online social network layer

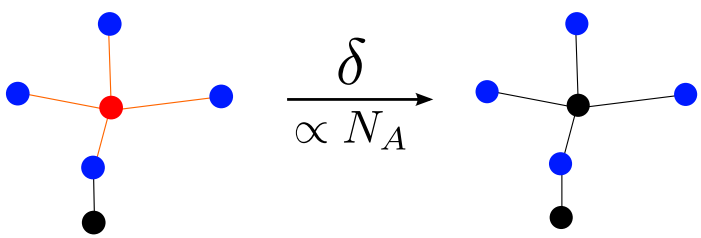
Underlying contact network layer



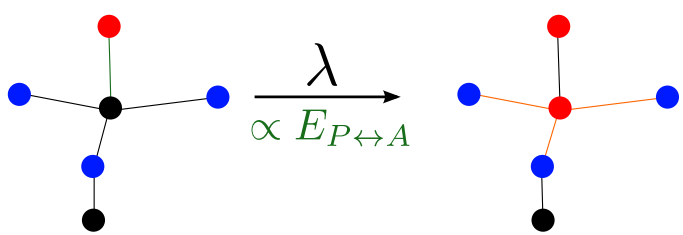
Mass media effect



Viral activation



Deactivation



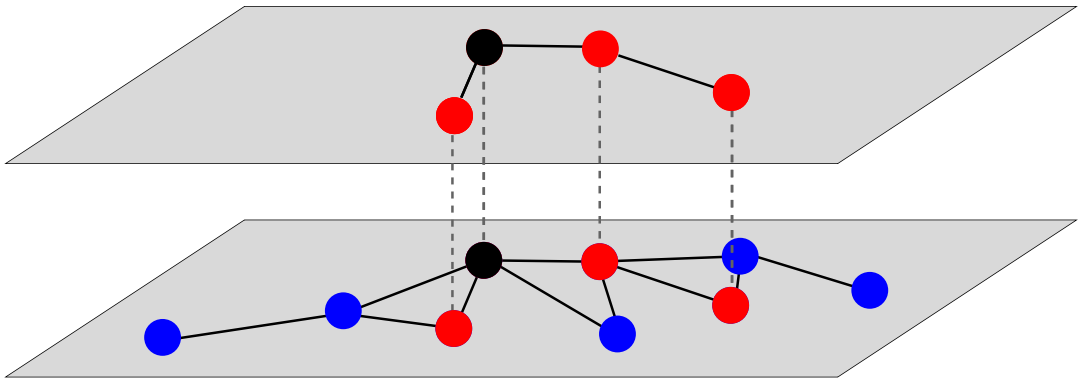
Viral reactivation

MODEL MECHANISM

S: Susceptible
only offline
can be activated

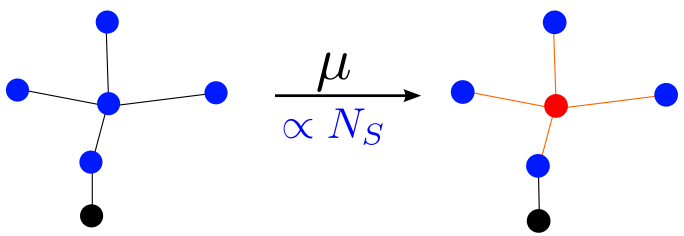
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate

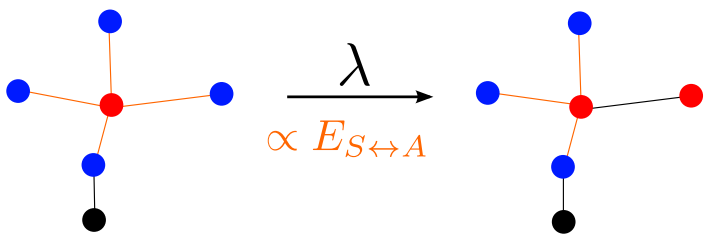


Online social network layer

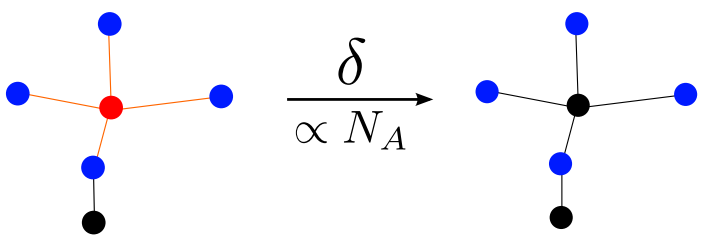
Underlying contact network layer



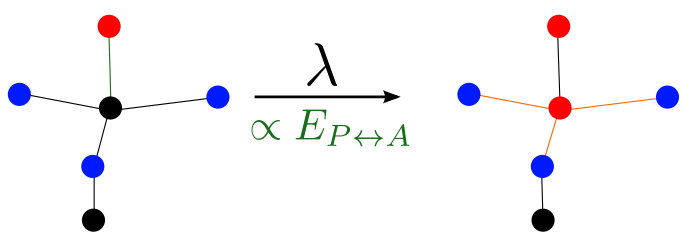
Mass media effect



Viral activation



Deactivation



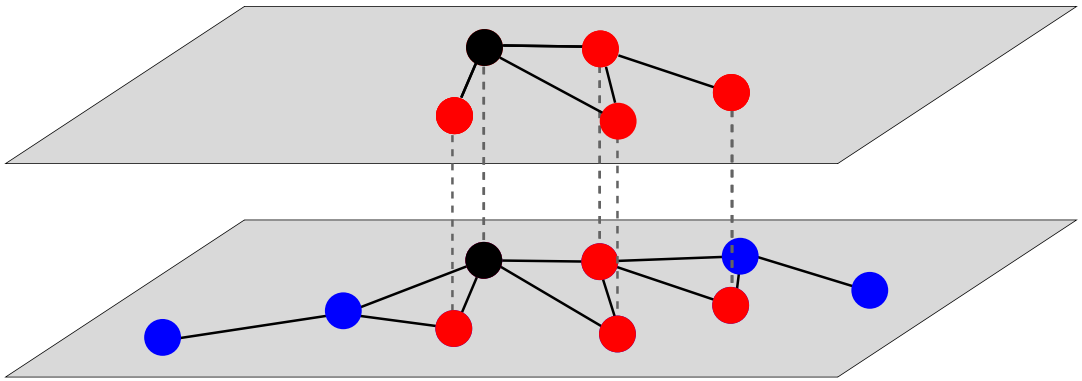
Viral reactivation

MODEL MECHANISM

S: Susceptible
only offline
can be activated

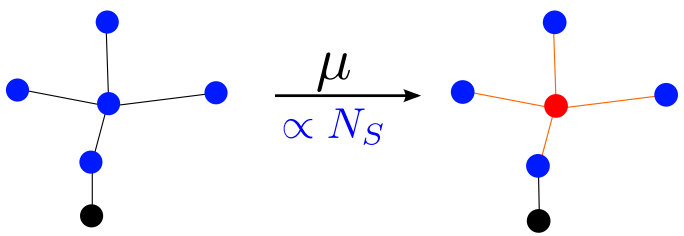
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate

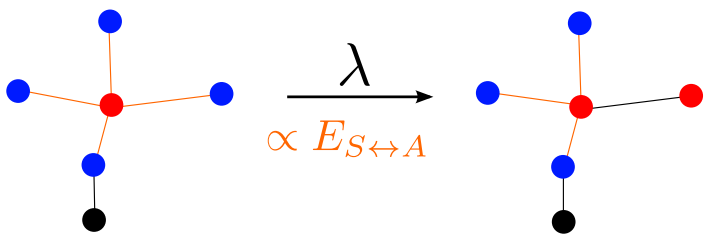


Online social network layer

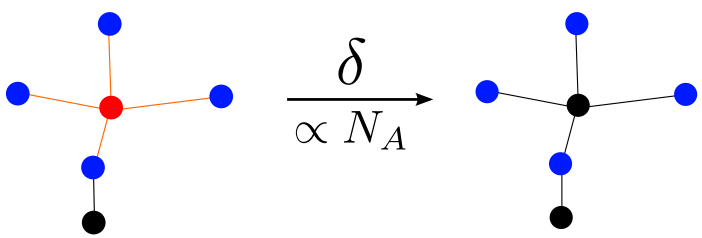
Underlying contact network layer



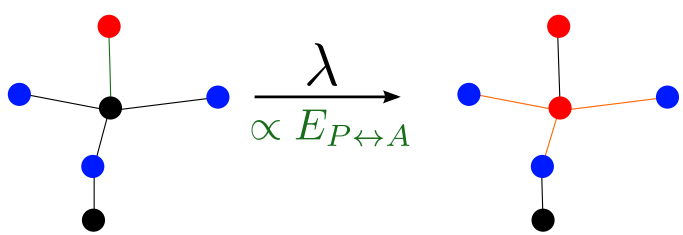
Mass media effect



Viral activation



Deactivation



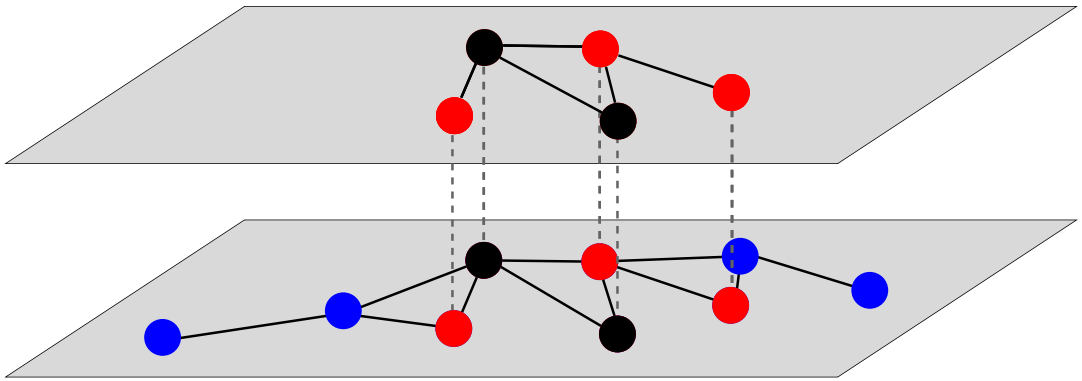
Viral reactivation

MODEL MECHANISM

S: Susceptible
only offline
can be activated

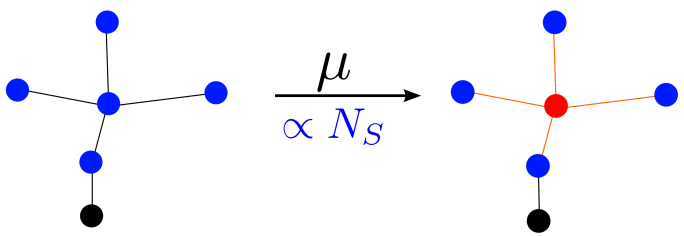
P: Passive
online and offline
do not activate

A: Active
online and offline
do activate

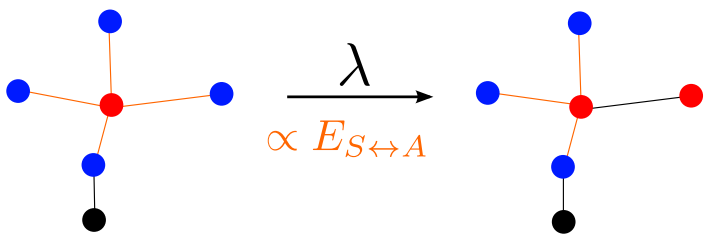


Online social network layer

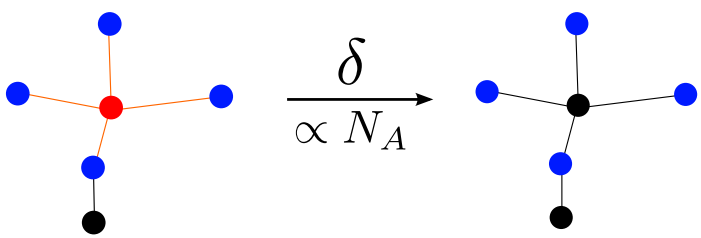
Underlying contact network layer



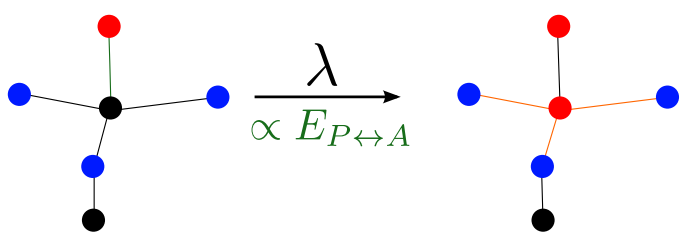
Mass media effect



Viral activation



Deactivation



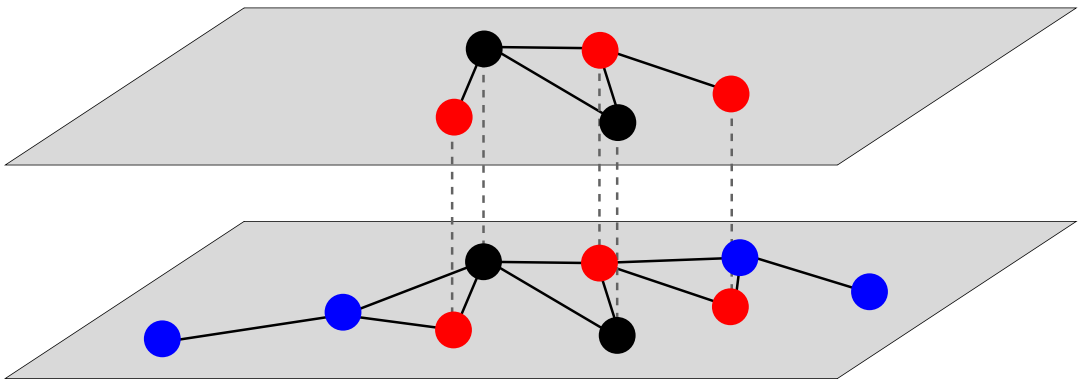
Viral reactivation

MODEL MECHANISM

S: Susceptible
only offline
can be activated

P: Passive
online and offline
do not activate

A: Active
online and offline
do activate



Online social network layer

Underlying contact network layer

Mass media effect

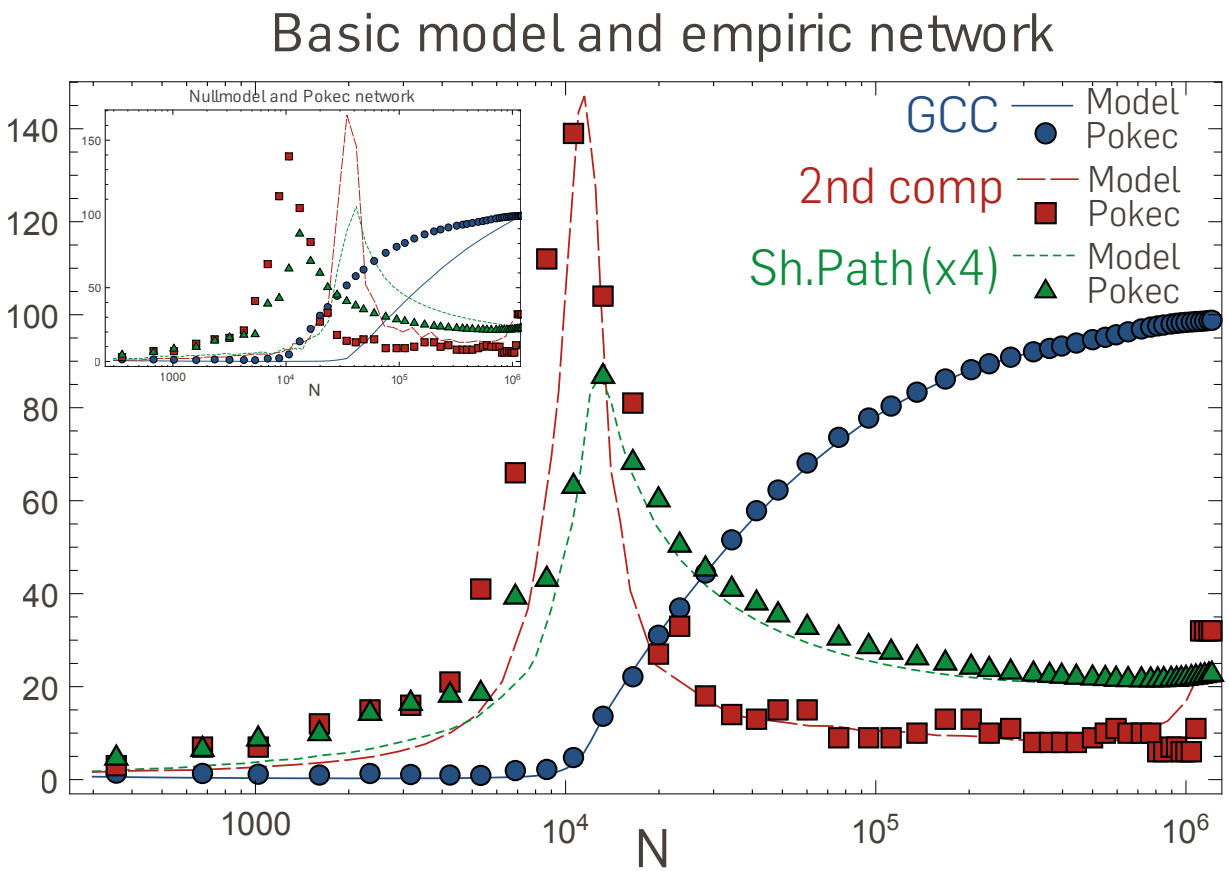
Deactivation

Viral activation

Viral reactivation

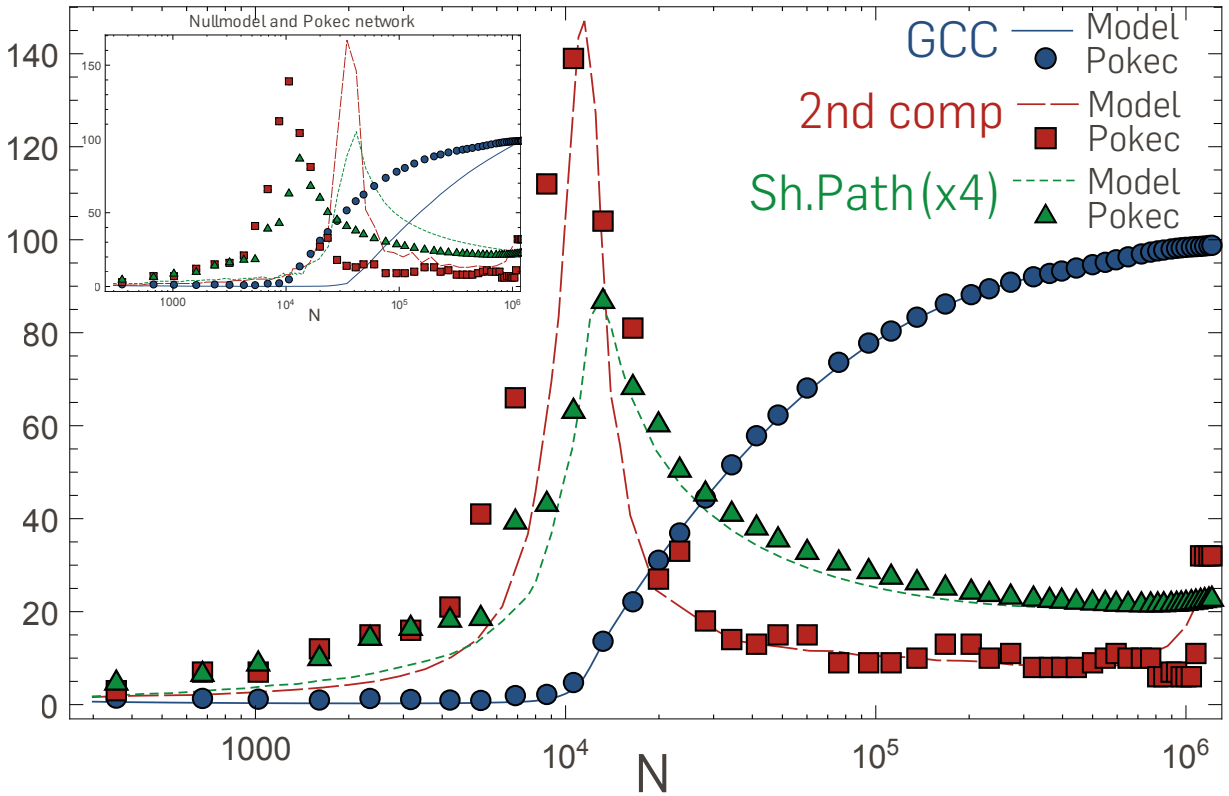
Can we **reproduce the empirical evolution** with **viral** and **mass media** dynamics?

RESULTS FROM VIRALITY MASS MEDIA MODEL



RESULTS FROM VIRALITY MASS MEDIA MODEL

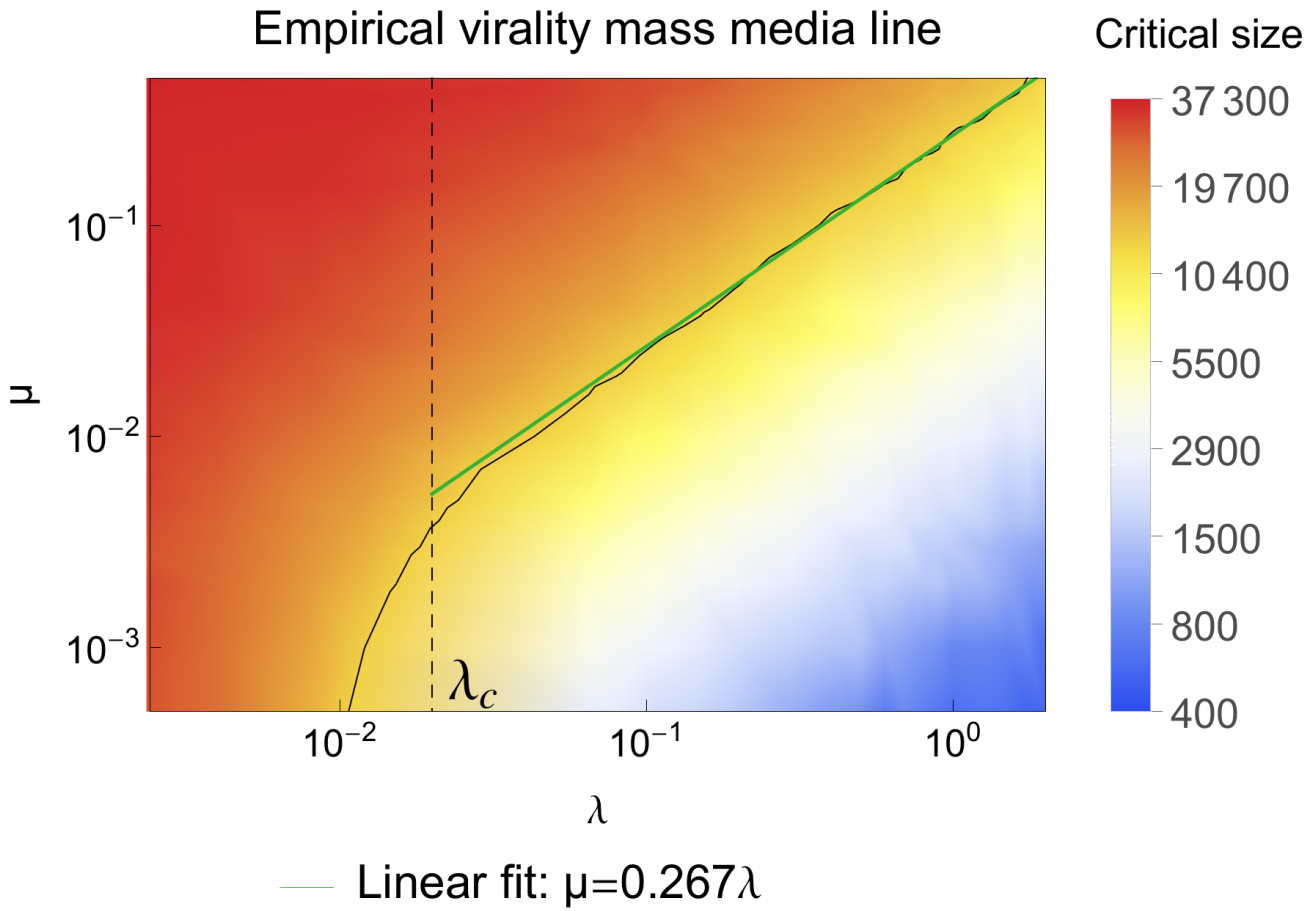
Basic model and empiric network



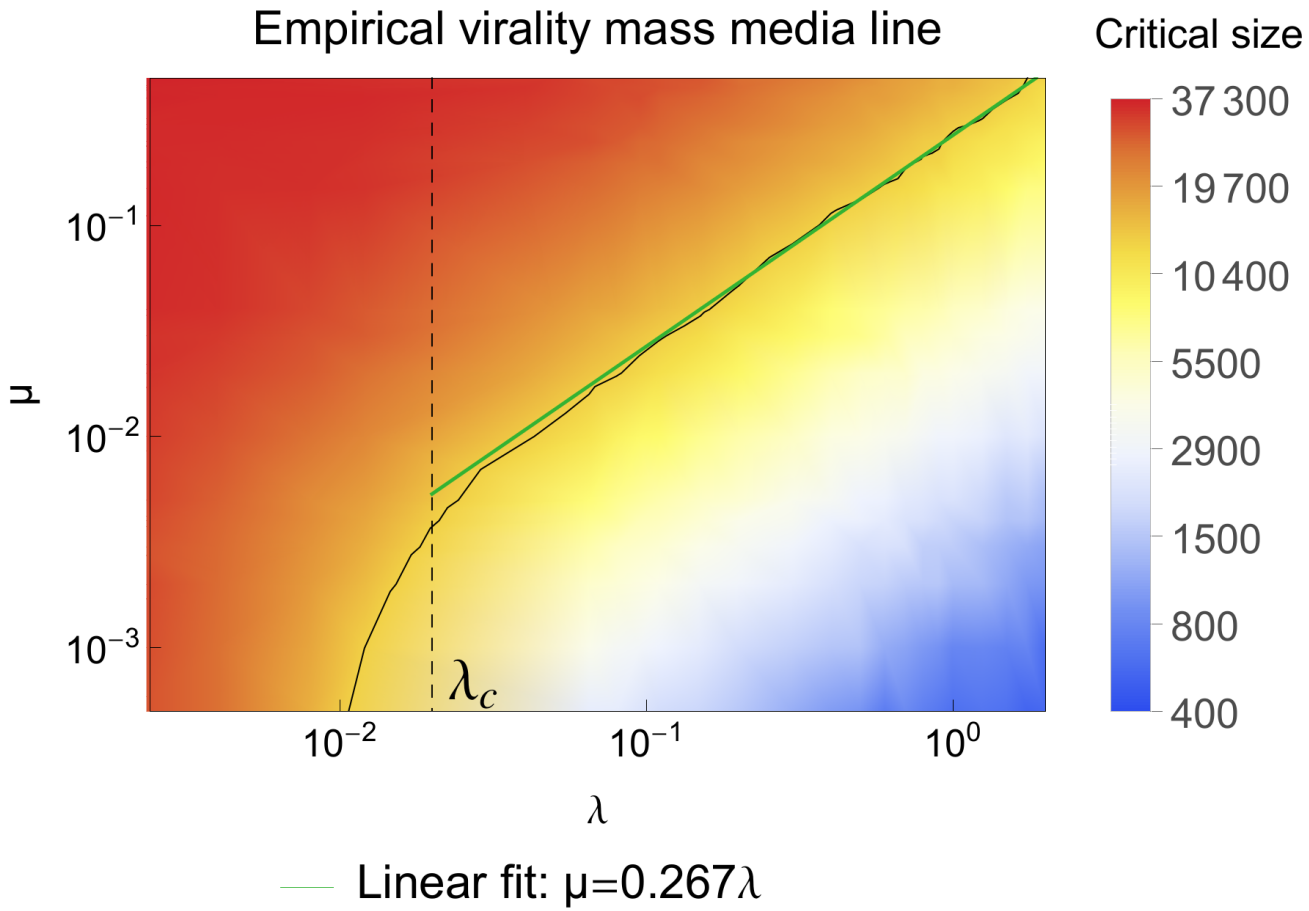
Interplay between **virality** and **mass media** dynamics is the underlying principle of the OSN evolution.

RESULTS FROM VIRALITY MASS MEDIA MODEL

BALANCE BETWEEN VIRALITY AND MASS MEDIA INFLUENCE



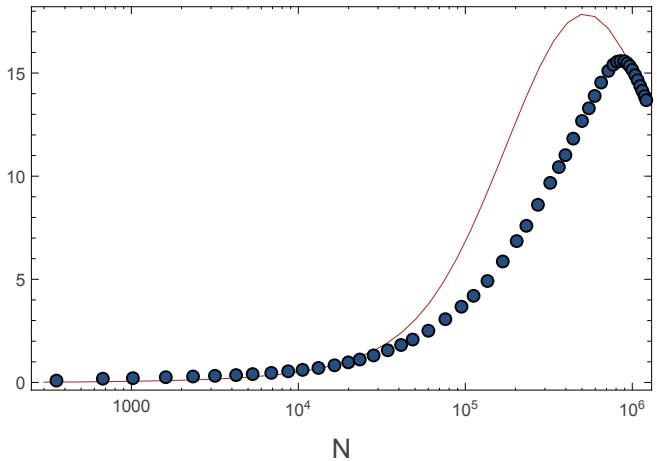
BALANCE BETWEEN VIRALITY AND MASS MEDIA INFLUENCE



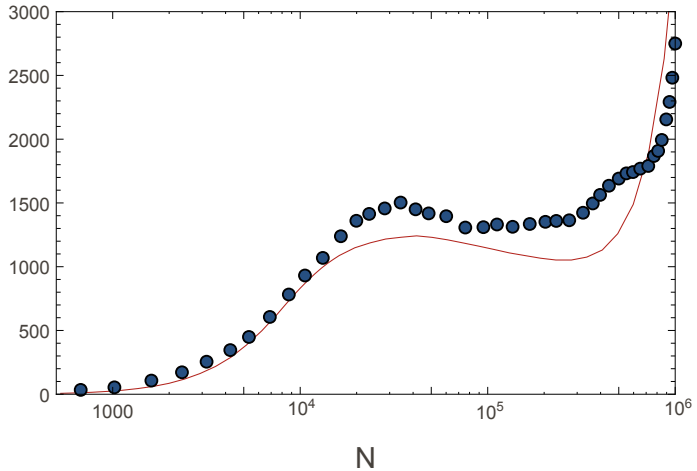
Virality is **four times** stronger than the **mass media** effect.

TOPOLOGICAL EVOLUTION

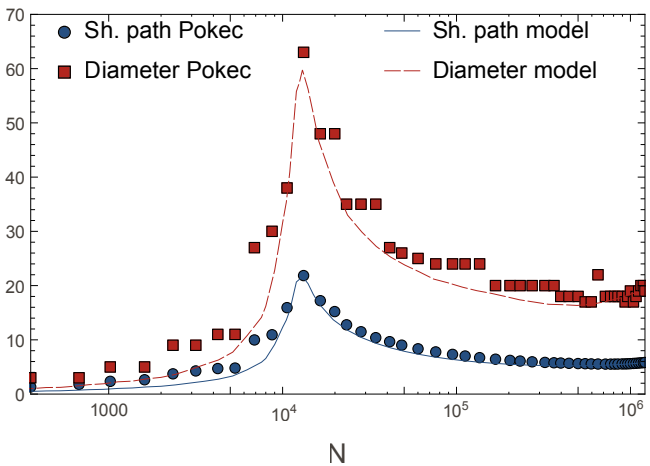
Growing **mean degree**



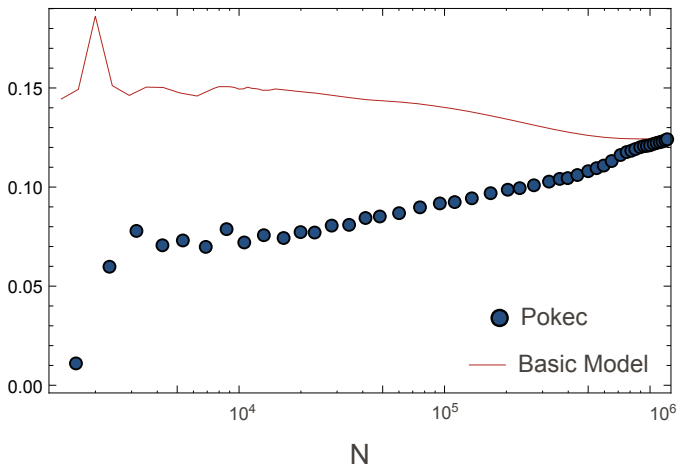
Number of **components**



Shrinking **pathlength**



Decreasing clustering



SUMMARY: VIRALITY MASS MEDIA MODEL

- **Social structure** underlying the OSN matters
- **Mechanisms: Virality** and **mass media** influence
- **Quantification** of **viral** and **mass media** effect
- Global topology ✓
- Local topology ✗

EXTENDED MODEL

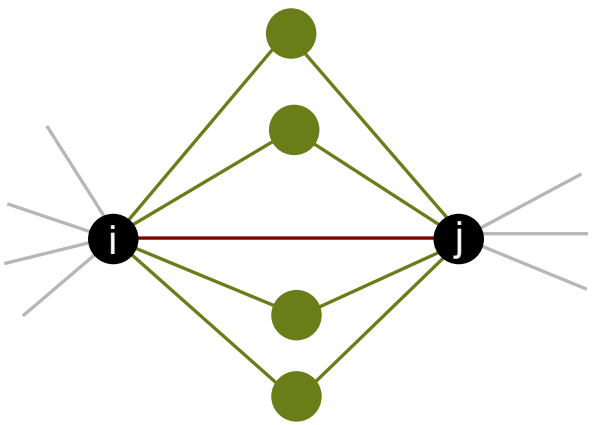
LOCAL NEIGHBORHOOD STRUCTURE

»It is argued that the degree of **overlap of two individuals' friendship networks** varies directly with the **strength of their tie** to one another.«[1]

[1] Quoted from »The strength of weak ties«, M. Granovetter, 1973

LOCAL NEIGHBORHOOD STRUCTURE

»It is argued that the degree of **overlap of two individuals' friendship networks** varies directly with the **strength of their tie** to one another.«[1]

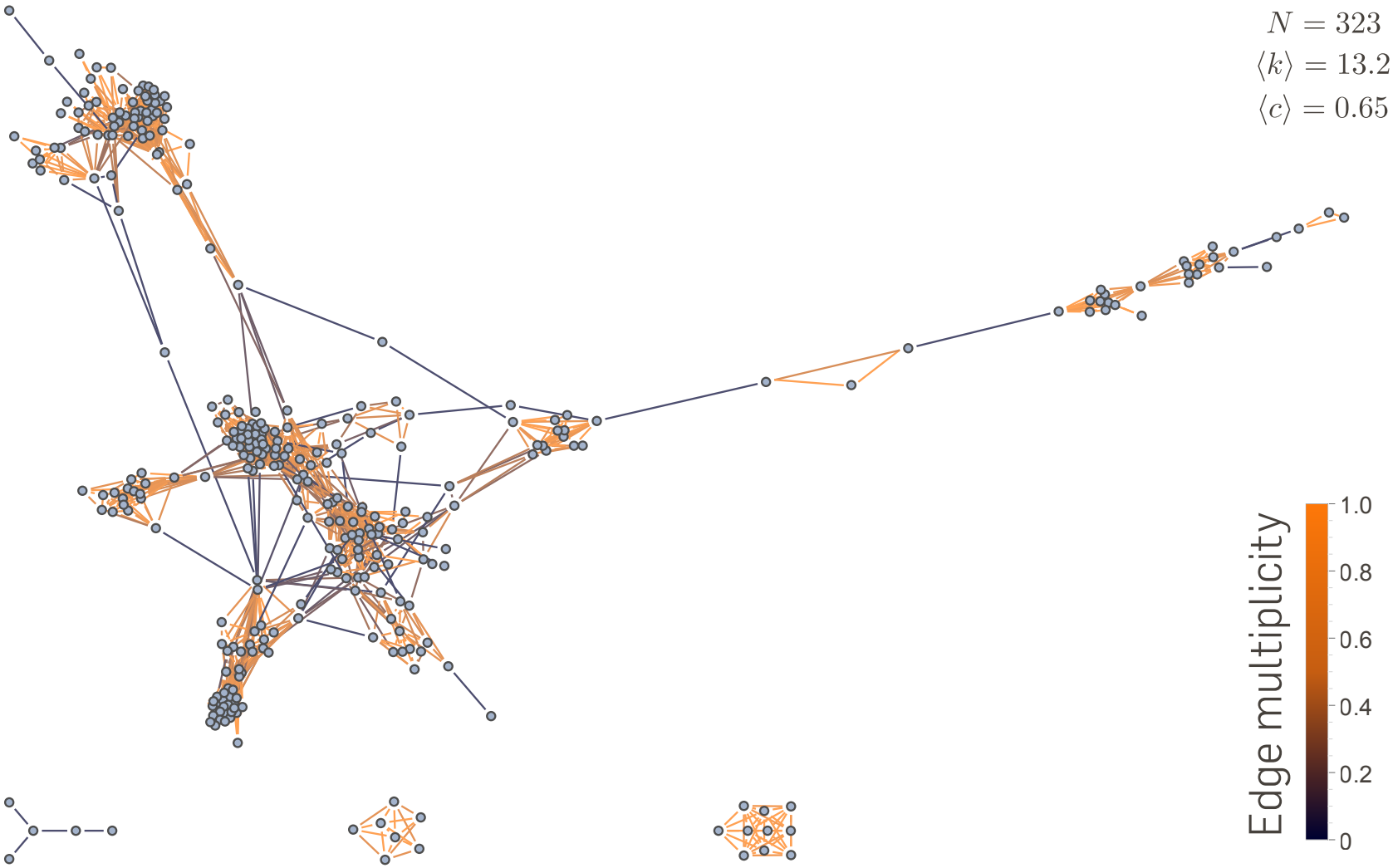


Edge multiplicity is given my the number of the **common neighbors** of the nodes at the end of the edge.

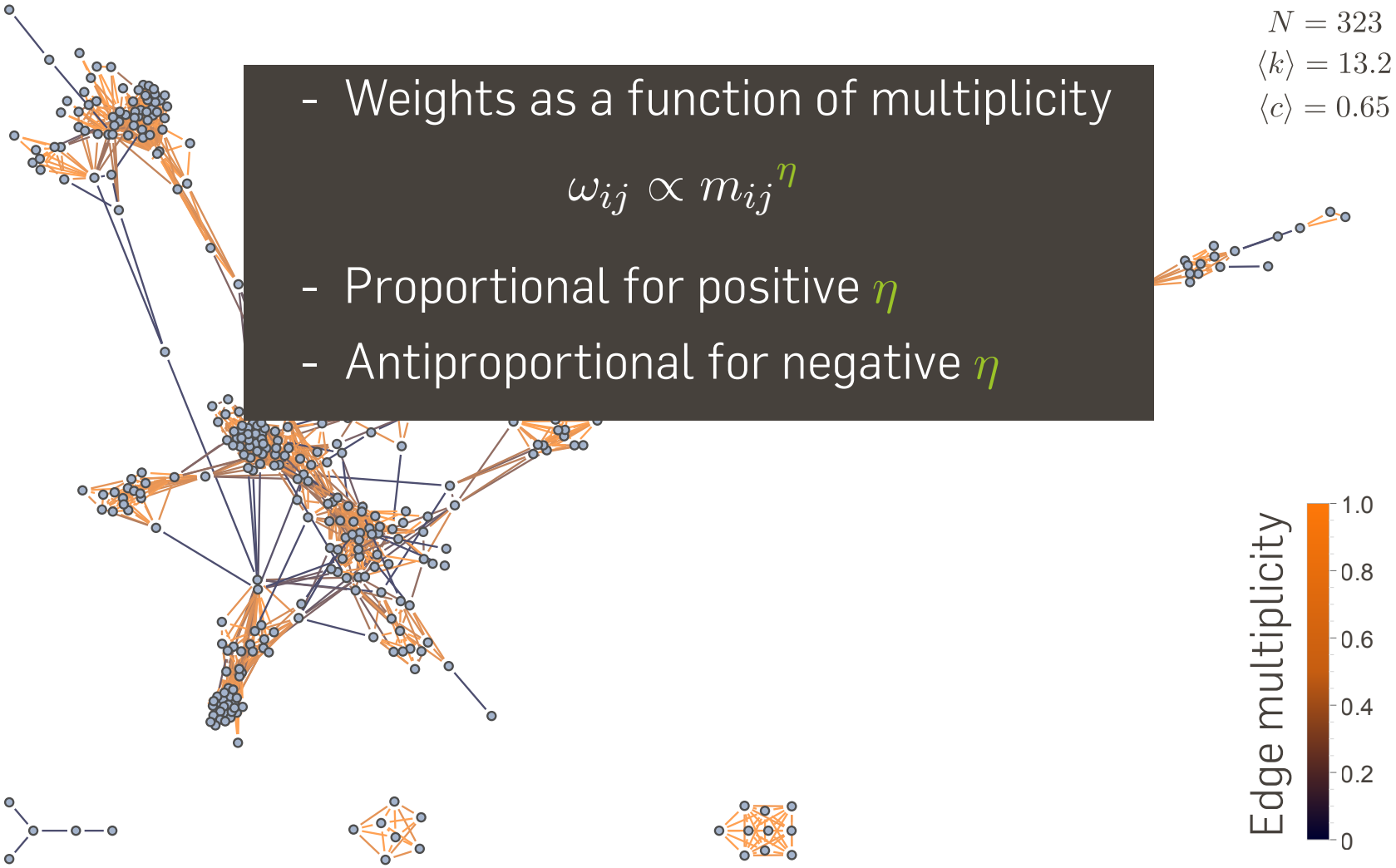
$$m_{ij} = \frac{N[\blacksquare]}{\min[k_i, k_j]}$$

[1] Quoted from »The strength of weak ties«, M. Granovetter, 1973

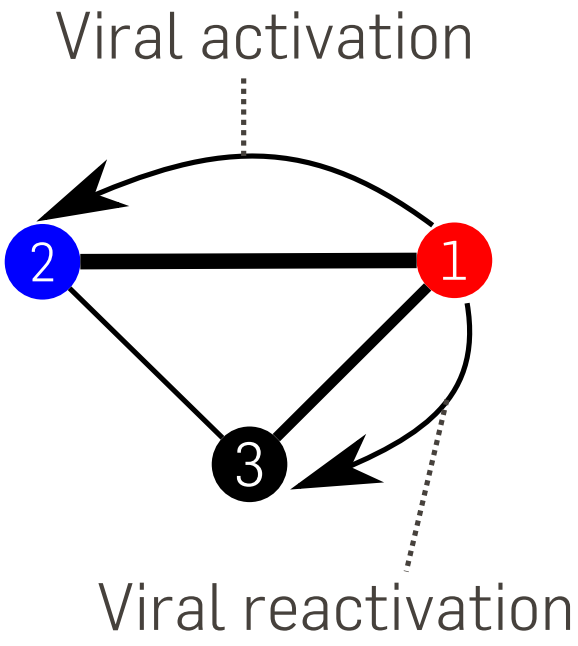
MULTIPLICITY AND LINK WEIGHTS



MULTIPLICITY AND LINK WEIGHTS

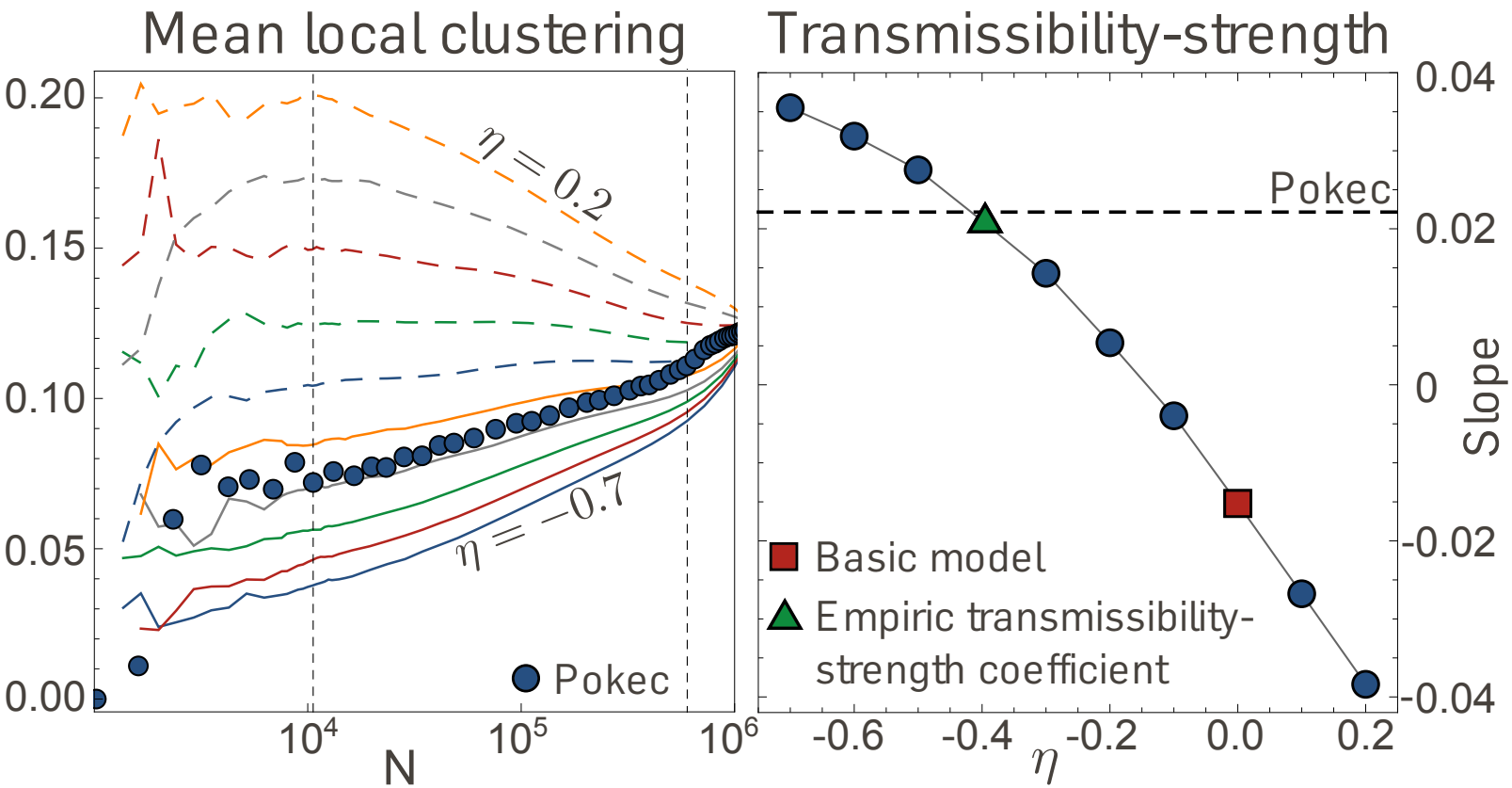


LOCAL TRANSMISSIBILITY OF THE VIRAL EFFECT

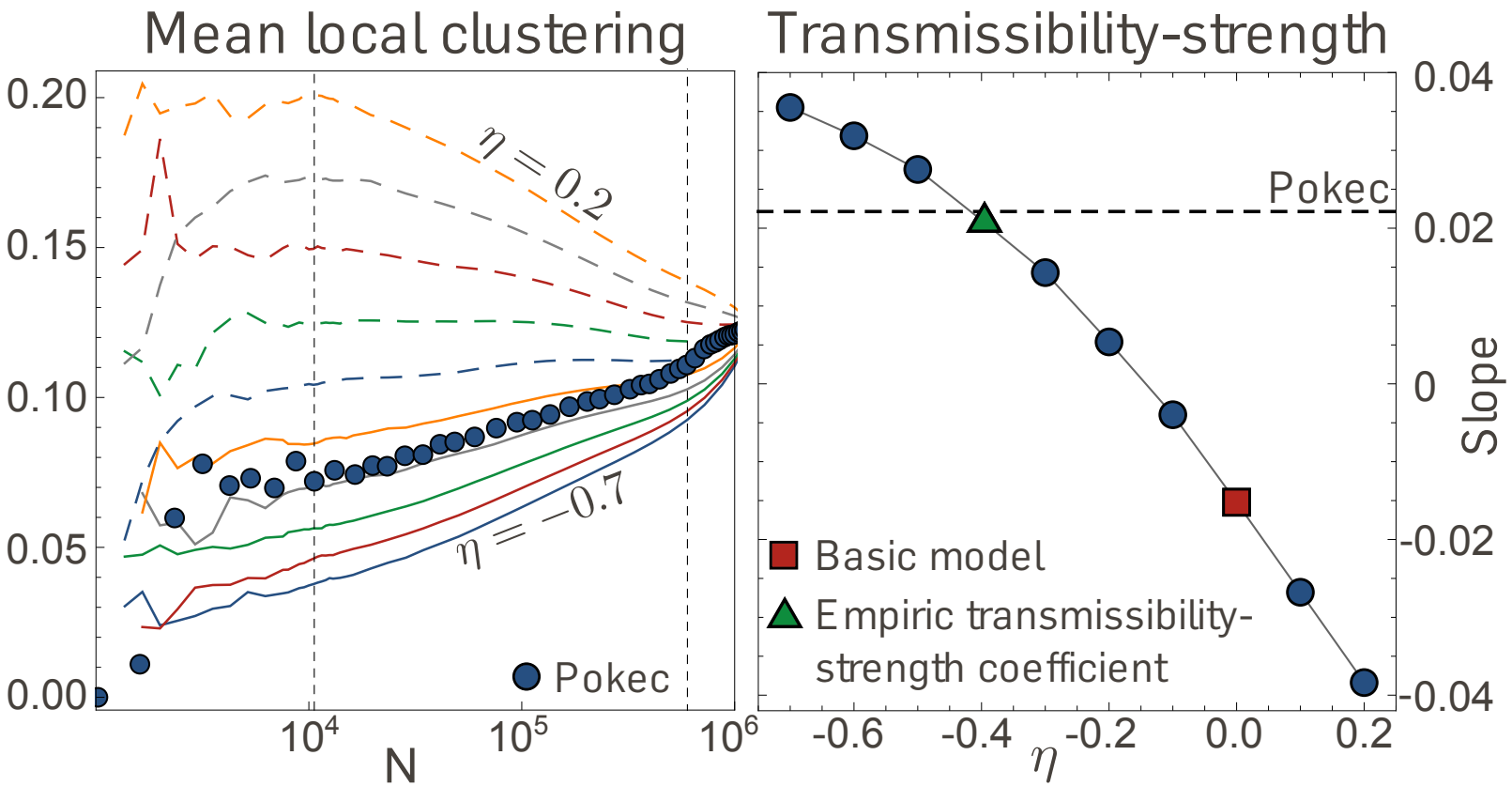


- **Viral activation** $(1 \leftrightarrow 2)$ occurs with **rate** $\lambda\omega_{1,2}$
- **Viral reactivation** $(1 \leftrightarrow 3)$ occurs with **rate** $\lambda\omega_{1,3}$
- **Mass media** effect and **de-activation** are not affected by assigned weights

TRANSMISSIBILITY AND SOCIAL TIES



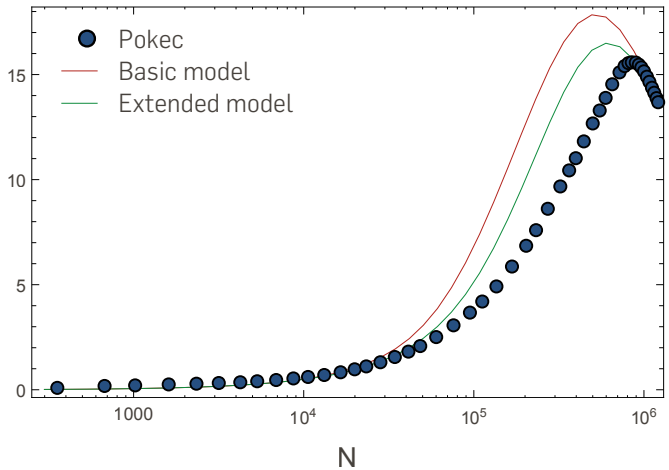
TRANSMISSIBILITY AND SOCIAL TIES



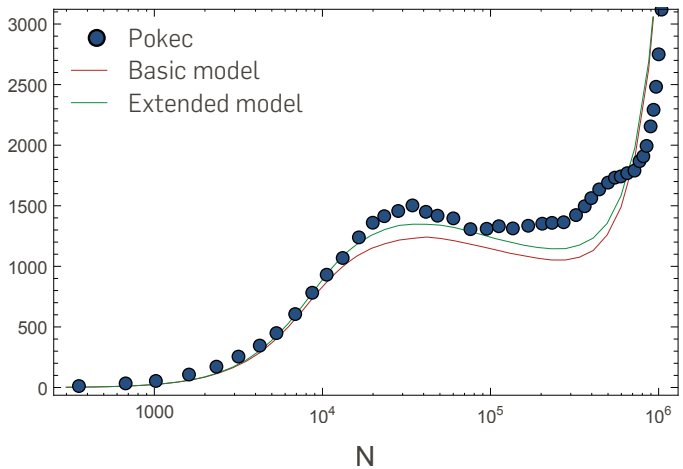
Weak ties exhibit a higher transmissibility for the **viral effect** than strong ties.

COMPARISON BASIC & EXTENDED MODEL

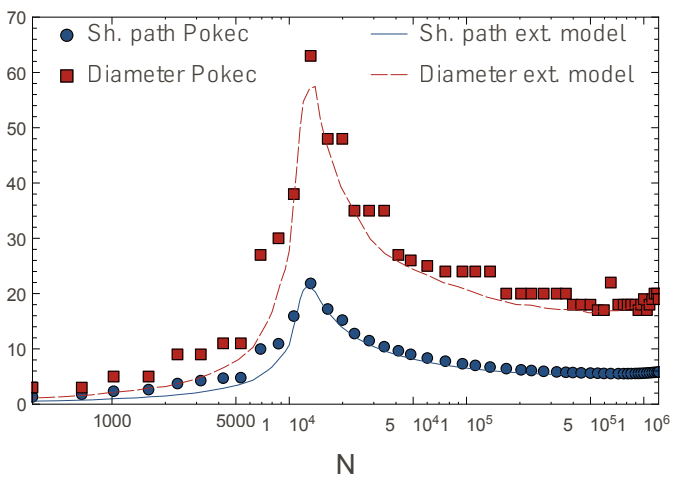
Growing **mean degree**



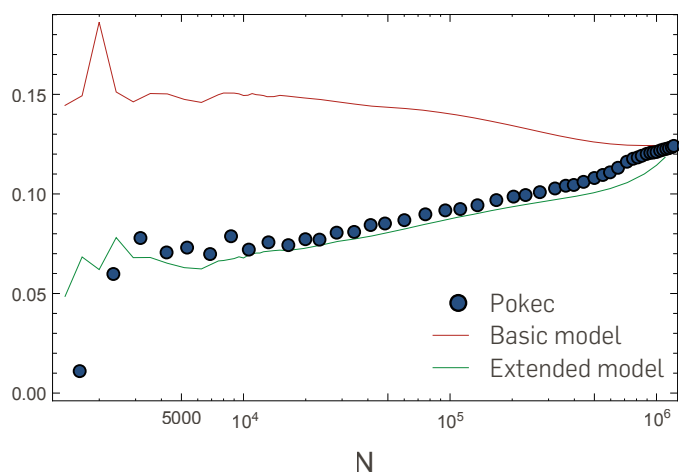
Number of **components**



Shrinking **pathlength**



Increasing clustering



SUMMARY: EXTENDED VIRALITY MASS MEDIA MODEL

- **Strength** of ties function of **multiplicity**
- **Viral** effect favors **weak ties**
- Global topology ✓
- Local topology ✓

SUMMARY & DISCUSSION

SUMMARY: TAKE HOME MESSAGES

- We need **models** to understand the underlying **mechanisms**
- **Validation concept:** Reproduction of empirical evolution
- **Two-layer** model: **Identify** & **quantify virality** and **mass media** influence as underlying **mechanisms**
- Local topological evolution: **higher viral transmissibility** for **weak ties**

OUTLOOK & FURTHER RESEARCH

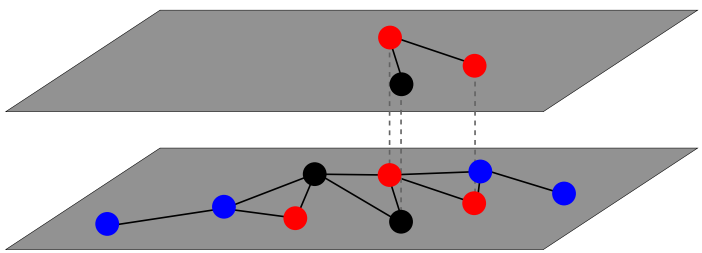


Figure: Single OSN layer.

OUTLOOK & FURTHER RESEARCH

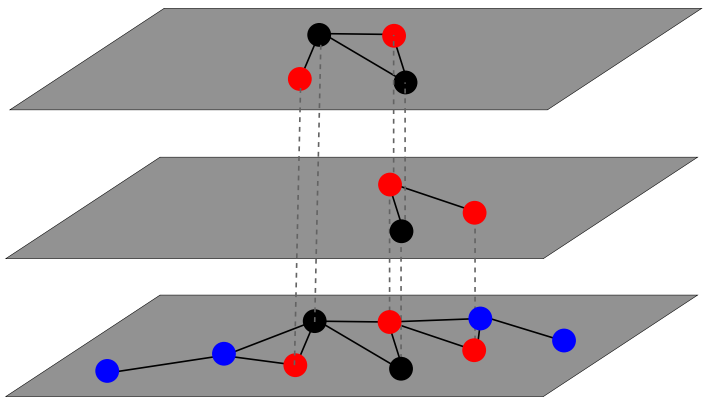


Figure: Multiple OSN layers.

OUTLOOK & FURTHER RESEARCH

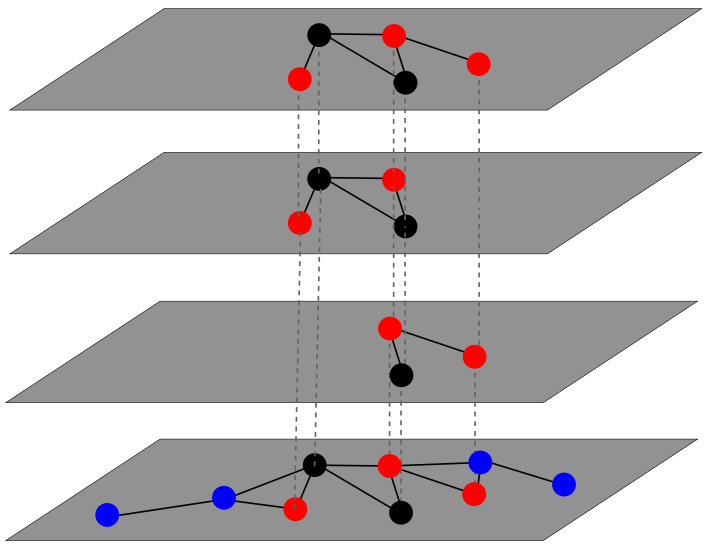


Figure: Multiple OSN layers.

OUTLOOK & FURTHER RESEARCH

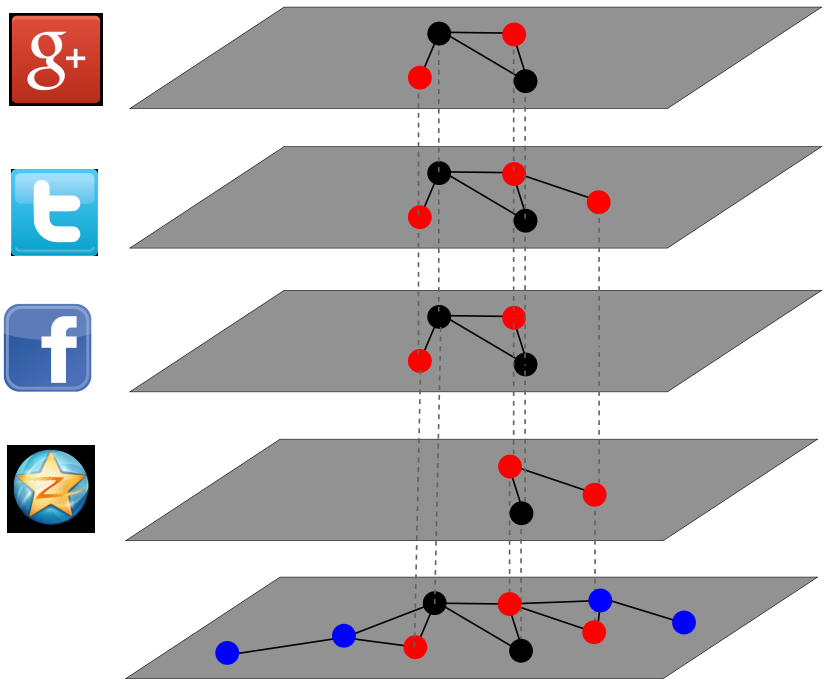


Figure: Multiple OSN layers.

Multiple OSN layers:

- Competition
- Parameters $\vec{\lambda}$, $\vec{\mu}$
- For instance:

$$\vec{\lambda} = \vec{\lambda}\left(\vec{A}, \vec{P}, \vec{S}\right)$$

$$\vec{\mu} = \vec{\mu}\left(\vec{A}, \vec{P}, \vec{S}\right)$$

- Distribution of network sizes, activities etc...

OUTLOOK & FURTHER RESEARCH

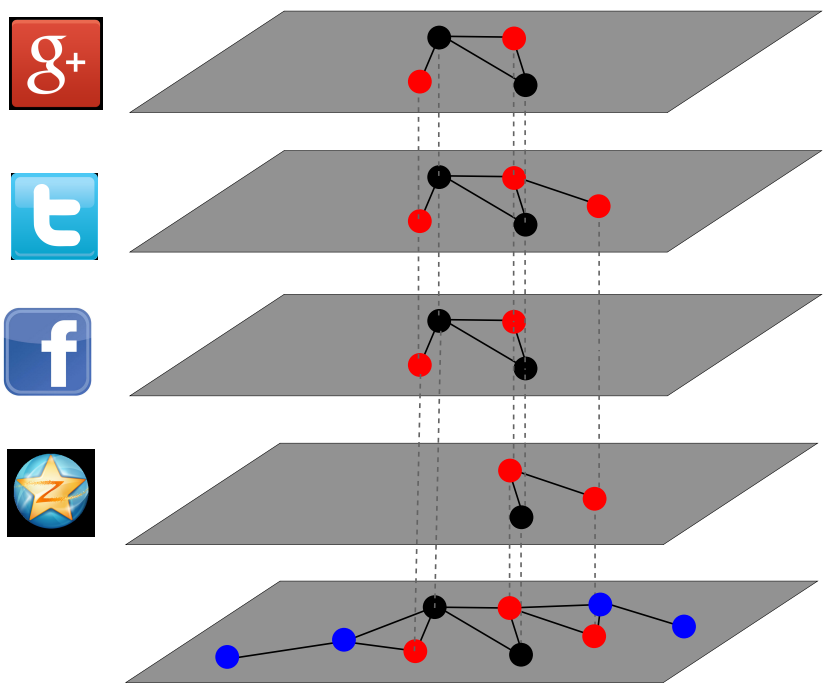


Figure: Multiple OSN layers.

Multiple OSN layers:

- Competition
- Parameters $\vec{\lambda}$, $\vec{\mu}$
- For instance:

$$\vec{\lambda} = \vec{\lambda}\left(\vec{A}, \vec{P}, \vec{S}\right)$$

$$\vec{\mu} = \vec{\mu}\left(\vec{A}, \vec{P}, \vec{S}\right)$$

- Distribution of network sizes, activities etc...

Ongoing research: Can we reproduce the distribution of empiric OSN sizes with a **competition model**?

QUESTIONES, COMMENTS & MORE

L. Takac and M. Zabovsky.
»Data analysis in public social networks.«
International Scientific Conference and International Workshop Present Day Trends of Innovations, 2012.

K. Kleineberg and M. Boguñá.
»Trade-off between virality and mass media influence in the topological evolution of online social networks«.
In progress.

Please don't hesitate to contact me.

Kaj Kolja Kleineberg
kkl@correu.ffn.ub.edu