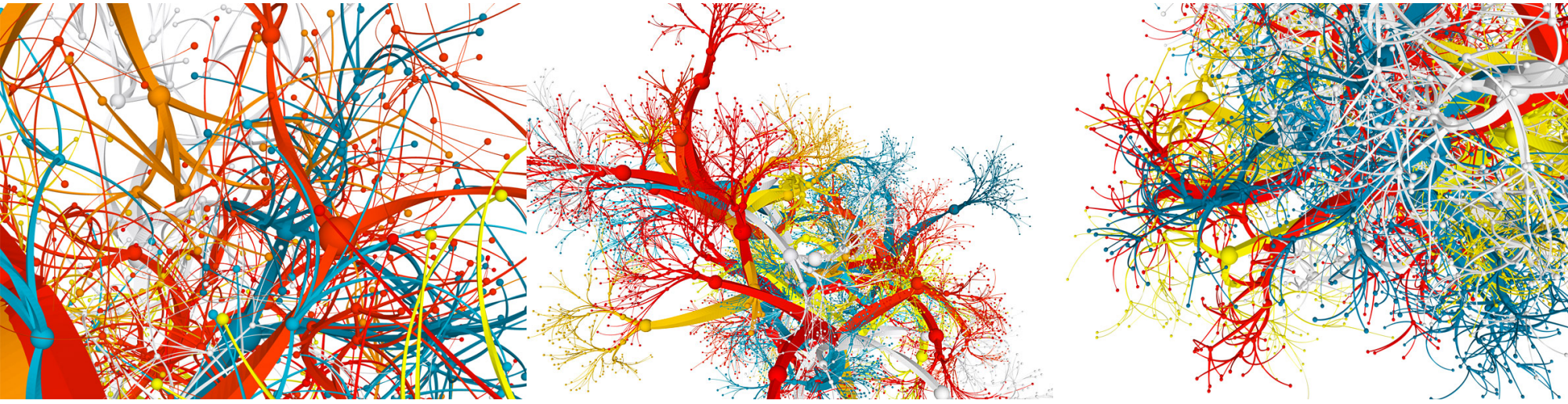
# Community subgraph densification

Róbert Pálovics



June 26, 2014

# PERSONAL

- Robert Palovics
- rpalovics@ilab.sztaki.hu
- Supervisor: András Benczúr
- BSc and MSc degree in Theoretical Physics
  - TU Budapest
- Phd student in Mathematics (2nd year)
  - TU Budapest
  - InfoLab @ Hungarian Academy of Sciences
    https://dms.sztaki.hu/en/
    http://www.sztaki.hu/department/INFOLAB/
- computer science:)

# RESEARCH INTEREST

- Information (epidemic) spreading in networks
  - Cascades in online social networks
  - **Community densification laws**
- Models of complex networks & large graphs
  - Accelerated growth of networks
  - Densification laws
- Recommender systems (RS)
  - Online (temporal) recommendations
  - Temporal prediction and evaluation
  - Online collaborative filtering
  - Context-based RS
  - Location based RS
  - Using social information in RS

# INFORMATION SPREAD IN NETWORKS

Network + Diffusion process $\leftrightarrow$ Measurements

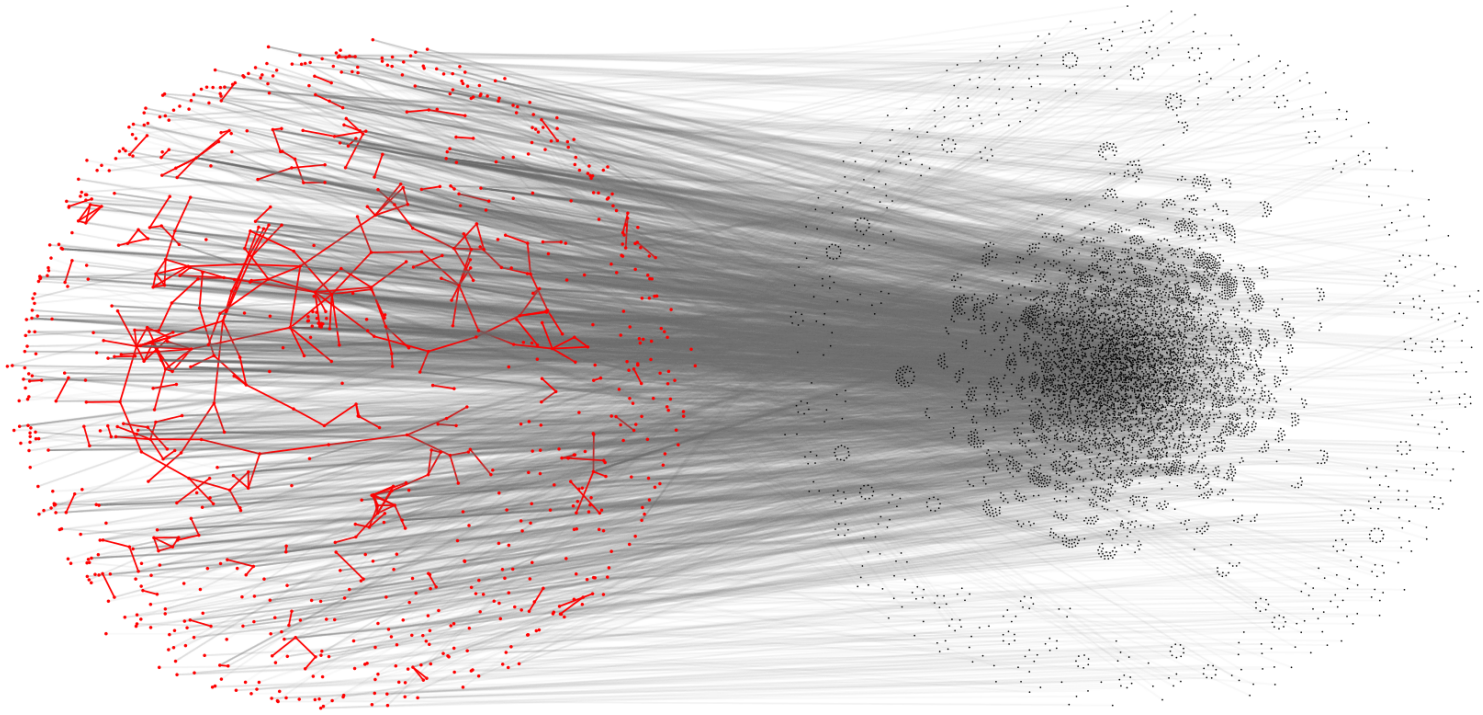
- diffusion $\leftrightarrow$ observable time series
- fixed network + time series

# COMMUNITY DENSIFICATION LAW

- Users adopt a given behavior $a$ after each other
- $G(a, t) = \{$subgraph of users who adopted $a$ before $t\}$

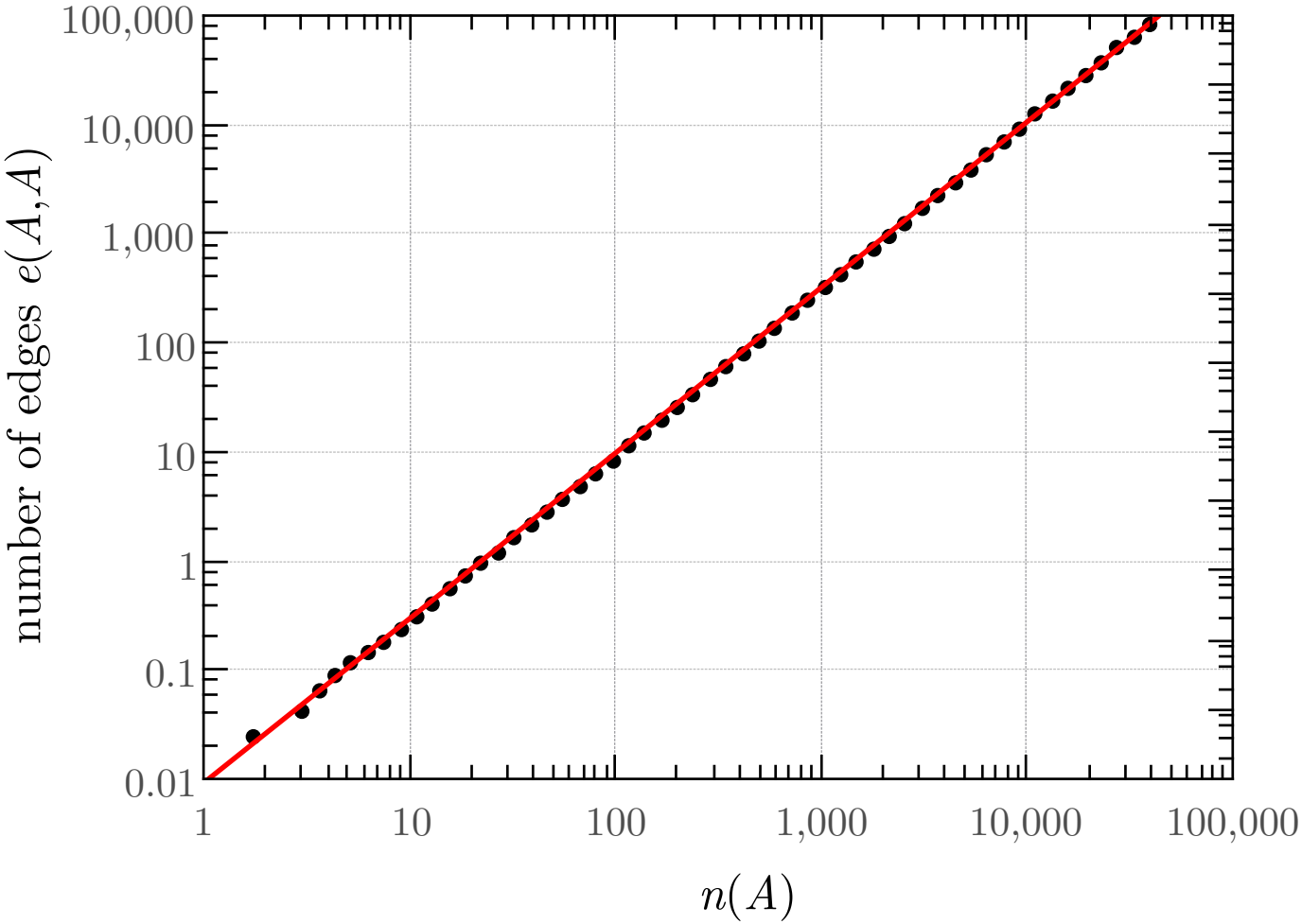
**Datasets**

- artists in Last.fm
- hashtags in Twitter

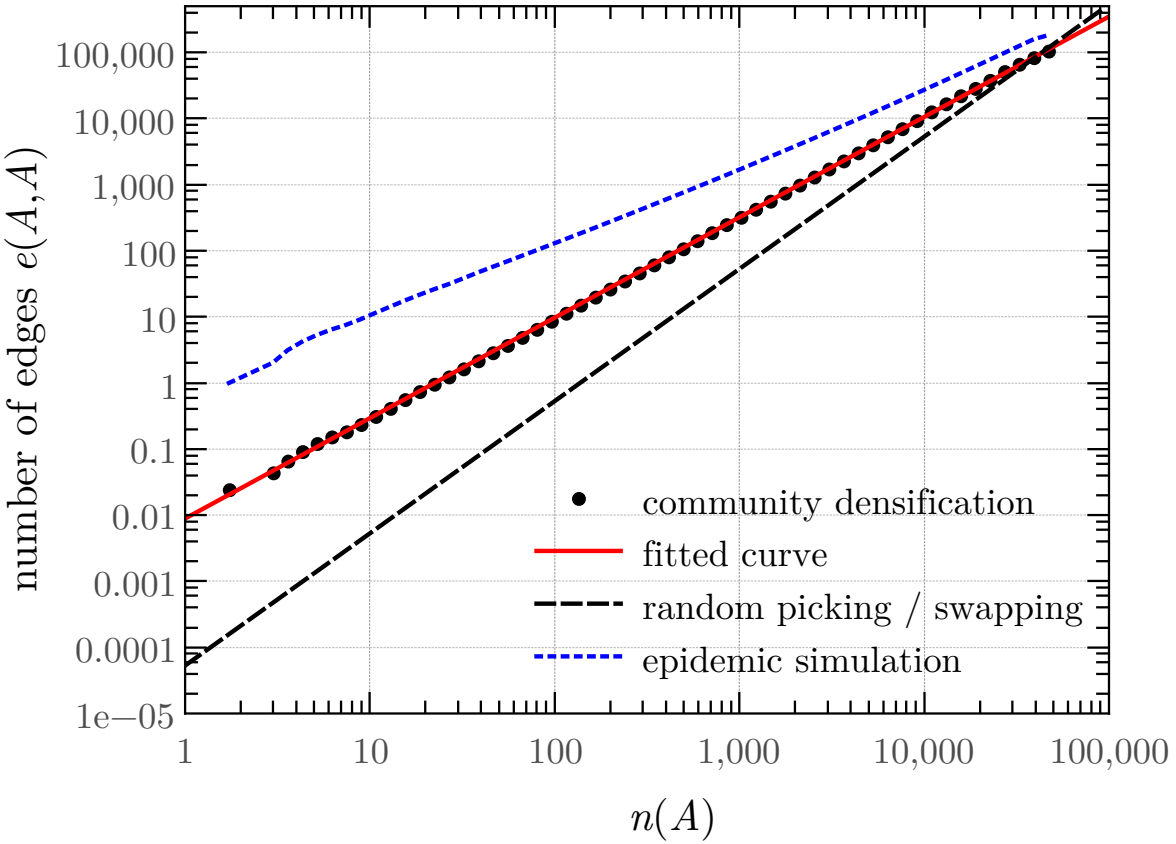


# COMMUNITY DENSIFICATION LAW

- The number of edges $e(a,t)$ is power-law function of the number of nodes $n(a,t)$ in the subgraph with exponent $\gamma < 2$.

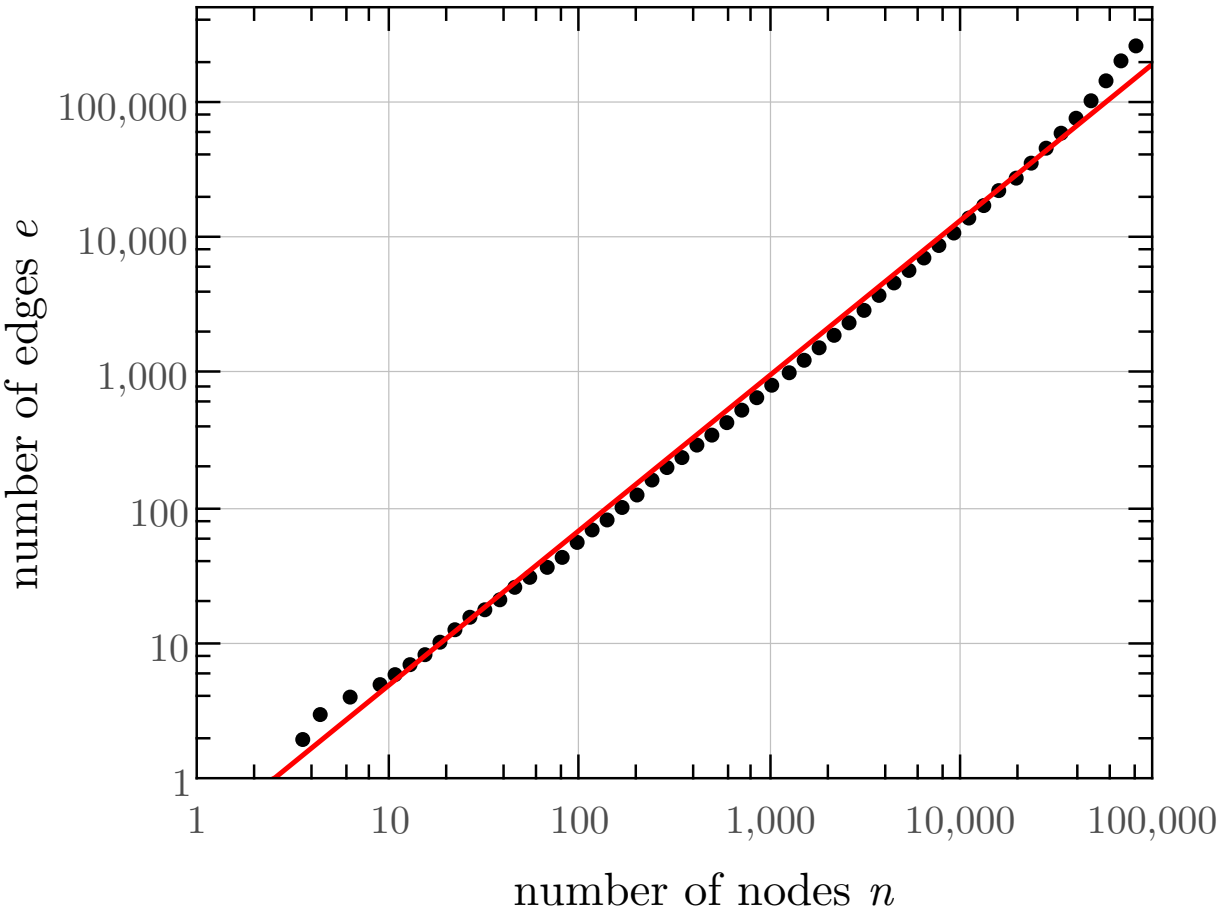


# Comparison of processes




- $\overline{d} = \frac{e}{n}$ $\qquad \rho = \frac{2e}{n(n-1)} \sim \frac{e}{n^2}$
- Densification vs. sparsification
- Maximum spread

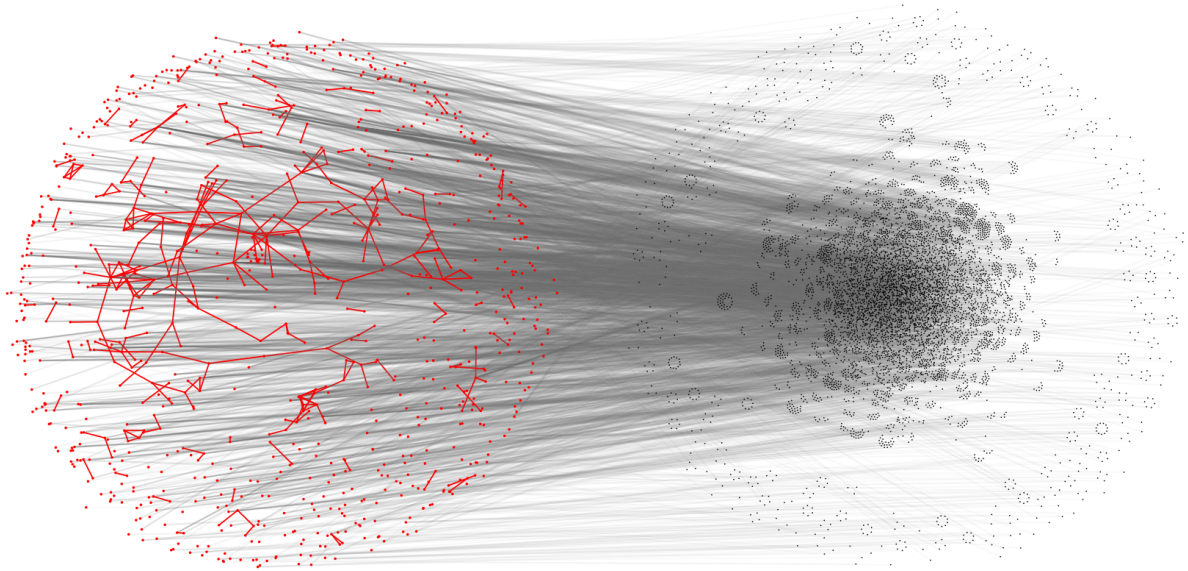
# ACCELERATED GROWTH OF NETWORKS



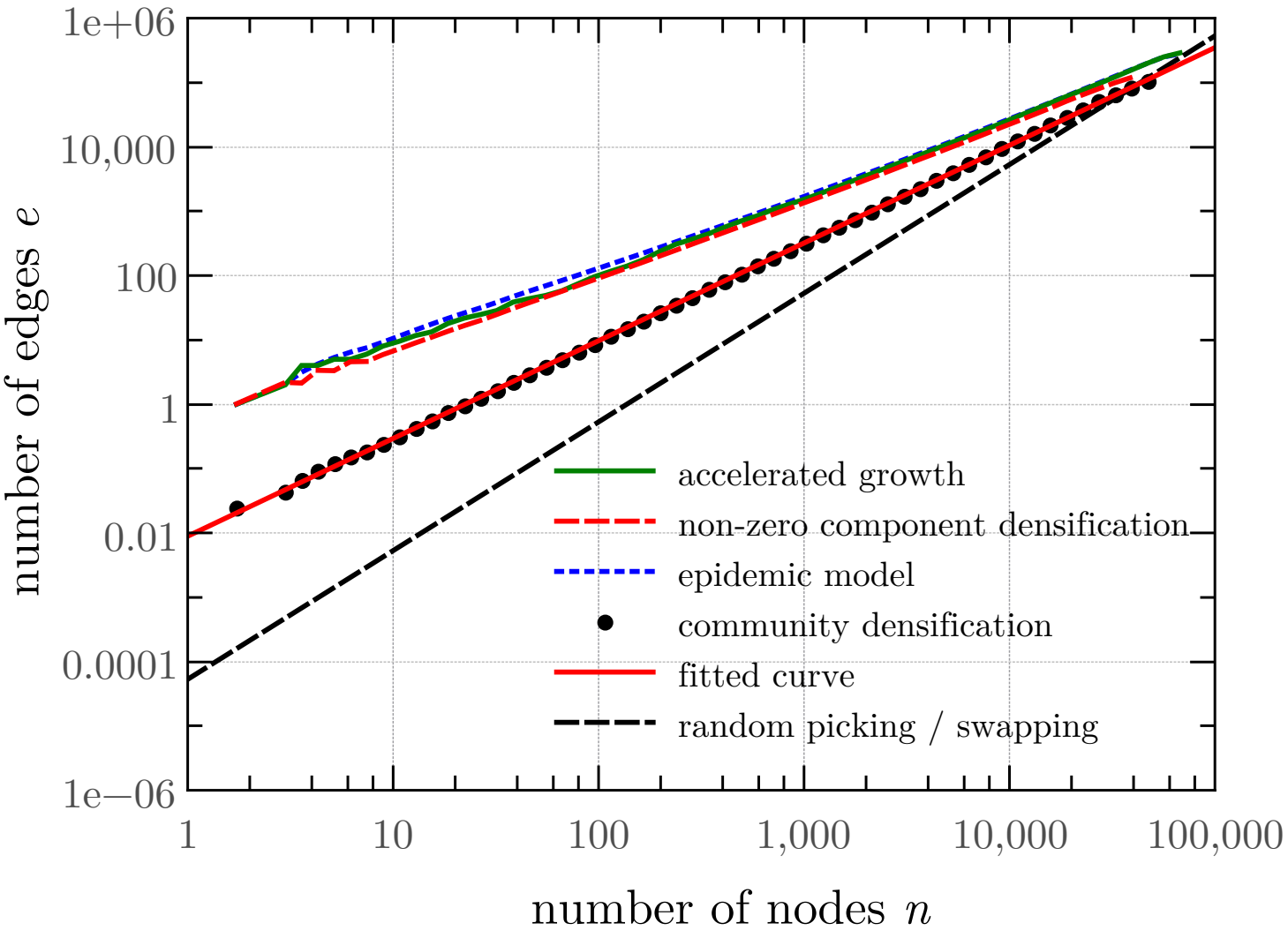

- Growing network, no information diffusion
- $L(t) \propto t^{a+1}$ $\quad$ $e(n) \propto n^{\beta}$

# Non-isolated nodes

- Power law fraction of nodes with at least one edge within the community, with exponent $\delta > 1$.
- The edge number in a community as the function of the number nodes with at least one edge also follows power law ($\beta'$).
- $\beta = \beta'$ (!)



# Summary


1e+06
10,000
100
1
0.01
0.0001
1e−06
number of edges $e$
1
10
100
1,000
10,000
100,000
number of nodes $n$
accelerated growth
non-zero component densification
epidemic model
community densification
fitted curve
random picking / swapping

**Network discovery process**

- Information spreading over a network and the dynamic growth of the network are similar and closely related processes
- The network itself can be considered as a community in a hidden social network

**Work in progress**

- Develop a network model that describes this effect
- Develop an information spreading model
- **Is the degree sequence sufficient (swapping, $\beta$-model)**?