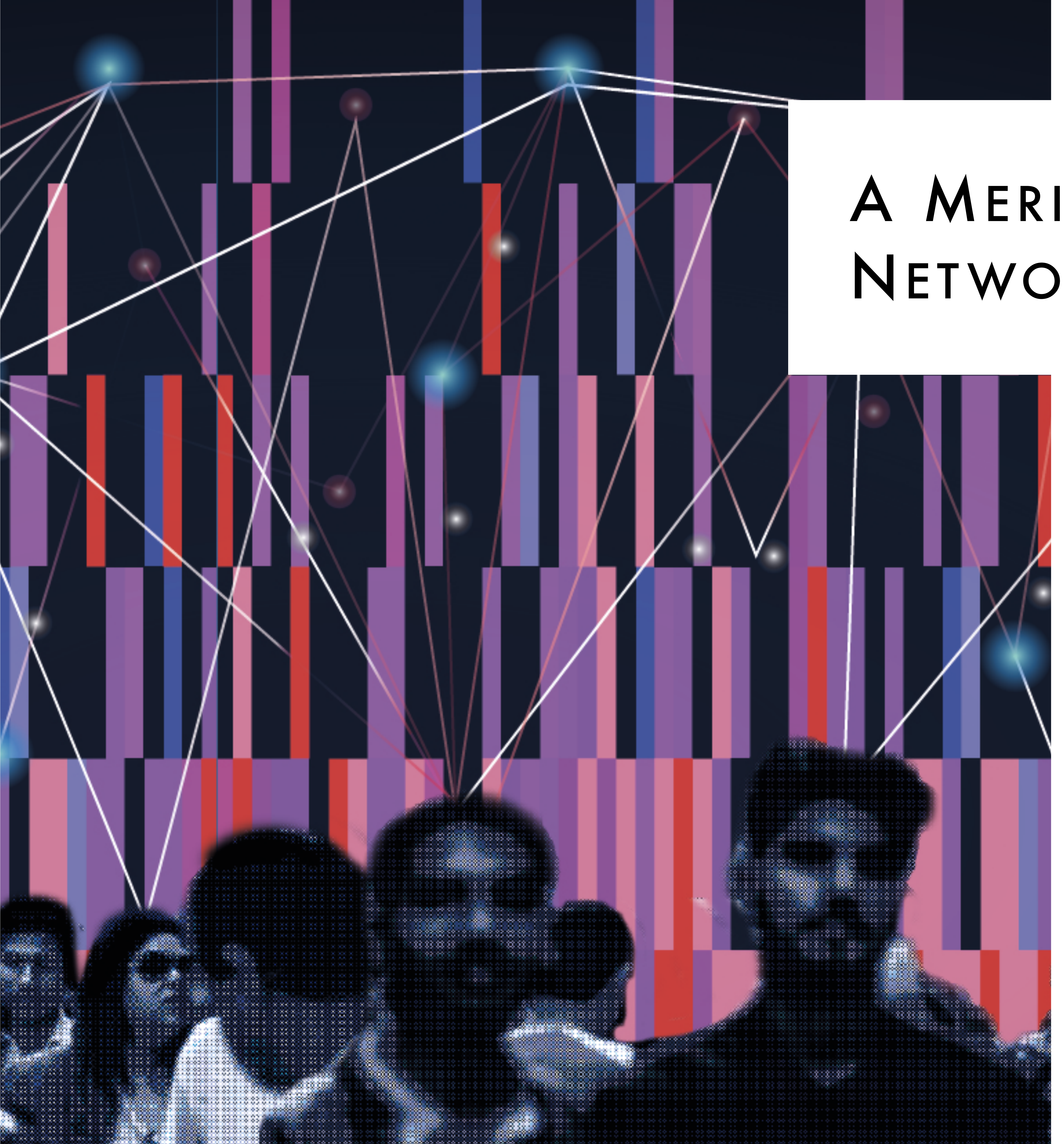


A Meritocratic Network Formation Model

Workshop on Dynamics in Social and Economics Networks,
59th IEEE Conference on Decision and Control

December 12-13, 2020

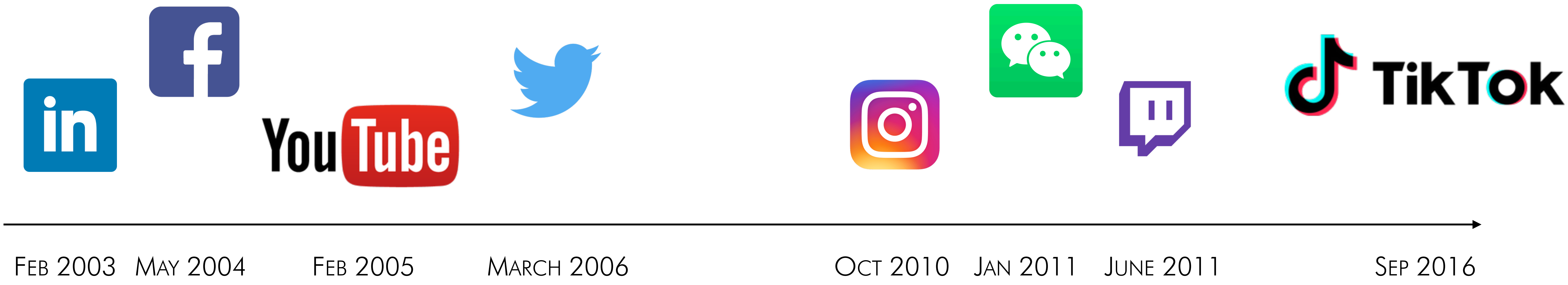
Nicolò Pagan

INSTITUT FÜR AUTOMATIK

ETH zürich

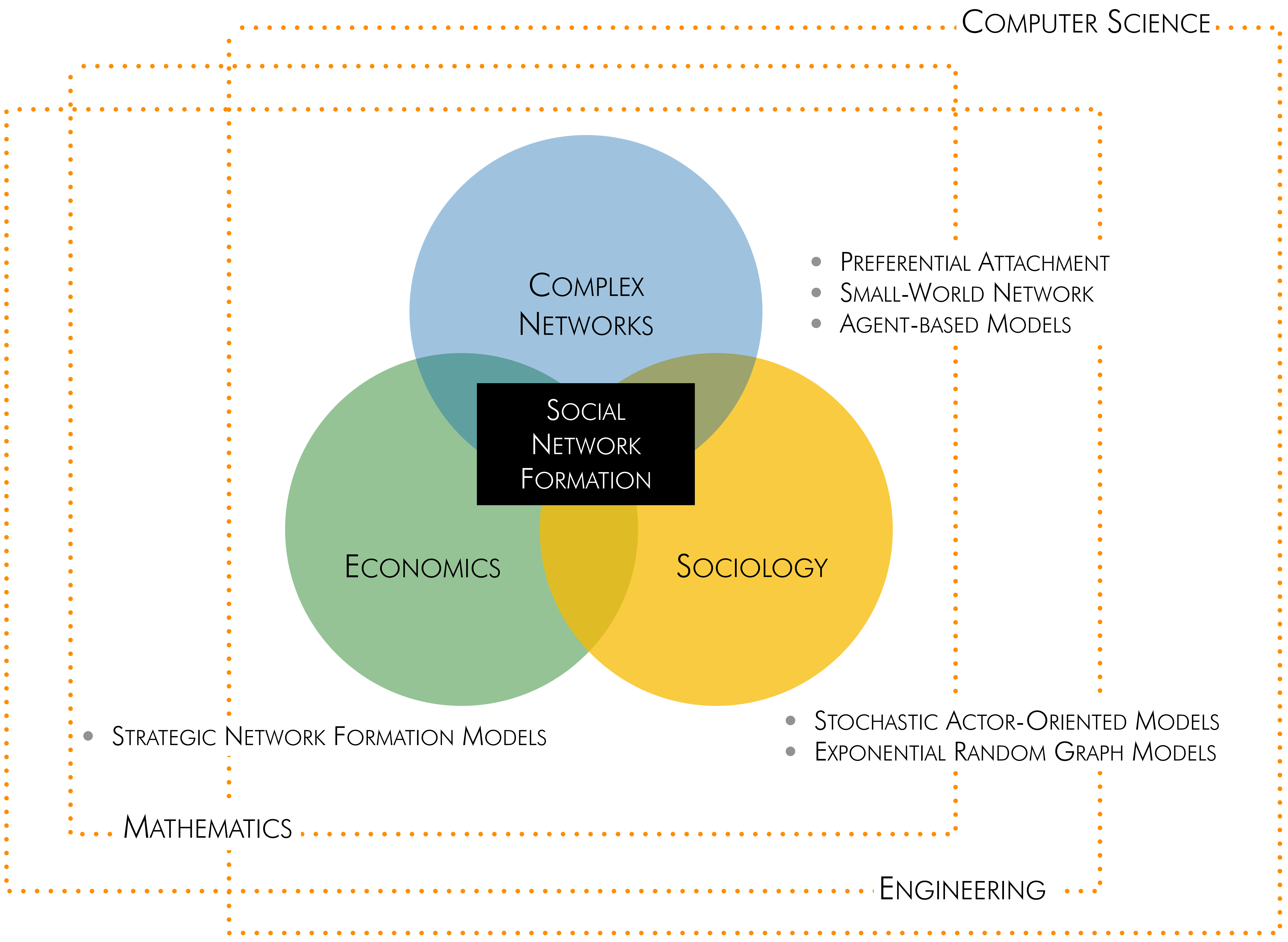
Motivation

- Online Social Networks are ubiquitous.



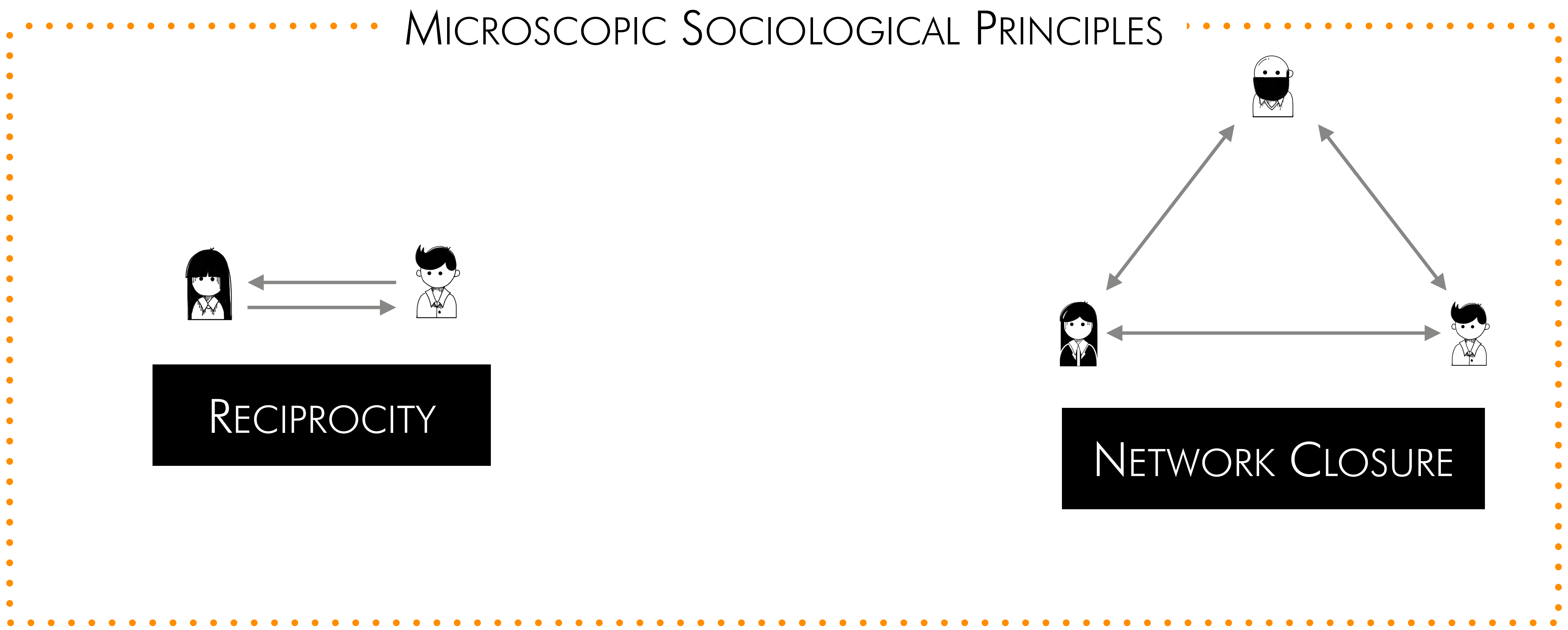
MOTIVATION

- Online Social Networks are ubiquitous.
- Researchers from different communities have started considering the problem of ***Social Network Formation***.



Motivation

- Online Social Networks are ubiquitous.
- Researchers from different communities have started considering the problem of *Social Network Formation*.
- Sociologists found that actors' decision on "Offline" Social Networks are driven by sociological principles.

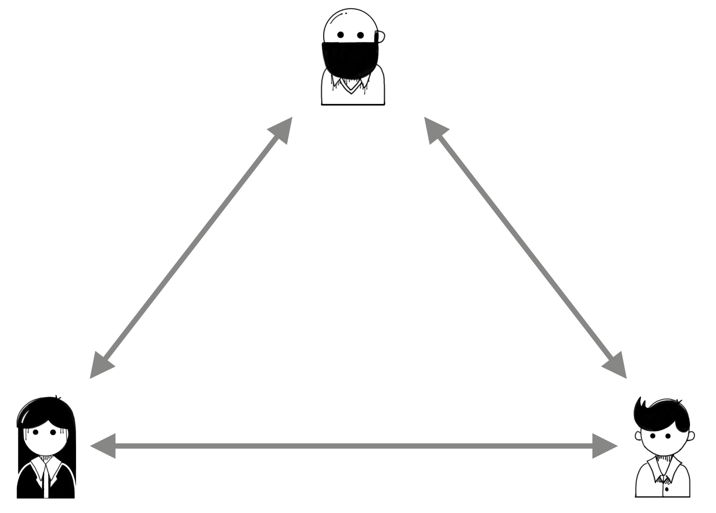


Motivation

- Online Social Networks are ubiquitous.
- Researchers from different communities have started considering the problem of ***Social Network Formation***.
- Sociologists found that actors' decision on "Offline" Social Networks are driven by sociological principles.
- "Old-generation" Online Social Networks try to mimic "offline" Social Networks.

Microscopic Sociological Principles

Reciprocity



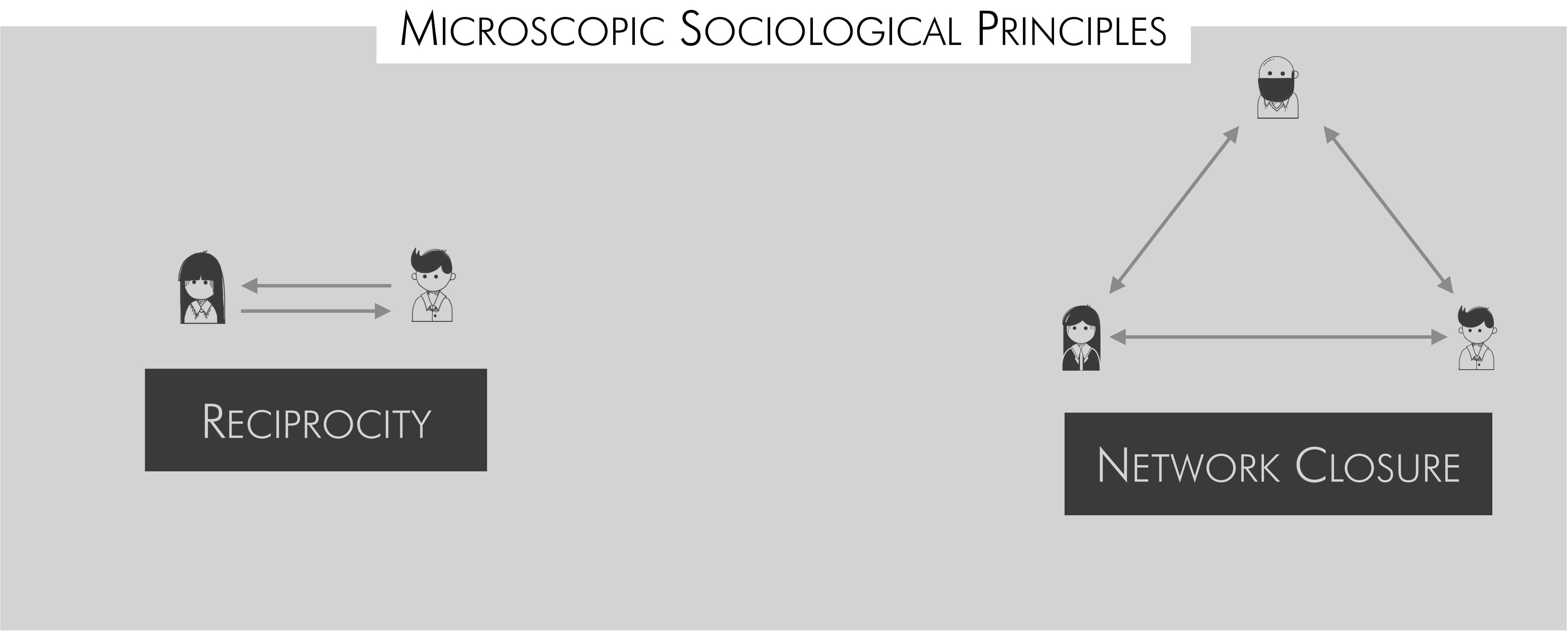
Network Closure

Reciprocity	100%
Network Closure	40%

S. Teng, M. Yeh and K. Chuang, "Toward understanding the mobile social properties: An analysis on instagram photo-sharing network". 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

Motivation

- Online Social Networks are ubiquitous.
- Researchers from different communities have started considering the problem of *Social Network Formation*.
- Today's most successful Online Social Networks show *limited reciprocity* and *network closure*.

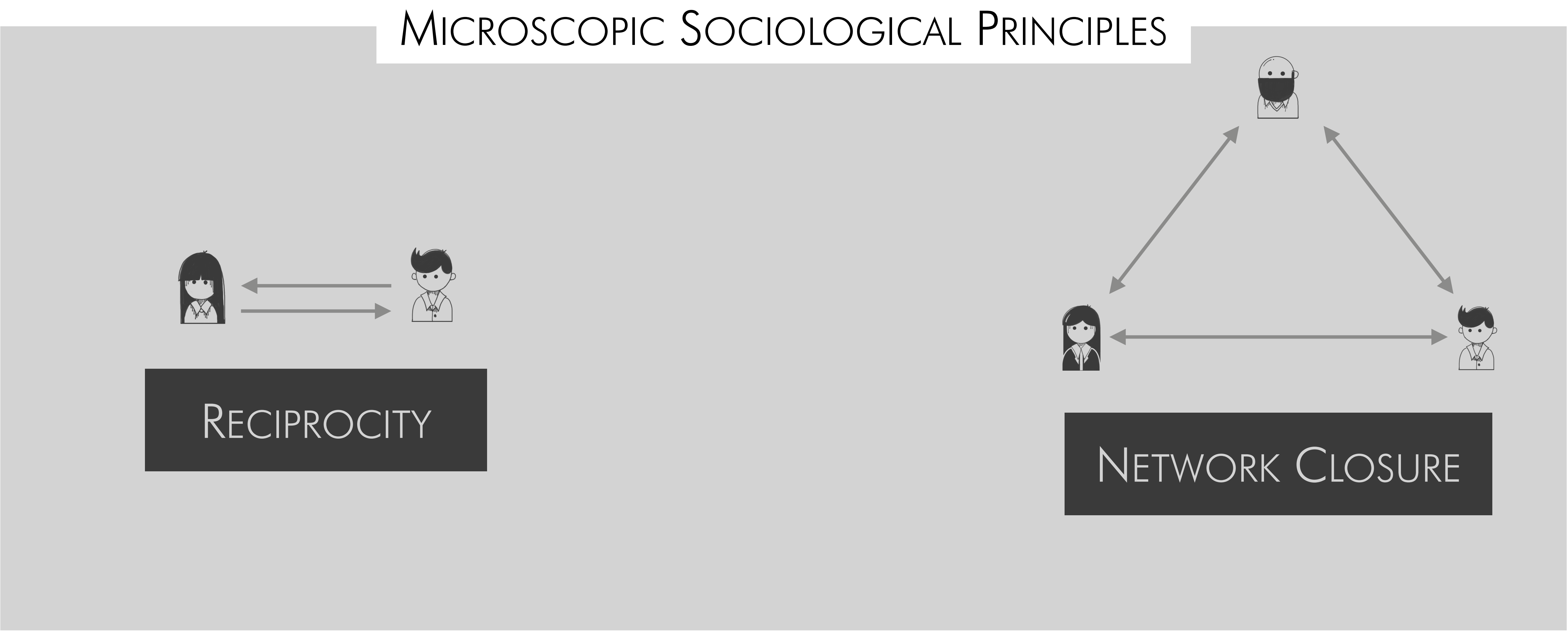


Reciprocity	100%	22%	14%
Network Closure	40%	23%	8%

S. Teng, M. Yeh and K. Chuang, "Toward understanding the mobile social properties: An analysis on instagram photo-sharing network". 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

Motivation

- Online Social Networks are ubiquitous.
- Researchers from different communities have started considering the problem of *Social Network Formation*.
- Today's most successful Online Social Networks show *limited reciprocity* and *network closure*.
- Rather, they are based on *User-Generated Content* (typically, images or videos), which is widely unexplored in the literature.



UGC-Based Platforms

They have a huge impact in our daily lives, e.g.:

- Activism;
- Travel planning;
- Food trends;
- Social Media Marketing;
- Spreading of (mis)information;
- …

QUALITY-BASED NETWORK FORMATION MODEL

Observations

- Actors have different interests, e.g, travelling, food, politics, …

Observations

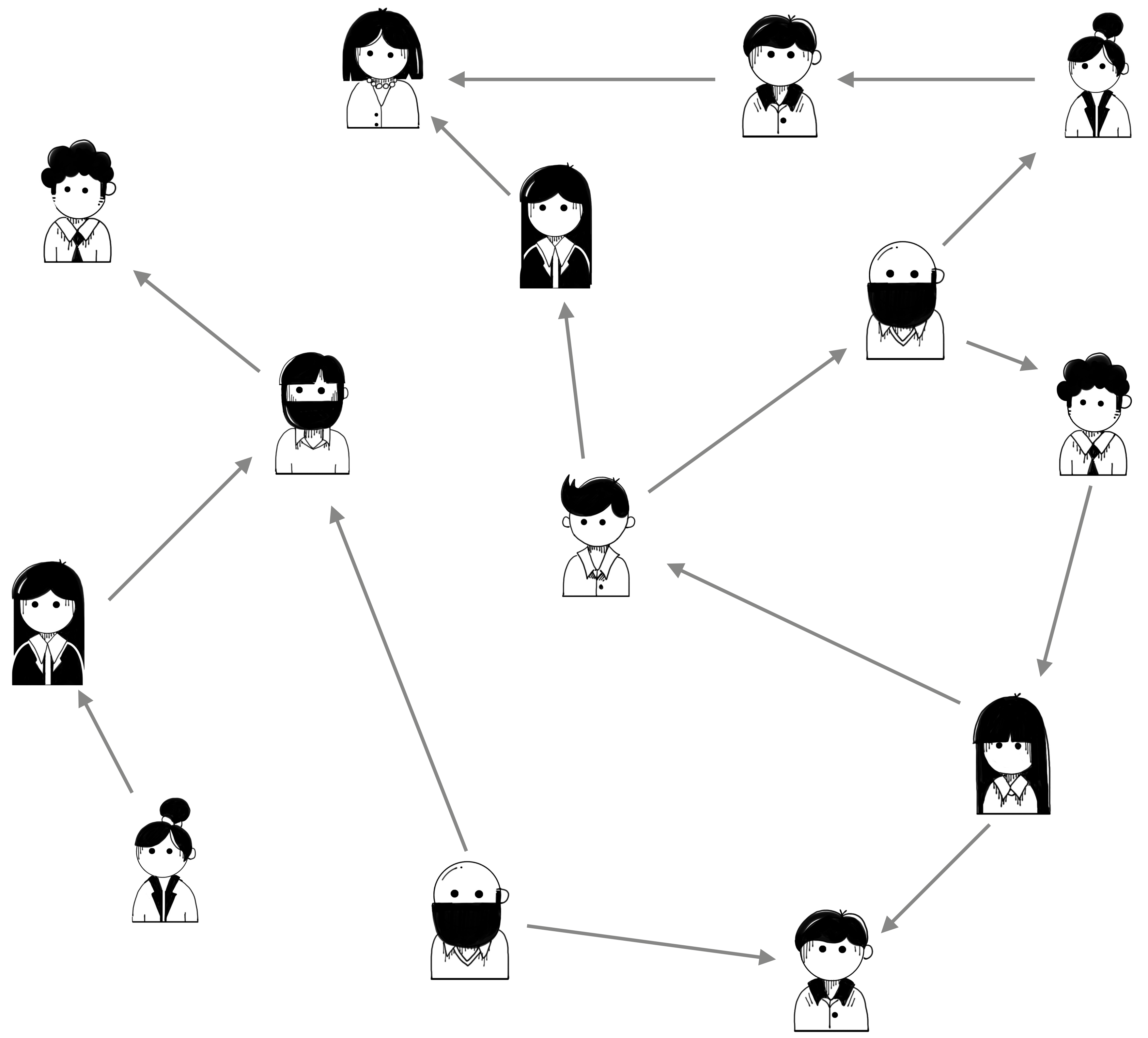
- Actors have different interests, e.g, travelling, food, politics, …
- Actors provide *User-Generated Content* (of potentially different quality).

Observations

- Actors have different interests, e.g, travelling, food, politics, …
- Actors provide *User-Generated Content* (of potentially different quality).
- Actors decide whom they want to follow.
  Directionality: followees $\neq$ followers

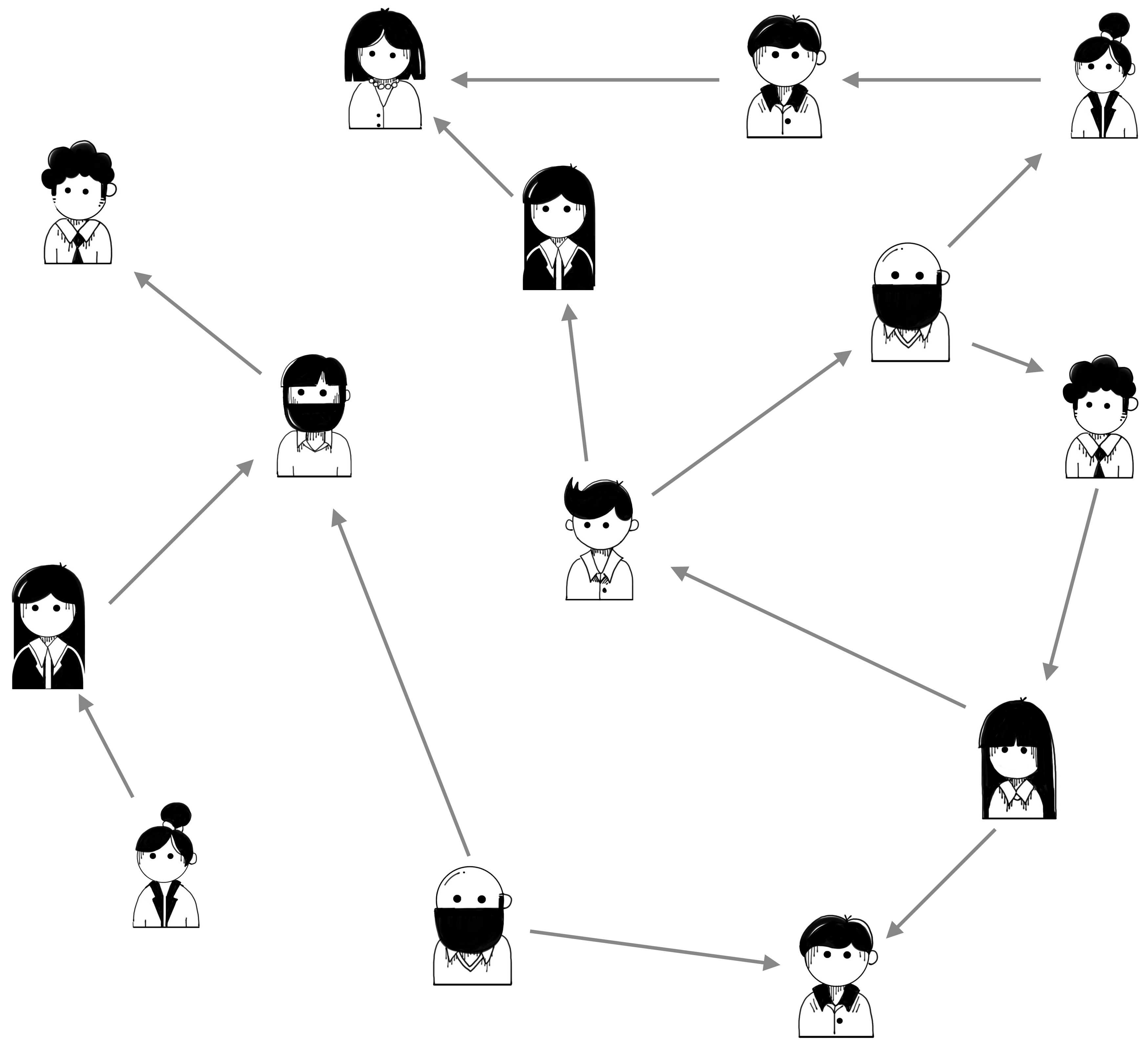
Quality-Based Network Formation Model

Directed Unweighted network $\mathcal{G}$ with $\mathcal{N} = \{1,\ldots,N\}$ agents.
$a_{ij} \in \{0,1\}$ means i follows (or not) j.



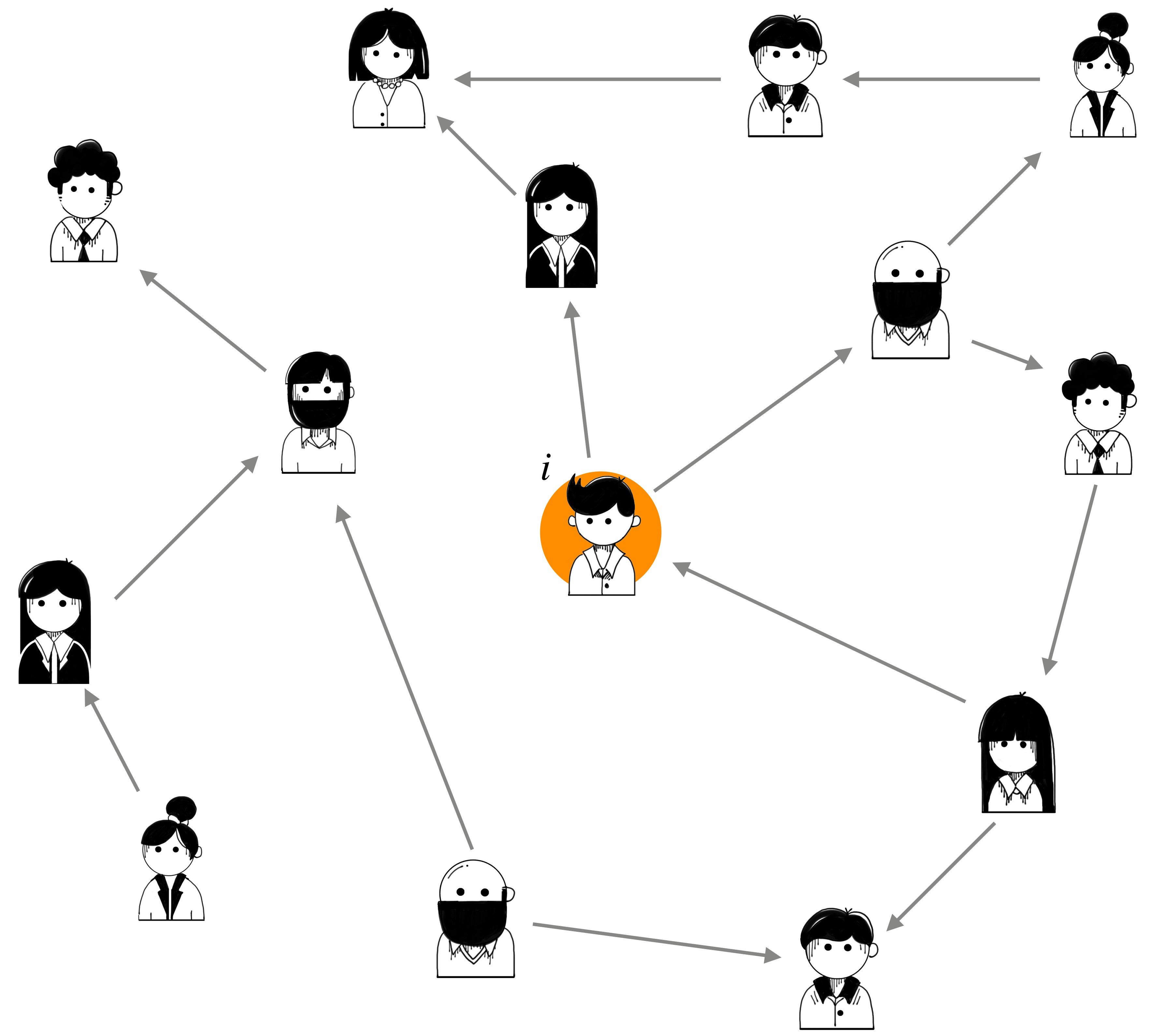
Quality-Based Network Formation Model

- **Directed Unweighted** network $\mathcal{G}$ with $\mathcal{N} = \{1,\ldots,N\}$ agents.
  $a_{ij} \in \{0,1\}$ means i follows (or not) j.
- All the agents share a common interest for a specific topic.



Quality-Based Network Formation Model

- **Directed Unweighted** network $\mathcal{G}$ with $\mathcal{N} = \{1,\ldots,N\}$ agents. $a_{ij} \in \{0,1\}$ means i follows (or not) j.
- All the agents share a common interest for a specific topic.
- Each agent i is attributed with a **quality parameter** $q_i \in [0,1]$ describing the likelihood her content will be liked by others. Wlog, we assume $q_1 > q_2 > \ldots > q_N$.

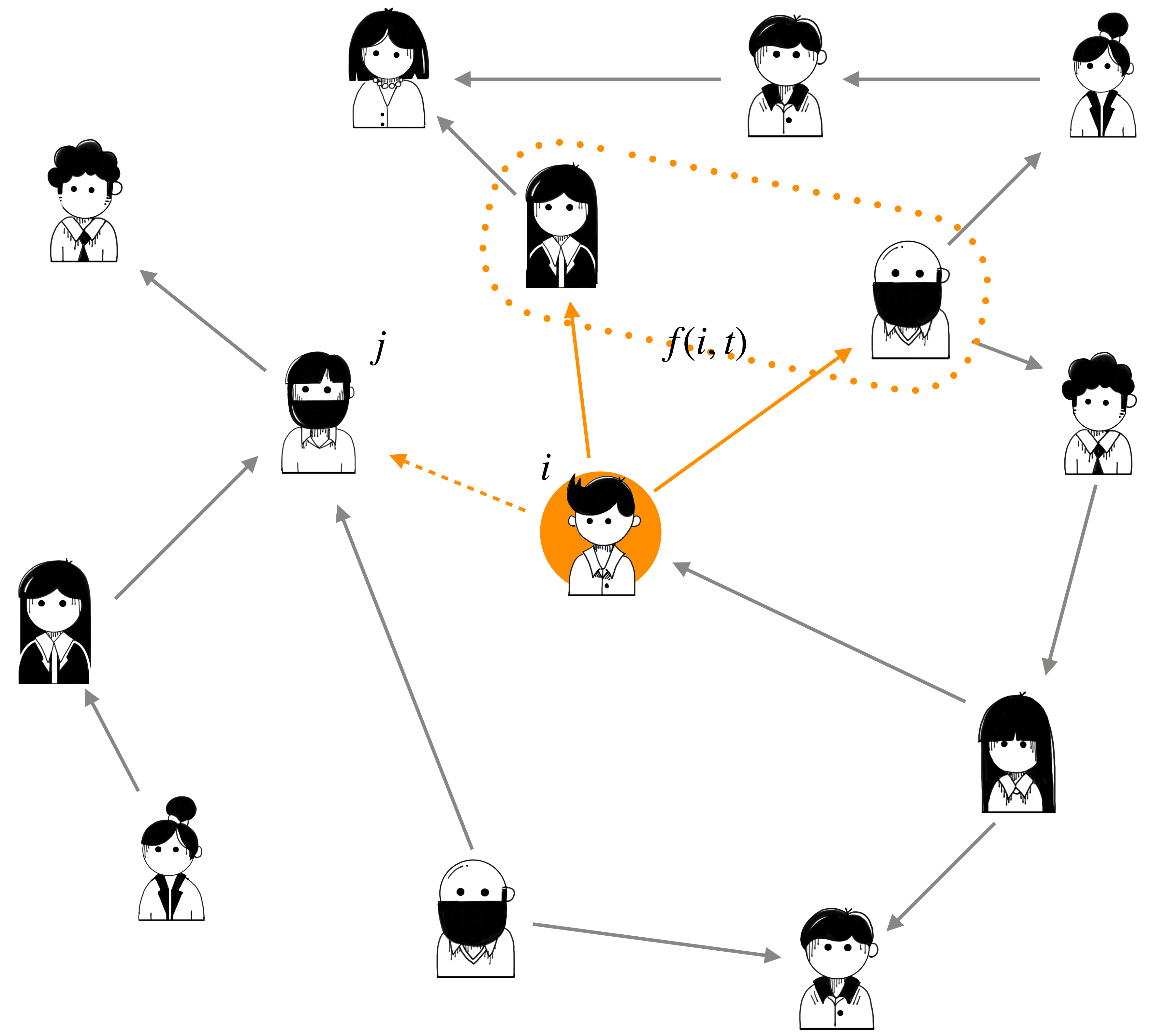


Quality-Based Network Formation Model

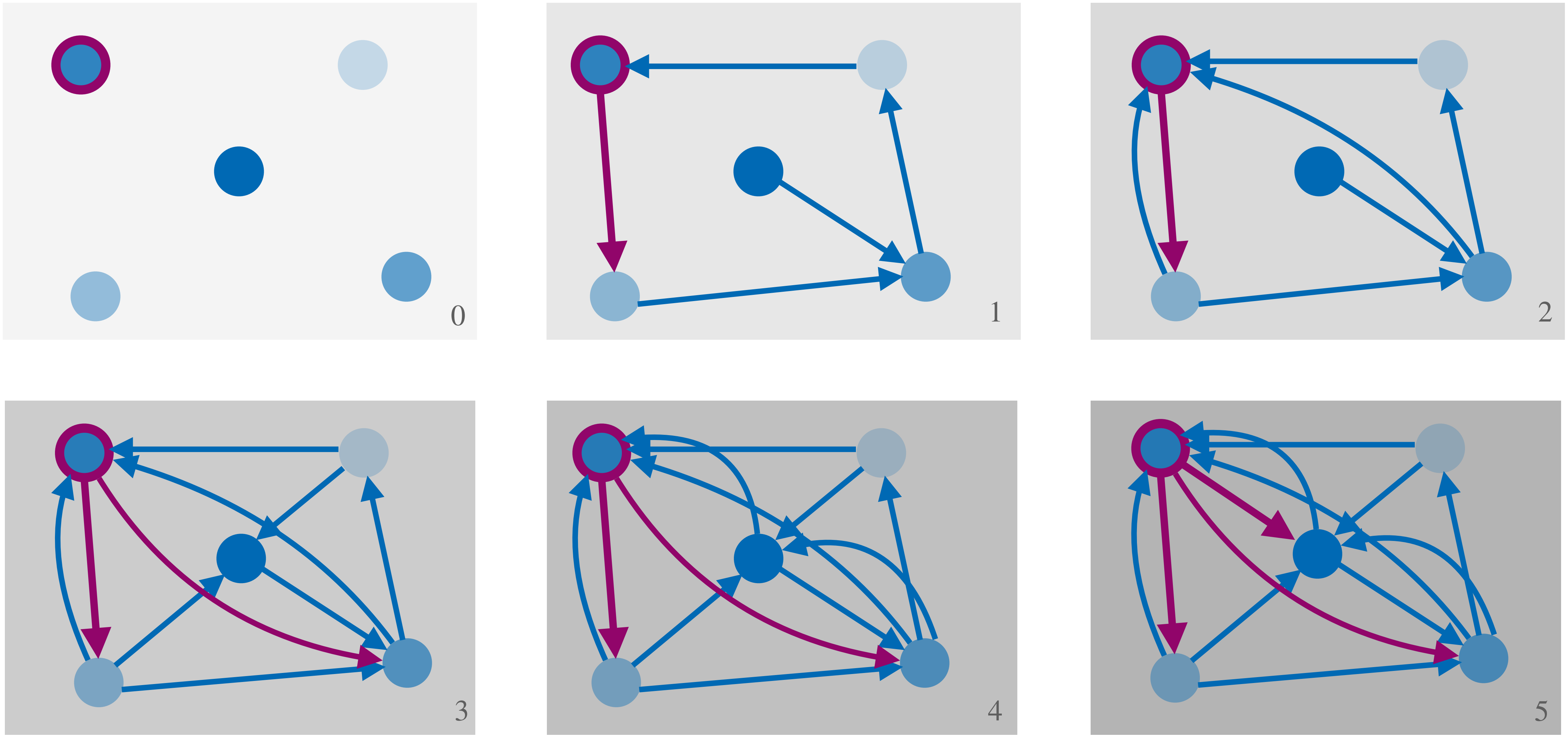
- *Directed Unweighted* network $\mathcal{G}$ with $\mathcal{N} = \{1,\ldots,N\}$ agents. $a_{ij} \in \{0,1\}$ means i follows (or not) j.
- All the agents share a common interest for a specific topic.
- Each agent i is attributed with a *quality parameter* $q_i \in [0,1]$ describing the likelihood her content will be liked by others. Wlog, we assume $q_1 > q_2 > \ldots > q_N$.
- Each agent i sequentially meets another distinct agent j (*uniformly at random*) and decides whether to start following j or not according to:

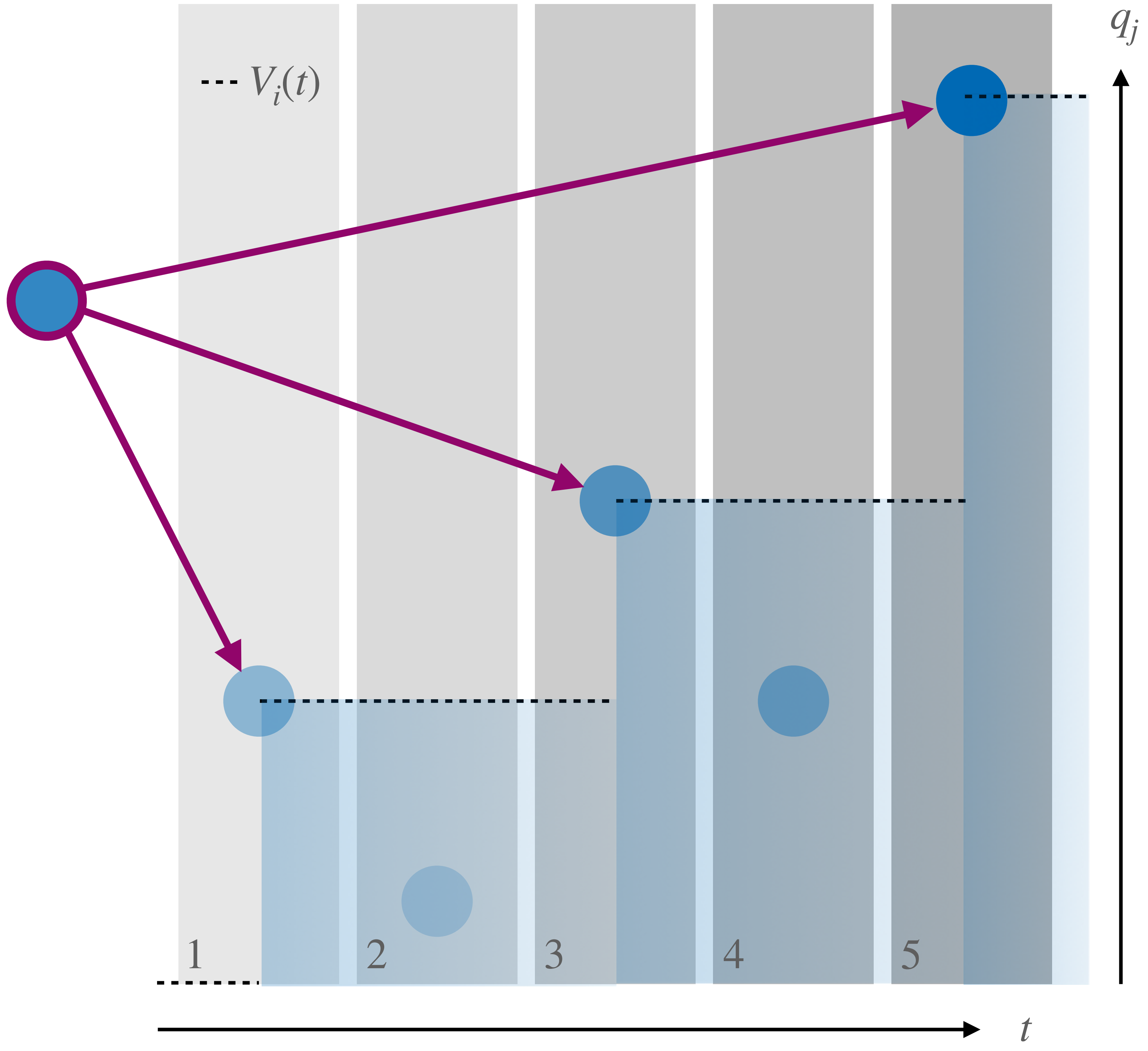
$$a_{ij}(t+1) = \begin{cases} 1, & \text{if } q_j > V_i(t) := \max_{j \in f(i,t)} q_j, \\ a_{ij}(t), & \text{otherwise,} \end{cases}$$

where $f(i,t) = \left\{ j, \ \text{s.t.} \ a_{ij}(t) = 1 \right\}$ denotes the set of i's followees.



Quality-Based Network Formation Model





RESULTS

Convergence

Theorem.

For any quality vector $\mathbf{q} = [q_1, \ldots, q_N]$, the network reaches an equilibrium almost surely.
At equilibrium, every node except node 1, follows node 1.

Numerical Simulation

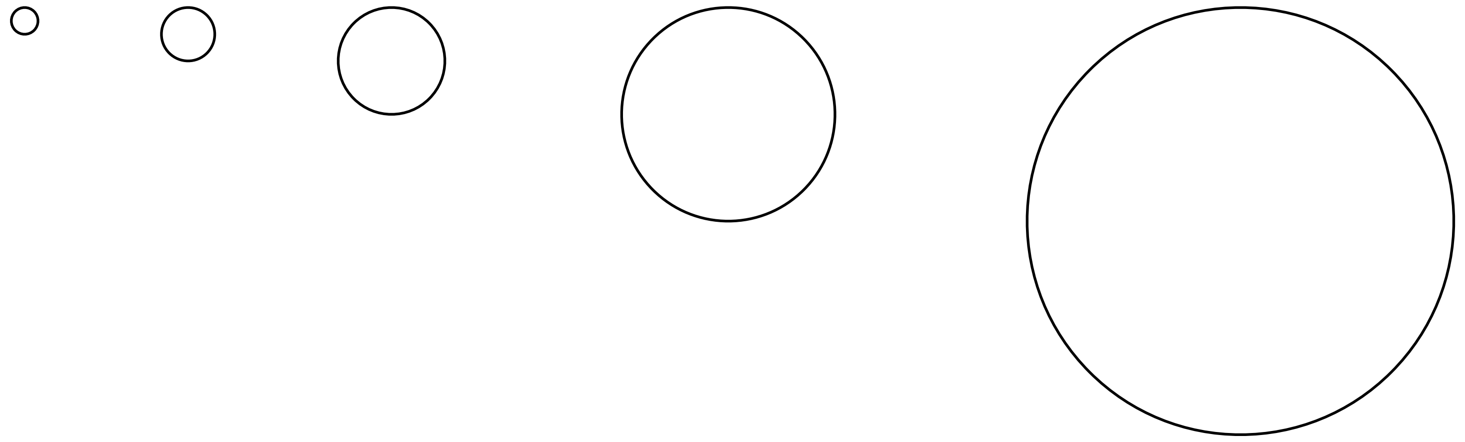
Theorem.

```
For any quality vector q = [q_1, ..., q_N], the network reaches an equilibrium almost surely.
At equilibrium, every node except node 1, follows node 1.
```

A ***meritocratic*** network formation process.

Quality

Indegree



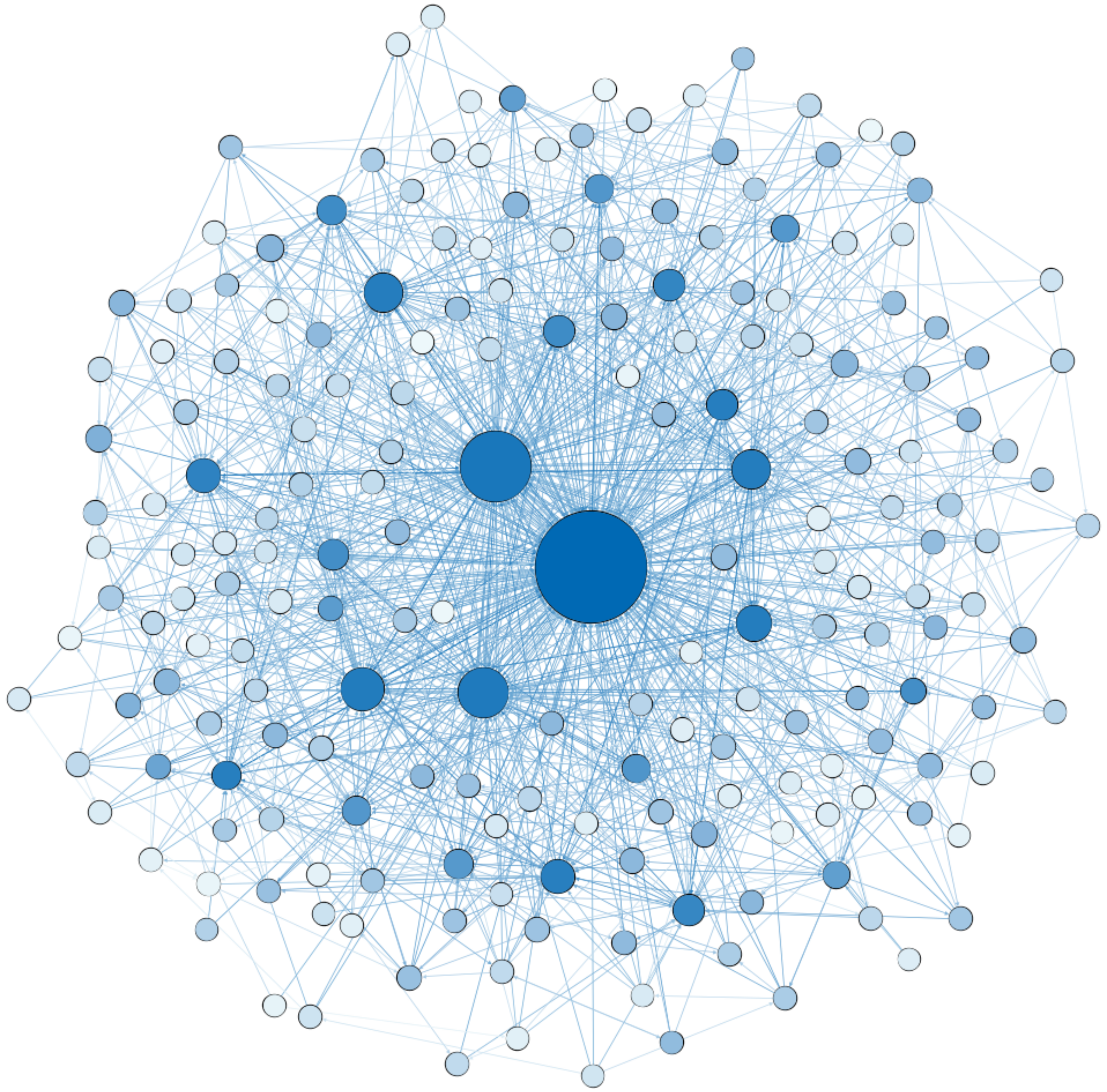


Fig. *Numerical* simulation with 200 agents. Colors are associated with quality, node's dimension is proportional to the indegree.

Indegree Distribution

Theorem.

At equilibrium, the expected indegree follows a Zipf's law: $$\mathbb{E}(d_i^{in,\star}) = \begin{cases} N-1, & \text{if } i=1, \\ \frac{N}{i}, & \text{otherwise.} \end{cases}$$

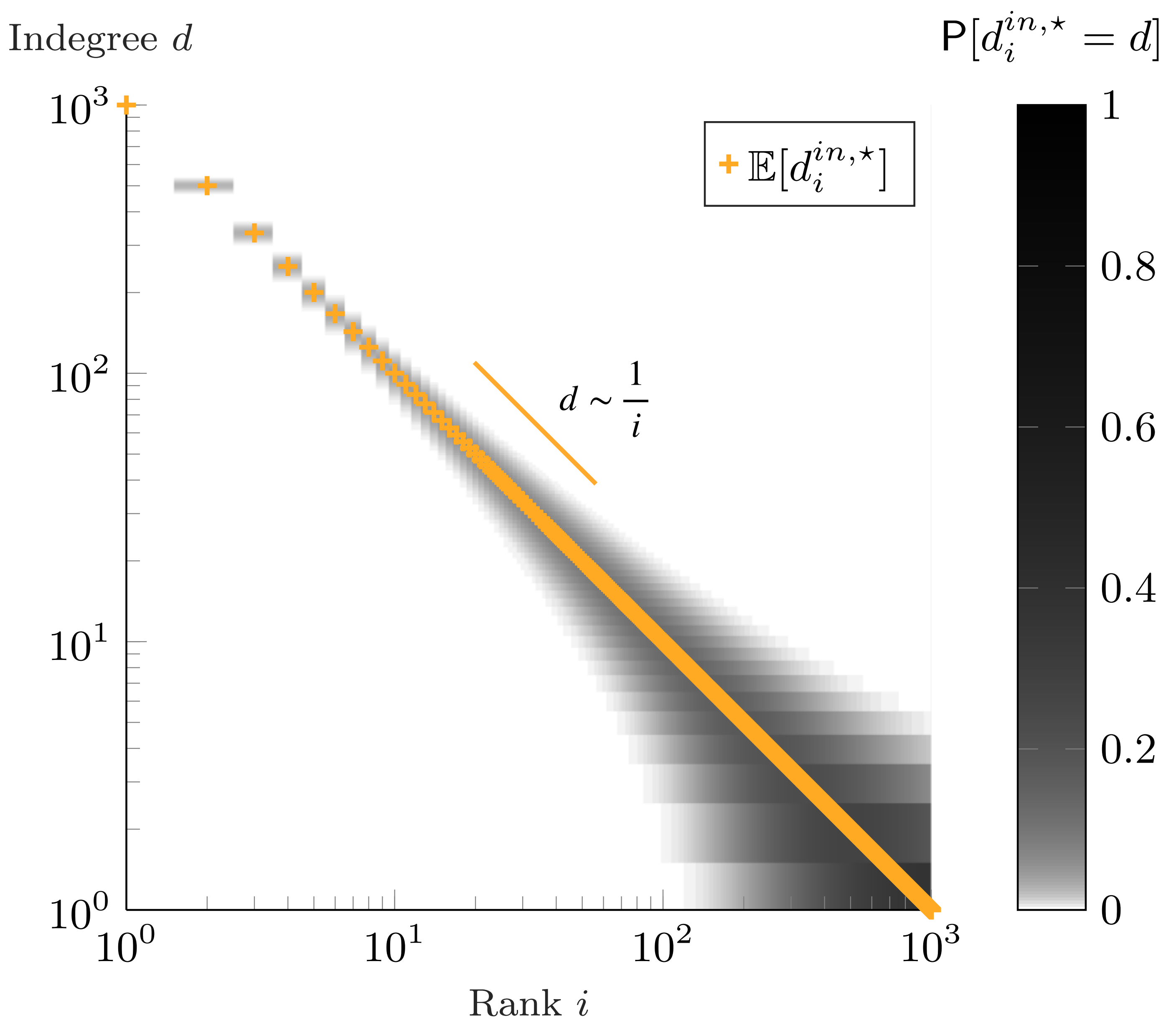


Fig. *Theoretical* probability density function of the nodes' indegree in a network of N=1000 agents.

Indegree Distribution

Theorem.

At equilibrium, the expected indegree follows a Zipf's law: $\mathbb{E}(d_i^{in,\star}) = \begin{cases} N-1, & \texttt{if } i=1, \\ \frac{N}{i}, & \texttt{otherwise.} \end{cases}$

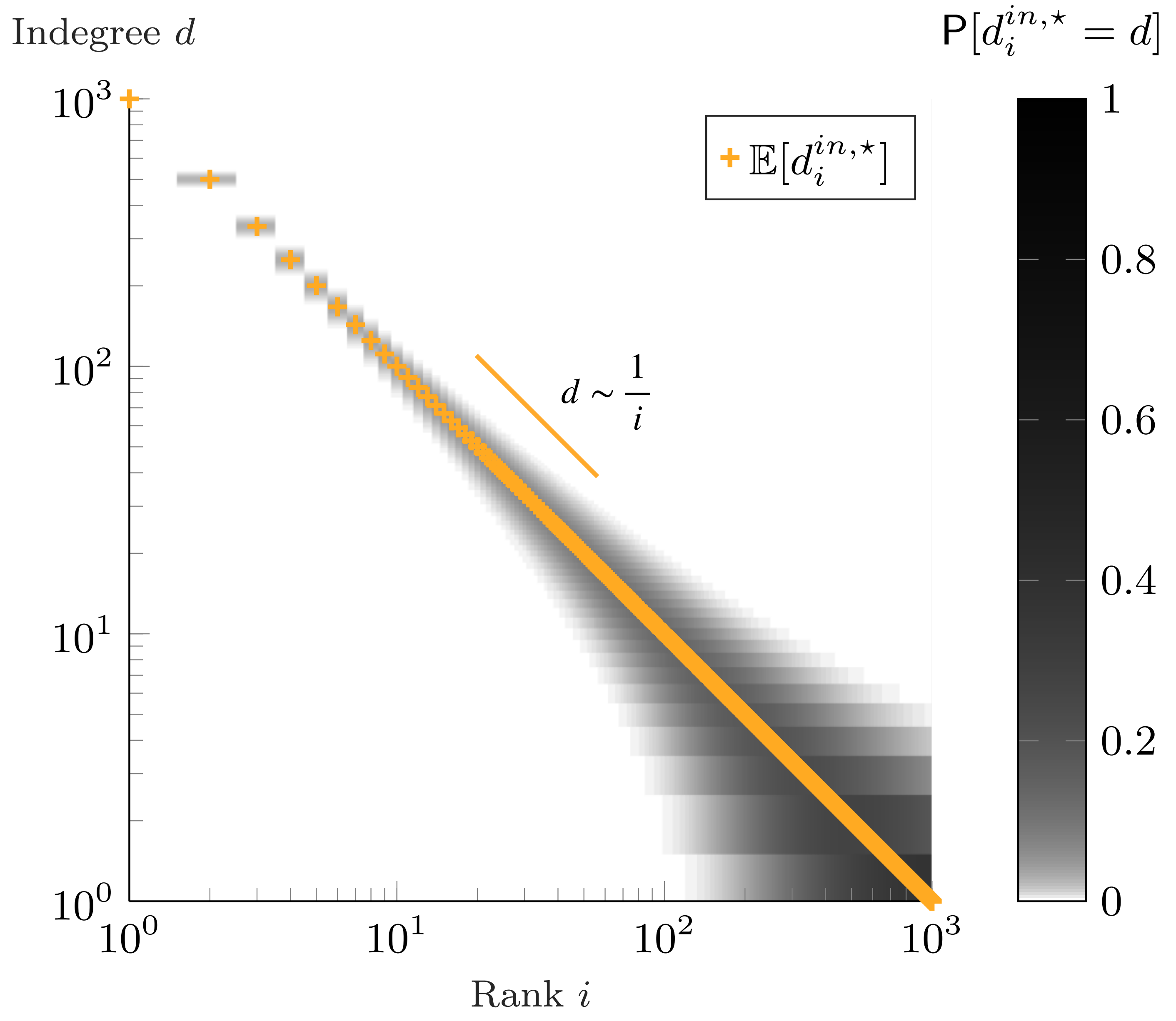


Fig. *Theoretical* probability density function of the nodes' indegree in a network of N=1000 agents.

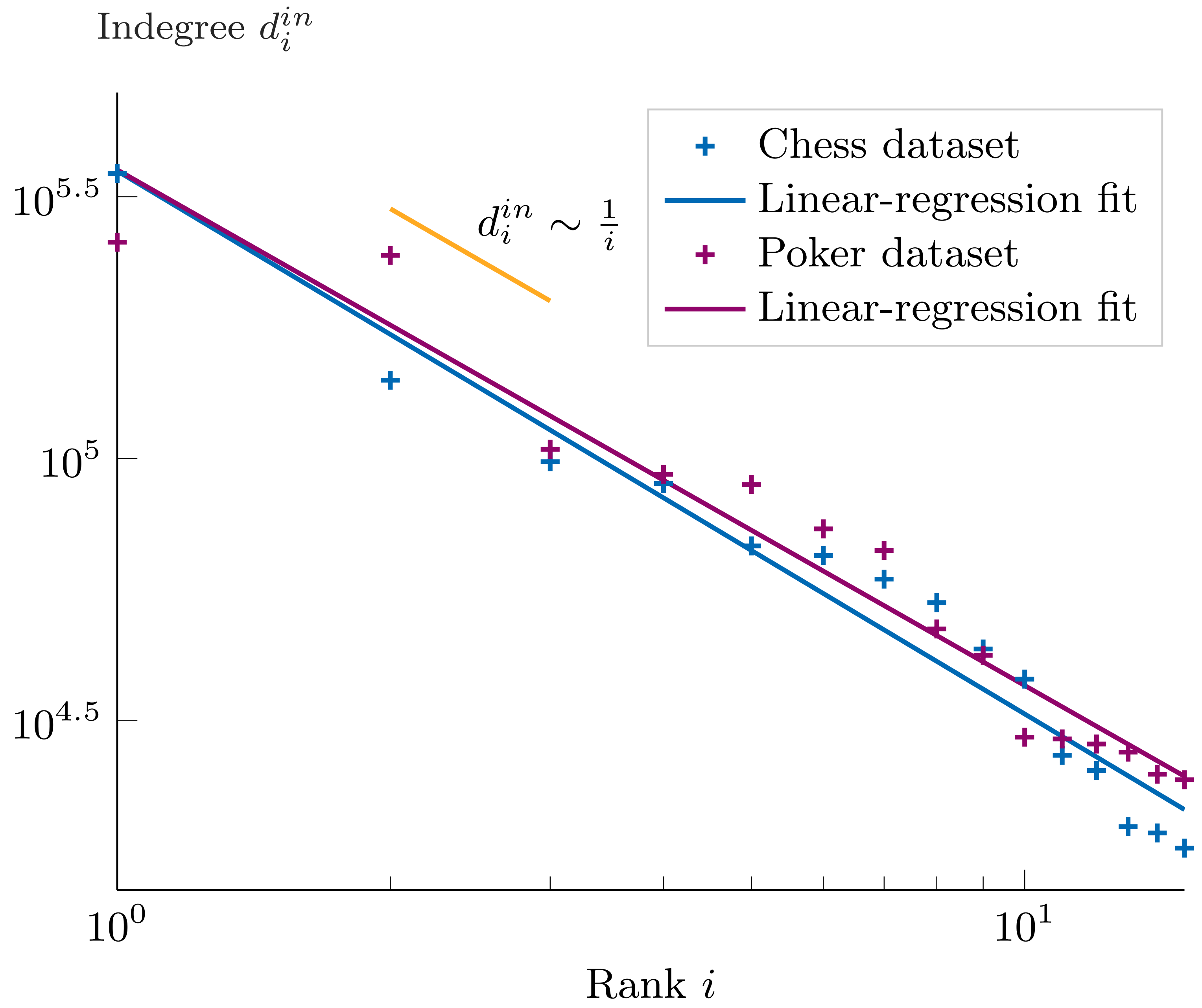


Fig. The *empirical* results from our collected Twitch data-sets are aligned with the theoretical prediction.

Indegree Distribution

Theorem.

At equilibrium, the expected indegree follows a Zipf's law: $\mathbb{E}(d_i^{in,\star}) = \begin{cases} N-1, & \text{if } i=1, \\ \frac{N}{i}, & \text{otherwise.} \end{cases}$

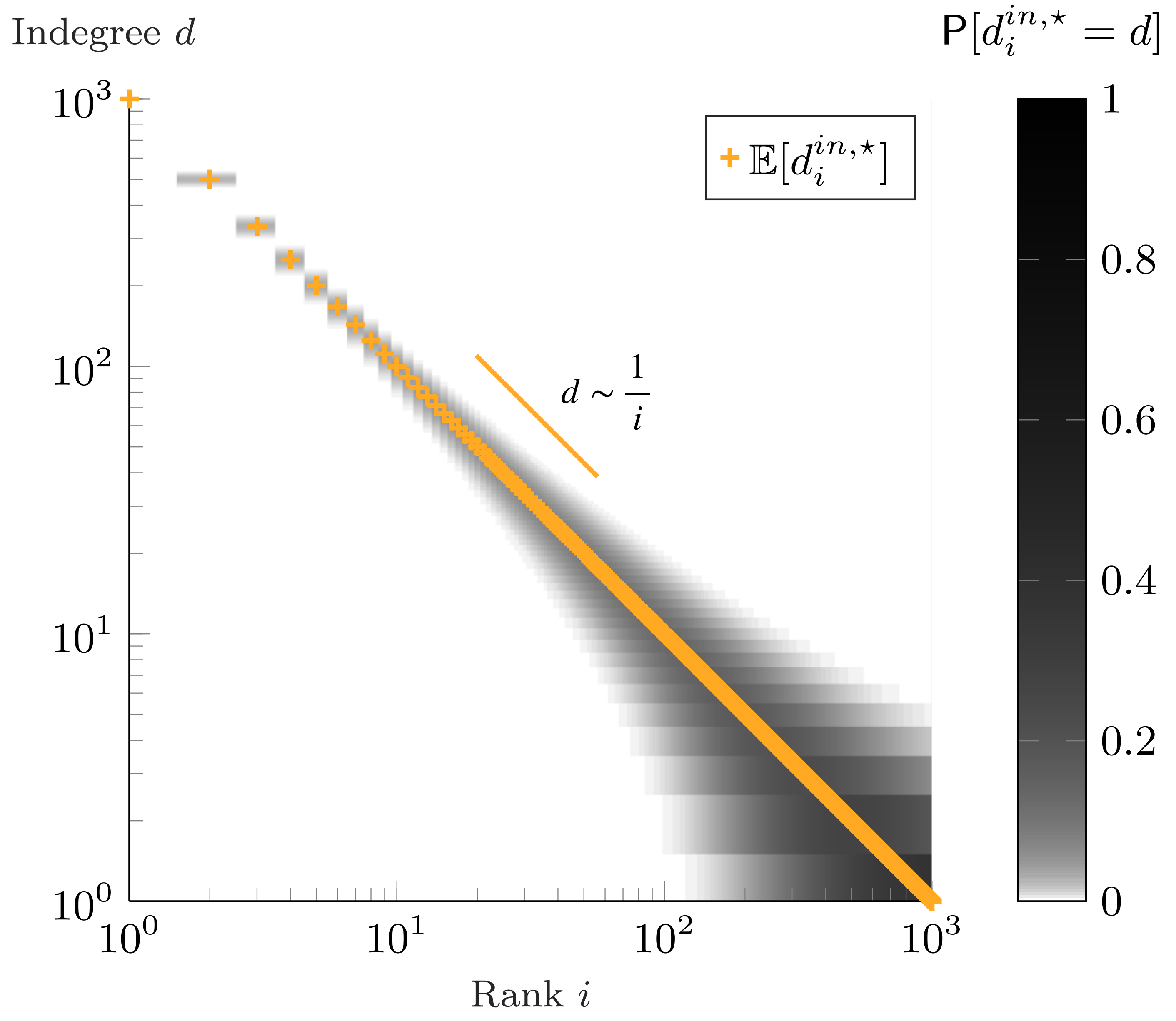


Fig. *Theoretical* probability density function of the nodes' indegree in a network of N=1000 agents.

$$\mathsf{P}[d^{in,\star} = d] = \frac{1}{N}\sum_{i=1}^{N} \mathsf{P}[d_i^{in,\star} = d]$$

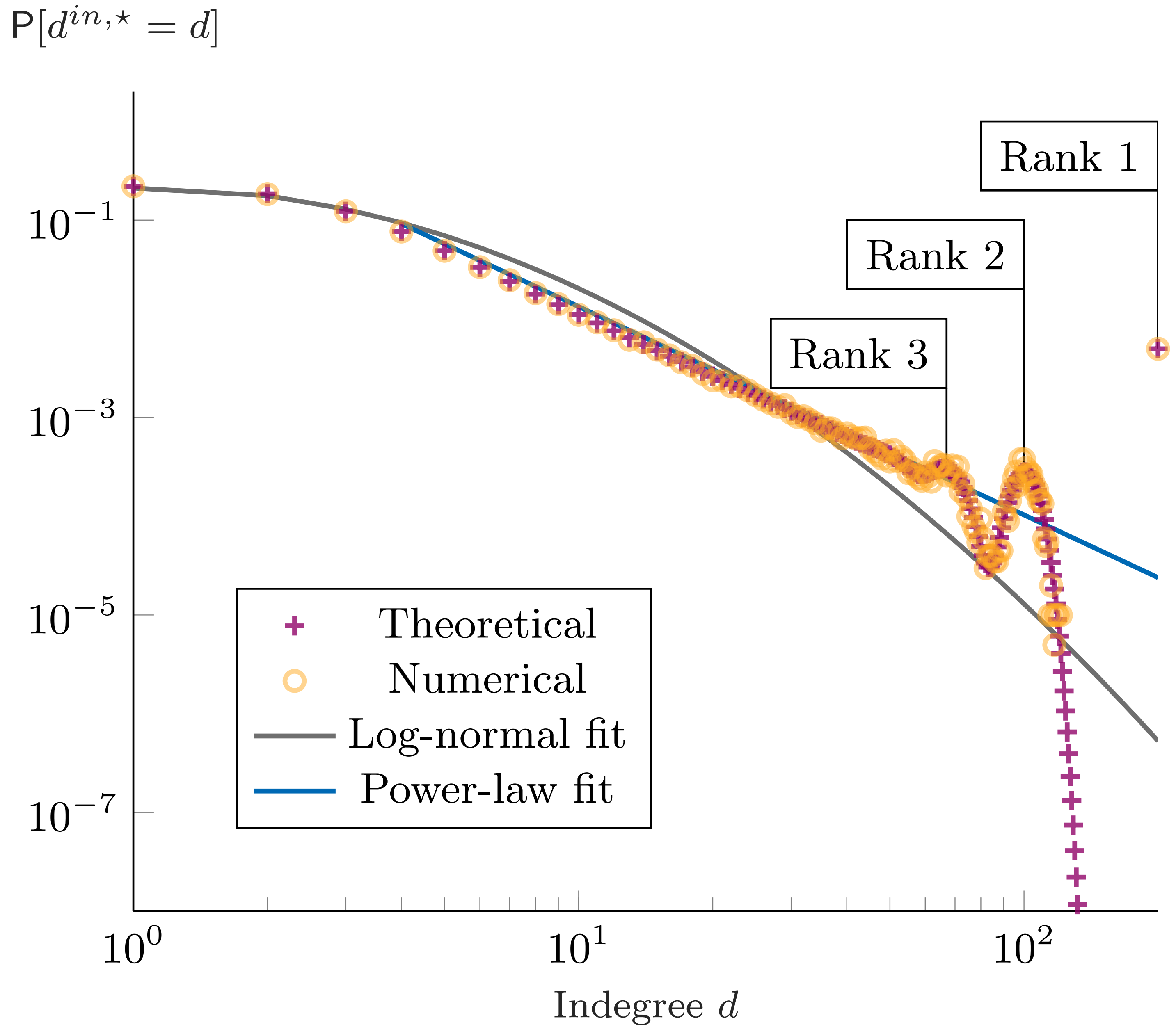


Fig. *Theoretical* Probability density function of the *average* nodes' indegree in a network of N=1000 agents.

Audience Overlap

Definition.

We define the Audience Overlap index as follows:

$$O(i,j) := \frac{\left|F(i) \cap F(j)\right|}{\left|F(i)\right|} \in [0,1],$$

if $\left|F(i)\right| > 0$, and 0 otherwise, where $F(i)$ denotes the set of followers of i.

$F(i), |F(i)| = 4$

$F(j), |F(j)| = 16$

$F(i) \cap F(j)$, $|F(i) \cap F(j)| = 2$

$O(i,j) = 1/2$

$O(j,i) = 1/8$

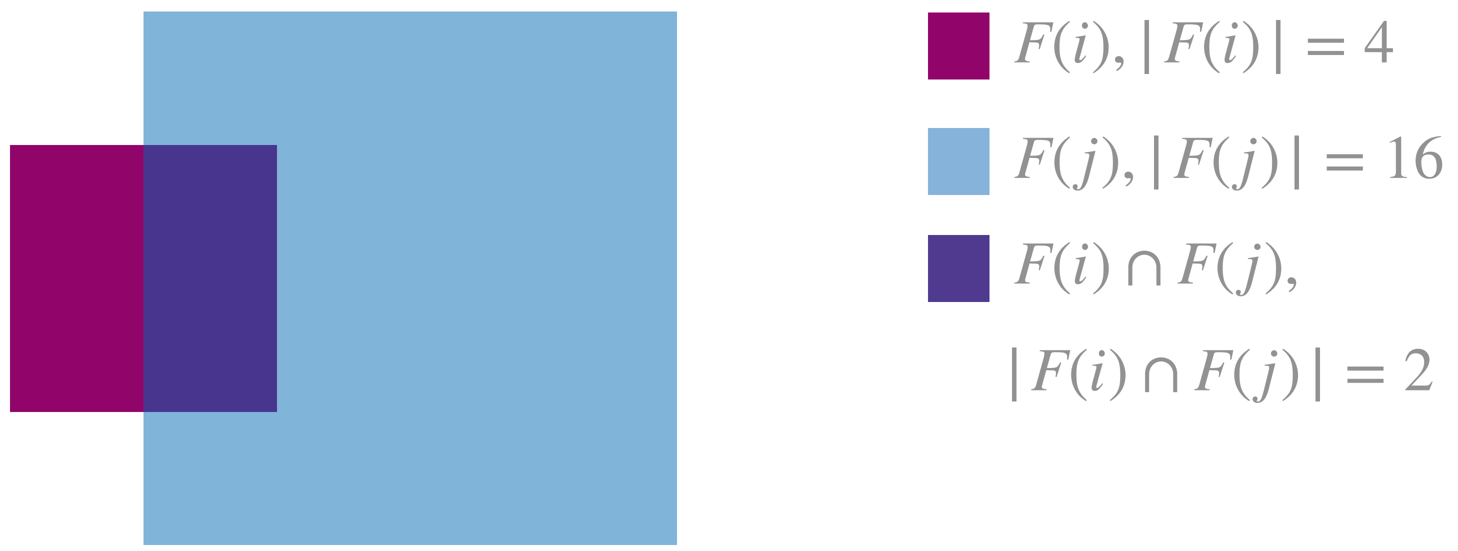
Audience Overlap

Definition.

We define the Audience Overlap index as follows:

$$O(i,j) := \frac{|F(i) \cap F(j)|}{|F(i)|} \in [0,1],$$

if $|F(i)| > 0$, and 0 otherwise, where $F(i)$ denotes the set of followers of i.



$O(i,j) = 1/2$

$O(j,i) = 1/8$

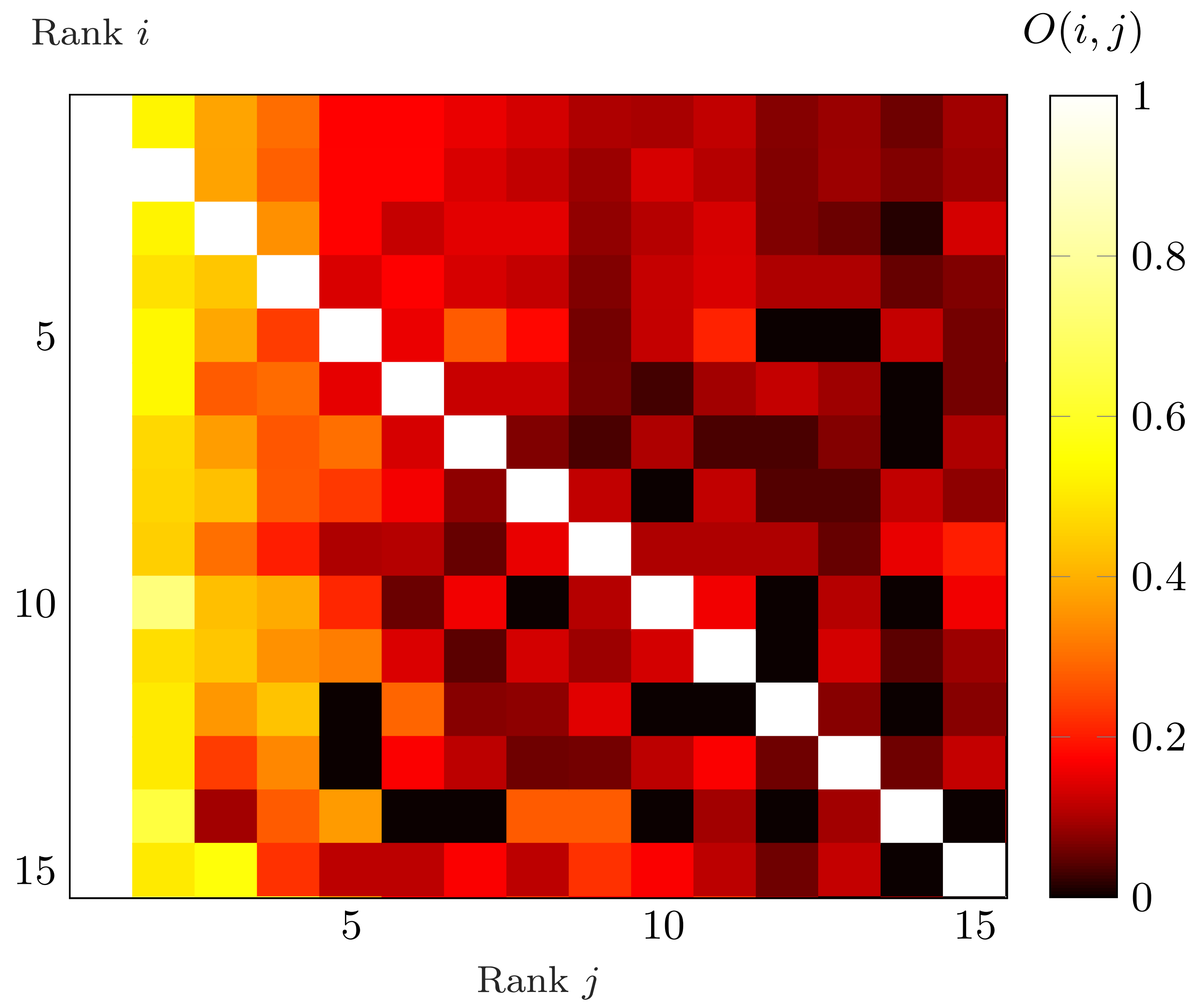


Fig. *Numerical* results on the followers' overlap among the top nodes. The result is obtained with N=1000 agents, upon reaching equilibrium.

Audience Overlap

Definition.

We define the Audience Overlap index as follows:

$$O(i,j) := \frac{\left|F(i) \cap F(j)\right|}{\left|F(i)\right|} \in [0,1],$$

if $\left|F(i)\right| > 0$, and 0 otherwise, where $F(i)$ denotes the set of followers of i.

$F(i), |F(i)| = 4$

$F(1), |F(1)| = 16$

$F(i) \cap F(1),$

$|F(i) \cap F(1)| = 4$

$O(i,1) = 1$

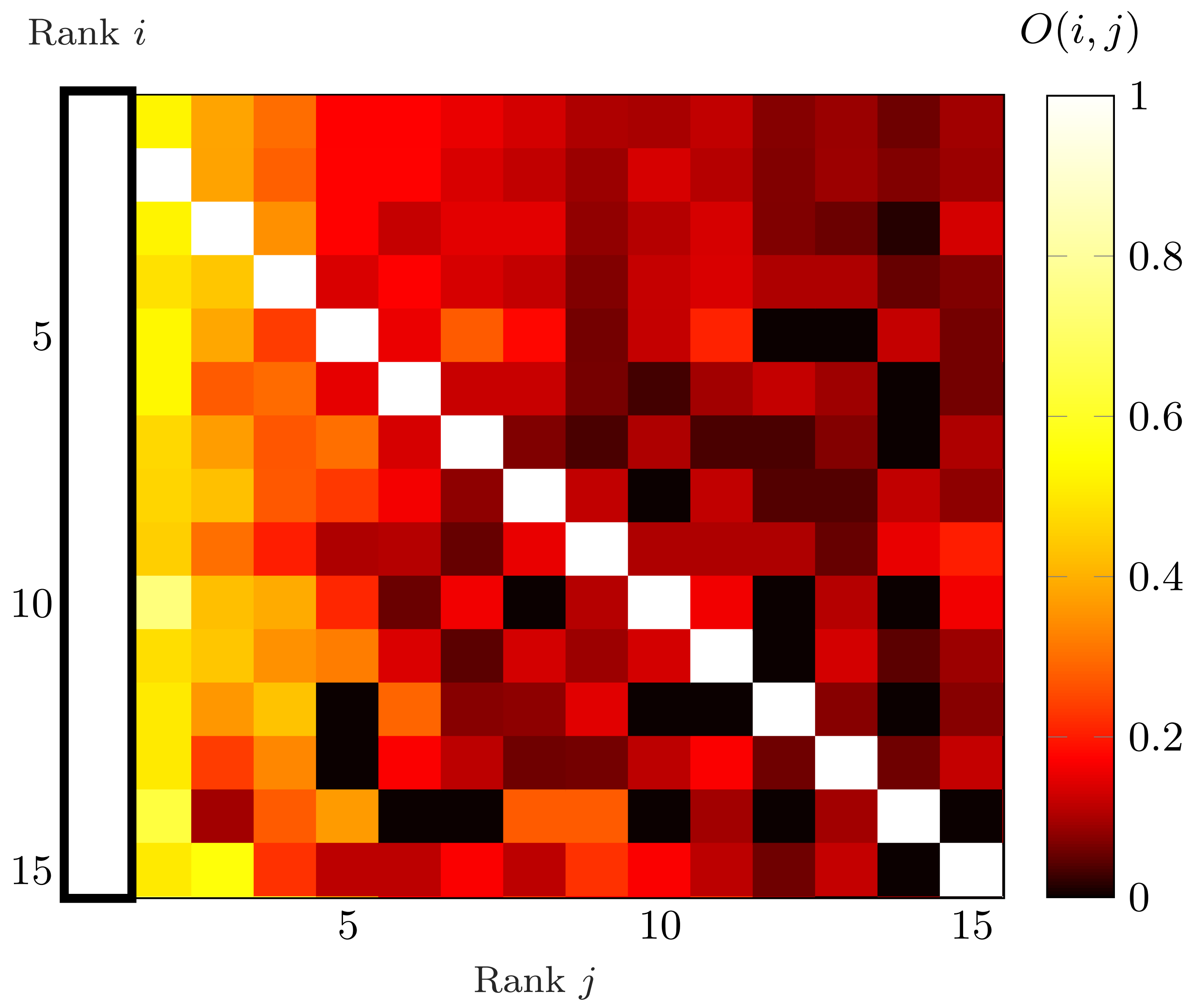


Fig. *Numerical* results on the followers' overlap among the top nodes. The result is obtained with N=1000 agents, upon reaching equilibrium.

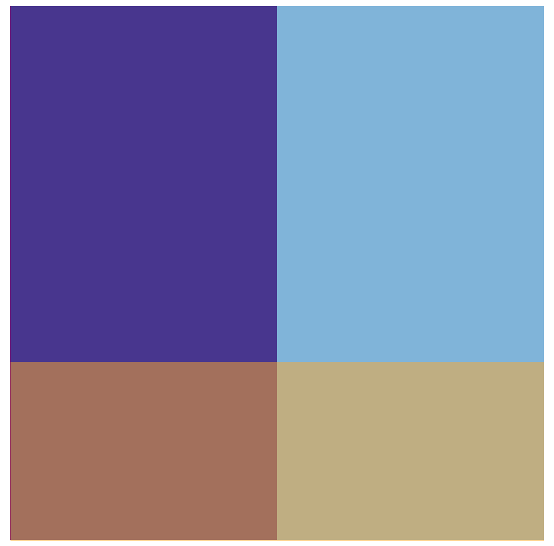
Audience Overlap

Definition.

We define the Audience Overlap index as follows:

$$O(i,j) := \frac{\left|F(i) \cap F(j)\right|}{\left|F(i)\right|} \in [0,1],$$

if $\left|F(i)\right| > 0$, and 0 otherwise, where $F(i)$ denotes the set of followers of i.



$F(1), |F(1)| = 12$

$F(2), |F(2)| = 6$

$F(3), |F(3)| = 4$

$O(1,2) = 1/2$

$O(1,3) = 1/3\ldots$

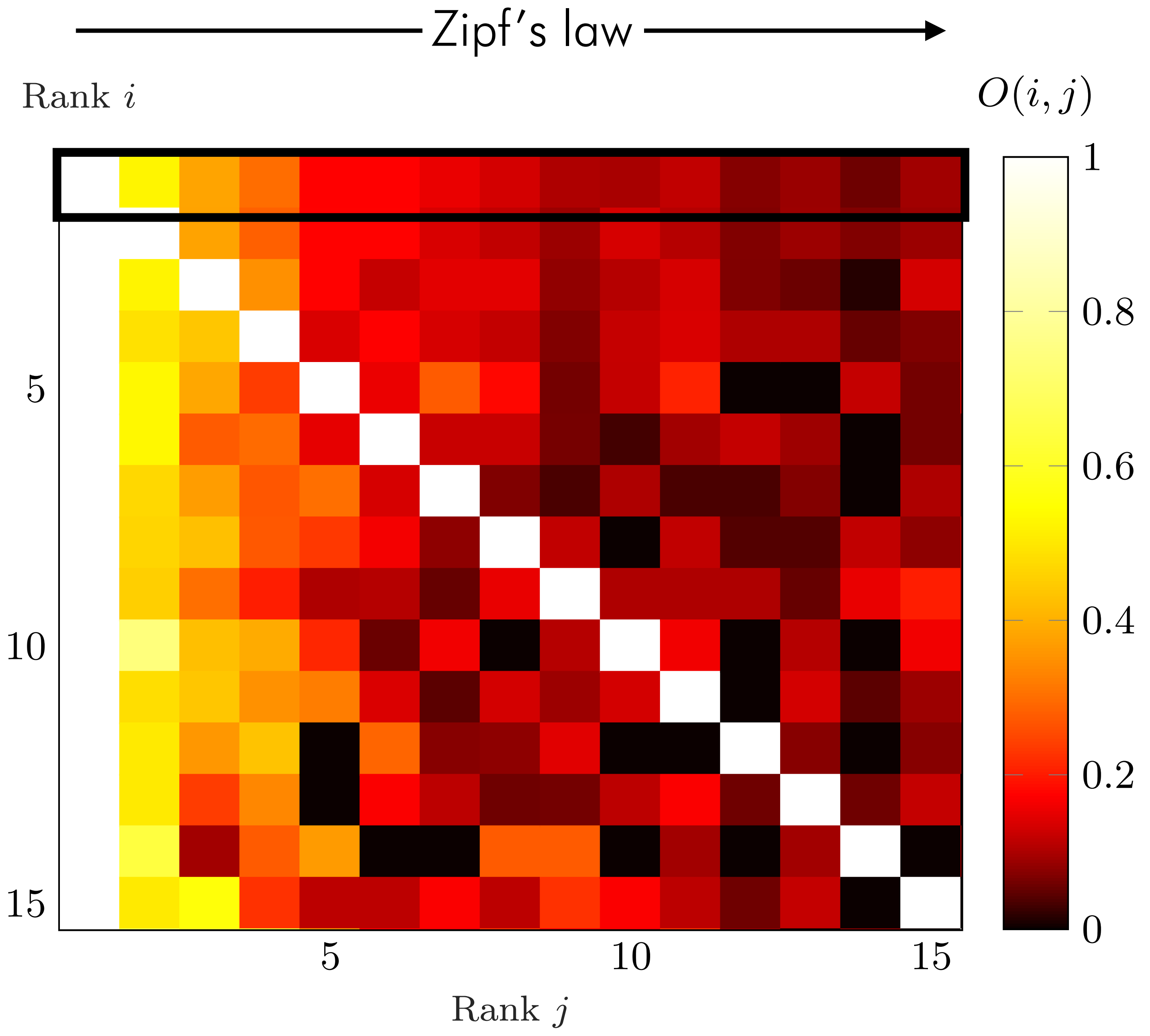


Fig. *Numerical* results on the followers' overlap among the top nodes. The result is obtained with N=1000 agents, upon reaching equilibrium.

Audience Overlap

Definition.

We define the Audience Overlap index as follows:

$$O(i,j) := \frac{\left|F(i)\cap F(j)\right|}{\left|F(i)\right|} \in [0,1],$$

if $\left|F(i)\right| > 0$, and 0 otherwise, where $F(i)$ denotes the set of followers of i.

$O(1,2) = 1/2$

$O(1,3) = 1/3\ldots$

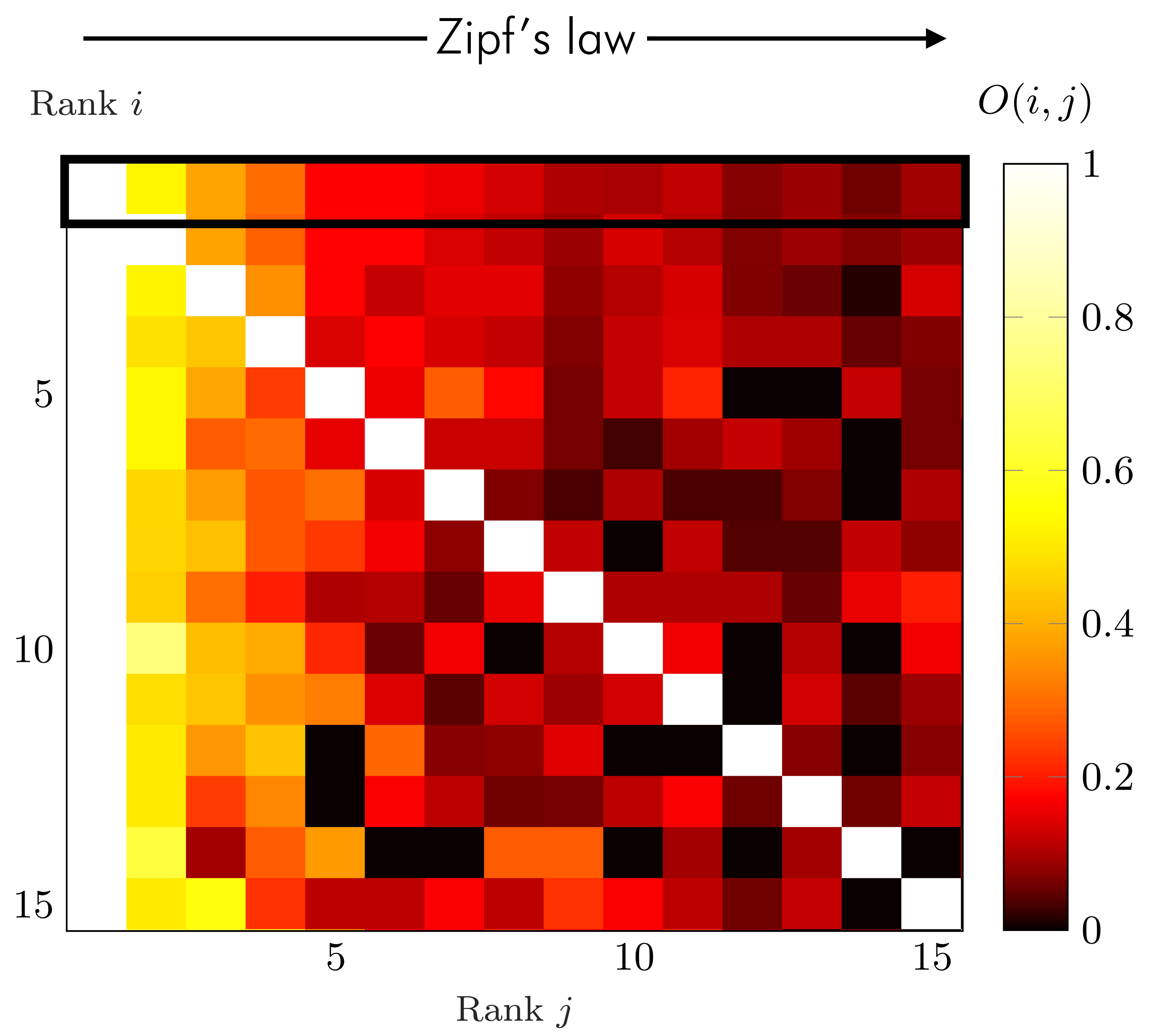


Fig. *Numerical* results on the followers' overlap among the top nodes. The result is obtained with N=1000 agents, upon reaching equilibrium.

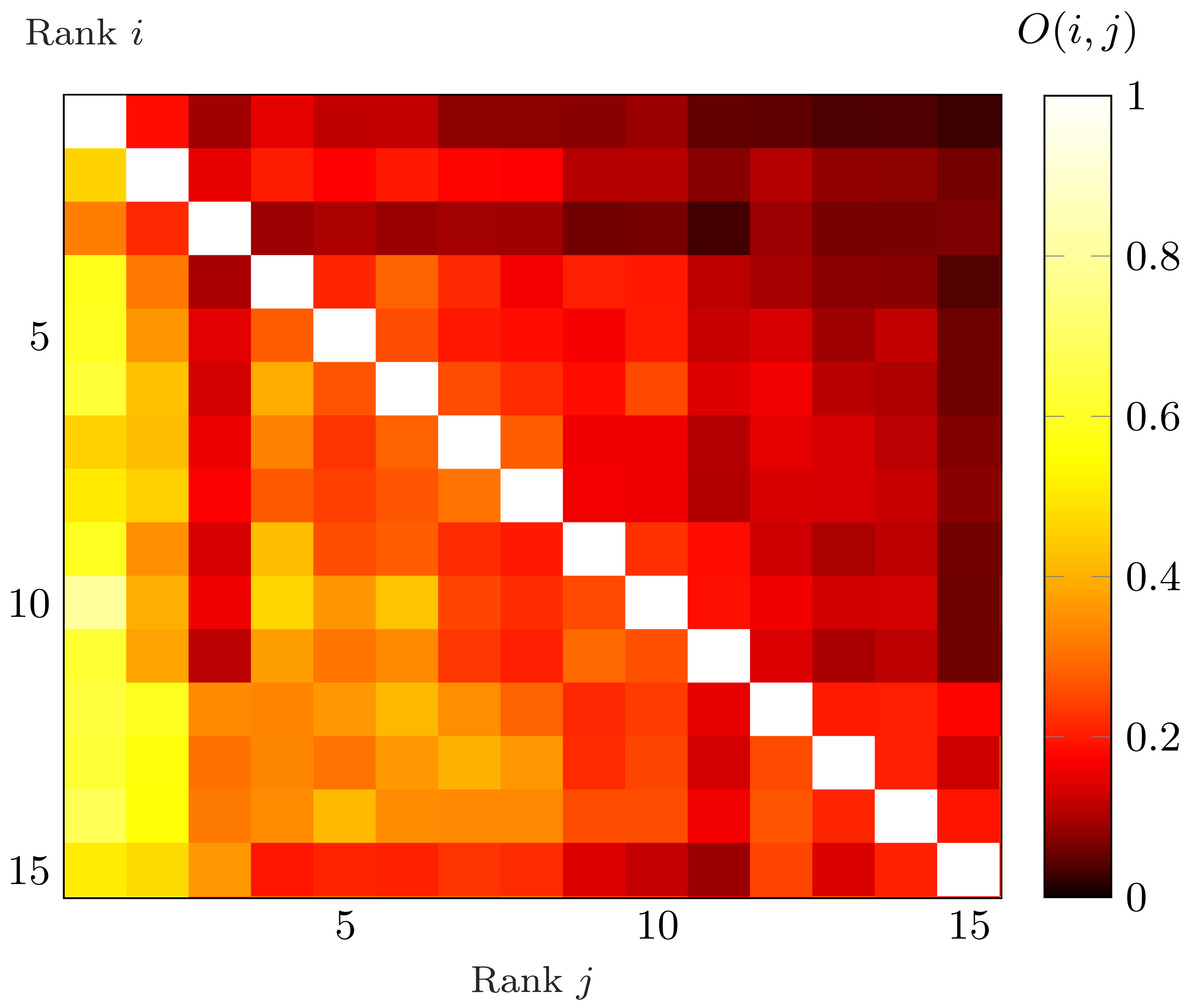


Fig. *Empirical* results from our collected data-set.

Outdegree Distribution

- The outdegree probability density function is *homogeneous* for all the nodes.

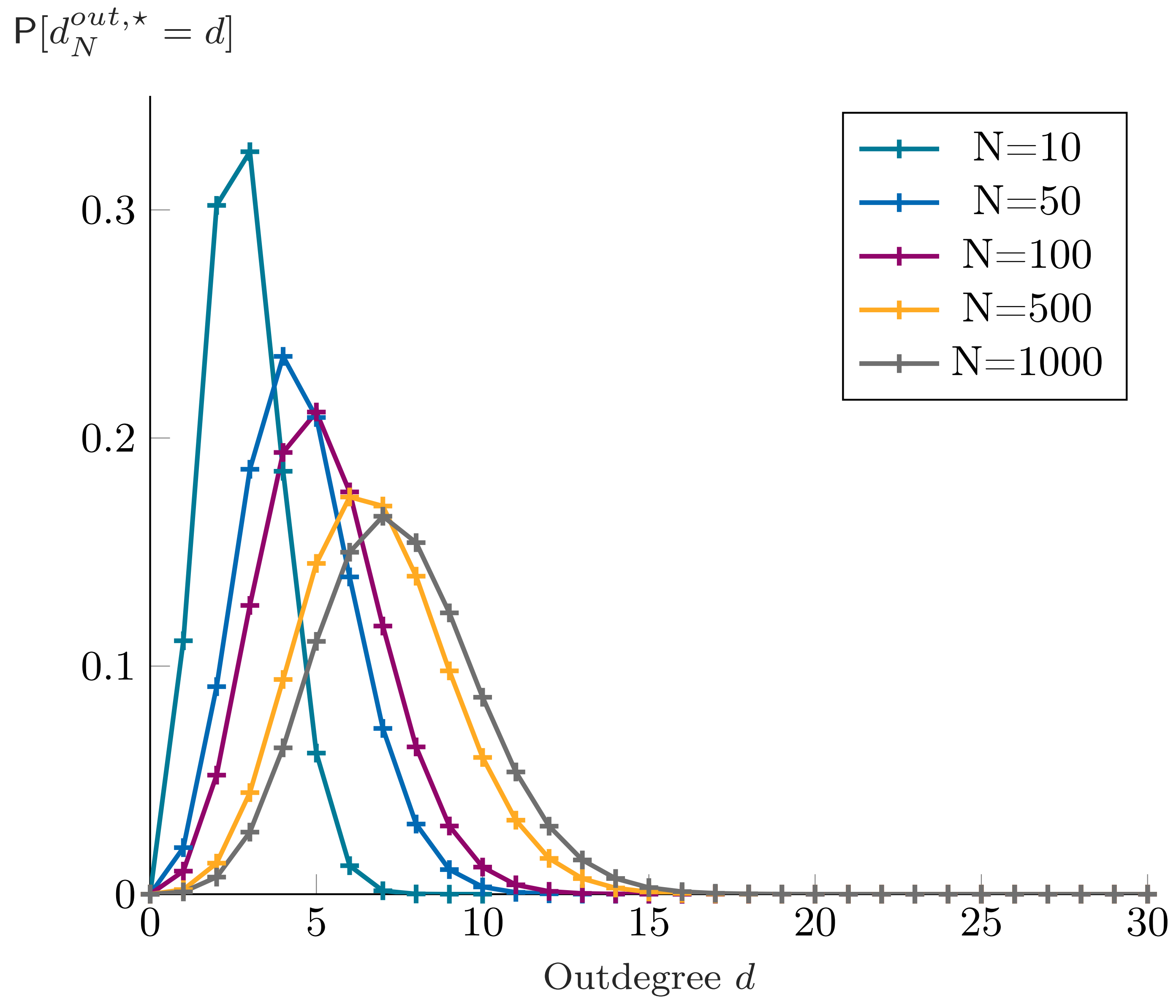


Fig. *Theoretical* outdegree distribution for different network size.

Outdegree Distribution

The outdegree probability density function is **homogeneous** for all the nodes.

THEOREM.

The nodes' expected outdegree in a network of N agents equals the (N-1)th harmonic number:

$$\mathbb{E}[d_N^{out,\star}] = \sum_{k=1}^{N-1} \frac{1}{k}.$$

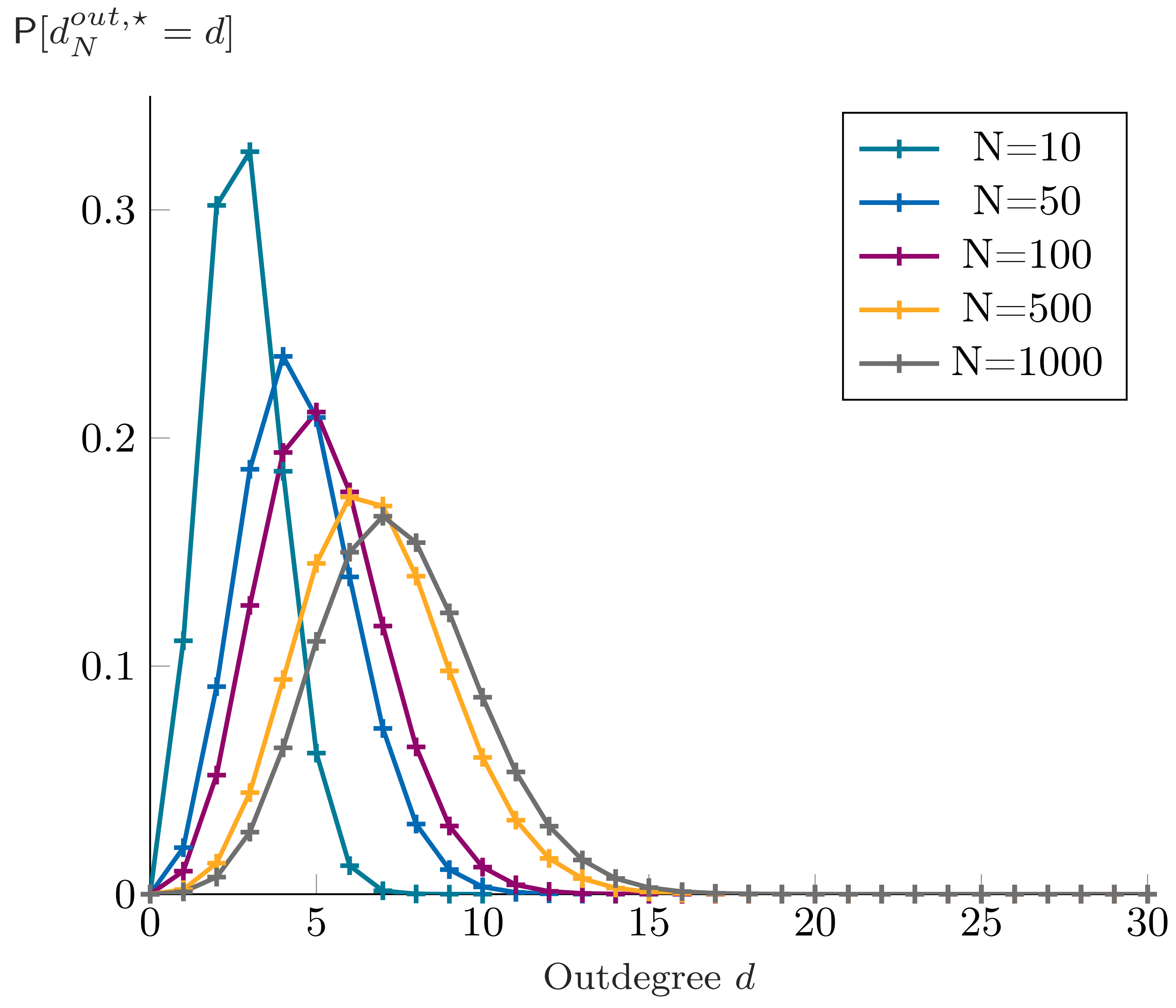


Fig. *Theoretical* outdegree distribution for different network size.

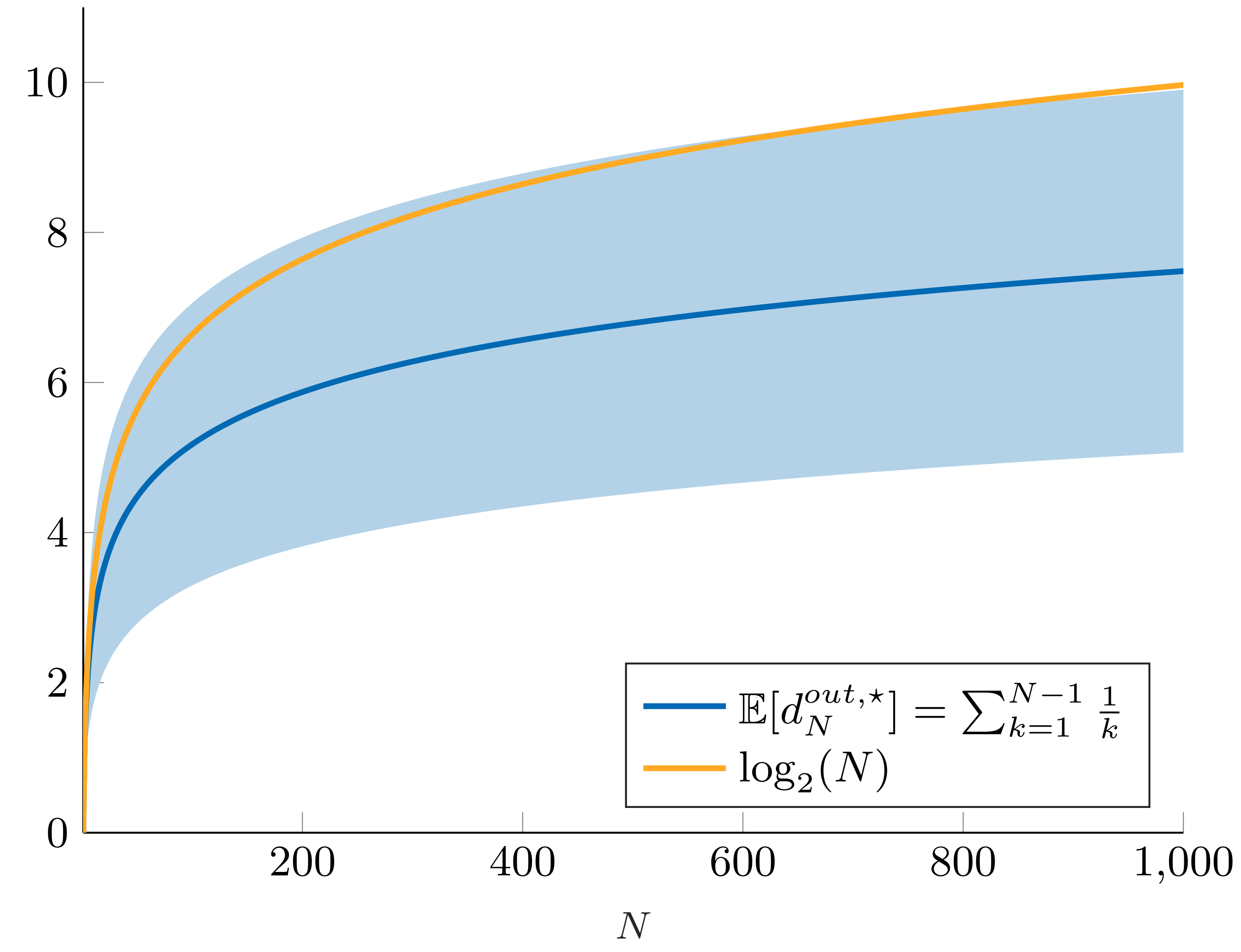


Fig. *Theoretical* expected outdegree as a function of the number of agents N.

Outdegree Distribution

The outdegree probability density function is ***homogeneous*** for all the nodes.

THEOREM.

```
The nodes' expected outdegree in a
network of N agents equals the (N-1)th
harmonic number:
```

$$\mathbb{E}[d_N^{out,\star}] = \sum_{k=1}^{N-1} \frac{1}{k}.$$

The empirical distribution is:

- concentrated in the region $d \in [0,10]$;
- but it seems to underestimate the low-outdegree nodes.

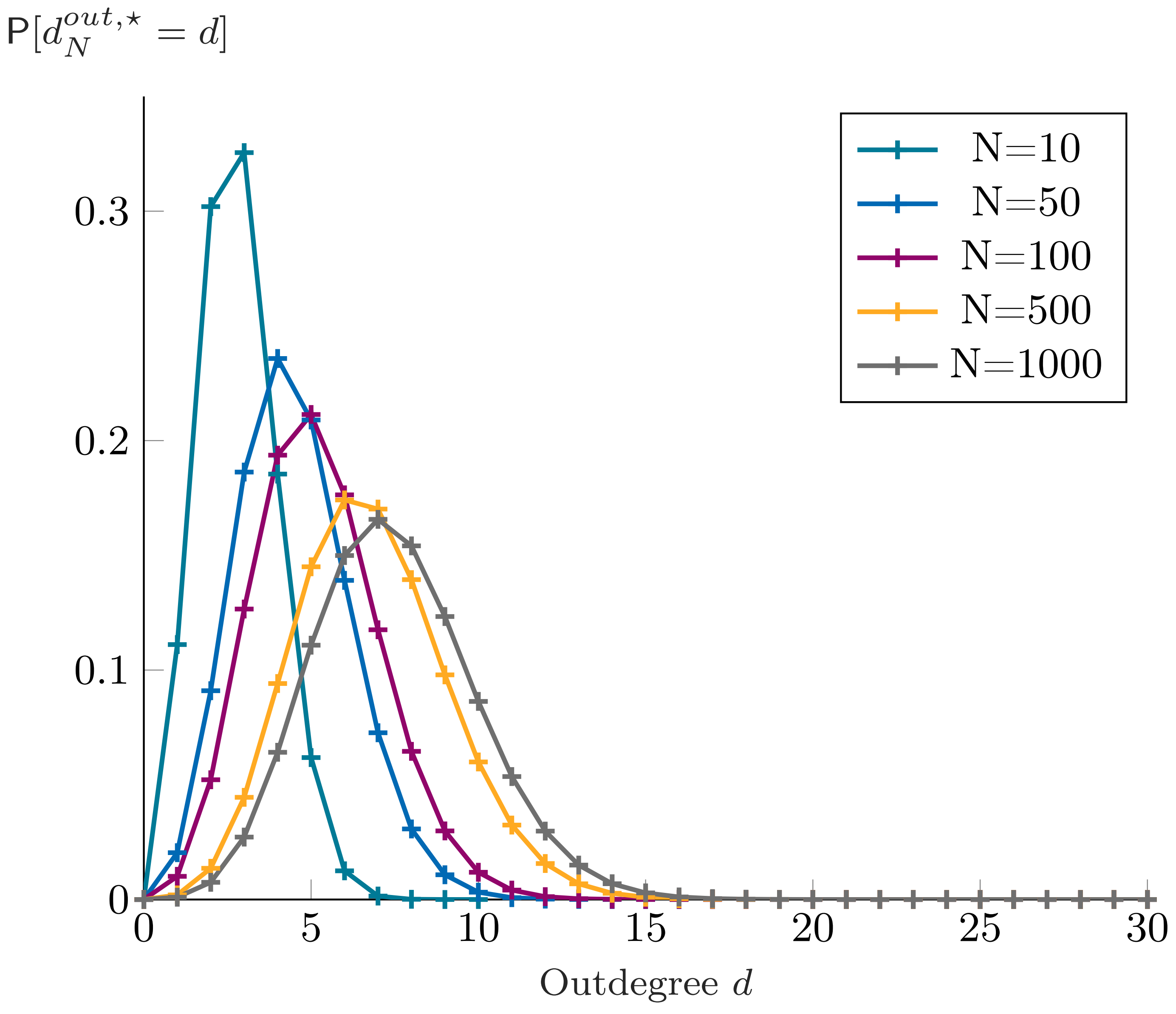


Fig. *Theoretical* outdegree distribution for different network size.

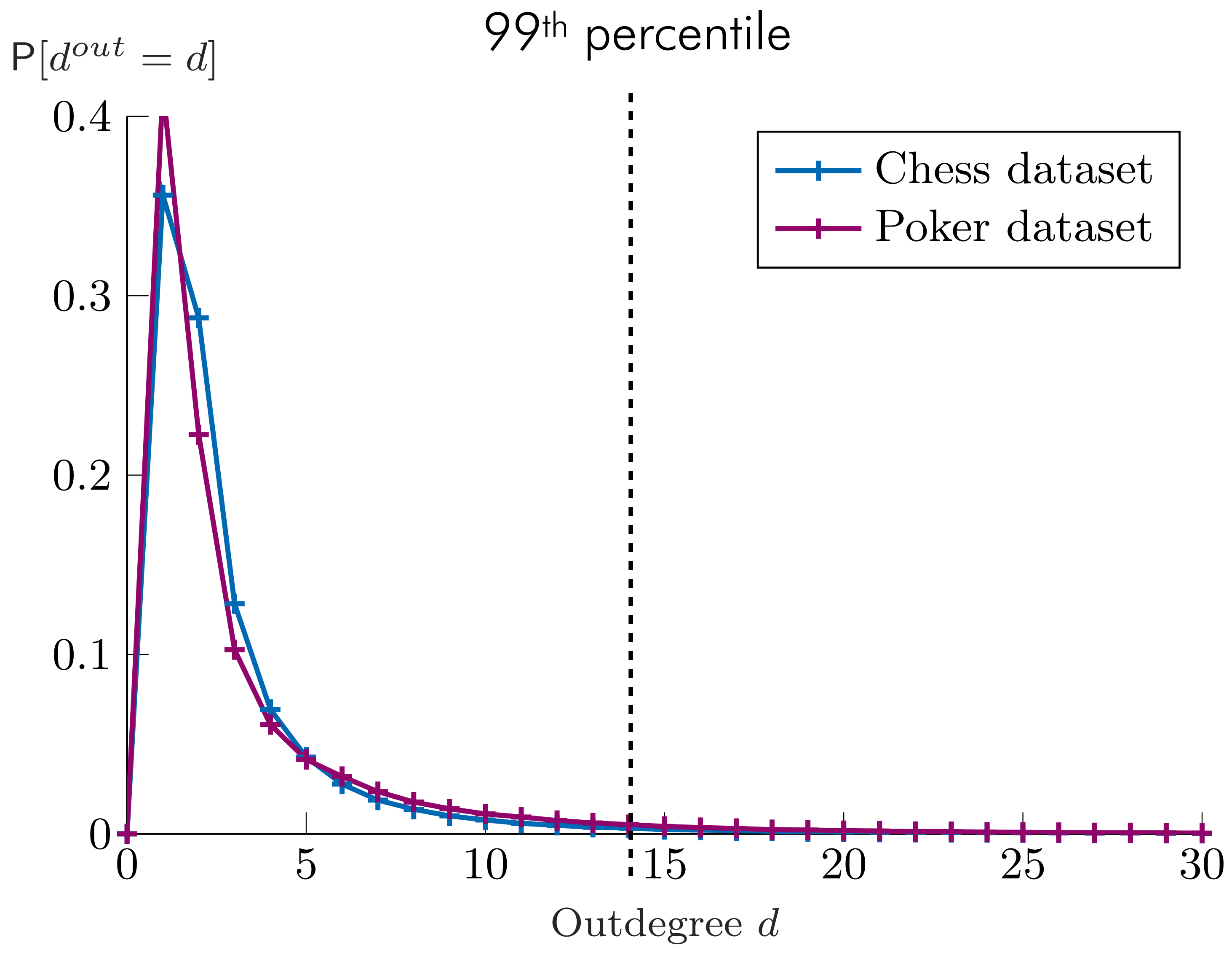


Fig. *Empirical* distribution of the outdegree in the collected data-sets.

Outdegree Distribution

- The outdegree probability density function is ***homogeneous*** for all the nodes.

THEOREM.

```
The nodes' expected outdegree in a
network of N agents equals the (N-1)th
harmonic number:
```

$$\mathbb{E}[d_N^{out,\star}] = \sum_{k=1}^{N-1} \frac{1}{k}.$$

- The empirical distribution is:
  - concentrated in the region $d \in [0,10]$;
  - but it seems to underestimate the low-outdegree nodes.
- This might be the effect of the ***recommendation systems***, as it is more pronounced for the followers of the top nodes.

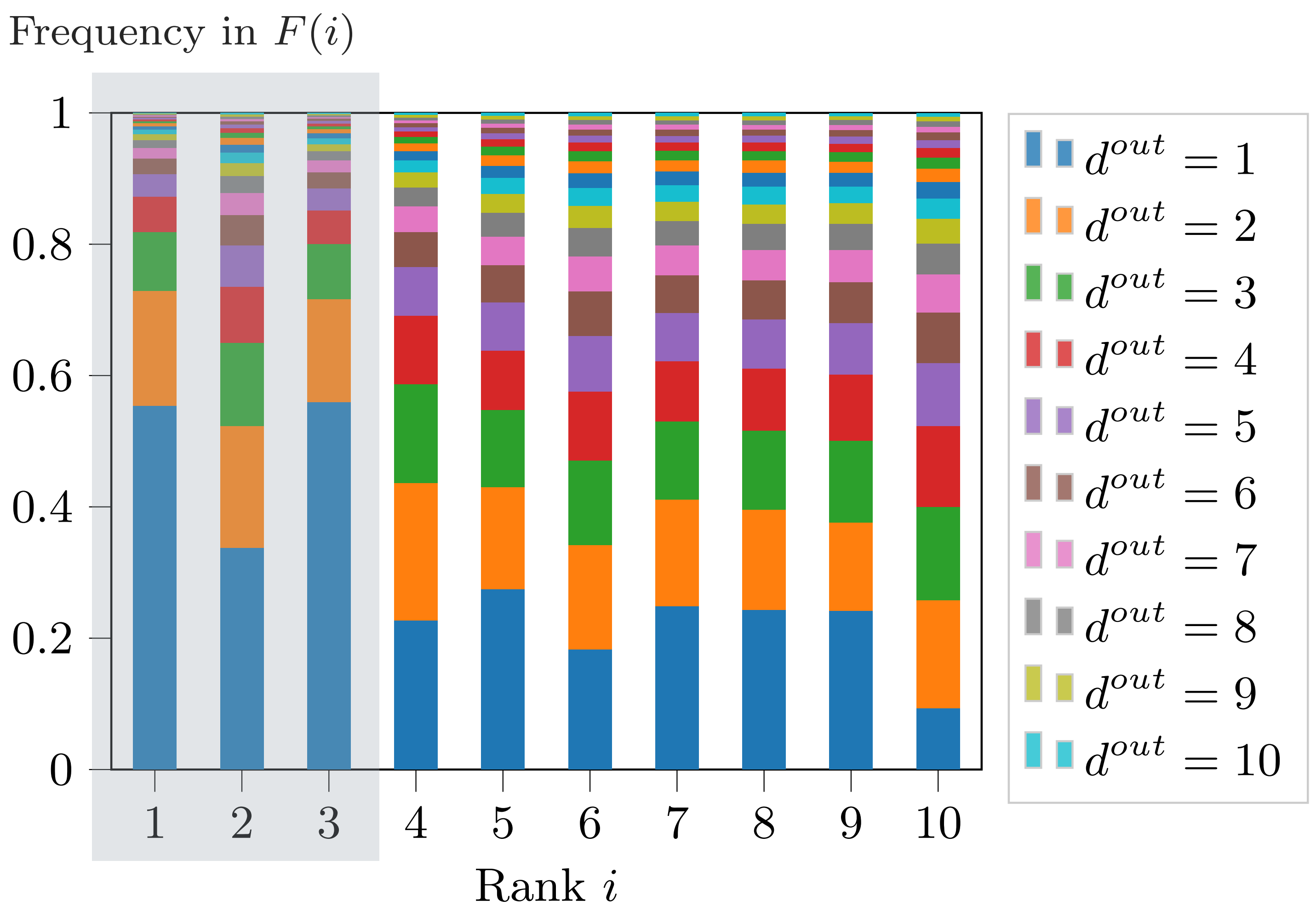


Fig. For each column i, the ***empirical*** distribution of the outdegree of the followers of node i (in the collected data-sets).

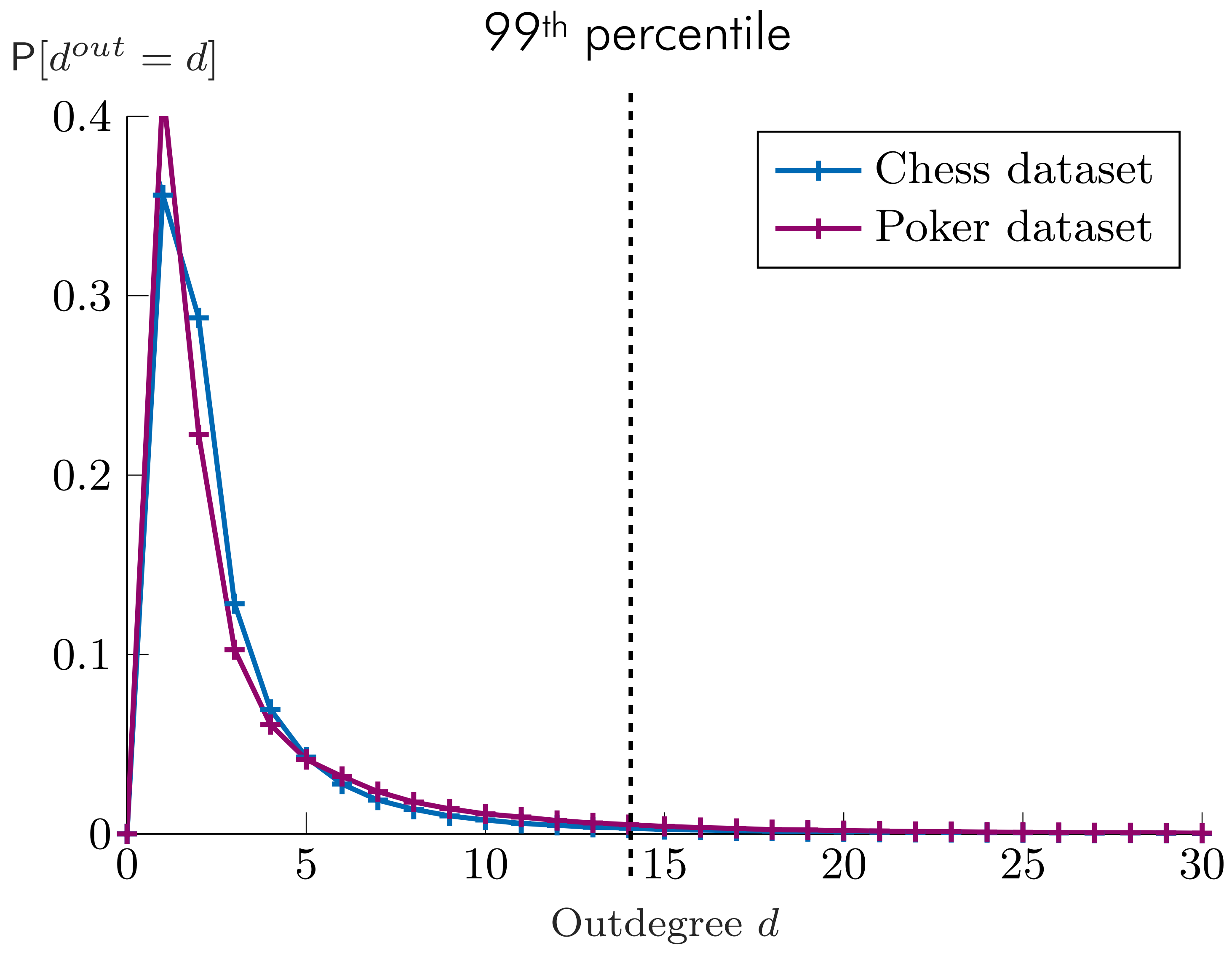


Fig. ***Empirical*** distribution of the out degree in the collected data-sets.

CONCLUSIONS

Summary and Open Directions

- We proposed a *quality-based* approach to model the social network formation on *UGC-based platforms*.
- We numerically *simulated* the dynamics and we studied the *theoretical properties* of the network dynamics and equilibria in terms of:
  - Indegree and Outdegree distribution;
  - Audience Overlap Index.
- Our results are cross-validated with two data-sets collected from the *Twitch* platform.
- The simple quality-based rule leads to a *Zipf's law* in the expected indegree of the *social media influencers*.

Summary and Open Directions

- We proposed a ***quality-based*** approach to model the social network formation on ***UGC-based platforms***.
- We numerically ***simulated*** the dynamics and we studied the ***theoretical properties*** of the network dynamics and equilibria in terms of:
  - Indegree and Outdegree distribution;
  - Audience Overlap Index.
- Our results are cross-validated with two data-sets collected from the ***Twitch*** platform.
- The simple quality-based rule leads to a ***Zipf's law*** in the expected indegree of the ***social media influencers***.

- Study the network properties in terms of, e.g., ***information spreading***, environmentally friendly habits adoption, …
- Validate the model through ***longitudinal*** network data and potentially enrich the quality-based rule.
- Combine quality attributes with ***socio-strategic incentives***.
- Model the role of the platforms' ***recommendation systems*** and study the interplay between individuals' and platforms' behavior.

Nicolò Pagan
pagann@ethz.ch
https://people.ee.ethz.ch/~pagann/

Dr. Wenjun Mei

Prof. Florian Dörfler

NICOLÒ PAGAN

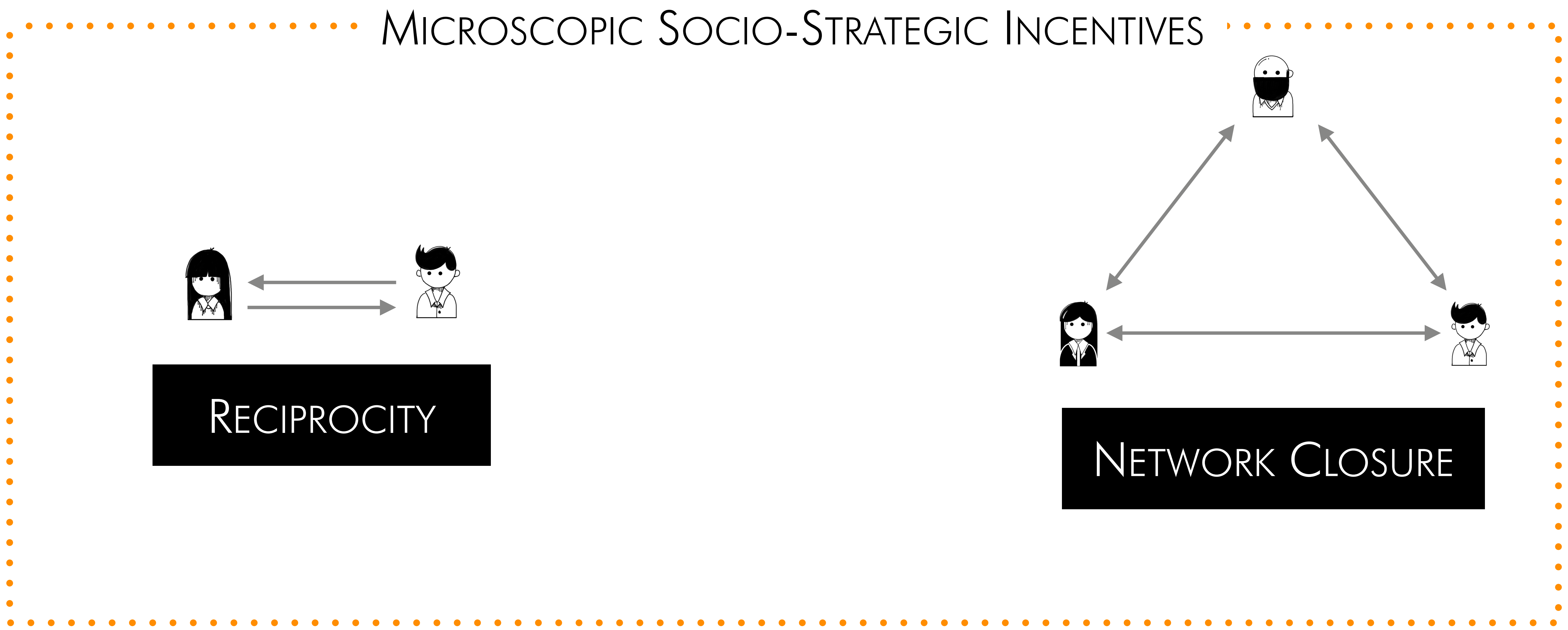
INSTITUT FÜR AUTOMATIK IFA

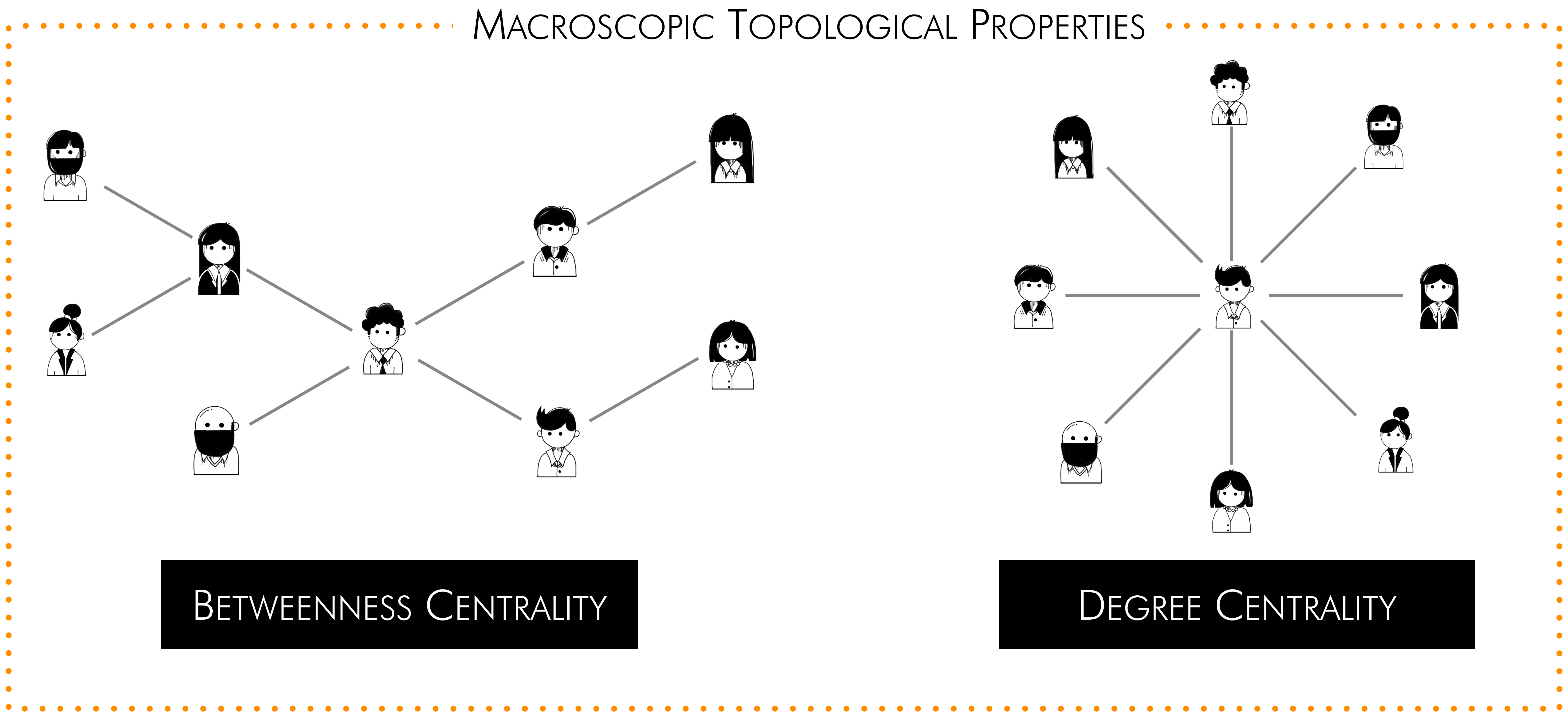
ETH zürich

BACK UP SLIDES

Motivation

- Online Social Networks are ubiquitous.
- Researchers from different communities have started considering the problem of *Social Network Formation*.
- "Old-generation" Online Social Networks try to mimic offline Social Networks.





Theoretical Results

Theorem.

The probability of node i having indegree $d_i^{in}(t) = d \in [0, N-1]$ after t time-steps is given by

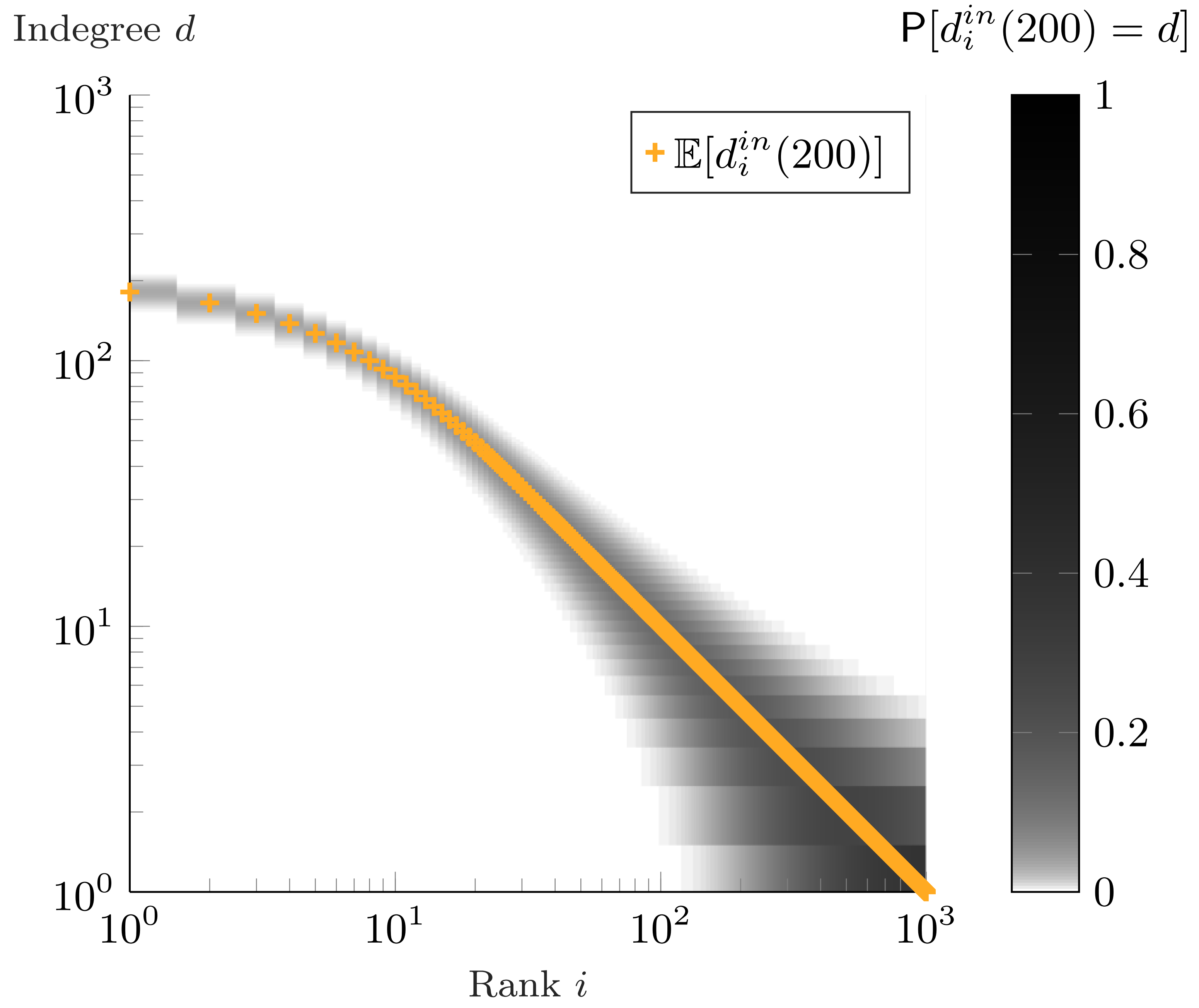
$$\mathsf{P}(d_i^{in}(t) = d) = \sum_{k=0}^{d} \binom{i-1}{k} \bar{p}_i^k \left(1 - \bar{p}_i\right)^{i-1-k} \binom{N-i}{d-k} \underline{p}_i^{d-k} \left(1 - \underline{p}_i\right)^{N-i-(d-k)},$$

where we omitted the time-step dependency on $\bar{p}_i$and $\underline{p}_i$ and:

$$\mathsf{P}(a_{ji}(t) = 1) = \begin{cases} \bar{p}_i(t) := \frac{1}{i-1}\left(1 - \left(\frac{N-i}{N-1}\right)^t\right), & \texttt{if } j < i \\ \underline{p}_i(t) := \frac{1}{i}\left(1 - \left(\frac{N-i-1}{N-1}\right)^t\right), & \texttt{if } j > i. \end{cases}$$

Thus, the expected indegree of node i after t time-steps reads as:

$$\mathbb{E}(d_i^{in}(t)) = \frac{N}{i} - \left(\left(\frac{N-i}{N-1}\right)^t + \frac{N-i}{i}\left(\frac{N-i-1}{N-1}\right)^t\right).$$

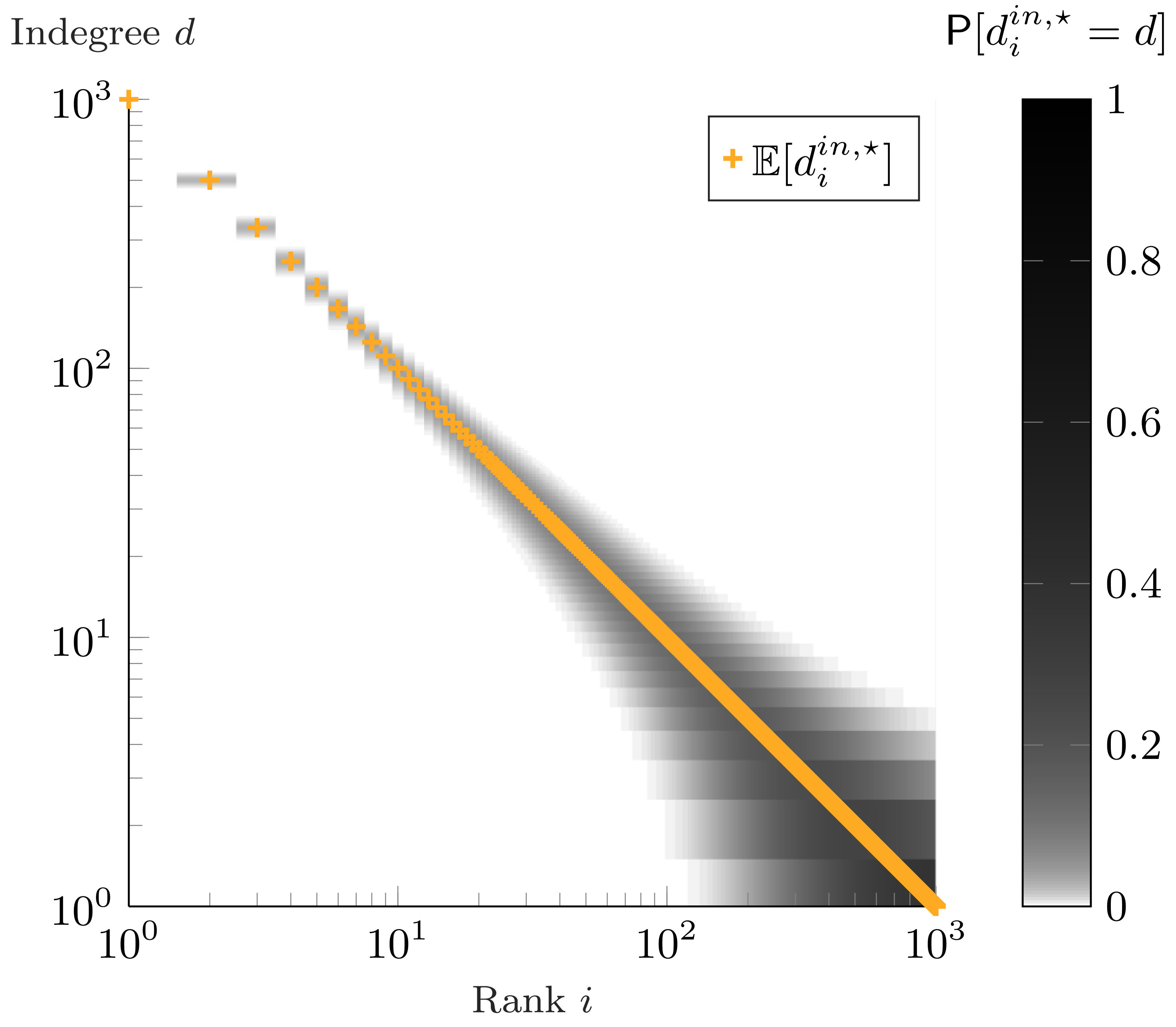


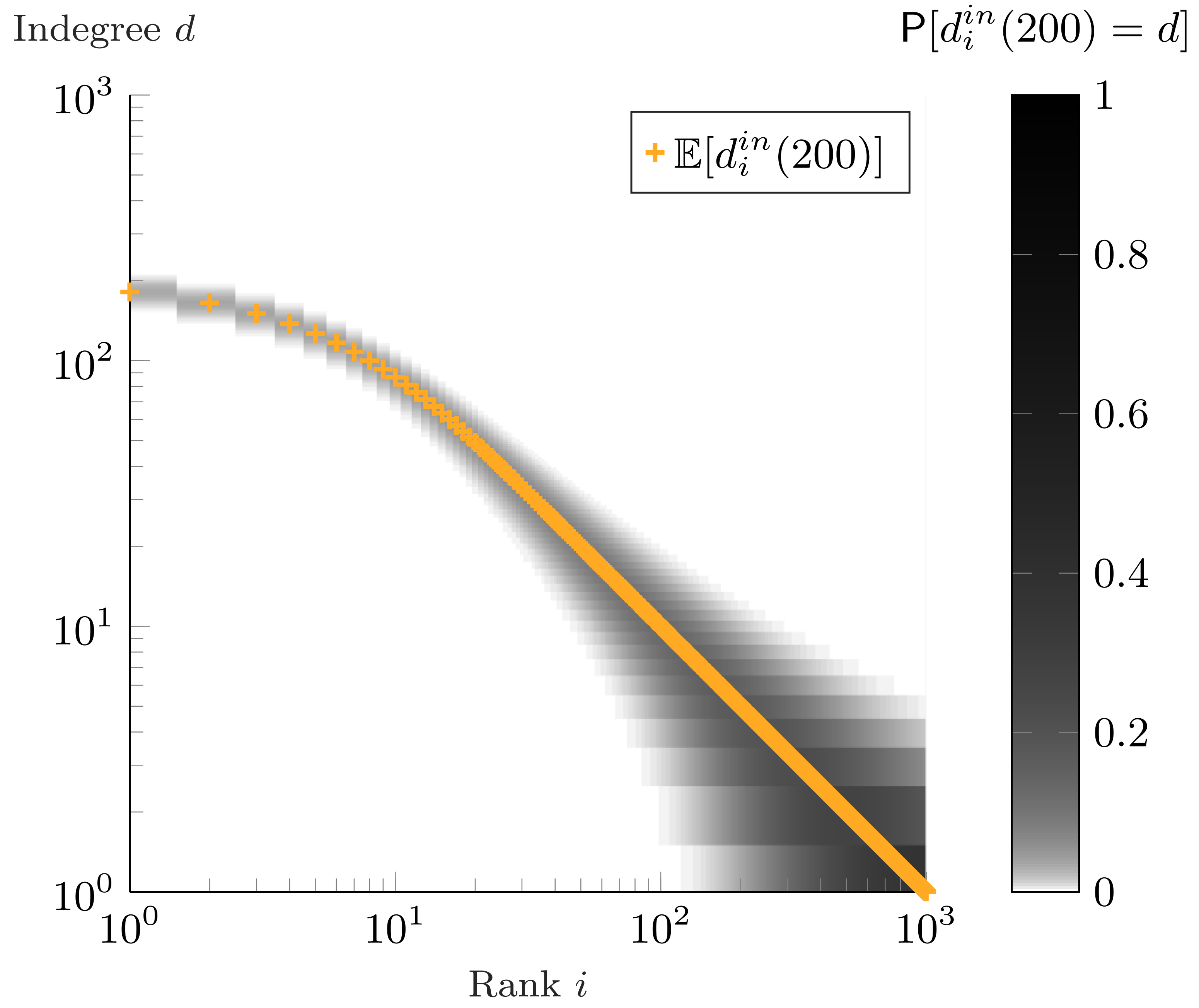
Indegree Distribution

Theorem.

At equilibrium, the expected indegree of node i reads as:

$$\mathbb{E}(d_i^{in,\star}) = \begin{cases} N-1, & \text{if } i=1, \\ \frac{N}{i}, & \text{otherwise.} \end{cases}$$





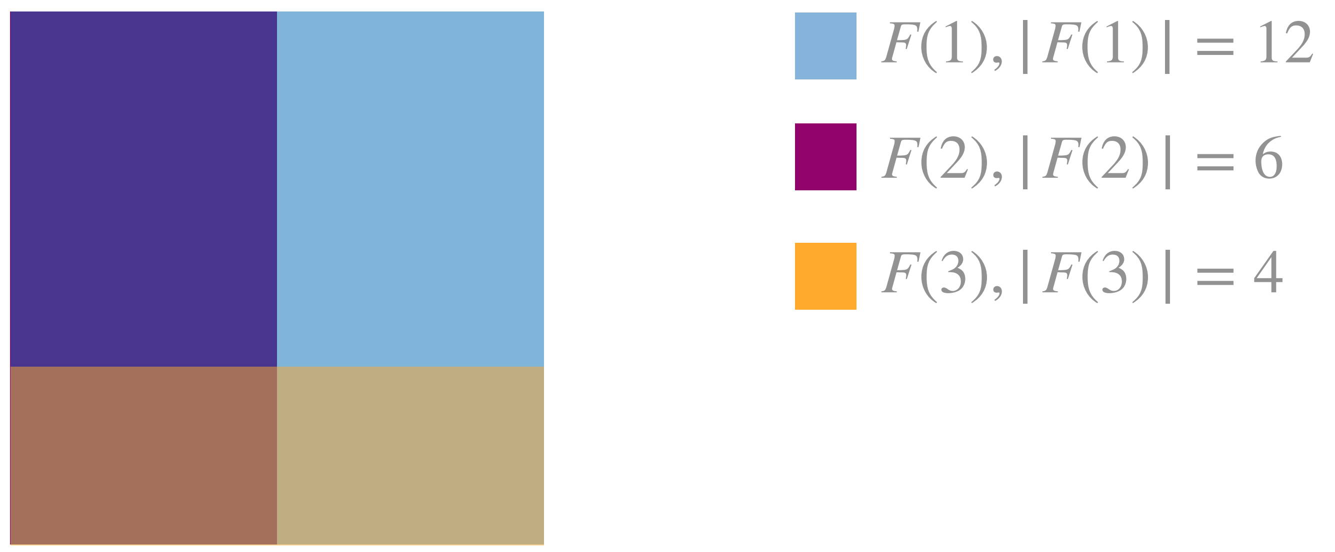
Audience Overlap

Definition.

We define the Audience Overlap index as follows:

$$O(i,j) := \frac{|F(i) \cap F(j)|}{|F(i)|} \in [0,1],$$

if $|F(i)| > 0$, and 0 otherwise, where $F(i)$ denotes the set of followers of i.



$O(1,2) = 1/2$

$O(1,3) = 1/3\ldots$

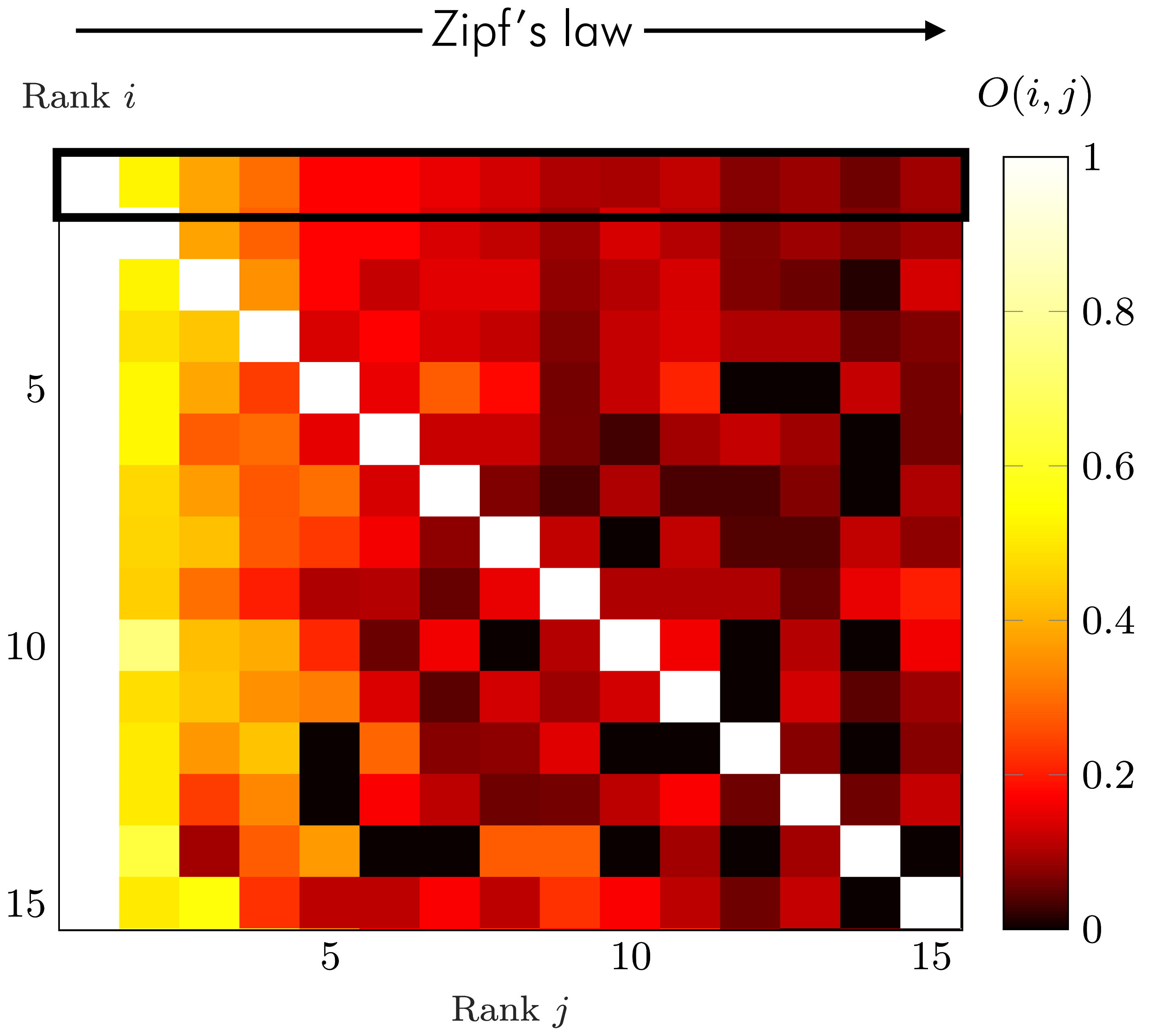


Fig. Numerical results on the followers' overlap among the top nodes. The result is obtained with N=201 agents, upon reaching equilibrium.

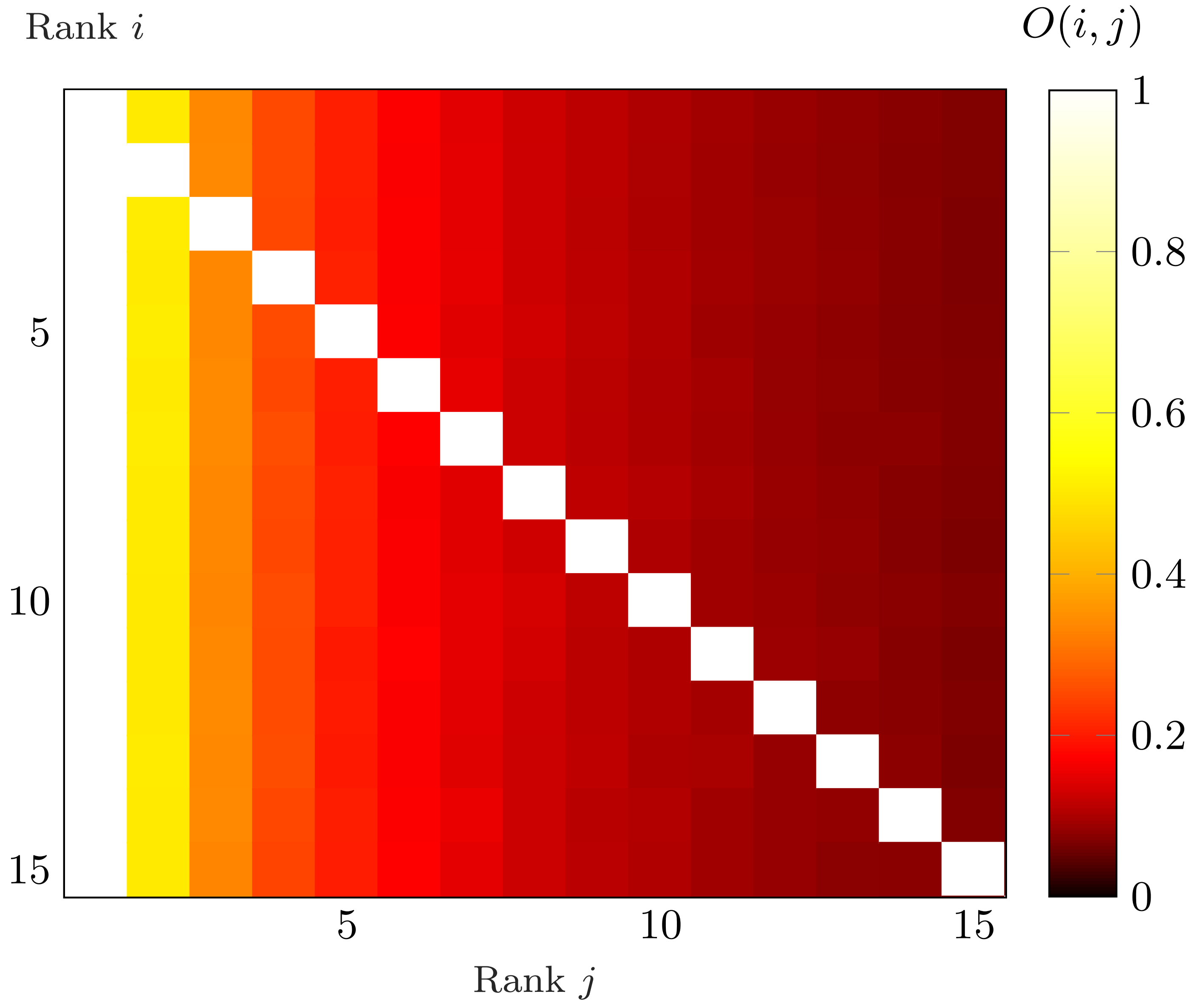


Fig. *Average* numerical results on the followers' overlap among the top nodes. The results are obtained from 1000 simulations with N=201 agents, upon reaching equilibrium.

Audience Overlap

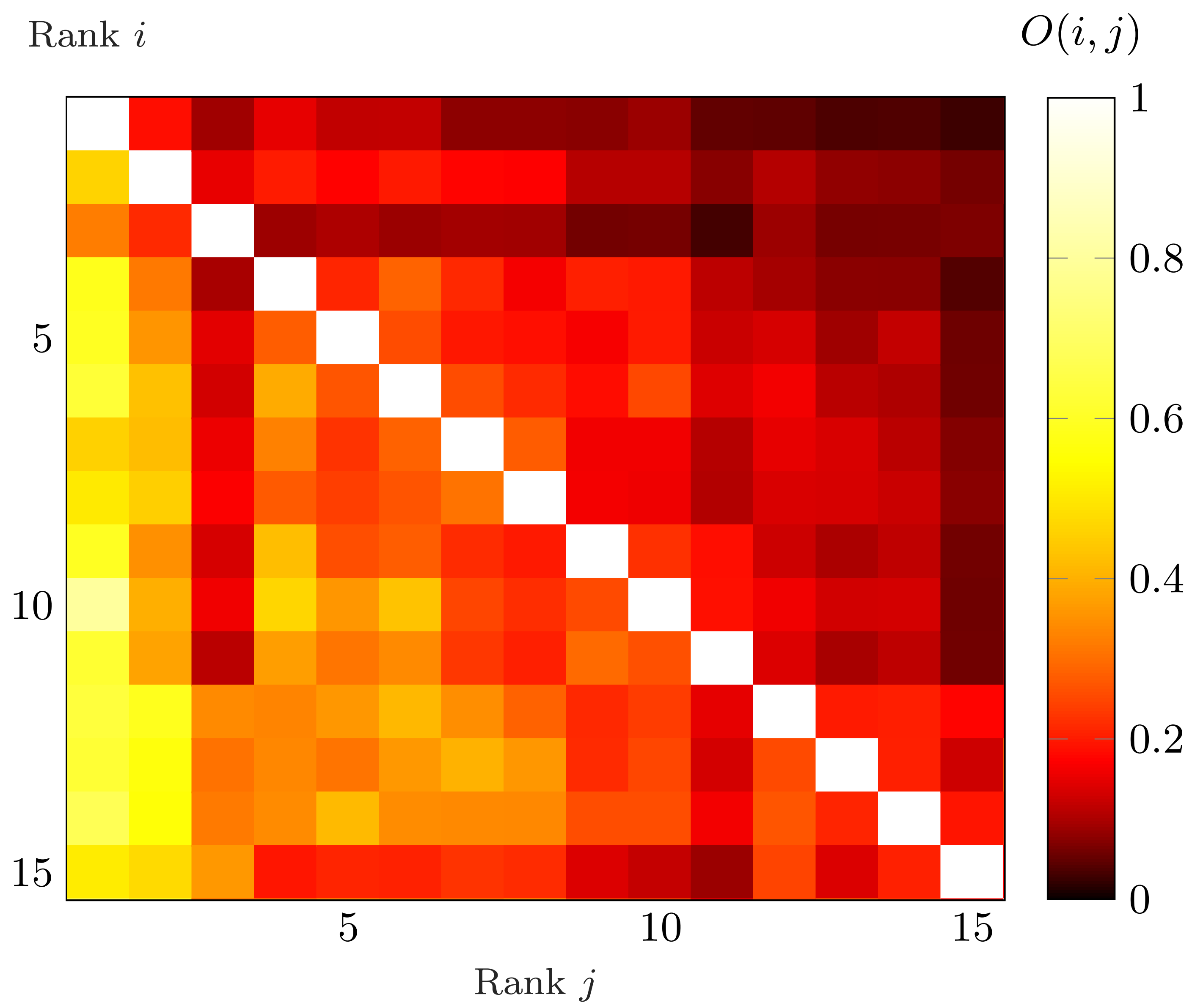


Fig. *Empirical* results from our collected data-set.

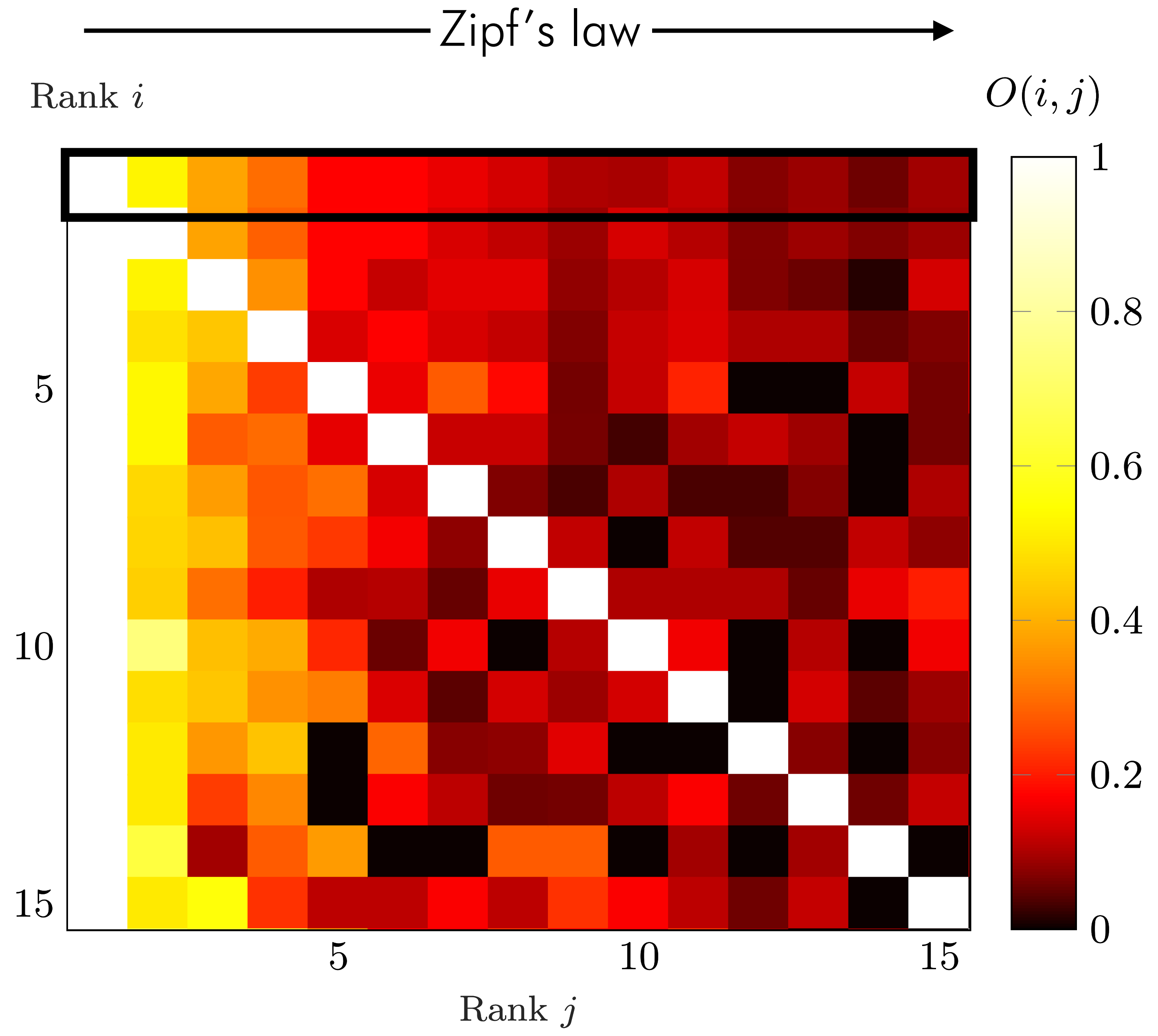


Fig. Numerical results on the followers' overlap among the top nodes. The result is obtained with N=201 agents, upon reaching equilibrium.

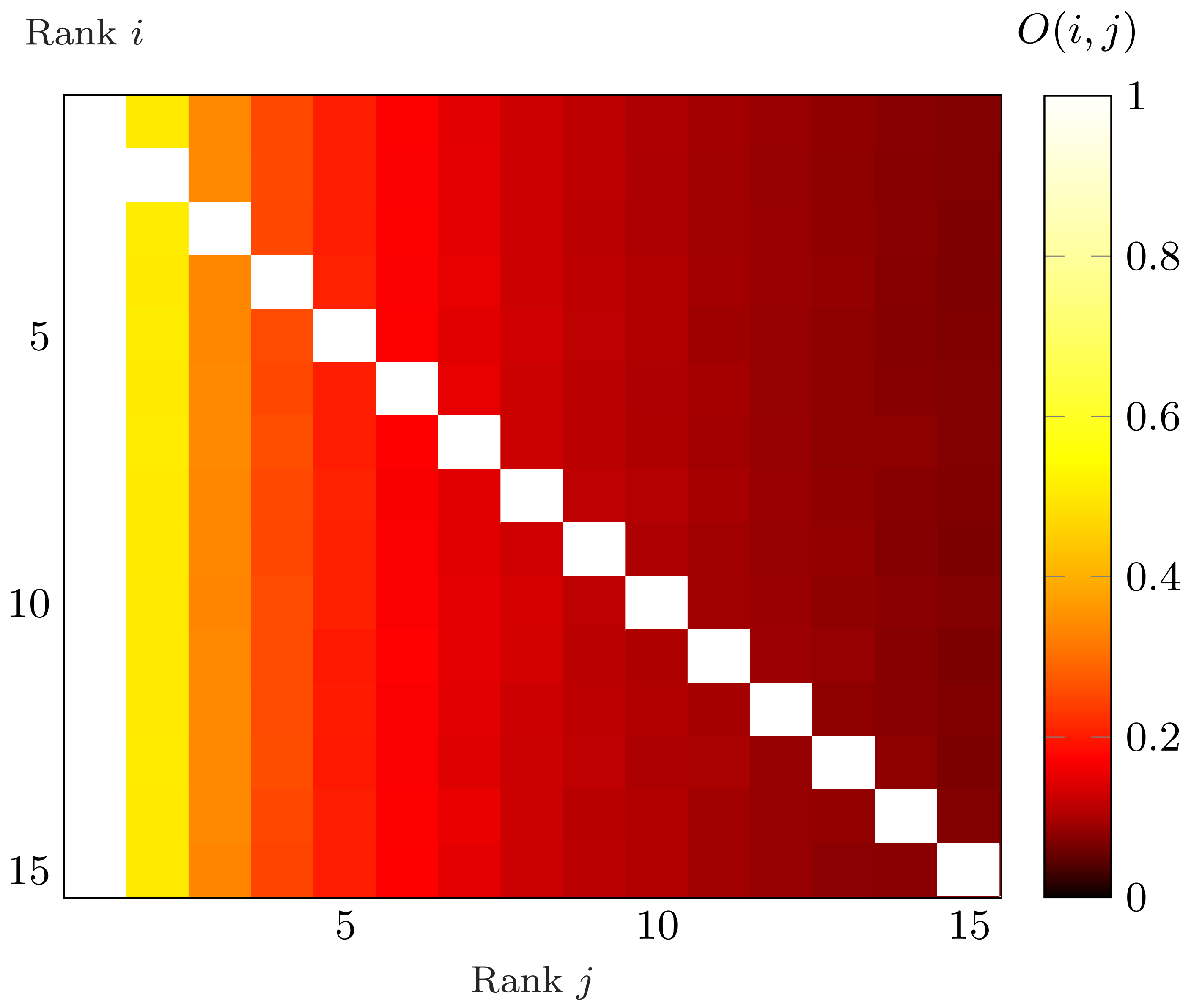


Fig. *Average* numerical results on the followers' overlap among the top nodes. The results are obtained from 1000 simulations with N=201 agents, upon reaching equilibrium.

Audience Overlap

Definition.

We define the Audience Overlap index as follows:

$$O(i,j) := \frac{|F(i) \cap F(j)|}{|F(i)|} \in [0,1],$$

if $|F(i)| > 0$, and 0 otherwise, where $F(i)$ denotes the set of followers of i.

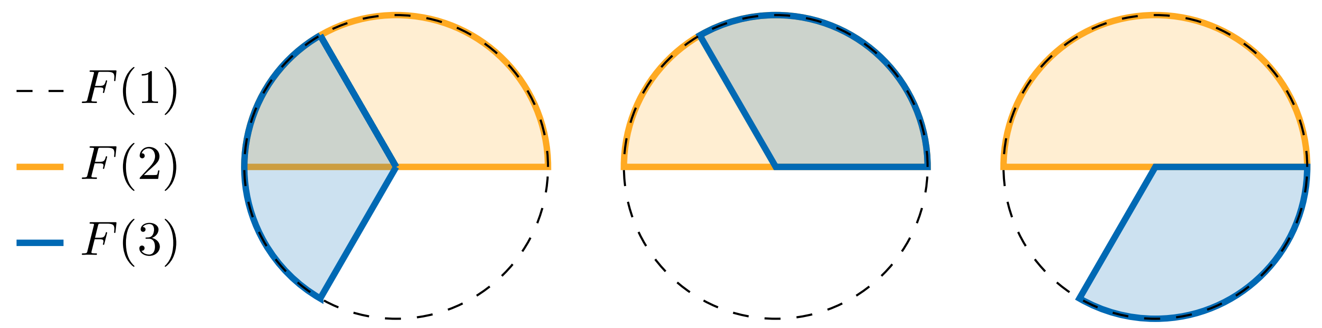


Fig. Sketch of the overlap between the followers' sets of the top 3 nodes.

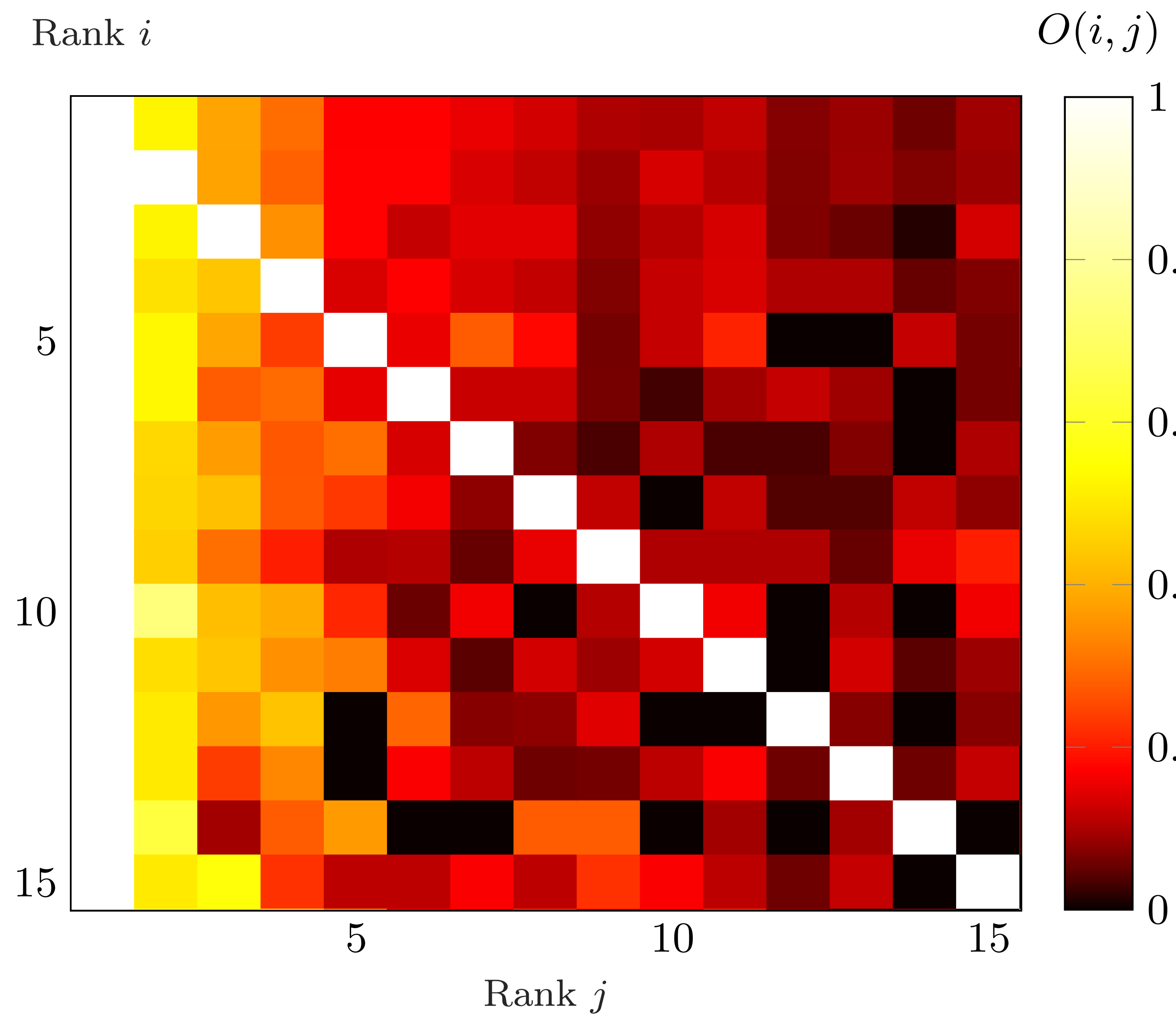


Fig. Numerical results on the followers' overlap among the top nodes. The results are obtained from 1000 simulations with N=201 agents, upon reaching equilibrium.

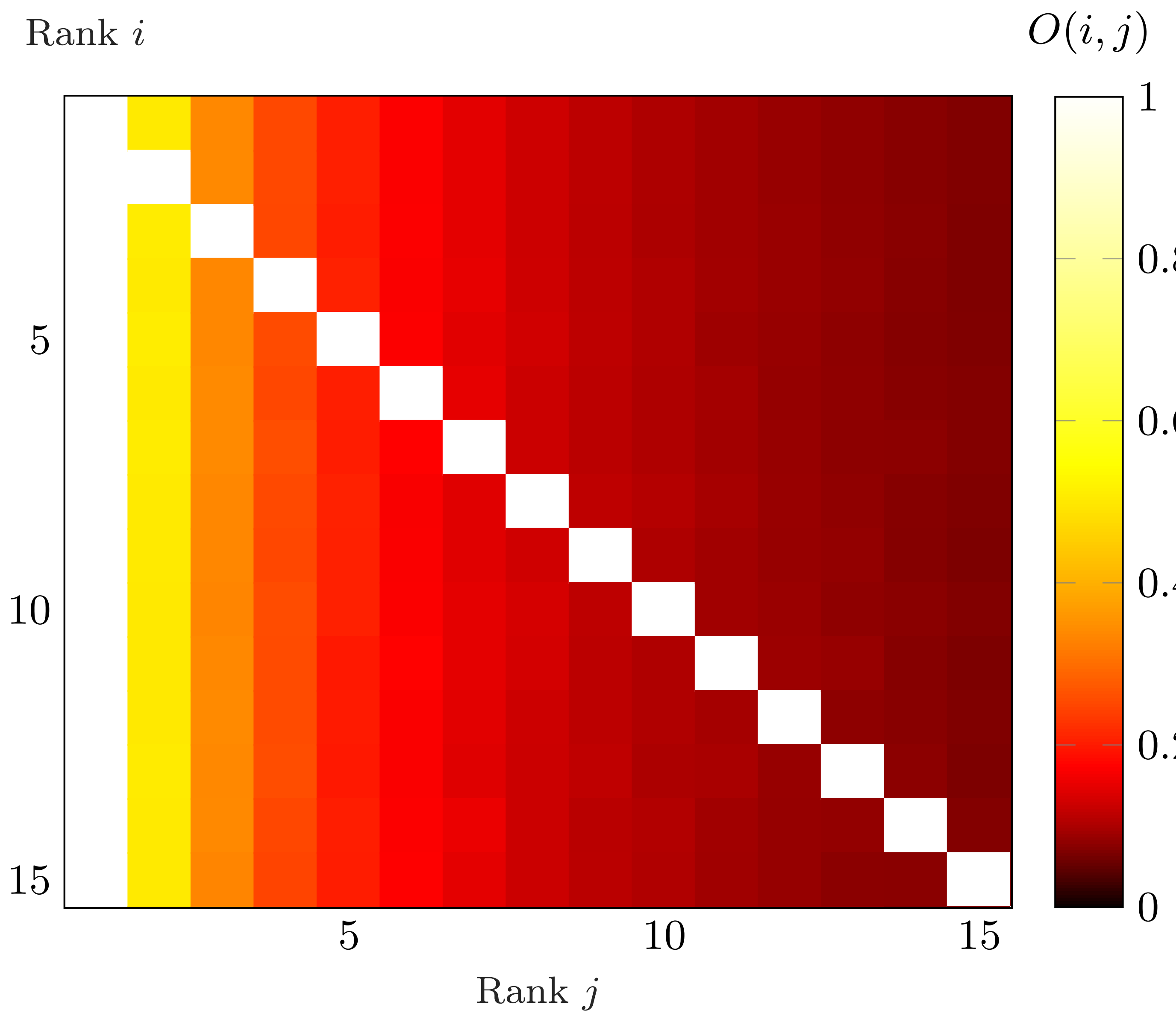


Fig. Average numerical results on the followers' overlap among the top nodes. The results are obtained from 1000 simulations with N=201 agents, upon reaching equilibrium.

Audience Overlap

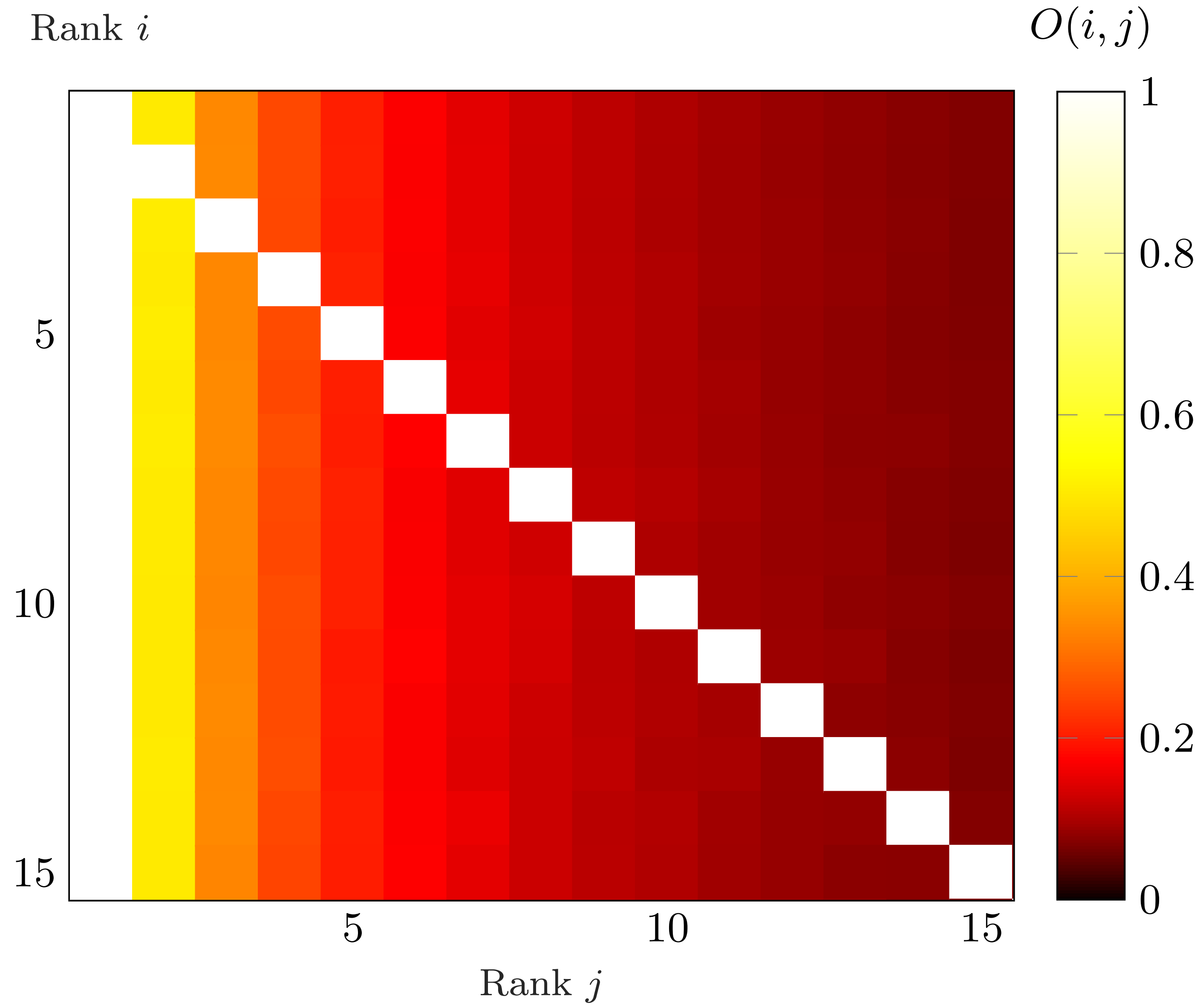


Fig. Average numerical results on the followers' overlap among the top nodes. The results are obtained from 1000 simulations with N=201 agents, upon reaching equilibrium.

Definition.

We define the Audience Overlap index as follows:

$$O(i,j) := \frac{|F(i) \cap F(j)|}{|F(i)|} \in [0,1],$$

if $|F(i)| > 0$, and 0 otherwise, where $F(i)$ denotes the set of followers of agent i.

Audience Overlap

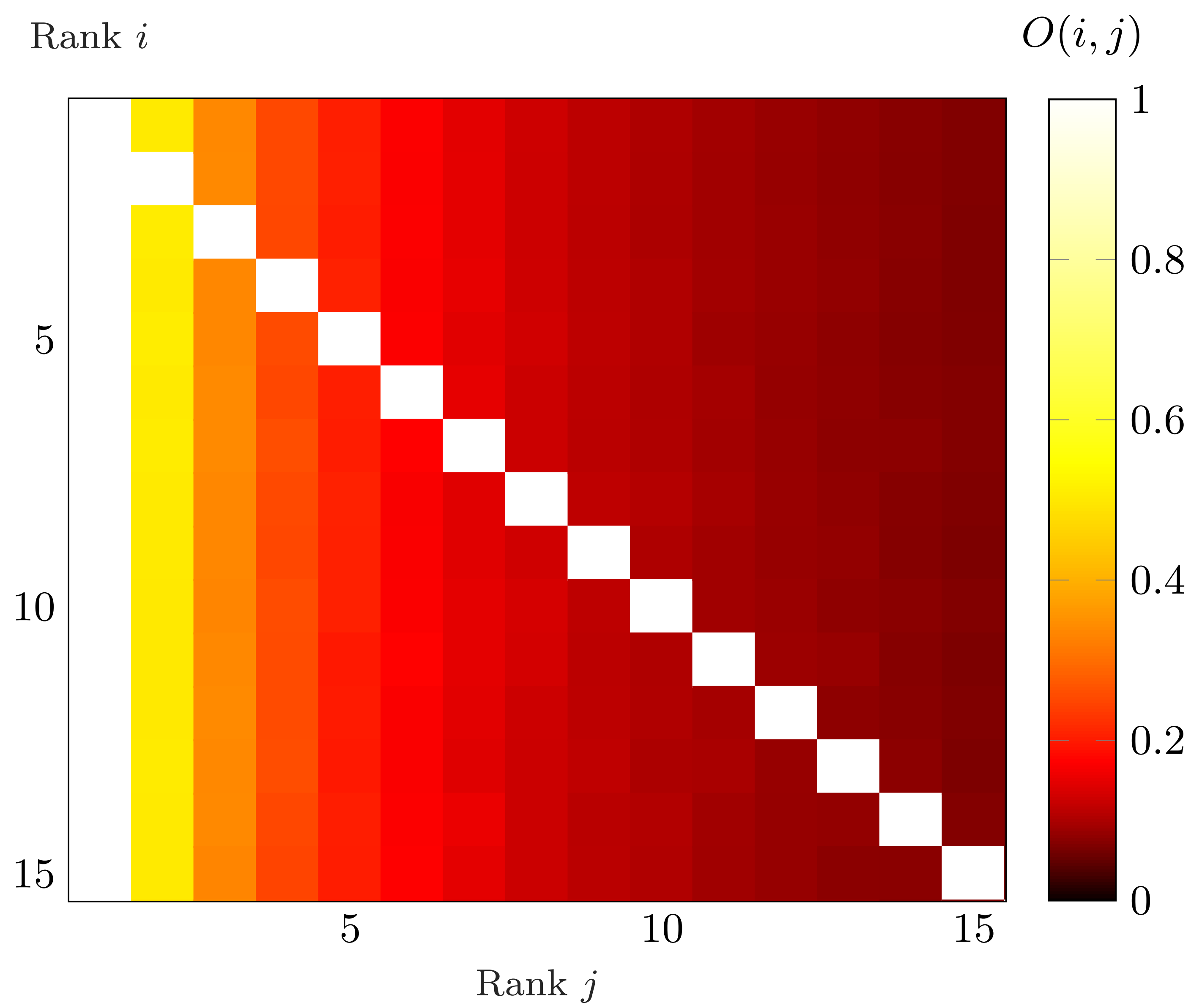


Fig. Average numerical results on the followers' overlap among the top nodes. The results are obtained from 1000 simulations with N=201 agents, upon reaching equilibrium.

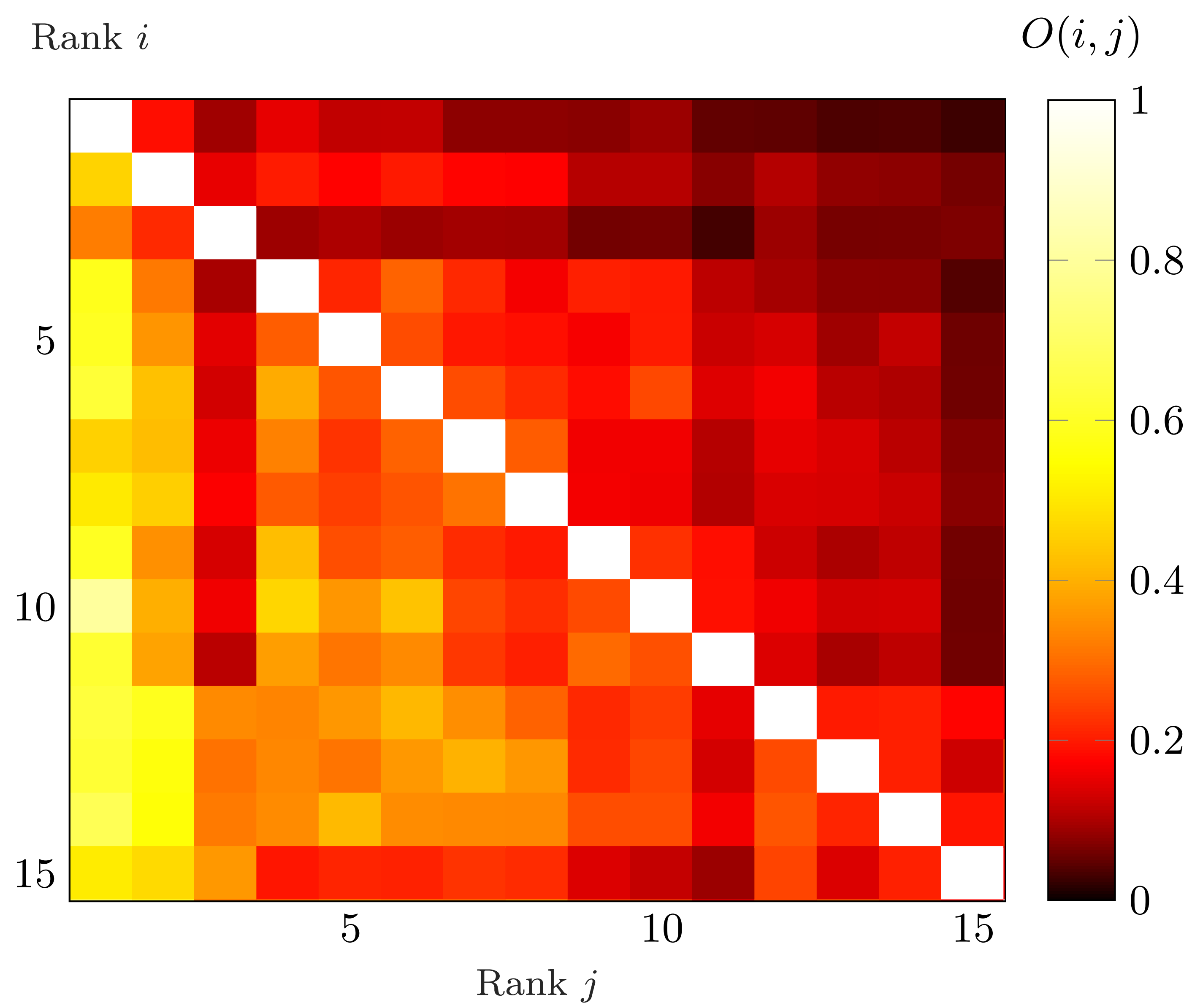


Fig. Average numerical results on the followers' overlap among the top nodes. The results are obtained from our collected data-set.

Audience Overlap

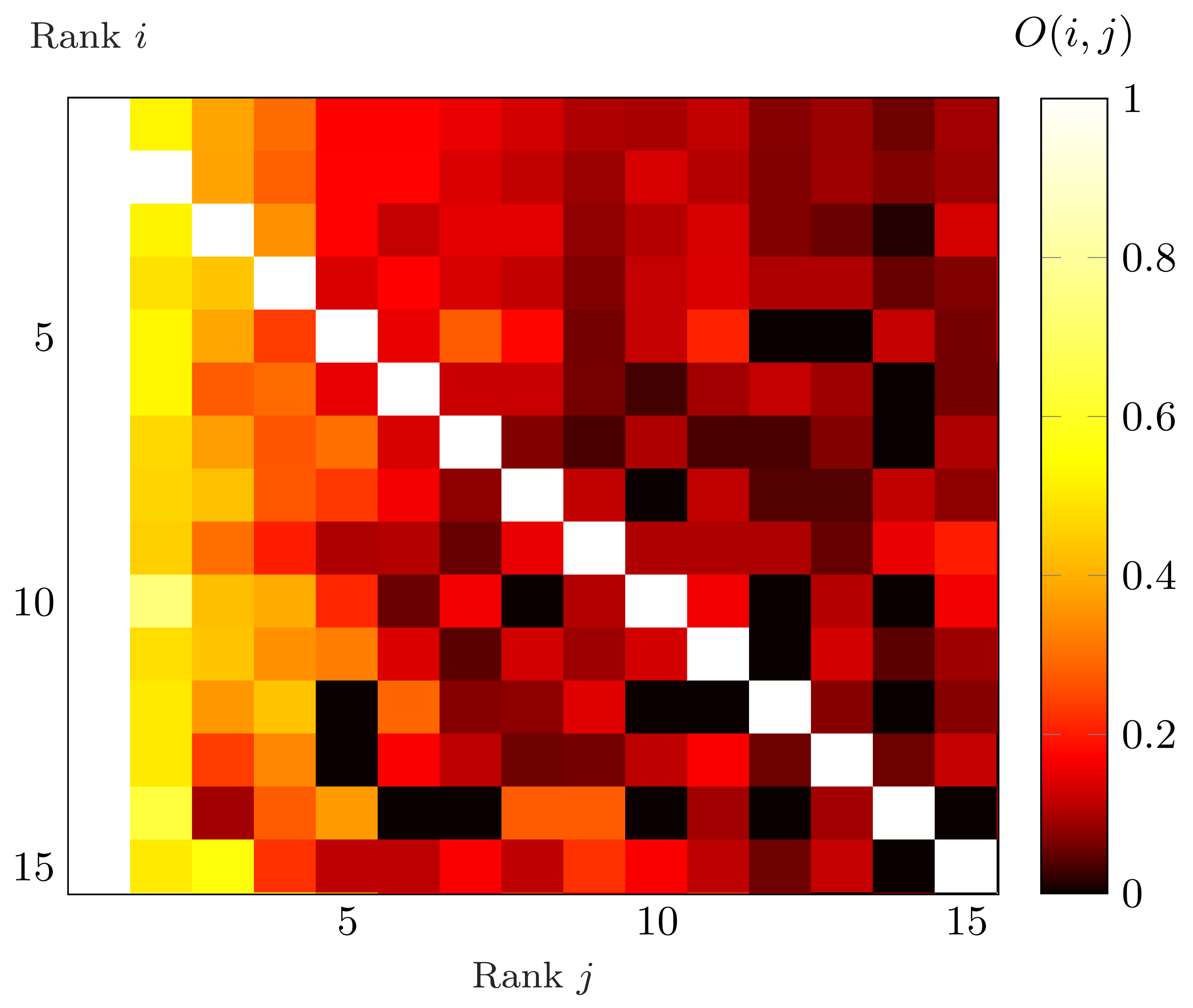


Fig. Numerical results on the followers' overlap among the top nodes. The results are obtained from 1000 simulations with N=201 agents, upon reaching equilibrium.

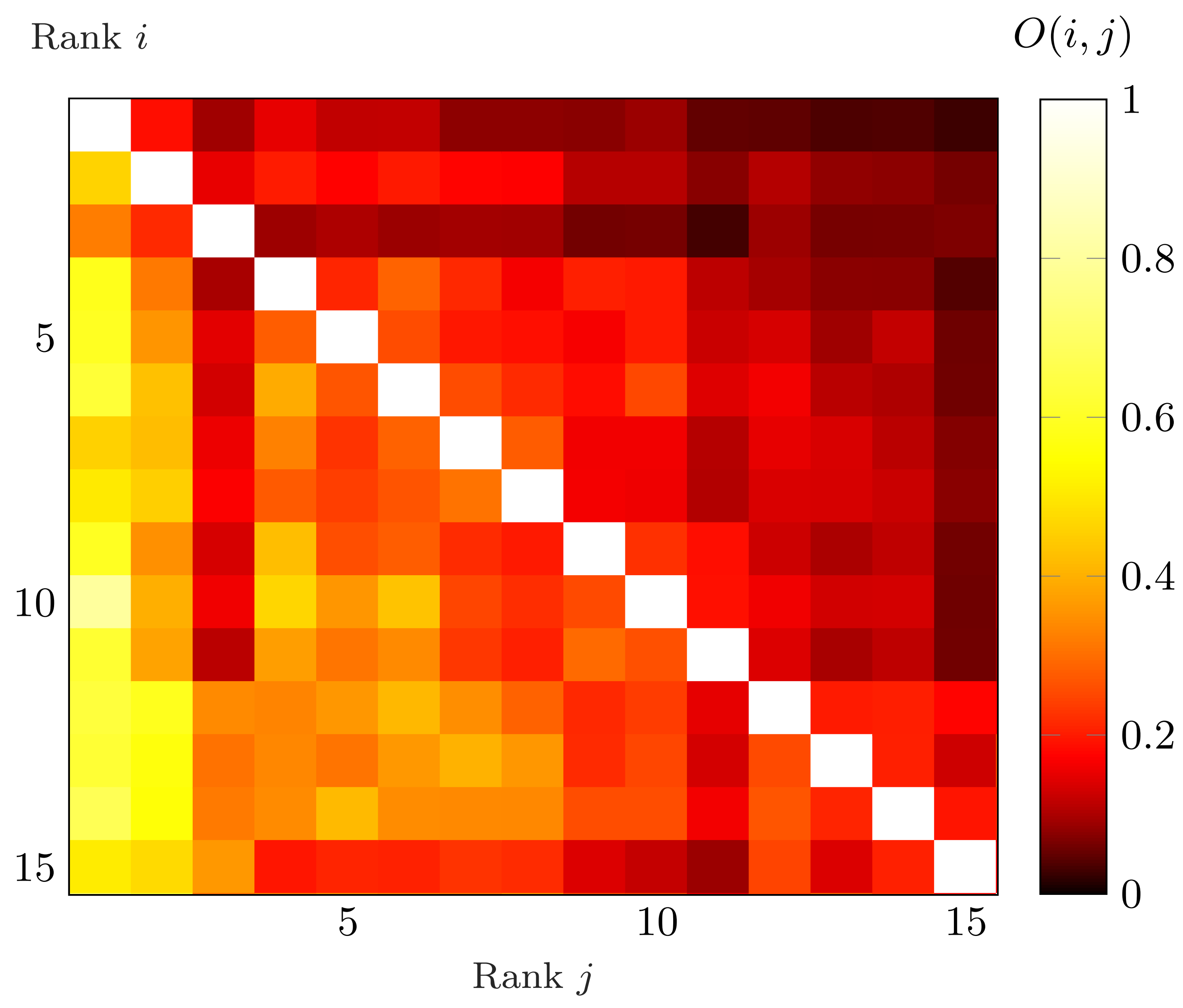


Fig. Average numerical results on the followers' overlap among the top nodes. The results are obtained from our collected data-set.

Simulation Results

Theoretical Results

Empirical Results

Outdegree Distribution

- The outdegree probability density function is homogeneous for all the nodes.

Theorem.

The nodes' expected outdegree in a network of $N \geq 2$ agents equals the $(N-1)$-th harmonic number:

$$\mathbb{E}[d_N^{out,\star}] = \sum_{k=1}^{N-1} \frac{1}{k}.$$

- Compared to empirical data, the theoretical distribution seems to underestimate the low-outdegree nodes.
- This might be the effect of the recommendation systems, as it is more pronounced for the followers of the top nodes.

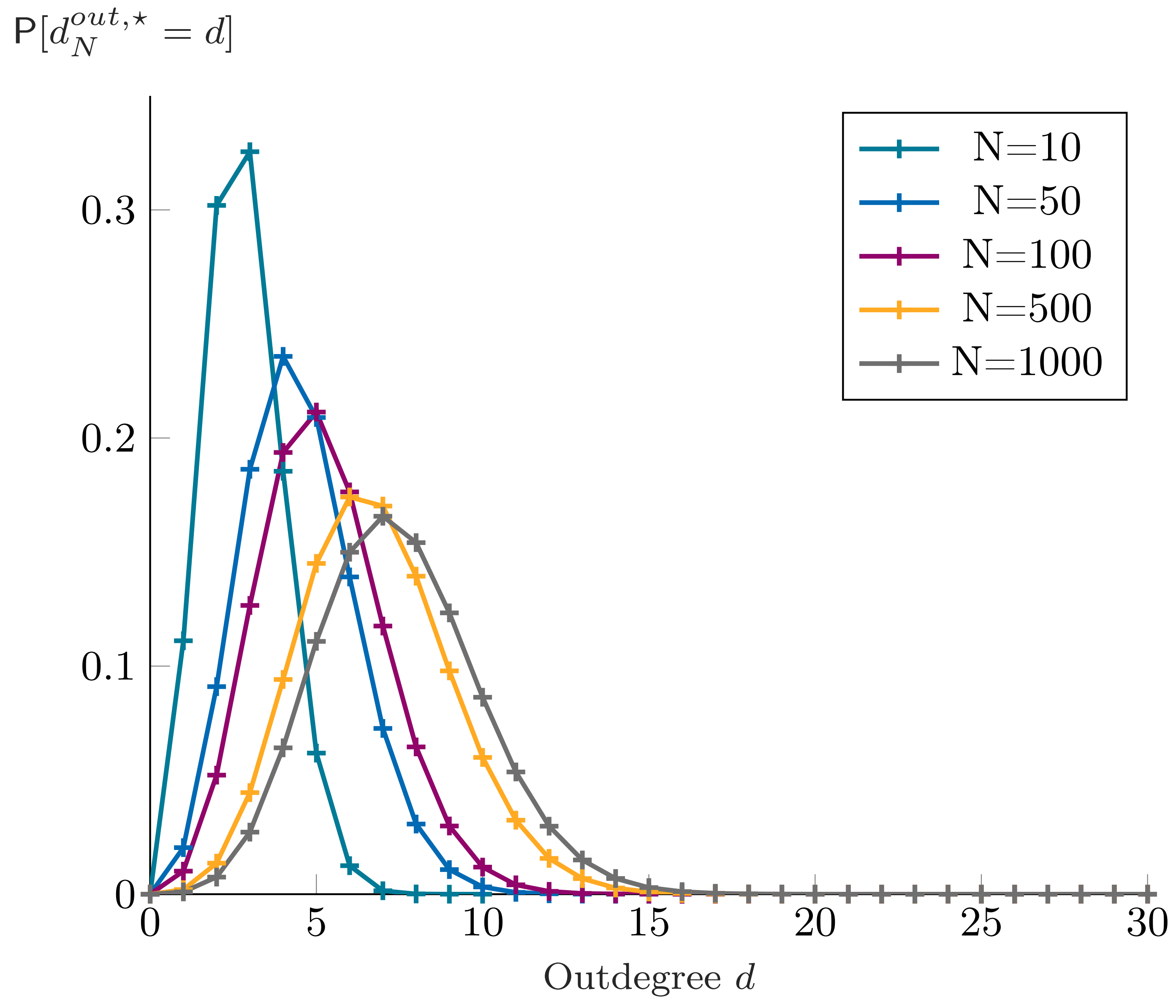


Fig. Analytical outdegree distribution for different network size.

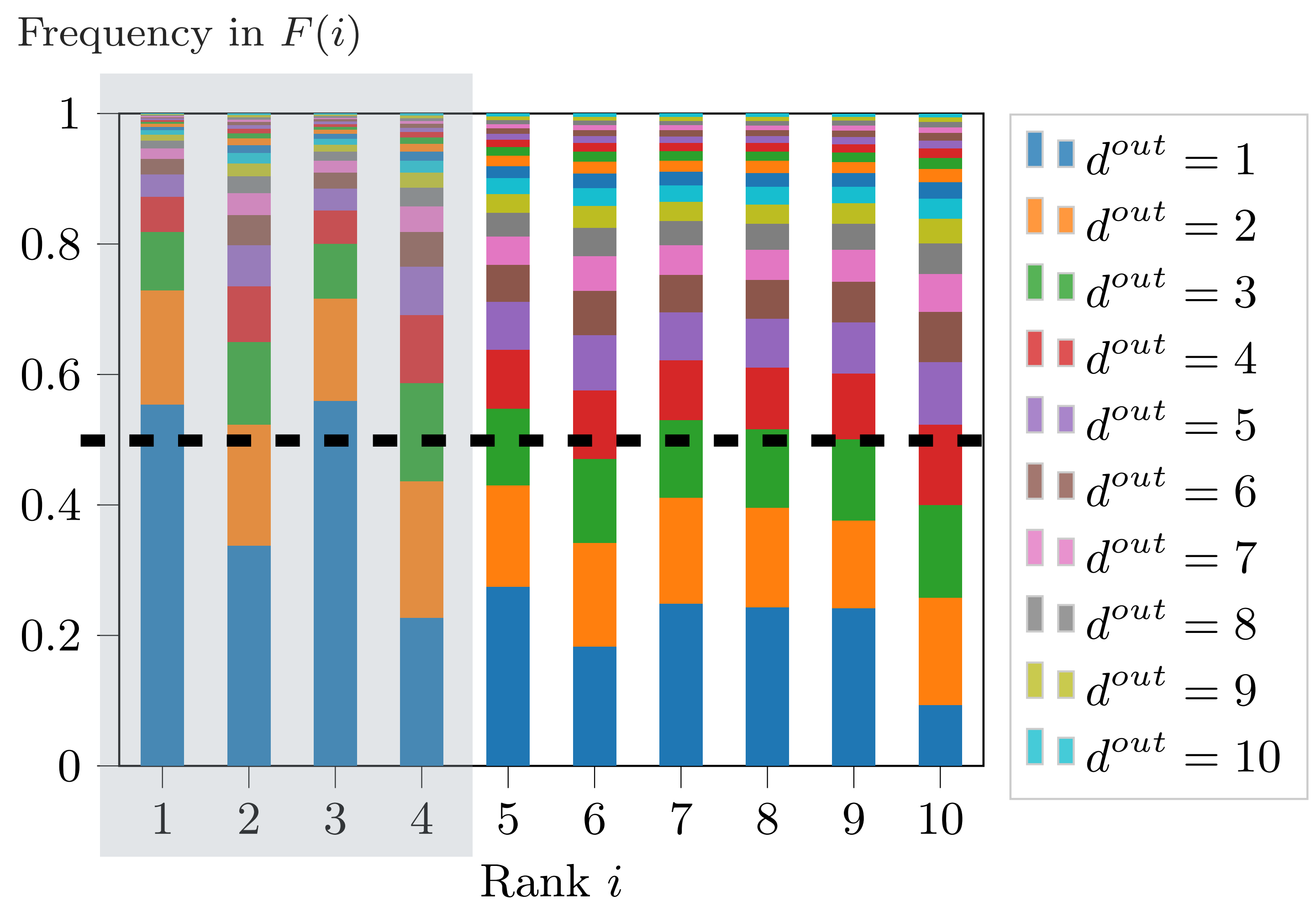


Fig. For each column i, the empirical distribution of the outdegree of the followers of node i (in the collected data-sets).