network science institute

Complexity Society Lab

The impact of network structure on social contagion and neural ODE generalization

Moritz Laber

✉ **laber.m@northeastern.edu**
MPI IS, Tübingen, October 6, 2026

About me

I’m a PhD candidate at the Network Science Institute at Northeastern University in Boston.

My background is in physics (Karlsruhe Institute of Technology) and computational science (University of Vienna / Complexity Science Hub Vienna).

About me

Complex systems

- (higher-order) network models
  (typicality, hypergraph null models)
- dynamical processes
  (contagion, percolation, and random walks)
- statistics and measurement
  (homophily, centrality, ranking)

Machine learning

- Graph machine learning
  (graph neural ODEs, uncertainty quantification)
- Weight space learning
  (weight generation and model editing)
- Reasoning in LLMs
  (scaling)

Information Access Equality, Homophily, and Hypergraphs

Effects of higher-order interactions and homophily on information access inequality


Laber, M.*


Dies, S.*


Ehlert, J.*


Klein, B.


Eliassi-Rad, T.

Communications Physics 9 13 (2026)
(arXiv:2506.00156)

* equal contribution

A guide to higher-order homophily


Laber, M.

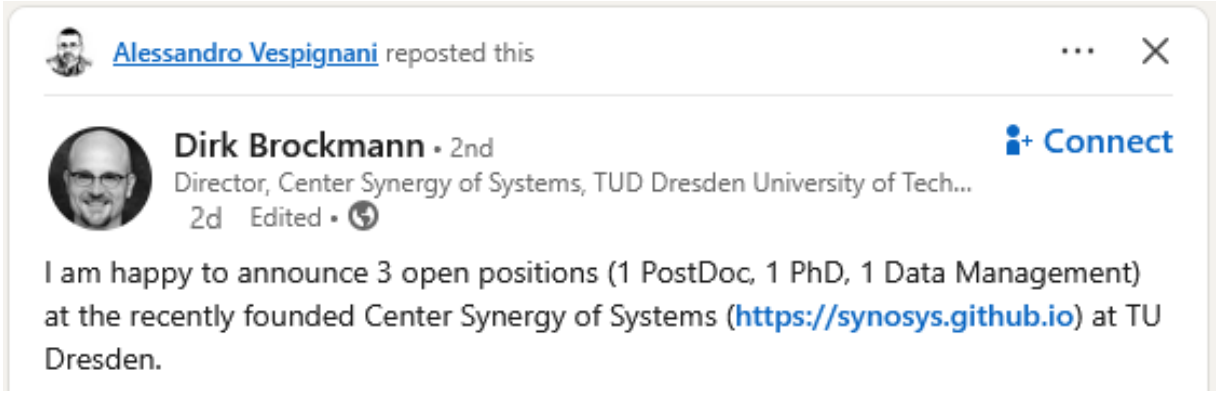

Klein, B.

Invited Chapter in *Social Networks beyond Pairwise Interactions* (Ed. Iacopini I. & Battiston F.)
(arXiv:2606.02537)

Information Access Equality

Information is a resource[1,2].

Information Access Equality[3,4] means that members of different groups have equal chance to obtain a new piece of information.

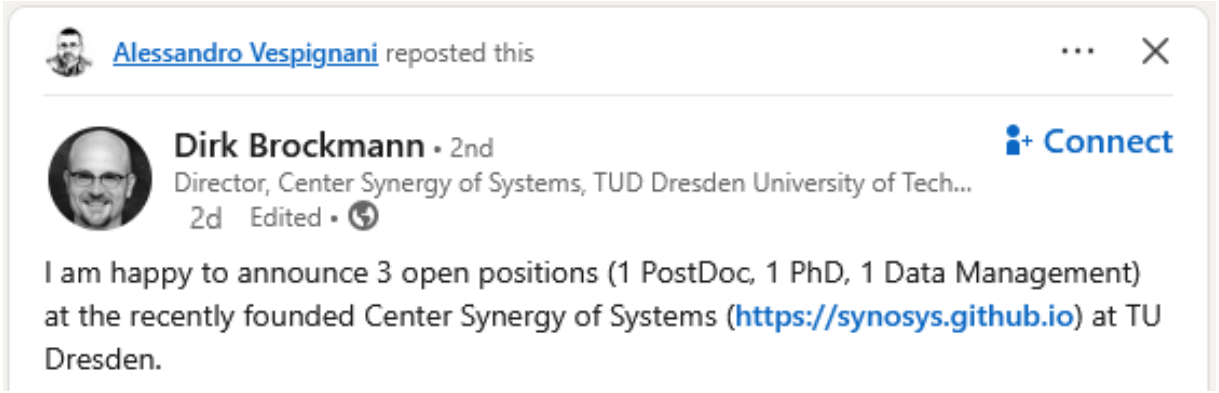


1 Burt; *Res. Org. Behav.* (2000)
2 Karthik et al.; *Science* (2022)
3 Wang, Varol, Eliass-Rad; A*ppl. Net. Sci.* (2022)
4 Stoica, Han, Chaintreau; WWW (2020)
5 Ali et al.; *IEEE TKDE* (2023)

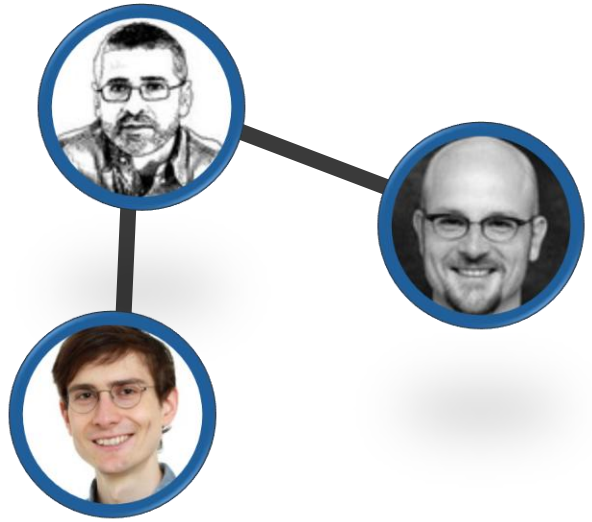
Information Access Equality

Information is a resource[1,2].

Information Access Equality[3,4] means that members of different groups have equal chance to obtain a new piece of information.







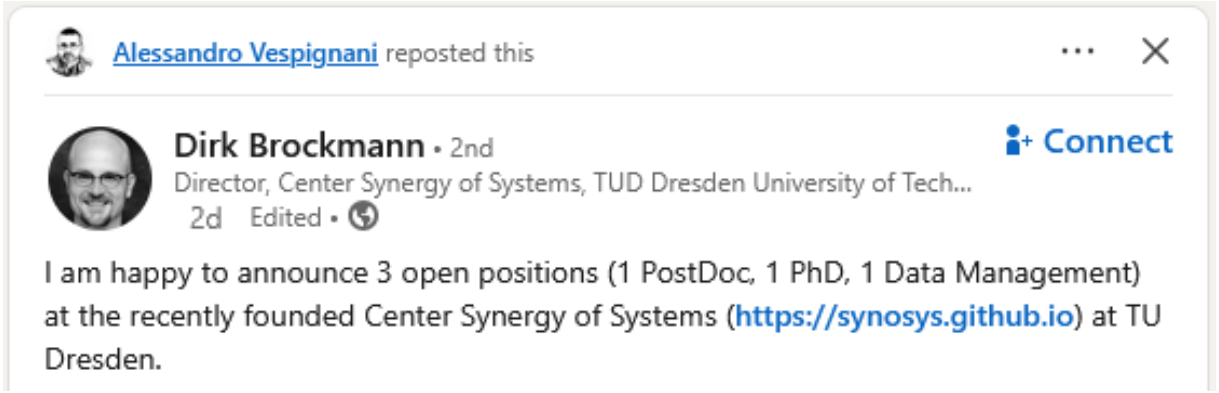
1 Burt; *Res. Org. Behav.* (2000)
2 Karthik et al.; *Science* (2022)
3 Wang, Varol, Eliass-Rad; A*ppl. Net. Sci.* (2022)
4 Stoica, Han, Chaintreau; WWW (2020)
5 Ali et al.; *IEEE TKDE* (2023)

Information Access Equality

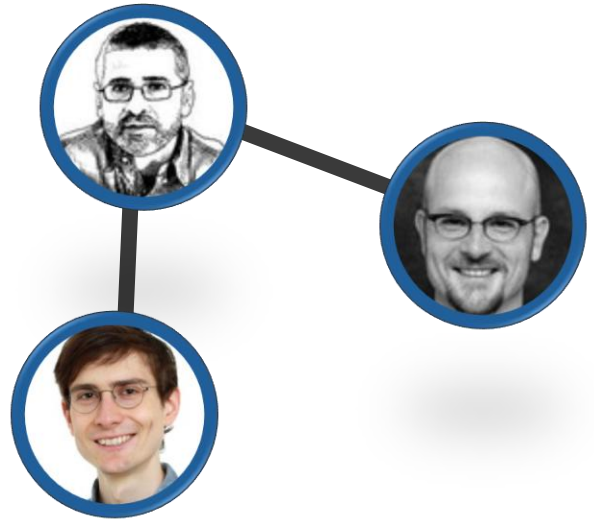
Information is a resource[1,2].

Information Access Equality[3,4] means that members of different groups have equal chance to obtain a new piece of information.

Information is a special resource as
- its value is time-dependent[5].
- it spreads through networks.







1 Burt; *Res. Org. Behav.* (2000)
2 Karthik et al.; *Science* (2022)
3 Wang, Varol, Eliass-Rad; *Appl. Net. Sci.* (2022)
4 Stoica, Han, Chaintreau; WWW (2020)
5 Ali et al.; *IEEE TKDE* (2023)

Higher-Order interactions

System structure matters for many dynamical processes, e.g., contagion, diffusion, opinion dynamics.[1]

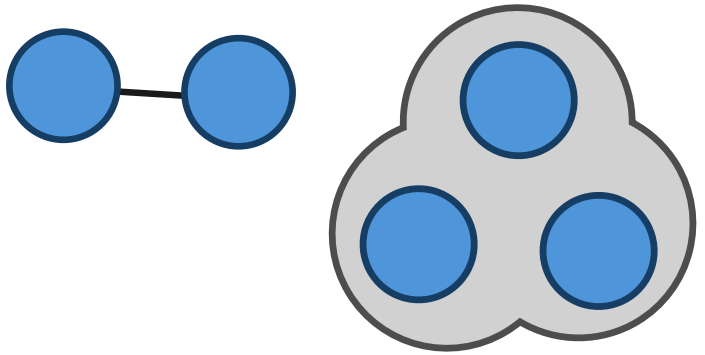
1 Barrat (2008) *Dynamical Processes on Complex Networks*
2 Battiston et al. (2020) Phys. Rep.
3 Battiston et al. (2026) Nat. Rev. Phys.
4 Iacopini et al. (2019) Nat. Comms.
5 St-Onge et al. (2025) arXiv:2507.02758

Higher-Order interactions

System structure matters for many dynamical processes, e.g., contagion, diffusion, opinion dynamics.[1]

Graphs encode structure in terms of pairwise interactions. Hypergraphs are an alternative representation of structure that takes *groups* as the fundamental unit of interactions[2,3].

In many mathematical models this gives rise to new phenomena (e.g., explosive transitions[4]).



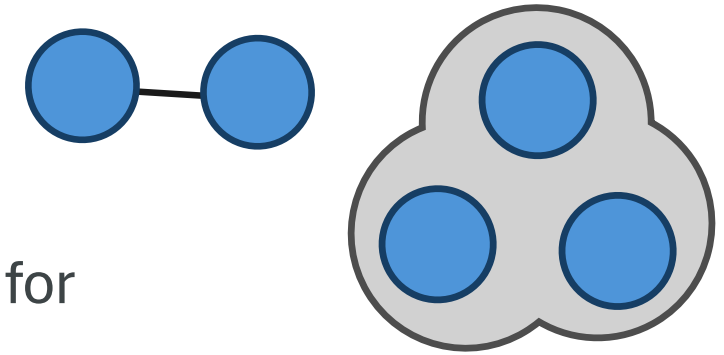
1 Barrat (2008) *Dynamical Processes on Complex Networks*
2 Battiston et al. (2020) Phys. Rep.
3 Battiston et al. (2026) Nat. Rev. Phys.
4 Iacopini et al. (2019) Nat. Comms.
5 St-Onge et al. (2025) arXiv:2507.02758

Higher-Order interactions

System structure matters for many dynamical processes, e.g., contagion, diffusion, opinion dynamics.[1]

Graphs encode structure in terms of pairwise interactions. Hypergraphs are an alternative representation of structure that takes *groups* as the fundamental unit of interactions[2,3].

In many mathematical models this gives rise to new phenomena (e.g., explosive transitions[4]).



But it using hypergraphs for modeling raises also new challenges for interpretation and connecting to theory in the social sciences[5].

1 Barrat (2008) *Dynamical Processes on Complex Networks*
2 Battiston et al. (2020) Phys. Rep.
3 Battiston et al. (2026) Nat. Rev. Phys.
4 Iacopini et al. (2019) Nat. Comms.
5 St-Onge et al. (2025) arXiv:2507.02758

Homophily

One thing that determines who people interact with are their “attributes”. Homophily is the idea that people who are similar interact more[1,2].



1 McPherson et al. (2021) Ann. Rev. Soc.
2 Lawrence & Shah (2020) Acc. Manag. Ann.
3 Rivera et al. (2010) Ann. Rev. Soc.
4 Ertug et al.(2022) Acc. Manag. Ann.

Homophily

One thing that determines who people interact with are their “attributes”. Homophily is the idea that people who are similar interact more[1,2].



Homophily has been documented in many settings and with respect to different attributes (e.g., gender, race, political leaning, socio economic status, etc.), and studies also explore its effects on outcomes and influence as a tie formation mechanism.[1,3,4]

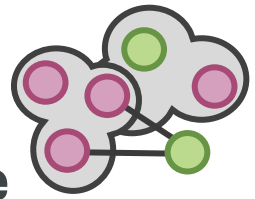
1 McPherson et al. (2021) Ann. Rev. Soc.
2 Lawrence & Shah (2020) Acc. Manag. Ann.
3 Rivera et al. (2010) Ann. Rev. Soc.
4 Ertug et al.(2022) Acc. Manag. Ann.

Higher-order Homophily and Information Access

What is the **impact of higher-order interactions and homophily on information access equality**?

Higher-order Homophily and Information Access

What is the **impact of higher-order interactions and homophily on information access equality**?

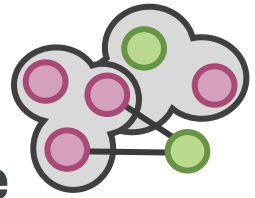


Structure

We propose a hypergraph model that has a fixed value of homophily and tunable degree homophily.

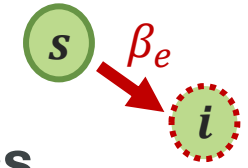
Higher-order Homophily and Information Access

What is the **impact of higher-order interactions and homophily on information access equality**?



Structure

We propose a hypergraph model that has a fixed value of homophily and tunable degree homophily.

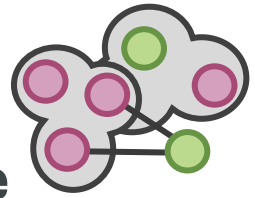


Dynamics

We generalize exiting social contagion models to account for group-dependent transmission.

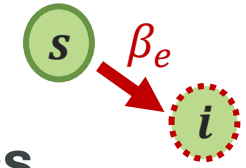
Higher-order Homophily and Information Access

What is the **impact of higher-order interactions and homophily on information access equality**?



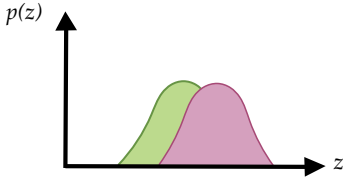
Structure

We propose a hypergraph model that has a fixed value of homophily and tunable degree homophily.



Dynamics

We generalize exiting social contagion models to account for group-dependent transmission.



Measurement

we conduct simulations on synthetic and real-world hypergraphs and measure the effect of homophily, degree-heterogeneity, and class-imbalance on fair information access.

Homophily

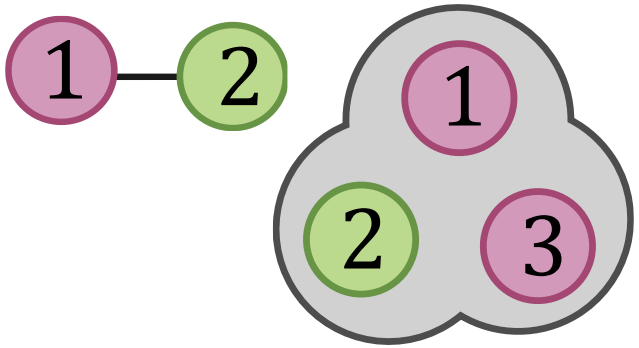
Homophily is the **above-chance** occurrence of interactions between similar nodes.

Homophily

Homophily is the **above-chance** occurrence of interactions between similar nodes.

Connectivity

(Hyper)edges indicating, e.g., co-authorship, discussions, teams, …

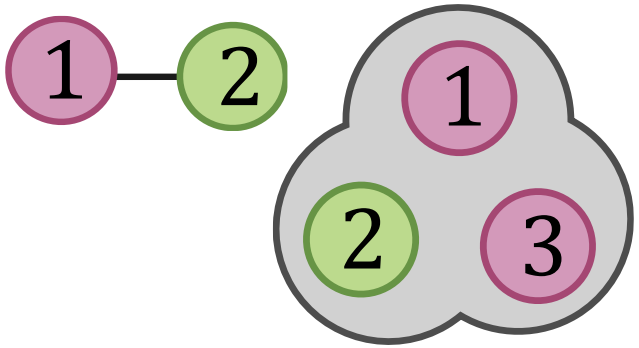


Homophily

Homophily is the **above-chance** occurrence of interactions between similar nodes.

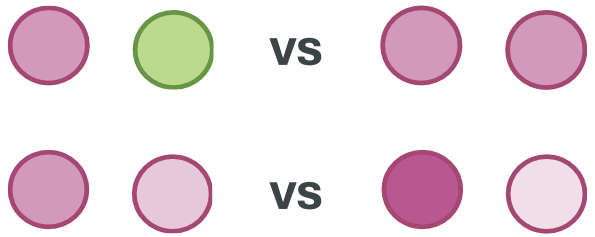
Connectivity

(Hyper)edges indicating, e.g., co-authorship, discussions, teams, …



Similarity

- *discrete attributes*: equality
- *continuous attributes*: inner products and distances

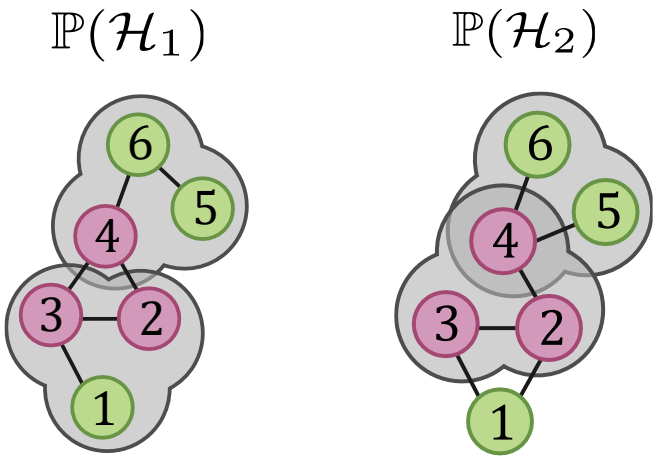


Homophily

Homophily is the **above-chance** occurrence of interactions between similar nodes.

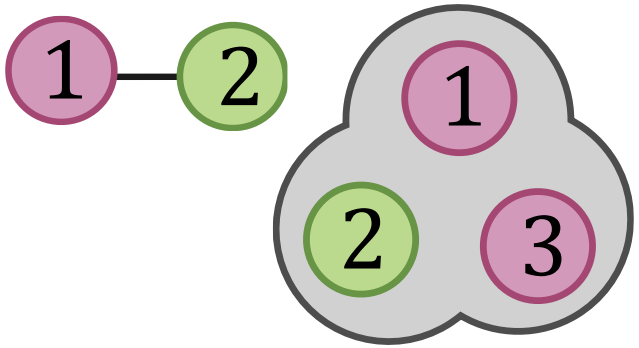
Null model

preserve relative class sizes, and structural properties, e.g., average degree.



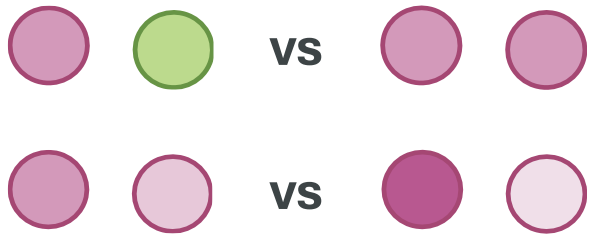
Connectivity

(Hyper)edges indicating, e.g., co-authorship, discussions, teams, …



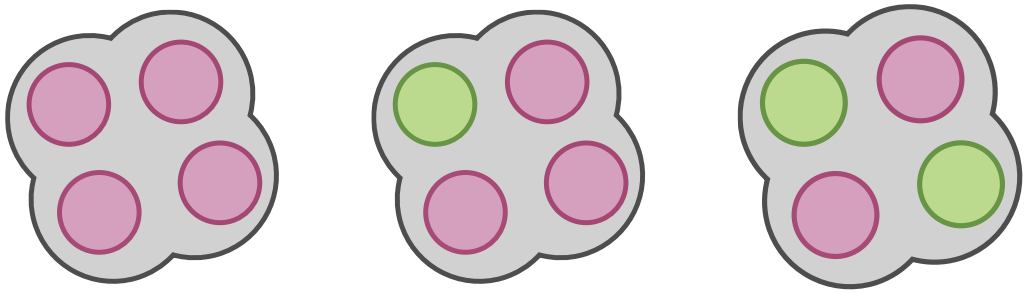
Similarity

- *discrete attributes*: equality
- *continuous attributes*: inner products and distances



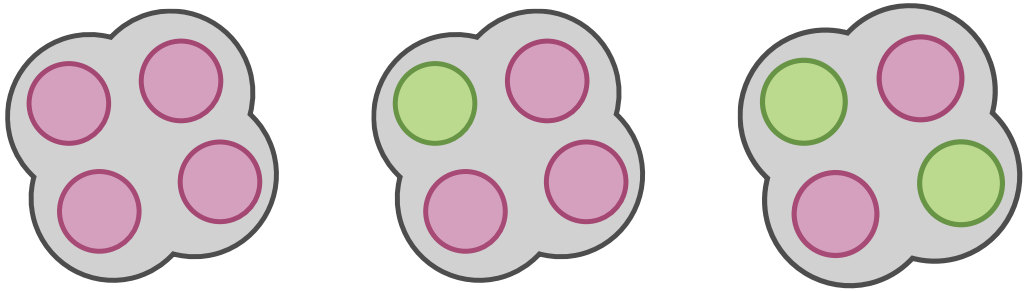
Higher-Order Homophily

Do **within-class** hyperedges make sense?

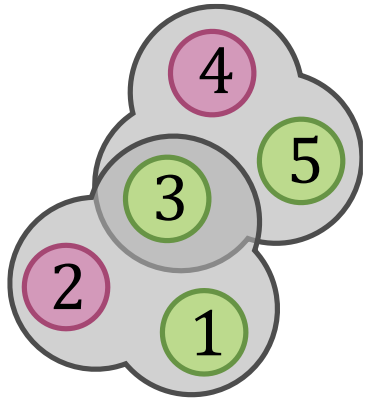


Higher-Order Homophily

Do **within-class** hyperedges make sense?

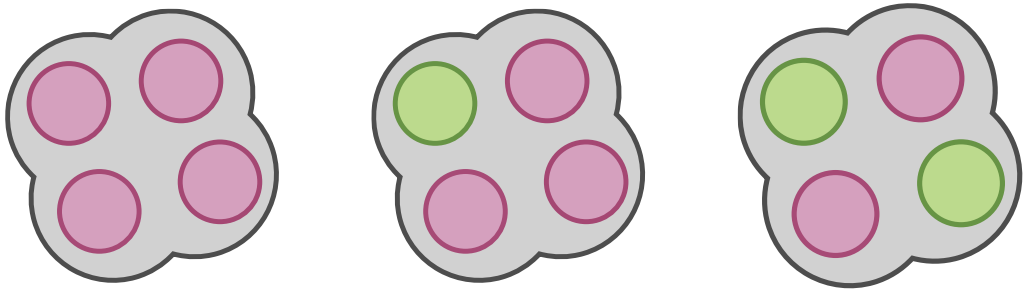


There are combinatorial **constraints**.

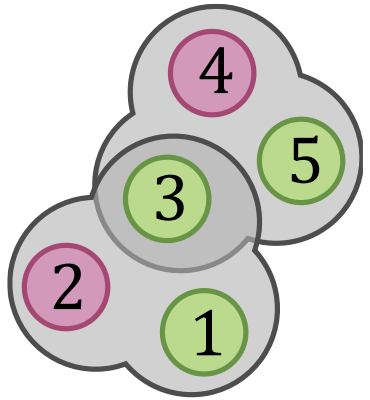


Higher-Order Homophily

Do **within-class** hyperedges make sense?



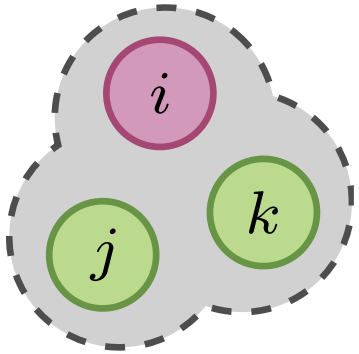
There are combinatorial **constraints**.



There are **a lot** of possible hyperedges.

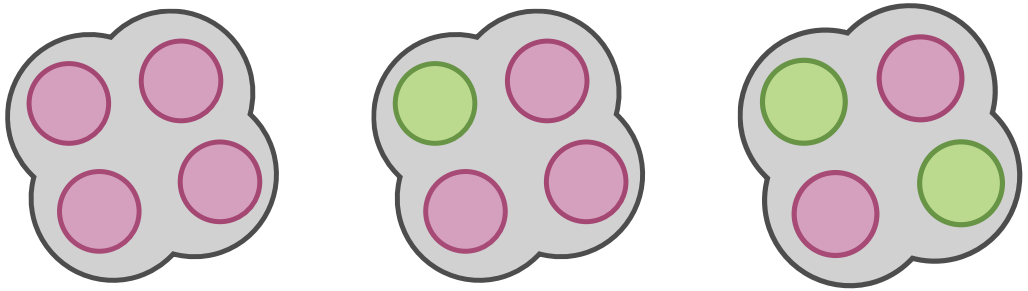
$$E^{\max}_{n,\alpha} = \binom{n}{\alpha}$$

$$\frac{E^{\max}_{1000,4}}{E^{\max}_{1000,2}} \approx 8 \cdot 10^4$$

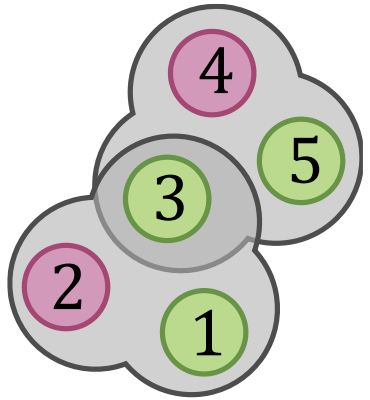


Higher-Order Homophily

Do **within-class** hyperedges make sense?



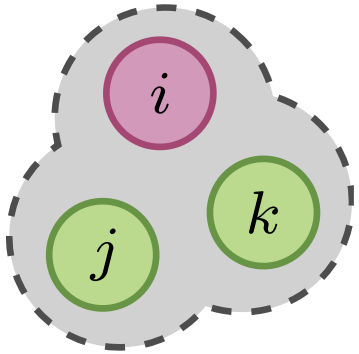
There are combinatorial **constraints**.



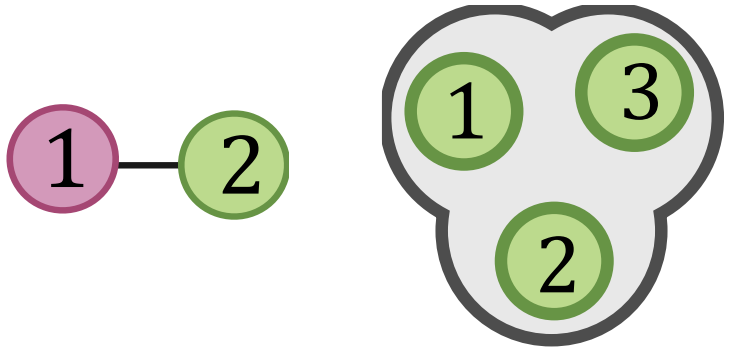
There are **a lot** of possible hyperedges.

$$E^{\max}_{n,\alpha} = \binom{n}{\alpha}$$

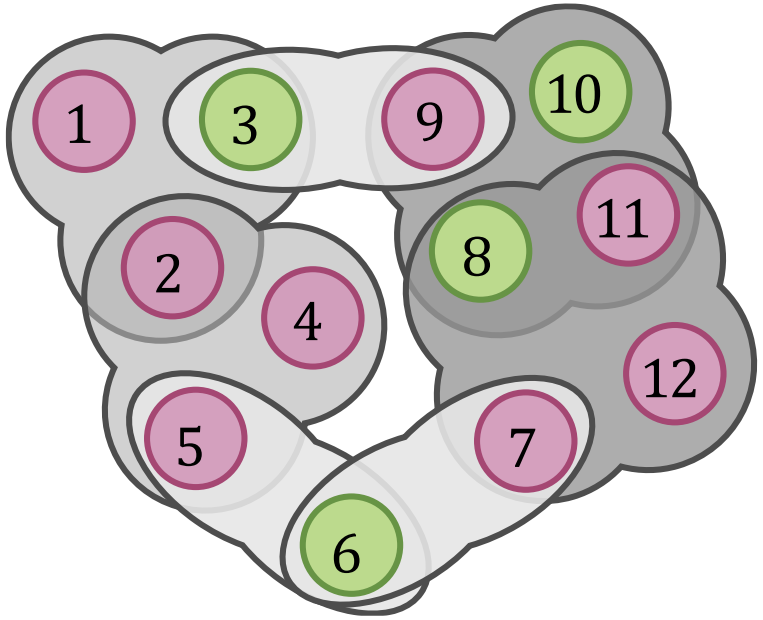
$$\frac{E^{\max}_{1000,4}}{E^{\max}_{1000,2}} \approx 8 \cdot 10^4$$



Mixing patterns can **vary by hyperedge size**.



Hypergraph Homophily



Properties

$N = 12$ number of nodes

$c \in \{0, 1\}$ two classes

$N_0 = 8$ number of class 0 nodes

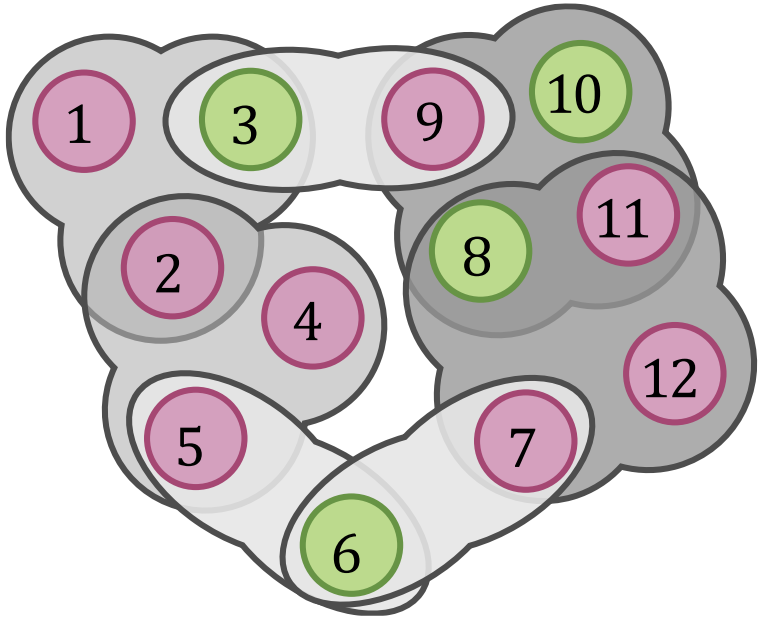
$N_1 = 4$ number of class 1 nodes

$E = 7$ number of hyperedges

$\alpha \in \{2,3,4\}$ hyperedge sizes

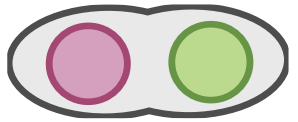
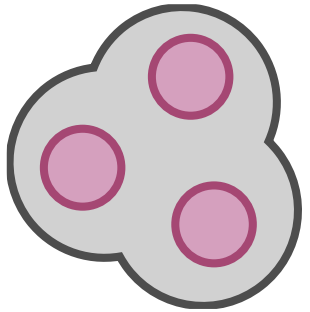
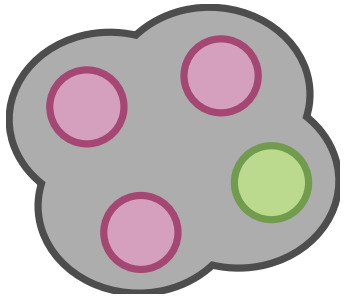
Veldt et al. Sci. Adv. **9** 1 (2023)

Hypergraph Homophily



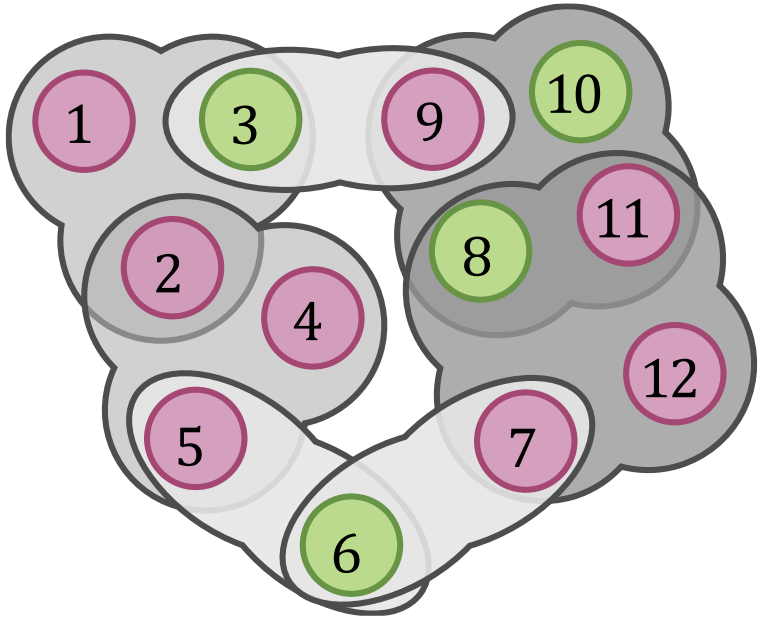
Number of hyperedges of type

$$E_{\alpha,t}^{(c)} = |\{e \in E : t_c(e) = t \wedge |e| = \alpha\}|$$

	Size	Type
	$\alpha = 2$	$t_0(e) = 1$ $t_1(e) = 1$
	$\alpha = 3$	$t_0(e) = 3$ $t_1(e) = 0$
	$\alpha = 4$	$t_0(e) = 3$ $t_1(e) = 1$

Veldt et al. Sci. Adv. **9** 1 (2023)

Hypergraph Homophily



$$a_{c,\alpha,t}^{\mathrm{VBK1}} = \frac{tE_{\alpha,t}^{(c)}}{\sum_{t'=1}^{\alpha} t' E_{\alpha,t'}^{(c)}}$$

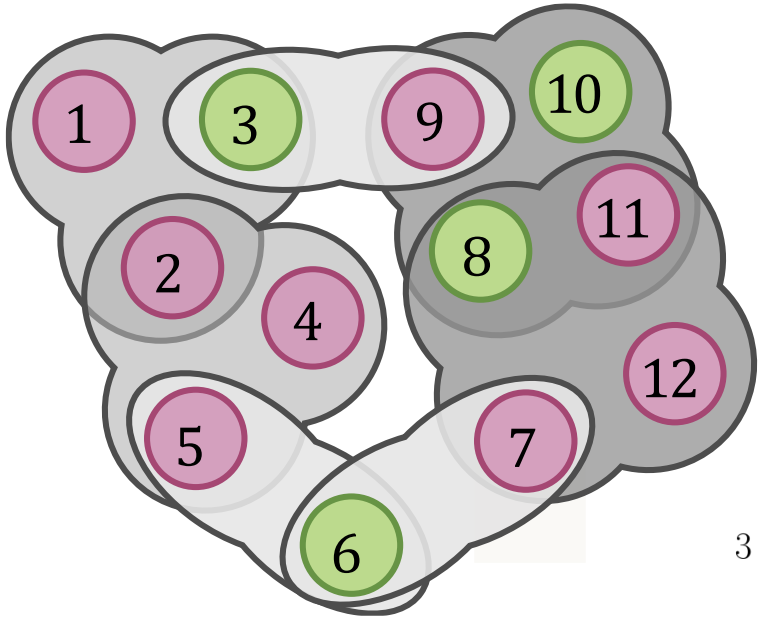
$$b_{c,\alpha,t}^{\mathrm{VBK1}} = \frac{\binom{N_c-1}{t-1}\binom{N-N_c}{\alpha-t}}{\binom{N-1}{\alpha-1}}$$

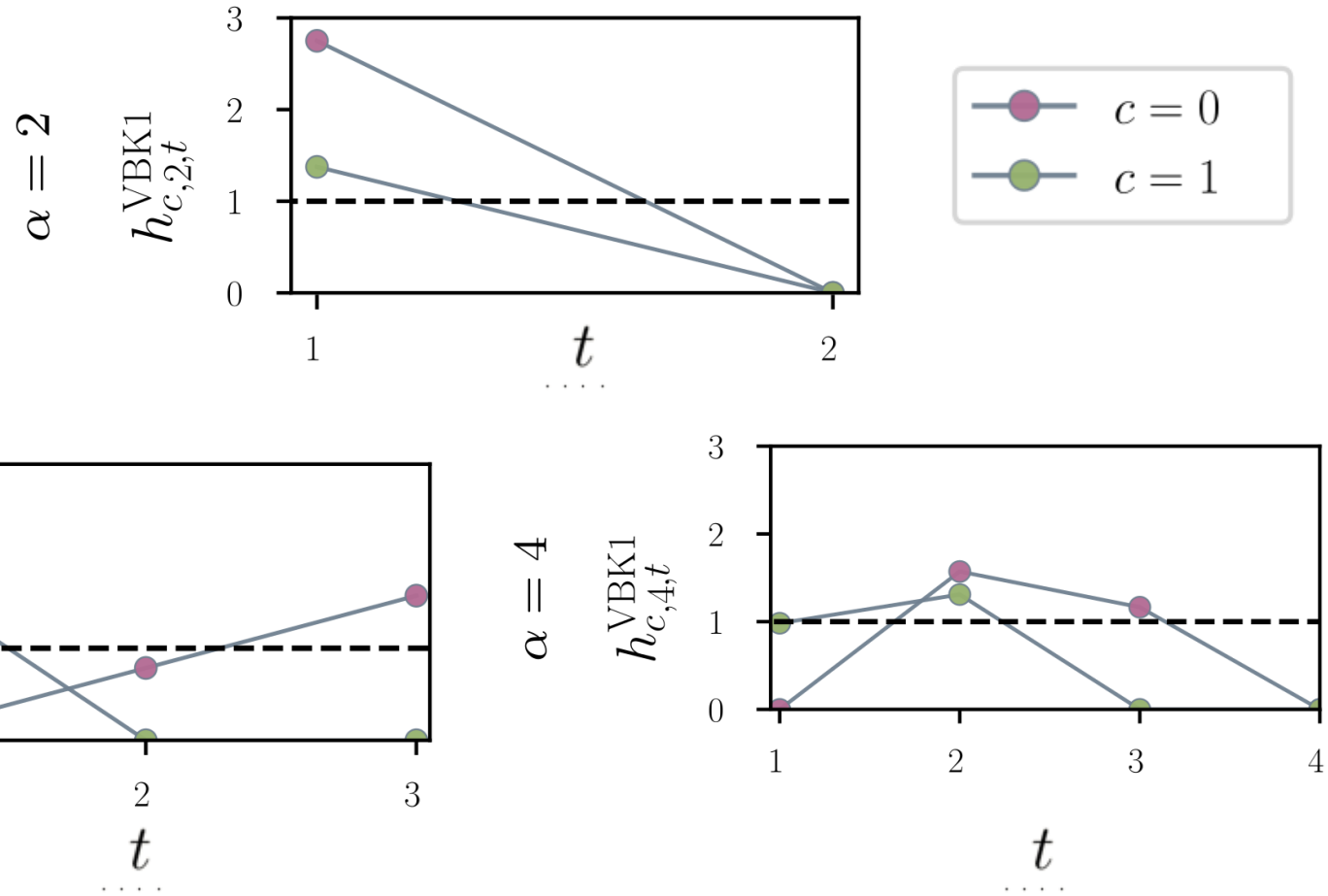
$$h_{c,\alpha,t}^{\mathrm{VBK1}} = \frac{a_{c,\alpha,t}^{\mathrm{VBK1}}}{b_{c,\alpha,t}^{\mathrm{VBK1}}}$$

c	α	t	$E_{\alpha,t}^{(c)}$	$a_{c,\alpha,t}^{\mathrm{VBK1}}$	$b_{c,\alpha,t}^{\mathrm{VBK1}}$	$h_{c,\alpha,t}^{\mathrm{VBK1}}$
0	2	0	0	–	–	–
		1	3	3/3	4/11	11/4
		2	0	0/3	7/11	0
	3	0	0	–	–	–
		1	0	0/5	6/55	0
		2	1	2/5	28/55	11/14
		3	1	3/5	21/55	22/14
	4	0	0	–	–	–
		1	0	0/5	4/165	0
		2	1	2/5	42/165	33/21
		3	1	3/5	85/165	99/85
		4	0	0/5	35/165	0
1	2	0	0	–	–	–
		1	3	3/3	8/11	11/8
		2	0	0/3	3/11	0
	3	0	0	–	–	–
		1	1	1/1	28/55	55/28
		2	0	0/1	24/55	0
		3	0	0/1	3/55	0
	4	0	0	–	–	–
		1	1	1/3	56/165	55/56
		2	1	2/3	84/165	55/42
		3	0	0/3	24/165	0
		4	0	0/3	1/165	0

Veldt et al. Sci. Adv. **9** 1 (2023)

Hypergraph Homophily





Veldt et al. Sci. Adv. **9** 1 (2023)

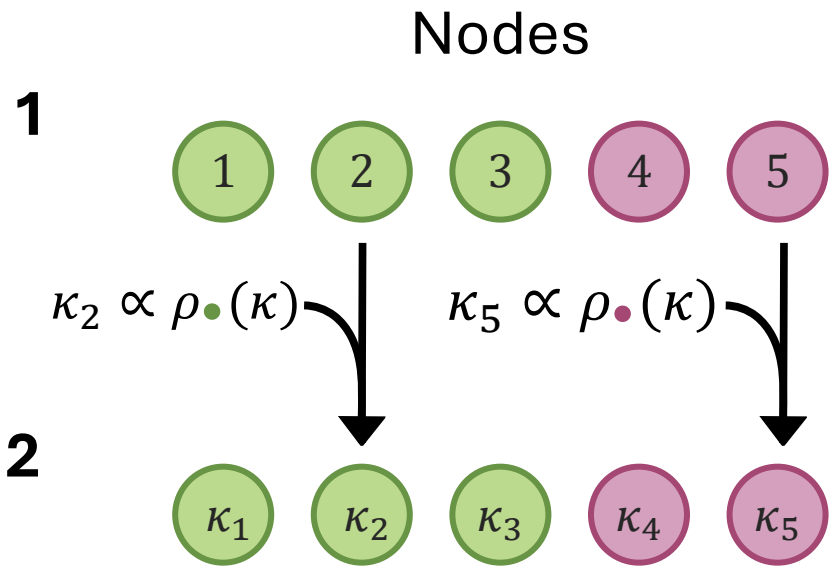
The H3 Model

Nodes

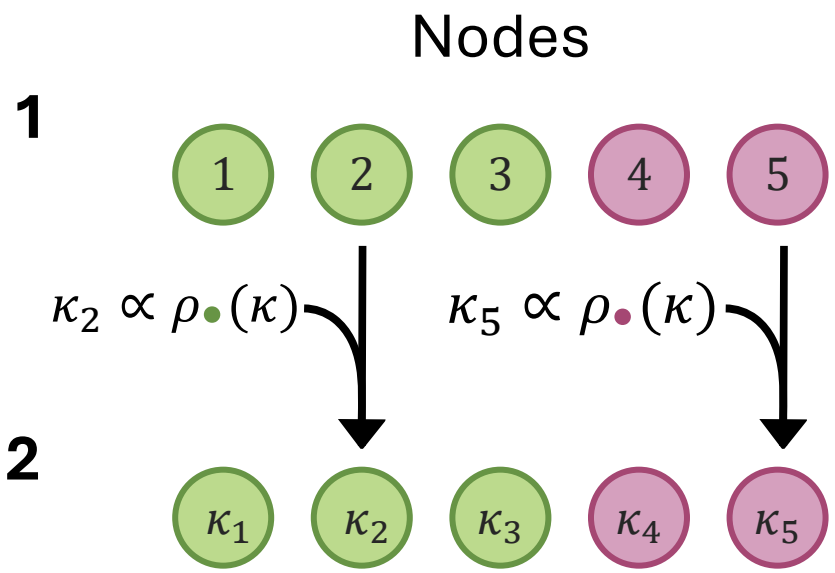
1

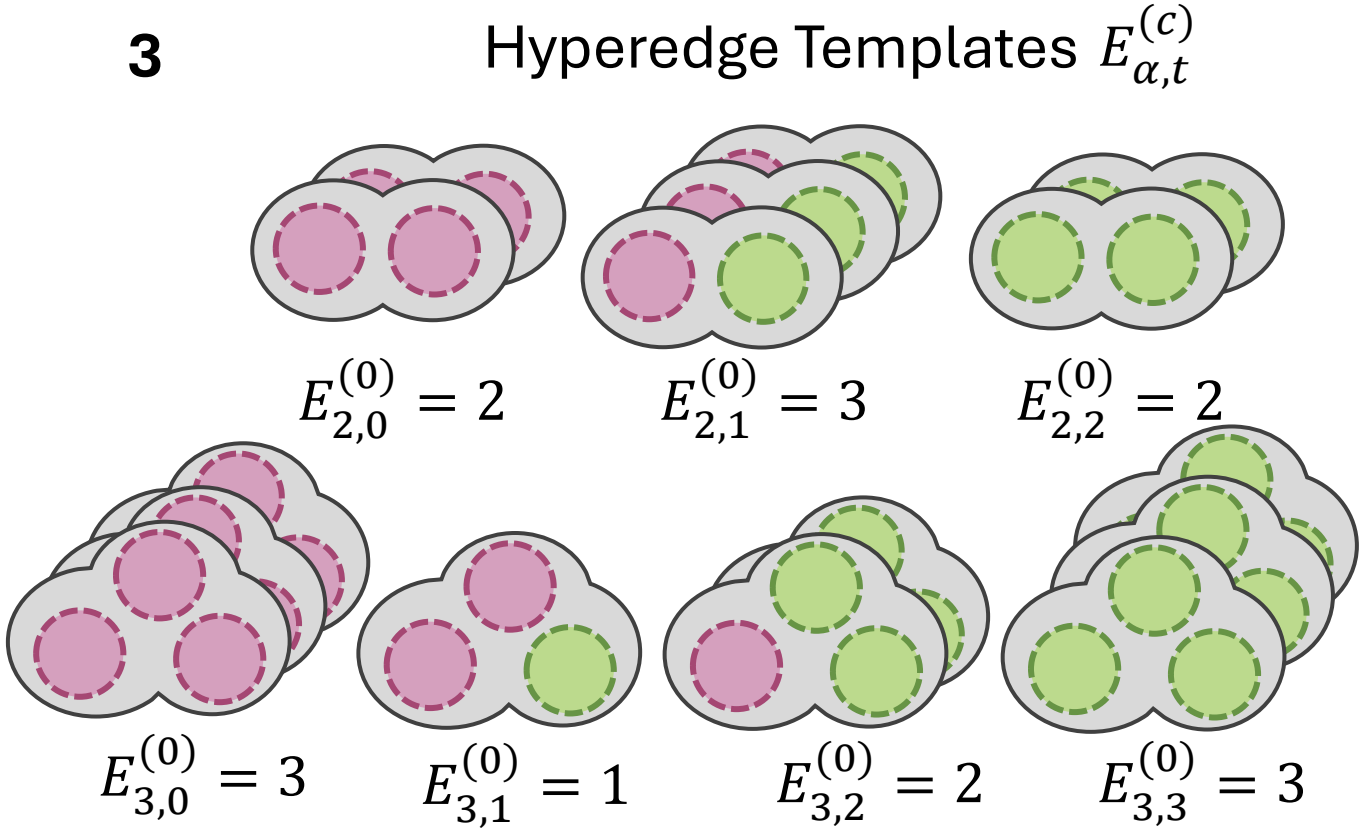


The H3 Model

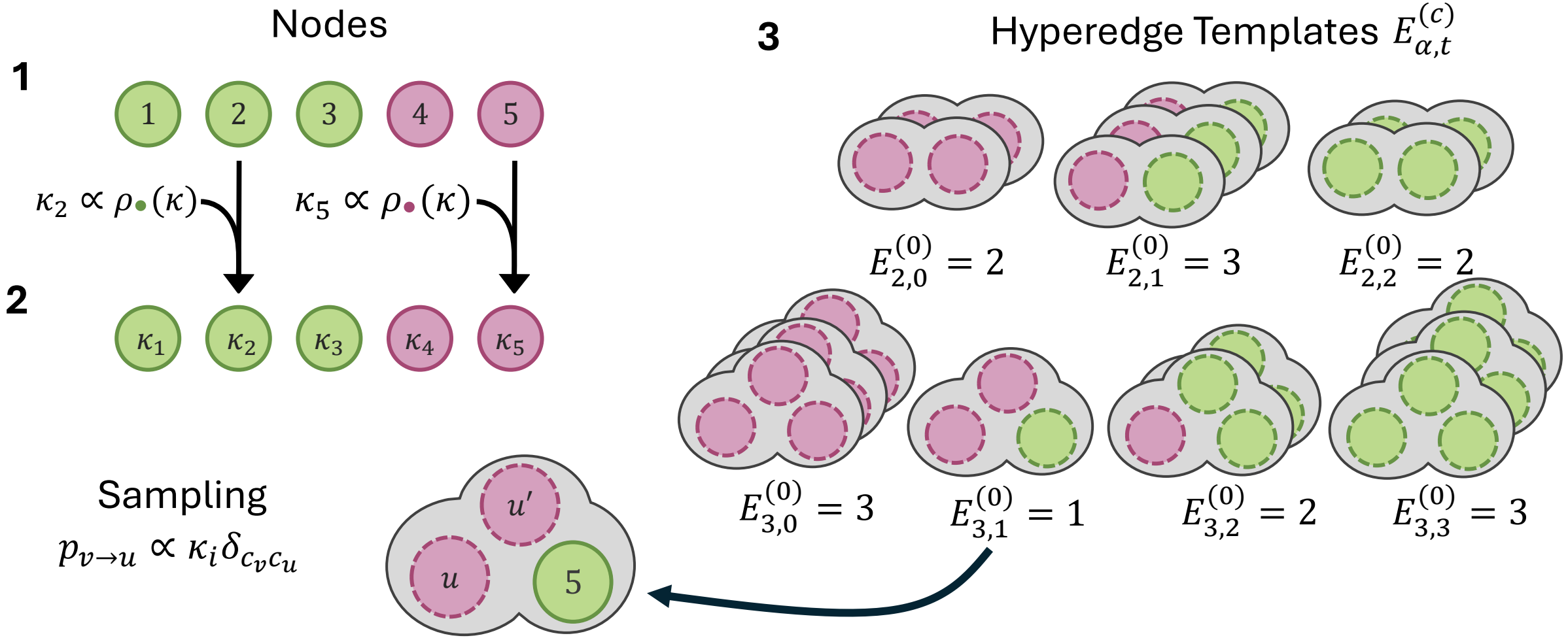

Nodes
1
1
2
3
4
5
$\kappa_2 \propto \rho_{\bullet}(\kappa)$
$\kappa_5 \propto \rho_{\bullet}(\kappa)$
2
κ_1
κ_2
κ_3
κ_4
κ_5

The H3 Model


Nodes
1
1
2
3
4
5
$\kappa_2 \propto \rho_\bullet(\kappa)$
$\kappa_5 \propto \rho_\bullet(\kappa)$
2
κ_1
κ_2
κ_3
κ_4
κ_5


3
Hyperedge Templates $E^{(c)}_{\alpha,t}$
$E^{(0)}_{2,0} = 2$
$E^{(0)}_{2,1} = 3$
$E^{(0)}_{2,2} = 2$
$E^{(0)}_{3,0} = 3$
$E^{(0)}_{3,1} = 1$
$E^{(0)}_{3,2} = 2$
$E^{(0)}_{3,3} = 3$

The H3 Model


Nodes
1
1
2
3
4
5
$\kappa_2 \propto \rho_\bullet(\kappa)$
$\kappa_5 \propto \rho_\bullet(\kappa)$
2
κ_1
κ_2
κ_3
κ_4
κ_5
3
Hyperedge Templates $E^{(c)}_{\alpha,t}$
$E^{(0)}_{2,0} = 2$
$E^{(0)}_{2,1} = 3$
$E^{(0)}_{2,2} = 2$
$E^{(0)}_{3,0} = 3$
$E^{(0)}_{3,1} = 1$
$E^{(0)}_{3,2} = 2$
$E^{(0)}_{3,3} = 3$
Sampling
$p_{v\to u} \propto \kappa_i \delta_{c_v c_u}$
u'
u
5

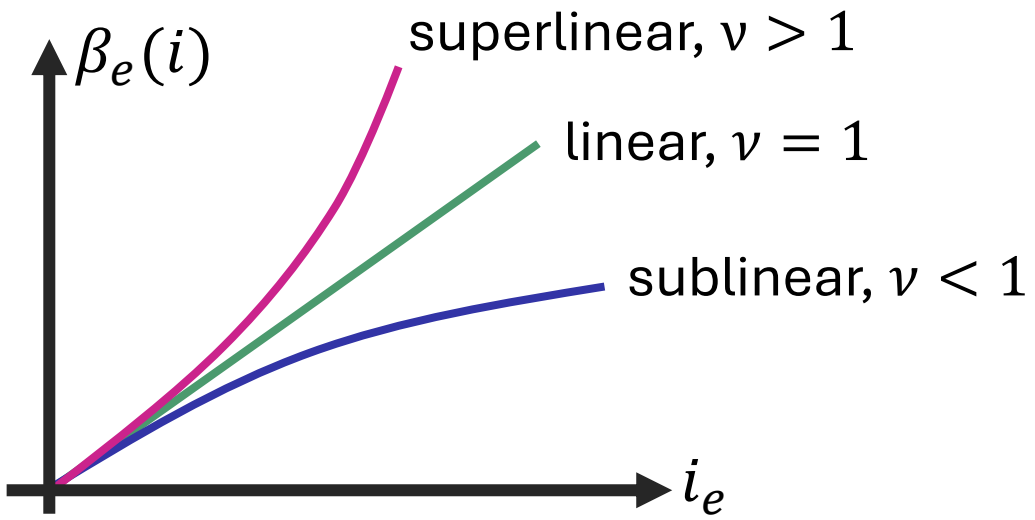
Social Contagion on Hypergraphs

In contagion models on hypergraphs [1,2] the transmission probability depends non-linearly on the number of informed individuals in an hyperedge.

Rate β_e at which uninformed individuals in edge e acquire information:

$$\beta_e = \lambda i_e(t)^{\nu} s_e(t)$$

Complex Contagion: Transmission rates grow non-linearly with the number of informed nodes.



1 St-Onge et al.; *Comms.Phys.* (2022)
2 de Arruda, Aleta, Moreno; *Nat.Rev.Phys.* (2024)

Social Contagion on Hypergraphs

In contagion models on hypergraphs [1,2] the transmission probability depends non-linearly on the number of informed individuals in an hyperedge.

Rate $\beta_e^{(g)}$ at which uninformed individuals in edge e and group g acquire information:

$$\beta_e^{(g)} = \left(\lambda_{\text{in}} i_{e,g}^{\nu_{\text{in}}}(t) + \lambda_{\text{out}} i_{e,g'}^{\nu_{\text{out}}}(t)\right) s_{e,g}(t)$$

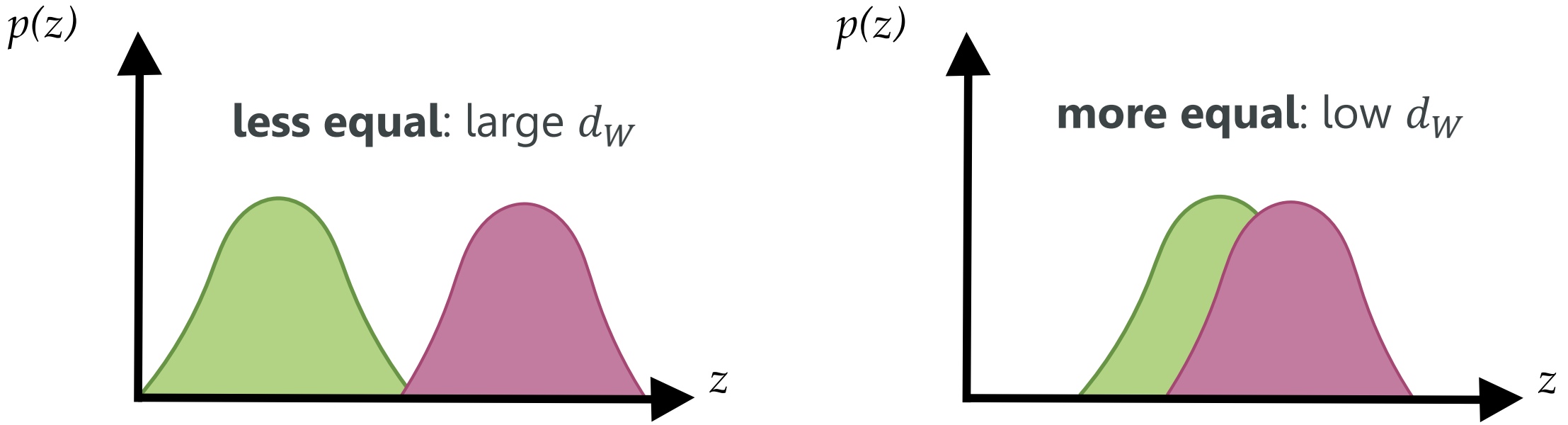
Complex Contagion: Transmission rates grow non-linearly with the number of informed nodes.

Asymmetry: Different parameters for the interaction with in and out group.

1 St-Onge et al.; *Comms.Phys.* (2022)
2 de Arruda, Aleta, Moreno; *Nat.Rev.Phys.* (2024)

Measuring Information Access Equality

Information access equality can be measured by the Wasserstein distance $d_W(Z^{\bullet}, Z^{\bullet})$ between distributions of the order $\{z_i\}_{i\in[n]}$ in which nodes are informed.



Measuring Information Acquisition

Acquisition fairness[1] $\alpha(f)$ measures which group is better in accessing information.

It is the fraction of informed nodes in the minority compared to the fraction of all informed nodes.

$$\alpha(f) = \left\langle \frac{i_\bullet(t_f)/n_\bullet}{i(t_f)/n} \right\rangle_{\mathrm{U}[\nu]}$$

1 Zappala et al. (2024) arXiv:2407.17441

Measuring Information Acquisition

Acquisition fairness[1] $\alpha(f)$ measures which group is better in accessing information.

It is the fraction of informed nodes in the minority compared to the fraction of all informed nodes.

$$\alpha(f) = \left\langle \frac{i_\bullet(t_f)/n_\bullet}{i(t_f)/n} \right\rangle_{\mathrm{U}[\mathcal{V}]}$$

t_f is the time to inform a fraction f of all nodes

Seeds chosen uniformly at random among all nodes $\mathcal{V}$.

1 Zappala et al. (2024) arXiv:2407.17441

Measuring Information Acquisition

Acquisition fairness[1] $\alpha(f)$ measures which group is better in accessing information.

It is the fraction of informed nodes in the minority compared to the fraction of all informed nodes.

$$\alpha(f) = \left\langle \frac{i_\bullet(t_f)/n_\bullet}{i(t_f)/n} \right\rangle_{\mathrm{U}[\nu]}$$

A value $\alpha(f) > 1$ majority advantage and $\alpha(f) < 1$ minority advantage.

1 Zappala et al. (2024) arXiv:2407.17441

Measuring Information Dissemination

Diffusion fairness[1] $\delta(f)$ measures which group is better in disseminating information.

It is the expected time it takes to reach a fraction f of all nodes when seeds are chosen in the minority compared to when they are chosen in the majority.

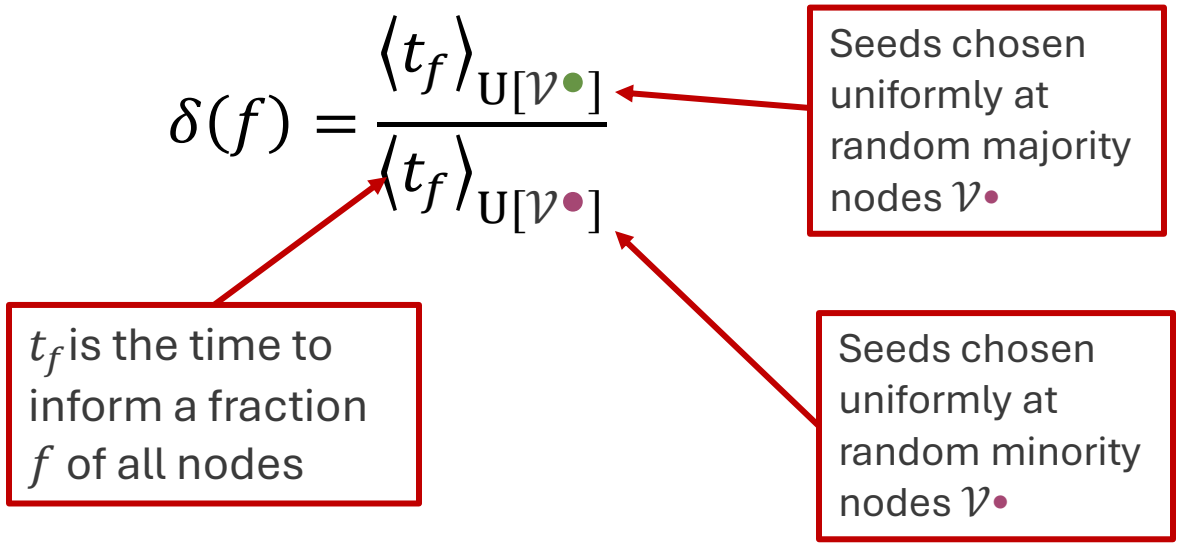
$$\delta(f) = \frac{\langle t_f \rangle_{\mathrm{U}[\mathcal{V}\bullet]}}{\langle t_f \rangle_{\mathrm{U}[\mathcal{V}\bullet]}}$$

1 Zappala et al. (2024) arXiv:2407.17441

Measuring Information Dissemination

Diffusion fairness[1] $\delta(f)$ measures which group is better in disseminating information.

It is the expected time it takes to reach a fraction f of all nodes when seeds are chosen in the minority compared to when they are chosen in the majority.



1 Zappala et al. (2024) arXiv:2407.17441

Measuring Information Dissemination

Diffusion fairness[1] $\delta(f)$ measures which group is better in disseminating information.

It is the expected time it takes to reach a fraction f of all nodes when seeds are chosen in the minority compared to when they are chosen in the majority.

$$\delta(f) = \frac{\langle t_f \rangle_{\mathrm{U}[\mathcal{V}\bullet]}}{\langle t_f \rangle_{\mathrm{U}[\mathcal{V}\bullet]}}$$

A value $\delta(f) > 1$ majority advantage and $\delta(f) < 1$ minority advantage.

1 Zappala et al. (2024) arXiv:2407.17441

Scenarios

Homophily:

- *neutral*: as expected if edges form at random.
- *homophilous*: more edges with only one type of node.
- *heterophilous*: less edges with only one type of node.
- *real-world*: as found in empirical hypergraphs.
- *mixed:* different patterns at each edge size.

Dynamics:

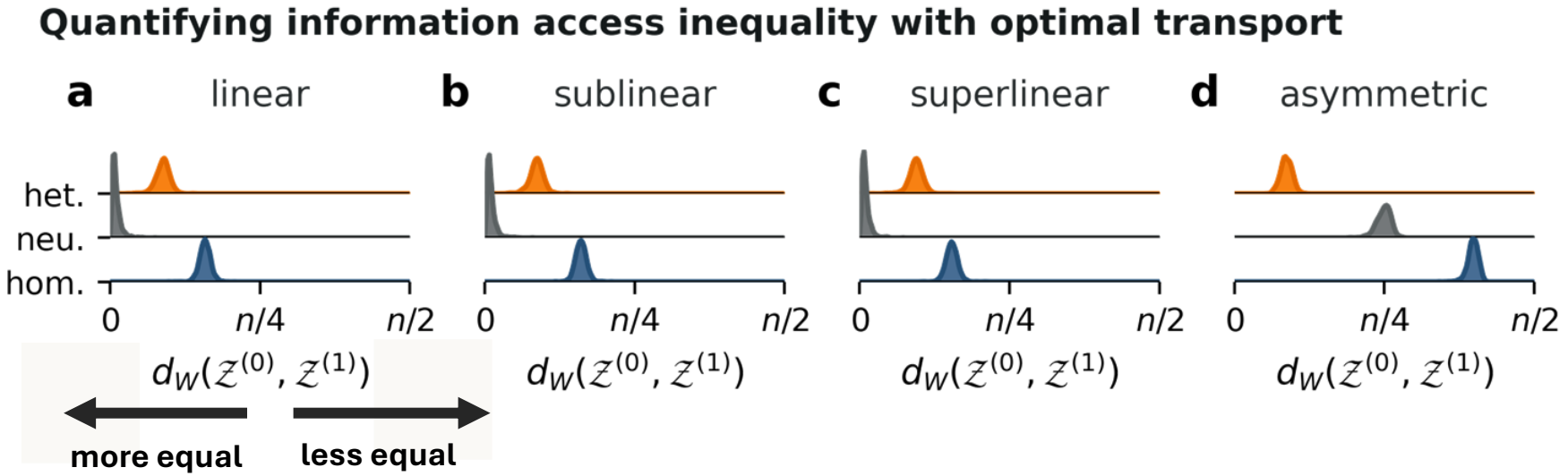
- *symmetric, linear:* similar to simple contagion on graphs.
- *symmetric, superlinear*: social reinforcement.
- *symmetric, sublinear*: social inhibition..
- *asymmetric, non-linear*: in-group social reinforcement, out-group social inhibition

We use $n_{\mathrm{hg}} = 10^3$ hypergraphs per setting with $n = 10^4$ nodes, 75% majority, and average degree $\bar{k} = 11$ and hidden degrees sampled from a Pareto with pdf-exponent $\gamma = 2.9$

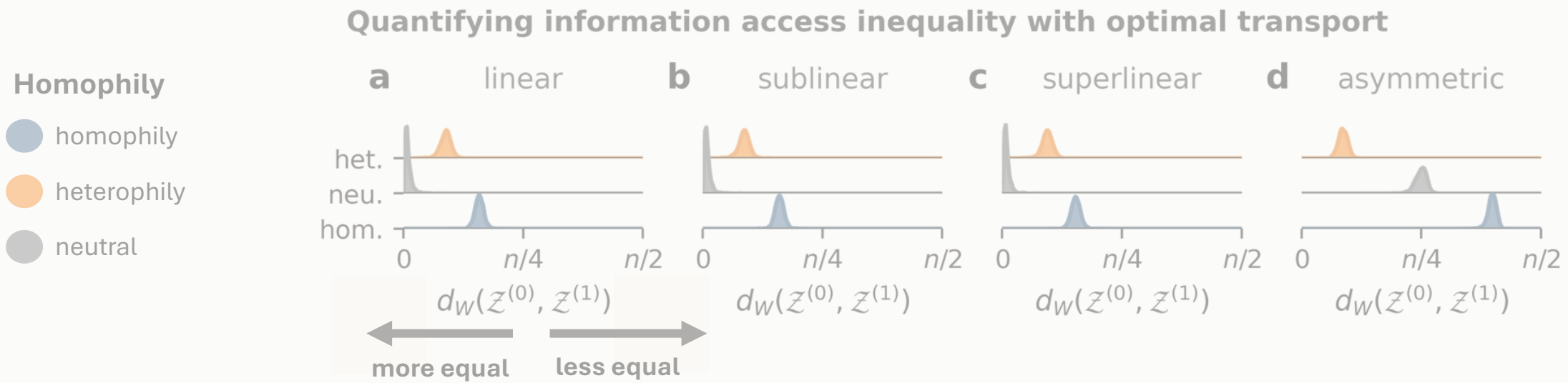
Information Access Equality in Hypergraphs

Homophily

- homophily
- heterophily
- neutral



Information Access Equality in Hypergraphs



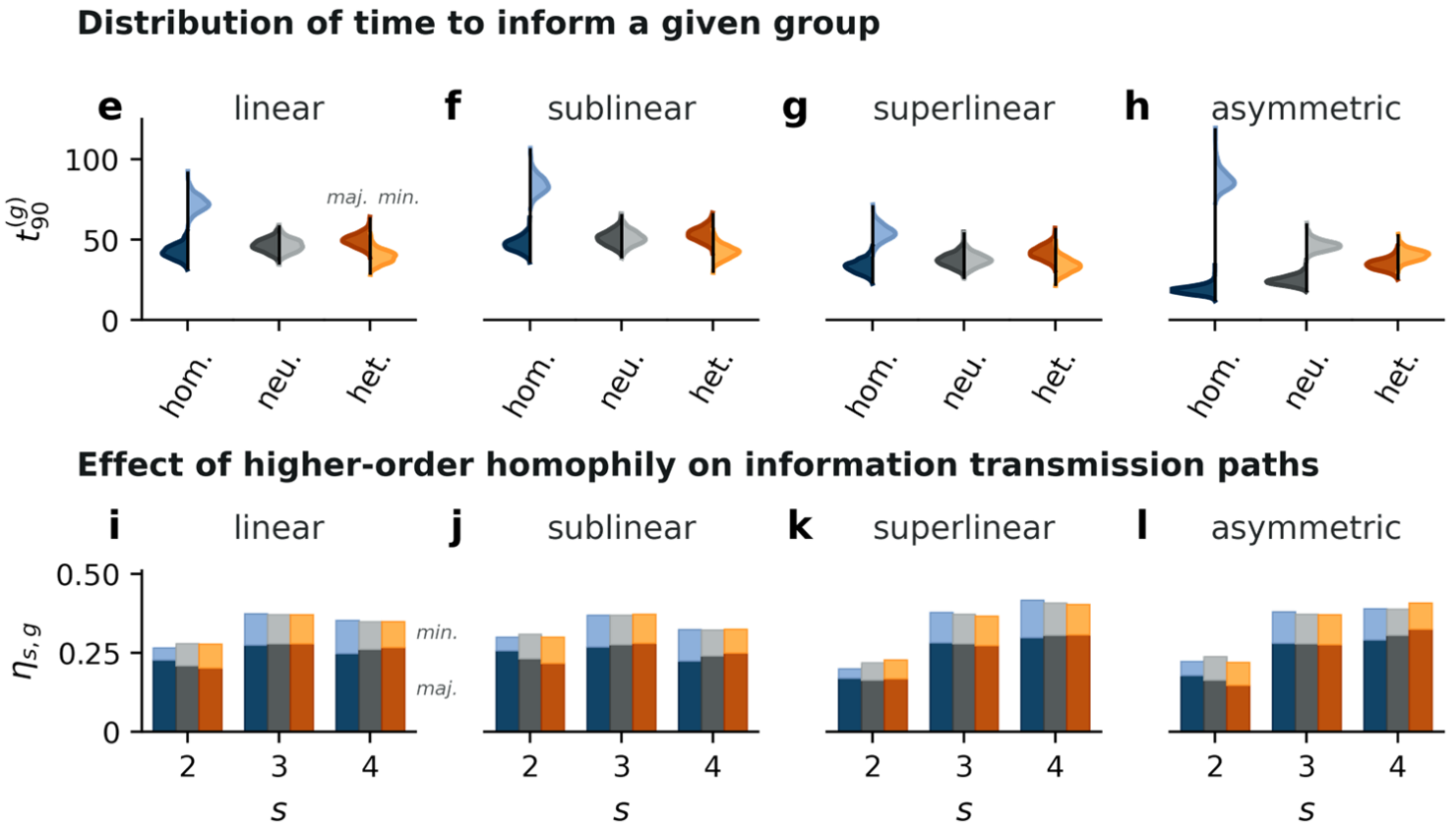
Takeaway

- Asymmetric transmission leads to the largest inequality irrespective of homophily pattern.
- Homophilous hypergraphs show largest inequality.
- Observed inequality is similar under linear, sublinear, and superlinear contagion.

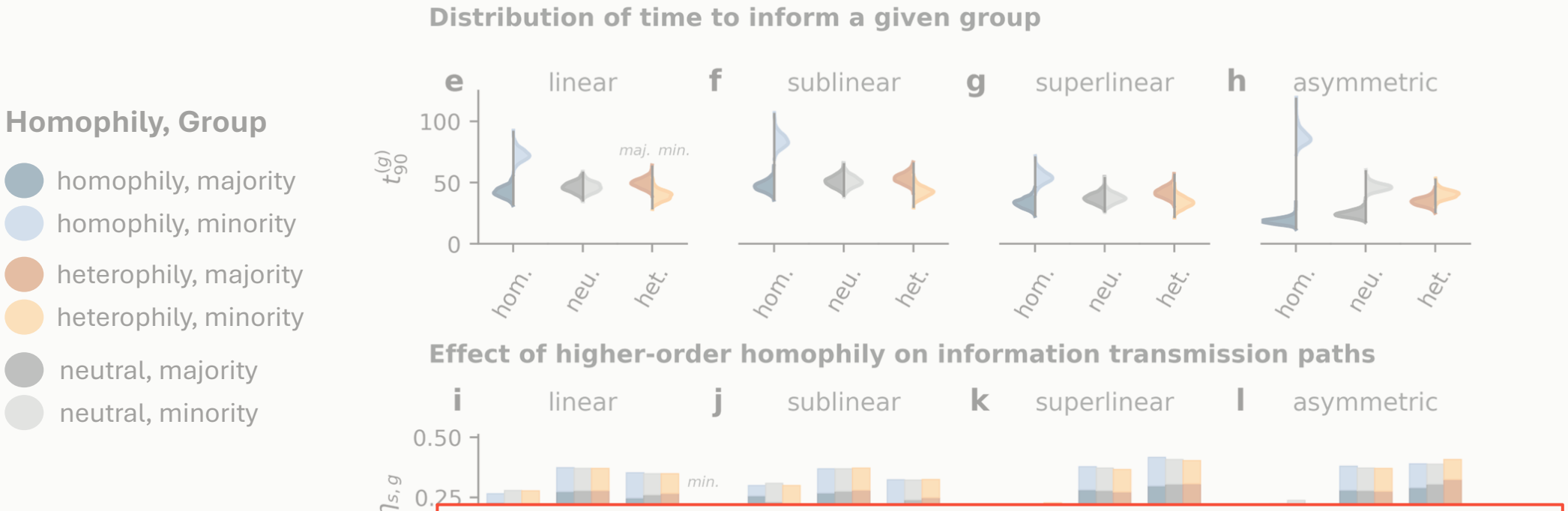
Differences in mechanism

Homophily, Group

- homophily, majority
- homophily, minority
- heterophily, majority
- heterophily, minority
- neutral, majority
- neutral, minority



Differences in mechanism



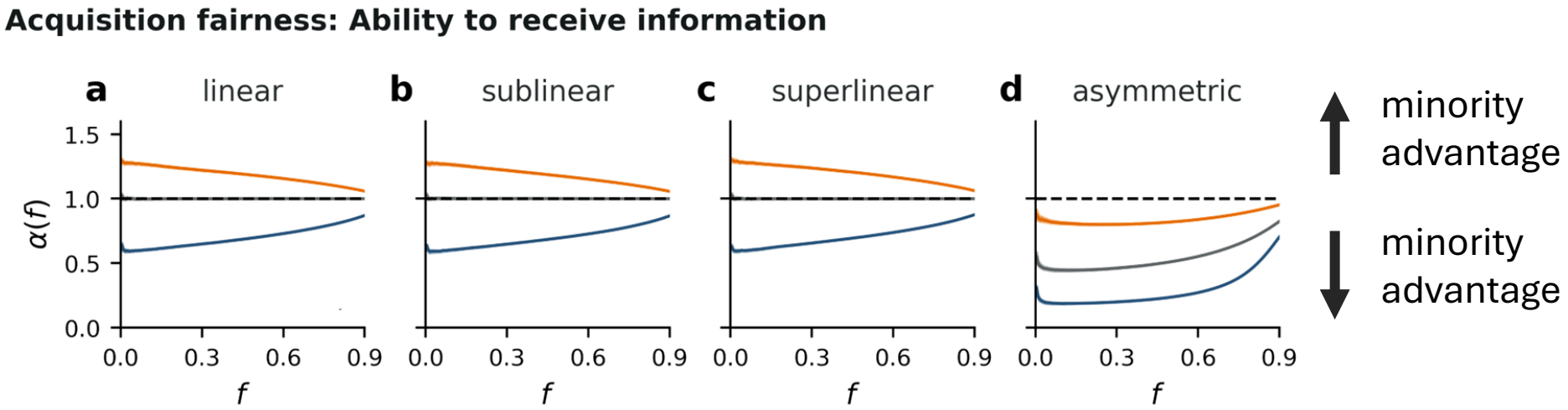
Takeaway

- Time-scale and mechanism of information spread differ in all cases.
- Superlinear contagion prefers large edges.

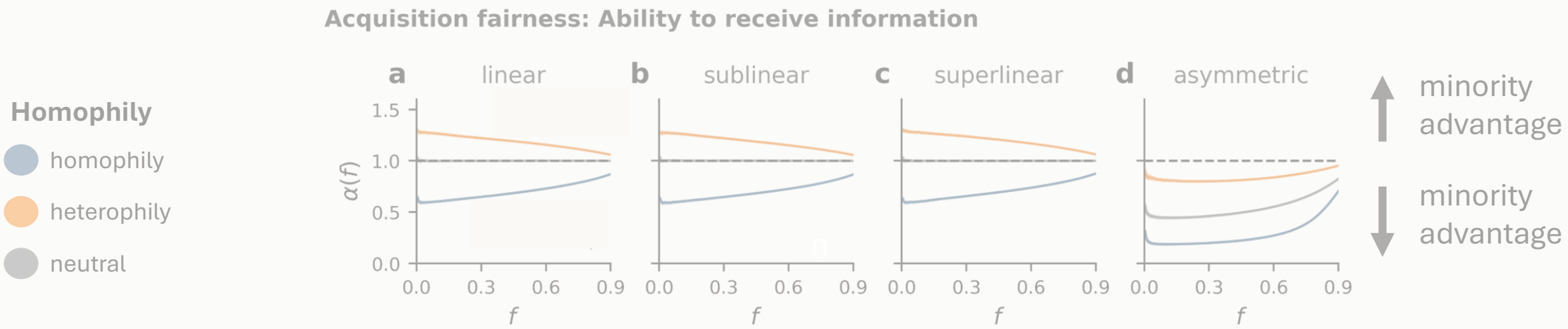
Differences in Information Acquisition

Homophily

- homophily
- heterophily
- neutral



Differences in Information Acquisition

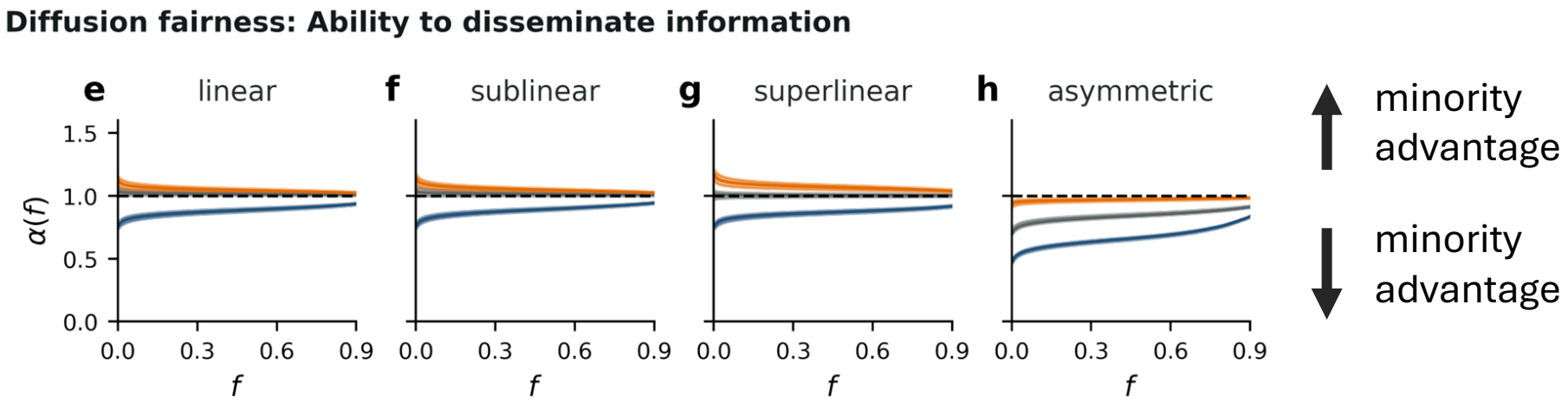


Takeaway: Homophily favors the majority, heterophily favors the minority in terms of *information acquisition*.

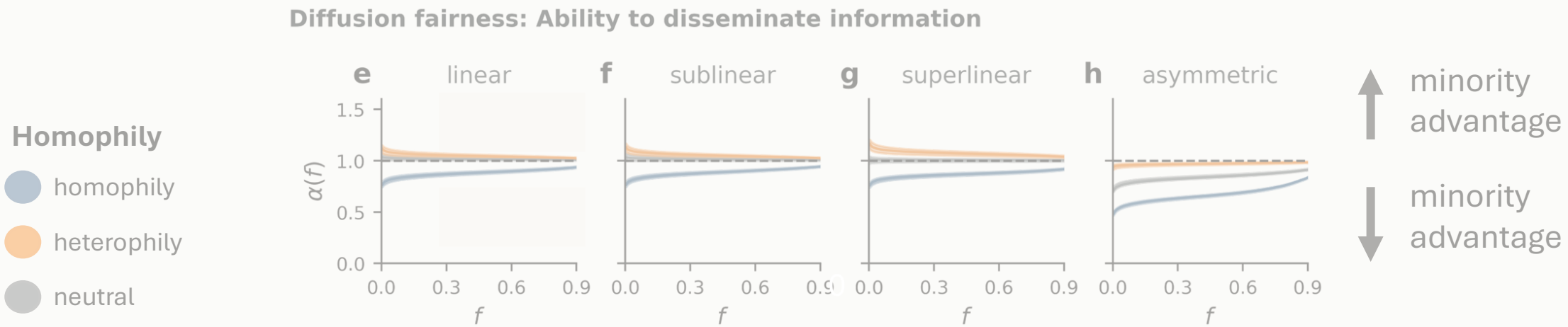
Differences in Information Dissemination

Homophily

- homophily
- heterophily
- neutral



Differences in Information Dissemination

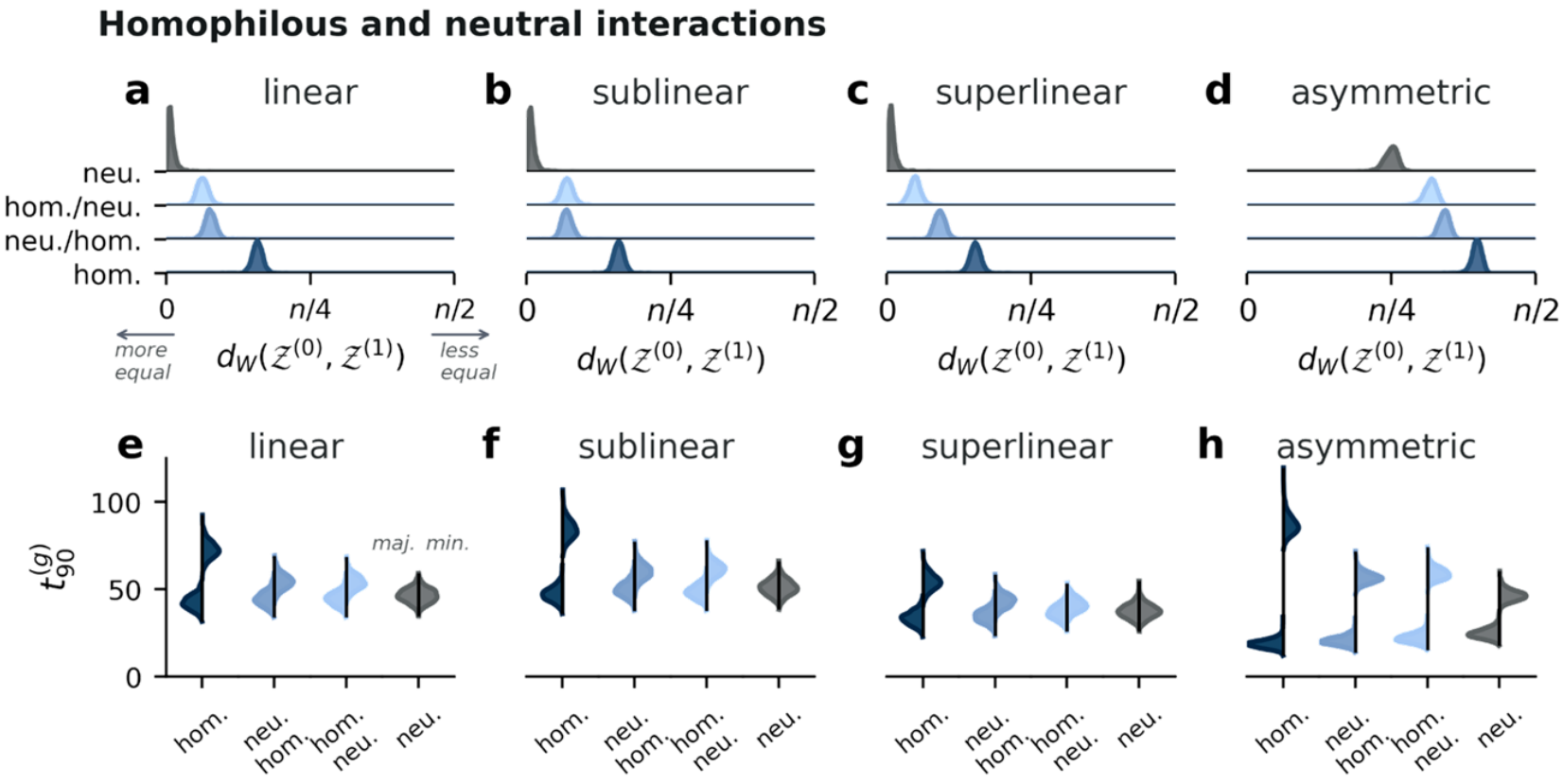


Takeaway: Homophily favors the majority, heterophily favors the minority in terms of *information dissemination*.

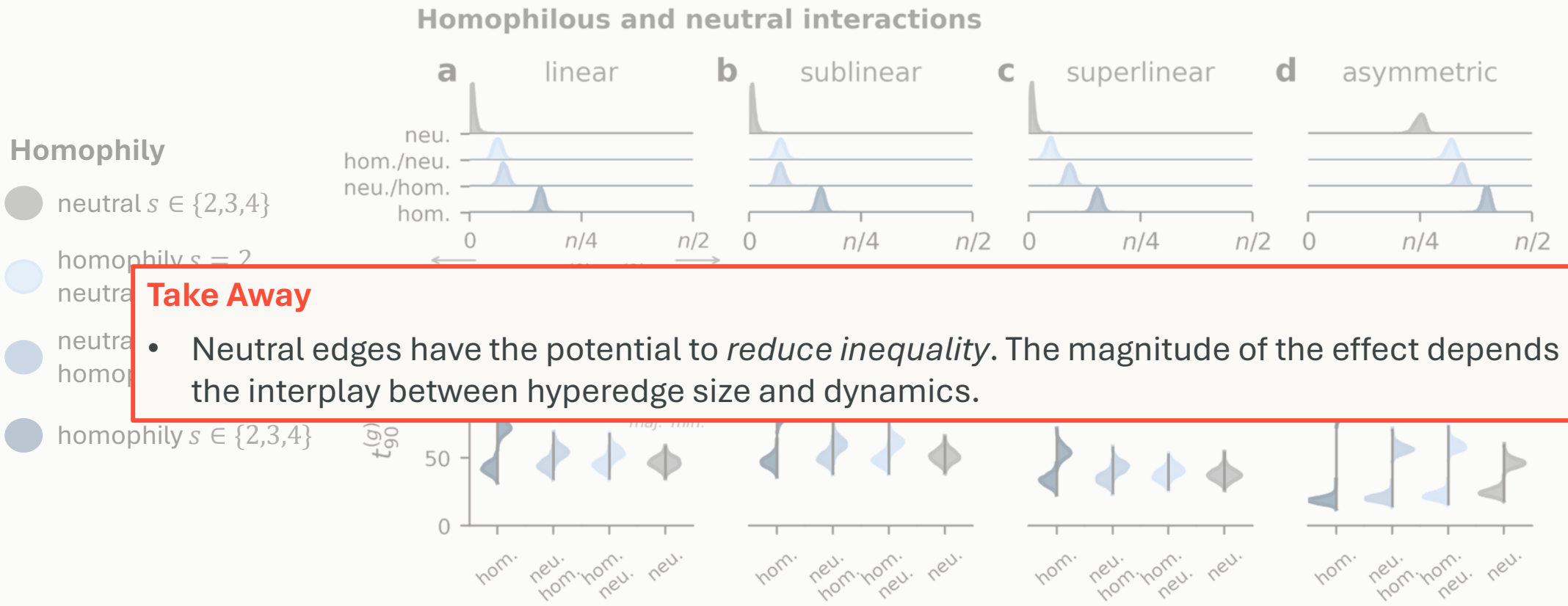
Mixed Patterns Homophily

Homophily

- neutral $s \in \{2,3,4\}$
- homophily $s = 2$ neutral $s \in \{3,4\}$
- neutral $s = 2$ homophily $s \in \{3,4\}$
- homophily $s \in \{2,3,4\}$



Mixed Patterns Homophily



Take Away

- Neutral edges have the potential to *reduce inequality*. The magnitude of the effect depends on the interplay between hyperedge size and dynamics.

Conclusions

The interplay of higher-order interactions, homophily, and non-linear social interactions shapes information access equality.

In particular, in-group preference (asymmetric transmission) and homophily lead to large information access inequality.

The presence of neutral interactions in mixed homophily patterns can reduce inequality, suggesting a novel type of structural intervention.

When do neural ODEs generalize on complex networks?

When do neural ODEs generalize on complex networks?

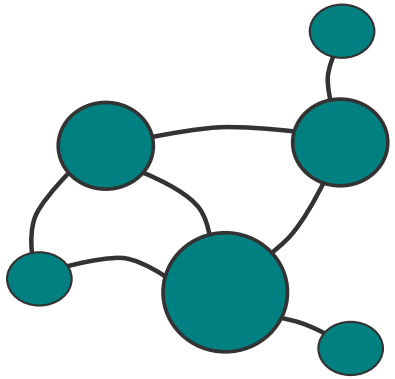




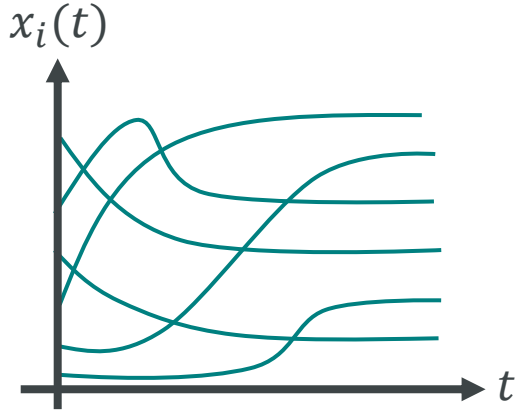
Laber, M. Klein, B. Eliass-Rad, T.

under revision at Nature Communications
(arXiv:2602.08980)

Dynamics from Data



graph A

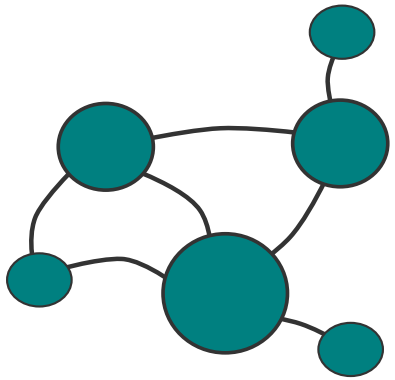


node-wise timeseries $x_i(t)$

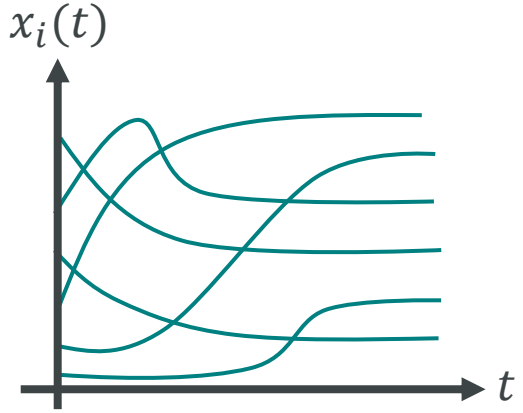
Dynamics from Data

Classical ODEs:
Handcraft differential equations using domain knowledge.

$$\frac{dx_i}{dt} = f(x_1, \dots, x_n)$$



graph A

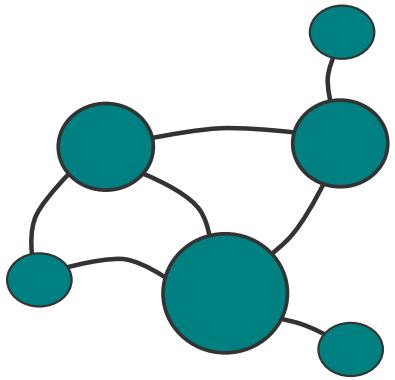


node-wise timeseries $x_i(t)$

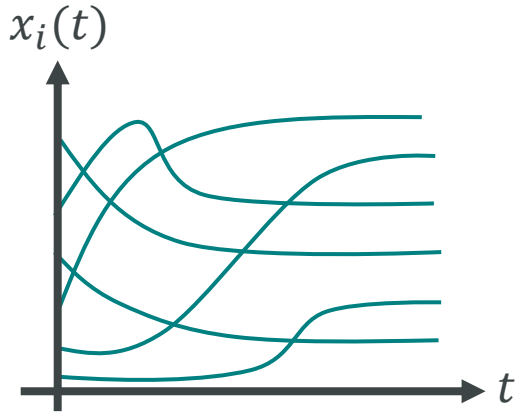
Dynamics from Data

Classical ODEs:
Handcraft differential equations using domain knowledge.

$$\frac{dx_i}{dt} = f(x_1, \dots, x_n)$$



graph A



node-wise timeseries $x_i(t)$

Neural ODEs [1,2,3]**:**
Learn differential equations from timeseries data.

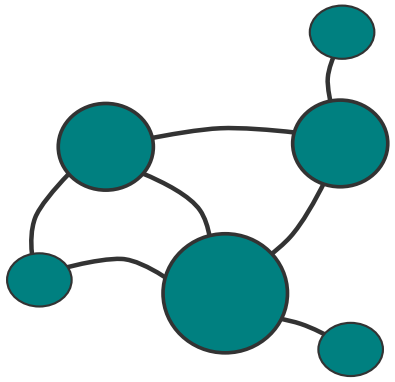
$$\frac{dx_i}{dt} = f_\omega(x_1, \dots, x_n)$$

[1] *Chen et al.* NeurIPS (2018), [2] *Poli et al.* AAAI (2022), [3] *Kidger* PhD Thesis (2022)

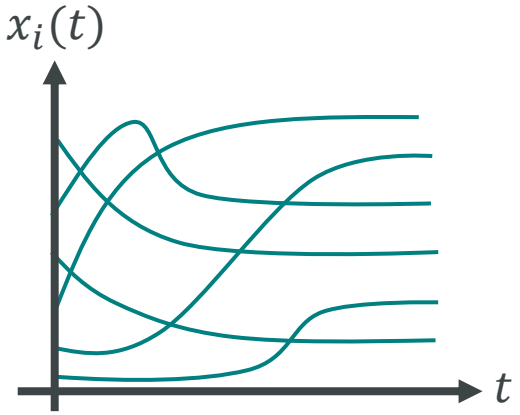
Dynamics from Data

Classical ODEs:
Handcraft differential equations using domain knowledge.

$$\frac{dx_i}{dt} = f(x_1, \dots, x_n)$$



graph A



node-wise timeseries $x_i(t)$

Neural ODEs [1,2,3]**:**
Learn differential equations from timeseries data.

$$\frac{dx_i}{dt} = f_\omega(x_1, \dots, x_n)$$

Questions: How do the performance and robustness of neural ODEs depend on the network? [4]

[1] *Chen et al.* NeurIPS (2018), [2] *Poli et al.* AAAI (2022), [3] *Kidger* PhD Thesis (2022)

[4] *Vasiliauskaite & Antulov-Fantulin* Comms Phys. 7 1 (2024)

Contributions

We train neural ODEs on **different dynamical systems** and **graphs from the $\mathbb{S}^1$ model** with different structural characteristics

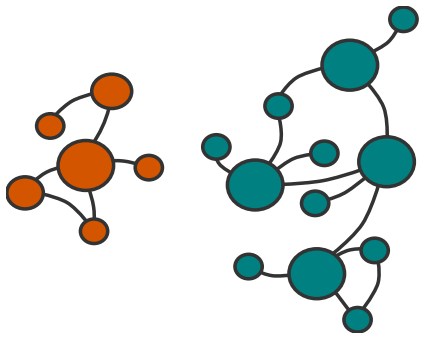
Contributions

We train neural ODEs on **different dynamical systems** and **graphs from the $\mathbb{S}^1$ model** with different structural characteristics

We evaluate their **performance and robustness** in terms of four criteria:

Generalization:

Graph Size



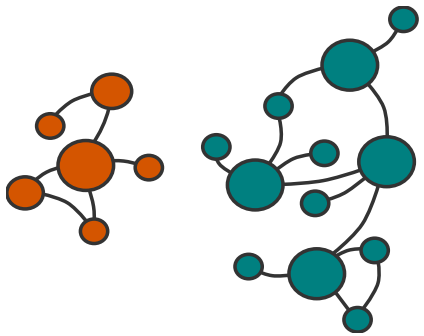
$n^{\text{train}} \ll n^{\text{test}}$

Contributions

We train neural ODEs on **different dynamical systems** and **graphs from the $\mathbb{S}^1$ model** with different structural characteristics

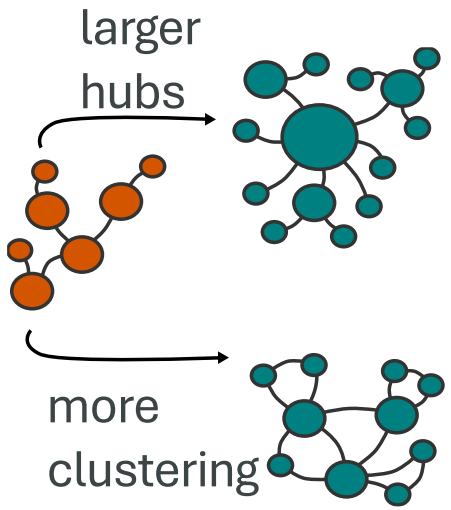
We evaluate their **performance and robustness** in terms of four criteria:

Generalization: Graph Size



$n^{\text{train}} \ll n^{\text{test}}$

Generalization: Graph Properties

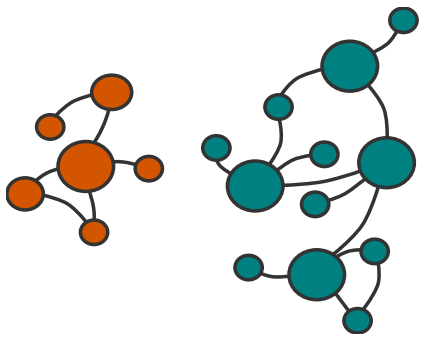


Contributions

We train neural ODEs on **different dynamical systems** and **graphs from the $\mathbb{S}^1$ model** with different structural characteristics

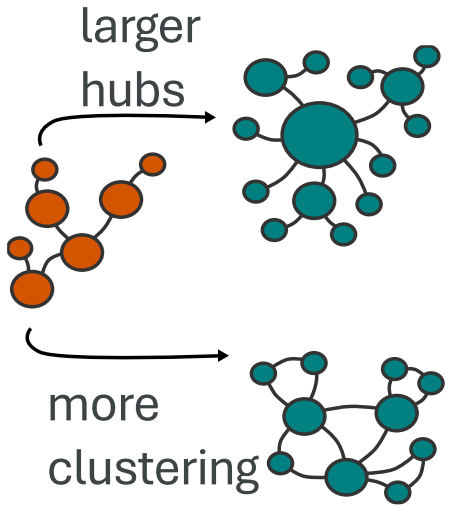
We evaluate their **performance and robustness** in terms of four criteria:

Generalization: Graph Size

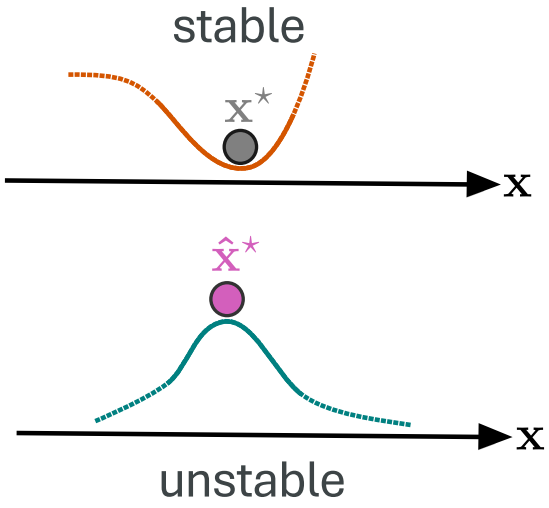


$n^{\text{train}} \ll n^{\text{test}}$

Generalization: Graph Properties



Robustness: Local Stability

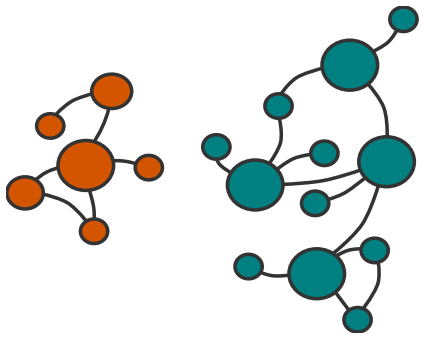


Contributions

We train neural ODEs on **different dynamical systems** and **graphs from the $\mathbb{S}^1$ model** with different structural characteristics

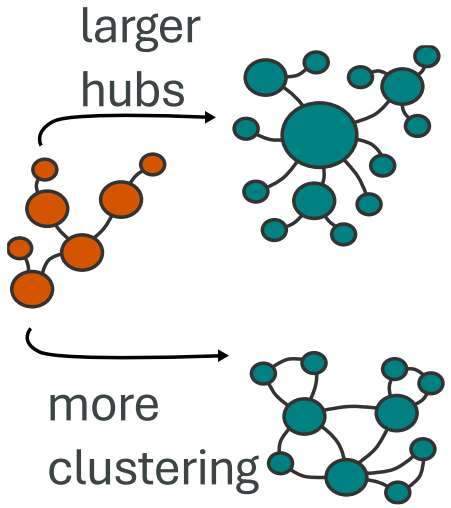
We evaluate their **performance and robustness** in terms of four criteria:

Generalization: Graph Size

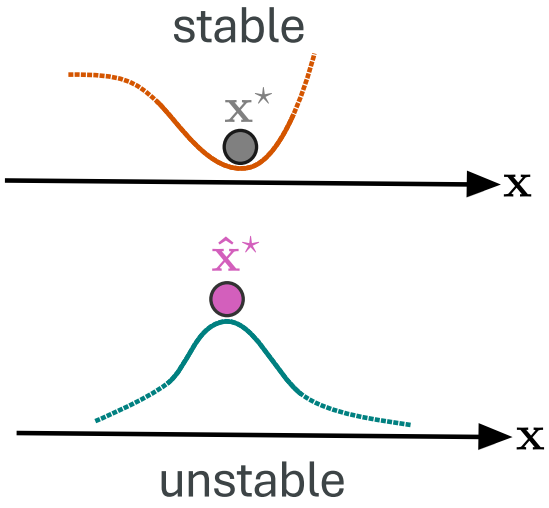


$n^{\text{train}} \ll n^{\text{test}}$

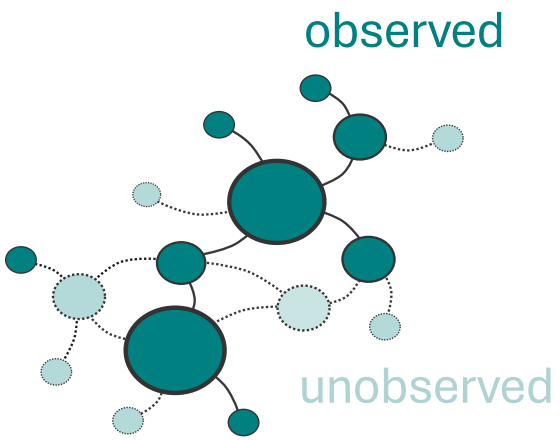
Generalization: Graph Properties



Robustness: Local Stability



Robustness: Missing Nodes



Background: Neural ODEs

Neural ODEs are ordinary differential equations in which the vector field is parameterized by a neural network.

$$\frac{dx_i}{dt} = f_\omega(t, x_1(t), \dots, x_n(t)) \qquad x_i(t=0) = \tilde{x}_i$$

Background: Neural ODEs

Neural ODEs are ordinary differential equations in which the vector field is parameterized by a neural network.

$$\frac{dx_i}{dt} = f_\omega(t, x_1(t), \dots, x_n(t)) \qquad x_i(t=0) = \tilde{x}_i$$

Applications:

- Continuum limit of architectures with residual connections [1].

[1] *Chen et al.* NeurIPS (2018)

Background: Neural ODEs

Neural ODEs are ordinary differential equations in which the vector field is parameterized by a neural network.

$$\frac{dx_i}{dt} = f_\omega(t, x_1(t), \dots, x_n(t)) \qquad x_i(t=0) = \tilde{x}_i$$

Applications:

- Continuum limit of architectures with residual connections [1].
- Part of generative models, i.e., continuous normalizing flows [1,2].

[1] *Chen et al.* NeurIPS (2018) [2] *Grathwohl et al.* ICLR (2019)

Background: Neural ODEs

Neural ODEs are ordinary differential equations in which the vector field is parameterized by a neural network.

$$\frac{dx_i}{dt} = f_\omega(t, x_1(t), \ldots, x_n(t)) \qquad x_i(t=0) = \tilde{x}_i$$

Applications:

- Continuum limit of architectures with residual connections [1].
- Part of generative models, i.e., continuous normalizing flows [1,2].
- Modeling of dynamical systems from data [3,4].

[1] *Chen et al.* NeurIPS (2018) [2] *Grathwohl et al.* ICLR (2019) [3] *Rackauckas et al.* arXiv:2001.04385 (2020) [4] *Gao et al.* Nat.Comms. 15 (1) 6029 (2024)

Background: Neural ODEs

Neural ODEs are ordinary differential equations in which the vector field is parameterized by a neural network.

$$\frac{dx_i}{dt} = f_\omega(t, x_1(t), \ldots, x_n(t)) \qquad x_i(t=0) = \tilde{x}_i$$

Applications:

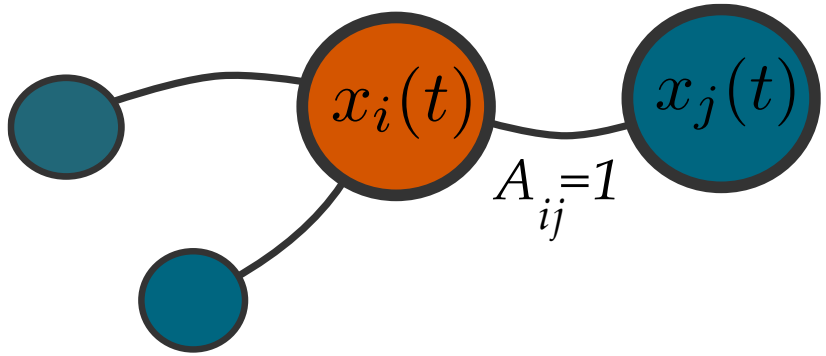
- Continuum limit of architectures with residual connections [1].
- Part of generative models, i.e., continuous normalizing flows [1,2].
- Modeling of dynamical systems from data [3,4].

[1] *Chen et al.* NeurIPS (2018) [2] *Grathwohl et al.* ICLR (2019) [3] *Rackauckas et al.* arXiv:2001.04385 (2020) [4] *Gao et al.* Nat.Comms. 15 (1) 6029 (2024)

Background: ODEs on Graphs

Many dynamical systems on graphs are well described by [1]

$$\frac{dx_i}{dt} = f\left(x_i(t)\right) + \sum_{j \in \mathcal{V}} A_{ij} h^{\text{ego}}\left(x_i(t)\right) h^{\text{alt}}\left(x_j(t)\right)$$

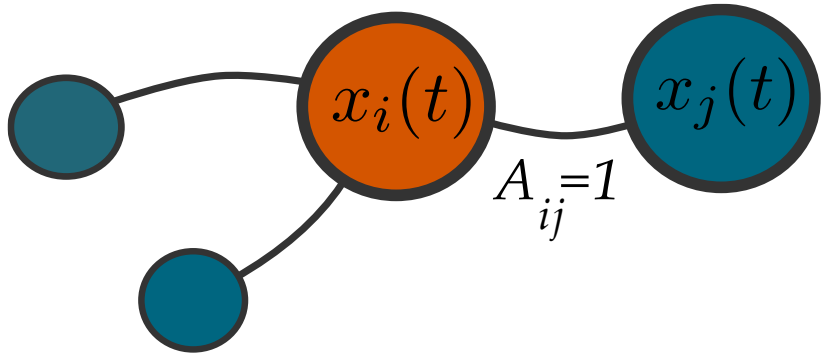


[1] *Barzel & Barabási* Nat. Phys. 9 673-681 (2013)

Background: ODEs on Graphs

Many dynamical systems on graphs are well described by [1]

$$\frac{dx_i}{dt} = f\left(x_i(t)\right) + \sum_{j \in \mathcal{V}} A_{ij} h^{\text{ego}}\left(x_i(t)\right) h^{\text{alt}}\left(x_j(t)\right)$$



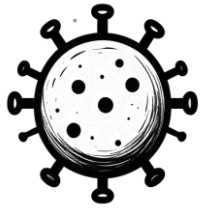
Example: SIS-Model

$$\frac{dx_i}{dt} = -\mu x_i(t) + \sum_{j \in \mathcal{V}} A_{ij}\, \beta\left(1 - x_i(t)\right) x_j(t)$$

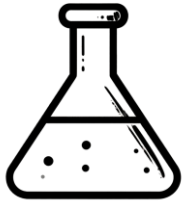
[1] *Barzel & Barabási* Nat. Phys. 9 673-681 (2013)

Setup: Dynamics

We use synthetic data from **five dynamical systems** from different domains [1,2].



Susceptible-Infected-Susceptible (SIS)



Mass-Action Kinetics (MAK)



Michaelis-Menten Model (MM)



Birth-Death Process (BD)



Neuronal Dynamics (ND)

[1] *Meena* et al. Nat. Phys. 19 1033 (2023) [2] *Hens et al.* Nat. Phys. 15 403-412 (2019)

Neural ODEs on Graphs

Replace f, h^{ego}, h^{alt} by separate neural networks f_ω, h_ω^{ego}, h_ω^{alt} to obtain a Neural ODE:

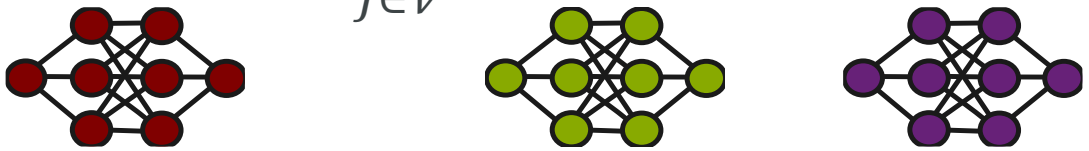
$$\frac{dx_i}{dt} = f\big(x_i(t)\big) + \sum_{j \in \mathcal{V}} A_{ij} h^{\text{ego}}\big(x_i(t)\big) h^{\text{alt}}\Big(x_j(t)\Big)$$

Neural ODEs on Graphs

Replace f, h^{ego}, h^{alt} by separate neural networks f_ω, h_ω^{ego}, h_ω^{alt} to obtain a Neural ODE[1,2,3]:

$$\frac{dx_i}{dt} = f\left(x_i(t)\right) + \sum_{j \in \mathcal{V}} A_{ij} h^{\text{ego}}\left(x_i(t)\right) h^{\text{alt}}\left(x_j(t)\right)$$

$$\frac{dx_i}{dt} = f_\omega\left(x_i(t)\right) + \sum_{j \in \mathcal{V}} A_{ij}\, h_\omega^{\text{ego}}\left(x_i(t)\right) h_\omega^{\text{alt}}\left(x_j(t)\right)$$

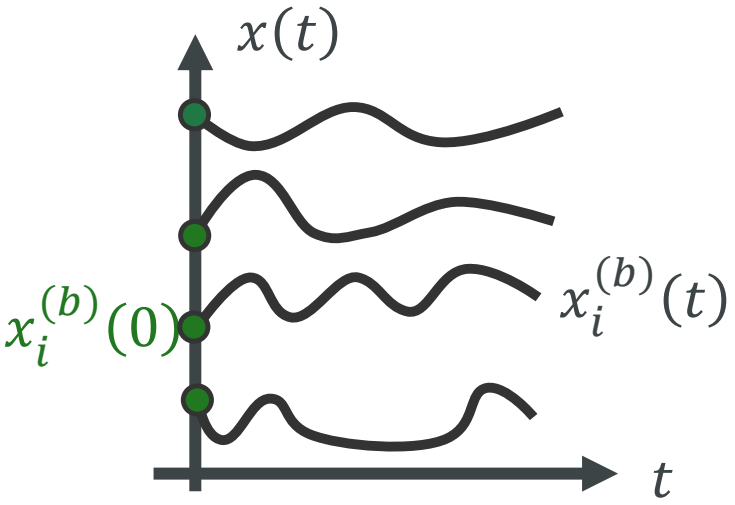


[1] *Poli et al.* AAAI (2022) [2] *Gao et al.* Nat.Comms. 15 (1) 6029 (2024) [3] *Vasiliauskaite & Antulov-Fantulin* Comms Phys. 7 1 (2024)

Training Neural ODEs on Graphs

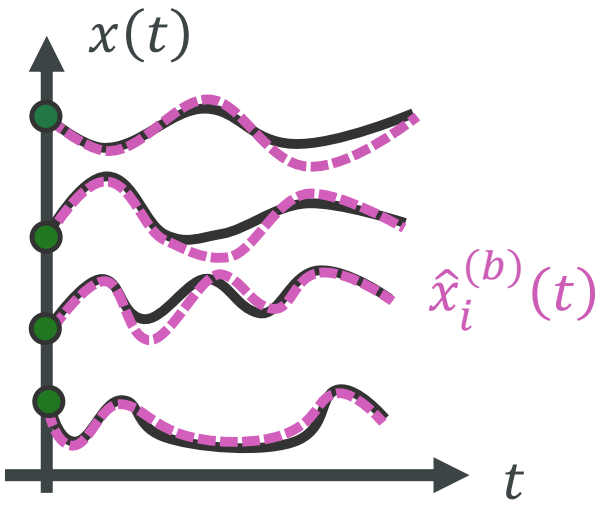
Data

node-wise time series data starting from different initial conditions.



Prediction

numerical integration with a standard ODE solver.



Optimization

Stochastic Gradient Descent with gradients obtained via backpropagation.

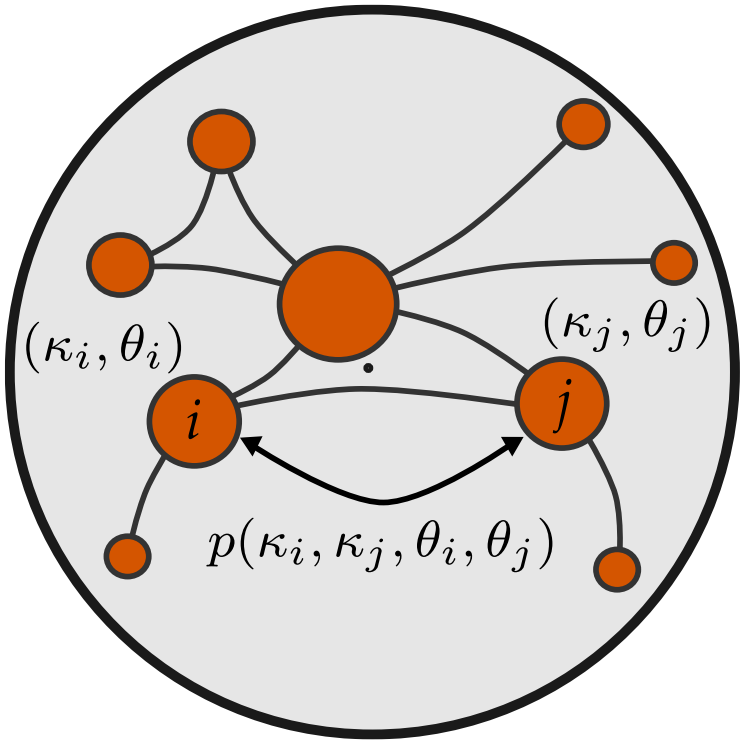
$$\omega \leftarrow \omega - \eta \nabla_\omega \mathcal{L}(x, \hat{x})$$

Background: $\mathbb{S}^1$ model

We use the $\mathbb{S}^1$ model [1,2] to generate graphs with different properties.

The model has **four parameters**:

- the *number of nodes* n
- the *average degree* $\bar{k}$
- the *exponent* γ of the degree distribution
  → controls degree heterogeneity
- The *inverse temperature* β
  → controls the clustering coefficient



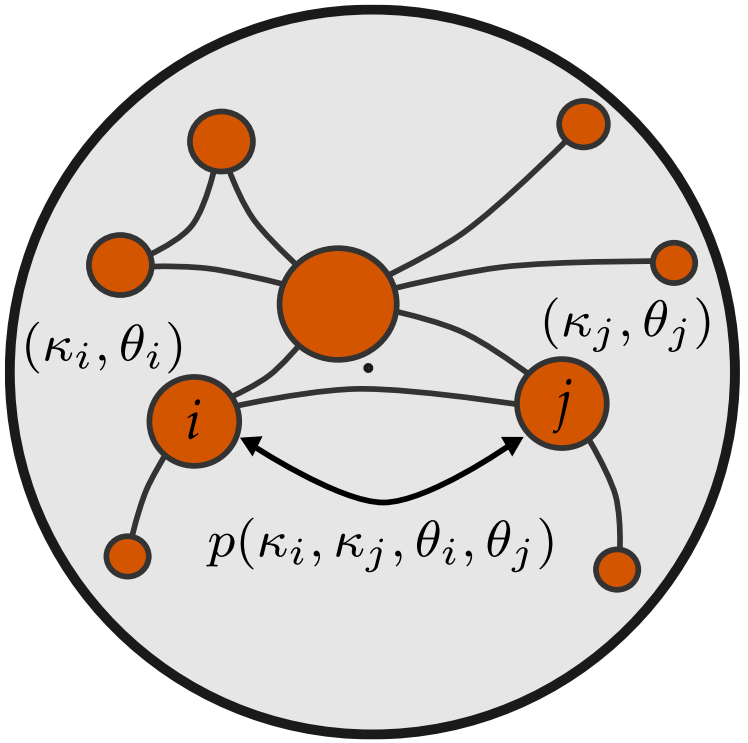
[1] *Serrano* et al. PRL 100, 078701 (2008) [2] *Krioukov* et al. PRE 82 036106 (2010)

Background: $\mathbb{S}^1$ model

We use the $\mathbb{S}^1$ model [1,2] to generate graphs with different properties.

The model has **four parameters**:

- the *number of nodes* n
- the *average degree* $\bar{k}$
- the *exponent* γ of the degree distribution
  → controls degree heterogeneity: **small γ → large hubs**
- The *inverse temperature* β
  → controls the clustering coefficient: **large β → strong clustering**



[1] *Serrano* et al. PRL 100, 078701 (2008) [2] *Krioukov* et al. PRE 82 036106 (2010)

Setup

Graphs

- number of nodes: $n \in \{64, \dots, 8192\}$
- average degree $\bar{k} = 10$
- exponent $\gamma \in [2.1, \dots, 3.9]$
- inverse temperature $\beta \in [0.1, \dots, 4.1]$

Dynamics

- SIS-Model (SIS),
- Mass-Action Kinetics (MAK)
- Birth-Death Process (BD)
- Neural Dynamics (ND)
- Michaelis-Menten (MM) model

Model

- 3 MLPs with layers 3 with 128 neurons.
- Dormand-Prince 4(5) ODE solver

Evaluation

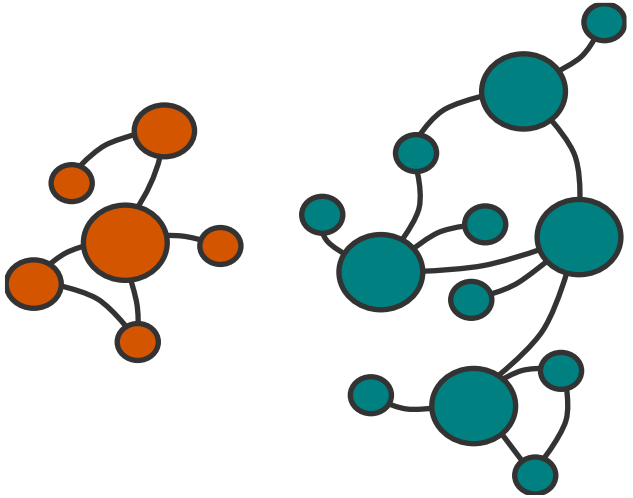
- mean absolute error $\mathcal{L}_{\text{mae}}$
- 100 test graphs

Generalization: Larger Graphs

Question: Can Neural ODEs, trained on small graphs, generalize to larger ones?

Motivation

- Size generalization is highly desirable for efficiency.
- Theoretical and empirical evidence for size generalization in (graph) machine learning is mixed [1,2].

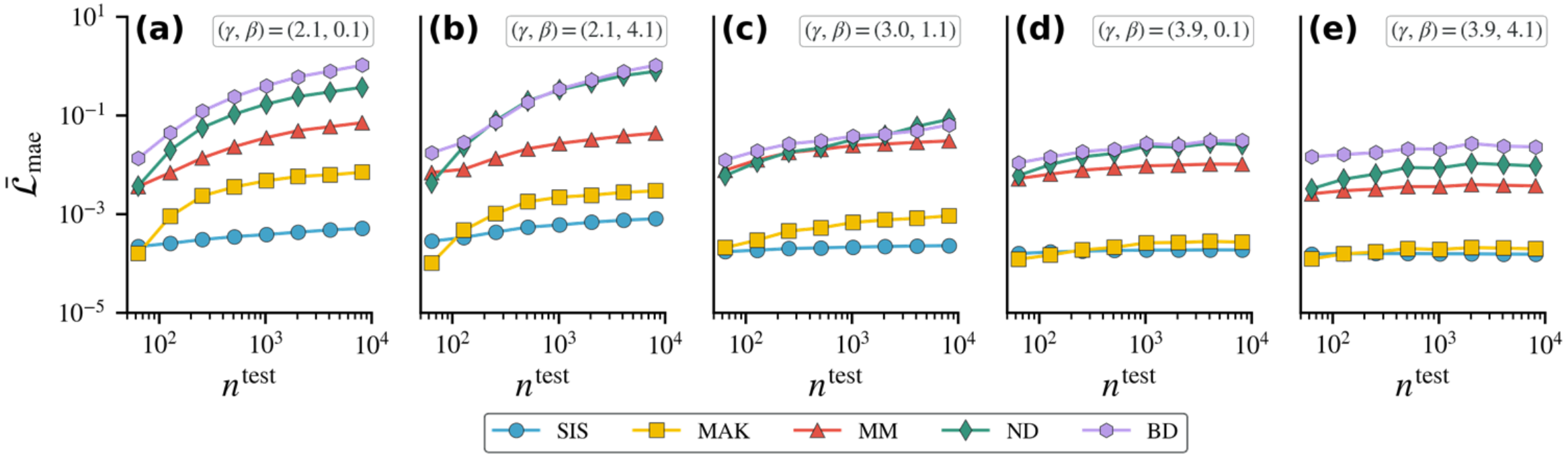


$n^{\text{train}} \ll n^{\text{test}}$

[1] *Yehudai et al.* ICML (2021) [2] *Anil et al.* NeurIPS (2022)

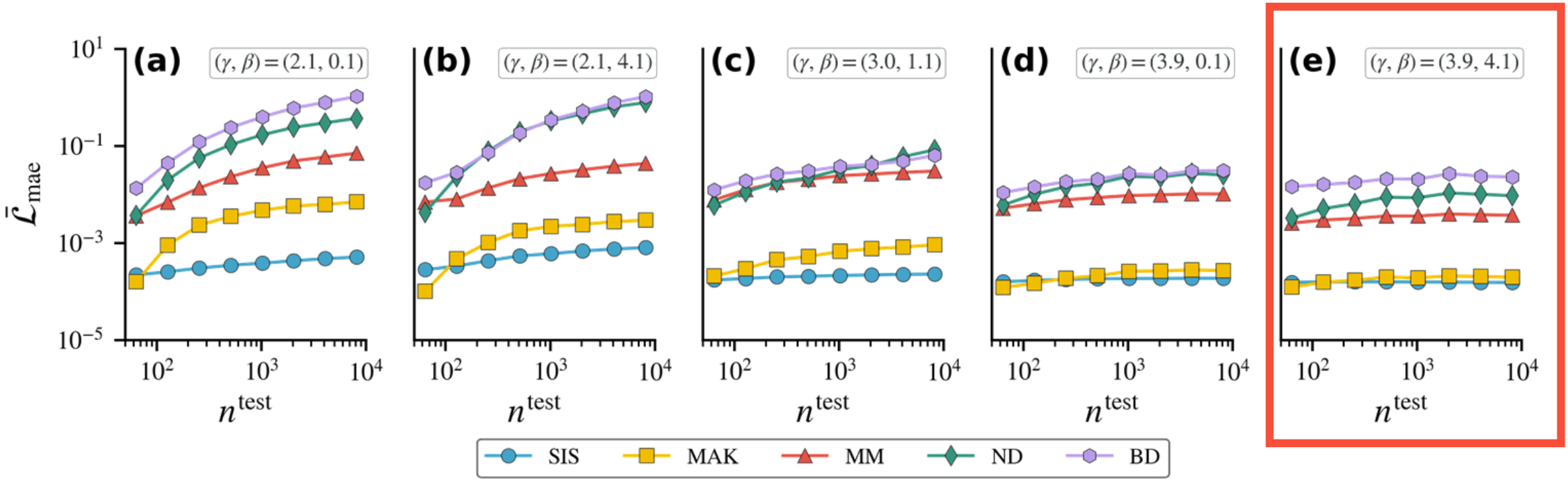
Generalization: Larger Graphs

Generalization to larger graphs is **possible for low degree heterogeneity.**
Clustering has limited influence on size generalization.



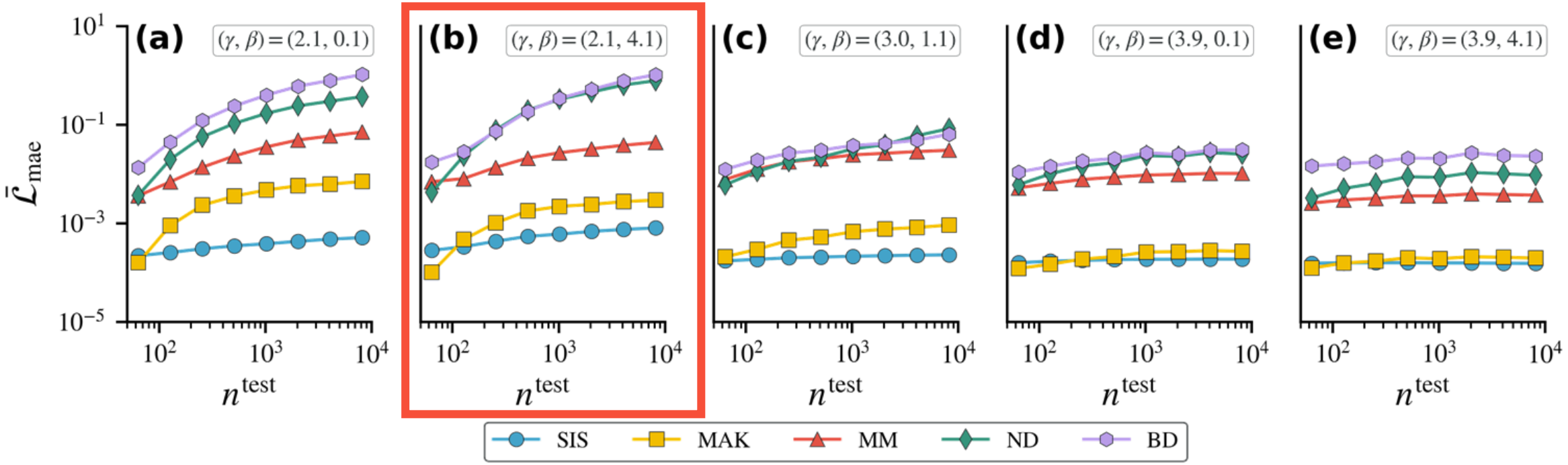
Generalization: Larger Graphs

Generalization to larger graphs is **possible for low degree heterogeneity.**
Clustering has limited influence on size generalization.



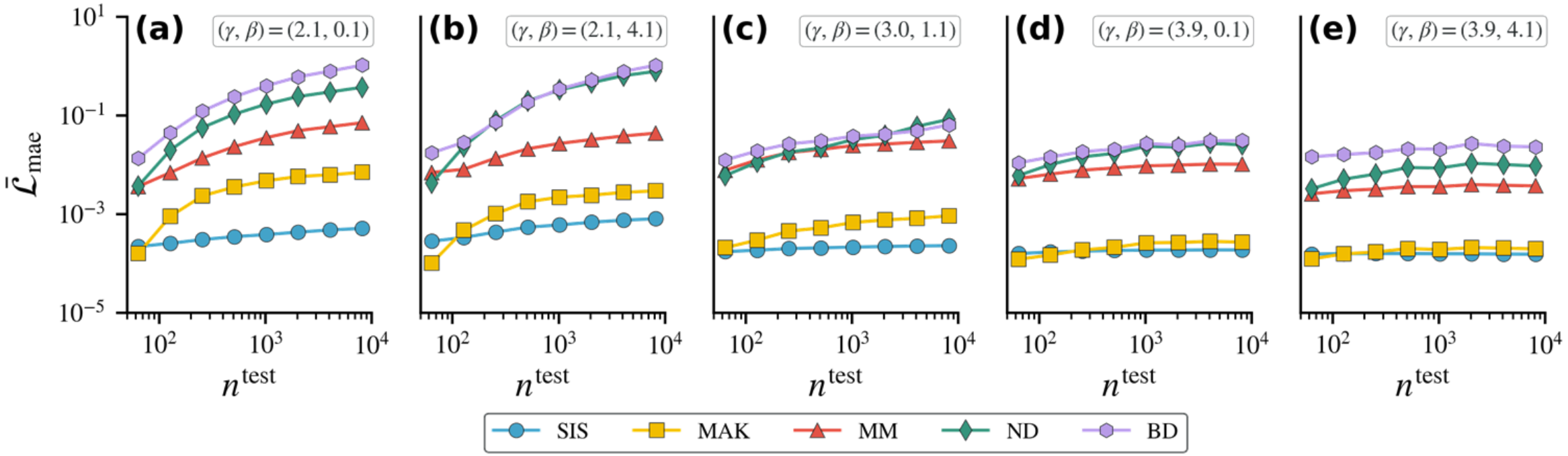
Generalization: Larger Graphs

Generalization to larger graphs is **possible for low degree heterogeneity.**
Clustering has limited influence on size generalization.



Generalization: Larger Graphs

Generalization to larger graphs is **possible for low degree heterogeneity.**
Clustering has limited influence on size generalization.



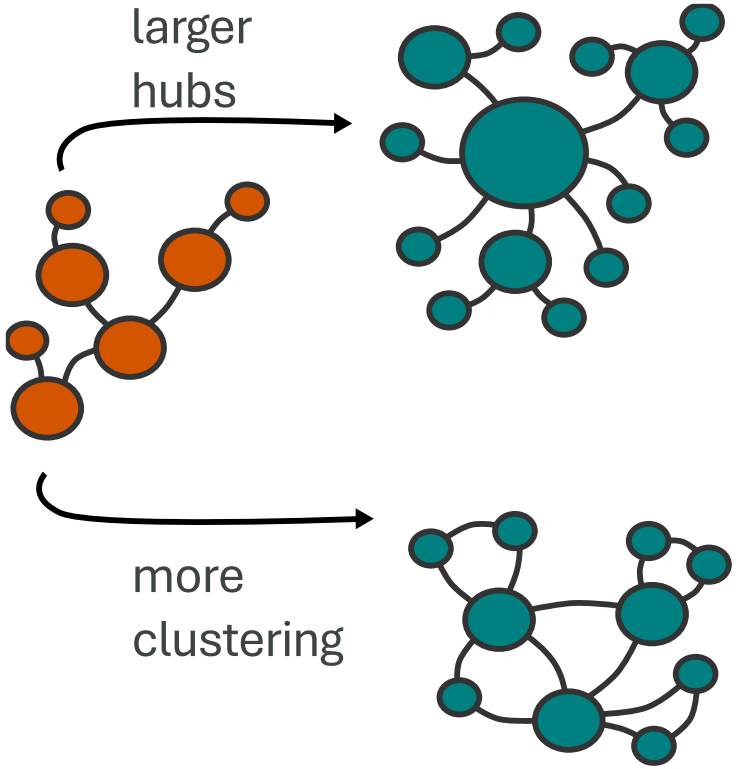
Generalization: Structurally Different Graphs

Question: Can Neural ODEs generalize to networks with different properties?

Motivation

- Many complex systems constantly evolve, causing potential mismatch between graph properties at training and test time.

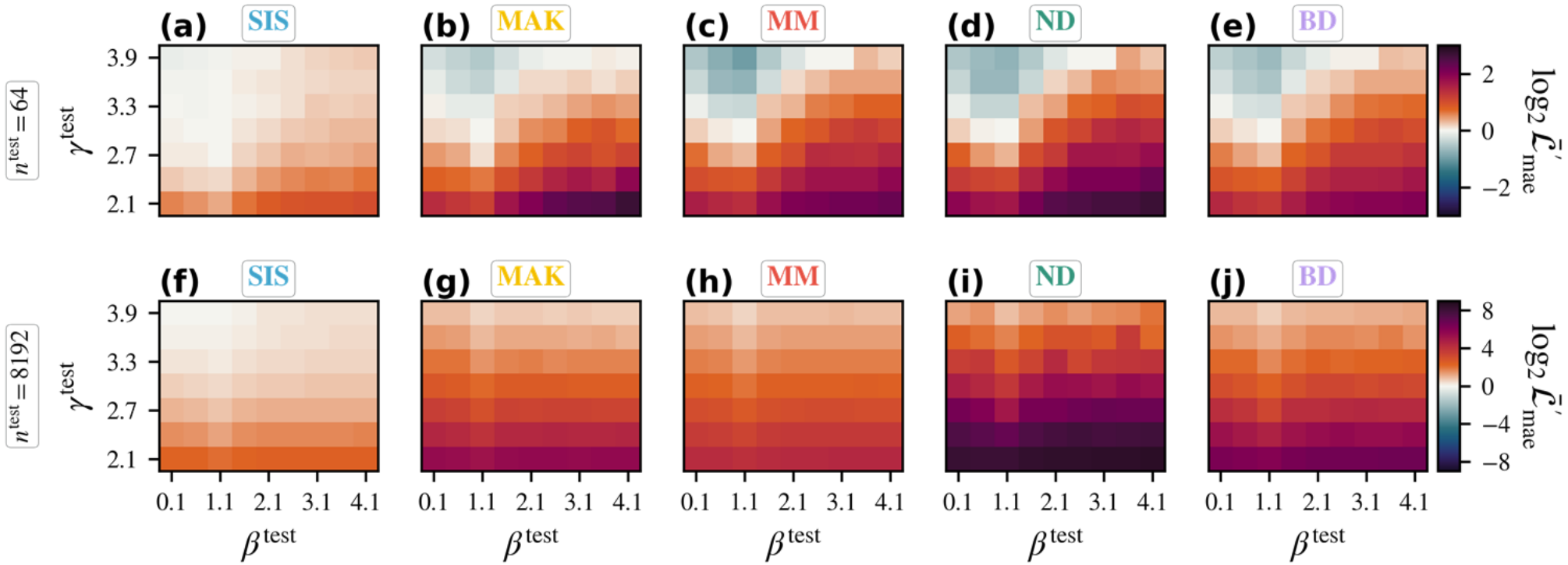
- Graph machine learning models are usually evaluated on a few benchmark datasets. Only recently, have random graph ensembles been recognized as a possible testbed [1,2,3].

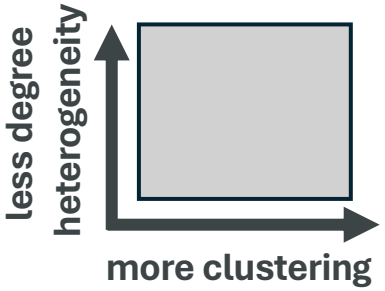


[1] *Palowitch et al.* KDD (2022) [2] *Aliakbarisani et al.* arXiv:2406.02772 (2024) [3] *Vasiliauskaite & Antulov-Fantulin* Comms Phys. 7 1 (2024)

Generalization: Structurally Different Graphs

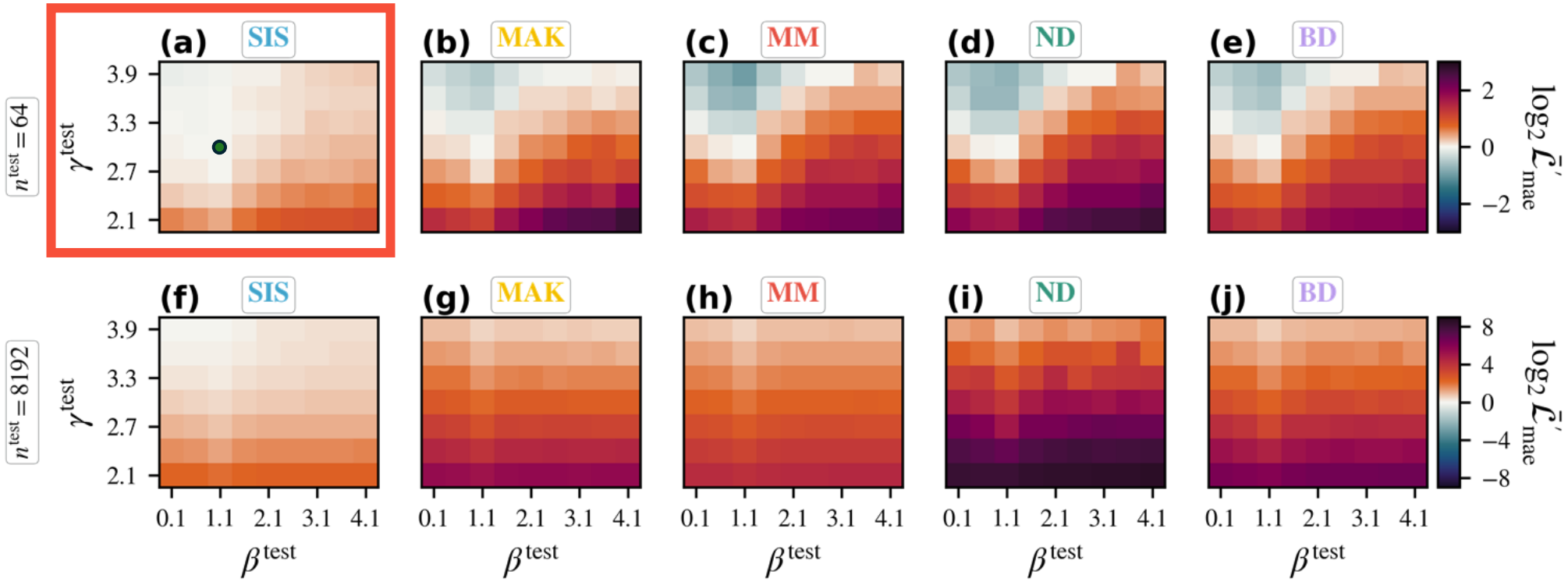
Performance is **good on less clustered**, and **less degree heterogeneous graphs**, but degrades with larger degree heterogeneity.

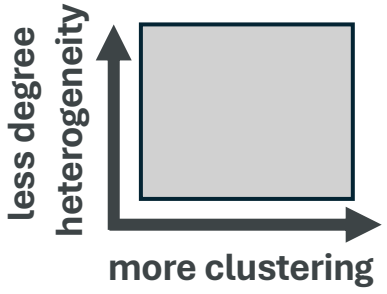




Generalization: Structurally Different Graphs

Performance is **good on less clustered**, and **less degree heterogeneous graphs**, but degrades with larger degree heterogeneity.

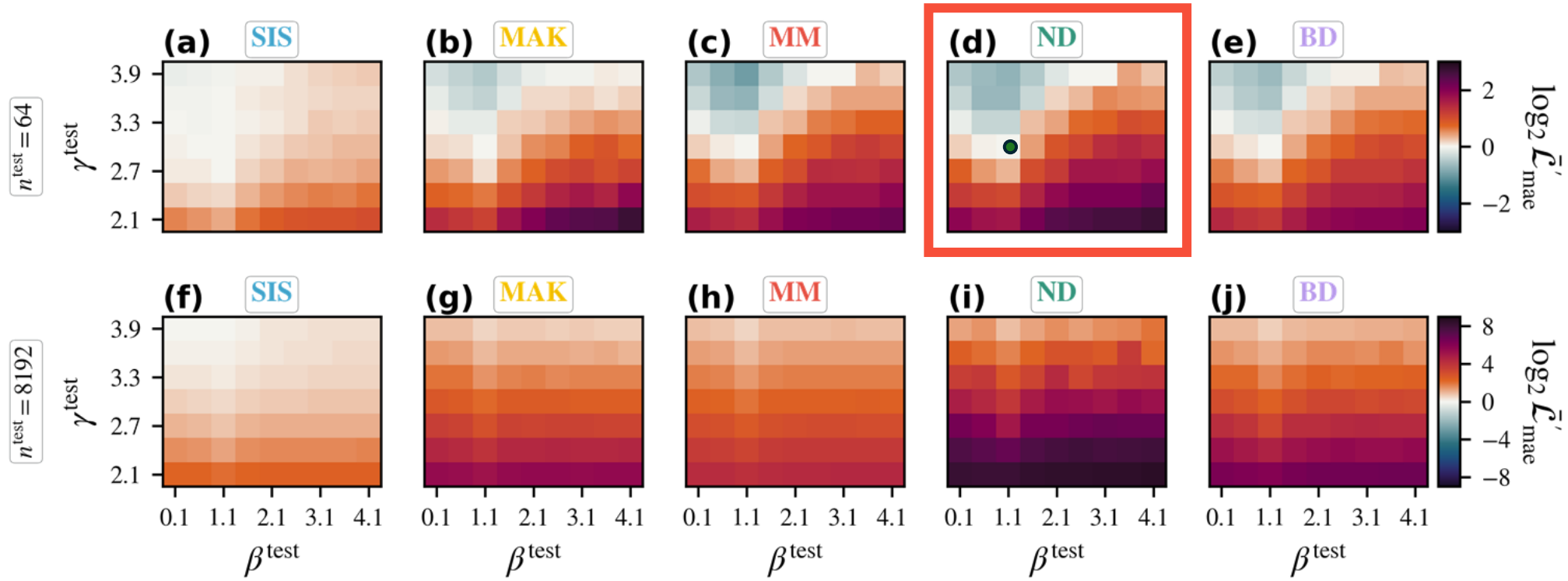


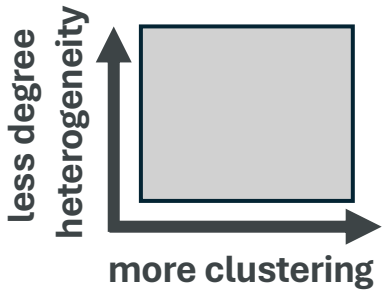


Training Parameters $(\gamma^{\text{train}}, \beta^{\text{train}}) = (3.0, 1.1)$

Generalization: Structurally Different Graphs

Performance is **good on less clustered**, and **less degree heterogeneous graphs**, but degrades with larger degree heterogeneity.

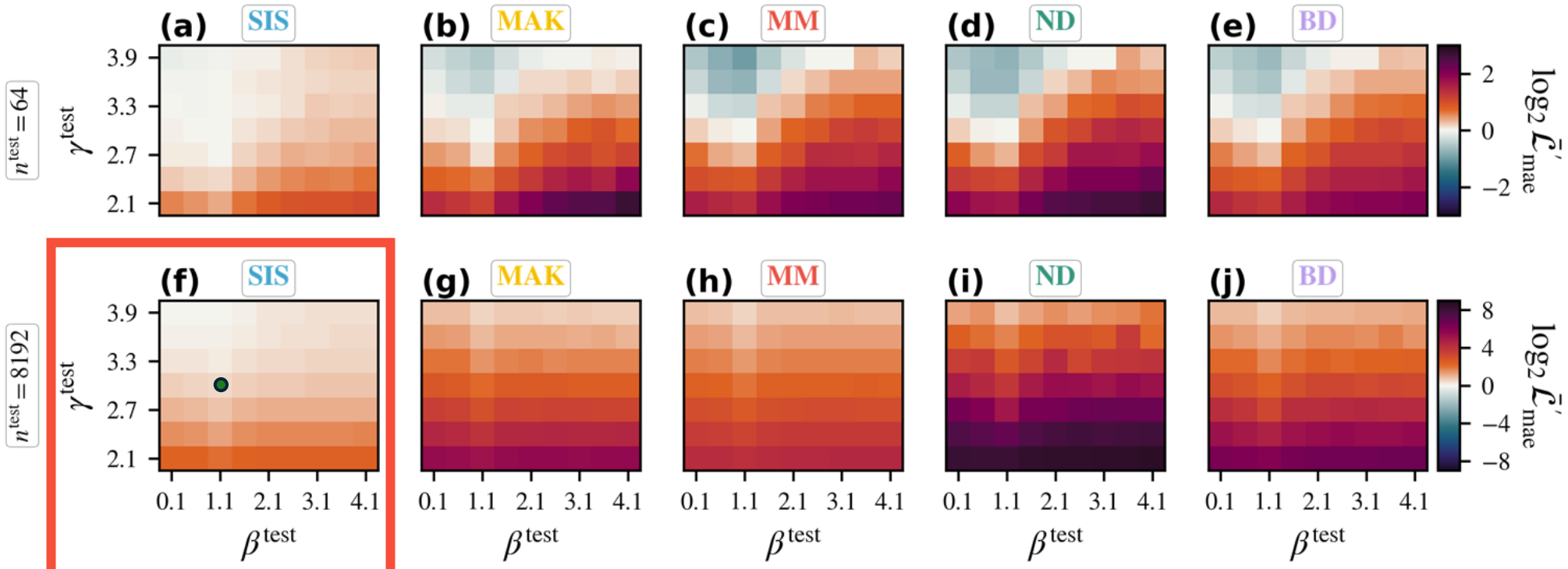


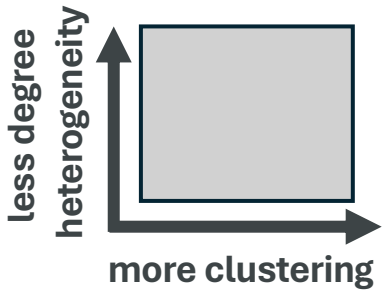


Training Parameters $(\gamma^{\text{train}}, \beta^{\text{train}}) = (3.0, 1.1)$

Generalization: Structurally Different Graphs

Performance is **good on less clustered**, and **less degree heterogeneous graphs**, but degrades with larger degree heterogeneity.

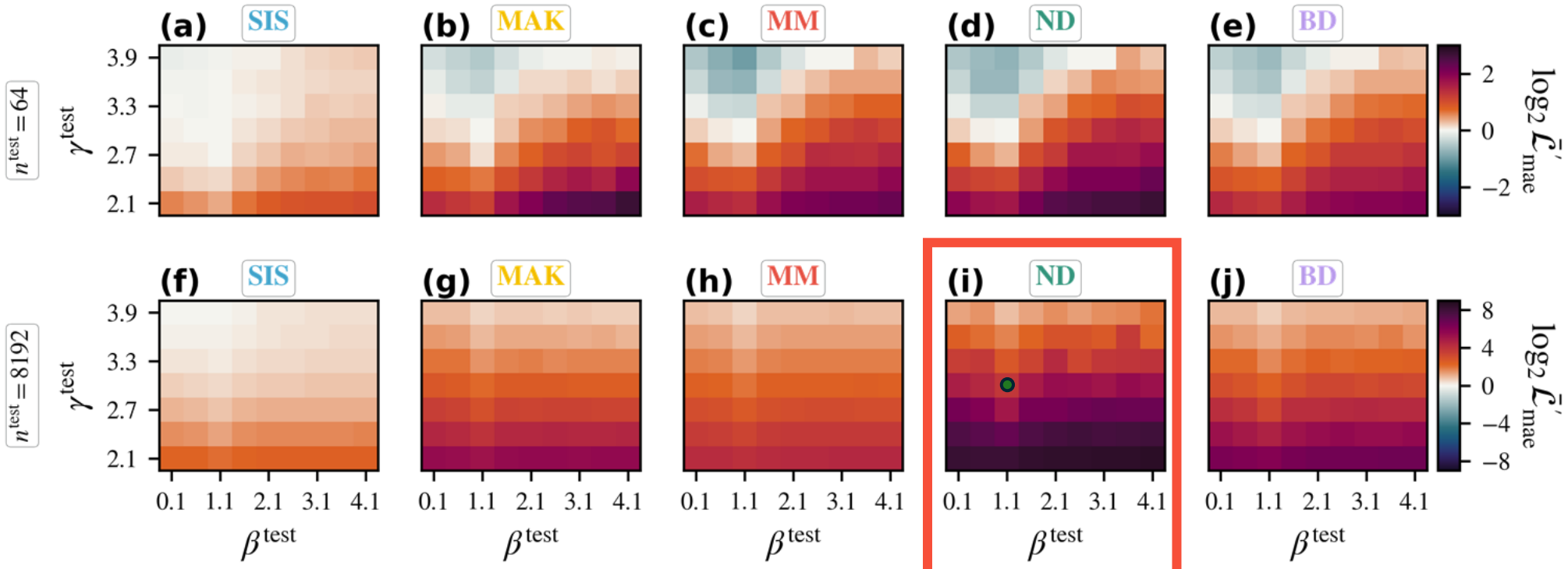


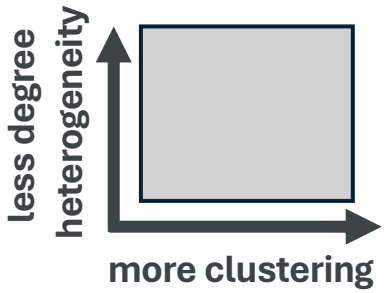


Training Parameters $(\gamma^{\text{train}}, \beta^{\text{train}}) = (3.0, 1.1)$

Generalization: Structurally Different Graphs

Performance is **good on less clustered**, and **less degree heterogeneous graphs**, but degrades with larger degree heterogeneity.

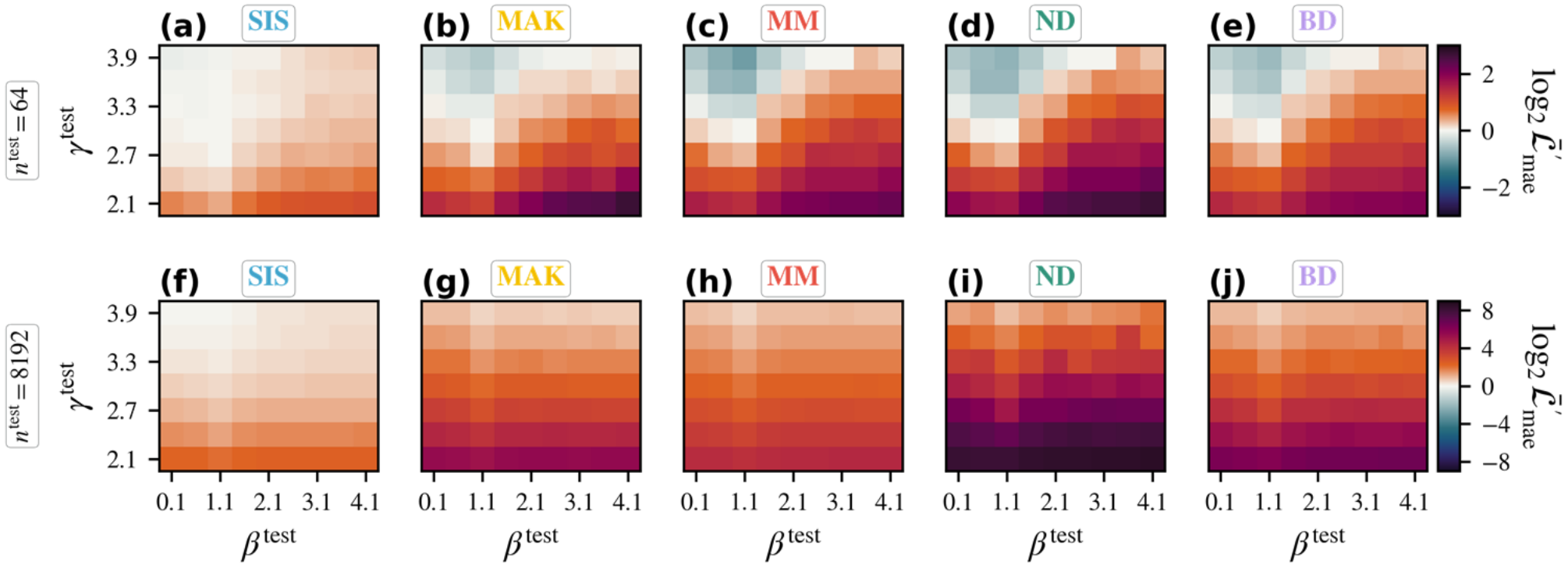


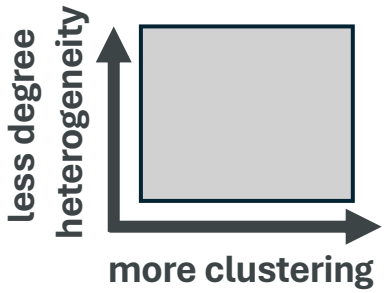


Training Parameters $(\gamma^{\text{train}}, \beta^{\text{train}}) = (3.0, 1.1)$

Generalization: Structurally Different Graphs

Performance is **good on less clustered**, and **less degree heterogeneous graphs**, but degrades with larger degree heterogeneity.



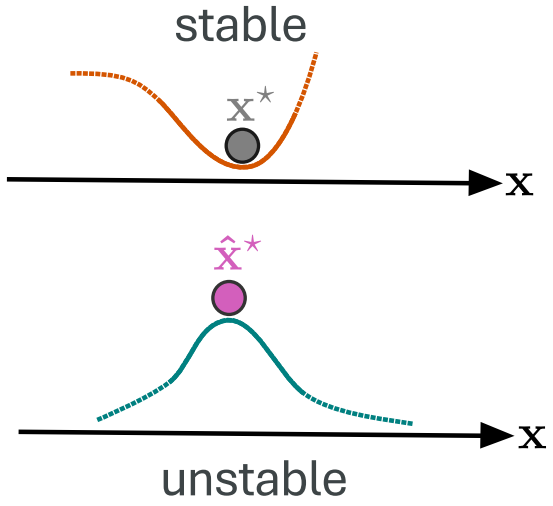


Local Stability

Question: Do neural ODEs reproduce local stability properties?

Motivation

- Stability is one of the most fundamental properties of a dynamical system. A good data-driven approximation should be able to capture stability properties.



Local Stability

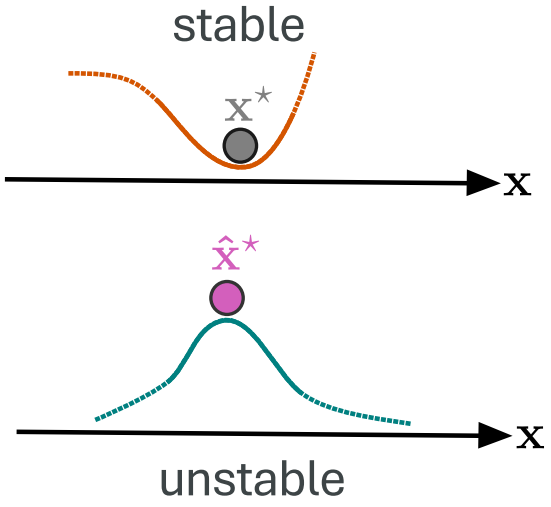
Question: Do neural ODEs reproduce local stability properties?

Motivation

- Stability is one of the most fundamental properties of a dynamical system. A good data-driven approximation should be able to capture stability properties.

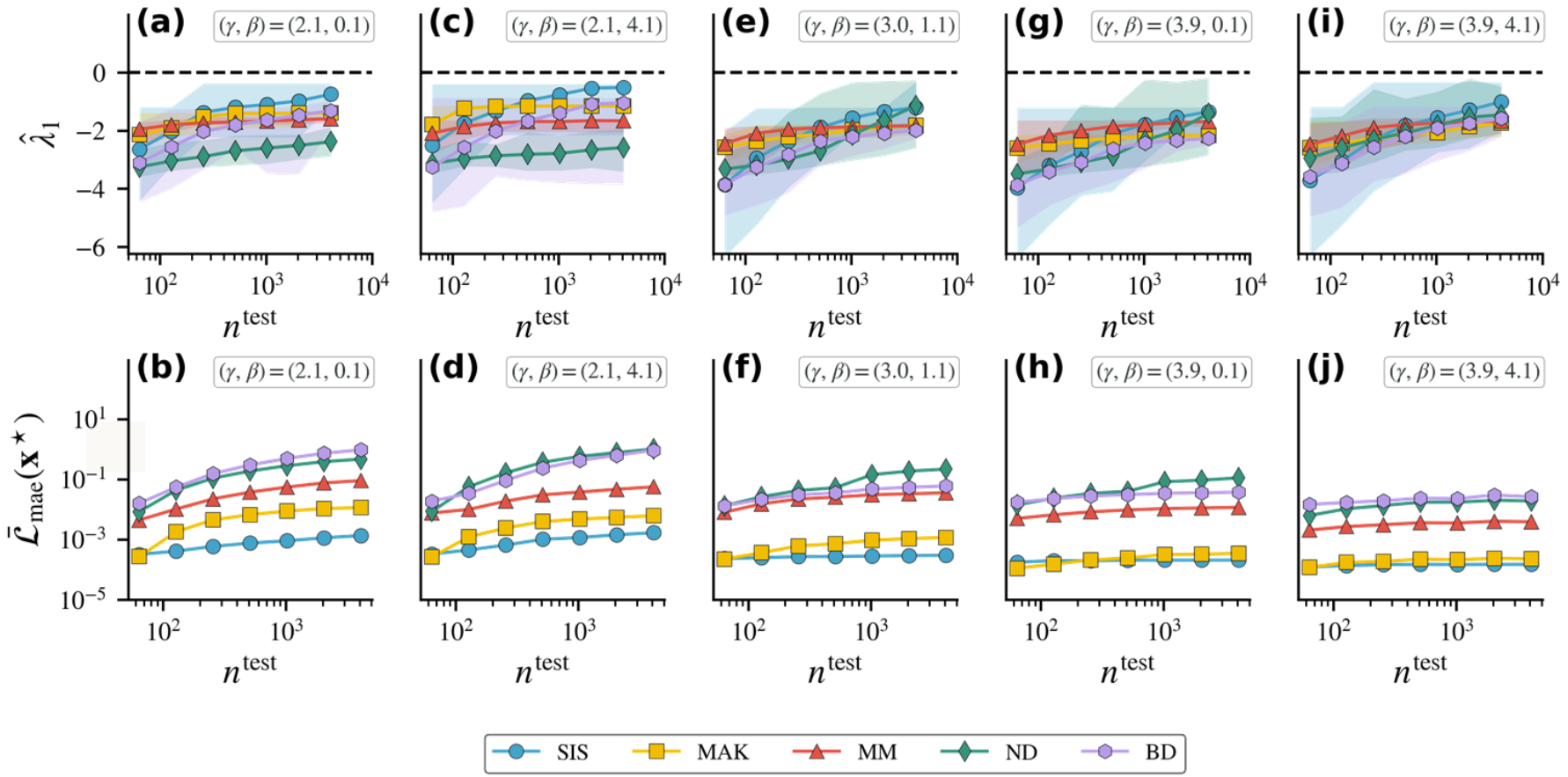
Background

- The local stability of a dynamical system at a given point $x^\star$is determined by the largest real-part λ_1 of the eigenvalues of the Jacobian at $x^\star$.



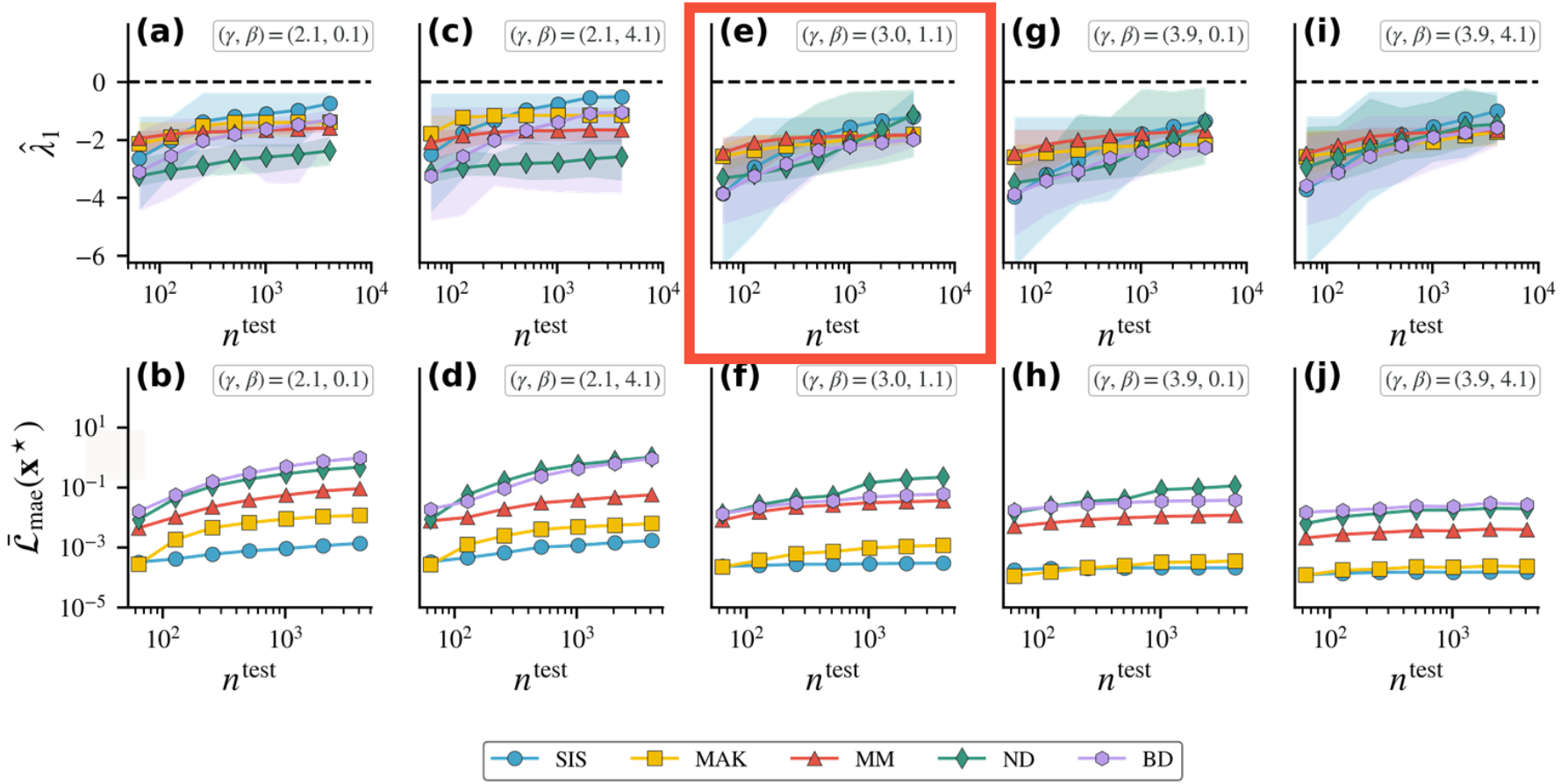
Local Stability

All models **approach stable fixed points**, but **potentially the wrong ones**.



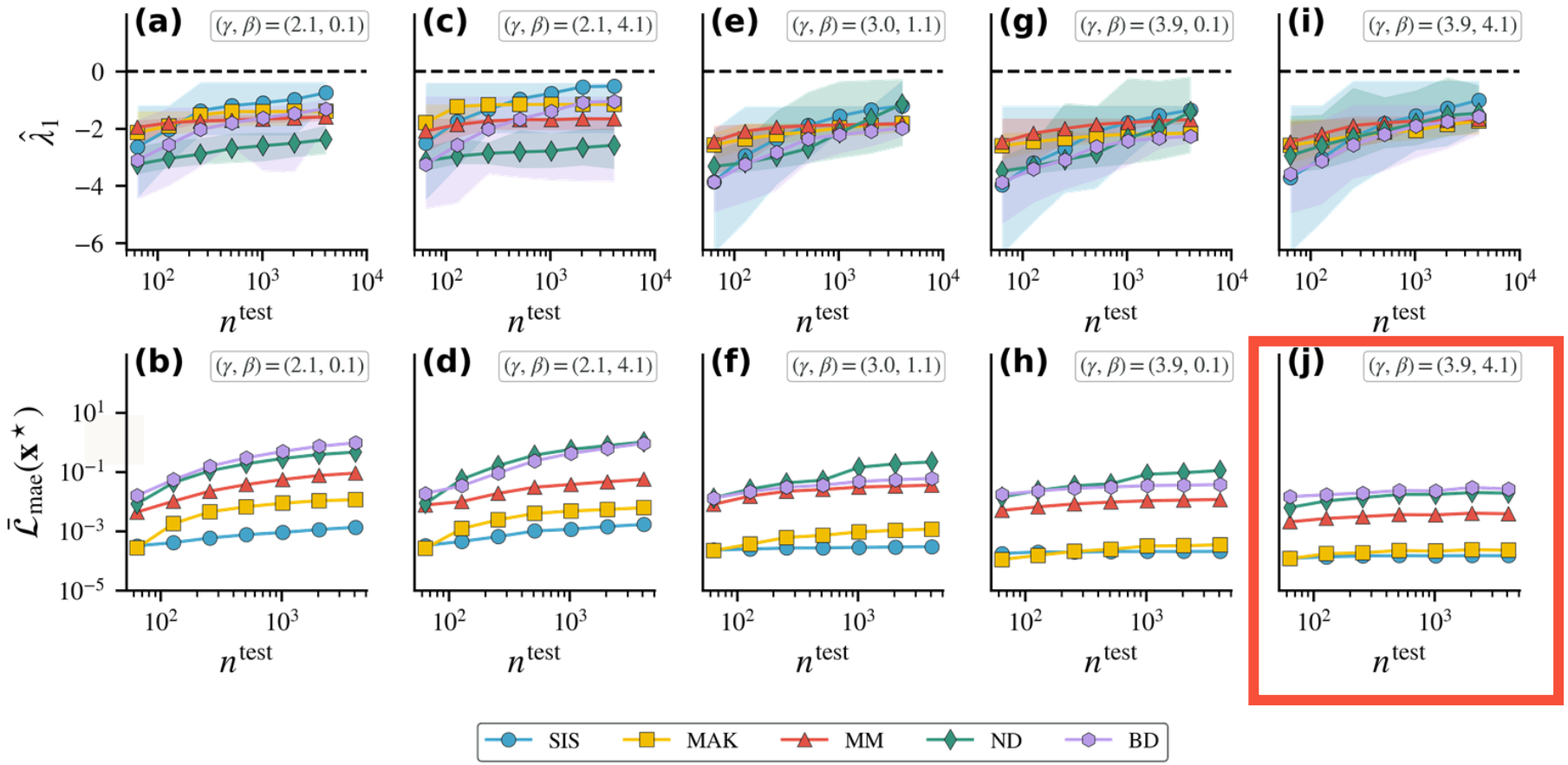
Local Stability

All models **approach stable fixed points**, but **potentially the wrong ones**.



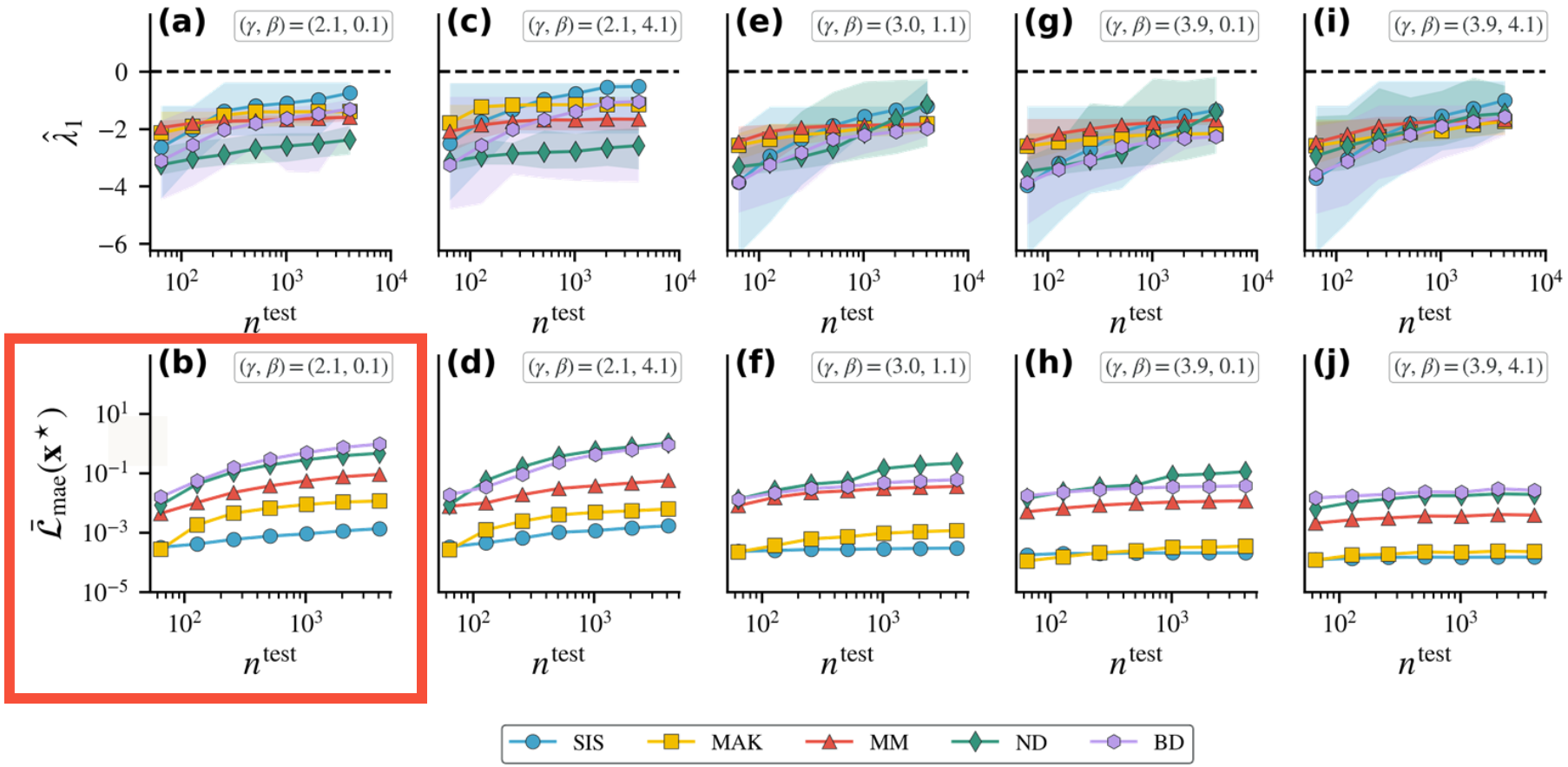
Local Stability

All models **approach stable fixed points**, but **potentially the wrong ones**.



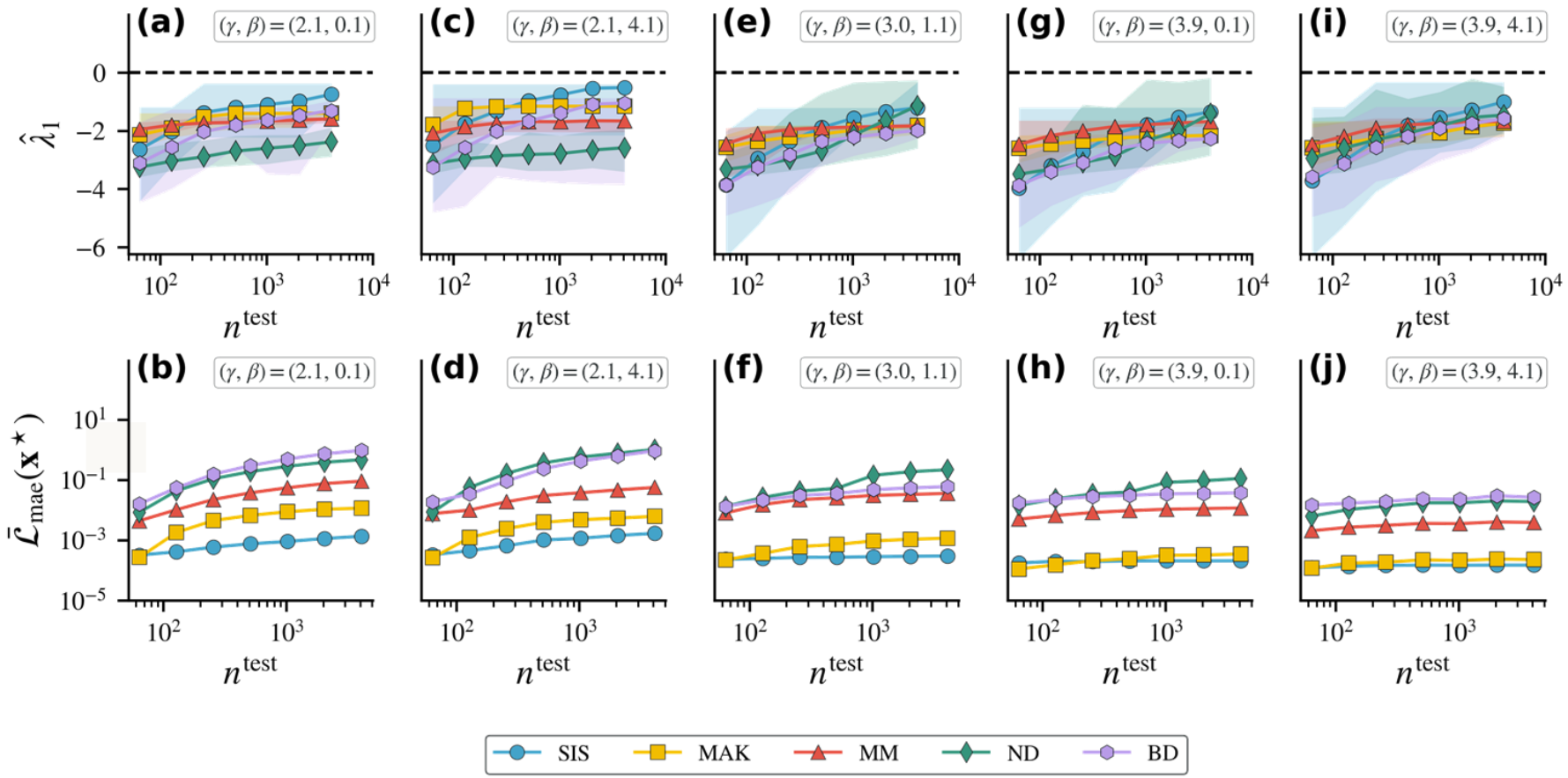
Local Stability

All models **approach stable fixed points**, but **potentially the wrong ones**.



Local Stability

All models **approach stable fixed points**, but **potentially the wrong ones**.

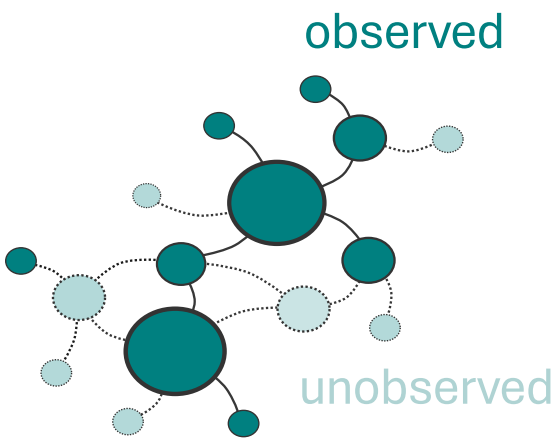


Missing Nodes

Question: Are neural ODEs robust to unobserved nodes at test time?

Motivation

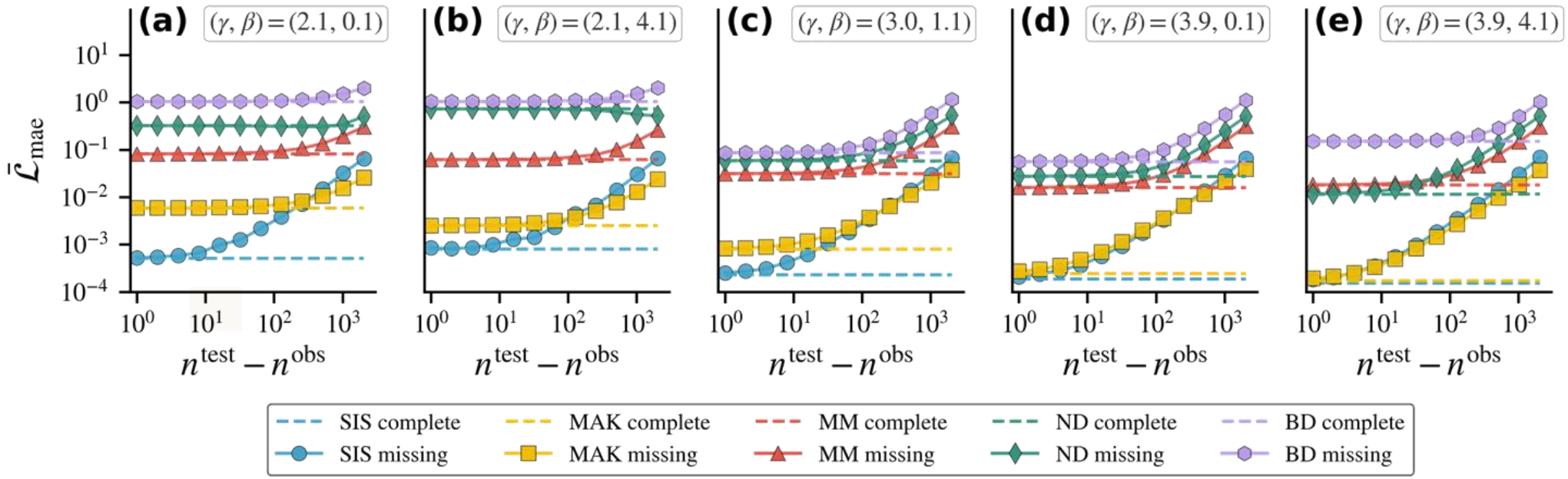
- In a large system, we might not have access to time series on all nodes.
- How the function of a system changes under node removal is a classic topic in network robustness [1].



[1] *Artime et al.* Nat. Rev. Phys. 6 114-131 (2024)

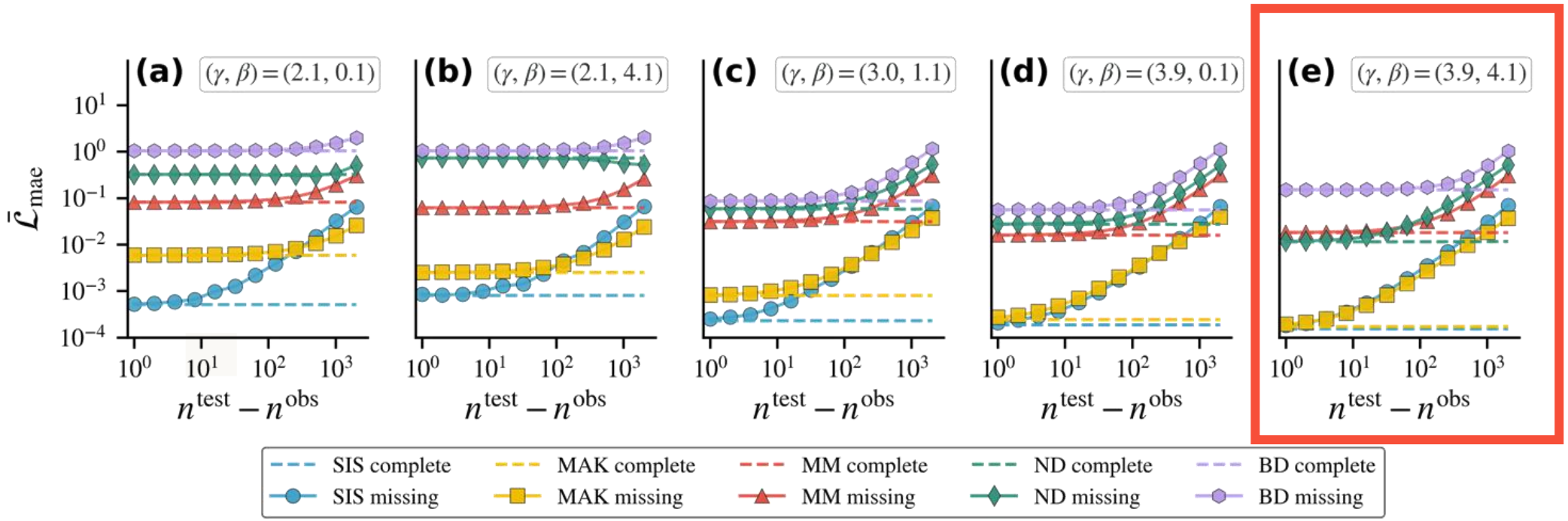
Missing Nodes

Performance degrades fast as nodes are missing.



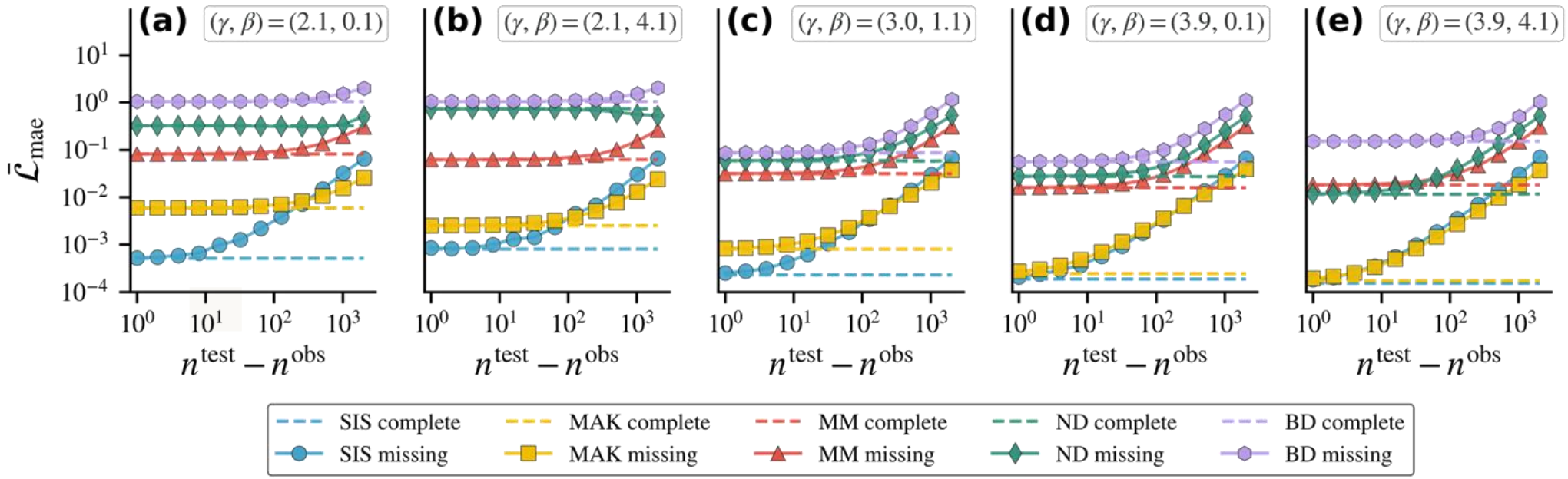
Missing Nodes

Performance degrades fast as nodes are missing.



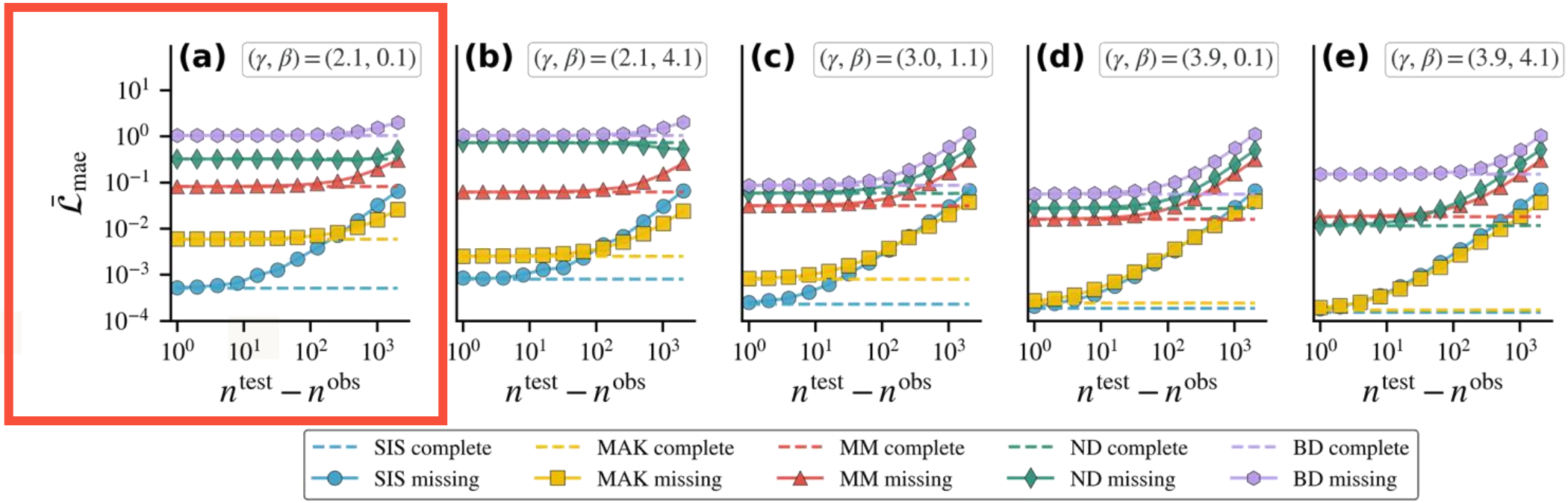
Missing Nodes

Performance degrades fast as nodes are missing.



Missing Nodes

Performance degrades fast as nodes are missing.



Conclusion

- Degree heterogeneity is a key factor for understanding generalization and robustness of graph neural ODEs.
- The work shows that random graph models are a way to systematically explore the space of possible data instances in graph machine learning.
- This work treats predictions as exogeneous to the complex system that is being modeled, but in a follow up work I am trying to relax this assumption.

Copyright

© 2026 Laber. Licensed under CC BY 4.0. If you use these materials, please cite:

Laber*, Dies*, Ehlert*, Klein & Eliassi-Rad *Effects of higher-order interactions and homophily on information access inequality* Communications Physics 9 13 (2026)

Laber, Eliassi-Rad & Klein *When do neural ordinary differential equations generalize on complex networks?* arXiv:2602.08980 (2026)

Laber & Klein *A Guide to Higher-Order Homophily* arXiv:2606.02537 (2026)