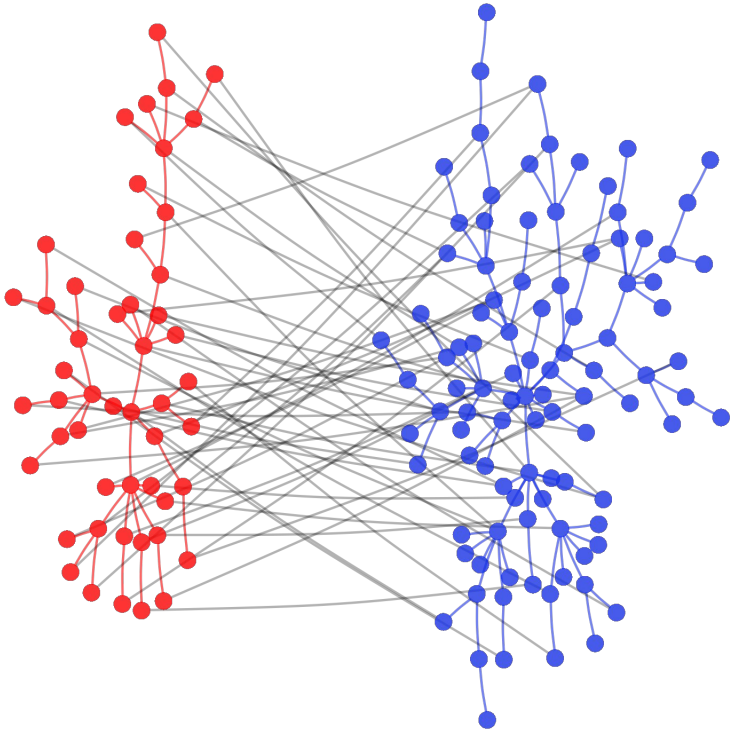
Topic discovery & diversity in a social network - *a toy model*

Tuan Pham

Net in Ecol & Evol Spr 21

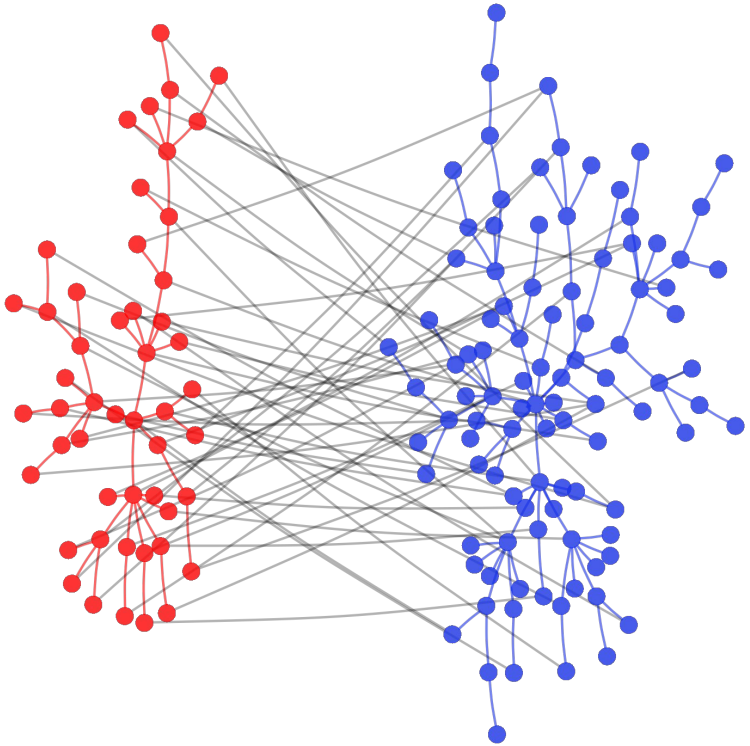
Project presentation



Why does it matter?

- Possibly (most definitely) not ...
- Diversity of topics (knowledge) as a result of self-learning and social influence
- Bigger picture would be be to consider growing/dynamic changes of agents and networks
- A fun *unrealistic* thought experiment: consider the *apocalypse*

Set up



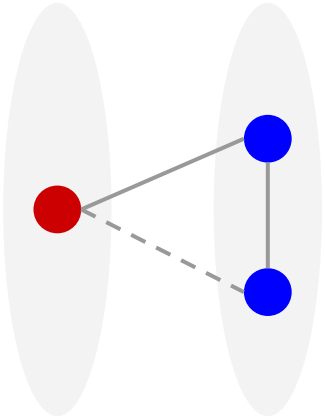
agents topics

A : adj. mat. **agent graph**

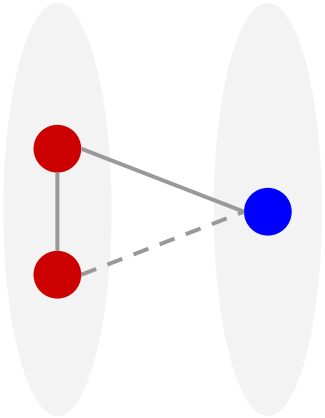
T : adj. mat. **topic graph**

τ : adj. mat. **learnt topics**

How to update



α - “rabbit hole”



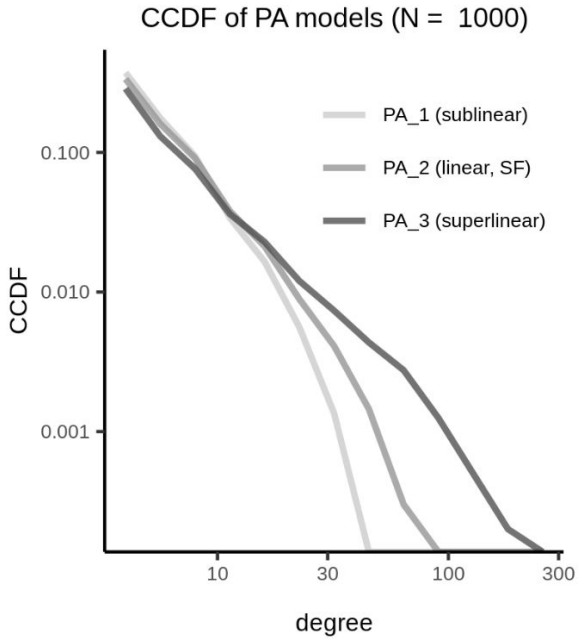
β - “recommender”

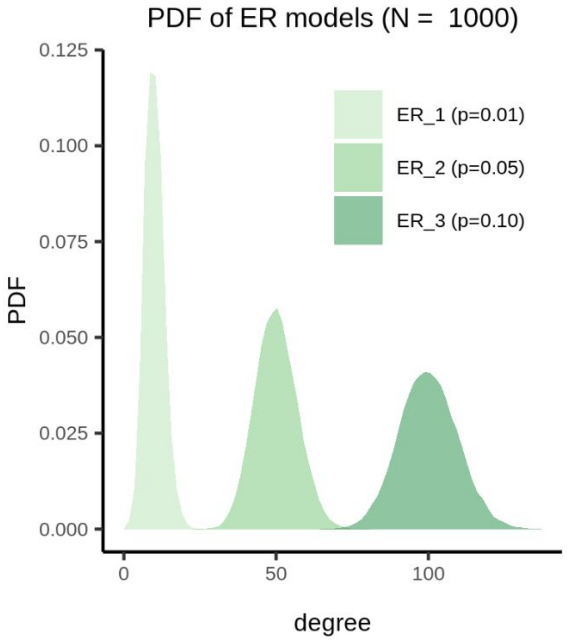
new topic = related to the known topics or learnt from friends

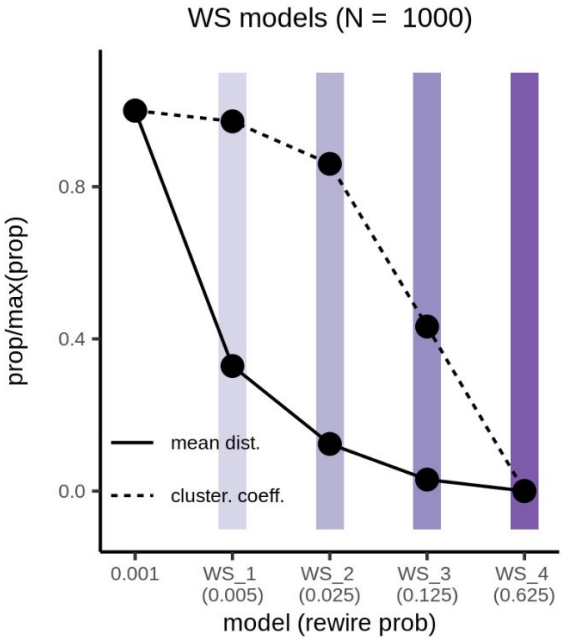
$$P = \alpha\psi\left([[T\tau]_\star - \tau]_\star\right) + \beta\psi\left([[\tau A]_\star - \tau]_\star\right)$$
$$\tau(t+1) \leftarrow \tau(t) + \text{sample}(P)$$
$$[x]_\star = 1 \text{ if } x > 0, 0 \text{ otherwise}$$
$$\psi(X) \text{ as column norm. for matrix } X$$

currently ignore serendipity, wandering & forgetting, strengths, directions

Intralayer *block* model generation







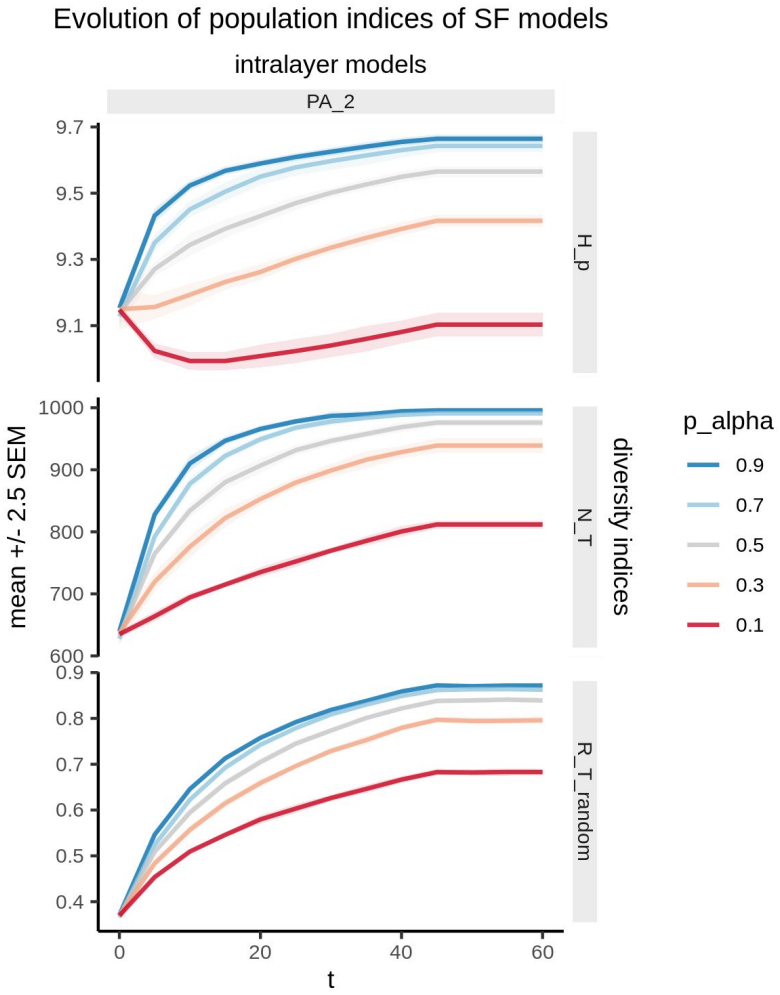
Diversity metric

Population

H_p: Topic entropy

N_T: # of topics

R_T: robustness due to removal of agents



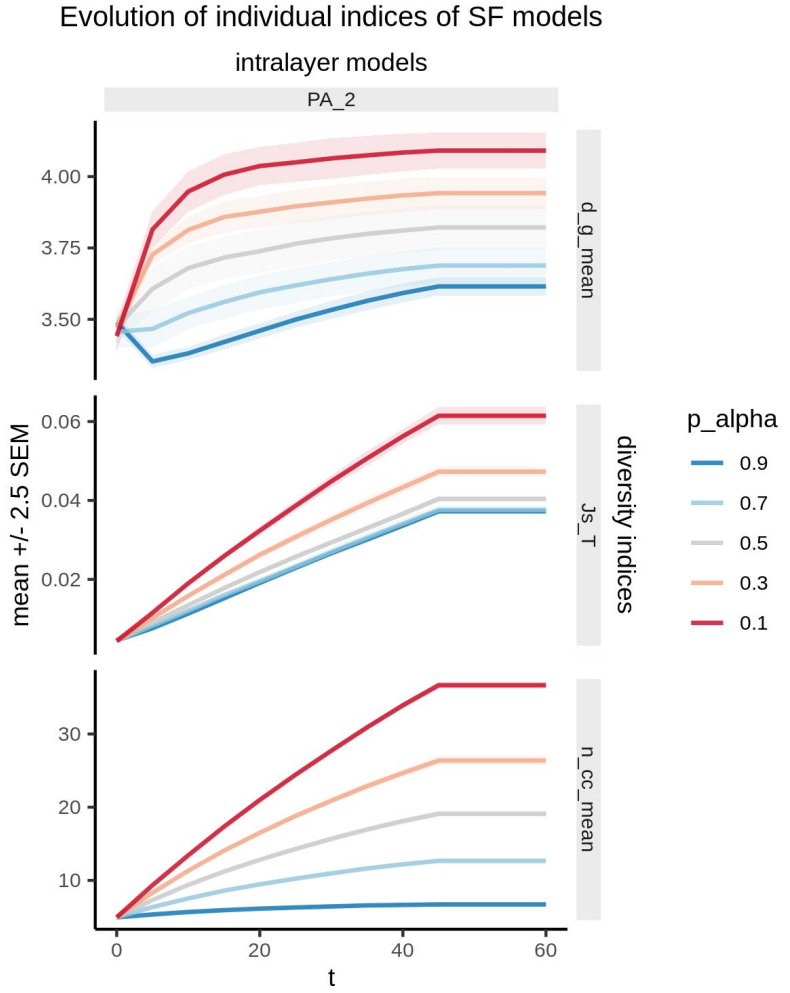
Diversity metric

Individual

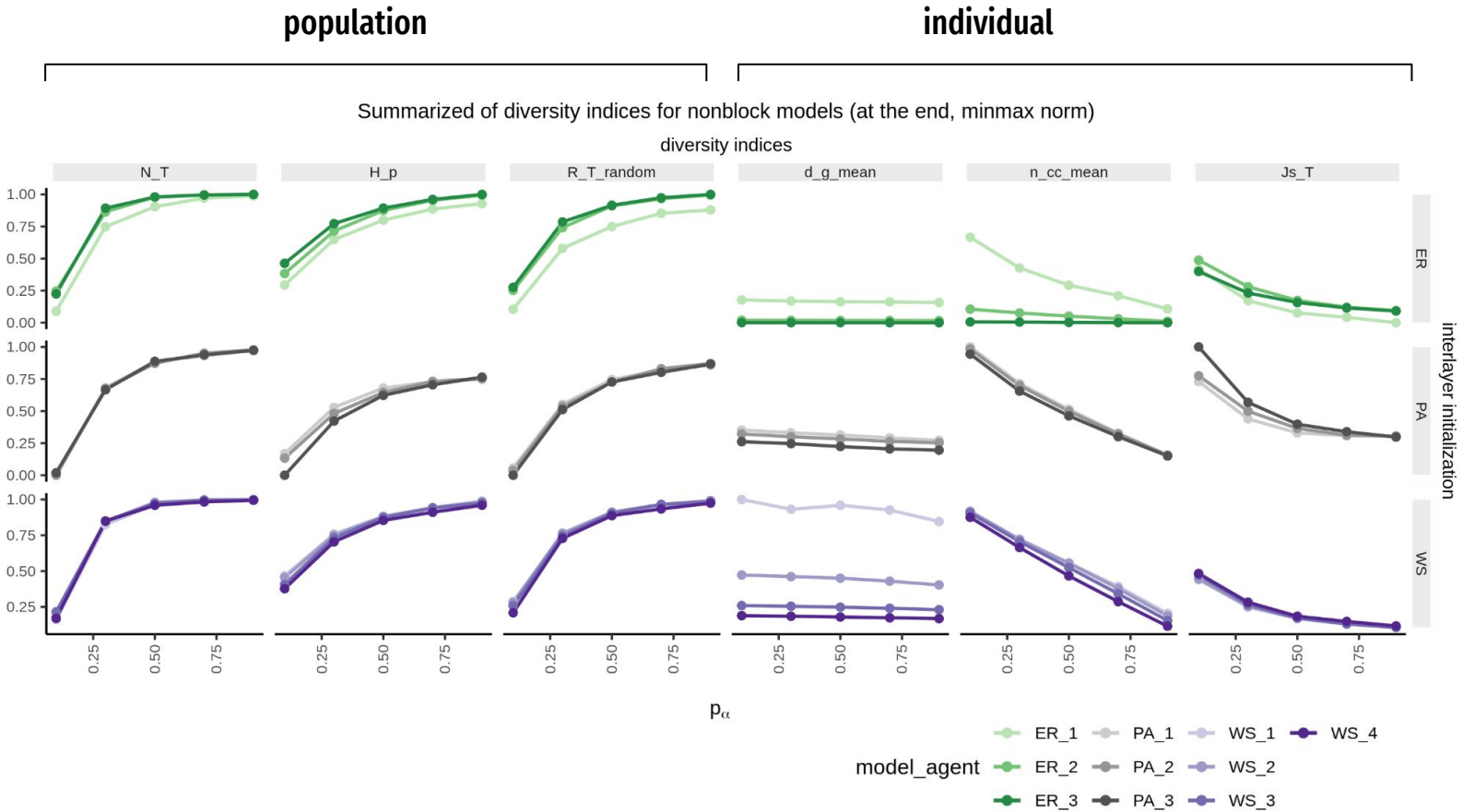
d_g: mean distance between agent's topic nodes in topic graph

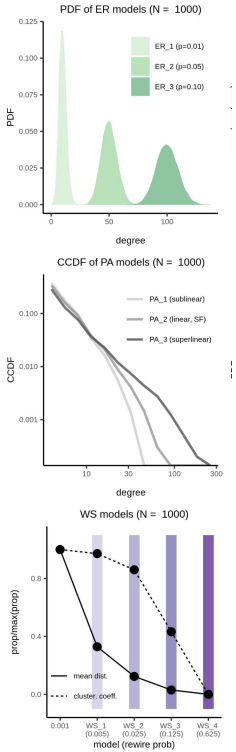
Js_T: mean pairwise overlap between topics of different agents

n_cc: # of conn comp in induced topic subgraphs



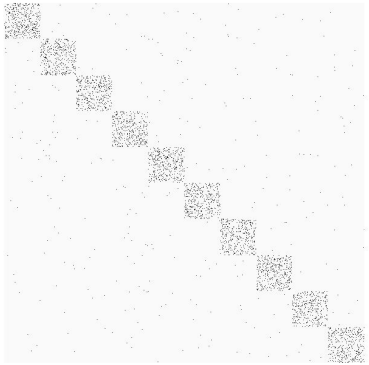
Across different intralayer *nonblock* models



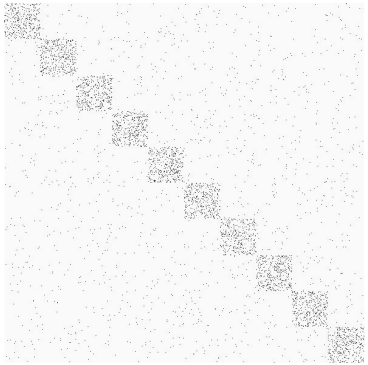


Intralayer *block* model generation

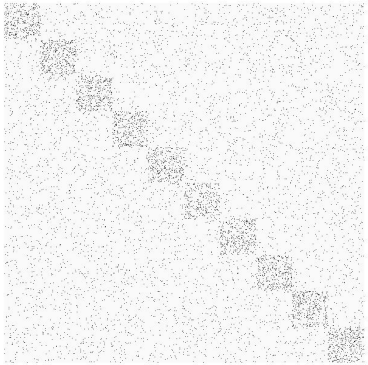
SBM_1
(within=0.1, between=0.001)



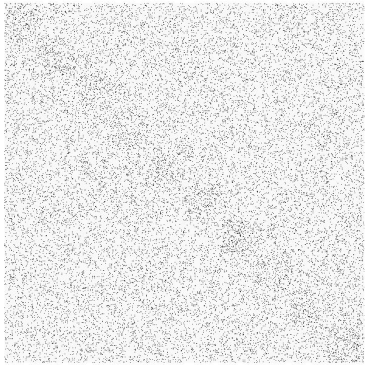
SBM_2
(within=0.1, between=0.004)



SBM_3
(within=0.1, between=0.016)



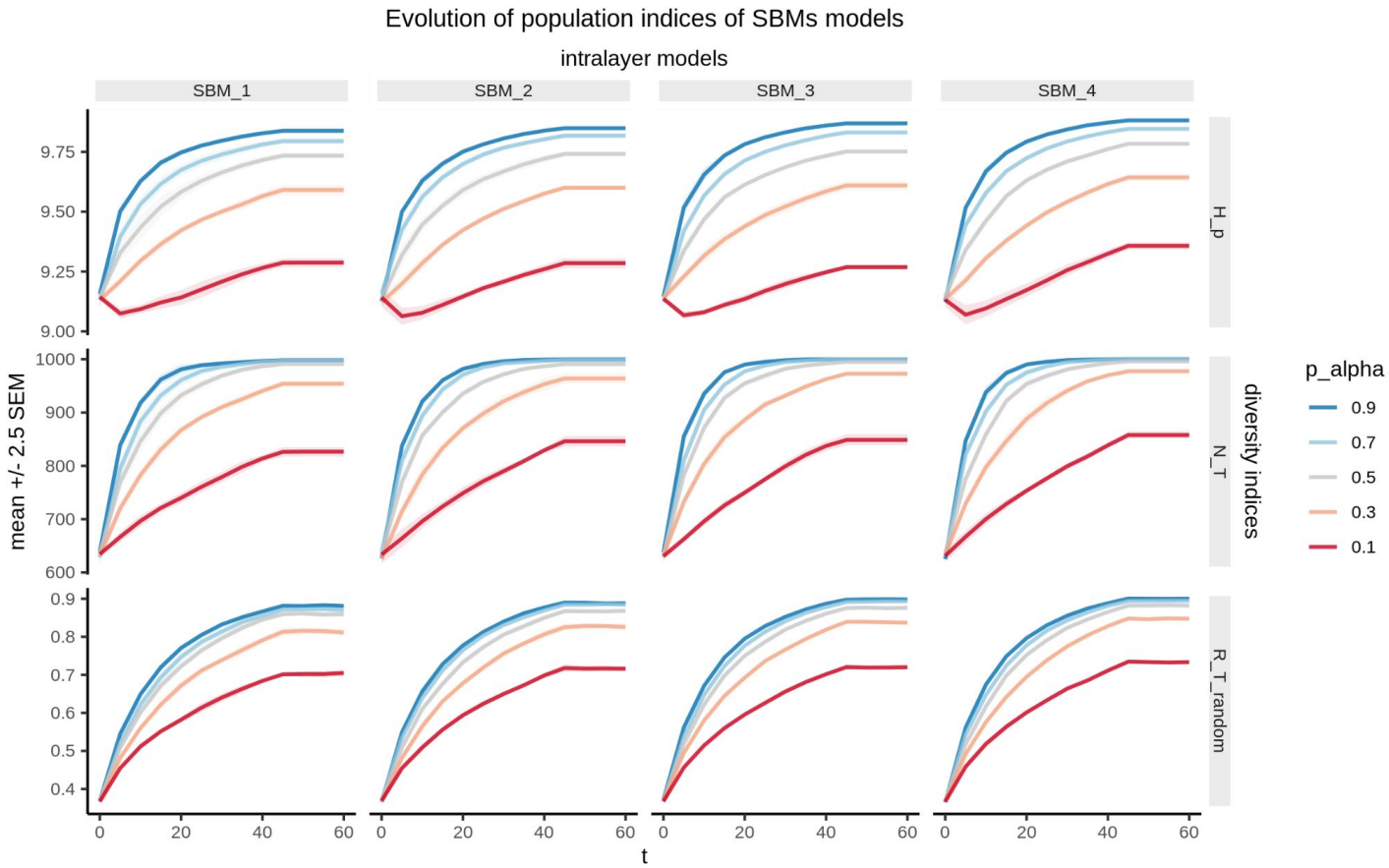
SBM_4
(within=0.1, between=0.064)



lower modularity

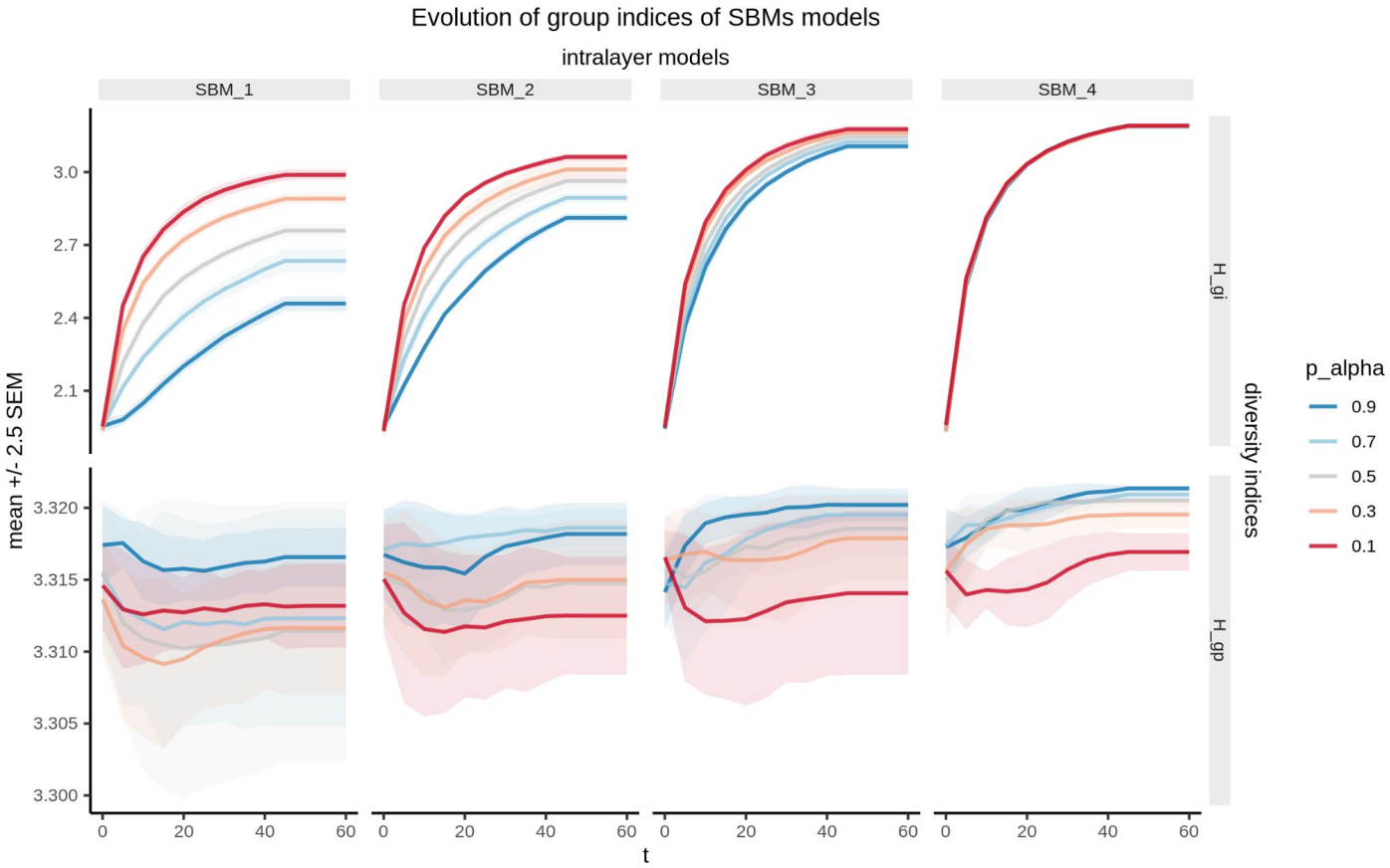
Population diversity

lower modularity

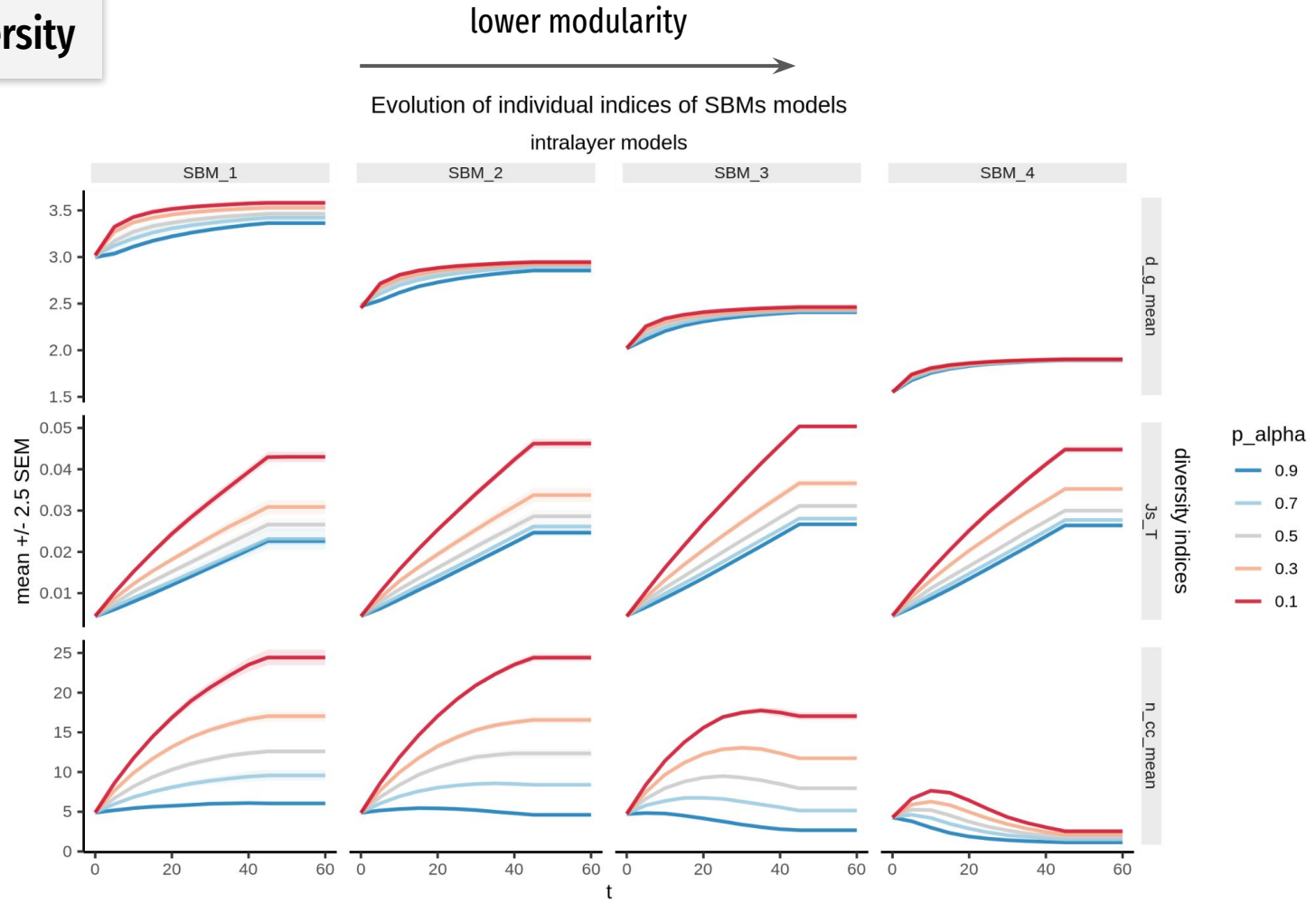


Group diversity

lower modularity



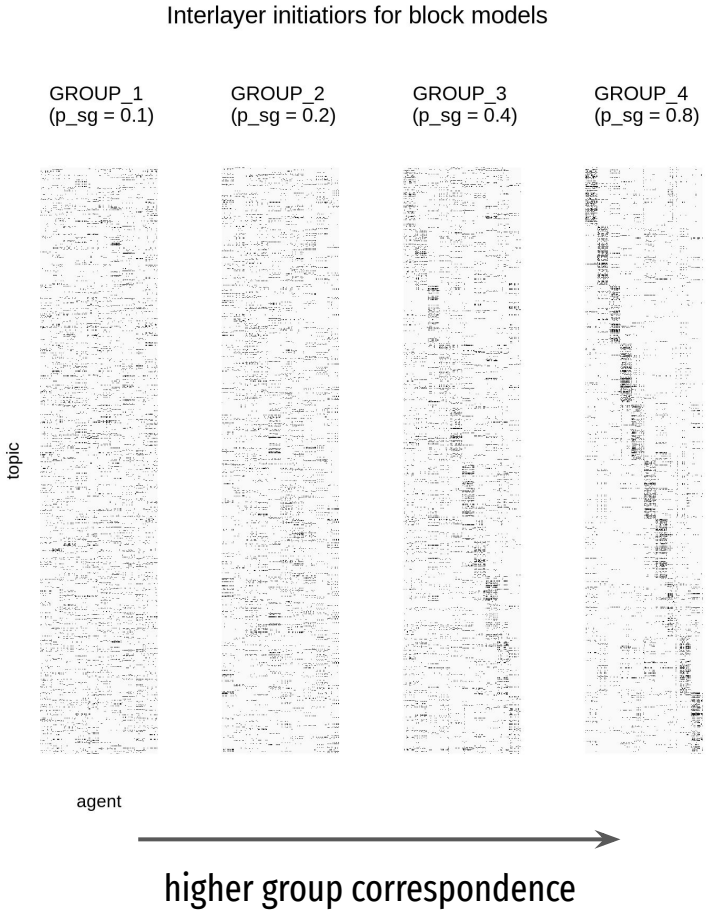
Individual diversity



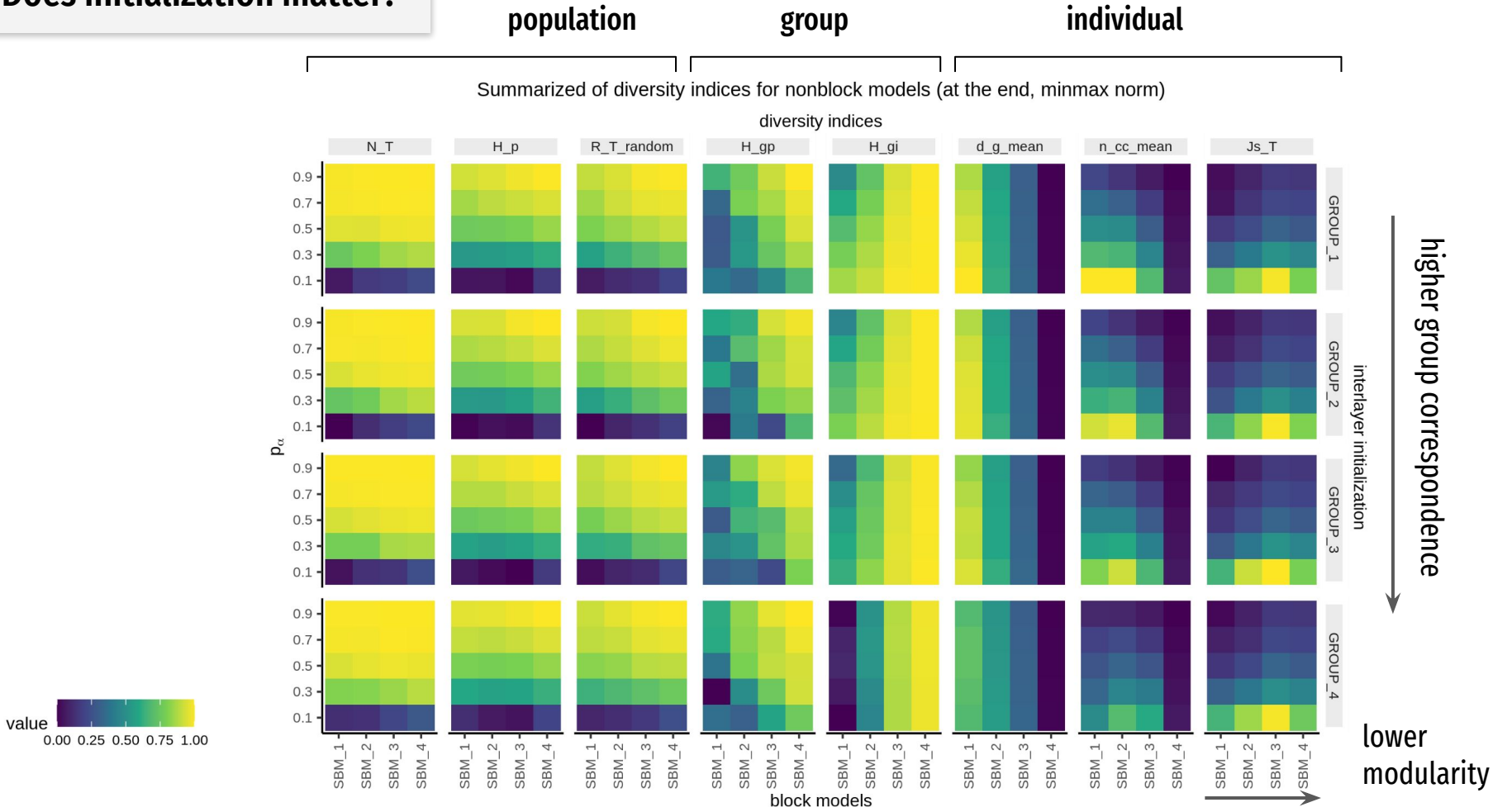
Does initialization matter?
Group correspondence

10 groups of topics
10 groups of agents

At initialization, connection between groups with prob **p_sg**



Does initialization matter?



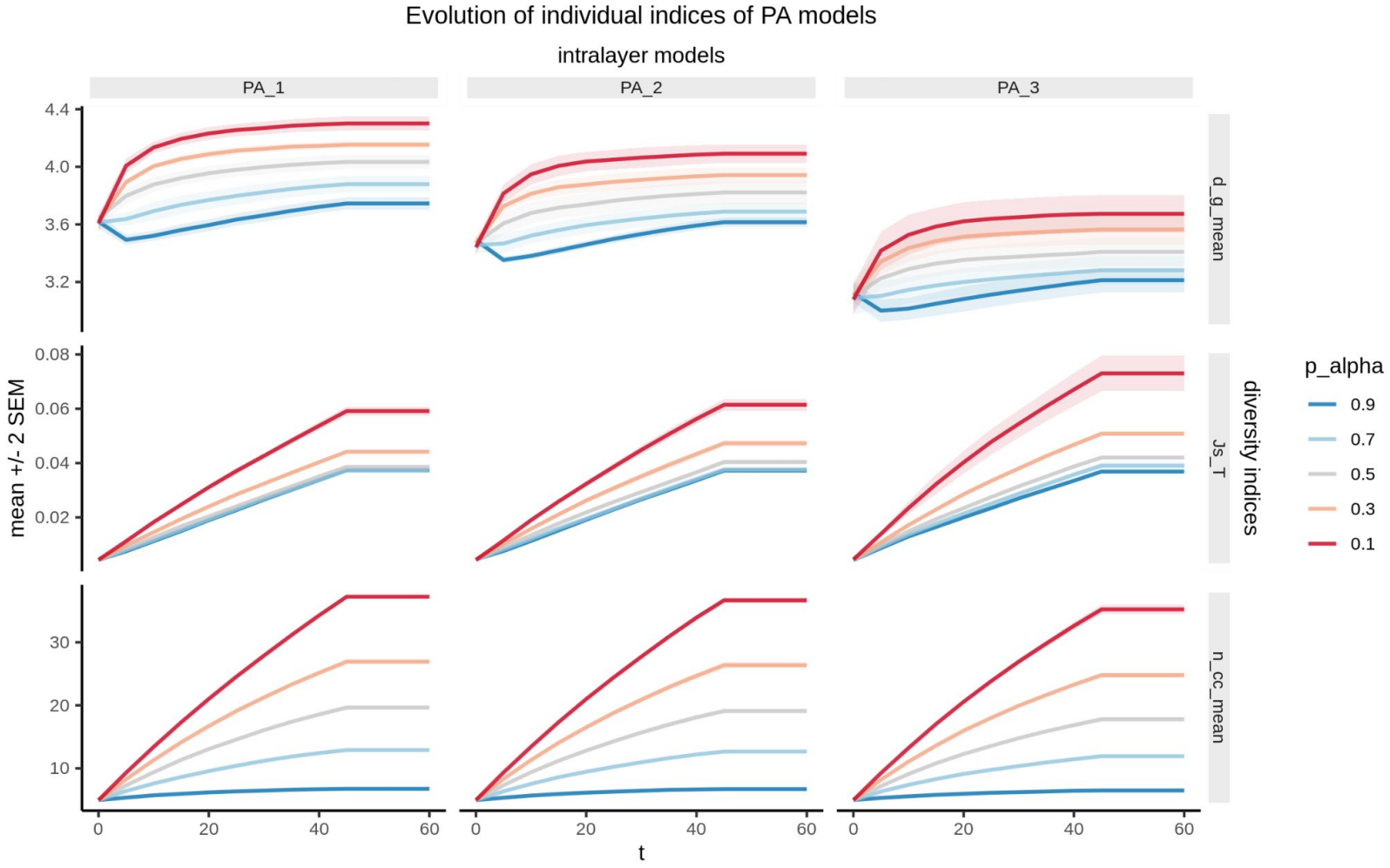
Conclusion

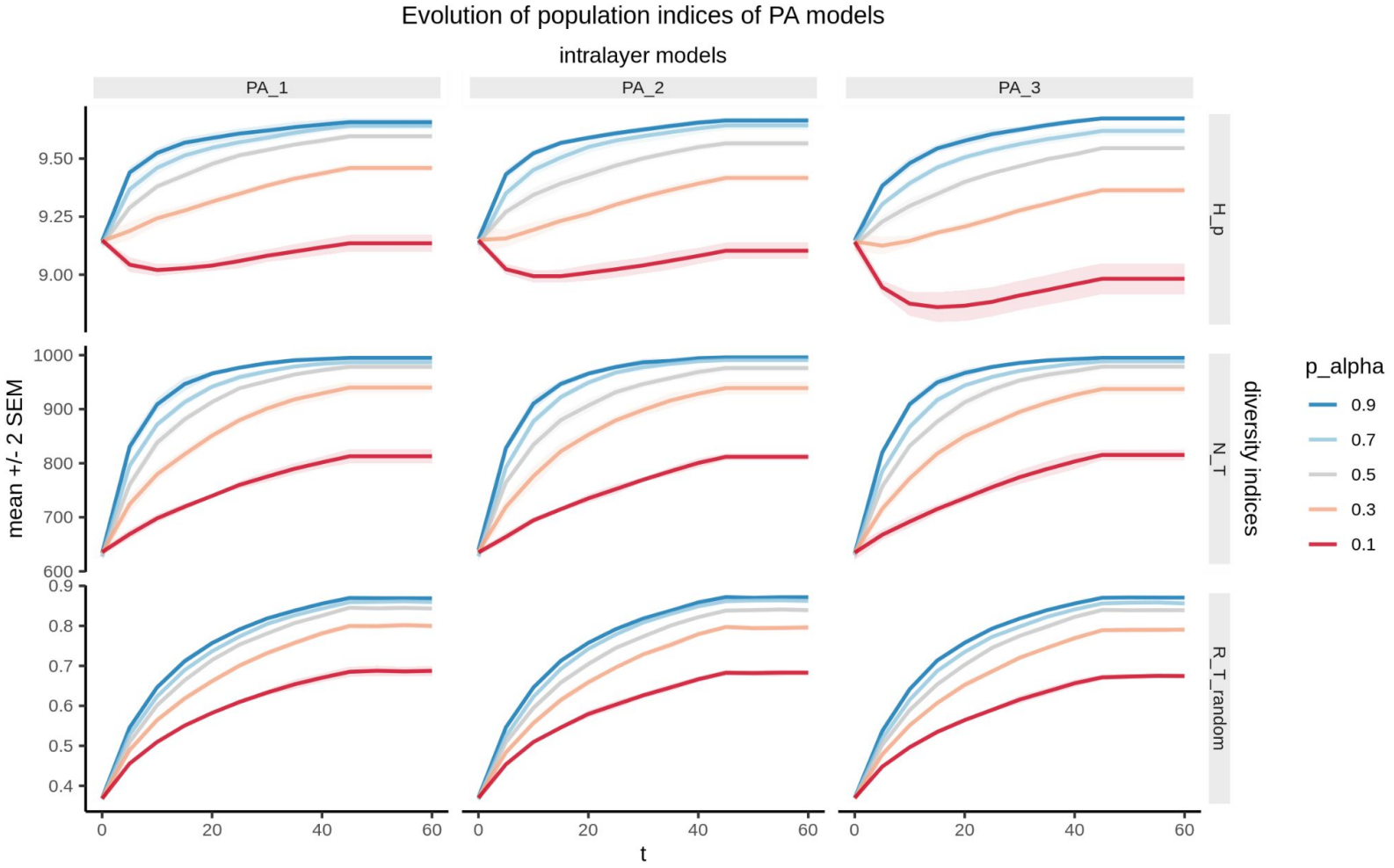
- Rabbit hole generally increases pop. div., reccomm increases ind. div.
- Group modularity and initial correspondence generally decreases diversity

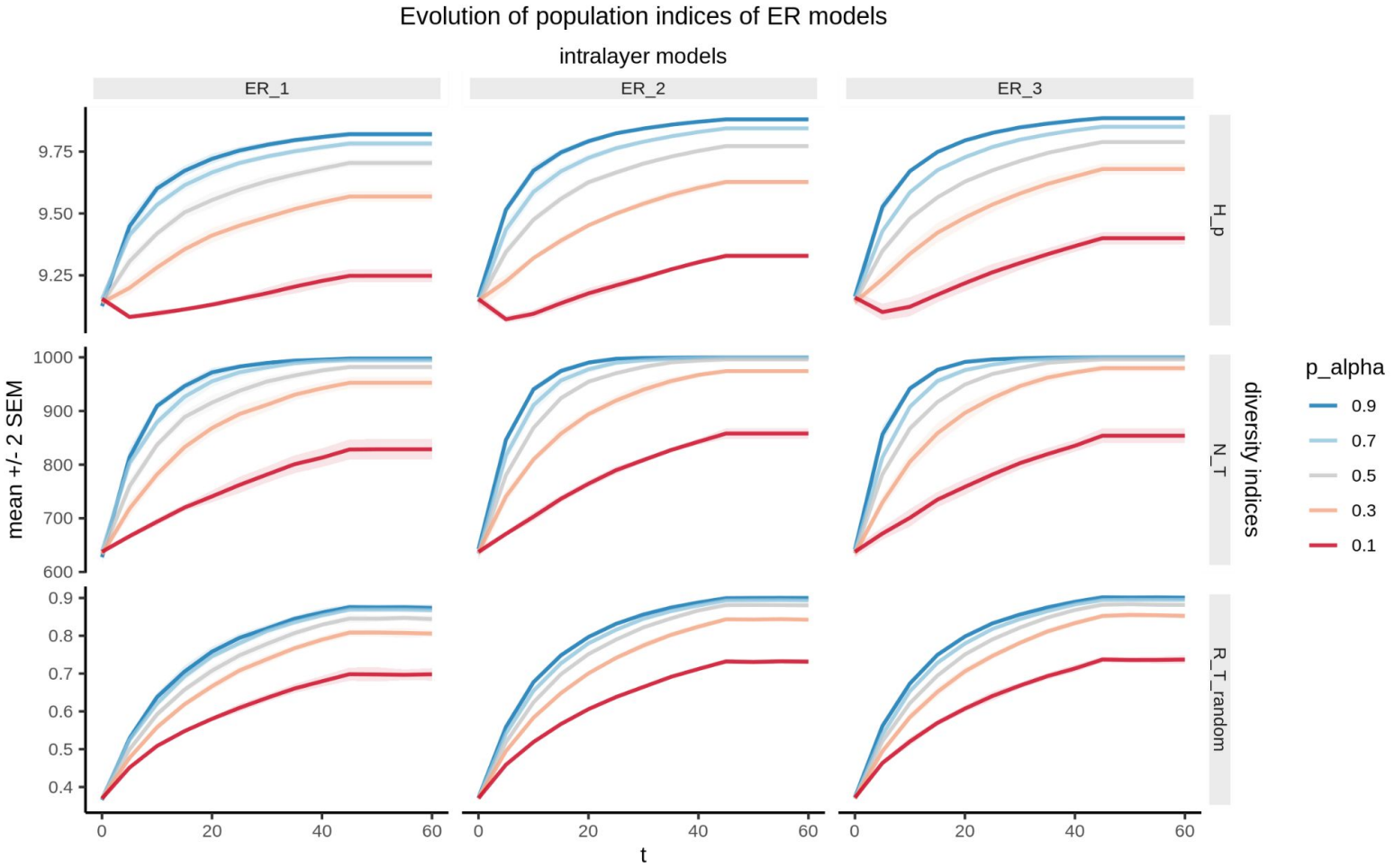
Future directions

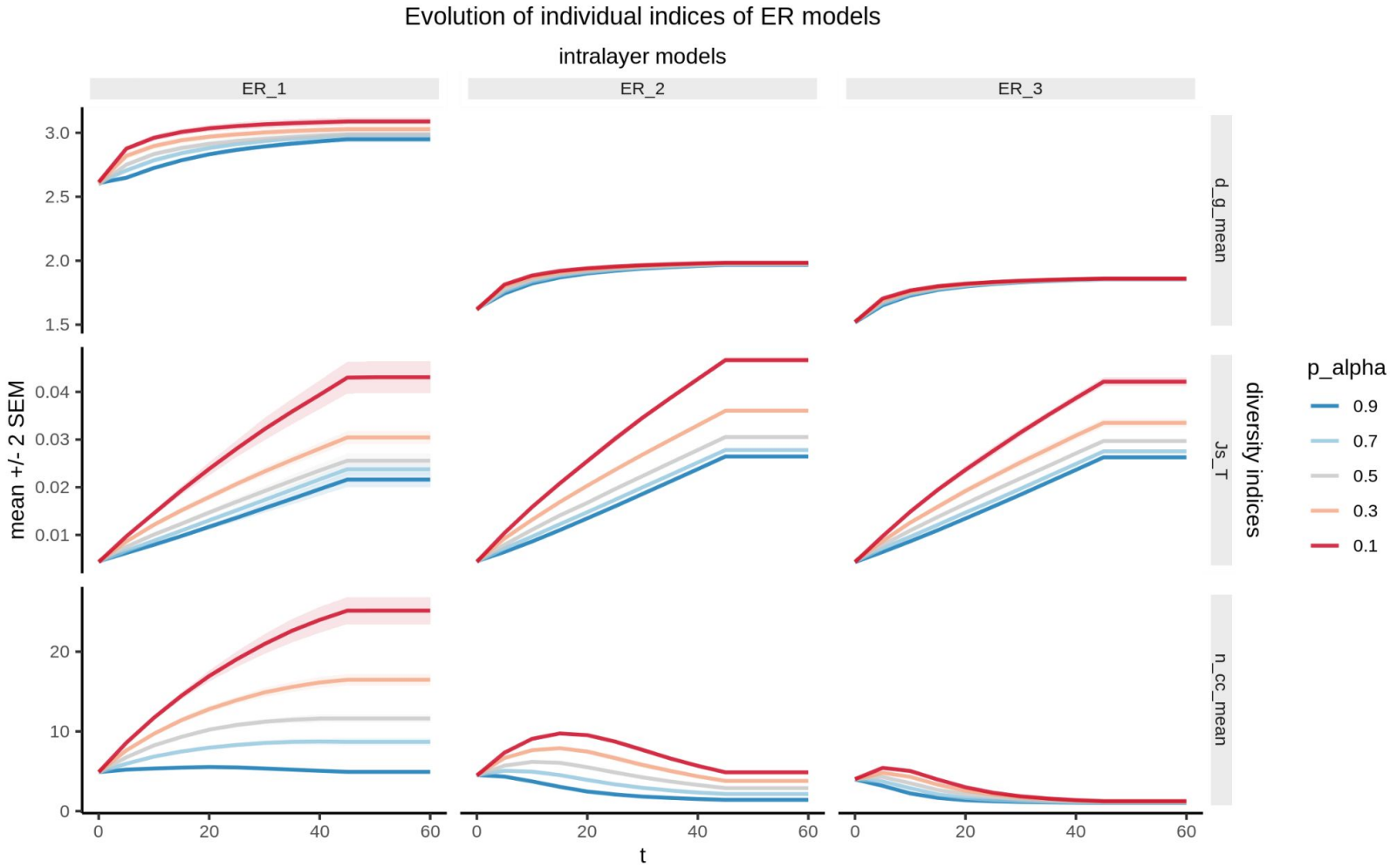
- Modularity in the *block* models
- Distribution of *specialists* and *generalists*
- Subsample real networks (FB + Wiki) or connected papers
- Other probabilities (serendipity/wandering, forgetting)
- Consider strengths + direction in networks + cost and bias
- Figure out if analytical results are possible
- Growing networks

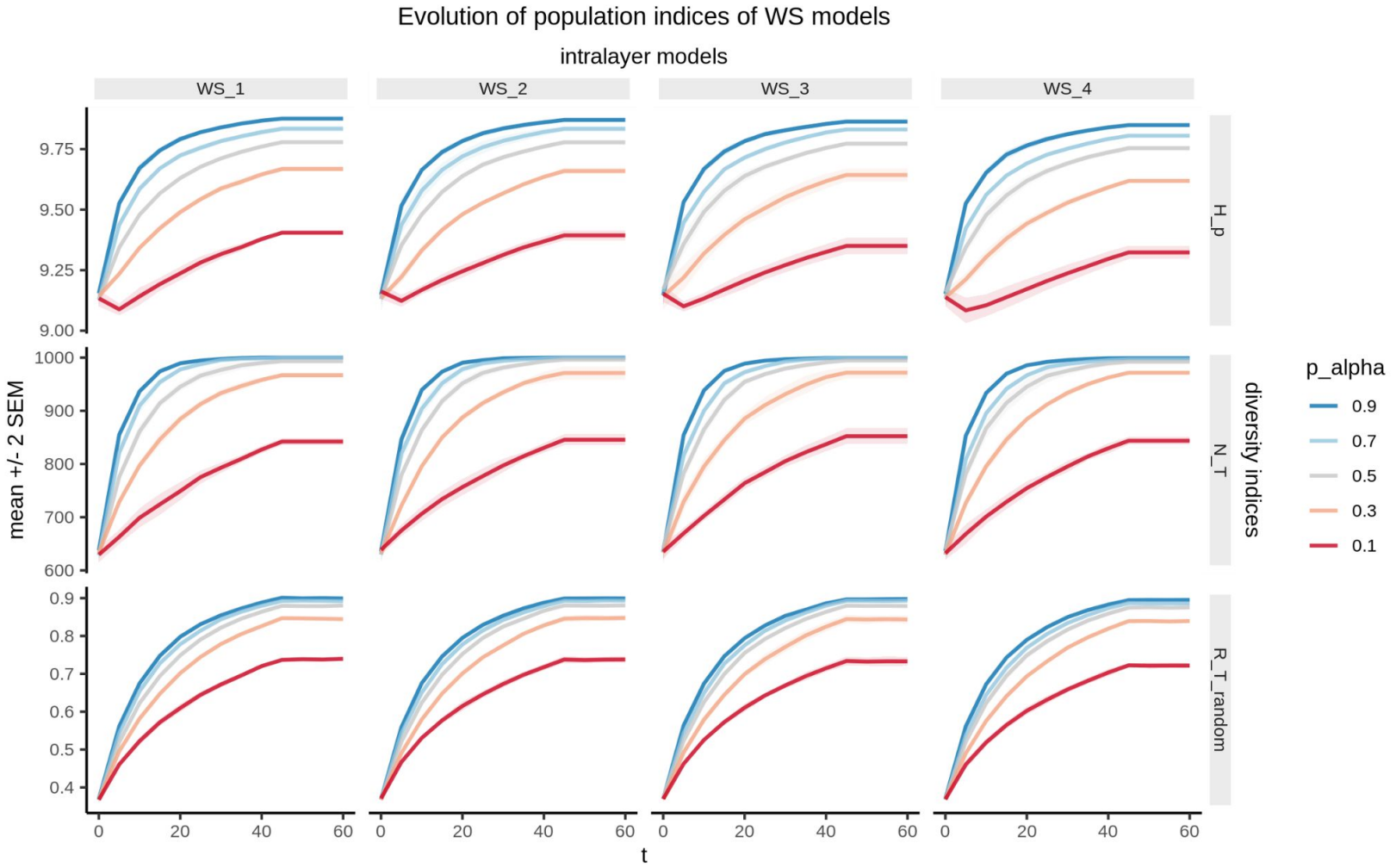
Thank you!

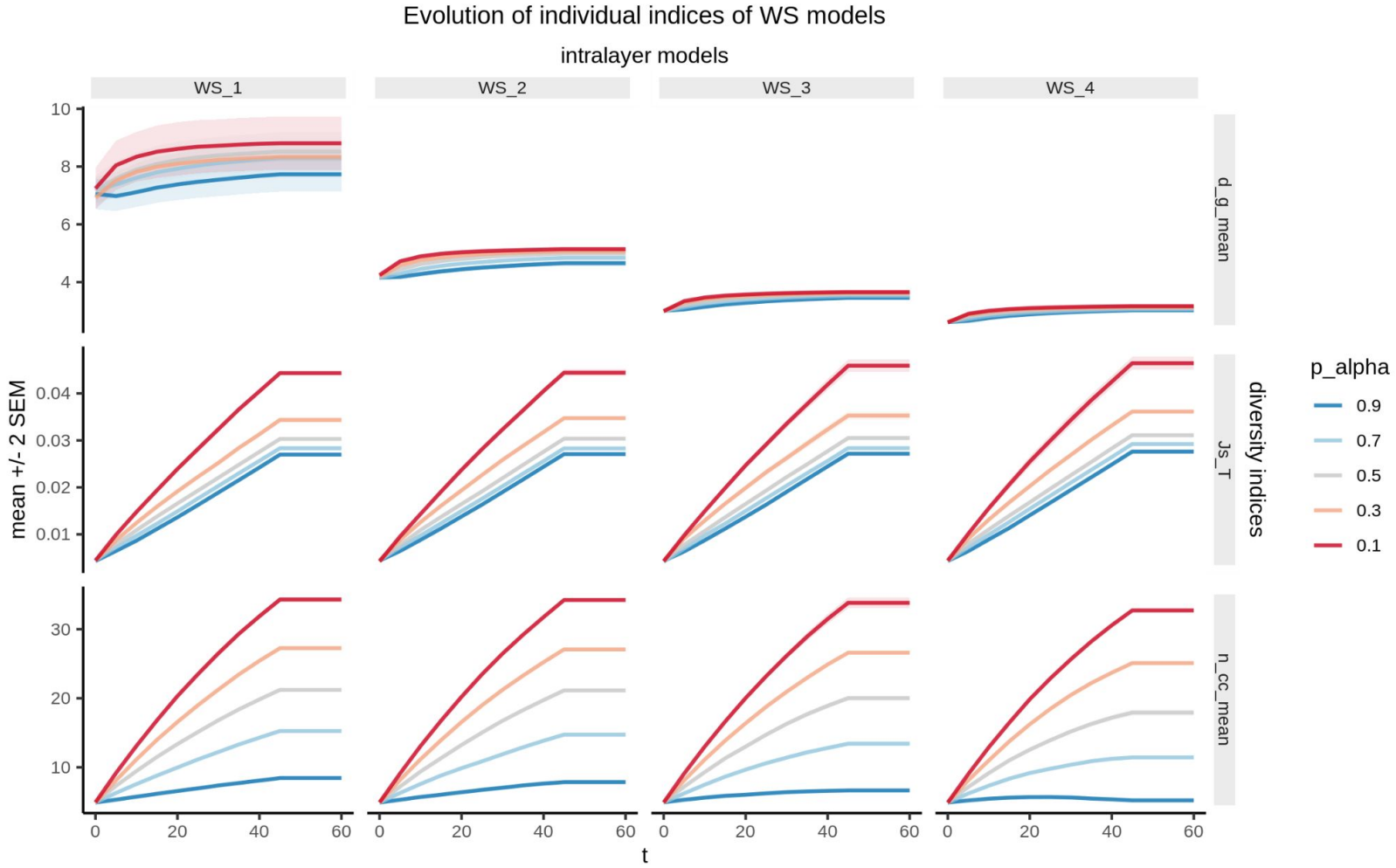

Evolution of individual indices of PA models
intralayer models
PA_1
PA_2
PA_3
d_g_mean
Js_T
n_cc_mean
diversity indices
mean +/- 2 SEM
t
p_alpha
0.9
0.7
0.5
0.3
0.1
4.4
4.0
3.6
3.2
0.08
0.06
0.04
0.02
30
20
10
0
20
40
60


Evolution of population indices of PA models
intralayer models
PA_1
PA_2
PA_3
H_p
N_T
R_T_random
diversity indices
mean +/- 2 SEM
t
p_alpha
0.9
0.7
0.5
0.3
0.1

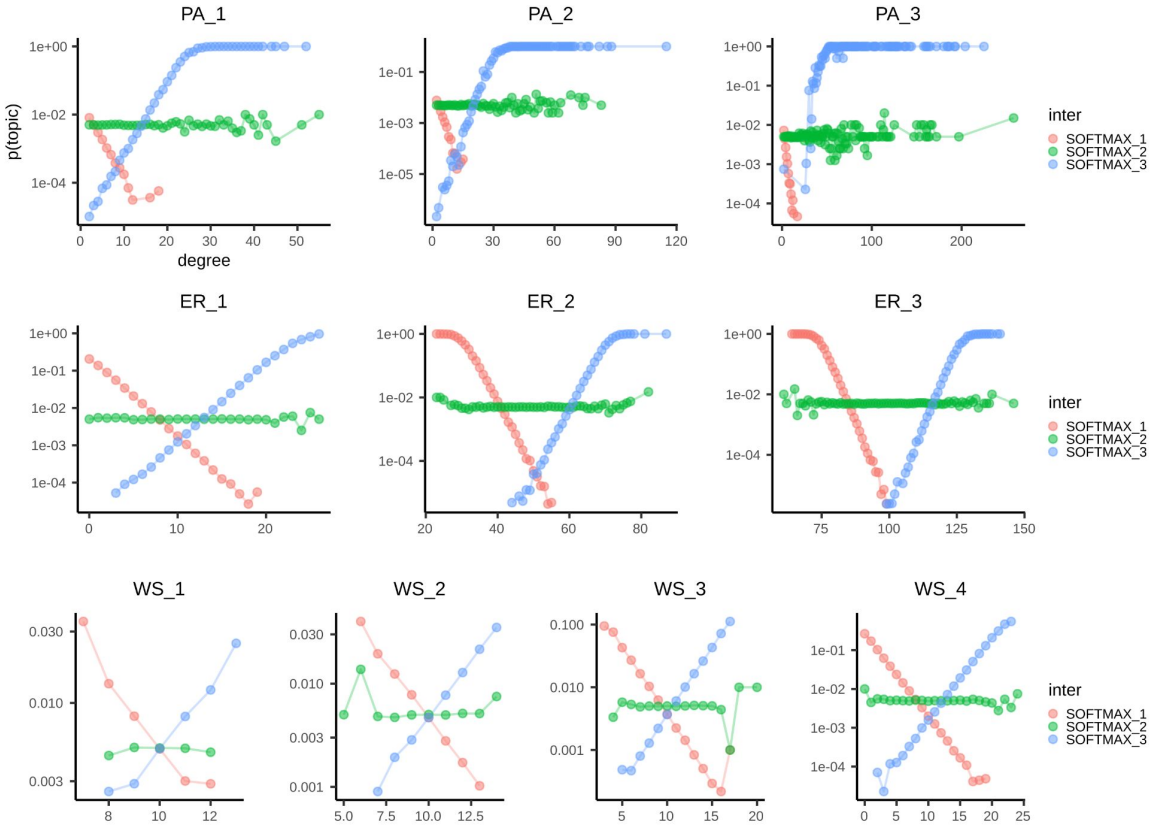

Evolution of population indices of ER models
intralayer models
ER_1
ER_2
ER_3
H_p
N_T
R_T_random
diversity indices
mean +/- 2 SEM
t
p_alpha
0.9
0.7
0.5
0.3
0.1

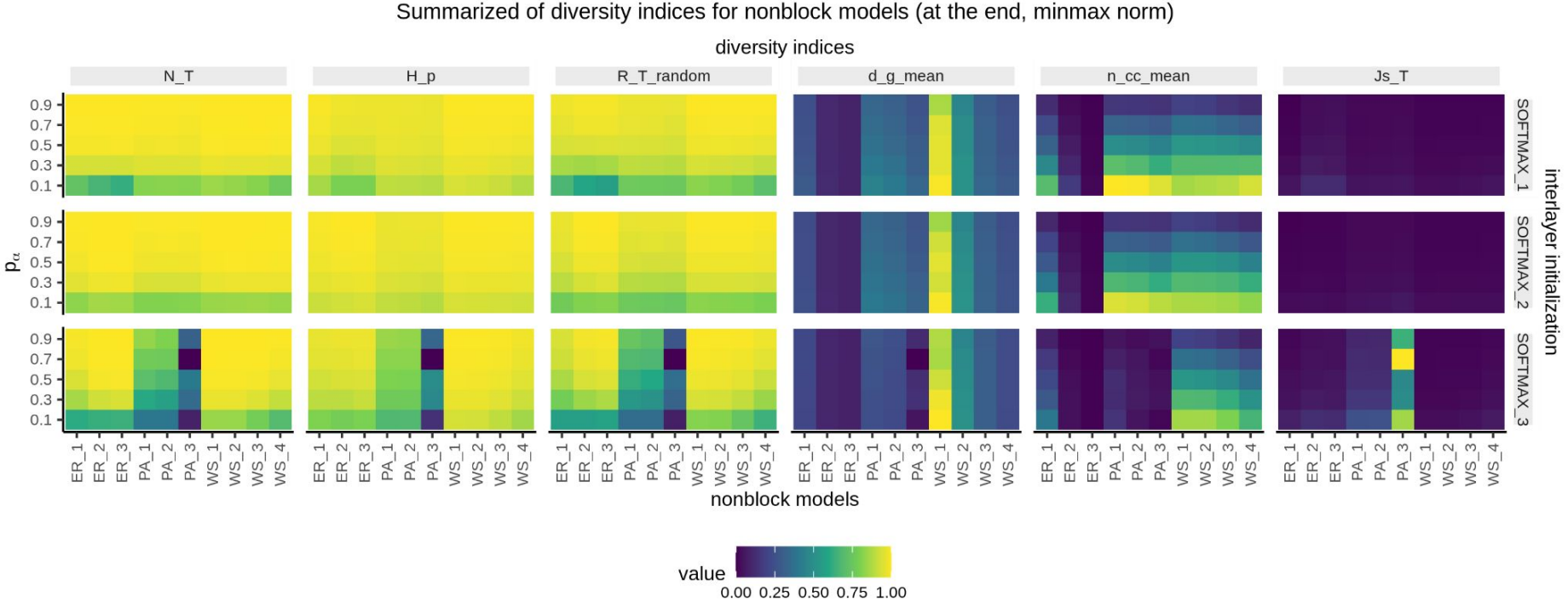

Evolution of individual indices of ER models
intralayer models
ER_1
ER_2
ER_3
d_g_mean
Js_T
n_cc_mean
diversity indices
mean +/- 2 SEM
t
p_alpha
0.9
0.7
0.5
0.3
0.1


Evolution of population indices of WS models
intralayer models
WS_1
WS_2
WS_3
WS_4
H_p
N_T
R_T_random
diversity indices
mean +/- 2 SEM
t
p_alpha
0.9
0.7
0.5
0.3
0.1


Evolution of individual indices of WS models
intralayer models
WS_1
WS_2
WS_3
WS_4
d_g_mean
Js_T
n_cc_mean
diversity indices
mean +/- 2 SEM
t
p_alpha
0.9
0.7
0.5
0.3
0.1

Interlayer initiatiors for nonblock models




Summarized of diversity indices for nonblock models (at the end, minmax norm)
diversity indices
N_T
H_p
R_T_random
d_g_mean
n_cc_mean
Js_T
SOFTMAX_1
SOFTMAX_2
SOFTMAX_3
interlayer initialization
p_α
0.9
0.7
0.5
0.3
0.1
ER_1
ER_2
ER_3
PA_1
PA_2
PA_3
WS_1
WS_2
WS_3
WS_4
nonblock models
value
0.00 0.25 0.50 0.75 1.00