

Supplemental Material for Hills, T. T. **Cognitive network enrichment not degradation explains the aging mental lexicon and links fluid and crystallized intelligence**

2024-04-11

The primary purpose of the Supplemental Material is to demonstrate that the results presented in the main text are robust to a wide variety of manipulations with respect to environment construction and learning. I have chosen representative examples which clearly demonstrate the key effect of these variations. The results are similar across environment types. The code available online allows further exploration of the other environments or assumptions. All results are based on 100 simulations for a given set of parameters. This was more than sufficient to generate robust and reliable patterns in the data.

1 Network size

The size of the network is inconsequential to the qualitative patterns of results presented in the main manuscript. However, the network size is consequential to rate at which enrichment effects saturate. Here I provide an example using the scale-free small world network with $M = 5$ and $\gamma = 1$, with network sizes of 500, 1000, and 2000. The weighted edge distribution in each environment is constructed by sampling 2000 pairs of nodes as described in the main text. This produces the results shown in Figure 1.

The results show the rising of entropy and the falling of similarity judgment as before. However, we can also see that as the number of nodes increases this slows the rise of entropy and the fall of similarity.

Note: For those experimenting with the code, as the number of nodes increases environments with the same number of associations will produce sparser representations. This reaches a computational boundary when the representation's giant component contains fewer than the 20 pairs required for the similarity computation. For this reason, I only show the scale-free small-world here. But denser small-word and Erdős-Renyi random graphs produce the same relationship as the network size increases.

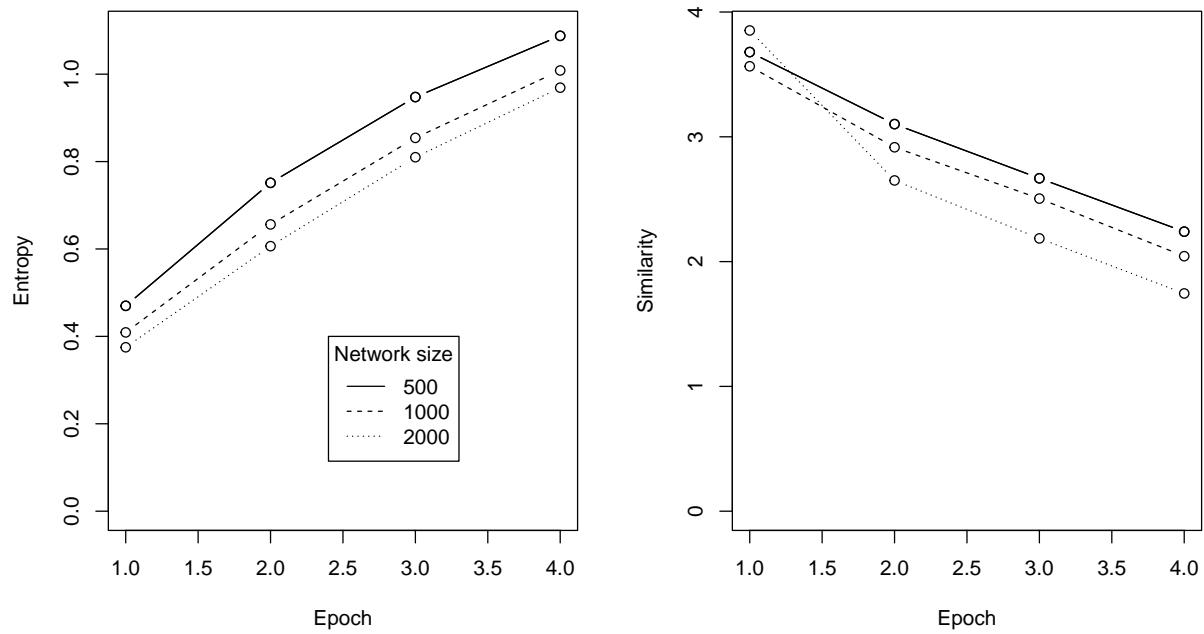


Figure 1: Adjusting the environment size. Entropy and similarity measures computed from the cognitive representation at different epochs of development for various network sizes holding the number of sampled pairs constant. Epochs are 250 learning events each.

2 Weighted density distribution in the environment

Figure 2 shows how increasing the number of associative relations in the network influences entropy and similarity. The number of concepts (nodes) in the environment is fixed and the number of associative samples used in constructing the weighted edge distribution is increased. This uses the scale-free small world network with $M = 5$ and $\gamma = 1$, with 500 nodes and 2000, 5000, and 10000 pairs sampled to construct the environment. As above, $\beta = .1$.

The results show rising entropy and falling similarity. Changes in the weighted edge distribution have a limited influence on these measures. This is largely because the learning advantage for strongly weighted edges is only relatively amplified by increasing the number of associations. When a sufficient number of samples are already in place to establish this relative difference in the environment, then adding additional weights to edges has a limited effect on learning.

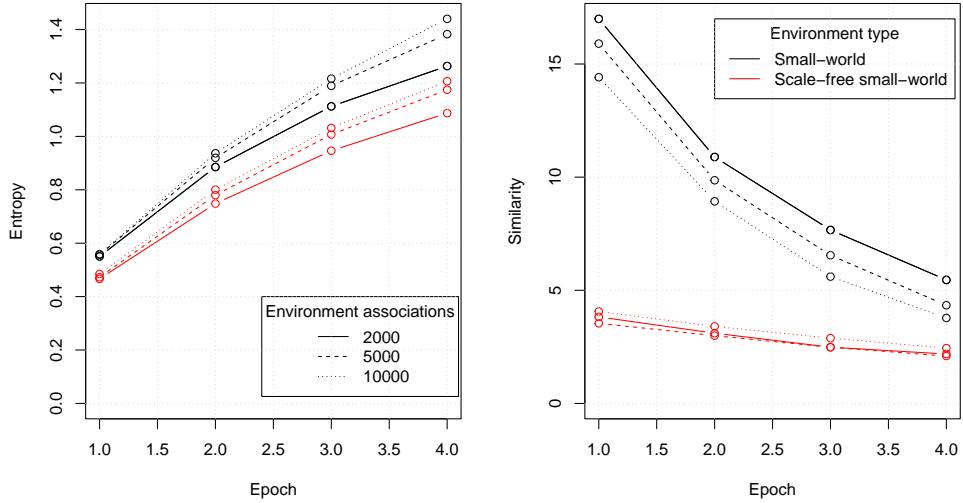


Figure 2: Varying the distribution of weighted edges in the environmental network. Edges are sampled to create the environment network with 2000, 5000, or 10000 samples, always using 500 concept nodes. Epochs are 250 training events long.

3 Learning events

Figure 3 varies the number of learning events. This uses the scale-free small world network with $M = 5$ and $\gamma = 1$ and learning rate is set to $\beta = 0.1$. The number of learning events per epoch varies from 250, to 500, to 5000.

The general pattern of rising entropy and falling similarity is preserved for all cases. However, as the amount of learning increases relative to a fixed environment size, maximum entropy and minimum similarity are reached more quickly. This is consistent with the previous results showing that increasing the network size lowered the rate at which these values changed. This suggests a saturation effect, where with sufficient learning learners can maximize the available entropy, which in turn maximizes diffusion leading to lower similarity judgements.

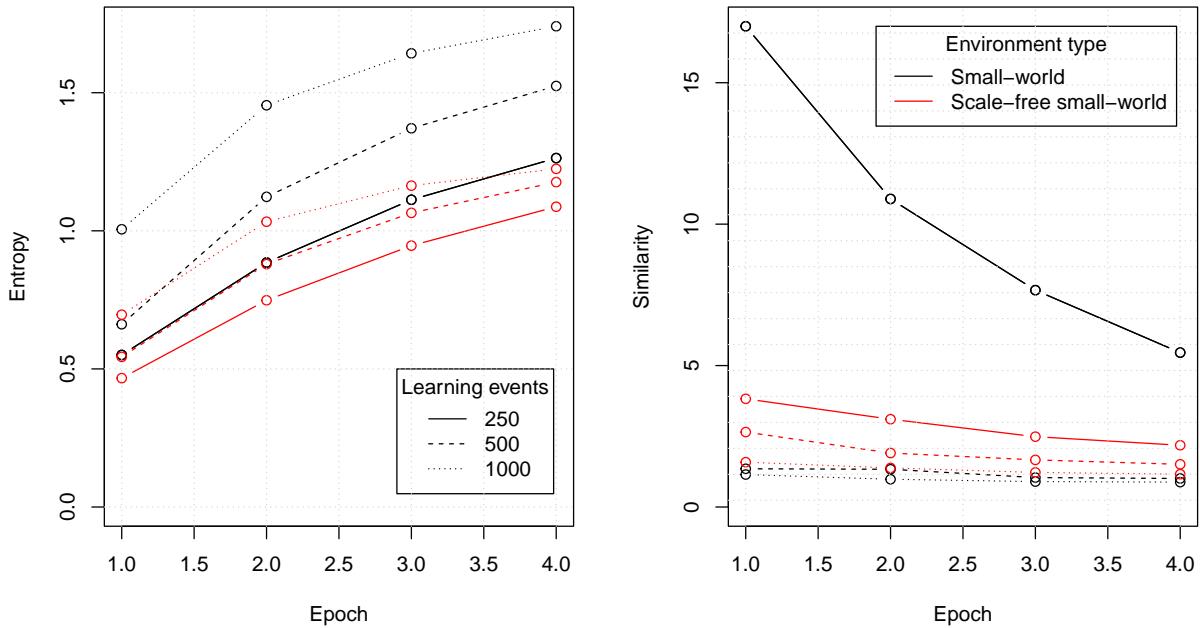


Figure 3: Entropy and similarity measures computed from the cognitive representation at different epochs of development for environmental networks with different numbers of learning events per epoch.

4 Learning rates

The learning values in the main text vary from $\beta = 0.01$ to 0.10 . Figure 4 shows the results for values ranging from $\beta = 0.01$ to 1.0 . This uses the scale-free small world environment as above, with $M = 5$ and $\gamma = 1$, and 250 learning events per epoch.

The results show that entropy always rises and similarity always falls with increasing learning. However, we can also see evidence of an inverted-U function, with the slowest rates of rising entropy and falling similarity occurring when learning is either very slow or very fast. Intermediate values of learning (near $.5$) have the highest entropy and lowest similarity overall.

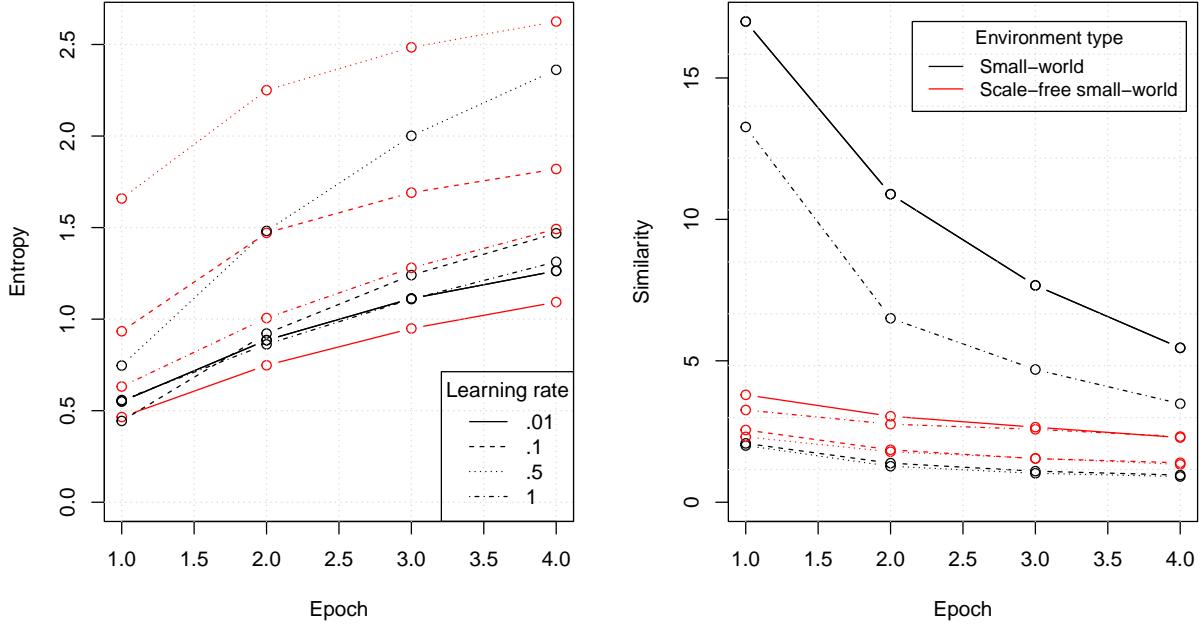


Figure 4: Varying learning rates. Entropy and similarity measures computed from the cognitive representation at different epochs of development for a variety of β values.

5 Rescorla-Wagner context cue

The Rescorla-Wagner model contains a context cue that is present during all associative learning events (Ramscar et al., 2017; Rescorla & Wagner, 1972). This means that all learning events contain the context cue in addition to the associative cue and target. This is incorporated into the model as a compound cue (see the online code). This has been criticized as sometimes creating odd behavior because cues can become positively associated with a target even if they never occur together as a result of the context cue carrying residual positive and negative predictive values for various cues (Miller, Barnet, & Grahame, 1995). Removing this context cue, however, does not alter the pattern of results provided in the main manuscript. Figure 5 shows the results for entropy and similarity with the context cue removed. This uses the scale-free small world environment as above, with $M = 5$ and $\gamma = 1$, and 250 learning events per epoch. The results show close overlap between learning with and without the context cue.

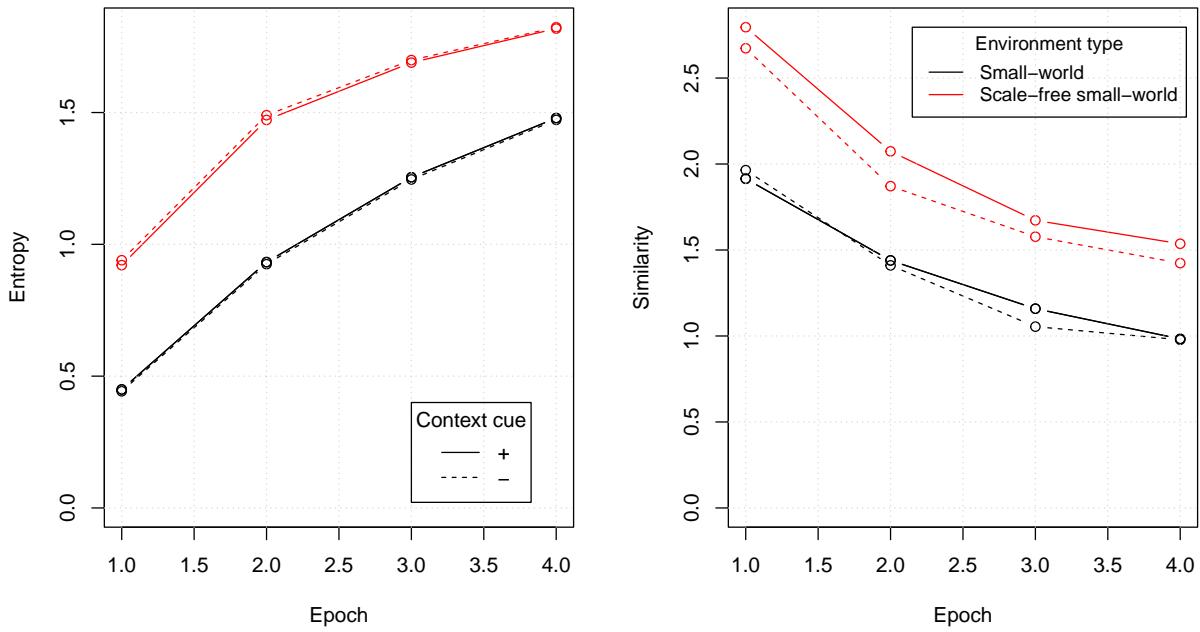


Figure 5: Removal of the context cue. The results for entropy and similarity computed using Rescorla Wagner with the context cue present or removed.

6 Directed versus undirected representations

The main text uses an undirected representation network. This occurs because the learning algorithm always assigns the same item from the pair as the cue, and treats the edge weight as undirected. Another approach is to assign the cue and target at random and produce a directed network. Here I show that the directed and undirected representation networks produce the same qualitative pattern. This uses the scale-free small world environment as above, with $M = 5$ and $\gamma = 1$, and 250 learning events per epoch, with $\beta = .1$.

Figure 6 shows that the difference between undirected and directed networks is similar to the pattern for increasing the network size or altering the number of associations. A directed network requires more learning to encode edges, which reduces the rate at which entropy rises early in development. Undirected edges lead to faster declines in similarity as they lead activation to diffuse away in the network more quickly than a corresponding directed network. Again, this appears to be a consequence of how quickly learning saturates the number of possible to-be-learned associative edges in the network, which will happen more quickly for undirected than directed networks.

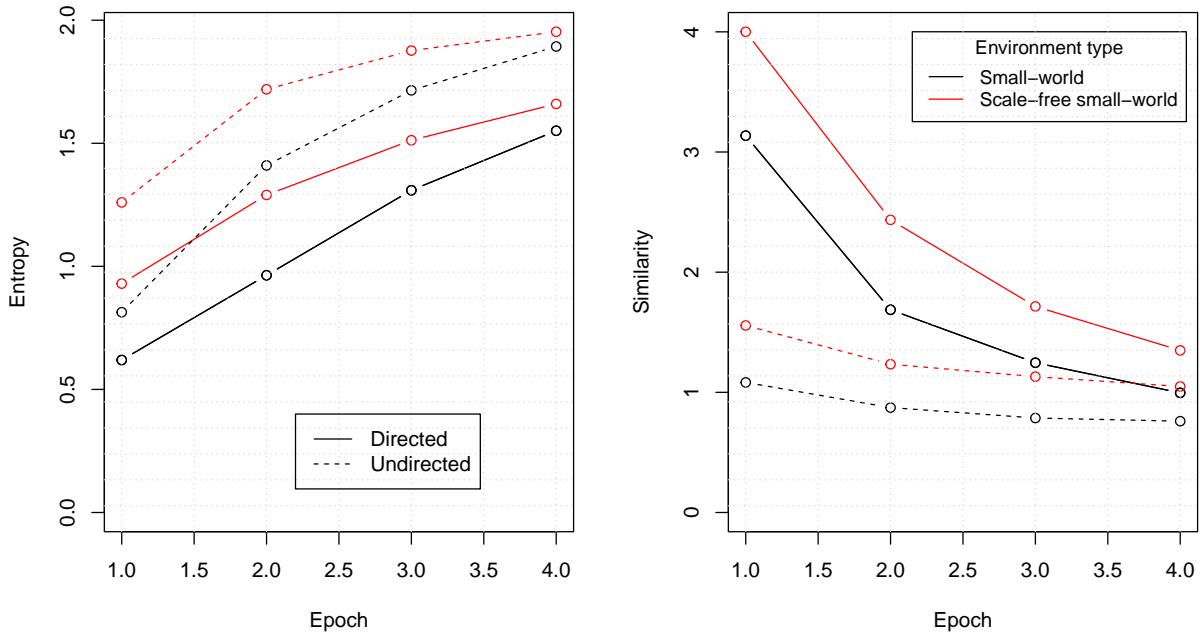


Figure 6: Directed and undirected representation networks. The results for entropy and similarity computed using Rescorla Wagner with the context cue present or removed.

7 Negative edges

Negative edges are produced in the original Rescorla-Wagner configuration when the context cue takes on positive predictive values with items it occurs with (in combination with other cues) and then imparts some of this positive value to other cues as negative predictive strengths. This happens because the positive context value predicts the occurrence of an outcome that does not appear (in the new context), and the new cue becomes negatively associated with the outcome it does not appear with. More intuitively, if AB predicts X and AC predicts Y (but not X) then C must negatively predict X to account for the prior observation that A predicts X. This is a standard outcome of prediction error models and more commonly associated with inhibition in classical conditioning.

As Rescorla (1988) points out, this learning of dis-associative information is required to account for the observation that base rates of the unconditioned stimulus matter to learning. We do not only learn on occasions when things co-occur. We also learn when things do not co-occur. Ramscar et al. (2017) use this feature to account for the observation that older adults primarily have difficulty learning new paired associates that they have previously learned over a lifetime of experience are not associated.

Negative association strengths are a common feature of prediction error models and the Rescorla-Wagner model in particular, so they are included in the simulation in the main text. However, the enrichment effects in the main article do not rely on negative edges. As demonstrated in Figure SM5, removing the context cue and the negative edges it produces does not affect the enrichment effects.

Nonetheless, it may be useful to consider the number of negative edges and their strengths. This information is provided in Table 1. This table is similar to the one shown in the main text but also includes the number of negative edges (Neg. edges) and the average node strength including negative edges (Strength w/ neg). There are many more negative edge weights than positive edge weights. This is consistent with the sparse environmental network in which most things predict the absence of other things. The negative edge weights are sufficiently small that including them in the computation of node strength has only a marginal influence on the age-related pattern in the results.

Table 1: Statistics for the environments and growing representations.

Type	Measures	Environment	Epoch			
			1	2	3	4
ER random	Nodes	499.84	305.15	415.16	459.33	478.76
	Edges	1984.16	234.82	449.25	941.92	1832.35
	Strength	8.003	0.010	0.020	0.029	0.039
	Degree	7.94	0.94	1.80	3.77	7.33
	ASPL	3.239	0.056	0.091	0.033	0.018
	C	0.016	0.002	0.004	0.019	0.043
	M	0.282	0.952	0.766	0.647	0.583
	Neg. edges	0.0	381.6	3067.2	8168.2	14322.5
	Strength w/ neg	8.003	0.010	0.019	0.028	0.037
	Nodes	442.23	172.12	248.13	295.28	327.32
Scale-free	Edges	1390.60	219.51	755.67	1570.71	2430.35
	Strength	9.042	0.009	0.016	0.022	0.028
	Degree	6.29	0.99	3.42	7.10	10.99
	ASPL	3.141	0.028	0.016	0.012	0.011
	C	0.382	0.158	0.182	0.201	0.214
	M	0.111	0.597	0.535	0.500	0.475
	Neg. edges	0.0	419.2	1321.8	2300.2	3234.2
	Strength w/ neg	9.042	0.009	0.015	0.021	0.026
	Nodes	499.81	305.43	415.35	459.63	478.88
	Edges	1941.61	234.30	447.05	936.21	1803.15
Small-world	Strength	8.003	0.010	0.020	0.029	0.039
	Degree	7.77	0.94	1.79	3.75	7.22
	ASPL	3.551	0.050	0.099	0.034	0.019
	C	0.049	0.006	0.012	0.028	0.051
	M	0.653	0.957	0.800	0.707	0.670
	Neg. edges	0.0	375.6	3026.5	8089.5	14138.0
	Strength w/ neg	8.003	0.010	0.019	0.028	0.037
	Nodes	445.21	176.64	253.54	301.05	333.10
	Edges	1193.40	199.44	666.05	1414.21	2225.90
	Strength	8.981	0.009	0.017	0.024	0.030
Scale-free small-world	Degree	5.36	0.90	2.99	6.35	9.99
	ASPL	3.612	0.038	0.019	0.014	0.012
	C	0.379	0.165	0.190	0.208	0.223
	M	0.611	0.744	0.717	0.706	0.700
	Neg. edges	0.0	369.9	1278.9	2327.7	3344.8
	Strength w/ neg	8.981	0.009	0.016	0.022	0.028

Note:

Measures are averaged over 1000 environments and cognitive representations learned from those environments over four epochs of 250 learning events each. Nodes indicates the total number of non-isolates in the environment or cognitive representation. Strength is the sum of the edge weights. Degree is the number of associations. ASPL is the average shortest path length. C is the mean local clustering coefficient. M is the modularity indicating community coherence. Neg. edges is the number of negative edges. Strength w/ neg is the average strength including negative edges.

8 Free association parameter exploration

The parameter exploration varies the number of participants, cues, and threshold values in the free association network construction. Each environment is constructed with 500 concept nodes and 2000 pairs sampled to produce edge weights. Learning takes place over 4 epochs of 250 learning events each, with $\beta = .1$. Each cell is the outcome of 100 simulations starting with the randomly generated environment.

8.1 Enrichment

Cells represent the difference between Epoch 4 and Epoch 1, with positive values indicating rising trends.

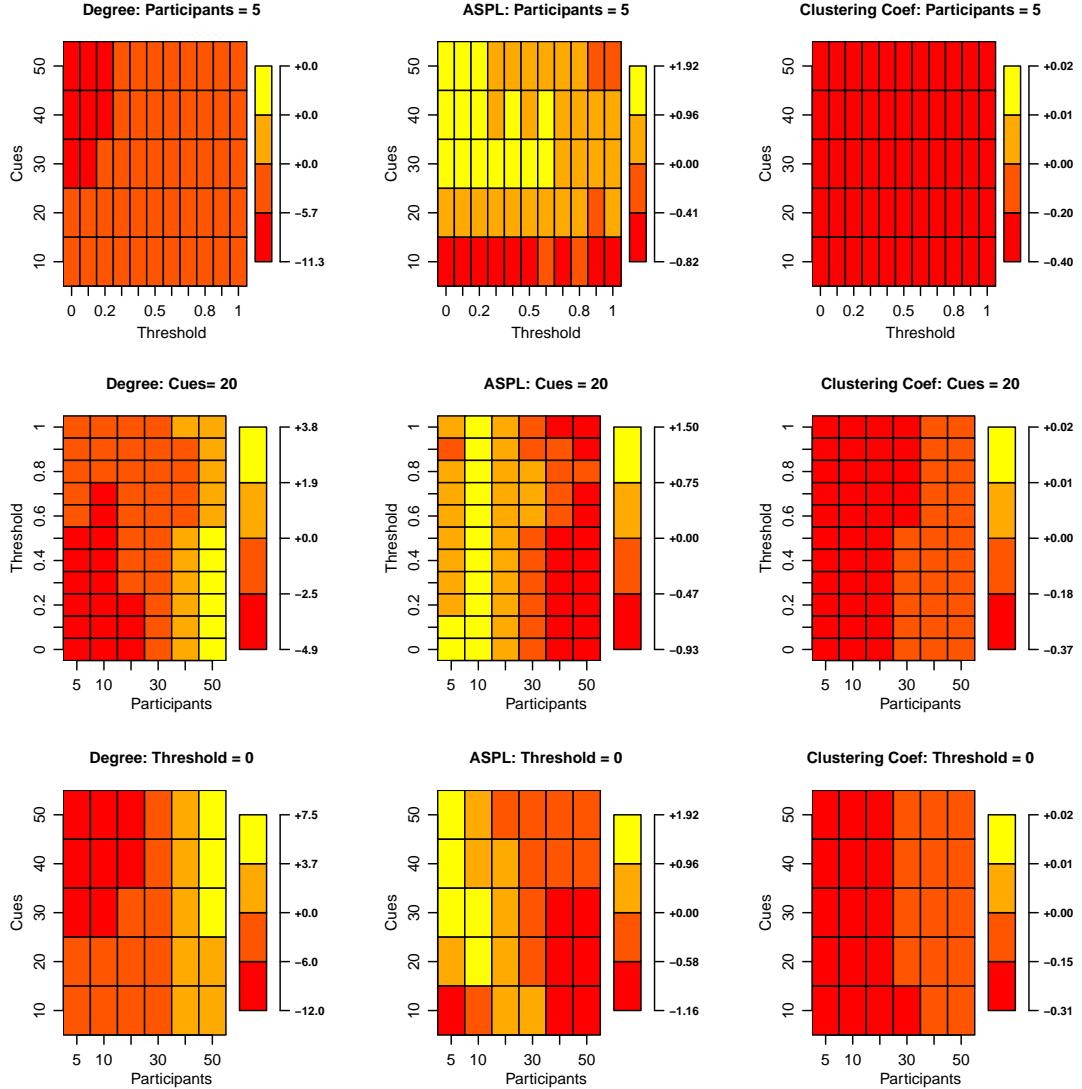


Figure 7: Weighted Erdös-Renyi random graph environment. Enrichment parameter exploration for free association network.

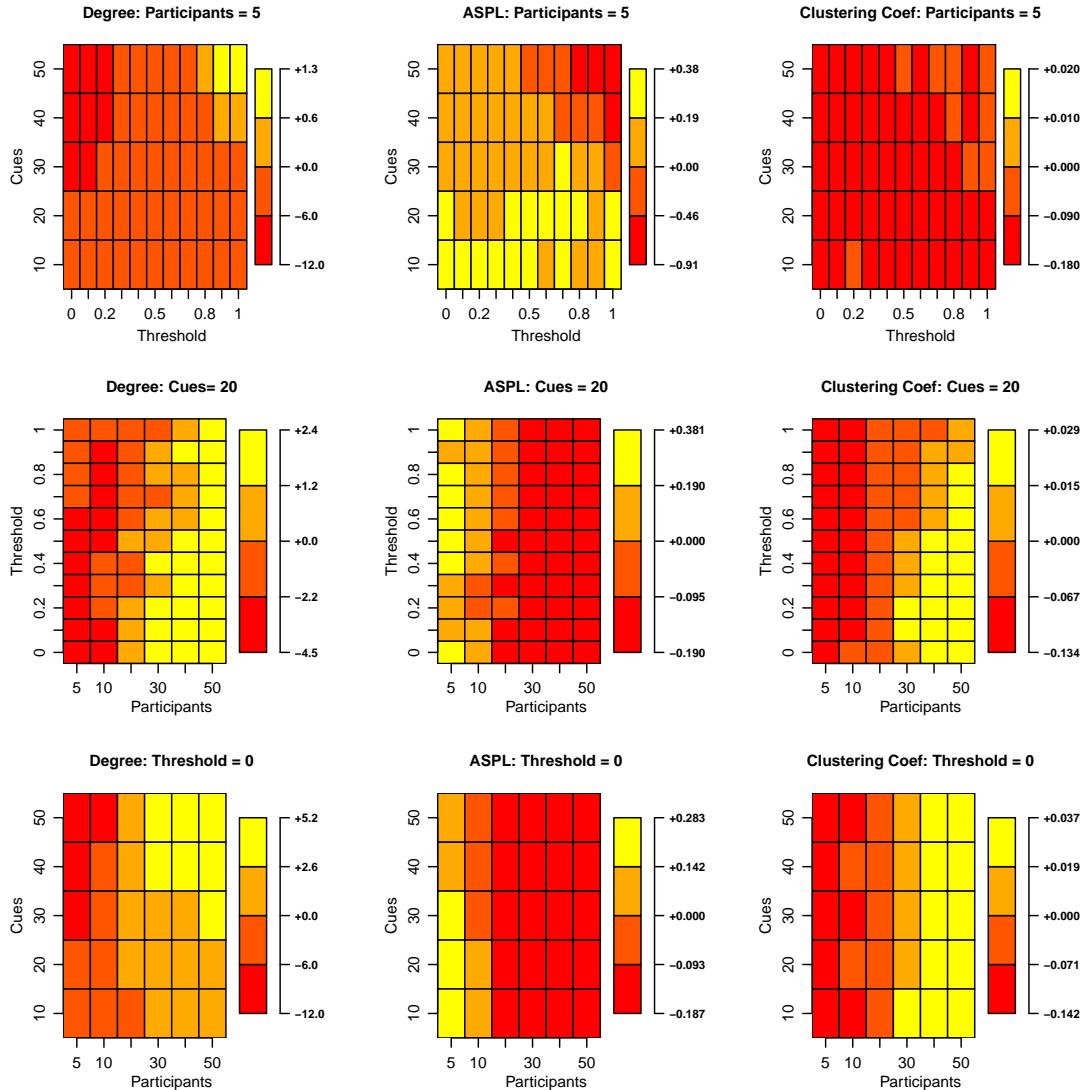


Figure 8: Weighted scale-free environment. Enrichment parameter exploration for free association network.

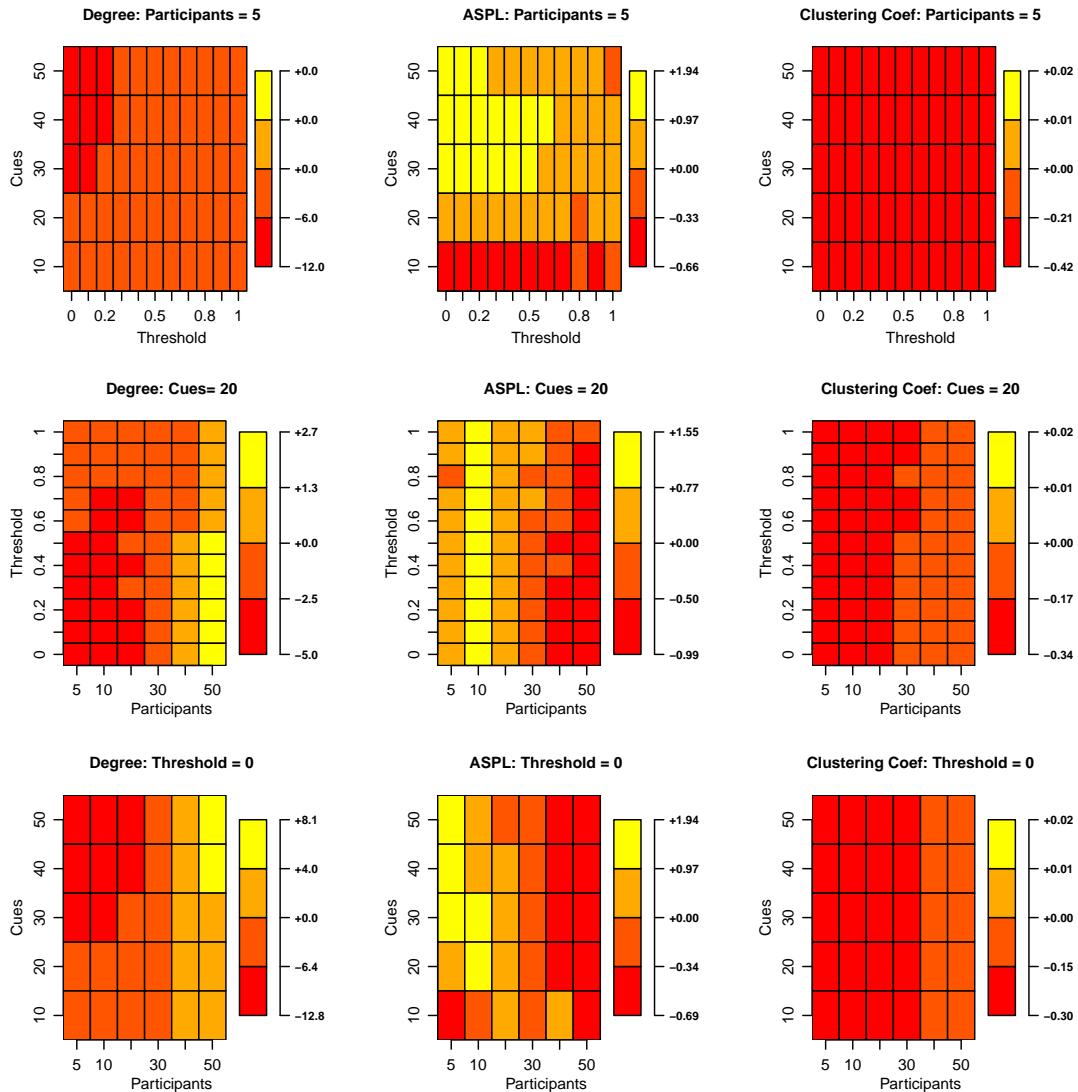


Figure 9: Weighted small-world environment. Enrichment parameter exploration for free association network.

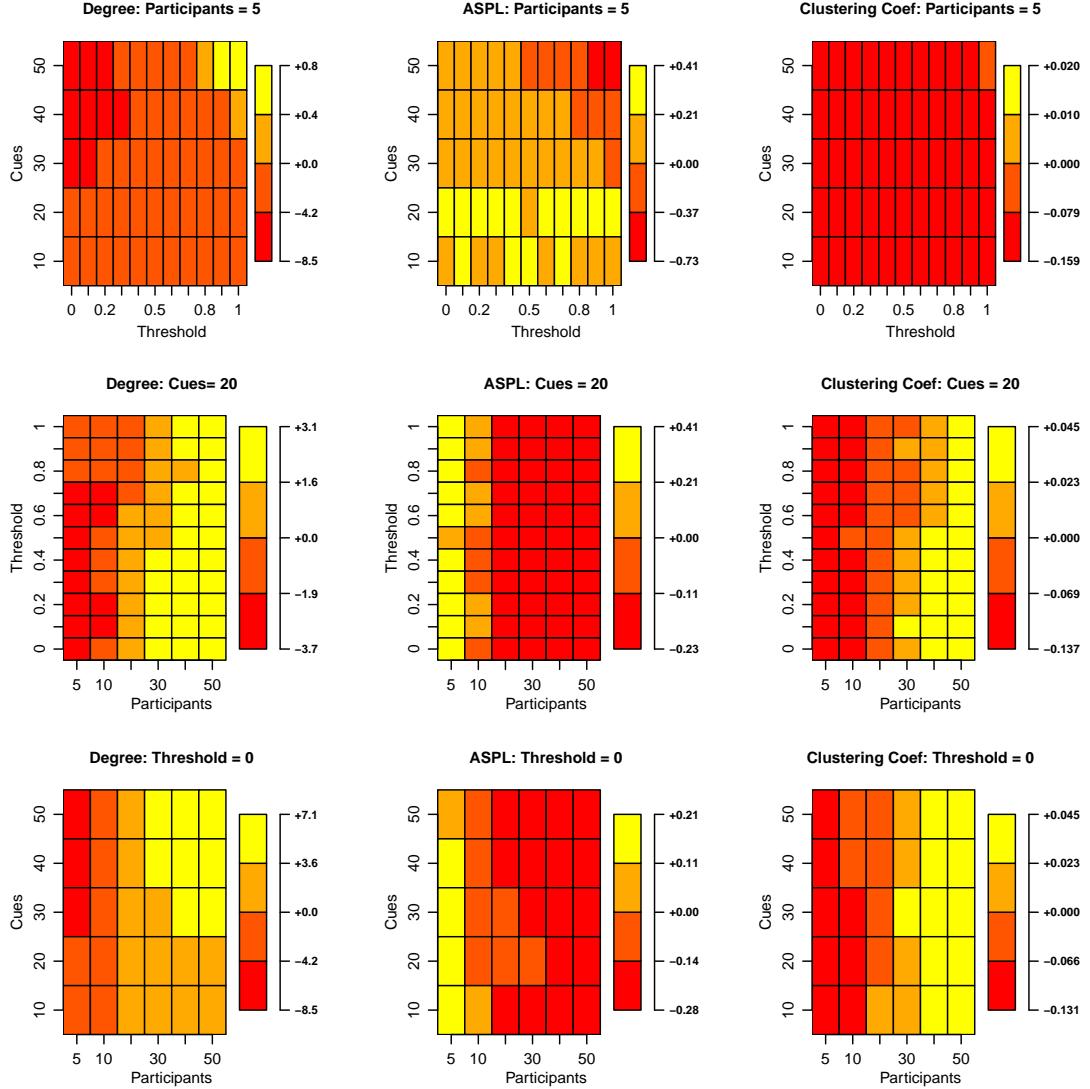


Figure 10: Weighted scale-free small-world environment. Enrichment parameter exploration for free association network.

8.2 Degradation

Cells represent the difference between 30% degradation and 0% degradation starting from the outcome of epoch 4. Positive values indicate rising trends. Networks are as for the enrichment parameter exploration. Each cell is the outcome of 100 simulations starting with the randomly generated environment.

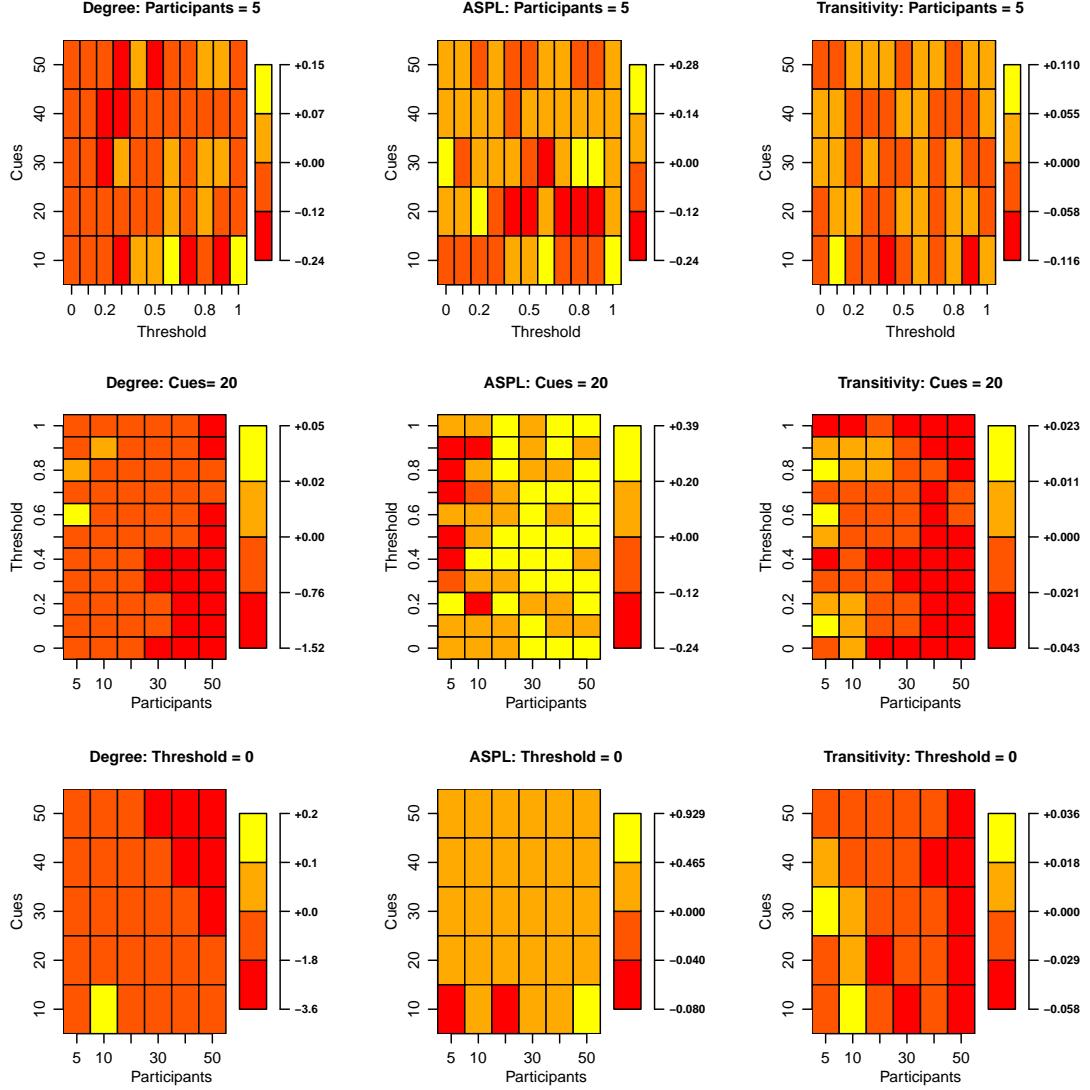


Figure 11: Degradation parameter exploration for free association network with the weighted Erdős-Renyi random graph environment.

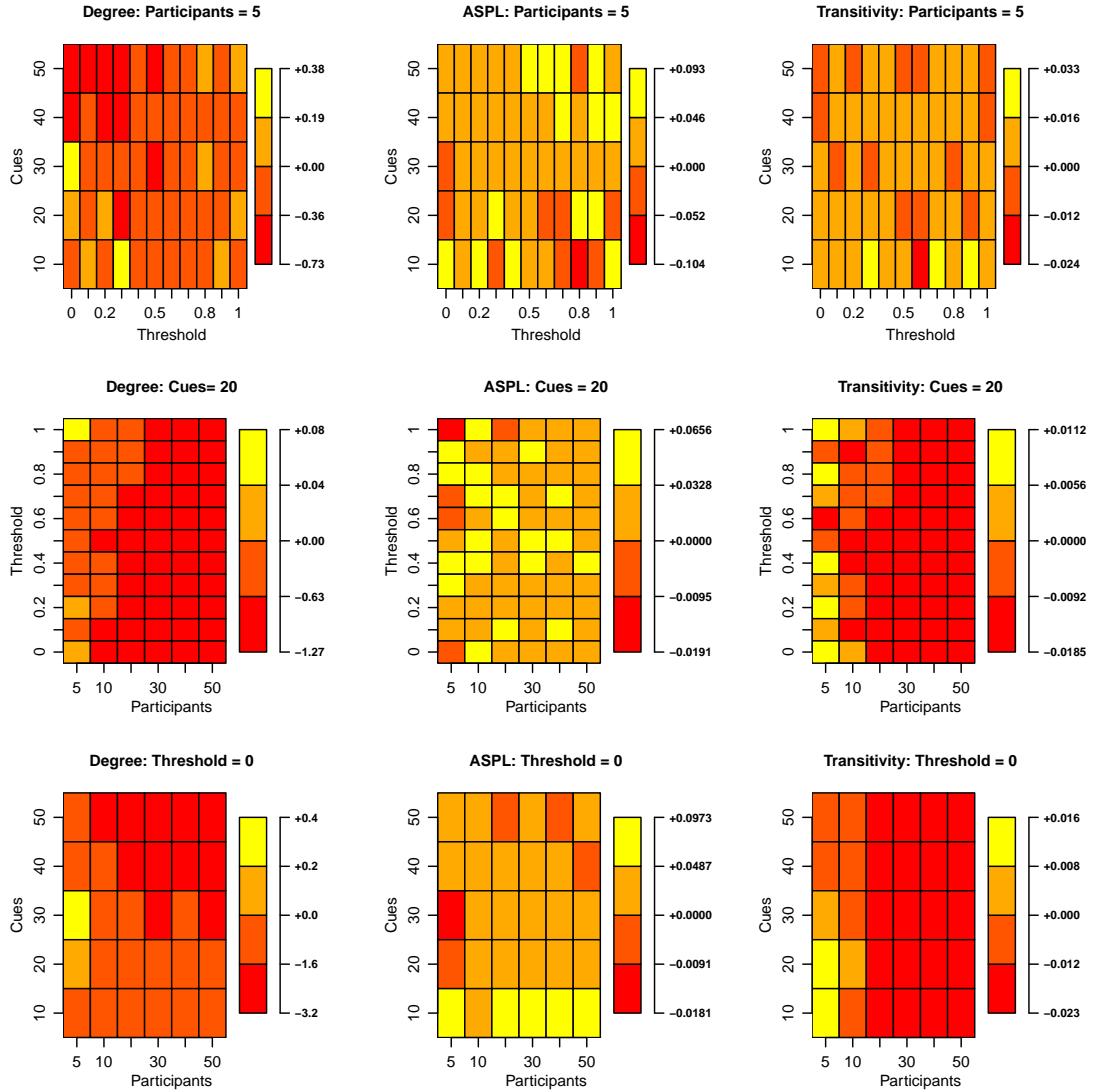


Figure 12: Degradation parameter exploration for free association network with the weighted scale-free environment.

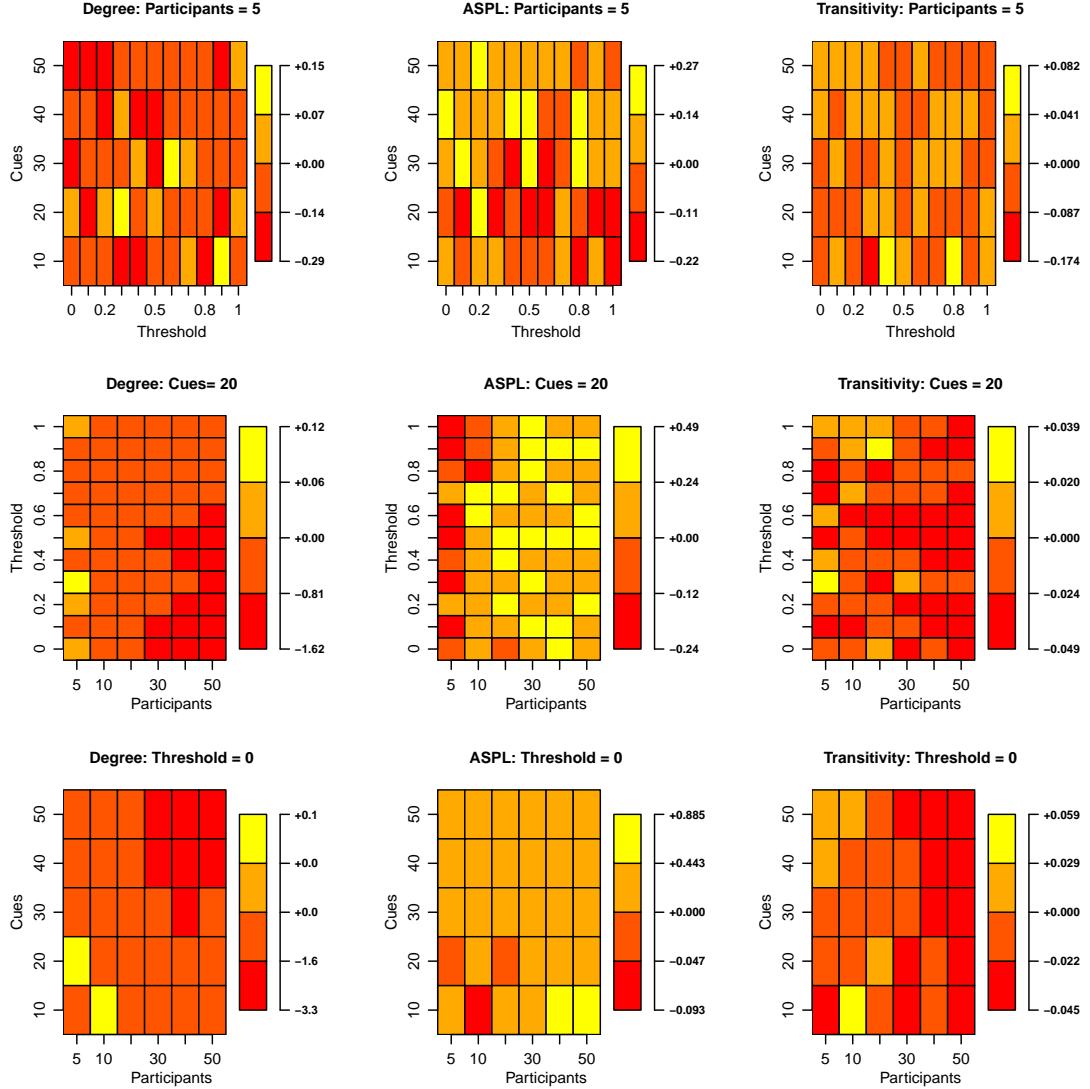


Figure 13: Degradation parameter exploration for free association network with the weighted small-world environment.

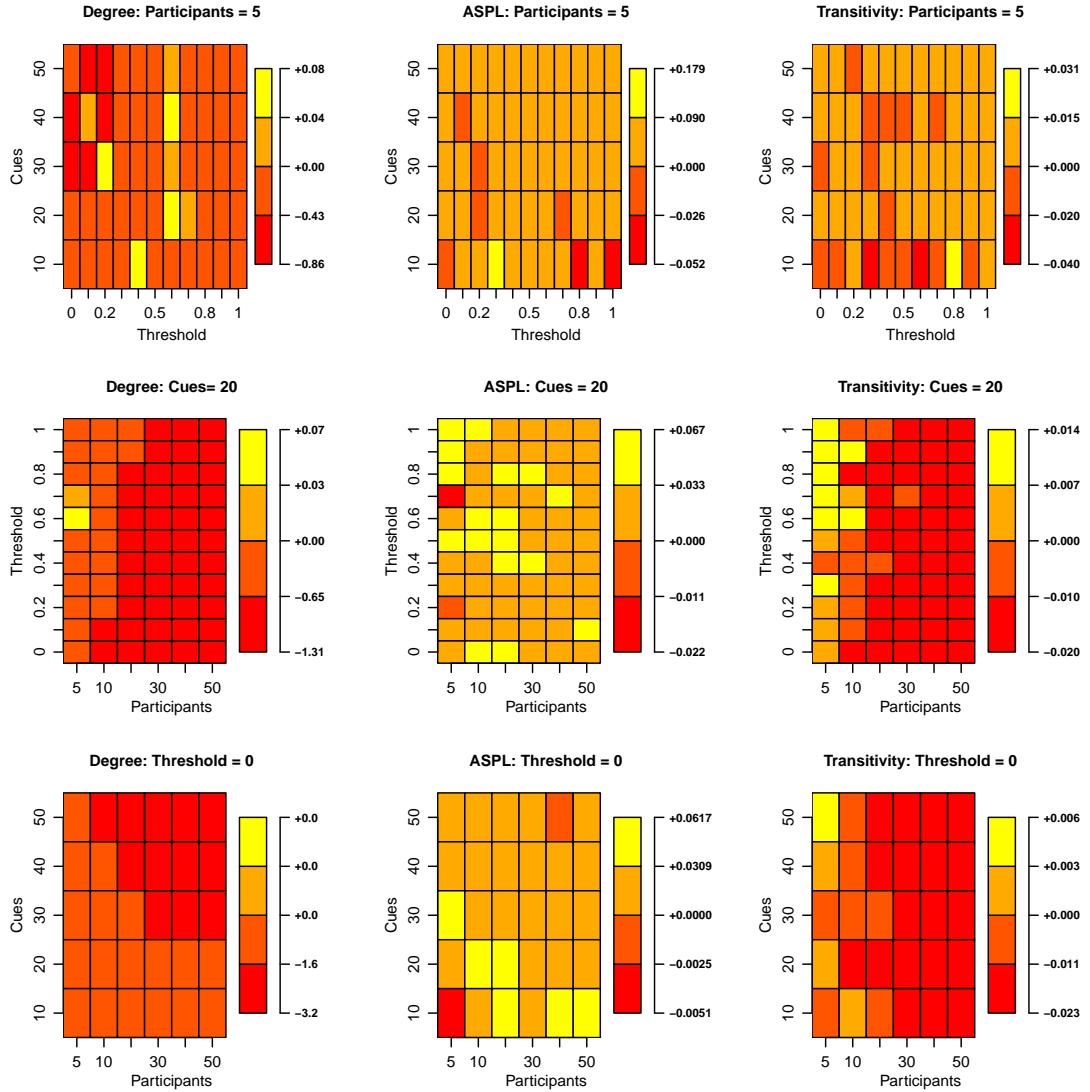


Figure 14: Degradation parameter exploration for free association network with the weighted scale-free small-world environment.

References

Rescorla, Robert A. 1988. "Pavlovian Conditioning: It's Not What You Think It Is." *American Psychologist* 43 (3): 151–60.