

Problem Sheet 4

Learning on Topological Data:
Over-squashing and Over-smoothing

Topology and Geometry in Machine Learning

Lecture 4 — Solutions

Conventions. Throughout, $G = (V, E)$ denotes a finite, simple, undirected and connected graph, with at least two vertices, until self-loops are explicitly added. Let

$$A = (A_{uv}), \quad d_v = \sum_u A_{vu}, \quad D = \text{diag}(d_v), \quad L = D - A, \quad N = D^{-1/2} L D^{-1/2}.$$

Each sum over $\{u, v\} \in E$ counts an undirected edge once. For any node signal $f \in \mathbb{R}^{|V|}$, we denote its *Dirichlet Energy* as,

$$\mathcal{E}(f) := f^\top L f = \sum_{\{u, v\} \in E} (f_u - f_v)^2.$$

Remark. The exercises are marked in increasing difficulty by *, **, ***.

Exercise 1 (Normalized Laplacians) Add one unit self-loop at every vertex of $G = (V, E)$. This amounts to modifying the corresponding matrices as:

$$\tilde{A} = A + I, \quad \tilde{D} = D + I, \quad \tilde{L} = \tilde{D} - \tilde{A}, \quad \tilde{N} = \tilde{D}^{-1/2} \tilde{L} \tilde{D}^{-1/2}.$$

- (a, *) Show that $\tilde{L} = L$, but in general $\tilde{N} \neq N$. Prove that N and $\tilde{N}$ are symmetric. By substituting $f = D^{-1/2} z$, and then $f = \tilde{D}^{-1/2} z$, prove that both matrices are positive semidefinite.
- (b, **) Use $(a - b)^2 \leq 2(a^2 + b^2)$ to show

$$0 \leq z^\top N z \leq 2 \|z\|_2^2.$$

Deduce that every eigenvalue ν of N belongs to $[0, 2]$, i.e., $0 \leq \nu \leq 2$.

- (c, ***) Let $d_{\max} = \max_v d_v$. Prove the stronger augmented bound

$$0 \leq z^\top \tilde{N} z \leq \frac{2d_{\max}}{d_{\max} + 1} \|z\|_2^2.$$

Conclude that all eigenvalues $\tilde{\nu}$ of $\tilde{N}$ are strictly below 2, i.e., $0 \leq \tilde{\nu} < 2$.

- (d, *) Identify the kernels of N and $\tilde{N}$, using connectedness. Normalize their zero eigenvectors to unit length. Deduce that $P = I - \tilde{N}$ has a simple eigenvalue 1 and all other eigenvalues in $(-1, 1)$.
- (e, **) For the graph consisting of a single edge, compute N , $\tilde{N}$ and their spectra. Starting from $x = (1, 0)^\top$, compare $(I - N)^t x$ and $P^t x$. What obstruction to convergence have the self-loops removed?

Solution.

- (a) The identity terms cancel, so $\tilde{L} = L$.

Both normalized matrices have the form BLB , with B diagonal. Since L and B are symmetric, $(BLB)^\top = B^\top L^\top B^\top = BLB$.

Their quadratic forms are $\mathcal{E}(D^{-1/2}z)$ and $\mathcal{E}(\tilde{D}^{-1/2}z)$, respectively, and hence are nonnegative:

$$z^\top Nz = z^\top D^{-1/2}LD^{-1/2}z = (D^{-1/2}z)^\top LD^{-1/2}z = f^\top Lf \geq 0, \quad \forall f = D^{-1/2}z \in \mathbb{R}^n$$

And the same applies to $\tilde{N}$.

(b) Applying the definition of the normalized Laplacian:

$$z^\top Nz = z^\top D^{-1/2}LD^{-1/2}z = \mathcal{E}(D^{-1/2}z)$$

Therefore, for $f = D^{-1/2}z$,

$$\begin{aligned} \mathcal{E}(f) &= \sum_{\{u,v\} \in E} (f_u - f_v)^2 \\ &\leq 2 \sum_{\{u,v\} \in E} (f_u^2 + f_v^2) \\ &= \sum_{u \in V} \sum_{v \in \mathcal{N}(u)} (f_u^2 + f_v^2) \\ &= \sum_{u \in V} \sum_{v \in \mathcal{N}(u)} f_u^2 + \sum_{u \in V} \sum_{v \in \mathcal{N}(u)} f_v^2 \\ &= 2 \sum_{v \in V} d_v f_v^2 = 2\|z\|_2^2 \end{aligned}$$

where we use that $f_u = \frac{z_u}{\sqrt{d_u}}$ and that summing over edges is the same as (half) summing over vertices and then neighbours.

For an eigenvector $z \neq 0$ of N , $Nz = \nu z$, we obtain:

$$0 \leq z^\top Nz = \nu \|z\|_2^2 \leq 2\|z\|_2^2$$

and in consequence, $0 \leq \lambda \leq 2$, as desired.

(c) Now set $\tilde{f} = \tilde{D}^{-1/2}z$. The first steps of the previous part still apply, and the energy still sums only over the original edges, since a self-loop has zero difference:

$$\tilde{f}_u^2 - \tilde{f}_u^2 = 0.$$

Thus

$$z^\top \tilde{N}z = \mathcal{E}(\tilde{f}) \leq 2 \sum_v d_v \tilde{f}_v^2 = 2 \sum_v \frac{d_v}{d_v + 1} z_v^2 \leq \frac{2d_{\max}}{d_{\max} + 1} \|z\|_2^2 < 2\|z\|_2^2$$

for $z \neq 0$. The third equality comes from the fact that the function

$$g(x) = \frac{x}{x+1}$$

is monotonically increasing, so since $d_v \leq d_{\max} = \max\{d_v : v \in V\}$, we get

$$\frac{d_v}{d_v + 1} \leq \frac{d_{\max}}{d_{\max} + 1}, \quad \forall v \in V.$$

For the monotonicity of the function, notice that:

$$g'(x) = \frac{x+1-x}{(x+1)^2} > 0.$$

The eigenvector argument gives the spectral bound again. If $\tilde{z}$ is an eigenvector of $\tilde{N}$, we have:

$$0 \leq z^\top \tilde{N} z = \tilde{\nu} \|z\|_2^2 < 2 \|z\|_2^2$$

which allows to conclude that $\tilde{\nu} \in [0, 2)$.

(d) The energy $\mathcal{E}(f)$ vanishes exactly when $f_u = f_v$ along every edge. Connectedness implies that f must be constant throughout the whole graph.

Hence, from the relations above between the Dirichlet energy and the normalized Laplacians used as bilinear forms, we can conclude

$$\ker N = \text{span}\{D^{1/2}\mathbf{1}\}, \quad \ker \tilde{N} = \text{span}\{\tilde{D}^{1/2}\mathbf{1}\}.$$

The respective unit vectors are

$$p = \frac{D^{1/2}\mathbf{1}}{\sqrt{\sum_v d_v}}, \quad q = \frac{\tilde{D}^{1/2}\mathbf{1}}{\sqrt{\sum_v (d_v + 1)}}.$$

Write $0 = \tilde{\nu}_1 < \tilde{\nu}_2 \leq \dots \leq \tilde{\nu}_n < 2$. The eigenvalues of P are $\mu_i = 1 - \tilde{\nu}_i$, so $\mu_1 = 1$ is simple and $|\mu_i| < 1$ for $i \geq 2$.

(e) Here $D = I$ and

$$N = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad \tilde{N} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Their spectra are $\{0, 2\}$ and $\{0, 1\}$. Without loops, $I - N$ swaps coordinates, giving $(1, 0)^\top$ for even t and $(0, 1)^\top$ for odd t . With loops, P has every entry $1/2$ and $P^t x = (1/2, 1/2)^\top$ for all $t \geq 1$. The eigenvalue -1 of the original propagation caused the oscillation.

Exercise 2 (Conductance, graph partitions and Cheeger inequalities) For a nonempty proper subset $S \subset V$, recall the definitions

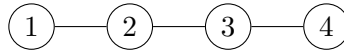
$$\partial S = \{\{u, v\} \in E : u \in S, v \notin S\}, \quad \text{vol}(S) = \sum_{v \in S} d_v, \quad \phi(S) = \frac{|\partial S|}{\min\{\text{vol}(S), \text{vol}(V \setminus S)\}}.$$

Set $\phi_G = \min_S \phi(S)$. This exercise uses the original graph without self-loops, and ν_2 denotes the second-smallest eigenvalue of N .

Linear algebra tool. Changing variables $z = D^{1/2}f$ in the Rayleigh principle gives

$$\nu_2 = \min_{\substack{f \neq 0 \\ f^\top D \mathbf{1} = 0}} \frac{\mathcal{E}(f)}{f^\top D f}.$$

The orthogonality condition here is $\sum_v d_v f_v = 0$; for z it is $z \perp D^{1/2} \mathbf{1}$. Let G be the path on four vertices $1 - 2 - 3 - 4$:



(a, *) **Compute cuts.** Compute $\phi(S)$ for $S = \{1\}, \{2\}, \{1, 2\}, \{1, 3\}$ and $\{1, 4\}$. Use complementary sets and symmetry to determine ϕ_G . Why does minimizing the number of crossing edges alone not identify the best cut?

(b, *) **Run a spectral sweep.** You may use

$$\nu_2 = \frac{1}{2}, \quad Nu = \frac{1}{2}u, \quad u = (-\sqrt{2}, -1, 1, \sqrt{2})^\top.$$

Compute the scores $f = D^{-1/2}u$. Sort vertices by increasing f_v and evaluate the conductance of each nonempty proper prefix. Let γ be the smallest value. Check numerically that

$$\frac{\nu_2}{2} \leq \phi_G \leq \gamma \leq \sqrt{2\nu_2}.$$

(c, ***) **Prove the conductance lemma from the lecture.** For a general cut put $a = \text{vol}(S)$, $b = \text{vol}(V \setminus S)$ and $y = \mathbf{1}_S - \frac{a}{a+b} \mathbf{1}$. Show that

$$y^\top D \mathbf{1} = 0, \quad \mathcal{E}(y) = |\partial S|, \quad y^\top D y = \frac{ab}{a+b}.$$

Deduce the lemma and then the first Cheeger bound:

$$\nu_2 \leq \frac{\text{vol}(V) |\partial S|}{\text{vol}(S) \text{vol}(V \setminus S)}, \quad \frac{\nu_2}{2} \leq \phi_G.$$

Solution.

(a) The degrees are $(1, 2, 2, 1)$. The requested values are

S	$\{1\}$	$\{2\}$	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$
$ \partial S $	1	2	1	3	2
$\min(\text{vol}(S), \text{vol}(S^c))$	1	2	3	3	2
$\phi(S)$	1	1	1/3	1	1

Every singleton and its complementary triple has conductance one. The three pair partitions are represented in the table, so $\phi_G = 1/3$. The cuts $\{1\}$ and $\{1, 2\}$ both cross one edge, but the second separates two larger volumes, i.e., to compute the conductance, you need to take into account both

the edges between the sets in the partition and the volumes of the sets in the partition.

(b) The scores are $(-\sqrt{2}, -1/\sqrt{2}, 1/\sqrt{2}, \sqrt{2})^\top$, giving order 1, 2, 3, 4. The three prefix conductances are 1, 1/3, 1, hence $\gamma = 1/3$ and the chain is $1/4 \leq 1/3 \leq 1/3 \leq 1$.

(c) The two values of y are $b/(a+b)$ on S and $-a/(a+b)$ outside. Thus

$$y^\top D\mathbf{1} = \frac{ab - ba}{a+b} = 0, \quad y^\top Dy = \frac{ab^2 + ba^2}{(a+b)^2} = \frac{ab}{a+b}.$$

Each crossing edge has difference one and all other edges have difference zero, giving $\mathcal{E}(y) = |\partial S|$. The Rayleigh principle yields

$$\nu_2 \leq |\partial S| \frac{a+b}{ab}$$

proving the first of the bounds after minimizing over S .

The second one is obtained as follows: if $a \leq b$, the right-hand side is $(1 + a/b)\phi(S) \leq 2\phi(S)$; the case $b \leq a$ is symmetric, minimizing over S yields the second bound.

Exercise 3 (Rewiring a bottleneck) Let G consist of two four-cycles

$$a_0 - a_1 - a_2 - a_3 - a_0, \quad b_0 - b_1 - b_2 - b_3 - b_0,$$

joined by the single edge $\{a_0, b_0\}$. Write $S = \{a_0, a_1, a_2, a_3\}$. You have a budget of two extra edges. Compare:

- **Across-community rewiring:** add $\{a_1, b_1\}$ and $\{a_3, b_3\}$, obtaining G_{cross} .
- **Within-community rewiring:** add $\{a_1, a_3\}$ and $\{b_1, b_3\}$, obtaining G_{inside} .

Compute conductances before adding self-loops.

- (a, *) Draw the three graphs. For each, compute $|\partial S|$, $\text{vol}(S)$, $\text{vol}(S^c)$ and $\phi(S)$. Which use of the edge budget improves this bottleneck? Explain why extra edges can also *decrease* the conductance of a fixed cut.
- (b, **) Compute the distance from a_2 to b_2 in each graph. What is the minimum number of local message-passing layers needed for a_2 to possibly depend on x_{b_2} ? Explain why shorter distance alone does not guarantee good predictions.
- (c, ***) For each graph form its own $P = (D + I)^{-1/2}(A + I)(D + I)^{-1/2}$ and propagate scalar features by $h^{(t)} = P^t x$. Show that

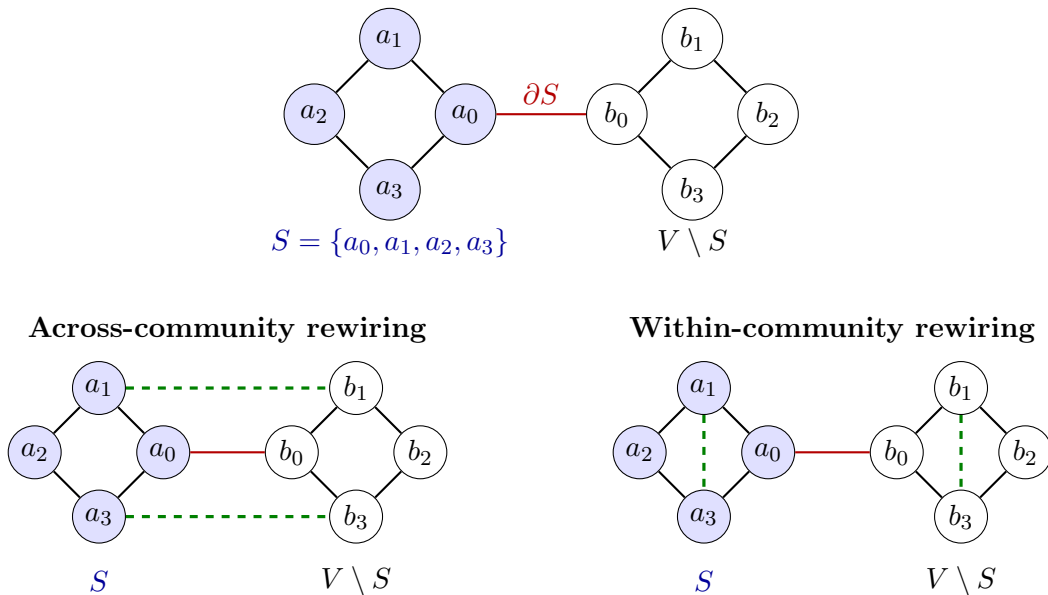
$$\frac{\partial h_{a_2}^{(t)}}{\partial x_{b_2}} = (P^t)_{a_2 b_2}.$$

At $t = 3$, prove that this derivative is zero for G and G_{inside} . For G_{cross} , enumerate all length-three walks from b_2 to a_2 and compute the derivative exactly. *Hint:* a walk contributes the product of its three propagation entries; self-loops cannot help a length-three walk if the distance is already three.

- (d, ***) Does improving $\phi(S)$ prove that the global conductance ϕ_G increases? Does adding edges increase every propagation entry? Compare $P_{a_1 a_2}$ in G and G_{cross} . Explain the possible tension between improved communication, normalization and over-smoothing.

Solution.

(a)



In each cycle, internal edges contribute eight to its volume. The original bridge adds one per side. Each new cross edge adds one per side; each internal edge adds two on its own side. Thus

Graph	$ \partial S $	$\text{vol}(S)$	$\text{vol}(S^c)$	$\phi(S)$
G	1	9	9	$1/9$
G_{cross}	3	11	11	$3/11$
G_{inside}	1	11	11	$1/11$

Cross edges enlarge the boundary and improve this cut's conductance. Internal edges increase its denominator without enlarging its boundary, so its conductance decreases.

(b) The original distance is five: two steps from a_2 to a_0 , one bridge step, and two steps from b_0 to b_2 . The internal diagonals do not shorten either two-step segment, so the distance remains five. The cross edges allow $a_2 - a_1 - b_1 - b_2$, of length three. There is no length-one or length-two route, so the new distance is three. These distances are necessary layer counts for possible dependence. Actual influence and prediction quality also depend on aggregation, feature dimension, learned parameters and the target task.

(c) Differentiating the linear map gives its corresponding matrix entry.

Matrix multiplication expresses this entry as a sum of products along length- t walks, with loops allowing stationary steps. Indeed, for two layers we have:

$$(P^2)_{uv} = \sum_w P_{uw} P_{wv}$$

and generally, for layer t ,

$$(P^t)_{uv} = \sum_{w_1, \dots, w_{t-1}} P_{uw_1} P_{w_1 w_2} \dots P_{w_{t-1} w}.$$

Hence, to have non-zero contribution in the sum above, there must exist a walk of length t from u to v .

No three-step walk reaches a vertex at distance five, so the first two graphs give zero. In G_{cross} the only length-three walks are

$$b_2 - b_1 - a_1 - a_2, \quad b_2 - b_3 - a_3 - a_2.$$

The endpoint augmented degrees are 3, and each intermediate augmented degree is 4. Each walk contributes

$$\frac{1}{\sqrt{3 \cdot 4}} \frac{1}{\sqrt{4 \cdot 4}} \frac{1}{\sqrt{4 \cdot 3}} = \frac{1}{48},$$

so the derivative is $1/24$.

(d) A single cut value only gives an upper bound on the minimum over all cuts; one must check other cuts or prove an additional bound to compare global conductances. For the existing edge,

$$P_{a_1 a_2}(G) = \frac{1}{3}, \quad P_{a_1 a_2}(G_{\text{cross}}) = \frac{1}{\sqrt{12}} < \frac{1}{3}.$$

New edges create routes, but increased degrees can reduce the weights of old routes. Improved cross-community communication can also mix features more quickly, potentially accelerating over-smoothing. Neither a better cut value nor a shorter distance alone proves a particular convergence rate or an improvement on every learning task.

Exercise 4 (Exact GCN limits and repeated input injection) Use the notations of Exercise 1 and additionally define

$$P = I - \tilde{N}, \quad Q = \sum_v (d_v + 1), \quad q = \frac{\tilde{D}^{1/2} \mathbf{1}}{\sqrt{Q}}.$$

We analyze the linear GCN from the slides: identity activation and $W^{(t)} = I$. Initially take scalar features $h^{(0)} = x \in \mathbb{R}^n$. Let $\tilde{N}\Phi_i = \tilde{\nu}_i\Phi_i$ be an orthonormal eigenbasis with $\Phi_1 = q$, and write $a_i = \Phi_i^\top x$, $\mu_i = 1 - \tilde{\nu}_i$.

- (a, *) **The usual propagation: exact nodewise limit.** For $h^{(t+1)} = Ph^{(t)}$, derive the finite-depth eigenbasis expression and prove

$$h^{(\infty)} = qq^\top x, \quad h_v^{(\infty)} = \frac{\sqrt{d_v + 1}}{Q} \sum_u \sqrt{d_u + 1} x_u.$$

Are these values equal at every node? What rescaling makes them equal? For the path $1-2-3$ with $x = (1, 0, 0)^\top$, compute the three limiting values.

- (b, **) **Rate of collapse.** Let $\rho = \max_{i \geq 2} |\mu_i|$, from Exercise 1 we know that $\rho < 1$. Prove the following inequality:

$$\|h^{(t)} - qq^\top x\|_2 \leq \rho^t \|x - qq^\top x\|_2.$$

- (c, ***) **Injecting the initial features.** For $0 < \alpha < 1$ and $\beta = 1 - \alpha$, consider

$$h^{(t+1)} = \beta Ph^{(t)} + \alpha x, \quad h^{(0)} = x.$$

Write the update at a single node and expand the first two iterations.

Then prove, by induction, that

$$h^{(t)} = (\beta P)^t x + \alpha \sum_{k=0}^{t-1} (\beta P)^k x.$$

Derive the coefficient of each Φ_i at depth t , summing the geometric series explicitly.

Solution.

- (a) Since $P\Phi_i = \mu_i\Phi_i$,

$$h^{(t)} = \sum_{i=1}^n a_i \mu_i^t \Phi_i.$$

Exercise 1 gives $\mu_1 = 1$ and $|\mu_i| < 1$ for $i \geq 2$, so only $a_1\Phi_1 = qq^\top x$ remains.

Expanding the entries gives the requested nodewise formula.

The values are proportional to $\sqrt{d_v + 1}$, and are not generally equal. However, the rescaled values $h_v^{(\infty)} / \sqrt{d_v + 1}$ are equal.

For the three-vertex path,

$$q = \frac{(\sqrt{2}, \sqrt{3}, \sqrt{2})^\top}{\sqrt{7}}, \quad h^{(\infty)} = \frac{1}{7}(2, \sqrt{6}, 2)^\top.$$

- (b) Observe that

$$h^{(t)} = qq^\top x + \sum_{i=2}^n a_i \mu_i^t \Phi_i.$$

Hence, orthonormality gives

$$\|h^{(t)} - qq^\top x\|_2^2 = \sum_{i=2}^n a_i^2 \mu_i^{2t} \leq \rho^{2t} \sum_{i=2}^n a_i^2.$$

Taking square roots yields the bound. Only the single scalar $q^\top x$ survives, encoded along the known vector q ; every component orthogonal to q disappears. With several features, this is a rank-one projection in the node dimension.

(c) At a node,

$$h_v^{(t+1)} = \beta \sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{h_u^{(t)}}{\sqrt{(d_v + 1)(d_u + 1)}} + \alpha x_v.$$

The first two steps are

$$h^{(1)} = \beta Px + \alpha x, \quad h^{(2)} = \beta^2 P^2 x + \alpha \beta Px + \alpha x.$$

For all $t \geq 0$, with the sum empty at $t = 0$, we now prove by induction that

$$h^{(t)} = (\beta P)^t x + \alpha \sum_{k=0}^{t-1} (\beta P)^k x.$$

The base case is immediate. Multiplying the formula by βP shifts the summation indices to $1, \dots, t$; adding αx restores the $k = 0$ term and proves the next case.

We now write the input and the layer- t features in the same orthonormal eigenbasis, to obtain explicit formulas for the coefficients. Let:

$$x = \sum_{i=1}^n a_i \Phi_i, \quad h^{(t)} = \sum_{i=1}^n c_i^{(t)} \Phi_i, \quad c_i^{(0)} = a_i.$$

Substituting these expansions into $h^{(t+1)} = \beta P h^{(t)} + \alpha x$ and using $P \Phi_i = \mu_i \Phi_i$ gives

$$\sum_i c_i^{(t+1)} \Phi_i = \sum_i (\beta \mu_i c_i^{(t)} + \alpha a_i) \Phi_i.$$

Comparing coefficients, each mode evolves independently according to

$$c_i^{(t+1)} = \beta \mu_i c_i^{(t)} + \alpha a_i.$$

The existing coefficient is multiplied by $\beta \mu_i$, and each layer adds a fresh contribution αa_i . Expanding the first two steps,

$$\begin{aligned} c_i^{(1)} &= a_i(\beta \mu_i) + \alpha a_i, \\ c_i^{(2)} &= a_i(\beta \mu_i)^2 + \alpha a_i(\beta \mu_i) + \alpha a_i. \end{aligned}$$

Continuing this expansion yields

$$c_i^{(t)} = a_i(\beta \mu_i)^t + \alpha a_i \sum_{k=0}^{t-1} (\beta \mu_i)^k.$$

The first term comes from the initial state; the sum collects the later injections, propagated for different numbers of layers. The finite geometric-series identity is

$$\sum_{k=0}^{t-1} r^k = \frac{1 - r^t}{1 - r}, \quad r \neq 1.$$

Here $r = \beta\mu_i$. Since $0 < \beta < 1$ and $|\mu_i| \leq 1$, we have $|\beta\mu_i| < 1$, so the denominator is nonzero. Therefore

$$c_i^{(t)} = a_i \left[(\beta\mu_i)^t + \alpha \frac{1 - (\beta\mu_i)^t}{1 - \beta\mu_i} \right],$$

which gives

$$h^{(t)} = \sum_{i=1}^n a_i \left[(\beta\mu_i)^t + \alpha \frac{1 - (\beta\mu_i)^t}{1 - \beta\mu_i} \right] \Phi_i.$$

For the zero-frequency mode: $\tilde{\nu}_1 = 0$, hence $\mu_1 = 1 - \tilde{\nu}_1 = 1$. Using $1 - \beta = \alpha$, its coefficient is exactly

$$\beta^t + \alpha \frac{1 - \beta^t}{1 - \beta} = \beta^t + \frac{\alpha}{\alpha} (1 - \beta^t) = 1.$$

Thus $c_1^{(t)} = a_1$ at every layer. The initial-state contribution decays as β^t , but the accumulated injections supply exactly the missing amount. For every mode, letting $t \rightarrow \infty$ gives

$$c_i^{(t)} \longrightarrow \frac{\alpha}{1 - \beta\mu_i} a_i = \frac{\alpha}{\alpha + \beta\tilde{\nu}_i} a_i.$$

The zero-frequency component is unchanged, while each other initial component retains a strictly positive fraction of its coefficient.